ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 5, May 2015

Facial Expression Recognition Dealing With Different Expression Variations

Abhijeet S. Tayde[1], Prof. A. S. Deshpande[2]

Electronics and Tele-communication (Signal Processing), Savitribai Phule University of Pune[1]

JSPM's Imperial College of Engineering and Research, Wagholi, Pune, Maharashtra, India[2]

Abstract: Automatic facial expression analysis is one of the an interesting and challenging problem and also impacts important applications in many areas such as human–computer interaction and data-driven animation. An important step for successful facial expression recognition is to deriving an effective facial representation from original face images. For this, empirically evaluate facial representation majorly based on statistical local features, Local Binary Patterns for person-independent facial expression recognition. Several machine learning methods are systematically examined on several databases. Local Binary Pattern features are effective and much more efficient for facial expression recognition. And formulating Boosted-LBP further to extract the most discriminant LBP features and the best recognition performance is obtained by using Support Vector Machine classifiers with Boosted LBP features. An investigation on LBP features for low-resolution facial expression recognition which is a critical problem but seldom addressed in the existing work. According to observation of experiments that LBP features perform not only stably but also robustly over a useful range of low resolutions of face images and yield promising performance in compressed low-resolution video sequences captured in real-world environments.

Facial expressions are one of the most critical sources of variation in face recognition, especially in the frequent case where only a single sample per person is available for enrollment. Some methods that improve the accuracy in the presence of such variations are still required for a reliable authentication system. Because of this, we address the problem with an analysis by- synthesis-based scheme in which a number of synthetic face images with different expressions are produced. For this an animatable 3D model is generated for each user based on 17 automatically located landmark points and the contribution of these additional images in terms of the recognition performance is evaluated with three different techniques such as principal component analysis(PCA), Linear Discriminant Analysis(LDA) and local binary patterns(LBP) on face recognition. Significant improvements are achieved in face recognition accuracies for each algorithm.

Keywords: LDA, LBP, PCA.

1. INTRODUCTION

The most powerful, natural and immediate means for human beings to communicate their emotions and intensions is the facial expressions. Automatic facial expression analysis is a challenging problem and impacts important applications in many areas like human–computer interaction and data-driven animation. Due to wide range of applications, automatic facial expression recognition gain much more attention. Though much progress has been made and recognizing facial expression with a high accuracy remains difficult due to the subtlety, complexity and facial expressions variability.

In successful facial expression recognition, deriving an effective facial representation from original face images is a major step. The two common approaches to extract facial features are: geometric feature-based methods and appearance-based methods. Geometric features present the shape and locations of facial components which are extracted to form a feature vector that represents face geometry. It usually requires accurate and reliable facial feature detection and tracking which is quite difficult to accommodate in many situations. With appearance-based methods, image filters like Gabor wavelets applied to either the whole face or specific face regions to extract the appearance changes of the face. Because of their superior performance the major works on appearance-based methods have focused on using Gabor-wavelet representations. Thus it is both time and memory intensive to convolve face images with a bank of Gabor filters to extract multi-scale and multi-orientational coefficients.

In this empirically study facial representation based on Local Binary Pattern (LBP) features for person-independent facial expression recognition. LBP features were proposed originally for texture analysis and recently have been introduced to represent faces in facial images analysis. Some of the most important properties of LBP features are tolerance against illumination changes and their computational simplicity.

We evaluate the generalization ability of LBP features across different databases. The limitation of the existing facial expression recognition methods is that they attempt to recognize facial expressions from data collected in a highly controlled environment given high resolution frontal faces. However in real-world applications such as smart meeting and visual surveillance, the input face images are at low resolutions. Low-resolution images in real world environments make real-life expression

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 5, May 2015

recognition much more difficult. Recently, a first attempt to recognize facial expressions at low resolutions. The effects of different image resolutions for each step of automatic facial expression recognition are take into cosideration. In this investigation on LBP features for low-resolution facial expression recognition is carried out. Experiments on different image resolutions shows LBP features perform stably and robustly over a useful range of low resolutions of face images. The better performance on real-world compressed video sequences illustrated their promising applications in real-life environments. The extended version of our previous work as follows:

