IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-9, Sept. 2018 E-ISSN: 2347-2693

Simulation Based Exploration of SKC Block Cipher Algorithm

S. Wilson[1*], A. Lenin Fred[2]

[1]Manonmaniam Sundaranar University, Tirunelveli, India
[2]Mar Ephraem College of Engineering and Technology, Elavuvillai, Marthandam, India

Available online at: www.ijcseonline.org

Accepted: 22/Aug/2018, Published: 30/Sept/2018

Abstract— Video based Face Recognition (VFR) has significantly more challenges when compared to Still Image-based Face Recognition (SIFR). The objective of this paper is to identify faces in video more precisely. In this paper, the minute details of the face are identified by block based technique. It is classified using neural network. The proposed method is tested with four publicly available datasets: Multiple Biometric Grand Challenge (MBGC), Face and Ocular Challenge Series (FOCS), Honda/UCSD and UMD Comcast10 datasets. The proposed method achieves higher recognition rate when compared to other recent methods.

Keywords— Keyframe, Block matching algorithm, face recognition

I. INTRODUCTION

Nowadays, video cameras are widely used for surveillance and mobile devices. Hence there are large amount video data being captured. Compared to still images, videos usually contain more information, e.g., temporal and multi-view information. Video surveillance gives more benefits in terms of security and law enforcement. It is highly desirable to build surveillance systems along with face recognition techniques to automatically detect faces.The majority of existing face recognition algorithms focuses on still image face recognition. Video-based Face Recognition (VFR) research is still in research [1].

Recently, face recognition system have ignored the uniqueness of videos when it is extended from SIFR to VFR. One of the major problems in VFR is its unsolvable severe image blur [2]. The reason is thatlack of real-world video training data, and existing still image databases are usually blur-free. The next problem is pose variations and occlusion which are partially solved in SIFR by ensemble modeling [3]. But this cannot be directly extended to VFR. A common practice in model ensembles is to train models separately for holistic face images and for patches cropped around facial components. Model fusion is then performed offline at the feature [4] or score level [5]. However, the performance of model ensembles is reduced by its significantly increasing time cost, which is impractical for VFR since each video usually contains dozens or even thousands of frames.

Frame quality evaluation is used for key frame selection from video, such that only a subset of best quality frames (keyframes) is selected for efficient face recognition [6]. In addition to frame quality evaluation methods, a number of recent VFR studies attempt to make use of redundant information contained between video frames. Algorithms that fall in this category include sequence-based methods, dictionary-based methods, and image set-based methods [7]. The sequence-based methods aim to extract person-specific facial dynamics from continuous video frames [8], which means that they rely on robust face trackers.

The dictionary-based methods construct redundant dictionaries using video frames and employ sparse representation-based classifiers for classification [9]. Due to the large size of the constructed dictionaries, the dictionary-based methods are often inefficient. The image set-based methods model the distribution of video frames using various techniques such as affine/convex hull [10], linear subspace [1], and manifold methods [11]. Then distribution similarity is measured between frames to match two image sets. The downside of image set modeling is that it is sensitive to the variable volume of video frames and complex facial variations that exist in real-world scenarios [12]. Extracting high-quality face representations has always been a core task in face recognition.

The remaining of the paper is organized as follows: Section II briefly gives an overview of the related works. Section III gives the proposed System Architecture. Section IV explains the algorithms used in the proposed Method. Section V demonstrates the proposed method with some experiments which is followed by Conclusion in Section VI.

II. RELATED WORK

As mentioned earlier, video frames suffer from severe image blur because of the relative motion between the subjects and the cameras. Two types of methods have been introduced to reduce the impact of image blur: deblur-based methods and blur-robust feature extraction-based methods [13]. The former method first estimates a blur kernel from the blurred image and then deblurs the face image prior to feature extraction. However, the estimation of the blur kernel is challenging, as it is an ill-posed problem. In case of blur-robust feature extraction methods, Ahonen et al. [14] developed blur-insensitive Local Phase Quantization (LPQ) descriptor for facial feature extraction, which has been widely used in VFR applications [2]. Furthermore, similar to still face images, faces in video frames exhibit rich pose, illumination, and expression variations and occlusion; therefore, existing studies tend to directly extend feature extractors designed for SIFR to VFR.

