Maximum Confidence Hidden Markov Modeling for Face Recognition
Jen-Tzung Chien* and Chih-Pin Liao

Department of Computer Science and Information Engineering, College of Electrical Engineering and Computer Science, National Cheng Kung University
*Email: jtchien@mail.ncku.edu.tw


Face recognition has been known as one of key applications to build high-performance surveillance or information security systems [1][2][3][4]. The statistical paradigm based on hidden Markov model (HMM) has been successfully developed for face recognition [3]. Generally, HMM is effective to represent time-series signal such as speech data where time alignment is a critical concern. It can be extended to align two-dimensional (2D) image signal and model spatial variations, e.g. facial expressions, orientations and beard, etc. Figure 1 displays the embedded HMM representation for face images with different facial features.

In this work, we concern two issues in establishing HMMs for face recognition. The first issue involves the hybrid process of feature extraction and model estimation.

Secondly, we develop a new discriminative training criterion derived by statistical hypothesis test theory. By maximizing the confidence towards accepting the hypothesis that sub-images are from target HMM state against the hypothesis that sub-images are from competing HMM states, the maximum confidence (MC) criterion is exploited for estimation of MC-HMMs, where the feature transformation W is naturally embedded. The facial features and HMM parameters are jointly estimated for discriminative face recognition. In recognition phase, we present a doubly Viterbi segmentation to obtain optimal state
and mixture component sequences. The implementation of MC-HMM estimation procedure is shown in Figure 2. Initial HMM parameters are estimated. Blocks of input images are realigned into super states and embedded states through Viterbi segmentation. Having the optimal state and mixture component sequences, we calculate HMM parameters corresponding to super states and embedded states. Transformation matrix $W$ is then calculated. The training procedure of MC-HMM converges after several iterations.

![Flowchart](image)

**Figure 2:** Implementation procedure of MC-HMM training.

In the experiments, we compare the segmentation of FERET facial images [4] using ML-HMM and MC-HMM through Viterbi state alignment. ML-HMM denotes the maximum likelihood HMM, which is known as the baseline model. In the comparison, MC-HMM with feature dimension reduction from $d=36$ to $d=16$ is considered. Typically, HMM state represents spatial characteristics in image data. State alignment shall be perfect if HMM parameters are well trained. As demonstrated in Figure 3, state alignment using MC-HMM is much better than that using ML-HMM. MC-HMM with reduced feature dimension obtains quite good alignment not only in characterizing the vertical facial segments via super states but also the horizontal tiny textures via embedded states. We can align image blocks into the best states with maximum confidence. With the discriminative model, classification performance is assured for future test data.
In addition to ML-HMM and MC-HMM, we examine the performance of minimum classification error (MCE)-HMM, which is known as popular discriminative training algorithm.

Figure 3: State alignment or segmentation of FERET face images using ML-HMM (Top Row) and MC-HMM (Bottom Row).

Figure 4: Recognition accuracies of using different feature dimensions and class numbers.
For comparative study, non-HMM methods using Eigenface and Fisherface [1] are implemented. The class numbers $C=50$ and $C=100$ are investigated. In Figure 4, HMM methods are significantly better than non-HMM methods. MCE-HMM and MC-HMM do outperform ML-HMM. MC-HMM achieves the best classification performance. For the case of $C=50$, ML-HMM obtain recognition accuracy 89%, which is improved to 92.4% using MCE-HMM and 94.4% using MC-HMM with $d=16$. When MC-HMM with $d=36$ is implemented, we achieve the accuracy as high as 95.6%. Eigenface and Fisherface only attain accuracies of 80% and 81.3%, respectively. Similar results can be obtained for $C=100$.

References: