MULTI-EXPOSURE IMAGE FUSION: A PATCH-WISE APPROACH

Kede Ma and Zhou Wang

Dept. of Electrical and Computer Engineering, The University of Waterloo
Email: {k29ma, zhou.wang}@uwaterloo.ca

ABSTRACT

We propose a patch-wise approach for multi-exposure image fusion (MEF). A key step in our approach is to decompose each color image patch into three conceptually independent components: signal strength, signal structure and mean intensity. Upon processing the three components separately based on patch strength and exposedness measures, we uniquely reconstruct a color image patch and place it back into the fused image. Unlike most pixel-wise MEF methods in the literature, the proposed algorithm does not require significant pre/post-processing steps to improve visual quality or to reduce spatial artifacts. Moreover, the novel patch decomposition allows us to handle RGB color channels jointly and thus produces fused images with more vivid color appearances. Extensive experiments demonstrate the superiority of the proposed algorithm both qualitatively and quantitatively.

Index Terms— Multi-exposure fusion, image enhancement, perceptual image processing

1. INTRODUCTION

Natural scenes often contain luminance levels that span a very high dynamic range (HDR), whose visual information may not be fully captured by a normal camera with a fixed exposure setting [1]. Multi-exposure image fusion (MEF) alleviates the problem by taking multiple images of the same scene under different exposure levels and synthesizing a low dynamic range (LDR) image from them. The resulting fused image is expected to be more informative and perceptually appealing than any of the input images. An example is given in Fig. 1. Compared with the typical HDR imaging pipeline, MEF bypasses the intermediate HDR construction step and directly yields an LDR image for normal displays.

Since first introduced in 1984 [3], MEF has attracted considerable interests from both academia and industry. Most existing MEF algorithms are pixel-wise methods that typically take the form of

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

where K is the number of input images in the multi-exposure source sequence, $\mathbf{W}_k(i)$ and $\mathbf{X}_k(i)$ indicate the weight and intensity values at the i-th pixel in the k-th exposure image, respectively; $\mathbf{Y}$ represents the fused image. A straightforward extension of this approach in transform domain is to replace $\mathbf{X}_k(i)$ with transform coefficients. The weight map $\mathbf{W}_k$ often bears information regarding structure preservation and visual importance of the k-th input image at a pixel level. With specific models to quantify this information, existing MEF algorithms differ mainly in the computation of $\mathbf{W}_k$. In 1994, Burt and Kolczynski applied Laplacian pyramid decomposition [3] to MEF, where $\mathbf{W}_k$ is computed from local coefficient energy and the correlation between pyramids [4]. Mertens *et al.* [5] defined contrast, color saturation and well exposure measures to compute $\mathbf{W}_k$. The fusion is done in a multiresolution fashion. Edge preserving filters such as bilateral filter [6], guided filter [7] and recursive filter [8] are applied to retrieve edge information and/or refine $\mathbf{W}_k$ in [9], [10] and [11] respectively. Song *et al.* [2] incorporated MEF into a MAP framework by first estimating the initial image with the maximum visual contrast and scene gradient, and then suppressing reversals in image gradients. Another MAP based approach embedded perceived local contrast and color saturation [12]. Gu *et al.* [13] extracted pixel-level gradient information from the structure tensor and smoothed it to compute $\mathbf{W}_k$. A similar gradient-based MEF method is proposed in [14]. By exploiting the gradient direction, the method is able to handle dynamic scenes that have moving objects. A

(a) Source image sequence by courtesy of Erik Reinhard

(b) Song12 [2]

(c) Proposed

Fig. 1. Demonstration of MEF.

detail-enhanced MEF is proposed in [15] on the basis of [5]. A relevant work in [16] divided input images into several non-overlapping patches and selected the ones with the highest entropy as the winners. The blocking artifact is reduced by adopting a pixel-wise blending function. A main drawback of most pixel-wise MEF algorithms is that the weight map $\mathbf{W}_k$ is often very noisy and may create a variety of artifacts if directly applied to the fusion process. Thus, most of the algorithms resort to certain ad-hoc remediation efforts either by pre-processing $\mathbf{X}_k$ (such as histogram equalization [11]) or post-processing $\mathbf{W}_k$ (such as smoothing [5, 13, 15] and edge preserving filtering [9–11]).

