Locality Preserving Projection-based Learning Methods for Human-Robot Collaborative Tasks

Thesis submitted by

Saibal Ghosh

Doctor of Philosophy (Engineering)

Department of Electrical Engineering,
Faculty Council of Engineering & Technology
Jadavpur University
Kolkata, India

2025

JADAVPUR UNIVERSITY
FACULTY OF ENGINEERING AND TECHNOLOGY

INDEX NO. **285/21/E** of **2021**

1. **Title of the Thesis:** "Locality Preserving Projection-based Learning Methods for Human-Robot Collaborative Tasks"

2. **Name, Designation, and Institution of the Supervisor/s:**

   a) **Prof. (Dr.) Amitava Chatterjee**
   Full Professor, Department of Electrical Engineering
   Jadavpur University, Kolkata–700032

   b) **Prof. Sugata Munshi**
   Retired Professor, Department of Electrical Engineering
   Jadavpur University, Kolkata–700032

3. **List of Publications:**

   a) **International Journal Papers (Published)**

      i. **Saibal Ghosh**, Pritam Paral, Amitava Chatterjee, and Sugata Munshi, "**Histogram refined local ternary pattern-based bilateral LPP for vision sensor-based robot navigation guidance under challenging environments**," *IEEE Sensors Letters*, vol. 7, no. 6, pp. 1-4, Jun. 2023, DOI: 10.1109/LSENS.2023.3272832. (Current IF: **2.2**).

      ii. **Saibal Ghosh**, Amitava Chatterjee, and Sugata Munshi, "**Visual cue-aided human supervised robot navigation guidance in photometrically challenging environments using adaptive spatial-feature kernel-guided bilateral LPP**," *Measurement Science and Technology*, vol. 34, no. 10, 105404, Jul. 2023, DOI: 10.1088/1361-6501/ace2dd. (Current IF: **3.4**).

      iii. **Saibal Ghosh**, Pritam Paral, Amitava Chatterjee, and Sugata Munshi, "**Rough entropy-based fused granular features in 2-D locality preserving projections for high-dimensional vision sensor data**," *IEEE Sensors Journal*, vol. 23,

Saibal Ghosh
10/11/2025

Amitava Chatterjee 10/11/25
Professor
Electrical Engg. Deptt.
Jadavpur University
Kolkata - 700 032

Sugata Munshi 10.11.2025
Retd. Professor
Electrical Engg. Deptt.
Jadavpur University
Kolkata - 700 032

no. 16, pp. 18374-18383, Aug. 2023, DOI: 10.1109/JSEN.2023.3288113. (Current IF: **4.5**).

iv. **Saibal Ghosh**, Pritam Paral, Amitava Chatterjee, and Sugata Munshi, "**A novel 2-D robust LPP-based approach using density-based neighborhood granulation for challenging visual cue detection**," *IEEE Sensors Journal*, vol. 25, no. 5, pp. 8665-8673, Mar. 2025, DOI: 10.1109/JSEN.2025.3529208. (Current IF: **4.5**).

b) **International Journal Papers (Communicated)**

i. **Saibal Ghosh**, Amitava Chatterjee, and Sugata Munshi, "**UaBMA-OLPP: A novel manifold learning technique for sEMG-based hand movement and object grasp recognition**," (Under Review).

ii. **Saibal Ghosh**, Amitava Chatterjee, and Sugata Munshi, "**sEMG-based hand movement recognition with robust possibilistic neighborhood graph-aided orthogonal LPP**," (Under Review).

4. **List of Patents: Nil**

5. **List of Presentations in National/International/Conferences/Workshops:**

a) **International Conference Papers (Presented and Published)**

i. **Saibal Ghosh**, Amitava Chatterjee, Sugata Munshi, and Anjan Rakshit, "**Flag-based human supervised robot guidance using locality preserving projections**," in *Proc. 2nd International Conference on Emerging Frontiers in Electrical and Electronic Technologies (ICEFEET)*, IEEE, Jun. 2022, pp. 1-6, DOI: 10.1109/ICEFEET51821.2022.9848242.

ii. **Saibal Ghosh**, Amitava Chatterjee, and Sugata Munshi, "**Local tetra pattern-based bilateral LPP in robot navigation guidance systems: A robust approach against vision sensor noises**," in *Proc. 6th International Conference on Condition Assessment Techniques in Electrical Systems (CATCON)*, IEEE, Dec. 2022, pp. 322-327, DOI: 10.1109/CATCON56237.2022.10077630.

iii. **Saibal Ghosh**, Pritam Paral, Pubali De, Amitava Chatterjee, and Sugata Munshi, "**Multi-color multi-shape visual cue recognition using a hybrid approach of cosine similarity-based 2DLPP and granular computing**," in *2nd*

Amitava Chatterjee 10/11/2025 Sugata Munshi 10.11.2025

Saibal Ghosh
10/11/2025

Professor
Electrical Engg. Deptt.
Jadavpur University
Kolkata - 700 032

Retd Professor
Electrical Engg. Deptt.
Jadavpur University
Kolkata - 700 032

International Conference on Recent Advances in Artificial Intelligence & Smart Applications (RAAISA), Lecture Notes in Networks and Systems, vol. 1380, Springer, Aug. 2025, pp. 469-482, DOI:10.1007/978-981-96-5822-0_37.

Saibal Ghosh
10/11/2025

Amitava Chatterjee
10/11/2025
Professor
Electrical Engg. Deptt.
Jadavpur University
Kolkata - 700 032

Sugata Munshi
10.11.2025
Retd Professor
Electrical Engg. Deptt.
Jadavpur University
Kolkata - 700 032

JADAVPUR UNIVERSITY
FACULTY OF ENGINEERING AND TECHNOLOGY

STATEMENT OF ORIGINALITY

I, **Shri Saibal Ghosh**, registered for the Ph.D. (Engineering) degree on 28th July 2021, do hereby declare that this Thesis entitled **"Locality Preserving Projection-based Learning Methods for Human-Robot Collaborative Tasks"** contains literature survey and original research work done by the undersigned candidate as part of doctoral studies.

All information in this Thesis has been obtained and presented in accordance with existing academic rules and ethical conduct. I declare that, as required by these rules and conduct, I have fully cited and referred all materials and results that are not original to this work.

I also declare that I have checked this Thesis as per the Policy on Anti Plagiarism, Jadavpur University, 2019, and the level of similarity as checked by iThenticate software is 2 %.

Signature of the Candidate: Saibal Ghosh

Date: 10/11/2025

Certified by Supervisor(s):

(Signature with date, seal)

1. Amitava Chatterjee 10/11/2025

(Prof. Amitava Chatterjee)

Professor
Electrical Engg. Deptt.
Jadavpur University
Kolkata - 700 032

2. Sugata Munshi 10.11.2025

(Prof. Sugata Munshi)

Retd Professor
Electrical Engg. Deptt.
Jadavpur University
Kolkata - 700 032

JADAVPUR UNIVERSITY
FACULTY OF ENGINEERING AND TECHNOLOGY

CERTIFICATE FROM THE SUPERVISOR/S

This is to certify that the Thesis entitled **"Locality Preserving Projection-based Learning Methods for Human-Robot Collaborative Tasks"** submitted by **Shri Saibal Ghosh**, who got his name registered on 28th July 2021 for the award of **Ph.D. (Engineering)** degree of Jadavpur University, is absolutely based upon his own work under the supervision of **Prof. Amitava Chatterjee** and **Prof. Sugata Munshi** and that neither his Thesis nor any part of the Thesis has been submitted for any degree/diploma or any other academic award anywhere before.

1. ..... 10/11/2025

(Prof. Amitava Chatterjee)

Signature of the Supervisor

and date with Official seal

Professor
Electrical Engg. Deptt.
Jadavpur University
Kolkata - 700 032

2. ..... 10.11.2025

(Prof. Sugata Munshi)

Signature of the Co-Supervisor

and date with Official seal

Retd Professor
Electrical Engg. Deptt.
Jadavpur University
Kolkata - 700 032

Acknowledgments

"I am a part of all that I have met;
Yet all experience is an arch wherethro'
Gleams that untravell'd world"
— **Alfred**, **Lord Tennyson**, *Ulysses*

This Thesis has been far more than an academic endeavor; it has been a long and evolving passage through moments of challenge, reflection, growth, and perseverance. Along the way, the experience has been interwoven with both joy and difficulty, much like a winding path marked by both sunlight and shadow. It is difficult to capture in words the depth of gratitude I feel for those who, knowingly or unknowingly, helped me navigate this journey. Acknowledging every contribution is no easy task, for the support came in many forms; through guidance, patience, encouragement, debate, and sometimes simply through presence. I am fully aware that in expressing thanks, I may fall short of conveying the full weight of what I owe to others. This work is the result of many hands and minds coming together; yet, *any shortcomings in it are mine alone*.

Among all the acknowledgments, this one holds the deepest significance in my heart. In the first place, I want to thank **My Parents—Mr. Mitan Kumar Ghosh** and **Mrs. Mallika Ghosh**, whose unwavering faith, silent strength, and selfless sacrifices have been the cornerstone of my pursuit of a research career. It was their support that gave me the courage to step away from earlier professional commitments and dedicate myself entirely to this long academic path. Their constant presence, emotional encouragement, and willingness to shoulder countless responsibilities at home; ranging from daily chores to managing personal and family matters; allowed me the time, focus, and peace of mind needed to carry out this work. Without their love, patience, and unspoken acts of care, this Thesis would not have been possible.

At the very beginning, I wish to extend my heartfelt gratitude to my supervisors, **Prof. Amitava Chatterjee** and **Prof. Sugata Munshi**, for their unwavering trust in my potential

and for offering me the opportunity to pursue my doctoral research under their esteemed guidance. From the earliest discussions of my research ideas, they welcomed me with openness and encouraged me to explore my academic interests with independence and confidence. Throughout the course of my Ph.D., their insightful feedback, constructive criticism, and continual encouragement were invaluable. They not only nurtured my growth as a researcher but also supported my academic pursuits, including participation in conferences and scholarly forums. Their steady guidance and understanding were especially reassuring during the more difficult phases of my journey. This work would not have come to fruition without their constant support and mentorship.

I also wish to express my deepest gratitude to the late **Prof. Anjan Rakshit**, whose early mentorship played a pivotal role in shaping the foundation of this research. Although he has not been my formal supervisor, his guidance during the initial stages; particularly in the development and conceptualization of the hardware-based systems, was instrumental. His insights, encouragement, and willingness to share his expertise provided the technical grounding and direction that this work ultimately built upon. The time and thought he invested in those formative discussions have had a lasting impact, and without his support during those crucial early phases, this Thesis would not have taken the shape it has today.

Beyond my supervisors, I would like to extend my sincere thanks to all the members of my doctoral Research Advisory Committee (RAC), including **Prof. Madhubanti Maitra**, **Head of the Department, Electrical Engineering, Jadavpur University**, and **Prof. Debangshu Dey**. I am also grateful to the faculty members of the **Instrumentation & Cyber Physical System** section (formerly known as **Measurement & Instrumentation**) of the Department of Electrical Engineering, Jadavpur University, including **Prof. Palash Kumar Kundu**, **Prof. Biswajit Bhattacharyya**, **Prof. Mita Dutta**, and **Prof. Gautam Sarkar**, for their academic support and encouragement throughout my research journey. Their thoughtful feedback, probing questions, and constructive observations helped me critically evaluate my work from multiple perspectives. Their involvement not only enriched the quality of my research but also guided me in refining various aspects of this Thesis through careful revision and reflection.

Additionally, I would also like to gratefully acknowledge the support of the former Heads of the Department of Electrical Engineering, Jadavpur University, during my Ph.D. tenure—**Dr. Kesab Bhattacharya**, **Dr. Saswati Mazumdar**, and **Dr. Biswanath Roy**. Their continuous encouragement and administrative support greatly facilitated the smooth progress

of my research work. I would also like to extend my sincere thanks to all the **Current and Retired Faculty Members** of the Department of Electrical Engineering, Jadavpur University, many of whom I had the opportunity to interact with during the course of my Ph.D. journey. Their constant support, thoughtful guidance, career advice, and occasional mentorship have been immensely valuable; both in shaping my research approach and in navigating the broader academic landscape. Their words of encouragement and willingness to engage in meaningful discussions have left a lasting impact on my professional growth.

I am sincerely thankful to the **All India Council for Technical Education (AICTE)**, **Ministry of Education** (formerly known as the **Ministry of Human Resource Development**), **Government of India**, for granting me the opportunity to pursue my doctoral research under the prestigious "AICTE Doctoral Fellowship (ADF)" scheme. The support received through this initiative extended far beyond the monthly fellowship; covering essential research-related expenditures, including publication fees for extended-length journal articles, conference registration charges for international forums, and several other academic and Thesis-related expenditures. This comprehensive backing significantly facilitated the smooth execution and dissemination of my research work.

I am deeply grateful to the **Hon'ble Vice-Chancellor**, **Registrar**, **Pro-Vice-Chancellor** (who also serves as the **ADF Nodal Officer**), and **Dean of Faculty of Engineering and Technology (FET)** of Jadavpur University, as well as to the faculty members who have previously served or are currently serving as **ADF Coordinators**. Their continuous support and timely intervention in resolving various issues related to fellowship disbursement and other administrative matters during my Ph.D. have been of immense help, ensuring that I could focus on my research without unnecessary obstacles.

During my Ph.D. tenure, I had the opportunity to visit several reputed academic and research institutions, including **NIT Durgapur**, **CSIR-CMERI Durgapur**, **IIT Kharagpur**, and **IIT Roorkee**; for conferences, workshops, and academic interactions. These visits proved to be intellectually enriching and personally inspiring. I had the privilege of meeting **Esteemed Faculty Members**, **Distinguished National and International Guests**, **Fellow Ph.D. Researchers**, **Scientists**, and other **Academic Staff**. Their valuable insights, thoughtful discussions, and generous sharing of knowledge significantly broadened my perspective, helped refine my understanding of various research areas, and motivated me to strive for academic excellence. I sincerely thank all of them for their time, openness, and meaningful academic

exchanges that contributed to my growth as a researcher.

The **Instrumentation & Cyber Physical System** Laboratory of the Electrical Engineering Department at Jadavpur University became a second home to me over the course of my Ph.D. journey. The entire ecosystem, comprising faculty members, support staff, and fellow researchers, offered not just an intellectually stimulating environment but also a sense of belonging and emotional support that helped me grow as both a researcher and an individual. I remain especially grateful to my fellow lab members for the countless hours of collaborative work, thought-provoking idea exchanges, and shared challenges. First of all, I would like to express my heartfelt gratitude to **Dr. Pritam Paral**, my former labmate who completed his Ph.D. from our laboratory, for his unwavering support and guidance throughout my doctoral journey. Beyond offering valuable career advice and thoughtful suggestions during crucial decisions, he actively contributed to various aspects of my research work, providing technical support, helping troubleshoot challenges, and offering constructive feedback on key ideas. His involvement made a significant difference in shaping the direction and quality of my work. I would like to extend my heartfelt thanks to my current labmates, especially **Mrs. Antara Ghosh** and **Mr. Anadi Biswas**, for standing by my side over the years and offering their unwavering support during difficult times. I am also sincerely grateful to my other labmates—**Ms. Samriddhi Maulik**, **Mr. Soumik Ghosh**, **Mr. Furkan Habib**, **Mr. Jayanta Digar**, and **Mr. Niladri Dutta**, for their constant encouragement and cooperative spirit throughout this journey.

I am also thankful to the former Ph.D. scholars of our lab, who were my early labmates—**Mrs. Rashmi Rekha Sahoo**, **Dr. Pubali De**, **Dr. Saptarshi Chatterjee**, **Dr. Sayanti Chaudhuri**, **Dr. Biswarup Ganguly**, and **Dr. Amitava Halder**, for their continued support, mentorship, and for creating a collaborative and intellectually enriching environment that played a key role in shaping my research journey. Their experience, encouragement, and willingness to help at every stage made the lab a welcoming and inspiring place to grow.

I would also like to thank the former postgraduate students from our lab, particularly **Mr. Aninda Sundar Mondal** and **Mr. Bisal Sarkar**, with whom I had the opportunity to collaborate closely on various research activities. Their dedicated efforts and teamwork significantly contributed to several research works within and beyond the scope of this Thesis. I am also grateful to other postgraduate students who had been my earlier labmates, such as—**Mr. Debarshi Brahma**, **Mr. Ashwin Rai**, **Mr. Barundeb Mondal**, **Ms. Susmita Bhattacharyya**,

Mr. Debasmit Dey, and **Mr. Rohan Das**, for their companionship, enthusiasm, and support during the formative years of my research.

I would like to sincerely acknowledge the support of the technical staff and non-teaching members associated with the **Instrumentation & Cyber Physical System** laboratory, namely **Mr. Debabrata Mondal**, **Mr. Santanu Biswas**, **Mr. Subhajit Paul**, and **Mr. Biswanath Adhikari**. Their friendly attitude and consistent assistance made it easier for me to settle into the laboratory environment during the early phase of my Ph.D. Whether it was arranging essential resources for experiments, extending a helping hand when unexpected issues arose, or volunteering for experimental studies as a subject, their cooperation was always prompt and reassuring, and I remain deeply grateful for their presence throughout my research journey. I am also thankful to the **Non-Teaching and Technical Staff** members from other sections of the Department of Electrical Engineering, Jadavpur University, whose occasional assistance and cooperative attitude have been helpful in various academic and laboratory-related matters.

I wish to express my thankfulness to the **Academic Support Staff**, **Administrative Personnel**, and **Library Staff** from the Department of Electrical Engineering, **Staff from the Administrative Office: Aurobindo Bhavan** of Jadavpur University, **Staff from the Research Section** of Jadavpur University, and **Staff from the Faculty of Engineering and Technology (FET) Office**, whose continuous support in handling academic, research, and administrative matters ensured the smooth progress of my work over the years. In addition, I am grateful to the members of the **Research Scholars' Organizations** of Jadavpur University for their ongoing support and advocacy on various academic and administrative issues that arose during the course of my Ph.D. Their efforts in representing and addressing the concerns of research scholars made a significant difference in navigating institutional processes more smoothly.

A special note of appreciation goes to all my **Fellow Ph.D. Peers** from this University—across various laboratories and departments—whose camaraderie, shared experiences, and open exchange of interdisciplinary ideas have been both motivating and intellectually stimulating. Engaging with them on topics ranging from technical tools and paper writing strategies to broader research discussions has been an invaluable part of my learning process. Together, we have exchanged thoughts and ideas not only confined to this Thesis but extending well beyond it; some of which directly contributed to shaping the direction and refinement of this work, while many others broadened my perspective and helped me grow both intellectually and personally. From late-night conversations about biology and space science

to passionate debates on sports, art, cinema, or politics, these interactions formed a vibrant and enriching backdrop to the more structured academic journey. Many ideas born out of these exchanges remain to be nurtured and explored in the future. This dynamic community of thinkers has played a fundamental role in shaping my journey, and I am deeply thankful for the inspiration, clarity, and companionship they have provided throughout. I also wish to extend my appreciation to all my other **Co-researchers** and **Technical Collaborators**, both within and outside the scope of this Thesis, with whom I had the privilege to work and co-author research publications. Their contributions, insights, and teamwork have played a meaningful role in my academic journey.

I am thankful to all the **Undergraduate Students** of Jadavpur University and other external institutions who undertook their project work under my guidance and in collaboration with me. Mentoring them was a truly enriching experience; one that not only allowed me to share my knowledge but also gave me the opportunity to learn from their fresh perspectives and enthusiasm. I am also grateful to the **Undergraduate Students** whom I had the privilege of teaching in both theory and laboratory classes during my Ph.D. tenure. Their critical thinking, inquisitive questions, and thoughtful feedback not only challenged me to communicate more clearly but also enriched my own understanding and deepened my engagement with the subject matter. I am sincerely grateful to all the **Individual Volunteers** who generously participated as subjects in my experimental studies throughout the course of my Ph.D.; their participation was invaluable to the success of my work.

I am also deeply thankful to all my other **Friends**, beloved **Juniors**, and **Seniors** from the Jadavpur University campus who remained connected with me over the years. Their continued support, warm presence, and timely words of encouragement during difficult times have meant more than words can express. Their friendship has provided strength and comfort through every high and low of this journey. Additionally, I gratefully acknowledge the silent yet essential contributions of the many individuals across campus and its peripherals whose work supported my day-to-day life; be it the ever-helpful **Security Guards**, **Canteen Staff**, **Internet and Computer Support Person**, **Photocopy Shop Operators**, other **Maintenance Personnel**, or those who kept the campus running smoothly in the background. Their everyday efforts created a welcoming and functional environment that made it possible to focus on research without hassle.

I would also like to take this opportunity to thank the **Familiar Faces from My Local**

Neighborhood, including my current and former **Roommates** at my residence in Jadavpur. Their presence has made day-to-day life outside the academic environment much smoother and more comforting. Whether it was arranging daily meals, assisting with housekeeping needs, or simply helping with small but essential tasks, the support from the surrounding community and local shopkeepers played a silent yet significant role in maintaining balance during my Ph.D. journey. I am especially grateful for the moments of light-hearted conversation and relaxation shared with **Local Friends**; it was these refreshing interactions that often brought much-needed relief at the end of demanding days.

I am also grateful to my extended family i.e., **Relatives and Family Friends**, who stood by me with their good wishes, encouragement, and practical help whenever needed. Their thoughtful gestures, timely words of reassurance, and understanding during busy or difficult phases provided comfort and strength. The warmth and support of this wider circle of family have played an essential, though often quiet, role in sustaining me through the years of this journey.

I would also like to express my special thanks to the **Childhood Friends** from my hometown neighborhood, whose constant support, concern, and long-standing bond have always brought comfort and strength, both near and afar. I am equally thankful to the **Friends from My Undergraduate Days**, with whom I have shared countless memories, formative experiences, and unwavering friendships that continue to uplift me. Additionally, I extend my sincere gratitude to all my **Other Friends and Well-wishers** I have met along the way over the years; each of whom, in their own way, has offered encouragement, motivation, or a kind word during moments when I needed it most. A special mention goes to my **Travel Companions**; friends with whom I grew closely connected through numerous travels, treks, and shared journeys. Experiencing the highs and lows of the road together not only brought us closer in those moments but also created a deeper bond that carried over into the ups and downs of life itself. These shared adventures, journeys, and life's many triumphs and trials experienced with all of them, have forged lasting connections that I continue to cherish deeply. Their presence in my life has been a source of resilience and joy throughout this journey.

With a heavy heart, I wish to remember my beloved pet dog, **Chhotu**, who left us last year. She had incidentally come into my life when she was just a tiny pup and soon found a warm shelter in our home, becoming an inseparable part of the family. Her unconditional love, quiet companionship, and intuitive understanding brought immense comfort during some of the

difficult phases of my Ph.D. journey. There were countless moments when I shared my thoughts and emotions with her, and though she could not speak, her eyes always reflected empathy and unwavering presence. I miss her dearly and hope that, from her heavenly abode, she continues to shower her love and blessings on us as she always did.

At last, I would like to pay my sincere tribute to this esteemed center of learning—my **Alma Mater, Jadavpur University**, Kolkata, India, which has been nothing less than a second home to me since my undergraduate days. My bond with this campus runs deep, shaped by years of learning, growth, and countless memories. Such is the connection that I could never stay away for long, choosing to pursue both my postgraduate and doctoral studies here. Even in the years to come, I hope to remain meaningfully associated with this remarkable institution in some capacity or form. I wholeheartedly wish Jadavpur University continued global recognition, ever-growing success, and the continued nurturing of brilliant minds in the years ahead.

I sincerely wish for the continued betterment of the overall **Research Ecosystem in India**; one that nurtures and empowers brilliant researchers, scientists, academicians, and educators. May it foster a culture of innovation, critical thinking, and academic excellence that not only contributes meaningfully to the nation but also creates a lasting impact on the global stage. I hope to see Indian research and education earn ever-greater international recognition in the years to come, as it continues to shape knowledge and inspire future generations.

Finally, I bow in humble gratitude to the **Almighty & Nature**; the silent force that governs all creation, offering strength in stillness, clarity in chaos, and balance in the journey. May I continue to remain grounded in its wisdom, guided by its rhythm, and nurtured by its infinite grace as I walk forward on the path of learning and contribution.

Saibal Ghosh 10/11/2025

Saibal Ghosh
Ph.D. Researcher,
AICTE Doctoral Fellow,
Department of Electrical Engineering,
Jadavpur University, Kolkata, India

Dedicated to all the Warriors on the battlefield.

Dedicated to all the Animals and Plants in the universe, who cannot speak, but feel — may there always be love and empathy for you.

> We are just an advanced breed of monkeys on a minor planet of a very average star. But we can understand the Universe. That makes us something very special.

..... **Stephen William Hawking**

⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆

The Road Not Taken

Two roads diverged in a yellow wood,
And sorry I could not travel both
And be one traveler, long I stood
And looked down one as far as I could
To where it bent in the undergrowth;

Then took the other, as just as fair,
And having perhaps the better claim,
Because it was grassy and wanted wear;
Though as for that the passing there
Had worn them really about the same,

And both that morning equally lay
In leaves no step had trodden black.
Oh, I kept the first for another day!
Yet knowing how way leads on to way,
I doubted if I should ever come back.

I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I—
I took the one less traveled by,

And that has made all the difference.

····· **Robert Frost**

⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆

Abstract

In recent years, the convergence of human-in-the-loop (HITL) systems and intelligent automation has led to a significant rise in the development of interactive systems where humans and machines engage in seamless communication and perform tasks in the same working space. Human-computer interaction (HCI), human-machine interface (HMI), and human-robot interaction (HRI) form the foundational pillars of this multidisciplinary advancement, each addressing distinct yet overlapping aspects of interaction. HCI traditionally focuses on how human users interact with computational systems through graphical, tactile, or voice-based modalities, facilitated by various sensing interfaces. HMI extends this paradigm to encompass a broader range of electromechanical systems, including industrial devices and wearable sensors. Whereas, HRI is a more specialized domain within this spectrum that explores the dynamic interplay between humans and robotic agents. It emphasizes bidirectional perception, shared control, and adaptive learning mechanisms to foster intuitive and responsive interaction in real-world scenarios.

As these paradigms mature and are deployed in working environments, *human-robot collaborative tasks* (HRCTs) gain greater prominence, particularly in the domains requiring physical and cognitive cooperation between humans and robots. In such collaborative scenarios, robots are no longer mere tools for isolated tasks but are intelligent partners capable of interpreting human intent and responding to nuanced environmental cues. This advancement requires robust perception modules, context-aware decision-making, and mutual adaptability to manage uncertainty and ensure safety under various complex environments. The integration of diverse sensor modalities that capture physiological and behavioral human cues offers promising avenues for bridging the gap between human intention and robotic execution. Such meaningful interaction facilitates shared autonomy through various learning tasks for an assistive robot in applications such as rehabilitation, assistive manipulation, navigation, and various other collaborative operations in regular and unstructured environments.

In the context of human-robot interaction (HRI), the choice of sensor modalities plays a very crucial role in interpreting human cues. It enables responsive collaboration between human users and robotic agents and ensures safety in shared environments. Sensor modalities can be broadly classified into two main categories i.e., contact-based and non-contact-based, each offering specific advantages depending on the working environment. Contact-based or wearable sensors, such as data gloves, magnetic trackers, surface electromyography (sEMG), inertial measurement units (IMUs), accelerometers, etc., require physical attachment to the user's body to detect human intent in the form of muscle activity, kinematic movement parameters, applied force/torque, etc. These provide rich physiological information that is crucial for precise recognition of hand gestures, hand movement, hand activity, and other motion intent estimation. These sensors can provide high signal fidelity and better temporal resolution, making them particularly effective for capturing muscle activity, joint dynamics, and force-related human cues. Such cues are critical for tasks involving fine motor control such as prosthetic manipulation, assistive rehabilitation, health monitoring, gesture recognition, etc. However, these modalities require precise placement, good connectivity, and stability during data acquisition, and may cause user discomfort during prolonged and continuous use. Moreover, these are also susceptible to motion artifacts, electrode displacement, and signal drift over time, which can affect the data quality. Additionally, some systems may require calibration or recalibration across sessions or users due to inter-session and inter-subject variability, increasing complexity in the setup and handling. The need for direct skin contact, especially in bio-signal sensors like sEMG, can lead to skin irritation and introduce variability in skin-electrode contact resistance, especially during long-term deployments.

On the contrary, non-contact sensors, such as RGB cameras, infrared thermographic imagers, or depth sensors can acquire human intent information from a distant place without direct physical contact. These modalities enable vision-based cue detection, body posture analysis, human action recognition, and gesture tracking, making them particularly advantageous in scenarios where minimum physical interference and user comfort are desired. These are especially essential in scenarios such as remote monitoring, robot guidance in dynamic environments, teleoperation in hazardous environments, activity recognition in ambient assisted living, and interaction support in socially assistive robotics. While non-contact systems offer flexibility and ease of deployment, they are susceptible to environmental variations like lighting conditions, occlusions, or background clutter, which can degrade the

recognition accuracy.

This Thesis investigates both non-contact approaches, such as vision sensing for visual cue interpretation in robot navigation guidance and assistive robotics environments, and contact-based modalities, such as sEMG for muscular activation analysis in prosthetic control and rehabilitation strategies, across separate studies. These complementary explorations highlight how different sensing strategies can be effectively designed for specific human-robot interaction or human-machine interface contexts, contributing to the development of robust and application-specific HRI/HMI frameworks suitable for deployment in real-world environments.

Despite the advancements in sensor technologies and data acquisition systems for human-robot interaction, several practical issues are encountered in real-world deployments. One of the foremost issues is sensor-generated and environmental noise, which can significantly impair the signal or image quality and deteriorate their subsequent processing. Vision-based systems are particularly vulnerable to photometric irregularities such as lighting conditions, contrast variations, and background clutter, all of which can disrupt the consistency in visual cue detection. Similarly, wearable and contact-based sensors such as sEMG, IMU, and accelerometer are prone to motion artifacts, crosstalk, sensor noise, and electrode displacement. These interferences often introduce random fluctuations or irregular distortions that reduce the reliability of feature extraction, especially during prolonged sessions or dynamic task execution in unstructured environments. Another critical challenge lies in intra-subject and inter-subject variability. Visual patterns, physiological signals, and behavioral expressions of intent can vary widely across individuals due to anatomical, neuromuscular, and habitual differences. Visual cues may be inconsistently performed or perceived due to human variability and contextual differences across social and geographical boundaries. With wearable sensors, even within the same subject, variations may occur across sessions due to muscle fatigue, inconsistent sensor placement, or deviated motion dynamics. This inherent diversity in data characteristics makes the development of generalizable models challenging and necessitates adaptive or personalized strategies. Additionally, background clutter in the environment can significantly affect the performance of non-contact modalities like vision sensors. Complex or dynamic backgrounds make it challenging to isolate relevant visual features, such as hand gestures or body posture, which are critical for accurate interaction modeling. This affects the robustness of visual feature extraction and can lead to higher false positives or recognition failures in cluttered or real-world environments.

A further concern is the high dimensionality of sensor data, particularly when dealing with large-sized image frames or rich signals such as EMG waveforms. Processing such data in real-time is computationally expensive and often impractical for embedded or robotic platforms. This highlights the need for effective dimensionality reduction (DR) techniques that can retain the discriminative and structural information of the data while reducing the computational burden. The use of manifold learning techniques, such as those explored in this Thesis, provides a promising direction for balancing information preservation and computational efficiency. In particular, this Thesis thoroughly investigates the family of linear approximation-based manifold learning-inspired approaches, with a primary focus on the *locality preserving projection* (LPP), which effectively preserves the local structural information of the data in the projected subspace.

Moreover, a persistent challenge remains in the generalization of models across different application scenarios. Algorithms trained on controlled environments or synthetic datasets often underperform when exposed to unstructured environments with unseen conditions. This lack of knowledge transferability necessitates the development of adaptive learning strategies that can dynamically adjust to changing scenarios, sensor drifts, and user-specific variations without requiring exhaustive retraining and offering more robustness to the system. Eventually, real-world deployment of HRI systems must keep a balance between accuracy and computational latency. High recognition accuracy is essential for reliable interaction, but the requirement of computational resources must be sufficiently low to support the practical deployment of such algorithms on embedded systems. This trade-off becomes particularly crucial in scenarios involving real-world control, shared autonomy, or safety-critical tasks. In the context of developing countries, where access to high-performance computing resources may be limited, the need for cost-effective, energy-efficient, and resource-constrained solutions becomes even more pronounced. Consequently, designing models that are both efficient, reliable, and will be able to work well without relying on costly hardware remains a major challenge in building practical and widely usable interaction strategies for HRI systems.

As mentioned above, dimensionality reduction techniques play a crucial role in simplifying the input representations by projecting the original high-dimensional data onto lower-dimensional subspaces. They need to do so while preserving the essential features of the data, required for interpretation and classification in the reduced space. Effective dimensionality reduction not only enhances computational efficiency but also improves model

generalizability and robustness by filtering out redundant or irrelevant information from the data. Traditional global dimensionality reduction methods such as principal component analysis (PCA), linear discriminant analysis (LDA), and independent component analysis (ICA) have been widely used due to their ease of implementation and computational efficiency. However, these techniques often work under the assumption of global linearity and may fail to preserve the local geometric relationships inherent in complex sensor data. Being linear dimensionality reduction techniques, these techniques rely on the assumption that the underlying structure of the data can be effectively captured using linear transformations. However, sensor data in HRI applications such as vision-based cues or sEMG signals often exhibit nonlinear characteristics due to various factors like anatomical differences, complex motion dynamics, and environmental hazards. In such cases, linear techniques may be inadequate for uncovering the true latent structure of the data. This has led to the exploration of nonlinear dimensionality reduction approaches, particularly the family of manifold learning techniques such as isomap, locally linear embedding (LLE), and Laplacian eigenmaps (LE), which attempt to discover low-dimensional embeddings that preserve the intrinsic geometry of the data. However, many of these methods lack an explicit mapping function, making it difficult to handle new samples under real-world implementations. To address these challenges, this Thesis focuses on locality preserving projections (LPP), a DR technique that explores the nonlinear local structure of the data by linear approximation strategies, while enabling efficient computation and generalization to new data points. LPP captures the intrinsic local structure of the high-dimensional data by constructing a nearest-neighbor graph and projecting the data onto a subspace that preserves such local relationships. Unlike PCA or LDA, LPP does not seek global variance maximization or class separability alone, but rather emphasizes preserving intrinsic neighborhood proximity, making it highly effective in scenarios involving dynamic motion cues, irregular visual cues, sensor disturbances, and human variability. Its ability to produce an explicit linear mapping and enable straightforward projection for the new test samples are critical for real-world deployment in embedded systems.

While the locality preserving projection offers clear advantages in preserving local geometric structures of high-dimensional data, several issues still persist with the traditional form of this technique, under irregular data conditions. One of the major concerns lies in the sensitivity of graph construction to noise and outliers. Traditional LPP relies on Euclidean distance-based similarity graphs, which can be heavily influenced by lighting variation and

background clutter in vision-based systems, or artifacts arising from sensor disturbances and motion dynamics. This reliance on fixed-distance kernels makes the resulting weight matrix vulnerable under non-ideal conditions, thereby reducing the discriminative power of the projected subspace. Another important aspect is that standard LPP-based approaches ignore the feature-specific importance of the data. Across various sensing modalities in HRI applications, not all the features contribute equally to human intent recognition or task discrimination. However, the conventional similarity graph used in LPP construction does not account for this relevance or discriminative capacity of individual features from the data. It focuses solely on spatial proximity between data samples, thereby discarding valuable information from the feature domain. As a result, the generated subspace may remain inadequate to emphasize semantically meaningful components of the data, especially in multimodal sensor data with heterogeneous characteristics. Additionally, the similarity matrix in standard LPP is restricted to Euclidean geometry, which assumes that the underlying samples lie on a flat, linear manifold. In practical scenarios, such as with biological signals e.g., sEMG, or with natural visual scenes, the data often resides on complex, nonlinear manifolds (e.g., Riemannian). The Euclidean assumption thus limits LPP's ability to fully encode spatial and contextual relationships among samples, leading to suboptimal embedding, especially in cases of data corruption. Encoding of nonlinear data structures requires models that are capable of adapting to curvature and topology in the data space. Furthermore, the use of inflexible similarity matrices, such as fixed Gaussian kernels, restricts LPP's adaptability across different sensing conditions and modalities. A fixed kernel may not effectively model relationships across varying sensor anomalies, noise levels, or inter-class/intra-class separations. The absence of adaptive distance modeling mechanisms limits LPP's scalability in heterogeneous datasets, such as those containing new data samples with previously unseen corruption. This creates a strong necessity for context-aware variants of LPP that can dynamically modulate the similarity function based on the statistical and geometric properties of the input data.

Taken together, these limitations motivate the need for robust, adaptive, and semantically aware extensions of LPP that can handle the diverse and noisy nature of HRI data. Several variants are developed in this Thesis work by incorporating adaptive spatial kernels, granular computing-aided kernel fusion, noise-resilient distance metrics (e.g., Euler, Grassmannian), uncertainty-aware similarity fusion, sparsity-inducing regularizations, etc. These adaptations not only enhance class separability and subspace stability but also ensure that the learned

embeddings remain interpretable and computationally viable for deployment in real-world, embedded platforms inside HRI frameworks. The entire work can be broadly perceived in two research verticals i.e., (i) LPP-based strategies for vision sensor-based modalities and (ii) LPP-based strategies for wearable sensor-based modalities. The augmented LPP variants from both these research verticals majorly focus on offering more accurate, reliable, robust, and computationally inexpensive solutions. Their performance was extensively evaluated across both vision and wearable sensor modalities, and their performance outcomes are summarized below.

The proposed LPP-based frameworks demonstrated notable performance across both vision and wearable sensor modalities. For vision-based symbolic cue data, the conventional LPP achieved about 99% accuracy under normal illumination but deteriorated to nearly 55% under the darkest condition. The proposed ALPSK-BLPP framework improved this to approximately 71%, while granular computing-based extensions such as REGF-2DLPP and dNG-2DRLPP further enhanced performance up to 81% and 85%, respectively. Under noisy conditions, the proposed LTrP-BLPP algorithm achieved around 85% accuracy at 35% salt-and-pepper noise density and 89% under speckle noise with $variance = 0.35$, outperforming the traditional LPP by $12 - 15\%$ accuracy margins. Similarly, variants like HRLTP-BLPP maintained about $80 - 84\%$ accuracy under simultaneous illumination degradation (dark level-2) and noise corruption (either of salt-and-pepper or speckle noise), and dNG-2DRLPP achieved $\sim 82\%$ under Gaussian noise ($variance = 0.35$). For wearable sEMG data, the proposed UaBMA-OLPP and RPNG-OLPP models respectively achieved around 88% and 89% recognition accuracies, showing higher resilience than the classical LPP ($\sim 81\%$) framework. The robust RPNG-OLPP variant maintained $77 - 84\%$ accuracy even under noisy conditions for the sEMG dataset.

Contents

List of Figures

List of Tables

List of Abbreviations

{Numeric}

1D	**O**ne-**D**imensional
2D	**T**wo-**D**imensional
2DLDA	**T**wo-**D**imensional **LDA**
2DLPP	**T**wo-**D**imensional **LPP**
2DLPP-L1	ℓ_1-norm-based **2DLPP**
2DPCA	**T**wo-**D**imensional **PCA**
2DRLPP	**T**wo-**D**imensional **R**obust **LPP**
3D	**T**hree-**D**imensional

{A}

ADC	**A**nalog-to-**D**igital **C**onverter
AI	**A**rtificial **I**ntelligence
ALPR	**A**daptive **L**ocality **P**reserving **R**egression
ALPSK	**A**daptive **L**ocality **P**reserving **S**patial **K**ernel
ANG	**A**daptive **N**eighborhood **G**raph
ANN	**A**rtificial **N**eural **N**etwork
AUC	**A**rea **U**nder the **C**urve

{B}

BEM	**B**inarized **E**igenvalue **M**ap
BLE	**B**luetooth **L**ow **E**nergy
BLPP	**B**ilateral **L**ocality **P**reserving **P**rojection

BMA **B**ayesian **M**odel **A**veraging

{C}

CCA **C**anonical **C**orrelation **A**nalysis

CCC **C**anonical **C**orrelation **C**oefficient

CDC **C**olor-**D**istance **C**riterion

CG **C**risp **G**ranulation

CP **C**enter **P**ixel

{D}

DC **D**irect **C**urrent

DBI **D**avies–**B**ouldin **I**ndex

DL **D**eep **L**earning

DLPP **D**iscriminant **L**ocality **P**reserving **P**rojection

DR **D**imensionality **R**eduction

DWT **D**iscrete **W**avelet **T**ransform

{E}

ECRB **E**xtensor **C**arpi **R**adialis **B**revis

ECRL **E**xtensor **C**arpi **R**adialis **L**ongus

ECU **E**xtensor **C**arpi **U**lnaris

ED **E**xtensor **D**igitorum

EEG **E**lectro**e**ncephalo**g**raphy

EER **E**qual **E**rror **R**ate

EMD **E**mpirical **M**ode **D**ecomposition

EMG **E**lectro**m**yo**g**raphy

{F}

FA **F**actor **A**nalysis

FAR **F**alse **A**cceptance **R**ate

FCU **F**lexor **C**arpi **U**lnaris

FCR **F**lexor **C**arpi **R**adialis

FFT **F**ast **F**ourier **T**ransform

FN **F**alse **N**egative

FNR **F**alse **N**egative **R**ate

FP **F**alse **P**ositive

FPGA **F**ield-**P**rogrammable **G**ate **A**rray

FPR **F**alse **P**ositive **R**ate

FREAK **F**ast **R**etina **K**eypoint

FrFT **F**ractional **F**ourier **T**ransform

{G}

GFD **G**ranular **F**eature **D**ecoding

GFRE **G**ranular **F**usion-based **R**ough **E**ntropy

GrC **G**ranular **C**omputing

{H}

HOG **H**istogram of **O**riented **G**radient

HCI **H**uman-**C**omputer **I**nteraction

HITL **H**uman-**i**n-**t**he-**L**oop

HRC **H**uman-**R**obot **C**ollaboration

HRI **H**uman-**R**obot **I**nteraction

HMI **H**uman-**M**achine **I**nterface

{I}

ICA **I**ndependent **C**omponent **A**nalysis

IDE **I**ntegrated **D**evelopment **E**nvironment

IEEE **I**nstitute of **E**lectrical and **E**lectronics **E**ngineers

IMF **I**ntrinsic **M**ode **F**unction

IMU **I**nertial **M**easurement **U**nit

Isomap **Iso**metric Feature **Map**ping

{K}

KLDA **K**ernel **L**inear **D**iscriminant **A**nalysis

KPCA **K**ernel **P**rincipal **C**omponent **A**nalysis

{L}

LBP **L**ocal **B**inary **P**attern

LDA **L**inear **D**iscriminant **A**nalysis

LE **L**aplacian **E**igenmap

LLE **L**ocally **L**inear **E**mbedding

LPP **L**ocality **P**reserving **P**rojection

LSP **L**ocal **S**kew **P**attern

LTP **L**ocal **T**ernary **P**attern

LTSA **L**ocal **T**angent **S**pace **A**lignment

{M}

MCC **M**atthews **C**orrelation **C**oefficient

MMG **M**echano**m**yo**g**raphy

MSLPP **M**ulti-**S**cale **L**ocality **P**reserving **P**rojection

MSVM **M**ulticlass **S**upport **V**ector **M**achine

{N}

NIL **N**ormal **I**llumination **L**evel

NMF **N**on-negative **M**atrix **F**actorization

NPE **N**eighborhood **P**reserving **E**mbedding

NPP **N**eighborhood **P**reserving **P**rojection

NPV **N**egative **P**redictive **V**alue

{O}

ODLPP **O**rthogonal **D**iscriminant **L**ocality **P**reserving **P**rojection

OLPP **O**rthogonal **L**ocality **P**reserving **P**rojection

OVA **O**ne-**v**s-**A**ll

OVO **O**ne-**v**s-**O**ne

{P}

PA **P**rincipal **A**xis

PC **P**ersonal **C**omputer

PCA **P**rincipal **C**omponent **A**nalysis

PNG **P**ossibilistic **N**eighborhood **G**raph

{Q}

QTD **Q**uad-**T**ree **D**ecomposition

{R}

RE **R**ough **E**ntropy

ReLSR **Re**targeted **L**east **S**quare **R**egression

RGB **R**ed, **G**reen, **B**lue

ROC **R**eceiver **O**perating **C**haracteristic

RPNG **R**obust **P**ossibilistic **N**eighborhood **G**raph

RTC **R**eal-**T**ime **C**lock

{S}

SCG **S**patio-**C**olor **G**ranule

SD **S**tandard **D**eviation

sEMG **S**urface **E**lectro**m**yo**g**raphy

SKLPP **S**upervised **K**ernel **L**ocality **P**reserving **P**rojection

SoC **S**ystem **o**n **C**hip

SRCC **S**pearman's **R**ank **C**orrelation **C**oefficient

SVM **S**upport **V**ector **M**achine

SVR	**S**upport **V**ector **R**egression
{T}	
TN	**T**rue **N**egative
t-SNE	**t**-Distributed **S**tochastic **N**eighbor **E**mbedding
TP	**T**rue **P**ositive
TPR	**T**rue **P**ositive **R**ate
{U}	
UDLPP	**U**ncorrelated **D**iscriminant **L**ocality **P**reserving **P**rojection
ULDA	**U**ncorrelated **L**inear **D**iscriminant **A**nalysis
{W}	
WLAN	**W**ireless **L**ocal **A**rea **N**etwork
{Z}	
ZC	**Z**ero **C**rossing

Chapter 1

Introduction

The field of human-robot collaboration (HRC) has witnessed rapid advancements over the past few decades, driven by the demand for intelligent and intuitive machines and robots that can operate seamlessly in dynamic and unstructured environments. From industrial co-working scenarios to assistive robotics for healthcare and rehabilitation, the integration of sensing, interpretation, and decision-making mechanisms has become a fundamental requirement in enabling robots to understand and respond to human cues effectively.

One of the key challenges in such collaborative environments lies in the accurate and robust interpretation of human inputs, acquired through diverse sensing modalities. These inputs may arise from vision-based cues, such as body pose, hand gestures, or semiotic visual markers captured through optical sensors, or from physiological signals acquired via contact-based modalities, such as surface electromyography (sEMG). However, both these sensory streams are inherently vulnerable to sensor noises and are subjected to various real-world disturbances; including poor illumination, occlusions, motion blur in visual data [1], signal variability and artifacts due to improper sensor placement, muscle fatigue, and inter/intra-subject differences in EMG data [2].

To address these issues, dimensionality reduction and feature extraction techniques have played a pivotal role in obtaining discriminative representations from high-dimensional raw sensor data. Classical global methods like principal component analysis (PCA) [3] and linear discriminant analysis (LDA) [4] often fail to capture the nonlinear structures embedded in such data. In contrast, nonlinear manifold learning techniques such as isometric feature mapping (Isomap) [5], Laplacian eigenmaps (LE) [6], and linearly approximated techniques e.g., locality preserving projection (LPP) [7] provide more appropriate alternatives by preserving neighborhood structures and local geometries during projection.

Recent extensions of LPP, such as uncorrelated discriminant LPP (UDLPP) [8], orthogonal LPP (OLPP) [9], fast orthogonal LPP (FOLPP) [10], bilateral LPP (BLPP) [11],

multi-scale LPP (MSLPP) [12], two-dimensional LPP (2DLPP) [13], and two-dimensional robust LPP (2DRLPP) [14], have further enhanced the performance against various environmental irregularities [15]. However, despite these advancements, several challenges remain unaddressed. Many existing variants still rely on spatial similarity measured via Euclidean distance alone, making them sensitive to outliers, occlusions, and structural noise in both vision and physiological data [11]. Furthermore, most formulations lack adaptability in integrating complementary feature domains (e.g., spatial and range), and hence do not effectively capture multi-channel information [16]. They also may fail to work with the same efficacy under varying lighting conditions, motion blur, and sensor aging [17]. In wearable sensing, such as sEMG-based applications, inter-subject variability, poor signal-to-noise ratio, and muscle fatigue effects continue to degrade performance [18], especially when classical graph construction strategies [7] are used without adaptive weighting or sparsity constraints. To address these limitations, this Thesis proposes several novel LPP-based frameworks designed specifically for robust subspace learning in real-world, diverse sensing environments. Collectively, these methods represent a significant advancement over traditional LPP approaches by integrating data-driven kernel construction, multi-channel fusion, sparsity-enforcing constraints, adaptive graph modeling, etc.

This Thesis explores the development and application of such LPP-based dimensionality reduction frameworks for two complementary yet distinct domains within HRC: (i) visual cue-driven robot navigation guidance using monocular vision sensors, and (ii) human intent recognition through hand activities using sEMG signals. By proposing novel variants of LPP and evaluating them in real-world experimental setups, the work aims to contribute toward more resilient and deployable perception systems for the next-generation collaborative robots.

1.1 Background and Motivation

Human-robot collaboration (HRC) represents a major shift in how autonomous systems and humans share, interact, and co-function within dynamic workspaces. Unlike traditional robotic systems confined to structured environments and repetitive tasks, collaborative robots (cobots) are being increasingly deployed in unstructured real-world scenarios; ranging from industrial premises and service robotics to assistive and rehabilitative applications. These applications

demand not only physical proximity and safety but also semantic understanding and adaptive interpretation of human behavior and intent [19].

The fundamental goal of HRC lies in enabling intuitive and natural interaction between humans and robots. This necessitates multimodal sensing frameworks that can reliably capture diverse forms of human input such as hand gestures, body posture, or physiological activity, and interpret them with low latency and high precision [20]. Vision-based sensing [21] and wearable/contact-based techniques [22] have emerged as two dominant categories in this regard, each offering unique benefits with unavoidable and distinct challenges. While visual sensors enable non-invasive, wide-field monitoring of gestures and movements, they are highly sensitive to environmental conditions like illumination or occlusion [23]. Conversely, physiological signals such as surface electromyography (sEMG) provide a direct means of understanding muscle activity and user intention but suffer from variability, noise, and user-dependent inconsistencies [24].

In such distinct modality systems, a critical computational requirement is the ability to extract robust and compact representations from high-dimensional, noisy, and redundant sensor data [25]. Dimensionality reduction techniques address this challenge by mapping raw sensor input to lower-dimensional subspaces that retain the most discriminative information. However, classical approaches such as PCA or LDA often fail to preserve the local, nonlinear structures intrinsic to real-world human-robot interaction data. This Thesis builds upon these challenges to develop robust DR techniques, particularly locality preserving projection (LPP) and its novel extensions, aimed at enhancing perception and decision-making pipelines in both vision-guided and EMG-driven collaborative robotic systems.

1.1.1 Human-Robot Collaboration in Real-World Environments

Human-robot collaboration (HRC) has rapidly evolved from isolated research verticals to widely deployed systems across industrial, military, domestic, and healthcare domains. Closely related to this is human-robot interaction (HRI), a broader field that investigates the communication, behavior, and decision-making dynamics between humans and robots. As the boundaries between human and robotic workspaces shrink, robots are no longer confined to restricted spaces or fixed trajectories. Rather, they are now expected to operate in close physical proximity with humans by interpreting their commands, adapting to uncertainties,

and ensuring safety in real-world setups. This transition has been driven by the growing demand for intelligent machines that can function seamlessly in unstructured and dynamic environments, performing tasks such as navigation assistance, object manipulation, assembly operation, heavy-duty jobs, and rehabilitative support [26].

Real-world HRC systems must accommodate a wide range of variability, including changes in lighting, occlusion, cluttered backgrounds, uncertain human behavior, senor placement issues, and sensor degradation. Unlike simulated environments, physical deployment often exposes systems to unpredictable photometric conditions, erratic signal artifacts, and hardware anomalies. Collaborative robotics in such contexts requires the integration of robust perception systems and adaptive decision-making strategies to interpret multimodal human cues with high reliability [27].

Vision-based human guidance is a widely adopted modality in HRC. It allows users to communicate their intent through body movements, hand gestures, hand activity, or symbolic markers. In particular, specially designed visual markers can be used as a viable means in assistive robot navigation tasks, where users provide easily recognizable, spatially distinct gestures for directional control. These visual cues are mapped to semantic commands via lookup tables, enabling real-time control of mobile robots in indoor and outdoor environments. Such approaches benefit from being contactless and intuitive, but their robustness is often challenged by low-cost sensor limitations, variable lighting conditions, and background interference [28].

Beyond vision, wearable physiological sensing methods, such as surface electromyography (sEMG) have enabled more nuanced forms of hand activity recognition, especially in applications involving prosthetics, rehabilitation, and human-robot cooperative manipulation [2]. By decoding muscle activity patterns, sEMG-based HRC systems allow users to perform precise manipulator-driven tasks or to communicate abstract control signals to the robot. These can be achieved by acquiring and analyzing sEMG patterns corresponding to different hand activity gestures, through feature extraction and classification frameworks. Such recognition forms a critical input layer in HRC systems, where the identified hand intent can be mapped to corresponding manipulator actions, such as gripping objects, activating tools, or executing specific assembly tasks. In prosthetic applications, the same framework can be used to interpret user intention for initiating predefined grasp patterns or motion sequences. In rehabilitative contexts, the system can monitor patient effort, detect intended motor function,

and provide adaptive robotic assistance or feedback during therapy. Such systems have shown significant promise in scenarios such as upper-limb assistive robotics and stroke recovery therapy [29]. However, inter-subject variability, sensor noise, and real-time constraints remain major bottlenecks in the reliable deployment of such systems [18].

Despite their complementary advantages, both vision and wearable sensing-based HRI/HRC modalities demand robust computational frameworks that can handle high-dimensional data, nonlinearities, environmental and sensory noise, etc. In response to these challenges, this Thesis aims to develop novel dimensionality reduction frameworks that improve the interpretability and resilience of sensor-driven HRC systems under realistic constraints.

1.1.2 Challenges in Vision-Based Sensing Systems

Vision-based sensing has become an integral component of modern human-robot interaction (HRI) and human-robot collaboration (HRC) frameworks due to its non-intrusive nature, wide spatial range, and intuitive interpretability. Optical sensors such as RGB cameras, infrared imagers, depth sensors, and stereo vision systems enable robots to perceive and interpret human gestures, body posture, and visual markers. This facilitates a wide range of collaborative tasks, from navigation and object handling to intent recognition through visual cue interpretation.

However, the practical deployment of vision-based HRI/HRC systems in real-world environments encounters several challenges. One of the most prominent issues is photometric variability, where uncontrolled lighting conditions, shadows, and reflections significantly degrade the quality of the captured image data. Such degradation adversely affects the robustness of feature extraction techniques, making it difficult to accurately recognize gestures or visual markers in poorly illuminated ambience [23]. Furthermore, background clutter and occlusion which are common in unstructured environments, can corrupt the visual features, leading to ambiguous or failed detection.

Low-cost monocular cameras, though easily accessible and lightweight, suffer from limitations like resolution constraints, image quality, pixel density, and sensor noises; all of which can hinder the precise localization and recognition of visual cues. In the context of semantic visual marker-based navigation systems, the spatial configuration of the marker can be distorted by camera tilt, variable distance from the sensor, partial occlusion, and ambient illumination. This poses significant challenges in interpreting visual cues consistently and

reliably across users and scenarios.

Traditional global feature extraction techniques, such as principal component analysis (PCA) [3], often fail to maintain local structural representations under such visual uncertainties. While a few local texture descriptors like local binary patterns (LBP) [30], local ternary patterns (LTP) [31], and their variants offer some robustness, their effectiveness weakens at higher noise levels or non-uniform lighting conditions. Additionally, many conventional methods do not consider channel-specific variations, often discarding rich color information by converting RGB inputs into grayscale, leading to the loss of discriminative details [32].

These challenges necessitate the development of dimensionality reduction or feature extraction techniques from the vision data that are resilient to illumination variation, sensor noise, and spatial distortion. As investigated in this Thesis, integrating multi-domain similarity information, adaptive spatial kernel modeling, channel-aware feature granulation, fuzzy rule-based local feature decoding, etc., within the framework of novel dimensionality reduction approaches, offers a promising direction toward achieving robust visual interpretation in real-world HRC tasks.

1.1.3 Challenges in Wearable and Contact-Based Sensing Systems

Wearable and contact-based sensing systems have emerged as vital tools in human-robot interaction (HRI), enabling the acquisition of physiological and biomechanical signals that are closely linked to human intent. Among these, surface electromyography (sEMG) is widely used for interpreting muscle activation patterns, providing a direct means to decode motor intentions and enable user-adaptive control of robotic systems. Such approaches are particularly relevant in assistive robotics, prosthetics, and rehabilitation, where capturing user intent is essential for seamless collaboration between humans and robots [15].

Despite their potential, wearable sensing systems are inherently challenged by several practical limitations when deployed in real-world environments. One major issue is the inter-subject and intra-subject variability in sEMG signals. It arises from anatomical differences, muscle fatigue, electrode placement inconsistencies, and variations in skin-electrode impedance over time. These factors can cause substantial fluctuations in signal quality and pattern, even for the same activity performed by different users or repeated across sessions. Such variability makes it difficult to generalize automatic learning models

across individuals without extensive calibration.

Another critical concern is the presence of inherent distortions in sEMG recordings, especially under dynamic or unconstrained ambient settings. Motion artifacts, environmental interference, and signal degradation due to sweat, movement, or electrode displacement can contribute substantially to noise contamination in sEMG signals. These adversely affect the robustness of dimensionality reduction algorithms designed for hand activity classification tasks. In addition to signal-level challenges, wearable systems introduce ergonomic and usability issues, such as discomfort from prolonged usage, constraints on natural movement, and dependency on frequent recalibration. These factors can limit their long-term usage or user adherence in real-world scenarios, particularly in sensitive domains like hand prosthetics and clinical rehabilitation.

To address these multi-dimensional challenges, this Thesis investigates robust subspace learning methods that enhance the reliability and generalizability of sEMG-based activity recognition under variable and noisy conditions. By incorporating adaptive graph structures, sparsity constraints, and manifold fusion strategies into the dimensionality reduction pipeline, the proposed methods aim to mitigate the effects of signal inconsistency and improve cross-subject performance in wearable sensing-based HRC systems.

1.1.4 Limitations of Classical Dimensionality Reduction Techniques

In sensor-driven human-robot collaboration (HRC) systems, high-dimensional data is often generated by vision sensors (e.g., RGB cameras) and physiological signals (e.g., sEMG). These data streams are often redundant, noisy, and complex in structure. To extract meaningful, compact representations for subsequent tasks like classification or recognition, dimensionality reduction (DR) techniques are widely employed. Classical DR methods such as principal component analysis (PCA) [3], linear discriminant analysis (LDA) [4], and independent component analysis (ICA) [33] have long served as foundational tools in this space due to their simplicity and efficiency.

However, these traditional methods exhibit several critical limitations when applied to real-world, multimodal HRC scenarios. Firstly, most classical DR techniques rely on global linear assumptions, which fail to capture the nonlinear, locally structured variations inherent in human gestures, visual marker patterns, or muscle activation signals. For example, PCA

projects data based on variance maximization without regard to class structure or neighborhood preservation, leading to poor discriminability in complex datasets. Similarly, LDA assumes Gaussian class distributions and seeks to maximize inter-class separability using global statistics, which often breaks down under noise, occlusion, or class imbalance.

Another fundamental shortcoming lies in their inability to preserve local geometric relationships among neighboring data points. Such relationships are critical for the accurate classification of subtle visual markers or complex muscle activities. In contrast, dimensionality reduction techniques such as Laplacian eigenmaps (LE) [6] and locality preserving projection (LPP) [7] are better suited for such tasks, as they exploit the local neighborhood structure of the data space to retain intrinsic relationships during projection.

Moreover in vision-based problems, classical DR methods typically operate on vectorized representations of 2D image data, thereby destroying the spatial structures and increasing computational complexity. This transformation can cause the loss of important spatial correlations, especially in hand gesture or visual cue recognition frameworks. Two-dimensional extensions like 2DPCA [34] and 2DLDA [35] were introduced to address this, but they too fall short in robustness under adverse conditions such as illumination changes or non-uniform backgrounds.

In wearable sensing, methods like PCA and ICA lack robustness against signal variability and artifacts introduced by muscle fatigue, electrode shifts, or motion-induced interference. As these methods often treat all features equally, they fail to reduce the influence of noisy or redundant components, leading to degraded performance in sEMG-based activity recognition.

To overcome these limitations, the research presented in this Thesis builds on the foundations of locality-based dimensionality reduction approaches while introducing explicit subspace learning frameworks that are robust to real-world distortions, noise, and user variability. By integrating novel approaches like adaptive kernel construction, channel-wise image handling, granular computing-based segmentation modeling, sparsity-enforced regularization, and possibilistic graph modeling, the proposed techniques improve both interpretability and generalization in sensor-driven HRC tasks.

1.2 Role of Dimensionality Reduction in Human-Robot Interaction Strategies

Human-robot interaction (HRI) systems often operate on high-dimensional data originating from multimodal sources such as vision sensors and physiological signals. These data streams are rich in temporal, spatial, and contextual information but are also inherently redundant, noisy, and computationally expensive to process in their raw form. For real-time interpretation and robust decision-making in HRI scenarios, it becomes essential to transform this high-dimensional input into compact, informative representations that preserve task-relevant characteristics while suppressing distortions, variability, and noise.

Dimensionality reduction (DR) techniques play a critical role in this transformation process by mapping raw sensor data into lower-dimensional subspaces that retain the essential structural information needed for application-specific processing tasks such as classification, regression, trajectory planning, and robotic guidance. In vision-based HRI systems, DR techniques must facilitate the recognition of visual cues under variable lighting, and in the presence of background clutter, sensor noise, etc. Similarly, in wearable sensing applications, such as sEMG-based activity recognition, effective DR techniques should help manage inter-subject variability, signal artifacts, and computational complexity, and improve the generalizability of recognition models. Classical approaches like principal component analysis (PCA) and linear discriminant analysis (LDA) have long been employed for dimensionality reduction due to their mathematical simplicity and interpretability. However, their reliance on global variance or class-level statistics often limits their ability to preserve local structural relationships in the data. These relationships are crucial in capturing subtle variations in visual marker patterns or muscle activity types. To overcome these limitations, nonlinear manifold learning-based and manifold-inspired linear approaches have gained attention for their ability to model complex, locally varying geometries in high-dimensional spaces.

Among these, locality preserving projection (LPP) and its extensions have emerged as particularly well-suited for sensor-driven HRI tasks. LPP preserves the intrinsic neighborhood structure of the data and provides an explicit mapping that facilitates newer samples in real-time projection [7]. Building on this foundation, recent advances including multi-kernel LPP [11], LPP with similarity measures in nonlinear spaces [16], LPP with orthogonality constraint

[9], LPP with uncorrelated discriminant [8], LPP with 2D data handling [13], etc., have further improved the resilience of subspace learning models against environmental noise, signal degradation, and user variability.

This section explores the evolution of dimensionality reduction techniques from traditional global methods to nonlinear manifold learning with their linear approximations, and subsequently to their multi-dimensional robust variants. It leads to customized projection strategies such as LPP and its tailored variants for robust HRI modalities under real-world constraints.

1.2.1 Traditional Global Dimensionality Reduction Approaches

Traditional global dimensionality reduction (DR) techniques have long served as foundational tools in machine learning, pattern recognition, and signal processing. These methods aim to transform high-dimensional input data into a lower-dimensional subspace while preserving as much of the original structural correlation as possible. Their global nature implies that they primarily operate based on overall data distributions or class-level statistics, often without explicitly considering local geometric relationships between neighboring data points.

Principal component analysis (PCA) is among one of the most widely used global DR methods, which identifies orthogonal directions i.e., principal components in the feature space that capture the maximum variance in the data [3]. By projecting the data onto a reduced set along these directions, PCA facilitates efficient representation and noise suppression. Despite its simplicity and effectiveness in well-structured datasets, PCA is inherently unsupervised and does not incorporate class label information, making it suboptimal for classification tasks where discriminative feature learning is crucial. Moreover, it ignores local spatial relationships among data samples which is very crucial in context-aware high-dimensional data projections. Linear discriminant analysis (LDA) extends the unsupervised DR approaches by introducing class supervision [4]. LDA seeks to maximize the ratio of between-class scatter to within-class scatter, thereby enhancing class separability in the projected space. It relies on class means and global covariance matrices, assuming that the class distributions are linearly separable and follow Gaussian statistics. While effective in certain scenarios, LDA often struggles when the number of samples per class is limited (small-sample-size problem) or when data lies on nonlinear manifolds which is quite common in vision and EMG-based human-robot interaction

(HRI) strategies. Another global method, independent component analysis (ICA), focuses on finding statistically independent components by maximizing non-Gaussianity in the data distribution [33]. ICA has been particularly explored in biomedical signal processing, such as electroencephalography (EEG) and electromyography (EMG), for artifact removal or source separation [39]. However, ICA requires strong assumptions about source independence and is highly sensitive to noise and dimensionality, making it less suitable for complex, noisy, and unstructured data modalities.

Despite their mathematical simplicity and computational efficiency, these global approaches are limited in their ability to capture nonlinear structures or local neighborhood relationships, which are crucial aspects while analyzing real-world sensor data with high variability, clutter, noise, etc. For instance, in vision-guided HRI, gestures or marker patterns may vary subtly depending on lighting, pose, or user-specific execution. Similarly, in wearable EMG applications, inter-subject differences and signal distortions introduce complex nonlinearities that are not well modeled by such global statistics alone. These limitations have led to the growing adoption of nonlinear manifold learning and their linear approximations, which aim to uncover the intrinsic low-dimensional geometry of the data. While these approaches improve representation quality, many suffer from the lack of an explicit mapping function, hindering their applicability in real-time or unseen data scenarios. This motivates the development of localized, projection-based subspace learning frameworks, such as locality preserving projection (LPP) [7], which combine the benefits of manifold learning with linear approximation and explicit transformation capabilities. These are discussed in detail in the following subsections.

1.2.2 Emergence of Manifold-Inspired Learning Techniques

The limitations of classical global dimensionality reduction techniques in modeling real-world, complex sensor data have led to the rise of nonlinear manifold learning techniques and their linear approximations. These methods are built on the assumption that high-dimensional data, such as visual gestures or physiological signals spread over a low-dimensional, nonlinear manifold, embedded within the ambient space. Instead of relying on global variance or class separation statistics, manifold-inspired methods aim to uncover the intrinsic geometry of the data by preserving local neighborhood relationships, geodesic proximity, or reconstruction

characteristics during dimensionality reduction.

Some of the prominent approaches from the manifold learning family can be named: isometric feature mapping (Isomap) [5], locally linear embedding (LLE) [36], and Laplacian eigenmaps (LE) [6]. Isomap seeks to preserve the spatial relationship between all pairs of points on the manifold by approximating them through shortest-path computations on a neighborhood graph. It is particularly useful for data with global nonlinear structure but is sensitive to graph connectivity and outliers. LLE attempts to preserve the local linear structure of the data by assuming that each point and its neighbors lie on a locally linear patch of the manifold. LLE computes low-dimensional embeddings that best reconstruct these relationships. Laplacian eigenmaps use spectral graph theory to construct embeddings that preserve local similarities between neighboring data points by solving a graph Laplacian-based eigenvalue problem.

While these techniques are effective in revealing the manifold structure and producing meaningful embeddings, they often lack an explicit mapping function from the high-dimensional input space to the reduced subspace. This absence of a parametric or closed-form transformation limits their applicability in real-time systems or new data streaming scenarios, where unseen data points must be projected efficiently without retraining the entire model. Additionally, their reliance on fixed neighborhood graphs can lead to instability when applied to noisy or cluttered data, which is common in HRC applications such as vision-based cue recognition or EMG signal interpretation.

In human-robot interaction tasks, nonlinear manifold learning has shown promise in enhancing the recognition of complex patterns, such as hand gestures with subject-wise variation, visual markers with inconsistent configuration, and physiological signals affected by electrode shifts or user-specific characteristics. However, their non-parametric nature and computational resource consumption restrict their deployment in embedded, time-constrained environments, especially where model interpretability and generalization to new samples are critical.

To overcome these limitations, several linear approximations of nonlinear manifold methods have been proposed. Notable among them are neighborhood preserving embedding (NPE) [37], neighborhood preserving projection (NPP) [38], and locality preserving projections (LPP) [7], which attempt to retain the advantages of manifold learning while offering explicit linear mappings. These methods are particularly useful in HRC systems where real-time intent recognition, adaptability, and computational feasibility are essential. The next subsection

focuses on LPP and its several recent extensions, which form the backbone of the subspace learning models proposed in this Thesis.

1.2.3 Locality-based Learning: LPP and Its Variants

Locality preserving projection (LPP) is a prominent subspace learning method that serves as a linear approximation of the nonlinear Laplacian eigenmap problem [7]. It addresses the key limitations of both global dimensionality reduction techniques (e.g., PCA [3], LDA [4]) and nonlinear manifold learning methods (e.g., LLE [36], Isomap [5]), by combining neighborhood-preserving capabilities with an explicit linear mapping. This makes LPP especially suitable for interpreting human-robot interactive cues, where real-time projection of high-dimensional sensor data onto a lower-dimensional manifold is essential for intent recognition or control through classification or estimation.

LPP constructs a graph-based Laplacian matrix to preserve local geometric structures among data samples. Specifically, it builds a similarity graph using a k-nearest neighbor (kNN) scheme and computes a weight matrix that encodes proximity between data points, often based on a Gaussian heat kernel. The objective is to solve a generalized eigenvalue problem that minimizes the distance between neighboring points in the projected space, effectively preserving the local structure of the data manifold in the reduced subspace [7]. Unlike purely nonlinear methods, LPP produces an explicit projection matrix $\mathbf{W}$, allowing newer data samples $\mathbf{x}$ to be mapped to $\mathbf{y}$ via a linear transformation $\mathbf{y} = \mathbf{W}^\top \mathbf{x}$, a property that is vital for real-time and embedded applications.

Despite these advantages, standard LPP still faces limitations in noise sensitivity, class separability, and feature redundancy, particularly in the context of real-world HRI scenarios. For instance, in vision-based systems, photometric variability, background clutter, and occlusion can distort the spatial structure of visual cues, reducing the reliability of LPP-based representations. Similarly, in wearable sEMG systems, inter-subject variability, motion-induced signal artifacts, and sensor noise affect the stability of the LPP-projected feature space.

To overcome these shortcomings, several enhanced variants of LPP have been discussed and evaluated in this Thesis. For example, discriminant locality preserving projection (DLPP) [40] has been one of the earliest improvements on LPP, which incorporated class label information to improve class separability while preserving locality. To address

redundancy and overfitting, orthogonal DLPP (ODLPP) [41] was introduced which imposes orthogonality constraints, thereby reducing the correlation between projection vectors. Later, supervised kernel LPP (SKLPP) [42] was proposed which combines the label information with nonlinear kernel mappings to enhance discrimination in complex feature spaces. To handle data with higher intra-class variance and distributional irregularities, Monte Carlo discriminant LPP (MC-DLPP) [43] was developed which employs stochastic sampling to build robust class structures. Multiscale LPP (MSLPP) [12] on the other hand, extended LPP by modeling local relationships at multiple neighborhood scales, capturing fine-to-coarse variations in data. Orthogonal enhanced LPP (OELPP) [44] added further regularization and orthogonality constraints to improve feature decorrelation in noisy environments. In another work, LPP with autoencoder (LPPAE) [45] was augmented, which integrates LPP with deep autoencoder architectures to capture nonlinear manifolds while retaining explicit low-dimensional embeddings. Recognizing the importance of maintaining a two-dimensional data structure, particularly in image analysis, 2DLPP was proposed as a matrix-based extension that avoids vectorization and preserves spatial correlations better [13]. This was further augmented into ℓ_1-norm-based 2DLPP (2DLPP-L1) [46] and Robust 2DLPP (2DRLPP) [14], which improved robustness against outliers and noise through sparsity-enforcing norms and regularized formulations.

Collectively, these variants illustrate the growing trend of adapting LPP-based formulations for robustness, discriminative capability, and computational feasibility in practical sensing environments. However, many of these methods still face challenges such as reliance on fixed graph structures, sensitivity to parameter tuning, and loss of spatial or feature-level interpretability. These limitations motivate the need for further advancements, particularly in the context of noisy, irregular, and heterogeneous data encountered in vision- and EMG-based HRI systems.

1.3 Vision Sensing-Based Robot Guidance: State-of-the-Art

Vision sensing has become an integral component in enabling context-aware and intuitive robot guidance in human-robot interaction (HRI) frameworks. Compared to other sensing modalities, visual sensors offer a contactless, spatially rich, and semantically interpretable

stream of information, making them particularly well-suited for collaborative robotics, assistive systems, and indoor navigation. In such scenarios, robots rely on visual cues which are often provided by human users through body gestures, visual markers, or other structured patterns, to infer the intent and execute subsequent tasks accordingly.

Early approaches to visual robot guidance primarily focused on human-supervised systems, where symbolic gestures or predefined visual patterns were employed to convey interpretable commands to the assistive robots. These cues are often encoded using simple hand gestures or fabricated markers using geometric shapes or colored objects held in the user's hand, which are then detected by the onboard cameras mounted on mobile or assistive robots. The extracted features are interpreted through rule-based systems or template-matching methods to trigger predefined robot behaviors. While such systems are inherently intuitive and inexpensive to deploy, they remain susceptible to various real-world factors such as inconsistent illumination, occlusions, camera viewpoint variation, and visual background clutter.

However, in practical deployment, vision-based robot guidance systems must operate under dynamic lighting conditions, sensor noise, and hardware constraints which are regular in embedded systems. The raw image data obtained from low-cost monocular cameras often suffer from intensity degradation, pixel corruption, and quantization noise, all of which hinder the robustness and repeatability of the visual cue recognition system. These issues are particularly pronounced in cluttered indoor environments where the distinction between foreground (e.g., cue marker) and background becomes visually ambiguous. These issues impact the accuracy of traditional feature descriptors such as histogram of oriented gradients (HOG) [47], Gabor filters [47], and local binary patterns (LBP) [30].

To address these challenges, there has been a growing focus on the development of robust subspace learning methods, particularly those based on graph-based topology and manifold learning principles, such as locality preserving projections (LPP) [7] and its two-dimensional extensions [13], [14]. By modeling local geometric structures and preserving spatial consistency in projection spaces, these methods offer improved resilience to noise, illumination changes, and non-linear variations in visual formation. Moreover, newer approaches have explored the integration of color-channel granulation, multi-dimensional kernel fusion, fuzzy-based feature decoding, adaptive graph construction, etc., to better capture the visual cue semantics under realistic conditions.

