Finding the correspondence points in images of multi-views

Yau, Wei-Yun; Teoh, Eam Khwang; Anvar, Seyed Mohammad Hassan

2012

Anvar, S. M. H., Yau, W.-Y., & Teoh, E. K. (2012). Finding the Correspondence Points in Images of Multi-Views. 2012 Eighth International Conference on Signal Image Technology and Internet Based Systems (SITIS), 275-280.

<https://hdl.handle.net/10356/79904>

<https://doi.org/10.1109/SITIS.2012.48>

© 2012 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at: [<http://dx.doi.org/10.1109/SITIS.2012.48>].

Downloaded on 26 Apr 2023 17:58:32 SGT

Finding the Correspondence Points in Images of Multi-Views

Seyed Mohammad Hassan ANVAR

School of Electrical & Electronic Engineering Institute for Infocomm Research
Nanyang Technological University
Singapore
seye0005@e.ntu.edu.sg

Wei-Yun YAU

*A*STAR*
Singapore
wyyau@i2r.a-star.edu.sg

Eam Khwang TEOH

School of EEE
Nanyang Technological University
Singapore
eekteoh@ntu.edu.sg

Abstract—Recently, many studies have been conducted to find the correspondence points between two images. However they are useful if they employed in the application for finding the correspondences between two images of the same scene and they fail if the number of true matches between two images was in small quantity in compare with all the potential correspondence points found. In this paper, we have presented an approach which uses both the geometry and appearance of features and utilizes the discriminative features to find the real correspondence points even if the number of true matches was small portion of all the correspondence points found between two comparing images. The experimental results show that the proposed method also stabilizes very fast where other methods are still searching for the correspondences.

Keywords—matching method; correspondence points; discriminative features; geometric constraint;

I. INTRODUCTION

Finding the real correspondence points between two images of different views has been in demand in many computer vision applications. There are many matching techniques have been introduced during the past decade [1], [2], [3], [4], [5], [6]. Among them, the method proposed by Mikolajczyk and Schmid [5] and the one introduced by Lowe [6] are very famous. Mikolajczyk and Schmid [5] proposed to find the correspondence points between two images of similar scene using scale and affine invariant interest point detector based on Harris detector [7]. They used cross correlation to discard the low score matches. Lowe [6] proposed another matching method based on scale invariant features. In his method, after finding keypoints in two comparing images, the appearance descriptors from features of the first image are compared to the appearance descriptors of the second image. Finally, the nearest neighbor of each feature's descriptor is considered as the best match if its distance was significantly small in compare with the distance of the second closest match. These matching algorithms are not completely perfect and they return many incorrect correspondence points. Thus, a refinement method should be used to reject outliers and select the real correspondence points between two images. Typically, the optical flow between two comparing images is considered for matching refinement and a method such as RANSAC method [8] or Hough transform [9] is performed to find the parameters of the optical flow. Hough transform is useful only when we have a model of object and we desire to refine the matches

between the model and the enquiry image. RANSAC is more general than Hough transform and it has been used widely in many computer vision applications; however, it has some deficiencies which make it useless in some case studies. Although many studies have been conducted to solve RANSAC issues [10], [11], [12], it fails when the number of real correspondence points is small in compare with the all matches found between two images [5]. Thus, any matching technique based on RANSAC is not able to find real correspondence points between the pair of images with major number of dissimilarities. This problem intensifies when the number of real correspondence are very small in compare with all the potential correspondence points found between two images. The main issue of these matching and refinement techniques is that they all use only appearance descriptor to find the correspondence points. Dorko and Schmid [13] have shown that the appearance of a descriptor from a patch of object could have many similarities to the one from noise and background area if the descriptor was not discriminative. Therefore, other constraints should be considered to find the reliable matching points between image pairs. In this paper we have presented an approach which uses appearance and geometry information of matched points besides considering the more discriminative features to find the real correspondence points between comparing images.

