Multiple Object Tracking Using SIFT Features and Location Matching

Seok-Wun Ha[1], Yong-Ho Moon[2]

[1,2] *Dept. of Informatics , Engineering Research Institute, Gyeongsang National University, 900 Gazwa-Dong, Jinju, Gyeongnam, Rep. of Korea*
swha@gnu.ac.kr, yhmoon5@gnu.ac.kr

Abstract

In recent, object recognition and tracking systems have been developed that use local invariant features from Shift Invariant Feature Transform algorithm. Most of them are implemented by distance matching of descriptor features between the reference and the next consecutive frame image. Among the matched keypoints generated from SIFT descriptor matching, there are some mismatched keypoints when keypoint location information is considered. These location-mismatched keypoints could be affected to object tracking perfomance. To achieve a stable tracking it is necessary that these are detected and discarded in tracking action. In this paper a robust object tracking system is presented that strengthen stability in tracking by eleminating location-mismatched keypoints. Experimental results show that a stable and robust tracking can be achieved in a test video sequence includes multiple objects.

Keywords: *SIFT, descriptor, location-matched, keypoint, tracking.*

1. Introduction

Tracking of objects in a video is an important research topic in video analysis applications such as computer vision, surveillance, and scene understanding. Various types of local feature descriptors such as SIFT[1], GLOH[2], SURF[3] and others[4-5] were used to represent characteristics of objects. Especially SIFT(Shift Invariant Feature Transform) proposed by D. G. Lowe[1] was known as an algorithm that generates local features robust to changes in image scale, noise, illumination and local geometric distortion.

To apply this algorithm to object tracking SIFT keypoints of objects are first extracted from a reference image and stored in a database. Then SIFT keypoints from object in next image are compared with the stored reference keypoints based on Euclidean distance of their feature vectors and an object is recognized that are featured by the matched keypoints. But among the matched keypoints there are mismatched keypoints in their location coordinates and these mismatched generate an unstable tracking or out of tracking.

Lowe[6] presented in his object recognition implementation that all matches are rejected in which the distance ratio is greater than 0.8, which eliminates 90% of the false matches while discarding less than 5% of the correct matches. Therefore the 10% of the false matches could be a serious impact to erase an unstable tracking in small objects in particular.

In this paper, we propose a robust method to increase stability of the tracking by detecting the location-mismatched keypoints and restraining their participation to the keypoint matching. And we applied a tracking window that determines the size and position of the

window by reflecting the distance ratio between two location-matched keypoints on the objects each the reference and the tracking frame images.

This paper is organized as follows: In section 2 the SIFT algorithm is introduced which is widely used in object recognition system. Section 3 introduces the proposal method utilizing the location information of keypoints. Section 4 presents the procedure for object detection and location matching. Section 5 shows the application results of the proposed method to the object tracking in a video sequence. Finally, conclusions and future works are given in section.

2. Scale Invariant Feature Transform

All SIFT algorithm proposed by Lowe[6] have the major stages of computation used to generate the set of image features:

1) Scale-space extrema detection: The first stage of computation searches over all scales image locations. It is implemented efficiently by using a difference-of-Gaussian function to identify potential interest points that are invariant to scale and orientation.

2) Keypoint localization: At each candidate location, a detailed model is fit to determine location and scale. Keypoints are selected based on measures of their stability.

3) Orientation assignment: One or more orientations are assigned to each keypoint location based on local image gradient directions. 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.

4) Keypoint descriptor: The local image gradients are measured at the selected scale in the region around each keypoint. These are transformed into a representation that allows for significant levels of local shape distortion and change in illumination.

In the last step, we can earn the descriptor vectors that are composed of histograms computed from gradient magnitudes and orientations of neighbor points in window around each keypoint. For most applications using SIFT feature these descriptor vectors are applied to the distance matching between keypoints in two interesting objects. But some of the matched keypoints are not matched in their locations and these are necessary to be discarded not to be affected to the stable object tracking. In this paper we did not change the properties of the SIFT and just used the location information of the keypoints.

3. The Proposed Algorithm

In this paper, we propose an algorithm to achieve a stable multiple objects tracking by only using location-matched keypoints among the candidate keypoints generated from SIFT processing. In our algorithm, object tracking is performed by a temporal rectangle window around the object at the reference frame and the current frame. Key concept is to find and exclude the location-mismatched keypoints among the candidate keypoints from SIFT.

