Applied Soft Computing 53 (2017) 396–406

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Face recognition under pose and illumination variations using the combination of Information set and PLPP features

Madasu Hanmandlu, Soniya Singhal*

Electrical Engineering Department, Indian Institute of Technology, New Delhi, India

ARTICLE INFO

Article history:
Received 11 June 2015
Received in revised form
17 November 2016
Accepted 1 January 2017
Available online 13 January 2017

Keywords:
Face recognition
Mamta-Hanman entropy function
Local Principal Independent Components (LPIC)
Pseudo-inverse Locality Preserving Projections (PLPP)
Information sets
Non- linear Shannon transform
Hanman transform
Weber face
Linear directional patterns (LDP)

ABSTRACT

This paper presents a new approach for face recognition under pose and illumination variations. The concept of information set is presented and the features based on this are derived using the Mamta-Hanman entropy function. The properties of an adaptive version of this entropy are given and nonlinear Shannon transform and Hanman transform which area higher form of information set are formulated. The information set based features and the nonlinear Shannon transform features are separately combined with the Pseudo-inverse Locality Preserving Projections (PLPP) for improving their effectiveness. The performance of the combined features is compared with that of the holistic approaches on four face databases (two FERET, one head pose image, and Extended Yale face database). The features from the combination of nonlinear Shannon transform and PLPP give consistent performance on the three databases tested whereas the well known features from the literature show good performance on one or two databases only.

1. Introduction

There are numerous applications centered around face recognition of which mention may be made of secure access to buildings, airports, ATM; general identity verification of passports, banking, licenses, employee IDs; tagging in images and videos; and surveillance. Moreover with the ever growing security concerns, the importance of face biometric has increased many-fold. Thus we see CCTVs in all commercial places to look for criminals in real time. Encouraged by these commercial applications of the automatic face recognition (AFR) we look for the most effective techniques.

Searching a person from images or videos needs a robust face recognition algorithm as it is highly susceptible to variations in illumination and poses. Hence, face recognition under the unconstrained conditions and from video footages is an active area of research now-a-days. Human beings have the ability to recognize faces even after several years, with or without objects on face, under varied expressions, poses and lighting conditions by remembering the previously stored information on the persons. This capability is very hard to provide to the machines. However the advances in the machine learning have given an impetus to cope with the face recognition under the unconstrained conditions.

A lot of work has been done on face recognition in the past few years and many face recognition techniques have been developed as can be seen in survey papers [28–30]. Most of them deal with the frontal faces under the controlled lighting conditions, with or without expressions and objects on the faces [5–8,9,10,11,13,14]. The main problem for the accurate face recognition is caused due to variation in pose, illumination, expression, occlusion and aging. A robust automatic face recognition system for pose variation is needed for security and surveillance purposes. Passwords based on retina, finger prints, voice and face recognition are considered to be more secure than PIN numbers, cards etc. which can be stolen or misplaced easily. As face is non-invasive acquiring a facial image is not difficult. Moreover the noise doesnt pose problems unlike occlusion and low resolution while processing the facial images.

1.1. Geometry based approaches

There are three categories of approaches for face recognition: the geometry based, holistic and hybrid approaches. The geom-

* Corresponding author.
E-mail addresses: mhmandlu@gmail.com (M. Hanmandlu), soniyatech90@gmail.com (S. Singhal).

http://dx.doi.org/10.1016/j.asoc.2017.01.014
1568-4946/

etry based approaches use expert-engineered geometry-based features; however, the holistic approaches use the entire image and/or generic high-level image features as input to the classification algorithms. The third category uses both geometric and holistic features.

The first category includes approaches such as Elastic Bunch Graph Model (EBGM) [1] and 3D face models. The study on various face recognition algorithms by Du and Ward [3] reveals that the best results are due to Blanz and Vetter [2] on FERET. A statistical morphable model of 3D faces learned from a set of textured 3D scans of heads is used. It estimates 3D shape and texture of a face from a single image. A set of six to eight standard feature points such as the corners of the eyes, the tip of the nose, corners of the mouth, ears, and around three points on the contour (cheeks, chin, and forehead) is selected. If any of these points fail to be located in some pose, then lesser number of points is used. Gourier et al. [26] have used two first-order and three second-order Gaussian derivatives at each pixel with respect to its neighborhood. These are normalized by the characteristic scale to derive robust facial features and clustered into regions of an appropriate facial structure thus making them robust towards illumination. The pose is estimated based on the relative image positions of salient image structures. As can be noticed the authors of [26] move from holistic features to geometric features. Limitations of geometry based approaches are: either they are based on the initial facial points selected or they need good equipment setups. Finding fiducial points automatically is in itself a challenging task. We are not pursuing the geometry based category in the present work but are attempting at the illumination and pose invariance.

1.2. The holistic approaches

The holistic approaches that are appearance based require dimensionality reduction. These techniques deal with the pixel intensities of the face only. They eliminate the background by adopting manual means or automatic methods so that the unwanted regions have no role in the training of the model. The plenoptic function or Eigen Light Field proposed by Gross et al. [4], makes use of radiance values emitted from an object in all directions called light fields. It is similar to Eigenface space construction but rather than using images as the input, the function uses light fields to project the training data. The advantage of this approach is that it can recognize any pose from a few images in the training set.

One of the most common methods, Principal Component Analysis (PCA) also called Eigenfaces method is by Turk and Pentland [8,9]. In this method high dimensional face images represented as n-dimensional vectors are reduced into lower d-dimensional ($d \ll n$) vectors based on the maximum variance. This new subspace on which the higher dimensional vectors are projected is known as the face space. Independent Component Analysis (ICA) is another Eigenface approach. ICA finds better basis vectors than those of PCA by considering higher order relationships among pixels. It minimizes both second-order and higher-order dependencies in the input data. The basis vectors so found make the data statistically independent when projected on them. Two architectures of ICA by Bartlett et al. [10] are: (i) *Architecture I* – that treats images as random variables and pixels as outcomes and helps find spatially local basis images for the faces, (ii) *Architecture II* – that treats pixels as random variables and images as outcomes, producing a factorial code representation. Both the architectures give better performance than that of PCA when tested on FERET dataset.

Linear Discriminant Analysis (LDA) is a Fisherface approach [11,12]. It minimizes the within-class differences and maximizes the between-class differences. In this regard we have two matrices: the first is between-class scatter matrix S_B and the second is within-class scatter matrix S_W. The goal is to minimize S_W and maximize S_B or maximize the ratio $|S_B|/|S_W|$. Etemad and Chellappa [11] have compared LDA with PCA. The limitations of these two methods are as follows: LDA aims at preserving the global structure while employing the local structure of importance in the recognition tasks. PCA and ICA lack in encoding the discriminant information.

Pang et al. [5] have proposed Neighborhood Preserving Projections (NPP) which unlike PCA and LDA utilizes the neighborhood information to learn global structure. It modifies the Locally Linear Embedding (LLE) by introducing a linear transform matrix. Li et al. [6] have used Discriminative Uncorrelated Neighborhood Preserving Projections (DUNPP). It preserves within the class neighborhood geometry structure and maximizes the distance between different classes. The extracted features are statistically uncorrelated. Wang et al. [7] have proposed Fisher Locality Preserving Projections (FLPP) by using the Maximum scatter difference criterion (MSDC) in the objective function of LPP. It not only preserves the local structure but also makes use of the class information for classification. FLPP is shown to outperform PCA, LDA, MSDC, NPP and LPP in [7].