II. PROPOSED MATCHING FRAMEWORK

In the proposed approach, interested points and their descriptors are extracted for two comparing images using scale invariant features. Common appearance features are discarded by computing likelihood ratio for each feature. Lowe's matching approach [6] is applied to these features in order to find the potential correspondence points between these two images. After that, the geometric registration is applied to find the real match points among image pairs. We have explained these steps with more detail in the following sections.

A. Distinctive Features and Common Features

All scale invariant features extracted by Lowe's method [6] are not informative and may be derived from noise or background clutter. Only a few features can represent the object very well. Unfortunately, those non-distinctive features impress the informative features so much and change

the overall result of matching dramatically. In this section, we aim to find the most informative features, which are expected to be the most discriminative parts of an object. It is very likely that one feature from a part of object and a feature from a part of clutter background emerge the same appearance [13]. It indicates that this appearance is not so discriminative to represent the object patches. Typically, a reliable feature's appearance is the one which represents a patch of that object with high probability. Let feature f_i has been extracted from an image which includes the object of interest. Two hypotheses exist. First, the feature f_i comes from a patch of object ($f_i = 1$) or the other hypothesis that it is from background ($f_i = 0$). In order to determine these hypotheses we need to evaluate the likelihood ratio test $\psi(f_i)$ in (1) for this feature. If the obtained value in this test for the feature f_i was greater than one, it says the probability that this feature comes from a patch of the interest object is higher than it comes from background clutter.

$$\psi(f_i) = \frac{p(f_i = 0|\Gamma)}{p(f_i = 1|\Gamma)} \quad (1)$$

where Γ is some observations on feature f_i . To evaluate this ratio we have used Maximum Likelihood Estimation (MLE) as follows: Let feature f_i be a binary random variable that might be originated from a part of object ($f_i = 1$) or non-object ($f_i = 0$) area. We can assign a Bernoulli distribution with parameter θ to it that determines its uncertain value as it is shown in (2).

$$p(f_i|\theta) = \theta^{f_i} (1 - \theta)^{1-f_i} \quad (2)$$

Let we have a sequence of N observation Γ over the feature f_i in the training data. The likelihood for the given Γ is obtained by (3) assuming that the observations are independent.

$$p(\Gamma|\theta) = \prod_{j=1}^N \theta^{f_i^j} (1 - \theta)^{1-f_i^j} = \theta^M (1 - \theta)^{(N-M)} \quad (3)$$

where M is the number of times that feature f_i has represented the object area in the observation Γ . When we have sufficient number of observation data or N is a large number, we can estimate the distribution parameter θ using MLE estimation in (4).

$$\hat{\theta}^{MLE} = \underset{\theta}{\operatorname{argmax}}(p(\Gamma|\theta)) \quad (4)$$

We assess the maximum likelihood estimate (MLE) to estimate the probability in (1) as follows: Let feature $f_{(1,i)}$ be a feature from the first image I_1 which compared to all N_2 features in the second image I_2 . If this feature was a distinctive feature of the object, it should be unique and the occurrence of it in the other part of the object would be very limited. Ideally, there would only be one such feature and for symmetrical parts, only two features can be found. Thus, we can estimate its probability as $p(f_{(1,i)} = 1|\Gamma) = \frac{2}{N_2}$. But if this feature originated from the common appearance parts of the image, the number of similarities found for it could be any number, β . Thus,

we can estimate its probability as $p(f_{(1,i)} = 0|\Gamma) = \frac{\beta}{N_2}$. Substituting these estimations into (1) indicates that all features with similarity number (β) greater than 2 can be regarded as common appearance features and tagged as such. To compute the similarity number of feature $f_{(1,i)}$ in the first image, its appearance descriptor $f_{(1,i)}^A$ is compared to the appearance descriptors of all features in the second image $f_{(2,j)}^A$ and $j \in (1, \dots, N_2)$, within a global threshold $((f_{(1,i)}^A - f_{(2,j)}^A)^2 < T^a)$. This procedure is then repeated for each feature found in the second image, which is then compared to all the features in the first image. Finally, all common features found are removed. To find an optimum value for T^a several image pairs from different databases such as FERET have been investigated. We have observed that a value between 0.2-0.3 is an optimum value for T^a . A greater value yields elimination of the distinctive features whereas a lower threshold lets noise and common features to appear in the correspondence points. Since this threshold corresponds to features appearance descriptor, its value is fixed as long as the features have been extracted by SIFT method in descriptor has 128 bins (standard SIFT feature).

