VISUAL SALIENCE AND STACK EXTENSION BASED GHOST REMOVAL FOR HIGH-DYNAMIC-RANGE IMAGING

Zijie Wang[1,2], *Qin Liu*[1,†], *Takeshi Ikenaga*[2]

[1] Software Institute, Nanjing University, China
[2] Graduate School of Information, Production, and Systems, Waseda University, Japan
[†] Email: qinliu@software.nju.edu.cn

ABSTRACT

High-dynamic-range imaging (HDRI) techniques are proposed to extend the dynamic range of captured images against sensor limitation. The key issue of multi-exposure fusion in HDRI is removing ghost artifacts caused by motion of moving objects and handheld cameras. This paper proposes a ghost-free HDRI algorithm based on visual salience and stack extension. To improve the accuracy of ghost areas detection, visual salience based bilateral motion detection is introduced to measure image differences. For exposure fusion, the proposed algorithm reduces brightness discontinuity and enhances details by stack extension, and rejects the information of ghost areas to avoid artifacts via fusion masks. Experiment results show that the proposed algorithm can remove ghost artifacts accurately for both static and handheld cameras, remain robust to scenes with complex motion and keep low complexity over recent advances including patch based method and rank minimization based method by 20.4% and 63.6% time savings on average.

Index Terms— High-dynamic-range imaging (HDRI), visual salience, motion detection, ghost removal, exposure fusion

1. INTRODUCTION

The dynamic range of an actual scene is far higher than sensor limitation of consumer cameras. Therefore, high-dynamic-range imaging (HDRI) techniques are proposed to enlarge the range of luminance, which benefit photography, realistic rendering, feature detection [1], etc. A host of studies have focused on HDRI via multi-exposure fusion [2][3] and tried to remove ghost artifacts caused by moving objects and camera motion. In general, these approaches can be classified into two categories.

The first category, alignment based method, assumes that each input image has similarity to the HDR image and removes ghosts by non-rigid registration using their local properties such as correspondence estimation. Bogoni [4] firstly proposed local motion estimation by introducing optical flow to warp the images. Kang [5] made improvements by introducing hierarchical homography. Jinno [6] employed Markov Random Fields (MRF) to solve the estimation problem. Zimmer [7] reduced brightness influence by estimating correspondence in gradient domain. Sen [8] made a significant progress by using patchmatch [9]. However, high complexity is required for optimization and mismatch causes new artifacts.

The second category, motion detection based method, assumes that moving objects affect different regions in different images, and detects motion areas to reject the ghost artifacts before HDR synthesizing. Reinhard [10] firstly used weighted normalized variance to detect motion. Gallo [11] measured ghost effects by deviation from linearity. Heo [12] detected ghost regions by the joint probability density of multi-exposure images and energy minimization. Zhou [13] proposed a unified framework for moving object detection by detecting contiguous outliers in low-rank representation. Lee [14] formulated ghost region detection as a rank minimization problem. Zhang [15] proposed an area extension method to avoid ghost artifacts remained around moving objects. These approaches can keep low complexity and avoid introducing new artifacts. However, some algorithms do not introduce local properties to motion detection; furthermore, there remain artifacts of brightness discontinuity and detail loss.

The proposal makes a further work on the second category, focuses on motion detection accuracy, reduces brightness discontinuity caused by deficient fusion information and keeps low complexity. It makes the following contributions:

- Measure ghost areas by bilateral motion detection using color distance and visual salience constraint, which introduce information of local features and consistency.
- Refine exposure fusion method [3] by extending multiple LDR images with estimated camera response function (CRF) and introducing weight maps to reduce artifacts of brightness discontinuity and enhance details.

2. FRAMEWORK

The main idea of proposal is measuring image differences with a priority of visual salience which demonstrates how no-

This work was supported by the Natural Science Foundation of Jiangsu, China (Grant No. BK20130588) and KAKENHI (16K13006).

