Quality Assessment of Degraded Natural Images due to Changes in Weather Conditions

Thesis submitted by

RATNADEEP DEY

DOCTOR OF PHILOSOPHY (Engineering)

Department of Computer Science and Engineering
Faculty Council of Engineering & Technology
Jadavpur University,
Kolkata-700032, India
2024

JADAVPUR UNIVERSITY
KOLKATA-700032, INDIA

INDEX NO.:87/19/E

1. TITLE OF THE THESIS:

Quality Assessment of Degraded Natural Images due to Changes in Weather Conditions

2. NAME, DESIGNATION, INSTITUTION OF THE SUPERVISORS:

Dr. Debotosh Bhattacharjee

PROFESSOR,

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING,
JADAVPUR UNIVERSITY, KOLKATA-700032

3. LIST OF PUBLICATIONS:

a) JOURNAL:

i. Bera, A., Dey, R., Bhattacharjee, D. et al. "Spoofing detection on hand images using quality assessment". Multimed Tools Appl 80, 28603–28626 (2021). https://doi.org/10.1007/s11042-021-10976-z

ii. Dey, R., Bhattacharjee, D. & Krejcar, O. "Detection of images degraded by rain using image quality assessment". Multimed Tools Appl (2022). https://doi.org/10.1007/s11042-022-13041-5

iii. Dey, R., Bhattacharjee, D. "No-reference image blurriness assessment using divisive normalization". SIViP (2022). https://doi.org/10.1007/s11760-022-02179-2

iv. Dey, R., Bhattacharjee, D, "OIVID: A Novel Image Descriptor for Full-Reference Image Assessment". Communicated

b) CONFERENCE:

i. Ratnadeep Dey, Debotosh Bhattacharjee, Mita Nasipuri, "Image De-noising using Generative Adversarial Network", Advances in Intelligent Systems and Computing, Springer.

ii. Ratnadeep Dey, Debotosh Bhattacharjee, "Single Image De-raining Using GAN for Accurate Video Surveillance," Intelligence Enabled Research, Springer, Singapore, pp. 7-11, 2020.

iii. Ratnadeep Dey, Debotosh Bhattacharjee, Mita Nasipuri, "Object Detection in Rainy Condition from Video Using YOLO Based Deep Learning Model", Advanced Computing and Systems for Security, Springer, Singapore, pp. 121-131, 2020.

iv. Dey, R., Bhattacharjee, D., Kollmann, C., Krejcar, O. (2023). Classification of Breast Tumor from Ultrasound Images Using No-Reference Image Quality Assessment. In: Basu, S., Kole, D.K., Maji, A.K., Plewczynski, D., Bhattacharjee, D. (eds) Proceedings of International Conference on Frontiers in Computing and Systems. Lecture Notes in Networks and Systems, vol 404. Springer, Singapore. https://doi.org/10.1007/978-981-19-0105-833

4. LIST OF PRESENTATIONS:

i. Paper name - "Single Image De-raining" using GAN for accurate video surveillance" in First Doctoral Symposium on Intelligence Enabled Research (DoSIER) on 19th and 20th October, 2019

ii. Paper name - "Object Detection in Rainy Condition from Video Using YOLO Based Deep Learning Model" in 7th ACSS Doctoral Symposium on 28th and 29th February, 2020

PROFORMA-1

"Statement of Originality"

I RATNADEEP DEY, registered on 29/05/2019, do hereby declare that this thesis entitl
"Quality Assessment of Degraded Natural Images due to Changes in Weather Conditions" contai
a literature survey and original research work done by the undersigned candidate as part
Doctoral studies.

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

I also declare that I have checked this thesis as per the "Policy on Anti Plagiarism, Jadavp
University,2019",and the level of similarity checked by iThenticate software is 6%.

Signature of Candidate:

Date: 09/05/2024

Certified by Supervisor(s):

(Signature with date, seal)

09/05/2024

Professor
Computer Sc. & Engg. Department
Jadavpur University
Kolkata-700032
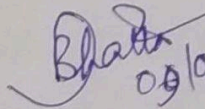
1.

CERTIFICATE FROM THE SUPERVISOR/S

This is to certify that the thesis entitled "Quality Assessment of Degraded Natural Images due to Changes in Weather Conditions" submitted by Ratnadeep Dey, who got his name registered on 29/05/2019 for the award of Ph.D.(Engg.) degree of Jadavpur University is absolutely based upon his own work under the supervision of Prof. Debotosh Bhattacharjee (Department of Computer Science Engineering, Jadavpur University) and that neither his thesis nor any part of the thesis has been submitted for any degree/diploma or any other academic award anywhere before.

Bhatta 09/05/2024

_____

Signature of the Supervisor
and date with Office Seal

Professor
Computer Sc. & Engg. Department
Jadavpur University
Kolkata-700032

Dedicated to my Family

Acknowledgements

I am deeply indebted to many people who have contributed to making the completion of this thesis work possible. Firstly, I would like to express my earnest gratitude and respect to my supervisor, **Prof. Debotosh Bhattacharjee,** for his relentless guidance, tremendous support, endless care, invaluable suggestions, and constant monitoring during this work.

I would also like to express my sincere respect to Prof. Mita Nasipuri, Former Coordinator of the Centre for Microprocessor Application for Training, Education and Research (CMATER) Laboratory, Jadavpur University (JU), Prof. Mahantapas Kundu, Coordinator of CMATER Laboratory, JU, Prof. Ram Sarkar, Prof. Subhadip Basu and Prof. Nibaran Das of Department of Computer Science and Engineering, Jadavpur University, for providing me constant inspiration, support and useful suggestions during the course of this thesis work.

I would like to thank Prof. Nandini Mukhopadhyay, HOD, Department of Computer Science and Engineering, Jadavpur University, for making all the departmental facilities available and for her valuable comments.

I am grateful to my fellow researchers and lab mates Dr. Sourav Pramanik, Dr. Kaushiki Roy, Dr. Koushik Dutta, Dr. Nirmal Das, Dr. Neelotpal Chakraborty, Dr. Samir Malakar, Dr. Anup Kumar Halder, Dr. Soumyendu Sekhar Bandyopadhyay and Dr. Soumi Pal of CMATER Laboratory, Jadavpur University.

Finally, I owe my encompassing debt to my parents, my wife and some very close friends whose love, support, encouragement, good wishes, and inspiration enabled me to complete this thesis.

Date: 09/05/2024

Place: Kolkata

Ratnadeep Dey.

(Ratnadeep Dey)

Contents

LIST OF FIGURES

LIST OF TABLES

Chapter 1:

Introduction

Image Quality Assessment is the field of research where the quality of an image has been estimated. This is an essential part of image processing research. Accessing the quality of the input image is mandatory for obtaining the best possible processing outcome. This thesis concentrates on the research field, namely image quality assessment. It contributes novel image quality assessment matrices and some image quality assessment-based application systems to solve problems from different domains like healthcare, industrial automation, surveillance systems, and weather classification.

In computer vision, scenes have been captured with image and video sensors. The captured scenes are stored in computer memory as images and videos, further analyzed like our human visual system. However, some noise can distort the captured image and video frames during capture. Processing of the distorted image is not suitable for further processing. Therefore, an assessment of the captured image before processing is required. Various degradation models can distort images. The degradation models of the distorted image have been discussed first. The degradation of images can be caused by two factors: technical and environmental. Both types of degradation models are discussed in this chapter. After discussing the degradation model, the need for quality assessment of input images was thoroughly discussed. The research domain of Image Quality Assessment has been introduced, followed by the discussion of the image degradation model and the need for quality assessment. The process of Image Quality Assessment has been discussed here. After these discussions, the scope of the dissertation has been stated. Finally, the organization of the whole thesis has been enunciated.

1.1 – Image Degradation Model

In the era of rapidly growing technology, images are essential. Many real-life problems can be solved by image analysis. Image analysis can quickly solve problems with medical diagnosis, security issues, weather forecasts, etc. Images contain much information, which helps draw essential conclusions. According to these conclusions, problems can be solved very quickly and efficiently. Therefore, input images must be treated appropriately before being used for the abovementioned problems. If the input images are erroneous, then their analyses become wrong. Thus, input image quality is essential. The quality of the input

images can be decreased for several reasons. Image degradation can be done in two ways - Degradation of Image due to technical reasons and Degradation of Image due to Environmental Stimuli.

1.1.1 – Degradation of Image due to technical reasons

There are different types of distortions present. Those are - blurring, impulse noise, Additive Gaussian Noise, Special Correlation noise, quantization noise, distortions due to compression, chromatic aberration, etc. Blurring is the most common type of distortion in those types of distortions. An image can be blurred for many reasons – defocusing of the camera, movement of the acquisition device, high-intensity light, etc. Different types of distortion degrade input images in various ways. For example, when an image is blurred, the variation in intensity values will be low, and objects in the image cannot be adequately separable from the backgrounds. Therefore, the information in the image goes very low. Analyses of those blurred images may lead to erroneous conclusions. Thus, in the case of analyzing an image, it is essential to check whether the image is blurred or not, and an image quality metric that can measure the blurriness of an image is beneficial in this regard.

Some of the technical reasons for which images can be degraded are as follows –

(i) **Noise due to acquisition system** – Image quality can be reduced due to acquisition systems. The X-ray images encountered Poisson noise; Speckle noise is observed in the case of ultrasound images.

(ii) **Quality degradation due to transmission** – The images are captured using acquisition devices. Then, the images are transmitted for further processing. Transmission channels can add noise. Images are compressed at the time of transmission. This compression results in some loss of detail.

(iii) **Defocusing of the camera** – Image quality can be degraded due to defocusing of the image capturing unit. It causes blurriness and low contrast in images. In the case of microscopic images, focusing is very important. The objects in microscopic images are very tiny. Object detection becomes very tough if the focus is not correct.

1.1.2 – Degradation of Image due to Environmental Stimuli

Other than technical reasons, images can be degraded by environmental stimuli like rain, fog, mist, snow, etc. According to physical properties and effects, the weather conditions can be categorized into two categories – (i) steady and (ii) dynamics. Fog, Haze, and Mist fall in the category of steady. In this case, the sizes of water droplets are small and vary between 1-10 μm. The droplets cannot be seen individually through human eyes. However, the presence of tiny droplets changes the intensity of the image. Many models are proposed to describe those effects in an image. In the case of dynamic weather conditions, the size of the droplets present in the air is between 0.1 – 10 mm. Therefore, the droplets can be seen through cameras. Rain and Snow come into the dynamic weather degradation category. In this case, the droplets come down with high velocity with complex trajectories. Motion blur is a widespread effect in this situation. In the case of steady weather degradation, the vision is affected by the aggregated impacts of many droplets. Therefore, modeling such weather conditions is comparatively easy. On the other hand, modeling dynamic weather conditions is complicated because each droplet changes vision with its different velocity and trajectory.

In a country like India, where rain is widespread rather than snow. Therefore, this thesis focuses on the images affected by rain. Each raindrop is spherical in shape. They work as a lens. The lights are refracted and reflected through the droplets, making sharp intensity patterns in the image. The raindrops are in motion in times of rain. Depending on the velocity, the shape of the droplets changes. According to Beard and Chuang [1], the shape of the raindrop can be defined as

$$r(\theta) = a\left(1 + \sum_{n=1}^{10} c_n \cos(n\theta)\right) \qquad (1.1)$$

where a = undistorted shape, c_n = coefficients depending on the radius, and Θ = polar elevation angle. The changes in a droplet shape can observe different types of distortions. The distortions caused by rain are categorized in two ways – (i) the presence of rain strikes. This can happen due to the Spatial-temporal correlation of many falling water droplets on the image plane and (ii) the visual effect of rain. The intensity of image pixels altered due to the droplets.

Images distorted by technical reasons or environmental stimuli need special attention before further processing. The assessment of the quality of images is very much needed, and the need for quality assessment will be discussed thoroughly in the next section.

1.2 – The need for Quality assessment

In the previous subsection, different image degradation models have been discussed. Some of them occur due to technical reasons, and environmental stimuli degrade some of them. Captured images or video frames degraded by those degradation models are hard to analyze, and there are chances of misinterpretation of analysis. Therefore, a quality assessment of those images is mandatory before processing any input images and video frames. Computer vision-based research performs difficult and sensitive applications where misinterpreted image and video frames affect the system poorly. That may lead to casualties, loss of resources, misleading business goals, etc. Image quality assessment is the only way to handle these losses in computer vision-based applications. Some of the very prominent scenarios can be video surveillance systems. It is well known that the successful implementation of video surveillance systems helps to reduce the chance of terrorist attacks in public areas. Terrorist attacks in a public area can be a reason for a vast number of causalities. Therefore, implemented video surveillance systems must be robust, accurate, and efficient. A small interpretation of video frames captured by surveillance cameras can lead to miseries. At the same time, surveillance cameras are generally implemented in open areas. For that reason, there is a massive chance of the captured scenes being affected by different environmental stimuli like rain, fog, mist, dust, snow, etc. Those environmental stimuli affect the surveillance video frames tremendously. If those affected video frames have been analyzed, there is a considerable chance of mistakes. Those mistakes can be very costly. Therefore, before processing surveillance video frames, quality assessment is mandatory. The introduction of image quality assessment must be very fruitful in analyzing surveillance video frames efficiently. A similar scenario can be found in the field of healthcare research. Computer vision-based disease diagnosis system is an exciting and respectful research domain. In this research domain, images of human organs, tissues, cells, etc., captured using different imaging modalities can be analyzed to diagnose affected regions. Various imaging modalities like X-ray, ultrasound, MRI, CT Scan, and optical image sensors have been used for capturing those images, and every imaging modality produces some technical errors in

capturing images discussed in subsection 1.1.1. Like the video surveillance system, analyzing distorted images must lead to the misdiagnosis of diseases of human beings, and this misinterpretation may be the reason for the loss of human life. The application of image quality assessment in healthcare research is appreciated in developing an efficient disease diagnosis system.

Another important application of the quality assessment process is identifying visual disputes. Everyday scenes can be disturbed due to different stimuli. The method of application can identify the presence of a stimulus. For example, natural scenes can be distorted by rain strikes. Quality assessment of the rainy image helps to identify the presence of rain in the natural scenes. Therefore, rainy images can be identified using image quality assessment. In this case, the image quality metrics can be utilized as features. Specific image quality metrics can be developed to identify rainy images. The quality assessment process in the video surveillance system also helps identify rainy video frames. Therefore, identifying rainy situations can be identified efficiently.

Similarly, in the case of medical image processing, quality assessment can be helpful in the diagnosis of various diseases. In the case of breast tumor detection, it is observed that the presence of the tumor in the breast region affects the scene of normal breast scenes captured using ultrasound image modality. Quality assessment of the captured scenes helps to identify the presence of breast tumors. Therefore, it can be said that quality assessment is vital in medical image processing to diagnose several diseases in the human body. Quality assessment has an excellent impact on the heavy electrical and electronics industries. In these industries, the quality of the product can be tested using artificial intelligence-based computer vision techniques. In the steel industry, different faults are found in the product. These faults can be identified efficiently by applying quality assessment to the captured images of the final product. There are several other possible areas in the industry where the quality assessment of images becomes the most acceptable technique for solving real-world challenges.

Quality assessment is needed for image processing research to obtain the quality of the input image. Before processing images and video frames, an assessment of quality is recommended to achieve the best processing results. On the other hand, quality assessment is beneficial in identifying natural stimuli like rain, fog, mist, and snow in the natural scene. Along with that, quality assessment can be helpful in the diagnosis of several diseases, and in heavy industries,

quality assessment is required to ensure the best quality product. The vast area of applications makes quality assessment very necessary for researchers.

1.3 – Image Quality Assessment

According to section 1.2, quality assessment of captured scenes is highly needed. The captured scenes are called images, and assessing the quality of the images is the process called image quality assessment. Images are the input of this process, and measured values are the output. Image quality assessment research mainly creates metrics that can help assess the quality of images. Videos can be evaluated similarly. Videos are a collection of frames, and each frame can be considered an image. Therefore, assessing the video is similar to determining the quality of a bunch of images. The quality metrics designed for images can help assess the quality of videos. There are two main types of image quality assessment processes. The subjective image quality assessment and objective image quality assessment are discussed in the following subsections.

1.3.1 – Subjective Image Quality Assessment

The subjective quality metrics [2] are obtained manually. In this case, human beings are appointed to judge images' quality and are asked to rate input images. Image quality metrics are obtained based on the ratings of the referees. This process is very costly and takes a considerable amount of time. Though this metric type is very appropriate, this approach is unpopular. This is a manual process. Human participants have measured the quality of images. Human intelligence has been utilized in this process. The image quality databases have been developed using this process. However, this quality assessment process is not in this thesis's scope.

1.3.2 – Objective Image Quality Assessment

The object image quality assessment [3] uses computing devices. Image quality assessment is the process of quality assessment. The objective image quality assessment refers to creating algorithms for computing devices to predict the perceived quality of the visual scenes. The perceived quality is mainly the human perception of visual scenes. The human visual system is the ultimate predictor of visual scenes. The process of objective image quality assessment is the process that helps computing devices achieve the capability of human visual systems. The main goal of this process is to assess the quality of visual scenes similar to human perception. Different algorithms have been designed for this goal. The algorithms have been

evaluated by comparing their prediction with human performance. The visual scenes can be captured and stored to create images. Algorithms have been designed to assess the quality of the image. The quality score obtained by the object image quality assessment process has been compared with the human-predicted quality score. All image quality datasets provide a subjective quality score along with the images. The subjective scores are known as the Mean Opinion Score (MOS). The correlation between the predicted quality scores and the corresponding MOS value must be calculated to evaluate the image quality assessment algorithm. Different types of standard correlation measures, like Spearman's rank-ordered correlation coefficient (SROCC), the linear correlation coefficient(LCC), root-mean-squared error (RMSE), and outlier ratio (OR), have been used. The quality assessment algorithm becomes best when the obtained quality score highly correlates with the MOS value of the image.

The objective quality metric is further classified into three types – Full Reference, Reduce Reference, and No-Reference. In the entire reference approach, the original image, and the test image, are both available, and the quality of the test image can be calculated by comparing both images. In reduced reference, the original image is not present. Instead of the original image, some features of the original image are present. These original image features are compared with the same features obtained from the test image, and image quality metrics are computed. In No reference image quality metric analysis, original images, or any features of original images are not present. The quality assessment is done entirely on the test image itself. It can be easily concluded that the three types of objective quality analysis approach and no-reference image quality assessment approach are most relevant in real-time problems because an original image or any features of the original image are not always available.

Much research has been done on the three objective image quality assessment approaches. The research works in the literature have been discussed in chapter 2. This thesis addresses the objective image quality assessment, and the scope of this thesis is discussed in the next section of this chapter.

1.4 – Scope of the Dissertation

This dissertation mainly addressed the research domain of image quality assessment. The main goal of the dissertation is to provide algorithms to assess the quality of degraded natural images due to weather degradation. This dissertation achieves this goal by proposing novel algorithms to assess the quality of images degraded by natural stimuli. However, the

dissertation discusses the other aspects of image quality assessment research and research on different segments of image processing, like medical image processing, image de-raining, and object identification. The complete scope of the dissertation has been discussed in this section.

The dissertation has a broad scope and various orientations. Image quality assessment of natural images degraded due to weather degradation is the main focus of this dissertation. Other sub-areas of image quality assessment and different sub-domains of image processing have been addressed about the key topic. Discussing different orientations may help express the overall scope of the dissertation. The dissertations to be discussed are contribution to research, application orientation, technological orientation, and experimental data.

Research Contribution –

The scope of the dissertation in the orientation of research contribution has to be addressed. Mainly, the research areas where the dissertation contains novel contributions have been enlisted below –

- The main research area addressed here is objective image quality assessment, also known as image quality assessment. In section 1.3.2, it has been stated that three techniques are available in the literature to assess the quality of an image. They are – Full reference image quality assessment, Reduced reference image quality assessment, and no reference image quality assessment. This dissertation's research contribution is limited to full-reference and no-reference image quality assessment. A reduced reference image quality assessment has been excluded from the research contribution.
- This dissertation mainly assesses natural images degraded by weather degradation—different environmental stimuli like fog, rain, snow, dust, mist, etc. Rainy image assessment is the main focus area of this dissertation.
- De-raining is another scope of this dissertation, where rain strikes have been removed from the rainy images to retrieve the actual scene.
- Object Detection is another research area addressed in this dissertation.
- The video surveillance system is also incorporated here.
- Medical image processing is another area of research that the dissertation claims to make a novel contribution to.

Application-based orientation –

The research contributions addressed in this dissertation have an orientation toward application areas. The application areas, which have been demonstrated using the novel research contribution, are as follows –

- A rainy image detection system has been implemented using a no-reference image quality assessment approach and evaluated with the publicly available dataset.
- A weather classification system has been implemented using full reference image quality assessment and experimented to check the efficiency.
- A blurriness assessment system from the input image using no-reference image quality assessment has been developed.
- An application system to protect against spoofing attacks on hand biometry has been implemented using a no-reference image quality assessment approach.
- The detection of a faulty product in a steel manufacturing company was fabricated using a full-reference image quality assessment approach.
- A System for object detection from rainy scenes captured through a video surveillance camera has been implemented.
- An application system has been built to remove rain strikes from rainy images.

Technological Orientation –

This dissertation's research contributions and application systems have been fabricated using different technological approaches. These approaches are enlisted as follows –

- Statistical analysis of different degrees has been used to implement the different contributions presented in this dissertation.
- Feature engineering is another process used to create applications. A novel feature extraction method has been contributed to this dissertation.
- Standard machine learning algorithms have been used here to develop several application systems for this dissertation.
- A deep learning model, You Only Look Once (YOLO), has been used to detect objects from rainy situations captured through a video surveillance camera.
- Another deep learning model, the Generative Adversarial Network (GAN), has been used to implement the application system to remove rain strikes from rainy images.

Orientation towards experimental data –

This dissertation incorporates several research contributions and application-based systems evaluated using different types of images and video data. The various data modalities used in this dissertation are as follows –

- Natural Images are the primary data modalities used for the experiments. Natural images are the data containing details of natural scenes captured using optical image sensors.
- Natural images degraded by weather conditions are another data modality utilized in different experiments stated in the dissertation.
- Surveillance video is another data modality addressed in this dissertation.
- Hand biometric data has been used in the experiment related to the application system implementation to protect against spoofing attacks.
- Ultrasound imaging has been needed for fabricating medical image processing-based research.
- X-ray imaging has been used in the experiment on fault identification in the steel manufacturing industry.

It has been seen that the current dissertation has a broad scope with four prominent orientations. The main contributions of this dissertation have been stated in the following section.

1.5 – Main Contribution

The dissertation's complete scope has been discussed in the previous subsection. From the elaborated discussion, the main contributions of this dissertation have been pointed out in this section. The overall contribution of this dissertation can be classified into two segments – the first one is the contribution to the literature on image processing, and the second part of the contribution to the application areas of image processing. Contributions to the literature on image processing have been enlisted as follows –

1. No reference image quality assessment metric has been designed to assess the quality of rainy images.

2. Novel metric named No-Reference Image Blurriness Estimation Metric (NIBEM)

3. Novel Spatial feature-based feature extractor algorithm coined as Orientate Intensity Velocity feature extractor (OIVFE)

4. Orientate Intensity Velocity Full Reference Image Quality (OIVFRIQ) metric

These contributions have been discussed throughout the dissertation. The organization of the dissertation has been discussed in the next section.

1.6 – Organization of the Dissertation

The current chapter introduces the dissertation. This chapter presents the background of the image quality assessment research, including a discussion on the image degradation model, the need for quality assessment, and the process of image quality assessment-based research. The scope of the dissertation and main contributions have been included here. The rest of the dissertation has been organized into six more chapters. Chapter 2 discusses the related research works available in the literature. Chapter 3 discusses a very elegant sub-area of image quality assessment research: Full reference image quality assessment. In this chapter, a novel entire reference image quality assessment algorithm has been discussed. The algorithm has been designed to assess the quality of natural images degraded by environmental stimuli. However, the algorithm's applicability has been proved by applying this algorithm in areas like medical image processing and industrial automation. Related experimental results have been included in the chapter. Chapter 4 addresses another sub-area of image quality assessment: no reference image quality assessment. A novel image quality assessment algorithm concentrating on blur-type distortion has been described here. Chapter 5 mainly focuses on the quality assessment of rainy images. A novel algorithm for rainy image quality assessment has been described in this chapter. Additionally, two more research areas related to the rainy image have been addressed in this chapter. They are de-raining rainy images and object detection in rainy images. Chapter 6 presents research on applying image quality assessment other than the natural image degraded by weather degradation. A novel system to protect against spoofing attacks in the field of hand biometrics using image quality assessment has been addressed in this chapter. In addition, the application of image quality assessment in healthcare has contributed. Chapter 7 concludes the dissertation by discussing the summary of the dissertation and rendering the future scope of the dissertation.

Chapter 2:
Literature Survey

This chapter discusses research in the field of image quality assessment found in the literature. The research works in the field of image quality assessment can be classified into three classes – Full reference Image Quality Assessment (FRIQA), Reduced Reference Image Quality Assessment (RRIQA), and No-Reference Image Quality Assessment (NRIQA). The classification tree of the objective image quality assessment research is represented in Figure 2.1. These three classes are further divided into subclasses depending on the methodology used to create the algorithms. State-of-the-art research works belonging to each class and subclasses have been discussed accordingly.
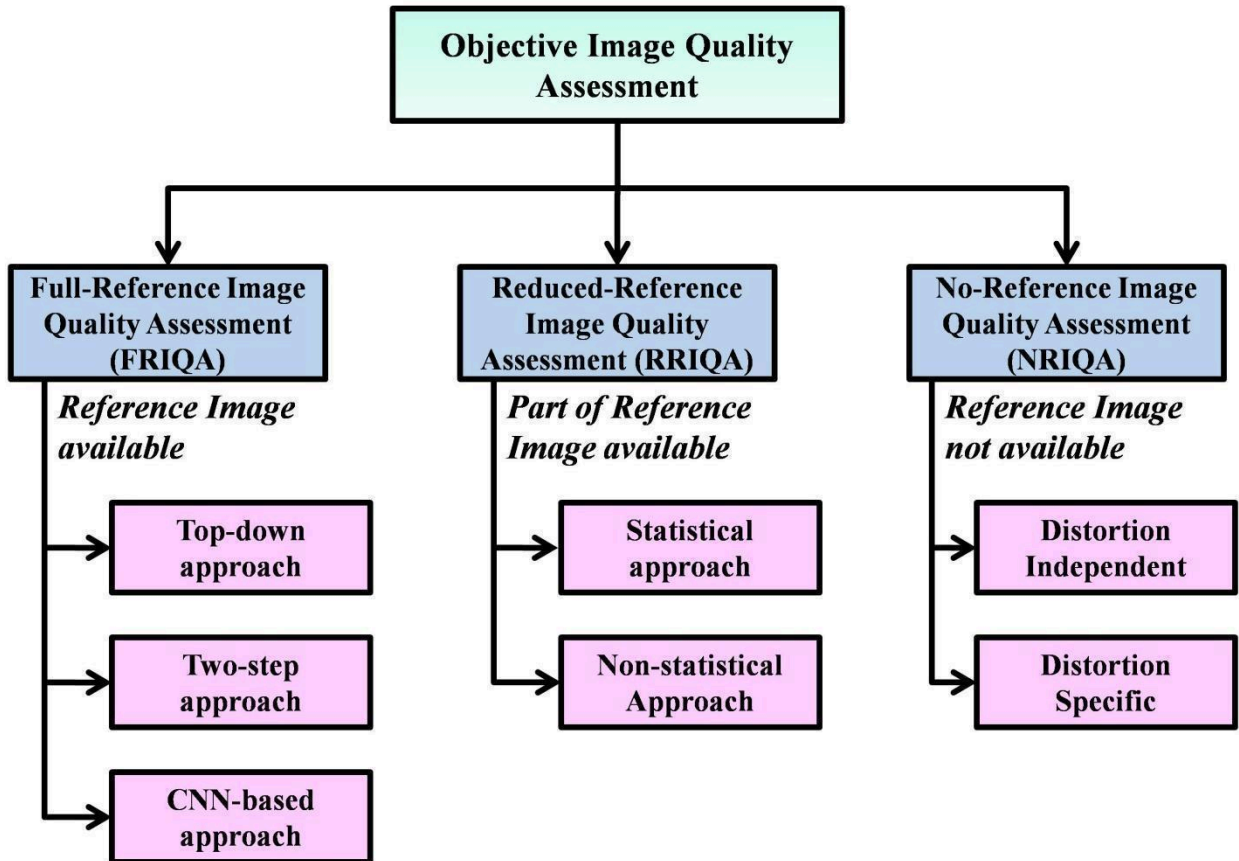
Figure 2.1: Classification Tree of Objective Image Quality Assessment Research

2.1 – Full Reference Image Quality Assessment

Full reference image quality assessment (FRIQA) compares the test image with the reference image. In image 2.2, the process of FRIQA has been visualized.

Figure 2.2: Overview of the process of Full-reference Image Quality Assessment

According to this process, the test image's quality must be analyzed by comparing it with the reference image. Full reference quality assessment is a primitive type of image quality analysis. The research in image quality assessment has begun with the full-reference approach. In this approach, the test image is compared to the corresponding reference image

to determine the quality of the test image. The preliminary idea in the full-reference image quality assessment research was to compare the two images pixel by pixel. According to this approach, evaluation metrics Mean Square Error (MSE) [4], Root Mean Square Error (RMSE) [5], Peak Signal to Noise Ratio (PSNR) [6], etc, from signal processing are used to compare the test image with the reference image. The approach, however, failed to produce the desired outcome. The main reason behind the failure of the approach is that the approach cannot emphasize human perception. The Human Visual System is the optimal evaluator of image quality. Therefore, human perception should have been included in the research of the objective image quality assessment [7].

Several full-reference image quality metrics, including human perception, are available in the literature. Universal Image Quality Index [8] was the first image quality metric proposed according to this idea. This metric included – luminance, contrast, and structural distortions in methodology. This metric is independent of the viewing condition and the observer's perception. The metric was modified to generate a well-known full reference image quality Structural Similarity Index metric (SSIM) [9]. The SSIM mimicked the functionality of the human visual system (HVS) for the first time. This was a milestone in this field of research. The visual quality of the test image was analyzed using the SSIM metric, which calculated the changes in contrast and luminance along with the structural information of the image. The metric was modified a lot to achieve more accuracy. Some simple modifications on the primary SSIM metric resulted from metrics like three-component SSIM [10]. This metric was generated based on the observation that the human visual system (HVS) can find differences between the texture and smooth areas. This metric concentrates on three features of an image – edge, smooth region, and texture. The other two modified versions of SSIM are simplified fast SSIM [11] and gradient-based SSIM [12]. The SSIM metric was also established in different domains. Some examples of such research work are Information-Weighted SSIM (IWSSIM) [13], Edge Strength Similarity (ESSIM) [14], Complex Wavelet SSIM (CWSSIM) [15], and Multi-Scale SSIM (MSSSIM) [16], etc. Other than the SSIM, another famous metric is Feature Similarity (FSIM) [17]. The metric compares the phase congruency and gradient magnitude of the test image along with the reference image to assess the quality of the test image. The Structural Texture Similarity Metric (STSIM) [18] finds texture similarity between two images. Other than the similarity measures, Visual Information Fidelity [19,20] based full reference metrics were proposed. Singular Value Decomposition (SVD) [21] was another approach in full-reference image quality assessment where the Euclidian distance

between the singular values of the 8 × 8 block of the test image and the corresponding reference image was calculated. The above-mentioned Full reference Image Quality metric (FRIQM) development approach can be coined as a top-down approach, where the complete HVS functionality has to be modeled using some global understanding.

There is another type of full-reference approach where the image quality is assessed in two steps. In the first step, quality maps of the test image and the corresponding reference image are calculated locally. Then, in the next step, the quality score is calculated from the local quality measures using a specific pooling strategy. Different approaches, like content, gradient, and illumination-based, were used to measure the local quality of the test image and the corresponding reference image. Then, several pooling strategies, like average pooling, standard deviation-based pooling, weighted average pooling, etc, were proposed in the literature. The GMSD [22] metric is one of the best FRIQA algorithms under the two-step approach. This work employs Gradient Magnitude Similarity (GMS) to find pixel-by-pixel differences between the reference and test images. After that, the comparison result was pooled using a novel method named the standard deviation of the GMS map. Other metrics [23-28] were also present and developed using the two-step approach. A genetic programming-based work [29] was proposed, which follows the two-step approach. This work was based on SSIM, and this handcrafted metric was formulated using genetic programming. Z. Tang [30] proposed another research work under the two-step approach. In the first step, features from the spatial and frequency domain were extracted, and in the second step, the random forest was used for pooling. Another metric, visual saliency with color appearance and gradient similarity (VCGS) [31], was proposed under the two-step approach. This work used color appearance similarity, chrominance similarity, and gradient similarity as features.

Another approach of FRIQA is the CNN-based approach, where Convolutional Neural Networks (CNN) have been used to extract features from the reference image and test image. JND-SalCAR [32] was proposed to extract the features related to the perception of HVS using the Saliency-Channel Attention Residual Network. This work was based on the Just Noticeable Difference (JND) feature of the human visual system. Another CNN-based work was proposed by W. Kim et al. [33]. In this work, the quality of the test image was measured using the representation of dynamic error and pooling of the errors using visual sensitivity-based quality. Another Full Convolutional Network-based work, SPSIM [34], was

proposed to measure the quality of Screen content images. The screen content image is a combination of text and pictures. In another work [35], a long-range dependencies model based on FRIQA was proposed, where a deep neural network was employed to highlight the long-range dependencies between the local artifacts of the test image and high-order spatial pooling applied to the image. This approach enhances the generalizability of the FRIQA algorithm.

Three categories of FRIQA algorithms have been discussed here, and research works under these categories are enlisted in Table 2.1. However, the main drawback of the FRIQA is the dependency on the reference images. All algorithms of this type are irrelevant due to the absence of the reference image. Therefore, the applicability of the FRIQA is very low because the reference image is not available for all real-life applications.

Table 2.1: Overview of State-of-the-art FRIQA methods

Name of the metric/ algorithm	Methodology	Category
Universal Image Quality Index [8]	Distortions can be modeled based on the lack of correlation distortion due to illumination and contrast.	*Top-down approach*
Structural Similarity Index metric (SSIM) [9]	Structural similarity between the test and reference images was computed to create a final metric score.	*Top-down approach*
Three-component SSIM [10]	Three specific structures – edge, texture, smooth region - of the test image and reference image were calculated using this approach.	*Top-down approach*
Simplified fast SSIM [11]	This was a fast and less complex method modifying the basic SSIM	*Top-down approach*
Gradient-based SSIM [12]	This modified version of SSIM includes gradient computations in similarity measures and is useful for handling blurry images.	*Top-down approach*

Information Weighted SSIM (IWSSIM) [13]	In this work, a novel information content-weighted pooling was introduced	*Top-down approach*
Edge Strength Similarity (ESSIM) [14]	This work was based on the edge strength of the test image and reference image.	*Top-down approach*
Complex Wavelet SSIM (CWSSIM) [15]	This work was done in the wavelet domain. SSIM was developed in the wavelet domain	*Top-down approach*
Multi Scale SSIM (MSSSIM) [16]	Structural Similarity between the test and reference images was calculated using multiple scales. A low-pass filter with factor 2 was used to scale the images.	*Top-down approach*
Feature Similarity (FSIM) [17]	This work was different from the SSIM method. In this work, two features, namely Phase congruency and Gradient Magnitude, were extracted from test and reference images for comparison.	*Top-down approach*
Structural Texture Similarity Metric (STSIM) [18]	The texture of the test image and reference image was compared to generate the metric. Texture similarity was the main focus of this work.	*Top-down approach*
Visual Information Fidelity [19]	Natural Scene Statistics IQA used Visual Information Fidelity to compare the test and reference images.	*Top-down approach*
Information Fidelity Criterion [20]	Loss of image information in the distorted image was the main concept of this work. Information Fidelity Criteria were calculated from the test image and reference image.	*Top-down approach*

Singular Value Decomposition (SVD) [21]	This work calculated the Euclidian distance between the singular values of 8×8 block of the test image and the corresponding reference image.	*Top-down approach*
Gradient Magnitude Similarity Deviation (GMSD) [22]	Gradient Magnitude Similarity (GMS) was first employed in this work to find pixel-by-pixel differences between reference and test images. After that, the comparison result was pooled using a novel method named the standard deviation of the GMS map.	*Two-step approach*
Gradient information-based IQM [23]	Harris' response was used to measure the gradient information from the test and reference images.	*Two-step approach*
Geometric structure distortion model [24]	This work finds that the geometric structure was extracted from the test and reference images.	*Two-step approach*
Gradient-based Structural Similarity [25]	Structural similarity between the test and reference images was calculated using image gradient.	*Two-step approach*
Harmonic mean-based pooling [26]	A novel pooling strategy was proposed as a harmonic means to create FRIQA.	*Two-step approach*
Adaptive Visual Perception [27]	Structural information of an image was measured using Adaptive visual perception.	*Two-step approach*
Information Content Weighting [28]	This work was proposed on the hypothesis that the optimal perceptual weights and local information content are proportional to each other.	*Two-step approach*

Generic programming based SSIM [29]	This work was based on SSIM, and this handcrafted metric was formulated using genetic programming.	*Two-step approach*
Combination of spatial and frequency domain-based FRIQA [30]	. In the first step, features from spatial and frequency domains were extracted, and in the second step, a random forest was used for pooling	*Two-step approach*
VCGS [31]	This work used color appearance similarity, chrominance similarity, and gradient similarity as features.	*Two-step approach*
saliency-channel attention residual network based on the just-noticeable-difference (JND-SalCAR) [32]	It extracted the features related to the perception of HVS using the Saliency-Channel Attention Residual Network. This work was based on the Just Noticeable Difference (JND) feature of the human visual system.	*CNN based approach*
dynamic receptive field generation-based image quality assessor (DRF-IQA) [33]	In this work, the quality of the test image was measured using a representation of dynamic error and pooling of the errors using visual sensitivity-based quality.	*CNN based approach*
SPSIM [34]	This work was proposed to measure the quality of Screen content images. The screen content image is a combination of text and pictures.	*CNN based approach*

Long-range dependency model-based FRIQA [35]	A long-range dependencies model based on FRIQA was proposed, where a deep neural network was employed to highlight the long-range dependencies between the local artifacts of the test image and high-order spatial pooling applied to the image.	*CNN based approach*

2.2 – Reduced Reference Image Quality Assessment

The reduced reference image quality assessment (RRIQA) is another approach in the Objective image quality assessment research area. The overview of this approach is depicted in Figure 2.3. In this approach, part of the reference image is available instead of a complete reference image.
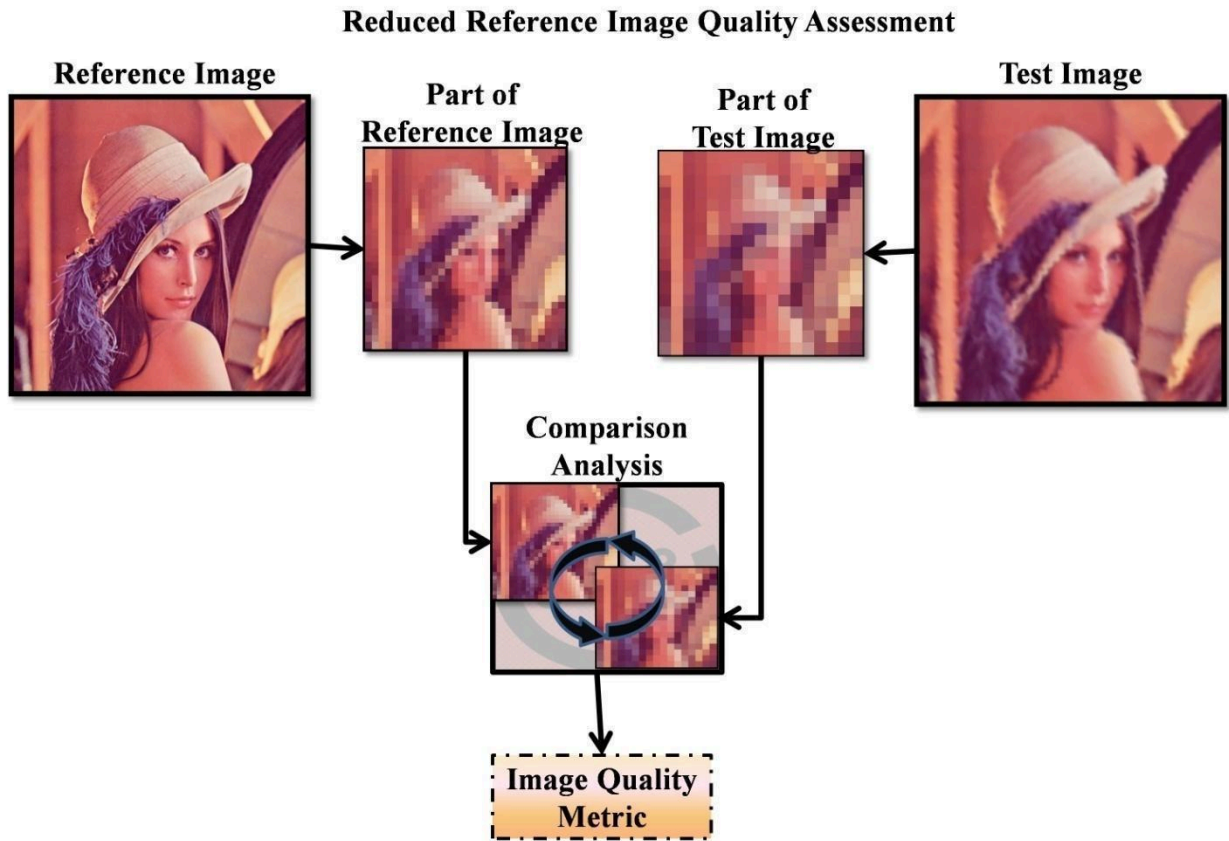
Figure 2.3: Overview of the process of Reduced Reference Image Quality Assessment (RRIQA)

The reduced reference image quality assessment is not a very popular research approach. Therefore, limited research contributions have been found in this area of research. This approach has remained untouched in this dissertation. However, few research works on this approach have been discussed in this section due to the discussion.

