IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 10, 2021, accepted January 29, 2021, date of publication February 4, 2021, date of current version February 11, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3057167

Multiview Ghost-Free Image Enhancement for In-the-Wild Images With Unknown Exposure and Geometry

RIZWAN KHAN[1,2], ADEEL AKRAM[3], AND ATIF MEHMOOD[4]
[1]School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China
[2]Wuhan National Laboratory of Opto-electronics, Wuhan 430074, China
[3]School of Information Engineering (Big Data), Xuzhou University of Technology, Xuzhou 221018, China
[4]School of Artificial Intelligence, Xidian University, Xi'an 710126, China

Corresponding author: Adeel Akram (dradeel@xzit.edu.cn)

This work was supported by the Xuzhou University of Technology.

ABSTRACT The multiview low dynamic range images captured with sparse camera arrangement under ill-lighting conditions contain highlighted and shadow regions due to over-exposed and under-exposed regions. The processing of these images produces contrast distortion, and it is challenging to maintain relative brightness with color consistency. Moreover, the disparity map estimation faces the challenges of holes and artifacts due to a wide baseline and poor visibility, with a shared view of vision. In this article, we propose a multiview ghost-free image enhancement strategy for in-the-wild images with unknown exposure and geometry. We address the complex geometric alignment problem for a wide variational baseline among multiple sparsely arranged cameras. The features among multiple viewpoints are detected and matched for the image restoration. The restored image contains highlighted and shadow regions with a color imbalance problem. We synthesize virtual images following the intensity mapping function, which compensates for the relative brightness and color distortions. Finally, we fuse all the images to obtain high-quality images. The proposed method is more frequent and feasible for future multiview systems with varying baselines without relying on disparity maps. The experimental results demonstrate that the proposed method outperformed the state-of-the-art approaches.

INDEX TERMS Multi-view images, feature matching, virtual images, exposure fusion.

I. INTRODUCTION

High quality images are essential for many applications in image processing and computer vision. The advancement in display technology has paved the path for multiview capturing devices, and recent imaging system involves more than one cameras, comprising stereo (two cameras) and multiview camera systems (MVCs). In the case of monitoring, surveillance, or even capturing a social event, the multiple cameras are pointed towards a common point, and the quality of the captured image varies due to the in-the-wild environmental illumination conditions. The shared view of vision and non-uniform illumination gives rise to backlit conditions with exigent glitches, where over-and-under-exposed regions coexist in the same scene.

The images captured with multiview cameras system (MVCs) suffer back-lighting issues, and non-uniform illumination distributions cause highlighted and shadow regions. Capturing the entire dynamic range of a scene under ill-lit (weaklit and backlit) conditions is almost impossible through ordinary digital cameras. The high dynamic range [1] (HDR) methods require multiple images with additional information (e.g., exposure time, ISO values) and prone to tone mapping artifacts. The alternative exposure fusion [2] methods also require more than one image. In the case of multiple sparsely arranged cameras, the disparity map estimation is another challenge, where wide baseline and complex geometric alignment in dim illumination give rise to holes and artifacts. Classic HDR and exposure fusion methods are susceptible to motion artifacts that give rise to additional challenges of deghosting. It limits the performance of the associated systems, and challenges arise from the special visual and display applications.

The associate editor coordinating the review of this manuscript and approving it for publication was Naveed Akhtar.

Capturing high-quality images in low lighting conditions is challenging and defies the camera setting. The direct amplification through histogram equalization [3] and gamma corrections [4] distort the global structure of image. These methods work without considering the illumination and produce vulnerable results with over-and under-exposures. The illumination aware strategies are proposed to manipulate the illumination following retinex (RT) decomposition principle [5]. The key assumption is of RT is to divide the input image into reflection and illumination components [6], [7]. The network-based methods [8]–[10] relies on image pairs or large scale training datasets [11], [12]. However, acquiring large scale datasets and image pairs of the same scene is challenging. Various learning-based methods [13]–[16] and high dynamic range and super-resolution imaging (SR-HDR) methods [17] have been proposed to solve the problem. Trocoli *et al.* proposed the multi-exposure stereo method [18], which is extended to the multiview HDR method [19], but inaccurate disparity estimation results in holes and artifacts. Extensive optical flow algorithms are available to solve the alignment problems [20], [21], but optical flow are unreliable when images are captured with varying exposures [15], [22]. The traditional HDR technique [1] require more than one images and fail in the case of robust exposure variations [23]. Exposure fusion-based methods [2], [24] plays a key role to increase the dynamic range of the output images [25]. But the direct fusion of ill-lit images with over-and-under-exposed regions results in the undesired highlight and shadow artifacts. Multiple exposure fusion methods and traditional HDR methods are also susceptible to motion artifacts. The general restoration based methods [11], [26], [27] are designed to enhance the contrast for under-exposed regions and maintain the over-exposed regions at the same level. These methods handle the single image and adaptive towards multiview low dynamic range imaging problems. In the case of a multiple-camera system, the scene's complex geometry gives rise to an alignment problem for the complex non-rigid motions. When outdoor and in-the-wild scenarios are considered, where both exposure time and geometric relationship among multiview cameras are unknown, no previous works have been proposed for this challenge to the best of our knowledge.

In this article, we focus on the above challenge of ill-lit image enhancement for the in-the-wild multiview (MV) backlit images with unknown exposure time and geometry. The input images are captured with a sparse camera arrangement, so we enclose the geometric characteristic of input images and detect and match the accurate feature points. We use these matched feature points to estimate the image restoration function to obtain multiview LDR images. The restored MVLDR images contain highlighted and shadow regions. Therefore, we generate virtual images with intermediate exposure using an intensity mapping function, merged the input, virtual, and restored low dynamic range (LDR) images to acquire the high-quality output. In our method, only one image is required for each viewpoint, where we solve the problem for standard photography to be more beneficial for future interactive visual application.

II. RELATED WORK

The restoration of ill-lit images in an active research area and many enhancement schemes have been proposed in the past few decades. We briefly overview the state of the art approaches in this domain.

A. DIRECT ENHANCEMENT AND ILLUMINATION AWARE METHODS

The visibility of the low light images can be enhanced directly using histogram equalization (HE) methods [3], [28], which amplify the pixels energy directly. The contrast limited adaptive enhancement methods [29], [30] has been proposed to encounter the problem of unnatural transitions, and methods are classified as global and local methods [31]. The HE methods and its follow-ups [32]–[36], [36] aims to enhance the contrast of the image, and map the histogram to avoid truncation problem. The gamma correction methods [4], [37], [38] also tries to map the pixel values individually in a non linear manner and promote the brightness to the dark pixels. These methods enhance the contrast of the low light images and work irrespective of the neighboring pixels' energy and frequently produce over-and under-enhancement with color and contrast distortions. The main limitation of the direct enhancement methods is that these methods work irrespective of considering the illumination and produce vulnerable results in real-world in-the-wild scenarios.

The illumination awareness plays a key role in the contrast enhancement of the low light LDR images. The key assumption of these methods is the Retinex based decomposition of the input images into reflection and illumination [39]. The Retinex methods, such as single scale retinex [5] and multi-scale retinex [40]. manipulate the illumination and reflection of the input images to handle the structure and texture distortion. The output of these methods looks over enhance with unnatural transformations. The non-uniform illumination in the image requires an illumination dependent robust mechanism. The variational energy models [41], [42] have also been proposed for the he non-uniform illumination distribution in the images. The algorithm known as NPE [43] is proposed to preserve the naturalness and contrast. The author in [27] proposed an illumination based feature fusion strategy for weak lit images (FEW) to adjust the initial illuminations, but it scarifies the details in the extremely short exposure and rich texture regions. The non-uniform illumination in the real world environments distorts the quality of the image [44] where some statistical prior plays a vital role. Guo *et al.* propose an illumination aware method [7] to estimate the illumination map to initialize the enhancement following total variation prior. The noise also exists in its worst form in the under-exposed regions, and inaccurate illumination smoothness amplifies the noise. The authors in [26] presented joint image denoising and enhancement (JED) approach working on the sequential decomposition. The order of denoising operation

limits many algorithms' performance because the denoising before enhancement produces blurry output. In contrast, the denoising after the enhancement removes the useful features along with noise. The structure revealing low light image enhancement (SLIME) [6] method solve the noise problem with robust retinex model but produce an inconsistent reflection.

B. LEARNING BASED AND HDR IMAGING METHODS

A large number of methods have been proposed in the literature for the low level vision tasks [45]–[47], super-resolution [48], enhancement and denoising [49], [50] tasks. A learning base restoration (LBR) method is proposed in [11] to enhance backlit images. It is a classifier based approach based on image division into front-and-back-lit regions. The general deep learning-based methods for the enhancement of low light images mainly follow the retinex decomposition principle. A deep retinex method [8] is proposed to enhance the low light images which suffer artificial texture transformations. A follow up of this method famous as KinD [9] is proposed to kindle the darkness with a denoising operation but fails to maintain the contrast and color balance in-the-wild non-uniform illumination conditions. The author in [14] proposes a method for the enhancement of undesirably illuminated images and propose to generate HDR by using a single image. But it produces an undesired noisy effect while compensating for the illumination, which sacrifices the realistic appearance. Relatively less work is devoted to multiview HDR, especially for the cameras' wide baseline.

The high dynamic range methods capture the same scene's exposures to recover the radiance map [1], [51]. The technique proposed by Debevec *et al.* is considered the most widely used technique for recovering radiance maps from multiple LDR images captured with an ascending exposure time. However, it is not suitable for robust exposure variations and fails to estimate the true weighting for extremely bright and dark regions; for details, readers are referred to O'Malley *et al.* [23]. Moreover, the traditional HDR pipeline involves complex steps to estimate camera response function (CRF), require exposure time, and ISO information, unknown to ordinary users in advance. This HDR approach is extended to stereo HDR [52]–[55] and multiview HDR [19], [56]–[58] with small baseline and a linear camera arrangement. Kalantari. *et al.* utilized optical flow algorithms to generate multiview HDR images, but the output suffers noise and artifacts [22]. The optical flow algorithms are not reliable when the images are captured with varying exposures [15]. The extant HDR approaches mainly rely on the estimation of disparity map and are suitable for small baseline among the cameras. Moreover, these methods are susceptible to motion, and tone mapping artifacts because they merge multiple frames and construct an intermediate HDR image, which is tone-mapped for LDR display.