The subsequent subsections delve into the evolution of human-supervised visual cue

systems, the impact of photometric and sensor-induced degradations, and the role of advanced LPP-based models in enhancing the robustness and generalizability of vision-guided robot navigation frameworks.

1.3.1 Human-Supervised Visual Cue Systems

In human-robot collaborative scenarios, direct supervision by human users through visual signals remains a widely adopted mechanism for robot control and navigation, particularly in assistive or constrained indoor environments. These systems rely on structured visual inputs, such as hand-held objects, colored markers, or patterned arrangements; that can be intuitively understood by users and interpreted by robotic platforms equipped with vision sensors. By minimizing the need for complex interfaces or wearable devices, such systems provide a low-cost and accessible channel for communicating high-level, intelligent instructions to the robot.

For example, such visual guidance schemes can be facilitated by hand gestures, body postures, human actions, artificial visual markers, etc., as demonstrated in several assistive robotic application frameworks. In these systems, the human user waves his/her hands in a specific direction or orientation with a specific gesture or holds an artificially fabricated visual structure e.g., with a stick and a shaped marker at specific orientations or positions relative to the robot's field of view. Each unique spatial arrangement corresponds to a predefined control command; such as forward motion, angular shift, left turn, right turn, or stop, thereby enabling the robot to interpret human intent through symbolic gesture cues. The simplicity of such gesture vocabulary makes it especially well-suited for scenarios involving elderly or mobility-impaired users, where the ease of interaction is of utmost importance.

From a sensing standpoint, monocular RGB cameras are typically mounted on the robot to capture incoming visual frames in real time. These inputs are processed to extract spatial and appearance-based features from the visual cue patterns. The system then performs cue recognition through a combination of preprocessing, segmentation, dimensionality reduction, and classification. In early-stage implementations, rule-based decision-making systems or template-matching methods were used to associate input patterns with corresponding actions. However, such methods have been shown to be vulnerable under practical conditions due to their sensitivity to illumination constraints, background variation, occlusion, and photometric

inconsistencies [23].

To address these challenges, computational frameworks developed in this Thesis, aim to introduce enhancements such as fine-grid granulation, adaptive spatial kernels, kernel fusion strategies, and robust graph-based subspace learning techniques to better capture the salient characteristics of visual cues in complex environments. These developments have improved recognition accuracy in scenarios involving poor lighting, sensor noise, and background clutter. Nevertheless, the reliance on structured markers implies that system performance remains dependent on the clear visibility and separability of the cues from the surrounding context, which can be an important aspect of implementing granular segmentations in this regard.

In practical applications, human-supervised visual cue systems have been successfully integrated with IoT-enabled assistive robots, wherein camera streams are processed either onboard or remotely, and the robot's navigation responses are updated in real time. This approach has proven effective in constrained indoor spaces such as laboratories, smart homes, and assistive living. However, further robustness is still needed to accommodate unstructured backgrounds, variable cue patterns, and different user approaches to cue presentation. Overall, human-supervised visual cue systems offer a pragmatic and interpretable interface for guiding robot behavior in shared spaces. Their continued evolution, particularly through robust feature encoding and dimensionality reduction techniques, will be central to scaling such systems beyond structured test environments into everyday use.

1.3.2 Photometric Degradation and Sensor Uncertainties

Vision-based robot guidance systems deployed in real-world environments often face significant performance degradation due to variations in ambient lighting and inherent limitations of low-cost camera sensors [49]. These photometric degradations, including uneven illumination, shadows, and specularities can distort both the appearance and geometry of visual cues, thereby compromising the consistency and accuracy of visual feature extraction. In human-supervised cue systems that rely on symbolic markers or gestures, such degradations often lead to incorrect segmentation, blurred boundaries, or misclassification, especially when the background is visually cluttered or when the cue partially overlaps with similar-colored regions [32].

A typical challenge arises when the robot is in transition between zones with differing lighting conditions, such as from well-lit areas to dimly lit zones. Under these conditions,

the pixel intensities of the same cue pattern can vary significantly across frames. When using monocular RGB sensors, which are commonly selected for their low cost and ease of integration, the absence of depth and the reliance on intensity-based information make the system especially sensitive to such variations. Moreover, hardware-related sensor noise such as electronic noise, analog-to-digital converter (ADC) quantization error, and lens distortion further complicates the task of reliable feature extraction [50]. These uncertainties get intensified over time due to sensor aging, physical wear and tear, and temperature fluctuations, all of which affect the optical fidelity of the captured image stream.

Traditional image descriptors like LBP, HOG, or Gabor filters show limited robustness in such environments [11]. Their local neighborhood computations are often highly sensitive to changes in absolute pixel intensity, and they generally lack mechanisms to distinguish between actual cue structures and photometric artifacts. While some degree of invariance can be achieved using measures like histogram equalization or color space transformation, these preprocessing techniques alone do not suffice when both spatial and chromatic distortions co-occur in real-time.

To mitigate these issues, recent approaches have explored adaptive feature encoding, granular color channel modeling, and spatial-feature kernel fusion, which more effectively capture structural patterns even in low-fidelity conditions. For example, previous approaches like histogram refinement or feature-induced bilateral weighting were introduced to improve robustness under photometric inconsistencies, demonstrating more reliable performance than traditional LPP or PCA in indoor HRI tasks.

Despite these improvements, achieving consistent cue recognition across varying environmental conditions remains an open challenge. Any practical vision-based robotic system must account for these photometric and sensor-induced uncertainties through both algorithmic resilience and adaptive sensing strategies. The next section explores how advanced LPP-based models have been extended to address such limitations in real-world deployment contexts.

1.3.3 Advanced LPP Models for Robust Visual Cue Recognition

While locality preserving projection (LPP) offers a computationally viable and theoretically strong method for preserving local structure in high-dimensional data, its standard formulation remains vulnerable to various challenges encountered in real-world visual sensing interfaces.

In particular, when applied to human-robot interaction (HRI) [51] scenarios involving symbolic visual cues, the effectiveness of LPP can degrade due to environmental disturbances such as illumination variability, background clutter, occlusions, and sensor-induced noise.

To improve robustness under such conditions, several advanced variants of LPP [7] have been developed to improve its robustness, generalization, and structural preservation. One line of enhancement focuses on enforcing orthogonality constraints to reduce redundancy among projection directions and improve numerical stability. Among these, orthogonal LPP (OLPP) [9] introduces orthogonal bases to ensure uncorrelated projections. Similarly, fast orthogonal LPP (FOLPP) [10] and orthogonal enhanced LPP (OELPP) [44] add regularization to improve scalability and noise resilience, especially when dealing with high-dimensional data or limited training samples.

Another important direction is the development of two-dimensional LPP-based models that preserve the matrix structure of image data. The traditional LPP operates on vectorized inputs, thereby discarding the spatial layout of pixel intensities. To address this, 2D-LPP [13] was introduced, allowing projections to operate on rows or columns of image matrices directly. This was followed by 2DLPP-L1 [46], which replaces the ℓ_2-norm in the objective function with ℓ_1-norm, offering improved robustness to outliers. Building further on these approaches, Robust 2D-LPP (2DRLPP) [14] incorporates additional regularization terms to control model complexity and suppress noise-induced variations in spatial structure.

Additional advancements in LPP modeling involve leveraging more adaptive or discriminative mechanisms. For instance, Monte Carlo discriminant LPP (MC-DLPP) [43] introduces stochastic neighborhood sampling to reduce sensitivity to data imbalance and enhance class separability. Multiscale LPP (MSLPP) [12] aggregates features over multiple neighborhood sizes to capture structural variations across different spatial scales. In parallel, LPP with autoencoder (LPPAE) [45] embedding combines graph-based locality preservation with the representational strength of deep autoencoders to improve the performance in highly nonlinear environments.

Despite these improvements, several key challenges continue to exist. Most of the variants in use, rely on fixed or static similarity graphs, which may not accurately reflect the complex variations in real-world visual data, particularly under challenging lighting or sensor conditions. Additionally, many approaches treat spatial and photometric features separately, lacking an integrated mechanism to jointly model both types of information. This becomes problematic

in HRI environments where visually similar gestures or markers must be distinguished based on subtle color or texture cues, often in the presence of cluttered background or environmental interference.

These observations highlight the need for more adaptive, semantically aware, and noise-tolerant projection frameworks, particularly those capable of integrating spatial structure, photometric variation, and uncertainty into a unified model. In practical vision-guided HRI systems, where real-time decisions must be made based on noisy and variable inputs, such improvements are essential for reliable cue recognition and robust robot control.

1.4 Surface EMG-based Human Intention Recognition

In human-robot collaboration (HRC), accurately inferring user intent is a prerequisite for enabling intuitive and adaptive robotic behavior and assistantship. While vision-based systems offer an external perspective on user motion or gesture cues, they face challenges in scenarios involving occlusion, poor lighting, or constrained visual fields. As an alternative and complementary approach, surface electromyography (sEMG) provides a direct, physiological channel for recognizing user intent by capturing the electrical activity generated by muscle contractions. This makes it particularly relevant for applications in assistive robotics, prosthetic control, and rehabilitative technologies, where real-time decoding of motor intention is essential [52].

Surface EMG signals are typically acquired using non-invasive electrodes placed on the skin surface above the target muscle groups. The recorded signals reflect cumulative motor unit action potentials, offering insight into the neuromuscular dynamics associated with voluntary movement [53]. In the context of HRC, these signals can be analyzed to identify distinct hand gestures, grasp types, or movement primitives, which can then be mapped to robot control commands. sEMG-based systems help enable fine-grained and low-latency control, which are especially effective in tasks that involve hand-based manipulation or joint-level coordination.

Despite its potential, sEMG signal processing is also not free from practical challenges on frequent occasions. The acquired data is highly sensitive to electrode placement, muscle fatigue, inter-subject anatomical variability, and motion-induced artifacts, making it difficult to achieve consistent classification performance across users or sessions [54]. Additionally, the signals

are inherently non-stationary, requiring robust feature extraction and representation techniques that can adapt to temporal variation and signal degradation. These concerns necessitate the development of algorithms that are not only accurate but also computationally efficient and generalizable to new conditions.

In recent years, there has been increasing emphasis on dimensionality reduction methods that can extract meaningful features from high-dimensional sEMG recordings. Classical techniques such as PCA and LDA have been explored for compressing signal features while preserving class-relevant information [55]. However, their limitations in modeling nonlinear manifolds and adapting to noisy or imbalanced data, have prompted the adoption of more advanced methods; including graph-based and manifold-preserving projections. Among these, locality preserving projection (LPP) [7] and its extensions have gained attention due to their ability to maintain neighborhood structure in the reduced space, which is particularly important in modeling subtle variations across myogesture classes. Moreover, the application of orthogonality-constrained LPP variants has shown improved robustness in sEMG-based classification problems. These methods help suppress noise, reduce redundancy, and ensure stability in the learned representations. Such formulations are particularly advantageous when working with high-dimensional time-series features, as is often the case in EMG-based recognition systems. This Thesis aims to introduce further advancements in such approaches by investigating avenues such as adaptive graph construction, possibilistic weighting, and multi-metric fusion, that expand the applicability of subspace learning methods in real-world wearable sensing scenarios.

As sEMG continues to gain relevance in HRC applications, especially for upper-limb activity recognition and prosthetic intent decoding, the development of resilient, interpretable, and deployable feature learning models remains a critical area of research. The following subsections explore the fundamental challenges associated with sEMG signal interpretation, the role of dimensionality reduction in handling its complexity, and recent advancements in LPP-based methods designed to improve classification performance under realistic conditions.

1.4.1 Challenges in sEMG Signal Interpretation

Surface electromyography (sEMG) presents a compelling modality for capturing voluntary muscle activity and decoding human intention in real-time human-machine interface (HMI)

tasks. However, despite its advantages in terms of responsiveness and direct physiological relevance, the interpretation of sEMG signals remains a complex task, particularly in the context of wearable sensing in HMI or HRI. A range of technical, anatomical, and environmental factors combine to make sEMG a highly variable and noise-sensitive signal source, posing several challenges to robust and generalizable intention recognition.

One of the foremost issues of concern is the non-stationary nature of sEMG signals corresponding to muscle movements [53]. Unlike controlled stimuli in laboratory environments, voluntary muscle activity in daily tasks is affected by a wide range of conditions; including user fatigue, joint angle variation, and biomechanical constraints. These can alter the statistical characteristics of the myosignals over time. These changes also introduce intra-class variability, making it difficult for classifiers to maintain consistent performance across sessions or over extended use. In addition, electrode placement sensitivity is a critical factor. Small changes in sensor positioning due to contact inconsistencies, slippage during movement, or individual anatomical differences, can significantly alter the amplitude and pattern of the signal morphology. Even when placed over the same muscle group, slight positional deviations may cause crosstalk from adjacent muscles or hinder the capture of the desired activation patterns altogether. This results in inter-session and inter-subject variability, which complicates the training of generalized models suitable for user-independent operations. Motion artifacts and environmental noise are further sources of degradation. Movement of the skin or cable relative to the electrode can generate low-frequency artifacts that overlap with the signal of interest, especially during dynamic arm or hand motions [56]. Additionally, interference from nearby electronic devices or improper grounding can introduce electromagnetic noise, particularly in unsupervised or mobile settings. Sometimes, the myoresponse itself is characterized by low signal strengths, especially during subtle or fine motor actions where muscle contraction is weak. This necessitates sophisticated preprocessing and filtering strategies, which, if not properly tuned, can suppress informative components of the signal along with the noise.

Beyond hardware-level concerns, there are algorithmic and modeling challenges as well. sEMG signal features tend to exhibit high intra-class variation and inter-class similarity, particularly for gestures involving similar muscle groups or overlapping motor synergies. This places a significant burden on feature extraction and dimensionality reduction techniques, which must balance discriminability with robustness to signal distortion. Standard feature representations such as root mean square (RMS), zero crossing rate, or waveform length [57],

though computationally efficient, often lack the descriptive power to distinguish such subtle activity patterns, particularly under noisy or variable conditions. Additionally, deep learning approaches which have demonstrated excellent performance in recent years, typically require large-scale, user-specific datasets to discern effectively. This imposes a practical challenge in most wearable or assistive robotics environments compared to the classical strategies. The high variability of sEMG signals often leads to overfitting in data-driven models, unless regularized carefully or augmented with domain-specific priors.

Altogether, these challenges underscore the need for robust, low-latency, and adaptable signal processing pipelines that can handle uncertainty, reduce dependence on sensor configuration, and generalize across users and use cases. The following sections explore how dimensionality reduction techniques, particularly those that preserve local structure and minimize sensitivity to outliers, have been adapted to address these issues in sEMG-based intention recognition systems.

1.4.2 Dimensionality Reduction for sEMG Data

Surface electromyography (sEMG) signals, while rich in neuromuscular information, are inherently high-dimensional and exhibit significant inter- and intra-subject variability. These characteristics, coupled with sensor noise, motion artifacts, and anatomical diversity, pose serious challenges for consistent feature extraction and classification. As such, dimensionality reduction (DR) is a critical step in sEMG-based hand activity recognition systems. It enables the transformation of raw or high-dimensional features into a compact, discriminative subspace that enhances computational efficiency and improves class separability, particularly in real-time or embedded robotic applications.

Traditional approaches in sEMG analysis employed classical DR techniques such as principal component analysis (PCA) [55], linear discriminant analysis (LDA) [55], independent component analysis (ICA) [58], and empirical mode decomposition (EMD) [58]. PCA identifies orthogonal directions of maximum variance, whereas LDA seeks linear projections that maximize between-class variance while minimizing within-class spread. ICA has been used to extract statistically independent components from surface EMG signals, particularly in tasks involving source separation or artifact removal. Likewise, EMD has been applied as a data-driven approach for decomposing non-stationary sEMG signals into intrinsic mode

functions (IMFs), often as a preprocessing step for feature extraction. However, both PCA and LDA are limited in their capacity to preserve local structural relationships in the data, especially when gestures involve subtle neuromuscular variations or when signals are distorted by noise or non-stationarity. While ICA and EMD offer advantages in noise suppression and signal interpretation, their outputs are not always directly suited for classification tasks and may require additional dimensionality reduction layers to ensure compact and discriminative representations. These limitations of traditional methods are particularly pronounced in wearable sensing environments, where signal characteristics shift due to fatigue, repositioning, or variability in electrode contact.

To address these limitations, research has been carried out toward manifold learning-inspired DR techniques, which preserve the intrinsic geometry of high-dimensional sEMG data. Among them, locality preserving projections (LPP) [7] has been a well-suited approach due to its ability to maintain neighborhood structures during projection while offering an explicit linear transformation matrix, especially critical for real-time classification tasks. In the context of EMG, manifold learning-inspired approaches have shown promise in retaining discriminative patterns embedded in muscle activation signals, particularly when local variations among gestures or subjects are subtle. Especially, the LPP-based methodologies and their augmented variants are thoroughly investigated in this Thesis.

Several refinements to LPP have been reported in the earlier literature to address its core limitations, particularly its sensitivity to noise and the rigidity of its Euclidean distance-based similarity graph. For example, orthogonal LPP (OLPP) [9] introduces orthogonality among projection vectors, ensuring uncorrelated and more interpretable components while improving numerical stability. It is especially useful in EMG-based hand movement recognition tasks as it enhances feature decorrelation and improves classifier generalization. Multi-kernel fusion-based LPP approaches [16] have been adopted to capture the data similarity in various nonlinear manifolds. To model the similarity graph in a flexible way, more adaptive approaches, such as adaptive neighborhood graph (ANG) [59] and possibilistic neighborhood graph (PNG) [60] have been introduced, to improve the fidelity of neighborhood modeling. Unlike conventional k-NN graphs or Gaussian kernels, PNG adapts to the data's intrinsic distribution, reducing sensitivity to outliers. In EMG applications, such adaptive graphs can be used to improve the projection robustness by better capturing the dynamic local dependencies in noisy biosignals. Other studies have extended the similarity modeling beyond Euclidean

geometry. For instance, incorporating cosine similarity [61], Euler-based distances [62], and Grassmannian manifold representations [63] have led to more resilient graph construction, particularly for high-dimensional data that exhibit both magnitude and directional variability.

Despite these advancements, practical deployment in wearable systems still demands dimensionality reduction models that are computationally efficient, adaptable to user variability, and robust to signal distortion. The ability to extract informative, sparse, and low-dimensional features from EMG data, without sacrificing classification accuracy, is essential for scaling these systems in real-world HRI and assistive robotics applications. These considerations form the basis for the continued exploration of orthogonality-enforced, graph-adaptive, and manifold-aware LPP variants, which are extremely important in robust sEMG signal analysis.

The next subsection discusses how robust LPP-based frameworks are developed and tailored for hand activity recognition, offering further improvements in noise resilience and inter-subject generalization.

1.4.3 Robust Orthogonal LPP Models for Hand Activity Recognition

Accurate recognition of hand activity from surface electromyography (sEMG) signals requires feature representations that are compact, discriminative, and robust to signal variability. Given the non-stationary and user-dependent nature of sEMG, traditional global dimensionality reduction methods often fall short in preserving the subtle intra-class variations while maintaining class-level discriminability. In this context, locality preserving projection (LPP) has emerged as a promising tool due to its ability to retain the local geometric structure of the data in a reduced subspace [7]. Its linear formulation and capacity for out-of-sample generalization make it particularly suitable for real-time data streams where new data must be processed efficiently.

However, the standard LPP algorithms have their known limitations. Their reliance on unconstrained optimization leads to projection vectors that may be highly correlated, reducing the interpretability and numerical stability of the resulting features. This redundancy is especially problematic in high-dimensional EMG feature spaces, where noise and signal overlap between muscle activations are common. To address this, researchers have introduced orthogonality constraints into the LPP framework, leading to what is broadly known as orthogonal locality preserving projection (OLPP) [9]. OLPP enhances LPP by ensuring that

the learned projection directions are mutually orthogonal, thereby eliminating redundancy and improving feature decorrelation. This property can be particularly beneficial in biosignal classification tasks where overlapping motor unit activations can result in high feature correlation. By enforcing orthogonality, the projected subspace becomes more stable and interpretable, which in turn improves classifier performance; especially in tasks involving a high number of gesture classes or overlapping activation patterns. Orthogonal enhanced LPP (OELPP) [44] introduces an additional regularization term to further improve the conditioning of the projection matrix and enhance robustness against noise, particularly in situations involving highly variable or corrupted input features. On the other hand, discriminant LPP (DLPP) [40] incorporates label information into the projection process, aiming to maximize class separability while preserving local manifold structure. Although DLPP improves discriminative capability, it can still suffer from redundant projections in high-dimensional settings. To overcome this, a hybrid approach orthogonal discriminant LPP (ODLPP) [41] extends DLPP by incorporating orthogonality constraints, thereby combining the strengths of supervised discriminative modeling with enhanced numerical stability.

Despite these improvements, a key challenge remains in the construction of the similarity graph that defines neighborhood relationships in LPP, OLPP, and other variants. Traditional LPP relies on fixed Euclidean distance-based graphs, such as k-nearest neighbors or Gaussian heat kernels [11]. These approaches assume homogeneous local distributions, which do not accurately reflect the heterogeneity present in sEMG data, particularly across subjects or sessions. Misrepresentation in the graph structure can lead to loss of critical inter-class boundaries or poor preservation of intra-class topology. To mitigate this, recent studies have investigated adaptive or data-driven graph construction techniques. Graphs constructed using possibilistic weighting, for instance, allow for greater flexibility in modeling uncertainty and local density variations in the data [60]. Such methods reduce the undue influence of outliers and better preserve meaningful relationships in noisy or sparse signal segments. Additionally, incorporating alternative similarity metrics, such as cosine distance or projections onto Grassmannian manifolds, has shown promise in capturing directional and structural patterns that Euclidean distance may overlook [61].

While these directions mark substantial progress, there remains an ongoing need for robust orthogonal projection frameworks that integrate graph adaptability, sparsity, and multi-geometry modeling altogether to effectively handle signal degradation, inter-subject

variability, and class ambiguity. These requirements are particularly demanding in HRI applications, where reliability and generalization across users and conditions are crucial [64]. The development of such models forms a natural extension of the previous research works and serves as a motivating foundation for the contributions presented in subsequent chapters.

1.5 Research Gaps

While significant progress has been made in developing perception-driven models for human-robot collaboration (HRC), several open challenges continue to adversely affect their robustness and adaptability in real-world deployment. The review of prior work across vision-based robot guidance and sEMG-based human intention recognition reveals that existing methods often perform well in controlled environments but exhibit reduced generalizability under real-world variability; such as sensor noise, illumination changes, signal drift, or subject-specific differences.

In the context of vision-guided navigation, human gesture, and symbolic cue systems have demonstrated their practicality, particularly when simple and repeatable markers are used to coordinate movement commands. However, the effectiveness of these systems declines significantly under inconsistent lighting, partial occlusion, and cluttered backgrounds [65]. Many existing dimensionality reduction and feature encoding techniques fail to maintain robustness under such visual disturbances, and models relying on hand-crafted descriptors often lack the adaptability required for continuous operation across sessions or users.

Similarly, sEMG-based recognition systems, despite their ability to directly capture neuromuscular intent, face persistent issues related to electrode placement variability, low signal strengths, inter-subject anatomical differences, etc [66]. Feature extraction techniques, especially those based on global or linear projection models, frequently fall short in handling the non-stationarity and redundancy characteristic of EMG signals. Although recent manifold learning-inspired frameworks have shown promise in preserving local signal structure, challenges related to outlier sensitivity and generalization still remain an open challenge.

Beyond these sensing-specific concerns, a broader gap persists in the robustness and adaptability of manifold learning-inspired algorithms when applied to real-world HRC systems [67]. Although many advanced projection models proposed in the literature demonstrate strong

performance under controlled or idealized conditions, their effectiveness tends to degrade in the presence of real-world complexities such as signal corruption, environmental variability, and heterogeneous input quality. Most existing methods rely on fixed or static graph constructions, which fail to account for data uncertainty, local distribution shifts, or inconsistencies arising from sensor noise, occlusion, or user-specific variation. Moreover, these models often lack mechanisms to handle incomplete, imprecise, or ambiguous data; a common occurrence in wearable or vision-based sensing environments. The absence of adaptive graph modeling [59], context-sensitive similarity measures, and robust outlier mitigation limits their generalizability and makes them less suited for deployment in dynamic, real-world collaborative settings, where input conditions are rarely uniform or predictable.

These observations suggest the need for more resilient, adaptable, and computationally efficient solutions that can bridge the divide between theoretical locality-based dimensionality reduction and practical, sensor-level deployment in collaborative robotics. The subsections that follow, identify specific gaps in three key domains: vision-based guidance, EMG-based intention modeling, and the real-world implementation of manifold learning-inspired techniques.

1.5.1 Gaps in Visual Cue-Based Navigation Systems

Visual cue-based navigation systems have become a widely explored modality in human-robot interaction (HRI), owing to their intuitive interface and contactless sensing mechanism [68]. In symbolic navigation frameworks, such as those involving hand-held visual markers, human users provide directional cues to mobile robots through predefined visual patterns. These systems are often deployed in assistive robotics and indoor navigation, where cost-effectiveness and ease of use are prioritized. However, despite their conceptual simplicity, several practical limitations continue to hinder the robustness and scalability of such systems beyond controlled environments.

One major limitation arises from sensitivity to photometric variation. As shown in experimental studies across multiple investigations, changes in ambient lighting, shadows, or background illumination can significantly distort the appearance and interpretation of visual cues. Many traditional feature extraction methods, such as local binary patterns (LBP) or gradient-based descriptors [11], do not generalize well under non-uniform lighting or

low-contrast conditions. This leads to unreliable classification or misinterpretation of gesture patterns, especially in environments with reflective surfaces or mixed indoor lighting.

Another challenge is related to visual ambiguity and background interference in dynamic visual frames. In real-world settings, cues may overlap with background elements of similar color or texture, making it difficult to isolate the marker from its surroundings [69]. The reliance on fixed segmentation thresholds or color-based heuristics in many earlier systems, contributes to classification errors, particularly when the robot's field of view is not consistently aligned with the cue, or when partial occlusion occurs due to the user's hand movement.

Additionally, current systems often depend on low-cost monocular cameras, which, while accessible, lack depth information and are susceptible to lens distortion, motion blur, and limited field of view [1]. These limitations reduce the spatial consistency of the captured cues, especially when users are at varying distances from the sensor or when movement is rapid and unconstrained. As a result, models trained under one configuration may fail when utilized under different camera orientations or user positions.

From a modeling standpoint, many visual navigation pipelines still rely on global or hand-crafted feature representations, which struggle to capture local semantic variations across different cue configurations. This is particularly problematic when the same visual structure is used to convey multiple classes under varying conditions, leading to significant intra-class variation. While dimensionality reduction techniques like PCA or LDA can improve computational efficiency, they do not retain neighborhood structure or spatial relationships critical to recognizing structurally similar but semantically distinct cues.

Furthermore, existing manifold learning-inspired approaches, including standard LPP and its linear extensions, often assume ideal imaging conditions and use Euclidean distance-based similarity graphs that fail to adapt to the nuanced photometric and spatial distortions present in real-world scenes [61]. These graph structures do not account for varying feature density, misalignment, or spatial uncertainty, which limits their effectiveness when visual cues are partially degraded or presented inconsistently.

Collectively, these gaps highlight the need for more photometrically resilient, spatially aware, and graph-adaptive models capable of accurately interpreting symbolic visual commands under varying and unpredictable sensing conditions. The solution space requires subspace learning techniques that can integrate local visual structure, channel-specific features, and robust similarity modeling, to generalize effectively across diverse indoor navigation scenarios.

These motivations provide the basis for the more advanced methods discussed in subsequent chapters.

1.5.2 Gaps in Robust EMG Feature Modeling

Surface electromyography (sEMG) has emerged as a compelling modality for decoding user intention in human-robot interaction (HRI), particularly in the context of hand movement recognition and assistive robotic control [2]. However, the inherently noisy, user-dependent, and non-stationary nature of sEMG signals presents significant challenges for consistent and reliable classification. While various feature extraction and dimensionality reduction techniques have been proposed earlier, achieving robustness across users, sessions, and environmental conditions remains an open problem.

A major challenge in sEMG-based systems is variability across sessions and subjects. Differences in electrode placement, skin impedance, and muscle morphology lead to considerable variation in the statistical properties of EMG signals [54]. Features that perform well for one individual or session, may degrade rapidly when the sensor is repositioned or when applied to another subject. Even subtle shifts in the position or orientation of the electrodes can affect signal quality and introduce crosstalk from neighboring muscles.

Another persistent issue is the presence of signal noise and motion artifacts, which further complicates feature modeling [56]. Movements of the limb or displacement of the sensor can introduce low-frequency artifacts that mask the underlying motor unit activity. Traditional feature representations such as root mean square (RMS) value, zero-crossing rate, or waveform length are simple to compute but often lack the discriminative power required to distinguish between gestures that engage overlapping muscle groups [57]. This is especially problematic in tasks involving fine motor control, where subtle differences in muscle activation must be preserved despite the background noise.

Dimensionality reduction methods like PCA and LDA have been employed [55] to compress feature space and improve class separability, but these techniques are sensitive to outliers and do not preserve the local neighborhood structure inherent in EMG data. Moreover, their assumptions of linearity and global data structure make them ill-suited for capturing the complex, non-linear patterns that arise in real-world EMG signals, particularly when data is collected across different postures, force levels, or fatigue states.

While locality preserving projection (LPP) and related manifold learning-inspired approaches offer a promising alternative by capturing local geometric relationships, they too exhibit limitations. Most notably, LPP relies on fixed graph construction based on Euclidean distances, which may not accurately represent the similarity structure of high-dimensional EMG features. This is problematic when signal magnitudes vary due to inter-subject or inter-session differences, as the graph may fail to reflect true physiological similarity, leading to poor generalization [70].

Furthermore, many LPP variants do not account for uncertainty, feature noise, or outlier presence in their graph formulation [60]. As a result, they are vulnerable to even minor fluctuations in signal quality and cannot dynamically adjust the affinity matrix to accommodate shifts in the data distribution. Although recent works have explored the use of orthogonal projection constraints and alternative similarity metrics; such as cosine distance or kernel-based fusion [16], there remains a lack of unified frameworks that integrate these ideas to address the full spectrum of such challenges. In addition, current models often overlook the temporal dynamics of sEMG signals. Feature extraction is typically performed on short, overlapping windows, with little regard for the sequential structure of the muscle activation pattern. This limits the system's ability to capture transitions between gestures or to handle continuous movement intent, which is critical in naturalistic human-robot interaction.

These limitations underscore the need for robust, generalizable, and physiologically informed feature modeling frameworks that can handle noisy, non-stationary, and user-variable sEMG input. Such models should incorporate adaptive graph learning, orthogonality enforcement [9], and metric diversity to improve ruggedness and ensure interpretability. The development of these capabilities forms the basis for subsequent chapters, which explore new directions in subspace learning for EMG-based intention recognition.

1.5.3 Gaps in Real-World Deployment of Locality-based DR Algorithms

Manifold learning-inspired DR techniques, particularly those supported by graph-based formulations such as locality preserving projection (LPP) and its enhanced variants, have gained increasing prominence in the domains of vision sensing and signal processing applications. These methods are designed to uncover the intrinsic low-dimensional structure of high-dimensional data by preserving local neighborhood relationships. While their

effectiveness has been demonstrated in controlled experimental settings, their translation to real-world applications; especially within embedded and wearable human-robot interaction (HRI) [2] systems, remain limited due to several practical and methodological constraints.

One key limitation arises from the rigidity of graph construction mechanisms. Most manifold learning approaches rely on static, distance-based graphs, often constructed using k-nearest neighbors or Gaussian heat kernels that assume smooth local continuity in the data. However, in real-world sensing environments, data distributions are often irregular and non-uniform due to noise, user variability, and environmental interference [11]. A fixed neighborhood structure may misrepresent these underlying relationships, particularly when the features exhibit nonlinear distortions, dynamic range shifts, or missing values. This undermines the graph's ability to capture meaningful locality, reducing the overall quality of the learned subspace.

Moreover, the majority of existing manifold-inspired models do not account for uncertainty or data confidence during graph formation. In wearable systems or mobile robotics, input quality can fluctuate due to motion artifacts, lighting variations, sensor drift, or occlusion [64]. Algorithms that treat all data points as equally reliable are prone to propagate errors in the affinity matrix, leading to unstable projections and degraded classification performance. While some extensions have incorporated alternative similarity measures, such as cosine similarity or kernel-based metrics, many still lack a principled mechanism to weigh or reconfigure the graph adaptively based on visual context or signal quality.

Another persistent challenge lies in the lack of scalability and generalization. Many manifold learning techniques were originally designed for batch-mode operation, where the full dataset is available upfront. In contrast, real-world HRI systems require online adaptability, with the ability to project unseen data in real time [67]. Although LPP and its linear variants do offer an explicit projection matrix, numerous graph construction methods, and non-linear embeddings do not extend naturally to streaming data or require full recomputation when new samples are introduced. This restricts their deployment in continuously operating systems or user-specific adaptation scenarios.

Together, these research gaps suggest that for manifold-inspired methods to become truly viable in practical HRI deployments, they must evolve beyond fixed-graph, single-kernel, and offline formulations toward adaptive, uncertainty-aware, efficient, and interpretable models. The subsequent chapters of this Thesis aim to address these gaps by exploring projection

frameworks that integrate graph regularization, orthogonality constraints, multi-metric fusion, and real-time adaptability within a unified learning structure.

1.6 Research Objectives

The review of current literature and methodologies across vision-based robot guidance and sEMG-based intention recognition, highlights several limitations of existing approaches that hinder the deployment of perception-driven human-robot collaboration (HRC) systems in real-world settings [64]. While symbolic cue-based visual systems offer intuitive interfaces and sEMG enables direct access to neuromuscular intent, both modalities suffer from signal degradation, inter-user variability, and environmental inconsistency. Existing dimensionality reduction techniques, particularly those based on classical global models or static graph learning, often fall short in providing representations that are both discriminative and robust under such practical constraints.

The primary goal of this Thesis is to develop a unified subspace learning framework that enhances the *reliability*, *generalizability*, and *robustness* of multimodal human intent recognition systems under real-world sensing conditions. This includes designing projection models that not only preserve local neighborhood relationships but are also resilient to visual and physiological noise, adaptable to changing input distributions, and interpretable for downstream classification tasks.

The specific research objectives of this Thesis are as follows:

1. To identify the limitations of existing dimensionality reduction techniques, including PCA, LDA, LPP, and their variants, when applied to vision- and sEMG-based data in HRC tasks, especially under noisy, occluded, or dynamically varying environments.

2. To examine the effect of spatial and photometric distortion on the recognition performance of symbolic visual cues, and to evaluate the shortcomings of hand-crafted and global feature extraction methods in capturing structural consistency and class discriminability.

3. To explore the challenges of robust EMG feature modeling, particularly in signal-specific irregularities, and to assess how existing subspace learning methods respond to these non-stationary conditions.

4. To investigate robust subspace projection strategies with an emphasis on preserving discriminative structure and mitigating redundancy in high-dimensional feature spaces, within the context of real-time gesture and hand movement classification.

5. To evaluate the role of adaptive graph construction and metric fusion in improving manifold-inspired learning, by enabling better neighborhood modeling across heterogeneous modalities such as RGB image cues and sEMG signal envelopes.

6. To design and validate a set of robust frameworks that integrate more strict regularization, spatial-feature granulation, and multi-geometry similarity modeling, aimed at improving recognition accuracy across visual and physiological modalities under uncertain sensing conditions.

7. To demonstrate the real-world applicability of the proposed methods through extensive experimental evaluation on datasets acquired under realistic constraints, involving varying lighting conditions, background clutter, muscle fatigue, and user-specific gesture variations.

By fulfilling these objectives, the Thesis aims to contribute to the development of scalable, interpretable, and sensor-independent subspace learning models suitable for next-generation collaborative robotic systems that operate reliably outside of controlled laboratory environments.

1.7 Key Contributions

This Thesis addresses the core challenges of human intention recognition in multimodal human-robot collaboration (HRC) systems by proposing a set of novel and robust subspace learning frameworks. The work spans both *vision-based symbolic cue interpretation* and *sEMG-based hand activity recognition*, targeting issues such as sensor noise, inter-subject variability, photometric distortion, and structural ambiguity in high-dimensional data. The contributions are built upon a detailed analysis of the limitations in existing dimensionality reduction techniques and are validated under diverse real-world constraints. The key contributions of this research are summarized below:

1. A novel vision-based symbolic cue recognition scheme is conceptualized and implemented, wherein structured flag-stick gestures are used to guide mobile robotic platforms in indoor

environments. This system enables intuitive and low-cost interaction, with a focus on cue interpretability and semantic clarity under variable sensing conditions.

2. A robust feature projection model, *Adaptive Locality Preserving Spatial Kernel-Guided Bilateral LPP (ALPSK-BLPP)* is proposed to address moderate illumination variability in visual sensing environments. By integrating spatial, feature kernels, and adaptive graph learning, the model enhances locality preservation and class separability under photometric distortions.

3. A suite of local texture encoding-based models, *Local Tetra Pattern-based BLPP (LTrP-BLPP)*, *Histogram Refined Local Binary Pattern-based BLPP (HRLBP-BLPP)*, and *Histogram Refined Local Ternary Pattern-based BLPP (HRLTP-BLPP)* are introduced for enhanced feature extraction under regular to moderately challenging illumination conditions. These models effectively preserve textural consistency while suppressing visual noises, resulting in improved recognition performance under sensor degradation.

4. A *Granular Computing-aided 2DLPP framework (REGF-2DLPP)* is developed for robust visual cue recognition in severely challenging illumination conditions. The model leverages rough entropy-based granulation and intelligent feature decoding to retain relevant spatial information even under low-light and distorted conditions. This model is further enhanced to a *Cosine Similarity-based Variant (REGF-c-2DLPP)*, that works particularly well when the color and shape variation of the visual marker is more pronounced.

5. A *Multichannel Granular and Fuzzy Decoding-based robust model (dNG-2DRLPP)* is proposed by embedding density-based neighborhood granulation and fuzzy feature encoding within a two-dimensional robust projection space. This scheme is shown to be highly effective under severe photometric degradation and external noise, making it suitable for deployment in embedded robotic vision systems.

6. A novel sEMG-based hand activity recognition scheme is conceptualized, and designed to support real-world human-machine interface and human-robot interaction scenarios. The framework enables reliable decoding of muscular intent for applications such as robotic manipulator control, upper-limb prosthetic actuation, and rehabilitation assistance. By mapping sEMG signal patterns to discrete motor actions, the system facilitates intuitive, user-driven control in wearable or assistive platforms.

7. An uncertainty-aware graph learning model, *Uncertainty-Aware Bayesian Model Averaging-based OLPP (UaBMA-OLPP)* is presented to handle inherent irregularities in sEMG signals. The method integrates geometric diversity through multiple similarity measures and performs adaptive graph fusion for stable low-dimensional representations.

8. A possibilistic and sparse projection model, *Robust Possibilistic Neighborhood Graph-aided OLPP (RPNG-OLPP)* is introduced to jointly address signal irregularity and external noise corruption in sEMG feature modeling. The model employs $\ell_{2,1}$-norm-based sparsity and possibilistic weighting to improve generalization across sessions and users.

These contributions collectively advance the state-of-the-art in subspace learning for multimodal human intention recognition and lay the foundation for resilient, adaptive, and deployable HRC systems capable of functioning under practical constraints.

1.8 Thesis Outline

The core contributions of this Thesis, as outlined in the previous section, form the basis for the subsequent chapters. The following section provides an overview of the Thesis structure and describes how each chapter is interconnected.

- ✓ **Chapter 2** presents the conceptualization and implementation of a novel vision-based robot navigation guidance scheme using symbolic human-supervised, semantic cues. The chapter discusses the design of a flag-stick visual marker system for intuitive robot guidance, along with the visual protocol definition and rule-based framing. It also details the data acquisition framework, including the imaging setup, participant instructions, cue placement strategies, and environmental settings used for building a custom visual marker/gesture dataset under varied illumination conditions.

- ✓ **Chapter 3** elaborates on the development and evaluation of the *Adaptive Locality Preserving Spatial Kernel-Guided Bilateral LPP (ALPSK-BLPP)* model, a projection framework designed to strengthen visual cue recognition under moderately challenging illumination conditions. This chapter introduces a joint spatial-feature guided learning strategy, where the spatial affinity between image regions is combined with local feature similarity to construct a more reliable similarity graph. The model adaptively

tunes the neighborhood graph using a learned kernel weighting mechanism, allowing it to dynamically emphasize more discriminative spatial relationships and de-emphasize unreliable or noisy connections.

✓ **Chapter 4** focuses on the development of a class of local texture encoding-based feature projection models, specifically *Local Tetra Pattern-based BLPP (LTrP-BLPP)*, *Histogram-Refined Local Binary Pattern-based BLPP (HRLBP-BLPP)*, and *Histogram-Refined Local Ternary Pattern-based BLPP (HRLTP-BLPP)*. These models are designed to address the limitations of global feature representations under regular to moderately distorted visual sensing conditions, such as mild illumination changes, background clutter, and sensor noise. The core motivation stems from the observation that local texture patterns, when extracted and encoded effectively, carry discriminative structural information that remains relatively invariant to photometric noise and background inconsistencies.

✓ **Chapter 5** presents two granular computing-based dimensionality reduction frameworks, *Rough Entropy-based Granular Fusion-aided 2DLPP (REGF-2DLPP)* and its *Cosine Similarity-augmented variant REGF-c-2DLPP*, specially designed for visual cue recognition under severely degraded illumination conditions and marker variability. The granular computing framework allows the system to isolate structurally stable regions from noisy or low-informative areas of the images. These methods address the limitations of conventional LPP variants in retaining relevant structural information when visual data is corrupted by poor lighting, background ambiguity, or visual distortion in the shape and color of symbolic cues.

✓ **Chapter 6** introduces the *Density-based Neighborhood Granulation-aided 2D robust LPP (dNG-2DRLPP)* model; a density-based neighborhood granulation approach inside the framework of 2D robust locality preserving projection, designed to address extreme photometric degradation, structural noise, and visual ambiguity in vision-based HRI systems. This chapter builds upon the limitations observed in earlier LPP variants by proposing a novel granular computing strategy that tightly integrates spatial structure with local feature saliency under uncertainty. The model dNG is especially highlighted in this study as unlike conventional granulation models which assume uniform or regular granule formation, the dNG strategy allows the model to adaptively prioritize stable, texture-rich regions of the visual cue while suppressing less informative, noisy, or

background-dominated areas.

✓ **Chapter 7** explores the alternate interaction vertical i.e., wearable physiological sensing, introducing a novel sEMG-based hand activity recognition framework tailored for real-world human-machine interface and human-robot interaction (HMI/HRI) applications. Unlike traditional gesture recognition approaches that rely on camera-based inputs or surface-level features, this framework is built on the premise of directly decoding neuromuscular activation patterns from surface electromyography (sEMG) signals to infer motor intent with high temporal precision and robustness. The proposed scheme is conceptualized to facilitate naturalistic and intuitive interaction in practical scenarios such as robotic manipulator control, prosthetic actuation, and rehabilitation feedback systems.

✓ **Chapter 8** presents the *Uncertainty-Aware Bayesian Model Averaging-based OLPP (UaBMA-OLPP)* framework, developed to improve the robustness and generalizability of sEMG-based hand activity recognition in the presence of inter-subject variability and signal irregularity. This model addresses a core challenge in wearable physiological sensing, that the same hand activity can yield substantially different signal patterns across users and sessions due to anatomical differences, muscle fatigue, electrode placement shifts, or varying force levels. The key innovation in UaBMA-OLPP lies in its integration of multi-geometry similarity modeling through a novel Bayesian model averaging (BMA) approach.

✓ **Chapter 9** illustrates the *Robust Possibilistic Neighborhood Graph-aided OLPP (RPNG-OLPP)* framework, designed to improve the resilience of sEMG-based subspace learning against both internal signal irregularities and external noise contamination. While conventional locality preserving projection (LPP) models are effective in preserving neighborhood structure under ideal conditions, their performance tends to degrade when the input signal is corrupted by sensor noise, muscle crosstalk, or varying contraction effort levels; a common occurrence in real-world wearable settings. The core contribution of RPNG-OLPP lies in its possibilistic graph construction strategy, by introducing possibilistic degrees that reflect the confidence with which two samples are considered neighbors. Additionally, RPNG-OLPP employs an $\ell_{2,1}$-norm regularized OLPP formulation, which enhances projection sparsity and robustness.

✓ **Chapter 10** concludes the Thesis by summarizing the key findings, synthesizing

cross-domain insights, and evaluating the effectiveness of the proposed subspace learning models across both vision-based and sEMG-based human-robot interaction (HRI) systems. The chapter reflects on the robustness, adaptability, and interpretability of the developed frameworks under real-world sensing constraints. The chapter concludes by outlining promising directions for future research, emphasizing the potential for extending LPP-based techniques toward dynamic, scalable, and context-aware applications in next-generation wearable and robotic systems.

Chapter 2

Visual Symbolic Cue-based System for Robot Guidance Applications

2.1 Introduction

In recent years, vision-based systems have gained significant prominence in human-robot interaction (HRI) due to their ability to support intuitive, contactless communication. These systems are particularly useful in indoor, semi-structured environments where physical constraints and user convenience must be considered with higher priority [51]. In the earlier days, traditional industrial applications e.g., assembly tasks, used to rely on fixed and pre-defined instructions. With the evolution of assistive and collaborative robotics, HRI applications now demand dynamic interaction schemes that adapt to human inputs in real time.

A critical consideration of such interaction frameworks is, which sensor modalities are involved in capturing the human intent. Wearable sensing modalities such as inertial measurement units (IMUs), haptic gloves, surface electromyography (sEMG), magnetic trackers, or pressure-sensitive sensors have been widely explored for their ability to provide detailed motion and physiological information [22]. However, these contact-based systems often pose practical constraints. Their reliance on direct attachment to the human body introduces discomfort, restricts natural movement, and may reduce user compliance over extended periods. Moreover, the cost, maintenance requirements, and operational hazards of such sensors can limit their feasibility in scalable or resource-constrained deployments.

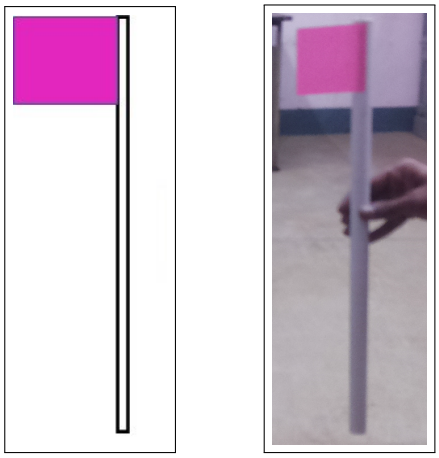
To address these limitations, vision sensors offer a non-contact alternative that is both cost-effective and user-friendly. A wide range of visual strategies has been explored for human-robot interaction, including natural hand gestures, full-body postures, and facial expressions, as well as artificial visual markers [71]. While gesture and posture-based systems

provide intuitive interaction, they often suffer from variability in human appearance, articulation speed, and viewpoint alteration, which can significantly affect recognition accuracy and system reliability. In contrast, artificial visual markers such as symbolic patterns or structured objects offer a controlled and repeatable alternative. These markers can be designed with well-defined geometries, colors, and orientations, which help to reduce ambiguity and enhance the robustness of recognition under varied environmental conditions. In this context, this Thesis proposes a vision-based symbolic cue framework, wherein human participants use hand-guided flag-stick combinations to communicate navigational guidance commands. Each marker configuration corresponds to a distinct directional instruction, enabling the robot to interpret spatial intent in a structured, non-verbal, and contactless manner.

The experimental setup consists of a low-cost, differentially driven mobile robot equipped with a Raspberry Pi 3 Model B+ processing unit and an RPi Camera Module v2, incorporating an 8-megapixel Sony IMX219 vision sensor [72]. The system has been developed and deployed in the Instrumentation & Cyber Physical System Laboratory, Electrical Engineering Department at Jadavpur University, Kolkata, India [1], [49]. Visual cue data are collected through supervised sessions in which participants present various flag-stick markers within the field of view of the robot's onboard camera. The visual markers present seven discrete cues: *Top-left* (TL), *Middle-left* (ML), *Bottom-left* (BL), *Neutral* (N), *Top-right* (TR), *Middle-right* (MR), and *Bottom-right* (BR), which are designed to represent corresponding directional commands, spanning a 180° angular range.

To reflect real-world photometric challenges in the acquired dataset, experiments are conducted under varied photometric conditions by systematically changing the ambient illumination. Additionally, multiple volunteers of different ages and genders contribute to the dataset, introducing natural variations in hand position, orientation, and cue presentation. These considerations help create a comprehensive and diverse visual database that covers the real-world variability in human-robot interaction scenarios.

This chapter details the design of the symbolic visual cue system, the development of the robot guidance protocol, and the environmental and procedural considerations involved in acquiring the visual dataset. The resulting data serve as the foundation for subsequent analysis, including the development of robust cue recognition strategies under several environmental constraints, explored in subsequent chapters.



(a) (b)

Figure 2.1: (a) A schematic design illustrating the prototype structure of the flag-stick marker. (b) Real-world implementation of the marker held by a participant during robot guidance tasks.

2.2 Design of the Flag-Stick Visual Marker System

The flag-stick visual marker system is developed as a symbolic interface for facilitating human-to-robot interpretable communication in an indoor environment. The core idea is to utilize a simple, low-cost physical object; a vertical stick mounted with a flag, as a visual marker, which can be easily manipulated by human users to indicate directional cues. Each marker consists of a rigid stick approximately 30–40 cm in length, with a planar flag affixed at one end. A schematic prototype of the flag-stick marker and its practical deployment are shown in Figs. 2.1a and 2.1b, respectively. The marker is handheld and freely maneuvered by the user to generate a set of distinguishable spatial configurations. These configurations are designed to occupy different relative positions within the robot's field of view and are interpreted by the robotic system as distinct cue categories.

The use of a stick-and-flag combination offers several practical advantages. It ensures consistent visual geometry, and maintains a clear separation between the hand and the marker, thereby improving the segmentation. It also reduces the likelihood of misclassification due to hand posture variability which is quite common in human-body gesture-based systems. Furthermore, the simplicity of the hardware design supports ease of replication across users without specialized training or instrumentation.

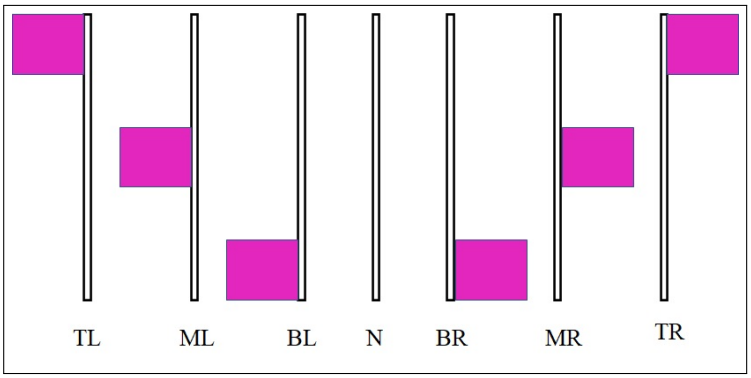


Figure 2.2: Sample visual cue templates corresponding to the TL, ML, BL, N, BR, MR, and TR classes.

Seven discrete orientations of the marker are defined here as valid cues. These spatial arrangements correspond to relative placements of the flag-stick within the left, center, and right zones of the camera frame, across top, middle, and bottom vertical regions. This spatial discretization forms the basis of the symbolic cue vocabulary used for subsequent robot guidance.

2.3 Robot Guidance Protocol and Cue Semantics

To translate and operationalize the flag-stick cues into actionable robot movements, a semantic protocol was developed that links each visual marker orientation with a corresponding directional command. The symbolic cue system comprises seven unique spatial classes: top-left (TL), middle-left (ML), bottom-left (BL), neutral (N), bottom-right (BR), middle-right (MR), and top-right (TR). These cues are interpreted as direction indicators, enabling the robot to align its trajectory accordingly. To illustrate the cue design more clearly, Fig. 2.2 presents the prototype templates for all seven flag-stick orientations, where PA denotes the principal axis of the camera's field of view.

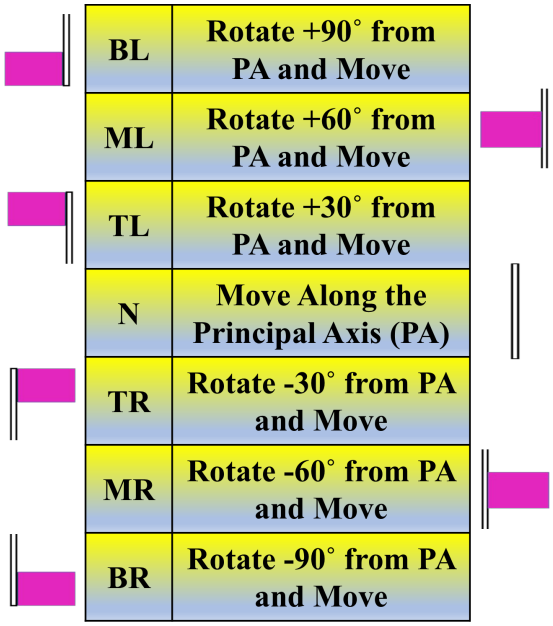


Figure 2.3: A lookup table linking human-guided visual cues to their respective robot navigation commands.

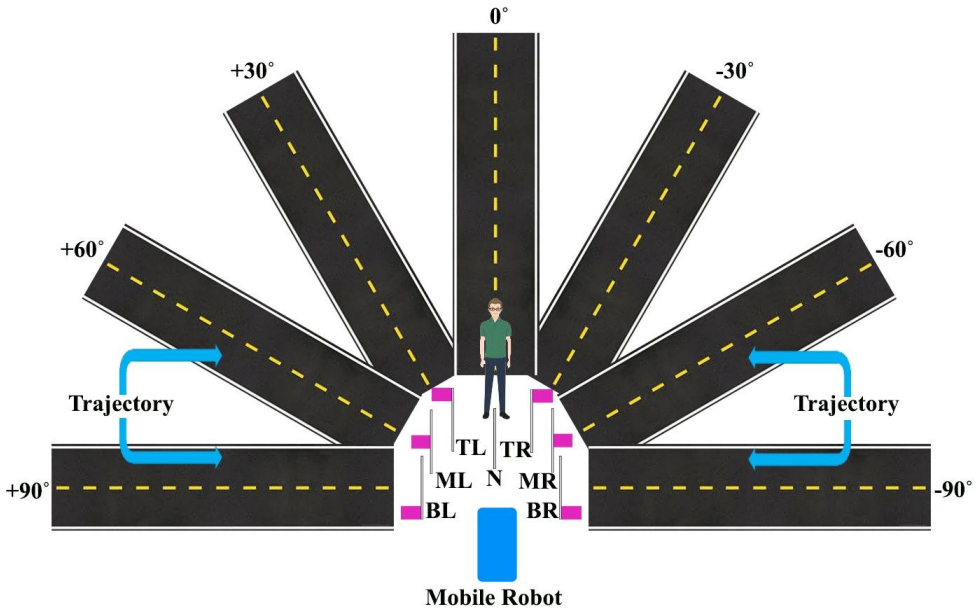


Figure 2.4: Prototype navigation map showing the robot's predefined movement directions based on visual cues.

The neutral (N) position represents a forward movement along the robot's principal axis (PA) of the camera. The other cues are mapped symmetrically about this central direction, forming a span of $\pm 90^{\circ}$ in angular displacement. Specifically: TL, ML, and BL correspond to

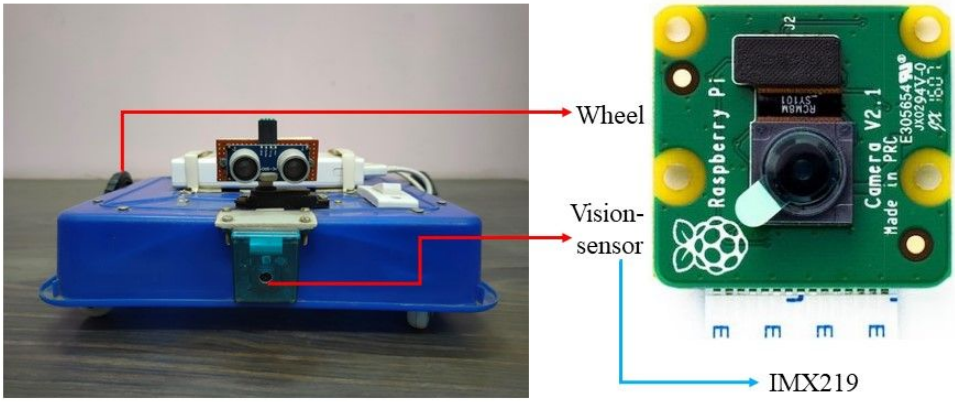


Figure 2.5: The differentially driven wheeled mobile robot utilized in the experiments and its associated Sony IMX219 vision sensor module.

angular deviations of $+30°$, $+60°$, and $+90°$, respectively, to the left of the neutral direction. Similarly, TR, MR, and BR respectively correspond to $-30°$, $-60°$, and $-90°$ deviations to the right. Fig. 2.3 depicts the corresponding navigation guidance instructions associated with these visual cues, in the form of a lookup table.

This mapping allows the robot to execute seven distinct motion directions in response to the observed cue. During runtime, the robot classifies the incoming visual cue and accesses a lookup table that translates the class into a pre-defined angular displacement command. This semantic mapping ensures that the robot can react deterministically to user-intended navigational triggers. A prototype trajectory map depicting all feasible regulated directions for robot navigation is shown in Fig. 2.4.

2.4 Data Acquisition Framework and Environmental Settings

The entire system was deployed and tested in an indoor semi-structured environment within the Instrumentation & Cyber Physical System Laboratory of the Electrical Engineering Department, Jadavpur University. A mobile robot platform was used, comprising a Raspberry Pi 3 Model B+ as the onboard processing unit, paired with an RPi Camera Module v2 containing an 8-megapixel Sony IMX219 sensor [72]. The differentially driven wheeled mobile robot with the vision sensor system is shown in Fig. 2.5. The camera provided a forward-facing view

Table 2.1: AVERAGE ILLUMINATION INTENSITY RECORDED ACROSS DIFFERENT LIGHTING SCENARIOS

Illumination Condition	**Average Illuminance (Lux)**
NIL	14.910
DIL-1	13.405
DIL-2	11.880
DIL-3	8.500
DIL-4	4.710

aligned with the robot's principal axis and captured all human-supervised cues within its field of vision.

Human participants from various age groups were invited to perform the visual cues. They held and presented the flag-stick structure in front of the robot, either in a standing or seated posture. No constraints were imposed on the users regarding their position, orientation, or distance from the camera, as long as the cue remained fully visible in the camera's frame. This ensured the dataset captured a range of naturalistic variations in spatial arrangement and hand articulation.

To evaluate the performance of the system under different environmental conditions, data were recorded across five illumination levels: one standard condition (*Normal Illumination Level* or NIL) and four progressively darker setups (*Dark Illumination Level*, DIL-1 through DIL-4). An initial set of data has been obtained under the normal illumination conditions inside the laboratory room which encompasses four sets of artificial illumination sources. Furthermore, dark level-1, dark level-2, dark level-3, and dark level-4 photometric conditions are generated in the presence of three, two, one, and no artificial light sources, respectively. Illumination levels were quantitatively measured using a Metravi 1310 digital illuminance meter. The variation of natural and artificial photometric levels over a certain period of the day is shown in 2.6, recorded with the illuminance meter. The average illuminance levels recorded during data acquisition under each photometric condition are listed in Table 2.1. These conditions closely resemble real-world photometric variations commonly observed in indoor environments.

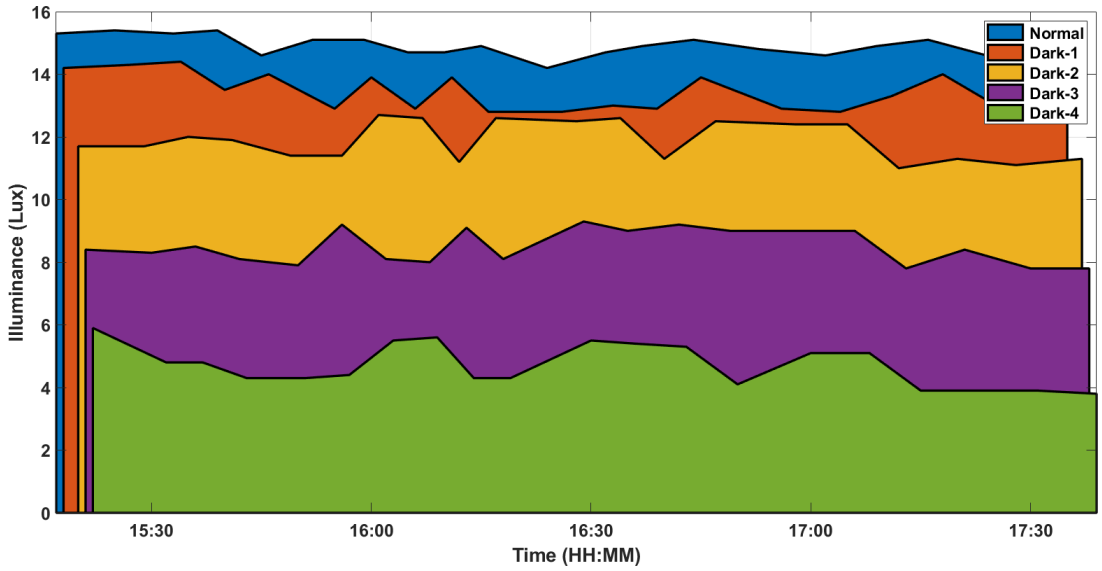


Figure 2.6: Variation of daylight illuminance and artificial illuminance at photometric levels NIL, DIL-1, DIL-2, DIL-3, and DIL-4.

2.5 Description of Acquired Visual Dataset

The visual dataset collected in this study forms the empirical foundation for the recognition and analysis modules developed in subsequent chapters. The primary dataset comprises seven visual cue classes, each represented by flag-stick orientations performed by human users. Each cue class consists of 150 unique images, captured under regular lighting condition NIL, resulting in a total of 1050 images for the total 7 categories. In a similar manner, testing images have been acquired under four different conditions DIL-1, DIL-2, DIL-3, and DIL-4, from the same participants, keeping other environmental settings unchanged. The camera resolution was fixed, and minimal preprocessing was applied to preserve the authenticity of the raw data. The class labels were cyclically assigned as TL, ML, BL, N, BR, MR, and TR. The visual variabilities in the dataset are inherent from factors such as hand positioning, camera distance, and lighting changes, reflecting real-world HRI environments. One set of sample visual cues from all seven classes captured under five distinct lighting conditions is presented in Fig. 2.7.

The inclusion of artificial marker variations addresses human-induced inconsistencies in gesture-based interfaces. Unlike hand gestures, which may vary in scale, articulation, and orientation, between users, fixed-shaped and consistently colored visual markers provide a more stable and repeatable recognition target. This design choice enhances inter-user consistency, particularly in uncontrolled environments or applications involving non-expert users.

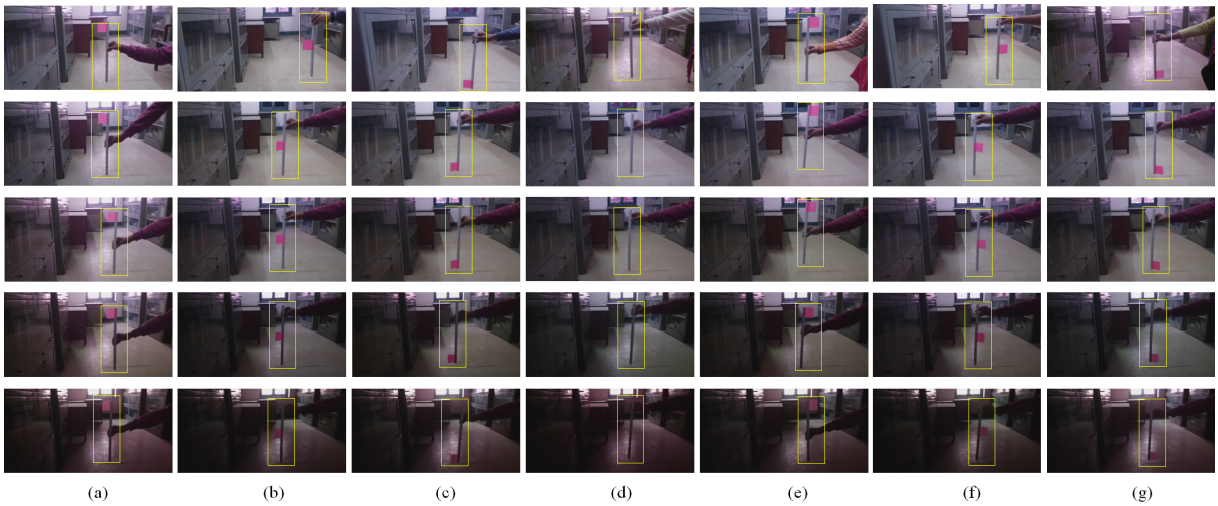


Figure 2.7: Representative real-world visual cues at normal illumination conditions (top row), dark level-1, dark level-2, dark level-3, and dark level-4 (bottom row) from the classes (a) TL. (b) ML. (c) BL. (d) N. (e) TR. (f) MR. (g) BR.

2.6 Summary

This chapter presented the design, implementation, and deployment of a vision-based symbolic cue system for mobile robot guidance in human-robot interaction contexts. A flag-stick marker interface was developed to translate spatial hand gestures into discrete navigational commands. The system design included a clear mapping between cue semantics and robot trajectory, realized through a structured seven-class orientation model. Here, data acquisition was performed using a Raspberry Pi-based mobile robot setup across five controlled lighting conditions, resulting in a comprehensive dataset comprising 1050 labeled images at NIL and test images at DILs. This dataset serves as the empirical basis for the recognition and projection techniques discussed in subsequent chapters.

Chapter 3

Adaptive Spatial Kernel-Guided LPP for Illumination-Robust Cue Recognition

3.1 Introduction

As human-robot collaboration (HRC) becomes increasingly embedded in real-world applications, the need for intuitive, reliable, and context-aware interaction frameworks has grown significantly. Among various sensing modalities, vision-based systems have gained prominence due to their non-contact nature and ease of utilization in indoor and semi-structured environments. Symbolic visual cue systems, such as those based on structured markers, offer a practical interface for communicating spatial commands to robots. However, their performance is often hindered by real-world challenges [73] such as illumination variability, background clutter, and visual ambiguity.

While the previous chapter established the empirical foundation and dataset description, the current chapter focuses on the algorithmic advancement required for robust cue recognition. Traditional global dimensionality reduction techniques are often inadequate in preserving the intrinsic local structures of high-dimensional visual data, particularly under photometric degradation. Locality preserving projections (LPP), a graph-based approach for dimensionality reduction with the linear approximation of a nonlinear manifold, addresses this limitation by retaining local neighborhood information during projection. Hence, this chapter first introduces an LPP-based visual cue detection framework, highlighting its advantages over global methods in terms of classification performance and subspace preservation [74].

However, as observed, standard LPP suffers from certain drawbacks when applied to real-world HRI scenarios. Its similarity graph relies solely on fixed spatial proximity based on Euclidean distance, which can be highly sensitive to noise, illumination shifts, and geometric

distortions. Moreover, it does not incorporate the range-domain (feature intensity) information [11] that often carries meaningful semantic cues under deteriorated lighting. To overcome these limitations, this chapter introduces a more robust spatial similarity measure; namely, the *Adaptive Locality Preserving Spatial Kernel* (ALPSK) and its corresponding dimensionality reduction algorithm ALPSK-BLPP, which jointly optimizes spatial similarity using adaptive graph learning and incorporates feature-based photometric similarity through bilateral filtering.

The ALPSK model learns a class-discriminative projection by dynamically adjusting spatial weights and enforcing margin-based constraints, while also selecting informative features via row-sparsity regularization. It is further extended to ALPSK-BLPP, a bilateral LPP scheme where both adaptive spatial and feature-domain similarities are fused into a unified kernel-guided projection framework. This progression from the baseline LPP to a more resilient ALPSK-BLPP, forms the core of this chapter, both conceptually and experimentally.

Through extensive evaluations under varying illumination conditions, the proposed schemes demonstrate significant improvements in classification accuracy and robustness. This chapter thus establishes the efficacy of adaptive kernel modeling in symbolic visual cue recognition, laying the groundwork for more generalizable perception modules in HRI systems.

3.2 LPP-Guided Baseline Visual Cue Detection Scheme

In the context of real-life vision-sensor-enabled human-robot interaction systems, the robust extraction of semantically meaningful visual cues becomes critical, especially under dynamically changing, unstructured, and photometrically adverse environments. The proposed visual cue detection scheme addresses these challenges by employing locality preserving projections (LPP) as the core of its feature extraction strategy, enabling the system to encode intrinsic local structures of the input data into a lower-dimensional manifold. To the best of available knowledge from the literature, the application of LPP in the context of human-supervised artificial visual marker-based robot navigation guidance has not been previously reported. This Thesis presents a novel flag-stick-based supervision strategy for collaborative robot navigation, employing an LPP-driven feature extraction and dimensionality reduction framework. The proposed approach is designed to operate effectively under real-world conditions and can be generalized to other vision-guided human-robot interaction

(HRI) systems [51].

3.2.1 Problem Formulation

In real-world human-robot interaction systems, vision-sensor-based approaches offer a non-intrusive and cost-effective alternative to wearable sensor technologies for capturing human intent. However, such systems often operate under dynamically changing and unstructured surroundings, where visual data is frequently affected by hindrances, such as poor illumination, background clutter, and low image clarity. These adversities significantly impede the extraction of reliable and discriminative features from the acquired images. Conventional feature extraction techniques that rely on global geometric information, such as principal component analysis (PCA), often fail to capture the local structural characteristics essential for distinguishing between similar visual patterns under such degraded conditions [13]. This necessitates the use of feature extraction methods that are sensitive to the intrinsic local manifold structures of the image data.

To address this issue, the problem is formulated as a vision-sensor-based real-life visual cue recognition task for human-supervised robot navigation. The objective is to extract compact and discriminative representations from visual inputs that encode different orientations of a flag-stick marker held by the human supervisor. Each orientation corresponds to a specific navigational command for the mobile robot. The challenge lies in developing a robust and efficient dimensionality reduction scheme that preserves critical local information while enabling high classification accuracy under real-world constraints.

Locality preserving projections (LPP) is adopted in this context as a graph-based embedding technique that preserves local neighborhood structures within the data. The proposed solution aims to construct a reliable mapping from the high-dimensional image space to a low-dimensional feature space, where class separability is retained. This enables effective classification of visual cues into discrete navigation commands using a suitable classifier, thereby facilitating accurate robot guidance in collaborative human-robot interaction scenarios. The complete system architecture is illustrated in the form of a block diagram in Fig. 3.1.

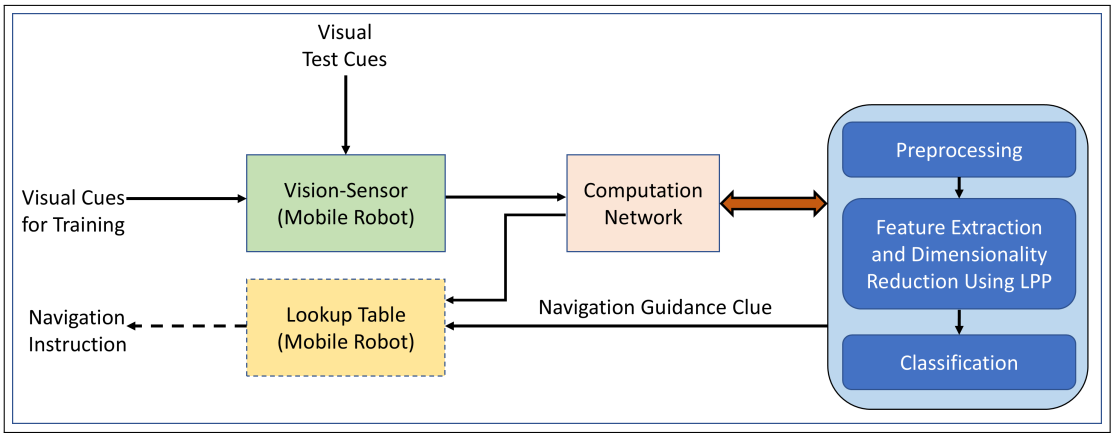


Figure 3.1: A block diagram illustrating the architecture of the proposed LPP-based visual cue detection and robot navigation guidance system.

3.2.2 Preprocessing and Lower-Dimensional Projection

The acquired dataset used in this study consists of practical images capturing distinct orientations of a flag-stick combination, each corresponding to a specific navigation command for the robot. These visual cues were recorded under varying environmental conditions throughout the day and other ambient irregularities. A detailed description of the dataset acquisition, environmental settings, and class distribution has already been provided in Chapter 2. To prepare the images for feature extraction, a minimal preprocessing step is applied. Each RGB image is resized to a lower resolution to reduce computational overhead and then converted to grayscale to simplify the visual information while retaining the essential structural patterns. It ensures that the feature extraction algorithm operates on a compact and consistent representation of the data, thereby improving efficiency without compromising performance.

After preprocessing, each two-dimensional grayscale image matrix is reshaped into a one-dimensional vector $\mathbf{a}_i \in \mathbb{R}^D$ and subsequently projected into a lower-dimensional space $\mathbf{y}_i \in \mathbb{R}^d$ using the LPP projection matrix $\mathbf{W} \in \mathbb{R}^{D\times d}$. LPP is a graph-based embedding technique that constructs an affinity graph from the original high-dimensional space $\mathbb{R}^D$ and preserves its structure in the lower-dimensional manifold $\mathbb{R}^d$. This method ensures that the local neighborhood relationships among data points are maintained after projection. As a result, the key geometric characteristics of the visual features are retained even in the reduced space, allowing for a significant decrease in dimensionality, while preserving the intrinsic properties of the data. Thus, the computational burden of subsequent classification tasks is

substantially reduced, and the development of efficient human-supervised robot navigation systems is facilitated.

