Engineering and Technology Journal
Journal homepage: engtechjournal.org

Word Retrieval based on FREAK Descriptor to Identify the Image of the English Letter that Corresponds to the First Letter of the Word

Ekhlas F. Naser [a*]

[a] Department of Computer Science, University Of Technology, Baghdad, Iraq, 110022@uotechnology.edu.iq

Submitted: 25/12/2019 Accepted: 21/04/2020 Published: 25/12/2020

KEYWORDS

FREAK, SUSAN, Harris, image retrieval, Manhattan distance.

ABSTRACT

For the reason of colossal technological developments, the requirement of image information methods became a significant issue. The aim of this research was to retrieve the word based on Fast Retina Key-points (FREAK) descriptor .The suggested system consists of four stages. In the first stage, the images of English letters are loaded. Points are detected via SUSAN in the second stage. FREAK used in the third stage and then a database was created containing 26 English letters. The image to be tested was entered and the points are extracted in the fourth stage and then Manhattan distance was used to calculate the distance between the value of the test image descriptors and all the values of the descriptors in a database. The experimental results show that the precision and the recall values were high for retrieval of the words when using SUSAN because it extracts a large number of interest points compared to the Harris method. For example, for the letter H was 104 with SUSAN while it was 42 for Harris, therefore; the precision for retrieval of the word Hour was 89% and recall was 93% when using SUSAN while precision was 77% and recall was 80% when using Harris.

How to cite this article: E. F. Naser, "Word Retrieval based on FREAK Descriptor to Identify the Image of the English Letter that Corresponds to the First Letter of the Word," Engineering and Technology Journal, Vol. 38, Part B, No. 03, pp. 150-160, 2020.
DOI: https://doi.org/10.30684/etj.v38i3B.1511

1. Introduction

The retrieval of an image is the study field that can be concerned with searching, scanning, and recapturing digital images of the expanded database. CBIR can be seen as a quick and dynamic advancement research region for the retrieval of an image domain. It is a mechanism for images retrieving by the similarity of a collection. Retrieval can be dependent on the characteristics that automatically elicitation from the images. Numerous systems of CBIR that depend on descriptors' features are established and advanced. A feature captures a certain visual property of an image. A descriptor can be encoded of an image in a way that can be allowed to compare and match with other images. The descriptors of image features in general could be either local or global. The global descriptors of the feature can be described as the visual content of the whole image, while local features can be described as an image's spot, for example, a small set of content image pixels. The superiority of a global descriptor elicitation is increasing feature extraction speed and calculating the similarity. Global features remain too solid for image representation. Especially, they can more sensitive to position and consequently may fail in identifying significant visual features. The approaches of local feature supply best retrieval efficiency and big discriminative force in resolving vision troubles than global features [1]. SUSAN is a special detector that offers considerably higher performance. It is very fast in detecting corners as key-points and consumes only a fraction of the time available during detection of a corner with more features, and on low power hardware [2]. FREAK was proposed as a fast retina's key-point descriptor [3]. The retina's organization imitated, utilizing a circular grid that receptive ranges are suggested of several sizes. The dissimilarity of intensity among receptive ranges pairs can be calculated and classified as a binary vector. Especially the focus of the receptive range is maximal at the center of the pattern, similar to the fovea of a retina. FREAK was estimated at matching function that shows high detection's performance for an object. The descriptors of BRISK [4] and DAISY [5] compare intensities pairs via utilizing a circular manner. Compare to descriptors' state art like SIFT, BRISK, or SURF, it outperformed them, whilst being faster and simpler. For a modern descriptor like CS-FREAK [6], a fundamental grid simplified via lessening the sensitive ranges' numbers and the nearness density will encode boosting the precision of matching. For a different job type, FREAK used for video action realization via extending to a descriptor that encoded motion this can be known as MoFREAK [7]. The descriptors biologically inspired have fundamentally been utilized for recognition of an object task [5]. For filtering, the difference of Gaussian (DoG) can simulate the performance of the retina that utilized for classification of the texture. In this action, it can be proposed a modern descriptors group of bio-inspired for scene job categorization [8]. The aim of this research was to retrieve the word based on Fast Retina Key-points (FREAK) descriptor to identify the image of the English letter that corresponds to the first letter of the word.

