

Research Article

Maximum Entropy Threshold Segmentation for Target Matching Using Speeded-Up Robust Features

Mu Zhou,¹ Xia Hong,^{1,2} Zengshan Tian,¹ Huining Dong,² Mingchun Wang,¹ and Kunjie Xu³

¹Chongqing Key Laboratory of Mobile Communications Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

²Chongqing Laboratory of Material Physics and Information Display, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

³Graduate Telecommunications and Networking Program, University of Pittsburgh, Pittsburgh, PA 15260, USA

Correspondence should be addressed to Mu Zhou; zhoumu@cqupt.edu.cn

Received 21 April 2014; Revised 2 August 2014; Accepted 11 August 2014; Published 26 August 2014

Academic Editor: Adam Panagos

Copyright © 2014 Mu Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes a 2-dimensional (2D) maximum entropy threshold segmentation (2DMETS) based speeded-up robust features (SURF) approach for image target matching. First of all, based on the gray level of each pixel and the average gray level of its neighboring pixels, we construct a 2D gray histogram. Second, by the target and background segmentation, we localize the feature points at the interest points which have the local extremum of box filter responses. Third, from the 2D Haar wavelet responses, we generate the 64-dimensional (64D) feature point descriptor vectors. Finally, we perform the target matching according to the comparisons of the 64D feature point descriptor vectors. Experimental results show that our proposed approach can effectively enhance the target matching performance, as well as preserving the real-time capacity.

1. Introduction

In recent decades, the image target matching not only plays a significant role in many research fields, like the computer vision and digital image processing [1], but also has been widely used in a variety of military and civil applications [2], such as the image target detection, autonomous navigation, 3-dimensional reconstruction, target and scene recognition, and visual positioning and tracking. The image target matching involves two main categories of matching algorithms: gray correlation-based algorithm [3] and feature-based algorithm [4]. Gray correlation-based algorithm is based on the calculation of image similarities and the searching for the extreme values of similarities by using the optimal parameters in transformation model. However, feature-based algorithm mainly relies on the matching of the feature parameters extracted from images (e.g., the points, lines, and surfaces in images). In the condition of slight distortion of gray and geometry, although a large amount of computation cost is

required by gray correlation-based algorithm, it normally outperforms feature-based algorithm, in terms of accuracy, robustness, and antinoise ability. However, in the serious distortion condition, feature-based algorithm is much preferred due to the lower false matching rates and better robustness for gray changes, image deformation, and occlusion.

The basic motivation of addressing 2DMETS based SURF in this paper is to improve the cost efficiency and matching accuracy further. In concrete terms, due to the smaller sizes of descriptors in integral images (e.g., each descriptor in SURF only contains 64 bins which is half the size of the descriptor in SIFT), 2DMETS based SURF requires lower computation cost for the detecting and matching of feature points compared to the conventional SIFT [5]. There are two main steps involved in 2DMETS based SURF: (i) performing 2DMETS to construct 2D gray histogram and (ii) conducting feature point searching and target matching by SURF.

The rest of this paper is structured as follows. In Section 2, we give some related works. In Section 3, the detailed steps of

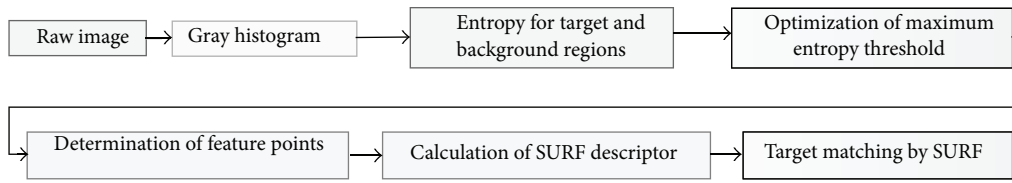


FIGURE 1: Flow chart of 2DMETS based SURF.

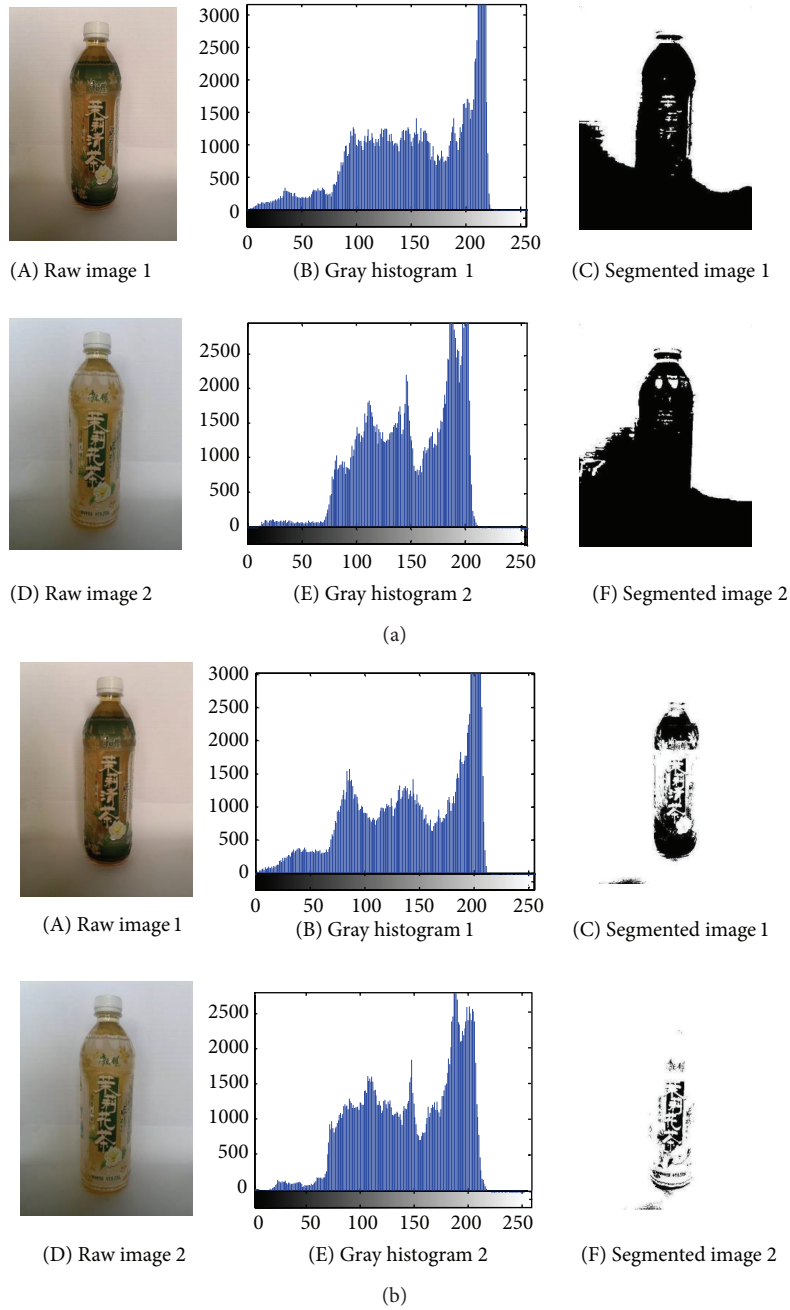


FIGURE 2: (a) Otsu segmentation on images in group 1. (b) 2DMETS on images in group 1.

