

PAS-MEF: MULTI-EXPOSURE IMAGE FUSION BASED ON PRINCIPAL COMPONENT ANALYSIS, ADAPTIVE WELL-EXPOSEDNESS AND SALIENCY MAP

Diclehan Karakaya, Oguzhan Ulucan, Mehmet Turkan

Department of Electrical and Electronics Engineering, Izmir University of Economics, Izmir, Turkey

ABSTRACT

High dynamic range (HDR) imaging enables to immortalize natural scenes similar to the way that they are perceived by human observers. With regular low dynamic range (LDR) capture/display devices, significant details may not be preserved in images due to the huge dynamic range of natural scenes. To minimize the information loss and produce high quality HDR-like images for LDR screens, this study proposes an efficient multi-exposure fusion (MEF) approach with a simple yet effective weight extraction method relying on principal component analysis, adaptive well-exposedness and saliency maps. These weight maps are later refined through a guided filter and the fusion is carried out by employing a pyramidal decomposition. Experimental comparisons with existing techniques demonstrate that the proposed method produces very strong statistical and visual results.

Index Terms— High dynamic range, multi-exposure image fusion, principal component analysis, saliency map, guided filtering

1. INTRODUCTION

High dynamic range (HDR) technology aims at producing high quality images similar to human perception. However, the dynamic range gap between high contrast scenes and low dynamic range (LDR) capture/display devices causes information loss in highlights and shadows [1]. In order to minimize distortions and detail loss, i.e., to capture and display high quality images, there are three main approaches: (*i*) using HDR compatible capture and display devices, (*ii*) employing tone-mapping operators to map HDR onto LDR displays, and (*iii*) using multi-exposure LDR fusion (MEF) to create HDR-like content for LDR screens [2]. The user-grade technology manufacturers generally prefer to use MEF to obtain high quality LDR images, since MEF has significantly lower cost than hardware-based solutions; and with MEF, it is also possible to avoid tone-mapping related problems such as low-subjective contrast and color saturation [3, 4].

MEF mainly aims at keeping the most informative parts of each exposure image via extracting weight maps, and then it blends them into a single HDR-like image [1]. There are several MEF studies present in the literature. In the milestone study of Mertens *et al.* [1], a weight map extraction scheme is proposed which is based on contrast, saturation and well-exposedness. The fusion of the input stack is carried out by taking the Gaussian pyramid of weight maps and the Laplacian pyramid of exposures, which is inspired from Burt and Kolczynski [5]. In a recent study of Lee *et al.* [6], a weight map is formed by employing an adapted version of well-exposedness of Mertens and a second map is characterized via the gradient information in each exposure. Finally, a pyramidal image decomposition is employed to carry out the fusion process. In Li and Zhang [7] (Li18), convolutional neural networks (CNNs) are employed for MEF. The first layer of a pretrained classification network is used for feature extraction. Then, local visibility and temporal consistency maps are

extracted and adopted for the weighted fusion operation. In Liu and Leung [8], a method is proposed for both MEF and decolorization. In this study, the local gradient information of each exposure is extracted and provided to a CNNs model. The proposed algorithm produces satisfying results, however it can operate on stacks consisting of three exposures only. In Hayat and Imran [9], the weight map characterization scheme of Li and Kang [10] is modified with the dense-SIFT descriptor, and guided filtering is used to eliminate discontinuities and noise in these maps. Finally, image fusion is conducted via a pyramid decomposition approach. In the study of Li *et al.* [11] (Li20), the method of Ma *et al.* [12] is investigated and improved by forming adopted weight maps via signal strength, signal structure and mean intensities. Recursive downsampling and processing are included into the model and halo effects are significantly reduced. In a recent study of Ulucan *et al.* [2], a MEF algorithm is developed which is based on linear embeddings of images and watershed masking. Weights maps are first characterized via linear embeddings of exposure image patch spaces, while preserving local geometry of the sampled manifold structure. These weight maps are then refined via watershed masking to highlight most informative parts of each exposure in the input stack. Lastly, the fusion process is performed through weighted averaging.