2. SUGGESTED METHODOLOGY

The suggested methodology for retrieval of the word had four major stages. The suggested system for retrieving a word consists of four stages.

1) In the first stage, 26 images of the English letters are loaded on the projection screen.
2) Corners are detected in the literal images as the interest points using the SUSAN corner detector in the second stage.
3) Corners are described in literal images using FREAK descriptor in the third stage and then a database was created containing 26 English letters, trained to find the number of descriptors for each letter and the words that correspond to each English letter.
4) In the fourth stage, the literal image to be tested was entered and the numbers of corners are extracted by using the SUSAN detector and then extracting the number of descriptors for this image by employing the FREAK descriptor. Using the Manhattan distance method to calculate the distance between the values of the test image descriptors and all the values of the descriptors in the database to identify the type of character in this image and then the word is retrieved from the database which begins with the letter that was identified. The figure should be referenced in Figure 1.

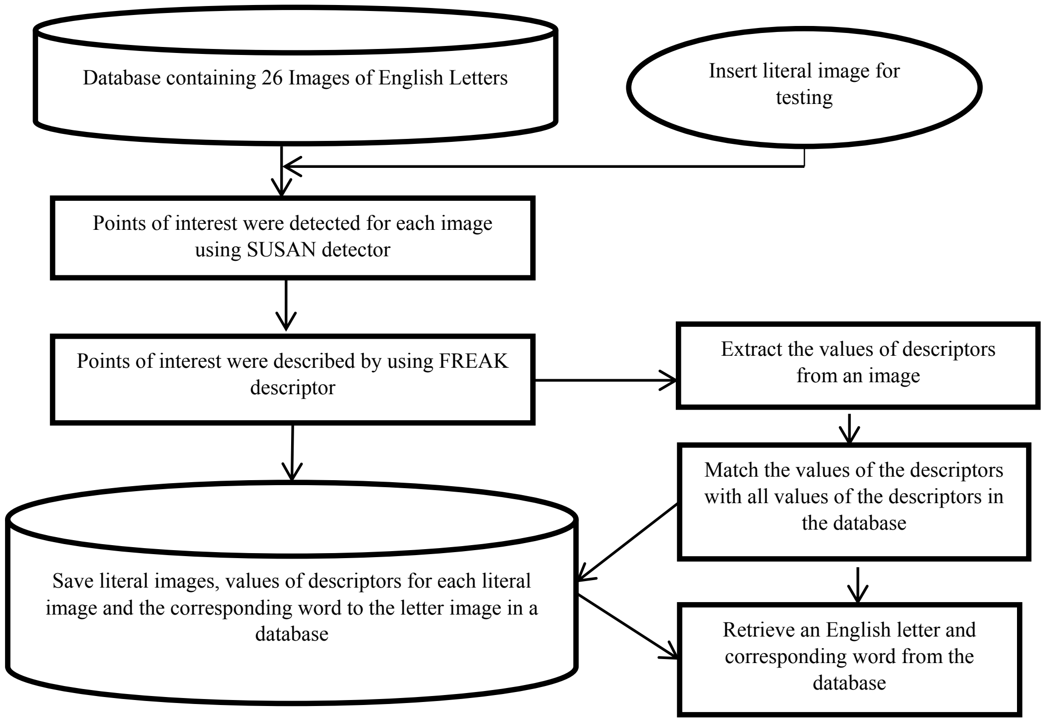


Figure 1: Block diagram of the proposed system for word retrieval

I. Feature Detection Using SUSAN

Smallest Univalue Segment Assimilating Nucleus (SUSAN) can cover the edge finding of an image, filtering of noise and corner finding. For detection of feature, SUSAN puts a circular mask on the pixel to be tested the nucleus. If every pixel's brightness in the mask can compare within the brightness of that mask's nucleus then a mask's zone can define that has a similar or same nucleus' brightness [9]. For detecting the image's corner via SUSAN, the effective function of similarity can be given in Eq. (1) [9].

