Prediction and Validation of Indexing Performance for Biometrics

Suresh Kumar R.*[†] Bir Bhanu[†] Subir Ghosh[‡] Ninad Thakoor[†]

Abstract

The performance of a recognition system is usually experimentally determined. Therefore, one cannot predict the performance of a recognition system a priori for a new dataset. In this paper, a statistical model to predict the value of k in the rank-k identification rate for a given biometric system is presented. Thus, one needs to search only the topmost k match scores to locate the true match object. A geometrical probability distribution is used to model the number of non match scores present in the set of similarity scores. The model is tested in simulation and by using a public dataset. The model is also indirectly validated against the previously published results. The actual results obtained using publicly available database are very close to the predicted results which validates the proposed model.

keywords: Object identification, Performance prediction, Rank-k identification rate, geometric distribution model.

1. Introduction

Two important operational tasks of a biometric system are authentication and identification. Several indicators have been proposed for measuring the performance of the above tasks which include False Accept Rate, False Reject Rate, Receiver Operating Characteristic, Rank-k identification/recognition rate (R_k) and Cumulative Match Characteristic (CMC) Curve. Out of these, the last two measures, namely, Rank-k identification rate and Cumulative Match Characteristic Curve have been specifically proposed for biometric identification systems. Rank-k identification rate provides the number of times the correct object is present in the top k most likely candidates. CMC Curve provides a plot of the rank-k identification rate against k [6]. These results are usually evaluated empirically for a given dataset. A fundamental problem in the empirical evaluation is that the results for a new dataset cannot be predicted [4]. What is fundamentally missing here is a theoretical approach that would predict the identification performance for any biometric system. This paper focuses on developing such an approach for predicting the performance of a biometric system. The results of our approach are validated on a publicly available dataset.

Given a collection of objects (gallery [7]) this paper predicts the value of k for a specific R_k so that the matching object could be found among the top k matches provided by the recognition system. It should be noted that k is an index position in the sorted list of similarity scores provided by the system. For a given gallery and the corresponding matching algorithm, this is achieved by collecting a set of similarity scores for various probe objects. The similarity scores are then processed and modeled using a geometric distribution. Once the parameter of the geometric distribution has been estimated, the bounds of the values of k are predicted. In addition, this paper examines the change in the value of k as the gallery size is increased. It has been found that the value of k expressed as a fraction of the size of the gallery remains more or less a constant (with a slight decrease) when the gallery size is increased to a very large value. The important terms used in this paper are defined in Table 1.

2. Related Work and Contributions

2.1. Related Work

To the best of our knowledge the problem of predicting bounds of the match score indices in a biometric identification system has not been addressed earlier and, therefore, a direct comparison with published research is not possible. The following describe the related work on indexing as well as on prediction. While comparing our work with existing indexing related work, it should be noted that while the end results of an indexing algorithm and our work are the same, the approaches are quite different. Our work is independent of a specific biometric and uses match scores produced by a matching algorithm on a given dataset and predicts the value of k for a given R_k, whereas a typical indexing algorithm is tailor made to a specific biometric and does not predict the results on a different dataset.

Bhanu *et al*. [1] describe indexing using minutiae triplets and its performance is reported for NIST-4 fingerprint database [12]. Cappelli *et al*. [3] describe a new hash based indexing method to speed up fingerprint identification in large databases and results are reported on several databases including NIST-4. Daugman [5] presents binomial models that used only the non-match scores to estimate the probability that a false match never occurs [8]. Grother and

---

*Email: suresh.kumar@email.ucr.edu

[†]Center for Research in Intelligent Systems, UC, Riverside

[‡]Dept of Statistics, University of California, Riverside

Table 1: Definitions

Gallery	A collection of objects.
Probe	A query object whose matching object *is present* among the gallery objects.
Similarity score	A real number representing the similarity between two objects. A similarity score may be either a match score or a non-match score. Depending on the way the similarity score is computed, the smaller the score the more similar the objects or vice versa. In this paper a low value of score is preferred.
Match score	The similarity score obtained when two matching (similar) objects are compared.
Non-match score	The similarity score obtained when two dissimilar objects are compared.
Indexing	A method of re-organizing a set of data such that object retrieval from a gallery becomes easier.
Trial	Any experiment generating n similarity scores.
Index, Rank	In a sorted set of similarity scores, the position of the given similarity score.

