A REAL TIME AUTOMATIC ACCESS CONTROL SYSTEM BASED ON FACE AND EYE CORNERS DETECTION, FACE RECOGNITION AND SPEAKER IDENTIFICATION

Ziyou Xiong, Yunqiang Chen, Roy Wang and Thomas S. Huang

Beckman Institute for Advanced Science and Technology
University of Illinois at Urbana-Champaign
Urbana, IL 61801
E-mail: {zxiong, chenyq, rwang, huang}@ifp.uiuc.edu

ABSTRACT
We have recently developed a real time automatic access control system using both face and voice identification. We show that combining these two modalities achieves high recognition accuracy, especially in difficult recognition scenarios. Our system also supports online training of the user models, i.e., a person who has not been registered in the database can be added on the spot in real time. This system is composed of a face and eye corners detection module, a face recognition module, a speaker identification module and a user interface module. We present the algorithm of each module with more emphasis on our information maximization based face and eye corners detection and the user interface design. We discuss some issues related to the integration of components into the total system. We report the experimental results of the system performance together with comparison with those of some other systems.

1. INTRODUCTION AND RELATED WORK
Most of the proposed approaches focus on mono-modal recognition based on a biometric feature (face, voice, iris, · · ·). Relatively high recognition rates were obtained for different modalities such as face recognition and speaker identification[1][2]. However, they still do not meet the very low error rates that can be tolerated in many of the access control applications such as access to buildings, computers, surveillance and intrusion detection. It has been shown that combining different modalities leads to more robust system performance[3].

We have studied face detection in the information maximization prospective and developed a fast system using Information-Based Maximum Discrimination (IBMD)[4]. The high accuracy of face and eye corners detection enables us to deal well with the translation sensitivity of the eigenface method for face recognition. Recognition with facial analysis alone has already produced very good results. To achieve higher recognition rate, we also use the speaker identification to assist face recognition.

The rest of the paper is organized as follows. A short summary of our IBMD-based approach is presented in Section 2. The new person recognition system is presented in Section 3 that performs multi-modal person recognition in real-time video based on this method together with speaker identification. Experimental results on face detection and person recognition are detailed in Section 4. Conclusion and future work are presented in Section 5.

2. OVERVIEW OF OUR FACE AND EYE CORNERS DETECTION ALGORITHM
2.1. Theory of Our Algorithm
We want to decide which probability distribution is more likely in generating the observation image \( O \) to conclude whether \( O \) is face or non-face using the maximum likelihood (ML) test:

\[
L(O) = \frac{P_F(O)}{P_N(O)}
\]

(for notation convenience, \( O \) for observation, \( F \) for "face" and \( N \) for "non-face")

Our assumption is that the permuted version of \( O = (o_1, \cdots, o_n) \), \( O' = (o_{s_1}, \cdots, o_{s_n}) \) comes from a \( k_{th} \) order Markov process \( X' = (X'_1, \cdots, X'_n) = (X_{s_1}, \cdots, X_{s_n}) \) where \( s_1, \cdots, s_n \) is a permuted sequence from \( 1, \cdots, n \). The task that follows is to find such an optimal permutation.

This optimal permutation is the one that maximizes a cost function which we choose to be the Kullback divergence (cross entropy) between \( P_F(O') \) and \( P_N(O') \)

\[
H_{F||N}(X') = \sum_{O'} P_F(O') \log \frac{P_F(O')}{P_N(O')}
\]

where \( H_{F||N}() \) denotes cross entropy. If we assume \( O' \) comes from a \( 1_{st} \) order Markov process, using recursively
the chain rule of the Kullback divergence:

\[ H_{F|N}(Z_n, \ldots, Z_1) = \sum_{i=n}^{1} H_{F|N}(Z_i|Z_{i-1}, \ldots, Z_1) \quad (3) \]

and the afore-mentioned Markovian property, we have

\[ H_{F|N}(X') = H_{F|N}(X_{s_1}) + \sum_{i=2}^{n} H_{F|N}(X_i|X_{s_{i-1}}) \quad (4) \]

Therefore, we want to find optimal \( S^* = (s_1^*, \ldots, s_n^*) \) such that

\[ H_{F|N}(X' = X(S^*)) \geq H_{F|N}(X(S)) \quad \forall S \quad (5) \]

where \( S \) denotes a permutation of sequence \((1, \ldots, n)\) and \( X(S) \) denotes a permutation of \( X = (X_1, \ldots, X_n) \) according to the permutation in \( S \).

