# Learning Models for Predicting Recognition Performance

Rong Wang, Bir Bhanu
Institute Center for Research in Intelligent Systems
University of California, Riverside
Riverside, California 92521, USA
{rwang, bhanu}@vislab.ucr.edu

## Abstract

*This paper addresses one of the fundamental problems encountered in performance prediction for object recognition. In particular we address the problems related to estimation of small gallery size that can give good error estimates and their confidences on large probe sets and populations. We use a generalized two-dimensional prediction model that integrates a hypergeometric probability distribution model with a binomial model explicitly and considers the distortion problem in large populations. We incorporate learning in the prediction process in order to find the optimal small gallery size and to improve its performance. The Chernoff and Chebychev inequalities are used as a guide to obtain the small gallery size. During the prediction we use the expectation-maximum (EM) algorithm to learn the match score and the non-match score distributions (the number of components, their weights, means and covariances) that are represented as Gaussian mixtures. By learning we find the optimal size of small gallery and at the same time provide the upper bound and the lower bound for the prediction on large populations. Results are shown using real-world databases.*

## 1. Introduction

Recognition systems can classify images, signals or other types of measurements into a number of classes. In this paper we mainly focus on biometric recognition systems. Biometrics can be fingerprint, palm, face, gait, signature or speech. Usually a biometric recognition system consists of three stages: image acquisition, feature extraction and matching. Distortion often occurs in these stages and is caused by sensor noise, feature uncertainty, feature occlusion, and feature clutter.

Before we can evaluate the performance of a recognition algorithm on large populations we need to answer some fundamental questions. Since the algorithm performance of recognition systems is usually based on limited data, it's difficult to estimate this performance for additional data: the limited test data may, after all, not accurately represent the larger population. When we use a small gallery to estimate the algorithm performance on large populations how can we find the optimal size of the small gallery and how accurate is the estimation? Since the prediction is based on the same recognition algorithm, we can give the confidence interval for the performance estimation of the large population [1]. The confidence interval can describe the uncertainty associated with the estimation. This gives an interval within which the true algorithm performance for the large population is expected to fall, along with the probability that it is expected to fall there [2]. Guyon et al. [1] propose guaranteed estimators to determine the test size for the independent identical distribution recognition error and the correlated recognition error, along with the assumption of the underlying probability distribution.

Grother et al. [3] introduce the joint density function of the match score and the non-match score to estimate both the open-set and the closed-set identification performance. They assume that the match score and the non-match score are independent and their distributions are the same for large populations.

Estimation of the match score and the non-match score distributions are very important for prediction. Grother et al. [3] use the Monte Carlo sampling method to linearly interpolate the match score and the non-match score lookup tables. Johnson et al. [4] use the count method to compute the error probability for a given match score.

In this paper we use a generalized prediction model that integrates a hypergeometric probability distribution model explicitly with a binomial model which takes into account distortion that may occur in large populations. The prediction model provides performance measurements as a function of rank, large population size, the number of distorted images, and similarity score distributions. While we use the expectation-maximum (EM) algorithm to estimate the match score and the non-match score distributions, we introduce learning to feed back similarity scores to increase the small gallery size. In this way we can find the optimal size of the small gallery to predict the large population performance. Meanwhile, we provide upper and lower bounds for the prediction performance for the large population. In this paper we use two different statistical methods–Chernoff's inequality and Chebychev's inequality–to obtain the relationship between the small gallery size and the confidence interval given a margin of error.

Our paper is organized as follows. Contributions are presented in section 2. In section 3 we describe the details of the integrated model, the procedure of learning for similarity score distributions in the prediction, and the statistical methods to find the optimal sample size. Experimental results are shown in section 4. The integral model with learning is tested on the NIST-4 fingerprint database. Conclusions are given in section 5.

## 2. Contributions

1) We use a generalized prediction model that integrates a hypergeometric probability distribution model explicitly with a binomial model which takes into account any dis-



Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05)
1550-5499/05 $20.00 © 2005 IEEE

tortion that may occur in large populations. Our distortion model includes feature uncertainty, feature occlusion and feature clutter. In the prediction model we use the EM algorithm to estimate match score and non-match score distributions and find the number of components automatically. For each component we can get its mean, covariance, and weight which represent the underlying Gaussian mixture model. For a subset of the biometric database if the error between the prediction and the actual performance is larger than a margin of error then we feedback the similarity scores to the EM algorithm until the error is smaller than the margin of error.

