6-1
MVA2011 IAPR Conference on Machine Vision Applications, June 13-15, 2011, Nara, JAPAN

MCM: An Efficient Geometric Constraint Method for Robust Local Feature Matching

Kai Chen Yi Zhou Qi Zheng Xiaokang Yang Li Song
Institute of Image communication and Information Processing
Shanghai Jiaotong University,China
Kchen@sjtu.edu.cn

Abstract

Local feature matching is not robust to extract correct correspondences under many conditions, such as images with general deformations and repetitive patterns. To solve this problem, this paper proposes a new geometric constraint method: Maximal Clique Matching (MCM). In MCM, the global geometric constraint problem can be expressed as the maximal clique problem in graph theory. MCM starts from building a geometric correspondence graph (GCG) based upon the pairwise geometric information in local features, and then an efficient heuristic approximation algorithm is developed to get the global geometric relationships by finding the maximal cliques in GCG. Given the characteristics of the global optimality of maximal cliques, MCM is robust to occlusion, clutter, deformations and repetitive patterns. We evaluated the method using two public datasets. Results show that our method outperforms other up-to-date techniques.

1 Introduction

Local features have been very successful in solving a wide variety of problems in computer vision, from image matching, image retrieval, stereo-vision to object recognition. However, lack of global information may cause ambiguities under some conditions, such as repetitive patterns (such as chessboard and building with windows) and general deformations (such as viewpoint change).

Our goal is to improve the accuracy of image matching using local features in the presence of repetitive patterns and general deformation.

This paper proposes a novel geometric constraint method: Maximal Clique Matching (MCM). Our method can be categorized as pairwise constraint method. In MCM, geometric correspondence graph (GCG) is built upon pairwise geometric constraint extracted from local features. Then we develop an efficient approximation heuristic algorithm to find all the global geometric relationships. Given the characteristics of the global optimality of maximal cliques, MCM is robust to repetitive patterns by effectively ignoring outliers.

The rest of paper is organized as follows. Section 2 discusses the related work. Section 2 introduces our method. Section 3 describes two experiments in comparison with several recent methods and analyzes the results. Section 4 concludes.

2 Related Works

To overcome the general deformation and repetitive patterns, some methods use global information to enrich the descriptors. SIFT descriptor with global content in [2] adds curvilinear shape information from a much larger neighbor.

Other methods use global information in matching process. Sample consensus methods such as RANSAC [3] and PROSAC [4] estimate the parameters of a prior geometric consistency model by sample correspondences selected from two images. But RANSAC and PROSAC can not work well with images with repetitive patterns, because of too many outliers. In [5], reinforcement matching scheme are employed on the affine-invariant log-polar elliptical bin. Relaxation method [6] is a probabilistic matching framework which iteratively update initial probabilities base on a compatibility function.

Leordeanu and Hebert [1] present a spectral method for finding consistent correspondences between two feature sets. Our method can be categorized as pairwise constraint method. Our method differs from previous method as follow: (1) We introduce pairwise geometric constraint based on local features to build the correspondence graph; (3) We develop an efficient approximation heuristic algorithm to match multiple objects with repetitive patterns by finding the maximal cliques in graph.

3 Maximal Clique Matching

3.1 Geometric Correspondence Graph

Given a image and a target image, let $X = \{x_1,...,x_n\}$ and $Y = \{y_1,...,y_m\}$ denote two sets of feature points extracted from two images. Let define an undirected and unweighted graph $G = (V, E)$ as GCG, where $V = \{v_1,...,v_k\}$ is the vertex set of G, and $E \leq V \times V$ is the edge set of G. Let $N(v) = \{ j \in V : e_{vj} = 1 \}$ denote the neighborhood of v in G.

For the local features, the spatial location, scale and rotation are the local geometric information of an individual feature point. In our method, relative orientation and relative distance are used to describe the local geometric relationship between two points in the same image, as shown in Figure 3. x_1 and x_2 are two feature points in the left image. α_1 and α_2 are the relative orientations. d_1/r_{11} and d_1/r_{12} are the relative distances.

Then we use the semi-local geometric relationship to evaluate the pairwise geometric constraint between two correspondences. The pairwise geometric consistency should obey the following rules:

$$\left|(\alpha_1-\alpha_2)-(\beta_1-\beta_2)\right|<\varepsilon_1 \qquad (1)$$

$$1/\varepsilon_2<\left(d_1/r_{11}\right)/\left(d_2/r_{21}\right)<\varepsilon_2 \qquad (2)$$

$$1/\varepsilon_2<\left(d_1/r_{12}\right)/\left(d_2/r_{22}\right)<\varepsilon_2 \qquad (3)$$

Since the local geometric information is not stable enough for deformation, the threshold ε_1 and ε_2 should be high enough that the most correct matches are kept. Another two specific cases should be concerned. First, in the case that two features have the same location but different orientation, the spatial distance is zero. So (2) and (3) do not need to be compute. Second, two correspondences may have the same point in a certain image. In this case, those two correspondences definitely are not consistent with the same weak geometric constraint.

