International Journal of Scientific & Engineering Research, Volume 6, Issue 1, January-2015 30
ISSN 2229-5518

A FACE RECOGNITION SYSTEM BY HIDDEN MARKOV MODEL AND DISCRIMINING SET APPROACH

Mrs. SWATI Y.RAUT , Mrs.DIPTI.A.DOSHI

ABSTRACT—Different approaches have been proposed over the last few years for improving holistic methods for facerecognition. Some of them include color processing, different face representations and image processingtechniques to increase robustness against illumination changes. There has been also some research about thecombination of different recognition methods, both at the feature and score levels. Embedded hidden Markov model(E-HHM) has been widely used in pattern recognition.The performance of Face recognition by E-HMM heavily depends on the choice of model parameters.In this paper, we proposea discriminating set ofmulti E-HMMs based face recognition algorithm. Experimental results illustrate that compared with the conventional HMMbased face recognition algorithm the proposed method obtainbetter recognition accuracies and higher generalization ability.

Keywords: Embedded hidden Markov model (E-HMM),Face recognition, Pattern recognition, Discriminating set, Generalization ability.

1 INTRODUCTION

In recent years, a large number of methods have been investigated for automatic face recognition [1]. Face recognition has attracted attention from the research and industrial communities with a view to achieve a "hands-free" computer controlled systems for access control, security and surveillance. As a baby one of our earliest stimuli is that of human faces. We rapidly learn to identify,characterize and eventually distinguish those who are near and dear to us. Thisskill stays with us throughout our lives.As humans, face recognition is an ability we accept as commonplace. It is only whenwe attempt to duplicate this skill in a computing system that we begin to realize thecomplexity of the underlying problem. Understandably, there are a multitude of differingapproaches to solving this complex problem. And while much progress hasbeen made many challenges remain.

HMM attracts more and more attention because of its effectiveness for face recognitionand facial expression recognition. The basic theory of HMM is founded by Baum at the end of 1960s.After1980s, the HMM is well known and applied in speech and printing recognitionsuccessfully [2] [3].

The earliest HMM of face is built by F. Samaria in 1994 [4].It is a onedimensionalmodel using pixel-intensity as the observation vectors. Later, a kindof HMM with two-dimensional discrete cosine transform (2D-DCT) coefficients asobservation vectors are proposed by Nefianwhich can be used more effectivelyin scale invariant systems and offer a more flexible framework for face recognition [5].However, the image is a two-dimensional (2D) array so HMM may lose a lot of spatialinformation. 2D-HMM is proposed for face recognition as an enhanced versionof HMM. But its application is limited due to high computational complexity andmemory requirement. Then the embedded HMM (E-HMM) is introduced for facerecognition by Nefian [6]. The E-HMM not only can extract main information of the2D images, but is also robust for pose and environment variation with a receivable-complexity.However, the E-HMM with different parameters will generate different informationexpression and recognition performance, so reasonable selection of modelparameters turns into an urgent problem. Aiming at this, a new face recognitionalgorithm is proposed based on E-HMM and Discriminating set approach. By selectingmany perfectandsundry models, the sensitivity of recognition performance toE-HMM is reduced and the generalization ability of the face recognition algorithmis enhanced.

2 LITERATURE SURVEY

2.1 Hidden Markov Model

Hidden Markov model is a variant of a finite state machine [7]. However, unlike finite state machines, they are not deterministic. A normal finite state machine emits a deterministic symbol in a given state. Further, it then deterministically transitions to another state. Hidden Markov models do neither deterministically, rather they both transition and emit under a probabilistic model.

2.2 Elements of Hidden Markov Models

In order to characterize an HMM completely, following elements are needed.
N=Thenumber of states of the model.
M=The number of distinct observation symbols per state.
T=length of observation sequence, length of observation sequence i.e. the number of symbols observed 1,2,...,N will denote the N urns respectively denotes the state in which we are at time t .
$V = \{v_1, \ldots v_m\}$ the discrete set of possible observation symbols.

ISSN 2229-5518

Π= { Π_i }= P(i_1= i), the probability of being in state i at the beginning of the experiment i.e. at t=1.

A = {a_{ij}} where a_{ij} = P(i_{t+1}=j | i_t=i), the probability of being in state j at time t+1 given that we were in state i at time t.

B = { $b_j(k)$}, $b_j(k)$ = P(v_k at t | i_t = j), the probability of observing the symbol $v_k$given that we are in state j.

