Performance Prediction for Multimodal Biometrics

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

Abstract

Sensor fusion is commonly used to improve the detection and recognition performance of a pattern recognition system. In this paper we propose a prediction model to predict the performance of a sensor fusion system. In particular, we answer two questions associated with the performance prediction in a sensor fusion system: (a) Given the characteristics of the individual sensors how can we predict the performance of the fusion system? (b) How good the prediction is? We provide the Cramer-Rao bounds for the prediction model. We carry out experiments on the publicly available database XM2VTS that has speech and face data.

1. Introduction

Sensors are used to obtain the signatures of biometrics such as fingerprint, palm, face, gait, signature, speech, etc. In order to improve the performance of a biometric recognition system sensor fusion from different modalities are commonly used to exploit the complementary information. By fusing different sensors, we may achieve the following benefits: availability of more meaningful information; reduction of false alarm; broadening the range of scenarios that the system will function correctly.

Usually sensors fuse at the data level, feature level, match score level, and decision level. Fusion at the match score level is widely used because match scores can be normalized and combined by different rules. Feature level is believed to be very promising since feature sets can provide more information about the input biometrics than other levels. But different feature sets are sometimes in conflict and may not be available which make the feature level fusion more challenging than other levels of fusion. Kittler et al. [5] develop a common theoretical framework for sensor fusion. They prove that the recognition problem can be formulated as the Bayesian rule which can be represented as the product rule and the sum rule under certain assumptions. Also the other rules such as min rule, max rule, median rule, and majority vote rule, which are used at the match score level can be developed from the Bayesian rule.

It's not always true that more sensors are better than less number of sensors. If two biometric systems have significantly different performance and each system has its own cross-over point [2] where the false reject equals to the false accept, then the fusion performance is worse than the single stronger sensor [1]. Then a question arises about how can we predict the sensor fusion performance? Or how can we find the optimal fusion combination to meet the biometric system requirement? Daugman [1] discusses the sensor fusion at the decision level. He gives the probability of a false accept under the disjunction rule and the probability of a false reject under the conjunction rule. He concludes that a strong biometric is better to be left alone than in combination with a weaker one. Poh et al. [7] propose a measurement F-ratio which is related to the *equal error rate* (EER) to find the optimal fusion candidate. They verify their approach on the BANCA multi-modal biometric database. Keller et al. [4] analyze the sensor fusion by the fuzzy set theory at the decision level. They use the d-metric which is the ratio of the probability of detection and the probability of false alarm to predict the sensor fusion performance.

In this paper we present a novel prediction model which is based on the theory that the likelihood ratio is proportional to the probability of correct recognition which is equal to the area under the receiver operating characteristic (ROC) curve. We use the Fisher measurement as our discriminability measurement to predict the sensor fusion system performance. Unlike the traditional application of this theory in the classification field, we apply it in the performance prediction of the sensor fusion system to find the optimal sensor fusion combination. We also give the Cramer-Rao bounds for the estimation. The paper is organized as follows. The details of the prediction model and the Cramer-Rao bounds are given in Section 2. Experimental results which verify our prediction model on the publicly available multi-modal database XM2VTS are shown in Section 3. Conclusions are given in Section 4.

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

2. Technical Approach

2.1. Discriminability measurement

There are two classes in a recognition system : match and non-match. We denote match as s and non-match as n. Assume that x is the similarity score which defines the degree of the similarity for two biometrics. Then $f(x|n)$ is the probability density function given n is true, and $f(x|s)$ is the probability density function given s is true. We assume that $f(x|n)$ and $f(x|s)$ are Gaussian. Then we have

$$f(x|n) = \frac{1}{\sqrt{2\pi\delta_n^2}} exp\left[-\frac{(x-m_n)^2}{2\delta_n^2}\right] \quad (1)$$

and

$$f(x|s) = \frac{1}{\sqrt{2\pi\delta_s^2}} exp\left[-\frac{(x-m_s)^2}{2\delta_s^2}\right] \quad (2)$$

where m_n and δ_n are the mean and standard deviation when the hypothesis n is true, and m_s and δ_s are the mean and standard deviation when the hypothesis s is true.

