MATCHING OF INTEREST POINT GROUPS WITH PAIRWISE SPATIAL CONSTRAINTS

E. S. Ng, N. G. Kingsbury

Signal Processing & Communications Laboratory
Department of Engineering, University of Cambridge, U.K.
{esn21,ngk10}@cam.ac.uk

ABSTRACT

We present an algorithm for finding robust matches between images by considering the spatial constraints between pairs of interest points. By considering these constraints, we account for the layout and structure of features during matching, which produces more robust matches compared to the common approach of using local feature appearance for matching alone. We calculate the similarity between interest point pairs based on a set of spatial constraints. Matches are then found by searching for pairs which satisfy these constraints in a similarity space. Our results show that the algorithm produces more robust matches compared to baseline SIFT matching and spectral graph matching, with correspondence ratios up to 33% and 28% higher (respectively) across various viewpoints of the test objects while the computational load is only increased by about 25% over baseline SIFT. The algorithm may also be used with other features apart from SIFT.

Index Terms— Object matching, SIFT, Spatial constraints

1. INTRODUCTION

The search for accurate correspondences between features of images is an important and challenging problem in computer vision. Low level features, such as SIFT [1], are commonly matched based on local appearance alone without considering important information such as the spatial information of interest points associated with these features. Matching based on local appearance alone may be inadequate for complex scenes, since the presence of multiple features with similar appearance is quite likely. Improvements can be made if the spatial information of nearby interest points is considered, as it provides important constraints on the structure and layout of features which can be used for matching. The main challenge lies in designing an algorithm that considers spatial constraints along with feature appearance for matching, and studying what improvements can be achieved.

Spatial constraints, in addition to local feature appearance, should produce more robust matches, since interest points may be considered to match only when they satisfy the spatial constraint while also matching in appearance. This should reduce the number of false matches produced. Previous works have presented several approaches to solve the correspondence problem using spatial constraints. One approach is to formulate the problem using graphical models. In [2], a subgraph matching technique was proposed where a template graph was approximately matched to weighted adjacency graphs of the search images. Torresani *et al.* [3] also formulated the problem as an energy minimisation graph matching problem and solved the problem using a dual decomposition technique. Enqvist *et al.* [4] proposed a graph method based on vertex cover to solve for correspondences using pairwise constraints. Berg *et al.* [5] solved the problem using quadratic programming to minimise a cost function representing the similarity of matching features as well as the geometric distortion between pairs of corresponding points. Leordeanu and Herbet [6] proposed a spectral technique to find the best matching clusters in graphs measuring the pairwise similarities of points. The algorithm was shown to perform faster than graph methods.

Our work is most closely related to [6]. We study the use of spatial constraints for matching by calculating the relationships between interest point pairs and collecting them in a similarity space. A matching algorithm is proposed to search for the best matching subset of interest points which satisfy a set of spatial constraints in the similarity space. To test our algorithm, we performed experiments to compare the algorithm with baseline SIFT and spectral graph matching [6]. However, the algorithm can also be used effectively with other local features apart from SIFT.

2. MATCHING WITH SPATIAL CONSTRAINTS

In this section, we introduce a matching algorithm that searches for modes in a similarity space which describes the spatial relationships between interest point pairs.

2.1. Pairwise relationships based on spatial constraints

Consider an arbitrary group G of M interest points. We calculate the pairwise spatial relationships of each interest point in G with the rest, thus forming a $M \times M$ matrix. We consider two measures between each pair of points. The line joining each pair of interest points can be represented as a vector:

$$\hat{x}_{u,v} = \delta_{u,v} \exp(j\theta_{u,v}) \tag{1}$$

where u, v is a pair of interest points, $\hat{x}$ is the vector between the two points, with $\delta_{u,v}$ the length and $\theta_{u,v}$ the orientation of the vector. The first measure A_1, is defined as:

$$A_1(u, v) = \begin{pmatrix} \phi_u - \theta_{u,v} \\ \phi_v - \theta_{u,v} \end{pmatrix} \tag{2}$$

