ISSN 23499842(Online), Volume 2, issue 1 March 2017
International Journal of Innovative Trends and Emerging Technologies

FEATURE BASED MOVING OBJECT DETECTION AND TRACKING

Monisha R[1], Muthuselvam M[2]
[1,2] *Department of Electronics, Madras Institute of Technology, Chennai- 600044, India*
E-Mail : monisharj618@gmail.com

Abstract

In the field of computer vision Object detection and tracking arechallenging tasks for many applications. It is essential for Security and surveillance systems in industry, hospitals and other places. Object Tracking involves the detection of interesting moving objects in a frame of a video sequence and tracking of such objects in the consecutive frames. In this paper, a Robust tracking method for reliably detecting and tracking the foreground moving object in a sequence of images is proposed. The proposed method uses Speeded Up Robust Features algorithm for object detection and tracking. The method is entirely based on the local features of an object. Distinct feature points can be detected using the Speeded Up Robust Features method which is invariant to rotation and spatial movement. Feature descriptors are extracted from this feature points. The proposed method performs the descriptor matching between consecutive frames. Experimental results confirm the robustness of the proposed method.

Index terms: *object detection, SURF, interest points, Haar wavelet, feature extraction, descriptor matching, object tracking.*

1. Introduction

Computer vision is being applied in a large number of applications. It has become a vast area of research for image analysis. Image analysis for real-time application mainly involves the detection of an object while it is moving. It is used in applications such as surveillance, crowd control, traffic monitoring, security, counting the number of vehicles, parking system. Computer vision is concerned with the understanding of information from a single image or sequence of images. It is sure that image processing algorithms can be applied to individual frames of a video sequence. Generally, the two consecutive frames are closely related. Object tracking is a valuable task within the field of computer vision. The increase in need for automated video analysis has generated a major interest in object detection and tracking. Object tracking involves detecting an object in a frame and tracking it in the consecutive frames. Object detection involves detection of an object of a particular class like face, car, bicycle and human. Tracking of an object becomes complex with respect to partial and full occlusion, scene illumination changes and real-time requirements. Nowadays robust, accurate and high performance approach is a challenging task.

In this paper robust object tracking method based on SURF algorithm is proposed. The essential thing is feature extraction. Features are more important to identify an object. There are a number of methodologies for object tracking based on features such as SIFT, SURF, BRISK. SURF and SIFT are invariant to scale and rotational changes. Visual features such as specific color, the shape of an object are easy to detect, recognize and track. But at different poses and illumination conditions, it is very difficult to achieve accurate detection. So local features of the objects are used. SURF is preferred for the use in real-time implementation behalf of its speed due to the usage of integral images.

2. Literature Survey

So far, many feature based tracking algorithms are proposed. Dr. Senthil Kumar K et al. [2] uses Continuously Adaptive Mean Shift algorithm based on color probability distribution and histogram color model of an image. It adapts to object's variable size by altering its window based on the object movement. Motion based optical flow technique is proposed by Sara Minaeian et al. [1] for outdoor crowd detection as well as they use Histogram of Gradients for human detection and landmark-based localization of target without the need for camera calibration. Jiawei He et al. [4] proposed a framework which uses Kalman filter and Mean shift algorithm along with Bhattacharyya co-efficient for occlusion recovery Accuracy and anti-occlusion robustness was achieved by using this method. Samar D. Gajbhiye et al. [3] discusses that real-time object tracking is important in video surveillance systems. It is entirely based on the concepts of RGB to HSV Conversion, Segmentation, Pre-Processing techniques based on morphological operations and Bounding Box creation. The efficient method under background clutter and illumination changes for extraction of feature points was proposed by Wenying Wang et al. [5] where the motion vector is calculated for each matched pair. Manual selection of Region of Interest reduces the calculation burden. Herbert Bay et al. [6] explained a local feature extraction method named as Speeded Up Robust Features for the extraction of feature points.ZHOU Dan et al. [7] proposed a two-stage matching and template update method to deal with the problem of object appearance changes. Reliable matching is performed by matching individual features to the database of features from aknown object using fast nearest-neighbor algorithm is showed by David G. Lowe [8].

The rest of the paper is organized as follows, In section III the detailed description of SURF algorithm is explained. The proposed system design is given in section IV. In section V We show the simulation results. Finally, we discussed the conclusion and future work in section VI.

