HYBRID-UNIFORM MIXTURE MODEL BASED ITERATIVE ROBUST CODING FOR IMAGE RECOGNITION PROBLEMS USING THERMAL IMAGING

Thesis Submitted in the Partial Fulfilment of the Requirements for the Degree of

MASTER OF ELECTRICAL ENGINEERING

by

ANINDA SUNDAR MONDAL

(Registration No. : **136717 of 2016-17)**

(Exam Roll No. : M4ELE23012**)**

Under the guidance of

Dr. AMITAVA CHATTERJEE

Electrical Engineering Department

JADAVPUR UNIVERSITY

KOLKATA, INDIA

2021-2023

HYBRID-UNIFORM MIXTURE MODEL BASED ITERATIVE ROBUST CODING FOR IMAGE RECOGNITION PROBLEMS USING THERMAL IMAGING

Thesis Submitted in the Partial Fulfilment of the Requirements for the Degree of

MASTER OF ELECTRICAL ENGINEERING

by

ANINDA SUNDAR MONDAL

Class Roll No. : **002110802009**

Registration No. : **136717 of 2016-17**

JADAVPUR UNIVERSITY

Under the guidance of

Prof. (Dr.) AMITAVA CHATTERJEE

Electrical Engineering Department

Faculty of Engineering and Technology

JADAVPUR UNIVERSITY

KOLKATA- 700032, W.B. INDIA

2021-2023

JADAVPUR UNIVERSITY

FACULTY OF ENGINEERING AND TECHNOLGY

ELECTRICAL ENGINEERING DEPARTMENT

KOLKATA- 700032, INDIA

CERTIFICATE OF RECOMMENDATION

This is to certify that, Mr. **ANINDA SUNDAR MONDAL** (Registration No. 136717 of 2016-17), has completed and submitted his thesis entitled, "**HYBRID-UNIFORM MIXTURE MODEL BASED ITERATIVE ROBUST CODING FOR IMAGE RECOGNITION PROBLEMS USING THERMAL IMAGING**", in partial fulfillment of the requirements for the degree of "Master of Electrical Engineering" of Jadavpur University. The thesis work has been carried out by him under my guidance and supervision. The project, in my opinion, is worthy of its acceptance.

SUPERVISOR

_____

Prof. (Dr.) Amitava Chatterjee
Professor
Electrical Engineering Department,
Jadavpur University

COUNTERSIGNED

_____

Prof. (Dr.) Biswanath Roy
Head, Dept. of Electrical Engineering,
Faculty of Engineering and Technology,
Jadavpur University

_____

Prof. (Dr.) Ardhendu Ghoshal
Dean,
Faculty of Engineering and Technology,
Jadavpur University

JADAVPUR UNIVERSITY

KOLKATA- 700032, INDIA

FACULTY OF ENGINEERING AND TECHNOLOGY

ELECTRICAL ENGINEERING DEPARTMENT

CERTIFICATE OF APPROVAL*

*The foregoing thesis is hereby approved as a creditable study of the **Master of Electrical Engineering** under the Electrical Engineering department and presented in a manner satisfactory to warrant its acceptance as a prerequisite to the degree for which it has been submitted. By this approval, the undersigned does not necessarily endorse or approve any statement made, opinion expressed, or conclusion therein but approve this thesis only for the purpose for which it is submitted.*

Final Examination for Evaluation of the Thesis

1. _____

2. _____

(Signature of Examiners)

*Only in case the thesis is approved

DECLARATION OF ORIGINALITY AND COMPLIANCE OF ACADEMIC ETHICS

I hereby declare that the thesis entitled "**HYBRID-UNIFORM MIXTURE MODEL BASED ITERATIVE ROBUST CODING FOR IMAGE RECOGNITION PROBLEMS USING THERMAL IMAGING**" contains a literature survey and original research work as part of the course of Master of Engineering under Electrical Engineering department. All the information in this document has been obtained and presented in accordance with academic rules and ethical conduct.

I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name	:	**ANINDA SUNDAR MONDAL**
Class Roll no.	:	**002110802009**
Registration no.	:	**136717 of 2016-17**
Thesis Name	:	**HYBRID-UNIFORM MIXTURE MODEL BASED ITERATIVE ROBUST CODING FOR IMAGE RECOGNITION PROBLEMS USING THERMAL IMAGING**
Signature with date	:	

ACKNOWLEDGEMENTS

I would like to sincerely thank my supervisor, **Professor Amitava Chatterjee** from the Department of Electrical Engineering at Jadavpur University, for granting me the valuable opportunity to engage in this captivating research field. I am deeply appreciative and indebted to him for his constant guidance and recommendations during the project. His recognition and support throughout the undertaking served as a great source of confidence for me. I am incredibly fortunate to have Professor Amitava Chatterjee as my thesis supervisor, and I will always be grateful for his mentorship.

I would like to extend my heart thank **Professor Sugata Munshi**, **Professor Debangshu Dey, Professor Biswajit Bhattacharyya, Professor Mita Dutta, Professor Gautam Sarkar,** and **Professor Palash Kr. Kundu** from the Department of Electrical Engineering at Jadavpur University for their unwavering support, counsel, and inspiration throughout my coursework. Their encouragement, guidance, and motivation have been invaluable to me.

I am also thankful to **Professor Biswanath Roy**, Head, Department of Electrical Engineering, Jadavpur University, for providing the necessary facilities to carry out this research work.

I would like to seize this moment to express my sincere gratitude to **Mr. Saibal Ghosh**, a Research Scholar at Jadavpur University, for his invaluable contributions to this project. I am also grateful to the other research scholars of the Measurement laboratory, including **Mrs. Antara Ghosh, Mrs. Sayanti Chaudhuri, Mr. Anadi Biswas** and others, for their unwavering support throughout my research journey. Their assistance has been greatly appreciated during the entire duration of my project.

I would like to express my heartfelt gratitude to my dear friends, **Mr. Ashwin Rai, Mr. Rohon Das** and other PG scholars in the E.E. Department for their unwavering support, inexplicable encouragement, and assistance have been invaluable to me throughout this journey. Additionally, I am thankful to the undergraduate students of Electrical Engineering at Jadavpur University who volunteered to assist with the database acquisition for my thesis work. Lastly, I extend my deepest appreciation to my parents for their constant support and motivation, without whom I would not have reached this point.

I am immensely grateful to the divine force, the inherent essence of nature, for safeguarding my well-being and enabling the successful completion of this undertaking.

Thanking all,

Aninda Sundar Mondal

ABSTRACT

In a human-robot coexisting environment, vision-sensing-based hand gesture recognition is considered a significant contemporary research problem in collaborative robotics. The recognition task becomes more challenging in complex environments, where accurate identification of specific hand gestures is crucial for correct robot navigation and task execution. The complexities in such challenging environments may include factors like improper illumination, presence of unwanted objects, occlusions, shadows, silhouette, and specularities.

To address challenges in visual pattern recognition, regression analysis-based techniques have been widely utilized. These techniques aim to reconstruct a query error image or signal based on available training data. However, in the presence of substantial signal corruption, as mentioned earlier, traditional parametric and non-parametric statistical regression methods often yield inferior results. Robust regression approaches are specifically designed to mitigate the adverse effects of complex noises and challenges on the signal recovery process. These approaches enhance the system models to effectively attenuate the impact of such adversities.

Furthermore, in this study, vein pattern recognition of the human forearm is also considered. Vein pattern recognition is an additional area of focus, where robust regression methods can be applied to improve the accuracy and reliability of recognition algorithms, particularly in challenging environments.

In the context of the above discussion, this thesis investigates the application of a robust regression technique called Laplacian Uniform Mixture (LUM). LUM is a robust regression model based on sparse coding that is used for signal or image reconstruction. It is an enhanced version of the Regularized Robust Coding (RRC) model. The LUM model employs an iteratively reweighted regularized approach to achieve an optimal solution. The weight update process is guided by conventional logistic functions in the case of LUM. This thesis thoroughly examines the effectiveness of these techniques, particularly when dealing with training dictionaries that consist of real-world, corrupted hand gesture images.

Environmental considerations, particularly challenging illumination circumstances that can compromise the system's functionality, can significantly impact the recognition ability of vision-based systems utilized in human-robot interaction (HRI). A solution based on far infrared thermal imaging is proposed in this study to mitigate this limitation of RGB cameras in collaborative robotics

applications. The standard American manual alphabet (AMA) library from American Sign Language (ASL) is employed to interpret user input from humans and provide relevant guidance instructions. Sparse representation-based modeling is utilized to construct a robust framework for the entire system, incorporating advancements in image processing. A novel Student's t-uniform mixed error distribution model is introduced to accurately characterize the distribution of coding error residuals. By leveraging the properties of the Student's t-distribution, which exhibits a prominent peak at zero and elongated tails on both sides for higher coding residuals, the error distribution is effectively captured. Through extensive case studies involving normal-conditioned and occluded thermal sign images, the effectiveness of the proposed distribution model in far infrared vision-based sign language detection is demonstrated, even in the presence of environmental challenges such as low-light vision and block occlusion.

Another work of this thesis explores the utilization of thermal images of the human forearm for vein pattern recognition. Two recognition methods, Laplacian Uniform Mixture (LUM) and the proposed Student's t Uniform Mixture (TUM), are investigated. The LUM method captures the complex distribution of vein patterns using a mixture model. However, limitations may arise in challenging lighting conditions or occlusions. In contrast, the proposed method, incorporating the unique properties of the Student's t-distribution, enhances robustness against environmental challenges. Extensive experiments on a dataset of thermal forearm images demonstrate the superior performance of the TUM method in terms of accuracy and robustness. This study contributes to the advancement of reliable and efficient biometric identification systems based on thermal vein patterns.

Finally a retrospection of the entire thesis work is done, and the future research scopes are projected as well.

LIST OF CONTENTS

LIST OF ABBREVIATIONS

	ABBREVIATION	EXPLICATION
1.	HCI	Human Computer Interaction
2.	HRI	Human Robot Interaction
3.	RRC	Regularized Robust Coding
4.	LUM	Laplacian Uniform Mixture
5.	AMA	American Manual Alphabet
6.	ASL	American Sign Language
7.	TUM	Student's t Uniform Mixture
8.	HRC	Human Robot Collaboration
9.	EMG	Electromyography
10.	DSP	Digital Signal Processing
11.	EEG	Electroencephalogram
12.	GUI	Graphical User Interfaces
13.	PCA	Principal Component Analysis
14.	SVM	Support Vector Machine
15.	PSO	Particle Swarm Optimization
16.	HOG	Histogram of Oriented Gradients
17.	ANN	Artificial Neural Network
18.	LSTM	Long Short Term Memory
19.	CNN	Convolutional Neural Network
20.	MLP	Multi-Layer Perceptron
21.	IMUs	Inertial Measurement Units
22.	SRC	Sparse Representation based Classification
23.	DOA	Direction-of-Arrival
24.	KSRC	Kernel Sparse Representation Based Classification
25.	CRC	Collaborative Representation based Classification
26.	OP-CRC	Optimized Projection for Collaborative Representation based Classification
27.	KCRC	Kernel Collaborative Representation Classification
28.	CRP	Collaborative Representation based Projections
29.	RSC	Robust Sparse Coding
30.	MAP	Maximum a Posteriori
31.	GUM	Gaussian Mixture Model
32.	LSM	Laplacian Scale Mixture

33.	HQ	Half-Quadratic
34.	FIR	Far Infrared imaging
35.	MLE	Maximum Likelihood Estimation
36.	PDF	Probability Density Function
37.	CERS	Correntropy-based Sparse Representation
38.	PALM	Primal form of the Augmented Lagrangian Method
39.	FISTA	Fast Iterative Shrinkage-Thresholding Algorithm
40.	IRC	Iterative Robust Coding

LIST OF FIGURES

LIST OF TABLES

CHAPTER-1

INTRODUCTION

- ***BACKGROUND***
- ***MOTIVATION OF THE THESIS***
- ***CONTRIBUTIONS IN THE THESIS***
- ***STRUCTURE OF THE THESIS***

1.1. BACKGROUND

The potential and advantages of human-robot collaboration (HRC), which has recently captured the attention of the scientific and business worlds, are being recognized. By being assisted, weariness and tension may be decreased by the robots, human skills can be strengthened, workplace accidents can be minimized, and rehabilitative therapies can be automated, and the overall standard of life can be enhanced [1]. On another note, expertise can be provided by humans, information can be imparted, and robot functionality can be overseen, thereby adding a degree of flexibility to the method and aiding in the successful completion of a wide range of activities [2]. The concept of distant robotics is currently being evolved into the more useful idea of collaborative robotics, in which people will be assisted and supported by robots in both industrial and household applications.

A wide range of topics is currently being spanned by human-robot interaction (HRI) research and design.

There are essentially four fields of application for HRI:

1. Mundane jobs are performed by robots under human supervision. These jobs involve handling components on production assembly lines and delivering goods, parts, mail, and medications to offices, hospitals, and warehouses. The devices, known as telerobots, have the ability to automatically execute a specific set of tasks based on a computer program. They can also sense their surroundings and their own joint positions and communicate this information back to a human operator, who updates the computer program as necessary.

2. Vehicles in space, the air, the ground, and the sea can be remotely controlled to perform unusual jobs in dangerous or inaccessible settings. If objects are moved and manipulated in a remote physical environment in response to continuous control motions made by the remote human, they are referred to as teleoperators. A machine is considered a telerobot if a computer is periodically reprogrammed by a human supervisor to carry out portions of the overall task.

3. Passengers are carried by automated vehicles, such as commercial aircraft and automated rail and road vehicles.

4. Social contact between humans and robots, including robots that entertain, educate, comfort, and assist the elderly, autistic, and crippled, is facilitated [3].

In collaborative robotics, the existence of a method for information to be shared between humans and robots working together in the same space is required. These kinds of communication can be classified into two main categories: verbal interactions and non-verbal interactions. Verbal interaction encompasses speech acts, situated language, emotive speech, and a fixed set of prefabricated directives. On the other hand, non-verbal interaction techniques can manifest in various forms, including head nods, head movements, facial expressions, and hand gestures [4]. Over the years, hand gestures have received significant attention from scholars across all communication channels. Hand and other gestures have been one of the earliest forms of communication between humans and other species, dating back before any form of vocal interaction was invented. They are considered a natural result of human instinct [5]. While various forms of interpersonal interaction, such as written, spoken, and visual, vary greatly across social, cultural, and geographical regions, hand gestures and various types of human body gestures convey essentially the same message universally. Hand gestures used during interaction encompass different forms, including salutations, indicating, miming, commanding, and others, each conveying a distinct message and course of movement [6]. Certain hand gestures are employed by military service or navy workers to convey specialized information [5]. Due to their universal understanding, transparency, and accessibility, gestures are increasingly being adopted more quickly in the creation of interaction scenarios between humans and robots (HRI) or human-computer interaction (HCI). Non-verbal movements in sign language are utilized by individuals to express their thoughts and feelings. However, comprehension of sign language poses significant challenges for non-signers. The recognition of sign language has long been a prominent area of study. In the initial stages, communication with a robot was solely through programming, requiring substantial effort. Gesture recognition has been developed in conjunction with advancements in science and robotics.

1.2. MOTIVATION OF THE THESIS

Gestures can originate from any body movement or state, although they are typically initiated from the hand or facial region. Gesture recognition is a method by which the body language of humans can be understood by a computer. As a result, the need for text interface and Graphical User Interfaces (GUIs) has decreased. A gesture is an action that needs to be observed by someone else and is intended to convey a message of some sort. The gesture is commonly perceived as a movement of a body part, particularly the hand or head, which is utilized to express an idea or message. The technology for gesture recognition available today is considerably more recent.

For the purpose of sensing gestures and commanding robots, several methods have been devised. One commonly used technique for identifying hand motions is the Glove-based technique. This technique involves the utilization of a sensor mounted on a glove that accurately records hand movements. Hand- and body-gesture technology-based gaming gesture interfaces need to be designed to be both popular and profitable. The suitability of each gesture-recognition algorithm for automatic hand gesture identification depends on the user's cultural background, the application area, and the surroundings. Therefore, no single approach for automatic hand gesture identification is suitable for all applications.

A wide range of uses is attributed to hand gesture-driven robotic systems, including telemedicine and surgery, motion control, computer games, multimedia device control, smart television, automated wheelchairs, and articulated robotic arms. Various adaptations of gesture-controlled systems are being implemented, such as in wheelchairs and equipment designed for individuals with disabilities. To perform various tasks, a traditional wheelchair user must be capable of manipulating the wheelchair's controls, pedals, and levers. Although some contemporary systems feature virtual interfaces with digital screens, each button on the display has its designated location, requiring the user to press them individually, making such systems unsuitable for individuals with disabilities. In contrast, mechanical accessories or projected icons are not required for hand gesture-controlled devices [7]. Skeletal hand tracking, which is being developed for virtual reality and augmented reality applications, is an emerging example of gesture-based motion capture [9]. Robots can utilize the programmed gesture dictionary to respond to detected gestures from a person's bare or gloved hand using a simple automated mechanism. Consequently, even individuals with limited ability can effectively control the entire system [7]. In addition to hand gestures, gesture-operated wheelchairs have also incorporated human facial movement, head movement, mouth opening and closing, voice, etc. [8]. However, hand gestures have been found to be more effective overall due to their various benefits, including accessibility, ease of interpretation, and resistance to surrounding noises, among others.

To understand the behavior of the domestic environment, human robots require channels/modalities through which they can interact. The effectiveness of an interface primarily depends on the number and diversity of its inputs and outputs, which are communication channels that facilitate interaction between users and the human robot. Each individual channel is referred to as a modality.

Data on human hand gestures are primarily collected using two modalities. One modality is a wearable sensor, while the other modality is a vision-based interface. Hand measurements, such as joints orientation, hands position, and hand velocity, are extracted using wearable/sensor-based approaches, which can be achieved through the use of microcontrollers and specific sensors. However, these approaches are costly and require specialized equipment, appropriate environment, and training for optimal utilization of the systems. Despite their reliable qualities, these devices can sometimes cause discomfort for participants as they need to be worn on the body, adding to the wearer's burden [10]. Over the years, many researchers have opted for vision-based systems as an alternative to address these issues associated with wearable sensors, such as data gloves and sensors. Vision-based techniques offer advantages such as contactless recognition and the absence of specialized instruments with limited mobility. Due to their simplicity and affordability, vision-based technologies are gaining increasing popularity [10]. In vision-based approaches, cameras are utilized to identify gestures based on attributes like hand texture and color, while overcoming challenges such as varying lighting conditions, complex backgrounds, shadows, camera movement, user variations, self-occlusions, and unrestricted participant movement in the scene [11]. However, vision-based systems still face challenges in unstructured environments and poor working conditions. This highlights the need for a reliable hand gesture detection system, which drives a significant portion of researchers to work in this field. Nonetheless, currently, there is no vision-based hand gesture detection system that is deemed sufficiently reliable to address all the aforementioned issues.

Sign language recognition is an application of hand gesture recognition. The improvements in human-computer interface technologies have primarily been developed for hearing people and rely on vocalized speech, leaving American Sign Language (ASL) users in the Deaf community unable to benefit from these advancements. Deaf individuals often rely on this natural language to communicate with each other [12]. Despite significant progress in ASL identification using video or wearable gloves, concerns regarding privacy have been raised regarding their use in homes, and wearable gloves significantly impede daily life. There are two established methods of communication between deaf individuals and hearing individuals who do not understand sign language: text messaging. However, the high cost of interpreters for everyday conversations results in a loss of independence and privacy for deaf individuals. Written communication is much slower compared to speaking or using sign language, and facial expressions made during speech or sign

language use are lost when communicating through text. Thus, a less expensive and more effective method is required to facilitate communication between hearing and deaf individuals.

Vein pattern recognition has emerged as a fascinating and promising area of research in the field of biometrics. The unique patterns of veins present in an individual's hand or finger provide a distinct and highly secure means of personal identification. This technology holds great potential for a wide range of applications, including access control systems, forensic investigations, healthcare management, and financial transactions.

One of the key motivations behind vein pattern recognition is its inherent security and reliability. Unlike other biometric modalities such as fingerprints or iris scans, vein patterns are internal features that are invisible to the naked eye and difficult to forge or replicate. The complex network of veins beneath the skin forms a highly individualistic pattern, making it highly resistant to fraudulent activities.

Another important motivation for vein pattern recognition is its non-intrusive and contactless nature. The technology utilizes near-infrared light to capture the vein patterns, eliminating the need for physical contact with any device or surface. This makes it hygienic and suitable for individuals with sensitive skin or cultural preferences.

The human forearm vein pattern is the focus of this thesis, utilizing thermal imaging technology. By capturing thermal images, the vein patterns can be analyzed in a non-invasive and contactless manner, offering comfort and hygiene to the users. The thermal contrast between the veins and surrounding tissues provides distinct features for reliable vein pattern recognition. The stability of vein patterns over time makes them a suitable biometric modality for long-term and accurate identification. Moreover, thermal imaging allows for vein pattern visualization in challenging lighting conditions and obstructed views, expanding the potential applications of this technology [13].

In this thesis, the utilization of thermal imaging technology for human forearm vein pattern recognition is explored. Through the passive capture of thermal images, the vein patterns can be analyzed, offering a non-invasive and hygienic approach. The distinct features derived from the thermal contrast between the veins and surrounding tissues contribute to reliable vein pattern recognition. The stability of vein patterns over time ensures long-term accuracy in identification. Furthermore, the versatility of thermal imaging enables vein pattern visualization in various environments, including challenging lighting conditions and obstructed views. These findings

contribute to the advancement of vein pattern recognition and its potential applications in security and authentication systems [14].

1.3. CONTRIBUTIONS IN THE THESIS

In previous works, the use of normal cameras for the acquisition of image data has been considered inadequate due to their dependence on the environment. Inadequate, incorrect, or challenging lighting conditions can result in the presence of shadows, specularities, and other undesirable effects in imaging. Visuals can be partially obscured due to unwanted obstructions, the free movement of the user's hand within the area, the spatial relationship between a person and a robot, and other factors.

In this thesis, two main issues frequently encountered in real-world vision-based hand gesture recognition systems have been addressed: difficult photometric circumstances and images with occlusion. New datasets have been created with the assistance of a Thermal Imaging Camera for data acquisition. The visibility of objects in an image is determined by the backdrop illumination, which varies depending on the working environment. The use of a thermal camera has allowed the easy removal of this obstacle. Previous attempts to recognize hand gestures have utilized depth images in low-intensity environments [15], as well as low-resolution RGB images under changeable lighting conditions [16] and [17]. However, when it is completely dark, RGB cameras fail to capture any objects in the imaging scene. The distance at which an object can be detected is influenced by the quality and resolution of the camera. In contrast, the use of a thermal camera enables the detection of objects in total darkness, as long as they are not the same temperature as the surrounding air, thereby providing added flexibility for data collection. Two types of datasets have been created in this thesis: (i) American Sign Language (starting from 1 to 10) and (ii) Vein Pattern of both forearms.

In order to address the challenges of insufficient luminosity and occlusions in real-world vision sensing for hand gesture identification, as well as the difficulties in human-assisted robot navigation and collaborative robotics, the LUM (Laplacian Uniform Mixture) method is employed. These diverse and challenging environments are highly common, particularly in underdeveloped nations, making it essential for the model development to consider this fundamental problem. While the cost of a thermal camera is slightly higher, it is a recommended solution for the aforementioned circumstances.

It is shown in this work that the Laplacian Uniform Mixture (LUM) robust regression general-

purpose algorithm can be effectively applied to resolve this real-world robotics issue in the area of human-robot collaboration. Furthermore, it is demonstrated in the present thesis how unique weight thresholding concepts can be used in conjunction with the fundamental LUM algorithm [referred to as TUM (Student's t Uniform Mixture) algorithms] to propose and construct enhanced versions of the LUM algorithm to address these issues. A weight thresholding variation has been proposed for TUM, and its utility has been aptly illustrated by thorough case studies in two difficult real-world settings: (i) the sign language thermal images without adding occlusion and (ii) the sign language thermal images with adding occlusion.

The focus of this thesis is on the vein pattern recognition of the human forearm, where a thermal camera is employed as the imaging device. The vein patterns are captured and analyzed using computational methods, including the LUM (Laplacian Uniform Mixture) algorithm and the proposed TUM (Student's t Uniform Mixture) algorithm. The LUM algorithm, known for its robust regression capabilities, is utilized for vein pattern analysis, while the TUM algorithm incorporates unique weight thresholding concepts to enhance the performance of the LUM algorithm. These algorithms enable accurate and reliable vein pattern recognition using thermal imaging technology. Previous works [13] [14] have demonstrated the potential of vein pattern recognition in various environments, and the utilization of thermal imaging technology further expands its applicability. Through this research, advancements in vein pattern recognition using thermal cameras will be explored, contributing to the field's existing knowledge and potential applications.

In this thesis, thorough investigations have been conducted into the circumstances when the experiment's temperature constraint has been further tightened. Since thermal images were utilized, the images are sensitive to temperature. Specifically, in certain situations, the temperature of the environment may reach levels that can be characterized as warm. Under these highly variable heat conditions, the recognition mechanism must exhibit greater durability. Consequently, some modifications were made to the parameters of the Laplacian uniform mixture method. In the Laplacian uniform mixture, the weight function has three variable parameters. By adjusting the values of the parameters c (representing a uniform distribution for the outlier model), b (the scaling parameter for the Laplacian distribution), and λ (Lambda) (a Lagrange multiplier), improved accuracy and enhanced robustness are achieved in the system.

To address such situations, a new model called Student's t uniform mixture has been proposed in this thesis, inspired by the error distribution model of Laplacian Uniform mixture. The Laplacian

distribution was replaced with the probability density function known as the Student's t distribution. It was discovered that a mixture model based on the Student's t and Uniform distribution is more suitable when considering coding errors in sparse representation due to dense error. In this process, optimal and adaptive shrinkage operators were considered, and the outcomes were carefully evaluated. The system's performance was further enhanced by a brand-new adaptive hard thresholding operator that is suggested in this thesis.

1.4. STRUCTURE OF THE THESIS

The entire thesis work is documented in different chapters, as briefly mentioned below:

Chapter 1 explains the context of gesture detection systems in HCI, the driving forces behind the study, the contributions made to this particular project, and the thesis' structure.

Chapter 2 presents a literature review of hand gesture based robot navigation, sparse representation, regularized coding, collaborative representation based classification, robust sparse coding, regularized robust coding, Gaussian uniform mixture, and laplacian uniform mixture.

Chapter 3 provides a prospective problem statement that is based on the ideas and knowledge gleaned from the literature review.

Chapter 4 explains the thermal imaging sign language hand gesture database that was used for this thesis' experimental work.

Chapter 5 records a thorough investigation of the laplacian uniform mixing in relation to the identification of sign language in thermal images. There is a detailed mathematical model and algorithm defined, and experimental findings are used to validate it.

Chapter 6 provides a newly created model for thermal image sign language recognition called student’s uniform mixture. This chapter presents examples of various weight functions together with the relevant algorithms and supporting experimental research.

Chapter 7 introduce a work focuses on vein pattern recognition using thermal images of the human forearm. The LUM and newly proposed TUM methods are employed, showcasing accurate and reliable results.

Chapter 8 concludes the thesis with a quick summary remark and a forecast of potential future working areas.

CHAPTER-2

LITERATURE SURVEY

- ***INTRODUCTION***
- ***HAND GESTURE BASED ROBOT NAVIGATION***
- ***SPARSE REPRESENTATION***
- ***COLLABORATIVE REPRESENTATION BASED CLACIFICATION***
- ***ROBUST SPARSE CODING***
- ***REGULARIZED ROBUST CODING***
- ***GAUSSIAN UNIFORM MIXTURE***
- ***LAPLACIAN UNIFORM MIXTURE***
- ***SUMMARY***

2.1. INTRODUCTION

In this chapter, a brief analysis of the domains essential to solve the problem specified in chapter one is presented. Firstly, a survey of prior studies on hand gestures-based robot navigation in collaborative robotics is conducted. In this section, the extent to which the current HCI solutions provide robustness against unfavorable circumstances that arise in gesture recognition scenarios is discussed.

A variety of problem-solving techniques that enhance performance by incorporating resilience into already-existing systems are presented in this thesis. For each subject required for the solution, a supplemental literature review has been conducted. As a result, the following sections include surveys on sparse representation, Collaborative Representation based Classification, Robust sparse coding, Regularized robust coding, Gaussian uniform mixture, and Laplacian uniform mixture, respectively.

2.2. HAND GESTURE BASED ROBOT NAVIGATION

When analyzing the earlier works of gesture recognition systems, two primary aspects are observed. Firstly, all of the interaction modes that can be used to capture gesture data have not yet been tested by researchers. Secondly, the type of analysis performed and the overall performance of the system are significantly influenced by different kinds of data acquisition systems. In the broad field of computation methods, increased interest has been observed among scholars over the years. In this section, several data gathering methods and computational strategies that have been employed in gesture detection systems in the past are briefly discussed.

2.2.1. Data Acquisition Mode

Data about participants' hand gestures can be gathered in a number of methods. The majority of methods are classified either as a vision-based or a non-vision-based approach. The non-vision-based approach uses interface tools such as data gloves, motion sensors, inertial measurement units (IMUs), magnetic tracking sensors, electromyography (EMG), etc. to gather data about hand gestures [18]-[22], [10]. Haptic gloves were employed by Ma *et al.* [6] in their teleoperation studies. Wang *et al.* [23] accomplished this using bending sensor incorporated data gloves. Data is collected using data gloves and magnetic trackers [10]. Since data gloves [24] have sensors built in to recognize hand and finger motions and their orientations [5], they can be very

effective at capturing hand gestures. But not all participants may find these to be comfortable to wear. By tracking the fingertips and markers, fingertip markers on a hand may identify different hand shapes [25], [26]. However, once again, participants can find it difficult to put on the markers, especially the wheelchair-using individuals with disabilities. Another common technique used nowadays to collect gesture data is depth sensing utilizing a Kinect sensor, although it comes at a significant cost [27].

As was already mentioned, wearable technology offers dependable features, but because they are often strapped to the user's body, they can occasionally cause discomfort. Additionally, these sophisticated sensors are quite expensive, which frequently causes them unsuitable for cheap gesture detection systems [10].

When the gesture types are other types, such as face movement, and not hand-based, techniques like electromyography (EMG) [28] or electroencephalogram (EEG) [29] can be used to capture gesture data. The tasks have to be carried out in certain situations by trained instructors. Additionally, performing tasks like EMG or EEG can occasionally put a mental strain on the users. These sensors are not ideal for systems that must be cost-effective due to their high price [8].

However, by gathering the data using cameras and imaging sensors, the vision-based approach gets around those drawbacks. For the purpose of obtaining the real-time hand gesture images, Raheja *et al.* [30] used a high resolution camera integrated with a digital signal processing (DSP) unit. A serial connection module and five megapixel webcam-based system were developed in [5]. However, research projects employing this strategy have run into a number of difficulties that reduce the effectiveness of current systems, including inconsistent lighting, motion blur, the background clutter, and hands occlusion.

In comparison to RGB-based hand gestures, thermal imaging-based hand gestures are considered more reliable [31], [32], and [33]. For the purpose of recognizing hand gestures for sign language digits, a proposal is made for utilizing a thermal picture with a pixel size of 32×32 and deep learning [31]. However, the performance of the utilized camera is constrained for many long-range applications due to its extremely poor resolution. Creating reliable hand gesture detection systems with such low-resolution cameras also presents difficulties. To address these challenges, a reliable hand gesture detection system is developed using deep learning and high-resolution thermal imaging.

2.2.2. Computational Methods

Hand gesture recognition is an important aspect of sign language interpretation and human-robot interaction. In the context of sign language, hand gesture robot navigation refers to a robot's ability to recognize and interpret hand gestures made by humans as navigation commands, after which the robot navigates itself in the environment based on the recognized gestures. There are several computational methods for hand gesture recognition and robot navigation, and the choice of method depends on the application requirements. Some commonly used methods include:

These methods use cameras to capture the hand gesture and then process the image to recognize the gesture. The Viola-Jones algorithm is an example of a computational method for hand gesture recognition that has been used in the design of a reliable classifier by Zhang *et al.* [8]. The algorithm utilizes Haar-like features [34] and the AdaBoost learning algorithm to create the classifier. Haar-like features concentrate on the contrast ratio between bright and dark areas in a kernel, which makes the system relatively tolerant to noise and variations in lighting conditions. However, the method may not perform adequately when the images of the hand gestures are partially obstructed. Other computer vision-based methods include contour tracking, template matching, and color-based segmentation.

Support vector machine (SVM) is a widely used classification algorithm in gesture detection, but it faces some challenges, particularly in multi-class classification problems. To address these issues, researchers have explored the use of particle swarm optimization (PSO) as an alternative for optimizing and fine-tuning system parameters [35]. However, traditional PSO algorithms are prone to being trapped in local optima. To overcome this challenge, Wang *et al.* [23] presented an improved PSO algorithm for SVM classification.

These methods use sensors such as accelerometers and gyroscopes to capture hand gestures. The data from these sensors are then processed using machine learning algorithms to recognize the gesture.

These methods use machine learning algorithms to train the system to recognize hand gestures. The system is trained on a dataset of hand gesture images or sensor data, and then the trained model is used to recognize the gesture. Hand gesture recognition for sign language using machine learning-based k-means++ algorithm and HOG descriptor has been studied by researchers [36]. Additionally, Xie and Cao [37] explored the use of artificial neural networks (ANN) to train hand gesture images.

Several works have been done in the area of hand gesture robot navigation for sign language interpretation. For example, in the work by Chen *et al.* [38], a deep learning-based method was used

for recognizing hand gestures in sign language. The system was able to recognize ten different hand gestures. The recognized gestures were then used as navigation commands to move a robot in a simulated environment.

In another work by Liu *et al.* [39], a sensor-based method was used for hand gesture recognition in sign language. The system used an accelerometer and a gyroscope to capture the hand motion, and a machine learning algorithm was used to recognize the gesture. The system was able to recognize five different hand gestures.

In a more recent work by Chen *et al.* [40], a hand gesture recognition system based on computer vision and deep learning was developed for sign language interpretation. The system was able to recognize ten different hand gestures. The recognized gestures were then used as navigation commands to move a robot in a real-world environment.

As hand gesture recognition is essential in sign language interpretation and human-robot interaction so with thermal imaging-based methods emerging as a reliable option in low-light conditions. Computational methods for hand gesture recognition and robot navigation vary based on the application requirements. Some commonly used methods include: Image Processing-based Methods and Machine Learning-based Methods.

Several works have been done in the area of hand gesture robot navigation for sign language interpretation using thermal imaging. For example, in the work by Velez-Rojas *et al.* [41], a convolutional neural network (CNN) was used for hand gesture recognition using thermal images. The system was able to recognize six different hand gestures. The recognized gestures were then used as navigation commands to move a robot in a simulated environment.

In another work by He *et al.* [42], a combination of image processing and machine learning-based methods was used for hand gesture recognition using thermal images. The system used image processing techniques to segment the hand region from the thermal image, and then a machine learning algorithm was used to recognize the gesture. The system was able to recognize ten different hand gestures.

In conclusion, hand gesture robot navigation in the area of thermal image sign language is an emerging field in human-robot interaction. Several computational methods, including image processing-based methods and machine learning-based methods, have been used for hand gesture recognition using thermal images. Several works have been done in this area, and the choice of method depends on the application requirements.

2.3. SPARSE REPRESENTATION

Sparse representation has recently gained popularity in the fields of pattern recognition and computer vision. A specific kind of redundant representation called sparse representation, often known as sparse coding, is growing in popularity every day, especially in the context of tackling several signal processing and image processing issues. It is applied to image super-resolution [43], motion segmentation [44] and supervised denoising [45]. Wright *et al.* have applied sparse representation to classification so that the can utilize sparse representation based classification (SRC) technique. [46]. A test sample is represented for SRC as a sparse mixture of training samples, and the sparse representation coefficient is calculated by resolving the sparse representation issue. The test sample is put into the category that minimizes the difference between its own and the one reconstructed with the training samples from this category. In trials involving facial recognition, SRC demonstrates its efficacy.

