ISSN 2277-2685
IJESR/July. 2023/ Vol-13/Issue-3/1-7
Ojerinde Oluwaseun Adeniyi *et. al.*, / International Journal of Engineering & Science Research

Using a Hierarchical Keypoint Model for Object Recognition

Ojerinde Oluwaseun Adeniyi [1], Saliu Adam Muhammed [1], MohammedAbubakar Saddiq [2], Ekundayo Ayobami [1]

[1] Department of Computer Science, Federal University of Technology, Minnna, Niger State, P.M.B.65, Nigeria
[2] Department of Electrical/Electronic Engineering, Federal University of Technology, Minnna, Niger State, P.M.B.65, Nigeria
* Corresponding author: Abdulmalik Danlami Mohammed, +2349069148660, Corresponding author ORCID:

ABSTRACT:
It is important to detect keypoints and calculate their descriptions before engaging in local keypoints matching between a pair of images for object identification. Accurately describing landmarks is crucial for many vision-based applications, including 3D reconstruction and camera calibration, structure from motion, picture stitching, image retrieval, and stereo pictures. To address this issue, this paper presents (1) UFAHB, a robust keypoints descriptor using a cascade of Upright FAST -Harris Filter and Binary Robust Independent Elementary Feature descriptor, and (2) a comprehensive performance evaluation of UFAHB descriptor and other state-of-the-art descriptors using a dataset extracted from images captured under different photometric and geometric transformations (scale change, image rotation, and illumination variation). The gathered experimental results show that the integration of the UFAH and BRIEF descriptors is fast in execution time and robust against variations in illumination.

KEYWORDS: Concepts include image landmarks, feature detection, feature characterization, Datasets of images, image retrieval, and image recognition

1. Introduction

Many computer vision applications revolve on describing visual keypoints to facilitate object recognition and object tracking. Applications in computer vision such as pose prediction, 3D reconstruction and camera calibration, structure from motion, picture stitching, image retrieval, and stereo images have benefited from keypoints description. A descriptor's duty is to characterize the surrounding intensity distribution of pixels around a place of interest. Therefore, a robust and distinguishable descriptor may improve the efficiency of various vision-based applications, including object detection, picture retrieval, and 3D reconstruction. Because of their limited processing power, creating computer vision-based applications for mobile phones has always been a difficult undertaking. However, this has sparked a new line of inquiry into how to implement computer vision and image processing on low-memory devices, such as smartphones. The end result of this line of inquiry is a variety of techniques for extracting feature descriptions from an image's architecture.

In order to facilitate the development and deployment of various computer vision-based applications on low processing devices (such as smartphones), new techniques have been developed to decompose the whole picture structure into a subset of descriptors.Several recent publications have suggested ways to enhance the calculation of picture keypoints and their description in an effort to make the process faster and more robust against common image changes including zooming, panning, rotating, changing the lighting, and blurring. The orientated FAST and Rotated BRIEF suggested in [1] are two examples of such efforts. Keypoints for the Fast Retina Display [2] and the Binary Robust Invariant Scale [3] are two such examples.

We augmented the Upright FAST-Harris Filter introduced in [4] with the Binary Robust Independent Elementary

Feature descriptor presented in [5] to obtain robust description of keypoints with low computational cost. In the first step of this expansion's cascading technique, keypoints are identified using the Upright

2. The image is first processed using a FAST-Harris filter, and then a descriptor based on the keypoints in its immediate vicinity is computed using the method of Binary Robust Independent Elementary Features. Finally, we use a dataset culled from photographs taken in a variety of lighting and camera settings to evaluate UFAHB's performance in comparison to other state-of-the-art descriptors.

