Multi-exposure Image Fusion using Patch-based Component Decomposition

[1]Dharmika A, [2]M Gnanapriya
[1]PG Scholar, Dept. of ECE, Gokula Krishna College of Engineering, Sullurpet
[2]Professor & Vice Principal, Dept. of ECE, Gokula Krishna College of Engineering, Sullurpeta

Abstract—Multi exposure image fusion is always a challenge in task in image processing. The multiple images with the different image content, when mixed using a fusion formula generate different effects. One of the most prominent effects is ghosting effect. Ghost in effect occur even in capturing of images. The smallest ghosting effect may be treated as image blur. To handle ghosting effect as well as many other affects that are generated in the process of fusion are treated in the proposed technique. The proposal scheme introduces a completely new representation that may be explorer for the for many different applications. First the input images are decomposed into several patches. As the fusion involves multiple input images the special correlated patches are further grouped into a class. Individual patches of the class are decomposed into three logical components named strength structure and intensity. These components are calculated for all the patches of the class. Now using the rule of fusion these logical components are derived for the whole class. The decomposition of a patch into logical components is unique as well as invertible hence using the generated components patches restored. Simulation results prove the superiority of the scheme proposed.

Index Terms— multi-exposure image fusion, patch, image decomposition, deghosting, tone mapping

1 Introduction

Image luminance values and captured for regular camera may acquire only a finite range of a practical scene. But in general, there exist infinite number of intensity levels in a natural scene. In addition to the quantization error, there is a fundamental limitation of capturing only finite appearances. Because of this, the captured image may contain only limited quality. This limited quality is in proportion to the human visual system [1]. Hence the captured image is sufficient for human intervention as well as many practical applications. But, when autonomous systems are using the images fed to them, to take decisions, the images need to be of high quality. This is mainly used in industries. In addition to industrial applications, scientific applications are also existed where the quality of image to be the maximum [2].

Towards this end, high dynamic range images came into existence. High dynamic range images capture all possible appearances that are available in a physical scene. The format of high dynamic range images is different from regular low dynamic range images. In addition to the format, the projection on to displays is also a complex task [3]. Hence specially manufactured displays are used to view high dynamic range images. These displays are complex and expensive. Hence this high dynamic range images are generally converted into low dynamic range images. This process is termed tone mapping [4]. The main feature of tone mapping is the quality is not compromised but the format and structural image file is modified so that it can be projected on regular display which is available with every personal computer and mobile phones [5].

The high dynamic range image will have multiple exposure content as well as multiple luminous characteristics. This gives a clue that using fusion which is a process of mixing multiple low-quality images to form a high-quality image, a number of low dynamic range images may generate another low dynamic range image but with rich content [6].

Image fusion itself has many applications like aerospace, medical and industrial. But utilizing fusion in tone mapping is a completely new approach where the tone mapping has applications in many other areas [7]. Fusion is generally done in spatial or transform domain. When fusion is implemented in spatial domain, a pixel in the output image corresponds to the pixel value in all the input images in the corresponding Geo-spatial location. Hence the spatial fusion techniques are pixel wise methods. Transform domain fusion techniques transform the input images into specific domain like a wavelet domain, Fourier domain etc. In the transform domain, the fusion rule will be applied on spatial image and transform domain will be generated. Output image is generated by inverting the content in the transformer domain. Hence, in transform domain also the fusion output is generated using pixel wise operations. In the proposed method, a new transform domain technique is proposed where there is no pixel-to-pixel relation. The method may be treated as a new kind of representation i.e., domain which is generated from it patches of input image.

