Threshold for Face Classes in Face Recognition

A dissertation submitted in partial fulfillment of the requirements for the Completion of degree of Master of Technology in Computer Science

By

Naveen Kumar Kukkadapu

(2006 – 2008)

Indian Statistical Institute

Kolkata

Under the supervision of

Prof. C. A. Murthy

Machine Intelligence Unit

<u>INDIAN STATISTICAL INSTITUTE</u>

203, Barrackpore Trunk Road

Kolkata – 700108

Certificate

This is to certify that the dissertation entitled **“Threshold for face classes in Face Recognition”** submitted by Naveen Kumar Kukkadapu towards partial fulfillment for the degree of M.Tech in Computer Science at Indian Statistical Institute, Kolkata, embodies the work done under my supervision.

Dated:

Signed:

(Prof.C.A.Murthy)

Supervisor

Countersigned:

External Examiner

Acknowledgement

It is my privilege and pleasure to convey my gratitude to my guide **Prof.C.A.Murthy** for his support and guidance throughout the duration of dissertation without which it would have been impossible to complete this work.

Naveen Kumar

Contents **Page no**

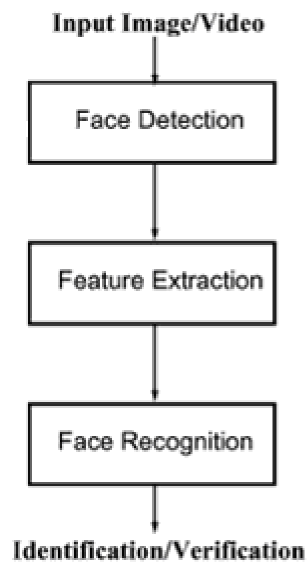
Chapter-1

Introduction

Task of Face Recognition:

Given still or video images of a scene, identify or verify one or more persons in the scene using a stored database of faces.
The solution to the problem involves the segmentation of faces (face detection) from image of a scene, feature extraction from face region, recognition.

A basic structure of a Face Recognition System is shown in the following diagram.



Basic Structure of a generic Face Recognition System

The main parts are typically face detection and face recognition which can it self be decomposed in normalization, feature extraction and classification steps.

Face detection aims to determine whether or not there are any faces in the image and, if present, return the face location while the goal of face localization is to estimate the position of a single face.

Feature extraction is to find a specific representation of the data that can highlight relevant information. At the feature extraction stage, the goal is to find an invariant representation of the face image. Usually, an image is represented by a high dimensional vector containing pixel values (holistic representation) or a set of vectors where each vector contains gray levels of a sub-image (local representation). There are different methods existing in the literature.

i) Holistic Methods
ii) Structural Matching Methods
iii) Hybrid Methods

i) ***Holistic methods:*** These methods use the whole face region as the raw input to the system for extracting the features of a face image.

Principal Component Analysis:

One of the feature extraction techniques, based on Principal Component Analysis (PCA), was first used for face recognition by Turk and Pentland [4]. The aim of PCA is to find a representation of the data minimizing the reconstruction error. The PCA finds the orthogonal directions that account for the highest amount of variance. The data is then projected into the subspace spanned by these directions. In practice, the principal component axes are the eigenvectors of the covariance matrix of the data. The corresponding eigen values indicate the proportion of variance of the data projections along each direction.

Linear Discriminant Analysis:

Another feature extraction method used in face recognition is based on Linear Discriminant Analysis (LDA), also known as Fisher Discriminant Analysis [5].

The LDA subspace holds more discriminant features than the PCA subspace. LDA finds a subspace in which the variability is maximized between different class data, and at the same time where variability in the same class data (face images of the same identity) is minimum.

We define the *within-class scatter matrix* as S_w and *between-class scatter matrix* as S_b. The goal is to maximize the between-class measure while minimizing the within-class measure. One way to do this is to maximize the ratio $\frac{\det(S_b)}{\det(S_w)}$. Intuitively, for face recognition, LDA should outperform PCA because it inherently deals with class discrimination. However, Martinez and Kak [5] have shown that PCA might outperform LDA when the number of samples per class is small.

ii) ***Structural matching methods:*** In these methods, local features such as the eyes, nose, and mouth are first extracted and their locations and local statistics (geometric and/or appearance) are fed into a structural classifier.
E.g. Pure Geometry Methods

iii) Hybrid Methods: Use both local features and whole face region to recognize a face.