Once the lower-dimensional features are extracted using the LPP projection, the next step involves classifying the visual cues into one of the predefined navigation classes. Various machine learning-based classification algorithms such as k-nearest neighbors (KNN) [75], artificial neural network (ANN) [76], and support vector machine (SVM) [77] can be employed for this purpose. In this work, a multiclass support vector machine (MSVM) has been chosen, owing to its robustness and effectiveness in handling high-dimensional data. While a standard SVM is inherently suitable for binary classification, it can be extended to multiclass scenarios by decomposing the original problem into multiple binary sub-problems. Several decomposition strategies are available, such as one-vs-one (OVO) [78], one-vs-all (OVA) [79], and directed acyclic graph (DAG) [80] schemes. In the present work, the OVO approach is adopted, where a separate binary classifier is trained for every possible pair of classes, and the final prediction is made through a majority voting mechanism.

3.2.3 Mathematical Formulation of LPP

Locality preserving projections (LPP) is a linear, unsupervised dimensionality reduction technique that aims to preserve the intrinsic local geometric structure of the data during the transformation from a high-dimensional to a low-dimensional space [7].

Let the training dataset be represented as $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N] \in \mathbb{R}^{D\times N}$, where each $\mathbf{a}_i$ is a D-dimensional vectorized image sample. The objective of LPP is to find a transformation matrix $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_d] \in \mathbb{R}^{D\times d}$ that maps the input data into a d-dimensional space ($d \ll D$), yielding the projected data as follows:

$$\mathbf{y}_i = \mathbf{W}^\top \mathbf{a}_i \tag{3.1}$$

To preserve the local similarity between samples, LPP constructs a similarity matrix $\mathbf{S}$, whose elements are defined as:

$$s_{ij} = \begin{cases} \exp\left(-\dfrac{\|\mathbf{a}_i - \mathbf{a}_j\|_F^2}{\tau}\right) & \text{if } \mathbf{a}_j \in \mathcal{N}_k(\mathbf{a}_i) \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

where $||\cdot||_F$ denotes the Frobenius norm, τ is a regulating parameter, and $\mathcal{N}_k(\mathbf{a}_i)$ denotes the k-nearest neighbors of data point $\mathbf{a}_i$. Subsequently, the objective function to be minimized by

LPP is defined as:

$$\min_{\mathbf{W}} \sum_{i,j} \left\| \mathbf{W}^\top \mathbf{a}_i - \mathbf{W}^\top \mathbf{a}_j \right\|^2 s_{ij} \tag{3.3}$$

Introducing matrix notation, let the diagonal matrix $\mathbf{D}$ be defined with entries $D_{ii} = \sum_j s_{ij}$ and the corresponding Laplacian matrix is $\mathbf{L} = \mathbf{D} - \mathbf{S}$. The minimization problem in Equation (3.3) can then be reformulated as:

$$\min_{\mathbf{W}} \ \mathrm{tr}(\mathbf{W}^\top \mathbf{A}\mathbf{L}\mathbf{A}^\top \mathbf{W}) \quad \text{subject to } \mathbf{W}^\top \mathbf{A}\mathbf{L}\mathbf{A}^\top \mathbf{W} = 1 \tag{3.4}$$

Solving this optimization leads to the generalized eigenvalue problem:

$$\mathbf{A}\mathbf{L}\mathbf{A}^\top \mathbf{W} = \lambda \mathbf{A}\mathbf{D}\mathbf{A}^\top \mathbf{W} \tag{3.5}$$

If $\mathbf{A}\mathbf{D}\mathbf{A}^\top$ is non-singular, Equation (3.5) can be equivalently expressed as:

$$(\mathbf{A}\mathbf{D}\mathbf{A}^\top)^{-1} \mathbf{A}\mathbf{L}\mathbf{A}^\top \mathbf{W} = \lambda \mathbf{W} \tag{3.6}$$

The solution to this eigenvalue problem yields the transformation matrix $\mathbf{W}$, which is used to project the original high-dimensional data into a lower-dimensional LPP feature space while preserving local neighborhood information. The algorithm of the entire LPP-based projection scheme is given in Algorithm 3.1.

3.2.4 Experimental Results and Discussion

To evaluate the effectiveness of the proposed LPP-based visual cue detection framework, experiments have been undertaken on the acquired dataset comprising seven distinct visual classes corresponding to different flag-stick orientations. The entire dataset has been split randomly into training and testing sets in a ratio of 7 : 3. The performance has been analyzed in terms of classification accuracy and the corresponding degree of dimensionality reduction.

In the case of LPP, the parameters have been chosen empirically. The number of nearest neighbors k has been set to 6 as per the general guideline given in the earlier literature [11], and the regulating parameter τ used in the construction of the similarity matrix has been chosen from the empirical range $[150, 200]$. The target dimensionality d for the LPP projection has been determined such that the projected subspace retains the maximum possible energy from the original input vectors. The proposed scheme has achieved significant dimensionality reduction without compromising classification performance. A mean accuracy of 99.69% has been attained while reducing the input dimensionality by approximately 97.62%. Table 3.1 presents

Algorithm 3.1 Locality Preserving Projection (LPP)

Input: Training image set $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N] \in \mathbb{R}^{D \times N}$, regulating parameter τ, number of nearest neighbors k, number of LPP basis vectors d

Output: Transformed feature matrix in the LPP subspace $\mathbf{Y}$

STEP-1 Construct the similarity matrix $\mathbf{S}$ using the k-nearest neighbors. Compute s_{ij} as:

$$s_{ij} = \begin{cases} \exp\left(-\dfrac{\|\mathbf{a}_i - \mathbf{a}_j\|_F^2}{\tau}\right), & \text{if } \mathbf{a}_j \in \mathcal{N}_k(\mathbf{a}_i) \\ 0, & \text{otherwise} \end{cases}$$

STEP-2 Compute the diagonal weight matrix $\mathbf{D}$ with entries:

$$D_{ii} = \sum_j s_{ij}$$

STEP-3 Compute the graph Laplacian matrix:

$$\mathbf{L} = \mathbf{D} - \mathbf{S}$$

STEP-4 Solve the generalized eigenvalue problem:

$$\mathbf{ALA}^\top \mathbf{W} = \lambda \mathbf{ADA}^\top \mathbf{W}$$

STEP-5 Select the d eigenvectors corresponding to the smallest non-zero eigenvalues to form the projection matrix $\mathbf{W}$.

STEP-6 Project the input data to the LPP subspace:

$$\mathbf{Y} = \mathbf{W}^\top \mathbf{A}$$

STEP-7 Return $\mathbf{Y}$.

the detailed classification accuracy across various levels of dimensionality reduction using LPP. It is observed that even with a dimensionality reduction of over 98%, the accuracy has remained consistently above 99%, demonstrating the robustness and efficacy of the approach.

For comparison, principal component analysis (PCA) [3] has also been tested on the same dataset. As shown in Table 3.2, PCA has achieved a maximum classification accuracy of 99.05% with a dimensionality reduction of 98.50%. However, with an increase in the reduction level beyond 99% (e.g., 99.18%), its performance has dropped to 86.67%, in contrast to LPP which maintains high accuracy even under significant reduction. This illustrates that LPP better

Table 3.1: CLASSIFICATION PERFORMANCE OF LPP IN TERMS OF ACCURACY (MEAN ± STANDARD DEVIATION) AT VARYING LEVELS OF DIMENSIONALITY REDUCTION

Dimensional Reduction (%)	**Classification Accuracy (%)**
99.92	17.46 ± 0.644
99.84	27.30 ± 0.615
99.60	76.83 ± 0.432
99.44	94.60 ± 0.321
99.21	98.09 ± 0.319
98.81	99.05 ± 0.276
98.41	99.05 ± 0.287
98.02	99.37 ± 0.198
97.62	**99.69 ± 0.109**
97.22	99.69 ± 0.112
96.83	99.37 ± 0.184
96.43	99.37 ± 0.213
96.03	99.05 ± 0.251

Table 3.2: CLASSIFICATION PERFORMANCE OF PCA IN TERMS OF ACCURACY (MEAN ± STANDARD DEVIATION) AT VARYING LEVELS OF DIMENSIONALITY REDUCTION

Dimensional Reduction (%)	**Classification Accuracy (%)**
99.73	28.25 ± 0.598
99.59	55.24 ± 0.511
99.46	71.43 ± 0.419
99.32	80.95 ± 0.377
99.18	86.67 ± 0.382
99.05	92.69 ± 0.316
98.91	96.83 ± 0.295
98.78	98.09 ± 0.270
98.64	98.09 ± 0.254
98.50	**99.05 ± 0.184**
98.37	98.73 ± 0.219
98.23	98.41 ± 0.232
98.10	98.09 ± 0.260

preserves discriminative features essential for classification under reduced dimensions as well. This performance comparison is depicted in Fig. 3.2, in a graphical manner. While PCA shows

Table 3.3: COMPARISON OF CLASSIFICATION ACCURACY AMONG DIFFERENT DIMENSIONALITY REDUCTION TECHNIQUES

Method	Accuracy (%)
PCA	99.05
Randomized PCA	99.05
Incremental PCA	98.41
Kernel PCA	99.05
Sparse PCA	99.37
LDA	99.37
LPP	**99.69**

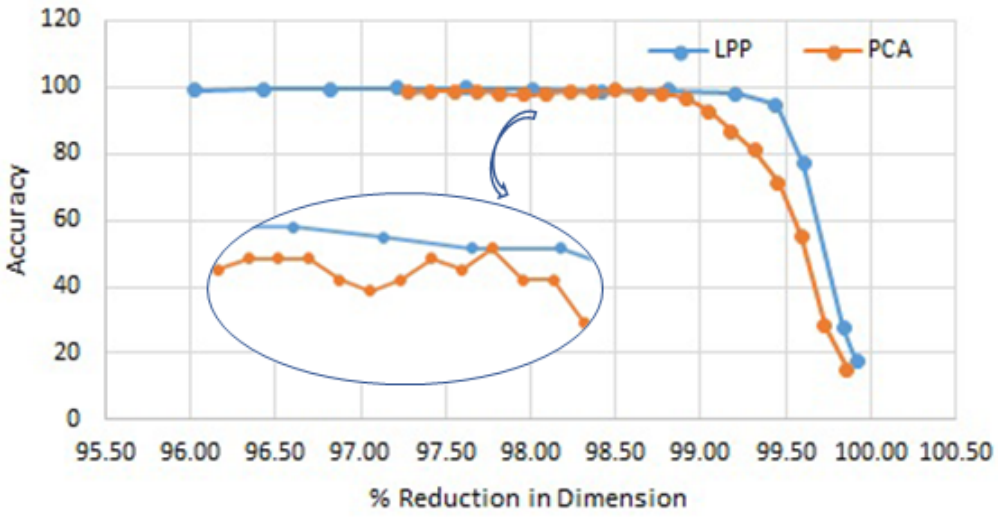
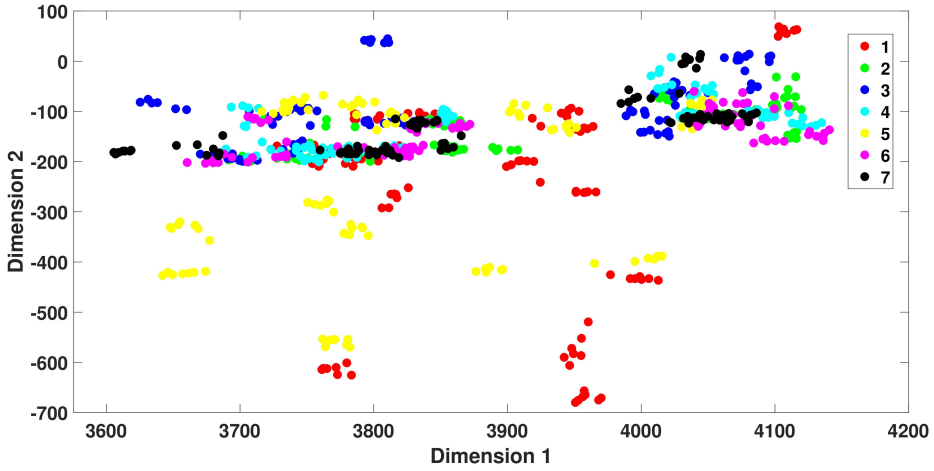


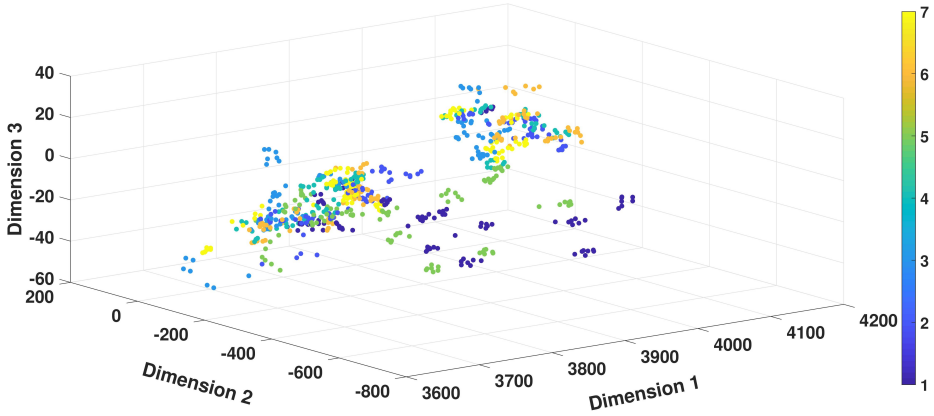
Figure 3.2: Comparison of classification performance between LPP and PCA as a function of increasing dimensionality reduction, highlighting tolerance to dimensional loss.

a rapid decline in accuracy beyond a certain threshold, LPP maintains performance stability, demonstrating better tolerance to dimensionality reduction.

A comparative study with other state-of-the-art dimensionality reduction techniques has also been undertaken. Table 3.3 presents the classification accuracy obtained using LPP [7], PCA [3], randomized PCA [81], incremental PCA [82], kernel PCA [83], sparse PCA [84], and LDA [4]. Among all these methods, LPP has achieved the highest classification accuracy, reinforcing its suitability for the task. Additionally, the feature distributions in the reduced subspace are visualized in Fig. 3.3, where the LPP-transformed features are plotted in both 2D and 3D spaces. The separation between classes is visibly clearer in 3D, contributing to higher classification performance. This further affirms that LPP offers a good trade-off between



(a)



(b)

Figure 3.3: Projection of high-dimensional input data using LPP onto a (a) 2D subspace and a (b) 3D subspace.

dimensionality reduction and feature separability.

3.2.5 Key Insights

In this section, a *locality preserving projection* (LPP)-based framework was presented for visual cue detection in a human-supervised robot navigation system. The proposed method effectively reduces the dimensionality of high-resolution visual data while preserving the local geometric structure, thereby enabling efficient and accurate classification of flag-stick orientation cues in real-world scenarios. Experimental results demonstrated that the LPP-based scheme achieves high classification accuracy even with substantial dimensionality reduction, outperforming

conventional methods such as PCA and LDA in terms of both robustness and tolerance to information loss. This confirms the suitability of LPP for vision-sensor-based human-robot interaction tasks involving structured and moderately varying visual environments.

However, while the proposed framework performs well on the controlled dataset, its robustness may be reduced in practical scenarios, involving severe photometric alterations e.g., diminished illumination. In such tough conditions, the preservation of local structures alone may not be sufficient. Hence, more advanced techniques, such as adaptive spatial kernel construction and the incorporation of discriminative feature information are necessary to escalate the performance. The performance of LPP under such challenging conditions and consequently the extensions of these advanced LPP techniques and their integration into the proposed pipeline, are discussed in the following sections. The current work thus provides a strong baseline for vision-based cue detection using LPP, while also highlighting the need for further robustness improvements to handle more complex real-world environments.

3.3 Adaptive Spatial Kernel Modeling

As discussed in the preceding section, LPP can be viewed as a linear approximation of Laplacian eigenmaps and has demonstrated superior performance in various recognition tasks under demanding conditions, by preserving the local geometric structure of high-dimensional data. However, there are two fundamental limitations in the traditional formulation of LPP which restrict its performance in photometrically complex environments. *First*, LPP encodes only spatial proximity using Euclidean distance to construct the graph Laplacian. While sufficient in ideal scenarios, this approach is highly susceptible to distortions caused by spatial transformations, such as translation, rotation, and scaling, as well as the presence of noise and outliers [11]. Consequently, the resulting similarity weights may not accurately reflect the true relationships between samples. *Second*, the standard LPP framework does not incorporate any information from the feature (or range) domain. In illumination-compromised environments, intensity variations carry crucial semantic information that should be integrated into the feature extraction process to improve robustness.

To address these limitations, this Thesis proposes an improved locality preserving approach that integrates two key components. *Firstly*, an adaptive weight learning method

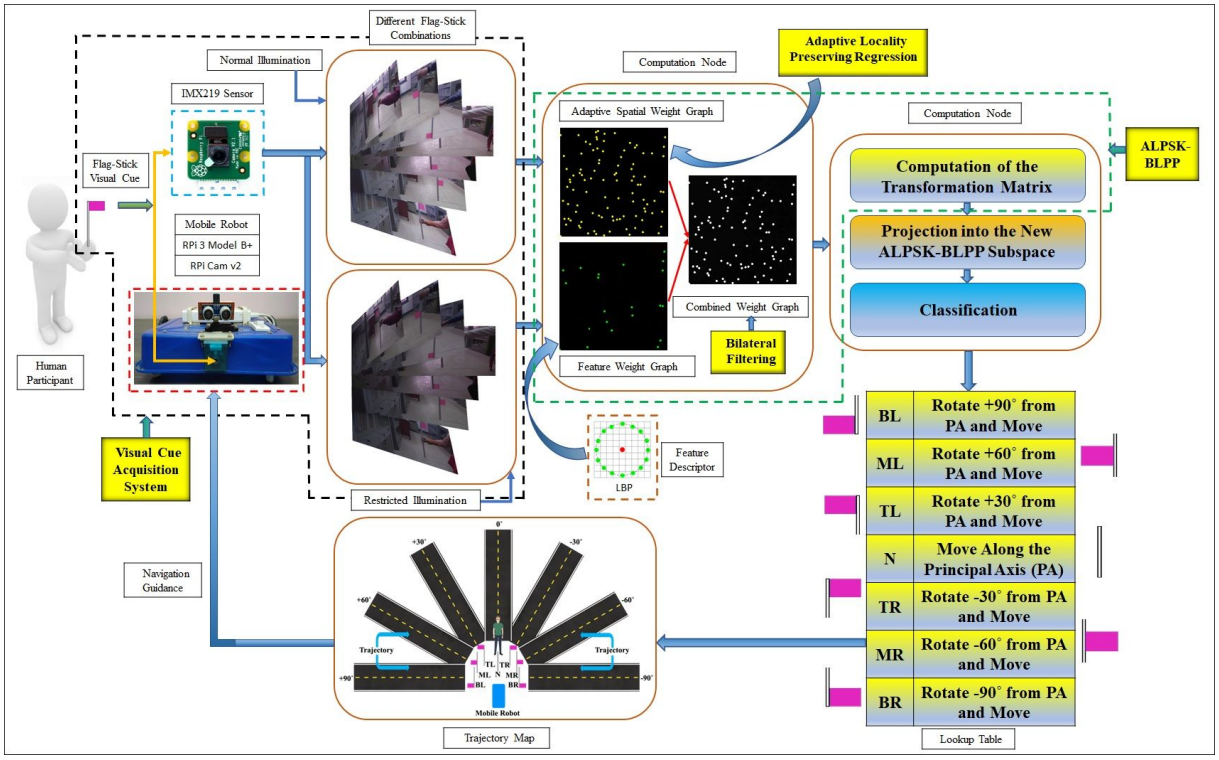


Figure 3.4: A block diagram illustrating the complete flag-stick visual cue-based robot navigation guidance system under human supervision, implemented using the ALPSK-BLPP framework.

based on retargeted least squares regression (ReLSR) [85] is introduced to refine the spatial similarity graph, making it less sensitive to outliers and more representative of the true neighborhood structure [70]. *Secondly*, the concept of bilateral filtering is incorporated to construct a combined similarity kernel that takes into account both spatial proximity and photometric similarity [11]. The resulting approach, termed *Adaptive Locality Preserving Spatial Kernel-Guided Bilateral LPP* (ALPSK-BLPP), leverages this dual-domain similarity to construct a more robust weight matrix. The adaptive spatial kernel mitigates spatial distortions, while the feature-domain kernel captures illumination-based intensity variations.

This enhanced modeling scheme is implemented and evaluated in a vision-sensor-based robot navigation system guided by human-provided flag-stick cues under changing photometric conditions. The formulation and effectiveness of the proposed ALPSK-BLPP approach are discussed in the subsequent sections. The block diagram of the overall ALPSK-BLPP scheme is presented in Fig. 3.4.

3.4 Mathematical Formulation of Adaptive Spatial Kernel

Let the input training dataset be denoted by $\mathbf{A} \in \mathbb{R}^{D\times N}$, where each column vector represents a visual cue sample of dimensionality D, and N is the total number of training samples. Suppose the j^{th} sample from the i^{th} class is denoted as $\mathbf{a}_j^i$, and the total number of distinct classes is c. It is generally expected that samples which are close neighbors in the original high-dimensional space, are likely to belong to the same class. Therefore, the goal is to preserve these neighborhood relationships in the low-dimensional subspace as well [86]. Now, there are various linear regression-based methods that can be used for encoding the nearest neighbor information in LPP formulations [87]. Traditional linear regression techniques offer mostly pre-constructed similarity graphs for performing the projection learning in LPP. However, such methods apply uniform weighting to all neighbor pairs, ignoring the subtle differences in similarity levels among them. This uniformity often leads to suboptimal projections, especially in the presence of overlapping or noisy data. To address this limitation, an adaptive strategy is incorporated for constructing the similarity graph here. In this approach, the graph weights are not fixed but are learned adaptively using a retargeted regularized optimization framework. Specifically, a graph regularization problem is formulated that jointly learns the projection matrix $\mathbf{\Omega}$ and the weight coefficients $s_{j,k}^i$ between neighboring sample pairs $\mathbf{a}_j^i$ and $\mathbf{a}_k^i$. The corresponding optimization problem is given by:

$$\min_{\mathbf{\Omega},\mathbf{S}} \lambda_1 \sum_{i=1}^{c} N_i \sum_{j,k=1,k\neq j}^{N_i} \left(s_{j,k}^i\right)^2 \left\|\mathbf{\Omega}^\top \mathbf{a}_j^i - \mathbf{\Omega}^\top \mathbf{a}_k^i\right\|_2^2 + \left\|\mathbf{Y} - \mathbf{A}^\top \mathbf{\Omega}\right\|_F^2 \tag{3.7}$$

$$\text{s.t.} \quad \sum_{k=1,k\neq j}^{N_i} s_{j,k}^i = 1, \quad s_{j,k}^i \geq 0$$

where s.t. stands for *subject to*, indicating the set of constraints applied to the optimization problem. Here, $\mathbf{Y} \in \mathbb{R}^{N\times c}$ refers to the label matrix corresponding to the training set $\mathbf{A}$, $\mathbf{\Omega} \in \mathbb{R}^{D\times c}$ is the transformation matrix, and $s_{j,k}^i$ is the required weight value for regularization of the distance between samples $\mathbf{a}_j^i$ and $\mathbf{a}_k^i$ in the discriminant subspace. The multiplier λ_1 works as the penalty parameter, which controls the strength of graph regularization, preserving local neighborhood relationships during adaptive spatial similarity learning. The introduction of the constraint in (3.7) ensures that all the classes are treated equally in preserving their

nearest neighbor structures. All the graph weights $s^i_{j,k}$ are calculated adaptively from the latent discriminative subspace instead of the complex original space. This helps the model in capturing intrinsic local nearest-neighbor relationships between the input samples. To improve the discriminability in the projection plane, retargeted learning approach can be incorporated [85]. With this, the graph regularization model can be re-written as:

$$\begin{aligned} &\min_{\mathbf{\Omega},\mathbf{S},\mathbf{Z}} \lambda_1 \sum_{i=1}^{c} N_i \sum_{j,k=1,k\neq j}^{N_i} \left(s^i_{j,k}\right)^2 \left\|\mathbf{\Omega}^\top \mathbf{a}^i_j - \mathbf{\Omega}^\top \mathbf{a}^i_k\right\|_2^2 + \left\|\mathbf{Z} - \mathbf{A}^\top \mathbf{\Omega}\right\|_F^2 \\ &\text{s.t.} \quad \sum_{k=1,k\neq j}^{N_i} s^i_{j,k} = 1, \quad s^i_{j,k} \geq 0, \quad \mathbf{Z}_{i,l_i} - \max_{j\neq l_i} \mathbf{Z}_{i,j} \geq 1 \end{aligned} \tag{3.8}$$

Here, $\mathbf{Z} \in \mathbb{R}^{N\times c}$ represents the target matrix with $l_i \in \{1,2,\ldots,c\}$ being the true label index for the i^{th} sample. In this way, margins between the true and false classes are maximized beyond the value of 1, which indeed increases the separability of the data in the target space [85]. Consequently, the discriminative projection could be learned by the adaptively computed target matrix $\mathbf{Z}$ in a more flexible manner. In real-world HRI applications like this, especially when the data is acquired under unstructured and dynamically challenging environments, it possesses a much higher dimension with several noise-like redundant features. To alleviate this redundancy, an $\ell_{2,1}$ row-sparsity norm term with a penalty parameter λ_2 can be introduced as [70]:

$$\begin{aligned} &\min_{\mathbf{\Omega},\mathbf{S},\mathbf{Z}} \lambda_1 \sum_{i=1}^{c} N_i \sum_{j,k=1,k\neq j}^{N_i} \left(s^i_{j,k}\right)^2 \left\|\mathbf{\Omega}^\top \mathbf{a}^i_j - \mathbf{\Omega}^\top \mathbf{a}^i_k\right\|_2^2 + \lambda_2 \|\mathbf{\Omega}\|_{2,1} + \left\|\mathbf{Z} - \mathbf{A}^\top \mathbf{\Omega}\right\|_F^2 \\ &\text{s.t.} \quad \sum_{k=1,k\neq j}^{N_i} s^i_{j,k} = 1, \quad s^i_{j,k} \geq 0, \quad \mathbf{Z}_{i,l_i} - \max_{j\neq l_i} \mathbf{Z}_{i,j} \geq 1 \end{aligned} \tag{3.9}$$

Here, λ_2 regulates the $\ell_{2,1}$-norm-based sparsity term, enforcing feature selection and suppressing redundant or noisy components. The use of $\ell_{2,1}$-norm is quite prominent in discriminative feature selection and feature extraction problems [88]. In the case of an ℓ_1-norm constraint, the solution of the corresponding optimization problem forces some elements of the vector to be zero. Whereas, in the case of the $\ell_{2,1}$-norm, all the elements corresponding to the i^{th} row of $\mathbf{\Omega}$ will be zero as the row-sparsity constraint will be $\|\mathbf{\Omega}_{i,:}\|_2 = \sqrt{\Omega^2_{i,1} + \Omega^2_{i,2} + \cdots + \Omega^2_{i,c}} = 0$. As the i^{th} row of the projection matrix becomes zero, the corresponding feature of the original input sample also gets eliminated and is not selected during the feature extraction process.

As the row-sparsity constrained graph regularization problem formulated above contains three unknowns $\mathbf{\Omega}$, $\mathbf{S}$, and $\mathbf{Z}$, it cannot be solved with analytical approaches [70]. However, this can be effectively solved in an iterative manner, obtaining the local optimal solution of the

variables. To solve Equation (3.9) iteratively, the corresponding cost function can be defined as:

$$\mathscr{C}(\mathbf{\Omega}) = \lambda_1 \sum_{i=1}^{c} N_i \sum_{j,k=1,k\neq j}^{N_i} \left(s_{j,k}^i\right)^2 \left\|\mathbf{\Omega}^\top \mathbf{a}_j^i - \mathbf{\Omega}^\top \mathbf{a}_k^i\right\|_2^2 + \lambda_2 \|\mathbf{\Omega}\|_{2,1} + \left\|\mathbf{Z} - \mathbf{A}^\top \mathbf{\Omega}\right\|_F^2 \tag{3.10}$$

This can be further simplified as:

$$\mathscr{C}(\mathbf{\Omega}) = \lambda_1 \operatorname{tr}(\mathbf{\Omega}^\top \mathbf{\Phi}\, \mathbf{\Omega}) + \lambda_2 \|\mathbf{\Omega}\|_{2,1} + \left\|\mathbf{Z} - \mathbf{A}^\top \mathbf{\Omega}\right\|_F^2 \tag{3.11}$$

where $\operatorname{tr}(\cdot)$ denotes the matrix trace operator, and the matrix $\mathbf{\Phi}$ is defined as:

$$\mathbf{\Phi} = \sum_{i=1}^{c} N_i \sum_{j,k=1,k\neq j}^{N_i} \left(s_{j,k}^i\right)^2 (\mathbf{a}_j^i - \mathbf{a}_k^i)(\mathbf{a}_j^i - \mathbf{a}_k^i)^\top \tag{3.12}$$

The optimal solution of $\mathbf{\Omega}$ is obtained by setting $d\mathscr{C}(\mathbf{\Omega})/d\mathbf{\Omega} = 0$, assuming $\mathbf{S}$ and $\mathbf{Z}$ are fixed. This leads to:

$$\mathbf{A}(\mathbf{A}^\top \mathbf{\Omega} - \mathbf{Z}) + \lambda_1 \mathbf{\Phi}\, \mathbf{\Omega} + \frac{\lambda_2}{2} \mathbf{\Delta} \mathbf{\Omega} = 0 \tag{3.13}$$

which results in:

$$\mathbf{\Omega} = \left(\mathbf{A}\mathbf{A}^\top + \lambda_1 \mathbf{\Phi} + \frac{\lambda_2}{2} \mathbf{\Delta}\right)^{-1} \mathbf{A}\mathbf{Z} \tag{3.14}$$

Here, $\mathbf{\Delta} \in \mathbb{R}^{D\times D}$ is a diagonal matrix with its elements defined as:

$$\Delta_{i,i} = \frac{1}{\sqrt{\sum_{j=1}^{c} \mathbf{\Omega}_{i,j}^2}} \tag{3.15}$$

With $\mathbf{\Omega}$ and $\mathbf{Z}$ fixed, the weight coefficients $\mathbf{S}$ are updated by solving the following constrained quadratic minimization:

$$\min_{\mathbf{S}} \sum_{i=1}^{c} N_i \sum_{j,k=1,k\neq j}^{N_i} \left(s_{j,k}^i\right)^2 \left\|\mathbf{\Omega}^\top \mathbf{a}_j^i - \mathbf{\Omega}^\top \mathbf{a}_k^i\right\|_2^2 \quad \text{s.t.} \quad \sum_{k=1,k\neq j}^{N_i} s_{j,k}^i = 1, \quad s_{j,k}^i \geq 0 \tag{3.16}$$

The optimal closed-form solution for $s_{j,k}^i$ is obtained as [70], [89]:

$$s_{j,k}^i = \frac{1}{\left\|\mathbf{\Omega}^\top \mathbf{a}_j^i - \mathbf{\Omega}^\top \mathbf{a}_k^i\right\|_2^2} \left(\sum_{p=1,p\neq j}^{N_i} \frac{1}{\left\|\mathbf{\Omega}^\top \mathbf{a}_j^i - \mathbf{\Omega}^\top \mathbf{a}_p^i\right\|_2^2} \right)^{-1} \tag{3.17}$$

Similarly, with $\mathbf{\Omega}$ and $\mathbf{S}$ fixed, the target matrix $\mathbf{Z}$ is computed by solving the following constrained optimization problem:

$$\min_{\{\mathbf{Z}_{i,l_i} - \max_{j\neq l_i} \mathbf{Z}_{i,j} \geq 1\}} \left\|\mathbf{Z} - \mathbf{A}^\top \mathbf{\Omega}\right\|_F^2 \tag{3.18}$$

This is a specific form of a constrained quadratic programming problem, which is solved using the ReLSR algorithm [85]. Fig. 3.5 illustrates the complete flowchart of the optimization algorithm used to solve the graph minimization problem in Equation (3.9) for computing the adaptive similarity matrix.

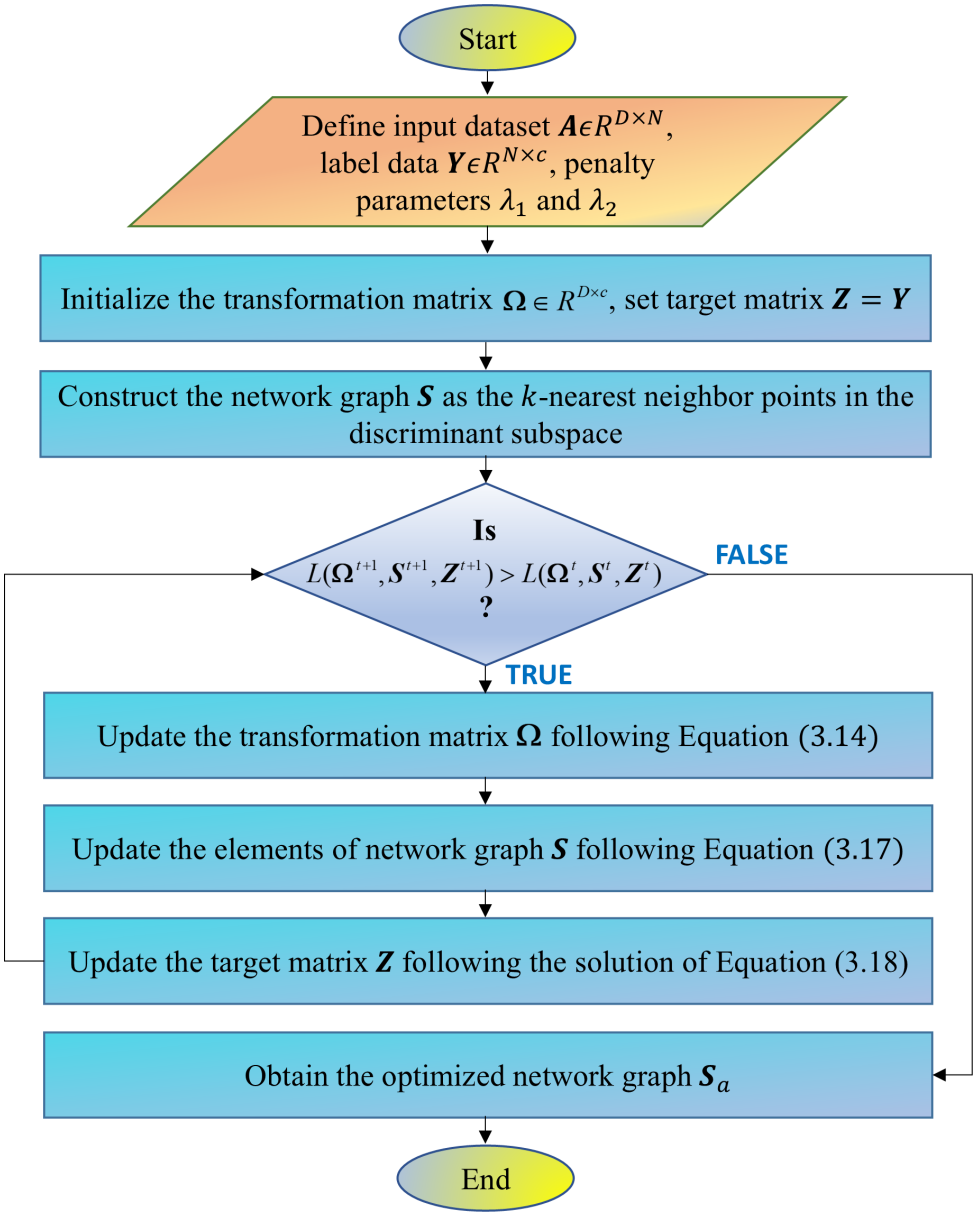


Figure 3.5: Flowchart illustrating the procedure for computing the adaptive spatial weight matrix $\mathbf{S}_a$ through the graph-based optimization framework.

3.5 Feature Learning and Multi-Kernel Fusion

In the case of conventional LPP [7], the similarity matrix in the Euclidean space is computed as:

$$s_{ij} = \begin{cases} \exp\left(-\dfrac{\|\mathbf{a}_i - \mathbf{a}_j\|_F^2}{\tau}\right), & \text{if } \mathbf{a}_j \in \mathcal{N}_k(\mathbf{a}_i) \\ 0, & \text{otherwise} \end{cases} \tag{3.19}$$

where $\mathcal{N}_k(\mathbf{a})$ is the set of k-nearest neighbors of $\mathbf{a}$. Here, the spatial weight s_{ij} is calculated entirely based on the Euclidean distance between samples $\mathbf{a}_i$ and $\mathbf{a}_j$. τ controls the size of the spatial neighborhood. Similar to this, the feature i.e., range weight between samples on the intensity axis in the feature domain can be computed as [11]:

$$r_{ij} = \begin{cases} \exp\left(-\dfrac{\|f(\mathbf{a}_i) - f(\mathbf{a}_j)\|_F^2}{\sigma}\right), & \text{if } \mathbf{a}_j \in \mathcal{N}_k(\mathbf{a}_i) \\ 1, & \text{otherwise} \end{cases} \tag{3.20}$$

σ is the parameter for regulating the weight between two pixels produced by the intensity difference. Any standard feature descriptor can be utilized to extract the feature $f(\mathbf{a}_i)$ from the image $\mathbf{a}_i$; e.g., local binary patterns (LBP) has been considered here. This photometric similarity or range weight r_{ij} between the pixel values in the feature domain can effectively capture the illumination variation in the original dataset. Instead of simple image intensities, an advanced set of features has been selected here to construct the r_{ij} kernel, which makes it more appropriate to be addressed as a feature kernel rather than a range kernel.

Now, in bilateral filtering, the output intensity $I_o(\mathbf{a}_i)$ corresponding to an image $\mathbf{a}_i$ and its intensity $I(\mathbf{a}_i)$ can be computed around its spatial neighborhood $\mathcal{N}_k(\mathbf{a}_i)$ as:

$$I_o(\mathbf{a}_i) = \frac{1}{\omega_p} \sum_{\mathbf{a}_j \in \mathcal{N}_k(\mathbf{a}_i)} I(\mathbf{a}_j) f_r(\|I(\mathbf{a}_i) - I(\mathbf{a}_j)\|) g_s(\|\mathbf{a}_i - \mathbf{a}_j\|) \tag{3.21}$$

Here, f_r is the feature kernel for smoothing intensity differences and g_s is the spatial (domain) kernel for smoothing the spatial differences in coordinates. ω_p is the normalized weight parameter, defined using the spatial closeness and intensity difference as:

$$\omega_p = \sum_{\mathbf{a}_j \in \mathcal{N}_k(\mathbf{a}_i)} f_r(\|I(\mathbf{a}_i) - I(\mathbf{a}_j)\|) g_s(\|\mathbf{a}_i - \mathbf{a}_j\|) \tag{3.22}$$

Employing a similar idea from bilateral filtering, where the combined weight $s_{ij} = g_s \times f_r$ is the product of spatial weight and feature weight, the combined weight prior to LPP can be computed. In the proposed method, the adaptive spatial weight is obtained from the algorithm mentioned in Fig. 3.5 instead of simple Euclidean weights as given in Equation (4.21). Let the adaptive spatial weight obtained be defined as $(s_a)_{ij}$, as discussed in Section 3.4. Consequently, the modified adaptive spatial-feature weight is computed as:

$$t_{ij} = (s_a)_{ij} \cdot r_{ij} \tag{3.23}$$

3.6 LPP-Based Low-Dimensional Projection Scheme

With the input training data $\mathbf{A} \in \mathbb{R}^{D\times N}$ comprising flag-stick visual cues, the objective of LPP is to find a suitable projection matrix $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_d] \in \mathbb{R}^{D\times d}$ which can map each input data $\mathbf{a}_i \in \mathbb{R}^D$ from the higher-dimensional space D to a lower-dimensional space d. The individual transformed images are obtained as $\mathbf{x}_i = \mathbf{W}^\top \mathbf{a}_i \in \mathbb{R}^d$ [7]. Owing to the concept of LPP, the optimization problem for ALPSK-BLPP can be formulated as:

$$\min_{\mathbf{W}} \sum_{i=1}^{N}\sum_{j=1}^{N} \left\|\mathbf{a}_i - \mathbf{a}_j\right\|^2 t_{ij} = \sum_{i=1}^{N} \sum_{\mathbf{a}_j \in \mathcal{N}_k(\mathbf{a}_i)} \left\|\mathbf{W}^\top \mathbf{a}_i - \mathbf{W}^\top \mathbf{a}_j\right\|^2 t_{ij} \tag{3.24}$$

Following some standard derivations, the minimization problem in (3.24) can be reformulated in the matrix form as [7]:

$$\min_{\mathbf{W}} \operatorname{tr}(\mathbf{W}^\top \mathbf{A}\mathbf{L}\mathbf{A}^\top \mathbf{W}) \quad \text{s.t.} \quad \mathbf{W}^\top \mathbf{A}\mathbf{D}\mathbf{A}^\top \mathbf{W} = 1 \tag{3.25}$$

where $\mathbf{L} = \mathbf{D} - \mathbf{T}$ is the Laplacian matrix and $\mathbf{T}$ is the weight matrix with its entries t_{ij}. $\mathbf{D}$ is a diagonal matrix with its elements defined as:

$$D_{ii} = \sum_{j=1}^{m} t_{ij} \tag{3.26}$$

Equation (3.25) can be equivalently represented in the form of a generalized eigenvalue decomposition problem such as [11]:

$$\mathbf{A}\mathbf{L}\mathbf{A}^\top \mathbf{w} = \lambda \mathbf{A}\mathbf{D}\mathbf{A}^\top \mathbf{w} \tag{3.27}$$

Given the non-singularity of $\mathbf{A}\mathbf{D}\mathbf{A}^\top$, the eigenvalue decomposition problem in (3.27) can be expressed as:

$$\left(\mathbf{A}\mathbf{D}\mathbf{A}^\top\right)^{-1} \mathbf{A}\mathbf{L}\mathbf{A}^\top \mathbf{w} = \lambda \mathbf{w} \tag{3.28}$$

The solution to the above eigenvector problem derives the required ALPSK-BLPP transformation matrix $\mathbf{W}_{\text{ALPSK-BLPP}}$, by which the data can be projected from the original higher-dimensional domain to the ALPSK-BLPP reduced domain. The entire algorithm of ALPSK-BLPP is shown in the form of a flowchart in Fig. 3.6. To deal with the singularity in $\mathbf{A}\mathbf{D}\mathbf{A}^\top$, PCA is introduced in the form of a preprocessing step before conducting ALPSK-BLPP, as shown in the algorithm.

The experimental evaluation using the flag-stick visual cue dataset has been organized into five distinct phases. In the first phase, only the images captured under normal lighting conditions were used for both training and testing purposes. In the subsequent four phases, while the

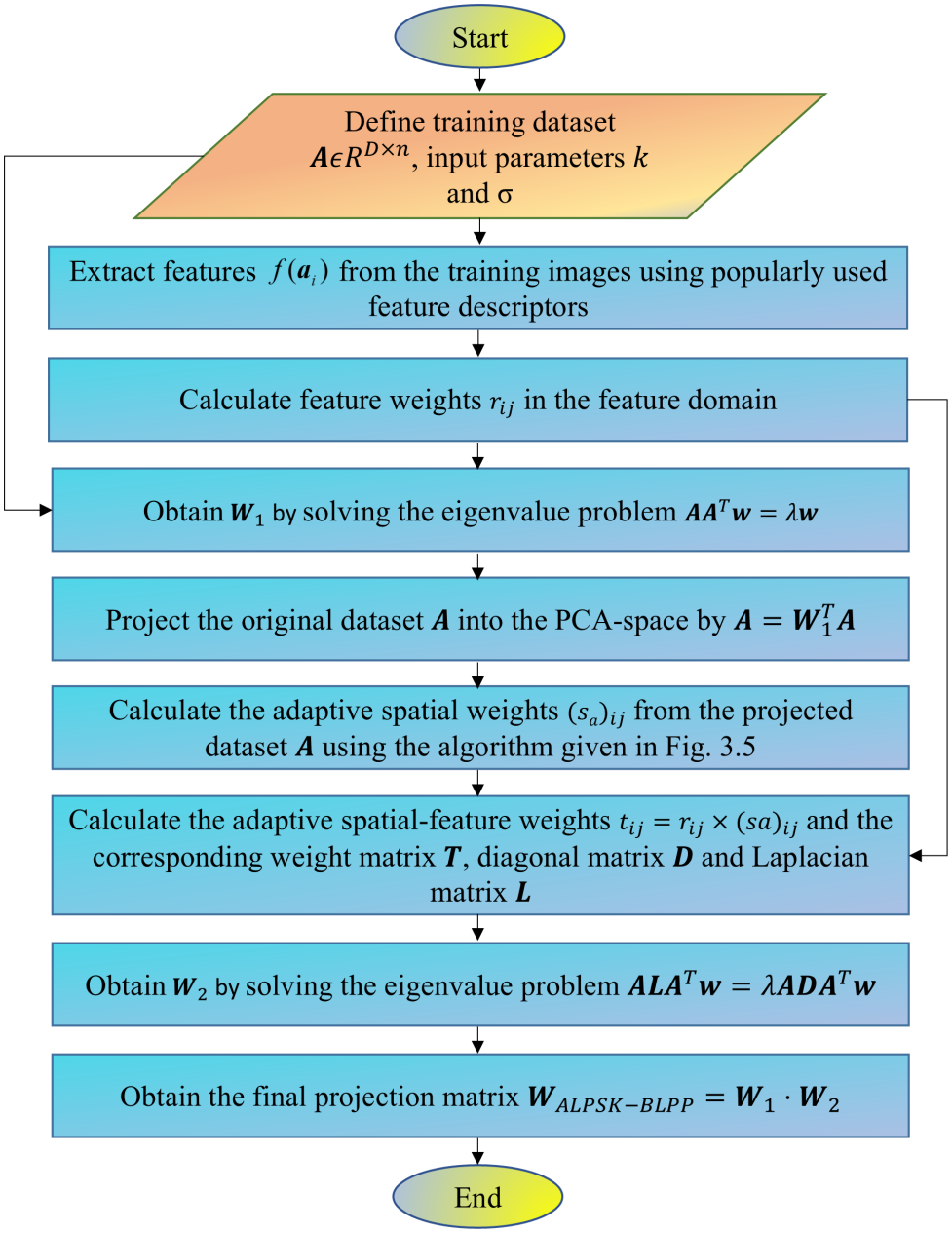


Figure 3.6: Flowchart depicting the proposed ALPSK-BLPP algorithm for deriving the projection matrix $\mathbf{W}_{\text{ALPSK-BLPP}}$.

training set has remained unchanged, the testing data has been replaced with samples collected under reduced illumination settings; specifically dark level-1, dark level-2, dark level-3, and dark level-4 conditions. A visual depiction of the dataset distribution under each lighting scenario was provided earlier in Fig. 2.7 in Section 2.4 of Chapter 2. The dataset acquired under normal photometric settings comprises 1050 images, with 150 samples from each of the seven visual cue classes. Throughout all experiments, a fixed training-to-testing ratio of 7 : 3 has been maintained. The number of nearest neighbors k for calculating the feature weights

has been set to 6. The dimension d of the reduced subspace has been selected in such a way that the energy content of the input vector has been maximally preserved within the chosen principal components. Local binary patterns (LBP) has been employed as the default feature descriptor to extract feature values $f(\mathbf{a}_i)$ and compute the corresponding range weights r_{ij} in the feature domain. Classification of test samples has been performed using a support vector machine (SVM) classifier.

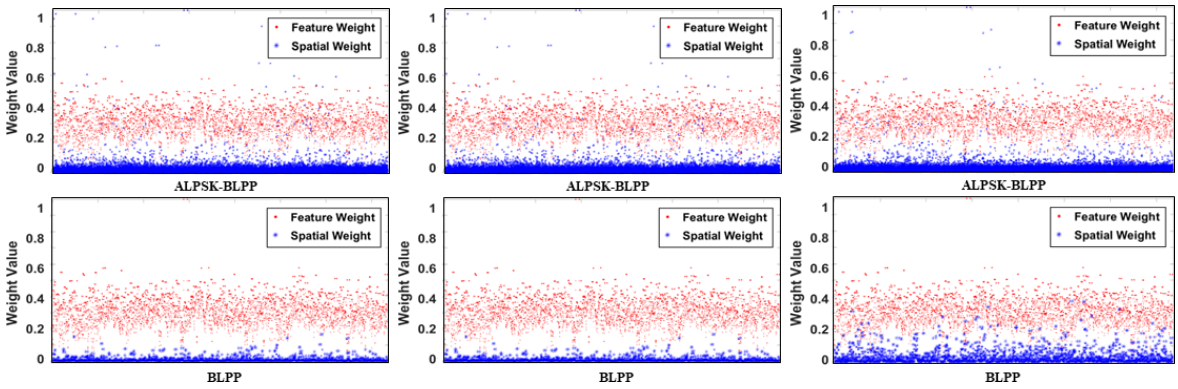
3.7 Experimental Results and Discussion

A comparative analysis of different popular dimensionality reduction methods, including their augmented forms using the proposed ALPSK-BLPP approach, under varying photometric conditions, is presented in Table 3.4. Initially, both the training and testing samples have been taken from data captured under regular lighting conditions. Under this setting, the majority of the state-of-the-art algorithms have produced satisfactory classification accuracy. Comparable results have been obtained for PCA, LDA, and LPP, while the remaining augmented methods have achieved accuracy exceeding 99.90%. This strongly validates the reliability and effectiveness of the proposed robot navigation guidance framework using flag-stick cues, which can be implemented cost-effectively with a straightforward user interface and integrated with state-of-the-art algorithms. To assess the robustness of the system under more challenging real-world conditions, the testing dataset has progressively been altered to include samples with degraded illumination. Four additional phases of experimentation were performed using images captured under these darkened conditions, which naturally introduce noise, outliers, and non-linear distribution of data points. This setup enables a more comprehensive evaluation of the robustness of locality-preserving dimensionality reduction strategies, particularly in the context of adaptive weighting mechanisms.

Consistent with this, the performance of each algorithm has been assessed under these low-light settings. Traditional dimensionality reduction methods such as PCA and LDA exhibited noticeable performance deterioration. However, significant improvements in recognition rates were evident when locality-preserving methods were employed. For instance, under dark level-2 lighting, standard LPP [7] achieved an accuracy of 84.127%, reflecting the influence of outliers and the non-linear nature of the sample distribution in vision-based

Table 3.4: CLASSIFICATION ACCURACY (% ± STD. DEV.) OF VARIOUS DIMENSIONALITY REDUCTION METHODS ACROSS DIFFERENT PHOTOMETRIC CONDITIONS

Method	Normal Illum.	Dark-1	Dark-2	Dark-3	Dark-4
PCA	99.048 ± 0.184	70.253 ± 0.352	46.349 ± 0.478	39.885 ± 0.512	32.012 ± 0.586
LDA	99.365 ± 0.165	81.296 ± 0.296	73.333 ± 0.368	65.320 ± 0.483	49.678 ± 0.504
LPP	99.683 ± 0.105	92.362 ± 0.255	84.127 ± 0.349	70.773 ± 0.392	55.710 ± 0.454
ALPR	99.939 ± 0.020	94.667 ± 0.211	88.571 ± 0.293	73.654 ± 0.322	61.175 ± 0.413
BLPP	99.965 ± 0.011	96.329 ± 0.159	94.603 ± 0.196	75.221 ± 0.250	62.394 ± 0.386
ALPR + LPP	99.965 ± 0.009	97.112 ± 0.107	95.556 ± 0.188	76.058 ± 0.198	65.943 ± 0.365
ALPSK-BLPP	**99.971 ± 0.009**	**98.875 ± 0.091**	**97.778 ± 0.165**	**81.269 ± 0.175**	**71.778 ± 0.310**



(a) (b) (c)

Figure 3.7: Comparison of feature weights and adaptive spatial weights generated by ALPSK-BLPP and BLPP across varying illumination conditions: (a) Normal, (b) Dark level-2, and (c) Dark level-4.

data collected in dark environments. By comparison, methods utilizing adaptive spatial kernel regression [70] and spatial-feature kernel-based LPP [11] have attained improved recognition rates of 88.571% and 94.603%, respectively. When adaptive locality preserving regression (ALPR) [70] was applied prior to LPP, the recognition accuracy further increased to 95.556%. Ultimately, the proposed ALPSK-BLPP framework outperformed all other approaches in both accuracy and resilience, reaching a maximum recognition accuracy of 97.778%.

Table 3.5: VARIATION IN FEATURE AND SPATIAL WEIGHT VALUES FOR ALPSK-BLPP AND BLPP UNDER DIFFERENT ILLUMINATION CONDITIONS

Method	**Illum. Cond.**	**Feature Weights**			**Spatial Weights**		
		Range $(w_f^{\max} - w_f^{\min})$	**Mean** $\bar{w}_f$	**SD** σ_{w_f}	**Range** $(w_s^{\max} - w_s^{\min})$	**Mean** $\bar{w}_s$	**SD** σ_{w_s}
ALPSK-BLPP	Normal	0.9447	0.2977	0.0809	1	0.0014	0.0076
	Dark-1	0.9433	0.2961	0.0810	1	0.0012	0.0078
	Dark-2	0.9488	0.2979	0.0812	1	0.0015	0.0079
	Dark-3	0.9472	0.2965	0.0815	1	0.0013	0.0079
	Dark-4	0.9455	0.2986	0.0817	1	0.0014	0.0078
BLPP	Normal	0.9368	0.2954	0.0804	0.178	0.0126	0.0579
	Dark-1	0.9411	0.2972	0.0807	0.213	0.0155	0.0587
	Dark-2	0.9395	0.2962	0.0811	0.267	0.0171	0.0688
	Dark-3	0.9464	0.2968	0.0812	0.324	0.0271	0.0880
	Dark-4	0.9483	0.2978	0.0816	0.389	0.0382	0.1053

In the ALPSK-BLPP method, the joint influence of both spatial and feature weight components contributes meaningfully to the overall performance. In contrast, in the case of BLPP, the spatial weights are fixed as per Equation (4.21). With an increasing number of outliers, the spatial distance between pixels tends to increase, thereby reducing the spatial weight values. ALPSK-BLPP, however, computes adaptive spatial weights that remain effective with substantial values, even under photometric distortions. Fig. 3.7 compares the spatial and feature weights produced by BLPP and ALPSK-BLPP under three lighting conditions. The variations of feature and spatial weights for both these methods under different illumination conditions are summarized in Table 3.5. Here, *Range* indicates the difference between the maximum and minimum weight values $(w^{\max} - w^{\min})$, *Mean* denotes the average value $\bar{w}$, and *SD* represents the corresponding standard deviation σ_w. The subscripts f and s refer to feature and spatial weights, respectively. While feature weights remain relatively stable, spatial weights show considerable variation in the case of BLPP. In contrast, ALPSK-BLPP exhibits less fluctuation in spatial weights under changing illumination, as reflected by the respective ranges, means, and standard deviations (SDs). This substantiates ALPSK-BLPP to be a more reliable method for computing spatial weights in the LPP framework, especially

$\lambda_1 \uparrow$ / $\lambda_2 \rightarrow$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
1.0	92.698	92.381	93.016	93.968	93.651	93.968	94.603	94.603	94.603	93.651
0.9	92.698	94.286	93.968	93.016	93.651	93.016	94.921	94.603	94.286	93.968
0.8	94.286	94.286	93.333	93.333	93.016	94.286	93.651	93.968	94.921	94.921
0.7	95.873	94.921	95.238	95.238	94.921	95.556	95.556	95.238	95.556	95.556
0.6	**97.778**	**97.778**	96.191	96.191	96.191	96.191	**97.778**	96.508	97.460	97.460
0.5	97.143	95.873	**97.778**	96.508	95.556	**97.778**	95.873	95.873	97.460	95.556
0.4	96.508	96.825	96.508	96.508	97.460	97.460	97.143	96.825	**97.778**	**97.778**
0.3	95.238	93.968	93.968	94.603	95.238	95.873	96.191	96.191	95.873	96.191
0.2	92.064	92.064	92.381	93.016	94.286	94.286	94.921	94.603	93.968	92.381
0.1	87.302	86.984	88.254	87.619	87.937	87.619	86.667	86.032	87.619	88.254

Figure 3.8: Classification accuracy (in %) of ALPSK-BLPP across varying values of regularization parameters λ_1 and λ_2 within the range [0.1, 1].

for photometrically degraded input images.

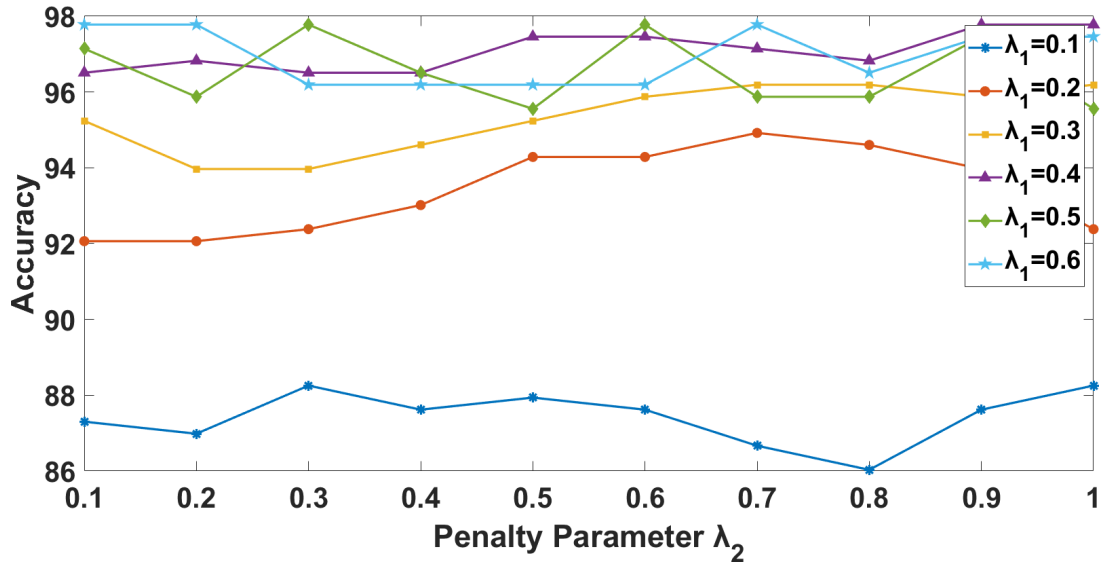


Figure 3.9: Surface plot illustrating the variation in classification accuracy (in %) with respect to changes in λ_1 and λ_2.

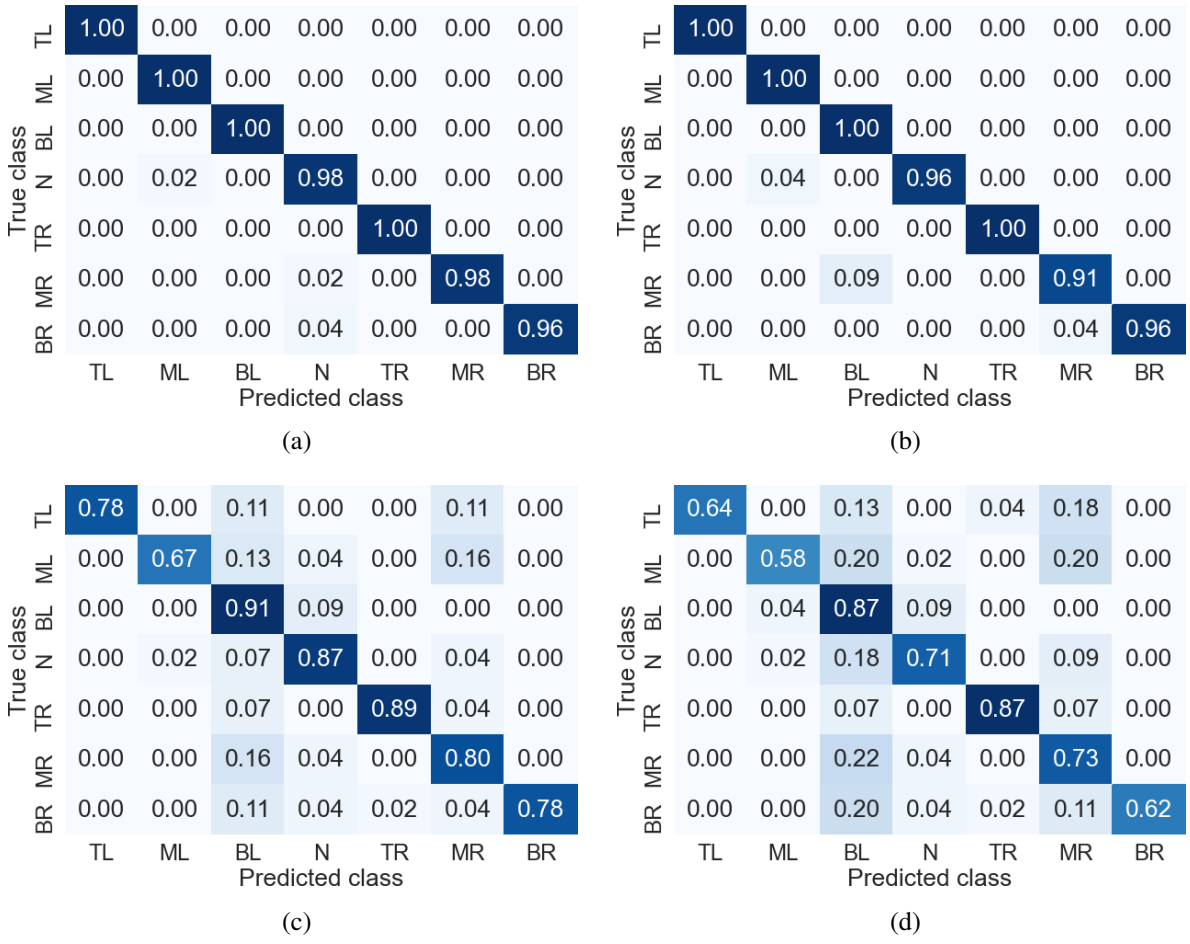


Figure 3.10: Confusion matrices illustrating the performance of the proposed ALPSK-BLPP algorithm under different illumination conditions: (a) Dark-1, (b) Dark-2, (c) Dark-3, and (d) Dark-4.

Table 3.6: CLASSIFICATION METRICS OF ALPSK-BLPP UNDER DIFFERENT ILLUMINATION CONDITIONS

Parameter	**Normal**	**Dark-1**	**Dark-2**	**Dark-3**	**Dark-4**
Sensitivity	99.68	98.73	97.46	81.27	71.75
Specificity	99.95	99.79	99.58	96.88	95.29
Precision	99.69	98.78	97.56	85.60	80.03
NPV	99.95	99.79	99.58	96.92	95.37
FPR	0.05	0.21	0.42	3.12	4.71
F1-Score	99.68	98.74	97.46	82.00	73.08

While implementing ALPSK-BLPP, the performance of the model is also influenced by the proper choice of its regularization parameters. In the objective function described by Equation (3.9), two penalty terms, λ_1 and λ_2, play a crucial role. These parameters are responsible for preserving nearest neighbor structures and for performing feature selection, respectively, within the graph-based regularization model [70]. To evaluate the sensitivity of the algorithm with respect to these parameters, each of them has been initially varied across a wide range of values, from 10^{-3} to 10^{3}. Based on this, a finer range of [0.1, 1] was identified for both λ_1 and λ_2. A grid search was then performed in this refined interval to determine optimal parameter values [70]. The effect of varying these parameters on classification accuracy at dark level-2 is depicted in Fig. 3.8. A graphical surface view of how ALPSK-BLPP's accuracy changes with λ_1 and λ_2 is shown in Fig. 3.9.

The classification performance of ALPSK-BLPP under each of the four darkened lighting conditions is shown using confusion matrices in Fig. 3.10. The detailed classification metrics—*sensitivity*, *specificity*, *precision*, *negative predictive value* (NPV), *false positive rate* (FPR), and *F1-score*, for all photometric levels are provided in Table 3.6. These experimental findings reinforce the theoretical motivation for ALPSK-BLPP as a resilient algorithm suitable for deployment in real-world HRI environments with photometric challenges. In settings where image data suffers from illumination degradation, the results highlight the importance of: (i) incorporating adaptive measures in spatial weight computation within LPP to handle non-linearity and noise, and (ii) integrating intensity similarity information from the feature domain.

3.8 Summary

This chapter presented a comprehensive methodology for robust visual cue recognition in human-supervised robot navigation systems. Beginning with a baseline approach based on locality preserving projections (LPP), the work addressed the limitations of conventional low-dimensional projection methods under variable and challenging photometric conditions. The LPP-guided scheme demonstrated improved neighborhood preservation compared to global information-based techniques, validating its effectiveness in semi-structured environments with moderate variation. Notably, the normal LPP method performed well under standard lighting conditions, but showed a marked decline in recognition accuracy across increasingly darkened environments, starting from dark-1 through dark-4 scenarios.

To extend this robustness to more adverse real-world scenarios, an adaptive spatial kernel modeling framework was developed. The proposed ALPSK method introduced an adaptively learned similarity structure that replaced the fixed Euclidean-based weighting with an optimized, data-driven graph. By incorporating class-wise margin constraints and row-sparsity regularization, ALPSK enabled both enhanced discriminability and feature selection within the projected space. The inclusion of range-domain feature weights, guided by bilateral filtering principles, further refined the locality encoding to accommodate photometric inconsistencies in the image data.

The chapter culminated in the formulation of ALPSK-BLPP, a unified framework combining adaptive spatial and feature-based similarities into a joint projection model. This scheme demonstrated a clear improvement over standard LPP, especially under dark-1 and dark-2 conditions, where it maintained high classification accuracy. However, its performance under dark-3 and dark-4 conditions, though improved, was still suboptimal. These results point to the limitations of current methods under extreme photometric degradation and highlight the need for more resilient algorithmic strategies. Furthermore, evaluating the robustness of these methods under externally induced noise conditions remains an open avenue. These aspects present meaningful opportunities for further improvement and are explored in subsequent chapters. Taken together, the techniques presented in this chapter form a scalable and effective solution for visual perception in collaborative robotic systems. They also serve as a foundational component for subsequent work on multimodal and context-aware interaction strategies, which

are explored in the following chapters.

Chapter 4

Local Pattern Encoding-based Kernel Adoption Strategies in LPP

4.1 Introduction

Visual recognition systems operating in real-world human-robot interaction (HRI) scenarios often encounter input data, degenerated due to illumination fluctuations, sensor-induced noise, and structural ambiguities [65]. Conventional global projection techniques, while working effectively under ideal imaging conditions, tend to falter when confronted with such distortions. To address the issue, this chapter presents a comprehensive investigation into local pattern encoding-based kernel adoption strategies designed to augment the performance of locality preserving projections (LPP) [7] under adverse visual conditions.

The central idea is to employ robust local feature descriptors that can effectively encode spatial microstructures, edge orientations, and statistical properties of the local image neighborhoods. These descriptors are then utilized to build photometric-aware kernel matrices that enhance the LPP framework. This chapter explores various local pattern-based descriptors, such as local binary patterns (LBP) [30], local derivative patterns (LDP) [90], local ternary patterns (LTP) [31], and local tetra patterns (LTrP) [91], to investigate their effectiveness under illumination degradation and in noise corrupted scenarios. In particular, LTrP, a four-directional encoding scheme, is found to work well in environments involving impulsive and multiplicative noise. Furthermore, a detailed discussion is carried out on histogram refinement schemes that further improve descriptor granularity, using local structural and statistical properties such as skewness (local skew pattern or LSP) and structural variation (binarized eigenvalue map or BEM) [92].

The chapter then introduces histogram refined local binary pattern (HRLBP) and

histogram refined local ternary pattern (HRLTP)-based encoding strategies. These two frameworks are built upon classical LBP and LTP descriptors and incorporate histogram refinement to enhance their robustness under varying photometric and noise conditions. In particular, the HRLBP approach demonstrates strong performance under moderate lighting shifts, while HRLTP is shown to be resilient in hybrid degradation cases, involving both noise and dim illumination. All local pattern descriptors are integrated into a bilateral locality preserving projection (BLPP) framework to ensure spatial-feature preservation during dimensionality reduction. The resulting schemes are rigorously evaluated across a range of degraded datasets to validate their resilience and class separability.

4.2 Feature Encoding with Local Tetra Patterns (LTrP)

Robust and meaningful local feature extraction is essential for accurate pattern recognition in real-world human-robot interaction systems, mainly when input data is affected by visual ambiguities and sensor inconsistencies. With datasets that are degraded and contaminated with various noise artifacts, extraction of meaningful and robust features from high-dimensional input data becomes a challenging task. To address this challenge, this work explores the use of *local tetra patterns* (LTrP) [91]; a noise-resilient, four-directional encoding scheme, specially suited for visual cue recognition under degraded sensor conditions.

4.2.1 Local Feature Descriptors

Under such adverse conditions as mentioned above, traditional subspace projection-based dimensionality reduction techniques often fail to deliver satisfactory performance. In contrast, locality-aware dimensionality reduction methods like locality preserving projections (LPP) provide a more suitable alternative in these scenarios. LPP captures the intrinsic geometrical relationships present in the original data and projects them into a lower-dimensional subspace while retaining local structure. This technique effectively embeds the neighborhood pixel distribution from the input data into the reduced space.

However, standard LPP only incorporates spatial proximity between neighboring data points using distance metrics, which are then employed to construct the projection kernel. In

this way, the local variation in pixel intensity due to sensor degradation and noises cannot be captured. To overcome this limitation, bilateral filtering-inspired LPP (BLPP) leverages the idea of combining both spatial and feature-based kernels derived from the input data [11]. Various standard descriptors, such as local binary patterns (LBP) [30], histograms of oriented gradients (HOG) [47], and local phase quantization (LPQ) [93] can be used to generate the feature kernel for this purpose. To explore beyond simple intensity-based encodings, Zhang *et al.* proposed local derivative patterns (LDP), which extend LBP using first-order directional derivatives [90]. Further building upon it, Tan *et al.* introduced local ternary patterns (LTP) [31], which improve discrimination and robustness to noise, making them a generalized, less sensitive form of the LBP descriptor [30].

Nevertheless, LBP and its variants like LDP and LTP perform feature coding in only two directions, limiting their ability to encode complex edge distributions. Extending the directional encoding beyond two axes can potentially improve overall system performance. Murala *et al.* proposed local tetra patterns (LTrP), a more discriminative four-directional local descriptor, initially developed for content-based image retrieval tasks [91]. In the current study, LTrP has been adopted to extract salient and noise-resilient local features from vision sensor data affected by impulse and multiplicative noise. The extracted LTrP-based feature representations are then fused with the spatial information to construct the final projection matrix using the proposed local tetra pattern-based bilateral LPP scheme (LTrP-BLPP).

4.2.2 Mathematical Foundation of LTrP

Local tetra pattern (LTrP) captures the spatial texture of local neighborhoods by determining directionality with respect to the central pixel's intensity. For a given image $\mathbf{A}$, the first-order derivatives in the horizontal (0°) and vertical (90°) directions for a neighborhood pixel g_p are represented as $\mathbf{A}^1_{0^\circ}(g_p)$ and $\mathbf{A}^1_{90^\circ}(g_p)$, where g_p is the gray value of the neighbor pixels around the central pixel g_c. Assuming g_h and g_v denote the neighboring gray values in the horizontal and vertical directions of g_c, the first-order derivatives are defined as [91]:

$$\mathbf{A}^1_{0^\circ}(g_c) = \mathbf{A}(g_h) - \mathbf{A}(g_c) \tag{4.1}$$

$$\mathbf{A}^1_{90^\circ}(g_c) = \mathbf{A}(g_v) - \mathbf{A}(g_c) \tag{4.2}$$

The directionality of the central pixel g_c is then encoded as:

$$\mathbf{A}_d^1(g_c) = \begin{cases} 1, & \text{if } \mathbf{A}_{0^\circ}^1(g_c) \geq 0 \text{ and } \mathbf{A}_{90^\circ}^1(g_c) \geq 0 \\ 2, & \text{if } \mathbf{A}_{0^\circ}^1(g_c) < 0 \text{ and } \mathbf{A}_{90^\circ}^1(g_c) \geq 0 \\ 3, & \text{if } \mathbf{A}_{0^\circ}^1(g_c) < 0 \text{ and } \mathbf{A}_{90^\circ}^1(g_c) < 0 \\ 4, & \text{if } \mathbf{A}_{0^\circ}^1(g_c) \geq 0 \text{ and } \mathbf{A}_{90^\circ}^1(g_c) < 0 \end{cases} \tag{4.3}$$

These values 1 to 4 define the four possible orientation directions of encoding that the central pixel can possess. The second-order tetra pattern for the central pixel, having N neighboring pixels, is then expressed as [91]:

$$\text{LTrP}^2(g_c) = \left\{\phi(\mathbf{A}_d^1(g_c), \mathbf{A}_d^1(g_1)), \phi(\mathbf{A}_d^1(g_c), \mathbf{A}_d^1(g_2)), \ldots, \phi(\mathbf{A}_d^1(g_c), \mathbf{A}_d^1(g_N))\right\} \tag{4.4}$$

where the function $\phi(\cdot)$ is defined as:

$$\phi(\mathbf{A}_d^1(g_c), \mathbf{A}_d^1(g_N)) = \begin{cases} 0, & \text{if } \mathbf{A}_d^1(g_c) = \mathbf{A}_d^1(g_N) \\ \mathbf{A}_d^1(g_N), & \text{otherwise} \end{cases} \tag{4.5}$$

To compute higher-order tetra patterns, it is necessary to use previously calculated derivative terms along both horizontal and vertical directions. Thus, the n-th order LTrP for g_c is expressed as [91]:

$$\text{LTrP}^n(g_c) = \left\{\phi(\mathbf{A}_d^{n-1}(g_c), \mathbf{A}_d^{n-1}(g_1)), \phi(\mathbf{A}_d^{n-1}(g_c), \mathbf{A}_d^{n-1}(g_2)), \ldots, \phi(\mathbf{A}_d^{n-1}(g_c), \mathbf{A}_d^{n-1}(g_N))\right\} \tag{4.6}$$

Higher-order LTrPs tend to become more sensitive to noise, which makes the second-order LTrP more suitable for extracting robust features from noisy images [91]. In contrast to local binary pattern (LBP), local derivative pattern (LDP), or local ternary pattern (LTP), which encode pixels into two (0 or 1) or three (-1, 0, 1) states, LTrP can represent them in four directional classes, enabling richer local feature encoding.