Xiaofei et al. [13] have proposed Laplacian face approach called Locality Preserving Projections (LPP) which maps face images into a face subspace for analysis. As compared to PCA and LDA that only see the Euclidean structure of face space, LPP not only preserves the neighborhood structure but also obtains a face subspace containing the essential face structure. Being a linear mapping, it shares the properties of the nonlinear techniques like Laplacian Eigen maps.

Rong et al. [14] have proposed the dimensionality reduction algorithm, an improvement over LPP. It is called Pseudo-inverse Locality Preserving Projections (PLPP) that uses Moore-Penrose pseudo inverse matrix to find the inverse of singular matrices resulting from the under-sampled problems from Eigen equation. The simultaneous diagonalization of three matrices reduces the time complexity.

Giorghiades et al. [27] attempt pose and illumination invariant face recognition. To this end they consider a small number of training images under different illumination conditions and reconstruct the shape and albedo of a face which help determine the pose space. This pose space is sampled such that each pose corresponds to a set of illumination conditions termed as illumination cone which is approximated by a low-dimensional space whose basis vectors are estimated using the generative model.

An important class of methods is HMM based; some related papers are surveyed here. In [32] the face images are converted into a sequence of blocks modeled by HMM on which DCT is applied followed by PCA for dimensionality reduction. Each class is associated with a HMM as this method pertains to one sample face recognition. 2D DWT is used for feature extraction and 1D HMM for classification in [33]. The face recognition method in [34] uses 2D-DCT coefficients as the features and HMM considers hair, forehead, eyes, nose and mouth as its states for modeling. In the proposed approach, we avoid segmenting a face into these regions by partitioning into windows. 2D-distributed HMM (2D-DHMM) is used in [35] by extending the algorithms of EM (Expectation-Maximization), GFB (Genera Forward-Backward), and Viterbi for image segmentation and classification. In contrast to the HMM based approaches that require learning of model parameters our approach computes the needed parameters directly from the windows.

1.3. Hybrid approaches

Bae and Kim [15] have used facial (geometric) features and Eigen (holistic) features to train a neural network model for real time face detection and recognition. Karimi and Task [16] have employed the global features such as eyes, mouth, nose, virtual top of head and chin and their ratios, which together are termed as hybrid

facial features for the determination of age and gender. Jangid et al. [31] have proposed holistic features comprising information set based features derived from non-overlapping sub images of a face and geometric features comprising fiducial points and contour features. The hybrid of these two feature types after feature reduction by 2DPCA is employed for illumination invariant face recognition based on a single training image. The hybrid approaches involve multiple stages of processing. In the initial stages, face segmentation is done. Most of the face segmentation approaches are either based on color information or the detection of edges. Using the color models such as RGB, HSV or YCbCr, it is possible to extract skin region from images by setting certain color threshold ranges. But the limitation of these approaches is that they select objects that fall in these color ranges. Also, owing to variation in skin colors of people, deciding about the range is difficult. Another way to segment a face is by using edge detection or contouring methods to compute the contrast between the background and face. This approach needs to do a lot of rectification while determining and removing the unwanted edges. In this work we are not pursuing face segmentation rather a face is partitioned into windows for extracting the information set based features.

1.4. Motivation for the present work

The above face recognitions suffer from computational complexity and low accuracy under unconstrained conditions such as pose and illumination variations. The first motivation is to seek robust features capable of handling the unconstrained conditions. Though robust features have been used in [26] and [27] but they involve high computation complexity. The second motivation is to reduce time complexity of the existing face recognition methods. To this end we explore the computationally simpler information set based features that have been found robust under the unconstrained conditions in [19]. In addition to these, we will also investigate the effectiveness of the higher form of the information set based features. Next we will combine Pseudo-inverse Locality Preserving Projections (PLPP) with the information set based features as this combination is shown to reduce the computational time drastically. Note that Local Principal Independent Components (LPIC) approach presented in [19] is similar to PLPP but it has no provision for preserving the neighborhood structure for the dimensionality reduction.

The paper is organized as follows: Section 2 gives the formulation of Information set based features. Section 3 presents the adaptive Mamta-Hanman entropy function and its properties. It also provides the derivation of non-linear Shannon transform and Hanman transform for face recognition. Section 4 describes PLPP algorithm for the dimensionality reduction, the illumination invariant approach and the matching algorithm. Section 5 discusses the results of experiments on four databases. The conclusions are given in Section 6.

2. Derivation of information set features

The concept of information set is introduced by Hanmandlu [17] based on the information theoretic entropy function named as the Hanman-Anirban entropy function which owes its origin to Hanmandlu and Anirban Das in [18]. This concept has been utilized by Mamta and Hanmandlu in [19] to derive information set based features such as energy, effective information, sigmoidal or multi quadratic features from non-overlapping windows of an ear image. These features are employed in Local Principal Independent Components (LPIC) for the dimensionality reduction. The reduced feature set from each ear of a user is used for authentication.

2.1. The concept of information set

This concept arises from finding the uncertainty in either the probability values, which lead to probabilistic uncertainty or the property/attribute values which lead to possibilistic uncertainty. The 2D Mamta-Hanman entropy function in [20] which is the generalization of the Hanman-Anirban entropy function is defined as

$$H = \sum\sum p_{ij}^{\alpha} e^{-\left(cp_{ij}^{\gamma}+d\right)^{\beta}} \tag{1}$$

where c, d,α, β and γ are the constant parameters and p_{ij} is the probability. For simplicity we take $\gamma = 1$ in Eq. (1) that represents the probabilistic uncertainty. This entropy is amenable to represent both probabilistic and possibilistic uncertainties. In this paper we are concerned with the possibilistic uncertainty in the gray levels of a sub image which is shown to be better than the probabilistic uncertainty in [17].

We term the subimage as information source and the graylevels as the information source values to prepare the ground for representing the possibilistic uncertainty by Eq. (1). To this end we replace the probability p_{ij} with the information source values (I_{ij}) in Eq. (1), which then becomes

$$H = \sum\sum I_{ij}^{\alpha} e^{-(cI_{ij}+d)^{\beta}} \tag{2}$$

We now choose the parameters of the exponential gain function such that it takes the form of a membership function whose role is to fit an appropriate possibility distribution for the information source values.

Selecting, $c = \frac{1}{I_{max}}$, $d = 1 - \frac{I_{avg}}{I_{max}}$ and $\beta = 1$ in (2) leads to

$$H = \sum\sum I_{ij}^{\alpha} \exp\{-(1-\mu_{ij}^{\beta})\} \tag{3}$$

where $0<\alpha<2$ and the membership function μ_{ij} is $\mu_{ij} = \frac{|I_{ij}-I_{avg}|}{I_{max}}$.

Taking the first order approximation in the exponential gain function in (3) leads to,

$$H = \sum\sum I_{ij}^{\alpha} \mu_{ij}^{\beta} \tag{4}$$

Definition of Information set: The set of information values $\{I_{ij}^{\alpha}\mu_{ij}^{\beta}\}$ is called the information set such that each information value is a product of the information source value and the corresponding membership function value. The values of α and β need to be selected appropriately.

2.2. Some properties of information sets

1) The information values can be modified using any function.

For instance we can apply a sigmoid function on the information values. By this the information values are forced to follow a different distribution.

2) The membership function can be empowered to act as an agent.

The membership function can be empowered with the capabilities that are beyond the scope of a fuzzy set. For example, the complement of a membership function can be an agent. Any intuitionist membership function can also be an agent. The membership function can be selected from other information source values. Thus the agent extends the scope of a fuzzy set.

