Chapter 2

Related Literature Review and Preliminaries

As discussed in Chapter 1, the main contribution of this dissertation is the design and implementation of efficient image dehazing and low-light image enhancement methods for real-time applications. Therefore, in this Chapter, a concise review of the existing works related to image dehazing and low-light image enhancement is presented. Quantitative indicators used to evaluate the performance of the proposed methods are also discussed in this Chapter.

2.1 Image Dehazing

Haze adversely affects the performance of real-time computer vision systems by restricting the visibility of the objects in the captured image [34]. This degradation in visibility may lead to the failure of such systems. A real-time dehazing system can improve a computer vision system's overall performance and make it more resilient even in hazy weather. However, the additional real-time dehazing system should not affect the performance of other subsystems and should be cost-effective. VLSI implementation of such real-time dehazing systems is extremely challenging as it requires performing sophisticated mathematical operations in numerous iterations, which consume extensive hardware resources.

2.1.1 Literature review

Image haze removal algorithms can be broadly categorized into three categories: prior-based dehazing methods, non-prior-based dehazing methods, and deep learning-based methods. A concise review of these methods is presented below

1. *Prior-based dehazing methods* - Various prior-based dehazing methods exist in literature. Tan proposed a dehazing method that was independent of image geometry [35]. His method was based on two main observations *viz.* contrast of the dehazed image is more than that of the hazy image, and variations in atmospheric light are smooth. A cost function based on the above two observations was formulated and optimized. However, this method suffers from halos due to patch-based dehazing. Fattal presented a dehazing method based on the assumption that surface shading of the objects in an image and their transmission are uncorrelated in a local region [36]. Since his method uses the extrapolation technique, it may fail when the input data lacks a sufficient signal-to-noise ratio. A novel, prior-based image dehazing method was proposed in [37] that exploits the statistics of haze-free images. This method is known as Dark Channel Prior (DCP) and is based on an interesting observation that within a local region of natural outdoor images, some of the pixels in the region have very low intensity, at least in one of the RGB channels. Though the DCP method is simple and effective, it produces halos around edges and requires some complex post-processing techniques for haze-free image restoration. Another prior-based image dehazing method was introduced in [38], which used a linear model to estimate scene depth based on the pixel's brightness and saturation. However, this method estimates transmission inaccurately wherever atmospheric conditions are non-homogeneous. A robust airlight estimation approach based on the finding of the most haze-opaque region is presented in [39]. Although an efficient airlight estimation method is proposed in this work, the statistical filtering and sorting process makes it computationally intensive. A linear dependency between saturation and the reciprocal of brightness is explored in [40], which is further used for transmission estimation using local distributions of image pixels and

some boundary constraints. Although it achieves impressive dehazing performance and clear details in the dehazed image, its performance degrades under nonhomogeneous haze. A depth-based positive quasi-linear relationship between hazy and haze-free images is established in [41]. The slope of the linear formula, along with the atmospheric light, is used to solve the atmospheric scattering model for efficient dehazing. This method prevents oversaturation of the dehazed images. Other prior-based dehazing algorithms exist in the literature [42]–[45]. But DCP being a simple and effective method, many researchers proposed modifications in DCP to enhance its performance [46], [47]. Several filtering techniques have also been proposed to refine the transmission map obtained by DCP. Joint bilateral filtering [48], guided image filtering [49], and median filtering [50] are some of the popular refinement methods proposed by researchers for image dehazing using DCP.

