

DCT based Multi Exposure Image Fusion

O. Martorell, C. Sbert and A. Buades

*Department of Mathematics and Computer Science & IAC3, University of Balearic Islands,
Cra. de Valldemossa km. 7.5, 07122 Palma, Illes Balears, Spain*

Keywords: Multi-exposure Images, Image Fusion, DCT Transform.

Abstract: We propose a novel algorithm for multi-exposure fusion (MEF). This algorithm decomposes image patches with the DCT transform. Coefficients from patches with different exposure are combined. The luminance and chrominance of the different images are fused separately. Details of the fused image are finally enhanced as a post-processing. Experiments with several data sets show that the proposed algorithm performs better than state-of-the-art.

1 INTRODUCTION

Multi-exposure fusion (MEF) is a technique for combining different images of the same scene acquired with different exposure settings into a single image. The natural light has a large range of intensities which a conventional camera cannot capture. By keeping the best exposed parts of each image, we can recover a single image where all features are well represented.

High Dynamic Range (HDR) imaging from a exposure sequence is usually confused with MEF. In HDR, the irradiance function of the image has to be built, and in order to do so, the camera response function has to be estimated. Most methods use the algorithm by Malik and Debevec (Debevec and Malik, 2008). Finally, to be displayed, the estimated HDR function has to be converted into a typical 8 bit image. This problem is known as tone-mapping (Reinhard et al., 2005).

All proposed MEF algorithms combine the set of images, somehow choosing for each pixel the one with better exposition. State of the art methods express this choice as a weighted average depending on common factors as exposure, saturation and contrast, e.g. Mertens et al. (Mertens et al., 2009). The pixel values or their gradient might be combined. In the case that pixel gradient is manipulated, a final estimate has to be recovered by Poisson editing (Pérez et al., 2003). Robustness of methods is achieved by using pyramidal structures or working at the patch level instead of the pixel one.

We propose a multi-exposure image fusion algorithm inspired by MEF Mertens's algorithm (Mertens

et al., 2009) and the Fourier Burst Accumulation (FBA) algorithm proposed by Delbracio and Sapiro (Delbracio and Sapiro, 2015). Instead of combining gradient or pixel values, we fuse the DCT coefficients of the differently exposed images. The algorithm decomposes the image into patches and computes the DCT. The coefficients of patches at the same location but different exposure are combined depending on its magnitude. This strategy is used to combine patches of the luminance. The chrominance values are fused separately as a weighted average at the pixel level.

The novel method for the fusion of the luminance images is presented in section 3.1. The chrominance fusion algorithm is described in section 3.2. Section 3.3 proposes a detail extraction technique to enhance the fused result. Finally, in section 4 some results and comparisons with other image fusion algorithms are presented.

2 RELATED WORK

There exists an extensive amount of research in MEF. The first distinction among MEF algorithms is whether they combine pixel color values or pixel gradients. Secondly, pixel-wise algorithms may be regularized or made robust by using multi-scale pyramids or patch based strategies. Finally, the fusion method might be applied directly to RGB channels, to a decomposition of luminance and chrominance, or a decomposition into basis and detail images. This latter case permits to additionally apply an enhancement of the detail part.

Pixel-wise methods, directly applied to RGB components, compute a weighted average

$$u(x) = \sum_{k=1}^K w_k(x) I_k(x) \quad (1)$$

where K is the number of images in the multi-exposure sequence, $I_k(x)$ is the input value at x position in the k -th exposure image and $w_k(x)$ is the k -th weight at this position. The weight map w_k measures information such as edge strength, well-exposedness, saturation, etc. Mertens's algorithm (Mertens et al., 2009) is the widely used algorithm in this class. It computes three quality metrics:

- *Contrast C*: the absolute value of the Laplacian filter applied to the grayscale of each image.
- *Saturation S*: measures the standard deviation of R , G and B channels.
- *Well-exposedness E*: measures the closeness to the mid intensity value.

These metrics are combined into a weight map for each image

$$w_k(x) = \frac{C_k(x)^{\alpha_c} S_k(x)^{\alpha_s} E_k(x)^{\alpha_e}}{\sum_{j=1}^K C_j(x)^{\alpha_c} S_j(x)^{\alpha_s} E_j(x)^{\alpha_e}}, \quad k = 1, 2, \dots, K.$$

Other methods compute fusion by a weighted average, but with a different configuration to the proposed one by Mertens et al. For example, Liu et al. (Liu and Wang, 2015) propose to use the SIFT descriptor at each pixel to compute the weights, which are later refined and merged by a multiscale procedure.

