PHASE-BASED TECHNIQUES FOR IMAGE AND VIDEO PROCESSING APPLICATIONS

A DISSERTATION SUBMITTED TO
THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

By
Serdar Çakır
December 2017

Phase-Based Techniques For Image And Video Processing Applications
By Serdar Çakır
December 2017

We certify that we have read this dissertation and that in our opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

A. Enis Çetin(Advisor)

B. Uğur Töreyin

S. Serdar Kozat

R. Gökberk Cinbiş

Uğur Güdükbay

Approved for the Graduate School of Engineering and Science:

Ezhan Karaşan
Director of the Graduate School

ABSTRACT

PHASE-BASED TECHNIQUES FOR IMAGE AND VIDEO PROCESSING APPLICATIONS

Serdar Çakır
Ph.D. in Electrical and Electronics Engineering
Advisor: A. Enis Çetin
December 2017

In this thesis, phase information is utilized to address several issues in image processing applications; namely image quality assessment, image contrast enhancement, and visual object tracking. The classical two-dimensional (2D) mel-cepstrum features, which ignore the phase information by design, are enhanced with image phase to form the 2D complex mel-cepstrum features. While integrating the phase information with the existing ceptral features, the unwrapping of phase information is carried out. The 2D complex mel-cepstrum features are fed into a regression scheme to map the feature matrices to subjective scores for the assessment of image quality. A Fourier domain approach for contrast enhancement of microscopy images is developed. The enhancement framework determines the frequency components in which the phase transitions are significant. The significant spectrum components are amplified by a factor depending on the level of transitions. In this way, phase variations are translated into amplitude changes which directly contribute to the enhancement process. Selective variation, which is an extension to the classical total variation framework, is introduced to determine the appropriate parameter set for the enhancement framework. The selective variation scheme evaluates the variations of the image in the high-frequency regions. A visual object tracking scheme based on image phase information is proposed. The main aim of the proposed scheme is to reduce the computational complexity of cross-correlation based matching frameworks. Starting from the derivation of normalized cross-correlation function, the tracking solution is simplified to a phase minimization problem under certain assumptions. The utilization of look-up tables for phase shifts enables a further decrease in computational cost.

Keywords: Image Phase Information, Phase Spectrum, Fourier Transform, 2D Complex Mel-Cepstrum, Image Quality Assessment, Contrast Enhancement, Selective Variation, Visual Object Tracking.

ÖZET

İMGE VE VİDEO İŞLEME UYGULAMALARI İÇİN FAZ TABANLI YAKLAŞIMLAR

Serdar Çakır
Elektrik ve Elektronik Mühendisliği, Doktora
Tez Danışmanı: Prof. Dr. A. Enis Çetin
Aralık 2017

Bu tezde faz bilgisi, imge kalite değerlendirmesi, imge karşıtlık arttırılması ve görsel nesne takibi gibi imge işleme uygulamalarında bir takım sorunları gidermek için kullanılmıştır. Tasarımı gereği, faz bilgilerini göz ardı eden klasik iki boyutlu (2B) mel-kepstrum öznitelikleri, imge fazıyla zenginleştirilerek 2B karmaşık mel-kepstrum özniteliklerini oluşturmuştur. Faz bilgisi ile mevcut kepstral öznitelikler birleştirilirken, ters faz geçişlerini önlemek üzere faz bilgisinin açılması gerçekleştirilmiştir. Görüntü kalite değerlendirmesi için, 2B karmaşık mel-kepstrum öznitelikleri, bir bağlanım yöntemine beslenmiş ve öznitelik matrisleri ile öznel puanlar arasında bir eşleştirme gerçekleştirilmiştir. Mikroskopik imgelerin karşıtlığının arttırılması için bir Fourier bölgesi yaklaşımı geliştirilmiştir. Karşıtlık arttırma yöntemi, faz geçişlerinin baskın olduğu frekans bileşenlerini belirlemektedir. Önemli tayf bileşenleri, geçiş seviyesine göre belirlenen bir faktörle kuvvetlendirilir. Bu yolla fazdaki geçişler, karşıtlık arttırma yönteminde doğrudan gözlemlenebilecek genlik değişimlerine çevrilmektedir. Karşıtlık arttırma yöntemi için uygun parametre kümesini belirlemek için, klasik toplam değişim yönteminin bir uzantısı olarak geliştirilen seçici değişim yöntemi tanıtılmıştır. Seçici değişim yöntemi, imgenin yüksek frekans bölgelerindeki değişimlerini değerlendirir. Gerçekleştirilen diğer bir çalışmada, imge faz bilgisine dayanan görsel nesne takip yöntemi önerilmiştir. Önerilen yöntemin ana amacı çapraz ilinti tabanlı eşleşme yöntemlerinin hesap karmaşasını azaltmaktır. Düzgelenmiş çapraz ilinti fonksiyonunun türetilmesinden yola çıkılarak, nesne takibi problemi, bazı varsayımlar altında faz en küçükleme problemine dönüştürülmüştür. Faz kaymaları için arama tablolarının kullanılması, hesaplama maliyetinde daha fazla azaltmayı mümkün kılmıştır.

Anahtar sözcükler: İmge Faz Bilgisi, Faz Tayfı, Fourier Dönüşümü, 2B Karmaşık Mel-Kepstrum, İmge Kalite Değerlendirmesi, Karşıtlık Arttırma, Seçici Değişim, Görsel Nesne Takibi.

Acknowledgement

Firstly, I would like to express my sincere gratitude to my advisor Prof. Dr. A. Enis Çetin for the continuous support of my Ph.D. study and related research, for his patience, motivation, and guidance. I could not have imagined having a better advisor and mentor for my Ph.D. study. He is indeed "more than an advisor" to me.

Besides my advisor, I would like to thank the rest of my thesis monitoring committee; Assoc. Prof. Dr. B. Uğur Töreyin, and Assoc. Prof. Dr. S. Serdar Kozat, for their insightful comments and encouragement throughout my Ph.D. study.

I would like to express my gratitude to the other members of Ph.D. Defence Committee, Assist. Prof. Dr. R. Gökberk Cinbiş, and Prof. Dr. Uğur Güdükbay, for reading, reviewing and making fruitful suggestions on the thesis.

I want to thank TÜBİTAK BİLGEM İLTAREN for creating opportunity to carry out my Ph.D. studies. I also want to thank all of my colleagues in the laboratory for discussing technical issues and giving ideas throughout my research.

I am thankful to Dr. Deniz Cansen Kahraman and Prof. Dr. Rengül Çetin Atalay for providing me the microscopy images used in my Ph.D. studies.

I can not forget to thank Mürüvet Parlakay who have helped me on every single issue I have faced. Thank you for your all help in complicated paperworks and formalities during my studies at Bilkent.

I want to thank Dr. Tayfun Aytaç for encouraging me throughout my Ph.D. study, proofreading my manuscripts, and for his inspiring remarks about my research and life.

I also want to thank my colleague Oğuzcan Dobrucalı for fruitful discussions

and chit chats in the office.

I want to thank the former members of the Bilkent SPG, Dr. Kıvanç Köse and Dr. Osman Günay for welcoming me and making me feel at home when I first arrived at Bilkent. Also, I want to thank all members of the Bilkent SPG for being kind and sharing.

I want to express my gratitude to Uzeler family for accepting me as their son and encouraging me throughout my Ph.D. study. It would be hard to concentrate on the research without their support and prayers.

I want to express my deepest thanks to my dear wife, best friend, and my better half Hande. She never stopped supporting me in desperate times and cheered me up. Also, thank you for being an editor, a proofreader, and a respectful colleague.

I would like to express my sincere gratitudes to my grandfather İsa Oylu for raising me, for his life-long advices, prayers and endless support.

Words fail to express my sincerest and deepest gratitude to my parents, Nevin & Hayrettin Çakır, for raising, supporting and encouraging me ever since my first breath. I wish I could have celebrated my accomplishments with my dad, nevertheless I know he will be watching me from above. I want to dedicate this thesis to him and his memory.

In this limited space, I may have forgotten to utter some of my friends' and colleagues' names however this does not mean that they are less valuable to me. Thanks to all of them.

Contents

List of Figures

List of Tables

To the memory of my beloved father Hayrettin ÇAKIR
and grandfather İsa OYLU ...

Chapter 1

Introduction

Fourier analysis plays an important role in estimating the behavior of signals in the frequency domain. The transform domain components, the magnitude and the phase of the Fourier transform, have different characteristics for signal representation. Although the phase and the magnitude components coexist in the spectrum in a perfect harmony, researchers have carried out large scale experiments to test the superiority of the components to each other [7,8]. Consequently, the phase component was shown to be the dominant factor on signal representation [7, 9]. Therefore, the phase information is preferred for the representation of both one dimensional and multi-dimensional signals.

In the literature, phase information is widely utilized for speech processing applications [10–16]. Speech enhancement, the process of enhancing speech which has been corrupted by additive noise, is an important field in speech processing. Previously, there had been a prejudice among researchers that the phase spectrum contains little useful information [17] or no information [16] for short time windows (20-40 ms). Conversely, recent advancements have shown that phase-spectrum based approaches can introduce significant performance increases in speech enhancement [11, 12]. In [12], the authors investigate short-time Fourier analysis-modification-synthesis (AMS) based speech enhancement approaches, which try to enhance the quality of speech by making use of the short-time phase

spectrum. In the experiments, both oracle (where undistorted phase spectrum is available) and non-oracle scenarios were considered. Significant quality improvements were achieved in the results for the noisy magnitude spectrum, especially if the undistorted phase spectrum was known. The improvements on the speech quality decrease significantly when only the noisy spectrum is available. This brings out the necessity of better phase spectrum estimation algorithms to make use of the short-time phase spectrum. In another work, parameter estimation in single channel speech enhancement is carried out by utilizing the phase information [15]. The authors focus on both two stages (parameter estimation and signal reconstruction) of the typical single-channel speech enhancement scheme. Instead of using the traditional Wiener filtering based approach, a minimum mean square error (MMSE) estimator is utilized in the parameter estimation stage. The MMSE estimator uses the undistorted ("clean") phase information as prior information. In the signal reconstruction stage, noise-free phase information is used instead of the traditional use of noisy phase information. The experimental results show that the utilization of phase information in both stages of the classical speech enhancement framework improves the signal quality better than the amplitude based approaches. Again, single channel speech enhancement framework [15] requires a successful phase estimation algorithm which is also the case for the method proposed in [12].

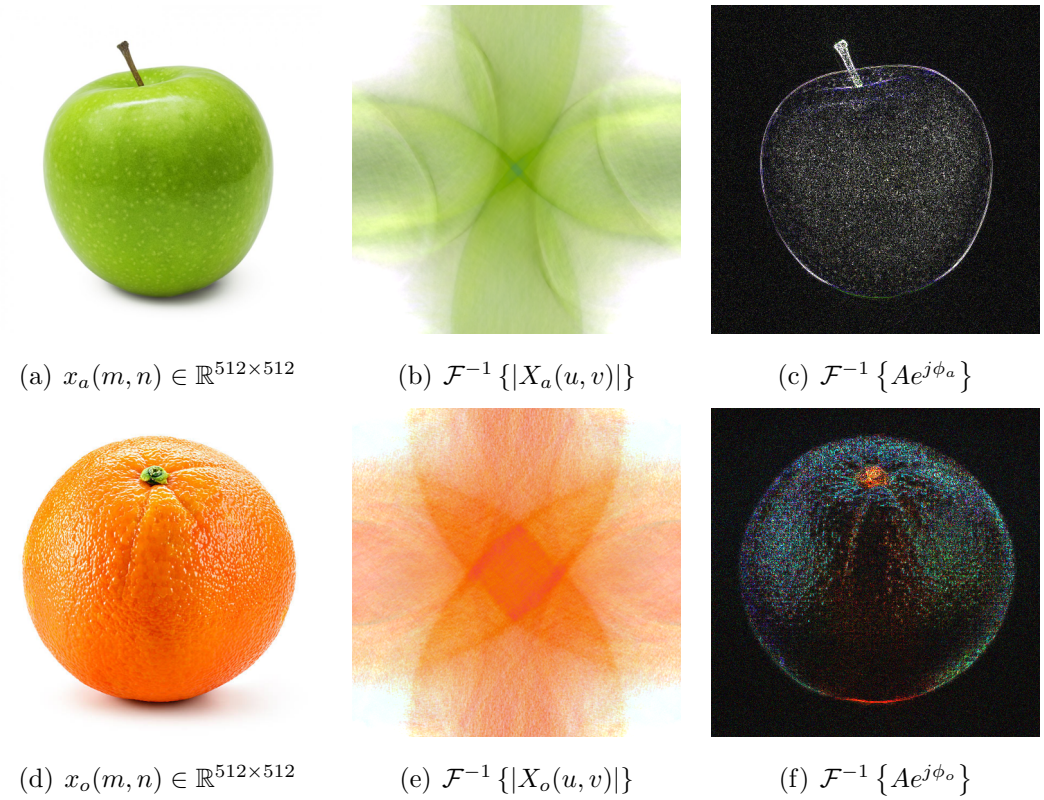
Speech recognition is another important field in which the phase information is utilized [10, 13, 14]. In [10], the effects of impurity in the phase spectrum on the recognition error rate have been analyzed. The relation between the recognition error rate and the amount of phase error is investigated for both low and high signal-to-noise ratio (SNR) scenarios. The authors state that, the effect of phase on the recognition error rate is small at high SNRs. Conversely, the effect of phase on the recognition error rate can be significant at low SNRs [10]. In another speech recognition study [13], a phase spectrum based approach is proposed as an alternative to the traditional magnitude-only spectral subtraction approach. The proposed approach seeks the possibility of perfect reconstruction of clean speech magnitudes from noisy samples by using the phase information. The authors have found that the error between the estimate and the clean speech depends on

the phase difference between the noisy and clean signals. Therefore, two alternatives have been proposed to perform subtraction in the frequency spectrum. The spectrum of clean speech can be obtained by either estimating the phase information of the noise or clean speech and using a tangent-based interpolation scheme to establish the clean speech spectrum. The experimental results show that a performance improvement of 20% has been achieved by using the exact phase information within a proof of concept experimental scheme. Phase-based feature representation framework is introduced for speech recognition [14]. In the proposed framework Autoregressive (AR) model extraction, group delay function (GDF) computation, power spectrum compression, and scale augmentation operations are carried out. A high-resolution estimate of the power spectrum is achieved by using the AR model together with the GDF .The two step compression scheme is similar to the mel-frequency ceptral coefficients (MFCC) without logarithmic mapping of the output energies. In the last stage, the scale information which is based on the Hilbert transform relations is used to enhance the phase-based feature vector. The experimental studies show that the phase-based features provide significant reductions on the error rates in the presence of additive and convolutional noises [14].

Image phase information has been also utilized for image processing applications [18–22] due to its representation capability. Oppenheim and his colleagues carried out a series of experiments on 2D images to analyze the effects of phase information in image synthesis [7, 8, 19]. In order to illustrate the importance of phase information in image representation, we repeated and expanded the experiments carried out by Oppenheim et.al. Let $x_a(m,n) \in \mathbb{R}^{M\times N}$ and $x_o(m,n) \in \mathbb{R}^{M\times N}$ be the images of an apple and an orange, respectively. The Fourier Transforms (FT) of these objects are calculated as follows:

$$\begin{aligned} \mathcal{F}\{x_a(m,n)\} &= X_a(u,v) = |X_a(u,v)|\, e^{j\phi_a} \\ \mathcal{F}\{x_o(m,n)\} &= X_o(u,v) = |X_o(u,v)|\, e^{j\phi_o} \end{aligned} \tag{1.1}$$

where $\mathcal{F}$ is the FT operator. The first experiment is evaluating the inverse FT by using the magnitude and phase information separately. The synthesis results of the first experiment are presented in Figure 1.1.



(a) $x_a(m,n) \in \mathbb{R}^{512\times512}$ (b) $\mathcal{F}^{-1}\{|X_a(u,v)|\}$ (c) $\mathcal{F}^{-1}\left\{Ae^{j\phi_a}\right\}$

(d) $x_o(m,n) \in \mathbb{R}^{512\times512}$ (e) $\mathcal{F}^{-1}\{|X_o(u,v)|\}$ (f) $\mathcal{F}^{-1}\left\{Ae^{j\phi_o}\right\}$

Figure 1.1: The illustration of magnitude-only and phase-only synthesis of the input images. The $\mathcal{F}^{-1}$ is inverse FT operator and $A = 100$ is a scaling factor.

By looking at the results presented in Figure 1.1, it can be observed that magnitude information is not sufficient for object representation. On the other hand, edges and even textural details are preserved in the synthesis using phase information even if the spectral magnitude is totally ignored and set to a constant value. The first experiment emphasizes the importance of image phase information for object representation. In order to analyze the impact of the FT components on the representation capability, an additional experiment has been performed using the image pair $x_a(m,n)$ and $x_o(m,n)$. In this experiment, the signal synthesis is carried out by using the magnitude spectrum of the first image with the phase spectrum of the other image, and vice versa. The synthesis results of the experiment are presented in Figure 1.2.

(a) $\mathcal{F}^{-1}\left\{|X_a(u,v)|\, e^{j\phi_o}\right\}$ (b) $\mathcal{F}^{-1}\left\{|X_o(u,v)|\, e^{j\phi_a}\right\}$

Figure 1.2: The illustration of image synthesis using the magnitude of the first image and phase information of the second image, and vice versa.

The synthesis results presented in Figure 1.2 show that the phase information is the dominant term in object representation. In the synthesis, the phase spectrum contributes to nearly all of the details of the object from which it was extracted and suppresses the magnitude information coming from the other object. In this way, the object in the synthesis result is more similar to the one from which the phase information is extracted.

In order to show the importance of phase information for signal representation, Oppenheim et. al. proposed an iterative algorithm to perfectly reconstruct the signal from its phase information [1]. The problem of reconstructing a signal using only the phase information is a robust inverse problem. The original image can be reconstructed from its phase within a scale factor when the support of the input signal is known [23]. A Papoulis-Gerchberg type of iterative algorithm for a given phase information was developed by Oppenheim et. al. In [1], the signal is reconstructed by successively imposing the support and phase information on iterates in space and Fourier domains. The iterative phase-only reconstruction technique [1] can be summarized as follows: Let $x[n] \in \mathbb{R}^N$ be a one dimensional real signal with $x[0] \neq 0$. The signal $x[n]$ is padded with zeros up to the size M

($M \gg N$) to form the signal $x_p[n] \in \mathbb{R}^M$. Then, the M-point Discrete Fourier Transform (DFT) is calculated and the phase of the input signal $\phi_x(k)$ is obtained. The iterative algorithm starts with an initial guess of the DFT magnitude which can be also set as a unit vector. In the second stage, the signal $X_{p+1}(k)$ is obtained by using the original phase information $\phi_x(k)$. Then, the M-point inverse DFT (IDFT) is computed. In the next iteration of the iterative algorithm, after the computation of the IDFT, the signal constraints should be applied on the signal $x_p[n]$. The signal constraints are the elements which are known a priori due to the zero-padding process. Then, the M-point DFT is computed and the phase of $Y_p(k)$ is replaced with the original phase information $\phi_x(k)$. The current iteration is finalized by computing the IDFT and the algorithm proceeds to next iteration. The algorithm converges to the original signal after some iterations. The flow diagram of the iterative algorithm is presented in Figure 1.3.

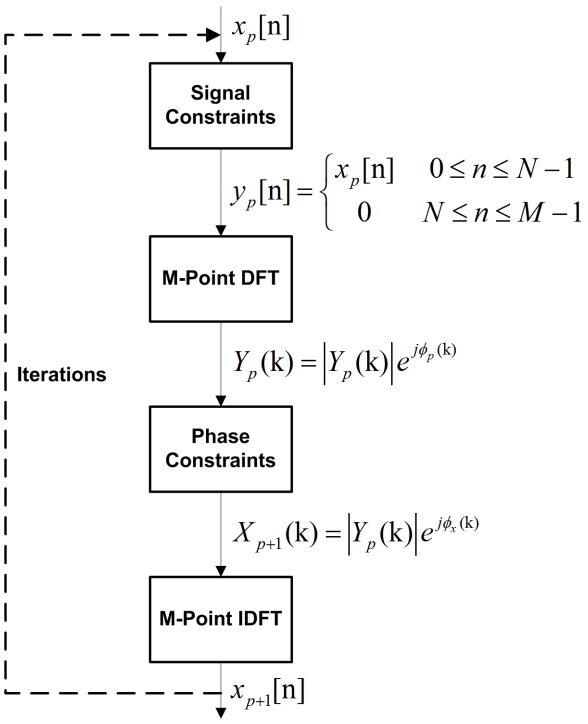


Figure 1.3: Flow diagram of the iterative phase-only reconstruction framework [1].

Phase information can be also utilized for image watermarking [20]. The method places the watermark on the most significant frequency components of an image. The reason behind this methodology is quite simple. If any operation

tries to damage the watermark, it may also unavoidably damage the image itself due to the occupation of significant frequency components by the watermark. The idea behind this approach is also consistent with the use of image phase in the transform domain due to the importance of phase as a major contributor in image representation. It should be also noted that, the phase information should be manipulated carefully by preserving the conjugate-symmetric nature of the spectrum.

In [24], a novel Fourier-based approach called WARP was introduced to match and retrieve similar shapes in images. The proposed technique contains two major stages: description by image phase information and Dynamic Time Warping (DTW) for image matching. The Fourier phase information enables a more representative description of the object boundaries compared to the amplitude-based schemes. Also, the DTW distance provides a more robust matching framework even in the presence of phase shifts. The performance of the WARP technique is superior than the performances of the selected baseline Fourier descriptors in terms of precision and recall levels [24].

Phase-based techniques have been widely utilized for matching applications starting from the work in [25]. Phase-only correlation (POC) [26] is one of the most popular matching frameworks in the literature. In [21], the authors proposed an efficient matching algorithm by using the 2D DFTs of input images for iris recognition. The experimental evaluations show that the use of phase spectrum of iris images provides higher accuracies of iris recognition even though a simple matching algorithm is utilized. In the article, the authors have introduced a coding scheme (2D Fourier Phase Code) in order to reduce the dimensionality of the data and to prevent the visibility of iris images due to privacy issues.

In [22], the phase spectrum is used for localization of salient areas in images. The authors demonstrated that the phase spectrum of Fourier Transform (PFT) provides better results compared to the Spectral Residual (SR) based scheme. Moreover, the proposed PFT-based scheme is computationally more efficient than the SR-based method. Actually, the PFT technique provides phase-only reconstruction of the input image followed by a smoothing filter as a saliency map. In

order to obtain spatio-temporal saliency maps, each pixel in the image is represented by additional features to form a "quaternion" composed of intensity, color, orientation and motion features.

In work [27], the authors make use of the phase-only FT (POFT) for defect detection in a production line. First of all, a defect-free control set is constituted by aligning images of products into an array. Then, the POFT is used to extract salient regions from the images in the control set. The same procedure is applied for the test images that may contain defects. In the article, the authors theoretically prove that POFT can eliminate the periodic components caused by the concatenated defect-free regions in the test images. Moreover, the POFT seems to significantly amplify the components that have nonperiodic and sparse behavior caused by the production defects. In the second stage of their algorithm, the regions that contain defects are determined by comparing the local image statistics near the salient regions of both the test and control sets. The experimental results show that the POFT-based defect detection achieves better performance compared to the baseline detection techniques [27].

1.1 Motivation

Image phase information has been a popular area of research due its significant advantages for signal representation. Researchers have been trying to develop new algorithms based on phase spectrum to overcome the limitations of magnitude-based approaches. Moreover, phase-based algorithms have changed the long standing belief about the superiority of the amplitude spectrum in signal representation. Due to its numerous advantages, we have been focusing our research on phase-based frameworks for the aim of achieving better representation, higher accuracy and computational efficiency.

The goal of this thesis is to develop phase-based approaches to address several issues in image processing applications which are described in the following chapters. In Chapter 2, 2D complex mel-ceptrum features [28] are introduced to be

utilized for the image quality assessment problem. The proposed scheme makes use of image phase information to improve the representation capability of the 2D mel-ceptrum features which were previously used for face recognition [29, 30] and image quality assessment [31, 32] applications.

In Chapter 3, contrast enhancement of microscopy images are performed using a phase-based transform domain approach [33]. The proposed framework is inspired by the design of the phase contrast microscopy developed by F. Zernike [34]. The proposed enhancement technique tries to convert the variations in phase to changes in amplitude in order to enhance the contrast of cell structures. The observations and experimental evaluations show that the proposed phase-based enhancement scheme provides significant increase in contrast of microscopy images while suppressing the background pixels.

In the fourth chapter, an automatic target tracking scheme based on image phase information is introduced. The main aim of the proposed technique is to reduce the computational complexity of the cross-correlation based matching schemes without sacrificing the tracking accuracy. Starting from the traditional normalized cross-correlation computation, the tracking solution is simplified to a phase minimization problem. The proposed tracking framework makes use of truth-tables for phase variations corresponding to translational shifts occurring in the target model. In this way, a computationally efficient tracking technique is obtained.

Conclusions are made and the contributions are stated in the last chapter of the thesis.

Chapter 2

2D Complex Mel-Cepstrum Based Image Quality Assessment

In this chapter, a full-reference perceptual visual quality metric (PVQM) which contains feature detection and feature pooling operations is proposed. A new set of features, the 2D complex mel-cepstrum, for feature extraction is introduced. The proposed two dimensional (2D) complex mel-cepstrum features contain image phase information which is obtained using the complex logarithm of the Fourier transform in mel-scale. Feature pooling is carried out using support vector regression (SVR) which is a machine learning framework. The rest of the chapter is organized as follows. In Section 2.2, the 2D mel-cepstrum extraction procedure is reviewed. The proposed 2D mel-complex cepstrum feature extraction scheme [28] is described in Section 2.3. The SVR-based feature pooling regime is presented in Section 2.4. Experimental studies are presented in Section 2.5.

2.1 Related Work On Image Quality Assessment

Visual quality assessment is an important process that plays a key role in the modeling, implementation, optimization and testing of image/video processing algorithms and multimedia services. Despite the recent advances in coding technologies, the transmitted data suffer from both lossy source encoding and the losses induced by transmission channels [35]. This results in a degradation of quality in the images and videos. Therefore it is crucial to establish a criteria such that the perceived quality can be measured. Since it is considered that "quality" is generally a subjective term, given that the number of subjects is sufficiently large, it can be stated that the most accurate and reliable tool in the assessment of visual quality is the subjective evaluation [36]. However, the subjective evaluation is impractical, tedious, and unsuitable to be used in real-time applications. Moreover, the environmental conditions of the subjects and the contents of the images [37] directly affect the evaluation process. Therefore, it is important to develop an objective quality measure and significant amount of research has been carried out in order to obtain such a measure [38]. The visual quality metrics can be divided into two main categories; "signal fidelity measures" and "PVQM"s.

The signal fidelity measures including mean absolute error (MAE), mean square error (MSE), signal-to-noise ratio (SNR), peak SNR (PSNR) are simple and well-defined measures [39]. However, these measures perform poorly when the noise is not additive [40]. Although these measures respond appreciably to the changes in picture quality, it has been acknowledged by Girod that the signal fidelity measures do not exactly correspond to the human visual system (HVS) [41]. The inconsistency between the MSE-based measures and the perceived quality was also illustrated by Dosselmann and Young through the equal MSE hypersphere concept [42].

In spite of the advances in the objective image quality assessment problem [43,

44] a compact solution seems difficult to achieve due to the complex, multi-disciplinary nature of the problem. The PVQMs can be defined as objective models which try to predict subjective visual quality scores. The PVQM models have been studied in two main categories: vision-based [45] and signal-driven [35] approaches. In the first category, the HVS is intended to be modeled using the psychophysical properties and physiological knowledge. Due to the special structure of the HVS system, it processes the achromatic and chromatic signals separately, possesses special cells for distinctive orientations and has separate pathways for speed-varying objects. Therefore, most of the vision-based models make use of the color (the opponent color space [46], CIELAB space [47], YC_bC_r space [48] and YOZ space [49]), spatial (Gaussian pyramid [50], steerable pyramid filters [51]) and temporal [52] signal decomposition of the images.

Another common technique used to evaluate the visual quality is the detection of features/attributes and artifacts. Contrast is one of the most important attributes [53–55], since the HVS perceives the signal contrast rather than the signal strength. Local luminance change [48], block-based DCT [49] and motion-based [56] approaches have been proposed for the evaluation of image contrast. Another important artifact occurring in the image, especially when utilizing block-based DCT coding, is "blockiness". The blockiness effect is significant under low bit-rate conditions due to the usage of different quantization sizes in neighboring DCT blocks [57, 58]. Another blockiness measure which does not take into consideration the any other artifacts is proposed for video sequences [59]. Blurring is one of the most common artifact occurring around the edges in images. The blurring effect can be easily determined under the availability of a reference image [48] as well as in the case when the reference image is not available [60, 61].

In order to model the HVS, the just noticable distortion (JND) is proposed which states that not every change results in a significant variation by means of visibility. The JND term refers to a certain visibility threshold below which a small change can not be perceived by the majority (75%) of the subjects [62]. The JND can also be determined in other frequency domains including Laplacian pyramid decomposition [63] and discrete wavelet transform (DWT) [64]. In an image, not every difference receives the same amount of attention level since HVS

selects the most important/informative portion of the incoming signal. The visual attention (VA) approaches try to model the awareness/responsiveness of the HVS to the changes occurring in the different parts of an image [65, 66].

The PVQMs can be grouped into additional three sub-categories based on their dependency on the reference information, namely full-reference, reduced-reference, and no-reference PVQMs. The PVQMs which use some part of the reference signal instead of using it as a whole are called reduced-reference metrics [67]. If the whole reference signal is used in the evaluation of the PVQM, then this PVQM is called a full-reference metric [68]. The no-reference metrics only use the test signal and therefore they are also called as no-reference metrics [57, 69].

Developing a PVQM using a signal-driven approach has recently become more popular since vision-based modeling is computationally more expensive. Also, signal-driven methods do not try to develop an extensive HVS model. Instead, they try to obtain a quality measure using the signal itself and extracted features/characteristics that lie in different domains. However, a metric evaluated using a signal-driven method may also take into account the psychophysical aspects of the HVS.

The structural similarity index (SSIM) developed by Wang et. al. is one of the most common full-reference metrics available in the literature [70]. The method tries to model the structural losses in a given image. The multiscale extension of SSIM, namely MS-SSIM [71], achieves better performance than ordinary SSIM. As the first version of the SSIM, the Universal Quality Index (UQI) [72] had been proposed by Wang and Bovik to model the loss of quality in terms of "loss of correlation", "luminance distortion" and "contrast distortion" components. Information theory has also been applied to image quality assessment and the visual information fidelity (VIF) [73] metric that tries to quantify the information shared by the reference and distorted images has been proposed by Sheikh and Bovik. Venkata et. al. proposed the noise quality measure (NQM) [74] full-reference framework that tries to model the degradations in the distorted images using reference images. A multi stage metric, visual SNR (VSNR) [75] is proposed by Chandler and Hemami for image quality assessment. In the first stage of

VSNR computation, the visibility of the distortion in the image is examined with a contrast thresholding based scheme. In the second stage, perceived contrast and visual property of global precedence based framework is applied when the distortion in the image is above the threshold.

Feature extraction based PVQMs are developed in order to model the degradations in an image. Local variance and correlation-based structural schemes, which are important milestones in image quality assessment [70], can also be considered as a feature extraction based method. Moreover, Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT)-based transform domain features are also used to measure the quality of the images [76]. Promising results obtained by DCT and DWT-based frameworks encourage researchers to develop quality measures that work in the transform domain [77–79]. As another feature extraction scheme, Shnayderman et. al. used SVD to define a quality measure for the images degraded by different levels and different types of distortions [80]. The method computes the distance between the eigenvalues of image sub-regions on the reference and distorted images. The distances calculated at different subregions are accumulated in order to obtain a quality score. In another work, Narwaria and Lin used the SVD-based features together with a machine learning framework [81].

After the features are extracted from the images, a final objective quality score based on the selected features has to be determined. The process of selecting or integrating features to obtain a quality score is generally called "feature pooling", which is an important concept in the PVQM framework. Feature pooling is performed by using simple summation [82], weighted combination [83] and other types of pooling regimes [46] in the literature. Recently, regression-based machine learning techniques have been used for feature pooling. However, machine learning frameworks were used to define quality measures [84–86] in the literature. Due to their generalization capabilities and the ability to handle high dimensional data, they have potential to provide better solutions for PVQM.

Recently, Narwaria and Lin have used regression-based machine learning techniques for feature pooling [81]. However, the utilization of machine learning theory in the image quality assessment problem has been explored by researchers who have proposed machine learning approaches to define quality measures [84, 86]. Due to their generalization capabilities and ability to handle high dimensional data, machine learning based techniques have the potential to provide better solutions for PVQM. A full-reference PVQM evaluation composed of feature detection and pooling stages is proposed by Narwaria and Lin [87]. In the feature detection stage, SVDs of the original and perturbed images are calculated. After detection of the features, feature pooling is carried out using the SVR. The main aim is to evaluate the deviations present in the singular vectors of the original and perturbed images.

Recently, convolutional neural networks (CNNs) have become very popular amongst researchers due to their great success in image classification and recognition [88]. In the presence of a large dataset, CNN-based frameworks achieve vast performance improvements compared to conventional methods [89]. One of the most important advantages of the CNN framework is its applicability to raw images and ability to form features through the training process. In this way, the CNN automatically generates descriptive features instead of using "hand-crafted" features. By making use of the deep structures, complex mapping functions can be generated easily by the CNN without taking into consideration specific domain knowledge. Due to their aforementioned advantages, researchers have been trying to utilize deep learning structures for the image quality prediction problem [90]. However, the major obstacle in utilizing CNNs in the IQA problem is that the IQA image datasets contain very few samples. Compared to the CIFAR-100 [91](80 million labelled samples) and ImageNet [92](50 million labelled samples) datasets, the IQA image datasets contains at most several thousands of labeled data. Since the performance of deep learning frameworks rely highly on the size of the training samples, a deep network can easily overfit on the samples of a IQA image dataset.

In order to overcome the insufficient training samples problem, several strategies have been implemented by the researchers. One strategy is to use the data

augmentation techniques to diversify the images by applying various rotation, cropping, and mirroring operations. However, applying data augmentation techniques to the image result in significant changes on the perceived quality of the image. Therefore, one can not directly assign the score of original image to the images which are formed after data augmentation techniques. Another strategy to cope with the small sample size problem is to divide each image into small patches to increase the number of training samples. Since the observers determine a single quality score in the subjective evaluations for the whole image, the local ground-truth quality scores for the image patches are not available. Due to the non-stationary nature of the distortions and the quality perceived by the observers, one can not simply assign the local image patches to the same score which was previously obtained for the whole image.

The researchers have also utilized unsupervised learning frameworks such as restricted Boltzman machine (RBM) and autoencoders [93] with deep structures to deal with the insufficient training sample set problem. By making use of the unsupervised systems, the deep learning frameworks can be trained on large datasets without requiring the groundtruth information. Due to their design, the unsupervised frameworks extract structural shape-based features while suppressing the small details. Conversely, IQA frameworks directly work on the small differences between the reference and distorted images to find a mapping function between the level of distortion and quality score. Yet, a generalizable and successful solution for the utilization of the deep unsupervised learning frameworks for the IQA problem have not been reported. By taking into account the aforementioned problems, researchers have proposed deep learning based no-reference and full reference IQA schemes to compete with state-of-the-art IQA techniques.

The first application of the deep learning structures to the no-reference IQA problem is the utilization of a shallow CNN model to generate features [94]. The authors divide the input image into small patches to increase the data size. Each image patch is assigned with the same quality score in the training phase. In another work, a pre-trained deep CNN framework on ImageNet dataset [95] is used for the IQA problem. The last layer of the pre-trained model is replaced by a regression layer which performs mapping between the features and subjective

quality scores. Similar to the work [94], the image is divided into patches and each patch is assigned with the same quality score [95]. A deeper CNN-based learning framework is also utilized for the no-reference IQA problem [96]. The patchwise training (as in [94, 95]) and weighted average patch aggression strategies are proposed for training the deep CNN structure. In order to cope with the insufficient training sample set problem, a two stage deep CNN framework is proposed [97]. The first stage of the algorithm is a pre-training stage which generates proxy quality scores by making use of conventional full-reference IQA metrics. In the second stage, the feature vectors corresponding to each image patch are aggregated and then a regression scheme is performed to create mappings for the subjective scores [97].

