High Dimensional Video-based Face Recognition

Shailaja Arjun Patil, Pramod Jagan Deore

Abstract—High dimensional data is the challenging task in Video-based Face Recognition system. Due to the curse of dimensionality, it needs a more memory space and more processing time (training or testing time). We have proposed a novel approach of concatenation of Graph Wavelet (GW) and Multi-Radius Local Binary Pattern (MRLBP) to VFR. After pre-processing step, the combination of GW and MRLBP provide a flexible model to extract the data features of video and image face database. Independent component Analysis (ICA) is used to reduce these data features. Euclidean distance (ED) is used for matching the data features. The experiments has been done with different face databases (Casia database for image to image recognition and NRC-IIT & HONDA-UCSD for video to video recognition). Experimental results show that, our system achieves better performance, more accuracy, less processing time and less memory space than other VFR algorithms on challenging, high dimensional video face databases and thus advancing the state-of-the-art.

Keywords: High dimensional data, Video-based Face Recognition (VFR), Graph Wavelet (GW), Multi-radius Local Binary Pattern (MRLBP), Independent Component Analysis (ICA), Euclidean distance (ED).

I. Introduction

In daily lives humans perform routinely and effortlessly the task of Face recognition. Wide availability of powerful and low-cost desktop and embedded computing systems has created an enormous interest in automatic processing of digital images and videos in a number of applications, including biometric authentication, surveillance, human-computer interaction and multimedia management [1], [2]. The concept of face recognition is from 1963. Woodrow Wilson Bledsoe is called as a father of face recognition. Bledsoe developed a system that could classify photos of faces by hand [3]. After 1963, many researchers are trying to work on many challenges of face recognition. The performance of face recognition systems has improved significantly since the first automatic face recognition system was developed by Kanade in 1973 [4]. Face Recognition is roughly classified into two broad categories face recognition through images and face recognition through videos. There are many challenges like illumination variation, occlusion, facial expression, resolution variation, high dimensional data and pose variation. In this paper our task is to work on high dimensional data in video databases. In order to evaluate the performance of our system, we use different databases like NRC-IIT [5], Honda-UCSD database [6] for video to video face recognition, for image to image face recognition Casia database [7]. Face recognition algorithms are distinguish between model based and appearance based algorithms [8], [9], [10], model based algorithms use 2D or 3D face models and appearance based algorithms directly use image pixels or features extracted from image pixels [1].

Zhiwu Huang et. al propose a novel Point-to-Set Correlation Learning (PSCL) method, and experimentally show that it can be used as a promising baseline method for V2S/S2V face recognition on COX Face DB. The VFR is far from mature especially compared with face recognition from still images. They suggest more efforts should be made to advance the real-world video-based face recognition applications [11]. Arif Mahmood et. al propose novel multi-order statistical descriptors which can be used for high speed object classification or face recognition from videos or image sets. They apply the proposed algorithm on image set and video face database and periocular biometric identification, object category recognition and hand gesture recognition. They have done their experiments on six benchmark data sets validate that the their proposed method achieves significantly better classification accuracy with lower computational complexity than the existing techniques. Their proposed compact representations can be used for real-time object classification and face recognition in videos [12]. Himanshu S. Bhatt, Richa Singh and Mayank Vatsa presents a VFR algorithm that computes a discriminative video signature as an ordered list of still face images from a large dictionary. In their paper [13], as a future research direction, they have a plan to improve the performance at lower false accept rates for security purpose. Therefore, it requires more computational time as compared to still face recognition algorithms. Another future research direction is to reduce the computational time of the proposed algorithm. From the literature survey, we have planned to propose the new idea of concatenation of GW and MRLBP to overcome the challenge of high dimension and computational time.

High dimensional data arises nowadays in a wide variety of applications, it rule rather than exception in areas like information technology, bio-informatics or astronomy. The word "high-dimensional" refers to the situation where the number of unknown parameters which are to be estimated is one or several orders of magnitude larger than the number of samples in the data [14]. Now a days there is a tremendous increase of data acquisition of audio, images, videos, medical/biological data, industrial processes and social networks. Automatic analysis becomes critical for industries, science medicine, Internet search and new services [15]. High-dimensional statistics refers to statistical inference when the number of unknown parameters p is of much larger order than sample size n, that is $p \gg n$. This encompasses supervised regression and classification models where the number of

Shailaja Arjun Patil, Electronics and Telecommunication Engineering, R. C. Patel Institute of Technology, Shirpur, Dist: Dhule, India, Email: shailajapatil@rcpit.ac.in

Pramod Jagan Deore, Electronics and Telecommunication Engineering, R. C. Patel Institute of Technology, Shirpur, Dist: Dhule, India, Email: hodetc@rcpit.ac.in

covariates is of much larger order than n, unsupervised settings such as clustering or graphical modeling with more variables than observations or multiple testing where the number of considered testing hypotheses is larger than sample size.

High-dimensional statistical inference comes into play whenever the number p of unknown parameters is larger than sample size n typically, we have in mind that p is an order of magnitude larger than n, denoted by $p \gg n$. Most often, we associate a setting where we have more variables than n. High dimensional statistics has relations to other areas. The methodological concepts share some common aspects with non-parametric statistics and machine learning, all of them involving a high degree of complexity making regularization necessary. An early and important book about statistics for complex data is Breiman et al. [16] with a strong emphasis placed on the CART algorithm. The influential book by Hastie et al. [17] covers a very broad range of methods and techniques at the interface between statistics and machine learning, also called "statistical learning" and "data mining". From an algorithmic point of view, convex optimization is a key ingredient for regularized likelihood problems which are a central focus of the book and such optimization arises also in the area of kernel methods from machine learning, cf. Scholkopf and Smola [18]. It include some deviations where non-convex optimization or iterative algorithms are used. Regarding many aspects of optimization, the book by Bertsekas [19] has been an important source for our use and understanding. Furthermore, the mathematical analysis of high-dimensional statistical inference has important connections to approximation theory, cf. Temlyakov [20], in particular in the context of sparse approximations.