The second type of methods use the gradient of the images, and then recovers the solution by Poisson image editing (Pérez et al., 2003). For several algorithms, just the most convenient gradient is chosen, for example Kuk et al. (Kuk et al., 2011). Other methods combine all gradients, for example (Raskar et al., 2005; Zhang and Cham, 2012; Gu et al., 2012; Sun et al., 2013). Sun et al. (Sun et al., 2013) additionally filter the weight maps to make fusion more robust. Ferradans et al. (Ferradans et al., 2012) first apply an optical flow algorithm to register the images before combining the gradient.

Paul et al. (Paul et al., 2016) fuse differently the luminance and the chrominance values, the YCbCr (Gonzalez and Wood,) color space is used. The luminance is combined by using gradient based fusion. However, chromatic components are fused by direct pixel-wise averaging, where weight depends on how far is a pixel from a gray value.

Pixel wise methods are regularized by using pyramidal structures or working at the patch level. Burt

and Adelson (Burt and Adelson, 1983) proposed the Laplacian pyramid decomposition for image fusion. This decomposition was adopted by Mertens et al. (Mertens et al., 2007) algorithm. The weight maps are decomposed into a Gaussian pyramid $\mathbf{G}(w_k)^l$ and the input images into a Laplacian pyramid $\mathbf{L}(I_k)^l$, l denotes the level in the pyramid decomposition. For each level l the blended coefficients of the Laplacian pyramid are computed as

$$\mathbf{L}(u)^l(x) = \sum_{k=1}^K \mathbf{G}(w_k)^l(x) \mathbf{L}(I_k)^l(x). \quad (2)$$

Finally, the pyramid $\mathbf{L}(u)^l(x)$ is collapsed to obtain the fused image u .

Patch based methods use as minimal unity small image windows. The decision is more robust than pixel wise algorithms since much more values are involved. The first patch based algorithm was proposed by Goshtasby (Goshtasby, 2005). The image is divided into several non-overlapping blocks and the ones with the highest entropy are selected. Blocking artifacts are reduced by a blending function. In (Zhang et al., 2017a), Zhang et al. use patch correlation to detect motion and define the average distribution, which is applied with a multi-scale procedure. Zhang et al. (Zhang et al., 2017b) use super-pixel segmentation to detect motion. This detection permits to replace non corresponding parts by the specified reference image. Finally, a multi-scale approach uses gradient magnitude of each pixel to define average configuration. Recently Ma et al. (Ma et al., 2017) proposed to decompose the image patches into three components: signal strength, signal structure and mean intensity. Then, they fuse each component separately. Moreover, the direction of the signal structure provides information for deghosting.

3 PROPOSED FUSION ALGORITHM

We propose a novel algorithm for multi-exposure fusion adopting the Fourier aggregation model (Delbracio and Sapiro, 2015) as main tool. Compared to most existing MEF algorithms, we apply a different procedure for the luminance and chrominance components of the images. We use the YCbCr (Gonzalez and Wood,) color transformation for this separation. The luminance channel (Y) contains most geometry and image details while the chrominance (Cb, Cr) channels carry information about color.

3.1 Luminance Fusion

For the Fourier transform, it is well known that the magnitude decay of coefficients is related to the smoothness of the function. That is, the sharpness of detail information of an image is related to the amount of significant Fourier coefficients.

A local interpretation of the decay of the Fourier or DCT transform, indicates that under/over exposed patches will have Fourier coefficients of smaller magnitude, due to the lack of high frequency information. For shake removal, one supposes the existence of a true spectrum, modified in each image by a different kernel. For MEF, one can fuse the DCT coefficients in order to recover the full transform.

We locally apply a weighted average of each frequency depending on its magnitude, permitting to recover the most significant ones, and thus the exposed details. This strategy does not apply to the zero frequency Fourier coefficient, i.e. the mean. Large zero frequency coefficients correspond to over-exposed images, then applying the same weighted combination would simply overexpose the whole image. We use the Mertens algorithm (Mertens et al., 2009) to set the zero frequency coefficient.