where ϕ is the dominant orientation of the feature associated with each interest point. $A_1(u, v)$ is the relative orientation of the two features to the orientation of the vector $\hat{x}_{u,v}$. This is a useful measure, since we expect the relative orientation to be approximately the same for a corresponding pair of interest points in different scenes. Apart from A_1, the features of the interest point pair are also stored in A_2:

$$A_2(u, v) = \begin{pmatrix} f_u \\ f_v \end{pmatrix} \tag{3}$$

where f is the local feature (e.g. SIFT) of the interest points u and v. This is required, since the local features of a corresponding pair of interest points should match. Thus for each group G, we have two measures A_1 and A_2, containing the relative orientation difference and the feature pair respectively for all pairwise combinations of interest points in G.

2.2. Pairwise spatial matching

Considering two groups of interest points from two different images, G_X and G_Y, we first calculate the pairwise similarity based on A_1 and A_2 described previously. We then define a pairwise similarity space based on interest point pairs, where $X_{u,v}$ and $Y_{p,q}$ are the interest point pairs (u, v) and (p, q) in G_X and G_Y respectively. As defined in (1), each pair of interest points can be represented as a vector $\delta\ \exp(j\theta)$. We can define the pairwise spatial relationship as the log-ratio:

$$\begin{aligned} \kappa + j\rho &= \ln\left(\frac{\delta_{u,v}\exp(j\theta_{u,v})}{\delta_{p,q}\exp(j\theta_{p,q})}\right) \\ &= \ln\frac{\delta_{u,v}}{\delta_{p,q}} + j(\theta_{u,v} - \theta_{p,q}) \end{aligned} \tag{4}$$

The pairwise similarity space is then defined as $\mathcal{S}(\rho, \kappa)$, where ρ is the difference in orientation of the vectors between interest point pairs (i.e. rotation) and κ is the log-ratio of the distance between interest point pairs (i.e. scale change). An illustration of a pairwise match is shown in Figure 1. A similarity score ψ is then defined for all possible pairwise matches. First, we consider the orientation differences in (2) and calculate the orientation consistency χ as:

$$\chi_{u,p} = \frac{\cos(\phi_u - \theta_{u,v} - \phi_p + \theta_{p,q}) + 1}{2} \tag{5}$$

where $\phi_u - \theta_{u,v}$ is from $A_1(u, v)$, and $\phi_p - \theta_{p,q}$ from $A_1(p, q)$, as defined in (2). $\chi_{u,p}$ thus measures the orientation consistency between interest points u and p, and $\chi_{v,q}$ can be calculated similarly. Next, we compare the pairwise similarity of features $\gamma_{u,p}$, defined as:

$$\gamma_{u,p} = \exp\left(-\|f_u - f_p\|^2/2\sigma^2\right) \tag{6}$$

where f_u is the feature in X matched to f_p in Y and σ is suitably chosen. We found that $\sigma = 1$ worked well when the f vectors were normalised for unit l_2-norm. Similarly, $\gamma_{v,q}$ can be calculated for f_v and f_q. When a pair of interest points have similar local feature appearance, we expect $\gamma \approx 1$. This is the case for χ in (5) as well, since the difference in orientation of features should remain consistent for an actual pair of matches. The similarity score $\psi_{\{(u,p),(v,q)\}}$ which combines the orientation consistency and feature similiarity is then defined as:

$$\psi_{\{(u,p),(v,q)\}} = \frac{\chi_{u,p}\gamma_{u,p} + \chi_{v,q}\gamma_{v,q}}{2} \tag{7}$$

Hence, $\psi_{\{(u,p),(v,q)\}}$ has a value close to unity when the interest point pair has a consistent orientation difference as well as feature similarity. Votes ψ are collected in the similarity space $\mathcal{S}(\rho, \kappa)$ for all interest point pairs in G_X and G_Y. The pairwise matches can then be found by searching for modes or regions of high density in $\mathcal{S}$. Here, we use a mean shift mode estimator [7] that searches for modes in $\mathcal{S}$, with ψ the weight of resulting peaks in $\mathcal{S}$. Histogram-based methods can also be used as an alternative here.

