HYBRID METHOD FOR MULTI-EXPOSURE IMAGE FUSION BASED ON WEIGHTED MEAN AND SPARSE REPRESENTATION

Takao Sakai, Daiki Kimura, Taichi Yoshida, and Masahiro Iwahashi

Dept. Electrical Electronics Information Engineering, Nagaoka Univ. of Tech.,
Nagaoka, Niigata, 940-2137 Japan.
Email: {sakai, kimura, yoshida}@iwalab.org, iwahashi@vos.nagaokaut.ac.jp

ABSTRACT

We propose a hybrid method for multi-exposure image fusion in this paper. The fusion blends some images capturing the same scene with different exposure times and produces a high quality image. Based on the pixel-wise weighted mean, many methods have been actively proposed, but their resultant images have blurred edges and textures because of the mean procedure. To overcome the disadvantages, the proposed method separately fuses the means and details of input images. The details are fused based on sparse representation, and the results keep their sharpness. Consequently, the resultant fused images are fine with sharp edges and textures. Through simulations, we show that the proposed method outperforms previous methods objectively and perceptually.

Index Terms— Multi-exposure image fusion, high-dynamic-range imaging, weighted mean, sparse representation.

1. INTRODUCTION

When we take photos of natural scenes that include very dark and very bright regions, their digital images often lose details of these regions. In general, commonly used digital cameras have narrower ranges of luminance than natural scenes [1, 2]. We cannot obtain details of regions whose luminance is outside camera ranges. These regions are commonly called saturation regions.

To clearly represent scenes without saturation regions, multi-exposure image fusion has been proposed [3–22]. It fuses some images into one desired image. The input images are obtained by taking photos of the same scene with different exposure times, and the locations of their saturation regions are different. Hence, the fused image fully represents the scene without saturation regions.

Methods for multi-exposure image fusion are mainly classified into two types: weighted mean and gradient cascade. The former type has been actively studied and includes the greatest number of methods [3–19]. These methods fuse images by pixel-wise weighted mean. Various procedures for the weight calculation have been proposed. Recently, some methods have aimed to prevent visual artifacts, such as motion blurs and ghosts [12–19]. The artifacts are caused by movement of objects, and recent methods try to align objects in the same locations. Consequently, these methods reduce artifacts and produce natural images. In the gradient cascade type, few methods have been proposed [20, 21]. These methods choose the maximum gradients of input images at each pixel, and the resultant gradient field is defined as gradients of the fused image. Finally, the gradients are transformed to the spatial domain, and the result is the fused image. They produce fine edges and textures in fused images.

Each of the two types has problems with fused images. Due to the mean procedure, weighted mean methods produce blurred images. In particular, edges and textures of their resultant fused images are blurred. With the other type, errors caused by noise and saturation are spread all over the image via the transformation from the gradient domain to the spatial domain, and the spreading amplifies the errors. Consequently, unnatural regions occur in fused images.

Recently, sparse representation is widely used as fundamental technique in image processing [23–25], because it can approximate images to have sharp edges and textures without slight variations such as noises. Several image applications based on sparse representation have been proposed, and achieve excellent results [23–25]. In the multi-exposure image fusion, a method based on sparse representation has also been proposed [22]. The method divides mean values and residual components of each input image by patch unit, averages the values, and fuses the components based on sparse representation to produce sharp fused components. Unfortunately, since the averaging procedure is poor and the fusion is affected by saturation regions, fused images are visually blurred and artifacts often occur.

To overcome the problems of previous methods, we propose a hybrid method for multi-exposure image fusion based on weighted mean and sparse representation. The proposed method produces averages and details of fused images by using weighted mean and sparse representation, respectively. The details mean edges, local contrasts, and textures. Due to the weighted mean method, the resultant average images

This work was supported by JSPS KAKENHI Grant Number 26889031

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

are visually natural. For fusion of detail components, we use the proposed selection method (which includes sparse representation) to avoid blurs and effects of saturation regions. Due to the proposed fusion, the resultant details have sharp edges and textures. Consequently, the proposed method produces fine fused images without artifacts, and we show that the proposed method outperforms previous methods through simulations objectively and perceptually. We assume that the object alignment is already finished by previous methods in this paper, because the several alignment methods have been proposed, and show their efficacy [2, 18]. Nowadays, tasks for multi-exposure image fusion should be divided to be easy solved. Hence, to show an efficacy of the proposed method for the fusion, we take the assumption in this paper.

