Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

A Robust Wavelet Based Feature Extraction Method for Face Recognition

Iman Makaremi
Department of Electrical and Computer Engineering
University of Windsor
Windsor, Ontario, N9B 3P4, Canada
i.makaremi@ieee.org

Majid Ahmadi
Department of Electrical and Computer Engineering
University of Windsor
Windsor, Ontario, N9B 3P4, Canada
hmadi@uwindsor.ca

***Abstract*— In this paper, we propose a wavelet based feature extraction method with a high tolerance to white Gaussian noise. This method is also computationally efficient. Along with an HMM classifier, this method is used for face recognition. High recognition rates in the presence of white Gaussian noises with different variances show this technique as a promising feature extraction method.**

I. INTRODUCTION

Face recognition has been one of the hottest topics in pattern recognition in the last decade. Demands for reliable automatic face recognition systems (FRS) have been growing exponentially by introducing machines with human-machine interfaces and rising of security issues. In parallel, different factors such as illumination, pose, and facial expression make face recognition a challenging problem. In addition, noise is an inevitable problem that one should deal with in designing a FRS.

There are many different approaches for reducing noise effect in a FRS. Filtering images before feeding them to the FRS is one of the effective and well-established approaches for this purpose. However, adding an extra preprocessing step to the whole algorithm makes it slow and expensive. On the other hand, there are other methods which deal with the noise issue in the feature extracting step. The feature extraction parts of these methods are designed to be robust to noise, and basically have a lower sensitivity to a specific range of noise. For example, Rashidy Kanan et. al. [1] proposed a feature extraction method which uses genetic algorithm to select optimal features among a set of Pseudo Zernike Moment invariant (PZMI) features. They studied the effect of Gaussian noise with different variances, and illustrated that additive noise does not affect the recognition rate significantly. Jadhav et. al. [2] used Radon and discrete cosine transforms for feature extraction. They exploited Radon transform to enhance low frequency components. This also helped make the method robust to noise. Their results show that the robustness of their method to noise is high.

In [3], we proposed a computationally efficient wavelet based feature extraction method, but in further studies, we found this method very sensitive to noise. Therefore, attempt has been made to make this method robust to noise. We will show that the improved method is robust to noise while it is still computationally efficient. In the next section, this method is described in detail. Section III deals with Hidden Markov Models (HMM) and its application as a classifier. Experimental results are presented in section IV. Finally, conclusions are presented in section V.

II. WAVELET BASED FEATURE EXTRACTION

Wavelet transform [4] has been established as a powerful tool in signal processing. Giving a multiresolution analysis in both time and frequency domain and having different alternatives for the basis function makes wavelet transform a better candidate than Fourier transform [4, 5] for many applications including JPEG2000 [6]. A growing number of publications that deal with hardware implementation of wavelet transform [7-10] are another proof of its applicability.

Performing wavelet transform on a 2D signal, one gets four different matrices of coefficients. One of these matrices is called the approximation coefficient matrix. For a multiresolution analysis, one can apply the wavelet transform on the approximation coefficient matrix. The other three are detail coefficient matrices which are details of the image in three different angles (0, 45, and 90). These matrices are also known as horizontal, diagonal, and vertical details, and are very sensitive to noise. They become deteriorated in presence of noise easily, while the approximation coefficients are more robust to it. The reason is that the detail coefficient matrices contain the high-frequency content of the signal while the approximation coefficient matrix includes the low-frequency contents. Figure 1 shows the decompositions of a face image when it was added with different levels of zero-mean white Gaussian noise (δ=0, 1e-3, 5e-3, 1e-2, 5e-2, and 1e-1). It should be mentioned that all throughout this paper, the intensity of a pixel is represented in the range from zero to one. The information in detail coefficients was almost lost even the variance was as low as 1e-3, while the information in the approximation coefficient was retrievable when δ=5e-2, however it is difficult to recognize the face when the δ=1e-1. Therefore, we only use the approximation coefficients for feature extraction.

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

In order to perform the feature extraction, the 2D wavelet transform of the faces is calculated up to three levels, and the approximation coefficients of different levels are used for further process. In this paper, we studied the effect of number of involving levels in the accuracy rate of the FRS.

