International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 - 5161

Available at http://inpressco.com/category/ijcet

Research Article

Tracking and Recognition of Objects using SURF Descriptor and Harris Corner Detection

J.Jasmine Anitha[Ȧ*] and S.M.Deepa[Ȧ]

[Ȧ]Nehru Institute of Engineering and Technology(Anna University) ,Coimbatore, India,

Accepted 01 March 2014, Available online 01 April 2014, **Vol.4, No.2 (April 2014)**

Abstract

Visual object tracking for surveillance applications is an important task in computer vision. Tracking of object is a matching problem. One main difficulty in object tracking is to choose suitable features and models for recognizing and tracking the target. SURF (Speeded Up Robust Features) algorithm is used here for continuous image recognition and tracking in video. The SURF feature descriptor operates by reducing the search space of possible interest points inside of the scale space image pyramid. SURF adds a lot of features to improve the speed in every step. The resulting tracked interest points are more repeatable and noise free. SURF is good at handling images with blurring and rotation. Corner detection is good for obtaining image features for object tracking and recognition. Interest points in an image are located using corner detector. By using harris corner detection algorithm along SURF feature descriptor, tracking efficiency is improved.

Keywords: *Video surveillance, object tracking, object recognition, motion estimation, interest points, edge orientation.*

1. Introduction

Video tracking is the process of locating a moving object (or multiple objects) over time using a camera. In video tracking an algorithm analyzes sequential video frames. Two major components of visual tracking are target representation and localization. Video tracking is a time consuming process depending the amount of data that is contained in given video. The main objective of video tracking is to associate target objects in consecutive video frames. Locating and tracking the target object depends on the algorithm. Robust feature descriptors such as SIFT(Sift Invariant Feature Transform), SURF, and GLOH (Gradient Localization Oriented Histogram) have become a core component in applications such as image recognition. As name suggests, SURF is a speeded-up version of SIFT. SURF descriptor is three times as fast as SIFT feature descriptor. SURF descriptor is preferred for its fast feature extraction. Quality of object recognition is important to the real-time tracking requirement, and the tracking algorithm should not interfere with the recognition performance.

SURF (Speeded Up Robust Features) algorithm is used for feature extraction and continuous image recognition and in video. It reduces the search space of possible interest points inside of the scale space image pyramid. The interest points tracked by SURF are resilient to noise. SURF is based on sums of 2D Haar wavelet responses and makes an efficient use of integral images. SURF feature tracks the objects by interest point matching and updating. It then continuously extracts feature for recognition. For feature extraction SURF uses the sum of the Haar wavelet response around the interest point. The main advantage of Haar wavelet transform is its calculation speed.

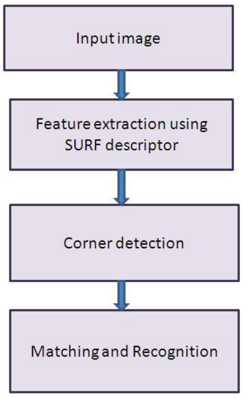


Fig.1 Flow chart of proposed method

The association can be especially difficult when the objects are moving fast relative to the frame rate. When the tracked object changes orientation over time, complexity increases. For these situations video tracking systems usually employ a motion model. The motion model describes how the image of the target might change for different possible motions of the object. Motion estimation is done through Harris corners and object recognition is done through robust features such as SURF feature descriptor. Harris corner detection is used for its computation speed. Harris corner detector is rotation and

*Corresponding author **J.Jasmine Anitha** is a PG Scholar(M.E-Communication Systems) and **S.M.Deepa** is Asst Prof

scale invariant. Using SURF descriptor along with harris corner detection improves tracking efficiency and is invariant to illumination changes in images.

2. SURF algorithm overview

A correspondence matching is one of the important tasks in computer vision, and it is not easy to find corresponding points in variable environment where a scale, rotation, view point and illumination are changed. A SURF(Speeded Up Robust Features) algorithm have been widely used to solve the problem of the correspondence matching because it is faster than SIFT(Scale Invariant Feature Transform) with closely maintaining the matching performance.