As it can be deduced from above studies, each specific algorithm generally differs in the way of extraction and/or characterization of weight maps. Therefore, new weight map extraction methods will enlighten the path leading to a general map formation framework. To this end, in this paper, a novel MEF method is proposed to fuse static exposure stacks. The weight map extraction algorithm relies on principal component analysis (PCA), adaptive well-exposedness and saliency map features. These maps are later refined by a guided filter, and then exposure images are fused via a pyramidal decomposition. The proposed method is compared with well-known MEF algorithms and it demonstrates very strong outputs both statistically and visually. To the best of available knowledge, this is the first time that PCA is used for extracting weights in MEF. Furthermore, this is the first study that modifies the well-exposedness (i.e., indicating brightness) feature to be fully adaptive, while existing MEF algorithms employ a fixed parameter and/or constant for well-exposedness. Finally, it is worth mentioning here that saliency maps are used in this study to mimic the human visual system (HVS) and give larger weights to the “best” parts of images as in the primary visual cortex.

This paper is organized as follows. The proposed MEF technique is detailed in Sec. 2. Experimental results are presented and discussed in Sec. 3. Lastly, a brief summary and possible future directions for this study are given in Sec. 4.

2. THE PROPOSED MEF METHOD: PAS-MEF

A simple flowchart of the proposed method is given in Fig. 1. Given the input exposures, there are three branches for extracting PCA,

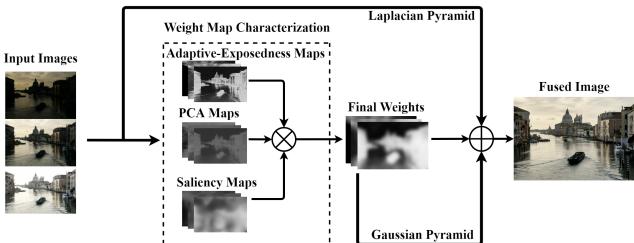


Fig. 1: A flowchart of the proposed MEF method.

adaptive well-exposedness and saliency maps. These maps are later combined to obtain final fusion weights in order to output a fused image via a pyramidal decomposition.

Weight extraction via PCA. PCA performs an orthogonal transformation to obtain linearly uncorrelated variables from possibly correlated data [13]. The correlated data is projected onto the PCA space by taking advantage of the eigenvectors of the covariance matrix. The first principal component is in the direction of the highest variance (first eigenvector), while the second principal component lies in the subspace perpendicular to the first one and the next principal components are computed similarly. The representations of data in the PCA space can be called as *scores*.

PCA has already been used in image fusion but, to the best of available knowledge, it has not been employed in MEF studies [14]. Therefore, it is investigated in this study first by vectorizing N number of gray-scale versions of exposure images $I_n, n = 1 \dots N$, into column vectors of the size $rc \times 1$ where r and c represent the number of rows and columns of each image, respectively. The obtained N column vectors are then stacked into the columns of an $rc \times N$ data matrix, in which there are rc observations with N variables each, for calculating the scores of observations via PCA. Subsequently, each variable-score vector is linearly normalized to have a range between [0 1], then reshaped back to an $r \times c$ matrix followed by a simple smoothing Gaussian filter. Finally, N number of PCA weight maps (P_n) are obtained with a sum-to-one normalization applied at each spatial position. As an example, the extracted PCA maps of the *Venice* stack are demonstrated in Fig. 2.

Weight extraction via adaptive well-exposedness. The well-exposedness feature is initially introduced by Mertens. For a given exposure image I_n , it is extracted for each red-green-blue channel separately using a Gaussian curve as $\exp(-\frac{(I_n - 0.5)^2}{2\sigma^2})$ where $\sigma = 0.2$. This weight aims at keeping pixel intensities which are not too close to 0 (under-exposed) or 1 (over-exposed), hence it favors pixels in well-exposed regions with intensity values close to 0.5. However, this exposedness feature sometimes can not sufficiently preserve bright regions of short-exposure images, as well as dark regions of long-exposure images [6]. To overcome this problem, Lee proposed an adaptive well-exposedness based on the mean of pixel intensities of exposure images. In this adaptive form, the constant values of 0.5 and σ of Mertens are replaced with some functions of the mean of pixel intensities of I_n and its neighboring exposure images I_{n-1} and I_{n+1} . However, there still exists a fixed constant parameter to calculate the σ value.