A schematic diagram of the proposed algorithm is shown in figure 1. Initially, rectangle windows around objects in the reference frame and candidate keypoins are generated using SIFT and their local features are stored. Then about the consecutive next frame, keypoints are generated by the same process and matching keypoints are selected on a distance ratio basis. These SIFT matched keypoints become candidates for tracking and among these candidates the location-

matched keypoints are determined. Finally, based on the location-matched keypoints a stable tracking for multiple objects is performed on the overall sequences in a test video.

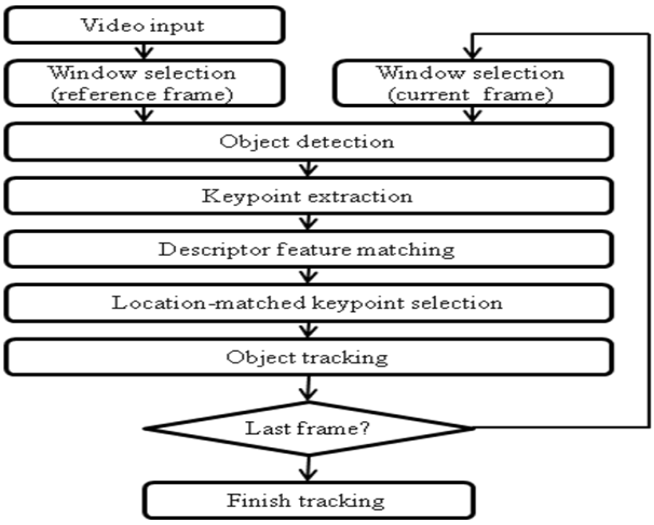

Fig. 1. A Schematic Diagram of Our Proposed System

4. Object Detection and Location Matching

4.1 Object Detection

Object tracking is performed by temporal tracking of a rectangle window around an object on a reference and a tracking window in continuous frame images.

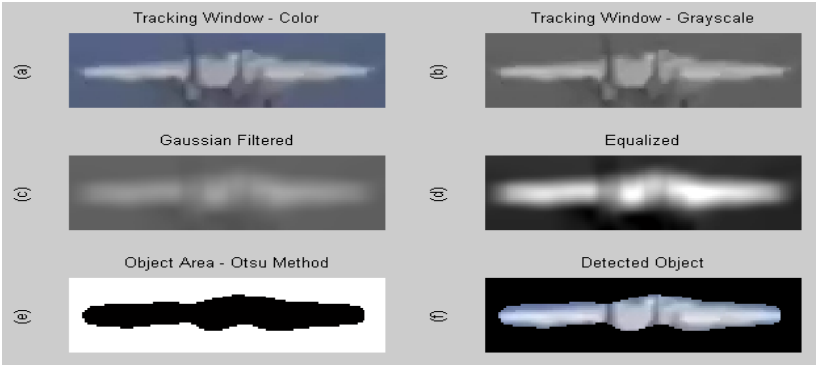

Fig. 2. Results on steps. (a) Original image in window, (b) Grayscale image, (c) Gaussian low-pass filtered image, (d) Equalized image, (e) Thresholded, and (f) Detected object.

Figure 2 shows the results on each steps First the object is detected in the window using the following steps:

1) convert the color images inside the reference and the tracking window to the grayscale images.

2) convolve the window area with a Gaussian low-pass filter, where the filter has a standard deviation and a size of 9.

3) equalize the low-pass filtered window to enhance the object area.

4) separate and detect the object from the equalized window using Otsu's method.[7]

4.2 Keypoint Extraction

In our method, in order to compare local features in both windows on the reference frame and current frame image it is needed to extract keypoints in the windows and it is done from SIFT processing as follow:

$$[im, des, loc] = SIFT(image) \quad (1)$$

Through this SIFT processing three types of data are provided. Data '*im*' includes pixel values of the test image, '*des*' has a matrix of descriptor vectors, and '*loc*' has location, scale, orientation values for all detected keypoints. Most cases of distance matching between keypoints using SIFT utilize the '*des*' data.

All of the keypoints from SIFT are separated to two parts of the object and the background. Because the background could be changed during movement of the object according to the next consecutive frames, it must be limited for the background keypoints to participate to the tracking on the whole sequence.

$$p_{keypoiint}(i, j) = \begin{cases} participate & if\ (i, j) \in object\_area \\ not & if\ (i, j) \in background\_area \end{cases} \quad (2)$$

Figure 3(a) shows that there exist multiple keypoints on the background among the keypoints generated from SIFT computation and 3(b) represents keypoints included in the object after the background keypoints eliminated.