Then we give the relationship between pairwise geometric constraint and GCG. In GCG, each vertex represents a matching correspondence. We define the projections as follow: $F:(x_i,y_j)\mapsto v_k$, $R:x_i\mapsto v_k$, $Q_X:v_k\mapsto x_i$ and $Q_Y:v_k\mapsto y_j$. F is bijection. R is injection. Q_X and Q_Y are surjection.

The vertices will be adjacent only if the correspondences those vertices represent are weak geometric consistent.

$$e_{ij}=\begin{cases}1, & if\ (1),(2)\\ 0, & otherwise\end{cases} \qquad (4)$$

The computational complexity of building graph is $O(k^2n^2)$, where n is the number of points in the query image and k is the number in KNN.

3.2 Finding Maximal Cliques

After GCG is built, the next step is to fill out the outliers and group the inliers by the global geometric consistency relationship. In GCG, the correct correspondences are likely adjacent to each other and form a large maximal clique. But incorrect ones establish links with others only accidentally. So they are likely in small maximal cliques. So the problem of finding all the correspondences consistent to the same global geometric constraint is equivalent to the problem of finding all the large maximal cliques in a graph.

Though finding the maximal cliques in a graph is an NP complete problem, many efforts have recently been directed to devising efficient heuristics algorithm. Algorithm *Best-in* is a greedy local search heuristic algorithm which is first proposed in [7]. The *Best-in* heuristics generate a maximum clique through the repeated addition of a vertex into a partial clique. A possible *Best-in* heuristic constructs a maximum clique by repeatedly adding a vertex that has the largest degree among the candidate vertices. In a graph, only those vertices that are the neighbors of all the vertices chosen before are probable in the maximal clique. So we can choose those vertices which are the neighbors of all the vertices chosen before to be the candidates.

In this application, there are two issues that differ from the heuristic algorithms. First, one-to-multiple and multiple-to-multiple object matches are required. So the algorithm has to find all the maximal cliques in the graph. For example, in Figure 2, Clique 1 and Clique 3 represent two objects. They have to be accepted. Second, in some images with repetitive patterns, such as chessboard or buildings, all the points are similar to each other. The subset of the points in query image may match another subset of the points in target image, though all the correspondences are incorrect. In Figure 2, Clique 2 is an example. Clique 2 has to be rejected.

We propose a novel algorithm *AllMaximalCliques* based on algorithm *Best-in* to find all the maximal cliques in the graph. To find all the large maximal cliques, we call *Best-in* to get the maximum cliques and delete some vertices in the graph iteratively. To avoid the second issue happening, we delete the vertices whose both points have already matched. The *AllMaximalCliques* is summarized in Table 1.

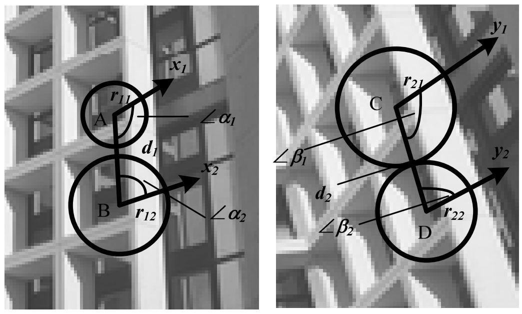

Figure 1. An example of the geometric weak constraint. x_1 and x_2 are two features in query image. y_1 and y_2 are two features in target image. r_{11}, r_{12}, r_{21} and r_{22} are the size of the scale. d_1 and d_2 are the spatial distance between two features in the same image.

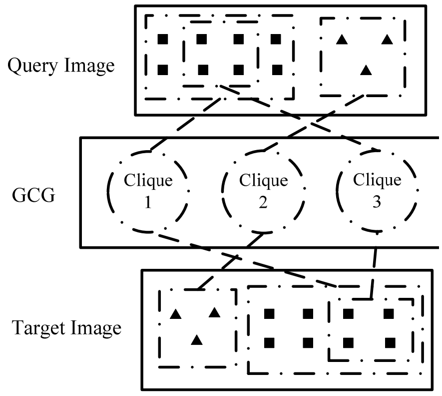

Figure 2. An example of multiple objects matching using GCG.