$O_t$will denote the observation symbol observed at instant t

λ =(A,B, Π) will be used as a compact notation to denote an HMM.

2.3 Embedded HMM

The Embedded HMM consists of a set of super states, along with a set of embedded states [8]. The super states may then be used to model two-dimensional data along one direction, with the embedded HMM modeling the data along the other direction. The elements of an embedded HMM are:

N=The number of super states, S= the set of super states, S={Si} 1≤ i ≤ N.

Π = {π, i} The initial super state distribution, where π are the probabilities of being in super state i at time zero.

A={a,ij}= super state transition probability matrix. Where a,ij is the probability of transitioning from super state i to super state i.

Λ= super states, and the ith super state, consisting of several embedded states, is defined as Λ = {Πi , Ai , Bi , 1 ≤ i ≤ Ns }, where Πi = {πik , 1 ≤ k ≤ N i } is the initial distribution of embedded states, N i is the number of states embeddedin the ith super state; Ai is the transition probability matrix of the embedded.

B_i =state probability matrix of the ith super state and $B_i = \{b_k^i(O_{x,y})\}$where b_k is distribution of observation vectors in the kth embedded state and theith super state.Bi is a finite mixture of the form:

$$b_k^i(o_{xy}) = \sum_{j=1}^{k} c_{kj}^i N(O_{xy}, \mu_{kj}^i, \Sigma_{kj}^i), \quad 1 \le K \le N^i, \quad (1)$$

where K is the number of mixture components, each of which is described by the Gaussian probability density functions (pdf) $N(O_{xv}, \mu_{ki}^i, \Sigma_{ki}^i)$ with mean vector μ_{ki}^i.

1. The Proposed Face Recognition Algorithm Based on E-HMM and discriminatingSet

A Good machine learning or pattern classification system should have a strong generalizationability, which means the ability to recognize or process the unknown thingsusing the obtained knowledge and techniques. Therefore, generalization ability isalways the ultimate problem concerned about in machine learning. Assembly learningis a new machine learning technique developed in last decades, where results ofmany algorithms are jointly used to solve a problem. The assembly learning cannotonly improve the classification or recognition accuracy, but also obviously improvethe generalization ability of the learning system. So it is regarded as one of thefourth fundamental research topics in recent years.

3.1 DiscriminatingSET

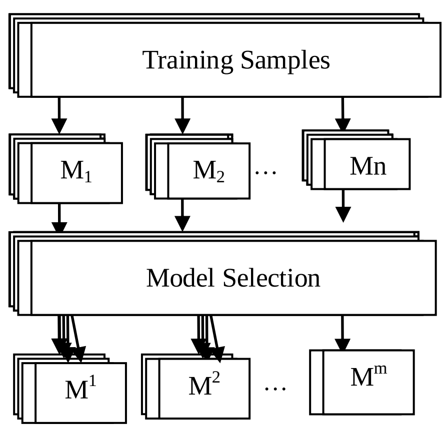
The discriminating set of the E-HMMsto improve the generalization ability is discussed here. Usually, different models constructed withdifferent parameters describe the different characteristics of the face. For example,the smaller sampling windows emphasize the local information, while the largerones pay attention to the global information. Based on the idea of "Many couldbe better than all", for the model ensemble, it is the more the better, but needto select diverse models with higher accuracy to set. Therefore, a novel facerecognition algorithm is proposed based on E-HMM and discriminating set, andthe scheme is shown Fig.1.

The algorithm is divided into two parts: the first part is training module, whichtrains n models and selects m divers models with higher accuracy for discriminating set; the second section is the assembly module, which is used to groupthe m selected models for face recognition.

In Fig.1,M_1,M_2, . . . , $M_n$are models of faces with different parameters, denoted as $M = (O; \lambda)$, where

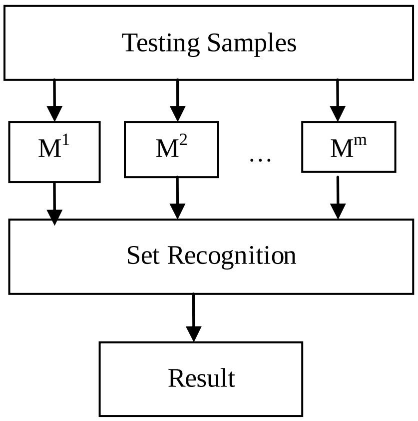
$O = (o1, o2, \ldots, o_T \mid P \times L, \Delta x \times \Delta y, N_{2D}–DCT)$ (2)

is the feature information of the model; $\lambda = (\Pi, A, \Lambda \mid Ns, N^i, K)$ is the structureinformation of the model. Therefore, model$M$ is determined by the above six groupsof parameters