3) The higher form of information sets called transforms can be derived based on the information values.

As shown in the sequel we can derive the Shannon transform using the Mamta-Hanman entropy function.

4) The spatially varying and time varying information sources can be represented easily by the information sets.

For instance the histogram that gives the spatial variation of gray levels in an image can be converted into the information set. Here the gray levels become the information source values and the frequency of occurrence takes the role of a membership function.

5) Both probabilistic and possibilistic uncertainties can be represented by the entropy function.

The information source values have to be probabilities in the entropy function for unearthing the probabilistic uncertainty and the property/attribute values for unearthing the possibilistic uncertainty.

2.3. Information set based features

We will describe the information set based features as derived in [20]. It is important to note that our unit of information is either the information value , $I_{ij}^{\alpha}\mu_{ij}^{\beta}$ or the complement information value, $I_{ij}^{\alpha}\bar{\mu}_{ij}^{\beta}$.

2.3.1. Energy feature

As per Property-2 we can use the complement of membership function, $\bar{\mu}_{ij}$ as the agent so that the complement information value is:$I_{ij}^{\alpha}\mu_{ij}^{-\beta}$. By summing these values for $\beta=2$ over the indices i and j the k^{th} energy feature is obtained from the k^{th} window as:

$$E_k = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} I_{ij}^{\alpha}\bar{\mu}_{ij}^{2} \tag{5}$$

2.3.2. Sigmoid feature

As per Property-1 of information set, the information value $(I_{ij}^{\alpha}\mu_{ij}^{\beta})$ can be modified by applying a function on it. Note that the modification of information values makes them more efficient. The sigmoid feature is derived from the information value by taking it as a function of the sigmoid function:

$$S_k = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{I_{avg}}{1+e^{-I_{ij}^{\alpha}\mu_{ij}^{\beta}}} \tag{6}$$

2.3.3. Effective information

The Effective information source value is computed from the information values $\left(I_{ij}^{\alpha}\mu_{ij}^{\beta}\right)$ in the k^{th} window using the centroid method as

$$I_k = \frac{\sum_i\sum_j I_{ij}^{\alpha}\mu_{ij}^{e}}{\sum_i\sum_j \mu_{ij}^{e}} \tag{7}$$

where $\mu_{ij}^{\beta} = \mu_{ij}^{e} = e^{-\left(\frac{|I_{ij}-I_{ref}|}{f_h^2}\right)}$ is the exponential membership function with $\beta=1$ and I_{ref} can be taken as $I_{\max}$ and the fuzzifier is given by [21]:

$$f_h^2 = \frac{\sum_i\sum_j\left(I_{ij}-I_{ref}\right)^4}{\sum_i\sum_j\left(I_{ij}-I_{ref}\right)^2}$$

2.3.4. Multi quadratic feature

The multiquadratic function has the monotonically increasing or decreasing response from the center. Adapting (3) we have

$$H = \sum\sum I_{ij}^{\alpha}\exp\{-(1-\mu_{ij}^{\mathrm{M}})\} = \sum\sum I_{ij}^{\alpha}\mu_{ij}^{\mathrm{M}} \tag{8}$$

where $\mu_{ij}^{M} = \sqrt{I_{ij}^2+f_h^2}$. The multi quadratic information value is given by $I_{ij}^{\alpha}\mu_{ij}^{M}$ and the effective multiquadratic value computed from the centroid method for the k^{th} window is:

$$M_k = \frac{\sum_i\sum_j I_{ij}^{\alpha}\mu_{ij}^{M}}{\sum_i\sum_j \mu_{ij}^{M}} \tag{9}$$

We extract the above four feature types from face images using Eqs. (5),(6),(7),(9).

3. Adaptive Mamta-Hanman entropy function

Considering the parameters in Eq. (1) as functions rather than constants as per the original definition in [20], the adaptive form of H is written as

$$H = \sum\sum I_{ij}^{\alpha}e^{-(c_{ij}I_{ij}+d_{ij})^{\beta}} \tag{10}$$

Now the parameters c_{ij} and d_{ij} are variables. The 1D form of this function [20] is:

$$H = \sum_{i=1}^{n} p_i^{\alpha} I(p_i) \tag{11}$$

where $I(p_i) = e^{-\left(c_i p_i^{\alpha}+d_i\right)^{\beta}}$ with $c_i, d_i \in [0,1]$. We will prove some important properties of the adaptive entropy function. In order to simplify proofs we set $\alpha=1$.

3.1. Properties of adaptive entropy function

1.$I(p_i) = e^{-(c_i p_i+d_i)^{\beta}}$ is a continuous function for $\forall\ p_i \in [0,1]$; so $p_i e^{-(c_i p_i+d_i)^{\beta}}$ is also a continuous function being a product of two continuous functions and H being the sum of continuous functions is also a continuous function.

2. $I(p_i)$ is bounded. As $e^{-(c_i p_i+d_i)^{\beta}} < 1, p_i^{\alpha}e^{-(c_i p_i+d_i)^{\beta}}$ is bounded for $\forall\ i$; so is H bounded

3. As p_i increases, $I(p_i)$ decreases; so that we have $\frac{\partial I(p_i)}{\partial p_i} = -c_i\beta(c_i p_i+d_i)^{\beta-1}e^{-(c_i p_i+d_i)^{\beta}} < 0 \text{as } c,\beta>0.$

4. If $p_1 = p_2 = p_3 = \ldots = p_n = \frac{1}{n}$ then H is an increasing function of n.

$$H = \sum_{i=1}^{n}\frac{1}{n^{\alpha}}e^{-\left(\frac{c_i}{n}+d_i\right)^{\beta}} = \frac{1}{n^{\alpha-1}}e^{-\left(\frac{c_i}{n}+d_i\right)^{\beta}} \tag{12}$$

$$\frac{\partial H}{\partial n} = \frac{1}{n^{\alpha}}e^{-\left(\frac{c_i}{n}+d_i\right)^{\beta}}\left[(1-\alpha)+\frac{c_i\beta}{n}\left(\frac{c_i}{n}+d_i\right)^{\beta-1}\right] > 0 \tag{13}$$

Hence this is proved.

5. Note that $H = \sum_{i=1}^{n} p_i^{\alpha}e^{-(c_i p_i+d_i)^{\beta}}$ is a concave function where $p_i \in [0,1]$ and $\sum_{i=1}^{n} p_i^{\alpha} = 1$

To prove that this is concave the Hessian matrix must be negative definite. The Hessian is computed as follows:

$$\frac{\partial H}{\partial p_i} = p_i^{\alpha-1}e^{-(c_i p_i+d_i)^{\beta}\left[\alpha-\beta c_i p_i(c_i p_i+d_i)^{\beta-1}\right]} \tag{14}$$

$$\frac{\partial^2 H}{\partial p_i^2} = e^{-(c_i p_i+d_i)^{\beta}}\begin{bmatrix}\alpha(\alpha-1)p_i^{\alpha-2}-\beta c_i p_i(c_i p_i+d_i)^{\beta-1}\\ \left((\alpha-1)+\alpha p_i^{\alpha-2}\right)+\beta c_i^2 p_i^{\alpha-1}\\ (c_i p_i+d_i)^{\beta-2}(\beta p_i(c_i p_i+d_i)^{\beta}-\beta p_i+1)\end{bmatrix} \tag{15}$$

as c_i, p_i are in [0,1].

$$p_i = \frac{1}{n},\ \frac{\partial^2 H}{\partial p_i^2} = c_i\left(\frac{c_i}{n} - 2\right) < 0 \tag{16}$$

$$\frac{\partial^2 H}{\partial p_i \partial p_j} = 0 \quad \text{and} \quad H_F = \begin{bmatrix} \beta_1 & 0\ldots & 0 \\ 0 & \beta_2\ldots & 0 \\ & \ddots & \\ 0 & 0\ldots & \beta_n \end{bmatrix} \tag{17}$$

Where $\beta_i = c_i(c_i p_i - 2) < 0$, hence all the Eigen values of the Hessian matrix are negative. So the Hessian is negative definite and H is concave.

