World of Computer Science and Information Technology Journal (WCSIT)
ISSN: 2221-0741
Vol. 5, No. 2, 28-33, 2015

An efficient Method to recognise Human Faces from Video Sequences with Occlusion

Vijayalakshmi A.
Dept. of Computer Science
Christ University
Bangalore, India.

Pethuru Raj, PhD
Cloud Architect
IBM Global Cloud Center of Excellence (CoE)
IBM India Bangalore

Abstract - There are several research endeavors and results in aptly recognizing faces from video clips. However recognizing partially occluded and pose variant face images poses a greater challenge and concern for professionals in image and video processing fields. There are efforts underway for unearthing pragmatic algorithms and viable mechanisms for better recognition rate of occluded and pose-variant images. In this paper, a new method is proposed for efficiently recognizing human faces from videos with varying poses and partial occlusion. In the proposed model, a training dataset is created with faces of varying inclination and occluded region that are subjected to in-painting. A variety of features are being extracted using the discrete curvelet transform method from the captured faces in the training set and they are matched against the features from the videos for recognition. We have supplied the implementation results in this paper in order to demonstrate the efficiency of our approach and the algorithm leveraged.

Keywords- video based face recognition; discrete curvelet transform; In-painting and face recognition.

I. INTRODUCTION

Face recognition among the established biometrics mechanisms has gained a huge popularity because it is being used widely across for personal authentication and security purposes. Face recognition is typically non-intrusive and natural when compared to other biometric systems. Specific environments such as hotels, railway stations, airports, nuclear establishments, and other joints wherein people gather around and work are being minutely monitored. There are powerful security and surveillance cameras being manufactured in plenty and deployed in vital junctions and joints in order to ensure utmost security and safety for people and properties. One aspect of ensuing unbreakable security is to do face recognition of people proactively to nib any kind of adventurisms in the budding stage itself. Cameras capture both still as well as dynamic images of all kinds of visitors and their gestures, signals, purposes, and movements. Especially the face portion of people are being precisely captured and communicated to centralized databases and storage appliances in order to be algorithmically analyzed in time. The images captured and subjected to a series of deeper investigations are being stored in databases as well as in archives for posterior processing and lazy investigations.

Face recognition can operate in two different modes 1. Face verification 2. Face identification. Face verification is where a queried face image is matched against a template face image to check the identity. Face identification compares a queried face image against the entire template image in the database to check the identity of the queried face [1]. A face recognition system along with its modules is depicted in Figure 1 [1].

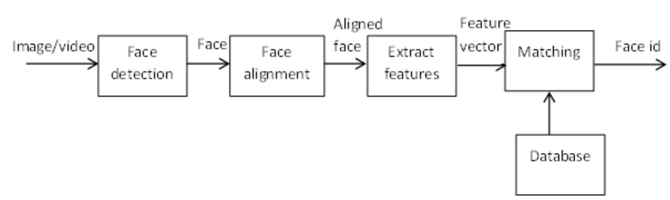

Figure 1. The Modules of face recognition system

There are research groups cogently working on the domain of face-recognition, which happens to be a prominent research area for security and privacy. The art and act of face-recognition is being done on still face as well as dynamic face images. However, there is a leaning towards video face images as they could carry and convey more relevant details. Even on video images, there are a couple of prickling challenges in precisely and perfectly recognizing human face images. That is the occlusion and pose variation are the most dominant issues attracting the attention of researchers across the globe. The reason for video-based face recognition gaining surging popularity when compared to still images-based face recognition in the current scenario is mainly because videos contain more decision-enabling information as compared with still images[1]. In case of videos, they include intrapersonal information from different frames that collectively capture different poses and occluded variations of faces [2]. The direction and the dimension aspects could be captured in videos. The fact of the matter is that video-based face recognition systems started from recognizing still images. Any improvement over still images-based recognition was by tracking the video frames. Tracking video frames could give abundant information as compared with still images. In this case, only faces were taken into consideration. The next advancement in recognition from videos is to consider multiple cues instead of considering only face portion. The next phase started with considering both temporal and spatial information. [3].

The Contribution

In this paper, a new model for face recognition is proposed and proved to be highly efficient. In this model, a training set is created with face samples with various inclination and face samples with occluded regions. The occluded regions are applied to in-painting and reconstructed. The features of these faces in the training set are generated using discrete curvelet transform. The features of the test image are compared with database images using ED.

