Recognizing faces with single sample per subject using fusion of transforms

Sujata G. Bhele
Priyadarshini College of Engineering, CRPF gate, Nagpur
sujata_bhele@yahoo.co.in

Vijay H. Mankar
Government Polytechnic, Bramhapuri
vhmankar@gmail.com

Abstract--Face recognition has attracted attention of the researchers. Face recognition becomes challenging if various factors are considered such as varying illumination, pose, facial expression and somewhat occlusion. The face recognition becomes more challenging if the single sample per person is available. In this paper a fusion of method is proposed to deal with single sample per person. Gabor transform is good for eliminating the orientation differences. An efficient ridgelet transform is proposed which effectively collects the meaningful rotational features. The results obtained from these two transforms are combined to classify the face image using support vector machine and distance based classifier. Experiments on FEI, JAFFE, ORL, UMIST, MUCT face datasets shows that the proposed method improves the performance in the scenario of one training sample per person.

Keywords- Gabor Wavelet, Feature extraction, Support Vector Machine

I. INTRODUCTION

Face recognition has been studied for last 30 years. It is widely used in computer vision area due to large number of applications such as security access control, personal identification to human computer communication. To develop a facial expression recognition system Gabor filter and learning vector quantization method was used by Shishir Bashyal et al. (2008). Gabor filter was used to extract the features from the JAFFE database images [1]. For extracting features Caifeng Shan et al. (2009) used Linear Binary Pattern and Support Vector Machine (SVM) as a classifier. This method is invariant to illumination changes and takes less time in comparison to other existing system. Linear Binary Pattern (LBP) is tolerant to changes in illumination and takes less time in computation [2]. Dahmane M. et al. (2011) proposed a general expression model (even when a person poses at different time). They used Histogram of oriented gradients (HOG) and Support Vector Machine (SVM). HOG is used to extract the appearance based features by gradient magnitudes considering a set of orientations [3]. Wenfei Gu. et al. (2012) used Gabor filter for feature extraction and classifier synthesis. The system is tested on person-independent expression recognition and variation in illumination. Gabor filters have the important property of robustness against slight object rotation, distortion and variation in illumination [4].

In this paper an efficient method for face recognition is proposed by combining the two transforms. Gabor transform is used to extract orientational features and an efficient implementation of ridgelet transform is presented to extract both directional and rotational features. These directional, rotational features are then combined to provide to a classifier. The advantage of the system is that above mentioned features are used to test the single sample per person problem.

The rest of this paper is organized as follows. Section 2 describes the proposed algorithm in detail. Section 3 describes classification methods. Section 4 performs extensive experiments on the well-known ORL, UMIST, JAFFE, FEI and MUCT face databases. Section 5 concludes the paper.

II. PROPOSED METHOD

In this work Gabor features are extracted from the face image. These features are then combined with the features obtained from efficient ridgelet transform which gives directional features and then it is given to classifier to classify the image.

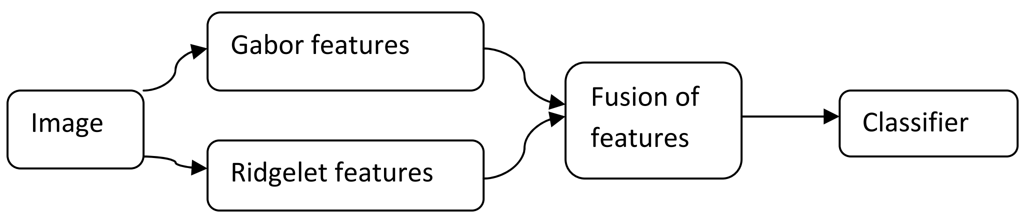


Figure1. Block Diagram of proposed method

A. Gabor Feature Extraction

Gabor filter is used as a feature generator in face recognition [5,6]. In the spatial domain, a two-dimensional Gabor filter is a Gaussian kernel function[7] modulated by a complex sinusoidal plane wave, defined as:

$$G(x,y) = \frac{f^2}{\pi\gamma\vartheta}\exp\left(-x'^2+\gamma^2y'^2\right)/2\sigma^2)\exp\left(j2\pi f x' + \emptyset\right) \quad (1)$$