In this study, a fully adaptive weight calculation is proposed for the well-exposedness scheme in order to allocate large weights for dark regions when the image is a long-exposure, as well as for bright regions when the image is a short-exposure. All computations are

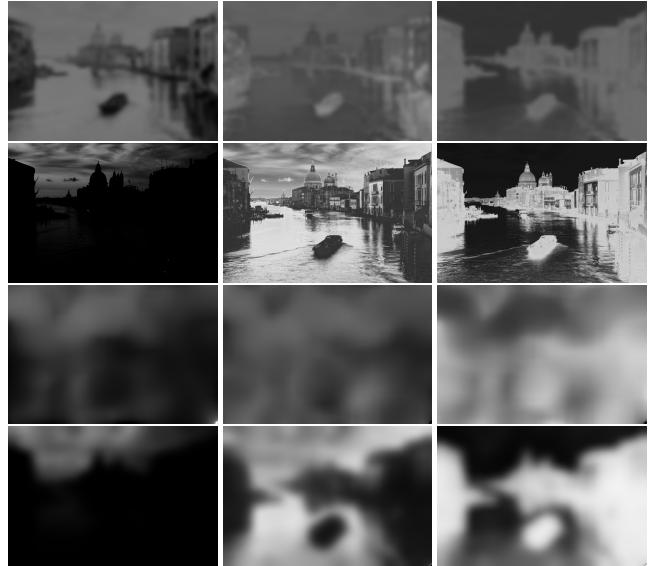


Fig. 2: Weights for *Venice*. (Top-to-bottom) PCA maps, adaptive well-exposedness weights, saliency maps, final fusion weights.

carried out on the luminance channel Y_n of I_n as given in Eqn. (1),

$$A_n = \exp\left(-\frac{\left(Y_n - (1 - \mu_{Y_n})\right)^2}{2\sigma_{Y_n}^2}\right) \quad (1)$$

where μ_{Y_n} and σ_{Y_n} represent the mean of pixel intensities and the standard deviation of Y_n , respectively. This finally leads to an adaptive algorithm since the Gaussian curve controlling parameters are extracted via self statistical information of each single exposure, and larger weights are given to the “best” luminance intensities of the input. The obtained adaptive well-exposedness weights (A_n) of the *Venice* stack are illustrated in Fig. 2.

Weight extraction via saliency map. When processed in the HVS, the attention that objects gather depends on the task at hand and stimulus-driven factors such as prominent colors [15]. In the literature, several computational models are proposed to mimic the HVS; thus to highlight salient regions, increase the visual appeal and quality of images.

In this study, saliency maps are used for assigning larger weights to regions which are more attractive to human observers. Since the design of a salient region detection algorithm is out of context, the technique introduced by Hou *et al.* [16] is integrated into the proposed MEF method. This saliency algorithm is based on a descriptor called *image signature*, which is the sign of the Discrete Cosine Transform (DCT) coefficients. Briefly, the DCT is first computed for each red-green-blue channel separately. Then, image reconstruction is carried out by calculating the inverse DCT of the sign of the DCT coefficients. The saliency maps (S_n) are finally obtained from the reconstructed image. For detailed information, the reader may refer to [16]. The saliency maps of the *Venice* stack are shown in Fig. 2.

Weight refinement and fusion. After all three weight maps are characterized, they are combined to form a single refined map for each exposure in the input stack given in Eqn. (2) as follows,

$$W_n = \text{GuidFilt}(P_n \times A_n \times S_n), \quad n = 1 \dots N, \quad (2)$$

where *GuidFilt* is an edge-aware (edge-preserving) smoothing filter called *guided filter* [17], which is generally used to eliminate pos-



Fig. 3: *Laurenziana*: (Left) Ulucan (0.989), (right) Proposed (0.991).

sible discontinuities and noise in weight maps, e.g., [9]. All these maps W_n are finally normalized to satisfy a sum-to-one constraint at each spatial position to form the final weights for fusion. The obtained final fusion weights of the *Venice* stack are given in Fig. 2.