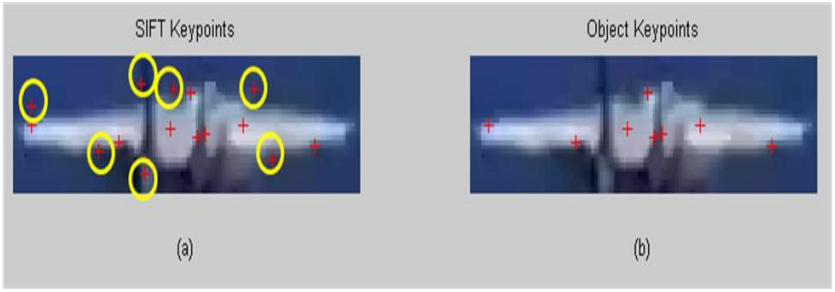

Fig. 3. Distribution of the keypoints - (a) Before and (b) after elimination of the background keypoints. Keypoints shown in (b) are used to matching and tracking. and a tracking window in continuous frame images.

4.3 Keypoint Matching

Matching between keypoints extracted from object area of the reference and the current window is measured by comparing the distance of the closest neighbor to that of the second-closest neighbor. The process to achieve matching is as follows:

1) Distance between the *m* keypoints of the reference object area and the *n* keypoints of the current object area are calculated from dot production of the descriptor vectors. distance d_{ij} between *i*th keypoint descriptor vector des_{Ri} in reference frame and *j*th keypoint descriptor vector des_{Cj} in current frame is as follow.

$$d_{ij} = cos^{-1}(des_{Ri} \bullet des_{Cj}) \quad (3)$$

2) Distances d_{ij} are sorted and distance ratios between the distance of the closest neighbor to that of the second-closest neighbor are calculated.

$$distRatio = \frac{the\ closest\ distance}{the\ second-closest\ distance} \quad (4)$$

3) And all matches are rejected in which the distance ratio is greater than 0.8, which 90% of the false matches while discarding 5% of the correct matches as Lowe[6].

$$match = \begin{cases} accept & if\ distRatio \le 0.8 \\ reject & if\ distRatio > 0.8 \end{cases} \quad (5)$$

The keypoints accepted from the above matching are stored and became candidates to participate in tracking.

4.4 Selection of Location-matched Keypoints

Among the candidate keypoints accepted from distance matching of previous section there exist multiple candidates mismatched considering their location in both windows. To find these location-matched keypoints location difference d_{RC} between the corresponding candidates in both windows is calculated and if the location difference d_{RC} is less than or equal to the threshold d_{th} the candidates are selected as a location-matched keypoints.

$$d_{RC} = \sqrt{(x_C - x_R)^2 + (y_C - y_R)^2} \quad (6)$$

$$location-matched = \begin{cases} yes & if\ d_{RC} \le d_{th} \\ no & if\ d_{RC} > d_{th} \end{cases} \quad (7)$$

where (x_R, y_R) is the location coordinate of the candidate in the reference frame and (x_C, y_C) is that of the corresponding candidate in the current frame.

In figure 4, the number of keypoints per frame about SIFT-matched, locatin-matched, and location-mismatched cases are plotted for the threshold $d_{th} = 3$. Figure 5 shows examples of the location-matched and location-mismatched case for two aircrafts in a video sequence. And the table 1 represents the average number of matched-keypoints on the overall frames according to a variety of d_{th}. From this table, we chose $d_{th} = 3$ as the appropriate value.

In figure 4(a), the average number of keypoints in case of SIFT-matched for entire sequence is 7.63 and this a few number is due to the small object. 4(b) and 4(c) shows that among this SIFT-matched keypoints in 4(a) some are location-matched and some are

location-mismatched, and the average number of location- matched keypoints is 4.73 and that of location- mismatched keypoints is 2.90. Therefore it is estimated that about 62% of the SIFT-matched keypoints will contribute to object tracking and they make the tracking to be stabled because the location–mismatched keypoints would be discarded in tracking .

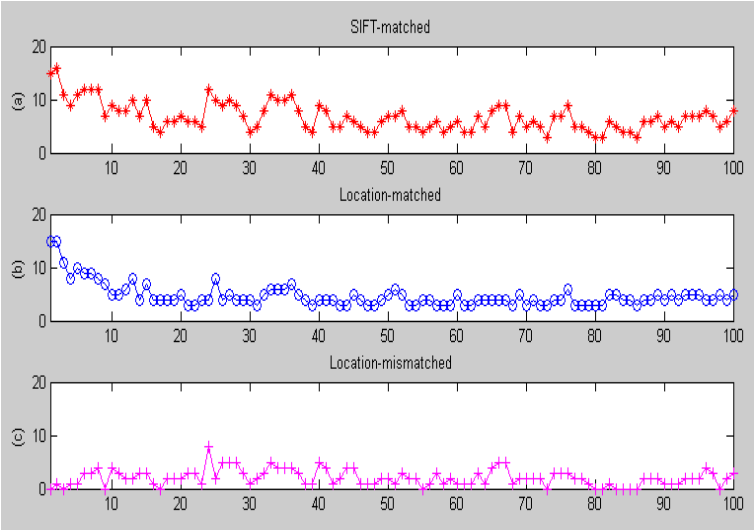

Fig. 4. The number of keypoints for 100 frames. (a) SIFT-matched, (b) Location-matched, and (c) Location-mismatched

