On the Performance Prediction and Validation for Multisensor Fusion

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

Multiple sensors are commonly fused to improve the detection and recognition performance of computer vision and pattern recognition systems. The traditional approach to determine the optimal sensor combination is to try all possible sensor combinations by performing exhaustive experiments. In this paper, we present a theoretical approach that predicts the performance of sensor fusion that allows us to select the optimal combination. We start with the characteristics of each sensor by computing the match score and non-match score distributions of objects to be recognized. These distributions are modeled as a mixture of Gaussians. Then, we use an explicit Φ transformation that maps a receiver operating characteristic (ROC) curve to a straight line in 2-D space whose axes are related to the false alarm rate (FAR) and the Hit rate (Hit). Finally, using this representation, we derive a set of metrics to evaluate the sensor fusion performance and find the optimal sensor combination. We verify our prediction approach on the publicly available XM2VTS database as well as other databases.

1. Introduction

Sensors are used to collect data and computer vision and pattern recognition techniques are used to recognize objects in the sensed data. In order to increase the recognition system performance, sensor fusion techniques are widely used today. By fusing different sensors, we may achieve increased accuracy, reduced false alarms and increased range of scenarios for which a system will function correctly.

There are four levels of sensor fusion [10]: data level, feature level, score level, and decision level. Data level fusion is the combination of unprocessed sensor data. When the sensors are alike, we can consider all the data suitable for the data level fusion. Feature level fusion 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. Fusion at the score level is widely used because match scores and non-match scores from different sensors can be normalized and combined by different score-level techniques. These techniques include product of likelihood ratios, logistic regression, sum of normalized scores, maximum of the normalized scores, and weighted sum of normalized scores [9]. Decision level fusion combines decisions made by different algorithms. Decision fusion methods include statistical method, voting method, fuzzy logic based method, etc.

Given the characteristics of the single sensors, how can we find the optimal sensor combination which gives the best recognition performance? The traditional approach is to try all possible combinations of sensors by performing exhaustive experiments to determine the optimal combination. In this paper, we present a theoretical approach to predict the sensor fusion performance that allows us to select the optimal sensor combination. In this approach, first, we use the characteristics of each sensor to compute the match score and non-match score distributions which are modelled as a mixture of Gaussians. Second, we decompose the *area under the ROC curve* ($AUROC$) of the fusion system to a set of $AUROC$s which are obtained from the combination of the components from the match score and non-match score distributions. Third, we use an explicit Φ transformation that maps a ROC curve to a straight line in 2-D space whose axes are related to the FAR and the Hit rate. Finally, using this representation, we derive a set of metrics to evaluate the sensor fusion performance and find the optimal sensor combination. By using this approach, we can determine the optimal sensor combination by computing the metrics instead of performing the exhaustive experiments. We show results on various databases.

2. Related Work and Contributions

2.1. Related Work

Kittler et al. [4] develop a common theoretical framework for fusing the decision of multiple sensors. According to the Bayesian theory, for different assumptions, the problem can be expressed as the product rule or the sum rule which constitute the base of the sensor fusion strategies. They prove that many commonly used combination rules such as the min rule, max rule, median rule, and majority vote rule can be developed from the basic decision rules. Their approach can be used to develop other combination strategies. Keller et al. [3] analyze the sensor fusion by the fuzzy set theory at the decision level. They use a d-metric which is the ratio of the probability of detection to the probability of false alarm to predict the sensor fusion performance. Daugman [1] discusses the sensor fusion at the decision level. He gives the probability of a false alarm under the disjunction rule and the probability of a false reject under the conjunction rule. He concludes that it is better to use a strong sensor alone than in combination with a weaker one.

