ISSN 2348 – 8034
Impact Factor- 3.155

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

DOMINANT ORIENTATION IN THE CLUSTERING OF SIFT KEYPOINTS EXTRACTED FROM BUILDINGS WITH A REPETITIVE STRUCTURE

Yong Cheol Kim[*1] **and Sunmin Lee**[2]
[*1,2]Department of Electrical and Computer Engineering, University of Seoul, Korea

ABSTRACT

Similarity matching of SIFT descriptors has been popularly used for object recognition. Conventional keypoint-to-keypoint matching is highly prone to error when the repetitive structure of a building generates a number of keypoints with similar descriptors. We propose a two-step clustering of SIFT keypoints, which can be used in cluster-to-cluster matching for building recognition. The first stage of clustering is on the 128-D local gradient vector and the second stage is on the 2-D coordinates and 1-D dominant orientation. Two-step clustering generates sub-clusters which are well separated both in location and in rotational symmetry. We achieved successful result in grouping of semantically homogeneous keypoints into clusters.

Keywords- *SIFT, mean-shift clustering, relaxation, matching, building recognition*

I. INTRODUCTION

With the expansion of smart mobile devices, the importance of localization of the mobile user is ever increasing. A possible application is to provide specific information about the building in an image taken by a smartphone. The first task for that purpose would be the recognition of a building in the mobile image. In a large city, it may seem implausible because there are a number of buildings which look alike. However, when a rough position of a smartphone is available from built-in GPS function, the search space would be reduced to just tens of buildings. Then, buildings in database images in the approximate position can be compared with the query image to choose the closest view.

We propose a two-step clustering of SIFT (Scale invariant feature transform) keypoints and cluster-to-cluster matching for building recognition. SIFT has been widely used in building detection as well as in object recognition, image stitching, gesture recognition and video tracking [1]. Local salient features are extracted as keypoints. The descriptor provides the information about the local geometrical shape of a keypoint. The gradients are the description for the local shape of a keypoint. They are measured in rotation-compensated image patch in scale-space. The descriptor in SIFT is known to be invariant both in scale and in rotation.

In the original SIFT scheme, the match of a keypoint is found by searching for one with the smallest distance between the gradient vectors. A pair of keypoints is decided to match if the distance ratio between the closest match and the second closest one is below a threshold. In (1), R is the descriptor vector for a keypoint. R_{1st} and R_{2nd} are the descriptor vectors for the closest and the second closest keypoint.

$$\mathrm{Dist}(R, R_{2nd}) < \tau \cdot \mathrm{dist}(R, R_{1st}) \quad (1)$$

Threshold-based SIFT match is effective when an object has fairly discriminative keypoints such that the match strength of a correct one is much larger than the second closest incorrect match. SIFT is highly effective in detecting objects with salient features, but not so powerful in matching modern buildings with repetitive structures such as windows or wall patterns. The descriptors for repetitive keypoints are so similar that a simple searching for the most similar one often fails to find the correct match. SIFT matching between building has inherent ambiguity since its structural repetitiveness produces many keypoints with similar descriptors. There is no guarantee that the candidate with the largest match strength is the correct match. Figure 1 shows the result of keypoint-to-keypoint matching between two images for a building. Solid lines represent the correspondence based only on the similarity of descriptor vectors. Most of them are wrong matches.

ISSN 2348 – 8034
Impact Factor- 3.155

In this paper, we present how to alleviate the problem of ambiguity in matching repetitive structures. We add new features to the conventional SIFT-based matching. First, we upgrade the matching unit, from a keypoint to a cluster of keypoints. When similar keypoints are grouped into a cluster by mean-shift clustering, the difficulty in choosing the closest keypoint out of several similar ones is relaxed if each cluster has a distinct mode vector. Second, we present the result of relaxation-based cluster-to-cluster match. We tested the proposed method on several images of buildings. Two-step clustering generates sub-clusters which are well separated both in location and in rotational symmetry. We achieved successful result in grouping of semantically homogeneous keypoints into a cluster.

