

Generalized RANSAC framework for relaxed correspondence problems

Wei Zhang and Jana Košecká
Department of Computer Science
George Mason University
Fairfax, VA 22030
{wzhang2,kosecka}@cs.gmu.edu

Abstract

Finding correspondences between two (widely) separated views is essential for several computer vision tasks, such as structure and motion estimation and object recognition. In the wide-baseline matching using scale and/or affine invariant features the search for correspondences typically proceeds in two stages. In the first stage a putative set of correspondences is obtained based on distances between feature descriptors. In the second stage the matches are refined by imposing global geometric constraints by means of robust estimation of the epipolar geometry and the incorrect matches are rejected as outliers. For a feature in one view, usually only one "best" feature (the nearest neighbor) in the other view is chosen as corresponding feature, despite the fact that several match candidates exist. In this paper, we will consider multiple candidate matches for each feature, and integrate this choice with the robust estimation stage, thus avoiding the early commitment to the "best" one. This yields a generalized RANSAC framework for identifying the true correspondences among sets of matches. We examine the effectiveness of different sampling strategies for sets of correspondences and test the approach extensively using real examples of *hard* correspondence problems caused by a large motion between views and/or ambiguities due to repetitive scene structures.

1 Introduction

Correspondence problem is one of the key problems in computer vision. Many vision tasks such as motion estimation and recognition require, or can greatly benefit from a set of corresponding points between the views. When trying to find correspondence¹ for a feature in one view, usually only the nearest neighbor (1-NN) in the second view which meets some matching criterion is selected as match. Once putative correspondences are obtained, usually robust

¹Points in different views are in correspondence when they are projection of same 3D point.

techniques (e.g. RANSAC² Algorithm) are used to impose global geometric constraints and identify the true correspondences.

In wide-baseline matching using scale and/or affine invariant features commonly used matching criteria are: distance between feature descriptors less than some threshold, similarity more than some threshold, or distance ratio between the nearest neighbor and the second nearest neighbor less than some threshold. Although in many instances the 1-NN rule is sufficient, it does have some deficiencies. As Figure 1 illustrates, the nearest neighbor might not be the correct correspondence when the local neighborhoods undergo large distortion, the descriptors change dramatically and are no longer sufficiently reliable. This is more likely when the scene contains similar repetitive structures, such as buildings and windows. As a result, if the ratio criterion is used, the number of true correspondences included in the correspondences set might be low. One way to increase the number of true correspondences is to loosen the matching threshold. However, it has been noticed that matches with lower matching scores have lower probability to be correct [1]. Moreover, the probability decreases very fast as the matching score goes down. Thus solely lowering the threshold is not a good solution for increasing correspondences. In other instances [2] multiple matches are kept, followed by additional verification based on an affine alignment to eliminate the incorrect matches. This stage is however quite computationally intensive if applied to all candidate matches.

Second, the sets of matches are often not symmetric, *i.e.*, $M(I_1, I_2) \neq M(I_2, I_1)$, where I_1 and I_2 are two images. This is due to the fact that multiple features in the query image can be matched to the same feature in the reference image, while every feature in the query image should only have one match in the reference image. Consequently, if the number of matches is used as similarity measure between images for recognition purposes, the fact that $Sim(I_1, I_2) \neq Sim(I_2, I_1)$ affects the recognition rate. In [3], the authors suggested to tackle the

²RANdom SAmple Consensus.

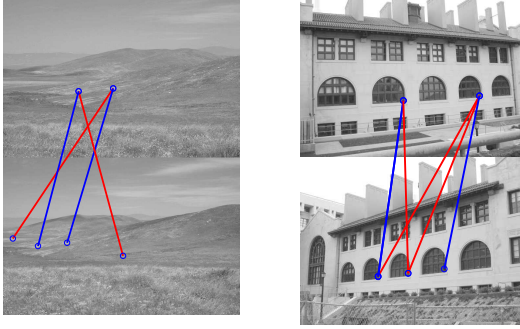


Figure 1: The closest matches in the second view (colored red) are not the true correspondences, even though the true correspondences (colored blue) also pass the matching threshold, they won't be retained by the 1-NN rule .

problem using bi-directional similarity measure, defined as $Sim(I_1, I_2) = (M(I_1, I_2) + M(I_2, I_1))/2$, which requires doubling of the computation time for finding matches.

