The Visual Computer (2021) 37:143–159
https://doi.org/10.1007/s00371-019-01788-2

ORIGINAL ARTICLE

Multi-level uncorrelated discriminative shared Gaussian process for multi-view facial expression recognition

Sunil Kumar[1] · M. K. Bhuyan[2] · Yuji Iwahori[3]

Published online: 1 January 2020

Abstract
In multi-view facial expression recognition, discriminative shared Gaussian process latent variable model (DS-GPLVM) gives better performance than that of linear and nonlinear multi-view learning-based methods. However, Laplacian-based prior used in DS-GPLVM only captures topological structure of data space without considering the inter-class separability of the data, and hence the obtained latent space is suboptimal. So, we propose a multi-level uncorrelated DS-GPLVM (ML-UDSGPLVM) model which searches a common uncorrelated discriminative latent space learned from multiple observable spaces. A novel prior is proposed, which not only depends on the topological structure of the intra-class data, but also on the local-between-class-scatter-matrix of the data onto the latent manifold. The proposed approach employs an hierarchical framework, in which, expressions are first divided into three sub-categories. Subsequently, each of the sub-categories are further classified to identify the constituent basic expressions. Experimental results show that the proposed method outperforms state-of-the-art methods in many instances.

Keywords Facial expression recognition · Multi-view learning · Local binary pattern · Local fisher discriminant analysis

1 Introduction

Recognition of human's emotion through facial expressions has many important applications including behavior recognition, human–computer interaction, security, psychology etc. [1,2]. In reality, there are infinite number of expressions, but Ekman and Friesen [3] defined a set of basic expressions, i.e., happy, surprise, disgust, sad, anger, and fear. Several research works have been reported to recognize these basic expressions from the frontal view [4–6]. However, another important research direction is the recognition of emotions from multi-view and/or arbitrary-view face images.

✉ Sunil Kumar
snk@iiitm.ac.in

M. K. Bhuyan
mkb@iitg.ernet.in

Yuji Iwahori
iwahori@isc.chubu.ac.jp

[1] ABV-IIITM Gwalior, Gwalior 474015, India

[2] Department of Electronics and Electrical Engineering, IIT Guwahati, Guwahati 781039, India

[3] Department of Computer Science, Chubu University, Kasugai, Aichi 487-8501, Japan

The existing multi-view or view-invariant facial expression recognition methods can be broadly classified into three main categories, which is based on how they deal with head-pose variations and expressions in 2D facial images [7].

In the first category of multi-view facial expression recognition (FER) methods [8–11], a view-specific classifier is learned for each of the views during training. For recognition, head-pose is first estimated, and then, corresponding view-specific learned classifier is applied. However, one major limitation is that these methods do not consider the correlation between different views of expressions. Since separate classifiers are learned for different views, so the classification would be suboptimal.

The second category of methods mainly follow three-step procedures, (i.e., head-pose estimation, head-pose normalization, and FER from the frontal pose) to recognize facial expressions from any poses or a discrete set of poses. Rudovic et al. localize 39 facial points on each of the non-frontal/multi-view facial images, and then head-pose normalization is done [12–14]. During head-pose normalization, the mapping functions between a discrete set of non-frontal poses and the frontal pose are learned. They proposed coupled Gaussian process regression-based framework, which considers pair-wise correlation between the

views in order to estimate a robust mapping function. However, learning of mapping functions is performed on observation space, and so, error occurring in mapping functions adversely affects classification accuracy. This is even more severe when high-dimensional noise affects the normalized features. View- normalization or multi-view facial feature synthesis proposed in [15] uses block-based texture features. These features are extracted from different views of facial images to learn the mapping functions between any two views. This consideration ascertains that the features are extracted from several off-regions, on-regions, and on/off-regions of a face, and subsequently weights are assigned for these regions. However, several unwanted features may be added to the observation space due to wrong weight allocation policy. Moreover, major limitation of these approaches is that the head-pose normalization and learning of expression classifier are carried out independently, and hence affect the overall classification accuracy.

The third category of methods [7,16–18] has significant advantages as a single classifier is used for all the views. As a result, these approaches bypass the first step, i.e., the head-pose estimation for pose-specific classifier. In [7], it is considered that different views of facial expressions are just different manifestations of the same facial expression, and hence the correlations between different views of expressions are considered during training. In this view, discriminative shared Gaussian process latent variable model (DS-GPLVM) is proposed to learn a single nonlinear discriminative subspace. More specifically, DS-GPLVM generalizes the characteristics of discriminative-GPLVM (D-GPLVM) [19] along with the shared GPs [20,21] to learn the discriminative manifold. Nevertheless, discriminative nature of Gaussian process depends on a kind of prior. In [19], a prior based on linear discriminant analysis (LDA) [22] was proposed to replace the standard spherical Gaussian-based prior. A more general prior based on the notion of the graph Laplacian matrix was proposed in [23–25]. However, the affect of between-class separation was not considered in the prior, and hence the latent manifold obtained by this approach may not be optimal. The same Laplacian-based prior is further generalized for multi-view in [7]. Furthermore, correlations between the latent positions of the manifold may exist, which may further affect the classification accuracy of the DS-GPLVM-based FER system [26]. Another extension of [27] is proposed in [28]. The authors of [28] imposed a view-similarity constraint to ensure projections of correlated views close to each other. This method may help to recognize facial expressions from other views which are not used in training. Recently, subspace clustering for unlabeled data has also been proposed in [29]. This may be useful in grouping a large class of unlabeled spontaneous expressions.

In view of searching an optimal subspace, we propose to extend uncorrelated discriminative shared Gaussian process latent variable model (UDSGPLVM) [6] to the multi-level UDSGPLVM (ML-UDSGPLVM) for multi-view FER. In ML-UDSGPLVM, a more generalized discriminative prior is proposed, which is based on graph Laplacian matrix [23] and a transformation matrix. The transformation matrix is derived from the local-between-class-scatter-matrix (LBCSM) of data [30,31]. The advantage of the proposed prior is that it can better infer the separability of data onto the manifold. The proposed prior depends on both the intra-class geometric structure of the data captured by Laplacian matrix and the local inter-class variability of the data inferred by LBCSM. Hence, the proposed prior is more efficient than the Laplacian-based prior [7]. Moreover, discriminative nonlinear latent manifold (feature space) obtained by Gaussian process might be correlated, and thus classification performed directly on correlated manifold reduces classification accuracy [26]. In our proposed ML-UDSGPLVM approach, we first transform the correlated manifold to the uncorrelated manifold via a kernel approach.

To implement multi-level classification scheme, expressions of multi-view face images are recognized in two steps as shown in Fig. 1. In the first step, all the basic expressions are grouped into three categories, namely Lip-based, Lip–Eye-based, and Lip–Eye–Forehead-based expressions. This classification of expressions is done on the basis of regions of a face which mostly contribute to an expression. Then, category-wise training and testing are performed using the proposed UDSGPLVM. In the second step, a separate UDSGPLVM is applied on each of the sub-categories

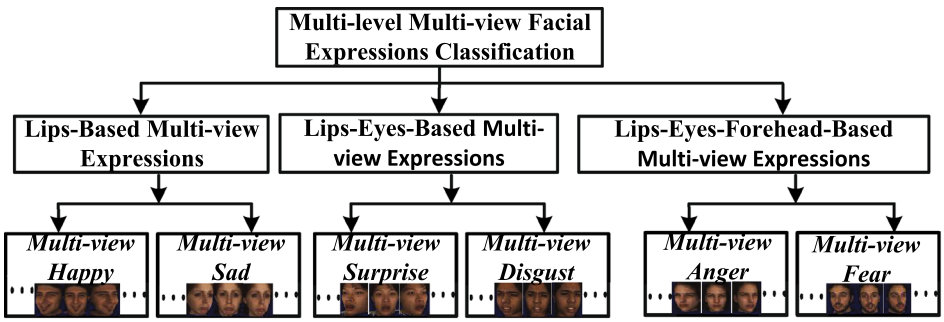


Fig. 1 Two-level separations of facial expressions. In first level, expressions are grouped into three categories namely Lips-based, Lips–Eyes-based and Lips–Eyes–Forehead-based. The second level shows the constituent basic expressions of each of the category expressions

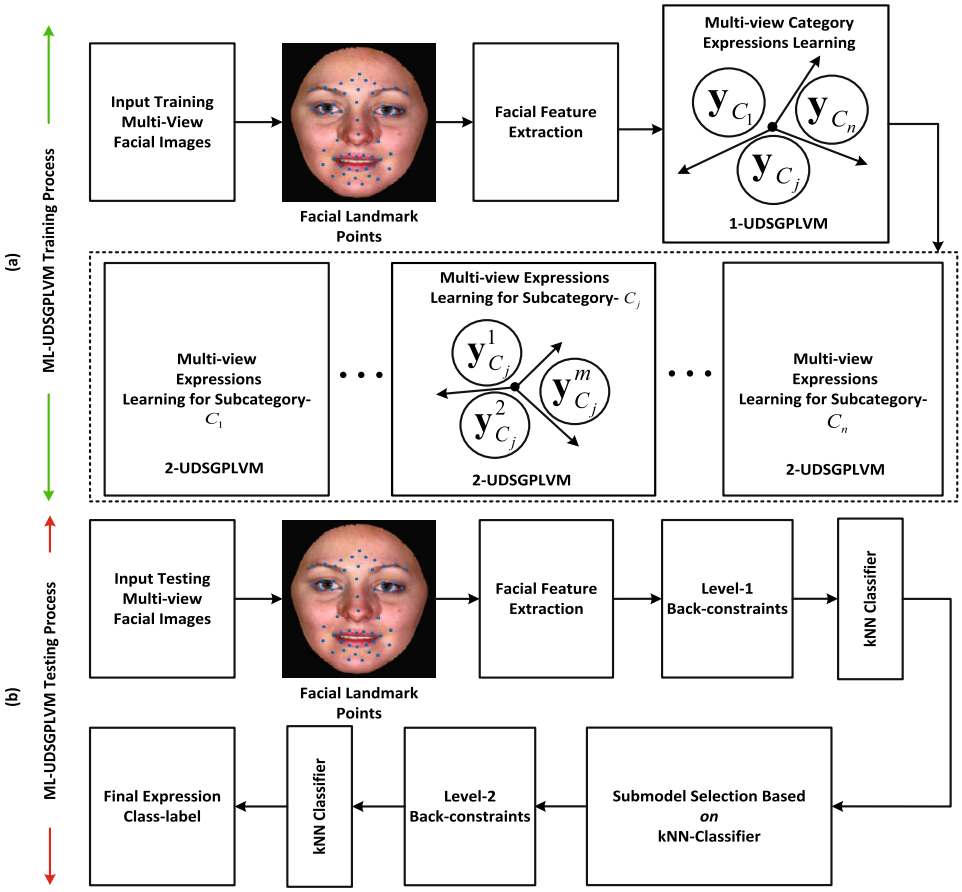


Fig. 2 Proposed multi-level DS-GPLVM for multi-view facial expression recognition: **a** training phase includes facial feature extraction and nonlinear dimensionality reduction using *l*-UDSGPLVM. 1-UDSGPLVM learns first level of discriminative features for group-level facial expression classification, and 2-UDSGPLVM comprises of distinct features for constituent expressions of the respective subgroups, **b** classification stages of the proposed scheme

to further classify the basic expressions embedded in the above-mentioned three category classes of expressions. The proposed 2-level UDSGPLVM follows this approach as against the method used in 1-level DS-GPLVM or simply DS-GPLVM.