We know that the likelihood ratio at one point is equal to the slope of the receiver operating characteristic (ROC) curve at this point. The area under the ROC curve is the probability of the correct match. So the likelihood ratio is proportional to the probability of the correct match [3]. Since the different decision criteria can get different likelihood ratio, here we use the Fisher measurement d as our discriminability measurement,

$$d = \frac{(m_s - m_n)^2}{\delta_s^2 + \delta_n^2} \quad (3)$$

2.2. Sensor fusion prediction model

Assume that we have q independent sensors, x_1, x_2,..., x_q. Each sensor represents an independent recognition system. If we fuse these q sensors then the performance of the fusion system can be expressed as the probability that the likelihood ratio of this fusion system be greater than a threshold k. This expression can be written as

$$PCM = P[l(x_1, x_2, ..., x_q) > k] \quad (4)$$

where PCM is the probability of correct match, and $l(x_1, x_2, ..., x_q)$ is the likelihood ratio of the fusion system. Since these sensors are independent then we have

$$\begin{aligned} PCM &= P[l(x_1, x_2, ..., x_q) > k] \\ &= P[l(x_1)l(x_2)...l(x_q) > k] \\ &= P[\textstyle\sum_{i=1}^q lnl(x_i) > lnk] \end{aligned} \quad (5)$$

Now we consider the sensor x_i. By definition we have

$$\begin{aligned} lnl(x_i) &= ln\frac{f(x_i|s)}{f(x_i|n)} \\ &= lnf(x_i|s) - lnf(x_i|n) \end{aligned} \quad (6)$$

If hypothesis n is true then the mean and variance of $lnl(x_i)$ can be obtained from equation (6).

$$E(lnl(x_i)|n) = \int_{-\infty}^{\infty} lnl(x_i) f(x_i|n)dx \quad (7)$$

$$E((lnl(x_i))^2|n) = \int_{-\infty}^{\infty} (lnl(x_i))^2 f(x_i|n)dx \quad (8)$$

Then the variance for $lnl(x_i)$ given n is

$$var(lnl(x_i)|n) = E((lnl(x_i))^2|n) - E(lnl(x_i)|n)^2 \quad (9)$$

If hypothesis s is true, then we have

$$E(lnl(x_i)|s) = \int_{-\infty}^{\infty} lnl(x_i) f(x_i|s)dx \quad (10)$$

$$E((lnl(x_i))^2|s) = \int_{-\infty}^{\infty} (lnl(x_i))^2 f(x_i|s)dx \quad (11)$$

Then the variance for $lnl(x_i)$ given s is

$$var(lnl(x_i)|s) = E((lnl(x_i))^2|s) - E(lnl(x_i)|s)^2 \quad (12)$$

Since $lnl(x_i)$ are independent, then we can get the mean and variance of $\sum_{i=1}^q lnl(x_i)$ under different hypothesis. If hypothesis n is true then the mean and variance of $\sum_{i=1}^q lnl(x_i)$ are

$$\begin{aligned} m_{n(q)} &= E(\textstyle\sum_{i=1}^q lnl(x_i)|n) \\ &= \textstyle\sum_{i=1}^q E(lnl(x_i)|n) \end{aligned} \quad (13)$$

$$\begin{aligned} \delta_{n(q)}^2 &= var(\textstyle\sum_{i=1}^q lnl(x_i)|n) \\ &= \textstyle\sum_{i=1}^q var(lnl(x_i)|n) \end{aligned} \quad (14)$$

If hypothesis s is true then the mean and variance of $\sum_{i=1}^q lnl(x_i)$ are

$$\begin{aligned} m_{s(q)} &= E(\textstyle\sum_{i=1}^q lnl(x_i)|s) \\ &= \textstyle\sum_{i=1}^q E(lnl(x_i)|s) \end{aligned} \quad (15)$$

$$\begin{aligned} \delta_{s(q)}^2 &= var(\textstyle\sum_{i=1}^q lnl(x_i)|s) \\ &= \textstyle\sum_{i=1}^q var(lnl(x_i)|s) \end{aligned} \quad (16)$$

Then the Fisher measurement of the sensor fusion system is

$$d_{f(q)} = \frac{(m_{s(q)} - m_{n(q)})^2}{\delta_{s(q)}^2 + \delta_{n(q)}^2} \quad (17)$$

The lower subscript f means the Fisher measurement for the fusion system.

2.3. Cramer-Rao bounds for the estimator

For the sensor fusion system when the hypothesis n is true then the similarity score x is as follows:

$$p(x; m_{n(q)}, \delta_{n(q)}^2) = \frac{exp\left[-\sum_{i=1}^N \frac{(x_i - m_{n(q)})^2}{2\delta_{n(q)}^2}\right]}{(2\pi\delta_{n(q)}^2)^{\frac{N}{2}}} \quad (18)$$

where q is the number of sensors, N is number of observations. Then we have the log-likelihood function

$$\begin{aligned} lnp(x; m_{n(q)}, \delta_{n(q)}^2) &= -\frac{N}{2} ln(2\pi\delta_{n(q)}^2) \\ &\quad - \textstyle\sum_{i=1}^N \left[\frac{(x_i - m_{n(q)})^2}{2\delta_{n(q)}^2}\right] \end{aligned} \quad (19)$$