6. Entropy H is maximum when all p_is are equal. In other words, $p_i = \frac{1}{n}, \forall\, i$

$$\text{That is},\quad (p_1, p_2, \ldots, p_n) = \left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) \tag{18}$$

$$\text{Cosequently},\quad \beta_i = c_i\left(\frac{c_i}{n} - 2\right) < 0 \forall i \tag{19}$$

7. The entropy is minimum if and only if all p_is except 1 are equal to zeros and a single $p_i = 1$

To make better representation of uncertainty, we will introduce its higher form.

3.2. Derivation of Shannon transform

A transform is a higher form of the entropy function and to formulate this transform entropy must be adaptive as only the adaptive entropy function is amenable for conversion into a transform. The gain function of an entropy function is a function of information source values whereas the gain function of a transform is a function of the information values. We employ the third and fourth properties of information sets to assess higher form of uncertainty in the information source values in a window of an image based on the initial uncertainty representation. We now invoke the adaptive Mamta-Hanman entropy function in Eq. (10) to convert it into Shannon transform. By taking $c_{ij} = \mu_{ij}$, $d_{ij} = -1$ and $\beta = 1$, in Eq. (10) this becomes

$$H_k = \sum_i \sum_j I_{ij}^{\alpha} e^{-(\mu_{ij} I_{ij} - 1)} \tag{20}$$

This can be shown to be equivalent to what we term as the non linear Shannon transform:

$$HS_k = -\sum_i \sum_j I_{ij}^{\alpha} log\left(\mu_{ij} I_{ij}\right) \tag{21}$$

This is called a transform because the information source values I_{ij} are modified by the information values, $\{I_{ij}\mu_{ij}\}$.This can be proved by taking the first order approximation of gain function in (20) which leads to $H_k = \sum_i \sum_j I_{ij}^{\alpha}(1 - (\mu_{ij} I_{ij} - 1)) = \sum_i \sum_j I_{ij}^{\alpha}(2 - \mu_{ij} I_{ij})$. It is higher form because the logarithmic gain function operates on the information values.

In some applications the use of complement of μ_{ij} in (21) improves the results as will be shown on face recognition. With this change Eq. (21) becomes

$$HNS_k = -\sum_i \sum_j I_{ij}^{\alpha} log\left(\bar{\mu}_{ij} I_{ij}\right) \tag{22}$$

To derive the above we have taken $c = \bar{\mu}_{ij} = 1 - \mu_{ij}$ in (22).

3.3. Derivation of Hanman Transform

The derivation of two forms of Hanman transform appears in [22] based on Hanman-Anirban entropy function. We will follow the same procedure to derive the first form of Hanman transform for k^{th} window from (10) by taking $c_{ij} = \mu_{ij}$, $d_{ij} = 0$ and β=1 in (10) as

$$HT_{k1} = \sum_i \sum_j I_{ij}^{\alpha} e^{-\mu_{ij} I_{ij}} \tag{23}$$

where μ_{ij} is given by (4). The second form is derived using the histogram of a sub image or window. By replacing I_{ij}^{α} by I_i^{α} and μ_{ij} by the probability p_i in (23) we obtain

$$HT_{k2} = \sum_i I_i^{\alpha} e^{-p_i I_i} \tag{24}$$

This is called the histogram $\left(I_i^{\alpha} vs. p_i\right)$ based Hanman transform or probabilistic transform whereas the first form is called the possibilistic transform. This can be used to represent the uncertainty associated with the spatially varying information source values. If we replace p_i by time intervals t_i then it represents the uncertainty in the time varying information source values. The usefulness of transforms will be demonstrated later in the results section. By the way we can also consider complement Hanman transform just as complement Shannon transform.

$$HCT_{k1} = \sum_i \sum_j I_{ij}^{\alpha} e^{-\bar{\mu}_{ij} I_{ij}} \tag{25}$$

In the above transforms the following Gaussian membership function is found more suitable than the earlier membership functions:

$$\mu_{ij} = e^{-\frac{\left(I_{ij} - Iavg\right)^2}{2\sigma^2}} \tag{26}$$

where σ is calculated as $\sigma = \sqrt{\frac{1}{n^2}\sum_i \sum_j \left(I_{ij} - I_{avg}\right)^2}$.

Here after HT_{k1} and HCT_{k1} are denoted by G_1 and G_2 respectively. The transforms have realistic applications. For example, we gather information about an unknown person of some interest to us. This is the first level of information (set) and then we evaluate him again to get the second level of information camped with the first level of information. The transforms can be used to evaluate not only the information source values but also the membership function values to see whether the selected membership function is appropriate.

To reduce the dimensionality of the feature vector formed from any of the above feature types, an algorihm for Pseudo- Inverse Locality Preserving Projections (PLPP) in [14] is outlined here.

4. Dimensionalty reduction and illumination invariant approach for matching

4.1. PLPP algorithm

For n feature vectors each of dimension m are arranged in the matrix of size $X \in R^{m \times n}$. The steps involved in PLPP are briefly described now, and shown in Fig. 1(a):

Step 1: Construct the weighted k nearest neighborhood-graph from $WG = (X, S)$ where x_i (each data sample) is the node and S is an affinity matrix that represents the local structure. An element of S matrix is computed from:

$$S_{ij} = e^{-\frac{\|x_i - x_j\|^2}{t}} \tag{27}$$

where t is a threshold, x_j is among the k nearest neighbors of x_i otherwise S_{ij} is 0. Instead of S_{ij}, we can use the cosine function to

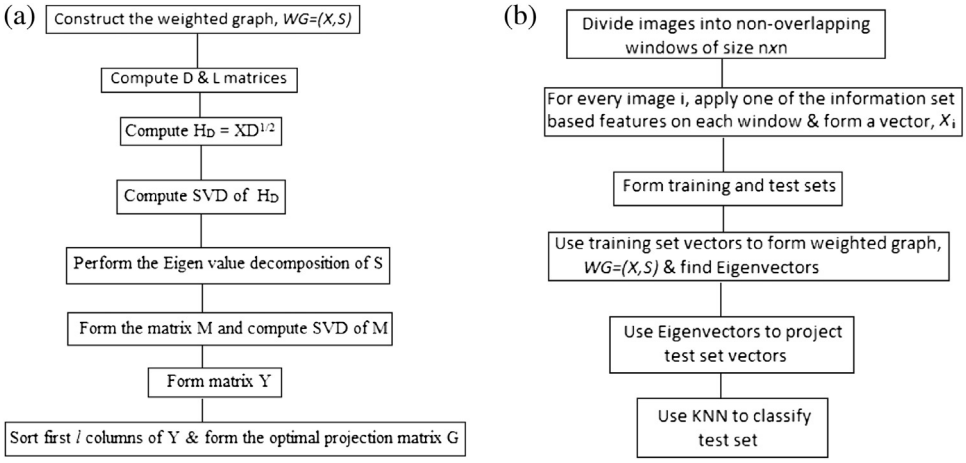

Fig. 1. Flowcharts of a) PLPP Algorithm, b) The Matching Algorithm.

assign weights to the edges. If x_i and x_j belong to the same class, then there exists an edge whose weight is:

$$cos(x_i, x_j) = x_i.x_j = x_i x_j^T \tag{28}$$

Step 2: Find D, L matrices using

$$D_{ii} = \sum_j S_{ij} \tag{29}$$

$$L = D - S \tag{30}$$

where D is a diagonal matrix and L is the Laplacian matrix.