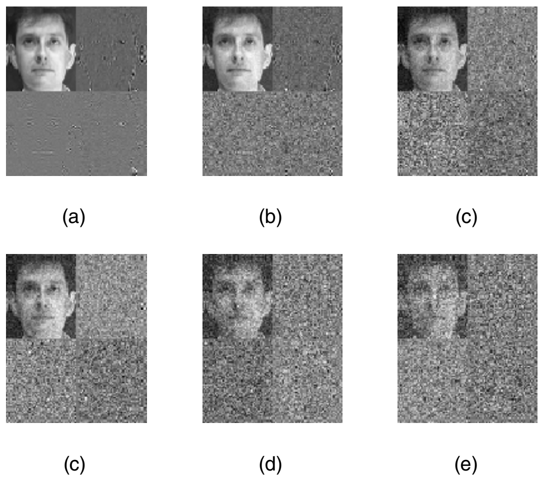

Figure 1. The wavelet decomposition of a face image when variance of the zero-mean Gaussian noise is (a) zero, (b) 1e-3, (c) 5e-3, (d) 1e-2, (e) 5e-2, and (f) 1e-1. The wavelet coefficients have been normalized between 0 and 1, and also resized for illustration.

A window with a width equals to the width of the coefficient matrices, and an arbitrary height of k is selected (See Figure 2). Placing it on top of the matrices, mean and variance of the content of the window is calculated. Then, the window is slid down based on an overlap factor (r) such that each two neighboring windows have r rows in common. The whole process is carried out till the bottom of the image is reached.

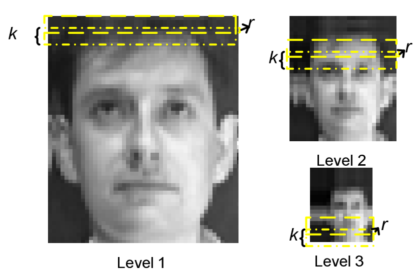

Figure 2. A wavelet decomposition and the top to bottom scanning for feature extraction. The wavelet coefficients have been normalized between 0 and 1, and also resized for illustration.

The outcomes of this part are sequences in $\mathbb{R}^2$, and have different lengths which depend on the height of matrices. If more than one level is used, the lengths of sequences from different levels are not equal. For example, the length of the sequence obtained from the approximation coefficient matrix from the first level is almost twice as long as the length of the sequence obtained from the second level. Therefore, to be able to put these sequences together to have a sequence in $\mathbb{R}^{2L}$ where L is the number of involved levels, a technique should be used to make different sequences of the same length. For this purpose, we used piecewise cubic Hermite interpolation (PCHIP) technique.

In this technique, the interpolants of the neighboring data points $(x_1, y_1), \ldots, (x_n, y_n)$ are constructed with

$$H_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^2(x - x_{i+1})$$

Then, $H_i(x)$'s are stitched together:

$$H(x) = \begin{cases} H_1(x) & \text{if } x_1 \leq x < x_2 \\ \vdots & \\ H_{n-1}(x) & \text{if } x_{n-1} \leq x < x_n \end{cases}.$$

Figure 3 shows the sequences after applying interpolation technique for having an equal length.

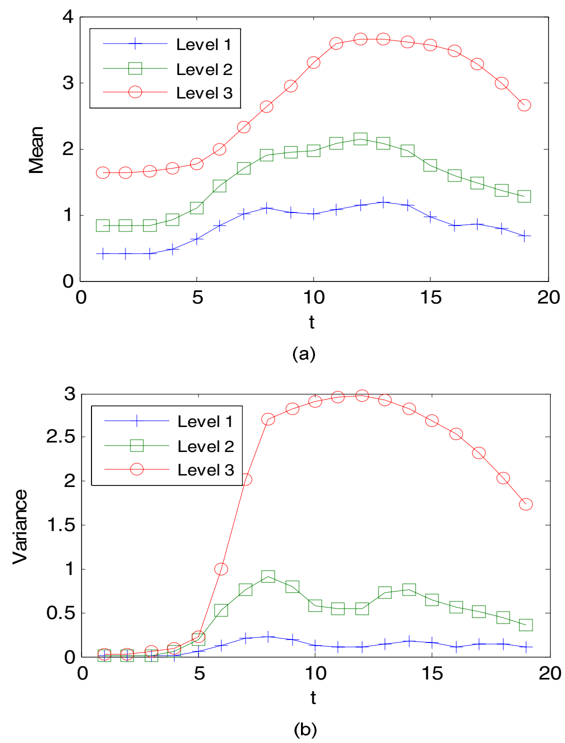

Figure 3. Extracted sequences from a face. (a) Mean, (b) Variance. The sequences extracted from higher level approximation coefficients have been made to be as long as the sequence from the first level with interpolation techniques.