Taigman et al. [15] formulated face representation learning as a face identification problem in CNN. Deep metric learning methods [4] are introduced to enhance the discriminative power of face representations. There are only a limited number of studies on CNN-based VFR. Recently, Huang et al. [2] introduced pre-training CNN models with a large volume of still face images and then fine-tuning the CNN models with small real world video databases. However, the fine-tuning strategyis suboptimal, as it only slightly adapts CNN parametersto the video data.

An image set classification method for the video-based face recognition problem was recently proposed in [16]. This method is based on a measure of between-set dissimilarity defined as the distance between sparse approximated nearest points of two image sets and uses a scalable accelerated proximal gradient method for optimization. This method is named as Sparse Approximated Nearest Points (SANP) method. In [17], Scene Change based Video based Face Recognition (SCVFR) method is proposed which extracted some features by grouping the frames in the video according to the pose and illumination. These features are given to neural network for classification. By grouping the frames, features for each pose are correctly extracted in this method. In Keyframe based Video based Face Recognition (KVFR) method [18], keyframes are created through two step process from which facial features are extracted and are given to convolution neural network.

In the proposed method, the video is summarized to keyframes. The keyframes are divided into blocks from which Gray – Level Co-occurence Matrix (GLCM) are obtained as features. Now, the feature block is compared using Block Matching algorithm (BMA) such as full search, diamond search, hexagon search and octagon search. The proposed method is compared with recent methods and it is proved that the proposed method works better than other methods.

III. SYSTEM ARCHITECTURE

The concept behind the proposed work is to identify the features within a specified distance around the pixel. For this, Block Matching Algorithm is used. As BMA suffers from increased computation time, the video is summarized to keyframes. Feature blocks are generated for each block in keyframe using GLCM. Then the faces in the video are recognized using BMA. The functional block diagram is shown in Fig. 1.

Initially, the video sequence is summarized to keyframes using the algorithm defined in [18]. Then the keyframe is reshaped to a standard size (128 x 128). This reshaping is done as the keyframe is to be divided into equal-sized blocks. The next step is to divide the keyframe into blocks of equal size 4 x 4. Features are extracted for each block in the keyframe. The faces in the video sequence are grouped according to the distance calculated from feature block.