The research works according to the RRIQA technique and can be classified into two classes: statistics-based approach and non-statistics-based approach. The statistics-based approach assumes an image can be represented using a statistical model. On the other hand, the non-statistical approach expresses an image without the statistical model.

First, the statistics-based approach will be discussed. In this work [36], a RRIQA metric was proposed based on the natural image statistics model in the wavelet domain. The

Kullback-Leibler distance was calculated between the marginal probability distribution of the wavelet coefficients extracted from the test image and the reference image. A standard Gaussian model was employed to summarize the wavelet features of two comparing images. In another work [37], a novel reduced-reference image quality assessment method has been developed using Divisive Normalization Transformation (DNT). The DNT is an image representation technique that is motivated statistically and perceptually. DNT can reduce the first-order correlation between neurons but can't reduce the higher-order statistical dependencies. This process has been shown to significantly reduce the statistical dependencies of the original linear representation and produce approximately Gaussian marginal distributions. This strong perceptual and statistical relevance motivated the researchers to switch from the linear wavelet transform domain to the DNT domain in designing the proposed RRIQA method. L. Ma et al. [38] utilized discrete cosine transform (DCT) distribution to model the statistical structures of the reference image and test image. A three-level coefficient tree was employed in this work to reduce and standardize the statistical features extracted from the two images. The reduced features were modeled for each sub-band of DCT using generalized Gaussian distribution (GGD). After that, features from the reference and test images were compared using city-block distance measurement. Another research work [39] presented a general-purpose RR-IQA based on Natural Image Statistics (NSS) modeling. The SSIM metric is efficient and acceptable in the full-reference image quality approach. This work was motivated by the working principle of SSIM. Modifications of the SSIM calculated in full reference image have been done to create the SSIM using the reduced reference image quality assessment approach. This strategy offers two advantages. First, the RR algorithm's development can naturally incorporate the successful design principle to construct SSIM. Second, it is much simpler to obtain training data when the algorithm design includes a supervised learning stage because SSIM can be easily computed instead of costly and time-consuming subjective evaluations. The work [40] focused on the distorted image's structural degradation to assess the test image's quality. At first, spatial correlation of image structure was extracted using Local Binary Pattern (LBP). Other than the LBP-based features, different types of structural patterns were calculated. The reference and test images were compared based on the structural information extracted from the images. Finally, the Support Vector Regression (SVR) was utilized to generate the final quality score. In another work [41], a two-level Haar wavelet transform in discrete form was first applied to extract information from the test and distorted images. Then, sparse representation was

leveraged from the sub-band images to extract the free energy-based features. These features were computed using a support vector regressor to produce the final quality score. In the work [42], sparse representation was used to analyze the visual information of an image. Natural scene statistics were used here to understand an image, such as the human brain's perception. This work was evaluated on stereoscopic images. At first, the reference image and test image were computed using sparse representation. Then, the visual information of these images was measured using the distribution statistics of the visual primitives extracted from the sparse representation. Natural scene statistics of these images were evaluated in the form of luminance coefficients, which were normalized locally. The difference between the test and reference images was computed using a support vector regression model based on the visual information and natural scene statistics extracted from the images.

Non-statistics-based RRIQA has been discussed now. In the work [43], reference images and test images were not described using statistical models. Instead, both images were converted into wavelet domains, and the entropy of these images was calculated. After that, a comparison was made between the extracted descriptions of the images. A visual information fidelity-based RRIQA algorithm was proposed in the work [44]. This work employed an observation of brain theory, where it was found that the human visual system is actively sensitive to the primary visual information and less sensitive to the perception and residual understanding of the image. In the design of this algorithm, these two characteristics were separately addressed. Information fidelity for the primary and residual visual information was measured and combined to create the proposed metric.

In the paper [45], a structural degradation model (SDM) was designed. Then, the quality score of an image is defined as a nonlinear combination, or SVM-based integration, of the distance between the structural degradation information of the original and distorted images. Accordingly, a new RR IQA approach using the SDM model is exploited. In another work [46], color information was used as reduced data. Color correlogram-based descriptor was used to measure the color distribution of an image. The color distribution of an image had been altered due to the presence of distortions. This algorithm has two parts. The first part tried to identify the type of distribution present in the test image. The next part predicted the quality of the test image by providing an objective quality score. The computational intelligence model process supported these two parts. In [47], a Reduced Reference-quality algorithm was designed especially for the Light Field (LF) images. The distortion present in the test image was predicted by analyzing the depth map of the image.

A subjective quality experiment was included in this work, which produced a quality dataset for LF images. In the work [48], a novel RRIQA method was proposed for Scene Content Images (SCI),

where the visual content of an image was measured using a perceptual hashing method. After applying joint pre-processing operations, the maximum gradient magnitude and related orientation information were computed for an input image's three Red, Green, and Blue channels. Features like normalized histograms and local frequency coefficients were extracted from the maximum gradient magnitude of the input image. Finally, a hashing-based method was used to generate the quality score.

Table 2.2: Overview of State-of-the-art RRIQA methods

RRIQA metric/algorithm	Methodology	Category
Wavelet domain Natural Image statistics model [36]	An RRIQA metric was proposed based on the natural image statistics model in the wavelet domain	*Statistical approach*
DNT based Image transform [37]	Divisive Normalization Transform was used to represent the images.	*Statistical approach*
DCT based [38]	Images were represented using divisive cosine transform.	*Statistical approach*
SSIM-based [39]	Structural similarity between the test image and the reference image was determined.	*Statistical approach*
Visual structure degradation model [40]	The visual structure of the images was computed, and the amount of degradation was measured using the degradation in visual structure.	*Statistical approach*
Multi-channel decomposition [41]	The free-energy principle was employed on the output of multi-channel decomposition of the input image.	*Statistical approach*
Sparse representation [42]	Images were converted using sparse representation, and natural scene statistics were employed.	*Statistical approach*

Entropic differencing [43]	Images were converted into wavelet domain first, and then entropic differencing was computed.	*Non-Statistical approach*
Visual Information Fidelity [44]	Visual Information fidelity was extracted from the reference image and test image.	*Non-Statistical approach*
Structural degradation modeling [45]	Structural degradation of images due to the presence of distortions has to be measured.	*Non-Statistical approach*
Color distribution-based [46]	This work is based on the changes of color information in the distorted images.	*Non-Statistical approach*
Depth map of Light Field [47]	The depth map of the Light Field was computed to create the metric	*Non-Statistical approach*
Perceptual Hashing [48]	Normalized histograms and local frequency coefficients were used as features.	*Non-Statistical approach*

2.3 – No-Reference Image Quality Assessment

As the name suggests, the No-reference image quality assessment (NRIQA) is independent of the reference image in the process of assessing the quality of an image. Figure 2.4 represents the overview of the process of NRIQA.

No Reference Image Quality Assessment

Reference Image

Not Available

Test Image

Analytical Process

Image Quality Metric

Figure 2.4: Overview of the process of No-Reference Image Quality Assessment (NRIQA)

Research work in the No-reference Image Quality Assessment (NRIQA) can be categorized into two categories – Distortion independent and Distortion specific. According to the distortion-independent NRIQA research work, metrics were developed to assess the general quality of an image. However, distortion-specific NRIQA metrics were designed for particular distortions like blurriness, haze, and luminance changes.

At first, state-of-the-art research works under the distortion-independent NRIQA have been discussed. A depth image-based rendering (DIBR) based on a no-reference image quality

assessment was proposed by K. Gu [49]. This work was done based on Multi-scale Natural Scene Statistics (MNSS). Two main characteristics were taken care of when creating this metric. The first one was the geometric distortions of the natural scenes. Local self-similarity characteristics were extracted from the image to find the geometric distortions.

The second characteristic was based on the multi-scale representation of the test image. For distorted images, the structural distribution of different scales was different. However, this characteristic was not found for distortion-free images. These two characteristics helped the authors to propose the novel NRIQA metric. The work [50] was based on the chromatic information of natural images. The quality of the test image was computed based on the color structure of the image. A color descriptor was proposed to establish the representation of the image. This descriptor fused four standard models of color images. They are - hue, saturation, opponent angle, and spherical angle. Distorting affected the images, and the distribution of these models had been changed. The amount of changes in images was quantified to create the metric. A multi-task learning method was used to combine two tasks, namely feature extraction from natural scene statistics (NSS) and prediction of quality score [51]. The features from the natural scene statistics were extracted using the process of deep learning.

The use of deep learning improved the ability for representation and generalization. The work [52] proposed a distortion-independent NRIQA metric using a video codebook. The input images were divided into patches. A Gabor filter was used from the patches to extract features that were the video codebook's elements. After that, the codebook encoded the statistics of input images by quantizing the feature space and accumulating histograms of the extracted patches. This process did not depend on the types of distortions.

A NRIQA metric was proposed [53] based on the macro-micro modeling of tensor domain dictionaries. At first, tensor decomposition was done on the input images to retain the color information. After that, a target dictionary was trained accurately with the principal components. Micro and macro features were characterized using a macro-micro model in the target dictionary space. The macro-micro model was established based on a systematic mathematical interpretation. A log-normal pooling strategy was employed to identify the effective features from the set of micro features.

Similarly, effective features from the macro-feature set were extracted. Bernoulli's law was used for this purpose. Finally, the effective features from macro and micro feature sets were

computed to generate the final score. Low-level and high-level statistical features were used in the work [54] to develop a novel NRIQA metric. The low-level and high-level statistical features were very effective in capturing distortions present in an image. The low features were extracted by subtracting the local means and normalized contrast coefficients. The high-level features were extracted from the mechanism of brain perception. These extracted features were integrated using a neural network, and the final quality score was generated. In [55], the optical feature selection process was employed to create the NRIQA metric. Optimal quality-aware features were used as paired-product (PP) features. These features were extracted from the input image's discrete cosine transform (DCT) coefficients. An attention mechanism-based NRIQA metric was proposed in [56]. This metric was designed based on multi-scale features. The features were extracted using a lightweight attention method based on decomposed large-kernel convolutions. A novel feature enhancement mechanism was proposed to simulate the human visual system (HVS).

Now, the research works toward the distortion-specific NRIQA have been discussed. H. R. Sheikh et al. [57] proposed an NRIQA algorithm for compressed images using the JPEG2000 technique. Compression of images produced distortion on images. This work tried to assess these distortions in a compressed image. Natural Scene Statistics (NSS) was used to measure the amount of distortion without the presence of a reference image. The authors of the paper observed that natural scenes hold nonlinear dependencies. However, the compression process alters the nonlinear dependencies. This alteration was quantified to create the proposed algorithm. K. Gu et al. [58] concentrated on the sharpness of an image. In this work, a novel NRIQA algorithm was proposed to estimate the sharpness of an image in the autoregressive (AR) parameter space. At first, AR coefficients were estimated locally using a pointwise way. Then, energy and contrast differences of the estimated parameters were calculated. Finally, the sharpness of the test image was obtained using a percentile pooling strategy. The work [59] tried to estimate two distortions in a test image - blockiness and luminance changes. Blockiness and luminance changes are prevalent types of distortions present in JPEG images. The test images were divided into 8 X 8 non-overlapping blocks and calculated for the final estimation metric. An auto-encoder-based NRIQA algorithm was proposed by S. V. R. Dendi et al. [60]. In this paper, different types of distortions were modeled, and based on the models, these distortions were quantified. The distortion map for every kind of distortion was extracted using a convolutional auto-encoder based on the SSIM structure of the test image. This work addressed four types of distortions: JP2K compression, JPEG compression,

blurriness, and additive white Gaussian noise. In the work [61], images distorted with multiple distortions have been quantified. A feature named a gradient-weighted histogram of a local binary pattern (LBP) was used to estimate the distortions present in the test image. The work proposed in [62] focused on the images affected by contrast distortion. The contrast of an image is an essential feature that can be identified by the human visual system very quickly. Slight changes in contrast can be seen as distorted images by human perception. This work developed a natural scene statistics model of the test images using the entropy and moment features. Alteration from the standard characteristics of the natural scene statistics model can be treated as distortion. This distortion was quantified using the support vector regression model. No reference image quality assessment was done on Haze images in the work [63]. At first, a haze distribution map was generated from the haze images, and based on the distribution map, a haze assessment metric was developed. The proposed metric was named HDM-based haze assessment (HDMHA). Another CNN-based NRIQA was proposed by P. Mahajan et al.[64], where the dense net was used as the base model. This model was fine-tuned using the dataset proposed in the same paper to produce the novel metric named deep Learning-based Haze Perceptual Quality Evaluator (DLHPQE).

Table 2.3: Overview of State-of-the-art NRIQA methods

NRIQA metric/algorithm	Methodology	Category
DIBR-based [49]	This work is based on Multi-scale Natural Scene Statistics (MNSS)	*Distortion Independent*
Statistics of Perceptual Color Descriptor [50]	This work is based on chromatic information from natural images	*Distortion Independent*
Naturalness-Aware [51]	In this work, a multi-task learning method combined two tasks: feature extraction from natural scene statistics (NSS) and prediction of quality score.	*Distortion Independent*

Visual Codebook [52]	The input images were divided into patches. A Gabor filter was used from the patches to extract features that were the video codebook's elements.	*Distortion Independent*
Macro-Micro Modeling [53]	The macro-micro model was established based on a systematic mathematical interpretation. A log-normal pooling strategy was employed to identify the effective features from the set of micro features. Similarly, effective features from the macro feature set were extracted. Bernoulli's law was used for this purpose.	*Distortion Independent*
Low-level and High-level Statistical Features [54]	The low features were extracted by subtracting the local means and normalized contrast coefficients. The high-level features were extracted from the mechanism of brain perception. These extracted features were integrated using a neural network, and the final quality score was generated.	*Distortion Independent*
Optimal Feature Selection [55]	Optimal quality-aware features were used as paired-product (PP) features. These features were extracted from the input image's discrete cosine transform (DCT) coefficients.	*Distortion Independent*
MAMIQA [56]	The features were extracted using a lightweight attention method based on decomposed large-kernel convolutions. A novel feature enhancement mechanism was proposed to simulate the human visual system (HVS).	*Distortion Independent*

JPEG 2000 compression [57]	Compression of images produced distortion on images. This work tried to assess these distortions in a compressed image. Natural Scene Statistics (NSS) was used to measure the amount of distortion without the presence of a reference image.	*Distortion Specific*
Sharpness Assessment [58]	A novel NRIQA algorithm was proposed to estimate the sharpness of an image in the autoregressive (AR) parameter space.	*Distortion Specific*
JPEG compression [59]	Blockiness and luminance changes are prevalent types of distortions present in JPEG images. This work tried to estimate two distortions in a test image - blockiness and luminance changes.	*Distortion Specific*
Distortion maps for JP2K, JPEG compression, blurriness, additive white Gaussian noise [60]	Different types of distortions were modeled, and based on the models, these distortions were quantified. The distortion map for every kind of distortion was extracted using a convolutional auto-encoder based on the SSIM structure of the test image.	*Distortion Specific*
Multiply Distorted [61]	A feature named gradient-weighted histogram of a local binary pattern (LBP) was used to estimate the distortion in the test image.	*Distortion Specific*
Contrast-distortion [62]	This work developed a natural scene statistics model of the test images using the entropy and moment features.	*Distortion Specific*
Haze-distortion [63]	At first, a haze distribution map was generated from the haze images, and based on the distribution map, a haze assessment metric was developed.	*Distortion Specific*

Hazy Natural Images [64]	The dense net was used as the base model. This model was fine-tuned using the dataset proposed in the same paper to produce the novel metric named deep Learning-based Haze Perceptual Quality Evaluator (DLHPQE).	*Distortion Specific*

In this dissertation, a novel no-reference image blurriness assessment metric has been contributed. Therefore, this section presents a survey on the research works for assessing blurriness.

A lot of no-reference image blur quality metrics have been proposed. This section provides a brief discussion about those previously proposed no-reference blur metrics. The previous works can be classified into three types of approaches – (i) edge-based approach, (ii) local analysis of image statistics, and (iii) global analysis of image statistics.

(i) Edge-based approach

In this approach, the main assumption is that when an image gets blurred, the edges present in that image become thicker than the edges present in the original image. A blur metric concerning just noticeable blur was present in [65]. In this work, the probability of blurriness was calculated by measuring edge width. The edge width was compared with a threshold value. The threshold was called just a noticeable blur (JNB). The comparison values of all edges in the test image were accomplished using a Minkowski metric based on a probability summation model. In [66], a similar approach was taken; in that case, the cumulative probability of edges was calculated. This metric was named cumulative probability of blur detection (CPBD). The higher value of the metric represents the sharper image. In [67], all vertical edges in the test image were extracted using the Sobel operator. Then, all edges' starting and ending points were identified to calculate edge width. Finally, the average width of the edges was calculated. This average value was taken as a blur metric. H. Tong et al. [68] assumed that due to blurriness in the natural image edge of type, Dirac-structure and the A-step structure disappeared. They transformed the test image using the Harr wavelet

transform, and the presence of Dirac-structure and A-step structure type edges were checked, and a blur metric was proposed. In [69], the blur metric was calculated depending on the standard deviation of the edge gradient magnitude profile and the value of the edge gradient magnitude using a weighted average. The weight was calculated from the contrast of the input image—N. G. Sadaka et al. [70] first extracted edge blocks. Then, block distortion was calculated using just a noticeable blur value. Saliency map weights were then calculated, and depending on the value, blur metric was presented. S. Varadarajan et al. [71] proposed a blur metric using an iterative edge refinement algorithm in the test image.

(ii)   Local analysis of image statistics

In this approach of blurriness metric analysis, the test image is analyzed part by part. Each part is analyzed locally, and then, depending on the local analysis, the final metric is proposed. J. Caviedes et al. [72] divided the test image into 8X8 blocks. If a block had an edge pixel, then a divisive cosine transform was applied to that block, and 2D kurtosis of the probability density of that block was calculated. Finally, the average kurtosis of all blocks is calculated, which is taken as a blur metric. In [73], the standard deviation of the maximum local variation of each pixel concerning its 8-neighborhood was used to measure the sharpness of the test image. The main observation behind the work in [74] is that the wavelet analysis across a set of scales results in locally linear phase segments intersecting in a single point. The sharpness was measured by examining its local phase's linearity and coherence.

(iii)   Global analysis of image statistics

This approach considers the total image to measure the blur metric. In [75], the geometry of the test image is analyzed in the Fourier domain. In this paper, a way of defining the Global Phase coherence of an image was proposed by comparing the likelihood of the image with the likelihood of all possible images sharing the same Fourier power spectrum. In [76], a probabilistic support vector machine was applied as a rough image quality evaluator. The gradient of the input image was used as a feature. Input images are classified into two classes – blur and sharp. After the rough classification, a detailed image was used to refine and form the final blur metric, applying wavelet decomposition to reveal spatial-spectral properties. K.

De et al. [77] proposed a blur metric in the frequency domain. Their main finding is that if blurriness in an image increases, the number of high-frequency components decreases.

In this chapter, a literature survey has been done on the three approaches of image quality assessment research. This survey helps to identify the areas where research contributions can be made. In the following chapters, the research contributions have been discussed.

Chapter 3:

Full reference Image Quality Assessment

3.1. Introduction

Full reference Image Quality Assessment is the best objective IQA technique for higher accuracy. In Chapter 1, under Section 1.4.2, the IQA technique FRIQA is introduced. According to this technique, the quality of the input image was assessed by comparing it with the corresponding reference image. The reference image is the benchmark image to obtain the best quality. The input image has to be compared with the reference image to assess the quality of the input image. In Chapter 2, under subsection 2.1, state-of-the-art research works related to the FRIQA have been presented. According to the discussion, FRIQA is a significant study area, and researchers are actively working in this area. Although the researchers proposed different FRIQA methods in the literature, there are limitations to each of them. These limitations open the scope for further contributions in this domain.

The general approach of the FRIQA technique is to describe the input image and the corresponding image first using an image descriptor and compare the two descriptions to create FRIQM. In Figure 3.1, the general approach for the FRIQA has been shown.
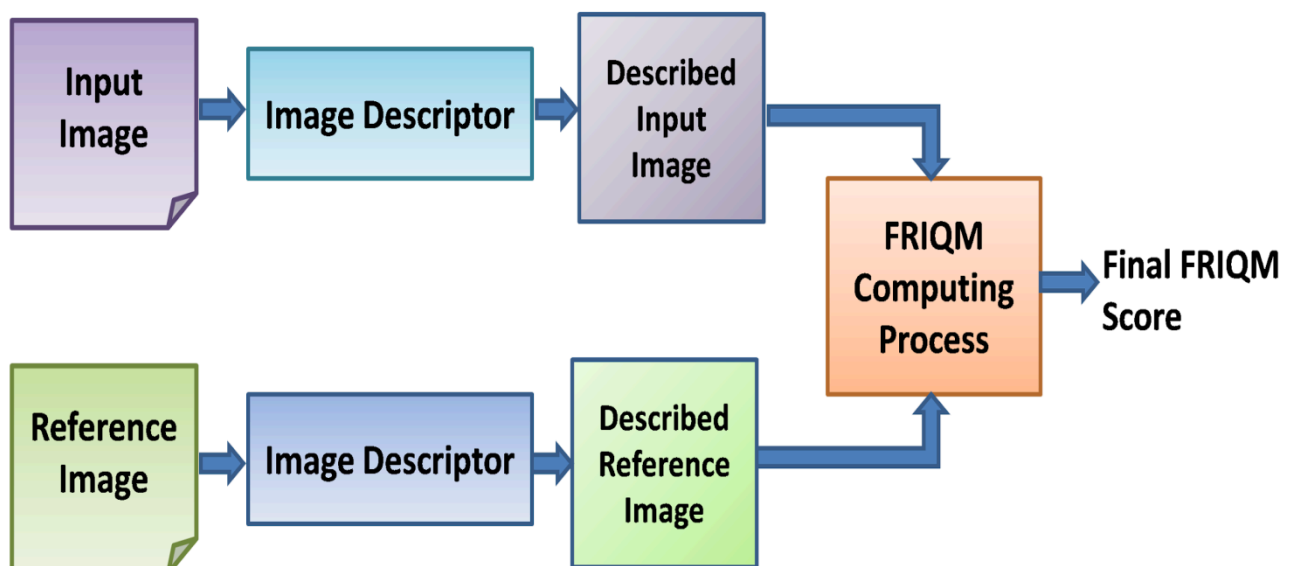
Figure 3.1: General Approach of FRIQA technique

According to Figure 3.1, the quality of the input image has to be assessed, and the FRIQM score is the output of the FRIQA technique. The input and corresponding reference images were described using an image descriptor. The described input and reference images are the input of the FRIQM computing process, and the final FRIQM score has been calculated. In the FRIQA technique, the image descriptor plays the most critical role. The main goal of the Image descriptor is to describe an image so that the information embedded into this image has been retained by removing the unnecessary parts. The usual representation of an image is suitable for human visual systems to extract information properly. However, the artificial computing process is not as capable as HVS. Therefore, images have to be described using image descriptors to understand the images better by artificial computing systems. Image descriptors are mathematical models. The intensity values of the images have been represented according to the mathematical models followed in the image descriptor process. The described images have been used for further processing.

This chapter introduces a novel image descriptor, Oriented Intensity Velocity Image Descriptor (OIVID), by considering image intensity as a vector quantity. This concept has been used in the design of image descriptors for the first time. The novel image descriptor has been evaluated on images of different types and modalities. It has been found that OIVID performs better than the state-of-the-art methods. Then, the image descriptor was applied in the FRIQA task, and a novel FRIQM named Oriented Intensity Velocity Full-reference-image Quality metric (OIVFQ) was developed. The OIVFQ performs significantly in the prediction of image quality based on the experiments conducted on three standard IQA databases.

The rest of the chapter contains sections, in which section 3.2 describes the novel OIVID and section 3.3 describes the OIVFQ. Section 3.4 includes the application of OIVID in identifying degraded images in different weather conditions. The final section of chapter 3.5 provides an overall discussion of the complete chapter.

3.2. Image Descriptor – OIVID

This section discusses the proposed image descriptor coined as Oriented Intensity Velocity Image Descriptor (OIVID). Image descriptors have been used to represent the images in an easily analyzed form. Image descriptors have been designed to extract the relevant information hidden in the image by ignoring the less relevant information. The image descriptor OIVID is a spatial image descriptor. The spatial feature descriptor algorithms

extract features from the intensity values of the input image pixels in the spatial domain. Our main intention in designing the OIVID is to use it in the FRIQA task. The special image descriptors are suitable for the IQA task for two reasons. The reasons are (i) the description of one image described by the spatial feature descriptor is distortion-specific, and (ii) the local description of the image is uniform over the entire image space. Although we have designed the OIVID for FRIQA purposes, the image descriptor can be used for other image-processing applications. We have evaluated the OIVID using the synthetic image and images from a public dataset of different image processing-based applications.

3.2.1. Motivation for designing OIVID

In the literature, various types of image descriptors are available. However, spatial image descriptors provide the best results for assessing an image because the output of the descriptors is distortion-specific, and the features extracted from the input image locally are uniform over the entire image. However, the existing spatial image descriptors cannot hold the detailed edge information and find the presence of distortions simultaneously. The OIVID has been designed to resolve these issues.

In traditional practice, the image pixels are considered coordinate points containing scalar values. However, by observation, it can be said that the intensity values also have directions based on the image's texture. Therefore, in designing the OIVID, the traditional concept of considering the image pixel as a scalar value has been altered, and the OIVID has been designed according to the idea of considering the image pixel as a vector. The Kinematic theory of Physics has been employed in the image space to create the OIVID.

3.2.2. Significance of OIVID

In the creation of OIVID, a novel measure named Velocity of Intensity (VoI) has been created by altering the traditional concept of considering the intensity values as scalar quantities by reckoning an image as a combination of vector quantities. At the same time, it has been observed that the vector quantities have velocities in different orientations within the image. The proposed OIVID describes an image according to the velocity of Intensity (VoI) measure in various orientations, and a new operator, Orientate Intensity Velocity (OIV), has been created to compute VoI in all orientations. The OIVID has achieved two notable properties: Distortion specificity and edge preservative. The two properties of the OIVID have been

justified theoretically and experimentally. The conceptual uniqueness and the two valuable properties distinguish the OIVID from the state-of-the-art image descriptors.

3.2.3. Methodology of OIVID

The Orientate Intensity Velocity Image Descriptor (OIVID) has been proposed based on the Velocity of Intensity (VoI) measure. The measure has been introduced to find the edge information and the presence of distortions in an image. Using VoI in designing OIVID makes the descriptor different from the existing image descriptors by achieving two exciting properties: edge preservative and distortion specific. However, the VoI was measured in a single direction. The Orientate Intensity Velocity (OIV) operator has been created to measure the VoI in all orientations. In the following subsections, the VoI and OIV have been described and derived accordingly. After that, the implementation process of the OIVID using VoI and OIV has been described.

3.2.3.1 Velocity of Intensity (VoI) measure and Orientate Intensity Velocity (OIV) operator

The 'Velocity of Intensity' (VoI) has been measured from an image by considering the image intensity values as vector quantities. In traditional practice, the image intensity values have been considered a scalar quantity. However, we have changed the traditional concept and considered the image intensity values as a vector quantity because it has directions along with the magnitude depending on an image's texture. An image can be regarded as a two-dimensional Euclidian space, and each coordinate point holds an intensity quantity. According to our concept, the intensity values have direction with magnitude.

Image intensity values have achieved motion within the Euclidean space along with the directions. The motion of intensity values ultimately depends on the texture of an image, and the motion of intensity is directly proportional to the texture of an image. Therefore, an image with high texture achieves high motion. Hence, an image's texture can be described by measuring the Velocity of Intensity (VoI), defined in definition 3 with equation 14, which can measure the motion of intensity and a specific direction. However, in an image, the intensity has motions and all orientations, as depicted in Figure 3.2. Therefore, measuring VoI along with a single direction is not enough to describe an image. Describing an image significantly and measuring VoI along with all directions is highly recommended. A new operator coined

as Orientate Intensity Velocity (OIV) has been introduced in this paper to measure the VoI in all orientations. The OIV has been defined in definition 4.
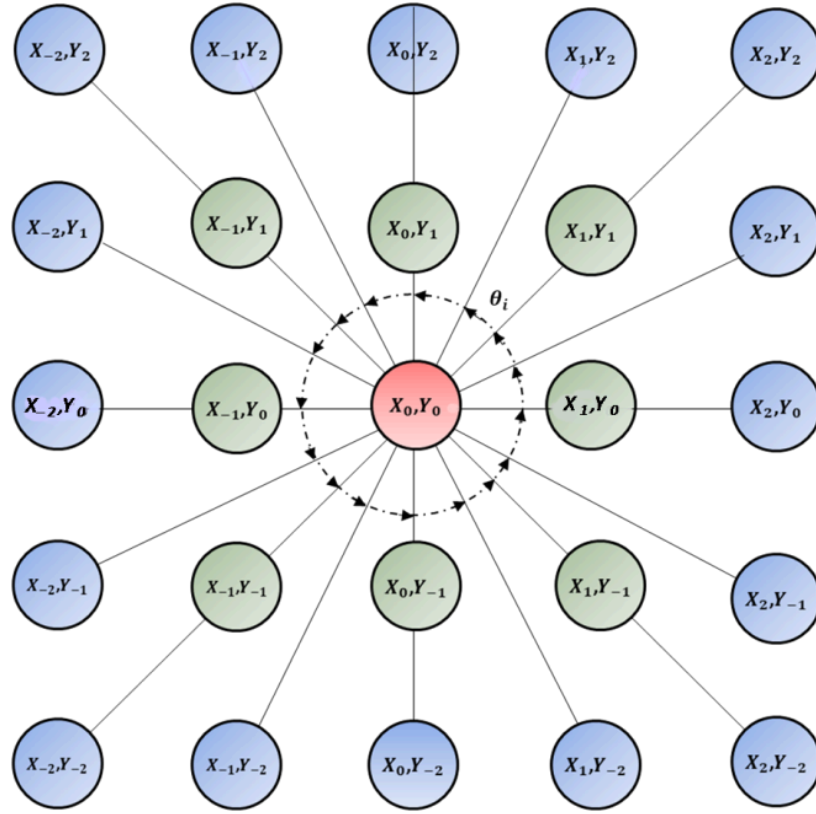
Figure 3.2: Motion of Intensity values according to the Pixel

3.2.3.2 Derivation of VoI and OIV

According to the Kinematics theory of Physics, the velocity of a moving particle is defined as the rate of change of motion in universal space. We can consider an intensity value as a moving particle in image space because the magnitude of intensity value changes along with directions due to the image's texture. The changes of magnitude can be altered depending on the directions as well. Therefore, the motion of intensity values VoI can be measured by computing the rate of change of Intensity motions in Euclidian space.

In mathematics, a Directional derivative is a process to find the rate of change of a continuous function along with a direction in real space. The directional derivative $D_{\vec{v}}$ of a continuous function f(x,y) along with a direction θ can be defined according to definition 1.

Definition 1: The directional derivative $D_{\vec{v}}$ of a function f(x,y) in the direction θ defined by $D_{\vec{v}}f(x,y)$ in equation 3.1. The $\vec{v}$ is the directional vector and $\vec{v} = \langle , \cos\cos\theta \rangle$.

$$D_{\vec{v}}f(x,y) = \frac{f(x+L\sin\theta, y+L\cos\cos\theta)}{L}$$

(3.1)

The main concern is digital images, which only contain discrete intensity values. Hence, an image cannot be treated as a continuous function. Therefore, the mathematical definition of the directional derivative cannot be applied to images directly. Definition 1 has been modified using a partial derivative to define the directional derivative, which can be applied to discrete functions like images. Before that, proving that a directional derivative can be defined using a partial derivative is necessary.

Proof of the concept 'Directional derivative can be defined using partial derivative.'

A variable term 'r' has been used here that denotes the difference of values between two discreet points of the function f(x,y). Therefore, the variation of the function f(x,y) can be represented by the function F(r) and can be defined as equation 3.2.

$$F(r) = f(x + \sin\sin\theta, y + r\cos\cos\theta)$$

(3.2)

Then, the rate of change of the function f(x,y) can be calculated as equation 3.3.

$$F'(r) = \frac{F(r+L)-F(r)}{L}$$

(3.3)

If we calculate the change and rate of change of the function f(x,y) for a single point (x_0, y_0), then equation 2 can be rewritten as equation 3.4.

$$F(r) = f(x_0 + r\sin\sin\theta, y_0 + \cos\cos\theta)$$

(3.4)

Putting r = 0 in equation 3, we can get equation 3.5.

$$F'(0) = \frac{F(L) - F(0)}{L}$$

$$(3.5)$$

After substituting the value of F(L) in equation 3.5, we can get equation 3.6

$$F'(0) = \frac{f(x_0 + L\sin\sin\theta, y_0 + L\cos\cos\theta\,) - f(x_0, y_0)}{L}$$

$$(3.6)$$

By comparing equations 3.1 and 3.6, we can generate equation 3.7.

$$F'(0) = D_{\vec{v}} f\left(x_0, y_0\right)$$

$$(3.7)$$

Now, the $F'(r)$ can be calculated using the chain rule. After applying the chain rule, we can get the following equations.

$$F'(r) = \frac{dF}{dr}$$

$$F'(r) = \frac{df}{dx}\frac{dx}{dr} + \frac{df}{dy}\frac{dy}{dr}$$

$$F'(r) = f_x(x, y) \sin\sin\theta + f_y(x, y) \cos\cos\theta$$

$$(3.8)$$

The $f_x(x, y)$ is the partial derivative of the function $f(x, y)$ along with the X-axis by maintaining the Y-axis constant and $f_y(x, y)$ is the partial derivative of the function f(x,y) along the Y-axis by remaining the X-axis constant. By putting r = 0, equation 3.8 becomes equation 3.9.

$$F'(0)= f_x\left(x_0, y_0\right) \sin \sin \theta \; + f_y\left(x_0, y_0\right) \cos \cos \theta \qquad (3.9)$$

By comparing equations 3.7 and 3.9, equation 3.10 has been obtained.

$$D_{\vec{v}}f\left(x_0, y_0\right) = f_x\left(x_0, y_0\right) \sin \sin \theta \; + f_y\left(x_0, y_0\right) \cos \cos \theta$$
$$(3.10)$$

Hence, we prove that the directional derivative can be calculated using partial derivatives.

The alternative definition of direction derivative using partial derivative is denoted as definition 2.

Definition 2: The $D_{\vec{v}}f(x, y)$ ***has been defined in terms of*** $f_x(x, y) + f_y(x, y)$***along with the direction*** $\vec{v} = \langle, \; \cos \cos \theta \, \rangle$ ***by equation 3.11.***

$$D_{\vec{v}}f(x, y) = f_x(x, y) \sin \sin \theta \; + f_y(x, y) \cos \cos \theta$$
$$(3.11)$$

Definition 2 is being used to develop the proposed 'Orientate Intensity Velocity Feature Descriptor' (OIVID).