In this paper, we set k and r equal to 5 and 2 respectively. We also adopted Daubechies wavelets because of their compact support and orthonormal nature [11]. We used Daubechies of order 4 (db4), because its better performance than other well-known wavelets in terms of computation time and recognition performance [12]. Cui et. al [13] also had the highest recognition rate with db4. We also tried db10, because we had the best overall recognition rate in [3] with this wavelet.

The proposed method takes 20 ms (one level of composition involved) to 40 ms (three levels of composition involved) to extract features from a 112×92 pixel image in MATLAB environment and with a 64X2 1.90GHz/Win XP.

III. HIDDEN MARKOV MODEL AS CLASSIFIER

Hidden Markov Model [14] is proven to be an effective tool in pattern recognition, and has been used as a promising classifier in face recognition [15-20]. HMMs as double stochastic process can be used to characterize the statistical properties of signals [14]. In fact, a signal is considered as a sequence of observation which can be observed directly. Because the output of our feature extraction method is a sequence, we utilized HMM as the classifier. There are basically two different kind of observation, discrete and continuous. In this paper, we use the continuous HMM since our extracted sequences from faces are continuous. Furthermore, it is not recommended to discretize the output as long as it is possible [14]. A continuous HMM, λ, is defined by the elements as follow:

- Q, the number of hidden states in the model
- T, length of sequences
- $\mathcal{S} = \{S_1, S_2, \dots, S_Q\}$, the finite set of possible hidden states.
- $\Pi = \{\pi_i\}$, the initial state probability distribution, where,

  $$\pi_i = P[q_1 = S_i], 1 \le i \le Q$$

  and $\sum_{i=1}^{Q} \pi_i = 1$.
- $\mathrm{A} = \{a_{ij}\}$, the state transition probability matrix, where

  $$a_{ij} = P[q_{t+1} = S_j | q_t = S_i], \qquad 1 \le i, j \le Q$$

  and $\sum_{j=1}^{Q} a_{ij} = 1, 1 \le j \le Q$.
- $\mathrm{B} = \{b_{j,t}\}$, the emission probability matrix, where

  $$b_{j,t} = P[O_t | q_t = S_j], 1 \le j \le Q, 1 \le t \le T\ .$$

  There are different approaches to define the emission probability for continuous observation. The most general representation of the PDF is a finite mixture of the form

  $$b_{j,t} = \sum_{m=1}^{M} c_{jm} \mathcal{N}(O_t, \mu_{jm}, U_{jm}), 1 \le j \le Q$$

  where c_{jm}, the mixture coefficient for the *m*th mixture in state *j* is always greater than or equal to zero, and summation over *m* should be equal to 1. $\mathcal{N}$ is a Gaussian function, and μ_{jm} and U_{jm} are the mean vector, and the covariance matrix of the *m*th mixture component in state *j* respectively.

To have a functional HMM for real-world applications, three basic problems should be solved. These problems are

- *Evaluation*: Calculating $P(O|\lambda)$
- *Decoding*: Choosing the state sequence that explains the observations.
- *Parameter Estimation*: Adjusting the model parameters.

HMMs can be used as classifiers in two different ways; path discriminant and model discriminant [21]. In path discriminant approach, only one HMM model is used for all classes, and different state sequences of the model distinguish classes. While in the model discriminant approach, a separate model is used for each class, and based on probability of output the class label is obtained.

$$c = \underset{1 \le i < C}{\operatorname{argmax}} [P(O|\lambda_i)]$$

Where, *C* is the total number of classes.

In this paper, we will use the second approach where for each individual class a distinct model will be built. We use Baum-Welch method [14] for training and considering that there are more than one sample in training set for each class, the modified version of this method is utilized [22].

IV. RESULTS AND DISCUSSION

AT&T (already known as ORL) is the database we have used in this paper. This database includes 400 different pictures of 40 individuals; 10 for each. Five out of ten photos were randomly selected for the train set, and the rest were put in the test set. For sensitivity evaluation, we only used noise-free samples in the train set, and generated 6 different test sets with different levels of white Gaussian noise. The variances of noise were 0 (noise-free), 0.001, 0.005, 0.01, 0.05, and 0.01.

