TOPOLOGY BASED AFFINE INVARIANT DESCRIPTOR FOR MSERS

Chenbo Shi[1], *Guijin Wang*[1], *Xinggang Lin*[1], *Yongming Wang*[2],*Chao Liao*[1],*Quan Miao*[1]

[1]Tsinghua University, Department of Electronic Engineering, Beijing
Tsinghua National Laboratory for Information Science and Technology
[2] Advanced Information Technology Institute,Beijing

ABSTRACT

This paper introduces a topology based affine invariant descriptor for maximally stable extremal regions (MSERs). The popular SIFT descriptor computes the texture information on a grey-scale patch. Instead our descriptor use only the topology and geometric information among MSERs so that features can be rapidly matched regardless of the texture in the image patch. Based on the ellipses fitting for the detected MSERs, geometric affine invariants between ellipses pair are extracted as the descriptors. Finally topology based voting selector is designed to achieve the best correspondences. Experiment shows that our descriptor is not only computational faster than SIFT descriptor , but also has better performance on wide angle of view and nonlinear illumination change. In addition, our descriptor shows a good result on multi sensor images registration.

Index Terms— local feature, topology, affine transformation, MSER

1. INTRODUCTION

With stability and repeatability, the local feature[1] has been used in various fields: image stitching, object recognition, image retrieval and automatic 3D reconstruction. Recently many robust and repeatable detectors, like MSER, have been proposed to obtain stable points or regions under notable transformations.

To describe the local features, various descriptors are invented,which can be categorized into two types: context-based and geometry-based[2]. Context-based descriptors make single detection distinctive, such as Filters, SIFT, PCA-SIFT, Shape-context and EBR[2]. The popular SIFT[3] descriptor uses HOG (Histogram of Oriented Gradients) to describe the local region. It can handle different situations such as scaling and rotation. However, it is not robust to the influence of nonlinear illumination and large view change. The computation cost of context-based descriptor is usually high. In the latter type, geometry information such as [4, 5] is extracted to construct descriptor. D. Tell et.al. combined the appearance and topology information to describe the corner points [5]. But their algorithm is not robust enough due to massive false alarm and only suitable to planar situation. Geometry hashing[4] of LAF (Local Affine Frames)[6] based on MSER[7] is presented to describe the region. It is used in matching[8] and object recognition[9, 10]. But point-pair in LAF is sensitive to noise and the computation time increases quadratically with the number of detected regions.

In this paper, a novel topology based affine invariant descriptor is proposed, which is independent on the images texture context. There are two main contributions: Firstly topological information between MSERs is adopted to describe regions. The regions are less sensitive to noise and more distinctive comparing to corner points. Secondly geometric affine constraints between ellipses are easily extracted to distinguish each region pair. Experiments shows that our descriptor outperforms SIFT in wide angle of view and nonlinear illumination change. Further, we explored our descriptor possibility of multi-sensor image registration. The preliminary result is quite promising. In addition, the computation time is 4 times faster than that of the generation of SIFT descriptor.

The remainder of this paper is organized as follows. In Section 2, the framework for topology of detected regions is represented. How to find proper region pairs and extract affine invariant relationship of ellipse pair is introduced in Section 3. Experimental results are shown in Section 4. Finally, we conclude this paper with a discussion in Section 5.

2. FRAMEWORK FOR DESCRIPTOR

The topological structure keeps invariant under different transformations. A pair of detected regions is the basic unit of topology in our description. The descriptor framework contains two steps:

1. Extraction: MSER detector is used to extract the robust regions. An ellipse is fitted to represent each region. Then suitable region-pairs are chosen from relative neighbors of each region presented in section 3.1. Affine descriptor of each ellipse-pair is calculated following section 3.2. In Fig1, the left frame shows a MSER set of A$\sim$F detected in the original image. As A's neighbors, B$\sim$F are a description of region A.

