A Survey of Multi-Focus Image Fusion with Guided Filtering Based on SIFT

Firdous Qureshi[1], Sriram Yadav[2], Avinash Sharma[3]

[1]M.Tech(CSE, Student), [2]Associate Professor (H.O.D), [3]Associate Professor
Department of CSE, MITS, Bhopal

Abstract - Current image coding with image fusion schemes make it hard to utilize external images for transform even if multi-focus images can be found in the cloud. To solve this problem, we propose a method of multi-focus image fusion scheme that is different from current image fusion scheme even on the ground. For this purpose, we use GFF-SIFT method which is combination of GFF (Guided Filtering based Fusion) and SIFT (Scale Invariant Feature Transform) method. A fast and efficient image fusion technique is proposed for creating a highly generated fused image through merging multiple corresponding multi-focus images. The proposed technique is based on a two-scale decomposition of an image into a low layer containing large scale variations, and a detail layer acquiring small scale details. A novel approach of GFF-SIFT method is proposed to make full use of spatial consistency for merge of the base and detail layers. Analytical results represent that the proposed technique can obtain state-of-the-art performance for image fusion of multispectral, multimodal, and multi exposure images.

Keywords: Multi-Focus image fusion, Guided Filtering, Scale-Invariant Feature Transform.

I. INTRODUCTION

Image fusion is an important technique for various image processing and computer vision applications such as feature extraction and target recognition. Through image fusion, different images of the same scene can be combined into a single fused image [1]. The fused image can provide more comprehensive information about the scene which is more useful for human and machine perception. For instance, the performance of feature extraction algorithms can be improved by fusing multi-focus remote sensing images [2]. The fusion of multi-exposure images can be used for digital photography [3]. In these applications, a good image fusion method has the following properties. First, it can preserve most of the useful information of different images. Second, it does not produce artifacts. Third, it is robust to imperfect conditions such as mis-registration and noise.

Image search has been demonstrated as a successful application on the Internet [6]. By submitting the description of one image, including semantic content [7], outline [8],[9], and local feature descriptors [12], one can easily retrieve many similar images. Near and partial duplicate image detection is a hot research topic in this field [12], [13]. However, the purpose of image search is not to generate an image from search results. In fact, reconstructing a given image from similar images is tougher than the image search itself.

II. SYSTEM MODEL

The basic system model of proposed method is strictly based on SIFT (Scale Invariant Feature Transform) and GFF (Guided Filtering Fusion) technique. SIFT Feature Descriptor detector has four main stages namely as Scale-space extrema detection, Key point localization, Orientation computation, Key point descriptor extraction. In SIFT, we Have to perform following process. Assume that, Take a 16 x16 window area that around interest node. Separate into a 4x4 grid of rectangular cells. Compute histogram of image gradient directions in each rectangular cell (8 bins each).

16 histograms x 8 orientations = 128 feature descriptors

Guided filtering fusion is a technique of obtain fused image through weight map construction. Basically, the basic steps of this technique is as follows

- Two Scale image decomposition
- Weight map construction
- Two Scale Image reconstruction

In first step, we have to consider two source multi-focus images, now find out base and detail layer using average filtering approach, these layers are further uses for weighted average of base and detail layer. In second step, Construct Weight map of an image on the basis of saliency map on source multi-focus image, then apply guided filter method, in this method source images are considered as guidance images and weight map images taken as input images. In third step, calculate weighted average of output of first step base-detail layer of source image and output of second step base-detail layer of weight map through guided filtering method.

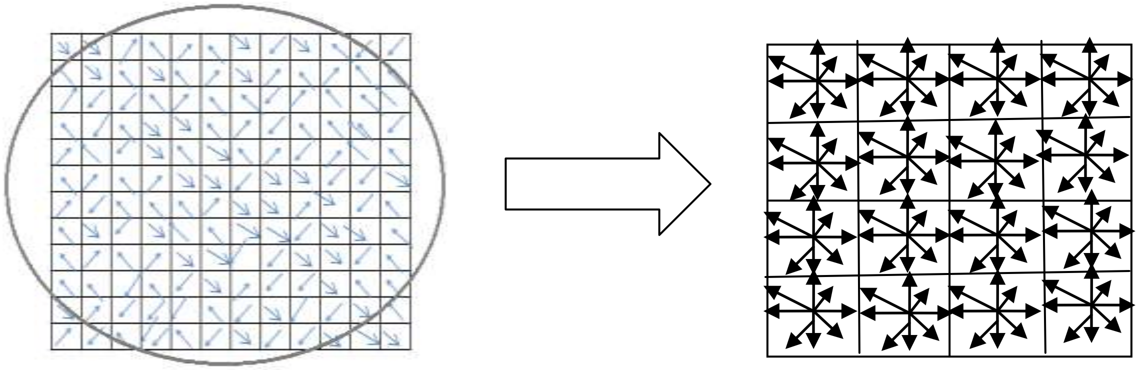

Fig 1: Process for finding key point descriptor from an image using SIFT procedure