Given the fact that 1-NN is not a reliable rule for finding correct correspondences, we propose to avert the early commitment to the nearest neighbor and leave the identification of the correspondences to later stage. With the proposed rule, all the features in the second view which pass the matching criterion are retained, thus for i^{th} feature f_i^1 in the first view, a match set $\{(f_i^1, f_j^2)\}, j = 1, \dots, n_i$, rather than one match will be associated with it, where n_i represents the number of matches for the i^{th} feature; n_i depends on how many features meet the matching criteria and is not fixed. The sets of matches of all the features will be the input of the second stage, where a generalized RANSAC framework will be used to identify the true correspondences and estimate the motion parameters.

Related work Significant amount of work on different feature detectors and descriptors has been proposed in the literature [4, 5, 6, 7]. The SIFT features proposed by [4] achieved best performance in the matching context based on comparison tests reported by Mikolajczyk and Schmid [8]. Ke and Sukthankar [9] developed the SIFT descriptor by applying PCA to image patches detected by SIFT demonstrating further improvement in their experiments. In [10], the author proposed to augment SIFT with a global context vector to handle scene with similar regions.

Despite the fact that great progress has been made in the past, in many domains matches obtained by comparing feature descriptors are often not reliable. In order to identify true correspondences, global geometric constraints (e.g. epipolar constraints) are usually used in connection with robust estimators. Examples include M-estimator [11], The Least Median of Squares (LMedS) Estimator [12]. RANSAC introduced by Fishler and Bolles [13] is the most

popular robust estimation technique used in computer vision community. Many works have been proposed to improve the standard RANSAC, which often requires large number of samples and has a costly hypothesis evaluation stage. Chum and Matas [14] suggested to improve the efficiency of standard RANSAC by a pre-evaluation called $T_{d,d}$ test. Even though the number of samples increased a lot, time is saved because only a fraction of data points is evaluated. In [15], the author proposed to select sample sets of adjacent points based on the assumption that the inliers will tend to be closer to one another than the outliers. Torr and Zisserman [16] have noticed that the traditional inlier count approach is faulty, since it treats all inliers equally (error terms for inliers are constant). They suggested using a log likelihood of solution as a support instead of the number of inliers. Tordoff and Murray [1] proposed to guide the sampling by the quality of matches to improve the sampling efficiency. The possibility of keeping multiple matches was also mentioned but not fully exploited. Based on the assumption that the similarity measure predicts correctness of a match better than a random guess, Progressive Sample Consensus [17] has been shown to be more efficient than RANSAC.

2 Keeping multiple matches

We describe our and test our method using SIFT keypoints and their associated SIFT descriptors, which have been shown to be superior in the previously mentioned comparison experiments. The proposed methodology is applicable to other existing features. Two features f_i^1 and f_j^2 are matched, when the cosine of the angle between their descriptors d_i^1 and d_j^2 is above some threshold τ_c . The cosine measure between two vectors v_1 and v_2 is defined as:

$$\cos(v_1, v_2) = \frac{v_1^T v_2}{\|v_1\|_2 \|v_2\|_2}. \quad (1)$$

Though τ_c is usually set to be fairly high, $\tau_c = 0.95$ in our experiments, still multiple features pass the threshold for some features. Alternatively one can use threshold related to the range of Euclidean distances between the two descriptors. We keep all the matches which pass the threshold, because experiments showed that matches with scores higher than the threshold have good chance to be true correspondences. If only the *best* one is selected, some true correspondences might be omitted.

