Research of Object Recognition and Tracking Based on Feature Matching

Hyochang Ahn and Sang-Burm Rhee*

Dept. of Applied Computer Engineering, Dankook Univ.,
Yongin-si, Gyeonggi-do, Republic of Korea
youcu92@dankook.ac.kr

Abstract. Object recognition and tracking are very important task in several computer vision applications in our life. Most of feature matching approaches have problems which are high computational complexity and weak robustness in various environments. In this paper, we proposed a low complexity and robust object recognition and tracking using advanced feature matching for real time environment. Our algorithm recognizes object using invariant features and reduces dimension of feature descriptor to deal with the problems. Our experiments demonstrate that our work is more fast and robust than the traditional methods and can track object accurately in various environments.

Keywords: object tracking, object recognition, feature matching.

1 Introduction

The video surveillance and security monitoring system have developed rapidly in order to monitor several circumstances of the public area in recent years, due to they have great powered performance and precision. One of most active research field related those system object recognition and tracking. In order to track a dynamic object in video sequence, firstly, feature points are extracted in the interest object, and then target object is recognized by extracted features[1][2]. Finally detected object tracking is carried out[3][4]. It is very important technology in computer vision. The object recognition consists of two main steps. One is extracting of interest feature point in target object. The other is matching corresponding point at target video sequence. In object recognition based on feature, extraction of accurate feature in target has influenced the performance of the object recognition. The performance of recognition can be improved by a large number of feature points which are extracted interest region. However, it is impossible to recognize the object in real time environment owing to increased computation complexity. On the other hand, it cannot carry out precise recognition. Therefor an algorithm of object recognition requires detecting accurate object and decreasing computational complexity in real time surveillance system. The interest object tracking is used CamShift[4][5][6]. CamShift is the object tracking method by color histogram as its target model[7][8]. In this paper, we proposed a low complexity and robust object recognition and tracking using

* Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Computer Science and Its Applications*,
Lecture Notes in Electrical Engineering 330, DOI: 10.1007/978-3-662-45402-2_152

advanced feature matching for real time environment. The rest of this paper is organized as follows. Section 2 gives a short survey of some related works for feature extraction. Section 3 concentrates on our proposed method for for object tracking by CamShift and object recognition by advanced feature extraction. The object recognition and tracking implements and experimental results follow in section 4. Finally, section 5 presents our conclusions and future work.

2 Related Works

The video surveillance system is mainly composed of object recognition and tracking. In object recognition, the most important thing is accurately feature extraction. Feature extraction methods are generally used SIFT and SURF in object recognition [9][10][11]. SIFT is a method of transformations into set of regional characteristics with robust feature against the object's size, rotation and projection[9]. SIFT method as image size increases will be plenty to calculate the amount of data, because of the high-dimension characteristic[12,13]. If the amount of data is increased, the method has the problem which increased amount of computation time as increasing volume of calculation[14]. SURF is presented a faster method using Fast-Hessian detector[15]. SURF has fast feature extraction and feature descriptors to reduce the complexity of the operation in feature extraction and matching step on SIFT method[8]. It is performed good result and high speed using feature extraction methods and feature descriptors by decreasing processing time. However, SURF is plenty of computational time for object recognition in real time. Thus, we need to development an algorithm to perform precise object recognition while reducing the amount of computational time. The CamShift is one of the most important algorithms for object tracking. The CamShift is an adaptation of the Mean Shift algorithm in computer vision. The primary difference between CamShift and MeanShift algorithm is that CamShift uses continuously adaptive probability distributions while MeanShift is based on static distributions, which are not updated unless the target experiences significant change in shape.

3 Proposed Method

The CamShift is an algorithm to track interest object for real time environment. The method is only used color features, so it is not robust to the surrounding environment and illumination. This algorithm has the problem that lost interest target when similar color existing in the background. In this case, there is a need for a way to find the object by using the feature points. SURF can find out target object using feature extracting and matching. Fast and accurate object recognition needs to find out matching points efficiently in real time environment. The proposed method extracts the features and find out the corresponding matching points in video sequence. Also we reduce dimension in feature descriptor to effectively decrease computational complexity to carry out object recognition.