2. *Non prior based dehazing methods* - A bi-histogram modification based haze density estimation in transmission map is proposed in [51], which uses two separate modules for haze density estimation and haze removal. However, this method may fail in the presence of non-homogeneous haze. Image dehazing method using some linear transformation is proposed in [52]. It is less computationally intensive and recovers the image efficiently, even at sudden edges. A fusion-based technique in which several under-exposed hazy images are merged using multi-scale Laplacian mixing to obtain the haze-free image is proposed in [53]. Dehazing based on multi-exposure fusion is proposed in [54]. In this method, the hazy image is blurred, and several underexposed versions of it are obtained using gamma corrections. These images with varying exposure levels are further fused using adaptive structure decomposition to obtain the dehazed image. The effectiveness of this method comes at the cost of high computational complexity. A near-infrared fusion model for dehazing is presented in [55], which combines the conventional dehazing model with some depth and color regularization factor to prevent color distortions in the dehazed image. However, dehazed images may contain halos. In [56], a multiscale fusion-based dehazing approach is introduced, which obtains white-balanced and

contrast-enhanced hazy images from the input hazy image. Further, luminance, chromaticity, and saliency maps are obtained from the hazy image. These maps provide the weights for the multiscale fusion of the white-balanced and contrast-enhanced hazy image to obtain a dehazed image. This method also works for night-time hazy images. Saturation-based image deahzing is reported in [57], which derives transmission for individual pixels using saturation of scene radiance. This technique is fast and efficient, as it does not require any refinement process. Saturation and intensity-based image dehazing is reported in [58], which uses a color line model to maintain smoothness and preserve edges.

3. *Deep-learning-based methods* - Several machine learning-based image dehazing networks have flourished recently due to the advancement in computing platforms and the availability of a large number of image datasets to train and test the model. DehazeNet [59], an end-to-end CNN, is highly efficient in image dehazing as compared to conventional and prior-based methods. However, the results could be better if atmospheric light is dynamically estimated. AOD-Net [60] enhanced its dehazing performance by estimating both transmission map and atmospheric light in a single unified model rather than separately estimating them. A multi-scale convolutional neural network is proposed in [61], which uses coarse and fine networks to roughly estimate transmission and then refine it. However, atmospheric light is also estimated separately in this method, which reduces its performance. PDR-Net [62] uses two separate networks, one for image dehazing and the other for image quality enhancement. This CNN architecture produces high-quality images by optimizing a multi-term loss function. RefineDNet [63] proposes a two-stage weakly supervised framework for image dehazing. In the first stage, the prior-based technique is employed to restore the haze-free image, and in the second stage, adversarial learning is applied to enhance the appearance of the dehazed image. A novel CNN-based image dehazing network (DeHamer) is proposed in [64], which integrates the features of CNN and Transformer to attain state-of-the-art dehazing performance. A semi-supervised generative adversarial network (GAN) is introduced

in [65], which aligns latent characteristics of the hazy image using adversarial learning and Kullback– Leibler divergence. Finally, it uses the local entropy of different hazy regions to estimate haze concentration, further aiding the domain alignment process. Supervised learning from synthetic and real datasets boosts the dehazing performance of this method. A hazy image can be considered as a combination of a clear image and a haze mask. Based on this consideration, an unsupervised network is proposed in [66], which separates these two components using heterogeneous networks. Further, it employs a multi-scale feature attention module to improve the visibility of the objects in dehazed images. In [67], a two-stage weakly supervised dehazing network is proposed that uses dark channel prior (DCP) [37] for visibility restoration and improves the naturalness of the restored image using GAN. Further, the performance of this method is enhanced using a fusion strategy. A contrastive learning approach introduced in [68] maximizes the content information between the reference image and the dehazed image. The global information from the reference hazy image and the entire dataset is utilized to improve the dehazing efficiency of the contrastive learning network. Most of the aforementioned techniques and algorithms require high-performance computing platforms or GPU accelerators for their implementation. Thus, a real-time hardware implementation of the aforementioned image dehazing methods is a tedious task.

