

Matching Image with Multiple Local Features

Yudong Cao

Pattern Recognition and Intelligent System Laboratory,
Beijing University of Posts and Telecommunications,
Beijing, China
Liaoning University of Technology,
Jinzhou, China
ydcdo@bupt.cn

Honggang Zhang, Yanyan Gao, Xiaojun Xu,
Jun Guo

Pattern Recognition and Intelligent System Laboratory,
Beijing University of Posts and Telecommunications,
Beijing, China
{zhhg, gaoyanyan, guojun}@bupt.edu.cn

Abstract— In this paper, we present the fusional feature composed of Affine-SIFT, MSER and color moment invariants. The fusional feature is more robust and distinctive than a single local feature. Instead of adding three local features together simply, an efficient two-level matching strategy is devised with the fusional feature, which speeds up the establishment of the local correspondences. To remove partial false positives, an affine transformation is estimated with the weighted RANSAC which decreases iteration times. The experimental results show that our approach can achieve more accurate correspondence. We prospect to apply the fusional feature and match strategy to image retrieval in the end.

Keywords- *image match; local feature; epipolar geometry; RANSAC*

I. INTRODUCTION

Image matching is a fundamental task for many applications of computer vision such as image retrieval.

Compared to the global feature, the local feature has been proven to be robust to background noise, occlusion and spatial deformations. In recent years, it is popular that each image is represented by a bag of words. Many local features are presented including SIFT [8], GLOH, SURF, MSER [7, 9], etc. A modified MSER proposed in [4] shows stable result. The joint use of two local features has emerged in [12], which is not sufficient.

In image match, finding reliable correspondences is critical. Firstly, initial match searches the correspondence for each keypoint in query image based on traversal; then, the accurate match removes some false correspondences (outliers) from the initial correspondences by epipolar geometry consistency constraint [5, 10, 11, 13]. Full geometry verification using RANSAC can make up for the drawback of the bag of words model ignoring spatial formation.

Our work is inspired by Dong Zhang and Weiqiang Wang [10]. We propose the fusional feature which is more robust and distinctive than a single local feature. Furthermore, a more efficient two-level image matching algorithm is devised based on the fusional feature, which speed up the establishment of the local

correspondences. Full geometry verification with RANSAC further deletes partial outliers.

II. THE FUSIONAL FEATURE DESCRIPTORS

Every local feature has its own advantages and disadvantages. The fusional feature makes full use of one type local feature's advantages making up for another type local feature's disadvantages. We represent an image with the fusional feature including Affine-SIFT, MSER and moment invariants.

SIFT features are invariant to image scale, rotation and translation by exploiting scale-space extrema and the local dominant orientation. The Affine-SIFT introduced in [1] is a fully affine invariant feature. It can undergo very large affine distortions, but ignoring color information of image.

MSER region can deform automatically with changing viewpoint, which is normalized as an ellipse with the center of the gravity of region as its keypoint. Usually MSER detector generates a relatively small number of regions in an image, and their repeatability and distinctness are higher [2] than that of the SIFT. Unfortunately, MSER is not robust to scale change.

MSER region is described by 18-dimension color moment invariants which are reliable and distinctive. The moment invariants are also robust to noise; in addition, they implicitly characterize the shape, the intensity and the color distribution of the region pattern in a uniform manner [3].

III. MATCHING STRATEGY

The three local features affect each other. The MSER detector provides stable and repeatable segmentation region of image. The Affine-SIFT descriptors are only computed inside the MSER region, and color moment invariants contribute to MSER matching. If there are not Affine-SIFT keypoints inside a MSER region, the MSER region will be ignored, which means color moment invariants are not computed on it, and the match for MSER is stopped. Affine-SIFT match only processes on the corresponding MSER regions. So the establishment of initial correspondences is speeded up. Apparently, the three local features are merged into the fusional features compatibly.

The detailed matching algorithm is described as follows.

A. Local Initial Matching

Some matching strategies, which are often used, contain threshold-based matching, nearest neighbor matching, nearest neighbor distance ratio matching and normalized cross-correlation [2, 3]. Especially nearest neighbor distance ratio is very efficient.