In our proposed method, we employed our earlier developed face model [32] to extract features only from the informative regions of a face, as most discriminative features are only attainable from the informative/active regions of a face [5,33,34]. The proposed method is elaborately discussed in the following sections.

2 Proposed methodology

Shape-based method is employed in our proposed method to extract texture features from the active/informative regions of a face. We proposed to use our earlier developed face model, as it was derived from informative regions of a face [32]. Subsequently, LBP features are extracted from a 15 × 15 block around each of the facial points of our proposed face model. Next, expressions are divided into three classes based on the movements of lips, eyes, and forehead as stated in [6,35]. The corresponding reduced non-linear subspace is learned using 1-UDSGPLVM as shown in Fig. 2a. Subsequently, a 2-UDSGPLVM is learned for each of the expressions embedded in each of the sub-categories. Hence, three different 2-UDSGPLVMs have to be learned for final level of classification. The class-label of the test sample obtained by the first level of ML-UDSGPLVM and *k*NN *i.e.,* 1-UDSGPLVM+*k*NN is used to select a specific 2-UDSGPLVM out of three 2-UDSGPLVMs. So, first level of classification is performed using 1-UDSGPLVM and *k*NN. The first level of classification is basically a three-class classification problem, and hence the classifier identifies the appropriate sub-category. Any specific expression is finally identified by 2-UDSGPLVM and *k*NN. Our proposed ML-UDSGPLVM is discussed in the following section.

3 Proposed ML-UDSGPLVM

In our method, a more accurate low-dimensional manifold is derived for multi-view FER. We first give a brief overview of DS-GPLVM [7]. The impact of the state-of-the-art priors on latent manifold is analyzed, and then we proposed a new prior to nullify some of the limitations of the existing priors. Finally, we introduce our proposed ML-UDSGPLVM model

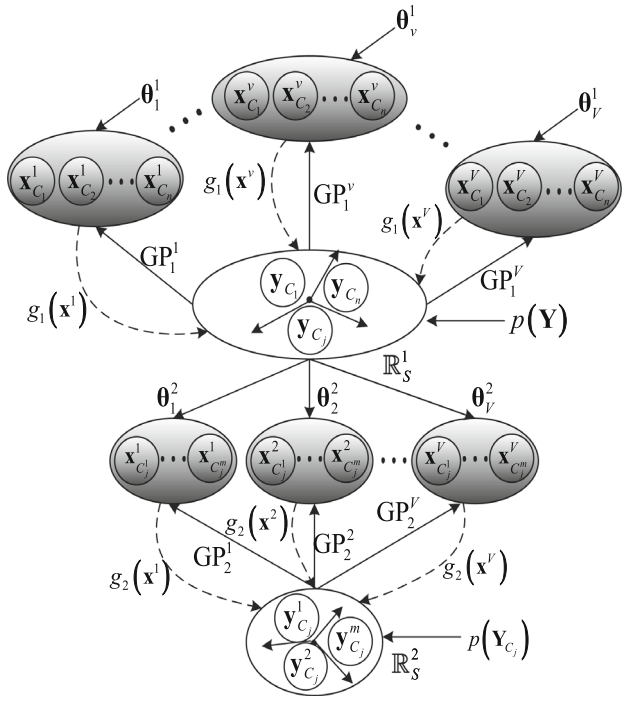


Fig. 3 Proposed ML-UDSGPLVM

as shown in Fig. 3, in which a more generalized discriminative prior is proposed. Also, uncorrelated constraint onto the latent manifold is imposed. All the steps of ML-UDSGPLVM are discussed below.

3.1 DS-GPLVM

The DS-GPLVM is a state-of-the-art approach in the field of multi-view FER [7]. More specifically, DS-GPLVM generalizes D-GPLVM [19] using the framework of shared GPs [20,21] to simultaneously learn a single nonlinear discriminative manifold of multiple observation spaces. The problem formulation of DS-GPLVM as a multi-view FER can be stated as follows:

Let $\mathbf{X} = \left\{\mathbf{X}^1, \mathbf{X}^2, \ldots, \mathbf{X}^V\right\}$ be the set of V observation spaces of size $VN \times D$, where N is the number of samples in each observation space and D is the dimension of each feature vector. Then, the objective of DS-GPLVM is to learn a single d-dimensional manifold $\mathbf{Y} \in \mathbb{R}^{N \times d}$ with $d << D$, which is assumed to be the shared information across all the views. The learning of low-dimensional manifold $\mathbf{Y}$ of DS-GPLVM and its mapping to the vth observation space $\mathbf{X}^v$ is modeled using the framework of shared GP. More specifically, it tries to learn the covariance function $k\left(\mathbf{y}_i, \mathbf{y}_j\right)$ of the shared manifold. In shared GP, each observation space is generated from the shared manifold via a separate Gaussian process, and hence the joint likelihood of the observed $\mathbf{X}$ given the shared manifold $\mathbf{Y}$ is factorized as follows:

$$p\left(\mathbf{X}|\mathbf{Y}, \boldsymbol{\theta}\right) = p\left(\mathbf{X}^1|\mathbf{Y}, \boldsymbol{\theta}_1\right) \cdots p\left(\mathbf{X}^V|\mathbf{Y}, \boldsymbol{\theta}_V\right) \tag{1}$$

where $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_V\}$ is the kernel parameters of the shared observation space. The vth factor of (1) represents likelihood of vth observation space $\mathbf{X}^v$ given the shared manifold $\mathbf{Y}$, i.e., $p\left(\mathbf{X}^v|\mathbf{Y}, \boldsymbol{\theta}_v\right)$, which is defined as:

$$p\left(\mathbf{X}^v|\mathbf{Y}, \boldsymbol{\theta}_v\right) = \frac{1}{\sqrt{(2\pi)^{ND}|\mathbf{K}_v|^D}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\mathbf{K}_v^{-1}\mathbf{X}^v\mathbf{X}^{v^{\mathrm{T}}}\right)\right\} \tag{2}$$

where $\mathbf{K}_v$ is the kernel covariance matrix associated with vth view of input space $\mathbf{X}^v$, whose (i, j)th element can be obtained using the covariance function $k\left(\mathbf{y}_i, \mathbf{y}_j\right)$ defined as the sum of the radial basis function (RBF) kernel, bias, and noise term. Hence, $k\left(\mathbf{y}_i, \mathbf{y}_j\right)$ can be represented as follows:

$$k\left(\mathbf{y}_i, \mathbf{y}_j\right) = \theta_{v1} \exp\left(-\frac{\theta_{v2}}{2}||\mathbf{y}_i - \mathbf{y}_j||^2\right) + \theta_{v3} + \frac{\delta_{i,j}}{\theta_{v4}} \tag{3}$$

where $\boldsymbol{\theta}_v = \{\theta_{v1}, \theta_{v2}, \theta_{v3}, \theta_{v4}\}$ are the kernel parameters of covariance function and $\delta_{i,j}$ is the Kronecker delta function. Finally, the distribution of shared manifold $\mathbf{Y}$ can be obtained by imposing a prior $p\left(\mathbf{Y}\right)$ over the shared manifold, and then applying the Bayes law. Thus, the posterior distribution of $\mathbf{Y}$ given $\mathbf{X}$ can be written as follows:

$$\begin{aligned} p\left(\mathbf{Y}, \boldsymbol{\theta}|\mathbf{X}\right) &= \frac{p\left(\mathbf{X}|\mathbf{Y}, \boldsymbol{\theta}\right) p\left(\mathbf{Y}\right)}{p\left(\mathbf{X}\right)} \\ &\propto p\left(\mathbf{X}|\mathbf{Y}, \boldsymbol{\theta}\right) p\left(\mathbf{Y}\right) \end{aligned} \tag{4}$$

The learning of the shared manifold is accomplished by minimizing the negative log-likelihood of the posterior distribution given in (4) with respect to the latent positions of the shared manifold $\mathbf{Y}$. The negative log-likelihood of (4) can be written as:

$$L_s = \sum_{v=1}^{V} L_v - \log\left(p\left(\mathbf{Y}\right)\right) \tag{5}$$

where L_v is given by:

$$L_v = \frac{D}{2}\ln|\mathbf{K}_v| + \frac{1}{2}\mathrm{tr}\left(\mathbf{K}_v^{-1}\mathbf{X}^v\mathbf{X}^{v^{\mathrm{T}}}\right) + \text{constant}. \tag{6}$$

3.2 Effect of priors on GPLVM

The effectiveness of GPLVM toward classification problem depends on the kind of prior for the manifold. In this direction, the first attempt was explored in [19], where a simple spherical Gaussian prior is replaced by a discriminative prior

based on LDA. Hence, it maximizes the between-class separability ($\mathbf{S}_\mathrm{b}$) and minimizes the within-class separability ($\mathbf{S}_\mathrm{w}$) of the latent space. The LDA-based prior is defined as:

$$p\left(\mathbf{Y}\right) = \frac{1}{Z_g}\exp\left\{-\frac{1}{\sigma_g}J^{-1}\left(\mathbf{Y}\right)\right\} \tag{7}$$

where $J\left(\mathbf{Y}\right) = \mathrm{tr}\left(\mathbf{S}_\mathrm{w}^{-1}\mathbf{S}_\mathrm{b}\right)$. In [24], a more general prior based on the notion of graph Laplacian matrix has been proposed. The Laplacian matrix of vth view is defined as:

$$\mathbf{L}^v = \mathbf{D}^v - \mathbf{W}^v$$

where $\mathbf{D}^v$ is a diagonal matrix with $\mathbf{D}^v_{ii} = \sum_j \mathbf{W}^v_{ij}$. The weight $\mathbf{W}^v_{ij}$ is defined as:

$$\mathbf{W}^v_{ij} = \begin{cases} \exp\left(-\frac{||\mathbf{x}^v_i - \mathbf{x}^v_j||^2}{\sigma^v}\right); & \text{if } c_i = c_j \\ 0; & \text{otherwise} \end{cases} \tag{8}$$

Also, [7] generalizes the Laplacian-based prior to obtain the prior for multi-view facial images. The net Laplacian matrix $\mathbf{L}_\mathrm{net}$ in [7] is obtained by summing all the normalize Laplacian matrices corresponding to each of the views. Hence, mathematically $\mathbf{L}_\mathrm{net}$ can be represented as:

$$\mathbf{L}_\mathrm{net} = \mathbf{L}^1_\mathrm{nor} + \mathbf{L}^2_\mathrm{nor} + \cdots + \mathbf{L}^V_\mathrm{nor} + \xi\mathbf{I} \tag{9}$$

where

$$\mathbf{L}^v_\mathrm{nor} = \left(\mathbf{D}^v\right)^{-1/2}\mathbf{L}^v\left(\mathbf{D}^v\right)^{-1/2}$$

Here, $\mathbf{I}$ indicates the identity matrix, and ξ is the regularization parameter which ensures positive-definiteness of $\mathbf{L}_\mathrm{net}$ [36]. Finally, the discriminative shared space prior is defined as:

$$p\left(\mathbf{Y}\right) = \prod_{v=1}^{V} p\left(\mathbf{Y}|\mathbf{X}^v\right)^{\frac{1}{V}} = \frac{1}{V\cdot Z_d}\exp\left\{-\frac{\beta}{2}\mathrm{tr}\left(\mathbf{Y}^\mathrm{T}\mathbf{L}_\mathrm{net}\mathbf{Y}\right)\right\} \tag{10}$$

where Z_d is a normalization constant and β (reciprocal of the variance) is the precision parameter.