Let σ_k be the fraction of points which have correct correspondence listed in the set of k matches. We have carried out an initial experiment to see the relationship between σ_k and k and obtained a cumulative distribution function (CDF). The probability distribution function (PDF) can be estimated by subtracting consecutive entries of CDF, each

entry of PDF represents the probability that k^{th} match is the true correspondence. The experimental results shown in Figure 2 were obtained based on an average of 5 widely separated image pairs. We can see that the 1^{st} nearest neighbor has the highest probability to be true correspondence, yet the probability is low and $\sigma_5 \approx 2\sigma_1$. Therefore, keeping multiple matches rather than only the *best* one can increase the number of true correspondences remarkably. In some scenes where repetitive structures are abundant, for example the building in Figure 1, the nearest neighbor can easily be some similar feature in a wrong location. In such a case σ_1 is relatively low and keeping multiple nearest neighbors is even more advantageous. The match set \mathcal{C}_i for the i^{th} feature is $\mathcal{C}_i = \{(f_i^1, m_1), (f_i^1, m_2), \dots, (f_i^1, m_{n_i})\}$, m_j is the index to features in the other view. Let the number of features which have nonempty match set to be N . The match sets between two views are $\{\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_N\}$, whose cardinality is $N_m = \sum_i n_i$.

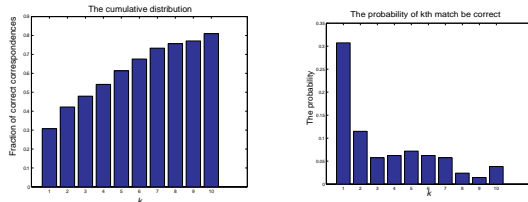


Figure 2: Left: the relationship between σ_k and k . Right: the probability that k^{th} match is the true correspondence.

Keeping a set of matches instead of one match requires some modifications to a standard RANSAC framework (described in next section). If applying RANSAC directly to the matching result, then it's possible that a sample contains several matches of the same feature in the first view, making it impossible to solve motion parameters. In the next section, we propose a generalized RANSAC framework to address the problem. We will show additional benefits of the generalized RANSAC by taking the distribution of matches into consideration.

3 Generalized RANSAC Algorithm

The standard RANSAC algorithm consists of two steps. First, it randomly selects M (a predetermined number) samples, the for each sample estimates a model hypothesis and finds the support (typically, the number of inliers) for this hypothesis. The hypothesis with the largest support is then chosen as a model and all its inliers are used to refine the model parameters. The inlier is defined as a data point whose residual is within some threshold T of the hypothesis. The idea is that those M samples include at least one sample which consists of only true correspondences, thus correct hypothesis can be obtained. Hence in order to

achieve confidence ρ that one such sample is obtained, the required number of samples M can be computed as:

$$M = \left\lceil \frac{\ln(1 - \rho)}{\ln(1 - (1 - \epsilon)^p)} \right\rceil \quad (2)$$

where ϵ is the fraction of outliers (false correspondences). In order to keep multiple matches per feature, we propose to generalize this procedure. The sampling stage is separated into two stages to account for the fact that there is only 1 correct correspondence for a feature in the first view. In the first stage the features from the first view are sampled, in the second stage the selected features' sets of matches are sampled. The generalized RANSAC algorithm is summarized in Algorithm 1.

Algorithm 1 The generalized RANSAC procedure.

1. Repeat for M_g (the number of required samples) times
 - (a) Select a sample of features (in the first view) which have a nonempty match set.
 - (b) For each feature in the sample, select one match from its match set.
 - (c) Estimate the hypothesis parameters based on the sample.
 - (d) Calculate the error of each match in \mathcal{C} with regard to the hypothesis.
 - (e) Determine the number of inliers consistent with the hypothesis.
 2. Select the model hypothesis with largest inlier support, then re-estimate model parameters using all the inliers.
-

When this framework is applied to the case where each match set only contains one match, the second sampling stage is straightforward and the whole procedure runs the same as the standard RANSAC. The hypothesis can be any model which is chosen to describe the relationship between two views, such as fundamental matrix F and homography H . The size of the sample selected in step 1(a) depends on the model. In our experiments, we use standard linear algorithms to estimate F and H . Therefore, the sizes of samples are 8 for F and 4 for H , respectively. For standard RANSAC, the required number of sample depends on the percentage of outliers ϵ as shown in Equation 2. For the generalized RANSAC, the required number of samples M_g is related to σ_k . What's more, M_g is related to the strategy of how to choose a sample. In the next sections we will discuss different sampling strategies for sampling the sets of matches.