Firstly, color moment invariants on normalized MSER regions are computed; then, each MSER region in query image is matched to the MSER region in the other image if the Mahalanobis-distance between the two moment invariants is minimal and below a predefined threshold p . Normalized correlation between the corresponding regions is used as a final check of similarity. Finally, we compare each Affine-SIFT descriptor inside MSER region of the query image with each Affine-SIFT descriptor inside corresponding MSER region of the other image based on nearest neighbor distance ratio method. If (1) is satisfied and the distance between keypoint and its nearest neighbor is below threshold l , the match is identified.

$$\text{dist}(A, B) / \text{dist}(A, C) < r. \quad (1)$$

where keypoints B and C in the other image are the first nearest neighbor (1-NN) and the second nearest neighbor (2-NN) of keypoint A in query image, and r is distance ratio threshold.

Generally, we prefer to believe that the match with a low threshold r and l is more accurate without other prior information. If the ratio computed between 1-NN and 2-NN is very high (say, tending to 1), the two nearest neighbors are almost equally near to the keypoint A so it is unreliable to select 1-NN as the correspondence of keypoint A , or on the other hand, two nearest neighbors are all far from the keypoint A so that we have best grounds to refuse. After finding all the matched keypoint pairs, we define them as the initial correspondences (ICs).

B. Global Geometric Matching

Some false correspondences are inevitable. These can also be caused by symmetries in the image, etc. We reduce these false positives in the initial correspondences through affine geometry model estimated with weighted RANSAC.

Concerning correspondences between two images projected from an object or scene, the point \mathbf{m} in the query image and the point \mathbf{m}' in the other image are related with $\mathbf{m}^T F \mathbf{m}' = 0$, where F is a 3×3 fundamental matrix, \mathbf{m} and \mathbf{m}' are represented with homogeneous image coordinate $(x_q, y_q, 1)^T$ and $(x_r, y_r, 1)^T$. Matrix F is estimated from initial

correspondences with RANSAC. The process is depicted as follows,

Step 1: Extract a sample of 7 correspondences from the initial correspondences set uniformly and at random.

Step 2: Compute the fundamental matrix F with current sample. Compute consistent correspondences set D with (2).

$$S(F) = \{(\mathbf{m}, \mathbf{m}') \in D \mid d^2(\mathbf{m}', F\mathbf{m}) + d^2(\mathbf{m}, F^T \mathbf{m}') < t^2\} \quad (2)$$

where the point \mathbf{m} (or \mathbf{m}') should lie on the corresponding epipolar line ideally, d denotes the distance between point \mathbf{m} (or \mathbf{m}') and the corresponding epipolar line, and t is threshold of distances. The all passed point pairs $(\mathbf{m}, \mathbf{m}')$ constitute consistent correspondences set D .

Step 3: If current set D is more than the previous one. The current matrix F and set D are preserved; and discard the previous set D and corresponding fundamental matrix F .

Step 4: Self-adaptive algorithm decides the iterative times, finally generates a maximum consistent correspondence set D containing all inliers.

Keypoint pairs with a larger similarity are more probably right match. According to the distance of keypoint pair $(\mathbf{m}, \mathbf{m}')$ in local initial match, the weightings are assigned to every initial correspondence. The subsample set of 7 correspondences is sampled from a prior distribution [10]. The weighted RANSAC algorithm can converge as soon as possible. Finally we define the maximum consistent correspondence set D as accurate correspondences (ACs).

C. The Analysis to Match Strategy

Since the interesting objects usually lie in the MSER regions, we can ignore the Affine-SIFT keypoints outside of MSER region, and select only Affine-SIFT keypoints inside MSER region. As the number of MSER regions is by far lower than that of Affine-SIFT, the match strategy based on established the corresponding MSER regions for Affine-SIFT descriptors reduces comparative times.

Apparently, we can also apply epipolar geometry consistency constraints to those MSER correspondences before matching Affine-SIFT. Epipolar geometry constraints can only eliminate partial outliers; thus, it is inappropriate to loose intentionally the thresholds in local initial match relying on the following epipolar geometry consistency constraint to eliminate all outliers, as has been certified in the experiment of Sec.IV.