4.2.3 LTrP-Encoded Low-Dimensional Projection Scheme LTrP-BLPP

Once the LTrP-coded local features are extracted from each input image, the next step involves embedding this high-dimensional data into a lower-dimensional subspace using the framework of bilateral filtering-aided locality preserving projections (BLPP) [11], as discussed earlier in Sections 3.5-3.6 in Chapter 3. This hybrid scheme, known as LTrP-BLPP, seeks to retain both the spatial coherence and local photometric similarity of the data during the projection process.

Let the input set of images be represented as $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_N] \in \mathbb{R}^{D \times N}$, where each $\mathbf{A}_i \in \mathbb{R}^D$ is a vectorized representation of the grayscale image, and N is the number of training samples. The goal is to compute a transformation matrix $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_d] \in \mathbb{R}^{D \times d}$ that projects each high-dimensional vector $\mathbf{A}_i$ into a lower-dimensional representation $\mathbf{x}_i = \mathbf{W}^\top \mathbf{A}_i \in \mathbb{R}^d$, with $d \ll D$.

In traditional LPP, the similarity between two samples $\mathbf{A}_i$ and $\mathbf{A}_j$ is evaluated based on Euclidean distance in the original space. In contrast, the LTrP-BLPP formulation enriches this by incorporating photometric feature similarity into the weight computation. The spatial similarity weight is defined as [7]:

$$s_{ij} = \begin{cases} \exp\left(-\frac{\|\mathbf{A}_i - \mathbf{A}_j\|_F^2}{\tau}\right), & \text{if } \mathbf{A}_j \in \mathcal{N}_k(\mathbf{A}_i) \\ 0, & \text{otherwise} \end{cases} \tag{4.7}$$

In parallel, the feature similarity is derived from the LTrP-coded descriptors, which effectively encode textural information with robustness to noise perturbations and outliers. Denoting $f(\mathbf{A}_i)$ as the LTrP feature representation of $\mathbf{A}_i$, the corresponding feature-based weight is given by:

$$r_{ij} = \begin{cases} \exp\left(-\frac{\|f(\mathbf{A}_i) - f(\mathbf{A}_j)\|_2^2}{\sigma}\right), & \text{if } \mathbf{A}_j \in \mathcal{N}_k(\mathbf{A}_i) \\ 0, & \text{otherwise} \end{cases} \tag{4.8}$$

Here, τ and σ are tuning parameters that control the influence of spatial and feature domain proximity, respectively. These two components are combined multiplicatively to yield the final joint similarity weight as [11]:

$$t_{ij} = s_{ij} \cdot r_{ij} \tag{4.9}$$

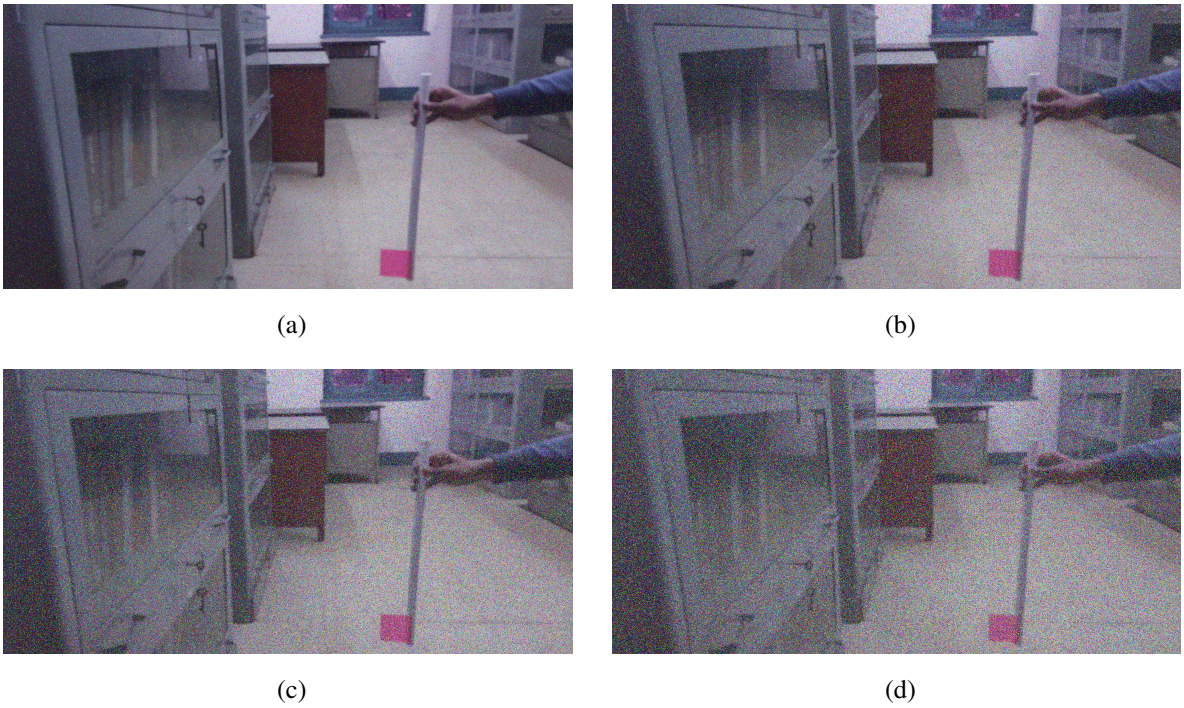
This combined weight forms the bilateral similarity matrix $\mathbf{T}$, whose entries t_{ij} reflect both spatial and feature domain affinities between neighboring samples. The diagonal degree matrix $\mathbf{D}$ is constructed with entries $D_{ii} = \sum_j t_{ij}$, and the corresponding Laplacian matrix is obtained as $\mathbf{L} = \mathbf{D} - \mathbf{T}$. Correspondingly, the optimization objective of LTrP-BLPP becomes:

$$\min_{\mathbf{W}} \operatorname{tr}(\mathbf{W}^\top \mathbf{A}\mathbf{L}\mathbf{A}^\top \mathbf{W}) \quad \text{subject to} \quad \mathbf{W}^\top \mathbf{A}\mathbf{D}\mathbf{A}^\top \mathbf{W} = \mathbf{I} \tag{4.10}$$

Equation (4.10) is equivalent to the following generalized eigenvalue problem:

$$\mathbf{A}\mathbf{L}\mathbf{A}^\top \mathbf{w} = \lambda \mathbf{A}\mathbf{D}\mathbf{A}^\top \mathbf{w} \tag{4.11}$$

Equation (4.11) yields the projection vectors $\{\mathbf{w}_1, \ldots, \mathbf{w}_d\}$ corresponding to the d smallest non-zero eigenvalues. These eigenvectors constitute the columns of the transformation matrix



(a) (b)

(c) (d)

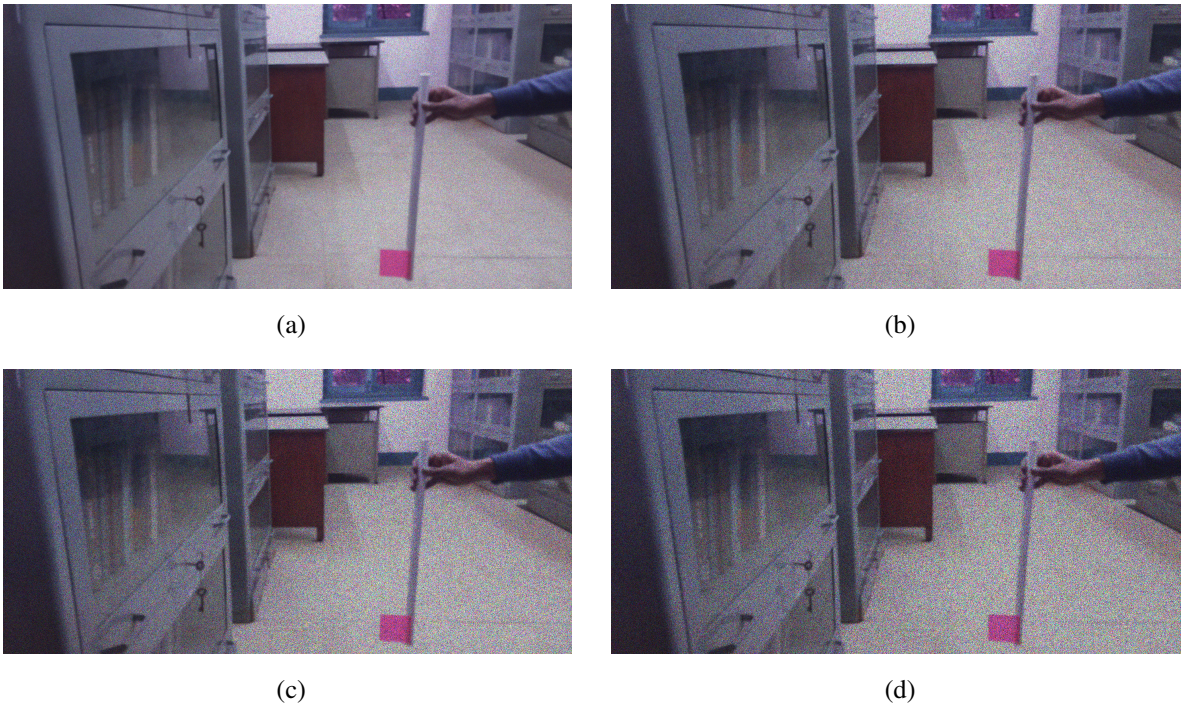
Figure 4.1: Sample flag-stick visual cue images corrupted with different levels of salt-and-pepper noise: (a) $\delta = 0.05$, (b) $\delta = 0.15$, (c) $\delta = 0.25$, and (d) $\delta = 0.35$.

$\mathbf{W}$. The final low-dimensional representation $\mathbf{x}_i = \mathbf{W}^\top \mathbf{A}_i$ encapsulates both the spatial arrangement and LTrP-coded local textures, making the proposed scheme particularly effective in handling noise and illumination-induced outliers. This projection, when followed by a suitable classifier (such as multiclass SVM), enables robust recognition of symbolic visual cues embedded in vision-sensor-acquired data, as investigated through empirical evaluations in the subsequent sections.

4.2.4 Noise-Corrupted Visual Cues for Evaluation

Noise generated in vision sensors represents a distinct type of electronic distortion that causes irregular variations in image brightness and color fidelity. These distortions may stem from inconsistencies within the image sensor or faults in the camera's internal processing circuitry. Such anomalies are often the result of sensor aging or prolonged wear and tear in hardware components.

To emulate realistic sensor degradation, this study includes two prominent types of noise disturbances: impulse noise and multiplicative noise [94]. Impulse noise, commonly referred



(a) (b)

(c) (d)

Figure 4.2: Sample flag-stick visual cue images corrupted with different levels of speckle noise: (a) $S^2 = 0.05$, (b) $S^2 = 0.15$, (c) $S^2 = 0.25$, and (d) $S^2 = 0.35$.

to as salt-and-pepper noise, arises due to abrupt disturbances in image signals [95], and manifests as random bright and dark spots in an image. This type of corruption introduces sparsely distributed white pixels in the dark regions and black pixels in the bright areas, producing a granular distortion pattern [96]. To mimic real-world distortions often encountered in sensor data due to issues like analog-to-digital converter (ADC) faults and transmission bit inconsistencies, salt-and-pepper noise with varying intensity levels has been introduced into the flag-stick image samples. Fig. 4.1 displays representative visual cues affected by salt-and-pepper noise across multiple intensity levels. For a noise density of $\delta\%$, nearly $\delta\%$ of the total pixel values in the corresponding input image become corrupted as a result of this simulated degradation.

Multiplicative noise, on the other hand, is typically caused by spurious signals that become amplified during image acquisition or processing stages. Speckle noise, a specific category of multiplicative noise, is frequently observed in radar and medical imaging contexts. Potential sources of speckles in vision systems include particle accumulation on the camera sensor surface and abrupt variation in the gain of the sensor array [97]. To emulate the grainy noise patterns that typically arise due to interference in coherent imaging setups, as well as from

Table 4.1: EVALUATION OF CLASSIFICATION ACCURACY FOR VARIOUS DIMENSIONALITY REDUCTION METHODS ON THE ORIGINAL VISUAL CUE DATASET

Algorithm	**Classification Accuracy (%)**
PCA	99.048
LDA	99.365
LPP	99.683
LBP-BLPP	99.965
LDP-BLPP	99.965
LTP-BLPP	99.965
LTrP-BLPP	99.965

effects like diffuse surface reflections and scattering particles, speckle noise with varying levels of variance has been incorporated into the system. Fig. 4.2 illustrates several visual cue samples that have been artificially degraded using speckle noise at varying variance levels. The parameter S^2 represents the variance of the multiplicative noise, which is uniformly distributed with mean 0 and variance S^2, where higher values of S^2 induce more severe speckle distortion in the image.

These controlled noise conditions help simulate real-world degradation scenarios and are employed in this work to rigorously assess the noise resilience of the proposed LTrP-BLPP algorithm.

4.2.5 Experimental Results and Discussion

As mentioned earlier in the preceding Section 4.2.4, the experimental evaluation has been conducted using the flag-stick visual cue dataset. Three distinct sets of experiments were designed: one using the original, noise-free dataset; another with images degraded by salt-and-pepper noise; and a third set containing speckle noise-infiltrated samples. For each of these experimentations, a 70-30 train-test partitioning strategy was considered.

To realistically mimic the real-life working environment with sensor noises, the corresponding noise artifacts were intentionally added into both the training and testing image sets. This is motivated by practical scenarios where vision sensors undergo deterioration due to aging, mechanical wear, or environmental stress, leading to noise contamination in all captured

Table 4.2: CLASSIFICATION ACCURACIES (%) OF DIFFERENT DIMENSIONALITY REDUCTION METHODS UNDER VARYING SALT-AND-PEPPER NOISE DENSITIES

Algorithm	$\delta = 0.05$	$\delta = 0.10$	$\delta = 0.15$	$\delta = 0.20$	$\delta = 0.25$	$\delta = 0.30$	$\delta = 0.35$
PCA	68.254	64.127	61.270	55.556	53.651	49.841	48.889
LDA	57.143	55.238	52.063	48.254	40.635	36.190	33.333
LPP	87.302	83.810	80.952	77.143	76.508	73.333	70.794
LBP-BLPP	98.095	94.921	92.063	90.159	88.254	85.714	84.762
LDP-BLPP	98.095	95.238	92.063	90.476	89.206	86.032	84.762
LTP-BLPP	**98.413**	95.556	93.333	**91.111**	89.524	86.667	85.079
LTrP-BLPP	**98.413**	**97.460**	**94.286**	**91.111**	**89.841**	**87.302**	**85.397**

Table 4.3: CLASSIFICATION ACCURACIES (%) OF DIFFERENT DIMENSIONALITY REDUCTION METHODS UNDER VARYING SPECKLE NOISE VARIANCES

Algorithm	$S^2 = 0.05$	$S^2 = 0.10$	$S^2 = 0.15$	$S^2 = 0.20$	$S^2 = 0.25$	$S^2 = 0.30$	$S^2 = 0.35$
PCA	66.984	63.809	62.222	61.269	55.873	52.064	48.889
LDA	56.190	51.429	49.841	47.619	43.175	40.952	38.413
LPP	90.476	89.206	87.302	85.714	81.905	78.730	77.778
LBP-BLPP	**99.365**	97.460	95.873	95.238	90.476	89.841	87.302
LDP-BLPP	**99.365**	97.460	95.873	95.556	91.429	90.159	87.302
LTP-BLPP	**99.365**	97.778	96.190	95.873	91.111	90.476	87.619
LTrP-BLPP	**99.365**	**98.730**	**98.730**	**98.413**	**94.603**	**92.381**	**89.206**

data. Consequently, both training and testing samples would inherently carry such distortions, making it essential to embed noise during both phases for faithful algorithm evaluation.

For the original clean dataset, Table 4.1 reports the classification performance of several widely-used dimensionality reduction methods, evaluated in terms of recognition accuracy. While most standard techniques achieved over 99% accuracy, the BLPP-based methods that incorporated feature descriptors surpassed even 99.90% in performance. To mimic common sensor anomalies such as those arising from analog-to-digital conversion errors or transmission bit faults, salt-and-pepper noise of varying densities was artificially introduced, as shown earlier in Fig. 4.1. Table 4.2 presents the performance of the evaluated methods under increasing levels of salt-and-pepper noise, with $\delta\%$ being the noise density. To simulate textured noise patterns typical of coherent imaging systems; caused by interference, scattering particles, or diffuse reflections, speckle noise with different variances was added, as shown earlier in Fig. 4.2.

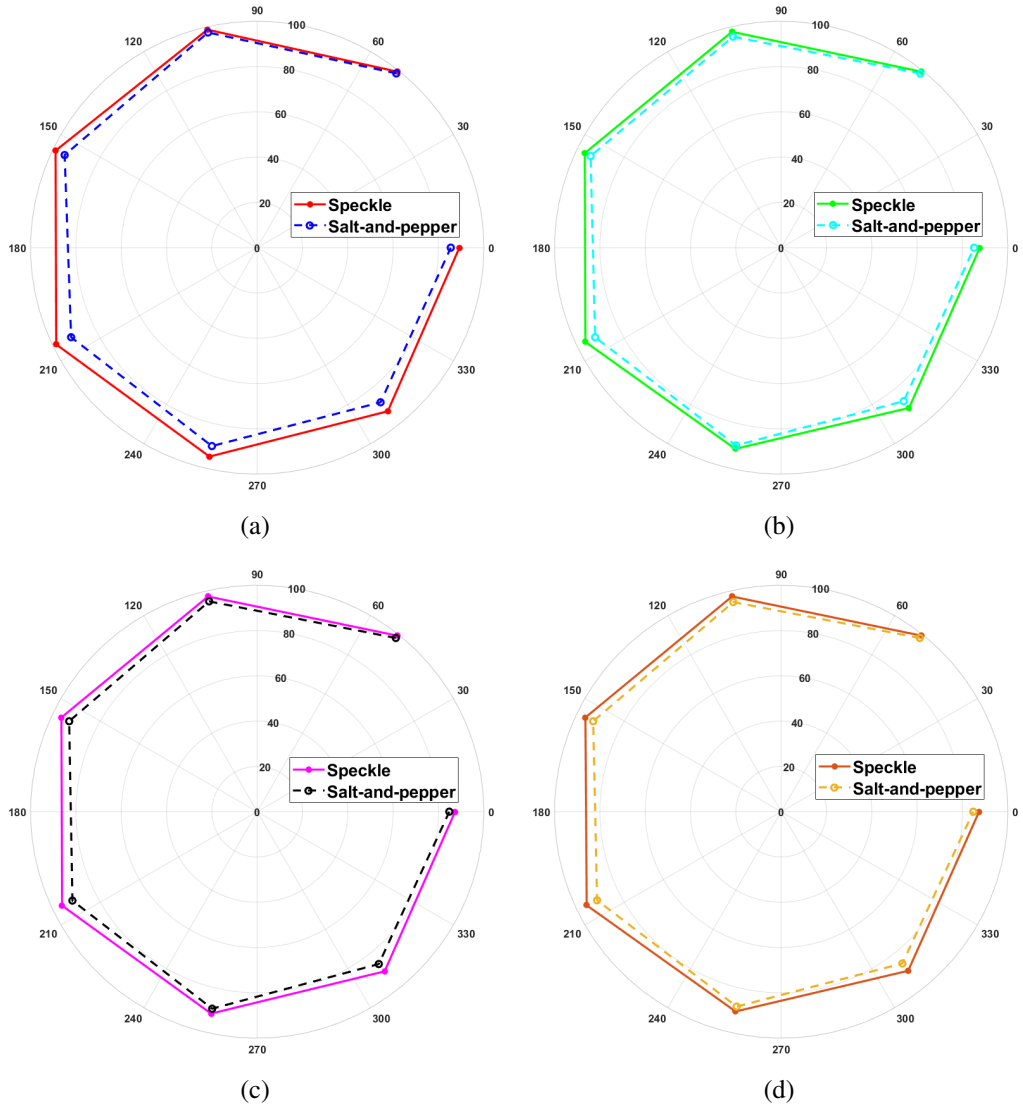


Figure 4.3: Polar plots illustrating the classification accuracy of various local descriptor-based algorithms across increasing levels of salt-and-pepper and speckle noise: (a) LTrP-BLPP, (b) LTP-BLPP, (c) LDP-BLPP, and (d) LBP-BLPP.

The corresponding results are detailed in Table 4.3, where S^2 denotes the variance level of the introduced noise profile.

Both comparative tables clearly demonstrate the robustness of BLPP algorithms augmented with texture descriptors. As noise levels increase, traditional methods like PCA and LDA exhibit notable drops in recognition accuracy, particularly at higher corruption levels. Although LPP initially outperforms PCA and LDA in terms of classification robustness, its performance also declines significantly with increasing noise intensity. In contrast, the BLPP variants with local descriptors such as LBP, LDP, LTP, and LTrP maintain much higher accuracy

across noise densities. Notably, for salt-and-pepper noise with a density of $\delta = 0.20$, these methods have still achieved over 90% classification accuracy. Among all these methods, LTrP-BLPP has consistently outperformed others at every noise level evaluated, demonstrating superior resilience to sensor-induced distortions. When tested with speckle noise, the robustness of the proposed method becomes even more evident. All competing local pattern-based BLPP approaches retained over 90% accuracy up to a noise variance of $S^2 = 0.30$.

The classification trends of the four BLPP-based methods—LTrP-BLPP, LTP-BLPP, LDP-BLPP, and LBP-BLPP, across various noise levels are illustrated in Fig. 4.3. In these polar graphs, noise intensity levels within the range $[0, 0.35]$ are mapped angularly over a 360° circular span. These experimental findings establish that the proposed LTrP-BLPP framework exhibits an escalation in robustness and consistent performance in the presence of substantial sensor noise, affirming its effectiveness for real-world visual recognition tasks.

4.3 Histogram Refinement Scheme for Feature Enhancement

To improve the discriminative capacity of histogram-based feature descriptors, a histogram refinement scheme has been introduced in the work of Tiwari *et al.* [92]. This approach aims to subdivide the global histogram features by categorizing image pixels based on local statistical or structural properties, thereby enhancing the retrieval and recognition performance in the presence of visually similar textures. The entire scheme of histogram refinement is developed based on two components: local skew pattern (LSP) and binarized eigenvalue map (BEM), utilizing their neighborhood-based local properties. These two components are discussed in the following sections.

4.3.1 Histogram Refinement via Local Skew Pattern (LSP)

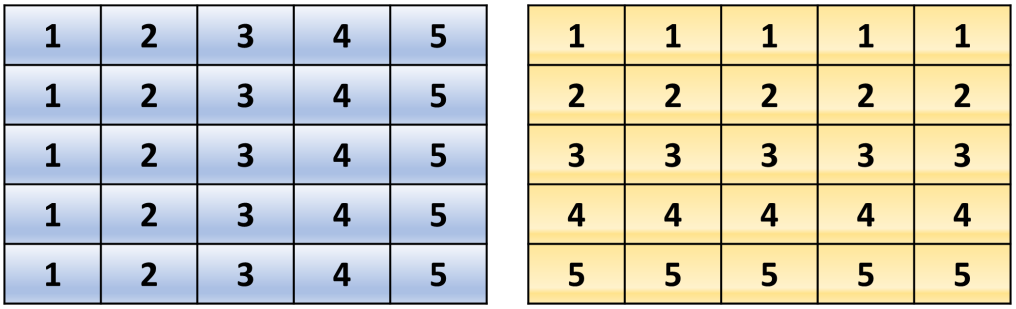
The *local skew pattern* (LSP) aims to enrich the discriminative capacity of feature descriptors by incorporating statistical information regarding the asymmetry of local intensity distributions. For traditional descriptors like LBP, even with similar encodings, the neighborhood distributions can differ notably, suggesting that standard descriptors may miss subtle structural distinctions there [92]. To address this, the LSP framework introduces a statistical analysis of local

neighborhoods by focusing on a specific measure i.e., *skewness*, which captures the asymmetry of pixel value distributions relative to their mean.

The objective is to enhance the separability of texture features by classifying each pixel based on the nature of skewness in its surrounding region. Skewness, as defined in statistical terms, measures how much a distribution deviates from symmetry. A positively skewed distribution is characterized by a longer tail extending towards higher intensity values, whereas a negatively skewed distribution displays a tail that stretches toward lower intensity values. This classification strategy yields a binary map across the image, which is termed the Local Skew Pattern. Let $\mathscr{P}_{g,h}$ denotes the set of gray-level intensities centered at pixel (g,h) and including its 8-connected neighborhood. The LSP at (g,h) is then defined using a median-mean comparison rule as follows:

$$\mathrm{LSP}(g,h) = \begin{cases} 0, & \text{if Median}(\mathscr{P}_{g,h}) \geq \text{Mean}(\mathscr{P}_{g,h}) \\ 1, & \text{otherwise} \end{cases} \tag{4.12}$$

Pixels assigned with values 1 and 0 in Equation (4.12) approximately represent the neighborhoods with positively and negatively skewed intensity distributions, respectively. Accordingly, partitioning of the global texture descriptor histogram into two segments based on LSP classification is anticipated to enhance inter-image separability, thereby improving the performance of content-based image encoding. However, upon applying histogram refinement, two of those pixels now contribute to a sub-histogram derived from regions where LSP equals 1, while the remaining two are associated with the sub-histogram corresponding to LSP value 0. In essence, the LSP-based refinement divides the original local descriptor histogram into two distinct subsets; one representing positively skewed regions and the other representing negatively skewed neighborhoods. This classification and subdivision of pixel patterns introduce a new degree of granularity into the overall descriptor. The final descriptor is then constructed by concatenating the two refined histograms into a single histogram vector. During the matching phase, the histogram refinement ensures that comparisons are made between patterns derived from neighborhoods of similar skewness, thereby embedding vital information about local distribution characteristics into the similarity computation.



(a) (b)

Figure 4.4: Kernels used for computing the covariance matrix over a 5×5 local neighborhood: (a) Kernel $\mathcal{K}^{g\_\texttt{cent}}$ for estimating the g-coordinate of the centroid, and (b) Kernel $\mathcal{K}^{h\_\texttt{cent}}$ for estimating the h-coordinate of the centroid.

4.3.2 Histogram Refinement via Binarized Eigenvalue Map (BEM)

This section describes the methodology for generating the binarized eigenvalue map (BEM) using principal component analysis (PCA), applied over the local neighborhood of each pixel. The main goal of the proposed histogram refinement approach using BEM is to incorporate structural characteristics present in the local region of the image. To achieve this, PCA is performed within the neighborhood around every pixel, a process referred to as local PCA. In the current approach, the covariance matrix required for local PCA is derived using a set of second-order image moments [98]. Image moments are statistical measures that represent the intensity-weighted distribution of pixel values within a given local neighborhood. The kernels employed to compute the elements of the covariance matrix within a 5×5 pixel window centered at each pixel, are depicted in Fig. 4.4.

Assuming $A_{g,h}$ to be the grayscale image under consideration, the second-order moments that contribute to the construction of the covariance matrix, need to be calculated. Variance along the g-direction, h-direction, and covariance between g- and h-directions are obtained as follows:

$$G_{g,h}^{\texttt{var}} = \sum_{p=1}^{m}\sum_{q=1}^{m}(\mathcal{K}_{p,q}^{g\_\texttt{cent}} - G_{g,h}^{\texttt{cent}})^2 \cdot A_{(g+p-\frac{m-1}{2},h+q-\frac{m-1}{2})} \tag{4.13}$$

$$H_{g,h}^{\texttt{var}} = \sum_{p=1}^{m}\sum_{q=1}^{m}(\mathcal{K}_{p,q}^{h\_\texttt{cent}} - H_{g,h}^{\texttt{cent}})^2 \cdot A_{(g+p-\frac{m-1}{2},h+q-\frac{m-1}{2})} \tag{4.14}$$

$$\mathtt{Cov}_{g,h} = \sum_{p=1}^{m}\sum_{q=1}^{m}(\mathcal{K}_{p,q}^{g\_\mathtt{cent}} - G_{g,h}^{\mathtt{cent}})(\mathcal{K}_{p,q}^{h\mathtt{cent}} - H_{g,h}^{\mathtt{cent}}) \cdot A_{(g+p-\frac{m-1}{2},h+q-\frac{m-1}{2})} \tag{4.15}$$

where

$$G_{g,h}^{\mathtt{cent}} = \frac{\sum_{p=1}^{m}\sum_{q=1}^{m}\mathcal{K}_{p,q}^{g\_\mathtt{cent}} \cdot A_{(g+p-\frac{m-1}{2},h+q-\frac{m-1}{2})}}{\sum_{p=1}^{m}\sum_{q=1}^{m}A_{(g+p-\frac{m-1}{2},h+q-\frac{m-1}{2})}} \tag{4.16}$$

$$H_{g,h}^{\mathtt{cent}} = \frac{\sum_{p=1}^{m}\sum_{q=1}^{m}\mathcal{K}_{p,q}^{h\_\mathtt{cent}} \cdot A_{(g+p-\frac{m-1}{2},h+q-\frac{m-1}{2})}}{\sum_{p=1}^{m}\sum_{q=1}^{m}A_{(g+p-\frac{m-1}{2},h+q-\frac{m-1}{2})}} \tag{4.17}$$

Equations (4.16) and (4.17) define the G and H coordinates of the centroid, calculated for the local neighborhood centered at the pixel location (g,h). While formulating the Equations (4.13) to (4.17), a square neighborhood of dimensions $m \times m$ pixels has been assumed for computation. Using these centroid and moment values, the local covariance matrix is subsequently formulated as follows:

$$\mathtt{Cov}(A_{g,h}) = \begin{bmatrix} G_{g,h}^{\mathtt{var}} & \mathtt{Cov}_{g,h} \\ \mathtt{Cov}_{g,h} & H_{g,h}^{\mathtt{var}} \end{bmatrix} \tag{4.18}$$

Performing eigenvalue decomposition on the covariance matrix defined above results in two distinct eigenvalues. This decomposition is carried out at each pixel location in the image, and the principal eigenvalue obtained from each pixel's neighborhood is used to construct an eigenvalue map. Thus, the eigenvalue map generated is of the same spatial dimensions as the original grayscale image. To normalize the values in this map, each principal eigenvalue is divided by the intensity of the center pixel around which the corresponding local neighborhood has been considered. It can be analytically demonstrated that, in regions of uniform intensity, this normalization yields a constant principal eigenvalue. In fact, for such constant-intensity regions, the resulting normalized value is dependent solely on the size of the neighborhood window used for local PCA.

4.3.3 Hybrid Frameworks for Histogram Refinement

This section outlines the hybrid frameworks for the effective integration of local skew pattern (LSP) and binarized eigenvalue map (BEM) in the histogram refinement process [92]. Three

such frameworks have been investigated, each differing in how LSP and BEM are combined with the local texture descriptors to enhance feature representation. The *first framework* adopts a parallel refinement strategy. In this design, the input image is processed independently through both LSP and BEM-based refinement pipelines. Specifically, the original histogram of local descriptors is refined separately using LSP and BEM, resulting in two refined histograms from each path. These four histograms, two from LSP and two from BEM are then concatenated to form the final histogram feature vector.

The second and third frameworks implement serial refinement mechanisms, wherein the output of one refinement stage is used as input to the next. In both of these serial frameworks, either the LSP or the BEM is first encoded at the pixel level, creating a binary pattern for each pixel based on its eight nearest neighbors. This binarization enables histogram computation using uniform binary patterns, similar to the technique employed for local descriptors like LBP or LTP. In the *second framework*, the histogram of the coded BEM is first concatenated with the original local descriptor histogram. This concatenated feature vector is then subjected to histogram refinement using the LSP-based classification of pixels.

In contrast, the *third framework* reverses the order. Here, the LSP is initially encoded and its histogram is combined with that of the original local descriptors. The resulting concatenated histogram feature is subsequently refined using the BEM-driven binarization strategy. Both serial frameworks aim to inject complementary information from LSP and BEM into the final representation, but do so, through differing order and processing paths.

Among these, *framework* 3 offers the most significant improvement in performance, as the LSP captures fine details while BEM provides stronger structural differentiation, in sequential order.

4.4 Feature Encoding with Histogram Refined Local Binary Patterns (HRLBP)

Accurate feature representation under varying illumination and sensor-induced distortions is essential for visual cue recognition in real-world human-robot interaction situations. While traditional texture descriptors such as local binary pattern (LBP) offer a compact and efficient encoding of micro-structures, their performance can deteriorate in the presence of illumination

variations and fine structural distortions. As a redressal, the present work incorporates the histogram refinement strategy, as discussed earlier, into the LBP framework, thereby yielding *histogram refined local binary pattern* (HRLBP) features with enhanced discriminative capability. The refinement process involves the classification of pixels based on local statistical or structural characteristics, such as skewness or eigenvalue variation, which allows the original LBP histogram to be subdivided into more semantically coherent segments. This subdivision enables finer discrimination between visually similar texture regions and introduces resilience to localized noise and illumination shifts.

The effectiveness of the HRLBP descriptor is evaluated under three distinct and moderately challenging lighting conditions: normal illumination, dark condition level-1, and dark condition level-2. These dark-level conditions are deliberately introduced to simulate practical conditions, where vision sensors operate under insufficient ambient light or dynamic exposure settings. The goal is to assess whether the HRLBP encoding, with its histogram-aware enhancement, can maintain robust performance where conventional LBP features may struggle due to underexposure or signal degradation. The following subsections present the mathematical formulation of LBP, followed by the proposed histogram refinement scheme applied to LBP, and its subsequent BLPP-based projection scheme, culminating in experimental evaluation under the aforementioned illumination conditions.

4.4.1 Mathematical Background of LBP

The local binary pattern (LBP) is one of the most widely adopted and computationally efficient texture descriptors for local feature extraction in grayscale images. Originally introduced by Ojala *et al.* [30], the LBP operator captures the micro-texture structure around each pixel by thresholding its neighboring pixel values with respect to the intensity of the center pixel. Given a central pixel A_c and its P surrounding neighbors located on a neighborhood of radius $\mathfrak{R}$, the LBP code is mathematically defined as:

$$\mathrm{LBP}_{P,\mathfrak{R}} = \sum_{p=0}^{P-1} s(A_p - A_c) \cdot 2^p, \tag{4.19}$$

where A_p represents the gray-level intensity of the p^{th} neighbor, and $s(x)$ is a sign function defined as:

$$s(x) = \begin{cases} 1, & \text{if } x \geq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{4.20}$$

The resulting binary string is then converted into a decimal number, which uniquely represents the local texture pattern at that pixel location. Typically, the operator is applied in a 3×3 window resulting in 8-bit binary codes (i.e., $P = 8, R = 1$), allowing up to 256 possible patterns. To achieve robustness against rotation and noise, the LBP descriptor is often further improved using variants such as rotation-invariant LBP, uniform LBP, and completed LBP. A common practice is to encode only those binary patterns with at most two transitions (uniform patterns), thus significantly reducing the histogram dimensionality while preserving discriminative power. This compact and effective representation has found wide utility in various image processing applications including texture classification, face recognition, gesture detection, and biomedical image analysis due to its simplicity and effectiveness in capturing local spatial variations.

4.4.2 HRLBP-Encoded Low-Dimensional Projection Scheme HRLBP-BLPP

Following the extraction of histogram refined LBP (HRLBP) features from each image sample using *framework* 3 of the proposed histogram refinement pipeline, the resulting high-dimensional feature representation is projected into a lower-dimensional subspace through the bilateral locality preserving projection (BLPP) [11] scheme, as described earlier in Section 3.5 and Section 3.6 of Chapter 3. The hybrid method developed here, referred to as HRLBP-BLPP, aims to retain both the spatial continuity and feature-level similarity present in the original data during dimensionality reduction.

Let the collection of all input images in vectorized form, be represented by $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_N] \in \mathbb{R}^{D \times N}$, where each column vector $\mathbf{A}_i \in \mathbb{R}^D$ corresponds to a vectorized form of the input image, and N denotes the number of training instances. The objective is to compute a projection matrix $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_d] \in \mathbb{R}^{D \times d}$ that transforms each high-dimensional image vector $\mathbf{A}_i$ into a lower-dimensional representation $\mathbf{x}_i = \mathbf{W}^\top \mathbf{A}_i \in \mathbb{R}^d$, where $d \ll D$.

Unlike conventional LPP, where neighborhood similarity is determined using Euclidean

distance in the spatial domain alone, the HRLBP-BLPP model incorporates both spatial and HRLBP feature similarities for computing local affinities. The spatial weight between two samples $\mathbf{A}_i$ and $\mathbf{A}_j$ is computed as:

$$s_{ij} = \begin{cases} \exp\left(-\dfrac{\|\mathbf{A}_i - \mathbf{A}_j\|_F^2}{\tau}\right), & \text{if } \mathbf{A}_j \in \mathcal{N}_k(\mathbf{A}_i) \\ 0, & \text{otherwise} \end{cases} \tag{4.21}$$

Simultaneously, the feature-level similarity is derived from the histogram refined HRLBP descriptor. Let $h(\mathbf{A}_i)$ denote the HRLBP feature vector associated with image $\mathbf{A}_i$. The feature similarity weight is then defined as:

$$r_{ij} = \begin{cases} \exp\left(-\dfrac{\|h(\mathbf{A}_i) - h(\mathbf{A}_j)\|_2^2}{\sigma}\right), & \text{if } \mathbf{A}_j \in \mathcal{N}_k(\mathbf{A}_i) \\ 0, & \text{otherwise} \end{cases} \tag{4.22}$$

Here, τ and σ are scalar parameters that regulate the decay of similarity in the spatial and feature domains, respectively. These two weight values are multiplicatively combined to form the final bilateral affinity:

$$t_{ij} = s_{ij} \cdot r_{ij} \tag{4.23}$$

The resulting joint affinity matrix $\mathbf{T} = [t_{ij}]$ reflects the mutual similarity of neighboring samples with respect to both spatial configuration and histogram-refined LBP features. The diagonal matrix $\mathbf{D}$ is computed with diagonal entries $D_{ii} = \sum_j t_{ij}$, and the corresponding graph Laplacian is defined as $\mathbf{L} = \mathbf{D} - \mathbf{T}$. The optimization problem for HRLBP-BLPP becomes:

$$\min_{\mathbf{W}} \operatorname{tr}(\mathbf{W}^\top \mathbf{A}\mathbf{L}\mathbf{A}^\top \mathbf{W}) \quad \text{subject to} \quad \mathbf{W}^\top \mathbf{A}\mathbf{D}\mathbf{A}^\top \mathbf{W} = \mathbf{I} \tag{4.24}$$

Solving this leads to the generalized eigenvalue problem:

$$\mathbf{A}\mathbf{L}\mathbf{A}^\top \mathbf{w} = \lambda \mathbf{A}\mathbf{D}\mathbf{A}^\top \mathbf{w} \tag{4.25}$$

The solution yields the projection directions $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_d\}$ corresponding to the d smallest non-zero eigenvalues. These form the columns of the optimal transformation matrix $\mathbf{W}$. The final low-dimensional representation $\mathbf{x}_i = \mathbf{W}^\top \mathbf{A}_i$ encodes both spatial layout and HRLBP-enhanced local texture statistics. This fusion enables the HRLBP-BLPP framework to preserve class-discriminative structural patterns while reducing the sensitivity to lighting and sensor noise variations, thereby contributing to more robust visual cue recognition.

Table 4.4: CLASSIFICATION ACCURACY OF COMPETING METHODS UNDER NORMAL ILLUMINATION

Method	Accuracy (%)
PCA	99.048 ± 0.184
LDA	99.365 ± 0.165
LPP	99.683 ± 0.105
HOG-BLPP	99.690 ± 0.102
Gabor-BLPP	99.745 ± 0.080
LBP-BLPP	99.965 ± 0.011
HRLBP-BLPP	$\mathbf{99.969 \pm 0.009}$

4.4.3 Experimental Results Under Moderately Challenging Illumination Conditions

For the purpose of experimentation, the full image dataset was randomly split into training and testing subsets, maintaining a consistent $70:30$ ratio. The testing phase employed a set of images under more adverse photometric conditions, as previously illustrated. To compute both spatial and range-based similarity weights, the number of nearest neighbors was fixed at $k = 6$. The target dimensionality d of the projected subspace was determined such that it retains the most significant principal components carrying maximum energy from the input vectors.

Table 4.4 presents a comparative analysis of multiple algorithms tested using the image database collected under well-lit, normal lighting conditions. Under this standard illumination, all benchmark algorithms demonstrated notably high classification accuracy. Specifically, classical dimensionality reduction techniques such as PCA [3], LDA [4], and LPP [7] each achieved over 99% accuracy. The BLPP [11] variants using LBP, and histogram refined LBP, further improved upon this, with classification accuracies exceeding 99.90%. These experimental findings emphasize the practical applicability and effectiveness of the proposed approach in flag-stick cue-based robot navigation guidance framework in regular working environments.

In environments where lighting is compromised, the effectiveness of traditional feature extraction and dimensionality reduction strategies significantly declines. To evaluate the

Table 4.5: CLASSIFICATION ACCURACY OF COMPETING METHODS ON PHOTOMETRICALLY CHALLENGED DATASET

Method	**Accuracy (%) at DIL-1**	**Accuracy (%) at DIL-2**
PCA	70.253 ± 0.352	46.349 ± 0.478
LDA	81.296 ± 0.296	73.333 ± 0.368
LPP	92.362 ± 0.255	84.127 ± 0.349
HOG-BLPP	94.426 ± 0.241	88.889 ± 0.234
Gabor-BLPP	95.217 ± 0.270	90.476 ± 0.215
LBP-BLPP	96.329 ± 0.159	94.603 ± 0.196
HRLBP-BLPP	$\mathbf{97.235 \pm 0.136}$	$\mathbf{95.238 \pm 0.193}$

robustness of the proposed approach under such adverse conditions, another round of testing was performed. Table 4.5 shows the comparative results of the same set of methods applied to the photometrically deteriorated dataset DIL-1 and DIL-2, previously described in Section 2.4 and Section 2.5, in Chapter 2. In these challenging conditions, the HRLBP-BLPP model surpassed all other evaluated algorithms in terms of classification accuracy and robustness. The histogram refinement technique within HRLBP-BLPP effectively captures local patterns even under disrupted pixel distributions caused by poor lighting. By incorporating both Euclidean-based spatial weights and feature-driven locality weights, the method shows considerable robustness with degraded visual inputs.

A confusion matrix, generated for HRLBP-BLPP under the condition DIL-2 is shown in Fig. 4.5, which illustrates the classification results in terms of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). Several classification performance indicators based on these values were also calculated, as detailed in Table 4.6. The following

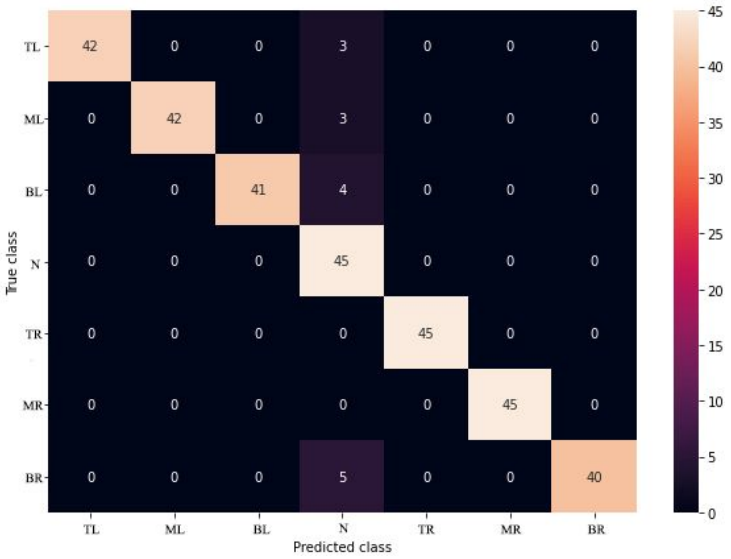


Figure 4.5: Confusion matrix illustrating the classification performance of the HRLBP-BLPP method.

equations were used to compute various classification metrics:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \tag{4.26}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{4.27}$$

$$\text{Sensitivity (Recall)} = \frac{TP}{TP+FN} \tag{4.28}$$

$$\text{Specificity} = \frac{TN}{TN+FP} \tag{4.29}$$

$$\text{Negative Predictive Value (NPV)} = \frac{TN}{TN+FN} \tag{4.30}$$

$$\text{False Positive Rate (FPR)} = 1 - \text{Specificity} \tag{4.31}$$

$$F1\text{-score} = \frac{2 \times \text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \tag{4.32}$$

$$\text{Matthews Correlation Coefficient (MCC)} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TN+FP)(TP+FN)(TN+FN)}} \tag{4.33}$$

$$\text{Cohen's Kappa } (\kappa) = \frac{2(TP \cdot TN - FP \cdot FN)}{(TP+FP)(FP+TN)+(TP+FN)(FN+TN)} \tag{4.34}$$

Table 4.6 summarizes the values of these indices attained for HRLBP-BLPP and its competing counterparts. Notably, the HRLBP-BLPP method consistently achieved superior scores across all measures, underscoring its resilience and suitability for real-world scenarios with moderately challenging illumination.

Table 4.6: A STUDY OF VARIOUS CLASSIFICATION METRICS OBTAINED WITH HRLBP-BLPP WITH THE DATASET ACQUIRED AT DIL-2

Metric (%)	**HRLBP-BLPP**	**LBP-BLPP**	**Gabor-BLPP**	**HOG-BLPP**	**LPP**
Precision	96.43	96.08	91.74	90.18	92.20
Recall	95.24	94.60	90.48	88.89	84.13
Specificity	99.21	99.10	98.41	98.15	97.35
NPV	99.22	99.12	98.42	98.15	97.46
FPR	0.79	0.90	1.59	1.85	2.65
F1-score	95.47	94.88	90.75	89.18	85.67
MCC	94.93	94.30	89.41	87.57	84.81
κ-score	80.56	77.96	61.11	54.63	35.19

4.5 Feature Encoding with Histogram Refined Local Ternary Patterns (HRLTP)

The development of noise-resilient and illumination-aware feature encoders is vital for ensuring consistent recognition accuracy in real-world situations, where visual data often suffers from degradation. The previously discussed schemes, LTrP-BLPP and HRLBP-BLPP, have shown promising results, especially LTrP-BLPP demonstrating robustness against sensor-induced noise artifacts, and HRLBP-BLPP performing effectively under moderate lighting deficiencies. However, real-life image acquisition conditions are rarely affected by a single source of degradation. More commonly, a combination of low-light environments and signal distortions co-occur due to the aging of sensors, analog-to-digital (ADC) conversion errors, and ambient interference.

To address such combined distortions, the present study proposes a histogram refined local ternary pattern (HRLTP)-based bilateral locality preserving projection (BLPP) framework. The HRLTP descriptor enhances the original LTP encoding [31], which itself is a generalization of LBP designed to be less sensitive to random noise while preserving edge structures. By integrating the histogram refinement [92] strategy into the LTP computation, the HRLTP descriptor improves its capacity to distinguish between subtle variations in texture caused by simultaneous illumination drops and additive or multiplicative noise.

The proposed HRLTP-BLPP scheme is specifically tailored to handle image sets captured under hybrid dilapidating conditions i.e., those affected both by reduced ambient light and by salt-and-pepper or speckle noise. The experimental evaluation conducted in this study incorporates visual cue data acquired under such conditions, combining the effects of photometric challenges and sensor noise. The integration of histogram-based local neighborhood characterization with ternary pattern encoding, aims to yield an enriched feature space with improved class separability. The following subsections provide the mathematical modeling of LTP, and the full HRLTP-BLPP framework, followed by a detailed performance analysis.

4.5.1 Mathematical Modeling of LTP

Local ternary pattern (LTP) is a noise-tolerant variant of the local binary pattern (LBP) and is particularly successful in representing image textures under uncertain or degraded conditions. Introduced by Tan and Triggs [31], the LTP encoding scheme is built upon a simple yet robust pixel-thresholding rule, offering better resilience against small intensity fluctuations that typically arise from noise.

For a given central pixel A_c in a grayscale image A and its neighborhood pixels A_p, where $p \in \{0,1,\ldots,P-1\}$, the ternary pattern is generated using the following decision rule:

$$\mathrm{LTP}(A_p,A_c) = \begin{cases} +1, & \text{if } A_p \geq A_c + \tau \\ 0, & \text{if } |A_p - A_c| < \tau \\ -1, & \text{if } A_p \leq A_c - \tau \end{cases} \tag{4.35}$$

Here, τ is a user-defined threshold that determines the dead zone for pixel comparison. The pixel pairs where the difference is within the interval $[-\tau,+\tau]$ are mapped to a neutral ternary state, making the descriptor less sensitive to minor changes in illumination or sensor noise.

The full ternary pattern thus consists of three possible states: $+1$, 0, and -1. To make this pattern computationally feasible and compatible with standard histogram-based frameworks, it is commonly split into two separate binary patterns. One binary pattern captures the positive transitions ($+1$ vs. others), while the second binary map encodes the negative transitions (-1 vs. others). These two binary patterns are then individually processed to obtain two corresponding histograms [92], which are concatenated to produce the final LTP histogram descriptor. This

decomposition not only preserves the edge information in both gradient directions but also enhances robustness against outlier pixel intensities.

The LTP operator is generally applied over a local 3×3 window, though it can be extended to larger or circular neighborhoods for rotation invariance and scale adaptability. The resulting descriptor offers a richer representation of texture by capturing fine directional patterns while ignoring insignificant pixel variations [31]. This formulation of LTP provides a solid foundation for developing more robust texture encoders such as the histogram-refined LTP (HRLTP), which combines ternary encoding with structural refinement for handling hybrid distortions.

4.5.2 HRLTP-Encoded Low-Dimensional Projection Scheme HRLTP-BLPP

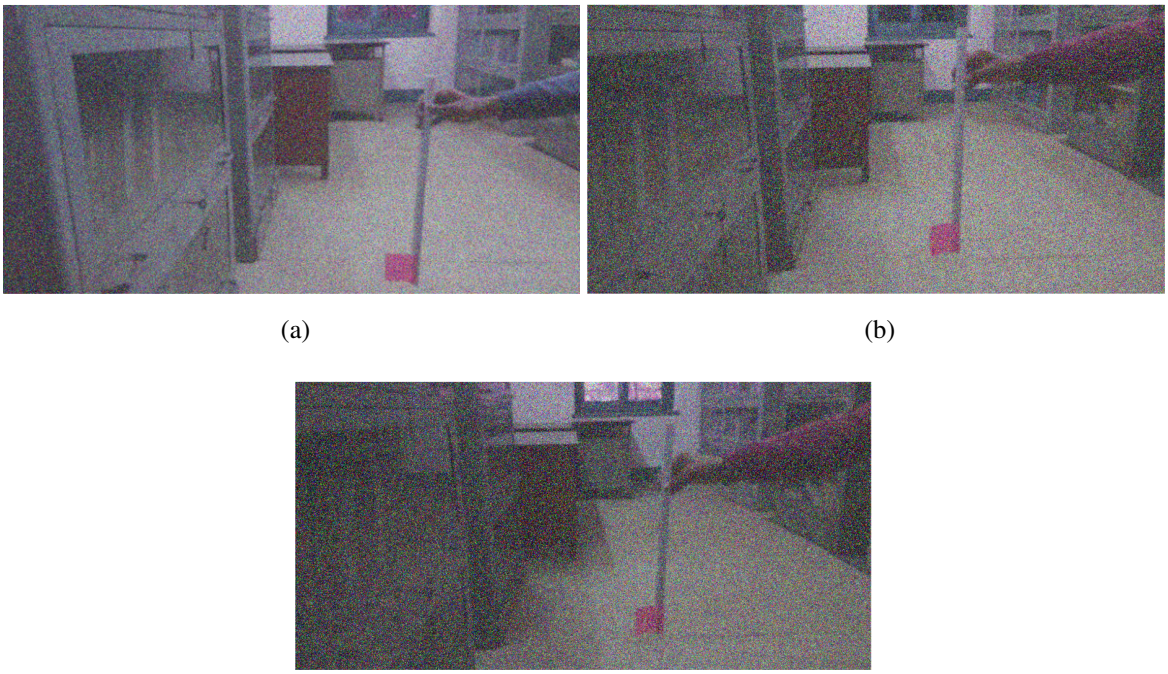
Following the extraction and refinement of LTP features, HRLTP features are obtained using the framework described earlier in Section 4.4.2, a compact and discriminative low-dimensional representation is obtained by applying the bilateral locality preserving projection (BLPP) technique [11]. This step ensures that both spatial proximity and feature-level similarity of the HRLTP descriptors are retained during dimensionality reduction.

Similar to HRLBP, the feature-level similarity information is obtained from the histogram refined HRLTP descriptor. With $g(\mathbf{A}_i)$ defining the HRLTP feature vector for the image $\mathbf{A}_i$, the feature similarity kernel is constructed as:

$$r_{ij} = \begin{cases} \exp\left(-\dfrac{\|g(\mathbf{A}_i) - g(\mathbf{A}_j)\|_2^2}{\sigma}\right), & \text{if } \mathbf{A}_j \in \mathcal{N}_k(\mathbf{A}_i) \\ 0, & \text{otherwise} \end{cases} \tag{4.36}$$

In line with earlier formulations, the bilateral weights are computed by combining the spatial similarity of image instances and the closeness of their HRLTP-based histograms. The feature similarity weight r_{ij} is multiplied with the spatial similarity weight s_{ij} to obtain the combined spatial-feature similarity weight $t_{ij} = s_{ij} \times r_{ij}$. These affinities are used to construct a graph-preserving projection matrix that maps high-dimensional HRLTP features to a lower-dimensional subspace, thereby improving classification performance under jointly degraded visual conditions such as noise and low illumination.

Following this, a similar projection strategy as discussed earlier in Section 4.4.2 is employed to obtain the transformation matrix $\mathbf{W} \in \mathbb{R}^{D \times d}$ and the corresponding projected



(a) (b)

(c)

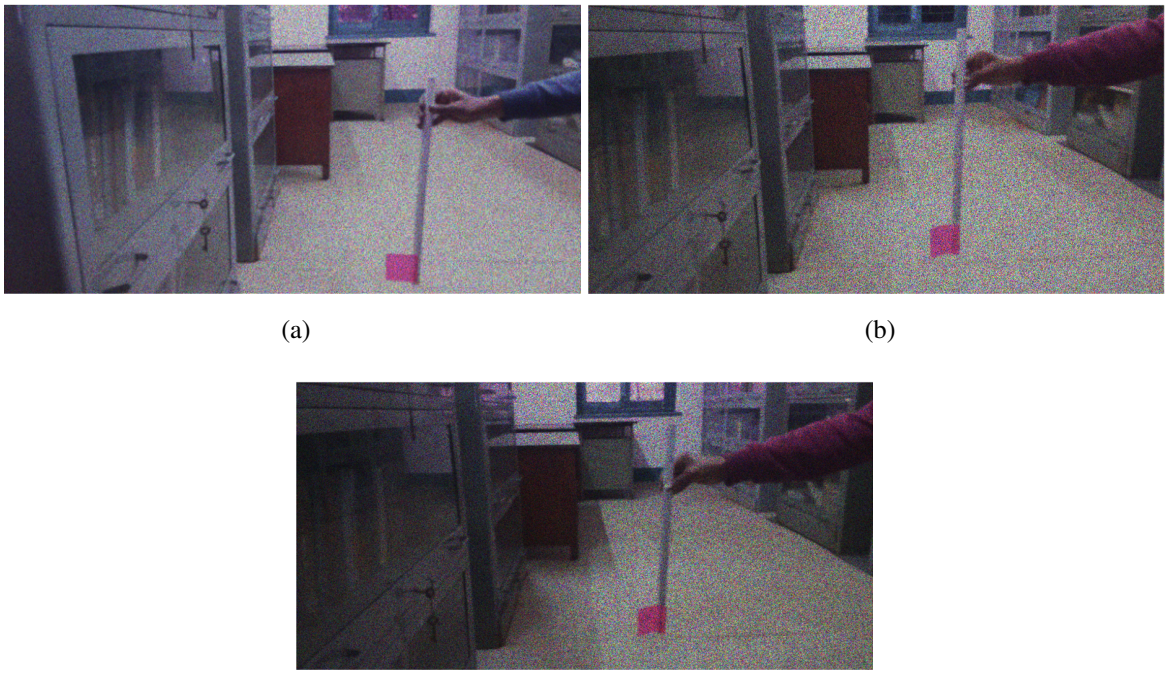
Figure 4.6: Flag-stick visual cue samples contaminated by salt-and-pepper noise with density $\delta = 0.35$ at illumination levels: (a) NIL, (b) DIL-1, and (c) DIL-2.

features as $\mathbf{x}_i = \mathbf{W}^\top \mathbf{A}_i$. This HRLTP-BLPP pipeline facilitates the learning of compact features that are not only structure-aware and noise-tolerant but also photometrically robust, making it well-suited for real-world HRI tasks involving hybrid data distortions.

4.5.3 Visual Cues with Moderate Illumination Degradation and Noise Corruption

To thoroughly evaluate the robustness of the proposed HRLTP-BLPP framework, a challenging visual cue dataset has been prepared, encompassing both moderate photometric deterioration and sensor-induced noise corruption. Initially, visual data acquisition has been carried out under standard lighting conditions inside the laboratory environment, referred to as the normal illumination level (NIL). The complete dataset at NIL comprises a total of 1050 image frames distributed across seven distinct classes. Following this, experimentation with the same setup has been repeated under systematically reduced lighting conditions.

Specifically, two levels of darkened lighting have been considered in this work. Dark illumination level-1 (DIL-1) is achieved by switching off one of the primary light sources, and



(a) (b)

(c)

Figure 4.7: Flag-stick visual cue samples contaminated by speckle noise with variance $\sigma^2 = 0.35$ at illumination levels: (a) NIL, (b) DIL-1, and (c) DIL-2.

dark illumination level-2 (DIL-2) is created by switching off two light sources. To quantitatively assess the lighting variations, an illuminance meter (Metravi 1310) has been used to measure the ambient brightness in Lux. The recorded average illuminance values were 14.910 Lx at NIL, 13.405 Lx at DIL-1, and 11.880 Lx at DIL-2, respectively, indicating controlled yet realistic degradation in lighting.

To emulate real-world distortions commonly encountered due to aging sensors, ADC quantization errors, or external interferences, two categories of noise have been artificially introduced into the dataset. These include impulse noise, simulated using salt-and-pepper noise with varying noise density δ, and multiplicative noise modeled as speckle noise with variance σ^2. Fig. 4.6 illustrates representative samples of flag-stick visual cues captured under NIL, DIL-1, and DIL-2 conditions, each degraded with salt-and-pepper noise at a fixed noise density of $\delta = 0.35$. In a similar fashion, Fig. 4.7 depicts the same set of illumination-conditioned images, this time affected by speckle noise characterized by a variance of $\sigma^2 = 0.35$.

This augmented dataset allows for a rigorous evaluation of the HRLTP-BLPP pipeline, highlighting its capability to preserve local structural features and maintain class separability in

Table 4.7: RECOGNITION ACCURACY (%) OF DIFFERENT DIMENSIONALITY REDUCTION APPROACHES UNDER VARYING LIGHTING CONDITIONS

Method	NIL	DIL-1	DIL-2
PCA	99.048	70.253	46.349
LDA	99.365	81.296	73.333
LPP	99.683	92.362	84.127
LBP-BLPP	99.965	96.329	94.603
LDP-BLPP	99.965	97.112	95.732
LTP-BLPP	99.965	97.798	96.334
HRLBP-BLPP	99.969	97.235	95.238
HRLDP-BLPP	99.970	97.554	95.995
HRLTP-BLPP	**99.972**	**98.107**	**96.887**

visual cues, even under concurrent photometric degradation and stochastic noise contamination. The investigation showcases the algorithm's potential for deployment in real-world robot guidance tasks where environmental reliability cannot be guaranteed.

4.5.4 Experimental Results and Discussion

To comprehensively evaluate the efficacy of the proposed HRLTP-BLPP framework, a series of experiments has been conducted under realistic environmental conditions. The dataset has been divided randomly with a consistent train-test split ratio of 70%-30% for each visual cue class considered in the study. Initially, the system's recognition capabilities have been tested across varying illumination levels; namely normal illumination level (NIL), dark illumination level-1 (DIL-1), and dark illumination level-2 (DIL-2). The training set included only samples captured under NIL, while testing was extended to all three illumination settings. Recognition results yielded by multiple techniques, including standard dimensionality reduction methods like PCA, LDA, and LPP, as well as several bilateral LPP variants with and without histogram refinement, are summarized in Table 4.7. Among these, the HRLTP-BLPP consistently yielded superior accuracy across conditions.

As illumination worsened, a noticeable drop in performance has been observed for the

Table 4.8: RECOGNITION ACCURACY (%) OF DIMENSIONALITY REDUCTION TECHNIQUES WITH SALT-AND-PEPPER NOISE DENSITY $\delta = 0.20$ UNDER VARYING LIGHTING CONDITIONS

Method	NIL	DIL-1	DIL-2
PCA	55.556	47.332	40.766
LDA	48.254	41.088	37.642
LPP	77.143	70.841	62.089
LBP-BLPP	90.159	87.502	84.805
LDP-BLPP	90.476	87.959	85.313
LTP-BLPP	91.111	88.222	85.775
HRLBP-BLPP	90.627	88.641	86.304
HRLDP-BLPP	90.925	89.965	87.538
HRLTP-BLPP	**92.353**	**90.447**	**88.295**

other competing models. However, the proposed HRLTP-BLPP framework demonstrated better stability with minimal performance degradation, showcasing its robustness to adverse lighting. Furthermore, Tables 4.8 and 4.9 display performance results when the test data is corrupted by additional sensor noise at a certain level. Specifically, these include scenarios with added salt-and-pepper and speckle noise, simulating more complex real-world conditions involving both photometric variation and sensory distortion. As expected, the presence of noise has resulted in a decline in accuracy across all methods. Nevertheless, HRLTP-BLPP has managed to preserve better recognition capability than its peers.

To further assess the sensitivity of the model to noise severity, experiments have been carried out under a range of salt-and-pepper noise densities and speckle noise variances, as presented in Tables 4.10 and 4.11, respectively. For salt-and-pepper noise with a density $\delta\%$, approximately $\delta\%$ of the pixel intensities in the image were randomly corrupted, mimicking hardware-induced faults such as ADC errors and transmission issues. The evaluations considered noise densities from 5% up to 35%. In the case of speckle noise, characterized by a variance parameter σ^2, the disturbances represent granular texture patterns typically originating from factors like coherent interference, reflective diffusion, or particulate scattering. Speckle variances were tested up to $\sigma^2 = 0.35$ to observe the degradation trend. As anticipated, recognition accuracy deteriorated progressively with increasing noise intensity.

Table 4.9: RECOGNITION ACCURACY (%) OF DIMENSIONALITY REDUCTION METHODS WITH SPECKLE NOISE VARIANCE $\sigma^2 = 0.20$ UNDER DIFFERENT LIGHTING CONDITIONS

Method	**NIL**	**DIL-1**	**DIL-2**
PCA	61.269	55.814	42.905
LDA	47.619	40.127	36.913
LPP	85.714	78.632	71.097
LBP-BLPP	95.238	92.157	89.970
LDP-BLPP	95.556	92.934	90.485
LTP-BLPP	95.873	93.255	90.699
HRLBP-BLPP	96.047	93.823	90.765
HRLDP-BLPP	96.461	93.917	90.890
HRLTP-BLPP	**96.910**	**94.075**	**91.149**

Table 4.10: RECOGNITION ACCURACY (%) OF HRLTP-BLPP UNDER VARIOUS DENSITIES OF SALT-AND-PEPPER NOISE AT DIFFERENT ILLUMINATION LEVELS

Noise Density (δ)	**NIL**	**DIL-1**	**DIL-2**
0.05	98.632	96.162	94.794
0.10	95.751	93.311	91.528
0.15	93.445	91.165	89.602
0.20	92.353	90.447	88.295
0.25	90.196	87.825	84.654
0.30	87.823	85.263	82.689
0.35	86.206	83.498	80.350

Despite these challenges, HRLTP-BLPP consistently outperformed other algorithms in all test scenarios. Even in the most adverse conditions involving simultaneous heavy noise and dim lighting, the method maintained recognition accuracy exceeding 80% and 84%, respectively for salt-and-pepper and speckle noise. This underscores its practical applicability in complex real-world environments as considered here.

Table 4.11: RECOGNITION ACCURACY (%) OF HRLTP-BLPP UNDER DIFFERENT SPECKLE NOISE VARIANCE LEVELS AT VARIOUS ILLUMINATION CONDITIONS

Noise Variance (σ^2)	**NIL**	**DIL-1**	**DIL-2**
0.05	99.365	97.885	95.644
0.10	98.257	96.272	94.136
0.15	97.709	95.338	93.435
0.20	96.910	94.075	91.149
0.25	92.823	90.471	88.195
0.30	90.903	88.356	86.168
0.35	88.142	86.111	84.595

4.6 Summary

This chapter introduced a set of local pattern encoding-based projection frameworks designed to enhance the robustness of feature extraction in visual cue recognition tasks under noisy and photometrically degraded environments. Starting with the LTrP descriptor, a direction-aware local texture encoder, the study explored how such encodings can be fused with spatial affinity in the bilateral LPP framework to build noise-resilient subspace projections.

To further improve descriptor granularity, histogram refinement strategies using LSP and BEM were incorporated, allowing better discrimination of image regions based on local statistical and structural variations. These refinements were then applied to classical descriptors like LBP and LTP, yielding HRLBP and HRLTP schemes, respectively. The HRLBP-BLPP method exhibited high classification accuracy under moderately challenging illumination conditions, while HRLTP-BLPP proved to be particularly effective under simultaneous low-light and noise interference.

Comprehensive experiments across various states of degradation, have demonstrated that the proposed hybrid schemes outperform standard projection and descriptor combinations, establishing their utility in real-world robotic vision systems where input distortions are commonplace. However, the development of even more resilient techniques remains an open challenge, especially for recognition under severely impaired lighting conditions and high-variance additive noise such as Gaussian noise, which tends to obscure subtle visual

features across the entire image. These advanced strategies are explored in the following chapters.

Chapter 5

Rough Entropy-based Fused Granular Feature Extraction Strategies in Two-dimensional LPP

5.1 Introduction

High-dimensional visual data originating from real-world human-robot interaction (HRI) scenarios often suffer from challenges such as illumination variability, sensor-induced distortions, and structural ambiguities [49]. Most commonly, dimensionality reduction (DR) methods are employed to simplify these complex data representations by mapping them onto lower-dimensional manifolds that retain essential information for downstream tasks like recognition and classification. Classical linear DR techniques, such as principal component analysis (PCA) [3], factor analysis (FA) [99], and linear discriminant analysis (LDA) [4], are widely utilized. Nonetheless, in situations where the data exhibits complex geometric structures, linear models often fail to uncover the true underlying relationships. To address this shortcoming, kernel-based variants such as kernel PCA (KPCA) [83] and kernel LDA (KLDA) [100] have been introduced to facilitate nonlinear feature learning. Alternatively, manifold learning-inspired techniques have shown effectiveness in extracting nonlinear low-dimensional embeddings of data. Examples include isometric feature mapping (Isomap) [5], locally linear embedding (LLE) [36], Laplacian eigenmaps (LE) [6], and local tangent space alignment (LTSA) [101]. However, most of these manifold-inspired methods lack an explicit mapping function, which restricts their ability to process unseen data (known as the out-of-sample problem). To circumvent this, linear approximations of manifold learning like neighborhood preserving projection (NPP) [38], neighborhood preserving embedding (NPE) [37], and locality

preserving projections (LPP) [7], have been proposed. LPP tackles the aforementioned challenge by linearly approximating Laplacian eigenmaps and allowing the projection of high-dimensional samples into a lower-dimensional ambient space. LPP preserves local neighborhood structures and has shown success in pattern and face recognition applications. Over time, several enhancements to LPP have emerged, such as discriminant LPP (DLPP) [40], orthogonal DLPP (ODLPP) [41], and supervised kernel LPP (SKLPP) [42], each aiming to improve discrimination or incorporate class information. Despite its merits, conventional LPP requires flattening two-dimensional data into one-dimensional vectors, which can obscure inherent spatial structures and cause a drop in performance, particularly in image-based datasets. This transformation also aggravates the small-sample-size problem, leading to optimization difficulties.

To mitigate these issues, 2D extensions such as 2DPCA [34] and 2DLDA [35] were proposed, enabling feature extraction directly from image matrices. Similarly, 2DLPP [13] preserves the Euclidean spatial structure of images while simplifying Laplacian matrix computation. Nevertheless, 2DLPP presents two limitations: (1) it relies exclusively on spatial similarity for constructing the kernel, which may not reflect true relationships in cases of intensity variations; and (2) high-dimensional real-world data often involve uncertainties and imprecise elements that impair local similarity computations. In particular, as discussed in the previous chapter, severely challenging illumination conditions can amplify such image uncertainties to a great extent.

To address these limitations, this work introduces a rough entropy-based granular fusion (REGF) framework, for extracting robust visual features within the framework of 2DLPP. This method incorporates additional feature-driven similarity, alongside spatial information. To tackle indiscernibility and uncertainty in the image space, the proposed framework leverages granular computing (GrC) [102], which encapsulates complex data traits as collections of semantically similar granules. The REGF model utilizes rough entropy (RE)-based granulation, combining the strengths of both homogeneous (crisp granulation, CG) [102] and non-homogeneous (quad-tree decomposition, QTD) [103] techniques to describe vague boundary regions in RGB images. The granular fusion strategy retains channel-wise color distinctions that grayscale conversions often discard. The combined spatial and feature-based approach improves Laplacian graph construction by incorporating both spatial and feature-driven similarity measures.

5.2 Traditional Approaches in Granular Computing

Granular computing (GrC) offers a robust theoretical framework for processing information by organizing it into granules i.e., clumps of data elements grouped based on similarity, functionality, or indistinguishability [102]. Over the years, traditional GrC techniques have been effectively applied in domains such as image processing, pattern recognition, and decision analysis [104], [105]. These classical approaches primarily rely on structured mechanisms like crisp granulation and hierarchical partitioning to model and interpret uncertain or imprecise information. Such strategies aim to simplify complex data by decomposing it into manageable subsets, allowing for more interpretable and computationally efficient analysis. This section outlines the foundational methodologies in granular computing, highlighting their operational principles, benefits, and inherent limitations in the context of visual data representation. Based on these, a novel granular fusion scheme is proposed here which can be suitably used as a robust feature descriptor inside the framework of 2DLPP [13].

5.2.1 Crisp Granulation (CG)

In the domain of granular computing (GrC), rough set theory is extensively employed to manage ambiguous data in images, particularly when it comes to challenges related to distinguishing objects [102]. According to this theory, a primary set is approximated using two definite sets called the lower and upper approximations [106]. Such rough set-based techniques have seen various applications spanning image analysis, pattern classification, and tracking of humans or objects [107], [104]. To address the uncertainty that exists between the background and foreground regions of an image, Pal *et al.* introduced a measure known as rough entropy (RE), which estimates the degree of ambiguity in segmentation [102]. Instead of treating individual pixels, the image is divided into a set of non-overlapping granules [108]. By maximizing this entropy, the discernibility between foreground and background is enhanced, leading to an effective threshold for segmentation.

Let $X \subseteq U$ represent the target set to be approximated using attribute set P. If X belongs to a single class, rough sets allow expressing X using equivalence classes defined by P. However, in general, X can contain information that cannot be distinguished based on the set of attributes

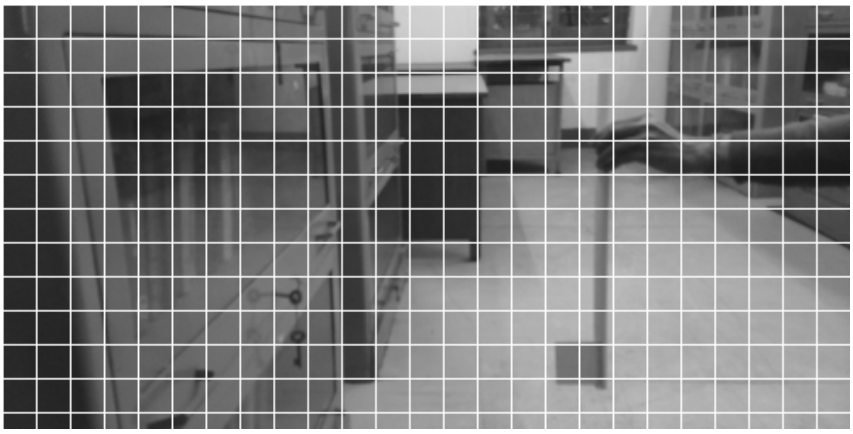
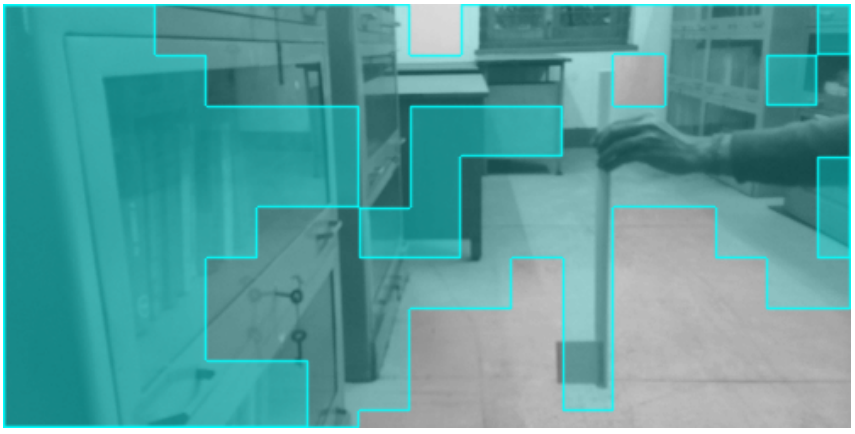


Figure 5.1: Demonstration of equal-sized granule formation using crisp granulation for a sample flag-stick visual cue.

(a)



(b)

Figure 5.2: Upper approximation (*lighter* tone) and lower approximation (*darker* tone) using crisp granulation for the (a) Foreground region (in *yellow*) and (b) Background region (in *cyan*) in a sample flag-stick image.

