Using Hidden Markov Models and Wavelets for face recognition

M. Bicego, U. Castellani, V. Murino

Dipartimento di Informatica, Università di Verona

Ca' Vignal 2, Strada Le Grazie 15

37134 Verona, Italia

{bicego,castellani,murino}@sci.univr.it

Abstract

In this paper, a new system for face recognition is proposed, based on Hidden Markov Models (HMMs) and wavelet coding. A sequence of overlapping sub-images is extracted from each face image, computing the wavelet coefficients for each of them. The whole sequence is then modelled by using Hidden Markov Models. The proposed method is compared with a DCT coefficients-based approach [9], showing comparable results. By using an accurate model selection procedure, we show that results proposed in [9] can be improved even more. The obtained results outperform all results presented in the literature on the Olivetti Research Laboratory (ORL) face database, reaching a 100% recognition rate. These performances proves the suitability of HMMs to deal with the new JPEG2000 image compression standard.

1. Introduction

Face recognition is undoubtedly an interesting research area, growing in importance in recent years, due to its applicability as a biometric system in commercial and security applications. These systems could be used to prevent unauthorized access or fraudulent use of ATMs, cellular phones, smart cards, desktop PCs, workstations, and computer networks. The appealing characteristic of a face recognition system is that, differently from fingerprint or iris biometric systems, it represents a not invasive control tool.

A large literature is available on this topic (for a review see [4]). The first approaches, proposed in the 70's, were based on geometric features [8]. Subsequently, one of the well known face recognition algorithms is the so-called *Eigenface* method [22, 25], that uses the *Principal Component Analysis* to project faces into a low-dimensional space, where each face can be expressed as a linear combination of the eigenfaces. This method showed to be not very robust to variations of the face orientation, and one solution to this problem was provided by the view-based eigenspace method introduced in [15]. Another important approach is the *Elastic Matching* method [25, 10], which proved to be invariant to expression changes. The idea is to build a lattice on image faces (rigid matching stage), and to apply, at each point of the lattice, a bank of Gabor filters. In case of variations of expression, this lattice can warp to adapt itself to the face (elastic matching stage).

To the best of our knowledge, the best results obtained on standard database (as ORL [20]) were proposed in [5] and [9], using Hidden Markov Model-based approaches, obtaining an almost perfect classification accuracy. More in detail, in [5] a pseudo 2D HMM was used for classifying faces: face images are described using the DCT (Discrete Cosine Transform) coefficients, computed on a set of partially overlapped sub-images. One of the most interesting feature of this method is its direct applicability to JPEG (Joint Photographic Experts Group) images, without any need of decompressing them. The other technique, proposed in [9], uses standard one-dimensional HMMs trained on sequences of DCT coefficients extracted from an image. This method, even if really effective, does not exploit all the potentialities of the HMM approach. In particular, the important question of model selection is completely disregarded. This issue consists of the choice of the model size, *i.e.*, the number of states. In this paper we will show that an accurate model selection, obtained with the method proposed in [2], allows the further improvement of the classification accuracy of the HMM approach.

With the advent of the new JPEG standard, the so called JPEG2000 [7], which uses wavelet coding [23], it is reasonable asking if this new standard can be treated efficiently as the older one in terms of recognition accuracy. In this paper, a comparison between DCT coding and wavelet coding is presented, aimed at evaluating the effectiveness of the HMMs in modelling faces using these two types of compression methods. Experimental evaluation on standard ORL database shows that HMMs are really effective in recognizing faces using both DCT and wavelet coefficients.

IEEE
COMPUTER
SOCIETY

We obtain a 100% classification accuracy on ORL database, and this result represents the best accuracy ever obtained on that database, largely better than Neural Networks, Eigen-Face and Elastic Matching approaches.

The rest of the paper is organized as follows. In Section 2, Hidden Markov Models are briefly resumed, and, in Section 3, the DCT-HMM approach proposed in [9] is described. Section 4 reports a brief introduction of the wavelet approach for image compression, and, in Section 5, a comparison between DCT and wavelet methods is discussed. Finally, in Section 6, conclusions are drawn.