B. Finding Real Correspondence Points

Without discarding common features, correspondence pairs from non-object parts may occur in the matching results. Thus, these undesired points yield unreliable correspondence points among image pairs. Once the common features are reduced from the extracted features, Lowe's feature matching method [6] is applied to them. This amendment not only removes the common features, but also results some distinctive parts of object to be revealed in the matching procedure that were not detected before. Lowe's method returns many potential correspondence points between two images. Assume that P potential correspondence points have been found between two images. Let $\{f_{(1,k)}\} \overset{k}{\leftrightarrow} \{f_{(2,k)}\}$ and $k \in (1, \dots, P)$ be the P potential matched pairs found between feature $f_{(1,k)}$ from the first image I_1 and feature $f_{(2,k)}$ from the second image I_2 . Each feature has a geometric information includes position, scale and orientation, like the vector $f_{1,k}^G = (X_{1,k}, Y_{1,k}, \sigma_{1,k}, \phi_{1,k})$ for features in the first image I_1 and the vector $f_{2,k}^G = (X_{2,k}, Y_{2,k}, \sigma_{2,k}, \phi_{2,k})$ for the features in the second image I_2 . To find reliable correspondence points between image pair we need to define a reference point in each image. First we consider pair $\{f_{(1,k)}\} \overset{k=1}{\leftrightarrow} \{f_{(2,k)}\}$ to be a potential reference point for its associated image. The geometry of all the other matched pairs $\{f_{(1,r)}\} \overset{r}{\leftrightarrow} \{f_{(2,r)}\}$ and $r \in (1, \dots, P) \& r \neq 1$ are normalized to the scale, position and orientation of the potential reference point in that image using the affine transformation. Those matched pairs which overlay each other after normalization are considered as members for the potential reference points $\{f_{(1,k)}\} \overset{k=1}{\leftrightarrow} \{f_{(2,k)}\}$ and we count their numbers in H_k . With varying k from 1 to P we obtain a histogram for H_k . The bin with maximum value ($K_m = \underset{k}{\operatorname{argmax}}(H_k)$) is the winner and we consider

Figure 1: Depicts a graphical representation of applying the affine transformation to two sets of correspondence points for finding the real correspondence points.

it as reference points $\{f_{(1,k)}\} \xleftrightarrow{k=K_m} \{f_{(2,k)}\}$ and all its member are considered as reliable correspondence points between the comparing image (I_1, I_2). The graphical representation of explained approach is demonstrated in **Figure1**. In this figure, (a) and (b) are two sets of potential correspondence points between two images represented as features' vectors. Let feature vector 1 (**Figure1(a)**) be a real match for feature vector 1' (**Figure1(b)**); thus, after applying the affine transformation to the two sets of correspondence points, the geometry of them should remain consistence with regards to the transformation based on this feature. For a correct match, we require a minimum three points from one plane to be matched to three points from another plane. Since each pair contains information about location, scale and orientation, a minimum of two pairs of the real correspondence points are sufficient to produce a correct match. However, with higher number of matched pairs, this will increase the confidence level of the matching operation.