Poh and Bengio [7] propose a measurement F-ratio which is related to the *equal error rate* (EER) to find the optimal fusion candidate under the assumption that the match score and non-match score are single Gaussian distributed. They verify their approach on the BANCA multimodal database. Wang and Bhanu [11] present a prediction model which is based on the likelihood ratio to predict the sensor fusion performance. They derive the Fisher measurement for the sensor fusion system to predict the system performance. In their approach, they model the match score and non-match score as the single Gaussian distributions. We list the above approaches and our approach proposed in this paper in Table 1.

Table 1. Prediction approaches for sensor fusion performance.

Authors	Approach	Comments
Kittler et al. (1998) [4]	Prove that the commonly used fusions rules can be developed from the sum rule and product rule by the Bayesian theory	Decision level
Daugman (2000) [1]	Provides the probability of false alarm and false reject under the conjunction rule and disjunction rule based on the statistical decision theory	Decision level
Keller et al. (2001) [3]	Derive a metric based on the fuzzy theory.	Decision level
Poh and Bengio (2004) [7]	Derive a F-ratio metric based on the statistical approach	Score level, Single Gaussian distribution assumption
Wang and Bhanu (2006) [11]	Derive the Fisher measurement based on the likelihood ratio of the sensors	Score level, Single Gaussian distribution assumption
This paper	Derive a metric based on Φ transformation and ROC curve decomposition	Score level, Gaussian mixture model assumption

2.2. Contributions

The specific contributions of this paper are:

1) Our prediction approach is based on the score level fusion. We assume that the match score and the non-match score distributions are mixture of Gaussians. Based on this we decompose the $AUROC$ of the fusion system to a set of $AUROC$s which are obtained from the combinations of the components from the match score distribution and the non-match score distribution.

2) We use an explicit Φ transformation that maps a ROC curve to a straight line in 2-D space whose axes are related to the FAR and the Hit. We derive a set of metrics to evaluate the sensor fusion performance and find the optimal sensor combination.

3) We verify our prediction approach on the publicly available XM2VTS, NIST-4, ear, face and gait (video) databases. We compare this approach with the previous prediction model presented in [11].

3. Technical Approach

The $AUROC$ can be used as a measurement of a system performance [5]. In order to make the prediction approach better applicable to real data, we assume that the match score and non-match score distributions are Gaussian mixture models. We decompose the $AUROC$ of the fusion system to a set of $AUROC$s which are obtained from the combinations of the components from the match score distribution and the non-match score distribution. Since $AUROC$ is not easy to calculate, we apply a Φ transformation to map a ROC curve to a straight line in 2-D space whose axes are related to the FAR and the Hit rates. Based on this line, we derive a set of metrics to evaluate the sensor fusion performance.

3.1. Decomposition of the area under the ROC curve

Let there be two classes for a recognition system: match and non-match. We denote $f(x)$ as the match score distribution and $g(x)$ as the non-match score distribution. We assume that $f(x)$ and $g(x)$ are Gaussian mixtures. Then, we have $f(x) = \sum_{i=1}^{m} \alpha_i f_i(x)$ and $g(x) = \sum_{j=1}^{n} \beta_j g_j(x)$. Where m and n are the number of components, α_i, β_j are component proportions, $\sum_{i=1}^{m} \alpha_i = 1$ and $\sum_{j=1}^{n} \beta_j = 1$. For each component, we have $f_i(x) \sim N(m_{s_i}, \delta_{s_i}^2)$ and $g_i(x) \sim N(m_{n_i}, \delta_{n_i}^2)$, where m_{s_i} and δ_{s_i} are mean and standard deviation for match score distribution, m_{n_j} and δ_{n_j} are mean and standard deviation for non-match score distribution. For a criterion λ, we can obtain $FAR = \sum_{j=1}^{n} \int_{\lambda}^{\infty} \beta_j g_j(x)dx$ and $Hit = \sum_{i=1}^{m} \int_{\lambda}^{\infty} \alpha_i f_i(x)dx$. We know that the $AUROC$ can be used as a performance measurement. Then, we have