SIFT uses visual pyramids to find candidate points and filters each layer according to the Gauss law with increased Sigma values and finds differences. SURF on otherhand uses Hessian Matrix to select the candidate points in different sizes. SURF uses Haar wavelet filters and the integral of the image to speed up the filtering operation. SURF is good at handling images with blurring and rotation. The method is very fast because of the use of an integral image where the value of a pixel (x,y) is the sum of all values in the rectangle defined by the origin and (x,y).

SURF uses integer approximation. To detect features the Hessian matrix (H) is assembled, where L_{xx} is the convolution of the second derivative of a Gaussian with the image at the point. Hessian matrix is represented as,

$$H = \begin{bmatrix} L_{xx} & L_{xy} \\ L_{xy} & L_{yy} \end{bmatrix} \quad (1)$$

SURF uses different scales of Gaussian masks, while the scale of image is always unaltered.

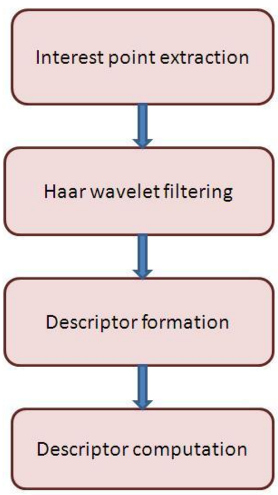


Fig.2 Process of SURF descriptor

The SURF descriptor is designed to be scale invariant. To ignore scale the descriptor is sampled over a window that is proportional to the window size with which it was detected. If a scaled version of the feature is in another image the descriptor for that feature will be sampled over the same relative area.

2.1 Tracking interest points

The interest point detection algorithm requires the estimation of relative motion. The tracking process produces fairly accurate results and only a small number of iterations suffices to reject most of the false tracking results When more than one interest point are detected in this region, we need to choose the one that best matches. The image corners are similarly transformed to the new image.

2.2 Haar wavelet filtering

A Haar-like feature uses adjacent rectangular regions at a specific location in a detection window and sums up the pixel intensities in each region and the difference between these sum is calculated. This difference categorizes the subsections of an image. For feature description, SURF uses Wavelet responses in horizontal and vertical direction around the point of interest. The position of these rectangles acts like a bounding box to the target object. For each subsection of the image the Haar-like feature is calculated. This difference is then compared to a threshold that separates non-objects from objects.

2.3 Making SURF descriptor

Since SURF is rotation invariant, rotation is handled by finding the direction of the feature and rotating the sampling window to align with that angle. Construct a square region centered around the feature point. The size of window is 20sX20s that is taken around the detected interest point, where s is the size. Once the rotated neighborhood is obtained it is divided up into 16 sub squares. Each sub square is again divided into 4 squares.

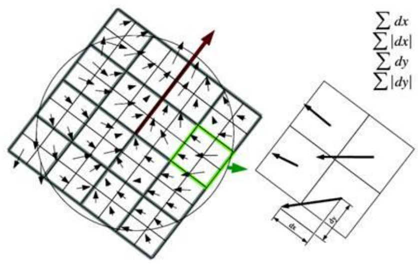


Fig.3 Graphical representation of SURF descriptor

2.4 Descriptor Computation

For each sub-region compute Haar wavelet responses in horizontal and vertical directions and a vector,V which includes summation of d_x, $|d_x|$, d_y, $|d_y|$ is formed. Derivatives in the x and y directions are taken in these final squares. The descriptor for the sub square is the sum of the x derivatives over its four quadrants, and similarly for y derivative. The total descriptor has 4 values. The vector is normalized to length 1 and is the feature descriptor. The vector gives the SURF feature descriptor with total 64 dimensions. Lower the dimension, higher the speed of computation and matching, but provide better distinctiveness of features.

3. Harris corner detection algorithm

Corner detection is an approach used to extract certain kinds of features and image contents. Corner detection is used in motion detection, image registration, video tracking, panorama stitching, and object recognition etc. The Corner Detection block finds corners in an image using the Harris corner detection. Harris detector considers the differential of the corner score with respect to direction directly, instead of using shifted patches.