2. Hidden Markov Model

A discrete-time Hidden Markov Model λ can be viewed as a Markov model whose states cannot be explicitly observed: each state has associated a probability distribution function, modelling the probability of emitting symbols from that state. More formally, a HMM is defined by the following entities [16]:

- $S = \{S_1, S_2, \cdots, S_N\}$ the finite set of the possible hidden states;

- the transition matrix $\mathbf{A} = \{a_{ij}, \ 1 \leq j \leq N\}$ representing the probability to go from state S_i to state S_j,

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

with $a_{ij} \geq 0$ e $\sum_{j=1}^{N} a_{ij} = 1$;

- the emission matrix $\mathbf{B} = \{b(o|S_j)\}$, indicating the probability of the emission of the symbol o when system state is S_j; in this paper continuous HMM were employed: $b(o|S_j)$ is represented by a Gaussian distribution, *i.e.*

$$b(o|S_j) = \mathcal{N}(o|\mu_j, \Sigma_j). \qquad (1)$$

where $\mathcal{N}(o|\mu, \Sigma)$ denotes a Gaussian density of mean μ and covariance Σ, evaluated at o;

- $\boldsymbol{\pi} = \{\pi_i\}$, the initial state probability distribution, representing probabilities of initial states, i.e.

$$\pi_i = P[q_1 = S_i] \ 1 \leq i \leq N$$

with $\pi_i \geq 0$ and $\sum_{i=1}^{N} \pi_i = 1$.

For convenience, we denote an HMM as a triplet $\boldsymbol{\lambda} = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi})$.

The training of the model, given a set of sequences $\{\mathbf{O}_i\}$, is usually performed using the standard Baum-Welch re-estimation [16], able to determine the parameters $(\mathbf{A}, \mathbf{B}, \boldsymbol{\pi})$ that maximize the probability $P(\{\mathbf{O}_i\}|\boldsymbol{\lambda})$. In this paper the training procedure is stopped after the convergence of the

likelihood. The evaluation step, *i.e.* the computation of the probability $P(\mathbf{O}|\boldsymbol{\lambda})$, given a model $\boldsymbol{\lambda}$ and a sequence $\mathbf{O}$ to be evaluated, is performed using the *forward-backward procedure* [16].

3. The DCT-HMM approach

In this section, the method proposed in [9] is detailed. In this approach, HMMs were used in a standard manner: one model is trained for each class, using standard Baum Welch algorithm, and the subsequent classification was performed using standard Maximum Likelihood classification rule, *i.e.* , assigning an unknown item to the class whose model shows the highest likelihood. In this paper, differently than in [9], where the model selection issue was disregarded, the model size is carefully estimated, using the technique proposed in [2]. This technique is able to deal with drawbacks of standard general purpose methods, like those based on the *Bayesian inference criterion* (BIC) [21], *i.e.*, computational requirements, and sensitivity to initialization of the training procedure. The basic idea is to perform a "decreasing" learning, starting each training session from a "nearly good" situation, derived from the result of the previous training session by pruning the "least probable" state of the model. As shown in the experimental session, this permits the improvement of the method performances presented in [9].

The strategy used to obtain the data sequence from a face image consists of two steps. In the the first step, a sequence of sub-images of fixed dimension is obtained by sliding over the face image a square fixed size window, in a raster scan fashion, with a predefined overlap. The procedure for scanning the image is visualized in Fig. 1. The second step con-
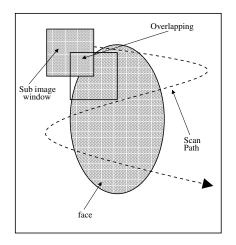

Figure 1. Sampling scheme to generate the sequence of sub-images.

sists of applying the 2D DCT to each gathered sub image. The obtained coefficients are scanned in a zig-zag fashion, similarly to the method used for the JPEG coding. Only few of these coefficients are retained, determining the dimensionality of the observation. By applying this step to all the sub-images of the sequence, we finally get the actual sequence observation. Its dimensionality will be $D \times T$, where D is the number of the DCT coefficients retained, and T is the number of sub-images gathered in the sample scanning operation.

4. The Wavelet coding

The wavelet transform is a methodology emerged in the last years, useful in many applications, especially in the field of image compression. Wavelet-based coding provides substantial improvements in picture quality at higher compression ratios, with respect to standard DCT transform. Over the past few years, a variety of wavelet-based schemes for image compression have been developed and implemented [17]. Because of the many advantages, the compression technologies used in the upcoming JPEG-2000 standard [7] are all based on the wavelet technology.

Wavelets could be defined as a mathematical tool for hierarchically decomposing functions. The wavelet transform is aimed at describing a function in terms of a coarse overall shape, plus details that range from broad to narrow. More formally, wavelets are functions defined over a finite interval and having an average zero value. The basic idea is to represent any arbitrary function $f(t)$ as a superposition of a set of basis functions. These basis functions, or *baby wavelets*, are obtained from a single prototype wavelet called the *mother wavelet*, by dilations or contractions (scaling) and translations (shifts).

