Received: 30 January 2022, Revised: 06 March 2022, Accepted: 11 March 2022, Online: 28 March 2022

DOI: https://dx.doi.org/10.55708/js0103017

Cascaded Keypoint Detection and Description for Object Recognition

Abdulmalik Danlami Mohammed [*,1], Ojerinde Oluwaseun Adeniyi [1], Saliu Adam Muhammed [1], Mohammed Abubakar 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, drmalik@futminna.edu.ng
Corresponding author ORCID: https://orcid.org/0000-0002-0217-7411

ABSTRACT: Keypoints detection and the computation of their descriptions are two critical steps required in performing local keypoints matching between pair of images for object recognition. The description of keypoints is crucial in many vision based applications including 3D reconstruction and camera calibration, structure from motion, image stitching, image retrieval and stereo images. This paper therefore, presents (1) a robust keypoints descriptor using a cascade of Upright FAST -Harris Filter and Binary Robust Independent Elementary Feature descriptor referred to as UFAHB and (2) a comprehensive performance evaluation of UFAHB descriptor and other state of the art descriptors using dataset extracted from images captured under different photometric and geometric transformations (scale change, image rotation and illumination variation). The experimental results obtained show that the integration of UFAH and BRIEF descriptor is robust and invariant to varying illumination and exhibited one of the fastest execution time under different imaging conditions.

KEYWORDS: Image keypoints, Feature detectors, Feature descriptors, Image retrieval, Image recognition, Image dataset

1. Introduction

The description of image keypoints is at the core of many computer vision applications since it simplifies the task of object recognition and object tracking. Some of the computer vision applications where keypoints description has been found useful include pose estimation, 3D reconstruction and camera calibration, structure from motion, image stitching, image retrieval and stereo images. The job of a descriptor is to describe the intensity distribution of neighbouring pixels around an interest points. As a result, the performance of many vision- based applications such as object recognition, image retrieval and 3D reconstruction can be enhanced with a stable and distinctive descriptor. The development of computer vision based application on mobile phones in time past, has been a challenging task due their low processing power. This has, however, led to a new research direction in image processing and computer vision on low memory devices such as the smart phones. The outcome of such research direction is the development of different methods for describing interest points from an image structure. These new methods of decomposing the whole image structure into a subset of descriptors reduce computational burden that would otherwise make the process of development and deployment of many computer vision based application cumbersome on a low processing devices (e.g smartphones).In recent time, few works have been proposed to improve the computation of image keypoints and their description with the aim of achieving real time performance and invariance to image transformations such as scale change, image rotation, illumination variation, and image blurring. Some of these works include the oriented FAST and Rotated BRIEF proposed in [1]. Binary Robust Invariant Scale Keypoint presented in [2] and the Fast Retina Keypoints proposed in [3].

In order to achieve robust description of keypoints with minimal computation, we expanded the Upright FAST-Harris Filter proposed in [4] to include Binary Robust Independent Elementary Feature descriptor proposed in [5]. The expansion is a cascaded approach in which keypoints are first detected using the Upright

FAST-Harris filter follow by the computation keypoints descriptor around its' neigbourhood based on Binary Robust Independent Elementary Feature descriptor. Finally, we compare the performance of UFAHB against other state of the art descriptors using dataset extracted from images captured under different imaging conditions.