Let assume we have a multi-exposure sequence of image luminances, supposed to be pre-registered, which we denote by Y_k , $k = 1, 2, \dots, K$. Since each image might contain well exposed areas, we apply the combination locally. We split the images Y_k into partially-overlapped blocks of $b \times b$ pixels, $\{B_k^l\}$, $l = 1, \dots, n_b$, n_b the number of blocks. We propose to fuse the non-zero frequencies of each block as follows:

$$\hat{B}^l(\xi) = \sum_{k=1}^K w_k^l(\xi) \hat{B}_k^l(\xi), \quad \xi \neq 0, \quad l = 1, 2, \dots, n_b, \quad (3)$$

where \hat{B}_k^l denotes the DCT transform of the block B_k^l and the weights $w_k^l(\xi)$ are defined depending on ξ as,

$$w_k^l(\xi) = \frac{|\hat{B}_k^l(\xi)|^p}{\sum_{n=1}^K |\hat{B}_n^l(\xi)|^p}, \quad \xi \neq 0. \quad (4)$$

The parameter p controls the weight of each Fourier mode. If $p = 0$ the fused frequency is just the arithmetic average, while for $p \rightarrow \infty$ each fused frequency takes the maximum value of the frequency in the sequence.

For $\xi = 0$, which corresponds to the mean intensity of the block B^l , the application of such a procedure would correspond to average those blocks with highest mean and therefore an over exposed image. Let YM be the luminance of the result of applying Mertens's algorithm (Mertens et al., 2009) to

the multi-exposure image sequence. We split it into blocks BM^l , $l = 1, \dots, n_b$ and define the zero frequency mode of each block as

$$\hat{B}^l(0) = \overline{BM^l} \quad l = 1, \dots, n_b \quad (5)$$

where $\overline{BM^l}$ denotes the mean value of the block BM^l .

Finally, if \mathcal{F}^{-1} denote the inverse Fourier transform we obtain the fused blocks,

$$B^l(x) = \mathcal{F}^{-1}(\hat{B}^l(\xi)), \quad l = 1, \dots, n_b. \quad (6)$$

Since blocks are partially overlapped, the pixels in overlapping are averaged to produce the final image.

3.2 Chrominance Fusion

We adopt a similar strategy to (Paul et al., 2016), and directly combine for each pixel, the values of the Cb, Cr channels at the same coordinates. The Cb, Cr components have a range from 16 to 240, being 128 the absence of color information. That is, if both chromatic components are equal to 128, the image is gray. In order to maximize the color information, color components further from 128 are privileged.

Let us denote the chrominance channels of the input multi-exposure sequence by Cb_k and Cr_k , $k = 1, 2, \dots, K$. The fused chrominance channels are represented as follows

$$Cb(x) = \sum_{k=1}^K w_k^b(x) Cb_k(x), \quad Cr(x) = \sum_{k=1}^K w_k^r(x) Cr_k(x) \quad (7)$$

We will use a non linear weight on the difference to the gray. We use an exponential kernel,

$$w_k^c(x) = \frac{\exp\left(\frac{(Cc_k - 128)^2}{\sigma^2}\right) - 1}{\sum_{j=1}^K \left(\exp\left(\frac{(Cc_j - 128)^2}{\sigma^2}\right) - 1\right)}, \quad c \in \{b, r\}. \quad (8)$$

The value of σ is set to $\sigma = 128$.

3.3 Enhancement

Recent methods (Li et al., 2012; Singh et al., 2014; Li et al., 2017) perform an additional enhancement to the fused details. This is achieved by first decomposing the images into a base and detail images. These components are fused independently and when combined an additional parameter permits to enhance the detail part.

We apply the detail enhancement separately of the MEF fusion chain. We use the screened-Poisson equation (Morel et al., 2014a; Morel et al., 2014b), to separate the details.

$$\lambda u(x) - \Delta u(x) = -\Delta f(x) \quad x \in \Omega \quad (9)$$



Figure 1: Exposure fusion comparison. From top to bottom and left to right: Mertens et al. (Mertens et al., 2009), Raman et al. (Raman and Chaudhuri, 2009), Gu et al. (Gu et al., 2012), Li et al. (Li et al., 2012), Li et al. (Li and Kang, 2012), Li et al. (Li et al., 2013), Ma et al. (Ma et al., 2017) and our result.

with homogeneous Neumann boundary condition. In (Morel et al., 2014a), the authors show that the solution of the screened Poisson equation (9) acts as a high-pass filter of f when λ increases, containing the details of f . The difference $f - u$ contains the basis image, responsible for the geometry.