3.1 Corner point

A corner point is the intersection of two edges. Corners represents a large variation in the gradient in an image. Corner is a point in which the directions of two edges *change*. Hence, the gradient of the image in both directions have a high variation, that can be used to detect it.

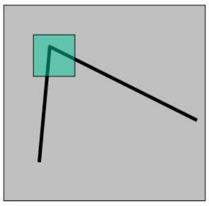


Fig.4 Corner point repsentation

A corner point can be easily recognized by looking into a window while shifting the window. For example a corner point is shown above. Shifting the window in any direction should give large change in intensity.

3.2 Edge and corner Differentiation

Edge is nothing but side or boundary of an object. For a square, its fours side are represented as its edges. Meeting point of tewo edges is a corner indicating a square has four corners.

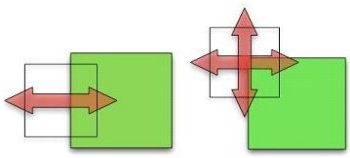


Fig.5 Edge vs Corner

In an image if the window is over a flat region, then there will be obviously be no intensity change when the window moves as shown in the above figure. If the window is over an edge there will only be an intesity change if the window moves in one direction. If the window is over a corner then there will be a change in all directions, and therefore we know there must be a corner. The Harris corner detector measures the strength of detected corners.

3.3 Formulation

Harris corner detector is based on the local autocorrelation function of a signal. The local auto-correlation function measures the local changes of the signal by patches shifted in different directions. Algorithm of harris corner detector is:

- To find partial derivatives from intensity of an image
- Compute corner response(R)
- Find local maxima in the corner response

In harris corner detection x and y derivatives is computed for an image. Product of derivative is determined for each pixel. Next sum of product is computed. Harris matrix in defined at each pixel(x,y). Corner response is computed and local maxima in corner response is found out. Auto correlation function is also called as summed square difference(SSD). For a point (x.y), its auto correlation function is represented as,

$$S(x,y) = \sum_u \sum_v w(u,v)\ (I(u+x, v+y) - I(u,v))^2 \qquad (2)$$

Where I(u+x,v+y) is approximated by taylor expansion. I_X and I_Y are known are partial derivatives such that

$$I(u+x,v+y) = I(u,v) + I_x(u,v)x + I_y(u,v)y \qquad (3)$$

The partial derivates can be calculated from image with a filter as [-1,0,1] and [-1,0,1]. Let Ω_1 and Ω_2 be two eigen values of autocorrelation function S(x,y). Auto-correlation matrix(H) captures intensity structure of the local neighborhood and it measure intensity based on the eigenvalues.

Three cases arrives:

- If both eigen values are high => Interest point(corner) is detected
- If one eigenvalue is high => Then it is contour
- If both eigen values are small => It is uniform region

Corner response is characterized based on eigen values. Corner reponse is represented as,

$$R = Det(H) - K(\ Trace(H))^2 \qquad (4)$$

Where H is the autocorrelation matrix and K is a constant such that K= 0.04- 0.06. Det(H) is the product of eigen values and trace(H) is the sum of eigen values. R depends only on the eigen values of H. R value is larger for corner, small for flat region and negative for edge.

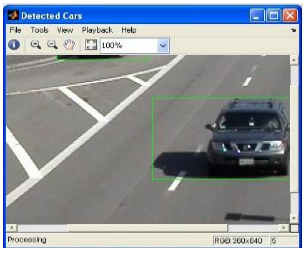
Fig.6 Image before corner detection

Fig.7 Corner points detected after harris corner detection

Result for corner detection is shown above. Harris corner detection algorithm is applied for a sample image shown in the fig 5and the fig 6 shows detected corner points in the sample image marked as green.

4. Results and discussions

First step is initialization. During initialization, we compute the full SURF feature descriptors from the first video image and match them. Then interest points are detected using corner detection. From a video sequence of frames are created features are extracted from objects using SURF descriptor. Corner points are detected using harris corner detection method. Features are matched in images and object is tracked and recognized.