The deep learning based structures have been also proposed for full-reference IQA problems. A dual path CNN framework is used for quality prediction of images using non-aligned images of similar scenes as the reference [98]. Normalized patches of reference and degraded images are fed into a dual path CNN framework which shares the same parameters. Then future vectors are concatenated before being exposed to the regression process. A pre-trained deep CNN framework on ImageNet is also utilized for the full-reference IQA problem [99]. The reference and distorted image pairs are fed into the CNN whose output layers are used as feature maps. A pooling strategy is carried out to pool the local similarities between the reference and distorted feature maps to obtain quality scores. A deep CNN-based framework is used to learn the visual sensitivity weights by making use of the spatial characteristics in local regions of the distorted image [100]. The distorted image and corresponding objective error map are fed into the deep CNN framework to produce weight maps. The extracted weight maps are used as multiplication factors for the objective error maps.

Although the deep learning frameworks obtain promising results on the IQA problem, insufficient training samples remain a major obstacle to achieve generalizable and competitive models for IQA. Still, shallow mapping engines such as support vector regressors (SVR) provide high performance on legacy datasets. However, the deep learning frameworks have the potential to set the performance standards to higher levels once a reliable large-scale IQA dataset is created.

As a good example of shallow mapping engines, Narwaria et. al. proposed another PVQM framework [31] which is similar to the one proposed in their earlier study [87]. The authors used 2D mel-cepstrum features instead of SVD in the feature extraction stage of the PVQM, [31]. The 2D mel-cepstrum features, which originate from the cepstral analysis have been used for image representation by Cakir and Cetin [29, 30]. Cepstral feature extraction techniques have drawn attention due to their representative power in speech processing applications including speech and sound recognition and speaker identification [101–104]. In the literature, the 2D cepstrum, which is the natural extension of 1D cepstrum, was used for shadow detection, echo removal, automatic intensity control, enhancement of repetitive features and cepstral filtering [105–107]. The 2D mel-cepstrum based image quality assessment (IQA) scheme proposed by Narwaria et. al. was shown to outperform the baseline techniques [31], however, the method does not consider the image phase information that may be a significant component in the quality assessment problem.

The Phase of the FT is extremely important in image representation [7,9]. Due to its significance in image representation, FT phase is often utilized in various image processing applications including image quality assessment [32,108]. These techniques try to integrate the image phase information with the magnitude by using a rule-based scheme.

2.2 2D Mel-cepstrum Feature Extraction

2D Mel-cepstrum has originated from the ordinary 2D cepstrum of a 2D signal. 2D cepstrum $\Psi(a, b)$ of a 2D image $y(a, b)$ can be defined as follows:

$$\Psi(a, b) = F_2^{-1}(log(|Y(u, v)|^2)) \tag{2.1}$$

where (a, b) denotes 2D cepstral "quefrency" coordinates, F_2^{-1} denotes 2D Inverse Discrete-Time Fourier Transform (IDTFT) and $Y(u, v)$ is the 2D Discrete-Time Fourier Transform (DTFT) of the image $y(a, b)$. In general, the Fast Fourier Transform (FFT) algorithm is preferred to compute the DTFT.

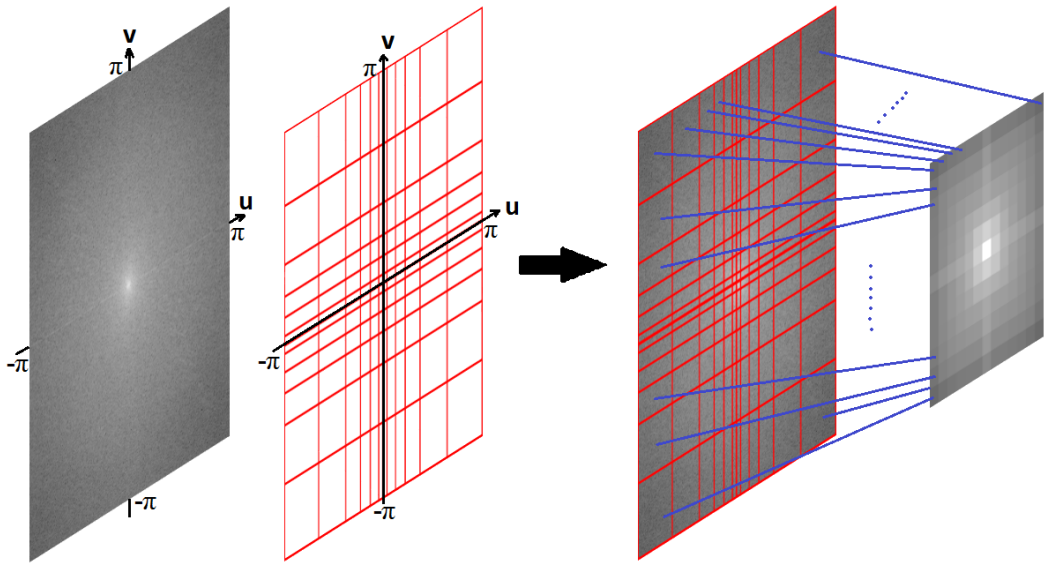


Figure 2.1: A representative non-uniform 2D mel-cepstrum grid in the DTFT domain. Cell sizes are smaller at low frequencies compared to high frequencies. The resultant matrix has smaller dimensions than the original spectrum since each frequency bin produces only one coefficient.

The 2D mel-cepstrum divides the DTFT domain into different sized regions by making use of the structural nature of the Fourier Transform. In other words, the DTFT domain is divided into non-uniform bins in a logarithmic manner as shown in Figure 2.1 and the energy $|G(m,n)|^2$ of each bin is computed as in Equation 2.2.

$$|G(m,n)|^2 = \sum_{u,v \in B(m,n)} |Y(u,v)|^2 \tag{2.2}$$

Here, $Y(u,v)$ is the Discrete Fourier Transform (DFT) of $y(a,b)$ and $B(m,n)$ is the cell of the logarithmic grid corresponding to the $(m,n)^{th}$ location. The bin sizes are smaller at low frequencies compared to high-frequencies.

The concept of dividing the transform domain into different sized regions is similar to the mel-cepstrum computation in speech processing. Similar to speech signals, most natural images are low-pass in nature. Therefore, there is more signal energy at low-frequencies compared to high frequencies. Logarithmic division of the DFT grid in such a manner emphasizes high frequencies. After DFT gridding, 2D mel-frequency cepstral coefficients $\Psi^{mel}(p,q)$ are computed using inverse

DFT as in Equation 2.3.

$$\Psi^{mel}(p,q) = F_2^{-1}(log(|G(m,n)|^2)) \tag{2.3}$$

Due to the logarithmic grid shown in Figure 2.1, the size of the forward DFT used to compute $Y(u,v)$ is larger than the size of the inverse DFT (IDFT). Similar to the use of mel-cepstrum in speech processing, to emphasize certain bands, different weights may also be assigned to different frequency bins. The dimensions of the 2D mel-cepstrum sequence, which is computed using the IDFT, are smaller than the original image because several DFT values are grouped together in each bin. The steps of the 2D mel-cepstrum based feature extraction scheme are summarized as follows (note that the image is selected to be a square image for ease of representation):

- $N \times N$ 2D DFT of input images are calculated. Before the DFT calculation, the images are padded with zeros up to the size that is the closest power of 2 in order to take advantage of the FFT algorithm during the DFT computation.
- The non-uniform DTFT grid is applied to the resultant DFT matrix and the energy $|G(m,n)|^2$ of each cell is computed. Each cell of the grid can also be weighted with a coefficient. The new data size is M by M where $M \leq N$
- Logarithm of cell energies $|G(m,n)|^2$ are computed.
- 2D IDFT of the M by M data is computed to get the M by M mel-cepstrum sequence.

In an image, the contributions to the high frequencies are generally made by edges and structural features. For the purpose of extracting better representative features, the cells of the 2D DFT grid, which correspond to the high-frequency components, are multiplied with higher weights compared to those used for the low-frequency cells in the grid. This ensures that high-frequency components are further emphasized.

An important feature is the invariance of the cepstrum to pixel amplitude changes. In this manner, robustness to illumination invariance can be achieved. Let $Y(u,v)$ denote the 2D DTFT of a given image $y(a,b)$ and $y_2(a,b) = \alpha y(a,b)$ has a DTFT $\alpha Y(u,v)$ for any real constant α. The log spectrum of $\alpha Y(u,v)$ is given as follows

$$log(|\alpha Y(u,v)|) = log(|\alpha|) + log(|Y(u,v)|) \tag{2.4}$$

and the corresponding cepstrum is given as follows

$$\Psi_2(a,b) = \kappa\delta(a,b) + \Psi(a,b) \tag{2.5}$$

where $\delta(a,b) = 1$ for $a = b = 0$ and $\delta(a,b) = 0$ otherwise. Therefore, the cepstrum values except at the $(0,0)$ location (DC term) do not vary with amplitude changes. Since the FT magnitudes of $y(a,b)$ and $y(a-c, b-d)$ are the same, the 2D cepstrum and mel-cepstrum are shift invariant features.

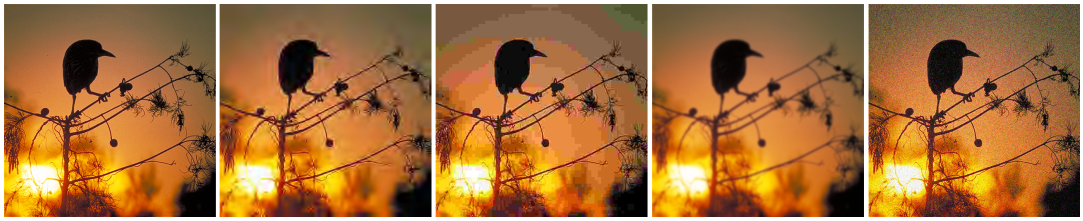
Another important characteristic of the 2D cepstrum is symmetry with respect to $\Psi(a,b) = \Psi(-a,-b)$. As a result, only a half of the 2D cepstrum or MxM 2-D mel-cepstrum coefficients are enough when IDFT is used. The 2D mel-cepstrum method, which originated from the ordinary 2D cepstrum, also has the same properties.

A more detailed discussion and derivations for 2D mel-cepstrum can be found in our earlier studies [29, 30]. Due to its representative properties and the success in the face recognition problem [29], the 2D mel-cepstrum feature extraction scheme were proposed for image quality assessment problem and achieved promising results [31].

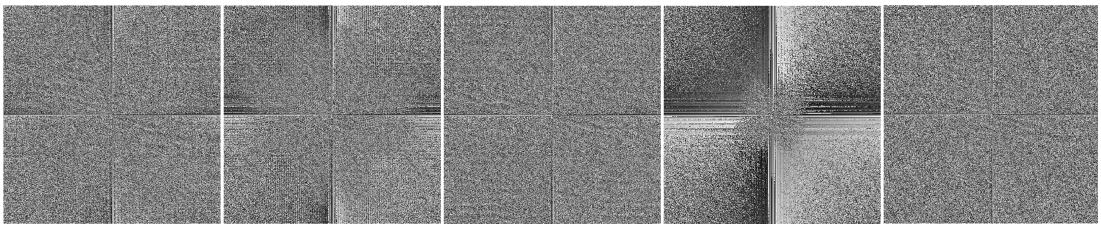
2.3 2D Complex Mel-Cepstrum based Feature Extraction

Although the 2D mel-cepstrum feature extraction scheme has achieved promising results for image representation [29–31], the method completely ignores the image

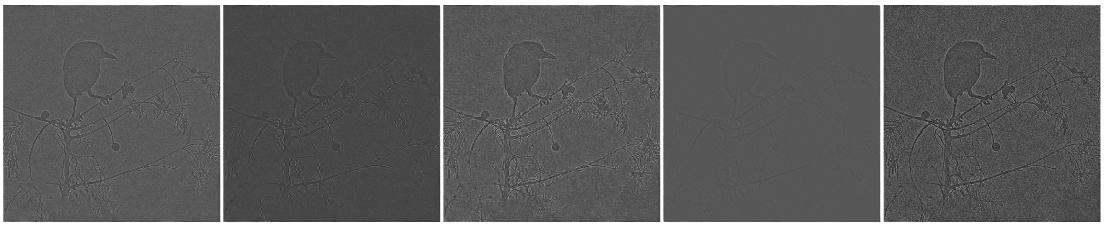
phase information that may possess superior representation capability. The FT phase information of a set of images and the synthesis results obtained by using the approach proposed by Oppenheim et. al. for these images, are presented in Figure 2.2.



(a) Reference Image, JPEG2000 compressed image, JPEG compressed image, Gaussian blurred image, and the noisy image (Image Size: 512×512).



(b) FT Phase Information of reference, JPEG2000 compressed, JPEG compressed, Gaussian blurred, and the noisy images.



(c) Image synthesis using only image phase information. Magnitude of the FT is set as unit matrix.

Figure 2.2: The illustration of FT phase information and phase-only synthesis of reference and distorted images.

By looking at the synthesis results presented in Figure 2.2, it can be observed that the image phase information contains descriptive high frequency characteristics, since the synthesis image preserves most of the edges and texture details of the input image. Moreover, high frequency characteristics of the image are still present in the image phase information after exposure to different types of

distortions. The illustrations of FT phase matrices in Figure 2.2 also show that different types of distortions cause the FT phase information to distribute in different manners in the spectrum. Since we deal with different types and different levels of distortion in image quality assessment problem, the FT phase information which disperses differently in the spectrum under different degradation types, may be a useful feature for signal representation. In addition, frequency binning using a non-uniform grid in the 2D mel-cepstrum computation can be applied to FT phase information to extract phase characteristics in local frequency bands. Therefore, due to its possible contribution to signal representation and its discriminative distribution in the spectrum, our aim is to integrate FT phase information with the 2D mel-cepstrum features in order to use it in a feature-based image quality assessment scheme. Here, we try to seek the possibility of deriving 2D complex mel-cepstrum features-based on complex cepstrum theory [7].

One dimensional (1D) complex cepstrum was defined by Oppenheim and his coworkers [7, 109, 110]. The natural 2D extension of the complex cepstrum was also proposed to be used in image processing applications such as echo removal, automatic intensity control and enhancement of repetitive features [106]. Recalling that $Y(u, v)$ is the DTFT of the image $y(a, b)$, $Y(u, v)$ may contain artifacts commonly known as the "cross-effect" due to the implicit assumption about the image's periodicity. However, natural images contain significant discontinuities especially across the frame borders that may cause cross-shape artifacts in the Fourier spectrum. Therefore, before further processing, the image is exposed to a "periodization" step [111] to eliminate the artifacts that degrade the spectrum. The periodization algorithm [111] used herein decomposes the input image into two components, one of which is a periodic image while the other is a smooth component except for sharp transitions near the frame borders. The main idea behind the algorithm is finding a linear decomposition which provides a periodic component whose spectrum is close to the spectrum of the input image as well as being artifact-free. However, the latter component captures all the edge artifacts arising from the implicit periodization of the input image. When the input image is directly padded with zeros in order to produce an image of size power of twos, this padding operation may cause undesired artifacts in the image. However, prior

periodization eliminates the undesired effects caused by direct padding. $Y_p(u,v)$ is calculated as the DTFT of the periodized image $y_p(a,b)$ and used in the 2D complex cepstrum $\Psi^{\mathbb{C}}(a,b)$ computation as follows:

$$\begin{aligned}\Psi^{\mathbb{C}}(a,b) &= F_2^{-1}\left\{\log\left(Y_p(u,v)\right)\right\} \\ &= F_2^{-1}\left\{\log\left(\left|Y_p(u,v)\right| e^{j\phi(u,v)}\right)\right\} \\ &= F_2^{-1}\left\{\log\left(\left|Y_p(u,v)\right|\right) + \log\left(e^{j\phi(u,v)}\right)\right\} \\ &= F_2^{-1}\left\{\log\left(\left|Y_p(u,v)\right|\right)\right\} + F_2^{-1}\left\{j\phi(u,v)\right\}\end{aligned} \tag{2.6}$$

In the final calculation step of Equation 2.6, we have the real 2D cepstrum of the image on the first summand, and the inverse Fourier transform of the image phase on the second summand. The right summand is due to the complex logarithm operation which has to be continuous. Therefore the discontinuities in phase should be unwrapped to compute the complex cepstrum correctly. In the literature, several phase unwrapping algorithms have been developed to calculate the complex cepstrum in an appropriate manner [110, 112–117]. The phase unwrapping is carried out by adding multiples of 2π to appropriate phase locations for a smooth phase function. The phase unwrapping algorithm used in this work is based on the local phase gradient. Each element of the phase is checked by whether the sign of the phase gradient is consistent with its consecutive neighbors. If the sign of the phase gradient changes, the phase element is modified by adding/subtracting integer multiples of 2π. Otherwise, the phase element remains unchanged. This way, undesired phase transitions are eliminated and the phase is modified to be smooth.

In this chapter, the 2D complex mel-cepstrum is defined for feature extraction from images. The 2D mel-cepstrum is known to be an efficient feature extraction technique in face recognition [29] and image quality assessment [31] problems. The 2D complex mel-cepstrum is computed similar to the 2D mel-cepstrum by combining discrete Fourier domain coefficients in logarithmic bins. However, both magnitude and phase terms are averaged in the 2D complex mel-cepstrum. As in regular 2D complex cepstrum, phase terms in a given logarithmic bin must be unwrapped before averaging. Let $\tilde{\phi}(u,v)$ be the unwrapped phase of the image FT. In each logarithmic bin, unwrapped phase coefficients are averaged as follows:

$$\theta\left(m,n\right)=\frac{1}{\mu_{m,n}}\sum_{u,v\in B(m,n)}\tilde{\phi}(u,v) \tag{2.7}$$

where $\theta\left(m,n\right)$ is the mean phase value of the $(m,n)^{th}$ bin and $\mu_{m,n}$ is the number of 2D DFT coefficients in the corresponding bin $B(m,n)$. By using the logarithmic bin energy values and phase information defined in Equations 2.2 and 2.7, the proposed 2D complex mel-cepstrum coefficients ($\Psi^{\mathbb{C}mel}(p,q)$) are calculated as follows:

$$\begin{aligned}\Psi^{\mathbb{C}mel}(p,q) &= F_2^{-1}\left\{\log\left(\left|G(m,n)\right|e^{j\theta(m,n)}\right)\right\}\\ &= F_2^{-1}\left\{\log\left(\left|G(m,n)\right|\right)+\log\left(e^{j\theta(m,n)}\right)\right\}\\ &= F_2^{-1}\left\{\log\left(\left|G(m,n)\right|\right)\right\}+F_2^{-1}\left\{j\theta(m,n)\right\}\\ \Psi^{\mathbb{C}mel}(p,q) &= W\cdot\Psi^{mel}(p,q)+F_2^{-1}\left\{j\theta(m,n)\right\}\end{aligned} \tag{2.8}$$

The 2D complex mel-cepstrum computation integrates the $\Psi^{mel}(p,q)$ with the inverse Fourier transform of the unwrapped phase information as shown in Equation 2.8. In order to establish an appropriate balance between the energy and phase terms in the computation (Equation 2.8), $\Psi^{mel}(p,q)$ should be calculated by using the appropriate weight. The weight W used in $\Psi^{mel}(p,q)$ computation is calculated as follows:

$$\begin{aligned}W &= \arg\min_{W^*}\left\{\left|\bar{\theta}-W^*\cdot\bar{\Psi}^{mel}\right|\right\}\\ \bar{\theta} &= \tfrac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\theta(i,j)\\ \bar{\Psi}^{mel} &= \tfrac{1}{N^2}\sum_{p=1}^{N}\sum_{q=1}^{N}\Psi^{mel}(p,q)\end{aligned} \tag{2.9}$$

In this way, the domination of either the energy or phase components is avoided in the overall 2D complex mel-cepstrum computation.

The flow diagram of the proposed 2D complex mel-cepstrum feature extraction scheme [28] is presented in Figure 2.3. The implicit periodization is carried out as a preprocessing step before the computation of cepstral coefficients.

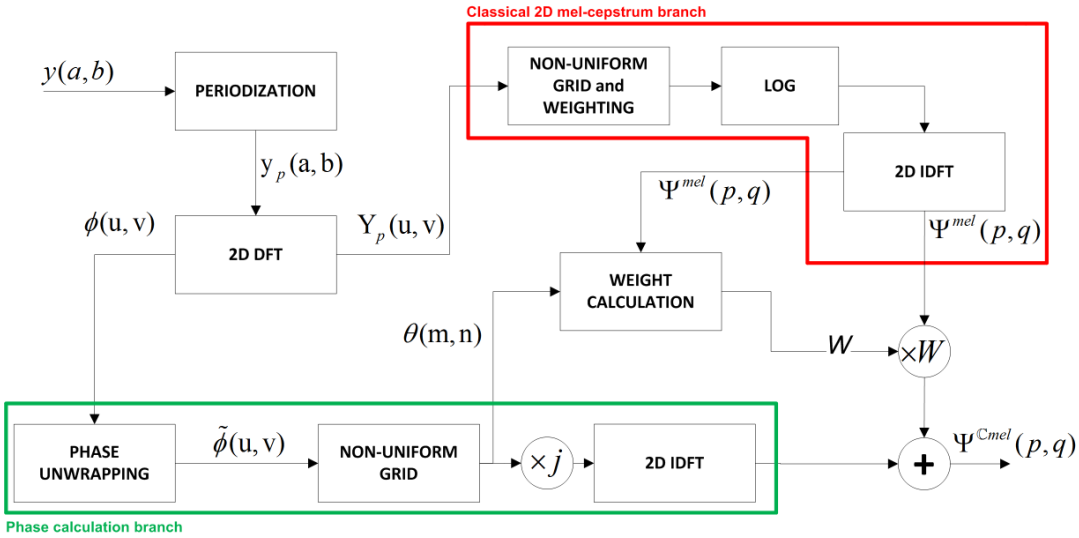


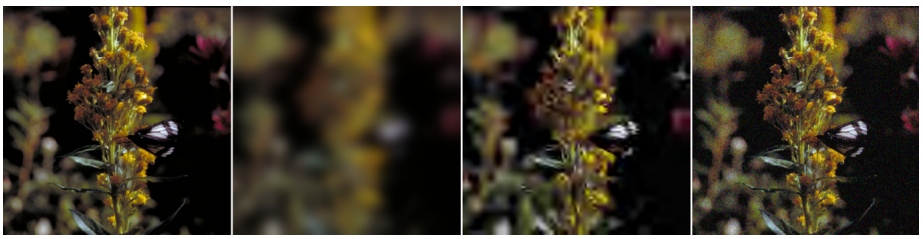
Figure 2.3: Flow diagram of the proposed 2D complex mel-cepstrum feature extraction scheme.

By looking at the flow diagram presented in Figure 2.3, one can say that the proposed 2D complex mel-cepstrum scheme introduces a new phase branch (lower part of Figure 2.3) in the calculations. The classical 2D mel-cepstrum computation ($\Psi^{mel}(p,q)$) is carried out in the right branch of Figure 2.3. At the end, the 2D complex mel-cepstrum features are obtained by the integration of the classical 2D mel-cepstrum with the phase information. Similar to the classical mel-cepstrum computation, the proposed technique also provides a reduced feature representation by making use of non-uniform binning in both magnitude and phase components.

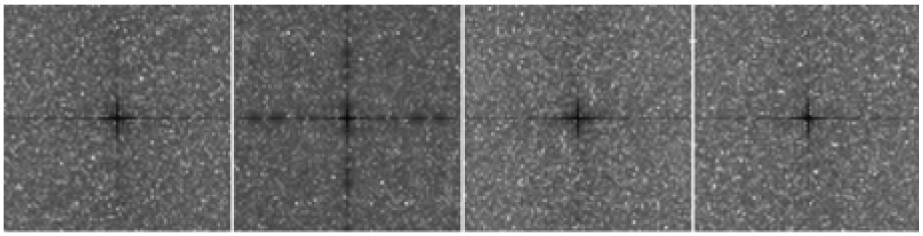
Recall that the 2D cepstrum is symmetric with respect to $\Psi(a,b) = \Psi(-a,-b)$ and enables a further dimensionality reduction in the feature space. This is caused by the conjugate-symmetric structure of the FT coefficients and it is clear that the magnitude of the complex FT pairs are the same. But, the proposed 2D complex mel-cepstrum features do not provide dimensionality reduction since the phase information is not symmetric with respect to the DC term, i.e. $(0,0)$ frequency location. This is due to the phase differences between the conjugate-symmetric FT pairs and the nature of the phase unwrapping algorithm. The main aim of the phase unwrapping algorithm is eliminating the local discontinuities in the

image phase. While eliminating these discontinuities, it does not try to preserve the overall symmetry since it may damage the phase information.

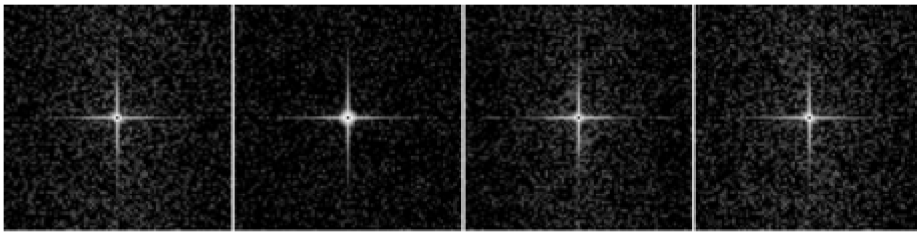
In Figure 2.4, examples of 2D complex mel-cepstrum coefficient matrices extracted from both reference and distorted images are presented. The images are taken from CSIQ image dataset [118] and the magnitude of the 2D complex mel-cepstrum matrices are calculated and then saturated for visualization since the proposed features are complex valued matrices.



(a) Reference Image, Gaussian blurred image, JPEG2000 compressed image and the noisy image (Image Size: 512×512).



(b) 2D mel-cepstrum coefficients corresponding to reference, Gaussian blurred, JPEG2000 compressed, and the noisy images (Feature Size: 85×85).



(c) 2D complex mel-cepstrum coefficients corresponding to reference, Gaussian blurred, JPEG2000 compressed, and the noisy images (Feature Size: 85×85).

Figure 2.4: Example of 2D mel-cepstrum and 2D complex mel-cepstrum features extracted from reference and distorted images.

2.4 Feature Pooling with Support Vector Regression

In order to calculate objective quality scores from the cepstral feature space, a regression framework is carried out to perform mapping between the high dimensional feature space and already available subjective scores. The regression framework used in this study is support vector regression (SVR), namely ϵ-SVR. Instead of re-implementing the SVR algorithm, the LIBSVM package [119] which provides an effective implementation and easy-to-use software is used. The SVR accepts the features in terms of vectors, and therefore, the cepstral feature matrices ($\Psi^{mel}(p,q)$, $\Psi^{\mathbb{C}mel}(p,q)$) are converted into vectors before the SVR algorithm. Let ϑ_R and ϑ_D be the cepstral feature vectors extracted from the reference and distorted images, respectively. The image degradations, i.e distortions, are modeled by defining the feature difference vector as in Equation 2.10.

$$x = |\vartheta_R - \vartheta_D| \tag{2.10}$$

More specifically, let $x_i,\ i = 1, 2, \ldots, n_{tr}$ denote the difference feature vector corresponding to the i^{th} image in the training dataset where n_{tr} denotes the number of images used for training. In ϵ-SVR, the main aim is to find a function f that maps the feature vectors to subjective quality evaluation scores (s_i) while pushing the standard deviation to a value smaller than ϵ. While satisfying these requirements, the function $f(x_i)$ should also be as flat as possible [120]. The function to be learned is $f(\mathrm{x}) = \beta^T\varphi(\mathrm{x}) + \gamma$ where $\varphi(\mathrm{x})$ is a nonlinear function of x, β is weight vector and γ is a bias term. The weight vector β is calculated as provided in Equation 2.11.

$$\begin{gathered} |s_i - f(x_i)| \leq \varepsilon \\ \beta = \sum_{i=1}^{n_{sv}} (\eta_i{}^* - \eta_i)\, \varphi(x_i) \end{gathered} \tag{2.11}$$

Here, n_{sv} denotes the number of support vectors obtained in the regression

process and $\eta_i{}^*$, η_i are the Lagrange multipliers used in the function optimization. If the optimization problem is solved, the function to be learned becomes as given in Equation 2.12.

$$\begin{aligned} f(x) &= \beta^T \varphi(x) + \gamma \\ &= \sum_{i=1}^{n_{sv}} (\eta_i{}^* - \eta_i)\, \varphi(x_i)^T \varphi(x) + \gamma \\ &= \sum_{i=1}^{n_{sv}} (\eta_i{}^* - \eta_i)\, K(x_i, x) + \gamma \\ &= \sum_{i=1}^{n_{sv}} (\eta_i{}^* - \eta_i) \exp\left(-\rho \|x_i - x\|^2\right) + \gamma \end{aligned} \tag{2.12}$$

Note that, the kernel function $K(x_i, x)$ is selected as the radial basis function (RBF) kernel.

After the function is learned, the quality score Q is obtained by $Q = f(\mathrm{x})$. In this way, the SVR maps the high dimensional image data to a single quality score.

2.5 Experimental Studies

In this thesis, the proposed technique is compared with baseline techniques over eight different image quality databases. Each feature extraction scheme is followed by a feature pooling phase that is carried out by using SVR. The quality scores obtained for each feature extraction scheme and image quality database are compared with three correlation-based measures. In the following subsections, the experimental procedures are explained in detail.

2.5.1 Image Quality Databases

In the performance evaluation of different IQA schemes, eight different image databases (A57 [121], TID2008 [122], IRCCyN/IVC [123], LIVE [124],

CSIQ [118], Toyoma [125], VCL@FER [126] and WIQ [127]) are used. Basic properties of each dataset, namely; number of reference images, number of distorted images, number of distortion types, number of distortion levels, image properties and information about subjective evaluation are listed in Table 2.1. Some of the attributes in Table 2.1 are stated as "Varying" because these attributes may change for different reference images.

2.5.2 Experimental Procedures and Results

In the experiments, a k-fold cross validation procedure is followed and the image databases used in the experiments are divided into equally-sized partitions. These partitions are formed such that none of the partitions have distorted images obtained using the same reference image in order to prevent testing with the images used in the training phase. In k-fold cross validation, the overall dataset is divided into k partitions. At each experimental step, the partition used in the test phase is changed and the remaining $k-1$ partitions are used in the training phase of the SVR. The experiments are repeated until every single partition is used in the test phase. At the end of the experiments the results obtained for each repetition are used to calculate the overall average of the results. The A57, TID2008, IRCCyN/IVC, LIVE, CSIQ, Toyama, VCL@FER and WIQ image databases are divided into 3, 5, 10, 6, 6, 7, 5, and 7 partitions, respectively. For each image dataset, we tried to construct partitions which contain the same number of images, depending on the number of reference images present in the database.

The quality measures produced by the proposed and the baseline techniques are evaluated by measuring the similarity between these measures and the subjective scores. In order to measure this similarity, three commonly used criteria, namely; Pearson Linear Correlation Coefficient (C_P), Spearman's Rho Correlation Coefficient (C_S), and Kendall's Tau Correlation Coefficient (C_K) are used. The objective quality metric resulting in the highest values of these correlation

Table 2.1: Information about image quality datasets. PD: Perceived Distortion, MOS: Mean Opinion Score, DMOS: Difference Mean Opinion Score.

Database	# Reference Images	# Distortion Types	# Distortion Levels	# Distorted Images	Image Properties			Subjective Evaluation	
					Image Size	Color	Bit Depth	Score Type	Scale
A57	3	6	3	54	512×512	Gray Scale	8-bit	PD	[0,1]
TID2008	25	17	4	1700	512×384	RGB	24-bit	MOS	[0,9]
IRCCyN/IVC	10	4	Varying	185	512×512	RGB	24-bit	MOS	[1,5]
LIVE	29	5	Varying	779	768×512	RGB	24-bit	DMOS	[1,100]
CSIQ	30	6	4/5	866	512×512	RGB	24-bit	DMOS	[0,1]
Toyoma	14	2	6	168	768×512	RGB	24-bit	MOS	[1,5]
VCL@FER	23	4	6	552	Varying	RGB	24-bit	MOS	[1,100]
WIQ	7	1	Varying	80	512×512	Gray Scale	8-bit	DMOS	[1,100]

coefficients with the subjective scores is determined to be the best quality assessment metric. Also, to establish a mapping between the objective quality metric results and subjective scores, a 5-parameter logistic function [128] is used before the evaluation of correlation-based similarity criteria. The performance of the proposed scheme is compared with the baseline techniques consisting of PSNR, VSNR [75], VIF [73], SSIM [70], MS-SSIM [71], UQI [72], NQM [74], Independent Feature Similarity (IFS) [129], Visual Saliency based Index (VSI) [130] and 2D mel-cepstrum (2DMC) [29, 31]. The results obtained for each objective measure corresponding to each image dataset are presented in Table 2.2. In order to increase the readability of the results, similarity measures providing the 1^{st} and 2^{nd} best performance, are written in bold font.

In order to visualize the correlation values, C_p's corresponding to each objective metric and image dataset are illustrated in Figure 2.5.

Table 2.2: Performance of the objective quality metrics for different image quality databases.