2. Related work

A wide range of keypoints detectors and their descriptors are proposed in the literature. For example, the Oriented Fast and Rotated Brief also referred to as ORB is a fast and robust local keypoints detector and descriptor proposed in [1]. The algorithm uses the FAST keypoints detector to detect corners in an image and subsequently employs the Harris edge filter to order the FAST keypoints. The orientation of the detected keypoints is computed using the intensity centroid while the keypoints are described using a rotated Binary Robust Independent Elementary keypoint. The Binary Robust Invariant Scale Key points referred to as BRISK is proposed in [2]. It is a scale invariant feature detector in which keypoints are localized in both scale and image plane using the modified version of FAST. In [2], the strongest keypoints are found in octaves by comparing 8 neighbouring scores in the same octave and 9 scores in each of the immediate neighbouring layers above and below. In BRISK, keypoints are described by computing a weighted Gaussian average over a selected pattern of points around the points of interest and thus achieves invariance to rotation. BRISK is however regarded as a 512 bit binary descriptor. Fast Retina keypoints (FREAK) is proposed in [3]. It is an improvement over the sampling pattern and binary comparison test approach between points of BRISK. The pattern of FREAK is motivated by the retina pattern of the eye. However, in contrast to BRISK, FREAK employs a cascade approach for comparing pairs of points and uses 128 bits as against the 512 bits obtained in BRISK to enhance the matching process. The Binary Robust Independent Elementary Feature (BRIEF) is one of the first binary descriptors and which was presented in [5]. The descriptor(BRIEF) works by building a bit vector from the result of comparing the intensity patterns of a smoothed image. Even though BRIEF does not measure keypoint orientation, it still can tolerate a small image rotation. BRIEF is computationally efficient and faster in comparison to BRISK and FREAK. The Scale-Invariant Feature Transform (SIFT) is a scale and rotation invariant feature detector and descriptor that is proposed in [6]. SIFT has a wide area of applications in object recognition, image stitching, stereo image, image tracking and 3D reconstruction. The generation of a set of image keypoints using SIFT method involves a stage-filtering approach that includes detection of scale-space extrema, key points localization and key points description. SIFT uses a 4 x 4 sub-region to divide the gradient location and 8 different orientations set aside for the gradient angles. The dimension of SIFT descriptor is 128. Speeded Up Robust Feature also referred to as SURF is a keypoints descriptor that is motivated by SIFT. SURF is proposed in [7] – a robust keypoints detector and descriptor based on the Hessian matrix. It has a wide area of applications that include object recognition, camera calibration, image registration, 3D reconstruction and objet tracking. SURF is computationally efficient with a high degree of repeatability, robustness and distinctiveness compared to other detectors including SIFT. The Harris detector or Harris edge filter as proposed in [8] works by finding keypoints in image area in which the matrix of the second order derivatives has two large eigenvalues. In [9], Features from Accelerated Segment Test detector also known as FAST is proposed. FAST works by comparing the intensities value of a pixel with its circular neighborhood pixels. The Local Binary Pattern is a local descriptor that works by acquiring the intensity value of an image in a small neighborhood around a central pixel. The local binary pattern consists of a string of bits in which each pixel in the neighborhood is represented by one bit. These binary patterns are rarely used directly instead they are first quantized and transformed into a histogram. LBP was made famous by the work presented in [10].

3. Methodology

The block diagram of our cascaded Upright FAST Harris and BRIEF method is shown in Figure 1. In the diagram, keypoints from input images are detected using the U-FAH method and for every keypoint extracted, its descriptions around the neighborhood are computed using the BRIEF method

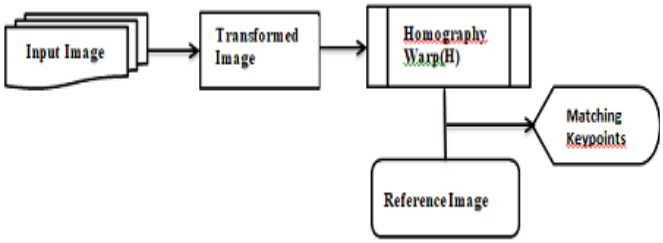


Figure 1: Schematic Diagram of keypoints matching between pair of images

As depicted in Figure1, keypoits and their descriptions are computed for both the transformed image and the reference images using U=FAHB. Subsequently, Homography warp is computed for the transformed image in order to align each of the transform images with the reference image. To keep the text clean and concise, we have a complete discussion of our proposed method in section 3.2 of this paper.

3.1. Dataset

In this work, dataset from real images that represent different types of scenes (see Figure 2) are extracted and

the recall and 1-precision criterion with regard to matching descriptor are used to evaluate the performance of U-FAHB against other state of the art descriptors.

