Object Recognition Based on Sift Features and A Novel Feature Matching Algorithm

[1] Ehsan Hosein Kalatehjari, [2] Mohammad Mehdi Hosseini, [3] Alireza Behrad

[1] Department of Electronic and computer, Islamic Azad University , Shahrood branch, Shahrood, Iran, Ehsankl@yahoo.com

[2] Department of Electronic and computer, Islamic Azad University , Shahrood branch, Shahrood, Iran, Hosseini_mm@iau-shahrood.ac.ir

[3] Faculty of Engineering , Shahed University Tehran, Iran, behrad@shahed.ac.ir

Abstract

Vision based object recognition is the task of finding object in an image or video sequence by comparing them with an image of that object. Images can be taken from different viewpoints, different scales or even when they are translated or rotated. In this paper, a new algorithm is proposed for object recognition. First the feature points are extracted using scale invariant feature transform (SIFT), then an initial matching is established using photometrical features of the points. A novel geometric based method is used for optimizing matching points. After removing false matches, by using proposed algorithm, remaining key points can be categorized and each distinctive group is assigned to a separated object. The group with maximum likelihood is considered as desirable object. Important innovation in our method is the use of geometrical constraints of feature points between two matching images to find correct matching.

Keywords: *Object Recognition, Geometrical Constraints, SIFT features*

١. Introduction

The automatic object recognition in one of the most important aspects in the machine vision industry for the purpose of inspection, registration and manipulation. In most of applications, it is necessary to detect the desirable object regardless to illumination variation, image scaling and rotation. Therefore several methods have been proposed for the task of object recognition. Schiele [١] used a technique based on multidimensional histograms of the responses of a vector of local linear neighborhood operators and showed that this technique can be used to determine the most probable objects in a scene, independent of the object's position, image-plane orientation and scale. Olshausen and Field [٢] represented receptive fields of simple cells as basic functions that are similar to Gabor filters, and demonstrated that these basis functions can be learnt from images by sparse coding, that maximizes information preservation and sparsity of response. Serre et al. [٣] proposed a method based on a set of features and feed-forward models of object recognition that are resembled in V١ and V٢ of the visual cortex.

Feature extraction is a main stage in most of object recognition methods. One approach needs to detect and match point features like corners which can be extracted using methods like Harris corner detector [۴]. SIFT method is a scale invariant point feature detector [۵], which is employed in several object recognition algorithms. Yang and et al. el [۶] proposed an effective framework based on SIFT-type feature extraction to perform distributed object recognition using a network of smart cameras and a computer as the base station. The method utilizes the available computational power on the smart sensors to locally extract and compress SIFT features to represent individual camera views. In particular, they showed that between a network of cameras, high-dimensional SIFT histograms share a joint sparse pattern corresponding to a set of common features in ٣-D. In [٧] a new approach based on combined set of color descriptors was utilized which outperformed intensity-based SIFT features and improved object recognition accuracy. They studied the invariance properties and the distinctiveness of color descriptors in a structured way and explored the analytical invariance properties of color descriptors. According to new researches, SIFT features are proper feature points that can be used for object recognition. This approach transforms an image into a large collection of local feature vectors that each of features is invariant to image translation, scaling and rotation. These features are partially invariant against illumination changes and affine or ۳D projection.

The main difficulty with SIFT feature extraction and matching is the large number of false matches. In this paper we present a novel geometric method to refine false matches and feature classification.

The paper is organized as follows: section ۲ states feature point extracting and descriptor vector which is used for feature extraction. Section ۳ introduces our proposed method and explain the criteria used in our implementation for feature grouping. Section ۴ discusses the results and open issues for future research.

۲. Feature extraction

To describe an object in the image, some special and unique characteristic of that object are needed so in the first step of the algorithm, feature points are extracted using SIFT method from both of input image and original image. These features are partially invariant against image scaling and rotation. In this section, we briefly review scale invariant feature point's detector, descriptor and the criteria used to measure matching cost.