Step 3: Form $H_D = XD^{1/2}$.

Step 4: Compute SVD of H_D from:

$$H_D = U \begin{pmatrix} \Sigma_d & 0 \\ 0 & 0 \end{pmatrix} V^T = U_1 \Sigma_d V_1^T \tag{31}$$

where U and V are unitary matrices, $\Sigma_d \in R^{rxr}$ is a diagonal matrix, $r = \text{rank}(H_D)$. $U_1 \in R^{mxr}$ and $V_1 \in R^{rxn}$ are orthogonal column matrices.

*Step 5:*Perform the Eigen value decomposition of S as $S = Q\Lambda Q^T$, where vectors $Q \in R^{nxn}$, is orthogonal.

Step 6: Form the matrix M as follows.

$$M = \Sigma_d^{-1} U_1^T X Q \Lambda^{1/2} \tag{32}$$

*Step 7:*Compute SVD of M.

*Step 8:*Form Y matrix using

$$Y = \left[U_1 \Sigma_d^{-1} P, U_2\right] \tag{33}$$

where $P \in R^{rxr}$ and U_2 results from colu mn-wise partitionitioing of $U = (U_1, U_2)$.

*Step 9:*Sort the first 'l ' columns of Y, where l is the number of vectors chosen and then form the optimal projection matrix G.

We will use this algorithm in the matching of a test sample with the training samples by computing their feature vectors.

4.2. Illumination invariant approach

As per Weber's law in [23] the ratio of the smallest perceptual change in stimulus to that in the background remains a constant. This implies that stimuli are perceived in the relative terms rather than in the absolute terms. An illumination insensitive representation of a face image $I(x, y)$, named Weber Face, $WI(x, y)$ is computed as the ratio of the local intensity variation to the background of $I(x, y)$:

$$WI(x, y) = arctan\left(\alpha \sum_{i \in A} \sum_{j \varepsilon A} \frac{I(x, y) - I(x - i\Delta x, y - j\Delta y)}{I(x, y)}\right) \tag{34}$$

where $A = \{-1, 0, 1\}$ and α is a parameter for adjusting (magnifying or shrinking) the intensity differences among the neighboring pixels and an increase in α thickens the edges.

The Weber face is subjected to Local Directional Patterns (LDP) to find the significant edge directions. The resultant Weber-LDP images become the input in our face recognition system. The information set based features are extracted from these images and then PLPP is applied for the dimensionality reduction.

4.3. The Matching Algorithm

This algorithm comprises the following steps, shown in Fig. 1(b):

Step 1: Divide each training image into non overlapping windows of size $n \times n$.

Step 2: Extract one of the feature types from each window and form the training feature vector.

Step 3: Construct the weighted graph as in Step 1 of PLPP algorithm.

Step 4: Find the Eigenvectors using weighted graph, *WG*.

Step 5: Repeat Steps 1 and 2 on the test images.

Step 6: Project the test images using Eigenvectors found in Step 9 of PLPP algorithm.

Step 7: Use KNN algorithm to recognize the unknown test face.

5. Results of experimentation

The experiments are carried out on Intel core i5 processor with 2.67 GHz and 3GB of RAM. We have used MATLAB R2010a as our coding platform and PhD toolbox [24,25] for LDA. The matching algorithm is tested on three datasets. In order to ascertain the performance of the proposed face recognition method in terms of accuracy and the processing time, k-fold cross validation is done. The window size of 13×13 is selected by experimentation. The recognition accuracy is counted based on the correct classification of the unknown subjects in one run. The correct classification averaged over all runs gives the average accuracy.

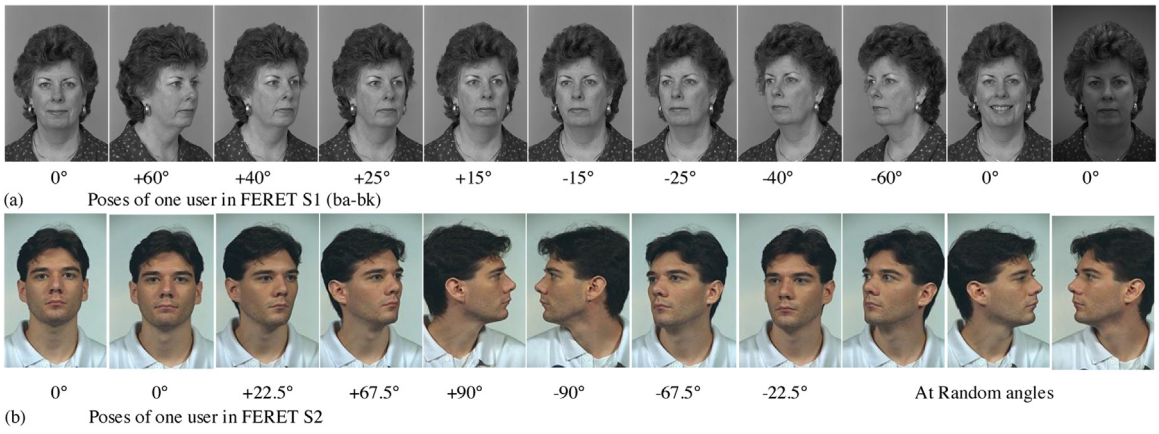

Fig. 2. Poses in FERET S1 and S2 Datasets.

5.1. The databases used

The two subsets of FERET database differ in the pose angles. The third database having pose variations is called Head Pose Image Database. The fourth database is called Extended Yale face database.

SET 1 (S1): One subset of FERET contains a total of 200 subjects with 11 gray images (subject numbers 19, 29, 268, 647, 700, 761, 1013–1206 of FERET gallery). Note that 6 out of these 200 subjects contain more than 11 images per subject. We have used only 11 images of which 9 poses (ba-bi) are used for comparing the proposed features with the existing features. The images of one subject with the pose angles are shown in Fig. 2(a). These images containing different frontal poses with varying illuminations are denoted by (ba-bk). The last two images of Fig. 2(a), frontal position with expression (bj) and frontal pose with illumination (bk) are not used in the experiments since we are concerned with only the pose variations.

The original size of the images is 256×384 pixels. The 8 poses of a subject are taken for the training set and one for the testing. The proposed algorithm is run 9 times for each pose during testing and the mean accuracy is computed.

SET 2 (S2): The other subset of FERET is a color dataset denoted by S2 where a total of 174 users each with 11 poses (the subject numbers include 562–739 of FERET gallery, some subjects are missing in the gallery so the total is 174).The poses vary from the frontal to the quarter left and to the right, half left to the right, profile left and from the right to some random angles. The various poses of one subject are shown in Fig. 2(b). The size of an image is 128×192 pixels.