A pyramidal decomposition is applied to blend the input stack in order to further avoid artifacts such as halo effects at sharp texture and color changeovers [5, 1]. In detail, the Laplacian pyramid (L) is employed to decompose each input exposure into ℓ -levels of distinct resolutions, while the Gaussian pyramid (G) to carry out a similar operation for final fusion weights. The blending operation is applied at each pyramidal level, and as a result a fused Laplacian pyramid is obtained for the fused image given in Eqn. (3) as follows,

$$L\{F^\ell\} = \sum_{n=1}^N G\{W_n^\ell\} \times L\{I_n^\ell\} \quad (3)$$

where the fused pyramid $L\{F^\ell\}$ is finally collapsed to acquire the final fused image F .

3. EXPERIMENTAL RESULTS

The proposed MEF algorithm (PAS-MEF) is compared to Mertens [1], Lee [6], Li18 [7], Liu [8], Hayat [9], Ulucan [2] and Li20 [11] over 13 image stacks obtained from datasets in [18, 19, 20]. All experiments are performed on an AMD Ryzen(TM) 5 3600x CPU @ 3.80 GHz 6-core 16GB RAM machine using MATLAB R2020a. All competing algorithms are employed with their default settings, including the proposed method, without any optimization. A statistical performance analysis is performed through the multi-scale structural similarity framework for MEF, i.e., MEF-SSIM [19]. MEF-SSIM is a perceptual quality assessment metric, which takes both the global luminance consistency and the local structure preservation into account to produce statistical results in the range [0 1]. A score closer to 1 indicates a better perceptual quality.

The obtained statistical scores of all algorithms are reported in Table 1, in which the bottom three rows present the average accuracy (avg), standard deviation (std) and average execution time (run-time) of each method. It can be clearly observed that PAS-MEF produces highly competitive results (i.e., being in the second best place) and surpasses six of the competing state-of-the-art MEF approaches on average. In addition, these statistical results indicate that the standard



Fig. 4: *Tower*: (Left) Ulucan (0.982), (right) Proposed (0.983).

deviation of PAS-MEF scores is the smallest (together with Li20) and the computational complexity is very conceivable.

A side-by-side visual comparison of the fusion outputs of PAS-MEF and Ulucan is given for the *Laurenziana* stack in Fig. 3. In this particular example, PAS-MEF reaches the highest MEF-SSIM score when compared to other competing techniques. It can be clearly seen that the sky has a more natural color in the output of PAS-MEF. Furthermore, the details on the rooftops are better preserved in the result of this technique, while the tree in the middle has more vivid colors in Ulucan. Another visual comparison for the *Tower* stack is demonstrated in Fig. 4. The grass and flowers on the foreground are much better recovered in the proposed PAS-MEF, while clouds are better preserved and the tower has vivid colors in Ulucan.

In Fig. 5, the fusion results are presented for the *Mask* stack. PAS-MEF produces the highest MEF-SSIM score (together with Hayat) for this exposure sequence. When compared to Li20 which has lower brightness and less information in several parts in the fused image, the proposed method clearly preserves the details of the building and the mask. On the other side, the sky has a more plausible color in Li20.

In Fig. 6, the fusion outputs of PAS-MEF and Hayat are compared for the *Kluki* stack. Although PAS-MEF has its lowest MEF-SSIM score for this exposure sequence among other stacks in the dataset, the sky has more well-settled colors in blue regions, and the rooftop of the house and the grass have more vivid colors when compared to Hayat, whose MEF-SSIM is slightly higher.