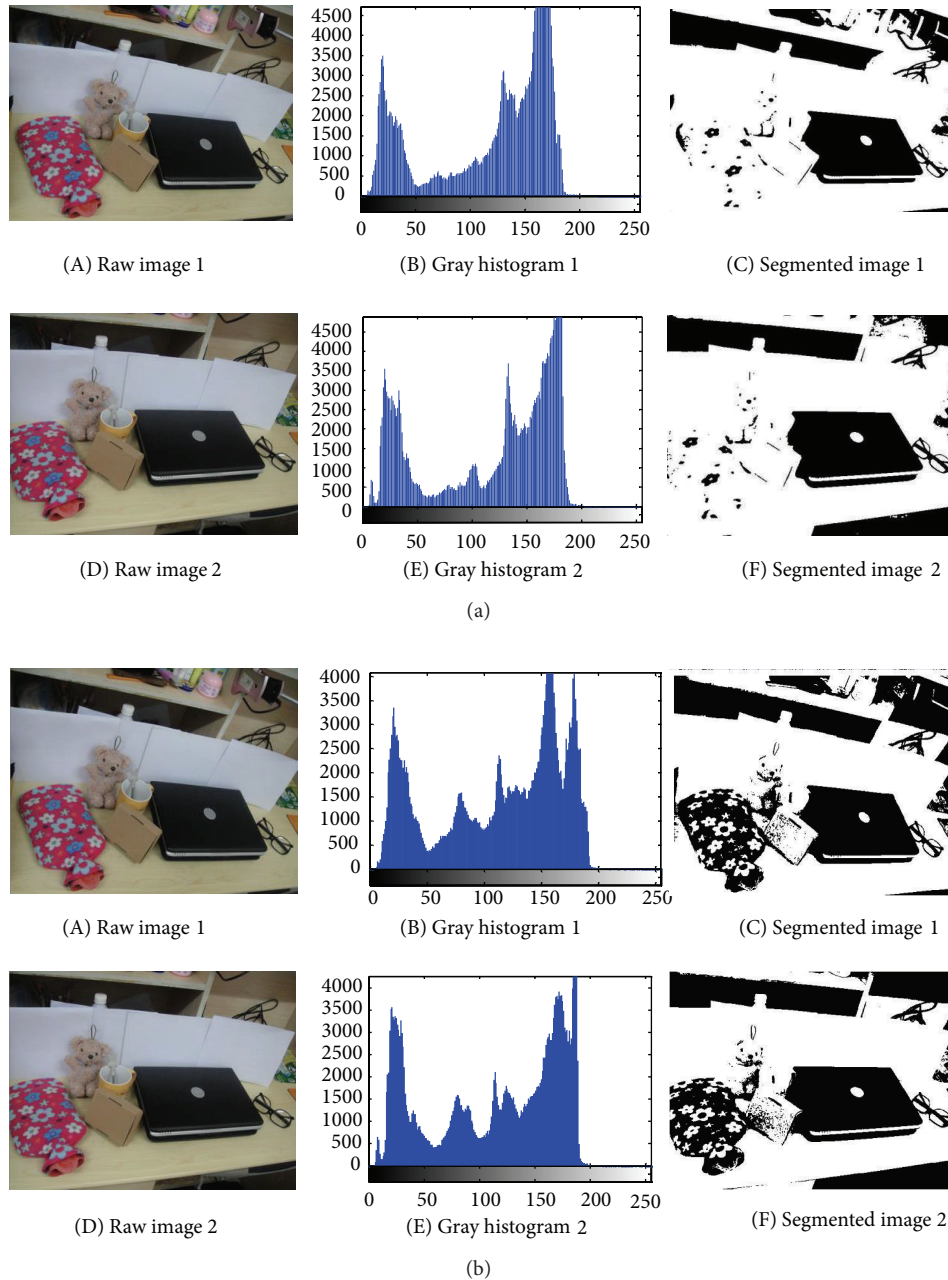


FIGURE 3: (a) Otsu segmentation on images in group 2. (b) 2DMETS on images in group 2.

2DMETS based SURF are discussed. Experimental results are provided in Section 4. Finally, Section 5 concludes this paper and presents some future directions.

2. Related Work

As the first representative work on image target matching, the authors in [6] proposed the cross correlation algorithm to conduct the target matching in remote multispectral and multitemporal images by using the fast Fourier transform. The sequential similarity detection algorithm (SSDA)

addressed in [6] can not only effectively eliminate the unmatched points, but also remarkably save the cost for image matching. Rosenfeld and Kak in [7] used a new concept of cross-correlation-based target matching which relies on the similarities of the gray areas in different templates. Harris operator which was introduced in [8] can extract the angular feature points with good performance in computation cost and stability. Harris and Stephens in [8] illustrated the four main factors involved in image target matching: feature spaces, searching spaces, searching strategies, and similarity measures. Considering medical image deformation property, Collignon et al. in [9] and Viola and Wells in [10] analyzed