IV. EXPERIMENTAL RESULTS

In the experiments, we discard the small MSER regions which are less discriminative and the big regions which are less repeatable. The global geometric match method is only applied to the Affine-SIFT keypoints. The threshold t is set to $2\sqrt{2}$.

TABLE I. FIVE IMAGE PAIRS MATCHING WITH OUR APPROACH.

	Image Pairs				
	<i>bike</i>	<i>tree</i>	<i>wall</i>	<i>building</i>	<i>woman</i>
ICs	197	79	138	87	21
ACs	137	36	66	42	13
ratio	0.695	0.456	0.478	0.482	0.619

The first three are from INRIA dataset. The last two can be seen in Fig.1 and 2.

Tab.1 summarizes the number of established initial correspondences (ICs) and the number of accurate correspondences (ACs) that were found to be consistent with the epipolar geometry. The high ratio between the accurate correspondences and initial correspondences guarantees the fast RANSAC termination. Generally more than 40% of initial correspondences are found epipolar geometry consistent [15]. It shows that our initial match strategy is also efficient. The last two visual effects of match can be seen from Fig. 1 and 2.



Figure 1. Building with resolution 352x288. The matching of ICs (top row), the matching of ACs (down row). (similarly hereinafter)

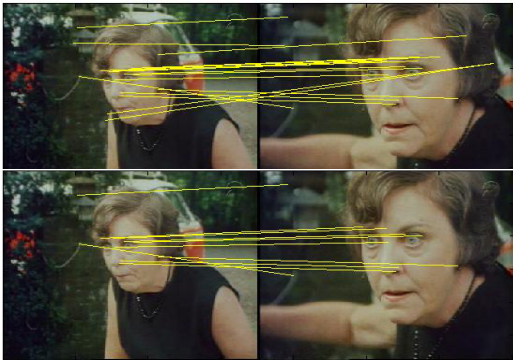


Figure 2. Woman keyframes extracted from video.

We continue to test the algorithm on some image pairs which even have serious changes. Original image pairs are presented in Fig.3, most of which are from the INRIA Copydays dataset. Tab.2 summarizes their match results.

TABLE II. MATCH RESULTS OF SOME IMAGE PAIRS.

	Image Pairs							
	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>
ICs	3667	752	127	697	35	74	314	370
ACs	3652	742	119	569	24	41	76	64
ratio	0.99	0.98	0.94	0.82	0.69	0.55	0.24	0.17

Original image pairs can be seen in Fig.3.

After analyzing the results in Tab. 2, we conclude that the number of accurate correspondences (ACs) between image pairs with blur, rotation and foreshortening is relatively small. In fact, it is difficult to find precise match position in the blurry images. The rotation and foreshortening probably seriously deform the images so many initial correspondences (ICs) don't be in accord with epipolar geometry consistency constraint again.



Figure 3. Some image pairs have zoom, blur and foreshortening etc.

The threshold of r and l can be adjusted up to select more matches or down to select only the most reliable. We gain different initial correspondences set through varying r and l on boat image pair match. Homography H of the boat images is provided by the INRIA dataset [2]. Ideally, if $m' = Hm$ is satisfied, a correspondence (m, m') is true correspondence. Formula (3) is designed for this experiment.

$$d^2(\mathbf{m}', H\mathbf{m}) + d^2(\mathbf{m}, H^{-1}\mathbf{m}') < h^2 \quad (3)$$

where d denotes the Euclidean distance between two points. If we hope that the distance between \mathbf{m}' and the ground-truth match of \mathbf{m} is less than 2.5 pixels, or vice versa, the threshold h should be set to $2.5\sqrt{2}$. We estimate the proportion of true correspondences (TCs) from the accurate correspondences (ACs) following the different initial correspondences set size with (3); then, draw the relation curve on Fig.4. The result shows that the correct rate decreases with the increase in the number of the initial correspondences (ICs).