3. **Related work**

4. Many different methods for detecting keypoints and classifying them have been presented. For instance, ORB (Oriented quick and Rotated Brief) is a local keypoints detector and descriptor presented in [1] that is both quick and resilient. The Harris edge filter is used to rank the FAST keypoints that were previously detected using the FAST keypoints detector. Keypoints are defined using a rotated Binary Robust Independent Elementary keypoint, and their orientation is calculated using the centroid of intensity. A Robust Invariant Scale for Binomial Data In [2], the use of a key point system called BRISK is suggested. Keypoints are localized in the scale and image planes with the help of the modified FAST, making this detector a scale-invariant feature detector. By comparing 8 neighboring scores inside the same octave and 9 scores in each of the immediate neighboring layers above and below, the strongest keypoints in octaves are identified in [2]. BRISK's description of keypoints is rotation-invariant since it computes a weighted Gaussian average over a pattern of points chosen around the points of interest. Despite this, BRISK is often understood to be a 512-bit binary descriptor. A method called Fast Retina keypoints (FREAK) is presented in [3]. It's a step up from BRISK's point-to-point sampling and binary comparison tests. FREAK's design was inspired by the retina pattern of the human eye. FREAK, in contrast to BRISK, utilizes a cascade mechanism to compare pairs of points and only 128 bits of data rather than the 512 bits produced by BRISK. One of the first binary descriptors, the Binary Robust Independent Elementary Feature (BRIEF) was introduced in [5]. By comparing the intensity patterns of a smoothed picture, the descriptor(BRIEF) generates a bit vector. BRIEF can still work even if the picture is slightly rotated since it does not assess keypoint orientation. When compared to BRISK and FREAK, BRIEF is both computationally efficient and speedier. In [6], a scale- and rotation-invariant feature detector and descriptor called the Scale-Invariant Feature Transform (SIFT) is introduced. SIFT may be used for a variety of purposes, including 3D reconstruction, image tracking, stereo imaging, and object detection. Using the SIFT technique, a group of picture keypoints may be generated using a stage-filtering strategy that identifies scale-space extrema, locates keypoints, and describes keypoints. SIFT divides the gradient location into 8 equal sub-regions using a 4x4 grid.

5. directions reserved for the gradient orientations. SIFT descriptors have a 128-dimensionality. The accelerated robust feature, or SURF, is a SIFT-inspired keypoints descriptor. In [7], we present SURF, a Hessian-matrix-based keypoints detector and descriptor. Object tracking, 3D reconstruction, camera calibration, and picture registration are just some of its many uses. When compared to other detectors like SIFT, SURF is more computationally efficient while maintaining high levels of repeatability, resilience, and uniqueness. Finding keypoints in picture region where the matrix of second order derivatives has two big eigenvalues is the basis for the Harris detector or Harris edge filter developed in [8]. In [9], a detector called FAST (Features from Accelerated Segment Test) is suggested. FAST is able to do its job by comparing a pixel's intensity value to those of its surrounding, concentric pixels. As a local descriptor, the Local Binary Pattern learns the intensity value of an image in a compact region around a center pixel. Each pixel in the immediate area is represented by a single bit in the local binary pattern. Instead of using these binary patterns in their original form, quantization and transformation into a histogram is often performed first. The paper in [10] is mostly responsible for popularizing LBP.

6. **Methodology**

The components of our Upright FAST Harris and BRIEF method are shown in Figure 1 as a block diagram. U-FAH is used to detect landmarks in input images, and BRIEF creates descriptions of the surrounding area for each one.
The block diagram of our cascaded Upright FASTHarris and BRIEF method is shown in Figure 1. In the diagram, keypoints from input images are detected usingthe U-FAH method and for every keypoint extracted, its descriptions around the neighborhood are computed using the BRIEF method

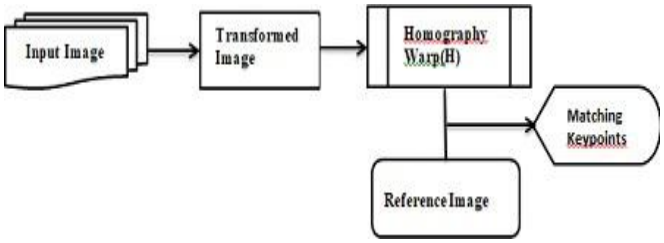

Figure 1: Schematic Diagram of keypoints matching between pair of images

Figure 1 shows how U=FAHB is used to determine keypoints and their respective descriptions for both the converted picture and the reference image. Each altered picture is then aligned with the reference image through a Homography warp computation. Section 3.2 of this article provides a comprehensive overview of our suggested technique while maintaining the document's cleanliness and brevity.

