# DWT-HMM based Face Recognition System

Vinayadatt V. Kohir* and U. B. Desai †
Dept. of Elec. Engg., IIT-Bombay
Powai, Mumbai, INDIA 400 076
email: {vvkohir,ubdesai@ee.iitb.ernet.in}

## Abstract

A transform domain approach coupled with Hidden Markov Model (HMM) for face recognition is presented. An ergodic HMM is trained and used for recognition with the transformed vectors of input image. The transformed vectors are obtained by applying DWT (Discrete Wavelet Transform) to the subimages generated by subsampling the input face image. The method is tested with ORL (Olevitti Research Laboratory) face database and in house, SPANN (Signal Processing and Neural Networks) laboratory face database . The recognition rate of 99.5% for ORL face database and 98.75% for SPANN face database is achieved.

Further, we present the recognition performance of the system for different sampling window size and varying overlap.

*Key words:* Face Recognition, HMM, DWT

## 1 Introduction

The present work on face recognition is carried out from the statistical analysis stand point. Intuitively, one feels that the face image can be segmented into visually distinguishable regions like mouth, eye, forehead, chin etc. It would have been a straight forward template matching problem, had there been a clear demarcation of these regions. Unfortunately, this is not the case. It is expected that an HMM would capture a "fuzzy" segmentation of the face, and thereby facilitate face recognition. HMMs were first used for face recognition by Samaria and Young [?].

Our face recognition system (Fig. 1)has two steps : First step is to build (or train) HMMs with the training face images. For, every subject to be recognized (i) DWT coefficients are computed for an $x \times y$ overlapping subimage of all the training face images (see Fig. 2), and these coefficients are arranged to for observation vector, Fig. 3. (ii) Using DWT coefficients an HMM is trained to model the face image. In the second step (see Fig. 1), i.e. recognition step, the unknown face image DWT coefficients are computed as in the training step. These coefficients are then used to compute the likelihood function; namely, $P[O, Q/\lambda_i]$. The recognized face corresponds to that $i$ which yields a maximum for $P[O, Q_i^*/\lambda_i]$, where $Q_i^* = \arg\max_Q P[O, Q/\lambda_i]$.

The method is tested on two face databases, namely ORL face database and SPANN face database. The best recognition performance of 99.5% for ORL face database and 97.5% for SPANN face database are obtained. Investigation is carried out to determine the dependence of recognition rate on the number HMM states and the overlap used to generate subimage sequence.

The training step is performed only once, and can be updated by training a new HMM for a new subject to be recognized.

To substantiate the findings, eigenface based method is used for face recognition yielding recognition rate of 88% for ORL face database and 78% for SPANN database.

In the following Subsection, related previous work is described. Section 2 gives a brief description on HMM and DWT. Section 3 presents the proposed approach. Experimental results are presented in Section 4.

### 1.1 Previous Work

Substantial effort has gone into understanding the face recognition problem and developing algorithms for face recognition. A detailed survey is provided in [?]. Due to space limitation we have not been able to include many useful references cited in [?]. Most of the algorithms exploit either template matching, Eigenfaces [?], artificial neural networks [?] or probabilistic models like HMM [?] or a combination of these paradigms [?, ?].

Eigenface based face recognition scheme of Turk and Pentland [?] proposes face recognition based on principal component analysis. An image is treated as a 2D random signal, and the covariance matrix of of

*On leave from P.D.A. College of Engg., Gulbarga.
†Corresponding author

2D face image is decomposed to extract principal components (Eigenfaces). The scheme assumes that face images and images other than face, are disjoint. The beauty of this method is reduction in the complexity of storage and computation. Authors report an accuracy in the range of 64% to 96%. The test set is of 2500 face images of 16 individuals digitized under different lighting conditions, head orientations and resolutions.

In an another approach Samaria and Young adopt, a top down HMM model for a face [?]. A separate HMM is trained for every subject to be recognized. The scheme uses grey level pixel values of the face image for training and recognition. With a test set of 50 face images of 24 individuals, success rate reported is 84%. All the training and test images were of the same size with no background.

## 2 Background

Briefly, basic principles of HMM and DWT are hilighted in this section.