- Empirically evaluate LBP features for person-independent facial expression recognition. Several different machine learning methods are exploited to classify expressions on several databases. LBP features were previously used for facial expression classification and more recently presented an extended LBP operator to extract features for facial expression recognition, however these existing works were conducted on a very small database using an individual classifier. In contrast, here comprehensively study LBP features for facial expression recognition with different classifiers on much larger databases.
- Investigation of LBP features for low-resolution facial expression recognition, a critical problem but seldom addressed in this work. Evaluate the performance on different image resolutions and also conduct experiments in real-world compressed video sequences. Compared to other work, LBP features provide just as good or better performance, so are very promising for real-world applications.

2.0 FACE FEATURE EXTRACTION APPROACH

The first part of the identification process is Face feature extraction. Before describing the featuring process, the another important operation related to face recognition must be mentioned. A proper image face registration is essential for a good face-recognition performance. This face registration process perform by using some facial detection algorithms and some image pre-processing operations may be necessary which are mentioned further. First, the original face images have to be converted to the grayscale form. After that, some contrast and illumination adjustment operations are performed. All face images must be processed with the same illumination and contrast. Therefore, some histogram equalization operations are performed on these images to obtain a satisfactory contrast. Also, the facial images are often corrupted by various types of noise. So, process them with the proper low-pass filters, for noise removal and restoration. The enhanced face images are now ready for the featuring process. A Gabor filter-based face feature extraction is proposed and tries to obtain some feature vectors which provide optimal characterizations of the visual content of facial images. For this reason, choosing the two-dimensional Gabor filtering a widely used image processing tool for feature extraction. Besides face recognition, Gabor filters are successfully used in many other image processing and analysis domains, such as: image smoothing, coding, texture analysis, shape analysis, edge detection, fingerprint and iris recognition. The Gabor filter (Gabor Wavelet) represents a band-pass linear filter whose impulse response is defined by a harmonic function multiplied by a Gaussian function. Thus, a bi-dimensional Gabor filter constitutes a complex sinusoidal plane of particular frequency and orientation modulated by a Gaussian envelope. It achieves an optimal resolution in both spatial and frequency domains.

This approach designs 2D odd-symmetric Gabor filters for face image recognition with the following form:

$$G_{\theta_k,f_i,\sigma_k,\sigma_y}(x,y) = exp\left(-\left[\frac{x_{\theta_k}^2}{\sigma_x^2} + \frac{y_{\theta_k}^2}{\sigma_y^2}\right]\right).\cos\left(2\pi f_i x_{\theta_k} + \varphi\right) \quad (1)$$

where $x_{\theta_k} = x.cos\theta_k + y.sin\theta_k$, $y_{\theta_k} = y.cos\theta_k - x.sin\theta_k$, f_i provides the central frequency of the sinusoidal plane wave at an angle θ_k with the x – axis, σ_x and σ_y represent the standard deviations of the Gaussian envelope along the two axes, *x* and *y*. By setting the phase $\phi = \pi / 2$ and compute each orientation as $\theta_k = \frac{k\pi}{n}$, where k = {1 …. n}. The 2D filters $G_{\theta_k,f_i,\sigma_k,\sigma_y}$ given by relation (1) represent a group of wavelets which optimally captures both local orientation and frequency information from a digital image. In this, each face image is filtered with $G_{\theta_k,f_i,\sigma_k,\sigma_y}$ at various orientations, frequencies and standard deviations. Thus, the design of Gabor filters for facial recognition needs an appropriated selection of those filter parameters.