P. This challenge originates the idea of P-lower and P-upper approximations for describing the entire target set, as defined below.

$$\underline{P}(X) = \{x \mid [x]_P \subseteq X\} \tag{5.1}$$

$$\overline{P}(X) = \{x \mid [x]_P \cap X \neq \emptyset\} \tag{5.2}$$

The uncertainty in approximating X is quantified by the roughness index [106]:

$$R = 1 - \frac{|\underline{P}(X)|}{|\overline{P}(X)|} \tag{5.3}$$

If the set X is perfectly distinguishable by attributes in P, then $R = 0$. The more indistinct X becomes, the higher this measure rises. For some particular images, based on some predefined properties, such as the gray-level bands τ, the outer and inner approximations are denoted as $\overline{F}_\tau$, $\underline{F}_\tau$, $\overline{B}_\tau$, and $\underline{B}_\tau$, for the foreground (F) and background (B) regions, respectively. The associated rough entropy (RE) is then calculated as [102]:

$$\mathrm{RE}_\tau = -\frac{e}{2}\left[R_{F_\tau}\log_e(R_{F_\tau}) + R_{B_\tau}\log_e(R_{B_\tau})\right] \tag{5.4}$$

Maximizing this entropy helps determine the optimal threshold τ^* for foreground–background segmentation:

$$\tau^* = \arg\max_\tau \mathrm{RE}_\tau \tag{5.5}$$

The effectiveness of this threshold is also influenced by the size of the granules chosen. While smaller granules offer more accurate approximations, they also increase computational overhead. Further implementation details regarding threshold estimation can be found in [102]. Fig. 5.1 illustrates a CG-based equal-sized granule formation on a representative flag-stick marker image in grayscale. Typically, crisp granulation is carried out on the grayscale version of color images. In Fig. 5.2a and Fig. 5.2b, for a representative visual cue, CG-based upper and lower approximations for the foreground and background segments are demonstrated, respectively.

5.2.2 Quad-Tree Decomposition (QTD)

In the context of granular computing, granules may vary in both shape and size [109]. While CG creates uniform granules across the entire image, such homogeneity may not be ideal in practical scenarios, where local regions exhibit different levels of separability [103]. To accommodate

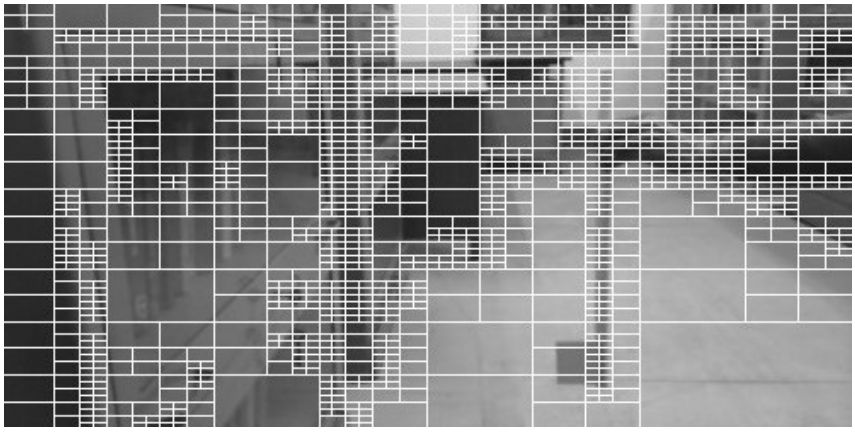


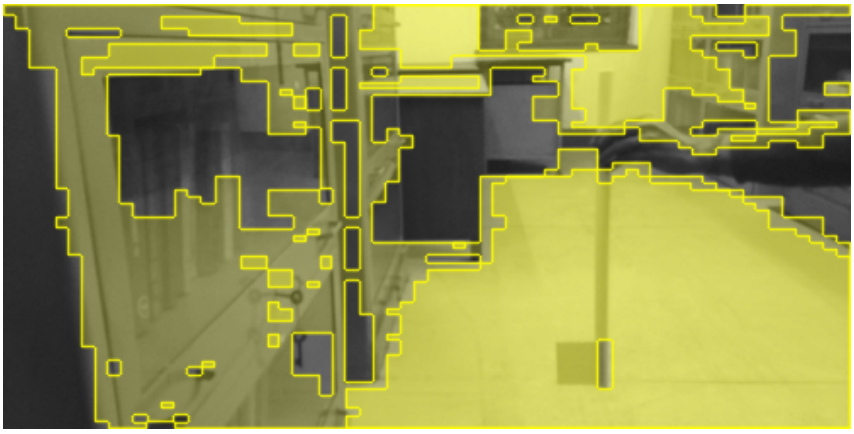
Figure 5.3: Demonstration of unequal-sized granule formation using quad-tree decomposition for a sample flag-stick visual cue.

this variability, Chakraborty *et al.* [103] proposed the use of quad-tree decomposition (QTD), which constructs granules of unequal size by evaluating both gray-level similarity and spatial proximity [110].

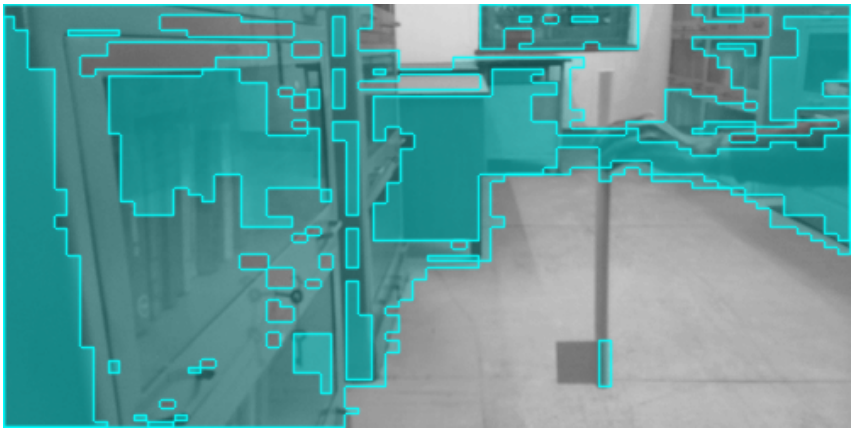
The process begins by selecting a window from the image, within which QTD is applied based on the gray-level differences of the contained pixels. The decomposition continues recursively, splitting the selected region into four smaller blocks until the intensity range (i.e., the difference between the highest and lowest gray values within a block) falls below a specified threshold. This recursive partitioning divides the image into quadrants, and the subdivision continues, based on the local gray-level uniformity. Let q_1 and q_3 represent the first (25^{th} percentile) and third (75^{th} percentile) quartiles of the gray-level histogram of the image. Then, the quad-tree decomposition threshold is computed as:

$$\text{QT}_{Th} = \frac{q_3 - q_1}{2} \tag{5.6}$$

The granule formation, driven by repeated division into quadrants and guided by this threshold, results in a set of variably sized blocks. Fig. 5.3 illustrates the unequal-sized granules obtained via QTD for a sample flag-stick image. Using this approach, foreground and background approximations for the same sample image are derived. These approximations, created through QTD, are presented in Fig. 5.4a and Fig. 5.4b, where the foreground regions are shown in yellow and background regions in cyan, with upper and lower approximations marked using lighter and darker shades, respectively.



(a)



(b)

Figure 5.4: Upper approximation (*lighter* tone) and lower approximation (*darker* tone) using quad-tree decomposition for the (a) Foreground region (in *yellow*) and (b) Background region (in *cyan*) in a sample flag-stick image.

5.3 Proposed Granular Fusion Scheme REGF

As highlighted earlier, segmentation approaches based on granulation prove to be quite effective in describing ambiguous image regions, especially for boundary extraction. However, irrespective of whether granules are of fixed or variable size, it is important that they encapsulate meaningful groupings of similar content. In the case of homogeneous crisp granulation (CG), if larger granules are used, they can potentially fail to capture informative segments, while excessively small granules might introduce irrelevant details and increase computational complexity. Furthermore, the general guideline provided in [102] for selecting an appropriate granule size often struggles with images that possess non-normal or multimodal distributions;

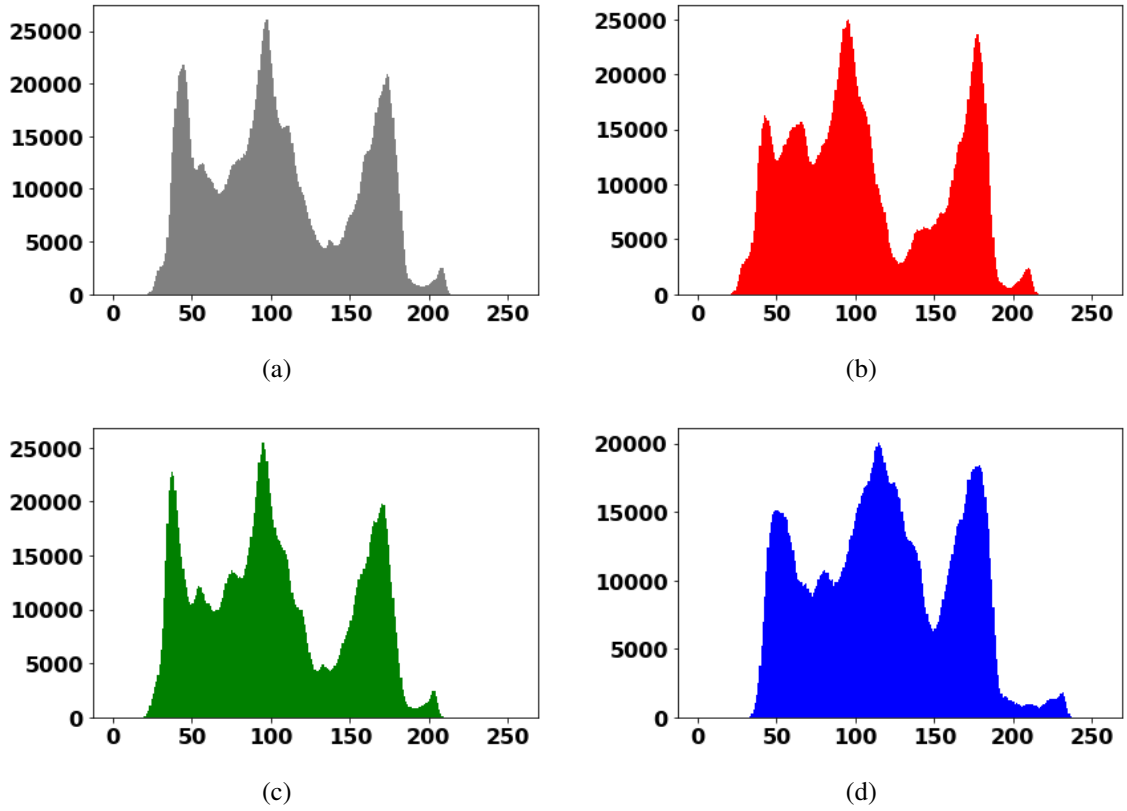


Figure 5.5: Histogram distributions exhibiting multimodal characteristics for a sample flag-stick image in (a) Grayscale, (b) Red, (c) Green, and (d) Blue channels.

such as the real-world flag-stick images acquired in this work, which are illustrated in Fig. 5.5.

This brings forth a key challenge in the granulation scheme CG. The determination of optimal granule sizes in CG becomes less straightforward when dealing with such complex histogram profiles. On the contrary, granulation through quad-tree decomposition (QTD), being non-uniform, adapts more naturally to foreground-background boundary estimation in images [103]. Here, the size and depth of granules are governed by a threshold which is computed based on the gray-level distribution's empirical quartiles, as explained earlier in Equation (5.6). However, experimental observations suggest that such empirical thresholds are not sufficiently robust when applied to multimodal flag-stick images, which often exhibit irregular clustering behavior and locally uneven intensity distributions. This can cause important features to be missed in certain regions during the decomposition. Moreover, both CG and QTD are typically applied to grayscale images, where pixel intensity serves as the key feature for granulation. However, converting RGB images into grayscale often leads to a significant loss of chromatic information, particularly in areas with distinct color contrasts but similar grayscale intensities

[32]. This can introduce further inaccuracies during granule formation.

To overcome these limitations, a novel fusion strategy is proposed here, that minimizes the uncertainty arising from vague boundaries; whether due to spatial inconsistencies or color ambiguities. This is achieved by integrating the rough set approximations generated separately from homogeneous (i.e., CG) [102] and non-homogeneous (i.e., QTD) [103] granulation approaches. Furthermore, to better preserve both local and contextual information across the image, the proposed method avoids converting the image to grayscale. Instead, granulation is performed individually on each of the RGB color channels, which often exhibit different gray-level distributions. To this end, rough set-based granules for each color channel are separately computed first, and then aggregated together to build the complete color image approximation. Specifically, for an image A, two sets of spatio-color granules—$\{\mathbb{UF}(\cdot)\}$ and $\{\mathbb{LF}(\cdot)\}$ are obtained, representing upper and lower approximations of the foreground, respectively. These are defined as:

$$\begin{aligned}\{\mathbb{UF}_{(k)}\} &= \{\mathbb{UF}^{R}_{(k)}\} \cup \{\mathbb{UF}^{G}_{(k)}\} \cup \{\mathbb{UF}^{B}_{(k)}\} \\ \{\mathbb{LF}_{(k)}\} &= \{\mathbb{LF}^{R}_{(k)}\} \cup \{\mathbb{LF}^{G}_{(k)}\} \cup \{\mathbb{LF}^{B}_{(k)}\}, \quad k \in \{\mathrm{CG}, \mathrm{QT}\}\end{aligned} \tag{5.7}$$

Similarly, for background, the corresponding upper and lower approximations are defined as:

$$\begin{aligned}\{\mathbb{UB}_{(k)}\} &= \{\mathbb{UB}^{R}_{(k)}\} \cup \{\mathbb{UB}^{G}_{(k)}\} \cup \{\mathbb{UB}^{B}_{(k)}\} \\ \{\mathbb{LB}_{(k)}\} &= \{\mathbb{LB}^{R}_{(k)}\} \cup \{\mathbb{LB}^{G}_{(k)}\} \cup \{\mathbb{LB}^{B}_{(k)}\}, \quad k \in \{\mathrm{CG}, \mathrm{QT}\}\end{aligned} \tag{5.8}$$

In the equations above, $i \in \{R, G, B\}$ denotes the red, green, and blue channels respectively, and the sets $\{\mathbb{UF}^{i}(\cdot)\}$, $\{\mathbb{LF}^{i}(\cdot)\}$, $\{\mathbb{UB}^{i}(\cdot)\}$, and $\{\mathbb{LB}^{i}(\cdot)\}$ represent the granules corresponding to the rough approximations (foreground and background) over each channel. Thereafter, the granules obtained via CG and QTD are combined using an intersection operation to produce the fused approximations: $\overline{f\mathbb{F}}$, $\underline{f\mathbb{F}}$, $\overline{f\mathbb{B}}$, and $\underline{f\mathbb{B}}$, which represent the fused upper and lower approximations of foreground and background, respectively:

$$\begin{aligned}\overline{f\mathbb{F}} &= \{\mathbb{UF}_{(\mathrm{CG})}\} \cap \{\mathbb{UF}_{(\mathrm{QT})}\} \\ \underline{f\mathbb{F}} &= \{\mathbb{LF}_{(\mathrm{CG})}\} \cap \{\mathbb{LF}_{(\mathrm{QT})}\} \\ \overline{f\mathbb{B}} &= \{\mathbb{UB}_{(\mathrm{CG})}\} \cap \{\mathbb{UB}_{(\mathrm{QT})}\} \\ \underline{f\mathbb{B}} &= \{\mathbb{LB}_{(\mathrm{CG})}\} \cap \{\mathbb{LB}_{(\mathrm{QT})}\}\end{aligned} \tag{5.9}$$

This fusion essentially retains only those pixels that are commonly included in both CG and QTD approximations, thereby integrating the benefits of both granulation strategies. The

same logic applies to all four fused regions mentioned above. Next, the roughness measures corresponding to the fused foreground and background models, denoted by $R_{f\mathbb{F}}$ and $R_{f\mathbb{B}}$, respectively, are obtained as follows:

$$R_{f\mathbb{F}} = 1 - \frac{|\underline{f\mathbb{F}}|}{|\overline{f\mathbb{F}}|} = \frac{|\overline{f\mathbb{F}}| - |\underline{f\mathbb{F}}|}{|\overline{f\mathbb{F}}|} \tag{5.10}$$

$$R_{f\mathbb{B}} = 1 - \frac{|\underline{f\mathbb{B}}|}{|\overline{f\mathbb{B}}|} = \frac{|\overline{f\mathbb{B}}| - |\underline{f\mathbb{B}}|}{|\overline{f\mathbb{B}}|} \tag{5.11}$$

Here, $|\cdot|$ represents the cardinality of a given set. Using these roughness values, the *granular fusion-based rough entropy* (GFRE) is calculated using a formulation similar to that of traditional RE:

$$\text{GFRE} = -\frac{e}{2}\left[R_{f\mathbb{F}}\log_e(R_{f\mathbb{F}}) + R_{f\mathbb{B}}\log_e(R_{f\mathbb{B}})\right] \tag{5.12}$$

The qualitative assessment and detailed effectiveness of the proposed granular fusion approach are illustrated and further validated in Section 5.5 .

5.4 Proposed Dimensionality Reduction Framework REGF-2DLPP

While traditional granular computing techniques have demonstrated effectiveness in reducing visual ambiguity and managing uncertainty, their standalone application often falls short when integrated with manifold-inspired learning frameworks. To bridge this gap, the REGF-2DLPP model was introduced as a hybrid dimensionality reduction approach that strengthens the two-dimensional locality preserving projection (2DLPP) through granulated feature augmentation. This method incorporates a rough entropy-based granulation scheme, termed REGF, which leverages both crisp granulation (CG) [102] and quad-tree decomposition (QTD) [103] to improve the robustness of feature encoding. By fusing granule-derived feature similarity with spatial neighborhood information, REGF-2DLPP constructs an enriched Laplacian graph that preserves local structure while also encoding appearance-level variations. This section presents the formulation and working mechanism of the REGF-2DLPP-based projection model, emphasizing its role in improving projection reliability under visually uncertain and photometrically degenerated conditions. The entire REGF-2DLPP scheme for

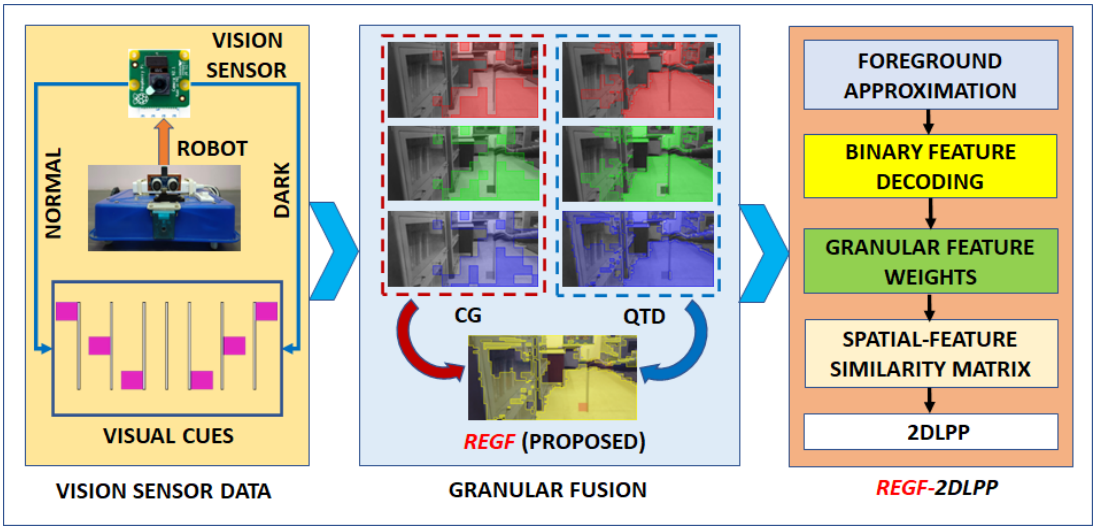


Figure 5.6: Block diagram of REGF-2DLPP: A robust visual cue recognition-based robot navigation guidance scheme.

visual cue recognition is presented in Fig. 5.6.

5.4.1 Granulated Feature Decoding

Following the granular fusion process, a feature decoding method is essential to derive a representative feature histogram from the input image. For this purpose, the fused upper approximation of the foreground, denoted as $\overline{f\mathbb{F}}$, is utilized to extract the relevant foreground content. Since the upper granular approximation maximally approximates the region of the foreground, it provides a comprehensive representation of all related granules, making it highly suitable for use in feature extraction. Additionally, as $\overline{f\mathbb{F}}$ consolidates the strengths of both homogeneous (CG) and non-homogeneous (QTD) granulation, it serves as a robust foundation for constructing image descriptors. To generate a binary map from the granulated image, pixels corresponding to the granules in $\overline{f\mathbb{F}}$ are set to 1, while all others are set to 0. Next, to extract local patterns from this binarized image as feature values, a sliding neighborhood of size 3×3 is applied across the entire image. Inspired by the local binary pattern (LBP) method, for any center pixel p_c within a neighborhood, the adjacent pixel immediately to its right is considered the first neighbor. The pixel located directly above this first neighbor becomes the second neighbor, and the remaining six are indexed thereafter in a counterclockwise sequence [30]. This configuration produces an 8-bit binary string for each central pixel within the image, as shown in Fig. 5.7. The corresponding decoded feature for the neighborhood centered at p_c is

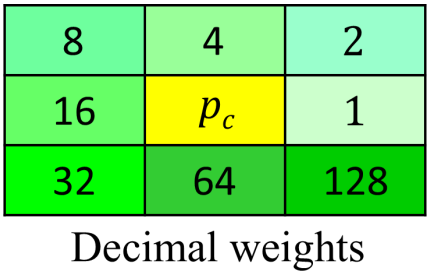


Figure 5.7: Positional binary weights assigned to the 3×3 neighborhood centered at pixel p_c for feature decoding.

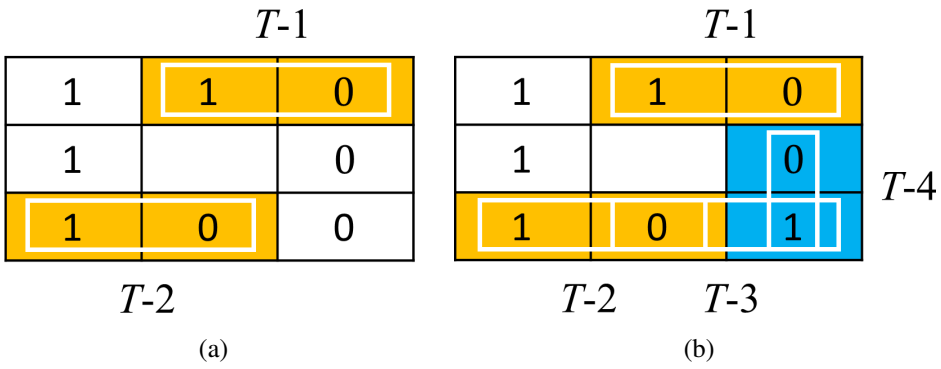


Figure 5.8: Example binary patterns for a 3×3 local window: (a) A uniform 8-bit binary string, and (b) A non-uniform 8-bit binary string.

computed using:

$$\phi_{p_c} = \sum_{n=1}^{8} 2^{n-1} \cdot p_n \tag{5.13}$$

Here, p_n refers to the binary value of the n-th neighbor surrounding p_c in the 3×3 neighborhood. The uniformity of these binary strings is important for preserving critical local structures such as corners and edges [30]. A binary string is labeled non-uniform if it contains more than two transitions between 0 and 1 (or vice versa). Fig. 5.8a and 5.8b present examples of uniform and non-uniform binary patterns, respectively, where transitions are marked as $T-1$, $T-2$, and so on. In total, there are $2^8 = 256$ possible binary strings that can be generated. Among these, only 58 patterns are limited to two transitions, making them uniform patterns as per definition. Owing to that, the output representation is reduced to a histogram with 59 bins only; one for each of the 58 uniform patterns and an additional bin that captures all non-uniform patterns. As a result, the final feature vector for any segmented image contains 59 elements. The steps for this entire feature decoding process are summarized in Algorithm 5.1.

Algorithm 5.1 Granulated Binary Feature Decoding

Input: Input image $\mathbf{A}$, fused upper foreground granules $\overline{f\mathbb{F}}$, local neighborhood size $n \times n$

Output: Normalized binary feature vector $\boldsymbol{\phi}$

STEP-1 Identify all pixel positions $\mathscr{L}$ in $\mathbf{A}$ corresponding to the granules in $\overline{f\mathbb{F}}$.

STEP-2 Binarize the image by assigning value 1 to all locations in $\mathscr{L}$ and 0 elsewhere.

STEP-3 For each center pixel p_c in the binarized image:

- Select the $n \times n$ local window centered at p_c.
- Generate an 8-bit binary string from the neighboring pixels.
- Compute the decimal equivalent using:

$$\phi_{p_c} = \sum_{n=1}^{8} 2^{n-1} \cdot p_n$$

STEP-4 Build a histogram with $(n^2-1)(n^2-2)+3$ number of bins.

STEP-5 Normalize the histogram to obtain the final binary feature vector $\boldsymbol{\phi}$.

STEP-6 Return $\boldsymbol{\phi}$.

5.4.2 Granulated Feature Induced 2DLPP

Let us consider the training dataset $\mathbf{A} \in \mathbb{R}^{pN \times q}$, which contains N flag-stick visual cue images denoted by $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_N$, where each image $\mathbf{A}_i \in \mathbb{R}^{p \times q}$. Due to the high dimensionality of the data, it becomes difficult to achieve effective recognition with latency using conventional methods. To manage this, a linear mapping is applied to project each $p \times q$ image onto a lower p-dimensional space [13]. Let $\mathbf{w} \in \mathbb{R}^{q \times 1}$ be a unitary column vector which facilitates the projection. The 2DLPP transformation for the i-th sample is defined as:

$$\mathbf{z}_i = \mathbf{A}_i \mathbf{w}, \quad i = 1, 2, \ldots, N \tag{5.14}$$

Here, $\mathbf{z}_i$ represents the p-dimensional projected feature vector of the image $\mathbf{A}_i$. 2DLPP is a two-dimensional counterpart of traditional LPP, which extracts spatial structure directly from image matrices rather than their flattened 1D versions. The similarity graph is constructed based on spatial proximity among the image samples. Two nodes i and j are considered connected if their respective images $\mathbf{A}_i$ and $\mathbf{A}_j$ are close in spatial space and thus are called neighbors. Conventionally, this similarity is determined using the Euclidean distance between the vectorized forms of $\mathbf{A}_i$ and $\mathbf{A}_j$ in $\mathbb{R}^{pq}$. However, in 2DLPP, the Frobenius norm between the

original matrices is used instead:

$$\delta = \|\mathbf{A}_i - \mathbf{A}_j\|_F^2 \tag{5.15}$$

Based on this metric, the k-nearest neighbors of each sample are selected to establish the similarity space. Typically, a heat kernel function is employed to populate the similarity matrix:

$$r_{ij} = \begin{cases} \exp\left(-\dfrac{\delta}{t_r}\right), & \text{if } \mathbf{A}_j \in \mathcal{N}_k(\mathbf{A}_i) \\ 0, & \text{otherwise} \end{cases} \tag{5.16}$$

Here, $\mathcal{N}_k(\mathbf{A}_i)$ denotes the k nearest neighbors of $\mathbf{A}_i$, and t_r is a user-defined kernel parameter. This matrix $\mathbf{R} = [r_{ij}]$ encodes only the spatial closeness between samples. However, such spatial metrics are susceptible to noise, distortions, and geometric variations, which can affect the reliability of the learned similarity structure [11]. To mitigate this, feature-based similarity weights t_{ij} are computed in the feature domain ψ using the granulated binary features obtained earlier (Equation (5.13)):

$$t_{ij} = \begin{cases} \exp\left(-\dfrac{\delta(\psi(\mathbf{A}_i) \sim \psi(\mathbf{A}_j))}{t_f}\right), & \text{if } \mathbf{A}_j \in \mathcal{N}_k(\mathbf{A}_i) \\ 0, & \text{otherwise} \end{cases} \tag{5.17}$$

In this equation, t_f is a scaling factor that regulates the sensitivity of the kernel to variations in feature intensity. The final similarity matrix $\mathbf{S}$ is then obtained by element-wise multiplication of the spatial and feature similarity matrices:

$$s_{ij} = r_{ij} \cdot t_{ij} \tag{5.18}$$

With the complete similarity matrix $\mathbf{S} = [s_{ij}]$, the generalized eigenvalue problem for 2DLPP is formulated as [13]:

$$\mathbf{A}^\top (\mathbf{L} \otimes \mathbf{I}) \mathbf{A}\mathbf{w} = \lambda \mathbf{A}^\top (\mathbf{\Delta} \otimes \mathbf{I}) \mathbf{A}\mathbf{w} \tag{5.19}$$

Here, $\mathbf{\Delta}$ is a diagonal matrix with entries $\Delta_{ii} = \sum_j s_{ij}$, and the Laplacian matrix is defined as $\mathbf{L} = \mathbf{\Delta} - \mathbf{S}$. The training data matrix $\mathbf{A}$ stacks all N image matrices as $\mathbf{A} = [\mathbf{A}_1; \mathbf{A}_2; \ldots, \mathbf{A}_N] \in \mathbb{R}^{pN \times q}$. $\mathbf{I}$ denotes the $p \times p$ identity matrix and $\otimes$ represents the Kronecker product. Solving this eigenvalue problem yields a set of projection vectors $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_q$ corresponding to eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_q$. Among these, d number of eigenvectors associated with the smallest non-zero eigenvalues are selected to form the projection matrix:

$$\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_d] \tag{5.20}$$

Using $\mathbf{W}$, each input image $\mathbf{A}_i$ is projected into a lower-dimensional subspace as:

$$\mathbf{Z}_i = \mathbf{A}_i\mathbf{W}, \quad i = 1,2,\ldots,N \tag{5.21}$$

Once the transformation is applied to both training and test data, the resulting low-dimensional representations $\mathbf{Z}_i$ can be used as input to any suitable classifier for recognition tasks.

5.5 Experimental Results and Discussion

To evaluate the proposed REGF approach, a step-by-step illustration is presented using a representative flag-stick image belonging to the bottom-left (BL) class. **Stage 1** starts by applying both CG and QTD to each individual color channel (R, G, and B). This yields upper foreground approximations for the R, G, and B channels, denoted as $\{\mathbb{UF}^R_{(\mathrm{CG})}, \mathbb{UF}^G_{(\mathrm{CG})}, \mathbb{UF}^B_{(\mathrm{CG})}\}$ and $\{\mathbb{UF}^R_{(\mathrm{QT})}, \mathbb{UF}^G_{(\mathrm{QT})}, \mathbb{UF}^B_{(\mathrm{QT})}\}$, Following CG and QTD, respectively. Since the histogram distribution varies across the three color channels (as shown earlier in Fig. 5.5), the mean of the minimum base widths from each channel is calculated first. Thereafter, following the guideline in [102], the CG granule size is set to half of that mean base width to achieve a balanced crisp granulation strategy. This approach takes into account the channel-wise multimodal distribution and provides a more data-driven selection of granule size. Foreground and background approximations derived from both CG and QTD, across all three RGB channels, are visualized in the *first three rows* of Fig. 5.9. Next, in **Stage 2**, unified RGB granules are computed using Equations (5.7) and (5.8), combining the approximations for all color channels separately for CG and QTD. These results are shown in the *fourth row* of Fig. 5.9. **Stage 3** is the final step involved here, which performs granular fusion as per Equation (5.9), integrating both CG- and QTD-based information. The fused results are displayed in the *fifth row* of Fig. 5.9. In all these visualizations, lighter shades indicate upper approximations, while darker shades correspond to lower approximations.

It can be observed that the fused granules preserve the strengths of both CG and QTD. In regions with clearer structure, large granules from CG dominate, whereas, in more uncertain areas, smaller granules from QTD provide finer detail. This fusion reduces ambiguity along foreground-background boundaries and enhances rough entropy (RE) in the image. Fig. 5.10 presents the average RE values computed over the entire flag-stick dataset, comparing four granulation schemes: CG, QTD, grayscale fusion, and RGB-based fusion. The results show

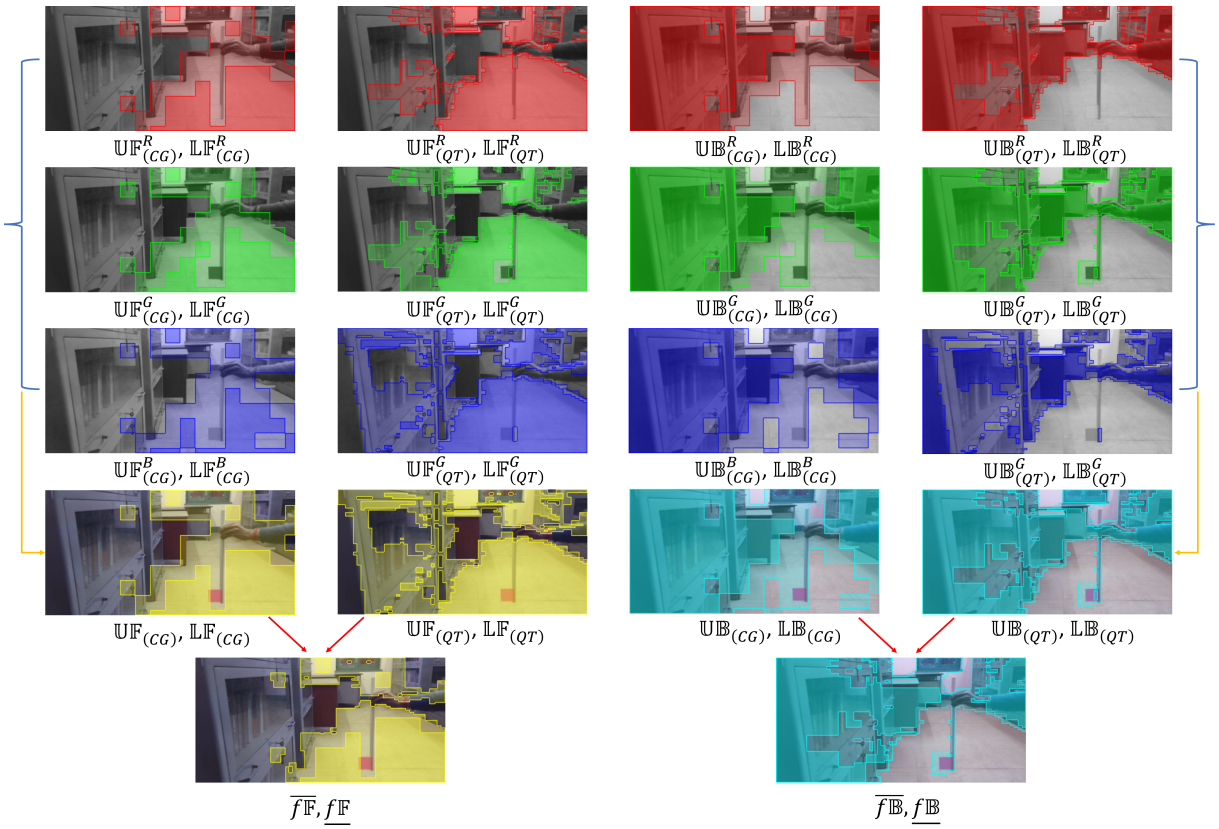


Figure 5.9: Stepwise qualitative representation of the REGF method, where foreground granules are depicted in *yellow* and background granules in *cyan*. *Lighter* and *darker* shades indicate the upper and lower approximations, respectively.

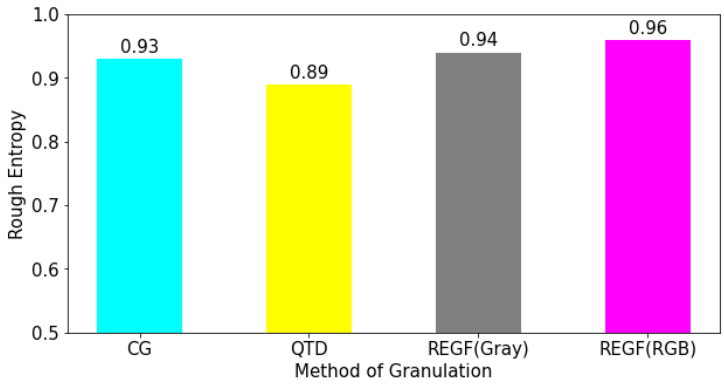


Figure 5.10: Comparison of average rough entropy (RE) values obtained from the flag-stick image dataset across the four granulation techniques evaluated.

that the proposed RGB-based fusion method achieves the highest RE values, indicating better granule segmentation performance.

To further evaluate effectiveness of the REGF-2DLPP framework, classification experiments were conducted using the flag-stick visual cue dataset introduced in Section

Table 5.1: CLASSIFICATION ACCURACY (%) OF VARIOUS DR TECHNIQUES UNDER DIFFERENT ILLUMINATION CONDITIONS

DR Technique	NIL	DIL-1	DIL-2	DIL-3	DIL-4
PCA	99.048	70.253	46.349	39.885	32.012
LDA	99.365	81.296	73.333	65.320	49.678
LPP	99.683	92.362	84.127	70.773	55.710
BLPP	99.965	96.329	94.603	75.221	62.394
2DLPP	99.968	98.223	97.114	83.623	73.521
CG-2DLPP	99.972	98.983	98.189	86.718	77.843
QTD-2DLPP	99.973	99.230	98.777	88.361	79.950
REGF-2DLPP	**99.980**	**99.436**	**99.167**	**89.480**	**81.565**

Table 5.2: SPEARMAN'S RANK CORRELATION COEFFICIENTS (SRCC) FOR DIFFERENT DR TECHNIQUES UNDER VARYING ILLUMINATION CONDITIONS

DR Technique	DIL-1	DIL-2	DIL-3	DIL-4
Raw Images	0.9602	0.9577	0.7905	0.7296
LPP	0.9649	0.9583	0.8157	0.7312
BLPP	0.9696	0.9695	0.8460	0.7748
2DLPP	0.9780	0.9752	0.8918	0.8302
CG-2DLPP	0.9854	0.9765	0.9135	0.8662
QTD-2DLPP	0.9872	0.9783	0.9244	0.8815
REGF-2DLPP	**0.9885**	**0.9811**	**0.9319**	**0.9081**

2.5 in Chapter 2. First, the method is evaluated under normal lighting conditions (NIL), where the dataset is randomly partitioned into 70% training and 30% testing samples. Then the REGF-2DLPP projection scheme is employed, as discussed in Section 5.4.2 , to extract features from both sets, followed by classification using a support vector machine (SVM) classifier. Under NIL, the proposed model achieves over 99% recognition accuracy, confirming its effectiveness in practical scenarios involving visual marker-based robot guidance. To

Table 5.3: COMPARISON OF REGF AND OTHER FEATURE DESCRIPTORS WITHIN THE 2DLPP FRAMEWORK UNDER VARYING ILLUMINATION CONDITIONS

Feature Descriptor	**NIL**	**DIL-1**	**DIL-2**	**DIL-3**	**DIL-4**
HOG	99.968	98.345	97.308	83.827	73.934
Gabor	99.968	98.491	97.566	83.960	74.251
LBP	99.968	98.586	97.776	84.256	75.778
FREAK	99.970	98.871	97.932	84.738	76.346
LTrP	99.972	98.985	98.317	86.820	78.140
A-KAZE	99.972	99.046	98.464	87.102	78.728
CG	99.972	98.983	98.189	86.718	77.843
QTD	99.973	99.230	98.777	88.361	79.950
REGF (Proposed)	**99.980**	**99.436**	**99.167**	**89.480**	**81.565**

assess robustness in more challenging environments, the model is tested under moderately to extremely reduced lighting levels (DIL-1 to DIL-4). Table 5.1 reports recognition accuracy for REGF-2DLPP and several comparative dimensionality reduction (DR) techniques across different lighting conditions. As the illumination decreases, most baseline methods exhibit significant performance drops. However, REGF-2DLPP maintains a strong performance even in poorly lit environments. To quantify how the projection methods handle visual degradation in low-dimensional subspaces, Spearman's rank correlation coefficient (SRCC) analysis [111] is performed in both the image and projected feature spaces. The results are presented in Table 5.2. The SRCC values in the raw image space decline substantially as lighting deteriorates from DIL-1 to DIL-4. By contrast, projecting data to the low-dimensional subspace enhances correlation, revealing the benefit of manifold-inspired DR techniques in preserving intrinsic structure. Among all tested methods, REGF-2DLPP demonstrates the highest SRCC in the DIL-4 scenario, reflecting its superior resilience to illumination-induced distortion.

Additionally, the REGF-2DLPP model is also benchmarked against various feature extraction strategies incorporated within the 2DLPP framework. Table 5.3 presents a comparative study including descriptors like HOG [47], Gabor features [48], LBP [30], FREAK [112], LTrP [91], and A-KAZE [113]. Among all approaches, the proposed REGF technique yields the best performance, significantly outperforming not only traditional descriptors but also the individual CG and QTD methods when used in isolation.

5.6 Proposed Dimensionality Reduction Framework REGF-c-2DLPP

Although the REGF-2DLPP model has demonstrated notable improvement in preserving local structure through the integration of granulated feature similarity, its reliance on Euclidean distance in neighborhood construction makes it susceptible to magnitude variations and photometric inconsistencies. In particular, when the dataset encounters appearance-based variations e.g., object shapes and colors, these issues tend to occur more often. In response to this observation, the REGF-c-2DLPP framework introduces a cosine similarity-augmented extension that strengthens the construction of the Laplacian graph by emphasizing directional coherence between image matrices and their associated features. By incorporating cosine-based similarity measures in both spatial and granulated feature domains, the proposed model augments the discriminative capability of subspace projections under complex real-world conditions, such as illumination variability and appearance diversity. This section introduces the theoretical formulation and operational pipeline of REGF-c-2DLPP, highlighting its benefits in addressing high-dimensional ambiguity and structure-preserving projection challenges, especially for variations in the object appearance.

5.6.1 Motivation

The previously developed REGF-2DLPP scheme demonstrated promising results in visual cue recognition under photometric variations, as discussed in the previous sections. The original framework primarily focused on fixed-shape, fixed-color flag-stick cues, relying solely on Euclidean similarity in the spatial domain for constructing the projection kernel. However, Euclidean distance becomes increasingly unreliable in high-dimensional spaces, particularly when the data are affected by noise, illumination inconsistencies, or color distortions. In this regard, the variability in color and shape of the visual markers are studied in this section, which are also critical aspects in real-world human-robot collaborative tasks.

Considering the above factors, an improved approach REGF-c-2DLPP is developed here. This enhanced model integrates a cosine similarity-based distance metric into the 2DLPP formulation to improve neighborhood encoding in both the spatial and feature domains. Unlike traditional Euclidean distance, cosine similarity captures directional alignment between data

matrices, offering improved robustness against magnitude variations and outliers. To handle diverse color and shape patterns, the previous granular computing-based fusion strategy is employed, which extends to RGB channels individually. This allows the framework to effectively extract and preserve local structure from multi-color, multi-shape flag-stick cues. As a result, REGF-c-2DLPP builds upon the strengths of its predecessor while incorporating a new cosine similarity-based approach, thereby offering a more generalizable and robust solution for visual cue recognition in unconstrained, real-world HRC environments.

5.6.2 Projection Scheme of REGF-c-2DLPP

To extract the granulated features from an image and utilize them for subsequent visual cue recognition, the upper-approximated foreground region, denoted as $\overline{f\mathbb{F}}$, is considered to contain the most informative granule structures. This region, $\overline{f\mathbb{F}}$, encompasses all essential foreground granules, exploiting the combined advantages of both homogeneous and heterogeneous granulation, making it ideal for feature representation. A dedicated feature decoding strategy, previously detailed in Algorithm 5.1, is employed to transform the thresholded granule set $\overline{f\mathbb{F}}$ into usable feature values. To project these features into a reduced-dimensional space, the two-dimensional locality preserving projections (2DLPP) technique is used [13].

Let the training dataset comprise image matrices $\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_p$, where each sample $\mathcal{I}_i \in \mathbb{R}^{m \times n}$ represents a distinct flag-stick visual cue and P is the total number of samples. In standard LPP, spatial similarity between samples is calculated using the Frobenius norm, i.e., $\|\mathcal{I}_i - \mathcal{I}_j\|_F^2$. However, due to the limitations of Euclidean distance in high-dimensional feature spaces [11], this work proposes the use of cosine similarity for capturing local neighborhood information more robustly. Cosine similarity focuses on the directional alignment of the feature vectors rather than their magnitudes, making it less sensitive to intensity scaling and noise. This approach reduces the impact of inflated magnitude differences between image matrices, especially under high-dimensional or noisy conditions. Algorithm 5.2 outlines the procedure for computing cosine similarity between two image matrices. The resulting spatial similarity is denoted as $(d_{ij})_{\text{sp}}$.

Let $\xi(\mathcal{I}_i)$ represent the granulated feature vector extracted from the image $\mathcal{I}_i$ using the REGF scheme described in Section 5.3. The corresponding feature-based similarity is then

Algorithm 5.2 Cosine Similarity Measure for Image Matrices

Input: High-dimensional image matrices $\mathcal{I}_i$ and $\mathcal{I}_j$

Output: Cosine distance d_{ij}

STEP-1 Normalize both matrices using their spectral norms:

$$\mathcal{I}_i \leftarrow \mathcal{I}_i / \|\mathcal{I}_i\|_2, \quad \mathcal{I}_j \leftarrow \mathcal{I}_j / \|\mathcal{I}_j\|_2$$

STEP-2 Compute the dot product:

$$\text{dot}_{ij} = \mathcal{I}_i \cdot \mathcal{I}_j^{\top}$$

STEP-3 Evaluate row-wise cosine similarity:

$$(d_{ij})_{\text{all}} = \frac{\text{dot}_{ij}}{\|\mathcal{I}_i\|_2 \cdot \|\mathcal{I}_j\|_2}$$

STEP-4 Compute the average similarity:

$$d_{ij} = \text{mean}\left(\text{mean}\left((d_{ij})_{\text{all}}\right)\right)$$

STEP-5 Return d_{ij}.

computed using cosine similarity as:

$$(d_{ij})_{\text{ft}} = \frac{\xi(\mathcal{I}_i)^T \cdot \xi(\mathcal{I}_j)}{\|\xi(\mathcal{I}_i)\|_2 \cdot \|\xi(\mathcal{I}_j)\|_2} \tag{5.22}$$

Finally, both spatial and feature domain similarities are fused to form a joint similarity kernel:

$$(d_{ij})_{\text{tot}} = (d_{ij})_{\text{sp}} \times (d_{ij})_{\text{ft}} \tag{5.23}$$

The term $(d_{ij})_{\text{sp}}$ serves as the input kernel for the 2DLPP algorithm, as elaborated in [13]. After applying the REGF-c-2DLPP transformation, the feature data is projected into a compact subspace, which is then fed into a classifier to compute recognition accuracy.

5.6.3 Visual Cue Dataset with Varying Colors and Shapes

The experimental dataset used in this study, has been acquired at the Instrumentation and Cyber Physical System Laboratory, Department of Electrical Engineering, Jadavpur University, India. A custom-designed, cost-efficient two-wheel mobile robotic platform has been employed for data collection, as previously discussed in Section 2.4 of Chapter 2. A Raspberry Pi (Model B+) as the primary computational unit, has been embedded in the robot, while a Raspberry

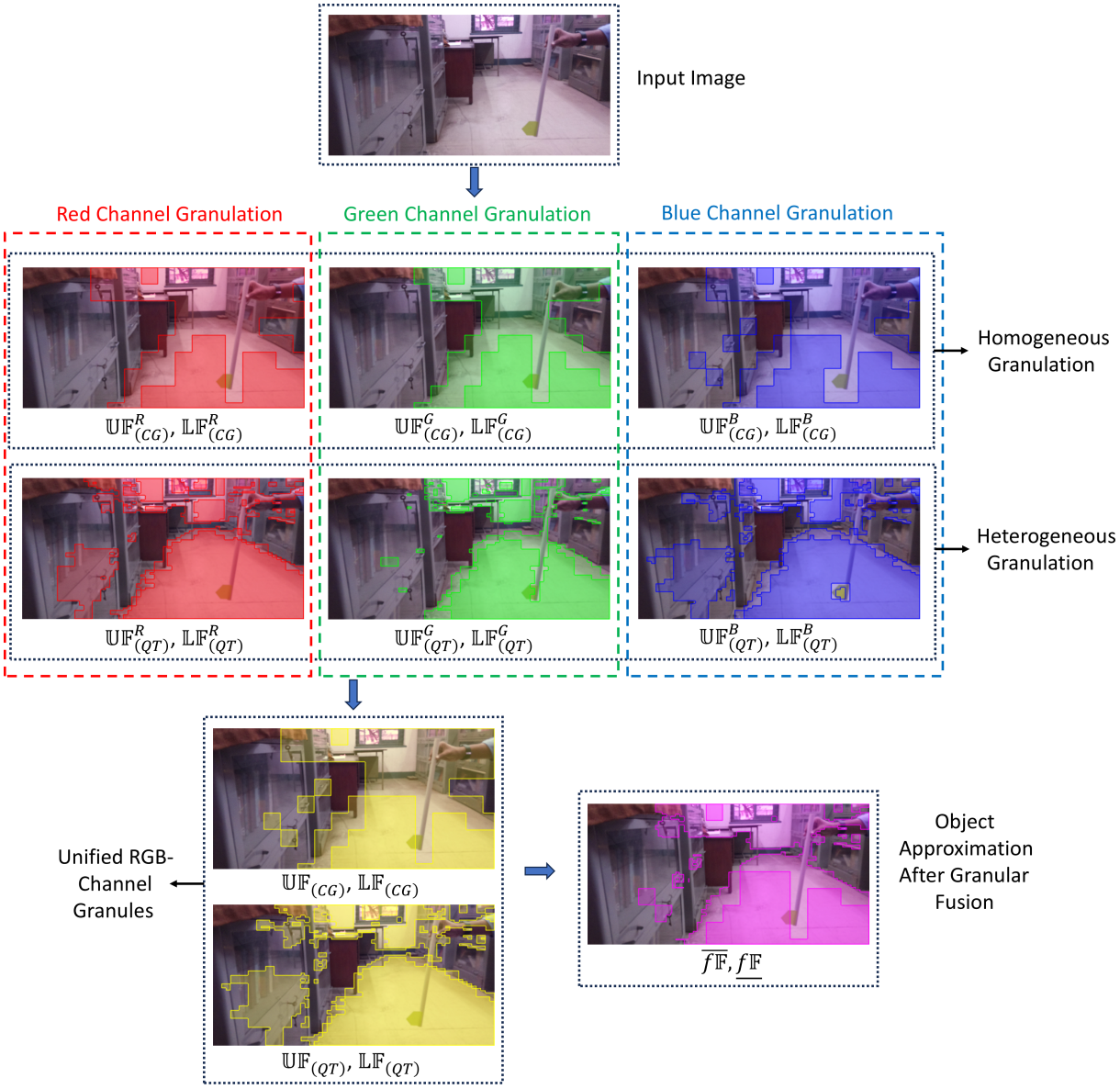


Figure 5.11: Illustrative flowchart of the REGF methodology applied to a sample visual cue featuring a lemon-lime colored flag. The *darker* and *lighter* shades represent the lower and upper object approximations, respectively.

Pi Camera Module v2, equipped with an 8-Megapixel Sony IMX219 image sensor, has been mounted on the robot for real-time visual data acquisition.

The visual cues were designed as hand-held flag-stick configurations demonstrated by human volunteers under a human-robot interaction (HRI) framework. Both male and female participants, spanning diverse age groups, have been engaged in the cue presentation process. Each volunteer displayed the visual cues either while sitting or standing, without any fixed constraint on distance, as long as the flag-stick remained within the camera's field of view. The cue set consisted of seven classes based on spatial positioning: top-left (TL), middle-left (ML),

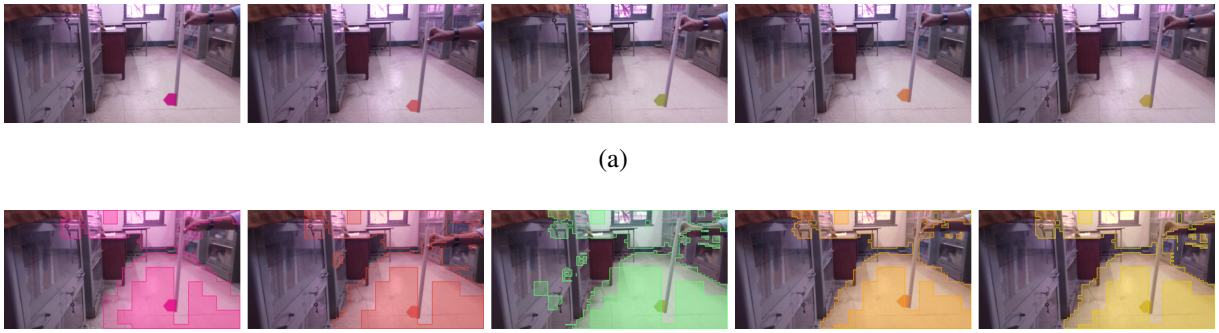
Figure 5.12: Sample flag-stick visual cues illustrating five different color variations incorporated in the dataset.

Figure 5.13: Sample flag-stick visual cues illustrating three different shape variations included in the dataset.

bottom-left (BL), neutral (N), bottom-right (BR), middle-right (MR), and top-right (TR). In contrast to the earlier dataset described in Section 2.5 of Chapter 2, various color and shape combinations of flags were used for each cue class to enhance visual diversity. Instead of only incorporating fixed-shape, fixed-color flag-stick patterns, the current dataset introduces significant augmentation by incorporating:

- **Color variability:** Five distinct color shades—rose pink, vermilion, lemon-lime, orange, and Munsell yellow. Sample flag-stick cues corresponding to each color variation are shown in Fig. 5.12.
- **Shape diversity:** Three different flag shapes—rectangle, triangle, and arrow. Representative flag-stick cues corresponding to each shape variation are shown in Fig. 5.13.
- **Lighting variability:** Similar to the earlier dataset, visual cues have been captured under five controlled lighting conditions, starting from standard indoor lighting and progressively decreasing to dark ambient states, designated as: Normal illumination level (NIL), Dark illumination level-1 (DIL-1), Dark illumination level-2 (DIL-2), Dark illumination level-3 (DIL-3), and Dark illumination level-4 (DIL-4).

These variations make the present visual cue database more comprehensive and representative of real-world environments compared to previous datasets. Furthermore, the introduction of multi-color and multi-shape stimuli improves the system's adaptability in diverse HRC contexts.



(a)

(b)

Figure 5.14: REGF-based processing outcomes for multi-colored flag-stick images: (a) Original input images, and (b) Corresponding foreground approximations derived using REGF. *Darker* and *lighter* shades indicate lower and upper approximations, respectively.

The cues have been strategically linked to robot navigation instructions to simulate actionable interaction modules for real-time robotic applications, as discussed earlier in Section 2.3 of Chapter 2.

5.6.4 Experimental Results and Discussion

This section provides a detailed evaluation of the proposed REGF-c-2DLPP scheme through classification experiments performed on the enhanced visual cue dataset introduced in Section 5.6.3. For experimental consistency, the dataset was partitioned such that 70% of the total samples were allocated for training, while the remaining 30% were retained for testing purposes. The REGF-c-2DLPP technique was used to map both training and test samples into a lower-dimensional subspace, and a support vector machine (SVM) classifier has then been applied to assess cue recognition performance. Fig. 5.11 illustrates the complete granulation pipeline using a representative flag-stick image characterized by an arrow shape and lemon-lime color, depicting the progressive transformation through CG and QTD-based granulations with lower and upper foreground approximations. Fig. 5.14 presents the REGF output for flag-stick images of multiple colors, highlighting the effectiveness of the granulation technique across chromatic variations. Table 5.4 reports classification accuracies obtained for the proposed REGF-c-2DLPP scheme in comparison with other state-of-the-art dimensionality reduction methods under various lighting environments.

As evident from the results, REGF-c-2DLPP consistently surpasses the performance

Table 5.4: COMPARATIVE PERFORMANCE ANALYSIS OF DIFFERENT DIMENSIONALITY REDUCTION TECHNIQUES ON THE MULTI-COLORED, MULTI-SHAPED FLAG-STICK CUE DATASET ACROSS VARYING LIGHTING CONDITIONS

Method	**NIL**	**DIL-1**	**DIL-2**	**DIL-3**	**DIL-4**
PCA	96.24	67.91	42.45	33.18	26.67
LDA	96.92	78.41	69.24	58.10	43.35
LPP	97.35	89.53	79.96	63.29	49.22
BLPP	97.41	93.57	89.89	69.47	56.13
2DLPP	97.58	94.90	91.71	76.88	65.17
c-2DLPP	97.75	95.27	92.80	78.63	67.39
CG-c-2DLPP	97.96	95.33	93.18	80.11	71.16
QTD-c-2DLPP	98.11	96.09	93.52	80.71	73.37
REGF-2DLPP	98.05	96.00	93.80	80.92	74.73
REGF-c-2DLPP (Proposed)	**98.20**	**96.48**	**94.74**	**81.19**	**75.62**

Table 5.5: COMPARATIVE PERFORMANCE ANALYSIS OF REGF AND OTHER FEATURE DESCRIPTORS WITHIN THE FRAMEWORK OF C-2DLPP, ON THE MULTI-COLORED, MULTI-SHAPED FLAG-STICK CUE DATASET UNDER DIFFERENT ILLUMINATION CONDITIONS

Feature Descriptor	**NIL**	**DIL-1**	**DIL-2**	**DIL-3**	**DIL-4**
HOG	97.15	94.81	91.07	76.56	67.29
Gabor	97.15	94.72	91.15	76.49	68.44
LBP	97.17	94.88	91.66	77.62	69.52
CG	97.77	95.25	93.49	79.72	71.45
QTD	97.82	95.93	93.75	80.30	73.50
REGF	**98.20**	**96.48**	**94.74**	**81.19**	**75.62**

of existing techniques across all lighting conditions. This performance advantage becomes increasingly prominent under low-illumination scenarios, indicating greater resilience of the proposed approach to photometric degradation. To further examine the strength of the proposed REGF-based granulated feature extraction in conjunction with the c-2DLPP framework,

additional comparisons have been carried out with several commonly used feature descriptors, as summarized in Table 5.5. Once again, the REGF descriptor delivers the most favorable classification outcomes, even when the lighting environment deteriorates severely. These results collectively affirm the enhanced robustness of the REGF-c-2DLPP technique, both in terms of feature encoding and neighborhood structure preservation, particularly under conditions where visual ambiguity, lighting irregularities, and appearance variability are present.

5.6.5 Key Insights

This study presented a cosine-enhanced variant of the granulated feature-assisted two-dimensional locality preserving projection, termed REGF-c-2DLPP, designed for visual cue recognition in dynamic human-robot collaboration (HRC) environments. The proposed method was specifically developed to investigate the utility and limitations of the prior REGF-2DLPP technique, particularly in terms of handling complex, multi-appearance visual cues affected by photometric distortions. In contrast to conventional LPP-based schemes that rely on Euclidean distance, the REGF-c-2DLPP approach introduces a cosine similarity-based kernel to better capture neighborhood consistency in both spatial and feature domains. This design shift allows the method to retain intrinsic relationships among high-dimensional image matrices, even when those matrices differ in magnitude due to variations in illumination or object appearance.

To support robust subspace learning, a granulated feature representation was extracted from RGB image samples using rough entropy-based foreground approximations. This combination of structure-aware and granule-sensitive representation enabled the model to effectively manage ambiguous visual patterns and illumination inconsistencies. The method was validated on a custom-constructed flag-stick marker dataset featuring multiple colors, shapes, and lighting conditions, where it demonstrated strong classification performance compared to several baseline and state-of-the-art dimensionality reduction techniques.

5.7 Summary

Effectively managing high-dimensional data in real-world environments remains a critical challenge, particularly in human–robot interaction (HRI) tasks. Dimensionality reduction (DR) techniques are essential in such contexts, as they enable the projection of complex data onto lower-dimensional manifolds, facilitating tasks such as classification and clustering. Traditional DR methods, including locality preserving projection (LPP) and its two-dimensional counterpart (2DLPP), rely predominantly on the spatial distribution of the input data. In this work, a novel feature fusion and extraction strategy, termed REGF-2DLPP, has been introduced. This method integrates the principles of granular computing by combining crisp granulation (CG) and quad-tree decomposition (QTD), thereby reducing the dependence of 2DLPP solely on spatial features. Additionally, to advance the capabilities of REGF-2DLPP, the REGF-c-2DLPP model has been introduced as a cosine similarity-augmented extension. By embedding cosine distance within both spatial and granulated feature domains, REGF-c-2DLPP improves local neighborhood encoding and achieves greater resilience against illumination-induced distortions and magnitude inconsistencies. This variant further enhances subspace learning for multi-color, multi-shape visual cues, offering more reliable performance in dynamic and visually heterogeneous HRI situations.

The principal contribution of both aforementioned techniques stems from the underlying granular framework, REGF. This addresses the individual limitations associated with both homogeneous and heterogeneous granulation approaches. In doing so, it enhances the representation of foreground-background boundaries in an image by capturing both structured and unstructured uncertainty. The granular fusion strategy developed in this work is based on the assumption that granules exhibit regular shapes. However, it is acknowledged that irregular granule shapes are more commonly observed in natural conditions, and extending the current method to accommodate such irregular formations presents a promising avenue for future research. Furthermore, the design of neighborhood granules that consider both spatial proximity and chromatic similarity is also identified as a potential direction for extending the current work. Such enhancements are expected to improve robustness, particularly in scenarios where spatio-color outliers may otherwise hinder accurate representation.

Building on these insights, the next chapter aims to explore more robust granular learning

frameworks that can effectively operate under highly challenging visual conditions, such as severe illumination degradation and the presence of random noise corruption.

Chapter 6

Density-based Neighborhood Granulation-aided Feature Extraction Strategies in Two-dimensional Robust LPP

6.1 Introduction

As discussed in the concluding section of Chapter 5, the REGF-based granulation approach, despite its promising performance, exhibits certain limitations under increasingly adverse environmental conditions. In situations with severe illumination variability or substantial noise corruption, its effectiveness reduces due to its reliance on regular granule structures, limited adaptability in handling spatial ambiguity, and simplified membership modeling. These shortcomings demand the development of more resilient and adaptive schemes capable of maintaining high recognition performance in unfavorable environmental conditions. Motivated by these observations, the current chapter presents a novel density-driven granulation framework, that integrates fuzzy feature decoding and a robust LPP-based dimensionality reduction approach to improve the reliability and generalizability of visual cue recognition systems.

The complete computational architecture of the proposed visual cue detection framework, *density-based neighborhood granulation-aided two-dimensional robust LPP* (dNG-2DRLPP), is illustrated as a flowchart in Fig. 6.1. As depicted in the diagram, the proposed dNG framework systematically extracts granular information from input images as feature descriptors. This approach integrates the variation in pixel intensities (range information) into the projection matrix to achieve effective low-dimensional embedding. In contrast to existing approaches such as 2DLPP [13], 2DLPP-L1 [46], and 2DRLPP [14], which rely solely on

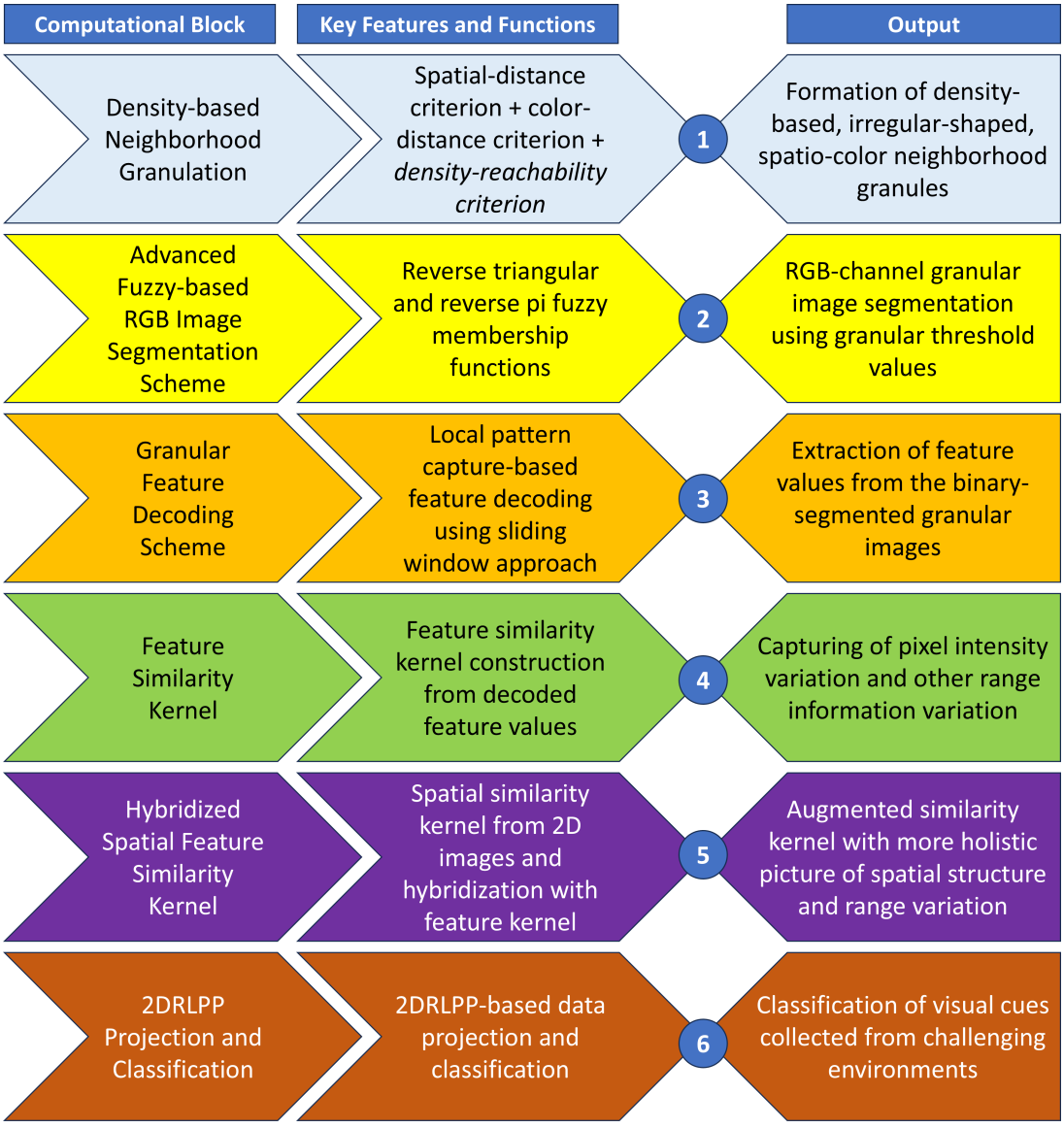


Figure 6.1: Schematic representation outlining the workflow of the proposed method dNG-2DRLPP for challenging visual cue detection.

spatial similarity for constructing the projection matrix, this method incorporates both spatial and range similarity features. This dual similarity modeling becomes particularly useful when noise and outliers introduce ambiguity between foreground and background regions or when boundaries become indistinct.

Under such scenarios, decomposing images at the granular level using granular computing (GrC) methods, followed by decoding the obtained granular features, has shown to be highly beneficial for capturing essential data characteristics. Previous studies have confirmed the efficacy of GrC-assisted methods in various vision-based applications such as object detection, tracking, and other HRI-relevant tasks. For instance, Chakraborty *et al.* [103] applied GrC for

detecting moving objects in a spatio-temporal context. Similarly, Chakraborty and Pal [105] employed GrC along with rough rule-based models to track hand movements in video data. The integration of GrC with Z-numbers and deep neural networks has also demonstrated its potential in motion detection and object recognition [104]. In another line of work, Paral *et al.* [114] leveraged GrC-based feature extraction techniques for gesture recognition using thermal imagery. However, most of these prior models rely on the assumption that granules are either homogeneous [102] or non-homogeneous [103], but are regular in shape; an assumption that may fail under real-world adversarial conditions [105].

This leads to the motivation to employ an irregular-shaped granule formation mechanism, namely *density-based neighborhood granulation* (dNG), which reflects the inherent complexity and ambiguity of real-world environments more realistically. Moreover, the density-based thresholding adopted in the proposed dNG scheme is designed to filter out outlier elements from the data more precisely and efficiently, compared to earlier neighborhood-based granulation strategies.

6.2 System Framework

This work focuses on building an assistive robotic framework that can execute various navigation tasks by interpreting vision-based cues provided by human users. A detailed schematic of the multi-stage robot guidance system, which integrates intelligent hardware and software modules, is illustrated in Fig. 6.2. The entire setup has been custom-developed at the Instrumentation and Cyber Physical System Laboratory, Department of Electrical Engineering, Jadavpur University. In this experimental framework, an 8 MP Sony IMX219 RGB camera is attached to the mobile robot to capture user-generated visual cues, as previously shown in Fig. 2.5 of Chapter 2. The robot operates using a Raspberry Pi 3 Model B+ as its main processing unit, which includes a 1.4 GHz 64-bit quad-core processor and supports dual-band WLAN and Bluetooth 4.2/BLE communication [72]. The camera module (RPi v2) features a depth of field ranging approximately from 10 cm to infinity, with a horizontal field of view of 62.2° and a vertical field of view of 48.8°. A remote workstation can communicate with the robot's hardware in real-time through 2.4 GHz and 5 GHz IEEE 802.11.b/g/n/ac WLAN protocol. Secure data transmission is ensured using a four-way handshake-based network authentication

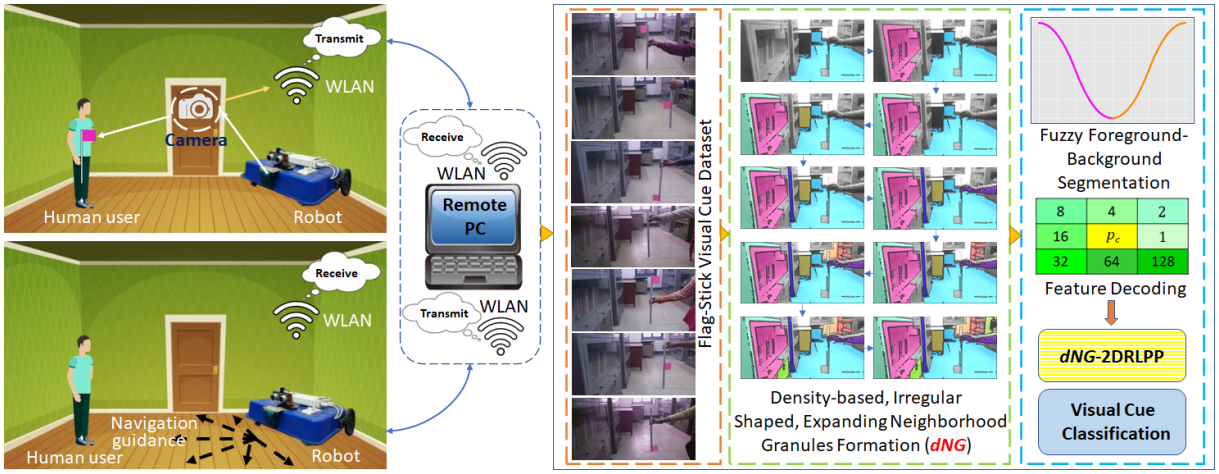


Figure 6.2: A block diagram illustrating the overall framework of the proposed robot navigation system guided by visual cues.

protocol. The real-time visual data stream is transmitted in half-duplex mode to a remote PC, which acts as a client terminal of the WLAN server. This communication setup enables real-time visual recognition tasks, leveraging previously trained models in offline mode. To carry out the computational tasks, the software is developed in Python IDE utilizing *OpenCV*, *NumPy*, and *SciPy* libraries.

For cue generation, seven distinct flag-stick patterns have been designed using a plain rod with an attached flag. These cue patterns correspond to different directional gestures: top-left (TL), middle-left (ML), bottom-left (BL), neutral (N), top-right (TR), middle-right (MR), and bottom-right (BR). Interpretations of these visual gestures in the form of navigational commands are previously outlined in Fig. 2.3 of Chapter 2. Initial data acquisition was performed under standard indoor lighting, referred to as normal illumination level (NIL). Subsequently, the lighting was systematically reduced to create more photometrically degraded conditions, designated as DIL-1 through DIL-4. The average illumination measurements for NIL, DIL-1, DIL-2, DIL-3, and DIL-4 have been recorded as 14.910 lx, 13.405 lx, 11.880 lx, 8.500 lx, and 4.710 lx, respectively. A comprehensive description and visual depiction of this image dataset are presented earlier in Section 2.5 of Chapter 2.

Besides poor illumination, other sources of image degradation such as sensor noise are frequently encountered in real-world applications. Typical causes include sensor aging, hardware wear-and-tear, random signal disturbances, reflections, scattering particles, analog-to-digital conversion (ADC) errors, or bit-level transmission errors. To emulate these effects, zero-mean additive Gaussian noise has been synthetically added to the acquired image

data here [94]. Gaussian noise is one of the most common noise types in digital imaging, arising from thermal effects, electronic interference, and insufficient illumination, factors that are highly relevant to the current study [115].

6.3 Irregular-Shaped Granulation

Granulation-based data processing frameworks have proven to be highly effective in addressing a wide range of modern machine learning challenges. Within the domain of granular computing (GrC) [102], the foundation is built upon rough set theory, which offers a robust way to represent ambiguity and uncertainty present in image data, particularly when dealing with object indiscernibility. Granular approximation techniques derived from rough sets have found successful applications across various fields, including image analysis, pattern classification, and feature extraction [104]. Specifically, in human-robot interaction (HRI), the GrC framework has demonstrated practical utility in solving detection and tracking-related problems [104], [107]. Pal *et al.* [102] initially introduced the concept of rough entropy (RE), a fundamental metric for images that facilitates the construction of homogenous, uniformly sized granules. Maximizing this RE metric leads to minimized uncertainty and vagueness in object boundary regions, which in turn supports more accurate object extraction.

