ISSN: 0067-2904

Object Tracking and matching in a Video Stream based on SURF and Wavelet Transform

Ekhlas Falih Nasser*, Abdul Alameer Abdulla Karim
Department of Computer Sciences, University of Technology, Baghdad, Iraq.

Abstract

In computer vision, visual object tracking is a significant task for monitoring applications. Tracking of object type is a matching trouble. In object tracking, one main difficulty is to select features and build models which are convenient for distinguishing and tracing the target. The suggested system for continuous features descriptor and matching in video has three steps. Firstly, apply wavelet transform on image using Haar filter. Secondly interest points were detected from wavelet image using features from accelerated segment test (FAST) corner detection. Thirdly those points were descripted using Speeded Up Robust Features (SURF). The algorithm of Speeded Up Robust Features (SURF) has been employed and implemented for object in video stream tracking and matching. The descriptor of feature in SURF can be operated by minimizing the space of search for potential points of interest inside the scale space image pyramid. The tracked interest points that are resulted are more recurrence and pother free. For dealing with images that contain blurring and rotation, SURF is best. Fast corner detector can be employed along SURF method to build integral images .The integral images can be used to enhance the speed of image matching. The features that are extracted from video images are matched using Manhattan distance measure. Apply the algorithm of FAST corner detection along SURF descriptor of feature; tracking and matching adequacy is better, fast and more efficient than Scale Invariant Feature Transform SIFT descriptor. The experimental outcomes displayed that the time that SURF could be taken for matching is less than the time that SIFT could be taken ,the SURF accuracy depends on number of key-points which are extracted from each frame. SURF key-points are less than SIFT key-points; therefore, SURF key-points could be considered optimal in the process of matching accuracy.

Keywords: Corner detection, FAST, Wavelet Transform, adaptive gray difference threshold, SURF, Integral Image

تتبع ومطابقة الجسم في سلسلة الفيديو بأستخدام "سيرف" وتحويل المويجة

أخلاص فالح ناصر*، عبدلأمير عبدالله كريم
قسم علوم الحاسوب، الجامعة التكنولوجية، بغداد، العراق

الخلاصة

في الرؤيةِ بالحاسوبِ، يكون تتبع الجسمِ بصرياً عمل مهم لمراقبة التطبيقاتِ. مشكلةً المطابقه تكون بتتبع نوعِ الجسمِ. الصعوبه الرئيسيه في تتبع الجسم هو اختيار الصفات وبناء النماذج المناسبه لتمييز وتتبع

*Email: ekhlas_uot1975@yahoo.com

الهدف.النظام المقترح لوصف الصفات ومطابقتها بأستمرار على الفيديو يتكون من ثلاث خطوات. اولا يتم تطبيق تحويل المويجه على الصوره بأستخدام Haar فلتر.ثانياً يتم أكتشاف النقاط المهمه في الصورة المضغوطه (wavelet) بأستخدام كاشف زاوية الصفات لأختبار المقطع السريع (FAST) .ثالثاً يتم وصف تلك النقاط المهمه بأستخدام الواصف تسريع الصفات القوي (SURF) . أستخدمت خوارزمية (SURF) ونفذت لغرض تتبع الجسم ومطابقته على السلسله الفيديويه.واصف الصفه (SURF) يستطيع العمل عن طريق تقليل فضاء البحث عن النقاط المهمه المحتمله داخل هرم فضاء الصوره. النقاط المتتبعه المهمه الناتجه كانت خاليه من الضوضاء والتكرار.عند التعامل مع الصور التي تحوي على تشويه وتدوير، يكون (SURF) هو الأفضل.كاشف الزاويه السريع نستطيع استخدامه على طول طريقة (SURF) لبناء الصور التكامليه.يتم استخدام الصور التكامليه لتحسين سرعة مطابقة الصوره.يتم مطابقة الصور المستخرجه من الفيديو باستخدام مقياس المسافه (Manhattan). عند تطبيق خوارزمية كاشف الزاويه السريع مع واصف الصفه (SURF) ; فان كفاءة التتبع والمطابقه تكون الأفضل ،سريعه واكثر كفاءه مقارنة مع واصف الصفه ذات مقياس التحويل الثابت(SIFT). أظهرت النتائج التجريبية أن الوقت الذي يمكن ان تأخذه (SURF) للمطابقة أقل من الوقت الذي يمكن ان تأخذه (SIFT)، وتتوقف دقة (SURF) على عدد النقاط الرئيسية المستخرجة من كل (Frame) . نقاط مفتاح (SURF) أقل من نقاط مفتاح (SIFT).لذلك يمكن اعتبار نقاط مفتاح(SURF) مثالية في عملية مطابقة الدقة.