3.1 Sampling strategies

The standard RANSAC algorithm samples the set of correspondences uniformly with the underlining assumption that the probability of each match being the true correspondence is equal. The assumption is generally invalid as pointed out in [1]. The authors instead suggested sampling matches according to their match quality, taking into account the fact that matches with higher scores are more likely to be true correspondences. By choosing appropriate sampling strategy, RANSAC can be expected to run more efficiently.

For the generalized RANSAC procedure, the second sampling stage 1(b) which samples within the match set is relatively straightforward. Corresponding point can be selected randomly from the match set assuming each match has roughly the same probability to be true correspondence. Alternatively the sampling can be guided by a matching score as mention above. This however requires knowledge of a mapping between the matching score and the probability of true correspondence. Currently in our experiments, the match set is sampled uniformly; for a match in \mathcal{C}_i , the probability that it is selected out of \mathcal{C}_i is $\frac{1}{n_i}$. The first sampling stage 1(a) requires more careful thought, since the match sets of different features have different cardinalities. Let the probability that i^{th} feature f_i^1 is selected be p_i , several sampling strategies can be considered as described next.

3.1.1 Proportional Sampling

With this strategy, we set p_i proportional to the size of its match set $p_i = \frac{n_i}{N_m}$. As a result, if each match set is uniformly sampled, the probability that a candidate correspondence is selected would be:

$$\frac{n_i}{N_m} \times \frac{1}{n_i} = \frac{1}{N_m}. \quad (3)$$

Clearly, each individual match has the same probability to be selected. Thus sampling based on this strategy means uniformly sampling set of all possible correspondences, with the guarantee that no same feature from the first image will appear in the same sample. The underlining assumption of this sampling strategy is the same as the standard RANSAC algorithm, where all the matches have the same probability to be true correspondences. Let this probability be $1 - \epsilon_g$ and probability that a selected match is an outlier would be ϵ_g . If σ_k is assumed to be the same for all the match sets, the probability that a selected match is an outlier is:

$$\epsilon_g = \frac{\sum_{i=1}^N \sigma_{n_i}}{\sum_{i=1}^N n_i}. \quad (4)$$

Given ϵ_g , it is relatively easy to obtain number of samples M_g in this case, which is of the same form as in Equa-

tion 2. We have obtained a preliminary relationship between σ_k and k in Figure 2 based on a small set of images. Even though more thorough experiments are needed to obtain a more accurate relationship, we do see an approximate trend here: $\sigma_1 > \sigma_2/2 > \dots > \sigma_k/k$. σ_1 is exactly the ϵ for the standard RANSAC algorithm. This means that if the matching criterion is the same (the same threshold), the proposed rule will return the match set with more fraction of outliers (because the total number of matches increases). This does not seem desirable. Note however, that with the same matching threshold, the proposed rule returns more true correspondences. If 1-NN rule is used instead, a lower threshold must be used to return the same number of true correspondences. In other words, we can use a higher matching threshold for the proposed rule (retain multiple matches) and still get a comparable number of true correspondences. Note the outlier percentage is not necessarily higher. The reason is the following: the probability of true correspondences decreases very fast as threshold decreases; in order to return the same number of true correspondences, the match set returned by 1-NN rule would contain much more outliers, probably even more than that of the proposed rule. Using the proposed rule and this sampling strategy, the generalized RANSAC procedure would not require more samples than using 1-NN rule and standard RANSAC.

