www.ijraset.com
IC Value: 13.98
Volume 4 Issue III, March 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Discriminative Shared Gaussian Process Based On Latent Variable Model – An Approach for Facial Expression Recognition

R.keerthanadevi[1], R.Sureshkumar[2]
[1]PG scholar, [2]Assistant Professor, Department of CSE
Tejaa Shakthi Institute of Technology for Women, Coimbatore

Abstract - Facial expression is an important channel for human communication and can be applied in many real applications. One critical step for facial expression recognition (FER) is to accurately extract emotional features. Existing methods for multi-view and/or view-invariant facial expression recognition typically perform separation of the observed expression using either classifiers learned separately for each view or a single classifier learned for all views. However, these approaches ignore the fact that different views of a facial expression are just different manifestations of the same facial expression. By accounting for this idleness, we can design more efficient classifiers for the target task. This paper proposes a discriminative shared Gaussian process latent variable model (DS-GPLVM) for multi-view and view- invariant separation of facial expressions from multiple views [1]. In this model, first learn a discriminative manifold shared by multiple views of a facial expression. Subsequently, we perform facial expression separation in the expression manifold. Finally, separation of an observed facial expression is carried out either in the view-invariant manner (using only a single view of the expression) or in the multi-view manner (using multiple views of the expression). The proposed model can also be used to perform mixture of different facial features in a principled manner. DS-GPLVM is proposed on both posed and impulsively displayed facial expression from three publicly available datasets (Multi-PIE, labeled face parts in the wild and static facial expression in the wild). And the results show that this model outperforms the modern methods for multi-view and view-invariant facial expression separation, and several modern methods for multi-view learning and feature mixture.
Keywords - Facial Expression Recognition, Discriminative Shared Gaussian Process, Latent Variable Model

I. INTRODUCTION

Facial expression recognition (FER) has been dramatically developing in recent years, thanks to advancements in related fields, especially machine learning, image processing, and human recognition [1]. Accordingly, the impact and possible usage of automatic FER have been growing in a wide range of applications, including human-computer interaction, robot control, and driver state surveillance. However, to date, robust recognition of facial expressions from images and videos is still a challenging task due to the difficulty in accurately extracting the useful emotional features. These features are often represented in different forms, such as static, dynamic, point-based geometric or constituency-based appearance. Facial movement features, which include feature position and shape changes, are generally caused by the movements of facial elements and muscles during the course of emotional expression. The facial elements, especially key elements, will continuously change their positions when subjects are expressing emotions. As an outcome, the same feature in different images usually has different positions. Most existing methods deal with imagery in which the depicted persons are relatively still and exhibit posed expressions in a nearly forward pose. However, many real-world applications relate to impulsive interactions (e.g., meeting summarization, political debates analysis, etc.), in which people tend to move their head while being recorded. Furthermore, depending on the camera position, facial images can be taken from multiple views. For these reasons, there is an ever rising need for computerized systems that can correctly perform multi- view and view-invariant facial expression recognition.

The main dispute here is to perform decoupling of the inflexible facial changes due to the head-pose and flexible facial changes due to the expression, as they are non-linearly coupled in 2D images. Another dispute is how to effectively make use of the information from multiple views (or different facial features) in order to ease the expression classification. Thus, accounting for the fact that each view of a facial expression is just a different manifestation of the same underlying facial expression related content is expected to result in more effective classifiers for the target task. These focus mainly on recognition of facial expressions of the six basic emotions. Based on how they deal with variation in head-pose (view) and expressions in 2D images, they can be divided into: (i) methods that perform view-invariant, i.e., per-view, (ii) methods that perform the view normalization before performing FER and (iii) methods that learn a single classifier using data from multiple views. However, the main downside of these approaches is that they fail to clearly model relationships between different views. This, in turn, results in classifiers that are less robust for the target task, but also more intricate in the case of large number of views/expressions. All this can efficiently be ameliorated using the modeling strategy of multi-view leaning methods.

This paper introduces the Discriminative Shared Gaussian Process Latent Variable Model (DS-GPLVM) for multi-view and view-invariant FER [1].