Let f denote the fused image by applying the method introduced in sections 3.1, 3.2. The proposed detail enhancement consists in solving the equation (9), and thus splitting the image into a detail u and geometry $f - u$ parts. By recomposing the fused image with an additional detail enhancement we obtain

$$\hat{f}(x) = f(x) - u(x) + \alpha u(x) = f(x) + (\alpha - 1)u(x). \quad (10)$$

4 EXPERIMENTAL RESULTS

In this section we compare the proposed method with state of the art algorithms for exposure fusion. We compare with Mertens et al. (Mertens et al., 2009), Raman et al. (Raman and Chaudhuri, 2009), Gu et al. (Gu et al., 2012), Li et al. (Li et al., 2012), Li et al. (Li and Kang, 2012), Li et al. (Li et al., 2013) and Ma et al. (Ma et al., 2017). All results except for the last method were taken from the dataset provided in (Zeng et al., 2014) and (Ma et al., 2015). This database (Zeng et al., 2014; Ma et al., 2015) contains seventeen input images with multiple exposure levels (≥ 3) together with fused images generated by eight state-of-the-art image fusion algorithms. The results from Ma et al. (Ma et al., 2017) were computed with the software downloaded from the author's page. In all cases, default parameter settings are adopted.

Our results were computed using the same parameters for all tests in this section. The parameters for the weighted average of the Fourier modes and the chrominance channels are set to $p = 7$ and $\sigma = 128$ respectively. Finally for the enhancement we set $\lambda = 0.5$ and $\alpha = 1.5$.

Figure 1 displays the results of all methods in the "Belgium House" data set. A zoomed detail of these images is displayed in Figure 2. The results by Raman et al. (Raman and Chaudhuri, 2009) and Gu et al. (Gu et al., 2012) are quite dark and blurred. The fused image by Li et al. (Li et al., 2013), having a better luminance than the two previous methods, it has large regions with a darker value than expected, as for example in the wall below the painting. The rest of the methods have a good global illumination. However, looking closer to the details in Figure 2, we might observe that many outdoor details in Mertens et al. (Mertens et al., 2009), Li et al. (Li et al., 2012) and Ma et al. (Ma et al., 2017) are overexposed, losing its definition. We might also observe that details in Li et al. (Li et al., 2012) are excessively enhanced, making them look like unnatural. Our result seems to be the best compromise having a natural look and a good definition of details.

Figure 3 and 5 respectively display the results with the "House" and "Cadik" tests and corroborate observations on previous figure. Figures 4 and 6 shows zoomed details of figure 3 and 5, respectively. For the "House" set, it must be noted the excessive detail and color enhancement of Ma et al. (Ma et al., 2017), the poor luminance balance of Li et al. (Li and Kang, 2012) and the detail blur in Li et al. (Li et al., 2013). For the "Cadik" set, let note that only Li et al. (Li et al., 2013) and our result are able to correctly avoid

