Pose Invariant Face Recognition Using HMM and SVM with PCA For Dimensionality Reduction

[#1]Mr. K.Seetharaman, [#2]Mr. N. Palanivel, [#3]R. Indumathi

[#1] *Associate Professor, Computer Science and Engineering,*
Annamalai University,
Chidambaram, Tamilnadu, India.
[#2]*Assistant Professor, Department of Computer Science,*
Manakula Vinayagar Institute of Technology,
Pondicherry, India.
[#3] *PG Scholar, Department of Computer Science,*
Manakula Vinayagar Institute of Technology,
Pondicherry, India.
[#1]kseethaddeau@gmail.com
[#2]npalani76@gmail.com
[#3]indu5490@gmail.com

Abstract—**An embedded system is presented in which face recognition and facial recognition for Human-Robot Interaction are implemented. To detect face with a fast and reliable way, HMM combined with SVM algorithm is used. The full pose from face recognition data base is considered to detect the face recognition. Performance of the face recognition reaches to 99.617%. The two main advantages of our method are that it does not require manually selected facial landmarks or head pose estimation. In order to improve the performance of our pose normalization method in face recognition, an algorithm is presented for classifying whether a given face image is at a frontal or non frontal pose. In addition to the proposed method, a pre processing state of detecting the face is included here. The images which are non faces are detected and eliminated from the database is an another main advantage in of this work.**

*Keywords —**Belief propagation, frontal face synthesizing, Markov random fields, pose-invariant face recognition.***

I. INTRODUCTION

Human beings express their emotions in everyday interactions with others. Emotions are frequently reflected on the face, in hand and body gestures, in the voice, to express our feelings or liking. Recent Psychology research has shown that the most expressive way humans display emotions is through face recognitions. Mehrabian indicated that the verbal part of a message contributes only for 7% to the effect of the message as a whole, the vocal part for 38%, while face recognitions for 55% to the effect of the speaker's message. Emotions are feeling or response to particular situation or environment. Emotions are an integral part of our existence, as one smiles to show greeting, frowns when confused, or raises one's voice when enraged. It is because other emotions and react based on that face recognition only enriches the interactions. Computers are "emotionally challenged". They neither recognize other emotions nor possess its own emotion. To enrich human-computer interface from point-and-click to sense-and-feel, to develop non intrusive sensors, to develop lifelike software agents such as devices, this can express and understand emotion. Since computer systems with this capability have a wide range of applications in different research arrears, including security, law enforcement, clinic, education, psychiatry and Telecommunications. There has been much research on recognizing emotion through face recognitions. In emotional classification there are two basic emotions are there, Love-fear. Based on this we classify the emotion into positive and negative emotions. The six basic emotions are angry, happy, fear, disgust, sad, surprise. One more face recognition is neutral. Other emotions are Embarrassments, interest, pain, shame, shy, anticipation, smile, laugh, sorrow, hunger, curiosity.

It indicates that only 7% of message is due to linguistic language, 38% is due to paralanguage and 55 % of message is communicated by face recognitions. This implies that the face recognition is a major modality in human face-to-face communication. Imagine that, when designing the Human Computer Interfaces (HCI), the face recognitions seems to be a major factor for improving the communicability of message, even in human-machine communication.

Recognition of human face recognition by computer is a key to develop such technology. In recent years, much research has been done on machine recognition of human face recognitions. Cross-cultural psychological research on face recognitions indicates that there may be a small set of face recognitions that are universal. The first methodologically sound studies, and concluded that the emotions "Happiness, Anger, Sadness, Disgust, Surprise and Fear" are shown and interpreted in all human cultures in the same way.

One of the means of showing emotion is through changes in face recognitions. But are these face recognitions of emotion constant across cultures? For a long time, Anthropologists and Psychologists had been grappling with this question. However the views were varied and there was no general consensus. Subjects from western and eastern cultures and reported that the face recognitions of emotions were indeed constant across cultures. A critique questioning that claims of universal recognition of emotion from face recognitions. Since then, it has been considered as an established fact that the recognition of emotions from face recognitions is universal and constant across cultures. Cross cultural studies have shown that, although the interpretation of face recognitions is universal across cultures, the face recognition of emotions though facial changes depend on social context. For example, when American and Japanese subjects were shown emotion eliciting videos, they showed similar face recognitions. However, in the presence of an authority, the Japanese viewers were much more reluctant to show their true emotions through changes in face recognitions. They used only six emotions, namely happiness, sadness, anger, surprise, disgust and fear. In their own words: “the six emotions studied were those which had been found by more than one investigator to be discriminable within any one literate culture”. These six face recognitions have come to be known as the ‘basic’, ‘prototypic’ or ‘archetypal’ face recognitions. Since the early 1990s, researchers have been concentrating on developing automatic face recognition systems that recognize these six basic face recognitions.

