International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol. 10, No. 6 (2017), pp. 13-22
http//dx.doi.org/10.14257/ijsip.2017.10.6.02

Expression Invariant Face Recognition System

Deepti Ahlawat[1*] and Vijay Nehra[1]

[1]Department of Electronics and Communication Engineering
Bhagat Phool Singh Mahila Vishwavidyalaya, Khanpur Kalan, Sonipat, Haryana, India
[*]deeptijaglan@gmail.com, nehra_vijay@yahoo.co.in

Abstract

Face recognition as one of the biometric technique is one of the interesting topic in the field of computer vision and pattern recognition. In the present study, attempt has been made in this direction to address the issues and challenges in face recognition system viz. pose, illumination, facial expressions etc . So, a method is presented which converts input face image with an arbitrary expression into its corresponding neutral facial expression image. In the present study, the method applies deep neural network to train the system to locate the landmarks in the face image. The landmark points create an intermediate triangular mesh and warp the image using affine transform. This warped image is classified using Support Vector Machine (SVM) and k-means classifier. Finally, results are compared and recognition accuracy is determined for different expressions. The result obtained from the investigated work is compared with state-of-art algorithms. For the proposed method the recognition rate is 97.3% for JAFFE database and 97.8% for CK database which outperforms the other state-of-art methods. The present work is tested on JAFFE and CK databases.

Keywords*: Expression, Invariance, Warping, Face Recognition, Support Vector Machine*

1. Introduction

Since last two decades, facial recognition has received substantial attention in the biometrics unlike other biometrics techniques. It is a non-intrusive method and can be used without subject’s knowledge. A great effort has being devoted to automate facial recognition system, but it still remains a challenging problem. Despite of more than thirty years of deep research work, the performance of the state-of-the-art of many face recognition systems is greatly affected. This is because human facial appearance has potentially large variations due to pose, illumination, facial expression, facial hair, aging, occlusion due to accessories [1]. The research on face recognition has being progressing with continuous visible improvements made from past to present.

Face recognition has numerous applications viz. public identification, verification of credit cards and driving license, criminal justice system and forensic, law enforcement and security applications to name only a few [2-4]. The typical applications of face recognition are shown in Figure 1.

ISSN: 2005-4254 IJSIP

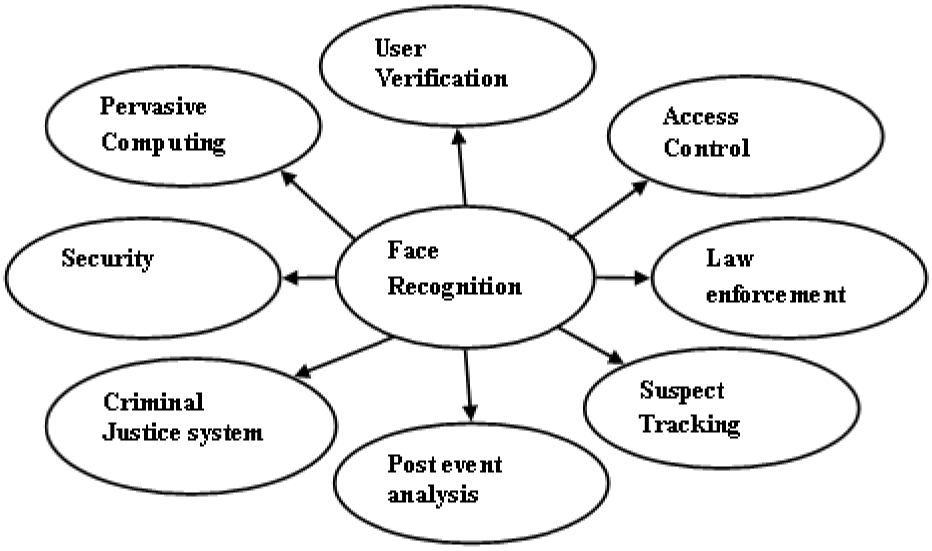

Figure 1. Typical Applications of Face Recognition [4]