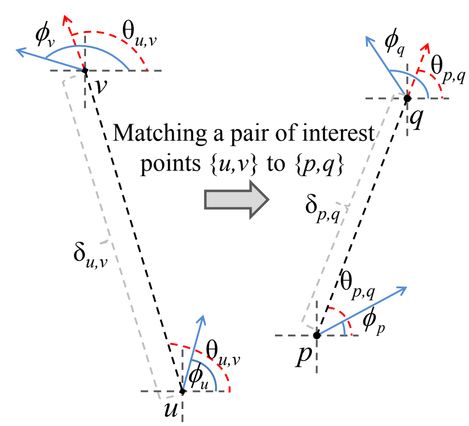


Fig. 1: Matching a pair of interest points u, v to a second pair p, q. $\theta_{u,v}$ is the direction of the vector between u, v, and $\delta_{u,v}$ is the distance between u, v (similarly for $\theta_{p,q}$ and $\delta_{p,q}$). ϕ_u, ϕ_v, ϕ_p, ϕ_q are feature orientations at interest points u, v, p, q.

3. PAIRWISE SPATIAL MATCHING ALGORITHM

Interest points which are far apart are likely to belong to separate objects and hence exist independently of each other, resulting in weaker spatial constraints between them. Thus, we consider the use of local interest point groups for matching, such that spatial constraints will only be employed over a local neighbourhood. To form interest point groups, we consider adjacent windows having 75% area overlap with each other in an image. Windows containing more than two interest points are then considered as interest point groups. We find that in general, 100 overlapping windows per image produced good results experimentally.

Along with the proposed pairwise spatial matching, we propose a matching algorithm based on the interest point groups formed. The algorithm, which uses SIFT [1] as the local feature, is summarised here. Consider two images X and Y, we perform an initial match between SIFT features of X and Y [1], such that a matching pair of features varies by a factor of less than 0.3 times of each other. This helps to reduce computational complexity since we can now consider fewer interest point pairs in the later stages of the algorithm. N and M groups of interest points are formed based on the initial matches, where G_{X_n} and G_{Y_m} are groups in X and Y indexed by $n = 1 \dots N$ and $m = 1 \dots M$. We then match the features in G_{X_n} to those in G_{Y_m} using a distance ratio threshold of 0.4 as defined in [1]. The measures A_1 and A_2 are calculated for all pairwise combinations of interest points in G_{X_n} that match to points in G_{Y_m}.

Here, matching the features in G_{X_n} and G_{Y_m} further reduces computational complexity before we consider the pairwise spatial constraints, since the number of pairwise combinations will be reduced after matching. A similarity space $\mathcal{S}_{n,m}(\rho, \kappa)$ is then defined and the mean shift mode estimator is used to find the pairwise combination of interest points with the maximum score. This is repeated across all G_{Y_m} for each G_{X_n}. Thus, for each G_{X_n}, we form M similarity spaces, each containing the similarity score of matching interest points in G_{X_n} and G_{Y_m}. The set of interest points that are best matched is found using the mean shift mode estimator. For each G_{X_n}, we select the group in Y with the highest similarity score as the correct set of matching interest points. In addition, we only accept groups with similarity score greater than a threshold $\tau = 0.7$. This is repeated for each G_{X_n}. The list of matching interest points between X and Y can then be found. The bandwidth of the mean shift mode estimator is set to the standard deviation of votes in $\mathcal{S}$.