Let the motion of a moving particle in the Euclidian space be represented by the function s(x,y). Then, according to the rule of kinematics, the velocity of the moving particle can be defined as the rate of change in the motion of a moving particle. Therefore, the velocity $u(x, y)$can be calculated by deriving the function $s(x, y)$according to the definition represented in equation 3.3.

$$u(x, y) = \; s'(x, y)$$
$$u(x, y) = \; \frac{s(x+L, y+L) - s(x,y)}{L}$$
$$(3.12)$$

However, the definition of $u(x, y)$ stated in equation 3.12, the velocity of a moving particle cannot be found appropriately because equation 3.12 does not consider the direction of the

velocity in this calculation. Therefore, the velocity of a moving particle in the direction θ can be calculated using definition 2 of the directional derivatives. Then the $u(x, y)$ can be determined using equation 3.13.

$$u(x, y) = s_x(x, y) \sin \sin θ + s_y(x, y) \cos \cos θ$$

$$(3.13)$$

An image can be represented as a Euclidian two-dimensional space of dimension X × Y. Each point in the space with coordinates (x,y) is called a Pixel, and each pixel contains a vector quantity intensity $I_{x,y}$. As discussed earlier, the intensity has motion according to an image's texture. The intensity value of a pixel C with the coordinate (x_0, y_0) can be represented as $I_{x0,y0}$. The intensity achieves motions according to the directions shown in Figure 3.2. The motion of intensity M(x,y) can be computed using the VoI measure defined in definition 3.

Definition 3: Velocity of Intensity (VoI), denoted as $V_θ(M(x,y))$, is the measure of intensity motion M(x,y) along with the directional vector $\langle \sin \sin θ, \cos \cos θ \rangle$ can be defined as equation 3.14.

The motion of the Ix,y can be represented as a function M(x,y). Then, the VoI of intensity values can be measured by calculating the derivative of the M(x,y) using definition 2 and equation 3.14.

$$V_θ(M(x, y)) = D_{\vec{v}} M(x, y)$$
$$V_θ(M(x, y)) = M_x(x, y) \sin \sin θ + M_y(x, y) \cos \cos θ$$

$$(3.14)$$

The $M_x(x, y)$ and $M_y(x, y)$ can be defined in equations 3.15 & 3.16 accordingly.

$$M_x(x, y) = \frac{∂M(x,y)}{∂x}$$

$$(3.15)$$

$$M_y(x, y) = \frac{∂M(x,y)}{∂y}$$

$$(3.16)$$

Equation 3.14 defines the VoI in general terms. When, the velocity can be calculated in the horizontal direction, then in equation 3.14, θ = 0°. Then, the equation is like equation 3.17.

$$V_0(M(x, y)) = M_x(x, y) \sin \sin 0 + M_y(x, y) \cos \cos 0$$
$$V_0(M(x, y)) = M_y(x, y) \tag{3.17}$$

and intensity velocity along with the vertical direction can be calculated by substituting $\theta = 90°$ and be like equation 3.18.

$$V_{90}(M(x, y)) = M_x(x, y) \sin \sin 90 + M_y(x, y) \cos \cos 90$$
$$V_0(M(x, y)) = M_x(x, y) \tag{3.18}$$

Equations 3.17 and 3.18 signify that the VoI only finds intensity velocity along a specific direction. However, our main goal is to measure VoI in all pixel orientations. Thus, we have developed an operator, Orientate Intensity Velocity (OIV), which computes the VoIof an intensity value Ix,y in all orientations concerning a given pixel. In definition 4, we have formally defined the OIV operator.

Definition 4: The Orientate Intensity Velocity (OIV) denoted as $V_0(I_{x,y})$ *of* $I_{x,y}$ *computed by equation 3.19.*

$$V_0\left(I_{x,y}\right) = \frac{1}{\eta} \Sigma_\theta V_\theta(M(x, y)) \tag{3.19}$$

Where the η is a scalar value, it denotes the number of orientations taken to find the OIV of an image and the value of θ ranges from 0° to 360°.

3.2.3.3. *Implementation of OIVID*

The OIVID has been implemented using VoI and OIV. We have developed an image filter mask based on the definition of the OIV. The mask (F_θ^η) is represented in equation 3.20.

$$F_\theta^\eta = \{0 \quad for\ (R) = \theta \quad \frac{1}{\eta}\ for\ (R) \neq \theta\ and\ R \geq 0 \quad -\frac{1}{\eta}\ for\ (R) \neq \theta\ and\ R < 0 \tag{3.20}$$

where, θ is the direction, η is the total number of orientations, and the mask size is ($\eta \times \eta$). The variable R is defined in equation 3.21.

$$R = \frac{r_y}{\sqrt{r_x^2 + r_y^2}} \text{ and } r_y = y_{i \pm d} - y_i \text{ and } r_x = y_{i \pm d} - y_i \text{ where d} = 1 \text{ to } \tag{3.21}$$
m -2

The OIVID has been computed on image I(x,y) by equation 3.22.

$$OIVID(I(x, y)) = F_{\theta}^{\eta} * I(x, y)$$

<div align="right">(3.22)</div>

3.2.4. Properties of OIVID

The OIVID has achieved two essential properties, which are as follows.

Properties 1: The OIVID can represent images by preserving complete edge information.

Property 2: The OIVID can describe images specific to the distortions present in the image.

3.2.4.1. Justification for the property 1 of OIVID

An edge is defined as the sequence of contiguous pixels where pixel intensity changes notably. Therefore, it can be said that, in the image plane, the sections where the velocity of intensity value is maximum are the edges of the image. Therefore, edges can be extracted by finding the intensity velocity of each image pixel. The highest intensity velocity extracted is the edge in the image.

In traditional practice, edges can be extracted by calculating the gradient of an image. In the gradient calculation, the variations of intensity values can be measured in the horizontal and vertical directions and then averaged. Therefore, the gradient cannot always extract the edges properly. The traditional gradient magnitude calculation cannot handle the edges properly because it only considers the edges in an image's vertical and horizontal directions. Therefore, traditional gradient calculation cannot reflect the complete edge information of an image. The minute changes in intensity values cannot be reflected through the conventional gradient magnitude calculation. Therefore, the traditional gradient magnitude calculation cannot extract the smoother edges in inclined directions. Specifically, in the case of distorted images, the edges are softer than the normal images. Therefore, special attention is mandatory for the edges of all directions in distorted images.

The proposed OIVID can solve the shortcomings of image gradient calculation. We provide a mathematical proof of the high-edge preservative property of the proposed OIVID.

A simple image I(x,y) in the two-dimensional space can be defined by equation 3.23.

$$I(x, y) = ax^2 + by^2 + c$$

(3.23)

First, the gradient of image G(I) is calculated in equation 3.24.

$$G(I) = \frac{1}{2}\left(\frac{\partial I(x,y)}{\partial y} + \frac{\partial I(x,y)}{\partial x}\right)$$

(3.24)

By solving equation 3.24, we can get equation 3.25.

$$G(I) = \frac{1}{2}(2x + 2y)$$

$$G(I) = (x + y)$$

(3.25)

Equation 3.24 defines the edge information of the image I(x,y) by applying G(I).

We apply the OIV to image I(x,y) defined in equation 3.23. We consider the value of η = 5 and the orientations of the intensity velocity along with five directions as 0°, 30°, 45°, 60°, and 90°. Then, we can get equation 3.26 by applying definition 4.

$$V_0(I) = \frac{1}{5}(2x(\sin \sin 0 + \sin \sin 30 + \sin \sin 45 + \sin \sin 60 + \sin \sin 9$$

(3.26)

By simplifying equation 3.26, we can get equation 3.27.

$$V_0(I) = \frac{2}{5}\left(1 + \frac{1}{2} + \frac{1}{\sqrt{2}} + \frac{\sqrt{3}}{2}\right)(X + y)$$
$$V_0 = 1.11(x + y)$$

(3.27)

Equation 3.27 represents the extracted edges of image I(x,y) using our proposed OIVID.

If we compare the extracted edge information from image I(x,y) by applying the traditional image gradient calculation and our proposed OIVID by comparing the equations 3.25 and 3.27, then we can say that VO(I)>G(I).

Therefore, it can be proved that property 1 of our proposed OIVID.

3.2.4.2. Justification for the property 2 of OIVID

Property 2 of the OIVID has been justified over Proposition 1.

Proposition 1: Let f(x) and g(x) be two functions, and differentiation exists between both of the functions. F(x) is the convolution of the f(x) and g(x) defined in equation 3.28, where * represents the convolution operation.

$$F(x) = f(x) * g(x)$$

(3.28)

Then, by the property of the convolution, it can be said that the derivation of F(x), which is represented as $F'(x)$ can be represented by equation 3.29, where the $f'(x)$ is the first-order derivative of the function f(x).

$$F'(x) = f'(x) * g(x)$$

(3.29)

Proposition 1 can be used to prove property 2.

Let's assume that image I(x,y) can be distorted by the distortion function η(x,y). Then, the distorted image D(I) can be defined by equation 3.30.

$$D(I) = I(x, y) * η(x, y)$$

(3.30)

If we apply the OIVID denoted in equation 3.19 on the distorted image D(I), we can get equation 3.31 by considering the range of θ = 0° to 90°.

$$V_O(D) = \frac{1}{η} \sum_{θ=0°}^{90°} V_θ(D(I))$$

(3.31)

By substituting the derivation of D(I) from equation 3.30 into equation 3.31, we can get equation 3.32.

$$V_O(D) = \frac{1}{η} \sum_{θ=0°}^{90°} V_θ(I(x, y) * η(x, y))$$

(3.32)

49

Now, by applying definition 3 from equation 3.14 in equation 3.32, we can get equation 3.33.

$$V_\theta(I(x,y) * \eta(x,y)) = (I(x,y) * \eta(x,y))'_x \sin\sin\theta + (I(x,y) * \eta(x,y))'_y \cos \quad (3.33)$$

Where, $(I(x,y) * \eta(x,y))'_x$ is the first-order derivative concerning x and $(I(x,y) * \eta(x,y))'_y$ is the first-order derivative with respect to y.

Applying proposition 1 in equation 3.33, we can get equation 3.34.

$$V_\theta(I(x,y) * \eta(x,y)) = \left(I(x,y)'_x * \eta(x,y)\right)\sin\sin\theta + (I(x,y)'_x * \eta(x,y))\cos co$$

$$(3.34)$$

If we substitute the definition of $V_\theta(I(x,y) * \eta(x,y))$ from equation 3.34 into equation 3.32, we can get equation 3.35.

$$V_O(D) = \frac{1}{\eta}\sum_{\theta=0°}^{90°}\left(I(x,y)'_x * \eta(x,y)\right)\sin\sin\theta + (I(x,y)'_x * \eta(x,y))\cos\cos\theta$$

$$(3.35)$$

Then, it can be said that the features extracted by the OIVID are specific to the distortion $\eta(x,y)$ and property 2 of the proposed OIVID has been proved.

3.2.5. Results of the Experiment conducted using OIVID

The experiment conducted to justify the OIVID has two parts. At first, the OIVID was evaluated using the synthetic image, and then images from the public dataset were used.

3.2.5.1. Experiment with Synthetic Image

A synthetic image was used for the experiment, as shown in Figure 3.3.a. The image contains two white inclined lines that cross each other on a black background. This image has been used because the intensity values change between two pixels seen in the image. The traditional gradient magnitude operator finds the changes in intensity values in an image, and the OIVID does the same thing. The main goal of this experiment is to signify the OIVID

over the traditional gradient magnitude operator. The traditional gradient operator finds the intensity changes in the image only in the horizontal and vertical direction, and as a result, the operator returns only the edge information. However, the OIVID finds the intensity changes of an image in all orientations and returns other necessary information along with the edge information. Figure 3.3.a has been designed to experiment.
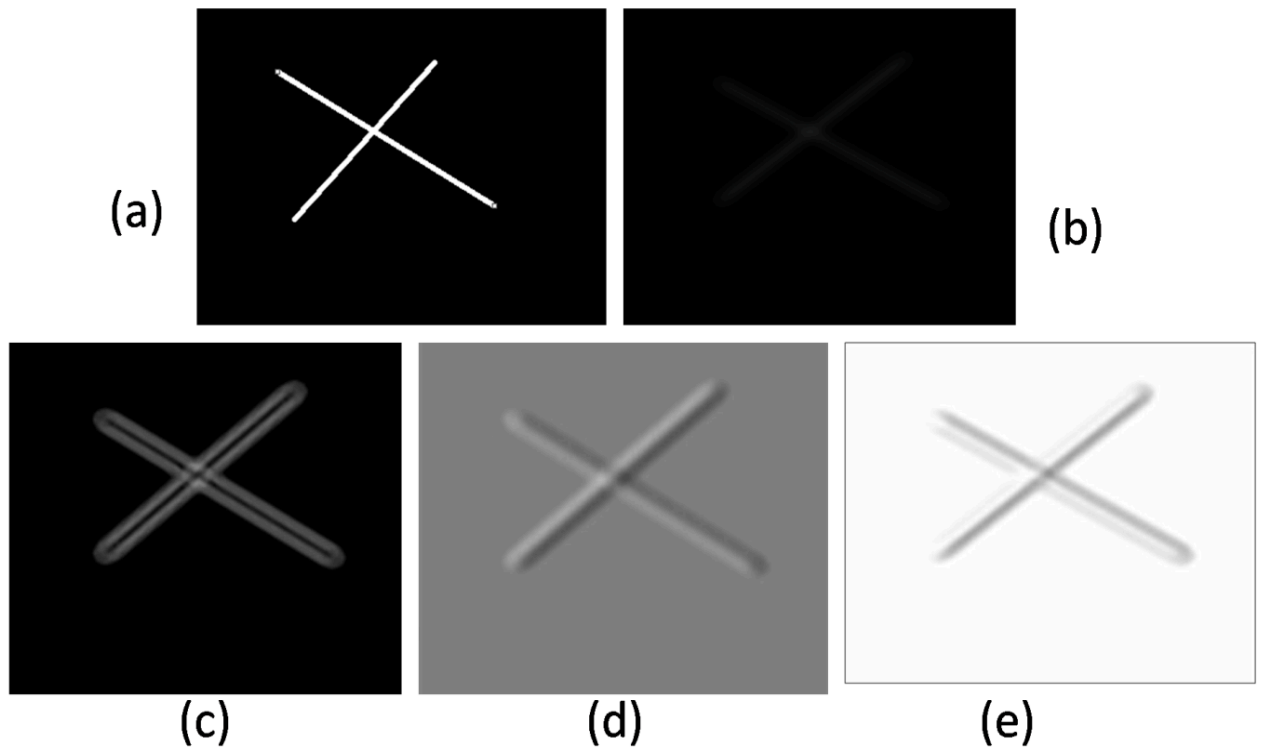
Figure 3.3: Experiment results of the experiment on Synthetic image

The synthetic image has been distorted by adding additive noise artificially to make the experiment more challenging. Figure 3.3.b shows the synthetic image after adding the noise. Traditional gradient and OIVID are applied to the distorted image shown in Figure 3b, and the resultant images are shown in Figures 3c and 3d accordingly. The difference between Figure 3d and 3c is shown in image 3e. Figure 3e contains some positive intensity values. That means the output of OIVID contains more information than the output of a traditional gradient tool. Another fascinating observation from Figures 3a and 3d is that Figure 3c only shows the edges; however, Figure 3d completely describes the image.

3.2.5.2. Experiment with Images from Public Datasets

The OIVID has been evaluated on public datasets as well. We chose datasets from two crucial industrial applications of image processing: steel manufacturing and Biomedical. The experiment on the two application domains justifies that the proposed OIVID can be applied independently of image processing application domains. Another reason for choosing the three datasets is that the datasets contain images of different modalities. The dataset from the steel manufacturing industry includes the grayscale image, and the Bio-medical dataset contains the ultrasound dataset. The OIVID has been applied to the images from all three datasets, and for all the datasets, OIVID performs excellently. The performance of the OIVID has been compared with the benchmark image descriptors, and according to the performance metric values shown in Table 3.1, OIVID outperforms all of the methods.

i) Application of OIVID in the Steel Industry

The proposed OIVID has been applied to detect defects in the hot-rolled steel strip, a critical image-processing application in the Steel industry. The NEU [78] hot-rolled steel surface defect detection database has been used for detecting steel surface defects. This database contains grayscale images from six defect classes: Crazing, Inclusion, Patches, pitted surface, rolled-in-scale, and scratches. The OIVID has been applied to the images, shown in Figure 3.4, and a multiclass kNN classifier has been used for the classification task. In Table 3.1, the performance of the OIVID in this application is shown in the columns entitled DB1.
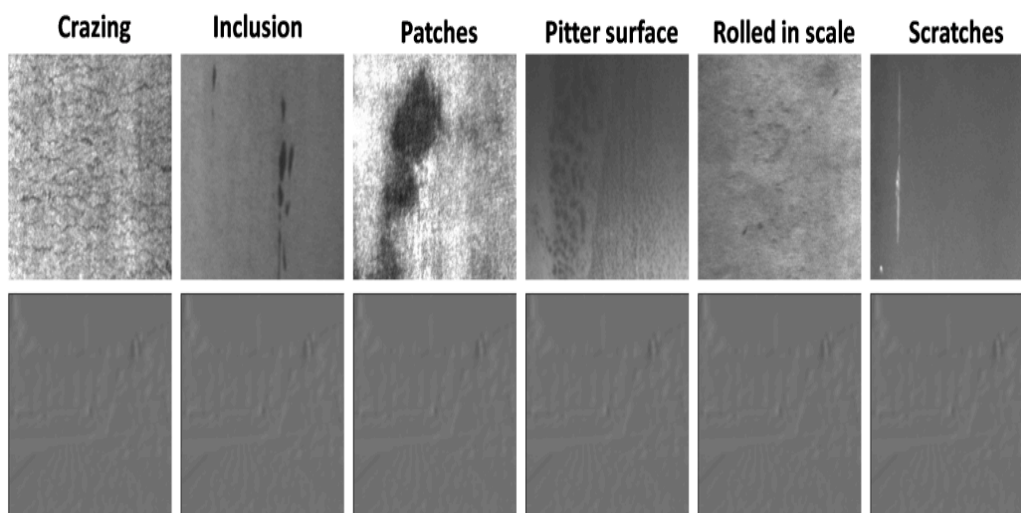
Figure 3.4: Results of OIVID on different classes of NEU dataset

ii) Application of OIVID in the Bio-medical

Biomedical is another significant application area of image processing. The OIVID has been applied to detect malignant breast tumors. The Breast Ultra Sound Imaging (BUSI) [79] dataset has been used for the experiment. The database contains images from three classes, namely – standard breast images, breast images containing benign tumors, and breast images containing malignant tumors. Similar to the previous work, input images have been described using OIVID, shown in Figure 5, and a classification task has been done using the kNN classifier. Table 1 displays the performance of the OIVID in the column with the title DB2.
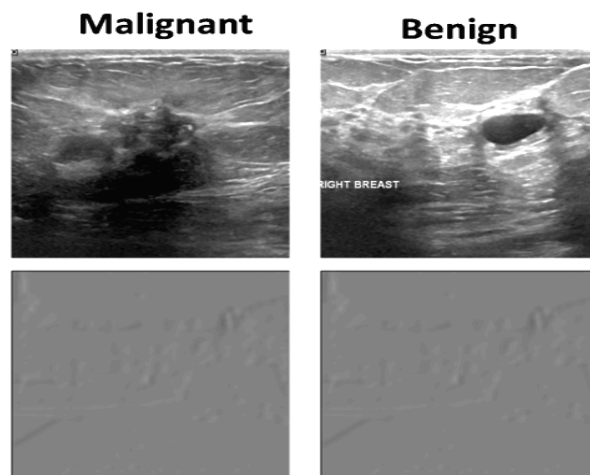
Figure 3.5: Results of OIVID on different classes of BUSI dataset

Table 3.1: Performance of the OIVID in two separate areas of applications and comparison with the other state-of-the-art methods

Image Descriptor	Accuracy		Precision	
	DB 1	DB 2	DB 1	DB 2
DAISY [80]	0.82	0.84	0.82	0.79
LIOP [81]	0.71	0.72	0.75	0.76
WLD [82]	0.72	0.74	0.75	0.71
DASC [83]	0.86	0.88	0.87	0.83
LGFA [84]	0.86	0.88	0.87	0.83
Oszust [85]	0.86	0.88	0.87	0.83
PLGM [86]	0.86	0.88	0.87	0.83
PLGA [87]	0.86	0.88	0.87	0.83
PLGF [88]	0.89	0.90	0.92	0.93
OIVID	**0.91**	**0.93**	**0.94**	**0.95**
DB1 - NEU dataset, DB2 – BUSI dataset				

3.3. Full Reference Image Quality Metric – OIVFQ

The OIVID, described in the previous section, has been applied to design a novel 'Orientate Intensity Velocity Full-reference-image Quality-metric' (OIVFQ). The complete design process of OIVFQ is demonstrated in Figure 3.6. According to the figure, OIVID has been applied to the input and corresponding reference images. Both of the outputs were compared after that. A pooling method was used to compare the results and produce the final OIVFQ score.
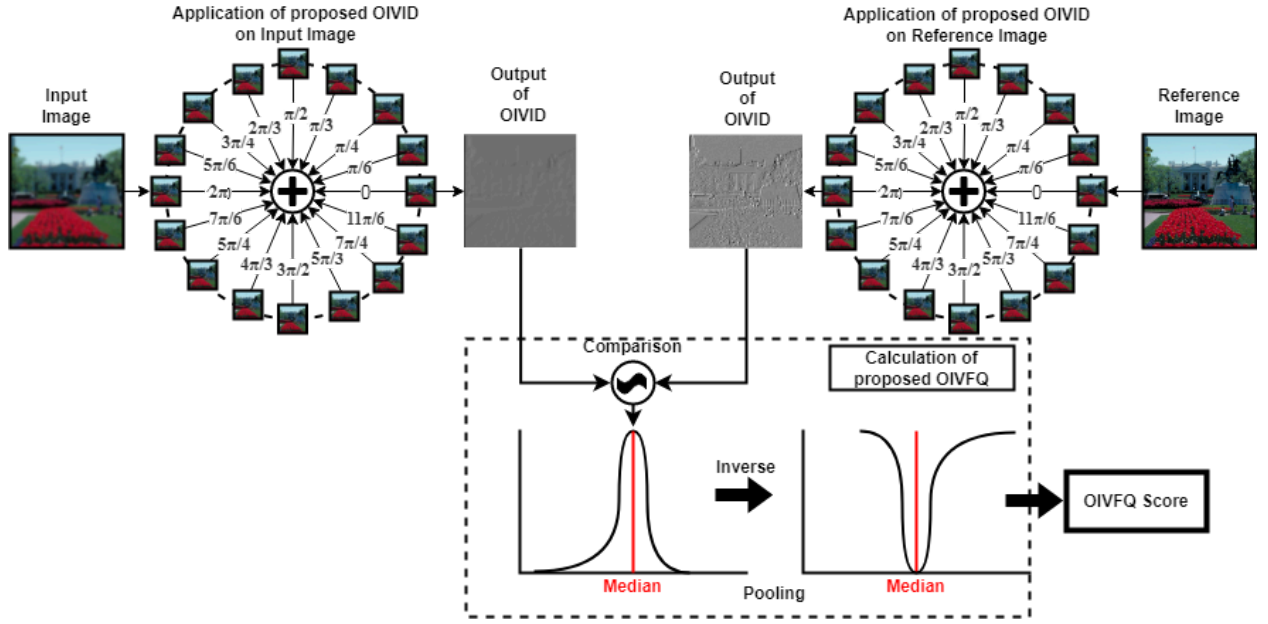
Figure 3.6: The overall methodology of the proposed research work

3.3.1. Methodology for designing OIVFIQ

Three steps have to be followed to determine the OIVFQ of an image using the OIVID. They are (i) extraction of local quality measures from the input image and corresponding reference image, (ii) comparison of local quality measures, and (iii) pooling.

1) Extraction of local quality measures from the input image and corresponding reference image

Let us consider the input image. $I_T(x, y)$ and the reference image $I_R(x, y)$. We have applied our proposed OIVID to extract the local features from the $I_T(x, y)$ and $I_R(x, y)$. The OIVID can be applied for this task due to its two essential properties. We have used equation 22 to extract the local quality measures. The OIVID of the $I_T(x, y)$ and $I_R(x, y)$ can be denoted by the terms $OIVID(I_T)$ and $OIVID(I_R)$ accordingly and calculated using equation 22 and defined by equations 3.36 and 3.37 accordingly.

$$OIVID\left(I_T\right) = F_\theta^n * I(x, y)$$

(3.36)

$$OIVID\left(I_R\right) = F_\theta^n * I(x, y)$$

(3.37)

2) Comparison of local quality measures

55

After extracting the local quality measures from the input and reference images using OIVID, the similarity map between the two images must be calculated. The similarity map Orientate Intensity Velocity similarity map (OIVSM) has been calculated using equation 3.38, where κ is a small non-zero value constant.

$$OIVSM = \frac{2\, OIVID(I_T) OIVID(I_R) + \kappa}{[OIVID(I_T)]^2 + [OIVID(I_R)]^2 + \kappa}$$

(3.38)

3) Pooling

We have used the pooling strategies to extract the final matric value from the distribution of the similarity map. According to the statistical methods, three properties must be calculated to describe a distribution. The properties are – the central tendency of the distribution, dispersion of the distribution, and the shape of the distribution. Different measures can calculate each of the properties. The mean, median, and mode can measure the central tendency of a distribution. The variance and standard deviation can estimate the dispersion of a distribution. The measure of skewness and kurtosis can measure the shape of a distribution. We have used seven pooling strategies separately. However, we get the best result for the 'Median' pooling strategy. Therefore, the proposed OIVFQ has been calculated using the median pooling strategy. The final OIVFQ has been calculated using equation 3.39.

$$OIVFQ = 1 - \frac{1}{2} \left(OIVSM(i)_{\lfloor (n+1)/2 \rfloor} + OIVSM(i)_{\lceil (n+1)/2 \rceil} \right)$$

(3.39)

3.3.2 Experiment to evaluate the OIVFQ

In this sub-section, the proposed OIVFQ has been evaluated on the public image quality assessment datasets, namely CSIQ [89], LIVE [90], and TID2013 [91]. Each database provides a subjective score for each image, taken as the standard quality score. Scores of the proposed OIVFQ have been calculated for each image from all the databases, and two types of correlation coefficients, Pearson's Correlation and Spearman's Rank Correlation, between the subjective score and the OIVFQ. The range of the OIVFQ lies between 0 and 1. A value nearer to 1 depicts the best-quality images. The evaluation of the proposed OIVFQ has been done in the following ways.

1) Experiment with different degrees of distortions of a single image to see how the OIVFQ changes according to the subjective score

This experiment has been conducted to see how the OIVFQ changes according to the degree of distortion. We have used the CSIQ image quality public database for this experiment. The CSIQ dataset contains distorted images in five degrees for each form of distortion and each reference image. Therefore, the CSIQ dataset is suitable for this experiment.

In Figure 3.7, we have shown the reference image, its corresponding distorted images of five degrees from low to high, and their corresponding DMOS value and the OIVFQ values. Figure 4 shows that the proposed OIVFQ value changes according to the degree of distortions, and these values correlate with the DMOS values. In this Figure, the outputs of applying OIVID on distorted images have been shown. It is very clear from the figure that the two metric values are highly correlated.
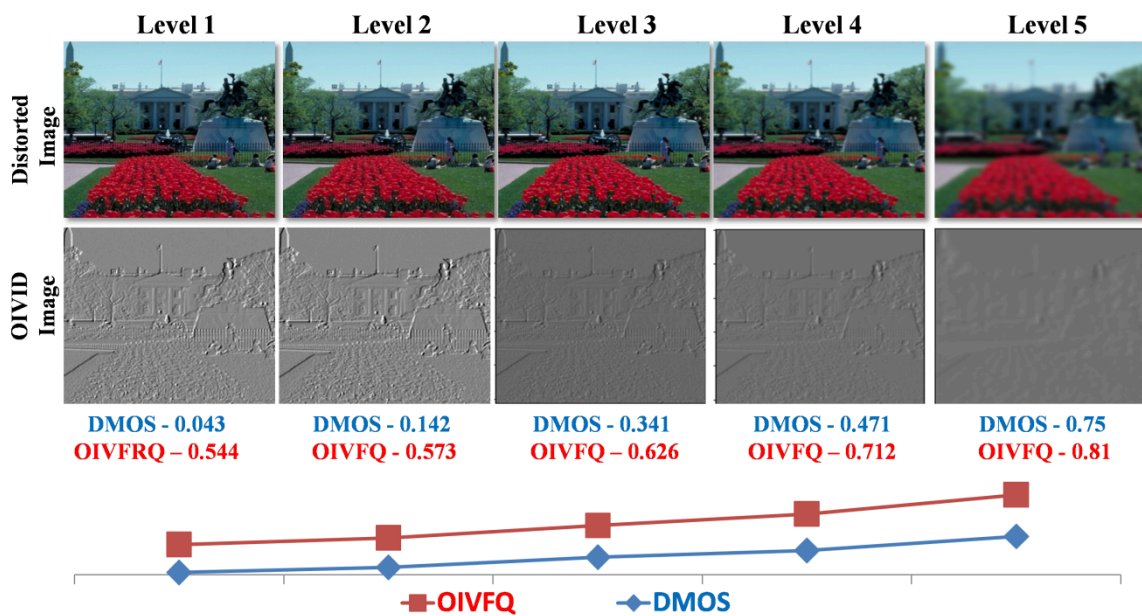
Figure 3.7: Experimental results on distorted images of different degrees

2) Experimental results on different IQA databases

We have calculated OIVFQ values for each image from three datasets, CSIQ, LIVE, and TID2013, and measured the correlation between the OIVFQ values and corresponding DMOS values using SRC and PROCC. We have compared the performance of the proposed OIVFQ with fourteen state-of-the-art methods. In Figure 3.7, the first row shows the comparison results based on SRC, and the second row shows the comparison results based on PROCC. In each figure, numbers are denoted in each row and column, and each number corresponds with a method. If the row methods perform better than those in the column, the

block has been colored green. If the opposite situation occurs, the block is colored red. If the performance of the row method and column methods is equal, then the block has been colored blue. In all three figures, it has been shown that the proposed OIVFQ performs better than the other competing methods.
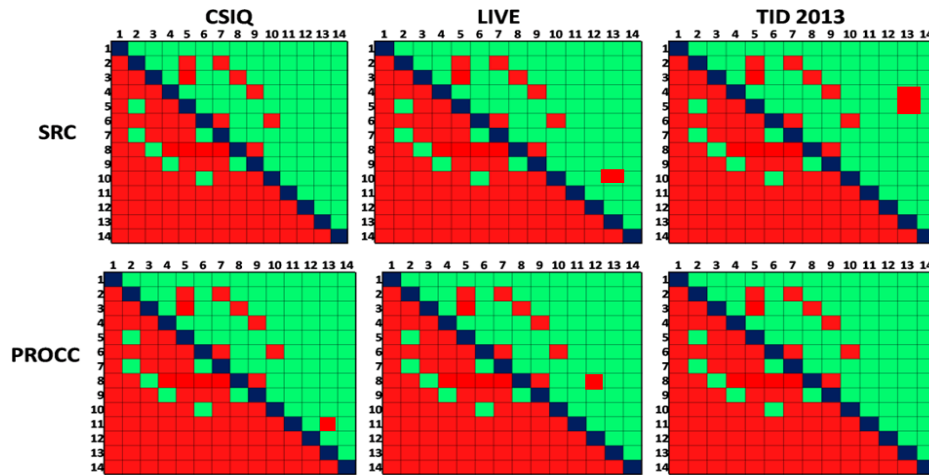
1. OIVFQ 2. GMSD 3. FSIM 4. IWSSIM 5. GMSM 6. GS 7. MSSSIM 8 MAD 9. VIF 10. SSIM 11. GSSIM 12. SSD. 13. IFC 14. PSNR

Figure 3.8: Experimental results on distorted images of different degrees

3.4. Application of OIVID identifying images degraded by different weather conditions

Weather classification is another central area of application of image processing. Natural stimuli like rain, snow, haze, and excessive sun rays have degraded images. Identifying the naturally degraded image is explicitly an exciting area of research. A multiclass weather classification database [92] has been used in this experiment. It is a four-class classification problem: sunny, foggy, rainy, cloudy, and snowy. The classification operation is performed using the kNN classifier on the images processed by OIVID, shown in Figure 3.7. The performance of the OIVID can be found in the columns with the label DB3 of Table 3.2.
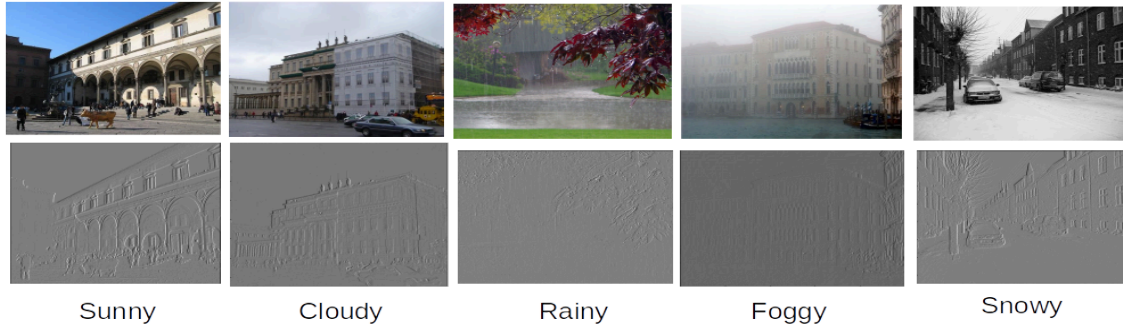
Figure 3.9: Results of OIVID on Weather classification dataset

Table 3.2: Performance of the OIVID in The weather classification-based areas of applications and Comparison with the OTHER State-of-the-art methods

Image Descriptor	Accuracy DB3	Precision DB3
DAISY	0.81	0.79
LIOP	0.74	0.76
WLD	0.76	0.71
DASC	0.82	0.83
LGFA	0.82	0.83
Oszust	0.82	0.83
PLGM	0.82	0.83
OIVID	**0.92**	**0.95**
DB 3 – Weather classification dataset		

3.5. Discussion

This chapter proposes a spatial image descriptor, the Orientate Intensity Velocity Image descriptor (OIVID), based on the Velocity of Intensity (VoI) measure. The VoI helps to represent the texture of an image. No other existing Image Descriptor applies this concept in their design, and they find difficulties in describing images independent of application areas and image modalities. Mainly, two properties, distortion specificity, and edge preservative,

59

help the OIVID perform better than the other related descriptors, and the advancement of the OIVID has been determined by the experiments conducted on images from different modalities and different industrial application areas. In addition, these two properties help to apply the OIVID in image quality assessment tasks, and a novel Orientate Intensity Velocity Full-reference-image Quality-metric (OIVFQ) has been proposed in this paper. The performance of the OIVFQ has been evaluated on three public image quality datasets and outclasses thirteen benchmark image quality metrics according to two correlation metrics.

This chapter contributes to two crucial areas of image processing research: image descriptive design and image quality assessment. These two processes are mandatory for any image processing application. Our contributions in these domains are very significant, and the concepts we established also open new research paradigms. Still, both the OIVID and OIVFQ have areas for future modification.

Chapter 4:

No-reference Image Assessment

4.1. Introduction

In the previous chapter, we discussed the Full-reference Image Assessment approaches. Section 1.4.2 states three ways of objective image quality assessment techniques. This chapter examines one of the critical and impactful image quality assessment techniques coined as the No-reference image quality assessment belonging to the Objective image quality assessment. As the name suggests, this image quality assessment technique is independent of reference images. According to this approach, an input image has been assessed without comparing it with the reference image. Although full-reference image quality assessment is a very accurate technique for determining an image, this technique entirely depends on the reference image. However, the presence of a reference image is not always achievable in real-world applications, and the full-reference image quality assessment technique does not meet the expected outcome without the availability of the reference images. Therefore, the reliance on reference images is one of the significant drawbacks of the full-reference image quality assessment method. The No-Reference Image Quality Assessment Technique can overcome this shortcoming by abolishing the dependency on the reference image, and this technique is applicable in real-life image assessment.

The input image is the input for the No-reference image quality assessment technique, and the objective quality score is the technique's output. As stated earlier, the input image has been processed solely, and the basic properties of an image have to be assessed to determine whether they have been altered. If any distortion affects an image, then the original intensity values of the input images have been changed because the distortion function has been convolved with the image function and produces some function different from the actual image function. In the no-reference image quality assessment technique, the altered image function has been assessed to determine the type of distortion that impacts the image and its degree of distortion. There are several state-of-the-art techniques under this type of image quality assessment technique to assess the quality of an image. In section 2.3, these techniques have been discussed.

Blurring is a common distortion type that frequently distorts input images in real-world scenarios. This chapter introduces a novel image quality assessment metric, the No-Reference Image Blurriness Estimation Metric(NIBEM). This is a no-training-based, no-reference image quality metric that performs the global analysis of image statistics. The following sections explain the novel metric NIBEM in detail.

4.2 Motivation of the NIBEM

There are different types of distortions present. Those are - blurring, impulse noise, Additive Gaussian Noise, Special Correlation noise, quantization noise, distortions due to compression, chromatic aberration, etc. Blurring is the most common type of distortion in those types of distortions. An image can be blurred for many reasons – defocusing of the camera, movement of the acquisition device, the presence of high-intensity light, etc. When a vision is blurred, the variation in intensity values will be below, and objects in the image cannot be adequately separable from the backgrounds. Therefore, the information in the image goes very low. Analyses of those blurred images may lead to erroneous conclusions. Thus, it is crucial to check whether the image is blurred when analyzing it. An image quality metric that can measure the blurriness of an image is beneficial in this regard. The NIBEM is a metric that can assess the blurriness of an image.

4.3 No-reference Image Blurriness Metric (NIBEM)

This chapter presents a no-training-based, no-reference blur metric NIBEM by global analysis of image statistics. This metric calculation is done entirely based on the input test image itself. No comparison with reference images is made when calculating our proposed metric. No feature set of the reference image is needed to analyze our proposed metric. Thus, the metric is ultimately a no-reference image quality metric. No classification method has been used to calculate the metric. No training operation is done here. Therefore, it is a no-training-based metric. The input test image has been transformed by a divisive normalization transform (DNT). The global analyses of the transformed coefficients have been performed to define the proposed blur metric—the value of the metric increases with the blurriness of an image. The value of the metric is between zero and one. When the image is completely blurred, the value is one, and the value is zero for a sharp image.