Figure 2: Image pairs representing different imaging conditions such as: (a)scale changes, (b) image rotation, (c) viewpoint changes and (d) illumination change. These image are standard images for evaluation as proposed in[11]

Figure 2(b) shows a pair of image rotations. The first image is referred to as the reference image, while the second image is obtained by rotating the camera optical axis. The angle of rotation in this case is 30 degrees; the average rotation angle in the dataset. Figure 2(c), shows a pair of image under varying view in which the first image referred to as reference, the second image is obtained by changing the camera position at 20 degree. The pair of image under illumination is shown in Figure 2(d). The illumination pair is obtained by a decreasing illumination

3.2. Upright FAST- Harris Filter with BRIEF

The modular approach employed to the design of Upright FAST-Harris Filter as proposed in [4] offers the benefit to combine UFAH with other descriptors. However, given the low computational resource of a mobile phone, it is important to combine UFAH with a computationally efficient descriptor. In this paper therefore, we consider the Binary Robust Independent Elementary Feature descriptor proposed in [5] due to its computational efficiency and speed. A complete discussion on UFAH can be found in [4]. Hence, in this paper, and for clarity, we restrict the discussion of our cascade approach to Binary Robust Independent Elementary Feature only.

3.2.1. Binary Robust Independent Elementary Feature

Binary Robust Independent Elementary Feature descriptor, referred to as BRIEF is a light-weight and simple to use descriptor of an image patch that is composed of a binary intensity test. The intensity test τ of a given smoothed image patch p is defined as follows:

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

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

$$f_n(p) = \sum_{1 \le i \le n} 2^{i-1}\, \tau(p; x_i, y_i) \quad (2)$$

While different types of test distributions are proposed in [5], we employed the Gaussian distribution around the center of the image patch for a better performance. From our test result, we observed that BRIEF descriptor with 512 length gave a better performance compared to the 256 employed in ORB. In order to reduce the noise associated with individual pixel when performing the binary test operation, a smoothing operation is applied to the image patch. Here, we use an integral image similar to the one used in [8] to perform the image smoothing operation.

3.3. Result of Matching Pair of Images using U-FAHB method

The dataset from images captured under scale change (see Figure 2a) is extracted by varying the camera zoom in the range 0 and 2.5 scale ratio. For all transformed image, the homography warp H that align each of the transformed images with the reference image is computed.

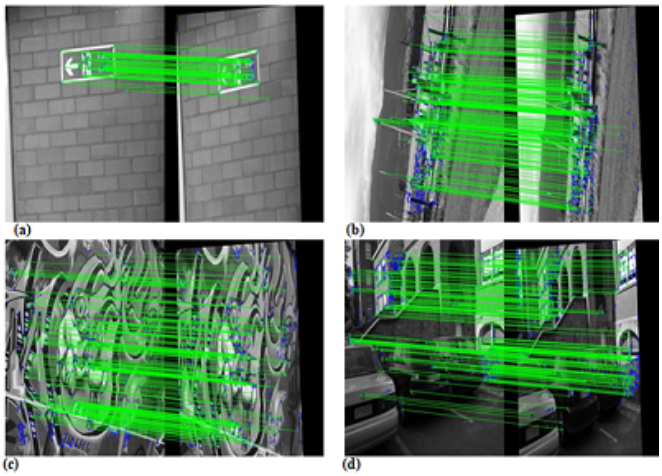


Figure 3: The result of matching descriptors from pair of image under (a) Scale change (b) Image rotation (c) View change and (d) Varying illumination

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

The dataset from images captured under rotation (see Figure 2b) is extracted by rotating the camera optical axis. In this experiment, the angle of rotation of the second image from the reference image is given as 30 degree representing the average rotation angle in this experiment. For each descriptor in the reference image, its nearest neighbor descriptor in the second image is computed and then cross checks their consistency in both directions to reduce false matches. The result of matching the first image 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 reference image, its nearest neighbor descriptor in the second image is computed and then cross checks their consistency in both directions to reduce false matches. Figure 3(c) shows the result of matching the first image with the second image under view change.