Today, one of the vital problems in the recognition of a person is emotions and verbal communication which cause variations in the facial expressions. Expressions have an adverse effect on the system accuracy and is the cause of failure for many face recognition algorithms. When we want to identify a person, facial expressions make the recognition difficult to identify a person. The expression images can be transformed into images that are same as neutral images which are used for training. However, it is not sure that the image is transformed correctly due to lack of appropriate information. Many researchers have investigated methods to improve the face recognition by removing the facial expressions to obtain a neutral face *i.e.* making the face expression-invariant [6-13]. Hence to develop a robust face recognition algorithm which is insensitive to expression variations is one of the greatest challenges in this field.

In the present study, efforts have been made to address the issues of the expression variations and devise an expression invariant face recognition system using SVM and k-Means classifier.

2. Related Work

In this section, record of earlier prominent research work related to this study is presented. The review is mainly for expression invariant techniques. Table1 gives a brief overview of the work related to the present study.

Table 1. Overview of State of Art of Expression Invariant Techniques

Techniques	Researchers
Expression-Invariant Techniques	Liu *et al.* (2003), Ramachandran *et al.* (2005), Lee *et al.* (2008), Tasai *et al.* (2009), Hsieh *et al.*(2009), Petpairote *et al.* (2013), Varma *et al.* (2014), Biswas et. al. (2014), Patil et. al. (2016)

Keeping in view the research gap identified from the related work, a method for expression invariant face recognition is proposed . After a brief overview of the related work, in the forthcoming section, methodology adopted for carrying out study is presented.

3. Material and Method

The aim of this study is to design a single image based face recognition system which converts an expression face to neutral face and also to recognize to which individual the face belongs to. The major steps for present investigation are depicted in Figure 2. The various steps involved in the implementation of system architecture are as follow:

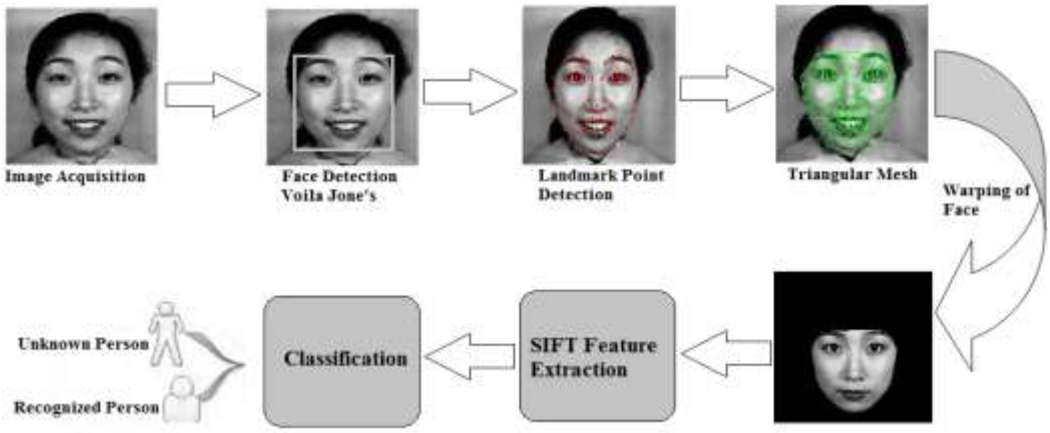

Figure 2. Flow Diagram of Present Investigation

3.1. Face Detection

After preprocessing of the facial image, face is detected from the database image. The work done by Voila and Jones in 2000's has being a most efficient and successful face detector which is able to run in real time also. In a given test image, the goal of face detection algorithm is to detect whether a face is present or not. If the face is present it returns the locations of the face. In this study, basic Adaboost Viola and Jones algorithm is used for face detection which is the fastest and most successful approach for real time systems [23].