Database	Crit.	Objective Quality Measures										
		PSNR	VSNR	VIF	SSIM	MS-SSIM	UQI	NQM	IFS	VSI	2DMC	Proposed
A57	C_P	0.7073	**0.9502**	0.6228	0.4257	0.8507	0.5584	0.8178	0.8995	0.9057	0.9242	**0.9383**
	C_S	0.6189	**0.9355**	0.6223	0.4143	0.8394	0.5398	0.7968	0.8617	0.9030	0.9016	**0.9312**
	C_K	0.4309	**0.8031**	0.4589	0.2854	0.6478	0.3936	0.5890	0.6842	0.7373	0.7490	**0.7883**
TID2008	C_P	0.5734	0.6823	0.8084	0.6413	0.8424	0.6473	0.6135	**0.8810**	**0.8762**	0.8226	0.8235
	C_S	0.5834	0.7015	0.7491	0.6272	0.8526	0.5851	0.6236	**0.8903**	**0.8979**	0.8026	0.8027
	C_K	0.4256	0.5323	0.5861	0.4562	0.6539	0.4255	0.4600	**0.7009**	**0.7123**	0.6122	0.6152
IRCCyN /IVC	C_P	0.6698	0.7306	0.8940	0.7885	0.8920	0.8255	0.7857	**0.9401**	0.9104	0.9294	**0.9481**
	C_S	0.6427	0.7347	0.8876	0.7750	0.8827	0.8196	0.7769	**0.9326**	0.8979	0.9174	**0.9270**
	C_K	0.4784	0.5469	0.7031	0.5887	0.6978	0.6202	0.5796	0.7695	0.7194	**0.7830**	**0.7955**
LIVE	C_P	0.9078	0.9427	**0.9666**	0.9424	0.9560	0.9177	0.8713	**0.9599**	0.9559	0.9509	0.9570
	C_S	0.9124	0.9465	**0.9695**	0.9479	0.9624	0.9136	0.8627	**0.9645**	0.9623	0.9501	0.9536
	C_K	0.7346	0.7989	**0.8477**	0.8006	0.8342	0.7495	0.6799	**0.8417**	0.8328	0.8148	0.8269
CSIQ	C_P	0.8000	0.7685	0.8996	0.7929	0.8937	0.8292	0.7773	**0.9576**	0.9279	0.9360	**0.9545**
	C_S	0.8033	0.7786	0.8935	0.7795	0.8847	0.8236	0.7719	**0.9582**	0.9423	0.9281	**0.9497**
	C_K	0.6022	0.5930	0.7101	0.5921	0.7005	0.6250	0.5745	**0.8165**	0.7857	0.7724	**0.8079**
Toyama	C_P	0.6440	0.8755	**0.9186**	0.7901	0.8919	0.7202	0.8746	0.8710	0.7131	0.9039	**0.9300**
	C_S	0.6153	0.8660	**0.9114**	0.7809	0.8872	0.7181	0.8721	0.8668	0.7015	0.8875	**0.9155**
	C_K	0.4476	0.6833	0.7383	0.5883	0.7044	0.5336	0.6831	0.6772	0.5159	**0.7422**	**0.7736**
VCL@ FER	C_P	0.8186	0.4186	0.6363	0.8419	0.8178	0.7902	**0.9396**	0.8394	0.6429	**0.9310**	0.9272
	C_S	0.8053	0.4620	0.6348	0.8419	0.8354	0.7871	**0.9398**	0.8637	0.6005	**0.9290**	0.9241
	C_K	0.6141	0.3243	0.4482	0.6479	0.6501	0.5895	**0.7752**	0.6736	0.4563	**0.7624**	0.7556
WIQ	C_P	0.7609	0.6537	0.7121	0.7881	0.7575	0.6967	0.8161	0.6404	0.8243	**0.8845**	**0.9429**
	C_S	0.6257	0.5945	0.6549	0.7195	0.7146	0.6084	0.7644	0.5795	**0.8043**	0.7811	**0.8289**
	C_K	0.4626	0.4193	0.5025	0.5261	0.5326	0.4360	0.5803	0.4119	0.6170	**0.6442**	**0.7124**

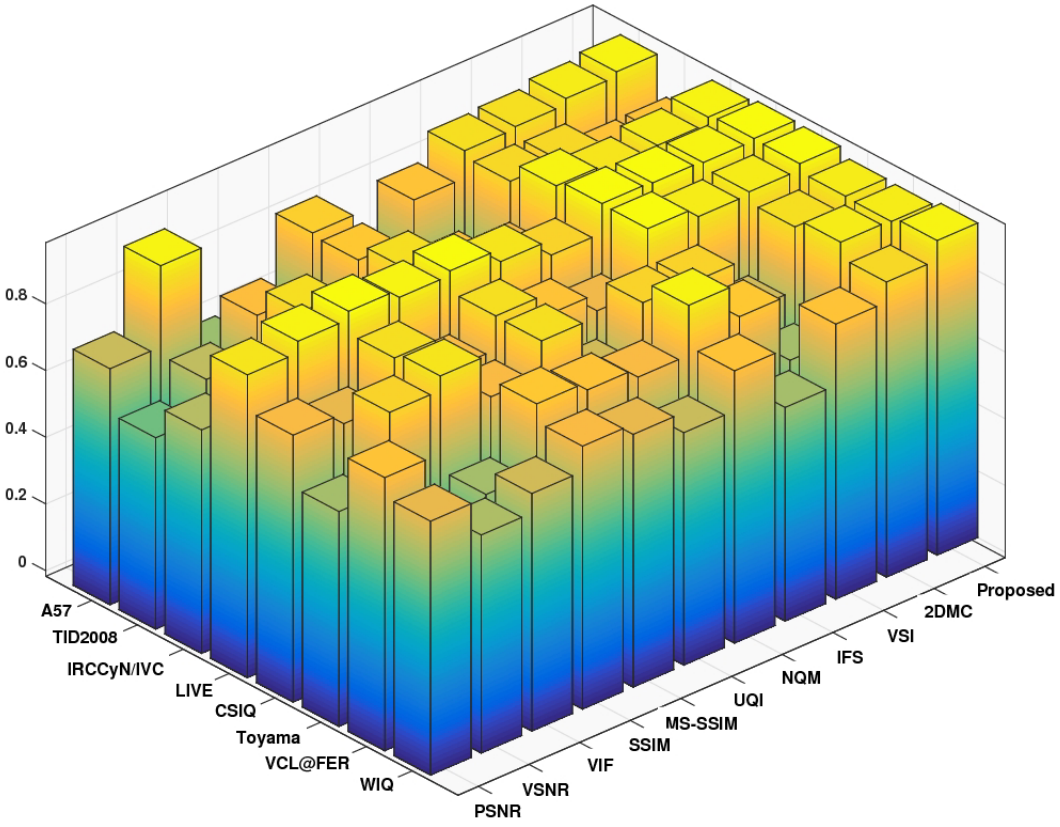


Figure 2.5: The illustration of C_p values corresponding to each dataset and objective quality measure.

By looking at the results presented in Table 2.2 and in Figure 2.5, one can say that PSNR provides the worst performance on all of the databases. In the literature, it is generally stated that PSNR can not model the HVS adequately [41], and therefore, the low performance of the PSNR metric is expected. The VSNR outperformed other objective measures in the A57 database but it didn't provide promising results on the remaining databases. The VIF metric is another objective measure that showed significant performance variations on different image databases. The VIF, outperforming other objective measures

on LIVE and Toyama databases, showed unsatisfactory performance on the remaining databases. Structural similarity based measures, namely SSIM and MS-SSIM, provided moderate performance on the image databases. However, MS-SSIM outperformed other objective measures (except IFS and VSI) on the TID 2008 database. The UQI measure didn't show a satisfactory performance on any database. The NQM, especially efficient on additive type of noise, performs quite well on the VCL@FER database. However, it fails to provide acceptable performance on the other datasets. IFS measure outperforms the other objective measures on the TID 2008, IVC, LIVE and CSIQ databases. However, the performance of the IFS scheme decreases dramatically on the remaining databases. The visual saliency based VSI method provides satisfactory performance on TID 2008 and LIVE image databases but the performance of the VSI method doesn't meet the expectations on the remaining image datasets. Different than the other objective measures, cepstral methods, namely 2D mel-cepstrum (2DMC) and the proposed 2D complex mel-cepstrum, achieved acceptable performance on all of the databases. The proposed technique outperformed other baseline techniques especially on IVC, CSIQ, Toyama, and WIQ image databases.

In order to evaluate the overall performance of each objective metric, the similarity results listed in Table 2.2 are used to compute the average performance. The averaging is carried out in two ways; ordinary and weighted averaging. In ordinary averaging, all similarity measures corresponding to each image dataset have equal impact on the final result. However on the weighted averaging, the similarity measures corresponding to each dataset are multiplied with the weights that are proportional to the number of distorted images in the corresponding dataset. The average performance results of each objective measure are presented in Table 2.3. The standard deviation is another important criterion for the consistency of the objective quality measures. The measure that deviates less than the other measures can be determined to be more consistent. Therefore, the standard deviation of the similarity measures are also presented in Table 2.3. As another performance criterion, the correlation results of the objective quality measures are sorted to obtain a ranking for each correlation type (C_P, C_S, C_K). The method that has the highest correlation value achieves the 1^{st} rank. For

example, by looking at the results provided in Table 2.2, VSNR obtains the 1^{st} rank for C_P measure on A57 image dataset. Similarly, each objective quality metric obtains a ranking between $1, 11$ for each correlation type on each of the image datasets. The average ranking is calculated by averaging the individual rankings through correlation types and image datasets. The average ranking of the correlation values are prensented in the last row of Table 2.3.

The average correlation results corresponding to each objective quality measure are displayed in Figure 2.6. In Figure 2.6, the minimum and maximum correlation values are denoted by upper and lower triangles. The line between the triangles denote the value interval for each correlation coefficient. Also, the length of the interval gives an impression about the deviation values listed in Table 2.3.

Table 2.3: The average, standard deviation and average ranking of the performance of the objective quality metrics.

Averaging Scheme	Criteria	Objective Quality Measures										
		PSNR	VSNR	VIF	SSIM	MS-SSIM	UQI	NQM	IFS	VSI	2DMC	Proposed
Ordinary Averaging	Average C_P	0.7352	0.7528	0.8073	0.7514	0.8628	0.7482	0.8120	0.8736	0.8446	**0.9103**	**0.9277**
	Average C_S	0.7009	0.7524	0.7904	0.7358	0.8574	0.7244	0.8010	0.8647	0.8387	**0.8872**	**0.9041**
	Average C_K	0.5245	0.5876	0.6244	0.5607	0.6777	0.5466	0.6152	0.6969	0.6721	**0.7350**	**0.7594**
	Deviation C_P	0.1007	0.1648	0.1256	0.1453	0.0559	0.1076	0.0902	0.0970	0.1041	**0.0382**	**0.0406**
	Deviation C_S	0.1135	0.1569	0.1320	0.1488	0.0657	0.1253	0.0883	0.1147	0.1193	**0.0580**	**0.0527**
	Deviation C_K	0.1052	0.1592	0.1374	0.1395	0.0786	0.1146	0.0890	0.1239	0.1229	**0.0656**	**0.0635**
Weighted Averaging	Average C_P	0.7203	0.7246	0.8366	0.7619	0.8722	0.7594	0.7562	**0.9029**	0.8658	0.8915	**0.8991**
	Average C_S	0.7182	0.7387	0.8112	0.7535	0.8762	0.7309	0.7559	**0.9084**	0.8715	0.8784	**0.8855**
	Average C_K	0.5429	0.5731	0.6474	0.5782	0.6962	0.5553	0.5788	**0.7418**	0.7073	0.7132	**0.7261**
	Deviation C_P	0.1018	0.1672	0.1290	0.1457	0.0567	0.1082	0.1060	0.1013	0.1063	**0.0426**	**0.0497**
	Deviation C_S	0.1148	0.1575	0.1336	0.1497	0.0683	0.1255	0.0992	0.1228	0.1237	**0.0587**	**0.0559**
	Deviation C_K	0.1069	0.1598	0.1394	0.1406	0.0807	0.1149	0.0961	0.1318	0.1279	**0.0691**	**0.0717**
Average Ranking		9.1667	7.7917	5.5833	7.8750	5.0417	8.5833	7.1250	4.2083	4.8333	**3.4583**	**2.3333**

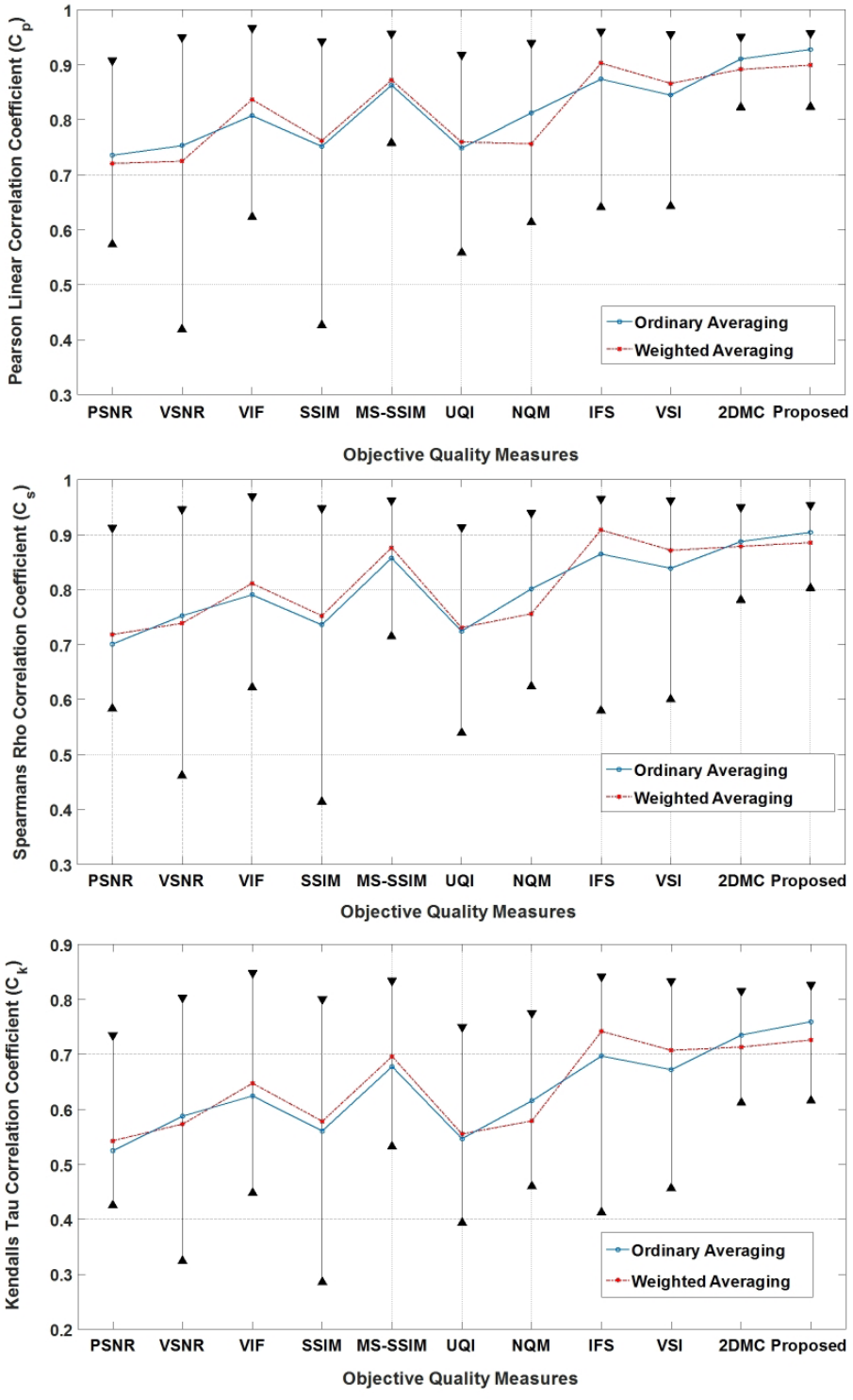


Figure 2.6: Average correlation results corresponding to each objective quality measure.

The results presented in Table 2.3 and Figure 2.6 show that the proposed metric obtains promising results in the overall performance evaluations. The proposed metric outperforms baseline techniques when the performance is evaluated using ordinary averaging. When weighted averaging is used in performance evaluation, the IFS metric obtains better results than the proposed regime due to the size of TID2008 dataset in which the IFS outperforms other metrics. However, IFS metric does not achieve a consistent regime throughout the datasets. The proposed quality scheme not only achieves promising results in average similarity measures, but also provides a consistent regime, i.e low standard deviation of performance on the overall databases. As expected, the baseline techniques showing large performance deviations on different datasets, result in higher deviation values which can be observed in Table 2.3. The proposed complex cepstrum based measure also achieves the highest average ranking that is the another sign of better performance. The overall performance evaluations show that cepstral methods, especially the proposed 2D complex mel-cepstrum, provide significant performance increases on certain databases while obtaining acceptable performance on the remaining databases.

In order to investigate the generalization capability of the proposed quality assessment scheme, a new experiment is carried out by using the IVC dataset. Recall that the IVC dataset contains 10 reference images. In the experiment, the algorithm is trained with a single image by changing the reference image at each time. When a single reference image is used in training, the remaining 9 images are used for testing. This way, the SVR based algorithm is trained with very few samples. In order to quantify the performance of the new experiment, average correlation coefficients are calculated. The average correlation values (C_p, C_s, C_k) obtained by the proposed algorithm on the IVC dataset are 0.8665, 0.8609, and 0.6704, respectively. It is obvious that, the performance of single-image training scheme is worse than the classical $k-1$ partition based training. However, the proposed feature based IQA framework provides promising results even in the case where there is a single image in the training set. This experiment revealed the generalization capability of the proposed method which may enable an efficient representation under circumstances which lack sufficient training data.

In order to further analyze the performance of objective quality metrics on different distortion types, another set of experiments are carried out. The LIVE image dataset [124], containing five different distortion types namely, JPEG2000 compression (JP2K), JPEG compression (JPEG), additive white noise (WN), Gaussian blur (GB) and Fast Fading (FF), is used for this single distortion based evaluation.

Single distortion based quality evaluation is performed using the LIVE image dataset because the objective quality metrics which provide the leading performance on the overall quality evaluations (IFS, VSI, 2DMC and Proposed) suffer performance losses in the LIVE image dataset. In this way, the effect of each distortion type on the overall performance can be examined. The performance of the VIF metric for different distortion types is also evaluated since the VIF metric outperforms the other objective metrics in the overall quality evaluation of the LIVE image dataset. The correlation results corresponding to each degradation type are listed in Table 2.4.

Table 2.4: The average, standard deviation and average ranking of the performance of the objective quality metrics.

Distortion Type	Criteria	Objective Quality Measures				
		VIF	IFS	VSI	2DMC	Proposed
JPEG2000	C_P	**0.9696**	**0.9694**	0.9604	0.9509	0.9571
	C_S	**0.9597**	0.9564	**0.9605**	0.9314	0.9456
	C_K	**0.8209**	0.8131	**0.8153**	0.7894	0.8124
JPEG	C_P	**0.9846**	**0.9778**	0.9761	0.9620	0.9636
	C_S	**0.9791**	**0.9728**	0.9712	0.9532	0.9585
	C_K	**0.8551**	**0.8474**	0.8449	0.8385	0.8466
WN	C_P	0.9858	**0.9883**	0.9835	**0.9859**	0.9852
	C_S	0.9786	**0.9805**	0.9785	0.9796	**0.9803**
	C_K	0.9231	**0.9324**	0.9265	0.9172	**0.9301**
GB	C_P	0.9728	0.9665	0.9527	**0.9735**	**0.9791**
	C_S	**0.9688**	0.9654	0.9562	0.9675	**0.9792**
	C_K	0.8863	0.8763	0.8447	**0.8914**	**0.9090**
FF	C_P	**0.9650**	0.9515	0.9430	0.9466	**0.9537**
	C_S	**0.9614**	**0.9459**	0.9403	0.9445	0.9454
	C_K	**0.8341**	**0.8123**	0.7865	0.7889	0.8046

By looking at the results presented in Table 2.4, one can conclude that the VIF

metric coincides well with the subjective scores. In other words, the correlation results corresponding to the VIF metric are higher than the result obtained by other objective measures. Except the results obtained for WN and GB, the VIF outperforms other metrics for the remaining distortion types. The results in Table 2.4 also reveal that the proposed complex mel-cepstrum technique obtained comparable results with the VIF metric. Moreover, it outperforms the baseline metrics in the presence of GB type of distortion. Although the proposed technique achieves satisfactory performance on WN, GB and FF type of degradation, it fails to provide acceptable performance on degradations caused by compression (JPEG2000, JPEG). Taking into consideration the overall quality evaluations presented in Table 2.2 and 2.3, one can conclude that slight performance losses in single distortion results lead to performance reduction on the overall quality evaluation.

The experiments also revealed that the proposed 2D complex mel-cepstrum outperforms classical 2D mel-cepstrum features in nearly all of the tests by making use of image phase information which contains descriptive features of the image. Therefore, our motivation to enhance 2D mel-cepstrum features with the image phase is validated through the experiments.

2.6 Summary

In this chapter, magnitude-based classical 2D mel-cepstrum [29, 30] features are enhanced by making use of the image phase information in order to propose 2D complex mel-cepstrum features [28] for a full-reference IQA framework. Since the classical 2D mel-cepstrum feature extraction scheme provides a reduced representation, the computation steps carried out for magnitude information are repeated for the image phase. After obtaining the 2D complex mel-cepstrum features by combining the classical 2D mel-cepstrum with the image phase, a regression framework is utilized to map the high-dimensional feature space to the subjective quality scores. In the experiments, the proposed 2D complex mel-cepstrum based IQA framework is compared with the baseline IQA algorithms

on various image databases. Experimental results show that the proposed algorithm provides satisfactory performance compared to the baseline techniques while providing a consistent regime through the datasets.

Chapter 3

Contrast Enhancement Of Microscopy Images Based On Image Phase Information

In this chapter, we introduce an image enhancement framework [33] by inspiring Phase Contrast Microscopy (PCM) in order to enhance the visual details of the images of transparent specimens. To obtain transformations between phase and amplitude differences in our image enhancement framework, the Discrete Fourier Transform (DFT) of the image is computed and image phase information is extracted. Then the frequency components where the phase transitions are large, are amplified with large weights to transform the phase changes into amplitude variations. In addition, a small amount of phase shift is added to the high frequency components. This process eliminates the artifacts on smooth background regions. After the calculation of the inverse DFT, an enhanced image is constructed. The last step of the overall enhancement process is the modified version of the well-known Total Variation (TV) approach [131]. In denoising and deconvolution applications TV is minimized [131, 132]. In this thesis, the TV functional around edges are maximized. Since microscopes have the most expensive optical equipment, there is almost no noise in microscopic images. As a result, edges in the image are enhanced by maximizing the TV within given

constraints. We introduce the concept of Selective Variation (SV) method which maximizes the variations in local regions of interest to achieve sharp transitions around edges. To the best of our knowledge, this is the first study maximizing the $\ell 1$-norm of a cost functional within a region of interest.

In the literature, most image contrast enhancement methods are in the spatial domain [133]. Also, several attempts have been made to utilize the contrast enhancement process in the frequency domain [4,5]. The proposed method which tries to convert the phase variations into the amplitude changes, aims to improve the image contrast and enhance the visual details in the image by utilizing a Fourier domain approach.

3.1 Related Work On Contrast Enhancement Techniques

Contrast enhancement is generally defined as the process of adjusting the dynamic range of the image pixel intensity distribution. The process of contrast enhancement has been widely used in various image processing, pattern recognition, and computer vision applications [134]. In the real scenarios, one can face digital images with low contrast or unbalanced brightness due to malfunctioning of capturing device or external factors such as reflections and filters. Illumination variation is another important obstacle in obtaining an image with a balanced contrast. The contrast enhancement techniques are often used as a preprocessing scheme to increase the performance of subsequent algorithms such as image analysis, object recognition, and etc. The main aim of this process is to increase the visibility of objects of interest in order to assist user oriented or automated tasks including object analysis, detection, segmentation, and recognition.

Contrast enhancement frameworks can be categorized into three main groups, namely pixel-based, transform domain based, and histogram based techniques. In the literature, most pixel-based schemes analyze the HVS in order to generate a

comprehensive neuro-physiological model and to achieve structural understanding. As a pixel-based method, a multi-scale scheme that utilizes anisotropic diffusion equation for the contrast enhancement of mammographic images has been implemented [135]. In another pixel-based method, an image functional is proposed and then minimized to obtain the color enhanced version of the input image [136]. They also implemented a numerical model for the classical gradient descent algorithm which is applied to this image functional. The authors show that such a numerical algorithm coincides with the automatic color enhancement (ACE) method which provides a perceptual-based color enhancement model [136]. In [137], the image captured by the HVS is decomposed into two multiplicative components, namely reflectance and illumination. Based on the Retinex hypothesis [137], several Retinex-based contrast enhancement algorithms, which directly work on the reflectance component, have been proposed [6,138,139]. These methods mainly differ from each other in their choice of filters which are utilized to estimate the illumination component. Although Retinex-based techniques reveal most of the suppressed details efficiently, they may introduce significant artificiality to the image.

The second family of contrast enhancement algorithms are based on transform domain approaches. These methods try to take advantage of transform domains to implement a contrast enhancement scheme. In a discrete wavelet transform (DWT) based scheme [140], the authors make use of the wavelet coefficients together with a nonlinear operator to utilize both enhancement and denoising simultaneously. In another transform domain approach, the histogram of the logarithmic orthogonal transformations such as FT and discrete cosine transform (DCT) are used to propose several image enhancement schemes [141]. The authors performed a group of experiments to measure the performance of the "logarithmic transform histogram matching", "logarithmic transform histogram shifting", and "logarithmic transform histogram shaping using Gaussian distributions" enhancement techniques [141]. A DCT based algorithm which performs operations on chromatic components was proposed for color image enhancement [142]. The DCT based technique does not share the same path with

the classical enhancement algorithms which operate on the luminance component [142]. The curvelet transform is also utilized for the contrast enhancement problem [143]. Since curvelet transform represents edges better than the classical wavelet based approaches, it may provide a good solution for multi-scale edge enhancement tasks. Experimental results show that, the curvelet transform provides better contrast enhancement especially in the presence of noise [143]. As a Fourier domain approach, homomorphic filtering is also utilized for contrast enhancement purposes [4]. The homomorphic filtering is basically a logarithmic transformation followed by a high pass filter. As a special high-pass filter, homomorphic filtering may introduce some Gibbs-like artifacts or amplify the noise in the image [4].

As the third family of contrast enhancement techniques, histogram based techniques perform modifications in the image histogram to improve the contrast [134]. Early global Histogram Equalization (HE) techniques directly modify the histogram without considering the artifacts which may occur on the enhanced image. Moreover, these techniques fail to enhance the local details. Numerous attempts have been made to modify the histogram locally and adaptively in order to eliminate the artifacts on the enhancement results [3, 144–147]. Adaptive HE methods [146, 147] divide the image into sub-blocks and perform HE within these sub-blocks. However, when combining the sub-blocks, a "checker-board" effect may occur on the boundaries of the blocks due to local contrast changes. In order to eliminate grid-shaped artifacts, overlapping blocks can be used at the expense of increasing computational complexity and memory. An efficient implementation called contrast limited adaptive HE (CLAHE) [3, 144], was proposed to overcome the limitations of HE approaches. The CLAHE technique limits the contrast and eliminates the artifacts caused by the mapping of two close gray-scale values to significantly different values. The CLAHE method is usually preferred in offline applications due to its high computational cost. Therefore, there have been several attempts which try to realize the algorithm efficiently for real-time applications [148, 149]. The HE-based techniques have been utilized in many medical imaging applications including chest radiography [150, 151], computer tomography (CT) [152], mammography [153], and cell microscopy [154–156]. Researchers

have widely used HE-based approaches as a powerful pre-processing tool for improving visibility and balancing the contrast in cell images [156–158].

Recently, deep learning frameworks have attracted researchers due to their groundbreaking design and great successes in object classification problem [88]. The researchers have been trying to utilize deep learning frameworks for various applications including the contrast enhancement problem [159]. The authors make use of an offline trained deep auto encoder as a feature extraction scheme to extract signal features from low-contrast images [159]. Then, they use these features to provide feedback to the brightening algorithm to prevent over-amplifying/saturating the bright parts in the images. Although the deep learning based enhancement algorithm obtains promising results in several images, a computationally more efficient scheme can be realized to perform such a brightness adjustment algorithm.

PCM [34] is a well-established optical technique to improve the contrast of the images of the transparent specimens. The method provides contrast enhancement for stained biological materials where the refractive index differences between cellular components cause phase differences of light and these phase differences are transformed into differences in amplitude of light, i.e., areas which are light and dark and can be easily observed [160]. In this way, PCM enhances the image contrast by converting phase differences into amplitude differences.

3.2 Contrast Enhancement Based On Fourier Phase Information

Inspired by the brilliant design of the PCM [34], we propose a contrast enhancement technique by utilizing the image phase information to increase the weight of frequencies where phase variations occur. In general, most natural images are low-pass by nature. In other words, spectral energy is very dense at low frequencies compared to high frequencies in natural images. On the other hand,

sharp transitions and highly structured textures contribute to the high frequency components of an image. In order to emphasize the high frequency components containing important hidden details, it is possible to amplify the high frequency DFT coefficients in Fourier domain. However, frequency-based modifications in Fourier domain should be carried out carefully because improper spectrum manipulations may degrade the natural harmony of the image and cause undesired artifacts in the image domain. Moreover, the modifications must not degrade the conjugate-symmetric nature of the Fourier coefficients.

In the proposed image enhancement algorithm, the high frequency components are emphasized by taking into consideration the image phase. Let $x\,[m,n] \in \mathbb{R}^{M\times N}$ be the $M \times N$ dimensional image. Each step of the proposed technique is summarized below:

The algorithm starts by calculating the two dimensional $\hat{M} \times \hat{N}$-point Discrete Fourier Transform (2D-DFT) $X[u,v]$ of the image $x\,[m,n]$. The DFT parameters should be selected as $\hat{M} = 2^p \geqslant M,\ \hat{N} = 2^q \geqslant N$ in order to take advantage of the Fast Fourier Transform (FFT) algorithm where p and q are positive integers. It is important to increase the FFT size by selecting p and q as large as possible for reducing the spectral noise in the enhancement process. Let $X[u,v]$ be:

$$\begin{aligned} &X[u,v] = |X[u,v]|\, e^{j\phi[u,v]} \\ &u = 0,1,\ldots,\hat{M}-1 \ \&\ v = 0,1,\ldots,\hat{N}-1 \end{aligned} \tag{3.1}$$

A small amount of phase shift (θ) is added to the high frequency components of the image spectrum which are determined by a high-pass mask $H[u,v]$ with

cut-off frequency indices u_c and v_c:

$$H[u,v] = \begin{cases} 0, & u \leqslant u_C - 1 \;\&\; v \leqslant v_C - 1 \\ 0, & u \leqslant u_C - 1 \;\& \\ & \hat{N} - v_C + 1 \leqslant v \leqslant \hat{N} - 1 \\ 0, & \hat{M} - u_C + 1 \leqslant u \leqslant \hat{M} - 1 \;\& \\ & v \leqslant v_C - 1 \\ 0, & \hat{M} - u_C + 1 \leqslant u \leqslant \hat{M} - 1 \& \\ & \hat{N} - v_C + 1 \leqslant v \leqslant \hat{N} - 1 \\ 1, & \text{otherwise} \end{cases} \tag{3.2}$$

Since low frequency components correspond to the smooth and constant regions of an image, adding phase shift to these components may degrade the quality of the image by resulting in undesired artifacts on smooth regions. However, modifying the phase at high frequencies may reveal the suppressed details in the spectrum. Therefore, the phase shift θ is added only to the high frequencies which are determined by the mask $H[u,v]$, in order to preserve the conjugate-symmetric nature of the spectrum. After the phase shift operation, the spectrum of the image becomes as follows:

$$X_P[u,v] = |X[u,v]|\, e^{j(\phi[u,v] + H(u,v)\theta\, \mathrm{sgn}(\phi[u,v]))} \tag{3.3}$$

where *sgn* is the sign operator, which is required to preserve the conjugate symmetry property of DFT. The image phase information is crucially important in image representation [1, 7–9] In order to emphasize the high frequencies which may contain suppressed important details, a weighting scheme is proposed based on the transitions of image phase information between neighboring frequency components. The phase transitions are obtained by computing the phase gradient along horizontal and vertical axes. The phase gradient $\partial_\phi[u,v]$ is computed as follows:

$$\partial_\phi[u,v] = |\phi[u,v] - \phi[u-1,v]| + |\phi[u,v] - \phi[u,v-1]| \tag{3.4}$$

After computing the gradient of the image phase, neighboring frequency components which provide large phase transitions are determined. Let S be the

set of spatial locations of the frequency pairs at which the corresponding phase transitions exceed the threshold τ_ϕ:

$$S = \{(u, v) | \partial_\phi[u, v] > \tau_\phi\} \tag{3.5}$$

The phase transition based weight matrix $W_p(u, v)$ is obtained by using the set S as follows:

$$W_p[u, v] = \begin{cases} \alpha \partial_\phi[u, v], & \text{for } (u, v) \in S \\ 1, & \text{otherwise} \end{cases} \tag{3.6}$$

where α is a fixed scaling factor. The $W_p[u, v]$ weight matrix may have sharp transitions due to direct thresholding; which may cause undesired artifacts in the reconstruction. To eliminate sharp transitions in the weight matrix $W_p[u, v]$, a 2D-Gaussian function ($g[u, v]$) with parameters μ and σ is utilized to smooth the discontinuities as follows:

$$W_f[u, v] = W_p[u, v] * g[u, v] \tag{3.7}$$

where "$*$" is the convolution operator. Similar to the phase addition step described in Equation 3.3, only high frequency components are exposed to weighting in the spectrum by using the high-pass mask $H[u, v]$. As a result, the final weight matrix W is obtained as follows:

$$W[u, v] = \begin{cases} W_f[u, v], & \text{if } H[u, v] = 1 \\ 1, & \text{otherwise} \end{cases} \tag{3.8}$$

The weight matrix $W[u, v]$ is used to amplify the high frequency components in a clever manner in order to reveal the suppressed details existing at the high frequency components. An example final weight matrix $W[u, v]$ is shown in Fig 3.1.

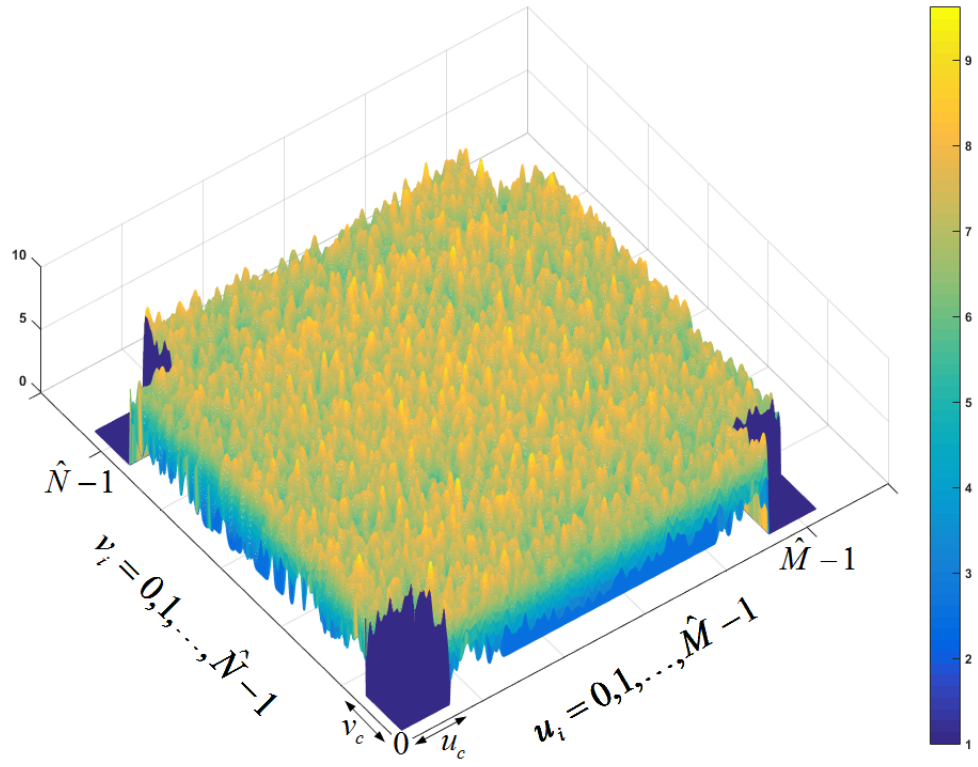


Figure 3.1: $\hat{M} \times \hat{N}$ sample weight matrix obtained for $\alpha = 1.5$, $u_c = v_c = 50$, $\tau_\phi = \pi$, $\mu = 13$, $\sigma = 2$, $\hat{M} = 1024$, $\hat{N} = 1024$.

The weight matrix $W[u, v]$ is applied to the spectrum matrix $X_P[u, v]$ which has been exposed to phase shifts. As a result, the spectrum of the input image is modified in terms of both magnitude and phase after the computation of $X_C[u, v] = W[u, v]X_P[u, v]$.

Inverse DFT (IDFT) of $X_C[u, v]$ is calculated to obtain the contrast-enhanced image. Here, the IDFT is realized using the inverse FFT (IFFT) algorithm:

$$x_C[m, n] = IDFT\{X_C(u, v)\} \tag{3.9}$$

where $x_C[m, n]$ is the contrast enhanced version of the original image $x[m, n]$.

The level of enhancement of the reconstructed image $x_C[m, n]$ is affected by

several parameters including the phase filter cut-off frequency indices u_c, v_c, scaling coefficient α, phase shift θ, phase gradient threshold τ_ϕ, and parameters of the Gaussian smoothing function μ, σ. Therefore, it is important to select the parameters appropriately by maximizing the image contrast. In this study, a Total Variation (TV) based measure is used to maximize the image contrast in a grid search framework. The TV of the original image $x[m,n]$ can be defined as follows [131, 132, 161, 162]:

$$\mathrm{TV}\,(x) = \sum_{m,n} |x[m+1,n] - x[m,n]| + \sum_{m,n} |x[m,n+1] - x[m,n]| \tag{3.10}$$

Instead of maximizing the TV over the entire image, we select "busy" regions of the image and carry out the optimization process over selected regions. In other words, background pixels are not included in the optimization process. Since only partial image information is used in the calculation of "variation", we define "Selective Variation" (SV) as follows:

$$\begin{aligned} SV\,(x_C) = S[m,n] \cdot \Bigg(& \sum_{m,n} |x_C[m+1,n] - x_C[m,n]| \\ & + \sum_{m,n} |x_C[m,n+1] - x_C[m,n]| \Bigg) \end{aligned} \tag{3.11}$$

where the selected pixels are obtained as follows:

$$S[m,n] = \begin{cases} 1, & x[m,n] > \frac{1}{b^2} \sum\limits_{i=m-\frac{b}{2},}^{m+\frac{b}{2}} \sum\limits_{j=n-\frac{b}{2},}^{n+\frac{b}{2}} x[i,j] + \beta \\ 1, & x[m,n] < \frac{1}{b^2} \sum\limits_{i=m-\frac{b}{2},}^{m+\frac{b}{2}} \sum\limits_{j=n-\frac{b}{2},}^{n+\frac{b}{2}} x[i,j] - \beta \\ 0, & \text{otherwise} \end{cases} \tag{3.12}$$

where b denotes the overlapping local window size and β is an additive constant which determines a deviation limit for each pixel from the average values of their local blocks. In other words, the pixels which are greater than β plus the average value of a block, are selected.