A simple yet very useful model for high-dimensional data is a linear model

$$Y_i = \mu + \sum_{j=1}^{p} \beta_j X_i^{(j)} + \varepsilon_i (i = 1.....n), \tag{1}$$

with $p \gg n$. It is intuitively clear that the unknown intercept μ and parameter vector $\beta = (\beta_1....\beta_p)^T$ can only be estimated reasonably well, based on n observations, if β is sparse in some sense. High-dimensional statistical inference is possible, in the sense of leading to reasonable accuracy or asymptotic consistency, if

$$log(p).(sparsity(\beta)) \ll n \tag{2}$$

depending on how we define sparsity and the setting under consideration. Early progress of high-dimensional statistical inference has been achieved a while ago: Donoho and Johnstone [21] present beautiful and clean results for the case of orthogonal design in a linear model where $p = n$. A lot of work has been done to analyze much more general designs in linear or generalized linear models where $p \gg n$, as occurring in many applications nowadays, cf. Donoho and Huo [22], Donoho and Elad [23], Fuchs [24] and many other references given later. Much of the methodology and techniques relies on the idea of ℓ_1-penalization for the negative log-likelihood, including versions of such regularization methods. Such ℓ_1-penalization has become tremendously popular due to its computational attractiveness and its statistical properties which reach optimality under certain conditions. Other problems involve more complicated models with e.g. some non-parametric components or some more demanding likelihood functions as occurring in e.g. mixture models. High-dimensional data applications include text mining, pattern recognition in imaging, astronomy and climate research.

A. Overview of Graph Wavelet

Graphs provide a very flexible model for representing data in many domains. Many networks such as biological networks [25], social networks [26], [27] and sensor networks [28] etc. have a natural interpretation in terms of finite graphs with vertices as data-sources and links established based on connectivity, similarity, ties etc. The data on these graphs can be visualized as a finite collection of samples, which we term graph-signals. For example, graphs can be used to represent irregularly sampled datasets in Euclidean spaces such as regular grids with missing samples. In many machine learning applications multi-dimensional datasets can be represented as point-clouds of vectors and links are established between data sources based on the distance between their feature-vectors [29], [30]. The sizes (number of nodes) of the graphs in these applications can be very large, which presents computational and technical challenges for the purpose of storage, analysis etc. In some other applications such as wireless sensor-networks, the data-exchanges between far-off nodes can be expensive (bandwidth, latency, energy constraints issues). Therefore, instead of operating on the original graph, it would be desirable to find and operate on smaller graphs with fewer nodes and data representing a smooth approximation of the original data. wavelet transform-based techniques would seem well suited to provide efficient local analysis, a major obstacle to their application to graphs is that these, unlike images, are not regularly structured.

A GW can be denoted as $G = (V, E)$ with vertices (or nodes) in set V and links (or edges) as tuples (i, j) in E. The graphs considered are undirected graphs without self-loops and without multiple edges between nodes. The edges can only have positive weights. The size of the graph $N =\mid V \mid$ is the number of nodes and the geodesic distance metric is given as $d(v, m)$, which represents sum of edge weights along the shortest path between nodes u and v, and is considered infinite if u and v are disconnected. The j-hop neighborhood $N_{j,n} = \{v \in V : d(v, n) \leq j\}$ of node n is the set of all nodes which are at most j-hop distance away from node n. Algebraically, a graph can be represented with the node-node adjacency matrix A such that the element $A(i, j)$ is the weight of the edge between node i and j (0 if no edge). The value d_i is the degree of node i, which is the sum of weights of all edges connected to node i and $D = diag(\{d_i\})$ denotes the diagonal degree matrix whose i^{th} diagonal entry is di.

A two-channel wavelet filter bank on a graph provides a decomposition of any graph-signal into a low pass (smooth) graph-signal and a high pass (detail) graph-signal component. The two channels of the filter banks are characterized by the graph-filters $\{H_i, G_i\}_{i\in\{0,1\}}$ and the down sampling operations β_H and β_L as shown in Fig. 1. The transform H_0 acts

as a low pass filter, i.e., it transfers the contributions of the low-pass graph-frequencies, which are below some cut-off, and attenuates significantly the graph-frequencies which are above the cut-off. The high pass transform H_1 does the opposite, i.e, it attenuates significantly, the graph frequencies below some cut-off frequency. The filtering operations in each channel are followed by down sampling operations β_H and β_L, which means that the nodes with membership in the set H store the output of high pass channel while the nodes in the set L store the output of low pass channel. For critically sampled output we have: $\mid H \mid + \mid L \mid = N$ [31] (Fig. 1).

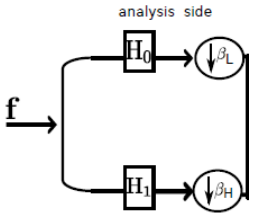

Fig. 1. Graph Wavelet (GW) [31].

B. Multi Radius Local Binary Pattern (MRLBP)

The basic local binary pattern operator, introduced by Ojala et al. [32], [33], was based on the assumption that texture has locally two complementary aspects, a pattern and its strength. In that work, the LBP was proposed as a two-level version of the texture unit to describe the local textural patterns. The original version of the local binary pattern operator works in a 3×3 pixel block of an image. The pixels in this block are thresholded by its center pixel value, multiplied by powers of two and then summed to obtain a label for the center pixel. As the neighbourhood consists of 8 pixels, a total of 2^8 = 256 different labels can be obtained depending on the relative gray values of the center and the pixels in the neighbourhood. An example of an LBP image and histogram are shown in Fig. 2, an illustration of the basic LBP operator is shown in Fig. 3.

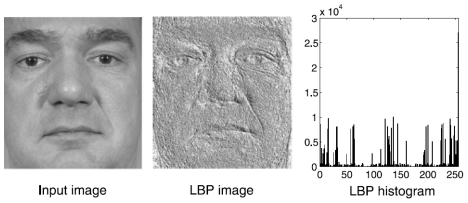

Fig. 2. Example of an input image, the corresponding LBP image and histogram

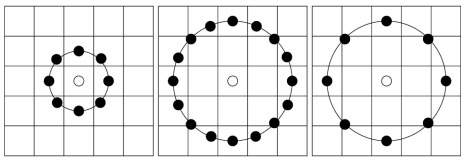

Fig. 3. The circular (8, 1), (16, 2) and (8, 2) neighbourhoods. The pixel values are bilinearly interpolated whenever the sampling point is not in the center of a pixel

The framework of multi-radius analysis, has been developed by the computer vision, image analysis and signal processing communities with complementary motivations from physics and biological vision. The motivation for having a multi-radius representation of the face image comes from the basic observation that real-world objects are composed of different structures at different scales. In this section, a simple but powerful texture representation, called Multi-Radius Local Binary Pattern (MRLBP), is proposed for face recognition. This multi-radius representation based LBP can be obtained by varying the sample radius R and combining the LBP frames or images. It has been suggested for texture classification and the results for this application show that its accuracy is better than that of the single scale local binary pattern method. In general, this Multi-Radius LBP representation method can be realised in a way that it can be accomplished by increasing the radius of the operator. Moreover, this kind of feature has been proven to be important for face detection under different conditions. In summary, increasing the radius of the LBP operator, while keeping the size of the lobe constant overcomes this problem. In our system, the size of the lobe is set to be one pixel. Thus, by sliding a set of LBP operators of different radii over an image and combining their results, a multi-radius representation capable of capturing non-local information can be extracted [34], [35].