2 Related Work

Most of the existing multi-exposure fusion methods devise the pixels of fused image using the corresponding pixels in the input images with a specified weight.

$$\hat{\mathbf{X}}(i) = \sum_{k=1}^{K} \mathbf{W}_k(i)\mathbf{X}_k(i)$$

Here $\mathbf{X_k}$ is a kth input image, K is the total number of input images and $\hat{\mathbf{X}}$ is fused output image. From the relation, it is very clear that each and every pixel of the output image is carried from the corresponding pixel of input images [8]. The weight function defines the strength of a pixel of a specific input image to appear in the output image. If the weight of pixel of a specific input image is less, then that particular input image will have less effect on the output fused image.

Similarly, if the weight of a pixel of a specific input image is high, then there will be a significant effect of that specific input image on the output image.
The transform domain fusion techniques are not different from the above method. In spatial domain, the pixel values are weighted by different quantity and fused image is generated, in transform domain techniques, transform domain coefficients are weighted by different quantity and fused images are generated [9].

In 1980's, a different scheme called Laplacian pyramid is proposed where the input image is decomposed by a pyramid which converts the image from the regular domain to a detailed form [10]. In this technique, once all the images are converted into a detailed form this detailed information will be merged using some fusion rule. Based on the rule of fusion a number of variations are proposed. Maximum rule, minimum rule, minmax rule, average rule etc., are the rules devised to implement fusion [11].

3 Proposed Technique

Image fusion schemes mainly suffer from ghosting effect. The ghosting effect is attributed to pixel-to-pixel operations involved in the image fusion technique. Ghosting occurs when either the objects in the image or capturing device like camera is in motion. The ghosting effect will create disturbance to the object itself as well as to the information that is contained in the background of the object. If the amount of ghosting is high, then in many cases, the object as well as information available in the background of the object are irrecoverable.

The proposed method does not involve any pixel-to-pixel operations, instead this method treats the image information in a different manner. First, the image is divided into patches. Let's consider an image with the dimension 512x512. If we intend to divide this image into patches of 16x16 then we will get 1024 patches. Patch separation will be carried on each input image. Once separation is completed, the spatial related patches maybe group into a class. If we have three images as input, of dimension 512x512, and if we have generated 1024 patches, then there will be 1024 classes; with each class contain three patches. Then each patch of each class will be logically divided into three components. These components correspond to signal strength, structure, and intensity. The extraction of these logical components for each patch is shown below mathematically.

$$\begin{aligned} x_k &= \left\| x_k - \mu_{xk} \right\| . \frac{x_k - \mu_{xk}}{\left\| x_k - \mu_{xk} \right\|} + \mu_{xk} \\ &= \left\| \tilde{x}_k \right\| . \frac{\tilde{x}_k}{\left\| \tilde{x}_k \right\|} + \mu_{xk} \\ &= c_k . s_k + l_k \end{aligned}$$

Here, μ_{xk} is the mean value, and $\tilde{x}_k = x_k - \mu_{xk}$, $\tilde{x}_k$ denotes a mean-subtracted patch, c_k is scalar, $c_k = \left\| \tilde{x}_k \right\|$, s_k is an unit vector, $s_k = \frac{\tilde{x}_k}{\left\| \tilde{x}_k \right\|}$, and l_k is a scalar, $l_k = \mu_{xk}$.

Here '*c*' represents signal strength, '*s*' represents structure and '*l*' represents intensity. These logical components will be calculated for every patch in the image. Then for every specific class of patches, fusion rule will be applied so as to generate an individual strength, structure and intensity components for each class. Hence, a class of three patches will now have and unified strength, structure, and intensity. When all these logical components of each class are inverted to generate the patches, the patches of fused image will arrive. The fusion rules are given below.

$$\hat{c} = \max_{1 \le k \le K} c_k = \max_{1 \le k \le K} \left\| \tilde{\mathbf{x}}_k \right\|$$

$$\hat{\mathbf{s}} = \frac{\bar{s}}{\left\| \bar{s} \right\|} \text{ and } \bar{s} = \frac{\sum_{k=1}^{K} S(\tilde{\mathbf{x}}_k) \mathbf{s}_k}{\sum_{k=1}^{K} S(\tilde{\mathbf{x}}_k)}$$

$$\hat{l} = \frac{\sum_{k=1}^{K} L(\mu_k, l_k) l_k}{\sum_{k=1}^{K} L(\mu_k, l_k)}$$

Here, S(.) and L(.) functions define percentage of patch in the fused image.