4. EXPERIMENTAL RESULTS

In our experiments, we tested three algorithms; 1) the proposed algorithm *pw-match* as specified earlier along with the defined parameters, 2) the baseline SIFT matching algorithm using only local feature appearance and 3) the spectral technique in [6] *sp-match*, which considers pairs of SIFT features to study the performance of using spatial constraints for matching. Here, we set the distance ratio threshold for baseline SIFT matches to 0.4. We adopted the evaluation framework in [8] and selected 25 objects from the database provided for testing, given in [9]. The correspondence ratio is calculated for all objects at viewpoint increments of $5°$ from $-45°$ to $45°$:

$$\text{correspondence ratio} = \frac{\sum \text{actual correspondences}}{\sum \text{total matches}} \quad (8)$$

More details of the framework can be found in [8]. Based on our test results in Figure 2, we observe that *pw-match* produced correspondence ratios that are up to 33% and 28% higher (respectively) when compared to baseline SIFT and *sp-match* across all viewpoints. *pw-match* also produces higher correspondence ratio compared to SIFT when we vary the distance ratio threshold, as shown in Figure 2. More importantly, the improvement in correspondence ratio is higher at larger viewpoint changes, which implies that the use of spatial constraints results in more robust matches being found between scenes with larger viewpoint changes in them. In general, *pw-match* produces approximately 35% fewer total matches compared to baseline SIFT.

In addition, we tested the algorithms using images from the Zurich Building Image Database (ZuBud) [10]. Some results are shown in Figure 3. Since the ground truth for actual correspondences is not available, we compared the results visually and marked the false matches by inspection. From Table 1, we observe that *pw-match* generally produces fewer false matches compared to baseline SIFT and *sp-match*, along with higher correspondence ratio. Thus, our results as a whole suggest that the spatial constraints of the proposed algorithm generally produce more robust matches compared to using local feature appearance alone. More details of our results can be found at [9].

Table 1: Matching results for 15 buildings in ZuBud database

Results	baseline	*sp-match*	*pw-match*
Total matches	2199	2033	1483
Positive matches	1913	1830	1421
False matches	286	203	62
Correspondence ratio	0.870	0.900	0.958

5. CONCLUSIONS

The matching of features based only on local appearance may be insufficient in many instances, since many false matches may be produced especially in complex scenes where many features have similar appearance. By considering the spatial relationships between pairs of interest points, we account for the structure and layout of features which can improve the matches produced. In this paper, we have presented an algorithm which uses spatial constraints to produce more robust matches. A mean shift mode estimator is used to search for

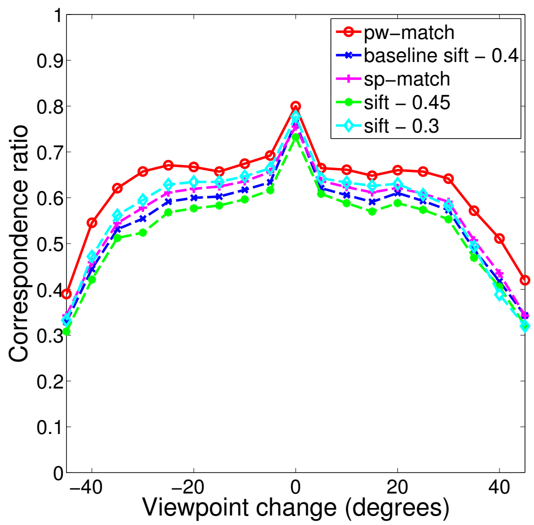
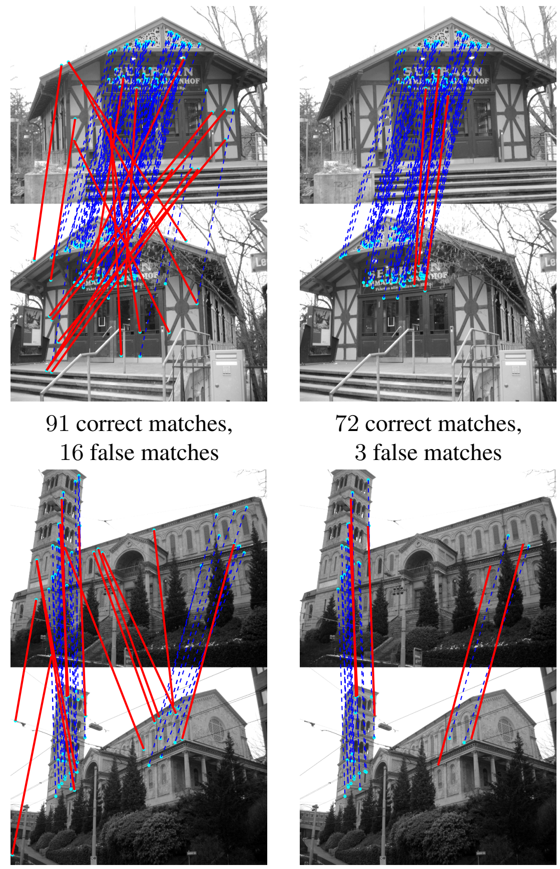