However, the general problem associated with the multi-radius analysis is the high dimensionality of the representation combined with the small training sample size. It limits the total number of LBP operators to at most 3. One of the approaches is to employ a feature selection technique to minimise redundant information. We propose another method (GW + MRLBP) which achieves a dimensionality reduction by feature extraction.

Certainly, extracting a multi-radius representation by using a set of LBP operators of different radii may give an unstable result because of noise effect, but this problem can be minimised by using aggregate statistics, such as histogram. There are several advantages in summarising the LBP results in the form of histogram. First, the statistical summary can reduce the feature dimension from the image size to the number of histogram bins. Secondly, using histogram as a set of features is robust to image translation and rotation to a certain extent and therefore the sensitivity to mis-registration is reduced. Finally, although the contribution to the histogram of the unstable LBP responses due to noise is small, it can be further reduced by controlling the number of histogram bins and /or projecting the histogram in other spaces, such as ICA. Zhao et al. [36] have proposed to combine the local binary pattern representation with Kernel Fisher Discriminant Analysis in order to improve the face verification performance of LBP.

C. Two-dimensional Transfer Function (2D-TF)

The 2D Transfer Function (2D-TF) is widely used as a standard way to evaluate the performance of video and digital imaging system because it can provide an objective and quantitative expression of imaging quality, as well as the

capability of calculation from the data. 2D-TF is defined as the magnitude of Optical Transfer Function (OPT) which is the Fourier transform of the incoherent Point Spread Function (PSF). Using discrete Fourier transform to numerically approximate the Fourier transform, the 2D-TF could be calculated as:

$$2DTF = DFT[PSF] = \sum_{n=0}^{N-1} y_n e^{-ikn\frac{2\Pi}{N}} \quad (3)$$

where $k \in [0, N-1]$ and y_n is the position of the n^{th} pixel. Thus, 2D-TF allows for the simplified description of the spatial resolution capabilities in imaging system. When considering the performance based on an image, the 2D-TF also defined as the contrast between a given special frequency and low frequencies, which usually measures the intensity of black and white lines, as shown in Equation 4. I_{max} and I_{min} are the maximal and minimal intensity from the database:

$$2DTF = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \quad (4)$$

For this study, face images from the special subset of Casia dataset have a reference beside the face region. Equation 3 is applied to face region [37]. The transfer function is stable for our image and video face database, it correspond two-dimensional systems.

D. Our Face Descriptor for High Dimensional Analysis

To achieve a more comprehensive description of local facial patterns, the LBP operators with different numbers of sampling points and various neighborhood radii can be combined. The Multi-radius LBP were introduced for facial description [38], to reduce sensitivity of LBP-based face representations to the scale of face images (Fig. 4). It proved that a boosted

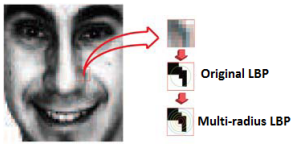

Fig. 4. Multi Radius LBP operator (MRLBP) [38].

classifier of Multi-Radius LBP consistently outperforms that of single scale LBP, and the selected LBP bins are distributed at all scales on the Video Face database. In our approach, we combine the Multi-Radius Local Binary Pattern (MRLBP) representation with GW [34]. GW based MRLBP here is not a specific extension of the original LBP, but denotes a set of approaches that combine GW and the MRLBP features in various ways. It is concluded that GW and the MRLBP features are mutually complementary, since MRLBP captures small appearance details while Graph features encode the facial shape over a broader range of scales. The two types of features can be fused at the feature level, the matching score level as well as the decision level. Such fusion schemes require that the GW and MRLBP features are extracted from the raw image or frames in the parallel way. Local Binary Pattern operators at R scales are first applied to a face image. This generates a grey level code for each pixel at every dimension. The resulting LBP frames or images, shown in Fig. 5, are cropped to the same size and divided into non-overlapping sub-regions, $M_0, M_1, ..M_{J1}$. The regional pattern histogram for each scale is computed. As illustrated in Fig. 5, for a face frame or image, multiple Graph Feature Maps (GFM) are computed by convolving the image with the multi-radius and multi-orientation Graph filters. Each GFM is then divided into

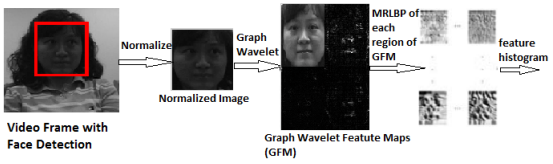

Fig. 5. The illustration of Graph Wavelet (GW) and Multi-Radius Local Binary Patter (MRLBP)

small non-overlapped regions from which MRLBP histograms are extracted and concatenated into a feature histogram for GFM. Moreover, the feature histograms extracted from all GFM are concatenated into a single feature histogram as the final facial representation [34]. In many similar serial method using both wavelets and MRLBP. It first adopts wavelets to decompose the images into four frequency images: low frequency, horizontal high frequency, vertical high frequency and diagonally high frequency, as the inputs of the original LBP. The method gives promising results for Video-based Face recognition.

$$h_{P,r,j}(i) = \sum_{x',y' \in M_j} B(LBP_{P,r}(x', y') = i) \quad (5)$$

$B(v) = \{1, v \geq 0$
$B(v) = \{0, v < 0$
B(v) is a Boolean indicator. The set of histograms computed at different scales for each region, M_j, provides regional information. L is the number of histogram bins. By concatenating these histograms into a single vector, we obtain the final regional face descriptor presented in equation 6.

$$f_j = [h_{P,1,j}, h_{P,2,j}, ..., h_{P,R,j}] \quad (6)$$

This regional facial descriptor can be used to measure the face similarity by fusing the scores of local similarity of the corresponding regional histograms of the pair of images being compared. However, by directly applying the similarity measurement to the multi-radius LBP histogram, the performance will be compromised. The reason is that this histogram is of high dimensionality and contains redundant information. The dimension of the descriptor can be reduced by employing the Independent component analysis (ICA). we refer to the use of ICA to produce statistically independent compressed images. It generates compressed data with minimum mean-squared reprojection error, ICA minimizes both second-order and higher-order dependencies in the input. ICA is used to extract the statistically independent information as a prerequisite to derive discriminative facial features. Thus a regional discriminative facial descriptor, d_j , is defined by projecting the histogram information, f_j , into ICA space W_j^{ica}, i.e.

$$d_j = (W_j^{ica})^T f_j \quad (7)$$

This discriminative descriptor, d_j , gives 4 different levels of locality: 1) the local binary patterns contributing to the histogram contain information at the pixel level, 2) the patterns at each scale are summed over a small region to provide information at a regional level, 3) the regional histograms at different scales are concatenated to produce multi-radius information, 4) the global description of face is established by concatenating the regional discriminative facial descriptors of GW and MRLBP. The diagram of our proposed system is shown in Fig. 6. Our results in this paper show that combining Multi-Radius Local Binary Pattern Histogram (MRLBP) with GW is more robust for high dimensional Video-based Face Recognition System.