In calculating the NIBEM metric, it has been found that, for some reason, the global analysis of image statistics approaches the other two categories - edge-based approach and Local

analysis of image statistics. The edge-based method has not been considered because the edge-based approaches do not always provide appropriate scenarios. In a wholly blurred image, edge detection is not always proper. In a blurred vision, the intensities are averaged. Therefore, the intensity difference between two image pixels goes low. As a result, edge detection operators fail to identify edges correctly, mainly horizontal edges. This method fails in fragile edges in the original image because the edges disappear entirely in the distorted image. In [93], the authors found the Dirac-structure and A-step structure edges to calculate the blurriness. However, the edges of the roof structure do not change after blurring. Suppose an original image contains all roof structure edges according to their approach. In that case, no blurriness can be measured after blurring because the distorted image has no D structure and A-step structure type edges.

In the local analysis approach, images are analyzed part by part. There is the possibility that a picture is not completely blurred. A tiny portion is obscured, but other factors are sharp. This can be possible for defocusing the camera. The small portion is not in focus, but the different parts are. In local analysis, though, the distorted portion was found. However, this distortion does not affect the final measure because this malformed part has minimal impact, according to the other parts.

The metric NIBEM has been developed based on theorem 1, which has been described as follows -

Theorem 1: *When a blurred image is transformed using the Divisive Normalization Transform, the distribution of the transformed coefficients becomes thin and like a spike distribution, rather than the distribution of the divisive normalized coefficients of the sharp, non-blurred image. The distribution shrinks as the amount of blurriness increases in the image.*

It was previously observed that when a sharp, non-distorted image is transformed using a divisive normalization transform, the distribution of transformed coefficients follows a normal distribution. Whether a distorted image is transformed by divisive normalization, then the transformed coefficient does not follow normal distribution [94]. In [94], the authors showed the effects of different distortions on DNT distribution but did not justify the behavior of the images. Our work concentrates on blurred images, and our proposed theorem specifically emphasizes blurred images. The statement of theorem one is depicted pictorially

63

in Figure 4.1. The thin, highly peaked distribution can be observed because of blurriness. The intensity values of an image become averaged. Therefore, the values of intensities lie very near to medians. Divisive normalization transforms intensity values in a way that the intensities usually distribute. However, as the intensities lie nearer to the medians in the blurred image, the distribution shrinks and looks like a spike after divisive normalization. As the blurriness increases, the shape changes more. The blue line represents the distribution of the DNT coefficients, and the red line represents the expected Gaussian curve. In the global analysis of image statistics approach, the test image is transformed into another domain, and the quality analysis is done. We have used divisive normalization change for this purpose.
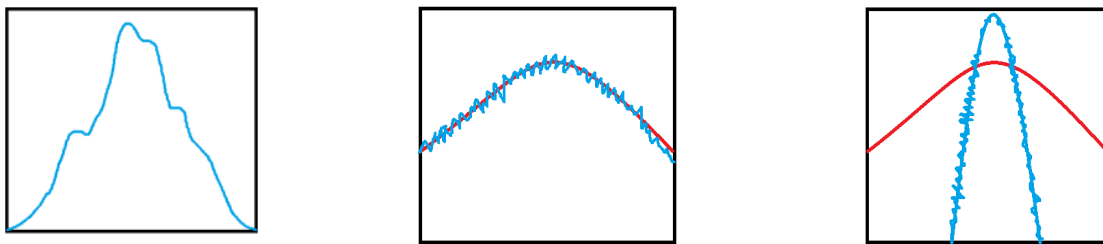
(a) distribution of an image that is nonlinear

(b) After divisive normalization, the distribution follows Gaussian (in case the image is sharp, non-blurred).

(c) When the image is blurred, the distribution of the divisive normalized coefficients looks thin in shape.

Figure 4.1: Comparison of the distributions of divisive normalized coefficients of the blurred image and non-blurred image

The use of divisive normalization transformation has several benefits. The reasons why we chose divisive normalization are as follows –

(i)   Divisive normalization transform provides lower dependencies between sub-bands of steerable pyramid decompositions [95]. The distribution of coefficients of sub-bands becomes similar by applying divisive normalization.

(ii)   The Divisive normalization transform can explain the adaptation of neural responses due to variations in visual surroundings [96].

(iii)   It explains the visual masking effect [97].

These properties of divisive normalization transform to make it an excellent model for estimating various distortion types of images.

Figure 4.2 shows how the proposed blur metric is measured. The input image was first transformed using the divisive normalization transform. The transform is done in two steps – steerable pyramid decomposition followed by normalization. Then, the transformed coefficients are used to calculate the blur metric. The metric calculation process has three further stages. They are – MAD calculation, Estimation according to orientation, and Estimation of the metric sequentially.
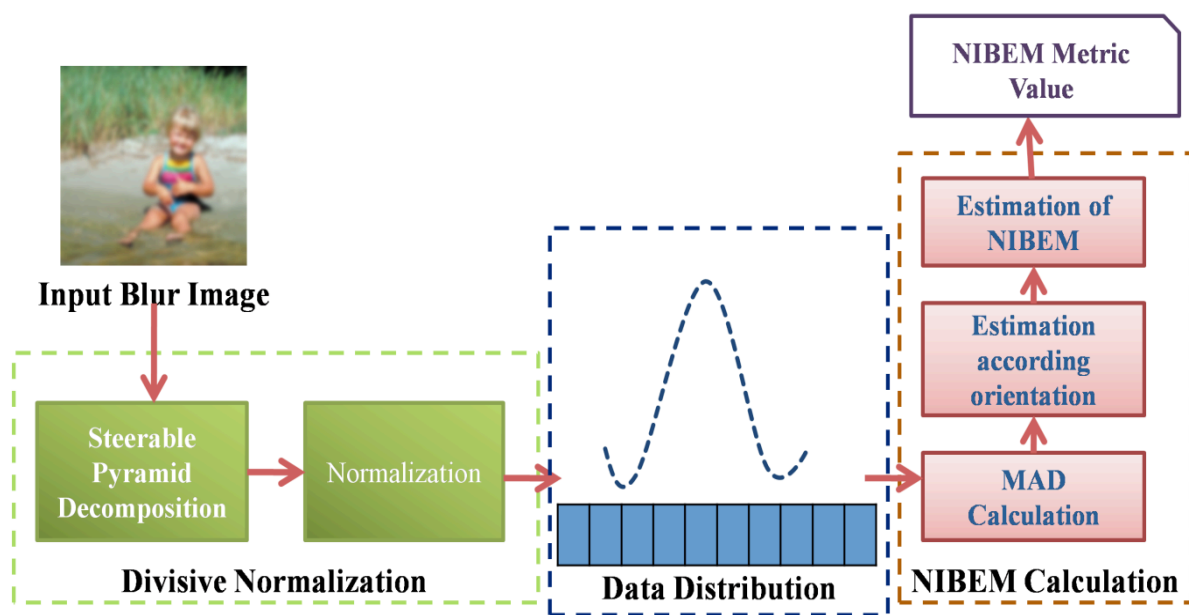
Figure 4.2: Methodology to calculate the proposed metric

4.4 Methodology

The working methodology of developing the NIBEM metric has two parts. The first part is the divisive normalization transform of the input image, and the second part is Metric calculation.

4.4.1 Divisive normalization transform

The divisive normalization transformation has two parts: steerable pyramid decomposition and normalization.

4.4.1.1 Steerable pyramid decomposition

The Steerable pyramid decomposition [98] is a linear multi-scale, multi-orientation image decomposition. The basis functions of the steerable pyramid are k-th order directional derivative operators. Therefore, the decomposed image comes into different sizes (factor of 2) and k+1 orientations.

The steerable pyramid decomposition overcomes the limitation of orthogonality constraints. Therefore, wavelet decomposition is more advantageous than wavelet decomposition in our work. The images decomposed by steerable pyramid decomposition span a rotation-invariant subspace, and the complete decomposition results in a tight frame. The decomposition performs a polar-separable decomposition, independently representing each scale and orientation. It obeys the generalized form of Parseval's Equality. The representation is translation-invariant and rotation-invariant. Thus, decomposition has been used to check distortions in all scales and orientations. However, the representation of decomposed images in all scales and orientations is entirely independent. Then, there is no chance of over-completeness in the case of distortion estimation. The input images have been decomposed into three scales and four orientations. Then, distortion estimation on each sub-band of the decomposition was performed. The decomposed images of one input image are shown in Figure 4.3. The decomposed images are named in the form $I_{i,j}$; where i is the scale number and j is the orientation number. Three scales and four orientations have been used in this work because it is seen that steerable pyramid decomposition in three scales and four orientations provides the best result [99].

1,1	1,2	1,3	1,4
2,1	2,2	2,3	2,4
3,1	3,2	3,3	3,4
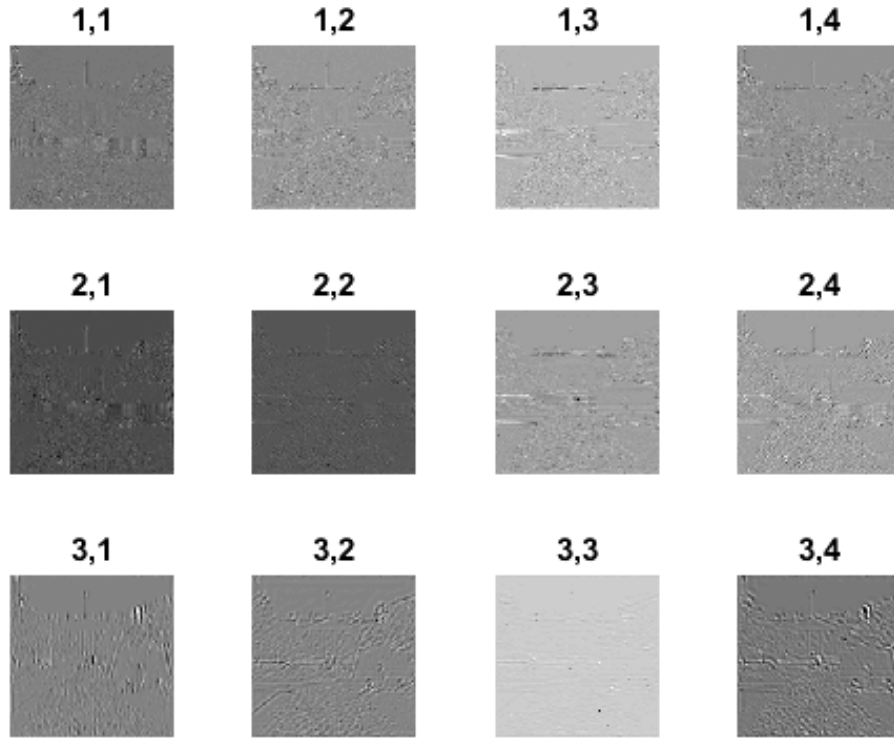
Figure 4.3: Sub-bands of Steerable pyramid decomposed images in three scales and four orientations

4.4.1.2 Normalization step

The coefficients obtained from steerable pyramid decomposition have been normalized in this step. Let x be the coefficient obtained from the steerable pyramid decomposition. Then, the divisive normalization coefficient x' can be defined as

$$x' = \frac{x}{n} \tag{4.1}$$

where n is the positive normalization factor, calculated using the Gaussian scale Mixture (GSM) [100] model.

Let X be a random vector with length L. Then, X can be said to be GSM if it follows the equation

67

$$X \triangleq qV \tag{4.2}$$

Where $\triangleq$ denotes equality in probability, V = zero-mean random Gaussian vector with covariance C_V, q is a scalar random variable called a mixing multiplier.

The GSM model can be explained more specifically in terms of the density of random variables, then the following equation exists.

$$p_X(X) = \int \frac{1}{[2\pi]^{\frac{L}{2}} \left| q^2 C_V \right|^{\frac{1}{2}}} exp\left(- \frac{X^T C_V^{-1} X}{2q^2}\right) p_q(q) dq \tag{4.3}$$

$p_X(X)$ is the density of vector X and $p_q(q)$ is the density of the mixing multiplier q.

In simple words, the Gaussian scale mixture model converts the density of a random vector into a Gaussian distribution but on a different scale. From equations 4.2 and 4.3, it can be said that in the Gaussian scale mixture model, the density of random vectors is combined with the help of a mixing multiplier to form a Gaussian distribution. It is also seen in [9] that the Gaussian scale mixture model is helpful when applied to marginal and joint statistics of wavelet coefficients of natural images. For those reasons, we have used the Gaussian scale mixture model in our work.

In the general form of GSM, the mixing multiplier q is assumed to be a continuous scalar random variable. The value of q is fixed for a fixed location. However, it varies for different scales and orientations. When the value of q is fixed, vector X becomes a zero-mean Gaussian vector with covariance $q^2 C_V$. That means the mixing multiplier q can convert the distribution of random vector X into a zero-mean Gaussian distribution.

The normalization factor n in the divisive normalization transform can be defined as mixing multiplier q. The mixing multiplier q can be estimated from the neighboring coefficient vector X. The coefficient cluster moves step by step as a sliding window across a sub-band. Maximum likelihood estimation is used to estimate the mixing multiplier q. The estimated q denoted as q_e can be defined as

$$q_e = \arg \arg \max_q \{\log \log p(q) \} \tag{4.4}$$

$$q_e = \arg \arg \min_q \{L \log \log q + \frac{x^T C_v^{-1} x}{2q^2} \} \tag{4.5}$$

$$q_e = \sqrt{\frac{x^T C_v^{-1} x}{L}} \tag{4.6}$$

The q_e is used as a normalizing factor in the divisive normalization transform, and equation 4.1 becomes

$$x' = \frac{x}{q_e} \tag{4.7}$$

In this work, for each coefficient x' of steerable pyramid decomposition, the vector X in equation 6 is taken from the neighboring coefficients. The length of X is 13 here. Nine coefficients are from the same sub-band – the coefficient itself and its eight neighbors; three are the same location coefficients of the same scale, but other orientations and one are the same location coefficient of the parent sub-band.

Proof of Theorem 1 stated in section 4.3

It has explained how the divisive normalization transformation on input images has been accomplished so far. When the input image becomes blurred, we discuss what happened in that case. We have proposed a definition of Blurry image in Definition 1 and a lemma in Lemma 1. We have proved them, and the Definition and lemma prove theorem 1 accordingly.

Let an image be defined as a two-dimensional function I(x,y) and the blurriness using the function B(x,y), where x & y are two coordinates in Cartesian space.

When an image gets blurred, the blurred image $I_B(x,y)$ can be expressed as equation 4.8.

$$I_B(x, y) = I(x, y) * B(x, y) \tag{4.8}$$

Where * is the convolution operation.

In this chapter, a new definition of the blurred image is proposed based on the variance of an image. The proposed definition of blurred image is as follows.

Definition 1: Images with minimum variance are Blurred Images.

The definition of the blurred image can be justified in the following process.

Justification of Definition 1: In the literature, different blur functions are available – Gaussian blur, motion blur, mean blur, etc. Although functional variations exist, any blurring operation minimizes the intensity differences between the two pixels in an image. Blurring an image means low pass filtering on the image. That means $B(x, y)$ is a low-pass filter. As a result, the high-intensity values of image pixels are reduced, and the difference between the maximum and minimum intensity values becomes low. Then, irrespective of the types of blurring functionality, as a general, the blurred image can be defined mainly as the minimization of the variance of an image as per equation 4.9.

This definition of the blurred image can be derived as follows.

Consider X as the vector representing the intensity values of an image, and Var() is the variance of the vector X.

Let us consider an image of size m × n and its intensity value of the pixel (i,j) represented as $x_{i,j}$. Then the vector $X = [x_{i,j}]_{i=1, j=1}^{i=m, j=n}$

The blurring means minimizing the deviation of intensity values of image pixels from each other. Therefore, the blur function can be written as

$$I_B(x, y) = \left(\sum_{i=1, j=1}^{i=m, j=n} (x_{i1,j1} - x_{i2,j2})^2 \right)$$

$$= \frac{2m^2n^2}{2m^2n^2} \left(\sum_{i=1, j=1}^{i=m, j=n} (x_{i1,j1} - x_{i2,j2})^2 \right)$$

$$= 2m^2n^2 \left(\frac{1}{m^2n^2} \sum_{i=1, j=1}^{i=m, j=n} \frac{1}{2} (x_{i1,j1} - x_{i2,j2})^2 \right)$$

$$= \arg \arg \left(\frac{1}{m^2n^2} \sum_{i=1, j=1}^{i=m, j=n} \frac{1}{2} (x_{i1,j1} - x_{i2,j2})^2 \right)$$

As the vector X contains equally likely variables, then the above equation can be written as

$$I_B(x,y) = \arg \arg \min(Var(X)) \tag{4.9}$$

According to the proposed process, the blurred image $I_B(x,y)$ further decomposed through a steerable pyramid. In this decomposition, rotational derivation of the image is done according to the scale factor. We can show the derivation for scale 1 on the blurred image $I_B(x,y)$.

At first, the derivation is done with the orientation 0°, as shown in equation 4.10.

$$I_{B_1}^{0°} = \frac{d}{dx} I_B(x,y) \tag{4.10}$$

If the derivation is done on the orientation 90°, then.

$$I_{B_1}^{90°} = \frac{d}{dy} I_B(x,y) \tag{4.11}$$

If we want to generalize this thing, then we can say that.

$$I_{B_s}^{\theta} = \cos\cos\theta \, I_{B_s}^{0°} + \sin\sin\theta \, I_{B_s}^{90°} \tag{4.12}$$

Where θ is the orientation factor, and the s is the scaling factor.

The $I_{B_s}^{\theta}$ is the output of steerable pyramid decomposition, and the Gaussian mixture model has been applied to form a uniform representation stated in equation 4.2. The vector X stated in equation 4.2 contains the values from $I_{B_s}^{\theta}$ and the mixing multiplayer q is also estimated from the vector X in equations 4.4 to 4.6.

As the Gaussian mixture model suggests, the vector V must be Gaussian depending on the vector X and the mixing multiplayer q. The shape of the distribution of the vector V depends entirely on the nature of the vector X as q is a latent variable and q is estimated from X. Equation 4.3 shows that the probability density function of the vector X. This function represents the vector V, and it must be multivariate as the image is a 2D signal. According to the multivariate Gaussian distribution rule, the shape of the distribution becomes thin and peaky if the signal's variance decreases.

Lemma 1: The variance of the steerable pyramid decomposition of blurred images is less than the distribution of the steerable pyramid decomposition of the non-blurred image.

If Lemma 1 is proven, then we can prove Theorem 1.

Proof of Lemma1: To prove this, it has to be confirmed the variance of $I_{B_s}^{\theta}$ is lower than the variance of I_s^{θ}. I_s^{θ} is the steerable pyramid decomposition of the non-blurred image.

By putting the value of $I_B(x,y)$ from equation 4.9 into equation 4.12, we can get

$$I_{B_s}^{\theta} = \cos\cos\theta (\arg\, arg\, min(Var(X)))_s^{0°} + \sin\sin\theta\ (\arg\, arg\, min(Var(X)))$$

(4.13)

An alternative description of I_s^{θ} is as equation 4.14

$$I_s^{\theta} = \cos\cos\theta (I(x,y))_s^{0°} + \sin\sin\theta\ (I(x,y))_s^{90°}$$

(4.14)

The blurred image I_B has a lower variance than the sharp image I, and due to steerable decomposition, the variance of I_B decreases as a series of derivations is done on it as per equation 4.13. By comparing the equations 4.13 and 4.14, it can be clearly stated that the variance of $I_{B_s}^{\theta}$ is less than the variance of I_s^{θ}. Therefore, it is proved that the blurred image distribution DNT coefficients are thin in shape and peaky.

4.4.2 Metric Calculation

After divisive normalization, the distribution of normalized coefficients must be in Gaussian for sharp images. However, in the case of a distorted image, this property does not follow. In blurred images, the distribution goes very thin to its median and shapes like a spike. As the blurriness increases, the shape of the allocation becomes thinner. Therefore, the blurriness of a test image has to be estimated to measure the deviation of the distribution from the median of the distribution. The deviation becomes lower as the blurriness goes high. Median Absolute deviation is a statistical measure that can be helpful in this regard.

The Median Absolute deviation (MAD) [101] is the robust estimator of dispersion. It is the median of absolute deviation of medians. It can be expressed as

$$MAD = \frac{1}{N} \sum_{i=1}^{n} \left| x_i - \underline{x} \right| \qquad (4.15)$$

We calculate the Median Absolute deviation of each sub-band of steerable pyramid decomposition $MAD_{i,j}$; where i mentions the scale and j mentions the sub-band's orientation. Then, we find the average Median Absolute deviation for each orientation M_j.

$$M_j = \frac{1}{3} \sum_{i=1}^{3} MAD_{i,j} \qquad (4.16)$$

Then, we find the average of all M_j.

$$MET = \frac{1}{4} \sum_{i=1}^{4} M_j \qquad (4.17)$$

The proposed metric for blurriness estimation $METRIC_{BLUR}$ is then defined as

$$METRIC_{BLUR} = \frac{(100 - MET)}{100} \qquad (4.18)$$

4.5 Result & Discussion

Two publicly accessible subject-rated natural image databases, CSIQ [89] & LIVE [90], have been used to validate the NIBEM metric. The CSIQ database has 150 blurred images, and the LIVE database has 145 blurred images. The metric calculation was applied to every blurred image, and the results were compared with the DMOS value provided in the database.

Some test results have been provided here that prove that theorem 1 and the metric calculation are accurate.

4.5.1 Experimental proof of the Theorem 1 of 4.3

The first assumption of this work is that when a non-distorted image is transformed by divisive normalization transform, then the distribution of normalized coefficients follows a Gaussian distribution. However, in distorted images, the distribution of normalized transform does not follow the Gaussian distribution.

Table 4.1 includes some statistical measures of the distribution of normalized coefficients for a non-distorted image and its five levels of blurred images.

Table 4.1: Statistical measures for the original image and its five-level of distorted versions

Skewness	-0.0055	0.0079	-0.0283	-0.0455	-0.97	-3.5419
Standard deviation	15.3960	12.9952	10.6232	6.8548	1.95	0.8844
Variance	237.0363	168.8758	112.8523	46.9885	3.8	0.7822
MAD	9.1550	7.7475	6.3438	4.1158	1.0595	0.2529
	Original Image	Blur level1	Blur Level2	Blur Level3	Blur Level4	Blur Level5

The table shows the value of four statistical parameters: skewness, standard deviation, variance, and median absolute deviation. It is known that the skewness value is very near zero for the Gaussian distribution. In the table, the second row shows the skewness of all tested images. It can be seen very clearly from Table 4.1 that the skewness value of the original image is very near to zero; however, as the blurriness increases, the skewness increases. That means the distribution loses its Gaussian shape. The third, fourth, and fifth row shows the distribution's standard deviation, variance, and median absolute deviation. It is shown that when the values decrease, the blurriness increases. The third measure is the measure of dispersion, which means they say how thick the distribution is. The decreasing importance of these measures ensures that the distribution is very thin and looks like a spike. The values mentioned in the table easily prove that our assumption is valid. We have used the CSIQ database for this purpose because, in the CSIQ database, each original image has four to five levels of distortion for each type of distortion. That means one original image has been distorted in four to five levels from distortion. Therefore, it is easy to show the changes in the shape of the distribution by the statistical parameters.

4.5.2 Validation of the proposed metric with the comparison of the subjective quality metric

Table 4.2 compares the Difference mean opinion score (DMOS) value provided in the CSIQ database with the proposed metric values. The DMOS is a subjective quality metric. The DMOS is calculated by calculating the difference between the raw rating score of the original

image and the test image. In the table, the values of five shots are shown. The images are identical but blurred with different levels.

In general, the DMOS value increases with the increment of distortions. In Table 4.1, it is shown that our metric value increases as the DMOS value rises. Therefore, we can conclude that our proposed metric can estimate the blurriness of distorted images. It is also seen that the value of the metric lies between zero and one. The metric's value is nearer to zero in level 1 blur and increases with the blurriness increases, and in the highest level of blur, the value is close to 1.

Table 4.2: Comparison between the DMOS value and the proposed NIBEM metric value

NIBEM (Proposed)	0.2283	0.4448	0.4855	0.6203	0.9790
DMOS	0.081	0.216	0.363	0.571	0.913
	Blur Level 1	Blur Level 2	Blur Level 3	Blur Level 4	Blur Level 5

4.5.3 Validation of the proposed metric with the comparison of the objective quality metric

The correlation coefficient (CC) and Spearman rank-order correlation coefficient (SROCC) were calculated between the subjective score DMOS and the databases and the objective score proposed. Images from both databases have been used for comparison purposes. The CC calculates the prediction accuracy, and the SROCC measures the prediction monotonicity. Then, the CC and SROCC values with existing techniques were compared. The performance of the proposed metric has been compared with the other four existing techniques. Table 4.3 shows the comparison result, and it is seen that our proposed metric is better than other metrics.

Table 4.3: Comparison with existing methods

Database	Metric	JNB [102]	CPBD [103]	BIBLE [104]	Wavelet-based [105]	NIBEM (proposed)
CSIQ	CC	0.8161	0.8955	0.9622	0.9062	0.9711
	SROCC	0.7872	0.9182	0.9622	0.9169	0.9656
LIVE	CC	0.8061	0.8818	0.9403	0.8163	0.9520
	SROCC	0.7624	0.8853	0.9132	0.8010	0.9324

In this chapter, a novel no-reference blurriness metric NIBEM has been proposed. A new definition of blurriness in terms of variance has been included, and a theorem on the behavior of blurriness of the image has been presented in this chapter. The metric can estimate the blurriness accurately as the comparison with DMOS values proves the statement. The NIBEM can be used in different image processing applications like surveillance and security and medical image processing, where the best quality input image is the primary concern.

There is ample scope for further modification of the metric. The median absolute deviation value was used to estimate the metric. Additional statistical analysis of the distribution makes the metric more accurate. In another aspect, statistical analysis of divisive normalization coefficients can classify different types of blurring effects. The other types of distortions can be estimated further according to the proposed method.

Chapter - 5
Rainy Image Processing

5.1 Introduction

In Chapters 3 and 4, two basic image quality assessment techniques, namely full-reference and no-reference IQA, have been discussed, respectively. In these two chapters, the process of determining image quality is discussed in relation to IQA. In addition, a novel image descriptor and related FRIQM have been introduced in Chapter 3, along with a novel NRIQM in Chapter 4. In the two previous chapters, image assessment has been discussed broadly. This chapter particularly focuses on images distorted by environmental stimuli like rain. Rain is a very complex environmental stimulus that can distort an image in a very critical way. Other environmental stimuli like fog, mist, and haze statically distort an image. However, rain dynamically distorts an image. In the case of a rainy image, the water droplets are visible, and the droplets move downward. Additionally, the presence of water droplets decreases the overall image quality. Therefore, processing a rainy image includes two points in consideration: firstly, the presence of water droplets with their motion, and secondly, the distortion caused by the presence of water droplets.

In this chapter, three significant fields of rainy image processing have been discussed. First, the method for assessing the quality of rainy images has been explained, followed by image draining and object detection in rainy images.

Subsection 5.2 contains a discussion on the quality assessment of rainy images. In this subsection, one novel no-reference image quality metric has been introduced that is specially designed to determine the quality of rainy images. This subsection presents a novel method of detecting rainy images based on the IQA metric designed for rainy images. A deep learning-based method for detecting rainy images was also discussed to evaluate the performance of the novel image quality metric. Subsection 5.3 discusses a generative adversarial network (GAN)-based image draining method. The rain strikes have been removed from the rainy images in this process. In subsection 5.4, a deep learning-based method capable of detecting objects in rainy images has been introduced. Visibility in rainy images decreases depending on the severity of the rain. Therefore, it is challenging to detect

objects in those rain-affected images. A You Only Look Once (YOLO)-based method has been introduced to detect objects from rainy images.

5.2 – Image quality assessment for the images degraded by rain

Images can be degraded for two reasons – (i) internal reason and (ii) external reason. Degradation of images due to internal reasons occurred after the acquisition of a scene. Image degradation due to external reasons has occurred before image acquisition. That means the image acquisition system captures scenes degraded by atmospheric conditions. The internal reasons for image degradation include misbehaving of acquisition devices, degradation of images at the time of quantization, or addition of noises during transmissions. The image acquisition system encounters some noises in acquired images. The x-ray acquisition system causes Poison noise.

Similarly, the Ultrasound image acquisition system encounters Speckles noise. The incorrect setting of the acquisition system also results from some noises. For example, defocusing the camera causes blurring and low contrast of images. In addition to that, the image can be generated at the time of transmission, storing, processing, compression, etc.

Noise like blurring, salt and pepper noise, additive white Gaussian noise, and noise due to jpeg compressions are the most common in the degraded images due to internal reasons. In this case, it has been considered that the image had no degradation before acquisition. On the other hand, images are degraded by external reasons where the scenes are already degraded before acquisition. The scenes that are to be captured have already been degraded. The natural images can be degraded mainly by atmospheric conditions like Rain, Haze, Fog, Mist, Poor illumination, Dust, etc. These reasons for scene degradation are entirely natural, and there is no procedure to remove those reasons. Images degraded by weather conditions need some special attention to deal with.

The atmospheric conditions that can degrade the visibility of a scene can be classified into two classes depending on the water droplets in the atmosphere. The two classes of atmospheric conditions [106] are – (a) Steady atmospheric conditions and (b) Dynamic atmospheric conditions. The two classes of atmospheric conditions cause the scenes to have lower visibility.

(a) Steady atmospheric conditions – According to these conditions, the images degraded only in one way. Fog, Mist, and Haze are the atmospheric conditions that fall in this class of atmospheric conditions. In these cases, the size of water droplets present in the air is in the range of 1 to 10 μm, which is a minimal quantity. Therefore, the water droplets cannot be seen by the naked eye. The presence of water droplets in the atmosphere obstructs the scenes. Thus, the visibility of the scenes becomes very poor.

(b) Dynamic atmospheric conditions - The size of water droplets varies between 0.1 and 10 mm. The water droplets are sufficiently large and can be seen normally. Rain and Snow are two atmospheric conditions named Dynamic weather conditions. In both cases, the droplets are seen clearly. In the case of the snow, snowflakes degrade the scene's visibility, and rain strikes are the reason for the degradation of scenes due to rain. Large water droplets create a layer of obstruction in the scenes. Therefore, the objects present in the scene cannot be correctly identified. This atmospheric condition creates another way of degradation besides the large weather droplets. Rain strikes and snowy flakes also add some noise to the scenes.

It is seen clearly that Steady atmospheric conditions degrade scenes in one way, and Dynamic weather conditions degrade scenes in two ways. Therefore, handling images distorted by Dynamic atmospheric conditions is very tough.

This research mainly concentrates on the images degraded by the dynamic atmospheric condition of Rain. A novel technique has been proposed to detect rainy images.

Degradation in images means the degradation in the quality of the image. Each type of degradation causes some specific kind of image quality degradation. Therefore, it is evident that rain causes the same kind of quality degradation in images. Then, image quality analysis can be a way of detecting images degraded by images. Image quality analysis is the area of research where the quality of images can be analyzed. The image quality analysis is of different types: subjective image quality analysis and objective image quality analysis. The subjective image quality analysis was conducted by employing physical human observers. The opinions of human observers about the quality of an image can be used to measure the quality of an image. However, it is a costly process and has very low applicability. The objective image quality analysis was done using computing devices. This approach does not include human observers. This is a less costly process and has very high applicability. However, human beings are the absolute evaluator of image quality.

For that reason, the main challenge of subjective image quality analysis is to evaluate the quality of the image as human beings. The objective quality metric is further classified into three types [107] – Full Reference, Reduce Reference, and No-Reference. In the full reference approach, the original image, and the test image, are both available, and the quality of the test image can be calculated by comparing both images. In reduced reference, the original image is not present. Instead of the original image, some features of the original image are present. These original image features are compared with the same features obtained from the test image, and image quality metrics are computed.

In No-reference image quality metric analysis, original images or any features of original images are not present. The quality assessment test image is completely based on the test image itself. This characteristic of this technique makes it applicable to this research work. Because of the unavailability of the reference image in a real-time system, the no-reference image quality assessment technique becomes the most appropriate choice. No-reference IQA metrics are used to detect rainy images in this work.

A new quality metric has been developed in this work, designed to assess the quality of an image degraded by rain. Three classifiers, namely Support Vector Machine, Random Forest, and k-Nearest Neighbor, have been used to classify the rainy and clear images. MobileNet deep learning model has been used for baseline experiments as well. The rainy image detection system has been validated using a public dataset containing 1400 images, including 700 rainy images. The main methodology of this work has been presented in the following subsection.

5.2.1 Methodology

The system is used to detect rainy images using a feature-based approach. In addition, the detection of rainy images has been done using a data-driven approach. A comparison was then provided between the two approaches. The main research contribution is in the feature-based approach, where a technique has been proposed to detect rainy images using image quality assessment and a no-reference image quality metric specially designed for rainy images. The methodology is depicted in Figure 5.1.
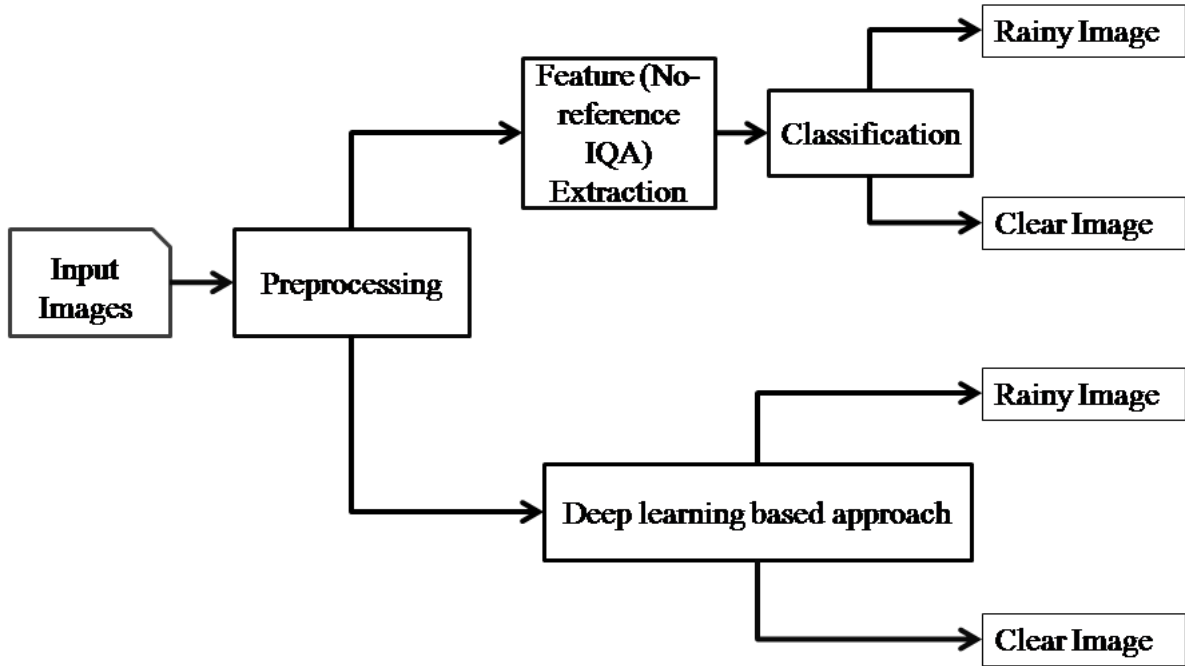
Figure 5.1: Methodology of the rainy image detection system

5.2.1.1 Feature-based Approach

The feature-based Rainy image detection technique consists of two processing steps. The first processing step is used to extract features from the images, and the second processing step is used to classify images according to the extracted features.

5.2.1.1.1 Feature Extraction

No reference image quality metrics have been used as features in this research work. These image quality metrics have been used because these types of metrics cannot depend on the corresponding reference images. According to the full reference image quality approach, the reference and test images are compared directly for quality assessment. The reduced reference quality assessment technique has reduced dependency on reference images. In this approach, instead of actual reference images, features of reference images have been compared with the same features of the test image. Therefore, both of the quality assessment approaches have a dependency on reference images. However, getting reference images for all Rainy images is impossible in the practical application. Then, it is not feasible to use full-reference and reduced-reference image quality metrics as features for detecting images degraded by Rain. The no-reference image quality metric has no dependency on reference

images. Therefore, no-reference image quality metrics are the best choices as features for detecting images degraded by Rain.

Five state-of-the-art no-reference image quality metrics as features and newly created features have been used. The features are discussed as follows.

(i) <u>BLind Image Integrity Notator using DCT Statistics (BLIINDS)</u> [108]

BLIINDS is a distortion-independent, no-reference image quality metric that correlates well with human perception. This metric follows the natural scene statistics model for image quality analysis. The input image is transformed into a Divisive Cosine Transform (DCT) domain and uses the DCT statistics to extract features, and those features have been associated with the help of probabilistic theory to measure the quality metric.

According to calculating the quality metric BLIINDS, the input images have been partitioned into small blocks of n×n. Then, each input image block is treated as image patches and processed with a 2-D DCT transform. This transformation has been done locally in correspondence with the Human Visual System (HVS). Then, the Gaussian density model was applied to each block to fit the distribution of the DCT coefficients. Applying the generalized density model divides the DCT distribution of each block into three different oriented sub-bands – low-frequency sub-bands, mid-frequency sub-bands, and high-frequency sub-bands. Various features related to human perception can be extracted from each block's generalized Gaussian density distribution. The features are related to the shape of the generalized Gaussian distributions, frequency variations, orientations, and energy sub-band ratio measures. The features' final quality score is then predicted using the Bayesian probabilistic model. The quality score helps to identify degraded images. The distortion-independent, fast-processing, HVS-correlated no-reference quality metric is helpful for our purpose. Therefore, we have selected BLIINDS as a feature for detecting images degraded by Rain.