III. EXPERIMENTAL RESULT

A. Data Description

The first image dataset have been used to evaluate the matching algorithm is the dataset that Mikolajczyk and Schmid [14] have used to evaluate the performance of different feature extraction methods in different scenarios. This dataset contains eight sets of different scenes, which may be blurred, zoomed, rotated or changed in illumination condition and viewpoints in six different levels. JPEG compression effects are also included. The images have different resolution form 765×512 to 1000×700 pixels. To make sure that the proposed matching algorithm is reliable, the worst-case images of each category are selected, which means the pair with maximum variation is selected in each category. The second data sources, which used to evaluate our matching algorithm, are Multi-view evaluation images provided by Strecha et al. [15]. These images have wide view variation. The third dataset, which have been used for finding reliable correspondence points among faces of multi-views is the standard color FERET images [16]. The images of this dataset are resized to the resolution 256×384 pixels and converted to grayscale. This dataset contains single face of different persons. It consists of 1209 persons with various age, gender and ethnicity. The average 10 images per person were captured which cover a wide range of face rotation from left profile to oblique, frontal and right profile.

Figure 2: Depicts the result of finding correspondence points in rotated and scaled views of objects or scenes.

B. Matching Objects and Scenes

In order to find the reliable correspondence points, first the SIFT features are extracted in two comparing images based on the method proposed by Lowe [6]. We tuned the parameters such that at least $1k$ features were extracted from each image. In the next step, common features are reduced from extracted features in both images using the likelihood ratio test in (1). Pruned features are then participated in Lowe's matching algorithm [6] to determine the potential correspondence points between the comparing images. Matching refinement performs by applying the affine transformation (described in previous section) to the potential correspondence points. Finally, the proposed method returns the real reference points between images pair if the number of matching points, found between them, was greater than 3 points; otherwise, the method returns nothing. This is the privilege of our method to the other methods such as RANSAC that our method returns NULL when the number of real correspondence points are so small which are not detectable but others always return many false matching points or it is better to say that they fail in this case. Although the process seems very complicated, it takes less than a second to find the real correspondence points between two comparing images at the resolution 256×384 pixels with non-optimized Matlab code using an ordinary PC.

The results of applying the matching procedure to the images of different scenes and objects have been demonstrated in **Figure2** and **Figure3**. The robustness of the algorithm over zoomed-rotated images, 180 rotation and different viewpoints is examined in **Figure2**. The number of matched points found for image pairs in **Figure2(a)** and (b) were 86 and 1004 points, respectively, without any error. It is inferred that if two comparing images follow an affine transformation, the number of real correspondence

Figure 3: Depicts the result of finding correspondence points in the different views of objects or scenes.

points found is large. However, the algorithm is able to find correspondence points between images that have nonlinear transform such as wide view difference as it is shown in **Figure3**. Our method has found 10 points and 118 points for image pairs in **Figure3(a)** and **(b)**, respectively, without any error. The main issue of the image pair in **Figure3(b)** is the repeated textures such as windows that usually make confusion for the matching techniques to detect the correct correspondence points. As the proposed method overlays two comparing images in the feature level and considers the relative geometry between features, the problem of confusing among similar texture is solved.

C. Matching Faces from Multi-Views

Faces of different people are not exactly the same. Due to gender, ethnicity and genetic factors the facial features differ from one face to another face in spite of similarities between the geometry of facial patches which is almost constant in all face images. This difference yields face image a wide in-class variation. Thus, it is not easy to find many true correspondence points between two comparing faces. However, it is possible to find some patches of one face to be matched to the similar patches in the other face image. In this case the number of real correspondence points between image pairs is very small such that ordinary methods fail to find these real correspondence points. We have applied our proposed method to these groups of images to show the strength of our method when the number of real correspondence points is a small portion of all the correspondence points found between image pair. In our experiment, we have considered two scenarios: First, face image pairs in the same view and second, face image pairs from different views which have less true correspondence points. We have compared our method with three variations of RANSAC methods (RANSAC 8-