A specific signal is here represented as a linear combination of various basis functions to have majority of representation coefficients to be zero or of modest magnitude. These representational structures appear to be quite helpful in dictionaries learning issues [47]. Many different signal and image processing issues have been solved using the sparse representation model, such as array signal processing issues with direction-of-arrival (DOA) estimation, [48] etc. In the field of signal reconstruction, compressed sensing and sparse modelling are both highly helpful techniques [49]. Sparse coding is utilized in mathematical applications like data clustering. Also, decomposition and denoising of signals can be done with sparse representation. Each signal can be represented by a sparse combination of basis elements, also known as atoms, producing corresponding columns in the dictionary matrix. The co-occurrences matrix of atoms and signals was obtained using sparse representation in [50] and was subsequently employed in probabilistic subspace clustering. Sensor applications can benefit greatly from sparse Bayesian learning. In sensor networks, sparse recovery algorithms are used to calibrate and estimate sensor drifts [51]. High resolution radar imaging systems use sparsity driven algorithms, which have shown great success in image reconstruction issues [52]. In [53], is suggested to use the kernel sparse representation based classification (KSRC) technique. Samples are first mapped onto a high dimensional feature space, and using a kernel trick, SRC is then carried out in this new feature space. They cannot directly execute KSRC because samples in the high dimensional feature space are unknown. They provide a strategy to address the issue of sparse representation in the high dimensional feature space in order to get over this

challenge.

2.4. COLLABORATIVE REPRESENTATION BASED CLACIFICATION

A more efficient approach to face recognition than Sparse Representation based Classification (SRC) is Collaborative Representation based Classification (CRC). In reality, by using different norms for the coding residual and coefficient, SRC may be seen as a particular instance of CRC that can be tailored for a variety of applications. However, Jun Yin *et al.* [54] introduced a cutting-edge dimensionality reduction method called Optimized Projection for Collaborative Representation based Classification (OP-CRC), which is closely related to CRC, to improve the performance of CRC. In particular, CRC uses the minimum reconstruction residual based on collaborative representation as the decision criterion. OP-CRC is intended to enhance this method.

In their discussion of how SRC operates, Lei Meng *et al.* [55] demonstrate how much more important the collaborative representation process used in SRC is to the accuracy of its face classification. The SRC is a particular instance of collaborative representation based classification (CRC), which can take on multiple forms by using various coding coefficient and residual norms. More specifically, the degree of facial feature discrimination is connected to the l_1 or l_2 norm characterization of coding coefficient, whereas the l_1 or l_2 norm characterization of coding residual is related to the robustness of CRC to outlier face pixels. To confirm the face recognition performance of CRC with various instantiations, numerous experiments were carried out. The research into sparsity-based pattern classification is greatly advanced by SRC (including S-SRC and R-SRC), which exhibits extremely intriguing strong FR performance. While Yang and Zhang [57] learned a Gabor occlusion dictionary to significantly lower the computational cost when dealing with face occlusion, Gao *et al.* [56] proposed the kernel sparse representation for FR as an SRC-inspired alternative. While Qiao *et al.* [59] learned a subspace to preserve the l_1-graph for FR, Cheng *et al.* [58] presented the l_1- graph for image classification. Yang *et al.* integrated linear pyramid matching and sparse coding for picture classification in [60]. In SRC, it is assumed that face images are aligned, and solutions to misalignment or issues with changes in posture have also been suggested. For instance, paper [61]'s approach is invariant to image-plane translation, and paper [62]'s method was created to handle misalignment and illumination variations. Low-rank decomposition was utilized by Peng *et al.* [63] to align a collection of linearly connected images with severe corruption. To improve SRC-based pattern categorization, dictionary learning

techniques [64]-[67] were also developed.

Despite having interesting classification results, SRC and CRC's intrinsic classification mechanism is still unknown. A probabilistic framework for collaborative representation is proposed in [68] and accordingly it is possible to define and calculate the likelihood that a test sample belongs to the collaborative subspace of all classes. Due to this, we introduce the ProCRC, a probabilistic collaborative representation-based classifier that jointly maximizes the likelihood that a test sample belongs to each of the many classes. Recent years have seen a significant increase in the use of CRC and its variants in a variety of applications, including face recognition and hyper spectral image classification.

Recently, Zhang *et al.* [69] developed a face recognition approach called Collaborative Representation based Classification (CRC), indicating that l_1-norm sparsity is not what makes SRC powerful. CRC also represents a test sample using all the training samples. The regularized word is the primary distinction between SRC and CRC. The objective function of CRC adopts l_2-norm but not l_1-norm, which greatly reduces the complexity of CRC compared to SRC. Kernel Collaborative Representation Classification (KCRC) [70] executes CRC in the new space after mapping the data into a higher dimensional space where various classes are more easily distinguishable. Yang *et al.* suggested Collaborative Representation based Projections (CRP) [71] as a dimensionality reduction technique for face recognition using collaborative representation. Utilizing collaborative representation coefficients, CRP creates an l_2-graph and uses graph embedding to reduce dimensionality. It is quicker than dimensionality reduction techniques based on sparse representation and could sustain the collaborative representation-based reconstruction relationship.

2.5. ROBUST SPARSE CODING

In now a days, face recognition has been successfully implemented using sparse representation (or coding) based classification (SRC). The l_2-norm or l_1-norm of the coding residual is used to determine the representation fidelity in SRC, where the testing picture is represented as a sparse linear mixture of the training samples. This type of sparse coding model really presupposes that the residual coding data has a Gaussian or Laplacian distribution, which may not be precise enough to capture genuine coding errors. In [72] by modelling the sparse coding as a sparsity limited robust regression problem, they suggest a novel approach known as the robust sparse coding (RSC).

The SRC approach exhibits remarkable robustness when dealing with corruption and occlusion. In order to deal with the errors and pose variation in FR, Huang *et al.* proposed a sparse representation recovery method that is constant to image plane transformation in [77], while Wagner *et al.* proposed a sparse representation based method in [74]. In order to significantly minimize the size of the occlusion dictionary and increase the FR accuracy, Yang and Zhang [75] employed Gabor features in SRC. Several studies have attempted to modify the sparsity constraint in the sparse coding model to address concerns about the effectiveness of the l_1-norm constraint in characterizing signal sparsity. Example: Liu *et al.* [76] modified the sparse coefficient by including a nonnegative constraint. In order to enforce the sparsity requirement, Wang *et al.* [80] employed the weighted l_2-norm, whereas Gao *et al.* [73] used a Laplacian term of coefficient. In order to create terms for sparsity regularization, Ramirez *et al.* [78] have proposed a framework for universal sparse modelling. The sparsity regularization terms were also created using Bayesian approaches [79].

In the [72] work, they presented a robust sparse coding (RSC) paradigm to increase the robustness and efficacy of sparse representation.

2.6. REGULARIZED ROBUST CODING

Regularized Robust Coding (RRC) is a sparse coding method that combines the sparsity-inducing l_1-norm with robust loss functions to handle outliers and corruptions in data. The method was first proposed by Cao *et al.* in 2013 [81] as an extension of the l_1-norm based sparse coding model. RRC has been applied in various fields such as computer vision, image processing, and signal processing, as a robust and effective method for feature extraction, image restoration, and denoising.

When applied to practical FR systems, a sparse coding model's assumption that the coding residual follows a Gaussian or Laplacian distribution may not be sufficient to adequately explain the coding residual. The sparsity constraint on the coding coefficients, meanwhile, raises the computational expense of SRC significantly. A brand-new coding scheme is provided in the [93] study called regularized robust coding (RRC). Except for the l_1-norm fidelity in [94] and [95], the

correntropy based Gaussian-kernel fidelity in [96] and [97], and the earlier work in [98], very few works have been presented in the scheme of sparse representation. The fidelity term significantly affects the outcome of the coding process. To define the fidelity term with the l_2- or l_1-norm really assumes that the coding residual $e = y - Ax$ has a Gaussian or Laplacian distribution from the perspective of maximum a posteriori (MAP) estimation. However, in practice, such an assumption might not be accurate, particularly if the query images contain occlusions, corruptions, and expression variations. The complicated variance of occlusion may prevent the Gaussian kernel-based fidelity term from performing well in FR with occlusion, despite claims that it is robust to non-Gaussian noise made in [96] and [97] and used in [96] and [97] [99]. It is presented a regularized robust coding (RRC) model to improve the FR's resistance to occlusion, pixel corruption, disguises, and huge expression changes, among other threats in [93].

Sparse coding has been shown to be effective in solving image classification and reconstruction problems. However, in tasks such as clustering and image classification, where data locality is critical, the usefulness of sparsity has been questioned. In such cases, data locality is more important than achieving a better reconstruction. To address this issue, researchers have proposed using a locality regularization term to supplement or even replace sparse representation in these tasks. This approach places greater emphasis on preserving data locality information, which can be crucial for accurately classifying images and other forms of data. By balancing the tradeoff between sparsity and data locality, this technique has the potential to improve the performance of various image analysis tasks. For these kinds of issues, [87] presented a locality-constrained linear coding strategy. [88] created a new graph regularized locality-constrained coding model using the Laplacian regularizer and the locality-constrained coding strategy. In issues involving compressed sensing and visual tracking, sparse representation governed by l_1 -regularized least square is employed [89]. In addressing problems based on compressive sensing, [90] addressed the regularization norm of l_0 and reported a reduction in computing complexity. Research has shown that the success of sparsity-based image reconstruction models is more strongly related to the collaborative representation of corrupted images, rather than the sparsity constraint imposed on coding coefficients [91]. This means that a key factor in achieving accurate reconstruction of images is the ability of the model to incorporate information from multiple similar images, rather than simply relying on sparse coding. Therefore, many recent advances in sparsity-based image reconstruction have focused on improving the collaborative representation of images, through

techniques such as dictionary learning and low-rank modeling. Yang *et al.* [92] have suggested a regularized resilient coding approach based on this concept.

Notable related works include the work of Zhang *et al.* [82] on robust image retrieval using RRC, the work of Zhang *et al.* [83] on RRC-based multi-view learning, and the work of Zhang *et al.* [84] on RRC-based face recognition. In addition, recent studies have shown that RRC can be further improved by incorporating additional regularization terms, such as graph-based regularization and non-local self-similarity.

For example, the work of Zhang *et al.* [85] proposed a graph-based regularized RRC method for image restoration, while the work of Zhang *et al.* [86] introduced a non-local self-similarity regularization term to enhance the performance of RRC-based image denoising.

2.7. GAUSSIAN UNIFORM MIXTURE

Sparse Representation-based Classification (SRC) is a technique that has been proposed for robust face recognition. In this method, the testing image is represented as a sparse linear combination of the training samples, and the quality of the representation is measured by the l_2-norm or l_1-norm of the coding residual. However, the assumption that the coding residual follows a Gaussian or Laplacian distribution may not be sufficient to describe the coding residual in practical FR systems, which can impact the effectiveness of SRC.

To address this issue, researchers have proposed modifications to the SRC algorithm that can improve its robustness and adaptability. For instance, some researchers have suggested using a non-Gaussian prior distribution for the coding residual, such as the Cauchy distribution, which can better model the heavy-tailed nature of the residual distribution in practical FR systems. Additionally, some researchers have proposed incorporating additional features such as facial landmarks or texture information to enhance the accuracy and robustness of the SRC method.

Then a new mixture model is proposed, which is called Gaussian Mixture Model (GUM). In GUM, outliers are predicted to have a uniform distribution and inliers to have a Gaussian distribution.

Among the most popular robust statistical techniques include M-estimators, sampling techniques, trimming techniques, and robust clustering. M-estimators aim to lessen the impact of

high residual values by minimizing the sum of a positive-definite function of the residuals. With no evidence of convergence for the majority of M-estimators, the minimization is carried out using weighted least squares approaches. By resolving a set of equations that are specific to a subset of data that was picked at random with sampling techniques like least-median-of-squares and random sample consensus (RANSAC) [112] estimate the model parameters. The fundamental disadvantage of sampling techniques is that they require complicated data-sampling procedures and are laborious to utilize while estimating a large number of parameter. Ranking residuals and de-weighting data points connected to significant residuals are two features of trimming approaches. Typically, they are transformed into a (non-linear) weighted least squares optimization problem, where the weights are changed after each iteration to produce iteratively re-weighted least squares problems. Combining the concepts of robust statistics and mixture models, a robust mixture model has been proposed, which involves a Gaussian mixture with a uniform noise component to model outliers [110], [113]. Notably, recent studies have shown that modelling outliers with a uniform component can yield excellent results [113], [111].

In the [114] work, they provide a novel technique known as DeepGUM: a deep regression model that uses a Gaussian-uniform mixture model to make it robust to outliers. They arrive at an optimization technique that alternates between supervised training on cleaned data using stochastic gradient descent and unsupervised outlier detection using expectation-maximization. Without having to manually impose any threshold on the percentage of outliers in the training set, DeepGUM is able to adapt to an outlier distribution which is continuously changing.

The Gaussian Uniform Mixture (GUM) model has been studied extensively in the context of outlier detection and anomaly detection. Here are a few related works on GUM: S. Sultani *et al.* [100] proposed a GUM model for anomaly detection in crowded scenes. They use a Gaussian distribution to model the normal behavior of the crowd, and a uniform distribution to model the rare events that are considered anomalies. J. Xu *et al.* [101] proposed an adaptive online anomaly detection method based on a GUM model. The method adapts the model parameters dynamically to changing data distributions and can detect anomalies in real-time. S. S. Lee *et al.* [102] proposed a GUM model for detecting malicious attacks in network traffic. They use a Gaussian distribution to model the normal behavior of the network traffic, and a uniform distribution to model the rare events that are considered malicious.

2.8. LAPLACIAN UNIFORM MIXTURE

The Laplacian Uniform Mixture (LUM) model was first proposed by Xie and Zha in 2004 as a probabilistic model [103]. The model assumes that the data is a mixture of two distributions: a Laplacian distribution representing the majority of the data and a uniform distribution representing the outliers or rare events. Since then, the LUM model has been used in various applications, including anomaly detection, image processing, and computer vision, as a way to robustly model.

Notable related works include the work of De *et al.* [104] on robust anomaly detection using the LUM model. In this paper, the authors propose a robust anomaly detection method based on the LUM model. The method uses a Laplacian distribution to model the inliers and a uniform distribution to model the outliers. Another work of Li *et al.* [105] on robust face recognition using a LUM model-based sparse representation. In addition, recent studies have shown that the LUM model can be used in combination with other techniques, such as sparse representation and deep learning, to achieve state-of-the-art performance in various tasks.

For example, the work of Liu *et al.* [106] proposed a sparse representation-based method for image denoising using the LUM model. In this paper, the authors propose a texture segmentation method based on the LUM model. The method uses multistage features and a Laplacian distribution to model the texture regions and a uniform distribution to model the non-texture regions. The work of Huang *et al.* [107] extended the LUM model to a more general framework, called the Laplacian Scale Mixture (LSM) model, which can better capture the scale variability of the data. The work of Zhang *et al.* [108] proposed a LUM model-based method for detecting and segmenting retinal blood vessels in fundus images.

J. Zhang *et al.* [109] propose a feature extraction method for hyper spectral image classification based on the LUM model. The method uses a Laplacian distribution to model the spectral signatures and a uniform distribution to model the background noise.

Overall, these studies demonstrate the usefulness of the LUM model in various applications, particularly in robust anomaly detection, image processing, and computer vision tasks. The LUM model's ability to capture the heavy-tailed nature of the data using the Laplacian distribution, combined with a uniform distribution to model outliers, has shown to be effective in these applications.

2.9. SUMMARY

The literature survey presented in the preceding section had the aim of providing a thorough understanding of the computational domains, problem frameworks, mathematical formulations, algorithms, and data handling strategies, as well as limitations and future prospects relevant to this thesis work. A detailed and organized overview of these topics was offered, and references for further study, if necessary, were included. The survey was conducted to ensure that a comprehensive understanding of the background and context of this thesis work is obtained by the reader.

The upcoming chapter will build upon this literature exploration and formulate a potential problem statement and model. Insights gained from the literature survey will be drawn upon to identify the research gap that this thesis work aims to address. Through the formulation of a suitable problem statement and model, the aim of the thesis work is to contribute to the existing literature and expand the scope of knowledge in this field.

CHAPTER-3

PROBLEM STATEMENT

- ***INTRODUCTION***
- ***PROBLEM FORMULATION***
- ***SUMMARY***

3.1. INTRODUCTION

As previously discussed in Section 1.2 of *Chapter* 1 and Section 2.2 of *Chapter* 2, vision-based hand gesture recognition systems are often preferred over wearable sensor-based systems due to their user-friendly interface and lower cost. However, significant challenges can be faced by these systems when operating in uncontrolled, real-world environments where it can be difficult to extract reliable features from visual sensing modules.

An approach to addressing these challenges is to incorporate thermal imaging technology into hand gesture recognition systems. Heat signatures from objects, including human hands, can be captured by thermal cameras, providing a more reliable and robust way to detect and interpret gestures in uncontrolled environments [12], [114]. This is particularly useful for sign language recognition, where traditional visual sensing modules can find it challenging to accurately recognize gestures due to environmental occlusions such as clothing, accessories, or other obstructions.

The accuracy of hand gesture recognition systems can be improved, and the impact of environmental factors can be reduced by combining thermal imaging with computer vision and machine learning algorithms. The effectiveness of collaborative robotics can be enhanced with this approach, where accurate gesture recognition is critical for improving overall efficiency and enabling seamless human-robot interactions. However, thermal imaging has some limitations and challenges. For instance, fine details of hand shapes and movements may not always be captured by thermal images, which can make it difficult for machine learning algorithms to accurately classify gestures. These challenges can lead to issues with the robustness of hand gesture recognition systems in collaborative robotics, especially when it comes to accurately detecting and interpreting gestures in practical, unpredictable scenarios.

Human Forearm Vein Pattern based Biometric Authentication is a state-of-the-art method that employs a thermal camera to conduct experiments and investigations. By utilizing the thermal camera, the vein patterns present on the human forearm are captured and analyzed for authentication purposes. This innovative approach offers a contactless and non-intrusive solution for biometric identification. The unique vein patterns serve as a reliable and distinctive feature,

ensuring enhanced security and accuracy in authentication systems. Through the utilization of thermal imaging technology, this method enables precise and robust authentication in diverse applications, including access control systems, identity verification processes, and secure digital transactions.

3.2. PROBLEM FORMULATION

In the present thesis, efforts have been made to solve three varieties of image recognition problems using the concepts of thermal imaging. Especially thermal imaging is useful for implementing these systems in challenging situations. The problems considered in these three categories are:

(i) Sign language recognition using thermal images in low-light situations and night vision;

(ii) Sign language recognition using thermal images in low-light situations and night vision with occlusion; and

(iii) Vein pattern recognition using thermal images.

The first problem is focused on the identification of sign language in low-light situations using thermal imaging. Thermal imaging can be utilized to record thermal data that is independent of lighting conditions in low-light settings. Issues with poor or improper lighting, unwelcome obstructions, and unwelcome shadows can be solved by using thermal pictures. However, using thermal images to detect visual hand gestures in hot environments can be difficult, and accuracy can be reduced when it is challenging to distinguish the hand's thermal signature from the environment.

The second problem's focus is on the identification of sign language using thermal pictures in low-light situations with occlusion. Partial occlusions may result from erroneous spatial interactions between the user and the robot, free participant movement, and the apparition of undesired moving items over the scene.

The focus of the third problem is on the identification of vein patterns using thermal imaging. Thermal imaging can be used to detect the thermal signature of veins in the hand for biometric

identification purposes. However, there are obstacles in vein pattern recognition using thermal imaging, such as the possibility of poor image quality due to various factors like hand positioning, movement, and temperature variations. Additionally, accurate and reliable vein patterns for identification purposes can be difficult to obtain in individuals with thin or less visible veins. False positives or false negatives can also occur due to variations in the thermal patterns of veins over time or changes in the individual's health or medication use.

After the descriptions of the three problems under consideration, let us discuss the computational methodologies that will be explored, proposed, implemented and investigated in detail within the scope of this thesis. In the field of visual hand gesture detection using thermal images, low-quality images due to image corruptions can often be encountered. To address this issue, the use of robust regression analysis approaches has been proposed by researchers to reconstruct corrupted image matrices. One such approach, originally developed for face recognition problems, is the Laplacian uniform mixture (LUM) [116] method. LUM utilizes regularized regression coefficients to robustly regress a given signal and is considered suitable for the particular types of corrupted hand gesture databases commonly encountered in this field.

The use of the LUM algorithm in solving hand gesture recognition problems in collaborative robotics is explored and implemented here in detail. It is well known that the error minimization problem can be converted into an iteratively robust coding problem by LUM [116], and its effectiveness can be improved through the incorporation of a suitably formulated weight update function. The thesis will also explore how different existing weight functions and newly proposed ones can be employed for our image recognition problem at hand i.e. how they canes enhance the robustness of the hand gesture recognition scheme.

To summarize, the LUM method is a robust regression analysis approach that can effectively reconstruct corrupted image matrices in visual hand gesture detection using thermal images. By utilizing the LUM algorithm and incorporating innovative weight update functions, we can improve the accuracy and robustness of hand gesture recognition systems in collaborative robotics.

An alternative distribution, such as the student's t distribution function, has been proposed to improve the robustness of the reconstruction process. The student's t uniform mixture, a new mixture

model, has been proposed by combining the student's t distribution function with the uniform distribution. The student's t distribution function is a statistical distribution that is more robust than the Gaussian distribution function and is capable of handling heavy-tailed data more effectively. Statistical modeling and inference often use this distribution. The Student's t Uniform Mixture (TUM) model can effectively handle image corruptions that result in heavy-tailed error distributions by combining the student's t distribution function with the uniform distribution. This method can handle a wide range of image corruptions and environmental factors, making it a promising approach for improving the robustness of hand gesture recognition systems in collaborative robotics.

Under certain conditions, such as extremely poor illumination and high temperature, errors in an image can be inaccurately represented by one-dimensional, pixel-based error models. The CESR algorithm is an iterative method used to detect errors in an image and assign weights to pixels based on those errors. Non-Gaussian noise and outliers present in image data can be effectively handled by this method. The algorithm is based on the half-quadratic (HQ) optimization framework, specifically the multiplicative form of HQ-based optimization (HQ-M), which was first proposed in paper [117]. In this framework, the errors in the image are estimated and corrected through an iterative process. Additionally, an algorithm based on the additive form of the HQ framework (HQ-A) [118] was proposed by He *et al.* [117] that can similarly estimate and correct errors in image data. As mentioned in prior studies [119], the effectiveness of using Gaussian distributions as the inlier model for image matching errors may be limited. The Laplacian distribution has been shown to provide a higher likelihood compared to Gaussian when modeling the inlier distribution (which will be discussed in upcoming chapters). Therefore, it is hypothesized that a mixture model with Laplacian as the inlier distribution could result in better robustness than previous methods. To account for unpredictable outliers, a uniform distribution is considered as the corresponding outlier model, similar to a previous study [120]. Thus, the LUM model was proposed in [116]. Although originally proposed for face recognition problems, the utilization of the LUM model in solving visual hand gesture recognition problems for collaborative robotics using thermal images is explored in the present study.

Various types of possible image corruptions that can occur in real-life human-computer interaction scenarios are considered in the study, and the effectiveness of LUM against them is evaluated. The most effective approach for various circumstances is determined by investigating

different threshold or shrinkage operations. In addition to utilizing the LUM method for gesture recognition, the study proposes and evaluates a new Student's t uniform mixture model. The proposed approach is thoroughly examined in view of the corrupted hand gesture database, and a detailed mathematical formulation and explanation of the proposed methods and their effectiveness in recovering the original hand gesture images from the corrupted thermal images are provided.

A Schematic Diagram of Hand Gesture Recognition System:

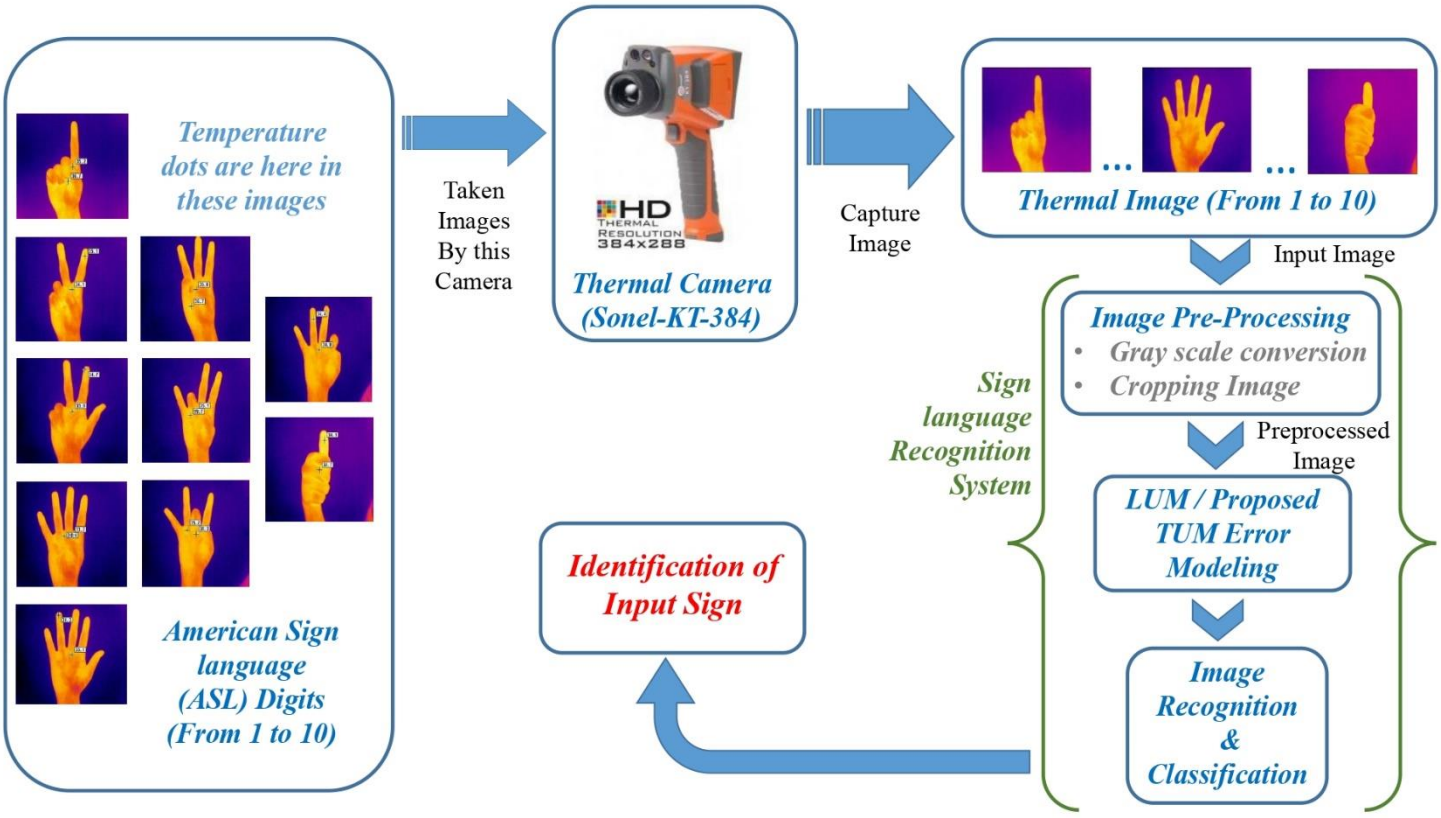


Fig. 3.1. Hand gesture recognition schematic diagram

For vein pattern recognition, the human forearm vein is utilized as a substitute for American Sign Language, while maintaining the same processes as depicted in the diagram.

3.3. SUMMARY

In summary, this passage discusses the use of thermal imaging technology in addressing the challenges faced by vision-based hand gesture recognition systems in uncontrolled environments. Thermal cameras capture heat signatures, providing a reliable and robust method for detecting and interpreting gestures, particularly beneficial for sign language recognition. The combination of thermal imaging with computer vision and machine learning algorithms improves accuracy and reduces the impact of environmental factors, enhancing collaborative robotics.

The passage also highlights the application of thermal imaging in human forearm vein pattern-based biometric authentication, offering a contactless and non-intrusive solution for identification purposes. Vein patterns serve as reliable and distinctive features, ensuring secure and accurate authentication in various applications.

The passage further presents three image recognition problems addressed using thermal imaging: sign language recognition in low-light and occluded situations, as well as vein pattern recognition. The Laplacian uniform mixture (LUM) method, known for reconstructing corrupted image matrices, is explored and implemented, along with innovative weight update functions, to improve the robustness of hand gesture recognition systems.

Additionally, the passage introduces the Student's t uniform mixture (TUM) model, combining the student's t distribution with the uniform distribution, to handle heavy-tailed error distributions in image corruptions. The CESR algorithm, based on half-quadratic optimization, is discussed as an effective approach for detecting errors and handling non-Gaussian noise and outliers in image data.

Overall, the passage emphasizes the computational methodologies employed to address image recognition problems using thermal imaging. The LUM and TUM models, along with threshold and shrinkage operations, are evaluated for their effectiveness in recovering original hand gesture images from corrupted thermal images. These findings contribute valuable insights to the field of hand gesture recognition in collaborative robotics.

CHAPTER-4

SIGN LANGUAGE DATABASE

- ***INTRODUCTION***
- ***DATA ACQUISITION***
- ***COMPOSITIONAL STRUCTURE OF DATABASE***
  - ***THERMAL IMAGE CAPTURING***
  - ***DATABASE 1***
  - ***DATABASE 2***
- ***SUMMARY***

4.1. INTRODUCTION

The hand gesture database used in this thesis was acquired under challenging circumstances to investigate the robustness of recognition schemes with greater reliability while considering environmental constraints. The acquisition was carried out in the Electrical Measurement and Instrumentation Laboratory of the Electrical Engineering Department at Jadavpur University in Kolkata, India, using a thermal camera (Sonel KT-384). The goal of collecting data under these specific conditions was to create databases for testing performances of some new algorithms under development and to demonstrate that those algorithms can perform well even in difficult circumstances.

4.2. DATA ACQUISITION

The KT-384 [121] thermal imager (Manufacturer: Sonel®, Poland), a completely radiometric camera, based on far infrared imaging (FIR) technology, was used to capture the images of the faces and body parts of human subjects. The thermal imager is displayed in Fig. 4.1 below.

Fig. 4.1. Sonel® KT-384 [121]

It has a micro bolometric matrix type non-cooled detector with a thermal resolution of 384×288 pixels and a thermal sensitivity of less than 0.08 °C [121]. It offers nine distinct palettes for image acquisition, as shown in Fig. 4.2 maintaining the same palate order and numbering provided in the thermal imaging camera.

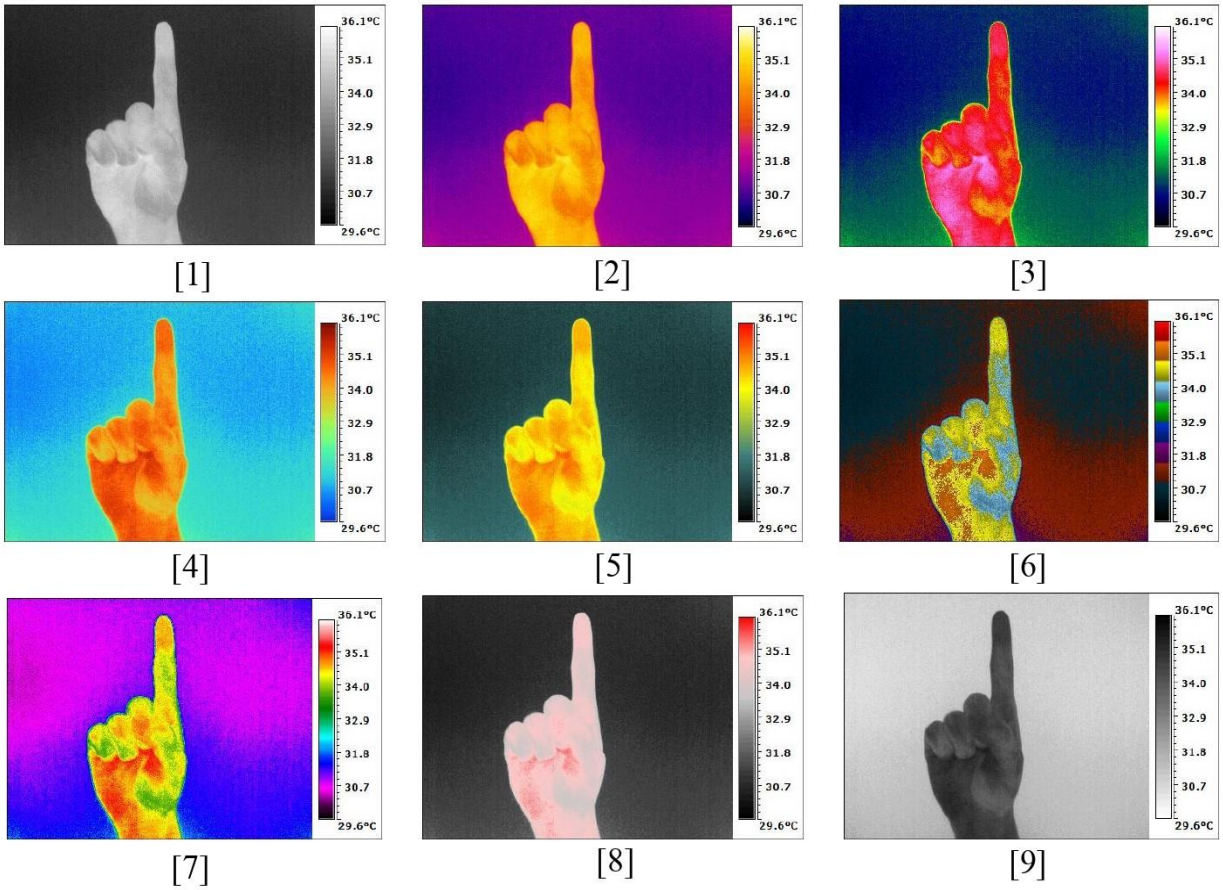


Fig. 4.2 Nine palettes of Sonel® KT-384 from 1 to 9

In the lab setting, where the humidity level was approximately 63% [122] and the temperature range was between 22-24 °C, images were captured without any pre-processing and added to the database. Palette 1, which captures each image as a gray scale image, was utilized for the database. Additionally, a database was created using the Iron Bow Palette shown in Fig. 4.2, named as the IR scale palette, for further reference in the thesis. The compositional structure of the databases and pre-processing details will be discussed in the following sections.

4.3. COMPOSITIONAL STRUCTURE OF THE DATABASES

The images in the gesture database were collected under diverse illumination conditions. During the data acquisition process, participants were provided with general guidance about placing their hands, and they were not restricted to keeping their finger at a fixed position from the camera. This approach ensured that the database was more realistic and reflective of practical

scenarios, where a robot must be able to detect the position of a person's finger in any reasonable proximity within its field of view. To construct the gesture database, 10 different hand gestures, namely "one," "two," "three," "four," "five," "six," "seven," "eight," "nine," and "ten," were collected from each participant.

4.3.1. Thermal Image Capturing

The hands in photographs were moved around within the frame to capture a variety of motions. Occasionally, participants moved their hands naturally or were instructed to do so. The posture in which participants held their hands when taking photographs caused natural movements. Fig. 4.3 illustrates different types of hand motions. The initial position is shown in Fig. 4.3(a), Fig. 4.3(b) demonstrates what happens when the hand is intentionally bent and a small portion of the face enters the frame, and Fig. 4.3(c) (naturally moved) illustrates what happens when the hand moves naturally towards the body after being stationary for an extended period. This approach created natural variations and increased the number of variations in the dataset for the same hand movement.

(a) (b) (c)

Fig. 4.3. Frame-specific Example of Hand and Body movement: (a) Hand position 1, (b) Hand position 2, (c) Hand position 3

Variations are induced by any movement of the hand or fingers. Fig. 4.4 illustrates an example of differences. While some individuals struggle to keep their fingers apart, as seen in Fig. 4.4(a), others can do so, as seen in Fig. 4.4(b). The dataset's diversity is enhanced by including individuals with various hand shapes and levels of movement, resulting in natural variance. This approach makes every gesture more varied, providing the computer with more information to consider. If the algorithm has not received adequate training data from various examples, it is more difficult to interpret an image in a real-world scenario. The participants' diversity of nationalities also increases the possibility of obtaining various gesture shapes. Their upbringing and line of work

may also influence the nature and flexibility of their hands. The dataset was created using a consistent background to make the hand stand out and highlight its differences from others. Although the quality of the photos may vary based on the conditions, it is noted that people are naturally warmer than the interior room where the images were taken. The issue of hand motions appearing frigid is illustrated in Fig. 4.4. The problem that arises when the camera perceives the hand as cold is depicted in Figs. 4.4(c), 4.4(d), and 4.4(e). Although the photographs were taken from January to March and the ambient temperature was lower than during the spring and summer, there are a number of possible explanations for the purple fingers. People may have been outside strolling before taking pictures, making their skin colder than the reference wall in the background. Lower body temperatures may also occur depending on where in the world they are. A few °C will be seen in thermal pictures, despite the fact that there shouldn't be any major effects. Wet hands are more likely to be colder or closer to room temperature due to individual characteristics, and other factors might also be relevant, but they will all change the dataset and make it more difficult for computers to learn. Therefore, the quality and effectiveness of the suggested algorithms can be assessed more easily using such a broad dataset. In order for the hands to be effectively captured, it is necessary for the arms to be held at the sides in front of the camera while capturing thermal photographs by people.

