Abstract—An image captured in dark environment usually has ambient illumination, but the image looks dark and noisy. However, the use of flash can introduce unwanted artifacts such as sharp shadows at silhouettes, red eyes, and non-uniform brightness in the image. We propose a new framework to enhance photographs captured in dark environments by combining the best features from a flash and a no-flash image. We use sparse and redundant dictionary learning based approach to denoise the no-flash image. A weighted least squares framework is used to transfer sharp details from the flash image into the no-flash image. We show that our approach is simple and able to generate better images than that of the state-of-the-art flash/no-flash fusion method.

I. INTRODUCTION

The practical problem usually faced by professional photographers while capturing photographs under low illumination environment is whether to use the flash or not. Under low illumination setting, an important goal of the photographer is to capture and reproduce the visual richness of a scene. A picture captured in such a dim environment usually has a warm atmosphere due to ambient illumination, but the image becomes dark and noisy if the camera is set to a short exposure time with a high camera gain (see Fig. 1). Thus no-flash image tends to have a relatively low signal to noise ratio (SNR) while it contains the natural ambient lighting of the scene.

Flash photography was invented to mitigate these problems, wherein the scene is illuminated by the artificial light and cameras with shorter exposure time and less sensor gain is used to produce relatively sharp and noise free images of a scene under low illumination [1]. However, the use of flash can introduce unwanted artifacts in the image such as sharp shadows at silhouettes, red eyes, and non-uniform brightness in the image. The basic idea of flash/no-flash photography is to fuse both the flash and no-flash images to get the best representation of the scene.

The primary objective of our approach is the learning of an over-complete dictionary which has been shown to be the state-of-the-art tool for denoising [2]. The color channels of no-flash image generally contain much lesser noise and edges compared to that of the luminance channel. Therefore, we use a simple non-local means technique to remove noise from the color channels [3]. The weighted least squares algorithm is shown to achieve better edge preservation for computational photography applications while not introducing any halo artifacts [4]. We employ this approach for detail transfer from flash image.

In this paper, we present a new framework to enhance images captured in dark scenes by combining the features from flash and no-flash image. Our contributions in this paper are: 1) A new technique for enhancing no-flash images by using sparse and redundant dictionary is proposed. 2) We transfer high frequency details from flash images into no-flash image using a weighted least squares framework.

In section II, we briefly describe the previous work on Flash/no-flash photography. In section III, we discuss the proposed framework to enhance images captured in dark environments. Experimental results are given in section IV along with comparison and discussion. Section V concludes the paper highlighting some challenges ahead.

II. RELATED WORK

Several approaches to solve the aforementioned problem of flash/no-flash photography have been proposed in the literature ([1],[5],[6],[7]). In [5], Eisemann and Durand employ bilateral filter [8] for denoising no-flash image and decompose each images into layers corresponding to the illumination and sharp details respectively. These layers are used to give the ambient tone to the flash image. On the other hand, Petschnigg et al. attempt to denoise no-flash image using joint bilateral filter and transferring details from the flash image to the no-flash image [1]. Agrawal et al. focus on removing flash artifacts by projecting the flash image gradient to the no-flash image gradients but their approach does not account for the no-flash image containing severe noise [9]. Seo and Milanfar use iterative guided filtering approach to denoise the no-flash image, while transferring details from the flash image [7]. Krishnan and
Fergus introduce the concept of dark flash to capture pictures in low-light conditions [6]. They utilised the correlations between images recorded at different spectral bands to denoise the ambient image and restore fine details. Bertalmo and Levine propose a variational method for automatically combining an exposure-bracketed pair of images within a single image that retains the colors and sharp details from the actual scene [10].

The methods proposed in ([1], [5], [6], [7]), ignore the image blurring due to camera shake or scene motion by acquiring the image pairs using a tripod setup or choosing sufficiently fast shutter speed. In practice, however, the images captured under low light conditions using hand-held camera often encounters blur due to the long exposure time and camera shake. Lu Yuan et al. propose an image deblurring approach using residual and gain-controlled deconvolution process to significantly reduce blur present in an image [11].

III. PROPOSED APPROACH

Consider a no-flash image and a flash image (Fig. 3 (a, b)). The proposed approach can be split into three steps - denoising luminance component of no-flash image, detail transfer from flash to no-flash image, and finally addition of processed color components. Fig. 2 shows the steps of the algorithm (larger version provided in the supplementary material).

A. Denoising

We convert the no-flash image in Fig. 3 (a) from RGB to CIE-Lab space. We remove noise from the image in two steps - first from the L component and then from the color channels a and b. The L component of the no-flash image patches $I_{nf}$ each of size $p$ can be denoised using an over-complete dictionary $D \in \mathbb{R}^{P \times k}$. We use K-SVD algorithm for designing an over-complete dictionary ($k > p$) using sparse representation, trained on an initial dictionary constituting the overlapping patches of the L component of $I_{nf}$ [12] as shown in Eq. 1.

$$\min_{D,A} \|I_{nf} - DA\|_F^2 \quad \text{subject to} \quad \|\alpha_i\|_0 \leq N$$ \hspace{1cm} (1)

where $N$ is the maximum number of non-zero coefficients in each column ($\alpha_i$, $i = 1, 2, \ldots, k$) of the coefficient matrix $A$.

We use a scalar bias term $\beta$ that is multiplied to all the elements of coefficient matrix $A$. The motivation for introducing this scalar is that, if there was more light in the scene, details would be better captured. This bias term controls the brightness of the scene and helps to enhance information that might be masked otherwise.

