World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:5, No:11, 2011

Face Recognition using Features Combination and a New Non-linear Kernel

Essam Al Daoud

Abstract**—**To improve the classification rate of the face recognition, features combination and a novel non-linear kernel are proposed. The feature vector concatenates three different radius of local binary patterns and Gabor wavelet features. Gabor features are the mean, standard deviation and the skew of each scaling and orientation parameter. The aim of the new kernel is to incorporate the power of the kernel methods with the optimal balance between the features. To verify the effectiveness of the proposed method, numerous methods are tested by using four datasets, which are consisting of various emotions, orientations, configuration, expressions and lighting conditions. Empirical results show the superiority of the proposed technique when compared to other methods.

Keywords**—**Face recognition, Gabor wavelet, LBP, Non-linear kerner

I. INTRODUCTION

OVER the past decades, great attention has been paid to face recognition task. It might be one of the most important tasks in the pattern recognition applications. There are various applications for face recognition systems such as searching surveillance images, access control, minimizing fraud by verifying identity, inmate tracking and verifying identity of a person prior to vote [1-3]. Feature extraction or descriptors is an essential step in face recognition system. Liu et. al. proposed two new sparse descriptors, they showed that by combining sparse LBP and sparse HoG more discriminative features can be extracted from a single image[4]. Prabhu and Savvides used 3D generic Elastic model to construct 3D image for each subject from 2D image. Their model can be applied to uncontrolled real-word images where variations in expression, illumination and pose are encountered [5]. Wagner et. al. aligned a test image to frontal part of the training faces by using tools from sparse representation [6]. Wolf et. al. captured local similarities by designing a family of novel descriptors. Moreover, to enhance the classification performance, they used unlabeled background samples [7]. Bartlett et. al. applied a sigmoidal neurons and a version of ICA to two different architectures, they showed that both representations are suitable for face recognition across days and changes in expression [8].

II. TOW-DIMENSIONAL GABOR WAVELET

The frequency and orientation of the tow-dimensional Gabor wavelet are similar to human visual system, it contains very rich information about the structure of the image and the texture representation. It has been successfully used with numerous image processing techniques such as fingerprint recognition, handwritten numerals recognition, face Recognition and texture segmentation[9]. 2D Gabor function is defined as follows:

$$g(x,y)=\left(\frac{1}{2\pi\sigma_x\sigma_y}\right)\exp\left(-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)+2\pi jWx\right) \quad (1)$$

The Gabor wavelet is:

$$g_{m,n}(x,y)=a^{-m}g(x',y'),\quad a>1$$
$$x'=a^{-m}(x\ \cos\ \theta+y\ \sin\ \theta)$$
$$y'=a^{-m}(-x\ \sin\ \theta+y\ \cos\ \theta)$$

Where $j=\sqrt{-1}$, W is the radial frequency, σ_x and σ_y are the scaling parameters, m and n are integers, a^{-m} is to ensure that the energy is independent of m, $\theta=n\pi/K$ is the orientation of the filter and K is the number of orientations. Thus the Gabor wavelet transform of the image *I(x, y)* is

$$GW_{m,n}=\iint I(x_1,y_1)\overline{g_{m,n}}(x-x_1,y-y_1)dx_1dy_1 \quad (2)$$

in this paper we use 48 Gabor wavelet features, which are the concatenation of the values $\{\mu_{m,n},\sigma_{m,n},S_{m,n}\}$, where m=1,…,4, n=1,…,4, and

$$\mu_{m,n}=\iint\left|WG_{m,n}(x,y)\right|dx\ dy \quad (3)$$

$$\sigma_{m,n}=\sqrt{\iint\left(\left|WG_{m,n}(x,y)\right|-\mu_{m,n}\right)^2dxdy} \quad (4)$$

$$S_{m,n}=\iint\left(\frac{\left|WG_{m,n}(x,y)\right|-\mu_{m,n}}{\sigma_{m,n}}\right)^3 \quad (5)$$