We denote the parameter vector $\theta = [m_{n(q)}, \delta_{n(q)}^2]^T$. The Fisher information matrix is

$$J = \begin{bmatrix} -E\left[\frac{\partial^2 lnp(x;\theta)}{\partial m^2_{n(q)}}\right] & -E\left[\frac{\partial^2 lnp(x;\theta)}{\partial m_{n(q)}\partial \delta^2_{n(q)}}\right] \\ -E\left[\frac{\partial^2 lnp(x;\theta)}{\partial \delta^2_{n(q)}\partial m_{n(q)}}\right] & -E\left[\frac{\partial^2 lnp(x;\theta)}{\partial (\delta^2_{n(q)})^2}\right] \end{bmatrix} \quad (20)$$

We compute the derivatives and get the Fisher information matrix

$$J = \begin{bmatrix} \frac{N}{\delta^2_{n(q)}} & 0 \\ 0 & \frac{N}{2\delta^4_{n(q)}} \end{bmatrix} \quad (21)$$

According to the Fisher information matrix we have

$$var(\hat{m_{n(q)}}) \geq \frac{\delta^2_{n(q)}}{N} \quad (22)$$

$$var(\hat{\delta^2_{n(q)}}) \geq \frac{2\delta^4_{n(q)}}{N} \quad (23)$$

By the similar process we can get the Cramer-Rao bounds for the prediction model when the hypothesis s is true.

$$var(\hat{m_{s(q)}}) \geq \frac{\delta^2_{s(q)}}{N} \quad (24)$$

$$var(\hat{\delta^2_{s(q)}}) \geq \frac{2\delta^4_{s(q)}}{N} \quad (25)$$

3. Experimental Results

The publicly available multi-modal database *extended multi modal verification for teleservices and security applications* (XM2VTS) contains face and speech data from 295 subjects which are divided into a set of 200 clients, 25 evaluation impostors, and 70 test imposters [6]. There are eight baseline systems which are denoted as (feature classifier). For each baseline system, there are two data sets: development set and evaluation set. Table 1 shows the baseline experiments, labels and the Fisher measurement on both of the data sets. The features of the face baseline experiments are FH (face image concatenated with its RGB histogram), DCTs (DCTmod2 features extracted from 40×32 pixel face image), and DCTb (DCTmod2 features extracted from 80×64 pixel face image). The speech baseline experiment features are LFCC (linear filter-bank cepstral coefficient), PAC (phase auto-correlation), and SSC (spectral subband centroid). The classifier we used in these experiments are the *multi-layer perceptron* (MLP) and *Gaussian mixture model* (GMM). The details of these features and classifiers are provided in [6].

3.1. Prediction of the sensor fusion system

Since we have eight baseline experiments we randomly combine two of them so we have $C^8_2 = 28$ combinations. Figure 1 shows these combinations. For example, the combination 1 is the baseline system $b2$ combines with the baseline system $b3$.

For each combination we apply equation (13), (14), (15), and (16) on the development set and evaluation set of the baseline experiments. Then we use these values in equation (17) and get the prediction d for the fusion system. For each combination we normalize the similarity scores in the development set and evaluation set. Here we apply the sum rule, product rule, min rule, median rule, and max rule to get the match and non-match scores for the fusion system. Then we compute the mean and variance of the similarity scores and apply equation (3) to get the actual d for each combination. We list the top 7 combinations for the prediction and actual experiments on the development set and the evaluation set in Table 2. From Table 2 we can see that for the development set the best combination which we get from the prediction model is 12 ($b1$ and $b4$). Meanwhile, according to the actual performance the best combination for the sum rule is 14 ($b1$ and $b6$), product rule is 9 ($b1$ and $b8$), min rule is 9 ($b1$ and $b8$), median rule is 14 ($b1$ and $b6$), and max rule is 12 ($b1$ and $b4$).

We use three criteria: rand statistic, Jaccard coefficient, and Fowlkes and Mallows index [8] to measure the degree of agreement between the prediction and the actual results. These criteria are between 0 and 1. The larger the values the greater is the agreement between them. There are two classes in the prediction and actual results. The first class is the top 7 combinations and the second class is the last 21 combinations. Since we have 28 combinations the total number of possible pairs is 378. Table 3 lists the criteria for the agreement between the prediction and actual results

Table 1. Baseline experiments and their labels.