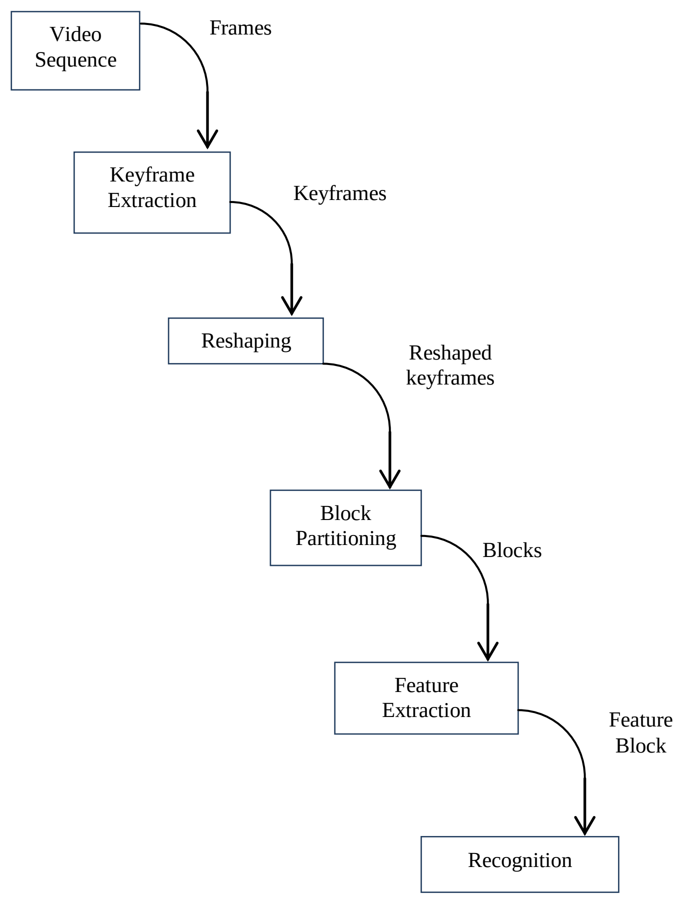


Fig. 1 System Architecture

The proposed method uses GLCMas features for each block. The GLCM is one of the most popular second-order statistical featured Texture property. Haralick in [19] described the use of GLCMs for texture analysis. This analysis is carried out separately for each block of size N x N in the keyframe. The G x G gray level co-occurrence matrix Pd for a displacement vector defined as d = (dx, dy) (polar coordinates, d = (r,θ): The entry (i,j) of Pd is the number of occurrences of the pair of gray levels i and j which are d distance apart. It is given by

$$P\_d\ (i,j)= |\{\{(r,s),(t,v)\} I(r,s)=i, I(t,v)=j\}|$$

where(r,s),(t,v)∈ N x N (t,v)= (r+ dx,s+ dy) and | . | is the cardinality of the set. Instead of using the number of occurrences, we used the probability of occurrence. Thus, we can define the normalized co-occurrence matrix P,P : G x G → [0,1] for an image as

$$P(i,j)=(p\_d\ (i,j))/R$$

where R is the number of pixels in the frame video. Once we obtain the GLCM, we need to form the feature vectors by extracting texture parameters such as those proposed by Haralick [19]. These include the Angular Second Moment (ASM), dissimilarity, correlation, entropy and sum of squares. In this work, sum of squares is used.

$$SOS = \sum_{i,j=0}^{N-1} P(i,j)(i-j)^2$$

IV. BLOCK MATCHING ALGORITHM

The features obtained in the previous section are termed as feature block. This feature block is created for each keyframe of the video sequence. The feature block obtained from the keyframeis compared with feature block generated from other nearest keyframes. The comparison is done using Block Matching Algorithm. There are several block matching algorithms developed in the last two decades.

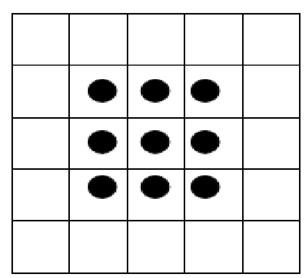
Block matching can only be implemented for the picture having a single object moving in the training picture to form corresponding objects in the testing picture. To implement block matching algorithm, testing picture is to be divided into a matrix of ‘blocks’ that are then compared with corresponding block in the training picture to create a vector. The search area is defined around a block for a search parameter of p (which is usually taken to be 7 pixels on all four sides of the corresponding macro block in the training picture but can vary as per the movement in the pictures). The larger the motions, the larger are search parameter p.

For each block in the current keyframe, one reference keyframe that is the most similar to current blockk is sought in the searching range of size [-P, P]. There are many cost function to compare the blocks like Mean Square Error (MSE), Mean Absolute Difference (MAD), Sum of Absolute Difference (SAD) etc. Among the various cost functions, the one that is less computationally expensive is the Mean Absolute difference (MAD) [20] and is given by the formula:

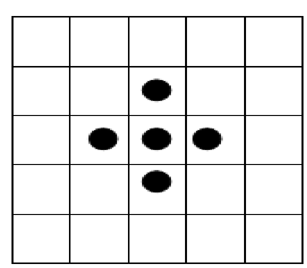
$$MAD = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} |C_{ij} - R_{ij}|$$

Where M and N is the size of the block, Cij and Rij are the pixels being compared in current block and reference block, respectively.

