REDUCING AMBIGUITY IN FEATURE POINT MATCHING BY PRESERVING LOCAL GEOMETRIC CONSISTENCY

Ouk Choi & In So Kweon

Korea Advanced Institute of Science and Technology
Division of Electrical Engineering, School of Electrical Engineering & Computer Science
335 Gwahangno, Yuseong-gu, Daejeon 305-701, Republic of Korea

ABSTRACT

In this paper, feature point matching is formulated as an optimization problem in which the uniqueness condition is constrained. We propose a novel score function based on homography-induced pairwise constraints, and a novel optimization algorithm based on relaxation labeling. Homography-induced pairwise constraints are effective for image pairs with viewpoint or scale changes, unlike previous pairwise constraints. The proposed optimization algorithm searches for a uniqueness-constrained solution, while the original relaxation-labeling algorithm is appropriate for finding many-to-one correspondences. The effectiveness of the proposed method is shown by experiments involving image pairs with viewpoint or scale changes in addition to repeated textures and nonrigid deformation. The proposed method is also applied to object recognition, giving some promising results.

Index Terms— wide-baseline stereo, affine regions, pairwise constraints, relaxation labeling

1. INTRODUCTION

Feature point matching is one of the fundamental problems in computer vision, where applications include 3D reconstruction, object recognition, and content-based image retrieval.

One of the difficulties in feature point matching is that the regions projected from the same preimage, such as the same part of an object, are different if their viewpoints are different. There have been efforts to resolve this problem by detecting regions covariant with the underlying viewpoint change and by describing the regions invariantly so that the regions projected from the same preimage will have similar descriptors [1, 2, 3].

Although covariant detection and invariant description have enabled feature points to be matched in the presence of significant viewpoint or scale changes, there remains the inherent problem of *ambiguity*. The descriptors of the regions projected from different preimages may be similar because the detected regions are often too small to include sufficient distinguished textures. For this reason, matches based on the most similar descriptors are not always correct.

In this paper, we aim to reduce the ambiguity in feature point matching for an image pair by finding a set of correct matches from a set of candidate matches. Correct matches satisfy the *uniqueness condition*, namely that a feature point in the first image corresponds to at most one feature point in the second image and vice versa, whereas candidate matches may not satisfy this uniqueness condition. Our main assumption is that object surfaces are locally smooth, at least between two feature points, which allows us to use homography-induced pairwise constraints between correspondences.

Approaches that use pairwise constraints [4, 5] have received a great deal of attention because of their ability to match feature points of nonrigid objects. In these approaches, feature point matching is usually formulated as an optimization problem aiming not only to maximize a score function but also to satisfy the uniqueness condition. Efforts have been made to develop good optimization techniques but less effort has been applied to pairwise constraints. Pairwise constraints have been designed to fix the distance or orientation between feature points, but it is questionable whether they remain effective when there are significant scale or viewpoint changes between images.

We define pairwise constraints based on the similarity of *local feature transformations* [6] between matches, where the term *local feature transformation* denotes a homography that transforms the neighborhood region of a feature point to the corresponding neighborhood region. In this paper, we show that homography-induced pairwise constraints enable robust feature point matching between image pairs with viewpoint or scale changes. In addition, we propose an optimization algorithm that finds one-to-one correspondences for image pairs with high ambiguity and nonrigid deformation.

This work has been funded by the IT R&D program of MIC/IITA [2006-S-028-01, Development of Cooperative Network-based Humanoids Technology] and the Korean MOST for NRL program [Grant number M1-0302-00-0064].

2. DETECTION OF CANDIDATE MATCHES

In this section, we explain how to obtain a set of candidate matches between images, and describe our notation.

We use affine invariant features (MSER-SIFT) [2, 1], although the proposed method can be generalized to use other features, because significant viewpoint changes between images are assumed. A feature point $\mathbf{x}$ (the gravitational center of an MSER) in image I is tentatively matched to a feature point $\mathbf{x}'$ in image I' if the SIFT vector distance d is smaller than a threshold value τ_d $(= 0.5)$. We also limit the number N of candidate matches for reasons of computational tractability, such as $N_{max} = 20000$, by selecting at most the N_{max} matches that have the smallest SIFT vector distances.