Phillips [7] present the prediction of recognition performance of large sized biometric galleries using a binomial model under the assumption that the match score distribution and the non match score distribution are independent. This approach has been followed by others [11] for recognition with an additional assumption that the match and non match score distributions remain the same when the gallery size is increased. In [2], Boshra *et al*. present a theory for predicting object recognition which is verified on synthetic aperture radar data.

2.2. Contributions

The contributions of this paper are as follows:

1. We developed a foundational approach for predicting identification performance, applicable to a variety of biometrics. The approach is rooted in statistical performance characterization. Such a theoretical approach does not exist in the biometric or object recognition field in general.
2. Given the match scores of a biometric, using the geometric distribution, we can predict the indexing performance for a given Rank-k identification rate.
3. The approach presented is independent of a specific biometric and can adapt to any matching algorithm.
4. We carry out both direct and indirect validation of our results in two publicly available databases.

While the prediction using a known (“nice”) distribution, (geometric distribution in this paper) is not new, this is the first paper to theoretically model the performance for Rank-k identification and fills the void in the biometric, computer vision and patter recognition field.

3. Technical Approach

An overview of the technical approach is shown in Figure 1. The input to our system is a collection of similarity scores produced by different probe objects which are noisy versions of the objects present in the gallery. Here we assume a closed-set recognition system [7]. The similarity scores are processed by our system and a geometric distribution model is built. The parameter p of the geometric distribution model when estimated from the given data, enables us to predict the value of k for a any given recognition rate R_k. The details of the approach are given below.

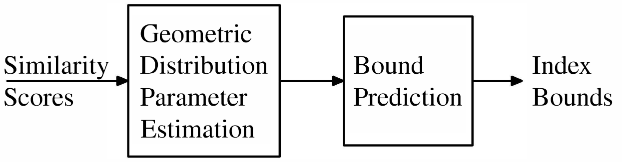

Figure 1: Overview of the technical approach

Let N be the size of the gallery. When a single probe is presented to the system, it is matched against all the N objects of the gallery and, thus, N similarity scores are produced. This constitute one trial. Let S represent the set of similarity scores. S contains one match score and $N-1$ non-match scores. Let S be sorted in the ascending order of the score values. Here we assume that the lower the similarity score, the better the match is. In a different recognition system, the opposite may be true and our approach can be applied by normalizing the similarity scores. Let X be a random variable representing the count of the non match scores present in S before the first match score. If the recognition system is ideal, the value of X would be zero.

We want to model X in order to predict its value for a given R_k. This is the same as predicting the value of k for a specific R_k. In order to build a statistical model, we examine the properties of X, which are:

1. X represents a count and hence $X \in \mathbb{Z}$, the set of integers.
2. $X \in [0, N-1]$.
3. Many a time X may have value zero. This is because, often, for a well designed recognition system, the correct match score is the first one in the set of sorted similarity scores.

Considering the above properties one can see that X can be modeled using geometric distribution. It models the number of trials needed before the first success in repeated Bernoulli trials [10]. In our case, each element in the set S can be as-