II. Proposed High Dimensional Video-based Face Recognition System

In this segment, we propose the complete architecture of high dimensional video-based face recognition system (VFR) in detail as shown in Fig. 6. The proposed system consists of three separate modules. These modules are; (1) detection of the face in the given video frame using Viola-Jones detector (2) feature extraction using the concatenation of GW and MRLBP to represent the face and (3)recognizing the test face from the train video face database.
The database (Casia, IIT-NRC, HONDA-UCSD) consist of male/female videos. In one video clip it consist of 24 frames/second, nearby 300 frames in each Video. Video is a combination of frames/images with respect to time. In this approach we have done a experiment on video databases. We have taken selected frames from all videos of databases, it's advantage is recognition time and memory space required is less, accuracy is 96% to 100%, For training and testing

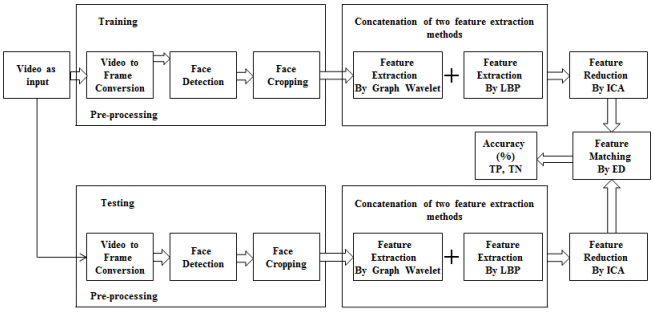

Fig. 6. Proposed high dimensional Video-based Face Recognition System

purpose, we consider the selected frames. We propose a high dimensional VFR system based on GW and MRLBP for Video face databases. For this high dimensional data GW is very effective. The concatenation of MRLBP and GW gives the reduced extracted features. We apply GW and MRLBP for feature extraction, The motivation behind MRLBP is that, MRLBP is one of the good algorithms that deal with VFR. In this approach GW is apply on face image to lower the dimension then the face image is divided into a small non-overlapping blocks or regions, where a histogram of the MRLBP for each region (block) is constructed. The similarity of two images are then computed by summing the similarity of histograms from corresponding regions. For this reason, GW and MRLBP concatenated features are more suitable for face recognition across high dimensional. These concatenated features are applied to the Independent Component Analysis (ICA) for feature reduction purpose. The euclidean distance classifier is used for face recognition.

A. Face Detection

The process of locating the face in a given image and to separate it from the remaining background is called the face detection. Several approaches have been proposed to achieve this with different techniques [39]. Nevertheless, almost most approaches work effectively for frontal faces, where both the eyes are present in the face image. In contrast, if the performance of skin-segmentation based face detection is investigated it proves to a larger degree of variation in poses. After detecting the face next step is face cropping.
Once the face is cropped from the input video frame/image, the classification of the face image is another most important module for our VFR system.

B. Feature extraction by concatenation of GW and MRLBP

After the face detection and face cropping, to represent the face in terms of feature vector to make a machine learning model there is a need of feature extraction technique. In this fragment we discuss the feature extraction through fusion of GW and MRLBP, this fusion is applied on cropped faces. Once the features have been extracted, we studied it for a large number of images or frames and we came to a conclusion, that being the reason for choosing these fusion of the features for high dimensional data. Independent Component Analysis focuses on independent and non-Gaussian components, Higher-order statistics and Non-orthogonal transformation. A signal(x) is generated by linear mixing(A) of independent components(s) ICA is a statistical analysis method to estimate those independent components(z) and mixing rule(W)

$$W * X = U = X = A * U \tag{8}$$

$$W^{-1} = A \tag{9}$$

In Independent Component Analysis a one variable can not be estimated from other variables, it is independent. By Central Limit Theorem, a sum of two independent random variables is more gaussian than original variables distribution of independent components are nongaussian. To estimate Independent Components, z should have nongaussian distribution, i.e. we should maximize nonguassianity. ICA is a Proper to blind source separation or classification using ICs when class id of training data is not available. All the images are easily discussed in dimensional Euclidean space, called image space. It is natural to adopt the base to form a coordinate system of the image space, where it corresponds to an ideal point source with unit intensity location. Thus an image is converted to the gray level at that pixel, is represented as a point in the image space, and is the coordinate with respect to that face image. The origin of the image space is an image whose gray levels are zero everywhere. Although the algebra of the image

space can be easily formulated. The Euclidean distance of images (i.e. the distance between their corresponding points in the image space) could not be determined until the metric coefficients of the basis are given. The metric coefficients are defined as the scalar product and the angle. Euclidean distance converts images into vectors according to gray levels of each pixel and then compares intensity differences pixel by pixel. Here we compare the GW and MRLBP features of test data with GW and MRLBP features of train data using Euclidean distance from which we recognize test video data with the help of video train data. The formula to calculate Euclidean distance is given by,

$$d(x,y) = \sqrt{\sum_{i=1}^{k}(m_i - n_i)^2} \tag{10}$$

Where mi = train image pixel & ni = test image pixel

C. Database Description

The Honda-UCSD dataset [6] contains 20 different subjects distributed over 59 videos. It contains male as well as female videos. Each video sequence is recorded in an indoor environment at 15 frames per second and each lasted for at least 15 seconds. We extracted the faces in these video sequences using Viola and Jones method [39]. We resized the gray-scale images to 20×26 pixels and applied histogram equalization. Fig. 7 shows cropped and re-sized face images from Honda dataset.