3.1 Object Tracking

Continuously adaptive mean shift (CamShift) algorithm is an adaptation of the Mean Shift algorithm for object tracking [3][4]. The basic idea of the CamShift method is that perform MeanShift operations for frame of the video sequence. The CamShift method makes the result of last frame to be as the initial value of the next frame for MeanShift algorithm, and performs those step iterative [5][6]. In CamShift, the important process is what can adaptively adjusted target region in order to continue the tracking, when the target size is changed. The CamShift algorithm can be described in the following steps[5][6]. Step 1 is that determine the initial position and size of the search window. Step 2 is that calculate the distribution of intensity in the center of the search window. Step 3 is that repeat the MeanShift algorithm. Step 4 is that change into the center of distribution of intensity that is obtained the position and size of the search window in the following frame at step 3. Final step is that repeat step 2 to 5. After object tracking, the size and angle of the target in the image, can calculate the first and second moment of distribution of intensity in the search window. It can be expressed as the equation (1).

$$
\begin{aligned}
Z_{20} &= \sum x \sum y[x^2 I(x, y)] \\
Z_{02} &= \sum x \sum y[y^2 I(x, y)] \\
Z_{11} &= \sum x \sum y[xyI(x, y)]
\end{aligned}
\tag{1}
$$

where $I(x, y)$ denotes the pixel value at (x, y) in the image, and x and y is represented the range of search window. For the convenience of calculation, the intermediate variables, a, b and c, define as equation (2).

$$
a = \frac{M_{20}}{M_{00}} - x_c^2 \qquad b = 2\left[\frac{M_{11}}{M_{00}} - x_c y_c\right] \qquad c = \frac{M_{02}}{M_{00}} - y_c^2 \tag{2}
$$

where $(x_c,\ y_c)$ denotes the center of search window.

The horizontal and vertical size and the angle which are detected distribution of intensity in the search window is calculated as following equation (3).

$$
\begin{aligned}
l &= \sqrt{\frac{(a+c)+\sqrt{b^2+(a-c)^2}}{2}} \\
w &= \sqrt{\frac{(a+c)-\sqrt{b^2+(a-c)^2}}{2}} \\
\theta &= \frac{1}{2}\arctan\left[\frac{b}{a-c}\right]
\end{aligned}
\tag{3}
$$

3.2 Object Recognition

First step of our scheme is quickly extract features using the extractor based on Hessian matrix. In this case, the extractor based on Hessian matrix extract features of images for changing of various scale by resize of box filter without changing of image scale. Hessian matrix can be obtained by convolution of the second derivative of Gaussian filter and an image, and it can be expressed as the equation (4) [15].

$$H(X,\sigma)=\begin{bmatrix} L_{xx}(X,\sigma), & L_{xy}(X,\sigma) \\ L_{xy}(X,\sigma), & L_{yy}(X,\sigma) \end{bmatrix} \tag{4}$$

where $L_{xx}(X,\sigma)$ denotes the convolution of the second derivative of Gaussian filter and an input image at the point of $X = I(x, y)$ in an input image having a scale of σ . In addition, $L_{xy}(X,\sigma)$ and $L_{yy}(X,\sigma)$ is represented convolution of the second derivative of Gaussian filter and an input image for xy direction (diagonal) and y direction (vertical). However, there will be occurred problems which increased processing time and slow down its speed as large amount of processing time. The method uses the box filter adapted approximation of convolution of the second derivative of Gaussian to solve problem increasing processing time[16].

The feature descriptor of accurate and robust feature points for rotation is needed to extract a lot of complexity. The conventional algorithms using 64-dimension descriptor have superior performance for exactly finding out feature points. However the methods are not suitable for real time environment, since computational complexity of them for extracting the feature points is high. Therefore, the reduction of dimension in feature descriptor is necessary to effectively decrease computational complexity to carry out object recognition in real time environment[12].

The reduction of dimension in feature descriptor is used calculated direction vector through scale s to determine dominant orientation and expanding its window. First, the rectangle window is made up by a circular neighborhood of radius 15σ around the feature points.

At this time, a rectangular window centered on the direction of the feature points according to the position the dominant orientation. The interest region is divided into sub-region as shown in Figure 3. 3×3 sub-region is re-divided into 5×5 sub-region. As equation (5), the feature descriptor 18($3\times3\times2$) dimension descriptor in segmented regions of makes up with two feature vectors

$$V_{sub} = [\sum dx, \sum dy] \tag{5}$$

We calculate sum of the Haar wavelet responses in horizontal(dx) and vertical (dy) direction. Our method generates suitable accurate and robust descriptor for scale and rotation through expansion of orientation window and reduction of dimensions in real time environment. It also decreases computation complexity.

4 Experimental Result

We have experimented object recognition and tracking method under windows with the Core i3 CPU@3.30GHz and 4GB RAM memory in Windows 7, and have implemented in Visual Studio 2013. We have test many video sequence. The computing task focuses on target modeling and matching, and concentrates feature point detection and matching, and is in proportion to tracking window size. Figure 4~8 have shown the result images through proposed method in real time environment.