3.3 ML-UDSGPLVM model

In the previous section, we introduced the impact of state-of-the-arts priors on the Gaussian processes. In this section, we derive a more generalized discriminative prior function. Also, influences of the prior function on the likelihood function are analyzed to obtain a more accurate posterior distribution. The prior based on Laplacian matrix ($\mathbf{L}_\mathrm{net}$) given in (10) essentially preserves the within-class geometric structure of the data. It uses RBF kernel to obtain weights between the data samples. So, it can also handle the multi-modalities present in the data. However, this approach did not consider the impact of between-class variability while defining the prior, and hence the prior proposed in (10) makes the GP suboptimal for classification. But, the impact of between-class-scatter matrix is crucial for all sorts of classification problems. So for our proposed prior, we incorporate a centering transformation matrix $\mathbf{B}$. This matrix is derived based on local-between-class-scatter-matrix ($\mathbf{S}_\mathrm{lb}$) as defined in [31]. Our proposed prior considers the joint impact of net Laplacian matrix ($\mathbf{L}_\mathrm{net}$) and the net $\mathbf{B}$, i.e., $\mathbf{B}_\mathrm{net}$ onto the shared manifold. The reason behind the use of local-between-class-scatter-matrix in the proposed method is that it can also handle the multi-model characteristics of the data. Mathematically, for vth view, $\mathbf{S}^v_\mathrm{lb}$, the LBSCM can be represented as follows:

$$\begin{aligned}\mathbf{S}^v_\mathrm{lb} &= \frac{1}{2}\sum_{i,j=1}^{N}\mathbf{W}^v_{lb,ij}\left(\mathbf{x}^v_i - \mathbf{x}^v_j\right)^\mathrm{T}\left(\mathbf{x}^v_i - \mathbf{x}^v_j\right) \\ &= \mathbf{X}^{v^\mathrm{T}}\mathbf{B}^v\mathbf{X}^v\end{aligned} \tag{11}$$

where $\mathbf{B}^v = \mathbf{D}^v_{lb,ii} - \mathbf{W}^v_{lb,ij}$ and $\mathbf{D}^v_{lb,ii} = \sum_j \mathbf{W}^v_{lb,ij}$. The term $\mathbf{W}^v_{lb,ij}$ is defined as follows [30,31]:

$$\mathbf{W}^v_{\mathrm{lb},ij} = \begin{cases} \left(\frac{1}{N} - \frac{1}{n^v_c}\right)\exp\left(-\frac{||\mathbf{x}^v_i - \mathbf{x}^v_j||^2}{\sigma^v_i\sigma^v_j}\right); & \text{if } c_i = c_j \\ \frac{1}{N}; & \text{otherwise} \end{cases} \tag{12}$$

The parameter n^v_c is the number of samples that belongs to cth-class in vth-view, and σ^v_i is the local scaling around $\mathbf{x}_i$ in vth view, which is defined as $\sigma^v_i = ||\mathbf{x}^v_i - \mathbf{x}^{vk}_i||_2$. The term $\mathbf{x}^{vk}_i$ is the k-nearest neighbor of $\mathbf{x}^v_i$. We use $k = 7$ in our proposed work [37]. Thus, the proposed regularized net-local-between-class-transformation matrix $\mathbf{B}_\mathrm{net}$ is defined as:

$$\mathbf{B}_\mathrm{net} = \mathbf{B}^1_\mathrm{nor} + \mathbf{B}^2_\mathrm{nor} + \cdots + \mathbf{B}^V_\mathrm{nor} + \xi\mathbf{I} = \sum_v \mathbf{B}^v_\mathrm{nor} + \xi\mathbf{I} \tag{13}$$

where

$$\mathbf{B}^v_\mathrm{nor} = \left(\mathbf{D}^v_{\mathrm{lb},ii}\right)^{-1/2}\mathbf{B}^v\left(\mathbf{D}^v_{\mathrm{lb},ii}\right)^{-1/2}$$

Finally, the proposed prior for ML-UDSGPLVM is defined as:

$$p\left(\mathbf{Y}\right) = \frac{1}{V\cdot Z_d}\exp\left\{-\frac{\beta}{2}\mathrm{tr}\left(\frac{\mathbf{Y}^\mathrm{T}\mathbf{L}_\mathrm{net}\mathbf{Y}}{\mathbf{Y}^\mathrm{T}\mathbf{B}_\mathrm{net}\mathbf{Y}}\right)\right\} \tag{14}$$

Hence, the proposed prior is more general and suitable for classification as compared to the earlier priors [7,19]. So,

class separation in the low-dimension manifold is being learned from the class separability of all the views. Additionally, it can also preserve the local structure of the data on the reduced manifold. Incorporating the proposed prior in (5), the proposed negative log-likelihood of ML-DSGPLVM is given by:

$$L_s = \sum_{v=1}^{V} L_v + \frac{\beta}{2}\mathrm{tr}\left(\frac{\mathbf{Y}^{\mathrm{T}}\mathbf{L}_{\mathrm{net}}\mathbf{Y}}{\mathbf{Y}^{\mathrm{T}}\mathbf{B}_{\mathrm{net}}\mathbf{Y}}\right) \tag{15}$$

where L_v is defined in (6). To obtain the optimal latent space, we need to find the derivative of (15) w.r.t $\mathbf{Y}$, which is given as:

$$\frac{\partial L_s}{\partial \mathbf{Y}} = \sum_{v=1}^{V} \frac{\partial L_v}{\partial \mathbf{Y}} + \frac{\beta}{2}\varphi\left(\mathbf{Y}\right) \tag{16}$$

where

$$\boldsymbol{\varphi}\left(\mathbf{Y}\right) = \left(\frac{2\mathbf{L}_{\mathrm{net}}\mathbf{Y}}{\mathbf{Y}^{\mathrm{T}}\mathbf{B}_{\mathrm{net}}\mathbf{Y}}\right) - \left(\frac{2\mathbf{B}_{\mathrm{net}}\mathbf{Y}}{\mathbf{Y}^{\mathrm{T}}\mathbf{B}_{\mathrm{net}}\mathbf{Y}}\right)\left(\frac{\mathbf{Y}^{\mathrm{T}}\mathbf{L}_{\mathrm{net}}\mathbf{Y}}{\mathbf{Y}^{\mathrm{T}}\mathbf{B}_{\mathrm{net}}\mathbf{Y}}\right) \tag{17}$$

As the GP follows an iterative procedure to find the optimal latent space, we need to evaluate $\boldsymbol{\varphi}\left(\mathbf{Y}\right)$ in each of the iterations which is computationally expensive. Also, latent states obtained by this approach is fluctuating, and hence convergence rate will be slower than that of LPP-based prior [7]. To overcome these limitations of our proposed method, the proposed prior is slightly modified:

$$p_{\mathrm{mod}}\left(\mathbf{Y}\right) = \frac{1}{V \cdot Z_d}\exp\left\{-\frac{\beta_1}{2}\mathrm{tr}\left(\mathbf{Y}^{\mathrm{T}}\mathbf{L}_{\mathrm{net}}\mathbf{Y}\right) + \frac{\beta_2}{2}\mathrm{tr}\left(\mathbf{Y}^{\mathrm{T}}\mathbf{B}_{\mathrm{net}}\mathbf{Y}\right)\right\} \tag{18}$$

The corresponding proposed negative log-likelihood and its derivative w.r.t. latent space $\mathbf{Y}$ can be reformulated as follows:

$$L_s^{\mathrm{mod}} = \sum_{v=1}^{V} L_v + \frac{\beta_1}{2}\mathrm{tr}\left(\mathbf{Y}^{\mathrm{T}}\mathbf{L}_{\mathrm{net}}\mathbf{Y}\right) - \frac{\beta_2}{2}\mathrm{tr}\left(\mathbf{Y}^{\mathrm{T}}\mathbf{B}_{\mathrm{net}}\mathbf{Y}\right) \tag{19}$$

$$\frac{\partial L_s^{\mathrm{mod}}}{\partial \mathbf{Y}} = \sum_{v=1}^{V} \frac{\partial L_v}{\partial \mathbf{Y}} + \left(\beta_1\mathbf{L}_{\mathrm{net}} - \beta_2\mathbf{B}_{\mathrm{net}}\right)\mathbf{Y} \tag{20}$$

This representation is simple, and also it allows smooth convergence of the latent space. This is due to the absence of denominator terms, which change the latent space abruptly. Hence, the proposed method is comparatively more suitable than the existing methods in terms of obtaining optimal latent subspace. This directly improves the recognition accuracy.