3.2. Detection of Landmark Points

Landmark points or fiducial points are pixel coordinates which define geometry of face. These coordinate points are a set of facial salient points usually located on the corners and outer mid points of the lips, corners of the eyebrows, corners of the eyes, corners of the nostrils, tip of the nose, and the tip of the chin [12]. In previous works, authors using model based face recognition algorithms generated a grid over a face and for that some reference points (on an average 64-80 points) were required and researchers marked those points manually or semi automatically by Bunch Graph Matching algorithm [14].

In this study, a facial image with 81 landmark points is chosen having 19 for chin, 18 for lips (which includes mouth corners, upper lip and lower lip), 10 for nose, 18 for both the eye and 16 for both the eyebrows as illustrated in the Figure 3.

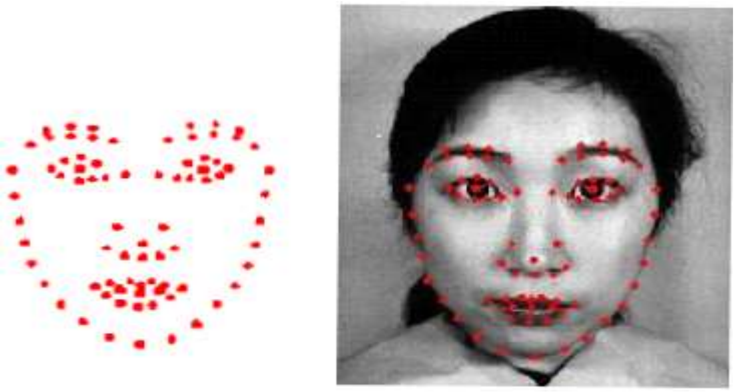

Figure 3. 81 Landmark Points on Reference Face Image for JAFFE Dataset

3.3 Facial Structure: Triangular Mesh

A set of triangular accord between sets of facial landmarks are used to pre-compute a triangular mesh for all the faces. This mesh is created by using Delaunay triangulation, which is a unique construction where no vertex from any triangle may lie within the circumcircle of any other triangle. Once a general facial mesh has been generated unique facial meshes can be created for all faces.

3.4 Interpose between Landmark Points

Facial mesh formed in previous step on the reference image and on the test image can now be compared. To morph an image between two resulting meshes interpose an intermediate facial mesh and wrap both test and reference faces to it. The intermediate facial mesh is generated by weighing between both sets of facial landmarks using equation (1): for all landmarks i

$$l_{m,i} = (1 - t) * l_{r,i} + t * l_{t,i} \tag{1}$$

where $l_{m,i}$ is the ith interpolated mid-point landmark, $l_{r,i}$ is ith reference image landmark, $l_{t,i}$ is *ith* test image landmark, t is weight between 0 and 1, where zero represents reference image's facial landmarks and one represents the test image's landmark points.

3.5 Warping Towards Intermediate Mesh

After calculating an intermediate facial mesh, the reference and test image are warped using thin plate splines based warping technique. The Thin Plate Spline (TPS) method was made popular by Fred L. Bookstein in the context of biomedical image analysis [16]. The process of using thin plate splines in image warping involves minimizing a bending energy function for a transformation over a set of given nodes or landmark points. While warping an image, all landmark points will be used to calculate the bending energy function which is then used to interpolate and transform the pixels into a warped image. Figure 4 shows the geometry change, when warping a facial expression (fear) to a reference face. The purpose to use this algorithm is to transform an expression face image to neutral face by mapping the landmarks of expression face to one preset standard face of neutral landmarks.