Mertens *et al.* proposed an algorithm for multi-exposure fusion using three quality aggregates (color, contrast, and saturation) for weight maps estimation [2]. Exposure fusion methods provide an alternative to the traditional HDR pipeline, produce a high-quality image without relying on CRF, and avoid tone mapping artifacts [24], [25]. These methods create a final high dynamic range image by fusing the best-exposed areas in the bracketed input sequence based on Laplacian pyramids. After that, Laplacian and Gaussian pyramids are decomposed, and finally, multiple exposures are blended to obtain the fused images. Quite often, capturing multiple exposures under robust light variations is also challenging due to shaking and motion effects. It gives rise to misalignment, ghosting, halos, and artifacts in the resulting image. It is also challenging to acquire multiple frames of the weakly and non-uniformly illuminated scene. The situation becomes worse if the images are captured by using an ordinary mobile camera or CCTV devices. Moreover, the direct fusion of the ill-lit images with over-and-under-exposed regions results in the undesired highlight and shadow artifacts.

Deep learning-based exposure fusion method is proposed in [59], but it has limited practical implication due to image resizing. HDR from a single LDR image [13] and contrast enhancement methods are also proposed using learning approaches [60]. The end to end trained deep Retinex network [8] and HDR-Net [10] relies on the paired training dataset, whereas adversity of obtaining training pairs limit the practical implication of these methods. A recent zero-reference deep curve estimation (Zero-DCE) method [12] is also proposed to formulates light enhancement as a task of image-specific curve estimation with a deep network, without relying on the image pairs. Although these methods produce high-quality images by utilizing low-contrast multi-exposure images with high-quality reference images, the preparation of large-scale datasets is challenging. Whereas, this method can't preserve details for over-exposed or large exposure areas in the image. In the case of multiple available exposures of the same scene, frame averaging is also a frequent blur-free method for registered frames without object motion. This restriction makes the frame averaging a problematic approach. Most of the proposed techniques produce undesired artifacts in the case of dark and short-exposure images. Multi-exposure-fusion and traditional HDR techniques work for bracketed exposures or a sequence of images with a small difference of exposure. Although researchers have proposed multiview HDR imaging techniques, halos and artifacts' problem due to disparity and pixel mismatch in low lighting conditions remained a challenge among the camera's wide baseline. On the other fusion of multiple frames produce ghosting artifacts.

The proposed method utilizes a single image per viewpoint instead of capturing bracketed exposures and produce a ghost-free image with a higher dynamic range. Our method involves the restoration of multiview low dynamic range images captured with unknown exposure time and geometry, using multiple sparsely arranged cameras. In our methods, we detect and match accurate feature points among the adjacent viewpoints and remove the color and contrast distortion

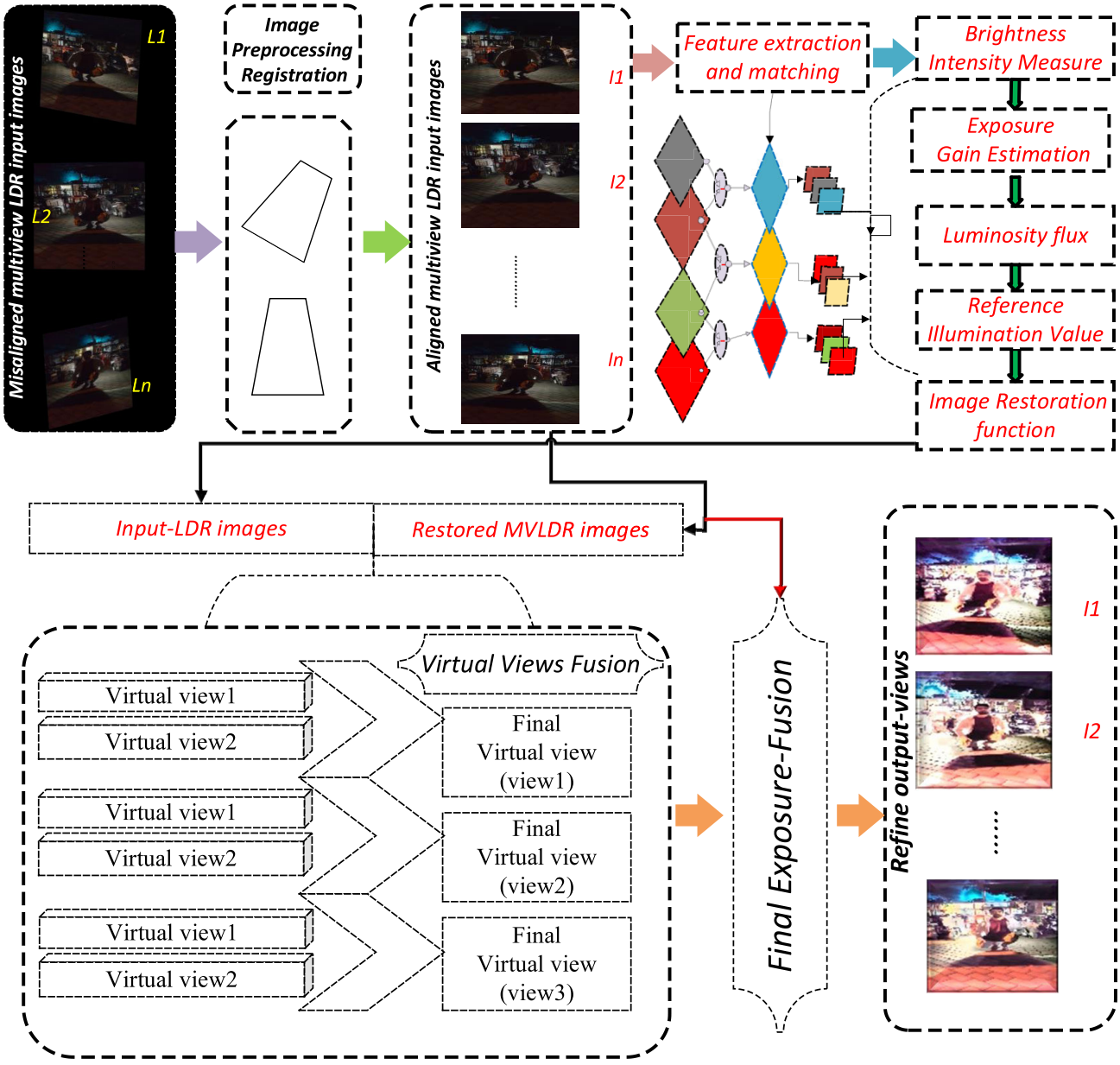


FIGURE 1. Architecture of the proposed scenario. Images are captured with multiview camera system. The geometric features are enclosed for accurate feature matching. The virtual images are produced and fused with the input and restored multiview low dynamic range images to acquire a high dynamic range.

by generating intermediate virtual views and dispense the challenge of disparity map estimation.

III. THE PROCESSING PIPELINE

A. THE PROPOSED METHOD

The proposed method combines the corresponding weighted illumination information gathered from the neighbouring viewpoints to preserve the relative brightness for the lowlight input images. In this regard, we first improve the feature points (FPs) detection and matching for accurate exposure gain (ϕ_g) estimation, amongst the adjacent input images. A weighted reference value ($\Re$) updates the exposure gain among corresponding views based on the matched feature points. The framework of our method is presented in Fig. (1). The color distortions arise due to varying exposures in the restored MVLDR images. In order to solve this problem, we produce virtual views via the intensity mapping function by utilizing input views and restored MVLDR images. Finally, the images per viewpoint $Z_i v_i$ (i = 3) include input images, restored MVLDR images per viewpoint, and final virtual views are merged via exposure fusion. Gaussian pyramid of the l-th level luminance component is $G\left\{Y_i^l(p)\right\}$ and weight map Gaussian pyramid for each view is defined as $G\left\{P_i^l(p)\right\}$ with the l-th level Laplacian pyramid $L\left\{Z_i^l(p)\right\}$.

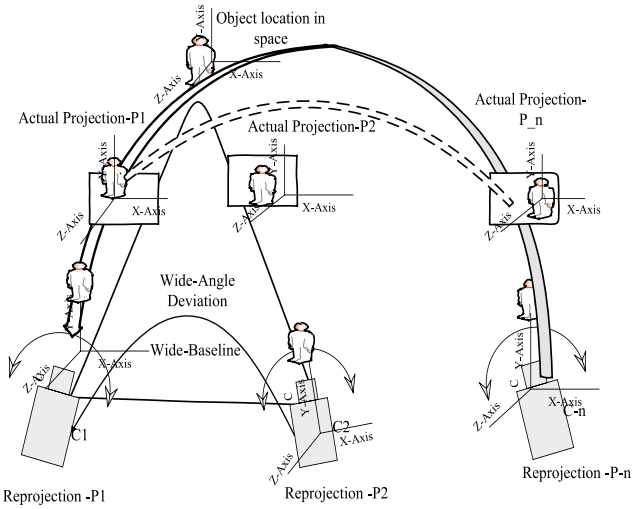


FIGURE 2. The alignment of a multicamera system with the proposed scenario.

In order to acquire the final weight $W^{(l)}{}_i(p)$ for input images at the l-th level, the guided image GGIF smooths these weight maps for input images, and fusion produces a refined output image following Eq. (1), with a higher dynamic range.

$$W^{(l)}{}_i(p) = GG(G\{P^{(l)}{}_i(p)\}, G\{Y^{(l)}{}_i(p)\}) \tag{1}$$

B. THE PROPOSED MULTIVIEW SCENARIO

In this work, we propose a scenario to enhance MVLDR images captured under ill-lighting conditions with multiple sparse camera arrangement, as shown in Fig. (2). The multiple non-linearly arranged cameras ($C_i; i = 1, 2 \ldots N$), capture multi-exposure input images ($Z_i; i = 1 \ldots\ldots N$), with exposure times t_i for exposures $X_s = e_1 \ldots\ldots e_N$, across multiple viewpoints ($V = Vi; i = 0 \ldots\ldots N$), with a variable baseline among the cameras. The proposed scenario doesn't impose any binding on the multi-camera system. Our method enhances the low light images with robust exposure variation and produces a high dynamic range image from these images by using a single image per viewpoint. We demonstrate the scenario with three sparsely arranged cameras C_1, C_2, C_3 to capture three images from three viewpoints, as shown in Fig. (2).