1. Introduction

Scale Invariant Feature Transform (SIFT) and Speeded Up Robust Features SURF are powerful descriptors for feature and they became an essence ingredient in image recognition applications, and registration of an image. They are fit to predomination with important amounts of image inconstancy because these descriptors are steady under consideration and alteration in lighting. Simultaneously, discriminative strength can be checked by representing points of feature as vectors of high-dimensional [1]. Presenting an image for testing and a video stream as an input, it would like to distinguish the image name that present in the stream of video. Matching an image versus a database is executed for determining recent image which can be revealed in the stream of video, image tracking is executed to determine a position and the name of an image which has the similar object in the successive video frames. Strong feature points with descriptors of high dimensional can be achieved best for recognition of an image, therefore it can compute and track them at rates of interacting frame. For strong features tracking over video frames, Battiato and Skrypnyk and Lowe can execute pairwise matching for an image by giving any two sequent frames of video. The stream of an algorithm of this approach can be illustrated in Figure-1. Matching of frame-to-frame has main drawback which is wasted in computation because it does not take advantage of video's coherence. For tracking purposes only the features are utilized most of the time, which is no required to achieve recognition move on an image, unless an important current frame's modification is discover. A crossbred algorithm that employed for estimation of motion is implementing through features of lightweight like Harris corners or FAST corners and confession of an image is implementing cross strength attributes. It can run a criterion algorithm for matching an image versus the database of an image to discover what is ready in a scene when adequacy motion is cumulative for a step of recognition [2]. The solution is that, at corner positions which are applied for tracking, compute the strength descriptors [3]. These methods are reliable in tracking situation, because corners are not scale invariant which needs elicitation many descriptors per corner. It can decouple the computation of descriptor from detection of interest point therefore; descriptors of feature can be computed just for the aim of image recognition [4].

2. Related Works

The related work can be depended on two parts:

2.1 Robust Feature Detectors and Descriptors.

Many approaches for detection point of interest that are invariant to scale, such as Lowe offer Difference-of-Gaussians in SIFT, Matas introduce Maximally Stable Extended Regions (MSER), Mikolajczyk and Schmid introduce Harris-Affine and Hessian-Affine corners, and Bay introduce Hessians approximated using Haar-basis. Mikolajczyk execute a comparison research of comprehensive upon detectors.Strong descriptor algorithms occupy region detectors outcome, build a

canonical frame, and then for each region elicitation a feature vector of high-dimensional. The organization of descriptors is invariant to change in scale lighting, and rotational. SIFT and SURF is very popular examples. There are several trials to fast the descriptor computation algorithm. A reliable SIFT execuation on graphics processing unit (GPU) by Sinha. SIFT computation speed up using integral images proposed by Grabner [5].

2.2 **Applications of Feature Descriptors.**

There are numerous researchers which could be employed descriptors of feature in the recognition of object and retrieval of an image. For example, SIFT for efficient key frame retrieval in video that can be applied by Sivic and Zisserman. An object identification method based fast kernel employing SIFT features that can be proposed by Grauman and Darrell. Pyramidal k-means to build a tree for searching local attributes for quickly and active retrieval of an image, that can be used by Nist´er and Stew´enius .Also numerous trials to trace descriptor of robust attributes. For example, SIFT attributes tracking like to Figure-1 which is proposed by Skrypnyk and Lowe [4] and Battiato [3]. Other algorithms which can alter the position of descriptor computation by using various detectors of interest point.