$$c(r,r_0) = \exp\left\{-\left[\frac{I(r)-I(r_0)}{t}\right]^6\right\} \qquad (1)$$

Where *r*0 is nucleus's pixel coordinates; *r* is another points' pixel coordinates within the mask; *c*(*r*, *r*0) representing comparison function outcome; *I*(*r*) explain gray value of pixel point; *t* representing the gray difference's threshold that employed for determining the ability size of feature space and minimal contrast which can be detected via the algorithm of SUSAN [9]. And SUSAN region can be given in Eq. (2) [9].

$$n(r_0) = \sum_{r \in c(r_0)} c(r,r_0) \qquad (2)$$

II. Features Description uses FREAK

The FREAK was a binary descriptor that calculated relies on the brightness comparison produced tests for a quantity of sampling positions over a key-point [10]. FREAK can consist of the following steps: -

A. Sampling Pattern

The pattern's sampling adopted via the descriptor of FREAK is inspired biologically via a retinal manner in the eye. So, the points' sample constructs the basis for calculating the descriptor of FREAK is arranged in a sampling manner as illustrated in Fig. 2.N points sample existing about the given key-point has been smoothed via a Gaussian kernel before calculating the descriptor is, here, the kernel's size is diverse with respect to the sampling point's location for simulating a retina behavior of the human in identity to the visual system for human. FREAK descriptor sampling points, hence, illustrate the receptive fields' centers [10]. This can illustrate by Eq. (3).

$$P_i = P(x_i, y_i) = L_{ri}(x_i, y_i) \quad (3)$$

Where

$$L_{ri}(x, y) = I(x, y) * G_{ri}(x, y, \sigma_{ri}) \quad (4)$$

In Eq. (4) [10], I(x, y) the input pixels of an image, G_{ri} (x, y,σ_{ri}) explains the kernel of the Gaussian in the i-th receptive domain from i = 1 to N and L_{ri}(x, y) explains the input image's version which can be smoothed. The i-th point's sampling P_i conformable to the i-th receptive domain center (ri) and can be known within the predefined coordinates (x_i, y_i) of the model's sampling where i = 1 to N [10].

B. Descriptor Building

The descriptor of FREAK is constructing depend on the comparisons intensity between various pairs of sampling points' smoothed such as receptive fields' centers which can be defined by the following. Suppose a pair of sampling points Pa = (Pi, Pj), where i # j, i and j ∈ {1 to N} as illustrated in Figure 2. The approach of FREAK can define a binary encoded comparison of intensity, s(Pa) on the pair as illustrated in Eq.(5) [10].

$$S(P_a) = \begin{cases} 1, & \text{if } P_i > P_j \\ 0, & otherwise \end{cases} \quad (5)$$

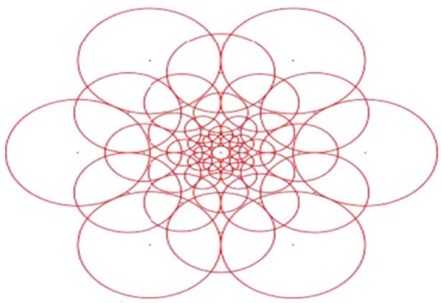


Figure 2: FREAK sampling pattern [3].