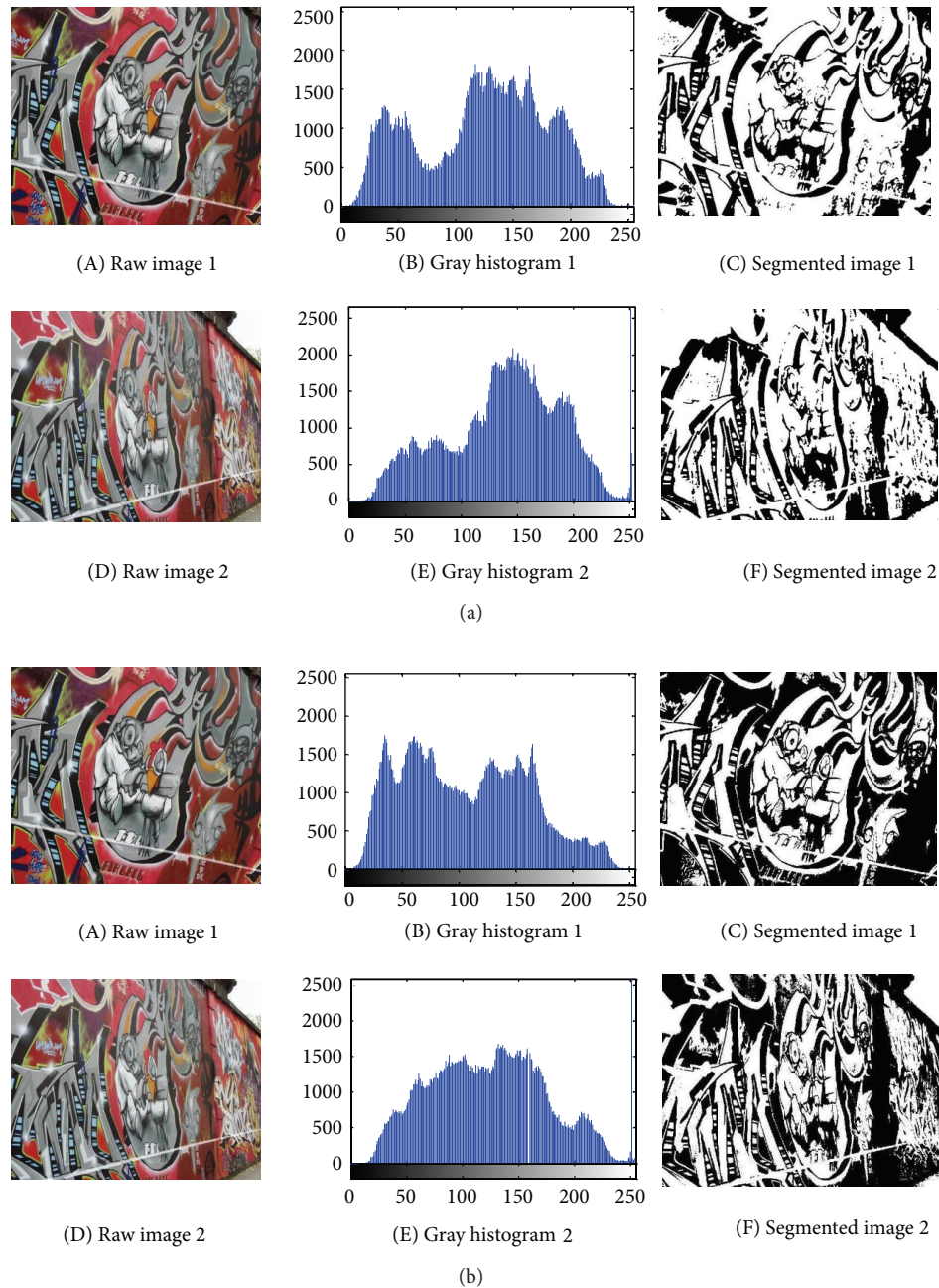


FIGURE 4: (a) Otsu segmentation on images in group 3. (b) 2DMETS on images in group 3.

the mutual information-based medical image target matching. By conducting normalization of rotation and translation to obtain the affine invariant, SIFT [11] was proved to perform well with respect to image rotation, transformation, and zooming [12]. One of the most popular ways to represent local features as the histogram of gradient locations and orientations was introduced in [13].

In recent decades, many institutes and universities proposed a variety of enhanced approaches for image target matching, like the principle component analysis-based SIFT (PCA-SIFT), Harris-SIFT, affine SIFT (ASIFT), shape SIFT (SSIFT), and speeded-up robust features (SURF).

The descriptors in PCA-SIFT can effectively reduce the number and dimensions of feature points. In concrete terms, the descriptors in PCA-SIFT encode salient aspect of image gradients into the neighborhood of feature points and then normalize the gradient patches by using the PCA approach [14]. Harris-SIFT relies on the Harris operators to extract feature points and calculate descriptors [5]. There are two camera axis parameters, latitude angle and longitude angle, considered in ASIFT [15]. Based on the global shape context, SSIFT was applied to recognize the Chinese characters in the images contaminated by complex circumstance in [16]. By using the integral images for the image convolution, SURF