2) By learning we can find the optimal size of a small gallery, and at the same time we can give the upper bound and the lower bound for a large population prediction. We use the Chernoff inequality and the Chebychev inequality to determine the small gallery size which is related to the margin of error and the confidence interval. Then we can give the confidence interval for the prediction performance.

3) The results are shown on a large data set of fingerprint images.

# 3. Technical Approach

Figure 1 provides the conceptual diagram of our system. For a given biometric recognition system whose size is $M$, we randomly pick $n$ images to be our small gallery. By identification we can get a set of match scores and non-match scores for this small gallery. Then we use the expectation-maximization (EM) algorithm to estimate distributions of the match score and the non-match score. Based on these distributions we use our prediction model which integrates a hypergeometric probability distribution model explicitly with a binomial model to estimate the recognition system performance for a large population whose size is $M_1$ where $M_1 < M$. We assume the prediction performance on $M_1$ is $\hat{p}$. From the recognition system we can obtain the match score and the non-match score for $M_1$, then compute the actual recognition performance $p$ for $M_1$. $\tilde{e}$ is the error between the prediction performance and the actual performance, where $\tilde{e} = |\hat{p} - p|$. If $\tilde{e}$ is larger than the margin of error $e$ then we feed back match scores and non-match scores to the EM algorithm to estimate the similarity score distributions again. Otherwise we increase $M_1$, the size of the large population, and repeat this process until the $M_1$ increases to $M$. We will explain each part of the diagram in detail in this section.

## 3.1. Prediction Model

Our two-dimensional prediction model considers the distortion problem which conforms with reality. Assume we have two kinds of different quality biometric images, group #1 and group #2. Group #1 is a set of good quality biometric images without distortion. Group #2 is a set of poor quality biometric images with distortion. In general, the size of these two groups are $n_1$ pairs and $n_2$ pairs. We randomly pick $n$ pairs of images from group #1 and group #2. Then the pair number of distorted images $y$ which are chosen from group #2 should follow the hypergeometric distribution

$$f(y) = \frac{\mathrm{C}_{n-y}^{n_1}\mathrm{C}_{y}^{n_2}}{\mathrm{C}_{n}^{n_1+n_2}} \tag{1}$$

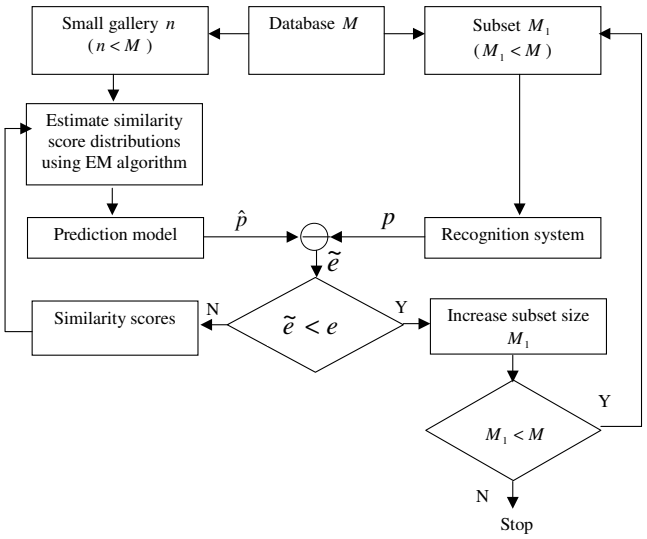


Figure 1: Diagram of the prediction system

where $n_1 + n_2$ is the total number of images in these two groups and $n-y$ is the number of images chosen from group #1.

These $n$ pairs of images are our small gallery. We split them into the gallery and the probe set. For each image in the probe set we compute the similarity scores with every image in the gallery. Then we have one match score and $n-1$ non-match scores for this image. Here we assume that the match score and the non-match score are independent. With these similarity scores we can use the EM algorithm to estimate the match score and the non-match score distributions.

From above we know that the similarity score distributions depend not only on the similarity scores but also on the number of images with distortion. Here we assume $ms(x|y)$ and $ns(x|y)$ represent the distributions of match scores and non-match scores given the number of distorted images. Assume if the similarity score is higher then the biometrics are more similar. The error occurs when a given match score is smaller than the non-match score. For a given number of distorted images the probability that the non-match score is greater than or equal to the match score $x$ is $NS(x)$ where

$$NS(x) = \int_x^\infty ns(t|y)f(y)dt \tag{2}$$

Then the probability that the non-match score is smaller than the match score is $1 - NS(x)$.