2. Matching: Each region-pair might have multiple correspondences. Best correspondences are determined by combining all the matched pairs. First, a kd-tree on ellipse de-

scriptor is built to get coarse matches. Then we use a selector to pick up the candidate with maximal probability. For example in Fig.1, corresponding pairs are linked by a dotted line. AB is corresponding to A'B' and A"B". A WTA (Winner-takes-all) scheme is designed to choose the best correspondence A'.

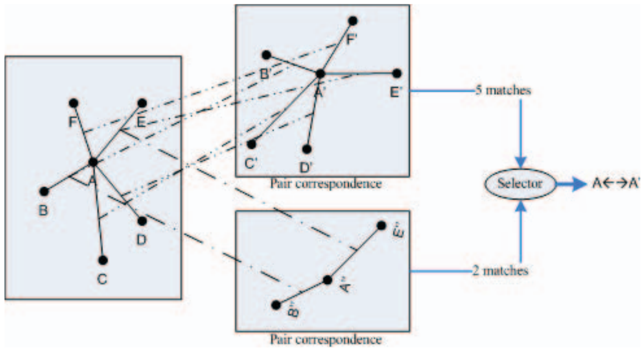

Fig. 1. An example for the descriptor framework. (left) MSER regions in the original image; (middle) correspondences in second image; (right) a selector to pick up the best candidates.

3. DESCRIPTOR FOR DETECTED REGIONS

3.1. Region pair selection

In order to obtain stable neighbors and reduce searching cost, we propose a new distance measure called relative neighbor. In Fig.2 we set the upper left corner as the origin and each dot indicates the position of detected region's center. Different grey level dots indicate the distance from the origin. The selection steps are as follows:

1. Regions are detected as shown in Fig.2(a).

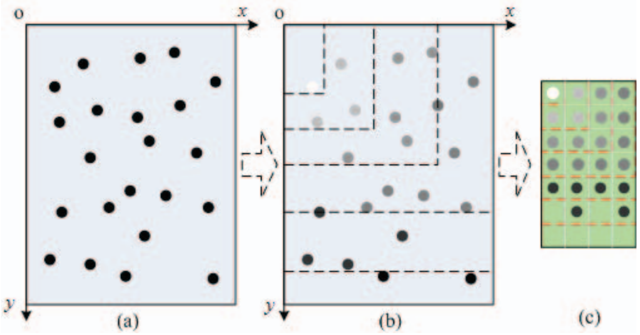

Fig. 2. Region pair selection. (a) detected regions; (b) resort regions and blocks; (c) relative neighbor matrix.

2. All regions are sorted by $max(x, y)$, where (x, y) is the center of each detected region. The regions are divided into some blocks from the origin (see Fig.2(b)). The number of the blocks is determined by: $N_{block} = \sqrt{nr} + 1$, where n is the number of detected regions and r is the length-width ratio.

3. Within each block, the value of $-x + y$ indicates the region distribution. Each region can be arranged to a position in the relative matrix as Fig.2(c). Ellipse pairs can be easily selected from the relative neighbors.

Following [8], some constraints are set to remove unstable pairs and reduce the complexity. Narrow regions are unstable in an affine transformation because of the high error of ellipse fitting. Consequently a filter is designed to remove pairs with significant area differences.

3.2. Affine invariants of ellipse pair

Affine-invariant property is achieved by ellipse fitting with parameters $E(x, y, a, b, \beta)$ for a single region [3], which is more stable to noise and transformation. Ellipse-pairs can be regulated into the same shape by normalization (with only rotational difference) as shown in Fig.3. In the normalized space of ellipse A/A', the parameters of transformed B and B' are identical: $d = d', \phi = \phi'$, where d is the distance between new ellipses centers and ϕ is the angle between new B's major axis and the line of centers.