$$\begin{aligned} AUROC &= \int_0^1 (1 - FAR)dHit \\ &= \int_0^1 (1 - \textstyle\sum_{j=1}^{n} \beta_j FAR_j) d(\sum_{i=1}^{m} \alpha_i Hit_i) \\ &= \textstyle\sum_{i=1}^{m} \sum_{j=1}^{n} AUROC_{ij} - 1 \end{aligned} \tag{1}$$

where $AUROC_{ij} = \int_0^1 (1 - \beta_j FAR_j) d(\alpha_i Hit_i)$. For example, if $m = 3$, $n = 2$, then the $AUROC$ can be decomposed as six $AUROCs$ whose axes are the components of the match score distribution and the non-match score distribution. Figure 1 shows the decomposition of the area under the ROC curve for Gaussian mixtures. We can see that the $AUROC = \sum_{i=1}^{3}\sum_{j=1}^{2} AUROC_{ij} - 1$. The sum of the bold parts in Figure 1 is 1.

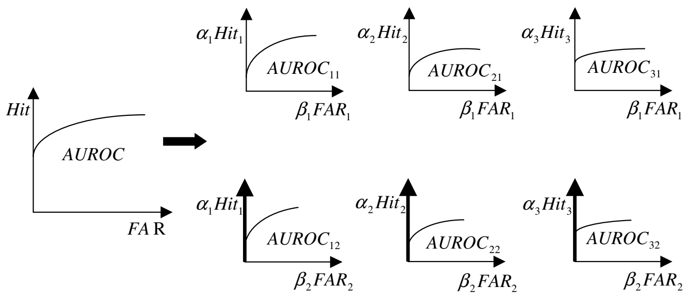

Figure 1. Decomposition of the area under the ROC curve for the Gaussian mixture model.

3.2. Fusion prediction

Now, we consider the components of the match score and non-match score distributions. We know that

$$\beta_j FA_j = \beta_j \Phi(Z_{Nj}) \tag{2}$$

$$\alpha_i Hit_i = \alpha_i \Phi(Z_{Si}) \tag{3}$$

where Φ is a standard normal distribution, $Z_{Nj} = \frac{\lambda - m_{n_j}}{\delta_{n_j}}$, and $Z_{Si} = \frac{\lambda - m_{s_i}}{\delta_{s_i}}$.

Since Φ is a monotonically increasing function, the relationship between $\beta_j FA_j$ and $\alpha_i Hit_i$ can be represented by $\beta_j Z_{Nj}$ and $\alpha_i Z_{Si}$. From equation (2) and equation (3), we can see that if we increase the value of Z_{Nj} then $\beta_j FAR_j$ will be decreased. If we decrease the value of Z_{Si}, then $\alpha_i Hit_i$ will be increased. For a fixed $\beta_j FAR_j$, the value of $\alpha_i Hit_i$ is higher, then $AUROC$ is greater. Thus, the $AUROC$ can be evaluated by $\beta_j Z_{Nj} - \alpha_i Z_{Si}$. We can get the linear relationship between $\beta_j Z_{Nj}$ and $\alpha_i Z_{Si}$ as

$$\alpha_i \cdot Z_{Si} = \frac{\alpha_i}{\beta_j} \cdot \frac{\delta_{n_j}}{\delta_{s_i}} (\beta_j \cdot Z_{Nj}) + \alpha_i \cdot \frac{m_{nj} - m_{si}}{\delta_{s_i}} \tag{4}$$

Figure 2 shows the transformation of the ROC curve to a straight 2-D linear.

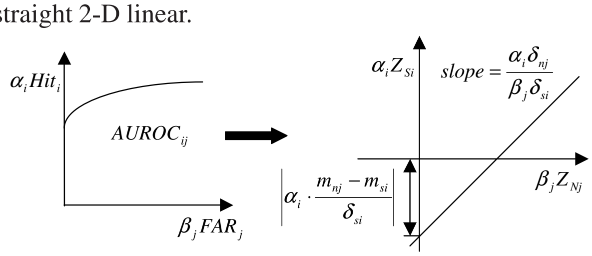

Figure 2. Transformation of the ROC curve to a straight 2-D line.