We denote detected candidate matches by:

$$\mathbf{m}_i = (\mathbf{x}_i, \mathbf{x}'_i),\ i = 1 \dots N, \tag{1}$$

where $\mathbf{x}_i$ and $\mathbf{x}'_i$ are tentatively corresponding feature points between images I and I'. For each $\mathbf{m}_i$, we consider not only a photometric observation d_i (the SIFT vector distance between $\mathbf{x}_i$ and $\mathbf{x}'_i$) but also a geometric observation $\mathbf{H}_i$, where $\mathbf{H}_i$ is the local feature transformation from $\mathbf{x}_i$ to $\mathbf{x}'_i$ [6, 7]. If necessary, we write $\mathbf{m}_i = (\mathbf{x}_i, \mathbf{x}'_i, \mathbf{H}_i)$ to include $\mathbf{H}_i$.

3. PAIRWISE GEOMETRIC CONSISTENCY SCORE

In this section, we define a pairwise geometric consistency score based on homography-induced pairwise constraints.

First, a pairwise transformation error e_{ij} between any two candidate matches $\mathbf{m}_i$ and $\mathbf{m}_j = (\mathbf{x}_j, \mathbf{x}'_j, \mathbf{H}_j)$ is defined as the sum of the symmetric errors produced by transforming a feature point using the local feature transformation of the other match:

$$\begin{aligned} e_{ij} &= e_i(j) + e_j(i), \\ e_i(j) &= \|\mathbf{x}'_j - \mathbf{H}_i(\mathbf{x}_j)\| + \|\mathbf{x}_j - \mathbf{H}_i^{-1}(\mathbf{x}'_j)\|, \\ e_j(i) &= \|\mathbf{x}'_i - \mathbf{H}_j(\mathbf{x}_i)\| + \|\mathbf{x}_i - \mathbf{H}_j^{-1}(\mathbf{x}'_i)\|, \end{aligned} \tag{2}$$

where $\mathbf{H}(\mathbf{x})$ denotes the Euclidean coordinates of $\mathbf{x}$ transformed by $\mathbf{H}$.

The error e_{ij} will be small if $\mathbf{H}_i$ and $\mathbf{H}_j$ are similar, and large if they are dissimilar. If $\mathbf{m}_i$ and $\mathbf{m}_j$ are detected from a smooth surface, then we can expect the error e_{ij} to be small, although the converse is not always true.

Finally, we define a pairwise geometric consistency score $f_g(e_{ij})$ as a smoothly decreasing function of e_{ij}, so that a geometrically consistent pair of matches can have a large value:

$$f_g(e_{ij}) = \exp(-e_{ij}^2 / 2\sigma_g^2), \tag{3}$$

where σ_g is a parameter that can be computed adaptively. Currently, our chosen σ_g value is:

$$\sigma_g = \frac{1}{N} \sum_{i=1}^{N} (\min_{j=1\dots N}(e_{ij})), \tag{4}$$

where min denotes the minimum value.

4. FORMAL PROBLEM DEFINITION

In this section, we formulate feature point matching as an optimization problem with the uniqueness constraint.

Let S denote the set of candidate matches:

$$S = \{\mathbf{m}_i = (\mathbf{x}_i, \mathbf{x}'_i) : i = 1 \dots N\}. \tag{5}$$

Two subsets of S can be defined for each $\mathbf{m}_i$. The first is a conflicting set $S_{\neg i}$, whose elements are all incorrect if $\mathbf{m}_i$ is correct, and $\mathbf{m}_i$ is incorrect if at least one of whose element is correct, via the uniqueness condition:

$$S_{\neg i} = \{\mathbf{m}_k = (\mathbf{x}_k, \mathbf{x}'_k) : \mathbf{m}_k \in S - \{\mathbf{m}_i\},\ \mathbf{x}_k = \mathbf{x}_i \text{ or } \mathbf{x}'_k = \mathbf{x}'_i\}. \tag{6}$$

The second subset is a supporting set S_i, defined as:

$$S_i = \{\mathbf{m}_j = (\mathbf{x}_j, \mathbf{x}'_j) : \mathbf{m}_j \in S - \{\{\mathbf{m}_i\} \cup S_{\neg i}\},\ e_{ij} < \tau_g\}, \tag{7}$$

where τ_g usually takes the value $3\sigma_g$.