C. RESTORATION OF MULTIVIEW-LDR IMAGES

The restoration and enhancement of MV-LDR images demand optimal weighting strategy. The coexisting over-and-under-exposed regions in the same scene results in objectionable artifacts and fusion of these images produce highlighted and shadow regions. The difference threshold can be estimated as a ratio between constraint stimulus δL to background Luminance L, $(C = \delta L/L)$ [61]. This work mainly focuses on suitable input images to accomplish fusion requirements to maintain exposedness, contrast, and color consistency. We utilize the input exposure-gain to handle the target problem in an organized manner to restore the MVLDR images. It is important to note that our scenario involves multiple non-linearly arranged cameras; hence we require a geometric alignment. First, we enclose the geometric characteristics of the multiview input images captured from multiple viewpoints for accurate feature detection. We utilize the FPs in the estimation of our image restoration function (IRF) among the multiple viewpoints. We introduce a weighted exposure-gain-based IRF to restore the LDR images for each viewpoint. In the case of multiple camera systems and ill-illumination conditions, robust exposure variation appears in the images because cameras are sparsely arranged. One light source appears in multiple images due to a shared view of vision among multiple views. In order to achieve the desired dynamic range, the proposed framework is elaborated in the following key-points.

- Input images are captured with sparsely arranged multiple cameras as shown in Fig. (3.a), where the geometric characteristics improve the matching and detection of the feature points among viewpoints. First of all, we calibrate and rectify multiview low exposure and low dynamic range (LDR) input images to get geometric alignment. The calibration and rectification process results in correspondence among all of the viewpoints for the reference viewpoint with respect to the adjacent viewpoints. A large number of outliers may contribute to the false matches as shown in Fig. (3.b). In order to acquire the exact feature point matching, we have matched the projection of the ideal image plane by rectification process as shown in Fig. (2). We determine the matched feature points for the viewpoints V_i across the images. In order to minimize the faulty matching, we utilized the reference base image adjustment techniques, and multiview geometric alignment process [62]. A reference image is selected for ideal plane adjustment concerning the horizontal and vertical alignments for multiple viewpoints. The rectification process limits the faulty matches and outliers are removed significantly, as shown in Fig. (3.c). After that, we restrict the range of horizontal and vertical boundaries to remove the outliers to achieve meticulous feature points, shown in Fig. (3.d). We employ a scale-invariant feature transform (SIFT) technique [63] for feature detection and matching. Now, we can utilize these matched FPs to estimate our image restoration function. The process of IRF is expressed and summarized in the following key-points.
- In order to estimate IRF, first we calculate brightness value ($\psi_{v,m}$) of all pixels (ϱ) in matched feature points (m) for viewpoints (V_i) with a corresponding view point ($V_{i\pm1}$) and calculate the mean value with respect to each pixel (Υ) in all matched feature points (M), whereas $\{\Upsilon \in \varrho\}$.

$$\psi_{v,m} = \sum_{\Upsilon \in \varrho} \frac{\psi^{\Upsilon}_{v,m}}{M} \tag{2}$$

- Now, we determine *exposure-gain* (ϕ_g) as a mean value difference (ϕ_{g-min}), based on illumination of matched

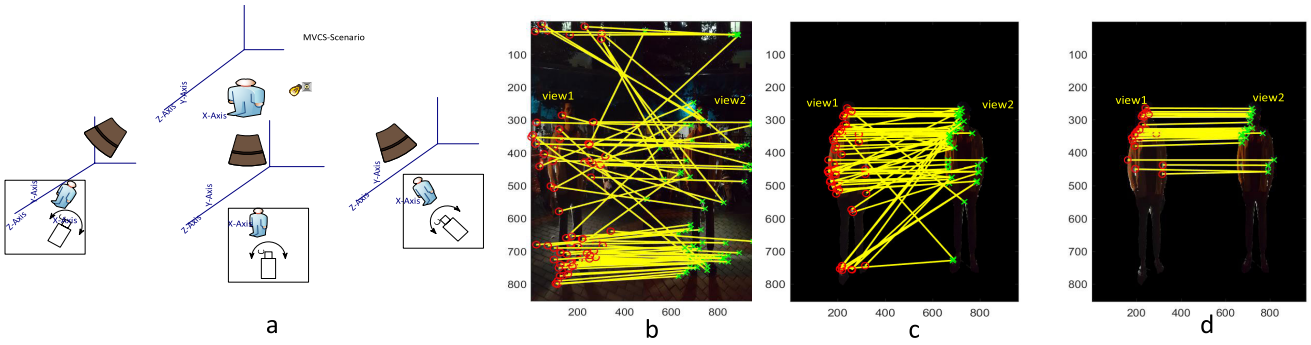


FIGURE 3. (a) Three viewpoints capturing scenario for an outdoor scene *boy & moto*. The comparison (b) and (c) without proposed strategy with a large number of outliers; (d) shows the accurate matching with the proposed strategy.

FPs in viewpoints V_i.

$$\phi_g = \frac{\sum_{m=1}^{M}(\psi_{v,m} - \psi_{v\pm1,m})}{M} \tag{3}$$

The manipulation of input exposure levels provides quite a practical demonstration to understand the luminance of input images. It can be seen through Eq. (2), where the value of brightness provides an estimate for the input exposure gain in Eq. (3). The optimal increment provides continuous up-gradation and manipulation based on input exposure gain to facilitate image luminance trade-off. Our scenario involves multiple images from multiple viewpoints, and the estimated gain depends on the feature points in the adjacent views. The feature points among each of the two viewpoints vary due to the non-uniform illumination distribution. To avoid possible distortion in the restored MV images, we design a reference value function $\Re$. This value depends on the exposure gain and the total value of brightness intensity, i.e., termed as luminosity flux ϕ_s.

$$\phi_s = \frac{\sum_{m=1}^{M}(\psi_{v,m} + \psi_{v\pm1})}{M} \tag{4}$$

In order to fully utilized the impact of the brightness intensity, our reference value function provides a supplemental consistency. The exposure gain value in Eq. (3) and luminosity flux value in Eq. (4) provides the midpoints (p_{ϕ_g}) and (p_{ϕ_s}) respectively, based on the corresponding maxima and minima. $\Re$, is expressed as below:

$$\Re = \sqrt[M-1]{\frac{p_{\phi_g}}{p_{\phi_s}}} \tag{5}$$

The product of this weighted reference value and exposure gain regulate the image restoration function φ, which is utilized to restore MVLDR images.

$$\varphi = \phi_g \Re \tag{6}$$

Exposure levels are dependent on multiple camera parameters sensor quality, zoom level, lens aperture, and shutter speed, including various other camera features. In theory, change in exposure time may have advantages and disadvantages for the image's amount of noise. By observing a sequence of exposures, we analyzed that short exposures are more vulnerable to fluctuations in the number of photons with respect to long exposures. In order to create HDR, the darkest frame is captured with an exposure of 1/8000 of a second, pretending that its noise is also amplified 8000 times [64]. A combination of these noisy frames will result in a more noisy image. At the same time, other problems such as halos and artifacts will also be more prominent in the case of multiple images due to the camera or object motion. Thus, to avoid the unrealistic condition of static image acquisition, which is required to avoid ghosting artifacts, we utilized a single input image strategy for each viewpoint. We have estimated the exposure gain in Eq. (3), which enables us to restore the MVLDR images. Considering the variational exposures in the same image, we utilized learning base restoration [11], which better handles the segmentation by utilizing the SVM classifier. This process results in the restoration of the output MVLDR-images with highlighted and shadow regions due to the robust and non-uniform illumination distribution. The unwanted artifacts and distortions in the structure and texture degrade the quality of the restored MVLDR images. In order to limit these artifacts, we produce virtual images and fuse them to achieve the desired quality.

D. VIRTUAL IMAGE GENERATION

The under-exposed, intermediate, and over-exposed images are required with known exposure time to acquire the complete dynamic range of the images. Recovery of MVHDRI from MVLDRI generally requires a camera response function (CRF) estimated through information acquired from these available multiple exposures. In the case of multiple images with multiple exposures (X_s), a global alignment is required to fulfil image registration requirements.

It is important to note that our scenario involves input images with unknown variational exposure and geometry. So far, we have solved the problem of geometric alignment and restored the MVLDR images for each viewpoint based on our IRF function. Now, in order to complete our fusion process, we synthesize virtual images with an intermediate exposure to remove color imbalances. It solves the problem

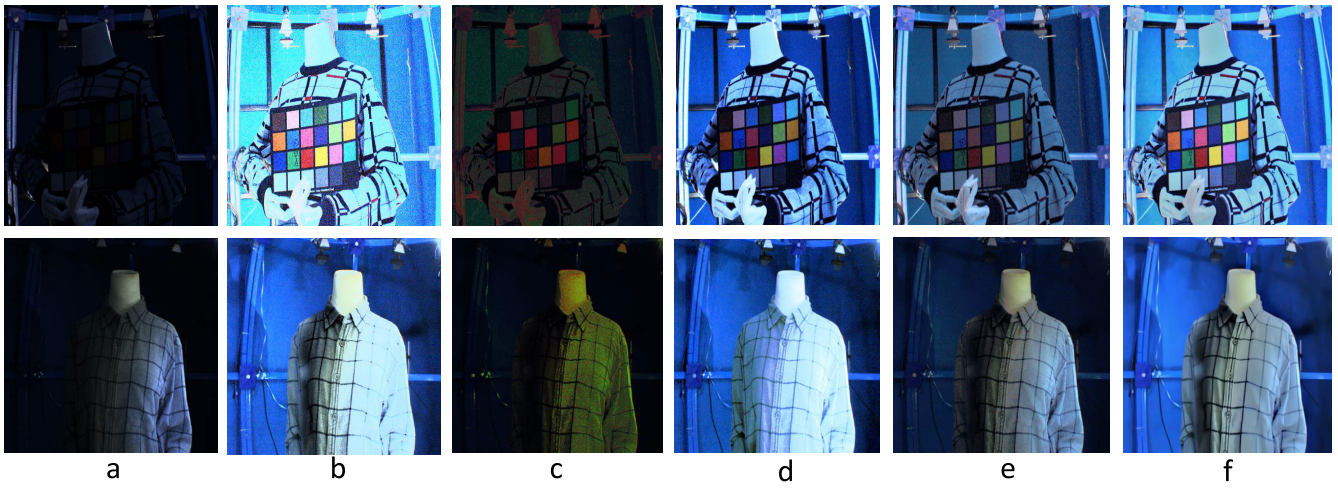


FIGURE 4. The generation of virtual views: (a) input model's images (b) restored images, (c) and (d) are two intermediate virtual images and (e) is the final virtual image, and (f) is the final output.

of highlighted and shadow regions for the large exposure ratio [65]. We use a *model* view-3, shown in Fig. (4) in the first row and another example *S-model* image in the second row with a partial lit condition in Fig. (4) to generate the virtual views (c) and (d) in the first and second row for these two images respectively. The final virtual image (e) is acquired as a weighted fusion of these two virtual images. The restored views (b) in the first and second row contain an over-exposure effect, which saturates the color and black-lining on the cloth. In contrast, the final image (f) represents a vivid color with complete details.