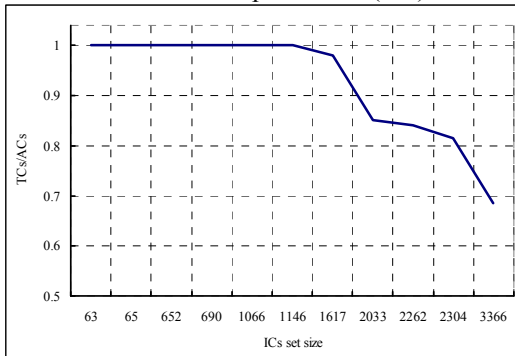


Figure 4. Experimental results for boat image pairs matching. (Images pairs can be available at <http://www.robots.ox.ac.uk/~vgg/research/affine>).

V. CONCLUSIONS

In this paper, the distinctive fusional feature is presented, which is more robust to spatial deformation, occlusion, scale change, viewpoint change and noise. The combination of Affine-SIFT and MSER with moment invariants will be more helpful to local object matching than full object matching. The two-level match strategy with the fusional feature is efficient.

It is prospective that applying the fusional feature to the inverted file index for image retrieval, for the much more geometry information contained in index will improve the performance of retrieval. It is a challenging task to compute visual words from the fusional features. The ellipse denoting MSER or other distinguished region can also be quantified and clustered according to [14]; then, a geometric vocabulary is learned. Some inspiration can be gained from [12]. Furthermore, a relevance score function needs be designed for ranking images. Attentively too high or low threshold and ratio in forementioned local match strategy will be harmful to precision and recall of retrieval. Generally we only apply affine geometry transform to top k returned images in image retrieval, i.e. re-ranking k candidate images.

ACKNOWLEDGMENT

This paper was partially sponsored by the grants from National High-tech 863 Project of China under Grant No. 2007 AA01Z417, the Fundamental Research Funds for the Central Universities, 111 Project of China (B08004) and Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry.

REFERENCES

- [1] J.M. Morel and G. Yu, "ASIFT: A new framework for fully affine Invariant Image Comparison," *SIAM Journal on Imaging Sciences*, 2009.
- [2] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.
- [3] T. Tuytelaars and V. Gool. Matching, "widely separated views Based on affine invariant regions," *Int'l Conf. Computer Vision*, 59(1), 61–85, 2004.
- [4] R. Kimmel and C. Zhang, "Are MSER features really interesting?" *Trans. IEEE on Pattern Analysis and Machine Intelligence*, 2009.
- [5] D. Chetverikov and Z. Megyesi, "Finding region correspondences for wide baseline stereo," *International Conference on Pattern Recognition*, 2004.
- [6] Fischler and Bolles, "Random sampling consensus—A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. Assoc. Comp. Mach.*, 24(6):381–395, 1981.
- [7] K. Mikolajczyk and T. Tuytelaars, "A comparison of affine region detectors," *International Journal of Computer Vision*, 65 (1–2) (2005) 43–72.
- [8] D.G. Lowe, "Distinctive image features from scale invariant keypoints," *International Journal of Computer Vision*, 60(2): pp.91–110, 2004.
- [9] J. Matas and O. Chum, "Robust wide baseline stereo from maximally stable extremal regions," *British Machine Vision Conference*, pp.384–393, 2002.
- [10] D. Zhang, Weiqiang Wang. Matching images more efficiently with local descriptors. *International Conference on Pattern Recognition*, 2008.
- [11] G. Carneira and Allan D. Jepson, "Pruning local correspondences using shape context," *International Conference on Pattern Recognition*, 2004.
- [12] Z. Wu and Q. Ke, "Bundling features for large scale parital-duplicate web image search," *IEEE Conference on Computer vision and Pattern Recognition*, 2009.
- [13] K. Palander and Sami S. Brandt, "Epipolar geometry and log-polar transform in wide baseline stereo matching," *International Conference on Pattern Recognition*, 2008.
- [14] M. Perdoch and O. Chum, "Efficient representation of local geometry for large scale object retrieval", *IEEE Conference on Computer vision and Pattern Recognition*, 2009.
- [15] J. Matas and S. Obdrzalek, "Local Affine frames for wide-baseline stereo," *International Conference on Pattern Recognition*, 2002.