The K-SVD algorithm has a parameter, $\sigma$, which estimates the additive noise in the image. The CCD noise/real noise is not additive and is dependent on various camera parameters such as ISO, shutter speed and aperture. Since it also strongly depends on image intensity level [13], we take the standard deviation of values at each pixel of the image and consider this value to be equal to $\sigma$. The result is the denoised $L$ component of the image as shown in Fig. 3 (c).

B. Detail Transfer

Many small scale details of the objects in the scene are well captured in the flash image due to proper lighting. The flash image is decomposed into a piece-wise smooth base layer and residual detail layer. This filtered detail layer would then be directly added to the denoised no-flash image ($L$ component). While performing the smoothing of the image strong edges become blur, which may introduce halos once the detail layer is merged with our denoised no-flash image. In order to avoid this, we use an edge preserving smoothing operator based on the weighted least squares optimization framework to get an edge preserved image [4]. This coarsened image is then subtracted from the original $L$ component of the flash image $I_f$ (Fig. 3 (e)) to obtain the detail layer which is of interest to us (Fig. 3 (f)). This detail layer is then directly added to the denoised no-flash image ($L$ component) to enhance the details. The next task at hand is the denoising of color channels and concatenating the two channels with this resultant $L$ component to obtain the final result in Lab space.

C. Color processing

Our aim is to retain the color and the ambient illumination of the scene while including details. The no-flash image has the desired colors and hence we use color channels of the no-flash image $I_{nf}$. The non-local means algorithm is applied for this purpose [3]. We use this algorithm for denoising of color components as we do not intend to preserve edges and want to remove only the noise. We do not apply this non local means algorithm to denoising of the $L$ component of no-flash image because it causes loss of both edges as well as information along with noise while averaging over pixel values. After applying non-local means algorithm for denoising we get the image in Fig. 3 (d), both color channels $a$ and $b$ planes are then concatenated with the earlier obtained $L$ channel with enhanced details to get the final result. Result is then converted to RGB color space as depicted in Fig. 3 (g).

IV. RESULTS AND DISCUSSION

In this section, we present and discuss the results obtained for a number of flash/no-flash image pairs. We compare the results of the proposed method with the results made available
by [1]. We trained the dictionary on a Intel(R) Core(TM) i5-3337u CPU at 1.80GHz, with 4GB RAM on a 64 bit Windows operating system using Matlab. We used OpenCV-Python library for denoising of color components using non-local means algorithm. We assigned values for parameters $\beta$ and $N$ to 0.0003 and 50 respectively obtained empirically by applying over various sample images. We fixed the number of iterations of K-SVD to 10, and trained on 512 blocks of the image each of size $8 \times 8$, chosen from a maximum of 260000 blocks. The time required for the execution of the algorithm depends on the amount of noise in the image. The running time for all the results along with image parameters are listed in TABLE I (columns 4, 3).

![Fig. 3: Intermediate Results: (a) No-flash Image, (b) Flash image, (c) Denoised L-channel of No-flash Image, (d) Denoised a & b channel of No-flash Image, (e) L-channel of Flash Image,(f) Detail layer, (g) Final Result.](image)

**TABLE I**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cave</td>
<td>706 $\times$ 774</td>
<td>15.04</td>
<td>77.90</td>
<td>0.77</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td>Pot</td>
<td>563 $\times$ 789</td>
<td>10.97</td>
<td>54.60</td>
<td>0.72</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>Puppet</td>
<td>574 $\times$ 782</td>
<td>10.00</td>
<td>52.27</td>
<td>0.67</td>
<td>0.70</td>
<td></td>
</tr>
<tr>
<td>Lamp</td>
<td>755 $\times$ 774</td>
<td>25.47</td>
<td>69.65</td>
<td>0.66</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>Carpet</td>
<td>636 $\times$ 780</td>
<td>14.14</td>
<td>59.35</td>
<td>0.78</td>
<td>0.80</td>
<td></td>
</tr>
</tbody>
</table>

Our method with these assigned values performs well for images with low noise content. For images with higher noise content, results obtained with same values of $\beta$ and $N$ are not as good as the ones with lower noise. These parameters can be adjusted to obtain better results, but at the cost of computation time because training of dictionary and denoising using K-SVD takes a longer time in such cases. Fig. 4 shows the flash, no-flash images and compares the existing results of [1] with results obtained using the proposed framework. It is evident from these results that, sharpness and brightness are more in results obtained by our method. In the cave image set in the first row, one could observe that the proposed method is able to depict more information in regions between the jars. In the puppet set in the fourth row, the proposed method produces the final image with higher contrast. Similar observation is made in the third row for the pot set.

We use a state-of-the-art no-reference objective image quality metric proposed in [14] in order to compare the results obtained by the two methods (TABLE I (columns 5,6)). As evident from the last 2 columns of TABLE I, the quality of the images generated using the proposed approach are better than that of [1].

**V. CONCLUSION**

We have proposed a novel approach to enhance the quality of photographs captured under dark environments by fusing flash/no-flash image pairs. The framework combines the best practices of modern image processing techniques in order to generate a high quality image which represents the poorly illuminated natural scene. It is evident from the results that the proposed approach leads to more realistic images of the scenes than the state-of-the-art method [1]. We validate the effectiveness of the approach using the scores obtained using a no-reference quality metric [14]. In the proposed framework, we have not considered shadows due to flash, which occur at depth discontinuities. We believe that, it should be possible to extend the technique to account for motion in the scene.

**REFERENCES**


Fig. 4: Datasets used in First Row: Cave, Second Row: Lamp, Third Row: Pot, Fourth Row: Puppet, Fifth Row: Carpet. First Column - No-flash images, Second Column - Flash images, Third Column - Petschnigg et al. [1], Fourth Column - Our Results.