۲.۱. SIFT Feature

In this stage maxima or minima of a difference- of- Gaussian are used for identifying location of suitable keypoints. For all image points, the value of cornerness function is computed in many successive resolutions or scales (δ_s). The discrete values of scale are distributed exponentially between the inner and outer limits $\delta_s = \delta_{\cdot} r^n$.

These values in ۳D space (x, y, δ), have a local extremum. The extermum points define scale invariant feature point after applying a proper threshold. The detector allocates each feature point a spatial location $X_p = (x_p, y_p)$ and a characteristic scale δ_p. Also, some descriptors need a standard orientation θ_{std} to make them rotation invariant. This orientation is usually the dominant local gradient orientation in a support region around the feature point. The local image gradients are measured at the selected scale in the region around each keypoint. Supported region is rotated such that assigned orientation to feature point lay on a canonical direction. Gradient direction at each sample point in support region is weighted by gradient magnitude. Then, the support region is divided to ۱۶ sub-regions (4×4). An orientation histogram with 8 bins is created by weighted gradient orientations of each sub region. Therefore, a vector with 16×8=128 components is obtained as the descriptor vector of each key point.

To compensate for the inaccuracy of the assigned characteristic scale a Gaussian weighting function is used to reduce weight of gradients that are far from the center of the support region.

Finally the descriptor vector is transformed into a representation that allows for significant levels of local shape distortion and change in illumination. All future operations are performed on image data that has been transformed relative to the assigned orientation, scale, and location for each feature, thereby providing invariance to these transformations.

۲.۲. Initial Matching

In this step descriptor vector is used to find feature point similarity between two images. The nearest neighbor (NN) method is used for finding Key point matching. The Euclidean distance for the invariant descriptor vector is described in the following equation:

$$DIS(j) = \sqrt{\sum_{i=۱}^{۱۲۸} (q_i - q_i^{'})^{۲}} \qquad (۱)$$

which q_i is related to the description matrix of tested feature, $q_i^{'}$ is related to features of reference image and J indicates which feature has been compared. A global threshold on closest distance doesn't have a good performance therefore, ratio between the distances of closest neighbor with respect to distance of the second-closest neighbor has proposed as a measure to reject outliers, by applying a threshold on this ratio significant number of false matches are rejected. At the same time, many correct matches are lost and many incorrect matches remain. In this step a local photometrical criterion in combination with a global geometrical method is used to obtain better results.

۴. Proposed method

After finding some candidate keypoint, they must be refined according to the common characteristics of them. This leads to one or more category that can be compared to the object feature points and make decision which category has the most similarity to that of the object. The geometrical characteristics between keypoints are used for feature categorizing. We proposed a new method for using geometrical characteristic that describe as follow.

۴.۱. Refining False matches using geometric constraints

The false matches obtained in the initial matches may have an unanticipated geometry distribution in two images, and for a loose reject value their number is significantly more than the number of correct matches. To remove false matches, we utilize a search approach which employs geometric distribution of match pairs for removing false matches.

We used geometric constraints, which are rotation, scale and translation invariant to remove false matches. We employ lines between feature points as well as their matches as it is shown in figure ۱, to apply proper geometric constraints. After extracting feature points in the input image and finding matching points in the reference image, the lines between feature points in each image are considered as the elements to apply geometrics constraints.

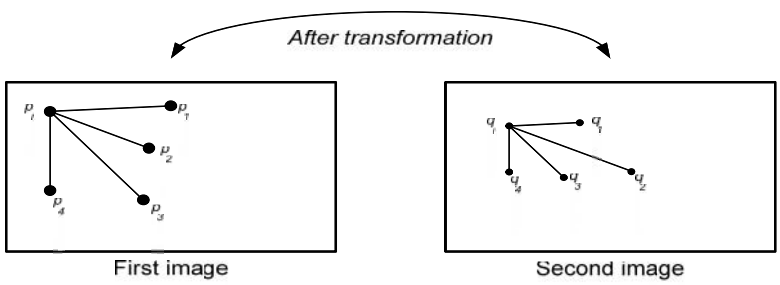


Figure ۱. Geometric constraints for removing false matches

۴.۲. Local scale consistency of lines

Suppose P={p1,p2,…,pM} are feature points extracted from input image, and Q={q1,q2,…,qM} are the corresponding feature points in the reference image, where (pi,qi) are matched pairs obtained using photometric in formation in section ۳. In the case of correct matches, the distances between two points in the input image changes with the same scale factor. We measure the scale factor between two lines as follows:

$$sca(i,j) = \frac{\left\| p_i - p_j \right\|}{\left\| q_i - q_j \right\|} \qquad (۲)$$