Posed face recognitions are the artificial face recognitions that a subject will produce when he or she is asked to do so. This is usually the case when the subject is under normal test condition or under observation in a laboratory. In contrast, spontaneous face recognitions are the ones that people give out spontaneously. These are the face recognitions that we see on a day to day basis, while having conversations, watching movies etc. Since the early 1990s, till recent years, most of the researchers have focused on developing automatic face recognition systems for posed face recognitions only. This is due to the fact that posed face recognitions are easy to capture and recognize. Furthermore, it is very difficult to build a database that contains images and videos of subjects displaying spontaneous face recognitions.

II. Background

Pose variations can be considered as one of the most important and challenging problems in face recognition. As the viewpoint varies, the 2D facial appearance will change because the human head has a complex non planar geometry. Magnitudes of variations of innate characteristics, which distinguish one face from another, are often smaller than magnitudes of image variations caused by pose variations . Popular frontal face recognition algorithms, such as Eigen faces or Fisher faces, usually have low recognition rates under pose changes as they do not take into account the 3D alignment issue when creating the feature vectors for matching. Existing methods for face recognition across pose can be roughly divided into two broad categories: (1) techniques that rely on 3D models, and (2) 2D techniques. In the first type of approaches, the morphable model fits a 3D model to an input face using the prior knowledge of human faces and image-based reconstruction. The main drawback of this algorithm is that it requires many manually selected landmarks for initialization. Furthermore, the optimization process is computationally expensive and often converges to local minima due to a large number of parameters that need to be determined.

Other general possible problems in existing system

The challenges to unconstrained face recognition in surveillance cameras are mainly due to the following reasons.

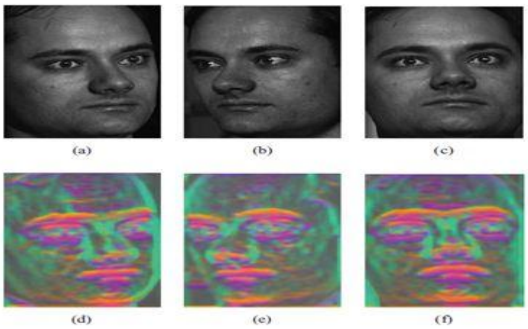
- Low resolution. In the video captured by surveillance cameras, the pixels that account for the faces are very limited. However, previous studies have shown that faces of size 64 64 are required for the existing algorithms to achieve good recognition accuracy.
- Arbitrary poses. Usually the subjects are moving freely. Consequently, it is not uncommon that the captured faces have different poses in different cameras.
- Varying lighting conditions. As the lighting is usually not uniform in the coverage area of the surveillance cameras, the illumination on the subject’s face could vary significantly as he/she moves (e.g., the subjects walks into the shade from direct sunshine).
- Noise and blurriness. The captured images are often corrupted by noise during transmission and the motion of the subjects usually introduces blurriness.

III. Problem Definition

Frontal Face Reconstruction Using Markov Random Fields

Given an input image I of a non frontal face and M training face images $T(k)$, $k = 1, . . . , M$ captured at the frontal pose, all of them are divided into the same regular grid of N overlapping patches of size $w \times h$. A set of M possible local warps $P_i = \{p(k)\ I : k = 1, . . . , M\}$ can be estimated for each patch I_i , by aligning it with the corresponding patches of the training images using the method. By aligning the patches in the non frontal views with the ones in the frontal views, the information about how the local patches are transformed as a result of the 3D rotation of the face. The goal of our algorithm is to find a globally optimal set of warps for all the patches in the input image such that predict the input face at the frontal pose by transforming these patches using the obtained warps. This problem can be turned into a discrete labeling problem with a well defined objective function using a discrete MRF. Note that in our approach, the training database need not contain the frontal images of the person in the input image I.