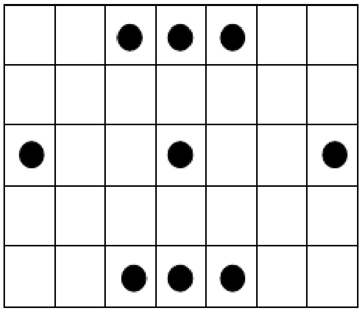
Full search [21] is the basic block matching algorithm which searches the query image block with all the blocks in the training image. Diamond Search [22] has no limit on the number of steps that the algorithm can take but the search should remain inside the defined search range. The end result should see a PSNR close to that of Full Search while computational expense should be significantly less. Diamond Search block matching algorithm uses four points around the center pixel. Hexagonal Search Pattern [23] uses six points for comparing the blocks. Octagon search pattern [24] uses 8 pixels around center pixel. All these search patterns are shown in Fig. 2.



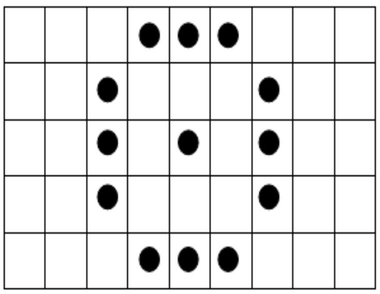
(a)



(b)



(c)



(d)

Fig. 2 (a) Full Search (b) Diamond Search (c) Hexagon Search (d) Octagon Search

The block matching algorithms generate motion vectors (Vx, Vy). The distance is calculated to identify how far the feature block is deviated from the current feature block. It is given by

$$d = (x - x_1)^2 + (y - y_1)^2$$

Where x,y indicates the current location i.e. (0,0), ⟦(x⟧_1,y_1) denotes deviated location. As the original image is resized to 128 x 128 and the image is divided into 4 x 4 equal sized blocks, total numbers of blocks obtained are 32 x 32. So, after calculating the distance, we have 1024 blocks.

To identify the face, mean value of the distance is calculated. It is given by

$$mean = \frac{d_1 + d_{2+\cdots d_{1024}}}{1024}$$

The class with less mean is matched as the recognized face.

V. EXPERIMENTAL RESULTS

The proposed method is tested on four publicly available datasets: the UMD dataset [25], the Multiple Biometric Grand Challenge (MBGC) dataset [26-27], the Honda/UCSD dataset [28] and the FOCS UT-Dallas Video.

a. UMD video

The UMD dataset consists of 12 videos recorded with a group of 16 subjects and it was collected in HD format. It contains sequences of subjects standing without walking toward the camera, which are referred as standing sequences, and sequences of each subject walking toward the camera, which are referred as walking sequences. The video sequences are segmented according to subjects and sequence types. After segmentation, 93 sequences are obtained in total: 70 standing sequences and 23 walking sequences.

b. MBGC Video version 1

The MBGC Video version 1 dataset (Notre Dame dataset) consists of 399 walking and 371 activity video sequences recorded of 146 subjects and were collected in SD format (720 × 480 pixels) and HD format (1440 × 1080 pixels). The 399 walking sequences consist of 201 sequences in SD and 198 in HD. For the 371 walking video sequences, 185 are in SD and 186 are in HD. A leave-one-out identification experiments on 3 subsets of the cropped face images from the walking videos were conducted. The 3 subsets are S2 (144 subjects, 397 videos), S3 (55 subjects, 219 videos) and S4 (54 subjects, 216 videos).

c. Honda/UCSD Dataset

The Honda Dataset [28] consists of 59 video sequences of 20 subjects. The experimental procedure presented in [29] is followed. The experiments are done in three cases of the maximum set length as defined in [29]: 50, 100 and full length frames. Image resolution is 20 × 20 pixels.

d. FOCS UT-Dallas Video

UT Dallas video sequences contain Face and Ocular Challenge Series (FOCS) [30]. The FOCS UT Dallas dataset contains 510 walking (frontal face) and 506 activity (non-frontal face) video sequences recorded from 295 subjects with frame size 720 × 480 pixels. The same leave-one-out tests are conducted on 3 subsets: S2 (189 subjects, 404 videos), S3 (19 subjects, 64 videos), and S4 (6 subjects, 25 videos) from the UT-Dallas walking videos.