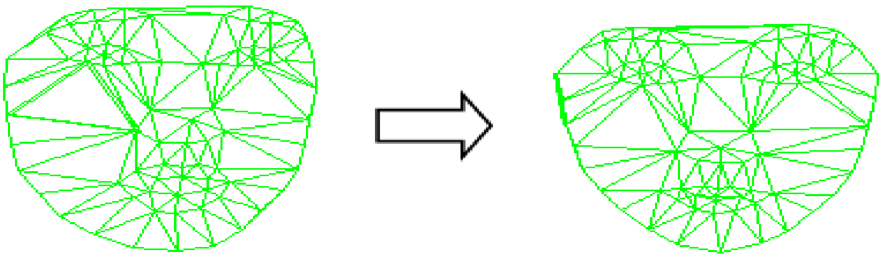

Figure 4. Geometry Change, when Warping a Facial Expression (Fear) to a Reference Face

Thine plate spline warping (TPSW) method estimates the new coordinates that minimizes energy between expression points and the neutral reference points given by the energy function in (9) that includes the smoothing term as given in equation (2):

$$I_f = \iint_{R^2}\left(\left(\frac{\partial^2 f}{\partial x^2}\right)^2 + 2\left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 + \left(\frac{\partial^2 f}{\partial y^2}\right)^2\right)dxdy \qquad (2)$$

The basis for solving the algorithm is given by a kernel function U using equation (3):

$$U = r^2 \log(r^2), \text{ where } r = (x_a - x_b)^2 + (y_a - y_b)^2 \qquad (3)$$

Given a set of source landmark points, we define P as a matrix of (3 * n) where n is the number of landmark points as depicted in equation (4):

$$P = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \dots & \dots & \dots \\ 1 & x_3 & y_3 \end{bmatrix}, 3 * n \qquad (4)$$

Using kernel function, let K be a n*n matrix defined by equation (5) :

$$K = \begin{bmatrix} 0 & U(r_{12}) & \dots & U(r_{1n}) \\ U(r_{21}) & 0 & \dots & U(r_{2n}) \\ \dots & \dots & \dots & \dots \\ U(r_{n1}) & U(r_{n2}) & \dots & 0 \end{bmatrix}, n * n \qquad (5)$$

Let Z is a 3*3 zero matrix as shown in equation (6):

$$Z = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, 3*3 \qquad (6)$$

Finally, a matrix L is defined by combination of K, P, P^T and Z as given in equation (7):

$$L = \begin{bmatrix} K & P \\ P^T & Z \end{bmatrix}, (n+3) * (n+3) \qquad (7)$$

Where, P^T is the matrix transpose operator and L is a $(n+3) * (n+3)$ matrix

The matrix L allows us to solve the bending energy equation. Inversing the matrix L on to another matrix Y, which is defined as $Y = (V|0\ 0\ 0)^T$, which is a column vector of length $(n+3)$ where V is defined as any n-vector $V = (v_1, v_2, \dots .. v_n)$. We derive a vector of n weights as $W = (w_1, w_2, \dots . w_n)$ and the coefficients a_1, a_x, a_y of affine transformation as shown in equation (8).

$$L^{-1}Y = (W|a_1\ \ a_x\ \ a_y)^T \qquad (8)$$

The final smoothing TPSW is obtained by minimizing the energy function given in equation (9) where parameter λ is smoothness constraint weight and I_f is smoothing term given in equation (2).

$$E_{TPS} = \sum_{i=1}^{n}(v_i - f(x_i, y_i)^2) + \lambda I_f \quad (9)$$

After obtaining an intermediate mesh from both reference and test faces, we interpolate between the faces to obtain a mixed face and finally we obtain a synthesized neutral face [16].

3.6 Feature Extraction and Classification

Feature extraction involves reduction of amount of resources required to explain a large set of data. Also initial data is suspected to be redundant, then it can be reduced to form a set of features. In the previous step, after obtaining a synthesized face, SIFT [19] is used for feature extraction with the advantage that SIFT features posses a strong robustness to illumination, pose and occlusion variations. It is widely and effectively used method for face recognition as it out performs Principal Component Analysis (PCA) [18], Independent Analysis component (ICA) [19], Linear Discriminant Analysis (LDA) [20] *etc.*

For the purpose of classification Support Vector Machine (SVM) is used as it is strongest of all classifiers specially for face recognition. SVM is a supervised machine learning method, and is a generalized linear classifier. SVM can solve the problems of high dimension and local minima effectively. The aim of SVM is to find the optimal hyperplane for classification purpose [21].