The assumption that σ_k is the same for all the match sets is actually inappropriate. If the match set of a feature only contains 1 match, it means that this match is very distinctive and it is likely to be the true correspondence. The distance ratio criteria proposed by Lowe [4] is based on the similar idea. The two keypoints are matched if their descriptors satisfy:

$$\frac{Dist(d, d_{1st})}{Dist(d, d_{2nd})} < \tau_r, \quad (5)$$

where $d \in \mathcal{R}^n$ is the descriptor to be matched and d_{1st} and d_{2nd} are the closest and the second closest descriptors from the model database, with $Dist(., .)$ denoting the Euclidean distance between two descriptors and τ_r is some threshold. On the other hand, for feature which has multiple matches, its matches are less likely to be true correspondence. Generally, the larger the size of the set of matches, the lower the probability that each individual match is the true correspondence. Therefore, the next sampling strategy seems more meaningful.

3.1.2 Uniform sampling

In this case, we set p_i to be the same for all features, $p_i = \frac{1}{N}$. As a result, if each match set is uniformly sampled, the probability that a single match is selected for hypothesis estimation would be:

$$\frac{1}{N} \times \frac{1}{n_i} = \frac{1}{Nn_i}. \quad (6)$$

Now the matches in matching sets with low number of points have higher probability of getting selected. For example, if the size of a match set is half as that of another match set, the first match in the first set has twice the probability of getting selected as the first match in the second set. Note that if 1-NN rule is used, the two matches would have same probability of being selected. The use of k matches and generalized RANSAC shows an advantage here, since the assumption that distinctive matches are more likely to be true correspondences is in general valid.

Clearly, the generalized RANSAC algorithm can run more efficiently (M_g is smaller) with this strategy, since its underlining assumption is more reasonable than the first one. To explain why M_g is smaller in this case, let's consider the following example. Consider two features in the first view, which get 2 and 4 matches in the second view, respectively. Assuming σ_k is the same for both of them (the first strategy), $\sigma_1 = 0.3$, $\sigma_2 = 0.5$, $\sigma_3 = 0.6$ and $\sigma_4 = 0.65$, then the probabilities that those 6 matches are true correspondence are $(0.3, 0.2)$ and $(0.3, 0.2, 0.1, 0.05)$, respectively. When taking into account that distinctive matches are more likely to be correct, it is reasonable that first two probabilities are higher. Suppose that they are 0.35 and 0.25, the probabilities are now $(0.35, 0.25)$ and $(0.3, 0.2, 0.1, 0.05)$. If sampling is based on the second strategy, the probability that a chosen match is correct would be $((0.35 + 0.25)/2 + (0.3 + 0.2 + 0.1 + 0.05)/4)/2 = 0.2325$. On the other hand, if sampling is based on the first strategy, the probability that a chosen match is correct would be the average of all those probabilities: $(0.35 + 0.25 + 0.3 + 0.2 + 0.1 + 0.05)/6 = 0.2083$. Thus the effective outlier percentage is lower for the second sampling strategy, and less samples are needed to assure a good hypothesis sample. The difference in the effective outlier percentage would be more evident when the size of match sets are more different. Large sets of matches can be penalized even more, yielding the third sampling strategy.

3.1.3 Inversely Proportional Sampling

In this case, p_i is inversely proportional to the size of its match set, $p_i = \frac{1}{\sum_i \frac{1}{n_i}}$. As a result, if each match set is uniformly sampled, the probability that a single match is selected for hypothesis estimation would be:

$$\frac{\frac{1}{n_i}}{\sum_i \frac{1}{n_i}} \times \frac{1}{n_i} = \frac{1}{n_i^2 \sum_i \frac{1}{n_i}}. \quad (7)$$