Different from the widely used pixel-wise approach of MEF in the literature, we work with image patches. Specifically, we first decompose a color image patch into three conceptually independent components: signal strength, signal structure and mean intensity, and determine each component respectively based on patch strength and exposedness measures. Such a novel patch decomposition enables us to handle RGB channels jointly so as to better make use of color information. As a result, the fused image has a more vivid color appearance. Another advantage of patch-wise approaches is their resistance to noise. As a result, unlike many existing approaches, the proposed method does not need significant ad-hoc pre/post-processing steps to improve the perceived quality or to suppress the spatial artifacts of fused images. Experiments demonstrate that the proposed algorithm creates compelling fused images both qualitatively and quantitatively.

2. PATCH-WISE MULTI-EXPOSURE FUSION

Let $\{\mathbf{x}_k\} = \{\mathbf{x}_k | 1 \le k \le K\}$ be a set of color image patches extracted from the same spatial location of the source sequence that contains K multi-exposure images. Here $\mathbf{x}_k$ for all k are column vectors of CN^2 dimensions, where C is the number of color channels in the input images and N is the spatial size of a patch. Each entry of the vector is given by one of the three intensity values in RGB channels of a pixel in the patch. Given any color patch, we first decompose it into three components: signal strength, signal structure and mean intensity

$$\begin{aligned} \mathbf{x}_k &= \|\mathbf{x}_k - \mu_{\mathbf{x}_k}\| \cdot \frac{\mathbf{x}_k - \mu_{\mathbf{x}_k}}{\|\mathbf{x}_k - \mu_{\mathbf{x}_k}\|} + \mu_{\mathbf{x}_k} \\ &= \|\tilde{\mathbf{x}}_k\| \cdot \frac{\tilde{\mathbf{x}}_k}{\|\tilde{\mathbf{x}}_k\|} + \mu_{\mathbf{x}_k} \\ &= c_k \cdot \mathbf{s}_k + l_k \,, \end{aligned} \tag{2}$$

where $\|\cdot\|$ denotes the l^2 norm of a vector, $\mu_{\mathbf{x}_k}$ is the mean value of the patch, and $\tilde{\mathbf{x}}_k = \mathbf{x}_k - \mu_{\mathbf{x}_k}$ denotes a mean-removed patch. The scalar $c_k = \|\tilde{\mathbf{x}}_k\|$, the unit-length vector $\mathbf{s}_k = \tilde{\mathbf{x}}_k / \|\tilde{\mathbf{x}}_k\|$ and the scalar $l_k = \mu_{\mathbf{x}_k}$ represent the signal strength, signal structure and mean intensity components of $\mathbf{x}_k$, respectively. Any patch can be uniquely decomposed by the three components and the processing is invertible. As such, the problem of constructing a patch in the fused image is converted to determining the three components separately and then inverting the decomposition.

We first determine the signal strength component. The visibility of the local patch structure largely depends on local contrast, which is directly related to signal strength. On one hand, the higher the contrast, the better the visibility. On the other hand, too large contrast may lead to unrealistic appearance of the local structure. Considering all input source image patches as realistic capturing of the scene, the patch that has the highest contrast among them would correspond to the best visibility under the realisticity constraint. Therefore, the desired signal strength of the fused image patch is determined by the highest signal strength of all source image patches:

$$\hat{c} = \max_{\{1 \le k \le K\}} c_k = \max_{\{1 \le k \le K\}} \|\tilde{\mathbf{x}}_k\| \,. \tag{3}$$

Different from signal strength, the structures of local image patches are denoted by unit-length vectors $\mathbf{s}_k$ for $1 \le k \le K$, each of which points to a specific direction in the vector space. The desired structure of the fused image patch corresponds to another direction in the same vector space that best represents the structures of all source image patches. A simple implementation of this relationship is given by

$$\bar{\mathbf{s}} = \frac{\sum_{k=1}^{K} S(\tilde{\mathbf{x}}_k)\, \mathbf{s}_k}{\sum_{k=1}^{K} S(\tilde{\mathbf{x}}_k)} \quad \text{and} \quad \hat{\mathbf{s}} = \frac{\bar{\mathbf{s}}}{\|\bar{\mathbf{s}}\|} \,, \tag{4}$$

where $S(\cdot)$ is a weighting function that determines the contribution of each source image patch in the structure of the fused image patch. Intuitively, the contribution should increase with the strength of the image patch. A straightforward approach that conforms with such intuition is to employ a power weighting function given by

$$S(\tilde{\mathbf{x}}_k) = \|\tilde{\mathbf{x}}_k\|^p \,, \tag{5}$$

where $p \ge 0$ is an exponent parameter.

Due to the construction of $\mathbf{x}_k$, Eq. (3) and Eq. (4) inherently take into account color contrast and structure. As an example, for uniform patches, the ones that contain strong color information are preferred to grayish ones, which usually results from under/over-exposure. By contrast, existing MEF algorithms that treat RGB channels separately may not make proper use of color information in a patch and often produce unwanted luminance changes.