Here we assume that the similarity score distributions are similar for the small gallery and the large population. If the size of the large population is $N$, then for the $jth$ image we can have a set of similarity scores, which includes one match score and $N-1$ non-match scores. We rank the similarity scores in decreasing order. Then for a given number of images with distortion the probability that the match score $x$ is at rank $r$ is given by the binomial probability distribution

$$\mathrm{C}_{r-1}^{N-1}(1 - NS(x))^{N-r}NS(x)^{r-1} \tag{3}$$

Integrating over all the match scores, for a given number of images with distortion the probability that the match scores



Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05)
1550-5499/05 $20.00 © 2005 IEEE

are at rank $r$ can be written as

$$\int_{-\infty}^{\infty} \mathrm{C}_{r-1}^{N-1}(1-NS(x))^{N-r}NS(x)^{r-1}ms(x|y)dx \quad (4)$$

We integrate over all the number of images chosen from group #2, the probability that the match scores are at rank $r$ can be written as

$$\int_{-\infty}^{\infty} \mathrm{C}_{r-1}^{N-1}(1-NS(x))^{N-r}NS(x)^{r-1} \sum_{y=0}^{n} ms(x|y)f(y)dx \quad (5)$$

In theory the match scores can be any values within $(-\infty, \infty)$. We get the probability that the match scores are within rank $r$ is

$$P(N,r) = \sum_{i=1}^{r}\int_{-\infty}^{\infty} \mathrm{C}_{i-1}^{N-1}(1-NS(x))^{N-i}NS(x)^{i-1} \sum_{y=0}^{n} ms(x|y)f(y)dx \quad (6)$$

Given that the correct match takes place above a threshold $t$, the probability that the match score is within rank $r$ becomes

$$P(N,r,t) = \sum_{i=1}^{r}\int_{t}^{\infty} \mathrm{C}_{i-1}^{N-1}(1-NS(x))^{N-i}NS(x)^{i-1} \sum_{y=0}^{n} ms(x|y)f(y)dx \quad (7)$$

When rank $r = 1$ then the prediction model with threshold $t$ becomes

$$P(N,1,t) = \int_{t}^{\infty} (1-NS(x))^{N-1} \sum_{y=0}^{n} ms(x|y)f(y)dx \quad (8)$$

In this model we make two assumptions: match scores and non-match scores are independent and large populations have distortion. In this model $N$ is the size of the large population whose performance needs to be estimated. A small sized gallery is used to estimate distributions of $ms(x|y)$ and $ns(x|y)$ .

## 3.2. Expectation-Maximization Algorithm

The EM algorithm is an iterative method to estimate the likelihood given good data [5]. We assume that the data distribution is a $c$-component mixture model $C = C_1, \cdots, C_c$, whose distribution can be written as

$$f(x) = \sum_{i=1}^{c} \pi_i f_i(x) \quad (9)$$

where $x$ is $d$-dimensional data, $f_i(x)$ are component densities and $\pi_i$ are component proportions. The component densities are specified by the parameter vector $\theta = (\theta_1, \cdots, \theta_c)$. Let the vector $\Psi$ contain all the unknown parameters in the mixture model

$$\Psi = (\pi_1, \cdots, \pi_{c-1}, \xi^T)^T \quad (10)$$

where $\xi$ contains all the parameters in $\theta_1, \cdots, \theta_c$. Here we rewrite (9) as

$$f(x;\Psi) = \sum_{i=1}^{c} \pi_i f_i(x;\theta_i) \quad (11)$$

Given a set of $N$ independent and identical distribution samples $\chi = \{x_1, \cdots, x_N\}$ from equation (11), the maximum likelihood (ML) estimation of the unknown parameter vectors $\theta_i$ can be obtained by the EM algorithm. We set the associated binary component-indicator vectors for $\chi$ as $Z = \{z_1, \cdots, z_N\}$, which is associated with the $N$ samples and indicates which component produces these samples. $z_{ji}$ means sample $x_j$ is produced by the $ith$ component. The complete data log-likelihood function is given by

$$logL(\chi, Z; \Psi) = \sum_{j=1}^{N}\sum_{i=1}^{c} z_{ji} log[\pi_i f_i(x_j;\theta_i)] \quad (12)$$

The EM algorithm produces a sequence of estimations $\hat{\Psi}(k)$ by proceeding iteratively in two steps (the E-step and the M-step) until some termination criterion is met.