The numbers of states and mixture components of the HMMs were fixed to 5 and 2 respectively, because the overall best results in [3] had been obtained with this size of model.

The algorithm was executed 20 times for each set, and the averages of the recognition rates are shown in Table I. We also tested the previous method's sensitivity to noise. The feature column shows the features with two-part names. The first part is the wavelet used for decomposition (db4 and db10), and the second part shows how many levels were used, e.g. L2 means that the approximation coefficients of two levels have been used. 'Full' in the second part refers to our previous proposed method [3] in which all of the coefficient matrices of the first level of decomposition would be used for feature extraction. As it is shown, the recognition rate is high in the noise-free case. However, the previous method is very sensitive to noise, and even in the presence of a weak noise, the recognition rate drops significantly. The reason is that the detail matrices were also used for feature extraction. Although they contain some information about the edges, and can be very informative, they are very sensitive to noise. On the other hand, features that only used approximation matrices are much more robust to noise. The recognition rates are more than %90 for white Gaussian noises with variances less than or equal to 0.01 (Figure 4).

The number of levels involved in feature extraction has a clear effect on the recognition rate and its robustness to noise. Features with more than one level of coefficients were more robust to more intense noises, while the recognition rates with features with one level of coefficients dropped significantly when δ was greater than 0.01. The recognition rates dropped to less than %50 for all of the features when noise variance was 0.1. In fact, faces at this level of noise are hardly recognizable.

Daubechies of order 4 performed better that Daubechies of order 10 when noise variance is not so high, while Db10 outperformed db4 in all levels for noises with variances greater than 0.05.

TABLE I. AVERAGE CLASSFICATION RATE WITH DIFFERENT FEATURES AND NOISES

Feature \ δ	*0*	*0.001*	*0.005*	*0.010*	*0.050*	*0.100*
DB4-Full	97.1	8.8	2.5	2.4	2.5	2
DB10-Full	96.9	6.5	2.8	2.7	2.4	1.8
DB4-L1	93.5	93.25	92.88	92.2	48.98	17.38
DB4-L2	96.5	95.43	95.45	95.25	81.9	35.65
DB4-L3	96.88	96.5	96.43	96.6	88.95	45.85
DB10-L1	93.88	93	93.4	91.98	57.8	21.78
DB10-L2	95.13	95.08	94.9	94.75	87.3	49.43
DB10-L3	95.63	95.08	95.13	94.85	89.73	56.7

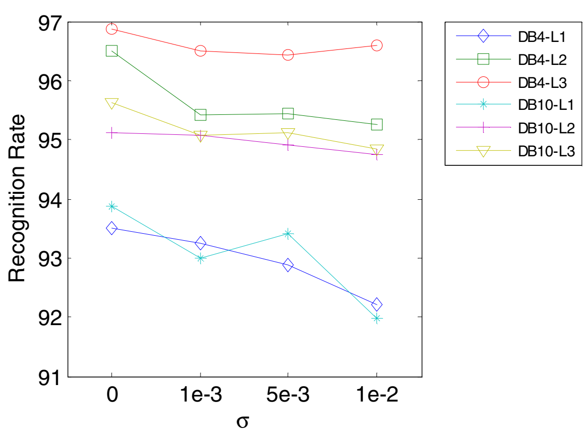

Figure 4. Average cassification rate with different features and noises.

V. CONCLUSION

A computationally efficient and robust to noise feature extraction method which is based on wavelet transform has been proposed in this paper. This method has been used along with HMM as the classifier for face recognition. The FRS has been examined with non-noisy and noisy faces from AT&T database, and very high classification rates have been obtained. Although using noisy samples in train set is common in order to adapt the classifier to noise, we only used noise-free samples in the train set. The results show that involving more levels of wavelet decomposition helps to have a more robust method. Even though it raises the computational effort, the feature extraction method is still fast enough with three levels of decomposition, and requires less than 40ms to extract features for one face image.

VI. REFERENCES

[1] H. Rashidy Kanan, K. Faez, "GA-based optimal selection of PZMI features for face recognition", Applied Mathematics and Computation, Vol. 205, pp. 706-715,.2008.