The offered comparison can form the basis to build the descriptor of FREAK F as illustrated in Eq. (6) [10].

$$F = \sum_{0 \le a < N} 2^a \; S(P_a) \quad (6)$$

The FREAK sampling manner can enable numerous pair wise comparisons which are leading to much great descriptor. Due to numerous pairs that may not be useful to describe image content, the writers implement an algorithm for training within the manner of sampling that can be illustrated in Fig. 3 for identifying beneficial pairs for descriptor building. The trained shape for the descriptor of FREAK can define 512 [3]**.** Pairs of the sampling that require to examine for computing the bit-string can be explained in Eq. (6) [3].Figure 3 illustrates the picked 512 sampling point pairs that classified into four clusters, each cluster has 128 pairs. Since pattern orientation over the global gradient, symmetric manner can be taken in these clusters.

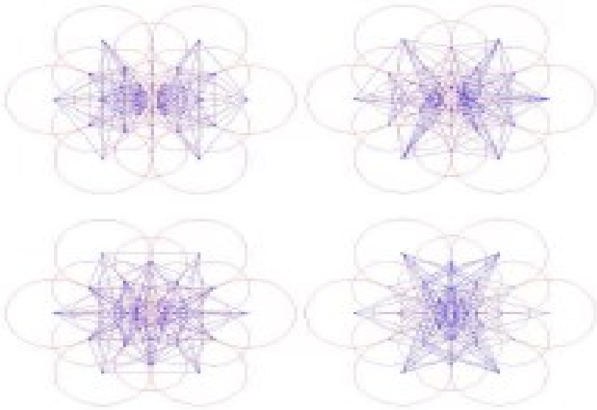


Figure 3: Explain four binary experiments clusters that compose FREAK. Peripheral receptive domains (top left), and central (bottom right) [3].

C. Orientation normalization

FREAK descriptor's orientation can be evaluated using 45 picked sampling pairs which are symmetrically ordered with regard to the sampling pattern's center as shown in Figure 4.

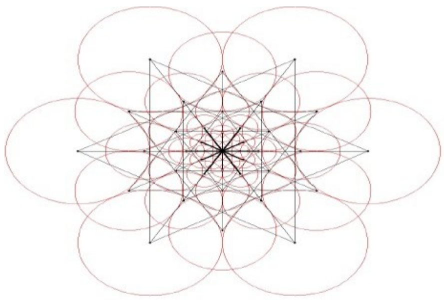


Figure 4: Pairs can be selected to calculate the key-point orientation [11].

Suppose G represents the group of whole the picked pairs, and suppose that local gradients have been calculated for all the picked points' sampling. The orientation o for a given key-point can be calculated in Eq. (7) [11].

$$O = \frac{1}{M} \sum_{\substack{P_i, P_j \in G \\ i \neq j}} (P_i - P_j) \frac{T(P_i) - T(P_j)}{\left\| T(p_i) - T(P_j) \right\|} \qquad (7)$$

Where M represents the pairs' number in T (P_i), and G which indicates a function that returns a two dimension vector of the receptive field's center spatial coordinates. For example, the coordinates' vector of the k-th sampling point T (P_k) = [x_k, y_k]. After elicitation the descriptors of an image, the Manhattan of those descriptors can calculate via Eq. (8). The Manhattan distance between two components can be calculated by the summation of the differences of their identical elements [11]. The distance formula among the points X= (X1, X2, etc.) within the points Y= (Y1, Y2, etc.) can calculate via Eq.(8) [12].

$$d = \sum_{i=0}^{n} |x_i - y_i| \qquad (8)$$

3. Suggested Algorithm for Word Retrieval

The suggested algorithm to retrieve a word can be explained as follows:

Input : Database containing 26 English letters images and corresponding words, test image of English letter **Output :** Retrieve the word in which the first letter is identical to English letter image
begin // For Training phase **Step1:- for** k=1 to 26