1. $Match_X=\varnothing$, $Match_Y=\varnothing$, $Matches=\varnothing$.
2. Call *Best-in* to find the maximum clique C in G.
3. If the $deg(C)<n$, finish searching the maximal cliques. Otherwise, continue.
4. $Matches=Matches\cup C$,
   $Match_X=Match_X\cup Q_X(C)$,
   $Match_Y=Match_Y\cup Q_Y(C)$.
5. $G=G\setminus\left(R\left(Match_X\right)\right)\cap\left(R\left(Match_Y\right)\right)$
6. Go to step 2.

Table 1. *AllMaximalCliques* Algorithm

Although our algorithm is an approximation algorithm, it can work well because the maximal cliques are always dominant in the application of local feature matching. In our experiment, we found even the correct correspondence is far below 1/10 among all, they can still be found.

The complexity of finding a maximal clique in the graph is $O(mk^2n^2)$ [21]. m is the number of all the maximal cliques. But n is far more than m and k. The computational complexity of our methods is in the same level $O(n^2)$ as that of some other up-to-date methods [5, 6].

4 Experiment

Data Set: We use two datasets: (1) A public available dataset[1] which is broadly used in testing local feature in typical scenes and transformations. The images are presented in Figure 3. (2) A datasets of image pairs with repetitive patterns proposed in [8]. The images are presented in Figure 4.

Other Methods: We compare our method with three methods: NNDR [10], REINF (reinforcement matching) [5] and RELAX (relaxation matching) [6]. We choose NNDR to show the distinctiveness of the descriptor alone.

Interest Points: We choose Harris-Affine [9] as detector and SIFT [10] as descriptor for better robustness to deformation. The Harris-Affine and SIFT codes are both implemented by Mikolajczyk which are provided online[2].

Evaluation Criterion: Correct matches are detected based on the homographies between the images. A couple of correspondence points (p, p') is said to be correct only if $|p'-Hp|<5$, where H is the transformation between two images.

In the first experiment, we choose four sequences from dataset which present two different scene types: structured (graffiti, boat) and textured (wall, bark), and three different transformations: viewpoint change (graffiti, wall), image rotation and scale change (boat, bark). In each sequence, we choose the first image to match the following four images with growing transformation.

In [5], the experiments show that REINF provide better matching rate than RANSAC and PROSAC. In [6], RELAX shows the great advantage than previous methods.

The experimental results of NNDR, REINF and RELAX are from [8]. We use the same code for local feature extraction. We also use the same criterion to evaluate the results. So our experimental results can be compared with the results of those methods.

The results are shown in Table 2. For 16 pairs of images, MCM gives more matches while the matching rate is superior or equal to that of all the other methods. For the other 3 pairs of images, MCM is the second among four methods, and gives only a few matches less than RELAX. Only for one pair of images, MCM gives the worst result.

Scene types: For textured scenes, all four methods can get high precision. Only a little improvement is obtained with our method. But for structured scenes, our method shows the great advantage in comparison with other methods, because the geometric relationship is more significant for structured scenes.

[1] http://www.robots.ox.ac.uk/~vgg/research/affine/

[2] http://www.robots.ox.ac.uk/~vgg/research/affine/

Transformations: Harris-Affine and SIFT are more robust to rotation and scale changes than to viewpoint change, so the global geometric information becomes more important for viewpoint change. That is why the improvement obtained by our method is greater for viewpoint changes than for rotation and scale changes.

We test on the Quad-Core AMD Opteron 800 MHz CPU. We keep 300 points with top cornerness in each image. The average matching time for a pair of images is 0.034 sec.

In the second experiment, we choose the same images and evaluation criterion as the experiments in [22]. The results of other methods are also from [22]. The results are shown in Table 3.

Our method gives both higher recall and precision for all four pairs of images. Especially for the structured scenes, our method outperforms other methods greatly. The average matching time for a pair of images is 1.25 sec. There are near 2000 feature points in every image.

5 Conclusion

This paper has proposed a novel graph-based geometric constraint method. The method is robust to repetitive patterns. The scheme expresses the geometric constraint problem as the maximal clique problem in a graph. This method utilizes a geometric correspondence graph based upon the pairwise geometric constraint information in local feature and an efficient approximation heuristic algorithm for finding the maximal cliques in GCG. We have made experiments for two different datasets. In comparison with several up-to-date methods, our method shows the better results and comparative computational complexity. The average matching time for a pair of images is 0.034 sec in 300 feature matching.

Acknowledgment

This work was supported by 973 Program of China with grant 2010CB731406,2010CB731401, and the 111 Project B07022.