4. Description of Dataset

In this study, we compare the results by using two publically available database. Japanese Female Facial Expression (JAFFE) [15] and Cohn-Kanade (CK) databases are used. JAFFE dataset has 213 gray scale images of 10 Japanese female models with 6 facial expressions + 1 neutral face. Size of each image is 256 * 256. Figure 5 shows the sample images of 10 subjects in the database.

Figure 5. Facial Images of 10 Subjects in the JAFFE Database [15]

CK database includes 486 sequences from 97 posers. Each sequence begins with a neutral face and proceeds to the peak expression. Figure 6 shows the sample images of the CK database.

Figure 6. Sample Images of Cohn-Kanade Database

5. Result and Discussion

The test and training images in database are first normalized to have same scale, position and rotation to mark landmark points on the neutral face and the test face. In this process the expression image is transformed to neutral image using warping technique. After transformation to neutral image SIFT is used for feature extraction and we compare the result using SVM and k-Means classifier for the two databases. Figure 7 depicts the neutral and expression face, cropped faces and their corresponding warped faces for the JAFFE dataset. Figure 7 (a) to (g) depicts the reference face and the expression face of the same subject and Figure 7 (h) to (m) represents the respective cropped faces. Further Figure 7 (n) to (t) show the morphed face which almost appear to be same as reference image.

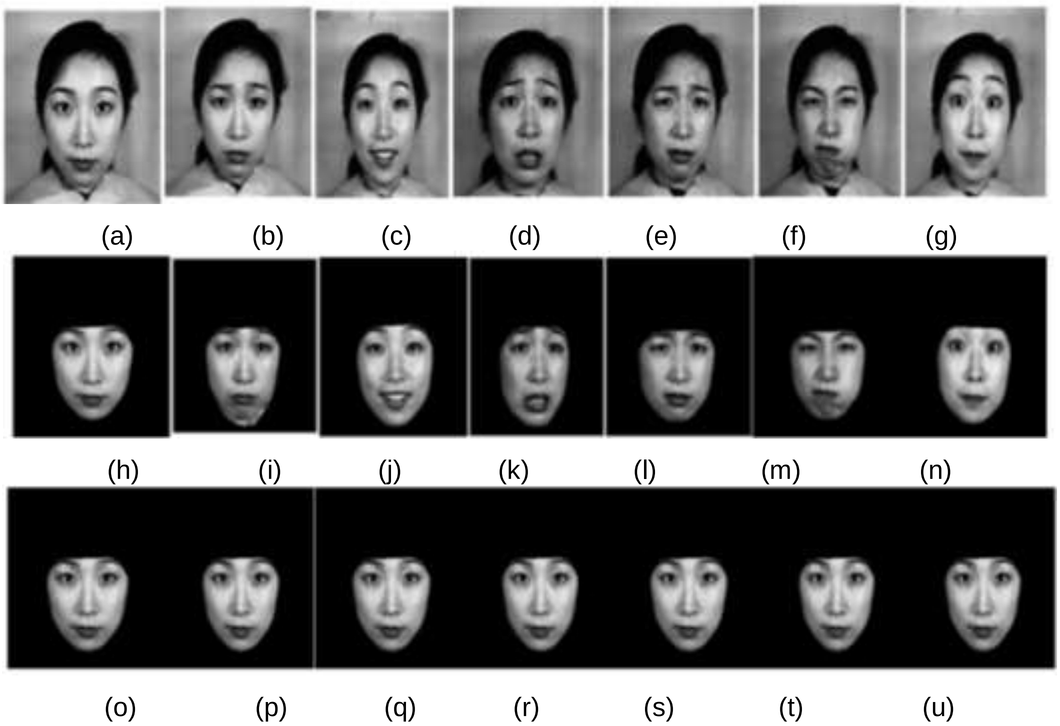

Figure 7: Examples of Warping of a Face; (a) Reference Face of Subject 1; (b), (c), (d), (e), (f), (g) Expression Face of Subject 1; (h), (i), (j), (k), (l), (m), (n) Cropped Images of Same Individual; (o), (p), (q), (r), (s), (t), (u) are the Images Warped to Neutral Face