The dataset from the images captured under varying illumination (see Figure 2d) is extracted by changing the camera aperture. Figure 3(d) shows the result of matching the descriptors from the first image with their nearest neighbor descriptor in the second image observed under varying illumination.

4. Performance Evaluation of Keypoint Descriptors

The joint performance of the Upright FAST-Harris Filter and the BRIEF descriptor is compared with the state of the art descriptors using the recall and 1-precision metrics. Given a pair of images, feature points and their description are computed for the reference images as well as 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.

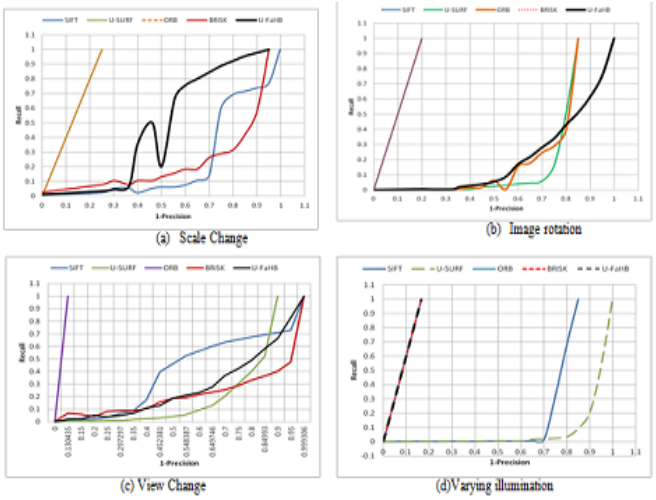


Figure 4: The Precision-recall curves for SIFT, U-SURF, ORB, BRISK and UFAHB descriptors using different dataset extracted from images observed under (a) Scale change (b) Image rotation (c) View change (d) Varying illumination

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

$$recall = \frac{number\ of\ positive\ matches}{number\ of\ corresponding\ regions} \qquad (3)$$

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 be expressed as:

$$1 - precision = \frac{no\ of\ false\ matches}{no\ of (+) matches + no\ of\ (-)\ matches} \qquad (4)$$

The recall vs 1-Precision curve for dataset from image observed under scale change is shown Figure 4(a). As can be observed from the graph, ORB and U-SURF have better performance on scale changes compare to SIFT, BRISK and UFAHB.

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 to have 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.

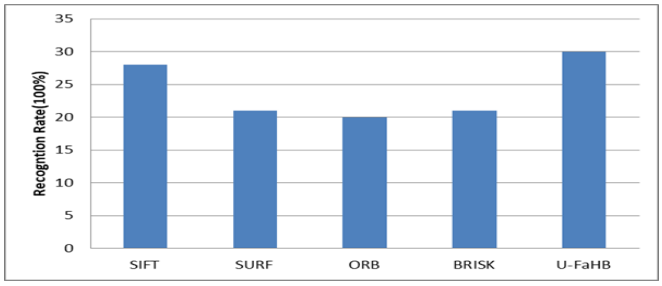


Figure 5: The recognition rate as obtained for all the algorithms (SIFT, SURF, ORB, BRISK and U-FaHB)

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

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

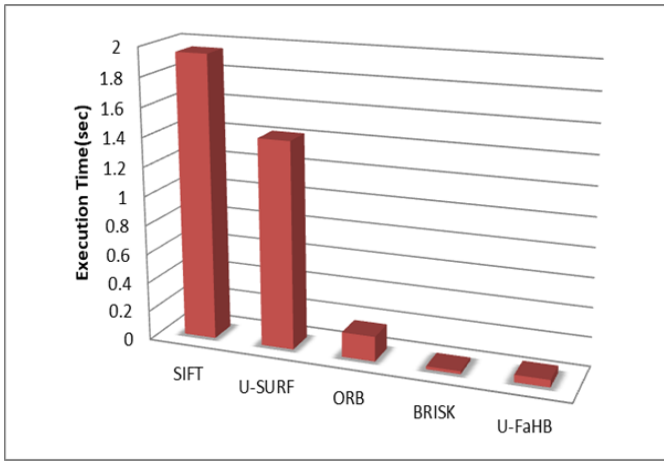


Figure 6: The Average execution time recorded for the different descriptors (SIFT, U-SURF, ORB, BRISK, U-FaHB)