SET 3 (S3): Head Pose Image Database (HPID) in [26] contains 15 subjects each with 93 images acquired in both vertical and horizontal directions. The original size of the images is 384×288 pixels. The vertical angles in the database include: +90°, +60°, +30°,+15°,0°,−15°,−30°,−60°,−90°.In this database, S3 a subject looks at the top and the bottom, and from right to left. The horizontal angles include: +90°, +75°,+60°, +45°,+30°,+15°,0°,−15°,−30°,−45°,−60°,−75°,−90°. Some poses of one subject of this database are shown in Fig. 3. The original size of the images is 384×288 pixels. Fig. 3 shows some samples of S3.

SET 4 (S4): Extended Yale Face Database B [27]: This database is illumination variant containing 28 subjects each with 9 poses under 65 illuminations.

Table 1
Classification Accuracy of the Proposed Features using EC.

Feature Type	S1 Accuracy	S2 Accuracy	S3 Accuracy
Nonlinear Shannon Transform	95.39%	84.95%	99.21%
Energy	94.89%	81.09%	99.36%
Sigmoid	**96.50%**	**86.31%**	**99.71%**
Effective Information	95.72%	80.93%	99.43%
Multi Quadratic	95.72%	80.72%	99.43%

Bold values represent the best results obtained on that Set/feature.

5.2. Results on databases S1-S3

Table 1 shows the accuracies obtained by applying the Euclidean classifier (EC) on the proposed features without using PLPP.

Table 2 gives the results of the combination of five feature types with PLPP on FERET S1, S2and S3. The time taken to extract the local features from the entire dataset is shown in Table 2. The setting of $\alpha = 0.8$ is found to give good results on three databases with three features (non linear Shannon transform, Energy, Effective information). But $\alpha = 1.6$ and $\alpha = 1.2$ are specifically suited to S1 database on sigmoid and multi quadratic features.

A comparison of Tables 1 and 2 shows that the combination the proposed features with PLPP improves the accuracy of results.

Table 3 gives the comparison of the earlier approaches on S1. In this SG features are tested on 10 poses (ba-bj) which lead to slightly better accuracy than that on 9 poses (ba-bi) for the same number of subjects. The 10th frontal pose with different expression is added to

Table 2
Combination of The Proposed Features with PLPP using EC.

Feature Type	S1		S2		S3	
	Accuracy	Feature Extraction Time (s)	Accuracy	Feature Extraction Time (s)	Accuracy	Feature Extraction Time (s)
Nonlinear Shannon (H)($\alpha = 0.8$)	95.89%	24.54	**93.26%**	18.83	**99.93%**	27.11
Energy (EN) ($\alpha = 0.8$)	97%	**22.82**	88.87%	**17.56**	**99.93%**	**24.81**
Sigmoid (SG) ($\alpha = 1.6$ for S1, $\alpha = 0.8$ for S2 & S3)	**97.44%**	24.97	89.60%	19.85	**99.93%**	27.62
Effective Information (EI) ($\alpha = 0.8$)	97%	31.93	84.12%	29.03	99.86%	34.68
Multi Quadratic (MQD) ($\alpha = 1.2$ for S1, $\alpha = 0.8$ for S2 & S3)	96.94%	30.71	84.38%	25.81	99.86%	38.53

Bold values represent the best results obtained on that Set/feature.

Fig. 3. Some poses of one subject of S3.

Table 3
Comparison of SG with The Earlier Approaches on S1.

Model	Number of subjects	Images per user (poses)	Accuracy
SG with PLPP	194	10(ba-bj)	**97.73%**
($\alpha = 1.6$)		9 (ba-bi)	97.25%
3D Morphable Model [2]	194	10(ba-bi,bk)	**95.9%**
SG with PLPP ($\alpha = 1.6$)		10(ba-bi,bk)	87.73%
Eigen Light Field [4]	200	9(ba-bi)	75%
SG with PLPP ($\alpha = 1.6$)		9(ba-bi)	**97.44%**

Bold values represent the best results obtained on that Set/feature.

both the training and the testing sets. It is compared with 3D morphable model with the poses (ba-bi, bk).The accuracy of SG features is reduced under illumination because the local characteristics of the image are affected due to changes in the illumination. However SG features fare well over Eigen Light Field features on the database of 200 subjects with an accuracy of 97.44%.

5.3. Comparison of results

Table 4 gives the comparison of the performance of proposed feature types with that of some holistic approaches. The comparison is made in terms of the percentage of correctly classified test samples. For each test sample, the Euclidean distance is calculated w.r.t. every training sample. The test sample is assigned to the class with the Euclidean minimum distance.

The highlighted values indicate the best recognition rates obtained. It can be seen from Table 4 that the mean processing times taken by nonlinear Shannon transform, MQD are the least. Next LPIC along with MQD on S1 and LPIC along SG on S2 and S3 take the second least time followed by PLPP. However LPIC features derived from the feature types (MQD & SG) have the reduced length. PLPP also reduces the time complexity by employing the pseudo-inverse matrix [14].

Though ICA gives a comparable accuracy on three databases, it is not suitable for the real time applications because it takes a lot of time to update its matrix. The mean processing times taken by ICA on S1 to S3: 1098.69 s≈ 18 min, 3341.1 s≈ 55 min and 3260 s ≈ 54 min approximately for one run are the maximum while the mean processing time of the proposed method is only 4.66 s for S1 and 3.29 s for S2, which are 200 and 1000 times smaller than the times taken by ICA for S1 and S2 respectively. The mean processing time taken by the proposed method is not even 0.5 s on S3.

Energy features take the least time to compute on all the three databases followed by the sigmoid function. The sigmoid and the non linear Shannon transform give the same accuracy on S3. The time taken to compute the sigmoid features is however more than that of non linear Shannon transform and its ROC is inferior to that of nonlinear Shannon transform as shown in Fig. 5(c).

The best accuracy is achieved by both nonlinear Shannon transform and SG features. The other methods such as ICA and PCA show higher accuracy on S1 and S2 but they need high computation time. On the other hand, LDA yields high accuracy on S3 because the number of subjects in S3 is very less and number of images per subject is very high as compared to those of S1 and S2. The processing time is the mean of the times taken for all 9, 11 or 93 runs corresponding to S1, S2 and S3 respectively. The processing timings of different methods are shown in Table 4. The non linear Shannon transform is the best while ICA is the worst in terms of the processing time.

5.4. Discussion of results

Table 1 gives the performance of the proposed features without PLPP on EC. Here nonlinear Shannon transform is somewhat inferior on S2. Table 2 demonstrates the consistent performance of the nonlinear Shannon transform features on three databases using PLPP.

Table 3 gives the comparison of the performance of the proposed approaches with that of earlier approaches. In Table 4, the non linear Shannon transform yields the best results of 93.26% and

Table 4
A Comparison of The Proposed Features with the Holistic Approaches.

Feature Types	S1		S2		S3	
	Accuracy	Mean Processing Time (s)	Accuracy	Mean Processing Time (s)	Accuracy	Mean Processing Time (s)
Nonlinear Shannon transform (α=0.8)	95.89%	4.66	**93.26%**	3.29	**99.93%**	0.3054
SG (α = 1.6 for S1, α = 0.8 for S2 &S3)	**97.44%**	4.75	89.60%	3.38	**99.93%**	0.3450
LPIC with MQD for S1& with SG for S2 and S3, for α = 1	94.06%	42.30	79.73%	50.57	98.64%	22.07
PCA	95%	296.73	69.80%	352.10	98.21%	224.48
ICA	96%	1098.69	86.52%	3341.1	91.11%	3260
LDA	78.28%	202.94	60.29%	287.90	**99.93%**	192.84
PLPP	92.56%	30.24	80.09%	33.41	87.38%	40.11

Bold values represent the best results obtained on that Set/feature.

