Journal of Engineering and Applied Sciences 12 (14): 3682-3686, 2017
ISSN: 1816-949X

Object Detection and Tracking Scheme Based on SURF and Difference Image

Dong-Chul Park
Department of Electronics Engineering, Myong Ji University, YongIn, Gyeonggi-do, Rep. of Korea

Abstract: An efficient scheme for object detection and tracking under a fixed camera environment with SURF (Speeded Up Robust Feature) algorithm and the difference image method is proposed in this study. Since, SURF provides local feature descriptors robust on scale, rotation and illumination changes, SURF algorithm is adopted for its advantages over SIFT (Scale Invariant Feature Transform) algorithm. In order to improve the speed and accuracy in detecting and tracking objects, the proposed algorithm combines SURF algorithm and the difference image method. Experiments on two video data sets imply that the proposed method can be an alternative to conventional methods because of the improvements in terms of the speed of operation and the tracking accuracy over conventional methods.

Key words: Object detection, tracking, SURF, interest points, difference image

INTRODUCTION

Designing security systems through camera images requires automatic detection and tracking algorithms with accuracy and speed. Especially, automatic detection of objects on camera images is one of the most important tasks in designing computer-controlled vehicles. Since, moving objects show different appearances in images, this task of object detection and tracking becomes a very challenging task in computer vision research (Thillainayagi *et al*., 2014; Wu *et al*., 2015). By adopting a binary classification method that classifies a local region into regions with an object or not, the object detection problem can be considered as a classifier design problem in computer vision research. Some of the important achievements for objection detection task include Histogram of Oriented Gradient (HOG) features and SIFT (Lowe, 2004; Dalal and Triggs, 2005; Athilingam *et al*., 2014). On the other hand, object tracking task requires different local features and descriptors related with different objects. One of the most widely used descriptors for different shapes of objects is SURF which is an improved version of SIFT (Bay *et al*., 2008). The difference image method is a very simple but powerful method for object tracking tasks. By using the difference between two consecutive image sequences, moving objects can be detected. Even though this difference image method is very simple method, its application is very limited because of its sensitivity to noise and illumination change. By combining SURF and difference image method, the object detection and tracking task can be achieved with accuracy and speed.

Literature review

Difference image: In order to find the region for objects from the background image, the image difference method uses the following Eq. 1:

$$h(x,y) = |g(x,y) - f(x,y)| \quad (1)$$

Where g(x, y) and f(x, y) denote the pixel values of the current and the previous frames at a location (x, y), respectively.

SIFT (Scale Invariant Feature Transform) and SURF (Speeded Up Robust Feature) algorithm: SURF algorithm is a variation of SIFT algorithm which can detect and describe local features. The SIFT descriptor has several invariant features including changes in scale, rotation and illumination. These invariant or partially invariant features of SIFT make SIFT a valuable tool for object detection and recognition. SIFT extracts keypoints first from given reference images as the feature points of the images. By matching these obtained features individually between target image and reference image, an object on a new image can be detected. The match process based on the matching information with subsets of keypoints for location, scale and orientation between reference image and target image can decide candidate objects. During the matching process, SIFT utilizes the Hough transform for determining consistent clusters and highly probable object detection. While SIFT gives an insightful tool for object recognition task, it suffers from operation speed and accuracy because of the high dimensionality of the descriptor in SIFT at the matching step (Bay *et al*., 2008). In order to overcome the speed and accuracy problems in

SIFT, SURF was introduced. SURF is based on the Hessian matrix with DoG approximation which is a simple Laplacian detector in order to achieve a higher speed operation. SURF utilizes 64-dimensional Haar-wavelet features around the interest point. The 64-dimensional Haar-wavelet features can reduce the computation time for feature calculation and matching process. The robustness of SURF can be achieved by using 64-dimensional Haar-wavelet features and a new indexing method with the Laplacian. SURF consists of the following 4 processes:

Generation of integral images: Since, a faster computation of box shape operators can be achieved by using integral images (Viola and Jones, 2001), the integral image $I_{\Delta}(x)$, x = (x, y) is formed first in SURF with the following definition:

$$I_{\Delta}(X) = \sum_{i=0}^{i\le x}\sum_{j=0}^{j<y} I(i, j) \quad (2)$$

Figure 1 shows how computation with integral images can save computation time.