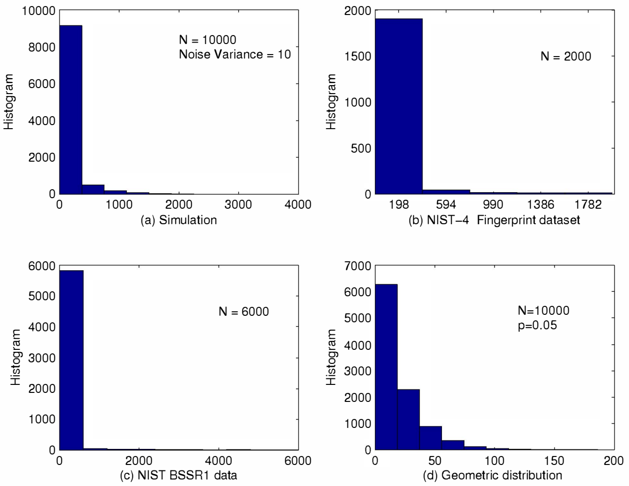

Figure 2: Histogram of the number of non match scores present before the match score (random variable X) obtained from (a) Simulation with gallery size = 10000, noise variance = 10 (b) NIST-fingerprint data [12] (c) NIST Biometric score set, Right index finger [9](d) geometric distribution with $p = 0.05, N = 10000$. Note the similarity of the histogram of data in (a), (b) and (c) and that of a geometric random variable in (d). Note that these figures provide visual justification for the selection of geometric distribution to model the random variable. The datasets in (a), (b), (c) and (d) are not related to each other.

sumed to be coming from a Bernoulli trial (a given similarity score may be a match score or a non match score) and if the score is a match score, we denote it as success and if the score is a non match score, we denote it as failure. Thus, geometric distribution can model the number of non match scores present before the first match score. Besides, our requirement of high probability for very small values of X is also satisfied by the geometric distribution.

The random variable X is a waiting-time random variable (we wait till the first match score is obtained) and the geometric distribution is a natural candidate model to describe the waiting-time random variable.

The suitability of geometric distribution can also be visually verified by observing the histograms of X obtained from experimental trials and the histogram of a geometrically distributed random variable as shown in Figure 2. Thus one may note that the candidacy of the geometric model is neither based on prior knowledge nor empirically obtained from the data. It is a first order formulation with promising results.

The reason for modeling the values of X as coming from a single geometric distribution, though the objects which produce the similarity scores are the instances of different gallery objects, is the fact that, the similarity score values are independent of the gallery/probe object names and the score never bears any information concerning the objects from which it is calculated. It only represents the closeness of the objects. To give an example, if we have an ideal matching algorithm, the 'distance' between an object and its noised version would be zero, irrespective of the object.

3.1. Estimation of geometric distribution parameter

Let X be a geometrically distributed random variable. Then

$$P(X = x) = p(1-p)^x \quad (1)$$
$$\text{where } x = 0, 1, \ldots$$

Let $x_1, x_2, \ldots, x_T$ be the realizations of T independent and identically distributed geometrical trials described by the model in eq. (1). The value of p can be estimated using maximum likelihood approach and it can be shown that the estimate is unbiased.

3.1.1 Estimating the value of k for $R_k = 95\%$

The CDF of the geometric distribution is given by

$$F(x) = 1 - (1-p)^{x+1} \quad (2)$$

We need to find out the range of values of x which occur in 95% of the experimental trials. As X is modeled as a geometric random variable, lower bound of x is obviously zero. The upper bound is given by the equation

$$F(x) = 1 - (1-p)^{x+1} = 0.95 \quad (3)$$

If there are x non match scores before the first match score, then

$$k \text{ (as defined in } R_k\text{)} = x + 1 \quad (4)$$

From eqs. (3) and (4), the predicted bounds for $R_k = 95\%$ can be obtained.

4. Experimental Results

The following sections describe the experimental results obtained from simulation and by using NIST 4 [12] fingerprint dataset. The unknown parameter p in eq. (1) is estimated using half the available data. The model is tested using the remaining data.