With regard to the mean intensity of the local patch, we take a similar form of Eq. (4)

$$\hat{l} = \frac{\sum_{k=1}^{K} L(\mu_k, l_k)\, l_k}{\sum_{k=1}^{K} L(\mu_k, l_k)} \,, \tag{6}$$

where $L(\cdot)$ is also a weighting function that takes the global mean value μ_k of the color image $\mathbf{X}_k$ and the local mean

value of the current patch $\mathbf{x}_k$ as inputs. $L(\cdot)$ quantifies the well exposedness of $\mathbf{x}_k$ in $\mathbf{X}_k$ so that large penalty is given when $\mathbf{X}_k$ and/or $\mathbf{x}_k$ are under/over-exposed. We adopt a two dimensional Gaussian profile to specify this measure

$$L\left(\mu_k, l_k\right) = \exp\left(-\frac{(\mu_k - 0.5)^2}{2\sigma_g^2} - \frac{(l_k - 0.5)^2}{2\sigma_l^2}\right), \quad (7)$$

where σ_g and σ_l control the spreads of the profile along μ_k and l_k dimensions, respectively.[1]

Once $\hat{c}$, $\hat{\mathbf{s}}$ and $\hat{l}$ are computed, they uniquely define a new vector

$$\hat{\mathbf{x}} = \hat{c} \cdot \hat{\mathbf{s}} + \hat{l}\,. \quad (8)$$

We extract patches from the source sequence using a moving window with a fixed stride D. The pixels in overlapping patches are averaged to produce the final output.

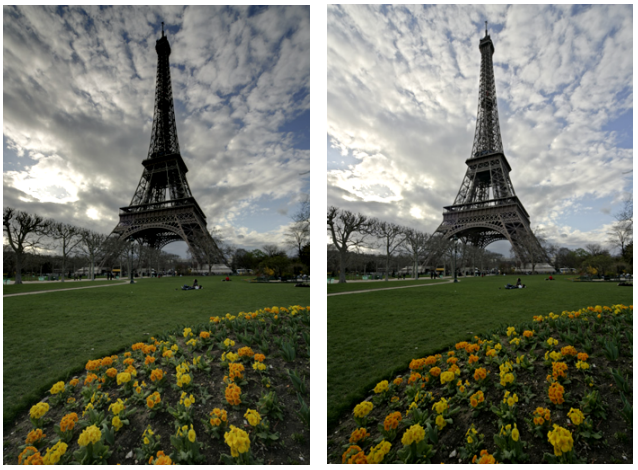
Throughout the paper, we set the patch size $N = 11$, the stride of moving window $D = \lfloor \frac{N}{5} \rfloor$, the exponent parameter $p = 4$, two spreads of Gaussian profile $\sigma_g = 0.2$ and $\sigma_l = 0.5$. Empirically, we find that the proposed algorithm is robust to variations of N and p, and a smaller value of σ_g relative to σ_l is important to produce more perceptually appealing results. The proposed method can be applied to grayscale images simply by setting $C = 1$.

3. EXPERIMENTAL RESULTS

We test the proposed method on a variety of static natural scenes with different numbers of exposure levels against eight existing MEF algorithms. For fair comparison, the same set of parameter values is used to produce all fused images as described previously. Due to space limit, only partial results are shown here. Nevertheless, the proposed algorithm is demonstrated to produce perceptually appealing results for all test sequences both qualitatively and quantitatively.

Fig. 1 shows the fused images produced by Song12 [2] and the proposed method on the "Balloons" sequence. We observe that the proposed method produces a more natural and vivid color appearance on the sky and the meadow regions. Moreover, it does a better job on structure preservation around the sun area. On the contrary, the fused image produced by Song12 [2] suffers from color distortions and detail loss. Besides, this pixel-wise method does not explicitly refine its weight map, and thus a noisy fused image may be produced on other sequences which are not shown here.

In Fig. 2, we compare Mertens09 [5] with the proposed method on the "Tower" sequence. The former algorithm performs the best on average in a recent subjective user study among eight MEF algorithms [17]. Compared with Mertens09 [5], we can clearly observe several perceptual gains on the fused image produced by the proposed method. For example, the structures of the tower at the top and the brightest cloud area are much better preserved. Also, the color appearance of the sky and the meadow regions is more natural and consistent with the source sequence.



(a) Mertens09 [5] (b) Proposed

Fig. 2. Comparison of the proposed method with Mertens09 [5]. Source sequence by courtesy of Jacques Joffre.