We define a hidden variable $p_i \in \{0, 1\}$ as the belief in the correctness of $\mathbf{m}_i$. $p_i = 1$ if $\mathbf{m}_i$ is correct, and $p_i = 0$ otherwise. We also define auxiliary sets of beliefs as:

$$\begin{aligned} P &= \{p_i = \text{belief for } \mathbf{m}_i : \mathbf{m}_i \in S\}, \\ P_{\neg i} &= \{p_k = \text{belief for } \mathbf{m}_k : \mathbf{m}_k \in S_{\neg i}\}, \\ P_i &= \{p_j = \text{belief for } \mathbf{m}_j : \mathbf{m}_j \in S_i\}. \end{aligned} \tag{8}$$

Finally, a score function $f(P)$ is defined as:

$$f(P) = \sum_{p_i \in P} p_i f_d(d_i) + \sum_{p_i \in P} \sum_{p_j \in P_i} p_i p_j f_g(e_{ij}). \tag{9}$$

The uniqueness constraint can be written as:

$$p_i + \sum_{p_k \in P_{\neg i}} p_k = K,\ \forall p_i \in P, \tag{10}$$

where K must be 1 if $p_i = 1$, and K may be one of 0,1, or 2 if $p_i = 0$. If $K = 1$ for all p_i such that $p_i = 1$, then we can say that the uniqueness constraint have been satisfied; in this case, $p_k = 0$ for all $\mathbf{m}_k$ in conflict with $\mathbf{m}_i$. Indeed, our updating rule introduced in the next section aims to satisfy this condition.

The score function is composed of two kinds of terms: a unary score $f_d(d_i)$ that favors local appearance similarity, and the pairwise score $f_g(e_{ij})$ that was defined in Section 3. In principle, the score function $f(P)$ has the same form as those of previous pairwise approaches [4, 5]. The novel part is our homography-induced pairwise geometric consistency score $f_g(e_{ij})$.

The unary score $f_d(d_i)$, which favors similar SIFT vectors, is designed as a smoothly decreasing function of the SIFT vector distance d_i, which varies in the range $d_i \in [0, 1]$:

$$f_d(d_i) = 1 - d_i. \tag{11}$$

Our problem is to find the P that maximizes $f(P)$. After finding an optimal P, the final solution $S^\star$ is decided as $S^\star = \{\mathbf{m}_i : \mathbf{m}_i \in S, p_i \in P^\star\}$, where $P^\star = \{p_i : p_i > p_k, \forall p_k \in P_{\neg i}\}$.

5. PROPOSED ALGORITHM

In this section, we introduce an algorithm to maximize the score function (9), while imposing the uniqueness constraint. The belief p_i is relaxed to take a real value from $[0, 1]$ to avoid combinatorial searching, and we propose a variant of the original relaxation labeling algorithm [8, 9] to maximize the score.

We define a local support q_i as a partial derivative of the score function [9]:

$$q_i = \frac{\partial f(P)}{\partial p_i} = f_d(d_i) + 2 \sum_{p_j \in P_i} p_j f_g(e_{ij}). \qquad (12)$$

By defining q_i as in (12), $p_i q_i$ becomes a contribution of $\mathbf{m}_i$ to the score $f(P)$. In addition, $p_i q_i$ can be considered as a confidence measure for $\mathbf{m}_i$, with a large $p_i q_i$ meaning that $\mathbf{m}_i$ is a good match in terms of local appearance similarity and geometric consistency. An idea that follows naturally is that p_i should be amplified if $p_i q_i$ is large.

An updating rule is given by:

$$p_i \leftarrow p_i q_i, \ \forall p_i \in P, \qquad (13)$$

$$p_i \leftarrow \frac{p_i}{p_i + \sum_{p_k \in P_{\neg i}} p_k}, \ \forall p_i \in P. \qquad (14)$$

Here, p_i is first replaced by $p_i q_i$, and then suppressed by the conflicting beliefs $p_k \in P_{\neg i}$.

The major difference from previous relaxation-labeling algorithms [8, 9, 5] is that the normalization (14) is not conducted for a fixed number of labels. The proposed algorithm searches for one-to-one correspondences, whereas previous relaxation labeling algorithms seek a label (a feature point in image I') for a site (a feature point in image I), which means that a feature point in I' may correspond to many feature points in I.

The proposed algorithm is similar to the cooperative algorithm [10] in the sense that both algorithms are based on the match-space formulation and search for one-to-one correspondences between images. Many interesting properties, including convergence to a uniqueness-constrained solution, can be shown for the proposed algorithm. However, we omit these them for reasons of space.

Relaxation labeling is known to be a local algorithm that finds a local maximum [9], so good initialization has been considered crucial in achieving a good solution. Although we cannot guarantee that a consistent solution can be produced from an arbitrary initialization, good results are produced from a naive uniform initialization, such as $p_i = 0.5$ for all $p_i \in P$. This uniform initialization was used for all the image pairs considered in this paper.

6. EXPERIMENTS

We first evaluated the proposed method using image pairs for a chessboard pattern with significant deformation, as shown in Fig. 1, which are examples with high ambiguity. We manually counted correct matches in the evaluation for the image pair in Fig. 1(c) because the nonrigid deformation prevented ground-truth parametric models from being used. For the planar image pair in Fig. 1(a), we used ground-truth homography in the evaluation.

For comparison with our *proposed* method, we implemented *texture*, a texture-descriptor-based method [11]. We also developed a *simple* method, which had a simple pairwise constraint similar to the last constraint in [4], coupled with the proposed optimization algorithm of Section 5. The fourth implementation, *spectral*, involved the homography-induced pairwise constraint coupled with the spectral method of [4]. For the three pairwise methods, namely *simple*, *spectral*, and *proposed*, we used the same initial set of candidate matches. For all four methods, including *proposed*, the detected matches were sorted in either ascending or descending order of an appropriate measure, such as the ascending order of the ratio between the best and the second-best dissimilarities in the *texture* method.

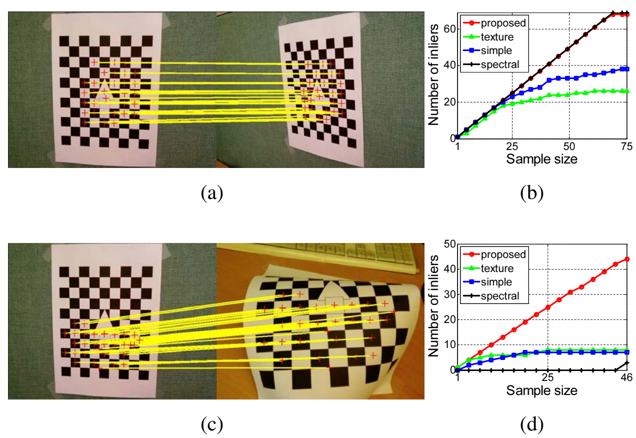

Fig. 1. A chessboard pattern with significant deformation. (a,c) The top 30 best matches detected by the proposed method. The yellow lines represent the detected matches. (b,d) Matching results for the image pair (a) and (c), respectively. The graphs show how many of the top k best matches are correct (k = Sample size).

Figures 1(b) and (d) give the quantitative matching results, with the graphs indicating how many of the top k best matches are correct. The *simple* method produced better results than the *texture* method for the image pair with a small-scale change (Fig. 1(a)). However, it did not produce better results for the image pair with a large-scale change (Fig. 1(c)). The *proposed* method produced the best results for both image pairs, while the *spectral* method failed to find correct correspondences for the image pair in Fig. 1(c), finding instead strongly structured false matches.

We applied the *proposed* method to other image pairs with ambiguity, such as those in Fig. 2, and it produced the best results. For all the tested image pairs except the one in Fig. 1(c), the *spectral* method produced similar results to those of the *proposed* method. Both the *proposed* method and the *spectral* method use the homography-induced pairwise constraint, and the experiments show that these two methods are more effective for the image pairs with scale or viewpoint changes than the *simple* method, which does not use the homography-induced pairwise constraint.