points fundamental matrix, RANSAC affine fundamental matrix and RANSAC Euclidean matrix). Since the number of real correspondence points between these types of images is very small, we have set the Lowe's matching parameter to obtain the maximum number of potential correspondence points between two comparing images to increase the chance of finding some real correspondence points. The results of comparing the mentioned methods with our proposed method in the face image pair of the same view is demonstrated in **Figure4**. **Figure4(a)** demonstrates the potential correspondence points between two images found by the matching algorithm proposed by Lowe [6]. Since Lowe's method only considers the feature appearance to find the correspondence, many found correspondence points are not correct correspondence. Hence, we have manually selected the real correspondence between these two images among all the potential matches as it is shown in **Figure4(b)**. We have tried to find the real correspondence points among the potential correspondence points by applying RANSAC method with 8 points fundamental matrix model, affine fundamental matrix model and Euclidean matrix model in **Figure4(c)**, **(d)** and **(e)**, respectively. But all the methods failed due to the low number of real correspondence points. The result of our proposed method is shown in **Figure4(f)**. In term of stability comparison between models, we have sketched the number of correct correspondence points found between two comparing images, the number of wrong correspondence points, precision and recall graphs versus time as they are shown in **Figure4(g)**, **(h)**, **(i)** and **(j)**, respectively. As it is seen, our method has stabilized after a while but others failed. The number of potential correspondence points and ground truth for the image pair after applying Lowe's matching algorithm and the number of false positives, true positives and mutual information index obtained from different methods after passing one second have been compared in **Table I**.

We have also evaluated the accuracy of our method versus others using mutual information metric described in [17], [18]. Mutual information index (μ) between the predicted label and the ground truth label is obtained using (5).

$$\mu = \sum_{y=0}^1 \sum_{t=0}^1 Z(y, t) \cdot \log \frac{Z(y, t)}{Z(y)Z(t)} \quad (5)$$

where $Z(y = 1, t = 1)$ represents the true positive normalized with the number of test points, $Z(y = 1, t = 0)$ represents the normalized false positive and so on. The two individual terms can be calculated by applying summation in one variable for instance $Z(y) = \sum_{t=0}^1 Z(y, t)$. This criterion shows that our proposed method performs much better than the other methods as it is depicted in **Figure4(k)**. In this graph method c, d and e refer to RANSAC with different models and f refers to our method. When image pairs come from different views, it is more difficult to find many real correspondence points between them. However, our method is able to

Figure 4: Shows the result of finding correspondence points between two face images of the same view. (a) Potential correspondence points obtained using Lowe's method; (b) Ground truth correspondence points found manually; (c), (d) and (e) show the results for RANSAC method with different models and (f) shows the results of our method.

Figure 5: Shows the result of finding correspondence points between two face images from different views. (a) Potential correspondence points obtained using Lowe's method; (b) Ground truth correspondence points found manually; (c), (d) and (e) show the results for RANSAC method with different models and (f) shows the results of our method.

Table I: Compares the Performance of Different Methods on Face Images of Multi-Views

Method	Figure 4				Figure 5			
	# Potential Correspondence	# Ground Truth	# True Positives	# False Positives	# Potential Correspondence	# Ground Truth	# True Positives	# False Positives
Our method (f)	1095	34	32	25	1094	12	9	10
RANSAC 8-Points (c)	1095	34	3	85	1094	12	0	100
RANSAC Affine (d)	1095	34	15	64	1094	12	0	65
RANSAC Euclidean (e)	1095	34	0	18	1094	12	0	20

find the small group of consensus points in the presence of heavy noise where other methods typically fail as it is demonstrated in **Figure 5**. A brief comparison between our method and RANSAC method with different models is provided in **Table I** based on the comparing image pairs in **Figure 4** and **Figure 5**. As it is illustrated in this table, the number of true matched points or ground truth is about 1% to 3% of all the potential points found between images. However, our method has returned the least number of false positives and the most number of true positives. Thus, the mutual information index ranked our method in much higher value than the other methods. Our method has also another privilege that if we set a minimum on the total number of returned correspondence points (false positives plus true positives), the results could be NULL for those cases which two comparing images has no similarities based on matched points found by Lowe's method. However, other methods typically return only false positive points in this case. We have repeated our experiment for 500 different face images from multi-views over 500×499 different image combinations and obtained the same results as above. It is concluded that the only suitable method for finding real correspondence