(a) (b) (c) (d) (e)

Fig.4.4. Example of difference between images, with position, gesture and temperature differences: (a)Good temperature difference, medium hand position, (b)Good temperature difference, good hand gesture, (c)Poor temperature difference, good hand gesture, (d)Poor temperature difference, good hand gesture, (e)Poor temperature difference, medium hand gesture.

4.3.2. Database: 1

The camera is placed at a fixed angle and distance for each image. A distance of 50 to 60 cm is maintained between the hand and camera to capture highly detailed photographs. Although it is technically possible to capture images from a greater distance. Fig. 4.4(a), 4.4(b) illustrate high-detail photographs in which all fingers are clearly visible with distinct skin temperatures. As the images are taken from further away, they begin to lose detail, as shown in Fig. 4.4(c).

In this experimental setup, a tripod is not used. The thermal camera is held by hand for capturing the image, while a chair is used for support to hold the thermal camera.

A total of 2000 images are acquired from 20 subjects, with 100 images taken from each subject. There are 10 classes/gestures, resulting in 10 images per class. The dataset comprises 4000 images, divided into grayscale and IR scale, with 2000 images in each. Each movement is separated into ten gestures, and there are 200 images for each gesture. The same person's collection of ten hand signs is illustrated in Figs. 4.6 and 4.7.

4.3.2.1. Preprocessing

The pre-processing techniques employed on the raw images acquired are depicted in Fig. 4.5 by a simple block diagram.

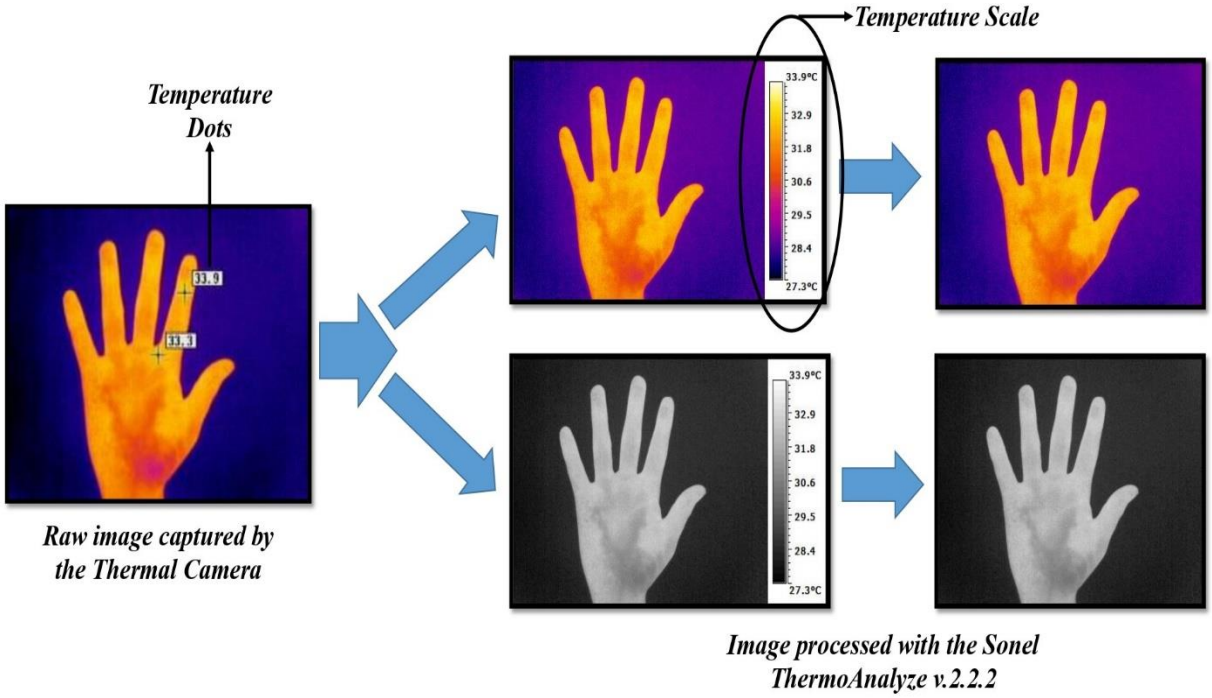


Fig. 4.5. Pre-processing employed for hand gesture raw thermal images acquired

The raw image displayed in Fig. 4.5 does not adhere to any standard color palette. However, after undergoing preprocessing, the images can be visualized using a standardized color palette, as depicted in Fig. 4.5.

In the first step of capturing the hand gesture images, thermal data was captured using a Sonel thermal camera. However, the captured images contained temperature dots that had to be removed to improve their quality.

Next, the Sonel ThermoAnalyze v-2.2.2 software was used to filter out the temperature dots from the raw thermal data. However, a temperature gauge bar was attached during the process of saving the images, which was needed to be removed.

For this purpose, a software was used to edit the images and remove the gauge bar. This step was crucial to ensure that the images were suitable for further processing and analysis and that they accurately represented the hand gestures being performed.

All images should show only the normal or personal variations for each person, and they should all resemble these movements as closely as possible. The collection consists of 200 grayscale and fusion images for each gesture, arranged by gesture. Images vary in size from fusion to grayscale. Fusion images take up more space, with each image occupying about 25 kB and the total size being 18 MB. Smaller than color photos are grayscale images, which add up to 3.73 MB for a single gesture and weigh in at 17 kB each. The dataset, which includes 20 participants, has an overall size of about 35 MB and provides some fine-grained thermal images suitable for image recognition. With a larger dataset and more data, higher accuracy can be achieved by the algorithm, but training time may increase. Integrating thermal images into any machine learning algorithm and computer platform will be easier, and the small memory requirement for integration and storage will be beneficial [123].

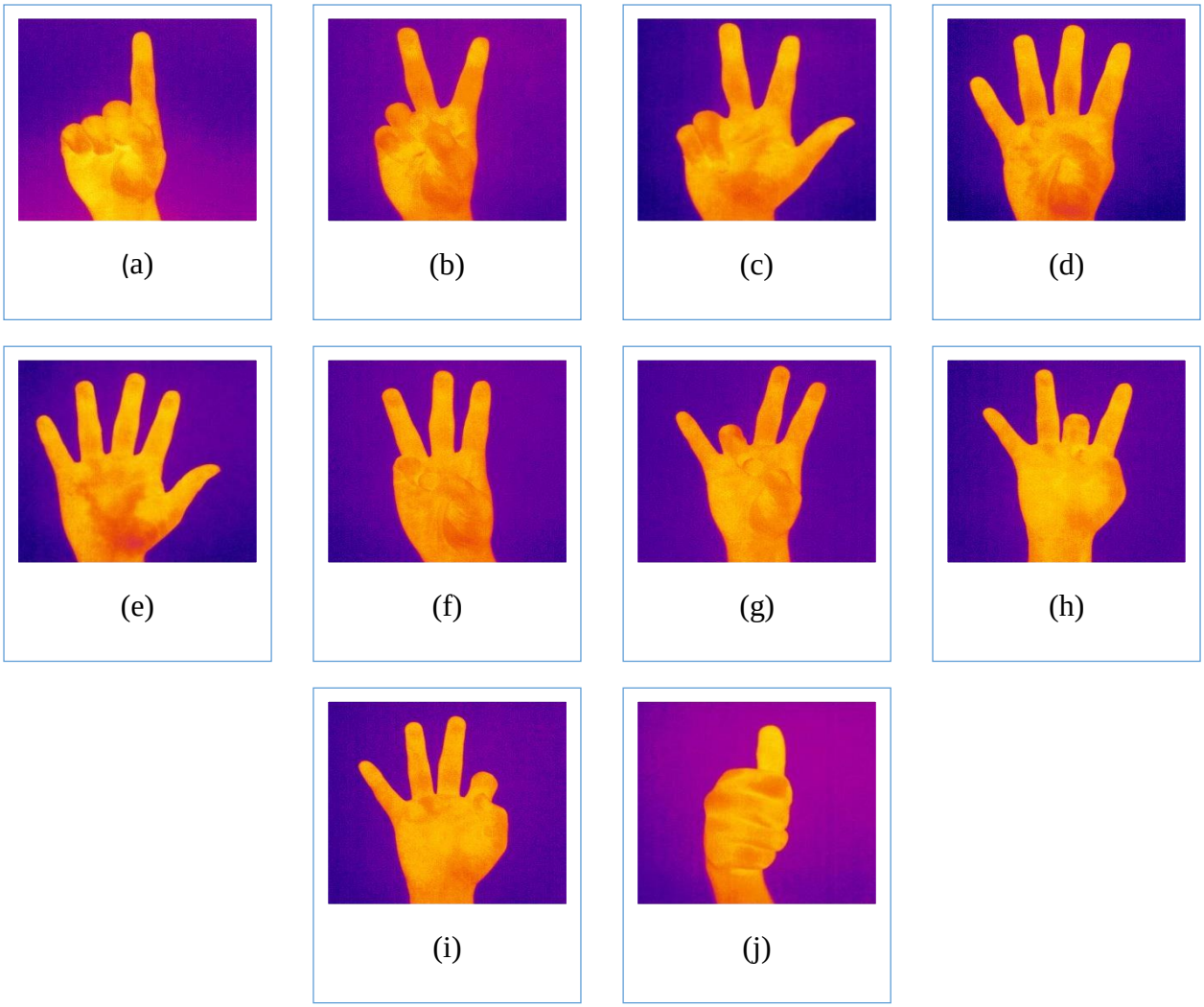


Fig. 4.6. A complete set of IR Scale thermal images: (a) IR image 1, (b) IR image 2, (c) IR image 3; (d) IR image 4, (e) IR image 5, (f) IR image 6, (g) IR image 7, (h) IR image 8, (i) IR image 9 and (j) IR image 10.

Fig. 4.7. A complete set of grayscale thermal images: (a) Grayscale image 1, (b) Grayscale image 2, (c) Grayscale image 3, (d) Grayscale image 4, (e) Grayscale image 5, (f) Grayscale image 6, (g) Grayscale image 7, (h) Grayscale image 8, (i) Grayscale image 9 and (j) Grayscale image 10.

4.3.3. Database: 2

A database of occluded images was created for further experiments, in addition to the hand gesture database described earlier. Rather than using natural occlusion, a computer-generated rectangle-shaped occlusion was added to the original images. The purpose of this was to test the recognition schemes' robustness later where the gesture images are captured with occlusion, which

can often occur in real-world scenarios.

Two levels of occlusions were created, referred to as Level-1 and Level-2, as depicted in Fig. 4.8 and Fig. 4.9, respectively. Level-1 was characterized by a block occlusion of 30%, while Level-2 exhibited a block occlusion of 40%.

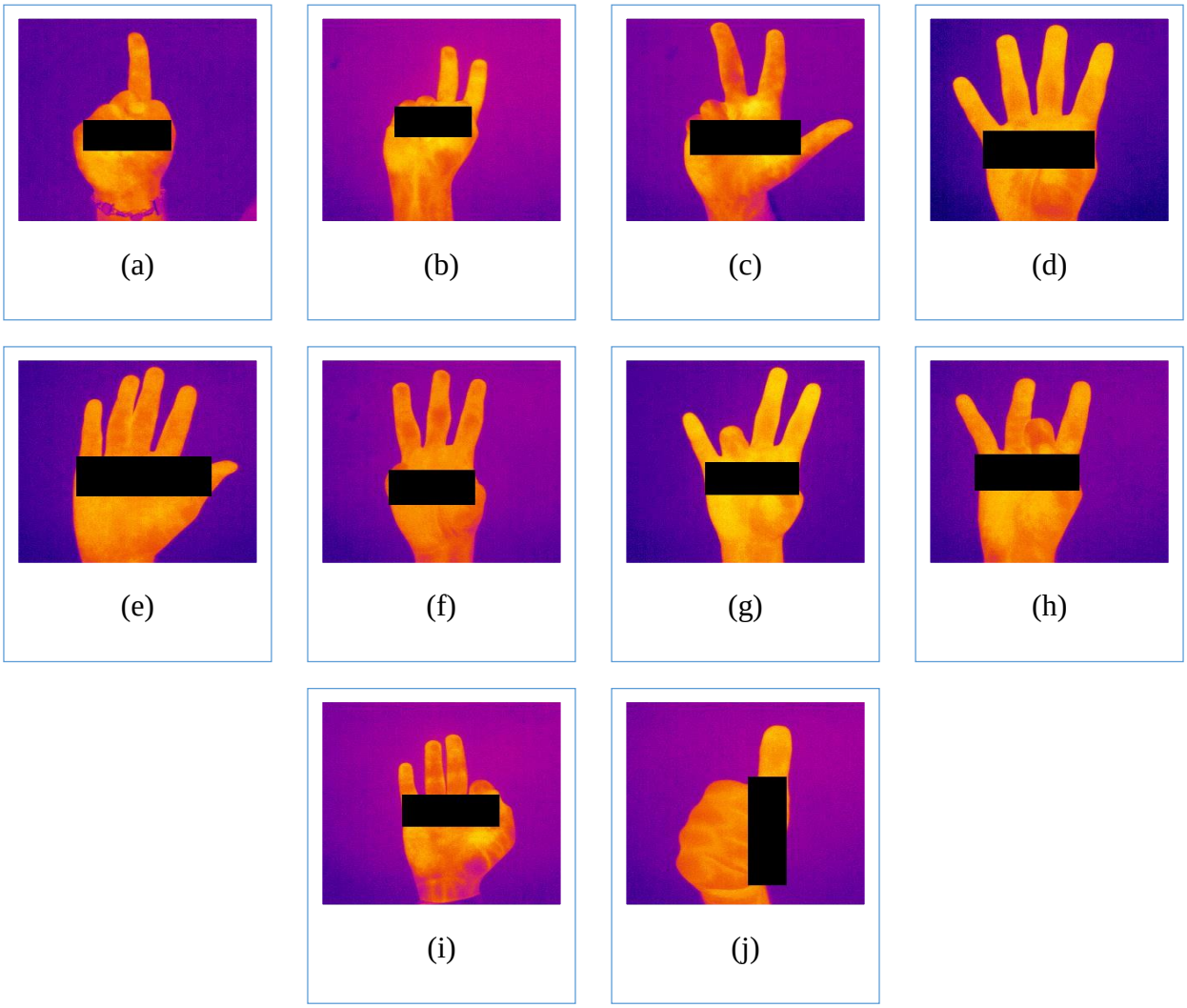


Fig. 4.8. Complete set of occluded IR scale hand sign thermal images at occlusion level 1

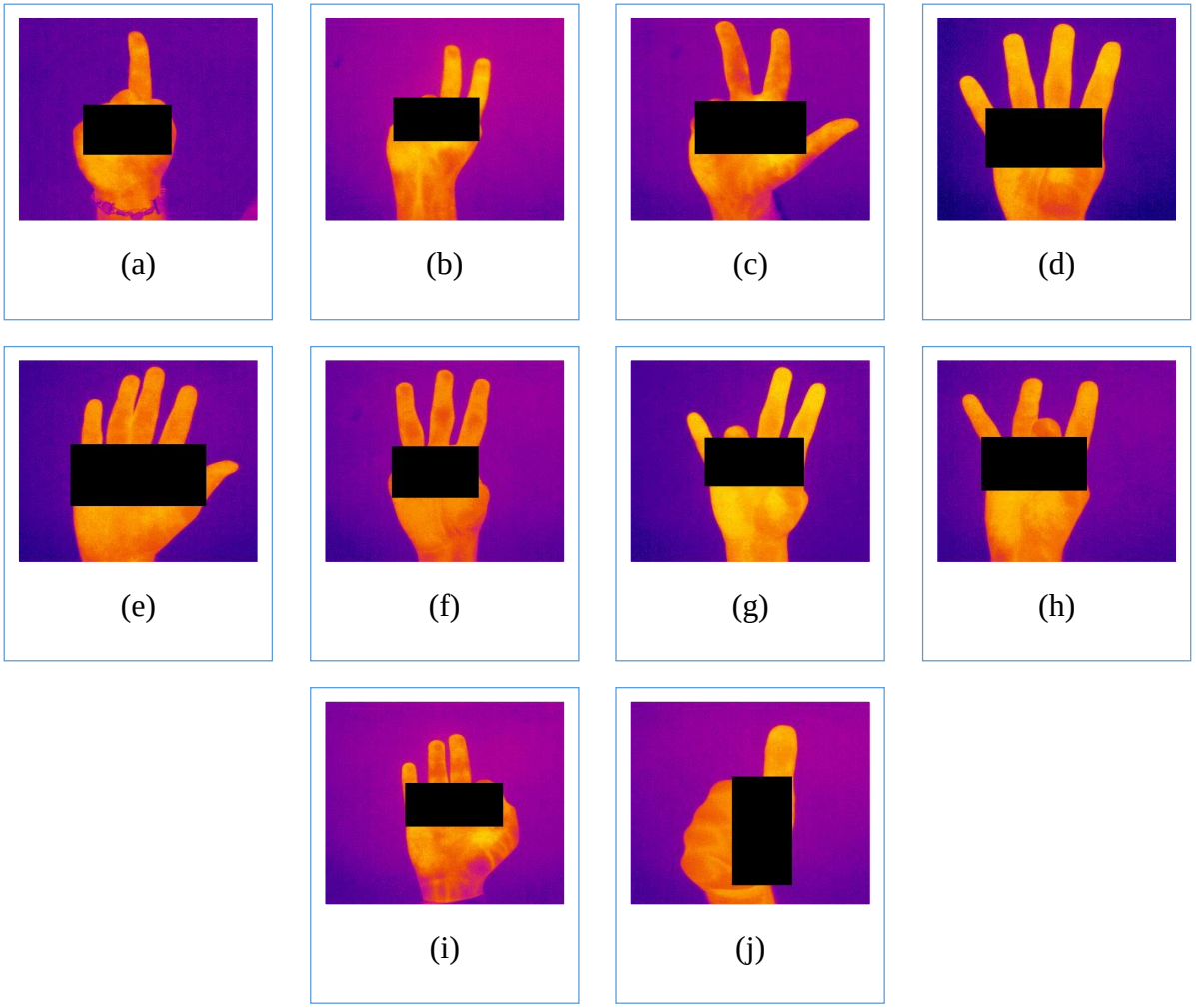


Fig. 4.9. Complete set of occluded IR scale hand sign thermal images at occlusion level 2

4.4. SUMMARY

In this chapter, detailed descriptions for creation of three databases have been discussed which comprises sign language based thermal hand gesture databases, both with and without occlusion, and creation of a thermal imaging of forearm vein pattern based biometric database. Next, detailed discussions will be presented about how hybrid-uniform mixture error modeling based iterative robust coding algorithms can be developed for these image recognition problems in challenging scenarios. In the following chapters, two different methodologies,

namely the recently proposed Laplacian-Uniform mixture (LUM) error model and the Student's-t-Uniform (TUM) error model, for iterative robust coding, will be discussed in detail, with a thorough analysis of their mathematical models, feasibility of implementation, limitations, and results. Proposed improvements to these methods are also presented and supported by experimental results. Overall, the thesis aims at developing effective, sophisticated algorithms which can be advantageously used for developing (i) a reliable sign language based hand gesture recognition system for collaborative robotics and (ii) a reliable biometric system using thermal forearm vein pattern imaging. The claims made in the thesis are validated by the use of experimental data.

CHAPTER-5

SIGN LANGUAGE BASED GESTURE RECOGNITION USING LAPLACIAN UNIFORM MIXTURE MODEL

- ***INTRODUCTION***
- ***REVIEW OF PREVIOUS WORK***
- ***LAPLACIAN UNIFORM MIXTURE-DRIVEN ITERATIVE ROBUST CODING***
- ***ITERTATIVELY REWEIGHTED*** $l_1 - l_1$ ***MINIMIZATION***
- ***THE WEIGHT FUNCTION***
- ***THE LUM ALGORITHM***
- ***CLASSIFICATION STRATEGY IN LUM***
- ***EXPERIMENTAL RESULTS***
- ***SUMMARY***

5.1 INTRODUCTION

In recent years, attention has been given to sparse representation based classification methods due to their ability to effectively capture the underlying structure of high-dimensional data. However, the traditional assumption of a Gaussian or Laplacian error distribution may not always hold true in practical scenarios, particularly when dealing with dense and gross outliers. Alternative error models such as the maximum correntropy criterion [129] or general M-estimators [117] or maximum likelihood [136] have been proposed in recent approaches, although a more insightful understanding of their suitability is still required. In this context, a mixture model based on the Laplacian and uniform distributions has been found to be more appropriate for handling dense errors. The proposed objective function introduces a nonconvex error term that cannot be properly addressed by the conventional iteratively reweighted least-squares formulation. An iterative local linear approximation method is proposed to optimize the objective function, which is optimal in the sense that it is the closest to the original function in the set of local convex approximations that preserve up to the first-order derivative. The iteratively reweighted $l_1 - l_1$ minimization algorithm is then proposed, which offers a cooperative approach to error detection and correction. Experimental results are provided to demonstrate the effectiveness of the proposed approach in addressing robust gesture recognition systems.

5.2 REVIEW OF PREVIOUS WORKS

SRC can represent a vectorized test image y as a sparse linear combination Ax while minimizing the $l_0 - norm \| x \|_0$, given a $m \times n$ dictionary A made up of n vectorized m-dimensional training images from all classes. When the genuine x is sparse [126], [127], the representation is distinct, which can be used for recognition. To address errors in the representation, the equality constraints $Ax = y$ is relaxed to $\| Ax - y \|_2 \leq \in, \in > 0$ [143]. In the signal processing and statistics communities [144], the analogous issue is known as basis pursuit denoising [125] and can be translated into the following $l_2 - l_1$ minimization problem:

$$\hat{x} = \arg\min_x \| Ax - y \|_2^2 + \lambda \| x \|_1 \tag{1a}$$

Here λ is a Lagrange multiplier. This is the stable version of SRC (S-SRC). The $l_1 - norm$

term enforces sparsity regularization, whereas the $l_2 - norm$ term in (1) denotes fidelity or error of the representation.

Let vector $e \in \mathbb{R}^m$ be the coding errors. A contaminated probe sample y can be represented by

$$y = Ax + e \tag{1b}$$

In this chapter, $\mathbb{R}^m$ denotes the space of m-dimensional real column vectors, $\mathbb{R}^{m \times n}$ denotes the space of $m \times n$ dimensional real matrixes. For a vector e and a matrix A, e_i signifies the i-th element of e, and a_i signifies the i-th column of A. The $l_2 - norm$, $l_1 - norm$, $l_{2,1} - norm$ and $l_p - norms$ of a vector $x \in \mathbb{R}^n$ are defined as $\| x \|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$, $\| x \|_1 = \sum_{i=1}^{n} | x_i |$, $\| x \|_{2,1} = \sum_{i=1}^{n} \| x_i \|_2$, *and* $\| x \|_p = (\sum_{i=1}^{n} x_i^p)^{\frac{1}{p}}$ respectively [159].

We present a broad framework for a common understanding of the regression-based techniques we examined. Hence the optimization problem can be represented as [159]:

$$\min_{e,x} \sum_{i=1}^{m} \rho(e_i) + \frac{1}{\alpha} \sum_{i=1}^{n} b_i \pi(x_i) \qquad s.t.\ e = y - Ax \tag{2}$$

Where $\sum_{i=1}^{m} \rho(.)$ represents a fidelity term to characterize e, $\sum_{i=1}^{n} \pi(.)$ is a penalty term to select features, b_i is a supervisory factor, and $1/\alpha$ is a tunable parameter.

5.2.1. Error Correction-Based Methods

These methods focus on formulating $\sum_{i=1}^{m} \rho(.)$ and assigning $\sum_{i=1}^{n} \pi(.)$ with $\| x \|_1$ and b_i with 1. Assuming e to be sparse, RSRC [145] used an identity matrix I to fit e and extended $y = Ax + e$ to $y = \bar{A}\bar{x}$, where $\bar{x} = [x_A; x_I]$ and $\bar{A} = [A, I]$. Let $e = Ix_I$. RSRC can be formulated as

$$\min_{e,x_A} \| e \|_1 + \frac{1}{\alpha} \| x_A \|_1 \qquad s.t.\ e = y - Ax_A \tag{3}$$

NMR [150] assumed e to be low-rank and assigned $\sum_{i=1}^{m} \rho(.)$ with $\| e \|_*$.

Without making any assumptions about the errors, HQ-A [151] measured e using the M-

estimator, with the corrected part coming from HQ-A optimization.

5.2.2. Regularization-Based Methods

These methods pay main attention to $\sum_{i=1}^{n}\pi(.)$ and b_i assign $\sum_{i=1}^{m}\rho(.)$ the l_2- or l_1-norm. In SRC [145], sparse codes x over A were produced by giving $\sum_{i=1}^{n}\pi(.)$ the l_1-norm. Without taking into account b_i, SRC accomplished sparse coding by

$$\min_{e,x} \| e \|_2^2 + \frac{1}{\alpha}\| x \|_1 \qquad s.t.\ e = y - Ax \tag{4}$$

By omitting b_i, GSC [146] assigned $\sum_{i=1}^{n}\pi(.)$ with $\| x \|_{2,1}$ to encourage the structural sparsity of x. To explore the locality of x, WSRC [147] and WGSC [148] computed b_i as

$$b_i = \exp(\| y - a_i \|_2^2 / \sigma^2), \qquad i = 1, 2, ...N. \tag{5}$$

Here σ is a kernel parameter. To better recover the sparsity of x, GISA [149] assigned $\sum_{i=1}^{n}\pi(.)$ the $l_p - norm$.

In earlier work [98], an MLE (Maximum likelihood Estimation) solution for reliable image representation was provided, which was motivated by the robust regression theory [153]-[154]. Re write A as $A = [r_1; r_2; ,.....,; r_n]$, where $r_i = i^{th}$ row of A, let $e = y - A\alpha = [e_1; e_2; ,...,; e_n]$, where $e_i = y_i - r_i\alpha,\ i = 1, 2, ..., n$. The so-called robust sparse coding (RSC) [98] was developed as the following l_1 -sparsity restricted MLE problem, where e_i has a probability density function (PDF) of $f_\theta(e_i)$, where θ indicates the unknown parameter set that defines the distribution. [let $\rho_\theta(e) = -\ln f_\theta(e)$]

$$\min_{\alpha} \sum_{i=1}^{n} \rho_\theta(y_i - r_i\alpha) \text{ s.t. } \|\alpha\|_1 \le \sigma \tag{6a}$$

The aforementioned RSC model, like SRC, takes the coding coefficients to be sparse and employs the l_1-norm to describe the sparsity. It has recently been suggested in [152] that the l_1-sparsity constraint on is not the key for the success of SRC [94], despite the fact that the l_1-sparsity requirement increases the complexity of RSC. In this study, proposed regularized robust coding (RRC), a more comprehensive paradigm. While RSC is one particular implementation of the RRC

concept, the RRC has a far higher potential for efficiency.

From the perspective of Bayesian estimation, specifically the maximum a posterior (MAP) estimation, let's look at the image representation problem. The MAP estimation of the coding vector α is $\hat{\alpha} = \arg\max_{\alpha} \ln P(\alpha \mid y)$, after coding the input image y over the supplied dictionary A. The Bayesian formula has led to this point

$$\hat{\alpha} = \arg\max_{\alpha} \{\ln P(y \mid \alpha) + \ln P(\alpha)\} \tag{6b}$$

5.2.3. Error Detection-Based Methods

When dense errors occur, the l_1 minimization imposed on the errors results in a sparsity constraint on the solution, which is undesirable in reality. Hence, the resilience of the quasi-norm concept in classification is utilized in [131]. The precise selection of the quasi-norm is largely heuristic, though. He *et al.* [129] introduced correntropy-based sparse representation (CERS) to increase robustness in sparse representation.

$$\hat{x} = \arg\max_{x} \sum_{j=1}^{m} g(y_j - A_j.x) - \lambda \| x \|_1 \tag{7}$$

Where $g(.)$ is a Gaussian kernel defined by $g(e) = \exp(-e^2 / (2\sigma^2))$, and y_j *and* A_j denote, respectively, the *j*th component of y and the *j*th row of dictionary A. It is assumed that $x \geq 0$ (component wise) for simplicity [129]. The Welsch M-estimator [130], [132] and correntropy have a close relationship. The repeatedly reweighted least-squares problem results from optimization and the final form can be given as:

$$\omega_j^t = g(y_j - A_j.x^t), \qquad j = 1, 2, ..., m \tag{8}$$

$$x^{t+1} = \arg\min_{x} \| \sqrt{W^t} (y - Ax) \|_2^2 + \lambda \| x \|_1 \tag{9}$$

Here t denotes the t-th iteration and $W^t = diag(\omega^t)$ is a diagonal matrix constructed from the weight vector ω^t with components ω_j^t. The CESR method iteratively finds the faults and gives pixels the appropriate weights. It has a strong ability to handle non-Gaussian noise and outliers.

The multiplicative form of the half-quadratic (HQ)-based optimization HQ-M taken into consideration in [117] is related to CESR. He *et al.* also proposed an algorithm that iteratively estimates and corrects the faults based on the additive form of the HQ framework (HQ-A). When

compared to R-SRC, HQ-A does not assume that errors e are sparse, making it better able to handle dense errors.

CESR was integrated by HQ-M [151] into the HQ optimization framework, where ω was dependent on the M-estimator that was chosen. ω Was given the Logistic function by RRC, where

$$\omega_j = \exp(-\mu e_j^2 + \mu\delta) / (1 + \exp(-\mu e_j^2 + \mu\delta)) \tag{10}$$

Where δ controls the demarcation point and μ controls the decreasing rate. Accordingly

$$\rho(e_j) = -\frac{1}{2\mu} \ln \frac{1 + \exp(-\mu e_j^2 + \mu\delta)}{1 + \exp \mu\delta} \tag{11}$$

Experimental findings [136] showed that RRC was superior to earlier approaches like SRC and CESR.

Despite their robustness to errors in image recognition, these methods are largely heuristic and treat error correction and detection as separate processes. In recent studies, attention has been given to low-ranked matrix decomposition in robust face recognition [135], [137], [138]. However, these methods are not well-suited to handle data with dispersed occlusions or corruptions. This issue has been addressed in subsequent studies, which have attempted to combine discontiguous error modeling with low-rank representation [128], [133]-[135]. Additionally, there have been new studies that utilize Laplacian-uniform mixture [116] to focus on error modeling in face images subjected to arbitrary contiguous occlusions or discontiguous corruptions. The LUM method will be briefly discussed in this context.

5.3. LAPLACIAN UNIFORM MIXTURE-DRIVEN ITERATIVE ROBUST CODING

Despite the robustness of previous M-estimator-based methods to gross errors, as cautioned by Huber [130], it is important to note that the concept of robustness primarily addresses deviations from a model at a small scale. However, when substantial contamination is present, it often signifies the requirement of a mixture model. This need for a mixture model might not be apparent in robust procedures, emphasizing the significance of data analysis and diagnostics.

It is worth mentioning that the density corresponding to equation (11) is predominantly heuristic in nature. To provide a more intuitive understanding, it can be observed that the weight ω_j in RRC (Robust Regularization based Coding [93]) can be equivalently derived from a Gaussian-uniform mixture (GUM) density of the errors.

$$f(e_j) = \alpha(1 + \exp(-\mu e_j^2 + \mu\delta)) \tag{12}$$

Here α is a normalization factor. The associated weight function is:

$$\omega_j = 2\mu \exp(-\mu e_j^2 + \mu\delta) / (1 + \exp(-\mu e_j^2 + \mu\delta)) \tag{13}$$

This function is equivalent to (10) up to a constant scaling factor.

In general, mixture models can be beneficial in situations where there is substantial contamination or deviations from a model. Robust procedures based on M-estimators may not always be sufficient in handling such situations, as they are typically designed to be robust to small deviations from a model. To address this, a mixture model approach can be used to account for both inliers and outliers in the data. By combining multiple probability distributions, a mixture model can better fit the distribution of the data, leading to more accurate and robust results. In particular, the proposed Laplacian-Uniform Mixture (LUM) model [116] has been shown to provide superior fitting performance compared to other density functions, especially in the presence of outliers. This can be helpful in a variety of applications, including image matching and sparse representation, where accurate and robust modeling of the data is critical. So, the LUM model is:

$$f(e_j) = \alpha\left(\exp\left(-|e_j|/\mathrm{b}\right) + \mathrm{c}\right) \tag{14}$$

Here $\mathrm{b} > 0$ corresponds to the parameter of the Laplacian component, $\mathrm{c} > 0$ is a constant corresponding to the outlier distribution (uniform over the range of values), and α is a normalization factor.

The capability of modeling the distribution of residuals around zero, which reduces the probability of false error detection, was demonstrated by the LUM model. The residuals mainly resulted from minor perturbations instead of outliers and corresponded to pixels that were significant in identity recognition. In the optimization algorithm, pixels with substantial residuals were deemed less reliable and given small weights, leading to their eventual exclusion from image matching. The proposed LUM exhibited better fitting to challenging residual distributions compared to other error models.

To account for the potential random dispersion of outliers in an image, the assumption of independent and identically distributed coding residuals is made. The formulation of the objective function involves both the negative log-likelihood function, denoted as $-\ln f\left(\mathrm{e}_j\right)$, and sparsity constraints. Alternatively, the objective function can be expressed as follows [116]:

$$x^* = \arg\min_x \sum_{j=1}^{m} \rho(e_j) + \lambda_0 \| x \|_1$$

$$s.t. \quad e = y - Ax \tag{15}$$

$$\rho\left(\mathrm{e}_j\right) = -\ln\left(\exp\left(-\left|\mathrm{e}_j\right| / \mathrm{b}\right) + \mathrm{c}\right) \tag{16}$$

The objective function of (15) comprises the first and second terms, which aim to ensure fidelity and sparsity of the representation, respectively. The parameter $\lambda_0 > 0$ is used to balance the importance of these terms in the optimization process.

5.4. ITERATIVELY REWEIGHTED $l_1 - l_1$ MINIMIZATION

The objective function (15) presents challenges in optimization due to its non-differentiability at $\mathrm{e}_j = 0$ and nonconvex nature. Directly optimizing this function becomes difficult. Although the first term bears resemblance to a typical M-estimator, using the conventional iteratively reweighted least-squares formulation $\omega_j^t \mathrm{e}_j^2$, where $\omega_j^t = \rho'\left(\mathrm{e}_j^t\right) / \mathrm{e}_j^t$, is not suitable. One of the reasons is that as the denominator e_j^t approaches 0 during the optimization process, ω_j^t can become arbitrarily large. This instability arises because the absolute value function $\left|\rho'\left(\mathrm{e}_j\right)\right|$ approaches a positive value $\left(1/\mathrm{b}\left(1 + \mathrm{c}\right)\right)$, as defined in (16). Consequently, the iterative procedure becomes unstable.

To gain a deeper understanding, the expansion of (16) is performed by utilizing the Taylor expansion up to the second order at the current estimation of e.

$$\rho(e_j) = \rho(|e_j^t|) + \rho'(|e_j^t|)(|e_j| - |e_j^t|) + \frac{1}{2}\rho''(|e_j^t|)(|e_j| - |e_j^t|)^2 + o(|e_j|) \tag{17}$$

Here $o(|e_j|)$ denotes the higher order terms.

One approach that can be considered is the relaxation of the constraint on the coefficient of the second-order term by replacing $\rho(|e_j^t|)$ in (17) with a positive coefficient $\omega(e_j^t) > 0$. This replacement aims to achieve a quadratic convex function of e_j in the second-order approximation, while ensuring a first-order derivative at the origin $\hat{\rho}_2'(0) = 0$, similar to the approach employed in [136]. As a result, the equation $\rho'(|e_j^t|) = \omega(e_j^t)|e_j^t|$ is obtained, which is equivalent to the

conventional M-estimator in effect. However, this approach also confronts the instability issue mentioned earlier.

