English Character Recognition from Video Stream based on Bag of Visual Words (BOVW)

Ekhlas Falih Nasser *Dr.Abdulameer A. Karim*

ekhlas_uot1975@yahoo.com *ameer_aldelphi@yahoo.com*

Abstract

Numerous digital images are available for printing the documents. The discussions are continuing for arriving to best algorithm for identifying the English letters. The suggested method has four steps. Firstly, apply wavelet transform on images of letters using Haar filter. Secondly interest points were detected using features from accelerated segment test (FAST) corner detection. Thirdly those points were descripted using Speeded up Robust Features (SURF). Fourthly, the clustering algorithm of moving k-means is employed to obtain bag of visual words (BOVW) and then build vocabulary and a histogram from visual words.The features of each visual word for video images and test image are matched using Manhattan distance measure. The suggested system was tested on three types of English letters font's databases (Time New Roman, Arial Black and Calibri) .Experimental outcomes show that the suggested method is more efficient and fast for matching and recognizing a letter than seven moment's method. The recognition time for BOVW is less than the seven moment's time and the BOVW accuracy depends on number of correct character recognition. BOVW have optimal accuracy in the process of recognition of letters.

Keywords: Haar filter, FAST, SURF, moving k-means clustering, Manhattan distance

تمييز الحرف الأنكليزي من سلسلة الفيديو بالأعتماد على حقيبة من الكلمات المرئية

م.أخلاص فالح ناصر **أ.م. د.عبدالأميرعبدالله كريم**

الجامعة التكنولوجيا / قسم الحاسوب

الخلاصة

هنالك الكثير من الصور الرقمية المتوفره للوثائق المطبوعة والبحث مازال مستمر للوصول الى افضل خوارزميه للتعرف على الحروف الأنكليزية. الطريقه المقترحه تتكون من أربع خطوات. اولا يتم تطبيق تحويل المويجه على الصورالحرفيه بأستخدام Haar فلتر. ثانياً يتم أكتشاف النقاط المهمه بأستخدام كاشف زاوية الصفات لأختبار المقطع السريع (FAST). ثالثاً يتم وصف تلك النقاط المهمه بأستخدام الواصف تسريع الصفات القوي (SURF). رابعاً خوارزمية k-means moving تستخدم لتجميع الصفات بشكل عناقيد(clusters) للحصول على حقيبه من الكلمات المرئيه وبعدها يتم بناء المعجم و مخطط للكلمات المرئيه.صفات كل كلمه مرئيه في صور الفيديو والصوره المختبره يتم مطابقتها باستخدام مقياس المسافه منهاتن. تم اختبار النظام المقترح على ثلاثة أنواع من قواعد بيانات الحروف الإنجليزية (تايم نيو رومان، أريال بلاك و كالبري). تبين النتائج التجريبية أن الطريقة المقترحة هي أكثر كفاءة وسرعة لمطابقة وتمييز الحرف مقارنة مع استخدام طريقة العزوم السبعه.الوقت الذي تأخذه حقيبة الكلمات المرئيه لتمييز الحرف يكون اقل من الوقت الذي تاخذه العزوم السبعه، دقة حقيبة الكلمات المرئيه يعتمد على عدد تمييز الحروف بصوره صحيحه. لذا تعتبر حقيبة الكلمات المرئيه ذات الدقه الأمثل الدقة المثلى في عملية تمييز الحروف.

الكلمات المفتاحية: تسريع الصفات القوي و تمييز الحروف و فلتر هار

1. Introduction

Employing digital documents images were increased in the recent years and they can be considered as resources for management and education in many applications. They are not dealing directly with digital images contents; because of they are complete digital images. For document's text accessing, it must be recognized text contents and divided it into symbols and letters, which the document wrote. The task which can be performed daily was called recognition of an object. It can be performed in the more varied circumstances, navigation to the sources of food, migration, predators' identification, mates' identification, etc. with efficiency remarkable [1]. Methods can be developed, which are capable for emulating different recognition of object forms that are evolving along with the need for constructing systems of an intelligent automated. Today's technology's main direction in industry and other activity domains also in these objects systems are represented in a convenient path for processing type, and these representations can be called *patterns*. The scientific discipline *that* deals with object classification *and* description methods is called *Recognition of Pattern* (RoP). Recognition of pattern is therefore a fruitful region of research, with doubled links to numerous other disciplines; including professionals from many regions [2].