NRC-IIT database consists of pairs of short video clips captured by an Intel web-cam mounted on the computer monitor. It shows a wide range of facial expressions and orientations. This database is downloaded from the FRiV technical website [5]. The details of the database are as follows: The video capture resolution is 160×120. Average file size: 1.5 MB, Average duration: 10-20 secs. Average total number of frames in a clip: 300.

CASIA Face Database Version 5.0 (or CASIA-FaceV5) contains 2,500 color facial images of 500 subjects. The face images of CASIA-FaceV5 are captured using Logitech USB camera in one session [7]. The volunteers of CASIA-FaceV5 include graduate students, workers, waiters, etc. All face images are 16 bit color BMP files and the image resolution is 640×480. Typical intra-class variations include illumination, pose, expression, eye-glasses, imaging distance, etc.

Fig. 7. (a) Honda-UCSD (b) NRC-IIT (c) Casia Database

D. Experimental Set-up

We compare the proposed algorithms with 11 object classification techniques including Canonical Correlation Analysis (DCC) [40], Covariance Discriminant Learning (CDL) [41], Manifold to Manifold Distance (MMD) [42], Regularized Nearest Points (RNP) [43], Manifold Discriminant Analysis (MDA) [44], Mean Sequence Sparse Representation Classification (MSSRC) [45], the Linear Affine Hull-based Image Set Distance (AHISD) [46], Sparse Approximated Nearest Points (SANP) [47], the Convex Hull-based Image Set Distance (CHISD) [46], and Set to Set Distance Metric Learning (SSDML) [48]. Standard implementations provided by the original authors are used in our experiments. However, $Hu's$ [47] implementation of MDA is used, while CDL is self-implemented. We use the standard experimental protocol defined previously by [41], [42], [44], [47] and [46], to conduct our experiments. We carefully choose the hyper parameters of each technique involved in our study. For DCC, a 10-dimensional subspace is used to represent image sets. Similarly, 10 maximum canonical correlations are used for discriminative learning. For MMD and MDA, we follow the recommendations of [42] and [44] to configure the hyper parameters. The ratio of Euclidean and geodesic distances is optimized for each dataset. We search different values in the range and report the best results. The top most canonical correlation is used to calculate the MMD. A search space is used to find the best number of connected nearest neighbours for geodesic distance in MDA and MMD. Similarly, a search space of (80%, 85%, 90%, 95%, 99%) is used to select the best value for the number of PCA basis used to represent each image set in AHISD, CHISD and SANP. Parameter C is set to 100 in the SVM optimization framework of CHISD. For RNP [43], 90% PCA energy is preserved and same weight parameters are used as in [43]. MSSRC [45] and SSDML [48] are parameters free.

For Honda-UCSD, NRC-IIT and Casia data sets, we used one image set from each class to construct the gallery while the remaining are used for testing. We performed experiments, each time randomly selecting gallery and test set combinations. For these dataset, we perform experiments based on the standard experimental protocol. Specifically, the experiments are designed in which the complete dataset is divided equally into two parts. Each part of one video consists of minimum 300 frames for each class. In each train-video data set and test-video dataset, gallery is constructed by randomly selecting five frames per class. The training video sets are further partitioned randomly into gallery sets. Specifically, 5 frames (train data set) or images are chosen for the gallery while the other frames or images (test data set) are set aside for validation. Experiments are repeated 5-folds with different combinations of gallery and validation sets in each fold.

The learning process of MLDA and KLDA require at least two samples from every class. Therefore, for the classes having only a single image set available in the gallery, we construct two disjoint image sets from the single one by randomly partitioning it. In our experiments, we preserve 100% energy of the basis, because all discarded basis had zero singular values. In GW [29] based classification, we use the Graph Wavelet feature maps (GFM) to report the results. We perform analysis of GW+MRLBP accuracy for the appropriate choice of the selected frames. Results shows accuracy variations as the order is changed from average 300 frames. For the Honda-

TABLE I
COMPARATIVE RESULTS FOR THREE DATABASES FOR DIFFERENT METHODS

Algorithms	Honda UCSD	NRC-IIT	Casia
DCC [40]	94.67%	93.61%	73.33%
MMD [42]	94.87%	93.19%	69.72%
MDA [44]	97.44%	97.06%	45.53%
CDL [41]	100%	95.83%	75.00%
AHISD [46]	89.74%	97.36%	51.52%
CHISD [46]	92.31%	96.41%	51.67%
SANP [47]	93.08%	96.94%	49.17%
MSSRC [45]	96.75%	97.05%	67.50%
SSDML [48]	89.41%	85.75%	73.20%
RNP [43]	95.95%	96.11%	50.21%
Fj+MLDA [12]	100.00%	97.36%	72.91%
Vj+MLDA [12]	100.00%	97.50%	79.58%
Fj+KLDA [12]	100.00%	97.50%	73.19%
Vj+KLDA [12]	100.00%	97.64%	80.00%
GW+MRLBP (our approach)	100.00%	100.00%	100.00%

UCSD, NRC-IIT and Casia datasets, the highest accuracy was achieved by selecting the 5 random or consecutive frames.

E. Frame Set Classification Results

TABLE I and summarizes the results of our Frame set classification experiments using the three benchmark datasets. In the case of Honda-UCSD dataset, GW+MRLBP combination of our proposed descriptors achieved 100% accuracy and outperformed the comparative methods. The accuracy of SANP, AHISD, and CHISD is lower compared to our proposed descriptor. The first image or frame set of each subject was chosen for gallery and the rest were used as probes. Also, we use 20×26 images in our experiments whereas the image size was 40×40 in [46] and [47]. The effect of high dimension, illumination, pose and resolution has been normalized using MRLBP filtering and increases speed of training and testing using GW. The proposed descriptors GW+MRLBP performed better than CDL which is based on 2-nd order statistic (TABLE I). AHISD, MDA and MSSRC also perform good on this dataset. These experiments show that the GW+MRLBP descriptor performs better than single order descriptors used in CDL, AHISD, MDA and MSSRC. Thus this combinations of proposed descriptor is best than the others. On these dataset, the combination of the proposed descriptors outperformed the existing methods (TABLE I). The image sets in this dataset are relatively more noisy and their structure cannot be perfectly estimated. Therefore, the structure based algorithms (DCC, CDL) perform poor compared to sample based algorithms (AHISD, CHISD, SANP, MSSRC, RNP). In contrast, the proposed descriptors combine both the sample as well as the structural properties of the image or frame sets and are therefore more accurate than the existing methods. Our use of the MRLBP histogram features increases the discrimination. Therefore, our algorithms achieve relatively higher accuracy than previously reported on this dataset [41], [47].