Moreover, test sample comes from the high-dimensional subspace that needs to be mapped onto the lower-dimensional latent manifold during the inference process of GPLVM. For this, back-constrain (learning of inverse mapping) has been defined such that the topology of data space is preserved in the latent manifold [38]. In [7], two kinds of back-constraints are defined for multi-views, namely independent back-projection (I_{bp}) and single back-projection (S_{bp}). For I_{bp}, separate inverse function is learned for each of the views, whereas for I_{bp}, a single inverse mapping function is learned from all the views to the shared space. They are defined as:

$$\mathbf{Y} = \begin{cases} \mathbf{K}_{ibc}^{v}\mathbf{A}_{ibc}^{v}; \forall v = 1, 2, \ldots, V : & \text{for } \mathrm{I}_{bc} \\ \left(\sum_{v=1}^{V} w_v\mathbf{K}_{bc}^{v}\right)\mathbf{A}_{sbc} = \mathbf{K}_{sbc}\mathbf{A}_{sbc} : & \text{for } \mathrm{S}_{bc} \end{cases} \tag{21}$$

where (i, j)th element of $\mathbf{K}_{ibc}^{v}$, i.e., $k_{bc}^{v}\left(\mathbf{x}_i^{v}, \mathbf{x}_j^{v}\right)$ which is given by:

$$k_{bc}^{v}\left(\mathbf{x}_i^{v}, \mathbf{x}_j^{v}\right) = \exp\left(-\frac{\gamma^{v}}{2}||\mathbf{x}_i^{v} - \mathbf{x}_j^{v}||^2\right) \tag{22}$$

$\mathbf{A}_{ibc}^{v}$ and $\mathbf{A}_{sbc}$ are the regression matrices and w_c is the weight corresponding to the vth view. Finally, these constraints are incorporated in the objective function (19), and then the minimization problem takes of the following form:

$$\begin{aligned} &\underset{\mathbf{Y},\theta_v,\mathbf{A}}{\arg\min}\, L_s^{\mathrm{mod}} + R\left(\mathbf{A}\right) \\ &\text{s.t.} \begin{cases} \mathbf{Y} - \mathbf{K}_{ibc}^{v}\mathbf{A}_{ibc}^{v} = \mathbf{0},\, v = 1, 2, \ldots, V \text{ for } \mathrm{I}_{bc} \\ \mathbf{Y} - \mathbf{K}_{sbc}\mathbf{A}_{sbc} = \mathbf{0},\, w_v \geq 0,\, \sum_v w_v = 1, \text{for } \mathrm{S}_{bc} \end{cases} \end{aligned} \tag{23}$$

where $R\left(\mathbf{A}\right)$ is a regularization term, which controls the over-fitting of the model to the data. An efficient way of solving this constraint optimization problem is given in [7], where the minimization problem is first divided into a set of sub-problems by employing alternative direction method (ADM) [39]. Next, an iterative approach (conjugate gradient algorithm) [40] is applied to solve each of the sub-problems separately with respect to their associated model parameters. We follow the same procedure to obtain the optimal latent manifold and other model parameters.

3.4 Uncorrelated latent space

In spite of using nonlinear-based approach to reduce the dimensionality of original feature space to the latent space, there may exist correlations between features. This may further affect the classification accuracy of the FER system. So in our proposed approach, instead of classifying directly from the correlated latent space, we first transform features of the shared space $\mathbf{Y}$ to the another shared space $\mathbf{Y}_{\mathbf{uc}}$, where features are uncorrelated. Then classification is performed. We obtain a nonlinear uncorrelated discriminative

manifold from the nonlinear correlated discriminative manifold (original latent manifold) via the transformation matrix $\boldsymbol{\chi} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_d]$. The columns of $\boldsymbol{\chi}$ are essentially the solutions (eigenvectors) of the following generalized eigenvalue equation corresponding to the first d lowest eigenvalues [26]:

$$\phi(\mathbf{Y})(\mathbf{L}_s + \mathbf{B}_s)\phi(\mathbf{Y})^{\mathrm{T}}\boldsymbol{v} = \lambda\phi(\mathbf{Y})\mathbf{G}\phi(\mathbf{Y})^{\mathrm{T}}\boldsymbol{v} \tag{24}$$

where $\phi(\mathbf{Y}) = \left[\phi(\mathbf{y}_1), \phi(\mathbf{y}_2), \ldots, \phi(\mathbf{y}_N)\right]$. $\mathbf{L}_s$ and $\mathbf{B}_s$ are the Laplacian and the local-between-class matrices, respectively [similar to (9], and (13)) obtained from the shared manifold. The matrix $\mathbf{G} = \mathbf{I} - (1/NV)\,\mathbf{e}\mathbf{e}^{\mathrm{T}}$, where $\mathbf{I}$ is an identity matrix and $\mathbf{e} = (1, 1, \ldots, 1)^{\mathrm{T}}$. Further, since eigenvectors of (24) should lie in the span of $\phi(\mathbf{y}_1), \phi(\mathbf{y}_2), \ldots, \phi(\mathbf{y}_N)$, there exists a vector $\boldsymbol{\alpha}_d$ such that $\boldsymbol{v}_d = \phi(\mathbf{Y})\boldsymbol{\alpha}_d$, where $\boldsymbol{\alpha}_d = \left[\alpha_1^d, \alpha_2^d, \ldots, \alpha_N^d\right]^{\mathrm{T}}$. Hence, for dth eigenvector, (24) can be rewritten in terms of $\boldsymbol{\alpha}_d$ as follows:

$$\phi(\mathbf{Y})(\mathbf{L}_s + \mathbf{B}_s)\phi(\mathbf{Y})^{\mathrm{T}}\phi(\mathbf{Y})\boldsymbol{\alpha}_d = \lambda\phi(\mathbf{Y})\mathbf{G}\phi(\mathbf{Y})^{\mathrm{T}}\phi(\mathbf{Y})\boldsymbol{\alpha}_d \tag{25}$$

Multiplying both side of (25) by $\phi(\mathbf{Y})^{\mathrm{T}}$ and by simple substitution, the following generalized eigenvalue equation is obtained:

$$\mathbf{M}(\mathbf{L}_s + \mathbf{B}_s)\mathbf{M}\boldsymbol{\alpha}_d = \lambda\mathbf{M}\mathbf{G}\mathbf{M}\boldsymbol{\alpha}_d \tag{26}$$

where $\mathbf{M} = \phi(\mathbf{Y})^{\mathrm{T}}\phi(\mathbf{Y})$ is the kernel matrix with $\mathbf{M}_{ij} = \exp\left(-||\mathbf{y}_i - \mathbf{y}_j||/\sigma\right)$. Let $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_d$ be the solutions of (26), then transformed uncorrelated nonlinear manifold can be obtained as follows:

$$\begin{aligned}\mathbf{Y}_{uc} &= [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_d]^{\mathrm{T}}\phi(\mathbf{Y}) \\ &= [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_d]^{\mathrm{T}}\phi(\mathbf{Y})^{\mathrm{T}}\phi(\mathbf{Y}) \\ &= [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_d]^{\mathrm{T}}\mathbf{M}\end{aligned} \tag{27}$$

Similarly, for a given new sample $\mathbf{y}^*$ of correlated manifold $\mathbf{Y}$, the corresponding position onto the uncorrelated manifold can be obtained using the following equation:

$$\mathbf{y}_{uc}^* = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_d]^{\mathrm{T}}[\mathbf{M}_{1*}, \mathbf{M}_{2*}, \ldots, \mathbf{M}_{N*}]^{\mathrm{T}} \tag{28}$$

where $\mathbf{M}_{k*} = \exp\left(-||\mathbf{y}_k - \mathbf{y}^*||/\sigma\right)$.

4 Experiments on BU3DFE dataset

The BU3DFE is a widely used dataset to evaluate the performance of multi-view and/or view-invariant FER methods. This database comprises of 3D facial images of Happy (HA), Surprise (SU), Fear (FE), Anger (AN), Disgust (DI), Sad (SA), and Neutral (NA) expressions. The database has 100 subjects, which includes 56% of female and 44% of male candidates. Also, expressions of BU3DFE dataset are captured at four different intensity levels ranging from onset/offset level to peak level of expression. As the database has 3D images, we first rendered the 3D face models using OpenGL to obtain the 2D textured facial images. 3D face model is first rotated by a user-defined angle, and then the corresponding 2D textured images are obtained. In our proposed approach, we obtained 2D facial images for seven views, i.e., −45°, − 30°, − 15°, 0°, 15°, 30°, and 45° yaw angles. A part of BU3DFE dataset is shown in Fig. 4, where a single subject is showing the happy expression for seven different viewing angles.

The validation of the proposed ML-UDSGPLVM algorithm is done on BU3DFE dataset. In our experiment, images from all the 100 subjects of BU3DFE dataset are employed. Also, expressions from all the intensity levels are considered for our experiment. So, altogether 1800 images per view, i.e., 1800 × 7 = 12,600 images are considered to evaluate the performance of our proposed method. Each view of the multi-view facial images comprises of six basic expressions, i.e., anger, disgust, fear, happy, sad, and surprise. For our experimentation, 300 images are taken for each of the expressions. The face part of 2D textured expressive images are manually cropped, and then down-sampled to get an image size of 160 × 140. Subsequently, the proposed 54 facial landmark points are localized. Localization of the facial points for the views − 45° and 45° are carried out manually, whereas images for the views (− 30°, − 15°, 0°, 15°, and 30°) are automatically annotated using active appearance model (AAM). Out of 54 landmark points, 5 stable points, i.e., left and right corners of the respective eyes, tip of the nose, and corners of the mouth are used to align the facial images using Procrustes analysis [41]. Finally, a grid of 15 × 15 is considered at each of the facial points to extract a feature vector from salient regions of a face. LBP^{u2} oper-

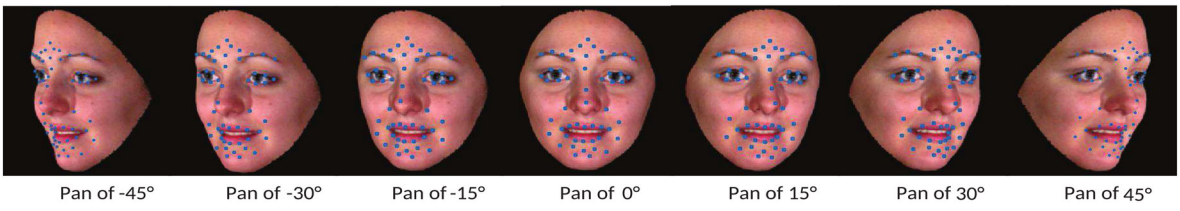


Fig. 4 An example of showing localization of landmark points (face model) on multi-view face images of BU3DFE dataset

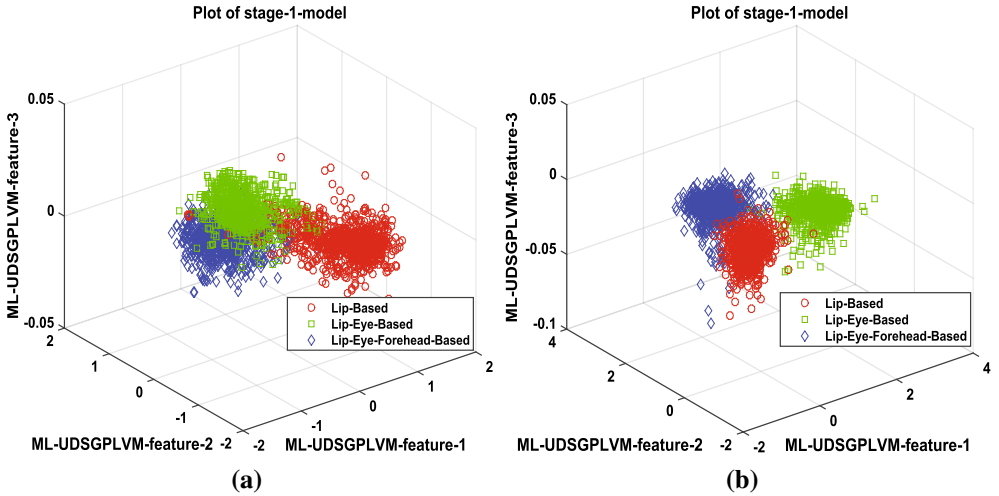


Fig. 5 3D distribution of test samples of three expressions: **a** trained by LBP, PCA, and ML-UDSGPLVM features, and **b** trained by LBP, LPP, and ML-UDSGPLVM features

ator is applied to each of the sub-blocks around each of the landmark points to obtain a feature vector. LBP^{u2} gives a 59-dimensional feature vector corresponding to each of the facial sub-regions, and hence the overall feature dimension for an image is $54 \times 59 = 3186$. The first level of dimensionality reduction in LBP-based appearance feature is performed using PCA, in which 95% of total variance of the data is preserved. As the features corresponding to the data (original feature space) are obtained for different views, so they may form altogether different clusters. Thus, the overall data space may be multi-modal. Hence, LPP-based dimensionality reduction approach would be more suitable in case of multi-view facial expression recognition. LPP-based dimensionality reduction technique is more capable in handling multi-modal data. In our proposed method, LPP-based approach is utilized to extract a set of discriminative features.

