MATCHING IMAGES MORE EFFICIENTLY WITH LOCAL DESCRIPTORS

Dong Zhang[1] Weiqiang Wang[1,2] Qingming Huang[1,2] Shuqiang Jiang[2] Wen Gao[2,3]
[1]*Graduate University of Chinese Academy of Sciences, Beijing, China*
[2]*Key Lab of Intell. Info. Process., Inst. of Comput. Tech., Chin. Acad. of Sci., Beijing, China*
[3]*Institute of Digital Media, Peking University, Beijing, China*
E-mail: {dzhang, wqwang, qmhuang, sqjiang, wgao}@jdl.ac.cn

Abstract

Image matching is a fundamental task for many applications of computer vision. Today it is very popular to represent two matched images as two bags of local descriptors, and the classic RANSAC based matching procedure is always exploited in the task. In this paper, we present a much efficient image matching approach based on sets of any local descriptors. A block-to-block strategy is devised to speed up the establishment of local correspondences. Additionally, the weighted RANSAC (w-RANSAC) technique is proposed to make the search of optimal global models converge faster. Comparative experiments with the RANSAC based paradigm show our approach can not only generate more accurate correspondences, but also double the matching speed.

1. Introduction

Image matching aims to automatically identify whether two digital images can establish consistent correspondences between primitives extracted from them. It is a fundamental technique in many applications of computer vision, e.g., object recognition, motion tracking, 3D reconstruction. The difficulties for image matching come from possible geometric transformations, illumination changes, as well as viewpoint variations between two images. The early correlation methods by comparing intensities in a window neighborhood are seldom used today, since they are sensitive to illumination changes, and geometric scaling, etc. Instead researchers investigate feature based approaches, such as keypoints [1, 2, 3], edges [4], regions [5, 6] etc. The keypoint based approaches are comprehensively investigated and become very popular today in vision communities, since many state-of-the-art local invariant descriptors are very robust to geometric transformation, as well as illumination and viewpoint change to some extent. In the paradigm, informative points called keypoints in two images are first located, and the neighborhood patches around the keypoints are represented by local invariant descriptors, such as geometric histogram [7], shape context [8], PCA-SIFT [9], SIFT [10]. Then correspondences between two bags of local descriptors representing two images are computed. It generally involves two stages: local matching and global matching. The local matching searches the initial correspondence for each keypoint only based on the distance of two descriptors in the feature space. Finally the initial correspondences are checked and verified through a global epipolar constraint. The global epipolar constraint model is estimated from the initial keypoint correspondences with possible noises by the algorithm called RANdom SAmple Consensus (RANSAC) algorithm [11, 12].

In this paper, we present a much more efficient image matching algorithm based on two bags of any local keypoint descriptors in the classic framework mentioned above. Specifically, we devise a block-to-block strategy to speed up the establishment of local correspondences by avoiding the exhaustive search way. Additionally, we present a weighted RANSAC (*w*-RANSAC) technique which makes the search procedure of optimal global models converge faster.

The paper is organized as follows. Section 2 presents our image matching approach in details. Section 3 reports the experimental results. Section 4 concludes the paper.

2. Our Image Matching Approach

We first briefly introduce the local invariant descriptor used in our matching approach in subsection

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

2.1 and the classic image matching method based on two bags of local descriptors in subsection 2.2. Then more descriptions are given to the block-to–block strategy for local matching in subsection 2.3 and the w-RANSAC technique in subsection 2.4.

2.1. PCA c-SIFT local descriptor

We present a hybrid local descriptor which integrates color and luminance information in [13]. Color is characterized by three color invariants *rgb*, hue, and $l_1 l_2 l_3$, and luminance is represented by the SIFT descriptor [9]. Different from the SIFT descriptor, the circular local patch is used and partitioned into three annular sub-regions for extracting color descriptor. For each annular sub-region, a weighted histogram for color features is evaluated. The complete descriptor is composed by joining the SIFT descriptor S to the color descriptor C weighted with a factor of λ, i.e. $(S, \lambda * C)$. Finally the principal component analysis (PCA) is further exploited to make the hybrid descriptor more compact, since the three color invariants may carry redundant correlated information. We call the generative descriptor as PCA c-SIFT descriptor.