We utilized input images and restored MVLDRI to produce two intermediate exposure virtual images from higher exposure to a lower one and from lower exposure to a higher one following the intensity mapping function (IMF) [66]. There exist one-to-many mapping correspondences for dark to bright images for all over-exposed pixels, and similarly, a one-to-many mapping from bright to dark images for all under-exposed pixels [65], [67]. Our two images of exposure e_1 and e_2 with an exposure ratio $k := e_2/e_1$ are expressed for the respective irradiance E_1 and E_2 with the relationship, $E_2 = k(E_1)$ [66]. Intensity measure different exposures for the same scene is the constrained information used to recover the response function [65]. The IMF recovery through a cross histogram analysis is called a compagram [68]. Intensity values are also associated with many other factors, such as noise, spatial quantization, and saturated pixels. Currently by ignoring these factors we calculate IMF and elaborate it with a multiview input image for the proposed three views represented as Z_1 to $\widetilde{Z}_1$, Z_2 to $\widetilde{Z}_2$, and Z_3 to $\widetilde{Z}_3$ and expressed for n-view points (Z_i to $\widetilde{Z}_i$) as follows.

$$Z_i(p) = \vee_i(\widetilde{Z}_1); \quad whereas : i = 1 \ldots N, \tag{7}$$

In order to produce the desired virtual image per view with an intermediate exposure, we utilized the input images along-with restored MVLDR images. We require two IMFs, which are produced from Z_i and $\widetilde{Z}_i$ for each viewpoint. Consider $\vee_{v_i}$ to be the virtual images per view with $\vee_i, i = 1, 2$ representing two intermediate virtual images, generated via IMF to avoid color distortions. A washout appears by averaging the over-exposed pixels, which results in a loss in dynamic range and image details. An effective weighting strategy is utilized following the scheme presented in [64] with weights $w_1(z) \rightarrow \widetilde{w_1(z)}$ of two virtual images adjusted for our scenario. The $\vee_i(.)$ represents virtual images with $\vee_1(.), \vee_2(.)$, which are two intermediate virtual images acquired via IMF calculated for an intermediate exposure time by using input and restored multiview images of each view point. Consider the MV-system IMF per view points (v_i, for i = 1…N) for Z_{v_i} to $\widetilde{Z_{v_i}}$ is $\vee_{v_{i,}13}$ and $\vee_{v_{i,}23}$ with exposure time Δt_i with values t_1, t_2, t_3, for the three viewpoints for each corresponding input to the restored image exposures $\widetilde{t_1}, \widetilde{t_2}, \widetilde{t_3}$. For every two images per view for Z_{v_i} to $\widetilde{Z}_{v_i}$ exposure time for intermediate virtual image is estimated as $\Delta t'_{3_{V_i}} = \sqrt{\Delta t_1 . \Delta t_2}$, whereas for proposed scenario $\Delta t'_{3_{v_1}}$, it depicts corresponding intermediate exposures for view points v_i ranges for $v_1, v_2\ v_3$, extendable for v_n. If camera response function (CRF) is F(.) IMF can be estimated by utilizing F(.), whereas it is a monotonically increasing function. The IMF from Z_{v_1} to an intermediate image Λ and from $\widetilde{Z}_{v_1}$ to intermediate image Λ is $\vee_{13}$ and $\vee_{23}$ respectively [66].

$$\begin{cases} \vee_{13}(z) = F(\Delta t_3 / \Delta t_1 F^{-1}(z)) \\ \vee_{23}(z) = F(\Delta t_3 / \Delta t_2 F^{-1}(z)) \end{cases} \tag{8}$$

IMF is one to many mapping from dark to bright and bright to dark images [68] for the underexposed pixels, hence in order to avoid the color distortions two intermediate virtual images are generated and fused following the weighting function in the [64] with the tradeoff parameter $\xi_U = 250$ and $\xi_L = 50$

for our weight adjustments.

$$w_1(z) = \begin{cases} 0; & \text{if } 0 \le z < \xi_L \\ 1 - 3{h_1}^2(z) + 2{h_1}^3(z); & \text{if } \xi_L \le z < 128 \\ 1; & \text{otherwise.} \end{cases} \quad (9)$$

$$\widetilde{w_1}(z) = \begin{cases} 1; & \text{if } 0 \le z < 255 \\ 1 - 3{h_1}^2(z) + 2{h_1}^3(z); & \text{if } 200 \le z < \xi_U \\ 0; & \text{otherwise} \end{cases} \quad (10)$$

$$h_1(z) = \frac{128 - z}{128 - \xi_L}, \quad h_2(z) = \frac{z - 200}{\xi_U - 200} \quad (11)$$

The desired intermediate exposure virtual image is produced by fusing these weights as follows;

$$\Lambda(p)|_{v_i} = \frac{\sum_{i=1}^{2} \widetilde{w}_i(Z_i(p))\widetilde{Z}_i(p)}{\sum_{i=1}^{2} \widetilde{w}_i(Z_i(p))} \quad (12)$$

The virtual images in Fig. (4) illustrate our scenario for one viewpoint to produce the final image. Every viewpoint in the scenario produces two virtual views, whereas the final intermediate exposure virtual image is obtained as a fusion of these two images. It is also important to note that our work in this article demonstrates the scenario for three viewpoints only, whereas our method is applicable for N-view points with N-cameras.

E. MULTI EXPOSURE FUSION

Recalling the whole process, we now have three images for each viewpoint, including MV-input, MV-restored, and MV-final virtual images. We utilize the fusion process mentioned in Eq. (1) to estimate the weights based on contrast, color, and saturation [2]. Our method uses the guided filter to refine the initial noisy weights and adopt pyramid decomposition to acquire the final image. After acquiring the l-th level of the Laplacian pyramid for each view input image $L\left\{Z_i^l(p)\right\}$ and weight map $W_i^l(p)$, the l-level of the fused pyramid for fused image R_i is acquired for each level and each view.

$$L\{R_i^{(l)}(p)\}|_{v_i=1}^{3} = \sum_{i=1}^{3} W^{(l)}{}_i(p) L\{Z^{(l)}{}_i(p)\} \quad (13)$$

The $L\left\{Z_i^l(p)\right\}$ is collapsed finally to acquire the resultant fused image R_i following Eq. (12). It is of immense importance to note that our method is applicable for N-viewpoints for the multi-camera system.

IV. DATASETS, EXPERIMENTS AND EVALUATIONS

A. THE PROPOSED DATASETS

The target problem involves multiview images captured with a large variational baseline among the sparsely arranged cameras. In order to prove the reliability of the proposed approach, we performed extensive experiments on indoor and outdoor datasets. We simulate the scenario mentioned in the subsection. III-B, to captured indoor and outdoor datasets. The capturing architecture considers the setting of practical scenarios of close-circuit-television-cameras (CCTVs) used for surveillance and monitoring purposes and illustrates the scenarios for daily life standard photography. In the outdoor dataset, we capture the scene from three different viewpoints by using three mobile phone cameras. In this case, the view of vision and light source is shared among multiple non-linearly arranged cameras. The four datasets named as '*boy&moto*' and '*boy*' are captured with a baseline of 100 *cm* among cameras with an approximate angle deviation of 30^0.

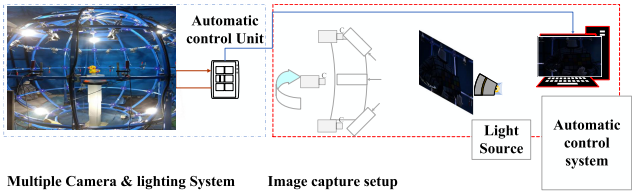


FIGURE 5. The architecture of the multiple camera and lighting system utilized in capturing the indoor datasets.

In order to capture the indoor dataset, we utilize the state of the art multiview camera system (MVCs), with an LED lighting control system shown in Fig. (5). The MVCs consist of two trigger lines to control the lighting-system and camera system. These trigger lines are independently controlled with an external computer system's help to avoid spatial misalignment. We lower the lighting projection behind the scene and create the aforementioned weaklit environment. The scene consists of a backlit *model*, captured from three different viewpoints with extremely short-unknown-exposure time. Firstly, we capture the three sets of *"model-images"* scenes with a baseline distance of 40 cm and an angle deviation of 30 degrees among three cameras. Secondly, we capture three *"emo"* scenes with a baseline of 60 cm, among the cameras with an angle deviation of 30 degrees. We utilized three *MV-CS30G, CREVIS-CAMERAS with 8mm-lens* in a controlled indoor environment. Taking advantage of the controlled environmental setting in our MVCs, we also generate the reference ground truth images (GT) for the indoor datasets. We capture three input exposures, i.e., low, intermediate, and high-exposure, and fuse these exposures via the exposure fusion method [2] to obtain the respective GTs for the *model and emo* images. The data-sets with preliminary implementation guideline are available at our github[1] repository.