(a)

ISSN 2229-5518



(b)

Fig.1 Proposed Face recognition algorithm based on EHM-Mand discriminatingset approach
a) Training algorithm b) Recognition algorithm

3.2 The discriminating set *algorithm of E-HMMs*

A discriminating set of E-HMMs is proposed in this section. Given a database with$L$ face images, each person has l images, which consists of l_1 gallery images and $l_2$probe images. The detailed algorithm is given as follows:

a) Suppose n models $\{M_1, M_2, \dots, M_n\}$are trained by the gallery images, they areused to recognize all the probe images and the corresponding recognition ratesare $r_i(i = 1, 2, \dots, n)$.
b) Sort the n models $M_i(i = 1, 2, \dots, n)$ according to descending order of $r_i$as$\{M_1', M_1', \dots M_n'\}$. Let Sbe the set of selected models and initialized as $S = \{M_1'\}$ and $Z = \{M_1', M_1', \dots M_n'\}$-$s$,$m$-$1$ models are selected by performingthe following steps:

1. $\varepsilon_t$is the number of faces wrongly recognized by the models in S.
2. For each model $M_i' \in Z$ *calculate* the number of the probe faces that arerecognized correctly by $M_i'$and wrongly by at least one of the models in$S$. The number is denoted as $\varsigma_i$which reflects the error correction abilityof M_i'.The model M_k'rectifying most images is selected.Model $M_k'$is selected based on the following rule:

$$k = arg\left\{\max_{M_i' \in Z}\left\{w1ri + w2\frac{\varsigma_i}{\ }\right\}\right\} \quad (3)$$

where$w1 + w2 = 1$, and $w1$ is the weight of face recognition rate, $w2$ is theweight of the rectifying rate.

3. $S \leftarrow S \cup M_K'$, $Z \leftarrow Z - M_K'$.
4. If $t < m$, let $t = t + 1$, then go to (a), else stop and m models$\{M^1, M^2, \dots, M^m\}$are selected.

3.3. *The ensemble of E-HMM based face recognition algorithms*

Using the m models $\{M^1, M^2, \dots, M^m\}$to recognize faces in the following method. Suppose there are H person's images in database containing one image per person, when the probe image is input, the m selected models are used to calculate the likelihood of it belonging to each person and a likelihood matrix is resulted:

$$\left| P(O^1|\lambda_1^1) \quad P(O^1|\lambda_2^1) \dots \quad P(O^1|\lambda_H^1) \right| \quad (4)$$

where the observation vector sequence of the probe image according to the model $M^i$is $O^i(i = 1, 2, \dots, m)$, the E-HMM of the model $M^i$is $\lambda_j^i(i = 1, 2, \dots, m; j = 1, 2, \dots, H)$, and $P(o^i|\lambda_j^i)$ $(i = 1, 2, \dots, m; j = 1, 2, \dots, H)$ is the likelihood of theprobe image corresponding to the jth person based on the ith model M^i.The weights of its m models for each face are $w^i(i = 1, 2, \dots, m)$ and $\Sigma_i w^i = 1$.There are many ways to set weight w^i, such as, average weighting, namely $wj = 1/m$; or setting weight according to the recognition rate:

$$w^i = \frac{\ }{m} \quad (5)$$

where$r^i$is the recognition rate of the ith model.In order to determine the identity of a given image, the likelihood matrix P istransformed into binary matrix

Pd:

$$P_d(O^i|\lambda_j^i = \{1 \quad if\ P(O^i|\lambda_j^i) = max_k\{P(O^i|\lambda_k^i)\} \quad (6)$$

Then, the decision is made according to the following equation:

$$D = \arg\left\{max_i\left\{\textstyle\sum_{i=1}^{m} w^i . P_d(O^i|\lambda_j^i)\right\}\right\} \quad (7)$$