2.2. Classic Image matching paradigm based on local invariant descriptors

In the image matching paradigm, each image is represented by a bag of local descriptors. For two images to be matched, we call one of them as a query image $I_q = \{q_i \mid i = 1, 2, ..., M\}$, and the other as a reference image $I_r = \{r_j \mid j = 1, 2, ..., N\}$, where q_i, r_j denote their corresponding local descriptors. In the classic paradigm, an initial matching procedure first starts up to establish point-to-point correspondences. In practice, the similarities for each q_i with each r_j are computed through an appropriate dissimilarity function $\varphi(q_i, r_j)$, and an candidate correspondence r_{c_i} in I_r for q_i is selected according to

$$c_i = \arg\min_j \varphi(q_i, r_j) \qquad (1)$$

If $\varphi(q_i, r_{c_i})$ is below a given threshold, the correspondence (q_i, r_{c_i}) between q_i and r_{c_i} is established, and otherwise no correspondence exists for q_i.

The results of initial matching may contain a few false correspondences called outliers. Then the RANSAC algorithm can be used to remove the noisy correspondences. The complete algorithm procedure is shown in Fig.1. In the step (c), the epipolar line l_r associated with p_r can be evaluated by $l_r = F p_r$. According to the epipolar constraint, the point p_q should lie on the epipolar l_r ($d_{q,r}$ =0) ideally, if (p_q, p_r) is a true correspondence. In practice, if the distance $d_{q,r}$ from p_q to l_r is small enough, the correspondence (p_q, p_r) is asserted. The distance $d_{q,r}$ is evaluated by

$$d_{q,r} = \sqrt{(l_r^T p_q)^2 / (l_{rx}^2 + l_{ry}^2)}, \qquad (2)$$

where l_r is represented by the vector $(l_{rx}, l_{ry}, l_{rz})^T$.

Input:
- G - the set of the correspondences generated by the initial matching procedure for two images I_q, I_r
- k - the number of iterations required
- t - the threshold used to identify a correspondence that fits the model well
- ρ - the number of correspondences required to assert two images are matched

Output: two images I_q, I_r are matched or not

Until k iterations have occurred
- (a) Draw a sample of 7 correspondences from the input set G of correspondences uniformly and at random
- (b) Compute fundamental matrix F by the 7-points algorithm
- (c) For each correspondence (p_q, p_r) except the sample,
  - Test the distance $d_{q,r}$ from point p_q to the epipolar line l_r associated with p_r.
  - If $d_{q,r} < t$, the correspondences is consistent.
- (d) If there are ρ or more consistent correspondences, assert two images are matched and the algorithm terminate

End

Assert two images are not matched.

Fig.1. image matching paradigm via RANSAC

2.3. Block-to-block strategy in local matching

Apparently, the similarity evaluations of $M * N$ times are involved in the classic paradigm. The block-to-block strategy (B2B) devised in our matching approach aims to reduce the times of such similarity evaluations through establishing the correspondence between region blocks.

Both query image I_q and reference image I_r are first partitioned into many small blocks. For instance, 8 by 8 rectangle blocks are used in our implementation.

The centroid of keypoints in each block is first evaluated. If the distance of two centroids for two neighboring blocks is lower than one third of their total edge length in alignment, the two blocks are merged into a bigger block, and the new centroid of keypoints in the generative block is evaluated for the next possible block mergence. When the iterative mergence procedure terminates, both images are partitioned into a sequence of region blocks. The simple mergence procedure can run in a much more efficient way than any clustering algorithm for keypoints,

Suppose B_i^q , $i = 1, 2, ..., n_q$ denote the blocks for query image I_q , and $B_j^r, j = 1, 2, ..., n_r$ for reference image I_r . For each block B_i^q , a very small portion of the keypoints in it (e.g., 5%) is randomly chosen as representatives. The correspondences in I_r for the representative keypoints are located through the exhaustive search way. Then the number of correspondences fell into each B_j^r , $j = 1, 2, ..., n_r$, is counted. The block in I_r with the maximum votes is labeled as the correspondence block for B_i^q , and is denoted as $B_{c_i}^r$.