It can be concluded from Table 1 that PAS-MEF reaches its highest MEF-SSIM score for the *Chinese Garden* stack and some visual fusion results are presented in Fig. 7. When compared to Lee, the sky region is more plausible in the proposed technique and overall a natural-looking image is obtained while avoiding any artifacts. In addition, it can be observed from Fig. 8 that both PAS-MEF and Lee output natural-looking images for *Venice*. However, the sky and buildings on the left contain more vivid colors in PAS-MEF, which led to a significantly higher MEF-SSIM score than Lee.

Further fusion examples are illustrated in Fig. 9 and Fig. 10 for *Lighthouse* and *OldHouse*, respectively. The rooftop of the light house, rocks and trees in the background have more striking colors in PAS-MEF, while a darker output is generated by Li20. Next in Fig. 10, Li20 contains very bright regions on the old house. The building has more well-settled colors in PAS-MEF which has the highest MEF-SSIM score when compared to other methods.

Table 1: MEF-SSIM scores for each exposure stack used in experiments. The highest scores are in boldface.

	Mertens	Lee	Li18	Liu	Hayat	Ulucan	Li20	PAS-MEF
<i>Arno</i>	0.991	0.987	0.948	0.985	0.985	0.986	0.990	0.989
<i>Chinese Garden</i>	0.989	0.990	0.977	0.988	0.993	0.991	0.994	0.993
<i>Church</i>	0.989	0.992	0.980	0.977	0.992	0.989	0.992	0.991
<i>Farmhouse</i>	0.981	0.979	0.974	0.978	0.984	0.978	0.986	0.981
<i>Flowers</i>	0.964	0.990	0.972	0.990	0.995	0.989	0.995	0.990
<i>Kluki</i>	0.980	0.974	0.957	0.973	0.980	0.963	0.983	0.979
<i>Laurenziana</i>	0.988	0.987	0.973	0.987	0.989	0.989	0.989	0.991
<i>Lighthouse</i>	0.980	0.979	0.962	0.985	0.974	0.975	0.978	0.982
<i>Mask</i>	0.987	0.990	0.975	0.985	0.992	0.987	0.991	0.992
<i>Office</i>	0.985	0.991	0.970	0.985	0.987	0.991	0.989	0.984
<i>OldHouse</i>	0.974	0.990	0.962	0.988	0.968	0.991	0.990	0.992
<i>Tower</i>	0.986	0.987	0.981	0.983	0.987	0.982	0.988	0.983
<i>Venice</i>	0.966	0.972	0.947	0.973	0.972	0.978	0.984	0.980
avg	0.981	0.985	0.967	0.982	0.984	0.983	0.988	0.986
std	0.009	0.007	0.011	0.006	0.009	0.008	0.005	0.005
run-time (sec)	0.36	0.38	0.93	50.68	0.89	2.91	0.31	0.66



Fig. 5: *Mask*: (Left) Li20 (0.991), (right) Proposed (0.992).



Fig. 8: *Venice*: (Left) Lee (0.972), (right) Proposed (0.980).



Fig. 6: *Kluki*: (Left) Hayat (0.980), (right) Proposed (0.979).



Fig. 9: *Lighthouse*: (Left) Li20 (0.978), (right) Proposed (0.982).



Fig. 7: *Chinese Garden*: (Left) Lee (0.990), (right) Proposed (0.993).



Fig. 10: *OldHouse*: (Left) Li20 (0.990), (right) Proposed (0.992).

4. CONCLUSION

MEF is commonly used for obtaining HDR-like high quality images and numerous studies are present in this field. In general, the existing methods differ in the weight map characterization process. In this study, a novel weight extraction method is introduced which is based on PCA, adaptive well-exposedness and saliency maps. The obtained weights are refined via a guided filter and then image fusion is carried out through a pyramidal decomposition. The proposed algo-

rithm presents very strong results both statistically and visually, and it outperforms several state-of-the-art MEF techniques. It is worth noting here that, to the best of available knowledge, this is the first study which incorporates PCA and fully adaptive well-exposedness into the MEF problem.

As a future direction, the proposed algorithm will be optimized for increasing its statistical and visual performance, and for further reducing the run-time complexity. Moreover, PAS-MEF will be extended to fuse dynamic scenes in the image stack.