2. Moment Invariants

Moment invariants are employed in numerous applications for recognition of pattern. Invariants of image's feature or shape's feature stay without modification if that shape or an image submit any mixing of the following alterations

- Modification in size (Scale)
- Modification in size position (Translation)
- Modification in Orientation (Rotation)
- Reflection

The invariants of moment are extremely beneficial method for features' eliciting from two-dimensional images. Invariants of moment are characteristics of regions which are connected in binary format of images, which are invariant to scaling, translating, rotation and.

Invariants of moment are beneficial because they can be determined by simple computed set of region characteristics which can be employed for classification of the shape and recognition of the part. Suppose a two-dimensional *F(x,* y) in a spatial domain [Ach 05].*Geometric moment* of order $p + q$ is illustrated in Eq.(1)

$$m_{p,q} = \sum_x \sum_y x^p y^q F(x,y) \quad \ldots (1)$$

For p , q = 0 , 1 , 2 , ... The moments' central are explained by

$$x_c = m_{1,0} / m_{0,0}$$

$$y_c = m_{0,1} / m_{0,0}$$

Where $m_{1,0}$ in Eq. (1) and ***(xc,*** yc) can be called the object's region center. The normalized moment's central that represented by $\eta_{p,q}$ can be defined by Eq.(2)

$$\eta_{p,q} = \mu_{p,q} / \mu^{\gamma}{}_{0,0} \quad \ldots (2)$$

Where

$$\gamma = p + q / 2 \quad \ldots (3)$$

A set of seven invariants can be derived from the second and third normalized central moments. This set of seven HU moment invariants (4) to (10) is invariant to translation, rotation, and scale change.

$$\phi 1 = \eta_{2,0} + \eta_{0,2} \quad \ldots (4)$$

$$\phi 2 = (\eta_{2,0} + \eta_{0,2})^2 + 4\eta_{1,1} \quad \ldots (5)$$

$$\phi 3 = (\eta_{3,0} - 3\eta_{1,2})^2 + (3\eta_{2,1} - \eta_{0,3})^2 \quad \ldots (6)$$

$$\phi 4 = (\eta_{3,0} + 3\eta_{1,2})^2 + (3\eta_{2,1} + \eta_{0,3})^2 \quad \ldots (7)$$

$$\begin{aligned}\phi 5 = {} & (\eta_{3,0} - 3\eta_{1,2})(\eta_{3,0} + 3\eta_{1,2})[(\eta_{3,0} + 3\eta_{1,2})^2 \\ & -3(\eta_{2,1} + \eta_{0,3})^2] + (3\eta_{2,1} - \eta_{0,3})(\eta_{2,1} + \eta_{0,3}) \\ & [3(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{2,1} + \eta_{0,3})^2]\end{aligned} \quad \ldots (8)$$

$$\phi 6 = (\eta_{2,0} + \eta_{0,2})[(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{2,1} - \eta_{0,3})^2] + 4\eta_{1,1}(\eta_{3,0} + \eta_{1,2})(\eta_{2,1} - \eta_{0,3}) \quad \ldots (9)$$

$$\begin{aligned}\phi 7 = {} & (3\eta_{2,1} - \eta_{0,3})(\eta_{3,0} + \eta_{1,2})[(\eta_{3,0} + \eta_{1,2})^2 \\ & -3(\eta_{2,1} + \eta_{0,3})^2] + (3\eta_{1,2} - \eta_{3,0})(\eta_{2,1} + \eta_{0,3}) \\ & [3(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{2,1} - \eta_{0,3})^2]\end{aligned} \quad \ldots (10)$$

3. Proposed methodology based on Bag of Visual Words (BOVW)

The suggested system in this paper for English character recognition is based on bag of visual words (BOVW). It has four steps to build BOVW. Firstly, get the images of letters from video series. Secondly, images of letters can be transformed to frequency domain by applying

Haar transform. Thirdly, the features of interests are elicited using FAST corner detection. After eliciting interest features from all images of letters, the interest features could be described using SURF descriptor. Fourthly, the algorithm of clustering that using moving k-means is employed for described features clustering to build visual words for each image of letter. Each visual word corresponds to the center of cluster, and then builds vocabulary from all visual words. An image of letter can be entered as query image and the numbers of features are elicited from a query. The elicited numbers of features are matched with all images of letters in vocabulary using Manhattan distance and discover the similar image of letter which is matching with query image of letter. After detected the similar image, it can recognize the name and a sequence of an image in that vocabulary. Figure 1 displays block diagram of the system for retrieve an image in a proposed method

.