It is proposed to approximate $\rho(e_j)$ by preserving only up to the first-order term of its Taylor expansion in equation (18).

$$\hat{\rho}_1^t(e_j) = \rho(|e_j^t|) + \rho'(|e_j^t|)(|e_j| - |e_j^t|) \tag{18}$$

Upon substituting equation (18) into equation (15) and eliminating terms that are not relevant to the argument being optimized, the resulting expression is derived.

$$x^{t+1} = \arg\min_x \|W^t e\|_1 + \lambda \|x\|_1$$

$$s.t. \quad e = y - Ax \tag{19}$$

Here W^t is a diagonal matrix with the diagonal element

LUM [116] assigned ω the Logistic function, in which

$$\omega_j^t = \frac{\exp(-|e_j^t|/b)}{\exp(-|e_j^t|/b) + c} \propto \rho'(|e_j^t|) \tag{20}$$

$e^t = y - Ax^t$, and λ is parameter that absorbs λ_0 and b. The optimization procedure defined by equation (19) involves solving a reweighted $l_1 - l_1$ minimization problem at each iteration t. The weight matrix W^t is responsible for assigning small values to outliers (with large errors $|e_j^t|$) and serves as an error detection mechanism. Additionally, it is widely recognized that $l_1 - l_1$ minimization exhibits robustness to gross errors and offers an error correction mechanism. This mechanism becomes particularly significant in challenging situations where the detection of gross errors by W^t may be limited, especially during the initial stages preceding the optimization of W^t .

5.5. THE WEIGHT FUNCTION

The mixture distribution $f(e_j)$ as described in equation (14) suggests that for foreground pixels, there exists a Laplacian distribution $f(e_j | \upsilon_j = 1) \propto \exp(-|e_j|/b)$, where υ_j represents the label of the j th pixel, with 1 indicating foreground and 0 representing the outlier. Subsequently, the posterior probability of the current pixel being foreground given e_j can be calculated as [116]:

$$P(\upsilon_j = 1 \mid e_j) = \frac{f(e_j \mid \upsilon_j = 1)P(\upsilon_j = 1)}{f(e_j)} \propto \omega_j^t \tag{21}$$

The weight function derived in LUMIRC is deemed statistically plausible, as it assigns weights to the pixels based on the posterior probabilities of the pixels being foreground. This weight assignment takes into account the errors as evidence within the LUM error modeling framework. [116]

5.6. THE LUM ALGORITHM

It is worth highlighting that CESR and HQ-M approaches, employing the Welsch M-estimator [117], yield a weight function (8) that corresponds to a Gaussian likelihood distribution, represented as $f(e_j \mid \upsilon_j = 1) \propto \exp(-e_j^2 / (2\sigma^2))$. Similarly, through analysis, it can be verified that the weight function (10) associated with RRC (Robust Regularization and Compensation) is equivalent to a posterior probability derived from the Gaussian-uniform mixture (GUM) model in (12).

However, it is crucial to consider the limitations of the Gaussian or GUM models when handling diverse outliers. These models may exhibit reduced effectiveness in accurately characterizing and accommodating the complex nature of outliers encountered in practical scenarios. In contrast, the proposed LUM (Laplacian Uniform Mixture) model demonstrates potential advantages in addressing the challenges posed by various outlier types.

The optimization problem represented by Equation (19) relates to a robust coding problem that needs to be addressed at each iteration. To effectively solve this problem, the primal form of the augmented Lagrangian method (PALM) [139] is adopted. PALM is particularly well-suited for tackling the extended l_1 minimization problem, as depicted in $\hat{c} = \arg\min_c \| c \|_1 \quad s.t. \; Bc = y$, by leveraging the inherent structure of the matrix B. The steps involved in solving the LUM minimization problem using iteratively robust coding are outlined in Algorithm 5.1 [116]

5.1 Algorithm Laplacian Uniform Mixture Model Based Iterative Robust Coding Algorithm

---

Input: Testing image***y***, Training data Matrix ***A***.

Output: ID ($\boldsymbol{y}$), that is, the identify the test image $\boldsymbol{y}$

1. Initialization:

$t = 0$,

$x^0 = \frac{1}{n}1$, where 1is a vector of all 1's,

$\beta^0 = 0$.

2. Repeat

3. $e^t = y - Ax^t$

4. Compute W^t with diagonal element $\omega_j^t = \frac{\exp(-|e_j^t|/b)}{\exp(-|e_j^t|/b)+c}$, *for* $j = 1, ..., m.$

5. $x^{\sim 0} = x^t, \beta^{\sim 0} = \beta^t$

6. $l = 0.$

7. **Repeat**

The solution of $e^{\sim l+1} = \arg\min_{e^{\sim}} \| e^{\sim} \|_1 + \frac{\upsilon}{2} \| d_1^l - e^{\sim} \|_2^2$ can be obtained explicitly by a 1-D shrinkage (soft thresholding) function [141], [139], [140]

$$e^{\sim l+1} = sign(d_1^l) \max\{|d_1^l| - 1/\upsilon, 0\} \tag{22}$$

Here $d_1^l = W^t(y - Ax^{\sim l}) + \beta^{\sim l}/\upsilon$ and here all the operation are performed component wise.

8. **Compute** $e^{\sim l+1}$ from the equation (22)

9. **Solve** $x^{\sim l+1} = \arg\min_{x^{\sim}} \frac{\upsilon}{2} \| d_2^l - W^t A x^{\sim} \|_2^2 + \lambda \| x^{\sim} \|_1$ (23)

Here $d_2^l = W^t y + \beta^{\sim l}/\upsilon - e^{\sim l+1}$, the equation (23) represents a common $l_2 - l_1$ minimization problem, which can be effectively solved by utilizing the fast iterative shrinkage-thresholding algorithm (FISTA) [124], as recommended in [139].

10. **Compute** $\beta^{\sim l+1} = \beta^{\sim l} + \upsilon(W^t(y - Ax^{\sim l+1}) - e^{\sim l+1})$

11. $l = l + 1$

12. **until Convergence.**

13. $x^{t+1} = x^{\sim l}$, $\beta^{t+1} = \beta^{\sim l}$.

14. $t = t + 1$.

15. until Convergence of the overall algorithm

5.7. CLASSIFICATION STRATEGY IN LUM

Upon the convergence of the iteratively robust coding problem, as outlined in Algorithm 5.1, a classification strategy is employed for the query test images. The classification strategy utilized in LUM [116], developed by Dajun Lin *et al.*, was adopted for robust face recognition via sparse representation. In this method, the test image y is classified as,

$$\text{(Class level)}\ \ ID(y) = \arg\min_i \{G_{cl}\} \tag{24}$$

Here, $G_{cl} = \| W^{t-1}(y - A\delta_i(x^t)) - e^{\sim l} \|_2^2$. The convergence of the overall algorithm is achieved by iterating over (24), and convergence can be verified by checking if $(\| x^{t+1} - x^t \|_2 / \| x^t \|_2)$ less than a small positive value is $\in$. To classify the test image y, the sparse solution must fulfill the condition $W^t y - e^{\sim} = W^t A x$. Consequently, the determination of y's identity is based on identifying the best reconstruction that can be achieved from the components in x corresponding to a single class, which can be expressed as follows:

$$ID(y) = \arg\min_i \| W^*(y - A\delta_i(x^*)) - e^{\sim *} \|_2^2 \tag{25}$$

The values at convergence are denoted by the superscript $*$, and the vector $\delta_i(x^*) \in \mathbb{R}^n$, where i represents the i th class, is a vector in which only the entries of x^* corresponding to that specific class are nonzero.

5.8. EXPERIMENTAL RESULTS

Now, this LUM based iterative robust coding algorithm is implemented for sign language based hand gesture recognition problems in challenging environments, using thermal imaging. The hand gesture thermal image database acquired in the laboratory of Jadavpur University, Kolkata is utilized for conducting experiments. The images in the database are captured under challenging illumination and temperature conditions. Here, the experiments consist of two different sets:

(i) Sign language recognition using thermal images in low-light situations and night vision,

(ii) Sign language recognition using thermal images in low-light situations and night vision with occlusion.

The entire database is randomly divided into training and testing sets with a ratio of 7:3, and the dictionary A is derived from the training dataset.

The results reported in this chapter reveal that LUM achieves higher recognition accuracy

compared to RRC and RSC for images captured under illumination and temperature constraints. Moreover, as the level of corruption in the images increases, particularly due to occlusions, the performance of LUM outperforms RRC. This indicates the superior capability of LUM in handling challenging scenarios with occlusions.

5.8.1. Using the Laplacian Uniform Mixture Model

To observe the advantages of mixture modeling, an illustrative example is presented in Fig. 5.1. The figure showcases the empirical distribution of errors in sparse representation and the effectiveness of different density models in fitting these errors. The training set and the test image utilized in this example are obtained from the sign language hand gesture database specifically acquired for this study, as described in *Chapter 4*. It should be noted that thermal images are employed for constructing the database.

To ensure accuracy and gain meaningful insights into how the distributions of coding residuals change in the presence of different types of outliers, the sparse coding coefficients are computed based on the clean image depicted in Fig. 5.1. This is accomplished using the $l_1 - 1_s$ algorithm [142], and these coefficients remain fixed throughout the analysis, following the approach adopted in [136].

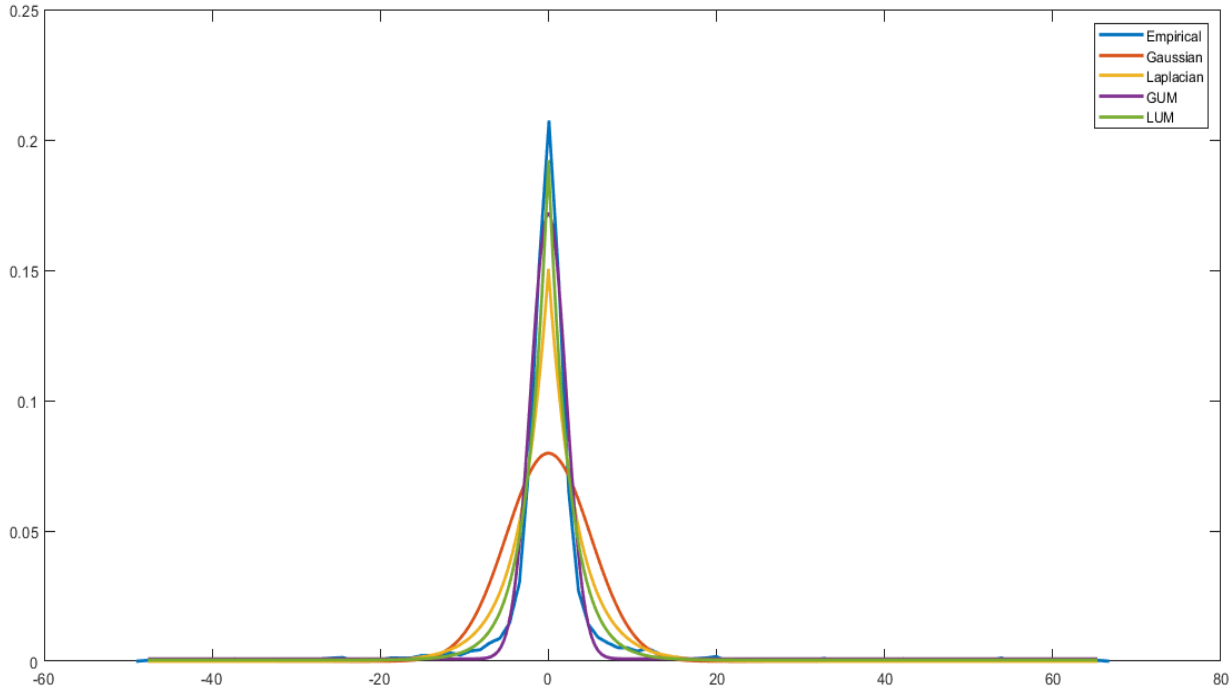


Fig. 5.1. Empirical, Gaussian, Laplacian, GUM and LUM based probabilistic distribution of residual errors for a single test image.

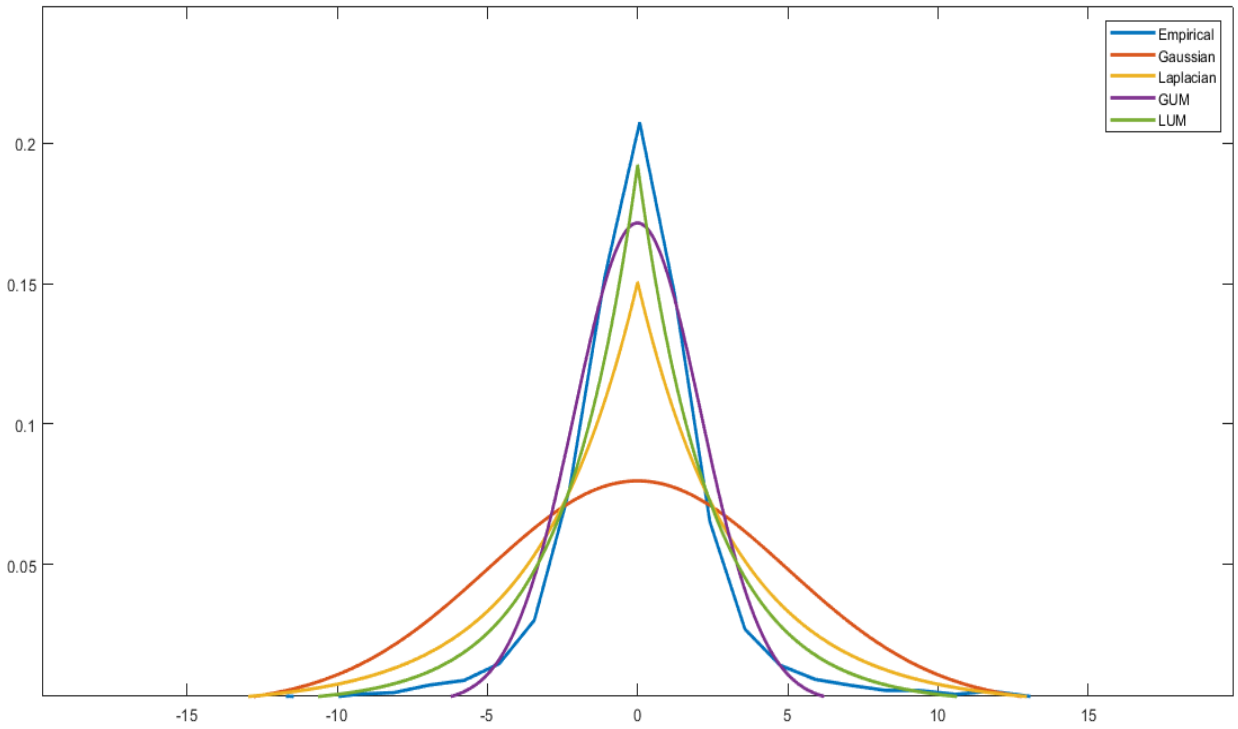


Fig. 5.2 Magnified view of Fig. 5.1 near the peak of Empirical and fit distributions of sparse representation errors for the test image without occlusion. Here LUM (Green line) is the best fitted model with Empirical.

The total log-likelihoods of the coding residuals estimated with established model developed are compared in this table, to evaluate the goodness of fit of density functions.

Table 5.1: The log likelihood of fit models

Model	***Gaussian***	***Laplacian***	***GUM***	***LUM***
Log-likelihood	20461.2	22500.1	22794.2	***22930.4***

The utilization of images with real occlusions resembling those in the AR face database was not feasible due to the absence of associated ground-truth clean images. In *chapter-4*, it was demonstrated that the Gaussian distribution and the Laplacian distribution fail to accurately model the coding residuals when outliers such as occlusion or corruption occur in the occluded database (level-1 and level-2). However, there is a relative improvement observed with the Laplacian distribution. Notably, the GUM (Gaussian-uniform mixture) exhibits an additional advantage in effectively modeling various coding errors, resulting in improved log-likelihoods.

Earlier research [119] has revealed that Gaussian distributions may not be highly effective as the inlier model for errors encountered in image matching. This is corroborated by the findings presented in table 5.1, where the Laplacian distribution outperforms the Gaussian distribution in terms of likelihood when modeling the inlier distribution. Based on the hypothesis, introducing a mixture model with the Laplacian distribution as the inlier model holds the potential for achieving enhanced robustness compared to previous methods. Given that outliers are assumed to be unpredictable in advance, a uniform distribution is employed as the corresponding outlier model, similar to the approach employed in [120].

5.8.2. Parameter Tuning

The LUM model provides the flexibility to fit the distribution by adjusting the parameters: ***b, c, and λ (Lambda)***. The parameter c represents the outlier model and takes values greater than 0, representing a uniform distribution. By varying **c from 0.5 to 3 in steps of 0.5**, different levels of outliers can be modeled, allowing the LUM model to handle varying degrees of extreme data points.

The parameter b, on the other hand, controls the scaling of the Laplacian distribution, which is used to model the inlier data. By adjusting **b from 0.5 to 3 in steps of 0.5**, the spread or scale of the Laplacian distribution can be modified. This parameter influences the width of the error distribution and plays a crucial role in capturing the variability of the inlier data.

Lastly, the parameter λ is a Lagrange multiplier, which governs the trade-off between fidelity to the data and sparsity of the representation. By varying **λ from 0.5 to 3 in steps of 0.5**, the LUM model can adapt to different levels of sparsity in the coding coefficients. This parameter allows for controlling the balance between accurately representing the data and promoting sparsity in the coding process.

By exploring different combinations of b, c, and λ within their specified ranges, the LUM model can be fine-tuned to fit the specific distribution characteristics of the data. This flexibility enables the model to effectively capture the underlying patterns and structures, leading to improved performance in various tasks, such as face recognition or image analysis.

In the earlier discussion, the variations of the parameters b, c, and λ (lambda) in the LUM model were examined. A systematic exploration was conducted, where these parameters were varied across a range of values to assess their impact on the model's performance.

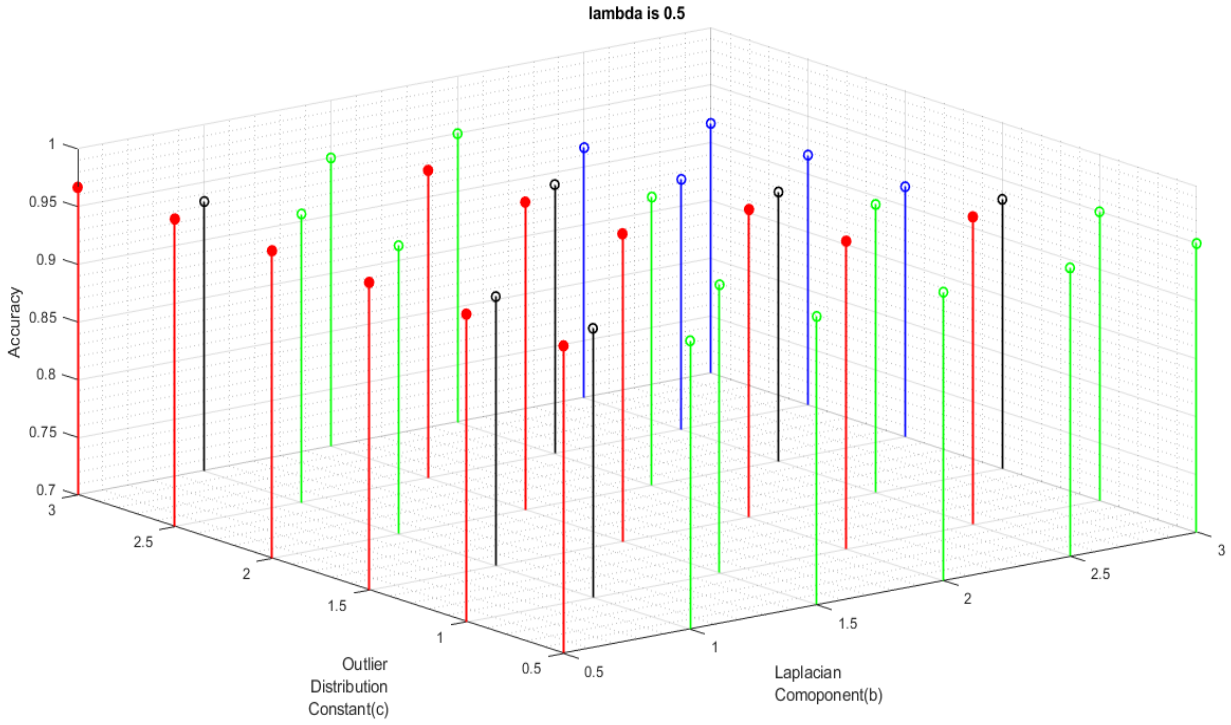
5.8.3. Results

The results obtained from applying the LUM model to the Non-Occluded entire database, which consisted of thermal hand gesture images, was split into training and testing sets in a ratio of 7: 3. The performance of LUM is presented in below Table 5.3. A total of 216 combinations of the parameters b, c, and λ were assessed. The graphical representation of the accuracy changes resulting from the variations in parameters b, c, and λ are presented in Fig. 5.3 (a), (b), (c), (d), (e) and (f). These plots provide visual insights into the relationship between the parameter values and the corresponding performance of the LUM model.

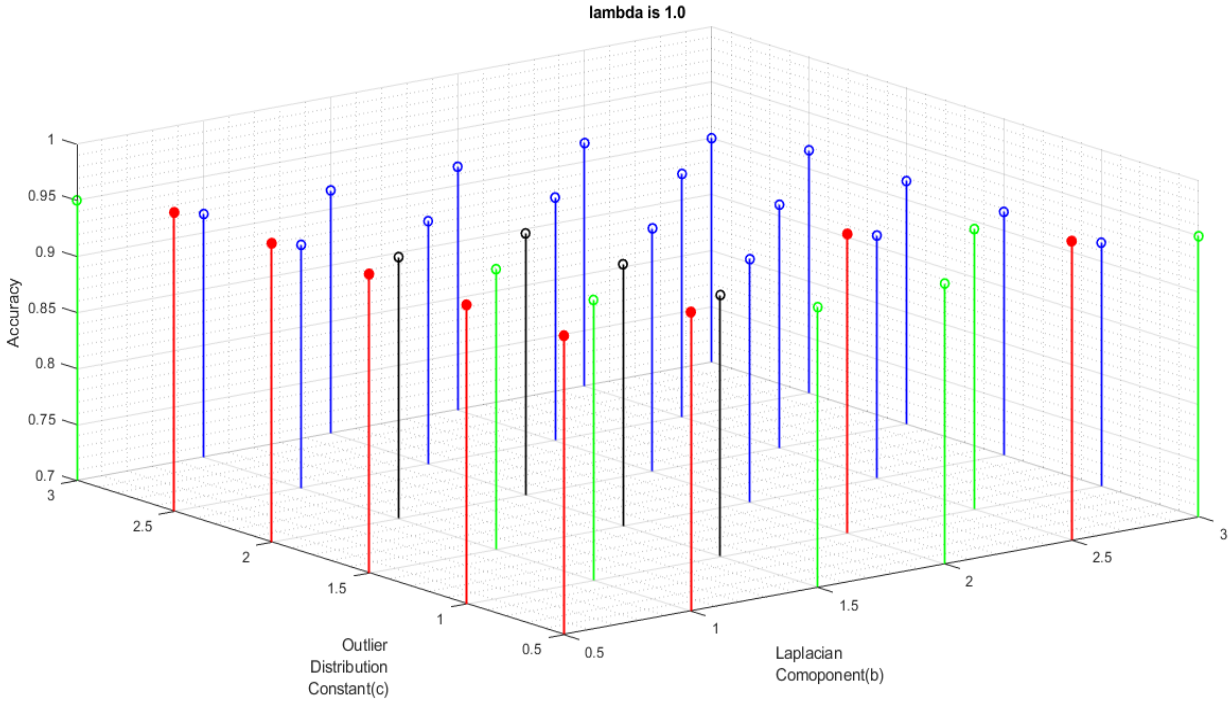
Table 5.2 represents the recognition rate range mapping table, which serves to comprehend the graphical representation of the LUM model's performance accuracy.

Table 5.2: Recognition Rate Range Mapping with Color for LUM

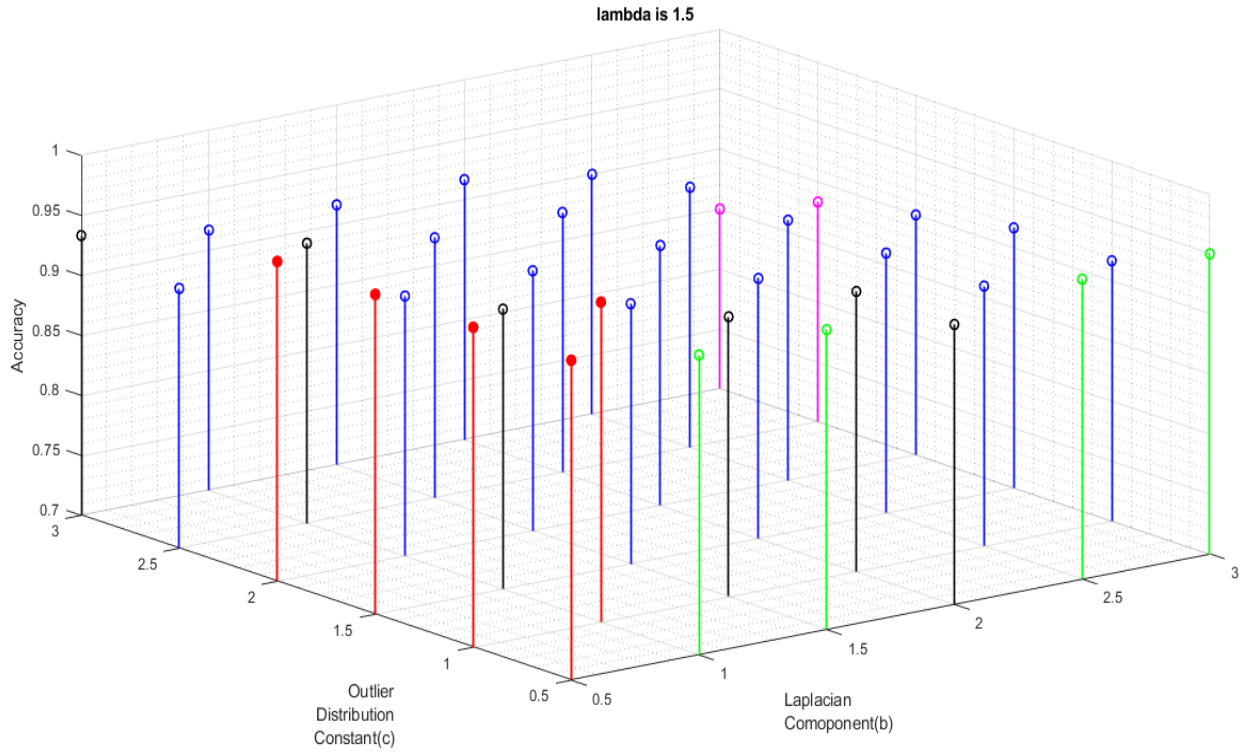
Recognition Rate Range	***Color***
Below 0.90	Magenta
0.90 to 0.93	Blue
0.93 to 0.95	Black
0.95 to 0.96	Green
Above 0.96	Red


lambda is 0.5
Accuracy
1
0.95
0.9
0.85
0.8
0.75
0.7
Outlier
Distribution
Constant(c)
3
2.5
2
1.5
1
0.5
Laplacian
Comoponent(b)
0.5
1
1.5
2
2.5
3

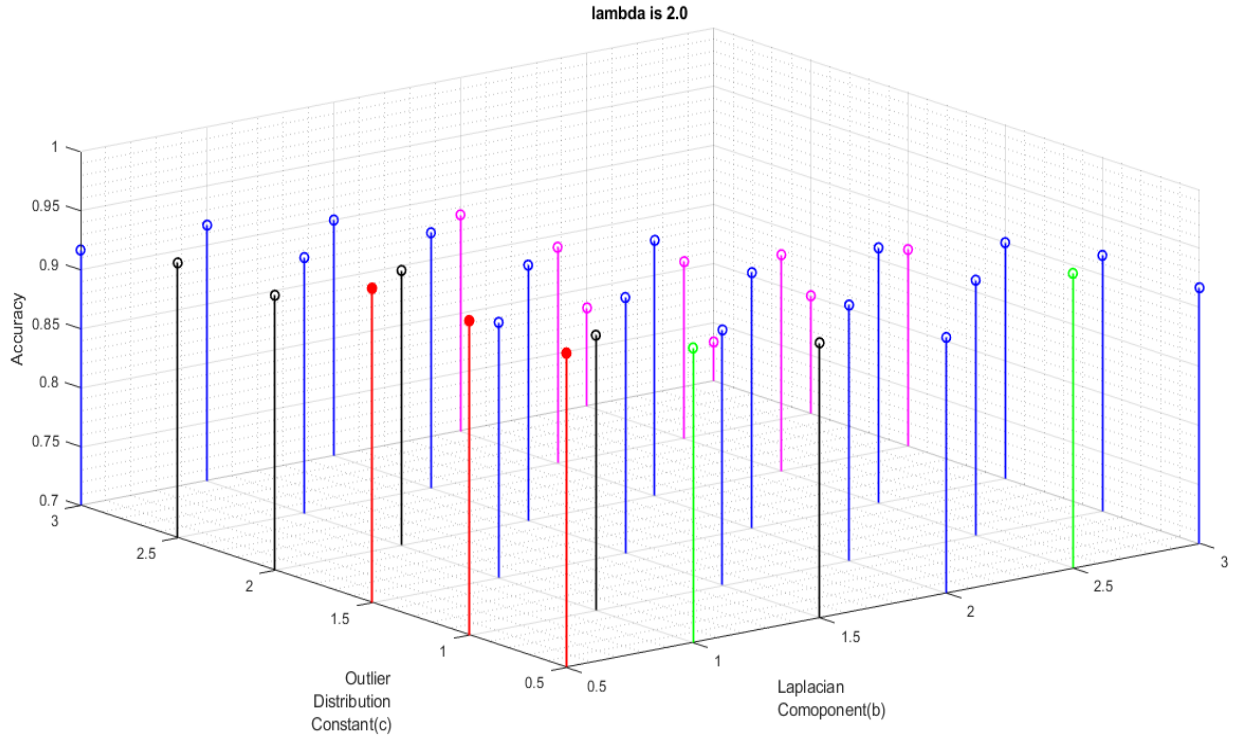
(a)


lambda is 1.0
Accuracy
1
0.95
0.9
0.85
0.8
0.75
0.7
Outlier
Distribution
Constant(c)
3
2.5
2
1.5
1
0.5
Laplacian
Comoponent(b)
0.5
1
1.5
2
2.5
3

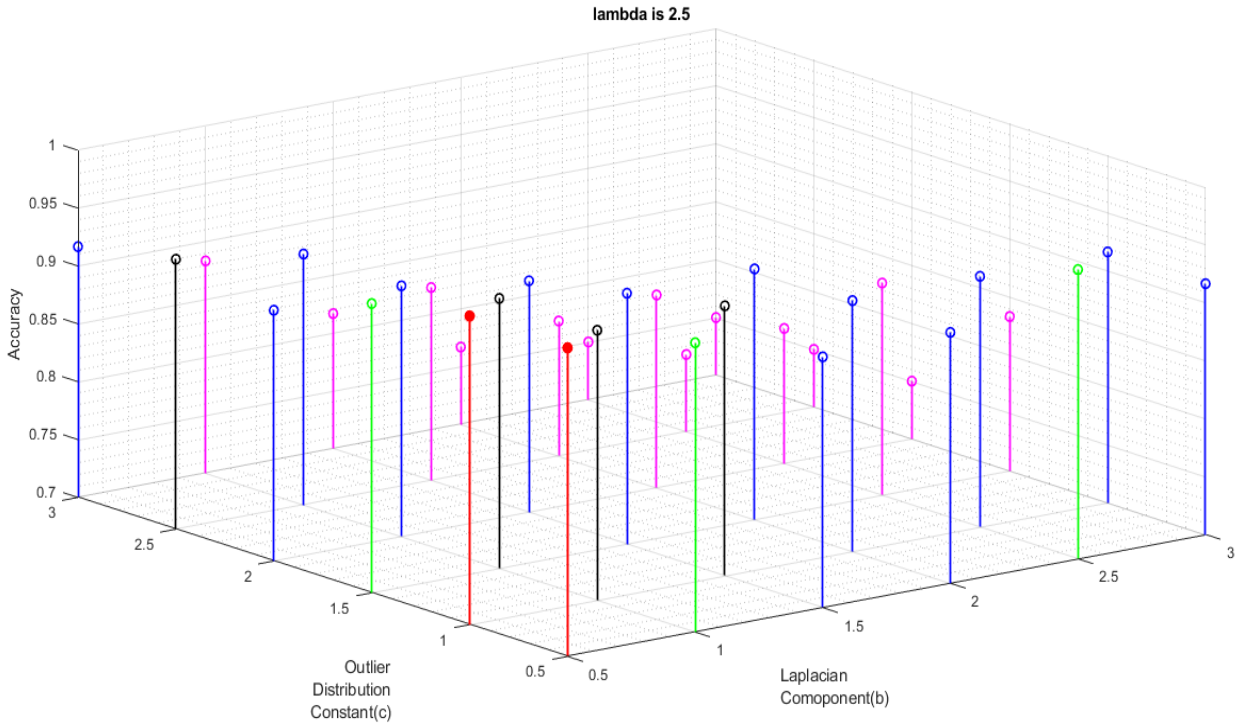
(b)


lambda is 1.5
Accuracy
1
0.95
0.9
0.85
0.8
0.75
0.7
3
2.5
2
1.5
1
0.5
Outlier
Distribution
Constant(c)
0.5
1
1.5
2
2.5
3
Laplacian
Comoponent(b)

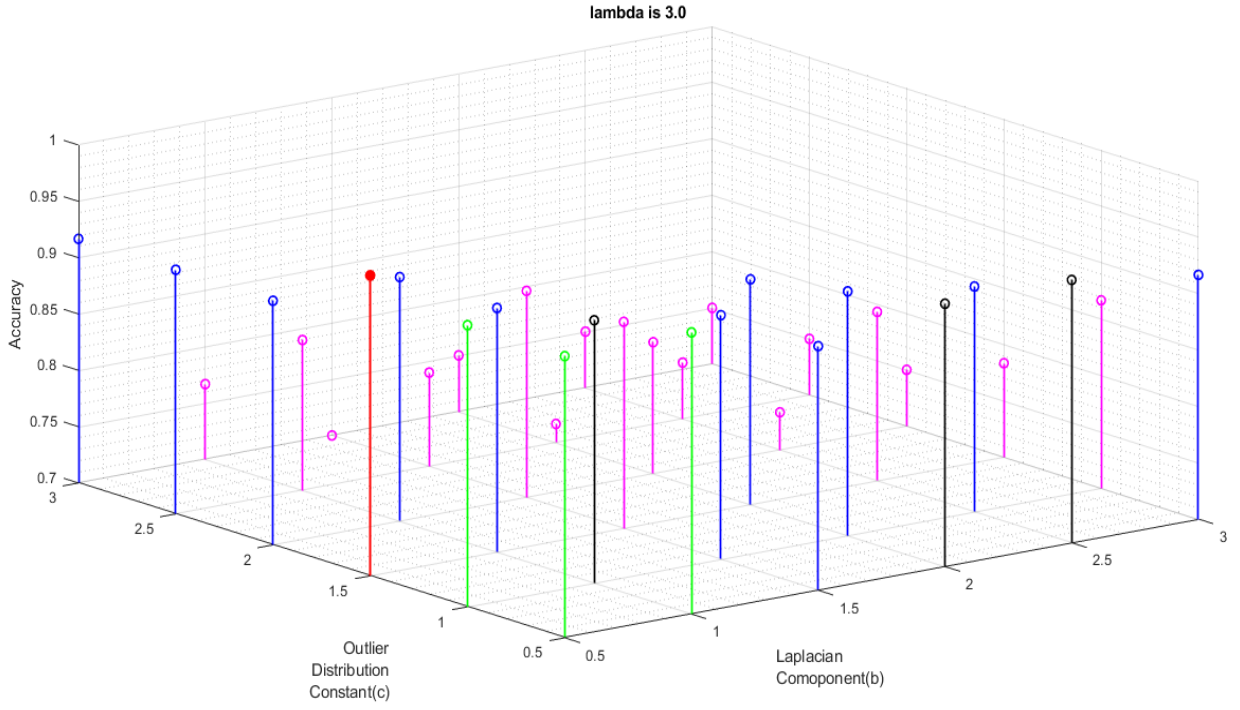
(c)


lambda is 2.0
Accuracy
1
0.95
0.9
0.85
0.8
0.75
0.7
3
2.5
2
1.5
1
0.5
Outlier
Distribution
Constant(c)
0.5
1
1.5
2
2.5
3
Laplacian
Comoponent(b)

(d)



(e)



(f)

Fig. 5.3. The graphical representations of the accuracy changes using LUM

Table 5.3: Performance of LUM with Non-Occluded Sign Language Thermal Image Datasets

Maximum Accuracy (in %)	***96.67%***
Minimum Accuracy (in %)	***75.00%***
Standard Deviation (in %)	***5.95%***
Mean (in %)	***90.88%***

In *chapter 4*, the occluded datasets were discussed to examine the improved robustness of the LUM method in practical scenarios. The experiment involved dividing the occluded data into two stages: level 1 and level 2 occlusion. Level-1 was characterized by a block occlusion of 30%, while Level-2 exhibited a block occlusion of 40%. Combinations of parameters b, c, and $\boldsymbol{\lambda}$ (lambda) are selected based on their ability to achieve higher recognition rates. The performance of the LUM model with Occluded Sign Language Thermal Images for dataset of occluded images are evaluated and presented in Table 5.4 and 5.5 for level-1 and level-2, respectively.