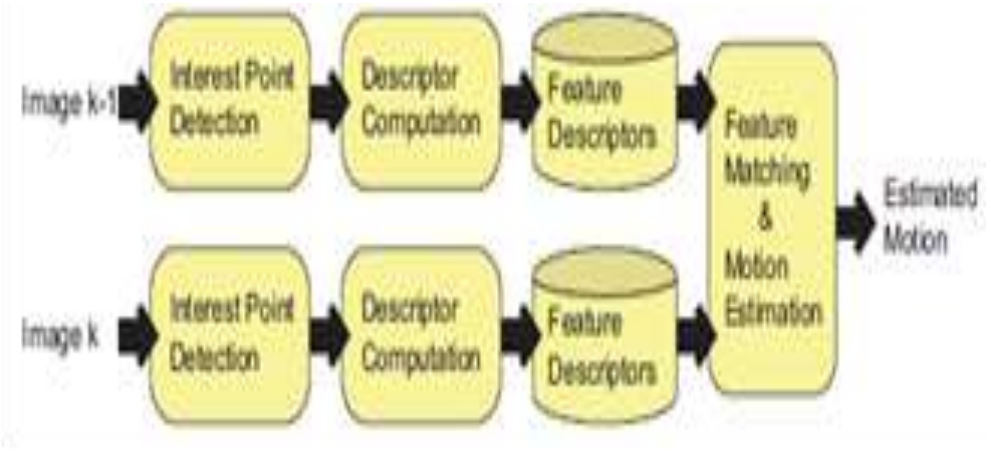


Figure 1- Feature Descriptor for Feature Matching and Motion Estimation.

3. Proposed Methodology

The suggested method in this paper for object tracking and matching based on SURF descriptor idea which looks the likeness among database of images. To track and match an object in database, the suggested method consists from major steps. Firstly, video frames series can be taken and these frames were transformed to frequency domain by applying Haar filter. Secondly, the features of interest are extracted using FAST corner detection from each frame. Thirdly, the interest features could be described using SURF descriptor and then tracked and matched using Manhattan measure. Figure- 2displays block diagram of the suggested system to identify an image label from video frames which has the similar object.

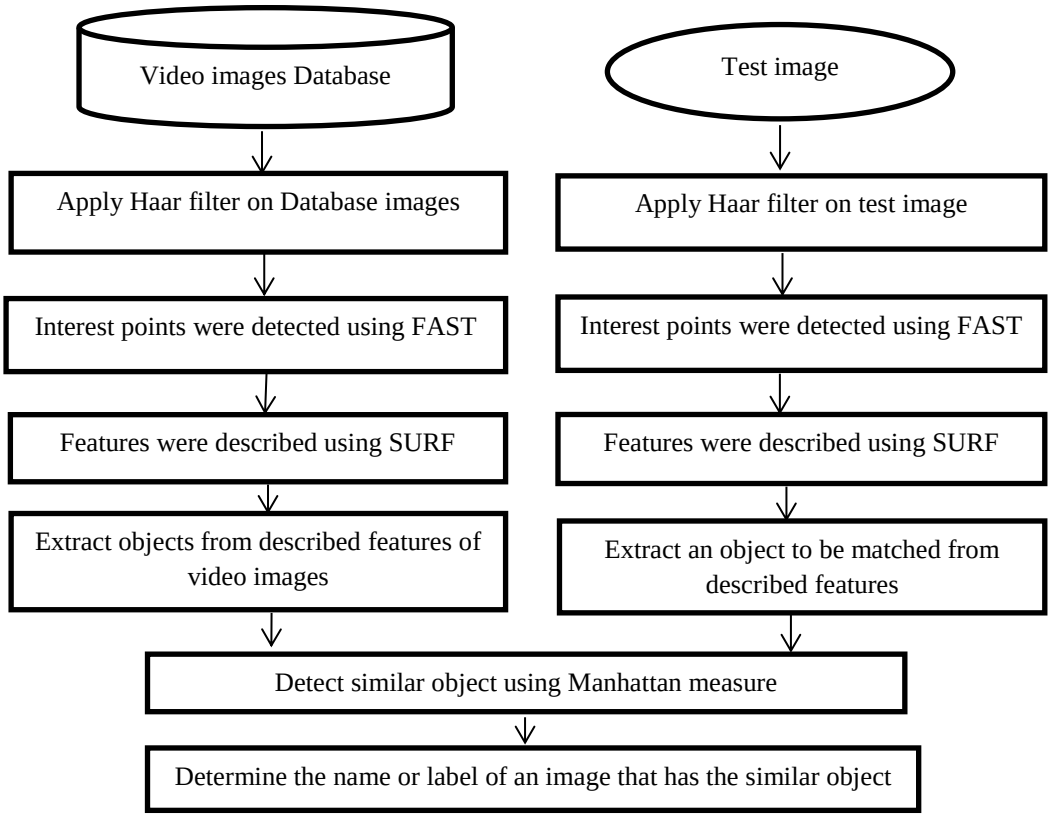


Figure 2- Block diagram of proposed system for object tracking and matching.