Thus, consider some proper variance values, a set of radial frequencies and a sequence of orientations. So, let the filter's parameters be $\sigma_x = 2, \sigma_y = 1$, where $f_i \in \{0.75,1.5\}$ and n = 1 which means $\theta_k = \left\{\frac{\pi}{5}, \frac{2\pi}{5}, \frac{3\pi}{5}, \frac{4\pi}{5}, \pi\right\}$.. Thus, create a 2D Gabor filter bank $\left\{G_{\theta_k,f_i,2,1}\right\} f_i \in \{0.75,1.5\} k \in \{1,5\}$ composed of 10 channels. The created filter set is applied to the input facial image by convolving the face image with each Gabor filter from set. Resulted Gabor responses are then concatenated into a three-dimensional feature vector. If *I* represent such a face image, having a [*X* ×*Y*] size, then its feature extraction can be expressed as follows:

$$V(I)[x,y,z] = V_{\theta(z),f(z),\sigma x,\sigma y}(I)[x,y] \quad (2)$$

And

$$V_{\theta(z),f(z),\sigma x,\sigma y}(I)[x,y] = I(x,y) * G_{\theta_k,f,\sigma_k,\sigma_y}(x,y) \quad (3)$$

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 5, May 2015

A fast 2D convolution could be performed using the Fast Fourier Transform, therefore formula (3) is equivalent with the following relation:

$$V_{\theta(z),f(z),\sigma x,\sigma y}(I) = FFT^{-1}\left[FFT(I).FFT\left(G_{\theta_k,f,\sigma_k,\sigma_y}\right)\right] \quad (4)$$

Therefore, for each facial image I obtain a 3D face feature vector V(I), having a [X × Y × 2n] dimension. This tridimensional feature vector constitutes a robust content descriptor of the input face. A certain face image (marked with a red rectangle) and its 10 Gabor representations that constitute the components of the corresponding feature vector, are displayed in Fig. 1.

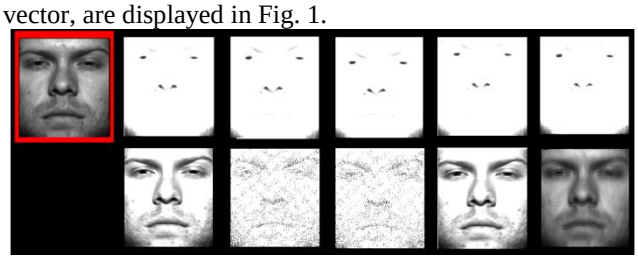

Fig. 1.: Human face and its 2D Gabor representations (feature vector components).

There are various metrics which can be applied to these feature vectors. Since the size of each vector depends on the size of the corresponding face image a resizing procedure has to be performed on the compared facial images, first. Then, some well-known metrics such as Euclidean distance or the sum of absolute differences (SAD) could be applied. The distance between these facial feature vectors using a squared Euclidean metric is computed. In the proposed system, the enrollment is assumed to be done in both 2D and 3D for each subject under a controlled environment – frontal face images with a neutral expression and under ambient illumination. The obtained 3D shape of the facial surface together with the registered texture is preprocessed, firstly to extract the face region. On the extracted facial surface, scanner-induced holes and spikes are cleaned and a bilateral smoothing filter is employed to remove white noise while preserving the edges. After the hole and noise free face model (texture and shape) is obtained, 17 feature points are automatically detected using either shape, texture or both according to the regional properties of the face. These detected points are then utilized to warp a generic animatable face model so that it completely transforms into the target face and the generic model with manually labeled 71 MPEG-4 points is suitable to simulate facial actions and expressions via an animation engine that is in accordance with MPEG-4 Face and Body Animation (FBA) specifications. Finally, in order to simulate the facial expressions on the obtained animatable model, an animation engine called visage is utilized. Multiple expression-infused face images are generated for each subject to enhance face recognition performance.

2.1. Data Preprocessing

3D scanner outputs are mostly noisy. The purposes of the preprocessing step can be listed as:

1. to extract the face region (same in 2D and 3D images);
2. to eliminate spikes/holes introduced by the sensor;
3. to smooth the 3D surface.