Classification Task:

The classification step consists of attributing a class label to the input data. Some of the well known classifiers are nearest neighbor classifier and minimum distance classifier. Current approaches for face recognition often make use of simple image similarity metrics such as the Euclidean distance between the feature vector of reference images and feature vector of the test image. Because of curse of dimensionality problem, the distance metrics are not computed in the image space but in an appropriate subspace such as PCA or LDA. Some more appropriate metrics have been proposed in the literature such as Mahalanobis distance, Normalized correlation.

Limitation of the above Methods:

All the above methods concern about the dissimilarity (i.e., generally the metrics like Euclidean distance or Mahalanobis distance) between representations of the query image and the training images. We classify the test face image into the class whose distance is the least from the training images. In this approach the test image always fall into one of the face classes of the database. This process of identification known as closed set identification

In case, if the test image is not a valid face image (i.e., image of an object or a scene or a face image which is not in our database), it will be classified to one of the face classes of our database, then this type of identification is called open set identification. Here the test image is said to be imposter to the system.

We have to find a threshold value on the dissimilarity value that will decide whether a test image is either a valid face or not.

Existing Method for finding the threshold values:

ROC (Receiver Operator Characteristics) Based Method:

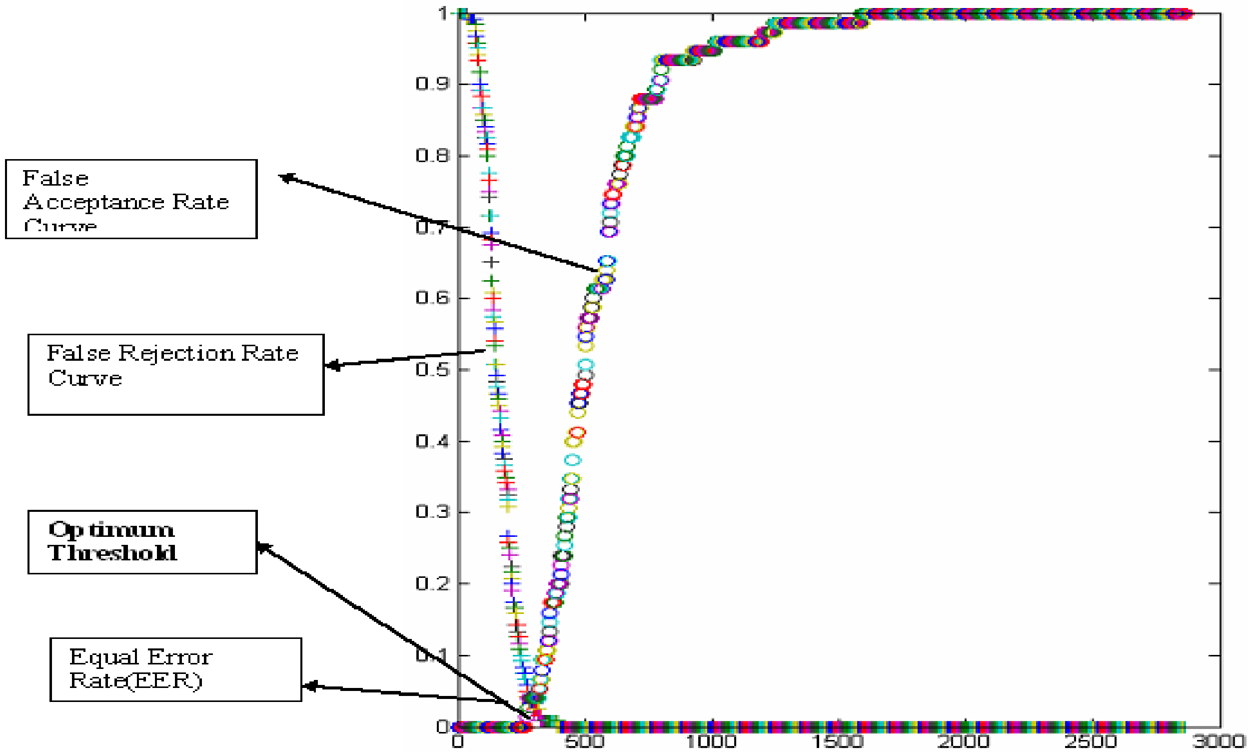
Mansfield et al [6] used a method of selecting a threshold based on false acceptance rate (FAR) and false rejection rate (FRR). Meanings of FAR and FRR [6] can be explained as follows.