3.1 Wavelet Transform using Haar Filter

Wavelet transforms based on sub-sampling low pass and high pass filters (Quadrature Mirror Filters (QMF)). By splitting the data into low pass band and high pass band with or without losing any information, matching the filters is done. Wavelet filters can be organized for applications of a broad range and numerous different sets of filters can be proposed for various applications. Wavelets are functions identified over a finite interval. The purpose from wavelet transform is to transform the data from **Time-space domain** to **Time-frequency domain** which can perform best compression results. There are a wide variety of popular wavelet algorithms, including Daubechies wavelets, Mexican Hat wavelets and Morlet wavelets. These algorithms have the disadvantage of being more expensive to calculate than the Haar wavelets. Haar wavelet is a simplest form of wavelets; the function is defined in Eq.1. The four bands are indicates to Low-Low (LL), Low-High (LH), High-Low (HL) and High-High (HH). It can potential to implement group of wavelet filters on LL band with self-path as implemented to the main image because it contains image-like information. An image dividing operation into sub-bands can be permanent as far as wished (based on an image resolution), probably for image compression it is commonly continued only to 4 or 5 levels[6] .Figure -3 shows a wavelet transform on gray scale image.

$$\varphi(x) = \begin{cases} 1 & 0 \le x < 1/2 \\ -1 & 1/2 \le x < 1 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

a **b**

Figures 3- a) Original image b) Two Dimensional Haar Transform

3.2 Corner Detector using FAST

Corners are important local attributes in images. These points have high drooping and lie in cross brightness of an image areas. In a diversity of image attributes, corners cannot be affected by lighting and can be have rotational constancy. Corners form 0.05% only from all image pixels. Without missing data of image, elicitation corners can reduce image's data processing [7]. Therefore, the detection of corner has factual value and it plays an important place for motion tracking, image matching, augmented reality, representation of an image and other different fields [8]. Tremendous techniques for detection the corners was suggested from multiple searchers. These techniques are divided into two groups: group of techniques focus on contour and the other group focus on intensity. Techniques focus on contour work at first to extract all contours from an image and then seek for points that have maximum diversity over those contours [9].Feature from an accelerated segment test (FAST) uses a Bresenhams algorithm for circle drawing with diameter of 3.4 pixels for trial mask. Trial 16 pixels compared to the nucleus's value for a complete accelerated segment. The criterion of corner should be more relaxed to block this broad trial. A pixel's criteria must be a corner based on an accelerated segment test (AST) which there must exist at least S pixels that have more brilliant circle connection or darker than a threshold .To reduce feature space of an image and increase the implementation speed of the suggested system, our algorithm was used an adaptive threshold **thr** and it can be computed using Eq. 2. Other values of 16 pixels are disregarded. So the value of S can be used to determine the detected corner at maximum angle [10].

$$thr=(Img_{max} - Img_{min}) /2 \quad (2)$$

where Img_{max} and Img_{min} are the largest and smallest gray value of whole image.

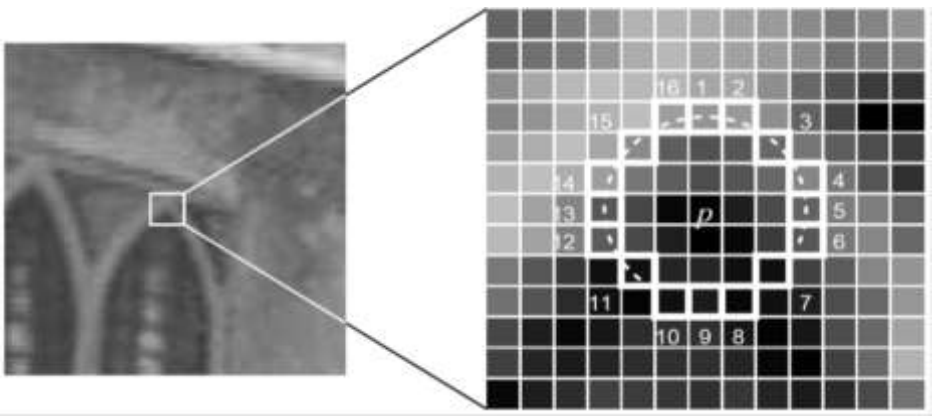


Figure 4-Image display the point of interest under a test and the circle of 16 pixel

Steps of FAST algorithm

1. From an image, chose a pixel p‟. IP represent pixel's intensity. This pixel can be specified as a point of interest or not. (Returning to Figure-4)).
2. Get **thr** from Eq. (2) that represents the value of threshold intensity.
3. Assume periphery a pixel p represents the center of circle which has16 pixels. (Brenham circle of radius 3.)
4. Need "N" exposure contiguous pixels out of the 16 pixels, either below or above IP by **thr** value, if the pixel wants to discover as a point of interest.
5. First match 1, 5, 9 and 13 of the circle pixels' intensity with IP to make an algorithm fast. From Figure-4, at least three of these four pixels should accept the norm of the threshold for this it subsist an interest point. P is not an interest point (corner) if at least three values of - I1, I5, I9 and I13 are not below or above IP + **thr** . For this, a pixel p can be rejected as a potential point of interest. Else if three pixels at least are up or down Ip + **thr**, for whole 16 pixels seek and check if 12 neighboring pixels drop in the norm.
6. A same procedure can iterate for whole image's pixels.