The Organization of the Paper

The Section I introduces the context for this research work. A study on existing research papers is briefly discussed in the section II, the proposed model is explained in the section III, the explanation of the algorithm is given in the section IV, Finally the implementation results of the proposed model is described in the section V and the conclusion is in the section VI.

II. LITERATURE SURVEY

Himanshu et al. [4] performed a work based on ranked list aggregation. They have proposed an algorithm where a dictionary consisting of various intra-personal variations of faces is used. The frames from input videos are captured and arranged based on the similarity measure of the faces. The ranked list is computed with respect to the frames in the video. Kendall tau distance measure is used to compare the video signatures. Discounted cumulative gain measure is being utilized in matching two videos which uses ranking of images in characterizing the individual in the video. The authors tested the efficiency of the proposed algorithms with Youtube database and MBGC v2 database.

Changubu Ho et al. [5] proposed a method to recognize face from video based on face patches. In this method, patches from face is cropped from the frames of the video and then compared with the original face and stitched together. The reconstructed face is used for recognition. Unlike other methods that make use of part of face for recognition, in this method, the entire face is reconstructed from patches and this whole face is used for recognition. The authors claim to use sparse representation as the method to recover the missing part of the face during reconstruction.

Ming du et al. [6] proposed an approach to recognize pose-variant faces using redundancy in multi-view video data. Multi-view video data makes use of faces with multiple viewpoints. Multi-view gives more information compared to the single view data. This method handles pose variation effectively. Viola jones algorithm is used in face tracking as it is the best-known algorithm for face tracking at this point of time. The features used in this method are LPP and LDA in the original space image and the SH+PCA. The similarity in this approach is measured using the reproducing kernel Hilbert space. The authors claim to have a better recognition rate when compared to traditional algorithms on multi view video database.

Deong Gong et al. [12] proposed a model where training faces from a particular subject is clustered semantically. Further on, clustered faces from different persons are again logically grouped according to their semantics. From each group, features are sliced and on these slices, the PCA+LDA algorithm is systematically applied. While testing, each individual frame is assigned to one of the groups in the training stage. Final matching is considered as the sum of the best matching from individual groups. Authors claim that there is a substantial improvement in recognition rate when compared to traditional approaches.

Yi-Chen chen et al. [13] have in their proposed model used joint sparsity-based approach for video-based face recognition. In their unique model, during the training stage, a face image is partitioned based on different pose and illumination condition and is placed in sub-dictionaries. Each of these sub dictionaries consists of the face in particular viewing condition. During the testing stage, the same partition is found for the input video and the joint sparse representation makes the decision of recognition easy. Authors claim that theirs bring a better recognition rate compared to the existing methods.

Sihago Ding [14] have, in their work, proposed a novel video-based face recognition algorithm known as sequential sample concensus using sequential sampling and updating scheme. In their proposed method, a training set is created and each individual's identity is tested against the training set using a probability mass function. According to the proposed method, for each testing, the frame numbers of samples are drawn from the training set to calculate the weights and are in turn used in updating the identity PMF. Authors claim that this method is robust and resilient against pose variation.

Comparing with still-based approaches, videos are becoming more challenging as it has been used extensively for surveillance activities. Videos are tending to provide more information when compared still image-based details. Videos provide the much-needed temporal continuity. The authors [15] have proposed an adaptive Hidden Markov Model (HMM)-based facial recognition algorithm. The authors [16] have proposed an average data modeling based on local preserving projections by making use of nearest neighbor graphs and eigenmaps. Even though the video-based face recognition is gaining popularity, their still exist disadvantages with respect to video-based face recognition. The major disadvantages in video-based face recognition are poor illumination, different pose variation, occluded face images, and varied expression. This work we have embarked and described in this paper is an attempt to propose a fresh model that can significantly increase the recognition rate of faces from a video sequence. Typically a video sequence contains faces with different poses and occludes different face images.