3. Overview Of Surf Algorithm

Speeded Up Robust Features (SURF) is a local feature detector and descriptor. The steps of surf algorithm contain three sections namely,

- Interest point detection
- Interest point description
- Interest point matching

ISSN 23499842(Online), Volume 2, issue 1 March 2017
International Journal of Innovative Trends and Emerging Technologies

The flowchart of the SURF algorithm is shown in the Fig. 1

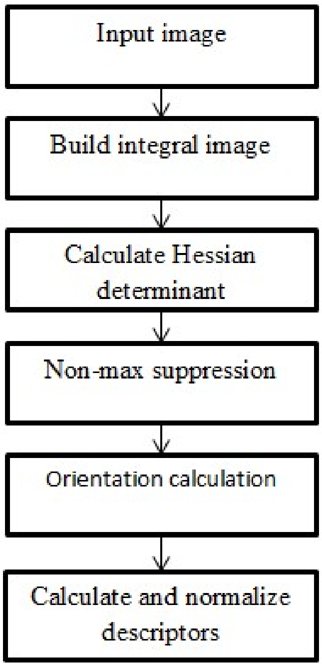


Fig. 1. Flow chart of SURF algorithm

A. Integral image

A summed area table is a structure which efficiently generating the sum of values in a rectangular subset of a grid. In the image processing domain, it is also known as an integral image. The value at any point (x, y) in the summed area table is just the sum of all the pixels above and to the left of (x, y) and it is given in (1),

$$I(x,y) = I(x,y)+I(x,y-1)+I(x-1,y)-I(x-1,y-1) \qquad (1)$$

B. Scalespace representation and interest point detection

Real world objects become meaningless entities after a certain range of values. A scale-space is the representation of an image at multiple scale levels. The scale-space can be represented by convolving the input image with a filter of increasing size at several levels. Here, the scale space is divided into octaves. Each octave is subdivided into levels.Fig.2 shows the scale space representation.

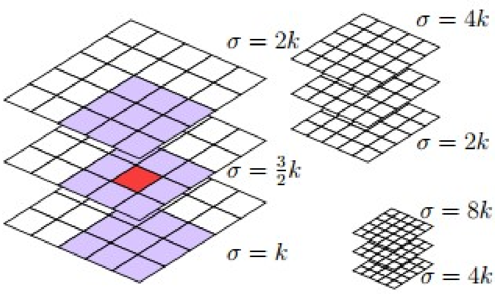


Fig. 2.scale space representation

A Fast-Hessian detector is used to find the feature points which are a local two-dimensional structure. For an image I, the hessian matrix at point X with scale of σ is given in (2),

$$(X,\sigma) = \begin{bmatrix} Lxx(X,\sigma) & Lxy(X, \\ Lxy(X,\sigma) & yy(X, \end{bmatrix} \qquad (2)$$

$Lxx(X,\sigma)$ and $Lxy(X,\sigma)$ is the convolution of Gaussian second order derivatives $\frac{\partial^2}{\partial x^2} g(\sigma)$ and $\frac{\partial^2}{\partial xy} g(\sigma)$with the image at points (x,y) as described in (4), (5) respectively. Gaussian second order derivative is given as follows,

$\frac{\partial^2}{\partial x^2} g(\sigma)$where $g(\sigma)$ is given as in (3),

$$g(\sigma)=\frac{1}{2\pi\sigma^2} e^{\frac{-(x^2+y^2)}{(2\sigma^2)}} \quad (3)$$

$$Lxx(X,\sigma) = I(X) * \frac{\partial^2}{\partial x^2} g(\sigma) (4)$$

$$Lxy(X,\sigma) = I(X) * \frac{\partial^2}{\partial xy} g(\sigma) (5)$$

Gaussian second order derivative needs to be discretized before calculating the convolution and it is done by Box filter.Fig. 3 shows the 9x9 box filter used for approximation of Gaussian second order partial derivative.

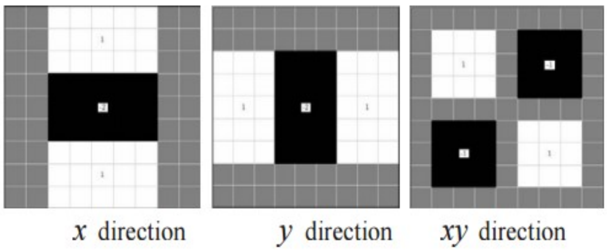


Fig. 3.Box filter representation

Dxx, Dyy and Dxyrepresentsthe convolution of box filter with the image.The determinant of hessian matrix which is used for deciding whether a point could be chosen as interest point or not and it is given in (6),

$$\det(Happrox) = DxxDyy - (0.9Dxy)^2 \qquad (6)$$

C. Non-maximum suppression

The determinant of hessian matrix is calculated at each scale and non-maximum suppression is carried out in 3x3x3 neighborhood inorder to find the maxima. If the determinant value is greater than zero, then that pixel is considered as an extreme point. Non-maximum suppression is performed to remove some extreme points. The interest points are the points which are extrema among 8 neighbors in the current level and 2x9 neighbors in the level above and below. These interest points are invariant to scale,rotation and illumination variations.