$$x' = x\cos\theta + y\sin\theta \quad (2)$$

$$y' = -x\sin\theta + y\cos\theta \quad (3)$$

Where *f is* the frequency of the sinusoidal factor, θ represents the angle, Φ is the phase offset, σ is the standard deviation and γ is the spatial aspect ratio. Gabor wavelet filter is used on whole image to extract the features from the face image. 1D feature vector is obtained from this feature extracted image and used for further processing. The face image is again divided into nine non-overlapping sub-images. The local features extracted are used to overcome the changes in some region of face. Feature vectors are extracted from each part and these are combined to form a single feature vector. Gabor features explains two desirable characteristic: spatial locality and orientation selectivity as in Fig 2

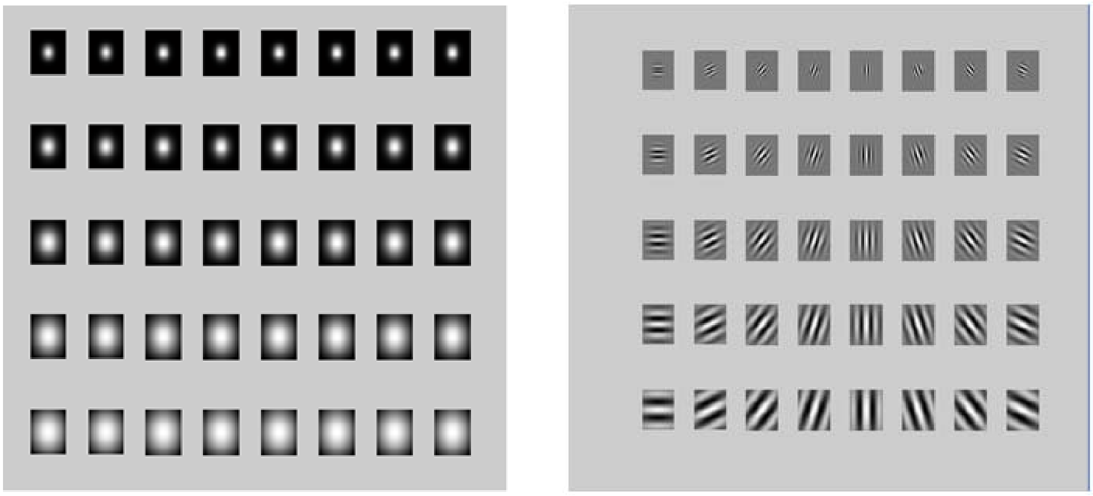


Figure 2. Gabor wavelets at five scales and eight orientations: (a) the magnitude of Gabor wavelet. (b) the real part of Gabor wavelet

$$f_g = f_{max}/\sqrt{2}^g$$

$$\theta_h = h\pi/8$$

$$\varphi_{g,h}(x,y) = \varphi_\pi(f_g, \theta_h, \gamma, \vartheta)$$

The value of parameters in Eq.(2) are given by $\gamma = \eta = 2$, *fmax* = 0.25 with $g \in \{0,\ldots,3\}$, $h \in \{0,\ldots,7\}$, generally. In the present approach both scale and orientation are taken as $(g,h) \in \in \{0,\ldots,7\}$, to extract more meaningful and discriminative features.

a. Feature Extraction from Gabor Wavelets

Each component of the 2D Gabor wavelet (at different scales and orientations) is convolved with the image given by

Eq.(4). Downsampling is performed by a factor of 4 in rows and columns of the absolute value of the filtered image. The feature vector is obtained by rearranging the downsampled value. The vector is normalized to have zero mean and unit variance. If $O_{g,h}$ denotes the feature vector from the filtered image at scale *g* and orientation *h*, then final feature (*F*) value in R_d is given by Eq.(5) [8]. For m×n image, the value of feature space dimension, d = m×n×g×h/(4×4).