F. Robustness of Experiments

We have used the Honda-UCSD dataset for robustness experiments. We first evaluated the proposed algorithm for its robustness to the number of samples available in each image and Video set for modeling. We randomly selected (25, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300) (frames or images) (for graph plotting we have taken 10 to 100 images per set) samples to form a set. The average recognition rates obtained in these experiments are shown in Fig. 8. The accuracy of the proposed algorithms is relatively lower when 12 frames or images per set were used. However, as the number of images used to construct a set increases, the accuracy of the proposed algorithms increases dramatically and GW+MRLBP achieves 100% recognition rates at 50 images or frames per set. All other algorithms exhibited very different behaviours with the increase in the number of images per set. SANP and CHISD obtained their maximum recognition rates of 96.92% and 97.69% respectively at 20 images per set and further increase in the number of images per set was mostly unfavourable for these two algorithms. The accuracy of MDA linearly increased until 60 images per set however, the accuracy dropped at 70 images followed by a liner increase. The accuracy of DCC increased in a piece-wise liner fashion. The accuracy of MMD increased with a big jump from 10 to 20 images per set however, it quickly reached a saturation value at 50 followed by a decreasing trend for 60 and 70 images per set. A second increasing trend followed for 80 and 90 images per set reaching a saturation level of 92.56% at 100 images per set. The maximum gain in accuracy for the proposed descriptors was till 50 images per set and a saturation level of 100% accuracy was reached at 60 images for all the variants of the proposed descriptors. This shows that 60 random frames per set are optimal for capturing the first and the second order statistics in this dataset. Moreover, the accuracy of the proposed algorithms monotonically increases with the increase in the number of images used to form a set and also have no negative effects of addition of more samples. In the next experiment, we evaluated the robustness of the proposed descriptors and compared it to other algorithms in a set-up similar to [46]. Using the Honda-UCSD dataset, In this experiment we consider all continuous frames (average 300 frames) of video dataset we constructed a clean gallery and test image sets each containing 300 continuous frames or images. This is to ensure that each set should be the same number of frames. In this experiment as we have taken all continuous frames, it increases the the training and testing time (increases time delay) due to MRLBP Histogram. To overcome this problem we combine MRLBP with GW. The results in TABLE III shows that the proposed descriptor GW+MRLBP exhibited robustness to time delay better than the other algorithms. As expected, the structure based techniques are more robust to time delay compared to the sample based techniques (AHISD, SANP, CHISD). This is because the holistic model of set structure has a smoothing effect which reduces the influence of time delay. In contrast, sample based algorithms usually generate interpolated samples from the original samples. This can lead to in-accurate representation. We have performed one more experiments. We set up the experiment such that (g-1)n_f samples are added to each test set, where g is the gallery size and n_f is the number of randomly selected frames from the other gallery sets. By varying n_f from 1 to 3 we added

TABLE II
COMPARISON OF THE AVERAGE ACCURACY OF DIFFERENT ALGORITHMS

Algorithms	$n_f = 1$	$n_f = 2$	$n_f = 3$
MMD [42]	93.83%	93.04%	89.74%
MDA [44]	97.44%	96.73%	95.73%
CDL [41]	98.72%	96.92%	94.62%
AHISD [46]	88.21%	87.31%	87.03%
CHISD [46]	92.11%	91.81%	91.03%
SANP [47]	92.82%	91.54%	91.16%
DCC [40]	93.59%	92.93%	92.31%
Fj+MLDA [12]	100.00%	100.00%	94.87%
Vj+MLDA [12]	100.00%	100.00%	98.12%
Fj+KLDA [12]	100.00%	98.97%	96.92%
Vj+KLDA [12]	100.00%	100.00%	99.49%
GW+MRLBP (our approach)	100.00%	100.00%	100.00%

15, 30 and 45 frames to each probe set. TABLE II shows a comparison of accuracy of different algorithms for these three challenging cases. The drop in the recognition rate of our proposed descriptors is significantly lower compared to the others. For example, in the case of the proposed GW+MRLBP algorithm, the drop in the recognition rate when $n_f = 3$ is 0.5% which is significantly less than the 5.38% drop of CDL and 1.92% drop of SANP. This experiment demonstrated the robustness of the proposed descriptor GW+MRLBP to time delay in the image sets.

In Fig. 8, we compare the recognition rate of different methods with different images or frames per set. We compare the proposed algorithm (GW+MRLBP) with different techniques like DCC [40], Covariance Discriminant Learning (CDL) [41], Manifold to Manifold Distance (MMD) [42], Regularized Nearest Points (RNP) [43], Manifold Discriminant Analysis (MDA) [44], Mean Sequence Sparse Representation Classification (MSSRC) [45], the Linear Affine Hull-based Image Set Distance (AHISD) [46], Sparse Approximated Nearest Points (SANP) [47], the Convex Hull-based Image Set Distance (CHISD) [46] and Set to Set Distance Metric Learning (SSDML) [48], the Multiple Linear Discriminant Analysis (MLDA), Kernel Linear Discriminant Analysis (KLDA), $F_j + KLDA$, $F_j + MLDA$, $V_j + MLDA$ and $V_j + KLDA$ [12].

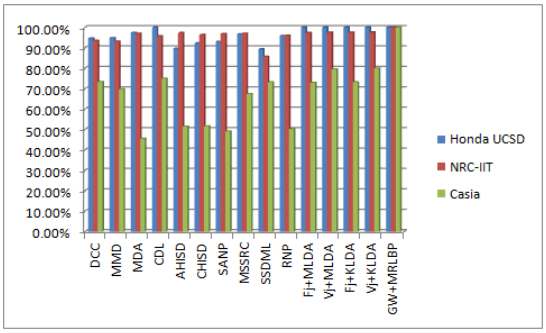

Fig. 8. Comparison graph of different methods for three data sets

G. Comparison of Computational Time

TABLE III, summarizes the average execution times of all algorithms in our study. The execution time is calculated for

TABLE III
COMPARISON OF THE EXECUTION TIMES (IN SECONDS) OF DIFFERENT ALGORITHMS.