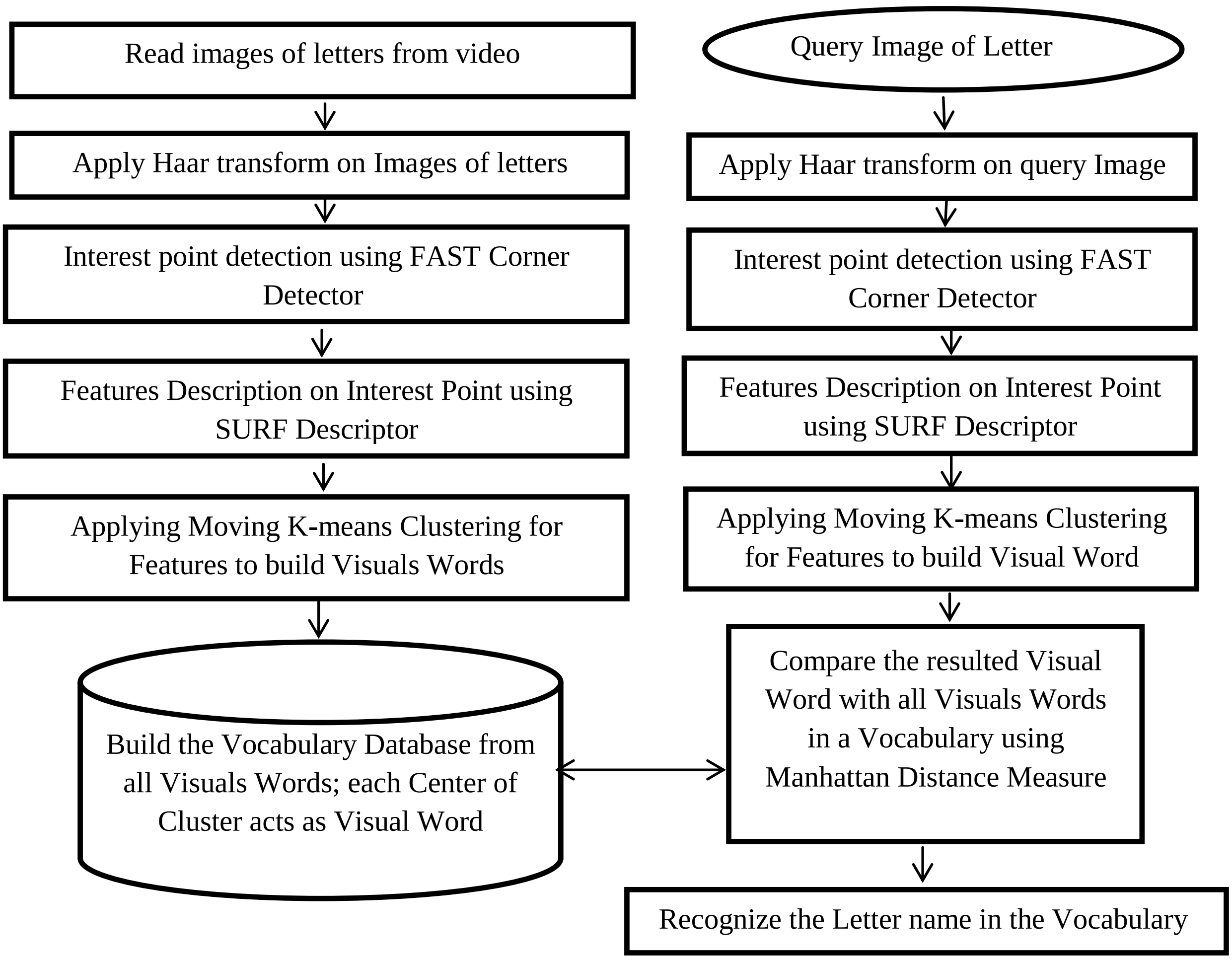


Figure (1): Block diagram of proposed system for character recognition

3.1 Haar Transform

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.11. 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[5].Figure (2) shows a wavelet transform on gray scale image.

$$\varphi(x) = \begin{cases} 1 & 0 \le x < 1/2 \\ -1 & 1/2 \le x < 1 \\ 0 & \text{otherwise} \end{cases} \quad \dots (11)$$

a b

Figures (2):Gray_Scale image a) Original image b) Two Dimensional Haar Transform

3.2 FAST Corner detection

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 [6]. 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 [7]. 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 [8].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. (12). Other values of 16 pixels are disregarded. So the value of S can be used to determine the detected corner at maximum angle [9].