Table 2 shows the recognition accuracy of different expression by using SVM and k-Means classifiers for JAFFE and CK datasets. It is observed that for the JAFFE database anger has the highest recognition accuracy of 100% which is followed by disgust which is much closer to anger *i.e.* 99.4% for SVM and 54.1% for k-Means and expression fear and surprise has the lowest recognition rate of 95.3% and 94.68% respectively for SVM and 55.6 and 53.8 respectively for k-Means, the recognition accuracy falls for different expressions because of large variations in the region of eyes and mouth of the face.

Table 2. Recognition Accuracy Using SVM Classifier and K-Means Classifier

Database	JAFFE		CK	
Expression	SVM (%Accuracy)	k-Means (%Accuracy	SVM (%Accuracy)	k-Means (%Accuracy)
Anger	100	61.8	99.14	63.4
Disgust	99.4	54.1	99.8	62.1
Fear	95.3	55.6	95.4	59.8
Happy	98.9	58.7	98.5	61.1
Sad	96.8	52.4	96.3	58.5
Surprise	94.68	53.8	94.8	57.3

On the similar basis for the CK dataset, disgust and anger has the highest recognition rate of 99.8% and 99.14% respectively for SVM. The lowest recognition accuracy of surprise is 94.8%. The average recognition rate of JAFFE database is 97.4% whereas of CK database is 97.8% for SVM classifiers. For k-Means the recognition rate is 56.8% and 60.3% for JAFFE and CK database respectively.

5.1. Comparison with State-of-art Algorithms

From the JAFFE dataset, neutral image is chosen as the reference image per person and rest of the images as test images. In Figure 8, the result obtained from the investigated work is compared with other techniques. The recognition rates of SVD, LLE-Eigen, FLD-PCA-ANN [13] and method in [8] are 92.96%, 93.93%, 84.9 and 96.6% respectively whereas for the proposed method the recognition rate is 97.3% which outperforms all the other methods discussed.

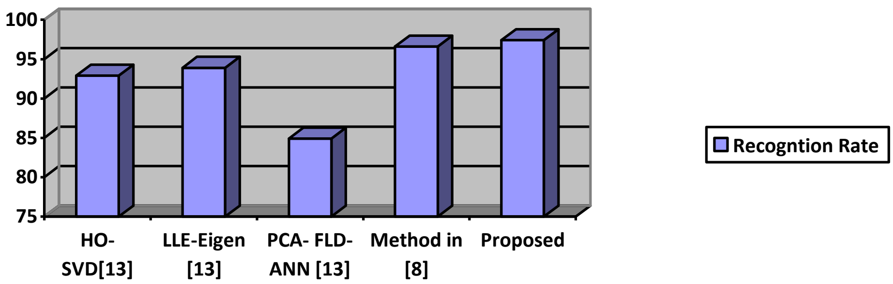

Figure 8. Performance Comparison on JAFFE Dataset

Similarly for the CK database, neutral image is chosen as the reference image per person and rest of the images are used for testing purpose. In Figure 9, the result obtained from the proposed work is compared with the state-of-art algorithms. The highest recognition rate is 97.8% which is offered by proposed method. Hence the proposed approach outperforms state-of-art methods with a wide range of expression variations in the image.