The experiments are carried out using 10-fold cross-validation strategy, and hence we first divide images of each of the views into 10 subsets. Out of which, 9 subsets are used to train the model, whereas the remaining set is used for testing. The experiments are repeated for 10 times such that testing subset is selected exactly ones in each iterations. Then, average accuracy is obtained for all the experiments. In all the experiments, we use 1-nearest neighbor (1-NN) classifier to evaluate the performance of the proposed method.

We used the same parameter settings as used in [7]. The parameters γ^v (back-projection parameter) are learned through leave-one-out cross-validation procedure. Finally, the optimum values of the two parameters, i.e., β (in our case β_1) and d (dimension of latent space) are found as $\beta = 300$ and $d = 5$. So, we used these parameter values to get optimal performance of our proposed ML-UDSGPLVM. The only extra parameter which is used in our proposed algorithm is the weight of the prior β_2, which controls the inter-class variance of the data onto the shared manifold. This parameter is learned experimentally by varying β_2 from 10 to 0.01, and found to be optimal at $\beta_2 = 0.8$.

The proposed ML-UDSGPLVM approach is a multi-level framework, where first level of proposed model, 1-UDSGPLVM is first trained by three sets of expression categories, i.e., Lip-based = {happy, sad}, Lip–Eye-based = {surprise, disgust}, and Lip–Eye–Forehead-based = {anger, fear}. Subsequently, a second level of Ml-UDSGPLVM, i.e., 2-UDSGPLVM is trained for the expressions and hence, three 2-UDSGPLVMs are trained in the second level of proposed ML-UDSGPLVM. Two different approaches, i.e., PCA and LPP are applied to reduce the dimensionality of LBP features. For this, both PCA and LPP are applied on 90% of the samples (i.e., 10-fold cross-validation strategy) of each of the views to obtain the principal directions, and subsequently those direction vectors are used to project both training and testing samples to the initial reduced subspace. In PCA, we reduce feature dimension in such a way that 95% variance of the data can be captured. In case of LPP, we restrict the feature set to 100-dimensional subspace. Finally, we apply our proposed ML-UDSGPLVM onto the reduced feature set to obtain a sufficiently lower-dimensional nonlinear discriminative subspace. Furthermore, features in the discriminative latent space may be correlated, and hence we perform another transformation on features of the correlated latent space. The first three components of ML-UDSGPLVM features are applied on two sets of features, i.e., LBP + PCA + ML-UDSGPLVM and LBP + LPP + ML-UDSGPLVM. The distribution of the test samples of all the views for these two cases is shown in Fig. 5a and b, respectively. These distribution plots show that first level of proposed ML-UDSGPLVM + LBP + LPP provides better separability than the combination of ML-UDSGPLVM + LBP + PCA [7]. The

Table 1 View-wise recognition rates (RR) for ML-UDSGPLVM on BU3DFE database

Ist-level of expression classes	Recognition rate (RR) (in %)							
	− 45°	− 30°	− 15°	0°	15°	30°	45°	Avg RR
Stage1 model evaluation using LBP + PCA + ML-UDSGPLVM features								
Lip-based	94.32	96.00	95.80	95.80	95.80	90.88	89.20	93.97
Lip–eye-based	96.08	92.70	95.00	98.30	91.70	92.53	95.84	94.59
Lip–eye–forehead-based	96.25	95.26	96.90	98.05	93.11	93.33	95.26	95.45
Average accuracy = 94.67%								
Stage1 model evaluation using LBP + LPP + ML-UDSGPLVM features								
Lip-based	98.02	98.30	98.20	99.20	99.01	99.20	97.00	97.30
Lip–eye-based	97.86	97.05	98.20	99.00	99.70	99.00	98.20	98.43
Lip–eye–forehead-based	98.50	98.20	99.00	99.00	98.30	98.30	98.20	98.50
Average accuracy = 98.07%								

view-wise average recognition rates for all the three types of expressions are shown in Table 1. From Table 1, it is clear that proposed LBP + LPP followed by ML-UDSGPLVM gives an improvement of about 4% as compared to LBP + PCA + ML-UDSGPLVM-based approach [7].

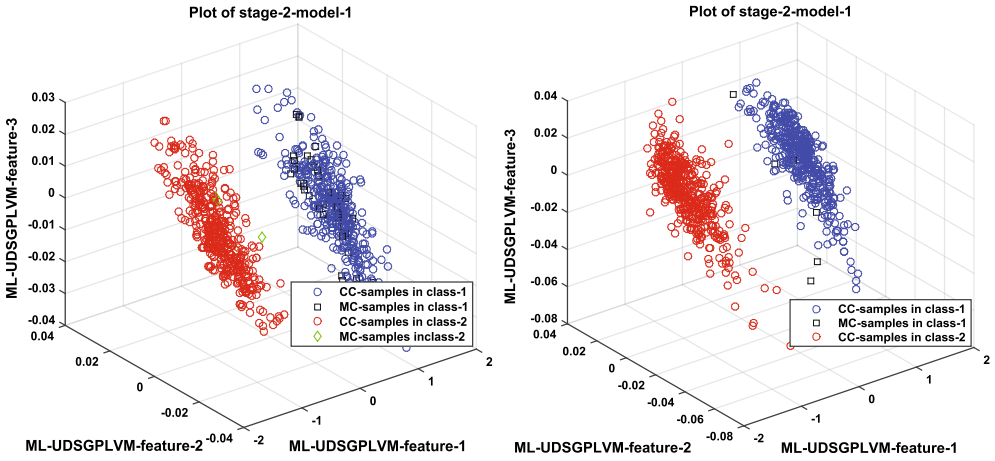
As discussed earlier, three-class problem is considered at the first stage of ML-UDSGPLVM. In the second stage, we need three 2-UDSGPLVM—one for each expression. Each 2-UDSGPLVM is trained using the same training samples of the respective expression class. For example, 2-UDSGPLVM corresponding to Lip-based expressions are trained using the samples of the respective sub-classes, i.e., happy and sad. Furthermore, the samples which were used for testing of 1-UDSGPLVM are again used for 2-UDSGPLVM. The samples which were misclassified in the first stage are tested by the respective 2-UDSGPLVM in the second level of ML-UDSGPLVM. So, misclassified samples of 1-UDSGPLVM (stage-1) and 2-UDSGPLVM (stage-2) are accounted for finding the overall misclassified samples. The misclassified samples are shown in Fig. 6. The overall view-wise classification accuracies for different basic expressions are shown in Table 2, and the corresponding distributions of test samples for two sets of features are shown in Fig. 6. It is even perceptually clear from the distribution plots that second level of ML-DSGPLVM, LBP + LPP provides better separability than that of LBP + PCA-based features, and the overall improvement of about 5%. Table 3 shows the recognition accuracy for different views, i.e.,(− 45°, − 30°, − 15°, 0°, 15°, 30°,and 45°) for the above-mentioned two feature sets. Table 4 shows the comparison of DS-GPLVM [7] and our proposed ML-UDSGPLVM. The performance of DS-GPLVM is evaluated by imposing it with LDA-based prior, LPP-based prior, and the prior proposed in (18). It is observed that the performance of DS-GPLVM with the proposed prior is better than LPP-based prior, and the improvement is even more significant (> 5%) than LDA-based prior. Our proposed ML-UDSGPLVM gives an overall average accuracy of 95.51%, which is about 3% better than the original DS-GPLVM (DS-GPLVM with LPP-based prior).

This significant improvement is due to the use of multi-level framework of uncorrelated DS-GPLVM. The proposed ML-UDSGPLVM on LBP + LPP-based feature gives the better performance as compared to DS-GPLVM.

Table 5 shows the comparison of several state-of-the-art multi-view learning-based methods [27,42–44] with the proposed ML-UDSGPLVM. In this, performance of MvDA is better than DS-GPLVM with LPP-based prior, and it is very close to DS-GPLVM with our proposed prior. Common spaces in all the multi-view-based linear approaches [27,42–44] were obtained by taking 98% of the total variance, which corresponds to 175 eigenvectors. This is relatively very high-dimensional common space than nonlinear DS-GPLVM latent space. Finally, our proposed ML-UDSGPLVM-based approach gives an overall improvement of about 2% than MvDA-based approach. In summary, the proposed ML-UDSGPLVM-based approach can efficiently find the low-dimensional discriminative shared manifold for multi-view FER.

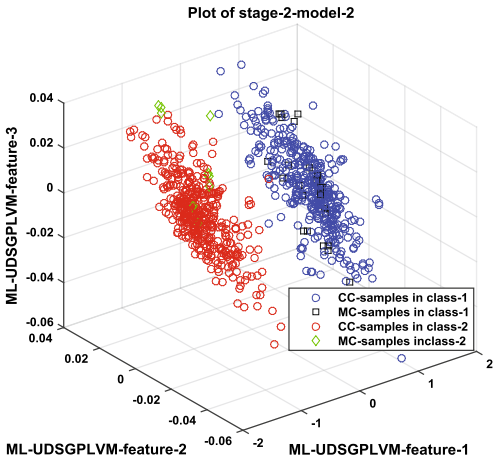
Experiments on KDEF dataset Images of BU3DFE dataset are synthetic, so we validate our proposed method by the images of multi-view KDEF dataset [45], and also by the dataset formed by combining images of both BU3DFE and KDEF datasets. Images of KDEF dataset are real and collected on controlled environment, whereas the combined dataset contains both synthetic and real images of facial expressions. The purpose of combining two datasets is to validate our proposed model on a large dataset.