$$thr=(Img_{max} - Img_{min})/2 \quad \dots (12)$$

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

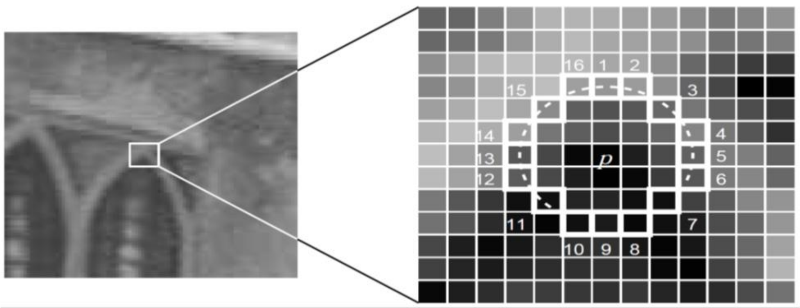


Figure (3): Image display the point of interest under a test and the circle of 16 pixels

3.2.1 Steps of FAST algorithm

The basic steps of FAST algorithm of detection the corners are illustrated bellow:-

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 (3)).
2. Get **thr** from Eq. (12) 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 (3), 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 is 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 [8].

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. (13).

$$U(x,y) = \sum_{i\le x, j\le y} u(i,j) \qquad \dots (13)$$

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 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 (4). Again every sub quadrates can be divided into 4 quadrates [10].

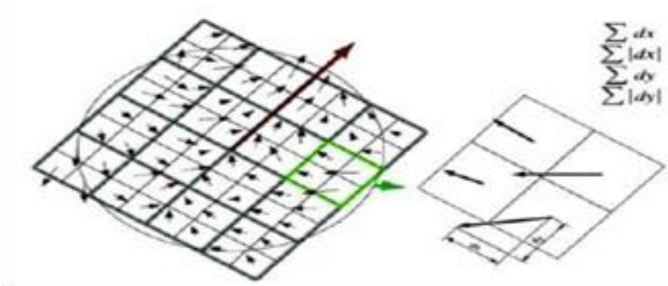


Figure (4): schematic impersonation for SURF descriptor

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 [10].

3.4 Clustering of features using Moving k-means

Clustering means collect elements of data from a data set to clusters of several likeness norms. In Content Based Image Retrieval (CBIR) systems likeness among the database of images is not see just the likeness between database images and query image that is used for retrieval . In this trouble of imposing computation time which is existing because comparison of features of query image with all images features in database. Clustering is employed to decrease the time of computation which considers likeness among the database images. There is no requiring after clustering for query image comparison with all images in database which decreases the time of computation and accuracy improvement. The algorithm of clustering for k-means is very straightforward for execution. The generations of clusters are not perfect in quality and time consuming for clusters generation is very imposing which is the problem in this way. An algorithm of clustering with moving k-means provides perfect time consuming and well clusters' quality to introduce cluster is less than the algorithm k-means [11]. An algorithm of clustering with moving k-means has been employed.The main steps of moving k-means clustering algorithm are illustrated bellow:-

1) The user has been specified the numbers of coveted clusters as input.

2) The set of factures that can be resulted from SURF descriptor are split randomly into number of coveted clusters based on Variance. The variance also tells something about the contrast.

$$\mathbf{variance} = \sqrt{\sum_{\mathbf{g}=0}^{\mathbf{L}-1}((\mathbf{g} - \mathbf{Mean})^{2} * \mathbf{p}(\mathbf{g}))} \quad \dots(\mathbf{14})$$

L :-is the total number of gray levels available , for example, for typical 8-bits image data l is 256 and range from 0 to 255.

g:-is the gray level value.

p (g)= N(g)/X*Y.

N (g):- is the number of pixels at gray level p.

P (g):- probability of each gray level value.

The midst point is treated as the cluster's centroid in every cluster. A distance among every data point to whole an initial centroid can be computed and a point of data is specified to cluster with aftermost centroid.

3) Through this, the cluster is enrolled to which data item can be specified and item distance of data for that cluster can be maintained. Cluster's centroids the can be recomputed.

4) From current after most cluster centroid ,compute a distance again for every data point if the distance is equal or lower than current nearest distance, a point of data in the selfsame cluster stays else data item distance is computed from whole the centroids.

5) The procedure continued until the convergence norm is not satisfied.