(1) E-step: Defines the conditional expectation of $Z$, whose elements are defined as $\tau_{ji} = E_{\hat{\Psi}(k)}(z_{ji}|\chi)$. By the Bayesian theory, it can be derived as

$$\tau_{ji} = \frac{\pi_i^{(k)} f_i(x_j;\hat{\theta_i^{(k)}})}{\sum_{h=1}^{c} \pi_h^{(k)} f_h(x_j;\hat{\theta_h^{(k)}})} \quad (13)$$

(2) M-step: Updates the estimation of $\Psi$ by

$$\hat{\Psi}(k+1) = argmax\Phi(\Psi;\hat{\Psi}(k)) \quad (14)$$

The updated expression for the component is

$$\pi_i^{(k+1)} = \frac{\sum_{j=1}^{N}\tau_{ji}^{(k)}}{N}(i = 1, \cdots, c) \quad (15)$$

When $c$ is unknown, we can select the value of $c$ according to some criterion function,

$$\hat{c} = argmax\Upsilon(\hat{\Psi}_c, c), c \in c_{min}, \cdots, c_{max} \quad (16)$$

where $\hat{\Psi}_c$ is the mixture parameter estimation when the model is assumed to contain $c$ components. The criterion function $\Upsilon(\hat{\Psi}_c, c)$ usually consists of two terms as log-likelihood of the data for the model and the penalty function.



Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05)
1550-5499/05 $20.00 © 2005 **IEEE**

### 3.3. Determine the Small Gallery Size

In this section we discuss the relationship between the confidence interval and the small gallery size. We use limited data to estimate the large population performance. Therefore the prediction value may be significantly accurate or not. This question can be mathematically expressed as

$$P\{|(p-\hat{p})| > e\} \leq (1-\alpha) \tag{17}$$

Where $\hat{p}$ is the predicted performance for the recognition system which can be obtained from our prediction model, $p$ is the actual performance of the recognition system, $e$ is the margin of error for the system, and $\alpha$ is the confidence interval. Then inequality (17) can be written as

$$P\{p > \hat{p} + e\} \leq (1-\alpha) \tag{18}$$

or

$$P\{p < \hat{p} - e\} \leq (1-\alpha) \tag{19}$$

Here we consider inequality (18) since inequality (19) is symmetric with inequality (18).

We assume that a recognition system recognizes biometrics with the probability $P\{X_i = 1\} = p$ and $P\{X_i = 0\} = 1-p$, where $X_i = 1$ means biometrics $X_i$ is recognized correctly, $X_i = 0$ means the opposite. According to the Chernoff inequality [6], let $X_1, X_2, \cdots, X_n$ be independent random variables. For any $X_i$, we have $P\{X_i = 1\} = p$ and $P\{X_i = 0\} = 1-p$, where $0 < p < 1$. We define the random variable

$$X = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{20}$$

For any $t \geq 0$ we have:

$$P\{X \geq E(X) + \frac{t}{n}\} \leq e^{-\frac{2t^2}{n}} \tag{21}$$

Comparing with inequality (18), we can get

$$1-\alpha = e^{-\frac{2t^2}{n}} \tag{22}$$

So,

$$t = \sqrt{-\frac{nln(1-\alpha)}{2}} \tag{23}$$

Thus, equation (21) becomes

$$P\{X \geq E(X) + \sqrt{-\frac{ln(1-\alpha)}{2n}}\} \leq \alpha \tag{24}$$

From inequality (18), we know that

$$e = \sqrt{-\frac{ln(1-\alpha)}{2n}} \tag{25}$$

Thus, we get

$$n = \frac{-ln(1-\alpha)}{2e^2} \tag{26}$$

In the above we assume that the recognition system can recognize biometrics with a certain distribution. If we do not know the underlying distribution of the recognition system then we can use the Chebychev inequality [6] which is distribution independent. Assume $X_1, X_2, \cdots, X_n$ are independent random variables defined as

$$X = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{27}$$

For any $\varepsilon \geq 0$, we have

$$P\{|X - E(X)| \geq \varepsilon\} \leq \frac{\sigma^2}{n\varepsilon^2} \tag{28}$$