We know that when the FAR is equal to the *false reject rate* (FRR), the criterion $\lambda^* = \frac{\delta_{s_i} m_{nj} + \delta_{n_j} m_{si}}{\delta_{s_i} + \delta_{n_j}}$. Then, we have

$$\beta_j Z_{Nj} - \alpha_i Z_{Si} = (\alpha_i + \beta_j) \frac{m_{si} - m_{nj}}{\delta_{s_i} + \delta_{n_j}} \tag{5}$$

Thus, according to equation (1), the $AUROC$ can be evaluated by

$$d = \sum_{i=1}^{m}\sum_{j=1}^{n} (\alpha_i + \beta_j) \frac{m_{si} - m_{nj}}{\delta_{s_i} + \delta_{n_j}} \tag{6}$$

We use equation (6) to evaluate the sensor fusion system performance instead of performing the exhaustive experiments to determine the optimal sensor combination.

4. Experimental Results

4.1. Databases

We evaluate our prediction approach on the *extended multi modal verification for teleservices and security applications* (XM2VTS) database, *NIST Special Database 4* (NIST-4) for fingerprint, ear database, face/gait video database.

The publicly available multi-modal database 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 include 5 face baseline systems and 3 speech baseline systems. These face baseline systems and speech baseline systems have different features and classifiers. We denote these face baseline systems as $b_1, \cdots, b_5$ and these speech baseline systems as b_6, b_7, b_8.

NIST-4 database consists of 2000 pairs of fingerprints. Each of the fingerprints has two different impressions. 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 480×512 pixels. We denote the fingerprint database as e_1.

The data in the ear database are captured by Minolta Vivid 300 camera. The camera outputs 3D range images which contain 200×200 grid points. There are 155 subjects which include 17 females, six subjects have earrings, and 12 subjects have their ears partially occluded by hair (with less than 10% occlusion). A total of six images per subject are recorded. Totally 902 shots are used for the experiments since some shots are not properly recorded. Every person has at least 4 shots. There are three different poses in the collected data: frontal, left and right. We denote the ear database as e_2.

Video data are obtained by a Sony DCR-VX1000 digital video camera recorder operating at 30 frames per second. There are 45 subjects who are walking in the outdoor condition and expose a side view to the camera. Each subject has two video sequences. The number of sequences per person varies from 2 to 3. The resolution of each frame is 720×480. The video data include gait data and frontal face data. We denote the gait database as e_3 and frontal face database as e_4.

4.2. Fusion performance evaluation

From the N baseline systems, we randomly pick q sensors, then all the possible sensor combinations would be $com = N + \sum_{i=1}^{q} \mathrm{C}_i^N$ which includes the single sensor possibilities.

Suppose we have N baseline systems. Before we fuse the baseline systems, we need to normalize the match scores and the non-match scores to insure that none of the baseline system will be dominant for the fusion combinations. We denote the match score of the baseline system as ms_i and the non-match score as ns_i, where $i = 1, 2, \ldots, N$. We use three normalization methods in our experiments: Min-Max, Z-score, and Tanh. We fuse the baseline systems by the sum rule and the max rule. We denote $ms_{fusion,j}$ and $ns_{fusion,j}$ as a set of normalized match scores and non-match scores after fusion, where $j = 1, 2, \ldots, com$. Then, based on $ms_{fusion,j}$ and $ns_{fusion,j}$, we can find the point where the FAR is equal to the FRR. The values of FAR and FRR at this point is called the *equal error rate* (EER) [2]. We know that the FAR is the probability that the non-match score is above criteria λ, where

$$FAR = P(ns_{fusion,j} > \lambda) = \frac{\#\text{ of non-match score above }\lambda}{\text{total }\#\text{ of non-match scores}} \tag{7}$$