Fig. 2: Correspondence ratio for viewpoint changes. The proposed algorithm *pw-match* has a higher ratio compared to the algorithms tested. The improvement is more significant at larger viewpoints, suggesting that spatial constraints can be used to produce more robust matches. Similarly, *pw-match* has a higher ratio as the distance ratio threshold is varied for baseline SIFT.

interest points with similar pairwise constraints between images, and the estimated modes correspond to matching subsets of interest points. Our results suggest that the proposed algorithm is capable of producing more robust matches than using only local SIFT features for matching, as well as the spectral technique in [6]. The proposed algorithm has approximately 25% higher computational time compared to the baseline SIFT algorithm due to the collection of votes in the similarity space. But since our algorithm produces more robust matches, the increased computational complexity is likely to be justified for applications where more robust matches and fewer false matches are required. In future work, we will develop methods of using pairwise spatial matching to provide improved object recognition and classification systems.



91 correct matches, 16 false matches | 72 correct matches, 3 false matches

35 correct matches, 13 false matches | 16 correct matches, 5 false matches

Fig. 3: Matching results for a pair of complex scene. False matches (red) can be observed for both the proposed algorithm (right) and baseline SIFT (left), however the proposed algorithm produces much fewer false matches, thus resulting in more robust matches.

6. REFERENCES

[1] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov 2004.

[2] P. Tu, T. Saxena, and R. Hartley, “Recognising objects using colour-annotated adjacency graphs,” *Lecture Notes in Computer Science; Shape, Contour and Grouping in Computer Vision*, pp. 246–263, 1999.

[3] L. Torresani, V. Kolmogorov, and C. Rother, “Feature correspondence via graph matching: Models and global optimisation,” *Proc. European Conference on Computer Vision*, pp. 596–609, 2008.

[4] O. Enqvist, K. Josephson, and F. Kahl, “Optimal correspondences from pairwise constraints,” *ICCV*, 2009.

[5] A. C. Berg, T. L. Berg, and J. Malik, “Shape matching and object recognition using low distortion correspondences,” *CVPR*, vol. 1, pp. 26–33, 2005.

[6] M. Leordeanu and M. Hebert, “A spectral technique for correspondence problems using pairwise constraints,” *ICCV*, pp. 1482–1489, 2007.

[7] D. Comaniciu and P. Meer, “Mean shift: A robust approach toward feature space analysis,” *PAMI*, vol. 24, no. 5, pp. 603–619, May 2002.

[8] P. Moreels and P. Perona, “Evaluation of features detectors and descriptors based on 3d objects,” *International Journal of Computer Vision*, vol. 73, no. 3, pp. 263–284, July 2007.

[9] E. S. Ng, “Matching of interest point groups with pairwise spatial constraints,” http://www-sigproc.eng.cam.ac.uk/ esn21/.

[10] H. Shao, T. Svoboda, and L. Van Gool, “Zubud – zurich buildings database for image based recognition,” Technical Report 260, Computer Vision Laboratory, Swiss Federal Institute of Technology, Apr 2003, http://www.vision.ee.ethz.ch/showroom/zubud.