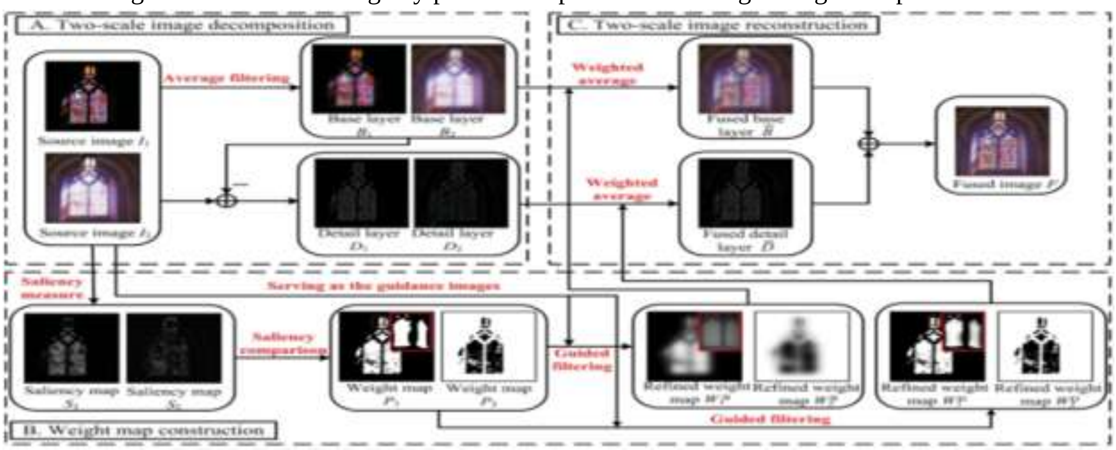

Fig 2: Process of obtain fused image using guided filtering fusion (GFF)

III. Previous Work

A large number of image fusion methods [4]–[7] have been proposed in literature. Among these methods, multi-scale image fusion [5] and data-driven image fusion [6] are very successful methods. They focus on different data representations, e.g., multi-scale co-efficient [8], [9], or data driven decomposition co-efficient [6], [10] and different image fusion rules to guide the fusion of co-efficient. The major advantage of these methods is that they can well preserve the details of different source images. However, these kinds of methods may produce brightness and color distortions since spatial consistency is not well considered in the fusion process. To make full use of spatial context, optimization based image fusion approaches, e.g., generalized random walks [3], and Markov random fields [9] based methods have been proposed. These methods focus on estimating spatially smooth and edge-aligned weights by solving an energy function and then fusing the source images by weighted average of pixel values. However, optimization based methods have a common limitation, i.e., inefficiency, since they require multiple iterations to find the global optimal solution. Moreover, another drawback is that global optimization based methods may over-smooth the resulting weights, which is not good for fusion.

SIFT descriptors, proposed by Lowe in [12], present distinctive invariant features of images that consist of location, scale, orientation, and feature vector. The scale and location of SIFT descriptors are determined by maxima and minima of difference-of-Gaussian images. One orientation is assigned to each SIFT descriptor according to the dominant direction of the local gradient histogram. The feature vector is a 128-dimension vector that characterizes a local region by gradient histogram in different directions. Since SIFT descriptors have a good interpretation of the response properties of complex neurons in the visual cortex [15] and an excellent practical performance, they have been extensively applied to object recognition, image retrieval, 3D reconstruction, annotation, watermarking, and so on.