II. ROBUSTNESS OF SIFT AGAINST ROTATION

Robustness of SIFT against rotation of objects is obtained through compensation by the magnitude of dominant orientation. Before the assignment of descriptor to a keypoint, the orientation of intensity gradient is computed at each pixel in the neighborhood patch around a keypoint. In the orientation histogram which is quantized in 10° step, the highest peak is chosen as the dominant orientation.

The 128-D gradient vector in a descriptor is a histogram of gradients measured with reference to this dominant orientation. As a result, the gradient vector in the descriptor is a rotation-compensated one. Simply speaking, between two images where one is the rotated version of the other, the descriptors for corresponding keypoints should be the same. This way, SIFT provides invariance against rotation. Figure 1 shows the patch around the matched keypoints between images with camera rotation. The dominant orientation of the keypoint in the left image is $\theta_o = 75^{\circ}$. The right image is rotated by $\theta_r = 10^{\circ}$. The dominant orientation of the corresponding keypoint in the right image is $\theta'_o \fallingdotseq \theta_o - 10^{\circ}$.

III. TWO STEP CLUSTERING OF SIFT KEYPOINTS

The ambiguity in matching can be relieved by using clusters of keypoints as the basic unit of matching. In [2], repetitive features of a large building are grouped into a cluster and then cluster matching is performed. The homography between the two images is estimated by applying RANSAC (random sample consensus) to the keypoints in matched clusters. In RANSAC, several candidates of corresponding pairs of keypoints are selected at random and the 3-D transformation matrix is computed for the correspondences. This process is repeated until we find the exact homography which can accommodate most of the keypoint-to-keypoint matching between the matched clusters. As a result, the computational cost of RANSAC is very high.

We take a different approach. As was done in [2], we group similar keypoints into a cluster and then perform cluster-to-cluster match instead of keypoint-to-keypoint match. The contribution of our work is that the keypoints in a (mother) cluster is clustered again into fine sub-clusters. The 132-D descriptor of a SIFT keypoint is composed of 128-D gradient vector, 2-D coordinates, 1-D scale parameter and 1-D dominant orientation parameter. One drawback of a cluster based only on the gradients is that it often does not correspond to a real entity. For example, the local gradient vector of some keypoints in a tree branch may happen to be very similar to that of keypoints in a window. In that case, the spatial distribution of a cluster may spread over contextually unrelated areas of the image. This problem can be remedied by clustering of keypoints in two steps. The first stage is clustering on the 128-D gradients and the second stage is on 2-D coordinates and dominant orientation.

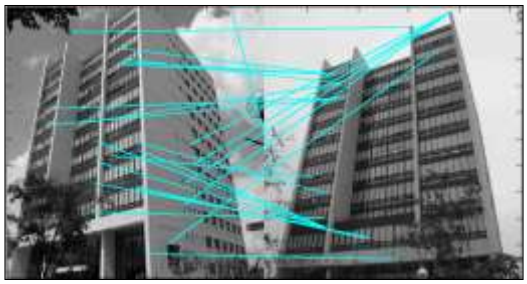

Figure 1. Errors in keypoint-to-keypoint matching based only on the similarity of gradient vector

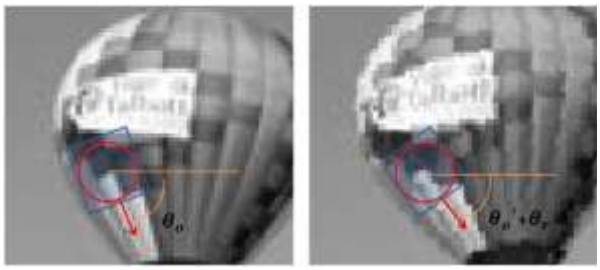

Figure 2. Dominant orientation of SIFT keypoints

ISSN 2348 – 8034
Impact Factor- 3.155

For the grouping of keypoints, mean-shift clustering is used, which is a nonparametric algorithm and thus does not require prior knowledge such as the number of clusters and the shape of the clusters [3]. Mean-shift clustering is a mode-seeking algorithm for feature-space analysis by locating the maxima of a density function. In mean-shift clustering, for n data points $x_i, i = 1,2,\cdots,n$ on a d-dimensional space R^d, the multivariate kernel density estimate obtained with kernel $K(x)$ and window radius h is:

$$f(x) = \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \quad (2)$$

When the kernel is radially symmetric, we define the kernel as follows:

$$K(x) = c_{k,d}\,k(\|x\|^2) \quad (3)$$

The mode vector of the density function is the point where the gradient of the density function, $\nabla f(x)$, is zero. The kernel estimation procedure is successively performed by computation of the mean-shift vector $\boldsymbol{m}_h(x^t)$ and translation of the window $x^{t+1} = x^t + \boldsymbol{m}_h(x^t)$.