2. REVIEW

2.1. Fusion Method based on Weighted Mean

In this paper, we use a successful method of multi-exposure image fusion based on weighted mean [10]. This method acts in the multi-resolution domain using the Gaussian and Laplacian pyramid [26].

The method consists of four processes: weight calculation, Laplacian pyramid decomposition of input images, Gaussian pyramid decomposition of weights, and fusion. Let $\hat{\mathbf{X}}$, $\mathbf{X}_l$, $\hat{\mathbf{W}}_l$, $\mathcal{L}(\cdot)$, and $\mathcal{G}(\cdot)$ be the fused image, the l-th input image, the normalized weights of $\mathbf{X}_l$ ($l = 1, 2, \cdots, L$), and the function for Laplacian and Gaussian pyramid decomposition, respectively, where L is the number of input images. The method fuses $\mathbf{X}_l$ with $\hat{\mathbf{W}}_l$ to produce $\hat{\mathbf{X}}$, which is defined as

$$\mathcal{L}(\hat{\mathbf{X}}) = \sum_{l=1}^{L} \mathcal{G}(\hat{\mathbf{W}}_l)\mathcal{L}(\mathbf{X}_l). \tag{1}$$

The weights of each pixel are calculated based on three measures: contrast, saturation and well-exposedness. The contrast values are defined as the absolute values of the transformed images, where each transformed image is calculated by applying the Laplacian filter to the greyscale version of the input image. The saturation values are the standard deviations within the colour channels at each pixel. The well-exposedness is derived from the function $f(x) = \exp(-(x - 0.5)^2/0.08)$. This function is separately applied to each channel at each pixel and the well-exposedness is obtained by multiplying the results. Finally, weights are calculated as a weighted multiplication of the results of the three measurements and normalized.

2.2. Sparse Representation

Sparse representation is the approximation of input signals with few pre-learned atoms. A set of the atoms is commonly called a dictionary. Let $\mathbf{x} \in \mathbb{R}^M$ be the input signals. $\mathbf{x}$ is approximated by a dictionary $\mathbf{D} \in \mathbb{R}^{M \times N}$ ($M < N$) as $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}$, where $\boldsymbol{\alpha} \in \mathbb{R}^N$ means sparse coefficients that are the solution of

$$\arg\min_{\boldsymbol{\alpha}} ||\boldsymbol{\alpha}||_0 \ \text{s.t.} \ ||\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}||_2 \leq \epsilon, \tag{2}$$

$||\cdot||_0$ and $||\cdot||_2$ are the l_0 and l_2 norm, respectively. From (2), $\boldsymbol{\alpha}$ has many zero and few non-zero numbers. (2) is a strict formulation of the sparse representation. In general, to easily realize sparse representation, the l_0 norm is relaxed into the l_1 norm. Similar to the conventional method, we use the l_1 version of (2) and the orthogonal matching pursuit method [27] as its solver in this paper.

3. PROPOSED HYBRID METHOD

3.1. Framework

The proposed method is a hybrid of weighted mean and sparse representation to separately produce average and texture components of fused images from input images. Its framework is shown in Fig. 1, where 'Conv. method' means the conventional method of the weighted mean in Sec. 2.1. First, we produce an initial version of the fused image by the weighted mean method. Next, since the initial image has blurred edges and textures, we subtract it from the input images to extract variational components, where the luminance level of the initial image is adaptively adjusted to the luminance levels of the input images. The variational components are fused based on the sparse representation in Sec. 2.2 to keep their sharpness. Finally, by adding the initial image and the fused components, the proposed method produces a fine fused image.