III. LOCAL BINARY PATTERNS

Local Binary Patterns (LBP) has achieved impressive classification results and it has been extended for various fields. The most attractive advantage of Local Binary Patterns is its invariance to scale and rotation changes. LBP process starts by comparing every pixel with the eight neighbors [10], the values of the LBP cells can be calculated as follows:

E. Al-Daoud is with faculty of Science and Information Technology, Computer Science Department, Zarka University, Jordan (Tel +962-796680005, e-mail:essamdz@zpu.edu.jo).

$$LBP(x,y)=\sum_{n=0}^{7} t(i_n - i_c)2^n \tag{6}$$

where i_n is the gray values of the eight pixels, i_c is the gray value of the center and

$$t(x)=\begin{cases}1 & if \quad x \ge 0\\ 0 & if \quad x<0\end{cases}$$

An illustration of this process is shown in Figure 1.

50	113	40
90	81	70
35	68	75

↓

0	1	0
1		0
0	0	0

00010100≡ 20

Fig. 1 The Original LBP Operator

To achieve rotation invariant, LBP is extended by allowing different radius and samples, for example $LBP_{16,2}^{u2}$ means 16 points on a circle of radius 2, the superscript u2 indicates to the uniform pattern. The general form can be expressed as follows [11]:

$$LBP_{N,R}^{u2}=\begin{cases}\sum_{n=1}^{N} t(i_n - i_c) & if \quad U(LBP_{N,R}) \le 2\\ N+1 & otherwise\end{cases} \tag{7}$$

where

$$U(LBP_{N,R}) = \left|t(i_N - i_c)t(i_1 - i_c)\right| + \sum_{n=2}^{N}\left|t(i_n - i_c) - t(i_{n-1} - i_c)\right|$$

Therefore the number of the extracted features by using $LBP_{N,R}$ and the histogram representation is at most *N+2*. In this paper, the concatenation of LBP_8, LBP_{16} and LBP_{24} are used, thus the total number of the features is 54.

IV. Support Vector Machine

SVM is a classification method tries to find the optimal hyperplanes between different classes. If the training data contains mislabeled points then a slack variable should be added to the SVM optimization problem [12]. Thus SVM problem can be expressed as following:

$$\text{Maximize} \ \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j X_i^T X_j \tag{8}$$

Subject to $0 \le \alpha_i \le C$

$$\sum_{i=1}^{N}\alpha_i y_i = 0$$

Where C measures the degree of misclassification, α's are the support vectors, y is the target and X is the feature vector.

If the data is not linearly separable, then the optimization problem must be extended to nonlinear case. Nonlinear SVM can be written as follows:

$$\text{Maximize} \ \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K(X_i, X_j) \tag{9}$$

Subject to $0 \le \alpha_i \le C$

$$\sum_{i=1}^{N}\alpha_i y_i = 0$$

Where K is a suitable kernel function. The decision function:

$$F(x)=\sum_{i=1}^{N}\alpha_i y_i K(x_i, x) + b \tag{10}$$

Popular predefined kernel functions such as:

- Radial basis function (RBF): $K(x,y)=e^{\left(-\gamma\|x-y\|^2\right)}$
- Polynomial: $K(x,y)=(x.y+1)^d$
- Sigmoid: $K(x,y)=\tanh(\kappa x.y + c)$

V. New Weighted Kernel

To insure that the solution of a SVM optimization problem converges to a unique solution, the kernels must be positive semi-definite. A symmetric kernel K is called positive semi-definite if and only if for any complex numbers c1, c2,…,cn we have:

$$\sum_{i,j=1}^{n} c_i c_j K(x_i, x_j) \ge 0 \tag{11}$$

The selection of a suitable kernel for a given application or for a set of features is still an open problem, for this reason we suggest, in the following proposition, a new kernel with free parameters:

Proposition: Let **A** be a diagonal matrix of size $n \times n$ and its elements are non-negative elements, then the kernel $K(X,Y)=(1+X^t AY)^d$ is positive semi-definite.