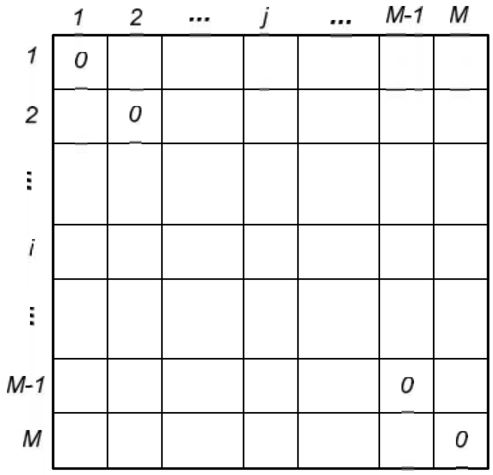


Figure ۲. SCM matrix to remove false matches

Where $\|.\|$ specify Euclidian distance. In the case of correct matches scale factor for all match pairs will be constants.

We utilize the property of the constant scale factor as a geometric constraint to remove false matches. Considering image distortion like image skew and ۳D camera rotation, image scale factor constancy may be not true for the entire image; therefore we designed a method that checks scale factor consistency locally.

Suppose that, there are M match pairs {(pk ,qk): k=۱,۲,..,M} between input and reference images. We consider Scale Consistency Matrix (SCM) as it is shown in Figure ۲. Each element of SCM matrix, SCM(i,j), stores a set of indices representing match pairs that are consistent with scale sca(i,j) as follows:

$$SCM\,(i\,,j\,) = \left\{k\ : \left|sca(i\,,j\,) - sca(i\,,k\,)\right| < \delta\, and\ k\ = ۱,۲,...,M\ \right\} \qquad (۳)$$

Where δ is a tolerance for comparison. As there is no line between i=j, therefore diagonal elements of SCM matrix are set to zero.

The Scale Consistency Matrix (SCM) is obtained, it is used to find false matches based on the following intuitions:

- When all matches are correct, the SCM matrix is symmetric. The asymmetry in the SCM implies false matches.
- *SCM(i,j)* sets with more elements are more likely to represent sets with correct matches and scale factors. We define p(i, j), the probability of validity for *SCM(i,j)* as follows:

$$p(i,j) = \frac{N_{SCM\,(i,j)}}{N_{\max}} \qquad (۴)$$
$$N_{\max} = \max(N_{SCM\,(i,j)}),\ i,j = ۱,۲,..,M$$

where $N_{SMC(i,j)}$ represents the number of elements in set $SCM\,(i,j)$.

- The set $S = SCM\,(i,j) \cap SCM\,(i,k)$ for $k \in SCM\,(i,j)$ defines the set that its elements are consistent geometrically by the match pairs (p_k,q_k) based on the scale factor.

Based on the mentioned intuitions, we have designed an algorithm which detects and removes false matches based on scale consistency of lines. The algorithm has the following stages:

۱. SCM matrix for *M* match pairs $\{(p_k,q_k)$: k=۱,۲,..,M} is calculated.
۲. The probability of validity for *SCM(i,j)* is calculated using the Eq. ۳ and set with lower probability are removed.

۳. The set $S = SCM(i,j) \cap SCM(i,k)$ for $k \in SCM(i,j)$ and different *i, j* values are calculated and the probability of the correct match (p_k, q_k) is estimated as follows:

۴.

$$p(k) = \frac{N_s}{N_{\max}}$$

where N_s is the number of elements in set S.