Over time, the notion of granule construction has evolved from producing uniformly sized, homogeneous segments to forming non-homogeneous granules of varying sizes that better reflect the natural complexity of cluttered visual scenes [103]. Chakraborty *et al.* [103] applied a quad-tree decomposition strategy to recursively generate such granules, aiding in the identification of moving objects across spatial and temporal dimensions. Despite their differences, both homogeneous and non-homogeneous granulation approaches (HmG and NHmG) are typically built on basic, regular geometric shapes; mainly rectangular granules. However, in practical environments, the formation of visual granules follows arbitrary and irregular patterns that depend on local uncertainties and contain no pre-established geometric structure [105]. To reflect this realistically, the idea of neighborhood-based granulation becomes a more suitable choice here [116]. Building on this, Chakraborty and Pal [105] proposed a granular rough rule-based model that forms arbitrarily shaped granules by taking both spatial and color proximity into account, to reduce the effect of indiscernibility in images. In their

formulation, the spatio-color granule (SCG) $\psi_{\text{SCG}}(p_i)$ centered around a pixel p_i is defined as follows [105]:

$$\psi_{\text{SCG}}(p_i) = \bigcup_{p_j \in U} : p_i \text{ and } p_j \text{ binary connected over } |c(p_j) - c(p_i)| < \rho \tag{6.1}$$

Here, $\psi_{\text{SCG}}(p_i)$ represents the SCG constructed around the center pixel p_i, where the set includes those neighboring pixels p_j that are both spatially connected to p_i and have color intensities within a threshold ρ. The function $c(p_i)$ denotes the color intensity of pixel p_i and ρ acts as the *color proximity threshold*. The underlying principles of neighborhood-based rough set theory and RE maximization remain consistent with prior works [102]. However, unlike earlier approaches [103], [102], this model does not rely on fixed granule dimensions or shapes, allowing for the creation of irregular and more context-adaptive segments [105].

Nonetheless, arbitrarily shaped granules may sometimes be generated from only a few outlier pixels that satisfy the spatial and color proximity requirements of the center pixel, termed *spatio-color* neighbors. These neighbors are spatially close and have similar intensity values compared to the center pixel. However, it has also been frequently observed that many pixels, although spatially adjacent via binary connectivity, fail to qualify for inclusion in a granule due to the violation of the *color-distance criterion* (CDC), as given in Equation (6.1), imposed by the color threshold ρ. Even when their color distributions resemble that of the center pixel, they are excluded.

This issue often leads to the formation of trivial or inconsistent granules. To address such limitations, this work introduces a new technique—*density-based neighborhood granulation* (dNG), that seeks to improve the granule formation process by ensuring greater consistency and reliability.

6.4 Density-Based Neighborhood Granulation

The primary goal of the proposed dNG framework is to generate arbitrarily shaped spatio-color granules by jointly considering spatial proximity and color similarity among image pixels. This technique is especially effective in segmenting intricate, non-uniformly shaped visual objects while efficiently removing noisy or outlier pixels from the granules. Such characteristics make it particularly advantageous for segmenting images in real-world conditions where noise, clutter, and ambiguity are prevalent. Furthermore, the density-based strategy offers scalable

performance even when applied to high-resolution image data. Let $\mathbf{I} \in \mathbb{R}^{m \times n}$ represent an image, where each pixel $p_i \in \mathbf{I}$. The dNG approach groups those pixels which possess a sufficient number of similar neighbors in terms of both spatial adjacency and color. Pixels located in sparse or dissimilar neighborhoods are labeled as outliers.

In addition to the color proximity threshold ρ, a user-defined parameter called the *density threshold* δ is introduced in this model. The threshold δ determines the minimum number of spatio-color similar neighbors required within an α-neighborhood around a center pixel $p_i \in \mathbf{I}$ for it to be considered a valid seed for granule formation. In this work, α is set to 8, implying the use of an eight-connected neighborhood. Inspired by [117], all pixels in $\mathbf{I}$ can be classified into three mutually exclusive categories—CorePixel, BorderPixel, and NoisePixel; based on the chosen values of α, ρ, and δ. These are defined as follows:

- A **CorePixel** is one that has at least δ spatio-color similar pixels within its α-neighborhood.
- A **BorderPixel** lies in the neighborhood of a CorePixel and satisfies spatial proximity, but does not contain δ number of spatio-color neighbor pixels in its own α-neighborhood.
- A **NoisePixel** (or outlier) is neither a CorePixel nor a BorderPixel.

Using these definitions, the concept of *density-reachability* is employed to facilitate granule expansion [117]. Let us consider two pixels, v (a CorePixel) and u (any other pixel in $\mathbf{I}$). Pixel u is said to be *directly density-reachable* from v with respect to parameters ρ and δ if u is located within the α-neighborhood of v and satisfies the color proximity condition i.e., $|c(u) - c(v)| < \rho$. Mathematically, this can be expressed as:

$$u \in N_\alpha(v), \quad \text{and} \quad |N_\alpha(v)| \geq \delta \tag{6.2}$$

Moreover, pixel u can be simply termed as *density-reachable* from v if there exists a chain of pixels $u_1, u_2, \ldots, u_n$ such that $u_1 = v$, $u_n = u$, and each u_{i+1} is directly density-reachable from u_i for $1 \leq i < n$, where each $u_i \in \mathbf{I}$. Two pixels u and v are considered *density-connected* if there exists a third pixel w such that both u and v are density-reachable from w under the same neighborhood and threshold constraints. An illustrative example of the density-reachability concept among image pixels is presented in Fig. 6.3. Suppose the orange pixel acts as the center pixel (CP) and the density threshold δ is set to 4. This indicates that at least four neighboring pixels must satisfy the color-distance criterion (CDC) within the CP's α-neighborhood, for

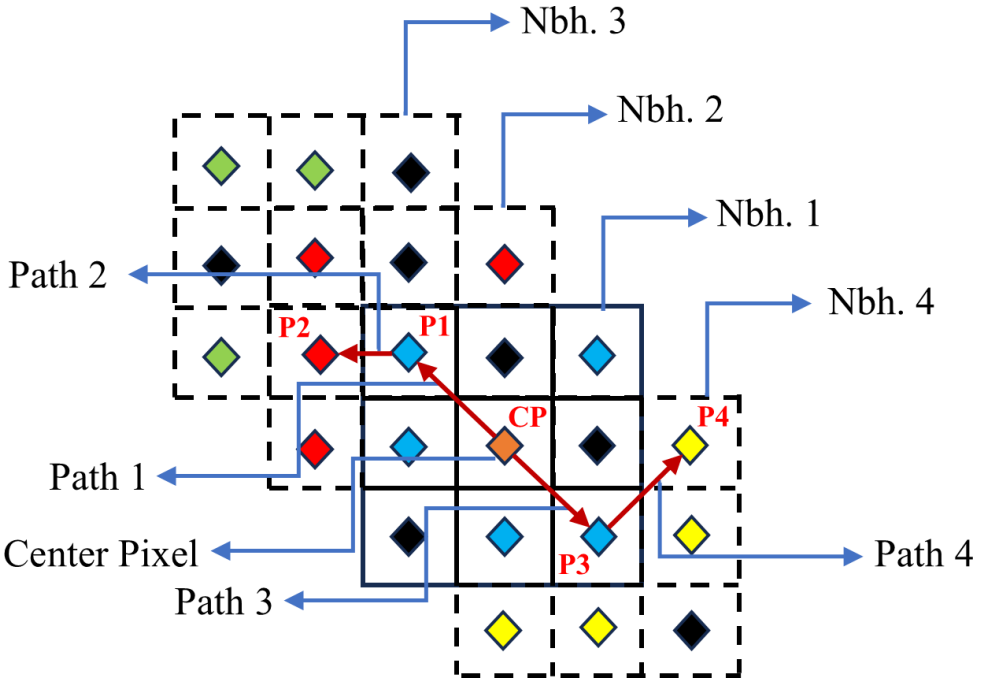


Figure 6.3: Visualization of the density-reachability concept applied to image pixels. Pixels that meet the color-distance condition (CDC) are indicated with colored diamonds, while non-satisfying pixels are displayed in black. (Here, "Nbh." refers to neighborhood.)

categorizing it to be a CorePixel. Among the neighborhoods depicted, *nbh*. 1 contains five cyan pixels fulfilling the CDC condition, making it eligible for CorePixel status. Similarly, neighborhoods *nbh*. 2 and *nbh*. 4 can be classified as density-reachable from the CP, while *nbh*. 3 does not meet the criteria. Pixels $P1$ and $P3$ are directly density-reachable from the CP via path segments 1 and 3, respectively. Further, pixels $P2$ and $P4$ are density-reachable via chained paths 1-2 and 3-4, respectively. Though P_2 and P_4 belong to different neighborhoods, they are density-connected through their mutual density-reachability from a common central pixel. A visual demonstration of this expanding granule formation based on the density-reachability condition is shown in Fig. 6.4.

A parameter sensitivity analysis has been conducted to evaluate the behavior of the proposed dNG method. The sensitivity of the parameters ρ and δ is assessed based on the number of generated granular segments and by using two standard evaluation indices—Beta index [118] and Davies–Bouldin index (DBI) [119]. The average quantitative measures visualized in Fig. 6.5, indicate that the proximity threshold ρ plays a more dominant role in determining the total number of granules formed. A higher number of segments corresponds to finer detail capture within the image, which tends to improve recognition performance but at the

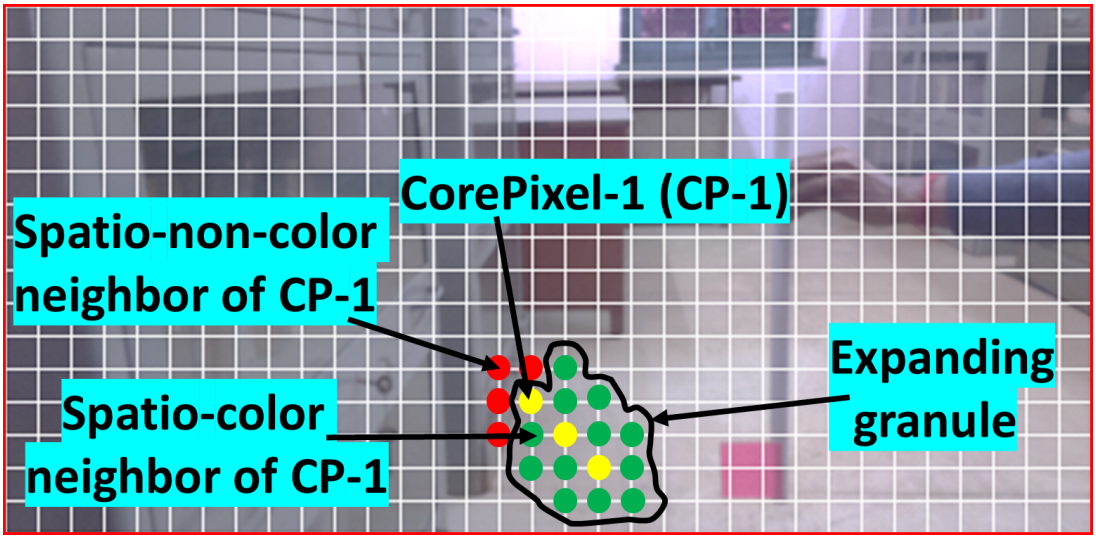


Figure 6.4: An example showcasing the expansion of granules in a representative flagstick cue image based on the applied density-reachability principle.

cost of increased computational burden. It was observed that a lower to moderately high value of $\delta \in [1, \alpha)$ generally yields the best results. A lower DBI combined with a higher Beta value indicates better segmentation quality. From the results, it is evident that $\delta = 5$ achieves the optimal performance, achieving both the highest global Beta value and the lowest DBI within the chosen parameter range, with ρ fixed at 5. The effect of the two primary parameters of the proposed dNG model i.e., the color proximity threshold (ρ) and the density threshold (δ), can also be illustrated using color-coded heatmaps. Figures 6.6a-6.6c present the average outcomes in terms of the number of generated granular segments, and the clustering evaluation parameters Beta and DBI value, respectively. An increased number of segmented granules generally leads to improved segmentation, enabling more precise and detailed thresholding of the image. From the heatmap analysis, it is evident that lower values of ρ and δ tend to yield better segmentation accuracy, although this improvement comes with a higher computational cost. Therefore, a balance is typically required between these two competing aspects, depending on the intended goal, whether the focus is on achieving better accuracy or faster processing speed. It is observed that the optimal Beta index is generally located in regions with smaller values of ρ. A higher Beta value is indicative of superior performance in terms of granule segmentation quality. On the contrary, improved results are associated with lower DBI values. However, in practical applications, the DBI metric often shows irregular behavior with no clear monotonic pattern. To handle this, specific ranges of ρ and δ are chosen where consistently higher Beta scores are observed. Within these selected regions, the configurations that correspond to the lowest DBI

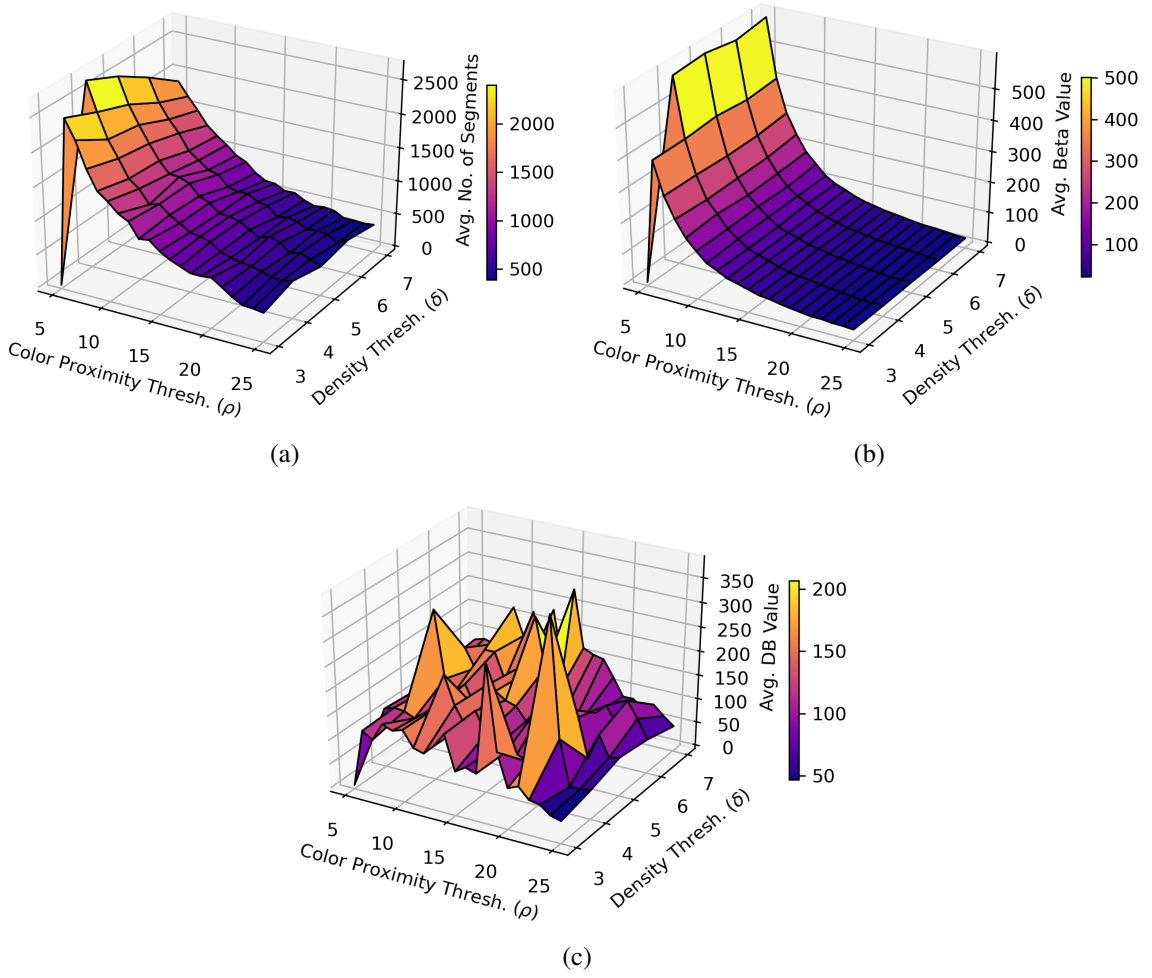


Figure 6.5: Evaluation of the proposed dNG model's parameter sensitivity, depicted through: (a) The number of segmented granules, (b) Beta index indicating segmentation quality, and (c) Davies–Bouldin index (DBI) index reflecting clustering compactness and separation.

values are considered optimal for the granulation model's performance.

In this work, the dNG operation is independently applied to each of the red (R), green (G), and blue (B) channels of the RGB image. This approach avoids the loss of important chromatic features that could occur during RGB-to-grayscale conversion, especially in regions where grayscale intensities are indistinguishable, but color differences are substantial [32]. For each RGB color channel, granules are computed separately and then unified into a single representation based on rough set approximations. For a given color image $\mathbf{A}$, the outer and inner approximations of the foreground are denoted by the sets $\{\mathbb{OF}\}$ and $\{\mathbb{IF}\}$ respectively, and are computed as:

$$\overline{\mathbb{F}} \equiv \{\mathbb{OF}\} = \{\mathbb{OF}^R\} \cup \{\mathbb{OF}^G\} \cup \{\mathbb{OF}^B\}, \quad \underline{\mathbb{F}} \equiv \{\mathbb{IF}\} = \{\mathbb{IF}^R\} \cup \{\mathbb{IF}^G\} \cup \{\mathbb{IF}^B\} \tag{6.3}$$

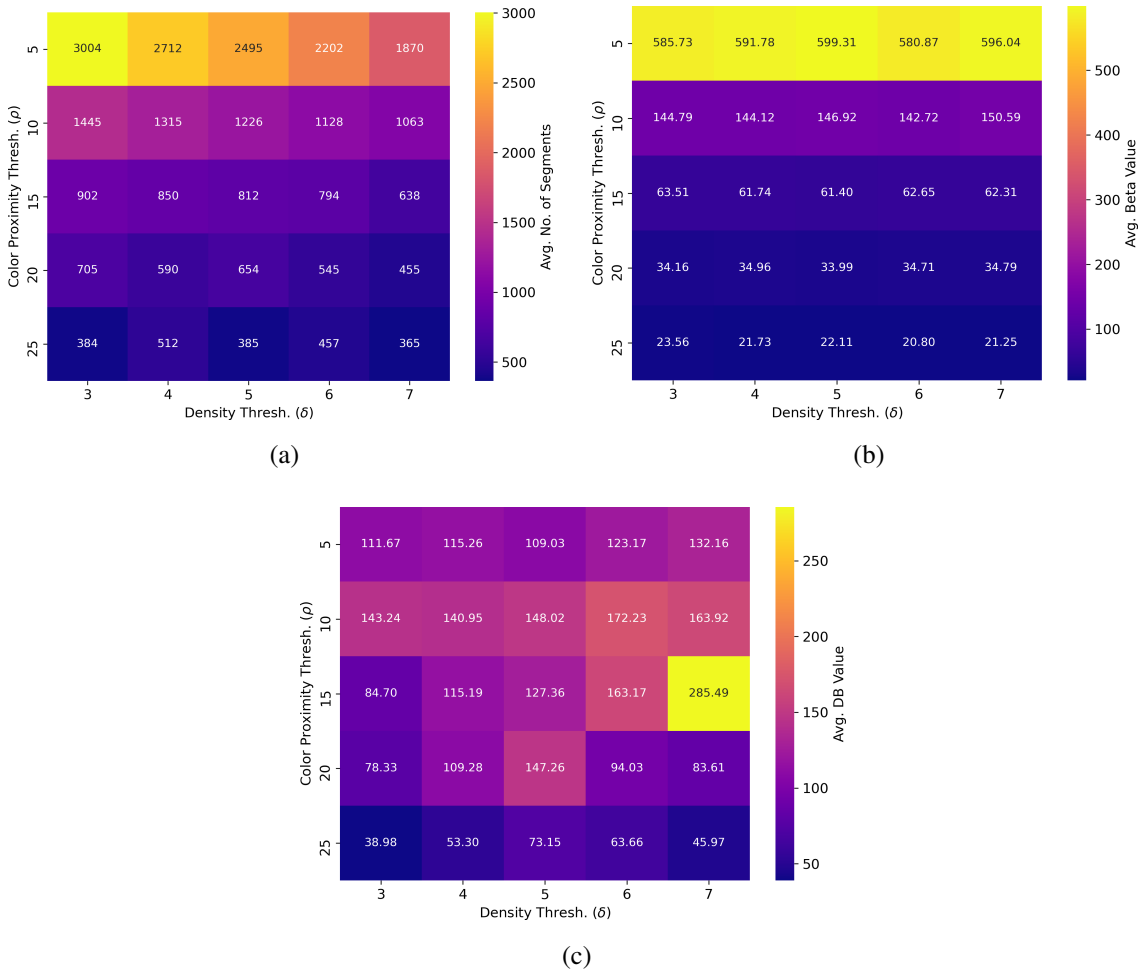


Figure 6.6: Heatmap visualization illustrating the parameter sensitivity of the proposed dNG model with respect to: (a) Total number of generated granular segments, (b) Beta index, and (c) Davies–Bouldin index (DBI).

Likewise, the outer and inner approximations for the background region are defined as:

$$\overline{\mathbb{B}} \equiv \{\mathbb{OB}\} = \{\mathbb{OB}^R\} \cup \{\mathbb{OB}^G\} \cup \{\mathbb{OB}^B\}, \quad \underline{\mathbb{B}} \equiv \{\mathbb{IB}\} = \{\mathbb{IB}^R\} \cup \{\mathbb{IB}^G\} \cup \{\mathbb{IB}^B\} \tag{6.4}$$

In the equations above, the superscripts R, G, and B denote the red, green, and blue channels, respectively. The sets $\{\mathbb{OF}^i\}$, $\{\mathbb{IF}^i\}$, $\{\mathbb{OB}^i\}$, and $\{\mathbb{IB}^i\}$ represent the outer and inner approximations for the foreground and background within each color channel. Using these approximations, the roughness measures of the foreground ($R_{\mathbb{F}}$) and background ($R_{\mathbb{B}}$) regions are calculated as [102]:

$$R_{\mathbb{F}} = 1 - \frac{|\underline{\mathbb{F}}|}{|\overline{\mathbb{F}}|} = \frac{|\overline{\mathbb{F}}| - |\underline{\mathbb{F}}|}{|\overline{\mathbb{F}}|}, \qquad R_{\mathbb{B}} = 1 - \frac{|\underline{\mathbb{B}}|}{|\overline{\mathbb{B}}|} = \frac{|\overline{\mathbb{B}}| - |\underline{\mathbb{B}}|}{|\overline{\mathbb{B}}|} \tag{6.5}$$

Here, $|\overline{\mathbb{F}}|$, $|\underline{\mathbb{F}}|$, $|\overline{\mathbb{B}}|$, and $|\underline{\mathbb{B}}|$ represent the cardinalities (i.e., the total number of elements) in the respective sets. The rough entropy (RE) of the image, based on the uncertainty in the foreground

and background approximations, is computed using the formula [102]:

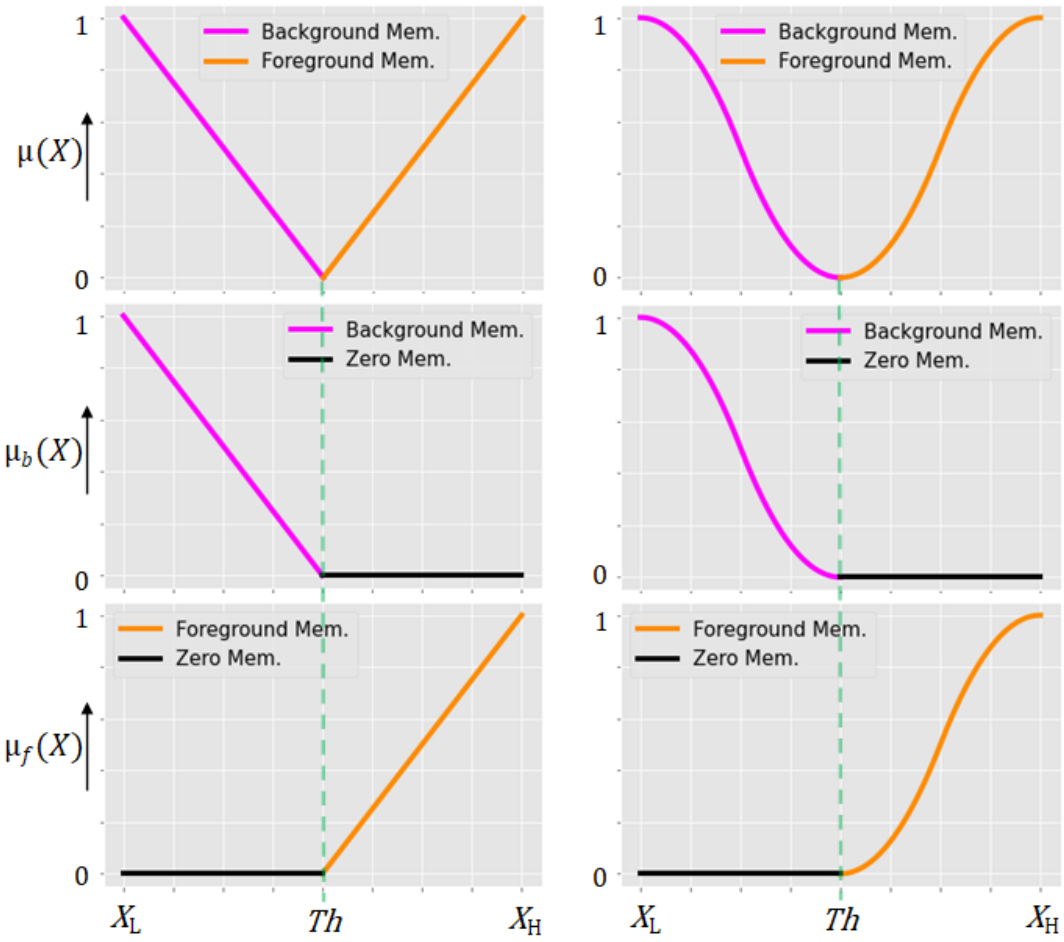
$$RE = -\frac{e}{2}\left[R_{\mathbb{F}} \cdot \ln R_{\mathbb{F}} + R_{\mathbb{B}} \cdot \ln R_{\mathbb{B}}\right] \tag{6.6}$$

To determine optimal foreground–background segmentation, this entropy measure is maximized with respect to the threshold values. For each color channel, the thresholds yielding the maximum RE are chosen and denoted as ThR, ThG, and ThB for the R, G, and B channels respectively.

6.5 Granular Feature Decoding

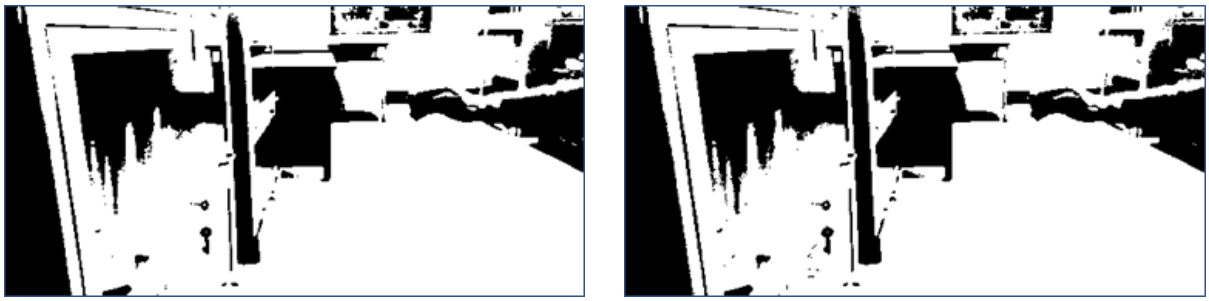
Once the irregularly shaped granules are formed, a decoding strategy is employed to extract the corresponding feature histogram from the image samples. As the foreground object in the image becomes increasingly ambiguous, the boundaries between foreground and background turn more vague and uncertain, posing challenges for segmentation based on crisp decision rules. To address this issue, fuzzy membership functions are incorporated here, for distinguishing between pixels belonging to the foreground and those in the background [114]. As a pixel p_i approaches its respective granulation threshold value Th, its likelihood of being categorized as a background pixel decreases, while correspondingly its membership in the foreground class increases. To model this gradual transition and classification between background and foreground pixels, two types of fuzzy membership functions are utilized: reverse triangular fuzzy membership (RTFM) and reverse pi fuzzy membership (RPFM). These membership functions are illustrated in Fig. 6.7. The threshold Th represents the channel-wise granulation threshold values, namely ThR, ThG, and ThB for the red, green, and blue color channels, respectively. X_L and X_H denote the minimum and maximum intensity values for image pixels, typically set to 0 and 255.

Using these parameters, the foreground and background membership functions $\mu_f(X)$ and $\mu_b(X)$ are computed for all three color channels. Based on the aggregated membership values across the R, G, and B channels, each pixel is classified as either a foreground or background pixel. This is done by comparing the cumulative foreground and background membership values. A pixel is marked as foreground if its total $\mu_f(X)$ is greater than or equal to $\mu_b(X)$, and as background otherwise. For thresholding the segmented image, foreground pixels are assigned a binary value of 1, and background pixels are assigned 0. Granular segmentation



(a) (b)

Figure 6.7: Fuzzy membership functions used for granular foreground–background segmentation, with (a) Reverse triangular fuzzy function, and (b) Reverse pi fuzzy function.



(a) (b)

Figure 6.8: Granular foreground–background segmentation output for a sample flag-stick visual cue, using (a) Reverse triangular fuzzy membership (RTFM), and (b) Reverse pi fuzzy membership (RPFM).

results for a representative flag-stick image using RTFM and RPFM are illustrated in Fig. 6.8. After segmentation, a local neighborhood window of size $n \times n$ is centered around each pixel to derive localized feature patterns. A feature descriptor ϕ_{p_c} is calculated for each central pixel p_c by aggregating the binary values of its surrounding pixels in a manner similar to the local binary pattern (LBP) approach [30]. Specifically, the feature value is computed as:

$$\phi_{p_c} = \sum_{k=0}^{n^2-1} 2^{k-1} \cdot p_k$$

where p_k represents the k^{th} pixel within the neighborhood. The total number of histogram bins generated from the image is given by $\left(n^2-1\right)\left(n^2-2\right)+3$. Finally, the resulting histogram is normalized to produce the final feature vector, which captures the local structure of the granulated image pattern. The entire granular feature decoding (GFD) process is summarized in Algorithm 6.1.

6.6 Granular Feature-Aided 2DRLPP

Let the training dataset be denoted by $\mathbf{A} \in \mathbb{R}^{m \times n \times N}$, consisting of N training image samples $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_N$, where each $\mathbf{A}_i \in \mathbb{R}^{m \times n}$. For high-dimensional image data, dimensionality reduction is essential to project samples into a more compact subspace that facilitates downstream tasks such as clustering or classification. In a two-dimensional unsupervised learning context, a projection matrix $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_d] \in \mathbb{R}^{n \times d}$ is learned, where each column vector $\mathbf{w}_k$ serves as a linear projection direction [13]. Using this projection matrix, each image is mapped to a lower-dimensional representation via the linear transformation:

$$\mathbf{X}_i = \mathbf{A}_i \mathbf{W} \tag{6.7}$$

Here, $\mathbf{X}_i$ denotes the low-dimensional embedding of the input image $\mathbf{A}_i$. A detailed description of the LPP formulation, including its mathematical derivation and the underlying graph-based embedding framework, has been presented earlier in Section 3.2.3 of Chapter 3. Nevertheless, to maintain notational consistency and facilitate convenient reference, the key mathematical expressions related to LPP are briefly revisited here. To compute $\mathbf{W}$ in the two-dimensional framework of LPP, the following optimization problem is defined [14]:

$$\min_{\mathbf{W}} \sum_{i,j=1}^{N} \left\|(\mathbf{A}_i - \mathbf{A}_j)\mathbf{W}\right\|_F^2 S_{ij}, \quad \text{s.t.} \sum_{i=1}^{N} \|\mathbf{A}_i \mathbf{W}\|_F^2 D_{ii} \tag{6.8}$$

Algorithm 6.1 Granular Feature Decoding (GFD)

Input: Color image **A**, R-G-B channel thresholds ThR, ThG, ThB, cutoff intensity bounds X_L, X_H, neighborhood size $n \times n$

Output: Normalized granular feature vector $\boldsymbol{\phi}$

STEP-1 For each pixel p in **A** and for each color channel $c \in \{R, G, B\}$:

- Compute foreground membership $\mu_f(X)$ using RTFM or RPFM.
- Compute background membership $\mu_b(X)$ using RTFM or RPFM.

STEP-2 Aggregate total foreground membership: $\mu_f^T(X) = \mu_f^R(X) + \mu_f^G(X) + \mu_f^B(X)$.

STEP-3 Aggregate total background membership: $\mu_b^T(X) = \mu_b^R(X) + \mu_b^G(X) + \mu_b^B(X)$.

STEP-4 Assign pixel p as:

- Foreground if $\mu_f^T(X) \geq \mu_b^T(X)$,
- Background otherwise.

STEP-5 Binarize the image: assign value 1 to all foreground pixels and 0 to background pixels.

STEP-6 For each center pixel p_c:

- Extract the $n \times n$ neighborhood centered at p_c.
- Compute the granular binary feature:

$$\phi_{p_c} = \sum_{k=0}^{n^2-1} 2^{k-1} \cdot p_k$$

  where p_k represents the k^{th} pixel in the local window.

STEP-7 Construct a histogram with $\left(n^2-1\right)\left(n^2-2\right)+3$ bins.

STEP-8 Normalize the histogram to obtain the final feature vector $\boldsymbol{\phi}$.

STEP-9 Return $\boldsymbol{\phi}$.

where $\mathbf{S} = \{S_{ij}\}$ is the similarity matrix and $\mathbf{D}$ is a diagonal matrix with $D_{ii} = \sum_{j=1}^{N} S_{ij}$. The matrix $\mathbf{S}$ encodes spatial proximity between samples and is constructed as:

$$S_{ij} = \begin{cases} \exp\left(-\dfrac{\|\mathbf{A}_i - \mathbf{A}_j\|_F^2}{\sigma_s}\right), & \text{if } \mathbf{A}_j \in \mathcal{N}_k(\mathbf{A}_i) \\ 0, & \text{otherwise} \end{cases} \tag{6.9}$$

Here, σ_s controls the influence of similarity weights, and $\mathcal{N}_k(\cdot)$ denotes the k nearest neighbors. However, spatial similarity alone may not be reliable under geometric distortions, sensor noise, or outliers [11], leading to degradation in the performance of manifold-inspired methods like

2DLPP. To mitigate the sole dependency on spatial structure, a complementary feature similarity matrix is introduced. It is computed using the dNG-based granular features as follows:

$$F_{ij} = \begin{cases} \exp\left(-\dfrac{\|\phi(\mathbf{A}_i) - \phi(\mathbf{A}_j)\|_2^2}{\sigma_f}\right), & \text{if } \mathbf{A}_j \in \mathcal{N}_k(\mathbf{A}_i) \\ 0, & \text{otherwise} \end{cases} \tag{6.10}$$

In the above formulation, $\phi(\mathbf{A}_i)$ denotes the granular feature vector extracted using the GFD algorithm, and σ_f is a regularization constant that adjusts feature-based weight sensitivity. Thereafter, the final similarity matrix S_{ij} is obtained by combining spatial and feature weights:

$$R_{ij} = S_{ij} \cdot F_{ij} \tag{6.11}$$

After incorporating similarity weights from both spatial and feature domains, and adopting the robust l_1-norm formulation of 2DRLPP [14], the revised optimization problem becomes:

$$\min_{\mathbf{W}^\top\mathbf{W}=\mathbf{I}} \sum_{i,j=1}^{N} \left\|(\mathbf{A}_i - \mathbf{A}_j)\mathbf{W}\right\|_1 R_{ij} + \tau\|\mathbf{W}\|_1, \quad \text{s.t.} \sum_{i=1}^{N} \|\mathbf{A}_i\mathbf{W}\|_1 D_i = 1 \tag{6.12}$$

Here, τ serves as the regularization parameter enforcing sparsity in the projection matrix $\mathbf{W}$, and the constraint $\mathbf{W}^\top\mathbf{W} = \mathbf{I}$ maintains orthonormality among the projection vectors. This optimization problem is solved iteratively using an l_1-norm minimization algorithm to derive the optimal projection vector $\mathbf{w}^*$. Repeating this process d times yields a set of optimal vectors $\{\mathbf{w}_h^*\}_{h=1}^d$, which are stacked to form the final transformation matrix $\mathbf{W}^* \in \mathbb{R}^{n\times d}$. Once learned, the transformation matrix $\mathbf{W}^*$ is used to project each training image $\mathbf{A}_i$ to a reduced representation $\mathbf{X}_i = \mathbf{A}_i\mathbf{W}^*$. In the testing phase, the same projection is applied to new input samples using the precomputed $\mathbf{W}^*$. The resulting low-dimensional feature vectors can then be passed on to any suitable classifier for evaluating the recognition accuracy of the proposed dNG-2DRLPP framework. A discussion of the computational complexity associated with the entire method dNG-2DRLPP is provided next.

• Computational Complexity Analysis

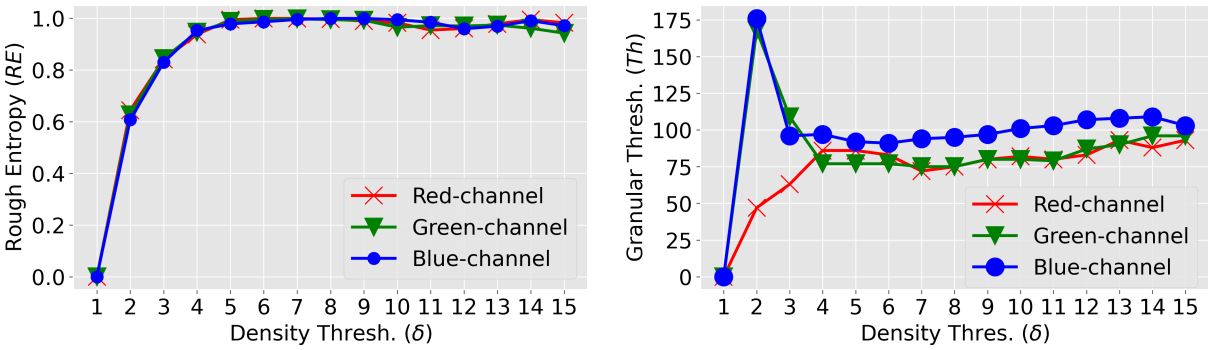
For an input image consisting of p pixels, where each pixel considers an α-connected neighborhood, the computational complexity of the dNG algorithm is $\mathcal{O}(\alpha \times p)$. The fuzzy segmentation stage applied to the color image introduces an additional complexity of $\mathcal{O}(p)$. In the granular feature decoding process, local computations are performed using α-pixel neighborhoods, which again contributes a complexity of $\mathcal{O}(\alpha \times p)$. Within the 2DRLPP

framework, the similarity matrix is built using exhaustive pairwise comparisons between all input samples. This results in a computational complexity of $\mathcal{O}(N^2)$, where N denotes the total number of training images. Once the similarity kernel is computed, generating a single orthogonal projection direction involves iterative optimization. If T represents the number of iterations required for convergence, the corresponding complexity is $\mathcal{O}(N^2 \times T)$. To compute a complete projection matrix consisting of d such projection vectors, the overall complexity becomes $\mathcal{O}(N^2 \times T \times d)$. Therefore, bringing all components together, the total computational complexity of the proposed dNG-2DRLPP pipeline can be approximated as:

$$\mathcal{O}\left(N(2\alpha p + p) + N^2(1 + T + T \cdot d)\right)$$

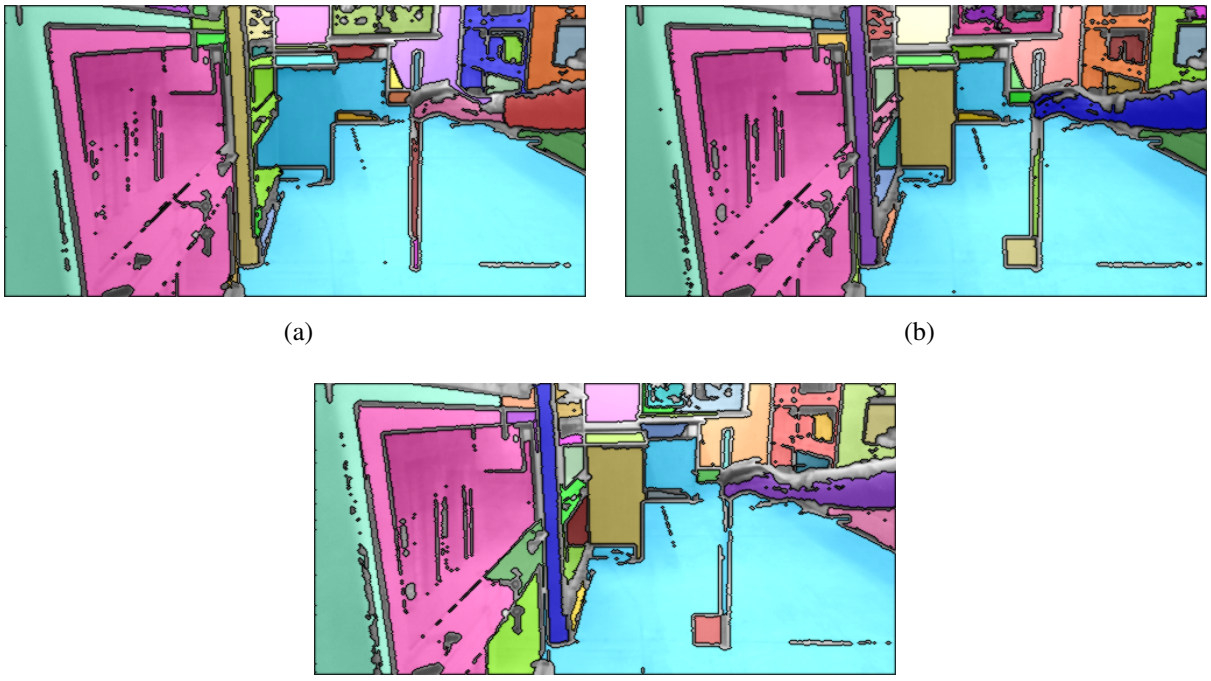
6.7 Performance Evaluation and Discussion

To assess the effectiveness of the proposed approach, a comprehensive set of experiments has been conducted using the previously described flagstick image dataset. Initially, the dNG algorithm has been applied to the input images. Notably, the RGB color images have been used in their original form rather than being converted to grayscale. This helps preserve significant chromatic details that might otherwise be lost during RGB-to-grayscale conversion, especially in regions where intensity levels are ambiguous, but color variations remain informative [32]. In this study, the red, green, and blue channels of the images have been processed individually, with the dNG procedure applied separately to each. As noted earlier, the choice of an appropriate density threshold δ plays a critical role in successful granule expansion during the dNG process. Accordingly, a systematic evaluation has been undertaken to determine optimal values of δ in each RGB channel, guided by the principle of maximizing rough entropy (RE) in the granulated images [102]. Simultaneously, the corresponding granulation thresholds for segmenting foreground and background regions have also been derived. Fig. 6.9a and 6.9b illustrate the variation of RE and threshold (Th) values with respect to different δ values across the three color channels. For this experimental setting, the optimal values of the density threshold that yield maximum RE have been found to be 7 for both red and green channels, and 9 for the blue channel. Using these optimal values, dNG produces irregularly shaped granules for each channel of the input flag-stick images. A sample image with the resulting granules for red, green, and blue channels is shown in Figs. 6.10a-6.10c, respectively, where different



(a) (b)

Figure 6.9: Variations in the (a) Rough entropy (RE) measure and (b) Granulation threshold values with respect to different density threshold (δ) settings.



(a) (b)

(c)

Figure 6.10: Example of a flag-stick visual cue image with irregularly shaped neighborhood granules generated in (a) Red channel, (b) Green channel, and (c) Blue channel.

color shades represent distinct granules. The size of the granules varies depending on the local spatio-color pixel density, while the shapes remain arbitrarily defined and non-uniform.

The recognition performance of the proposed dNG-2DRLPP method has then been evaluated under five varying lighting conditions, as previously explained in Section 2.5 of Chapter 2. A comparative analysis of recognition accuracy is presented in Table 6.1, highlighting the performance of dNG-2DRLPP against other state-of-the-art dimensionality

Table 6.1: CLASSIFICATION ACCURACY (%) OF VARIOUS DR METHODS UNDER DIFFERENT ILLUMINATION CONDITIONS

DR Method	NIL	DIL-1	DIL-2	DIL-3	DIL-4
LPP	99.683	92.362	84.127	70.773	55.710
BLPP	99.965	96.329	94.603	75.221	62.394
2DLPP	99.968	98.223	97.114	83.623	73.521
2DRLPP	99.975	98.905	98.455	88.349	80.106
HOG-2DRLPP	99.981	99.121	98.766	89.815	81.278
Gabor-2DRLPP	99.981	99.256	98.908	90.127	82.485
LBP-2DRLPP	99.981	99.483	99.179	90.632	83.792
HmG-2DRLPP	99.981	99.534	99.250	90.813	84.280
NHmG-2DRLPP	99.981	99.534	99.295	91.067	84.941
dNG-2DRLPP	**99.990**	**99.730**	**99.465**	**91.234**	**85.913**

reduction methods. Among these techniques, 2DRLPP demonstrates notable resilience under low-light conditions. When spatial and feature information are jointly incorporated, the recognition performance is further improved. To offer a comprehensive comparison, conventional feature descriptors such as histogram of oriented gradients (HOG), Gabor features, local binary patterns (LBP), and GrC-based descriptors (HmG and NHmG) have been evaluated alongside 2DRLPP. The proposed dNG feature descriptor outperforms all these methods, achieving over 85% accuracy even under the most restricted illumination settings.

Following the recent advancements of deep learning (DL) architectures in visual recognition tasks, a few of the popular DL models have been included for comparative evaluation in this study. These models deliver satisfactory recognition performance under normal illumination conditions (NIL), but their accuracy significantly declines when tested under adverse conditions, particularly at DIL-4. At the most restrictive lighting level (DIL-4), the recognition accuracies achieved by MobileNet, ShuffleNet, EfficientNetB0, and ResNet50 [120] are 76.569%, 72.220%, 78.796%, and 81.652%, respectively. A plausible reason for this observed drop in performance could be the limited size of the training dataset used in this study; a situation that is commonly encountered in practical, real-time data collection practices. Deep learning models typically demand large-scale datasets to effectively tune their parameters and

Table 6.2: CLASSIFICATION METRICS FOR dNG-2DRLPP UNDER DIFFERENT ILLUMINATION CONDITIONS

Metric	**NIL**	**DIL-1**	**DIL-2**	**DIL-3**	**DIL-4**
Accuracy (%)	99.990	99.730	99.465	91.234	85.913
Sensitivity (%)	99.683	99.365	99.048	91.429	86.032
Specificity (%)	99.947	99.894	99.841	98.571	97.672
Precision (%)	99.689	99.379	99.082	91.988	87.598
NPV (%)	99.947	99.895	99.842	98.571	97.673
FAR (%)	0.053	0.106	0.159	1.429	2.328
FRR (%)	0.317	0.635	0.952	8.571	13.968
EER (%)	0.186	0.186	0.186	2.407	3.704
F1-Score (%)	99.683	99.365	99.051	91.572	86.409
MCC (%)	99.632	99.264	98.901	90.234	84.353
Kappa Score (%)	98.704	97.407	96.111	65.000	42.963

Table 6.3: CLASSIFICATION ACCURACY (%) OF VARIOUS DR METHODS UNDER DIFFERENT GAUSSIAN NOISE VARIANCE LEVELS

DR Method	$\sigma^2 = 0.05$	$\sigma^2 = 0.10$	$\sigma^2 = 0.15$	$\sigma^2 = 0.20$	$\sigma^2 = 0.25$	$\sigma^2 = 0.30$	$\sigma^2 = 0.35$
LPP	85.097	81.824	79.695	76.317	75.950	73.034	70.127
BLPP	97.443	93.438	91.107	86.382	81.765	77.795	73.187
2DLPP	98.064	97.943	94.120	87.325	82.149	78.866	74.737
2DRLPP	98.864	98.640	96.769	90.950	86.678	82.757	78.843
HOG-2DRLPP	98.951	98.869	97.149	91.584	87.224	83.196	79.447
Gabor-2DRLPP	99.055	98.989	97.257	91.925	87.492	83.525	80.352
LBP-2DRLPP	99.206	99.138	97.547	92.381	87.835	84.261	80.783
HmG-2DRLPP	99.553	99.320	97.964	92.571	88.225	85.106	81.232
NHmG-2DRLPP	99.553	99.246	97.757	92.790	88.870	85.661	81.840
dNG-2DRLPP	**99.782**	**99.498**	**98.276**	**93.960**	**89.822**	**86.367**	**82.585**

generalize well to new data. Moreover, when DL networks are trained on clean, well-lit samples acquired under controlled settings but deployed on test data affected by adverse conditions like noise or low illumination, their performance tends to deteriorate significantly [121], [122]. This

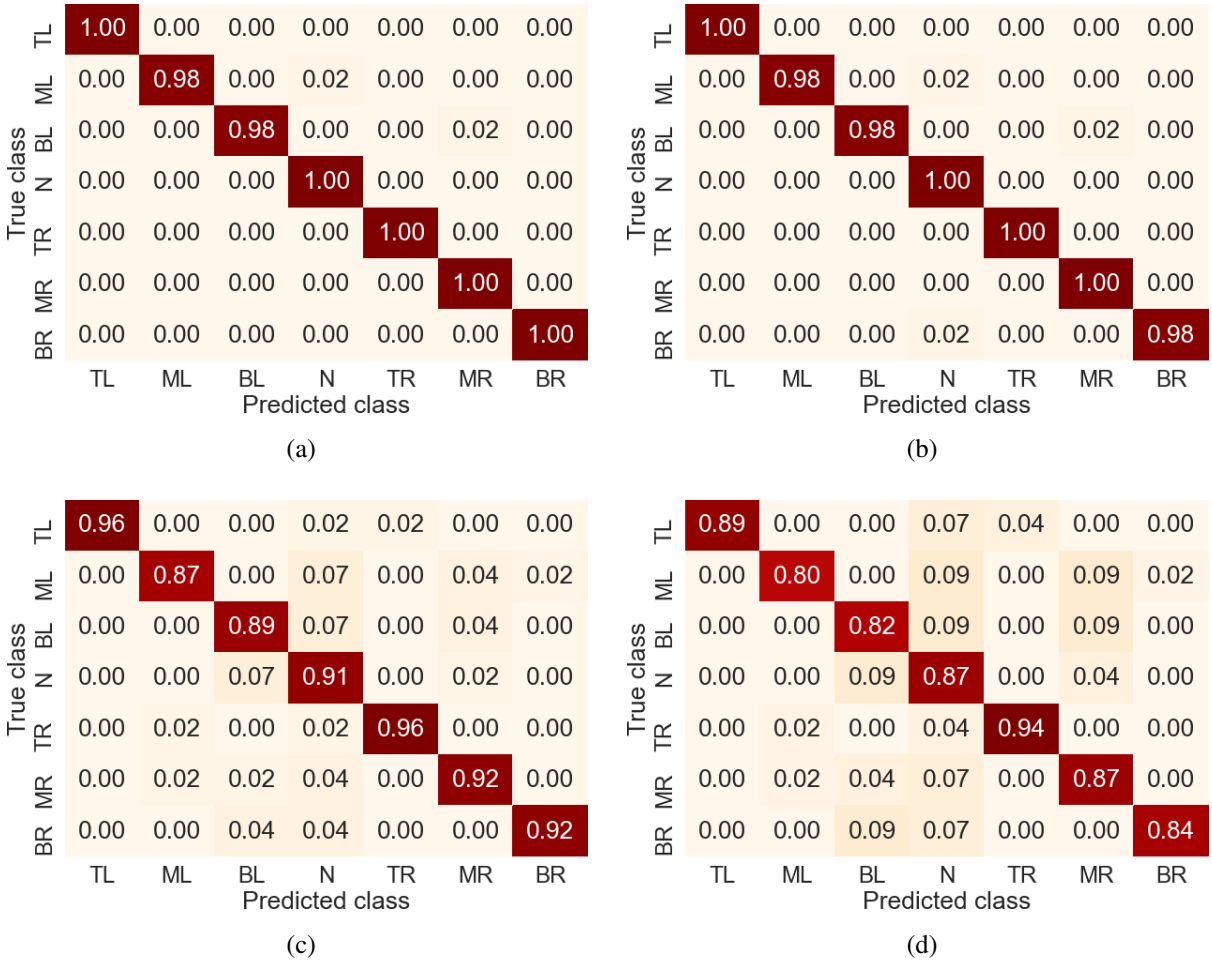


Figure 6.11: Confusion matrix representations of the dNG-2DRLPP algorithm under different illumination environments (a) DIL-1, (b) DIL-2, (c) DIL-3, and (d) DIL-4.

is a widely recognized issue in real-world applications, where unexpected variations in ambient lighting, sensor hardware, or scene structure can arise abruptly and unpredictably.

Further, a detailed performance evaluation of dNG-2DRLPP has been carried out using multiple classification metrics, as shown in Table 6.2. Metrics include accuracy, sensitivity, specificity, precision, negative predictive value (NPV), false acceptance rate (FAR), false rejection rate (FRR), equal error rate (EER), F1-score, Matthews correlation coefficient (MCC), and Cohen's kappa score. Results have been reported across all five lighting levels. Notably, even under the most difficult lighting condition (DIL-4), the model maintains strong robustness, yielding a low FAR of approximately 2.3% and an EER close to 3.7%. For more deeper insights into the class-wise performance, confusion matrices have been generated using the proposed dNG-2DRLPP method under the photometric conditions DIL-1, DIL-2, DIL-3, and DIL-4, and are illustrated in Figs. 6.11a-6.11d, respectively.

To evaluate the model's resistance to noise, flag-stick images have also been artificially corrupted with Gaussian noise at varying intensity levels. Experimental results for Gaussian noise with different variances (σ^2) are presented in Table 6.3. Here too, the proposed method demonstrates superior performance across all noisy conditions. The benefits of 2DRLPP become more pronounced as the noise intensity increases. The integration of the dNG-based granular features further enhances robustness. Even at a noise variance of $\sigma^2 = 0.35$, the model achieves an accuracy exceeding 82%, surpassing other competing methods.

6.8 Summary

A novel dimensionality reduction (DR) approach, grounded in granular computing (GrC), has been introduced in this study. To facilitate the extraction of informative features from challenging image datasets, a new density-driven neighborhood granulation strategy, termed dNG, has been developed. Unlike traditional granulation methods that assume regular-shaped granules in both homogeneous and non-homogeneous approaches, the proposed method performs image granulation using irregularly shaped segments. A density-reachability criterion has been formulated to more effectively exclude outlier regions during the granulation process. Subsequent to this granule segmentation, relevant features have been derived using a fuzzy rule-based local feature decoding technique. The decoded feature information has been further fused with a spatial similarity kernel, and the resultant kernel has been utilized in low-dimensional projections. Extensive experiments using a custom flag-stick image dataset acquired under a variety of real-world visual distortions, such as poor lighting and sensor noise, demonstrate the robustness of the proposed method. In this work, color proximity threshold and density threshold values have been obtained through a parameter sensitivity analysis using quantitative evaluation metrics. As a potential future direction, metaheuristic or other intelligent optimization algorithms could be employed to automatically determine the optimal values of the color proximity and density thresholds, thereby enhancing the adaptability and generalization capability of the proposed granulation framework, across diverse visual environments.

The progression across Chapters 3 to 6 reflects a systematic enhancement in algorithmic robustness, aimed at addressing increasingly challenging visual recognition scenarios. Beginning with foundational approaches that performed effectively under normal or moderately

distorted conditions, each subsequent chapter introduced refinements that targeted specific limitations, such as vulnerability to illumination changes, spatial ambiguity, and sensor noise. Among these, the proposed dNG-2DRLPP framework in this chapter, demonstrates superior resilience in severely degraded environments through its integration of density-driven granulation, fuzzy decoding, and robust subspace learning. This gradual evolution underscores the research objective of developing adaptive and reliable vision-based recognition systems for real-world applications. Looking forward, in the future, the framework can be extended to handle even more complex operational contexts, such as dynamic background interference, multi-modal sensory fusion, and abrupt photometric or geometric perturbations.

Chapter 7

sEMG-based Hand Activity Capturing System for Robot Guidance Applications

7.1 Introduction

Surface electromyography (sEMG) signals, representing the electrical activity of skeletal muscles, are being increasingly utilized in diverse applications including prosthetic rehabilitation, human-machine interface (HMI), and robotic manipulation. In particular, sEMG-based hand gesture and grasp recognition has gained prominence in assisting individuals with motor impairments, and in facilitating intuitive control in collaborative robotic systems. The recognition of muscular patterns through sEMG, enables efficient mapping of human intent to mechanical action, thereby enhancing task execution in assistive and industrial environments. The complexity of upper limb movement, involving multiple degrees of freedom across the elbow, wrist, and fingers, introduces significant intra-subject and inter-subject variability in the recorded EMG signals. These signals are often non-stationary, susceptible to noise, and affected by muscle fatigue, crosstalk, and anatomical variations. Consequently, effective preprocessing and dimensionality reduction (DR) techniques are essential to extract relevant features and lessen the computational burden associated with real-time classification.

To support experimental research in this domain, a custom laboratory-based sEMG data acquisition system has been developed, employing low-cost sensors and embedded modules such as Arduino Uno and ESP32. This setup captures myoelectric activity from key forearm muscles associated with functional hand movements. The acquired signals have been stored locally on a microSD card and also transmitted in real-time to an online server for remote monitoring. A set of ten representative hand movement gestures has been designed to reflect practical use cases ranging from everyday object manipulation to repetitive physical tasks.

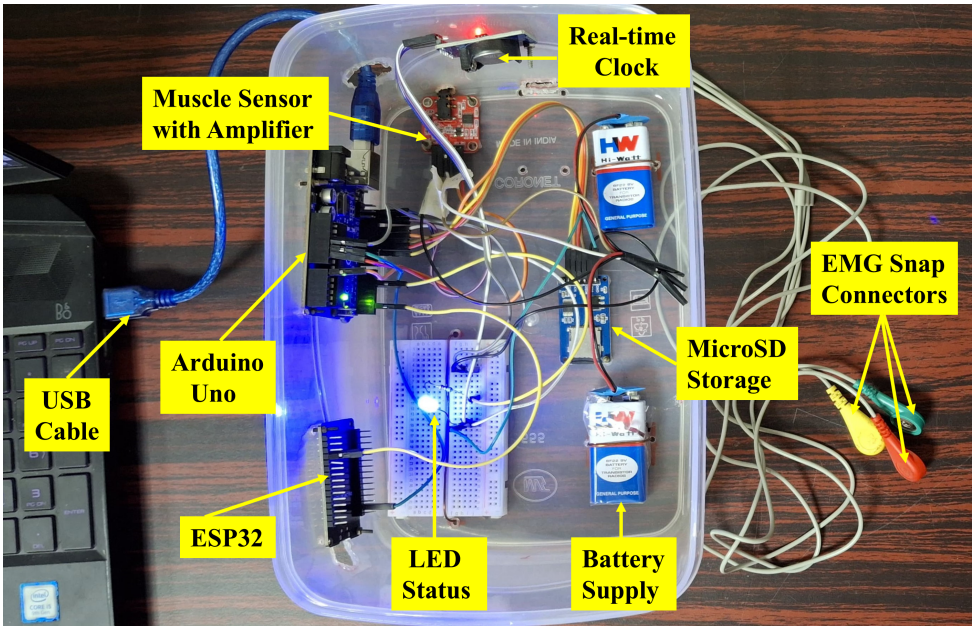


Figure 7.1: Custom-built sEMG signal acquisition setup developed in laboratory, highlighting and labeling its individual components.

The laboratory-acquired sEMG dataset along with benchmark datasets such as NinaPro, enable comprehensive validation of manifold-inspired DR frameworks for these applications. These methods aim to project the high-dimensional sEMG data into low-dimensional spaces that preserve local geometric structure while enhancing inter-class separability. Building on this foundation, the subsequent sections will describe the complete sEMG data acquisition framework, experimental settings, signal recording protocol, and dataset organization for subsequent analysis and recognition tasks.

7.2 Data Acquisition Terminal

All experimental activities in this study have been undertaken within the premises of the Instrumentation & Cyber Physical System Laboratory, Department of Electrical Engineering at Jadavpur University, India. A custom-designed, in-house sEMG signal recording setup, as shown in Fig. 7.1, has been employed for collecting the data. This system incorporates an EMG Muscle Sensor Module v3.0, which is capable of capturing muscle-generated electrical signals arising from physical movements. The sensor contains a built-in signal conditioning unit that processes the raw myoelectric signal by amplifying it and transmitting it as an analog voltage output to a microcontroller [123]. The sensor unit, which weighs approximately 30

grams and operates on a dual power supply of ± 9 volts. Three surface electrodes, connected through EMG snap leads, have been placed on the subject's body to record differential muscle activity. Two of them have been placed on the muscle of interest and one on a reference point. Specifically, the red lead is connected to an electrode placed centrally over the muscle belly of interest, while the green lead connects to another electrode located at one end of the same muscle. The reference electrode, associated with the yellow lead, is attached to a nearby bony or non-muscular neutral area that remains relatively inactive during the muscle contractions [123]. Prior to electrode placement, the skin surface has been thoroughly cleaned using alcohol wipes. For this setup, pre-gelled, single-use Ag/AgCl surface electrodes with integrated skin abrasives have been used to ensure consistent signal quality. The built-in conductive gel reduces skin-electrode impedance, enhancing signal transmission.

The complete data acquisition unit comprises the EMG sensor module with electrodes, an Arduino Uno microcontroller, a Wi-Fi-enabled ESP32 module, a DS3231 real-time clock (RTC), and a microSD storage interface. The sensor module is equipped with an onboard signal conditioning circuit that performs filtering and rectification of the acquired muscle activity signals for suppressing noise and external disturbances, to some extent. The processed signal is then amplified and converted into an analog voltage output for feeding into the microcontroller-based acquisition systems. These EMG output signals have then been routed to an analog input pin of the Arduino Uno via connecting wires. The Arduino's ATmega328P microcontroller samples the analog EMG signal using its built-in 10-bit ADC, converting it into a digital representation. This digitized data has been logged onto a local microSD card along with corresponding timestamps from the DS3231 RTC module. In parallel, the digital values have then been transmitted to the ESP32 module via TX-RX serial communication for real-time streaming. A unidirectional data flow has been established by connecting the Arduino's TX (transmit) pin to the RX (receive) pin of the ESP32. The ESP32 uses its built-in Wi-Fi to upload the data to a cloud server, enabling remote monitoring of real-time sEMG signals. For this study, the open-source IoT platform *ThingSpeak* has been employed for live visualization and storage of the transmitted data [124]. The real-time transmission of EMG data from the Arduino Uno microcontroller to the ThingSpeak server, facilitated by the ESP32 module, is illustrated in Fig. 7.2. This approach supports remote access and monitoring of sEMG recordings via any connected device. Additionally, data can be retrieved either directly from the local microSD storage or remotely via the ThingSpeak cloud interface for further analysis.

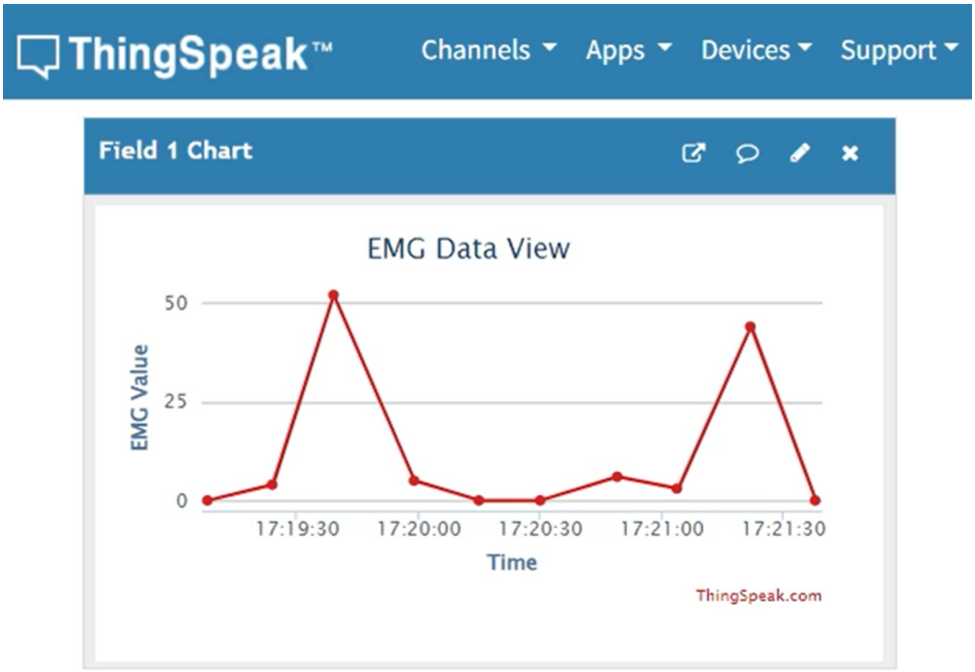


Figure 7.2: Real-time visualization of EMG signals on the ThingSpeak server, transmitted via the ESP32 module.

7.3 Anatomical Overview of Targeted Muscles

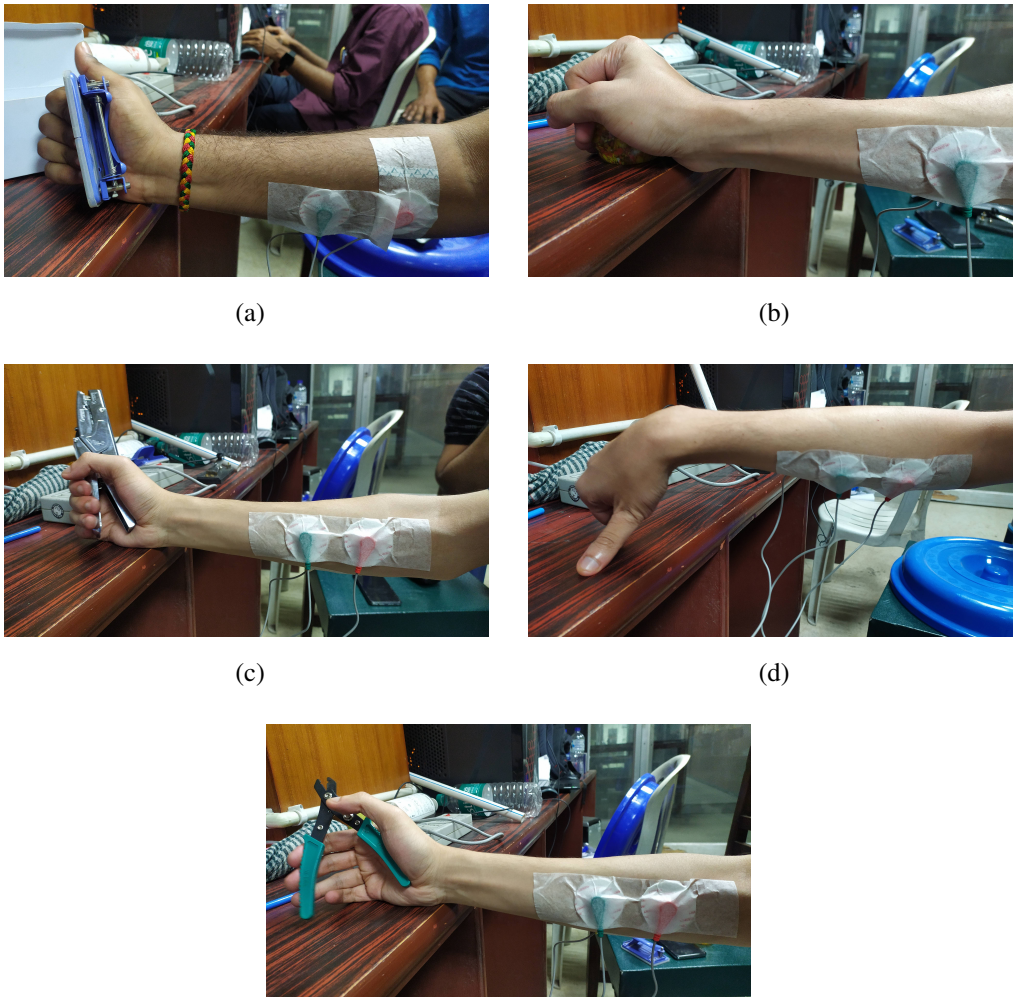
Understanding the anatomical layout of the forearm muscles and their associated movements is essential for the accurate placement of surface electrodes in sEMG acquisition. The functional roles of muscles are commonly described using specific physiological terms such as *flexion*, *extension*, *abduction*, and *adduction*. Flexion refers to the action of bending a joint, thereby decreasing the angle between articulating bones, such as curling the wrist toward the forearm. Extension is the opposite movement, where the joint angle increases, as in straightening the wrist. Abduction generally indicates movement away from the body's midline, while adduction refers to movement toward the midline. In the context of wrist mechanics, *wrist abduction* signifies wrist movement toward the thumb side, which is functionally equivalent to radial deviation, whereas *wrist adduction* means movement toward the little finger side, which corresponds to ulnar deviation. Additionally, anatomical references are made to directional terms: the anterior (volar) side refers to the front or palm-facing surface of the forearm, while the posterior (dorsal) side indicates the back or knuckle-facing surface. The radial (thumb) side corresponds to the lateral aspect of the forearm closer to the thumb, and the ulnar (pinkie) side refers to the medial aspect aligned with the little finger. These terms provide

a biomechanical basis for interpreting the muscular contributions of various forearm regions involved in voluntary hand and wrist movements.

The electrode placements target various upper forearm muscles involved in wrist and finger movements. These include the *flexor carpi radialis* (FCR), which lies on the anterior radial side of the forearm and supports wrist flexion and abduction on the radial side. Electrodes have been placed around 5-7 cm below the elbow crease, following the orientation of muscle fibers, longitudinally down the forearm [125]. Similarly, the *flexor carpi ulnaris* (FCU), located on the anterior ulnar side of the forearm and responsible for wrist flexion and adduction, has been targeted with electrodes positioned 4-6 cm distal to the elbow, over the muscle belly [126]. On the dorsal side, the *extensor carpi radialis longus* (ECRL) and *extensor carpi radialis brevis* (ECRB), which facilitate wrist extension and radial deviation, have been monitored by placing electrodes roughly 6-8 cm below the lateral epicondyle of the humerus [127]. The lateral epicondyle refers to a bony prominence located on the outer side of a bone, typically serving as a site for muscle and ligament attachment. The humerus is the long bone in the upper arm or brachium, located between the shoulder and the elbow. For recording from the *extensor carpi ulnaris* (ECU), responsible for wrist extension and ulnar deviation, the electrodes have been positioned around 8-10 cm distal to the elbow on the posterior ulnar region [128]. Additionally, the *extensor digitorum* (ED), situated centrally on the dorsal forearm and involved in finger extension, is targeted by placing electrodes longitudinally down the middle of the forearm, 6-8 cm below the lateral epicondyle.

7.4 Description of Acquired sEMG Dataset

In the conducted experiments, a total of ten distinct hand movement gestures have been considered, and the corresponding sEMG signals have been recorded for each class. Prior to data collection, informed consent has been obtained from all participating volunteers. The selected gesture classes include: (1) pulling an object, (2) pushing upward on a desk, (3) pressing downward on a desk, (4) pressing the thumb against a surface, (5) punching using a stapler, (6) pressing a paper punch, (7) grabbing a paperweight, (8) pressing a plier, (9) opening a bottle cap, and (10) twisting a towel. A few sample instances depicting the laboratory-based EMG acquisition process for five of these gestures are provided in Fig. 7.3.



(a) (b)

(c) (d)

(e)

Figure 7.3: Illustration of the EMG signal recording procedure for five representative hand activity classes: (a) Pressing a paper punch, (b) Grabbing a paperweight, (c) Punching with a stapler, (d) Pressing the thumb onto a desk, and (e) Pressing a plier.

Such sEMG-based systems offer practical relevance in rehabilitation by enabling continuous monitoring of motor-impaired patients and by facilitating focused muscle reconditioning. Moreover, in human-machine interface (HMI) contexts, accurate classification of these gestures promotes seamless and intuitive control over assistive robotic platforms, such as robotic manipulators, thereby improving functionality and enhancing the quality of life for individuals with disabilities. Beyond clinical applications, these systems can also help general users in collaborative tasks aided by assistive robots. For example, EMG-based recognition of actions such as pressing, punching, or twisting can serve as intuitive control signals for robots to execute corresponding physically demanding tasks. As an illustration, a person firmly grabbing

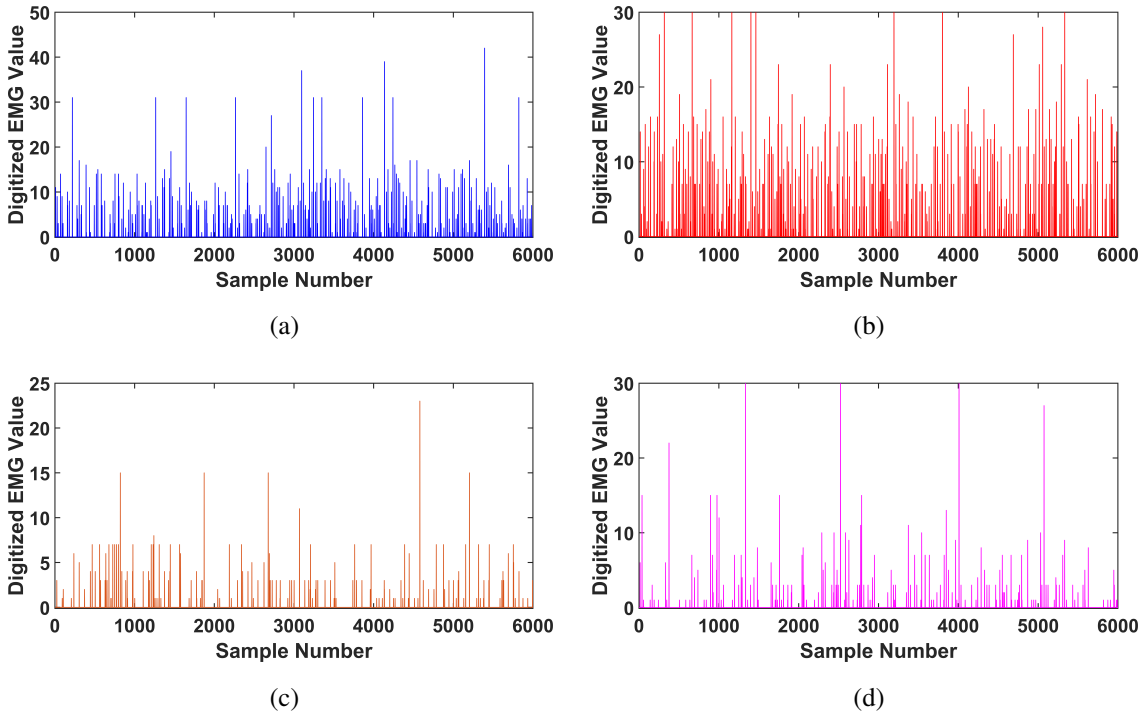


Figure 7.4: Representative EMG signal waveforms corresponding to four distinct hand movement categories.

an object or pushing something down on a floor may guide a robotic manipulator to carry out similar functions on heavy industrial machinery. This integration not only improves productivity and operational safety in shared human-robot workspaces but also unlocks new possibilities for intention-driven robotic control in complex and precision-demanding environments. Representative EMG signals from four different gesture categories are depicted in Fig. 7.4, each shown over a 12 second duration and sampled at a rate of 500 Hz.