Once the correspondence block in I_r for B_i^q is identified, the correspondences for the remaining keypoints in B_i^q is only searched in the keypoints belonging to the correspondence block $B_{c_i}^r$. Suppose block B_i^q , $i = 1, 2, ..., n_q$, in query image I_q contains m_i keypoints, $\sum_{i=1}^{n_q} m_i = M$, and its correspondence block $B_{c_i}^r$ contains n_{c_i} keypoints. If the percentage of representative keypoints is ζ , the initial matching procedure using the block-to-block strategy only involves dissimilarity evaluations of $\zeta MN + (1-\zeta)\sum_{i=1}^{n_q} m_i n_{c_i}$ times or so.

2.4. Weighted RANSAC

The results of initial matching generally contain some outliers (false correspondences). If the outliers are sampled to estimate fundamental matrix F , a wrong model is prone to be generated. Although the RANSAC algorithm exploits multiple trials as a robust mean to obtain a right model, anyone hopes the RANSAC algorithm could find a right model very soon, i.e., in a more efficient way. Yet in the classic image matching paradigm, the outliers have equal opportunities as true correspondences to be sampled. Apparently we expect a group of true correspondences can be sampled in very few cycles, so that the RANSAC algorithm can converge as soon as possible. Intuitively, keypoint pairs with a larger similarity are more probably true correspondences. So similarities of two keypoints in a correspondence provide useful apriori knowledge to direct the procedure of data sampling. Concretely, we associate each keypoint pair with a sampling probability p_k , $k = 1, 2, ..., L$, according to the Euclidean distance d_k of two keypoints, i.e.,

$$p_k = e^{-d_k * d_k} / \sum_{l=0}^{L} e^{-d_l * d_l} \qquad (3)$$

Thus, those correspondences with larger similarities have higher probabilities to be sampled, and a right model is more likely to be found within a limited number of cycles correspondingly. We call the modified RANSAC as the weighted RANSAC (w-RANSAC), in which weights associated with data points reflect the reliability or confidences of them for using them to estimating the global model.

3. Evaluation experiments

In this section, we evaluate the performance of our image matching approach by comparing it with the classic image matching paradigm described in subsection 2.2. The dataset consists of three groups of images from the publicly available INRIA dataset. The INRIA dataset contains many groups of images, and each group corresponds to a kind of transformation, e.g., blur, viewpoints, zoom, rotation, light, etc. Each group of images includes a reference image and multiple transformed images. The image data used in our experiments are boat (zoom plus rotation), graffiti (viewpoints) and cars (light). The performance is evaluated from two aspects: accuracy of correspondences and matching efficiency. The accuracy of correspondences is defined as the ratio of true correspondences (inliers) to total correspondences generated. The INRIA dataset provides homographies H from reference images to transformed images for each group of images. Thus, for a keyppoint p_q in reference images, we can obtain its ground-truth correspondence p_x in transformed images by $p_x = Hp_q$. In our experiments, if the Euclidean distance between correspondence p_r generated by the algorithms for p_q and the ground-truth p_x is less than a predefined threshold (e.g., 2.5 pixels), the system asserts (p_q , p_r) as a true correspondence (inlier). The matching efficiency is measured by the running time of

the related functions on a mobile computer with 1.4 GHz Intel Celeron M and 512MB main memory.