Firstly adopting the proposed method , the nose tip is detected; for each row, the position with the maximum z value is found and then for each column. The number of these positions is counted to create a histogram. The peak of this histogram is chosen as the column for the position of the vertical midline and the maximum point of this contour is identified as the nose tip. Using a sphere of radius 80mm and centered 10mm away from the nose tip in +z direction, the facial surface is cropped. Next, the existing spikes are removed by threshoulding. Spikes are frequent with laser scanners especially in the eye region. After the vertices that are detected as spikes are deleted, they leave holes on the surface together with other already existing holes (again usually around the eyes and eyebrows), they are filled by applying linear interpolation. Once the complete surface is obtained; a bilateral smoothing filter is employed to remove white noise while preserving the edges. This way, the facial surface is smoothed but the details hidden in high frequency components are maintained.

2.2. Automatic Landmarking

Bearing in mind that subject cooperation is required during the enrollment, the system based on the assumption of a well-controlled acquisition environment in which subjects are registered with frontal and neutral face images. In accordance with this scenario, goal to extract a subset (17 points) of MPEG-4 Facial Definition Parameters (FDPs) to be utilized. For the alignment of the faces with the animatable generic model. For the extraction process of the points, 2D and/or 3D data are used according to the distinctive information they carry in that particular facial region.

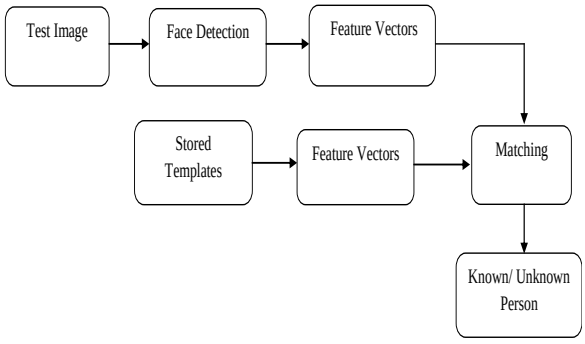

Fig. 2: Block diagram for facial expression recognition

For the extraction of the points, 2D and/or 3D data are used according to the distinctive information they carry in that particular facial region. The 17 facial interest points are detected in total, consisting of 4 points for each eye, 5 points for the nose and 4 points for the lips. These steps are detailed in the following:

1) Vertical Profile Analysis:

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 5, May 2015

The analysis done on the vertical profile constitutes the backbone of the entire system. It starts with the extraction of the facial midline and for this purpose; the nose tip is detected. Nose tip position allows to search for the eyes in the upper half of the face in order to approximately locate irises, so that the roll angle of the face can be corrected before any further processing.

2) Eye Regions:

The 3D surface around the eyes tends to be noisy because of the reflective properties of the sclera, the pupil and the eyelashes. On the other hand, its texture carries highly descriptive information about the shape of the eye. For that reason, 2D data is preferred and utilized to detect the points of interest around the eyes, namely the iris center, the inner and outer eye corners and the upper and the lower borders of the iris.

3) Nose Region:

Contrary to the eye region, nose region is extremely distinctive in surface but quite plain in texture. Start with the yaw angle of the face is corrected in 3D. For this purpose, the horizontal curve passing through the nose tip is examined. Ideally, an area under this curve should be equally separated by a vertical line passing through its maximum (assuming the nose is symmetrical).

3.0 CONSTRUCTINGTHE ANIMATABLE FACE MODELS

In order to construct an animatable face model for each enrolled subject;, a mesh warping algorithm based on the findings is proposed. The generic face model, with holes for the eyes and an open mouth is strongly deformed to fit the facial models in the database, using the TPS method. 17 points to be automatically detected together with the rest of the FDP points for MPEG-4 compliant animations are annotated for the generic face. MPEG-4 specifications and the mathematical background of the TPS method will be briefly explained before going into details about the proposed animatable face construction method.