The first clustering of keypoints based on 128-D gradient vectors generates several clusters. For a cluster to be a set of features which distinctively characterize the structure of a building, the keypoints in a cluster need to belong to a single local structure and consequently be spatially close to each other. The second clustering aims at this goal. Each cluster from the first stage is divided into smaller sub-clusters by the second clustering on the 2-D coordinates and dominant orientation.

Clustering on the 2-D coordinates generates sub-clusters which consist of spatially close keypoints. This will suffice if the purpose of clustering is simply dividing clusters into semantically homogeneous regions. However, if the clusters are to be used as the matching units in building recognition, we need to be able to distinguish rotationally symmetric features since the spatial distribution of such features is important characteristics of a building. Rotational symmetry of an object is the property of looking the same after a certain amount of rotation. Typical examples are the four corners in a building window. For rotationally symmetric features, the 128-D gradients are not distinguishable from each other since, in SIFT, gradients are measured in rotation-compensated patch. Clustering on the dominant orientation component helps to distinguish them.

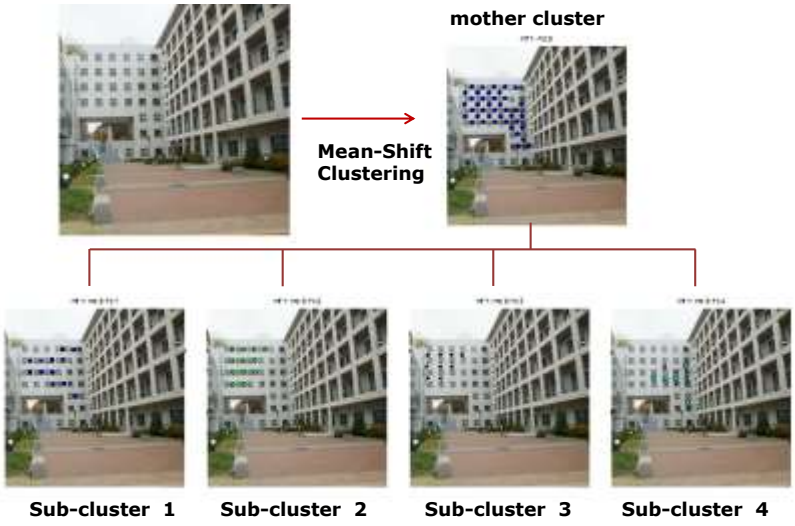

Figure 3. Result of two-step clustering

ISSN 2348 – 8034
Impact Factor- 3.155

IV. RESULTS OF CLUSTERING WITH DOMINANT ORIENTATION

In the first stage, mean-shift clustering of SIFT keypoints generates several mother clusters. In the second stage, the keypoints in each of them are mean-shift clustered again into fine sub-clusters. Figure 3 shows an example. The first stage of clustering generated nine mother clusters. The keypoints extracted from the window boundaries on front side of the left building constitute the eighth mother cluster, which is being divided into four sub-clusters.

An enlarged view of the sub-clusters is shown in figure 4. The second clustering is based on 2-D coordinates and dominant orientation. We can find that the sub-clusters either reside in separated space (between sub-cluster 3 and sub-cluster 4) or they have distinct characteristics of keypoints (between sub-cluster 1 and sub-cluster 3). Hence, rotationally symmetric keypoints are well separated.

A comparison of matching between mother clusters and between sub-clusters is shown in Fig.5 (a) and in Fig.5 (b). Matching between sub-clusters produces far more meaningful correspondence [4]. Relaxation-based matching in [5] is used to find the correspondence between clusters. Relaxation is a labeling procedure where ambiguities in matching of local features are iteratively reduced by exploiting the global consistency of contextual information of corresponding features [6].