$$O_{g,h} = I(x,y) * \varphi_{g,h}(x,y) \quad (4)$$

$$F = (O_{0,0}^T \, O_{0,1}^T \ldots . O_{7,7}^T)^T \quad (5)$$

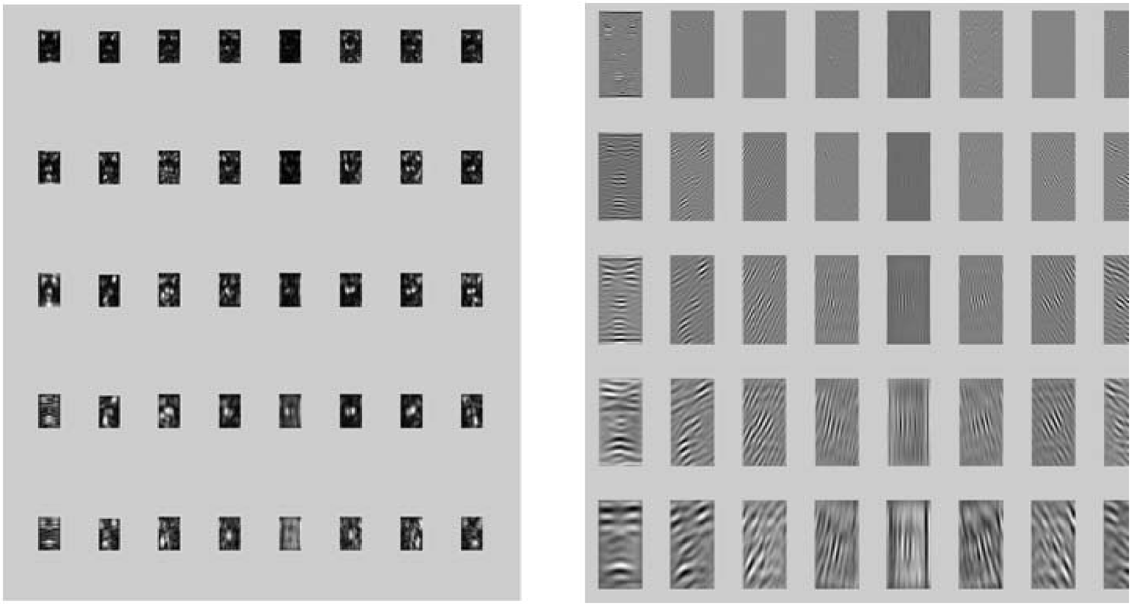


Figure 3. Gabor representation of a face image: (a) the magnitude; (b) the real part.

B. Ridgelet Transform

In this paper, we describe a system of orthonormal bases and frames for two dimensional digital data I(x, y). The system we describe Ridgelet packets, has many points in common with ridgelet analysis, while allowing for a more complex and oscillatory structure. Thus, the system contains bases, while allowing for a more complex and oscillatory structure. Thus, the system contains bases where the basis elements are, like ridgelets, highly orientation- selective, but which have greater degrees of oscillation either along or across the direction of primary orientation.

In this paper we described a general construction that contains the idea behind the ridgelets construction as a special case. The result will be a family of bases with a variety of interesting space-frequency localization properties. By modifying the ridgelet construction using wavelet packets in the ridge direction and wavelets in the angular direction, one induces on real space elements which are oscillatory ridgelets in the sense that they have angular localization features similar to the orthonormal ridgelet family – being nearly ridge fuctions –while being oscillatory in the ridge direction. On the other hand, by using wavelets in the ridge direction together with cosine packets in the angular direction one can produce effects which are more like brush strokes – bundles of line elements with given orientation, position, and textural cross-section.