### 2.1 Hidden Markov Model (HMM)

HMMs have been extensively exploited for speech recognition [?]. In recent years HMMs have also been used in vision. We cite a few papers illustrating the point : texture segmentation [?], face finding [?], object recognition [?], face recognition [?].

HMM is a doubly stochastic model for a 1D time series signal. It is associated with non-observable (hidden) states, and an observable sequence generated by the individual hidden states.

HMM is characterized by three parameters $(A, B, \Pi)$. Let $O = (o_1, o_2, \ldots, o_T)$ be the observed 1D signal at $T$ different observation instances and corresponding state sequence be $Q = (q_1, q_2, \ldots, q_T)$, where $q_t \in \{1, 2, \ldots, N\}$ , number of states. Let each of the observation be a vector of $D$ dimensions.

The HMM parameters $\lambda = (A, B, \Pi)$ are defined as follows :

$A$: is the transition probability matrix of size $N \times N$. The elements of $A$ are: $a_{ij} = P[q_{t+1} = j / q_t = i]$

$B$: is the emission probability matrix determining the output observation given that the HMM is in a particular state. Every element of this matrix;

$b_j(o_t) : \ 1 \leq j \leq N, \ \text{and} \ 1 \leq t \leq T$

is the posterior density of observation $o_t$ at time $t$ given that HMM is in state $q_t = s_j$.

$\Pi$: is the initial state distribution matrix of size $N \times 1$, with $i$-th entry: $\pi_i = P[q_1 = i]$, being the probability of being in state $i$ at the start of the observation.

### 2.2 DWT

Discrete Wavelet Transform (DWT) of function $f(\cdot) \in L^2(\mathcal{R})$ is defined as:

$DWT_{j,k} f = a^{-j/2} \int f(x) \overline{\psi(a^{-j}x - k)} dx$

where $\psi(\cdot)$ generated by dilation and translations:

$\psi_{j,k}(x) = a_0^{-j/2} \psi(a_0^{-j} x - k b_0)$

where $a_0 > 1$ and $b_0 > 0$ are fixed parameters, $j, k \in \mathcal{Z}$ cover the whole range of $\mathcal{Z}$. The parameter $a_0$ determines the width of $\psi_{j,k}(\cdot)$ and $n$ determines its translation. Generally, the values chosen for $a_0$ is 2 and for $b_0$ is 1; i.e. dyadic wavelet. The functions $\psi(\cdot)$ span the whole of $L^2(\mathcal{R})$ space forming an orthonormal basis functions. These basis functions $\psi(\cdot)$ are called as the wavelets. This concept can be easily extended to 2 dimetional case.

The DWT transforms spatial information to spatio-frequency information in the form of DWT coefficients. Further details are included in Appendix. Reader is advised to refer [?, ?] for detailed discussions on wavelets.

## 3 DWT-HMM Approach

Conventional HMMs are 1D data modelers. To facilitate them to be useful for 2D image applications, either they are to be redefined in two dimensional context, or the 2D data must be converted to 1D, without losing significant information.

In the present context, the second approach is used, i.e. 2D face image data is converted to 1D sequence using the DWT, and then the DWT coefficients are used for HMM training and testing. This crucial step of converting 2D image information to 1D vector is performed via data generation step discussed in the following Section.

### 3.1 Data Generation

Data generation consists of two steps. In the first step, a subimage of the face image is picked. In the next step, the DWT coefficients of the subimage are computed. These coefficients are then used to form an observation vector. This procedure is repeated on the sequence of overlapping subimages.

To generate the sequence of overlapped subimages, a square window is slid over the image in raster scan fashion from left to right with predefined overlap. Fig. 2 shows the sampling of a hypothetical image. At each location of the window over the image, the pixel values of the image are captured. These pixels represent a subimage of the original image. Next, the DWT coefficients of the subimage are computed. Only few coarser DWT coefficients are retained to form the observation vector (Fig. 3). This vector of DWT coefficients becomes the 1D observation vector for training HMM.

For every position of scan window we have a corresponding observation vector. The dimension of the observation vector is determined by the number of DWT

coefficients retained. In our experiments we have chosen the number to be 16. Each image gives an observation sequence, whose length is determined by the size of the sliding window and the amount of overlap between adjacent windows.

The 2D face image is now represented by a corresponding 1D spatio-frequency domain observation sequence, which is suitable for training HMM.