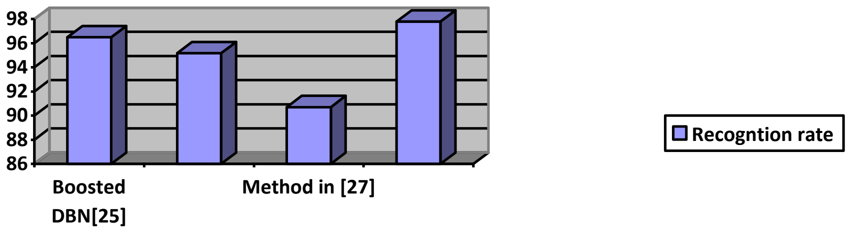

Figure 9. Performance Comparison on CK Dataset

6. Conclusion

In this investigation, a face recognition system which is invariant to facial expressions is presented. The present method converts an input face image with any arbitrary expression into its corresponding neutral facial image. The warped image is classified using SVM and k-means classifier and the results are compared with the reference face image and recognition accuracy is determined for different expressions of JAFFE and CK database. The result obtained from the investigated work is compared with state-of-art algorithms. For the proposed method the recognition rate is 97.3% for JAFFE database and 97.8% for CK database which outperforms the other methods. The work presented here requires only single image for training, which make it useful in security and surveillance applications. In summary, the present work is an attempt made to make a face recognition system invariant to expressions.

References

[1] S. Chintalapati and M. V. Raghunadh, "Illumination, Expression and Occlusion Invariant Pose-Adaptive Face Recognition System for Real-Time Application", International Journal of Engineering Trends and Technology(IJETT), **(2014)**.

[2] R. Chellappa, C. L. Wilson and S. Sirohey, "Human and machine recognition of faces: A survey", Proceedings of the IEEE, vol. 83, no. 5, **(1995)**, pp. 705-740.

[3] R. Jafri and H. R. Arabnia, "A Survey of Face Recognition Techniques", Journal of Information Processing Systems, vol. 5, no. 2, **(2009)**, pp. 41-68.

[4] W. Zhao, R. Chellappa, A. Rosenfeld and P. J. Phillips, "Face Recognition: A Literature Survey", ACM Computing Surveys, vol. 35, no. 5, **(2003)**, pp. 399-458.

[5] M. Ramachandran, S.K. Zhou, D. Jhalani, and R. Chellappa, "A Method for Converting a Smiling Face to A Neutral Face with Application to Face Recognition," Proceedings of conference ICAASSP, pp. ii/977-ii/980, vol. 2, **(2005)**.

[6] Y. Liu, K. L. Schmidt, J. F. Cohn, and S. Mitra, "Facial asymmetry quantification for expression invariant human identification," Computer Vision and Image Understanding: CVIU, vol. 91, **(2003)**, pp. 138-159.

[7] H.-S. Lee, and D. Kim, "Expression-Invariant Face Recognition by Facial Expression Transformations," Pattern Recognition Letters, vol. 29, no. 13,**(2008)**, pp. 1797-1805.

[8] P. Tasai, T.P. Tran, and L. Cao, "Expression-Invariant Facial Identification," Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, **(2009)**, pp. 5151-5155.

[9] C.-K. Hsieh, S.-H. Lai, and Y.-C. Chen, "Expression-Invariant Face Recognition with Constrained Optical Flow Warping," IEEE Transaction on Multimedia, vol. 11, no. 4, **(2009)**, pp.600-610.

[10] C. Petpairote and S. Madarasmi, "Face Recognition Improvement by Converting Expression Faces to Neutral Faces", ISCIT Conf., **(2013)**, pp. 439 – 444.

[11] Varma Rahul, Sandesh Gupta, and Phalguni Gupta. "Face recognition system invariant to expression." In International Conference on Intelligent Computing, Springer International Publishing, **(2014)**, pp. 299-307.

[12] Z. Zeng, M. Pantic, G. I. Roisman, T. S. Huang, "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions", IEEE Transactions on Pattern Analysis and Machine Intelligent, vol. 31, no. 1, **(2009)**.

[13] Patil, H.Y., Kothari, A.G. and Bhurchandi, K.M, "Expression invariant face recognition using semidecimated DWT, Patch-LDSMT, feature and score level fusion", Journal Applied Intelligence, vol. 44, no. 4, **(2016)**, pp. 913-930.