2.1.2 Existing Image Dehazing VLSI Architectures

A real-time haze removal method based on DCP is proposed in [69], which performs gradient detection to avoid halo artifacts arising due to inaccurate transmission estimation for pixels in a patch lying on the edge. Since this method uses a smaller window size of 3×3 for calculating the dark channel, it may inaccurately estimate the atmospheric light when white objects are present in an image. A faster image dehazing architecture is proposed in [70], but it may also suffer from inaccurate atmospheric light estimation due to a smaller window size. A video defogging method is proposed in [71], which also works for single-image dehazing. This method uses complex weights to calculate atmospheric light using gray levels of the hazy image and suppresses the halo artifacts by employing an edge-preserving filter. A similar dehazing architecture is proposed in [72] with a modified edge-preserving filter to eliminate halo artifacts in the recovered images. The atmospheric light is also dynamically adjusted to enhance the visual quality of the dehazed image. However, this technique also uses a smaller window size to calculate the dark channel, which makes it prone to inappropriate global atmospheric light calculation, which is further used to calculate local atmospheric light. An image dehazing engine with parallel processing cores for calculating airlight and transmission rate is presented in [73], which achieves better performance at the cost of extra hardware. Depth-based transmission and atmospheric light estimation are proposed in [74], where the depth of the scene is calculated and carefully calibrated to differentiate distant objects from white objects. A saturation-based image dehazing architecture is presented in [75], which computes transmission on a pixel-by-pixel basis. Though it consumes fewer hardware resources, its performance is still limited by the smaller window size used for atmospheric light estimation. Recently, a high-performance video dehazing accelerator based on mixed atmospheric prior is reported in [76]. Here, the initial atmospheric light map is estimated depending on the intensity of the hazy pixels being greater or less than a statistical threshold. Further, the atmospheric light and transmission map are refined to suppress artifacts using a fast-guided filter (FGF) and side window filter, increasing algorithmic complexity and hardware resource utilization.

2.2 Low-Light Image Enhancement (LLIE)

Low-light images suffer from poor visibility and color degradation. A real-time computer vision system is expected to work equally well under different illumination conditions. But low lighting and uneven illumination conditions can severely affect its performance because some information, such as edges and textures, may be lost during the image acquisition [77], [78]. This problem can be mitigated by integrating an efficient low-light image enhancement (LLIE) subsystem. However, real-time implementation of a reliable and cost-effective LLIE system is intricate.

2.2.1 Literature review

A concise review of LLIE methods is presented in [79]. Traditionally, histogram equalization (HE) methods [21]–[25] were used to adjust the dynamic range of the low-light image. However, these methods achieve limited enhancement as they neglect the overall features of the image. Retinex-based low-light image enhancement methods have gained more popularity than HE methods. Retinex techniques decompose a low-light image into reflectance and illumination channels. These channels are further enhanced and recombined to obtain a visually pleasing image. In [80] and [81], the illumination channel is obtained using a Gaussian filter. In [82], illumination is estimated by employing a bright pass filter (BPF), which uses average values of pixels greater than the center pixel value. To prevent detail suppression and preserve lightness order, bi-log transformation is used for illumination remapping. Simultaneous illumination and reflectance estimation using a weighted variational Retinex model is proposed in [83], which uses a minimization scheme to solve the proposed model. This variational model efficiently preserves fine details in reflectance and limits noise up to a certain level. However, its computational time is high. Another Retinex-based low-light image enhancement variational framework is proposed in [84]. Here, an RGB image is converted into HSV color space, and a guided image filter is applied to the V channel to obtain illumination. Further, contrast-limited adaptive histogram equalization (CLAHE) is applied to both illumination and reflectance to enhance visibility. Though it produces fruitful results, its computational complexity is high. An efficient LLIE technique, low-light image enhancement via illumination map estimation (LIME), is presented in [85], which imposes some structural information to refine the initial illumination channel using the Augmented Lagrangian Multiplier (ALM). The refined illumination is further gamma-corrected to obtain the enhanced image. If needed, it converts reflectance from RGB to YUV color space and employs the block-matching for 3D (BM3D) technique for denoising the reflectance. However, it may suffer from overexposure of restored images. Since Retinex-based enhancement algorithms are complex, a fast and efficient method with reduced computational cost is proposed in [86], where illumination is modified using a weight map to suppress halos. Further, gamma correction is used to enhance the brightness of the enhanced image. Based on Retinex, an edge-preserving filter for illumination estimation is presented in [87]. It limits the range of illumination and preserves details in the reflectance. Further, the illumination estimation process is accelerated using box filters and by reducing the number of window slides.