Algorithm	Training Time (Sec)	Testing Time (Sec)
DCC [40]	167.49	8.08
MMD [42]	313.57	78.32
MDA [44]	580.70	201.48
CDL [41]	345.88	13.08
AHISD [46]	N/A	18.10
CHISD [46]	N/A	190.61
SANP [47]	N/A	17.94
MSSRC [45]	N/A	30.82
SSDML [48]	400.01	21.87
RNP [43]	N/A	6.42
Fj+MLDA [12]	11.52	0.05
Vj+MLDA [12]	10.63	0.07
Fj+KLDA [12]	5.28	0.04
Vj+KLDA [12]	8.21	0.06
GW+MRLBP (our approach)	6.20	0.06

classifying one probe image set by matching with the 125 gallery image-sets in the Honda-UCSD, NRC-IIT and Casia dataset. The average time of experiments is reported for each algorithm. A Intel(R) core(TM) i3-2120 CPU, 3.30 GHz processor with 8 GB RAM and MATLAB implementations are used to conduct these experiments. These comparisons verify that all variants of the proposed descriptors are significantly faster than existing techniques. For example, the proposed GW+MRLBP is faster than different methods like CDL [41] and SANP [47] the Convex Hull-based Image Set Distance (CHISD) [46] and Set to Set Distance Metric Learning (SSDML) [48], the Multiple Linear Discriminant Analysis (MLDA), Kernel Linear Discriminant Analysis (KLDA), $F_j + KLDA$, $F_j + MLDA$, $V_j + MLDA$ and $V_j + KLDA$ [12] respectively. Our use of MRLBP histogram features increases the discrimination but makes the feature dimension d very high (d = 928). Therefore, the existing algorithms suffer from computational complexity as well as space complexity. However, even for such high dimensional features, all the variants of the proposed descriptors are significantly faster. This shows that the proposed descriptors have better scalability for high dimensional and large datasets. We have also significantly optimized the implementation of CDL to achieve faster execution times.

III. CONCLUSION

In this paper, we have proposed the concatenation of GW and Multi-Radius Local Binary Pattern (MRLBP) for high-dimensional video and image database. Dimensionality of the descriptors reduced using Independent Component Analysis (ICA). The proposed descriptors are compared with 14 existing algorithms on three datasets. Experimental results demonstrate that the proposed descriptors are computationally efficient, robust and highly accurate for video-based face recognition tasks. Experiments also demonstrate that the Multi-Radius Local Binary Pattern and GW descriptors are robust to small number of samples per set and the large number of samples per set in the probe sets as well in the gallery sets. In terms of execution time speed-up, the proposed descriptors are more

faster than the nearest competitor. Therefore, in future the proposed descriptors will potentially be used for real time face recognition with occlusion in videos.

ACKNOWLEDGMENT

This project is supported by University Grant Commission, Western Regional Office, Ganeshkhind, Pune-411007, Maharashtra, India. $(No.F. : 47 - 1110/14(WRO), Dated : February25, 2016)$
Authors are grateful to reviewers for their useful suggestions.