The multi- view learning strategy is presented to represent multi-view facial expression data on a common expression manifold. And use the idea of Shared GPs, the generative framework for discovers a non-linear subspace shared across unusual observation spaces (e.g., the facial views or feature representations) is used. Since our ultimate goal is the expression separation, we place a discriminative earlier, informed by the expression labels, over the manifold. The separation of an observed expression is then performed in the learned manifold using the kNN classifier. The proposed model is a simplification of the discriminative GP Latent Variable Models (D-GPLVM) for non-linear dimensionality reduction and separation of data from a single observation space. The learning of DS-GPLVM is carried out using the expression data from multiple views. Separation of an observed facial expression, however, can be carried out either in the view- invariant manner (in case only a single view of the observed expression is available at runtime) or in the multi-view manner (in case multiple views of the observed expression are available at runtime). The proposed model can also perform mixture of different facial features in order to improve view-invariant facial expression separation. In order to keep the model computationally well-mannered in the presence of large number of views, we propose a learning algorithm that splits the learning into different sub-problems (for each view), and then employs the Alternating Direction Method (ADM) [9] to optimize each sub-problem independently.

The contributions of this work can be summarized as follows. 1) We propose the DS-GPLVM for multi-view and/or view- invariant FER. The proposed model is a simplification of existing discriminative dimensionality decline methods from single to multiple observation spaces. This is, also, the first approach that exploits the multi-view learning strategy in the context of multi-view FER.2) We propose a novel learning algorithm for well-organized optimization of the model parameters that is based on the ADM strategy. This allows us to solve the model parameters' optimization problem for each-view, as a separate sub-problem, to perform parameter optimization for each view separately, consequential in the model being computationally well-organized even in the case of a large number of views. 3) The proposed DS-GPLVM is valid to a mixture of tasks (multi-view classification, multiple-feature fusion, pose-wise classification, etc.). Compared to modern methods for multi-view learning, which use linear techniques to align special views on a manifold, the DS-GPLVM is a kernel-based method, being able to learn non-linear correlations between different views. In contrast to modern methods for view-invariant and/or multi-view FER, the DS-GPLVM exploits dependencies between different views, humanizing the FER performance.

Finally, use the GPs as a basis for our (non-parametric) multi-view learning framework because, in contrast to majority of parametric models, it allows us to capture subtle details of facial expressions and preserve them on the expression manifold that is largely robust to the view/subject differences. Furthermore, due to the probabilistic nature of GPs, different types of priors can effortlessly be integrated into the model for multi-view learning (in our case, discriminative priors over the expression manifold). Last but not least, GPs are known for their ability to simplify quite well even from a small number of training data (on the order of several hundred) [10]. While this may not seem a big advantage when data are rich, it is of crucial importance for multi-view FER due to the lack of existing datasets containing annotated expressions and poses.

II. LITERATURE REVIEW

A regression-based scheme for multi-view facial expression recognition based on 2−D geometric features is presented in [3]. The problem by mapping facial points (e.g. mouth corners) from non-frontal to frontal view where further recognition of the expressions can be performed using a state-of-the-art facial expression recognition method is presented. To learn the mapping functions, the authors investigate four regression models: Linear Regression (LR), Support Vector Regression (SVR), Relevance Vector Regression (RVR) and Gaussian Process Regression (GPR). The extensive experiments on the CMU Multi- PIE facial expression database show that the proposed scheme out performs view-specific classifiers by utilizing considerably less training data.

Emotion recognition from facial images is a very active re- search topic in human computer interaction (HCI). However, most of the previous approaches only focus on the frontal or nearly frontal view facial images [4]. In contrast to the frontal/nearly-frontal view images, emotion recognition from non-frontal view or even arbitrary view facial images is much more difficult yet of more practical utility. To handle the emotion recognition problem from arbitrary view facial images, in this paper the author propose a novel method based on the regional covariance matrix (RCM) representation of facial images. They also develop a new discriminant analysis theory, aiming at reducing the dimensionality of the facial feature vectors while preserving the most discriminative information, by minimizing an estimated multiclass Bayes error derived under the Gaussian mixture model (GMM). An efficient algorithm to solve the optimal discriminant vectors of the proposed discriminant analysis method is also presented. The authors render thousands of multi-view 2D facial images from the BU-3DFE database and conduct extensive experiments on the generated database to demonstrate the effectiveness of the proposed method. It is worth noting that their method does not require face alignment or facial landmark point localization, making it very attractive.