III. THE PROPOSED MODEL

A. Proposed Model

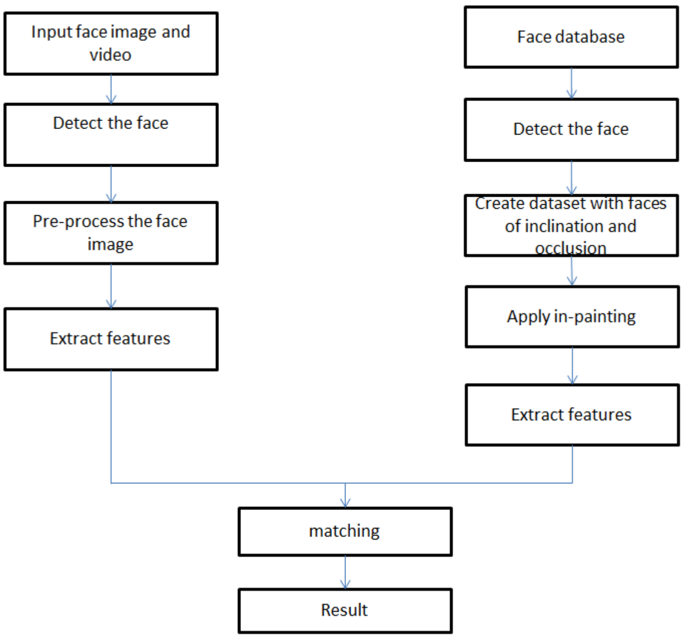

Figure 2. The Proposed model

1. The Image Database

There are various databases available for video-based face recognition. For experimenting on this proposed model, we have used YouTube database which has faces of celebrities. Videos are of low resolution and these videos are recorded at high compression rates. The YouTube database is being used in face recognition in this proposed method. This database is designed for face videos specifically for studying the problem of unconstrained face recognition in videos. The dataset has 3425 videos of 1595 different people. This database includes faces with large variations in the face pose and position, illumination, expression, and other different conditions. This is assumed to be an interesting database for face trackers and recognizers.

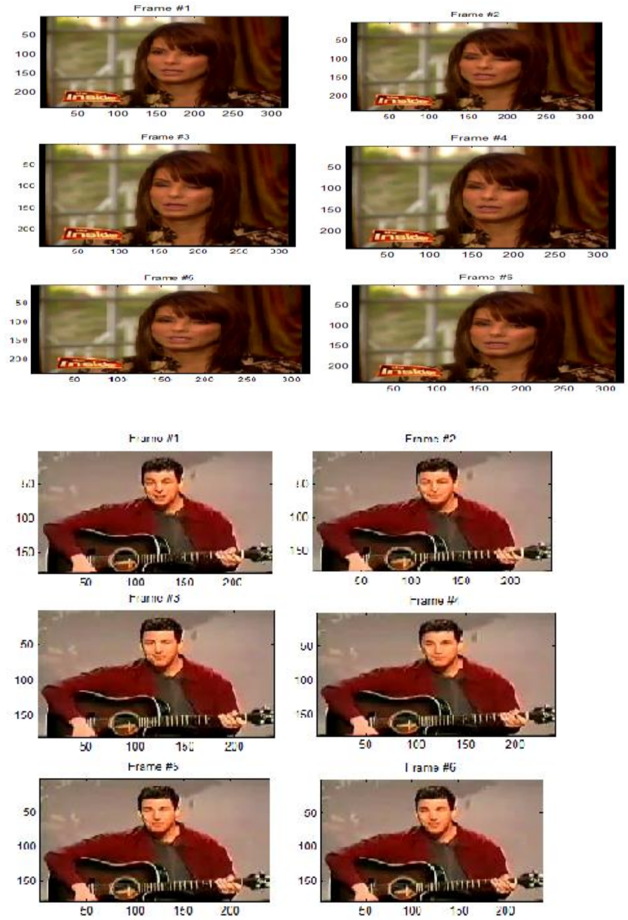

Figure 3. Frames from You Tube data base

2. Face Detection

Skin color is the most important feature of human face that can help in detecting face from the background details in a video. Processing or extracting skin color is faster than extracting any other feature from a face. Considering various faces with different skin colors, the major difference between the faces is the brightness not the chrominance. Adaboost is one of the best methods to detect the face and hence along with skin color extraction, the Adaboost algorithm is used in detecting the face from the video sequence.

Figure 4 Face detected from the video

3. Viola Jones Face Detection Algorithm

Viola Jones is considered as one of the best algorithms for the detection of face [7]. The first step in the Viola Jones algorithm is to select the Haar features. The five Haar features considered are given below.

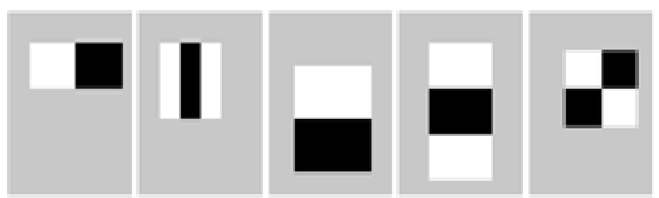

Figure 5. Haar features

All those images with variance lower than one and having less information are ignored. These features help in holding all the required information for the face. Once the Haar features are selected, Adaboost is used for feature extraction and the Adaboost algorithm is explained below in detail [8].