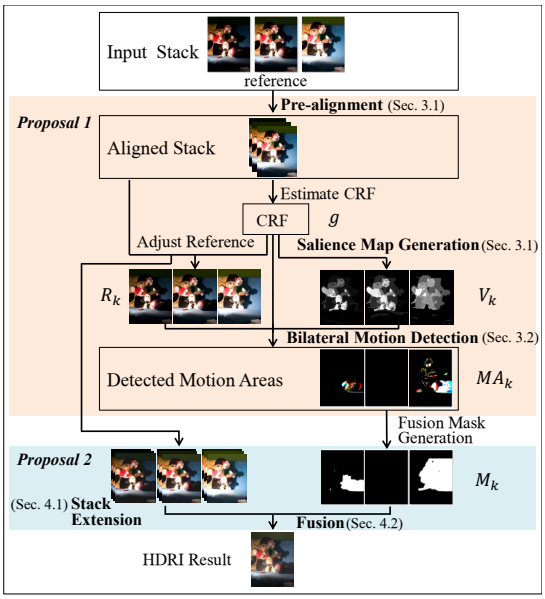

Fig. 1. Framework of the proposed algorithm

ticeable the foreground objects are. Thus the proposed algorithm computes the salience in bilateral motion detection and extends input stack to avoid fusion information deficiency.

Fig. 1 illustrates the framework of proposal which divides into two parts. For motion detection, the proposed algorithm makes a pre-alignment to remove camera motion and generate salience maps of multiple images; secondly, the method of Debevec [2] is employed to estimate CRF to map reference image to different exposures. The visual salience and color distance between two images are calculated in bilateral motion detection to detect moving areas, which are extended into fusion masks using graph cuts based segmentation [16]. For fusion, input stack are extended by estimated CRF; finally, exposure fusion method [3] is used to synthesize multi-exposure images with fusion masks.

3. VISUAL SALIENCE BASED BILATERAL MOTION DETECTION

3.1. Pre-alignment and Salience Map Generation

The proposed algorithm aims to use a set of N LDR images at different exposures and times $(L_1, \ldots, L_{ref}, \ldots, L_N)$ to reconstruct an HDR image R, and remove all moving objects in each input image except reference image to avoid ghost artifacts. For handheld cameras, there might be motion in full scene even static background. Therefore, the proposed algorithm use Evangelidis's method [17] based on enhanced correlation coefficient to make a fast pre-alignment. This area-based approach warps images by brightness constancy [18].

The salience maps then are generated using the method of Cheng[19] which measures the salience by region-based contrast. The method simultaneously evaluates global contrast differences and spatial weighted coherence scores. For

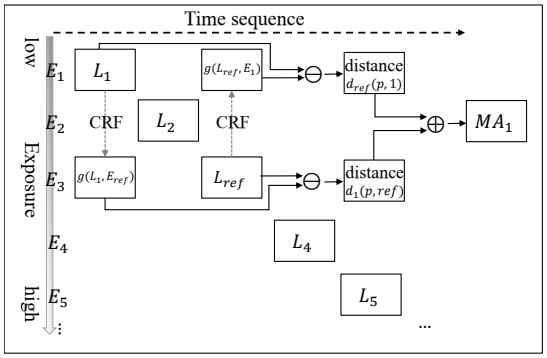

Fig. 2. Workflow of bilateral motion detection. The example shows the motion detection between L_1 and L_{ref}.

a region r_k, $w(r_k)$ denotes its weight, $\overline{r_k}$ denotes all other regions, D_r and D_s denote color and spatial distance, σ_s is strength threshold for spatial distance weighting (${\sigma_s}^2 = 0.4$ used in our proposal) and its visual salience $V(r_k)$ is defined as Eq. (1). The area is evaluated to be more salient when the pixel in salience maps is closer to 1.

$$V(r_k) = w_s(r_k) \sum_{r_k} e^{\frac{D_s(r_k, \overline{r_k})}{-{\sigma_s}^2}} w(\overline{r_k}) D_r(r_k, \overline{r_k}) \quad (1)$$

3.2. Bilateral Motion Detection

Fig. 2 shows details of bilateral motion detection between L_1 and L_{ref}, ref denotes the index of reference image. Before motion detection with other images, reference image needs to be adjusted to different exposures. Estimated CRF is shown as Eq. (2), the value of pixel p in L_k as $I(p,k)$, the exposure duration of L_k as $T(k)$, the film irradiance values for p as $E(p)$, the relationship among them as function g [2].