IV. Proposed Methodology

The proposed method of this paper is GFF-SIFT (Guided Filtering based Fusion with Scale Invariant Feature

Transform). Recently, edge-preserving filters [6], [8] have been an active research topic in image processing. Edge-preserving smoothing filters such as guided filter [4], weighted least squares [5], and bilateral filter [7] can avoid ringing artifacts since they will not blur strong edges in the decomposition process. Among them, the guided filter is a recently proposed edge-preserving filter, and the computing time of which is independent of the filter size. Furthermore, the guided filter is based on a local linear model, making it qualified for other applications such as image matting, up-sampling and colorization [8]. In this paper, the SIFT method applied on image for finding the keypoint descriptor. Images of one scene may be taken from different viewpoints or may suffer transformations such as rotation, noise etc [12]-[14]. So it is likely that two images of the same scene will be different. The task of finding similarity correspondences between two images of the same scene or object has thus become a challenging problem in a number of vision applications. Such applications range from image registration, camera calibration, object recognition, scene localization in navigation systems, image retrieval based search engines etc. For image matching, extraction of such information (i.e. features) is required from the images which can provide reliable matching between different viewpoints of the same image. Feature detection occurs within an image and seeks to describe only those parts of that image where we can get unique information or signatures (i.e. feature descriptors). During training, feature descriptors are extracted from sample images and stored. In classification, feature descriptors of a query image are then matched with all trained image features and the trained image giving maximum correspondence is considered the best match. Feature descriptor matching can be based on distances such as Euclidean, Mahalanobis or distance ratios. For detect feature descriptors, we use SIFT method [11]. SIFT Feature Descriptor detector has four main stages namely as

- Scale- space extrema detection
- Key point localization
- Orientation computation
- Key point descriptor extraction.

Now apply guided filter for image fusion. In theory, the guided filter assumes that the filtering output O is a linear transformation of the guidance image I in a local window ωk centered at pixel k.

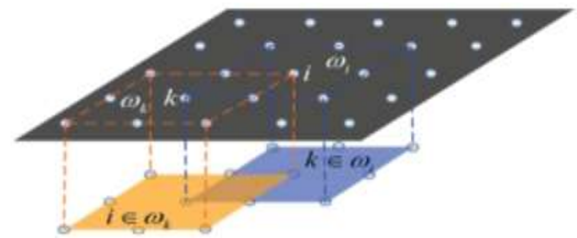

Fig 3: illustration of window choice

Oi = akIi +bk ∀ i ∈ ωk (1)

Where, ωk is a tile window of size (2r+1)×(2r+1). The linear value coefficients ak and bk are constant value in ωk and can be calculated by reducing the squared difference between the output picture O and the input picture P.

$$E(a_k, b_k) = \sum_{i\in\omega_k}((\mathrm{a}_k \mathrm{I}_i + \mathrm{b}_k - \mathrm{P}_i)^2 + ea^2{}_k) \tag{2}$$

Where, e is a regularization argument supply by the user. The coefficients value ak and bk can be directly solved by linear regression value [15] as follows:

$$a_k = \frac{\frac{1}{|\omega|}\sum_{i\in\omega_k} I_i P_i - \mu_k \overline{P_k}}{\delta_k + e} \tag{3}$$

$$b_k = \overline{P_k} - a_k \mu_k \tag{4}$$

where μk and δk are the mean and variance value of I in ωk respectively, |ω| is the no. of pixels in ωk, and Pk is the mean of P in ωk. Next, the result image can be calculated as per equation to (1). As shown in Fig. 1, all local value windows centered at pixel k in the window ωi that will contain pixel i. So, the cost of Oi in (1) will change when it is computed in different windows ωk. To solve this question, all the possible costs of coefficients ak and bk are first averaged. Then, the filtering output is estimated as follows:

$$O_i = \overline{a}_i I_i + \overline{b}_i \tag{5}$$

Where $\overline{a}_i = \frac{1}{|\omega|}\sum_{k\in\omega_i} a_k, \overline{b}_i = \frac{1}{|\omega|}\sum_{k\in\omega_i} b_k.$ In this chapter $G_{r.e}(\mathrm{P},\mathrm{I})$ is used to represent the guided filtering operation, where r and e are the parameters which decide the filter size and blur degree of the guided filter, respectively. Moreover, P and I refer to the input image and guidance image, respectively. Furthermore, when the input is a color image, the filtering output can be obtained by conducting the guided filtering on the red, green, and blue channels of the input image, respectively. And when the guidance image I is a color image, the guided filter should be extended by the following steps. First, equation (1) is rewritten as follows:

$$O_i = a_k^T I_i + b_k, \forall i \in \omega_k \tag{6}$$

Where, ak is a 3×1 coefficient vector and Ii is a 3×1 color vector. Then, similar to (3)–(5), the output of guided filtering can be calculated as follows:

$$a_k = (\sum_k + \mathrm{eU})(\frac{1}{|\omega|})\sum_{i\in\omega_k} I_i p_i - \mu_k \overline{P_k} \tag{7}$$

$$b_k = \overline{p_k} - a_k^T \mu_k \tag{8}$$

$$O_i = \overline{a_i}^T I_i + \overline{b_i} \tag{9}$$

where Σk is the 3×3 covariance matrix of I in ωk, and U is the 3×3 identity matrix.