B. EXPERIMENTS AND COMPARED METHODS

We perform the experiments using the images captured with multiple sparsely arranged cameras for controlled and uncontrolled scenarios, with variational baseline and angle deviations. The proposed approach utilizes a single input image per viewpoint and relies on the accurate feature points matching. The extant stereo and multiview high dynamic range reconstruction methods rely on disparity map estimation. In contrast, in the target problem, wide baseline and low lighting conditions make the disparity map highly unreliable. Moreover, the methods following traditional HDR

[1] https://github.com/imrizvankhan/MV-EFU

FIGURE 6. Subjective comparison of three views of outdoor boy&moto dataset. (a) Shows original multiview-LDR input images, (b) shows results for SLIMER [6], (c) FEW [27], (d) JED [26], (e) Deep Retinex [8] and (f) Shows results for the proposed method.

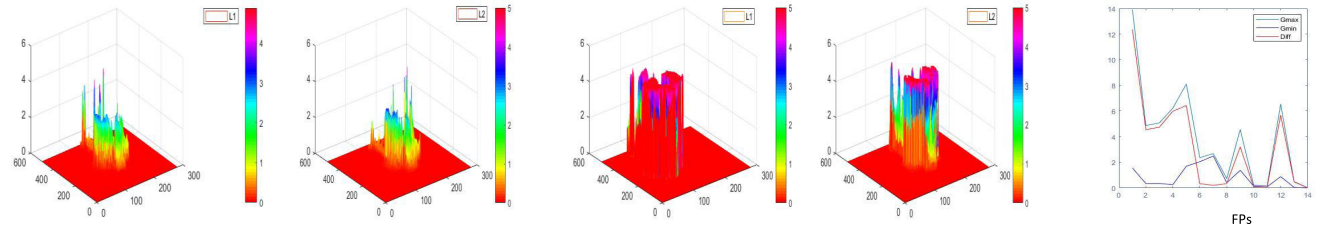


FIGURE 7. (a) shows HSV map for the luminance distribution before processing, (b) shows HSV map for the luminance distribution after processing. (c) Controlling gain among multiple views.

pipelines require more than one input image and require additional exposure time information. The single image base restoration techniques enhance the contrast for the under-exposed images and work for single view enhancement, and are adaptive towards the multiview image enhancement scenarios. We compare our method with state-of-the-art, SLIME [6], JED [26], Deep Retinex [8] and LBR [11], generation of high dynamic range illumination from a single image (GHDRI) [14] and a super-resolution and high dynamic range (SR-HDR) [17], KinD [9], HDR-Net [10], and Zero-DCE [12] methods. The proposed strategy effectively mitigates the irregularities associated with color and contrast distortions and produces a decent balance of smoothness. The HSV maps in Fig. (7 a & b), shows the luminance

distribution for view1 and view2 of *boy & moto* multiview LDR input images before and after the proposed strategy. The control over the intensity shown in Fig. (7. c) is obtained via exposure gain, based on the matched feature points. The overall experimental evaluation proves the significance and superiority of the proposed approach.

C. SUBJECTIVE AND OBJECTIVE EVALUATIONS

In this section, we demonstrate the performance of the proposed method through subjective and objective comparison on our indoor and outdoor datasets. The extant stereo and multiview HDR methods can hardly estimate the disparity map for the large angle and baseline among the cameras; therefore, these methods fail in the target problem. We compare our method with FEW [27], SLIMER [6], JED [26], GHDRI [14], Retinex [8], KinD [9] and SR-HDR [17] methods. We evaluate the performance of the proposed method in controlled and uncontrolled conditions by using indoor and outdoor datasets respectively. In the case of indoor scenario we design the reference GTs, therefore we relies on the reference image based quality assessment metrics, i.e., structure similarity index measure (SSIM) and peak signal to noise ratio (PSNR) [69]. We also utilized no reference image quality assessment metric for contrast distorted images (NQAC) [70] and a completely blind non-reference image quality estimation metric (NIQE) [71]. The lower value of NIQE ranks higher, whereas higher values of NQAC, PSNR, and SSIM ranks higher and vice versa. In the case of outdoor scenario we utilize two datasets, i.e., *boy&moto and boy*, and present the subjective comparison with SLIME [6], FEW [27] and JED [26] methods in Fig. (6) and Fig. (8). The corresponding objective quality metrics are shown in Tab. (1). In the case of indoor scenarios, we utilized the datasets *model* and *emo* captured by our MVCs. The comparison for corresponding subjective evaluation for our indoor *MVCs model-images* dataset is shown in Fig. (9) with respective objective evaluation metrics shown in Table. (2). Moreover, we illustrate the subjective quality by showing zoomed in fragments for view-3 of model-image in Fig. (10). We generalize our approach by using nine publicly available off-the-shelf datasets form Middlebury.[2] The datasets of Middlebury contain images captured from seven viewpoints (view0 to view6) with different illumination conditions (Illum), and exposures (Exp). The target problem is designed to handle the large baseline; therefore we select the possible distant views (i.e., view0, view3, view6). We select the images with lowest available exposure which include *Art Illum1/Exp0;(view0,view3 and view6)*, *Dwarves Illum1/Exp0;(view0,view3 and view6)* and *Drumsticks Illum1/Exp0;(view0,view3 and view6)*. The comparison with state of the art approaches, FEW [27], JED [26], SLIMER [6] and SR-HDR [17], is presented in term of subjective and objective metrics. The subjective analysis on *Art*, *Dwarves* and *Drumstick* images is shown in Fig. (13),

[2]https://vision.middlebury.edu/stereo/data

FIGURE 8. Comparison of outdoor scene *boy* with various other methods.

and the respective objective evaluation metrics are shown in Tab. (3). In order to generalize the efficiency of the proposed approach, we provide a quantitative comparison for the objective quality metrics in Fig. (12). The graph depicts the robustness of the proposed approach for reference and non-reference based assessments metrics. Furthermore we conduct extensive experiments to compare ours method with deep learning-based methods and present the average objective evaluation metrics in Table. (4) for KinD [9], HDR-Net [10], Retinex [8] and Zero-DCE [12] and ours method. The overall analysis shown in Fig. (12) and quantitative analysis for deep learning-based methods in Fig. (15) on indoor datasets demonstrate the superiority of the proposed approach. Additionally, the subjective comparison of the proposed approach with deep learning-based methods on two model images is shown in Fig. (14), demonstrating that the proposed method outperformed the others. The proposed method maintains the balance of color and contrast and recovers most details in the dark regions in term of brightness and visibility.

D. DISCUSSION

The overall comparison demonstrates that the proposed method outperformed state-of-the-art approaches in terms of subjective and objective evaluations. Our method preserves the spatial details with pleasant visual quality. In general, in the case of coexisting over-exposed and under-exposed regions, the over-exposed pixel saturates, and the under-exposed one suffers under enhancement. The resulting image

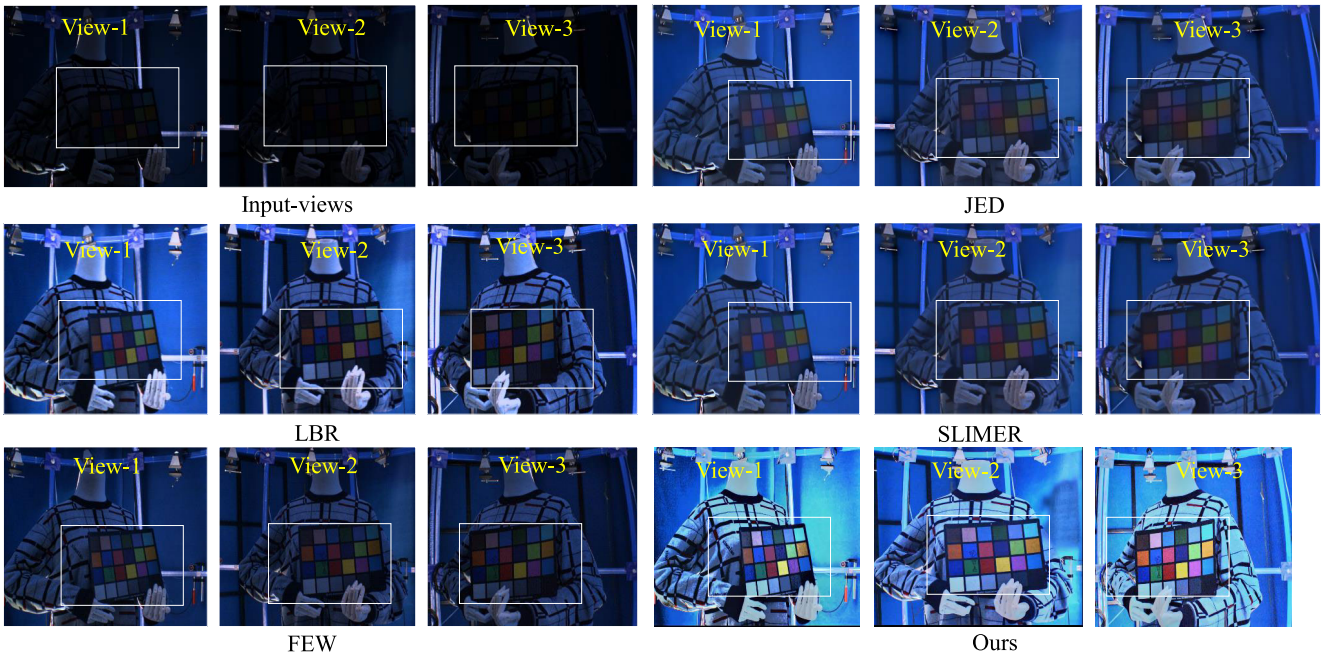


FIGURE 9. Comparison on model image views, first three images are three multiview low-dynamic-range input images. Next shows the comparison of JED, LBR, SLIMER, FEW and Ours.

TABLE 1. Comparison of Objective Quality Metrics, NIQE and NQAC for FEW, JED, SLIMER and Ours for Multiview Input Images (View1, View2, View3).