4 SIMULATION RESULTS

In this section, the simulation results of proposed scheme are presented. As mentioned earlier, different number of images are fed to the fusion system. The number of input images is immaterial as these images are decomposed as patches and only three components are generated out of all these patches. Simulations are carried out on a large number of images. Few results are shown in this chapter. Two input images are considered in three cases, three input images are considered in one case and five input images are considered in another case. The output image depends on the information available in the input images.

Though, the number of images is less, if the images contain needed information, then the quality of output image will be high. Even if there are large number of images with less information about the scene, then the output will be of less quality. The results on the test image sequences are given in Fig. 1 to 5.

Fig. 1 Simulation results on Image Sequence – 1: The Top three images are input images, the last image is fused output image

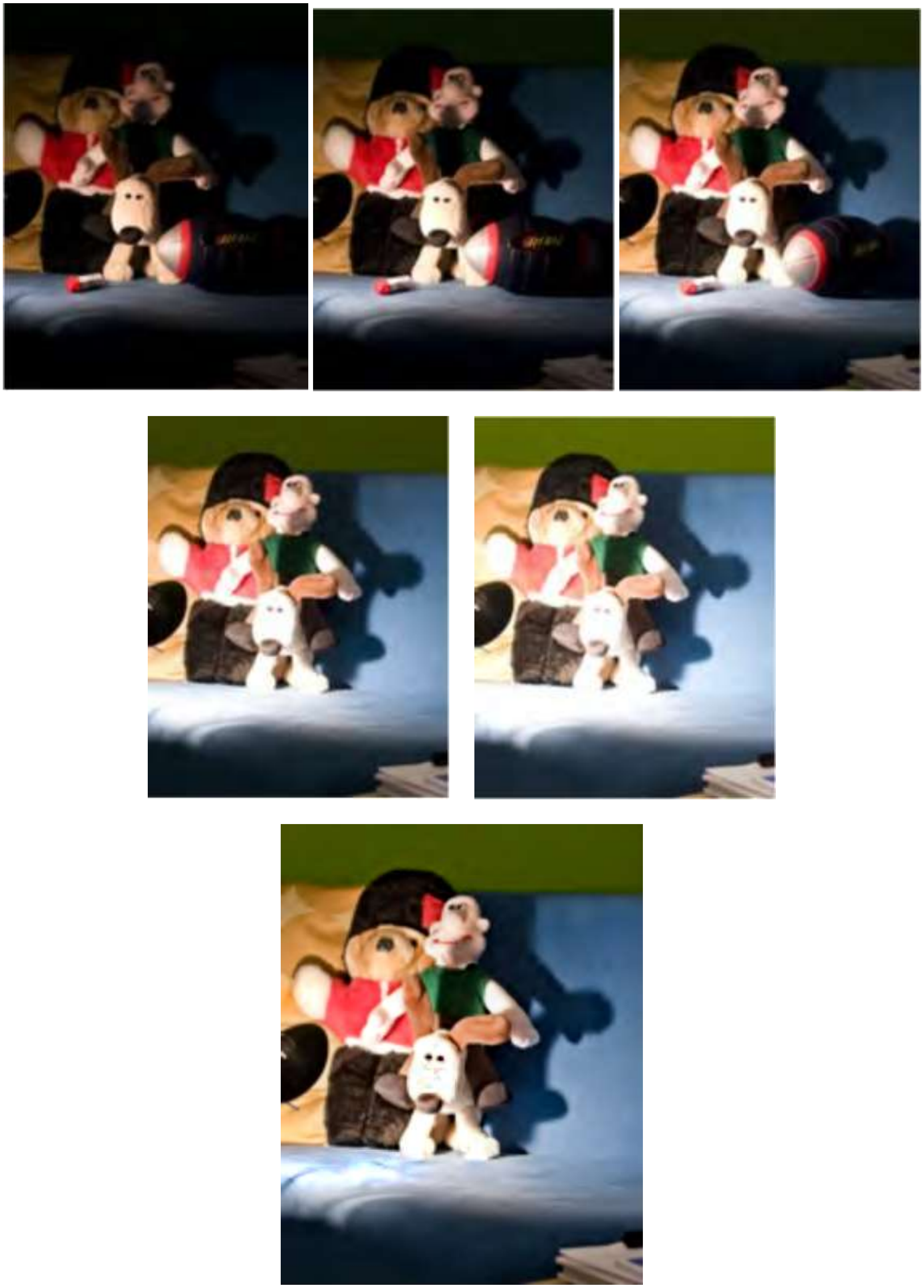

Fig. 2 Simulation results on Image Sequence – 2: The first five images are input images different exposure, the last image is the fused output image

Fig. 3 Simulation results on Image Sequence – 3: The first two images are input images, the last one is the fused output image.

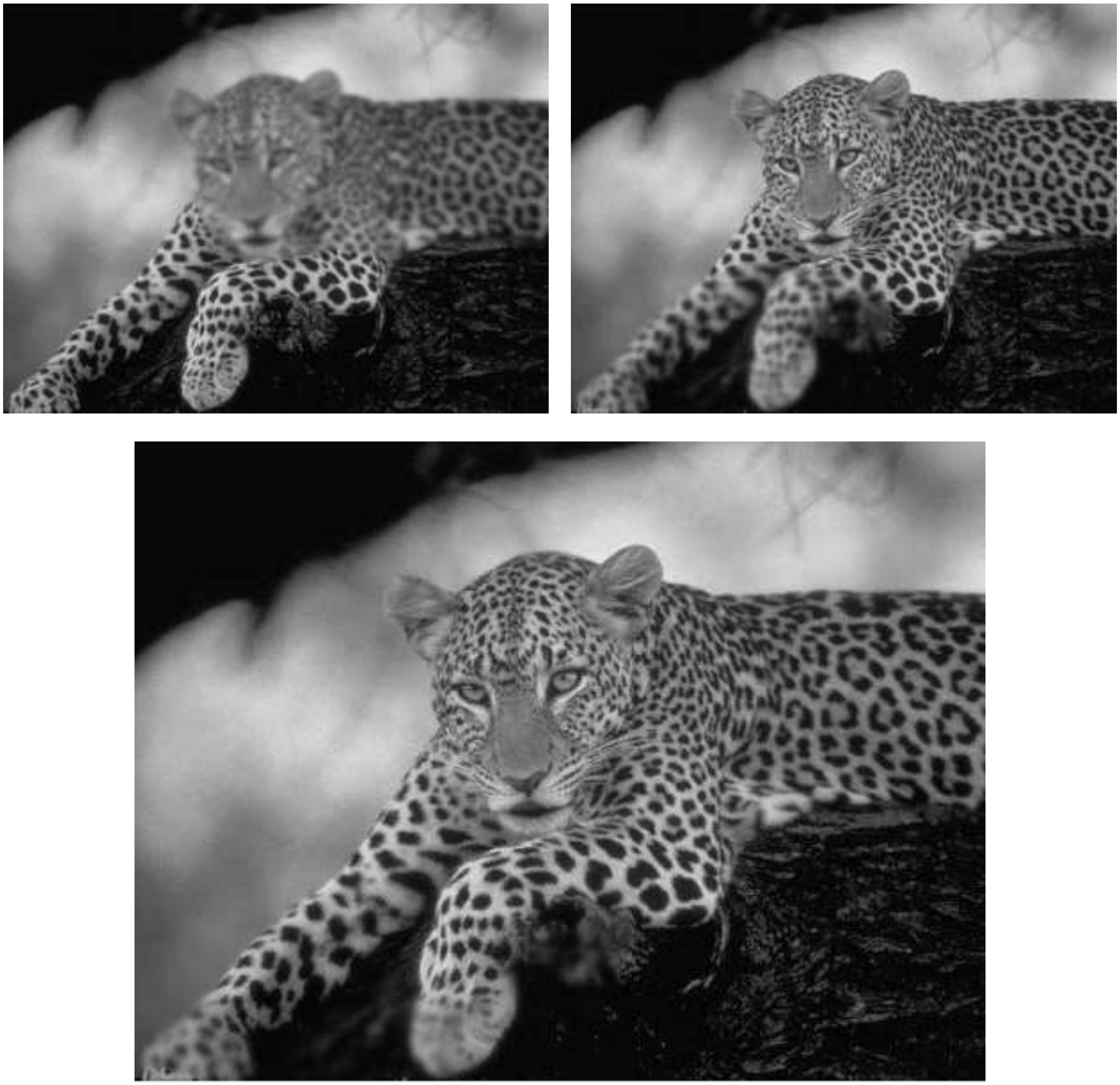

Fig. 4 Simulation results on Image Sequence - 4: The first two images are input images, the last one is the fused output image.

Fig. 5 Simulation results on Image Sequence – 5: The first two images are input images, the last one is the fused output image.

5 Conclusions

In this paper, a novel image fusion scheme is proposed using a domain which perform fusion without any pixel-to-pixel operation. The proposed method decomposes image patch into three logical components, which form the basis for a new kind of representation such as sparse representation. Sparse representation decompose image into patches, and patches will be grouped based on the features it has, as clusters. Clusters of high-quality test image set will be used to regularize the image at hand. In contrast, the proposed scheme suggests to extract three logical components from the patches in a unique manner. The attractive thing about this decomposition is that this decomposition is reversable. The decomposition involves, signal strength, structure and intensity. The fusion rule is applied on these components. Then, the resultant patch is constructed based these components. The simulation results prove that the performance is better than state-of-the-art methods.

REFERENCES

[1] K. Ma and Z. Wang, "Multi-exposure image fusion: A patch-wise approach," in *Proc. IEEE Int. Conf. Imag. Process.*, Sep. 2015, pp. 1717–1721.