The results are compared with recent methods such as SANP method [16], SCVFR method [17], and KVFR method [18]. Table 1 shows the recognition rate comparison of the proposed method and the recent methods.

Table 1 Recognition Rate comparison of the proposed method and recent methods

Dataset	Subset	Methods			
		SANP (%)	**SCVFR (%)**	**KVFR (%)**	**Proposed Method (%)**
UMD videos	S2	92.47	93.64	93.68	93.69
	S3, S4, S5	93.41	94.95	95.12	95.3
	S6	98.04	98.21	98.54	98.6
	Average	94.64	95.6	95.78	**95.86**
MBGC walking videos	S2	83.88	87.1	87.95	87.98
	S3	84.02	88.94	89.22	89.26
	S4	84.26	89.23	89.42	89.8
	Average	84.05	88.43	88.86	**89.01**
Honda dataset	50 Frames	84.62	95.21	95.68	95.75
	100 Frames	92.31	98.12	98.35	98.42
	Full Length	100	98.12	98.38	98.53
	Average	92.31	97.15	97.47	**97.57**
UT-Dallas walking videos	S2	48.27	60.42	65.35	65.4
	S3	60.94	78.88	80.17	80.35
	S4	68	81.18	81.95	82.1
	Average	59.07	73.49	75.82	**75.95**

From Table I, it is observed that the proposed method achieves higher recognition rate when compared to other recent methods in all the datasets. The proposed method achieves an average of 5% increase in recognition rate when compared to SANP method. When compared to SCVFR and KVFR methods, the proposed method achieves 1% increase in recognition rate. Fig. 3 shows the pictorial representation of the recognition rate comparison of the proposed method with other recent methods.

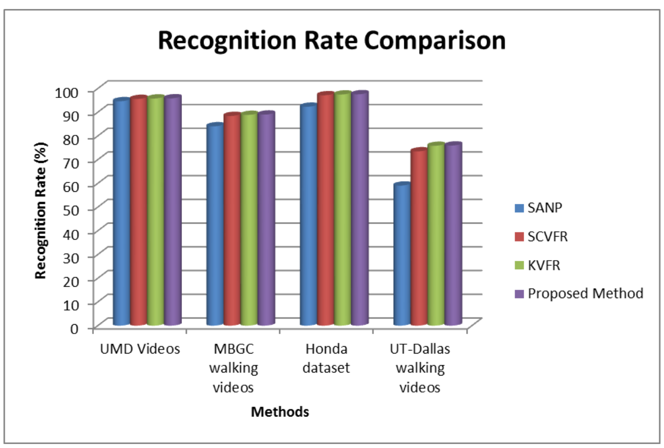


Fig. 3 Bar Chart representation of the comparison of proposed method with recent methods

Table II Increase in Recognition rate of the proposed method with other methods

Dataset	Increase in Recognition Rate (%)		
	SANP	SCVFR	KVFR
UMD Videos	1.14	0.18	0.08
MBGC Videos	4.81	0.43	0.15
Honda Dataset	5.16	0.32	0.10
UT-Dallas Walking videos	16.75	2.33	0.13
Average	6.97	0.82	0.11

From Table II, it is evident that the proposed method achieves 6.97% increase in recognition rate when compared to SANP method, 0.82% increase in recognition rate when compared to SCVFR method and 0.11% % increase in recognition rate when compared to KVFR method.

VI. CONCLUSION AND FUTURE SCOPE

Video based face recognition has the following challenges: more bluring in video and the selection of keyframe for feature extraction. The second challenge is solved by the proposed method by introducing a novel keyframe extraction process. The proposed method is tested on four publicly available datasets and compared with other methods. The experimental results proved that the proposed method achieves higher recognition rate when compared to other methods. The proposed method achieves slight increase in recognition rate when compared to other recent methods. In future, the proposed method can have various other texture features for classification.