3.1 MPEG-4 Specifications and Facial Animation Object Profiles*:*

MPEG-4 is an ISO/IEC standard developed by Moving Picture Experts Group which is a result of efforts
of hundreds of researchers and engineers from all over the world. Mainly defining a system for decoding audiovisual objects, MPEG-4 also includes a definition for the coded representation of animatable synthetic heads. In other words, independent of the model, enables coding of graphics models and compressed transmission of related animation parameters. The facial animation object profiles defined under MPEG-4 are often classified under three groups. Simple facial animation object profile: The decoder receives only the animation information and the encoder has no knowledge of the model to be animated.

- Calibration facial animation object profile:

The decoder also receives information on the shape of the face and calibrates a proprietary model accordingly prior to animation.

- Predictable facial animation object profile:

The full model description is transmitted. The encoder is capable of completely predicting the animation produced by the decoder. The profile most conformable to approach is the second one: calibration facial animation object profile, since aiming to calibrate the "generic" model according to the samples in database. System generates an animatable model by using 17 of 71 MPEG-4 specified face definition parameters (FDP) which are annotated automatically. The rest of the points are only marked on the generic model for animation.

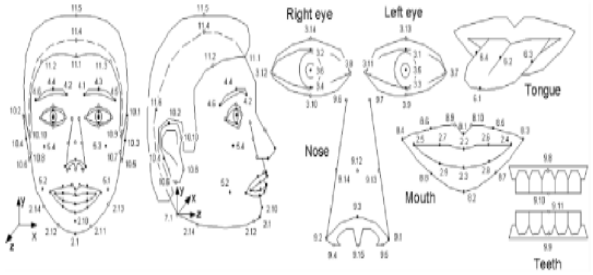

Fig. 3.: MPEG-4 Facial Definition Parameters.

In Fig. 3, the positions of the MPEG-4 FDP points are given. Most of these points are necessary for an MPEG-4 compliant animation system, except for the ones on the tongue, the teeth and the ears, depending on the animation tool structure.

4.0 A SUPERVISED FACE CLASSIFICATION METHOD

The next stage of the face identification process consists of feature vector classification. Here a supervised classification technique for these Gabor filter-based 3D feature vectors is proposed. Some of the popular supervised classifiers including *minimum distance classifier* and *K-Nearest Neighbour* (K-NN) classifier, can be used in this case. An extended version of minimum distance classifier, named the *minimum average distance classifier* is developed. First to create the training set of this supervised classifier. Now consider *N* authorized (registered) persons. Each of these registered users provides a set of faces of its own, which are included in the training set. Each face image from the training set represents a template face. Therefore, the model of the proposed training face set can be expressed as $\left\{\{F_j^i\} j = 1, \ldots n(i)\right\}$ i = 1 ... n where F_j^i represent jth template face of the ith user and n(i) is the number of training faces of the i_{th} user. A classification process creates N face classes, each class corresponding to a registered person. Then, one computes the

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 5, May 2015

training feature vector set as $\{V\{F_j^i\}j = 1, \dots n(i)\}\text{i} = 1 \dots \text{n}$
Also, we consider a set of input digital images to be recognized. Let us note them{I1…….Ik}. The classification approach inserts each of these input images in the class of the *closest* registered user, representing the user corresponding to the minimum *average distance*. An average distance value is computed as the mean of the distances between the feature vector of the input image and the feature vectors of the template faces corresponding to an authorized person. Minimum average distance classification process is expressed formally as follows:

$$Class(j) = arg\,{}^{min}/_{i \in [1,K]}\, i = 1 \frac{\sum_{t=1}^{n(i)} d(V(I_j), V(F_t^i))}{n(i)}, \forall j \in \quad (5)$$

where the result Class(j) ∈ [1, N] represents the index of the face class where Ij is inserted. Let C1……Cn be the resulted classes.
These obtained human face classes represent the face identification result. Input image is identified as a face of a registered person. Unfortunately, some of these identified images could not really represent the persons that associated with them. Some of them could not represent human face images at all.