### 3.2 HMM for Face Recognition

An ergodic HMM is trained for very subject to be recognised. The training of each of the HMM is carried out using the DWT based training sequence. Biorthagonal spline wavelets [?] are used in decomposing subimages.

#### 3.2.1 HMM Training Steps

**Step 1**: Cluster all $\Omega$ training sequences, generated from $\Omega$ number of face images of one particular subject, i.e. $\{O^{\omega}\}$, $1 \leq \omega \leq \Omega$, each of length $T$, in $N$ clusters using some clustering algorithm, say k-means clustering algorithm. Each cluster will represent a state of the training vector. Let them be numbered from 1 to $N$.

**Step 2**: Assign cluster number of the nearest cluster to each of the training vector. i.e. $t^{th}$ training vector will be assigned a number $i$ if its distance, say Euclidean distance, is smaller than than its distance to any other cluster $j$, $j \neq i$.

**Step 3**: Calculate mean $\{\mu_i\}$ and covariance matrix $\{\Sigma_i\}$ for each state (cluster).

$\mu_i = \frac{1}{N_i} \sum_{o_t \in i} o_t$ for $1 \leq i \leq N$

$\Sigma_i = \frac{1}{N_i} \sum_{o_t \in i} (o_t - \mu_i)^t (o_t - \mu_i)$ for $1 \leq i \leq N$

where $N_i$ is the number of vectors assigned to state $i$.

**Step 4**: Calculate $A$ and $\Pi$ matrices using event counting.

$$\pi_i = \frac{\text{No. of occurrences of } o_1 \in i}{\text{No. of training sequences,}\Omega} \text{for } 1 \leq i \leq N$$

$$a_{ij} = \frac{\text{No. of occurrences of } o_t \in i \text{ and } o_{t+1} \in j}{\text{No. of occurrences of } o_t \in i}$$

for $1 \leq i, j \leq N$ and for $1 \leq t \leq T-1$

**Step 5**: Calculate the $B$ matrix of probability density for each of the training vector for each state. Here we assume that $b_j(o_t)$ is Gaussian. For $1 \leq j \leq N$

$$b_j(o_t) = \frac{1}{(2\pi)^{\frac{D}{2}} |R|^{\frac{1}{2}}} exp[-\frac{(o_t - \mu_i)^t \Sigma_j^{-1} (o_t - \mu_i)}{2}]$$

where, $o_t$ is of size $D \times 1$.

**Step 6**: Now use the Viterbi algorithm to find the optimal state sequence $Q^*$ for each training sequence. Here, the state reassignment is done. A vector is assigned state $i$ if $q_t^* = i$.

**Step 7**: The reassignment is effective only for those training vectors whose earlier state assignment is different from the Viterbi state. If there is any effective state reassignment, then repeat Steps 3 to 6; else STOP and the HMM is trained for the given training sequences.

The details of Viterbi algorithm can be found in [?].

#### 3.2.2 Face Recognition

Once, all the HMMs are trained, then we can go ahead with recognition. For the face image to be recognized, the data generation step is followed as described in Section 3.1. The trained HMMs are used to compute the likelihood function as follows: Let $O$ be the DWT based observation sequence generated from the face image to be recognized,

**1:** Using the Viterbi algorithm, we first compute $Q_i^* = \arg\max_Q P[O, Q/\lambda_i]$

**2:** The recognized face corresponds to that $i$ for which the likelihood function $P[O, Q_i^*/\lambda_i]$ is maximum.

## 4 Experimental Evaluation

Two sets of face databases: ORL face database[1] and SPANN face database [2] are used to experiment. The face data bases have both male and female subjects with and without facial accessories like glasses, beard etc. There is no background for the face image. The lighting, head orientation or the facial expression are not monitored while capturing the face images.

ORL face database constitutes 40 subjects with 10 different poses each. These face images are 256 grey level images of size $92 \times 112$. 5 poses of a subject are used to train the HMM. The reaming 5 poses, which are not part of training are used for recognition. Thus, in all, the method is tested on 200 face images. Some of the training and testing face images are in Fig. 5.