۵. The index *k* is removed from the set $SCM(i,j)$, if $p(k)<thr$, where *thr* is a predefined threshold.

۶. *SM(i,j)* and *SM(j,i)* are set to $SCM(i,j) \cap SCM(j,i)$.

۷. The algorithm is repeated until no change in *SCM* occurs.

۴.۳. Rotation consistency between lines

Another parameter that is used for finding incorrect matching points is the rotation consistency. The angle between match lines in two images is almost constant under rotation, scale and translation. In this step, we use the following constraint for further removing false matches.

$$rot(i,j) = \measuredangle p_i p_j - \measuredangle q_i q_j = const \qquad (۵)$$

Where $\measuredangle p_i p_j$ represent the angle of vector $p_i p_j$ with respect to horizontal axes. Like the method mentioned before, we consider Rotation Consistency Matrix (RCM), which each element of RCM matrix, RCM (i,j), stores a set of indices representing match pairs that are consistent with rotation rot(i,j) as follows:

$$RCM(i,j) = \left\{ k : \left| rot(i,j) - rot(i,k) \right| < \delta_{۲} \ and \ k = ۱, ۲, ..., M \right\} \qquad (۶)$$

The algorithm to remove false matches using the rotation consistency is the same as scale consistency; however RCM matrix is used in the algorithm.

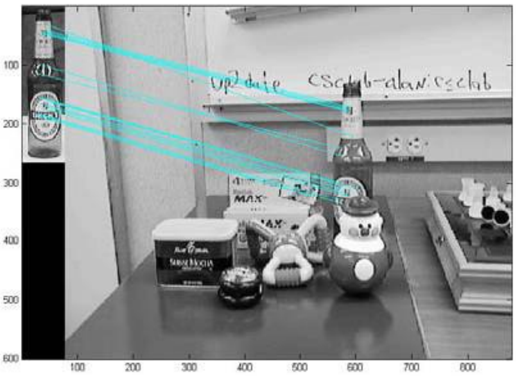
۵. Experimental result

For maximizing true matching results in occluded scenes or images with few correct matching points a new algorithm based on geometrical features is used. The difference between proposed algorithm and other algorithms is increasing in true match results. In exiting algorithms such as Hough transform algorithm or other algorithms that use geometrical features, matched features vote a hash table using rotation and scaling parameter. For finding correct matching points, a model with the maximum vote is picked and matching points that are consistent with the model are selected and other points are eliminated. These algorithm fails when there are high number of false matching points which may be consistent with each other. These matched features may be in different place of the image and different objects and therefore object recognition algorithm fails.

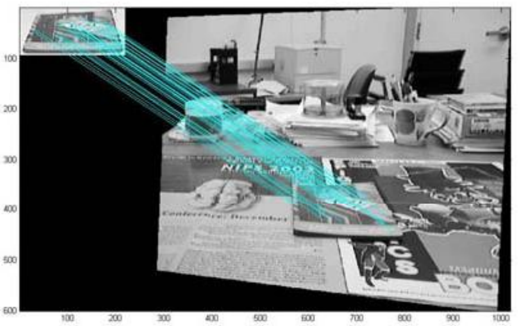
To handle this problem, a new algorithm to refine matching points based on geometrical constraints is introduced that is based on regularity between feature points in rigid object and considers adjacency of feature points. In other word, most of existing approaches doesn't consider adjacency of feature points to refine false match points which make them improper for object recognition.

The proposed algorithm was implemented using a Matlab program and tested with several images. To test the proposed algorithm and compare results with those of other methods, we used CSCLAB image database [۸] which consists of approximately ۵۰ reference objects with ۵۰۰ scenes with significant occlusion.

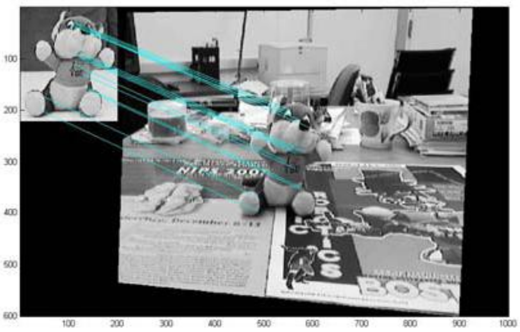
Figure ۳ shows the results of applying SIFT method on several test images. As the figure shows, several false matched points are obvious. Figure ۴ shows the results of refining false matches after applying the proposed algorithm. As shown in the figure, in the final images all matching points are located in the target objects, and false matching points between two objects are removed. Figures ۵ and ۶ show the robustness of the proposed algorithm in the case of rotation and scale change respectively. These figures show the correct matching rate of different algorithms with the change in rotation of the camera around Z axes ad scale change. These figure compare the results of the proposed method with Hough method [۶] and traditional geometrical refining [۷].



a) First image

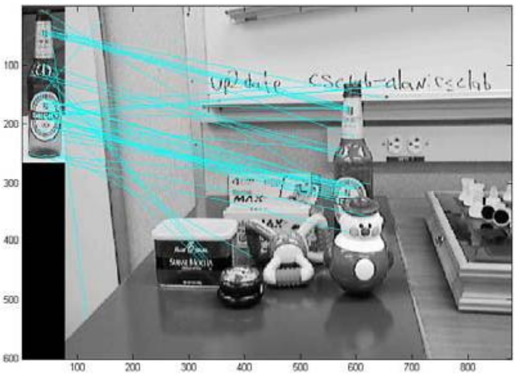


b) Second Image

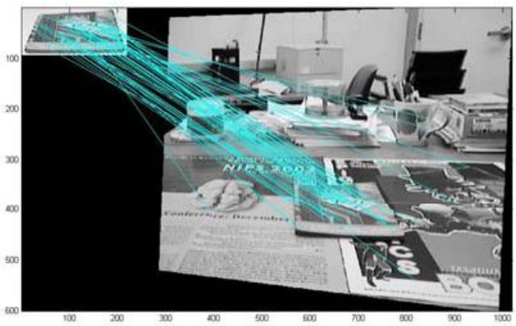


c) Third Image

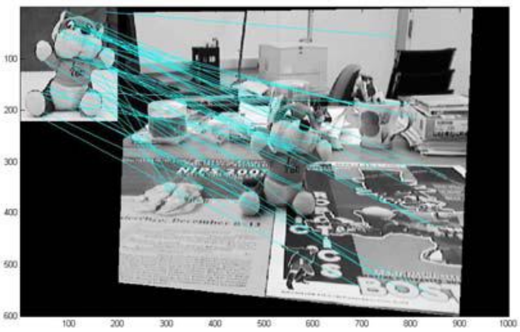
Figure ۴. Output of proposed Algorithm



a) First image



b) Second Image



c) Third Image

Figure ۳. output Sift Algorithm

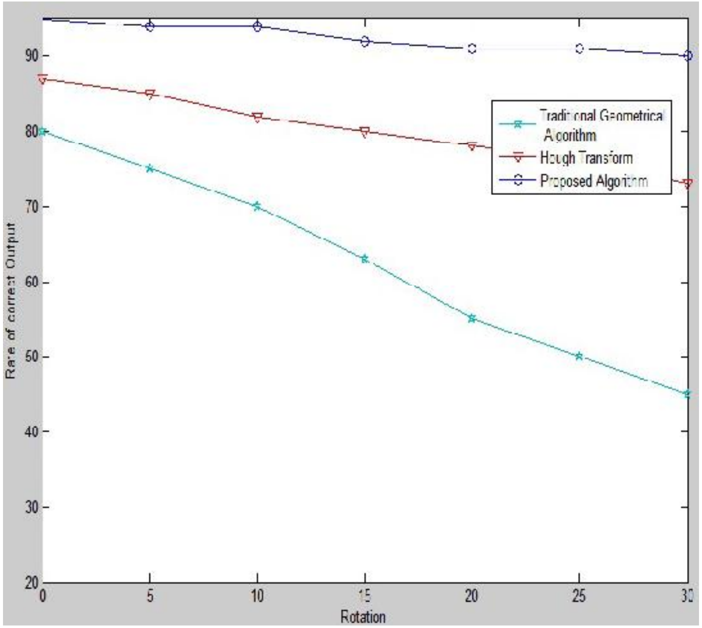


Figure ۵. Rate of correct matched points rate with the change of rotation angle.