In this paper, we proposed to modify the sequence extraction approach presented in the previous section, by substituting the DCT coding with the wavelet coding. In this case, we used the Haar wavelets [23], representing the simplest wavelet basis. We employed the non standard decomposition, which alternates between row and column processing, allowing a more efficient coefficients' computation. The proposed algorithm calculates the coefficients representing the image with a normalized two-dimensional Haar basis, sorting these coefficients in order of decreasing magnitude. Subsequently, the first M coefficients are retained, performing a lossy image compression. For a more complete treatment of wavelet image compression, see the De Vore paper [23]. As in the DCT case, the number of retained coefficients determines the dimensionality of the observation vector, while its length is determined by the number of sub images gathered.

5. Experimental results

In this section, wavelet and DCT approach are compared, in order to assess the HMM suitability in modelling wavelet coefficients. The experiments have been conducted on the ORL database [3], which consists of 40 subjects with 10 faces each. In Fig. 2, 10 examples of subjects from the ORL database are presented with different images per subject: one can notice that this database is characterized by illumination, pose and expression changes between images of the same subject.

Figure 2. 10 example faces from ORL face database.

One HMM is trained for each subject, using 5 images, while the remaining 5 are used for testing. Training was performed using standard Baum Welch technique, stopping the procedure after likelihood convergence. The number of states was carefully estimated, using the technique proposed in [2]. The adopted classification scheme was the usual one, *i.e.*, the maximum likelihood (ML) scheme. Experiments were repeated 20 times, in order to increase the statistical significance of the results. Sub-image size was fixed during all experiments to 16x16, and the number of the retained coefficients was varied (4, 8 and 12), so as the overlapping ratio (50% and 75%). Results are proposed in Table 1(a) and (b). From these tables is evident that Wavelet and DCT approaches perform equally well on this database. Regard-

Num. coeff.	DCT accuracy	Wavelet accuracy
4	98.6%	97.4%
8	99.4%	100%
12	100%	100%

(a)

Num. coeff.	DCT accuracy	Wavelet accuracy
4	97.9%	95.4%
8	99.2%	99.5%
12	99.6%	98.8%

(b)

Table 1. Comparison of accuracies obtained in the ORL database by DCT and Wavelet approaches, for different number of retained coefficients and for different overlap ratios: (a) overlap ratio = 50% and (b) overlap ratio = 75%.

Overlap ratio	Classification accuracy
50%	84.9%
75%	77.8%

Table 2. Classification accuracy using HMM and the "naive" signal.

Method	Error	Ref.	Year
Top-down HMM + gray tone features	13%	[19]	1994
Eigenface	9.5%	[22]	1994
Pseudo 2D HMM + gray tone features	5.5%	[18]	1994
Elastic matching	20.0%	[25]	1997
PDNN	4.0%	[12]	1997
Continuous n-tuple classifier	2.7%	[13]	1997
Top-down HMM + DCT coef.	16%	[14]	1998
Point-matching and correlation	16%	[11]	1998
Ergodic HMM + DCT coef.	**0.5%**	[9]	1998
Pseudo 2D HMM + DCT coef.	**0%**	[5]	1999
SVM + PCA coef.	3%	[6]	2001
Indipendent Component Analysis	15%	[24]	2002
Gabor filters + rank correlation	8.5%	[1]	2002
Wavelet + HMM	**0%**		2003

Table 3. Comparative results on ORL database. "Wavelet + HMM" represents the proposed method. The three best results are displayed in bold font.

6. Conclusions

In this paper, a new approach for face recognition has been proposed, based on HMMs and wavelet coding. A model selection procedure has been applied to the HMMs in order to automatically find the best model for the data. The effectiveness of HMMs in tackling the face recognition problem has been proved by the interesting results obtained using both "naive" and accurate features.

This method, compared to the DCT approach, reported similar results, confirming the HMM suitability to deal with the new JPEG2000 image compression standard. The obtained results outperform all results presented in the literature on the ORL database, reaching a perfect classification accuracy.

ing the performances obtained with the DCT coefficients, it is worthwhile to note that the use of the model selection permits to reach a classification accuracy of 100%, not obtained in [9] (99.5%).