We have three types of data sets. They are training, validation and test sets. Validation and test sets contain both the client images and imposter images. It is used to get hold of a threshold value. Then test set is used for verifying the performance of the threshold.

After dimensionality reduction of each observation, we calculate the FAR (False Acceptance Rate) and FRR (False Rejection Rate) using validation set for each of the threshold in a specified range.
Where FAR = Number of False Acceptances / Number of Imposter images.
FRR = Number of False Rejections / Number of Client images.

On a range of thresholds (from minimum to maximum), for each of them, we calculate both false acceptance rate and false rejection rate in classifying the validation set.

We consider a threshold as the final threshold using the Validation Set for which both the FAR and FRR are same (EER = Equal Error Rate).We call this threshold as "operating threshold". Test Set is used to verify the performance of the above obtained threshold. An example of an ROC curve is shown below which is performed on ORL database with minimum distance classifier.



ROC curve considered ORL Database as Training Set(40*5) and Validation set as ORL database(40*3) & Yale Database(as imposter set)(15*5) is shown above.

Limitations of ROC based Method:
If the number of imposter images is less in validation set, then we may not get a good estimate of number of false acceptances so is FAR.
On increasing the number of imposter images in the validation set, we get a good estimate of false acceptance rate. Note that the number of imposters (i.e., non-face images *or* face-images not belonging to our databases *or* images of objects) are infinite. But there is a computation problem when we consider validation set is bigger one. To overcome this limitation, we have to find a method which is not involving the Validation Set.

Our new proposal method based on **Set Estimation Technique [1] by U.Grenander.**

Chapter – 2

Set Estimation

Consistent Estimate of a Set:

Definition: Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of independent and identically distributed random vectors which follow some continuous distribution over the set $\alpha \subseteq \mathbb{R}^m$, where α is an unknown quantity. Let α_n be an estimated set based upon the random vectors $X_1, X_2, \ldots, X_n, \ldots$ Then α_n is said to be a consistent estimate of α, if $E_\alpha[\mu(\alpha_n \Delta \alpha)] \to 0$ as $n \to \infty$,
where Δ denotes symmetric difference, μ is the lebesgue measure and E_α denotes the 'expectation' taken under α.

Theorem: Let $X_1, X_2, \ldots, X_n, \ldots$ be independent and identically random vectors, which follow uniform distribution over $\alpha \subseteq \mathbb{R}^2$, where α is unknown .Let α be such that $cl(Int(\alpha)) = \alpha$ and $\mu(\delta\alpha) = 0$ where $\delta\alpha$ denotes the boundary of α, *cl* denotes closure, *Int* denotes the interior. Let $\{\varepsilon_n\}$ be a sequence of positive numbers such that $\varepsilon_n \to 0$ and $n\varepsilon_n^2 \to \infty$ as $n \to \infty$ .Let $\alpha_n = \bigcup_{i=1}^{n} \{x \in R^2 : d(x, X_i) \le \varepsilon_n\}$ where 'd' denotes the Euclidean Distance .Then α_n is a consistent estimate of α.

Suppose we consider $\varepsilon_{n1} = K\varepsilon_n$ for some constant $K>0$, where $\varepsilon_n \to 0$ and $n\varepsilon_n^2 \to \infty$ as $n \to \infty$ .Then the sequence $\{\varepsilon_{n1}\}$ also satisfies the property that $\boldsymbol{\varepsilon_{n1} \to 0}$ **and** $\boldsymbol{n\varepsilon_{n1}^2 \to \infty}$ **as** $\boldsymbol{n \to \infty}$**.**

There can be several ways of choosing the sequence $\{\varepsilon_n\}$ exist in the Literature .One of the mostly used one which is developed by Prof.C.A.Murthy [2] which is based on Minimal Spanning Tree (MST).

Let $S = \{X_1, X_2, \ldots, X_n\} \subseteq \Re^2$ .Considering the MST of set S, say G_n .Consider the weight of a edge as the Euclidean distance between the two vertices joining by it.

i) Find the length of G_n , l_n = Sum of the weights of the edges of MST.