In following figure the main processes of the proposed guided filtering based fusion method (GFF). First, an average filter is utilized to get the two-scale representations. Then, the base and detail layers are fused through using a guided filtering based weighted average method.

A. Two-Scale Image Decomposition

As shown in following figure, the source images are first decomposed into two-scale representations by average filtering. The lower layer of each given image is obtained in following way:

$$\mathrm{Bn} = \mathrm{In} * \mathrm{Z} \tag{10}$$

Where, In is the nth source image, Z is the average value of filter, and the size of the filter is conventionally fixed. Once the lower layer is found, the detail layer can be found by subtracting the lower layer from the source image.

$$\mathrm{Dn} = \mathrm{In} - \mathrm{Bn}. \tag{11}$$

The two-scale decomposition step concerns at separating each source image into a lower layer consisting the large-scale variations in intensity level and an another layer containing the small - scale details.

B. Weight Map Construction with Guided Filtering

The weight map is constructed as follows as per above figure. First, Laplacian filtering is applied to each given image to obtain the high-pass image Hn.

$$\mathrm{Hn} = \mathrm{In} * \mathrm{L} \tag{12}$$

Where, L is a 3×3 Laplacian filter matrix. Then, the local average of the value of Hn is used to design the saliency maps Sn.

Sn =|Hn|∗ grg,σg (13)

g is a Gaussian low-pass filter, and the parameters rg and σg are set to 5. The measured saliency maps support good characterization level of the saliency of detail information. Next, the saliency values are compared to determine the weight maps as follows:

$$P_n^k = \{^1_0 \; 1 \text{ if } S_n^k = \max(S_1^k, S_2^k, ..., S_N^k) \text{ , 0 otherwise}$$
(14)

Where N is number of given images, S_n^k is the saliency value of the pixel k in the nth image. However, the weight value obtain above are usually noisy and not aligned with object boundaries, which may produce artifacts to the fused image. Using spatial consistency is an effective way to solve this problem. Spatial consistency means that if two adjacent pixels have similar brightness or color, they will tend to have similar weights. A popular spatial consistency based fusion approach is formulating an energy function, where the pixels saliencies are encoded in the function and edge aligned weights are enforced by regularization terms, e.g., a smoothness term. This energy function can be then minimized globally to obtain the desired weight maps. However, the optimization based methods are often relatively inefficient. An interesting alternative to optimization based methods is proposed. Guided image filtering is performed on each weight map Pn with the corresponding source image In serving as the guidance image.

$$W_n^B = G_{r1,e1}(P_{n,}I_n)$$
$$W_n^D = G_{r2,e2}(P_{n,}I_n) \tag{15, 16}$$

Where, r1, e1, r2, and r2 are the parameters of the guided filter, W_n^B and W_n^D are the resulting weight maps of the base and detail layers. Finally, the values of the N weight maps are normalized such that they sum to one at each pixel k.

The motivation of the proposed weight construction method is as follows. According to (1), (3) and (4), it can be seen that if the local variance at a position i is very small which means that the pixel is in a flat area of the guidance image, then ak will become close to 0 and the filtering output O will equal to $\overline{P_k}$ i.e., the average of adjacent input pixels. In contrast, if the local variance of pixel i is very large which means that the pixel i is in an edge area, then ak will become far from zero. As demonstrated in [12], $\nabla O \approx \bar{a} \nabla I$ will become true, which means that only the weights in one side of the edge will be averaged. In both situations, those pixels with similar color or brightness tend to have similar weights. This is exactly the principle of spatial consistency.

Furthermore, as shown in above figure, the base layers look spatially smooth and thus the corresponding weights also should be spatially smooth. Otherwise, artificial edges may be produced. In contrast, sharp and edge-aligned weights are preferred for fusing the detail layers since details may be lost when the weights are over-smoothed. Therefore, a large filter size and a large blur degree are preferred for fusing the base layers, while a small filter size and a small blur degree are preferred for the detail layers.