Then we have

$$1-\alpha = \frac{\sigma^2}{n\varepsilon^2} \tag{29}$$

From the above equation we obtain

$$\varepsilon = \frac{\sigma}{\sqrt{n(1-\alpha)}} \tag{30}$$

From equations (28), (29), and (30) we have

$$P\{X \geq E(X) + \frac{\sigma}{\sqrt{2n(1-\alpha)}}\} \leq (1-\alpha) \tag{31}$$

Then

$$e = \frac{\sigma}{\sqrt{2n(1-\alpha)}} \tag{32}$$

So we have

$$n = \frac{\sigma^2}{2(1-\alpha)e^2} \tag{33}$$

We know that the Chernoff inequality is much tighter than the Chebychev inequality and the Chebychev inequality is distribution independent. From the above equations (33) and (26), we obtain the relationship between the small gallery size and the confidence interval given the margin of error.

In the above we give the small gallery size from mathematics. Meanwhile in our approach we incorporate the process of learning for similarity score distributions to find the optimal size of the small gallery. Also we can provide the upper bound and the lower bound for the performance prediction on the large population.

## 4. Experimental Results

In all the experiments we use fingerprints from the *NIST Special Database 4* (NIST-4). It consists of 2000 pairs of fingerprints. Each of them is labeled with an ID number preceded by an 'f' or an 's' which represents different impressions of the same fingerprint. The images are collected by scanning inked fingerprints from paper. The resolution of the fingerprint image is 500 DPI and the size of the image is $480 \times 512$ pixels.



Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05)
1550-5499/05 $20.00 © 2005 **IEEE**

## 4.1. Distorted Data

Usually the minutiae features are used for the fingerprint recognition which can be expressed as $f = (x, y, c, d)$, where $x$ and $y$ are the locations of the minutiae, $c$ is the class of the minutiae, and $d$ is the direction of the minutiae. We define the percentage of the minutiae with distortion for one fingerprint as $g$. Here we choose $g = 5\%$. Assume the number of minutiae is $num_j$. Usually one pair of fingerprints has a different number of minutiae so $j = 1, 2, \cdots, 4000$. We apply the distortion model [7] to these 2000 pairs of fingerprints as follows:

(a) Uncertainty: Uniformly choose $U = 5\% \times num_j$ minutiae features out of $num_j$ features and replace each $f_i = (x, y, c, d)$ with $f_i^{'}$ chosen uniformly at random from the set

$$\{(x', y', c', d'), (x', y') \in 4NEIGHBOR(x, y), \\ c' = c \pm 1, d' = d \pm 3^\circ\}$$

where $i = 1, 2, \cdots, U$.

(b) Occlusion: Uniformly choose $O = 5\% \times num_j$ minutiae features out of $num_j$ features and remove these minutiae.

(c) Clutter: Add $C = 5\% \times num_j$ additional minutiae, where each minutiae is generated by picking a feature uniformly at random from the clutter region. Here we choose the clutter region as

$$CR = \{(x, y, c, d), 50 \le x \le 450, 60 \le y \le 480, \\ c = \{0, 1, 2, 3, 4\}, 10^\circ \le d \le 350^\circ\}$$

In our experiments we use the uniform distribution as the uncertainty $PDF$ and the clutter $PDF$. The number of features with uncertainty, occlusion, and clutter is the same. We use the algorithm provided in [8] to extract minutiae and algorithm [9] for matching.

## 4.2. Estimate Distributions

In our experiments the EM algorithm is used to estimate the match score distribution and the non-match score distribution. The EM algorithm can find the number of components automatically [10] and for each component the EM algorithm can get its mean, covariance, and weight. In the learning process we feed back match scores and non-match scores to the small gallery. Then according to these different similarity scores the EM algorithm gives us a different estimation of distributions. Table 1 shows the estimation of the match score distribution with different number of the small gallery size. The distributions are represented by the Gaussian mixture model. For each component we have its mean, covariance, and weight. Figure 2 shows the match score distribution curves on different small gallery sizes.