Extraction of interest points: Interest points are obtained for the invariances to the changes in scale, rotation, illumination and image noise. Instead of using Gaussian filter, SURF utilizes the following Hessian matrix for both of location and scale features.

$$\tilde{H} = \begin{vmatrix} CG_{xx} & CG_{xy} \\ CG_{xy} & CG_{yy} \end{vmatrix} \quad (3)$$

Where CG_{xx}, CG_{xy} and CG_{yy} denote the convolutions of the Gaussian second order derivative with I(x, y). For faster computation, the following Hessian determinant is utilized (Bay *et al*., 2008):

$$\det(\tilde{H}) = CG_{yy} - (0.9CG_{xy})^2 \quad (4)$$

Figure 2 shows examples of the second order differentiation box filters. For scale invariant features, the image pyramids are used. In image pyramids, integral images and box filters are used in parallel fashion. Unlike SIFT, SURF can calculate the image pyramid in parallel fashion by using different box filters instead of iteratively reducing the size of the original image in SIFT. The interest points for scale invariance in the image are found by considering a predetermined threshold and the determinant of the Hessian matrix in the 3×3×3 window.

The interest points for rotation invariance are calculated by adopting the Haar-wavelet responses in

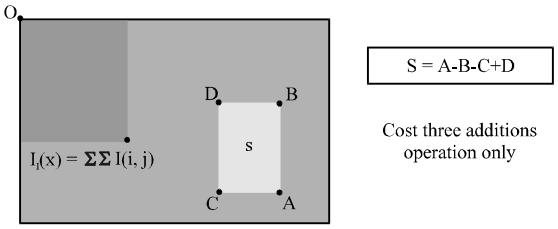

Fig. 1: Integral image calculation

Fig. 2: Examples of the second order differentiation box filters

horizontal and vertical directions for a neighborhood of radius size 6 s where s denotes the scale for the interest point obtained.

Generation of local feature descriptor: In order to form local descriptor components, a neighbourhood centred at the interest point with the size of (20×20 s) is first formed and the neighbourhood box is then split up 4×4 into subregions. Haar-wavelet responses are then calculated on these subregions. The Haar wavelet responses utilized in SURF are Σdx, $\Sigma|dx|$, Σdy and $\Sigma|dy|$. The resulting features on each 4×4 subregion are 64-dimensional descriptors.

Feature matching procedure: For efficient and fast matching process, SURF utilizes the sign of Laplacian for keypoints because its sign can discriminate bright spots from dark background. During matching procedure, the signs of Laplacian for interest points are first compared. When the signs of Laplacian for interest points match, the feature matching procedure then calculates the Euclidean distances among candidates. The nearest neighbour search method is utilized for determining matching points with the following constraint:

$$\frac{|P - P_1|}{|P - P_2|} < T \quad (5)$$

Where, P_1 and P_2 denote the nearest neighbor and the 2nd nearest neighbor points from the point of interest P, respectively. Note that we use in our experiments.

MATERIALS AND METHODS

In order to achieve accurate and fast performance in object detection and tracking scheme, the proposed method combines difference image method and SURF algorithm. The proposed method consists of two procedures; moving object detection and moving object tracking.

Moving object detection procedure: In detecting moving objects using SURF for obtaining local feature descriptors, it first requires to estimate the region where an object is located. The difference image between consecutive image frames is first found and labelling procedure is then applied to the difference image in order to discriminate different objects from background. The Grassfire transform (Blum, 1967) used in this case computes from a pixel to the borders of an image region can yield the boundary of an object in image. Based on the results obtained by the Grassfire transform, a proper box-shape boundary of a moving object can be obtained. The box-shape boundary can be considered as a moving object when the size of boundary is larger than a predetermined size. SURF algorithm is then applied to the detected box-shape boundary positions for obtaining local feature descriptors and the matching results between the local feature descriptors of the previous frame and current frame are saved as the matching points.