5. REFERENCES

- [1] T. Mertens, J. Kautz, and F. Van Reeth, “Exposure fusion: a simple and practical alternative to high dynamic range photography,” *Comp. Graph. Forum*, vol. 28, no. 1, pp. 161–171, February 2009.
- [2] O. Ulucan, D. Karakaya, and M. Turkan, “Multi-exposure image fusion based on linear embeddings and watershed masking,” *Signal Process.*, vol. 178, pp. 107791, January 2021.
- [3] A. O. Akyuz and E. Reinhard, “Color appearance in high-dynamic-range imaging,” *J. Electron. Imag.*, vol. 15, no. 3, pp. 033001, July 2006.
- [4] C. Kiser, E. Reinhard, M. Tocci, and N. Tocci, “Real time automated tone mapping system for HDR video,” in *IEEE Int. Conf. Image Process.*, 2012, pp. –.
- [5] P. J. Burt and R. J. Kolczynski, “Enhanced image capture through fusion,” in *IEEE Int. Conf. Comput. Vis.*, 1993, pp. 173–182.
- [6] S. Lee, J. S. Park, and N. I. Cho, “A multi-exposure image fusion based on the adaptive weights reflecting the relative pixel intensity and global gradient,” in *IEEE Int. Conf. Image Process.*, 2018, pp. 1737–1741.
- [7] H. Li and L. Zhang, “Multi-exposure fusion with CNN features,” in *IEEE Int. Conf. Image Process.*, 2018, pp. 1723–1727.
- [8] Q. Liu and H. Leung, “Variable augmented neural network for decolorization and multi-exposure fusion,” *Inf. Fusion*, vol. 46, pp. 114–127, March 2019.
- [9] N. Hayat and M. Imran, “Ghost-free multi exposure image fusion technique using dense SIFT descriptor and guided filter,” *J. Vis. Commun. Image Represent.*, vol. 62, pp. 295–308, July 2019.
- [10] S. Li and X. Kang, “Fast multi-exposure image fusion with median filter and recursive filter,” *IEEE Trans. Consum. Electronics*, vol. 58, no. 2, pp. 626–632, July 2012.
- [11] H. Li, K. Ma, H. Yong, and L. Zhang, “Fast multi-scale structural patch decomposition for multi-exposure image fusion,” *IEEE Trans. Image Process.*, vol. 29, pp. 5805–5816, April 2020.
- [12] K. Ma, H. Li, H. Yong, Z. Wang, D. Meng, and L. Zhang, “Robust multi-exposure image fusion: a structural patch decomposition approach,” *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2519–2532, February 2017.
- [13] S. Wold, K. Esbensen, and P. Geladi, “Principal component analysis,” *Chemom. Intell. Lab. Syst.*, vol. 2, no. 1-3, pp. 37–52, August 1987.
- [14] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, “Pixel-level image fusion: A survey of the state of the art,” *Inf. Fusion*, vol. 33, pp. 100–112, January 2017.
- [15] N. D. B. Bruce and J. K. Tsotsos, “Saliency, attention, and visual search: An information theoretic approach,” *J. Vision*, vol. 9, no. 3, pp. 5–5, March 2009.
- [16] X. Hou, J. Harel, and C. Koch, “Image signature: Highlighting sparse salient regions,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 194–201, January 2012.
- [17] K. He, J. Sun, and X. Tang, “Guided image filtering,” in *Eur. Conf. Comput. Vision*, 2010, pp. 1–14.
- [18] Y. Fang, H. Zhu, K. Ma, Z. Wang, and S. Li, “Perceptual evaluation for multi-exposure image fusion of dynamic scenes,” *IEEE Trans. Image Process.*, vol. 29, pp. 1127–1138, September 2019.
- [19] K. Ma, K. Zeng, and Z. Wang, “Perceptual quality assessment for multi-exposure image fusion,” *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3345–3356, June 2015.
- [20] I. Merianos and N. Mitianoudis, “Multiple-exposure image fusion for HDR image synthesis using learned analysis transformations,” *J. Imaging*, vol. 5, no. 3, pp. 32, February 2019.