Then we can consider the sequence $\varepsilon_n = \sqrt{\frac{l_n}{n}}$ which satisfy the property $\varepsilon_n \to 0$ and $n\varepsilon_n^2 \to \infty$ as $n \to \infty$.

With this ε_n as radius, consider all the points in each disc with centers as points in S which is the consistent estimate of the set α.

i.e., $$A_n = \bigcup_{x \in G_n} \{ y \in \Re^2 : d(x,y) \le \varepsilon_n \}.$$

(Also it is generalized to $\Re^m$ **)**

ii) Another way of choosing sequence $\{\varepsilon_n\}$,

$\varepsilon_n =$ (Maximum of the (n - 1) edge weights of MST) /2

Our Proposed 'ε_n' for a Single class: Consider the set of points from α **,**
S= $\{ \underset{\to}{x_1}, \underset{\to}{x_2}, \ldots, \underset{\to}{x_n} \} \subseteq \Re^m$ from a single class.

Where $\underset{->}{x_i} = (x_{i1}, x_{i2}, \ldots, x_{im})$**,** $i = 1, 2, 3, \ldots n$**.**

For each dimension, consider the coordinates of feature vectors in it's direction. (along each axis).After sorting these values, find the maximum consecutive difference among them in each dimension. Then find the half of the maximum of all these dimension wise maximums. This value is ***our proposed*** $\boldsymbol{\varepsilon_n}$ and the estimated set with this proposed radius value is,

$$A_n = \bigcup_{x \in S_n} \{ y \in \Re^m : d(x,y) \le \varepsilon_n \}.$$

Basic intuition of the above algorithm:

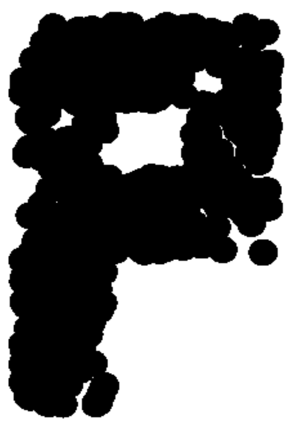
According to the already existing ε_n value which is based on Minimal Spanning Tree Edge weights [2], the above ε_n value is proposed. i.e., Our Proposed ε_n value is approximation of the (maximum of edge weights of MST)/2.

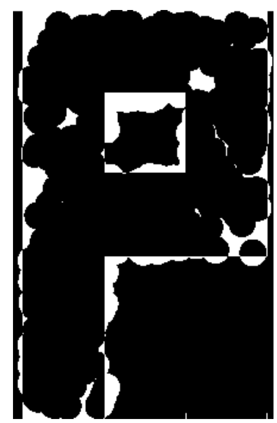
<u>Example of Set Estimation Technique:</u> Consider the data set containing all the points in the image of the letter "**P". **Our "**P"** set is taken as collection of squares. To select a point in A_n, we selected a random square first and then from the square a random point.

In this way, we considered a random number of points (A_n), say n, from the set,
α = ([0,1] X [0,5]) U ([1,2]X[2,3]) U ([1,2]X[4,5]) X ([2,3]X[2,5])

We have drawn the disks with the radius value which is equal to our proposed ε_n value with the points of A_n as centers. Then the estimated sets for n=500, 1000, 1500 are shown below. We can see that as the 'n' value increases the consistency of estimated set increases. i.e., the difference in area between the original "**P**" set and the estimated set decreases.

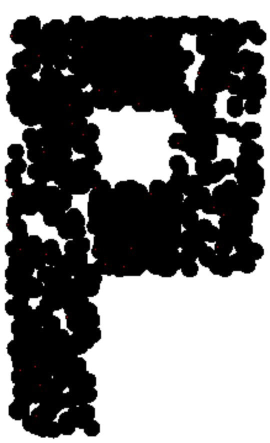
n= 500

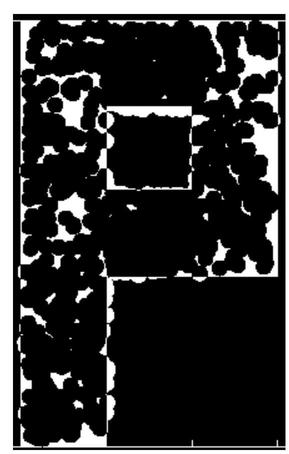




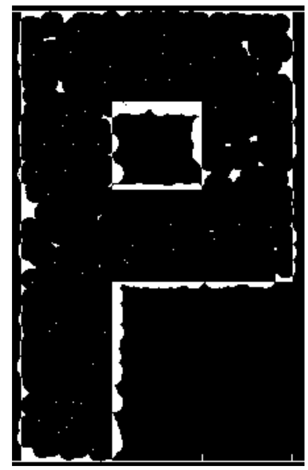
Estimated set   Difference between the original & estimated sets