D. Dominant Orientation

Once the feature points are detected a local image patch around that was detected. Descriptors are used for the detection of an extent of matching between the two images. Before finding the shape of the descriptor, an outline needs to be find which can be carried out by Haar Wavelets. It is the reason why the circular region is formed around the feature points.

Dominant orientation for the determined interest point is assigned by HaarWavelet response. Haar–like feature consider the adjacent regions at a specific location, sums up the pixel intensity in each region and calculate the difference between these sums. Haar wavelet is calculated in both x and y directions. The advantage of Haar-like feature is its calculation speed.

E. Descriptor Normalisation

After that, a square region is constructed along the dominant orientation for getting the descriptive information. Further, the square regions are split up into 4x4 sub-regions. Dividing the interest point region into 4x4 sub-regions for computing the SURF descriptor is shown in Fig. 4

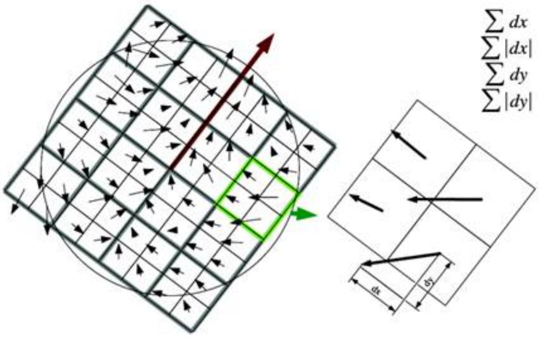


Fig. 4. Representation of SURF descriptor

For each sub-region,Haar wavelet response is weighted in bothhorizontal and vertical directions. Focusing at invariant to rotation$|dy|$ and $|dx|$are summed up to obtain $\Sigma|dy|$, $\Sigma|dx|$respectively. After the normalization of the eigenvector, a 4D vector is obtained.After the calculation of 4-dimensional vector, eigenvector for 16 sub-regions are calculated, thus 64D vector is constructed and it is described as in (7),

$$V= (\Sigma dx, \Sigma dy, \Sigma|dx|, \Sigma|dy|) \quad (7)$$

F. Keypoint Matching

The keypoints matching is based on the distance between the vectors. The best match for each key point is found by finding its nearest neighbor in the extracted keypoints from template images. We use the Euclidean distance as a similarity measure for finding the nearest-neighbor

4. Implementation Of The Proposed Method

SURF algorithm is not only be used for object detection, it can also be used for tracking. There are generally two methods to represent images namely, global features and local features. The global representation method produces a single vector with values that measure various aspects of the image such as color, texture or shape. The main goal of local feature representation is to distinctively represent the image based on some salient regions while remaining invariant to viewpoint and illumination changes. Thus the image is represented based on its local structures by a set of local feature descriptors extracted from a set of image regions called interest regions. Feature detection is usually performed as the first operation on the image, and examines every pixel to see if there is a feature present at that pixel. Flow chart for the proposed methodology is shown in Fig. 5.

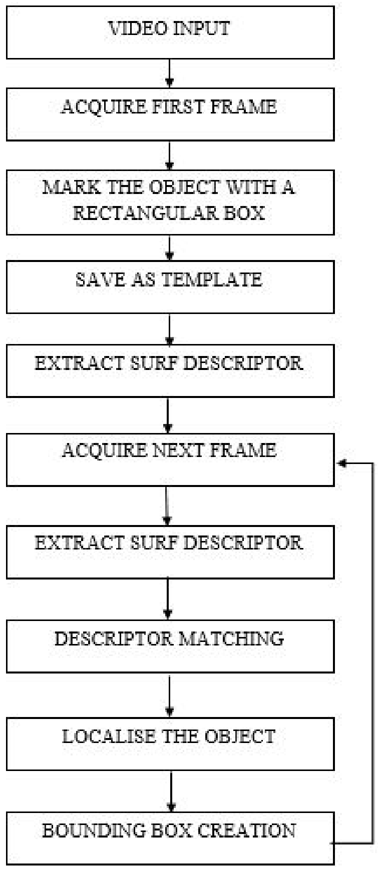


Fig. 5. Flow chart of the proposed algorithm

In our method, there is no prior information about the target. First frame in the video sequence is taken as input and mark the desired target with a rectangle box by the click of a mouse and consider this as a Region of Interest. This Region of Interest is saved as a reference image. SURF keypoints are detected for the reference image and corresponding SURF descriptor is extracted. The next frame is acquired and keypoints are detected. Corresponding SURF descriptor is extracted.