(ii) <u>Blind/ Referenceless Image Spatial QUality Evaluator (BRISQUE)</u> [109]

BRISQUE is another broadly used no-reference image quality metric. This metric is also used in the Natural Scene Statistics to identify the quality score. According to this approach, image quality analysis has been done in a spatial domain. There is no need to transform the test images into other domains like DCT or Wavelet. The image quality analysis is

distortion-independent. No distortion-specific features have been used to measure the quality score. The BRISQUE has very low computational complexity. Therefore, the applicability of the BRISQUE for identifying the distortion type is very high. This is why the BRISQUE is selected as a feature for detecting images degraded by Rain.

The input images have been transformed using Gabor filters. The Gabor filters are of different orientations and various frequencies. The variations in orientations and frequencies correlate the Gabor transformation with human visual perception. The Gabor filters are sensitive to texture identification and discrimination. The Gabor filter has been applied in four different orientations and two different frequencies. Two features from the transformed coefficients are extracted. The two features are the Energy and Entropy of the transformed coefficients. The two features were combined using linear least square regression modeling to find the final quality score.

There must be a texture change in rainy images, and the Gabor transformation is very sensitive to textures. Therefore, we have used the BRISQUE as a feature for our proposed technique.

(iii) Distortion Identification-based Image Verity and INtegrity Evaluation (DIIVINE) [110]

The DIIVINE is another distortion-independent, no-reference quality metric that depends on the Natural Scene Statistics and is measured in two steps. This metric is the most accurate as it can outperform standard full reference quality metrics. Therefore, this metric is an automatic choice as a feature for detecting images degraded by the Rain.

It is a two-step method. The first step of the method transforms the test image into a wavelet domain. In the next step, features were extracted to qualify the test image. The wavelet transform divides the test images into sub-bands. The features have been selected along with the coefficients of sub-bands. A total of six features have been used to calculate the final quality metric score. The features are – the variance of coefficients of all sub-bands, shape of coefficients of each sub-band and all sub-bands, correlation of all sub-bands, spatial correlation of sub-bands, and orientations of the sub-bands.

(iv) Naturalness Image Quality Evaluator (NIQE) [111]

This is another no-reference metric based on the natural scene statistics criteria. This is a no-training-based image quality metric where the quality scores have not been measured through supervised training. This metric measures the changes found in the statistical distribution of degraded images. The test images have been represented using the space domain natural scene statistics model. The features have been extracted from the representation of the images. The statistical representation of the degraded image is different from the statistical representation of normal distortion-free images.

(v) <u>Perception-based Image QUality Evaluator (PIQUE)</u> [112]

This metric used human perception to measure the quality of the test image. The metric has been calculated by processing the local patches of the test image. The test image has been partitioned into n × n blocks. The blocks are treated as image patches. The quality of the image was calculated for each block. The blocks have been classified according to the level of degradation. Then, an overall measurement was done by including the measurement of each block to provide the final metric.

(vi) <u>No-reference image quality metric designed for rainy images</u>

This metric has been calculated in two steps. The first step is the image decomposition in a model where the effect of rain can be extracted, and the second step is feature extraction. According to the features, the final quality metric has been calculated.

Image decomposition

In the case of rainy images, the water droplets are large enough to see through the naked eye, and the water droplets are coming down with a velocity. The dynamic water droplets create rain strikes. The rain strikes present in the rainy images also have different orientations. Because the trajectory of rain strikes depends on the speed and direction of air. We can describe the situation in mathematical terms.

An image can be represented using n^{th} order polynomial I(x,y) in a 2D space. The order of the polynomial of the function I(x,y) represents the texture complexity of an image. A natural image contains objects mainly with complex textures. Therefore, the degree of the polynomial is sufficiently large and cannot be equal to zero.

The effect of rain can be defined on an image as a windowing function. A window function is a function that has a large value for a small interval, and on the outside of the interval, the value of the function is zero. The rain strikes have the same functionality as an image. Intensity values are high for some points where the strikes are visible, and the intensity values are zero for other points where the rain strikes are not visible. Therefore, rain strikes can be expressed as a window function $W_R(x,y)$, where (x,y) denotes the coordinates.

Therefore, the rainy image R(x,y) is expressed as equation 5.1.

$$R(x, y) = I(x, y).W_R(x, y) \qquad (5.1)$$

According to the Adelson and Freeman Steerability Theorems [113], when a function can be expressed as the multiplicative of nth order polynomial and a window function, then the function can be steered in any direction with n+1 number of basis filters where n is the degree of the polynomial.

The steerability formula can be applied to the rainy images. The rainy images can be steered in different orientations using equation 5.2. The application of basis filters highlights the rain strikes present in the rainy images as a vector of coefficients. The generalized Gaussian Density model helps to distribute the coefficients. The presence of rain strikes changes the shape of the distribution of the coefficients.

$$R^\theta (x, y) = \sum_{i=1}^{N} \kappa_i(\theta) R_i^\theta (x, y) \qquad (5.2)$$

Where, $R^\theta(x,y)$ is the steered output of the function R(x,y) in the orientation θ, $R_i^\theta(x, y)$ are N basis functions, and $\kappa_i(\theta)$ are the steer coefficients for a specific orientation θ.

Feature Selection

Features have been selected from the normalized coefficients. Our main goal is to check the normalized coefficients' distribution shape. Therefore, we have selected those features that can identify the distribution of the coefficients. We have extracted two features. They are –

(i)Feature of symmetry –

The statistical parameter skewness is calculated from normalized coefficients of each orientation. Then, the average of all skewness values is taken as features. The skewness measures the symmetry of a curve. Therefore, this feature mainly measures the symmetry of the distribution of the coefficients.

(ii) Feature of flatness –

Another statistical measure, the kurtosis, has been used to understand the shape of the curve. It also tells about the tails of the curve. Therefore, we have calculated the kurtosis of normalized coefficients of all scales and all orientations, and an average of those was computed.

The final metric is measured by averaging the two feature values.

5.2.1.1.2. Classification

The previous subsection discusses the extraction of features. Six features have been extracted from an input image: no-reference image quality metrics. These features can help detect Rainy images. The detection of rainy images is a classification problem. The main goal of the classification process is to separate rainy images from clear images. Therefore, the classification of images is necessary to detect rainy images from clear images. In this subsection, the classification process has been discussed.

This classification problem is a two-class classification problem. The two separate classes are the Clear image and the Rainy image. Three machine learning-based classifiers have been used for the classification task, and the performance of the three classifiers has been compared. The classifiers that have been used are as follows –

- k Nearest Neighbor (K-NN) [114] classifier

- Random Forest (RF) [115] classifier

- Support Vector Machine (SVM) [116] classifier

5.2.1.2 Data-Driven Approach

A data-driven approach has recently been prevalent and has provided excellent performance. Therefore, a deep learning-based technique has been used for this research work. This work has developed a technique to detect images degraded by rain. Rainy image detection is a real-time process, and the response time of the rainy image detection system should be deficient because this technique can be incorporated with security and surveillance systems. As a contradiction, deep learning-based classification techniques are time-consuming and highly complex. It has been decided to use the MobileNet [117] classification model to minimize the contradiction. The MobileNet model is lightweight compared to other deep learning-based classification models. This model can even be used in an Android environment for classification tasks.

The MobileNet Architecture uses a two-stage convolution layer in place of the traditional convolution layer. The standard convolution layer has been divided into a depthwise convolution layer and a point-wise convolution layer. The depthwise convolution layer applies lightweight filtering on each image input channel. The pointwise convolution layer is a 1×1 linear filter to each input layer. The MobileNet version 2 (V2) has been used as a classification model for binary classification purposes and used for our work. This version of mobileNet uses a 3×3 depthwise convolution layer. Including a 3×3 depthwise convolution layer reduces the computational complexity 8 to 9 times from the complexity of the standard convolution layer with adjustable accuracy reduction. The adjustment was made by introducing the linear bottleneck layer. The MobileNet V2 has 32 convolution layers at first. Then, 19 residual bottleneck layers were used. ReLU6 has been used for optimization. Figure 8 shows the architecture of the MobileNet v2 model. In the end, average pooling has been used. The model has been used for a binary classification task. The two classes are clear images and rainy images.

5.2.2 Experimental Result and Discussion

In this subsection, the details of the experiment processes have been discussed. First, a brief description of the dataset used here is presented. The experiment is twofold. Firstly, the feature-based approach was validated, and then the experiment for the data-driven approach was done. The experimental setups for both experiments have been discussed, and the results

obtained from the experiments have also been highlighted. After that, the results from both experiments have been discussed to establish the significance of the developed technique.

5.2.2.1 Dataset

A public dataset [13] of rainy images has been used for the experiment. The dataset has been hosted by the Vision and Image Understanding (VIU) lab of Johns Hopkins University. The dataset has been created with the help of two other datasets – the UCID and BSD-500. This dataset consists of 1,500 images. The dataset consists of 1,400 images as a training set and 100 as a testing set. In the training set, there are 700 clear images of nature and 700 rainy images corresponding to the clear images. Therefore, each clear image has a corresponding rainy image. The test set contains 100 natural images. Figure 5.2 includes some of the clear images and corresponding rainy images. We have selected this dataset because this dataset has images of the same scenes for a clear day and a rainy day.
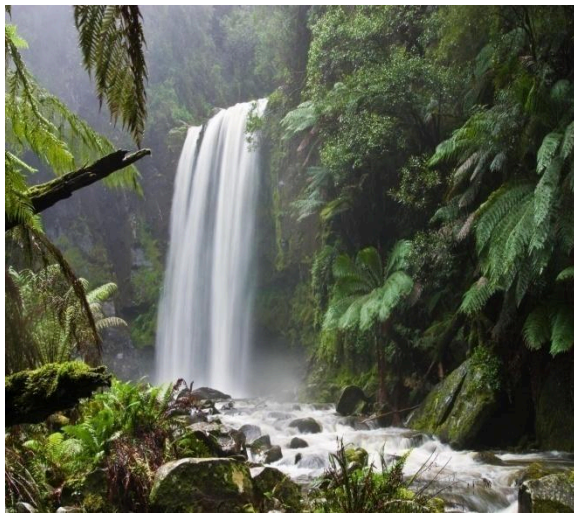
a) Clear Images   b) Rainy Images

Figure 5.2: Images of Dataset hosted by Vision and Image Understanding (VIU) lab

5.2.2.2 Experiment of Feature-based Technique

This experiment was conducted to classify clear images and rainy images. This is a two-class classification problem. The two classes are (i) clear images and (ii) rainy images. The class of clear images has been labeled as class 0, and the class of rainy images has been labeled as class 1. The features have been extracted from the images belonging to the two classes. Each feature vector has six entries. Therefore, each image is represented by a vector of seven values – the class label and the six extracted features. The image description vector has been calculated for all images present in the dataset. This image description data has been passed into the classifiers to train the classifiers.

Three classifiers have been used for classification purposes. At first, the k-NN classifier was used to classify the clear and rainy images. The number of initial neighbors was considered 3, and the distance measure used between the neighbors was the Euclidian distance measurement. After that, a Random Forest (RF) classifier was used to classify clear images and rainy images. It is well known that the Random Forest classifier uses the bagging tree method for classification. A bagging tree with a value of 50 has been used for this experiment.

Finally, the clear and rainy images were classified using a Support Vector Machine (SVM) classifier. A linear SVM classifier has been used for this experiment. The dataset has been partitioned in 7:3 manners, where 70% of the whole dataset has been used for training the classifiers, and 30% of images have been used for testing.

The performance of the feature-based technique was evaluated for the detection of rainy images by the classification error calculation of each classifier, and the average classification error of the three classifiers was the final evaluation score. The classification error rate has been calculated using two parameters: False Acceptance Rate (FAR) and False Rejection Rate (FRR). The FAR represents the percentage of rainy images classified as clear images, and the FRR denotes the percentage of clear images classified as rainy images. The two parameters, FAR and FRR, are combined to calculate another parameter named Half Total Error Rate (HTER). The HTER is calculated as the average of two parameters, FAR and FRR. Therefore, HTER = (FAR+FGR)/2.

Table 5.1 displays the results obtained from the experiment. The classification operation has been done two times. The number of rows signifies the number of experiments. In the first experiment, five features have been considered: no-reference image quality metrics, namely – BLIINDS, BRISQUE, DIVIINE, NIQE, and PIQUE. The second experiment has been conducted by including the metric proposed in this research work with the five previously mentioned features. Therefore, in the first experiment, the classification was done using five numbers of features, and in the second experiment, the classification was done using six numbers of features. Table 5.1 includes all FAR, FRR, and HTER parameters for each classification process conducted using k-NN, RF, and SVM classifiers for both experiments. The experimental results show that including the proposed metric significantly reduces the classification error rate.

Table 5.1: Comparison of results obtained from two experiments

IQM Features	k-NN			RF			SVM		
	FAR	FRR	HTER	FAR	FRR	HTER	FAR	FRR	HTER
1st Experiment	0.15	0.2	0.175	0.1	0.12	0.11	0.08	0.09	0.085
2nd Experiment	0.12	0.18	0.15	0.08	0.11	0.095	0.07	0.085	0.0775

5.2.2.3 Experiment of Data-driven Technique

At first, the training process was discussed. Tensorflow and Keras framework have been used for this experiment. A total of 1400 images have been used for each classification task. The dataset has been divided into 80:10:10 divisions. 80 % of the images were used for training, and 10% were used for validation and testing. Data augmentation techniques like horizontal flip, zoom, shearing, and rotation by 90° have increased the number of training samples.

Two classes of data were used for each of the four experiments. The class of clear images has been denoted by 0, and the class of rainy images has been denoted by 1. The MobileNet V2 has been trained with the images classified as 0 or 1. The images have been reshaped to adjust the dimension of the tensor. The dimension of the input tensor for our work is 224 × 224 × 3. That means the dimension of the reshaped image is 224 × 224. The number 3 denotes the number of color channels in the input images. RGB images have been used here.

Therefore, the color channels of our image are 3. The reshaping process is required to maintain the acceptable tensor dimension of the MobileNet v2 model.

The model has been trained for 100 epochs with batch size 8. The training process has two stages. In the first stage, 50 epochs were evaluated, and an uneven training and validation accuracy and loss graph was obtained. Then, all layers were frozen with the previous weights. Then, in stage 2, the saved model was trained for another 50 epochs with batch size 8. Figure 5.3 shows the validation, training accuracy, and loss graph according to the two-stage training process. All graphs show that the graph in stage 2 is more stable than the first stage. The MobileNet v2 model has been iterated 100 times because the best possible result comes for this number of iterations. The batch size 8 has been maintained because our database is not big enough. Therefore, it is necessary to train the model with the best possibilities.

a) Training Accuracy Vs. Validation Accuracy (Stage 1)	b) Training Accuracy Vs. Validation Accuracy (Stage 2)
c) Training Loss Vs. Validation Loss (Stage 1)	d) Training Loss Vs. Validation Loss (Stage 2)
Figure 5.3: Results of the training process in the Data-driven technique	
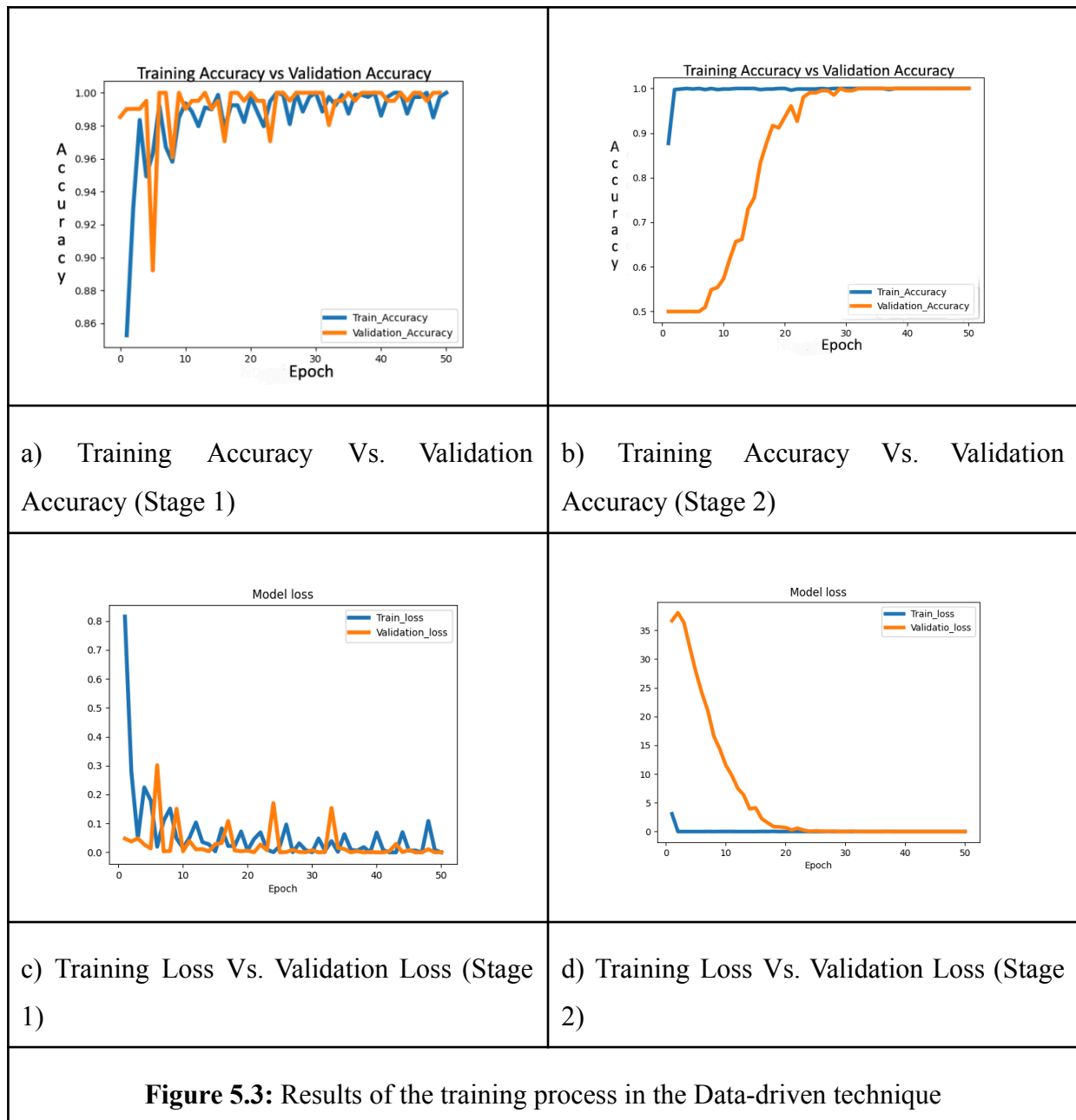
Testing process:

The trained model has been tested with 10% of images of the total dataset used for a single experiment. A confusion matrix was created for testing operations, and the experiment was evaluated using the metric accuracy of the classification task. The classification error rate was measured using the parameters of feature-based techniques: FAR, FRR, and HTER. Table 5.2 consists of the values of the parameters. It has been found that the classification error for the data-driven method is similar to the value obtained from the feature-based technique.

Table 5.2: Experimental results of the data-driven approach

	FAR	FRR	HTER
Experiment with a Data-driven approach	0.08	0.09	0.085

5.3 – Image De-raining

Video surveillance is a very crucial area of research. Accuracy is the primary criterion for this type of research because the video surveillance system ensures security issues. A less accurate surveillance system may lead to mass causalities in recent terrorism threatening the world. However, surveillance cameras are installed mainly in open areas. Therefore, the acquired data can be suffered from environmental stimuli. The degraded data lowers the performance of the surveillance system. It is necessary to restore the surveillance data for accurate analysis. This research work mainly concentrates on video surveillance data affected by rain. In this work, degraded rainy surveillance video data has been de-rained using the Generative Adversarial Network (GAN) model. The proposed model performs better than other types of CNN-based de-raining algorithms. Under dynamic conditions, rain-affected video surveillance data is the primary concern of the present work. This work uses a Generative Adversarial Network (GAN) [119] to de-rain surveillance video data. The GAN has been used to de-rain single images. Rain-affected video surveillance data have been de-rained, which is a very new work in this field. The work is inspired by the work of ID-CGAN [120] and has created a very similar model to de-rain video surveillance data. The model has been evaluated with a created dataset. The model performs well in comparison to other CNN-based de-raining algorithms.

5.3.1 Methodology

I. Goodfellow created a new research orientation in computer vision and pattern recognition by proposing the Generative Adversarial Network [119]. The GAN is mainly designed to generate data from random noise. The GAN consists of two networks – the generator and the discriminator. The discriminator tries to understand the data distribution of real images, while the Generator tries to generate data similar to real data from random noise. The discriminator

network evaluates the performance of the generator network and guides it in generating data similar to the real one.

5.3.1.1 The basic idea behind the work

Although the GAN was developed to generate new data, it can be used in image restoration. In this work, the discriminator model has been trained with video data of normal weather. The discriminator module learns how data is distributed from real data. Then, the rainy video frames are inserted into the generator module. The generator module tries to generate similar real data. The discriminator module is trained first, then the generator module. The generator module produces rain-free data after a couple of iterations. Figure 5.4 depicts the scenario of the working principle of the model.
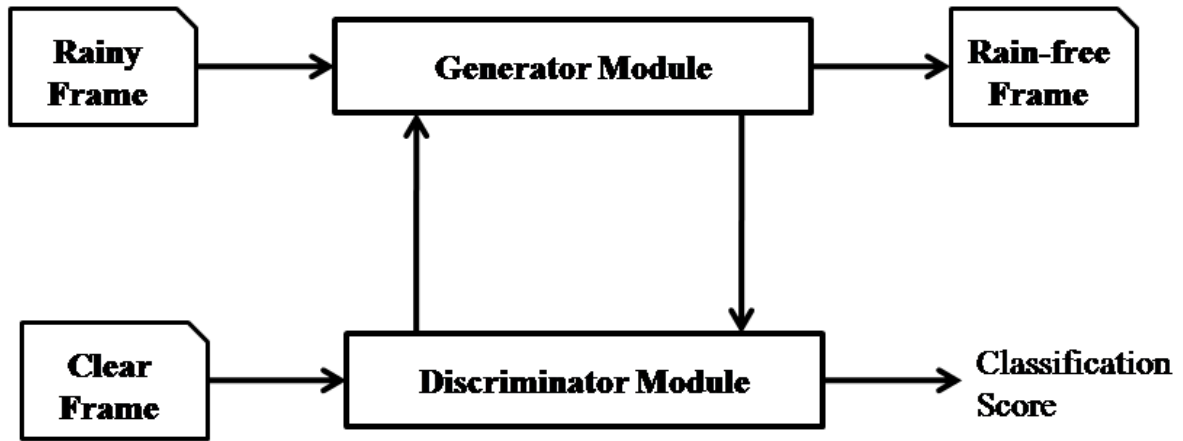
Figure 5.4: Basic working principle of the method

5.3.1.2 Network Architecture

The network architecture is very similar to the work proposed in [120]. The basic network used here is CGAN. The generator module consists of densely connected symmetric convolution networks. The discriminator module used here is a multi-scale network. Refined perceptual loss is used to stabilize the training procedure for GAN.

5.3.2 Result & Discussion

In this experiment, video data captured by a surveillance camera was used. Frames are extracted from the video data. A total of 527 rainy and 541 rain-free video frames have been used for training. About 250 video frames each for rainy and rain-free video frames are used

for testing. Figure 5.5a shows a rainy video frame, and Figure 5.5b shows a corresponding de-rained output. The model's output has been compared with the other state-of-the-art CNN-based de-raining techniques. Table 5.3 shows the comparison results. The comparison result ensures that the proposed model outperforms other de-raining techniques in the case of this video surveillance data.

Table 5.3: Comparison with state-of-the-art techniques

	Proposed Method	De-rain Net [121]	DID-MDN [122]
PSNR	**23.93**	22.09	21.28
SSIM	**0.8152**	0.7856	0.7265

(a) Input video frame   (b) Output video frame

Figure 5.5: Input and corresponding output video frame

5.4 – Object detection in rainy images

In recent years, terrorism has become a headache for security personnel around the world. In 2019, until August already, 1259 attacks and 5,436 casualties have been reported. In 2018, the number of attacks was 2,180, where the number of casualties was 9,776. The statistics help to understand the alarming situation in the globe. However, countries take necessary actions to prevent the nuisance, and such fatal activities cannot be stopped. Nowadays, all crowded public areas are under video surveillance to prevent terrorist attacks. Each

movement of objects, such as persons, vehicles, and animals, is monitored thoroughly for 24 hours. Manual monitoring of any public area is a tedious job, and there may be a possibility for misrecognition. Therefore, automation is needed for monitoring. An automatic surveillance system is needed to monitor public areas. Hence, research in this domain has become popular. Much work has been done in this area [123-129]. This object detection research applies traditional image processing techniques in the first stage. However, tremendous progress in deep learning forces researchers to apply deep learning in this domain. Deep learning seems useful in this domain.

There are some deep convolutional models for object detection present in the literature. The models are mainly classified in two ways – (i) two-stage detection and (ii) one-stage detection. In the case of two-stage detection, objects can be detected in two steps. In the first step, the probable object can be identified. In the next step, the bounding box is predicted. Finally, at the optimization stage, the optimal bounding boxes are adjusted around the objects. The images are divided into several grids in a single-stage object detection model. The grids containing objects are highlighted.

At last, desired objects are detected. RCNN [130], Fast RCNN [131], and Faster RCNN [132] are examples of two-stage object detection, whereas YOLO [133] and SSD [134] are examples of one-stage detection. These models can detect an object very accurately.

The object detection models are designed for object detection in reasonable visibility conditions. However, the scene's visibility may be degraded in a real-time application. Most of the surveillance cameras are installed in open spaces. Therefore, weather conditions affect vision. According to physical properties and effects, the weather conditions can be categorized into two categories – (i) steady and (ii) dynamics. Fog, Haze, and Mist fall in the category of steady. In this case, the sizes of water droplets are small and vary between 1-10 μm. The droplets cannot be seen individually through human eyes. However, the presence of tiny droplets changes the intensity of the image. Many models are proposed to describe those effects in the image.

In the case of dynamic weather conditions, the size of the droplets present in the air is between 0.1 – 10 mm. Therefore, the droplets can be seen through cameras. Rain and Snow come into the dynamic weather degradation category. In this case, the droplets come down at high velocity with complex trajectories. Motion blur is a widespread effect in this situation.

In the case of steady weather degradation, the vision is affected by the aggregated effects of many droplets. Therefore, modeling such weather conditions is comparatively easy. On the other hand, modeling dynamic weather conditions is complicated because each droplet changes vision with its different velocity and trajectory.

In a country like India, rain is more widespread rather than snow. Therefore, we are concentrating on rainy weather conditions. Each raindrop is spherical in shape. They work as a lens. The lights are refracted and reflected through the droplets, which create sharp intensity patterns in the image. The raindrops are in motion in times of rain. Depending on the velocity, the shape of the droplets changes. According to Beard and Chuang [1], the shape of the raindrop can be defined as

$$r(\theta) = a\left(1 + \sum_{n=1}^{10} c_n \cos(n\theta)\right) \tag{5.3}$$

where a = undistorted shape, c_n = coefficients depending on the radius, and Θ = polar angle of elevation. The changes in a droplet shape can observe different types of distortions. The distortions caused by rain are categorized in two ways – (i) the presence of rain strikes. This can happen due to the Spatio-temporal correlation of many falling water droplets on the image plane and (ii) the visual effect of rain. The droplets cause the intensity changes.

The discussion until now makes the readers feel that rain causes a complex degradation in vision. Therefore, object detection in rainy conditions is also a very tough job. As discussed earlier, many deep learning-based object detection models are present. In our case, we want to make a real-time system. Therefore, we are considering the object detection model, which can detect an object in less time. The YOLO model suits our criteria rather than the other models. The reason why we choose the YOLO model over other models is as follows.

1. Very Fast model – The YOLO model is high-speed to detect objects. The model's training and testing time is faster than the other models.

2. Real-time usability – The YOLO model can quickly be used in real-time applications because of its fastest computation ability. A pre-trained YOLO model can detect objects from webcam streaming at the current time.

3. Applicability – The YOLO model is convenient for system applications. The model is high-speed, and its ability to detect objects in real time makes it worthwhile in an application domain.

4. Classification Score – The model can detect an object by drawing a rectangular box. Besides, the model calculates the classification score and shows this score with the rectangular box. The classification score is the probability of identifying an object in a particular class. The classification score is useful in application because, according to the score, an application system can decide the success of detecting an object in that domain.

The YOLO model has been utilized in this research because of the abovementioned advantages. The YOLO model is mainly designed to detect objects in reasonable visible conditions. In this research work, the YOLO model has been used to detect objects in rainy conditions. The YOLO v3 model has been modified using the inductive transfer method to detect an object in rainy conditions.

5.4.1. Methodology

The methodology of this research work has been discussed in this subsection. In this work, objects distorted by rain are detected using video data. The input data is directly fed into the network for training. No pre-processing has been done. The YOLO v3 [133] model mainly inspired the basic deep learning block.

5.4.1.1 Basic Idea

The YOLO model treats the object detection problem as a regression problem. The model contains one convolutional neural network for detecting objects that can be trained end to end. A single image entirely feeds into the network. The network divides the image into grids of different sizes. If the center of an object falls into a grid, then the grid is responsible for detecting the object. Each grid further predicts the B number of bounding boxes and the confidence scores of each box. The confidence score is the probability of a box containing an object. Confidence is defined as $Pr\ Pr\ (Object) * IOU_{pred}^{truth}$. The Pr(Object) is the

probability of the presence of an object, and IOU is the intersection over the union between the predicted box and ground truth data. The confidence score is zero if the bounding box contains no object. Each bounding box contains five predictions – (i) X-coordinate, (ii) Y-coordinate, (iii) Height, (iv) Width, and (v) Confidence. Each grid cell computes a class probability Pr $Pr\ (Class_i|\ Object)$. At the time of testing, the class probability is multiplied by the confidence score of each bounding box of that grid. That provides the classification accuracy of each object. Equation 5.4 shows how the classification score is calculated.

$$Pr\ (Class_i|\ Object\ *\ \mathrm{Pr}\ Pr\ (Object)\ *\ IOU^{truth}_{pred} =\ \mathrm{Pr}\ Pr\ (Class_i)\ *\ IOU^{truth}_{pred} \qquad (5.4)$$

5.4.1.2 Network Architecture

The basic YOLO model has 24 convolutional layers and two fully connected layers. The network consists of 1 X 1 reduction layers followed by 3 X 3 convolutional layers. The basic YOLO model is further modified for better accuracy and speed. In our work, we have used the YOLO v3 model [28]. This model is more robust and accurate than the basic YOLO model. The YOLO v3 model is a hybrid of some models. They are YOLO v2, Darknet-19, and some portion of the residual network. There are shortcut connections in the layers. This network uses 3 X 3 and 1 X 1 convolutional layers like the basic YOLO model. However, the total number of CNN layers used in the YOLO v3 model is 53, much higher than those used in the YOLO basic model. The YOLO v3 is more powerful and accurate than the YOLO basic model. Instead of containing more layers than the basic model, the YOLO v3 is faster than the basic model. Table 5.5 describes the details of the network architecture of the YOLO v3 model.

5.4.1.3. Transfer Learning

Transfer learning is an approach of machine learning where one model generated for a specific task can be used in another task. This technique ensures the multi-tasking of a generated model. The form of transfer learning used in deep learning is called inductive transfer. This approach uses an existing deep neural model in a different but related task. In this work, we are using a pre-trained YOLO v3 model. Our experimental database is minimal according to in-depth learning requirements. Therefore, we are using the pre-trained model. The images distorted by the rain further train the pre-trained model. Deep learning-based approaches are data-driven procedures. The model gradually learns from training data

samples. We have tried to learn the YOLO v3 model by rain distorted images. We have chosen the YOLO v3 model over the basic YOLO model because the YOLO v3 model consists of Residual layers, as shown in Table 5.4. The layer computes de-convolution operations, which up-samples the tensor. As a result, the data that can be missed by downsampling is regained further. This mechanism is handy in the case of object detection from distorted images because the distorted images contain very little information. Therefore, it is necessary to upsample the tensor periodically for proper results.

Table 5.4: Network Architecture of YOLO v3 model

	Type	No. of Filters	Size of filters	Output
	Convolution	32	3×3	256×256
	Convolution	64	3×3 /2	128×128
1 ×	Convolution	32	1×1	
	Convolution	64	3×3	
	Residual			128×128
	Convolution	128	3×3/2	64×64
2 ×	Convolution	64	1×1	
	Convolution	128	3×3	
	Residual			64×64
	Convolution	256	3×3/2	32×32
8 ×	Convolution	128	1×1	
	Convolution	256	3×3	
	Residual			32×32
	Convolution	512	3×3/2	16×16
8 ×	Convolution	256	1×1	
	Convolution	512	3×3	
	Residual			16×16
	Convolution	1024	3×3/2	8×8
4 ×	Convolution	512	1×1	
	Convolution	1024	3×3	
	Residual			8×8
	Average pooling		Global	
	Connected Layer		1000	
	Softmax Classifier			

5.4.1.4. Training of the Network

The final layer of the architecture predicts the bounding box and calculates class probability. Therefore, the linear activation function is used for the final layer. The Leaky rectified linear activation function is used for other layers. Equation 5.5 defines the Leaky rectified linear activation function.

$$\emptyset(x) = \begin{cases} x, & if\ x>0 \\ 0.1x, & Otherwise \end{cases} \tag{5.5}$$

This model uses a sum square error in the output for optimization. It equals weight errors for large boxes and small boxes. Therefore, the square root of the bounding box width and height are directly predicted instead of the width and height. The model predicts multiple bounding boxes from each grid cell. During training, the model tries to find one box predictor to detect objects. The predictor with the highest IOU with the ground truth is responsible for the detection of objects. This is the way to specialize detectors. The predictors learn more and more features and become accurate through several pieces of training. Equation 5.6 shows the multipart loss function used for optimization during training.

$$\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \pi_{ij}^{obj}(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \pi_{ij}^{obj}(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + \tag{5.6}$$

5.4.2. Result and Discussion

The model has been trained with video frames extracted from the video dataset. The dataset consists of surveillance videos of rain for 3-5 minutes each. The videos are captured in rainy conditions during the day and night. The database consists of 600 video clips.

Frames have been extracted from the video clips. Then, ground truth was created using the LabelImg [134] tool. The network has been trained with frames extracted from videos. A total of 10,000 frames have been used to train the network. The model has been tested using rainy video clips. The developed model correctly detects most objects present in the video clips. A bounding box is drawn on the detected objects. The colors of bounding boxes are specific for specific classes. The bounding box is blue for the Person; the box is Magenta for the car and violet for the motorcycle. Classification scores of detected objects are shown with

the bounding boxes. The value of the classification score is between 0 and 1. Figure 5.6 shows some screenshots of the detection of objects from videos using our model. The figure shows that the model can correctly identify objects in both daytime and night-time.

The precision and recall were measured using the output. These performance evolution metrics are useful to predict the performance of a system. The Precision and Recall have been calculated by equations 5.7 and 5.8 where FP = False Positive, TP = True Positive, and FN =False Negative.

$$\Pr ecision = \frac{TP}{TP + FP} \qquad (5.7)$$

$$\mathrm{Re} call = \frac{TP}{TP + FN} \qquad (5.8)$$

The precision and recall value of the model (YOLO v3) has been compared with the basic YOLO model. Table 5.5 shows the comparison. It is shown that the proposed model inspired by YOLO v3 performs much better than the basic YOLO model. Therefore, we can conclude the discussion by stating that our proposed deep learning-based model with YOLO v3 can detect objects successfully from video frames distorted by rain.

Table 5.5: Comparison of basic YOLO and our proposed model

	Basic YOLO	Proposed Model Inspired by YOLO v3
Precision	78.44%	89.23%
Recall	74.33%	84.65%
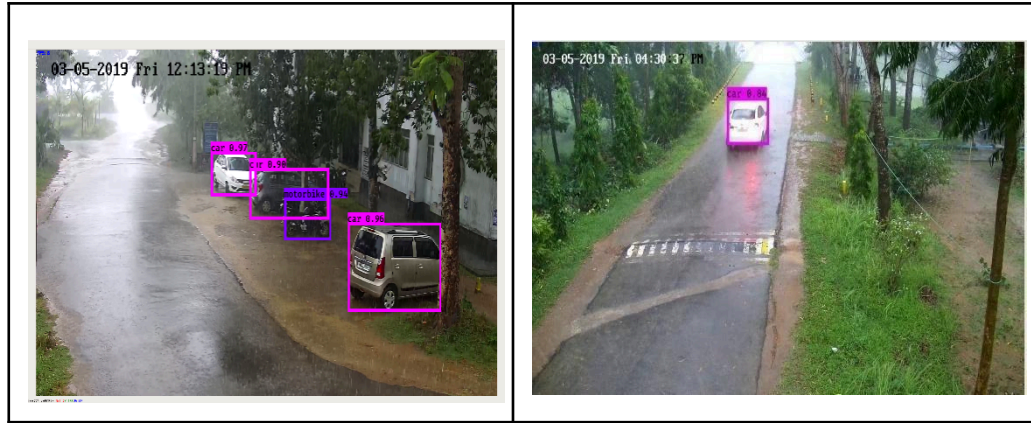
a. Objects detected in the daytime
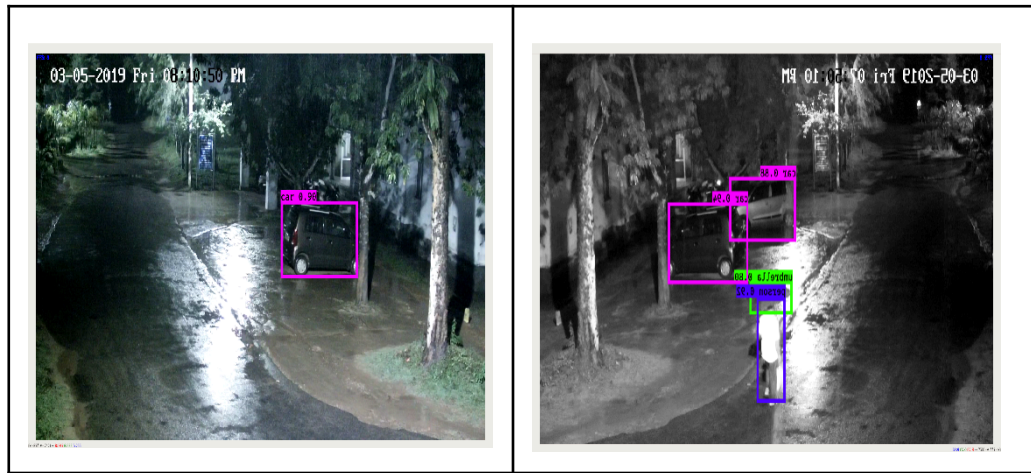
b. Objects detected in the nighttime

Figure 5.6: Screenshot extracted in time of testing from video clips

Degradation of images due to natural hazards like rain, fog, snow, mist, etc, is common when images and video have been captured in outdoor applications. Rain and snow are the two natural hazards that degrade the scenes very complexly. Due to rain and snow, scenes have been degraded in two ways – the raindrops and snowflakes present in the scenes and the scenes have been degraded with blurriness. The complex degradation model obtained from rain and snow makes the researchers interested in contributing to this research area.