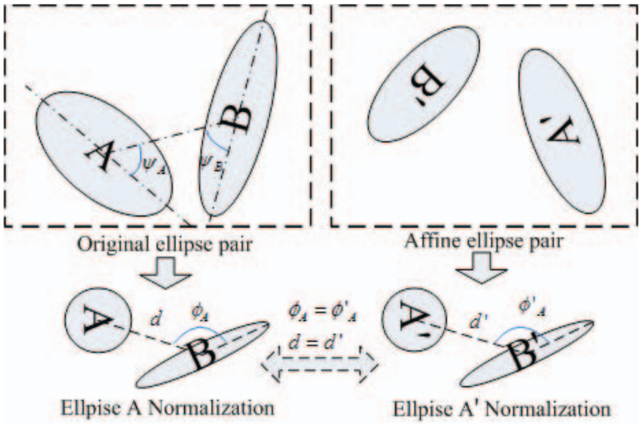

Fig. 3. Ellipse pair normalization under affine transformation.

we propose five affine invariant geometric parameters C_i, i=1∼5 to represent the ellipse pair. Firstly, the ratio of two ellipses' area is invariant:

$$C_1 = \frac{\pi a_A b_A}{\pi a_B b_B} = \frac{a_A b_A}{a_B b_B} \tag{1}$$

.

Secondly, we prove that the distance between centers of normalized ellipses is also constant under affine transform:

$$C_2 = d = \left\| S_{(\frac{1}{a_A},\frac{1}{b_A})} R_{(-\beta_A)} T_{(-x_A,-y_A)} \begin{bmatrix} x_B \\ y_B \end{bmatrix} \right\| \tag{2}$$

Where T, R, S denote translation, rotation and scaling respectively. Similarly, we get C_3 by swapping A and B subscript in Eq.2.

In the normalized space of ellipse A, ellipse B uses the same transformation as A. As a result, the transformed B ellipse is typically still unbalanced and tilted as shown in Fig.4.

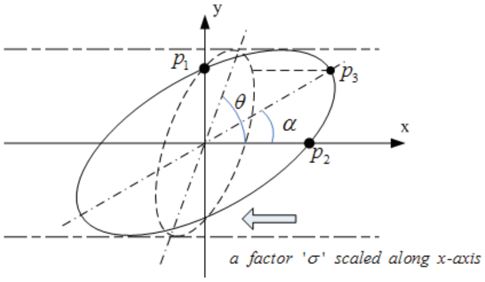

Fig. 4. Ellipse B in the norm space of ellipse A. The solid line ellipse is B and the dash line ellipse is transformed B.

Points p_1, p_2 is on edge of ellipse B with the parameters (α, a, b) where $\alpha = \beta_B - \beta_A$. After a factor $\sigma = a_A/b_A$ scaled along x-axis, the parameters of new ellipse B can be calculated via solving the following equations:

$$\begin{cases} \frac{\sin^2\theta}{a_n^2} + \frac{\cos^2\theta}{b_n^2} = \frac{1}{p_{1y}^2} = \frac{\sin^2\alpha}{a^2} + \frac{\cos^2\alpha}{b^2} \\ \frac{1}{\sigma^2}(\frac{\cos^2\theta}{a_n^2} + \frac{\sin^2\theta}{b_n^2}) = \frac{1}{p_{2x}^2} = \frac{\cos^2\alpha}{a^2} + \frac{\sin^2\alpha}{b^2} \\ a_n b_n = \frac{a_B b_B}{\sigma} \end{cases} \quad (3)$$

Relative angle ϕ in Fig.4 is:

$$C_4 = \phi_A = \theta_A + \sigma_A \arctan(\sigma_A \tan(\psi_A)) \quad (4)$$

Where ψ is the angle between the line connecting the centers of A and B and A's major axis in Fig.4. Similarly, C_5 can be obtained by swapping A and B subscript.