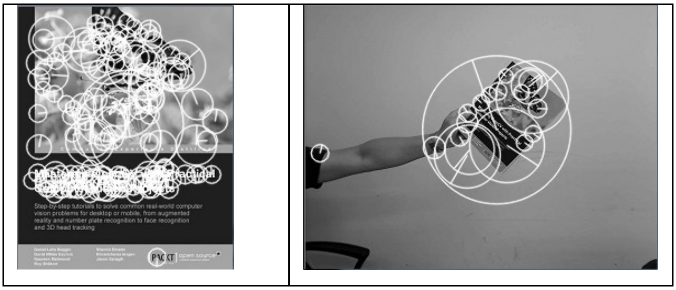

Fig. 1. Object feature detection result

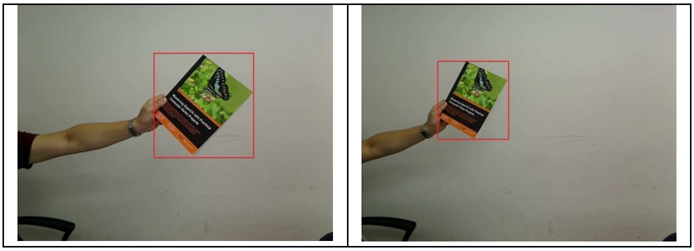

Fig. 2. Object recognition and tracking result

Table 1. Object recognition and tracking result

Algorithm	Recognition Time(sec)
Proposed	0.49
SURF	0.65
SIFT	4.82

Figure 1 describes object recognition result using our approach which is a low complexity and robust object recognition by advanced feature matching. Figure 2 shows object tracking result using our method which is improved CamShift. As the change of interest object, its region can adapt to resize of window correctly. Table 1 shows improved performance in the comparison with existing algorithms of the proposed algorithm.

5 Conclusion

In this paper, we proposed a low complexity and robust object recognition and tracking using advanced feature matching for real time environment. Our algorithm recognizes object using invariant features and reduces dimension of feature descriptor to deal with the problems. Our experiments demonstrate that our work is more fast and robust than the traditional methods and can track object accurately in various environments. As future researches, our method needs to detect object in simple environment and recognize multiple objects in real time environment.

References

[1] Yilmaz, A., Javed, O., Shah, M.: Object tracking: A survey. ACM Computing Surveys (CSUR) 38(4) (December 2006)
[2] Yang, H., Shao, L., Zheng, F., Wang, L., Song, Z.: Recent advances and trends in visual tracking: A review. Neurocomputing 74, 3823–3831 (2011)
[3] Kloihofer, W., Kampel, M.: Interest point based tracking. In: International Conference on Pattern Recognition (ICPR), pp. 3549–3552. ACM (2010)
[4] Wang, J., He, F., Zhang, X., Yun, G.: Tracking Object throuth Occuusions Using Improved Kalman Filter. In: IEEE 2010, vol. (6), pp. 223–228 (2010)
[5] Yue, Y., Gao, Y., Zhang, X.: An Improved Camshift Algorithm Based on Dynamic Background. In: The 1st International Conference on Information Science and Engineering (ICISE 2009), pp. 1141–1144 (2009)
[6] Zhang, J., et al.: The Target Tracking Method Based on Camshift Algorithm Combined with SIFT. Advanced Materials Research 186, 281–286 (2011)
[7] Ido, L., Michael, L., Ehud, R.: Mean Shift tracking with multiple reference color histograms. Computer Vision and Image Understanding (9), 400–408 (2010)
[8] David, E., Erich, B., Daniel, K., Anselm, S.: Fast and Robust Camshift Tracking. IEEE Transaction (8), 1–8 (2010)
[9] Juan, L., Gwun, O.: A Comparison of SIFT, PCA-SIFT and SURF. IJIP 3(4) (2009)
[10] Ha, S., Moon, Y.: Multiple Object Tracking Using SIFT Features and Location Matching. International Journal of Smart Home 5(4) (October 2011)
[11] Ke, Y., Sukthankar, R.: PCA-SIFT: A More Distinctive Representation for Local Image Descriptors. In: Proc. Conf. Computer Vision and Pattern Recognition, pp. 511–517 (June 2004)
[12] Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
[13] Brown, M., Lowe, D.: Invariant features from interest point groups. In: BMVC (September 2002)
[14] Stommel, M.: Binarising SIFT-Descriptors to Reduce the Curse of Dimensionality in Histogram-Based Object Recognition. International Journal of Signal Processing Image Processing and Pattern Recognition 3(1) (March 2010)
[15] Bay, H., Tuytelaars, T.: SURF: Speeded Up Robust Features. Computer Vision and Image Understanding (CVIU) 110(3), 346–359 (2008)
[16] Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple feature. In: CVPR(1), pp. 511–518 (December 2001)