If \( H(s_m|s_{m-1}) \) is thought to be a distance from vertex \( s_m \) to \( s_{m-1} \) and \( H(s_1) \) a distance from a fixed starting point (different from any of the \( n \) vertices \( 1, \ldots, n \)) to vertex \( s_1 \), then the physical meaning of \( H(S) \) is the maximum distance of a path starting from the fixed starting point, traverse each and every vertex \( 1, \ldots, n \) exactly once. The optimal solution gives the optimal traversing path \( s_1, \ldots, s_n \). Figure 1 is an illustration. This is closely related to the NP-Complete "Travelling Salesman Problem(TSP)" in graph theory where this optimal path is called the Hamiltonian path[5]. Note that this optimization problem, similarly to the TSP, in practice would not be solved exhaustively. However, a modified version of the Kruskal’s Algorithm for minimum-weight spanning tree[5][6] has shown be able to obtain sub-optimal, but very good results.

![Fig. 1. Finding the optimal path starting from the fixed node S and traversing each and every node exactly once. Dashed-line arrows always start from node S. Solid-line arrows are bi-directional. Each arrow is associated with a weight (i.e., distance). The sub-graph on the left with node V1, Vn is fully connected by the solid-line arrows. The one on the right is an illustration of the optimal path.](image)

The ML ratio test \( L(O) \), or equivalently \( L(O') \) is simplified to the following form:

\[ L(O') = \frac{P_F(O')}{P_N(O')} = \frac{\prod_{i=n}^{1} P_F(a_i^*|a_{i-1}^*)P_F(a_i^*)}{\prod_{i=n}^{1} P_N(a_i^*|a_{i-1}^*)P_N(a_i^*)} \quad (6) \]

Notice now only pair-pixel and single pixel statistics are needed to estimate the probabilities.

Once a suboptimal solution \( S' \) is found, the classifier is implemented by a look-up table that holds the logarithm of the likelihood ratios \( \log L_{s_i|s_{i-1}^*} \) and \( \log L_{s_i'|s_{i-1}^*} \) for \( i = 2, \ldots, n \) so that, given an observation vector \( x_i \), its log-likelihood ratio is computed very efficiently.

2.2. Eye Corners Detection

The aforementioned detection algorithm is also used to detect the 2 outer eye corners with positive examples to be eye images. For an online demonstration of the system on still images, please see the demo at http://www.ifp.uiuc.edu/~antonio/Demo/SelectImage.html.

3. A MULTI-MODAL PERSON RECOGNITION SYSTEM

The face and eye corners detection algorithm is utilized as the kernel of a recently developed multi-modal person recognition system. It achieves a detection rate of \( 11 \sim 12 \) fps for \( 360 \times 240 \) video[7] on a Pentium III 600 MHz processor. In comparison, Viola and Jones’s face detection based on Adaboost achieves a detection rate of 15 fps for \( 384 \times 288 \) video on a Pentium III 700 MHz processor[8]. Please see the video demo at http://www.ifp.uiuc.edu/~zxiong/Demo.html.

The computer interface of the new system and the interface description are shown in Figure 2.

3.1. Face Recognition Using PCA

The detected face region and the two outer eye corners are highlighted by the green square and two red crosses(+) respectively at the upper-right corner of the interface window. 100 normalized faces of size \( 24 \times 24 \) are cropped out based on the detected eye corner positions for each of the 24 users. After histogram equalization to minimize the effect of lighting condition difference, they are used to train the user models based on Principle Component Analysis(PCA), i.e., eigen-face based approach[9][10]. K-nearest neighbor classifier is used to recognize a new face as belonging to one of the prototypes in the database. Out of the K-nearest neighbors of 2400 training points, we establish the statistics of how many are from the training set of each trained user. This statistics is then transformed into probabilities of the test face image nearest to each training set.