The final enhanced image $x_C^*[m,n]$ is obtained by maximizing the SV functional as follows:

$$\begin{aligned} x_C^*[m,n] &= \underset{u_c,v_c,\theta,\tau_\phi,\alpha}{\arg\max}\, SV\left(x_C^P[m,n]\right) \\ \text{subject to } & u_c < \frac{\hat{M}}{2},\ v_c < \frac{\hat{N}}{2} \\ & \frac{\pi}{K_1} \leqslant \theta \leqslant \frac{\pi}{K_2}, \\ & c_1\pi \leqslant \tau_\phi \leqslant c_2\pi, \text{and} \\ & a_1 \leqslant \alpha \leqslant a_2 \end{aligned} \tag{3.13}$$

where K_1, K_2 determines the phase interval added to the high frequency components of the spectrum. The c_1, c_2 coefficients determine the interval of high phase transitions, and a_1, a_2 values are used to limit the scaling coefficient α used in the weighting stage. $x_C^P[m,n]$ is the projection of $x_C[m,n]$ onto the convex set C_E. The convex set C_E is a set of images whose high-pass energy is below a certain level ε and this set is determined as follows:

$$C_E = \left\{x[m,n]\colon \|\partial x[m,n]\|_2^2 \leq \varepsilon^2\right\} \tag{3.14}$$

where $\partial x[m,n] = x[m,n] * h[m,n]$ and $h[m,n]$ is a high-pass filter. Due to the *Projection Theorem*, the orthogonal projection which minimizes the distance between the image and the convex set C_E results in an image on the boundary of the ball with radius ε. 2D illustration of C_E and orthogonal projections are presented in Figure 3.2.

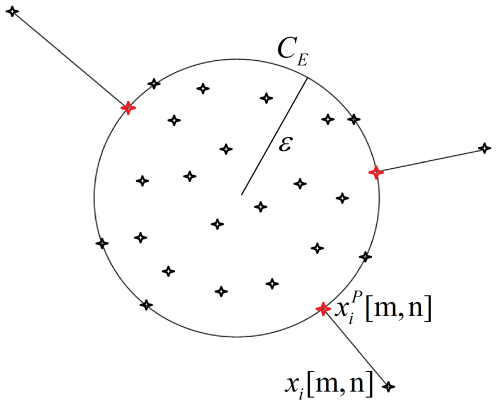


Figure 3.2: 2D illustration of the convex set C_E.

The projection $x_i^P[m,n]$ of the image $x_i[m,n]$ is evaluated as presented in Algorithm 1. In order to simplify the representation, we introduce a new vector notation where $x^P[k]$ represents the projection of the image vector $x[k]$ on the convex set C_E.

Algorithm 1: Projection on C_E

Input: $x[k] \in \mathbb{R}^{1\times M\cdot N}$
Output: $x^P[k]$
if $\|\partial x[k]\|_2 > \varepsilon$ **then**
  $\lambda = \|\partial x[k]\|_2 / \varepsilon$;
else
  $\lambda = 1$;
end
initialization;
$x^P[0] = x_0$;
for $0 < k < M \cdot N$ **do**
  $\partial x[k] = x[k] - x[k-1]$;
  $\partial x^P[k] = \partial x[k]/\lambda = \frac{x[k]-x[k-1]}{\lambda}$;
  $\mathrm{P}\left\{\partial x^P[k]\right\} = \partial x^P[k] = \frac{x^P[k]-x^P[k-1]}{\lambda}$;
  $x^P[k] = \lambda \partial x^P[k] + x^P[k-1]$
end

The projection onto C_E is carried out to limit the high-pass energy of the image which may grow with the increasing values of α. Therefore, an upper bound for α is set by constraining the high-pass energy and SV is utilized in the C_E set. The maximization of SV functional in Equation 3.13 is carried out to determine the aforementioned enhancement parameters $u_c, v_c, \theta, \tau_\phi, \alpha$. The selective maximization process introduced in Equation 3.13 is different than the widely used Total Variation (TV) approaches [131]. In the literature, most denoising and deconvolution applications are based on TV minimization [131, 132]. In this study, we use an alternative approach by maximizing the TV functional around the edges present in the image. Since microscopes use expensive optical equipment, the microscopy images are almost noise-free. As a result, edges in the image are highlighted by maximizing the TV within the constraints described above. The parameter optimization process can be performed only once for a certain family of microscopic images. However, one can repeat the optimization process for any individual image to obtain a custom parameter set for that image.

3.3 Experimental Studies

3.3.1 Image Dataset

To evaluate the performance of the proposed algorithm, microscopic images of human hepatocellular carcinoma cells (HCC) are captured. HCC cells (Huh7 and Mahlavu) were maintained in Dulbecco's Modified Eagle's Medium (DMEM) (Invitrogen GIBCO), supplemented with 10% fetal bovine serum (FBS) (Invitrogen GIBCO), 2 mM L-glutamine, 0.1 mM nonessential amino acids, 100 units/mL penicillin and 100 g/mL streptomycin at 37 °C in a humidified incubator under 5% CO_2. Before capturing the microscopy images, cells were seeded on 100 mm culture plates and grown until they reached 50-60% confluency. Microscopy images of Mahlavu cells were captured by using Euromex Oxion Inverso OX.2003-PL inverted light microscope. For Hematoxylin and Eosin staining, Huh7 cells were inoculated onto coverslips in six well plates. Cells were fixed with 100% ice cold methanol once they reached 50-60% confluency. After washing cells with cold 1x PBS, hematoxylin was added onto cells to be incubated for two minutes and washed with tap water. Then, cells were incubated in 1% acid alcohol (1 unit HCl in 100 units of ETOH) for thirty seconds and washing step was repeated. Next, cells were incubated with 1% ammonia solution for thirty seconds and washed. Finally eosin staining was carried out for one minute and washed. Cells were mounted onto slides using glycerol and visualized under Nikon Eclipse 50i upright microscope and Nikon Eclipse Ti-S inverted microscope using NIS-Elements software. When capturing images using the microscope, two levels of optical zooming were performed. 20x optical zoom provides a wide angle view of the HCC cells. 40x optical zoom enables a narrower field of view and detailed visualization of the organelles of the cells. Totally, 15 images of size 1360×1024 were captured using the inverted light microscope.

To increase the number and variability of the dataset, an additional image set is used [163]. This image set contains the microscopy images of native urine sediments. Before capturing the microscopy images, the urine sediments were centrifuged in 2000 rpm and the supernatant was removed and the sediment

resuspended to create a tenfold concentrated sample solution [163]. 14 sample microscopy images of native urine sediments were collected from a publicly available report [163] and the images in the document are of size 1213×910. Forming an image dataset from microscopy images captured using different imaging devices, environments and purposes enables us to test the proposed algorithm's generalization capability.

The grid search algorithm maximizing Equation 3.13 stops after covering the whole parameter space defined by the user. At the end, the image providing the maximum SV ($x_C^*[m,n]$) is determined as the output of the proposed enhancement framework. The following parameter space is covered by our scheme:

$$\begin{aligned} &u_c \in \left[1, \frac{\hat{M}}{2}\right],\ v_c \in \left[1, \frac{\hat{N}}{2}\right],\ \theta \in \left(\frac{\pi}{2^{10}}, \frac{\pi}{2^3}\right), \\ &\alpha \in (1,3),\ \tau_\phi \in (0.5\pi, 2.5\pi) \end{aligned} \tag{3.15}$$

The algorithm covers the above parameter space to search for the image that provides the largest SV measure. The proposed framework is implemented in MATLAB environment on a computer containing Intel(R) Core(TM) i5 3.40 GHz processor, 8 GB RAM running on Microsoft Windows 7 operating system. The enhancement of a single image of size 1024×1024 takes approximately 1.68 seconds. Experimental results show that the proposed algorithm obtains satisfactory enhancement results in microscopic images, if the parameters are selected as $u_c = v_c = 15$, $\theta = \pi/32$, $\alpha = 2$, $\tau_\phi = \pi$.

3.3.2 Objective Performance Measures

In order to obtain a fair and quantitative performance analysis, objective performance measures are utilized to measure the performance of the algorithms for image enhancement. In this work, "Color Enhancement Factor (CEF)" [164], Universal Image Quality Index (UIQI) [72], adaptive edge map (AEM), Histogram Flatness (HF) & Histogram Spread (HS) [165], Measure of Enhancement

(EME) [166], Measure of Enhancement by Entropy (EMEE) [166], Logarithmic Michelson Contrast Measure (AME) [166] and Logarithmic AME by Entropy (AMEE) [166] objective measures are used in the performance comparison.

The first performance measure used in the tests is the CEF measure. The CEF is based on the ratio between the colorfulness measures of the reference and enhanced images. The colorfulness is generally defined as the attribute of chrominance information perceived by a human observer. The image statistics based colorfulness measure was first proposed by Hasler and Süsstrunk [167]. The colorfulness measure is then modified by defining opponent color spaces [164]. Let R, B, and G be the red, green, and blue image channels, respectively. Based on the RGB color space, the opponent color spaces are defined as follows:

$$\alpha = R - G, \quad \beta = \frac{R+G}{2} - B \tag{3.16}$$

Based on the opponent color spaces, the colorfulness measure C is computed by using the first and second order statistics of α and β as follows:

$$C = \sqrt{\sigma_\alpha^2 + \sigma_\beta^2} + 0.3\sqrt{\mu_\alpha^2 + \mu_\beta^2} \tag{3.17}$$

where μ_α, μ_β, and σ_α, σ_β are the mean values and standard deviations of the opponent components α and β, respectively. Finally, CEF is obtained by computing the ratio between the colorfulness measures of the reference and images as follows:

$$CEF = \frac{C_{enh}}{C_{ref}} \tag{3.18}$$

The second performance measure used in the experiments is the UIQI [72]. As a full reference image quality measure, the UIQI measure is composed of three main components which evaluate the total quality as the combination of "loss of correlation", "luminance distortion" and "contrast distortion" as follows:

$$UIQI = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \cdot \frac{2\mu_x \mu_y}{\mu_x^2 + \mu_y^2} \cdot \frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} \tag{3.19}$$

where μ_x, μ_y, σ_x , and σ_y are the mean values and the standard deviations of the images x and y, respectively. Subsequently, these components are multiplicatively combined in order to generate a final quality score. The UIQI measure takes

values in the interval $[-1, 1]$ and higher values of this measure indicate better image enhancement.

The AEM measure is proposed to quantify the efficiency of algorithms in edge preservation on foreground pixels. The AEM measure can be defined as the ratio between the number of foreground edge pixels preserved in the enhanced image and the total number of foreground pixels on the reference image. In order to extract foreground pixels from the reference image, an adaptive local threshold based approach is utilized. The thresholding technique determines a threshold for each pixel by computing the mean around the k-neighborhood of each pixel. In this way, different regions of the image can be thresholded with different values which are calculated by an adaptive approach. Let $I(x, y)$ be the $M \times N$ reference image. The foreground pixels $T(x, y)$ of the reference image are determined as follows:

$$\begin{aligned} T(x,y) &= \begin{cases} 1, & I(x,y) \geqslant \mu(x,y) + \eta \\ 0, & I(x,y) < \mu(x,y) + \eta \end{cases} \\ \mu(x,y) &= \frac{1}{k^2} \sum_{i=x-k}^{x+k} \sum_{j=y-k}^{y+k} I(i,j) \\ x &= 1, \ldots, M \quad y = 1, \ldots, N \end{aligned} \tag{3.20}$$

where η is an additive constant which determines the minimum level of deviation from the mean to be considered as foreground pixel. In the tests, η is selected as 5. The matrix of foreground pixels $T(x, y)$ is used as a reference to quantify the edge preservation in the enhanced image. The edges of the enhanced image are extracted by using Sobel operator. Also, in order to eliminate the weak transitions, basic thresholding is utilized. At the end, the binary Sobel edge image $I_{Sobel,bin}(x, y)$ is obtained. The AEM measure, which quantifies the edge preservation rate, is computed as follows:

$$AEM = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} I_{Sobel,bin}(i,j) \cdot T(i,j)}{\sum_{i=1}^{M} \sum_{j=1}^{N} T(i,j)} \tag{3.21}$$

By its definition, the AEM measure takes values in the interval $[0, 1]$. Higher values of AEM measure correspond to better preservation accuracies.

Two additional objective performance measures based on image histogram are also used in the tests. One of the histogram based measures is the HF. The HF measure tries to quantify the level of contrast by calculating the ratio between the geometric mean and arithmetic mean of the histogram counts for each histogram bin. Let h_i be the histogram count for the i^{th} histogram bin. Using the h_i's, HF is computed as follows:

$$HF = \frac{\left(\prod_{i=1}^{n} h_i\right)^{1/n}}{\frac{1}{n}\sum_{i=1}^{n} h_i} \tag{3.22}$$

Since the geometric mean of a vector is less than or equal to its arithmetic mean, the HF measure takes values between the interval $[0, 1]$.

The other histogram based objective performance measure is HS. The HS measure can be defined as the ratio between the quartile distance and the total range of the histogram. The quartile distance is defined as the difference between the 3^{rd} (Qt_3) and 1^{st} (Qt_1) quartiles. The Qt_1 and Qt_3 correspond to the histogram bins at which the cumulative distribution function (CDF) of the histogram has 25% and 75% of the total histogram counts. The Qt_1 and Qt_3 of an image histogram are illustrated in Figure 3.3.

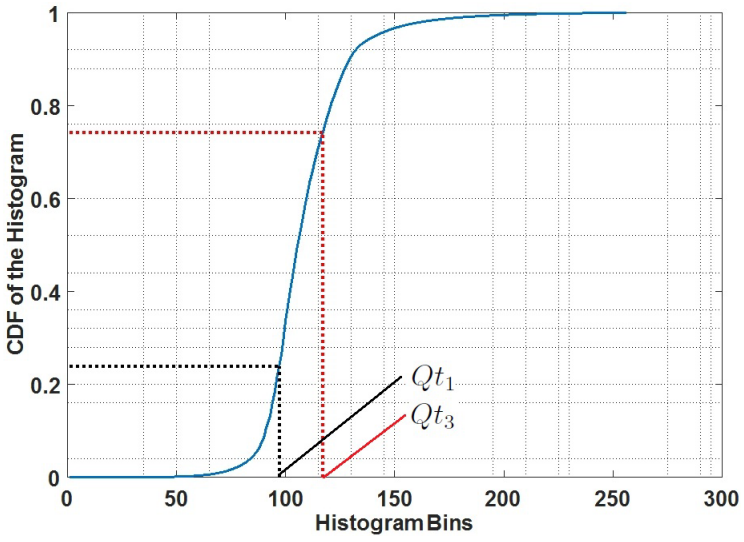


Figure 3.3: Illustration of quartiles Qt_1 and Qt_3.

By using the 1^{st} and 3^{rd} quartiles, HS measure is evaluated as follows:

$$HS = \frac{Qt_3 - Qt_1}{2^d - 1} \tag{3.23}$$

where d is the image bit depth. The images used in the tests are 8-bit images ($d = 8$). Since the quartiles lie between the interval $[0, 2^d]$, HS measure takes values between the interval $[0, 1]$.

In the experiments, an additional set of performance measures, which are based on the ratio between the maximum and minimum values of local image regions, are used. In these methods, the image is first divided into non-overlapping sub-blocks. Let $B_{k,l}$ be the k^{th}, l^{th} image sub-block. Also, let $I_{max;k,l}$ and $I_{min;k,l}$ be the maximum and minimum intensity values of the sub-block $B_{k,l}$. Using the local maximum and minimum values, the EME measure is calculated as follows:

$$EME = \frac{1}{RC} \sum_{k=1}^{C} \sum_{l=1}^{R} 20 \ln \left(\frac{I_{max;k,l}}{I_{min;k,l}} \right) = \frac{1}{RC} \sum_{k=1}^{C} \sum_{l=1}^{R} 20 \ln \left(CR_{k,l} \right) \tag{3.24}$$

where R and C are the number of sub-blocks in the vertical and horizontal directions, respectively. In other words, the image is divided into $R \times C$ non-overlapping sub-blocks.

Using the definitions introduced in the EME measure calculation, the EMEE measure is computed as follows:

$$EMEE = \frac{1}{RC} \sum_{k=1}^{C} \sum_{l=1}^{R} \alpha \left(CR_{k,l} \right)^{\alpha} \ln \left(CR_{k,l} \right) = \frac{1}{RC} \sum_{k=1}^{C} \sum_{l=1}^{R} Entropy(CR_{k,l}^{\alpha}) \tag{3.25}$$

Before defining the AME and AMEE measures, the Michelson contrast measure C_M, which is the basis element of AME and AMEE measures, is defined as follows:

$$C_{M;k,l} = \frac{I_{max;k,l} - I_{min;k,l}}{I_{max;k,l} + I_{min;k,l}} \tag{3.26}$$

where $C_{M;k,l}$ is the Michelson contrast measure of the (k^{th}, l^{th}) image sub-block. Using the definition of block-wise Michelson contrast measure $C_{M;k,l}$, the AME and the AMEE measures are computed as follows:

$$AME = -\frac{1}{RC} \sum_{k=1}^{C} \sum_{l=1}^{R} 20 \ln \left(C_{M;k,l} \right) \tag{3.27}$$

$$AMEE = -\frac{1}{RC}\sum_{k=1}^{C}\sum_{l=1}^{R}\alpha\left(C_{M;k,l}\right)^{\alpha}\ln\left(C_{M;k,l}\right) = -\frac{1}{RC}\sum_{k=1}^{C}\sum_{l=1}^{R}Entropy(C_{M;k,l}^{\alpha}) \tag{3.28}$$

Objective performance measures presented in this section provide quantitative evaluations for the enhancement level of the images. However, the enhancement quality of the images are determined by utilizing a "global" approach. With this approach, the enhancement quality is evaluated by a single score for the whole image. In addition to the evaluation of the enhancement performance by a global approach, a more comprehensive performance evaluation methodology is proposed. In this approach, the input image is divided into $N \times N$ non-overlapping sub-blocks. Then, the objective performance measures are calculated for each sub-block. Using the performance scores obtained for each sub-block, the "performance rate plot" of each method is evaluated. The performance rate plots are similar to the success rate plots used in visual object tracking applications [168]. In order to generate performance plots, a performance threshold vector τ_i^P is determined to control the level of enhancement quality. Then, the number of sub-blocks which provide quality scores larger than τ_i^P is counted. This process is repeated for each τ_i^P varying between the range of the corresponding performance measure. At the end, the performance rate is computed as the ratio of sub-blocks which provide quality scores larger than the τ_i^P. The performance rate plots enable a more efficient performance comparison scheme for the contrast enhancement problem. In the experiments each image is divided into 64×64 sub-blocks. Finally, the performance plots are obtained over 6888 sub-blocks. As an additional performance evaluation criterion, Area Under Curve (AUC) measure is also utilized. The AUC of the performance plot is an indicator to quantify the overall enhancement performance. The AUC measure of each enhancement technique is obtained by computing the area under the performance rate plots.

3.3.3 Baseline Enhancement Algorithms

In order to test the effectiveness of the proposed phase-based contrast enhancement framework, a comparative performance analysis is carried out by comparing the performance of the proposed technique with the performances of baseline enhancement methodologies. The proposed phase-based contrast enhancement framework is compared with several CLAHE techniques, frequency domain methods, and model based approaches. Three CLAHE frameworks [3], which model the results with different distributions (Rayleigh, uniform, and exponential) are used. In addition, a balanced CLAHE approach (BCLAHE) [2], which integrates the process of dynamic range compression and local contrast enhancement, is employed in the comparison. A frequency domain approach based on homomorphic filtering [169] is also used in the comparative analysis. Homomorphic filtering is a well-known technique to enhance images containing illumination variations [169]. A Fourier domain high-pass filtering based approach is also utilized for the comparison. The high-pass (HP) filter used in the tests is the flat version of the weighting scheme presented in Figure 3.1. The high-pass filter is designed with parameters $u_c = v_c = 15$ and $a = 4$ to increase the weight of high frequency components while the low frequency components remain unchanged. Another frequency domain method proposed by Aare Mällo (MEAM) [5] is also studied for comparison purposes. The MEAM method separates an input image into high and low spatial frequency components. A higher degree of control over the dynamic range compression and contrast enhancement can be realized by operating separately on the two components [5]. The last enhancement method used in the comparison is the Retinex based-enhancement framework [6]. The Retinex is a neuro-physiological model which is defined as the lightness and color perceived by the human vision system. This model is based on the receptive fields present in the human visual system. By using a dedicated spatial filter called "surround", the lateral opponent operation of the human visual system could be realized [6].

3.3.4 Experimental Results

Each image is processed by using the enhancement algorithms described in Section 3.3.3. In the first experiment, the proposed algorithm is compared with the baseline enhancement techniques using 15 microscopy images of human hepatocellular carcinoma cells. The enhancement outputs of the proposed and baseline algorithms for two samples of HHC cells are presented in Figure 3.4 and Figure 3.5.

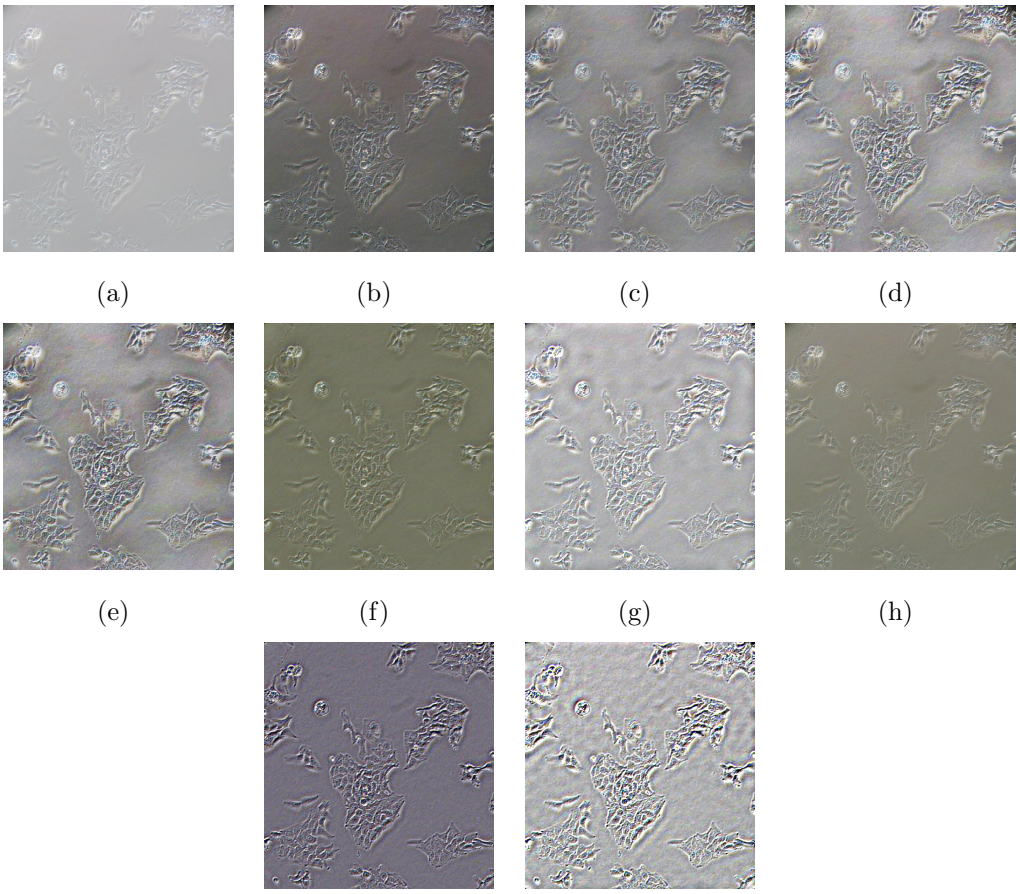


Figure 3.4: The enhancement results of a 20x HHC microscopy image obtained by the proposed and baseline enhancement algorithms. (a) Input image, (b) BCLAHE Image [2], (c) CLAHE Rayleigh [3], (d) CLAHE Uniform [3], (e) CLAHE Exponential [3], (f) Homomorphic Filtering [4], (g) HP Filtering, (h) MEAM [5], (i) Retinex [6], (j) Proposed.

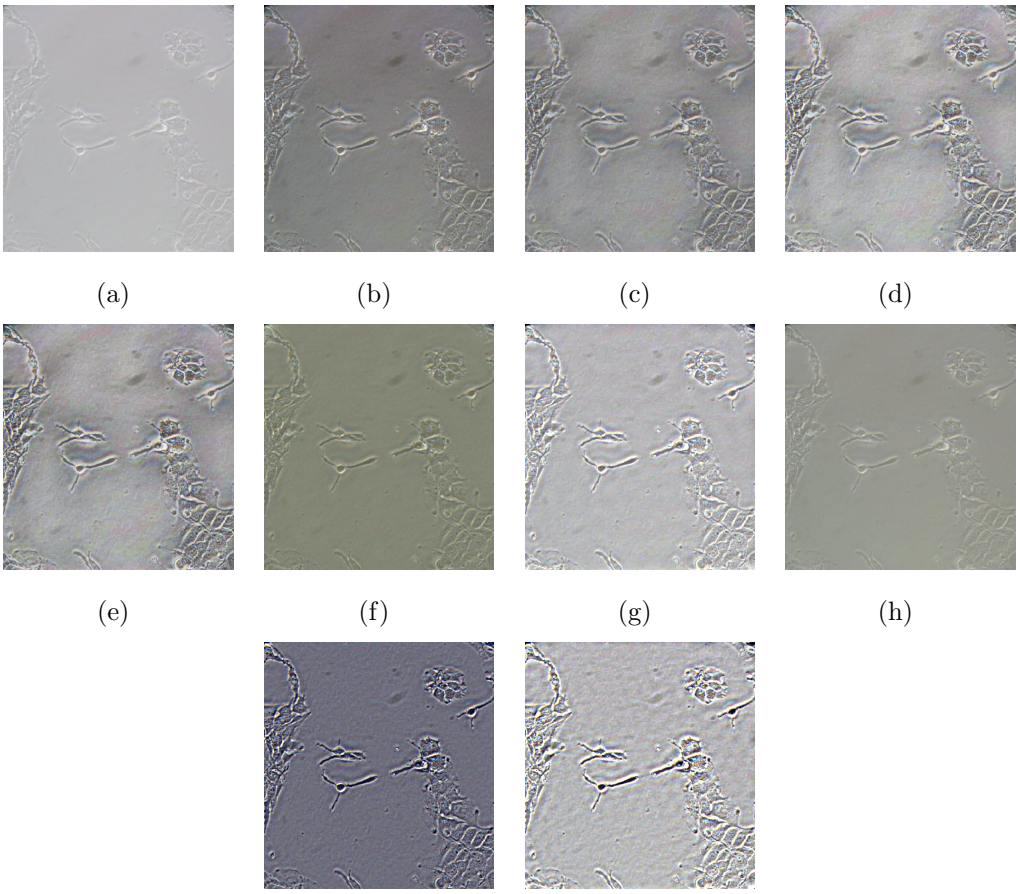


Figure 3.5: The enhancement results of a 40x HHC microscopy image obtained by the proposed and baseline enhancement algorithms. (a) Input image, (b) BCLAHE Image [2], (c) CLAHE Rayleigh [3], (d) CLAHE Uniform [3], (e) CLAHE Exponential [3], (f) Homomorphic Filtering [4], (g) HP Filtering, (h) MEAM [5], (i) Retinex [6], (j) Proposed.

In the second experiment, the proposed and the baseline techniques are utilized for the enhancement of 14 microscopy images of native urine sediments. The enhancement results for two sample microscopy images of native urine sediments are presented in Figure 3.6 and Figure 3.7.

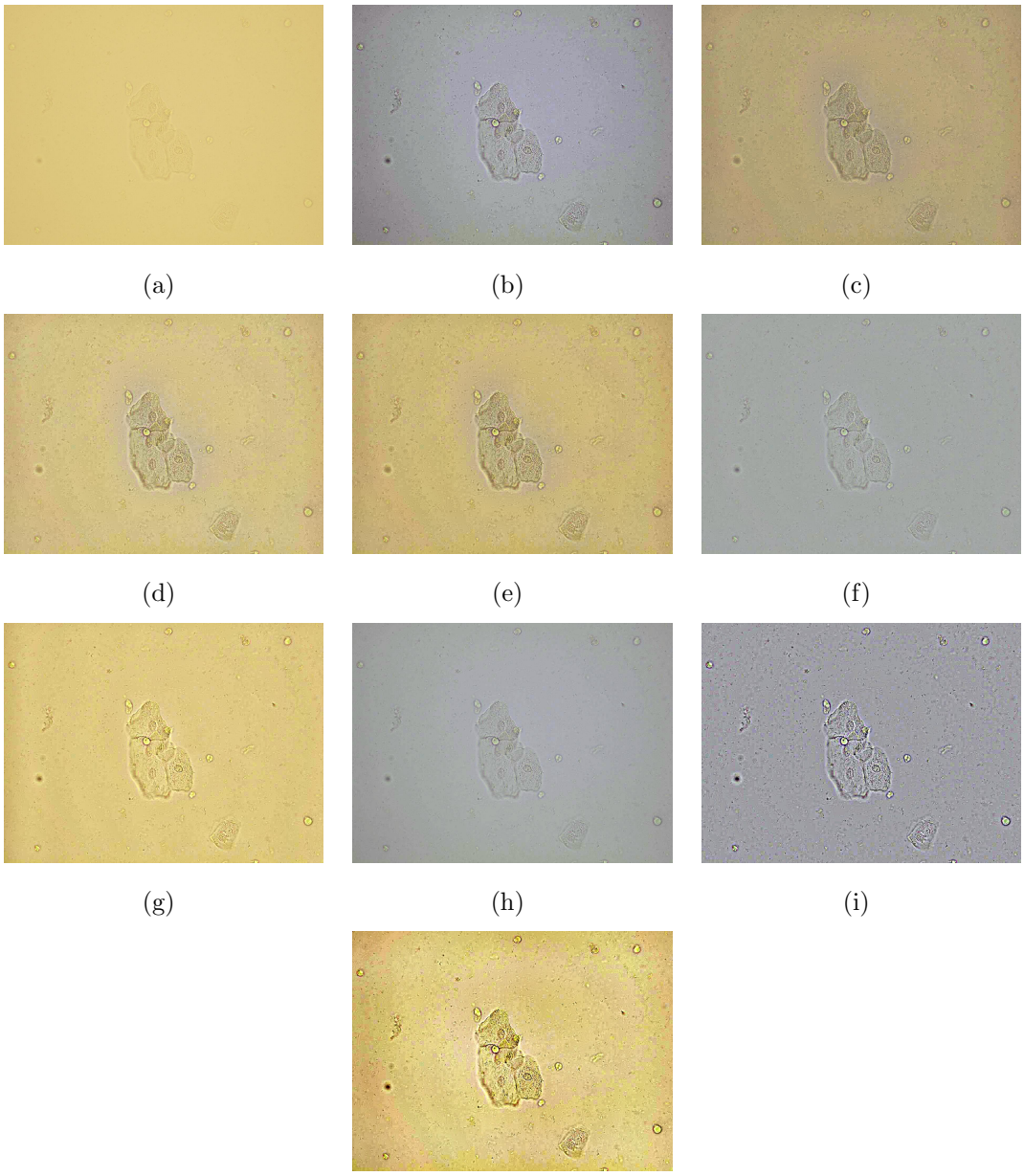


Figure 3.6: The enhancement results of a microscopy image of native urine sediment by the proposed and baseline enhancement algorithms. (a) Input image, (b) BCLAHE Image [2], (c) CLAHE Rayleigh [3], (d) CLAHE Uniform [3], (e) CLAHE Exponential [3], (f) Homomorphic Filtering [4], (g) HP Filtering, (h) MEAM [5], (i) Retinex [6], (j) Proposed.

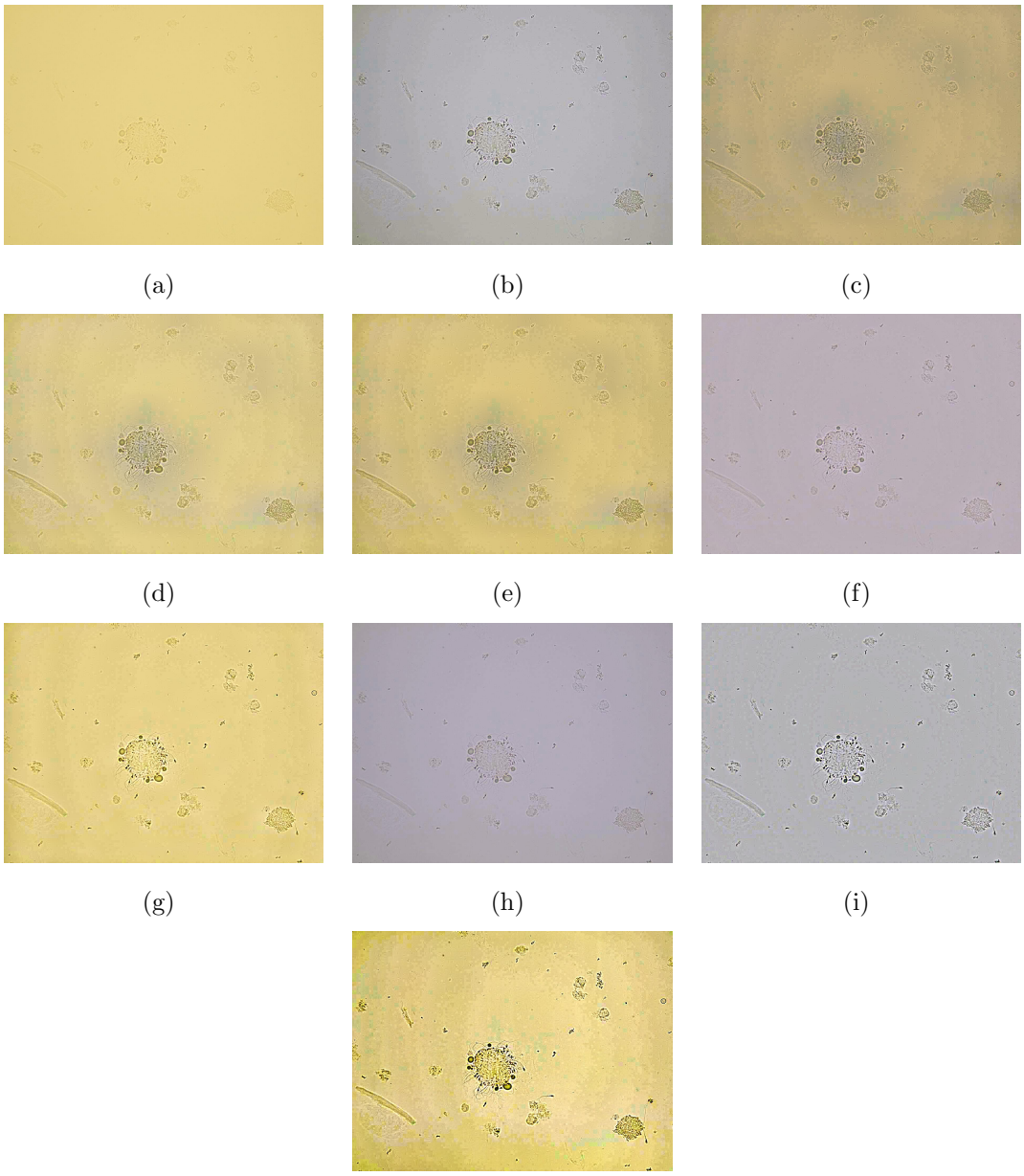


Figure 3.7: The enhancement results of a microscopy image of native urine sediment by the proposed and baseline enhancement algorithms. (a) Input image, (b) BCLAHE Image [2], (c) CLAHE Rayleigh [3], (d) CLAHE Uniform [3], (e) CLAHE Exponential [3], (f) Homomorphic Filtering [4], (g) HP Filtering, (h) MEAM [5], (i) Retinex [6], (j) Proposed.

In order to perform a comprehensive performance test, the performance analysis carried out on color images are repeated for monochromatic images. To obtain monochromatic images, RGB images are converted to the CIE XYZ color space [170]. The luminance component Y of the CIE XYZ space is used for the monochromatic images. The enhancement results for monochromatic microscopy image samples of native urine sediment and HHC cells are presented in Figure 3.8 and Figure 3.9.