Now matches in matching sets with small number of matches have higher probability of getting selected. We can see in the probability expression of the second strategy 6, the denominator contains a linear term in n_i , while the expression for the this strategy has a term quadratic in n_i . It

is straightforward to deduce that penalizing large sets more, will result in higher order terms of n_i . In the extreme case, only features with a single match will be selected. In such case, the probability that a single match is selected can be represented as:

$$\frac{1}{n_i^\infty \sum_i \frac{1}{n_i^\infty}}, \quad (8)$$

where n_i^∞ means the infinite power of n_i . The members of match sets with $n_i > 1$ will have zero probability to be sampled. If the assumption that the distinctive matches are likely to be true correspondences is valid, this last extreme sampling strategy would be the most effective one. The third strategy is also more effective than the second one, since the distinctive matches will have higher chance to be sampled. However, the assumption might not always be true. If we rely on the assumption and sample small match sets too excessively, it may turn out that not enough true correspondences can be obtained.

In practice, we adopt the second strategy, guaranteeing that the matches in large sets will still have a reasonable chance to be selected. We believe it is worthwhile to sacrifice some efficiency for better results. The suitability of sampling strategies also depends on the scene. For instance, for scenes with many repetitive structures, a good assumption is that all matches in a set are equally likely to be true correspondences and hence have similar probability, making the first strategy the most suitable this case.

4 Experimental Results

We have tested the proposed framework in several wide baseline matching and motion estimation scenarios, considering different motion models. For each image pair, the experiment is repeated 5 times to assure repeatability of the results. Due to randomized nature of the RANSAC framework, the identified inliers might not be the same all the time. However, the estimated motion parameters were consistent.

4.1 Estimation of fundamental matrix

For the estimation of the fundamental matrix we have used standard linear 8-point algorithm with normalization. The test images contained many repetitive structures, making the process of establishing correspondences using local features inherently ambiguous. While there are other approaches which could be used to overcome these inherent ambiguities by exploiting additional image structure, we have obtained good results without excessive number of RANSAC samples, as shown in Figure 3, 4 and 5. Given a fundamental matrix F hypothesis, correspondences are identified as inliers, if the Sampson error [18] is less than

the inlier scale (set to be 2 pixel in our experiments). The Sampson error for i^{th} correspondence given fundamental matrix F is defined as:

$$E_s = \frac{(\mathbf{x}_i^{2T} F \mathbf{x}_i^1)^2}{(F \mathbf{x}_i^1)_1^2 + (F \mathbf{x}_i^1)_2^2 + (F^T \mathbf{x}_i^2)_1^2 + (F^T \mathbf{x}_i^2)_2^2}, \quad (9)$$

where \mathbf{x}_i^1 and \mathbf{x}_i^2 are image coordinates of correspondences, and $(F\mathbf{x})_k^2$ represents the square of the k -th entry of the vector $F\mathbf{x}$.

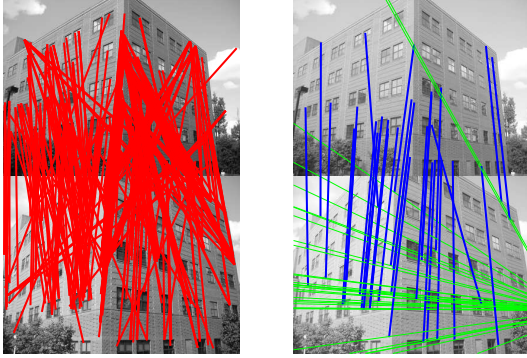


Figure 3: Left: matching result of the building with many repetitive structures. Right: identified inliers (connected with blue lines) using 2000 samples. Note some points seems to be matched to wrong positions, but their matches identified as inliers lie on the epipolar lines (green lines).