Table 5.4: Performance of LUM for occluded Level 1 dataset

Recognition Rate	***0.9167***
Accuracy (in %)	***91.67%***

Table 5.5: Performance of LUM for occluded Level 2 dataset

Recognition Rate	***0.8667***
Accuracy (in %)	***86.67%***

The present chapter showed how LUM algorithm can be employed for sign language based gesture recognition using thermal images. However, so far we have used Level 1 and Level 2 occluded datasets in non-occluded environments. In the next chapter more challenging scenario with occlusion will be considered. In the following chapter, the improvement of the previous results will be examined by employing the model proposed in this thesis, the Student's t Uniform Mixture model (TUM). The focus will be on evaluating the enhanced performance achieved through the

utilization of this model. The TUM model, which will be introduced next, will be employed to address the limitations identified with the LUM method, when employed in scenarios with occlusions. By using the TUM model, the aim will be to overcome the challenges associated with accurately modeling coding residuals in the presence of occlusions or corruptions. Through the passive utilization of the TUM model, the improvements observed in the likelihood estimation and the overall robustness of the model will be investigated. The significance of incorporating the TUM model in the context of image matching and recognition will be thoroughly discussed, providing insights into its effectiveness and potential benefits.

5.9. SUMMARY

This chapter introduced the LUM (Laplacian Uniform Mixture) method and its effectiveness in robust thermal sign language hand gesture recognition. The method utilizes an iterative robust coding approach, which has been proven highly effective in gesture recognition. LUM stands out for its ability to handle different types of outliers, such as improper illumination and occlusion. By assigning weights to pixels based on their residual errors, the iterative robust coding algorithm reduces the impact of outlier pixels with higher coding errors, ensuring robustness. LUMIRC (Laplacian Uniform Mixture Iterative Robust Coding) assigns smaller weights to coding residuals associated with gross errors, maintaining sparsity in the overall solution even when dense gross errors are present. The optimization process of the LUM model naturally leads to l_1-norm minimization of coding residuals [116]. Experimental results in this section support the improved image reconstruction achieved by LUMIRC, which preserves sparsity in coding residuals ($W^t e$, i.e., coefficients corresponding to the error dictionary I). Compared to other state-of-the-art methods like RRC, RSC, and previous versions [92], LUM demonstrates superior performance, positioning it as a promising choice for robust sign language hand gesture detection systems.

Given the insights provided in this chapter and the advantages associated with LUM, there is an important consideration that presents opportunities for researchers to further enhance its recognition rate. The weight function described in Equation (20) serves as a logistic function for weight updating in the iterative robust coding algorithm. However, being a monotonically

decreasing and continuous exponential function, this logistic function assigns weights to pixels in a smooth and continuous manner. Consequently, there is a risk of excessively penalizing pixels with higher residual errors, even if they are non-corrupted and significant in the images. Conversely, pixels with lower residual errors, but are actually corrupted, may be assigned higher weights due to the continuous nature of the logistic function. To address this concern, the thesis proposes more robust and error-tolerant weight functions that make more judicious choices when allocating weights to specific pixels. The subsequent chapter offers a comprehensive examination of these weight functions, including their mathematical formulations, accompanied by extensive experimental results.

CHAPTER-6

RECOGNITION OF SIGN LANGUAGE USING STUDENT'S T UNIFORM MIXTURE (TUM) MODEL

- ***INTRODUCTION***
- ***STUDENT'S T-UNIFORM MIXTURE-DRIVEN ITERATIVE ROBUST CODING***
- ***STUDENT'S UNIFORM MIXTURE FOR ERROR MODELING***
- ***WEIGHT FUNCTION MODELING***
- ***DISCUSSION ON THE WEIGHT FUNCTION***
- ***EXPERIMENTAL RESULTS***
- ***SUMMARY***

6.1. INTRODUCTION

The experimental results presented in [136] provided compelling evidence to support the superiority of RRC over previous methods such as SRC and CESR. Building on this progress, the previous chapter further demonstrated the exceptional performance of LUM compared to RRC. LUM, a Laplacian-uniform mixture approach, was specifically developed for error modeling in sparse representation, offering significant advancements in the field.

While previous methods have showcased their robustness in face recognition by effectively handling errors, their treatment of the error term has predominantly relied on heuristic approaches. This limitation calls for a more thorough analysis to better understand and address the error component. It is worth noting that these methods have also been applied to sign language thermal images, demonstrating their versatility across different domains.

Furthermore, the existing techniques have primarily treated error correction and error detection as separate entities, neglecting the potential benefits of their cooperative utilization within sparse representation. This represents an untapped opportunity for improving the overall performance and resilience of the methods.

Recently, low-rank matrix decomposition has gained considerable attention in the field of robust face recognition, as evidenced by the studies cited in [135], [137], and [138]. However, it is essential to acknowledge that the assumption of errors exhibiting a low-rank structure may not hold in practical scenarios, especially when face images are subjected to dispersed occlusions or corruptions. Recognizing this limitation, subsequent research endeavors have explored the combination of discontinuous error modeling with low-rank representation [128], [133], [134], and [135]. Despite these efforts, the methods still fall short in effectively handling arbitrary outliers, compromising their robustness.

In the previous chapter, the focus was on the discussion of error modeling in sparse representation using the Laplacian-uniform mixture (LUM) approach. Subsequently, an iterative robust coding (IRC) algorithm, referred to as LUMIRC, was developed, demonstrating the distinct property of error detection and error correction unification in a cooperative manner. In this chapter, an algorithm named Student's t Uniform Mixture model was developed, which is based on the student's t distribution. By utilizing the proposed algorithm, sparse representation is enhanced with exceptional robustness.

To address the aforementioned challenges, the proposed method in this chapter focuses

specifically on error modeling in sign language thermal images that are subject to arbitrary contiguous occlusions. By leveraging the insights gained from previous studies, the developed algorithm, based on the student's t distribution, introduces the **Student's t Uniform Mixture model (TUM)**. This novel approach enhances sparse representation with exceptional robustness, outperforming state-of-the-art methods.

6.2. STUDENT'S T-UNIFORM MIXTURE-DRIVEN ITERATIVE ROBUST CODING

According to the literature [116], the coding errors cannot be accurately modelled by Gaussian or Laplacian distributions in the presence of outliers like corruption or occlusion. When compared to only Gaussian distributions, GUM shows a better log-likelihood, especially when outliers are involved. As observed in image matching situations [119], while Gaussian distributions are excellent at modelling the outliers, they are less effective at modelling the inlier pixels. Laplacian models might perform better in these circumstances because of their sharper peak at zero, which is appropriate for fitting the inliers. A uniform distribution can be included to create a LUM model, which addresses the Laplacian model's representation of outlier pixels [116]. The modification of Student's t-distribution is offered for modelling the residual error [160] based on these observations regarding the accommodation of inliers and outliers in an error distribution curve. The Student's t-distribution is a good option, especially for modelling outlier data because of its distinctive features, which include its longest tail and more sparsity compared to Gaussian or Laplacian distributions [161]. Compared to Gaussian or Laplacian distributions, the usage of Student's t-distribution can better balance the system's sparsity and computing complexity [161]. The suggested solution achieves computational efficiency when compared to the previously mentioned methods. How effectively the inlier data may be fitted depends on the t-distribution curve's peak's degree of sharpness, which is governed by the parameter v. Higher values of result in distribution peaks that are higher, which enhances the fitting of inlier data. The student't distribution's probability density function (PDF) is described as follows:

$$tpdf = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)}\left(1+\frac{e_j^2}{v}\right)^{-\frac{v+1}{2}} \tag{1a}$$

$$\text{Let, } k = [\frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})} \frac{1}{\sqrt{v\pi}}]$$

$$\text{So now, } \quad tpdf = k((1+\frac{e_j^2}{v})^{-\frac{v+1}{2}}) \tag{1b}$$

Here, $c > 0$ is a constant corresponding to the outlier distribution (uniform over the range of values) and v is degree of freedom.

Subsequently, the development of an iterative robust coding (IRC) algorithm, referred to as TUMIRC, takes place, which possesses the unique characteristic of integrating error detection and error correction in a cooperative manner. The proposed algorithm significantly improves sparse representation by exhibiting exceptional robustness.

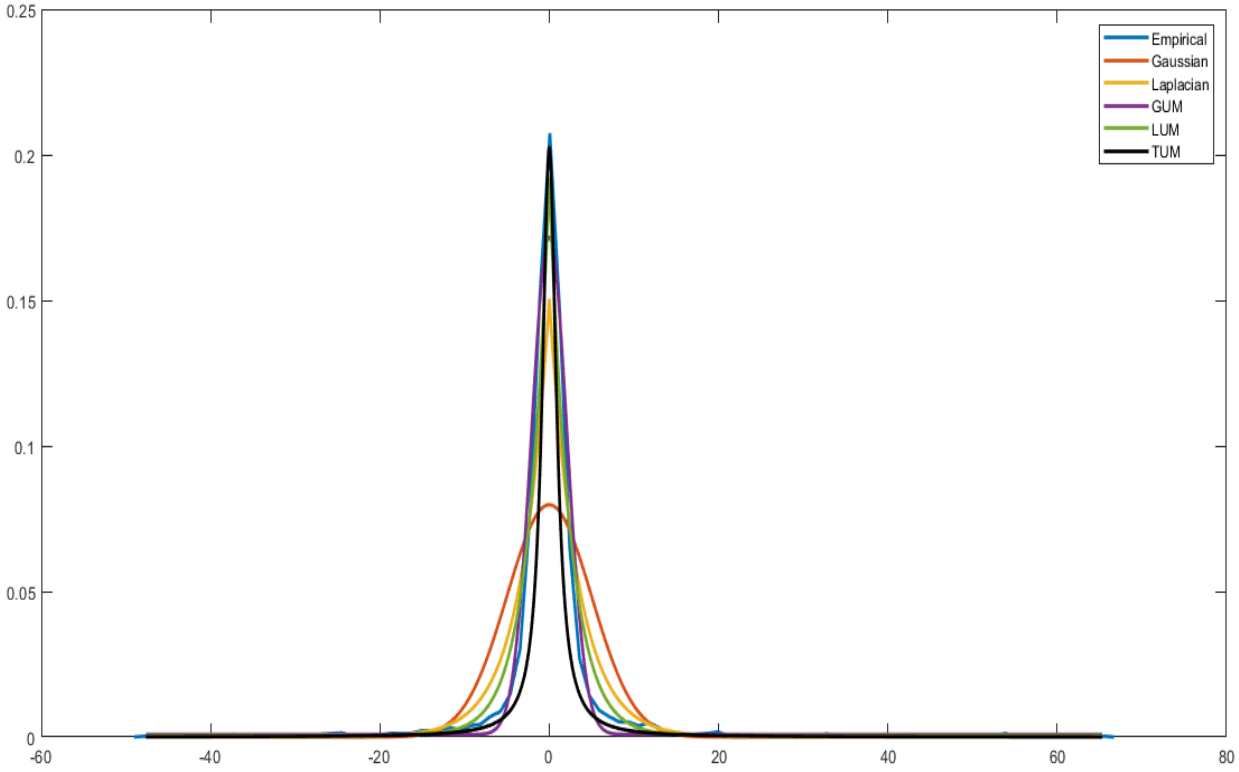


Fig. 6.1(a) Empirical and fit distributions of sparse representation errors for the test image without occlusion

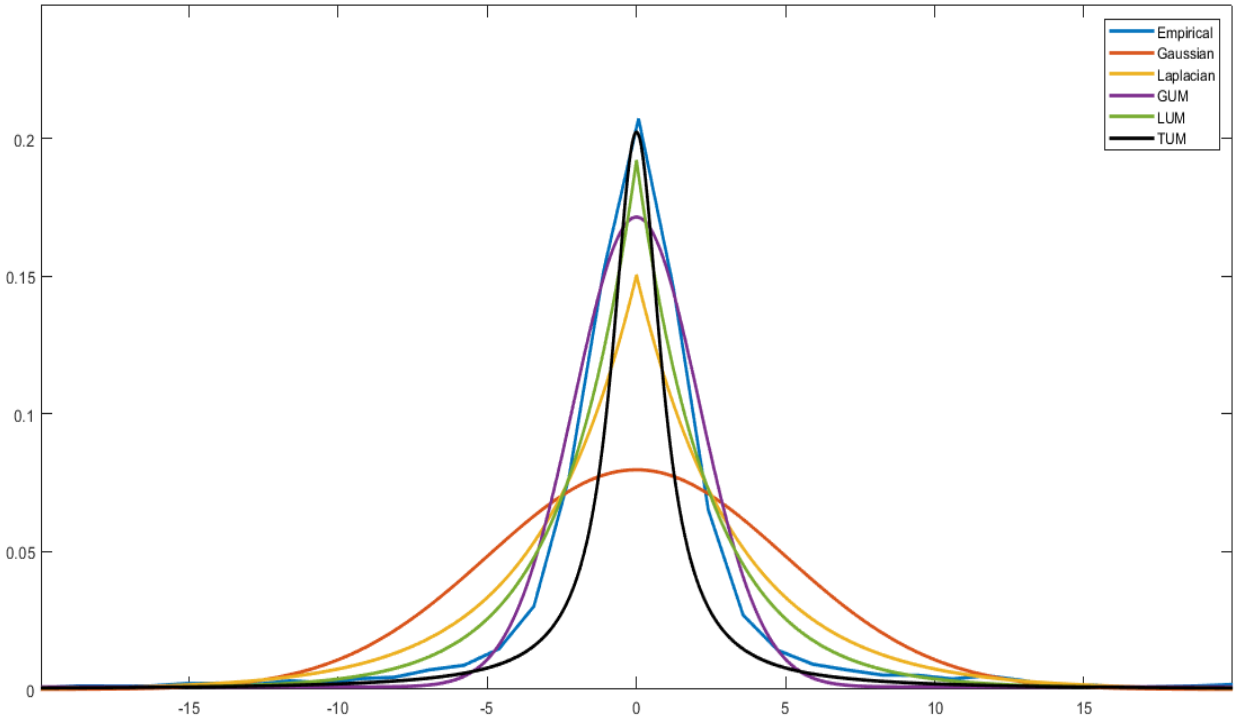


Fig. 6.1(b) Magnified view of Fig. 6.1(a) near the peak of Empirical and fit distributions of sparse representation errors for the test image without occlusion

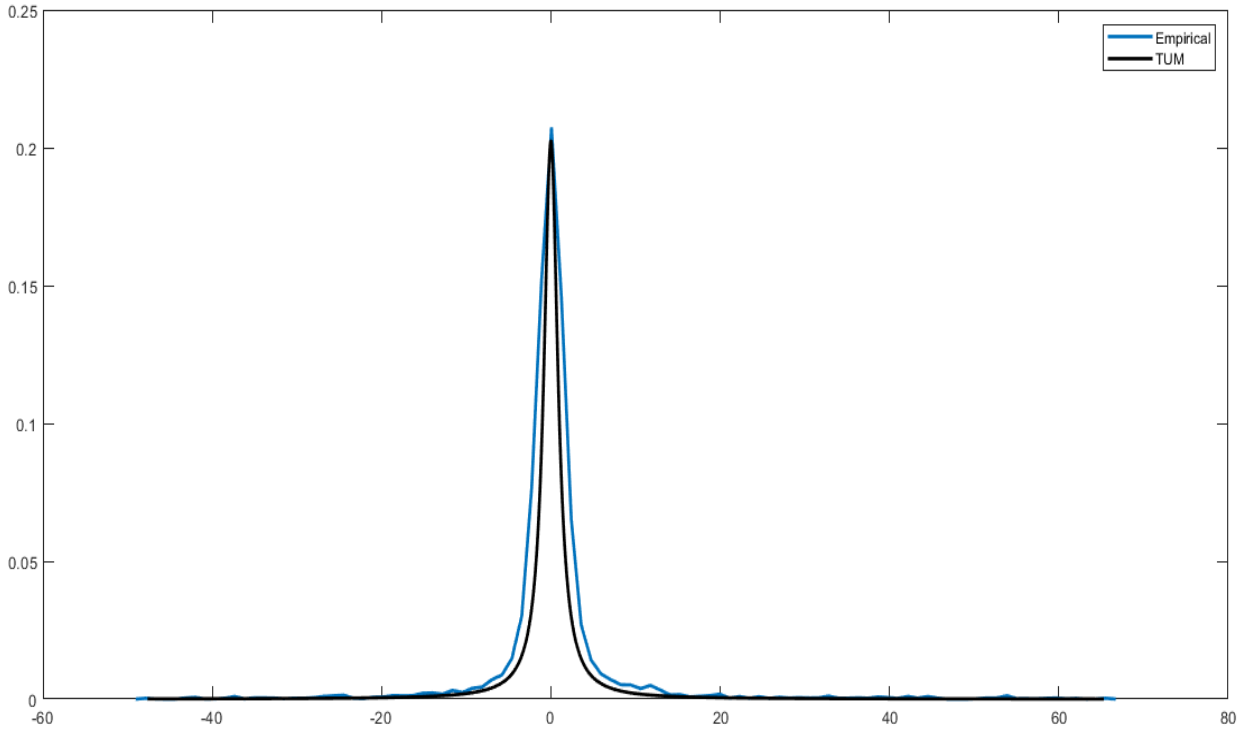


Fig. 6.1(c) Peaks of the Empirical and TUM (Best Fitted) based probabilistic distribution of residual errors for test image without occlusion

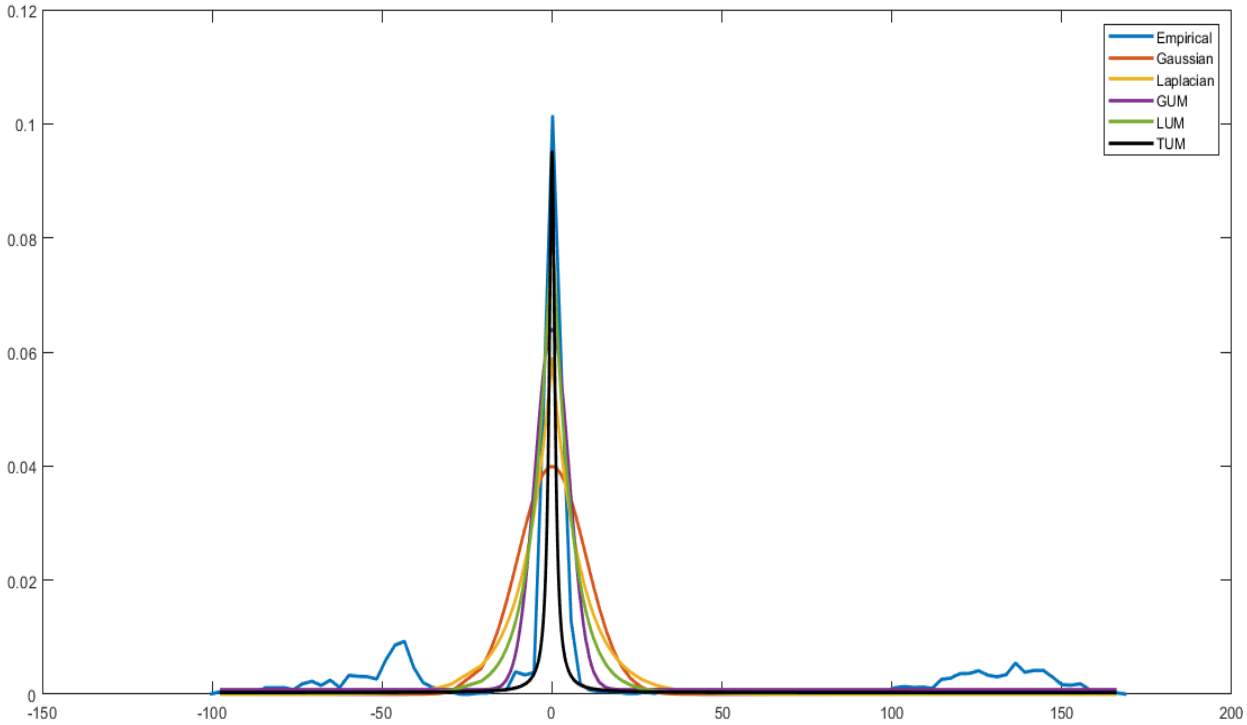


Fig. 6.1(d) Empirical and fit distributions of sparse representation errors for the test image with 30% block occlusion (Level-1)

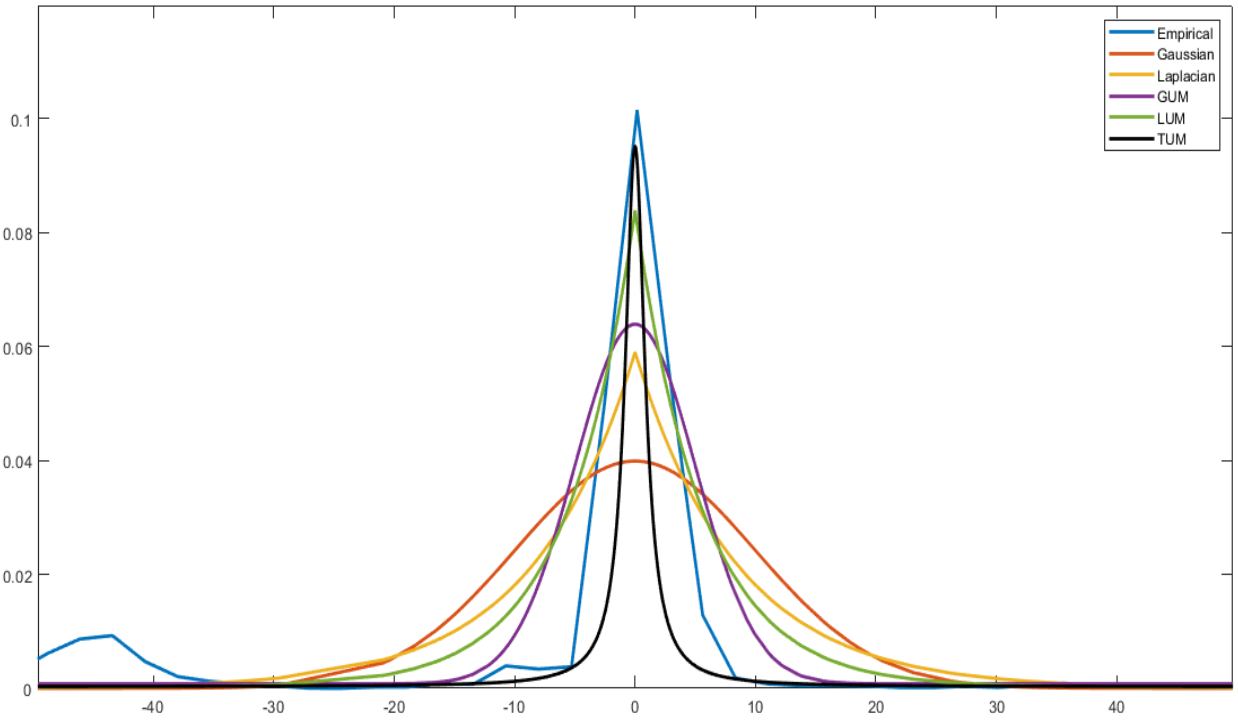


Fig. 6.1(e) Magnified view of Fig. 6.1(d) near the peak of Empirical and fit distributions of sparse representation errors for the test image with approximately 30% block occlusion (Level-1)

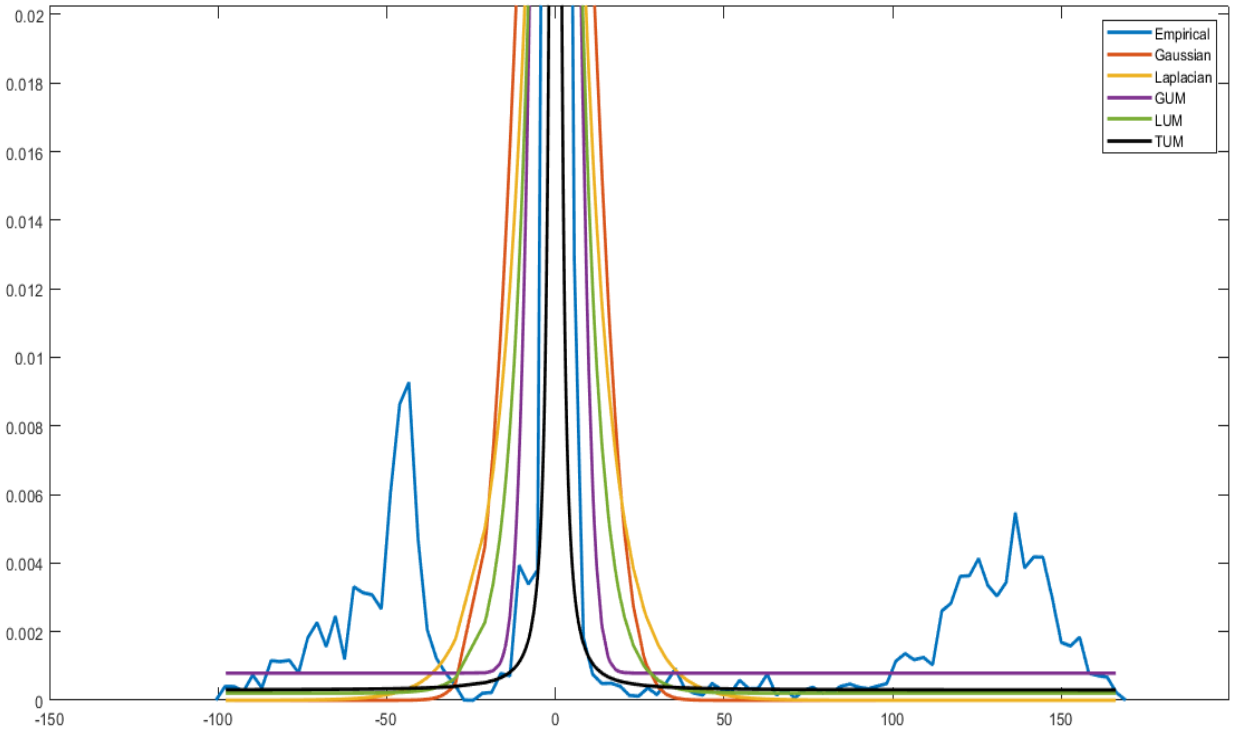


Fig. 6.1(f) Magnified view of Fig. 6.1(d) near the tails of Empirical and fit distributions of sparse representation errors for the test image with approximately 30% block occlusion (Level-1)

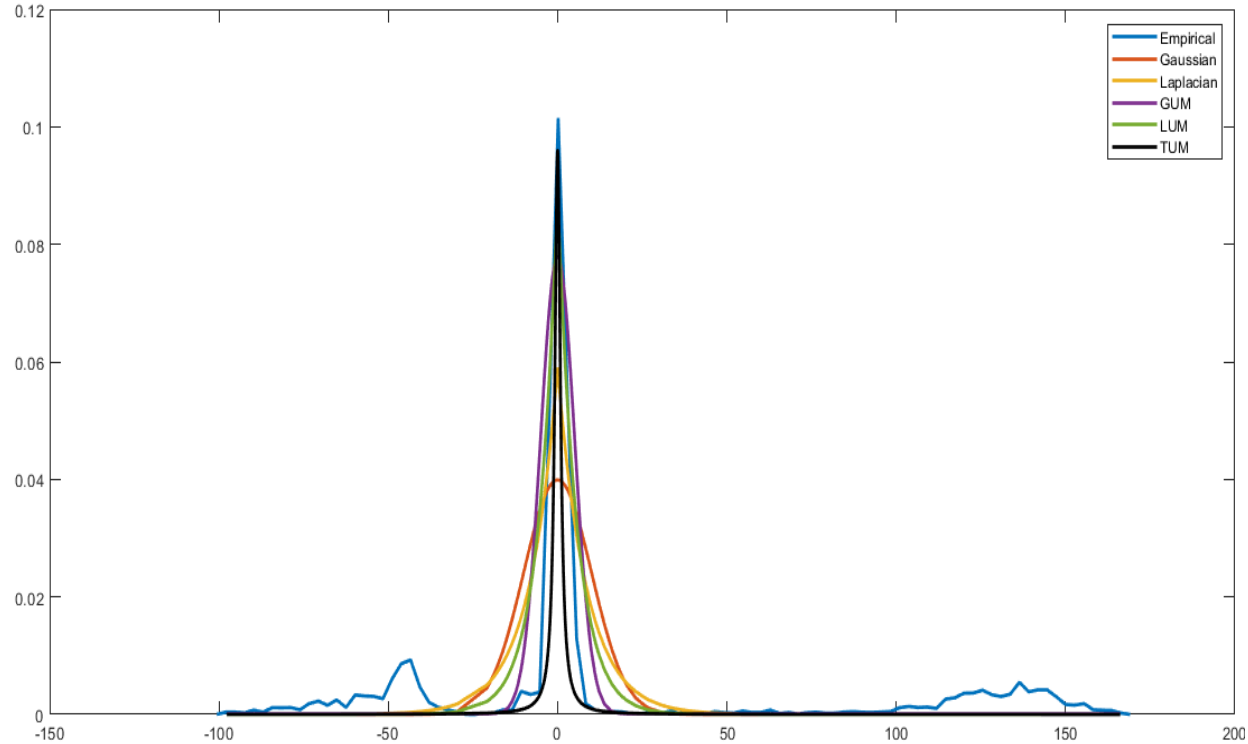


Fig. 6.1(g) Empirical and fit distributions of sparse representation errors for the test image with 40% block occlusion (Level-2)

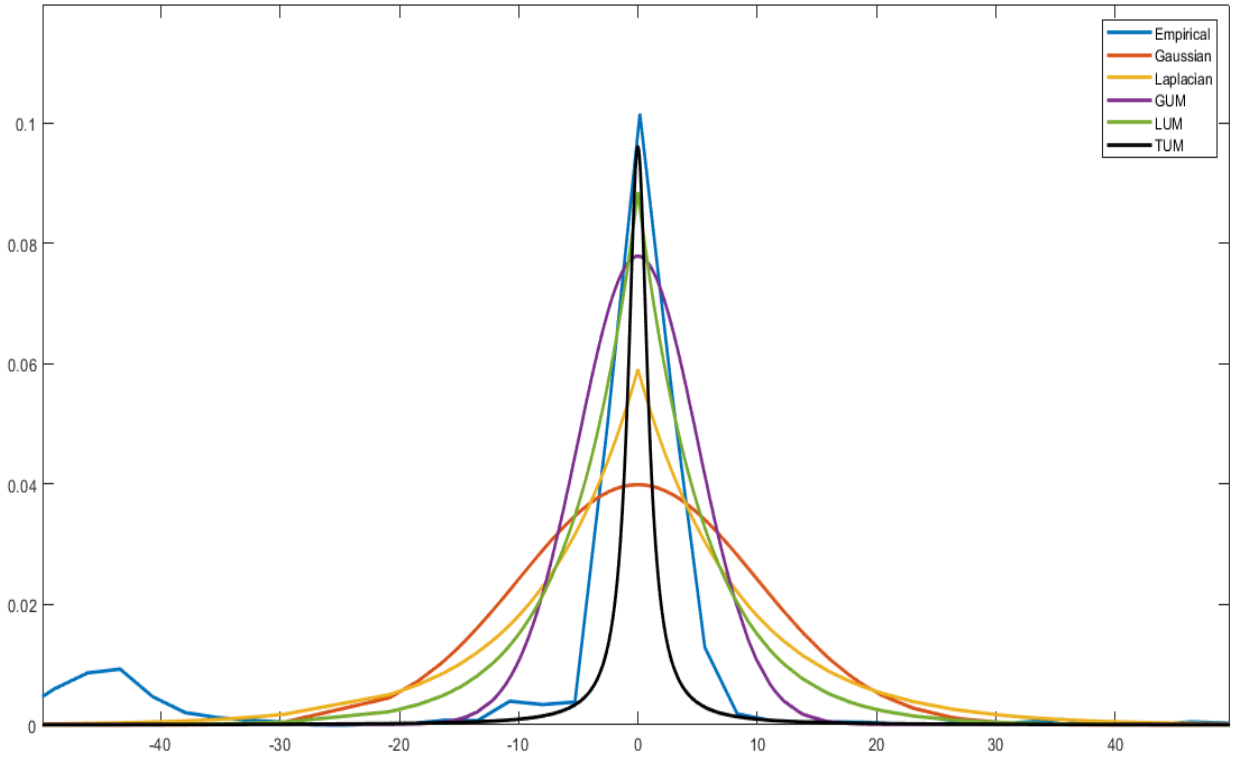


Fig. 6.1(h) Magnified view of Fig. 6.1(g) near the peak of Empirical and fit distributions of sparse representation errors for the test image with approximately 40% block occlusion (Level-2)

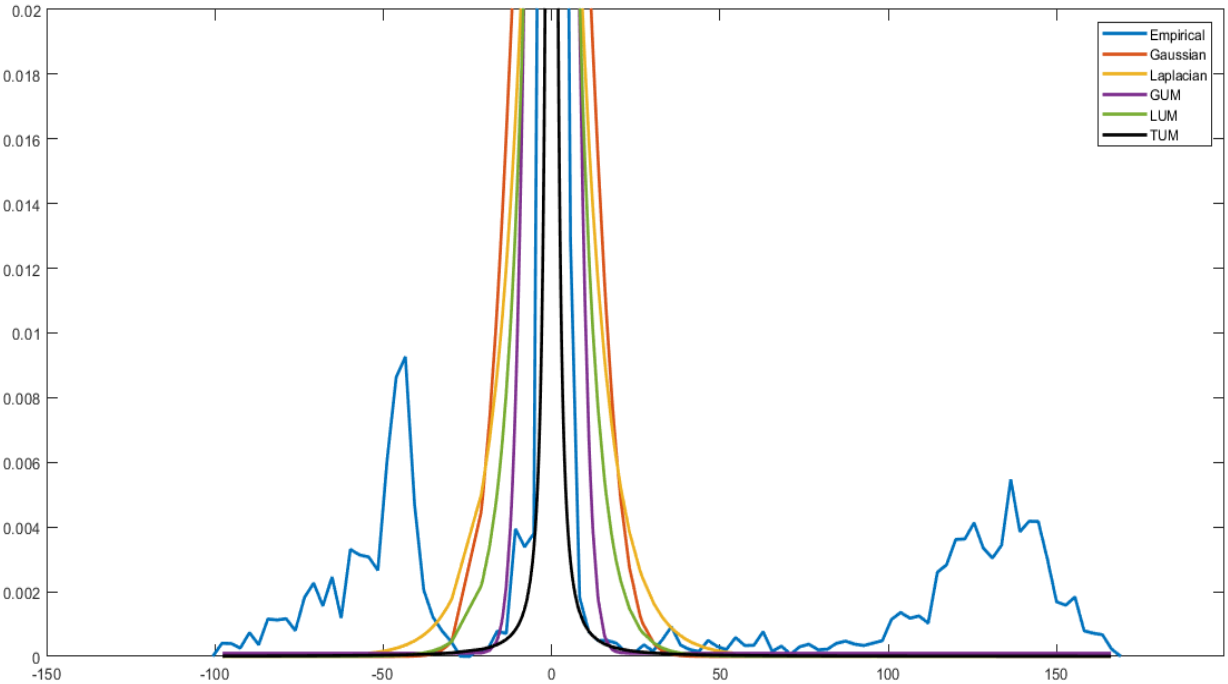


Fig. 6.1(i) Magnified view of Fig. 6.1(g) near the tails of Empirical and fit distributions of sparse representation errors for the test image with approximately 40% block occlusion (Level-2)