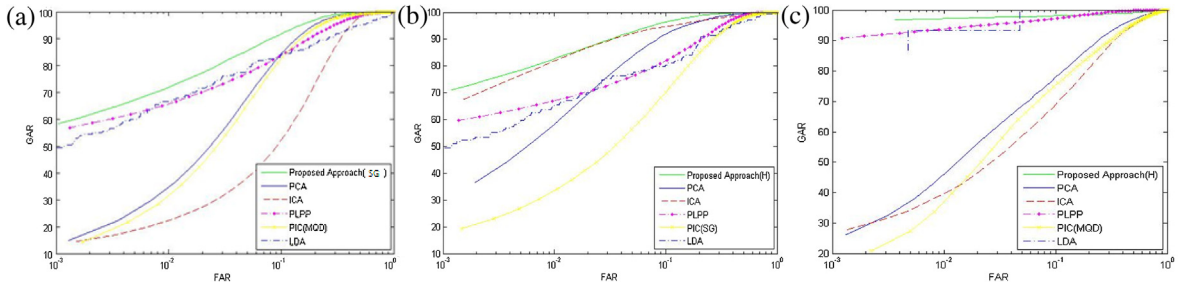

Fig. 4. ROC of various holistic methods on (a)S1, (b)S2 and (c)S3.

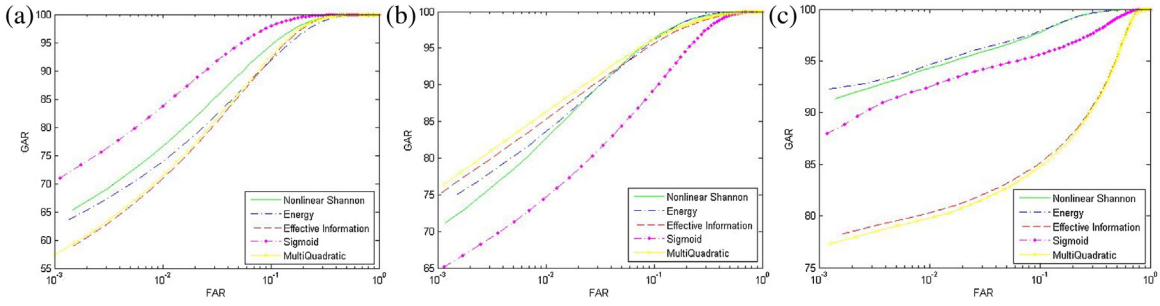

Fig. 5. ROC of the proposed features on (a)S1, (b)S2 and (c)S3.

99.93%on S2 and S3 respectively whereas SG features give the best result of 97.44% on S1. Thus the performance of Shannon transform is slightly inferior on S1 with GAR of 96.89%.

The Receiver Operating Characteristic (ROC) which is a graph between genuine acceptance rate (GAR) and false acceptance rate (FAR) is used to judge the performance of classifiers. The performance of the nonlinear Shannon transform is close to that of LDA on S3 where it displays an exceptional performance. LDA reports the least accuracy as can be noticed from its ROC because S1and S2 contain a large number of subjects while the nonlinear Shannon transform shows the best results on all three databases and takes lesser time than LDA.

A comparison of six holistic methods is given in Table 4 and shown in Fig. 4. ROCs of different features are shown in Fig. 5.The nonlinear Shannon transform (H) shows the best performance on S2 and S3 as in Fig. 5(b) and 5(c) with GARs of around 97.3% and 97.5% at FAR of 0.1% on S2 and S3 respectively. The sigmoid features show the best GAR of 98.1% at FAR of 0.1% followed by the non linear Shannon in Fig. 5(a).

5.5. Results on S4

The effectiveness of our approach is demonstrated on illumination variant (Extended Yale face) database S4. Only frontal pose under 65 illuminations is considered for the experiment. In this

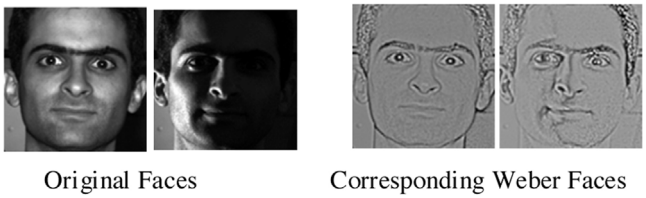

Fig. 6. Illumination invariant correction.

we convert the original into a Weber face. Let us see the effect of converting a face into a Weber face. Fig. 6 shows two images of a person from this database under different illuminations and the corresponding Weber faces. It is observed that the stark variation in the original face pair is reduced significantly in the Weber face pair, which is independent of the illumination component.

The results of our approach on Weber-LDP normalized images are given in Table 5. From this table it may be noted that non- liner Shannon transform yields good results with the complement of Exponential membership function whereas the best result of 100% comes from Hanman transform with the complement of Gaussian function.

We have also conducted tests on S1, S2 and S3 datasets to see the effect of changes in the resolution of images on the performance of the proposed features (Table 6). Two exactly the same images but

Table 5
Accuracy on S4.

Feature Types	Weber-LDP Normalized Images
Energy	99.89%
Sigmoid	99.89%
Effective Information	99.89%
MQD	99.89%
Non Linear Shannon	99.89%
HT with Exponential Its complement	99.62%, 97.53%
HT with Gaussian Its complement	82.75% (G_1) 100%(G_2)

Table 6
Accuracy of Resolution Invariant Approach.

Feature Types	S1	S2	S3
Non Linear Shannon	97.25%	93.68%	99.86%
Energy	96.68%	90.75%	99.93%
Sigmoid	97.59%	90.02%	99.93%
Effective Information	96.62%	82.08%	99.78%
MQD	96.45%	82.18%	99.71%

with different resolutions are partitioned into windows of different sizes such that the number of partitions is the same. In that case the features from the two images will differ by a scale parameter. Thus we can achieve resolution invariance by the proposed approach.

6. Conclusions

It is imperative to have a fast and robust method for face recognition as in surveillance applications. In this paper, the information-set based features are formulated to overcome the drawbacks of the prevailing holistic methods. The face recognition is attempted under pose, illumination and resolution variations by roping in Weber-LDP images for illumination invariance, PLPP for pose invariance in conjunction with the information set based features,

Combining PLPP with either the information set based features or with the non-linear Shannon transform based features improves the effectiveness and facilitates the dimensionality reduction. The results of the proposed features are better than

those obtained with other holistic methods in terms of both accuracy and processing time on S1-S3. The nonlinear Shannon transform features alone give the best results on S2 and S3 and consistent results on S1 whereas the multi quadratic features give the best results on S1.

ICA yields the second highest accuracy on S1 and S2 but is quite slow. The accuracies reported by LDA and the proposed features are the same on S3. LDA displays an exceptional performance on S3where number of images of each person is high

compared to the total number of persons. However, the time taken by LDA is very high as compared to that taken by the proposed features and its performance is less than that of ICA. The processing time of the proposed features is far less than that of ICA. The proposed features take from 3 to 4 s for processing while ICA takes 54–55 min on S1 and S2. The processing time is not even half a second on S3 while LDA takes 3 min. This makes the Shannon transform features suitable for all real time face recognition applications. Thus our method can be used in real time provided some improvements such as code optimization are included in the present work.

The proposed method takes care of variations in illumination by converting the images into Weber-LDP images. In this case Hanman transform with complement Gaussian membership function gives the best recognition accuracy on Extended Yale face database, S4.