The affine descriptor of region-pair should be normalized on each dimension. The distance between two descriptors is calculated as following:

$$D_{12} = \sum_{i=1}^{5} \mathrm{sgn}(\frac{C_{1i} - C_{2i}}{\max(C_{1i} + C_{2i}, 1.0)} - T_i) \quad (5)$$

where T_i is a presetting threshold on the i-th dimension. The acceptable matched pairs are those whose distance is below a setting threshold. Each matched pairs contains two matched regions according to the area ratio between them. The matched results obtained are called coarse matched regions in this paper.

3.3. Selector

The purpose of selector is to pick up the best correspondences from all coarse matched regions with the same source region, according to the matching probability of each candidate based on the similarity of coarse matches. For example in Fig.1, region A has two candidates A' and A" in region-pair match step. Because of the different topology of the candidates, point A' has 5 matched pair while A" only has 2. Then a WTA (Winner-takes-all) selector is designed to choose the best correspondence A'.

4. EXPERIMENTS

We implemented MSER detection according to [11]. The parameters in MSER detection is set to typical values. Some parameters for our descriptor are set as following: logical neighbor size for each region is $n_s = 8$; area ratio and distance constant tolerance is $T_{2,3} = 0.07$; relative angular tolerance is $T_{4,5} = 0.1$.

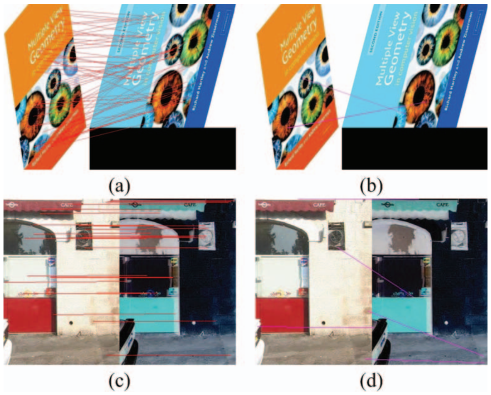

Fig. 5. Matching results. (a), (c) is based on the proposed descriptor while (b), (d) is based on SIFT descriptor.

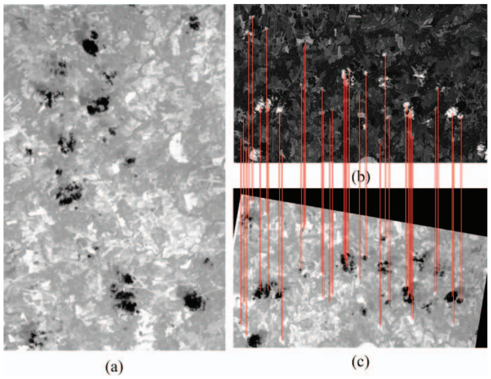

Fig. 6. The multi sensor example: (a) the thermal image; (b) the mid infra-red band image; (c) registration result.

Experiments exhibit our descriptor outperforms SIFT descriptor in various situations. As shown in Fig.5 there is a large affine transformation between images in (a) and (b) and the color of images is different. Our topological descriptor

(a) has more than 40 matched pairs while the SIFT descriptor (b) fails. Nonlinear illumination transformations, for example in pressing industry, has bad influence to matching. In Fig.5(c) and (d), a typical inverted color transformation with white balance is performed on the original image. The topology of the image is preserved but the descriptor based on the context is dissimilar. Our new descriptor obtains more than 30 corresponding results in (c) while the SIFT descriptor doesn't work. All matching lines between correspondences shown in figures are before the RANSAC processing.

We test our descriptor to match the multi sensor images. The input images shown in Fig.6 are the thermal (left) and the mid infra-red band (right-top) of the same place of the Landsat Thematic Mapper[12]. It is shown that the image transformation does not follow a simple model because of the different spectral reflectance of individual materials. The effects cannot be removed by normalization. However, transformed result (right-bottom) can still be obtained from topology by our descriptor. About 40 correct correspondences (the lines between (b) and (c)) are found after RANSAC and the computing speed is fast.