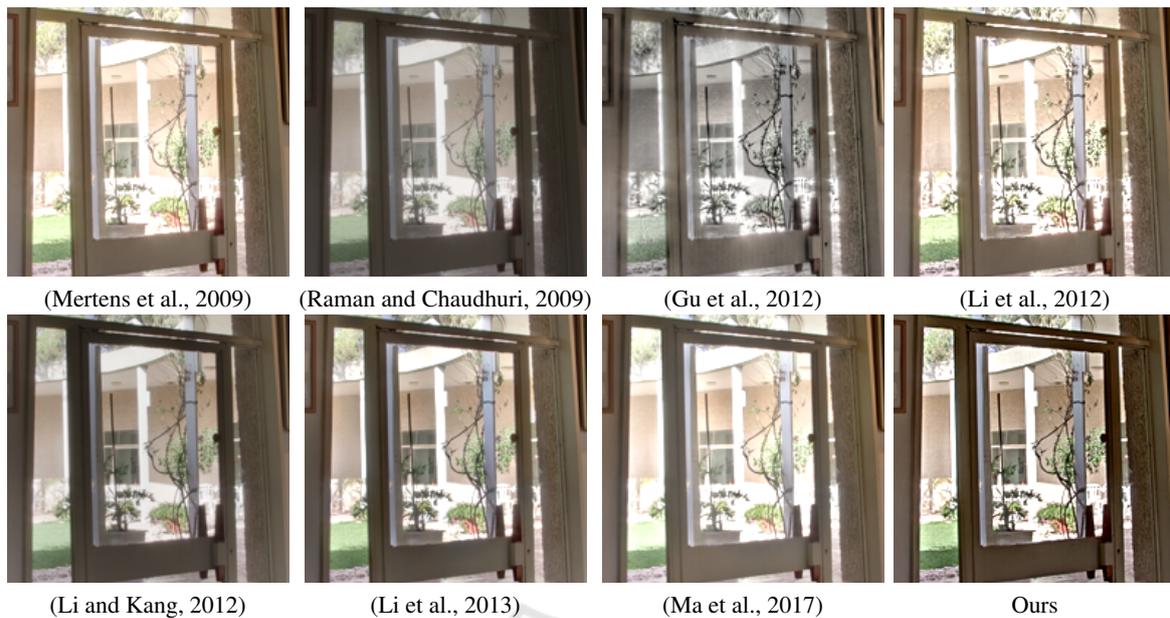


Figure 2: Detail of images in Fig. 1. We observe that many outdoor details in Mertens et al. (Mertens et al., 2009), Li et al. (Li et al., 2012) and Ma et al. (Ma et al., 2017) are overexposed. Details in Li et al. (Li et al., 2012) are excessively enhanced, making them look like unnatural. Our result looks natural and details are well defined.

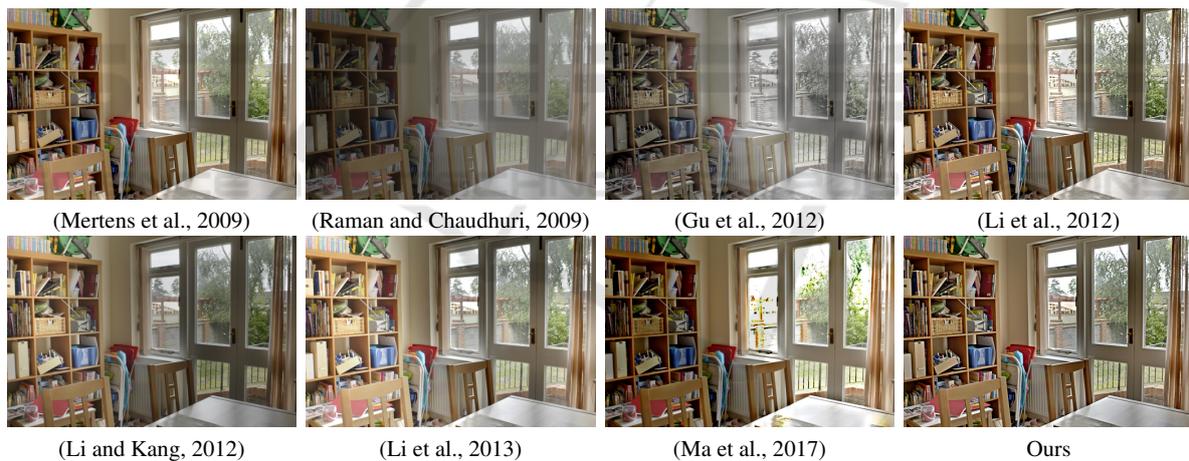


Figure 3: Exposure fusion comparison. From top to bottom and left to right: Mertens et al. (Mertens et al., 2009), Raman et al. (Raman and Chaudhuri, 2009), Gu et al. (Gu et al., 2012), Li et al. (Li et al., 2012), Li et al. (Li and Kang, 2012), Li et al. (Li et al., 2013), Ma et al. (Ma et al., 2017) and our result.

saturation near the lamp's light. However, details on Li et al. (Li et al., 2013) are excessively blurred.

5 CONCLUSION

We have presented a novel fusion algorithm for multi-exposure images. The block based method uses the DCT coefficient instead of the traditional gradient combination. Moreover, we propose to apply an enhancement to the fused image by applying the

screened Poisson equation.

Experiments have shown the improvement of the proposed algorithm compared to state of the art.