4. The Adaboost Algorithm

(Input)

(1) Training examples $Z = \{(x1, y1), \ldots ,(xN, yN)\}$,where $N = a + b$; of which a examples have $yi = +1$and b examples have $yi = -1$.

(2) The number M of weak classifiers to be combined.

1. (Initialization)

$w_i^{(0)} = \frac{1}{2a}$for those examples with y_i=+1or

$w_i^{(0)} = \frac{1}{2b}$for those examples with y_i=-1

2. (Forward inclusion)

For $m = 1, \ldots , M$:

(1) Choose optimal $hm$to minimize the weighted error.

(2) Choose αm according to$\alpha m=\log \frac{1-\epsilon M}{\epsilon M}$

(3) Update $w_i^{(m)} \leftarrow w_i^{(m)} \exp[-y_i \alpha_m h_m(x_i)]$ and normalize to $\sum_i w_i^{(m)} =1$

.

3. (Output)

Classification function: $H_M(x)= \frac{\sum_{m=1}^{M} \alpha_m h_m(x)}{\sum_{m=1}^{M} \alpha_m}$.

Class label prediction: $\hat{y}(x) = \mathrm{sign}[H_M(x)]$.

5. DISCRETE CURVELET TRANSFORM

The curvelet transform was developed by Candes and Donoho in 1999 and proved to be better when compared with the wavelet transform [8]. Facial images are generally 8 bits, which means they have 256 gray levels. In this type of images, any two very close regions that differ in pixel values will form edges. In case of a face, the edges thus formed are generally curved edges. Curvelets are proved to be good at curved singularities and hence extensively used in extracting edge-based features from face images more efficiently than wavelets. Curvelet transform makes use of the directional transform other than multi-scale time frequency. Candes and Donoho developed two types of curvelet transform, namely fast Fourier transform and fast curvelet transform. When close regions of faces are with conflicting pixel values may lead to creation of edges and as these are with faces, the edges are curved. Curvelets are good at approximating curved singularities; Curvelets can be used in extracting edge based features in facial images [10]. The research conducted by [17] proves to have a better result on face recognition with curvelets. The studies conducted in [18] prove to have a better result when curvelets are used to recognize face images. As the results prove to have a better recognition rate with feature extraction using curvelets, this approach is used in the proposed method.

6. Digital In-Painting

Digital in-painting is a technique of filling in the missing area using the information from the surrounding area of the hole. One of the in painting techniques is texture synthesis. Texture synthesis is based on Markov Random Field. It fills pixel by pixel using the information from neighboring pixel values. Texture synthesis methods are used in texture completion. Texture synthesis algorithms start filling the occluded region from an initial seed while preserving the local structures of the image [11].

IV. THE PROPOSED ALGORITHM

The proposed algorithm is designed in such a way that a predefined training set is created and the features extracted from the video are matched against the features from the training set.

Input: The face image to be recognized and the video in which to be recognized.

Output: The Face Marked Frames in which the face is present.

Creating a training set

Step 1: For the Face image, we extract the skin tones using Gabor features.

Step 2: Also the texture feature prediction is done for the case of fixed poses and most common lighting variations. This is done by inserting noise patterns and generating the multi face image set.

Step 3: With all multi-face images generated, use Discrete_curvelet_transform to extract the features vector for each image and added to training set.

Matching Phase:

Step 1: In each frame of video, detect using face detection algorithm – Voila Jones.

Step 2: Extract faces from the frames.

Step 3: Preprocess the face image by resizing to standard size.

Step 4: On the preprocessed face image, Gabor wavelet is executed to extract the skin texture. Also the Discrete curvlet transform is executed to extract the feature vector for the image.

Step 5: Matching is done against the training data set using minimum distance matching.

Step 6: Once it is matched the training image, provide the result. Also note the position where the face is found and input to the SIFT feature selection.

Step 7: Between subsequent frames, the SIFT matching is done to locate the recognized face and mark in the video frames.

V. RESULTS AND DISCUSSIONS

The aspect of face recognition from video images is there for constant surveillance and the use of it has been increasing day by day. Face recognition from video faces is beset with major problems of occlusion and pose variation. The proposed method is a step in solving this problem in video. A training set is created in the proposed model with multi face images. The features from the stored images in the dataset are extracted using Descrete_curvelet_transform, which is one of the best known algorithms for feature extraction from faces as face includes curved edges and curvelets are good at identifying and retrieving curved edges.