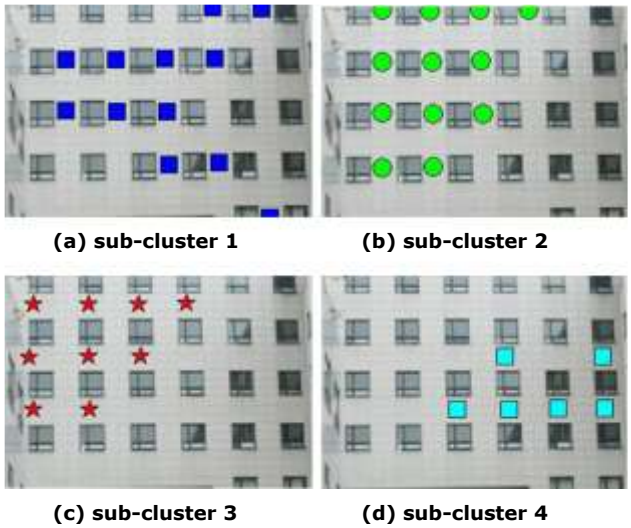

Figure 4. Sub-clusters either reside in separated space or have distinct dominant orientation.

(a) (b)

Figure 5. (a) Matching between clusters generated from conventional clustering (b) Matching of clusters from two-step clustering (Red circle represents a rough area of keypoint distribution)

ISSN 2348 – 8034
Impact Factor- 3.155

In relaxational matching, the initial probability of matching of a feature in one image is computed against all the features in the other image, based only on the similarity of features in both images. This matching probability is iteratively updated, by examining the structural consistency of matching pairs. Consistent pairs support each other as their matching probabilities get larger at each iteration.

V. CONCLUSION

Though SIFT is highly effective in detecting objects with salient features, it not so powerful in matching repetitive structures since descriptors for repetitive keypoints are very similar to each other. A simple searching for the one with the most similar descriptor often fails to find the correct match. If similar keypoints are grouped into a cluster, then cluster-to-cluster matching can lead to better performance in recognition than keypoint-to-keypoint matching.

We proposed two-step clustering of SIFT keypoints. In order to extract groups of semantically homogeneous keypoints, mean-shift clustering is performed in two steps. First, keypoints are clustered based on the similarity of the 128-D local gradient vectors. The keypoints within each mother cluster from the first step are mean-shift clustered again. In the second step, the proximity of 2-D coordinates and the similarity of the dominant orientation are the main factors of clustering. Close-by keypoints are grouped into a cluster, due to the proximity of 2-D coordinates. Dominant orientation is effective in separating rotationally symmetric keypoints, such as the four corners of a rectangular window. We tested the proposed clustering algorithm to several images for ten buildings in the University of Seoul. Database of 72 test images with 512 x 512 pixels have been taken at three months intervals. Each set of 24 images was shot under different illumination condition in each season. Relaxation-based matching on the two-step clusters significantly improved the quality of correspondence.

VI. ACKNOWLEDGEMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (grant number NRF-2013R1A1A2012745).

REFERENCES

1. *D. Lowe, "Distinctive Image Features from Scale Invariant Keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, November, 2004, pp.603–619*
2. *S. Ha, S. Kim and N. Cho, "Image Registration by Using a Descriptor for Repetitive Patterns," IEEE Proceedings of Visual Communications and Image Processing, pp. 1-6, San Diego, November, 2012*
3. *D. Comaniciu and P. Meer, "Mean Shift: a Robust Approach Toward Feature Space Snalysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.24, no.5, May, 2002, pp.603–619*
4. *S. Lee and Y. Kim, "Two-Step Clustering of SIFT Keypoints and Relaxation Based Matching of Clusters," 2015 ACM Research in Applied Computation Symposium, Prague, Czech Republic, October 9-12, 2015*
5. *R. Hummel and S. Zucker, "On the Foundations of Relaxation Labeling Processes," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 5, no. 3, May, 1983, pp. 267-287*
6. *S. Barnard and W. Thompson, "Disparity Analysis of Images", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.2, no.4, July, 1980, pp. 333–340*