Each of the expressions of KDEF dataset is captured from five different angles ranging from − 90° to 90° with an interval of 45°. In our experiment, expressions from three different views i.e., − 45°, 0°, and 45° are considered. In training, 2160 expressive images of 60 individuals are used, and remaining 360 facial images of 10 individuals are used for testing. The experimental results on KDEF dataset using

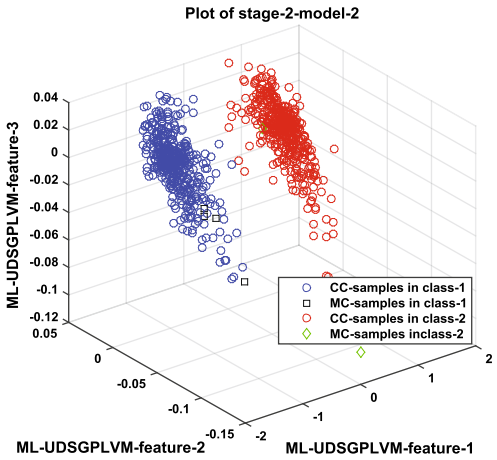


(a) Lip-based expressions

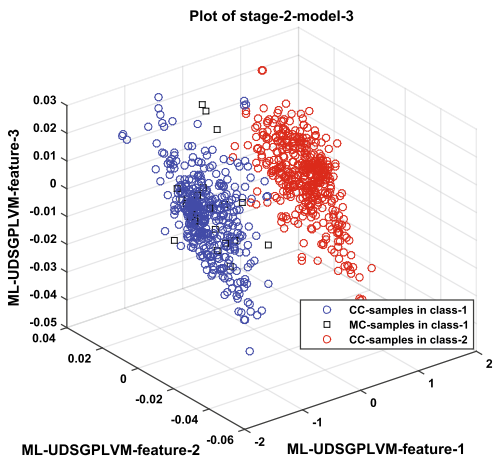
(b) Lip-based expressions



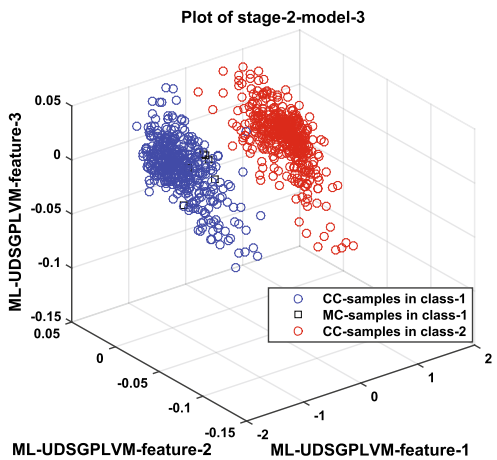
(c) Lip-Eye-based expressions



(d) Lip-Eye-based expressions



(e) Lip-Eye-Forehead-based expressions



(f) Lip-Eye-Forehead-based expressions

Fig. 6 3D distribution of test samples of the basic expressions. First these figures show the plot of test samples when 2-UDSGPLVM is applied on LBP followed by PCA, and the second column shows the distribution when 2-UDSGPLVM is applied on LBP followed by LPP-based features, respectively. CC and MC stand for correctly classified and miss-classified test samples

Table 2 View-wise expressions recognition rates (RR) for ML-UDSGPLVM on BU3DFE database

Model	Expressions	Recognition rate (RR) (in %)							
		− 45°	− 30°	− 15°	0°	15°	30°	45°	Avg RR
Stage2 model evaluation using LBP + PCA + ML-UDSGPLVM features									
Stage2-Model1	Happy	83.40	93.30	96.70	96.70	99.60	96.70	99.00	95.05
	Sad	93.30	73.40	76.70	96.70	80.00	90.00	83.40	84.78
Stage2-Model2	Surprise	99.10	90.00	96.70	98.00	93.30	99.80	93.40	95.75
	Disgust	93.30	93.40	96.70	96.70	83.30	99.20	93.30	93.70
Stage2-Model3	Anger	83.40	80.00	96.40	93.30	76.70	96.70	86.70	87.60
	Fear	80.00	73.30	76.70	96.70	83.30	90.00	90.00	84.28
Average accuracy = 90.20%									
Stage2 model evaluation using LBP + LPP + ML-UDSGPLVM features									
Stage2-Model1	Happy	99.80	99.50	99.50	99.80	83.40	99.60	99.20	97.25
	Sad	79.20	98.80	91.60	94.40	93.30	94.80	96.00	92.58
Stage2-Model2	Surprise	95.60	94.80	94.40	97.60	99.70	97.60	99.60	97.04
	Disgust	99.60	98.80	98.40	98.40	93.30	99.20	96.00	97.67
Stage2-Model3	Anger	95.60	96.00	98.40	96.40	83.40	96.40	99.60	95.11
	Fear	90.00	96.80	98.00	99.20	80.00	94.00	95.60	93.37
Average accuracy = 95.51%									

DS-GPLVM, ML-UDSGPLVM with and without proposed prior are shown in Table 6.

Experiment on BU3DFE + KDEF Combined datasets In this experiment, images from both the datasets are considered for training and testing. In training, 13230 facial images from BU3DFE dataset and 2160 facial images from KDEF dataset are used. The training images are captured from 130 subjects (70 from BU3DFE and 60 from KDEF) out of 170 subjects. A total of 6030 images from rest of the 40 (30 from BU3DFE and 10 from KDEF) subjects are used for testing. The experimental results on combined dataset are shown in Table 7.

Recently several deep-learning-based frameworks are proposed which give excellent performance in many Computer Vision applications [46,50–53]. However, one drawback of deep-learning framework is that it requires a large number of training samples, which may not be readily available in many applications. In case of limited training data, the performance of deep-learning-based approach is no longer superior to the DS-GPLVM-based method proposed in [7]. To be more fair, we did an experiment on our dataset by employing deep-learning framework as discussed in [46]. In our experiment, tenfold cross-validation strategy is employed. Hence, 11,340 samples out of 12,600 samples are used for training the deep neural network, and remaining 1260 samples are used for testing. The experiments are repeated for 10-times to calculate average accuracy. The average accuracies obtained by different deep-learning-based frameworks and our proposed method are shown in Table 8. It is observed that our proposed method can give better accuracy in contrast to convolution neural network (CNN),

Table 3 View-wise confusion matrices for six basic expressions

View	Anger	Disgust	Fear	Happy	Sad	Surprise
Confusion matrix obtained using LBP + PCA + ML-UDSGPLVM features						
Pan of − 45°						
Anger	83.4	0	0	0	13.3	3.3
Disgust	6.7	93.3	0	0	0	0
Fear	3.3	0	80	0	10	6.7
Happy	0	0	0	83.4	13.3	3.3
Sad	0	0	6.7	0	93.3	0
Surprise	0	0	0	0.9	0	99.1
Pan of − 30°						
Anger	80.0	0	0	0	6.7	13.3
Disgust	3.3	93.4	0	0	3.3	0
Fear	6.7	0	73.3	0	20	0
Happy	0	0	0	93.3	6.7	0
Sad	3.3	0	20	0	73.4	3.3
Surprise	0	0	0	0	10	90
Pan of − 15°						
Anger	96.4	0	3.0	0	0	0.6
Disgust	0	96.7	0	0	3.3	0
Fear	0	0	76.7	0	23.3	0
Happy	0	0	0	96.7	0	3.3
Sad	0	0	23.3	0	76.7	0
Surprise	0	0	0	0	3.3	96.7

Table 3 continued

View	Anger	Disgust	Fear	Happy	Sad	Surprise
Pan of 0°						
Anger	93.3	0	0	0	0	6.7
Disgust	0	96.7	0	0	0	3.3
Fear	0	0	96.7	0	3.3	0
Happy	0	0	0	96.7	3.3	0
Sad	0	0	3.3	0	96.7	0
Surprise	0	0	0	2	0	98
Pan of 15°						
Anger	76.7	0	0	0	10	13.3
Disgust	3.3	83.3	0	0	13.3	0
Fear	6.7	0	83.3	0	0	10
Happy	0	0	0	99.6	0	0.4
Sad	10	0	10	0	80	0
Surprise	0	0	0	0	6.7	93.3
Pan of 30°						
Anger	96.7	0	0	0	3.3	0
Disgust	0	99.2	0	0	0	0.8
Fear	0	0	90	0	10	0
Happy	0	0	0	96.7	3.3	0
Sad	0	0	10	0	90	0
Surprise	0	0	0	0.2	0	99.8
Pan of 45°						
Anger	86.7	0	0	0	13.3	0
Disgust	0	93.3	0	0	0	6.7
Fear	0	0	90	0	10	0
Happy	0	0	0	99	0	1
Sad	3.3	0	13.3	0	83.4	0
Surprise	0	3.3	0	0	3.3	93.4
Confusion matrix obtained using LBP + LPP + ML-UDSGPLVM features						
Pan of − 45°						
Anger	95.6	0	0	4.0	0	0.4
Disgust	0	99.6	0	0	0.4	0
Fear	0.4	0	90.0	0	9.6	0
Happy	0	0	0	99.8	0	0.2
Sad	0	0	20.4	0	79.2	0.4
Surprise	0	4.4	0	0	0	95.6
Pan of − 30°						
Anger	96.0	0	0	1.2	2.4	0.4
Disgust	0.8	98.8	0	0	0	0.4
Fear	0.8	0	96.8	0	1.2	1.2
Happy	0	0	0	99.5	0	0.5
Sad	0	0	1.2	0	98.8	0
Surprise	0	4.4	0	0	0.8	94.8

Table 3 continued

View	Anger	Disgust	Fear	Happy	Sad	Surprise
Pan of − 15°						
Anger	98.4	0	0	0.4	0.4	0.8
Disgust	0.4	98.4	0	0	1.2	0
Fear	0.8	0	98.0	0	1.2	0
Happy	0	0	0	99.5	0	0.5
Sad	0.4	0	8.0	0	91.6	0
Surprise	0	5.6	0	0	0	94.4
Pan of 0°						
Anger	96.4	0	0	1.2	0.8	1.6
Disgust	0.4	98.4	0	0	1.2	0
Fear	0	0	99.2	0	0.8	0
Happy	0	0	0	99.8	0.2	0
Sad	0	0	5.2	0	94.4	0.4
Surprise	0.4	2.0	0	0	0	97.6
Pan of 15°						
Anger	83.4	0	0	0	13.3	3.3
Disgust	6.7	93.3	0	0	0	0
Fear	3.3	0	80	0	10	6.7
Happy	0	0	0	83.4	13.3	3.3
Sad	0	0	6.7	0	93.3	0
Surprise	0	0	0.3	0	0	99.7
Pan of 30°						
Anger	96.4	0	0	0.8	1.2	1.6
Disgust	0.4	99.2	0	0	0.4	0
Fear	1.2	0	94.0	0	4.0	0.8
Happy	0	0	0	99.6	0.4	0
Sad	0.4	0	4.4	0	94.8	0.4
Surprise	0	1.6	0	0	0.8	97.6
Pan of 45°						
Anger	99.6	0	0	0	0.4	0
Disgust	0.4	96.0	0	0	0.4	3.2
Fear	0.4	0	95.6	0	4.0	0
Happy	0	0	0	99.2	0.4	0.4
Sad	0	0	4.0	0	96.0	0
Surprise	0	0.4	0	0	0	99.6

deep belief network (DBN), and a special DNN-based structure proposed in [46]. In fact, performance of DNN-based approaches (trained on limited dataset) is very much similar to view-wise multi-view FER methods. Hence for limited training dataset, DNN-based methods are not much benefited from the samples of additional views.