References

[1] M. Leordeanu, M. Hebert: A Spectral Technique for Correspondence Problems Using Pairwise Constraints. ICCV 2 (2005) 1482-1489

[2] E. N. Mortensen, H. Deng, L. Shapiro, A Sift Descriptor with Global Context, In *CVPR*, 2005.

[3] M. A. Fischler, R. C. Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *CACM*, 24(6): 381–395, 1981.

[4] O. Chum and J. Matas, Matching with PROSAC Progressive Sample Consensus, In *CVPR*, 2005.

[5] H. Deng, E. N. Mortensen, L. Shapiro, and T. G. Dietterich. Reinforcement matching using region context. In *CVPR Workshop*, 2006.

[6] D. Sidibe, P. Montesinos, and S. Janaqi. Fast and robust image matching using contextual information and relaxation. In *VISAPP*, 2007.

[7] D. S. Johnson. Approximation algorithms for combinational problems. In *JCSS*, 1974.

[8] D. Sidibe, P. Montesinos, and S. Janaqi. Matching Local Invariant Features with Contextual Information: An Experimental Evaluation. *ELCVIA* 7(1):26-39, 2008

[9] K. Mikolajcyk. and C. Schmid. An affine invariant interest point detector. In ICCV. 2002.
[10] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 2(60):91–110, 2004.

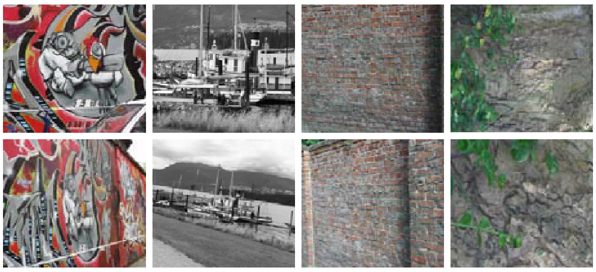

Figure 3. Test images for baseline matching. From left to right: Graffiti, Boat, Wall and Bark.

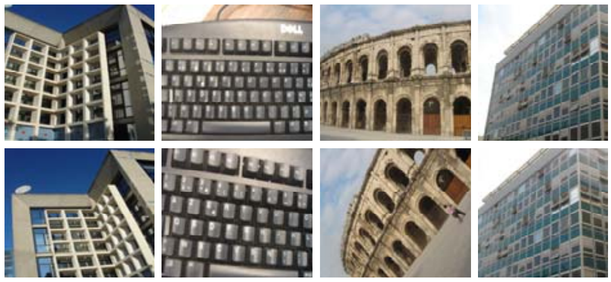

Figure 4. Test images for repetitive pattern matching. From left to right: Office, Keyboard, Arenas and Building.

Image	id	NNDR		REINF		RELAX		MCM	
		M	P	M	P	M	P	M	P
Graf	2	82	0.96	93	0.96	115	0.92	129	1
	3	58	0.4	68	0.4	40	0.67	70	1
	4	23	0.35	24	0.5	24	0.5	28	0.61
	5	13	0.08	13	0.08	8	0	10	0.2
Boat	2	70	0.98	91	0.98	150	0.98	145	1
	3	75	0.95	98	0.95	131	0.98	126	0.98
	4	22	0.88	29	0.88	34	0.92	50	0.92
	5	19	0.99	22	0.99	24	0.99	42	0.94
Wall	2	82	0.96	93	0.96	115	0.92	129	1
	3	58	0.4	68	0.4	40	0.67	70	1
	4	23	0.35	24	0.5	24	0.5	28	0.61
	5	13	0.08	13	0.08	8	0	10	0.2
Bark	2	35	0.97	34	0.97	52	0.98	71	1
	3	17	0.88	24	0.87	31	0.93	35	1
	4	3	1	8	1	9	0.9	17	0.88
	5	11	0.99	14	0.99	15	0.8	9	1

Table 2. Result of the first experiment. M means the number of matching correspondences. P means precision.

Scene	Method	M	R	P
Office (structured scenes)	NNDR	6	0.5	0.064
	REINF	16	0.5	0.17
	RELAX	38	0.66	0.53
	MCM	51	0.88	0.96
Keyboard (structured scenes)	NNDR	18	0.44	0.1
	REINF	18	0.44	0.1
	RELAX	40	0.66	0.37
	MCM	50	0.98	0.69
Arenas (textured scenes)	NNDR	353	0.94	0.48
	REINF	347	0.96	0.47
	RELAX	568	0.98	0.79
	MCM	627	1	0.89
Building (textured scenes)	NNDR	276	0.91	0.44
	REINF	300	0.92	0.48
	RELAX	420	0.98	0.72
	MCM	519	1	0.91

Table 3. Results of the repetitive pattern image matching. R means recall.