$$R(x,y) = f^{-1/2}\,((x' + y') - b)/a) \tag{6}$$

$$x^{'} = xcos\theta + ysin\theta \tag{7}$$

$$y' = -xsin\theta + ycos\theta \tag{8}$$

$$f = f_{max}/\sqrt{2}^{g}$$

$$b = f/\sqrt{2}$$

$$a = f/\sqrt{2}$$

Each component of the R(x,y) at different scales and orientations is convolved with the image given by Eq.(9). If $O_{g,h}$ denotes the feature vector from the filtered image at scale *g* and orientation *h*,

$$O_{g,h} = I(x,y) * \vartheta_{g,h}(x,y) \tag{9}$$

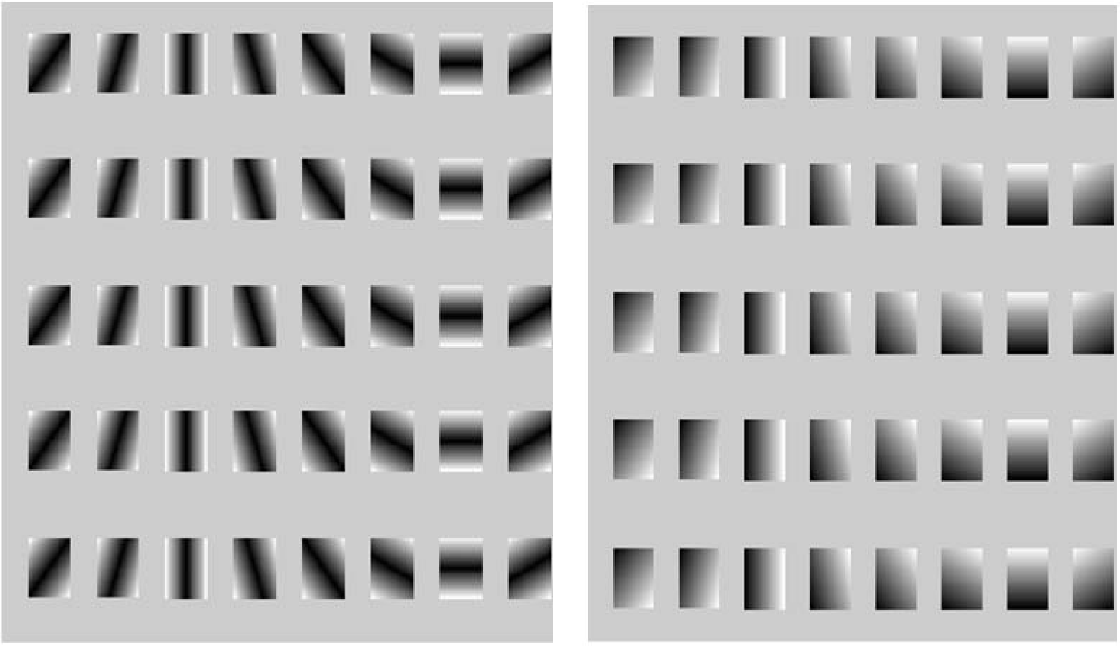


Figure 4. Ridgelet at five scales and eight orientations: (a) the magnitude (b) the real part.

III. CLASSIFIER

A. Support vector machines

After the feature extraction, we have designed a non-linear multi-class support vector machines (SVMs) to classify and recognize the image samples. The support vector machines were designed for binary-class classification problems [9,10]. Various binary-class SVMs can be combined to form multi-class SVMs for multi-class classification problems, like face recognition problem.

a. Multi-class support vector machines

Support vector machines are designed for binary pattern classification. Multi-class pattern recognition problems are commonly solved using a combination of binary SVMs and a decision strategy to decide the class of the input pattern. Each SVM is independently trained. Multi-class SVM can be implemented using one-against-all [10] and one-against-one [11] strategy. In this paper , we have implemented one-against-all strategy due to its less memory requirement, as discussed below. Let the training set (xi, ci) consists of N samples of M classes, where ci (ci ∈ 1, 2, . . .,M) represents the class label of the sample xi. An SVM is constructed for each class by discriminating that class against the remaining(M−1) classes. The number of SVMs used in this approach is M. A test pattern x is classified by using the winner-takes-all decision strategy, i.e., the class with the maximum value of the discriminant function f(x) is assigned to it. All the N training samples are used in constructing an SVM for a class. The SVM for class k is constructed using the set of training samples and their desired outputs,(xi, yi). The desired output yi for a training sample xi is defined as follows:

$$f(x) = \begin{cases} +1, & if\ c_i = k \\ -1 & if\ c_i \neq k \end{cases} \qquad (10)$$

The samples with the desired output yi = +1arecalled positive samples and the samples with the desired output yi =−1 are called negative samples.