(a) Shutao12 [11] (b) Proposed

Fig. 3. Comparison of the proposed method with Shutao12 [11]. Source sequence by courtesy of Tom Mertens.

Fig. 3 compares Shutao12 [11] with the proposed method on the "House" sequence. Shutao12 [11] treats RGB channels separately, making it difficult to properly make use of color information. As a result, the color in the uniform areas such as the walls and window frames, appears dreary. The global luminance of the fused image also changes drastically, where the left part of the image is clearly brighter than the right part. By contrast, the proposed method better preserves the color information and the overall appearance of the fused image is more appealing.

The comparison results between Li12 [15] and the proposed method on the "Belgium House" sequence is exemplified in Fig. 4. Li12 [15] is a detail-enhanced version of Mertens09 [5]. Detail enhancement does not necessarily result in perceptual gains especially when it neglects the realisticity constraint of camera acquisition. As a result, the fused image produced by Li12 [15] looks unnatural around edges, for example near the branches and window frames. The pro-

[1] Input multi-exposure images are normalized to [0,1].

Table 1. Performance comparison of the proposed method with existing MEF algorithms using the objective IQA model in [18]. The quality value ranges from 0 to 1 with a higher value indicating better perceptual quality. globalEng stands for a naïve method that linearly combines the input images using global energy as weighting factors.

Source sequence	[13]	[15]	[10]	[9]	[11]	[2]	globalEng	[5]	Proposed
Balloons	0.913	0.941	0.948	0.768	0.944	0.883	0.862	**0.969**	**0.963**
Cave	0.934	0.923	**0.978**	0.694	0.961	0.822	0.837	0.974	**0.980**
Chinese garden	0.927	0.951	0.984	0.911	0.982	0.878	0.928	**0.989**	**0.988**
Farmhouse	0.932	0.958	**0.985**	0.877	0.977	0.756	0.916	0.981	**0.983**
Lamp	0.871	0.933	0.934	0.864	0.937	0.817	0.887	**0.948**	**0.945**
Landscape	0.941	0.948	0.942	0.954	0.972	0.937	0.962	**0.976**	**0.991**
Madison Capitol	0.864	0.949	0.968	0.763	0.918	0.702	0.886	**0.977**	**0.974**
Office	0.900	0.954	0.967	0.907	0.972	0.919	0.955	**0.984**	**0.986**
Tower	0.931	0.950	**0.986**	0.895	0.984	0.178	0.912	**0.986**	0.981
Venice	0.889	0.937	0.954	0.892	0.952	0.845	0.913	**0.966**	**0.978**
Average	0.910	0.944	0.965	0.852	0.960	0.774	0.906	**0.975**	**0.977**

(a) Li12 [15]

(b) Proposed

Fig. 4. Comparison of the proposed method with Li12 [15]. Source sequence by courtesy of Dani Lischinski.

posed method produces the fused image with a more realistic appearance and little detail loss.

In order to evaluate the performance of MEF algorithms objectively, we adopt a recently proposed image quality assessment (IQA) model that well correlates with subjective judgements [18]. Although a number of IQA models for general image fusion have also been proposed [19–27], none of them makes adequate quality predictions of subjective opinions as reported in [17]. The details of these models can be found in an excellent review paper [28]. The model in [18] is based on the multi-scale structural similarity (SSIM) framework [29,30]. It keeps a good balance between local structure preservation and global luminance consistency. The quality value of the IQA model ranges from 0 to 1 with a higher value indicating better quality. The comparison results of the proposed method with eight existing MEF algorithms on ten source sequences are listed in Table 1, from which we observe that the proposed method produces comparable results with Mertens09 [5] in terms of the IQA model in [18], whose quality values are considerably higher than those of other MEF algorithms. Note that the model in [18] works with luminance component only and may underestimate the quality gain of the proposed method, for which producing a natural and vivid color appearance is one of the main advantages.

The computational complexity of the proposed method increases linearly with the number of pixels in the source sequence. Our unoptimized MATLAB code takes around 2.9 seconds to process a source sequence of size $341 \times 512 \times 3$.

4. CONCLUSION AND FUTURE WORK

MEF is a handy and practical image enhancement framework that is widely adopted in consumer electronics. Most existing MEF algorithms are pixel-wise methods, which often suffer from noisy weight maps. As a result, ad-hoc pre/post-processing steps are often involved in order to produce reasonable results. By contrast, the proposed method works with color image patches directly by decomposing them into three conceptually independent components and determining each component respectively based on patch strength and exposedness measures. Experiments demonstrate that the proposed method produces compelling fused images both visually and in terms of a recently proposed objective quality model [18].