SPANN database is the inhouse face database with face images stored in IRIS RGB format. Each subject has 7 different poses and are of size $128 \times 128$. The poses are not monitored precisely. All the face images of a subject are captured in one sitting. The face subjects chosen here are male or female, with or without glasses, with or without beard. Changes in facial expression as well face orientation are permitted. 20 subjects, at random, are selected for the experiment. The color images are converted to 256 grey level images to suit the method. 3 pose per subject are used to train HMM, and the remaining 4 poses of the subject are used to test. Representative face images are displayed in Fig. 6.

### 4.1 Training

For every subject a HMM is trained as follows:

**Step 1:** Sample the image with a square window, and take DWT of the sub-image enclosed. Scan the DWT

[1] http://131.111.20.59/pub/data/orl_faces.tar.Z

[2] http://www.khayal.ee.iitb.ernet.in/usr/SPANN_DATA_BASE/2D_Signals/Face/faces/

matrix, as indicated in Data generation step and select only coarser coefficients to form 1D vector. Complete sampling of the face image from top-left corner down upto right-bottom of the image with 75%. This gives an observation sequence.

**Step 2:** Repeat Step 1 for all the training images. This step gives the set of observation sequences for HMM training.

**Step 3:** Use the training algorithm described earlier to train the HMM.

### 4.2 Recognition

The recognition test is performed on 200 images (5 for each subject) of ORL face database and 80 images (4 for each subject) of SPANN face database. The test images are not part of training images.

**Step 1:** For the input image, generate the DWT based observation sequence. Note that the same size window, same amount of overlap and same number of DWT coefficients arranged in the same order, as used in the training phase, are used while generating this data.

**Step 2:** Use the Viterbi algorithm to decode the state sequence and find the state-optimized likelihood function for all the stored HMM, namely, $\forall\, i$, $P[O, Q_i^*/\lambda_i]$, where $Q_i^* = \arg\max_Q P[O, Q/\lambda_i]$, and $O$ is the dct-based observation sequence corresponding to the image to be recognized.

**Step 3:** Select that label of the HMM for which the likelihood function is maximum. By varying the size of sampling window, % overlap and number of HMM states the experiments are carried out with the same set of images.

We have also implemented the Eigenface based method of Turk and Pentland [?], and an HMM based method which works strictly on the image. This HMM based method is analogous to that of Samaria and Young [?], except that we use an ergodic HMM as opposed to [?], where they use a top-down HMM.

## 5 Results

Table ?? has recognition results for the subimage size of $8 \times 8$ and $16 \times 16$ with 50%, 75%, and 85% overlap. The recognition rate varies from 84% to 99.5% for ORL case and 80% to 98.75% for SPANN case. This implies that for the best recognition rate: for ORL case 1 image out 200 images is misrecognised and for SPANN case also out of 80 test images 1 is misrecognised.

Figure 4 gives an investigation result of the dependence of recognition rate on the number of HMM states. The graph is plotted for both ORL and SPANN face databases. The sampling window of $16 \times 16$ with 75% overlap is used to investigate the dependence.

Comparative results of proposed DWT-HMM scheme, Plain HMM and Eigenface methods are indicated in Table ??. The HMM approaches the proposed scheme gives much better recognition percentage then either the Eigenfaces based scheme or the HMM based scheme without DWT. We are able recognize faces which have significant changes in their expression, as well in their orientation.

To substantiate proposed DWT-HMM method, comparison is made with other methods (see Table ??). The % recognition and corresponding timings are as reported by the respective authors.

## 6 Conclusions

For the face data set considered, the results of the proposed DWT-HMM based scheme show substantial improvement over the results obtained from Eigenface based method and the non-DWT based HMM method. The recognition rate depends on overlap, window size and even the number of HMM states. There is no consistent pattern (see Fig. 4) of recognition rate versus number of HMM states.

## References


[1] R. Chellappa, C.L. Wilson and S.A. Sirohy, "Human and Machine Recognition of Faces: A Survey" , *IEEE Proceedings*, Vol. 83, No. 3, pp. 704-740, May 1995.

[2] A. Cohen, I. Daubecheis and J.-C. Feauveau, "Biorthogonal Bases of Compactly supported Wavelets", *Communications on Pure and Applied Mathematics*, vol. XLV, pp. 485–560, 1992.