$$g(I(p,k)) = \ln E(p) + \ln T(k) \quad (2)$$

Since g is assumed to be smooth and monotonic, the reference can be adjusted to different exposures by $g(L_{ref}, E_k)$ as R_k. $D(p,k)$ is the bilateral distance of pixel p between $I(p,k)$ and $R(p,k)$ as shown in Eq. (3). A weighted distance $d(p,k)$ for pixel p is calculated between L_k and R_k to measure the difference with both color and salience information. Image areas are divided into three parts when calculating $d(p,k)$ by well-exposedness of reference because estimated CRF gets inaccurate near the under-exposed and over-exposed parts. Eq. (4) shows the definition of three parts.

$$D_k = d_{ref}(p,k) + d_k(p,k) \quad (3)$$

$$\begin{cases} d(p,k) = d_l(p,k), p \in A_l & I(p,ref) \leqslant 1 \\ d(p,k) = d_w(p,k), p \in A_w & 1 < I(p,ref) < 254 \\ d(p,k) = d_h(p,k), p \in A_h & I(p,ref) \geqslant 254 \end{cases} \quad (4)$$

Eq. (5) and Eq. (6) show the distance calculation for under-exposed and well-exposed parts, and for over-exposed area is reversal to Eq. (5) as Eq. (7). For areas with more salience,

the distance is increased larger with th_v to limit the salience strength. The area near over-exposed parts always has a big contrast to other area around even it's flat in spatiality, so we compute the salience maps with threshold V_h by the exposure of the image to limit the priority th_v as Eq. (8).

$$d_l(p,k) = \begin{cases} |I(p,k)|, & k < ref \\ \max\{I(p,k) - R(p,k), 0\} & \\ +th_v V(p,k), & k > ref \end{cases} \quad (5)$$

$$d_w(p,k) = |R(p,k) - I(p,k)| + th_v V(p,k) \quad (6)$$

$$d_h(p,k) = \begin{cases} \max\{R(p,k) - I(p,k), 0\} & \\ +th_v V(p,k), & k < ref \\ |255 - I(p,k)|, & k > ref \end{cases} \quad (7)$$

$$V(p,k) = \begin{cases} V(r_k), & I(p,k) < V_h, p \in r_k \\ 0, & I(p,k) \geqslant V_h, p \in r_k \end{cases} \quad (8)$$

The proposed algorithm assumes that moving area is continuous, and the distance $D(p,k)$ between L_{ref} and L_k in well-exposed area is a Gaussian distribution. Then the problem can be solved with an MRF model as energy minimization. Define l_p and l_q as the label of pixel p and q with moving objects $'obj'$ and background $'bkg'$, E_D as the data cost, E_S as the smooth cost, and Eq. (9) - Eq. (13) show the definition of energy function. $R(l_p)$ is the cost flag of l_p that equals 0 when distance does not exceed threshold cost. The threshold cost equals the standard deviation of d_l, d_w and d_h as σ_l, σ_w and σ_h times a threshold β. α can be treated as camera noise for calculating the threshold cost in poorly-exposed areas.

$$E = \sum_{p \in P} E_D + \gamma \sum_{(p,q) \in N} E_S(l_p, l_q) \quad (9)$$

$$E_S(l_p, l_q) = \begin{cases} 0, & l_p = l_q \\ 1, & l_p \neq l_q \end{cases} \quad (10)$$

$$E_D = \begin{cases} R(l_p='obj') \times \gamma + R(l_p='bkg') \times |D(p,k) - \beta\sigma|, \\ k > ref, p \in A_l \bigcup p \in A_w \bigcup k < ref, p \in A_h \\ R(l_p='obj') \times \gamma + R(l_p='bkg') \times (\gamma + |D(p,k) - \alpha|), \\ k < ref, p \in A_l \bigcup k > ref, p \in A_h \end{cases} \quad (11)$$

$$R(l_p='obj') = \begin{cases} 0, D(p,k) > threshold \\ 1, D(p,k) \leqslant threshold \end{cases}$$

$$R(l_p='bkg') = \begin{cases} 0, D(p,k) \leqslant threshold \\ 1, D(p,k) > threshold \end{cases} \quad (12)$$

$$threshold = \begin{cases} \beta\sigma_l, & k > ref, p \in A_l \\ \beta\sigma_w, & p \in A_w \\ \beta\sigma_h, & k < ref, p \in A_h \\ \alpha, & k < ref, p \in A_l \bigcup k > ref, p \in A_h \end{cases} \quad (13)$$

This energy minimization can be solved via graph cuts, using the method of Boykov [20]. The pixel value of $'obj'$ in detected moving areas MA_k is 1 as shown in Fig. 3.

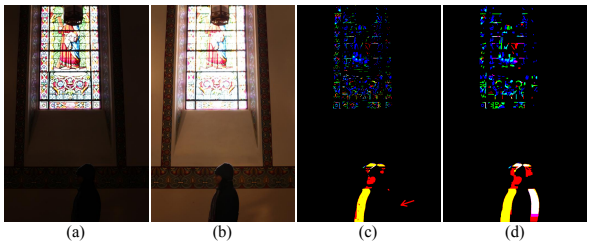

Fig. 3. Result of proposal 1: (a) reference (4th) image of dataset StainedGlass1 [8]. (b) the 5th image. (c) detected motion areas by Zhang's method [15]. (d) detected motion areas by proposed method.

4. STACK EXTENSION BASED EXPOSURE FUSION

4.1. LDR Stack Extension

Motion detection based method can obtain satisfying results when detected motion areas are in different regions of input sources. However sometimes large motion occurs, detected motion areas might be in the same region and very limited information of input sources can be used for fusion except the reference. This may cause brightness discontinuity and detail loss as shown in Fig. 6 (d)(i). In order to reduce this kind of artifacts, the proposed algorithm extends the input LDR stack before exposure fusion. Since the CRF has already been estimated, the input LDR images can be mapped to different exposures using the function g. L' denotes the images set after exposure extension, the proposed algorithm maps each input image L_k to adjacent exposure E_{k-1} and E_{k+1}, the extended result $L'_{k,k-1}$ and $L'_{k,k+1}$ can be expressed as Eq. (14).

$$L' \ni \begin{cases} L'_{k,k-1} = g(L_k, E_{k-1}), & 1 < k \leqslant N \\ L_k, & 1 \leqslant k \leqslant N \\ L'_{k,k+1} = g(L_k, E_{k+1}), & 1 \leqslant k < N \end{cases} \quad (14)$$

For a set of N LDR images, the proposed algorithm generates an extended stack of $3N$-2 images for the following exposure fusion, and the extended images $L'_{k,k-1}$ and $L'_{k,k+1}$ have the same fusion mask M_k as L_k.

4.2. Fusion

The proposed algorithm computes fusion masks and salience to modify the weight function of Mertens [3]. $\Lambda(p,k)$ denotes the final weight of pixel p in L'_k, $\{C,S,E\}$ and $\{W_C, W_S, W_E\}$ denote contrast, saturation, well-exposedness and their corresponding weight; the moving area label in fusion mask is $M(p,k)$. Eq. (15) is original weight function and Eq. (16) is the proposed with a constraint of fusion masks. Fig. 6 (d)(i)(e)(j) show the comparison between fusion results.

$$\Lambda(p,k) = (C(p,k))^{W_C} \times (S(p,k))^{W_S} \times (E(p,k))^{W_E} \quad (15)$$

$$\Lambda'(p,k) = \Lambda(p,k) \times (1 - M(p,k)) \quad (16)$$

5. EXPERIMENT RESULTS

The test platform is Intel(R) Core(TM) i7-3770 CPU @ 3.40GHz with 8.0 GB DDR3 RAM. 8 datasets are selected from Gallo's [11] and Sen's work [8] for evaluation. The proposed algorithm is tested with configuration parameters: $V_h = 230$, $th_v = 10.0$, $\alpha = 5.0$, $\beta = 1.9$.