## 4.3. Prediction Results

We randomly choose 50 pairs of fingerprints from fingerprint pairs of two levels of quality (high and low) as our small gallery following a hypergeometric distribution. We can get 50 match scores and 2450 non-match scores. After we obtain these similarity scores we use the EM algorithm to estimate the match score distribution and the non-match score distribution. Then we choose the subset size

Table 1: Match score distribution estimated by the EM algorithm.

| Size | Component # | Mean | Covariance | Weight |
|---|---|---|---|---|
| 100 | 2 | 17.152658 | 334.452802 | 0.535764 |
| | | 299.015489 | 55459.580193 | 0.450296 |
| 200 | 5 | 57.348298 | 1026.825771 | 0.160830 |
| | | 3.615611 | 20.189071 | 0.362406 |
| | | 585.278037 | 66686.529667 | 0.151087 |
| | | 206.327514 | 7334.980411 | 0.191394 |
| | | 27.106400 | 131.423073 | 0.133465 |
| 300 | 4 | 3.581775 | 21.569950 | 0.395165 |
| | | 420.142835 | 64933.952657 | 0.236481 |
| | | 35.420091 | 423.267100 | 0.228275 |
| | | 143.774430 | 3016.000039 | 0.139634 |

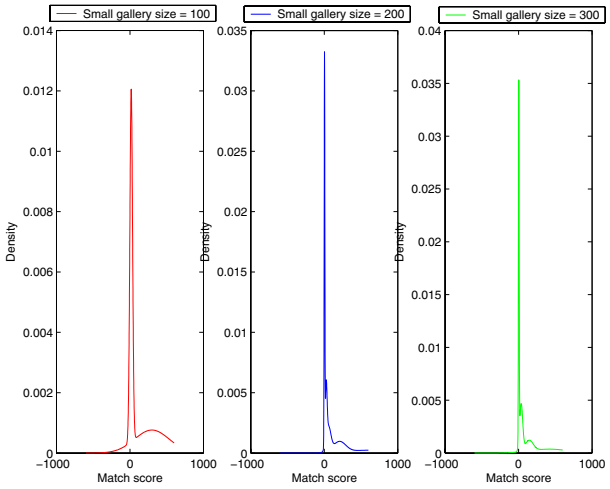


Figure 2: Match score distributions for different small gallery sizes.

$M_1 = 100$. We use 50 fingerprint pairs to predict the recognition performance for this subset. Here we set the margin of error $e = 0.06$. The prediction result is showed in Figure 3. From this curve we see that for the large population size 100 the error between the prediction performance and the actual performance is 0.137 which is larger than the margin of error.

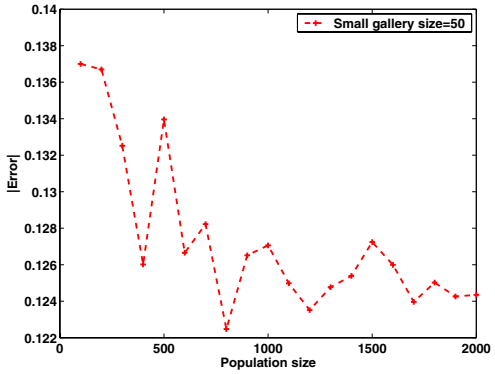


Figure 3: Absolute error between the prediction and the actual performance when the small gallery size is 50.

Now we apply learning to the prediction process. Figure 4 shows the error between the prediction and the actual performance decreases when the gallery size increases. When the small gallery size $n = 300$ the absolute error is smaller than the margin of error. At this point we can stop learning. We randomly feed back match scores from se-



Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05)
1550-5499/05 $20.00 © 2005 IEEE

lected fingerprint pairs and repeat this process seven times. Then we pick the maximum and the minimum prediction performance as our upper bound and lower bound for the prediction on the large population. Figure 5 gives the upper bound and the lower bound on the large population performance prediction. Since we have 2000 pairs of fingerprints, our actual recognition performance is shown in Figure 5. Beyond this population size we can give the bounds for the prediction. From Figure 5 it can be seen that the actual performance is within the upper bound and the lower bound except when the population size is very small. Our experiments show that when the small gallery size $n = 300$ then the prediction error is less than $5\%$.

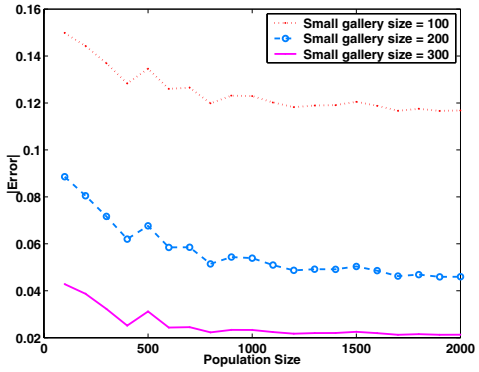


Figure 4: Absolute error between the prediction and the actual performance for different small gallery sizes.