Figure 5 shows the results of the keypoint detection for three cases of SIFT-matched, location-matched, and location-mismatched. In this figure the number of SIFT-matched keypoints in the object is 8 in total and among these only one keypoint is detected as a location-mismatched keypoint and the rest 7 keypoints are location-matched.

Table 1 presents the average number of the keypoints on the overall frames according to a variety value of the threshold d_{th}. Here the threshold means the tolerance pixel distance from the center of the keypoints and they are just applied to both cases of location-matched and mismatched. From this table we choose the value 3 as a threshold value and this illustrates that if the distance between two SIFT-matched keypoints are less than or equal to this threshold they are location-matched and if the distance are greater than the threshold they are location-matched.

Table 1. The average number of SIFT matched, location-matched, and location-mismatched keypoints on the overall frames according to a variety of d_{th}

	Threshold					
	1	2	3	4	5	6
SIFT-matched	7.69	7.62	7.63	7.63	7.63	7.63
Location-matched	4.01	4.51	4.73	4.91	4.95	5.13
Location-mismatched	3.68	3.11	2.90	2.72	2.68	2.50

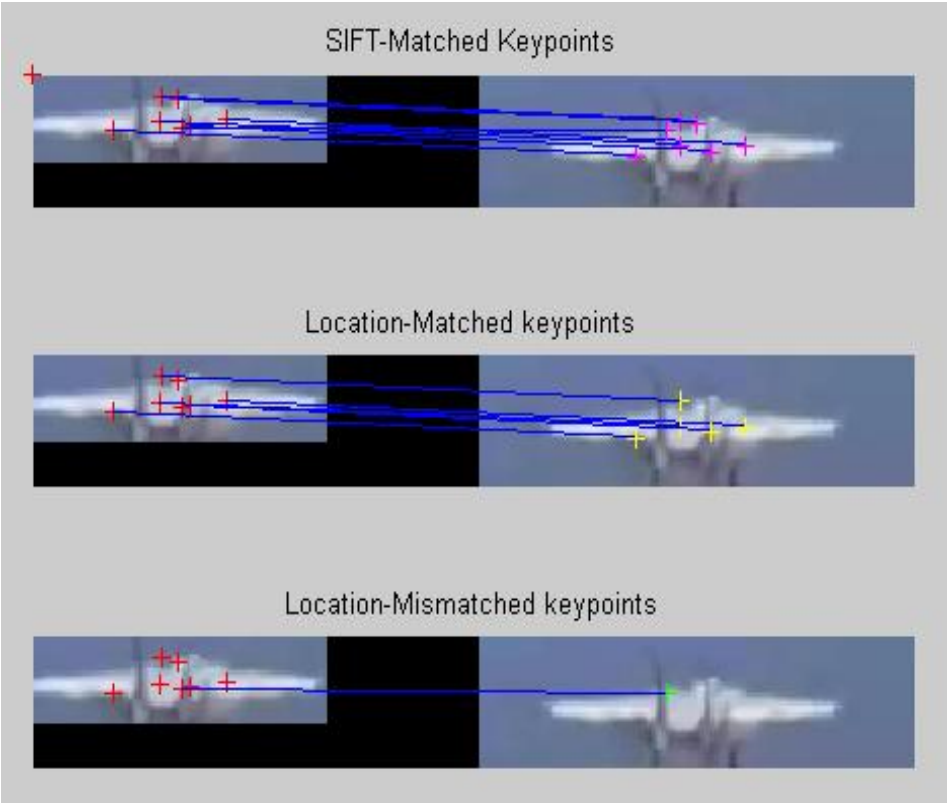

Fig. 5. Example of keypoints detection of SIFT-matched, Location-matched, and location-mismatched.

5. Experimental Results of Object Tracking

The proposed algorithm has been tested on a video sequence-'aircrafts'- includes two aircrafts'. The tested sequence has 132 frames of 376*504 pixels and two aircrafts are landing on the Mothership.