3.5 Building the Vocabulary

The total numbers of features which can be extracted from the descriptors are huge. To solve this problem, the feature descriptors are clustered by applying the clustering algorithm, such as moving K-Means technique to generate a visual vocabulary. Each cluster can be treated as a distinct visual word in the vocabulary, which is represented by their respective cluster centers. The size of the vocabulary is determined using the clustering algorithm. In addition, it depends on the size and the types of the dataset [12]. The frequency of each word can be considered as the number of features in the corresponding cluster.

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 [13].The distance's formula among a point **X**=(**X**1, **X**2, etc.) with a point **Y**=(**Y**1, **Y**2, etc.) is :

$$\mathbf{d} = \sum_{i=0}^{n} |x_i - y_i| \qquad \dots (15)$$

3.6 Proposed algorithm

The algorithm of the proposed algorithm is illustrated as:

Input : video stream of images of letters, image of letter to be test and recognized

Output : name of letter in the tested image of letter of video frames

Step1: 1) Enter AVI video and Covert images of letters of 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.**11**.

Step 3: Detect the interest points for **(H_imges**) and (**H_test**) using FAST corner detection with adaptive threshold based on Eq.12 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**.(13)**.

Step 5: Apply the algorithm of moving k-means clustering for all features descriptor of video frames and test image by using Eq.(**14**).

Step 6: //To Create vocabulary

6.1) Each center of cluster is treated as a distinct visual word for video frames image can be called **(video_visual_words)** and the center of cluster for test can be called **(test_visual_words)**.

6.2) build the vocabulary for all visual words from video **(video_visual_words)**.

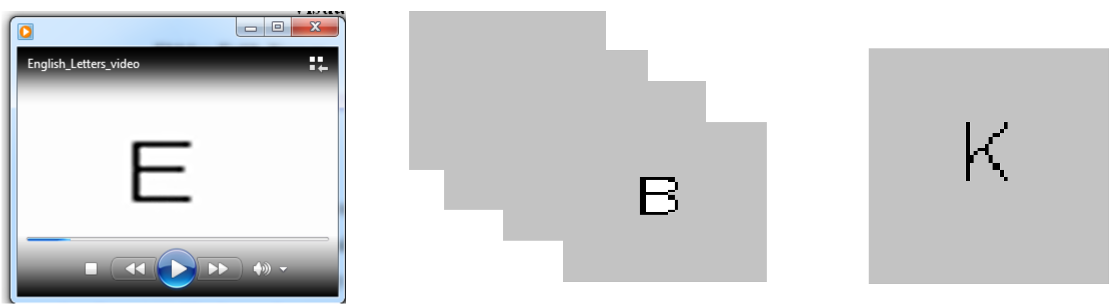
6.3) The frequency of each word can be considered as the number of features in the corresponding cluster.

Step 7: Track, match and recognize the letter features of the **(test_visual_word)** with the letters features of **(video_visual_words)** using Eq.**(15)** and recognize the name of the letter based on matching process.

4. Experimental results

The outcomes of suggested method are offered and discussed at this part. The suggested method is executed in C#. Three types of databases like ((Time New Roman, Arial and Calibri) video are employed for evaluation the suggested method. Database images are colored, and with size 256 × 256 pixels. The suggested method consists from multiple steps:-

1) At the first step, loading the video stream and tested image as shown in the figure (5) for Time New Roman database and figure (12) for Arial database.



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

Figure (5): Time New Roman Database, a) Input Video of Images of Letters b) Extraction Frames from vidro c) Input Test Image of Letter from Time New Roman Database

2) Second step exhibited initialization. During the initialization, it computes Haar transform on video frames and testing image as exhibited in figure (6) for Time New Roman database and figure (13) for Arial database.

a) **b)**

Figure(6): Haar transform on a) video frames b) test image

3) The interest points are detected in the third step using FAST corner detection as exhibited in figure (7) for Time New Roman database and figure (14) for Arial database.

a) **b)**

Figure (7): 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 in figure(8) for Time New Roman database and figure (15) for Arial database.

a) **b)**

Figure (8): SURF descriptor on a) video frames b) test image

5) In the fifth step, clustering the features that are resulted from SURF descriptor, visual words can be built from each cluster's center with its histogram and then construct bag of visual words as exhibited at figure (9) for Time New Roman database and figure (16) for Arial database.