4.1. Simulation

4.1.1 Technical details of simulation

The simulation was carried out on a gallery of variable number of objects, starting from 100 to 50,000. Each object was represented by 10 point features . The feature points were obtained from a uniformly distributed random variable in the interval $[0, 100]$. The probe objects were obtained by shifting the feature point locations of the gallery objects by adding a Gaussian noise of zero mean and different variances. This helps to evaluate the performance at different noise levels. The idea is not to identify all the noise levels that could be present in a biometric system because some of them are known and others are unknown but to represent

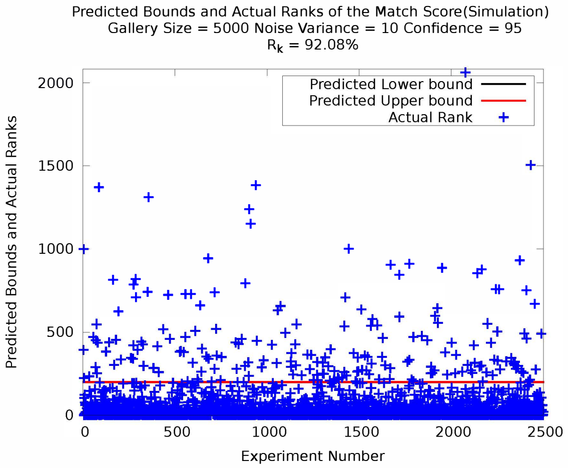

Figure 3: Simulation results on a gallery of size 5000. The model predicted that the top 201 match scores would contain the exact match 95% of the time. The actual success was 92%.

them in terms of Gaussian noise variance. The similarity score between two objects were obtained by finding the sum of the Euclidean distance between the closest feature points of the two objects.

4.1.2 Results on a gallery of size 5000

Our aim is to predict the upper bound of k (The lower bound of k is always one) for $R_k = 95\%$ so that by searching k top matches we could find the actual match object in 95% of the cases. 5000 probe objects were generated by adding Gaussian noise $N(0, 10)$ to each of the 5000 gallery objects. Out of these, 2500 probe objects were used to build the model. The model predicted the value of k to be 201 ($\sim$ 4% of the gallery size). The prediction was validated by testing using the remaining 2500 probe objects. It was found that the top 201 matches contained the actual match object in 92% cases. This result is shown in Figure 3. It should be noted that the probe objects used for training and testing the model were derived from different gallery objects, giving more credibility to the results.

4.1.3 Ten fold validation of results

To understand how stable the above results are, we performed a 10-fold cross-validation on the same gallery of size 5000, each time trying to predict k for $R_k = 95\%$. The mean and variance of the predicted value of k for $R_k = 95\%$ and the actual value of R_k were calculated. The mean value of k was found to be equal to 194 with almost zero variance. The average success of finding the exact matching object on searching the top 194 objects was 91.11% with a variance of 0.58. The low variance for the value k and for the accuracy of prediction shows the stability of our prediction model.

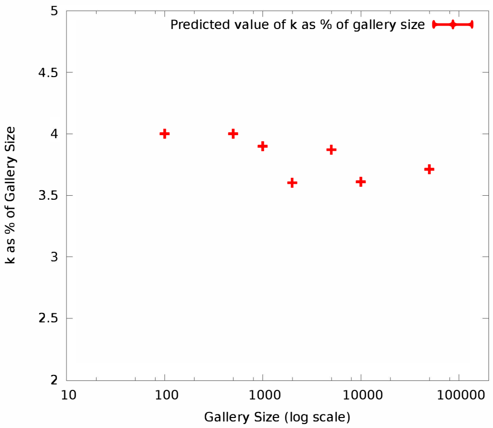

Figure 4: Variance plot of the predicted value of k expressed as fraction of the gallery size. The variance changes from 10^{-35} to 10^{-6}. The gallery sizes are 100, 500, 1000, 2000, 5000, 10000, and 50,000. k value predicted for $R_k = 95\%$. The validation of the prediction is shown in fig. 5.