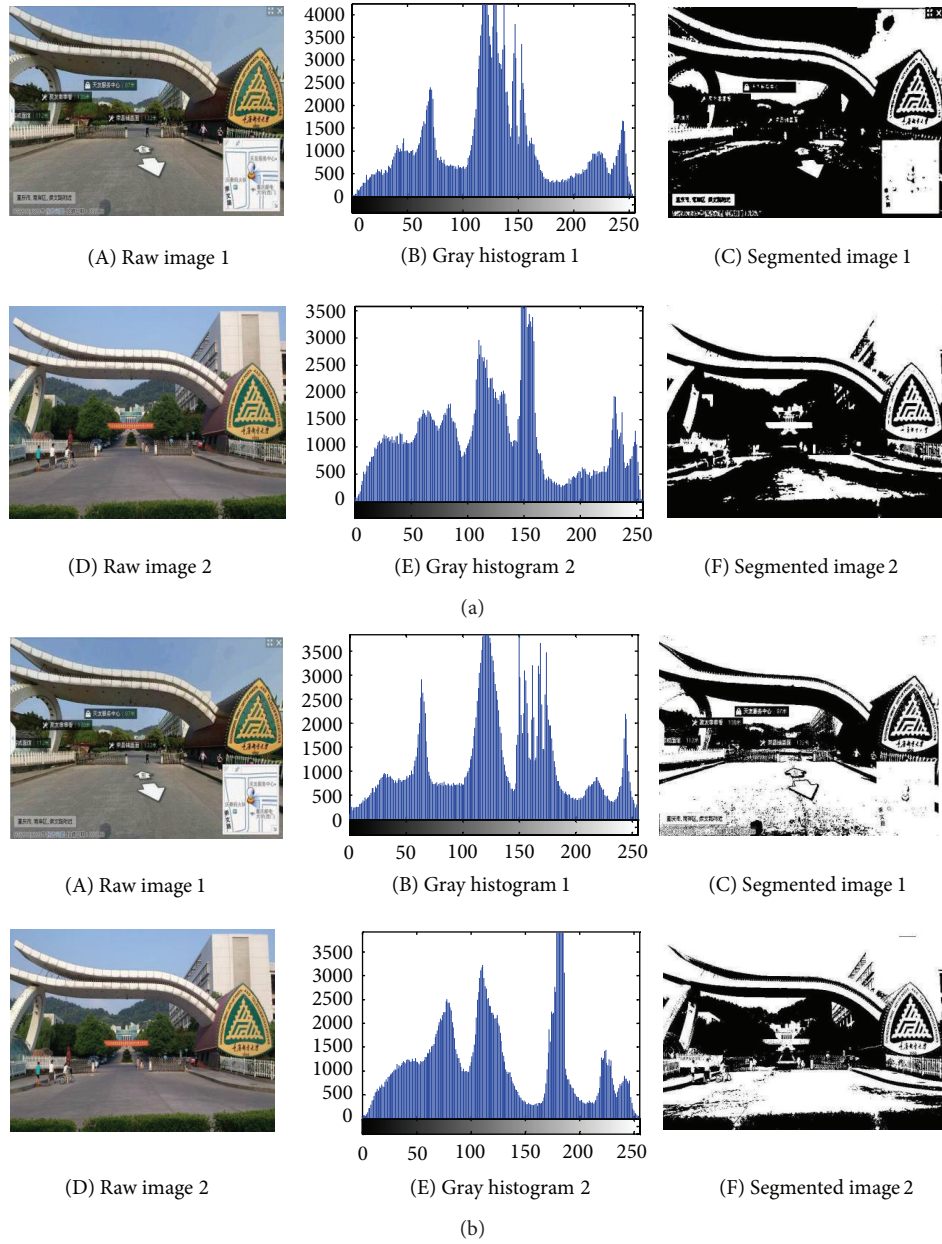


FIGURE 5: (a) Otsu segmentation on images in group 4. (b) 2DMETS on images in group 4.

only requires a small number of histograms to quantize the gradient orientations [17]. Sergieh et al. in [18] studied the way to reduce the number of required features by SURF, while preserving the high correct matching performance. Zhang and Hu in [3] invented the Fast-Hessian detectors for SURF from accelerated segment test (FAST) corner detector. Kai et al. in [19] proposed the normalized SURF to reduce the influence of huge difference on target matching. Juan and Gwun in [20] focused on panorama image stitching by the integration of SURF and multiband blending.

The abovementioned algorithms fail to carefully consider the interference of background noise and edge pixels on image target matching. To fix this problem, we propose the 2DMETS based SURF in this paper. 2DMETS based SURF

can be simply recognized as an integration of 2DMETS and SURF.

3. Steps of 2DMETS of SURF

3.1. Flow Chart. In 2DMETS based SURF, we first construct a 2D gray histogram based on the gray level of each pixel and the average gray level of its 8 neighboring pixels (or *neighborhood*). Then, we conduct image segmentation for the sake of mitigating the interference from background noise and edge pixels. Finally, we use SURF to conduct the target matching. The flow chart of 2DMETS based SURF for image matching is shown in Figure 1.

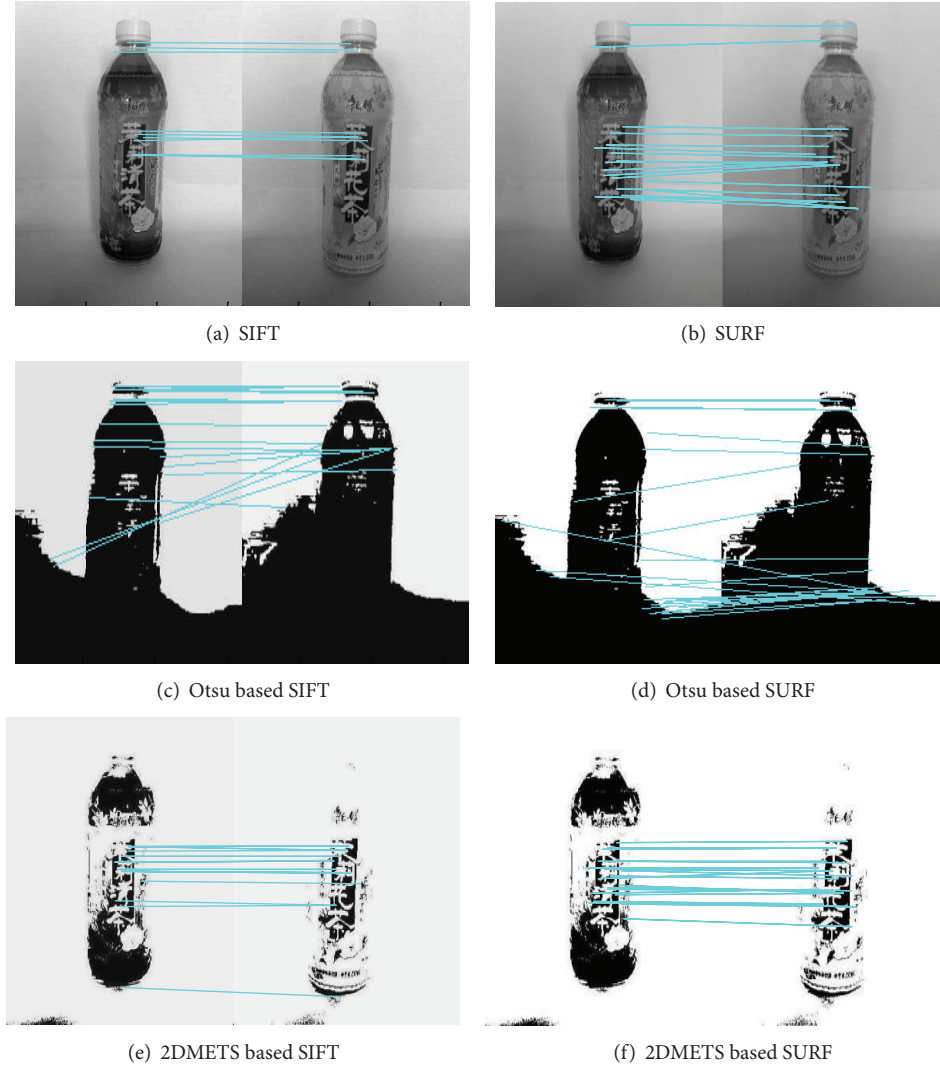


FIGURE 6: Target matching for images in group 1.

3.2. Gray Histogram Construction. For each raw image I (with gray levels L), the gray level of each pixel i ($i = 0, \dots, L-1$) and the average gray level of its neighborhood j ($j = 0, \dots, L-1$) form a pair of gray levels $[i, j]$. We can calculate the probability of each gray level pair $p(i, j)$ ($0 \leq p(i, j) \leq 1$) by

$$p(i, j) = \frac{r(i, j)}{M \times N}, \quad \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} p(i, j) = 1, \quad (1)$$

where $r(i, j)$ denotes the frequency of pair $[i, j]$; M and N stand for the number of pixels in horizontal direction and in vertical direction, respectively, in the raw image. Then, the gray histogram to be constructed in this paper can be recognized as a 2D histogram consisting of the frequencies of gray level pairs.