In recent years, machine-learning-based LLIE models have flourished. An adaptive CNN model for LLIE is presented in [88], where a multiscale attribute map is utilized to preserve the textures in the enhanced images. However, it may suffer from blurred edges in the restored images. A Retinex model-based enhancement network is introduced in [89]. It uses a separate network for reflectance decomposition and illumination enhancement. This end-to-end model performs very well in illumination adjustment. First, it feeds the low-light image to a subnetwork, Decom-Net, to obtain lighting-independent reflectance and structure-aware smooth illumination. Next, the illumination map is adjusted using another subnetwork, Enhance-Net, using multi-scale illumination adjustment. Further, it enhances reflectance by denoising it. However, it sometimes restores images that look unnatural. Another Retinex-based CNN model for LLIE is presented in [90]. This model learns by performing an end-to-end mapping between the dark and its corresponding bright image. The visual results produced by this model are rich in detail, but halos can appear in smooth regions. A highly effective GAN-based enhancement network proposed in [91]. This method employs a global-local discriminator to handle the lighting conditions adaptively. It imposes a self-regularized attention model using the illumination map and eliminates the need for paired training data. Further, it employs a global-local discriminator to handle the spatially varying lighting conditions adaptively. It uses U-net as a generator backbone, which has eight convolutional blocks. It mitigates the checkerboard artifacts using bilinear upsampling and a convolutional layer. However, it may produce color distortion up to some extent in restored images. Another Retinex-inspired, lightweight enhancement network is presented in [92], which simultaneously searches for an efficient architecture within a compact search space to estimate illumination and remove noise without paired or unpaired supervision. It uses a bilevel search strategy and distillation technique to reduce the processing time. However, sometimes, it may be victimized by overexposure of images. An end-to-end LLIE model proposed in [93]. It uses illumination as a constraint to produce visually enhanced images. It performs feature extraction based on the pre-trained VGG model. Further, channel attention mechanisms based on Res2Net adjust the light levels of poorly illuminated areas. It incorporates perceptual loss and structural similarity loss in its total loss function for better restoration of contrast and image details. Though it has better control over the exposure and saturation of the restored images, it may cause high brightness in them.

2.2.2 Existing LLIE VLSI Architectures

Most of the effective LLIE algorithms are computationally intensive and require large memory, especially machine learning-based algorithms. Methods using high-performance GPUs like [88], [89], [90], [91], [92], and [93], become not only power hungry but also costly. This restricts their suitability for real-time implementation on reconfigurable devices with limited resources. Researchers have tried to implement Retinex-based LLIE algorithms on FPGAs in real-time with several modifications to the base algorithm. FPGA implementation of multi-scale Retinex-based LLIE is proposed in [94], which replaces Gaussian filtering with blurring and substitutes logarithmic values with pre-calculated values. Still, it requires a large number of hardware resources. The algorithm proposed in [86] is modified to attain real-time implementation on the FPGA platform in [95]. Here, the hardware cost is reduced by employing approximate computations and replacing frame buffers with line buffers. Though this method restores the naturalness in the images efficiently, some details are lost in the decomposed reflectance. A real-time implementation of a tone mapping algorithm with halo suppression capability is presented in [96]. It efficiently controls the brightness of the low-light images by generating some key control parameters and employs a halo-reducing filter to suppress halo artifacts. However, large bit widths increase its hardware cost and power consumption. An adaptive histogram equalization-based FPGA implementation of low-light video enhancement is proposed in [97], which suppresses haze also. It utilizes a mean filter to suppress noise introduced by a single-color neighborhood space. With moderate hardware resource utilization, this method is capable of processing low-light images in real-time.