C. Two-Scale Image Reconstruction

Two-scale image reconstruction consists of the following two steps. First, the base and detail layers of different source images are fused together by weighted averaging

$$\overline{B} = \sum_{n=1}^{N} W_n^B B_n \tag{17}$$

$$\overline{D} = \sum_{n=1}^{N} W_n^D D_n \tag{18}$$

Then, the fused image F is obtained by combining the fused base layer $\overline{B}$ and the fused detail layer $\overline{D}$

$$F = \overline{B} + \overline{D} \tag{19}$$

GFF (Guided Filtering based Fusion) technique is based on generic SIFT method, which is used for finding the keypoint descriptor. Guided Filtering is used for find fused image from image descriptor. This method increases the accuracy of pixel ratio of fused image. Hence, we are explains the proposed GFF-SIFT method. The potential of the GFF-SIFT is illustrated on a 3D object recognition task using the Coil database. The images are either represents by a matrix of their pixel values (bitmap representation) or by a color histogram. In both cases, the proposed system requires feature extraction and performs recognition on images regarded as points of a space of high dimension. The feature extraction is perform by SIFT scheme. We also purpose an extension of the basic color histogram which keeps more about the information contained in the images.

The Algorithm of proposed method is explained below:

[fusimg] = GFF_SIFT(image1, image2)

Step 1: Consider the two multi-focus source images with same resolution.

Step 2: Now we find out the SIFT descriptors of each source image of cell array for images of image dataset. SIFT method perform the following sequence of steps for find the keypoint descriptors for texture feature.

Scale-Space Extreme Detection: The initial step of evaluation finds total all scale-space and different image area in image dataset nodes [4]. It is completely apply effectively by using a Difference-of-Gaussian (DoG) mapping to represents potential interest keypoints of feature descriptors which are scale invariant and orientation in image dataset nodes [6].

Keypoints Localization: All candidate area of image in selected ROI (Region of Interest), a detailed prototype is fit to analyze keypoints area and its scale-space [5]. Keypoints of image area in image ROI are chooses basis on calculate of existing stability [6].

Orientation Assignment: One or more orientations task are applied to each keypoints area based on local image data nodes gradient directions [2]. Each and every future image operations are implemented on image keypoint dataset which has been transformed relative to the applied orientation, scale, and location for each feature descriptor, hence providing invariance to these transformations in image data nodes [6].

Keypoints Descriptor: The local image gradients value are measured at the choose scale-space in the Region of Interest (ROI) around all keypoints in image dataset points [4]. These are transformed into a presentation that permits for significant levels of local shape, location and orientation and changes in illumination of image dataset points [6].

Step 3: Above step are perform in repeated form, then all the descriptor of images are store, Now apply Guided Filtering method for obtaining fused images.

Step 4: In Guided Filtering, apply two-scale image decomposition using average filter for find base and detail layer of each source image from image database.

Step 5: Now apply weight map construction, obtain saliency map from source images, now compare saliency map and feature descriptor corresponding images, then find out weight map of base and detail layer of each source image from image database.

Step 6: Now image reconstruction on the base layer and detail layer of each source image. Hence, the reconstruct image is taken as fused image.

V. IMPLEMENTATION

Experiments are performed on three image databases, i.e., the Petrovic´ database [6] which contains 50 pairs of images including aerial images, outdoor images (natural, industrial) and indoor images (with different focus points and exposure settings), the multi-focus image database which contains 10 pairs of multi-focus images, and the multi-exposure and multi-modal image database which contains 2 pairs of color multi-exposure images and 8 pairs of multi-modal images. The testing images have been used in many related papers [3]–[6], [2]–[6]. Firstly, We have to find keypoint descriptor using SIFT method. The proposed guided filtering based fusion method (GFF) is compared with seven image fusion algorithms based on Laplacian pyramid (LAP) [8], stationary wavelet transform (SWT) [9], curvelet transform (CVT), non-subsampled contour let transform (NSCT), generalized random walks (GRW) [3], wavelet-based statistical sharpness measure (WSSM) and high order singular value decomposition (HOSVD) , respectively. The parameter settings of these methods are as follows. Four decomposition levels, the "averaging" scheme for the low-pass sub-band, the absolute maximum choosing scheme for the band-pass sub-band and the 3×3 window based consistency check are adopted for the LAP, CVT, SWT, and NSCT method. Four decomposition levels with 4, 8, 8, 16 directions from coarser scale to finer scale are adopted for the NSCT method. Furthermore, the default parameters given by the respective authors are adopted for the GRW, WSSM and HOSVD based methods.