5.0 RESULTS AND DISCUSSIONS 8796271798

Different challenging data sets are used for evaluation. The detection performance is evaluated on data sets. Results are reported for data sets using the per-image methodology. Results are reported using performance at Equal Error Rate (EER). The classification performance is evaluated on different data sets. All the classification experiments for each data set are repeated 10 times with randomly selected training and test images (15 and 30). The average of per-class recognition rates is recorded for each run. The mean and standard deviation of the results from individual runs is reported as the final results. The framework is used in the classification experiments where the SIFT features are replaced with the proposed features. The texture classification performance is evaluated on 2 data sets – Brodatz, and KTH-TIPS2-a.
A 8 × 8 pixels block size is used for data set. For INRIA, Caltech Pedestrian, Caltech 101 and Caltech 256, a 16 × 16 pixels block size is used. A 50% overlap of blocks is considered. The histograms are normalized using L1-norm. The square root of the bins are then taken. Linear SVM classifier is used. During this classification experiments, the linear SVM classifiers are trained using a one-versus-all rule i.e., a classifiers trained to separate each class from the rest and a test image is assigned the label of the classifier with the highest response.
For the texture classification experiments, we follow the procedures for training and testing on Brodatz and
KTH-TIPS2-a. Global features are used for Brodatz and KTHTIPS2- a i.e. the entire image is represented by a single histogram. The DRLBP and DRLTP histograms are normalized first using L2-norm followed by L1-norm. In Similar way , we use a 3-nearest neighbor classifier with normalized histogram intersection as the distance measure between features.
For all data sets, a circular neighbourhood of radius 1 (R) and 8 (B) pixels is considered. The uniform pattern representation is used. For LTP and DRLTP in our experiments, the threshold, T , is 3 for INRIA and Caltech Pedestrian, 9 for UIUC Car, Caltech 101 and Caltech 256, 15 for Brodatz and 5 for KTH-TIPS2-a.

5.1 Performance Comparison of DRLBP and DRLTP Against LBP, LTP and RLBP

We compare the performance of DRLBP, DRLTP against LBP, LTP, RLBP on INRIA for detection and on Caltech 101 for classification. INRIA training set contains 2416 cropped positive images and 1218 uncropped negative images. The sliding image window size is 128 × 64 pixels. By randomly take 10 samples from each negative image to obtain a total of 12180 negative samples for training the linear SVM classifier. The Bootstrapping is performed across multiple scales at a scale step of 1.05 to obtain hard negatives which are added to the original training set for retraining.

The INRIA test set consists of 288 images. Images are scanned over multiple scales at a scale step of 1.05. A window stride is 8 pixels in the x and y directions. The miss rate (MR) against false positives per image (FPPI) (using log-log plots) is plotted to compare between different detectors. The *log-average miss rate* (LAMR) is used to summarize the detector performance which is computed by averaging the miss rates at nine evenly spaced FPPI rates in the range 10^{-2} to 10^{0}. If any of the curves end before reaching 10^{0}, the minimum miss rate achieved is used. It is seen that our proposed features outperform its predecessors. The RLBP underperforms LBP as there is a loss of information due to the mapping of LBP codes and their complements to the same code. DRLBP outperforms RLBPLTP outperforms LBP thanks to its robustness to noise and small pixel value fluctuations. Similarly, DRLTP outperforms LTP. In this overall, DRLTP performs the best at 29%. It is seen that our proposed features outperform its predecessors. DRLTP has a recognition rate of 72.59% while LTP has a recognition rate of 55.71% for the 15 training and test images case. This shows a significant gain of 17%. Furthermore for 30 training and test images case, the gain is 14%. Similarly, DRLBP has a gain of 1% and approximately 3% in comparison to RLBP and LBP for both cases. Furthermore, we also perform another experiment using 90% of the samples per class as training data with the remaining 10% as test data. There is a significant improvement in performance for all features. This is expected as there are more samples available for training which improves classification performance. The DRLTP still gives the best performance at 93.1%.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 5, May 2015