Step2:- select an image of the English letter **IMG[k]** from a database
Step 3:- points of interest are detecting for **(IMG[k])** by employing SUSAN corner detector and put the outcomes in **(IP_IMG[k])**
Step 4:- points of interest are describing using FREAK descriptor and put the results in **T[k]**.
Step 5:- Store **IMG[k]** , its values of descriptors **(T[k])** and the corresponding word in a database **(D)**.
End for // for k loop
Step 6: //For Testing and a word retrieval phase
6.1 Enter the tested image of English letter , apply step3 and step4 and then find the values of the descriptor **(VT)** for the tested image.
6.2 Compute the Manhattan distance between all values of descriptor for each letter in the database **(D)** and **(VT)** using Eq.(8) to determine the type of English letter.
6.3 Retrieve the corresponding word from the database **(D)** in which the first letter is identical to the English letter based on the results of step(6.2)
End.

4. Harris Detector for Features Extraction

Harris used for feature detection. This algorithm can be recognized by calculating the gradient of each pixel. The pixel can be taken as a corner if the absolute gradient values in two directions are both mighty. A Harris algorithm for detection of the corner can be calculated by Eq. (9) [13].

$$\mathrm{Reult} = \mathrm{detm}(G) - \mathrm{ktr}^2\,(G) \qquad (9)$$

$$G(a,b) = \begin{bmatrix} \mathrm{Img}_m^2\,(a,b) & \mathrm{Img}_{mn}(a,b) \\ \mathrm{Img}_{mn}(a,b) & \mathrm{Img}_m^2\,(a,b) \end{bmatrix} \qquad (10)$$

$$\mathrm{Img}_m^2\,(a,b) = A^2 \otimes \mathrm{GAUS}(a,b),$$
$$\mathrm{Img}_n^2\,(a,b) = B^2 \otimes \mathrm{GAUS}(a,b),$$
$$\mathrm{Img}_{mn}(a,b) = AB \otimes \mathrm{GAUS}(a,b),$$

$$\mathrm{GAUS}(a,b) = \frac{1}{2\pi} e^{\frac{-a^2+b^2}{2}} \qquad (11)$$

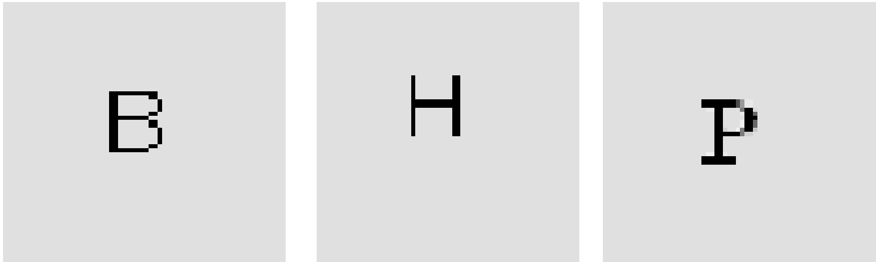
Where $\mathrm{Img}_m(a,b)$ and $\mathrm{Img}_n(a,b)$ are gray values partial derivatives in coordinate m and n at pixel (a,b), and $I_{mn}(a,b)$ represents the partial derivative jumbled of second-order; k represents practical value; GAUS(a,b) refers to a Gaussian function; A and B are the directional differentials at first-order, which can compute by convolving gray values and operators of difference in coordinate m and n. A function of Gaussian that used to minimize the noise effect due to the first-order of directional differentials which are critical to noise. Pick the pixel's point as a corner if R exceeds the values of the threshold [13].

5. Experimental Results

The suggested method outcomes are explained and discussed in this section. C# language is utilized to implement the suggested methodology. Twenty six kinds of English capital letters are utilized to evaluate the suggested method. Images in the database are JPEG black and white with a size of 200 × 200 pixels.