4 Experimental Results and Analysis

Experiments are conducted to demonstrate the effectiveness of the proposed methodon The Face Recognition Technology (FERET) database[16] and Georgia Tech Face Database (GTFD) [17],[18](http://www.face-rec.org/databases) were used. The images are atthe resolution of 92×112 pixels, 8-bit gray levels. For each subject, 5 images wereselected randomly as gallery images in training, the left ones are used as probeimages, so sample set A is gained. Then swap the gallery and probe images, sampleset B is obtained. Both A and B are utilized to determine E-HMMs of faces forassembly. The sample sets C, D, and E are selected from database, 5 imageswere selected randomly as gallery images for each subject, the left 5 images are probe images. C, D and E are different from A and B. In experiment, the number of people is $L = 40$, the number of images for each person is $l = 10$, and number of gallery images is $l_1 = 5$, the number of probe images is $l_2 = 5$. The model set is $\Omega = \{M_1, M_2, \dots, M_{2000}\}$, the range of parameters is set as follows: sampling window size: $2 \times 2 \le P \times L \le 31 \times 31$, sampling step: $1 \times 1 \le \Delta_x \times \Delta_y \le 26 \times 26$, 2D-DCT coefficients: $4 \le N_{2D-DCT} \le 24$,

ISSN 2229-5518

number of pdfs: $2 \le N_s \le 7$, number of state: $1 \le N^i \le 9$. In model selection module, the weight of face recognition rate is $w1 = 0.05$, and the weight of rectifying rate is $w2 = 0.95$.

4.1 Experiment of face recognition based on single model

In this subsection, the N models of Ω are used to recognize the sample set C, Dand E, and the model recognizing each sample set optimally is denoted as M_C, $M_D$and M_E. Model $M_W$was randomly selected from Ω.Model M_T is used as a reference model.The performance of each model, such as M_C, $M_D$and M_E, is optimal only for some certain dataset. This means that each model specializes in expressing a type of feature information rather than all features and the generalization ability of single model is not strong enough.Parameters of theabove models are shown in Table 1.

For the given FERET database, the recognition rates of above five models areshown in Table 2, and the result indicates that the performance of each model, such as M_C, $M_D$and M_E, is optimal only forsome certain dataset. This means that each model specializes in expressing a type of feature information rather than all features and the generalization ability ofsingle model is not strong enough.

TABLE 1

The models and their parameters.

Models	P×L	$\Delta_x \times \Delta_y$	$N_{2D\text{-}DCT}$	N_s	K
M_C	8×8	3×3	12	7	6
M_D	18×18	4×4	10	4	6
M_E	12×12	2×2	6	7	3
M_W	31×31	24×24	6	2	2
M_T	10×8	2×2	24	5	3

TABLE 2

The recognition rates of five models on three test samples (%).

Models	Sample set C	Sample set D	Sample Set E	Average
M_C	98.4	97.7	98.6	98.23
M_D	97.5	98.5	98.8	98.26
M_E	97.2	98.4	98.3	97.96
M_W	55.5	45.5	62.0	54.33
M_T	95.5	98.5	94.0	96

4.2. Experiment of the two different sets

As previously stated, weights for sets can be set averagely or according to therecognition rate of each component model. A set of experiments are performed inorder to examine the performance of E-HMMensemble based on these two weighting-methods respectively. Model sets $M_A$and$M_B$are selected from Ω using the methodintroduced in Sec. 3.2, and then, the models in each set are equally weighted toconstitute M_A^e and M_B^e, while $M_A^u$and $M_B^u$are formed by unequally weighting themodels. Sample sets C, D and E are recognized by these four models. The result isshown in Table3.The model assembly method obtains higher recognition rate than single models,with lower variance, which shows that the proposed method leads to better andmore stable recognition performance, that is to say, it can deal with new datawith higher accuracy, and has strong generalization ability. This means that theproposed algorithm makes good use of recognition and error correction abilityof each single model, and these component models are complemented with eachother so as to improve recognition results.

TABLE 3

Recognition results of $M_A$and M_B on three sample sets (%)

Models	Sample set C	Sample set D	Sample Set E	Average
M_A^u	98.4	100	99	99.16
M_B^u	98.2	99.5	98.33	98.67
M_A^e	95.0	98.5	97.5	97
M_B^e	95.5	97.5	94.5	95.83

4.3 Comparison experiment of face recognition algorithms

To verify the stability and generalization ability of the proposed algorithm theselective ensemble of model sets $M_A$and $M_B$with unequal weighting is testedrespectively on dataset C, D and E.Also, we compare the our proposedalgorithm with different face recognition methods such as EigenFace [9], FisherFace[10],Laplacianfaces [11] ,ML-HMM[12], MCE-HMM[13][14], MC-HMM[15], our proposed method gives highest recognition accuracy.For all the sample sets, the recognition rate of the proposed algorithm in this paper is higher than that of the traditional algorithms based on HMM. Comparison result is summarized in Table.