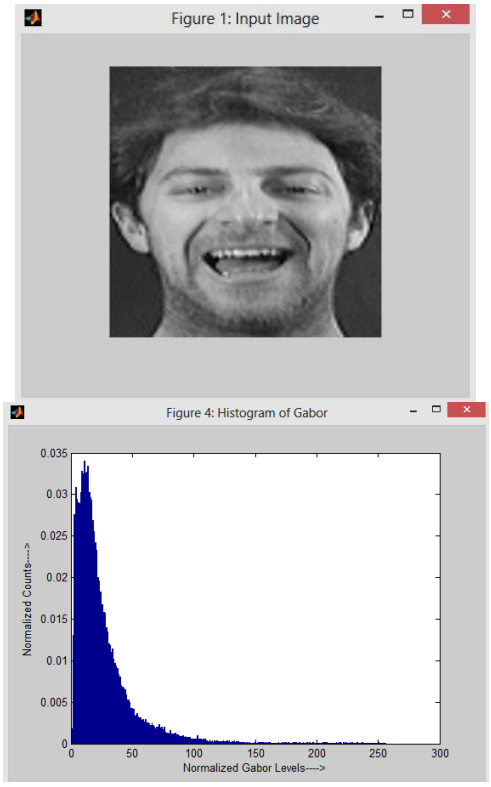

Fig 4: Input image use for face feature extraction and its histogram by using Gabor filter

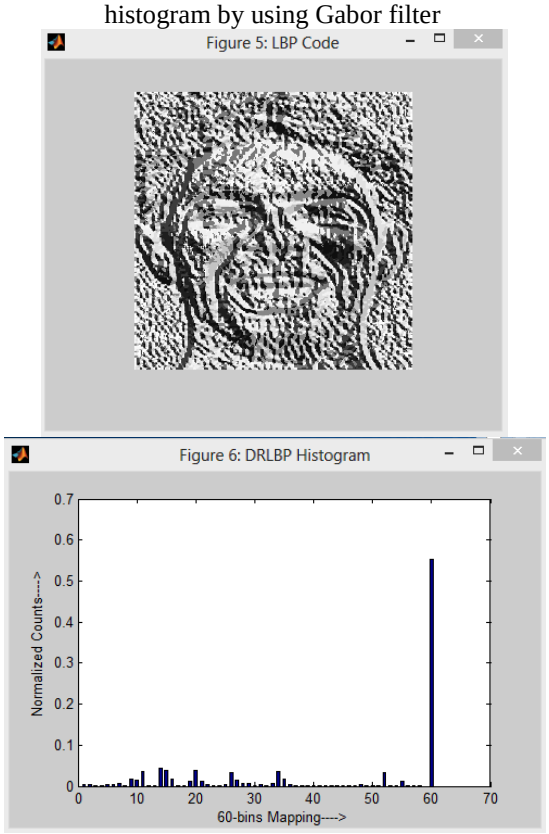

Fig.5: Extracted face features assembles to create 3D animatable model and its histogram using discriminant robust local binary pattern

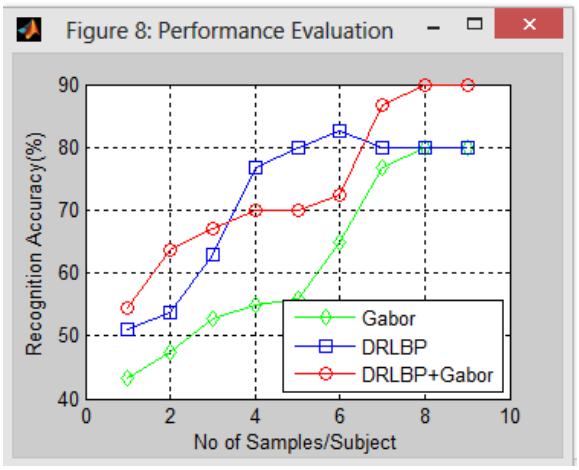

Fig:6: Performance evaluation of Gabor filter, DRLBP and its comparison

6.0 CONCLUSION

Based on the assumption of a fully-controlled environment for enrollment a face recognition framework is proposed in which the widely-encountered single sample problem for identification of faces with expressions is targeted by augmenting the dataset with synthesized images. Several expressions are simulated for each enrolled person on an animatable model which is specifically generated based on the 3D face scan of that subject.