This chapter's primary research is on the scenes degraded by rain. Three sub-areas of rainy image processing have been employed here. At first, rainy images were detected using the No-reference Image Quality assessment technique. A novel no-reference image quality metric has been proposed here to analyze the quality of a rainy image. The feature-based technique finds features from the images and uses the features for the classification of rainy

images from clear images. In this work, five state-of-the-art no-reference image quality metrics and the proposed image quality metric were used as features.

The classification process includes three classifiers: k–NN, RF, and SVM. The feature-based technique has been evaluated based on the classification error rate. Including the proposed metric with the other five state-of-the-art metrics for classification tasks minimizes the classification error rate. This observation established the significance of the proposed metric, and the classification error rate obtained from the experiments shows the intent of the proposed technique. A data-driven method for the detection of Rainy images has also been used. MobileNet V2 model has been utilized for this purpose and evaluates the process using the same parameter called classification error rate. As we all know, the data-driven method is superior to traditional feature-based methods. However, the performance of the feature-based technique is very similar to that of data-driven methods. This technique has a prominent application in the field of security and surveillance. Although there are many scopes of further development, the performance of the proposed technique fulfills the purpose of detecting rainy images very well.

After that, a method for de-raining was presented. The technique used to remove the rain strikes from the rainy scenes captured using a surveillance camera. This method utilizes a Generative Adversarial Network (GAN) for the de-raining operation. Although the GAN is designed to generate new data, it performs well in video de-raining. The de-raining method using the GAN network presented in this chapter proves better than the other CNN-based networks. However, this method consists of some loopholes that can be resolved to create a more accurate method.

Along with detecting rainy images and de-raining operations, other aspects of rainy image processing have been incorporated in this chapter. Detection of objects from rain-degraded scenes captured through surveillance cameras is an exciting area of research. Here, a deep-learning-based, data-driven system has been presented. CCTV video footage is the system's input and can detect objects successfully in rainy conditions. Object detection in normal vision is a broad research area that many researchers have contributed to. However, object detection in rainy conditions is an untouched paradigm. Much work has been done on rain image modeling, image de-raining, etc. However, no specific work on object detection is found in the literature. In this presented method, the YOLOv3 model has been utilized as the basic model, and the transfer learning approach has been used to detect an object from videos

in rainy conditions. This contribution is significant in this area and has obtained excellent results. However, such complex conditions exist where the presented system fails to detect objects. These failures must be part of future research on the current method.

This chapter illustrates the research area of rainy image processing. This research area has several sub-areas; the three most essential sub-areas, detection of rainy images, de-raining of rainy surveillance video frames, and detection of static and moving objects from surveillance video frames, have been concentrated. Separate methods for each sub-area have been presented as research contributions throughout the chapter. Among them, one method uses the feature-based process, and the other two are based on a deep learning-based model. The significance of the research areas and the applied methods makes the chapter most exciting and valuable. This chapter provides contributions in application areas along with contributions to image processing.

Chapter - 6

Image Quality and Spoofing Attack

6.1 – Introduction

In Chapters 3 and 4, two image quality assessment techniques, namely Full-reference IQA and No-reference IQA, have been discussed, respectively. Chapter 5 focuses on rainy image processing, which includes three significant subfields: quality assessment of rainy images, drainage, and object detection in rainy images. This chapter discusses two major applications: spoofing detection and medical image processing of the IQA.

Biometric systems are now used in almost all organizations for authentication purposes. However, the biometric system is always under the threat of spoofing attacks, which are unauthorized attempts to break the security of a biometric system. Spoofing is a very challenging aspect of a biometric system. Therefore, a separate process has to be included in the biometric system to protect it from spoofing attacks, and IQA techniques are very applicable to handling such spoofing attacks. In subsection 6.2, a novel spoofing attack prevention technique using the IQA technique for processing has been discussed.

Medical Image processing is a significant area of research related to diagnosing different diseases from different image modalities. IQA has a very prominent scope of application in diagnosing different diseases. In subsection 6.3, a novel technique to detect breast tumors, using the no-reference IQA technique, is introduced.

This section mainly focuses on the application areas of the IQA, and the following subsections discuss some applications of the IQA technique. Subsection 6.4 provides an overall discussion of the chapter.

6.2 –Spoofing Detection using Image Quality assessment

Biometrics is a secure tool for identity verification in different government, industrial, and forensic applications. The physiological and behavioral biometric modalities have been established and ameliorated over the decades [135]. While biometric authentication renders security against illegitimate identities, it is susceptible to spoofing attacks. In this presentation attack scheme, an imposter replicates a genuine user's biometric trait(s) to represent a legitimate person. Different spoofing methods (e.g., synthetic, manipulated, or reconstructed

traits, etc.) have been attempted on various modalities to undermine the reliability and security of the biometric systems [136]. A biometric system's legitimacy is compromised by different presentation attacks, also known as direct attacks [137].

As spoofing is a severe challenge to biometrics, necessary countermeasures should be addressed to maintain individualization privacy. The researchers have emphasized various forms of presentation attack detection (PAD) techniques, mainly for unimodal and multimodal environments with the face [137,138], fingerprint [139, 140], palm print [141], iris [142], speech [143, 144], and other traits. Several hardware and software-specific fundamental techniques are deployed for PAD in the biometric contexts. In the hardware-based method, an expensive sensor is employed for image acquisition from which the constraints about liveness, such as the thermal facial map, blood pressure, perspiration, hand-vein temperature, and other salient features are computed [145, 146, 147, 148, 149]. Technological advancements of the sensor devices offer to acquire thermal [150, 145, 146, 147, 149] and hyperspectral [151] images, which direct to devise high-security measures such as liveness detection to discard fake hand biometric samples. On the contrary, the software-based method discriminates between genuine and counterfeit images using a computable characteristic, such as visual quality assessment [152, 153, 154, 155, 156, 157, 158, 159]. Software-oriented methods are more advantageous than hardware (i.e., sensor-specific) approaches regarding the cost of sensors, and the former can be incorporated easily into a deployed authentication system. In general, the printed photos of a trait (e.g., face [137, 138]), synthetic biometric samples (e.g., silicone fingerprint [152], synthetic speech [160], etc.), and electronic screen display (e.g., iris images captured by an iPad [142]) are established spoofing vulnerabilities. Table 6.1 summarizes the findings.

Though hand geometry and other hand-based modalities are useful for identity verification in industry attendance maintenance and legal and forensic purposes [159], they are also exposed to direct attacks [161]. Spoofing attack detection (SAD) using the thermal hand image has been explored in [150]. Individual authentication is not adequate to strengthen the security of a deployed hand biometric system. Also, liveness recognition is a vital task before the actual authentication of a licit user. Hence, there is ample scope to develop a SAD method to deter fraudulent endeavors using hand images before a biometric authentication task, which is the main objective of this work.

Image quality assessment (IQA) is preponderantly applicable in several tasks of image processing, such as image compression, image enhancement, super-resolution, and others [162, 163, 164]. IQA is also suitable for software-based spoofing detection using biometric samples [136, 137]. Among several standard image quality metrics (IQM), gradient magnitude plays a vital role in describing the visual quality of generic images [162,166]. It is based on intensity variations in the edge maps. This metric has been addressed in various state-of-the-art approaches to discriminate between the original and distorted images [166].In this work, we have considered a threshold parameter to offer flexibility in gradient computation. The main advantage of this modification in gradient computation with thresholding is considering the extent of intensity difference between neighborhood pixels. This work presents the modified gradient similarity, i.e., threshold-based gradient magnitude similarity (GMS), for quality assessment in the context of anti-spoofing on hand biometrics. The proposed framework follows an electronic screen display-based anti-spoofing technique to hinder the susceptibility of attack using a fake hand, which is presented to a deployed verification system. This advanced system is pictorially ideated in Fig. 6.1. For experimentation, a fake hand dataset with natural degradation is created using a Canon EOS 700D camera from 255 subjects from the Bogazici University (BU) hand dataset [167, 168]. Three original and corresponding fake samples of an individual are shown in Fig. 6.2a and 6.2b.

Moreover, we have also created a fake dataset comprising the same population with artificial degradation by introducing three types of noises, namely, Gaussian blur, salt and pepper, and speckle noise. Samples of degraded hand images with these noises are shown in Fig. 6.2c. Ten full reference quality measures are computed for experiments using machine learning and deep learning techniques.

Table 6.1: Common biometric traits for spoofing attack, detected with software-based methods

Attack method	Biometric trait(s)	Spoof sample creation methodology
Printed photo [136,137,138,142]	Face, iris, fingerprint, Palm print	A high-quality printout of the trait is displayed to the sensor. The success of the attack depends on the quality of printed photos.
Electronic screen display [142]	Iris	The artefact is produced using a sensor from an enrolled real sample, displayed in an electronic display device such as an iPad (4th generation) or a Samsung Galaxy Pad.
Electronic screen display (**This work**)	Hand geometry	The fake hand dataset with natural degradation is created using a Canon EOS 700D camera from the original sample displayed on a laptop. Another fake hand dataset is made with additional artificial degradation with the Gaussian blur, salt and pepper, and speckle noise.
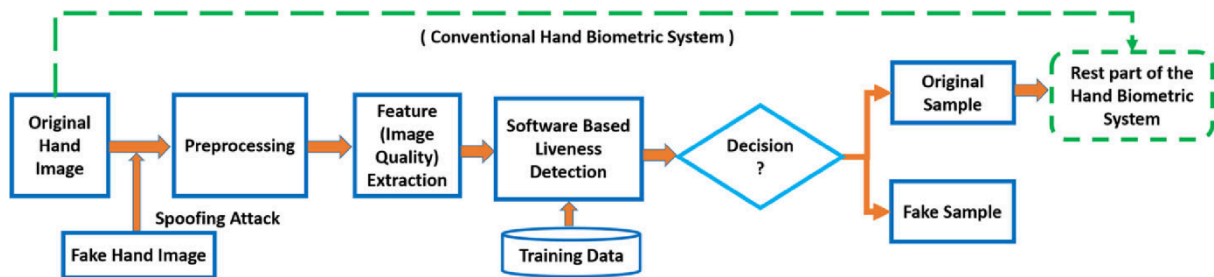
Figure 6.1: The proposed anti-spoofing method using hand images. It is pertinent before conventional biometric authentication (specified with a dotted line). A traditional biometric system considers a valid image to discriminate against legitimate and zero-effort imposters other than fake sample detection. Our proposed scheme uses The feature extraction module for conventional hand-crafted feature extraction or deep feature extraction using theMobileNetV2 [169] base CNN.

a

b

c

Figure 6.2: a Sample of original hand images of a subject (top-row), originally captured by a HP Scan jet 5300 c scanner at 45-dpi. **b** Fake hand images with natural degradation acquired from corresponding original hand images (middle-row) using a Canon EOS 700D camera at 72 dpi. **c** Fake hand images with artificial noise degradation (bottom-row) by introducing: (left) Gaussian blur, (middle) salt and pepper, and (right) speckle noise. Two fake samples are shown column-wise for each type of added noise.

The contributions of this work are summarized as follows:

– A new visual quality metric is proposed based on the gradient magnitude similarity. A total of ten quality metrics are computed to distinguish an original hand image from a fake sample.

– Two types of fake hand datasets are created with natural and artificial degradation for spoofing detection using the original images from the Bogazici University hand dataset.

– The proposed method is evaluated using conventional and deep learning methods for spoofing detection with various experimental scenarios.

– The experimental results signify that presentation attack detection using quality assessment can be effective for a hand biometric authentication system. This proposed method can reduce the susceptibility of a hand biometric system regarding spoofing detection.

6.2.1 Implementation

The visual quality of an image entails several fundamental characteristics of the image itself rather than its context. The generic attributes of an image are very significant for quality assessment [162], with applications in computer vision and image processing. Image quality measurement can be classified as subjective or objective [170]. In subjective evaluation, a human is involved in measuring the visual quality (i.e., blurriness assessment), maintaining the subjective evaluation standards such as ITU-R BT.500. On the contrary, objective assessment methods are generally based on the transformed (e.g., wavelet domain) and spatial(e.g., structural similarity) domains. Many of the aim estimates are mainly full reference(FR), reduced-reference (RR), and no-reference (NR) methods. Different features are established for the quality assessment of generic and unique images, such as biometric samples.

6.2.1.1 Image Quality Metric assessed

The Quality of input images is measured according to the method described in [136]. Also, five IQMs (indexed with (a)-(e)) are discussed next, and those metrics are also assessed in their work.

Firstly, an input color image is converted into a grayscale image I with N rows and M pixels columns, denoted as $I_{N \times M}$. A pixel is indexed by $(i, j) \in$ R2 with the row index $i = 1, 2, ..., N$ and column-index $j = 1, 2, ..., M$. Here, $N \times M$ is smoothed with a 3×3 low-pass Gaussian filter with the correlation kernel (standard deviation $\sigma = 0.5$), which produces a distorted image $\underline{I}_{N \times M}$. The pixels which lie outside the bound are replicated with the values of boundary pixels. The quality difference (Δ) between the original and distorted images is measured with respect to a metric one at a time, defined below:

$$\Delta(I, \underline{I}) = IQM(I) - IQM(\underline{I})$$

(6.1)

Mean Square Error (MSE): MSE is a simple and mostly used metric for IQA due to its mathematical simplicity. It is based on pixel-wise intensity differences:

$$MSE(I, \underline{I}) = \sum_{i=1}^{N} \sum_{j=1}^{M} (I_{i,j} - \underline{I}_{i,j})^2$$

(6.2)

Peak Signal to Noise Ratio (PSNR): PSNR is derived from the squared differences between pixel-wise intensities, and its underlying basis is MSE. It is attractive due to its simple mathematical easiness, like MSE. Both the MSE and PSNR do not consider the quality of visual information:

$$PSNR(I, \underline{I}) = 10\left(max(I^2)/MSE(I, \underline{I})\right)$$

(6.3)

Structural Content (SC): SC is another pixel-based metric that computes the ratio between the original and distorted images:

$$SC(I, \underline{I}) = \sum_{i=1}^{N} \sum_{j=1}^{M} (I_{i,j})^2 / \sum_{i=1}^{N} \sum_{j=1}^{M} (\underline{I}_{i,j})^2 \qquad (6.4)$$

Edges and corners represent significant information about an image. These are apposite for quality assessment too. The edge and corner differences are computed between I and $\underline{I}$.

Edge Difference (ED): The Sobel operator is applied to determine binary edge maps from which a total edge difference (absolute value) is calculated. Here, the edges are denoted with E and $\underline{E}$ which are computed from I and $\underline{I}$, respectively:

$$ED(I, \underline{I}) = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \left| E_{i,j} - \underline{E}_{i,j} \right| \qquad (6.5)$$

Corner Difference (CD): The Harris corner detector is followed to compute the number of corners NCR and $\underline{NCR}$ from binary images of natural and distorted images, respectively. It is delineated as a ratio related to the number of corners detected in I and $\underline{I}$:

$$CD(I, \underline{I}) = \frac{\left| N_{CR} - \underline{N}_{CR} \right|}{max(N_{CR}, \underline{N}_{CR})} \qquad (6.6)$$

Entropy Difference (EyD): Entropy, H, is a statistical measure to define a grayscale image's randomness. It is defined as $H = -\sum(p)$, where p represents the histogram counts regarding the image's grayscale intensities. This metric computes the entropy difference between I and $\underline{I}$:

$$EyD(I, \underline{I}) = \frac{|H(I) - H(\underline{I})|}{max(H(I), H(\underline{I}))} \qquad (6.7)$$

Structural Similarity Index Measure (SSIM): The similarity index measure relies on the luminance (*l*), contrast (*c*), and structural (*s*) perceptual qualities of the human visual system[25]. The mean SSIM is calculated to estimate the overall perceived image quality. The most straightforward formulation is given as:

$$SSIM(I, \underline{I}) = \left(l(I, \underline{I})^{\alpha}\right) \cdot \left(c(I, \underline{I})^{\beta}\right) \cdot \left(s(I, \underline{I})^{\gamma}\right) \qquad (6.8)$$

where, $l(I, \underline{I}) = \dfrac{2\mu_I \mu_{\underline{I}} + \varepsilon}{\mu_I^2 + \mu_{\underline{I}}^2 + \varepsilon}$, $c(I, \underline{I}) = \dfrac{2\sigma_I \sigma_{\underline{I}} + \varepsilon}{\sigma_I^2 + \sigma_{\underline{I}}^2 + \varepsilon}$, $s(I, \underline{I}) = \dfrac{\sigma_{II} + \varepsilon}{\sigma_I \sigma_{\underline{I}} + \varepsilon}$ and α, β, and γ are positive

constants for emphasizing each component relatively. The mean of I (μ_I), the mean of $\underline{I} (\mu_{\underline{I}})$, the variance of I (σI), the variance of $\underline{I}(\sigma_{\underline{I}})$,, and the covariance of I and $\underline{I}(\sigma_{I,\underline{I}})$, are vital parameters to determine the quality of real (I) and distorted ($\underline{I}$) images, which are considered as a set of pixel blocks. A small positive constant ε is considered to make the denominator non-zero. The SSIM estimates the quality of each pixel with reasonable accuracy. Its mathematical foundation is computationally efficient and appealing for quality analysis of generic images. It is a popular metric, and several variations have been developed for IQA, such as super-pixel structural similarity (SPSIM) [24,31].

Edge-Strength Similarity-Based Image Quality Metric (ESSIM): ESSIM represents the semantic information as the edge strength of an object regarding each pixel [171]. It computes whether a pixel belongs to the edge of a semantic object. The ESSIM is measured in horizontal, vertical, and diagonal directions. The directional derivatives determine the directional edge strength using a suitable Scharr kernel. The ESSIM estimates the visual fidelity between a reference and distorted image regarding the edge-strength map:

$$\mathrm{E}SSIM(I, \underline{I}) = \frac{1}{MN} \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{2e\left(I_{i,j}\right)e\left(I_{\underline{i,j}}\right) + \varepsilon}{e\left(I_{i,j}\right)^2 e\left(I_{\underline{i,j}}\right)^2 + \varepsilon} \qquad (6.9)$$

where e() computes the edge strength of corresponding pixels indexed with (i, j), and ε is a small positive constant, as defined earlier.

WAvelet-based SHarp features (WASH): The perceived quality is based on sharpness and zero-crossings in the wavelet domain [155]. The sharpness of an image is estimated from the energy in the wavelet sub-bands. The wavelet-based zero-crossings follow.

The Laplacian method for edge detection uses the second-order derivative of an image. Note that successive operations are followed to compute WASH (W). The similarity (v) metric is derived utilizing the sharpness (λ) of a reference, and the distorted image is defined as:

$$v(I, \underline{I}) = \frac{2I_{\lambda}I_{-\lambda} + \varepsilon}{I_{\lambda}^2 + I_{-\lambda}^2 + \varepsilon} \qquad (6.10)$$

where ε is defined above, the zero-crossing is defined using the edge structural similarity(ES) of the reference and distorted images in three wavelet sub-bands (namely, the W_{LH}, W_{HL}, and W_{HH}):

$$Z(I, \underline{I}) = \Pi_X E_S(I, \underline{I}) \ where \ \Pi_X E_S = \frac{\Sigma(NE_I \cap NE_{\underline{I}})}{\sqrt{\Sigma NE_I}\sqrt{\Sigma NE_{\underline{I}}}} \qquad (6.11)$$

where X denotes the sub-bands of single-level wavelet decomposition of the image; finally, the WASH is measured as $W(I, \underline{I}) = v(I, \underline{I})^{\gamma} + Z(I, \underline{I})^{1-\gamma}$ with $\gamma = 0.8$ for a higher weightage to sharpness.

6.2.1.2 Proposed gradient magnitude similarity (GMS)

Gradient conveys crucial visual information for IQA [162,165]. It reflects structural and contrasts differences of an image. Generally, gradient operators can account for a slight edge discontinuity due to illumination imperfection or manipulated artefact, which could not be reckoned only from a grayscale image. The general idea for computing the gradient is to apply convolution between a given image with a linear kernel operator. Commonly used filters include the Prewitt, Sobel, and Scharr filters (Figure 6.3). In [166], the Prewitt filters along the horizontal (x) and vertical (y) directions are used for gradient computation. For example, Prewitt's 3×3 operators for gradient computation are given in Figure 6.4. The convolution ($\otimes$)is applied to the input image I(i,j) with these operators:

$$Gx \ (i, j) = I \ (i, j) \otimes hx \ and \ Gy \ (i, j) = I \ (i, j) \otimes hy \qquad (6.12)$$

116
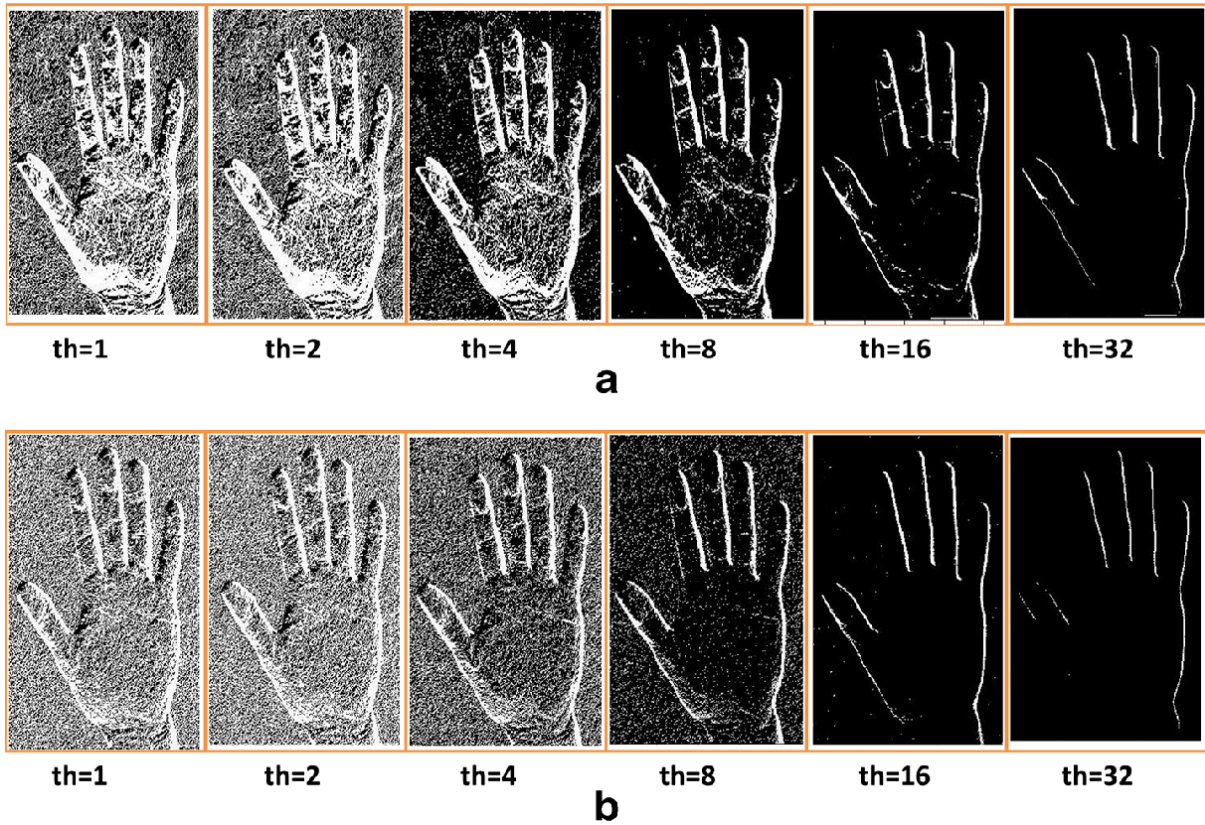
Figure 6.3: Binary images of the first images (column 1 in Fig. 6.2a and b) rendered with the variations of thresholds, for p=1,2,3,4, and 5.**a** Variations in the real hand images obtained from (6.13), (top row). **b** Deviations of the respective spoofed hand images (bottom row)

+1	0	-1
+1	0	-1
+1	0	-1

+1	+1	+1
0	0	0
-1	-1	-1

Figure 6.4: Prewitt's Operator to calculate gradient values hx(left)and hy(right)

A simple method has been proposed to compute the gradient values (Gx and Gy) by considering the local intensity difference between two adjacent pixels at the location I(i, j)and I(i, j+1) along the X-axis. The 1-D gradient operator hx= [1, 0,−1] is simplified as hx = [1,−1]. A control parameter *is* introduced to eliminate trivial intensity profile variations between consecutive pixels in the gradient computation. The modified 1-D masks are given as Gx= 1th[1,−1], and Gy= 1th[1,−1]T . Gx and Gy are akin to traditional gradient

117

magnitude *(GM ≈ Gx+ Gy)*when *th*=1. In our proposed method, the gradient maps are computed in a pixel-wise manner, defined as follows:

$$Gx\ (i,j) = \tfrac{1}{th}|I(i,j) - I(i,j + 1)|\ and\ Gy\ (i,j) = \tfrac{1}{th}|I(i,j) - I(i + 1,j)| \tag{6.13}$$

$$\Delta G_x = \sum_{i=1}^{N}\sum_{j=1}^{M} (G_x(i,j) - \underline{G}_x(i,j))^2, \Delta G_y = \sum_{i=1}^{N}\sum_{j=1}^{M} (G_y(i,j) - \underline{G}_y(i,j))^2 \tag{6.14}$$

Similarly, $\underline{G}_x$ and $\underline{G}_y$ are defined for $\underline{I}$ in the same manner. A higher value *is* a significant variation from high- to low-intensity between neighborhood pixels. It is defined as *th*= 2*p* where p=0, 1,...,5. The significance of changing *th* to determine the edge strength is illustrated in Fig. 3. The variations of intensities are presented in the binary images for both types of images. This method is efficient for geometric feature extraction based on contour profiles of the real hand images. For a higher *th*, the intensity profile variations in gradient magnitudes are remarkable [38,39]. This important modification in the gradient map computation facilitates the formulation of the proposed gradient similarity metric, and the influence of *th* in *GM* is described in the experiments. In addition, normalizing the overall quality with the average factor 1/*NM* can be considered another modification in the proposed mean GMS map (6.15)-(6.16). The average value considers the equal importance of each pixel in computing the image quality. Therefore, it computes the average gradient magnitude similarity value to estimate the image quality as:

$$GMS(I,\ \underline{I}) = \tfrac{1}{MN}\sum_{i=1}^{N}\sum_{j=1}^{M} \sqrt{\Delta G_x + \Delta G_y} \tag{6.15}$$

Alternatively, GMS can be defined as:

$$GMS(I,\ \underline{I}) = \frac{2\Delta G_x \Delta G_y + \varepsilon}{NM\ (\Delta G_x^2 + \Delta G_y^2 + \varepsilon)} \tag{6.16}$$

where a positive constant $\varepsilon = 10^{-5}$.

In summary, the novelty in formulating the mean GMS metric is thresholding on the gradient computation method using a control parameter *th*. All these ten IQMs are used for quality assessment between the original and fake hand images, described in the next section.

6.2.2 Experiment

The experiments are conducted to classify the authentic and spoofed hand images of 255 persons of the BU dataset. Three original left-hand images per person (i.e., 765 images) are considered for fake dataset creation. The fake hand datasets are created with two variations: (i) Fake hand dataset with natural degradation and (ii) Fake hand dataset with artificial degradation. A comprehensive description of various experiments with these fake image sets is summarized.

6.2.2.1 Fake Hand Dataset Creation

The original left-hand images (45 dpi) of 255 subjects of the BU database are chosen as the natural hand from which the fake samples are captured. The characteristics of the BU database are described in [34,35].

6.2.2.1.1 Fake hand dataset with natural degradation

The original dataset contains three color hand images per person, displayed on a laptop one at a time. An image is then acquired as the fake sample (72 dpi) for each real hand using a Canon EOS 700D camera. Figure 6.2a shows three real hands of a subject, and its corresponding fake samples are shown in Fig. 6.2b. The idea of creating such a fake sample is that the original image of the BU dataset is displayed on a laptop screen, and then an image is captured using the camera from the laptop screen. There is a fixed distance between the camera lens and the laptop screen. We have maintained the same distance for each capture. When creating a sample fake dataset, the lighting conditions, camera angle, and other imaging factors remain consistent. These fake images are distorted naturally. Due to surrounding environmental conditions during image acquisition, natural artefacts are introduced in the counterfeit samples. Therefore, this spoofing image dataset collection is denoted as a fake hand dataset with natural degradation, consisting of 765 fake samples.

6.2.2.1.2 Fake hand dataset with artificial degradation

Another fake dataset is created by introducing various artificial noises to the same original hand images, denoted as a fake hand dataset with artificial degradation. Three common noises, namely, Gaussian blur, salt and pepper noise, and speckle noise, were added to the

real images added with the input image. The samples are shown in Fig. 6.2c. One fake sample has been created from each actual sample for each category of artificial noise. Thus, this fake hand dataset comprises *the number of original images per hand × total number of subjects × the number of artificial noises* = 3 × 255 × 3 = 2295 spoofed images altogether.

We experimented on 765 original images and 765+2295=3060 fake hand images from both datasets. All the images are converted into grayscale and resized to 400 × 300 pixels before computing the defined IQMs.

6.2.2.2 Experimental analysis

The work is experimented with hand-crafted features using conventional machine learning techniques and deep features using CNN. MobileNetV2 [169] is used as a base network for the latter experiments. Thus, our experimental setup is two-fold.

6.2.2.2.1 Experimental results

The first experiment was conducted with naturally degraded fake hand images. Next, experiments with the artificially degraded images are conducted. Again, in the context of dataset size, this experimental setup is two-fold. In the first set of experiments, we considered 150 subjects with the original and corresponding spoofed samples as a random sub-population from 255 subjects. Next, the other set of experiments considers all the 255 subjects and related fake samples. Unless explicitly mentioned, the experimental description is based on the original and fake sub-population samples consisting of 150 subjects. The results of various experiments with this sub-population are presented in Tables 6.2, 6.3, 6.4 and 6.5 and pictorially illustrated in Figures 6.4, 6.5 and 6.6. The experimental outcomes with all 255 subjects are shown in Tables 6.6 and 6.7 and Figures 6.8 and 6.9.

The objective of the experimental arrangement is delineated as a binary classification to discriminate between the genuine hand images from fake samples using the k-nearest neighbour (k-NN), random forest (RF), and support vector machine (SVM) classifiers. Three (k=3) neighbours are considered using the Euclidean distance for the k-NN classifier. The RF classifier consists of an ensemble of decision trees [172]. The tree bagging method is used to classify the unknown test samples, and fifty bagged decision trees (RF=1 to 50) have been considered for each experiment. Linear and radial basis function (RBF) kernels are used in SVM classifiers during testing.

As only three samples per category are available, two real/original hand images and the corresponding two fake samples per subject are trained, with the remaining one of each category used for testing. In this fashion of choosing the training and testing samples, three different test cases are experimented with one at a time. Finally, the average of three classification results are considered and reported here. The classification errors are measured regarding the genuine false rate (FGR) and false fake rate (FFR).

– The FGR represents the percentage of spoofed samples classified as genuine.

– The FFR denotes the percentage of real images that are classified as fake.

The average of these two error rates is half the total error rate, defined as HTER=(FGR+FFR)/2. Before the experiment, IQMs were normalized on a [0, 1] scale using the min-max rule. Firstly, the quality metrics are evaluated independently to estimate the HTER by the classifiers; as mentioned earlier, the errors (%) are provided in Table 6.3. The results imply that SSIM, ESSIM, and GMS (*th*=8) perform satisfactorily. The performance of our proposed GMS is remarkable compared to other existing metrics. In particular, GMS offers errors comparable to state-of-the-art SSIM and ESSIM metrics. The accuracy of GMS depends on threshold (*th*) values to determine the gradient images. Thus, a good *th* is determined to minimize HTER, and the errors at various thresholds are given in Table 6.3.

In this regard, the errors of GMS are determined by two definitions in (6.15) and (6.16). GMS in c offers better results than (6.16). Thus, the following experiments are carried out with GMS at *th*=8, according to (6.15). GMS standalone cannot render zero error. Thus, significant metrics are collectively considered with the GMS one at a time, and improvements in results are observed. The HTERs of this test are given in Table 6.4. Mainly, a few random subsets containing

two, three, and four metrics are tested. It implies that instead of considering one metric for assessment, we consider two/three metrics as our feature matrix for error estimation. As mentioned, the metrics are normalized to the [0, 1] scale before the experiment. Therefore, the other quality features are concatenated in the feature matrix along the GMS in a simple incremental manner to minimize the error. In this way, we have concatenated different metrics

to reduce the errors. Also, all the metrics are evaluated together (Table 6.4) to observe any significant improvement in the results. In this test with 150 subjects, the SVM classifier with both kernels produces comparable errors. Combining all the metrics may not offer an optimal solution. Only three metrics can be effective for PAD in this testing context. With GMS, these three relevant metrics, SC and WASH, result in 0% HTERs except for the RF classifier. An alternative to WASH, the ESSIM, can be used as it can offer similar HTER. Thus, it is inferred that only three IQMs are sufficient for anti-spoofing regarding this experimental scenario's hand images of 150 persons. Notably, all the metrics are all together and can not produce the best results. Therefore, for the following experiments with 255 subjects, we have chosen the combination of WASH and GMS metrics to reduce the classification errors because the two metrics produce the minimum HTER for 150 subjects. The HTERs lie within 0.5% - 1.25% for different classifiers with all the ten metrics for 255 subjects (Table 6.4). In this regard, the proposed GMS produces 1.50% HTER using the k-NN and RF classifiers. The HTER difference between our proposed metric and all the metrics is at most 1% using the SVM classifier with RBF kernels. Thus, the performance of GMS (*th*=8) over other metrics is competitive. These results follow a similar trend as observed with 150 subjects. The HTER of the proposed GMS at *th*=8 is increased to 0.17% for 255 subjects (Table 6.4) compared with 150 subjects (Table 6.2) using the k-NN and RF classifiers.

Table 6.2: HTERs (%) estimation using the defined IQMs

IQM	k-NN			RF			SVM(linear)			SVM (RBF)		
	FFR	FGR	HTER	FFR	FGR	HTER	FFR	FGR	HTER	FFR	FGR	HTER
a) MSE	35.33	35.33	35.33	20	44	32	23.33	33.33	28.33	23.33	33.33	28.33
b) PSNR	35.33	35.33	35.33	30	34.67	32.33	24.67	30	27.33	24	31.33	27.67
c) SC	38	22	30	32	24	28	23.33	42.67	33	44.67	10.67	27.67
d) ED	37.33	38	37.67	34.67	33.33	34	56.67	20	38.33	27.33	34	30.67
e) CD	46	58.67	52.33	61.33	32.67	47	68.67	21.33	45	63.33	24	43.67
f) EyD	40	41.33	40.67	38	38	38	10	71.33	40.67	24.67	46	35.33
g) SSIM	**0**	**2**	**1**	**0**	**1.33**	**0.67**	**1.33**	**0.67**	**1**	**0.67**	**0.67**	**0.67**
h) ESSIM	2.67	2	2.33	0.67	1.33	1	1.33	0.67	1	1.33	0.67	1
i)WASH	6	1.33	3.67	5.33	1.33	3.33	3.33	10	6.67	6	2.67	4.33
j) Proposed GMS(the = 8)	**0**	**2.67**	**1.33**	**0**	**2.67**	**1.33**	1.33	2	1.67	1.33	2	1.67

The SSIM performs the best results. The ESSIM and proposed GMS have achieved results comparable to those of SSIM. The best results among individual IQM and the proposed GMS using different classifiers are marked in **bold** font.

Table 6.3 Experiments to determine threshold using two definitions of Gradient Magnitude Similarity (GMS)

th	k-NN			RF			SVM(linear)			SVM (RBF)		
	FFR	FGR	HTER	FFR	FGR	HTER	FFR	FGR	HTER	FFR	FGR	HTER
1	18	20.67	19.33	16.67	18	17.33	4.67	28	16.33	14.67	12.67	13.67
2	10	15.33	12.67	12	10	11	0.67	28	14.33	10.67	8.67	9.67
4	1.33	4.67	3	1.33	2.67	2	0.67	12.67	6.67	4	2.67	3.33
8	**0**	**2.67**	**1.33**	**0.67**	**2**	**1.33**	**1.33**	**2**	**1.67**	**1.33**	**2**	**1.67**
16	23.33	26.67	25	22	20.66	21.33	12	26	19	8	27.33	17.67
32	39.33	48	43.67	36	47.33	41.67	38.67	58.67	48.67	32	40.67	36.33
1	39.33	43.3	41.33	35.33	40	37.67	27.33	32.67	30	36	21.33	28.67
2	37.33	36.67	37	34.67	33.33	34	35.33	24.67	30	34	27.33	30.67
4	31.33	44.67	38	26.67	43.33	35	32	23.33	27.67	18.67	39.33	29
8	33.33	45.33	39.33	22.67	46	34.33	33.33	26	29.67	35.33	24.67	30
16	42.67	46	44.33	37.33	45.33	41.33	32	54.67	43.33	22	52.67	37.33
32	47.33	51.33	49.33	36.67	54	45.33	42	48	45	46.67	42.67	44.67

Top row-set: GMS (15), bottom row-set: GMS (16). The best results achieved using threshold variation (th=8) in both equations are marked with bold font

In another test with the support vectors, only two metrics are tested together, and particularly,

GMS is paired with a) PSNR and b) SC. These two cases are tested using the mentioned kernels and pictorial examples of the support vectors plotted in Figure 6.5. The GMS and PSNR are tested with a linear kernel SVM (Figure 6.5a), and GMS with SC is used for classification with the RBF kernel SVM (Figure 6.5b). Likewise, a combination of four metrics and all ten metrics are also evaluated. However, no improvement in accuracy has been observed in this experiment. Therefore, only three significant metrics (as specified above) are chosen for our experiments.