5. Discussion and Conclusion

All descriptors are fairly evaluated using a different dataset of images that are exposed to different photometric and geometric transformations including scale change, rotation, viewpoint change and illumination change. Their performances are analyzed using the precision and recall curve. The description time, which is critical for real time performance is recorded for each descriptor using the same experimental setup.

The recall and 1-precision curve for a pair of images under scale change between 0 and 2.5 is evaluated. In this test (see figure 3a), ORB outperformed the other descriptors, with BRISK and UFAHB following closely in that order. This however, shows that descriptors based on the bit pattern performed extremely well in situation where the scale of an image varies. The recall and 1-precision curves obtained for a pair of images under rotation is evaluated to demonstrate the invariant nature of different descriptors (see figure 3b). The image rotation is 0-30 degrees representing the average rotation in the dataset. SIFT and BRISK outperformed other descriptors followed by UFAHB, ORB and U-SURF. This thus indicate that both SIFT and BRISK perform extremely well under rotation. The recall and 1-precision curve under viewpoint change was evaluated between two images whose viewpoint angles lie between 0 and 20 degrees. Looking at the curve in figure 3c, it is obvious that ORB descriptor is not distinctive even though it is able to match correctly a small number of keypoints correspondences. On the other hand SIFT is highly distinctive compared to other detectors under viewpoint changes, followed by UFAHB and BRISK. The robustness of each descriptor to illumination change was evaluated on a pair of images with decreasing brightness. In this test as shown in figure 3d, UFAHB, ORB and BRISK outperformed the other descriptors showing the robustness of bit pattern to illumination changes.

In order to evaluate the potential of individual descriptor for real time performance, the execution time for each descriptor was analyzed (see table 1). The results obtained show that BRISK is the fastest descriptor and closely followed by UFAHB. Even though, the description time recorded by UFAHB is slow compare to BRISK, it boasts of accurate recognition since all sample points are involved in the matching process. SIFT and U-SURF are the slowest of the descriptors due to their computational complexity.

In conclusion, descriptors based on binary patterns are faster to execute under different imaging conditions. Therefore, the combination of UFAH and BRIEF is promising especially for devices with low computational power.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] E. Rublee et al., "An efficient alternative to SIFT or SURF," *International Conference on Computer Vision* pp. 2564- 2571, 2011, doi: 10.1109/iccv.2011.6126544.

[2] S. Leutenegger et al., "BRISK: Binary robust invariant scalable keypoints," *IEEE International Conference on Computer Vision (ICCV)*, pp. 2548-2555, 2011,doi: 10.1109/iccv.2011.6126542.

[3] 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.

[4] A. D. Mohammed, A. M. Saliu, I. M. Kolo, A. V. Ndako, S. M. Abdulhamid, A. B. Hassan and A. S. Mohammed, "Upright FAST-Harris Filter," i-manager's Journal on Image Processing, vol. 5, no. 3, pp. 14-20, 2018, doi: 10.26634/jip.5.3.15689.

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

[6] Lowe D. G., "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*,vol .60 no. 2, pp. 91-110,2004, doi:10.1023/b:visi.0000029664.99615.94.

[7] H. Bay et al., "Surf: Speeded up robust features," *European Conference on Computer Vision*, pp. 404-417, 2006, doi:10.1007/11744023_32.

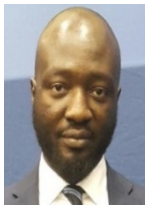
[8] Harris, M. Stephens, "A Combined Corner and Edge Detector," *Alvey vision conference* , pp. 147-151.,1988, doi:10.5244/c.2.23.

[9] Rosten E., Porter R., Drummond T., "Faster and better: A machine learning approach to corner detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 105-119., 2010, doi: 10.1109/tpami.2008.275.