B. Distance Classifiers

Most face recognition methods use distance measure to classify the face image. Images are projected to a lower-dimensional feature space. The distance between the feature space representation can be used as a basis for face recognition. Images are projected into an eigen space and is represented as vectors. The distance between the vectors of two images is the similarity of the images. The dissimilarity distance between the image projection and known projections is calculated, the face image is then classified as one of the faces with minimum distance.

a .Euclidean Distance Classifier

Euclidean distance is the most commonly used distance measure. The Euclidean distance between the image projection and known projections is calculated; the face image is then classified as one of these faces with minimum Euclidean distance. The Euclidean distance, d2 as follows:

$$d_2(x,y) = \sum_{i=1}^{l}(x_i - y_i)^2 \qquad (11)$$

where x, y in the data set X and xi, yi are the ith coordinates of x and y, respectively. This is a dissimilarity measure on X. The minimum possible distance between two vectors of X is 0 that is d0 = 0 (the distance of a vector
from itself).

b. City-Block Distance Classifier

City-Block Distance Classifier, Manhattan Distance Classifier, also called, rectilinear distance, L1 distance, L1 norm, Manhattan length.. It examines the absolute differences between the coordinates of a pair of objects as follows:

$$d_1(x,y) = \sum_{i=1}^{l} | x_i - y_i | \tag{12}$$

c.Cosine distance classifier

$$d_3(x,y) = \cos(\theta) = x.y/(\|x\|.\|y\|) \tag{13}$$

d. Correlation distance classifier

$$d_4 = \sum_{i=1}^{l}(1 - cor(\beta)) \tag{14}$$

ß- sample correlation between points

IV. EXPERIMENTS

The experiments are done using various databases such as : UMIST[12], ORL[13], JAFFE[14], FEI[15], MUCT[16].

JAFFE, MUCT and FEI database are used for expression variations; the ORL database is used for size and rotation variations; and the UMIST database is selected for pose variations.

The ORL database contains samples from 40 individuals, with 10 different subjects. For some subjects, the images were taken at different times. The facial expressions (open or closed eyes, smiling or non-smiling) and occlusion (glasses or no glasses) also change. There is also some variation in the scale of up to 10%. Fig. 6 shows five sample images of one person in the ORL database.

Figure6. Some images of one subject in the ORL database.

The UMIST database was built by the University of Manchester Institute of Science and Technology. It consists of 575 images from 20 people, each covering a wide range of poses from profile to frontal views. In the experiment, 19 poses per person were selected to form a new sub-database of 380 images. Fig. 7 shows some images of one person.

Figure7. Some images of one subject in the UMIST database.

JAFFE database have 210 images of frontal face showing seven emotions (Anger, Disgust, Fear, Happy, Sad, Surprise and Neutral) posed by 10 Japanese female models. Each emotion class has 30 images.

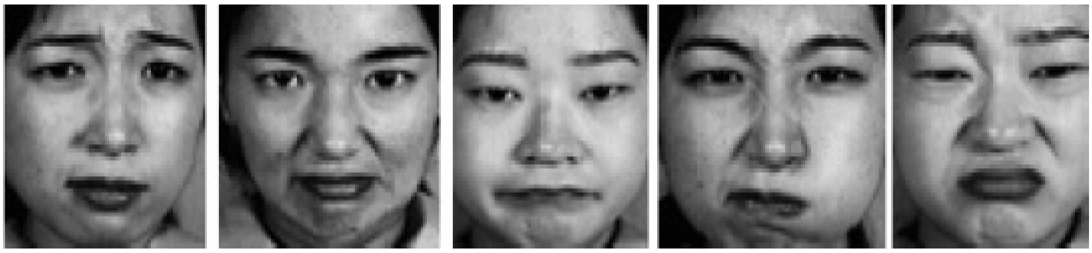


Figure 8.Some test images from JAFFE database

The MUCT database consists of 3755 faces with 76 manual landmarks. The database was created to provide more diversity of lighting, age, and ethnicity

Figure 9 Sample face images from MUCT database

FEI face database composed of only frontal face images previously aligned to a common template . Since the number of subjects is equal to 200 and each subject has two frontal images (one with a neutral or non-smiling expression and the other with a smiling facial expression), there are 400 full frontal face images manually registered to evaluate experiments on a controlled environment.