During this experiment, the RF classifier estimates the out-of-bag (OOB) errors with a subset of three good features. It computes the cumulative wrong classification probability average for OOB observations in the bootstrap dataset. The minimum OOB errors are a) SC, SSIM, and GMS: 6.41×10^{-4}; b) SC, ESSIM, and GMS: 1.4×10^{-3}; and c) SC, WASH, and GMS: 1.7×10^{-3}. These OOB errors are estimated to be within 1 to 50 decision trees. A graphical comparison of OOB errors with these combinations of metrics is shown in Figure 6.6a. It is clear that SC, SSIM, and GMS collectively perform the best for correct classification.
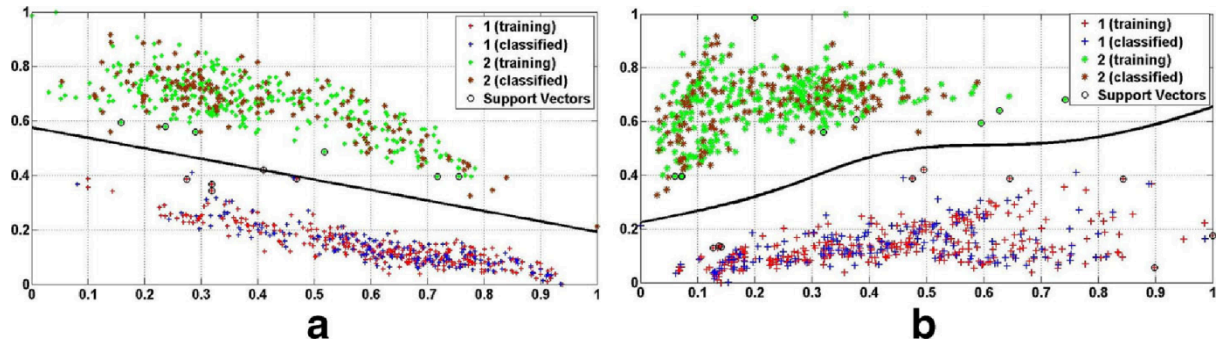
Figure 6.5: Classification with two IQMs: **a** PSNR and GMS using SVM with a linear kernel. **b** SC and GMS using SVM with an RBF (σ =1) kernel

Table 6.4: Different combinations of significant IQMs to minimize error (%)

Population	IQMs	kNN			RF			SVM (Linear)			SVM (RBF)		
		FFR	FGR	HTER	FFR	FGR	HTER	FFR	FGR	HTER	FFR	FGR	HTER
150 Subjects	b) PSNR j) GMS	0	0.67	0.33	0	2	1	0	0	0	0	0.67	0.33
	c) SC j) GMS	0	0.67	0.33	0	2	1	0	0.67	0.33	0	0.67	0.33
	g) SSIM j) GMS	0	2	1	0	1.33	0.67	1.33	1.33	1.33	1.33	1.33	1.33
	h) ESSIM j) GMS	0	2.67	1.33	0.67	2	1.33	0.67	2	1.33	0.67	2	1.33
	i) WASH j) GMS	0	2	1	0	0.67	0.33	0.67	1.33	1	0.67	1.33	1
	c) SC g) SSIM j) GMS	0	0.67	0.33	0	2	1	0	0.67	0.33	0	0.67	0.33
	c) SC h) ESSIM j) GMS	0	0	0	0.67	0.67	0.67	0	0	0	0	0	0
	c) SC i) WASH j) GMS	0	0	0	0.67	0	0.33	0	0	0	0	0	0
	c) SC h) ESSIM i) WASH j) GMS	0	0	0	2	1	0	0	0	0	0	0	0
	All	0	0	0	1.33	0.67	0	0	0	0.67	0	0.33	
225 Subjects	j) Proposed GMS th = 8	1	2	1.5	2	1	1.5	1.8	2	109	1.5	3	2.25
	i) WASH j) GMS	1	3	2	2	2	2	2	4	3	3	2	2.5
	All	0	1	0.5	0	1	0.5	0.75	1.7	1.23	1.2	1.3	1.25
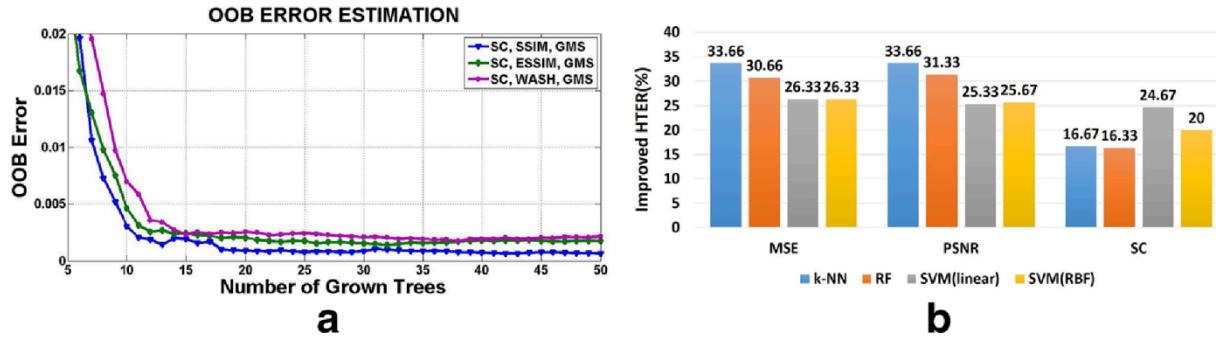
Figure 6.6: a OOB-error estimation using RF with the SSIM, ESSIM, and WASH chosen one at a time along with the SC and GMS such that only three IQMs are tested in each case. **b** The improvements in HTERs due to modification at the preprocessing stage before estimating the MSE, PSNR, and SC quality

In another experiment, we introduced a slight modification during the preprocessing stage before computing the defined metrics. In this test, six IQMs, particularly the MSE, PSNR, SC, ED, CD, and EyD, are calculated with this new preprocessing step. Firstly, the original hand image is considered, and its gradient image is computed at th=8 using (6.13). Next, the specified IQMs are extracted. Similarly, the fake images are smoothed with the same Gaussian filter, and the gradient image is calculated similarly to the original hands. The results are given in Table 6.5. This modification in the preprocessing of images for these metrics also produces more remarkable results than the errors reported in Table 6.2. In this test, reduced errors(%) are attained (Table 6.6), and significant gains in HTERs are shown in Figure 6.6b. However, minimal gains in HTERs are obtained using RF for ED (1%) and CD (2%), while no improvement has been observed for EyD by any classifier. Hence, the present preprocessing also effectively computes HTERs for existing IQMs. It justifies that our thresholded gradient computation can also be helpful at the preprocessing stage when computing other metrics.

The following experiment considered the original and fake hand samples of all 255 subjects. Similar to the aforementioned experimental strategy, three classification tests are conducted. Specifically, these tests discriminate between the authentic hand images and all three types of fake hand images, artificially degraded with the Gaussian blur, salt and pepper noise, and speckle noise. We used 765 original (real) and 765 fake (1530) images for each classification task. Here, we consider our proposed GMS metric at th=8 and all the metrics together for classification. The results are given in Table 6.6.

Table 6.5: Improvements of proposed GMS over six referred metrics: errors (%)

IQM	k-NN			RF			SVM (linear)			SVM (RBF)		
	FFR	FGR	HTER	FFR	FGR	HTER	FFR	FGR	HTER	FFR	FGR	HTER
a) MSE	0.67	2.67	1.67	0	3.33	1.67	2	2	2	2	2	2
b) PSNR	0.67	2.67	1.67	0	3.33	1.67	2	2	2	2	2	2
c) SC	12.67	14	13.33	11.33	12	11.67	9.33	7.33	8.33	10.67	4.67	7.67
d) ED	37.33	38	37.67	33.33	34	33.67	56.67	20	38.33	27.33	34	30.67
e) CD	46	58.67	52.33	59.33	34	46.67	68.67	21.33	45	63.33	24	43.67
f) EyD	40	41.33	40.67	36	40	38	10	71.33	40.67	24	67.46	35.33

The proposed GMS(th=8) remains fixed in each combination, and other metrics are concatenated with it one at a time. The HTERs lie within 0-1.25% considering the following conjugate metrics. Significant results for different combinations and classifiers are marked in **bold** font

Table 6.6: Errors (%) for artificially created fake images from 255 subjects with noise degradation

Noise	IQM	k-NN			RF			SVM (linear)			SVM (RBF)		
		FFR	FGR	HTER	FFR	FGR	HTER	FFR	FGR	HTER	FFR	FGR	HTER
Gaussian Blur	GMS(th=8)	0	0	0	10	10	10	0	0	0	0	0	0
	All	0	0	0	0	0	0	0	0	0	0	0	0
Salt and Pepper	GMS(th=8)	0	0	0	0	20	10	0	0	0	0	0	0
	All	0	0	0	0	10	5	0	0	0	0	0	0
Speckle	GMS(th=8)	0	0	0	10	20	15	0	0	0	0	0	0
	All	0	0	0	0	0	0	0	0	0	0	0	0

The performance (HTER) of the proposed GMS metric is excellent in discriminating artificially degraded fake and original samples using the SVM classifier with both types of kernel. Therefore, the proposed metric performs well for spoofing detection using fake hand samples with noise degradation. Next, the verification of actual and spoofed templates is experimented with regarding a decision threshold t_v, defined as:

$$t_v = \sqrt{\sum_{f=1}^{W} \frac{(U_f - V_f)^2}{\sigma^2}} \qquad (6.17)$$

where an IQM is denoted with f, trained samples of a subject are denoted by U, an unknown IQM of a claimer is represented with V, and f is the standard deviation of quality determinative feature f. The verification performance is calculated regarding the following:

– False Accept Rate (FAR): a spoofed matching score is lower than the threshold t_v.

– False Reject Rate (FRR): a legitimate matching score is more than the threshold t_v.

– Genuine Accept Rate (GAR) = 1 – FAR

The GAR and FAR depend on the *tv* and are plotted in a Receiver Operating Characteristic (ROC) curve [41] in Figure 6.7a. Alternatively, the GAR and FAR of genuine and spoofed samples are considered separately through the histograms with 100 bins to represent the quality template matching distances. These two histogram distributions are plotted in Figure 6.7b and c.

6.2.2.2.2 Deep learning experiment

In recent years, deep learning (DL) models have been widely used for solving image recognition problems. We have experimented with our spoofing detection approach on our fake datasets using the DL method. As conventional heavyweight CNNs (e.g., VGG-19, ResNet-50) follow deeper and more complex network architecture, stacking multiple convolution layers is time-consuming. To cope with the time constraint, we have used the MobileNetV2 model [36]. As stated above, it is a lightweight model compared to other deep models. It can also be used in Android to classify tasks like ours. For our experiment, a binary classification model is created leveraging the high-level feature maps of the MobileNetV2 base network using the Tensorflow and Keras frameworks. Like the aforementioned experimental strategy, four classification tests are conducted with all subjects. In these test cases, the dataset is split with an 8:1:1 ratio, where 80% of total images are used for training, 10% of images are for validation, and the remaining 10% for testing purposes. We have applied standard data augmentation techniques (i.e., horizontal flip, zoom (20%), shearing (20%), and rotation 90 degrees) with random variations. The original color hand image with 382 × 585 pixels is resized to 224 × 224 pixels.

We have used a simple transfer learning approach. We have adapted pre-trained Imagenet weights in base CNN for faster convergence. The Stochastic Gradient Descent (SGD) optimizer with a momentum of 0.90 is applied to reduce the categorical cross-entropy loss function, and the learning rate is 0.0001. The model is trained for 100 epochs with a mini-batch size of 8 in two steps. First, our model is trained for 50 epochs, resulting in an

uneven training and validation performance. Next, the model is trained further using the saved weights of step 1 for 50 more epochs with the same batch size.

Figure 6.8 clearly shows the training accuracy vs. validation accuracy for each type of fake sample for step 2. It is evident that the training behaviors of these tests are similar, and a slight variation is observed in the validation accuracy due to different fake samples with noise degradation. The trained model is tested with the remaining unseen 10% images of each experiment's total dataset. The classification accuracy(%) is given in Table 8. The confusion matrix of each experiment is shown in Figure 6.9. The deep network effectively distinguishes the real and fake hand images. From all the experiments, it is clear that naturally degraded images are more challenging to detect than artificial degradation with salt and pepper noise and speckle noise. However, Gaussian noise makes it more challenging to detect correct spoofing.
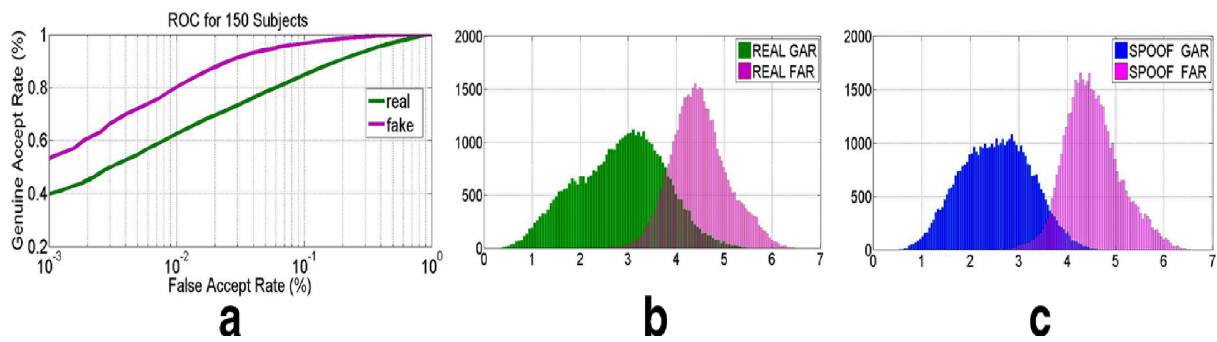
Figure 6.7: a Receiver Operating Characteristic (ROC) curve illustrates the errors during verification of the original and fake samples. Histogram distribution of the genuine accept rate (GAR), and the false accept rate (FAR) of (**b**) real hands and (**c**) spoofed hands
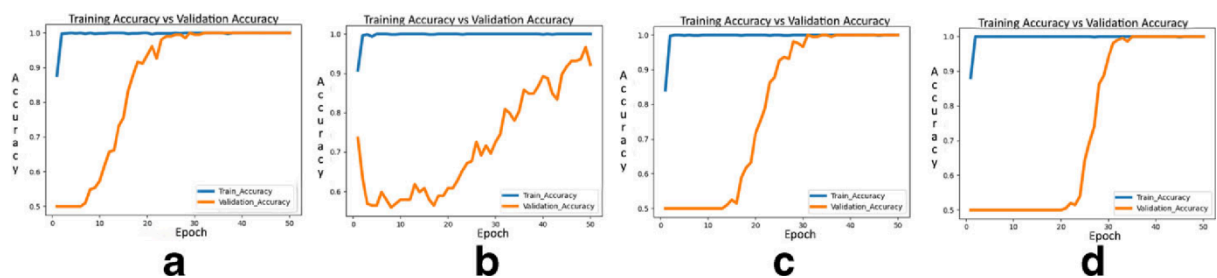
Figure 6.8: Classification results. (**a**) accuracy of original and naturally degraded fake images, (**b**) accuracy of original and degraded fake images using Gaussian blur, (c) accuracy of original and degraded fake images using salt and pepper, and (d) accuracy of original and degraded fake images using speckle noise

In [173], the thermal features of the hand are computed using a CNN architecture. It applies a score fusion of the RGB and thermal image matching modules to reach a final decision. According to our study, no comparable work for anti-spoofing on hand images using a software-based method such as the IQMs explored in this proposed work is available. However, there exist a few hardware-based liveness detection methods using hand images. Therefore, there is no scope for performance assessment compared to similar works on the same hand image dataset. There is further scope to experiment with more challenging fake samples for PAD. Spoofing detection using real-time hand images will be more demanding in a realistic situation. Our method is evaluated on a small dataset with only 255 subjects. Therefore, the fake hand dataset will be increased to a larger population and should be made publicly available for further exploration. Also, different sensors can be used for fake image acquisition and tested with our proposed method for better generalization. It will be interesting to explore further real-time fake samples in a mobile/android-based environment, which is currently a limitation of our approach. In the future, anti-spoofing detection on hand biometrics can be investigated with more challenging datasets using deep learning methods. Lastly, new quality metrics can be devised to improve hand biometric spoofing detection further.

Table 6.7: Error (%) estimation using MobileNetV2 with 255 subjects

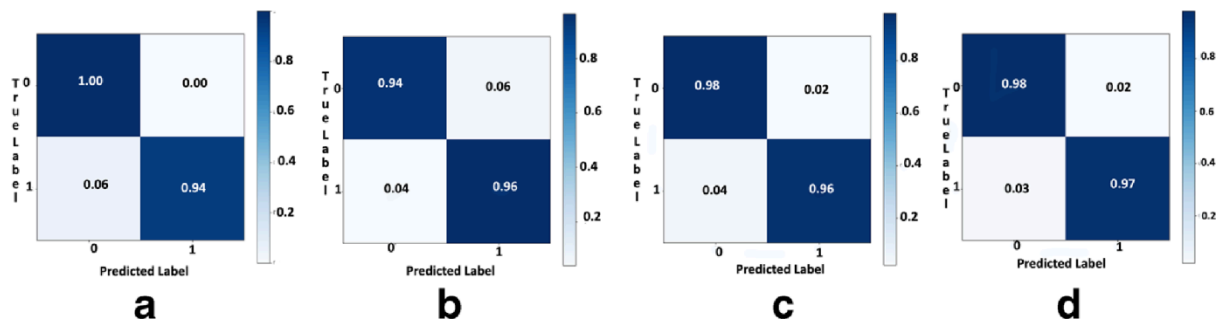| Noise	FFR	FGR	HTER
Natural	6	0	3
Gaussian Blur	6	4	5
Salt & Pepper	3	2	2.5
Speckle	2	3	2.5

Figure 6.9: Confusion matrix. **a** original and naturally degraded fake images, (**b**) original and degraded fake images using Gaussian blur, (**c**) original and degraded fake images using salt and pepper, and (**d**) original and degraded fake images using speckle noise

6.2.3 Discussion

This work, mainly a method for presentation attack detection (PAD) using the visual quality assessment, presents to defend illegitimate attempts for a hand biometric system. The defined quality metrics are very significant for anti-spoofing on real-hand images. A new gradient magnitude-based FR-IQM is described that renders satisfactory classification results. The proposed GMS defends electronic display-based spoofing attacks (i.e., natural degradation) on hand images with at most 3% HTER using traditional hand-crafted and deep-learning experimental settings. The fake samples are discriminated from the original samples using deep networks and conventional classification methods. This method can be helpful for software-based fake sample detection in a cost-efficient manner at an earlier stage of a deployed hand biometric system. However, no experiment for hand biometric authentication has been reported here, which is further scope to enhance this work such that the zero-effort imposters can also be detected successfully. In the future, necessary counter-measurements to deter spoofing attacks on hand biometrics can be included in this work.

6.3 – Application of Image Quality Assessment in Healthcare

Along with spoofing detection, Image Quality Assessment has another critical area of application coined as healthcare. A computer-aided Diagnosis system (CAD) has been proposed that is designed to detect malignant tumors in breast regions using Ultrasound images. Ultrasound imaging is a modality that has a low cost and lower health risk. Therefore, this imaging modality has been chosen, and breast ultrasound images have been classified between benign and malignant tumor classes. The image quality assessment approach has been used for this task. No reference image quality metrics have been used as features for the classification task. A public database of Ultrasound images of breast tumors has been used that contains 780 images. The classification of breast ultrasound images using image quality assessment is a novel approach, producing significant results.

Breast cancer occurs due to the unwanted growth of tissues in the breast. Mainly, the cells responsible for producing milk start growing abnormally very often, and this type of cancer is known as invasive ductal carcinoma. The other cells of the breast, called lobules, can also be affected. This type of breast cancer is known as invasive lobular carcinoma. As to the causes

of breast cancer, four reasons can be raised. The four probable reasons are – genetic, hormonal, environmental, and lifestyle. However, there are some prominent exceptions.

The unwanted growths in the breast tissues form tumors of different shapes and sizes. The tumors, in general, can be classified into two classes – benign and malignant. The benign tumors are not cancerous. The benign tumors do not have or have tiny live cells to divide and grow. Therefore, the growth of benign tumors is prolonged. The cells of benign tumors cannot invade, and this type of tumor can sufficiently respond to treatment.

On the other hand, malignant tumors are very fast-growing, destroying healthy cells. The cells present in malignant tumors invade through blood vessels. Malignant tumors contain cells of damaged DNA that can divide abnormally. Therefore, cells grow rapidly without the death of the old cell. The malignant tumor is cancerous and is a curse to human life. The identification of malignant tumors over benign tumors is an essential diagnosis step in diagnosing breast cancer. The appropriate detection of malignant breast tumors can save the lives of human beings.

Three general image-based breast cancer screening tests have been used [174]. They are – mammography, ultrasound imaging, and MRI. The MRI technique is not very appreciable due to its risk due to high radiation and colossal cost. Mammography is a more frequently used imaging technique for screening breast cancer. However, the case of dense breast mammograms does not provide great success. Ultrasound imaging can detect tumors in dense breasts, even in adolescent women. At the same time, an ultrasound imaging-based breast cancer screening system is less costly and has a shallow health risk. Therefore, ultrasound is one of the most frequently used breast cancer screening procedures nowadays.

This work is an image processing-based technique to detect malignant breast tumors over benign breast tumors using breast ultrasound imaging. We have used the image quality assessment technique for the classification task. No-reference image quality metrics have been used as features, and the two classifiers, k-NN and SVM, have been used for classification. A publicly available dataset containing 780 breast ultrasound images has been used to evaluate the system. The approach is novel because breast tumor classification using image quality assessment is not available in the literature, and the experiment has found promising results on the public dataset.

6.3.1 Methodology

The methodology of the proposed system has been discussed in this section. We have classified the benign tumor and malignant tumor from the breast ultrasound image using image quality assessment. No reference image quality metrics have been used as features for the classification task. Finally, the classification task has used two classifiers, k-NN and SVM. Figure 6.10 shows the complete workflow of the proposed methodology.
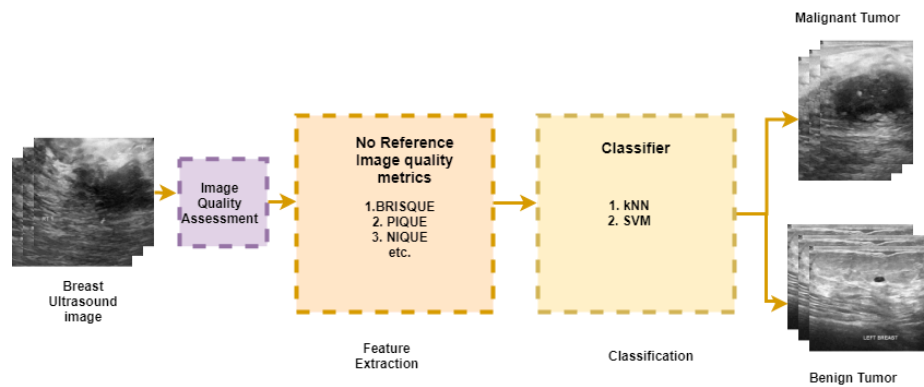
Figure 6.10: Working methodology of the proposed system

The Breast ultrasound image is the input of the system. The complete image has been processed. No segmentation of the tumor area from the input image has been done. The segmentation of the suspicious region raises the complexity of the system. Therefore, we avoid this segmentation process in a preprocessing step and extract features from the complete input image. The first step of our proposed system is image quality assessment.

6.3.1.1 Image Quality Assessment

Image quality assessment is the research field of image processing where the quality of the image has been assessed. In general, image quality can be assessed in three ways –full reference, reduced reference, and no reference. In the case of full-reference image quality assessment, the reference image was compared directly with the test image to assess the quality of the test image. Instead of the complete reference image, features of the reference image have been used for the quality assessment. The no-reference image quality assessment has no use of reference images. That means no-reference image quality is entirely independent of the reference image. A reference image for each breast ultrasound image is available in our case. Therefore, we have applied a no-reference image quality assessment for the proposed system.

6.3.1.2 Feature Extraction

We have used four no-reference image quality metrics as features. In this subsection, the metrics have been discussed briefly.

(i) Blind/Reference less Image Spatial QUality Evaluator (BRISQUE) [109]

BRISQUE is a no-reference image quality metric based on natural scene statistics. The metric has been calculated in the spatial domain. There is no need for domain transformation. In addition, it is a distortion-independent image quality metric. These characteristics of the metric are suitable for this work.

(ii) Naturalness Image Quality Evaluator (NIQE) [111]

This is another no-reference image quality metric created on the natural scene statistics criteria. This metric is a no-training-based image quality metric that was created without any supervised training process.

(iii) Perception-based Image QUality Evaluator (PIQUE) [112]

This is a no-reference image quality metric, which counts the human perception in assessing the quality of a test image. According to the methodology, the quality of local patches is calculated first. Then, the quality of the entire image is assessed.

(iv) Shannon Entropy [175]

Shannon entropy metric is the concept of information theory. This metric measures the amount of information that resides in an image. It follows a probabilistic approach for quantifying the amount of information in an image.

6.3.1.3 Classifiers

Our proposed system is classified into two classes: benign and malignant. We have used two classifiers for the classification task. We used two classifiers to compare the results. We have used two standard classifiers – k -Nearest Neighbor [114] and support vector machine [116].

6.3.2 Experimental Result

6.3.2.1 Dataset used

We have used a public dataset [79] containing 780 images, among which 487 images are benign, 210 are malignant, and only 130 are normal breast ultrasound images. The images were captured among women between the ages of 25 years to 75 years. The image resolution of the dataset is 500 × 500 pixels. We have used 170 images each from the benign class and malignant class for training, and 50 images from each class for testing — figure 6.11 displays images of benign breast tumors and malignant breast tumors.

Benign breast tumor

Malignant breast tumor

Figure 6.11: Breast ultrasound images used in the experiment
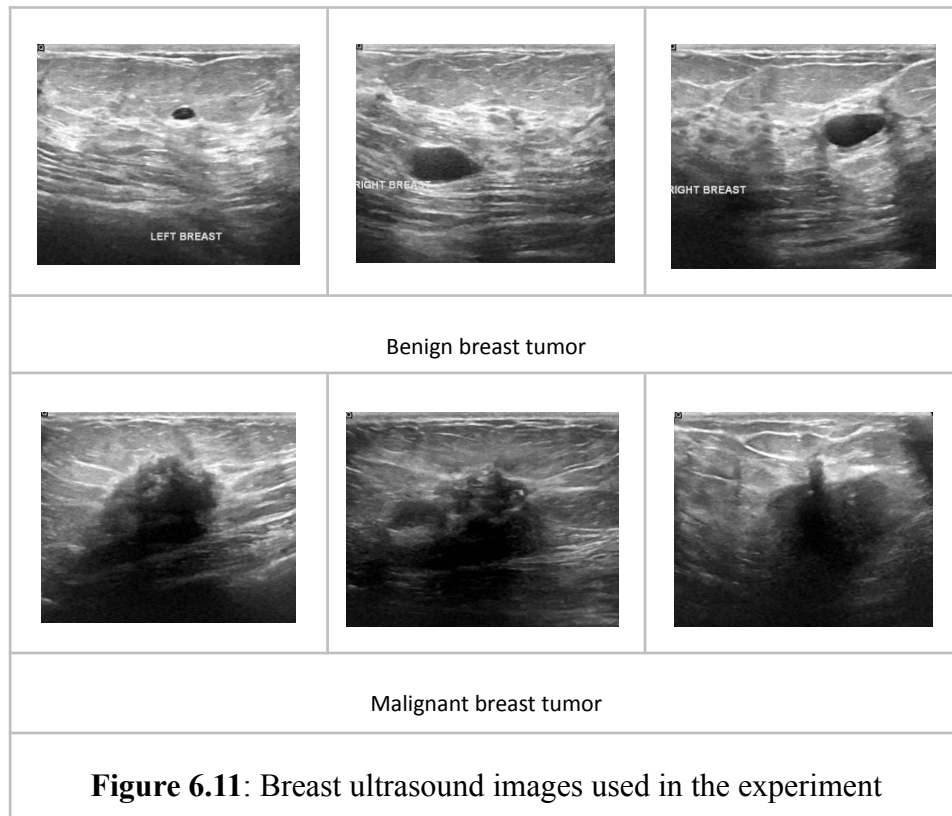
6.3.2.2 Evaluation metric used

We have used different evaluation metrics to evaluate our proposed system. The evaluation metrics are as follows.

		(6.18)
	$$Specificity = \frac{TN}{TN + FP}$$	(6.19)
	$$Sensitivity = \frac{TP}{TP + FN}$$	(6.20)
	$$F - measure = \frac{2 \times TP}{(2 \times TP) + FP + FN}$$	(6.21)
		Accuracy

In the above equations, the terminologies used are TP = True Positive, FP = False Positive, TN = True Negative, FN = False Negative,

6.3.2.3 Experimental Results

The experimental results are shown in Table 6.8. Table 6.8 consists of values obtained for the evaluation metric from our experiment. We have used two classifiers, and the results for each are shown. SVM performs better than the kNN. We have used a 0-1 scale for the evaluation metric values. We have compared our results with two state-of-the-art research works. In the paper [51], the authors used an ensemble learning process, and the LBP-based texture feature was used [52].

Table 6.8: Experimental results and comparison with the state-of-the-art methods

	Proposed Method		Ensemble Learning [175]	LBP based feature [176]
	kNN	SVM		
Accuracy	0.854	0.912	0.884	0.825
Specificity	0.885	0.924	0.903	0.81
Sensitivity	0.845	0.874	0.837	0.7
F-measure	0.778	0.811	0.805	0.8

6.3.3 Discussion

The proposed system tries to classify breast tumors using ultrasound breast images. The proposed system uses an image quality assessment technique to classify breast tumors. No-reference image quality metrics have been used as features, and two classifiers, kNN and SVM, have been used for the classification task. Utilizing the quality assessment approach is a very novel approach for breast tumor classification from ultrasound images. We have evaluated our proposed system with a public dataset containing 780 breast ultrasound images and experiment results showing that our proposed system performs significantly.

This chapter mainly concentrates on the application domains of the Image Quality Assessment technique. Although there are several applications, research contribution has been provided in two leading application domains – spoofing detection and healthcare.

In the spoofing attack, a method for presentation attack detection (PAD) using the visual quality assessment has been presented to defend against illegitimate attempts for a hand biometric system. The defined quality metrics are very significant for anti-spoofing on real-hand images. A new gradient magnitude-based FR-IQM has been described that renders satisfactory classification results. The proposed GMS defends against electronic display-based spoofing attacks (i.e., natural degradation) on hand images with at most 3% HTER using traditional hand-crafted and deep-learning experimental settings. The fake samples are discriminated from the original samples using deep networks and conventional classification methods. This method can be helpful for software-based fake sample detection in a cost-efficient manner at an earlier stage of a deployed hand biometric system. However, no experiment for hand biometric authentication has been reported here, which is further scope to enhance this work such that the zero-effort imposters can also be detected successfully. It is projected that further investigation will be conducted for countermeasures to deter spoofing attacks on hand biometrics.

In the research work on healthcare, a system has been contributed for classifying breast tumors using ultrasound breast images. The contributed system uses an image quality assessment technique to classify breast tumors. No-reference image quality metrics have been used as features, and two classifiers, kNN and SVM, have been used for the classification task. Utilizing the quality assessment approach is a very novel approach for breast tumor classification from ultrasound images. The system was evaluated using a public dataset containing 780 breast ultrasound images, and experiment results showed that the developed system performed significantly.

Chapter 7:
Conclusion

This chapter concludes the dissertation on image quality assessment of naturally degraded images. Although the main focus is weather-degraded images, related research has also been done. In this chapter, the summary of the complete dissertation will be discussed first. The current research works can then be steered in their future directions.

7.1 Summary of the Dissertation

Images are the primary resource for image processing-based research. Processing input images to obtain several analogies is the primary concern of image processing research. Therefore, before processing, the input image's quality must be assessed. Otherwise, the processing of low-quality images leads to the wrong analogy. In this dissertation, an essential and highly challenging research domain, image quality assessment, has been addressed. This dissertation broadly discusses the research domain with all of its orientations. Along with that, discussions on related research have been included in this dissertation.

The dissertation is distributed over seven chapters. The first chapter introduces the dissertation, where the image degradation model, the need for quality assessment, a brief overview of the research domain named image quality assessment, the scope of the dissertation, and the main contributions have been discussed. The second chapter discusses the research works found in the literature. The third chapter discusses an approach to image quality assessment, namely full-reference image quality assessment. In this chapter, a novel full reference image quality metric has been presented. Another approach to image quality assessment, namely no-reference image quality assessment, is discussed in chapter four. The next chapter deals with the processing of rainy images. The sixth chapter incorporates some applications of image quality assessment. The seventh chapter concludes the dissertation.

The introduction chapter of this dissertation started with the types of image degradation models. Two image degradation models have been discussed, one for technical reasons and the other for environmental stimuli. After that, the need for a quality assessment was analyzed. The research domain of image quality assessment has been instigated after the discussion on the need for quality assessment. Basic operations of image quality assessment with the three approaches, namely full-reference, reduced-reference, and no-reference of this

research method, have been discussed. The scope of the dissertation has been presented along with all its orientations. Finally, the main contributions are listed in the section.

Chapter 2, denoted as Literature Survey, as the name suggests, incorporates discussions on the research works of the recent past found in the literature. The literature survey was conducted using the three approaches of image quality assessment: full-reference, reduced-reference, and no-reference.

Research contribution has been discussed in this dissertation from Chapter 3 onwards. In Chapter 3, the research contribution in the subdomain is named full-reference image quality assessment. This chapter has contributed a feature extractor coined as an Oriented Intensity Velocity Feature Extractor (OIVFE). The superiority of the OIVFE has been proved in three application areas: weather classification, healthcare, and the steel manufacturing industry. The Oriented Intensity Velocity Full-reference image Quality (OIVFQ) metric has been implemented based on the novel feature extractor. The metric has been evaluated using publicly available image quality datasets. These appreciable contributions open a new paradigm in the full-reference image quality assessment-based research.

Chapter 4 contains the sole contribution to no-reference image quality assessment research. A novel No-reference Image Blurriness Metric (NIBEM) has been designed, and details of the implementation of this metric are the main topic of discussion in this chapter. This metric was developed based on the characteristic of blurred images represented by the Divisive Normalization Transform (DNT). A novel definition of blurry images has been incorporated in this chapter. The NIBEM has been evaluated using publicly available image quality datasets. This contribution raises the quality of the dissertation to a higher level.

Chapter 5 is entirely dedicated to the research on rainy image processing, which justifies the main focus of the dissertation. Three research areas have been addressed here. They are rainy image detection, object detection from the rain-affected video surveillance data, and de-raining of rainy images. The rainy images were identified using the no-reference image quality assessment approach, where a metric was contributed to dedicatedly assessing the quality of rainy images. Object detection from the rain-affected video surveillance data has been done using a deep learning model, namely, You Look Only Once (YOLO). The research work on deraining has been implanted using another deep learning model coined a

Generative Adversarial Network (GAN). This chapter is completely dedicated to the natural images degraded by rain.

Chapter 6 is the discussion platform where applications of image quality assessment have been demonstrated. Research contributions in hand biometry and health care have been incorporated into this chapter. One of the severe threats in hand biometry is spoofing attacks. A system to protect against spoofing attacks is highly recommended. An image quality assessment-based system has been contributed in this chapter to protect against spoofing attacks. Another crucial area that has been addressed in this chapter is breast cancer detection. A no-reference image quality assessment-based approach has been contributed here. This chapter deals with two significant application areas, the most application-oriented part of the dissertation.

Finally, Chapter 7 winds up the dissertation, where the summary of the dissertation has been stated, and the future scope of the current dissertation has been discussed. The following section directs some future orientations of the research contributions discussed throughout the dissertation.

7.2 Future Scope

Image quality assessment research has been addressed in this dissertation. Different approaches have been considered to address the research goal. This dissertation contains several contributions to the field of image processing research. However, there is scope for betterment in the future. The future scope of the dissertation is discussed below.

- As discussed in the introduction, image quality assessment research can be done using three general approaches. The three approaches are full-reference, reduced-reference, and no-reference image quality assessment. This dissertation's research contributions are limited to the two approaches – full-reference and no-reference image quality assessment. The reduced-reference approach remains untouched in this dissertation for addressing in the future.

- The current dissertation mainly focuses on the quality assessment of natural scenes degraded by different weather conditions. Different environmental stimuli, like rain, fog, mist, snow, etc., can degrade natural scenes. However, the dissertation stress on the rainy images. Separate chapters are solely dedicated to the processing of rainy

images. As part of the future scope of the current dissertation, thorough research on degraded natural scenes due to weather degradation other than rain can be conducted. Different aspects of those degraded images can be included in future research.

- Chapter 5 contains research contributions limited to detecting rainy scenes using no-reference image quality assessment. This research work can be extended to include different aspects. The intensity of rain can be detected from the rainy scenes. Similarly, detecting images in challenging situations can be an excellent future research scope.

- In Chapter 4, a no-reference image quality metric was used to assess the blurriness present in the image. However, there are different distortions possible that can degrade images. Creating no-reference image quality assessment metrics for distortions other than blurriness can be the future direction of current research work. Images may be distorted by more than one distortion type at a time, which is very common in real-life scenarios. Quality assessment of naturally distorted images, where the images have been distorted with multiple distortion types, can be a good research scope to address in the future.

- The current dissertation incorporates contributions to different application areas like weather classification, the steel manufacturing industry, healthcare, rainy image detection, and hand biometry using different image quality assessment approaches. The application-oriented research can be distributed over more application areas in the future. In Chapter 5, two application areas contributed to object detection from rainy scenes captured by the video surveillance system and the draining of rainy images. Similar application areas related to the degraded natural scenes by different weather conditions can be contributed in the future.

- Image quality assessment is one of the most important domains of computer vision-based research. Images are the main resources of this research domain, and images have been processed to extract information. As images are the main resource of this research domain, quality assessment is essential to see if the images are helpful for further processing or not. Depending on the quality assessment, it can be decided whether the input images can be processed. If the quality assessment is not done properly, then some cases may arise where the desired outcome cannot be achieved

even if the processing steps are completed properly. In today's world, many images are available online and processed for different purposes to solve real-world problems.