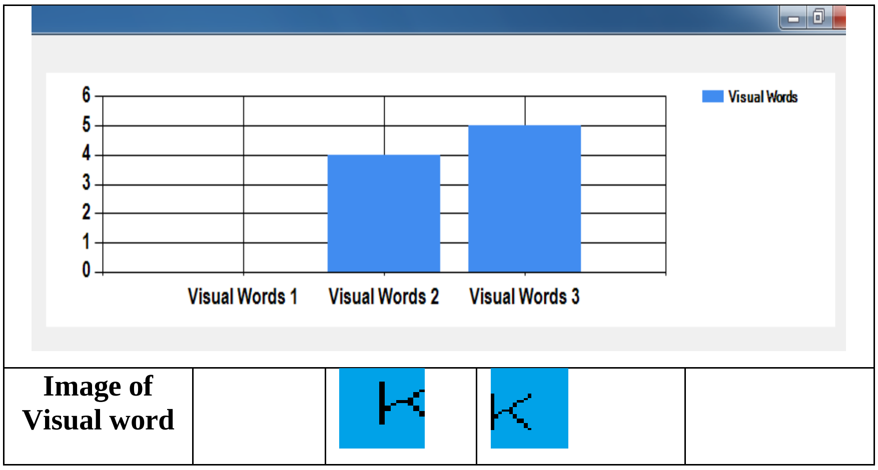


Figure (9):Histogram and Visual words for Letter K in Time New Roman Database the frequency of each word can be considered as the number of features in the corresponding cluster.

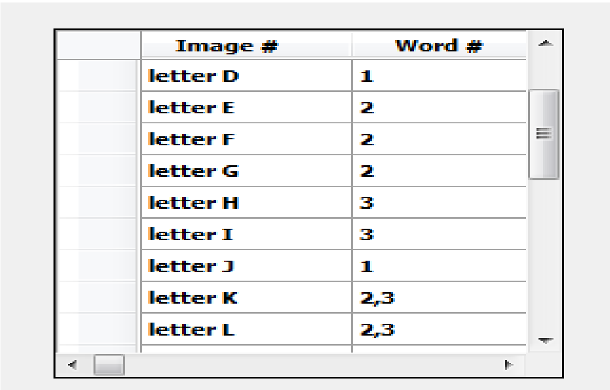
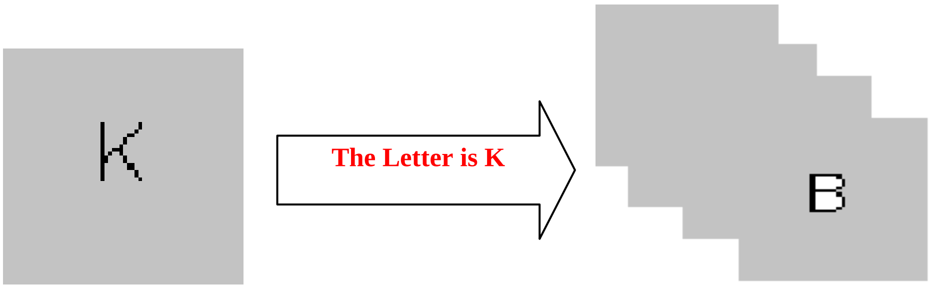


Image #	Word #
letter D	1
letter E	2
letter F	2
letter G	2
letter H	3
letter I	3
letter J	1
letter K	2,3
letter L	2,3

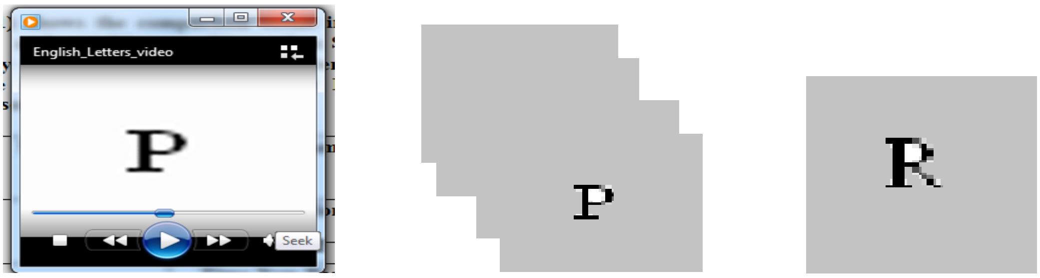
Figure (10):Vocabulary of all Images of Letters for Time New Roman Databse and Corsponding Visual Words Numbers

6) In the sixth step, test image recognition with video frames to identify the letter name in vocabulary as exhibited in figure (11) for Time New Roman database and figure (18) for Arial database.