For the animatable model generation, a generic model for which MPEG-4 FDPs are located manually is utilized. This is totally based on only 17 common points on both the generic and the target models, the generic model is first coarsely warped using the TPS method. By assuming the surfaces are close enough, new and denser point correspondences are formed by pairs with minimum distance and fine warping is applied. Finally, the texture is copied.

A sub-procedure on automatic detection of those 17 landmarks is presented utilizing both 2D and 3D facial data. For the simulation of facial expressions on the generated models, an animation engine, called visage|life™ is utilized. The facial images with expressions constitute a synthetic gallery, of which the contribution to the face recognition performance is evaluated on a PCA-based implementation. These experiments are conducted on two large and well accepted databases; FRGC and Bosphorus 3D face database.

The experiment results reveal that introduction of realistically synthesized face images with expressions improves the performance of the identification system. In addition to this, it is based on the evaluations on Bosphorus database. The imprecisions introduced by the proposed automatic landmarking algorithm has no adverse effect on the success rates thanks to the corrective property of the warping phase.

REFERENCES

1. L. D. Introna and H. Nissenbaum, "Facial recognition technology: A survey of policy and implementation issues," in *Report of the Center for Catastrophe Preparedness and Response*. New York, NY, USA: New York Univ., 2009.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 5, May 2015

2. N. Erdogmus and J.-L. Dugelay, "Automatic extraction of facial interest points based on 2D and 3D data," in *Proc. Electron. Imag. Conf. 3D Image Process. (3DIP) Appl.*, San Francisco, CA, USA, Jan. 2011, pp. 1–13.
3. N. Erdogmus and J.-L. Dugelay, "An efficient iris and eye corners extraction method," in *Proc. Joint IAPR Int. Workshop Struct., Synth., Statist. Pattern Recognit.*, Cesme, Turkey, 2010, pp. 549–558.
4. X. Liu, Y.-M. Cheung, M. Li, and H. Liu, "A lip contour extraction method using localized active contour model with automatic parameter selection," in *Proc. 20th Int. Conf. Pattern Recognit.*, 2010, pp. 4332–4335.
5. N. Erdogmus and J.-L. Dugelay, "An efficient iris and eye corners extraction method," in *Proc. Joint IAPR Int. Workshop Struct., Synth., Statist. Pattern Recognit.*, Cesme, Turkey, 2010, pp. 549–558
6. S. Agarwal, A. Awan, and D. Roth, "Learning to detect objects in imagesvia a sparse, part-based representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1475–1490, Nov. 2004.
7. T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.
8. H. Bay, A. Ess, T. Tuytelaars, and L. J. V. Gool, "Speeded-up robust features (surf)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.
9. O. Boiman, E. Shechtman, and M. Irani, "In defense of nearest-neighbor based image classification," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.
10. P. Brodatz, *Textures: A Photographic Album for Artists and Designers*.New York, NY, USA: Dover Publications, Aug. 1999.
11. B. Caputo, E. Hayman, and P. Mallikarjuna, "Class-specific material categorisation," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2. Oct. 2005, pp. 1597–1604.
12. J. Chen *et al.*, "WLD: A robust local image descriptor," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1705–1720, Sep. 2010.
13. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 886–893.
14. H. Deng, W. Zhang, E. Mortensen, T. Dietterich, and L. Shapiro, "Principal curvature-based region detector for object recognition," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
15. P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Trans. Pattern Anal. Mach.Intell.*, vol. 34, no. 4, pp. 743–761, Apr. 2012.
16. L. Fei-fei, R. Fergus, and P. Perona, "One-shot learning of objectcategories," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 594–611, Apr. 2006