3.2. Speaker Identification Based on MFCC and GMM

In the mean time, a speaker identification system is also developed running in parallel with face recognition. The speaker ID system uses Mel-frequency
3. cepstrum coefficients (MFCC)[2] as low level features and models each user’s speech as a Gaussian Mixture Model (GMM)[11]. Maximum likelihood is used as classifier to classify a new user’s speech captured by an on-the-desk microphone to be from one of the users in the database. Both the likelihoods from K-nearest neighbor classifier and maximum likelihood classifier are displayed in the two bar charts respectively in Figure 2. The final decision is made by simply the multiplication of the two sets of probabilities.

3.3. Online Training of both Face Model and Speaker Model

The system also supports online training. A new user who has not yet been in the database can have his/her model added to the database. 30 seconds of training speech is recorded to train his/her GMM model after the user issues the ”Add User” command and reads the text prompted on the screen. MFCC+GMM for speaker ID is not language dependent. The experimental subject can speak different languages both in training and testing.

While the user is reading the text or digits, face detection detects his/her face and prepares normalized faces for training his/her face model. The features of the new user(s) are extracted by projecting the faces to the PCA space trained using the original 24 users. The new user’s face image is added to the icon image set in the lower-center of Figure 2. The two bar charts and the list of user names are also updated to reflect the addition of a new user. After 30 seconds of training, the user can ask the system to test whether he/she can be recognized by the system. The user can also have his/her model deleted or replaced with a new one.

4. EXPERIMENTAL RESULTS

4.1. Face Detection Results on Image Databases

We use a face classifier to test sub-windows of $16 \times 14$ pixels. The preprocessing algorithm consists of a histogram equalization and a re-quantization procedure so that four grey levels are used to feed the classifier.

In [4], we have reported the face detection results on the CMU/MIT database. Figure 3 is the receiver operating characteristic (ROC). The vertical axis shows the correct answer rate, while the horizontal is the false alarm rate. It is one of the most robust and fastest face detectors reported in literature.
4.2. Results of the person recognition System

Testing of face detection results in real-time video is done by turning off the functionality of face recognition and speaker ID in Figure 2. 42 subjects have been tested for face detection in real-time video. Several video sequences of each subject have been taken on different days under different lighting conditions. Different sizes of faces have also been tested. Each sequence is between 1 minute and 5 minutes long.

Face detection results consistently show high detection accuracy (≈ 95%) and very low false alarm rate (< 0.1%). Given the near real-time face detection, a pattern of face tracking by the green square and the red crosses can be observed.

Experiments on face detection+face recognition alone without speaker identification and vice versa have shown nearly perfect recognition precision. Each of the 24 original subjects whose models are used in training is correctly recognized using one modality alone. The combination of two modalities shows 100% accuracy as a result of both strong recognition in each individual modality.

The person recognition system recognizes both users who have been already in the database and the new users who have been added to the database by online training with very high accuracy. In one demonstration, a set of 22 new people were added to the database one by one and all of them have been correctly recognized.

4.3. Known Problems

First, false detection of outer eye corner positions degrades the face recognition accuracy due to PCA’s sensitively to translation. Second, our face detection has problems detecting faces of people wearing very reflective glasses. Last but not least, like many of other systems ours can only handle small degree of in-plane rotation. A relatively large rotation of the user’s head will degrade the face detection greatly.

5. CONCLUSIONS AND FUTURE WORK

In this paper a real time automatic access control system is described using both face and voice identification. We present the algorithm of each module of this system with more emphasis on our information maximization based face and eye corners detection and the user interface design.

In the future, we will further improve the system on the problems listed above, improve eye corner detection accuracy, research on face recognition under greater variability and side-view face detection and recognition.

6. REFERENCES