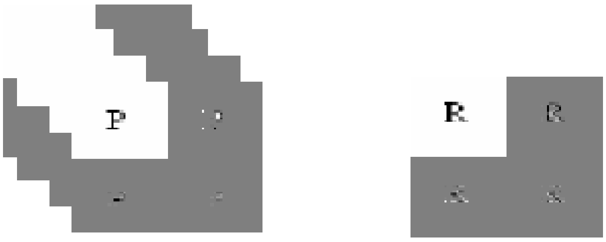
a) **b)**

Figure (11): The matching result of the proposed system between a) test image b) video frames



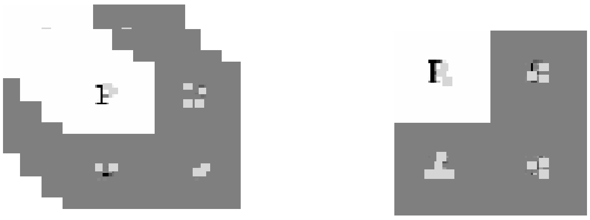
a) b) c)

Figure (12): Arial Black Database, a) Input Video of Images of Letters b) Extraction Frames from vidro c) Input Test Image of Letter from Arial Black Database



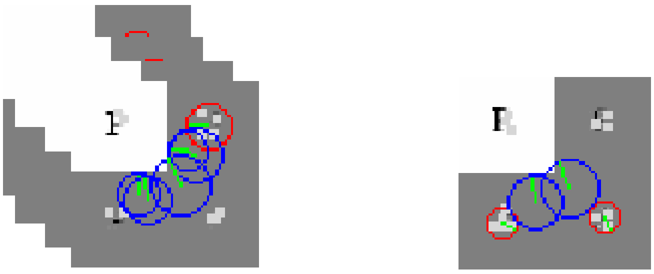
a) b)

Figure (13): Haar transform on a) video frame b) test image



a) b)

Figure (14): FAST corner detection on a) video frames b) test image



a) b)

Figure (15): SURF descriptor on a) video frames b) test image

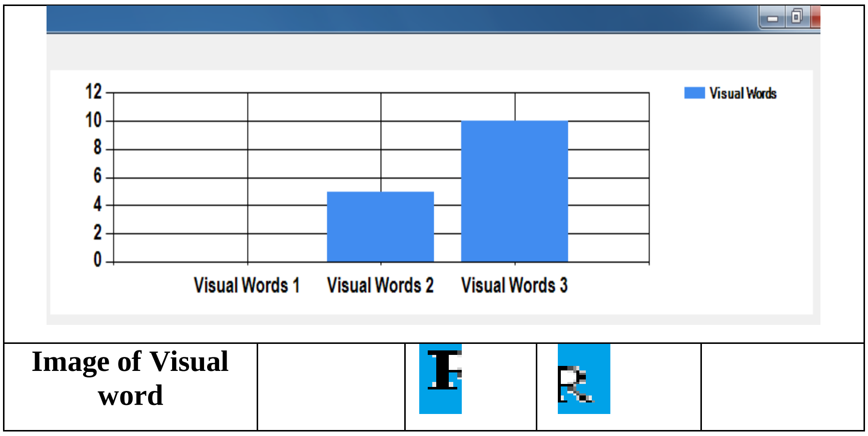


Figure (16):Histogram and Visual words for Letter R in Arial Black Database the frequency of each word can be considered as the number of features in the corresponding cluster.

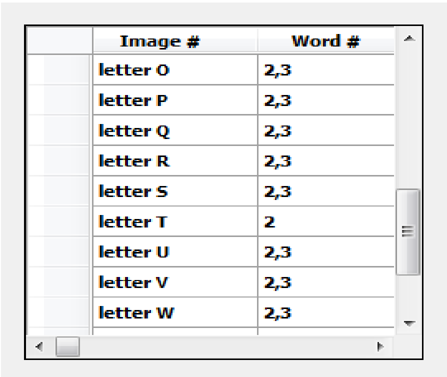


Image #	Word #
letter O	2,3
letter P	2,3
letter Q	2,3
letter R	2,3
letter S	2,3
letter T	2
letter U	2,3
letter V	2,3
letter W	2,3

Figure (17):Vocabulary of all Images of Letters for Arial Black Databse and Corsponding Visual Words Numbers

a) **b)**

Figure (18): The matching result of the proposed system between a) test image b) video frames

5. Proposed methodology analysis

In this section, it can analysis the proposed methodology based on consuming time for complete the recognition process and accuracy which has been depended on computing the precision performance that can be used for the correct recognition. The accuracy can be computed using equation (16).

$$precision = \frac{True\ positivies}{True\ positivies + False\ positivies} \qquad ...(16)$$

True Positives is the number of the images that are correctly retrieved from the image datasets, While False Positives is the number of images that are incorrectly recognized from the image database [12].