One problem with the holistic methods is that they need a sufficient training data to train the model properly. The two methods: 3D Morphable model and Eigen light field [2,4] are able to recognize poses even if the training dataset is less while the proposed features work efficiently if the training data is sufficiently large.

The reasons for better performance of our approach are: i) Our approach is local in nature leading to reduction in computation and time, ii) The information sets based features derived from uncertainty representation are robust and effective, and iii) The combination of higher form of Information (Shannon & Hanman transforms) set based features and PLPP provides consistent performance as shown by cross validation testing on our various databases.

The focus of our future work is to recognize faces even if the training data is less and in the event of facial expressions.

We have not made the real time comparison as it requires video processing whereas our approach is an off-line approach. As a future work we will also extend our approach to operate on videos in real time.

References

[1] S. David Bolme, Elastic Bunch Graph Matching, Computer Science Department, Colorado State University, Summer, 2003 (Master's Thesis).
[2] V. Blanz, T. Vetter, Face recognition based on fitting a 3D morphable model, IEEE Trans. On Pattern Anal. Mach. Intell. 25 (September (9)) (2003) 1063–1074.
[3] S. Du, R.K. Ward, Face recognition under pose variation, J. Franklin Inst. 343 (September (6)) (2006) 596–613.
[4] R. Gross, I. Matthews, S. Baker, Appearance-based face recognition and light-fields, IEEE Trans. Pattern Anal. Mach. Intell. 26 (April (4)) (2004) 449–465.
[5] Y.W. Pang, N.H. Yu, et al., Face recognition using neighborhood preserving projections, Proceeding Of Pacific-Rim Conference on Multimedia 3768 (2005) 854–864.
[6] Wei Li, Qiuqi Ruan, Gaoyun An, Jun Wan, Discriminative uncorrelated neighborhood preserving projection for facial expression recognition, IEEE 11th International Conference on Signal Processing (ICSP), 21–25 Oct. Vol. 2 (2012) 801–805.
[7] Guoqiang Wang, Yunxing Shu, Dianting Liu, Yanling Shao, Fisher locality preserving projections for face recognition*, J. Comput. Inf. Syst. 9 (8) (2013) 2993–3000.
[8] M. Turk, A. Pentland, Eigenfaces for recognition, J. Cog. Neurosic. 3 (1) (1991) 71–86.
[9] M.A. Turk, A.P. Pentland, Face recognition using eigenfaces, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 3–6 June, Maui, Hawaii, USA, 1991, pp. 586–591.
[10] M.S. Bartlett, J.R. Movellan, T.J. Sejnowski, Face recognition by independent component analysis, IEEE Trans. Neural Netw. 13 (November (6)) (2002) 1450–1464.
[11] K. Etemad, R. Chellappa, Discriminant analysis for recognition of human face images, J. Opt. Soc. Am. A 14 (No. 8) (1997) 1724–1733.
[12] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. fisherfaces, recognition using class specific linear projection, in: Proc. 4th European Conference on Computer Vision, 15–18 April, Cambridge, UK, 1996, pp. 45–58.
[13] Xiaofei He, Shuicheng Yan, Yuxiao Hu, P. Niyogi, Hong-Jiang Zhang, Face recognition using Laplacian faces, IIEEE Trans. on Pattern Analysis and Machine Intelligence Vol. 27 (2005) 328–340.
[14] Rong-Hua Li, Zhiping Luo, Guoqiang Han, Pseudo-inverse locality preserving projections, 11–14 Dec. 2009, Beijing, China, in: International Conference on Computational Intelligence and Security, Vol. 1, 2005, pp. 363–367.
[15] Hyeon Bae, Sungshin Kim, Real time face detection and recognition using hybrid information extracted from face space and facial features, Image Vision Comput. 23 (Issue 13) (2005) 1181–1191.
[16] V. Karimi, A. Tashk, Age and gender estimation by using hybrid facial features, in: 20th International Telecommunication Forum (TELFOR2012), Belgrade, Serbia November 20–22, 2012.
[17] M. Hanmandlu, Information sets and information processing, Def. Sci. J. 61 (No. 5) (2011) 405–407.
[18] M. Hanmandlu, A. Das, Content-based image retrieval by information theoretic measure, Def. Sci. J. 61 (No. 5) (2011) 415–430.
[19] Mamta, M. Hanmandlu, Robust ear based authentication using local principal independent components, Expert Syst. Appl. 40 (November (16)) (2013) 6478–6490.
[20] Mamta, M. Hanmandlu, A new entropy function and a classifier for thermal face recognition, Eng. Appl. Artif. Intell. 36 (2014) 269–286.
[21] M. Hanmandlu, D. Jha, R. Sharma, Colorimage enhancement for fuzzy intensification, Pattern Recognit. Lett. 24 (3) (2003) 81–87.
[22] Jyotsana Grover, M. Hanmandlu, Hybrid fusion of score level and adaptive fuzzy decision level fusions for the finger-knuckle-print based authentication, Appl. Softcomput. 31 (2015) 1–13.

[23] Biao Wang, Weifeng Li, Wenming Yang, Qingmin Liao, Illumination normalization based on weber's law with application to face recognition, IEEE Signal Process. Lett. 18 (8) (2011) 462–465.
[24] V. ˇ Struc, N. Pavešiíc, The complete gabor-Fisher classifier for robust face recognition, EURASIP Adv. Signal Process. (2010) (pp. 26).
[25] V. ˇ Struc, N. Pavešiíc, Gabor-Based kernel partial-Least-Squares discrimination features for face recognition, Informatica (Vilnius) 20 (1) (2009) 115–138.
[26] N. Gourier, D. Hall, J.L. Crowley, Estimating face orientation from robust detection of salient facial features, in: Proceedings of Pointing 2004, ICPR, International Workshop on Visual Observation of Deictic Gestures, Cambridge, UK, 2004.
[27] A. Georghiades, P. Belhumeur, D. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2001) 643–660.
[28] R. Chellappa, C. Wilson, S. Sirohey, Human and machine recognition of faces: a survey, Proc. IEEE 83 (No. 5) (1995) 027–98.
[29] T. Fromherz, Face Recognition: A Summary of 1995–1997, Intl. Computer Science Inst. ICSI TR-98-027 Univ. of California, Berkeley, 1998.
[30] A. Samil, P. Iyengar, Automatic recognition and analysis of human faces and facial expressions: a Survey, Pattern Recogn. 25 (1992) 65–77.
[31] B.L. Jangid, K.K. Biswas, M. Hanmandlu, G. Chetty, Illumination invariant efficient face recognition using a single training image, Adelaide, Australia, in: Digital Image Computing Techniques and Applications (DICTA- 2015), 23–25 November, vol. 3, 2015, pp. 1–7.
[32] S.K. Jameel, Face recognition system using PCA and DCT in HMM, Int. J. Adv. Res. Comput. Communi. Eng. 4 (1) (2015) 13–18.
[33] J.A. Ojo, S.A. Adeniran, One-sample face recognition using HMM model of fiducial areas, Int. J. Image Process. (IJIP) 5 (1) (2011) 58–68.
[34] A.V. Nefian, M.H. Hayes, Hidden markov models for face recognition, acoustics, speech and signal processing, Seattle, WA, in: Proceedings of the 1998 IEEE International Conference on, vol. 5, 1998, pp. 2721–2724.
[35] Xiang Ma, Dan Schonfeld, Ashfaq Khokhar, Image segmentation and classification based on a 2D distributed hidden Markov model, Proc. SPIE 6822, Visual Communications and Image Processing 2008, January 28 (2008) 68221F.