3.3 Features Description using SURF

SURF (Speeded Up Robust Features) can be widely employed for problem solving of the correspondence matching due to it was faster than SIFT (Scale Invariant Feature Transform) by briefness the showing of matching. To find candidate points, SIFT uses visual pyramids and based on the law of Gauss filters each layer with raise values of Sigma and determines differences. For image identification and matching, the proposed algorithm employs SURF descriptor for feature. Vectors of feature are elicitation by SURF which is stable to image rotation and scaling. Features can be matched using Manhattan distance measure. Local descriptors of SURF are better computational efficiency than local descriptors of SIFT because of integral images computed in SURF. At discrete locations, points of interest are chosen in the image such as corners. Every key point's neighborhood is represented by a vector of feature. The descriptor of feature has to be discriminative, robust to noise, errors' detection, deformations of geometric and photometric.

Finally the vectors of SURF descriptor are matched between various images. The matching is based on Manhattan dissimilarity.

To build feature space, SURF algorithm consists of various stages. These stages are detection of interest point, for each key point, SURF descriptor must be build, and descriptor matching [9].

3.3.1 Constructing Integral image

For SURF speed, integral images can be calculated. Image of integral is an intermediate representation and construct from the summation of image pixel values. It is also called as Summed Area Tables [10]. Integral image is given by Eq.3.

$$U(x,y) = \sum_{i \le x, j \le y} u(i,j) \qquad \ldots \qquad (3)$$

where u(i,j) represent the value of pixel at the locations i and j of the original image. U(x,y) represent the value of pixel at the locations x and y of the integral image.

3.3.2 Interest Point Detection

Fast Hessian feature detector can be used in SURF .It is based on the determinant of Hessian matrix. Hessian matrix consists of partial derivatives of two dimensional functions. Our algorithm uses FAST corner for detection the interest point and can be used in applications of real time.

3.3.3 Descriptor with SURF

Because SURF is stable to rotation, rotation can be processed by determining the direction of feature and rotating window's sampling to adjacency together with this angle. Build a quadrate area centered on the point of feature. Window volume that can take about **20sX20s** from the discovered interest point,where **s** represents the volume. When the rotated nearness is finding, it is split into 16 sub quadrates as illustrate in Figure- 5. Again every sub quadrates can be divided into 4 quadrates [11].

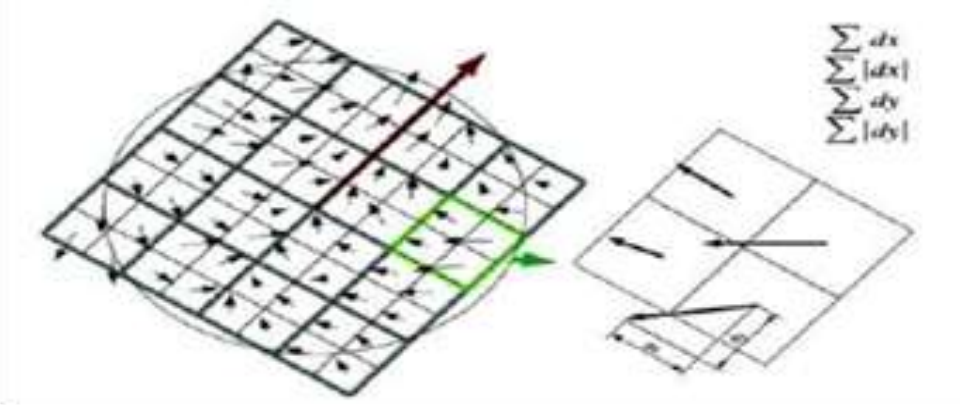


Figure 5- schematic impersonation for SURF descriptor

where dx and dy represent derivatives of x and y directions, |dx| and |dy| represent dx and dy normalization.