3.3. Optimization of Maximum Entropy Threshold. If we set the maximum entropy threshold at gray level pair $[s, t]$,

the segmented regions which contain the pairs $[i', j']$ ($i' = 0, \dots, s; j' = 0, \dots, t$) and $[i'', j'']$ ($i'' = s+1, \dots, L-1; j'' = t+1, \dots, L-1$), respectively, stand for the target and background regions. The normalization entropy with respect to target (H_1) and background (H_2) is calculated by

$$H_1 = \sum_{i=0}^s \sum_{j=0}^t \frac{p(i, j)}{p_1(s, t)} \lg \frac{p_1(i, j)}{p(s, t)} = \lg p_1(s, t) + \frac{h_1(s, t)}{p_1(s, t)},$$

$$p_1(s, t) = \sum_{i=0}^s \sum_{j=0}^t p(i, j),$$

$$h_1(s, t) = - \sum_{i=0}^s \sum_{j=0}^t p(i, j) \lg p(i, j),$$

$$H_2 = \sum_{i=s+1}^{L-1} \sum_{j=t+1}^{L-1} \frac{p(i, j)}{p_2(s, t)} \lg \frac{p_2(i, j)}{p(s, t)} = \lg p_2(s, t) + \frac{h_2(s, t)}{p_2(s, t)},$$

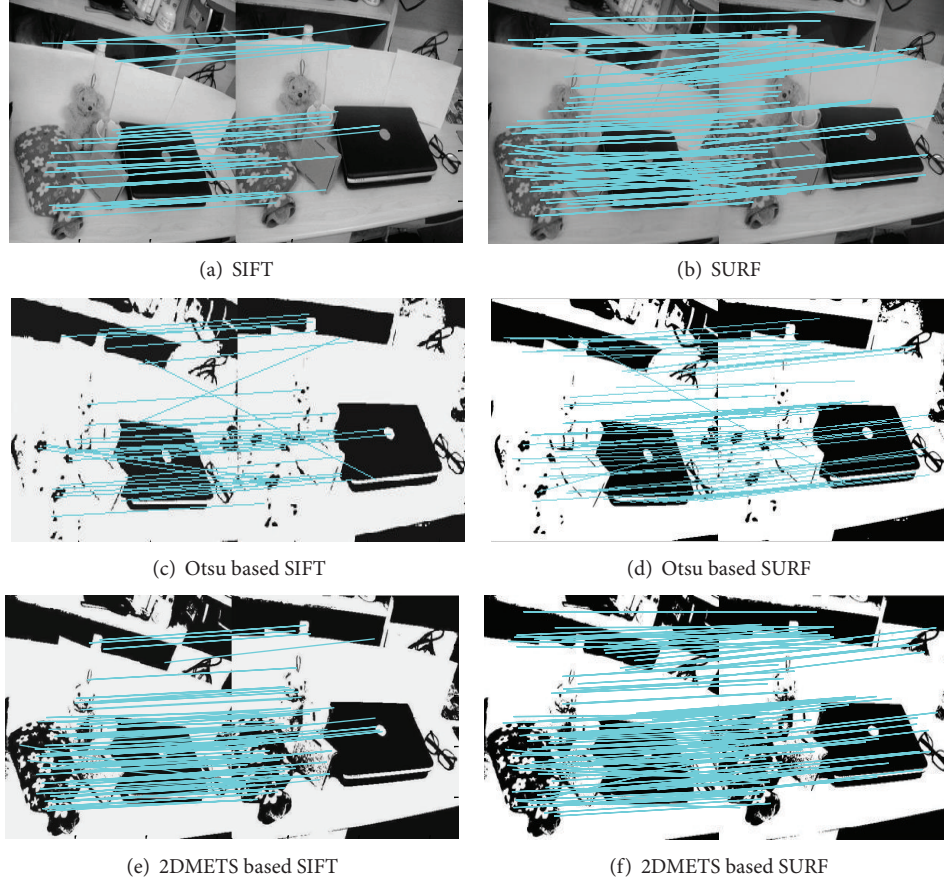


FIGURE 7: Target matching for images in group 2.

$$\begin{aligned}
 p_2(s, t) &= \sum_{i=s+1}^{L-1} \sum_{j=t+1}^{L-1} p(i, j), \\
 h_2(s, t) &= - \sum_{i=s+1}^{L-1} \sum_{j=t+1}^{L-1} p(i, j) \lg p(i, j).
 \end{aligned} \tag{2}$$

The total entropy ($H(s, t)$) for the target and background regions can be obtained by

$$H(s, t) = \lg p_1(s, t) p_2(s, t) + \frac{h_1(s, t)}{p_1(s, t)} + \frac{h_2(s, t)}{p_2(s, t)}. \tag{3}$$

We select the pair $[s^*, t^*]$ which results in the largest total entropy as the optimal maximum entropy threshold, such that

$$[s^*, t^*] = \arg \max_{0 \leq s \leq L-1, 0 \leq t \leq L-1} H(s, t). \tag{4}$$

3.4. Feature Point Determination. The three main steps involved in the determination of feature points are as follows: integral image construction, interest point detection, and Gaussian scale approximation.

After the Gaussian scales have been approximated, all the interested points can be detected. As the final step of

the feature point determination, we compare each interested point with its 26 neighboring pixels in a 3×3 region at the current and adjacent scales by the nonmaximum suppression approach and then localize the feature points at the interest points which have the local maximum or minimum values of box filter responses.

3.5. Calculation of SURF Descriptor. To guarantee the rotation invariance, each feature point is assigned by a reproducible orientation. By assuming that a feature point is found at scale s , Haar wavelet responses with the size $4s$ can be obtained for the neighboring pixels with radius $6s$. The Haar wavelet responses are weighted by the Gaussian scale with $\sigma = 2s$ and then represented as the points in a space centered at the feature point. The longest orientation vector is selected as the dominant orientation to be assigned to the descriptor.