In the experimental protocol, surface EMG signals have been collected from a total of 25 individual participants. A set of 10 hand movement gestures as previously mentioned, has been recorded across two separate recording sessions. For each gesture in every session, signals have been captured over a duration of 90 seconds, from which the central 80 seconds have been retained for inclusion in the final dataset, excluding transitional movement phases. The full 90-second recordings have been divided into three segments of 30 seconds each, with a 10-second resting interval between them, to minimize the onset of muscular fatigue. Initially, a narrow-band notch filter centered at the power line frequency has been applied to suppress any interference caused by power line noise. Following this, the mean value of each signal has been subtracted to eliminate baseline drift and remove any DC offset from the recorded EMG data.

7.5 Summary

This chapter presented the design and implementation of a surface electromyography (sEMG)-based data acquisition system tailored for hand activity recognition in robot guidance applications. A custom-built, low-cost acquisition setup has been developed using Arduino Uno and ESP32 modules, capable of both local storage and real-time transmission of sEMG data to a cloud server. Key anatomical muscles responsible for wrist and finger movements have been identified and selected for electrode placement. Relevant physiological and anatomical terms have been clarified and discussed in detail in this chapter to support an accurate and consistent sensor positioning protocol. The recorded gestures were selected to reflect practical, functional hand movements relevant to assistive and industrial contexts.

A laboratory dataset comprising ten distinct gesture classes was acquired from 25 volunteers, with recordings performed across multiple sessions under a standardized protocol. Signal preprocessing steps, including notch filtering and baseline correction, were applied to ensure high-quality data suitable for machine learning-based analysis. The captured sEMG signals and their corresponding gesture labels form a foundational resource for validating dimensionality reduction and classification methods, enabling intention-aware robotic systems that respond to natural muscular cues. Such categorized sEMG signals have direct implications in real-world HMI environments. In prosthetic control, accurate classification of muscle activity enables intuitive, proportional actuation of artificial limbs based on user intent. In rehabilitation, these systems support targeted muscle retraining and progress monitoring for individuals with neuromuscular impairments. Within collaborative robotics and manipulator-driven tasks, gesture-based control using sEMG facilitates seamless and hands-free operation of robotic arms in manufacturing, logistics, or assistive home settings. Thus, the developed system serves as a versatile platform for advancing sEMG-driven interaction across diverse application domains, bridging human intention with robotic execution.

Chapter 8

Uncertainty-aware Bayesian Model Averaging-based Spatial Kernel Modeling in Orthogonal LPP

8.1 Introduction

Surface electromyography (sEMG) signals, which capture the electrical activity of muscles during voluntary contractions, are widely used in upper limb activity recognition tasks, for human-machine interface (HMI), prosthetic control, and rehabilitation. The ability to decode complex hand gestures from sEMG signals, enables seamless communication between human intention and assistive robotic systems. However, the inherently stochastic and subject-dependent nature of sEMG signals, coupled with variations due to sensor placement, muscle fatigue, and inter-subject anatomical differences, poses significant challenges for reliable gesture recognition [54]. To overcome these challenges, dimensionality reduction (DR) techniques are employed to project high-dimensional, redundant, and inconsistent EMG features into a compact and discriminative subspace that preserves essential gesture-relevant information [55]. Among various DR approaches, locality preserving projections (LPP) [7] offer a linear framework that approximates the underlying nonlinear manifold, effectively retaining the intrinsic geometric structure of the data. LPP preserves local neighborhood relations of data points during projection, making it highly effective in capturing subtle variations in biosignal patterns. A further advancement, orthogonal locality preserving projections (OLPP) [9], introduces orthogonality constraints on the basis vectors, enhancing the diversity and robustness of the projection space.

Despite these advantages, a key limitation of LPP and OLPP lies in their reliance

on Euclidean distance-based similarity kernels, which are sensitive to outliers and may inadequately capture nonlinear spatial or directional relations in the signal space. This motivates the development of enhanced similarity measures and fusion mechanisms that can robustly represent spatial affinity across multiple geometrical perspectives. Addressing the isuue, this chapter presents a novel orthogonal LPP-based framework termed *UaBMA-OLPP*, which integrates uncertainty-aware Bayesian model averaging (UaBMA) for fusing similarity kernels derived from Euclidean [7], Euler (cosine-based) [62], and Grassmannian spaces [63]. Each of these manifolds captures complementary aspects of signal similarity: spatial proximity, directional alignment, and subspace-level geometric variation, respectively. UaBMA adaptively weights the contribution of each kernel by modeling their posterior probability including uncertainty measures in terms of classification performance, thereby generating a robust and discriminative similarity matrix for OLPP. Extensive experimental results, both with laboratory-collected EMG data and benchmark databases, validate the superior performance of the proposed technique in recognizing diverse hand movement gestures.

Through this contribution, the chapter aims to bridge the gap between traditional dimensionality reduction and adaptive kernel fusion by incorporating uncertainty modeling and BMA, thereby advancing the robustness and reliability of sEMG-based activity recognition in real-world HMI environments.

8.2 Review of Existing Literature

The operational pipeline of EMG-based hand movement recognition frameworks can be broadly broken down into four sequential phases: (1) acquisition of raw signals, (2) signal preprocessing, (3) extraction of relevant features, and (4) classification of movements. Zandigohar *et al.* introduced a Bayesian evidence-based method to estimate grasp classification probabilities by integrating sEMG and vision-derived data in prosthetic hand systems [129]. In a related approach, grasp classification outputs from EMG signals were first determined and subsequently utilized within visual object detection modules [130]. Other researchers have explored domain adaptation methodologies to enhance EMG-based classification tasks [131], [132]. Du *et al.* applied domain adaptation to classify high-density sEMG signals across different recording sessions, wherein the signals were transformed into grayscale image

formats and adaptive batch normalization was used to boost the performance of unsupervised models [132]. Xu *et al.* developed a spatial sliding window approach aimed at detecting spike patterns in high-density sEMG data [133]. In another work, Batayneh *et al.* [134] proposed a generalized regression neural network (GRNN) to estimate joint angles based on wrist motion using the NinaPro dataset. Esa *et al.* incorporated time-domain statistical features such as root mean square (RMS), zero crossing (ZC), and slope sign change (SSC) into an SVM classifier to monitor individual finger movements [57]. A new method involving channel selection and time-domain feature fusion was proposed in [135], where an SVM-based framework was utilized to classify EMG signal patterns. Approaches based on the frequency domain have also gained attention in similar contexts. Alazrai *et al.* [136] utilized discrete wavelet transform (DWT) for feature extraction and employed support vector regression (SVR) for estimating finger and wrist joint angles. Taghizadeh *et al.* [137] used the fractional Fourier transform (FrFT) as the primary feature extractor and classified finger movements with a k-nearest neighbor (kNN) algorithm. Zhai *et al.* [138] applied fast Fourier transform (FFT)-based features to develop a neuroprosthesis control system.

Ilyas *et al.* [55] presented an EMG-based study, where statistical features were extracted and then reduced using techniques like principal component analysis (PCA), linear discriminant analysis (LDA), and non-negative matrix factorization (NMF) for estimating finger joint angles. Similarly, Anam *et al.* [139] utilized PCA to project statistical features into lower-dimensional spaces before performing finger joint regression. Tapia *et al.* [58] adopted independent component analysis (ICA) and empirical mode decomposition (EMD) to extract motor activation cues from EMG signals. Guo *et al.* proposed a multimodal hand motion recognition system that fused sEMG and mechanomyography (MMG) features using an LDA-based classification methodology [140]. Saikia *et al.* [141] applied PCA and LDA to represent EMG features in low-dimensional spaces, for recognizing prosthetic finger movements. A further improvement was proposed in [142], where uncorrelated LDA (ULDA) was used for EMG feature reduction. Li *et al.* [143] introduced a graph-based method to estimate wrist angles and grasping forces from high-density EMG signals using simultaneous and proportional mapping.

8.3 Proposed Framework

This section introduces a novel subspace learning methodology termed UaBMA-OLPP, which aims to minimize feature redundancy and boost the discriminative quality of sEMG signal representations in the projected domain. OLPP strengthens class separability by ensuring orthogonality among the projection vectors, thereby improving the representation capacity of features in the transformed space. UaBMA operates within this OLPP framework and integrates manifold-specific similarity information through a Bayesian probabilistic approach.

8.3.1 Orthogonal Locality Preserving Projection (OLPP)

Let the input training dataset be denoted as $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N] \in \mathbb{R}^{D\times N}$, where each $\mathbf{z}_k \in \mathbb{R}^D$ represents a D-dimensional data point, and N is the total number of samples. In the LPP framework, the projection matrix $\mathbf{W} = [\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_d] \in \mathbb{R}^{D\times d}$ is designed to generate lower-dimensional representations as $\mathbf{y}_k = \mathbf{W}^T\mathbf{z}_k \in \mathbb{R}^d$. While Section 3.2.3 of Chapter 2 has already presented a comprehensive discussion on the LPP formulation, including its mathematical basis and graph-based embedding strategy, the core mathematical expressions are briefly summarized here to ensure consistency in notation and ease of reference. The objective is to preserve local proximity between data points by minimizing the following expression:

$$\min_{\mathbf{W}} \sum_{i=1}^{N}\sum_{j=1}^{N} \|\mathbf{z}_i - \mathbf{z}_j\|^2 S_{ij} = \sum_{i=1}^{N}\sum_{\mathbf{z}_j\in\mathcal{R}_k(\mathbf{z}_i)} \|\mathbf{W}^T\mathbf{z}_i - \mathbf{W}^T\mathbf{z}_j\|^2 S_{ij} \tag{8.1}$$

Here, $S = [S_{ij}] \in \mathbb{R}^{N\times N}$ denotes a symmetric similarity matrix, where S_{ij} captures the affinity between $\mathbf{z}_i$ and $\mathbf{z}_j$, and $\mathcal{R}_k(\mathbf{z}_i)$ represents the k-nearest neighbors of $\mathbf{z}_i$. This minimization is equivalent to the following trace minimization problem:

$$\min_{\mathbf{W}} \operatorname{tr}(\mathbf{W}^T\mathbf{Z}(\boldsymbol{\Delta} - \mathbf{S})\mathbf{Z}^T\mathbf{W}) \quad \text{s.t.} \quad \mathbf{W}^T\mathbf{Z}\boldsymbol{\Delta}\mathbf{Z}^T\mathbf{W} = \mathbf{1} \tag{8.2}$$

The constraint is introduced to avoid degenerate solutions. Here, $\boldsymbol{\Delta} = \operatorname{diag}[\Delta_{ii}]$ is a diagonal matrix with entries $\Delta_{ii} = \sum_{j=1}^{N} S_{ij}$, and $(\boldsymbol{\Delta} - \mathbf{S})$ corresponds to the graph Laplacian. Typically, $\mathbf{S}$ is computed using a Gaussian kernel over Euclidean distances:

$$S_{ij} = f_1(\mathbf{z}_i, \mathbf{z}_j) = \begin{cases} \exp\left(-\frac{\|\mathbf{z}_i - \mathbf{z}_j\|^2}{\tau}\right), & \mathbf{z}_j \in \mathcal{R}_k(\mathbf{z}_i) \\ 0, & \text{otherwise} \end{cases} \tag{8.3}$$

The kernel parameter τ controls the scale of similarity. Minimizing the objective ensures that in the reduced space, the similarity between projected points $\mathbf{y}_i$ and $\mathbf{y}_j$ matches with their high-dimensional counterparts, as well. This leads to the generalized eigenvalue problem:

$$Z(\mathbf{\Delta} - \mathbf{S})\mathbf{Z}^T\mathbf{W} = \lambda \mathbf{Z}\mathbf{\Delta}\mathbf{Z}^T\mathbf{W} \tag{8.4}$$

Assuming $\mathbf{Z}\mathbf{\Delta}\mathbf{Z}^T$ is invertible, it can simplified as:

$$(\mathbf{Z}\mathbf{\Delta}\mathbf{Z}^T)^{-1}\mathbf{Z}(\mathbf{\Delta} - \mathbf{S})\mathbf{Z}^T\mathbf{W} = \lambda \mathbf{W} \tag{8.5}$$

By solving this, the eigenvalues λ_i and the corresponding eigenvectors $\boldsymbol{\omega}_i$ are obtained, which are then used to construct the projection matrix $\mathbf{W}$.

In OLPP, the dataset $\mathbf{Z}$ is first transformed using PCA to remove null-space components. The PCA matrix $\mathbf{W}_{\text{PCA}}$ ensures uncorrelated projected features and maintains matrix rank. Here, the orthogonal basis is defined as:

$$\mathbf{W}^{(k-1)} = [\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_{k-1}] \tag{8.6}$$

$$\mathbf{B}^{(k-1)} = [\mathbf{W}^{(k-1)}]^T(\mathbf{Z}\mathbf{\Delta}\mathbf{Z}^T)^{-1}\mathbf{W}^{(k-1)} \tag{8.7}$$

The first orthogonal basis vector $\hat{\boldsymbol{\omega}}_1$ is computed from the minimum eigenvalue of $(\mathbf{Z}\mathbf{\Delta}\mathbf{Z}^T)^{-1}\mathbf{Z}(\mathbf{\Delta} - \mathbf{S})\mathbf{Z}^T$. Subsequent vectors $\hat{\boldsymbol{\omega}}_k$ $(k = 2, 3, \ldots, d)$ are derived using:

$$\mathbf{M}^{(k)} = \left\{ I - (\mathbf{Z}\mathbf{\Delta}\mathbf{Z}^T)^{-1}\mathbf{W}^{(k-1)}[\mathbf{B}^{(k-1)}]^{-1}[\mathbf{W}^{(k-1)}]^T \right\} (\mathbf{Z}\mathbf{\Delta}\mathbf{Z}^T)^{-1}\mathbf{Z}(\mathbf{\Delta} - \mathbf{S})\mathbf{Z}^T \tag{8.8}$$

Eigenvectors of $\mathbf{M}^{(k)}$ yield the orthogonal projection matrix $\hat{\mathbf{W}} = [\hat{\boldsymbol{\omega}}_1, \hat{\boldsymbol{\omega}}_2, \ldots, \hat{\boldsymbol{\omega}}_d]$, and the final projection matrix becomes $\mathbf{W}_{\text{OLPP}} = \mathbf{W}_{\text{PCA}}\hat{\mathbf{W}}$. The reduced feature matrix is then given by $\mathbf{Y} = \mathbf{W}_{\text{OLPP}}^T\mathbf{Z}$.

8.3.2 Similarity Kernels across Multiple Manifold Spaces

To determine an effective set of projection vectors $\boldsymbol{\omega}$ and ultimately obtain an informative low-dimensional embedding $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N] \in \mathbb{R}^{d \times N}$, selecting an appropriate similarity matrix $\mathbf{S}$ is of utmost importance. This matrix governs how local structural relationships among samples are retained during projection, typically by encouraging nearby points to remain close and distancing dissimilar points. This work employs three distinct similarity kernels constructed from different underlying manifolds: Euclidean space, Euler space, and the Grassmann manifold. These representations capture complementary aspects of spatial,

directional, and subspace-level affinities among EMG signal vectors. These are explained below in detail.

1) Euclidean Domain: In most conventional approaches, similarity matrices are computed using Euclidean distances. As previously discussed, the spatial distance between $\mathbf{z}_i$ and $\mathbf{z}_j$ in the Euclidean space is defined as:

$$f_1(\mathbf{z}_i, \mathbf{z}_j) = \|\mathbf{z}_i - \mathbf{z}_j\|_2^2 \tag{8.9}$$

This kernel emphasizes spatial proximity in the Euclidean domain and is typically applied to the k-nearest neighbors in LPP-based formulations. It captures the squared ℓ_2 norm between two data points, which is effective for measuring distances in flat, linear spaces. However, this measure can be sensitive to variations in scale and noise, and may fail to capture the underlying geometric structure when the data resides on a curved or nonlinear manifold. To overcome these limitations, alternative similarity formulations have been explored.

2) Euler Domain: Euclidean-based formulations tend to be sensitive to data irregularities and noise [62]. In these scenarios, incorporating directional information becomes crucial to differentiate between classes more effectively while mitigating the influence of outliers. With sample vectors $\mathbf{z}_i, \mathbf{z}_j \in \mathbb{R}^D$, in the Euler domain, the cosine-based similarity is defined as:

$$f_2(\mathbf{z}_i, \mathbf{z}_j) = D - \sum_{p=1}^{D} \cos\left(\nu\pi(\hat{\mathbf{z}}_i(p) - \hat{\mathbf{z}}_j(p))\right) \tag{8.10}$$

Here, $\hat{\mathbf{z}}_i$ and $\hat{\mathbf{z}}_j$ are ℓ_2-normalized representations of the respective vectors, with their p-th components denoted as $\hat{\mathbf{z}}_i(p)$ and $\hat{\mathbf{z}}_j(p)$. The scalar $\nu > 0$ controls the frequency scaling in the cosine operation and modulates the impact of outlier-induced large distances [144]. Functionally, this formulation corresponds to Andrew's M-estimate, which is known for its robust correlation scheme in matching problems [62]. From a complex number standpoint, Equation (8.10) can be equivalently rewritten as:

$$f_2(\mathbf{z}_i, \mathbf{z}_j) = \left\|\mathbf{c}_i - \mathbf{c}_j\right\|^2 \tag{8.11}$$

where the complex-valued vectors $\mathbf{c}_i$ and $\mathbf{c}_j$ are defined as:

$$\mathbf{c}_k = \frac{1}{\sqrt{2}} \exp(i\nu\pi\hat{\mathbf{z}}_k)$$

This transformation embeds each normalized vector into the Euler space, allowing the Euler-based similarity to be interpreted as the squared Euclidean distance between these complex embeddings.

3) Grassmannian Domain: While the Euclidean space provides a flat, coordinate-based geometry, suitable for conventional data structures, it does not capture the curvature inherent in more complex manifolds. To extend the LPP graph construction to a Riemannian context, data points are embedded into symmetric matrix spaces via the Grassmann manifold, which is well-suited to preserve local relationships in higher-order geometries [145]. Given $\mathbf{z}_i$ and $\mathbf{z}_j$ in $\mathbb{R}^D$, the corresponding Grassmann manifold $\mathcal{G}(p,D)$ is defined as the space containing all p-dimensional subspaces embedded within $\mathbb{R}^D$, where $0 \leq p \leq D$. For $p = 0$, the manifold degenerates to the Euclidean space, while $p = 1$ represents the set of all lines through the origin. A mapping from the Grassmann manifold to *symmetric matrix space* (SMS) is given as:

$$\Omega : \mathcal{G}(p,D) \rightarrow \text{SMS}(D), \quad \Omega(\mathbf{z}_i) = \mathbf{z}_i \mathbf{z}_i^T \tag{8.12}$$

The corresponding Grassmannian distance between $\mathbf{z}_i$ and $\mathbf{z}_j$ is then computed as:

$$f_3(\mathbf{z}_i, \mathbf{z}_j) = \left\{ \frac{1}{2} \left\| \Omega(\mathbf{z}_i) - \Omega(\mathbf{z}_j) \right\|_2^2 \right\}^{1/2} \tag{8.13}$$

To provide visual insight into how each kernel models similarity information, 50 EMG samples are selected at random, and their respective weight matrices; computed using Euclidean, Euler, and Grassmannian measures, are illustrated in Fig. 8.1.

8.3.3 Bayesian Model Averaging with Uncertainty Measure

Bayesian model averaging (BMA) serves as an effective strategy for ensemble learning, where predictions from multiple models are combined to enhance classification accuracy. It offers a principled probabilistic framework that assigns weights to each model's output based on how likely it is, given the observed data. Let $\mathcal{D}$ denote the available dataset, and let $\{H_1, H_2, \ldots, H_n\}$ represent a set of candidate models. The posterior probability of a specific model H_i given the dataset $\mathcal{D}$ is computed using Bayes' theorem as:

$$P(H_i|\mathcal{D}) = \frac{P(\mathcal{D}|H_i) \cdot P(H_i)}{P(\mathcal{D})} \tag{8.14}$$

Here, $P(\mathcal{D}|H_i)$ is the likelihood of the data under model H_i, and $P(H_i)$ is the prior belief in that model, while $P(\mathcal{D})$ is the marginal likelihood used for normalization. Suppose each model H_i yields a prediction $\hat{y}_i$ for a data instance. Then, the overall BMA prediction is a weighted sum

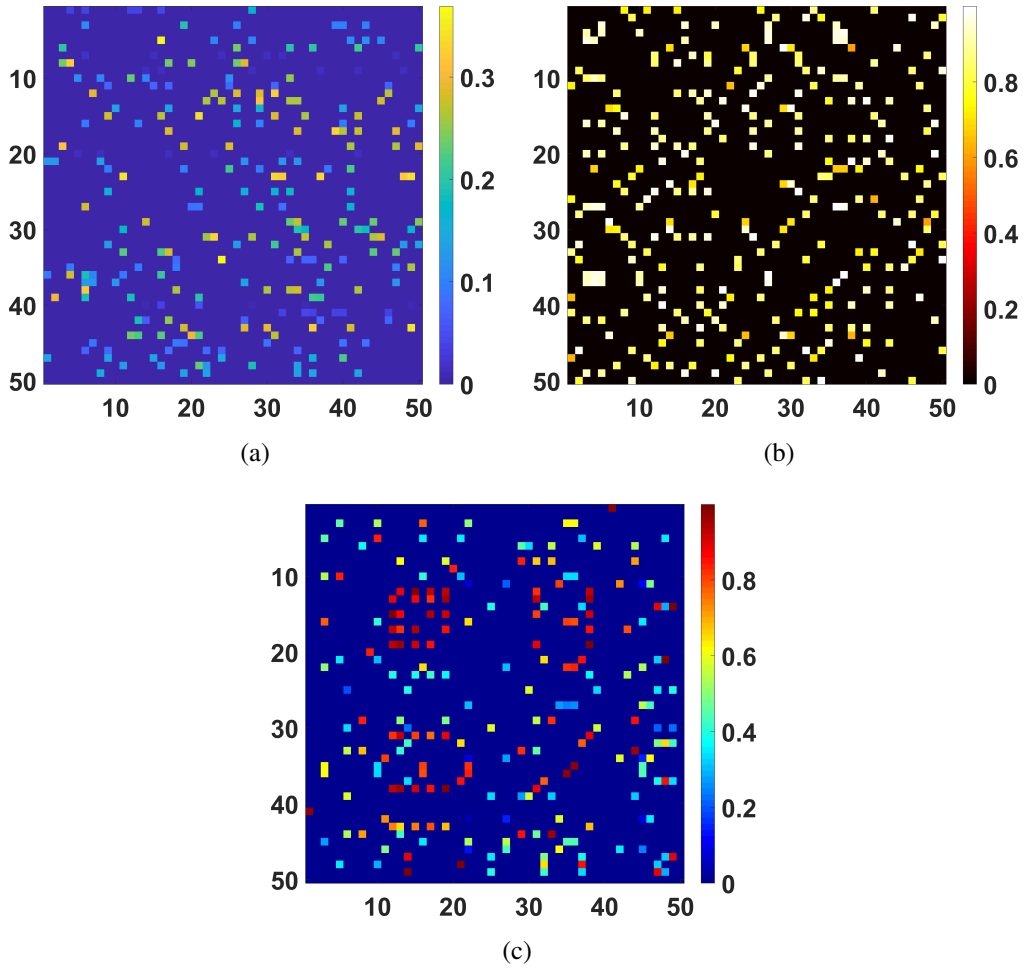


Figure 8.1: Visualization of similarity matrices derived from 50 randomly selected EMG samples using three different geometrical representations: (a) Euclidean-based kernel, (b) Euler domain kernel, and (c) Grassmannian manifold kernel.

of these predictions:

$$\hat{y}_{\text{final}} = \sum_{i=1}^{n} P(H_i|\mathcal{D}) \cdot \hat{y}_i \tag{8.15}$$

In the present study, the models H_1, H_2, and H_3 correspond to the similarity kernels derived from Euclidean, Euler, and Grassmannian domains, respectively. Their classification accuracies, denoted as α_1, α_2, and α_3, are used to estimate the likelihoods $P(\mathcal{D}|H_i)$. Assuming all models are equally probable a priori, the posterior reduces proportionally to the likelihood:

$$\hat{y}_{\text{final}} \propto \sum_{i=1}^{n} P(\mathcal{D}|H_i) \cdot \hat{y}_i \tag{8.16}$$

Algorithm 8.1 Uncertainty-aware Bayesian Model Averaging-based OLPP (UaBMA-OLPP)

Input: EMG dataset $\mathbf{Z}$, Gaussian kernel parameters τ_1, τ_2, τ_3, uncertainty parameters μ_i, σ_i for each model H_i

Output: Low-dimensional embedded feature matrix $\mathbf{Y}$

STEP-1 Apply PCA to the input dataset $\mathbf{Z}$ to eliminate null-space components and retain meaningful variation.

STEP-2 Construct the spatial similarity matrix $\mathbf{S}_{f_1}$ using the Euclidean kernel with parameter τ_1.

STEP-3 Compute the Euler similarity matrix $\mathbf{S}_{f_2}$ using complex cosine-based embedding with kernel parameter τ_2.

STEP-4 Derive the Grassmannian similarity matrix $\mathbf{S}_{f_3}$ using the subspace projection kernel with parameter τ_3.

STEP-5 Estimate the likelihood values $P(\alpha|H_i)$ using a Gaussian distribution $\mathcal{N}(\mu_i, \sigma_i^2)$ for each model's observed accuracy α, using Equation (8.18).

STEP-6 Compute the weighted posterior $P(H_i|\alpha)$ for each model using the Bayesian inference given in Equation (8.14).

STEP-7 Generate the final fused similarity matrix as a weighted combination: $\hat{\mathbf{S}} = \sum_{i=1}^{3} P(\alpha|H_i) \cdot \mathbf{S}_{f_i}$.

STEP-8 Apply OLPP using the fused similarity matrix $\hat{\mathbf{S}}$ to compute the projection matrix $\mathbf{W}_{\text{OLPP}}$.

STEP-9 Project the dataset: $\mathbf{Y} = \mathbf{W}_{\text{OLPP}}^T \mathbf{Z}$.

STEP-10 Return the reduced-dimensional feature matrix $\mathbf{Y}$.

Now, to construct a fused similarity matrix $\hat{\mathbf{S}}$, the BMA formulation is applied to the three kernels as follows:

$$\hat{\mathbf{S}} = P(\alpha|H_1) \cdot \mathbf{S}_{f_1} + P(\alpha|H_2) \cdot \mathbf{S}_{f_2} + P(\alpha|H_3) \cdot \mathbf{S}_{f_3} \tag{8.17}$$

Here, $P(\alpha|H_i)$ denotes the likelihood of achieving a specific classification accuracy α under model H_i. The primary goal of Bayesian model averaging (BMA) is to generate a weighted combination of multiple models in a way that enhances the overall classification performance of the ensemble.

To achieve this, the likelihoods corresponding to accuracy values from different models, denoted as $P(\alpha|H_i)$, are estimated using a framework that accounts for uncertainty in performance. This uncertainty-aware modeling approach evaluates the confidence associated

with each model's classification accuracy (or other relevant performance indicators) during the integration process, such as when aggregating similarity weight kernels. Instead of considering accuracy as a fixed scalar value, the method incorporates uncertainty by treating it as a random variable, allowing for a more realistic representation of potential fluctuations in performance. The effectiveness of each similarity kernel, representing a Bayesian hypothesis model, is not treated deterministically but is acknowledged as being susceptible to variation. Such variation may arise due to factors like limited training data, the presence of noise, or differences introduced by cross-validation procedures. To reflect this uncertainty, accuracy metrics are modeled as stochastic variables rather than constants. To capture the inherent variability in performance, statistical distributions, such as the *beta* or Gaussian distribution can be employed. In this work, the uncertainty around model performance is captured using the normal distribution, which serves as the basis for computing likelihoods associated with each accuracy score, as detailed below.

$$P(\alpha|H_i) \propto P(\alpha) \sim \mathcal{N}(\mu_i, \sigma_i^2) \tag{8.18}$$

where μ_i and σ_i denote the empirical mean and standard deviation of the accuracy for model H_i. By incorporating these distributions into the likelihood estimation, the BMA framework becomes sensitive to confidence levels associated with each kernel. This uncertainty-aware formulation helps to account for performance variability, making the combined similarity kernel $\hat{\mathbf{S}}$ more robust and reliable. For a clearer understanding, the comprehensive steps of the UaBMA-OLPP approach are detailed in Algorithm 8.1.

8.4 Experimental Results and Discussion

During the experimentation phase, sEMG signals have been collected from a total of 25 volunteers. Each participant has portrayed 10 distinct hand gestures across two separate recording sessions, as previously mentioned in Section 7.4 of Chapter 7. The entire dataset has been partitioned randomly using a 70 : 30 train-test split. Using the training set, the similarity matrices $\mathbf{S}_{f_1}$, $\mathbf{S}_{f_2}$, and $\mathbf{S}_{f_3}$, as well as the aggregated kernel $\hat{\mathbf{S}}$, have been computed based on the proposed UaBMA scheme. The orthogonal basis vectors have then been derived through UaBMA-OLPP, to perform the data projection. Fig. 8.2 displays the first six principal orthogonal basis vectors obtained using UaBMA-OLPP. These eigenvectors have been arranged

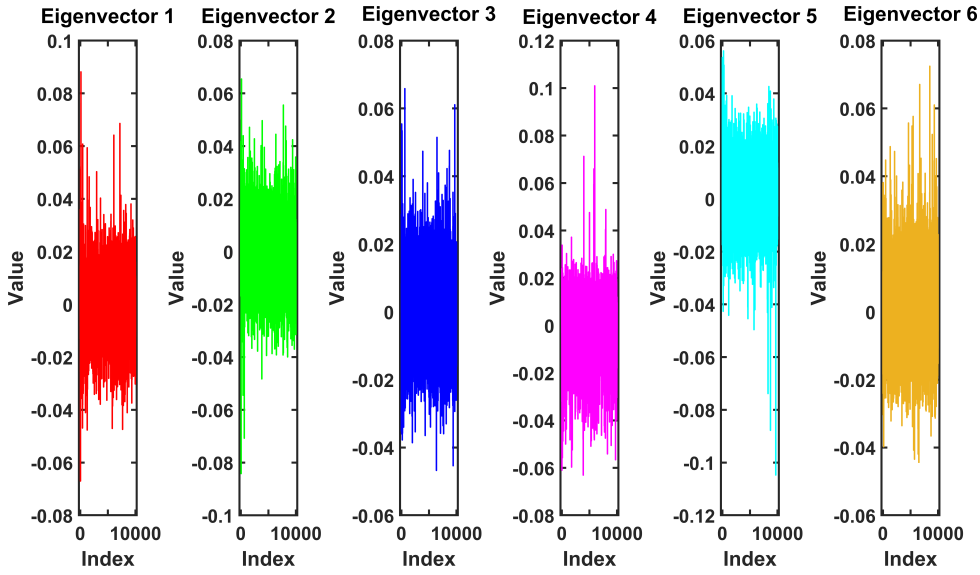


Figure 8.2: Visualization of the six most significant eigenvectors derived using the proposed UaBMA-OLPP method for dimensionality reduction.

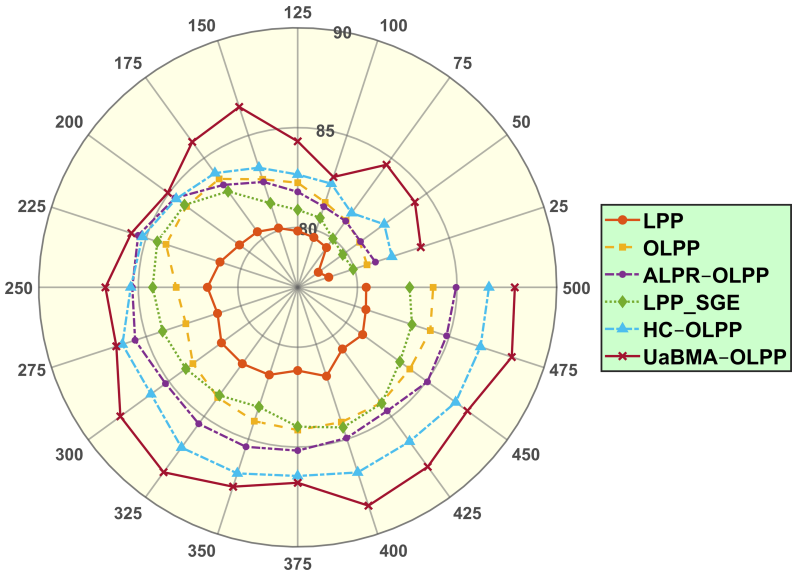
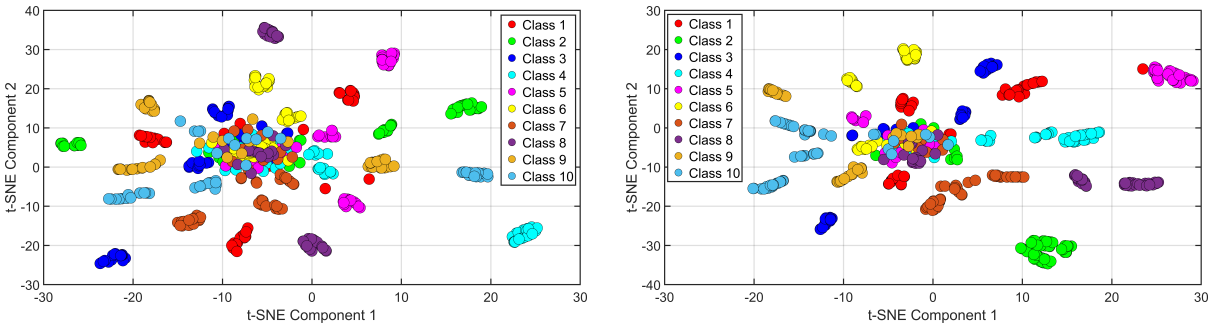


Figure 8.3: Polar plot illustrating the performance of different dimensionality reduction methods, where the angular axis indicates the number of eigenvectors and the radial axis corresponds to classification accuracy.

in order of their energy significance, highlighting the directions that most effectively preserve the intrinsic local structure of the dataset.

Determining the optimal number of eigenvectors for maximizing classification performance is non-trivial, as it involves a balance between two conflicting factors. Utilizing too few eigenvectors may result in the loss of critical information, while using too many can lead to overfitting and increased complexity due to the curse of dimensionality. To examine



(a) (b)

Figure 8.4: Two-dimensional t-SNE visualization of EMG data: (a) Representation in the original high-dimensional feature space, and (b) Corresponding distribution in the low-dimensional embedded space after projection.

this trade-off, a comparative study has been conducted involving the proposed UaBMA-OLPP and existing approaches such as LPP [7], OLPP [9], ALPR-OLPP [70], LPP_SGE [146], and HC-OLPP [16]. The variations in classification accuracy with respect to the number of principal eigenvectors are illustrated in Fig. 8.3 using a polar plot. This plot reveals distinct optimal dimensionalities for different DR methods, indicating diverse mechanisms of local structure preservation. The EMG feature spaces before and after projection are visualized using t-distributed stochastic neighbor embedding (t-SNE) in Fig. 8.4. It can be observed that the UaBMA-OLPP projection enhances both discriminability and class separability.

The low-dimensional features have been classified using both KNN and SVM classifiers. While both models have achieved commendable results, SVM has slightly surpassed KNN in terms of accuracy. Therefore, subsequent performance analyses presented in this chapter are based on the SVM classifier. Table 8.1 summarizes the classification accuracy, sensitivity, precision, and F1-score for all evaluated techniques. The property of orthogonality of the basis vectors has significantly contributed to performance improvements, and further advancement has been achieved through the fusion of multiple kernel weights. Table 8.2 reports the error metrics including the false positive rate (FPR), false rejection rate (FRR), equal error rate (EER), and false discovery rate (FDR). The proposed approach exhibits the lowest EER among the competitors, indicating a more balanced classifier with reduced misclassification for both positive and negative instances. Fig. 8.5 presents the confusion matrix for the proposed UaBMA-OLPP followed by SVM classification, further illustrating its robust performance across individual classes.

Table 8.1: COMPARISON OF CLASSIFICATION PERFORMANCE OF VARIOUS DIMENSIONALITY REDUCTION TECHNIQUES USING ACCURACY, SENSITIVITY, PRECISION, AND F1-SCORE

Method	Accuracy	Sensitivity	Precision	F1-Score
LPP	81.73	81.60	82.44	82.02
OLPP	84.15	84.08	84.61	84.34
ALPR-LPP	82.67	82.54	83.35	82.94
ALPR-OLPP	85.58	85.41	86.18	85.79
LPP_SGE	84.41	84.22	84.91	84.56
HC-LPP	85.36	85.29	85.86	85.57
HC-OLPP	87.27	87.13	87.77	87.45
UaBMA-OLPP	**88.50**	**88.42**	**89.18**	**88.79**

Table 8.2: COMPARISON OF CLASSIFICATION PERFORMANCE OF VARIOUS DIMENSIONALITY REDUCTION TECHNIQUES IN TERMS OF ERROR METRICS FPR, FRR, EER, AND FDR

Method	FPR	FRR	EER	FDR
LPP	10.02	18.40	11.65	17.57
OLPP	8.14	15.92	10.26	15.39
ALPR-LPP	9.93	17.46	10.72	16.64
ALPR-OLPP	6.61	14.59	7.15	13.81
LPP_SGE	7.77	15.78	9.71	15.09
HC-LPP	6.65	14.71	7.39	14.14
HC-OLPP	4.43	12.87	4.72	12.23
UaBMA-OLPP	**3.27**	**11.58**	**3.56**	**10.82**

Moreover, performance of the proposed technique has been further evaluated using three subsets of the publicly available NinaPro DB1 dataset, comprising recordings from 27 subjects [147], [128]. The NinaPro subsets DB1-E1, DB1-E2, and DB1-E3 consist of 12 unique finger movement tasks, 17 varied isometric and isotonic hand postures, and 23 different grasping and

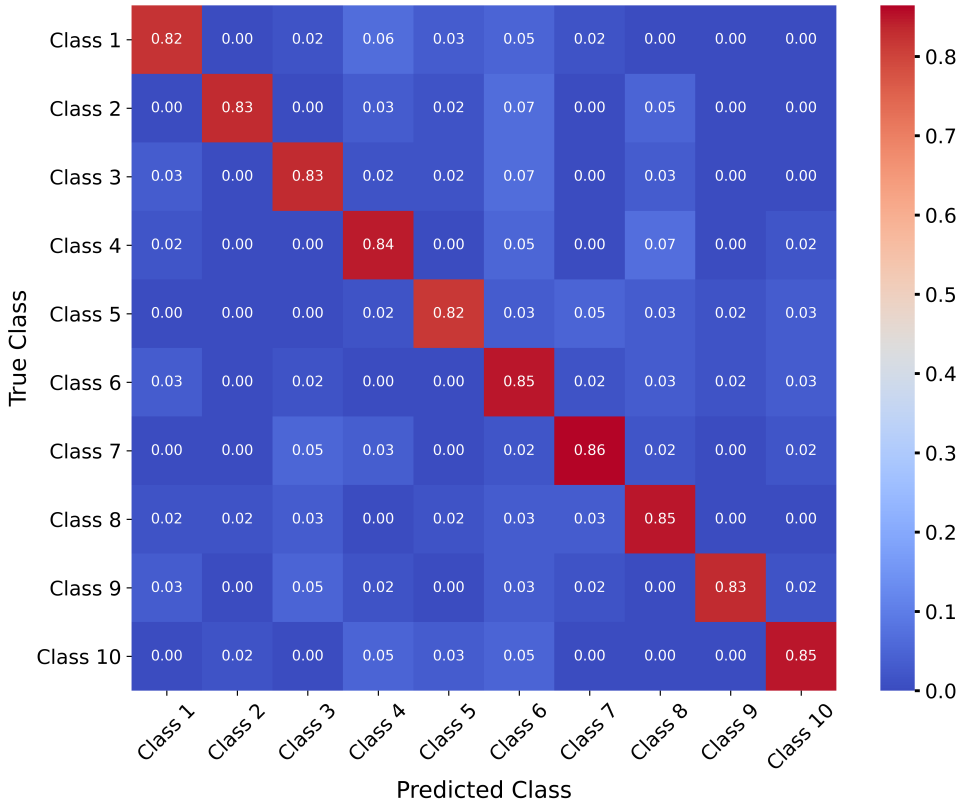


Figure 8.5: Confusion matrix resulting from SVM classification using features extracted via the proposed UaBMA-OLPP dimensionality reduction method.

functional motion exercises, respectively. A comparative study of the proposed method against several state-of-the-art dimensionality reduction techniques is reported in Table 8.3. It can be observed that UaBMA-OLPP has consistently outperformed all competing methods across all four evaluated datasets.

8.5 Summary

This study introduced a novel manifold-inspired framework for recognizing hand movements using surface EMG signals. The experimental findings across different categories of gestures and grasping motions validate the efficiency of the proposed method. The broader applicability of such a system is evident in two major domains: (i) intelligent healthcare, prosthetic systems, and rehabilitation technologies, and (ii) robotic systems, particularly in translating human hand gestures into robotic commands for executing complex or heavy-duty operations. The approach builds upon an augmented version of orthogonal locality preserving projections

Table 8.3: RECOGNITION ACCURACY ACROSS NINAPRO DB1 SUBSETS AND FULL DATASET

Method	DB1-E1	DB1-E2	DB1-E3	DB1 Overall
LPP	86.12	85.23	82.24	83.61
OLPP	91.31	91.63	88.15	86.54
ALPR-LPP	89.43	88.56	87.33	85.25
ALPR-OLPP	93.79	92.60	90.23	88.12
LPP_SGE	92.22	92.74	89.83	89.33
HC-LPP	94.16	93.85	90.05	91.90
HC-OLPP	95.89	95.52	92.45	93.10
UaBMA-OLPP	**97.18**	**96.68**	**94.47**	**94.75**

by incorporating a Bayesian model averaging strategy, to fuse multiple similarity kernels derived from diverse geometric representations. A fundamental principle behind LPP-based dimensionality reduction techniques lies in preserving local neighborhood relationships within high-dimensional feature spaces. Identifying the most appropriate geometric domain for effectively capturing this locality information has remained a pertinent research topic. This work attempted to bridge existing research gaps by proposing an adaptive, uncertainty-aware similarity modeling scheme that integrates contributions from both Euclidean and Riemannian geometrical frameworks.

Unlike the relatively straightforward distance computations in Euclidean spaces, measuring similarities within Riemannian manifolds, often involves complex and non-trivial formulations. Numerous advanced distance learning methods in Riemannian geometry exist and may offer more effective means of uncovering the intrinsic structure of data distributions. Exploring such sophisticated similarity metrics in manifold spaces presents a promising future direction, as highlighted by the insights gained from this research. Furthermore, there remains a need for developing more robust techniques, capable of maintaining reliable performance under challenging conditions—such as in the presence of sensor noise or variability in signal acquisition.

Chapter 9

Robust Possibilistic Neighborhood-based Spatial Graph Modeling in Orthogonal LPP

9.1 Introduction

Surface electromyography (sEMG) signals, which capture muscle activity through non-invasive electrodes, have gained considerable attention in recent years for their utility in prosthetic hand control, motor rehabilitation, and human-machine interface (HMI). These signals encapsulate neuromuscular dynamics and are capable of providing insight into voluntary motion intentions, enabling intuitive gesture-based control for assistive technologies. However, the intrinsic variability of sEMG, caused by inter-subject physiological differences, electrode displacement, muscle fatigue, and sensor-induced noise; poses substantial challenges to the reliability of gesture recognition systems [54]. Moreover, the high dimensionality of sEMG features often hampers real-time processing and robust classification, thereby necessitating effective dimensionality reduction (DR) operations.

Locality Preserving Projections (LPP) and its orthogonal variant (OLPP) have been extensively adopted for their ability to capture the intrinsic local geometric structure of data, thereby offering richer representation in reduced-dimensional subspaces. OLPP enforces orthogonality among the projection vectors, ensuring linear independence and reducing feature redundancy, which leads to improved numerical stability and more interpretable subspace representations [9]. However, the construction of its similarity graph traditionally relies on Euclidean distance-based kernels. These similarity measures are particularly sensitive to noise and outliers, often degrading the projection quality in non-ideal settings such as real-world

EMG recording environments.

To overcome these limitations, this chapter introduces a robust and adaptive spatial graph learning technique named *Robust Possibilistic Neighborhood Graph* (RPNG) inside the framework of OLPP. The proposed method augments OLPP by integrating an adaptive graph formulation scheme, which makes use of possibilistic similarity learning rather than probabilistic affinity assumptions. By eliminating the sum-to-one constraint imposed in conventional adaptive neighbor graphs, the model avoids bias in neighborhood relationships caused by noisy observations. Furthermore, a row-sparsity inducing $\ell_{2,1}$ norm constraint [70] is imposed on the projection matrix to automatically suppress irrelevant or weakly discriminative features, enhancing both robustness and interpretability.

This robust DR framework is evaluated using a customized, laboratory-acquired sEMG dataset under both normal and noisy conditions and benchmark NinaPro data. Comparative analyses with multiple baseline DR techniques demonstrate the superior recognition performance and resilience of the RPNG-OLPP approach. The method holds significant potential for real-world applications in prosthetic limb control, rehabilitation robotics, and gesture-driven HMI systems, particularly under adverse signal acquisition conditions.

9.2 Background and Motivation

Adaptive graph construction techniques have been extensively explored in recent years to improve neighborhood-based modeling in unsupervised and semi-supervised learning tasks. Among these, the Adaptive Neighbor Graph (ANG) model [59] presents an important advancement by learning data similarity and neighborhood relationships in a unified framework. In ANG, the similarity between data points is computed based on their local distances, with the constraint that the weights assigned to neighbors must sum to one. Although this probabilistic formulation ensures normalized affinities, it introduces a significant limitation—outliers or noisy samples may inadvertently receive high similarity scores due to the sum-to-one constraint, thus compromising graph fidelity.

To overcome this issue, the Possibilistic Neighborhood Graph (PNG) model [60] eliminates the normalization constraint and instead introduces a possibilistic regularization mechanism. This shift allows each data point to determine its neighbors independently

without being influenced by the presence of noise or outliers, thereby resulting in more reliable similarity estimation. Despite this improvement, PNG does not incorporate an explicit mechanism for enforcing sparsity in the learned representation. Consequently, the projection subspace may still be contaminated with redundant or weakly informative features, reducing both efficiency and discriminative power [148]. Another limitation of PNG, lies in its inability to account for feature relevance during projection. All features are treated equally, which limits the model's ability to emphasize more discriminative components. As a result, there is no interpretability-driven constraint to guide the selection of informative features, which is essential for robust and transparent decision-making.

To address these gaps, this work introduces the Robust Possibilistic Neighborhood Graph (RPNG) model, which incorporates an $\ell_{2,1}$ norm-based regularization on the projection matrix. This formulation enforces row-wise sparsity, enabling the model to suppress the contribution of irrelevant or noisy features while retaining the most informative ones [70]. The minimization of the $\ell_{2,1}$ norm results in the shrinking of rows associated with less useful features, effectively filtering them out from the learned representation. This not only enhances the robustness of the graph construction process but also improves interpretability, as the non-zero rows of the projection matrix correspond directly to features with high discriminative importance.

The integration of this sparsity-inducing mechanism into the possibilistic graph learning framework marks a significant step toward adaptive, noise-resilient, and semantically meaningful dimensionality reduction; particularly suited for high-variability data such as surface EMG signals used in human-machine interface scenarios.

9.3 Proposed Methodology

The proposed RPNG-OLPP framework is designed to achieve robust subspace learning and neighborhood graph construction for high-dimensional and noise-prone datasets such as surface EMG signals. The methodology combines the strengths of Orthogonal Locality Preserving Projection (OLPP) and Robust Possibilistic Neighborhood Graph (RPNG) to address key challenges in feature redundancy, noise sensitivity, and local structure preservation. In this scheme, OLPP is employed to obtain orthogonal projection vectors that preserve the intrinsic local geometry of the data, while ensuring linear independence and reducing redundancy

among the features in the projected space. Simultaneously, the RPNG model constructs an adaptive similarity graph by incorporating possibilistic weighting and sparsity constraints, thereby enhancing resilience against outliers and irrelevant features. The following subsections present detailed descriptions of the OLPP-based projection and the RPNG-driven similarity graph learning, which together form the core of the RPNG-OLPP dimensionality reduction approach.

9.3.1 OLPP-Based Low-Dimensional Projection

Orthogonal Locality Preserving Projection (OLPP) is a linear manifold-inspired learning technique that aims to retain the local neighborhood structure of high-dimensional data while ensuring orthogonality among the projection vectors. This orthogonality facilitates decorrelated feature extraction and enhances numerical stability. In this chapter, OLPP serves as the foundational projection scheme onto which the robust similarity structure, learned via the proposed RPNG model is applied. A comprehensive explanation of the OLPP formulation, including its mathematical derivation and graph-based embedding mechanism, has already been provided in Section 8.3.1 of Chapter 8. However, for consistency of notation and ease of reference, the essential mathematical expressions associated with OLPP are briefly outlined here. Let the input dataset be denoted as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m] \in \mathbb{R}^{D\times m}$, where each $\mathbf{x}_i \in \mathbb{R}^D$ represents a high-dimensional data sample. The goal of OLPP is to compute a projection matrix $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_d] \in \mathbb{R}^{D\times d}$ such that the projected data $\mathbf{U} = \mathbf{P}^T\mathbf{X}$ preserves the local geometric structure encoded in a similarity graph. The locality preserving objective function is formulated by minimizing the weighted pairwise distance:

$$\sum_{i=1}^{m} \sum_{\mathbf{x}_j \in \mathcal{N}_k(\mathbf{x}_i)} \|\mathbf{P}^T\mathbf{x}_i - \mathbf{P}^T\mathbf{x}_j\|_2^2 \cdot \omega_{ij} \tag{9.1}$$

where ω_{ij} denotes the similarity weight between $\mathbf{x}_i$ and $\mathbf{x}_j$, and $\mathcal{N}_k(\mathbf{x}_i)$ represents the k-nearest neighbors of $\mathbf{x}_i$. The above objective can be reformulated in matrix form as:

$$\min_{\mathbf{P}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{P}=1} \operatorname{tr}(\mathbf{P}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{P}) \tag{9.2}$$

where $\mathbf{L} = \mathbf{\Lambda} - \mathbf{\Omega}$ is the graph Laplacian, $\mathbf{\Omega} = [\omega_{ij}]$ is the similarity matrix, and $\mathbf{\Lambda}$ is a diagonal matrix with entries $\Lambda_{ii} = \sum_j \omega_{ij}$. In its standard formulation, the similarity matrix $\mathbf{\Omega}$ is constructed using a Gaussian kernel that relies entirely on the Euclidean distance between

pairs of data points, and is defined as:

$$\omega_{ij} = \text{dist}(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} \exp\left(-\dfrac{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{\xi}\right), & \text{if } \mathbf{x}_j \in \mathcal{N}_k(\mathbf{x}_i) \\ 0, & \text{otherwise} \end{cases} \tag{9.3}$$

The parameter ξ significantly influences the strength of the similarity weights by controlling how strongly nearby points are connected. Given the matrix $\mathbf{\Omega}$, minimizing the objective function defined in Equation (9.1), ensures that the projected points $\mathbf{u}_i = \mathbf{P}^T\mathbf{x}_i$ and $\mathbf{u}_j = \mathbf{P}^T\mathbf{x}_j$ maintain their relative proximity in the embedded space, preserving the neighborhood structure of the original high-dimensional data. The remaining details of the orthogonal basis vector formulation can be found in Section 8.3.1 of Chapter 8.

These core notations and formulations are directly leveraged in the RPNG-based graph modeling introduced in the subsequent subsection.

9.3.2 Robust Possibilistic Neighborhood Graph

Conventional dimensionality reduction techniques that rely on graph-based models often use Euclidean distance to compute similarity. However, such metrics are particularly susceptible to noise and outliers, which is a common issue in complex biomedical signals like EMG that are inherently nonstationary and contain multiple components. This makes it challenging to effectively capture the underlying temporal and spectral variations of the signal using static distance-based similarity measures. To address these limitations, adaptive methods that learn the similarity graph from the data have become more favorable. In the adaptive neighborhood graph (ANG) approach, similarity between data samples is modeled as a probability distribution based on their distance in a lower-dimensional space. The corresponding optimization problem is formulated as [59]:

$$\begin{aligned} \min_{\mathbf{\Omega},\mathbf{P}} \quad & \sum_i \sum_j \omega_{ij} \|\mathbf{P}^T\mathbf{x}_i - \mathbf{P}^T\mathbf{x}_j\|_2^2 + \alpha \cdot \mathcal{T}(\mathbf{\Omega}) \\ \text{s.t.} \quad & 0 \le \omega_{ij} \le 1, \quad \boldsymbol{\omega}_i \mathbf{1}^T = 1, \quad \mathbf{P}^T\mathbf{X}\mathbf{M}\mathbf{X}^T\mathbf{P} = I \end{aligned} \tag{9.4}$$

In this objective, the dissimilarity between samples is quantified by their Euclidean distance in the projected space. $\mathbf{1}$ represents a unity row vector of length m. The term $\mathcal{T}(\mathbf{\Omega})$ typically denotes the Frobenius norm of the similarity matrix, and α serves as a regularization parameter to control its influence. The Frobenius norm is incorporated as a regularization term to avoid

degenerate or trivial solutions. The constraint $\mathbf{P}^T\mathbf{X}\mathbf{M}\mathbf{X}^T\mathbf{P} = I$ is employed to retain the discriminative characteristics of the data, particularly when M is defined as $I - \frac{1}{n}\mathbf{1}^T\mathbf{1}$. The ANG framework is capable of mitigating the adverse influence of redundant features, making it suitable for processing high-dimensional data. The normalization constraint $\omega_i\mathbf{1}^T = 1$ ensures that all similarity weights assigned to a given point, sum to one. However, this sum-to-one constraint often leads to biased weights, especially when dealing with noisy or irrelevant neighbors. Moreover, if the regularization parameter α is chosen too large, the resulting similarity scores across neighbors may become too uniform, reducing the discriminative capability of the graph.

To overcome these issues, the Possibilistic Neighborhood Graph (PNG) replaces the sum-to-one normalization with a possibilistic regularization term defined as $\psi(\omega_{ij}) = (1 - \omega_{ij})^r$, where $r > 1$ is referred to as possibilistic exponent. This possibilistic formulation allows each similarity score to be optimized independently, free from the constraint of influencing or being influenced by other neighbors. The revised optimization objective for PNG becomes:

$$\begin{aligned} \min_{\mathbf{\Omega},\mathbf{P}} \quad & \sum_i \left(\sum_{\mathbf{x}_j \in \mathcal{N}_k(\mathbf{x}_i)} \omega_{ij} \|\mathbf{P}^T\mathbf{x}_i - \mathbf{P}^T\mathbf{x}_j\|_2^2 + \alpha_i \cdot \psi(\omega_{ij}) \right) \\ \text{s.t.} \quad & 0 \le \omega_{ij} \le 1, \quad \mathbf{P}^T\mathbf{X}\mathbf{M}\mathbf{X}^T\mathbf{P} = I \end{aligned} \tag{9.5}$$

Here, $\mathcal{N}_k(\mathbf{x}_i)$ denotes the k-nearest neighbors of the point $\mathbf{x}_i$. To determine the balance parameter α_i, the locality of samples around a data point $\mathbf{x}_i$ is considered, which allows the diversity of the neighborhood for each sample. The parameter α_i is adaptively computed as [141]:

$$\alpha_i = \max_j \left\{ \left\|\mathbf{P}^T\mathbf{x}_i - \mathbf{P}^T\mathbf{x}_j\right\|_2^2 \;\middle|\; \mathbf{x}_j \in \mathcal{N}_k(\mathbf{x}_i) \right\} \tag{9.6}$$

Although PNG provides a more robust similarity model by eliminating the probabilistic normalization, it lacks an explicit mechanism to enforce sparsity in the feature space. As a result, irrelevant or noisy features may still be retained in the learned low-dimensional subspace. Furthermore, since PNG treats all input features equally during the projection process, it cannot effectively highlight the more discriminative ones. The absence of a dedicated interpretability-promoting constraint also limits the model's ability to isolate and retain only the most important features.

To address these limitations, this work introduces a sparsity constraint based on the $\ell_{2,1}$ norm, which is applied directly to the projection matrix $\mathbf{P}$. This norm promotes row-wise sparsity, thereby allowing the model to selectively emphasize more informative features while

suppressing the less relevant ones. The $\ell_{2,1}$ norm is defined as $\|\mathbf{P}\|_{2,1} = \sum_{i=1}^{D} \|\mathbf{p}_{i,:}\|_2$, where D is the original feature dimension and $\mathbf{p}_{i,:}$ denotes the i-th row of P [70]. By minimizing this norm, the rows of the projection matrix associated with non-informative or noisy features, are pushed to shrink toward zero. This ensures that such features contribute minimally, to the final projection, thereby improving robustness and feature selection. In addition, the sparsity induced by this regularization term enhances the interpretability of the model by identifying and retaining only the most discriminative features. With this enhancement, the final objective function for the proposed Robust Possibilistic Neighborhood Graph (RPNG) is expressed as:

$$\begin{aligned} \min_{\mathbf{\Omega},\mathbf{P}} \quad & \sum_{i=1}^{m} \left(\sum_{\mathbf{x}_j \in \mathcal{N}_k(\mathbf{x}_i)} \omega_{ij} \|\mathbf{P}^T\mathbf{x}_i - \mathbf{P}^T\mathbf{x}_j\|_2^2 + \alpha_i \cdot \psi(\omega_{ij}) + \lambda \|\mathbf{P}\|_{2,1} \right) \\ \text{s.t.} \quad & 0 \le \omega_{ij} \le 1, \quad \mathbf{P}^T\mathbf{X}\mathbf{M}\mathbf{X}^T\mathbf{P} = I \end{aligned} \tag{9.7}$$

Here, $\lambda > 0$ is a regularization parameter that balances the trade-off between locality preservation and sparsity in feature selection. This optimization problem is jointly convex in both $\mathbf{\Omega}$ and $\mathbf{P}$, and needs to be solved using an alternating minimization strategy [60]. Consequently, it is solved by considering one variable at a time as given below.

1) Optimizing Ω while Keeping P Fixed

When the projection matrix $\mathbf{P}$ is kept constant, the optimization problem in Equation (9.7) simplifies to:

$$\min_{\mathbf{\Omega}} \sum_{i=1}^{m} \left(\sum_{\mathbf{x}_j \in \mathcal{N}_k(\mathbf{x}_i)} \omega_{ij} \left\|\mathbf{P}^T\mathbf{x}_i - \mathbf{P}^T\mathbf{x}_j\right\|_2^2 + \alpha_i \cdot \psi(\omega_{ij}) \right) \quad \text{s.t.} \quad 0 \le \omega_{ij} \le 1 \tag{9.8}$$

Based on the approach in [60], a closed-form expression for ω_{ij} can be derived as:

$$\omega_{ij}^{*} = \begin{cases} 1 - \left(\frac{h}{r}\right)^{\frac{1}{r-1}}, & \text{if } \mathbf{x}_j \in \mathcal{N}_k(\mathbf{x}_i) \\ 0, & \text{otherwise} \end{cases} \tag{9.9}$$

Here, the term h is defined as $h = \frac{\left\|\mathbf{P}^T\mathbf{x}_i - \mathbf{P}^T\mathbf{x}_j\right\|_2^2}{\alpha_i}$, which lies in the interval $[0,1]$ and remains unaffected by the optimization variable $\mathbf{\Omega}$. Once the values of ω_{ij} are computed, the similarity matrix is symmetrized and updated as $\mathbf{\Omega} = \frac{1}{2}(\mathbf{\Omega} + \mathbf{\Omega}^T)$. In this stage, the effect of the sparsity constraint does not reflect, since the optimization step for $\mathbf{\Omega}$ is independent of $\mathbf{P}$ which includes the $\ell_{2,1}$ term.

2) Optimizing P while Keeping Ω Fixed

With $\mathbf{\Omega}$ held constant, the objective becomes:

$$\min_{\mathbf{P}^T\mathbf{X}\mathbf{X}^T\mathbf{P}=I} \sum_{i=1}^{m} \sum_{\mathbf{x}_j \in \mathcal{N}_k(\mathbf{x}_i)} \omega_{ij} \left\| \mathbf{P}^T\mathbf{x}_i - \mathbf{P}^T\mathbf{x}_j \right\|_2^2 + \lambda \|\mathbf{P}\|_{2,1} \tag{9.10}$$

Here, the matrix $\mathbf{M}$ is assumed to be an identity matrix I [60]. Using standard linear algebraic reformulation, the first term is equivalently written as $2 \cdot \mathrm{tr}(\mathbf{P}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{P})$, where $\mathbf{L} = \mathbf{\Lambda} - \mathbf{\Omega}$ represents the graph Laplacian constructed from $\mathbf{\Omega}$. Consequently, the problem reduces to the following optimization form:

$$\min_{\mathbf{P}^T\mathbf{X}\mathbf{X}^T\mathbf{P}=I} \mathrm{tr}(\mathbf{P}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{P}) + \frac{\lambda}{2}\|\mathbf{P}\|_{2,1} \tag{9.11}$$

The first term, $\mathrm{tr}(\mathbf{P}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{P})$, ensures the neighborhood structure is preserved in the embedded space by optimizing the weighted distances, while the second term introduces sparsity by penalizing irrelevant features through the $\ell_{2,1}$ norm. The orthogonality constraint $\mathbf{P}^T\mathbf{X}\mathbf{X}^T\mathbf{P} = I$ avoids trivial solutions by enforcing decorrelation among the projected features. To facilitate optimization, the $\ell_{2,1}$ norm of the projection matrix is expanded as:

$$\|\mathbf{P}\|_{2,1} = \sum_{i=1}^{D} \sum_{j=1}^{d} \sqrt{p_{ij}^2} \tag{9.12}$$

Here, D and d denote the dimensionality of the original feature space and the projected feature space, respectively. An auxiliary diagonal matrix $\mathbf{\Delta}$ is introduced, with entries defined as $\Delta_{ii} = \frac{1}{\|\mathbf{p}_{i,:}\|_2}$. With this auxiliary matrix, the sparsity regularizer is reformulated as $\mathrm{tr}(\mathbf{P}^T\mathbf{\Delta}\mathbf{P})$. Subsequently, Equation (9.11) can be equivalently expressed as:

$$\min_{\mathbf{P}^T\mathbf{X}\mathbf{X}^T\mathbf{P}=I} \mathrm{tr}\left(\mathbf{P}^T\left(\mathbf{X}\mathbf{L}\mathbf{X}^T + \frac{\lambda}{2}\mathbf{\Delta}\right)\mathbf{P}\right) \tag{9.13}$$

This expression given above can be reformulated as a generalized eigenvalue problem, by defining $\mathbf{J} = \mathbf{X}\mathbf{L}\mathbf{X}^T + \frac{\lambda}{2}\mathbf{\Delta}$. With this, the optimization task reduces to:

$$\min_{\mathbf{P}^T\mathbf{X}\mathbf{X}^T\mathbf{P}=I} \mathrm{tr}(\mathbf{P}^T\mathbf{J}\mathbf{P}) \tag{9.14}$$

This minimization is equivalent to solving the following generalized eigenvalue problem:

$$\mathbf{J}\mathbf{v} = \mu(\mathbf{X}\mathbf{X}^T)\mathbf{v} \tag{9.15}$$

Here, the set $\{\mathbf{v}_i\}_{i=1}^n$ represents the eigenvectors, and the corresponding eigenvalues are denoted by $\{\mu_i\}_{i=1}^n$. Using this approach, eigenvectors are computed for the matrix product $(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{J}$. From this set, the d eigenvectors associated with the smallest eigenvalues are selected, and these form the projection matrix, specifically, the columns of $\mathbf{P}$. To help in understanding the overall

Algorithm 9.1 Robust Possibilistic Neighborhood Graph (RPNG)

Input: EMG dataset $\mathbf{X} \in \mathbb{R}^{D\times m}$, initial similarity matrix $\mathbf{\Omega}_0$, reduced dimension d, neighborhood size k, possibilistic exponent r, regularization weight λ

Output: Final adaptive robust similarity matrix $\mathbf{\Omega}^*$

STEP-1 Initialize the auxiliary diagonal matrix: $\mathbf{\Delta}_0 = I \in \mathbb{R}^{D\times D}$.

STEP-2 Compute the initial OLPP diagonal matrix $\mathbf{\Lambda}_0$ from $\mathbf{\Omega}_0$ and Laplacian matrix $\mathbf{L}_0 = \mathbf{\Lambda}_0 - \mathbf{\Omega}_0$.

STEP-3 Formulate the matrix: $\mathbf{J}_0 = \mathbf{X}\mathbf{L}_0\mathbf{X}^T + \frac{\lambda}{2}\mathbf{\Delta}_0$.

STEP-4 Solve the generalized eigenvalue problem: $\mathbf{J}_0\mathbf{v} = \mu(\mathbf{X}\mathbf{X}^T)\mathbf{v}$.

STEP-5 Select d eigenvectors corresponding to the smallest eigenvalues to form the initial projection matrix $\mathbf{P}_0$.

STEP-6 Repeat until convergence:

STEP-6.1 *Update* $\mathbf{\Omega}$*:*

- For each data point $\mathbf{x}_i$, find k-nearest neighbors: $\mathcal{N}_k(\mathbf{x}_i)$.
- Compute $\alpha_i = \max_j \left\{ \left\| \mathbf{P}^T\mathbf{x}_i - \mathbf{P}^T\mathbf{x}_j \right\|_2^2 \;\middle|\; \mathbf{x}_j \in \mathcal{N}_k(\mathbf{x}_i) \right\}$.
- For each neighbor $\mathbf{x}_j$, compute $h = \frac{\|\mathbf{P}^T\mathbf{x}_i - \mathbf{P}^T\mathbf{x}_j\|_2^2}{\alpha_i}$ and update: $\omega_{ij} = 1 - \left(\frac{h}{r}\right)^{\frac{1}{r-1}}$

STEP-6.2 Symmetrize the similarity matrix: $\mathbf{\Omega} = \frac{1}{2}(\mathbf{\Omega} + \mathbf{\Omega}^T)$.

STEP-6.3 *Update* $\mathbf{P}$*:*

- Update $\mathbf{\Delta}$ using $\Delta_{ii} = \frac{1}{\|\mathbf{p}_{i,:}\|_2}$.
- Recompute $\mathbf{\Lambda}$, $\mathbf{L}$, and update $\mathbf{J} = \mathbf{X}\mathbf{L}\mathbf{X}^T + \frac{\lambda}{2}\mathbf{\Delta}$.
- Solve $\mathbf{J}\mathbf{v} = \mu(\mathbf{X}\mathbf{X}^T)\mathbf{v}$ and compute the updated projection matrix $\mathbf{P}$.

STEP-7 Return the final similarity matrix: $\mathbf{\Omega}^*$.

procedure, the full RPNG-OLPP algorithm is presented in Algorithm 9.1.

9.4 Experimental Results and Discussion

This study has involved the recording of EMG signals from a group of 25 volunteers performing ten different hand gestures and muscular movements, conducted across two distinct recording sessions. Comprehensive details of this dataset can be found in Section 7.4 of Chapter 7. To

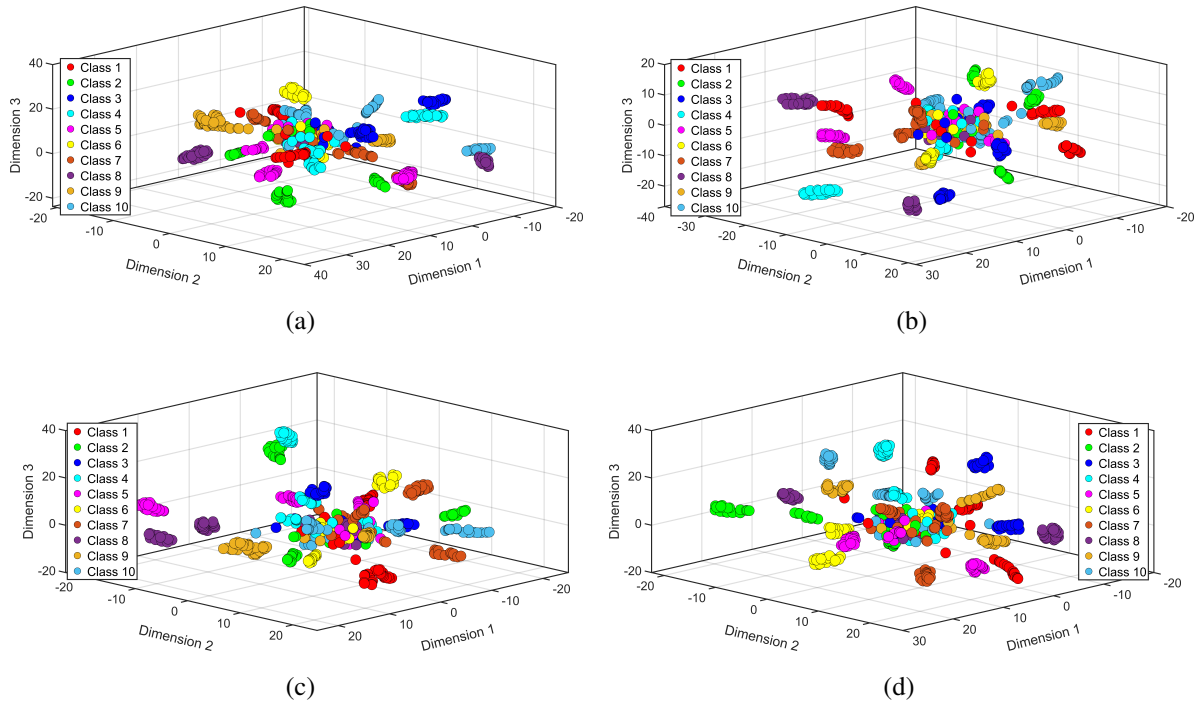


Figure 9.1: 3D representation of the low-dimensional embeddings obtained from the original high-dimensional EMG data using the dimensionality reduction techniques: (a) ALPR-OLPP, (b) LPP_SGE, (c) HC-OLPP, and (d) the proposed RPNG-OLPP.

remove unwanted electrical interference, a notch filter centered at the power line frequency has been applied to the recorded signals. Following this, mean value subtraction has been performed to eliminate baseline drift from the EMG signal recordings. For evaluation purposes, the dataset has been randomly partitioned into training and testing sets in a $70:30$ ratio. Using these processed EMG data, a robust and adaptively learned similarity graph has been constructed through the proposed RPNG method. Subsequently, Orthogonal Locality Preserving Projection (OLPP) has been employed to generate the orthogonal eigenvectors and derive the projection matrix for reducing the data dimensionality.

The effectiveness of the dimensionality reduction techniques, including ALPR-OLPP, LPP_SGE, HC-OLPP, and the proposed RPNG-OLPP, has been visually examined using 3D projections generated by t-distributed stochastic neighbor embedding (t-SNE), as shown in Fig. 9.1. These visualizations exhibit improved inter-class separability across all methods, with the proposed RPNG-OLPP showing the clearest separation. This result underscores the capability of the RPNG-OLPP framework to retain the most significant discriminative patterns within the transformed EMG data. To statistically assess the quality of the learned low-dimensional

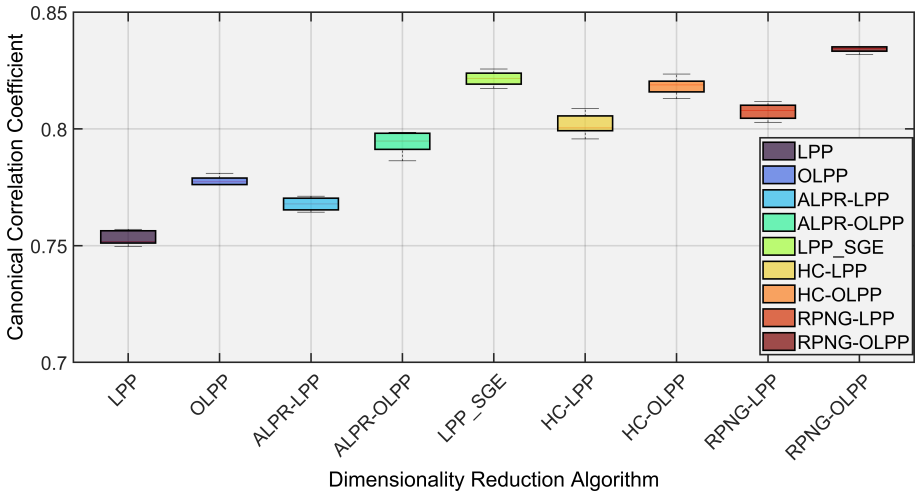


Figure 9.2: Box plot illustrating the canonical correlation coefficients (CCCs) between the original high-dimensional EMG signals and their corresponding low-dimensional representations obtained using different dimensionality reduction methods.