3.3.4 Computation for Descriptor

It computes Haar wavelet responses in horizontal and vertical directions for each sub-region and summation of dx, |dx|, dy, |dy| is formed and put in a vector V. For final squares, derivatives in the x and y directions are taken. The x derivatives summation over its four quadrants, similarly for y derivative is representing a descriptor for sub square. It has 4 values for total descriptor. Normalize V to length 1 and feature's descriptor. A vector supplies the feature descriptor of SURF with aggregate 64 dimensions. Providing good discriminative to features for lower dimension with maximum computation's speed and matching [11].

3.3.5 Feature Matching

Matching speed of feature is performed by a unique step of indexing for interest point which depends on the value of the Manhattan. Compute a distance that would be traveled through a Manhattan distance function to obtain with one point of data to another if a grid-like track is followed. The distance of Manhattan among two components is the summation of differences of their corresponding items [12].The distance's formula among a point **X**=(**X**1, **X**2, etc.) with a point **Y**=(**Y**1, **Y**2, etc.) is:

$$d = \sum_{i=0}^{n} |x_i - y_i| \quad (4)$$

3.4 Proposed algorithm

The algorithm of the proposed algorithm is illustrated as:

Input : video stream, image to be test and matched **Output :** name or label of the tested image in video frames
Step1: 1) Enter AVI video and Covert video stream into frames **(Imgs)** 2)Enter the test image for matching (**test**) **Step 2:** Compute Haar transform for frames (**(Imgs)** and test image (**test**) and put the result in **(H_imges**) and **(H_test**) respectively using Eq**.1**. **Step 3:** Detect the interest points for **(H_imges**) and (**H_test**) using FAST corner detection with adaptive threshold based on Eq.2 and put the results in **(D_H_Imgs)** and (**D_H_ test**) respectively. **Step 4:** Construct the integral images for both **(D_H_Imgs)** and (**D_H_ test**) using Eq**.(3)** **Step 5:** Track and match the object features of the test image with the features of video frames using Eq.**(4)** and determine the label or name of the occurrence of the test image on video frames

4. Experimental result

The outcomes of suggested method are offered and discussed at this part. The suggested method is executed in C#. Two types of databases like (Women conversation) video and (Tracery) video are employed for evaluation the suggested method. Database images are colored, and with size 320 × 240 pixels. The suggested method consists form multiple steps:-

1. At the first step, loading the video stream and tested image as shown in the Figure-6 for women conversation dataset and Figure-11 tracery dataset.

a) **b)** **c)**

Figure 6-Women conversation video, a) input video b) extraction frames for women conversation dataset c) input test image from women conversation data set

2. Second step exhibited initialization. During the initialization, it computes Haar transform on video frames and testing image as exhibited at Figure -7 for women conversation dataset and Figure- 12 tracery dataset.

(a) **b)**

Figure7-Haar transform on a) video frame b) test image.

3. he interest points are detected in the third step using FAST corner detection as exhibited at Figure -8 for women conversation dataset and Figure -13 tracery dataset.

(a) **b)**

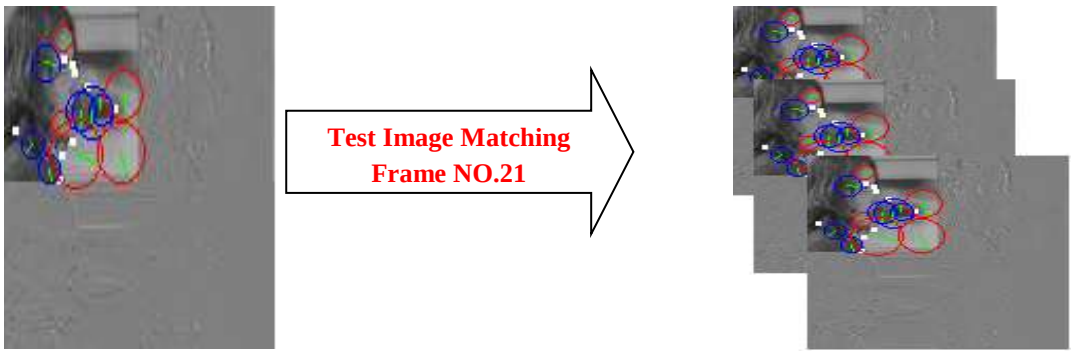
Figure 8- FAST corner detection on a) video frames b) test image

4. In the fourth step, complete SURF feature descriptor can be computed for the frames of video and tested image as exhibited at Figure-9 for women conversation dataset and Figure -14 tracery dataset.