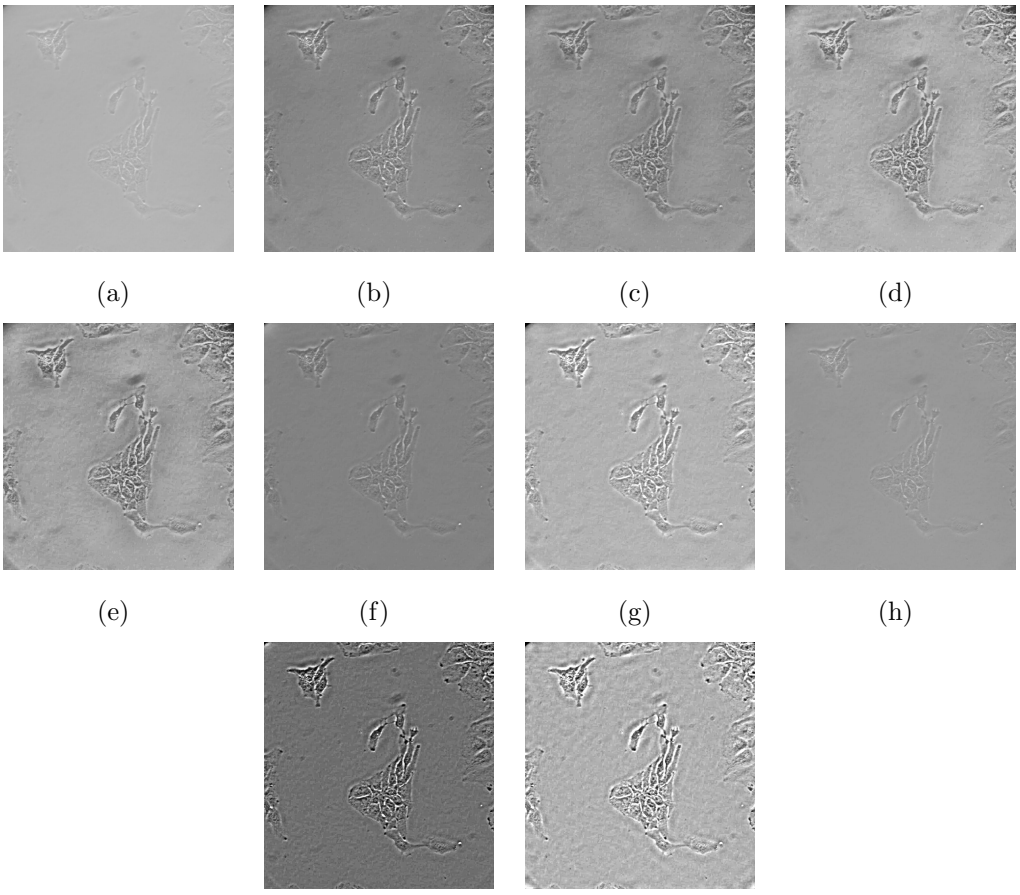


Figure 3.8: The enhancement results of a monochromatic 40x HHC microscopy image obtained by the proposed and baseline enhancement algorithms. (a) Input image, (b) BCLAHE Image [2], (c) CLAHE Rayleigh [3], (d) CLAHE Uniform [3], (e) CLAHE Exponential [3], (f) Homomorphic Filtering [4], (g) HP Filtering, (h) MEAM [5], (i) Retinex [6], (j) Proposed.

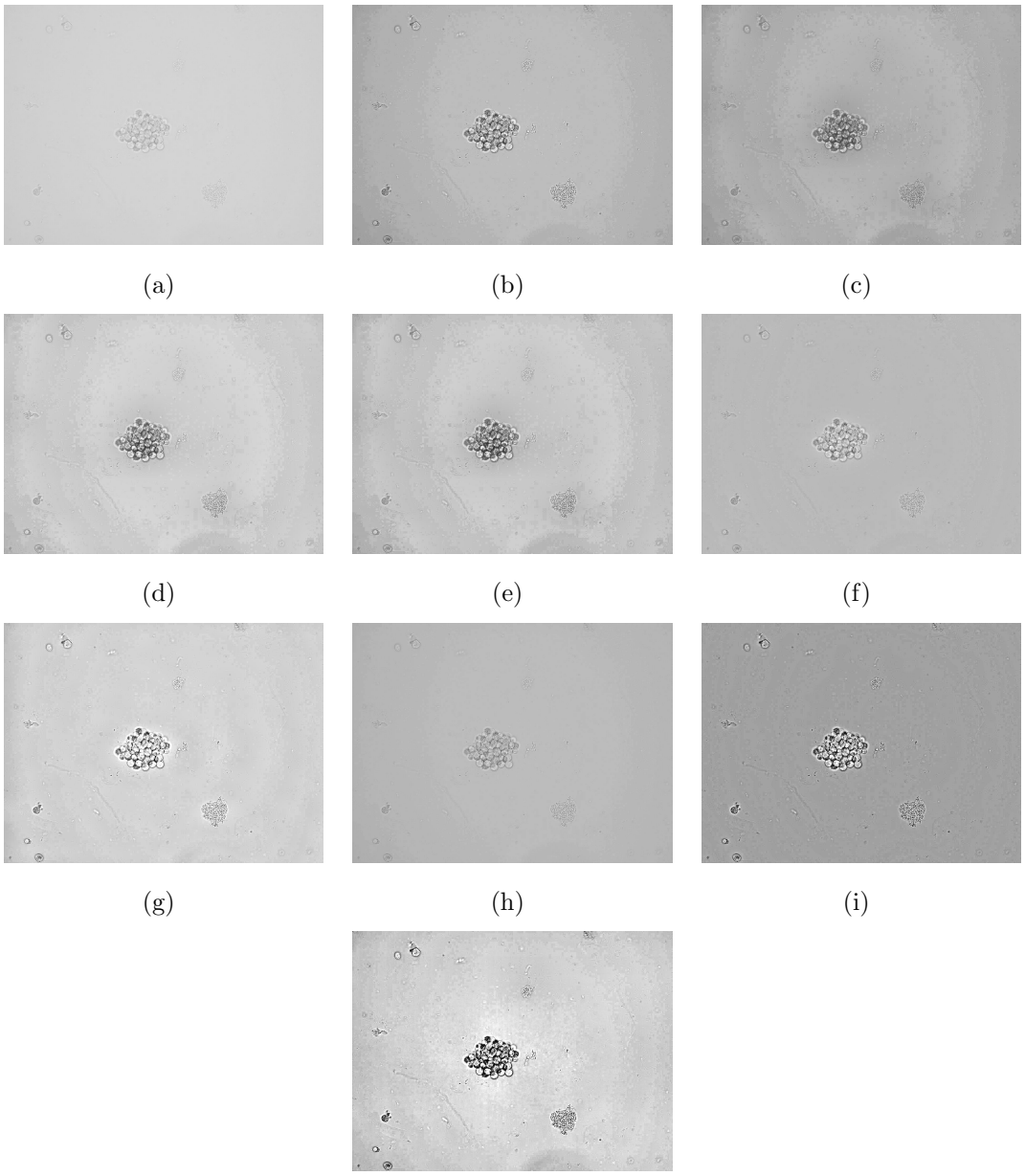


Figure 3.9: The enhancement results of a monochromatic microscopy image of native urine sediment by the proposed and baseline enhancement algorithms. (a) Input image, (b) BCLAHE Image [2], (c) CLAHE Rayleigh [3], (d) CLAHE Uniform [3], (e) CLAHE Exponential [3], (f) Homomorphic Filtering [4], (g) HP Filtering, (h) MEAM [5], (i) Retinex [6], (j) Proposed.

By examining the enhancement results presented in Figures 3.4 - 3.9, the

CLAHE exponential method, the Retinex algorithm, and the proposed phase-based enhancement framework provide noticeable improvements on image contrast. However, it can be easily observed that the proposed scheme suppresses the background better than the baseline techniques while providing higher contrast. Additionally, the proposed methodology preserves and even enhances the color content of the images. The Retinex approach also provides satisfactory performance on contrast enhancement while suppressing the background efficiently. However, the Retinex algorithm fails to preserve the color information in the enhancement results. Undesired effects on the image such as local shading and darkening can be eliminated more effectively by the proposed scheme. The proposed algorithm also provides more detailed and more significant cell structures which may increase the perceptual awareness of the operator/user.

As discussed in Section 3.3.2, objective performance metrics have been utilized to quantify the enhancement performance. In the objective evaluations, the performance of the proposed scheme is compared with the baseline techniques using the CEF [164], UIQI [72], AEM, HF&HS [165], EME, EMEE, AME, and AMEE [166] measures over 29 colored and monochromatic microscopy images. Since the CEF measure quantifies the color enhancement quality, it is not applicable for the tests carried out for monochromatic images. The average performance of each enhancement framework corresponding to each objective measure is presented in Table 3.1.

In Table 3.1, the highest two scores obtained for each objective performance measure are written in bold font. By looking at the results presented in Table 3.1, one can conclude that the proposed technique outperforms the baseline methods in most of the objective quality measures. The Retinex framework also provides satisfactory performance on several performance measures. The results presented in Table 3.1 also show that the enhancement techniques obtain similar results for color and monochromatic images. As stated in Section 3.3.2, the global approach, which is the enhancement evaluation of the whole image with a single score, is not an efficient method to quantify the enhancement quality. Although the results presented in Table 3.1, which are obtained by global approach, give an idea about the enhancement performance, a more comprehensive evaluation approach

Table 3.1: Average performance of the enhancement techniques on both RGB and monochromatic images.

	Objective Performance Measures																
	CEF	UIQI		AEM		HF		HS		EME		EMEE		AME		AMEE	
	RGB	RGB	CIE-Y	RGB	CIE-Y	RGB	CIE-Y	RGB	CIE-Y	RGB	CIE-Y	RGB	CIE-Y	RGB	CIE-Y	RGB	CIE-Y
BCLAHE [2]	1.295	0.901	0.921	0.360	0.362	0.053	0.049	0.052	0.0501	5.051	5.018	0.9842	0.9438	50.75	47.476	0.185	0.186
CLAHE Rayleigh [3]	1.708	0.950	0.947	0.347	0.336	0.031	0.025	0.055	0.055	5.596	5.104	0.519	0.444	48.804	49.431	0.216	0.206
CLAHE Uniform [3]	2.256	0.987	0.989	0.371	0.360	0.072	0.0583	0.072	0.069	6.413	5.846	0.758	0.593	47.83	48.604	0.217	0.209
CLAHE Exponential [3]	**2.346**	0.974	0.975	0.376	0.366	0.086	0.069	**0.076**	**0.073**	7.243	6.571	0.994	0.7245	45.581	46.682	**0.226**	**0.2182**
Homomorphic Filtering [4]	1.181	0.918	0.944	0.298	0.298	0.010	0.009	0.017	0.016	3.130	2.969	0.322	0.307	61.874	60.381	0.150	0.147
HP Filtering	2.163	**0.999**	**0.999**	0.377	0.377	0.059	0.057	0.047	0.043	5.593	5.258	1.310	0.953	51.047	52.001	0.200	0.195
MEAM [5]	0.915	0.948	0.970	0.259	0.260	0.008	0.007	0.025	0.023	2.253	2.225	0.229	0.224	66.459	63.158	0.122	0.122
Retinex [6]	2.321	0.853	0.856	**0.417**	**0.416**	**0.183**	**0.179**	0.045	0.041	**10.932**	**11.051**	**9.371**	**9.2442**	**38.912**	**38.609**	0.214	0.212
Proposed	**3.169**	**0.998**	**0.998**	**0.383**	**0.380**	**0.196**	**0.185**	**0.077**	**0.075**	**9.560**	**8.923**	**5.112**	**3.304**	**41.231**	**40.458**	**0.221**	**0.234**

is utilized by constructing performance rate plots for each objective measure. The performance rate plots corresponding to each objective performance measure and enhancement algorithm are presented in Figure 3.10 - Figure 3.18.

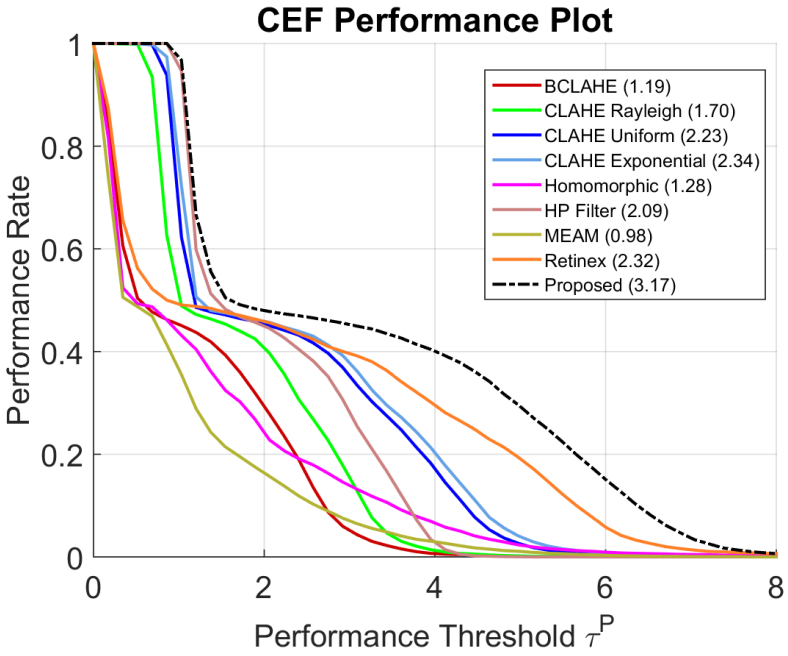


Figure 3.10: Performance rate plots of CEF measure. AUC measures for each enhancement framework are presented in the legend of the figure.

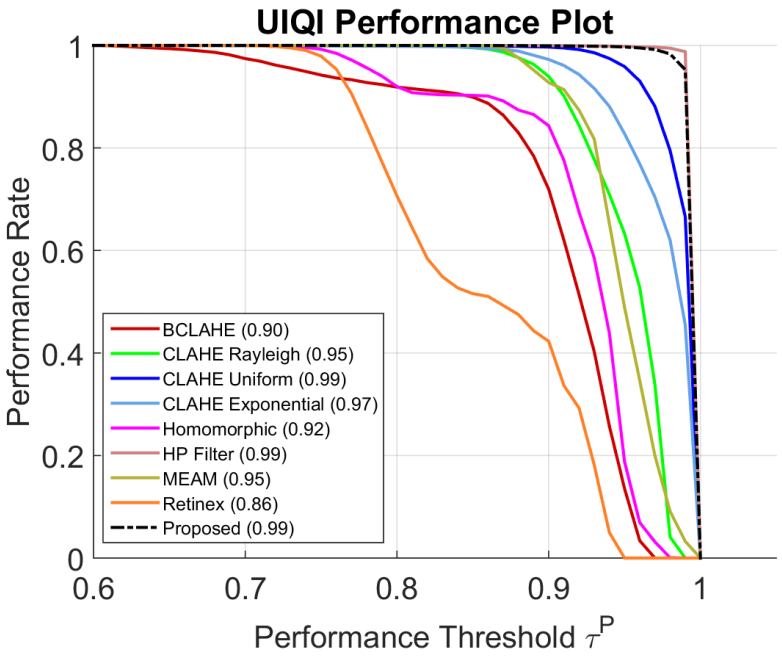


Figure 3.11: Performance rate plots of UIQI measure. AUC measures for each enhancement framework are presented in the legend of the figure.

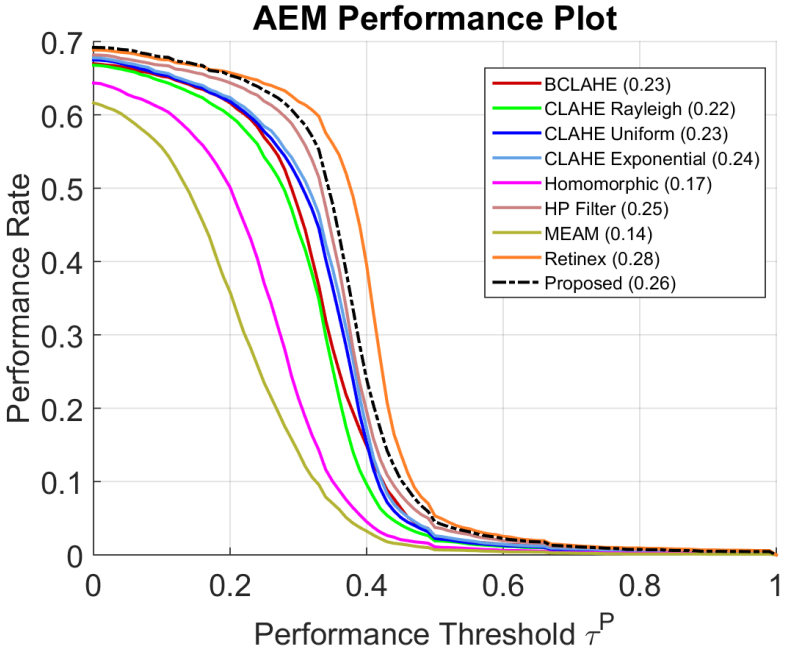


Figure 3.12: Performance rate plots of AEM measure. AUC measures for each enhancement framework are presented in the legend of the figure.

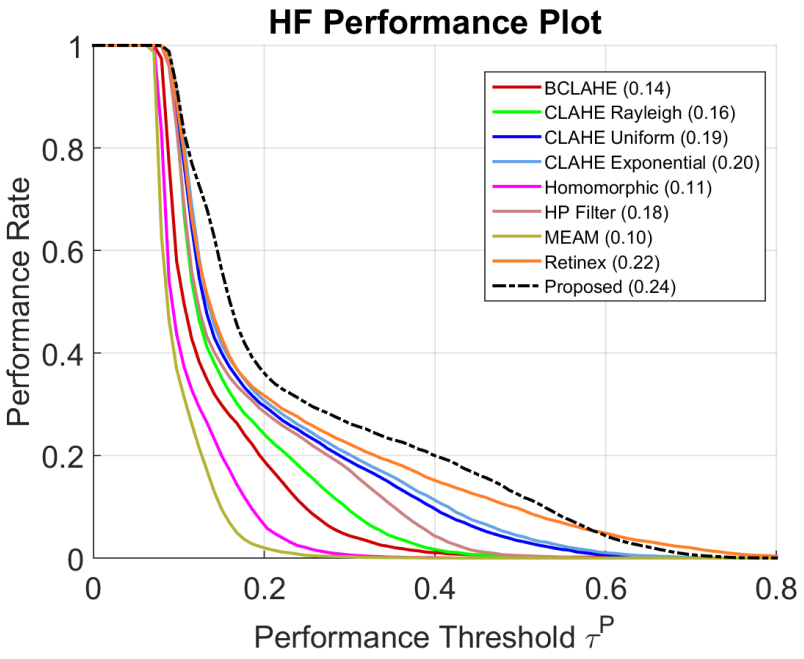


Figure 3.13: Performance rate plots of HF measure. AUC measures for each enhancement framework are presented in the legend of the figure.

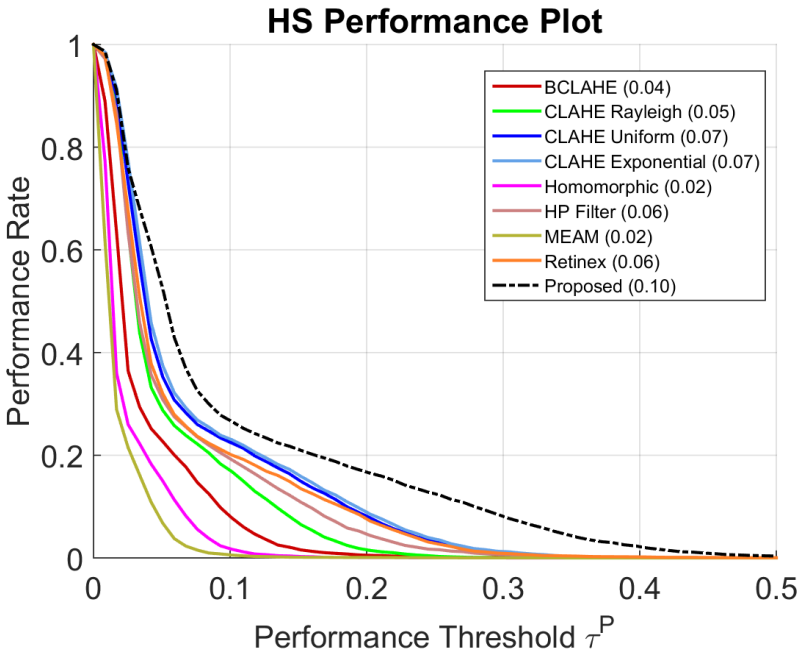


Figure 3.14: Performance rate plots of HS measure. AUC measures for each enhancement framework are presented in the legend of the figure.

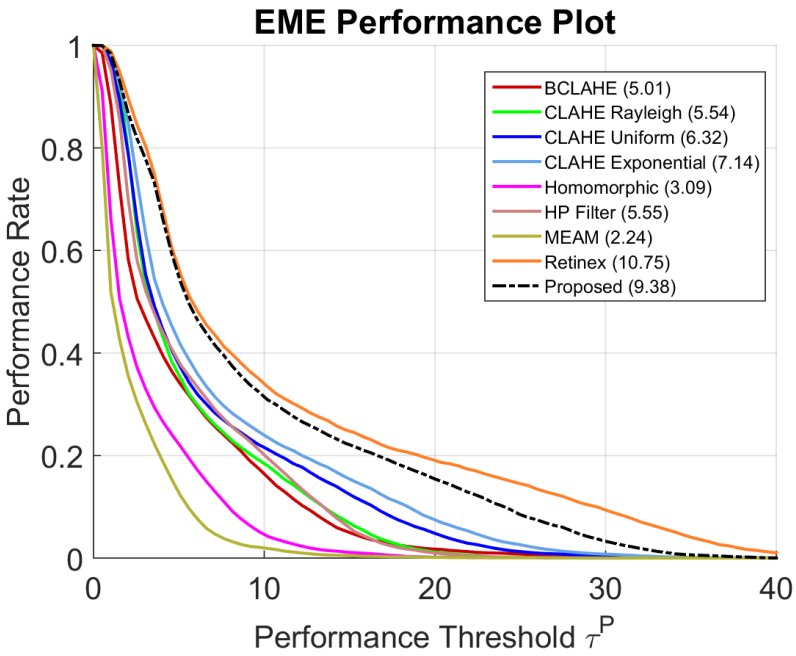


Figure 3.15: Performance rate plots of EME measure. AUC measures for each enhancement framework are presented in the legend of the figure.

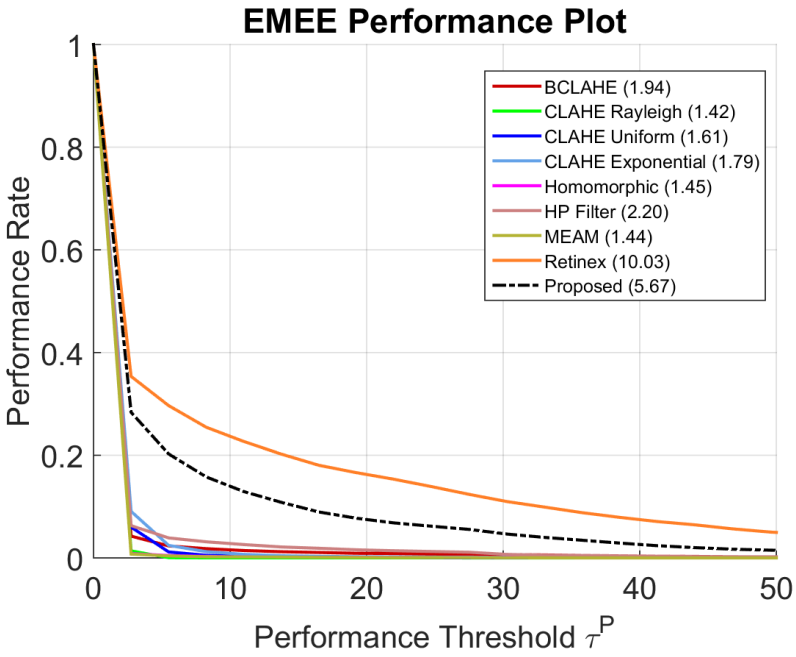


Figure 3.16: Performance rate plots of EMEE measure. AUC measures for each enhancement framework are presented in the legend of the figure.

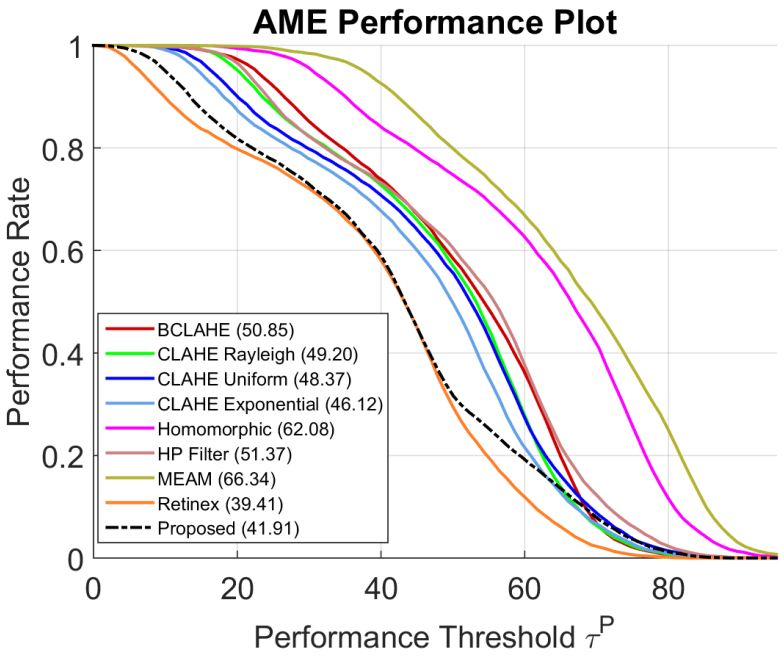


Figure 3.17: Performance rate plots of AME measure. AUC measures for each enhancement framework are presented in the legend of the figure.

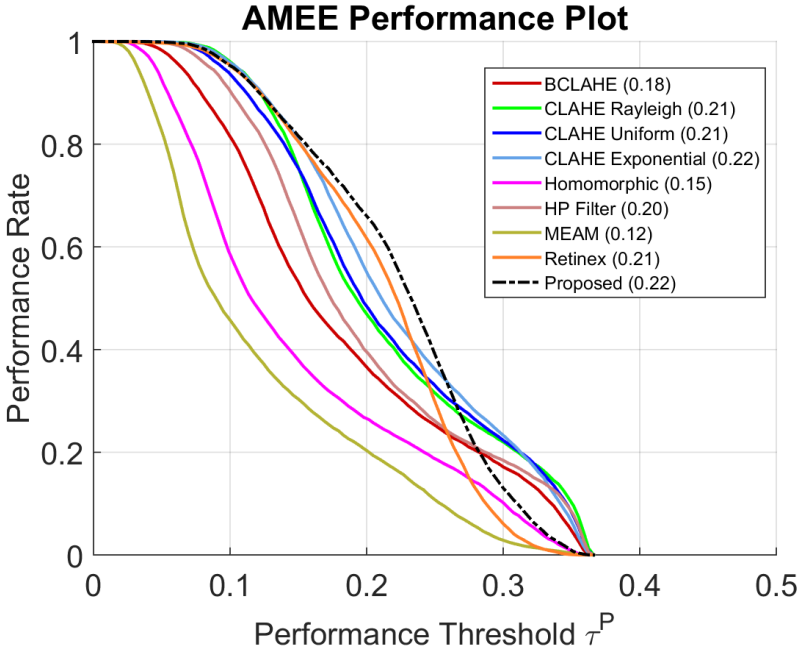


Figure 3.18: Performance rate plots of AMEE measure. AUC measures for each enhancement framework are presented in the legend of the figure.

The performance rate plots for the CEF measure (Figure 3.10) show that the proposed enhancement framework outperforms the baseline techniques on color enhancement. Also, the CLAHE exponential and Retinex methods obtain promising results on color enhancement. Since the AUC measure of the CLAHE exponential technique is slightly higher than the AUC measure of the Retinex framework, the CLAHE exponential technique can be stated as the contender to the proposed framework for the CEF performance measure. The UIQI performance plots in Figure 3.11 show that the HP filtering technique provides the highest UIQI scores together with the proposed phase-based enhancement technique. The CLAHE uniform method is a strong contender to the proposed algorithm and the HP filtering technique on UIQI measure. Surprisingly, the Retinex technique which obtains promising results throughout the test, performs poorly on the UIQI measure. This may be caused by the loss of chrominance information in the enhancement results of the Retinex method.

The performance rate plots for the AEM measure (Figure 3.12) show that the Retinex technique provides the best results for the AEM measure. The proposed phase-based scheme obtains comparable results with the Retinex method on the AEM measure. The AEM measure tries to quantify the capability of edge preservation. The Retinex technique and the proposed scheme obtain acceptable level of edge preservation performances in the tests. By looking at the histogram based performance plots (HF&HS) presented in Figure 3.13 and Figure 3.14, one can conclude that the proposed phase-based scheme outperforms the baseline techniques on histogram based performance measures. Although CLAHE based enhancement methods operate directly on the image histogram, the proposed algorithm provides better histogram flatness and histogram spread performances without using any histogram information. The results show the efficiency of the proposed algorithm in balancing the histogram by nature for better enhancement performance.

The EME and EMEE performance plots presented in Figure 3.15 and Figure 3.16 show that the Retinex method outperforms the competitive enhancement frameworks on EME and EMEE measures. The proposed phase-based

approach also provides satisfactory performance on the EME and EMEE performance plots by obtaining comparable results with the Retinex technique. The remanining enhancement techniques fail to provide acceptable performance on the EME and EMEE measures. By looking at the AME based performance plots presented in Figure 3.17, one can observe that Retinex algorithm outperforms other enhancement techniques on the AME measure. The proposed phase-based approach obtains comparable results with the Retinex algorithm on the AME measure. Note that, unlike other objective performance measures, lower values of AME correspond to better enhancement performance. The AMEE performance plots presented in Figure 3.18 show that the proposed enhancement framework achieves the highest performance on the entropy-based AMEE measure. The CLAHE exponential and the Retinex techniques obtain comparable results with the proposed framework for this measure.

Since the proposed framework is inspired from the translation concept between the phase and the magnitude information in the PCM, an additional experiment has been carried out by comparing the images enhanced by the proposed technique with the PCM images. Therefore, an additional image dataset is gathered by capturing images of hepatocellular carcinoma cells using two different microscopes. The microscopy images of the carcinoma cells were captured by Nikon Eclipse 50i upright microscope and Nikon Eclipse Ti-S inverted microscope using NIS-Elements software. The images that were captured by Nikon Eclipse 50i upright microscope are fed into the proposed enhancement scheme. The enhancement results are compared with PCM images which were captured by Nikon Eclipse Ti-S inverted microscope and they are presented in Figure 3.19.

By examining the images presented in Figure 3.19, we conclude that the proposed framework provides promising results by obtaining similar results with the PCM in contrast enhancement. Most of the cell structures and organelles in the enhanced image become more significant than the ones captured by regular microscopy. The proposed framework introduces some "halo effects" near the dark object boundaries. In this way, the proposed method introduces an additional contrast difference between the foreground and background regions which may be useful for visual cell analysis.

3.4 Summary

In this chapter, an image enhancement framework inspired by the PCM is proposed to improve the structural details in microscopy images. The proposed scheme [33] translates phase variations into amplitude changes in order to reveal the suppressed details at high frequency components. In this way, the highly structured regions of cell images become more significant than the background. The performance of the proposed technique is evaluated for two different types of cell microscopy images. Experimental results show that the proposed framework provides satisfactory enhancement performance on color and monochromatic microscopy images of both hepatocellular carcinoma cells and native urine sediments. Moreover, the proposed scheme outperforms the baseline contrast enhancement techniques on objective performance evaluations. The proposed method can be incorporated into microscopic imaging systems of inverted microscopes.

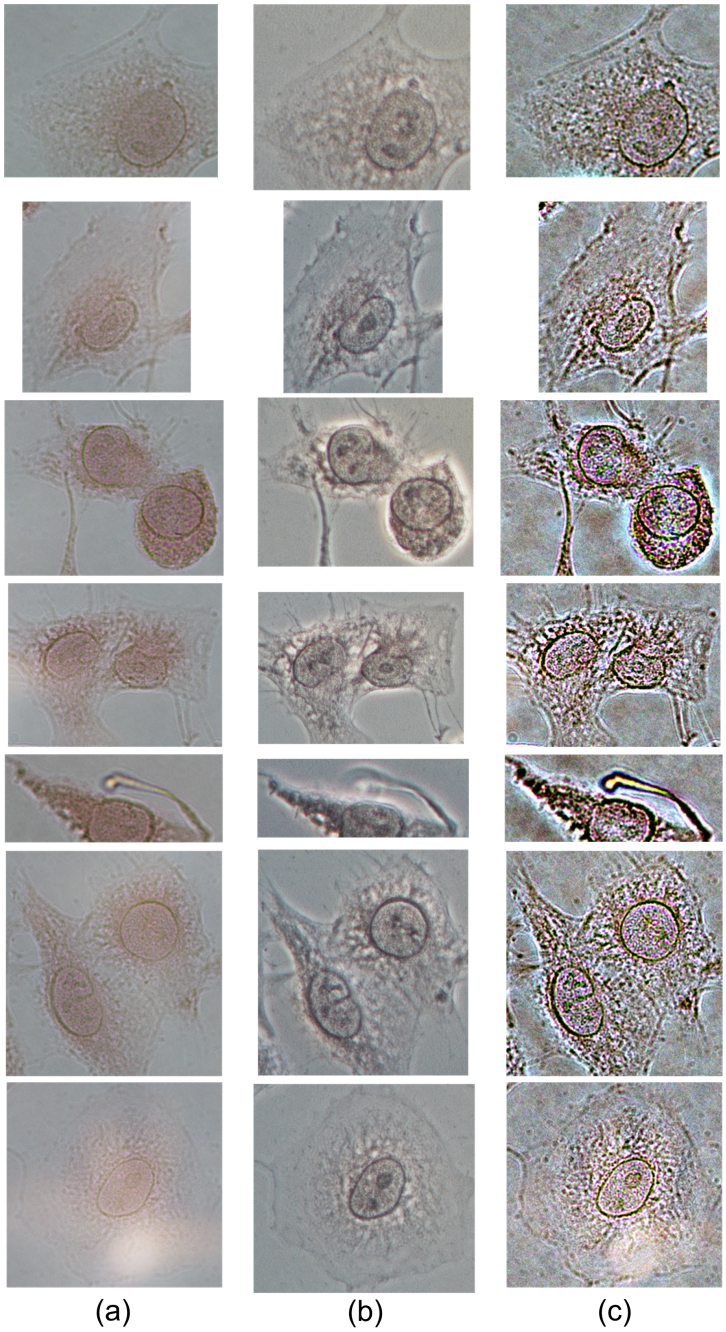


Figure 3.19: The comparison of the proposed technique's outputs with PCM images of hepatocellular carcinoma cells. (a) Images captured by Nikon Eclipse 50i upright microscope (40x). (b) Images obtained by Nikon Eclipse Ti-S inverted PCM microscope (40x). (c) Enhanced versions of the images in (a) using the proposed enhancement framework.

Chapter 4

Phase Based Visual Target Tracking

In this chapter, a computationally efficient visual target tracker scheme based on image phase is introduced. In visual tracking applications, moving objects may result in translational shifts in image domain in a small time window. Since the proposed tracking framework works in the Fourier domain, the translational shifts in the image space are converted to phase variations in the Fourier domain due to the "translational invariance" property of the FT. Therefore, the proposed algorithm estimates the spatial target position by checking the phase information of the FT. The rest of the chapter is organized as follows: In Section 4.1, related work in visual target tracking is presented. The proposed tracking scheme is explained in Section 4.2. Experimental studies and observations are provided in Section 4.3.

4.1 Related Work On Visual Target Tracking Techniques

Visual target tracking has been a popular research area in the field of computer vision, which is utilized in several applications including surveillance [171], military [172], transportation safety [173], human-computer interaction [174], and medical analysis [175]. The tracking problem is generally defined as the estimation of the desired target position given some initial conditions such as the initial position and size of the target. During target tracking, position estimation may become difficult due to occlusion by other objects, object shape deformations, motion blur caused by rapid movements, illumination variations, low contrast between foreground and background, and object scale changes. In order to overcome these difficulties, researchers have made extensive efforts to develop robust target tracking algorithms while taking into account the computational complexity [176].