[2] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski, *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. San Mateo, CA, USA: Morgan Kaufmann, 2010.

[3] 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.

[4] J.-Y. Lee, Y. Matsushita, B. Shi, I. S. Kweon, and K. Ikeuchi, "Radiometric calibration by rank minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 144–156, Jan. 2013.

[5] K. Ma, H. Yeganeh, K. Zeng, and Z. Wang, "High dynamic range image compression by optimizing tone mapped image quality index," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3086–3097, Oct. 2015.

[6] P. J. Burt, "The pyramid as a structure for efficient computation," in *Multiresolution Image Processing and Analysis*. Springer, 1984.

[7] K. Ma, K. Zeng, and Z. Wang, "Objective quality assessment for colorto-gray image conversion," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3345–3356, Dec. 2015.

[8] B. Gu, W. Li, J. Wong, M. Zhu, and M. Wang, "Gradient field multi-exposure images fusion for high dynamic range image visualization," *J. Vis. Commun. Imag. Represent.*, vol. 23, no. 4, pp. 604–610, May 2012.

[9] 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, Mar. 2009.

[10] Z. G. Li, J. H. Zheng, and S. Rahardja, "Detail-enhanced exposure fusion," *IEEE Trans. Image Process.*, vol. 21, no. 11, pp. 4672–4676, Nov. 2012.

[11] S. Raman and S. Chaudhuri, "Bilateral filter based compositing for variable exposure photography," in *Proc. Eurographics*, 2009, pp. 1–4.