The first extension to the standard BP is the use of *dynamic label pruning*. If the number of active labels for a node is greater than L_{max}, a user specified constant, label pruning will be applied to the node. The labels of a visited node are traversed in the descending order of relative belief $b_{rel\ i}(p_i)$, where the relative belief is defined as $b_{rel\ i}(p_i) = b_i(p_i) - b_{max\ I}$ and $b_{max\ i}$ is the maximum belief of node i . Those labels $p_i \in P_i$ with $b_{rel\ i}(p_i) > b_{prune}$ are selected as active labels for node i . b_{prune} is the label pruning threshold belief.



Furthermore, a label is declared as active only if it is not too similar to any of the already active labels in order to avoid choosing many similar labels and wasting a large part of the active label set. Two labels are considered similar if their normalized cross correlation is greater than a threshold *T* similar. Note that a minimum number of labels *L* min is always kept for each node. The complexity of updating the messages is reduced from *O(*|*L*|2*)* to *O(*|*L*max|2*)* by applying label pruning to BP . In addition, the speed of BP can also be improved by precomputing the reduced matrices of pairwise potentials.

The second improvement is the use of *message scheduling* to determine the transmitting order for a node based on the confidence of that node about its labels. *The node most confident about its label should be the first one to transmit outgoing messages to its neighbors* . The priority of a node is defined as *priority(i)* = 1|*Qi* | where |*Qi* | is the cardinality of the set *Qi* = {p*i* ∈ *Pi* : *b*rel *i* (p*i)* ≥ *b*conf}. *B*conf is the confidence threshold belief. By employing this message scheduling in BP, the node that has the most informative messages will transmit first in order to increase the confidence of its neighbors. This helps the algorithm to converge only after only a small, fixed number of iterations. Furthermore, message scheduling also makes the neighbors of the transmitting node more tolerant to label pruning.

IV. Scheme Description

1) Training Phase

In the network with training examples, which consist of a pattern of activities for the input units together with the desired pattern of activities for the output units how closely the actual output of the network matches the desired output. The network produces a better approximation of the desired output. The input to the training phase is a collection of images showing human faces. These images are also called as Face images. These face images are then passed through a feature extraction step. In the feature extraction step key attributes of the images are computed and stored as a vector called feature vector. These feature vectors define or represent the most important properties observed in the face image. Highest Eigen values are chosen.

There are two advantages of this step. First, the size of the data is reduced from the entire image to only a few selected important features. Second, the selection of features gives more structured information than just basic pixel values of the images.

2) Testing Phase

This phase can be performed to measure the classification rate. The inputs to this phase are the models that were build during training phase and the test images for which the emotions are to be recognized. Here again only the face region is used as rest of the image do not contribute information about the emotion. In a typical real time scenario the input image would be detected face image from an earlier face detection phase.

The first step here again would be a feature extraction phase where the key features from the face image are extracted. The extraction method must be same as the one used in the training phase. The output of this step is the feature vector of the face image that would then be subjected to a testing step. In the testing step the feature vector is tested against the models built during the training phase. The output of the testing step is a score that indicated the emotion that is detected by the model. This score is usually in the form of distance or probability and it defines which model was best suited for the feature vector extracted in the previous step.

In the testing step there are two possible ways that can be employed. The first possibility is in the case when one model was built per class of emotion. Here the feature vector is tested against all the models and their scores then define which model was the most suited one. The second possibility is the case when only one model was built for the entire set and a single score defines the possible emotion detected. During testing a simple image the approach can correctly classify it as the correct emotion expressed. In another case the approach can also wrongly classifies a sample image as the correct face. These cased constitute false positives. It could also be the case that the approach wrongly classifies a sample images as incorrect emotion expressed. These cased constitute false negatives

3) Preprocessing

Pre-processing block the face image can be treated with a series of pre-processing techniques to minimize the effect of factors that can adversely influence the face recognition algorithm. The most critical of these are facial pose and illumination.