The FRR is the probability that the match score is below criteria λ, where

$$FRR = P(ms_{fusion,j} < \lambda) = \frac{\#\text{ of match score below }\lambda}{\text{total }\#\text{ of match scores}} \tag{8}$$

FAR and FRR are functions of criteria λ. By the empirical approach, we can find

$$\lambda^* = argmin_i |FA(\lambda) - FR(\lambda)| \tag{9}$$

We use the *half total error rate* ($HTER$) to evaluate the system performance, where

$$HTER = \frac{FA(\lambda^*) + FR(\lambda^*)}{2} \tag{10}$$

We repeat this process for com times. According the the $HTER$ for each combination, we can find the optimal sensor fusion combination which has the minimum $HTER$.

4.3. Prediction for the XM2VTS database

We use the expectation-maximization (EM) algorithm to estimate the match score and non-match score distributions. Figure 3 shows these distributions for the eight baseline systems in the XM2VTS database. The component numbers for the eight match score distributions are 1, 1, 1, 1, 5, 4, 1, 2 and for the eight non-match score distributions are 1, 13, 18, 1, 15, 17, 18, and 18.

We use the Min-Max, Z-score, and Tanh normalization methods to normalize these eight baseline systems. Then, we randomly combine two or three baseline systems and get $8 + \mathrm{C}_2^8 + \mathrm{C}_3^8 = 92$ combinations. For each combination,

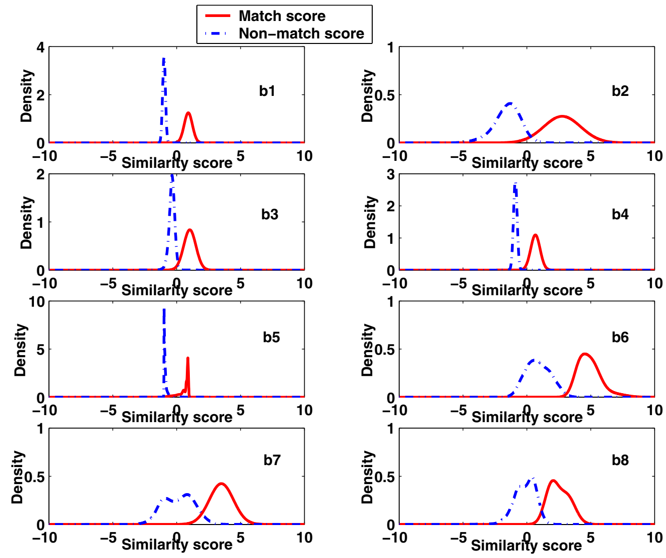

Figure 3. Match score and non-match score distributions for the baseline systems in the XM2VTS database.

we fuse the baseline systems by the sum rule and the max rule.

- **Fusion performance evaluation:** We use equation (7) and (8) to obtain the FAR and FRR for the fusion system. Then, we find the cross-point λ^* by using equation (9). Finally, we apply equation (10) to get the $HTER$ for the fusion systems. We repeat the above process for all the 92 possible combinations and rank the $HTER$ to find the optimal combination which has the minimum $HTER$.

- **Fusion performance prediction:** For each combination, we use the EM algorithm to estimate the match score and the non-match score distributions to get the component numbers, mean, variance, and weight for each component. Then, we apply equation (6) to obtain the d measurement which is used to predict the fusion system performance. We repeat the above process for all the 92 possible combinations and rank the value of d measurement to find the optimal combination which has the maximum d. We list the top 5 optimal baseline system combinations which got from the evaluation and prediction in Table 2. We observe that most of the top 5 optimal combinations from the evaluation can be found in the top 5 optimal combinations from the prediction.

We use the *Fowlkes and Mallows index* (FM) as the criteria [8] to measure the degree of the agreement between the fusion performance evaluation and prediction. The value of this criteria is between 0 and 1. The larger the value the greater is the agreement between them. Figure 4 shows the FM values for top 5 combinations for Min-Max, Z-score, and Tanh normalization methods and the sum and max fusion rules. We can see that the FM values are all greater than 0.92.