The proposed methodology involves multiple stages: -

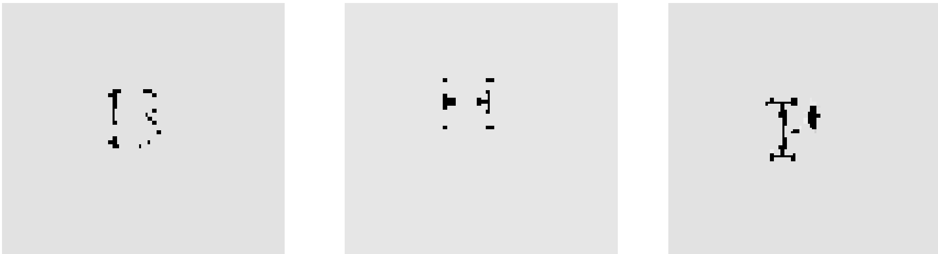
1) The English letters images are loaded on the project screen in the first stage of the proposed methodology as explained in Figure 5. For example, the samples of the images of English letters from the database, a) image of letter B, b) image of the letter H and c) image of letter P.



a) b) c)

Figure 5: For example, the samples of the images of English letters from a database, a) Image of Letter B, b) Image of Letter H and c) Image of the letter P

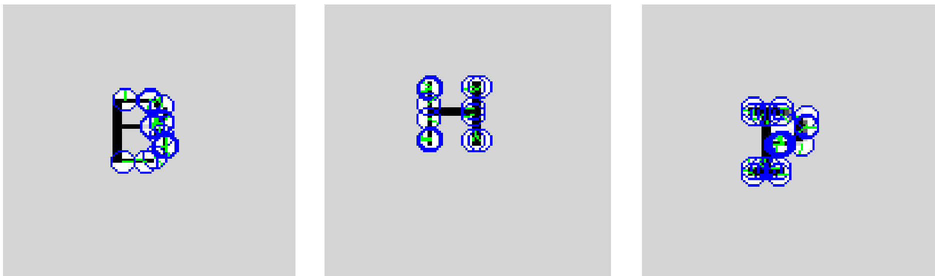
2) Corners are detected in literal images as points of interest using SUSAN corner detector in the second stage as illustrated in Figure 6 a) The detection of corners in the letter B, b) the detection of corners in the letter H and c) The detection of corners in the letter P.



a) b) c)

Figure 6: Detection of Corners stage using SUSAN, a) Letter B, Letter H and c) Letter P.

3) Corners are described in literal images using FREAK descriptor in the third stage as illustrated in Figure 7 a) The description of corners in the letter B, b) the description of corners in the letter H and c) The description of corners in the letter P.



a) b) c)

Figure 7: Description of Detected Corners stage using FREAK, a) Letter B, Letter H and c) Letter P.

A database was created containing 26 large English letters, trained to find the values of descriptors for each letter and the words that correspond to each English letter, as shown in Table (1).

TABLE I: A database is showing all images of English Letters with values of descriptors and corresponding word to each literal image.

Image Name	Sample values of Descriptors for each image	Corresponding Word
Letter A	13,10,5,…	Apple
Letter B	2,8,13,9,…	Baby
Letter C	12, 2,22,13,…	Car
Letter D	14,6,33,…	Day
Letter E	13,1,4,…	Edit
Letter F	9,2,2,..	File
Letter G	16,11,22,..	Green
Letter H	11,6,33,…	Hour

Letter I	9, 13,35,….	Image
Letter J	11,45,10,2,…	Japan
Letter K	7,12,8,40,…	Keys
Letter L	10,34,20,77	Lion
Letter M	18,1,15,…	Man
Letter N	8,33,27,…	Now
Letter O	12,1,17,…	Orange
Letter P	13, 6,33,…	Park
Letter Q	15,,8,41	Quit
Letter R	12,31,11,10	Red
Letter S	16,9,12,…	Stop
Letter T	7,27,11,…	Tree
Letter U	18,16,29,…	University
Letter V	11,30,38,…	Vertical
Letter W	14,18,22,…	Woman
Letter X	9,33,40 ,1,…	X-ray
Letter Y	9,28,19,…	Yellow
Letter Z	10,3,21,…	Zero

4) In the fourth stage, the literal image to be tested was entered , the values of the corners were extracted by using the SUSAN detector , extracting the values of the descriptors for the tested image by employing FREAK descriptor , using the Manhattan distance method to calculate the distance between the values of the test image descriptors and all the values of the descriptors in Table I to identify the type of letter in the tested image and then the word is retrieved from Table I, which begins with the letter that was identified.