On processing a video, upon face detection, features are extracted using Descrete_curvelet_transform and matching is done against the faces in the training set using minimum distance matching. It is found that when YouTube dataset is used in identifying faces, the accuracy rate of recognising faces is comparatively high with respect to the patch-based method. The following data shows the comparative study on the accuracy at which faces are recognised when compared with existing algorithms.

The algorithm is tested using videos from YouTube database. In the video sequence, both full faces and partly visible faces of each of the subjects are present with occlusions and with pose variation. We compare the recognition rates of the reconstructed face images to that of the full face examples from the video.

TABLE I THE COMPARISON OF ACCURACY WITH POSE VARIATION

Patch based	81%
Proposed	93%

TABLE II THE COMPARISON OF ACCURACY IN PRESENCE OF OCCLUSION

Patch based	75%
Proposed	91%

TABLE III THE COMPARISON OF ACCURACY IN PRESENCE OF DIFFERENT TYPES OF OCCLUSION

Algorithm	Hat	Glass	Binds
Patch based	80%	75%	73%
Proposed	92%	91%	87%

TABLE IV Time to recognize face in a two minute video

Patch based	20 seconds
Proposed	15 seconds

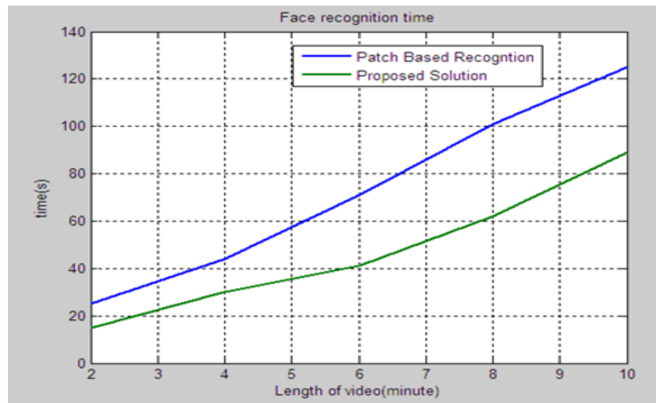

Figure 6. Performance analysis on time taken

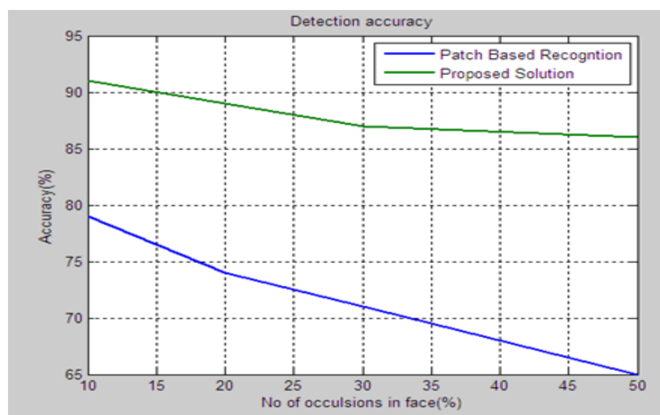

Figure 7. Detection Accuracy

VI. CONCLUSION

In this work, a new and easy-to-use model is proposed to recognise human faces from videos that are partially occluded and with variable inclination. Faces from video are detected and a training set is created with multiple faces of inclination and occlusion. The popular in-painting technique is applied in reconstructing the occluded region. The recognition-enabling features of these faces are meticulously extracted using discrete_curvelet_transform and the extracted features are detected with the face from the input video are subjected for matching. The result of this method shows that our algorithm provides a better recognition rate for partially occluded faces. We are considering a fresh set of innovations in video processing especially video analytics using the big and fast data analytics platforms in public cloud offerings.

References

[1] Dihong Gong; Kai Zhu; ZhiFeng Li; Yu Qiao, "A semantic model for video based face recognition," Information and Automation ICIA),2013 IEEE International Conference on , vol., no., pp.1369,1374, 26-28 Aug.2013

[2] Bhatt, H.S.; Singh, R.; Vatsa, M., "On rank aggregation for face recognition from videos," Image Processing (ICIP), 2013 20th IEEE International Conference on , vol., no., pp.2993,2997, 15-18 Sept. 2013

[3] Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. (2003). Face recognition: A literature survey. Acm Computing Surveys (CSUR), 35(4), 399-458.