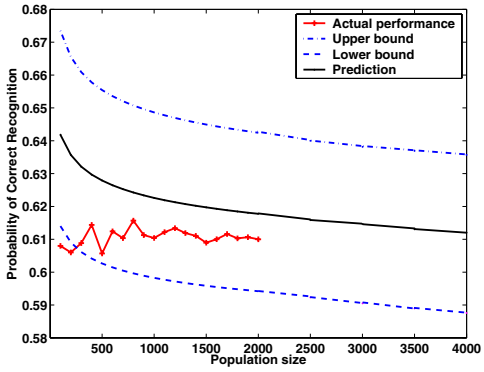


Figure 5: The upper bound and the lower bound on the large population when the small gallery size is 300. Note that the upper bound and the lower bound are within $5\%$.

Table 2: Values of the confidence interval, the margin of error, and the small gallery size

| 1-$\alpha$ | -ln$(1-\alpha)$ | e | n |
|---|---|---|---|
| 0.05 | 2.996 | 0.06 | 417 |
| 0.05 | 2.996 | 0.04 | 937 |
| 0.1 | 2.303 | 0.06 | 320 |
| 0.1 | 2.303 | 0.04 | 720 |
| 0.15 | 1.8971 | 0.06 | 264 |
| 0.15 | 1.8971 | 0.04 | 593 |

Since the Chernoff inequality is much tighter than the Chebychev inequality, we compare our learning small gallery size with the Chernoff inequality. Table 2 shows that the different confidence interval and the margin of error gives different small gallery size. From the table we ascertain that when the confidence interval $\alpha = 95\%$, margin of error $e = 0.06$ then the small gallery size $n = 416$. From our experiment for the same margin of error our small gallery size is 300. And the confidence interval is $\alpha = 97.5\%$. Statistical methods are independent of data which can give us a loose estimation of the small gallery size. Based on our own recognition system we can find the more accurate small gallery size by learning.

# 5. Summary and Conclusions

We focus on the fundamental problem of performance prediction for object recognition: what is the optimal size of the small gallery that can give good error estimation and how confident is the estimation. We use a generalized prediction model that integrates a hypergeometric probability distribution model with a binomial model explicitly and takes into account distortion in large populations. We incorporate learning in the prediction process to find the optimal small gallery size and provide the upper bound and the lower bound for the performance prediction on large populations. Meanwhile the Chernoff inequality and the Chebychev inequality are used as a guide to obtain the small gallery size and the confidence interval given a margin of error. Experimental results show that the small gallery size from the statistical methods are more loose than the learning method. Using a sufficient small gallery size we improve the prediction performance on the large population.

# References


[1] I. Guyon, and J. Makhoul, "What size test set gives good error rate estimates?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 20, NO. 1, pp.52-64, 1998.

[2] T. M. Mitchell, "Machine Learning," *McGraw,Hill*, 1997.

[3] P. Grother, and P. J. Phillips, "Models of large population recognition performance," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Volume 2, pp.68-75, 2004.

[4] A. Y. Johnson, J. Sun and A. F. Boick, "Using similarity scores from a small gallery to estimate recognition performance for large galleries," *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 100-103, 2003.

[5] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern Classification," *A Wiley-Interscience Publication*, 2000.

[6] A. M. Mood, F. A. Graybill, and D. C. Boes, "Introduction to the Theory of Statistics," *McGraw,Hill*, 1974.

[7] R. Wang, B. Bhanu, and H. Chen, "An Integrated Prediction Model for Biometrics," *Accepted by Audio- and Video-based Biometric Person Authentication (AVBPA) 2005*.

[8] B. Bhanu, M. Boshra,and X. Tan, "Logical templates for feature extraction in fingerprint images," *International Conference on Pattern Recognition (ICPR)* vol.3,pp.850-854,2000

[9] X. Tan, and B. Bhanu, "Robust fingerprint identification," *Proc. IEEE International Conference on Image Processing (ICIP)*, vol. 1, pp. 277-280, 2002.

[10] M. A. T. Figueiredo, and A. K. Jain, "Supervised learning of finite mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 24, NO. 3, pp.381-396, 2002.




Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05)
1550-5499/05 $20.00 © 2005 **IEEE**