Table 4 Comparison of proposed method with the state-of-the-art DS-GPLVM-based methods on BU3DFE dataset in terms of average recognition rates with average standard deviation

Methods	Recognition rate (RR) (in %)							
	− 45°	− 30°	− 15°	0°	15°	30°	45°	Avg RR
LBP + PCA + Shared features								
DS-GPLVM with LDA-based prior	81.04	76.20	74.79	83.87	81.04	79.23	78.83	79.29 ± 0.027
DS-GPLVM with [25]	84.56	79.36	85.08	93.98	82.20	80.55	77.54	83.32 ± 0.015
DS-GPLVM [7]	90.92	85.88	85.68	93.95	81.04	87.50	77.01	86.00 ± 0.021
DS-GPLVM with proposed prior	90.32	83.87	84.07	95.16	85.08	85.28	83.46	86.75 ± 0.021
ML-UDSGPLVM without proposed prior	87.57	86.98	91.66	92.65	90.08	85.28	82.46	88.10 ± 0.065
ML-UDSGPLVM with proposed prior	88.75	83.90	89.98	96.35	86.03	95.40	90.96	90.19 ± 0.011
LBP + LPP + Shared features								
DS-GPLVM with LDA-based prior	96.97	91.53	94.95	91.73	93.75	87.90	84.87	91.67 ± 0.025
DS-GPLVM with [25]	96.31	91.03	91.50	97.32	92.12	89.46	86.72	92.06 ± 0.015
DS-GPLVM [7]	96.37	90.12	91.33	97.37	92.74	91.53	88.50	92.56 ± 0.015
DS-GPLVM with proposed prior	95.86	92.13	94.15	96.87	95.86	91.73	89.61	93.75 ± 0.015
ML-UDSGPLVM without proposed prior	94.32	95.44	95.75	96.02	95.16	94.75	92.75	94.85 ± 0.017
ML-UDSGPLVM with proposed prior	93.30	97.45	96.71	97.63	88.85	96.93	97.66	95.51 ± 0.014

Table 5 Comparison of proposed method with the state-of-the-arts methods on BU3DFE dataset

State-of-the-art-methods										Proposed method
GMPCA	GMLDA	GMLPP	GMCCA	PW-CCA	MCCA	MvDA	D-GPLVM	DS-GPLVM		ML-UDSGPLVM
89.64	91.19	92.03	91.91	84.28	89.32	93.48	88.33	92.56		**95.51**

Table 6 Performance of ML-UDSGPLVM with and without proposed prior on KDEF dataset

Methods	Recognition rates (RR) (in %)			
	− 45°	0°	45°	Avg RR
DS-GPLVM	76.08	90.45	78.25	81.59 ± 0.018
ML-UDSGPLVM without proposed prior	79.88	96.44	84.06	86.79 ± 0.027
ML-UDSGPLVM with proposed prior	84.95	97.02	88.12	90.95 ± 0.022

Evaluation is done in terms of average recognition rates and average standard deviation

Table 7 Performance of different methods on combined datasets (BU3DFE + KDEF datasets)

Methods	Recognition rates (RR) (in %)							
	− 45°	− 30°	− 15°	0°	15°	30°	45°	Avg RR
DS-GPLVM	90.86	93.23	95.36	93.53	93.66	90.20	86.88	91.96 ± 0.021
ML-UDSGPLVM without proposed prior	94.41	95.42	95.70	96.10	95.16	94.75	92.74	94.89 ± 0.016
ML-UDSGPLVM with proposed prior	94.40	96.75	97.30	96.70	96.10	95.91	94.30	95.92 ± 0.013

Table 8 Comparison of proposed method with deep neural network (DNN)-based approaches. The models are trained and tested on BU3DFE dataset

Methods	Recognition Rate (RR) (in %)							
	− 45°	− 30°	− 15°	0°	15°	30°	45°	Avg RR
DNN-driven feature learning [46]	86.00	87.80	91.20	91.00	88.70	82.60	80.20	86.78 ± 0.022
DBN (2 hidden layers)	76.40	79.60	81.01	82.40	81.80	76.50	70.20	78.28 ± 0.028
CNN (3 convolution layers)	82.54	86.50	89.20	89.40	85.20	81.80	76.60	84.46 ± 0.025
Proposed method	93.30	97.45	96.71	97.63	88.85	96.93	97.66	95.51 ± 0.014

Table 9 Cross-data classification accuracy (in %) of the proposed method versus state-of-the-arts methods

Training dataset	BU3DFE dataset (View-Set-A)		KDEF dataset (View-Set-B)		BU3DFE + KDEF datasets (View-Set-A)	
Methods	Testing on BU3DFE	Testing on KDEF	Testing on BU3DFE	Testing on KDEF	Testing on BU3DFE	Testing on KDEF
MvDA	93.41	92.30	86.55	84.35	93.42	94.63
D-GPLVM	88.28	90.22	81.25	76.53	88.27	90.12
DS-GPLVM	92.86	91.34	88.33	81.59	92.86	92.68
MPCNN [47]	94.86	93.88	88.76	87.54	94.84	94.87
RGL [48]	93.35	89.21	90.36	**92.12**	93.36	94.85
MvGAN [49]	95.01	**96.76**	82.42	81.88	95.22	95.88
ML-UDSGPLVM	**95.51**	96.75	**91.04**	90.95	**96.10**	**95.91**

Bold face numerics indicate the best classification performance obtained by respective methods and the proposed method
View-Set-A = {−45°, −30°, −15°, 0°, 15°, 35°, 45°}
View-Set-B = {−45°, 0°, 45°}

Cross-data performance comparisons of the proposed method with the state-of-the-art methods are shown in Table 9. In this comparative analysis, we trained different models including our proposed model on BU3DFE dataset and/or KDEF dataset. Subsequently, on each of the trained models, testing accuracies for both the dataset taken one at a time are calculated. It can be observed from Table 9 that in most of the cases, the performance of the ML-UDSGPLVM is better than other state-of-the-art models.

5 Conclusion

In this paper, a multi-level framework of uncorrelated discriminative shared Gaussian process latent variable model ML-UDSGPLVM is proposed to obtain a single nonlinear uncorrelated discriminative shared manifold. More specifically, we proposed a novel prior with the help of Laplacian matrix and the local-between-class-scatter-matrix. The reason behind the use of between-class-separability matrix is that it can handle the multi-modal characteristics of multi-view data similar to Laplacian matrix. In our proposed ML-UDSGPLVM, instead of classifying a test sample directly on correlated shared space, we transform it to a nonlinear uncorrelated latent space, and then 1-NN classifier is used. Also, the proposed approach is multi-level framework—the expressions are first divided into three basic categories, i.e., expressions by only Lip, expressions by Lips–Eyes, and expressions by Lips–Eyes–Forehead, which are recognized by first level of ML-UDSGPLVM (1-UDSGPLVM). Subsequently, a separate second level of ML-UDSGPLVM (2-UDSGPLVM) is learned for each of the sub-classes. So, three 2-UDSGPLVMs have to be learned to reach final classification level. Expressions are first classified on 1-UDSGPLVM manifold, and the corresponding 2-UDSGPLVM manifold is used for final level of classification. This multi-level decision strategy inherently improves the recognition accuracy. The performance of our proposed ML-DSGPLVM is evaluated for six basic expressions obtained from seven different poses ($-45°$, $-30°$, $-15°$, $0°$, $15°$, $30°$, and $45°$) of BU3DFE dataset. ML-UDSGPLVM approach gives an average recognition rate of 95.51% with LBP + LPP-based features. So, our proposed scheme outperforms the state-of-the-art linear and nonlinear-based multi-view learning techniques.

Computational complexity is one of the major drawbacks of GPLVM. However, advantage of GPLVM is that it efficiently models data in a nonlinear low-dimensional subspace. In our proposed approach, we handle complexity issue of GPLVM for multi-view FER by decomposing the minimization problem defined in Eq. (23) into number of sub-problems, and subsequently, conjugate gradient algorithm is used to compute model parameters associated with each of the sub-problems. This approach makes the problem

tractable even if the number of views is increased. Computational complexity of our proposed algorithm can further be reduced by incorporating sparse approximation to full Gaussian process [54]. This process reduces original complexity of GPLVM, i.e., $O(N^3)$ to $O(k^2N)$, where k is the number of points retained in the sparse representation. Hence, the proposed method can be extended to recognize expressions from many views.

References

1. Bettadapura, V.: Face expression recognition and analysis: the state of the art (2012). arXiv preprint arXiv:1203.6722
2. Yan, J., Zheng, W., Xu, Q., Lu, G., Li, H., Wang, B.: Sparse kernel reduced-rank regression for bimodal emotion recognition from facial expression and speech. IEEE Trans. Multimed. **18**(7), 1319–1329 (2016)
3. Ekman, P., Friesen, W.V., Press, C.P.: Pictures of Facial Affect. Consulting Psychologists Press, Mountain View (1975)
4. Tie, Y., Guan, L.: A deformable 3-d facial expression model for dynamic human emotional state recognition. IEEE Trans. Circuits Syst. Video Technol. **23**(1), 142–157 (2013)
5. Kumar, S., Bhuyan, M., Chakraborty, B.K.: Extraction of informative regions of a face for facial expression recognition. IET Comput. Vis. **10**(6), 567–576 (2016)
6. Siddiqi, M.H., Ali, R., Khan, A.M., Park, Y.-T., Lee, S.: Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields. IEEE Trans. Image Process. **24**(4), 1386–1398 (2015)
7. Eleftheriadis, S., Rudovic, O., Pantic, M.: Discriminative shared Gaussian processes for multiview and view-invariant facial expression recognition. IEEE Trans. Image Process. **24**(1), 189–204 (2015)
8. Moore, S., Bowden, R.: Local binary patterns for multi-view facial expression recognition. Comput. Vis. Image Underst. **115**(4), 541–558 (2011)
9. Hu, Y., Zeng, Z., Yin, L., Wei, X., Tu, J., Huang, T.S.: A study of non-frontal-view facial expressions recognition. In: 19th International Conference on Pattern Recognition, 2008. ICPR 2008, pp. 1–4. IEEE (2008)
10. Hu, Y., Zeng, Z., Yin, L., Wei, X., Zhou, X., Huang, T.S.: Multi-view facial expression recognition. In: 8th IEEE International Conference on Automatic Face Gesture Recognition, 2008. FG '08, pp. 1–6 (2008)
11. Hesse, N., Gehrig, T., Gao, H., Ekenel, H.K.: Multi-view facial expression recognition using local appearance features. In: 2012 21st International Conference on Pattern Recognition (ICPR), pp. 3533–3536. IEEE (2012)
12. Rudovic, O., Pantic, M., Patras, I.: Coupled Gaussian processes for pose-invariant facial expression recognition. IEEE Trans. Pattern Anal. Mach. Intell. **35**(6), 1357–1369 (2013)
13. Rudovic, O., Patras, I., Pantic, M.: Coupled Gaussian process regression for pose-invariant facial expression recognition. In: Computer Vision–ECCV 2010, pp. 350–363. Springer (2010)
14. Rudovic, O., Patras, I., Pantic, M.: Regression-based multi-view facial expression recognition. In: 2010 20th International Conference on Pattern Recognition (ICPR), pp. 4121–4124, IEEE (2010)
15. Zheng, W.: Multi-view facial expression recognition based on group sparse reduced-rank regression. IEEE Trans. Affect. Comput. **5**(1), 71–85 (2014)
16. Tariq, U., Yang, J., Huang, T.S.: Multi-view facial expression recognition analysis with generic sparse coding feature. In: Computer Vision–ECCV 2012. Workshops and Demonstrations, pp. 578–588, Springer (2012)
17. Zheng, W., Tang, H., Lin, Z., Huang, T.S.: Emotion recognition from arbitrary view facial images. In: Computer Vision–ECCV 2010, pp. 490–503, Springer (2010)
18. Eleftheriadis, S., Rudovic, O., Pantic, M.: View-constrained latent variable model for multi-view facial expression classification. In: International Symposium on Visual Computing, pp. 292–303. Springer (2014)
19. Urtasun, R., Darrell, T.: Discriminative Gaussian process latent variable model for classification. In: Proceedings of the 24th International Conference on Machine Learning, pp. 927–934, ACM (2007)
20. Shon, A., Grochow, K., Hertzmann, A., Rao, R.P.: Learning shared latent structure for image synthesis and robotic imitation. In: Advances in Neural Information Processing Systems, pp. 1233–1240 (2005)
21. Ek, C.H., Lawrence, P.: Shared Gaussian Process Latent Variable Models. Ph.D. dissertation, PhD thesis (2009)
22. Christopher, M.B.: Pattern Recognition and Machine Learning, vol. 16(4). Springer, New York (2006)
23. Chung, F.R.: Spectral Graph Theory, vol. 92. American Mathematical Soc., New York (1997)
24. Zhong, G., Li, W.-J., Yeung, D.-Y., Hou, X., Liu, C.-L.: Gaussian process latent random field. In: AAAI, pp. 679–684 (2010)
25. He, X., Niyogi, P.: Locality preserving projections. In: Mozer, M.C., Jordan, M.I., Petsche, T. (eds.) Advances in Neural Information Processing Systems, pp. 153–160. MIT Press, Cambridge (2004)
26. Yu, X., Wang, X.: Uncorrelated discriminant locality preserving projections. IEEE Signal Process. Lett. **15**, 361–364 (2008)
27. Kan, M., Shan, S., Zhang, H., Lao, S., Chen, X.: Multi-view discriminant analysis. IEEE Trans. Pattern Anal. Mach. Intell. **38**(1), 188–194 (2016)
28. Hu, P., Peng, D., Guo, J., Zhen, L.: Local feature based multi-view discriminant analysis. Knowl. Based Syst. **149**, 34–46 (2018)
29. Peng, X., Feng, J., Xiao, S., Yau, W.-Y., Zhou, J.T., Yang, S.: Structured autoencoders for subspace clustering. IEEE Trans. Image Process. **27**(10), 5076–5086 (2018)
30. Sugiyama, M.: Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. J. Mach. Learn. Res. **8**(May), 1027–1061 (2007)
31. Rahulamathavan, Y., Phan, R.C.-W., Chambers, J.A., Parish, D.J.: Facial expression recognition in the encrypted domain based on local fisher discriminant analysis. IEEE Trans. Affect. Comput. **4**(1), 83–92 (2013)
32. Kumar, S., Bhuyan, M., Chakraborty, B.K.: An efficient face model for facial expression recognition. In: 2016 Twenty Second National Conference on Communication (NCC), pp. 1–6. IEEE (2016)
33. Zhong, L., Liu, Q., Yang, P., Liu, B., Huang, J., Metaxas, D.N.: Learning active facial patches for expression analysis. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2562–2569. IEEE (2012)
34. Liu, P., Zhou, J.T., Tsang, I.W.-H., Meng, Z., Han, S., Tong, Y.: Feature disentangling machine-a novel approach of feature selection and disentangling in facial expression analysis. In: Computer Vision–ECCV 2014, pp. 151–166. Springer (2014)
35. Nusseck, M., Cunningham, D.W., Wallraven, C., Bülthoff, H.H.: The contribution of different facial regions to the recognition of conversational expressions. J. Vis. **8**(8), 1–1 (2008)
36. Zhu, X., Ghahramani, Z., Lafferty, J.D.: Semi-supervised learning using gaussian fields and harmonic functions. In: Proceedings of the 20th International conference on Machine learning (ICML-03), pp 912–919 (2003)
37. Zelnik-Manor, L., Perona, P.: Self-tuning spectral clustering. In: Mozer, M.C., Jordan, M.I., Petsche, T. (eds.) Advances in Neu-

ral Information Processing Systems, pp. 1601–1608. MIT Press, Cambridge (2004)
38. Lawrence, N.D., Quiñonero-Candela, J.: Local distance preservation in the gp-lvm through back constraints. In: Proceedings of the 23rd International Conference on Machine Learning, pp. 513–520. ACM (2006)
39. Bertsekas, D.P.: Constrained Optimization and Lagrange Multiplier Methods. Academic press, Cambridge (2014)
40. Rasmussen, C.E.: Gaussian Processes for Machine Learning, vol. 1. MIT Press, Cambridge (2006)
41. Goodall, C.: Procrustes methods in the statistical analysis of shape. J. R. Stat. Soc. Ser. B (Methodol.) **53**, 285–339 (1991)
42. Hotelling, H.: Relations between two sets of variates. Biometrika **28**(3/4), 321–377 (1936)
43. Rupnik, J., Shawe-Taylor, J.: Multi-view canonical correlation analysis. In: Conference on Data Mining and Data Warehouses (SiKDD 2010), pp. 1–4 (2010)
44. Sharma, A., Kumar, A., Daume, H., Jacobs, D.W.: Generalized multiview analysis: a discriminative latent space. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2160–2167. IEEE (2012)
45. Dhall, A., Goecke, R., Lucey, S., Gedeon, T.: Static facial expression analysis in tough conditions: data, evaluation protocol and benchmark. In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pp. 2106–2112. IEEE (2011)
46. Zhang, T., Zheng, W., Cui, Z., Zong, Y., Yan, J., Yan, K.: A deep neural network-driven feature learning method for multi-view facial expression recognition. IEEE Trans. Multimed. **18**(12), 2528–2536 (2016)
47. Liu, Y., Zeng, J., Shan, S., Zheng, Z.: Multi-channel pose-aware convolution neural networks for multi-view facial expression recognition. In: 2018 13th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2018), pp. 458–465. IEEE (2018)
48. Kang, Z., Pan, H., Hoi, S.C., Xu, Z.: Robust graph learning from noisy data. IEEE Trans. Cybern. (2019)
49. Li, D., Li, Z., Luo, R., Deng, J., Sun, S.: Multi-pose facial expression recognition based on generative adversarial network. IEEE Access **7**, 143980–143989 (2019)
50. Mollahosseini, A., Chan, D., Mahoor, M.H.: Going deeper in facial expression recognition using deep neural networks. In: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1–10. IEEE (2016)
51. Kim, B.-K., Roh, J., Dong, S.-Y., Lee, S.-Y.: Hierarchical committee of deep convolutional neural networks for robust facial expression recognition. J. Multimodal User Interfaces **10**(2), 173–189 (2016)
52. Li, J.. Lam, E.Y.: Facial expression recognition using deep neural networks. In: 2015 IEEE International Conference on Imaging Systems and Techniques (IST), pp. 1–6. IEEE (2015)
53. Khorrami, P., Paine, T., Huang, T.: Do deep neural networks learn facial action units when doing expression recognition? In: Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 19–27 (2015)
54. Lawrence, N.D.: Large scale learning with the gaussian process latent variable model. Technical Report CS-06-05, University of Sheffield, 2006. 3, 4, 7, Technical Report (2008)

Sunil Kumar received the Ph.D. degree in Electronics and Communication Engineering from IIT Guwahati, Guwahati, India. He did his post-graduate from NIT Rourkela in a specialization of Electronics Systems and Communication. Currently, Dr. Sunil is working as Assistant Professor at ABV-Indian Institute of Information Technology and Management (ABV-IIITM), Gwalior, Madhya Pradesh, India. His research interest includes Image Processing and Computer Vision, Pattern Recognition, and Deep Learning.

M. K. Bhuyan received the Ph.D. degree in electronics and communication engineering from IIT Guwahati, Guwahati, India. He was with the School of Information Technology and Electrical Engineering, University of Queensland, St. Lucia, QLD, Australia, where he was involved in the postdoctoral research. He was a Researcher with the SAFE Sensor Research Group, NICTA, Brisbane, QLD, Australia. He was an Assistant Professor with the Department of Electrical Engineering, IIT Roorkee, Roorkee, India. In 2014, he was a Visiting Professor with Purdue University, West Lafayette, IN, USA. He is currently a Professor with the Department of Electronics and Electrical Engineering, IIT Guwahati. His current research interests include image/video processing and computer vision. Dr. Bhuyan was a recipient of the National Award for Best Applied Research/Technological Innovation, which was presented by Honorable President of India, the Prestigious Fullbright-Nehru Academic and Professional Excellence Fellowship, and the BOYSCAST Fellowship.

Yuji Iwahori received B.S. degree from Nagoya Institute of Technology, Japan in 1983, M.S. and Ph.D. degree from Tokyo Institute of Technology, Japan in 1985 and 1988. He joined Nagoya Institute of Technology in 1988 and became a professor of Nagoya Institute of Technology in 2002. He joined Chubu University (Japan) as a professor in 2004 and he was a head of Graduate Program of Computer Science, Chubu University and he is acting a Vice-Dean of College of Engineering, Chubu University. He was a researcher in UBC, Canada and he has a research collaboration with UBC since 1991. He also has research collaborations with IIT Guwahati since 2010 and with Chulalongkorn University, Thailand since 2014. He was awarded for "Excellence in Global Engagement" from Ohio State University, US in November 2017. His research interests include Computer Vision, Image Recognition, Neural Network, Bioinformatics and Biomedical Imaging.