In the literature, feature-based tracking approaches have been proposed in order to find an appropriate feature set which may improve the performance and the robustness of the target tracker. Depending on the image source and the visual and motion characteristics of the target, the selection of features may change. In the literature, color, edges, motion vectors, textures, or their combinations have been used to address the problem of target representation in tracking applications [177]. The features extracted from the target regions should have distinctive properties such as high differentiation property, invariance to illumination, scale and rotation, and robustness to partial occlusions and noise. In the literature, direct pixel based approaches or statistical features extracted from the pixel values are commonly used [178]. Additionally, kernel-based techniques are one of the most widely used and applied approaches in tracking applications [179–181]. Researchers have also proposed to use strong feature descriptors for target representation. Scale invariant feature transform (SIFT) is one of the most popular feature descriptors proposed in the literature [182] due to its invariance to scale, rotation, and intensity changes, and robustness against affine transformations.

SIFT descriptors have been used in a variety of applications including target tracking, target classification [183,184], image matching [185], and image mosaicing [186]. The SIFT features are more robust to background clutter, occlusions and noise compared to other interest point detectors [187, 188]. Another popular scheme, the covariance feature descriptor, which is relatively computationally more efficient, has been proposed for detection and classification [189]. Due to its high representation capability, the covariance descriptor has also been used in target tracking applications [190–192]. Also, more sophisticated tracker frameworks have been proposed for the visual tracking problem [168]. Although the feature based approaches and complex feature descriptors have been successfully used in tracking applications, the complexity of the solution and the processing times are crucially important for an efficient implementation. Unfortunately, the complex descriptor based approaches fail to satisfy the real-time requirement which is the key necessity in visual surveillance applications.

Rapid advances in deep learning structures and their successes in object recognition applications encourage researchers to utilize deep learning frameworks in various computer vision applications including visual target tracking. Deep learning structures are neural network based approaches which make use of cascaded hidden layers to extract complex hierarchies from the raw data. In particular, CNN achieves significantly higher results compared to the traditional solutions in object recognition [88]. Due to their aforementioned merits, researchers have also attempted to use them for target tracking applications.

One solution to integrate the CNN framework to the tracking problem is using a previously trained (offline) deep learning model [193, 194]. However, some of the offline solutions treat the CNN as a feature extractor mechanism [194–196] as a replacement for hand-crafted features. In [194], the CNN learns the filters using a large dataset of images. In another offline CNN mechanism, highly descriptive structural features extracted by CNN [195] are fed into an online SVM framework to learn the target appearance model. In [196], the authors implemented a transfer learning scheme to transfer the highly descriptive offline features obtained with a large dataset to the online tracking which contains limited number of data. Then they used a heuristic scheme to decide whether the current tracking

results are reliable or not.

The utilization of CNN for online learning based tracking schemes is far too complex compared to the offline learning frameworks. First of all, the deep learning schemes require vast number of training data to avoid overfit. However, there exist only a few positive instances of the object at the track initialization phase. Additionally, training a deep learning structure is computationally too complex for online object tracking applications. In [197], the authors utilize a candidate pool of multiple CNNs to cope with the changes in the appearance of the object. Each CNN in the pool utilizes specific kernels to achieve different representations for target/backround discrimination by using the available low-level cues. The training of these kernels are carried out by using just one instance in the initialization. Then, the kernels are updated frame by frame. At each frame, the most promising CNNs compete in the evaluation of hypothesis of the tracked object. The detection window for each frame is determined by the CNN which provides the highest score. Then, the CNN which provides the highest score is trained using a "warm-start back-propagation" scheme. Meanwhile, the remaining CNNs are not retrained and their models are not distorted by the current update [197]. Although, deep learning based algorithms provides successful results on various tracking benchmarks, their performances strongly depend on the training data which can be problematic in certain cases.

As an image matching framework, template matching based approaches have been utilized in many fields due to their easy implementation and low computational cost [198]. The idea behind these approaches is to measure the similarity between two signal/image patches. The matching process can be performed using various functions and methodologies to define a standard to quantify the matching quality. In the literature, cross-correlation operation has been widely preferred for template matching [199, 200] since it can be calculated in both time domain and frequency domain [201]. However, when the target template becomes larger, the complexity of the solution significantly increases. Therefore, several attempts have been made to reduce the required computational time of the normalized cross-correlation (NCC) algorithm [201–203].

In Section 4.2, we propose a computationally efficient visual target tracker scheme based on image phase information to reduce the computational complexity of the correlation-based matching techniques while preserving the tracking performance. The experimental results and observations are presented in Section 4.3.

4.2 Phase Based Visual Target Tracking Framework

Matching based techniques rely on a target model which is obtained by a target detection mechanism. The detector mechanism may be a user who determines the target region manually or an automatic target detector algorithm which extracts the target region automatically. After the target region is obtained by the detector, the matching-based tracking scheme tries to find regions similar to the target model in the following frames. While searching for the new target location, the tracker does not prefer to search the whole image in order to satisfy computational limitations. Instead of processing the whole image, it performs the search within a sub-window that is determined according to the position of the target in the previous image frame. This sub-window is generally defined as the "search region". The search region not only reduces the computational cost but also improves the tracking performance by ignoring objects which may have similar appearance to the target of interest. In order to measure the similarity between the regions in the search region and the target model, different similarity measures have been utilized. Sum of Squared Differences (SSD), Sum of Absolute Differences (SAD), and the NCC are widely used pixel-wise measures to define similarity between two regions [204].

The NCC technique is one of the most popular matching schemes in the literature due to its efficiency and applicability to a wide variety of applications. In

general, NCC is computed as follows:

$$\gamma[p,q] = \frac{\sum_{x,y} [s[x,y] - \bar{s}_{p,q}] \, [t[x-p, y-q] - \bar{t}\,]}{\left\{ \sum_{x,y} [s[x,y] - \bar{s}_{p,q}]^2 \sum_{x,y} [t[x-p, y-q] - \bar{t}\,]^2 \right\}^{1/2}} \tag{4.1}$$

where s is the search image, t is the target model, $\bar{t}$ is mean of the target model, and $\bar{s}_{p,q}$ is the mean of $s[x,y]$ which is the candidate target region in the search window. In Equation 4.1, the calculation of the numerator dominates the computational cost of the NCC despite the use of transform domain approaches [201]. Also, when the target region becomes large, the computation time of the NCC algorithm increases significantly. Here, our aim is to propose a computationally efficient solution to the image matching problem.

The proposed technique performs matching using the image phase information in the Fourier domain. The proposed scheme also relies on the following assumption: Most of the moving objects result in translational shifts in the image domain in a small time window. This assumption is valid when the frame rate is high. Since the proposed tracking framework works in the Fourier domain, the translational shifts in the image space are converted to phase variations in the Fourier domain due to the "translational invariance" property of the FT. Our solution is derived by starting from the general NCC formulation. The nominator of Equation 4.1 is actually the cross-correlation operation between the target template and search region. By using the well-known "cross-correlation theorem", the frequency domain equivalent of the cross-correlation process is computed as follows:

$$\Gamma[u,v] = S_{p,q}[u,v] T^*[u,v] \tag{4.2}$$

where $S_{p,q}[u,v]$ is the Fourier transform of the sub-region in the search region and $T^*[u,v]$ is the complex conjugate of the Fourier transform of the target model. $\Gamma[u,v]$ in Equation 4.2 can be rewritten as follows:

$$\Gamma[u,v] = |S_{p,q}[u,v]| \, e^{j\phi_{p,q}[u,v]} \, |T[u,v]| \, e^{-j\phi_T[u,v]} \tag{4.3}$$

where $\phi_{p,q}[u,v]$ and $\phi_T[u,v]$ are the phase of candidate target region in the search window and target model, respectively.

The moving objects in the scene result in translational shifts in a small time interval. Magnitudes of the Fourier transforms of the candidate region $|S_{p,q}[u,v]|$ and target model $|T[u,v]|$ are approximately the same because the target region may be exposed to only translational shifts between consecutive frames. In other words, comparing the phase information between the candidate region and target model can be sufficient for the target tracking problem. In the ideal case, the target model and the candidate region are perfectly matched and the phase difference between these regions becomes zero. Therefore, the matching problem can be converted to the following minimization problem:

$$\min \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} |\phi_{p,q}[u,v] - \phi_T[u,v]| \tag{4.4}$$

where M, N denote the size of 2D Discrete Fourier Transform (DFT) which are determined by the image size. Since the minimization problem in Equation 4.4 does not require any multiplications, the matching process can be performed efficiently. Moreover, there is no need to calculate the FFT over and over again for each image frame. When there is a translational shift in the image space, it corresponds to phase variations in the Fourier domain due to the translational invariance property of the FT. The phase variations corresponding to each translational shift can be computed as follows:

$$\phi_{shift}[u,v] = \phi_T[u,v] - 2\pi \left(\frac{um}{M} + \frac{vn}{N} \right) \tag{4.5}$$

The second term in Equation 4.5 can be stored in a look-up table to achieve computational savings. In other words, artificial translational shifts can be generated based on the target model t, and phase variations corresponding to each artificial translational shift can be stored in memory. This operation should be carried out once after the target detection process. In this way, the repetitive FFT calculation in each image frame is no longer necessary for the matching problem and a computationally efficient target tracker can be achieved.

In order to construct look-up tables for phase variations, artificial translational shifts can be formed as follows:

$$t_{(m,n)}[x,y] = t[x-m, y-n] \tag{4.6}$$

where m and n are integers in the interval $[-\epsilon, \epsilon]$. The phase variations $\phi_{T_{(m,n)}(u,v)}$ corresponding to each artificial translational shift can be obtained by computing the phase of the Fourier transform of $t_{(m,n)}$. Then, the target region in the current frame can be obtained by redefining the minimization problem in Equation 4.4 as follows:

$$\begin{aligned} &\phi_{T_{(m^*,n^*)}}[u,v] = \underset{m,n}{\text{minimize}} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \left| \phi_C[u,v] - \phi_{T_{(m,n)}}[u,v] \right| \\ &\textit{subject to } m, n \in [-\varepsilon, \varepsilon] \end{aligned} \tag{4.7}$$

where $\phi_C[u,v]$ is the phase information of the image region in the current frame where the target was located in the previous frame. The translational shift (m^*, n^*) providing the best match with the current target region $c(x,y)$ is determined as the offset between the current and previous target location. If the target of interest is located at (x_{t-1}, y_{t-1}) in the previous video frame, the new target location (x_t, y_t) can be obtained as follows:

$$\begin{aligned} x_t &= x_{t-1} + m^* \\ y_t &= y_{t-1} + n^* \end{aligned} \tag{4.8}$$

Also note that, there is no need to normalize the phase based matching function described in Equation 4.7. This is because the phase of the Fourier transform is not affected by the amplification factors which are all real valued.

The overall process is repeated for each video frame. The flow diagram of the proposed phase-based target tracker scheme is presented in Figure 4.1.

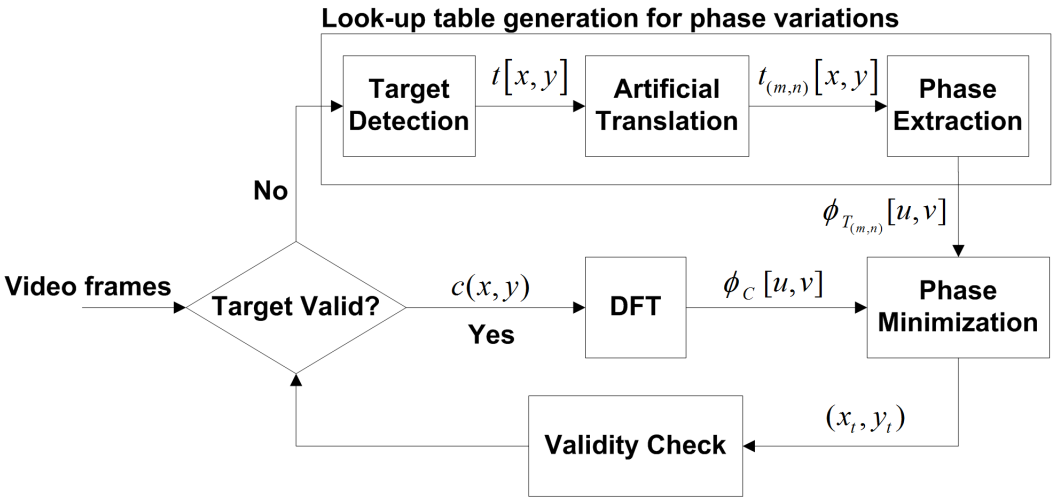


Figure 4.1: Flow diagram of the proposed phase based tracking scheme.

The tracking framework is initialized by target detection. The target detection could be carried out manually by an operator or an automatic target detection algorithm by processing the video frames. After the target region is determined, the target model is exposed to artificial translational shifts to generate look-up tables for phase variations. Here, look-up tables enable a more computationally efficient scheme for target tracking. By using these look-up tables, the current target position is obtained by solving the phase minimization problem defined in Equation 4.7. After the target region is obtained in the current frame, the validity of the target model is checked by comparing it with the one extracted by the target detection process. The target validation measure can be selected from a large variety of functions from pixel-wise error measures to image quality metrics. If the target region obtained by the tracker is valid, the tracking loop continues to operate in tracking mode. Otherwise, the tracking scheme switches to "reacquisition" mode. In reacquisition mode, the detection block tries to determine the target location by utilizing a manual or an automatic detection approach.

4.3 Experimental Studies

In the experiments, the performance of the proposed tracking scheme is tested on six image sequences containing two sea-surface targets, an aerial target and two ground vehicles in an urban environment. In order to measure the performance of the proposed tracker in terms of objective metrics, the ground-truth information (the location of the target on each video frame) is extracted manually by a tracking expert. In the following subsection, objective performance measures are defined to quantify the performance of tracking algorithms.

4.3.1 Objective Performance Measures

In order to evaluate the performance of the tracking techniques in quantitative terms, objective performance measures should be used. In each frame of the test video, the objective measures compare the target information determined by the tracker with the actual target location which was extracted manually by a tracking expert. In the literature, researchers have proposed objective measures to assess a quality level for tracking [168, 191]. In our earlier work, performance measures were proposed to quantify the tracking quality [191, 192]. These performance measures compare the bounding rectangles as well as the rectangle centers obtained by the tracker and the ground-truth information.

In this work, we prefer to use a commonly used tracking benchmark [168] in order to constitute a fair and more up-to-date performance evaluation. The success and precision rates defined in [168] are utilized for evaluating the tracker performance. In order to explain and determine the performance measures visually, the bounding rectangles obtained by the tracker and ground-truth information are illustrated in Figure 4.2.

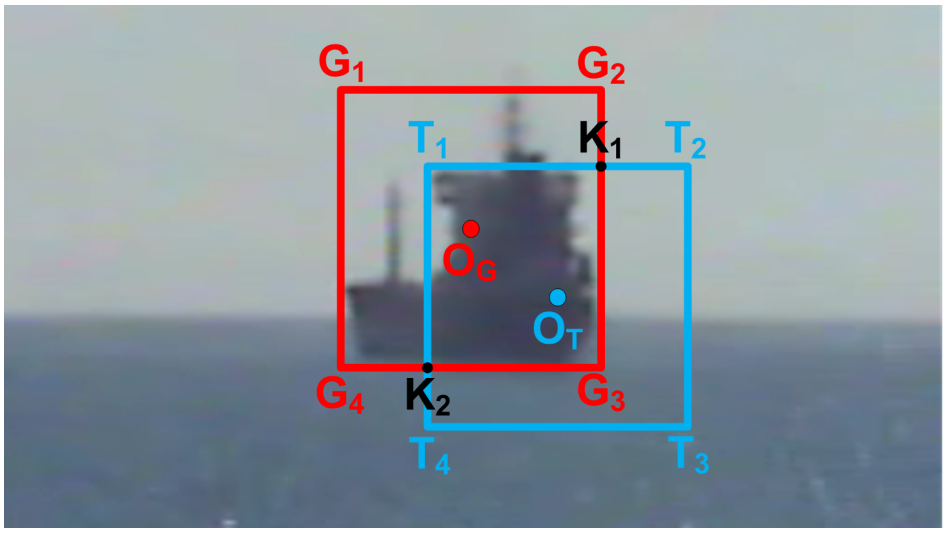


Figure 4.2: The illustration of bounding rectangles obtained by the tracker and data ground-truth information. Rectangle $G_1G_2G_3G_4$: Actual target gate obtained by ground-truth information. Rectangle $T_1T_2T_3T_4$: Target gate obtained by a tracking algorithm.

The first performance measure, success rate, is based on the area of the overlapping region between the ground truth information and the bounding box determined by the tracker. After computing the overlapping area, the "overlapping score" β is obtained by normalizing the overlapping area with the union of the target gates obtained by the tracker and ground truth information. By using the illustration presented in Figure 4.2, the overlapping score can be defined as follows:

$$\beta = \frac{Area\,(T_1K_1G_3K_2)}{Area\,(G_1G_2K_1T_2T_3T_4K_2G_4)} \tag{4.9}$$

The overlapping score β takes values between $(0, 1)$ depending on the location of the bounding boxes produced by tracker and ground truth information. If the target gate obtained by the tracker is exactly the same as the actual target gate β score becomes 1, which is a sign of perfect overlap. The overlapping score β is computed for each frame of the video. In order to generate success plots, a threshold τ_S is determined to control the level of overlapping. Then, the number of frames which provide a β larger than τ_S is counted. This process is repeated for each τ_S value varying between the interval $(0, 1)$. At the end, the success

rate is computed as the ratio of frames which provide β values larger than the τ_S which varies between the interval $(0, 1)$. Since the success rate determines the relation between the overlapping frame ratio and the overlapping threshold, it is generally called as "success plot".

The second performance measure, precision rate, is based on the pixel-wise error between the center locations of the bounding boxes (O_G and O_T in Figure 4.2) obtained by the tracker and ground truth information. To evaluate the precision rate, the first step is to calculate the Euclidean distance between the centers of the target gates produced by the tracker and ground truth information for each frame of the video. In order to generate precision plots, a threshold τ_P is determined to control the level of pixel-wise error between the centers of the bounding boxes. Then, the number of frames which provide a pixel-wise error smaller than the error threshold τ_P is counted. This process is repeated for each τ_P value varying between the interval $(0, 50)$. At the end, the precision rate is computed as the ratio of frames which provide pixel-wise error values smaller than the τ_P which varies between $(0, 1)$. The precision rate is also named as "precision plot".

In addition to success and precision rates, three ranking measures namely, area under curve (AUC), track maintenance (TM), and localization accuracy (LA) are used in order to rank the trackers according to their performances. The ranking measures have been widely utilized to quantify the overall performance evaluation of the trackers [168, 205]. The AUC and TM measures are extracted from the success rate while LA measure is derived from the precision plot. The AUC measure corresponds to the area under the success plot and TM measure is the percentage of frames in which a non-zero overlap occurs between the target gates obtained by the tracker and ground truth information. The LA measure is defined as the percentage of frames in which the pixel-wise error in localization of the target by tracker is below a certain threshold. In this work, the acceptable pixel-wise error is selected as 5 pixels. Therefore, the LA measure is determined as the precision value corresponding to 5-pixel error threshold.

4.3.2 Dataset

Evaluating a tracking scheme is a difficult task due to uncontrolled factors which may directly affect the tracking performance. An ideal tracking scenario without any external factors such as illumination change, partial occlusion, etc. may provide misleading results in the performance evaluation of a tracking scheme. Therefore, the selection of the test dataset plays a crucial role in the performance evaluation. In this context, in order to measure the performance of the proposed tracking scheme in an appropriate manner, different tracking scenarios containing complex background structures, illumination changes, partial occlusions, and motion blur due to the unstabilized video capture, have been considered. In the experiments, a database is constructed by using six videos containing outdoor captures of moving vehicles. After capturing the videos, targets which are intended to be tracked are determined. In order to create a controlled test set, the ground truth information is extracted manually by an expert. Here, the operator determines the target rectangle manually at certain frames using a customized software. The videos used in the experiments are captured by a visible band and an long-wave infrared (LWIR) camera.

In the first video ("Video_Seq_1"), there are 1000 image frames of a moving sea-surface platform where the platform is occluded by other sea-surface targets in certain frames. Video_Seq_1 contains a good scenario to test the robustness of the tracking algorithm when there is a partial occlusion on the target. The second video ("Video_Seq_2") contains 500 frames of a moving fishing boat on a complex background. The background is complex due to sea glints caused by the reflection of the sunlight on the waves. The third video sequence ("Video_Seq_3") contains a considerably fast moving aerial platform. The aerial platform is exposed to illumination changes in certain frames of the total 120 video frames. Moreover, the visibility is low in the captured frames due to the atmospheric conditions during the time of recording. The fourth video ("Video_Seq_4") contains 200 frames of a moving motorcycle in an urban environment. Since the video is captured in a populated area, the background contains other moving objects. Moreover, the capturing device is not stabilized and some of the video frames

are blurred due to undesired movements of the capturing device. The fifth video ("Video_Seq_5") contains 198 frames of a moving car in an urban environment. This video is captured by a LWIR camera and the video contains illumination changes and partial occlusions in certain frames. The last video (Video_Seq_6) used in the experiment contains 625 frames of an approaching ship captured by an unstabilized LWIR camera which is exposed to undesired vibrations throughout the video. This video, which is originally named as "boat1", is obtained from the Visual Object Tracking dataset (VOT-TIR2016) [206, 207]. Table 4.1 provides a brief summary about the videos used in the experiments.

Table 4.1: The basic properties of the videos used in the performance evaluation.

	Video/Scenario Properties				
	Type	Image Size	# of frames	Scenario	Platform
Video_Seq_1	Visible	640×480	1000	Nautical	Ship
Video_Seq_2	Visible	640×480	500	Nautical	Ship
Video_Seq_3	Visible	720×576	120	Aerial	Helicopter
Video_Seq_4	Visible	640×480	200	Ground	Motorcycle
Video_Seq_5	LWIR	320×240	198	Ground	Car
Video_Seq_6	LWIR	640×480	625	Nautical	Ship

4.3.3 Baseline Techniques

The proposed tracking scheme is compared with the baseline techniques using the objective performance measures over the six videos. The baseline techniques used for the comparisons are: The Discriminative Scale Space Tracker (DSST) [208], Fast Compressive Tracker (FCT) [209], Incremental Visual Tracker (IVT) [210], kernelized correlation filter (KCF) [211], sum of template and pixel-wise learners (Staple) tracker [212], Minimum Output Sum of Squared Errors (MOSSE) tracker [213], Spatially Regularized Discriminative Correlation Filters (SRDCF) tracker [214], and NCC [201] tracker. These techniques are well established and sophisticated tracking schemes which are used in tracking benchmarks.

4.3.4 Tracking Experiments on Different Scenarios

The proposed tracker scheme is tested by using the six videos in the dataset. The target gates produced by the tracker are stored at each frame instance to compare with the actual target gates in the ground-truth information. To visualize the tracking process, the target gates produced by the proposed tracking scheme are marked with red symbologies for each video sequence. Also, each video frame processed by the tracker is marked with the frame information to make the tracking scenarios practical to follow. Example target gates for "Video_Seq_1", "Video_Seq_2", "Video_Seq_3", "Video_Seq_4", "Video_Seq_5", and "Video_Seq_6" are presented in Figure 4.3, Figure 4.4, Figure 4.5, Figure 4.6, Figure 4.7, and Figure 4.8, respectively.

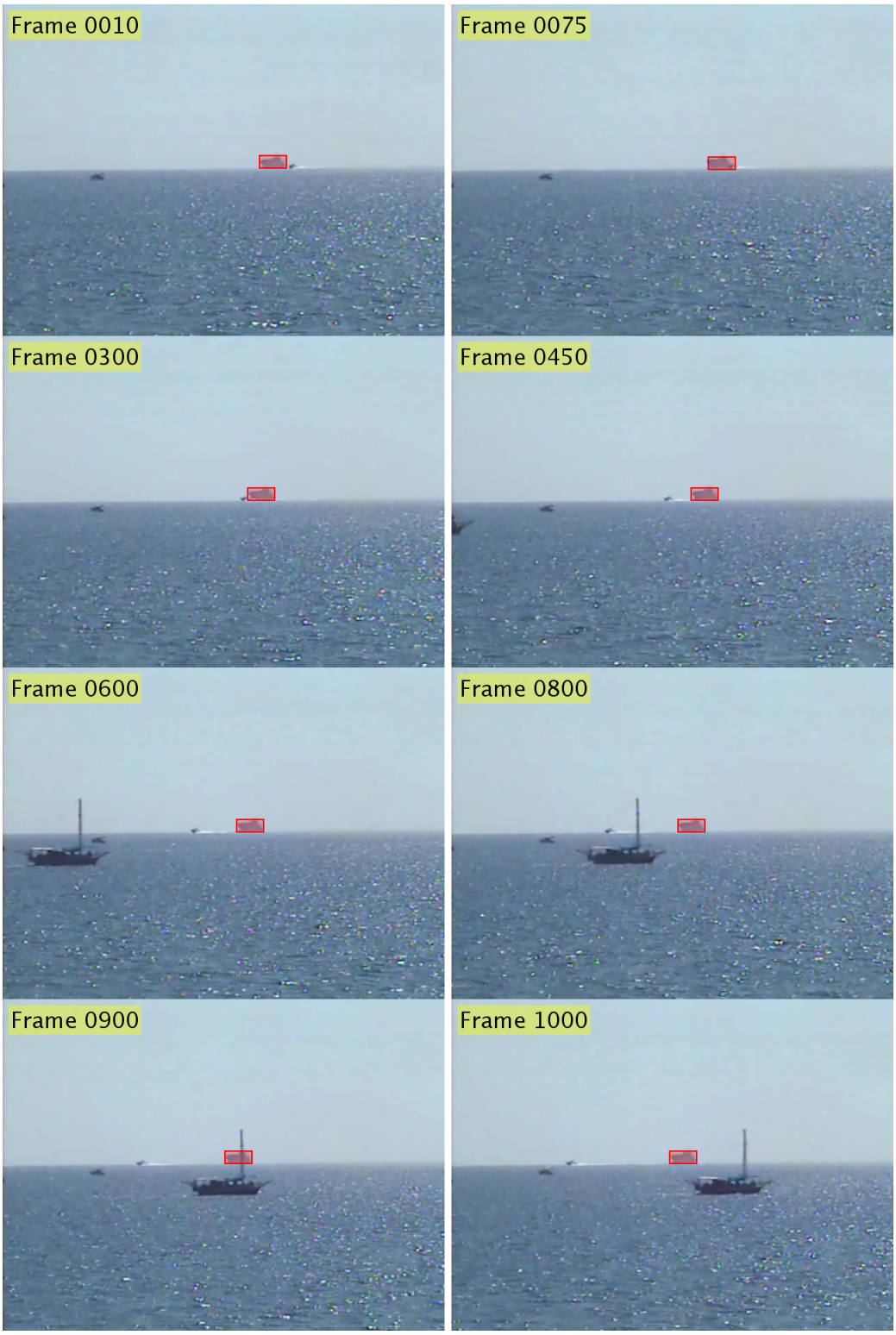


Figure 4.3: Sample target gates produced by the proposed tracking scheme in Video_Seq_1.

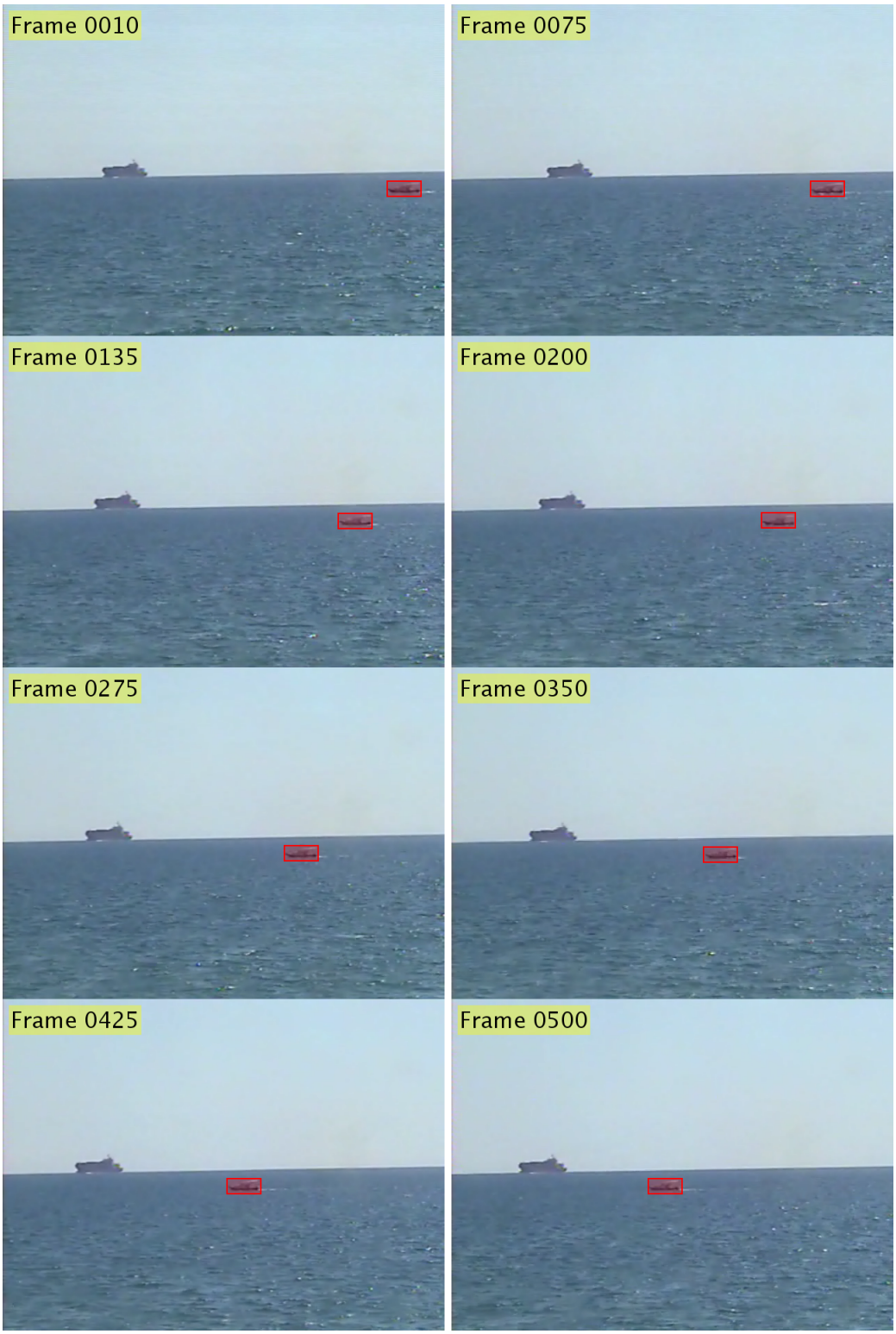


Figure 4.4: Sample target gates produced by the proposed tracking scheme in Video_Seq_2.

Figure 4.5: Sample target gates produced by the proposed tracking scheme in Video_Seq_3.

Figure 4.6: Sample target gates produced by the proposed tracking scheme in Video_Seq_4.

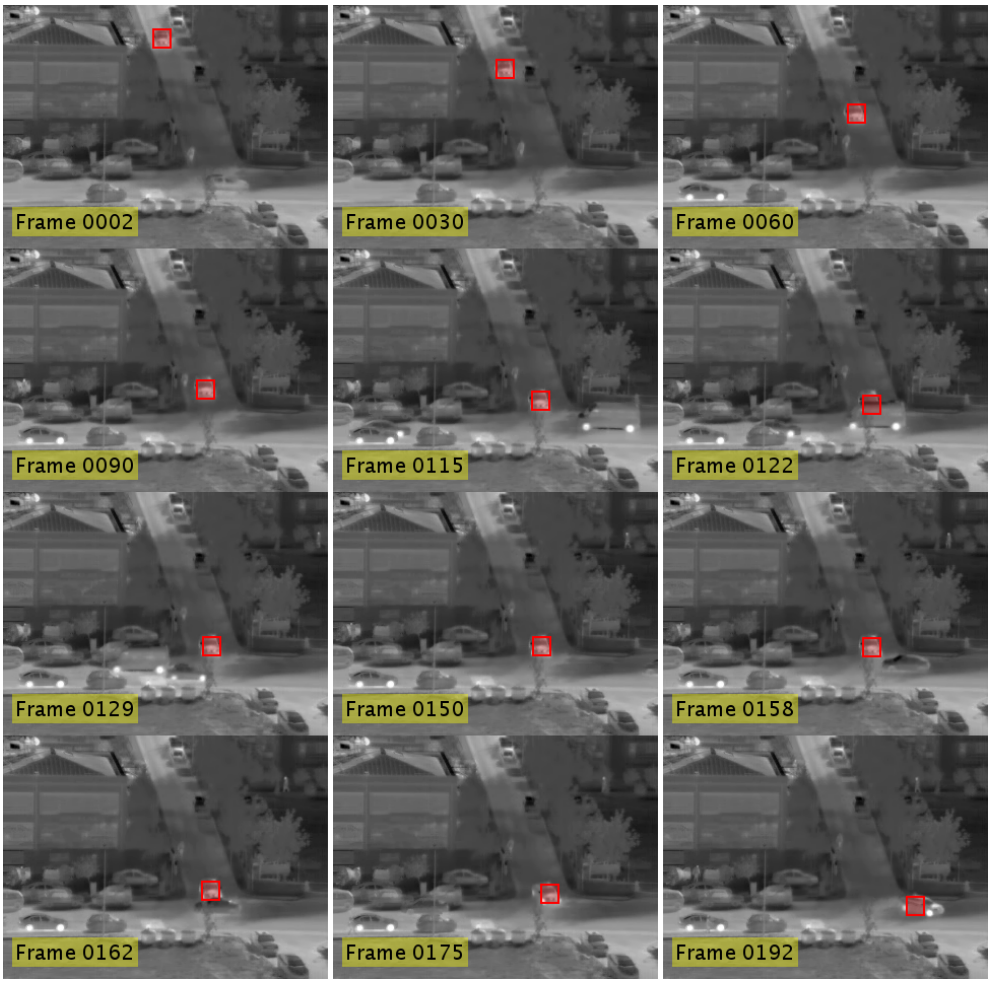


Figure 4.7: Sample target gates produced by the proposed tracking scheme in Video_Seq_5.