4.1.4 Results on different sizes of gallery

In order to understand how our prediction model performs for different gallery sizes, the procedures mentioned in sections 4.1.2 and 4.1.3 were repeated for gallery sizes 100, 500, 1000, 2000, 10000, and 50,000. The same number of trials were used for building the model and for validating it. The number of trials used for training and testing were 5000, 25000, 50000, 1000, 2500, 5000, and 25000 respectively. The mean value of k predicted for each of the different gallery sizes is shown in Figure 4. The vertical bar at each data point represents the variance obtained from the 10-fold cross validation which varied from 10^{-35} to 10^{-6}. It can be observed from the figure that we need to search approximately the top 4% of the gallery size similarity scores to find the match irrespective of the gallery size. Note the very low value of variance of the predicted k values. The measured value of R_k with variance obtained in 10-fold cross validation experimentation is shown in Figure 5.

4.2. Effect of noise on the prediction

In order to study the effect of noise on our prediction model, we carried our prediction on a gallery of size 5000. We choose noise variance to be 1, 3, 5, 7, and, 10 and predicted the value of k for $R_k = 95\%$. For each noise variance value, a 10-fold cross-validation was also carried out. The variance of predicted value of k with noise is shown in Figure 6. Even with large value of noise, the variance of the predicted value of k is extremely small and therefore the vertical bars representing the variance of the k values appear as points. As expected the mean value of k increases as noise increases. For low noise variance, $k = 1$, implying that the first score itself is the match score. For these values of k, 100% success is obtained. Thus, for low noise

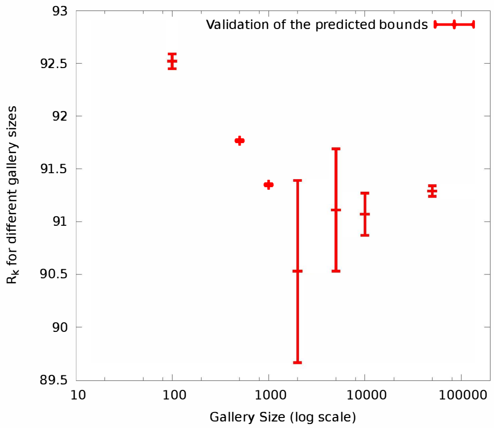

Figure 5: Validation of predicted bounds for $R_k = 95\%$. R_k values obtained for gallery sizes 100, 500, 1000, 2000, 5000, 10000, and 50,000 in simulation. The variance values are 0.07, 0.01, 0.01, 0.86, 0.58, 0.2, 0.05 respectively. These results were obtained with a 10 fold validation. By performing a larger fold validation in relation to the gallery size, a smoother variance plot can be obtained. The predicted bounds are shown in fig. 4.

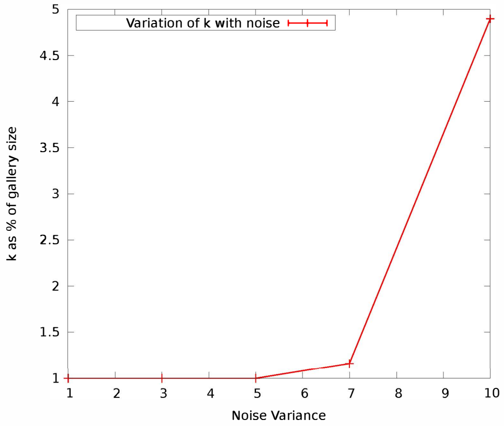

Figure 6: Effect of noise on the predicted value of k, for $R_k = 95\%$. The variance of k for different noise values in a 10 fold experimentation was so small that the vertical bars appear like points.

variance, our prediction over achieves the results. The actual value of R_k obtained under different noise variances is shown in Figure 7. Note the increasing bar size as the noise variance is increased. Even with a noise variance of 10, the achieved value of R_k is around 91.5% while the prediction was made for 95% R_k. Thus we see that the performance of our model does not deteriorate significantly in presence of high noise.