Figs. 4,5 are static camera datasets. Lee [14] and Zhang [15] remove most ghost artifacts but still some artifacts and over-exposed areas remain due to incorrect motion detection. Sen introduces new artifacts in the poorly-exposed area of reference due to mismatch. Figs. 6,7 are handheld camera dataset. Zhang cannot adapt to scene motion; hole artifacts occurs and poorly-exposed areas remain in Lee; Sen introduces artifacts in corresondence based optimization. In Fig. 7, the proposed algorithm removes most ghosts except the area around moving car because it is over-exposed in reference and motion information cannot be reconstructed from other images. Table 1 shows the detailed timecost. For static camera datasets, the proposed algorithm has a far less time cost, just 52.9% of Sen and 28% of Lee; for handheld camera datasets, although time cost changes noticeably depending on the complexity of pre-alignment such as the dataset "pianoman", the result is still acceptable. On average, the proposed method achieves 20.4% and 63.6% time savings against Sen and Lee respectively. Additional results and source code are provided in our project website [21].

6. CONCLUSIONS

Ghost removal is a major issue in HDRI. This paper proposes a robust method based on visual salience and stack extension. To improve the accuracy of ghost areas detection, visual salience based bilateral motion detection is introduced for measuring image differences. Furthermore, input stack is extended before exposure fusion to reduce brightness discontinuity and enhance details. Experiment results on static and handheld camera datasets show that our method provides significant gains in deghosting quality and robustness over state-of-the-art approaches, with low complexity about 20.4% and 63.6% time savings on average over recent advances [8][14].

Table 1. Timecost of different algorithms (seconds)

Dataset	Lee[14]	Sen[8]	Proposed
Static Camera			
Arch	374.3	165.4	91.59
Puppets	540.1	172.5	113.71
SculptureGarden	253.9	233.0	105.1
Forest	223	145.9	79.74
Average	347.8	184.4	**97.5**
Handheld Camera			
StainedGlass1	284.7	297.1	236.61
StainedGlass2	379.5	197.8	136.33
Pianoman	1033.1	210.3	338.76
GoingOutTheDoor1	791.7	334.9	312.64
Average	622	260	**256.1**

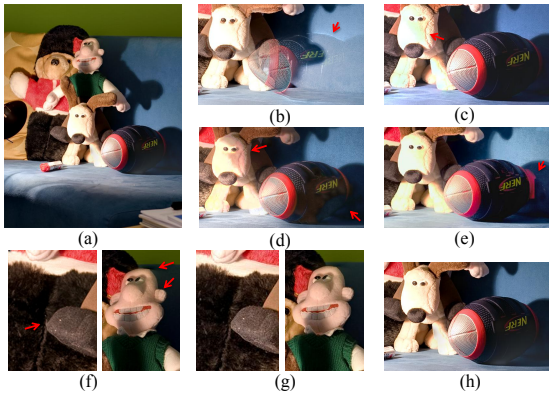

Fig. 4. Comparison of dataset Puppets: (a)(g)(h) Proposed. (b) Mertens. (c) Gallo. (d) Sen. (e) Lee. (f) Zhang.

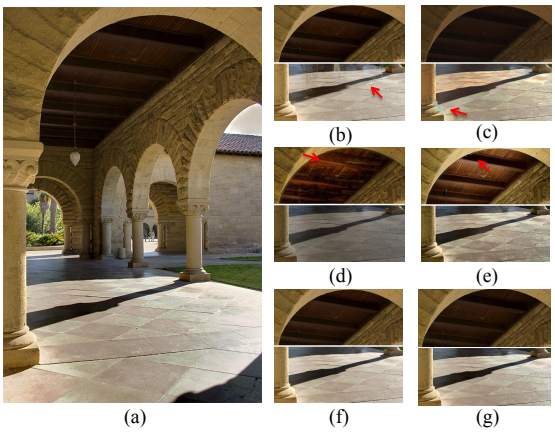

Fig. 5. Comparison of dataset Arch: (a)(g) Proposed. (b) Mertens. (c) Gallo. (d) Sen. (e) Lee. (f) Zhang.

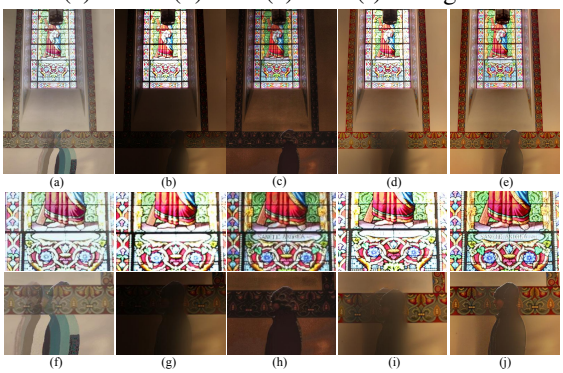

Fig. 6. Comparison of dataset StainedGlass1: (a)(f) Lee. (b)(g) Zhang. (c)(h) Sen. (d)(i) Fusion result based on Mertens [3]. (e)(j) Proposed.