HMMs are very suited in modelling faces. To confirm this fact we substitute, in the proposed system, the wavelet coding with a trivial coding, represented by the mean of the square window. Then we learn the classifier on this "naive" signal, obtaining the results presented in Table 2. The results obtained with this "naive" approach show that HMM is very effective in modelling faces, proposing results comparable to those obtained in the literature (see Table 3). Moreover, the wavelet coding and the DCT coding gives a fundamental surplus to enhance the HMM potentiality. The obtained results outperform, on the ORL database, all other methods proposed in the literature. This could be observed in the Table 3, which presents a comparison between published results obtained by the most important face recognition algorithms on the ORL database. The three best performances are displayed in bold fonts in the table, all based on Hidden Markov Models.

References

[1] O. Ayinde and Y. Yang. Face recognition approach based on rank correlation of gabor-filtered images. *Pattern Recognition*, 35(6):1275–1289, 2002.

[2] M. Bicego, V. Murino, and M. Figueiredo. A sequential pruning strategy for the selection of the number of states in Hidden Markov Models. *Pattern Recognition Letters*, 24(9–10):1395–1407, 2003.

[3] A. L. Cambridge. The olivetti research ltd database of faces. Downloadable from http://www.uk.research.att.com/facedatabase.html.

[4] R. Chellappa, C. Wilson, and S. Sirohey. Human and machine recognition of faces: a survey. *Proceedings of IEEE*, 83(5):705–740, 1995.

[5] S. Eickeler, S. Mller, and G. Rigoll. Recognition of jpeg compressed face images based on statistical methods. *Image and Vision Computing*, 18:279–287, March 2000.

[6] G. Guo, S. Z. Li, , and C. Kapluk. Face recognition by support vector machines. *Image and Vision Computing*, 19(9-10):631–638, 2001.

[7] ISO/IEC/JTC1/SC29/WG1 and N390R. Jpeg 2000 image coding system, March 1997. http://www.jpeg.org/public/wg1n505.pdf.

[8] T. Kanade. Picture processing system by computer complex and recognition of human faces, November 1973. doctoral dissertation, Kyoto University.

[9] V. V. Kohir and U. B. Desai. Face recognition using DCT-HMM approach. In *Workshop on Advances in Facial Image Analysis and Recognition Technology (AFIART)*, Freiburg, Germany, June 1998.

[10] C. Koutropoulos, A. Tefas, and I. Pitas. Morphological elastic graph matching applied to frontal face authentication under well-controlled and real conditions. In *International Conference on Multimedia Computing and Systems*, volume 2, pages 934–938, 1999.

[11] K. M. Lam and H. Yan. An analytic-to-holistic approach for face recognition on a single frontal view. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(7):673–686, 1998.

[12] S. Lin, S. Kung, and L. Lin. Face recognition/detection by probabilistic decision-based neural network. *IEEE Trans. Neural Networks*, 8(1):114–131, January 1997.

[13] S. M. Lucas. Face recognition with the continuous n-tuple classifier. In *Proceedings of British Machine Vision Conference*, September 1997.

[14] A. V. Nefian and M. H. Hayes. Hidden Markov models for face recognition. In *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2721–2724, Seattle, May 1998.

[15] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. *Computer Vision and Pattern Recognition*, pages 84–91, 1994.

[16] L. Rabiner. A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proc. of IEEE*, 77(2):257–286, 1989.

[17] S. Saha. Image compression - from DCT to wavelets: A review. *ACM Crossroads Magazine*, Spring 2000.

[18] F. Samaria. *Face recognition using Hidden Markov Models*. PhD thesis, Engineering Department, Cambridge University, October 1994.

[19] F. Samaria and A. Harter. Parameterisation of a stochastic model for human face identification. In *Proceedings of IEEE Workshop on Applications of Computer Vision*, Sarasota, Florida, December 1994.

[20] F. Samaria and A. Harter. Parametrisation of a stochastic model dor human face identification. In *Proc. IEEE Workshop on Applications of Computer Vision*, pages 138–142, 1994.

[21] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[22] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[23] R. D. Vore, B. Jawerth, and B. Lucier. Image compression through wavelet transform coding. *IEEE Trans. on Information Theory*, 38(2), 1992.

[24] P. C. Yuen and J. H. Lai. Face representation using indipendent component analysis. *Pattern Recognition*, 35(6):1247–1257, 2002.

[25] J. Zhang, Y. Yan, and M. Lades. Face recognition: eigenface, elastic matching and neural nets. *Proceedings of the IEEE*, 85(9), September 1997.

IEEE
COMPUTER
SOCIETY