VI. CONCLUSIONS

We have presented a novel approach is GFF-SIFT technique. The proposed technique find the keypoint descriptor using SIFT method then utilizes the filter to acquire the two-scale presentations, which is simple and effective. More importantly, the guided filter method is used in a desired way to make full use of the correlations between neighborhood pixels for weight optimization trends. Experiments explained that the proposed technique can well keep the original and complementary information of all input images. Encouragingly, the proposed technique is very robust to image registration. Furthermore, the proposed approach is computationally effective, making it quite qualify for desired applications. At last, how to get improve the effectiveness of the concern proposed method by adaptively choosing the parameters of the guided filter can be further researched.

VII. FUTURE WORK

The future work is as follows:

1. We can use SURF, CHOG, Fast SIFT or Dense SIFT method for find the keypoint descriptors of an image.
2. We can use other image fusion method which incooperate with above feature descriptor method.

REFERENCES

[1] A. A. Goshtasby and Nikolov, "Image fusion: Advances in the state of the art," Inf. Fusion, vol. 8, no. 2, Apr. 2007.
[2] Socolinsky and L. Wolff, "Multispectral image visualization through first-order fusion," IEEE Trans. Image Process., vol. 11, no. 8, Aug. 2002.
[3] Shen, Cheng, and A. Basu, "Generalized random walks for fusion of multi-exposure images," IEEE Trans. Image Process., Dec. 2011.
[4] Li, Kwok, I. Tsang, and Wang, "Fusing images with different focuses using support vector machines," IEEE Trans. Neural Network, Nov. 2004.
[5] Pajares and J. M. de la Cruz, "A wavelet-based image fusion tutorial," Pattern Recognition, Sep. 2004.
[6] Looney and Mandic, "Multiscale image fusion using complex extensions of EMD," IEEE Trans. Signal Process., vol. 57, no. 4, pp. 1626–1630, Apr. 2009.
[7] M. Kumar and S. Dass, "A total variation-based algorithm for pixel- level image fusion," IEEE Trans. Image Process., vol. 18, no. 9, pp. 2137–2143, Sep. 2009.
[8] F. C. Crow, "Summed-area tables for texture mapping," in Proc. SIG- GRAPH 84. 11th Annu. Conf. Comput. Graph. Interact. Tech., vol. 18, no. 3, pp. 207–212, Jan. 1984
[9] O. Rockinger, "Image sequence fusion using a shift-invariant wavelet transform," in Proc. Int. Conf. Image Process., vol. 3, Washington, DC, USA, Oct. 1997, pp. 288–291.
[10] J. Liang, Y. He, D. Liu, and X. Zeng, "Image fusion using higher order singular value decomposition," IEEE Trans. Image Process., vol. 21, no. 5, pp. 2898–2909, May 2012.
[11] M. Xu, H. Chen, and P. Varshney, "An image fusion approach based on markov random fields," IEEE Trans. Geosci. Remote Sens., vol. 49, no. 12, pp. 5116–5127, Dec. 2011.
[12] D. G. Lowe, "Distinctive image features from scale-invariant key- points," Int. J. Comput. Vis., vol. 60, pp. 91–110, 2004.
[13] Q. Zhang, Y. Chen, Y. Zhang, and Y. Xu, "SIFT implementation and optimization for multi-core systems," in Proc. IEEE Int. Parallel Distrib. Process., Apr. 2008, pp. 1–8.

[14] H. Feng, E. Li, Y. Chen, and Y. Zhang, “Parallelization and characteriza- tion of SIFT on multi-core systems,” in Proc. IEEE Int. Symp. Workload Characterizat., Sep. 2008, pp. 14–23.
[15] S. N. Sinha, J.-M. Frahm, M. Pollefeys, and Y. Genc, “GPU-based video feature tracking and matching,” in Proc. Workshop Edge Comput. Using New Commod. Archit., 2006, pp. 1–15.
[16] S. Heymann, K. Müller, A. Smolic, B. Froehlich, and T. Wiegand, “SIFT implementation and optimization for general-purpose GPU,” in Proc. Int. Conf. Central Eur. Comput. Graph., Visualizat. Comput. Vis., 2007, pp. 144–159.
[17] N. Cornelis and L. Van Gool, “Fast scale invariant feature detection and matching on programmable graphics hardware,” in Proc. IEEE Com- put. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops, Jun. 2008, pp. 1–8.
[18] T. Terriberry, L. French, and J. Helmsen, “GPU accelerating speeded-up robust features,” in Proc. 4th Int. Symp. 3D Data Process., Visualizat. Transmiss., 2008, pp. 1–8.