4.2 Estimation of homography

We also tested the proposed framework on planar scenes by estimating the homography between two views. In this case, the i^{th} correspondence is identified as inliers of a homography H if $|\mathbf{x}_i^2 - H\mathbf{x}_i^1| < T_H$, where T_H is the threshold set to be 3 pixels in our experiments. Figure 6 and Figure 7 show two examples. Even though repetitive structures are abundant, correct model can be estimated using 1000 samples at most. Note that using 1-NN rule and the standard RANSAC algorithm, correct model cannot be estimated with high confidence with such a small number of samples. Using 1-NN rule, 127 matches were obtained for Figure 6, 26 of them were correct. Using the same number (1000) of samples, out of 5 experiments, only 2 times could the correct homography be estimated. For Figure 7, 136 matches were obtained and 37 of them were correct. Although RANSAC could also identify correct matches using 1000 samples, the number of identified inliers using the proposed framework is consistently higher than in case of 1-NN matching criterion and the standard RANSAC algorithm. This brings the benefit of obtaining more accurate estimates.

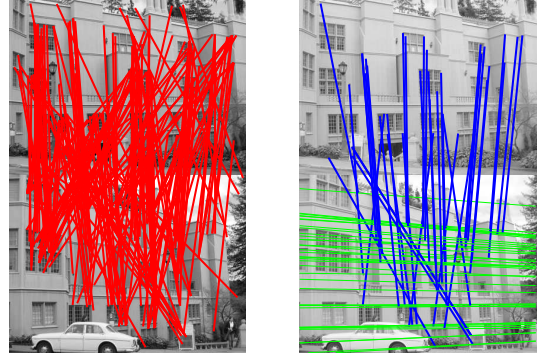


Figure 4: Left: matching result of an almost planar scene with many repetitive structures. Note the points on the small bulletin board are very important for the estimation of F because they are the only points out of the dominant plane. There were 209 matches and 72 of them were correct. Right: identified inliers (connected with blue lines) using 2000 samples. Note that some points seems to be matched to wrong positions, but their matches identified as inliers lie on epipolar lines (green lines). We also tried using 1-NN rule and standard RANSAC. 151 matches were obtained 60 of them were correct. With the same number (2000) of samples, only 1 out of 5 experiments correctly returned the matches on the bulletin board.

5 Conclusion and Future Work

In this paper, we proposed to retain a set of matches for each feature, instead of using a single *best* match. We described a generalized RANSAC algorithm to address the problem of sampling the set of matches and discussed different sampling strategies. We can see that the generalized RANSAC algorithm can not only incorporate matching scores, but can also take in the consideration the distribution of matches. Thus we claim that that it addresses the multiple matches problem favorably. The presented framework is suitable for solving the correspondence problem for widely separated views. In the future, we will try to investigate the influence of sampling strategies on the performance. In our work the distance between two descriptors was used as a criterion for evaluation of quality of the matches. Such criterion however can only provide a lower bound on the appearance of the local region associated with each feature and it can also happen that two features which are close together in the descriptor space are actually quite different. Ideally, residuals from an affine alignment of local neighborhoods could be used for assessments of the quality of the match, but this strategy is computationally expensive if applied to a large number of matches. The proposed approach can hence be viewed as an extension of standard RANSAC, to deal with multiple sets of matches, without incurring any additional cost on the first matching stage.

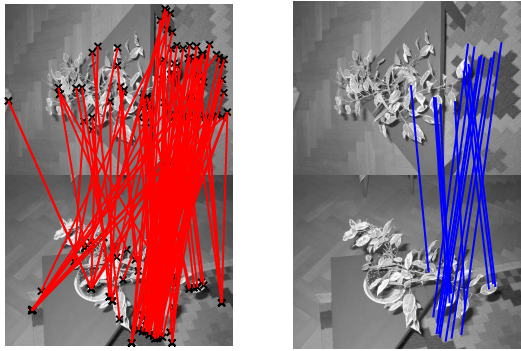


Figure 5: Left: The plant image pair is quite hard as pointed out in [17]. Right: identified inliers using 3000 samples.