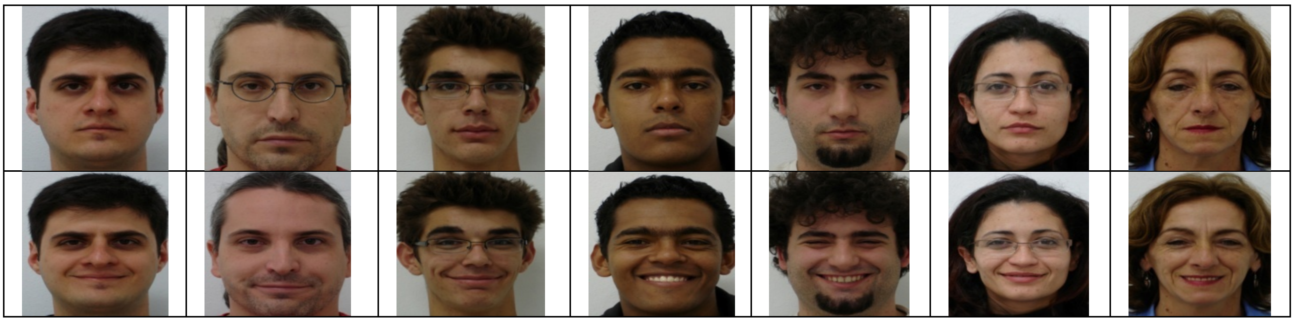
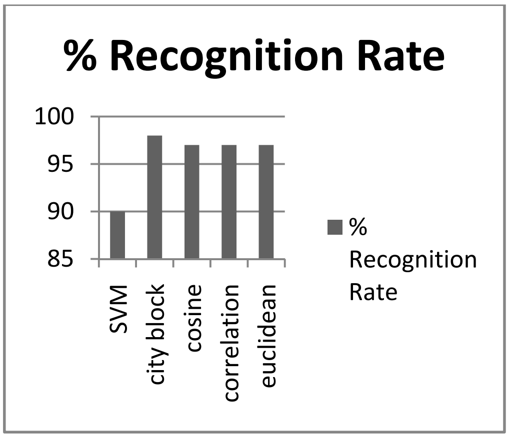


Figure 10 Sample face images from FEI database a)Training sample of some subjects b)Test samples of some subjects

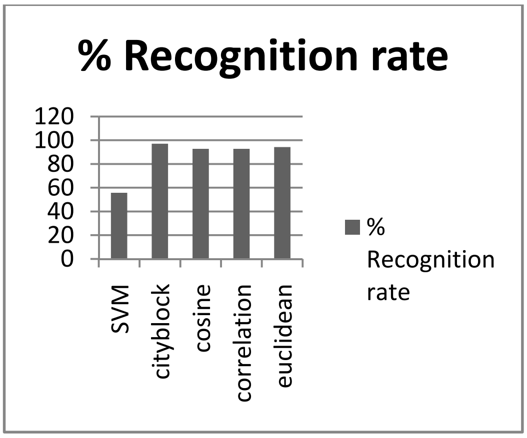
From all the databases the first image of each subject was selected for training and rest of the images for used for testing. In this paper, combined features from two algorithms are used to extract discriminant features. Alongwith SVM, various distance based classifiers are used to classify the input image. Three methods, which were proposed to solve the single training sample per person problem, are used for comparison. The first method is the projection approach [17]; the second is the singular value perturbation approach [18]; and the third is to divide an image into several non-overlap blocks [19]. In all the four schemes, 2D-FLDA is used to extract the discriminant features and the nearest classifier is used to classify the image. All the parameters in the three comparison algorithms are set as the same as those in [17,18] and [19], respectively.

Fig. 11 plots the recognition accuracies by different classifiers on the five databases.

