Probabilistic Fusion of Gait Features for Biometric Verification

Alex I. Bazin
School of Electronics and
Computer Science
University of Southampton
Southampton, UK
aib02r@ecs.soton.ac.uk

Lee Middleton
School of Electronics and
Computer Science
University of Southampton
Southampton, UK
ljm@ecs.soton.ac.uk

Mark S. Nixon
School of Electronics and
Computer Science
University of Southampton
Southampton, UK
msn@ecs.soton.ac.uk

Abstract – *This paper examines the fusion of various gait metrics in a probabilistic framework. Using three gait modalities we describe a process for determining probabilistic match scores using intra and inter-class variance models together with Bayes rule. We then propose to fuse these modalities based on established fusion rules with weights determined in a principled manner. Using a large publicly available database we show improvements through fusion, both in terms of verification accuracy and class separation; we also consider how the accuracy of each modality and the correlation between the modalities affects overall performance.*

Keywords: Fusion, Biometrics, Logistic function, Bayesian.

1 Introduction

Gait recognition is defined as the identification of a person through the pattern produced by walking. This field has produced significant interest over recent years, and through this work it has been shown that a subject's gait pattern is sufficiently unique for identification. Gait has particular advantages over other biometrics: it can be used at a distance, uses no additional skills on the part of the subject, and may be performed without the subject's awareness or active participation. All of these advantages make it particularly valuable in surveillance or security systems.

Recognition methods can be broadly divided into two groups, silhouette-based techniques and model-based techniques. Silhouette-based techniques [1] tend to offer speed and simplicity, but are only indirectly linked to gait and are difficult to normalise for noise or variations such as clothing. Model-based techniques [2] use the shape and dynamics of gait to guide the extraction of a feature vector. In this paper we utilize the dynamic method of Wagg and Nixon [2, 3] and two different silhouette-based static methods [1].

In many recognition approaches, recognition relies on distance metrics, typically Euclidean distance, i.e a form of nearest neighbour classifier, between two vectors: one of a known subject, i_C, and one of an unidentified subject, i_N. A classification decision can be made based on the Euclidean distance, d, between the two vectors. This approach has a number of weaknesses; principal amongst them is that it fails to exploit any knowledge of variation within the data, particularly which variation is due to changes between recordings of the same subject (intra-class variations) and which variation is due to changes between subjects (inter-class variations). The second important weakness of this technique is the uncertainty about the range or distribution of scores that will be produced. This is an important factor in verification (the one to one match between a new sequence and a stored template) and in data fusion. The Euclidean distance can struggle with verification since unimportant dimensions in the measurement vector can contribute greatly to the similarity score, exacerbating difficulty in finding a suitable threshold.

Data fusion combines metrics from various techniques to build one single identification score [4, 5, 6]. These methods struggle when using distance based metrics since the scale and distribution of scores across multiple techniques are unlikely to be similar. To compensate for this, score transformation is often used to approximate the posterior probability [7, 8]; however it is not clear whether these are good approximations of the posterior probability. In this paper we describe a more suitable method, where the posterior probabilities are calculated directly. Previous approaches to combining gait metrics [7, 9] have all relied on score transformation to allow data fusion, also only static fusion rules are used and no consideration is given to the imbalance of classifiers.

One solution for all of the problems described above would be to use the Bayesian classifier [10] to provide a probabilistic measure for the verification (or classification) decision. A well designed Bayesian classifier will take into account intra and inter-class variation, as well as providing well scaled (guaranteed between zero and one) and well distributed outputs. The use of Bayesian classifiers is unknown in gait recognition, although there have been some attempts to implement Bayesian classifiers in face recognition [11, 12, 13, 14]. Whilst these implementations have performed well in the classification task, initial experiments have suggested that similar performance is not achieved in the verification task. It would appear that the suitability of the Gaussian distribution for modelling these distributions has not been fully explored and that

the likelihoods arising from these models are poorly scaled and distributed [15].

In this paper we examine the application of our novel probabilistic framework to the task of data fusion. We describe our method for obtaining confidences in verification decisions using static and dynamic gait vectors. These confidences are combined using weighted fusion rules to assess the performance of our method using a large publicly available database [16]. We examine the effect of the correlation of classifiers on the efficacy of fusion.

2   Theory

2.1   Gait signature extraction

For our gait signatures we employ two methods, a dynamic method proposed by Wagg and Nixon [2, 3] and two static methods [1]. For the dynamic signature we follow the methods of Wagg and Nixon to extract a seventy-three dimensional measurement vector, i, using model based estimation. This signature derives from bulk motion and shape characteristics of the subject, articulated motion estimation using an adaptive model and motion estimation using deformable contours; examples of all of these processes can be seen in Fig. 1. After pre-processing to remove noise and background the sequence is edge detected with a Sobel operator. A motion compensated temporal accumulation algorithm [17] is used to extract the bulk motion of the subject in the horizontal plane. This is then filtered using template matching, leaving only motion due to the subject. Shape estimation is then performed using a more accurate model of the subject's shape.

Articulated motion is estimated as sinusoidal models of hip, knee, ankle and pelvic rotation. These provide a starting point for model adaptation of the subject's limb movements. An adaptive process for joint location is then applied to the sequence to form a more accurate and robust model of limb movement. This adaptive process is based on an iterative gradient descent model repeated until no changes occur over the entire sequence.

By feature selection, the processes described in [3] yield 45 parameters based on joint rotation models for the hip, knee and ankle and 18 parameters describing the subject's speed, gait frequency and body proportions. A further 10 parameters are extracted from the processes described in [2]. All of these parameters are normalised to make them size invariant.