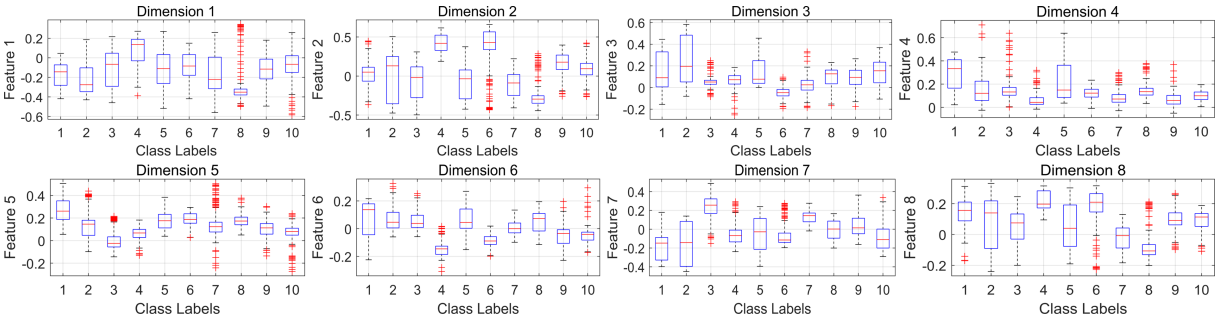


Figure 9.3: Box plots showing the feature value distributions across the first eight principal dimensions in the reduced space obtained after dimensionality reduction.

embeddings, canonical correlation analysis (CCA) has been conducted. CCA is a multivariate technique used to analyze the relationships between two variable sets. In this context, it serves to assess how well the structure of the original EMG signals is preserved in the reduced space. As the canonical correlation coefficient (CCC) reflects the extent to which the projected features maintain the original correlations; higher values imply stronger linear relationships. A comparative CCA evaluation between RPNG-OLPP and several other baseline dimensionality reduction methods is illustrated in Fig. 9.2.

Moreover, to examine how the RPNG-OLPP projected features are distributed across dimensions, box plots are presented in Fig. 9.3, with each subplot representing a distinct reduced dimension. These visualizations reflect the class-wise variability and spread in the low-dimensional space, offering insights into how well the projection maintains discriminative

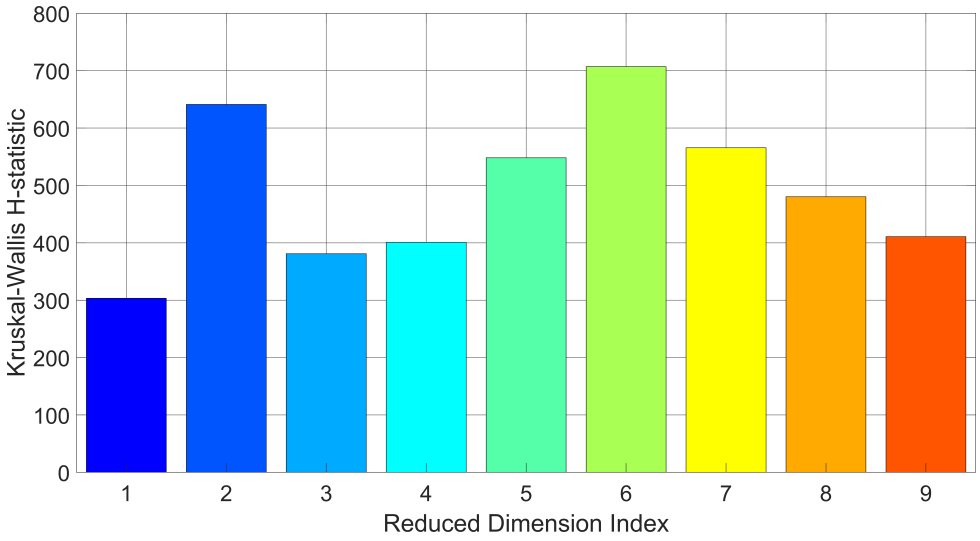


Figure 9.4: H-statistic scores computed from the Kruskal–Wallis test for the first nine individual dimensions in the reduced feature space.

structure among classes. To quantitatively evaluate the dimensionality reduction performance, the Kruskal-Wallis non-parametric statistical test has been conducted on each individual dimension. This test has been used to determine whether there are statistically significant differences in the distributions of projected features across different gesture classes. The computed H-statistic values for all dimensions are displayed in Fig. 9.4, where higher values indicate greater statistical divergence among the classes. Such enhanced H-statistics imply that the corresponding dimensions offer better separability between hand movement categories. Conversely, dimensions with relatively lower H-statistic values suggest that the projected features are less effective in distinguishing between classes. Thus, dimensions yielding higher H-scores are likely to play a more impactful role in improving classification outcomes.

After dimensionality reduction, classification has been carried out using both k-nearest neighbor (kNN) and support vector machine (SVM) classifiers. Although both classifiers have provided competitive results, the SVM classifier has consistently yielded slightly better performance and has been used for reporting the final classification outcomes. The detailed results for classification accuracy, precision, and recall across different methods are listed in Table 9.1. To further evaluate class-wise performance, receiver operating characteristic (ROC) curves were plotted for each individual gesture class, as illustrated in Fig. 9.5, along with their corresponding area under the curve (AUC) values. These ROC plots demonstrate the trade-off between true positive rate (TPR) and false positive rate (FPR) at different classification thresholds, with the average AUC achieved being 0.948. Performance analyses have also

Table 9.1: CLASSIFICATION ACCURACY, PRECISION, AND RECALL OF DIFFERENT DR METHODS UNDER NORMAL AND NOISY CONDITIONS

Method	Normal Condition			Noise Level 1			Noise Level 2		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall
LPP	81.73	82.44	81.60	70.65	71.05	70.84	64.60	64.54	64.83
OLPP	84.15	84.61	84.08	74.13	74.47	73.88	67.20	66.73	67.48
ALPR-LPP	82.67	83.35	82.54	72.45	72.77	72.81	66.41	67.20	66.75
ALPR-OLPP	85.58	86.18	85.41	76.33	76.32	77.21	70.34	69.75	69.81
LPP_SGE	84.41	84.91	84.22	77.70	77.43	77.29	70.66	70.21	70.22
HC-LPP	85.36	85.86	85.29	79.46	79.77	79.22	70.45	70.77	70.48
HC-OLPP	87.27	87.77	87.13	81.52	81.65	81.16	73.33	72.64	72.78
RPNG-LPP	86.12	86.54	86.15	80.61	80.82	80.33	72.59	71.76	71.95
RPNG-OLPP	**89.85**	**89.83**	**90.36**	**84.18**	**83.68**	**84.55**	**77.31**	**76.94**	**77.66**

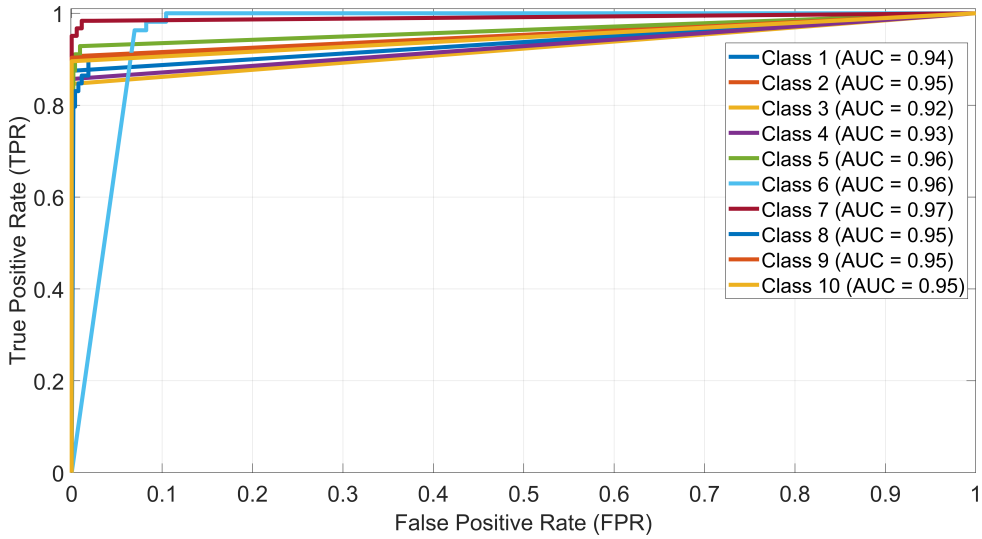


Figure 9.5: Receiver operating characteristic (ROC) curves for each individual class along with their corresponding area under the curve (AUC) values, obtained using the RPNG-OLPP method combined with SVM classification.

been carried out under three different conditions: original signals (normal), and two levels of additive zero-mean Gaussian noise—one with variance 0.05, and the other with variance 0.10. The results shown in Table 9.1 demonstrate that the RPNG-OLPP framework remains robust across increasing noise levels, with only a modest degradation in performance. This performance stability is attributed to the combined strength of orthogonal projection and robust adaptive similarity modeling inherent to the RPNG-OLPP method. Additional classification error metrics, including false acceptance rate (FAR), false negative rate (FNR), equal error rate (EER), and false discovery rate (FDR), are depicted in Fig. 9.6 using a heatmap representation.

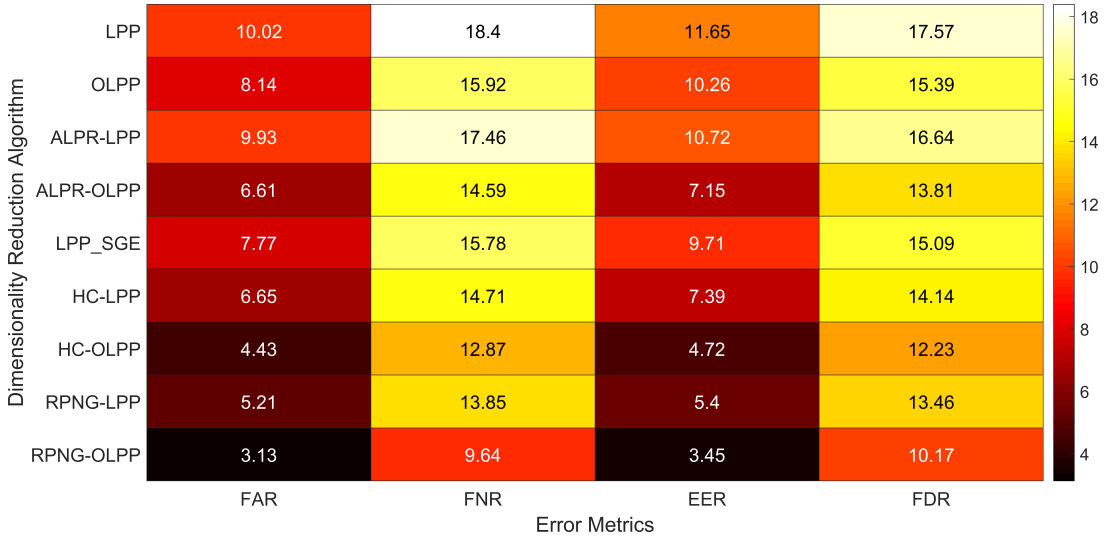


Figure 9.6: Heatmap comparison of multiple state-of-the-art dimensionality reduction techniques based on classification error metrics: false acceptance rate (FAR), false negative rate (FNR), equal error rate (EER), and false discovery rate (FDR).

Table 9.2: RECOGNITION ACCURACY (%) OF DIFFERENT DIMENSIONALITY REDUCTION METHODS ON BENCHMARK NINAPRO DB1 DATASET

Method	**DB1-E1**	**DB1-E2**	**DB1-E3**	**DB1 Overall**
LPP	86.12	85.23	82.24	83.61
OLPP	91.31	91.63	88.15	86.54
ALPR-LPP	89.43	88.56	87.33	85.25
ALPR-OLPP	93.79	92.60	90.23	88.12
LPP_SGE	92.22	92.74	89.83	89.33
HC-LPP	94.16	93.85	90.05	91.90
HC-OLPP	95.89	95.52	92.45	93.10
RPNG-OLPP	**97.52**	**97.10**	**95.32**	**95.35**

These metrics have been evaluated for various dimensionality reduction methods, to provide a broader perspective on classification reliability. Among all the approaches compared, the proposed RPNG-OLPP demonstrates a significantly reduced equal error rate (EER), suggesting a better balance between false positives and false negatives. This improved balance indicates that RPNG-OLPP yields more consistent classification performance by minimizing misclassification rates across both positive and negative prediction classes.

In addition to the primary dataset acquired in the laboratory, the performance evaluation

has been further extended to the publicly available NinaPro DB1 benchmark dataset, which includes EMG recordings from 27 individuals [128], [147]. Three different subsets of the NinaPro DB1 dataset have been considered for analysis: DB1-E1, which includes 12 distinct finger movements; DB1-E2, comprising 17 isometric and isotonic hand and wrist gestures; and DB1-E3, which contains 23 grasp and functional movement classes. The classification accuracies corresponding to these three exercise subsets, as well as the complete dataset, are summarized in Table 9.2. The results clearly demonstrate the effectiveness of the proposed robust RPNG-OLPP framework, which consistently surpasses other dimensionality reduction techniques across all subsets.

9.5 Summary

This chapter presented a robust and adaptive dimensionality reduction framework—Robust Possibilistic Neighborhood Graph integrated with Orthogonal Locality Preserving Projection (RPNG-OLPP), tailored for surface EMG signal analysis. The method effectively addresses limitations associated with traditional similarity graph construction and feature selection in OLPP by incorporating a possibilistic similarity modeling approach and enforcing row-wise sparsity through an $\ell_{2,1}$ norm regularization. This dual enhancement facilitates the suppression of irrelevant or noisy features while retaining the most discriminative components, thus ensuring improved interpretability and robustness in noisy and non-stationary sEMG environments.

The proposed approach was rigorously evaluated on a custom laboratory-acquired sEMG dataset under clean and noisy conditions, as well as on the widely used benchmark NinaPro DB1 dataset. Comparative analyses against several state-of-the-art dimensionality reduction methods demonstrated improved low-dimensional embedding and enhanced classification performance of RPNG-OLPP. Quantitative validation through canonical correlation analysis, Kruskal-Wallis testing, ROC-AUC curves, and multiple error metrics; including FAR, FNR, EER, and FDR, further substantiated its effectiveness. The results consistently highlighted the model's ability to maintain inter-class separability, reduce misclassification, and provide robust performance under varying acquisition conditions.

These findings underscore the potential of the RPNG-OLPP framework for deployment in real-world biomedical applications, including gesture-driven human-machine interfaces,

prosthetic control, and rehabilitation systems, particularly in cases where data reliability is compromised due to sensor or environmental disturbances. Looking ahead, future enhancements can explore the integration of metaheuristic or evolutionary optimization strategies for an automatic tuning of graph construction parameters and sparsity constraints. Additionally, hybrid models that integrate RPNG-OLPP with hierarchical feature extraction frameworks or temporal pattern modeling schemes may further enhance its ability to capture complex structures and dynamics within EMG signals. Expanding the framework's application to multimodal datasets, involving inertial or visual cues, could also broaden its utility in complex human-robot interaction scenarios.

Chapter 10

Conclusion

This chapter presents a comprehensive conclusion to the work undertaken in this Thesis, which focuses on the development of robust dimensionality reduction (DR) frameworks for human-robot collaboration (HRC) in real-world environments. The central aim was to design and evaluate enhanced locality preserving projection (LPP)-based dimensionality reduction techniques, capable of operating reliably across heterogeneous sensing modalities, including vision-based symbolic cue recognition and surface electromyography (sEMG)-based hand activity analysis. Throughout the Thesis, several novel algorithmic models were proposed, benchmarked, and validated under challenging operational conditions involving photometric degradation, sensor-induced noise, and subject-level variability. This chapter synthesizes the key findings from each domain, highlights the shared cross-modality insights, and points out the remaining research gaps that invite further exploration. Additionally, considerations for practical deployment and future research directions have been systematically discussed to provide a holistic closure to the Thesis.

10.1 Summary of Key Findings

This Thesis presented a comprehensive investigation into robust subspace learning strategies for human-robot collaboration (HRC) under real-world sensing constraints. The core objective was to develop and evaluate locality preserving projection (LPP)-based frameworks that can operate reliably across diverse sensing modalities; namely, vision-based symbolic cue detection and surface electromyography (sEMG)-based hand activity recognition.

In the visual sensing domain, the work began with the design and implementation of a symbolic cue-based robot navigation framework using structured flag-stick markers. A comprehensive dataset was acquired under systematically varied photometric conditions to

emulate real-world lighting variability. In addition to lighting-based adversities, situations involving external noise corruptions such as salt-and-pepper noise, speckle noise, and Gaussian noise were also introduced to simulate sensor imperfections and environmental interferences common in real-world robotic vision systems. Baseline experiments using traditional LPP confirmed its efficacy in preserving local geometrical information, while also exposing its limitations under degraded illumination. To address these, several advanced projection models were proposed, including:

- **ALPSK-BLPP**: An adaptive spatial-feature kernel-guided bilateral LPP model was designed to improve robustness under varying illumination conditions. It combines a spatial similarity kernel, learned using graph-regularized retargeted least squares regression (ReLSR), with a feature-domain kernel based on local binary patterns (LBP). The multiplicative fusion of these kernels constructs a bilateral similarity graph that captures both spatial structure and photometric similarity. This enhanced graph is used within the LPP framework to generate a discriminative subspace, yielding improved classification performance under challenging visual environments.

- **LTrP-BLPP**: A bilateral LPP model that employs Local Tetra Patterns (LTrP) to ameliorate robustness under sensor-induced noise in vision-based robotic perception. The four-directional encoding of LTrP captures both edge orientation and spatial gradients, making it effective in distinguishing structural details even under distorted or corrupted imaging conditions. These encoded features are integrated into a bilateral similarity graph that combines spatial and intensity-based information for improved neighborhood representation.

- **HRLBP-BLPP and HRLTP-BLPP**: These models incorporate histogram-refined local pattern encoders within a bilateral locality preserving projection framework to enhance resilience against photometric irregularities and sensor noise. HRLBP-BLPP utilizes histogram-refined local binary patterns, while HRLTP-BLPP extends this idea using local ternary patterns (LTP), which are more robust in near-uniform regions and under weak illumination gradients. Both models apply histogram refinement using local skew patterns and binarized eigenvalue maps to retain detailed local neighborhood structures. The resulting descriptors are embedded in a bilateral graph that fuses spatial and feature-domain similarities, enabling more reliable subspace projection under varying

lighting conditions and sensor noise.

Extensive evaluations across controlled lighting degradations, salt-and-pepper noise, and speckle noise demonstrated that the proposed models consistently outperformed traditional approaches. Notably, ALPSK-BLPP has demonstrated enhanced performance under moderate to severely challenging illumination conditions and HRLTP-BLPP showed robustness against simultaneous low-light and noise-induced settings, confirming their suitability for deployment in embedded robotic vision systems.

In addition to these, granular computing-inspired feature modeling approaches were introduced to improve the spatial characterization of visual cues, particularly in scenarios involving geometric distortion and regional degradation. These models focused on enhancing foreground-background segmentation and preserving local boundary integrity prior to subspace embedding. Specifically two new models were developed:

- **REGF-2DLPP**: A rough entropy-guided fused granulation approach that combines homogeneous (crisp granulation) and non-homogeneous (quad-tree decomposition) granular representations over RGB channels to improve 2DLPP-based dimensionality reduction. The fusion mechanism effectively captures foreground–background boundaries by reducing spatial and color uncertainty through rough entropy maximization, without converting to grayscale. A binary feature decoding scheme inspired by local binary patterns, is used to extract robust descriptors from the fused granulated map. This granulated feature space is then integrated with spatial similarity to build a hybrid kernel for 2DLPP, preserving structural integrity even under degraded illumination. Experimental results confirm that REGF-2DLPP achieves superior recognition accuracy and robustness in photometrically unfavourable HRI environments.

- **dNG-2DRLPP**: The core idea lies in density-based neighborhood granulation (dNG), which forms irregular-shaped granules based on spatial and color proximities. This enables more accurate modeling of ambiguous foreground–background transitions, especially in cluttered and low-contrast environments. A fuzzy-based decoding scheme further enhances the representation by assigning soft membership values to pixels based on their likelihood of belonging to the foreground or background. The resulting granular feature descriptors are fused with the spatial kernel to construct a hybrid similarity graph, which is then embedded into the robust two-dimensional LPP formulation using an

ℓ_1-norm optimization.

Both REGF-2DLPP and dNG-2DRLPP models were rigorously evaluated under complex visually impaired scenarios, including extreme low-light conditions and spatial variations. REGF-2DLPP exhibited enhanced foreground–background separation through rough entropy-based fused granulation, resulting in improved recognition for visual cues captured under severely challenging lighting conditions. On the other hand, dNG-2DRLPP, leveraging density-based irregular-shaped granulation and fuzzy feature decoding, demonstrated superior resilience by preserving discriminative structural details even under severe illumination deprivation and sensor-induced noise. The consistent performance of these models across adverse sensing conditions, validates their potential for robust real-world robotic visual perception in dynamic and uncertain human-robot interaction environments.

In the sEMG-based modality, two novel orthogonal LPP variants were introduced to improve hand activity recognition:

- **UaBMA-OLPP**: This model introduces a novel orthogonal locality preserving projection framework, reinforced by uncertainty-aware Bayesian model averaging. It integrates multiple similarity kernels derived from Euclidean space, Euler space, and the Grassmann manifold, thereby capturing spatial, directional, and structural characteristics of high-dimensional sEMG data. The model adaptively computes posterior probabilities based on classification performance and uncertainty, effectively balancing contributions from different manifold-based kernels. This results in a more discriminative subspace embedding that preserves local geometry, while mitigating the influence of inter-subject variability and signal irregularity. Extensive evaluations on both laboratory-acquired and benchmark datasets demonstrate its superior accuracy and robustness compared to existing DR techniques.

- **RPNG-OLPP**: A robust orthogonal locality preserving projection framework, augmented with a possibilistic neighborhood graph (PNG) for adaptive similarity estimation in noisy sEMG environments. Unlike traditional Euclidean-based graph construction, PNG models pairwise relationships using a possibilistic regularization term, increasing resilience against outliers. Additionally, an $\ell_{2,1}$-norm sparsity constraint is introduced on the projection matrix, promoting row-level sparsity that effectively prunes irrelevant or redundant features. This dual regularization enables the model to preserve the

intrinsic structure of high-dimensional EMG data while improving interpretability and robustness. Extensive evaluations under both clean and noise-corrupted conditions, confirm RPNG-OLPP's superior classification accuracy and structural consistency, making it highly suited for EMG-based prosthetic control and HMI applications.

Experimental results demonstrated that the proposed sEMG frameworks achieved superior generalization performance compared to classical LPP, PCA, and other graph-based DR methods. The RPNG-OLPP model, in particular, offered stable classification accuracy across subjects and sessions, even in the presence of physiological inconsistencies and sensor noise.

Overall, the Thesis contributes multiple novel dimensionality reduction algorithms that significantly enhance the interpretability, adaptability, and resilience of HRC systems operating in unstructured environments. The methods developed herein establish a robust foundation for scalable, multimodal perception frameworks in next-generation collaborative robotics.

10.2 Cross-Modality Insights

This Thesis explored both vision-based and surface electromyography (sEMG)-based sensing modalities to address robust human–robot collaboration (HRC) under real-world uncertainties. While the two modalities operate on fundamentally different physical principles i.e., visual pattern encoding versus bioelectrical signal decoding; they share common challenges in terms of high-dimensionality, noise susceptibility, and inter-class similarity. The cross-modality investigations conducted throughout the Thesis offer several important insights into how dimensionality reduction and subspace learning techniques can be generalized and adapted across domains.

Firstly, locality preserving projection (LPP)-based models proved effective in capturing intrinsic neighborhood structures in both modalities. However, while visual cues primarily benefited from spatial and structural feature enhancements such as bilateral spatial-feature kernels and granular computing; the sEMG domain required integration of statistical confidence and inter-subject variability handling through uncertainty-aware and possibilistic formulations. This highlights the modality-specific adaptations necessary for reliable feature embedding.

Secondly, the concept of hybrid kernel construction emerged as a unifying principle. In visual sensing, spatial and range similarity kernels were fused to handle

ambiguous foreground-background transitions, whereas, in sEMG, multiple geometrical representation-based neighborhood graphs were combined to represent signal confidence and reduce the impact of variability across subjects and sessions. This multi-kernel philosophy enabled more robust graph construction and ultimately more discriminative subspace mappings.

Finally, both sensing modalities underscored the importance of adaptive, data-driven model construction; moving beyond fixed projection frameworks toward those that incorporate uncertainty, sparsity, and heterogeneous similarity cues. These shared design patterns reflect a deeper convergence in multimodal signal processing, where robust subspace learning serves as a powerful abstraction for diverse HRC environments.

Together, these insights pave the way for integrated multimodal perception systems capable of fusing vision and myoelectric signals to achieve context-aware, resilient, and real-time collaborative robotic behavior.

10.3 Remaining Gaps and Future Research Directions

While the methods developed in this Thesis significantly advance robust dimensionality reduction and feature extraction for vision- and sEMG-based human-robot interaction (HRI) systems, several research gaps and limitations remain, that could guide future investigations.

One key limitation is the sensitivity of learned projections to domain shifts in uncontrolled environments. Although models like dNG-2DRLPP and RPNG-OLPP have shown resilience under photometric deterioration and physiological variability, their performance may degrade under dynamic, unseen operating conditions such as extreme occlusions, multiple human interactions, or abrupt lighting transitions. Domain adaptation strategies and self-supervised learning mechanisms may help extend generalization capabilities in such scenarios.

Secondly, while the incorporation of combined spatial-photometric similarity measures and probabilistic similarity measures significantly improved neighborhood modeling; these strategies largely remain hand-tuned. Automatic selection of projection parameters such as kernel weights, neighborhood size, or granulation thresholds; still requires specific guidelines or manual calibration. In future work, the integration of metaheuristic algorithms or reinforcement learning-based tuning mechanisms, could optimize these parameters adaptively in real-time.

Another important gap is the limited exploitation of temporal dependencies in sEMG

signals and sequential visual inputs. The current projection models primarily operate on spatial or instance-level data, overlooking sequential correlations that are critical in dynamic gesture recognition or continuous motion prediction. Incorporating temporal modeling techniques may offer a better abstraction of time-evolving patterns in multimodal sensor streams, especially for dynamically changing human cues.

Moreover, most of the experiments were conducted on datasets acquired in moderately controlled laboratory conditions. While this provided a consistent testbed for benchmarking, real-world deployment scenarios such as extremely cluttered domestic spaces, outdoor environments, or multi-agent collaborative settings introduce additional complexities. Expanding the current data acquisition framework to include more diverse environments and user profiles would offer stronger external validation of the proposed models.

Lastly, the current implementations are primarily offline and partially embedded. For full deployment in assistive or industrial robotic systems, future research should focus on real-time implementation, energy-efficient computation, and edge-based deployment of these subspace learning models using resource-constrained hardware such as field-programmable gate arrays (FPGAs) or low-power AI accelerators.

Addressing these gaps will not only refine the robustness and adaptability of HRC systems but also bring them closer to scalable, real-world applications in healthcare, smart manufacturing, and human-assistive robotics.

10.4 Deployment Considerations in HRI/HMI Systems

Translating the proposed dimensionality reduction techniques from laboratory settings to real-world deployment in human-robot interaction (HRI) or human-machine interface (HMI) systems, necessitates careful consideration of several practical factors. These include computational efficiency, hardware compatibility, adaptability to user and environmental variability, and integration with control architectures in robotic platforms.

From a computational perspective, many of the proposed models involve optimization, dynamic graph construction, and multi-channel feature extraction, which may introduce latency in real-time applications. Although these techniques demonstrated strong robustness under diverse noise and distortion conditions, their computational cost must be optimized for

low-power embedded platforms. Techniques such as sparse approximation, projection matrix pruning, or lightweight approximations of kernel functions may help reduce complexity without significantly compromising accuracy.

Regarding hardware deployment, the developed algorithms have been tested on embedded platforms such as Raspberry Pi with onboard vision sensor and Arduino-powered EMG data acquisition system, offering a proof-of-concept for real-time deployment. However, wider adoption would require implementation on more resource-constrained or application-specific hardware platforms such as FPGAs, systems on chips (SoCs), or neuromorphic processors. Efficient hardware-software co-design will be essential to balance energy efficiency, speed, and model robustness.

Environmental adaptability is another key concern. In vision-based systems, lighting variation, occlusion, and background clutter pose significant challenges, whereas in wearable sEMG systems, electrode placement, skin impedance variation, and muscle fatigue can affect signal quality. The proposed models incorporate measures, such as uncertainty modeling and granularity refinement to mitigate such effects. However, deployment in unpredictable environments would benefit from closed-loop feedback mechanisms and adaptive retraining based on environmental cues and user feedback.

Moreover, seamless integration with higher-level robot control systems is critical. While the Thesis focuses on perception and subspace learning, real-world HRI/HMI systems must also account for decision-making, actuation, and safety. Interfacing the learned low-dimensional representations with trajectory planning, intent inference, or shared autonomy modules would be an important step toward holistic deployment.

Finally, for user acceptability and scalability, future deployments should consider human factors such as intuitiveness, comfort, and learning curve. Wearable interfaces should be ergonomic and unobtrusive, while visual interfaces should be interpretable and transparent to the user. Continuous usability testing and co-design with target end-users (e.g., elderly, patients, industrial operators) will be essential in bridging the gap between laboratory innovations and societal impact.

10.5 Final Remarks

This Thesis presented a comprehensive investigation into robust subspace learning strategies tailored for human-robot interaction (HRI) systems, operating under real-world uncertainties. By addressing the limitations of conventional dimensionality reduction methods, the research introduced multiple advanced models, spanning bilateral projection schemes, granular computing-based feature extraction, uncertainty-aware subspace embedding, etc., for both vision-based and sEMG-based sensing modalities.

Building on these methodological advancements, each proposed model was designed to incrementally overcome challenges such as photometric degradation, sensor-induced noise, inter-subject variability, and foreground-background ambiguity. Through extensive evaluations across controlled and degraded conditions, the models demonstrated notable improvements in classification accuracy, structural consistency, and cross-scenario generalization. Importantly, the findings also revealed how underlying principles such as kernel fusion, sparse neighborhood modeling, and granular feature decoding can be adapted effectively across diverse sensory domains, as further reflected in their quantitative performance highlighted below.

Across both sensing modalities, the proposed LPP-based frameworks demonstrated strong and consistent performance under diverse degradation scenarios. In the vision domain, the enhanced bilateral and granular computing-based variants, such as ALPSK-BLPP, REGF-2DLPP, and dNG-2DRLPP, improved recognition accuracy from around 55% under severe illumination degradation to nearly 71% and $81-85\%$ respectively, across challenging visual conditions. Under noise-corrupted scenarios, models like LTrP-BLPP and HRLTP-BLPP maintained high stability, achieving accuracies above 80% under the toughest conditions considered. In the wearable sEMG domain, the proposed UaBMA-OLPP and RPNG-OLPP frameworks consistently achieved around $88-89\%$ recognition accuracies and retained $77-84\%$ accuracy under noisy conditions (by RPNG-OLPP). These results collectively highlight the robustness and adaptability of the proposed models across heterogeneous sensing environments, underscoring their potential for real-world deployment in human-robot interaction applications. To consolidate the key outcomes of all the proposed LPP-based frameworks, a comparative summary of the developed algorithms for vision-based image data and sEMG-based hand gesture data is presented in Tables 10.1-10.4.

Table 10.1: CLASSIFICATION ACCURACY (%) OF VARIOUS PROPOSED LPP-BASED FRAMEWORKS UNDER DIFFERENT ILLUMINATION CONDITIONS FOR THE VISUAL CUE DATASET

Method	NIL	DIL-1	DIL-2	DIL-3	DIL-4
ALPSK-BLPP	99.971	98.875	97.778	81.269	71.778
REGF-2DLPP	99.980	99.436	99.167	89.480	81.565
dNG-2DRLPP	99.990	99.730	99.465	91.234	85.913

Table 10.2: CLASSIFICATION ACCURACY (%) OF VARIOUS PROPOSED LPP-BASED FRAMEWORKS WITH SALT-AND-PEPPER NOISE DENSITY $\delta = 0.35$ UNDER VARYING LIGHTING CONDITIONS FOR THE VISUAL CUE DATASET

Method	NIL	DIL-1	DIL-2
LTrP-BLPP	85.397	81.212	77.587
HRLTP-BLPP	86.206	83.498	80.350

Table 10.3: CLASSIFICATION ACCURACY (%) OF VARIOUS PROPOSED LPP-BASED FRAMEWORKS WITH SPECKLE NOISE VARIANCE $\sigma^2 = 0.35$ UNDER DIFFERENT LIGHTING CONDITIONS FOR THE VISUAL CUE DATASET

Method	NIL	DIL-1	DIL-2
LTrP-BLPP	89.206	85.924	82.346
HRLTP-BLPP	88.142	86.111	84.595

Table 10.4: CLASSIFICATION ACCURACY (%) OF VARIOUS PROPOSED LPP-BASED FRAMEWORKS UNDER NORMAL AND NOISY CONDITIONS FOR THE SEMG DATASET

Method	Normal Condition	Noise Level 1	Noise Level 2
UaBMA-OLPP	88.50	82.77	74.68
RPNG-OLPP	89.85	84.18	77.31

Beyond the algorithmic development and performance aspects, the Thesis also laid a foundation for real-world deployment by highlighting computational trade-offs, hardware feasibility, and user-centric design considerations. The cross-modality insights drawn from this

work reinforce the potential of unified, data-driven perception frameworks that are resilient, interpretable, and adaptable across dynamic human-centric environments.

In summary, this research contributes to the growing body of knowledge in robust perception for collaborative robotics. The developed models and methodologies provide a meaningful step toward enabling intelligent machines that can interact seamlessly with humans in uncertain, unstructured environments. While several open problems remain, the directions outlined in this Thesis serve as a roadmap for future explorations into scalable, multimodal, and context-aware HRI systems.

References

[1] P. Paral, A. Chatterjee, and A. Rakshit, "Chaos-based random sampling for photometric invariant shoe detection with vision sensor in human–robot coexisting environments" *IEEE Sensors J.*, vol. 20, no. 10, pp. 5424–5434, May 2020. DOI: 10.1109/JSEN.2020.2969819.

[2] D. Xiong, D. Zhang, Y. Chu, Y. Zhao, and X. Zhao, "Intuitive human-robot-environment interaction with EMG signals: A review," *IEEE/CAA J. Auto. Sin.*, vol. 11, no. 5, pp. 1075–1091, May 2024. DOI: 10.1109/JAS.2024.124329.

[3] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991. DOI: 10.1162/jocn.1991.3.1.71.

[4] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997. DOI: 10.1109/34.598228.

[5] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000. DOI: DOI: 10.1126/science.290.5500.231.

[6] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 14, 2002, pp. 585–591. DOI: 10.7551/mitpress/1120.003.0080.

[7] X. He and P. Niyogi, "Locality preserving projections," in *Adv. Neural Inf. Process. Syst.*, vol. 16, MIT Press, pp. 1–8, 2004. DOI: 10.5555/2980539.2980616.

[8] X. Yu and X. Wang, "Uncorrelated discriminant locality preserving projections," *IEEE Signal Process. Lett.*, vol. 15, pp. 361–364, 2008. DOI: 10.1109/LSP.2008.919841.

[9] D. Cai, X. He, J. Han, and H.-J. Zhang, "Orthogonal Laplacianfaces for face recognition," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3608–3614, Nov. 2006. DOI: 10.1109/TIP.2006.881945.

[10] R. Wang, F. Nie, R. Hong, X. Chang, X. Yang, and W. Yu, "Fast and orthogonal locality preserving projections for dimensionality reduction," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 5019–5030, Oct. 2017. DOI: 10.1109/TIP.2017.2726188.

[11] X. Li, J. Pan, Y. He, and C. Liu, "Bilateral filtering inspired locality preserving projections for hyperspectral images," *Neurocomputing*, vol. 164, pp. 300–306, Sep. 2015. DOI: 10.1016/j.neucom.2015.01.021.

[12] J. Long, Q. Zhang, X. Lu, J. Wen, L. Zhao, and W. Xie, "Multi-scale locality preserving projection for partial multi-view incomplete multi-label learning," *Neural Netw.*, vol. 180, 106748, Dec. 2024. DOI: 10.1016/j.neunet.2024.106748.

[13] S. Chen, H. Zhao, M. Kong, and B. Luo, "2D-LPP: A two-dimensional extension of locality preserving projections," *Neurocomputing*, vol. 70, nos. 4–6, pp. 912–921, Jan. 2007. DOI: 10.1016/j.neucom.2006.10.032.

[14] W.-J. Chen, C.-N. Li, Y.-H. Shao, J. Zhang, and N.-Y. Deng, "2DRLPP: Robust two-dimensional locality preserving projection with regularization," *Knowl.-Based Syst.*, vol. 169, pp. 53–66, Apr. 2019. DOI: 10.1016/j.knosys.2019.01.022.

[15] H. Muramatsu, Y. Itaguchi, and S. Katsura, "Involuntary stabilization in discrete-event physical human–robot interaction," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 53, no. 1, pp. 576–587, Jan. 2023. DOI: 10.1109/TSMC.2022.3184960.

[16] N. Zhang, Y. Xu, Q.-X. Zhu, and Y.-L. He, "Improved locality preserving projections based on heat-kernel and cosine weights for fault classification in complex industrial processes," *IEEE Trans. Rel.*, vol. 72, no. 1, pp. 204–213, Mar. 2023. DOI: 10.1109/TR.2021.3139539.

[17] S. Murata, Y. Li, H. Arie, T. Ogata, and S. Sugano, "Learning to achieve different levels of adaptability for human–robot collaboration utilizing a neuro-dynamical system," *IEEE Trans. Cogn. Dev. Syst.*, vol. 10, no. 3, pp. 712–725, Sep. 2018. DOI: 10.1109/TCDS.2018.2797260.

[18] P. Khan, P. Ranjan, and S. Kumar, "AT2GRU: A human emotion recognition model with mitigated device heterogeneity," *IEEE Trans. Affect. Comput.*, vol. 14, no. 2, pp. 1520–1532, Apr.–Jun. 2023. DOI: 10.1109/TAFFC.2021.3114123.

[19] J. Y. C. Chen and M. J. Barnes, "Human–agent teaming for multirobot control: A review of human factors issues," *IEEE Trans. Hum.-Mach. Syst.*, vol. 44, no. 1, pp. 13–29, Feb. 2014. DOI: 10.1109/THMS.2013.2293535.

[20] M. A. Simão, O. Gibaru, and P. Neto, "Online recognition of incomplete gesture data to interface collaborative robots," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9372–9382, Dec. 2019. DOI: 10.1109/TIE.2019.2891449.

[21] D. Avola, M. Cascio, L. Cinque, G. L. Foresti, C. Massaroni, and E. Rodolà, "2-D skeleton-based action recognition via two-branch stacked LSTM-RNNs," *IEEE Trans. Multimedia*, vol. 22, no. 10, pp. 2481–2496, Oct. 2020. DOI: 10.1109/TMM.2019.2960588.

[22] T.-Y. Pan, W.-L. Tsai, C.-Y. Chang, C.-W. Yeh, and M.-C. Hu, "A hierarchical hand gesture recognition framework for sports referee training-based EMG and accelerometer sensors," *IEEE Trans. Cybern.*, vol. 52, no. 5, pp. 3172–3183, May 2022. DOI: 10.1109/TCYB.2020.3007173.

[23] P. Paral, A. Chatterjee, and A. Rakshit, "Vision sensor-based shoe detection for human tracking in a human–robot coexisting environment: A photometric invariant approach using DBSCAN algorithm," *IEEE Sensors J.*, vol. 19, no. 12, pp. 4549–4559, Jun. 15, 2019. DOI: 10.1109/JSEN.2019.2897989.

[24] J. Navallas, C. Mariscal, A. Malanda, and J. Rodríguez-Falces, "Understanding EMG PDF changes with motor unit potential amplitudes, firing rates, and noise level through EMG filling curve analysis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 32, pp. 3240–3250, 2024. DOI: 10.1109/TNSRE.2024.3452308.

[25] N. Rabin, M. Kahlon, S. Malayev, and A. Ratnovsky, "Classification of human hand movements based on EMG signals using nonlinear dimensionality reduction and data fusion techniques," *Expert Syst. Appl.*, vol. 149, 113281, Jul. 2020. DOI: 10.1016/j.eswa.2020.113281.

[26] M. Lagomarsino, M. Lorenzini, P. Balatti, E. D. Momi, and A. Ajoudani, "Pick the right co-worker: Online assessment of cognitive ergonomics in human–robot collaborative

assembly," *IEEE Trans. Cogn. Dev. Syst.*, vol. 15, no. 4, pp. 1928–1937, Dec. 2023. DOI: 10.1109/TCDS.2022.3182811.

[27] P. Paral, A. Chatterjee, A. Rakshit, and S. K. Pal, "Extended target tracking in human–robot coexisting environments via multisensor information fusion: A heteroscedastic Gaussian process regression-based approach," *IEEE Trans. Ind. Inform.*, vol. 19, no. 9, pp. 9877–9886, Sep. 2023. DOI: 10.1109/TII.2022.3232765.

[28] D. Wei, L. Chen, L. Zhao, H. Zhou, and B. Huang, "A vision-based measure of environmental effects on inferring human intention during human–robot interaction," *IEEE Sensors J.*, vol. 22, no. 5, pp. 4246–4256, Mar. 1, 2022. DOI: 10.1109/JSEN.2021.3139593.

[29] H. Yang, J. Wan, Y. Jin, X. Yu, and Y. Fang, "EEG- and EMG-driven poststroke rehabilitation: A review," *IEEE Sensors J.*, vol. 22, no. 24, pp. 23649–23660, Dec. 15, 2022. DOI: 10.1109/JSEN.2022.3220930.

[30] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002. DOI: 10.1109/TPAMI.2002.1017623.

[31] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, Jun. 2010. DOI: 10.1109/TIP.2010.2042645.

[32] D. Jia, J. Cao, W. Song, X. Tang, and H. Zhu, "Colour FAST (CFAST) match: Fast affine template matching for colour images," *Electron. Lett.*, vol. 52, no. 14, pp. 1220–1221, Jul. 2016. DOI: 10.1049/el.2016.1331.

[33] P. Comon, "Independent component analysis, a new concept?" *Signal Process.*, vol. 36, no. 3, pp. 287–314, Apr. 1994. DOI: 10.1016/0165-1684(94)90029-9.

[34] J. Yang, D. Zhang, A. F. Frangi, and J.-Y. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, Jan. 2004. DOI: 10.1109/TPAMI.2004.1261097.

[35] M. Li and B. Yuan, "2D-LDA: A statistical linear discriminant analysis for image matrix," *Pattern Recognit. Lett.*, vol. 26, no. 5, pp. 527–532, Apr. 2005. DOI: 10.1016/j.patrec.2004.09.007.

[36] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000. DOI: 10.1126/science.290.5500.2323.

[37] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 2, Beijing, China, 2005, pp. 1208–1213. DOI: 10.1109/ICCV.2005.167.

[38] Y. Pang, L. Zhang, Z. Liu, N. Yu, and H. Li, "Neighborhood preserving projections (NPP): A novel linear dimension reduction method," *Lect. Notes Comput. Sci.*, vol. 36, no. 1, pp. 117–125, 2005. DOI: 10.1007/11538059_13.

[39] H. Zhao, M. Chen, Y. Liu, X. Chen, P. Zhou, and X. Zhang, "Two-source validation of online surface EMG decomposition using progressive FastICA peel-off," *IEEE Trans. Biomed. Eng.*, early access, 2025. DOI: 10.1109/TBME.2025.3538338.

[40] W. Yu, X. Teng, and C. Liu, "Face recognition using discriminant locality preserving projections," *Image Vis. Comput.*, vol. 24, no. 3, pp. 239–248, Mar. 2006. DOI: 10.1016/j.imavis.2005.11.006.

[41] L. Zhu and S. Zhu, "Face recognition based on orthogonal discriminant locality preserving projections," *Neurocomputing*, vol. 70, nos. 7–9, pp. 1543–1546, Mar. 2007. DOI: 10.1016/j.neucom.2006.12.004.

[42] J. Cheng, Q. Liu, H. Lu, and Y.-W. Chen, "Supervised kernel locality preserving projections for face recognition," *Neurocomputing*, vol. 67, pp. 443–449, Aug. 2005. DOI: 10.1016/j.neucom.2004.08.006.

[43] Y.-L. He, K. Li, L.-L. Liang, Y. Xu, and Q.-X. Zhu, "Novel discriminant locality preserving projection integrated with Monte Carlo sampling for fault diagnosis," *IEEE Trans. Rel.*, vol. 72, no. 1, pp. 166–176, Mar. 2023. DOI: 10.1109/TR.2021.3115108.

[44] C. Zhu, H. Yang, X. Jin, K. Xu, and W. Shen, "Locality preserving projections-based spatiotemporal modeling of the temperature distribution of lithium-ion batteries," *IEEE Trans. Ind. Inform.*, vol. 20, no. 1, pp. 179–189, Jan. 2024. DOI: 10.1109/TII.2023.3257311.

[45] R. Ran, J. Feng, Z. Li, J. Wang, and B. Fang, "Locality preserving projections with autoencoder," *Expert Syst. Appl.*, vol. 242, 122750, May 2024. DOI: 10.1016/j.eswa.2023.122750.

[46] H.-X. Zhao, H.-J. Xing, X.-Z. Wang, and J.-F. Chen, "L1-norm-based 2DLPP," in *Proc. 2011 Chin. Control Decis. Conf. (CCDC)*, Mianyang, China, 2011, pp. 1259–1264. DOI: 10.1109/CCDC.2011.5968382.

[47] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, San Diego, CA, USA, 2005, pp. 886–893. DOI: 10.1109/CVPR.2005.177.

[48] Y. Pang, Y. Yuan, and X. Li, "Gabor-based region covariance matrices for face recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 7, pp. 989–993, Jul. 2008. DOI: 10.1109/TCSVT.2008.924108.

[49] S. Ghosh, P. Paral, and A. Chatterjee, "Weight thresholded regularized robust coding for vision sensing-based hand gesture detection in collaborative robotics," *Phys. Scr.*, vol. 98, no. 11, 116002, Oct. 2023. DOI: 10.1088/1402-4896/acfe5a.

[50] J. M. Bioucas-Dias and M. A. T. Figueiredo, "Multiplicative noise removal using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 7, pp. 1720–1730, Jul. 2010. DOI: 10.1109/TIP.2010.2045029.

[51] C. Y. Wong, L. Vergez, and W. Suleiman, "Vision- and tactile-based continuous multimodal intention and attention recognition for safer physical human–robot interaction," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 3, pp. 3205–3215, Jul. 2024. DOI: 10.1109/TASE.2023.3276856.

[52] S. M. Sarhan, M. Z. Al-Faiz, and A. M. Takhakh, "A review on EMG/EEG based control scheme of upper limb rehabilitation robots for stroke patients," *Heliyon*, vol. 9, no. 8, e18308, Aug. 2023. DOI: 10.1016/j.heliyon.2023.e18308.

[53] L. Bi, A. G. Feleke, and C. Guan, "A review on EMG-based motor intention prediction of continuous human upper limb motion for human–robot collaboration," *Biomed. Signal Process. Control*, vol. 51, pp. 113–127, May 2019. DOI: 10.1016/j.bspc.2019.02.011.

[54] J. P. Vásconez, L. I. Barona López, Á. L. Valdivieso Caraguay, and M. E. Benalcázar, "A comparison of EMG-based hand gesture recognition systems based on supervised and reinforcement learning," *Eng. Appl. Artif. Intell.*, vol. 123, 106327, Aug. 2023. DOI: 10.1016/j.engappai.2023.106327.

[55] Z. Ilyas, K. Anam, Widjonarko, C. Avian, A. Z. Muttaqin, and M. E. Ramadhan, "Evaluation of gated-recurrent unit for estimating finger-joint angle using surface electromyography signal," in *Proc. 9th Int. Conf. Electr. Eng., Comput. Sci. Informat. (EECSI)*, Jakarta, Indonesia, 2022, pp. 25–28. DOI: 10.23919/EECSI56542.2022.9946461.

[56] M. B. I. Reaz, M. S. Hussain, and F. Mohd-Yasin, "Techniques of EMG signal analysis: Detection, processing, classification and applications," *Biol. Proced. Online*, vol. 8, pp. 11–35, 2006. DOI: 10.1251/bpo115.

[57] N. M. Esa, A. M. Zain, and M. Bahari, "Electromyography (EMG) based classification of finger movements using SVM," *Int. J. Innov. Comput.*, vol. 8, no. 3, 2018. DOI: 10.11113/ijic.v8n3.181.

[58] C. Tapia, O. Daud, and J. Ruiz-del-Solar, "EMG signal filtering based on independent component analysis and empirical mode decomposition for estimation of motor activation patterns," *J. Med. Biol. Eng.*, vol. 37, pp. 140–155, 2017. DOI: 10.1007/s40846-016-0201-5.

[59] F. Nie, X. Wang, and H. Huang, "Clustering and projected clustering with adaptive neighbors," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. (KDD)*, New York, NY, USA, Aug. 2014, pp. 977–986. DOI: 10.1145/2623330.2623726.

[60] C. Gao, Y. Wang, J. Zhou, W. Ding, L. Shen, and Z. Lai, "Possibilistic neighborhood graph: A new concept of similarity graph learning," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 7, no. 6, pp. 1636–1650, Dec. 2023. DOI: 10.1109/TETCI.2022.3225173.

[61] Q. Yu, R. Wang, B. N. Li, X. Yang, and M. Yao, "Robust locality preserving projections with cosine-based dissimilarity for linear dimensionality reduction," *IEEE Access*, vol. 5, pp. 2676–2684, 2017. DOI: 10.1109/ACCESS.2016.2616584.

[62] T. Long, Y. Sun, J. Gao, Y. Hu, and B. Yin, "Locality preserving projection based on Euler representation," *J. Vis. Commun. Image Represent.*, vol. 70, 102796, Jul. 2020. DOI: 10.1016/j.jvcir.2020.102796.

[63] B. Wang, Y. Hu, J. Gao, Y. Sun, H. Chen, M. Ali, and B. Yin, "Locality preserving projections for Grassmann manifold," in *Proc. 26th Int. Joint Conf. Artif. Intell. (IJCAI)*, Aug. 2017, pp. 2893–2900. DOI: 10.48550/arXiv.1704.08458.

[64] D. Rodríguez-Guerra, G. Sorrosal, I. Cabanes, and C. Calleja, "Human–robot interaction review: Challenges and solutions for modern industrial environments," *IEEE Access*, vol. 9, pp. 108557–108578, 2021. DOI: 10.1109/ACCESS.2021.3099287.

[65] Q. Guo, Z. Yang, J. Xu, Y. Jiang, W. Wang, Z. Liu, W. Zhao, and Y. Sun, "Progress, challenges and trends on vision sensing technologies in automatic/intelligent robotic welding: State-of-the-art review," *Robot. Comput.-Integr. Manuf.*, vol. 89, 102767, Oct. 2024. DOI: 10.1016/j.rcim.2024.102767.

[66] O. W. Samuel *et al.*, "Intelligent EMG pattern recognition control method for upper-limb multifunctional prostheses: Advances, current challenges, and future prospects," *IEEE Access*, vol. 7, pp. 10150–10165, 2019. DOI: 10.1109/ACCESS.2019.2891350.

[67] P. Tsarouchi, S. Makris, and G. Chryssolouris, "Human–robot interaction review and challenges on task planning and programming," *Int. J. Comput. Integr. Manuf.*, vol. 29, no. 8, pp. 916–931, 2016. DOI: 10.1080/0951192X.2015.1130251.

[68] J. Fan, P. Zheng, and S. Li, "Vision-based holistic scene understanding towards proactive human–robot collaboration," *Robot. Comput.-Integr. Manuf.*, vol. 75, 102304, Jun. 2022. DOI: 10.1016/j.rcim.2021.102304.

[69] D. Wei, L. Chen, L. Zhao, H. Zhou, and B. Huang, "A vision-based measure of environmental effects on inferring human intention during human–robot interaction," *IEEE Sensors J.*, vol. 22, no. 5, pp. 4246–4256, Mar. 1, 2022. DOI: 10.1109/JSEN.2021.3139593.

[70] J. Wen, Z. Zhong, Z. Zhang, L. Fei, Z. Lai, and R. Chen, "Adaptive locality preserving regression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 1, pp. 75–88, Jan. 2020. DOI: 10.1109/TCSVT.2018.2889727.

[71] M. Oudah, A. Al-Naji, and J. Chahl, "Hand gesture recognition based on computer vision: A review of techniques," *J. Imaging*, vol. 6, no. 8, 73, Jul. 2020. DOI: 10.3390/jimaging6080073.

[72] "Raspberry Pi Documentation Webpage," accessed Mar. 2025. [Online]. Available: https://www.raspberrypi.com/products/raspberry-pi-3-model-b-plus/

[73] M. Y. Arafat, M. M. Alam, and S. Moh, "Vision-based navigation techniques for unmanned aerial vehicles: Review and challenges," *Drones*, vol. 7, no. 2, 89, Jan. 2023. DOI: 10.3390/drones7020089.

[74] Z. Huang, Z. Tang, X. Zhang, L. Ruan, and X. Zhang, "Perceptual image hashing with locality preserving projection for copy detection," *IEEE Trans. Dependable Secure Comput.*, vol. 20, no. 1, pp. 463–477, Jan.–Feb. 2023. DOI: 10.1109/TDSC.2021.3136163.

[75] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967. DOI: 10.1109/TIT.1967.1053964.

[76] F.-L. Fan, J. Xiong, M. Li, and G. Wang, "On interpretability of artificial neural networks: A survey," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 5, no. 6, pp. 741–760, Nov. 2021. DOI: 10.1109/TRPMS.2021.3066428.

[77] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Schölkopf, "Support vector machines," *IEEE Intell. Syst. Their Appl.*, vol. 13, no. 4, pp. 18–28, Jul.–Aug. 1998. DOI: 10.1109/5254.708428.

[78] Y. Liu, J.-W. Bi, and Z.-P. Fan, "A method for multi-class sentiment classification based on an improved one-vs-one (OVO) strategy and the support vector machine (SVM) algorithm," *Inf. Sci.*, vols. 394–395, pp. 38–52, Jul. 2017. DOI: 10.1016/j.ins.2017.02.016.

[79] J.-H. Hong, J.-K. Min, U.-K. Cho, and S.-B. Cho, "Fingerprint classification using one-vs-all support vector machines dynamically ordered with naïve Bayes

classifiers," *Pattern Recognit.*, vol. 41, no. 2, pp. 662–671, Feb. 2008. DOI: 10.1016/j.patcog.2007.07.004.

[80] H. Keskes and A. Braham, "Recursive undecimated wavelet packet transform and DAG SVM for induction motor diagnosis," *IEEE Trans. Ind. Inform.*, vol. 11, no. 5, pp. 1059–1066, Oct. 2015. DOI: 10.1109/TII.2015.2462315.

[81] V. Rokhlin, A. Szlam, and M. Tygert, "A randomized algorithm for principal component analysis," *SIAM J. Matrix Anal. Appl.*, vol. 31, no. 3, pp. 1100–1124, 2010. DOI: 10.1137/080736417.

[82] J. Weng, Y. Zhang, and W.-S. Hwang, "Candid covariance-free incremental principal component analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 1034–1040, Aug. 2003. DOI: 10.1109/TPAMI.2003.1217609.

[83] B. Schölkopf, A. Smola, and K.-R. Müller, "Kernel principal component analysis," in *Proc. Int. Conf. Artif. Neural Netw. (ICANN)*, 1997, pp. 583–588. DOI: 10.1007/BFb0020217.

[84] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *J. Comput. Graph. Stat.*, vol. 15, no. 2, pp. 265–286, 2006. DOI: 10.1198/106186006X113430.

[85] X.-Y. Zhang, L. Wang, S. Xiang, and C.-L. Liu, "Retargeted least squares regression algorithm," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 9, pp. 2206–2213, Sep. 2015. DOI: 10.1109/TNNLS.2014.2371492.

[86] J. Wen, N. Han, X. Fang, L. Fei, K. Yan, and S. Zhan, "Low-rank preserving projection via graph regularized reconstruction," *IEEE Trans. Cybern.*, vol. 49, no. 4, pp. 1279–1291, Apr. 2019. DOI: 10.1109/TCYB.2018.2799862.

[87] H. Xue, S. Chen, and Q. Yang, "Discriminatively regularized least-squares classification," *Pattern Recognit.*, vol. 42, no. 1, pp. 93–104, Jan. 2009. DOI: 10.1016/j.patcog.2008.07.010.

[88] Y. Yang, H. T. Shen, Z. Ma, Z. Huang, and X. Zhou, "$l_{2,1}$-norm regularized discriminative feature selection for unsupervised learning," in *Proc. 22nd Int. Joint Conf. Artif. Intell. (IJCAI)*, Jul. 2011, pp. 1589–1594. DOI: 10.5591/978-1-57735-516-8/IJCAI11-267.

[89] X. Li, M. Chen, F. Nie, and Q. Wang, "Locality adaptive discriminant analysis," in *Proc. 26th Int. Joint Conf. Artif. Intell. (IJCAI)*, Aug. 2017, pp. 2201–2207. DOI: 10.24963/ijcai.2017/306.

[90] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 533–544, Feb. 2010. DOI: 10.1109/TIP.2009.2035882.

[91] S. Murala, R. P. Maheshwari, and R. Balasubramanian, "Local tetra patterns: A new feature descriptor for content-based image retrieval," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2874–2886, May 2012. DOI: 10.1109/TIP.2012.2188809.

[92] A. K. Tiwari, V. Kanhangad, and R. B. Pachori, "Histogram refinement for texture descriptor based image retrieval," *Signal Process. Image Commun.*, vol. 53, pp. 73–85, Apr. 2017. DOI: 10.1016/j.image.2017.01.010.

[93] T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkilä, "Recognition of blurred faces using local phase quantization," in *Proc. 19th Int. Conf. Pattern Recognit. (ICPR)*, Tampa, FL, USA, 2008, pp. 1–4. DOI: 10.1109/ICPR.2008.4761847.

[94] X. Yu and J. Li, "Adaptive Kalman filtering for recursive both additive noise and multiplicative noise," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 3, pp. 1634–1649, Jun. 2022. DOI: 10.1109/TAES.2021.3117896.

[95] U. Tigga and J. Jha, "Image deblurring with impulse noise using alternating direction method of multipliers and Lucy-Richardson method," in *Proc. 8th Int. Conf. Comput. Intell. Commun. Netw. (CICN)*, Tehri, India, 2016, pp. 230–235. DOI: 10.1109/CICN.2016.52.

[96] S. Banerjee, D. Sarkar, D. Chatterjee, and S. R. Chowdhuri, "High-density salt and pepper noise removal from colour images by introducing new enhanced filter," in *Proc. 2021 Int. Conf. Intell. Technol. (CONIT)*, Hubli, India, 2021, pp. 1–6. DOI: 10.1109/CONIT51480.2021.9498402.

[97] T. Yassine, S. Ahmed, and N. Abdelkrim, "Speckle noise reduction in digital speckle pattern interferometry using Riesz wavelets transform," in *Proc. 2017 Int. Conf.*

Adv. Technol. Signal Image Process. (ATSIP), Fez, Morocco, 2017, pp. 1–4. DOI: 10.1109/ATSIP.2017.8075565.

[98] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Inf. Theory*, vol. 8, no. 2, pp. 179–187, Feb. 1962. DOI: 10.1109/TIT.1962.1057692.

[99] M. U. Ali, S. Ahmed, J. Ferzund, A. Mehmood, and A. Rehman, "Using PCA and factor analysis for dimensionality reduction of bio-informatics data," arXiv preprint, arXiv:1707.07189, Jul. 2017. DOI: 10.48550/arXiv.1707.07189.

[100] Q. Liu, H. Lu, and S. Ma, "Improving kernel Fisher discriminant analysis for face recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 42–49, Jan. 2004. DOI: 10.1109/TCSVT.2003.818352.

[101] Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimensionality reduction via tangent space alignment," *SIAM J. Sci. Comput.*, vol. 26, no. 1, pp. 313–338, 2004. DOI: 10.1137/S1064827502419154.

[102] S. K. Pal, B. U. Shankar, and P. Mitra, "Granular computing, rough entropy and object extraction," *Pattern Recognit. Lett.*, vol. 26, no. 16, pp. 2509–2517, Dec. 2005. DOI: 10.1016/j.patrec.2005.05.007.

[103] D. Chakraborty, B. U. Shankar, and S. K. Pal, "Granulation, rough entropy and spatiotemporal moving object detection," *Appl. Soft Comput.*, vol. 13, no. 9, pp. 4001–4009, Sep. 2013. DOI: 10.1016/j.asoc.2012.09.003.

[104] S. K. Pal, D. Bhoumik, and D. Bhunia Chakraborty, "Granulated deep learning and Z-numbers in motion detection and object recognition," *Neural Comput. Appl.*, vol. 32, pp. 16533–16548, 2020. DOI: 10.1007/s00521-019-04200-1.

[105] D. B. Chakraborty and S. K. Pal, "Neighborhood granules and rough rule-base in tracking," *Nat. Comput.*, vol. 15, pp. 359–370, 2016. DOI: 10.1007/s11047-015-9493-6.

[106] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning About Data*. Dordrecht, The Netherlands: Kluwer, 1991. DOI: 10.1007/978-94-011-3534-4.

[107] T. J. Vennila and V. Balamurugan, "A rough set framework for multihuman tracking in surveillance video," *IEEE Sensors J.*, vol. 23, no. 8, pp. 8753–8760, Apr. 15, 2023. DOI: 10.1109/JSEN.2023.3242007.

[108] L. A. Zadeh, "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Sets Syst.*, vol. 90, no. 2, pp. 111–127, Sep. 1997. DOI: 10.1016/S0165-0114(97)00077-8.

[109] W. Pedrycz, *Granular Computing: An Emerging Paradigm*. Heidelberg, Germany: Physica-Verlag, 2001. (Studies in Fuzziness and Soft Computing, vol. 70). DOI: 10.1007/978-3-7908-1823-9.

[110] B. Uma Shankar and D. Chakraborty, "Spatiotemporal Approach for Tracking Using Rough Entropy and Frame Subtraction," in *Pattern Recognition and Machine Intelligence (PReMI 2011)*, Lecture Notes in Computer Science, vol. 6744, Springer, 2011, pp. 267–273. DOI: 10.1007/978-3-642-21786-9_33.

[111] G. Finlayson, S. Hordley, G. Schaefer, and G. Y. Tian, "Illuminant and device invariant colour using histogram equalisation," *Pattern Recognit.*, vol. 38, no. 2, pp. 179–190, Feb. 2005. DOI: 10.1016/j.patcog.2004.04.010.

[112] A. Alahi, R. Ortiz, and P. Vandergheynst, "FREAK: Fast retina keypoint," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Providence, RI, USA, 2012, pp. 510–517. DOI: 10.1109/CVPR.2012.6247715.

[113] P. Alcantarilla, J. Nuevo, and A. Bartoli, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, Bristol, U.K., 2013, pp. 13.1–13.11. DOI: 10.5244/C.27.13.

[114] P. Paral, S. Ghosh, S. K. Pal, and A. Chatterjee, "Adaptive non-homogeneous granulation-aided density-based deep feature clustering for far infrared sign language images," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 9, no. 2, pp. 1269–1280, Apr. 2025. DOI: 10.1109/TETCI.2024.3510292.

[115] R. D. Gow *et al.*, "A comprehensive tool for modeling CMOS image-sensor-noise performance," *IEEE Trans. Electron Devices*, vol. 54, no. 6, pp. 1321–1329, Jun. 2007. DOI: 10.1109/TED.2007.896718.

[116] Q. Hu, D. Yu, J. Liu, and C. Wu, "Neighborhood rough set based heterogeneous feature subset selection," *Inf. Sci.*, vol. 178, no. 18, pp. 3577–3594, Sep. 15, 2008. DOI: 10.1016/j.ins.2008.05.024.

[117] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd Int. Conf. Knowl. Discov. Data Min. (KDD)*, Portland, OR, USA, Aug. 1996, pp. 226–231. https://dl.acm.org/doi/10.5555/3001460.3001507.

[118] R. W. Cooksey and G. N. Soutar, "Coefficient Beta and hierarchical item clustering: An analytical procedure for establishing and displaying the dimensionality and homogeneity of summated scales," *Organ. Res. Methods*, vol. 9, no. 1, pp. 78–98, Jan. 2006. DOI: 10.1177/1094428105283939.

[119] A. K. Singh, S. Mittal, P. Malhotra, and Y. V. Srivastava, "Clustering evaluation by Davies–Bouldin Index (DBI) in cereal data using K-means," in *Proc. 4th Int. Conf. Comput. Methodol. Commun. (ICCMC)*, Erode, India, 2020, pp. 306–310. DOI: 10.1109/ICCMC48092.2020.ICCMC-00057.

[120] A. Lee, G. Kim, S.-J. Hong, S.-W. Kim, and G. Kim, "Classification of dead cocoons using convolutional neural networks and machine learning methods," *IEEE Access*, vol. 11, pp. 137317–137327, 2023. DOI: 10.1109/ACCESS.2023.3338540.

[121] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Regularized robust coding for face recognition," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1753–1766, May 2013. DOI: 10.1109/TIP.2012.2235849.

[122] H. Zheng, D. Lin, L. Lian, J. Dong, and P. Zhang, "Laplacian-uniform mixture-driven iterative robust coding with applications to face recognition against dense errors," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 9, pp. 3620–3633, Sep. 2020. DOI: 10.1109/TNNLS.2019.2945372.

[123] "Muscle Sensor v3 Datasheet," accessed Mar. 2025. [Online]. Available: https://www.pololu.com/file/0J745/Muscle_Sensor_v3_users_manual.pdf

[124] R. Djehaiche, S. Aidel, A. Sawalmeh, N. Saeed, and A. H. Alenezi, "Adaptive control of IoT/M2M devices in smart buildings using heterogeneous wireless

networks," *IEEE Sensors J.*, vol. 23, no. 7, pp. 7836–7849, Apr. 1, 2023. DOI: 10.1109/JSEN.2023.3247007.

[125] H. Ghapanchizadeh, S. A. Ahmad, and A. J. Ishak, "Recommended surface EMG electrode position for wrist extension and flexion," in *Proc. IEEE Student Symp. Biomed. Eng. Sci. (ISSBES)*, Shah Alam, Malaysia, 2015, pp. 108–112. DOI: 10.1109/ISSBES.2015.7435877.

[126] L. A. Green, J. McGuire, and D. A. Gabriel, "Flexor carpi radialis surface electromyography electrode placement for evoked and voluntary measures," *Muscle & Nerve*, vol. 52, no. 5, pp. 818–825, Nov. 2015. DOI: 10.1002/mus.24631.

[127] A. O. Perotto, *Anatomical Guide for the Electromyographer*, 5th ed. Springfield, IL, USA: Charles C. Thomas, 2011.

[128] G. Li *et al.*, "Multi-view fusion network-based gesture recognition using sEMG data," *IEEE J. Biomed. Health Inform.*, vol. 28, no. 8, pp. 4432–4443, Aug. 2024. DOI: 10.1109/JBHI.2023.3287979.

[129] M. Zandigohar *et al.*, "Multimodal fusion of EMG and vision for human grasp intent inference in prosthetic hand control," *Front. Robot. AI*, vol. 11, Feb. 2024. DOI: 10.3389/frobt.2024.1312554.

[130] D. M. Perera and D. G. K. Madusanka, "Vision-EMG fusion method for real-time grasping pattern classification system," in *Proc. 2021 Moratuwa Eng. Res. Conf. (MERCon)*, Moratuwa, Sri Lanka, 2021, pp. 585–590. DOI: 10.1109/MERCon52712.2021.9525702.

[131] D. Lee, D. You, G. Cho, H. Lee, E. Shin, T. Choi, S. Kim, S. Lee, and W. Nam, "EMG-based hand gesture classifier robust to daily variation: Recursive domain adversarial neural network with data synthesis," *Biomed. Signal Process. Control*, vol. 88, pt. B, 105600, Feb. 2024. DOI: 10.1016/j.bspc.2023.105600.

[132] Y. Du, W. Jin, W. Wei, Y. Hu, and W. Geng, "Surface EMG-based inter-session gesture recognition enhanced by deep domain adaptation," *Sensors*, vol. 17, no. 3, 458, 2017. DOI: 10.3390/s17030458.

[133] Y. Xu, Y. Yu, Z. Zhao, C. Chen, and X. Sheng, "Cumulative spike train estimation for muscle excitation assessment from surface EMG using spatial spike detection," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 11, pp. 5335–5344, Nov. 2023. DOI: 10.1109/JBHI.2023.3309662.

[134] W. Batayneh, E. Abdulhay, and M. Alothman, "Comparing the efficiency of artificial neural networks in sEMG-based simultaneous and continuous estimation of hand kinematics," *Digit. Commun. Netw.*, vol. 8, no. 2, pp. 162–173, Apr. 2022. DOI: 10.1016/j.dcan.2021.08.002.

[135] M. Mohammadiazni, J. G. C. Alfaro, and A. L. Trejos, "Mitigate the effect of arm posture on electromyography pattern recognition," *IEEE J. Biomed. Health Inform.*, vol. 29, no. 4, pp. 2413–2424, Apr. 2025. DOI: 10.1109/JBHI.2024.3518978.

[136] R. Alazrai, M. I. Daoud, A. Khalifeh, N. Alnuman, Y. Mowafi, and D. Alabed, "A wavelet-based approach for estimating the joint angles of the fingers and wrist using electromyography signals," in *New Technologies to Improve Patient Rehabilitation*, Springer, 2019, vol. 1002, *Commun. Comput. Inf. Sci.*, pp. 29–44. DOI: 10.1007/978-3-030-16785-1_3.

[137] Z. Taghizadeh, S. Rashidi, and A. Shalbaf, "Finger movements classification based on fractional Fourier transform coefficients extracted from surface EMG signals," *Biomed. Signal Process. Control*, vol. 68, p. 102573, Jul. 2021. DOI: 10.1016/j.bspc.2021.102573.

[138] X. Zhai, B. Jelfs, R. H. M. Chan, and C. Tin, "Self-recalibrating surface EMG pattern recognition for neuroprosthesis control based on convolutional neural network," *Front. Neurosci.*, vol. 11, Art. no. 379, Jul. 2017. DOI: 10.3389/fnins.2017.00379.

[139] K. Anam, C. Avian, D. I. Swasono, A. Z. Muttaqin, and H. Ismail, "Estimation of finger joint movement based on electromyography signal using long short-term memory," in *Proc. Int. Conf. Comput. Eng., Netw., Intell. Multimedia (CENIM)*, Surabaya, Indonesia, 2020, pp. 86–90. DOI: 10.1109/CENIM51130.2020.9298023.

[140] W. Guo, X. Sheng, H. Liu, and X. Zhu, "Mechanomyography assisted myoelectric sensing for upper-extremity prostheses: A hybrid approach," *IEEE Sensors J.*, vol. 17, no. 10, pp. 3100–3108, May 15, 2017. DOI: 10.1109/JSEN.2017.2679806.

[141] A. Saikia, N. M. Kakoty, N. Phukan, M. Balakrishnan, N. Sahai, S. Paul, and D. Bhatia, "Combination of EMG features and stability index for finger movements recognition," *Procedia Comput. Sci.*, vol. 133, pp. 92–98, 2018. DOI: 10.1016/j.procs.2018.07.012.

[142] J. R. Torres-Castillo, C. O. López-López, and M. A. Padilla-Castañeda, "Neuromuscular disorders detection through time-frequency analysis and classification of multi-muscular EMG signals using Hilbert-Huang transform," *Biomed. Signal Process. Control*, vol. 71, pt. A, 103037, Jan. 2022. DOI: 10.1016/j.bspc.2021.103037.

[143] D. Li, P. Kang, Y. Yu, and P. B. Shull, "Graph-driven simultaneous and proportional estimation of wrist angle and grasp force via high-density EMG," *IEEE J. Biomed. Health Inform.*, vol. 28, no. 5, pp. 2723–2732, May 2024. DOI: 10.1109/JBHI.2024.3373432.

[144] S. Liwicki, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "Euler principal component analysis," *Int. J. Comput. Vis.*, vol. 101, pp. 498–518, 2013. DOI: 10.1007/s11263-012-0558-z.

[145] M. Harandi, C. Sanderson, C. Shen, and B. Lovell, "Dictionary learning and sparse coding on Grassmann manifolds: An extrinsic solution," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Sydney, NSW, Australia, 2013, pp. 3120–3127. DOI: 10.1109/ICCV.2013.387.

[146] X. Lu, J. Long, J. Wen, L. Fei, B. Zhang, and Y. Xu, "Locality preserving projection with symmetric graph embedding for unsupervised dimensionality reduction," *Pattern Recognit.*, vol. 131, 108844, Nov. 2022. DOI: 10.1016/j.patcog.2022.108844.

[147] M. Atzori *et al.*, "Building the Ninapro database: A resource for the biorobotics community," in *Proc. 4th IEEE RAS & EMBS Int. Conf. Biomed. Robot. Biomechatronics (BioRob)*, Rome, Italy, 2012, pp. 1258–1265. DOI: 10.1109/BioRob.2012.6290287.

[148] K. D. Koutroumbas, S. D. Xenaki, and A. A. Rontogiannis, "On the convergence of the sparse possibilistic C-means algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 1, pp. 324–337, Feb. 2018. DOI: 10.1109/TFUZZ.2017.2659739.

[149] L. Zelnik-Manor and P. Perona, "Self-tuning spectral clustering," in *Proc. 18th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, Vancouver, Canada, Dec. 2004, pp. 1601–1608. DOI: 10.5555/2976040.2976241.

Saibal Ghosh
10/11/2025

Amitava Chatterjee
10/11/2025

Professor
Electrical Engg. Deptt.
Jadavpur University
Kolkata - 700 032

Sugata Munshi
10.11.2025

Retd.-**Professor**
Electrical Engg. Deptt.
Jadavpur University
Kolkata - 700 032