A linear regression-based projection for multi-view facial expressions recognition (MFER) based on sparse features. While facial expression recognition (FER) approaches have become popular in frontal or near to frontal views, few papers demonstrate their results on arbitrary views of facial expressions [2]. Their model relies on a new method for multi-view facial expression recognition, where the authors encode appearance-based facial features using sparse codes and learn projections from non- frontal to frontal views using linear regression projection. Then they reconstruct facial features from the projected sparse codes using a common global dictionary. Finally, the reconstructed features are used for facial expression recognition. Their regression of sparse codes approach outperforms the state-of-the- art results on both protocols of BU3DFE dataset.

Novel framework for the recognition of facial expressions at arbitrary poses that is based on 2D geometric features. The authors address the problem by first mapping the 2D locations of landmark points of facial expressions in non-frontal poses to the corresponding locations in the frontal pose. Then, recognition of the expressions is performed by using any state-of-the-art facial expression recognition method (multi-class SVM) [6]. To learn the mappings that achieve pose normalization, they use a novel Gaussian Process Regression (GPR) model which they name Coupled Gaussian Process Regression (CGPR) model. Instead of learning single GPR model for all target pairs of poses at once, or learning one GPR model per target pair of poses independently of other pairs of poses, they propose CGPR model, which also models the couplings between the GPR models learned independently per target pairs of poses. The proposed method is the first one satisfying all: (i) being face-shape-model-free, (ii) handling expressive faces in the range from -45° to $+45^\circ$ pan rotation and from -30° to $+30^\circ$ tilt rotation, and (iii) performing accurately for continuous head pose despite the fact that the training was conducted only on a set of discrete poses.

Facial-expression data often appear in multiple views either due to head-movements or the camera position. Existing methods for multi-view facial expression recognition perform classification of the target expressions either by using classifiers learned separately for each view or by using a single classifier learned for all views. However, these approaches do not explore the fact that multi-view facial expression data are different manifestations of the same facial- expression-related latent content. To this end, we propose a Shared Gaussian Process Latent Variable Model (SGPLVM) for classification of multi-view facial expression data [7]. In this model, first learn a discriminative manifold shared by multiple views of facial expressions, and then apply a (single) facial expression classifier, based on k-Nearest-Neighbours (kNN), to the shared manifold. In their experiments on the Multi-PIE database, containing real images of facial expressions in multiple views, show that the proposed model outperforms the state- of-the-art models for multi-view facial expression recognition.

A research into facial expression recognition has predominately been applied to face images at frontal view only. Some attempts have been made to produce pose invariant facial expression classifiers. However most of the attempts have only considered yaw variations of up to 45° where all of the faces are visible. Little work has been carried out to investigate intrinsic potential of different poses for facial expression recognition. This is largely due to databases available, which typically capture frontal view faces images only. Recent database, BU3DFE and multi-pie allow emprical investigation of facial expression recognition for different viewing angles. A sequential two stage approach is taken for pose classification and view dependent facial expression classification to investigate the effects of yaw varaitions from frontal to profile views [11]. Local binary patterns and variations of LBPs as texture descriptions are investigated. Such features allow investigation of the influence of orientation and multi-resolution analysis for multi view facial expression recognition.

III. MATERIALS AND METHOD

The GPLVM is a generative model of the data, where a simple spherical Gaussian prior is placed over the manifold. However, this model can be adapted for classification by using a discriminative prior that encourages the latent positions of the examples of the same class to be close and those of different classes to be far on the manifold. This has firstly been explored, where a prior based on Linear Discriminant Analysis (LDA) is proposed in [7].

A learning algorithm that splits the learning into different sub-problems (for each view), and then employs the Alternating Direction Method (ADM) [9] to optimize each sub-problem separately.

Volume 4 Issue III, March 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

The main challenge here is to perform decoupling of the rigid facial changes due to the head-pose and non-rigid facial changes due to the expression, as they are non-linearly coupled in 2D images. Another challenge is how to effectively exploit the information from multiple views (or different facial features) in order to facilitate the expression classification. To overcome the challenges, a discriminative shared Gaussian process latent variable model (DS-GPLVM) for multi-view and view- invariant classification of facial expressions from multiple views is used [1].