points between face images of multi-poses is the proposed method

IV. CONCLUSIONS

We have introduced a matching algorithm which uses the information about appearance and geometry descriptors besides the discriminative features to find the real correspondence points between two images. The presented method is able to find the correct matches even if the number of true correspondence points between images is very small where other methods would simply fail. The other advantage of the proposed method is that it will return NULL results if it could not find any real correspondence points but other methods return only false correspondence points. This method also stabilizes very fast but other methods based on RANSAC are not able to finalize the results and they should be forced to stop their searching for the correspondences.

ACKNOWLEDGMENT

This research was supported by the Agency for Science, Technology and Research (A*STAR), Singapore.

REFERENCES

- [1] D. Fleck and Z. Duric, "Using local affine invariants to improve image matching," in *Pattern Recognition Proceedings (ICPR), 20th International Conference on*, 2010, pp. 1844–1847.
- [2] M. Liyong, S. Yude, F. Naizhang, and L. Zheng, "Image fast template matching algorithm based on projection and sequential similarity detecting," in *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), Fifth International Conference on*, 2009, pp. 957–960.
- [3] D. Chenguang, J. Song, and Z. Yongsheng, "Multiple-view matching ammgc model for three-line-array digital images," in *Information Science and Engineering (ISISE '08), International Symposium on*, vol. 2, 2008, pp. 139–142.
- [4] M. Perd'och, J. Matas, and O. Chum, "Epipolar geometry from two correspondences," in *Pattern Recognition Proceedings (ICPR), 18th International Conference on*, vol. 4, 2006, pp. 215–219.
- [5] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.
- [6] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [7] K. Mikolajczyk and C. Schmid, "Indexing based on scale invariant interest points," in *Computer Vision Proceedings (ICCV), Eighth IEEE International Conference on*, vol. 1, 2001, pp. 525–531.
- [8] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [9] R. C. Lo and W. H. Tsai, "Perspective-transformation-invariant generalized hough transform for perspective planar shape," *Pattern Recognition*, vol. 30, no. 3, pp. 383–396, 1997.
- [10] O. Chum and J. Matas, "Optimal randomized ransac," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 8, pp. 1472–1482, 2008.
- [11] J. Matas, O. Chum, and S. O. C. Ieee Computer, "Randomized ransac with sequential probability ratio test," in *Computer Vision Proceedings (ICCV), Tenth IEEE International Conference on*, 2005, pp. 1727–1732.
- [12] P. H. S. Torr, A. Zisserman, and S. J. Maybank, "Robust detection of degenerate configurations while estimating the fundamental matrix," *Computer Vision and Image Understanding*, vol. 71, no. 3, pp. 312–333, 1998.
- [13] G. Dorko and C. Schmid, *Selection of scale-invariant parts for object class recognition*, ser. Ninth IEEE International Conference on Computer Vision (ICCV), Proceedings. Los Alamitos: IEEE Computer Soc, 2003.
- [14] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.
- [15] C. Strecha, W. von Hansen, L. Van Gool, P. Fua, U. Thoennessen, and Ieee, "On benchmarking camera calibration and multi-view stereo for high resolution imagery," in *Computer Vision and Pattern Recognition Proceedings (CVPR), IEEE Conference on*, New York, 2008, pp. 2838–2845.
- [16] FERET, "The color feret database, online available at <http://www.nist.gov/itl/iad/ig/feret.cfm>," 2011.
- [17] D. J. Mackay, *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.
- [18] H. M. Wallach, "Evaluation metrics for hard classifiers," Cavendish Lab., Univ. Cambridge, Cambridge, online available at <http://www.inference.phy.cam.ac.uk/hmw26/>, Tech. Rep. Technical Report, Sep. 2006.