In the experiments, both approaches use the Euclidean distance of two PCA c-SIFT descriptors to reflect their dissimilarity, and the threshold is 0.42. For the RANSAC algorithm, k =50, t =2 and ρ =0.5*| G | respectively. Table 2 summarizes all the experimental results, including average time of local matching (**L**), average time of global matching (**G**), and average inliers/outliers(**IO**) . It can be observed that our approach generates nearly the same inliers as the classic paradigm but fewer outliers. It implies that the block-to-block strategy can eliminate false matches to some extent. For example, for a keypoint p_q in transformed images, if a more similar point p_r does not locate in the correspondence block of reference images, p_q and p_r cannot form a valid correspondence due to the block-to-block constraint. Additionally, the related results show that the block-to-block strategy also makes the initial matching procedure become more efficient than the counterpart in the classic matching paradigm. The experimental results also show that *w*-RANSAC algorithm doubles the convergence speed of the original counterpart. The block-to-block strategy is the main contributor to the speedup of matching procedure. From Table 1, we can easily compute the accuracy of correspondences 98.6% for our method and 96.0% for the classic paradigm. At the same time, the ratio of overall matching time for our method to the classic paradigm is 0.526. Thus our method is very efficient and does not degrade the accuracy of correspondence at the same time.

Table 1: Experimental results for the classic image matching paradigm vs. our approach.

Group ID		boat	graffiti	cars
L (ms)	Classic	1249	1009	458
	B2B	523	613	292
G (ms)	Classic	32	29	33
	w-RANSAC	17	15	17
IO	Classic	129/10	151/13	535/11
	Our method	118/3	138/5	514/3

4. Conclusion

In this paper, we present a new much efficient image matching approach which can be applied to any local keypoint descriptors. The experimental results show that the devised block-to-block matching strategy and the proposed *w*-RANSAC technique can not only generate more accurate correspondences, but also make the matching procedure more efficient. Combining the novel techniques with stat-of-the art keypoint detectors and descriptors is helpful to produce more effective and efficient solution to the issue of image matching.

Acknowledgements

The work is supported by the research start-up fund of GUCAS and by National Key Technologies R&D Program under Grant 2006BAH02A24-2.

5. References

[1] C. Harris, M. Stephens: "A combined corner and edge detector." *Fourth Alvey Vision Conference, Manchester*, UK, pp.147-151, 1988.

[2] A. Baumberg: "Reliable Feature Matching across Widely Separated Views." *IEEE Conference on Computer Vision and Pattern Recognition*, pp.774–781, 2000.

[3] K. Mikolajczyk, C. Schmid: "Indexing Based on Scale Invariant Interest Points." *IEEE International Conference on Computer Vision*, pp.525–531, 2001.

[4] W. Förstner: "Hierarchical Chamfer Matching: "A Parametric Edge Matching Algorithm." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.10, pp.849-865, 1988.

[5] T. Tuytelaars, L. Van Gool: "Wide Baseline Stereo Matching Based on Local, Affinely Invariant Regions." *British Machine Vision Conference*, pp.412–425, 2000.

[6] J. Matas, O. Chum, M. Urban, and T. Pajdla: "Robust wide baseline stereo from maximally stable extremal regions." *British Machine Vision Conference*, pp.384–393, 2002.

[7] A. Ashbrook, N. Thacker, P. Rockett, C. Brown: "Robust Recognition of Scaled Shapes Using Pairwise Geometric Histograms." *Proceedings of the 6th British conference on Machine vision*, Vol.2, pp.503-512, 1995.

[8] S. Belongie, J. Malik, J. Puzicha: "Shape Matching and Object Recognition Using Shape Contexts." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.24, pp.509-522, 2002.

[9] D.G. Lowe: "Distinctive image features from scale invariant keypoints". *International Journal of Computer Vision*, 60(2): pp.91-110, 2004.

[10] Y. Ke and R. Sukthankar: "PCA-SIFT: A more Distinctive Representation for Local Image Descriptors." *IEEE Conference on Computer Vision and Pattern Recognition*, Vol.2, pp.506-513, 2004.

[11] D.A. Forsyth, J. Ponce: *Computer Vision: A Modern Approach*. Electronic Industry Press, June, 2004.

[12] L. Qin, W. Gao: "Image Matching Based on a Local Invariant Descriptor." *International Conference on Image Processing*, Vol.3, pp.377-380, 2005.

[13] D. Zhang, W.Q. Wang, W. Gao: "An Effective Local Invariant Descriptor Combining Luminance and Color." *International Conference on Multimedia and Expro.*, pp.1507-1510, 2007