Example 1:- for the test image of the letter B, the proposed system retrieves the corresponding word "Baby" from Table1 that starts with B as illustrated in Figure 8.

Figure 8: Retrieved the word "Baby" from the database which corresponded the image of the letter image B

Example2:-for the test image of letter H, the proposed system retrieves the corresponding word "Hour" from Table I that starts by B as illustrated in Figure 9.

Figure 9: Retrieved the word "Hour" from the database which corresponded the image of the letter image H

Example3:- for the test image of the letter H, the proposed system retrieves the corresponding word "Hour" from Table I that starts with B as illustrated in Figure 10.

Figure 10: Retrieved the word "Park" from the database which corresponded the image of the letter image P

6. THE PERFORMANCE OF PROPOSED METHODOLOGY

The performance of the retrieval system is measured by employing the measurements of recall and precision. The recall gauged the system's capability of recovering whole images that are pertinent whilst the precision gauges the system's capability for recovering only the pertinent images [13].

Eq. (9) [14] is applied to calculate the accuracy of the retrieval adequacy:

$$P = \frac{P_T}{P_T + P_F} \qquad (9)$$

Where P_T explains how many images are correctly acquired from the data-sets of an image while P_F explains how many images are incorrectly acquired from the datasets of an image.

The recall performance of the retrieval system can be calculated using Eq.(10) [13], the parameter of M explains how many pertinent images are not recovered [14].

$$R = \frac{P_T}{P_T + M} \qquad (10)$$

Table II illustrates the numbers of interest points for both SUSAN and Harris detectors methods.

TABLE II: illustrates the numbers of interest points for both SUSAN and Harris

Image Name	Numbers of Interests Points	
	SUSAN	***Harris***
Letter B	10	33
Letter H	104	42
Letter P	64	21

TABLE III: Shows the values of precision and recall for three sample words (Baby, Hour and Park) when using both SUSAN and Harris detectors methods.

Image Name	The precision with using the different detectors		Recall with using the different detector	
	SUSAN	***Harris***	***SUSAN***	***Harris***
Baby	93%	81%	91%	83%
Hour	89%	77%	93%	80%
Park	84%	79%	98%	86%

7. THE RETRIEVAL SPEED OF PROPOSED METHODOLOGY

It was that the retrieval speed was higher when using the SUSAN detector compare with the Harris detector. The retrieval speed was computed via computes the time spent to detect the corresponding word to the English letter image and retrieve it from the database. Table IV illustrates the retrieval speed for the words when using both SUSAN and Harris detectors.

TABLE IV:illustrates the retrieval speed for both SUSAN and Harris detectors methods

Image Name	Retrieval Speed for SUSAN	Retrieval Speed for Harris
Letter B	0.0141	0.1349
Letter H	0.0079	0.0228
Letter P	0.0112	0.0717

8. Conclusions

The aim of this research was to retrieve the word based on the FREAK descriptor to identify the image of the English letter that corresponds to the first letter of the word. The suggested methodology employed a SUSAN detector for elicitation points of interest from an image of the English letter which is faster. This algorithm could realize by a circular mask holding the center pixel which would be known as the nucleus. SUSAN carries out fully until in the existence of noise. No derivatives for an image are employing and non-linear response can give powerful rejection of noise. The employing of controlling parameters is very simpler and minimal arbitrary. The FREAK descriptor is a binary that can be calculated via the comparison of brightness experiences outcomes in a sampling situation number about the key-point. The descriptor of FREAK is a powerful and more accurate descriptor. The experimental results show that the values of the precision and the recall were high for retrieval of the words when using the SUSAN method as illustrated in Table III because it extracts a large number of interest points compared to the Harris method as illustrated in Table II. Also, the retrieval speed was high for the SUSAN method because it takes less time as illustrated in Table IV.