n = 1000





Estimated set   Difference between the original&estimated set n = 1500



The symmetric difference ($A_n \Delta \alpha$) in estimated set and the original set is shown in white color. We can see that white color is decreasing as with the increasing of 'n' value.i.e. the estimated set tending towards the original set

Chapter 3

Proposed Methods for finding Threshold for Face Recognition

In the case of ROC based threshold, we get only one threshold for whole database. But the threshold which is based on set estimation technique is proposed for a single class. In order to get a single threshold representing whole database, we aggregate all the thresholds for each class. We can apply the proposed threshold to Face Recognition as a face class of a human being is a connected set. (path connected) First we calculate our proposed threshold for each face class of a given database. Then we can combine the threshold in the following ways.

i) Consider the mean of the thresholds of all the subjects (face classes)

i.e. mean (maximum_classwise (maximum_dimwise (consecutive_diff))/2)

ii) Consider the median of the thresholds of all the subjects (face classes)

i.e. Median (maximum_classwise (maximum_dimwise (consecutive_diff))/2)

Remarks: We have considered taking the mean as the way for our proposed method .But one may consider some other ways of combining the thresholds of face classes.

Chapter 4

Description of Dataset, Training Set, Validation Set and Test Set

Here we considered two different face databases named ORL and Yale Databases.

ORL Database: It contains face images of 40 different persons and for each person having 10 different images taken under different conditions.

Yale Database : It contains face images of 15 different persons and for each Person having 11 different images taken under different conditions.

We can form three different types of datasets namely Training Set, Test Set and Validation Set which are called TVT combination.

Training set is used to form a client space. Validation Set is used to get hold of a threshold (ROC Method).and the Test set is used to verify the performance of a threshold.

Validation Set and Test Set contains both the Client Images and Imposter Images. (Which are *either* non-face images *or* face images of persons not from the training set *or* images of objects like vehicles, Computer etc.,)

We have considered various types of TVT combinations using the above two different databases ORL and Yale as in the following table.

********Here are the different TVT combinations considered********

TVT	Training Set	Validation Set	Test Set
TVT1	ORL – 40*5	ORL – 40*3 Yale – 15*5	ORL-40*2 Yale – 15*6
TVT2	**ORL-20*5**(odd)	ORL - 20*3(odd) ORL – 20*5(even)	ORL -20*2odd) ORL – 20*5(even)
TVT3	ORL-20*5(**even**)	ORL - 20*3(even) ORL – 20*5(odd)	ORL -20*2even) ORL – 20*5(odd)
TVT4	ORL-20*5(odd)	ORL - 20*3(odd) ORL – 20*5(even) Yale -15*5	ORL -20*2odd) ORL – 20*5(even) Yale – 15 * 6
TVT5	Yale-15*5	Yale -15*3 ORL - 40*5	Yale – 15*3 ORL -40*5
TVT6	Yale-8*5(odd)	Yale – 8*3(odd) Yale – 7*5(even)	Yale – 8*3(odd) Yale – 7*6(even)
TVT7	Yale-7*5(even)	Yale – 7*3(even) Yale – 8*5(odd)	Yale – 7*3(even) Yale – 8*6(odd)

TVT8	Yale-8*5(odd)	Yale – 8*3(odd) Yale – 7*5(even) ORL – 40*5	Yale – 8*3(odd) Yale – 7*6(even) ORL – 40*5

Chapter 5

Dimensionality Reduction

Each face image in each database is cropped (removed some noisy part) and resized to the size 80x100 and then we consider each face image as a column vector.

On each database, by applying Principal Component Analysis, we reduced the dimension of it.

We have considered various dimensions such as 60, 80,100 for our experiments.