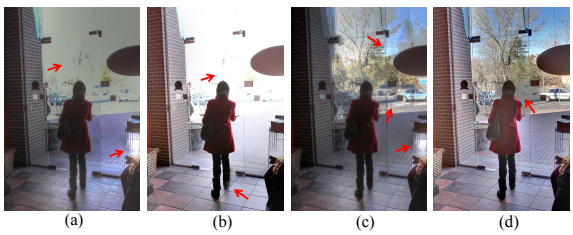

Fig. 7. Comparison of dataset GoingOutTheDoor1: (a) Lee. (b) Zhang. (c) Sen [8]. (d) Proposed.

7. REFERENCES

[1] G. Kontogianni, E.K. Stathopoulou, A. Georgopoulos, and A. Doulamis, "HDR imaging for feature detection on detailed architectural scenes," *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 40, no. 5, pp. 325, 2015.

[2] P.E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *ACM SIGGRAPH 2008 classes*. ACM, 2008, p. 31.

[3] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion: A simple and practical alternative to high dynamic range photography," *Computer Graphics Forum*, vol. 28, no. 1, pp. 161–171, 2009.

[4] L. Bogoni, "Extending dynamic range of monochrome and color images through fusion," in *Proc. International Conference on Pattern Recognition*. IEEE, 2000, vol. 3, pp. 7–12.

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

[6] T. Jinno and M. Okuda, "Motion blur free HDR image acquisition using multiple exposures," in *Proc. International Conference on Image Processing*. IEEE, 2008, pp. 1304–1307.

[7] H. Zimmer, A. Bruhn, and J. Weickert, "Freehand HDR imaging of moving scenes with simultaneous resolution enhancement," *Computer Graphics Forum*, vol. 30, no. 2, pp. 405–414, 2011.

[8] 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. on Graphics*, vol. 31, no. 6, pp. 203, 2012.

[9] C. Barnes, E. Shechtman, A. Finkelstein, and D. Goldman, "Patchmatch: A randomized correspondence algorithm for structural image editing," *ACM Trans. on Graphics*, vol. 28, no. 3, pp. 24, 2009.

[10] 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.

[11] O. Gallo, N. Gelfandz, W.C. Chen, M. Tico, and K. Pulli, "Artifact-free high dynamic range imaging," in *Proc. International Conference on Computational Photography*. IEEE, 2009, pp. 1–7.

[12] Y. Heo, K. Lee, S. Lee, Y. Moon, and J. Cha, "Ghost-free high dynamic range imaging," *Computer Vision–ACCV 2010*, pp. 486–500, 2011.

[13] X. Zhou, C. Yang, and W. Yu, "Moving object detection by detecting contiguous outliers in the low-rank representation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 597–610, 2013.

[14] C. Lee, Y. Li, and V. Monga, "Ghost-free high dynamic range imaging via rank minimization," *IEEE Signal Processing Letters*, vol. 21, no. 9, pp. 1045–1049, 2014.

[15] B. Zhang, Q. Liu, and T. Ikenaga, "Ghost-free high dynamic range imaging via moving objects detection and extension," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, 2015, pp. 459–462.

[16] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient N-D image segmentation," *International Journal of Computer Vision*, vol. 70, no. 2, pp. 109–131, 2006.

[17] G.D. Evangelidis and E.Z. Psarakis, "Parametric image alignment using enhanced correlation coefficient maximization," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1858–1865, 2008.

[18] R. Szeliski, *Computer vision: algorithms and applications*, Springer Science & Business Media, 2010.

[19] M.M. Cheng, N.J. Mitra, X. Huang, P.H. Torr, and S.M. Hu, "Global contrast based salient region detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569–582, 2015.

[20] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, 2001.

[21] [Online] Available: *https://github.com/U2FsdGVkX41/hdr-fusion*.