9. Future Works

The following future works can be added:-

1) It can use a suggested methodology to retrieve an object from a video frame.
2) It can use a suggested methodology to deal with small English letters, digits and special characters
3) It can use the FAST detector to increase the retrieval speed.
4) It can use SURF or SIFT descriptors to improve the retrieval system.
5) Multi-level neural networks and support vector machines can be used in the retrieval system.

References

[1] P. Srivastava and A. Khare, "Content-based Image Retrieval using Scale Invariant Feature Transform and moments," IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electrical Engineering (UPCON), 162-166, 2016.

[2] S. Mahammadi and N. Mahesh, "Image Mosaic Using FAST Corner Detection, " International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE), Vol. 1, No. 6, pp. 1-6, December, 2012.

[3] C.H. Gomez, K. Medathati. P.Kornprobst, V. Murino, and D. Sona, "Improving FREAK Descriptor for Image Classification," in International Conference on Computer Vision, Copenhagen, Denmark, pp.14-23,July 6-9, 2015.

[4] A. Alahi, R. Ortiz and P. Vandergheynst, "FREAK: fast retina keypoint", IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, pp. 510–517, June, 2012.

[5] O. Guclu and B. Ahmet, "A Comparison of Feature Detectors and Descriptors in RGB-D SLAM Methods," Image Analysis and Recognition, Computer Society Conference on Computer Vision and Pattern Recognition, Ankara/Turkey, pp. 297-305, July, 2015.

[6] E. Tola ,V. Lpetit and P. Fau, "Daisy: An efficient dense descriptor applied to wide-baseline stero," IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 32, No. 5, pp.815-830, May, 2009.

[7] J.,Wang, X.,Wang, X. Yang and A. Zhao, "CS-FREAK: an improved binary descriptor,"in Chinese Conference on Image and Graphics Technologies, Beijing, China, pp.129-136, June, 2014.

[8] C. Whiten, R. Laganiere, and G.-A. Bilodeau, "Efficient action recognition with MoFREAK," International Conference on Computer and Robot Vision (CRV), Ottawa, Canada, pp. 319–325, May , 2013.

[9] M. Ami, "A Survey on Object Based Image Retrieval using Local and Global Features," International Journal Of Engineering And Computer Science, ISSN:2319-7242, Vol. 3, No. 10, pp. 8643-8646, December, 2014.

[10] N. Alyuz, B. Gokberk, and L. Akarun, “3-d face recognition under occlusion using masked projection,” IEEE Transactions on Information Forensics and Security, vol. 8, no. 5, pp. 789–802, May, 2013

[11] D. -D. Nguyen, A. El Ouardi, E. Aldea, and S. Bouaziz, "HOOFR: An Enhanced Bio-Inspired Feature Extractor, " 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, pp.2977-2982, December, 2016.

[12] A.A. Karim and E.F. Nasser, "Image Retrieval from Video Streams Databases using Similarity of Clustering Histogram, " Al-Mansour Journal, No.29, pp.1-22, December, 2018.

[13] A.A. Karim and E. F. Nasser, "Improvement of Corner Detection Algorithms (Harris, FAST and SUSAN) Based on Reduction of Features Space and Complexity Time," Engineering & Technology Journal, Vol. 35, No. 2,Part B., pp.112-118, 2017.

[14] M. Alkhawlani, M.Elmogy, and H. Elbakry, "Content-based Image Retrieval using local Features descriptors and Bag-of-Visual Words," (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 6, No. 9, pp.212-219, 2015.