For each dimension, we have selected the 5 training images from each class with 10 different combinations. In each combination, the remaining images from the corresponding classes are considered for validation images and test images.

E.g.: From the ORL Database, consider for some i^{th} class (i = 1 to 40), a

Combination may be,

Training images are 1, 5,7,4,8.

Validation images are 2, 9

And the remaining are for testing purpose 3, 10.

Chapter 6

Results

We have applied our proposed threshold for different combinations of datasets described above. Finally we have tabulated the Performance of the the ROC based threshold and Our Proposed threshold in the terms of Recognition Rate which is defined as below. Let's say the test set is with ,

the number of Client images = M

the number of accepted Client Images = M_1

the number of Imposter Images = N

the number of Imposter Images accepted = N_1

Then the definition of Recognition Rate (RR) is $\frac{M_1 + (N - N_1)}{M + N} * 100$.

The results are shown below containing the performances of ROC based threshold and proposed threshold value. Except for the cases TVT1 and TVT3, in all the

Remaining cases the proposed threshold is giving the better results than ROC based threshold.

TVT	ROC based threshold	ROC based recognition rate(FA,FR)	Proposed (mean) method threshold	Proposed method threshold
TVT1	309	98.2353(0, 3)	138.1386	68.8235(0, 53)
TVT2	177	74.2857(26, 10)	91.1295	77.8571(1, 30)
TVT3	183	80.7143(20, 7)	104.6672	79.2857(0, 29)
TVT4	190	81.3043(38, 5)	91.1295	86.5217(1, 30)
TVT5	342	74.6939(57, 5)	287.4657	84.0816(27, 12)
TVT6	330	71.2121(17, 2)	272.9237	80.3030(8, 5)
TVT7	229	76.8116(13, 3)	207.7101	86.9565(6, 3)
TVT8	356	73.6842(68, 2)	272.9237	92.4812(15, 5)

Description of Results: Proposed threshold (i) average of the thresholds of the face classes, is better performing than the ROC based threshold except in TVT 1, 3 combinations.And also in the case of ROC based method, we get different values of Threshold for one Training Data (as validation set varies) But in the case of Proposed Method, we get only one threshold for one training set (irrespective of Validation Set)

Chapter 7

Conclusion and Discussion

In distance based classification systems, any test point will be classified into one of the classes even it is not a member of the training set. Similarly for face recognition systems using the dissimilarity based classification, if the test image is an imposter image then it will fall into one of the face classes. In that case, it is needed to put a threshold value basing on which one can say that the test image is a valid or not. In this thesis a new thresholding technique is proposed which is based on the set estimation technique. A face class of a person is a connected set. Because of this we can apply set estimation technique to find the threshold value basing on it. We have found the mean of the thresholds of each face class in a face database. We have used the minimum distance classifier for classification and different combinations of TVT (training, validation, test sets) and experimented our proposed threshold value in each case. It is performing better than ROC based threshold except in some cases of our experiments.

Chapter 8

References

[1] U.Grenander, Abstract Inference,. John Wiley, New York 1981

[2] Madhura Datta and C.A.Murthy, Intra-class Threshold Selection in Face Space using set Estimation Technique. (communicative paper)

[3] C.A.Murthy "On Consistent Estimation of Classes in $\Re^2$ in the context of Cluster Analysis", Ph.D Thesis, Indian Statistical Institute.

[4]M.Turk and A.Pentland. "Eigenfaces for Recognition," Journal of Cognitive Neuroscience.3 (1), pp 71-86, 1991.

[5] A. M. Martinez and A.C.Kak. "PCA versus LDA". In IEEE Transactions on pattern analysis and machine intelligence,Vol 23, 2001.

[6] A.J.Mansfield and J.L.Wayman, Best Practices in Testing and Reporting performance of Biometric Scientific Computing, National Physical Laboratory, Queens Road,Teddington, Middlesex , TW11 OLW,August 2002.

[7] P.J.Phillips, P.Grother, R.J.Micheals, D.M.Blackburn, E.Tabassi, and J.M.Bone, "FRVT 2002. Evaluation Report," March 2003,

http://www.frvt.org/DLs/FRVT_2002_Evaluation_Report.pdf

[8]. W. ZHAO, R. CHELLAPPA, P. J. PHILLIPS, A. ROSENFELD "Face Recognition: A Literature Survey" University of Maryland, 2003