Object:Boy&Moto	View1		View2		View3	
Methods	NIQE	NQAC	NIQE	NQAC	NIQE	NQAC
FEW	4.5109	2.8579	3.1779	2.9300	3.5882	2.7345
JED	5.5879	2.8944	4.6129	3.0855	5.1835	2.9684
SLIMER	6.0641	2.3867	5.9878	2.3242	6.1378	2.9559
Retinex	6.6231	2.8817	5.5545	2.4103	5.5131	2.4930
Ours	**3.5413**	**3.4780**	**2.9539**	**3.7516**	**3.2219**	**3.4730**
Object:Boy	View1		View2		View3	
Methods	NIQE	NQAC	NIQE	NQAC	NIQE	NQAC
FEW	2.9806	2.9896	2.9957	2.6927	2.8957	2.9657
JED	3.3249	2.9351	3.3410	2.9197	3.4037	2.9515
SLIMER	3.3581	2.9287	3.2594	2.9327	3.4890	2.9470
Retinex	4.3742	2.6034	5.5014	2.6523	5.3485	2.8596
Ours	**2.3172**	**3.4796**	**2.0574**	**3.4660**	**2.1281**	**3.5235**

TABLE 2. Average Values of the Objective Quality Evaluation Metrics for the Indoor Dataset Sample Input *model views* and *emo* Images.

Dataset	model-image-view1,view2, view3			emo-view1,view2, view3		
Methods	Average PSNR	Average SSIM	Average NIQE	Average PSNR	Average SSIM	Average NIQE
SLIMER	13.7365	0.6948	3.4554	17.6902	0.6073	3.5314
FEW	13.7730	0.7315	4.8097	14.5237	0.5931	3.9865
JED	12.1764	0.6404	3.9801	14.7358	0.6035	3.9538
LBR	11.7095	0.5538	4.5735	15.1210	0.4524	4.3754
IBHDR	11.1723	0.4261	5.5735	12.7536	0.3127	5.5023
Ours	**17.2647**	**0.7504**	**2.9376**	**17.8912**	**0.6351**	**2.9867**

contains highlight and shadow regions with contrast distortion. In contrast, we generate intermediate virtual views to handle these challenges. The traditional HDR methods and exposure fusion methods require multiple images of the same scene with the static acquisition's strict conditions, otherwise ghosting artifacts appear. Multiview image enhancement

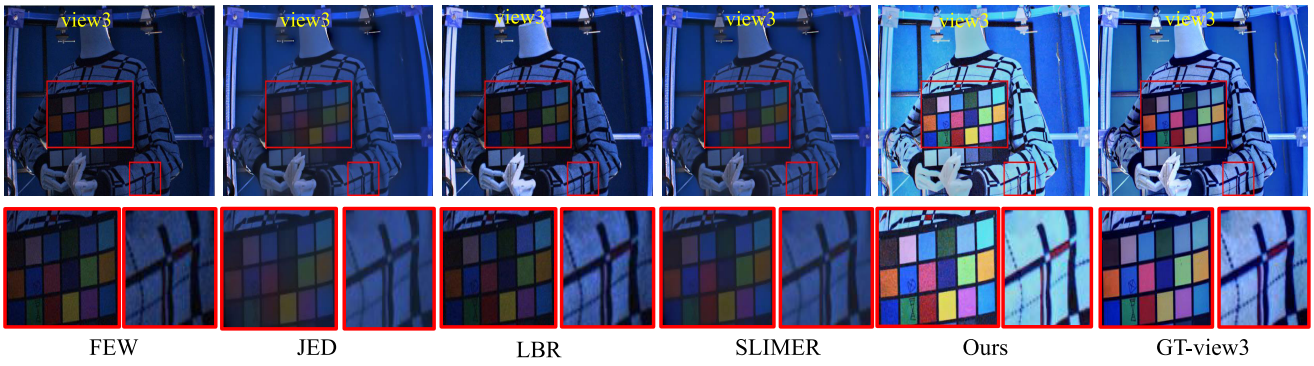


FIGURE 10. Subjective comparison of ours method with zoomed-in fragments of *model* with various other methods.

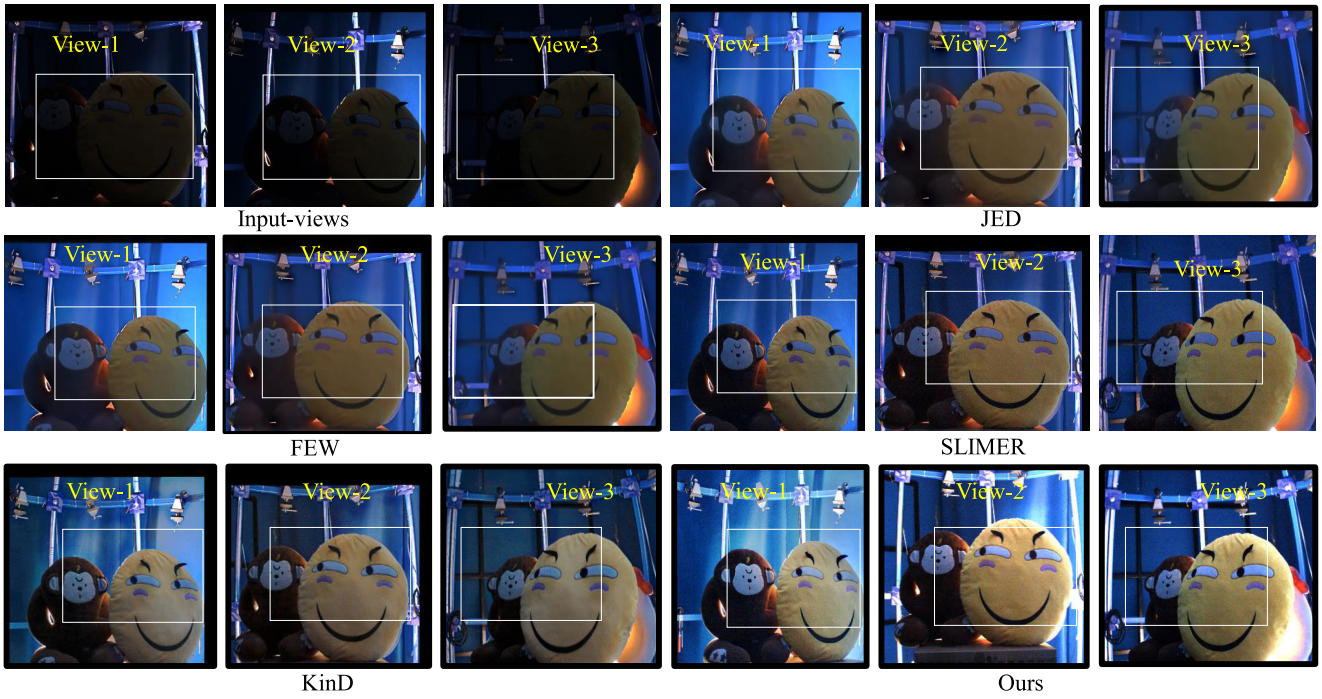


FIGURE 11. Subjective comparison of the indoor dataset, the input *emo images* results showing JED [26], FEW [27], SLIMER [6], a deep learning method KinD [9] and Ours.

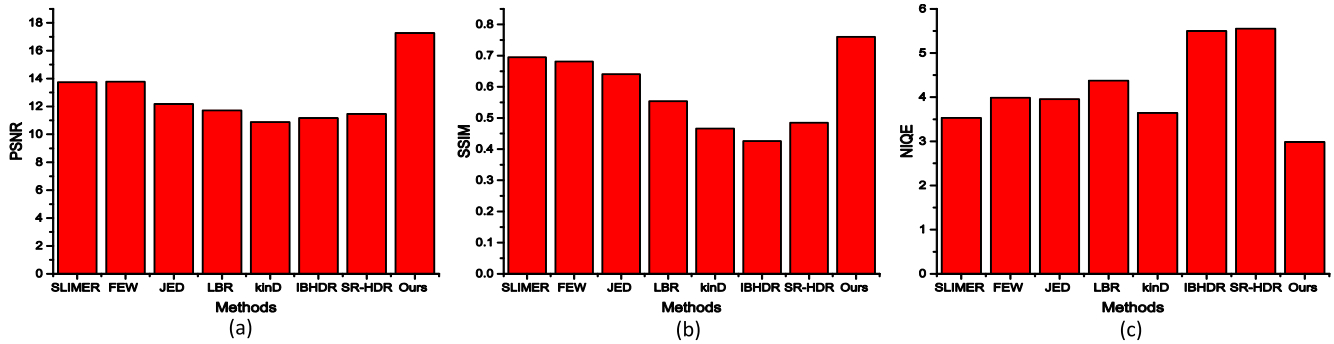


FIGURE 12. Quantitative analysis: showing the robustness of the proposed approach on all dataset images compared with state-of-the arts, (a) overall PSNR (b) SSIM (c) NIQE.

methods, whether traditional or deep learning base recent methods, all failed when it comes to estimating the disparity map for a wide varitaional baseline among the cameras. Moreover, the ill-lighting condition contributes to the holes and artifacts, which distort the final image's quality. The competitor methods are the enhancement methods, which either

FIGURE 13. Subjective comparison on the input Dwarve images from Middleburry datasets. Results shows, FEW method [27], SLIMER [6], JED [26] and proposed.

TABLE 3. Comparison of Objective Quality Metrics. Average PSNR and SSIM Evaluations for FEW, JED, SLIMER and Ours for Multi-View Input Images (View0, View3, View6) Using Middlebury Datasets.

Metrics	Average PSNR(3views)			Average SSIM(3views)		
Objects	Dwarves	Art	Drumstsick	Dwarves	Art	Drumstsick
FEW	12.7199	12.3265	12.7175	0.5790	0.5730	0.6384
JED	15.2069	15.9337	16.9011	0.75186	0.8274	0.8762
SLIMER	14.9128	16.2461	13.5002	0.7726	0.8528	0.5423
SR-HDR	11.4629	11.0286	12.0388	0.391833	0.413933	0.4845
Ours	**17.9632**	**17.8634**	**18.6896**	**0.7962**	**0.8692**	**0.8849**

TABLE 4. Comparison of Objective Quality Metrics on Indoor Datasets, Using Deep Learning Based KinD [9], HDR-Net [10], Retinex [8] Zero-DCE [12] and Ours.