The image given as input is processed to extract the features i.e., skin color, lip detection, face detection. The features are classified to recognize the face expression. The classification is done by applying beizer curve on lip and face to obtain the output. And it is processed to extract features and that can be classified to obtain output as facial expression. Finally, it is observed using classifiers. Then the classified image is normalized using normalizer. Then it is processed through classifier to obtain the projection. The problem is spitted into sub-problems and each sub-problem can be optimized separately. The architecture, context and data flow diagrams of DS-GPLVM are given in Figures 1, 2 and 3 respectively.

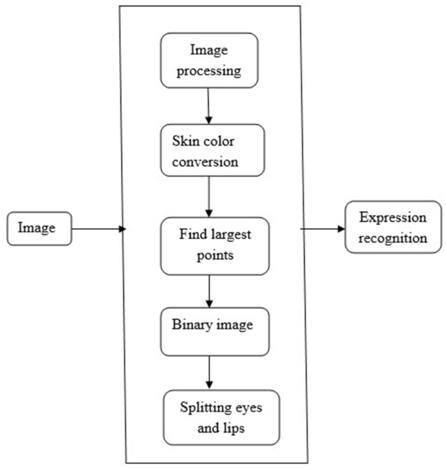

Fig.1 Architecture diagram of DS-GPLVM

Fig.2 Context diagram of DS-GPLVM

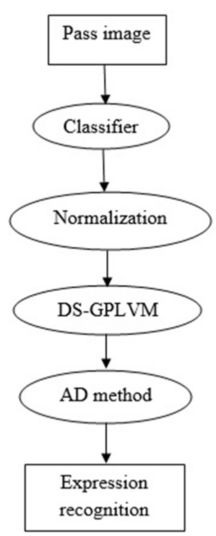

Fig.3 Data flow diagram of DS-GPLVM

Volume 4 Issue III, March 2016
ISSN: 2321-9653

IV. RESULTS AND DISCUSSION

In fig.4 for skin color segmentation, first we contrast the image. Then we perform skin color segmentation. Then, have to find the largest connected region. Then have to check the probability to become a face of the largest connected region. If the largest connected region has the probability to become a face, then it will open a new form with the largest connected region. If the largest connected regions height & width is larger or equal than 50 and the ratio of height/width is between 1 and 2, then it may be face.

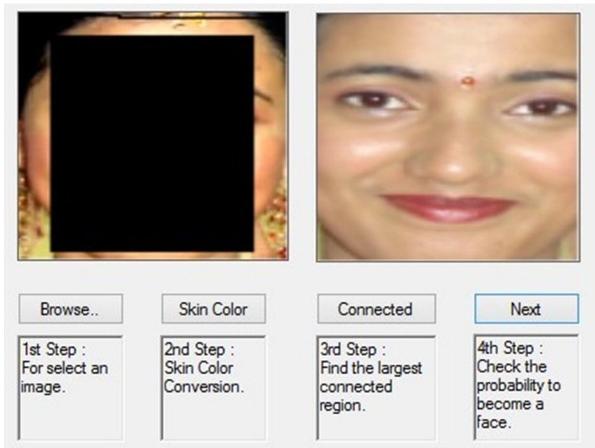

Fig.4 Skin color segmentation

For face detection, first convert binary image from RGB image. For converting binary image, calculate the average value of RGB for each pixel and if the average value is below than 110, replace it by black pixel and otherwise we replace it by white pixel. By this method, we get a binary image from RGB image. Then, try to find the forehead from the binary image. Start scan from the middle of the image, and then want to find continuous white pixels after a continuous black pixel. Then want to find the maximum width of the white pixel by searching vertical both left and right site. Then, if the new width is smaller half of the previous maximum width, then break the scan because if reach the eyebrow then this situation will arise. Then cut the face from the starting position of the forehead and its high will be 1.5 multiply of its width.

In the fig.5, it will be equal to the maximum width of the forehead. Then will have an image which will contain only eyes, nose and lip. Then will cut the RGB image according to the binary image. For eyes detection, convert the RGB face to the binary face. Now, consider the face width by W. Scan from the W/4 to (W-W/4) to find the middle position of the two eyes. The highest white continuous pixel along the height between the ranges is the middle position of the two eyes.