3.2. Extraction of Variational Components

The extraction of variational components consists of three steps: lowpass filtering, luminance adjustment, and subtraction. The initial image is purposely blurred by lowpass filtering to enhance the extracted variational components. If this process is skipped, values of the subtracted components are slight like noise, and the proposed fusion in Sec. 3.3 is impaired. In this paper, we use a two-dimensional Gaussian filter with size 5×5 and $\sigma = 1.5$ as the lowpass filter, where σ is its standard deviation. The luminance adjustment shifts the mean value of the blurred initial image to fit the mean values of the input images by patch unit. Let $\hat{\mathbf{x}}_{l,p}$ ($p = 1, 2, \cdots$) be the p-th adjusted patch for the l-th input image, and $\hat{\mathbf{x}}_{l,p}$ is defined as

$$\hat{\mathbf{x}}_{l,p} = \mathbf{x}_{0,p} + (\mu_{l,p} - \mu_{0,p}) \times \mathbf{1}, \tag{3}$$

where $\mathbf{x}_{0,p}$, $\mu_{l,p}$, $\mu_{0,p}$, and $\mathbf{1}$ are the p-th patch of the blurred initial image, mean value of the p-th patch in the l-th input image, mean value of $\mathbf{x}_{0,p}$, and vector whose components are 1, respectively. Finally, let $\mathbf{x}_{l,p}$ and $\mathbf{v}_{l,p}$ be the p-th patch in the l-th input image and its variational components, and $\mathbf{v}_{l,p}$ is calculated as $\mathbf{v}_{l,p} = \mathbf{x}_{l,p} - \hat{\mathbf{x}}_{l,p}$.

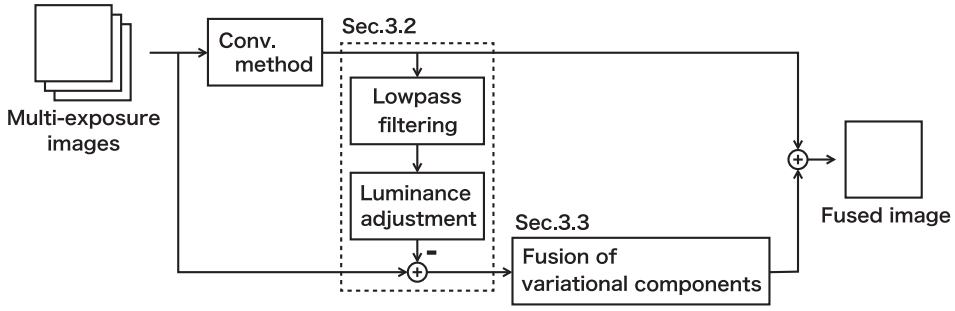

Fig. 1: Block diagram of proposed method.

3.3. Fusion of Variational Components

The proposed method fuses sparse coefficients of the variational components of input images by weighted mean, and the components of the fused image are calculated by the fused coefficients. To avoid the effect of saturation regions in input images and restrict the number of atoms used for the proposed fusion, we calculate the weight based on the variances of patches in the input images. Saturation regions generally have low variances, and we mainly utilize patches whose variances are high because these patches usually have important edges and textures. Therefore, when the patch variance is low, we determine that the weight is also low and vice versa.

Let $\hat{\mathbf{v}}_p$ and $\hat{\boldsymbol{\alpha}}_p$ be the variational components of the p-th fused patch and their sparse coefficients, respectively. Based on Sec. 2.2, $\hat{\mathbf{v}}_p = \mathbf{D}\hat{\boldsymbol{\alpha}}_p$, and we define $\hat{\boldsymbol{\alpha}}_p$ as

$$\hat{\boldsymbol{\alpha}}_p = \sum_{l=1}^{L} \hat{w}_{l,p}\boldsymbol{\alpha}_{l,p}, \tag{4}$$

where $\boldsymbol{\alpha}_{l,p}$ denotes sparse coefficients of $\mathbf{v}_{l,p}$, and $\hat{w}_{l,p}$ is the normalized weight defined as

$$\hat{w}_{l,p} = 1/S_p \times g(\sigma_{l,p}),$$
$$S_p = \sum_{l=1}^{L} g(\sigma_{l,p}),$$
$$g(x) = \begin{cases} 1 & \text{if } x \geq \tau + \epsilon \\ 1/\epsilon^2 \times (x-\tau)^2 & \text{if } \tau \leq x \leq \tau + \epsilon \\ 0 & \text{if } x \leq \tau \end{cases},$$