REFERENCES

- Burt, P. J. and Adelson, E. H. (1983). The laplacian pyramid as a compact image code. *IEEE TRANSACTIONS ON COMMUNICATIONS*, 31:532–540.
- Debevec, P. E. and Malik, J. (2008). Recovering high dy-

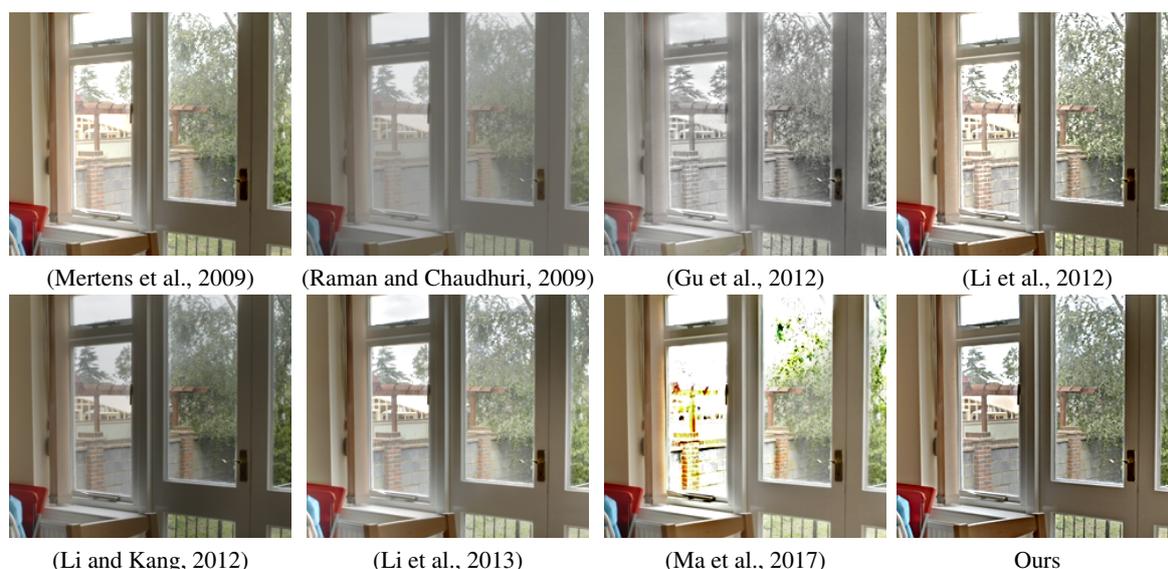


Figure 4: Detail of images in Fig. 3. We note the excessive detail and color enhancement of Ma et al. (Ma et al., 2017), the poor luminance balance of Li et al. (Li and Kang, 2012) and the detail blur in Li et al. (Li et al., 2013). Our result has no noticeable artifacts and a well detail definition.