8.1**. Dataset**

In this study, we collect a dataset from genuine photographs depicting various scenarios (see Figure 2) and use it to

U-FAHB's performance is measured against other state-of-the-art descriptors using the recall and 1-precision criteria with respect to matching descriptor.

where p(x) is the intensity of the pixel within the smoothedpatch p at point x. Here the outputs of the binary test are concatenated into a vector of n bits that is referred to as thedescriptor. This vector of n bit string can be defined as:

Upright FAST- Harris Filter with BRIEF

6.1.1. Upright FAST-Harris Filter (UFAH) has the advantage of being combinable with other descriptors according to the modular method provided for its construction in [4]. Given the limited processing power of a mobile device, however, it is essential to couple UFAH with a computationally efficient descriptor. In light of its computing efficiency and speed, the Binary Robust Independent Elementary Feature descriptor described in [5] is taken into account in this research. Refer to [4] for a comprehensive analysis of UFAH. Due to space constraints, we will only address our cascade method for the Binary Robust Independent Elementary Feature in this work.*Binary* ***Robust Independent Elementary Feature***

The BRIEF descriptor, which stands for Binary Robust Independent Elementary Feature, is a lightweight and straightforward descriptor of an image patch that relies on a binary intensity test. Here is how we characterize the intensity test of a specified patch p of a smoothed image:

6.2. In order to get the best results, we used a Gaussian distribution centered on the picture patch, rather than one of the other test distributions suggested in [5]. Based on our experiments, we found that the 512-length BRIEF descriptor outperformed the 256-length ORB descriptor. The picture patch undergoes a smoothing operation before undergoing the binary test process to lessen the noise associated with each individual pixel. In this case, we smooth the picture using an integral image, like the one in [8].

6.3. Result of Matching Pair of Images using U-FAHBmethod

Extracting the dataset from photographs acquired while zooming the camera anywhere from 0 to a scale ratio of 2.5 yields the results shown in Figure 2a. The homography warp H is calculated for every image transformation so that the final picture matches up perfectly with the original.

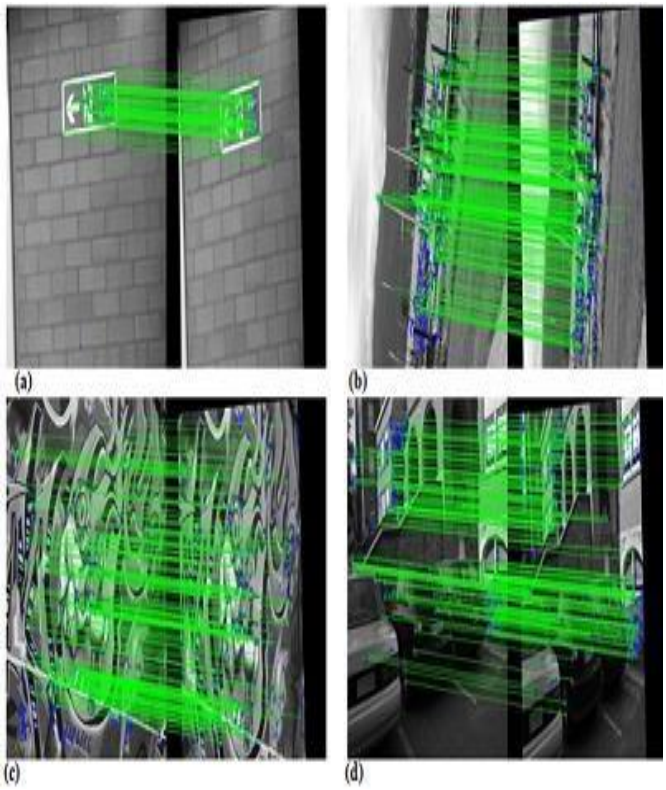

Figure 3: The result of matching descriptors from pair of image under (a) Scale change (b) Image rotation (c) View change and (d) Varying illumination

Figure 3(a) shows the result of matching the first imagewith the second image under scale change.