To perform better in the recognition stage, we apply a set of pre-processing techniques: the first step in pre-processing is to bring all images into the same color space and to normalize the size of face regions. This normalization process is critical to improving the final face recognition rate and we will later present some experimental results for our HMM-specific AFR.

a) Color to grayscale conversion : In most face recognition applications the images are single or multiple views of 2D intensity data, and many databases built for face recognition applications are available as grayscale images. From the four databases used in our experiments, 3 contained grayscale images (BioID, Achermann, UMIST) and one contained RGB images (FERET). Practical images will, naturally, be acquired in color as modern image acquisition systems are practically all color and to convert from color to grayscale, or intensity images of the selected face regions. In practice the intensity data may be available from the imaging system – many camera system employ YCC data internally and the Y component can be utilized directly. In other cases we may need to perform an explicit conversion of RGB data. Here a set of red, green and blue integer values characterize an image pixel.

b) Resizing: For a HMM-based face recognition system having a consistently sized face region is particularly important because the HMM requires regional analysis of the face with a scanning window of fixed size. A straightforward approach is to resize all determined face regions to a common size. To facilitate more efficient computation we seek the smallest sized face region possible without impacting the overall system recogntion rate. Some empirical data will be presented later to illustrate how different factors, including the size of normalized face regions, affect recognition rate.

c) Denoing: The face images consist of film artifact or labels (may be) That are removed using tracking algorithm. Here, starting from the first row and the first column, the intensity value, greater than that of the threshold value is removed from images. During removal of film artifacts, the image consist of salt and pepper noise. This can be filtered through mean or median filters. Then the image is given to the enhancement stage for the removing high intensity component and the above noise. This part is used to enhance the smoothness towards piecewise homogeneous region and reduce the edge-blurring effects. This proposed system describe the information of enhancement using weighted median filter for removing high frequency component.

d) Illumination normalization: One of the most important factors that influence the recognition rate of a system is illumination variation. In was shown in that variations in illumination can be more relevant than variations between individual characteristics. Such variations can induce an AFR system to decide that two different individuals with the same illumination characteristics are more similar than two instances f the same individual taken in different lighting conditions. Thus normalizing illumination conditions across detected face regions is crucial to obtaining accurate, reliable and repeatable results from an AFR. One approach suitable for face models which combine both facial geometry and facial texture such as active appearance models (AAM) is described by [Ionita 2008]. However as HMM techniques do not explicitly rely on facial geometry or textures it is not possible to integrate the illumination normalization within the structure of the model itself. Instead on a discrete illumination normalization process.

4) Feature detection using HMM

The first features used in face recognition performed with HMM were pixel intensities . The recognition rates obtained using pixel intensities with a P2D-HMM were up to 94.5% on the ORL database. However the use of pixel intensities as features has some disadvantages: firstly they cannot be regarded as robust features since: (i) the intensity of a pixel is very sensitive to the presence of noise in the image or to illumination changes; (ii) the use of all the pixels in the image is computationally complex and time consuming; and (iii) using all image pixels does not eliminate any redundant information and is thus a very inefficient form of feature. Another example of features used with EHMM for face recognition are KLT features. used with recognition rates of up to 98% on ORL database. The main advantage of using KLT features instead of pixel intensities is their capacity to reduce redundant information in an image. The disadvantage is their dependence of the database of training images from which they are derived .

The most widely used features for HMM in face recognition are 2D-DCT coefficients. These DCT coefficients combine excellent decorrelation properties with energy compaction. Indeed, the more correlated the image is, the more energy compaction increases. Thus a relatively small number of DCT coefficients contain the majority of information encapsulated in an image. A second advantage is the speed with which they can be computed since the basis vectors are independent of the database and are often pre computed and stored in an imaging device as part of the JPEG image compression standard. Recognition rates obtained when using 2D DCT with HMM can achieve 100% success on smaller databases such as ORL. The use of wavelets have not been previously used with HMMs for face recognition applications.

Feature extraction for both 1D and 2D HMMs was originally described . The method was subsequently adopted in the majority of HMM-based face recognition papers. This feature extraction technique is based on scanning the image with a fixed-size window from left-to-right and top-to-bottom. A window of dimensions h × w pixels begins scanning each extracted face region from the left top corner sub-dividing the image into a set number of h × w sized blocks.

On each of these blocks a transformation is applied to extract the characterizing features which represent the observation vector for that particular region. Then the scanning window moves towards right with a step-size of n pixels allowing an overlap of o pixels, where o = w − n. Again features are extracted from the new block. The process continues until the scanning window reaches the right margin of the image. When the scanning window reaches the right margin for the first row of scanned blocks, it moves back to the left margin and down with m pixels allowing an overlap of v pixels vertically. The horizontal scanning process is resumed and a second row of blocks results, and from each of these blocks an observation vector is extracted. The scanning process and extraction of blocks is depicted in Figure4.1.

