

A GLOBAL CORRESPONDENCE FOR SCALE INVARIANT MATCHING USING MUTUAL INFORMATION AND THE GRAPH SEARCH

Hyun-Ho Jeon, Andrea Basso and Peter F. Driessen

Department of Electrical and Computer Engineering
University of Victoria, Victoria BC, V8W 3P6, Canada
{hjeon, abasso, peter}@ece.uvic.ca

ABSTRACT

In this paper we propose a novel approach to find a global correspondence between two images by maximizing mutual information in the presence of large scale changes and rotations. Our approach combines the local descriptors with global search. We have tested our method on various test images and compared the matching performance with the SIFT descriptors. The experimental results show that the proposed local descriptor and graph-based search provides robust point matching for scale, rotation and illumination changes.

novel matching method that employs an invariant local similarity measure and the graph search technique.

There are two key steps of such procedure: 1) We build a mutual information based local description that is invariant to scale changes, rotations and illumination changes. 2) The best available set of compatible correspondences is found by maximizing mutual information through the graph search.

The rest of the paper is organized as follows: Section 2 describes related work for feature matching between images. Section 3 describes the proposed matching algorithm. Section 4 presents experimental results. Section 5 presents our conclusion.

1. INTRODUCTION

Finding feature correspondences between two or more images of the same scene has been one of the fundamental problems in computer vision such as image registration, object tracking and object recognition. Each point to be matched must be first detected. And then each point and its neighborhoods are described by feature descriptors so that it can be matched to the points in another image. It is important that the same points must be detected in both images and the detected points should be characterized as uniquely as possible.

Many approaches have been proposed to estimate feature correspondences between images. The main difficulties in this problem are the ambiguity in detecting the same points in both images and matching them based on their local features. We address the latter problem. Recent studies in the literature use local invariant descriptors to describe and match the image features under various changes [2][6]. These local descriptors enable the detected points to be more distinctive.

In this work, we define a procedure to identify a global correspondence between two images, which is optimal in the number of matched points and overall similarity of the determined matches. To solve this problem, we propose a

2. RELATED WORK

If the two images of the same scene differ largely by a camera's viewpoint, focal length, orientation, or illumination, conventional corner detectors may not detect the same points in both images very well and correlation based similarity measures tend to fail to correctly distinguish the detected points [1]. In order to find correspondences between the images under rotations and scale changes, many approaches have been developed in the literature. These approaches first detect feature points and then calculate feature descriptors for similarity measurement. Schmid and Mohr [8] showed that rotated images can be matched by using Harris' corner detector and rotationally invariant local descriptor. Dufournaud et al. [2] proposed a multi-scale framework. A high resolution image is smoothed at different scale levels and the scale ratio between two images is estimated by initial matching. Then, the higher resolution image is smoothed to the scale of a lower resolution image and feature points between two images are compared. Lowe [6] used the local extrema of difference-of-Gaussian in scale-space as feature points. He proposed a local descriptor, which is computed by local image gradients around feature points.

3. PROPOSED METHOD

This section considers the problem of scale and rotation invariant point matching using mutual information combined with graph search techniques. We first discuss mutual information as an invariant similarity measure. After that, we formulate the correspondence problem as a maximization of mutual information between two feature point sets and find the largest set of matching points through the graph search.

3.1 Mutual information as a similarity measure

Mutual information is a measure from information theory and it is the amount of information one variable contains about the other. Mutual information has been used extensively as a similarity measure in medical image registration [7] due to its robustness to illumination changes. Mutual information has also been used for correlation based stereo matching [3][5].

Mutual information between two random variables depends on the entropy and joint entropy of the two random variables. In the case of point matching, we can define a measurement window in each feature point. Then, the image pixels within the window can be discrete random variables [3]. If we assume X and Y are pixel values of two images, the mutual information between X and Y is defined as

$$MI(X,Y) = H(X) + H(Y) - H(X,Y) \quad (1)$$

where $H(X)$ and $H(Y)$ are the entropies of X and Y derived from the probability density functions, i.e.,

$$\begin{aligned} H(X) &= - \sum_{x \in W1} P(x) \log(P(x)), \\ H(Y) &= - \sum_{y \in W2} P(y) \log(P(y)) \end{aligned} \quad (2)$$

$H(X,Y)$ is the joint entropy of X and Y derived from the joint probability density function $P(x,y)$,

$$H(X, Y) = - \sum_{x \in W1} \sum_{y \in W2} P(x,y) \log(P(x,y)) \quad (3)$$

$W1$ and $W2$ are the measurement windows centered at feature points in two images. The probability density function is approximately estimated from the normalized 2-D joint histogram of gray-level pixels in the matching windows. It is given by, $P(x,y) = h(x,y)/N$, where $h(x,y)$ denotes the joint histogram between the two matching windows. N is the total number of pixels in the matching window. The probability densities $P(x)$ and $P(y)$ can then be obtained by a summation over the row and column directions of the joint density, respectively.

Conventional approaches of mutual information based

corner matching [3][5] cannot be used directly for our problem because the mutual information between two corresponding image regions is not invariant to scale changes and rotations. To achieve invariance, we incorporate the scale and the orientation information to the local mutual information measure. We extract feature points in both images by using the Scale Invariant Feature Transform (SIFT) detector [6]. The SIFT detector provides the location, scale and orientation of each feature point (termed as ‘‘keypoint’’). Before the computation of mutual information between two feature points, we change the size of the two matching windows based on the scale values of the SIFT keypoints. Then, the two windows are aligned by rotating one window to the direction of the other window’s dominant orientation. Our feature descriptor is rotation invariant since it is rotated to the keypoint’s orientation. Further the descriptor is scale invariant since it is constructed based on the scales of each point.

3.2 Graph search

We aim to find a global mapping between two images which has the largest number of correspondence pairs. In order to find such a mapping, we use a graph search which was proposed by Horaud and Skordas [9]. Based on their work, we establish a set of rules that allow us to apply topological constraints and suppress false matches. Let I_1 and I_2 be the two images. For each feature point in the image I_1 , we search a number of initial matches in the other image I_2 . These initial matching pairs constitute the individual nodes of the association graph. The association graph shows the relationship between the potential correspondence pairs and enables the determination of the largest correspondence set. Let the association graph $G = (N,E)$ be an undirected and unweighted graph, where $N = \{n_{ij}, i \in [1, \dots, N_1], j \in [1, \dots, N_2]\}$ is the set of nodes and $E = \{e_{sk}, s, k \in [1, \dots, M], s \neq k\}$ is the set of edges. N_1 and N_2 are the total number of candidates in image 1 and image 2, n_{ij} is a node representing a match between two candidates, e_{sk} denotes an edge connecting two nodes, and M is the total number of nodes. A node score B_{node} is computed for each node, which takes the similarity (i.e., mutual information) of that node.

After the nodes are built, nodes are linked by edges based on compatibility between nodes. Three rules are built to examine the node compatibilities. Consider two correspondence point pairs, $(p_1(i), p_2(j))$ and $(p_1(m), p_2(n))$, and their nodes n_{ij} and n_{mn} . Node n_{ij} is compatible with node n_{mn} if following three conditions are true:

Rule 1: $(i \neq j)$ and $(m \neq n)$

Rule 2: $\|p_1(i), p_1(m)\| \geq \|p_2(j), p_2(n)\|$

where $\|\cdot\|$ denotes the distance between two points.

Rule 3: $\text{Angle}(p_1(i), p_1(m)) \approx \text{Angle}(p_2(j), p_2(n))$

The first rule enforces one-to-one matching between feature points. Our second rule applies a distance constraint to the point pairs. As image I_1 is assumed to be assigned with a higher resolution image than image I_2 in this work, the spatial distance between two points in the image I_1 cannot be less than the distance between two corresponding points in the image I_2 . The third rule reflects the topological constraint. Angle between two points $p_1(i)$ and $p_1(m)$ with respect to the dominant orientation of the point $p_1(i)$ should not be changed for the corresponding two points in the image I_2 under scaling or rotation.

After building the association graph between the two images, correspondences between points are established by finding the largest set of matching points (i.e., maximal clique) in the graph. When multiple maximal cliques, $\{C_i\}_{i=1}^K$, are found, the best one $w(G)$ is determined by following evaluation:

$$w(G) = \max_i (\text{Score}(C_i)) \quad (4)$$

$$= \max_i \left(\sum_{n=1}^L B_{n,i} \right)^K$$

where L denotes total number of maximal cliques and K represents the size of maximal clique.

4. EXPERIMENTAL RESULTS

In this section, we present the experimental results based on the method described in section 3. Our matching method has been applied to the test image pairs, which contain large scale changes, rotations and brightness changes. We used real images from [4] and the Internet. We compare the evaluation results of our method and the SIFT descriptor. In both methods, the interest points are detected by the SIFT detector. Since the SIFT detector generates hundreds or thousands of key points per image, we pick the 100 points with the largest entropy in image 1 and search the matching points in image 2 using the descriptors to be compared. For the SIFT method, matching points are determined by searching the point pairs with minimum Euclidean distance in their keypoint vectors.

To determine the number of correct matches in each image pair, we need to know the ground truth pixel correspondences between image pairs. To compute the true correspondences, we use the geometric transformation matrices that are provided with the test images. In the experiments, an interest point p_2 in image 2 is considered as a correct match of a point p_1 in image 1 if the spatial distance between p_1 and p_2 is within a three pixel distance.

The correct matching rate is defined as in Eq. 5 and used to evaluate the performance.

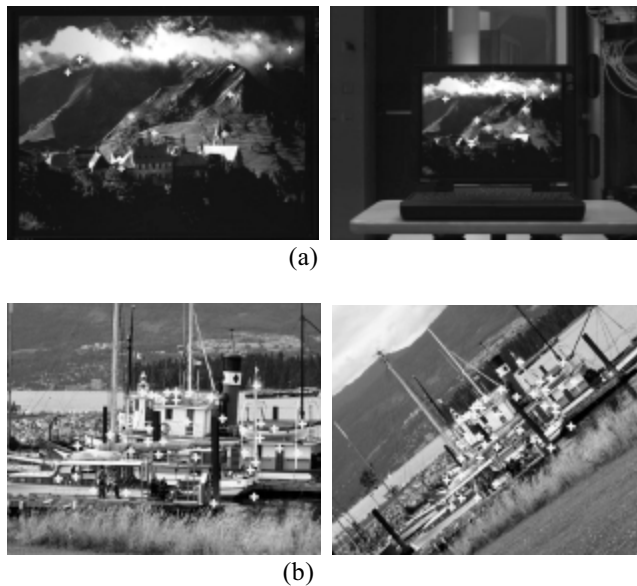


Fig. 1. Matching results for “Laptop” and “Boat” images by the proposed method. (a) The scale factor between the two images is 2.1. 25 points have correct matches. (b) The rotation angle is 40 degrees and scaling factor is 1.4. 43 points have correct matches.

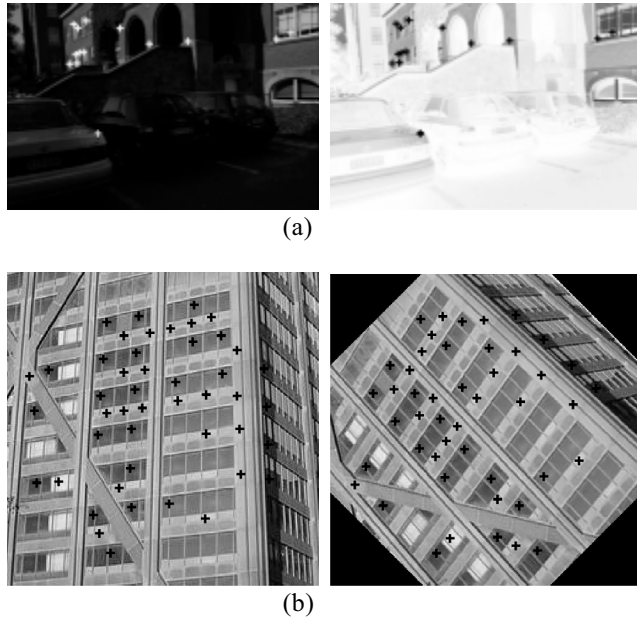


Fig. 2. Matching results for “Cars” and “Building” images. (a) The right image is synthetically generated by altering the intensity of the left image. 19 correct point matches are found. (b) The right image is generated by 45 degrees rotation of the left image. 47 points have correct matches.

$$r = \frac{\text{\# of correct matches}}{\text{\# of detected points in image 1}} \quad (5)$$

Fig. 1 shows the matching results of the proposed method for two different scenes, which contain scale and rotation changes. The scale factor in Fig. 1 (a) is 2.1, and 25 correct matches are detected. Fig. 1 (b) shows both rotation and scale changes, and 43 correct matches are found.

Fig. 2 shows the results for images containing brightness changes and locally similar structures. In Fig. 2 (a), the right image is generated by inverting the intensity of the left image. Although the intensities of the corresponding two points are quite different, the use of the mutual information enables correct matching between two images. In Fig. 2 (b), the right image has many points that are locally similar to the points in the left image. For example, there are many squares in both images having similar texture patterns. These locally similar areas produce ambiguities and result in many false matches. Since our method combines global constraint with local descriptors, we can effectively avoid false matches and find 47 correct matches for this scene.

Table 1 shows the comparison of the matching results with the SIFT descriptors. For images containing scale and rotation changes, the proposed method is comparable to the SIFT descriptor. For intensity changes, our method shows 16 % improvements of the correct matching rate. Since our local matching cost is based on mutual information, it is insensitive to illumination changes. For images containing many locally similar points, our method performs slightly better than the SIFT descriptor.

Table 1 Comparison of the matching results on the test images. Columns 2 and 3 show the number of correct matches for each image. The last column shows the improvements of the correct matching rates.

Image	Proposed	SIFT	Δr (%)
Laptop	25	29	- 4.0
Boat	43	44	- 1.0
Cars	19	3	+ 16.0
Building	47	39	+ 8.0

5. CONCLUSION

In this paper, we propose a novel approach to find a global correspondence between two images by maximizing mutual information in the presence of large scale changes and rotations. First, we present a local descriptor based on mutual information and showed that it is invariant to scale changes and rotations. Second, we formulate the correspondence problem as a maximization of mutual

information between two feature point sets and find the global correspondence by looking for the maximal clique in the graph search. Our test results show that mutual information can be used to the problem of scale invariant matching. If the feature points and their neighborhoods are distorted by more complicated models such as affine transformation, more robust algorithm should be used, which will be our future work.

6. REFERENCES

- [1] M. Brown and D. Lowe, "Invariant features from interest point groups", Proc. British Machine Vision Conference, pp.253-262, 2002.
- [2] Y. Dufournaud et al, "Image matching with scale adjustment", Computer vision and image understanding 93 (2004), pp.175-194, 2004.
- [3] G. Egnal, "Mutual information as a stereo correspondence measure", Technical report, Dept. of computer and information science, University of Pennsylvania, 2000.
- [4] <http://lear.inrialpes.fr/people/Mikolajczyk/Database>
- [5] M. Lourakis et al., "A graph-based approach to corner matching using mutual information as a local similarity measure", Proc. International Conference on Pattern Recognition, 2004.
- [6] D. Lowe, "Distinctive image features from scale-invariant keypoints", International Jr. of Computer Vision, 60(2), pp.91-110, 2004.
- [7] J. Pluim et al., "Mutual-information-based registration of medical images: A survey", IEEE Tr. On Medical Imaging, vol.22, no.8, pp.986-1004, 2003.
- [8] C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval", IEEE Tr. Pattern Anal. Mach. Intell., 19 (5) pp.530-534, 1997.
- [9] R. Horaud and T. Skordas, "Stereo correspondence through feature grouping and maximal cliques", IEEE Tr. Pattern Analysis and Machine Intelligence, vol.11, no.11, pp.1168-1180, 1989.