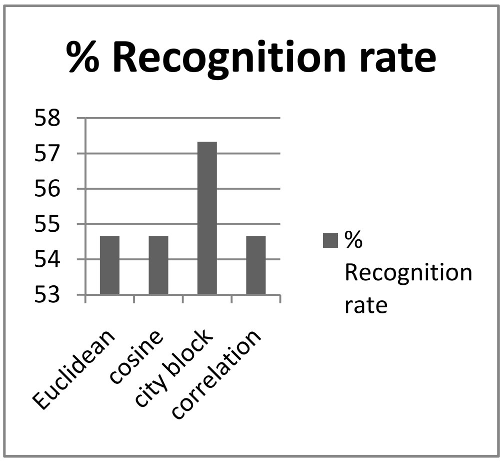
Table 1 lists the top recognition accuracy of the four methods. On ORL database, the top recognition accuracies are 44.17%, 46.39%, 70.83% and 75.556% for the four methods, respectively. On UMIST database, the top recognition accuracies are 18.61%, 52.5%, and 57.22% respectively. It can be seen that the proposed method achieves higher recognition accuracy than the other four methods.



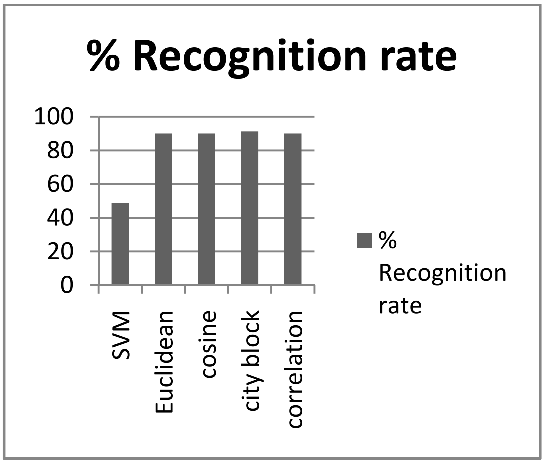
a)



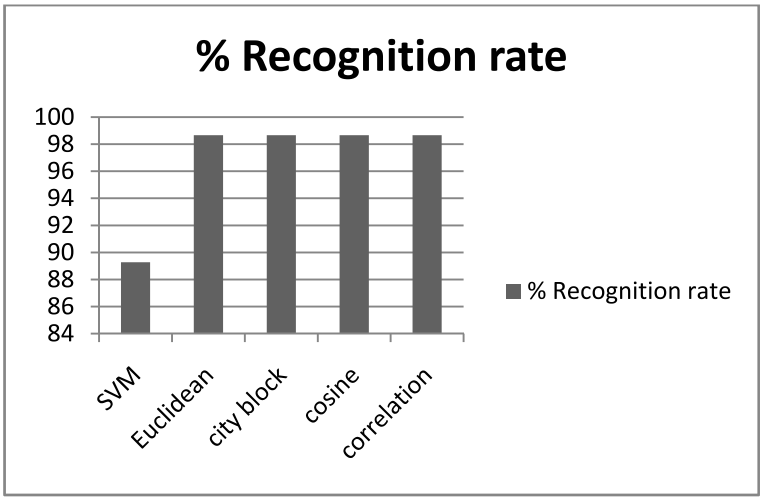
b)



c)



d)



e)

Figure.11 Recognition rate using various classifiers for various databases
a) FEI database b) JAFFE database c) UMIST database d) ORL database e) MUCT database

Table 1.The top recognition accuracies of the four methods on ORL databases

Database	Method in [17]%	Method in [18]%	Method in [19]%	Method in [20]%	Proposed Method
ORL	44.17	46.39	70.83	75.56	91.25

Table 2. The top recognition accuracies of the methods on UMIST databases

Database	Method in [17]%	Method in [18]%	Method in [19]%	Proposed Method
UMIST	18.61	52.50	57.22	57.33

Table 3 The top recognition accuracies of the proposed method by using various classifiers

Database	ORL(%)	UMIST(%)	MUCT(%)	FEI(%)	JAFFE(%
SVM	48.75	-	89.28	90	55.71
Euclidean	90	54.66	98.66	97	94.28
Cosine	90	54.66	98.66	97	92.85
City Block	91.25	57.33	98.66	98	97.14
Correlation	90	54.66	98.66	97	92.85