[3] I. Daubecheis, "*Ten Lectures on Wavelets*", SIAM, 1992.

[4] Joachim Hornegger , Heinrich Niemann, D.Pulus and G. Schlottke, "Object Recognition using hidden Markov Models", *Proc. of Pattern Recognition in Practice*, pp. 37-44, 1994.

[5] S. Lawarance, C.L. Giles, A.C. Tsoi and A.D. Back, "Face Recognition : A Convolutional Neural-Network Approach", *IEEE Trans. on Neural Networks*, Vol. 8, No. 1, pp. 98-113, Jan. 1997.

[6] S.H. Lin, S.Y. Kung and L.J.Lin, "Face Recognition/Detection by Probabilistic Decision Based Neural Network", *IEEE Trans. on Neural Networks*, Vol. 8, No.1, pp. 114-132, Jan.1997.

[7] S. M. Lucas, "Face Recognition with Continuous n-tuple Classifier", *BMVC'97*.

[8] L. R. Rabiner, "A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", *IEEE Proceedings*, Vol. 77, No. 2, pp. 257-285, 1989.

[9] A.N. Rajagopalan, K. Sunil Kumar, Jayashree Karlekar, Manivasakan, Milind Patil, U.B. Desai, P.G. Poonacha, S. Choudhuri, "Finding Faces in Photographs", *ICCV98*, pp. 640-645, Jan. 1998.

[10] F.S. Samaria and S. Young, "HMM-based Architecture for Face Identification", *Image and Vision Computing*, Vol. 12, No. 8, pp. 537-543, Oct. 1994.

[11] M. Turk and A. Pentland, "Eigenfaces for Recognition", *Journal of Neuroscience*, Vol. 3, No. 1, MIT 1991.

[12] Vinayadatt V. Kohir and U. B. Desai, "Face Recognition Using a DCT-HMM Approach", To appear in *proceedings of WACV'98*, Oct. 19-21, 1998, Prienceton, NJ, USA.

[13] Vinayadatt V. Kohir and U. B. Desai, "Face Recognition Using DCT-HMM Approach", Communicated to *Image Vision Computing* Journal, 1998.

[14] Yang He and Amalan Kundu, "Unsupervised Texture Segmentation Using Multichannel Decomposition and Hidden Markov Models", *IEEE Image Processing*, Vol. 4, No. 5, pp. 603-619, May 1995.

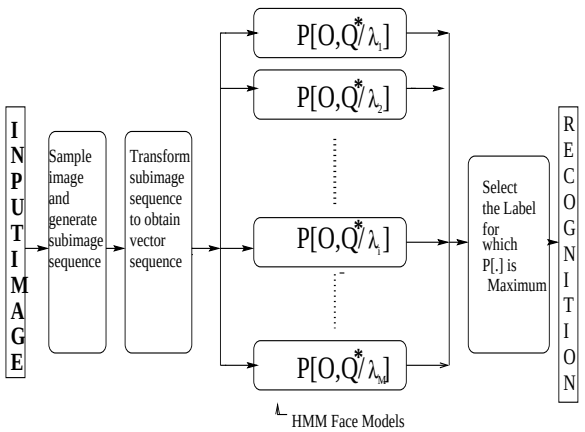



Figure 1: Proposed DWT-HMM face recognition system.

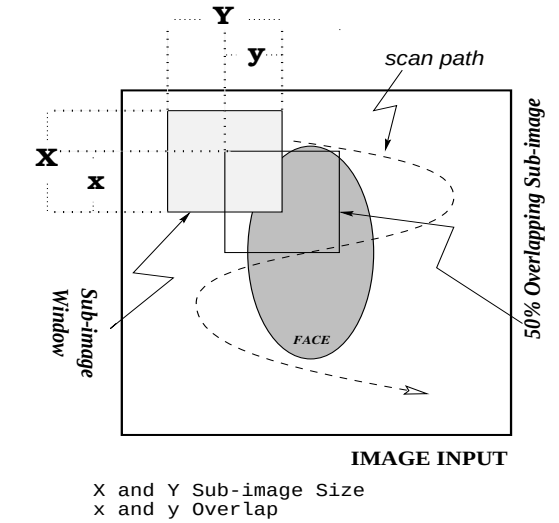



Figure 2: Subimage sequence generation by sampling the face image by sliding a square window.