Fig.5 Face detection

Then find the starting high or upper position of the two eyebrows by searching vertical. For left eye, search w/8 to mid and for right eye search mid to w – w/8. Here w is the width of the image and mid is the middle position of the two eyes. There may be some white pixels between the eyebrow and the eye. To make the eyebrow and eye connected, place some continuous black pixels vertically from eyebrow to the eye. For left eye, the vertical black pixel-lines are placed in between mid/2 to mid/4 and for right eye the lines are in between mid+(w-mid)/ 4 to mid+3*(w-mid)/ 4 and height of the black pixel-lines are from the eyebrow starting

Volume 4 Issue III, March 2016
ISSN: 2321-9653

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

height to (h- eyebrow starting position)/4. Here w is the width of the image and mid is the middle position of the two eyes and h is the height of the image. Then find the lower position of the two eyes by searching black pixel vertically. For left eye, search from the mid/4 to mid - mid/4 width. And for right eye, search mid + (w-mid)/ 4 to mid+3*(w- mid)/ 4 width from image lower end to starting position of the eyebrow. Then find the right side of the left eye by searching black pixel horizontally from the mid position to the starting position of black pixels in between the upper position and lower position of the left eye. And left side for right eye search mid to the starting position of black pixels in between the upper position and lower position of right eye. The left side of the left eye is the starting width of the image and the right side of the right eye is the ending width of the image. Then cut the upper position, lower position, left side and the right side of the two eyes from the RGB image.

For lip detection, determine the lip box. And consider that lip must be inside the lip box. So, first determine the distance between the forehead and eyes. Then add the distance with the lower height of the eye to determine the upper height of the box which will contain the lip. Now, the starting point of the box will be the ¼ position of the left eye box and ending point will be the ¾ position of the right eye box. And the ending height of the box will be the lower end of the face image. So, this box will contain only lip and may some part of the nose. Then will cut the RGB image according the box. So, for detection eyes and lip, only need to convert binary image from RGB image and some searching among the binary image.

In the lip box, there is lip and may be some part of nose. So, around the box there is skin color or the skin. So, convert the skin pixel to white pixel and other pixel as black. Also find those pixels which are similar to skin pixels and convert them to white pixel. Here, if two pixels RGB values difference is less than or equal 10, then we called them similar pixel. Here, we use histogram for finding the distance between the lower average RGB value and higher average RGB value. If the distance is less than 70, then use 7 for finding similar pixel and if the distance is better than or equal 70 then use 10 for finding similar pixel. So, the value for finding similar pixel depends on the quality of the image. If the image quality is high, use 7 for finding similar pixel and if the image quality is low, use 10.

So, in the binary image, there are black regions on lip, nose and may some other little part which have a little different than skin color. Then apply big connected region for finding the black region which contain lip in binary image. And sure that the big connected region is the lip because in the lip box, lip is the largest thing which is different than skin.

Then we have to apply Bezier curve on the binary lip. For apply Bezier curve, find the starting and ending pixel of the lip in horizontal. Then draw two tangents on upper lip from the starting and ending pixel and also find two points on the tangent which is not the part of the lip. For the lower lip, find two-point similar process of the upper lip. Use Cubic Bezier curves for draw the Bezier curve of the lip. We draw two Bezier curve for the lip, one for upper lip and one for lower lip.

In fig.6, we first have to remove eyebrow from eye. For remove eyebrow, search 1st continuous black pixel then continuous white pixel and then continuous black pixel from the binary image of the eye box. Then remove the 1st continuous black pixel from the box and then get the box which only contains the eye. Now, the eye box which contains only eye has some skin or skin color around the box. So, apply similar skin color like the lip for finding the region of eye. Then apply big connect for finding the highest connected region and this is the eye because in the eye box, eye is the biggest thing which is not similar to the skin color. Then apply the Bezier curve on the eye box, similar to the lip. Then we get the shape of the eye.

Fig.6 Apply beizer curve on lip and eyes

In our database, there are two tables. One table "Person" is for storing the name of people and their index of 4 kinds of emotion which are stored in other table "Position". In the "Position" table, for each index, there are 6 control points for lip Bezier curve, 6 control points for left eye Bezier curve, 6 control points for right eye Bezier curve, lip height and width, left eye height and width

Volume 4 Issue III, March 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

and right eye height and width. So, by this method, the program learns the emotion of the people.

In fig.7, we have to find the Bezier curve of the lip, left eye and right eye. Then convert each width of the Bezier curve to 100 and height according to its width. If the person's emotion information is available in the database, then the program will match which emotion's height is nearest the current height and the program will give the nearest emotion as output.