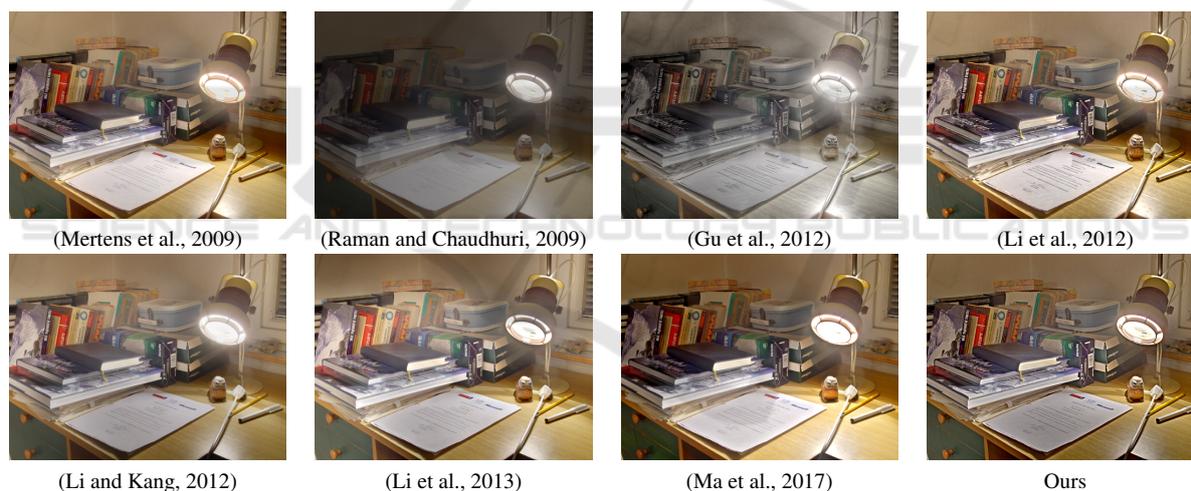


Figure 5: Exposure fusion comparison. From top to bottom and left to right: Mertens et al. (Mertens et al., 2009), Raman et al. (Raman and Chaudhuri, 2009), Gu et al. (Gu et al., 2012), Li et al. (Li et al., 2012), Li et al. (Li and Kang, 2012), Li et al. (Li et al., 2013), Ma et al. (Ma et al., 2017) and our result.

dynamic range radiance maps from photographs. In *ACM SIGGRAPH 2008 classes*, page 31. ACM.

Delbracio, M. and Sapiro, G. (2015). Removing Camera Shake via Weighted Fourier Burst Accumulation. *Image Processing, IEEE Transactions on*, 24(11):3293–3307.

Ferradans, S., Bertalmio, M., Provenzi, E., and Caselles, V. (2012). Generation of hdr images in non-static conditions based on gradient fusion. In *VISAPP (1)*, pages 31–37.

Gonzalez, R. C. and Wood, R. E. *Digital image processing*, 2nd edtn.

Goshtasby, A. A. (2005). Fusion of multi-exposure images. *Image Vision Comput.*, 23(6):611–618.

Gu, B., Li, W., Wong, J., Zhu, M., and Wang, M. (2012). Gradient field multi-exposure images fusion for high dynamic range image visualization. *Journal of Visual Communication and Image Representation*, 23(4):604 – 610.

Kuk, J. G., Cho, N. I., and Lee, S. U. (2011). High dynamic range (hdr) imaging by gradient domain fusion. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pages 1461–1464. IEEE.

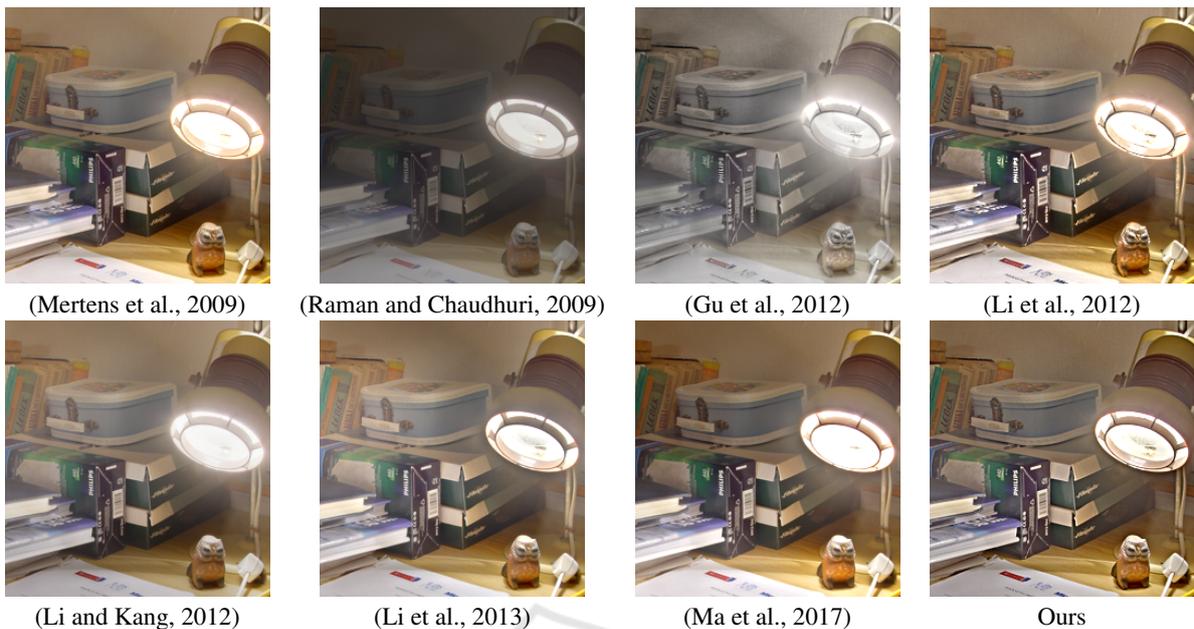


Figure 6: Detail of images in Fig. 5. Li et al. (Li et al., 2013) and our result are able to correctly avoid saturation near the lamp's light. However, details on Li et al. (Li et al., 2013) are excessively blurred.