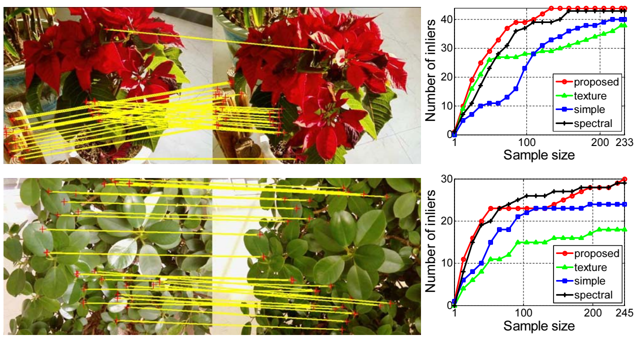

Fig. 2. Plant scenes with viewpoint changes. Left: The top 30 best matches detected by the proposed method. Right: Matching results for the image pair on the left.

Finally, we applied the proposed method to object recognition in a cluttered environment. For this experiment, we used images from KAIST-104 DB [12]. Fig. 3 shows some examples. Feature points from the cluttered background are often matched to feature points from the objects. For this reason, the success rate was 71.15%, classifying 74 images correctly in 104 query images [12]. We classified each cluttered query image by first matching it to all of the 104 data images using the proposed method, and then selecting the data image that maximized the score (9) of the matches in the largest GAM (Groups of Aggregated Matches) [7], where the score was divided by the number of feature points in the data image, to discourage an incorrect data image with a large number of feature points from getting a larger gain than a correct data image. For the fairness of the comparison, we used a fixed parameter ($\tau_g = 30$) for the clustering and the score computation. We could classify 85 images correctly, which is about a 10% improvement on state-of-the-art methods [12].

Fig. 3. Examples of query-and-data image pairs from KAIST-104 DB [12]. Matches detected by the proposed method are displayed for both of the image pairs.

7. CONCLUSIONS

In this paper, we proposed a robust feature point matching method that combines the advantages of the homography-induced pairwise constraint and a matching algorithm based on optimization. The homography-induced pairwise constraint was shown to be effective for image pairs with viewpoint or scale changes, and the proposed optimization algorithm found good solutions despite high ambiguity and nonrigid deformation between images. The proposed algorithm sometimes found a better solution than state-of-the-art algorithms such as the spectral method [4].

The major drawbacks of the proposed method are its high computational complexity and memory requirements for computing pairwise constraints, and we are currently investigating an effective method to reduce them.

8. REFERENCES

[1] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.

[2] J. Matas, O. Chum, U. Martin, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *BMVC*, 2002, pp. 384–393.

[3] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *IJCV*, vol. 65, no. 1–2, pp. 43–72, 2005.

[4] M. Leordeanu and M. Hebert, "A spectral technique for correspondence problems using pairwise constraints," in *ICCV*, 2005.

[5] Y. Zheng and D. Doermann, "Robust point matching for nonrigid shapes by preserving local neighborhood structures," *PAMI*, vol. 28, no. 4, pp. 643–649, 2006.

[6] O. Choi, H. W. Kim, and I. S. Kweon, "Simultaneous plane extraction and 2d homography estimation using local feature transformations," in *ACCV*, 2007.

[7] V. Ferrari, T. Tuytelaars, and L. Van Gool, "Simultaneous object recognition and segmentation from single or multiple model views," *IJCV*, vol. 67, no. 2, pp. 159–188, 2006.

[8] R. A. Hummel and S. W. Zucker, "On the foundations of relaxation labeling processes," *PAMI*, vol. 5, no. 3, pp. 267–287, 1983.

[9] S.Z. Li, *Markov random field modeling in image analysis*, Springer-Verlag, 2001.

[10] C. L. Zitnick and T. Kanade, "A cooperative algorithm for stereo matching and occlusion detection," *PAMI*, vol. 22, no. 7, pp. 675–684, 2000.

[11] P.-E. Forssén, "Maximally stable colour regions for recognition and matching," in *CVPR*, 2007.

[12] S. Kim, K. J. Yoon, and I. S. Kweon, "Object recognition using a generalized robust invariant feature and gestalt's law of proximity and similarity," *Pattern Recognition*, vol. 41, no. 2, pp. 726–741, 2008.