To get the dataset from the rotated photos (Figure 2b), we rotate the camera's optical axis. The average rotation angle in this experiment was 30 degrees, and that's what's stated for the angle at which the second picture was rotated relative to the reference image. Nearest neighbor descriptors are calculated for each reference image descriptor and applied to the second picture.

$$r(p; x, y) = \begin{cases} 1 & if\ p(x) < p(y) \\ 0 & otherwise \end{cases}$$

then cross checks their consistency in both directions to reduce false matches. The result of matching the firstimage with the second image observed under rotation is shown in Figure 3(b). The dataset from images captured under view change(see Figure 2c) is extracted by changing camera position from a front-parallel view to more foreshortening. The view point angle of the second image from the reference image is given as 20 degrees. For each descriptor in the

1-Precision on the other hand corresponds to the number of false matches in relation to the sum total of positive matches and false matches, which can beexpressed as:

reference image, its nearest neighbor descriptor in thesecond image is computed and then cross checks symmetry to lessen the number of mismatches. The outcome of this view-shift-induced image-matching is shown in Figure 3(c). By adjusting the camera's aperture, we may pull the dataset from photographs taken in different lighting conditions (Figure 2d). Descriptors from the first picture are matched with their closest neighbor descriptors from the second image obtained under different lighting conditions, and the result is shown in Figure 3(d)..

7. Performance Evaluation of Keypoint Descriptors

The joint performance of the Upright FAST-HarrisFilter and the BRIEF descriptor is compared with the stateof the art descriptors using the recall and 1-precisionmetrics. Given a pair of images, feature points and their description are computed for the reference images as wellas for the transformed images using the appropriate methods. For each keypoint in the reference image, a nearest neighbor in the transformed image is located followed by a consistency check in both directions to reduce the number of false matches. Subsequently, the number of positive matches and the false matches are counted and the results are plotted using the recall vs 1- Precision curve.

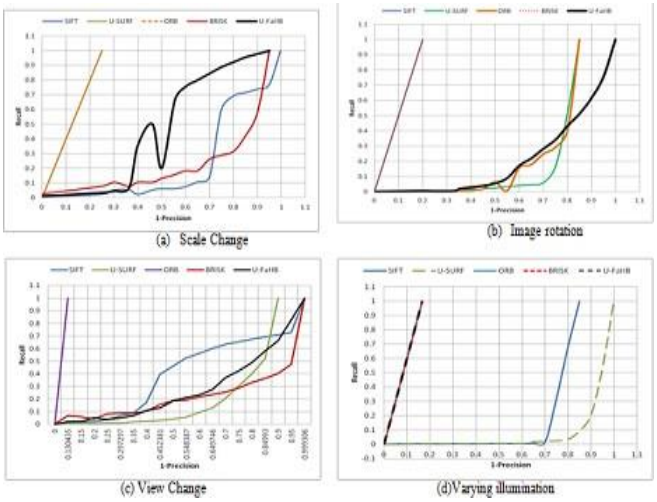

Figure 4: The Precision-recall curves for SIFT, U-SURF, ORB, BRISK and UFAHB descriptors using different dataset extracted from images observed under (a) Scale change (b) Image rotation (c) View change (d) Varying illumination

While recall in this context corresponds to the numberof positively matched regions in relation to the number of corresponding regions obtained for a pair of image and istherefore expressed as:

The recall vs 1-Precision curve for dataset from image observed under scale change is shown Figure 4(a). As canbe observed from the graph, ORB and U-SURF have better performance on scale changes compare to SIFT, BRISK andUFAHB. In Figure 4(b), the dataset extracted from image under rotation for all descriptors is plotted using a recall vs 1- Precision curve. The result shows both SIFT and BRISK tohave similar performance and better score on image rotation than the remaining descriptors. Figure 4(c) shows how well each descriptor has performed on a dataset extracted from pair of images observed under view point change. As observed from the graph in Figure 4(c), ORB descriptor has the highest score in terms of the number of correctly matched descriptors. The recall and 1- precision curve obtained for all descriptors using a dataset from image observed under varying illumination is shown in Figure 4(d). Here, ORB,BRISK and UFAHB have the best performance for a small number of keypoints regions detected in images of decreasing illumination.

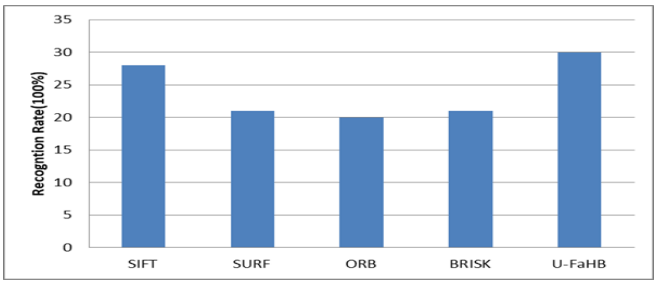

Figure 5: The recognition rate as obtained for all the algorithms (SIFT,SURF, ORB, BRISK and U-FaHB)

In Figure 5, the rate of recognition as observed by the different descriptors is shown. As can be deduced from graph, the Upright-FAST Harris combined with Binary Robust Independent Elementary Feature descriptor recorded the highest recognition rate as compare to the other descriptors.

Table 1: Description time in millisecond across all dataset

Descriptor	Average description time(ms)
SIFT	2.94643
U-SURF	2.52788
ORB	0.199153
BRISK	0.040877
UFAHB	0.086515

$$recall = \frac{number\ of\ positive\ matches}{number\ of\ corresponding\ regions}$$

Table 1 and Figure 6 show the average time it takes foreach descriptor to describe a feature region. From Table 1it can be observed that BRISK has the fastest description time. This is followed by UFAHB, ORB, U-SURF and SIFTin that order.

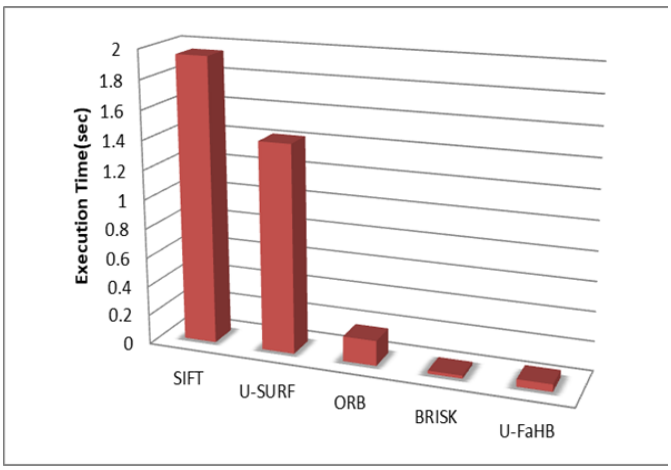

Figure 6: The Average execution time recorded for the different descriptors (SIFT, U-SURF, ORB, BRISK, U-FaHB)

8. Discussion and Conclusion

An independent dataset of photos that have been scaled, rotated, shifted in perspective, and had their lighting altered provides a fair evaluation of all descriptors. The accuracy and recall curve is used to evaluate their effectiveness. Using the same experimental setup, the description time, which is crucial for real-time performance, is recorded for each descriptor.