Moving object tracking procedure: When undesirable matching points for background or noises are found, errors in object detection results are inevitable. In order to avoid this situation, a new procedure to determining objects by using the centre of matching points is proposed in this study.

Matching points density information: In order to extract a matching points density information from matching points on target image, a mask with the size of N×N where each cell has a value of one is adopted. The mask is then convolved with each matching point at (x, y) and the resulting convolved map provides us Matching Points Density Function (MPDF), D(x, y). Note that D(x, y) shows the number of matching points at each matching point (x, y) within a certain boundary.

Centre of mass for matching points: In order to estimate the location of an object with stability, the moment functions are adopted (Mukundan and Ramakrishman, 1998):

$$M_{00} = \sum_x \sum_y D(x,y),\ M_{10} = \sum_x \sum_y xD(x,y),\ M_{01} = \sum_x \sum_y yD(x,y) \quad (6)$$

By using the moments in Eq. 6, the centre location of an object, (x_c, x_y) can be estimated as follows:

$$x_c = \frac{M_{10}}{M_{00}} \text{ and } y_c = \frac{M_{10}}{M_{00}} \quad (7)$$

RESULTS AND DISCUSSION

For evaluating the proposed method for real-time video image data, experiments are performed under the following computing environment:

- CPU: Intel Core i7-2600, 3.40GHz×CPU, RAM size: 4 GB, OS: Window 7 Enterprise, 64bit
- Camera resolution: 640×480, 30fps

The region of a moving object is first estimated by using the difference image method. In applying the difference method, a moving objet is detected only if the size of a moving object is >30×30 window size. Figure 3 shows an example of object detection results by applying the difference image method to image data. When applying SURF algorithm for object tracking problem, the tracking performance depends heavily on the parameters including the numbers of octaves and scales. Experiments show that the number of scales increases the number of interest points and tracking accuracy. However, the processing speed decreases as the number of scales increases because more interest points obtained with more scales require more computation time. In object tracking system, the tracking speed is as important as the tracking accuracy. In order to achieve a real-time operation condition, a set of parameters for a processing speed of more than 10 fps and a high tracking accuracy is found as: the number of octave is 2 and the number of scales is 3. Figure 4 shows an example of tracking results.

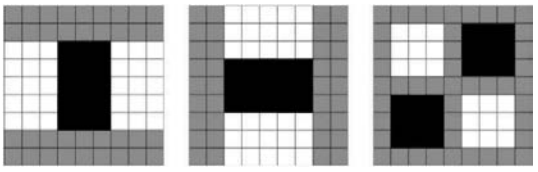

Fig. 3: Example of object detection results with the difference image method

Fig. 4: Example of tracking results

Fig. 5: Example of a pedestrian video data

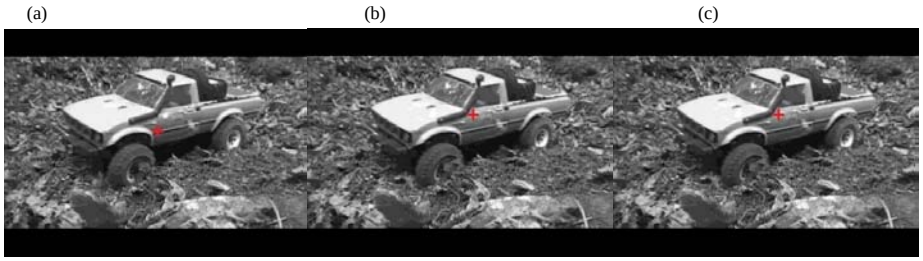

Fig. 6: Example of tracking results by different algorithms for truck video data; a) Difference image method; b) SURF algorithm; c) The proposed method

Table 1: Comparison of performances for pedestrian data

Algorithms	Accuracy (%)	Speed (fps)
Diff. image	89.4	32.6
SURF	78.7	12.4
The proposed	90.6	31.7

In Fig. 4, white dots are the interest points found and the red mark is the center of tracking object. Note that these images in Fig. 4 are with different sizes, orientations and camera perspectives.

In order to compare the performance of the proposed algorithm with conventional algorithms, two video image sequences, pedestrian video data and vehicle video data, are used for experiments.