Methods	kinD	HDR-Net	Retinex	Zero-DCE	**Ours**
SSIM	0.5092	0.4632	0.4568	0.6532	**0.7649**
PSNR	12.6894	10.5481	11.6654	15.7592	**18.2479**

suffer under enhancement or produce saturated regions for over-exposure. For example, the comparison on *boy&moto* in Fig. (6) demonstrates that SLIMER [6], FEW [27], and JED [26] are almost unable to preserve the image details for the outdoor low light scenario. On the other hand, the first three images in the third row in Fig. (6) show the deep learning-based deep-retinex method results. In this case, the artificial texture transformation distorts the contrast, texture, and image structure. The retinex approaches work on the principle of image decomposition and split input image into reflection and illumination. It is a quite reasonable approach to handle the illumination and reflection independently, but inaccurate illumination smoothness diverts the noises on reflection. The author in [6] propose a robust retinex model to handle the low lighting conditions but produce inconsistent reflection, limiting the quality of the final image. Therefore in the case of Retinex approaches, the contrast of the final image is distorted completely where the resulting images are over-textured with artificial effects. To better understand the color and contrast preservation, we utilize the color chart with our model image. The proposed method's vivid color preservation demonstrates that our method is capable of handling the color and contrast distortion effectively for the ill-illumination conditions, with largely varying exposures in the same scene. The comparison on the second indoor scene *emo* shows a back-lighting condition, where

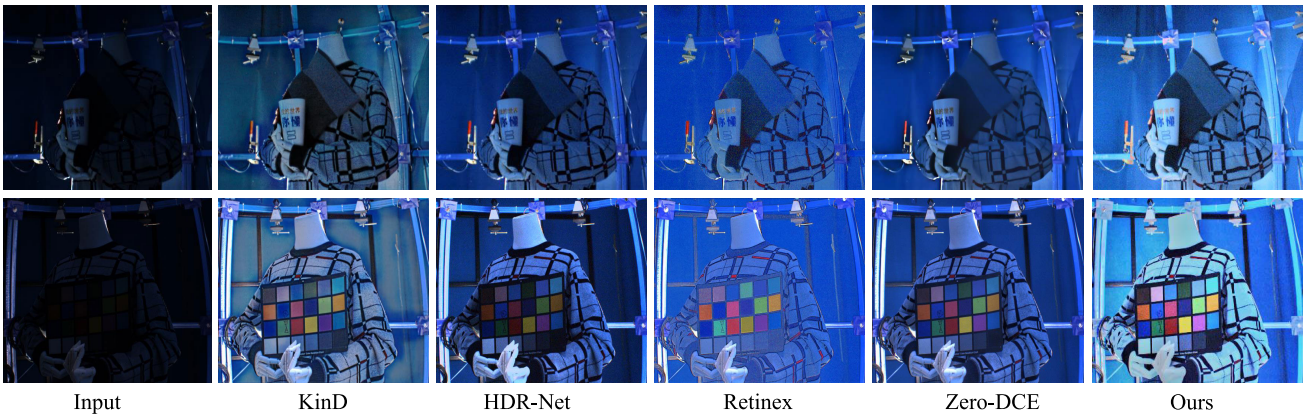


FIGURE 14. Subjective comparison on the model images. Results shown on deep-learning based methods, KinD [9], HDR-Net [10], Retinex [8], Zero-DCE [12] and Ours.

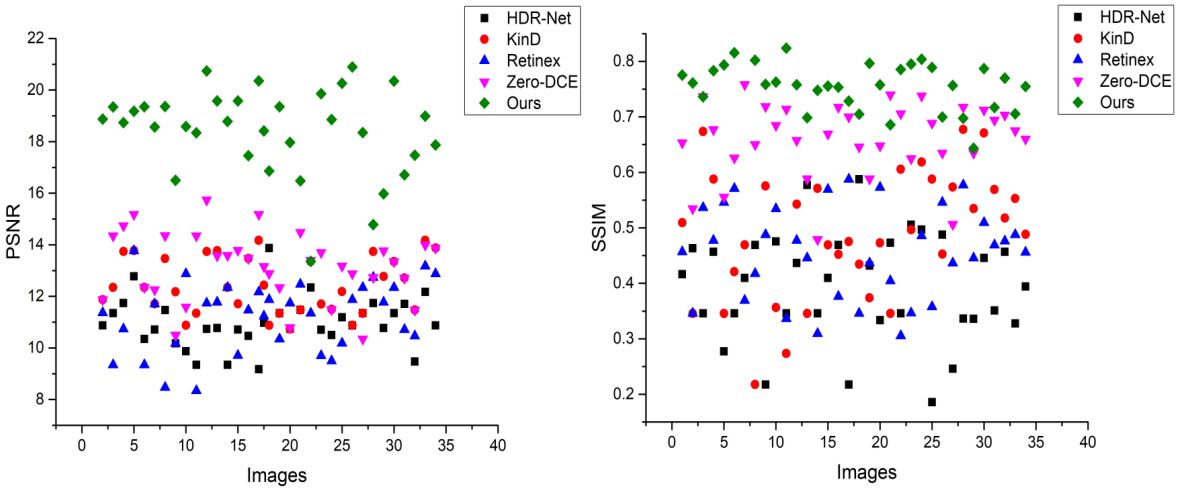


FIGURE 15. Quantitative analysis; comparison of objective metrics using various deep learning based methods, showing the robustness of the proposed approach on various indoor datasets.

the sharp point illumination source produces a large exposure region. The traditional enhancement methods fail in this scenario, and our method preserves the details without any over or under enhancement. The comparison on the input *emo-image* in Fig. (11) demonstrates that the competitors, JED, FEW, and SLIMER, and kinD hardly restore the image details. The deep learning-based method KinD [9] causes some imbalance in the background and washout in some pixels of the *emo* image, whereas our method preserves the details efficiently. Similarly, in the case of Middlebury datasets, the JED, FEW, and SLIMER, and the SR-HDR methods fails to preserve the details for the under-exposed regions. The subjective and objective results on the indoor and outdoor datasets demonstrate that the proposed method outperformed state-of-the-art approaches. Furthermore, during extensive experiments we noticed that the network based approaches hardly preserve the color and contrast for the robust exposure variations in the same scene. For example, the HDR-Net is designed in a heuristic manner and learn to enhance from synthesized training data. The Retinex model follows a decomposition approach but produces unnatural results, whereas HDR-Net seems unable to recover the dark regions. The proposed method recover the details in the under-exposed regions and maintain a balance of color and contrast for the robust exposure variations. Although the time complexity of the deep learning-based methods is less, but unfortunately the adversity of obtaining large scale and even paired training datasets limits the practical implication of these methods. We proposed a multiview enhancement method that is more adaptive to variational exposures and utilizes one image per viewpoint to produce a refined output image. The comparison demonstrates that the proposed method achieves a decent and pleasing visual quality more natural than the state-of-the-art approaches.

V. CONCLUSION

In this article, we proposed a method for the enhancement of ill-lit images captured with multiple sparsely

arranged cameras. The input images are captured with an unknown short exposure time and geometry. We produce multiview images with a higher dynamic range by using multiview low dynamic range ill-illuminated images. The traditional stereo and multiview methods rely on disparity map estimation, suspected to be almost impossible for wide baseline, large-angle deviations, and ill-lit conditions. We handle the complex geometric alignment of multiview cameras with rectification and calibration of the multiview low dynamic range images to acquire accurate features. We detect and match accurate feature points and estimate exposure gain to design an image restoration function. We dispense the challenges of color and contrast distortions via generating virtual views and follow exposure fusion strategy to obtain the final image. The proposed method can handle the challenges of multiple views captured with variational baselines and angle deviations without relying on disparity maps. We perform various experiments in controlled and uncontrolled scenarios on indoor and outdoor datasets. The comparison with state-of-the-art approaches demonstrates that our method outperformed and is more suitable for future high dynamic range capturing and interactive visual display applications.

REFERENCES

[1] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. ACM SIGGRAPH Classes*, 2008, p. 31.

[2] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion: A simple and practical alternative to high dynamic range photography," *Comput. Graph. Forum*, vol. 28, no. 1, pp. 161–171, 2009.

[3] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Comput. Vis., Graph., Image Process.*, vol. 39, no. 3, pp. 355–368, 1987.

[4] H. Farid, "Blind inverse gamma correction," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1428–1433, 2001.

[5] E. H. Land, "The retinex theory of color vision," *Sci. Amer.*, vol. 237, no. 6, pp. 108–129, 1977.

[6] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo, "Structure-revealing low-light image enhancement via robust retinex model," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2828–2841, Jun. 2018.

[7] X. Guo, Y. Li, and H. Ling, "LIME: Low-light image enhancement via illumination map estimation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982–993, Feb. 2017.

[8] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *Proc. Brit. Mach. Vis. Conf.*, 2018, pp. 1–12.

[9] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," 2019, *arXiv:1905.04161*. [Online]. Available: http://arxiv.org/abs/1905.04161

[10] M. Gharbi, J. Chen, J. T. Barron, S. W. Hasinoff, and F. Durand, "Deep bilateral learning for real-time image enhancement," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–12, Jul. 2017.

[11] Z. Li and X. Wu, "Learning-based restoration of backlit images," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 976–986, Feb. 2018.

[12] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1780–1789.

[13] S. Lee, G. H. An, and S.-J. Kang, "Deep chain HDRI: Reconstructing a high dynamic range image from a single low dynamic range image," *IEEE Access*, vol. 6, pp. 49913–49924, 2018.

[14] J. S. Park, J. W. Soh, and N. I. Cho, "Generation of high dynamic range illumination from a single image for the enhancement of undesirably illuminated images," *Multimedia Tools Appl.*, vol. 78, no. 14, pp. 20263–20283, Jul. 2019.

[15] S. Wu, J. Xu, Y.-W. Tai, and C.-K. Tang, "Deep high dynamic range imaging with large foreground motions," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 117–132.

[16] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 136–144.

[17] J. S. Park, J. W. Soh, and N. I. Cho, "High dynamic range and super-resolution imaging from a single image," *IEEE Access*, vol. 6, pp. 10966–10978, 2018.

[18] A. Troccoli, S. B. Kang, and S. Seitz, "Multi-view multi-exposure stereo," in *Proc. 3rd Int. Symp. 3D Data Process., Vis., Transmiss. (3DPVT)*, Jun. 2006, pp. 861–868.