a) **b)**

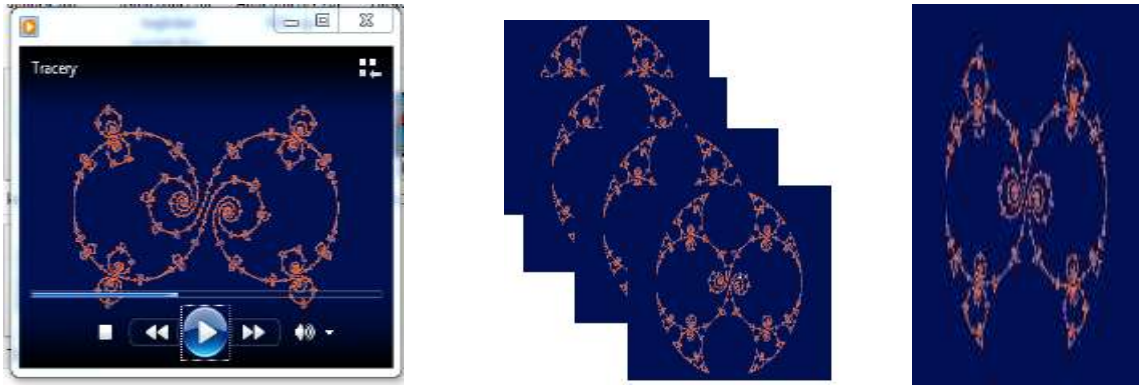
Figure 9- SURF descriptor on a) video frames b) test image

5. In the fifth step, tracking and matching between object of the test image and video frames objects to identify the frame label in video frames that has similar object as exhibited at Figure-10 for women conversation dataset and Figure -15 tracery dataset.



a) **b)**

Figure 10- The matching result of the proposed system between a) test image b) video frames



(a) **(b)** **(c)**

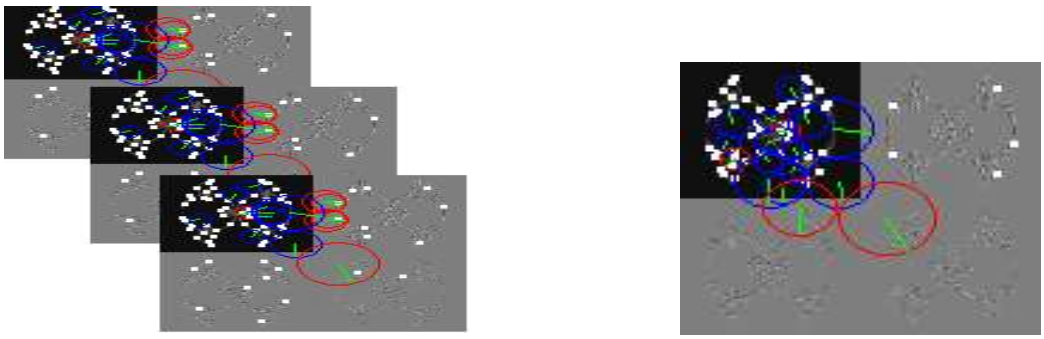
Figure 11- Tracery video a) input video , b) extraction frames for tracery dataset, c) input test image from tracery data set

a) **b)**

Figure 12- Haar transform a) video frame b) test image

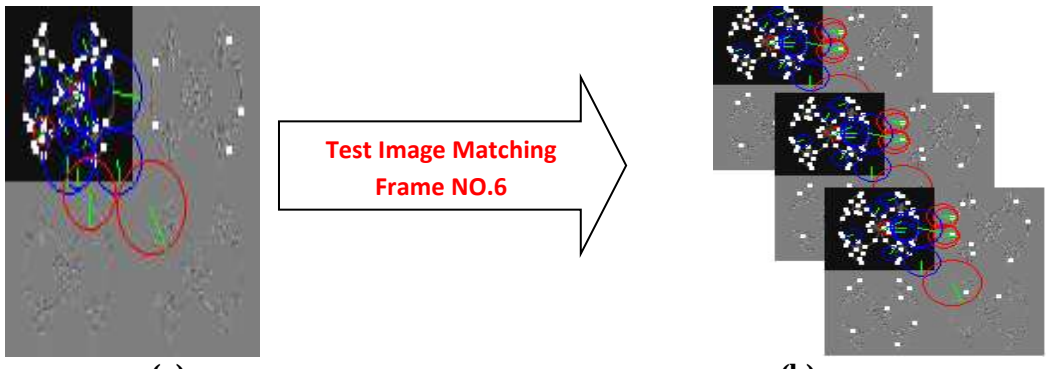
(a) **b)**

Figure 13- FAST corner detection on a) video frames b) test image



a) **b)**

Figure 14- SURF descriptor on a) video frames b) test image



(a) **(b)**

Figure 15- The matching result of the proposed system between a) test image b) video frames