Machine learning-based techniques generally address research problems in the Image Quality Assessment (IQA) domain. Recently, deep learning has been applied in all research areas under computer vision, ensuring excellent results. However, deep learning has some disadvantages as well. One disadvantage of the deep learning technique is data dependency. The deep learning-based system is completely data-driven. Therefore, a massive amount of similar data is needed to train the system. However, the performance of deep learning-based systems decreases in the problem domains where a sufficient amount of training data is unavailable. Another problem is that a deep learning model designed for solving a specific problem trained with a dataset cannot provide the best result if the model has been tested with another dataset.

The No-reference Image Quality Assessment is a research domain with few experimental datasets. Therefore, deep learning-based techniques are not suitable for addressing the problem. Another issue in the NRIQA research domain is the lack of actual data. The leading IQA dataset contains synthetic data. However, the problem is that a deep learning model trained with synthetic data cannot assess the quality of the images distorted by natural stimuli. Traditional machine learning and deep learning techniques are ineffective in this situation. Meta-learning is a paradigm that can be helpful in this regard. Meta-Learning is the approach to a learning system in which a learning system can learn from previous experiences. Therefore, there is a scope for including different insights into the learning system, and the learned model can learn a system in a generalized fashion. There are situations when a learning system cannot learn from small training data. Meta-learning is a procedure where some learned system is used to train another learned model, which has been a solution to problems with a small amount of experimental data. The No Reference Image Quality Assessment is a problem, and a meta-learning-based system can be applied here. Meta-learning-based image quality assessment can be the finest direction of the current dissertation in the future.

References

[1] K. V. Beard and C. Chuang, " A New Model for the Equilibrium Shape of Raindrops", Journal of the Atmospheric Sciences, Vol. 44 No. 11, 1987, pp. 1509-1524.

[2] P. Ye and D. Doermann, "Active Sampling for Subjective Image Quality Assessment," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014, pp. 4249-4256, doi: 10.1109/CVPR.2014.541.

[3] Bo Hu, Leida Li, Jinjian Wu, Jiansheng Qian, "Subjective and objective quality assessment for image restoration: A critical survey", Signal Processing: Image Communication, Volume 85, 2020, 115839, ISSN 0923-5965, https://doi.org/10.1016/j.image.2020.115839.

[4] U. Sara, M. Akter, and M. Uddin, "Image Quality Assessment through FSIM, SSIM, MSE, and PSNR—A Comparative Study" . *Journal of Computer and Communications*, 2019, **7**, 8-18. doi: 10.4236/jcc.2019.73002

[5] K. Tyagi, C. Rane, Harshvardhan, M. Manry, "Regression analysis", Artificial Intelligence and Machine Learning for EDGE Computing, Academic Press, 2022, Chapter 4, Pages 53-63, ISBN 9780128240540, https://doi.org/10.1016/B978-0-12-824054-0.00007-1.

[6] R. G. Deshpande, L. L. Ragha, and S. K. Sharma, "Video Quality Assessment through PSNR Estimation for Different Compression Standards", Indonesian Journal of Electrical Engineering and Computer Science, Vol. 11, No. 3, 2018, pp. 918~924 ISSN: 2502-4752, DOI: 10.11591/ijeecs.v11.i3.pp918-924

[7] Wang. Z, Bovik. A. C., "Mean squared error: love it or leave it? A new look at signal fidelity measures", IEEE Signal Process. Mag. 26 (1) (2009) 98-117.

[8] Zhou Wang and A. C. Bovik, "A universal image quality index," IEEE Signal Processing Letters, vol. 9, no. 3, pp. 81-84, March 2002, doi: 10.1109/97.995823.

[9] W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Process., vol. 13, no.

4, pp. 600–612, Apr. 2004

[10]   C. F. Li, and A. C. Bovik, "Three-Component Weighted Structural Similarity Index. Image   Quality and System Performance", VI, SPIE Proc. 7242, San Jose, CA, 19 January 2009, 1-9

[11]   M. -J. Chen and A. C. Bovik, "Fast structural similarity index algorithm," 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, TX, USA, 2010, pp. 994-997, doi: 10.1109/ICASSP.2010.5495310.

[12]   G. -h. Chen, C. -l. Yang and S. -l. Xie, "Gradient-Based Structural Similarity for Image Quality Assessment," 2006 International Conference on Image Processing, Atlanta, GA, USA, 2006, pp. 2929-2932, doi: 10.1109/ICIP.2006.313132.

[13]   Z. Wang, and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Transactions on Image Processing,* vol. 20, no. 5, pp. 1185-1198, 2011.

[14]   X. Zhang, X. Feng, W. Wang and W. Xue, "Edge Strength Similarity for Image Quality Assessment," in IEEE Signal Processing Letters, vol. 20, no. 4, pp. 319-322, April 2013, doi: 10.1109/LSP.2013.2244081.

[15]   M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik and M. K. Markey, "Complex Wavelet Structural Similarity: A New Image Similarity Index," in *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2385-2401, Nov. 2009, doi: 10.1109/TIP.2009.2025923.

[16]   Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals Syst. Comput.*, vol. 2. Pacific Grove, CA, USA, 2003, pp. 1398–1402.

[17]   L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[18]   J. Zujovic, T. N. Pappas and D. L. Neuhoff, "Structural similarity metrics for texture analysis and retrieval," *2009 16th IEEE International Conference on Image Processing (ICIP)*, Cairo, Egypt, 2009, pp. 2225-2228, doi: 10.1109/ICIP.2009.5413897.

 [19]   H. R. Sheikh, A. C. Bovik, and G. D. Veciana, "An information fidelity criterion for

image quality assessment using natural scene statistics," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005.

[20]   H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[21]   M. Esmaeilpour, A. Mansouri and A. Mahmoudi-Aznaveh, "A new SVD-based image quality assessment," *2013 8th Iranian Conference on Machine Vision and Image Processing (MVIP)*, Zanjan, Iran, 2013, pp. 370-374, doi: 10.1109/IranianMVIP.2013.6780013.

[22]   W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," IEEE Trans. Image Process., vol. 23, no. 2, pp. 684–695, Feb. 2014.

[23]   D. O. Kim, H. S. Han, and R. H. Park, "Gradient information-based image quality metric," IEEE Transactions on Consumer Electronics, vol. 56, no. 2, pp. 930-936, 2010.

[24]   G. Q. Cheng, J. C. Huang, C. Zhu, Z. Liu and L. Z. Cheng, "Perceptual image quality assessment using a geometric structural distortion model," 17th IEEE International Conference on Image Processing, 2010.

[25]   G. H. Chen, C. L. Yang and S. L. Xie, "Gradient-based structural similarity for image quality assessment," 13th IEEE International Conference on Image Processing, 2006.

[26]   X. Ding, Z. Zhang, X. Chen and Y. Huang, "A novel pooling strategy for Full Reference Image Quality Assessment based on harmonic means," *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, South Brisbane, QLD, Australia, 2015, pp. 1672-1676, doi: 10.1109/ICASSP.2015.7178255.

[27]   D. Narsaiah, R. S. Reddy, A. Kokkula, P. A. Kumar and A. Karthik, "A Novel Full Reference-Image Quality Assessment (FR-IQA) for Adaptive Visual Perception Improvement," *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, Coimbatore, India, 2021, pp. 726-730, doi: 10.1109/ICICT50816.2021.9358610.

145

[28] Z. Wang, and Q. Li, "Information content weighting for perceptual image quality assessment," IEEE Transactions on Image Processing, vol. 20, no. 5, pp. 1185-1198, 2011.

[29] I. Bakurov, M. Buzzelli, R. Schettini, M. Castelli and L. Vanneschi, "Full-Reference Image Quality Expression via Genetic Programming," in IEEE Transactions on Image Processing, vol. 32, pp. 1458-1473, 2023, doi: 10.1109/TIP.2023.3244662.

[30] Z. Tang, Y. Zheng, K. Gu, K. Liao, W. Wang and M. Yu, "Full-Reference Image Quality Assessment by Combining Features in Spatial and Frequency Domains," in IEEE Transactions on Broadcasting, vol. 65, no. 1, pp. 138-151, March 2019, doi: 10.1109/TBC.2018.2871376.

[31] C. Shi and Y. Lin, "Full Reference Image Quality Assessment Based on Visual Salience With Color Appearance and Gradient Similarity," in IEEE Access, vol. 8, pp. 97310-97320, 2020, doi: 10.1109/ACCESS.2020.2995420.

[32] S. Seo, S. Ki and M. Kim, "A Novel Just-Noticeable-Difference-Based Saliency-Channel Attention Residual Network for Full-Reference Image Quality Predictions," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 7, pp. 2602-2616, July 2021, doi: 10.1109/TCSVT.2020.3030895.

[33] W. Kim, A. -D. Nguyen, S. Lee and A. C. Bovik, "Dynamic Receptive Field Generation for Full-Reference Image Quality Assessment," in IEEE Transactions on Image Processing, vol. 29, pp. 4219-4231, 2020, doi: 10.1109/TIP.2020.2968283.

[34] J. Yang, Z. Bian, Y. Zhao, W. Lu and X. Gao, "Full-Reference Quality Assessment for Screen Content Images Based on the Concept of Global-Guidance and Local-Adjustment," in IEEE Transactions on Broadcasting, vol. 67, no. 3, pp. 696-709, Sept. 2021, doi: 10.1109/TBC.2021.3064266.

[35] M. Liu et al., "Long-Range Dependencies and High-Order Spatial Pooling for Deep Model-Based Full-Reference Image Quality Assessment," in IEEE Access, vol. 8, pp. 72007-72020, 2020, doi: 10.1109/ACCESS.2020.2984886.

[36] Z. Wang, E.P. Simoncelli, Reduced-reference image quality assessment using a wavelet-domain natural image statistic model, in: Proceedings of the SPIE Conference,

2005, pp. 149–159.

[37]   Q. Li and Z. Wang, "Reduced-Reference Image Quality Assessment Using Divisive Normalization-Based Image Representation," in IEEE Journal of Selected Topics in Signal Processing, vol. 3, no. 2, pp. 202-211, April 2009, doi: 10.1109/JSTSP.2009.2014497.

[38]   L. Ma, S. Li, F. Zhang, K.N. Ngan, Reduced-reference image quality assessment using reorganized DCT-based image representation, IEEE Trans. Multimedia 13 (4) (2011) 824–829.

[39]   A. Rehman, Z. Wang, Reduced-reference image quality assessment by structural similarity estimation, IEEE Trans. Image Process. 21 (8) (2012) 3378–3389.

[40]   J.J. Wu, W.S. Lin, Y.M. Fang, L.D. Li, G.M. Shi, Visual structural degradation based reduced-reference image quality assessment, Signal Process., Image Commun. 47 (2016) 16–27.

[41]   W. Zhu et al., "Multi-Channel Decomposition in Tandem With Free-Energy Principle for Reduced-Reference Image Quality Assessment," in IEEE Transactions on Multimedia, vol. 21, no. 9, pp. 2334-2346, Sept. 2019, doi: 10.1109/TMM.2019.2902484.

[42]   Z. Wan, K. Gu and D. Zhao, "Reduced Reference Stereoscopic Image Quality Assessment Using Sparse Representation and Natural Scene Statistics," in IEEE Transactions on Multimedia, vol. 22, no. 8, pp. 2024-2037, Aug. 2020, doi: 10.1109/TMM.2019.2950533.

[43]   R. Soundararajan, A. Bovik, RRED Indices: Reduced reference entropic differencing for image quality assessment, IEEE Trans. Image Process. 21 (2) (2012) 517–526.

[44]   J.J. Wu, W.S. Lin, G.M. Shi, A. Liu, Reduced-reference image quality assessment with visual information fidelity, IEEE Trans. Multimed. 15 (7) (2013) 1700–1705.

[45]   K. Gu, G.T. Zhai, X.K. Yang, W.J. Zhang, A new reduced-reference image quality assessment using structural degradation model, in: International Symposium on Circuits and Systems, Beijing, China, 2013, pp. 1095–1098.

[46] J. Redi, P. Gastaldo, I. Heynderickx, R. Zunino, Color distribution information for the reduced-reference assessment of perceived image quality, IEEE Trans. Circuits Syst. Video Technol. 20 (12) (2010) 1757–1769.

[47] P. Paudyal, F. Battisti and M. Carli, "Reduced Reference Quality Assessment of Light Field Images," in IEEE Transactions on Broadcasting, vol. 65, no. 1, pp. 152-165, March 2019, doi: 10.1109/TBC.2019.2892092.

[48] Z. Huang and S. Liu, "Perceptual Hashing With Visual Content Understanding for Reduced-Reference Screen Content Image Quality Assessment," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 7, pp. 2808-2823, July 2021, doi: 10.1109/TCSVT.2020.3027001.

[49] K. Gu, J. Qiao, S. Lee, H. Liu, W. Lin and P. Le Callet, "Multiscale Natural Scene Statistical Analysis for No-Reference Quality Evaluation of DIBR-Synthesized Views," in *IEEE Transactions on Broadcasting*, vol. 66, no. 1, pp. 127-139, March 2020, doi: 10.1109/TBC.2019.2906768.

.[50] D. Lee and K. N. Plataniotis, "Toward a No-Reference Image Quality Assessment Using Statistics of Perceptual Color Descriptors," in *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3875-3889, Aug. 2016, doi: 10.1109/TIP.2016.2579308.

[51] B. Yan, B. Bare and W. Tan, "Naturalness-Aware Deep No-Reference Image Quality Assessment," in *IEEE Transactions on Multimedia*, vol. 21, no. 10, pp. 2603-2615, Oct. 2019, doi: 10.1109/TMM.2019.2904879.

[52] P. Ye and D. Doermann, "No-Reference Image Quality Assessment Using Visual Codebooks," in *IEEE Transactions on Image Processing*, vol. 21, no. 7, pp. 3129-3138, July 2012, doi: 10.1109/TIP.2012.2190086.

[53] Y. Bai, Z. Zhu, G. Jiang and H. Sun, "Blind Quality Assessment of Screen Content Images Via Macro-Micro Modeling of Tensor Domain Dictionary," in *IEEE Transactions on Multimedia*, vol. 23, pp. 4259-4271, 2021, doi: 10.1109/TMM.2020.3039382.

[54] Y. Liu, K. Gu, S. Wang, D. Zhao and W. Gao, "Blind Quality Assessment of Camera Images Based on Low-Level and High-Level Statistical Features," in *IEEE*

Transactions on Multimedia, vol. 21, no. 1, pp. 135-146, Jan. 2019, doi: 10.1109/TMM.2018.2849602

[55] J. Beron, H. D. Benitez-Restrepo and A. C. Bovik, "Blind Image Quality Assessment for Super Resolution via Optimal Feature Selection," in *IEEE Access*, vol. 8, pp. 143201-143218, 2020, doi: 10.1109/ACCESS.2020.3014497.

[56] L. Yu, J. Li, F. Pakdaman, M. Ling and M. Gabbouj, "MAMIQA: No-Reference Image Quality Assessment Based on Multiscale Attention Mechanism With Natural Scene Statistics," in *IEEE Signal Processing Letters*, vol. 30, pp. 588-592, 2023, doi: 10.1109/LSP.2023.3276645.

[57] H. R. Sheikh, A. C. Bovik and L. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," in *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1918-1927, Nov. 2005, doi: 10.1109/TIP.2005.854492.

[58] K. Gu, G. Zhai, W. Lin, X. Yang and W. Zhang, "No-Reference Image Sharpness Assessment in Autoregressive Parameter Space," in *IEEE Transactions on Image Processing*, vol. 24, no. 10, pp. 3218-3231, Oct. 2015, doi: 10.1109/TIP.2015.2439035.

[59] Y. Zhan and R. Zhang, "No-Reference JPEG Image Quality Assessment Based on Blockiness and Luminance Change," in *IEEE Signal Processing Letters*, vol. 24, no. 6, pp. 760-764, June 2017, doi: 10.1109/LSP.2017.2688371.

[60] S. V. R. Dendi, C. Dev, N. Kothari and S. S. Channappayya, "Generating Image Distortion Maps Using Convolutional Autoencoders With Application to No Reference Image Quality Assessment," in *IEEE Signal Processing Letters*, vol. 26, no. 1, pp. 89-93, Jan. 2019, doi: 10.1109/LSP.2018.2879518.

[61] Q. Li, W. Lin and Y. Fang, "No-Reference Quality Assessment for Multiply-Distorted Images in Gradient Domain," in *IEEE Signal Processing Letters*, vol. 23, no. 4, pp. 541-545, April 2016, doi: 10.1109/LSP.2016.2537321.

[62] Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang and G. Zhai, "No-Reference Quality Assessment of Contrast-Distorted Images Based on Natural Scene Statistics," in *IEEE Signal Processing Letters*, vol. 22, no. 7, pp. 838-842, July 2015, doi: 10.1109/LSP.2014.2372333.

[63] X. Pan, F. Xie, Z. Jiang, Z. Shi and X. Luo, "No-Reference Assessment on Haze for Remote-Sensing Images," in *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 12, pp. 1855-1859, Dec. 2016, doi: 10.1109/LGRS.2016.2614890.

[64] P. Mahajan, V. Jakhetiya, P. Abrol, P. K. Lehana, B. N. Subudhi and S. C. Guntuku, "Perceptual Quality Evaluation of Hazy Natural Images," in *IEEE Transactions on Industrial Informatics*, vol. 17, no. 12, pp. 8046-8056, Dec. 2021, doi: 10.1109/TII.2021.3065439.

[65] Joshi, P., Prakash, S.: Continuous wavelet transform based no-reference image quality assessment for blur and noise distortions. IEEE Access **6**, 33871–33882 (2018). https://doi.org/10.1109/ACCESS.2018.2846585

[66] Rajchel, M., Oszust, M.: No-reference image quality assessment of authentically distorted images with global and local statistics. SIViP **15**, 83–91 (2021). https://doi.org/10.1007/s11760-020-01725-0

[67] Li, S., Ding, Y. & Chang, Y.: No-reference stereoscopic image quality assessment based on cyclopean image and enhanced image. SIViP 14, 565–573 (2020). https://doi.org/10.1007/s11760-019-01582-6

[68] Li, Q., Wang, Z.: Reduced-reference image quality assessment using divisive normalization-based image representation. IEEE Journal of Selected Topics in Signal Processing, **3**(2), (2009)

[69] Lyu, S., and Simoncelli, E. P.: Statistically and perceptually motivated nonlinear image representation," in Proc. SPIE Conf. Human VisionElectron. Imaging XII, Jan. 2007, vol. 6492, pp. 649207–1–649207–15

[70] Heeger, D.J.: Normalization of cell responses in cat striate cortex. Vis. Neural Sci. **9**, 181–198 (1992)

[71] Wainwright, M.J.: Visual adaptation as optimal information transmission. Vis. Res. **39**, 3960–3974 (1999)

[72] Simoncelli, E.P., Freeman, W.T., Adelson, E.H., Heeger, D.J.: Shiftable multi-scale transforms. IEEE Trans. Inf. Theory **38**(2), 587–607 (1992)

[73]   Simoncelli, E.P., Olshausen, B.: Natural image statistics and neural representation. Annu. Rev. Neurosci. **24**, 1193–1216 (2001)

[74]   Wainwright, M.J., Simoncelli, E.P.: Scale mixtures of Gaussians and the statistics of natural images. Adv. Neural Inf. Process. Syst. **12**, 855–861 (2000)

[75]   Pham-Gia, T., Hung, T.L.: The mean and median absolute deviations. Math. Comput. Model. **34**(7–8), 921–936 (2001)

[76]   Larson, E. C., and Chandler, D. M.: Most apparent distortion: full-reference image quality assessment and the role of strategy. Journal of Electronic Imaging, **19**(1), (2010). http://vision.okstate.edul/?loc=csiq

[77]   Sheikh, H. R., Wang, Z., Bovik, A. C., and Cormack, L. K.: Image and video quality assessment research at LIVE. [Online]. Available:https://live.ece.utexas.edu/research/quality/

[78]   https://www.kaggle.com/datasets/kaustubhdikshit/neu-surface-defect-database/code

[79]   https://www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset

[80]   E. Tola, V. Lepetit, and P. Fua, "Daisy: An efficient dense descriptor applied to wide-baseline stereo," IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 5, pp. 815–830, 2010

[81]   Z. Wang, B. Fan, and F. Wu, "Local intensity order pattern for feature description," in Proc. (IEEE) International Conference on Computer Vision, 2011, pp. 603–610.

[82]   E. Tola, V. Lepetit, and P. Fua, "Daisy: An efficient dense descriptor applied to wide-baseline stereo," IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 5, pp. 815–830, 2010.

[83]   S. Kim, D. Min, B. Ham, M. N. Do, and K. Sohn, "Dasc: Robust dense descriptor for multi-modal and multi-spectral correspondence estimation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 9, pp. 1712–1729, 2017.

[84]   H. Roy and D. Bhattacharjee, "Local-Gravity-Face (LG-face) for Illumination-Invariant and Heterogeneous Face Recognition," in *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 7, pp. 1412-1424, July 2016, doi:

10.1109/TIFS.2016.2530043.

[85]   M. Oszust, "Local Feature Descriptor and Derivative Filters for Blind Image Quality Assessment," in *IEEE Signal Processing Letters*, vol. 26, no. 2, pp. 322-326, Feb. 2019, doi: 10.1109/LSP.2019.2891416.

[86]   Bhattacharjee and H. Roy, "Pattern of Local Gravitational Force (PLGF): A Novel Local Image Descriptor," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 2, pp. 595-607, 1 Feb. 2021, doi: 10.1109/TPAMI.2019.2930192.

[87]   Bhattacharjee and H. Roy, "Pattern of Local Gravitational Force (PLGF): A Novel Local Image Descriptor," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 2, pp. 595-607, 1 Feb. 2021, doi: 10.1109/TPAMI.2019.2930192. PLGA

[88]   Bhattacharjee and H. Roy, "Pattern of Local Gravitational Force (PLGF): A Novel Local Image Descriptor," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 2, pp. 595-607, 1 Feb. 2021, doi: 10.1109/TPAMI.2019.2930192.

[89]   E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," Journal of Electronic Imaging, 19 (1), March 2010.

[90]   H.R.   Sheikh, Z.Wang, L. Cormack and A.C. Bovik, "LIVE Image Quality Assessment Database Release 2",   http://live.ece.utexas.edu/research/quality.

[91]   N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J.   Astola, B.   Vozel, K. Chehdi, M. Carli, F. Battisti, C.-C. Jay Kuo,   Color   Image   Database   TID2013: Peculiarities and Preliminary Results,   Proceedings of 4th Europian Workshop on Visual Information   Processing EUVIP2013, Paris, France, June 10-12, 2013, pp. 106-111.

[92] Lin Gao,   November 25, 2019,   "Five   class weather image dataset", IEEE Dataport, doi: https://dx.doi.org/10.21227/gv99-q408.

[93]   Kumar, N., Nachamai, M.: Noise removal and filtering techniques used in medical images. Oriental Journal of Computer Science & Technology; ISSN:0974–6471 **10**(1), 103–113 (2017)

[94]   Patel, N., Shah, A., Mistry, M., Dangarwala, K.: A study of digital image filtering

techniques in spatial image processing. International Conference on Convergence of Technology, IEEE, (2014)

[95] Webster, J.: Methods for image quality assessment. Wiley Encyclopedia of Electrical and Electronics Engineering, John Wiley & Sons (2015) https://doi.org/10.1002/047134608X.W8282

[96] Ferzli, R., and Karam, L. J.: A no-reference objective image sharpness metric based on the notion of Just Noticeable Blur (JNB). IEEE Transactions on Image Processing, **18**(4) (2009)

[97] Narvekar, N.D., and Karam, L. J.: An improved no-reference sharpness metric based on the probability of blur detection. International Workshop on Quality of Multimedia Experience, (2009)

[98] Marziliano, P., Dufaux, F., Winkler, S., Ebrahimi, T.: Perceptual blur and ringing metrics: application to JPEG2000. Signal Process. Image Commun. **19**, 163–172 (2004)

[99] Ferzli, R., and Karam, L. J.: No-reference objective wavelet based noise immune image sharpness metric. IEEE International Conference on Image Processing 2005, pp. I-405, (2005) https://doi.org/10.1109/ICIP.2005.1529773

[100] Chung, P.-C., Wavg, J. M., Bailey, R., Chien, S.-W., Chang;, S.-L.: A non-parametric blur measure based on edge analysis for image processing applications. Proceedings of the 2004 IEEE Conference on Cybernetics and Intelligent Systems Singapore, 1–3 (2004)

[101] Varadarajan, S., and Karam, L. J.: An improved perception-based no-reference objective image sharpness metric using iterative edge refinement. 2008 15th IEEE International Conference on Image Processing, pp. 401–404 (2008) https://doi.org/10.1109/ICIP.2008.4711776

[102] Li, Z., Liu, Y., Xu, J., and Du, H.: A no-reference perceptual blur metric based on the blur ratio of detected edges. 2013 5th IEEE International Conference on Broadband Network & Multimedia Technology, 2013, pp. 1–5 (2013) https://doi.org/10.1109/ICBNMT.2013.6823903.

[103]   Caviedesa, J., Obertib, F.: A new sharpness metric based on local kurtosis, edge and energy information. Signal Process.: Image Commun. **19**, 147–161 (2004)

[104]   Bahrami, K., and Kot, A.C.: A fast approach for no-reference image sharpness assessment based on maximum local variation. IEEE Signal Processing Letters, **21**(6), (2014)

[105]   Kržić, A. S., Đonlić, M., Pejčinović, M., and Seršić, D.: Image sharpness assessment based on local phase coherence and LAD criterion. 2016 International Conference on Systems, Signals and Image Processing (IWSSIP), 2016, pp. 1–4, https://doi.org/10.1109/IWSSIP.2016.7502724

[106]   Breiman LRandom forestsMach Learn20014553210.1023/A:1010933404324

[107] Caviedes J, Gurbuz S, "No-reference sharpness metric based on local edge kurtosis", Proc IEEE Int Conf Image Processing 200235356

[108]   Cunningham P, Delany SJ (2007) k-Nearest neighbour classifiers. Technical Report UCD-CSI-2007–4 March 27

[109] Freeman W, Adelson E (n.d.) The design and use of steerable filters. IEEE Transactions on Pattern Analysis and Machine Intelligence, https://doi.org/10.1109/34.93808

[110]   Garg K, Nayar SK (2007) Vision and Rain. International Journal of Computer Vision, Springer, https://doi.org/10.1007/s11263-006-0028-6

[111]   Ghadiyaram D, Bovik AC (n.d.) Blind image quality assessment on real distorted images using deep belief nets. IEEE Global Conference on Signal and Information Processing (GlobalSIP), https://doi.org/10.1109/GlobalSIP.2014.7032260

[112]   Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B (n.d.) Support vector machines. IEEE Intelligent Systems and their Applications, https://doi.org/10.1109/5254.708428

[113]Image Deraining Dataset (2020) https://engineering.jhu.edu/vpatel36/datasets/ (Last Access on 20th Decem ber 2020)

[114]   Jiang K et al. (n .d.) Decomposition makes better rain removal: An improved attention-guided deraining network, in IEEE Transactions on Circuits and Systems for Video

Technology, https://doi.org/10.1109/TCSVT.2020.3044887

[115]   Khare V, Shivakumara P, Kumar A, Chan CS, Lu T, Blumenstien M (2016) A quad tree based method for blurred and non-blurred video text frames classification through quality metrics. 23rd international conference on pattern recognition (ICPR)

[116]   Marziliano P, Dufaux F, Winkler S, Ebrahimi T (2004) Perceptual blur and ringing metrics: application to JPEG2000. Signal Process:ImageCommun 19(2):163–172

[117]   Meesters L, Martens JB (2002) A single-ended blockiness measure for JPEG-coded images. Signal Process 82(3):369–387

[118]   .Mittal A, Moorthy AK, Bovik AC (2012) No-reference image quality assessment in the spatial domain," IEEE Transactions on Image Processing

[119]   I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al., "Generative adversarial nets", *in Advances in Neural Information Processing Systems*, vol. 27, pp. 2672, 2014.

[120]   He Zhang, Vishwanath Sindagi, Vishal M. Patel, "Image De-raining Using a Conditional Generative Adversarial Network", https://doi.org/10.48550/arXiv.1701.05957

[121]   Xueyang Fu ; Jiabin Huang ; Delu Zeng ; Yue Huang ; Xinghao Ding ; John Paisley;

"Removing Rain from Single Images via a Deep Detail Network"; IEEE Conference on

Computer Vision and Pattern Recognition (CVPR), DOI: 10.1109/CVPR.2017.186, 09

November 2017.

[122]   He Zhang, Vishal M. Patel; "Density-aware Single Image De-raining using a Multi-

stream Dense Network"; arXiv:1802.07412 [cs.CV]

[123]   Q. Peng, W. Luo, G. Hong, M. Feng, Y. Xia, L. Yu, X. Hao, X. Wang, M. Li; "Pedestrian Detection for Transformer Substation Based on Gaussian Mixture Model and YOLO"; 2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), 15 December 2016

[124]   P. Ren, W. Fang, S. Djahel; "A novel YOLO-Based real-time people counting approach"; 2017 International Smart Cities Conference (ISC2), 02 November 2017.

[125]   W. Lan, J. Dang ; Y. Wang ; S. Wang; "Pedestrian Detection Based on YOLO Network Model"; 2018 IEEE International Conference on Mechatronics and Automation(ICMA), 08 October 2018.

[126]   W. Yang, Z. Jiachun; "Real-time face detection based on YOLO"; 2018 1st IEEE International Conference on Knowledge Innovation and Invention (ICKII); 10 December 2018.

[127]   R. Widyastuti, C. K. Yang; "Cat's Nose Recognition Us-ing You Only Look Once (Yolo) and Scale-Invariant Feature Transform (SIFT)"; 2018 IEEE 7th Global Conference on Consumer Electronics (GCCE), 13 December 2018.

[128]   J. P. Lin, M. T. Sun; "A YOLO-Based Traffic Counting System"; 2018 Conference on Technologies and Applications of Artificial Intelligence (TAAI); 27 December 2018.

[129]   Joseph Redmon, Ali Farhadi; "YOLOv3: An Incremental Improvement"; arXiv:1804.02767, 2018

[130]   R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 2014, pp. 580-587, doi: 10.1109/CVPR.2014.81.

[131]   R. Girshick, "Fast R-CNN," *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169.

[132]   R. Gavrilescu, C. Zet, C. Foşalău, M. Skoczylas and D. Cotovanu, "Faster R-CNN: an Approach to Real-Time Object Detection," *2018 International Conference and Exposition on Electrical And Power Engineering (EPE)*, Iasi, Romania, 2018, pp. 0165-0168, doi: 10.1109/ICEPE.2018.8559776.

[133]   Joseph Redmon, Ali Farhadi, *"YOLOv3: An Incremental Improvement", https://doi.org/10.48550/arXiv.1804.02767*

[134]   Liu, W. *et al.* (2016). SSD: Single Shot MultiBox Detector. In: Leibe, B., Matas, J.,

Sebe, N., Welling, M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science(), vol 9905. Springer, Cham. https://doi.org/10.1007/978-3-319-46448-0_2

[135] Banitalebi-Dehkordi M, Khademi M, Ebrahimi-Moghadam A, Hadizadeh H (2019) An image quality assessment algorithm based on saliency and sparsity. Multimed Tools Appl 78(9):11507–11526

[136] Bapat A, Kanhangad V (2017) Segmentation of hand from cluttered backgrounds for hand geometry biometrics. In: 2017 IEEE Region 10 Symposium (TENSYMP). IEEE, pp 1–4

[137] Barra S, De Marsico M, Nappi M, Narducci F, Riccio D (2019) A hand-based biometric system in visible light for mobile environments. Inf Sci 479:472–485

[138] Bartuzi E, Trokielewicz M (2018) Thermal features for presentation attack detection in hand biometrics. In: 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS). IEEE, pp 1–6

[139] Bera A, Bhattacharjee D, Nasipuri M (2014) Hand biometrics in digital forensics. In: Computational Intelligence in Digital Forensics: Forensic Investigation and Applications. Springer, pp 145–163

[140] Bera A, Bhattacharjee D, Nasipuri M (2017) Finger contour profile based hand biometric recognition. Multimed Tools Appl 76(20):21451–21479

[141] Bera A, Bhattacharjee D, Nasipuri M (2015) Fusion-based hand geometry recognition using Dempster-Shafer theory. Int J Pattern Recognit Artif Intell 29(05):1556005

[142] Bera A, Bhattacharjee D, Nasipuri M (2019) Finger biometric recognition with feature selection. Biometr Comput Recogn Registr:87

[143] Bera A, Bhattacharjee D (2020) Human identification using selected features from finger geometric profiles. IEEE Trans Syst Man Cybern Syst 50(3):747–761

[144] Bhilare S, Kanhangad V, Chaudhari N (2018) A study on vulnerability and presentation attack detection in palmprint verification system. Pattern Anal Appl 21(3):769–782

[145]   Biggio B, Fumera G, Marcialis GL, Roli F (2016) Statistical meta-analysis of presentation attacks for secure multibiometric systems. IEEE Trans Pattern Anal Mach Intell 39(3):561–575

[146]   Bondzulic B, Petrovic V, Andric M, Pavlovic B (2018) Gradient-based image quality assessment. Acta Polytech Hungarica 15(4)

[147]   Bong DBL, Khoo BE (2014) Blind image blur assessment by using valid reblur range and histogram shape difference. Signal Process Image Commun 29(6):699–710

[148]   Breiman L (2001) Random forests. Mach Learn 45(1):5–32

[149]   Chatterjee A, Singh P, Bhatia V, Prakash S (2018) A low-cost optical sensor for secured antispoof touchless palm print biometry. IEEE Sens Lett 2(2):1–4

[150]   Chen H, Valizadegan H, Jackson C, Soltysiak S, Jain AK (2005) Fake hands: spoofing hand geometry systems. Biometric Consortium

[151]   Chingovska I, Dos Anjos AR, Marcel S (2014) Biometrics evaluation under spoofing attacks. IEEE Trans Inf Forensic Secur 9(12):2264–2276

[152]   Chugh T, Cao K, Jain AK (2018) Fingerprint spoof buster: Use of minutiae-centered patches. IEEE Trans Inf Forensic Secur 13(9):2190–2202

[153]   Czajka A, Bulwan P (2013) Biometric verification based on hand thermal images. In: 2013 International Conference on Biometrics (ICB). IEEE, pp 1–6

[154]   Doi J, Yamanaka M (2005) Discrete finger and palmar feature extraction for personal authentication. IEEE Trans Instrum Measur 54(6):2213–2219

[155]   Dutagaci H, Sankur B, Yörük E (2008) Comparative analysis of global hand appearance-based person recognition. J Electron Imaging 17(1):011018

[156]   Farmanbar M, Toygar O (2017) Spoof detection on face and palmprint biometrics. SIViP 11(7):1253–1260

[157] Faundez-Zanuy M, Mekyska J, Font-Aragon`es X (2014) A new hand image database simultaneously acquired in visible, near-infrared and thermal spectrums. Cogn Comput 6(2):230–24

[158] Ferrer MA, Morales A, D´ıaz A (2014) An approach to swir hyperspectral hand biometrics. Inf Sci 268:319

[159] Galbally J, Marcel S, Fierrez J (2014) Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition. IEEE Trans Image Process 23(2):710–724

[160] Wu Z, Yamagishi J, Kinnunen T, Hanilc ¿ C, Sahidullah M, Sizov A, Evans N, Todisco M, Delgado H (2017) Asvspoof: the automatic speaker verification spoofing and countermeasures challenge. IEEE J Sel Top Signal Process 11(4):588–604

[161] Gamassi M, Lazzaroni M, Misino M, Piuri V, Sana D, Scotti F (2005) Quality assessment of biometric systems: a comprehensive perspective based on accuracy and performance measurement. IEEE Trans Instrum Measur 54(4):1489–1496

[162] Gao H, Miao Q, Yang J, Ma Z (2018) Image quality assessment using image description in information theory. IEEE Access 6:47181–47188

[163] Guan J, Yi S, Zeng X, Cham W-K, Wang X (2017) Visual importance and distortion guided deep image quality assessment framework. IEEE Trans Multimed 19(11):2505–2520

[164] Harvey J, Campbell J, Adler A (2018) Characterization of biometric template aging in a multiyear, multivendor longitudinal fingerprint matching study. IEEE Trans Instrum Meas 68(4):1071–1079

[165] Jaswal G, Kaul A, Nath R (2019) Multimodal biometric authentication system using hand shape, palm print, and hand geometry. In: Computational Intelligence: Theories, Applications and Future Directions-Volume II. Springer, pp 557–570

[166] Jia S, Guo G, Xu Z, Wang Q (2020) Face presentation attack detection in

mobile scenarios: A comprehensive evaluation. Image Vis Comput 93:103826

[167] Klonowski M, Plata M, Syga P (2018) User authorization based on hand geometry without special equipment. Pattern Recogn 73:189–201

[168] Korshunov P, Marcel S (2017) Impact of score fusion on voice biometrics and presentation attack detection in cross-database evaluations. IEEE J Sel Top Signal Process 11(4):695–705

[169] Liu A, Lin W, Narwaria M (2011) Image quality assessment based on gradient similarity. IEEE Trans Image Process 21(4):1500–1512

[170] Martini MG, Hewage CTER, Villarini B (2012) Image quality assessment based on edge preservation. Signal Process Image Commun 27(8):875–882

[171] Fourati E, Elloumi W, Chetouani A (2020) Anti-spoofing in face recognition-based biometric authentication using image quality assessment. Multimed Tools Appl 79(1-2):865–889

[172] Pinto A, Goldenstein S, Ferreira A, Carvalho T, Pedrini H, Rocha A (2020) Leveraging shape, reflectance and albedo from shading for face presentation attack detection. IEEE Trans Inf Forensic Secur 15:3347–3358

[173] Raghavendra R, Busch C (2015) Robust scheme for iris presentation attack detection using multiscale binarized statistical image features. IEEE Trans Inf Forensic Secur 10(4):703–715

[174] Vajapeyam S (2014) Understanding Shannon's Entropy metric for Information. Preprint arXiv:1405.2061

[175] Moon WK, Lee Y-W, Ke H-H, Lee SH, Huang C-S, Chang R-F (2020) Computer‑aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. Comput Methods Programs Biomed 190:105361. ISSN 0169-2607.https://doi.org/10.1016/j.cmpb.2020.105361

[176] Abdel-Nasser M, Melendez J, Moreno A, Omer OA, Puig D (2017) Breast tumor classification in ultrasound images using texture analysis and super-resolution methods. Eng Appl Artif Intell 59:84–92