V. CONCLUSION

This paper proposes an efficient method of face recognition which combines the two transforms. This method is invariant to varying pose, illumination, facial expression and slight occlusion. In this approach the input face is decomposed using gabor wavelet transform and an efficient ridgelet transform and the combination of features are used to represent the face. The recognition of faces is done using various distance measures, discriminant analysis and SVM classifier. The proposed method is efficient for single sample per person condition. From the results it can be concluded that the present system efficiently classifies the face using distance based classifiers.

REFERENCES

[1] Shishir Bashyal et al., “Recognition of facial expression using Gabor wavelet and learning vector quantization” ,Elsevier Trans. ,Engineering Application of Artificial Intelligence 21, 1056 1064, 2008
[2] Ceifeng Shan, et al., “Facial expression recognition based on Local Binary Patterns: A comprehensive Study”, Elsevier Science Ltd Trans., 2009
[3] M. Dahmane and J. Meunier, ”Emotion recognition using dynamic grid based HOG feature”, IEEE In. onf.,2011.
[4] Wenfei Gu, et al., “Facial expression recognition using radial encoding of local Gabor Feature and Classifier Synthesis”, Elsevier Ltd Trans., 2012
[5] M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, C. Von der Malsurg, R. Wurtz, and W. Konen, “Distortion invariant object recognition in the dynamic link architecture,” Computers, IEEE Transactions on vol. 42, no. 3, pp. 300–311, 2002.
[6] C. Liu, “Gabor-based kernel PCA with fractional power polynomial models for face recognition,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 26, no. 5, pp. 572–581, 2004.
[7] Mohammad Haghighat, Saman Zonouz, and Mohamed Abdel-Mottaleb, “Identification Using Encrypted Biometrics”, Springer - Verlag Berlin Heidelberg 2013, pp. 440-448.
[8] V. Struc and N. Pavesic, “Gabor-Based Kernel Partial- Least-Squares Discrimination Features for Face Recognition,” Informatica, vol.20(1), pp.115–138, 2009.
[9] C. Cortes, V. Vapnik, Support-vector network, Mach. Learn. 20 (1995) 273–297.
[10] V.N. Vapnik, Statistical Learning Theory, John Wiley & Sons, New York, 1998.
[11] S. Knerr, L. Personnaz, G. Dreyfus, Nurocosingle-Layer Learning Revisited: A Stepwise Procedure for Building and Training a Neural Network, Springer, 1990.
[12] UMIST Face Database, Daniel Graham <http://images.ee.umist.ac.uk/danny/database.html>, 2005.
[13] ORL database <http://www.cam-orl.co.uk/facedatabase.html>, 2002.
[14] M. Lyons, M. Kamachi and J. Gyoba. (1998, April). The Japanese Female Facial Expression (JAFFE) Database. [Online]. Available: http://www.kasrl.org/jaffe.html
[15] C. E. Thomaz and G. A. Giraldi. A new ranking method for Principal Components Analysis and its application to face image analysis, Image and Vision Computing, vol. 28, no. 6, pp. 902-913, June 2010.
[16] Stephen Milborrow, John Morkel,Fred Nicolls,” The MUCT Landmarked Face Database”
[17] S.C. Chen, D.Q. Zhang, Z.H. Zhou, Enhanced (PC)2A for face recognition with One training image per person, Pattern Recognit. Lett. 25 (10) (2004) 1173–1181.
[18] D. Zhang, S.C. Chen, Z.H. Zhou, A new face recognition method based on SVD perturbation for single example image per person, Appl. Math. Comput.163 (2) (2005) 895–907.
[19] S. Chen, J. Liu, Z.H. Zhou, Making FLDA applicable to face recognition with one sample per person, Pattern Recognit. 37 (1) (2004) 553–1555.
[20] Quan-xue Gao, Lei Zhang , David Zhang ,” Face recognition using FLDA with single training image per person Applied Mathematics and Computation”, 205 (2008) 726–734