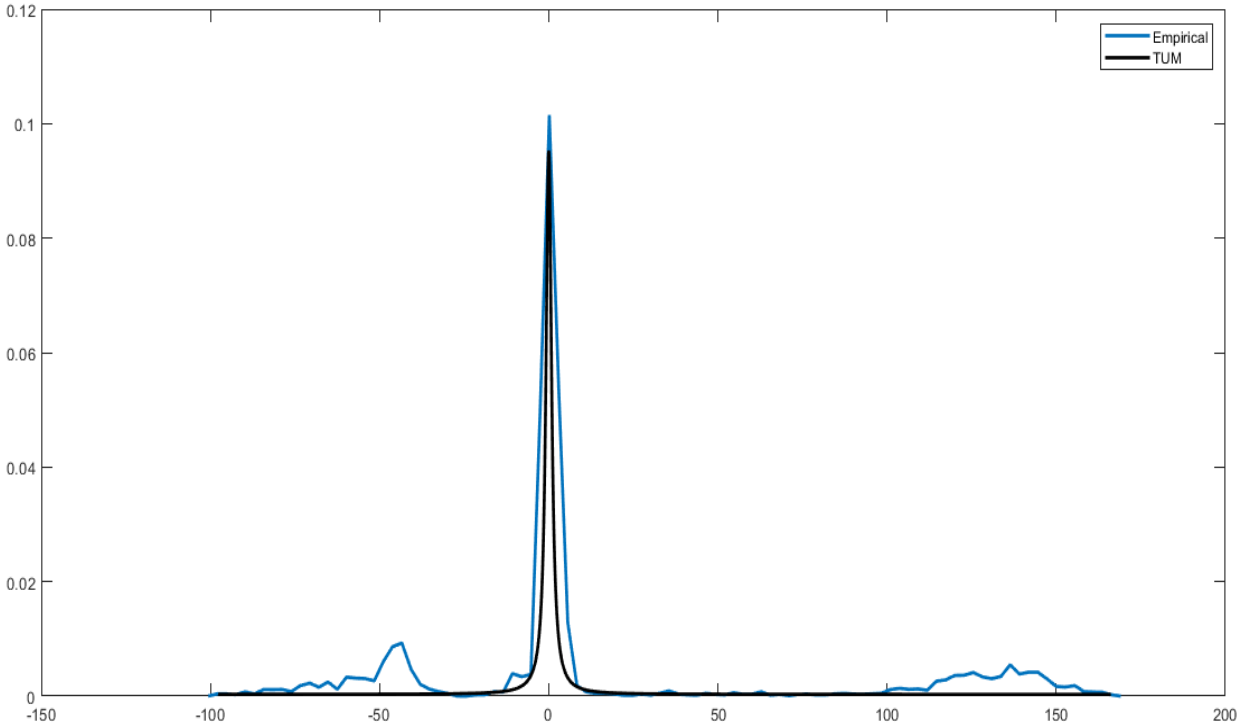


Fig. 6.1(j) Empirical and TUM (Best Fitted) based probabilistic distribution of residual errors for test image with approximately 30% block occlusion (Level-1)

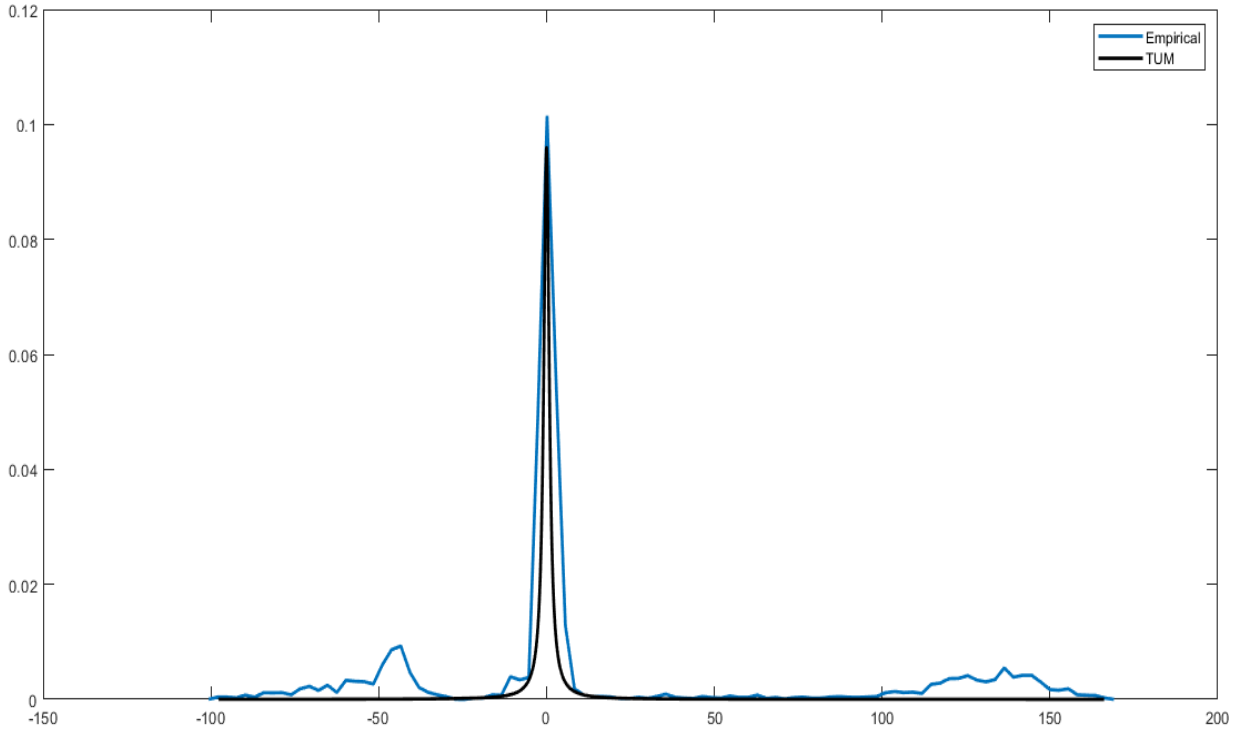


Fig. 6.1(k) Empirical and TUM (Best Fitted) based probabilistic distribution of residual errors for test image with approximately 40% block occlusion (Level-2)

Fig. 6.1. Empirical and fit distributions of sparse representation errors for a sample representative thermal hand sign language image with and without occlusion.

Table 6.1: The log likelihood of fit models without any occlusion

Model	***Gaussian***	***Laplacian***	***GUM***	***LUM***	***TUM***
Log-likelihood	20461.2	22500.1	22794.2	22930.4	23026.5

Table 6.2: The log likelihood of fit models with 30% block occlusion (Level-1)

Model	***Gaussian***	***Laplacian***	***GUM***	***LUM***	***TUM***
Log-likelihood	14177.1	15000.1	15508.3	16094.4	17269.0

Table 6.3: The log likelihood of fit models with 40% block occlusion (Level-2)

Model	***Gaussian***	***Laplacian***	***GUM***	***LUM***	***TUM***
Log-likelihood	13545.2	13750.4	14141.2	14701.1	15537.2

The log likelihoods of fit models without occlusion are shown in Table 6.1 and 30% and 40% block occlusion are shown in Table 6.2 and Table 6.3 respectively

In order to assess the advantages of mixture modeling, an example is considered in Fig. 6.1(a), where the empirical distributions of errors in sparse representation are depicted. The capability of fitting these errors is evaluated using the Gaussian, previously mentioned LUM, and proposed TUM models. To ensure accuracy and gain a comprehensive understanding of how the distributions of coding residuals change, the sparse coding coefficients are determined using the l_1-ls algorithm [142] and held fixed, following the approach employed in [136]. It is evident that the LUM model possesses an additional benefit in effectively modeling various coding errors, leading to improved log-likelihoods.

6.3. STUDENT'S T UNIFORM MIXTURE FOR ERROR MODELING

However, as revealed in previous research [119], it has been indicated that Gaussian distributions may not be highly effective as the inlier model for errors in image matching. This

observation is supported by the findings presented in Fig. 6.1 and Table 6.1, where a higher likelihood is obtained by the Laplacian distribution when modeling the inlier distribution compared to the Gaussian distribution. In the previous chapter, a mixture model named the Laplacian Uniform Mixture (LUM) model was introduced, incorporating the Laplacian distribution, to enhance robustness compared to previous methods. In Fig. 6.1(b), it is demonstrated that the accuracy of the TUM graph (black line) coincides more effectively with the peak of the empirical distribution (blue line) than the GUM (Purple), LUM (Green line), Laplacian (Yellow) And Gaussian (red line) models. It is hypothesized that the introduction of a mixture model with the Student's t distribution as the inlier model can lead to improved robustness compared to the previous LUM method. Considering the unpredictability of outliers, a uniform distribution is considered as the corresponding outlier model, similar to the approach employed in [120]. Hence, the TUM model is proposed as follows:

$$f(e_j) = \alpha(tpdf) + c \tag{2a}$$

$$f(e_j) = [\alpha k((1+\frac{e_j^2}{v})^{-\frac{v+1}{2}})] + c \tag{2b}$$

The best fitting performance is achieved by the proposed mixture model, TUM, as demonstrated in Fig. 6.1(c). A comparison of the total log-likelihoods of the coding residuals estimated with LUM is presented in the bottom tables. When considering the logarithmic scale, the improvement brought about by the proposed model is notably significant compared to other models, highlighting its robustness against outliers. For instance, in Fig. 6.1 Table 6.1, the log-likelihood of TUM exhibits an increment of 96.1 compared to that of LUM, corresponding to a likelihood multiplication of 5.0×10^{73} due to the logarithmic scale. From a statistical perspective, a model that better fits the distribution of coding errors, while possessing similar capacities, results in a more accurate recovery of the underlying sparse coefficients and a more reliable reconstruction of the clean sign language thermal image. This faithful sparse representation inherently offers discriminative properties and aids in revealing the true identity of the input pattern. It is important to note that none of the existing models accurately capture the side portions of the error distributions depicted in Fig. 6.1(f) and 6.1(i), which is considered reasonable as these regions correspond to outliers that are assumed to be unpredictable in the given problem. In Fig. 6.1(e) and Fig. 6.1(h) it is observed that the peaks of Empirical and fit distributions of sparse representation errors for the test image with 30% (Level-1) and 40% (Level-2) block occlusion respectively. The

proposed TUM model effectively models the distribution of residuals around zero, reducing the risk of false error detection. These residuals mainly arise from minor perturbations rather than outliers and correspond to pixels that play a significant role in identity recognition. In our optimization algorithm, pixels with large residuals are deemed less reliable and assigned small weights, eventually being excluded from image matching. Compared to other error models, the proposed TUM model demonstrates the best fit for challenging residual distributions, as shown in Fig. 6.1(j) and Fig. 6.1(k).

It is believed that the coding residuals are i.i.d. (independent and identically distributed) due to the random distribution of outliers throughout the image. The negative log-likelihood function $-\ln f(e_j)$ and sparsity constraints can be used create the objective function or equivalently as follows:-

$$x^* = \arg\min_x \sum_{j=1}^{m} \rho(e_j) + \lambda_0 \| x \|_1 \tag{3}$$

$$s.t.\ \ e = y - Ax$$

$$\text{Where, } \rho(e_j) = -\ln[\alpha k(1+\frac{e_j^2}{v})^{-\frac{v+1}{2}} + c] \tag{4}$$

The representation is made more faithful and sparse, respectively, by the 1st and 2nd term in the objective function (3).

$\lambda_0 > 0$ Is a balancing parameter.

6.4. WEIGHT FUNCTION MODELING

The objective function (3) is non-differentiable at $e_j = 0$ and nonconvex, which is hard to be optimized directly. Using the traditional iteratively reweighted least-squares formulation, $\omega_j^t = \rho'(\mathrm{e}_j^t)/\mathrm{e}_j^t$, is not suitable for the first term, despite its resemblance to a typical M-estimator. One reason is that when the denominator, e_j^t approaches zero during the optimization process, ω_j^t can become arbitrarily large. This occurs because the limit as j approaches zero of $\left|\rho'(\mathrm{e}_j)\right| = (1/b(1+c)) > 0$ for $\rho(e_j^t)$ defined in equation (4). Consequently, the iterative procedure becomes unstable.

The expansion of (4) is performed by utilizing the Taylor expression up to the second order at the

current estimation of e_j.

$$\rho(e_j) = \rho(|e_j^t|) + \rho'(|e_j^t|)(|e_j| - |e_j^t|) + \frac{1}{2}\rho''(|e_j^t|)(|e_j| - |e_j^t|)^2 + o(|e_j|) \quad (5)$$

Where, $o(|e_j|)$ denotes the higher order terms and,

1st derivative of equation (4) is shown below,

$$\rho'(|e_j^t|) = -\frac{\alpha k(-\frac{\upsilon+1}{2})(1+\frac{e_j^2}{\upsilon})^{-\frac{\upsilon+1}{2}-1}(\frac{2e^j}{\upsilon})}{[\alpha k(1+\frac{e_j^2}{\upsilon})^{-\frac{\upsilon+1}{2}} + c]}$$

$$= (\frac{\upsilon+1}{\upsilon})\frac{\alpha k(1+\frac{e_j^2}{\upsilon})^{-\frac{\upsilon+1}{2}}}{[\alpha k(1+\frac{e_j^2}{\upsilon})^{-\frac{\upsilon+1}{2}} + c]} \times \frac{e_j}{(1+\frac{e_j^2}{\upsilon})}$$

$$\rho'(|e_j^t|) = (1+\frac{1}{\upsilon})\frac{\alpha tpdf}{\alpha tpdf + c} \times \frac{e_j}{(1+\frac{e_j^2}{\upsilon})} \quad (6)$$

Denote the 2nd order approximation of $\rho(e_j)$ by $\hat{\rho}_2(e_j)$ because $\rho(e_j)$ is an increasing function of $|e_j|$ and $\rho''(|e_j^t|) < 0$, $\hat{\rho}_2(e_j)$ higher order approximations of $\rho(e_j)$ are nonconvex and hard to be optimized.

By replacing the higher order terms with a coefficient $\omega(e_j^t) > 0$, to make the 2nd order approximation a quadratic convex function of e_j with the 1st order derivative at the origin $\hat{\rho}_2'(0) = 0$, it may be possible to relax the constraint on the coefficient of the 2nd order term, similar to what has been done in [136].

This strategy will lead to the equation $\rho'(|e_j^t|) = \omega(e_j^t)|e_j^t|$, faces the aforementioned issue/instability issue and is equivalent to the standard M- estimator in use.

$\rho(e_j)$ is concave in $|e_j|$. By simply maintaining the first order term of its Taylor expression, it is proposed to approximate $\rho(e_j)$ in (6).

$$\hat{\rho}_1^t(e_j) = \rho(|e_j^t|) + \rho'(|e_j^t|)(|e_j| - |e_j^t|) \quad (7)$$

Substitution, (7) into (3) and removing terms unrelated to the argument to be optimized, we have

$$x^{t+1} = \arg\min \| W^t e \|_1 + \lambda \| x \|_1 \quad (8)$$

$$s.t.\ \ e = y - Ax$$

Here W^t is a diagonal matrix with the diagonal element.

$$\omega_j^t = (1+\frac{1}{\upsilon})\frac{\alpha tpdf}{\alpha tpdf + c} \times \frac{e_j}{(1+\frac{e_j^2}{\upsilon})} \propto \rho'(|e_j^t|) \quad (9)$$

$e^t = y - Ax^t$, and λ is a parameter that absorbs λ_0 and b. For each iteration on t, the optimization method described in (8) resolves a reweighted $l_1 - l_1$ minimization issue. The weight matrix W^t serves as an error detector by giving small values to outliers, (with large error $|e_j^t|$). Moreover, it is recognized that $l_1 - l_1$ minimization possesses robustness against gross errors and provides an error correction mechanism. This mechanism becomes particularly significant in situations where the detection of gross errors by W^t is challenging, especially during the initial stage well before the optimization of W^t.

6.5. DISCUSSION ON THE WEIGHT FUNCTION

The mixture distribution $f(e_j)$ defined in (2b) implies for foreground pixels a Student's t distribution $f(e_j | \upsilon_j = 1) \propto (1+\frac{e_j^2}{v})^{-\frac{v+1}{2}}$, where υ_j denotes the label of the jth pixel, 1 for foreground and 0 for the outlier. Consequently, given e_j, the posterior probability of the present pixel being in the foreground is

$$P(\upsilon_j = 1 | e_j) = \frac{f(e_j | \upsilon_j = 1)P(\upsilon_j = 1)}{f(e_j)} \propto \omega_j^t \quad (10)$$

This indicates that the weight function obtained in TUM iterative robust coding is statistically valid. In other words, TUMIRC assigns weights to the pixels based on the posterior probabilities of the pixels being foreground, taking into account the errors as evidence in the TUM error modeling approach.

The iterative robust coding algorithm has been previously addressed in *Chapter 5* and is also discussed in [116]. The algorithm of TUM is same as well as LUM. The overall algorithm iterates over (8) until convergence, which can be verified by $(\| x^{t+1} - x^t \|_2 / \| x^t \|_2) < \in$, where $\in$ is a small

positive value. To classify the test image y, from

$$c^{t+1} = \arg\min_{c} \| c \|_1, \quad s.t.\ Bc = W^t y \tag{11}$$

Where $c = [W^t e; \lambda x]$ and $B = [I, \frac{1}{\lambda} W^t A]$, that the sparse solution should satisfy $W^t y - e^{\sim} = W^t A x$.

Therefore, the identity of y can be determined by the best reconstruction that can be obtained from the components in x corresponding to a single class as follows:

$$ID(y) = \arg\min_i \| W^*(y - A\delta_i(x^*)) - e^{\sim *} \|_2^2 \tag{12}$$

Here the superscript $*$ indicates the values at convergence, while $\delta_i(x^*) \in \mathbb{R}^n$ is a vector whose nonzero components are only the entries of x^* corresponding to the i th class. The overall algorithm, denoted by TUMIRC.

6.6. EXPERIMENTAL RESULTS

The database comprises images captured under challenging illumination and temperature conditions, and the experiments are conducted using three different sets:

(i) Sign language recognition with thermal images in low-light and night vision conditions,

(ii) Sign language recognition with thermal images in low-light and night vision conditions, including occlusions.

The entire database is divided randomly into training and testing sets with a 7:3 ratio, and the dictionary A is derived from the training dataset.

The findings of this chapter demonstrate that TUM achieves higher accuracy in sign language recognition compared to LUM, RRC, and RSC for images captured under illumination and temperature constraints. Furthermore, as the level of occlusion, especially due to occlusions, increases in the images, TUM outperforms LUM in terms of performance. These results indicate the superior capability of TUM in handling challenging scenarios involving occlusions.

6.6.1. Using the Student's t Uniform Mixture Model

The sign language hand gesture database specifically created for this study, as detailed in *Chapter 4*, provides the training set and test images for this example. Notably, thermal images are utilized for constructing this database.

To ensure accuracy and gain valuable insights into the changes in coding residuals' distributions in the presence of various outliers, the sparse coding coefficients are calculated using

the -ls algorithm [142]. These coefficients remain fixed throughout the analysis, following the methodology employed in [136].

The comparison of total log-likelihoods of coding residuals estimated with previous models is presented in the table to assess the goodness of fit of density functions.

According to previous research, it has been found that Gaussian distributions may not be highly effective in accurately modeling errors encountered in image matching. This observation is consistent with the results presented in Table 6.1, where the Laplacian distribution demonstrated superior likelihood in modeling the inlier distribution compared to the Gaussian distribution. Building upon this knowledge, the introduction of a mixture model that incorporates the Laplacian distribution as the inlier model shows potential for achieving enhanced robustness compared to previous methods. To accommodate unpredictable outliers, a new method known as TUM (Student's t uniform mixture) has been developed, which incorporates a uniform distribution as the outlier model. This approach is similar to the methodology employed in related studies [120]. TUM distinguishes itself as the preferred choice due to its effective handling of outliers and its ability to achieve improved overall performance.

6.6.2. Parameter Tuning

The parameters v, c, and α (Alpha) can be changed in the LUM model to fit the distribution more precisely. The outlier model is represented by the parameter c, which accepts values greater than 0 to indicate a uniform distribution. The TUM model can accommodate a range of extreme data points with c varying by 0.00001, 0.001, 0.01, 0.1, and 1. Different levels of outliers can be modelled with c fluctuating between these values.

The degree of freedom is represented by the parameter v. The Laplacian distribution's spread or scale can be changed by changing v by 1, 3, 5, 7, or 10. This parameter has a significant impact on the error distribution's width and is essential for representing the variability of the inlier data.

The parameter α, which can vary by 0.01, 0.1, 1, 2, or 10, allows the TUM model to adjust to various degrees of sparsity in the coding coefficients. This parameter enables the adjustment of the trade-off between increasing sparsity in the coding process and accurately portraying the data. The TUM model can be adjusted to meet the particular distribution properties of the data by experimenting with various combinations of v, c, and α within their designated limits. This

adaptability allows the model to accurately capture the underlying structures and patterns, improving performance across a range of tasks like image analysis and hand gesture detection.

The prior discussion looked at the changes to the TUM model's parameters v, c, and α. To determine their effect on the performance of the model, these parameters were adjusted throughout a range of values in a methodical exploration.

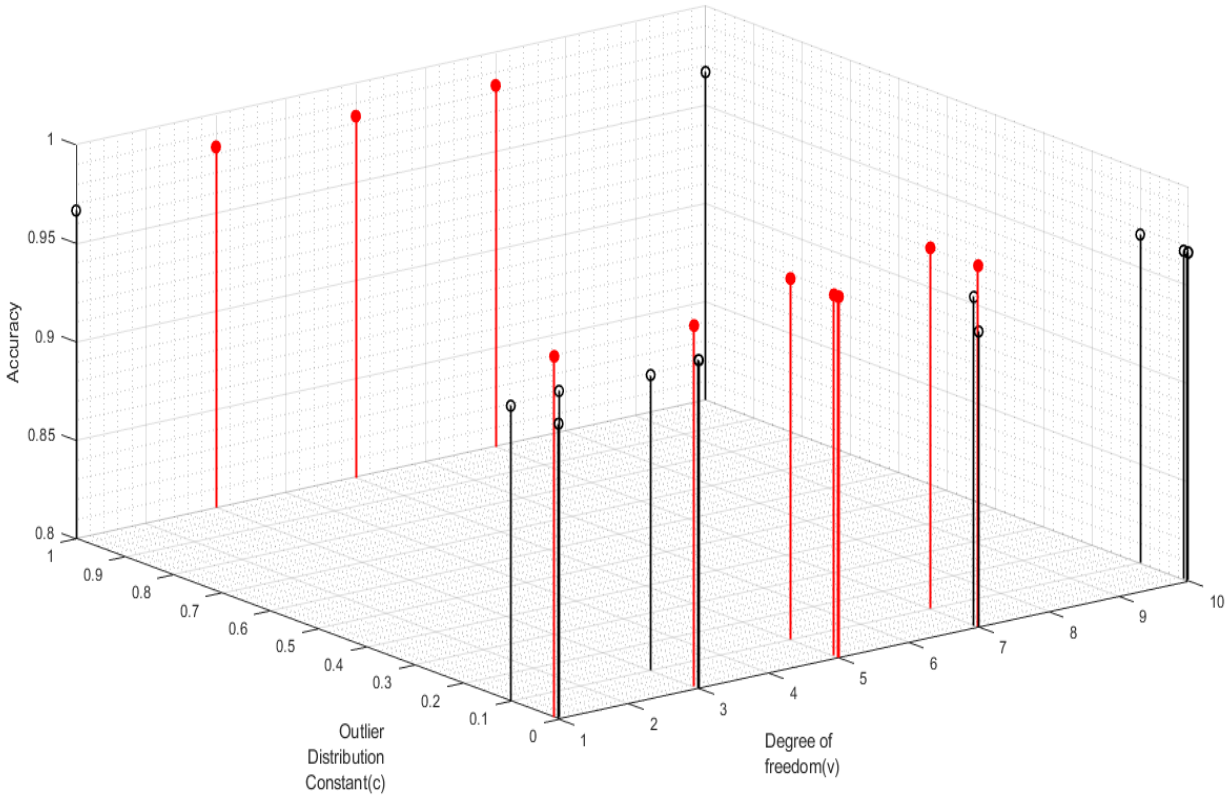
6.6.3. Results

Table 6.5 displays the findings from the application of the TUM model to the Non-Occluded database. There were 125 different combinations of the parameters v, c, and α that were evaluated using the dataset. In below Fig. 6.2 (a), (b), (c), (d) and (e) show a graphical representation of the accuracy changes brought on by changes in the parameters v, c, and α. These table 6.5 offer illustrative insights into the correlation between parameter values and associated TUM model performance.

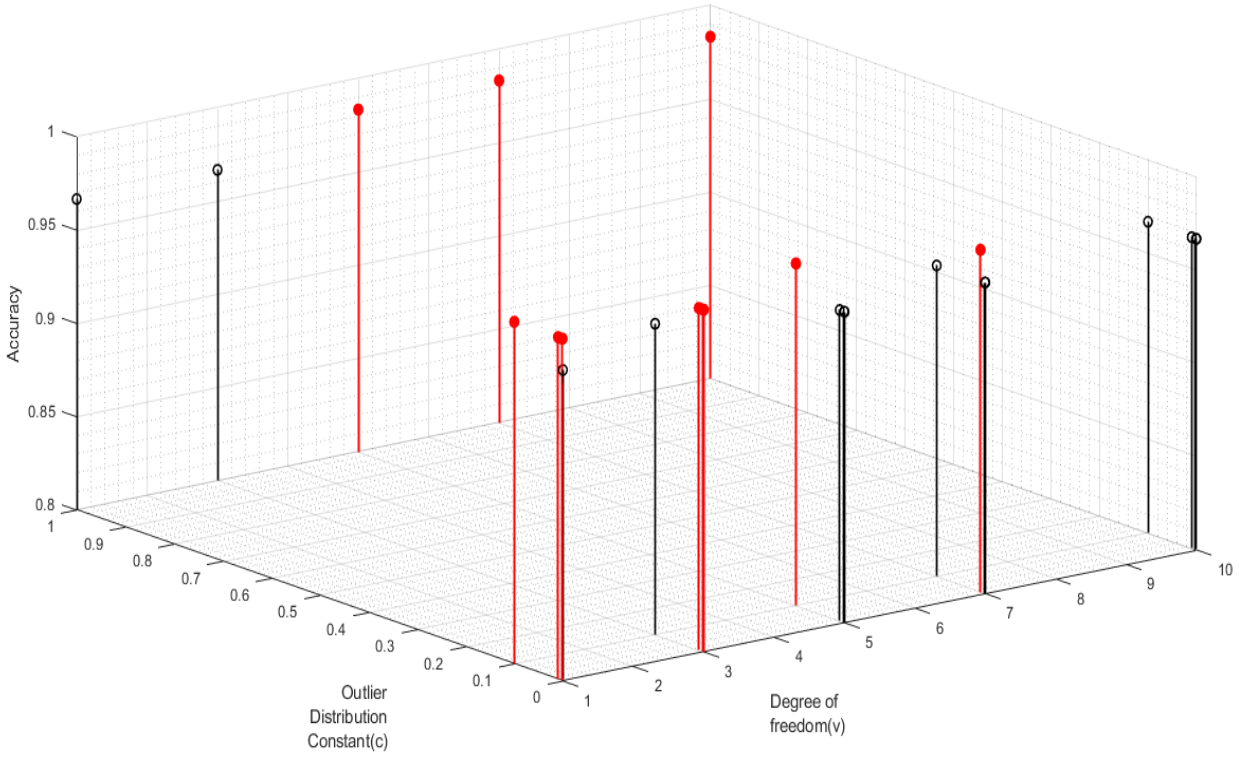
Table 6.4 represents the recognition rate range mapping table, which serves to comprehend the graphical representation of the TUM model's performance accuracy.

Table 6.4: Recognition Rate Range Mapping with Color for TUM

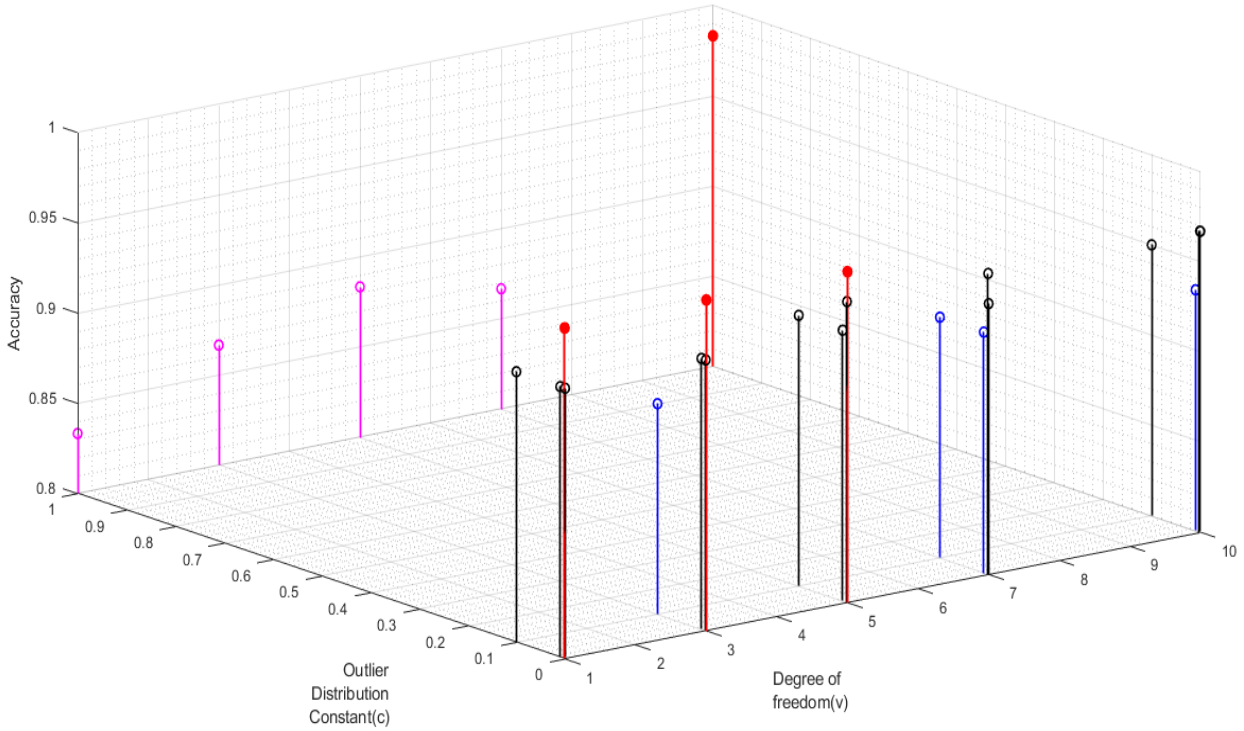
Recognition Rate Range	***Color***
Below 0.90	Magenta
0.90 to 0.95	Blue
0.95 to 0.97	Black
0.97 to 0.98	Green
Above 0.98	Red



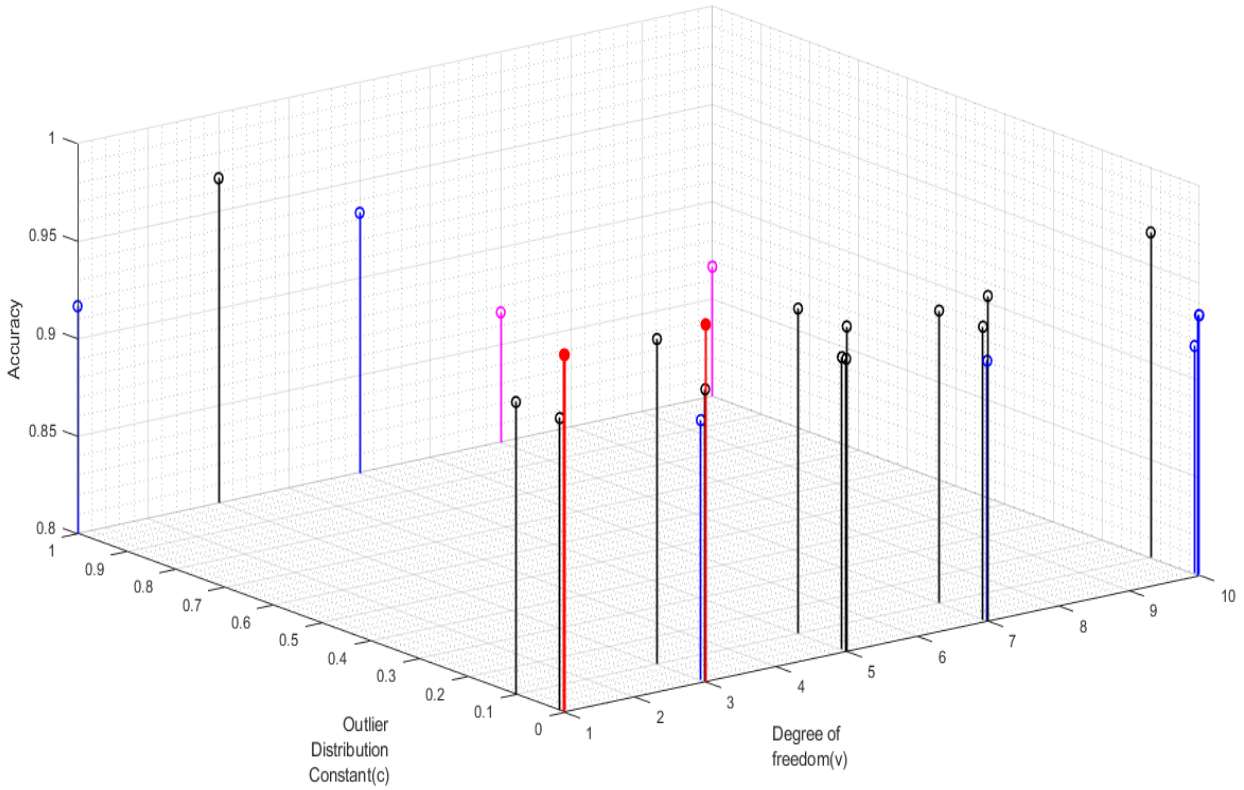
(a)- (Alpha: 0.01)



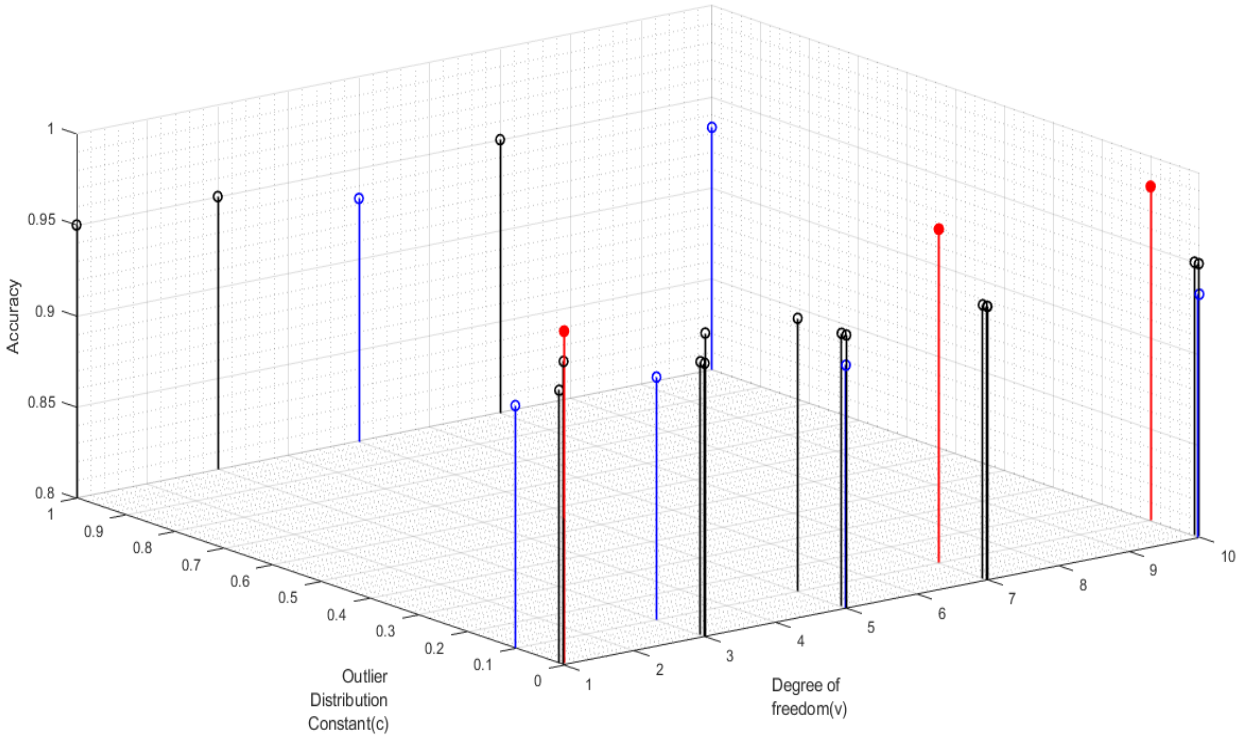
(b)- (Alpha: 0.1)



(c)- (Alpha: 1)



(d)- (Alpha: 2)



(e)- (Alpha: 10)

Fig. 6.2. The graphical representations of the accuracy changes using TUM

After the Graphical representation was analyzed, these results were obtained.