[2] D. V. Jadhav, R. S. Holambe, "Radon and discrete cosine transforms based features extraction and dimensionality reduction approach for face recognition", Signal Processing, Vol. 88, pp. 2604-2609, 2008.

[3] I. Makaremi, M. Ahamdi, "An efficient wavelet based feature extraction method for face recognition", International Symposium of Neural Networks, Wuhan, China, 2009.

[4] I. Daubechies, "Wavelet transforms and orthonormal wavelet bases", Different perspectives on wavelets, San Antonio, pp.1-33, 1993.

[5] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation." IEEE Transaction on Pattern Analysis and Machine. Intelligence.,Vol. 11, pp. 674–693, 1989.

[6] D. Taubman, and M. W. Marcellin, "JPEG2000 image compression: fundamentals, standards and practice", Boston, MA: Kluwer, 2002.

[7] Grzezczak, M. K. Mandal, and S. Panchanathan, "VLSI implementation of discrete wavelet transform", IEEE Transaction on VLSI Systems,. Vol. 4, pp. 421-433, 1996.

[8] G. Lafruit, F. Catthoor, J. P. H. Cornelis, and H. J. De Man, "An efficient VLSI architecture for 2-D wavelet image coding with novel image scan" IEEE Transaction on VLSI Systems, Vol. 7, pp. 56-68, 1999.

[9] P. McCanny, S. Masud, and J. McCanny, "Design and implementation of the symmetrically extended 2D wavelet transform", Proceeding of ICASSP'02, Vol.3, pp. 3108-3111, 2002.

[10] A. Shahbahrami, B. Juurlink, and S. Vassiliadis, "Implementing the 2-D wavelet transform on SIMD-enhanced general-purpose processors", IEEE Transaction on Multimedia., Vol. 10, pp. 43-51, 2008.

[11] P. J. Ohilips, H. Moon, S. A. Rizvi, P. J. Rauss, "The FERET ecaluation methodology for face recognition algorithms", IEE transactions on pattern analysis and machine intelligence, Vol. 22, pp.1090-1104, 2000.

[12] G. C. Feng, P.C. Yuen, D. Q. Dai, "Human face recognition using PCA on wavelet subband", SPIE journal of electronic imaging, Vol. 9, pp. 226-233, 2000.

[13] L. Cui, Y. Y. Tang, F. Liao, X. Du, "Face recognition based on discriminant waveletfaces", International Conference on Systems, Man, and Cybernetics, Montreal, Canada, 2007.

[14] L. Rabiner, "A tutorial on Hidden Markov Models and selected applications in speech recognition", Proc. of IEEE,Vol. 77, No. 2, pp. 257–286, February 1989

[15] V. V. Kohir and U. B. Desai, "Face recognition using DCTHMM approach", Workshop on Advances in Facial Image Analysis and Recognition Technology (AFIART), Freiburg, Germany, June 1998.

[16] H. Othman, and T. Aboulnasr, "Hybrid Hidden Markov Model for Face Recognition", IEEE Southwest Symposium on Image Analysis and Interpretation, Texas, USA, pp. 36-40, April 2000.

[17] T. Otsuka, and J. Ohya, "Recognition of facial expressions using HMM with continuous outputprobabilities", 5th IEEE Workshop on Robot and Human Communication, Tsukuba, Japan, pp.323-328, November, 1996.

[18] M. Bicego, U. Castellani, and V. Murino, "Using hidden Markov models and wavelets for face recognition", 12th International Conference on Image Analysis and Processing, pp. 52-56, September 2003.

[19] L. Hung-Son, and L. Haibo, "Face identification system using single hidden Markov model and single sample image per person", Proceeding of IEEE International Joint Conference on Neural Networks, Vol. 1, pp. 455-459, July 2004.

[20] P. Nicholl, D. Bouchaffra, A. Amira, and R.H. Perrott,"Multiresolution Hybrid Approaches for Automated Face Recognition", Second NASA/ESA Conference on AHS, pp. 89-96, August 2007.

[21] M.Y. Chen, A. Kundu, and S.N. Srihari, "Variable duration hidden Markov and morphological segmentation for handwritten word recognition", IEEE Transaction on Image Processing, Vol. 4, No. 12, pp. 1675-1688, 1995.

[22] L.R. Rabiner, and B.-H. Jung, "Fundamentals of Speech Recognition" Prentice Hall, Ch. 6, pp. 369-370, 1993.