The novel patch decomposition underlying the proposed method renders it highly flexible to include new features. First, by incorporating the direction of structure vector $\mathbf{s}_k$ into the construction of the weighting function $S(\cdot)$, we may be able to account for dynamic scenes. Second, by replacing $\hat{l}$ in the current computation with local patch mean values of some already fused images, the algorithm is transformed to a detail enhancement algorithm. The problem now is to find the best candidate fused image that combines with the proposed method to produce the best quality image. These issues will be investigated in our future work.

5. REFERENCES

[1] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski, *High Dynamic Range Imaging: Acquisition, Display, and Image-based Lighting*. Morgan Kaufmann, 2010.

[2] M. Song, D. Tao, C. Chen, J. Bu, J. Luo, and C. Zhang, "Probabilistic exposure fusion," *IEEE TIP*, 2012.

[3] P. J. Burt, *The pyramid as a structure for efficient computation*. Springer, 1984.

[4] P. J. Burt and R. J. Kolczynski, "Enhanced image capture through fusion," in *ICCV*, 1993.

[5] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion: A simple and practical alternative to high dynamic range photography," *Computer Graphics Forum*, 2009.

[6] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *ICCV*, 1998.

[7] K. He, J. Sun, and X. Tang, "Guided image filtering," in *ECCV*, 2010.

[8] E. S. Gastal and M. M. Oliveira, "Domain transform for edge-aware image and video processing," *ACM TOG*, 2011.

[9] S. Raman and S. Chaudhuri, "Bilateral filter based compositing for variable exposure photography," in *Proc. Eurographics*, 2009.

[10] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE TIP*, 2013.

[11] S. Li and X. Kang, "Fast multi-exposure image fusion with median filter and recursive filter," *IEEE Trans. on Consumer Electronics*, 2012.

[12] R. Shen, I. Cheng, and A. Basu, "QOE-based multi-exposure fusion in hierarchical multivariate Gaussian CRF," *IEEE TIP*, vol. 22, no. 6, pp. 2469–2478, 2013.

[13] B. Gu, W. Li, J. Wong, M. Zhu, and M. Wang, "Gradient field multi-exposure images fusion for high dynamic range image visualization," *Journal of Visual Communication and Image Representation*, 2012.

[14] W. Zhang and W.-K. Cham, "Gradient-directed multiexposure composition," *IEEE TIP*, 2012.

[15] Z. Li, J. Zheng, and S. Rahardja, "Detail-enhanced exposure fusion," *IEEE TIP*, 2012.

[16] A. A. Goshtasby, "Fusion of multi-exposure images," *Image and Vision Computing*, 2005.

[17] K. Zeng, K. Ma, R. Hassen, and Z. Wang, "Perceptual evaluation of multi-exposure image fusion algorithms," in *6th International Workshop on Quality of Multimedia Experience*, 2014.

[18] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *submitted to IEEE TIP*, 2015.

[19] G. Qu, D. Zhang, and P. Yan, "Information measure for performance of image fusion," *Electronics Letters*, 2002.

[20] C. S. Xydeas and V. S. Petrovic, "Objective pixel-level image fusion performance measure," in *AeroSense*, 2000.

[21] P.-W. Wang and B. Liu, "A novel image fusion metric based on multi-scale analysis," in *IEEE ICSP*, 2008.

[22] Y. Zheng, E. A. Essock, B. C. Hansen, and A. M. Haun, "A new metric based on extended spatial frequency and its application to DWT based fusion algorithms," *Information Fusion*, 2007.

[23] G. Piella and H. Heijmans, "A new quality metric for image fusion," in *IEEE ICIP*, 2003.

[24] N. Cvejic, A. Loza, D. Bull, and N. Canagarajah, "A similarity metric for assessment of image fusion algorithms," *International Journal of Signal Processing*, 2005.

[25] C. Yang, J.-Q. Zhang, X.-R. Wang, and X. Liu, "A novel similarity based quality metric for image fusion," *Information Fusion*, 2008.

[26] H. Chen and P. K. Varshney, "A human perception inspired quality metric for image fusion based on regional information," *Information Fusion*, 2007.

[27] Y. Chen and R. S. Blum, "A new automated quality assessment algorithm for image fusion," *Image and Vision Computing*, 2009.

[28] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganiere, and W. Wu, "Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: a comparative study," *IEEE TPAMI*, 2012.

[29] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE TIP*, 2004.

[30] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thirty-Seventh IEEE Asilomar Conference on Signals, Systems and Computers*, 2003.