Proof: since any diagonal non-negative matrix is positive semi-definite then from the definition, the matrix $X^t AY$ is positive semi-definite, and since a matrix of positive constant is positive semi-definite, the summation and the multiplication of a positive semi-definite matrix is again positive semi-definite, then $(1+X^t AY)^d$ is positive semi-definite.

The matrix **A** can be used to optimize the weight of the features. Grid search or genetic algorithm can be applied to find the suitable values of the diagonal matrix **A**. although this approach is time consuming, the values could be found once and can be used in several similar datasets and applications. If the feature vector is concatenation of features that are extracted by using **M** methods, then the values of the diagonal can be grouped to **M** sets. In this study, the feature vector is concatenation of two sets: 54 features from Gabor wavelet and 48 from LBP.

VI. Experimental Results

To test the suggested method, four databases are used: Yale face database (Yale), Faces95 from Essex university database, Indian Face Database (IFD) and AT&T database of faces (ATT), [13-16]. Figure 2 shows samples from each database. Table 1 summarizes the content of each database.

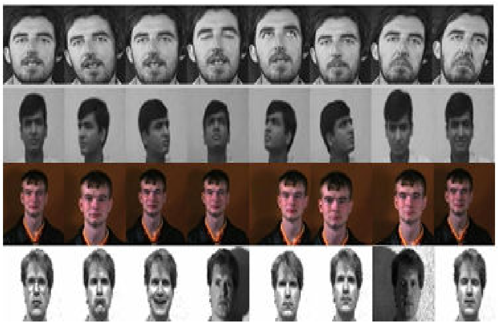

Fig. 2 Faces from ATT, Yale, IFD and Faces95

TABLE I
THE DESCRIPTION OF EACH DATABASE

Database	images	subjects	Variations
ATT	10	40	lighting and expressions
IFD	11	40	Face direction
Faces95	20	72	Distance and scale
Yale	165	15	expressions

Tables II-IV show the classification rate by using 10%, 30% and 70% of the dataset for training and the rest for testing. The proposed method is compared six methods: Principal component analysis (PCA), Linear discriminant analysis (LDA), Batch Linear discriminant analysis (Batch-ilda), Independent component analysis (ICA), Gabor –SVM and LBP – SVM. In all experiments, Matlab 10.0 is used to implement the seven methods. The number of Gabor-SVM features is 48 and the number of LBP-SVM features is 54, therefore, the number of the proposed method features is 102. The adopted diagonal matrix is $\{a_{11}=a, \ldots. a_{54,54}=a,\ a_{55,55}=b, \ldots, a_{102,102}=b\}$, thus the parameters that must be optimized are a and b. In our experiments, the optimal values are a=0.9 and b=0.7, where 10% of the data is selected randomly and grid search is used to find the optimal values.

TABLE II
THE CLASSIFICATION RATE USING 10% FOR TRAINING

Method	ATT	IFD	Faces95	Yale
LDA	97.1	85.6	86.0	91.4
PCA	93.9	80.6	79.4	88.5
BILDA	97.1	84.4	87.4	91.4
ICA	89.1	71.3	78.5	76.1
Gabor SVM	96.3	84.1	82.2	91.4
LBP SVM	95.8	81.7	85.2	88.4
Proposed	96.7	84.8	85.2	93.0

TABLE III
THE CLASSIFICATION RATE USING 30% FOR TRAINING

Method	ATT	IFD	Faces95	Yale
LDA	98.0	89.8	90.0	97.3
PCA	96.0	82.8	80.1	92.0
BILDA	98.0	90.0	90.0	97.3
ICA	90.5	80.2	82.5	86.6
Gabor SVM	97.0	87.1	87.3	92.0
LBP SVM	97.2	85.6	88.7	90.6
Proposed	97.5	89.4	90.2	95.2