τ and ϵ are parameters to design $g(\cdot)$, and $\sigma_{l,p}$ is the standard deviation of $\mathbf{x}_{l,p}$. Note that we define $1/0 \times 0 = 0$ in (4). As mentioned above, the function $g(\cdot)$ controls weights in accordance with the patch variances, shown in Fig. 2.

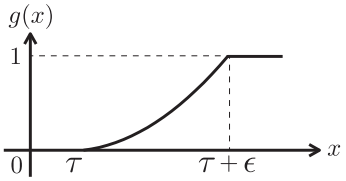

Fig. 2: Function to determine weights for fusion of variational components.

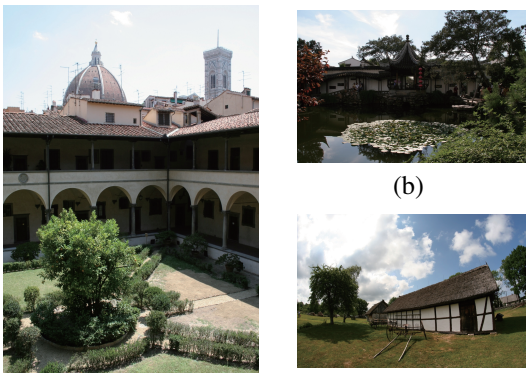

Fig. 3: Images of image sets at middle exposure time.

4. SIMULATION

The proposed method is compared with previous methods [10, 15] and the easyHDR. The weighted mean methods in Sec. 2.1 is chosen as conventional method [10], and a state-of-the-art weighted mean method with object alignment is used [15]. The easyHDR is a commercial software for multi-exposure image fusion. The source codes and the easyHDR are available on the author's website and www.easyhdr.com. In the proposed method, the pre-learned dictionary is generated from images of various exposure time and many scenes by the K-SVD [23], and the size of atoms is 5×5.

We use 10 image sets of natural scenes, and the tone-mapped image quality index (TMQI) [28] as an objective measure. The test sets include three or four images per scene. We show some images at middle exposure time in Fig. 3. The TMQI measures tone-mapped images based on a modified structural similarity method between images before and after tone-mapping and its statistical naturalness. The tone-mapped images are produced by compressing the luminance of their scenes into the desired range. The details of the im-

(a) Input multi-exposure image set

(b) [10] (c) Prop. (d) [15] (e) easyHDR

Fig. 4: Fused images.

ages are fully represented. Hence, the fused images seem to be the tone-mapped images, and the TMQI is valid as a measure for them. However, due to not obtaining images before tone-mapping, we can only use the statistical naturalness.

The proposed method averagely shows better scores in the TMQI than the others, as shown in Table 1, where 'Prop.' and 'Average' mean the proposed method and average values of 10 test sets, respectively. The values are in $[0, 1]$ and a higher one is better. The proposed method cannot always outperform the others, however it always shows high scores over 0.5, which is different from the others.

The proposed method obviously shows sharp edges and clear textures in both dark and bright regions than the others, as shown in Fig. 4. The result of Fig. 4 (b) seem to be a blurred version of the proposed ones. Fig. 4 (d) has wrong colour due to the false object alignment. The methods which reduce the ghost artifact by the object alignment such as [15] occasionally produce particular artifacts similar to Fig. 4 (d). The easyHDR unnaturally enhances colour and shows blurred edges. Therefore, we understand that the proposed method perceptually outperforms the others.

Table 1: TMQI scores.