[19] R. Orozco, C. Loscos, I. Martin, and A. Artusi, "HDR multiview image sequence generation: Toward 3D HDR video," in *High Dynamic Range Video*. Amsterdam, The Netherlands: Elsevier, 2017, pp. 61–86.

[20] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2004, pp. 25–36.

[21] Z. Chen, H. Jin, Z. Lin, S. Cohen, and Y. Wu, "Large displacement optical flow from nearest neighbor fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2443–2450.

[22] N. K. Kalantari and R. Ramamoorthi, "Deep high dynamic range imaging of dynamic scenes," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–144, 2017.

[23] S. M. O'Malley, "A simple, effective system for automated capture of high dynamic range images," in *Proc. 4th IEEE Int. Conf. Comput. Vis. Syst. (ICVS)*, 2006, p. 15.

[24] F. Kou, Z. Li, C. Wen, and W. Chen, "Multi-scale exposure fusion via gradient domain guided image filtering," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2017, pp. 1105–1110.

[25] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and A. C. Bovik, "Single-scale fusion: An effective approach to merging images," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 65–78, Jan. 2017.

[26] X. Ren, M. Li, W.-H. Cheng, and J. Liu, "Joint enhancement and denoising method via sequential decomposition," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2018, pp. 1–5.

[27] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, and J. Paisley, "A fusion-based enhancing method for weakly illuminated images," *Signal Process.*, vol. 129, pp. 82–96, Dec. 2016.

[28] 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," *J. Digit. Imag.*, vol. 11, no. 4, p. 193, Nov. 1998.

[29] A. M. Reza, "Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement," *J. VLSI Signal Process.-Syst. Signal, Image, Video Technol.*, vol. 38, no. 1, pp. 35–44, Aug. 2004.

[30] T. Celik, "Two-dimensional histogram equalization and contrast enhancement," *Pattern Recognit.*, vol. 45, no. 10, pp. 3810–3824, Oct. 2012.

[31] W. Wang, X. Wu, X. Yuan, and Z. Gao, "An experiment-based review of low-light image enhancement methods," *IEEE Access*, vol. 8, pp. 87884–87917, 2020.

[32] S.-D. Chen and A. R. Ramli, "Minimum mean brightness error bi-histogram equalization in contrast enhancement," *IEEE Trans. Consum. Electron.*, vol. 49, no. 4, pp. 1310–1319, Nov. 2003.

[33] J.-Y. Kim, L.-S. Kim, and S.-H. Hwang, "An advanced contrast enhancement using partially overlapped sub-block histogram equalization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 4, pp. 475–484, Apr. 2001.

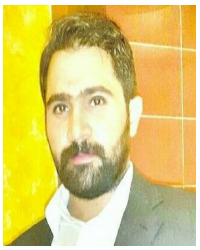
[34] M. Abdullah-Al-Wadud, M. H. Kabir, M. A. A. Dewan, and O. Chae, "A dynamic histogram equalization for image contrast enhancement," *IEEE Trans. Consum. Electron.*, vol. 53, no. 2, pp. 593–600, May 2007.

[35] Z. Xu, X. Liu, and N. Ji, "Fog removal from color images using contrast limited adaptive histogram equalization," in *Proc. 2nd Int. Congr. Image Signal Process.*, Oct. 2009, pp. 1–5.

[36] G. Yadav, S. Maheshwari, and A. Agarwal, "Contrast limited adaptive histogram equalization based enhancement for real time video system," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2014, pp. 2392–2397.

[37] S.-C. Huang, F.-C. Cheng, and Y.-S. Chiu, "Efficient contrast enhancement using adaptive gamma correction with weighting distribution," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 1032–1041, Mar. 2013.

[38] Y. Chang, C. Jung, P. Ke, H. Song, and J. Hwang, "Automatic contrast-limited adaptive histogram equalization with dual gamma correction," *IEEE Access*, vol. 6, pp. 11782–11792, 2018.
[39] E. H. Land and J. J. McCann, "Lightness and Retinex theory," *J. Opt. Soc. Amer.*, vol. 61, no. 1, pp. 1–11, 1971.
[40] 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 Trans. Image Process.*, vol. 6, no. 7, pp. 965–976, Jul. 1997.
[41] T. F. Chan, J. Shen, and L. Vese, "Variational PDE models in image processing," *Notices AMS*, vol. 50, no. 1, pp. 14–26, 2003.
[42] R. Kimmel and N. A. Sochen, "Geometric-variational approach for color image enhancement and segmentation," in *Proc. Int. Conf. Scale-Space Theories Comput. Vis.* Cham, Switzerland: Springer, 1999, pp. 294–305.
[43] S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3538–3548, Sep. 2013.
[44] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.
[45] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3291–3300.
[46] A. Akram, N. Wang, J. Li, and X. Gao, "A comparative study on face sketch synthesis," *IEEE Access*, vol. 6, pp. 37084–37093, 2018.
[47] A. Akram, N. Wang, X. Gao, and J. Li, "Integrating GAN with CNN for face sketch synthesis," in *Proc. IEEE 4th Int. Conf. Comput. Commun. (ICCC)*, Dec. 2018, pp. 1483–1487.
[48] D. C. Garcia, C. Dorea, and R. L. de Queiroz, "Super resolution for multiview images using depth information," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 9, pp. 1249–1256, Sep. 2012.
[49] L. Li, R. Wang, W. Wang, and W. Gao, "A low-light image enhancement method for both denoising and contrast enlarging," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 3730–3734.
[50] K. G. Lore, A. Akintayo, and S. Sarkar, "LLNet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognit.*, vol. 61, pp. 650–662, Jan. 2017.
[51] S. K. Nayar and T. Mitsunaga, "High dynamic range imaging: Spatially varying pixel exposures," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2000, pp. 472–479.
[52] M. Bätz, T. Richter, J.-U. Garbas, A. Papst, J. Seiler, and A. Kaup, "High dynamic range video reconstruction from a stereo camera setup," *Signal Process., Image Commun.*, vol. 29, no. 2, pp. 191–202, Feb. 2014.
[53] H.-Y. Lin and W.-Z. Chang, "High dynamic range imaging for stereoscopic scene representation," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 4305–4308.
[54] E. Selmanović, K. Debattista, T. Bashford-Rogers, and A. Chalmers, "Generating stereoscopic HDR images using HDR-LDR image pairs," *ACM Trans. Appl. Perception*, vol. 10, no. 1, pp. 1–18, Feb. 2013.
[55] Y. Chen, G. Jiang, M. Yu, Y. Yang, and Y.-S. Ho, "Learning stereo high dynamic range imaging from a pair of cameras with different exposure parameters," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1044–1058, 2020.
[56] W. A. Bruls and R. J. Muijs, "Generation of high dynamic range images from low dynamic range images in multiview video coding," U.S. Patent 9 098 906, Aug. 4, 2015,
[57] J. Bonnard, C. Loscos, G. Valette, J.-M. Nourrit, and L. Lucas, "High-dynamic range video acquisition with a multiview camera," in *Proc. 2nd Opt., Photon., Digit. Technol. Multimedia Appl.*, vol. 8436, 2012, Art. no. 84360A.
[58] J.-C. Chiang, P.-H. Kao, Y.-S. Chen, and W.-R. Chen, "High-dynamic-range image generation and coding for multi-exposure multi-view images," *Circuits, Syst., Signal Process.*, vol. 36, no. 7, pp. 2786–2814, Jul. 2017.
[59] K. R. Prabhakar, V. S. Srikar, and R. V. Babu, "DeepFuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4724–4732.
[60] J. Cai, S. Gu, and L. Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 2049–2062, Apr. 2018.
[61] G. Ekman, "Weber's law and related functions," *J. Psychol.*, vol. 47, no. 2, pp. 343–352, Apr. 1959.
[62] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[63] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[64] A. O. Akyüz and E. Reinhard, "Noise reduction in high dynamic range imaging," *J. Vis. Commun. Image Represent.*, vol. 18, no. 5, pp. 366–376, Oct. 2007.
[65] Y. Yang, W. Cao, S. Wu, and Z. Li, "Multi-scale fusion of two large-exposure-ratio images," *IEEE Signal Process. Lett.*, vol. 25, no. 12, pp. 1885–1889, Dec. 2018.
[66] M. D. Grossberg and S. K. Nayar, "Determining the camera response from images: What is knowable?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 11, pp. 1455–1467, Nov. 2003.
[67] Z. Li, J. Zheng, Z. Zhu, and S. Wu, "Selectively detail-enhanced fusion of differently exposed images with moving objects," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4372–4382, Oct. 2014.
[68] S. Mann, "Comparametric imaging: Estimating both the unknown response and the unknown set of exposures in a plurality of differently exposed images," in *Proc. CVPR*, vol. 1, 2001, pp. 842–849.
[69] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[70] J. Yan, J. Li, and X. Fu, "No-reference quality assessment of contrast-distorted images using contrast enhancement," 2019, *arXiv:1904.08879*. [Online]. Available: http://arxiv.org/abs/1904.08879
[71] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Nov. 2012.



RIZWAN KHAN received M.Eng. degree from the School of Electronic and Electrical Engineering, Superior University Lahore, Pakistan, in 2016. He is currently pursuing the Ph.D. degree with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China. He is also working with the Wuhan National Laboratory of Optoelectronics, Wuhan. His current research interests include computer vision, signal processing, image processing, machine learning, and deep learning.

ADEEL AKRAM received the M.S. degree in information technology from the Institute of Information Technology, Quaid-i-Azam University, Islamabad, Pakistan, in 2012, and the Ph.D. degree in computer science and technology from the School of Computer Science and Technology, Xidian University, Xi'an, China, in 2019. He is currently an Assistant Professor with the School of Information Engineering (Big Data), Xuzhou University of Technology, Xuzhou, China. His research interests include image processing, computer vision, pattern recognition, machine learning, and deep learning.

ATIF MEHMOOD received the B.S. degree in computer science from COMSATS University Islamabad, Pakistan, in 2015, and the M.S. degree in computer science from Riphah International University, Islamabad, in 2018. He is currently pursuing the Ph.D. degree in computer science and technology with the School of Artificial Intelligence, Xidian University, Xi'an, China. His research interests include machine learning and medical image processing.