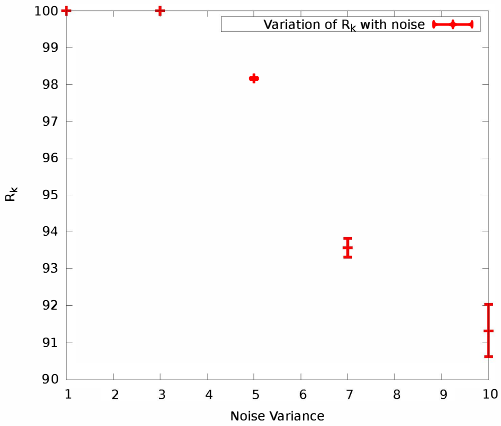

Figure 7: Effect of noise in the achieved value of R_k. k was predicted for $R_k = 95\%$. The variance of k was in the range 0 to 7×10^{-5}

4.3. Results on NIST-4 Fingerprint Dataset

The NIST-4 fingerprint database [12] consists of 2000 pairs of fingerprints. Using a fingerprint matching algorithm [1], 4 million similarity scores were generated resulting in 2000 sets of trial data (one probe object would produce one trial data containing 2000 similarity scores). This dataset was equally divided into two halves randomly and one half was used for building the model and the other half was used for testing the model. Our system predicted that we need to examine the top 215 match scores for getting the true match object 95% of the time. We found that we actually get 93.8% success by examining the top 215 scores. By conducting a 10-fold cross-validation experiment, we found that our model predicted the mean value of k to be 194 (9.7% of gallery) with almost zero variance for $R_k = 95\%$. By searching the top 194 match scores, the actual match object was found in 92.44% trials with variance 0.16.

In order to evaluate the performance of our prediction model on different gallery sizes of real data, we built smaller galleries of size 100, 200, 500, and 1000 from the original NIST-4 2000 gallery set and in each case carrying out 1000 trials. 10-fold cross-validation experimentation was done for each of these sizes. The predicted k value for $R_k = 95\%$ with its variance is shown in Figure 8. The R_k values achieved for various gallery sizes is shown in fig. 9.

4.4. Indirect validation of our model

As mentioned earlier in section 2.1, Bhanu *et al*. [1] and Cappelli *et al*. [3] have reported results on indexing performance on the NIST-4 dataset. Bhanu *et al*. report 83% success on examining the top 10% of the gallery while Cappelli et al. report 96% success. The success achieved using our prediction model is 92.44%. While Cappelli *et al*. report a higher success, they have also indicated that a 50-pixel border has been removed from the images before extracting the minutiae. It is also not clear whether the entire dataset was used for their experiment. We have reported results on the entire database without doing any preprocessing and our results are in the same ballpark. Besides, note that ours is a generic approach which is independent of a specific biometric and requires only the similarity scores produced by any matching algorithm whereas the algorithms used by others are specifically meant for fingerprint indexing. Thus the re-

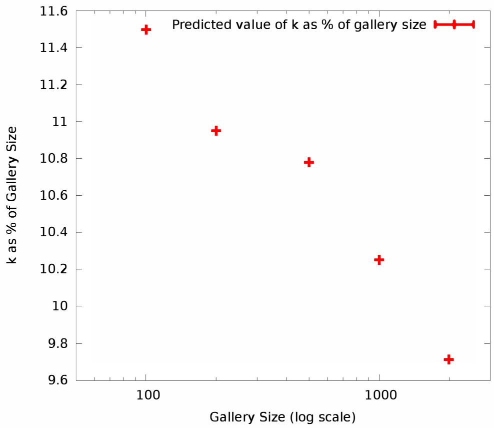

Figure 8: Variance plots of the predicted values of k for NIST4 Fingerprint data, for $R_k = 95\%$. The variance was in the range 2.5×10^{-5} to 10^{-4}. Gallery sizes 100, 200, 500, 1000, 2000. The validation of the predicted bounds is shown in fig. 9.

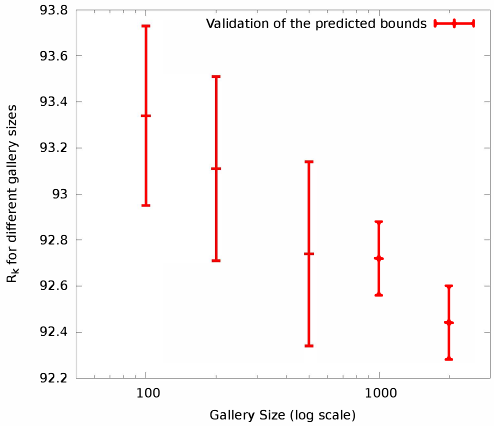

Figure 9: Validation of the predicted bounds for gallery sizes 100, 200, 500, 1000, 2000 on NIST4 fingerprint data. The bounds were predicted for $R_k = 95\%$. The predicted bounds are shown in fig. 8.

sults reported by Bhanu *et al*. [1] and Cappelli *et al*. [3] provide an indirect validation of our model.

5. Conclusions

We developed a model for predicting the value of k for a given rank-k identification rate. Our method is independent of specific biometrics used and can adapt to any matching algorithm. It just requires the match scores produced by different probe objects. The model was tested using simulation and found that even when the noise is increased, the model performs gracefully. The model was also tested on a public dataset available from NIST, of size 2000. We predicted the value of k for 95% rank-k identification rate. It was experimentally found that the predicted k value provided success $\sim 95\%$. The slight difference between the prediction and actual performance is due to the deviation of the data from the pure geometric distribution, which is a first order model. By varying the gallery size from a small value to a very large value, we found that the fraction of the gallery that should be examined when a new probe object is presented remains a constant or decreases slightly. The value of k decreasing as the gallery size is increased is a very good indication. It implies that the fraction of the gallery that needs to be examined for locating a matching object would not increase proportionately if the gallery size is increased. However in simulation, we found that the k values remains more or less a constant. The performance of our model was also validated indirectly with the results reported in the literature.

Applicable to any biometrics, and adaptable to any matching algorithm, our method provides a fundamental approach for predicting indexing performance.

6. Acknowledgements

This work was supported by the NSF grant 0915270. The contents and information do not reflect positions or policies of the US government.

References

[1] B. Bhanu and X. Tan. Fingerprint indexing based on novel features of minutiae triplets. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25:616–622, 2003.

[2] M. Boshra and B. Bhanu. Predicting performance of object recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22:956–969, 2000.

[3] R. Cappelli, M. Ferrara, and D. Maltoni. Fingerprint indexing based on minutia cylinder-code. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(5):1051 –1057, May 2011.

[4] G. Chollet, B. Dorizzi, and D. Petrovska-Delacrétaz. *Guide to Biometric Reference Systems and Performance Evaluation*, chapter Introduction–About the Need of an Evaluation Framework in Biometrics. Springer Verlag, Feb. 2009.

[5] J. Daugman. The importance of being random: statistical principles of iris recognition. *Pattern Recognition*, 36(2):279–291, 2003.

[6] D. Gorodnichy. Evolution and evaluation of biometric systems. In *IEEE Symposium on Computational Intelligence for Security and Defense Applications*, pages 1 –8, July 2009.

[7] P. Grother and P. J. Phillips. Models of large population recognition performance. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2:68–75, 2004.

[8] A. Y. Johnson, J. Sun, and A. F. Bobick. Using similarity scores from a small gallery to estimate recognition performance for larger galleries. *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, 0:100, 2003.

[9] NIST. Nist biometric score set. `http://www.itl.nist.gov/iad/894.03/biometricscores/`.

[10] A. Papoulis and S. U. Pillai. *Probability, Random Variables and Stochastic Processes*. Mc Graw Hill, 2002.

[11] R. Wang and B. Bhanu. Predicting fingerprint biometrics performance from a small gallery. *Pattern Recogn. Lett.*, 28:40–48, Jan. 2007.

[12] C. Watson and C. Watson. NIST special database 4: fingerprint database. *National Institute of standards and technology*, Mar. 1992.