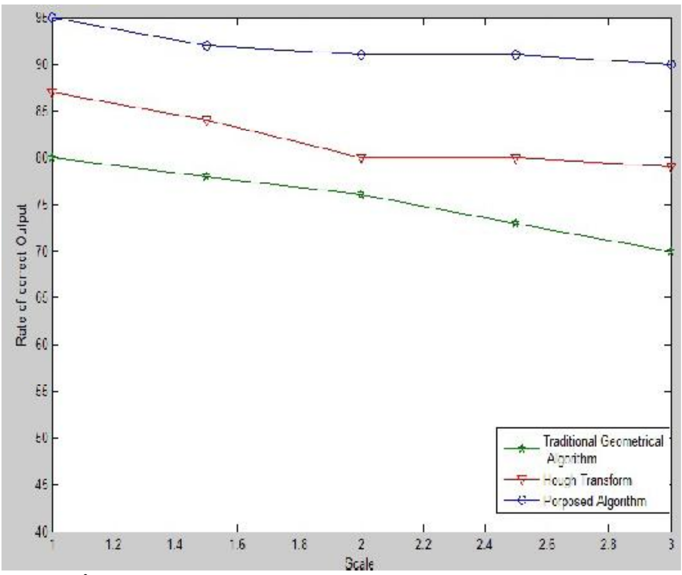


Figure ۶. Rate of correct matched points with the change of scale.

۶. Conclusion

As shown in the experimental results, proposed algorithm has promising results in different condition and is partially robust against rotation and scaling compared to other common algorithm such as Hough transform. Because in those algorithms, all the points have same effect in determining correct matching point regardless of adjacency. The proposed algorithm considers adjacency that groups feature points in different clusters and finds candidate cluster.

Grouping feature points leads to dividing image to smaller parts and therefore nonlinear changes in ۳-D rotation and scaling can be estimated with linear changes. The proposed algorithm was tested with several test images and results showed the efficiency of the proposed algorithm.

۱۰. References

[۱] Bernt Schiele and James L. Crowley, "The Robustness of Object Recognition to Rotation using Multidimensional Receptive Field Histograms," Proceedings of the 13th International Conference on Pattern Recognition,IEEE, Vienna , Austria, pp. 50 – 54, 1996.

[۲] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," Nature, 381, pp. 607-609, 1996.

[۳] T. Serre, L. Wolf and T. Poggio, "Object Recognition with Features Inspired by Visual Cortex," Proceedings of ۲۰۰۵ IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR ۲۰۰۵), IEEE Computer Society Press, San Diego, pp. ۹۹۴-۱۰۰۰, June ۲۰۰۵.

[۴] C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval," IEEE PAMI, pp. 530–۵۳۴, ۱۹۹۷.

[۵] David G. Lowe, "Object Recognition from Local Scale-Invariant Features," ICCV, The Proceedings of the Seventh IEEE International Conference, Greece, pp.۱۱۵۰-۱۱۵۷, ۱۹۹۹.

[۶] Subhransu Maji, C. Mario Christoudias, Trevor Darrell, Jitendra Malik, and S. Shankar Sastry, "Multiple-View Object Recognition in Band-Limited Distributed Camera Networks," Third ACM/IEEE International Conference on ICDSC, Como, pp. ۱-۸, ۲۰۰۹.

[۷] Van de Sande, K.E.A. Gevers, T. Snoek, C.G.M.," Evaluating Color Descriptors for Object and Scene Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.32, No. ۹, pp. ۱۵۸۲ - ۱۵۹۶, Sept. ۲۰۱۰.

[۸] http://csclab.ucsd.edu/labeledimages.php