Table 2. Top 5 optimal sensor combinations for the XM2VTS database: evaluation vs. prediction.

(a) Min-Max

Rank	Evaluation		Prediction	
	Sum	Max	Sum	Max
1	(b2,b6)	(b1,b4)	(b2,b3,b7)	(b1,b4)
2	(b2,b3,b6)	(b1,b4,b5)	(b2,b3,b6)	(b1,b4,b5)
3	(b2,b3,b7)	(b1,b5)	(b2,b6)	(b1,b3,b4)
4	(b2,b7)	(b1,b5,b6)	(b2,b7)	(b1,b4,b6)
5	(b2,b6,b8)	(b1,b3,b5)	(b2,b3,b8)	(b1,b5)

(b) Z-score

Rank	Evaluation		Prediction	
	Sum	Max	Sum	Max
1	(b2,b6)	(b1,b4,b5)	(b2,b3,b7)	(b1,b4)
2	(b1,b4)	(b1,b4)	(b3,b7)	(b1,b4,b6)
3	(b1,b4,b5)	(b1,b5)	(b2,b6)	(b1,b2,b4)
4	(b3,b7)	(b1,b3,b5)	(b1,b4,b5)	(b1,b4,b5)
5	(b2,b3,b7)	(b1,b2,b5)	(b1,b4)	(b1,b3,b4)

(c) Tanh

Rank	Evaluation		Prediction	
	Sum	Max	Sum	Max
1	(b2,b6)	(b1,b4,b5)	(b2,b3,b7)	(b1,b4)
2	(b1,b4)	(b1,b4)	(b3,b7)	(b1,b4,b6)
3	(b1,b4,b5)	(b1,b5)	(b2,b6)	(b1,b2,b4)
4	(b3,b7)	(b1,b3,b5)	(b1,b4,b5)	(b1,b4,b5)
5	(b2,b3,b7)	(b1,b2,b5)	(b1,b4)	(b1,b3,b4)

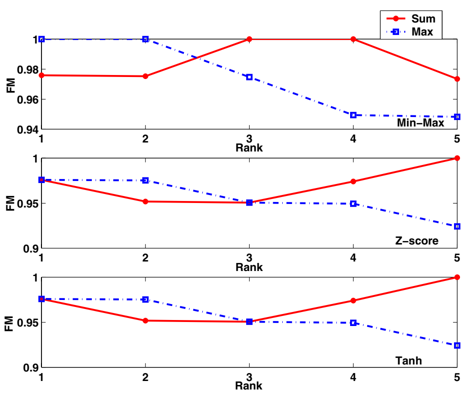

Figure 4. Fowlkes and Mallows index for top 5 sensor fusion combinations in the XM2VTS database.

4.4. Prediction for the NIST-4 database, ear database, and video database

We use the fingerprint, ear, gait, and frontal face as our four baseline systems. Note that the number of subjects in these systems are different, we repeat the data of biometrics to make sure that these four baseline systems have the same number of subjects. Figure 5 shows the match score and non-match score distributions which are estimated by the EM algorithm. The component numbers for the four match score distributions are 12, 11, 9, 9 and for the four non-match score distributions are 1, 17, 1, and 1.