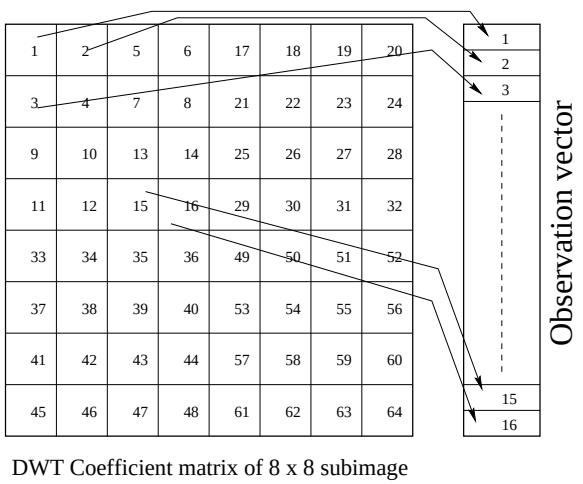
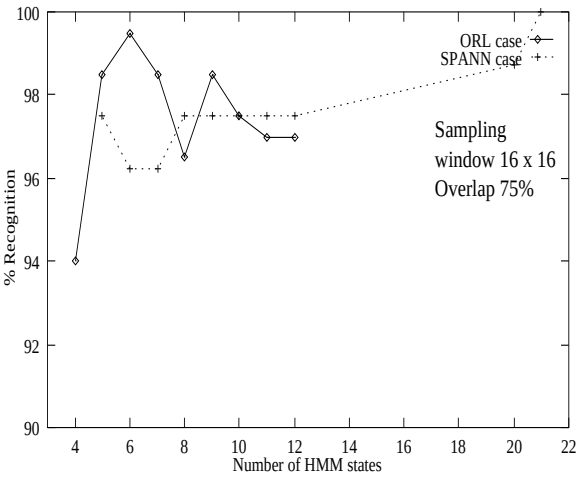



Figure 3: Vector generation from DWT coffients.




Figur 4: Plot of number of HMM states vs % Recognition.

Figure 5: Sample ORL face images. First four rows represent training face images and the remaining four represent corresponding test face images. Crossed face image of bottom row is misrecognised for the best case

Figure 6: Sample SPANN face images. First three columns represent training face images and the last four columns represent test face images used for recognition. Crossed image in the first row is the misrecognised face image for the best case

| Sampling window size | % Overlap | % Recognition | |
|---|---|---|---|
| | | SPANN Face Database | ORL Face Database |
| $8 \times 8$ | 50 | 83.33 | 84.0 |
| | 75 | 97.2 | 89.5 |
| $16 \times 16$ | 50 | 80.0 | 85.0 |
| | 75 | 97.5 | 98.5 |
| | 85 | 98.75 | 99.5 |

Table 1: Recognition results of DWT-HMM based face recognition for SPANN and ORL face database. The number of HMM states is 5

| | % Recognition | |
|---|---|---|
| | ORL Face Database | SPANN Face Database |
| DWT-HMM (85% overlap (and 16 coeffs.) | 99.5 | 98.75 |
| Plain HMM (75% overlap) | 84 | 81 |
| Eigenface | 88 | 77.5 |

Table 2: Comparative recognition results of proposed DWT-HMM method, eigenface method and Plain HMM. The best results are tabulated

| Method | % Recognition | Training Time | Recognition Time |
|---|---|---|---|
| SOM+CN [?] | 96 | 4hrs. | 0.5s |
| PDBNN [?] | 96 | 20min | $\leq$0.1s |
| n-tuple [?] | 86 | 0.9s | 0.025s |
| cont.n-tuple [?] | 96 | 0.9s | 0.025s |
| Top-Down HMM [?] | 84 | n/a | n/a |
| Pseudo-2D HMM [?] | 95 | n/a | 240s |
| Eigenface | 88 | 9.4s | 0.62s |
| DCT-HMM[?] (75% overlap,15 coeffs.) | 99.5 | 23.50s | 3.50s |
| DWT-HMM (75% overlap,16 coeffs.) | 98.5 | 44.02s | 6.94s |

Table 3: Comparative recognition results of some of the other methods as reported by the respective authors on ORL face database.