Experiments on pedestrian video data: Figure 5 shows video data clips with background, a front view of the pedestrian and a side view of the pedestrian. Table 1 summarizes detection and tracking results on the pedestrian video data. As shown in Table 1, the proposed method shows a favorable performance in terms of tracking accuracy and speed over SURF or difference image method. As far as the tracking speed is concerned, the proposed method is much faster than SURF and this is a very important feature for real-time applications.

Experiments on truck video data: Another video clip data used for experiments is obtained from youtube (11). Unlike the pedestrian video data, the truck video data has very complex background image and contains many interest points. Therefore, as expected, the difference image method shows very poor tracking accuracy. Table 2 shows a summary of performance comparison among

Table 2: Comparison of performance for truck data

Algorithms	Accuracy (%)	Speed (fps)
Diff. image	75.8	33.70
SURF	84.6	5.5
The proposed	87.8	25.2

different algorithms. As can be seen from Table 2, the proposed method yields a very comparable tracking accuracy with SURF algorithm while its tracking speed is about 5 times faster than SURF algorithm. Figure 6 shows example of tracking results by different algorithms for truck video data. The red-cross signs in the images denote the centers of tracking objects found. The difference image method shows significant errors in estimating the center when compared with the other two methods. This estimation error in the difference image method is considered as an effect of the complex background.

CONCLUSION

A moving object detection and tracking scheme based on SURF algorithm and the difference image method is proposed in this study. The proposed scheme takes advantages both of the difference image method for its fast speed of operation and SURF algorithm for its accurate recognition. The proposed scheme first finds probable areas where moving objects exist by using the difference image method. After an object is detected, interest points inside a window are then calculated by using SURF algorithm. In order to increase the

computation speed required for matching procedure in SURF, the centre of an object is then estimated by adopting the moment generating function. When an image frame doesn't have enough interest points from SURF algorithm, the matching process is replaced with the difference image method. Experiments on different sets of video image sequence data for the evaluation purpose show that the proposed scheme yields very accurate tracking results with a speed suitable for real-time operation. When compared with SURF algorithm-based tracking scheme, the proposed scheme shows very comparable accuracy with the SURF algorithm-based tracking scheme while its processing speed is far faster than the SURF algorithm-based tracking scheme. This advantageous feature of the processing speed for the proposed scheme results from the combined efforts of adopting difference image method and centre estimation of an object by using the moment generating function.

ACKNOWLEDGEMENT

This research was supported by 2014 Research Fund of Myongji University. The researcher thanks to Miso Jang for her help in preparing this manuscript.

REFERENCES

Athilingam, R., M.A. Rasheed, K.S. Kumar, A. Kaviyarasu and R. Thillainayagi, 2014. Target tracking with background modeled mean shift technique for UAV surveillance videos. Intl. J., 6: 805-814.

Bay, H., A. Ess, T. Tuytelaars and L. van Gool, 2008. Speeded-Up Robust Features (SURF). Comput. Vision Image Understand., 110: 346-359.

Blum, H., 1967. A Transformation for Extracting New Descriptors of Shape: Models for the Perception of Speech and Visual Form. 1st Edn., MIT Press, Cambridge, pp: 362-380.

Dalal, N. and B. Triggs, 2005. Histograms of oriented gradients for human detection. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognition, 1: 886-893.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision, 60: 91-110.

Mukundan, R. and K.R. Ramakrishnan, 1998. Moment functions in image analysis-theory and applications. World Scientific, Singapore, Asia, ISBN:978-02-3524-0, Pages: 151.

Thillainayagi, R., M.A. Rasheed, K.S. Kumar and R. Athilingam, 2014. Object detection using a novel YIQ model based image fusion for UAV aerial surveillance. Intl. J. Eng. Technol., 6: 1386-1393.

Viola, P. and M. Jones, 2001. Rapid object detection using a boosted cascade of simple feature. IEEE Conf. Comput. Vision Pattern Recogn., 1: 511-518.

Wu, S., R. Laganiere and P. Payeur, 2015. Improving pedestrian detection with selective gradient self-similarity feature. Pattern Recognit., 48: 2364-2376.