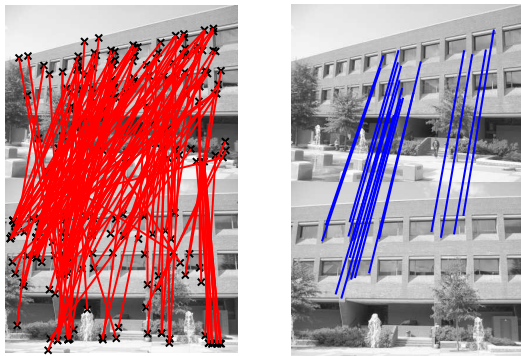


Figure 6: Left: matches of the facade of a building with many repetitive structures. There were 221 matches and 38 of them were correct. Right: identified inliers using 1000 samples.

References

- [1] Ben Tordoff and David W. Murray, “Guided sampling and consensus for motion estimation,” in *ECCV (1)*, 2002, pp. 82–98.
- [2] F. Schaffalitzky and A. Zisserman, “Multi-view matching for unordered image sets,” in *ECCV’02*, 2002, pp. 414–431.
- [3] M. Mattar, A. Hanson, and E. Learned-Miller, “Sign classification using local and meta-features,” in *CVAVI’05 (in conjunction with CVPR)*, June 2005.
- [4] D. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, 2004.
- [5] C. Schmid and R. Mohr, “Local greyvalue invariants for image retrieval,” *Pattern Analysis and Machine Intelligence*, vol. 19, pp. 530–535, 1997.
- [6] T. Tuytelaars and L. Van Gool, “Matching widely separated views based on affine invariant regions,” *IJCV*, vol. 59, 2004.
- [7] Jiri Matas, Ondrej Chum, M. Urban, and Tomas Pajdla, “Robust wide baseline stereo from maximally stable extremal regions,” in *BMVC’02*, 2002, pp. 384–393.

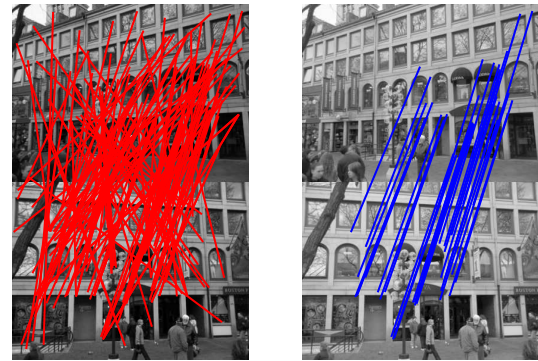


Figure 7: Left: another building facade with many repetitive structures. There were 168 matches and 44 of them were correct. Right: identified inliers using 1000 samples.

- [8] K. Mikolajczyk and C. Schmid, “A performance evaluation of local descriptors,” in *CVPR 2003*, 2003.
- [9] Y. Ke and R. Sukthankar, “Pca-sift: A more distinctive representation for local image descriptors,” in *Proceedings of IEEE Computer Vision and Pattern Recognition*, 2004.
- [10] E. Mortensen, H. Deng, and L. Shapiro, “A sift descriptor with global context,” in *CVPR’05*, June 2005.
- [11] P. J. Huber, *Robust Statistics*, John Wiley and Sons, 1981.
- [12] Peter J. Rousseeuw, “Least median of squares regression,” *Journal of the American Statistical Association.*, vol. 79, pp. 871–880, 1984.
- [13] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” in *ECCV’96*, 1996, pp. 683 – 695.
- [14] Jiri Matas and Ondrej Chum, “Randomized ransac with t(d,d) test,” in *BMVC02*, 2002, pp. 448–457.
- [15] DR Myatt, PHS Torr, SJ Nasuto, and RJ Craddock, “Napsac: High noise, high dimensional model parameterisation - its in the bag,” in *BMVC’02*, 2002, pp. 458–467.
- [16] P. Torr and A. Zisserman, “Mlesac: A new robust estimator with application to estimating image geometry,” *CVIU*, vol. 78, pp. 138–156, 2000.
- [17] O. Chum and J. Matas., “Matching with prosac: Progressive sample consensus,” in *CVPR’05*, June 2005.
- [18] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, ISBN: 0521540518, 2000.