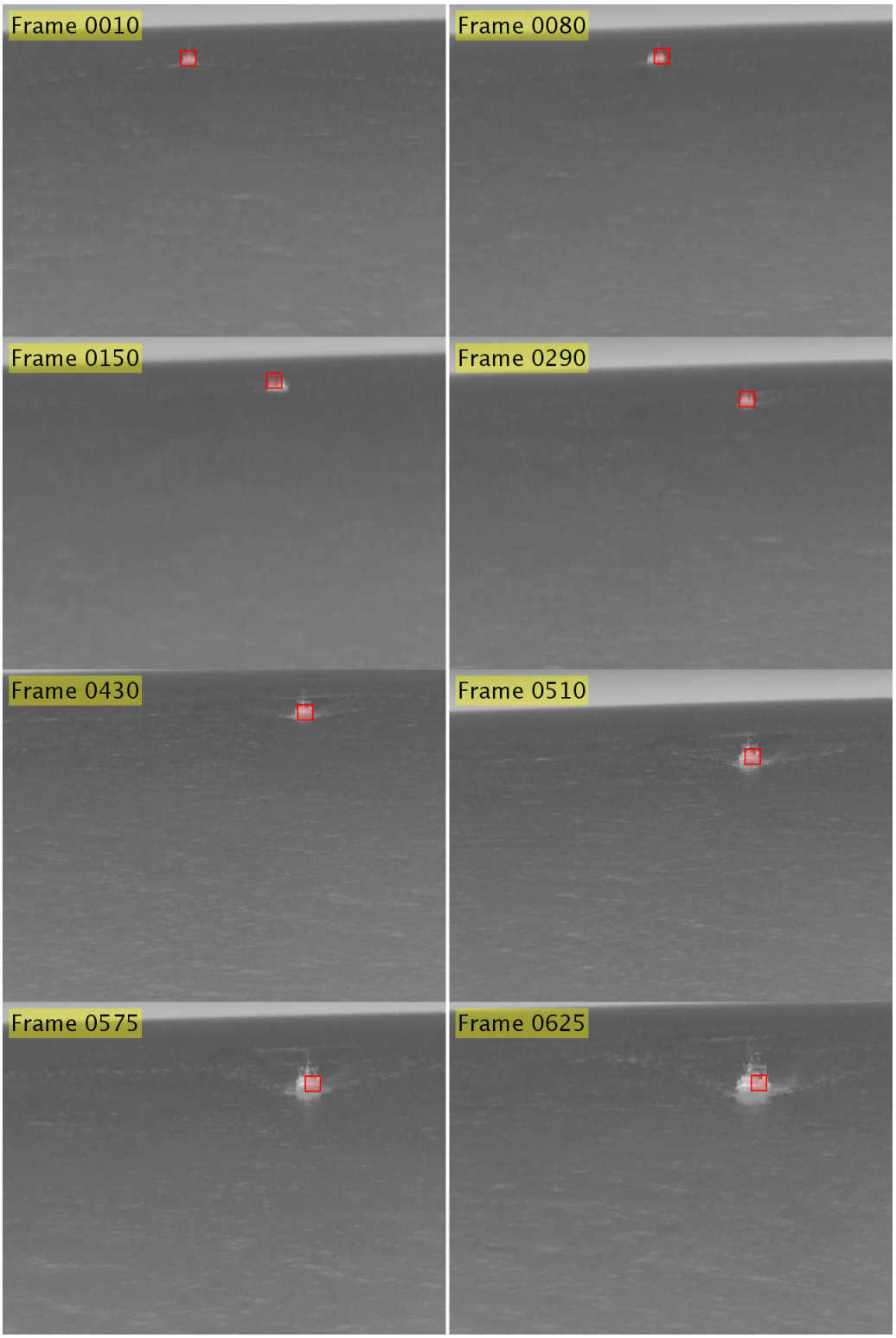
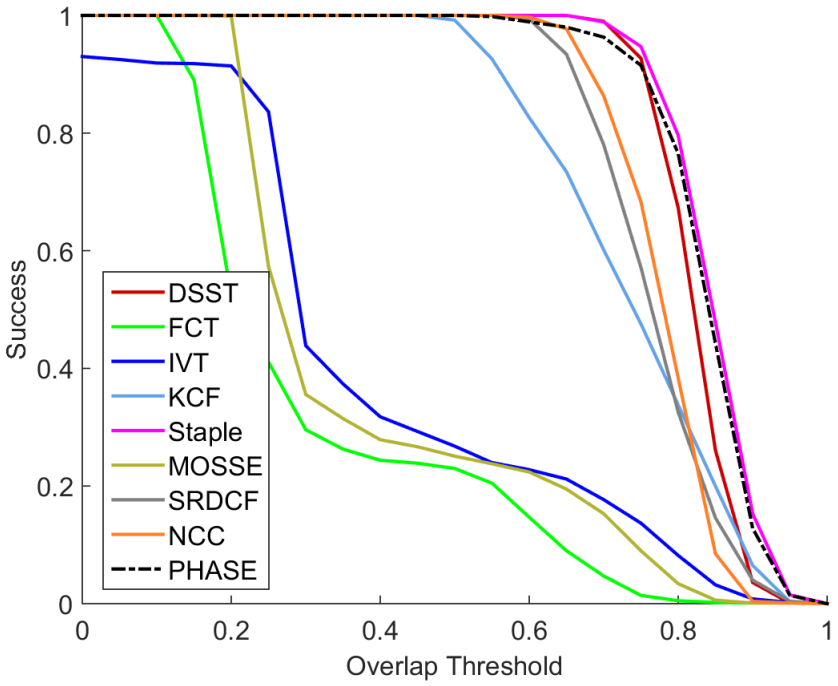


Figure 4.8: Sample target gates produced by the proposed tracking scheme in Video_Seq_6.

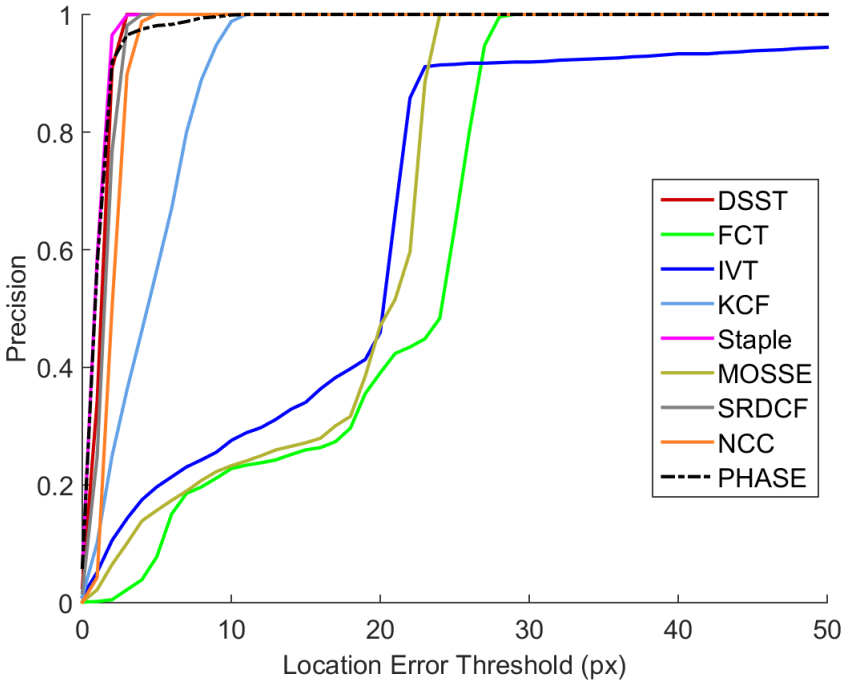
By looking at the results presented in Figures 4.3-4.8, one can observe that the proposed tracking scheme succeeds in tracking the target despite negative effects such as illumination changes, sudden camera movements, motion blur, and partial occlusions. In early frames of Video_Seq_1, the target is occluded by a moving speed boat that produces sea glints caused by the motor. Although the appearance of the target changes significantly, the phase-based technique has been able to track the target. In the last frames of the same video sequence, the occlusion source (sailboat) has more contrast than the target of interest. In other words, the clutter is more significant than the original target. However, the phase-based matching scheme preserves the lock on the desired target. Video_Seq_2 contains a complex background which changes rapidly due to sea-glints. However, the proposed phase-based tracker is able to track the target successfully throughout the frames. In Video_Seq_3, the target model changes rapidly due to illumination variations and motion blur in certain frames. In spite of these difficulties, the proposed tracking scheme does not lose the lock on the target throughout the entire scenario. Captured in an urban environment, Video_Seq_4 has a complex background which contains moving vehicles near the target of interest. Also, some of the frames in Video_Seq_4 are blurry due to the undesired rapid movements of the capturing device. Although the proposed scheme is able to track the target throughout the video, there are noticeable localization errors in some of the frames which cause performance loses in terms of objective metrics. Video_Seq_5 is an infrared image sequence captured in an urban environment. In Video_Seq_5, there are significant illumination changes and partial occlusions in certain frames. Moreover, the background of the video is complex due to different heat sources near the target of interest. In the last frames of the video, the target model is also exposed to sudden rotation changes caused by the maneuver of the target. The proposed tracker is able to track the target throughout the frames but noticeable tracking errors occur in some frames. Most of the frames in Video_Seq_6 are effected by Gaussian blur due to the undesired movements of the capturing device. Additionally, the scale of the target changes gradually since the target is approaching the capturing device during the video capture. Despite these difficulties, the proposed phase-based scheme is able to maintain the track on the target. However, the scale changes of the target result in degradations in tracking

performance of the proposed phase tracker.

We expand the experimental studies to compare the performance of the proposed tracking scheme with the baseline tracking techniques. The baseline tracking algorithms are tested with the same video sequences utilized for the proposed technique. The target gates produced by each tracker are stored for objective performance evaluation. By using the target gates obtained for each tracker, the success and precision rate performance measures are computed for each of the videos in the dataset. The performance plots for Video_Seq_1, Video_Seq_2, Video_Seq_3, Video_Seq_4, Video_Seq_5, and Video_Seq_6 are presented in Figure 4.9, Figure 4.10, Figure 4.11, Figure 4.12, Figure 4.13, and Figure 4.14 respectively.

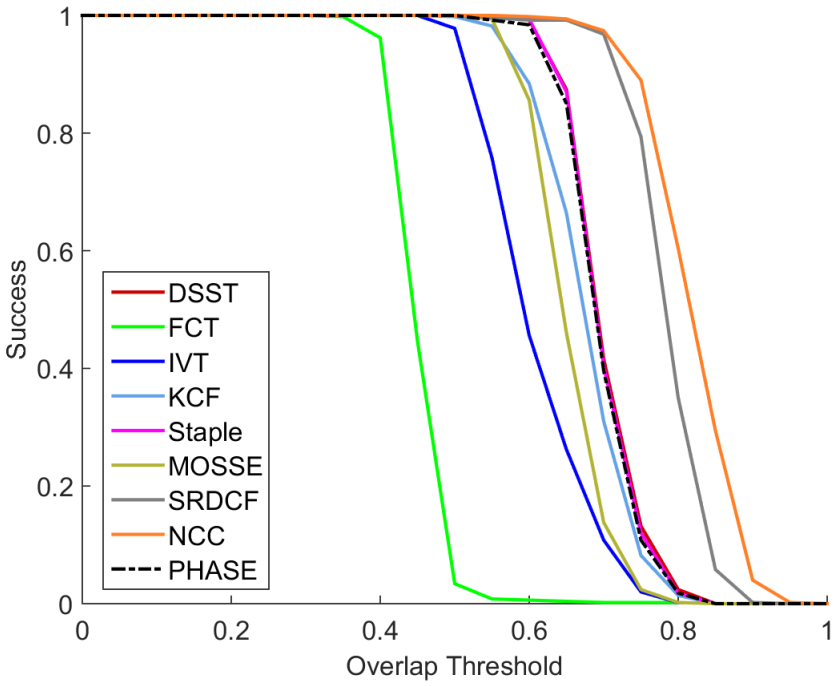


(a) Success Plot

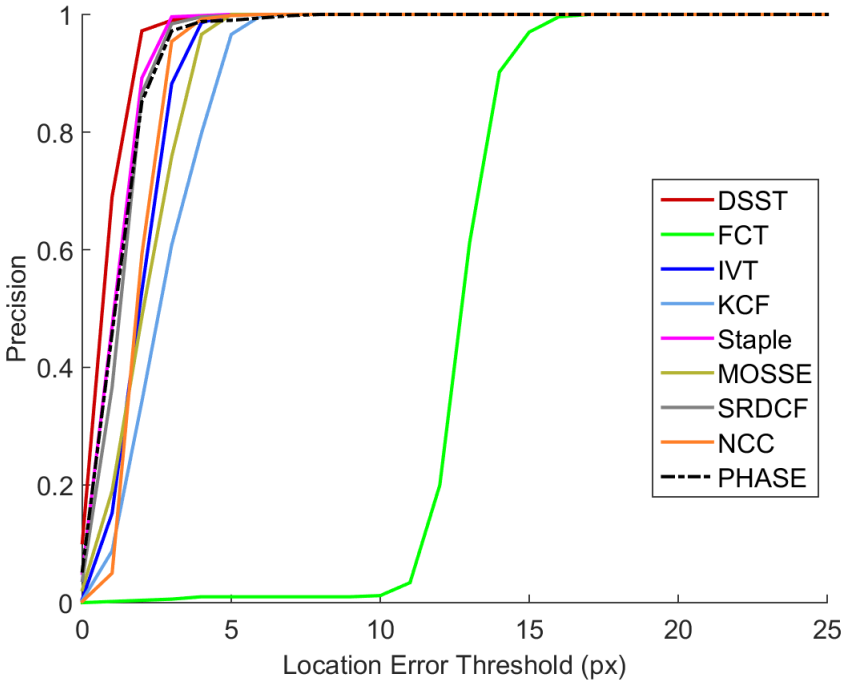


(b) Precision Plot

Figure 4.9: The success and precision plot of Video_Seq_1.

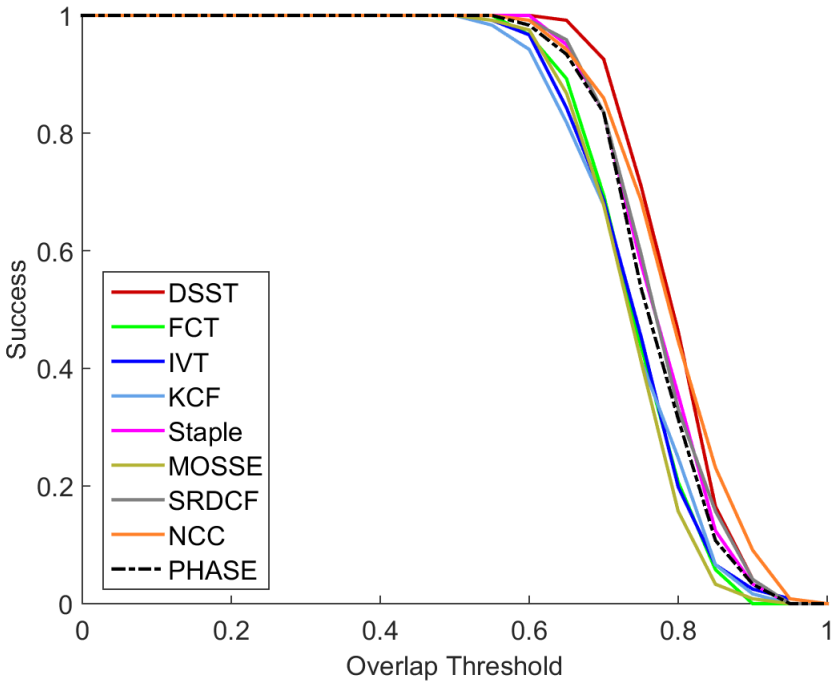


(a) Success Plot

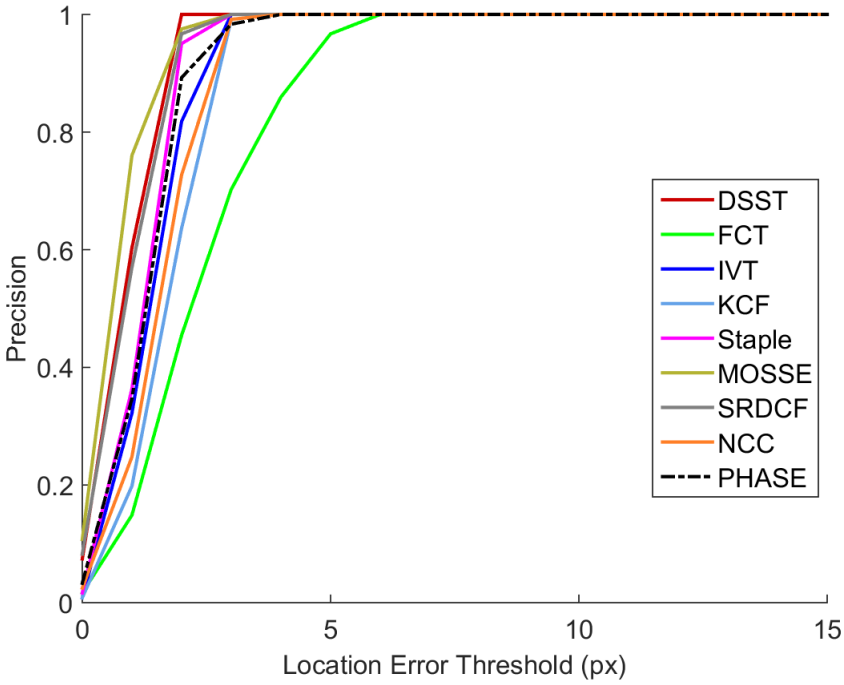


(b) Precision Plot

Figure 4.10: The success and precision plot of Video_Seq_2.

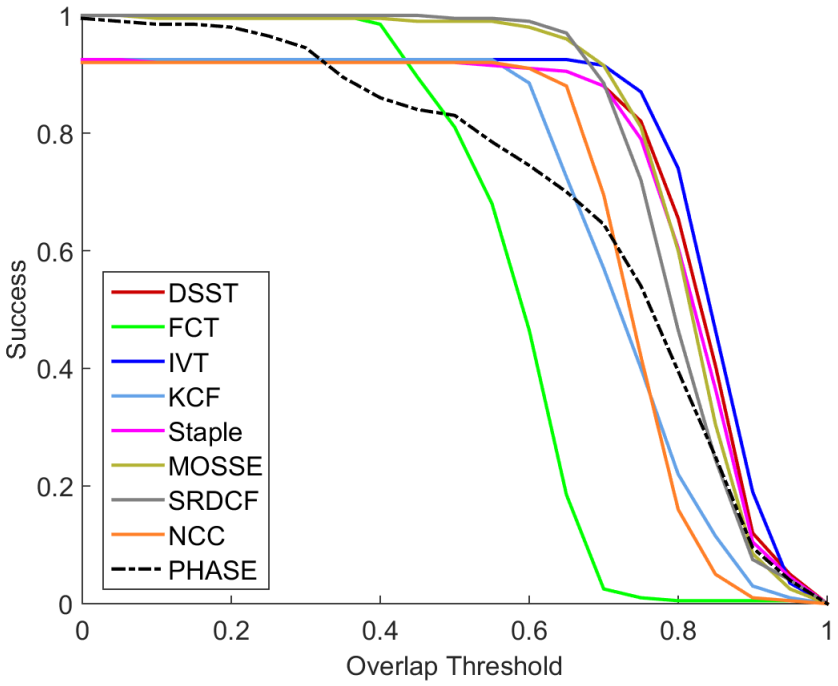


(a) Success Plot

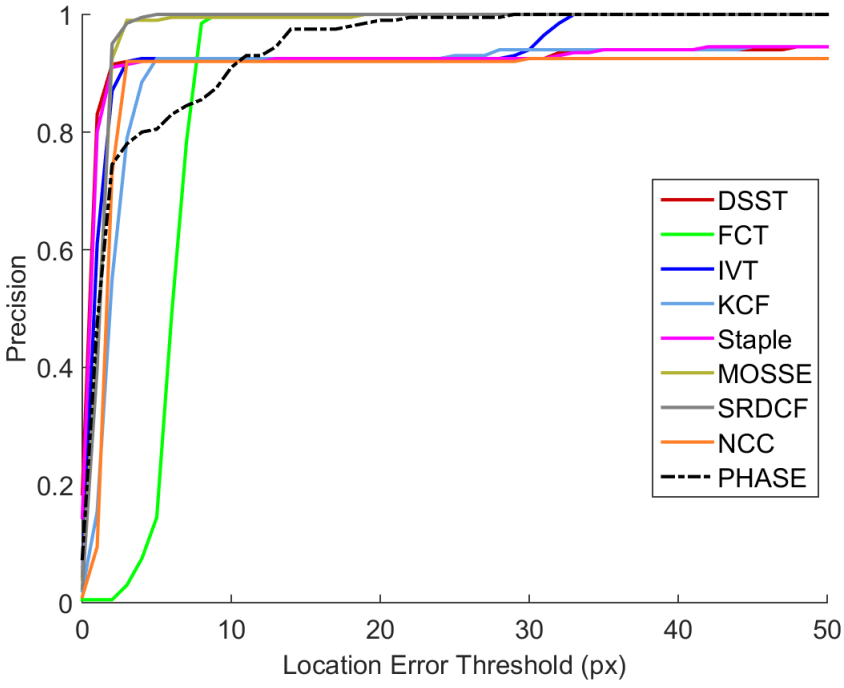


(b) Precision Plot

Figure 4.11: The success and precision plot of Video_Seq_3.

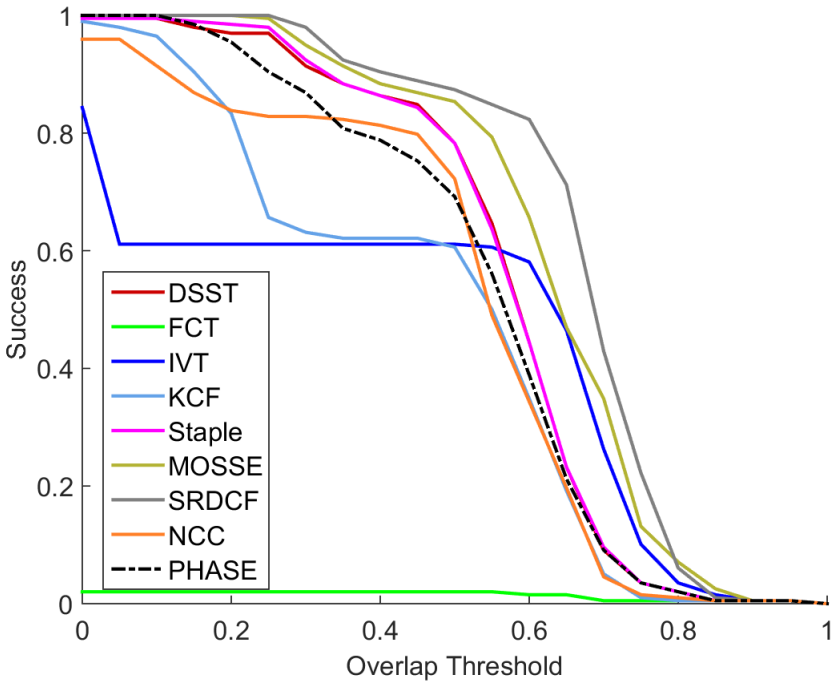


(a) Success Plot

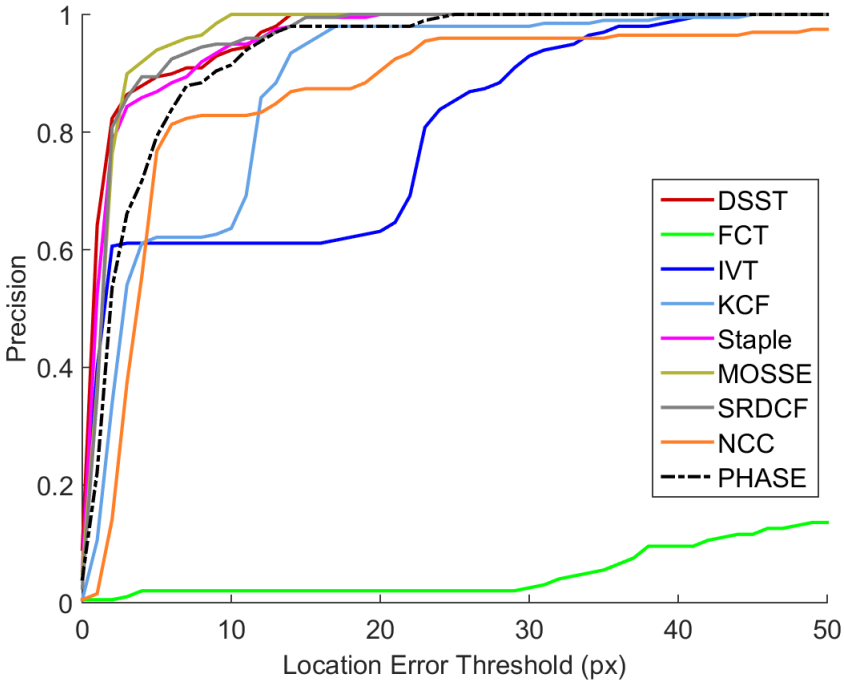


(b) Precision Plot

Figure 4.12: The success and precision plot of Video_Seq_4.

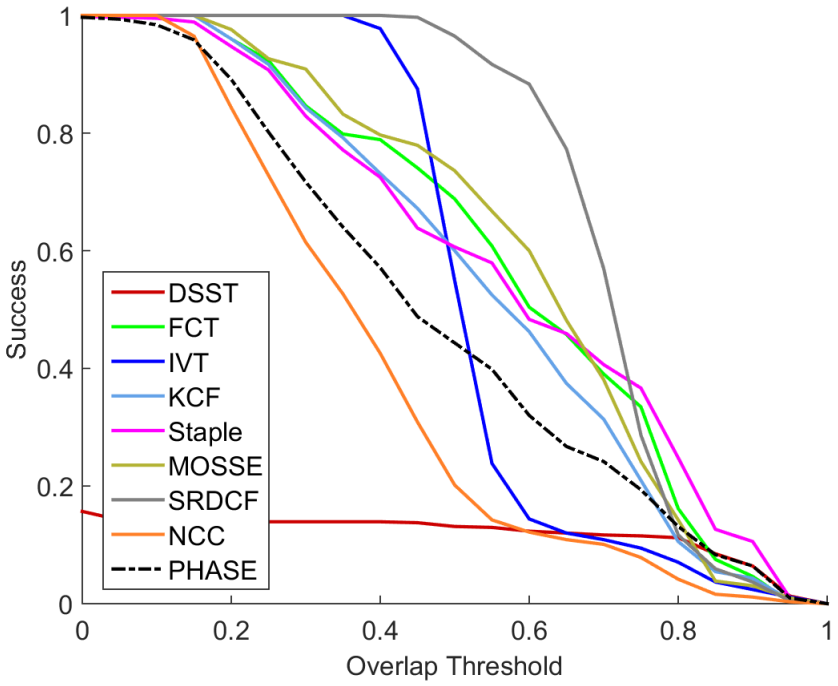


(a) Success Plot

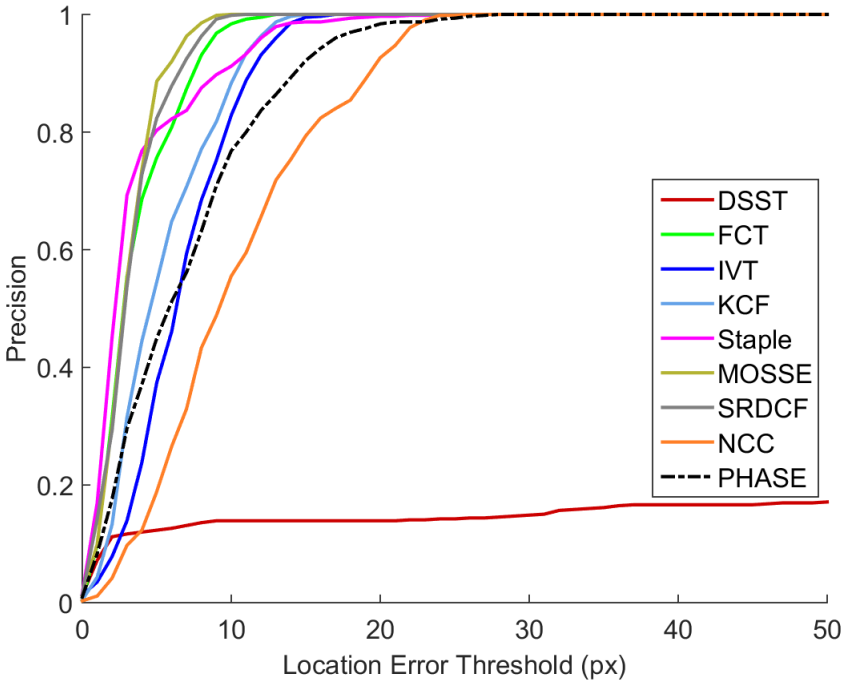


(b) Precision Plot

Figure 4.13: The success and precision plot of Video_Seq_5.



(a) Success Plot



(b) Precision Plot

Figure 4.14: The success and precision plot of Video_Seq_6.

By looking at the results presented in Figures 4.9-4.14, one can say the proposed tracking technique provides a consistent regime on the performance tests. In all of the video sequences, the proposed scheme tracks the target with acceptable success and precision rates. Since the video sequences contain negative factors such as illumination variations, motion blur, complex background and partial occlusions, the results provided by the proposed technique are quite satisfactory. Developed with the aim of reducing the computational complexity of the NCC algorithm, the proposed phase-based tracking scheme obtains comparable results with the NCC tracker. In some of the videos (Video_Seq_1 and Video_Seq_3), it even provides higher performance than NCC tracker. It is not surprising that some of the baseline tracking algorithms, which are sophisticated and computationally complex tracker frameworks, outperform the proposed tracking regime in the objective performance evaluations. Although it is simple by design, the proposed phase-based technique achieves satisfactory tracking accuracy while outperforming FCT, IVT and MOSSE trackers in the tests. Similar to the proposed phase-based tracking framework, the NCC tracker obtains satisfactory results in all of the scenarios.

In order to summarize the performance tests of the trackers, the overall performance of each tracker is evaluated by computing the ranking measures (AUC, TM, and LA) over the frames of all videos. The results obtained by the overall performance evaluations are presented in Table 4.2.

Table 4.2: The overall performance evaluations of video trackers.

	Ranking Method		
	Success		Precision
	AUC	TM	LA
DSST [208]	0.6055	0.7951	0.7860
FCT [209]	0.4186	0.9267	0.4439
IVT [210]	0.5138	0.9561	0.6507
KCF [211]	0.6545	0.9936	0.9350
Staple [212]	0.7235	0.9936	0.9694
MOSSE [213]	0.5476	1.0000	0.7093
SRDCF [214]	0.7487	1.0000	0.9958
NCC [201]	0.6639	0.9909	0.8760
Proposed	**0.6886**	**0.9989**	**0.9316**

The overall performance evaluations presented in Table 4.2 show that the proposed phase tracker provides comparable results with the baseline tracking frameworks. The proposed scheme obtains a satisfactory AUC result which is a sign of accurate localization of the target gate throughout the frames. Moreover, the phase tracker is able to track the target successfully by producing a very high TM score which corresponds to very few non-overlapping target gates produced by the tracker and the ground truth information. In the experiments, although the LA is evaluated at a tight error threshold of 10-pixels, the localization error of the target locations obtained by the proposed tracker are below 10 pixels in approximately 93% of the all video frames.

As an additional experiment, the proposed phase tracker is compared with the NCC tracker in terms of computational complexity. The main aim of this experiment is to evaluate and compare the computation times of the proposed tracker and the NCC tracker with respect to varying dimensions of the target model. The system specifications should be taken into account when discussing computation time. The proposed tracker framework is implemented at Matlab

environment on a computer containing Intel(R) Core(TM) i5 3.40 GHz processor, 8 GB RAM running on Microsoft Windows 7 operating system. Starting from a target model of size 32×32, the width and height of the target model are doubled at each step until we obtain a target model of size 512×512. At each step of the experiment, average time required to process a single frame by the trackers are computed over 100 frames and the maximum number of frames which can be processed within one second is estimated. The results are presented in the following table:

Table 4.3: Average time required to process a single frame by the proposed and NCC tracking algorithms. Here, the computation time of the look-up-table (LUT) extraction process for the proposed framework is included in the results.

Target Size	NCC	Proposed
32×32	0.0023 s (≈435 fps)	0.0015 s (≈667 fps)
64×64	0.0046 s (≈217 fps)	0.0034 s (≈294 fps)
128×128	0.0124 s (≈81 fps)	0.0088 s (≈114 fps)
256×256	0.0684 s (≈15 fps)	0.0268 s (≈37 fps)
512×512	0.2222 s (≈5 fps)	0.0867 s (≈12 fps)

By looking at the result presented in Table 4.3, one can conclude that the proposed phase-based technique is computationally more efficient compared to the classical NCC framework. Even for the largest target size (512×512), the proposed scheme performs in the limit of real-time requirements. As the complexity increases due to the target size, the processing time of the NCC algorithm grows exponentially. However, the processing time of the proposed phase-based tracker has a more acceptable regime compared to the NCC tracker.

4.4 Summary

In this chapter, image phase information is utilized to reduce the computational complexity of the template matching based tracking frameworks. The proposed

target tracker scheme provides a significant reduction in computation without sacrificing the tracking accuracy. Derived from the NCC calculation, the proposed framework can be an alternative to NCC-based trackers due its low computational cost, fast response and its ability to operate within real-time requirements. In the experiments, the proposed framework has obtained comparable results with the DSST [208], KCF [211], Staple [212], SRDCF [214], and NCC [201] trackers while outperforming the FCT [209], IVT [210], and MOSSE [213] tracking techniques. Although the proposed scheme is simple by design, it has the potential to cooperate with sophisticated object tracking frameworks which may lead to an efficient implementation in terms of both tracking performance and computational efficiency.

Chapter 5

Conclusions

In this thesis, image phase information is utilized for various image processing applications in order to improve the performance of a full-reference IQA framework, to develop an image contrast enhancement technique and to reduce the computational complexity of the NCC-based visual object tracking scheme for efficiency.

The 2D mel-cepstrum based image representation framework is enhanced with image phase information to obtain a superior image representation scheme for the image quality assessment problem. To integrate the image phase information with the 2D mel-ceptrum features, a phase-unwrapping algorithm is performed and a non-uniform grid is applied to represent neighboring frequency components with a common phase term. In this way, a reduced phase representation is obtained and it is combined with the 2D mel-cepstrum features by using appropriate weights which are obtained by minimizing the level of energy between the phase and cepstral coefficients. The new-phase based representation scheme is called "2D complex mel-cepstum" [28] and it is fed into a machine learning framework, SVR, to generate a mapping function between the feature matrices and the subjective scores. According to the experimental results, the new phase-based representation scheme has been found to be more effective than the classical 2D mel-ceptrum features in the assessment of image quality.

A Fourier domain contrast enhancement framework [33] is developed to improve the contrast and visibility of the microscopy images while suppressing undesired background information. Our algorithm's design is inspired by the PCM which was invented by F. Zernike [34]. In PCM, minute variations in light rays are translated into changes in magnitude which can be observed as contrast differences. Inspired by the PCM, we applied the idea of phase-magnitude relation to the Fourier transform of the image, although the definition of phase in our case is completely different from the phase in PCM. Operating in the Fourier domain, the proposed scheme makes use of the image phase information to transform phase variations into amplitude changes. In order to determine phase variations between the neighboring frequency components, phase gradient is computed in both vertical and horizontal directions. The frequency components which have significant phase transitions are amplified by factors which are proportional to the strength of the corresponding phase transitions. The frequency components where the phase transitions are negligible remain unchanged. The proposed algorithm has several parameters and these parameters are determined by an optimization framework based on TV maximization. However, using a maximization scheme based on ordinary TV may lead the algorithm to noisy and artificial results. Therefore, the concept of selective variation is introduced which performs maximization on the "busy" regions of the image. To best of our knowledge, the $\ell 1$-norm maximization of a cost functional within local regions is first proposed by our approach. Experimental results show that the proposed algorithm provides higher contrast enhancement performance while suppressing the background more effectively than the baseline contrast enhancement techniques.

Normalized cross-correlation (NCC) is a widely used image matching technique in visual tracking applications. However, when the target model becomes large, the computational complexity of the NCC scheme increases significantly. In this thesis, starting from the derivation of the NCC algorithm, the tracking problem is simplified to a phase minimization solution under certain constraints. By utilizing the phase information, the proposed tracking scheme makes use of the convolution theorem and shift invariance property of the Fourier Transform (FT). Moreover, the phase-based technique relies on the following fact: Most of

the moving objects result in translational shifts in the image domain in a small time window. Since the translational shifts in the image domain correspond to phase variations in the Fourier domain, the target region is determined only by checking the phase information. By forming look up tables which contain phase variations corresponding to each translational shift, repetitive FFT calculations are prevented and an efficient implementation is obtained. The experimental results indicate that the proposed phase-based tracking scheme obtains comparable performance with the baseline methods which are more sophisticated frameworks designed for competing in tracking benchmarks. Also, the proposed scheme introduces a significant decrease in the computational complexity which is our primary objective in this study.

Future work may include the utilization of phase based techniques to other image processing and computer vision applications. Since image phase contains more descriptive information compared to the magnitude, magnitude-based techniques can be enhanced or totally replaced with phase-based approaches in order to obtain more efficient implementations. Also, the phase-based object tracking framework can be adapted to on-plane rotations by making use of a log-polar transform. In this way, the rotational changes in object appearance can be converted into translations in the image domain which will correspond to phase changes in the frequency domain.

Bibliography

[1] A. V. Oppenheim, M. H. Hayes, and J. S. Lim, "Iterative procedures for signal reconstruction from phase," *Proc. SPIE*, vol. 0231, pp. 121–129, 1980.

[2] F. Branchitta, M. Diani, G. Corsini, and A. Porta, "Dynamic-range compression and contrast enhancement in infrared imaging systems," *Optical Engineering*, vol. 47, no. 7, pp. 076401–076401, 2008.

[3] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems IV* (P. S. Heckbert, ed.), pp. 474–485, San Diego, CA, USA: Academic Press Professional, Inc., 1994.