3.6. Target Matching. We adopt the Euclidean distance to evaluate the similarity of every two normalized SURF descriptors ($D_{i,j}$), as described in

$$D_{i,j} = \sqrt{\sum_{k=0}^{64} (X_i - X_j)^2}, \tag{5}$$

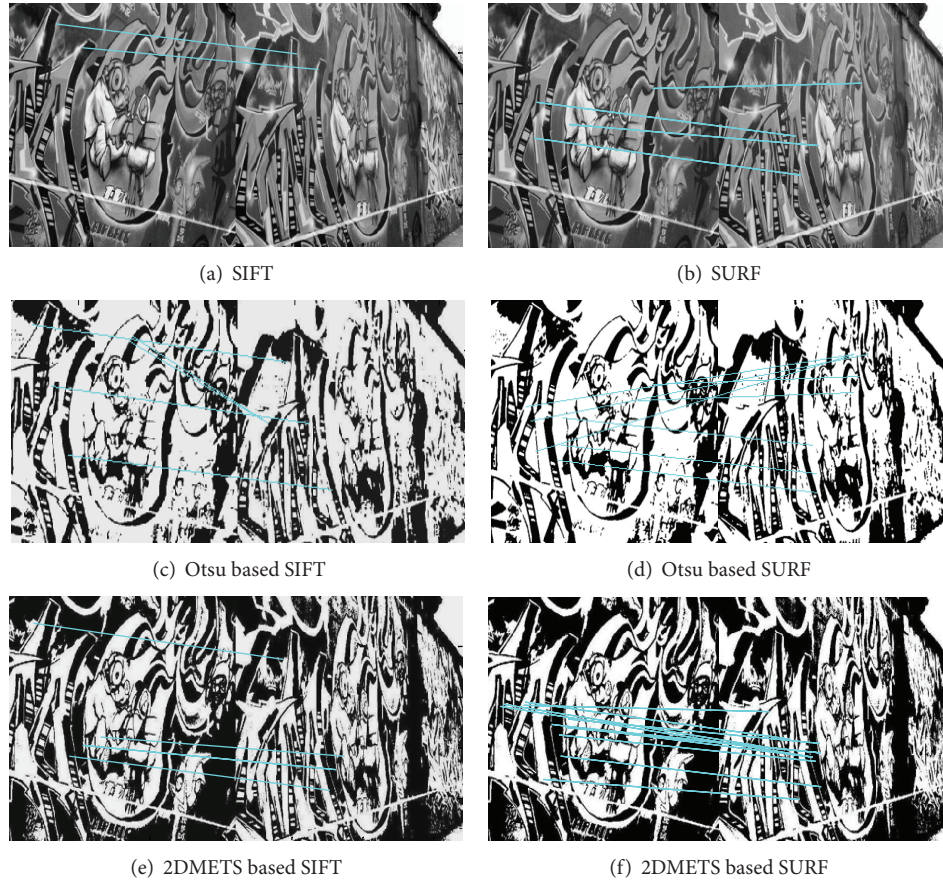


FIGURE 8: Target matching for images in group 3.

where X_i and X_j stand for the i th and j th normalized SURF descriptors in two different images. We calculate the Euclidean distances from each feature point in one of the two images to its first nearest neighbor (1st NN) and second nearest neighbor (2nd NN) in another image. The matching occurs when the ratio (1st NN)/(2nd NN) is larger than a given threshold. In our experiments, we set the threshold as 0.75. The larger threshold could result in the smaller number of matching points between these two images.

4. Experimental Results

4.1. Image Description. There are four groups of images selected for the testing: (i) group 1 (in Figure 2): indoor short-distance images containing one target and with slight difference of illumination intensity and angle rotation; (ii) group 2 (in Figure 3): indoor short-distance images containing multiple targets and with similar illumination intensity, but slight difference of angle rotation; (iii) group 3 (in Figure 4): outdoor long-distance images with great difference of angle rotation; this group of images is also used in [13, 14]; and (iv) group 4 (in Figure 5): image 1 is from the SOSO street view [21], while image 2 is taken by a SONY L26i cellphone. The interference of background noise in this group of images is

more significant compared to the previous three groups of images (e.g., the passing pedestrians).

4.2. Matching Results. First of all, we apply Otsu segmentation and 2DMETS to transform the raw images into black-and-white images in a uniform gray scale to mitigate the interference from background noise and edge pixels, as shown in Figures 2–5. By setting $L = 16$, we have 256 pairs of gray levels, as represented at horizontal coordinates in gray histogram, while the vertical coordinates stand for the frequencies of gray level pairs. Figures 2(a), 3(a), 4(a), and 5(a) show the segmentation results by Otsu, while Figures 2(b), 3(b), 4(b), and 5(b) show the results by 2DMETS.

Second, Figures 6, 7, 8, and 9 show the results of target matching by using SIFT, SURF, Otsu based SIFT, Otsu based SURF, 2DMETS based SIFT, and 2DMETS based SURF for each group of images. Last, the matching performance is compared in Tables 1, 2, 3, and 4.

4.3. Result Discussion

4.3.1. Repeatability. After the affine transformation, if there is a pair of feature points located at the same target in the two different images, a *correspondence* occurs.

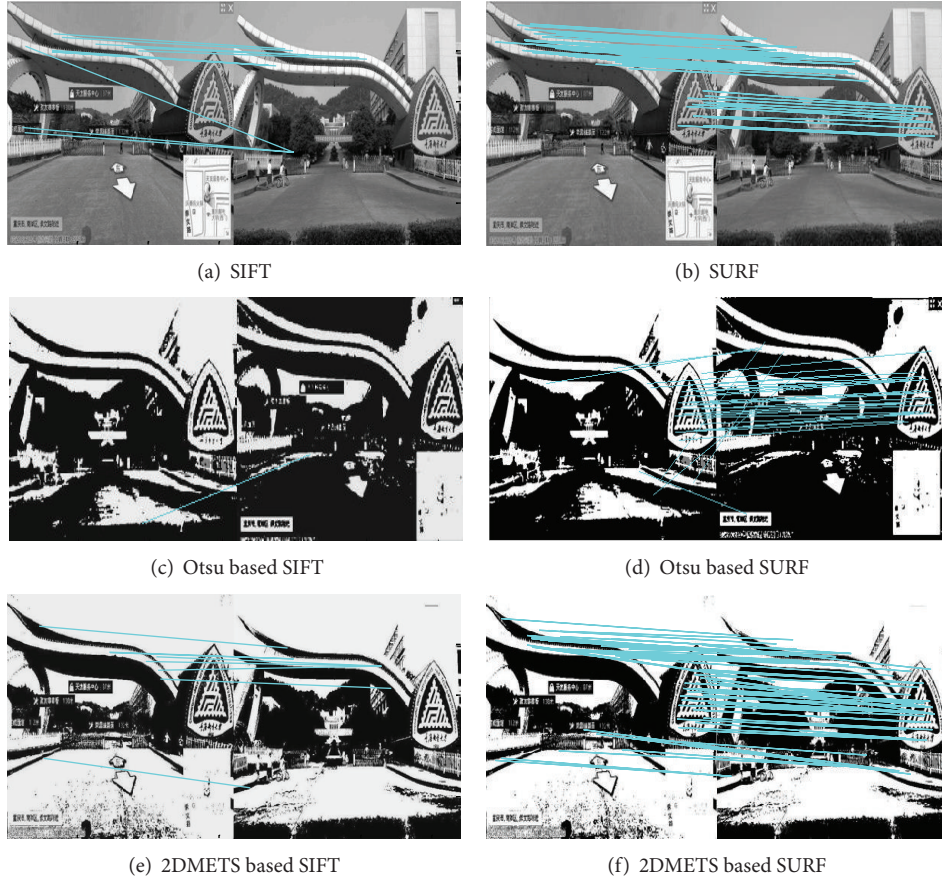


FIGURE 9: Target matching for images in group 4.