Table (1) shows the comparison result in term of time consuming and accuracy for recognition an object when Seven Moments and BOVW were employed. Accuracy recognition depends on number of correct matching between test image and database images. For example, letter K and R could be selected to display the outcome of comparison.

Method name	Object name	Time(sec) for recognition the letter	Accuracy recognition on letter
Seven Moments	Letter (K) in Time New Roman	2.991	83%
	Letter (R) in Arial Black	2.371	81%
BOVW	Letter (K) in Time New Roman	0.025	90%
	Letter (R) in Arial Black	0.011	93%

6. Conclusions

English letters recognition can be depended on multiple measurements. The suggested measurements method can be employed bag of visual word for recognition process. Bag of visual word is an efficient method for representation an image in the classification and recognition tasks. The suggested method deals with capital letters only and it was tested on three types of English letters font's databases (Time New Roman, Arial Black and Calibri). Experimental outcomes show that the suggested method is more efficient and fast for matching and recognizing a letter than recognition character using seven moments method. The time that BOVW could be taken for recognition a letter is less than the time that seven moments could be taken, the BOVW accuracy depends on number of correct character recognition. BOVW have optimal accuracy in the process of recognition of letters.

References

[1] Mandana Hamidi, Ali Borji, "**Invariance analysis of modified C2 features: case study hand written digit recognition**" , Machine Vision and Application, Vol.21, No.6, pp. 969-979, 2010.

[2] Nibaran Das, Ram Sarkar, Subhadip Basu, Mahantapas Kundu, Mita Nasipuri, Dipak Kumar Basu," **A genetic algorithm based region sampling for selection of local features in handwritten digit recognition application**", Applied

Soft Computing, Vol.12, No.5, pp.1592-1606, May 2012.

[3] Nusrat Jahan, Shahrul Nizam Yaakob and Phaklen Ehkan, **"Feature Extraction for Neural Network Pattern Recognition for Bloodstain Image Analysis ",** School of Computer and Communication Engineering, Malaysia, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 15) pp. 8583-8589 ,2016.

[4] Muralidharan R., Dr. Chandrasekar C.," **SCALE INVARIANT FEATURE EXTRACTION FOR IDENTIFYING AN OBJECT IN THE IMAGE USING MOMENT INVARIANTS** ",Journal of Engineering Research and Studies, E-ISSN 0976-7916, JERS/Vol.II/ Issue I/ pp. 99-103, January-March 2011.

[5] David Salomon, "**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, 2007.

[6] Wenjia Yang, Lihua Dou, Juan Zhang, Jinghua Lu, "**Automatic Moving Object Detection and Tracking in Video Sequences**", *SPIE Fifth International Symposium on Multispectral Image Processing and Pattern Recognition*, pp.676-712, 2007.

[7] Asif Masood, M. Sarfraz, "**Corner detection by sliding rectangles along planar curves**", *Computers & Graphics*, Vol. 31, pp.440-448, 2007.

[8] Jasmine J. Anitha and Deepa S.M., "**Tracking and Recognition of Objects using SURF Descriptor and Harris Corner Detection**", ȦNehru Institute of Engineering and Technology (Anna University) ,Coimbatore, India, Vol.4, No.2,pp.1-6, 2014.

[9] Bay H., Ess A., Tuytelaars T., and L., "**Speeded-up robust features (SURF)**", ,Comput. Vis. Image Understand., vol. 110, no. 3, pp. 346–359, 2008.

[10] Aswini C. and Chitra D., "**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, 2014.

[11] Shefalli and Balkrishan Jindal, "**Enhancing Content-based Image Retrieval using Moving K-Means Clustering Algorithm** ", International Journal of Computer Applications (0975 – 8887) Volume 102 – No.9, September 2014.

[12] Mohammed Elmogy and Hazem Elbakry," **Content-Based Image Retrieval using Local Features**

Descriptors and Bag-of-Visual Words ",Dept. of Information Technology, Faculty of Computers and Information, Mansoura University, Mansoura, Egypt , (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 6, No. 9,pp.212-219, 2015.

[13] Jiawei Han, and Micheline Kamber , "**Data mining, concepts and techniques**", Third Edition book, Morgan Kaufmann Publishers is an imprint of Elsevier.225 Wyman Street, Waltham, MA 02451, USA, 2012.