- Li, S. and Kang, X. (2012). Fast multi-exposure image fusion with median filter and recursive filter. *IEEE Transactions on Consumer Electronics*, 58(2).
- Li, S., Kang, X., and Hu, J. (2013). Image fusion with guided filtering. *IEEE Trans. Image Processing*, 22(7):2864–2875.
- Li, Z., Wei, Z., Wen, C., and Zheng, J. (2017). Detail-enhanced multi-scale exposure fusion. *IEEE Transactions on Image Processing*, 26(3):1243–1252.
- Li, Z. G., Zheng, J. H., and Rahardja, S. (2012). Detail-enhanced exposure fusion. *IEEE Trans. Image Processing*, 21(11):4672–4676.
- Liu, Y. and Wang, Z. (2015). Dense sift for ghost-free multi-exposure fusion. *Journal of Visual Communication and Image Representation*, 31:208–224.
- Ma, K., Li, H., Yong, H., Wang, Z., Meng, D., and Zhang, L. (2017). Robust multi-exposure image fusion: A structural patch decomposition approach. *IEEE Trans. Image Processing*, 26(5):2519–2532.
- Ma, K., Zeng, K., and Wang, Z. (2015). Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 24(11):3345–3356.
- Mertens, T., Kautz, J., and Reeth, F. V. (2007). Exposure fusion. In *Proceedings of the 15th Pacific Conference on Computer Graphics and Applications*, PG '07, pages 382–390, Washington, DC, USA. IEEE Computer Society.
- Mertens, T., Kautz, J., and Van Reeth, F. (2009). Exposure Fusion: A Simple and Practical Alternative to High Dynamic Range Photography. *Computer Graphics Forum*.
- Morel, J.-M., Petro, A.-B., and Sbert, C. (2014a). Automatic correction of image intensity non-uniformity by the simplest total variation model. *Methods and Applications of Analysis*, 21(1):107–120.
- Morel, J.-M., Petro, A.-B., and Sbert, C. (2014b). Screened Poisson Equation for Image Contrast Enhancement. *Image Processing On Line*, 4:16–29.
- Paul, S., Sevcenco, I. S., and Agathoklis, P. (2016). Multi-exposure and multi-focus image fusion in gradient domain. *Journal of Circuits, Systems and Computers*, 25(10):1650123.
- Pérez, P., Gangnet, M., and Blake, A. (2003). Poisson image editing. In *ACM Transactions on graphics (TOG)*, volume 22, pages 313–318. ACM.
- Raman, S. and Chaudhuri, S. (2009). Bilateral Filter Based Compositing for Variable Exposure Photography. In Alliez, P. and Magnor, M., editors, *Eurographics 2009 - Short Papers*. The Eurographics Association.
- Raskar, R., Ilie, A., and Yu, J. (2005). Image fusion for context enhancement and video surrealism. In *ACM SIGGRAPH 2005 Courses*, page 4. ACM.
- Reinhard, E., Ward, G., Pattanaik, S., and Debevec, P. (2005). *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting (The Morgan Kaufmann Series in Computer Graphics)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- Singh, H., Kumar, V., and Bhooshan, S. (2014). A novel approach for detail-enhanced exposure fusion using guided filter. *The Scientific World Journal*, 2014(659217):8.
- Sun, J., Zhu, H., Xu, Z., and Han, C. (2013). Poisson image fusion based on markov random field fusion model. *Information fusion*, 14(3):241–254.
- Zeng, K., Ma, K., Hassen, R., and Wang, Z. (2014). Perceptual evaluation of multi-exposure image fusion al-

- gorithms. In *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 7–12.
- Zhang, W. and Cham, W.-K. (2012). Gradient-directed multi-exposure composition. *IEEE Transactions on Image Processing*, 21(4):2318–2323.
- Zhang, W., Hu, S., and Liu, K. (2017a). Patch-based correlation for deghosting in exposure fusion. *Information Sciences*, 415:19–27.
- Zhang, W., Hu, S., Liu, K., and Yao, J. (2017b). Motion-free exposure fusion based on inter-consistency and intra-consistency. *Information Sciences*, 376:190–201.