TABLE 1: Matching performance for images in group 1.

Parameters	SIFT	SURF	Otsu based SIFT	Otsu based SURF	2DMETS based SIFT	2DMETS based SURF
Number of feature points (left/right images)	84/88	60/69	223/169	82/83	698/321	116/89
Number of correct matches	10	18	21	25	21	22
Number of incorrect matches	1	3	3	2	1	1

Then, we define *Repeatability* as the ratio between the number of *correspondence* and the minimal number of feature points ($\min(\text{feature}_{\text{img1}}, \text{feature}_{\text{img2}})$):

$$\text{Repeatability} = \frac{\text{Number of correspondence}}{\min(\text{feature}_{\text{img1}}, \text{feature}_{\text{img2}})}, \quad (6)$$

where $\text{feature}_{\text{img1}}$ and $\text{feature}_{\text{img2}}$ stand for the numbers of feature points in the two different images, respectively. The higher value of *Repeatability* indicates that the targets are more likely to be matched.

4.3.2. Match Score. *Match Score* is defined as the ratio between the number of *correct matches* and the value

$\min(\text{feature}_{\text{img1}}, \text{feature}_{\text{img2}})$. Obviously, the higher *Match Score* means that the targets are more likely to be matched correctly:

$$\text{Match Score} = \frac{\text{Number of correct matches}}{\min(\text{feature}_{\text{img1}}, \text{feature}_{\text{img2}})}. \quad (7)$$

4.3.3. Correct Matching Rate. We use *Correct Matching Rate* to examine the probability of the targets to be matched. *Correct Matching Rate* is defined as the ratio between the number of *correct matches* and the number of *total matches*.

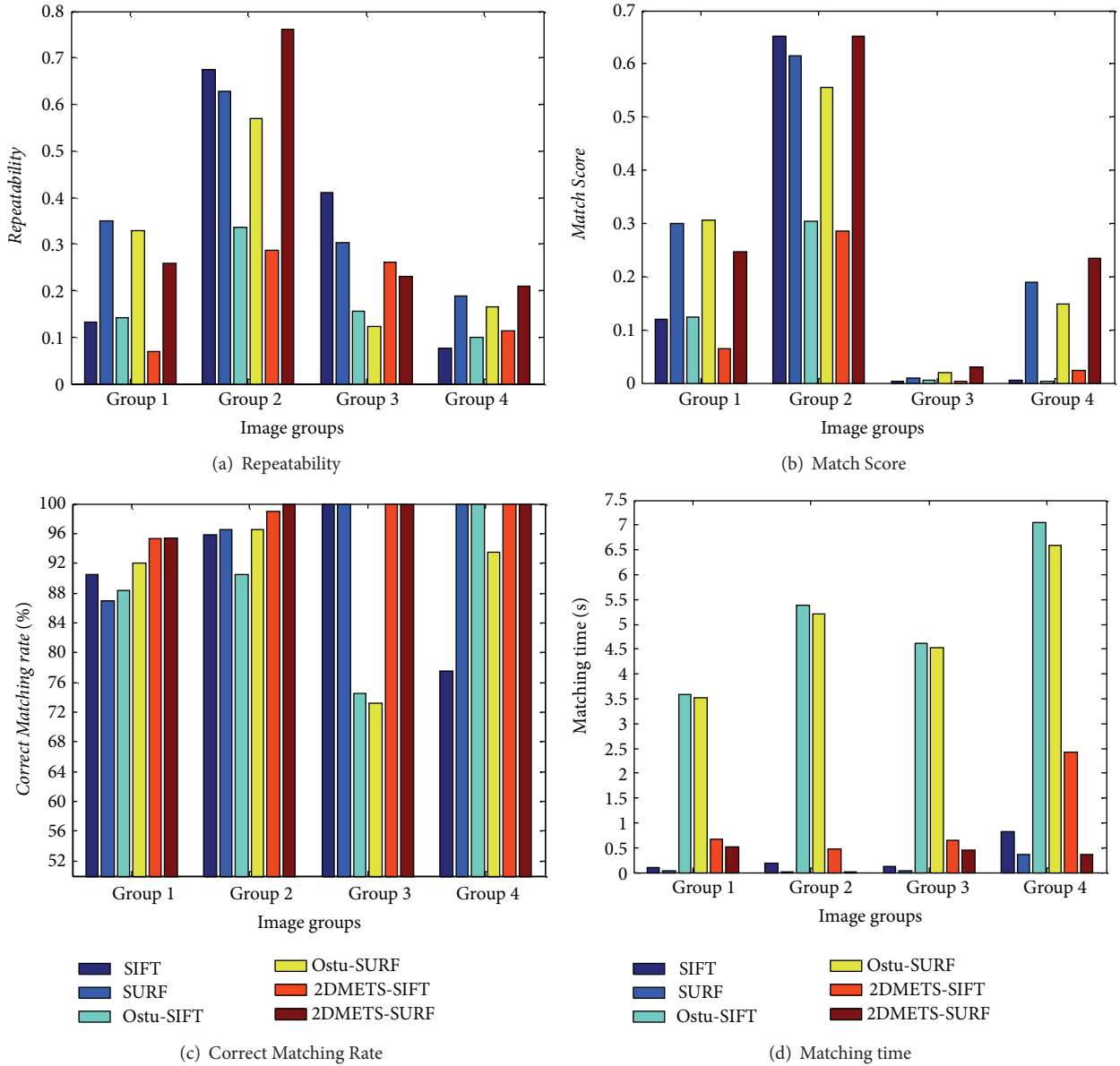


FIGURE 10: Repeatability, Match Score, Correct Matching Rate, and matching time for each group of images.

The higher *Correct Matching Rate* will result in the higher probability for correct matching:

$$\text{Correct Matching Rate} = \frac{\text{Number of correct matches}}{\text{Number of total matches}} \quad (8)$$

4.3.4. *Matching Time.* The matching time determines the real-time capacity of our proposed approach. We define it as the time cost for feature point searching and target matching. The *Repeatability*, *Match Score*, *Correct Matching Rate*, and the matching time for each group of images are shown in Figure 10.