Assume that image of the first frame I exist a SURF keypoint P. In next frame J we find two nearest keypoints P1 and P2 from P. We use the Euclidean distance to calculate the distance between two key-points and P3 represents the nearest distance and P4 represents the second nearest distance. If the ratio of P3 and P4 is less than the threshold, we consider keypoints P and P1 are matched.In this way, we can obtain the initial matching points between image I and image J. Matching formula is given in (8),

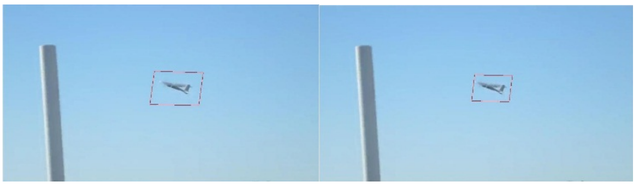
$$\frac{\text{nearest distance}}{\text{second nearest distance}} \leq \text{Threshold} \quad (8)$$

If the target is matched with reference image, then Bounding box is created around the target. Thus the object is tracked in consecutive video frames.

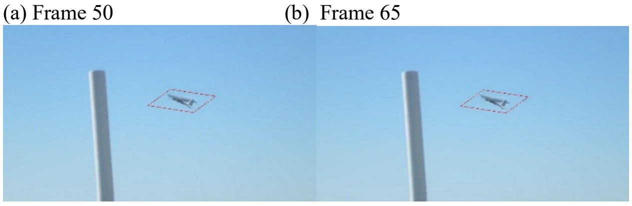
5. Simulation Results

A desktop computer with Pentium(R) Dual Core CPU Processor 2.10GHZ with 2GB RAM under MATLAB R2015a environment is used to run our proposed algorithm. The test is conducted on many video sequences. Here, the result of two video sequences are shown.

The first video sequence has 1791 frames and frame size of 450x854. Fig. 6 Shows the tracking result of first video sequence. Due the fast movement, the flight is accurately tracked in consecutive sequence of frames.



(a) Frame 50 (b) Frame 65

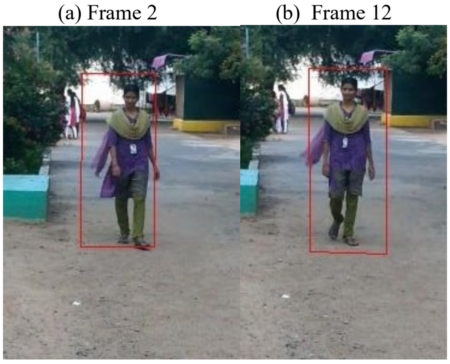


(c) Frame 80 (d) Frame 95

Fig. 6. Tracking result of a fast moving flight (a) 50th frame (b) 65th frame (c) 80th frame (d) 95th frame

The second video sequence has 100 frames and frame size of 1080x1920. Fig. 7 shows the tracking result of a specific person. The object is tracked accurately in consecutive frames

(a) Frame 2 (b) Frame 12



(c) Frame 27 (d) Frame 38

Fig. 7. Tracking result of a specific person (a) 02^{th} frame (b) 12^{th} frame (c)27^{th} frame (d) 38^{th} frame

The experimental result shows that the proposed algorithm is robust to detect the object. The performance depends on the accuracy of Region of Interest detection and interest point detection.

6. Conclusion

Simple and robust object detection and tracking method is presented in this paper. By using this method, in first step keypoints are detected successfully. The descriptor extraction and matching steps also showed good results. The object is detected and tracked from one frame to the other. It can be used in various applications, which would become the development of surveillance system, requires fast and robust methods for detecting and tracking moving objects. It can also be used for people counting and security systems.

For future work and as an extension of the work performed in this method, low pixel intensity levels in the image could also be considered which would also disrupt the performance of this algorithm. This algorithm has to be improved to track the object in a dynamic environment.

7. References

[1] Sara Minaeian, Jian Liu, and Young-Jun Son, (2016), “Vision-Based Target Detection and Localization via a Team of Cooperative UAV and UGVs”, IEEE Transactions on systems, man, and cybernetics: systems, vol. 46, no. 7.

[2] Dr.Senthil Kumar K, Mohamed Rasheed A and Vairamanikandan K, (2015), “Hardware Development and Implementation of an Object Tracking Algorithm for UAV Applications”, International Conference on Image Processing, Production and Computer Science, ISBN 97893-84422-22-6, pp.68-74.

[3] Samar D. Gajbhiye and Pooja P. Gundewar, (2015), “Object Tracking Using Embedded Platform for Video Surveillance”, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 7.

[4] Jiawei He and Yingyun Yang, (2015), “Improved Anti-occlusion tracking algorithm based on Face Detection”, 2015 978-1-4799-1980-2 © IEEE

[5] Wenying Wang, Yibo Zhou, Xucheng Zhu and Yuxiang Xing, “A Real-time Tracking Method Based on SURF”, 8th International Congress on Image and Signal Processing (CISP), Ministry of Education, Beijing, China, 2015, 978-1-4673-9098-9 ©2015 IEEE. 29.