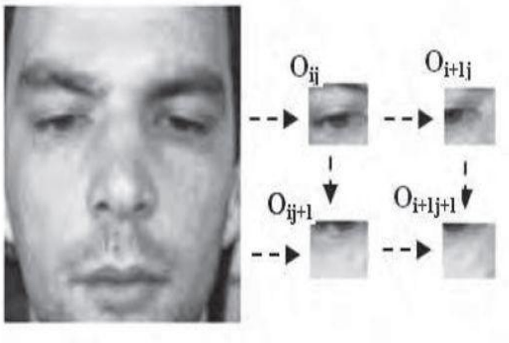


Figure 4.1 Scanning process and extraction of blocks

I. An optimal size for face image when using HMM is 128 × 128 pixels, although the recognition rates obtained for the smaller size of 96 × 96 pixels are quite close and may offer a better choice for applications where memory and computational efficiency are important, e.g. in handheld imaging devices;

II. The optimal number of Gaussian functions is 3, representing a trade-off between best precision and computational burden; but we remark that with 2 Gaussians a faster computation is achieved with almost the same recognition performance; the effects of varying the number of Gaussians across super stages was not considered in this research;

III. The optimal performance with 2D DCT is achieved by employing the first 9 coefficients, but we noted that using the first 4 coefficients gives an acceptable result as well and may be preferable where speed of computation and memory efficiency are important; no significant improvement was noted when we used Daubechies wavelets in place of DCT;

IV. Very good results were obtained for a reduced 2-4-4-2 EHMM topology applied on very small images: these results improve when increasing the number of training images per person, and recognition rates as high as 86.43% were achieved in our experiments. As no illumination normalization was used and these tests were performed on a combined database rather than a single standard database, the results which were obtained may be regarded as highly promising for real-world applications.

V. Three different illumination normalization techniques are used in the pre-processing phase of our recognizer. Investigated some non-standard combinations of these techniques to determine their suitability for pre-processing data for a HMM face recognition algorithm. The best results were obtained for CLAHE and HE with over 95% recognition rates. Very good performances were obtained when using a combination of CLAHE and logDCT, with 92.87% recognition rate, but also when using the more basic HE, with up to 92.5%. This analysis of various image normalization filters should pro- vide a useful baseline for future researchers in face recognition. One aspect we would have liked to investigate was the potential of such combining of illumination normalization filters to improve the performance of other well-known face recognition techniques such as PCA, ICA and AAM methods.

5) Concept of PCA

Principal Component Analysis (PCA) has been proven to be an efficient method in pattern recognition and image analysis, Recently, PCA has been extensively employed for face-recognition algorithms, such as eigenface and fisherface. The encouraging results have been reported and discussed in the literature. Many PCA-based face-recognition systems have also been developed in the last decade. Principal component analysis (PCA) is one of the most popular representation methods for face recognition. PCA also known as Karhunen-Loeve method is a technique commonly used for dimensionality reduction in computer vision, particularly in face recognition. A method called Eigen dent, based on PCA was used in face recognition. In PCA, the principal components of the distribution of faces or the eigenvectors of the covariance matrix of the set of face images are sought treating an image as a point in a very high dimensional space. These eigenvectors can be thought of as a set of features that together characterize the variation between face images.

Given an s-dimensional vector representation of each face in a training set of M images, Principal Component Analysis (PCA) tends to find a t-dimensional subspace whose basis vectors correspond to the maximum variance direction in the original image space. This new subspace is normally lower dimensional (t << s). New basis vectors define a subspace of face images called *face space*. All images of known faces are projected onto the face space to find a set of weights that describes the contribution of each vector.

To identify an unknown image, that image is projected onto the face space to obtain its set of weights. By comparing a set of weights for the unknown face to sets of weights of known faces, the face can be identified. If the image elements are considered as random variables, the PCA basis vectors are defined as eigenvectors of the scatter matrix S_T defined as

$$S_T = \sum_{i=1}^{M} (x_i - \mu).(x_i - \mu)^T$$

Where μ the mean of all images in the training is set (the *mean face*) and x_i is the *i*-th image with its columns concatenated in a vector. The projection matrix W_{PCA} is composed of *t* eigenvectors corresponding to *t* largest eigenvalues, thus creating a *t*-dimensional face space.

6) SVM based classification.