As can be seen from Figure 10, we can find that (i) the targets are very likely to be matched by 2DMETS based SURF

due to the high *Repeatability* achieved; (ii) there is slight influence on *Match Score* by using the SURF with and without 2DMETS; (iii) our proposed 2DMETS based SURF performs best in terms of *Correct Matching Rate*; and (iv) although a little extra time cost is required by 2DMETS processing, the real-time capacity can also be guaranteed by the proposed 2DMETS based SURF.

5. Conclusion

A novel 2DMETS based SURF proposed in this paper is proved to perform well in accuracy and computation cost for image target matching. Compared to the conventional SIFT, SURF, Ostu based SIFT, Ostu based SURF, and the enhanced

TABLE 2: Matching performance for images in group 2.

Parameters	SIFT	SURF	Otsu based SIFT	Otsu based SURF	2DMETS based SIFT	2DMETS based SURF
Number of feature points (left/right images)	151/86	210/161	152/169	175/144	478/508	221/189
Number of correct matches	56	99	46	80	136	144
Number of incorrect matches	2	2	5	2	1	0

TABLE 3: Matching performance for images in group 3.

Parameters	SIFT	SURF	Otsu based SIFT	Otsu based SURF	2DMETS based SIFT	2DMETS based SURF
Number of feature points (left/right images)	827/796	406/429	1570/1211	418/430	1756/1522	378/414
Number of correct matches	2	4	6	8	4	11
Number of incorrect matches	0	0	2	3	0	0

TABLE 4: Matching performance for images in group 4.

Parameters	SIFT	SURF	Otsu based SIFT	Otsu based SURF	2DMETS based SIFT	2DMETS based SURF
Number of feature points (left/right images)	342/182	316/296	339/458	328/269	355/407	296/322
Number of correct matches	11	56	1	40	8	62
Number of incorrect matches	3	0	0	4	0	0

2DMETS based SIFT, an effective improvement of *Correct Matching Rate* without significant loss in real-time capacity is possible, which indicates an important advantage for the time-efficient image processing applications. However, this paper mainly focuses on the target matching between gray images. We will pay more attention to the design of the accurate and cost-efficient image target matching approaches for color images in future.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors wish to thank the editor and all the reviewers for the careful review and the effort in processing this paper. This work was supported in part by the Program for Changjiang Scholars and Innovative Research Team in University (IRT1299), National Natural Science Foundation of China (61301126), Special Fund of Chongqing Key Laboratory (CSTC), Fundamental and Frontier Research Project of Chongqing (cstc2013jcyjA40041, cstc2013jcyjA40032, and cstc2013jcyjA40034), Scientific and Technological Research Program of Chongqing Municipal Education Commission (KJ130528, KJ1400413), Startup Foundation for Doctors of

CQUPT (A2012-33), Science Foundation for Young Scientists of CQUPT (A2012-77), and Student Research Training Program of CQUPT (A2013-64).

References

- [1] K. A. Peker, "Binary SIFT: fast image retrieval using binary quantized SIFT features," in *Proceedings of the 9th International Workshop on Content-Based Multimedia Indexing (CBMI '11)*, pp. 217–222, Madrid, Spain, June 2011.
- [2] G. Schroth, R. Huitl, D. Chen, M. Abu-Alqumsan, A. Al-Nuaimi, and E. Steinbach, "Mobile visual location recognition," *IEEE Signal Processing Magazine*, vol. 28, no. 4, pp. 77–89, 2011.
- [3] H. Zhang and Q. Hu, "Fast image matching based-on improved SURF algorithm," in *Proceedings of the International Conference on Electronics, Communications and Control (ICECC '11)*, pp. 1460–1463, Ningbo, China, September 2011.
- [4] J. C. Yoo and C. W. Ahn, "Image matching using peak signal-to-noise ratio-based occlusion detection," *IET Image Processing*, vol. 6, no. 5, pp. 483–495, 2012.
- [5] Q. Zhang, T. Rui, and H. S. Fang, "Particle filter object tracking based on Harris-SIFT feature matching," in *Proceedings of the International Workshop on Information and Electronics Engineering*, pp. 924–929, 2012.
- [6] D. I. Barnea and H. F. Silverman, "A class of algorithms for fast digital image registration," *IEEE Transactions on Computers*, vol. 21, no. 2, pp. 179–186, 1972.
- [7] A. Rosenfeld and A. C. Kak, Eds., *Digital Picture Processing*, Academic Press, New York, NY, USA, 1982.

- [8] C. J. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of 4th Alvey Vision Conference (AVC '88)*, pp. 147–151, Manchester, UK, August 1988.
- [9] A. Collignon, F. Maes, and D. Delaere, "Automated multi-modality image registration based on information theory," in *Information Processing in Medical Imaging*, pp. 263–274, 1995.
- [10] P. Viola and W. M. Wells III, "Alignment by Maximization of Mutual Information," *International Journal of Computer Vision*, vol. 24, no. 2, pp. 137–154, 1997.
- [11] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV '99)*, pp. 1150–1157, September 1999.
- [12] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [13] M. Amiri and H. R. Rabiee, "RASIM: a novel rotation and scale invariant matching of local image interest points," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3580–3591, 2011.
- [14] Y. Ke and R. Sukthankar, "PCA-SIFT: a more distinctive representation for local image descriptors," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, vol. 2, pp. II-506–II-513, July 2004.
- [15] G. Yu and J.-M. Morel, "A fully affine invariant image comparison method," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '09)*, pp. 1597–1600, Taipei, Taiwan, April 2009.
- [16] Z. Jin, K. Y. Qi, and Y. Zhou, "SSIFT: an improved SIFT descriptor for Chinese character recognition in complex image," in *Proceedings of the International Symposium on Computer Network and Multimedia Technology*, pp. 62–64, 2009.
- [17] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "SURF: speeded-up robust features," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.
- [18] H. M. Sergieh, E. Egyed-Zsigmond, M. Döller, D. Coquil, J. Pinon, and H. Kosch, "Improving SURF image matching using supervised learning," in *Proceedings of the 8th International Conference on Signal Image Technology and Internet Based Systems (SITIS '12)*, pp. 230–237, Naples, Italy, November 2012.
- [19] W. Kai, C. Bo, M. Lu, and X. Song, "Multi-source remote sensing image registration based on normalized SURF algorithm," in *Proceedings of the International Conference on Computer Science and Electronics Engineering (ICCSEE '12)*, pp. 373–377, March 2012.
- [20] L. Juan and O. Gwun, "SURF applied in panorama image stitching," in *Proceedings of the 2nd International Conference on Image Processing Theory, Tools and Applications (IPTA '10)*, pp. 495–499, July 2010.
- [21] <http://map.qq.com/pano=10081147130320163359700&heading=97&pitch=0&zoom=1>.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