As the fusion system evaluation and prediction in the XM2VTS database, we apply the Min-Max, Z-score, and Tanh normalization methods to normalize these four baseline systems. We randomly combine two or three baseline

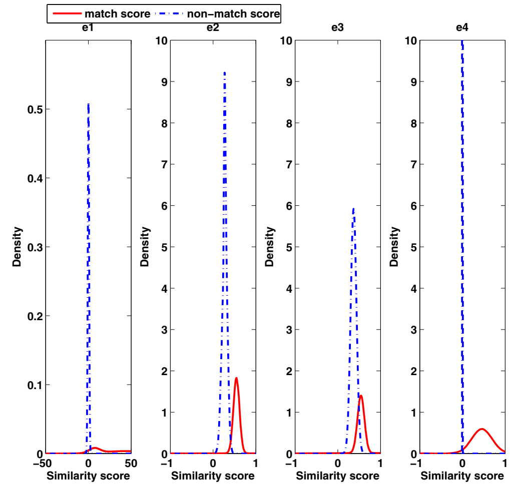

Figure 5. Similarity score distributions for the fingerprint, ear, gait, and frontal face: (a) Match score distributions. (b) Non-match score distributions.

systems and get $4 + C_2^4 + C_3^4 = 14$ combinations. For each combination, we fuse the baseline systems by the sum rule and the max rule. We use the same process as we mentioned in Section 4.3 to evaluate and predict the fingerprint, ear, gait, and frontal face fusion performance. We list the top 5 optimal baseline system combinations which are normalized by the Min-Max, Z-score, and Tanh methods and fused by the sum rule and max rule in Table 3.

We observe that most of the top 5 optimal combinations from the evaluation can be found in the top 5 optimal combinations from the prediction. Figure 6 shows the FM values for top 5 optimal combinations for Min-Max, Z-score, and Tanh normalization methods and the sum and max fusion rules. The curve is not smooth because we do not have enough data.

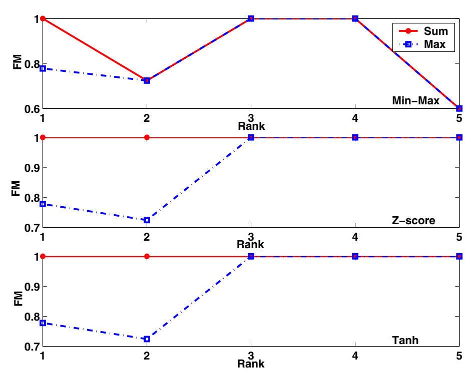

Figure 6. Fowlkes and Mallows index for top 5 sensor fusion combinations on the fingerprint, ear, and video databases.

4.5. Comparison with other prediction method

We apply the prediction model which is presented in [11] on the XM2VTS database. We use the Min-Max normalization method to normalize the baseline systems. We fuse these systems by the sum rule and the max rule. Then, we

Table 3. Top 5 optimal sensor combinations for for the NIST-4 database, ear database, and video database: evaluation vs. prediction.

(a) Min-Max

Rank	Evaluation		Prediction	
	Sum	Max	Sum	Max
1	(e1,e2)	(e2,e3,e4)	(e1,e2)	(e2,e3)
2	(e1,e2,e4)	(e2,e3)	(e1,e2,e3)	(e1,e2,e3)
3	(e1,e2,e3)	(e1,e2,e3)	(e1,e2,e4)	(e2,e3,e4)
4	(e2,e3)	(e1,e2,e4)	(e2,e3)	(e1,e2,e4)
5	(e2,e4)	(e2,e4)	(e2,e3,e4)	(e1,e2)

(b) Z-score

Rank	Evaluation		Prediction	
	Sum	Max	Sum	Max
1	(e1,e2,e3)	(e1,e2,e3)	(e1,e2,e3)	(e2,e3)
2	(e1,e2)	(e1,e2)	(e1,e2)	(e1,e2,e3)
3	(e2,e3)	(e2,e3)	(e2,e3)	(e1,e2)
4	(e1,e2,e4)	(e1,e2,e4)	(e1,e2,e4)	(e1,e2,e4)
5	(e2,e3,e4)	(e2,e4)	(e2,e4)	(e2,e4)