REFERENCES

[1] Stan Z. Li and Anil K. Jain, "Handbook of Face Recognition", ISBN 0-387-40595-X, Springer Science Business Media, 2005.
[2] Stan Z. Li and Anil K. Jain, "Handbook of Face Recognition", ISBN 978-0-85729-4, Springer Science Business Media, 2011.
[3] Woodrow Wilson Bledsoe, "A study to determine the feasibility of a simplified face recognition machine", Phd Thesis, King-Hurley research group, Washington, D. C., January 1963.
[4] T. Kanade, "Picture Processing by Computer Complex and Recognition of Human Faces", PhD thesis, Kyoto University, 1973.
[5] http://www.videorecognition.com/db/video/faces/cvglab/avi/
[6] vision.ucsd.edu/ leekc/HondaUCSDVideoDatabase/HondaUCSD
[7] http://biometrics.idealtest.org/downloadDB
[8] Huy Tho Ho and Rama Chellappa,"Pose-Invariant Face Recognition Using Markov Random Fields",IEEE Transactions on Image Processing, Vol. 22, Issue 4, pp. 1573-1584, 2013.
[9] Utsav Prabhu, Jingu Heo and Marios Savvides,"Unconstrained Pose-Invariant Face Recognition Using 3D Generic Elastic Models", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, Issue 10, pp. 1952 - 1961, 2011.
[10] Richa Singh, Mayank Vatsa, Arun Ross and Afzel Noore," A Mosaicing Scheme for Pose-Invariant Face Recognition", IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), Vol. 37, Issue 5, pp. 1212 - 1225, 2007.
[11] Zhiwu Huang, Shiguang Shan, Ruiping Wang, Haihong Zhang, Shihong Lao, Alifu Kuerban and Xilin Chen, "A Benchmark and Comparative Study of Video-Based Face Recognition on COX Face Database", IEEE Transactions on Image Processing, vol. 24, Issue. 12, pp. 5967-5981, 2015.
[12] Arif Mahmood, Muhammad Uzair and Somaya Al-maadeed, "Multi-order Statistical Descriptors for Real-time Face Recognition and Object Classification", IEEE Access, issue: 99, Jan. 2018.
[13] Himanshu S. Bhatt, Richa Singh and Mayank Vatsa, "On Recognizing Faces in Videos Using Clustering-Based Re-Ranking and Fusion", IEEE Transactions on Information Forensics and Security, vol. 9, Issue: 7, pp. 1056-1068, 2014.
[14] Peter Buhlmann and Sara van de Geer, "Statistics for High-Dimensional Data, Methods, Theory and Applications", Springer-Verlag Berlin Heidelberg, ISSN: 0172-7397, ISBN 978-3-642-20191-2, e-ISBN 978-3-642-20192-9, 2011.
[15] David L. Donoho, "High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality", AMS Conference on Math Challenges of the 21st Century, 2000.
[16] Breiman L., Friedman J., Olshen R. and Stone, "Classification and Reression Trees", Wadsworth, 1984
[17] Hastie T., Tibshirani R. and Friedman J, "The Elements of Statistical Learning; Data Mining, Inference and Prediction", Springer, New York, 2001.
[18] Scholkopf B. and Smola A., "Learning with Kernels:Support Vector Machines, Regularization, Optimization and Beyond", MIT Press, Cambridge, 2002.
[19] Bertsekas D., "Nonlinear Programming", Athena Scientific, Belmont, MA, 1995.
[20] Temlyakov V., "Nonlinear methods of approximation", Foundations of Computational Mathematics, 2008.
[21] Donoho D. and Johnstone I., "Ideal spatial adaptation by wavelet shrinkage", Biometrika, vol. 81, pp. 425-455, 1994.
[22] Donoho D. and Huo X., " Uncertainty principles and ideal atomic decomposition", IEEE Transactions on Information Theory, vol. 47, pp. 2845-2862, 2001.
[23] Donoho D. and Elad M., "Uncertainty principles and ideal atomic decomposition", Proceedings of the National Academy of Sciences, vol. 100, pp. 2197-2202, 2003.
[24] Fuchs J., "On sparse representations in arbitrary redundant bases", IEEE Transactions on Information Theory, vol. 50, pp. 1341-1344, 2004.
[25] M. Weber and S. Kube, "Robust perron cluster analysis for various applications in computational life science", In Comp Life, pp. 57-66, 2005.
[26] M. Crovella and E. Kolaczyk, "Graph wavelets for spatial traffic analysis", In INFOCOM 2003, vol. 3, pp. 1848-1857, 2003.
[27] M. Girvan and M. E. Newman, "Community structure in social and biological networks", Proc Natl Acad Sci U S A, vol. 99(12), pp. 7821-7826, June 2002.
[28] G. Shen and A. Ortega, "Optimized distributed 2D transforms for irregularly sampled sensor network grids using wavelet lifting", In ICASSP'08, pp. 2513-2516, April 2008.
[29] Sunil K. Narang and Antonio Ortega,"Perfect Reconstruction Two-Channel Wavelet Filter-Banks for Graph Structured Data", IEEE Transactions on Signal Processing (TSP), pp. 1-30, Dec. 2011.
[30] Bhushan D. Patil, Pushkar G. Patwardhan and Vikram. M. Gadre, "Eigenfilter approach to the design of One-dimensional and Multidimensional Two channel Linear Phase FIR Perfect Reconstruction filter banks", IEEE Transactions on Circuits and Systems-I, vol. 55, No. 11, pp. 3542-3551, Dec. 2008.
[31] Sunil K. Narang, Y. H. Chao and A. Ortega, "Graph-wavelet filterbanks for edge-aware image processing", SSP, Aug. 2012.
[32] T. Ojala, M. Pietikinen, and D. Harwood,"Performance evaluation of texture measures with classification based on Kullback discrimination of distributions", Proceedings of the 12th IAPR International Conference on Pattern Recognition (ICPR 1994), vol. 1, pp. 582 585,1994.
[33] T. Ojala, M. Pietikinen, and D. Harwood,"A Comparative Study of Texture Measures with Classification Based on Feature Distributions", Pattern Recognition, vol. 29, pp. 51-59,1996.
[34] Di Huang, Caifeng Shan, Mohsen Ardebilian, Yunhong Wang and Liming Chen, "Local Binary Patterns and Its Application to Facial Image Analysis: A Survey", IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 41, issue: 6, pp. 765-781, 2011.
[35] Di Huang, Caifeng Shan, Mohsen Ardebilian and Liming Chen, "Facial Image Analysis Based on Local Binary Patterns: A Survey", IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, Institute of Electrical and Electronics Engineers, vol. 4, pp.1-17, 2011.
[36] Jiali Zhao, Haitao Wang, Haibing Ren and Seok-Cheol Kee, "Lbp discriminant analysis for face verification", In Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, vol. 3, pp. 161167, 2005.
[37] Fang Hua, Peter Johnson, Nadezhda Sazonova, Paulo Lopez-Meyer, Stephanie Schuckers, "Impact of Out-of-focus Blur on Face Recognition Performance Based on Modular Transfer Function", 5th IAPR International Conference on Biometrics (ICB), DOI: 10.1109/ICB.2012.6199763, 2012.
[38] S. Yan, H. Wang, X. Tang and T. S. Huang, "Exploring feature descriptors for face recognition", in Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), pp. 629-632, 2007.
[39] Paul Viola and Michael Jones, "Robust Real-Time Face Detection", International Journal of Computer Vision, Kluwer Academic Publishers. Manufactured in The Netherlands, 57 (2), pp. 137-154, 2004.
[40] T. K. Kim, J. Kittler and R. Cipolla, "Discriminative learning and recognition of image set classes using canonical correlations", IEEE Transactions on PAMI, vol. 29, no. 6, pp. 1005-1018, 2007.
[41] R. Wang, H. Guo, L. Davis and Q. Dai, "Covariance discriminative learning: A natural and efficient approach to image set classification", CVPR-2012, pp. 2496-2503, 2012.
[42] R. Wang, S. Shan, X. Chen and W. Gao, "Manifold-manifold distance with application to face recognition based on image set", in Computer Vision and Pattern Recognition-2008, pp. 1-8, 2008.
[43] M. Yang, P. Zhu, L. Van Gool and L. Zhang, "Face recognition based on regularized nearest points between image sets", in IEEE International Conference on Automatic Face and Gesture Recognition-2013, pp. 1-7, 2013.
[44] R. Wang and X. Chen, "Manifold discriminant analysis", in Computer Vision and Pattern Recognition-2009, pp. 429-436, 2009.
[45] E. Ortiz, A. Wright and M. Shah, "Face recognition in movie trailers via mean sequence sparse representation-based classification", in Computer Vision and Pattern Recognition-2013, pp. 3531-3538, 2013.

[46] H. Cevikalp and B. Triggs, "Face recognition based on image sets", in Computer Vision and Pattern Recognition-2010, pp. 2567-2573, 2010.
[47] Y. Hu, A. Mian and R. Owens, "Face recognition using sparse approximated nearest points between image sets", IEEE Transactions on PAMI, vol. 34, no. 10, pp. 1992-2004, 2012.
[48] P. Zhu, L. Zhang, W. Zuo and D. Zhang, "From point to set: Extend the learning of distance metrics", in International Conference on Computer Vision-2013, pp. 2664-2671, 2013.

Shailaja Arjun Patil was born in Maharashtra, India. She received B.E. degree in electronics from the North Maharashtra University, Jalgaon, India, in 2000 and the M.E. degree in electronics from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India, in 2008. She is currently pursuing Ph.D. from North Maharashtra University Jalgaon. Her current research interests include image and video processing, biometrics.

Pramod Jagan Deore was born in Maharashtra, India. He received the B.E. degree in electronics from the North Maharashtra University, Jalgaon, India, in 1997. He received the M.E. and Ph.D. degrees from Shri Guru Gobind Singhji (SGGS) Institute of Engineering and Technology, Swami Ramanand Teerth Marathwada University, Nanded, in 1999 and 2007, respectively. He is a Professor of Electronics and Telecommunication Engineering Department at the R. C. Patel Institute of Technology, Shirpur, India. His research interests include interval arithmetic operations applications in robust control, image processing, bio-medical signal processing, microwave circuits and antennas. He has published 40 papers in national/international conferences/journals and he has co-authored two books. He is Life Member of ISTE.