[14] Wiskott, L., et al, "Face Recognition by Elastic Bunch Graph Matching", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, **(1997)**, pp. 775-779.
[15] Lyons Michael J., Akamatsu Shigeru, Kamachi Miyuki and Gyoba Jiro, "Coding Facial Expressions with Gabor Wavelets", IEEE International Conference on Automatic Face and Gesture Recognition, Nara Japan, IEEE Computer Society, **(1998)**, pp. 200-205.
[16] F. L. Bookstein, "Principle Warps: Thin-Plate Splines and the decomposition of deformations, IEEE Trans. on Pattern Analysis and Machine Intelligence", vol. 11, no. 6, **(1989)**.
[17] Lowe D., "Distinctive image features from scale-invariant key-points", Int. Journal of Computer Vision, vol. 60, no. 2, **(2004)**, pp.91-110.
[18] Turk M. and Pentland A., "Eigen faces for recognition", Journal of Cognitive Neuroscience, vol.3, no.1, (1991), pp.71-86.
[19] Hyvärinen Aapo and Oja Erkki, "Independent Component Analysis: Algorithms and Applications", Neural Networks Research Centre Helsinki University of Technology P.O. Box 5400, Finland, Neural Networks, 13(4-5), **(2000)**, pp. 411-430.
[20] J. Lu, K. Plataniotis and A. Venetsanopoulos, "Face recognition using LDA based algorithms", IEEE Transactions on Neural Networks, vol. 14, no. 1, **(2003)**, pp. 195-200.
[21] J. C. B. Christopher, "A tutorial on support vector machines for pattern recognition", Data mining and knowledge discovery, vol. 2, no. 2, **(1998)**, pp. 121-167.
[22] P. Kamencay M. Brenzan, D. Jelsovka and M. Zachariasova, "Improved Face Recognition Method based on segmentation Algorithm using SIFT-PCA", IEEE International Conference on Telecommunication and Signal Processing, **(2012)**, pp. 758-762.
[23] P. Viola and J. J. Michael, "Robust real-time face detection", International Journal of Computer Vision, vol.57, no. 2, **(2004)**, pp. 137-154.
[24] A. Biswas and M. K. Ghose, " Expression Invariant Face Recognition using DWT SIFT Features", International Journal of Computer Applications, vol. 92, no. 2, **(2014)**, pp. 30-32.
[25] P. Liu, S. Han, Z. Meng and Y. Tong, "Facial Expression Recognition via a Boosted Deep Belief Network", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA**, (2014)**, pp. 1805–1812.
[26] D. Ghimire and J. Lee, "Geometric Feature-Based Facial Expression Recognition in Image Sequences Using Multi-Class AdaBoost and Support Vector Machines", Sensors, vol. 13, **(2013)**, pp. 7714–7734.
[27] X. Jia, Y. Zhang, D. Powers and H. B. Ali, "Multi-classifier fusion based facial expression recognition approach", KSII Trans. Internet Inf. Syst., vol. 8, **(2014)**, pp. 196–212.

Authors

Deepti Ahlawat, Received her M.Tech degree in 2008 from N.C. College of Engg., Israna, Panipat under Kurukshetra University, Kurukshetra, India. She is now a Ph.D. scholar at BPSMV,Khanpur Kalan, Sonipat, India. Her research interest includes Pattern recognition and Image Processing.

Vijay Nehra, earned Ph.D degree in Electronics and Communication Engineering in 2009 from Maharshi Dayanand University, Rohtak. He is currently working as professor in the Electronics and Communication Engineering Department at the Faculty of Engineering and Technology, Bhagat Phool Singh Mahila Vishwavidayala, Khanpur Kalan, Sonipat, Haryana, India. Dr. Nehra has a professional experience of 14 years in teaching, research, curriculum planning, laboratory development, educational administration, planning, management and execution. He has over 45 publications in refereed journals and proceedings along with many articles in the field of engineering and engineering education. He is a life member of various professional societies such as ISTE, CSI, IETE, Institution of Engineers, Plasma Science Society of India.