TABLE 4

Comparison Results of Different Face recognition algorithms (%)

Models	Recognition rate
Proposed E-HMM	99.16
Eigenfaces	85.9
Fisherfaces	92.3
Laplacianfaces	93.2
ML-HMM	88.2
MCE-HMM	93.3
MC-HMM	97.5

5. Conclusion

ISSN 2229-5518

Based on the idea of discriminating set, a multiple E-HMMs based face recognitionalgorithm is proposed in this paper, which solves the model selection problemin E-HMM to some extent. The experimental result shows that this algorithmachieves higher recognition rate and stronger generalization ability. That is to say,this algorithm has a stronger ability to deal with new data. Of course, the assemblyof multiple E-HMMs will lead to high computational complexity. Fortunately, thenew algorithm is parallel virtually, which can be used to improve the efficiency byparallel computing.

References

[1] W. Zhao, R. Chellappa, A. Rosenfeld, P.J. Phillips, Face Recognition: A Literature Survey, ACM Computing Surveys, 2003, pp. 399-458

[2] Baum, L. E. & Petrie, T., Statistical inference for probabilistic functions of finite stateMarkov chains. *Annals of Mathematical Statistics*, vol. 37, 1966.

[3] Baum, L. E., An inequality and associated maximization technique in statisticalestimation for probabilistic functions of Markov processes, Inequalities, vol. 3, pp.1-8, 1972.

[4] F. Samaria, S. Young ,An HMM-based architecture for face identification. *Image and Computer Vision,* 12(8): 537-543, 1994.

[5] Nefian, A.V. & Hayes III, M.H., Face detection and recognition using hidden markov models. Image Processing, ICIP 98, Proceedings. 1998 International Conference*on*, Vol. 1, pp. 141–145, October 1998

[6] Nefian, A.V. & Hayes III, M.H. , An embedded hmm-based approach for face detection and recognition. *Acoustics, Speech, and Signal Processing, ICASSP '99.Proceedings,* IEEE International Conference, 6:3553–3556, March 1999.

[7] Rabiner, L., Juang, B, An introduction to hidden Markov models , IEEE ASSP Magazine, Jan 1986 Volume: 3 Issue:1 On page(s): 4 - 16 ISSN: 0740-7467

[8] V. N. Ara and H. H. Monson, "Face recognition using an embedded HMM," *Proc. IEEE Int'l Conf. Audio Video-based Biometric Person Authentication*, pp. 19–24(1999).

[9] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci.* **3**(1), 71–86 (1991).

[10] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 711-720, July 1997.

[11] X. He, S. Yan, Y. Hu, P. Niyogi and H. J. Zhang, "Face recognition using laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(3), 1–13 (2005).

[12] A.V. Nefian and M.H. Hayes III, "Maximum Likelihood Training of the Embedded HMM for Face Detection and Recognition," Proc. Int'l Conf. Image Processing, vol. 1, pp. 33-36, 2000.

[13] B.-H. Juang and S. Katagiri, "Discriminative Learning for Minimum Error Classification," IEEE Trans. Signal Processing,vol. 40, no. 12, pp. 3043-3054, 1992.

[14] S. Katagiri, B.-H. Juang, and C.-H. Lee, "Pattern Recognition Using a Family of Design Algorithms Based upon the GeneralizedProbabilistic Descent Method," Proc. IEEE, vol. 86, no. 11, pp. 2345-2373, 1998.

[15] Chien, J-T. & Liao, C-P. , Maximum Confidence Hidden Markov Modeling for Face Recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , Vol. 30,No 4, pp. 606-616, April 2008

[16] P.J. Phillips, H. Wechsler, J. Huang, and P.J. Rauss, "The FERET Database and Evaluation Procedure for Face-Recognition Algorithms," Image and Vision Computing, vol. 16, no. 5, pp. 295-306, 1998.

[17] L. Chen, H. Man, and A.V. Nefian, "Face Recognition Based Multi-Class Mapping of Fisher Scores," Pattern Recognition, vol. 38,pp. 799-811, 2005.

[18] A.V. Nefian, "Embedded Bayesian Networks for Face Recognition," Proc. IEEE Int'l Conf. Multimedia and Exp, vol. 2, pp. 133-136,2002.