The experiments are operated on two cases of SIFT-matched not considering location of the keypoints and location-matched considering location. Figure 5 represents object detection and tracking states in the window for 8, 26, 54, and 83th of total 132 frames. Figure 6(a) shows the original first frame image as a reference frame, 5(b) shows tracking in case of SIFT-matched not considering location, 5(c) shows tracking considering location. In this figure, we can find the more stable performance in case of tracking using location-matched keypoints.

(a) SIFT-matched tracking

(b) Location-matched tracking

Fig. 6. Example of multiple objects tracking of SIFT-matched, Location-matched.

And the distances between the center of the tracked object and the desired object center are measured. Figure 7 represents object evaluation for difference from desired center. In this figure, maximum distances from the expected center are 41 and 5 in SIFT-matched and location-matched each.

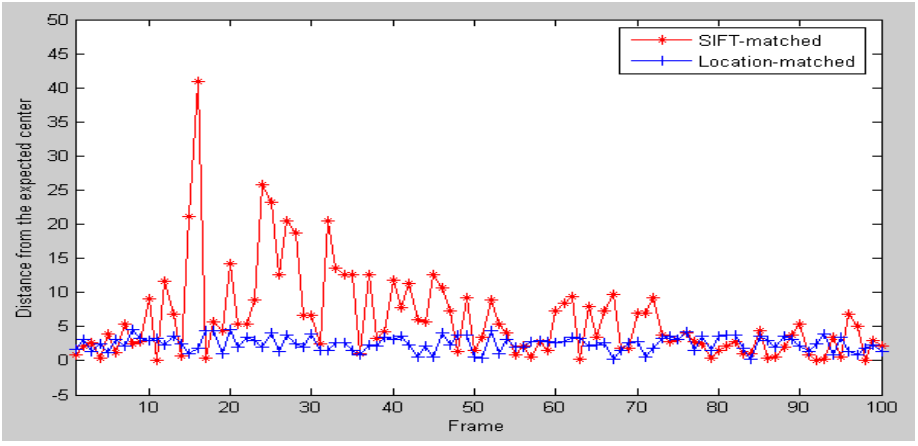

Fig. 7. Distance of the object centers from the expected center in SIFT-matched and location-matched tracking.

6. Conclusion

SIFT descriptor feature and its matching is widely used in object recognition and object tracking of computer vision fields because of its distinctive robust invariant characteristics. But according to distance ratio's value it's performance are not stable due to the false matched keypoints, and it is serious in case of small object tracking especially. So we proposed the simple stable tracking method to increase the tracking effect by discarding location-mismatched keypoints and by only participating location-matched keypoins in tracking. Based on the experimental results, it was founded that the proposed tracking algorithm has the more stable and robust performance. In future we would research a method that the tracking performance could be further improved by adaptively controlling the tracking window size according to variation of the object size.

Acknowledgments

This research was supported by the MKE (The Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency) (NIPA-2011-C1090-1031-0007) and This work was supported by the fund of Research Promotion Program(RPP-2009-000), Gyeongsang National University.

References

[1] Lowe, D. G., "Object Recognition from Local Scale-Invariant Features", Proc. Of the International Conference on Computer Vision, pp. 1150--1157 (1999)

[2] Mikolajczyk, K. and Schmid, C., "A Performance Evaluation of Local Descriptors", IEEE Transactions on Pattern Recognition (2004)

[3] Bay, H., Tuytelaars, T. and Gool, L. V., "SURF: Speeded Up Robust Features", Proc. Of the ninth European Conference on Computer Vision (2006)

[4] Ke, Y. and Sukthankar, R., "PCA-SIFT: A More Distinctive Representation for Local Image Descriptors", Computer Vision and Pattern Recognition (2004)

[5] Laptev, I. and Lindeberg, T., "Local Descriptors for Spatio-Temporal Recognition", Workshop on Spatial Coherence for Visual Motion Analysis. Vol. 3667, pp. 91—103 (2004)

[6] Lowe, D. G., "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, vol. 60-2, pp. 91--110 (2004)

[7] Otsu, N., "A threshold selection method from gray-level histogram", IEEE Transactions on systems Man Cybernet, SMC-8 pp. 62-66, (1978).

Authors

Seok-Wun Ha

Received the B.S.,M.S. and Ph.D. degrees in Department of Electronics from Busan National University , Korea in 1979, 1985, 2005 respectively. He was a research scholar of VISLab in University of California, Reverside from 2002 to 2003. Since 1993 he has been a professor in Department of Informatics and a member of Engineering Research Institute in Gyeongsang National University.

Yong-Ho Moon

Received the Ph.D. degree in Department of Electronics from Busan National University, Korea in 1998. Since 2007 he is a associate professor in Department of Informatics and a member of Engineering Research Institute in Gyeongsang National University.