REFERENCES

[1] Z. Huang, S. Shan, R.Wang, H. Zhang, S. Lao, A. Kuerban, and X. Chen, “A benchmark and comparative study of video-based face recognition on cox face database,” IEEE Trans. Image Process., vol. 24, no. 12, pp. 5967–5981, 2015.

[2] J. R. Beveridge, H. Zhang et al., “Report on the fg 2015 video person recognition evaluation,” in Proc. IEEE Int. Conf. Automatic Face and Gesture Recognit., 2015, pp. 1–8.

[3] Y. Sun, X. Wang, and X. Tang, “Deeply learned face representations are sparse, selective, and robust,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 2892–2900.

[4] Y. Sun, Y. Chen, X. Wang, and X. Tang, “Deep learning face representation by joint identification-verification,” in Proc. Adv. Neural Inf. Process. Syst., 2014, pp. 1988–1996.

[5] C. Ding, J. Choi, D. Tao, and L. S. Davis, “Multi-directional multi-level dual-cross patterns for robust face recognition,” IEEE

Trans. Pattern Anal. Mach. Intell., vol. 38, no. 3, pp. 518–531, 2016.

[6] J. Phillips, J. R. Beveridge, D. S. Bolme, B. Draper, G. H. Givens, Y. M. Lui, S. Cheng, M. N. Teli, H. Zhang et al., "On the existence of face quality measures," in Proc. IEEE Int. Conf. Biometrics, Theory, Appl. Syst., 2013, pp. 1–8.

[7] J. R. Barr, K. W. Bowyer, P. J. Flynn, and S. Biswas, "Face recognition from video: A review," Int. J. Pattern Recognit. Artif. Intell., vol. 26, no. 05, 2012.

[8] M. Bicego, E. Grosso, and M. Tistarelli, "Person authentication from video of faces: a behavioral and physiological approach using pseudo hierarchical hidden markov models," in Advances in Biometrics, 2006, pp. 113–120.

[9] Y.-C. Chen, V. M. Patel, P. J. Phillips, and R. Chellappa, "Dictionarybased face recognition from video," in Proc. Eur. Conf. Comput. Vis., 2012, pp. 766–779.

[10] Y. Hu, A. S. Mian, and R. Owens, "Face recognition using sparse approximated nearest points between image sets," IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 10, pp. 1992–2004, 2012.

[11] M. T. Harandi, C. Sanderson, S. Shirazi, and B. C. Lovell, "Graph embedding discriminant analysis on grassmannian manifolds for improved image set matching," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2011, pp. 2705–2712.

[12] M. Shao, D. Tang, Y. Liu, and T.-K. Kim, "A comparative study of videobased object recognition from an egocentric viewpoint," Neurocomputing, vol. 171, pp. 982–990, 2016.

[13] R. Gopalan, S. Taheri, P. Turaga, and R. Chellappa, "A blur-robust descriptor with applications to face recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 6, pp. 1220–1226, 2012.

[14] T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkila, "Recognition of blurred faces using local phase quantization," in Int. Conf. Pattern Recognit., 2008, pp. 1–4.

[15] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2014, pp. 1701–1708.

[16] Y. Hu, A. S. Mian, and R. Owens, "Sparse approximated nearest points for image set classification," IEEE Conference on Computer Vision and Pattern Recognition, pp. 27–40, 2011.

[17] Lenin, Wilson S, "An Efficient Key frame Extraction Method in Video based Face Recognition" IPASJ International Journal of Computer Science (IIJCS), Volume 6, Issue 2, February 2018, ISSN 2321-5992.

[18] Lenin, Wilson S, "An Efficient Key frame Extraction Method in Video based Face Recognition" IPASJ International Journal of Computer Science (IIJCS), Volume 6, Issue 2, February 2018, ISSN 2321-5992.