Recall and 1-precision curves are analyzed for a set of pictures with a scale variation between 0 and 2.5. As can be seen in Figure 3a, ORB fared best in this test, followed by BRISK and UFAHB. However, this demonstrates that bit-pattern-based descriptors function very well when an image's size fluctuates. To illustrate the rotation invariance of several descriptors, we compare the recall and 1-precision curves produced for a set of pictures that have been swapped. The average picture rotation in the dataset is between 0 and 30 degrees. The best performing characterizations were SIFT and BRISK, followed by UFAHB, ORB, and U-SURF. This proves that SIFT and BRISK are equally effective when rotated. We compared the recall and 1-precision curve for perspective changes between two pictures with viewpoint angles ranging from 0 to 20 degrees. Figure 3c shows that, although being able to accurately match a few keypoints correspondences, the ORB descriptor lacks distinctiveness. Under perspective shifts, however, SIFT stands out as the most distinguishable detector, followed closely by UFAHB and BRISK. On a set of photographs where one was progressively darker than the other, we tested how resistant various descriptors were to this variation in lighting. Test setup shown in figure 3d, UFAHB, ORB and BRISK outperformed the otherdescriptors showing the robustness of bit pattern to illumination changes.

The execution time of each descriptor was studied (table 1) to assess the potential of each descriptor for real-time performance. The acquired findings demonstrate that BRISK is the quickest descriptor, with UFAHB coming as a close second. UFAHB's description time is longer than BRISK's, but its accurate identification is a result of using all sample points in the matching process. Due to their computational complexity, SIFT and U-SURF are the slowest descriptors.

In conclusion, binary-pattern-based descriptors may be executed more quickly in a variety of imaging settings. Therefore, for devices with limited processing capability, the combination of UFAH and BRIEF has great promise.

References

One such paper is "An efficient alternative to SIFT or SURF," by E. Rublee et al., published in Proceedings of the 2011 International Conference on Computer Vision, pages 2564–2571; 2011; doi:10.1109/iccv.2011.6126544.

Based on the work of S. Leutenegger et al., "BRISK: Binary robust invariant scalable keypoints," IEEE ICCV, pp. 2548-2555, 2011,doi: 10.1109/iccv.2011.6126542.

A. Alah et al., "Fast retina keypoint," IEEE Conference on Computer Vision and Pattern Recognition, pp. 510-517, 2012, doi: 10.1109/cvpr.2012.6247715 [3].

[4] "Upright FAST- Harris Filter," i-manager's Journal on Image Processing, 5(3),14- 20,2018, doi: 10.26634/jip.5.3.15689; A. D. Mohammed, A. M. Saliu, I. M. Kolo, A. V. Ndako, S. M. Abdulhamid, A. B. Hassan, and A. S. Mohammed.

European Conference on Computer Vision, 2010, M. Calonder, V. Lepetit, C. Strecha, and P. Fua., "BRIEF:Binary Robust independent elementary features," doi:10.1007/978-3-642-15561-1_56.

In 2004, the International Journal of Computer Vision published an article by David G. Lowe titled "Distinctive Image Features from Scale-Invariant

Keypoints," which may be accessed online at doi:10.1023/b:visi.0000029664.99615.94.
European Conference on Computer Vision, pp. 404-417, 2006; H. Bay et al., "Surf: Speeded up robust features," doi:10.1007/11744023_32.
Alvey vision conference, pp. 147-151.,1988, Harris, M. Stephens, "A Combined Corner and Edge Detector," doi:10.5244/c.2.23.
"Faster and better: A machine learning approach to corner detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 1, pages 105-119, 2010, doi: 10.1109/tpami.2008.275; Rosten E., Porter R., and Drummond T.
[10] T. Ojola, T. Maenpana, D.Harwood, "Performance evaluation of texture measures with classification based on kullback discrimination of distribuitions," Proceedings of the 12th IAPR International Conference on Computer Vision and Image Processing, Vo 1, pp. 701-706.,1994, doi: 10.1109/icv.1994.576366.
IEEE Transaction on Pattern Analysis and Machine Intelligence, volume 27, issue 10, pages 1615–1630, 2005; K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," 2005; doi: 10.1109/tpami.2005.188.