Table 6.5: Performance of TUM with Non-Occluded Sign Language Thermal Image Datasets

Maximum Accuracy (in %)	***98.33%***
Minimum Accuracy (in %)	***83.33%***
Standard Deviation (in %)	***2.67%***
Mean (in %)	***95.76%***

This strategic decision aimed to identify specific combinations of v, c, and α that yielded significantly higher recognition rates compared to the previously used models. The selected parameter combinations were then applied to the dataset to achieve the maximum recognition rate. The performance of the TUM model with Non-Occluded Sign Language Thermal Images on the dataset of 2000 images is presented in Table 6.5.

Table 5.3 in *chapter* 5 demonstrates that the maximum accuracy achieved by the LUM model is comparatively lower than the maximum accuracy attained by the TUM model. In the case of larger datasets, it was observed that the recognition rate showed improvement when using the TUM model compared to the LUM model.

In Table 6.6, the comparison of the RRC, LUM and TUM models and the improvement of the TUM model using the same datasets for both are demonstrated. It can be observed that the TUM model exhibits improved performance compared to the RRC and LUM models.

Table 6.6: Comparison of RRC, LUM and TUM Model with Non-Occluded Sign Language Thermal Image Datasets

	RRC MODEL	**LUM MODEL**	**TUM MODEL**
Recognition Rate	0.9167	0.9667	0.9833
Accuracy (in %)	91.67%	96.67%	**98.33%**

6.6.4. For the Occluded Datasets

In *chapter* 4, the discussion focused on the occluded datasets. These datasets was created to assess the enhanced robustness of the TUM method in practical scenarios. The experiment involved the division of the occluded data into two stages: level 1 and level 2. Level-1 featured a block occlusion amounting to 30%, whereas Level-2 showcased a higher block occlusion of 40%. Combinations of parameters v, c, and α were selected based on their ability to achieve higher recognition rates. The performance of the TUM model with Occluded Sign Language Thermal Images for the dataset was evaluated and presented in Table 6.7 and 6.8 for level-1 and level-2, respectively.

Table 6.7: Performance of TUM for Level 1 occluded dataset

Recognition Rate	0.9667
Accuracy (in %)	96.67%

Table 6.8: Performance of TUM for Level 2 occluded dataset

Recognition Rate	0.9333
Accuracy (in %)	93.33%

In the occluded datasets, it is evident that the TUM model outperforms the previous LUM model. This improvement can be observed in Table 6.9, which presents a comparison between the RRC, TUM and LUM models for the level 1 and level 2 occluded datasets.

Table 6.9: Performance Comparison of RRC, LUM & TUM Model using occluded dataset Level- 1 and Level-2

	RRC MODEL		**LUM MODEL**		**TUM MODEL**	
	Level-1	**Level-2**	**Level-1**	**Level-2**	**Level-1**	**Level-2**
Recognition Rate	0.8833	0.8333	0.9167	0.8667	0.9667	0.9333
Accuracy (in %)	88.33%	83.33%	91.67 %	86.67 %	**96.67 %**	**93.33 %**

The analysis of the previous results reveals a notable improvement achieved by the TUM model compared to its predecessor. The utilization of the TUM model demonstrates enhanced performance and effectiveness in addressing the challenges encountered in the given context. This improvement is clearly evident when considering the observed outcomes and comparisons between the TUM model and the previous approaches like RRC, LUM. The significant advancements showcased by the TUM model further highlight its potential and efficacy in solving the specific problems at hand.

6.7. SUMMARY

In this chapter, a new concept was introduced for vision sensing-based thermal image sign language recognition in challenging environments, addressing a contemporary problem in collaborative robotics. The previous model, LUM (Laplacian-based Uniform Mixture) [116], served as a benchmark for comparison with the newly proposed model, TUM (Student's t Uniform Mixture). The LUM model, widely used in the field, provided a basis for evaluating the performance improvements achieved by the TUM model.

In addition to its successful performance in moderately challenging environments with poor illumination, the weight thresholded LUM models introduced in this chapter have opened up new possibilities for sign language recognition using thermal images. However, when faced with more demanding scenarios, the limitations of the LUM model become apparent. This is where the TUM model comes into play, surpassing the capabilities of the LUM model and offering improved

performance in handling various challenging situations.

The TUM model's ability to adapt to different levels of sparsity in coding coefficients, along with its incorporation of the Student's t distribution and uniform distribution, grants it a unique advantage in tackling more complex and challenging environments. By leveraging the flexibility and robustness of the TUM model, it becomes a valuable tool for addressing even more demanding scenarios, surpassing the limitations of the previous LUM model.

Through extensive experimentation and evaluation, it has become evident that the TUM model provides significant improvements over the LUM model, making it a promising choice for sign language recognition in challenging situations. Its superior performance in handling occlusions, extreme data points, and varying levels of corruption sets it apart as a reliable solution for advanced applications in collaborative robotics and other related fields.

CHAPTER-7

THERMAL IMAGING BASED VEIN PATTERN IDENTIFICATION USING LUM AND TUM ERROR MODELING

- ***INTRODUCTION***
- ***LUM AND TUM BASED VEIN PATTERN IDENTIFICATION***
- ***DATABASE CREATION***
- ***PREPROCESSING***
- ***EXPERIMENTAL RESULTS***
- ***SUMMARY***

7.1. INTRODUCTION

Vein pattern recognition using thermal imaging has emerged as a promising biometric authentication technique, offering advantages such as non-intrusiveness, high accuracy, and resistance to spoofing. This innovative approach involves capturing the unique thermal patterns emitted by veins beneath the skin's surface, which are influenced by the variation in blood flow and body temperature.

Several previous works have contributed to the development and advancement of this technology, laying the foundation for its current state. Liu *et al.* conducted a study in 2014, exploring the feasibility of vein recognition using thermal images [155]. They achieved high recognition rates by analyzing the thermal patterns of dorsal hand veins. Zhang *et al.* conducted a comprehensive review in 2017, providing insights into thermal infrared imaging techniques for vein pattern biometrics [156]. Their work guided the development of more robust vein recognition systems.

In 2019, Wang *et al.* introduced deep learning techniques to enhance vein recognition systems, achieving significant improvements in accuracy [157]. Huang *et al.* proposed a novel vein recognition framework in 2021 that utilized an infrared laser to enhance vein visibility, leading to improved recognition performance [158].

In the context of this thesis, vein pattern recognition of the human forearm using a thermal camera is being focused on. Experiments are conducted in the Measurement and Instrumentation Laboratory of the Electrical Department at Jadavpur University. The objective is to investigate and analyze vein patterns on the human forearm using thermal imaging. Infrared images of the forearm are captured using a thermal camera, revealing the unique thermal patterns emitted by the veins. Advanced algorithms and techniques are employed for vein extraction, segmentation, and feature extraction. The collected dataset is used to train and fine-tune models for accurate vein pattern recognition. Efforts are made to enhance the system's accuracy and efficiency by refining algorithms and incorporating machine learning techniques. Challenges such as image noise, variability in vein appearance, and occlusions are actively addressed. The controlled environment of the Measurement and Instrumentation Lab ensures consistency and accuracy in data collection.

7.2. LUM AND TUM BASED VEIN PATTERN IDENTIFICATION

In vein pattern recognition for computational techniques, both the previously described LUM (Laplacian Uniform Mixture) and the proposed TUM (Student's t Uniform Mixture) error models are utilized. A detailed discussion on the LUM model and the TUM model have already been conducted in the previous *chapter 5* and *chapter 6* respectively. These models have been thoroughly explored, and their algorithms have been implemented for this problem at hand in a comprehensive manner.

The LUM technique, previously described in this thesis and employed for sign language recognition, has not been explored in vein pattern recognition. The LUM model has shown promising results for authentication purposes. On the other hand, the proposed TUM technique aims to enhance recognition accuracy by leveraging thermal uniform mapping. This technique involves analyzing the thermal patterns emitted by veins on the forearm and creating a mapping that preserves the uniformity of these patterns. By reducing noise and enhancing the discriminative power of the extracted features, the TUM model improves vein pattern recognition performance.

The utilization of both the LUM and the proposed TUM models in vein pattern recognition highlights the ongoing efforts to advance the field. By integrating these techniques, researchers can explore the strengths and limitations of existing approaches while striving for more accurate and reliable vein recognition systems. Through the passive utilization of the previous LUM technique and the proposal of the TUM technique, vein pattern recognition for computational techniques aims to achieve higher accuracy and efficiency in the authentication process. The continuous exploration of these models contributes to the advancement of vein pattern recognition and its application in various fields.

A dataset specifically tailored for the vein pattern experiment has been successfully generated through the utilization of a thermal camera.

7.3. DATABASE CREATION

In *Chapter **3*** of this thesis, the main focus of discussion was to develop biometric recognition/classification algorithms based on vein pattern recognition using thermal images. A database was created containing vein patterns from 30 subjects by taking images of both hands' forearms.

One of the primary challenges in capturing vein patterns using thermal images is that sometimes the veins may not be clearly visible, which can impact the accuracy of the recognition algorithm. Additionally, age can also affect the visibility of veins, with older individuals potentially having thinner or less visible veins, making it more challenging to capture their vein patterns.

To address these challenges, a database of vein patterns was aimed to be created that included individuals of different ages and genders. The database included 30 subjects, with both male and female participants of varying ages. By capturing vein patterns from individuals of different ages and genders, the effectiveness of the vein pattern recognition algorithm could be evaluated across a diverse range of subjects, and any potential limitations or biases could be identified.

However, capturing vein patterns from older individuals presented a particular challenge due to the potential difficulty in capturing clear images of their veins. Despite this, the aim was to include individuals from different age groups in the database to ensure that the algorithm's performance could be evaluated across a broad range of age groups. This approach ensured that the vein pattern recognition algorithm would be robust and effective in identifying vein patterns from individuals irrespective of their age or gender, making it suitable for a wide range of practical applications.

To create the database, 40 thermal images were taken for each subject, with 20 images taken for the right forearm and 20 for the left forearm. Here each subject was an individual class. For each subject, the vein patterns of both the left and right forearms were captured, resulting in a total of 60 individual classes. 30 class was for the left forearm and 30 class for the right forearm. The database was created in two scales, one in gray and another one in inverse gray for both left and right forearm.

7.4. PREPORSSESSING

The preprocessing techniques employed for creating the vein pattern database were similar to those used in *Chapter 4* for generating the sign language database. A block diagram has been shown in Fig. 7.1 that describes the flow of data acquisition.

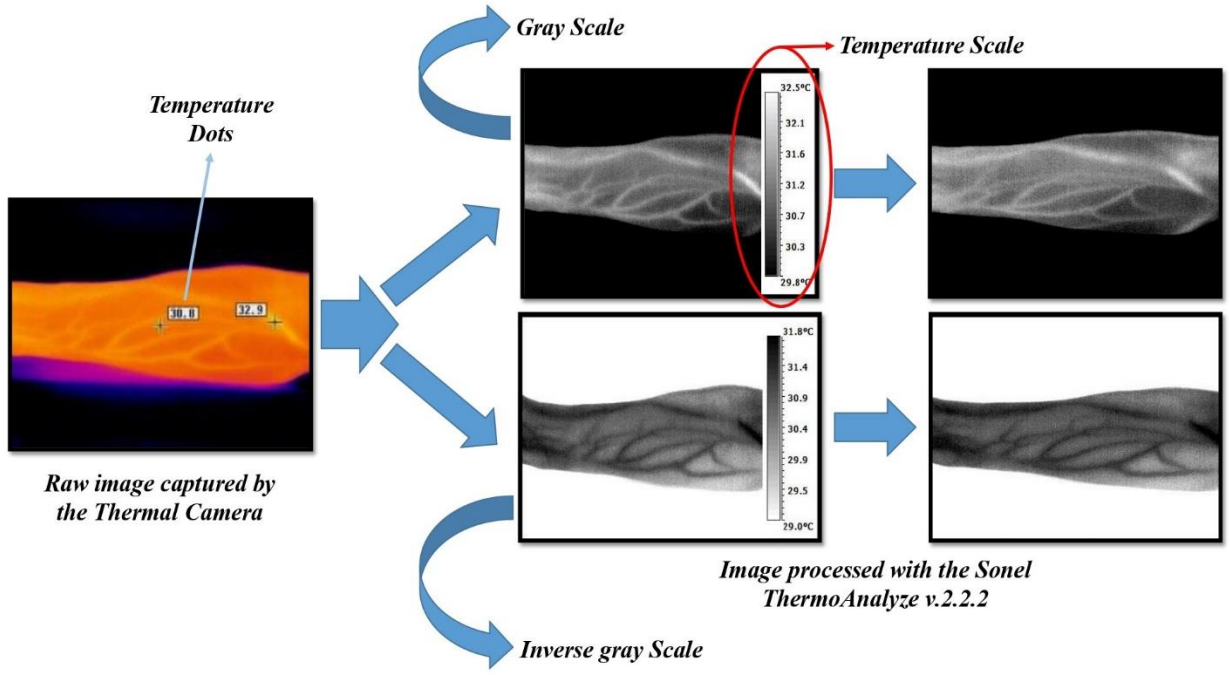


Fig. 7.1. Block diagram of Pre-processing employed for thermal images of human forearm.

In the **first step** of capturing the thermal human forearm images, thermal data was captured using a Sonel thermal camera. However, the captured images contained temperature dots that had to be removed to improve their quality. **Next**, the Sonel ThermoAnalyze v2.2.2 software was used to filter out the temperature dots from the raw thermal data. However, a temperature gauge bar was attached during the process of saving the images, which needed to be removed.

For this purpose, a software was used to edit the images and remove the gauge bar. This step was crucial to ensure that the images were suitable for further processing and analysis and that they accurately represented the hand gestures being performed.

The vein pattern databases are presented in four different variations:

1. Database of the left forearm in gray scale (Fig. 7.2).
2. Database of the left forearm in inverse gray scale (Fig. 7.3).
3. Database of the right forearm in gray scale (Fig. 7.4).
4. Database of the right forearm in inverse gray scale (Fig. 7.5).

Each figure represents the corresponding vein pattern database, showcasing the distinctive characteristics of the left and right forearm veins in both gray scale and inverse gray scale formats.

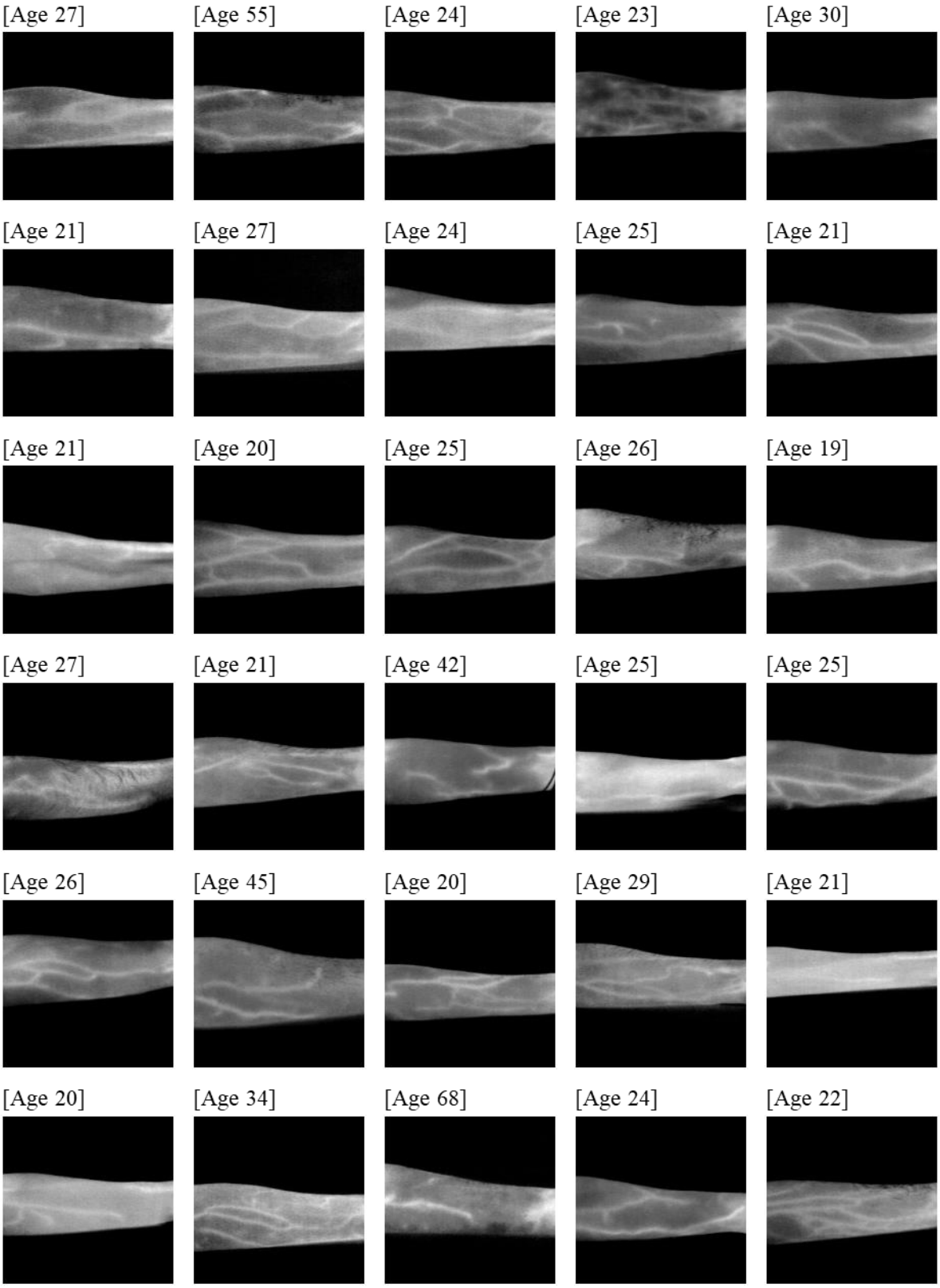


Fig. 7.2. Left forearm thermal vein pattern with different aged volunteers, shown in Gray scale

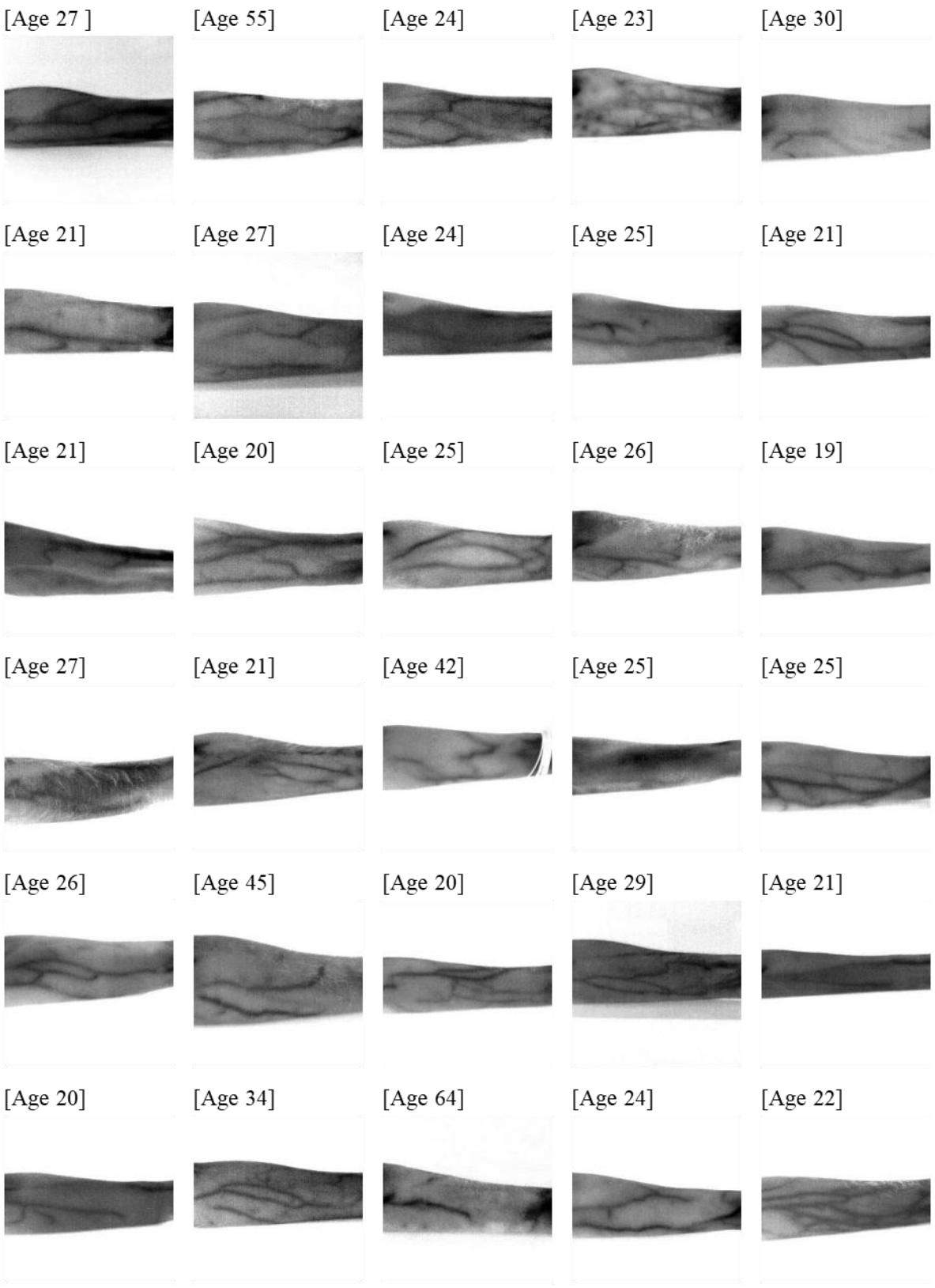


Fig. 7.3. Left forearm thermal vein pattern with different aged volunteers, in Inverse Gray scale

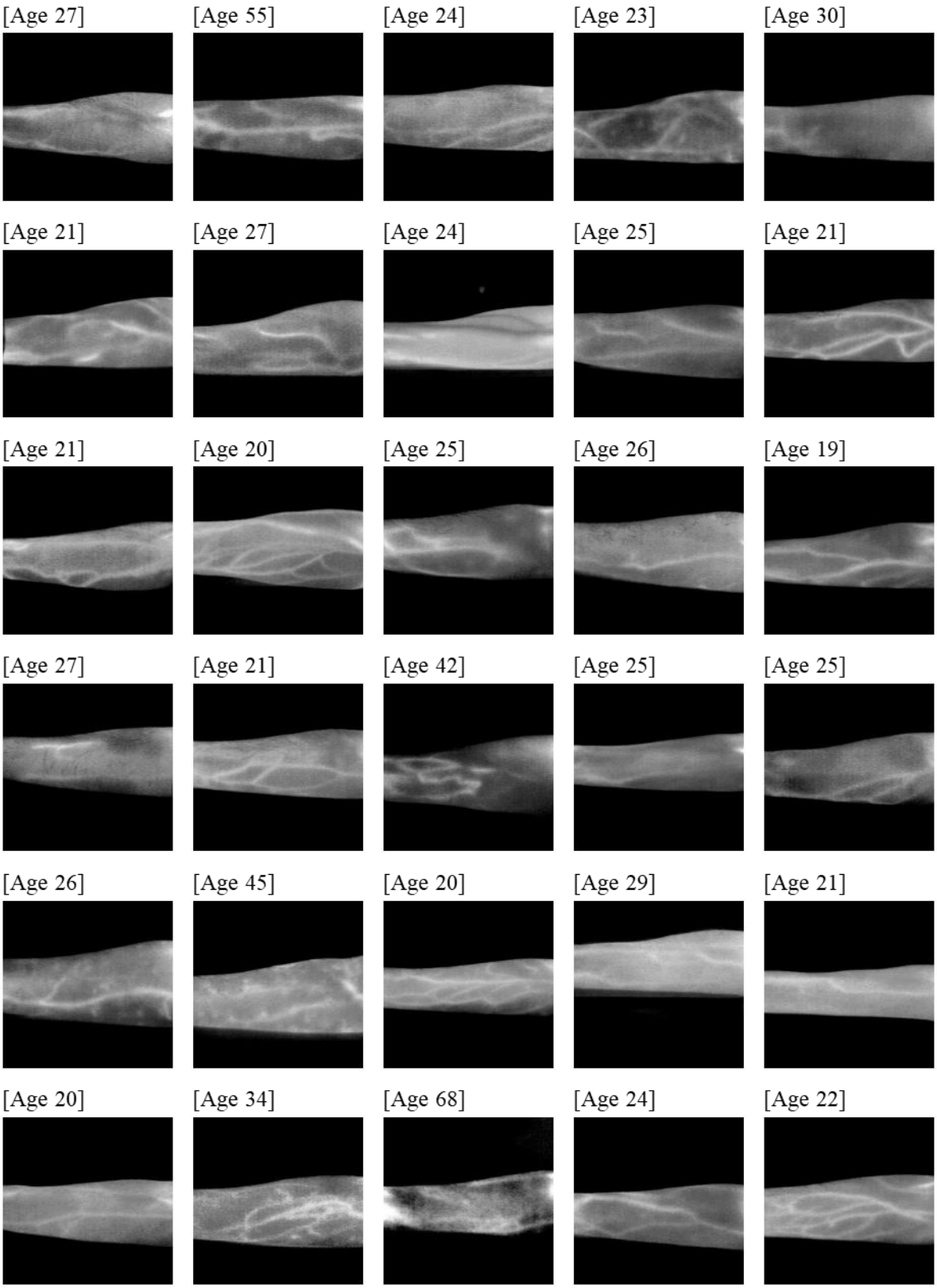


Fig. 7.4. Right forearm thermal vein pattern with different aged volunteers, in Gray scale

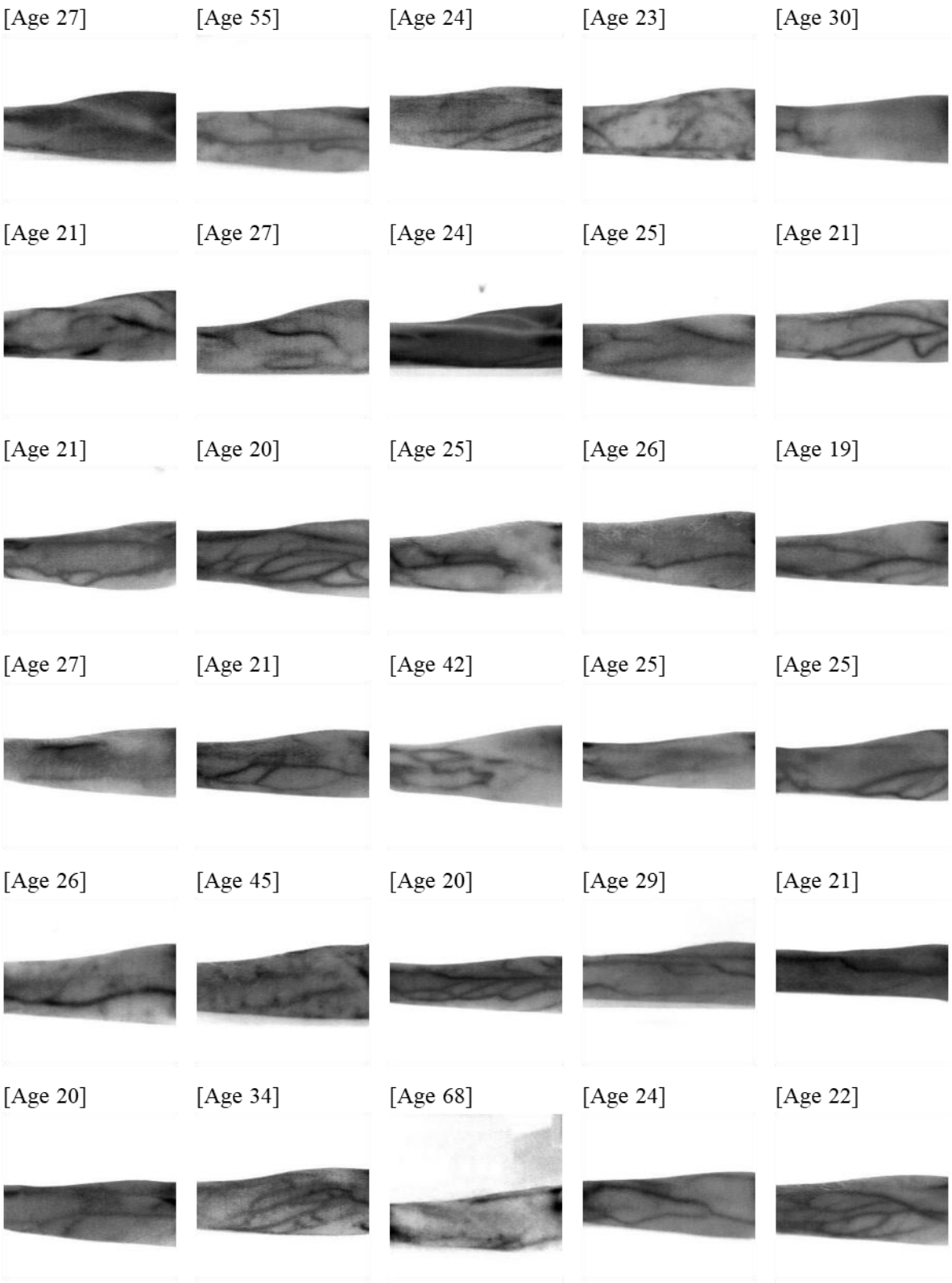


Fig. 7.5. Right forearm thermal vein pattern with different aged volunteers, in Inverse Gray scale

7.5. EXPERIMENTAL RESULTS

In this experimental study, the vein pattern database mentioned earlier is analyzed using the LUM (Laplacian Uniform Mixture) and TUM (Student's t Uniform Mixture) methods, which were extensively discussed in previous chapters. The LUM method, which has shown promising results in capturing vein patterns, and the proposed TUM method, designed to enhance recognition accuracy, are both employed in the analysis. The experimental results, obtained from both the left and right forearm, are presented in gray and inverse gray scales to evaluate the effectiveness and performance of the LUM and TUM methods in vein pattern recognition.

By tuning the parameters of the LUM and TUM methods, significant improvements were achieved for the vein pattern database. In the case of the LUM method, the parameters b, c, and λ (Lambda) were fine-tuned, while the parameters v, c, and α (Alpha) were adjusted for the TUM method, as discussed in previous chapters.

The results obtained from the parameter tuning process are presented in the following tables. These tables showcase the performance of the LUM and TUM methods in vein pattern recognition for the given database. The tables provide a comprehensive overview of the accuracy and effectiveness achieved by optimizing the parameters of each method.

The successful parameter tuning of the LUM and TUM methods has led to substantial improvements in vein pattern recognition for the dataset under investigation. These results demonstrate the significance of parameter optimization in enhancing the performance and reliability of vein pattern recognition systems.

At first, the performance results of the vein pattern database obtained using the LUM (Laplacian Uniform Mixture) error model are presented in Tables 7.1.

7.5.1. Performance Using LUM Model

Table: 7.1 Performance tables of vein database for both hands using LUM

Table:7.1(a) Performance of Database of the left forearm in gray scale	
Recognition Rate	0.8667
Accuracy (in %)	86.67%

Table:7.1(b) Performance of Database of the left forearm in inverse gray scale	
Recognition Rate	0.8722
Accuracy (in %)	87.22%

Table:7.1(c) Performance of Database of the right forearm in gray scale	
Recognition Rate	0.92.78
Accuracy (in %)	92.78%

Table:7.1(d) Performance of Database of the right forearm in inverse gray scale	
Recognition Rate	0.9056
Accuracy (in %)	90.56%

Next, the results obtained utilizing the TUM (Student's t Uniform Mixture) error model are presented in Tables 7.2.

7.5.2. Performance Using TUM Model

Tables: 7.2 Performance tables of vein database for both hands using TUM

Table:7.2(a) Performance of Database of the left forearm in gray scale	
Recognition Rate	0.9000
Accuracy (in %)	90.00%

Table:7.2(b) Performance of Database of the left forearm in inverse gray scale	
Recognition Rate	0.8833
Accuracy (in %)	88.33%

Table:7.2(c) Performance of Database of the right forearm in gray scale	
Recognition Rate	0.9333
Accuracy (in %)	93.33%

Table:7.2(d) Performance of Database of the right forearm in inverse gray scale	
Recognition Rate	0.9111
Accuracy (in %)	91.11%

The improvements over the results obtained in Table 7.1(a), Table 7.1(b), Table 7.1(c), and Table 7.1(d), obtained using LUM method, are demonstrated by the corresponding results presented in Table 7.2(a), Table 7.2(b), Table 7.2(c), and Table 7.2(d), respectively, obtained using TUM method. These tables provide a comprehensive representation of the advancements achieved through the implementation of the TUM method when compared to the results obtained using the LUM method.

Significant improvement is observed in the analysis of the previous results when the TUM (Student's t Uniform Mixture) model is employed, surpassing its predecessor, the LUM (Laplacian Uniform Mixture) model. The utilization of the TUM model not only enhances performance and effectiveness in addressing the encountered challenges but also demonstrates notable efficiency improvements compared to the LUM model. This improvement becomes evident when considering the observed outcomes and conducting comparisons between the TUM and LUM models. The substantial advancements showcased by the TUM model further emphasize its potential and efficacy in resolving the specific problems at hand, particularly when considering the vein pattern database used in this analysis.

7.6. SUMMARY

In this study, vein pattern recognition using computational techniques was thoroughly investigated, with particular emphasis on the effectiveness of the LUM (Laplacian Uniform Mixture) [116] and TUM (Student's t Uniform Mixture) methods. The vein pattern database was comprehensively analyzed, revealing notable outcomes and insights.

It was observed that both the LUM and TUM models exhibited promising results in capturing and recognizing vein patterns. Notably, the TUM model demonstrated a substantial improvement in performance and efficiency compared to the LUM model. This improvement is attributed to the meticulous optimization of parameters and algorithm refinement employed in the TUM model.

The findings obtained from this study underscore the potential of computational techniques

for vein pattern recognition in various domains, including biometric authentication and security systems. The successful implementation of the LUM and TUM models highlights the significance of parameter optimization and algorithm refinement, both of which contribute to achieving accurate and reliable vein pattern recognition.

Looking towards the future, there is a vast scope for further advancements in vein pattern recognition. Future research endeavors can explore novel computational techniques, such as deep learning algorithms, to enhance the accuracy and robustness of vein pattern recognition systems. Additionally, the integration of other modalities, such as multispectral imaging or 3D imaging, holds potential for expanding the applications and improving the recognition capabilities of vein pattern recognition.

In conclusion, this study has laid a strong foundation for advancing the field of vein pattern recognition. The future holds tremendous potential for further advancements and applications in this area, driven by continuous research and innovation in computational techniques for vein pattern recognition.

CHAPTER-8

CONCLUSION

- ***SYNOPSIS***
- ***FUTURE SCOPE***

8.1. SYNOPSIS

The thesis, "**HYBRID-UNIFORM MIXTURE MODEL BASED ITERATIVE ROBUST CODING FOR IMAGE RECOGNITION PROBLEMS USING THERMAL IMAGING**" reaches its conclusion at this point. The central focus of this thesis has been the creation of an affordable solution for hand gesture recognition that is well-suited for collaborative robotics. The widespread adoption of artificial intelligence and advanced computing techniques worldwide has greatly contributed to the concept of human-computer coexistence in various environments. However, in developing countries, the implementation of such human-computer interaction (HCI) environments necessitates cost-effective solutions. With this in mind, a visual sensor-based thermal camera, which is moderately priced compared to other high-cost sensors, has been chosen for the collaborative robotics tasks in this study. The primary objective has been to establish a robust mathematical framework for computation, rather than relying on expensive sensors. Additionally, significant attention has been given to addressing environmental challenges throughout this work, and existing mathematical models have been enhanced to effectively handle such situations.