	[10]	Prop.	[15]	easyHDR
Fig. 3 (a)	0.934	0.943	0.503	0.613
Fig. 3 (b)	0.623	0.610	0.216	0.163
Fig. 3 (c)	0.679	0.682	0568	0.738
Average	0.554	0.639	0.444	0.464

6. CONCLUSION

In this paper, we propose a hybrid method for multi-exposure image fusion based on weighted mean and sparse representation to produce average and variational components of fused images, respectively. Because of the advantages of weighted mean and sparse representation, the resultant fused images are visually natural and have sharp edges and textures. The proposed method shows better results than previous methods in the simulations objectively and perceptually. As future work, an alignment procedure considering the proposed algorithm will be introduced.

REFERENCES

[1] B. Hoefflinger, *High-Dynamic-Range (HDR) Vision*, New York: Springer-Verlag, 2007.

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

[3] S. Mann and R. W. Picard, "On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures," in *Proc. IS&T Conf.*, 1995, pp. 422–428.

[4] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. ACM SIGGRAPH*, 1997, pp. 369–378.

[5] T. Mitsunaga and S. K. Nayar, "Radiometric self calibration," in *Proc. IEEE Comput. Society Conf. Comput. Vision Patt. Recognition*, 1999, vol. 1, pp. 374–380.

[6] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," in *Proc. ACM SIGGRAPH*, 2002, pp. 257–266.

[7] A. A. Goshtasby, "Fusion of multi-exposure images," *Image Vision Comput.*, vol. 23, no. 6, pp. 611–618, 2005.

[8] X. Li, K. M. Lam, and L. Shen, "An adaptive algorithm for the display of high-dynamic range images," *J. Vis. Commun. Image Represent.*, vol. 18, no. 5, pp. 397–405, 2007.

[9] S. Zheng, W.-Z. Shi, J. Liu, G.-X. Zhu, and J.-W. Tian, "Multisource image fusion method using support value transform," *IEEE Trans. Image Process.*, vol. 16, no. 7, pp. 1831–1839, 2007.

[10] 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.

[11] R. Shen, I. Cheng, J. Shi, and A. Basu, "Generalized random walks for fusion of multi-exposure images," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3634–3646, 2011.

[12] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High dynamic range video," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 319–325, 2003.

[13] E. A. Khan, A. O. Akyiiz, and E. Reinhard, "Ghost removal in high dynamic range images," in *Proc. IEEE Int. Conf. Image Process.*, 2006, pp. 2005–2008.

[14] W.-C. Kao, "High dynamic range imaging by fusing multiple raw images and tone reproduction," *IEEE Trans. Consumer Electronics*, vol. 54, no. 1, pp. 10–15, 2008.

[15] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman, "Robust patch-based HDR reconstruction of dynamic scenes," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 203:1–203:11, 2012.

[16] T. Jinno and M. Okuda, "Multiple exposure fusion for high dynamic range image acquisition," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 358–365, 2012.

[17] M. Song, D. Tao, C. Chen, J. Bu, J. Luo, and C. Zhang, "Probabilistic exposure fusion," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 341–357, 2012.

[18] N. K. Kalantari, E. Shechtman, C. Barnes, S. Darabi, D. B. Goldman, and P. Sen, "Patch-based high dynamic range video," *ACM Trans. Graph.*, vol. 32, no. 6, 2013.

[19] 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, 2014.

[20] A. R. V.-Kóczy, A. Rövidz, and T. Hashimoto, "Gradient-based synthesized multiple exposure time color HDR image," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 8, pp. 1779–1785, 2008.

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

[22] J. Wang, H. Liu, and N. He, "Exposure fusion based on sparse representation using approximate K-SVD," *Neurocomputing*, vol. 135, pp. 145–154, 2014.

[23] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006.

[24] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. IEEE Int. Conf. Comput. Vision*, 2009, vol. 12, pp. 2272–2279.

[25] K. Zhang, X. Gao, D. Tao, and X. Li, "Single image super-resolution with multiscale similarity learning," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 24, no. 10, pp. 1648–1659, 2013.

[26] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Prentice Hall, 2002.

[27] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Proc. Asilomar Conf. Signals Syst. Comput.*, 1993, vol. 1, pp. 40–44.

[28] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 657–667, 2013.