2.3 Quantitative Evaluation Metrics

Quantitative evaluation metrics are used to evaluate the effectiveness of image enhancement techniques. These evaluation metrics are reference-based or no reference-based. In reference-based evaluation metrics, the enhanced image is compared to a reference image to measure the enhancement performance. Some of the commonly used reference-based evaluation metrics are peak signal-to-noise ratio (PSNR) [98], structural similarity (SSIM) [99], color difference CIEDE2000 [100], etc. On the contrary, no reference-based evaluation metrics measure the enhancement performance using the enhanced image itself. Some of the commonly used no-reference-based evaluation metrics are the naturalness image quality evaluator (NIQE) [93], blind/referenceless image spatial quality indicator (BRISQUE) [101], etc. These quantitative evaluation metrics help determine improvements in clarity, sharpness, contrast, and overall visual quality.

2.3.1 Peak Signal-to-Noise Ratio (PSNR)

PSNR [98] measures the ratio between the maximum possible power of a signal (image) and the power of noise that affects the quality of the signal. A higher PSNR value indicates better image quality, as it means less distortion and noise. It is often used to assess image denoising and compression techniques. It can be obtained mathematically as

$$PSNR = 10 * log(\frac{{f_{max}}^2}{MSE}), \tag{2.1}$$

where f_{max} is the maximum gray level (2^N for an N-bit image), and MSE represents the mean square error between the enhanced image and the reference image obtained as

$$MSE = \frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I_e(i,j) - I_r(i,j)]^2. \tag{2.2}$$

Here, I_e represents the enhanced image, I_r represents the reference image and $m \times n$ represents the image size.

2.3.2 Structural Similarity (SSIM)

SSIM [99] evaluates image quality by comparing structural information, contrast, and luminance between the original and enhanced images. Unlike PSNR, SSIM considers human visual perception, making it a more accurate metric for measuring visual quality. SSIM values range from -1 to 1, with 1 indicating a perfect match. It can be expressed as

$$SSIM = F(l, c, s), \tag{2.3}$$

where l is luminous, c is contrast, and s is structure.

2.3.3 CIEDE 2000

CIEDE2000 [100] is an advanced color difference metric used for color difference measurements between two images. CIEDE2000 considers factors such as lightness, chroma, hue, and their interactions to better align with human visual perception. It introduces corrections for non-linearity in color perception and includes weighting functions that adjust differences based on hue and chroma variations. CIEDE2000 ranges between 0 to 100. The smaller value indicates a lower color loss.

2.3.4 Naturalness Image Quality Evaluator (NIQE)

NIQE [93] is a no-reference image quality assessment (IQA) metric that measures the perceptual quality of an image without requiring a reference or original high-quality image. Unlike traditional metrics like PSNR or SSIM, which compare an enhanced image to a ground truth, NIQE evaluates image quality based on statistical models derived from natural images. NIQE works by analyzing the statistical properties of an image's local patches and comparing them to a learned model of high-quality natural images. It uses a multi-variate Gaussian model to assess distortions such as blurring, noise, and compression artifacts. A lower NIQE score indicates better image quality, as it means the image is closer to the natural image distribution. NIQE is widely used in image enhancement, denoising, and super-resolution applications, where subjective quality assessment is needed without access to a reference image.

2.3.5 Blind/Referenceless Image Spatial Quality Indicator (BRISQUE)

BRISQUE [101] is a no-reference image quality assessment (IQA) metric that evaluates the perceptual quality of an image without requiring a reference or original high-quality image. It is designed to measure distortions such as blurring, noise, compression artifacts, and contrast loss by analyzing the natural scene statistics (NSS) of an image. BRISQUE operates in the spatial domain by extracting statistical features from local image patches and comparing them to the statistical properties of high-quality natural images. It uses a machine learning model trained on human opinion scores to predict image quality. The BRISQUE score ranges from 0 to 100, where lower values indicate better image quality. Because of its efficiency and strong correlation with human visual perception, BRISQUE is widely used in image processing, enhancement, and quality assessment tasks.

2.4 Chapter summary

In this chapter, we discussed the necessity of real-time haze removal and low-light image enhancement systems. A brief summary of the existing techniques and related hardware architectures for image dehazing and low-light image enhancement, along with their merits and demerits, are presented in this chapter. Finally, the quantitative performance measurement metrics used to measure the efficiency of image enhancement algorithms are explained briefly in this chapter.