Table 1- shows the comparison result in term of time consuming and accuracy for tracking and matching an object when SIFT and SURF were employed. Accuracy tracking depends on number of key points that are extracted from video frames. Accuracy matching depends on number of correct matching between test image and database images. For example, frame number (10) was selected to display the outcome of comparison**.**

Method name	Object name	Time(sec) on frame no.10	Accuracy tracking on frame no.(10)	Accuracy matching on frame no.(10)
SIFT	Women conversation	1.991	56	70%
	Tracery	2.549	105	87%
SURF	Women conversation	0.025	10	85%
	Tracery	0.813	50	92%

5. Conclusion

- Mixing between FAST corners detector with SURF descriptor is implementing to betterment the qualification of the tracking.
- Since SURF is designed to be rotation invariant, it is responsible for fast feature elicitation and employs the filters of Haar wavelet that implement a fast operation for filtering. Haar can be intended for its fast calculation.
- Since SURF make use of integral images, it is good for processing blur images.
- FAST algorithm for detection of corner that simply discovers points of corner followed by SURF feature extraction can make results in excellent tracking.

5.1 Advantages of Suggested Method

- Haar filter is memory efficient, due to it could be computed in place without using a temporary array.
- Employing FAST algorithm for detection of corner over SURF descriptor of feature, tracking and matching adequacy is best, fast and more efficient than SIFT algorithm.

5.2 Disadvantage of Suggested Method

- It requires object to be geometrically rich and the performance degrades in case of objects having low geometrical features. In that scenario, contour based techniques outperform this technique.

References

1. Duy-Nguyen. **2009.** SURFTrac: Efficient Tracking and Continuous Object Recognition using Local Descriptors. Certified by IEEE PDF eXpress, March.
2. Sergios, Theodoridis and Konstantions, Koutroumbas. **2009.** Pattern Recognition. Fourth Edition, Academic Press is an imprint of Elsevier30 Corporate Drive, Suite 400, Burlington, MA 01803, USA.
3. Wagner, D., Reitmayr, G., Mulloni, A., Drummond, T. and Schmalstieg, D. **2008.** Pose tracking from natural features on mobile phones. In Proc. ISMAR, Cambridge, UK.
4. Liu, C., Yuen, J., Torralba, A. Sivic, J. and Freeman, W. T. **2008.** SIFT Flow: Dense correspondence across different scenes. In Proc. of the 10th European Conference on Computer Vision.
5. Manuel, Blum and Martin, Riedmiller. **2012.** A Learned Feature Descriptor for Object Recognition in RGB-Data. In IEEE International Conference on Robotics and Automation in Shanghai, China, pp.2-4.
6. David, Salomon. **2007.** Data Compression", the Complete Reference. Fourth Edition, Professor David Salomon (emeritus) Computer Science Department, California State University, Northridge, CA 91330 8281, USA, Springer-Verlag London Limited.
7. Wenjia, Yang, Lihua, Dou, Juan Zhang, Jinghua Lu. **2007.** Automatic Moving Object Detection and Tracking in Video Sequences. *SPIE Fifth International Symposium on Multispectral Image Processing and Pattern Recognition*, pp.676-712.

8. Asif, Masood, Sarfraz, M. **2007.** Corner detection by sliding rectangles along planar curves", *Computers & Graphics*, **31**: 440-448.
9. Jasmine, J., Anitha Ȧ* and.Deepa, Ȧ. S.M. **2014.** Tracking and Recognition of Objects using SURF Descriptor and Harris Corner Detection. ȦNehru Institute of Engineering and Technology (Anna University), Coimbatore, India, **4**(2): 1-6.
10. Bay, H., Ess, Tuytelaars, A. T. and Van Gool, L. **2008.** Speeded-up robust features (SURF). *Comput. Vis. Image Understand.*, **110**(3): 346–359.
11. Aswini, C., and Chitra, D. **2014.** Enhanced Logo Matching and Recognition using SURF Descriptor. Department of Computer Science and Engineering, P. A. College of Engineering and Technology, Pollachi, Tamil Nadu, India.
12. Jiawei, Han, and Micheline, Kamber. **2011.** *Data mining, concepts and techniques*. Third Edition book, Morgan Kaufmann Publishers is an imprint of Elsevier.225 Wyman Street, Waltham, MA 02451, USA.