[4] Bhatt, H.S.; Singh, R.; Vatsa, M., "On Recognizing Faces in Videos Using Clustering-Based Re-Ranking and Fusion," Information Forensics and Security, IEEE Transactions on , vol.9, no.7, pp.1056,1068, July 2014

[5] Changbo Hu; Harguess, J.; Aggarwal, J.K., "Patch-based face recognition from video," Image Processing (ICIP), 2009 16th IEEE International Conference on , vol., no., pp.3321,3324, 7-10 Nov. 2009.

[6] Ming Du; Sankaranarayanan, A.C.; Chellappa, R., "Robust Face Recognition From Multi-View Videos," Image Processing, IEEE Transactions on , vol.23, no.3, pp.1105,1117, March 2014

[7] Yi-Qing Wang, An Analysis of the Viola-Jones Face Detection Algorithm, Image Processing On Line, 4 (2014), pp. 128–148.http://dx.doi.org/10.5201/ipol.2014.104.

[8] S. Z. Li and A. K. Jain, Handbook of Face Recognition,2nd Edition,Springer, 2011.

[9] E. Cand`es, L. Demanet, D. Donoho, and L. Ying. Fast discrete curvelet transforms. j-MULTISCALE-MODELING-SIMUL, pages 861–899, 2006.

[10] Analysis of Recognition Accuracy Using Curvelet Tranform Arvind Kourav Dr. Prashant Singh Electronics & Comm.Globus, Engg.College Project Head ISC Software Pvt. Ltd., Bhopal, India Bhopal, India

[11] Mahalingam, V. V. (2010). Digital inpainting algorithms and evaluation.

[12] Dihong Gong; Kai Zhu; ZhiFeng Li; Yu Qiao, "A semantic model for video based face recognition," Information and Automation (ICIA), 2013 IEEE International Conference on , vol., no., pp.1369,1374, 26-28 Aug. 2013.

[13] Chen, Yi-Chen, et al. "Video-based face recognition via joint sparse representation." Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on. IEEE, 2013.

[14] X. Liu, T. Chen, "Video-Based Face Recognition Using Adaptive Hidden Markov Models", CVPR, pp. 340-345,2003.

[15] K. Lu, Z. Ding, J. Zhao, Y. Wu, "Video-based Face Recognition",Image and Signal Processing, pp.232-235, 2010.

[16] Radha, V., & Nallammal, N. (2011). Comparative Analysis of Curvelets Based Face Recognition Methods. In Proceedings of the World Congress on Engineering and Computer Science (Vol. 1).

[17] Mandal, T., Majumdar, A., & Wu, Q. J. (2007). Face recognition by curvelet based feature extraction. In Image Analysis and Recognition (pp.806-817). Springer Berlin Heidelberg.

[18] Yi-Chun Lee; Chin-Hsing Chen, "Face Recognition Based on Digital Curvelet Transform," Intelligent Systems Design and Applications, 2008. ISDA '08. Eighth International Conference on , vol.3, no., pp.341,345, 26-28 Nov. 2008

AUTHORS' PROFILE

Vijayalaskhmi A. currently serves as Assistant Professor in the Department of Computer Science at Christ University, Bangalore. She received her B. Sc. degree in Computer Science, Physics and Mathematics, Master of Science in Computer Science from Mangalore University and MPhil in Computer Science from Christ University. She is pursuing her Ph.D. in Computer Science of Christ University under the guidance of Pethuru Raj, PhD, Cloud Architect, IBM Global Cloud Centre of Excellence, IBM India, Bangalore. Her research interests include image and video processing with prominence on the development of enhanced face recognition algorithms from video.

Pethuru Raj has been working as a cloud architect in the IBM Global Cloud Center of Excellence (CoE), IBM India Bangalore. He finished the CSIR-sponsored PhD degree in Anna University, Chennai in the year 1997 and continued the UGC-sponsored postdoctoral research in the department of Computer Science and Automation, Indian Institute of Science, Bangalore. Thereafter, he was granted a couple of international research fellowships (JSPS and JST) to work as a research scientist for 3 years in two leading Japanese universities.

He has been contributing book chapters for a number of high-quality technology books that are being edited by internationally acclaimed professors. Currently in an association with another IBMer, he is writing a book on the topic "Smarter Cities: the Enabling Technologies and Tools" for the CRC Press, USA. He has collaborating with a few authors to publish a book on the title "High-Performance Big Data Analytics" to be published by Springer-Verlag, UK in the year 2015. He maintains an Information Technology (IT) portal at www.peterindia.net and his LinkedIn page is https://www.linkedin.com/in/peterindia