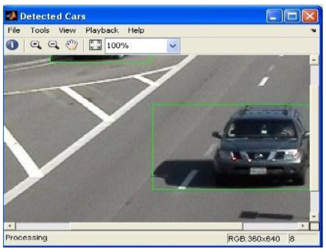
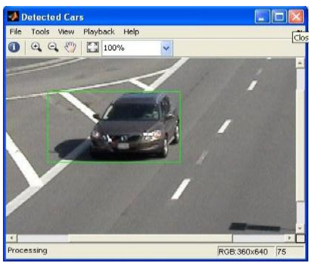


Fig.8 Detected car in frame number 5

Fig.9 Detected car in frame no 8

Car detection is done using SURF descriptor and corner detection. First foreground image is found in frames created. As camera is fixed, pixel values for background remains constant but on the otherhand pixel values of foreground object varies continuously.



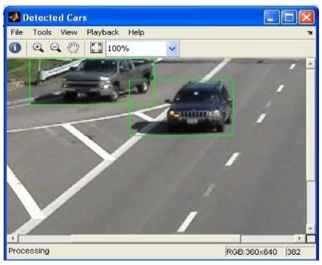


Fig.10 Detected car in frame no 75

Fig.11 Detected car in frame no 382

Background image is subtracted in consecutive frames and foreground image is extracted. Fig.9 and fig.8 are consecutive video frames created from video file. In these frames it is noticed that there is slight change in motion of the car. Thus pixel values changes for foreground image and pixel values for background remains same and is subtracted. Similarly fig.10 and fig.11 shows detected cars in frame number 75 and 382. Using harris corner detection interest points are extracted in detected foreground objects. Features are matched in consecutive frames and object is continuously tracked. Tracked cars are shown in green coloured bounding boxes as shown in the above figures.

Conclusions

Combined algorithm including SURF descriptor with harris corner detector is implemented which improves the tracking efficiency. SURF is responsible for fast feature extraction since it is desgined to be rotation invariant and is uses Haar wavelet filters which performs a fast filtering operation. Haar wavelet is meant for its fast calculation and interest points detection. SURF is good at handling blur images since it make suse of integral images. Thus SURF feature extraction is followed by harris corner detection algorithm which easily detects corner points thus results in excellent tracking.

References

E.Rosten and T.Drummond, (2005), Fusing points and lines for high performance tracking, *IEEE International Conference on Computer Vision*, volume 2.

E.Rosten and T.Drummond, (2006), Machine learning for high-speed corner detection, *European Conference on Computer Vision*, volume 1, pages 430–443.

D.Chekhlov, M.Pupilli, W.Mayol-Cuevas, and A. Calway, (2006), Realtime and robust monocular SLAM using predictive multi-resolution descriptors, *2nd International Symposium on Visual Computing*.

C.Harris and M.Stephens, (1998), A combined corner and edge detection, *Proc. of The Fourth Alvey Vision Conference*, pages 147–151.

Jialue Fan, Xiaohui Shen and Ying Wu,(2013), What Are We Tracking: A Unified Approach of Tracking and Recognition, *IEEE transactions on image processing*,vol. 22, no. 2.

C.Liu, J.Yuen, A.Torralba, J.Sivic, and W.T. Freeman,(2008), SIFT Flow:Dense correspondence across different scenes, *Proc. of the 10th European Conference on Computer Vision*.

D.Lowe, (2004), Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, 60(2):91–110

B.D.Lucas and T.Kanade, (1981), An iterative image registration technique with an application to stereo vision, *IJCAI81*, pages 674–679.

K.Mikolajczyk and C.Schmid, (2004) Scale & affine invariant interest point detectors, *Int. J. Comput. Vision*, 60(1):63–86

K.Mikolajczyk and C.Schmid, (2005), Performance Evaluation of Local Descriptors, *IEEE Trans. Pattern Anal. Mach. Intell*, 27(10):1615–1630