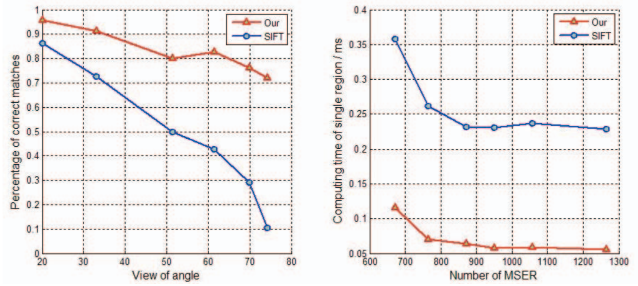

Fig. 7. Compared results with normalized SIFT descriptor on Graffiti dataset. left: the percentage of correct matches with different angle of view; right: the computation cost of each region descriptor.

Fig.7 illustrates the percentage of correctly matched pair and processing time cost on the Graffiti dataset[13] with different angle of view. The precision of the descriptor is shown in left frame. The detected results in large angle of view are more stable than normalized SIFT descriptor. At 70^o angle of view, the new descriptor can still obtain the transformation matrix, while normalized SIFT descriptor fails. As shown in right frame, the time cost of each topological descriptor is about 0.07ms running on a 3.0GHz Intel CPU. The speed is 1/4 of SIFT descriptor on the same platform.

5. CONCLUSION

In this paper, we propose a new descriptor based on the topology of MSERs. The descriptor depends on the geometric affine invariants between detected regions. Compared to the descriptors based on the image context, the new descriptor keeps more stable information about object structure. The proposed descriptor is less sensitive to nonlinear illumination and is more robust in structure images than the formers. Besides, the computation is reduced to below 1/4 of that of the normalized SIFT descriptor. Since MSER detected regions are sensitive to blurs, the new descriptor does not work very well on blurred images. We hope to extend our work to handle such situation in the future.

6. REFERENCES

[1] C. Schmid A. Zisserman J. Matas F. Schaffalitzky T. Kadir L. Van Gool K. Mikolajczyk, T. Tuytelaars, "A comparison of affine region detectors," *IJCV*, vol. 65, no. 1/2, pp. 43–72, 2005.

[2] Li Jing and N. M. Allinson, "A comprehensive review of current local features for computer vision," *Neurocomputing*, vol. 71, no. 10-12, pp. 1771–87, 2008.

[3] D. Lowe, "Distinctive image features from scale invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.

[4] Yehezkel Lamdan Wolfson and Haim, "Geometric hashing: A general and efficient model-based recognition scheme," *Proc.ICCV*, pp. pages 238 – 249, 1988.

[5] D. Tell and S. Carlsson, "Combining appearance and topology for wide baseline matching," in *ECCV 2002, 28-31 May 2002*.

[6] J. Matas, T. Obdrzalek, and O. Chum, "Local affine frames for wide-baseline stereo," in *ICPR, 11-15 Aug. 2002*.

[7] Chum O. Urban M. Pajdla T. Matas, J., "Robust wide baseline stereo from maximally stable extremal regions," *BMVC*, pp. 384–393, 2002.

[8] Ondrej Chum and Jiri Matas, "Geometric hashing with local affine frames," in *CVPR 2006, June 17-22, 2006*.

[9] M. Perd'och, O. Chum, and J. Matas, "Efficient representation of local geometry for large scale object retrieval," in *CVPR Workshops, 20-25 June 2009*.

[10] P. E. Forssen and D. G. Lowe, "Shape descriptors for maximally stable extremal regions," in *ICCV, 14-21 Oct. 2007*.

[11] D. Nister and H. Stewenius, "Linear time maximally stable extremal regions," in *ECCV 2008, 12-18 Oct.2008*.

[12] NASA Landsat Program, "Landsat.u.s. geological survey," *online: http://www.landcover.org*.

[13] mikolajczyk, "Feature detector evaluation sequences," *online: http://lear.inrialpes.fr/people/mikolajczyk/*.