Following an introductory session in ***Chapter 1*** and a comprehensive literature survey in ***Chapter 2, Chapter 3*** formulates a potential problem based on the insights gained from the previous chapters. ***Chapter 4*** provides an overview of the hand gesture database, which has been meticulously prepared in the Electrical Measurement and Instrumentation Laboratory of Jadavpur University. This database serves as the foundation for all subsequent experimental investigations conducted in this thesis. In ***Chapter 5***, the effectiveness of the robust regression-based method known as LUM is extensively examined within the context of the recognition system. Building upon the principles and components of LUM, ***Chapter 6*** introduces a newly proposed method named TUM, accompanied by detailed experimental studies. TUM offers a robust solution for the recognition system, particularly in highly challenging environments. ***Chapter 7*** marks the initiation of a new research endeavor focusing on thermal image vein pattern recognition on the human forearm, employing the LUM and TUM methods. This undertaking presents a significant challenge; however, the preliminary results obtained from the first attempt demonstrate promising outcomes. The initial findings indicate favorable performance, providing a solid foundation for further exploration and advancement of these two methods in vein pattern recognition. This work showcases the potential of LUM and TUM in tackling the complexities associated with thermal

image vein pattern analysis and lays the groundwork for future investigations in this area.

Based on the experimental findings, it has been observed that in situations where images suffer from moderate challenges such as poor illumination and occlusion, LUM demonstrates superior performance as a robust regression method compared to other contemporary alternatives. The use of thermal image datasets further enhances the productivity of LUM, contributing significantly to the overall objective of developing a cost-effective and robust hand gesture recognition solution in human-computer interaction (HCI) environments. Since thermal images were employed, the impact of illumination effects was minimized. Furthermore, the efficiency of LUM has been enhanced through parameter tuning, which is extensively discussed in this thesis. These investigations are supported by detailed mathematical models and experimental results.

In this thesis, a new model named TUM is introduced, building upon the foundation of LUM. Through the utilization of thermal image datasets, the accuracy of TUM surpasses that of the LUM model. This development highlights the potential of robust regression-based gesture recognition models, such as LUM and TUM, in providing effective and affordable solutions for challenging environments. The newly proposed method, TUM, based on LUM, elevates the recognition system to a higher level of robustness.

In conclusion, the comprehensive studies conducted in this thesis reveal that robust regression-based gesture recognition models, including LUM and TUM, have the potential to offer excellent moderate-cost solutions for demanding environments. The incorporation of thermal image datasets and the advancements made in the TUM model contribute to the system's scalability and enhanced robustness.

8.2. FUTURE SCOPE

The significance of HCI in various aspects of human life cannot be overstated. It encompasses a wide range of applications, both direct and indirect, that permeate numerous spheres of daily existence. From household chores to industrial processes, from aviation operations to medical practices, from education to entertainment, and from civil devices to defense systems, HCI has left its mark in every domain. Its contributions have played a pivotal role in advancing the pace and quality of our world today. The widespread adoption of HCI has served as a catalyst for continuous research efforts in this field, with scholars and practitioners consistently striving to make further advancements and innovations. The positive impact of HCI on a global scale has

served as a driving force for researchers to remain dedicated and engaged in this domain throughout the years.

In light of the significance of the research conducted on robust regression models in the field of HCI, it is essential to reflect on the work done and provide insights for future researchers.

Firstly, there is a need for further investigation into the potential extension of the proposed regression models to address different types of noise. In addition to challenges related to temperature and occlusions, it is important to consider various other real-world obstacles that can impact gesture recognition. This exploration may involve examining multiple real-world noises concurrently or creating synthetic complex noises to develop a more resilient gesture recognition model. By delving into these areas, future researchers can build upon the foundation laid by this study and make advancements in the field of robust regression models for HCI applications.

Secondly, furthermore, while the methods examined in this study exhibit superior performance in comparison to existing state-of-the-art techniques, there is still potential for further enhancements. There is an opportunity to explore the utilization of more error-tolerant weight functions, such as TUM, to advance the capabilities of the system. By investigating and implementing these alternative weight functions, researchers can explore avenues for improving the robustness and accuracy of the system. This endeavor can contribute to the continuous evolution and refinement of gesture recognition methodologies in the field of HCI.

Thirdly, the computational efficiency of the system emerges as another crucial factor to consider. Given the increasing demand for real-time applications in today's fast-paced world, HCI-based systems must exhibit sufficient speed to keep up with these requirements. It is worth noting that LUM has demonstrated faster computational performance compared to previous versions like RRC. However, the TUM models currently exhibit slightly slower computational processes in comparison to LUM, indicating the need for further improvements in terms of computational time. Nevertheless, there remains untapped potential for developing an even faster recognition scheme, as swift processes facilitate better synchronization among interacting entities, a pivotal criterion for successful HCI. Hence, exploring avenues to enhance the computational efficiency of gesture recognition systems holds promise for further advancements in the field.

Fourthly, in the thesis, a new vein pattern recognition task on the human forearm was initiated. The objective is to enhance the recognition rate by employing the LUM and TUM models, along with exploring the potential introduction of additional computational techniques for future studies. Furthermore, there is a requirement for further investigation into the extension of the proposed regression models to effectively address diverse types of noise. Besides the challenges associated with temperature variations and occlusions, it is crucial to consider a wide range of real-world obstacles that can significantly impact gesture recognition. By addressing these aspects, it is anticipated that the research will contribute to the development of more advanced and reliable systems in the field.

As a result of these considerations, it is anticipated that more efficient gesture detection systems will be developed in the near future, thereby contributing to the construction of improved HCI systems. Additionally, there is hope that the continued evolution of HCI systems will lead to a better and more convenient world for both humans and other species to inhabit. By incorporating advancements in technology and understanding human-computer interaction, it is envisaged that these systems will enhance the overall quality of life and simplify various aspects of daily existence.

BIBLIOGRAPHY

[1] A. Ajoudani, A. M. Zanchettin, S. Ivaldi, A. Albu-Schaffer, K. Kosuge, ¨ and O. Khatib, "Progress and prospects of the human–robot collaboration," *Auton. Robots*, vol. 42, pp. 957–975, Oct. 2017.

[2] O. Khatib, K. Yokoi, O. Brock, K. Chang, and A. Casal, "Robots in human environments: Basic autonomous capabilities," *Int. J. Robot. Res.*, vol. 18, no. 7, pp. 684–696, 1999.

[3] Thomas B. Sheridan,"Human–Robot Interaction: Status and Challenges," vol. 58, no. 4, pp. 525–532 Jun 2016,

[4] N. Mavridis, "A review of verbal and non-verbal human-robot interactive communication," *Robotics and Autonomous Systems*, vol. 63, part 1, pp. 22-35, Jan. 2015.

[5] S. Ganapathyraju, "Hand gesture recognition using convexity hull defects to control an industrial robot," in *Proc. Int. Conf. Inst. Cont. Auto*., pp. 63-67, Aug. 2013.

[6] M. Simão, P. Neto, and O. Gibaru, "Natural control of an industrial robot using hand gesture recognition with neural networks," in *Proc. IECON Ann. Conf. IEEE Indus. Elec. Soc.*, pp. 5322-5327, Oct. 2016.

[7] S. P. Kang, G. Rodnay, M. Tordon, and J. Katupitiya, "A hand gesture based virtual interface for wheelchair control," in *Proc. IEEE/ASME Int. Conf. Adv. Int. Mechat.*,pp. 778-783, Jul. 2003.

[8] Y. Zhang, J. Zhang, and Y. Luo, "A novel intelligent wheelchair control system based on hand gesture recognition," in *Proc. IEEE/ICME Int. Conf. Compl. Med. Engg*., pp. 334-339, May 2011.

[9] C. Ma, S. Zhang, A. Wang, Y. Qi and Ge Chen, "Skeleton-Based Dynamic Hand Gesture Recognition Using an Enhanced Network with One-Shot Learning."

[10] M. Simão and O. Gibaru, "Online recognition of incomplete gesture data to interace collaborative robots," *IEEE Trans. Indus. Elec.*, vol. 66, no. 12, pp. 9372-9382, Jan. 2019.

[11] Anju S R, S. Surendran, "A Study on Different Hand Gesture Recognition Techniques," vol. 3, no. 4, Apr. 2014, ISSN: 2278-0181

[12] J. Wu, L. Sun, and R. Jafari, "A Wearable System for Recognizing American Sign Language in Real-Time Using IMU and Surface EMG Sensors," vol. 20, no. 5, pp. 2168-2194, Sep. 2016.

[13] S. Joardar, A. Chatterjee, S. Bandyopadhyay, U. Maulik, "Multi-size patch based collaborative representation for Palm Dorsa Vein Pattern recognition by enhanced ensemble learning with modified interactive artificial bee colony algorithm," *Engineering Applications of Artificial Intelligence*, vol. 60, pp. 151-163, 2017

[14] S. Joardar, A. Chatterjee and A. Rakshit, "A Real-Time Palm Dorsa Subcutaneous Vein Pattern Recognition System Using Collaborative Representation-Based Classification,", vol. 64, no. 4, Apr. 2015.

[15] D. K. Vishwakarma and V. Grover, “Hand gesture recognition in low intensity environment using depth images,” in *Proc. Int. Conf. Intell. Sustain. Syst*., pp. 429–433, Dec. 2017,

[16] D. K. Vishwakarma, R. Maheshwari, and R. Kapoor, “An efficient approach for the recognition of hand gestures from very low resolution images,” in *Proc. 5th Int. Conf. Commun. Syst. Netw. Technol.*, pp. 467–471, Apr. 2015.

[17] D. K. Vishwakarma and R. Kapoor, “Simple and intelligent system to recognize the expression of speech-disabled person,” in *Proc. 4th Int. Conf. Intell. Hum. Comput. Interact. (IHCI)*, pp. 1–6, Dec. 2012.

[18] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, and J. Yang, ‘‘A framework for hand gesture recognition based on accelerometer and EMG sensors,’’ *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 41, no. 6, pp. 1064–1076, Nov. 2011.

[19] V. E. Kosmidou and L. J. Hadjileontiadis, ‘‘Sign language recognition using intrinsic-mode sample entropy on sEMG and accelerometer data,’’ *IEEE Trans. Biomed. Eng.*, vol. 56, no. 12, pp. 2879–2890, Dec. 2009.

[20] T. D. Bui and L. T. Nguyen, ‘‘Recognizing postures in vietnamese sign language with MEMS accelerometers,’’ *IEEE Sensors J.*, vol. 7, no. 5, pp. 707–712, May 2007.

[21] G. Fang, W. Gao, and D. Zhao, ‘‘Large-vocabulary continuous sign language recognition based on transition-movement models,’’ *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 37, no. 1, pp. 1–9, Jan. 2007.

[22] U. Cote-Allard, C. L. Fall, A. Drouin, A. Campeau-Lecours, C. Gosselin, K. Glette, F. Laviolette, and B. Gosselin, ‘‘Deep learning for electromyographic hand gesture signal classification using transfer learning,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 4, pp. 760–771, Apr. 2019.

[23] W. Zhi-heng, C. Jiang-tao, L. Jin-guo, and Z. Zi-qi, “Design of human-computer interaction control system based on hand-gesture recognition,” *in Proc. You. Aca. Ann. Conf. Chi. Asso. Auto*., pp. 143-147, May 2017,

[24] T. Takahashi and F. Kishio, “A hand gesture recognition method and its application,” *Systems and Computers in Japan*, vol. 23, no. 3, 1992.

[25] J. Davis and M. Shah, “Visual gesture recognition,” *IEE Proc. Vis. Image Signal Process*, vol. 141, no. 2, Apr. 1994.

[26] C. C. Lien and C. L. Huang, “The model-based dynamic hand posture identification using genetic algorithm,” *Machine Vision and Application*, vol. 11, pp. 107-121, 1999.

[27] Z. Ren, J. Yuan, J. Meng, and Z. Zhang, “Robust part-based hand gesture recognition using Kinect sensor,” *IEEE Trans. Multimedia*, vol. 15, no. 5, pp. 1110-1120, Aug. 2013.

[28] K. Choi, M. Sato, and Y. Koike, “Consideration of the embodiment of a new, human-centered interface,” *IEICE Trans. Inf. Sys.*, vol. E89D, pp. 1826-1833, Jun. 2006.

[29] B. Rebsamen, C. L. Teo, Q. Zeng, M. H. Ang Jr., E. Burdet, C. Guan, H. Zhang, and C. Laugier, “Controlling a wheelchair indoors using thought,” *IEEE Intelligent Systems*, vol. 22, no. 2, pp. 18-24, Mar. 2007.

[30] J. L. Raheja, G. A. Rajsekhar, and A. Chaudhury, "Controlling a remotely located robot using hand gestures in real time: A DSP implementation," in *Proc. Int. Conf. Wir. Net. Embed. Sys.*, Oct. 2016.

[31] D. S. Breland, S. B. Skriubakken, A. Dayal, A. Jha, P. K. Yalavarthy, and L. R. Cenkeramaddi, "Deep learning-based sign language digits recognition from thermal images with edge computing system," *IEEE Sensors J.*, vol. 21, no. 9, pp. 10445–10453, Feb. 2021.

[32] P. Kumar *et al*. Nus Dataset. Accessed: May, 2023. [Online]. Available: https://www.ece.nus.edu.sg/stfpage/elepv/NUS-HandSet/

[33] P. P. Kumar, P. Vadakkepat, and A. P. Loh, "Hand posture and face recognition using a fuzzy-rough approach," *Int. J. Humanoid Robot.*, vol. 7, no. 3, pp. 331–356, Sep. 2010.

[34] Q. Chen, N. D. Georganas, and E.M. Petriu, "Real-time vision-based hand gesture recognition using Haar-like features," in *Proc. IEEE Instr. Meas. Tech. Conf.*, pp. 1-6, May 2007.

[35] M. Elbes, S. Alzubi, T. Kanan, A. Al-Fuqaha, and B. Hawashin, "A survey on particle swarm optimization with emphasis on engineering and network applications," *Evol. Intel.*, vol. 12, pp. 113-129, Feb. 2019.

[36] M. R. Abid, E. M. Petriu, and E. Amjadian, "Dynamic sign language recognition for smart home interactive application using stochastic linear formal grammar," *IEEE Trans. Instr. Meas.*, vol. 64, no. 3, pp. 596-605, Mar. 2015.

[37] R. Xie and J. Cao, "Accelerometer-based hand gesture recognition by neural network and similarity matching," *IEEE Sens. Journ.*, vol. 16, no. 11, pp. 4537-4545, Jun 2016.

[38] Chen, S., Liu, Z., Zhao, Y., & Wu, Z.,"Hand gesture recognition for sign language using deep learning," *Journal of Intelligent & Fuzzy Systems*, vol. 36 no. 6, pp. 5877-5885, 2019

[39] Liu, Y., Lv, C., & Li, L. "Recognition of sign language gesture based on sensor fusion," *Journal of Physics: Conference Series*, vol. 1053, no. 1, pp. 12-51, 2018

[40] Chen, S., Liu, Z., & Wu, Z. "Hand gesture recognition system for sign language interpretation based on deep learning," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 1, pp. 1197-1205, 2021

[41] Velez-Rojas, M. A., & Escamilla-Ambrosio, P. J. "Hand gesture recognition using thermal images and convolutional neural networks," *International Journal of Innovative Computing, Information and Control*, vol. 15, no. 6, pp. 2385-2393, 2019

[42] He, Z., Yan, Y., Zhang, Y., & Zhang, L. "Hand gesture recognition based on thermal images for sign language translation". in *2020 39th Chinese Control Conference (CCC) IEEE*, pp. 7108-7113, 2020

[43] J. Yang, J. Wright, T. Huang, Y. Ma, "Image super-resolution as sparse representation of raw patches," *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2008.

[44] S. Rao, R. Tron, R. Vidal, Y. Ma, "Motion segmentation via robust subspace separation in the presence of outlying, incomplete, and corrupted trajectories," in *Proceedings of the*

IEEE International Conference on Computer Vision and Pattern Recognition, 2008.

[45] J. Mairal, G. Sapiro, M. Elad, "Learning multiscale sparse representations for image and video restoration," *SIAM MMS* ,vol. 7, no. 1, pp. 214–241, 2008

[46] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 31, no. 2, pp. 210–227, 2009

[47] M. Sadeghi, M. Babaie-Zadeh, and C. Jutten, "Dictionary learning for sparse representation: A novel approach," IEEE Sig. Proc. Lett., vol. 20, no. 12, pp. 1195-1198, Dec. 2013.

[48] X. Wu, W. Zhu, and J. Yan, "Direction of arrival estimation for off-grid signals based on sparse Bayesian learning," *IEEE Sens. Journ.*, vol. 16, no. 7, pp. 2004-2016, Apr. 2016.

[49] S. H. Fouladi, S. Chiu, B. D. Rao, and I. Balasingham, "Recovery of independent sparse sources from linear mixtures using sparse Bayesian learning," *IEEE Trans. Sig. Proc.*, vol. 66, no. 24, pp. 6332-6346, Dec. 2018.

[50] A. Adler, M. Elad, and Y. Hel-Or, "Probabilistic subspace clustering via sparse representations," *IEEE Sig. Proc. Lett.*, vol. 20, no. 1, pp. 63-66, Jan. 2013.

[51] Y. Wang, A. Yang, Z. Li, X. Chen, P. Wang, and H. Yang, "Blind drift calibration of sensor networks using sparse Bayesian learning," *IEEE Sens. Journ.*, vol. 16, no. 16, pp. 6249-6260, Aug. 2016.

[52] X. Zhou, B. Fan, H. Wang, Y. Cheng, and Y. Qin, "Sparse Bayesian perspective for radar coincidence imaging with array position error," *IEEE Sens. Journ.*, vol. 17, no. 16, pp. 5209-5219, Aug. 2017.

[53] Jun Yin, Zhonghua Liu, Zhong Jin, Wankou Yang, "Kernel sparse representation based classification," *Neurocomputing*, vol. 77, pp. 120-128, 2012.

[54] Jun Yin, Lai Wei, Miao Song, Weiming Zeng, "Optimized projection for Collaborative Representation based Classification and its applications to face recognition," *Pattern Recognition Letters*, vol. 73, pp. 83-90, 2016

[55] Lei Zhanga, Meng Yang, Xiangchu Feng, Yi Ma and David Zhang, "Collaborative Representation based Classification for Face Recognition."

[56] S.H. Gao, I.W-H. Tsang, and L-T. Chia, "Kernel Sparse Representation for Image classification and Face recognition," in *Proc. European Conf. Computer Vision*, 2010.

[57] M. Yang and L. Zhang, "Gabor Feature based Sparse Representation for Face recognition with Gabor Occlusion Dictionary," in *Proc. European Conf. Computer Vision*, 2010.

[58] B. Cheng, J. Yang, S. Yan, Y. Fu, and T. Huang, "Learning with l1-graph for image analysis," *IEEE Trans. Image Processing*, vol. 19, no. 4, pp. 858-866, 2010.

[59] L.S. Qiao, S.C. Chen, and X.Y. Tan, "Sparsity preserving projections with applications to face recognition", *Pattern Recognition*, vol. 43, no. 1, pp. 331-341, 2010.

[60] J. Yang, K. Yu, Y. Gong and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2009.

[61] J.Z. Huang, X.L. Huang, and D. Metaxas, "Simultaneous image transformation and sparse

representation recovery," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2008.

[62] A. Wagner, J. Wright, A. Ganesh, Z.H. Zhou, and Y. Ma, "Towards a practical face recognition system: robust registration and illumination by sparse representation," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2009.

[63] Y.G. Peng, A. Ganesh, J. Wright, W.L. Xu, and Y. Ma, "RASL: robust alignment by sparse and low-rank decomposition for linearly correlated images," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2010.

[64] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zissserman, "Learning discriminative dictionaries for local image analysis," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2008.

[65] Q. Zhang, and B. Li, "Discriminative K-SVD for Dictionary Learning in Face Recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2010.

[66] Z.L. Jiang, Z. Lin, L.S. Davis, "Learning A Discriminative Dictionary for Sparse Coding via Label Consistent KSVD," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2011.

[67] M. Yang, L. Zhang, X.C. Feng, and D. Zhang, "Fisher Discrimination Dictionary Learning for Sparse Representation," in *Proc. Int'l Conf. Computer Vision,* 2011.

[68] Sijia Cai, Lei Zhang, Wangmeng Zuo, Xiangchu Feng, "A Probabilistic Collaborative Representation based Approach for Pattern Classification."

[69] D. Zhang, M. Yang, X. Feng, "Sparse representation or collaborative representation: which helps face recognition" *In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (IEEE),* pp. 471–478, 2011.

[70] W. Liu, L. Lu, H. Li, W. Wang, Y. Zou, "A novel kernel collaborative representation approach for image classification," in *Proceedings of the IEEE International Conference on Image Processing (ICIP), Paris, France, IEEE*, pp. 4241– 4245, 2014.

[71] W.K. Yang, Z.Y. Wang, C.Y. Sun, "A collaborative representation based projections method for feature extraction, Pattern Recognit," vol. 48, pp. 20–27, 2015.

[72] Meng Yang, Lei Zhang, Jian Yang, David Zhang, "Robust Sparse Coding for Face Recognition"

[73] S. H. Gao, I. W. H. Tsang, L. T. Chia, and P. L. Zhao, "Local features are not lonely-laplacian sparse coding for image classification" in *CVPR*, pp. 626, 2010.

[74] A. Wagner, J. Wright, A. Ganesh, Z. H. Zhou, and Y. Ma., "Towards a practical face recognition system: Robust registration and illumination by sparse representation," in *CVPR*, pp. 625-630, 2009.

[75] M. Yang and L. Zhang,"Gabor feature based sparse representation for face recognition with gabor occlusion dictionary," in *ECCV*, pp. 625, 630, 631, 632. 2010.

[76] Y. N. Liu, F. Wu, Z. H. Zhang, Y. T. Zhuang, and S. C. Yan. "Sparse representation using nonnegative curds and whey". In *CVPR*, pp. 626, 2010.

[77] J. Z. Huang, X. L. Huang, and D. Metaxas. "Simultaneous image transformation and sparse

representation recovery," In *CVPR*, 625. 2008.
[78] I. Ramirez and G. Sapiro. "Universal sparse modeling. Technical report," *arXiv:1003.2941v1*[cs.IT], University of Minnesota, pp. 626, 2010.
[79] S. H. Ji, Y. Xue, and L. Carin. "Bayesian compressive sensing." *IEEE SP*, vol. 56, no. 6, pp. 2346-2356. 626 , 2008.
[80] J. J. Wang, J. C. Yang, K. Yu, F. J. Lv, T. Huang, and Y. H. Gong."Locality-constrained linear coding for image classification". in CVPR, pp. 626, 2010.
[81] Cao, X., Wei, L., Zhang, L., & Yang, Y., "Sparse coding with shifted Laplacian regularization for image classification", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* , pp. 919-926, 2013
[82] Zhang, L., Yang, M., & Feng, X., "Robust image retrieval via joint feature learning of multiple cues", *IEEE Transactions on Image Processing*, vol. 25, no. 6, pp. 2475-2490, 2016.
[83] Zhang, X., Zhang, L., & Shi, Y., "Multi-view learning via regularized robust coding", *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 11, pp. 5738-5750.
[84] Zhang, Y., Li, X., & Li, X., "Face recognition using regularized robust coding based on non-subsampled contourlet transform. Signal Processing", 175, 107736, 2020.
[85] Zhang, X., Zhang, L., & Yang, J., "Robust image restoration based on regularized sparse coding with adaptive graph", *IEEE Transactions on Image Processing*, vol. 26, no. 1, pp. 233-246, 2017.
[86] Zhang, X., Zhang, L., & Yang, J., "Regularized robust coding for image denoising via non-local self-similarity", *IEEE Transactions on Image Processing*, vol. 28, no. 6, pp. 2819-2831, 2019.
[87] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Proc. IEEE Comp. Soc. Conf. Comp. Vis. Patt. Recog.*, pp. 3360-3367, Jun 2010.
[88] T. Zhou, H. Bhaskar, F. Liu, and J. Yang, "Graph regularized and locality-constrained coding for robust visual tracking," *IEEE Trans. Circuits Sys. Video Tech.*, vol. 27, no. 10, pp. 2153-2164, Oct. 2017.
[89] X. Mei and H. Ling, "Robust visual tracking using ℓ1 minimization," in *Proc. IEEE Int. Conf. Comp. Vis.*, pp. 1436-1443, Oct. 2009.
[90] M. Javanmardi, M. Yazdi, and M. M. Shirazi, "Fast and robust L0-tracker using compressive sensing," in *Proc. Int. Conf. Patt. Recog. Image Ana.*, pp. 1-6, Mar. 2015.
[91] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?," in *Proc. Int. Conf. Comp*. Vis., pp. 471-478, Nov. 2011.
[92] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Regularized robust coding for face recognition," *IEEE Trans. Image Proc.*, vol. 22, no. 5, pp. 1753-1766, May 2013.
[93] Meng Yang, Member, Lei Zhang, Member, Jian Yang, Member, and David Zhang, "Regularized Robust Coding for Face Recognition," vol 22, no 5, pp.1057-7149, May 2013.

[94] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 31, no. 2, pp. 210–227, Feb. 2009.

[95] J. Wright and Y. Ma, “Dense error correction via l_1 minimization,” *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3540–3560, Jul. 2010.

[96] R. He, W. S. Zheng, B. G. Hu, and X. W. Kong, “A regularized correntropy framework for robust pattern recognition*,” Neural Comput*., vol. 23, no. 23, pp. 2074–2100, 2011.

[97] R. He, W. S. Zheng, and B. G. Hu, “Maximum correntropy criterion for robust face recognition,” *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 33, no. 8, pp. 1561–1576, Aug. 2011.

[98] M. Yang, L. Zhang, J. Yang, and D. Zhang, “Robust sparse coding for face recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 625–632, Jun. 2011.

[99] W. F. Liu, P. P. Pokharel, and J. C. Principe, “Correntropy: Properties and applications in non-Gaussian signal processing,” *IEEE Trans. Signal Process*., vol. 55, no. 11, pp. 5286–5298, Nov. 2007.

[100] "A Gaussian-Uniform Mixture Model for Anomaly Detection in Crowded Scenes" (2018).

[101] J. Xu *et al*., "Adaptive Online Anomaly Detection Using a Gaussian-Uniform Mixture Model", 2019.

[102] S. S. Lee *et al*, "Gaussian-Uniform Mixture Model for Detecting Malicious Attacks in Network Traffic", 2020.

[103] Xie, Y., & Zha, H., “Multilevel thresholding based on the maximum Tsallis entropy via a two-stage optimal graph method. Pattern recognition”, vol. 37, no. 11, pp. 2259-2268.

[104] De, S., Mandal, S., & Chaudhuri, B. B., “Robust anomaly detection using Laplacian mixture model”, *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 4, pp. 619-633.

[105] Li, Z., Liu, L., & Yang, J., “Robust face recognition using Laplacian mixture model-based sparse representation. Pattern Recognition Letters”, vol. 91, pp. 36-43, 2017.

[106] Liu, L., Li, Z., & Yang, J., “Robust image denoising via Laplacian uniform mixture model-based sparse representation”, *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 3164-3177, 2016.

[107] Huang, Y., Li, Z., Wang, X., & Yang, J., “Laplacian scale mixture model for data with heavy tails and outliers”, *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4811-4826, 2017

[108] Zhang, Y., Zhang, Z., Ma, J., & Wu, X. (2020), “Robust retinal vessel segmentation based on Laplacian uniform mixture model. Computers in Biology and Medicine”, 124, 103930.

[109] Zhang, J., Li, Y., Chen, C., & Tao, R., “Laplacian mixture model-based feature extraction for hyperspectral image classification”, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 2, pp. 564-575, 2018.

[110] Banfield, J.D., Raftery, A.E., “Model-based gaussian and non-gaussian clustering”, *Biometrics*, pp. 803–821, 1993.

[111] Coretto, P., Hennig, C., "Robust improper maximum likelihood: tuning, computation, and a comparison with other methods for robust Gaussian clustering", *JASA* 111, pp. 1648–1659 , 2016.
[112] Meer, P., Mintz, D., Rosenfeld, A., Kim, D.Y. "Robust regression methods for computer vision: A review", *IJCV* , vol. 6, no. 1, pp. 59–70, 1991.
[113] Gebru, I.D., Alameda-Pineda, X., Forbes, F., Horaud, R., "Em algorithms for weighted-data clustering with application to audio-visual scene analysis", *IEEE TPAMI* , vol. 38, no. 12, pp. 2402–2415, 2016.
[114] S. Lathuili, P. Mesejo, X. Alameda-Pineda, and R. Horaud, "DeepGUM: Learning Deep Robust Regression with a Gaussian-Uniform Mixture Model", arXiv:1808.09211v1 [cs.CV] Aug. 28, 2018.
[115] S. Z. Gurbuz, Ali C. Gurbuz, E. A. Malaia, D. J. Griffin, C. S. Crawford , Mohammad M. Rahman, E. Kurtoglu, R. Aksu, T. Macks, and R. Mdrafi, "American Sign Language Recognition Using RF Sensing," vol. 21, no. 3, Feb 1, 2021.
[116] 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," vol. 31, no. 9, Sep , 2020.
[117] R. He, W.-S. Zheng, T. Tan, and Z. Sun, "Half-quadratic-based iterative minimization for robust sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 36, no. 2, pp. 261–275, Feb. 2014.
[118] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Signal Process*., vol. 4, no. 7, pp. 932–946, Jul. 1995.
[119] D. Hasler, L. Sbaiz, S. Süsstrunk, and M. Vetterli, "Outlier modeling in image matching," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 25, no. 3, pp. 301–315, Mar. 2003.
[120] P. H. S. Torr, R. Szeliski, and P. Anandan, "An integrated Bayesian approach to layer extraction from image sequences*," IEEE Trans. Pattern Anal. Mach. Intell*., vol. 23, no. 3, pp. 297–303, Mar. 2001.
[121] Sonel KT-384 datasheet, Accessed in: May, 2023. [online]. Available. http://www.sonel.pl/sites/default/files/pl/ins/kt384_ins_104_plgb.pdf.
[122] Historical weather information, Accessed in: May, 2023. [online]. Available. https://www.timeanddate.com/weather/@1269628/historic?month=11&year=2022.
[123] D. S. Breland , A. Dayal , A. Jha , P. K. Yalavarthy, Om J. Pandey and L. R. Cenkeramaddi, "Robust Hand Gestures Recognition Using a Deep CNN and Thermal Images," vol. 21, no. 23, pp.1558-1748, Dec. 1, 2021.
[124] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci*., vol. 2, no. 1, pp. 183–202, 2009.
[125] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput*., vol. 20, no. 1, pp. 33–61, 1999.
[126] D. Donoho and M. Elad, "Optimal sparse representation in general (nonorthogonal) dictionaries via minimization," in *Proc. Nat. Acad. Sci. USA*, vol. 100, no. 5, pp. 2197–

2202, Mar. 2003.
[127] D. L. Donoho, “For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution.”
[128] K. Guo, L. Liu, X. Xu, D. Xu, and D. Tao, “GoDec+: Fast and robust low-rank matrix decomposition based on maximum correntropy,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2323–2336, Jun. 2018.
[129] R. He, W.-S. Zheng, and B.-G. Hu, “Maximum correntropy criterion for robust face recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1561–1576, Aug. 2011.
[130] P. Huber, *Robust Statistics*, 2nd ed. Hoboken, NJ, USA: Wiley, 2009.
[131] H. Jia and A. M. Martinez, “Face recognition with occlusions in the training and testing sets,” in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, pp. 1–6, Sep. 2008.
[132] W. Liu, P. P. Pokharel, and J. C. Principe, “Correntropy: Properties and applications in non-Gaussian signal processing,” *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5286–5298, Nov. 2007.
[133] M. Iliadis, L. Spinoulas, A. S. Berahas, H. Wang, and A. K. Katsaggelos, “Multi-model robust error correction for face recognition,” in *Proc. Int. Conf. Image Process.*, pp. 3229–3233, Sep. 2016.
[134] M. Liadis, H. Wang, R. Molina, and A. K. Katsaggelos, “Robust and low-rank representation for fast face identification with occlusions,” *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2203–2218, May 2017.
[135] J. Xie, J. Yang, J. Qian, Y. Tai, and H. Zhang, “Robust nuclear norm based matrix regression with applications to robust face recognition,” *IEEE Trans. Image Process*, vol. 26, no. 5, pp. 2286–2295, May 2017.
[136] 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.
[137] J. Yang, L. Luo, J. Qian, Y. Tai, F. Zhang, and Y. Xu, “Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 1, pp. 156–171, Jan. 2017.
[138] Z. Zhang, Y. Xu, L. Shao, and J. Yang, “Discriminative block-diagonal representation learning for image recognition,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 7, pp. 3111–3125, Jul. 2018.
[139] A. Y. Yang, Z. Zhou, A. G. Balasubramanian, S. S. Sastry, and Y. Ma, “Fast l_1-minimization algorithms for robust face recognition,” *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3234–3246, Aug. 2013.
[140] J. Yang and Y. Zhang, “Alternating direction algorithms for 1-problems in compressive sensing,” *SIAM J. Sci. Comput.*, vol. 33, no. 1, pp. 250–278, 2011.
[141] I. Daubechies, M. Defrise, and C. De Mol, “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” *Commun. Pure Appl. Math.*, vol. 57, no. 11,

pp. 1413–1457, Nov. 2004.
[142] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interiorpoint method for large-scale l_1-regularized least squares," *IEEE J. Sel. Topics Signal Process*, vol. 1, no. 4, pp. 606–617, Dec. 2007.
[143] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
[144] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. B, Methodol*., vol. 58, no. 1, pp. 267–288, 1996.
[145] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust Face Recognition via Sparse Representation," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 31, no. 2, pp. 210–227, Feb. 2009.
[146] J. Huang, F. Nie, H. Huang, and C. Ding, "Supervised and projected sparse coding for image classification," in *Proc. 27th AAAI Conf. Artif. Intell*., pp. 438–444, Jun. 2013.
[147] Z. Fan, M. Ni, Q. Zhu, and E. Liu, "Weighted sparse representation for face recognition," *Neurocomputing*, vol. 151, pp. 304–309, Mar. 2015.
[148] X. Tang, G. Feng, and J. Cai, "Weighted group sparse representation for undersampled face recognition," *Neurocomputing*, vol. 145, pp. 402–415, Dec. 2014.
[149] W. Zuo, D. Meng, L. Zhang, X. Feng, and D. Zhang, "A generalized iterated shrinkage algorithm for non-convex sparse coding", in *Proc. IEEE Int. Conf. Comput. Vis*., pp. 217–224, Dec. 2013.
[150] J. Yang, L. Luo, J. Qian, Y. Tai, F. Zhang, and Y. Xu, "Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 39, no. 1, pp. 156–171, Jan. 2017.
[151] R. He, W.-S. Zheng, T. Tan, and Z. Sun, "Half-quadratic-based iterative minimization for robust sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 36, no. 2, pp. 261–275, Feb. 2014.
[152] L. Zhang, M. Yang, and X. C. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *Proc. IEEE Int. Conf. Comput. Vis*., Nov. 2011, pp. 471–478.
[153] P. Huber, Robust Statistics. New York: Wiley, 1981.
[154] Z. Y. Zhang, "Parameter estimation techniques: A tutorial with application to conic fitting," *Image Vis. Comput.*, vol. 15, no. 1, pp. 59–76, 1997.
[155] Liu, X., *et al.*, "Thermal Infrared Imaging-Based Vein Recognition", *PLOS ONE*, vol. 9, no. 4, e94914, 2014
[156] Zhang, L., *et al.*, "Thermal Infrared Imaging for Vein Pattern Biometrics: A Review", *Sensors*, vol. 17, no. 8, 1899.
[157] Wang, Z., *et al.*, "Deep Learning-Based Vein Recognition Using Thermal Infrared Images", *IEEE Transactions on Circuits and Systems II*: Express Briefs, vol. 67, no. 2, pp. 396-400, 2019.

[158] Huang, K., *et al.*, "Infrared Laser-Based Thermal Imaging for Vein Pattern Recognition", *IEEE Access*, vol. 9, pp. 13470-13480, 2021

[159] K. Jing, X. Zhang, and X. Xu, "Double-Laplacian Mixture-Error Model-Based Supervised Group-Sparse Coding for Robust Palmprint Recognition."

[160] H. Zhang, Q. M. J. Wu, and T. M. Nguyen, "A Robust Fuzzy Algorithm Based on Student's t-Distribution and Mean Template for Image Segmentation Application," vol. 20, no. 2, Feb. 2013.

[161] Y. Huang, Y. Zhang, Y. Zhao, and J. A. Chambers, "A Novel Robust Gaussian–Student's t Mixture Distribution Based Kalman Filter," vol. 67, no. 13, Jul. 1, 2019.

THE

END