[19] R. Haralick, K. Shanmugan, and I. Dinstein, "Textural feature forimage classification," IEEE Trans. Systems, Man, Cybern., vol.SMC-3, no. 6, pp. 610–621, Nov. 1973.

[20] KatmelBelloulata, Shiping Zhu, and Zaikuo Wang "A Fast Fractal Video Coding Algorithm Using Cross-Hexagon Search for Block Motion Estimation" International Scholarly Research Network ISRN Signal Processing Volume 2011, Article lD 386128, (2011)

[21] J. Huska and P. Kulla, "Trends in block-matching motion estimation algorithms," Dept. of Radioelectronics, Slovak Univ. of Technology, Bratislava, Tech. Rep.

[22] S. Zhu and K.-K. Ma, "A new diamond search algorithm for fast block-matching motion estimation," in Proc. Int. Conf. Inf. Commun. Signal Process. (ICICS '97), vol. 1, Sep. 9–12, 1997, pp.292–296.

[23] C. Zhu, X. Lin, and L.-P. Chau, "Hexagon-based search pattern for fast block motion estimation," IEEE Trans. Circuits Syst. Video Technol.,vol. 12, no. 5, pp. 349–355, May 2002.

[24] Y. Liang, J. Liu, M. Du, "A cross octagonal search algorithm for fast block motion estimation", International Symposium on Intelligent Signal Processing and Communication Systems, Hong Kong, Dec. 13-16, 2005.

[25] R. Chellappa, J. Ni, and V. M. Patel, "Remote identification of faces: problems, prospects, and progress," Pattern Recognition Letters, vol. 33, no. 15, pp. 1849–1859, Oct. 2012.

[26] P. J. Phillips, P. J. Flynn, J. R. Beveridge, W. T. Scruggs, A. J. O'Toole, D. Bolme, K. W. Bowyer, B. A. Draper, G. H. Givens, Y. M. Lui, H. Sahibzada, J. A. Scallan III, and S. Weimer, "Overview of the multiple biometrics grand challenge," International Conference on Biometrics, 2009.

[27] National Institute of Standards and Technology, "Multiple biomertic grand challenge (MBGC)."

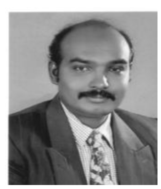
[28] K.-C. Lee, J. Ho, M.-H. Yang, and D. Kriegman, "Visual tracking and recognition using probabilistic appearance manifolds," Computer Vision and Image Understanding, vol. 99, pp. 303–331, 2005.

[29] Y. Hu, A. S. Mian, and R. Owens, "Sparse approximated nearest points for image set classification," IEEE Conference on Computer Vision and Pattern Recognition, pp. 27–40, 2011.

[30] O'Toole A.J, Harms J, Snow S.L, Hurst D.R, Pappas M.R, Ayyad J.H, Abdi H, "Recognizing people from dynamic and static faces and bodies: Dissecting identity with a fusion approach", Vision Research, Vol. 51, No. 1, 2005, pp.74-83.

Authors Profile

S.Wilson received his MCA degree from Manonmaniam Sundaranar University in Tirunelveli, India in 1998 with First class. On 2008, he passed out M.Phil degree in the field of computer science from the same university andin 2010 he completed M.Tech (Computer Science and Information technology) degree with First class. Also, he qualified in State Eligibility Test (SET) for lectureship on 2012 conducted by Bharathiar University, Coimbatore.



Prof. Dr. A. Lenin Fred received B.E in Computer Science and Engineering from Madurai Kamaraj University, India, in 1995 with First class. He has passed M.E. Degree in the same discipline in Madurai Kamaraj University, India and awarded in 2001 with First class. He received his Ph.D. (Doctor of Philosophy) in Computer Science and Engineering at Manonmaniam Sundaranar University, Tirunelveli, India in 2010. His research interest has widen over a variety of fields such as Information Technology, Digital Image Processing, Biometrics, Automatic Fingerprint Identification System and Biometrics.