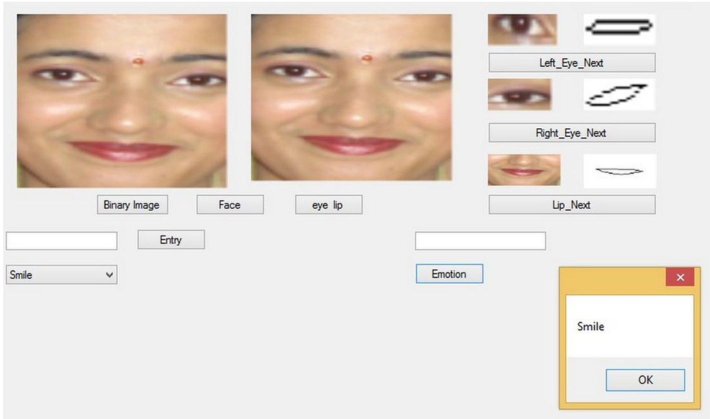

Fig.7 Emotion detection

If the person's emotion information is not available in the database, then the programs calculates the average height for each emotion in the database for all people and then get a decision according to the average height.

V. CONCLUSION

The DS-GPLVM model for learning a discriminative shared manifold of facial expressions from multiple views that is optimized for the expression classification. This model is a generalization of latent variable models for learning a discriminative subspace of a single observation space. As such, it presents a complete non-parametric multi-view learning framework that cans instantiate the rest of the compared non-linear single-view methods (i.e. D-GPLVM and GPLRF) [1]. As evidenced by results on posed and spontaneously displayed facial expressions, when compared to the state-of-the-art methods for supervised multi-view learning and facial expression recognition, modeling of the manifold shared across different views and/or features using the proposed framework considerably improves both multi and per- view/feature classification of facial expressions.

REFERENCES

[1] Stefanos Eleftheriadis, Ognjen Rudovic and Maja Pantic, "Discriminative Shared Gaussian Processes for Multi-view and View-Invariant Facial Expression Recognition", IEEE transactions on image processing, vol. 24, no. 1, January 2015.
[2] Mahdi Jampour, Thomas Mauthner and Horst Bischof "Multi-view Facial Expressions Recognition using Local Linear Regression of Sparse Codes", in Proc. 20th Computer Vision Winter Workshop, Austria, February 9-11, 2015
[3] O. Rudovic, I. Patras, and M. Pantic, "Regression-based multi- view facial expression recognition," in Proc. 20th Int. Conf. Pattern Recognit. (ICPR), Istanbul, Turkey, Aug. 2010, pp. 4121–4124.
[4] W. Zheng, H. Tang, Z. Lin, and T. S. Huang, "Emotion recognition from arbitrary view facial images," in Proc. Eur. Conf. Comput. Vis., 2010, pp. 490–503.
[5] N. Hesse, T. Gehrig, H. Gao, and H. K. Ekenel, "Multi-view facial expression recognition using local appearance features," in Proc. 21st Int. Conf. Pattern Recognit., Nov. 2012, pp. 3533–3536.
[6] O. Rudovic, M. Pantic, and I. Patras, "Coupled Gaussian processes for pose-invariant facial expression recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 6, pp. 1357–1369, Jun. 2013.
[7] S. Eleftheriadis, O. Rudovic, and M. Pantic, "Shared Gaussian process latent variable model for multi-view facial expression recognition," in Proc. ISVC, Jul. 2013, pp. 527–538.
[8] Ligang Zhang and Dian Tjondronegoro, "Facial Expression Recognition Using Facial Movement Features", IEEE transactions on affective computing, vol. 2, no. 4, 2011.
[9] D. P. Bertsekas, Constrained Optimization and Lagrange Multiplier Methods (Computer Science and Applied Mathematics), vol. 1. Boston, MA, USA: Academic, 1982.
[10] R. Urtasun and T. Darrell, "Discriminative Gaussian process latent variable model for classification," in Proc. 24th Int. Conf. Mach. Learn., Jun. 2007, pp. 927–934.
[11] S. Moore and R. Bowden, "Local binary patterns for multi-view facial expression recognition," Comput. Vis. Image Understand., vol. 115, no. 4, pp. 541–558, Apr. 2011.