TABLE IV
THE CLASSIFICATION RATE USING 70% FOR TRAINING

Method	ATT	IFD	Faces95	Yale
LDA	98.3	95.5	94.0	97.7
PCA	92.5	91.1	82.6	86.6
BILDA	97.5	93.8	90.0	97.7
ICA	92.5	92.2	87.3	73.3
Gabor SVM	95.8	93.3	89.3	88.8
LBP SVM	94.7	89.5	84.5	97.8
Proposed	96.0	93.6	89.1	97.9

It can be notice that the proposed kernel enhances the classification rates for any testing set and for all databases, for example, if 50% of datasets is used for training, then the best classification rates are 97.5, 89.4, 90.2 and 95.2 which can be achieved by using the suggested method.

VII. Conclusion

The performance of the proposed method has demonstrated on four databases that contain various images per individual, different facial emotions, orientations, configuration and expressions. The results are quite promising; the proposed approach outperforms other methods and is suitable to many real-world scenarios. It can be concluded that combining several descriptors and weighted features boost face recognition performance. An important direction for future work includes the use of the diversity of types, such as age , race, and gender.

References

[1] C. Tenllado, J.I. Gomez, J. Setoain, et al., “Improving face recognition by combination of natural and Gabor faces,” Pattern Recognition Letters, 31 (11): 1453- 1460, 2010.

[2] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma, “Robust Face Recognition via Sparse Representation,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31, no. 2, pp. 210-227, Feb. 2009.

[3] Li J,Yu.D.W ,Kuang G, “The Research on Face Recognition Approaches of Infrared Imagery [J]”, Journal of National University of Defense Technology, 28 (2), pp.73-76, 2006.

[4] N. Liu, J. Lai, and H. Qiu, "Robust Face Recognition by Sparse Local Features from a Single Image under Occlusion," 2011 Sixth International Conference on Image and Graphics, pp. 500-505, 2011.

[5] U. Prabhu, and M. Savvides, "Unconstrained Pose Invariant Face Recognition Using 3D Generic Elastic Models," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 33, no. 10, pp. 1952-1961, Oct. 2011.

[6] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, Y. Ma, :Towards a Practical Face Recognition System: Robust Alignment and Illumination by Sparse Representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, no. 99, DOI10.1109/TPAMI.2011.112 ,2011.

[7] Lior Wolf, Tal Hassner, Member, IEEE, and Yaniv Taigman "Effective Unconstrained Face Recognition by Combining Multiple Descriptors and Learned Background Statistics," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 33, no. 10, pp. 1978-1990, Oct. 2011.

[8] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face Recognition by Independent Component Analysis," IEEE Transaction on Neural Networks, vol. 13, no. 6, pp. 1450-1462, 2002.
[9] E. Meyers and L. Wolf, "Using Biologically Inspired Features for Face Processing," Int'l J. Computer Vision, vol. 76, no. 1, pp. 93-104, 2008.
[10] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: application to face recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 12, pp. 2037–2041, 2006.
[11] Prasetiyo, M. Khalid, R. Yusof, and F. Meriaudeau, "A Comparative Study of Feature Extraction Methods for Wood Texture Classification, " 2010 Sixth International Conference on Signal-Image Technology and Internet Based Systems, pp23-29, 2010.
[12] H. Zhang, A. Berg, M. Maire, and J. Malik, "SVM-KNN: Discriminative Nearest Neighbor Classification for Visual Category Recognition," Proc. IEEE CS Conf. Computer Vision and Pattern Recognition, 2006.
[13] http://cvc.yale.edu/projects/yalefaces/yalefaces.html (last Accessed August 2011).
[14] Faces95 from Essex university http://dces.essex.ac.uk/mv/allfaces/ (last Accessed August 2011).
[15] Indian Face Database http://vis-www.cs.umass.edu/~vidit/IndianFaceDatabase/ (last Accessed August 2011).
[16] AT&T http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html (last Accessed August 2011).Yale face database