[4] I. Belykh, *Homomorphic Filtering for Radiographic Image Contrast Enhancement and Artifacts Elimination*, pp. 423–430. Springer International Publishing, 2015.

[5] S. Weith-Glushko and C. Salvaggio, "Quantitative analysis of infrared contrast enhancement algorithms," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XVIII*, vol. 6543, p. 65430S, International Society for Optics and Photonics, 2007.

[6] Z.-u. Rahman, D. J. Jobson, and G. A. Woodell, "Retinex processing for automatic image enhancement," *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 100–110, 2004.

[7] A. Oppenheim and J. Lim, "The importance of phase in signals," *Proceedings of the IEEE*, vol. 69, no. 5, pp. 529–541, 1981.

[8] M. Hayes, J. Lim, and A. Oppenheim, "Signal reconstruction from phase or magnitude," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 6, pp. 672–680, 1980.

[9] A. E. Cetin and R. Ansari, "Convolution-based framework for signal recovery and applications," *Journal of Optical Society of America (JOSA) A*, vol. 5, no. 8, pp. 1193–1200, 1988.

[10] G. Shi, M. M. Shanechi, and P. Aarabi, "On the importance of phase in human speech recognition," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, pp. 1867–1874, Sept 2006.

[11] L. D. Alsteris and K. K. Paliwal, "Short-time phase spectrum in speech processing: A review and some experimental results," *Digital Signal Processing*, vol. 17, no. 3, pp. 578 – 616, 2007.

[12] K. Paliwal, K. Wójcicki, and B. Shannon, "The importance of phase in speech enhancement," *Speech Communication*, vol. 53, no. 4, pp. 465 – 494, 2011.

[13] T. Kleinschmidt, S. Sridharan, and M. Mason, "The use of phase in complex spectrum subtraction for robust speech recognition," *Computer Speech & Language*, vol. 25, no. 3, pp. 585 – 600, 2011.

[14] E. Loweimi, S. M. Ahadi, and T. Drugman, "A new phase-based feature representation for robust speech recognition," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 7155–7159, May 2013.

[15] P. Mowlaee and R. Saeidi, "On phase importance in parameter estimation in single-channel speech enhancement," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 7462–7466, May 2013.

[16] D. Wang and J. Lim, "The unimportance of phase in speech enhancement," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 30, pp. 679–681, Aug 1982.

[17] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proceedings of the IEEE*, vol. 67, pp. 1586–1604, Dec 1979.

[18] T. Huang, J. Burnett, and A. Deczky, "The importance of phase in image processing filters," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 23, pp. 529–542, Dec 1975.

[19] A. Oppenheim, J. Lim, G. Kopec, and S. Pohlig, "Phase in speech and pictures," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'79.*, vol. 4, pp. 632–637, IEEE, 1979.

[20] J. J. K. O. Ruanaidh, W. J. Dowling, and F. M. Boland, "Phase watermarking of digital images," in *Proceedings of 3rd IEEE International Conference on Image Processing*, vol. 3, pp. 239–242 vol.3, Sep 1996.

[21] K. Miyazawa, K. Ito, T. Aoki, K. Kobayashi, and H. Nakajima, "An effective approach for iris recognition using phase-based image matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 1741–1756, Oct 2008.

[22] C. Guo, Q. Ma, and L. Zhang, "Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, June 2008.

[23] D. Youla, "Generalized image restoration by the method of alternating orthogonal projections," *IEEE Transactions on Circuits and Systems*, vol. 25, pp. 694–702, September 1978.

[24] I. Bartolini, P. Ciaccia, and M. Patella, "Warp: accurate retrieval of shapes using phase of fourier descriptors and time warping distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 142–147, Jan 2005.

[25] C. Kuglin and D. Hines, "The phase correlation image alignment method," *Proc. Int. Conf. on Cybernetics and Society, 1975*, pp. 163–165, 1975.

[26] K. Takita, T. Aoki, Y. Sasaki, T. Higuchi, and K. Kobayashi, "High-accuracy subpixel image registration based on phase-only correlation," *IEICE transactions on fundamentals of electronics, communications and computer sciences*, vol. 86, no. 8, pp. 1925–1934, 2003.

[27] X. Bai, Y. Fang, W. Lin, L. Wang, and B. F. Ju, "Saliency-based defect detection in industrial images by using phase spectrum," *IEEE Transactions on Industrial Informatics*, vol. 10, pp. 2135–2145, Nov 2014.

[28] S. Cakir and A. E. Cetin, "Image quality assessment using two-dimensional complex mel-cepstrum," *Journal of Electronic Imaging*, vol. 25, no. 6, pp. 061604–061604, 2016.

[29] S. Cakir and A. E. Cetin, "Mel-cepstral feature extraction methods for image representation," *Optical Engineering*, vol. 49, no. 9, pp. 097004–097004–8, 2010.

[30] S. Cakir, "Cepstral methods for image feature extraction.," Master's thesis, Bilkent University, Department of Electrical and Electronics Engineering, Ankara, Turkey, 2010.

[31] M. Narwaria, W. Lin, and A. E. Cetin, "Scalable image quality assessment with 2d mel-cepstrum and machine learning approach," *Pattern Recognition*, vol. 45, no. 1, pp. 299 – 313, 2012.

[32] M. Narwaria, W. Lin, I. McLoughlin, S. Emmanuel, and L.-T. Chia, "Fourier transform-based scalable image quality measure," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3364–3377, 2012.

[33] S. Cakir, D. C. Kahraman, R. C. Atalay, and A. E. Cetin, "Contrast enhancement of microscopy images using image phase information," *IEEE Access*, Accepted for publication, January 2018.

[34] F. Zernike, "Phase contrast, a new method for the microscopic observation of transparent objects," *Physica*, vol. 9, no. 7, pp. 686 – 698, 1942.

[35] U. Engelke and H.-J. Zepernick, "Perceptual-based quality metrics for image and video services: A survey," in *3rd EuroNGI Conference on Next Generation Internet Networks*, pp. 190–197, 2007.

[36] "Recommendation ITU-R BT.500-11 Methodology for the subjective assessment of the quality of television pictures," 2002.

[37] I. van der Linde and R. M. Doe, "Influence of affective image content on subjective quality assessment," *Journal of Optical Society of America (JOSA) A*, vol. 29, no. 9, pp. 1948–1955, 2012.

[38] S. Winkler and P. Mohandas, "The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics," *IEEE Transactions on Broadcasting*, vol. 54, no. 3, pp. 660–668, 2008.

[39] A. Eskicioglu and P. Fisher, "Image quality measures and their performance," *IEEE Transactions on Communications*, vol. 43, no. 12, pp. 2959–2965, 1995.

[40] S. Karunasekera and N. Kingsbury, "A distortion measure for blocking artifacts in images based on human visual sensitivity," *IEEE Transactions on Image Processing*, vol. 4, no. 6, pp. 713–724, 1995.

[41] B. Girod, "Psychovisual aspects of image processing: What's wrong with mean squared error?," in *Proceedings of the Seventh Workshop on Multidimensional Signal Processing*, pp. P.2–P.2, Sep 1991.

[42] R. Dosselmann and X. D. Yang, "A comprehensive assessment of the structural similarity index," *Signal, Image and Video Processing*, vol. 5, no. 1, pp. 81–91, 2009.

[43] M. P. Eckert and A. P. Bradley, "Perceptual quality metrics applied to still image compression," *Signal Processing*, vol. 70, no. 3, pp. 177 – 200, 1998.

[44] T. N. Pappas and R. J. Safranek, "Perceptual criteria for image quality evaluation," in *Handbook of Image and Video Processing*, pp. 669–684, Academic Press, 2000.

[45] F. Lukas and Z. Budrikis, "Picture quality prediction based on a visual model," *IEEE Transactions on Communications*, vol. 30, no. 7, pp. 1679–1692, 1982.

[46] S. Winkler, "A perceptual distortion metric for digital color video," in *Proceedings of SPIE 3644, Human Vision and Electronic Imaging IV*, pp. 175–184, 1999.

[47] X. Zhang and B. A. Wandell, "Color image fidelity metrics evaluated using image distortion maps," *Signal Processing*, vol. 70, no. 3, pp. 201–214, 1998.

[48] W. Lin, L. Dong, and P. Xue, "Visual distortion gauge based on discrimination of noticeable contrast changes," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 7, pp. 900–909, 2005.

[49] A. B. Watson, J. Hu, and J. F. M. Iii, "DVQ: A digital video quality metric based on human vision," *Journal of Electronic Imaging*, vol. 10, no. 1, pp. 20–29, 2001.

[50] P. Burt and E. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 532–540, 1983.

[51] E. Simoncelli, W. Freeman, E. Adelson, and D. Heeger, "Shiftable multiscale transforms," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 587–607, 1992.

[52] R. Fredericksen and R. Hess, "Estimating multiple temporal mechanisms in human vision," *Vision Research*, vol. 38, no. 7, pp. 1023 – 1040, 1998.

[53] S. W. Kuffler, "Discharge patterns and functional organization of the mammalian retina," *J Neurophysiol*, vol. 16, pp. 37–68, Jan. 1953.

[54] E. Peli, "Contrast in complex images," *Journal of Optical Society of America (JOSA) A*, vol. 7, no. 10, pp. 2032–2040, 1990.

[55] Y.-K. Lai and C.-C. Kuo, "A haar wavelet approach to compressed image quality measurement," *Journal of Visual Communication and Image Representation*, vol. 11, no. 1, pp. 17 – 40, 2000.

[56] Z. Lu, W. Lin, X. Yang, E. Ong, and S. Yao, "Modeling visual attention's modulatory aftereffects on visual sensitivity and quality evaluation," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1928–1942, 2005.

[57] H. Wu and M. Yuen, "A generalized block-edge impairment metric for video coding," *IEEE Signal Processing Letters*, vol. 4, no. 11, pp. 317–320, 1997.

[58] L. Meesters and J.-B. Martens, "A single-ended blockiness measure for jpeg-coded images," *Signal Processing*, vol. 82, no. 3, pp. 369 – 387, 2002.

[59] T. Vlachos, "Detection of blocking artifacts in compressed video," *Electronics Letters*, vol. 36, no. 13, pp. 1106–1108, 2000.

[60] J. Caviedes and F. Oberti, "A new sharpness metric based on local kurtosis, edge and energy information," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 147 – 161, 2004.

[61] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," in *Proceedings of IEEE International Conference on Image Processing*, vol. 3, pp. III–57–III–60, 2002.

[62] W. Lin, *Computational Models for Just-Noticeable Difference (Ch. 9), Digital Video Image Quality and Perceptual Coding*. CRC Press, 2005.

[63] M. Ramasubramanian, S. N. Pattanaik, and D. P. Greenberg, "A perceptually based physical error metric for realistic image synthesis," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, (New York, NY, USA), pp. 73–82, ACM Press/Addison-Wesley Publishing Co., 1999.

[64] A. Bradley, "A wavelet visible difference predictor," *IEEE Transactions on Image Processing*, vol. 8, no. 5, pp. 717–730, 1999.

[65] H. E. Pashler, *The Psychology of Attention*. Cambridge, MA: MIT Press, 1998.

[66] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[67] S. Wolf, "Measuring the end-to-end performance of digital video systems," *IEEE Transactions on Broadcasting*, vol. 43, no. 3, pp. 320–328, 1997.

[68] H. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

[69] S. Gabarda and G. Cristóbal, "No-reference image quality assessment through the von Mises distribution," *Journal of Optical Society of America (JOSA) A*, vol. 29, no. 10, pp. 2058–2066, 2012.

[70] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[71] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *Proceedings of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 1398–1402 Vol.2, 2003.

[72] Z. Wang and A. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2002.

[73] H. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

[74] N. Damera-Venkata, T. Kite, W. Geisler, B. Evans, and A. Bovik, "Image quality assessment based on a degradation model," *IEEE Transactions on Image Processing*, vol. 9, no. 4, pp. 636–650, 2000.

[75] D. Chandler and S. Hemami, "VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images," *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2284–2298, 2007.

[76] M. Sendashonga and F. Lebeau, "Low complexity image quality assessment using frequency domain transforms," in *Proceedings of IEEE International Conference on Image Processing*, pp. 385–388, 2006.

[77] C.-Y. Wee, R. Paramesran, R. Mukundan, and X. Jiang, "Image quality assessment by discrete orthogonal moments," *Pattern Recognition*, vol. 43, no. 12, pp. 4055 – 4068, 2010.

[78] L. Junfeng, D. Wenzhan, P. Haipeng, and W. Huijiao, "Image quality assessment based on nonsubsampled contourlet transform," in *29th Chinese Control Conference*, pp. 2665–2670, 2010.

[79] Z. Haddad, A. Beghdadi, A. Serir, and A. Mokraoui, "Image quality assessment based on wave atoms transform," in *Proceedings of IEEE International Conference on Image Processing*, pp. 305–308, 2010.

[80] A. Shnayderman, A. Gusev, and A. M. Eskicioglu, "An SVD-based grayscale image quality measure for local and global assessment," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 422–429, 2006.

[81] M. Narwaria and W. Lin, "Objective image quality assessment based on support vector regression," *IEEE Transactions on Neural Networks*, vol. 21, no. 3, pp. 515–519, 2010.

[82] A. Shnayderman, A. Gusev, and A. Eskicioglu, "An svd-based grayscale image quality measure for local and global assessment," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 422–429, 2006.

[83] M. Miyahara, K. Kotani, and V. Algazi, "Objective picture quality scale (pqs) for image coding," *IEEE Transactions on Communications*, vol. 46, no. 9, pp. 1215–1226, 1998.

[84] P. Carrai, I. Heynderickz, P. Gastaldo, and R. Zunino, "Image quality assessment by using neural networks," in *Proceedings of IEEE International Symposium on Circuits and Systems*, vol. 5, pp. V–253–V–256, 2002.

[85] A. Bouzerdoum, A. Havstad, and A. Beghdadi, "Image quality assessment using a neural network approach," in *Proceedings of IEEE International Symposium on Signal Processing and Information Technology*, pp. 330–333, 2004.

[86] W. Ding, Y. Tong, Q. Zhang, and D. Yang, "Image and video quality assessment using neural network and svm," *Tsinghua Science and Technology*, vol. 13, no. 1, pp. 112–116, 2008.

[87] W. Lin and M. Narwaria, "Perceptual image quality assessment: recent progress and trends," in *Proceedings of SPIE 7744, Visual Communications and Image Processing*, pp. 1–9, 2010.

[88] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25* (F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds.), pp. 1097–1105, Curran Associates, Inc., 2012.

[89] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," in *BMVC 2014 - Proceedings of the British Machine Vision Conference 2014*, 05 2014.

[90] J. Kim, H. Zeng, D. Ghadiyaram, S. Lee, L. Zhang, and A. C. Bovik, "Deep convolutional neural models for picture-quality prediction: Challenges and solutions to data-driven image quality assessment," *IEEE Signal Processing Magazine*, vol. 34, pp. 130–141, Nov 2017.

[91] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," in *Computer Science Department, University of Toronto, Tech. Rep*, vol. 1, 01 2009.

[92] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, June 2009.

[93] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 1798–1828, Aug 2013.

[94] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1733–1740, June 2014.

[95] Y. Li, L. M. Po, L. Feng, and F. Yuan, "No-reference image quality assessment with deep convolutional neural networks," in *2016 IEEE International Conference on Digital Signal Processing (DSP)*, pp. 685–689, Oct 2016.

[96] S. Bosse, D. Maniry, T. Wiegand, and W. Samek, "A deep neural network for image quality assessment," in *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 3773–3777, Sept 2016.

[97] J. Kim and S. Lee, "Fully deep blind image quality predictor," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, pp. 206–220, Feb 2017.

[98] Y. Liang, J. Wang, X. Wan, Y. Gong, and N. Zheng, "Image quality assessment using similar scene as reference," in *Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part V* (B. Leibe, J. Matas, N. Sebe, and M. Welling, eds.), (Cham), pp. 3–18, Springer International Publishing, 2016.

[99] F. Gao, Y. Wang, P. Li, M. Tan, J. Yu, and Y. Zhu, "DeepSim: Deep similarity for image quality assessment," *Neurocomputing*, vol. 257, no. Supplement C, pp. 104 – 114, 2017.

[100] J. Kim and S. Lee, "Deep learning of human visual sensitivity in image quality assessment framework," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[101] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 4, pp. 357–366, 1980.

[102] H. Yang, D. Huang, and L. Cai, "Perceptually weighted mel-cepstrum analysis of speech based on psychoacoustic model," *IEICE Transactions on Information and Systems*, vol. E89-D, no. 12, pp. 2998–3001, 2006.

[103] V. Tyagi and C. Wellekens, "On desensitizing the mel-cepstrum to spurious spectral components for robust speech recognition," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 529 – 532, 2005.

[104] T. Kitamura and S. Takei, "Speaker recognition model using two-dimensional mel-cepstrum and predictive neural network," in *Proceedings of Fourth International Conference on Spoken Language*, vol. 3, pp. 1772 –1775 vol.3, 1996.

[105] B. U. Toreyin and A. E. Cetin, "Shadow detection using 2D cepstrum," in *Acquisition, Tracking, Pointing, and Laser Systems Technologies XXIII*, p. 733809, SPIE, 2009.

[106] J. K. Lee, M. Kabrisky, M. E. Oxley, S. K. Rogers, and D. W. Ruck, "The complex cepstrum applied to two-dimensional images," *Pattern Recognition*, vol. 26, no. 10, pp. 1579 – 1592, 1993.

[107] Y. Yeshurun and E. Schwartz, "Cepstral filtering on a columnar image architecture: a fast algorithm for binocular stereo segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 759–767, 1989.

[108] P. Skurowski and A. Gruca, "Image quality assessment using phase spectrum correlation," in *Proceedings of the International Conference on Computer Vision and Graphics: Revised Papers*, ICCVG 2008, pp. 80–89, Springer-Verlag, 2009.

[109] R. Schafer, *Echo removal by discrete generalized linear filtering*. PhD thesis, MIT, 1968.

[110] J. Tribolet, "A new phase unwrapping algorithm," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 25, no. 2, pp. 170–177, 1977.

[111] L. Moisan, "Periodic plus smooth image decomposition," *Journal of Mathematical Imaging and Vision*, vol. 39, no. 2, pp. 161–179, 2011.

[112] D. C. Ghiglia and L. A. Romero, "Minimum lp-norm two-dimensional phase unwrapping," *Journal of Optical Society of America (JOSA) A*, vol. 13, no. 10, pp. 1999–2013, 1996.

[113] K. Steiglitz and B. Dickinson, "Phase unwrapping by factorization," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 30, no. 6, pp. 984–991, 1982.

[114] G. Fornaro, G. Franceschetti, R. Lanari, E. Sansosti, and M. Tesauro, "Global and local phase-unwrapping techniques: a comparison," *Journal of Optical Society of America (JOSA) A*, vol. 14, no. 10, pp. 2702–2708, 1997.

[115] M. Costantini, "A novel phase unwrapping method based on network programming," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 3, pp. 813–821, 1998.

[116] J. Bioucas-Dias and G. Valadao, "Phase unwrapping via graph cuts," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 698–709, 2007.

[117] B. Marendic, Y. Yang, and H. Stark, "Phase unwrapping using an extrapolation-projection algorithm," *Journal of Optical Society of America (JOSA) A*, vol. 23, no. 8, pp. 1846–1855, 2006.

[118] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, p. 011006, 2010.

[119] C.-C. Chang and C.-J. Lin, "LIBSVM: A Library for Support Vector Machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 27:1–27:27, 2011.

[120] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2001.

[121] "A57dataset." http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html.

[122] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008 - A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics," *Advances of Modern Radioelectronics*, vol. 10, pp. 30–45, 2009.

[123] P. Le Callet and F. Autrusseau, "Subjective quality assessment IRCCyN/IVC database," 2005. http://www.irccyn.ec-nantes.fr/ivcdb/.

[124] 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.

[125] Z. M. P. Sazzad, Y. Kawayoke, and Y. Horita, "Image Quality Evaluation Database." http://mict.eng.u-toyama.ac.jp/database_toyama/.

[126] A.Zaric, N.Tatalovic, N.Brajkovic, H.Hlevnjak, M.Loncaric, E.Dumic, and S.Grgic, "VCL@FER Image Quality Assessment Database," *AUTOMATIKA*, vol. 53, no. 4, pp. 344—-354, 2012.

[127] U. Engelke, T. M. Kusuma, H.-J. Zepernick, and M. Caldera, "Reduced-Reference Metric Design for Objective Perceptual Quality Assessment in Wireless Imaging," *Signal Processing: Image Communication*, vol. 24, no. 7, pp. 525–547, 2009.

[128] H. Sheikh, M. Sabir, and A. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, 2006.

[129] H. wen Chang, Q. wen Zhang, Q. gang Wu, and Y. Gan, "Perceptual image quality assessment by independent feature detector," *Neurocomputing*, vol. 151, Part 3, no. 0, pp. 1142 – 1152, 2015.

[130] L. Zhang, Y. Shen, and H. Li, "VSI: A Visual Saliency-Induced Index for Perceptual Image Quality Assessment," *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4270–4281, 2014.

[131] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259 – 268, 1992.

[132] K. Kose, V. Cevher, and A. E. Cetin, "Filtered variation method for denoising and sparse signal processing," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3329–3332, March 2012.

[133] C. R. Nithyananda, A. C. Ramachandra, and Preethi, "Survey on histogram equalization method based image enhancement techniques," in *2016 International Conference on Data Mining and Advanced Computing (SAPIENCE)*, pp. 150–158, March 2016.

[134] A. K. Jain, *Fundamentals of Digital Image Processing.* Englewood Cliffs: Prentice-Hall, 1989.

[135] G. Boccignone and A. Picariello, "Multiscale contrast enhancement of medical images," in *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 2789–2792, Apr 1997.

[136] M. Bertalmio, V. Caselles, E. Provenzi, and A. Rizzi, "Perceptual color correction through variational techniques," *IEEE Transactions on Image Processing*, vol. 16, pp. 1058–1072, April 2007.

[137] E. H. Land and J. J. McCann, "Lightness and retinex theory," *J. Opt. Soc. Am.*, vol. 61, pp. 1–11, Jan 1971.

[138] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing*, vol. 6, pp. 965–976, Jul 1997.

[139] C. Xiao and Z. Shi, "Adaptive bilateral filtering and its application in retinex image enhancement," in *2013 Seventh International Conference on Image and Graphics*, pp. 45–49, July 2013.

[140] A. Laine, J. Fan, and W. Yang, "Wavelets for contrast enhancement of digital mammography," *IEEE Engineering in Medicine and Biology Magazine*, vol. 14, pp. 536–550, Sep 1995.

[141] S. S. Agaian, B. Silver, and K. A. Panetta, "Transform coefficient histogram-based image enhancement algorithms using contrast entropy," *IEEE Transactions on Image Processing*, vol. 16, pp. 741–758, March 2007.

[142] J. Mukherjee and S. K. Mitra, "Enhancement of color images by scaling the dct coefficients," *IEEE Transactions on Image Processing*, vol. 17, pp. 1783–1794, Oct 2008.

[143] J. L. Starck, F. Murtagh, E. J. Candes, and D. L. Donoho, "Gray and color image contrast enhancement by the curvelet transform," *IEEE Transactions on Image Processing*, vol. 12, pp. 706–717, June 2003.

[144] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. T. H. Romeny, and J. B. Zimmerman, "Adaptive histogram equalization and its variations," *Comput. Vision Graph. Image Process.*, vol. 39, pp. 355–368, Sept. 1987.

[145] Y.-T. Kim, "Contrast enhancement using brightness preserving bi-histogram equalization," *IEEE Transactions on Consumer Electronics*, vol. 43, pp. 1–8, Feb 1997.

[146] J.-Y. Kim, L.-S. Kim, and S.-H. Hwang, "An advanced contrast enhancement using partially overlapped sub-block histogram equalization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, pp. 475–484, Apr 2001.

[147] C.-C. Sun, S.-J. Ruan, M.-C. Shie, and T.-W. Pai, "Dynamic contrast enhancement based on histogram specification," *IEEE Transactions on Consumer Electronics*, vol. 51, pp. 1300–1305, Nov 2005.

[148] A. M. Reza, "Realization of the Contrast Limited Adaptive Histogram Equalization (CLAHE) for Real-Time Image Enhancement," *Journal of VLSI signal processing systems for signal, image and video technology*, vol. 38, no. 1, pp. 35–44, 2004.

[149] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas, and K. E. Muller, "Contrast-limited adaptive histogram equalization: speed and effectiveness," in *[1990] Proceedings of the First Conference on Visualization in Biomedical Computing*, pp. 337–345, May 1990.

[150] R. H. Sherrier and G. A. Johnson, "Regionally adaptive histogram equalization of the chest," *IEEE Transactions on Medical Imaging*, vol. 6, pp. 1–7, March 1987.

[151] Y. Li, W. Wang, and D. Yu, "Application of adaptive histogram equalization to x-ray chest images," in *Second International Conference on Optoelectronic Science and Engineering'94*, vol. 2321, pp. 513–515, International Society for Optics and Photonics, 1994.

[152] S. M. Pizer, J. D. Austin, J. R. Perry, H. D. Safrit, and J. B. Zimmerman, "Adaptive histogram equalization for automatic contrast enhancement of medical images," in *Application of Optical Instrumentation in Medicine XIV and Picture Archiving and Communication Systems*, vol. 626, pp. 242–251, International Society for Optics and Photonics, 1986.

[153] E. D. Pisano, S. Zong, B. M. Hemminger, M. DeLuca, R. E. Johnston, K. Muller, M. P. Braeuning, and S. M. Pizer, "Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms," *Journal of Digital Imaging*, vol. 11, no. 4, p. 193, 1998.

[154] N. S. P. Kong, H. Ibrahim, C. H. Ooi, and D. C. J. Chieh, "Enhancement of microscopic images using modified self-adaptive plateau histogram equalization," in *2009 International Conference on Computer Technology and Development*, vol. 2, pp. 308–310, Nov 2009.

[155] Y. Li, Y. Ishitsuka, P. N. Hedde, and G. U. Nienhaus, "Fast and efficient molecule detection in localization-based super-resolution microscopy by parallel adaptive histogram equalization," *ACS Nano*, vol. 7, no. 6, pp. 5207–5214, 2013.

[156] K. Sim, C. Tso, and Y. Tan, "Recursive sub-image histogram equalization applied to gray scale images," *Pattern Recognition Letters*, vol. 28, no. 10, pp. 120–1221, 2007.

[157] P.S.Hiremath, P. Bannigidad, and S. Geeta, "Article:automated identification and classification of white blood cells (leukocytes) in digital microscopic images," *IJCA,Special Issue on RTIPPR*, no. 2, pp. 59–63, 2010. Published By Foundation of Computer Science.

[158] J. Byun, M. R. Verardo, B. Sumengen, G. P. Lewis, B. S. Manjunath, and S. K. Fisher, "Automated tool for the detection of cell nuclei in digital microscopic images: Application to retinal images," *Molecular Vision*, vol. 12, pp. 949–960, Aug 2006.

[159] K. G. Lore, A. Akintayo, and S. Sarkar, "LLNet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognition*, vol. 61, no. Supplement C, pp. 650 – 662, 2017.

[160] S. Ruzin, *Plant Microtechnique and Microscopy*. Oxford University Press, 1999.

[161] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1, pp. 89–97, 2004.

[162] M. Tofighi, O. Yorulmaz, K. Köse, D. C. Yıldırım, R. Çetin Atalay, and A. E. Çetin, "Phase and tv based convex sets for blind deconvolution of microscopic images," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, pp. 81–91, Feb 2016.

[163] M. Benovska, J. Tumova, and O. Wiewiorka, "Microscopic analysis of urine [online]," tech. rep., Faculty of Medicine, Masaryk University ,Department of Laboratory Methods and University Hospital Brno Department of Clinical biochemistry, 2016.

[164] Y.-Y. Fu, *Color image quality measures and retrieval*. New Jersey Institute of Technology, 2006.

[165] A. K. Tripathi, S. Mukhopadhyay, and A. K. Dhara, "Performance metrics for image contrast," in *2011 International Conference on Image Information Processing*, pp. 1–4, Nov 2011.

[166] S. Gupta and R. Porwal, "Appropriate contrast enhancement measures for brain and breast cancer images," *International journal of biomedical imaging*, vol. 2016, 2016.

[167] D. Hasler and S. Süsstrunk, "Measuring Colourfulness in Natural Images," in *Proc. IS&T/SPIE Electronic Imaging 2003: Human Vision and Electronic Imaging VIII*, vol. 5007, pp. 87–95, 2003.

[168] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2411–2418, 2013.

[169] T. K. Agarwal, M. Tiwari, and S. S. Lamba, "Modified histogram based contrast enhancement using homomorphic filtering for medical images," in *Advance Computing Conference (IACC), 2014 IEEE International*, pp. 964–968, IEEE, 2014.

[170] M. Tkalcic and J. F. Tasic, "Colour spaces: perceptual, historical and applicational background," in *The IEEE Region 8 EUROCON 2003. Computer as a Tool.*, vol. 1, pp. 304–308 vol.1, Sept 2003.

[171] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *Proceedings of the IEEE*, vol. 90, pp. 1151–1163, Jul 2002.

[172] R. Clendenin and R. Freeman, "Optical target tracking and designating system," June 7 1983. US Patent 4,386,848.

[173] C. E. Smith, C. A. Richards, S. A. Brandt, and N. P. Papanikolopoulos, "Visual tracking for intelligent vehicle-highway systems," *IEEE Transactions on Vehicular Technology*, vol. 45, pp. 744–759, Nov 1996.

[174] R. Jacob and K. S. Karn, "Eye tracking in human-computer interaction and usability research: Ready to deliver the promises," *Mind*, vol. 2, no. 3, p. 4, 2003.

[175] J. Günther and A. Bongers, "3d motion detection and correction by object tracking in ultrasound images," Jan. 8 2013. US Patent 8,348,846.

[176] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM computing surveys (CSUR)*, vol. 38, no. 4, p. 13, 2006.

[177] B. Deori and D. M. Thounaojam, "A survey on moving object tracking in video," *International Journal on Information Theory (IJIT)*, vol. 3, no. 3, pp. 31–46, 2014.

[178] K. Nishida, T. Kurita, Y. Ogiuchi, and M. Higashikubo, "Visual tracking algorithm using pixel-pair feature," in *International Conference on Pattern Recognition*, pp. 1808–1811, Aug 2010.

[179] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 603–619, May 2002.

[180] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)*, vol. 2, pp. 142–149 vol.2, 2000.

[181] G. R. Bradski, "Computer vision face tracking for use in a perceptual user interface," 1998.

[182] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[183] H. Lee, P. G. Heo, J.-Y. Suk, B.-Y. Yeou, and H. Park, "Scale-invariant object tracking method using strong corners in the scale domain," *Optical Engineering*, vol. 48, no. 1, pp. 017204–017204–9, 2009.

[184] C. Park, K.-h. Baea, and J.-H. Jung, "Object recognition in infrared image sequences using scale invariant feature transform," in *Proc. SPIE*, vol. 6968, p. 69681P, 2008.

[185] C. Liu, J. Yuen, A. Torralba, J. Sivic, and W. T. Freeman, "SIFT Flow: Dense Correspondence across Different Scenes," in *European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part III* (D. Forsyth, P. Torr, and A. Zisserman, eds.), pp. 28–42, Springer Berlin Heidelberg, 2008.

[186] Y. Li, Y. Wang, W. Huang, and Z. Zhang, "Automatic image stitching using SIFT," in *International Conference on Audio, Language and Image Processing*, pp. 568–571, July 2008.

[187] H. P. Moravec, "Visual mapping by a robot rover," in *Proceedings of the 6th international joint conference on Artificial intelligence-Volume 1*, pp. 598–600, Morgan Kaufmann Publishers Inc., 1979.

[188] C. Harris and M. Stephens, "A combined corner and edge detector.," in *Alvey vision conference*, pp. 10–5244, Manchester, UK, 1988.

[189] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification," *Computer Vision–ECCV 2006*, pp. 589–600, 2006.

[190] F. Porikli, O. Tuzel, and P. Meer, "Covariance tracking using model update based on lie algebra," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 1, pp. 728–735, IEEE, 2006.

[191] S. Cakır, T. Aytaç, A. Yıldırım, and Ö. N. Gerek, "Classifier-based offline feature selection and evaluation for visual tracking of sea-surface and aerial targets," *Optical Engineering*, vol. 50, no. 10, pp. 107205–107205, 2011.

[192] S. Cakir, T. Aytaç, A. Yildirim, S. Beheshti, Ö. N. Gerek, and A. E. Cetin, "Salient point region covariance descriptor for target tracking," *Optical Engineering*, vol. 52, no. 2, pp. 027207–027207, 2013.

[193] J. Fan, W. Xu, Y. Wu, and Y. Gong, "Human tracking using convolutional neural networks," *IEEE Transactions on Neural Networks*, vol. 21, pp. 1610–1623, Oct 2010.

[194] N. Wang and D.-Y. Yeung, "Learning a deep compact image representation for visual tracking," in *Advances in Neural Information Processing Systems 26* (C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, eds.), pp. 809–817, Curran Associates, Inc., 2013.

[195] S. Hong, T. You, S. Kwak, and B. Han, "Online tracking by learning discriminative saliency map with convolutional neural network," in *International Conference on Machine Learning*, pp. 597–606, 2015.

[196] Y. Chen, X. Yang, B. Zhong, S. Pan, D. Chen, and H. Zhang, "CNNTracker: Online discriminative object tracking via deep convolutional neural network," *Applied Soft Computing*, vol. 38, no. Supplement C, pp. 1088 – 1098, 2016.

[197] H. Li, Y. Li, and F. Porikli, "DeepTrack: Learning Discriminative Feature Representations by Convolutional Neural Networks for Visual Tracking," in *Proceedings of the British Machine Vision Conference*, BMVA Press, 2014.

[198] R. Brunelli, *Template matching techniques in computer vision: theory and practice*. John Wiley & Sons, 2009.

[199] K. Briechle and U. D. Hanebeck, "Template matching using fast normalized cross correlation," in *Proc. SPIE*, vol. 4387, pp. 95–102, 2001.

[200] A. Goshtasby, S. H. Gage, and J. F. Bartholic, "A two-stage cross correlation approach to template matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, pp. 374–378, May 1984.

[201] J. P. Lewis, "Fast normalized cross-correlation," in *Vision interface*, pp. 120–123, 1995.

[202] J. Luo and E. E. Konofagou, "A fast normalized cross-correlation calculation method for motion estimation," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 57, pp. 1347–1357, June 2010.

[203] Y. Fouda and K. Ragab, "An efficient implementation of normalized cross-correlation image matching based on pyramid," in *2013 International Joint Conference on Awareness Science and Technology Ubi-Media Computing (iCAST 2013 UMEDIA 2013)*, pp. 98–103, Nov 2013.

[204] W. Ouyang, F. Tombari, S. Mattoccia, L. D. Stefano, and W. K. Cham, "Performance evaluation of full search equivalent pattern matching algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 127–143, Jan 2012.

[205] H. S. Demir and A. E. Cetin, "Co-difference based object tracking algorithm for infrared videos," in *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 434–438, Sept 2016.

[206] M. Kristan, J. Matas, A. Leonardis, T. Vojíř, R. Pflugfelder, G. Fernandez, G. Nebehay, F. Porikli, and L. Čehovin, "A novel performance evaluation methodology for single-target trackers," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 11, pp. 2137–2155, 2016.

[207] M. Kristan, A. Leonardis, J. Matas, M. Felsberg, R. Pflugfelder, L. Čehovin, T. Vojir, G. Häger, A. Lukežič, and G. Fernandez, "The visual object tracking vot2016 challenge results." Springer, Oct 2016.

[208] M. Danelljan, G. Häger, F. Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," in *British Machine Vision Conference, Nottingham, September 1-5, 2014*, BMVA Press, 2014.

[209] K. Zhang, L. Zhang, and M. H. Yang, "Fast compressive tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, pp. 2002–2015, Oct 2014.

[210] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," *International Journal of Computer Vision*, vol. 77, no. 1, pp. 125–141, 2008.

[211] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583–596, 2015.

[212] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, and P. H. Torr, "Staple: Complementary learners for real-time tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1401–1409, 2016.

[213] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2544–2550, 2010.

[214] M. Danelljan, G. Hager, F. Shahbaz Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4310–4318, 2015.