The basic idea behind the SVM classification technique is to identify the class of the input test vectors. This is a supervised learning algorithm, where the training vectors are used to train the system to map these training vectors in a space with clear gaps between them using some standard kernel functions and the input test vectors are mapped on to the same space to predict the possible class .

Given some training data D, a set of *n* points of the form

$$\mathcal{D} = \{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in \mathbb{R}^p,\, y_i \in \{-1, 1\}\}_{i=1}^{n}$$

where the y_i is either belonging to the class 1 or class −1, indicating the class to which the point X_i belongs. Each X_i is a *p*-dimensional real vector. Here it is needed to find the maximum-margin hyperplane that divides the points having y_i=1 from those having y_i= - 1. So any hyperplane can be written as the set of points $\mathbf{X}$ satisfying

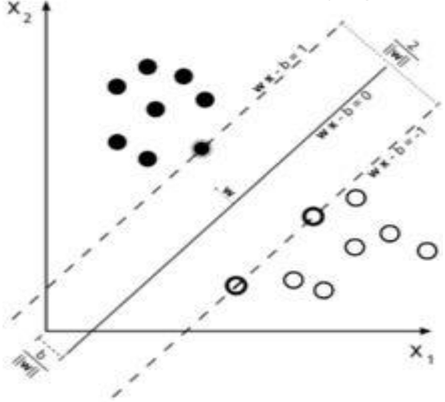


Figure 6.1 SVM Scenario

Maximum-margin hyperplane and margins for an SVM trained with samples from two classes. Samples on the margin are called the support vectors.

w.x – b = 0

where **.**denotes the dot product and $\mathbf{W}$ is the normal vector to the hyperplane. The parameter b/||w|| determines the offset of the hyperplane from the origin along the normal vector $\mathbf{W}$.

If the training data are linearly separable, then two hyperplanes can be selected in such a way that they separate the data and there are no points between them, and then tried to maximize their distance. The region bounded by them is called "the margin". These hyperplanes can be described by the equations

$$w.x – b =1$$

and

$$w.x – b = -1$$

At the testing phase, the data points Xi are separated using the following constraints

$w.x_i – b\geq 1$ for x_i of the first class or $w.x_i – b\leq 1$ for x_i of the second class.

V. Conclusion

An approach to face recognition in the static images with different poses is evaluated. This emotion analysis system implemented using HMM and SVM based RBF network for feature selection and classification. The design to recognize emotional face recognition in human faces using the average values calculated from the training samples. The system was able to identify the images and evaluate the face recognitions accurately from the images.

References

[1]. P. Ekman and E. L. Rosenberg, What the Face Reveals: Basic and Applied Studies of Spontaneous Face recognition Using the Facial Action Coding System. Oxford, U.K.: Oxford Univ. Press, 2005.

[2]. J. Russell and J. Fernandez-Dols, The Psychology of Face recognition. New York: Cambridge Univ. Press, 1997.

[3]. B. Golomb and T. Sejnowski, “Benefits of machine understanding of face recognitions,” in NSF Report—Face recognition Understanding, P. Ekman, T. Huang, T. Sejnowski, and J. Hager, Eds. Salt Lake City, UT, 1997, pp. 55–71.

[4]. M. Pantic, “Face for ambient interface,” in Ambient Intelligence in Everyday Life, vol. 3864, Lecture Notes on Artificial Intelligence. Berlin, Germany: Springer-Verlag, 2006, pp. 32–66.

[5]. M. Cohen, Perspectives on the Face. Oxford, U.K.: Oxford Univ. Press, 2006.

[6]. A.Young, Face and Mind. Oxford, U.K.: Oxford Univ. Press, 1998.

[7]. M. Pantic and L. J. M. Rothkrantz, “Toward an affect-sensitive multimodal human–computer interaction,” Proc. IEEE, vol. 91, no. 9, pp. 1370–1390, Sep. 2003.

[8]. Z. Zeng, M. Pantic, G. Roisman, and T. Huang, “A survey of affect recognition methods: Audio, visual and spontaneous face recognitions,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 1, pp. 39–58, Jan. 2009.

[9]. Y. L. Tian, T. Kanade, and J. F. Cohn, Handbook of Face Recognition. New York: Springer-Verlag, 2005.

[10]. M. Pantic and M. Bartlett, “Machine analysis of face recognitions,” in Face Recognition, K. Delac and M. Grgic, Eds. Vienna, Austria: I-Tech Educ. Publishing, 2007, pp. 377–416.