Baseline experiment	Label	Fisher Measurement	
		Development	Evaluation
(FH MLP)	b1	31.401	25.106
(DCTs GMM)	b2	5.6474	6.1931
(DCTb GMM)	b3	6.9723	8.0993
(DCTs MLP)	b4	16.068	13.163
(DCTb MLP)	b5	8.5408	6.0329
(LFCC GMM)	b6	7.9877	8.2195
(PAC GMM)	b7	4.8361	4.3385
(SSC GMM)	b8	4.5091	4.8422

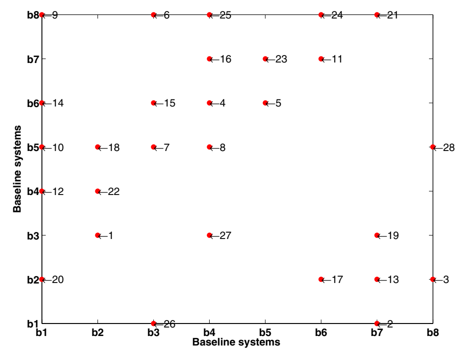

Figure 1. Combinations of the two sensor fusion.

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

Table 2. Top 7 optimal combinations for two sensor fusion system: actual performance vs. prediction.

(a) development set						
Rank	Prediction	Actual Experiments				
		sum	product	min	median	max
1	12	14	9	9	14	12
2	26	12	2	2	12	10
3	10	9	14	14	9	14
4	20	2	20	20	2	26
5	9	26	26	26	26	20
6	14	20	25	16	20	9
7	2	10	16	25	10	2

(b) evaluation set						
Rank	Prediction	Actual Experiments				
		sum	product	min	median	max
1	12	14	9	9	14	12
2	10	9	20	20	9	14
3	26	26	2	26	26	10
4	9	2	14	2	2	26
5	20	12	26	14	12	9
6	14	20	12	27	20	2
7	2	4	25	25	4	20

Table 3. Measurement of the agreement degree between the actual results and the prediction.

(a) development set					
Measurement	Sum	Product	Min	Median	Max
Rand	1	0.74613	0.74613	1	1
Jaccard	1	0.65591	0.65591	1	1
FM	1	0.79221	0.79221	1	1

(b) evaluation set					
Measurement	Sum	Product	Min	Median	Max
Rand	0.86243	0.86243	0.74603	0.86243	1
Jaccard	0.79767	0.79767	0.65591	0.79767	1
FM	0.88745	0.88745	0.79221	0.88745	1

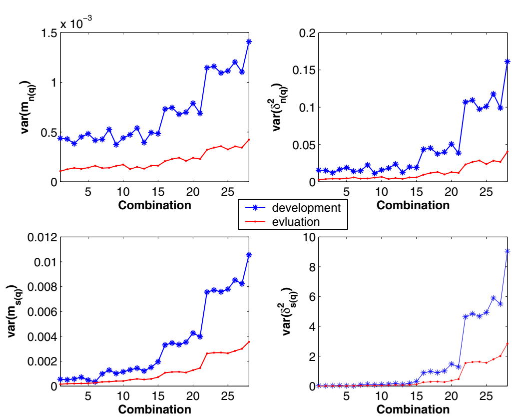

Figure 2. Cramer-Rao bounds for the prediction model.

for different fusion rules on the development set and the evaluation set. The results show that our prediction model can get a high degree of agreement with the actual results under different fusion rules.

3.2. Cramer-Rao bounds for the prediction model

According to equation (22), (23), (24) and (25) we can get the Cramer-Rao bounds for the prediction model. Figure 2 shows the Cramer-Rao bounds for the prediction of $m_{n(q)}$, $\delta^2_{n(q)}$, $m_{s(q)}$, and $\delta^2_{s(q)}$.

4 Conclusions

In this paper we proposed a prediction model which is based on the relationship among the area under the ROC curve, likelihood ratio, and discriminability measurement to predict the performance of a fusion system. We verify our prediction model on the publicly available multi-modal database XM2VTS which has the development set and the evaluation set. We use several criteria to evaluate the degree of the agreement between the prediction and actual results on both data sets. By using the prediction model we can find the optimal sensor fusion combination instead of doing the brute-force experiments. The technical approach presented here is applicable not only to biometric data but also to various other fusion problems in computer vision and pattern recognition.

References

[1] J. Daugman. Biometric decision landscapes. *Technical Report No. TR482*, 2000.

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

[3] D. M. Green and J. A. Swets. *Signal Detection Theory and Psychophysics*. 1966. John wiley and sons.

[4] J. M. Keller, S. Auephanwiriyakul, and P. D. Gader. Experiments in predictive sensor fusion. *Proc. SPIE*, 4394:1047–1058, 2001.

[5] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas. On combining classifiers. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(3):226–239, 1998.

[6] N. Poh and S. Bengio. Database, protocol and tools for evaluation score-level fusion algorithms in biometric authentication. *IDIAP research report*, 2004.

[7] N. Poh and S. Bengio. Towards predicting optimal fusion candidates: a case study on biometric authentication tasks. *Proc. Machine Learning and Multimodal Interaction Workshop*, pages 159–172, 2004.

[8] S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. 1998. Academic Press.

0-7695-2521-0/06 $20.00 © 2006 IEEE