(c) Tanh

Rank	Evaluation		Prediction	
	Sum	Max	Sum	Max
1	(e1,e2,e3)	(e1,e2,e3)	(e1,e2,e3)	(e2,e3)
2	(e1,e2)	(e1,e2)	(e1,e2)	(e1,e2,e3)
3	(e2,e3)	(e2,e3)	(e2,e3)	(e1,e2)
4	(e1,e2,e4)	(e1,e2,e4)	(e1,e2,e4)	(e1,e2,e4)
5	(e2,e3,e4)	(e2,e4)	(e2,e4)	(e2,e4)

evaluate and predict the fusion system performance by the approach presented in [11]. In this model, the prediction is independent of the fusion rules. We list the fusion system evaluation result and the prediction result in Table 4. We compute the FM values and compare them with the prediction approach presented in this paper. Figure 7 shows the FM values obtained by the prediction approach in [11] and this paper by the Min-Max normalization method. From Figure 7, we can see that our approach presented in this paper has higher FM values than the approach presented in [11]. That means the approach presented in this paper is more effective.

Table 4. Top 5 optimal combinations for sensor fusion system: evaluation vs. prediction.

Rank	Evaluation		Prediction
	Sum	Max	
1	(b1,b4,b6)	(b1,b4)	(b1,b3,b4)
2	(b1,b4,b7)	(b1,b4,b5)	(b1,b4,b5)
3	(b1,b6)	(b1,b3,b4)	(b1,b2,b4)
4	(b1,b4,b8)	(b1,b4,b6)	(b1,b4)
5	(b1,b4)	(b1,b5)	(b1,b4,b6)

5. Conclusions

In this paper, we present a theoretical prediction approach that predicts the performance of sensor fusion that allows us to select the optimal combination. We verify our prediction approach on the multi-modal XM2VTS database, NIST-4 fingerprint database, ear database and

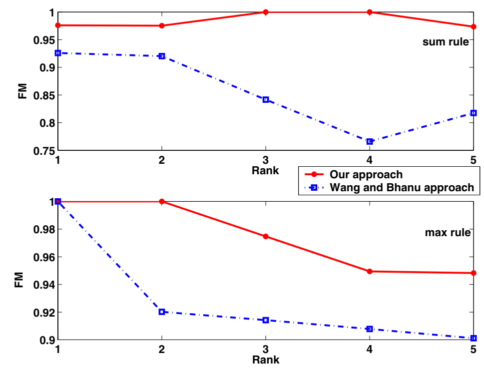

Figure 7. Fowlkes and Mallows index for top 5 optimal sensor combinations by our previous approach [11] and the approach presented in this paper on the XM2VTS database.

video databases. We use Fowlkes and Mallows index to evaluate the degree of the agreement between the fusion performance evaluation and prediction. The experimental results show that our prediction approach can predict the sensor fusion performance effectively. The technical approach presented here is applicable not only to sensor fusion 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

[2] R. O. Duda, P. E. Hart, and D. G. Stork. Pattern classifications, 2000. A Wiley-Interscience Publication. 4

[3] J. M. Keller, S. Auephanwiriyakul, and P. D. Gader. Experiments in predictive sensor fusion. *Proc. SPIE*, 4394:1047–1058, 2001. 2

[4] 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. 1, 2

[5] C. Marzban. The ROC curve and the area under it as performance measures. *Weather and Forecasting*, 19:1106–1114, 2004. 2

[6] N. Poh and S. Bengio. Database, protocol and tools for evaluation score-level fusion algorithms in biometric authentication. *IDIAP Research Report*, 2004. 3

[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. 2

[8] S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. 1998. Academic Press. 4

[9] B. Ulery, A. Hicklin, C. Watson, W. Fellner, and P. Hallinan. Studies of biometric fusion. *NISTIR 7346*, 2006. 1

[10] E. Waltz and J. Llinas. *Multisensor Data Fusion*. 1990. Artech House. 1

[11] R. Wang and B. Bhanu. Performance prediction for multimodal biometrics. *Proc. Int. Conf. on Pattern Recognition*, 3:586–589, 2006. 2, 5, 6