[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/icpr.1994.576366.

[11] K. Mikolajczyk, C. Schmid., "A performance evaluation of local descriptors," *IEEE Transaction on Pattern Analysis and and Machine Intelligence*, vol. 27, no. 10, pp. 1615-1630, 2005, doi: 10.1109/tpami.2005.188.

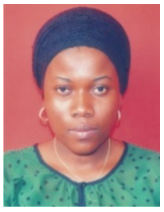
DR. ABDULMALIK DANLAMI MOHAMMED is a Lecturer in the Department of Computer Science at the School of Information and Communication Technology, Federal University of Technology, Minna. He received his PhD in Computer Science from The University of Manchester, United Kingdom, MSc in Computer Science from Belarussian National Technical University, Minsk, Belaruss and BSc in Computer Science from Saint Petersburg State Electro-Technical University, Saint Petersburg, Russia. He has published many academic papers in reputable International Journals, Conference Proceedings and Book chapters. Dr. Abdulmalik Mohammed is the Founder and CEO of Korasight Technovation hub. His research interest includes but not limited to Data Science, Feature Engineering for predictive models, Feature extraction and description for pattern recognition, the application of Machine learning and Deep learning techniques for Emerging Technologies such as the Internet of Things (IoT), Big Data, Computer Vision and Image processing. Dr. Abdulmalik Danlami Mohammed is a member of Nigeria Computer Society (NCS) and International Association of Engineers (IAENG).



DR. OLUWASEUN A. OJERINDE is a lecturer in the Department of Computer Science in the School of Information and Computer Technology in Federal University of Technology, Minna. He bagged his B.Sc. in Computer Technology at Babcock University in 2006. He received his M.Sc. in Mobile Communication System from Loughborough University in 2008. He also obtained his PhD in Mobile Communication System from Loughborough University in 2014. His research area is in Antenna, On-body systems, Multiple Input Multiple Output (MIMO) systems, spanning, Telecommunications, Networking and Radiation. He has worked on the effects of metallic objects on radiation for mobile devices. He is a member of CPN, IEEE and IET.

SALIU ADAM MUHAMMAD, received B. Tech. Mathematics/Computer Science from Federal University of Technology, Minna, Niger State- Nigeria, MSc. Computer Science from Abubakar Tafawa Balewa, Bauchi, Bauchi State Nigeria. He was a PhD. Student in Computer Science & Technology Department, School of Information Science and Electronic Engineering, Hunan University, Changsha, Hunan Province – PR. China, but could not complete the programme owing to financial issues. His PhD programme is currently in view. He was a lecturer in the Department of Mathematics/Computer Science and currently a lecturer in the Department of Computer Science, School of Information & Communication Technology, Federal University of Technology, Minna, Niger State – Nigeria. He has authored and co-authored Thirteen papers in Journals (National & International). He has also participated in ten Conferences (all national) – with four papers in Book of Proceedings, three presentations and three without presentation.

DR. ABUBAKAR SADDIQ MOHAMMED has not only valuable experience in broadcasting, computing and networking engineering but many years of experience in lecturing and research. He holds a Doctor of Philosophy (Ph.D) (Micro & Nano Electronics) from Belarussian State University of Informatics and Radioelectronics (BSUIR), Minsk, Republic of Belarus. He obtained an M.Eng. (Communication Engineering) and B.Eng. (Electrical, Computer & Electronics Engineering) both from Federal University of Technology, Minna. Nigeria. He is a member of Professional bodies among which are: The Council for



MRS EKUNDAYO AYOBAMI is a lecturer in the Department of Computer Science at the School of Information and Communication Technology of Federal University of Technology, Minna, Niger State, Nigeria. She received both B.Sc and M.Sc Computer Science from the University of Ilorin, Kwara State Nigeria. Her research interest includes but not limited to Data mining. Mrs Ekundayo Ayobami is a member of Nigeria Computer Society (NCS). She has published many academic papers in reputable journals and conferences.