For the first static signature the sequence undergoes the same pre-processing to remove distortion and background and silhouettes are obtained by connected component analysis and morphological operators. The output of this operation is a series of binary images of the subject silhouette over a complete gait cycle. These silhouettes are normalised to provide a common centre of mass and then summed to form an average silhouette across the complete gait cycle. This average silhouette yields a vector of 4,096 dimensions.

The second static signature is a 1,836 dimensional vector obtained from a block-based silhouette averaging method.
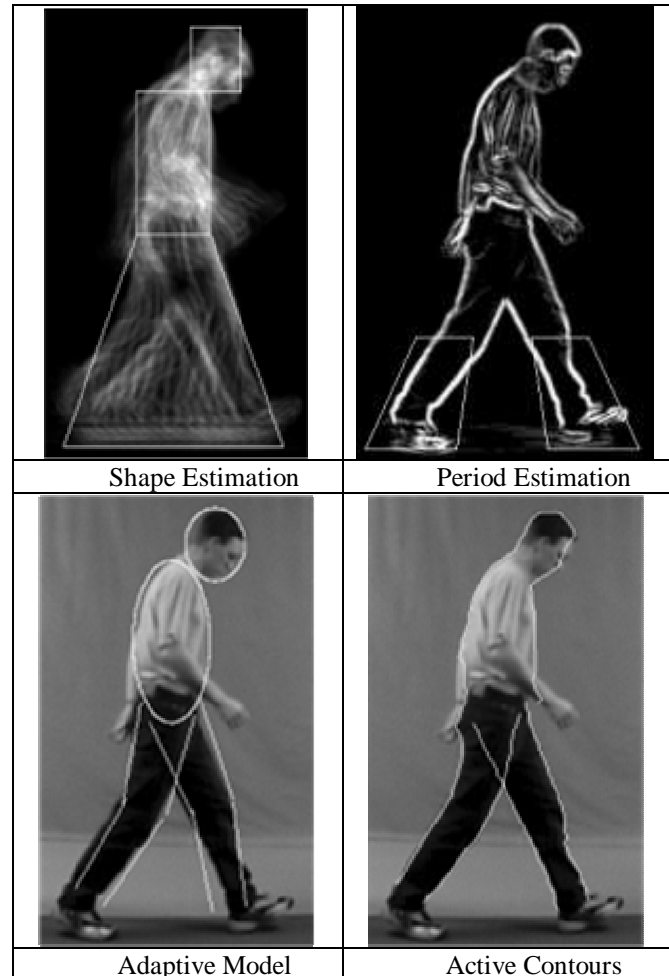

| Shape Estimation | Period Estimation |
| Adaptive Model | Active Contours |

Fig. 1 – Gait Parameter Extraction

2.2   Intra and inter-class variation

We seek to exploit our knowledge of the variation within the difference measure d to provide a probabilistic measure of whether the feature vectors i_C and i_N belong to the same subject. Specifically we wish to describe the variation in two ways, the variation that arises from differences in measurements from the same subject (intra-class variation) and the variation that is the result of differences between the measurements of different subjects (inter-class variance).

To describe this variance we take a corpus of training data that contains a number of subjects each with multiple measurement vectors. From this corpus we subtract every vector from all other vectors of the same subject to form our intra-class training set, D_C; we also

subtract every vector from every other vector in the corpus where the subjects are not identical, this forms the inter-class training set, D_I.

Having created our two training sets we then find the mean and variance of each set:

$$\mu = \frac{1}{N}\sum_{i=1}^{N}|d_i| \; \forall d \in D \qquad (1)$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\left(|d_i|-\mu\right)^2 \forall d \in D \qquad (2)$$

This process is undertaken for both the intra and inter-class training sets to give, μ_C, μ_I, σ^2_C, and σ^2_I. We justify the use of the variance rather than the covariance following Liu and Wechsler's work in face recognition [13] where they make the assumption that the covariance matrices are diagonal:

$$\Sigma = diag\left\{\sigma_1^2, \sigma_2^2, ..., \sigma_m^2\right\} \qquad (3)$$

Initial experiments performed on our data show that the covariance matrices are indeed sparse except on the diagonal and we concur with Liu and Wechsler's experiments showing no loss of performance using variance rather than covariance. We would expect this to be a general result for biometrics, since most components are constructed to be orthogonal to one another and hence have no covariate properties. If this assumption was invalid it would reduce the discriminant ability of our method.

2.3 Likelihood estimation

We wish to describe a new sequence's similarity to a stored sequence of a known subject in a probabilistic manner using the information on the mean and variance obtained in Eqs. (1, 2). To achieve this we must calculate the likelihoods of obtaining the distance d given either intra-class variation, $P(d/C)$, or an inter-class variation, $P(d/I)$, i.e that the subject is either a client or an impostor.

It is desirable to model these two distributions such that $P(d/C)$ tends to one with d less than μ_C, tending to zero as d increases beyond μ_C; conversely $P(d/I)$ should tend to zero with d less than μ_I and tend to one as d increases beyond μ_I. If the distributions of d from clients and impostors are slightly overlapping then the functions for $P(d/C)$ and $P(d/I)$ should appear as in Fig. 2. To achieve this distribution we have chosen to model $P(d/C)$ and $P(d/I)$ as logistic functions (Fig. 2) such that:

$$P(d\,|\,C) = 1 \bigg/ \left(1 + e^{\frac{d-\mu_C}{b_C}}\right) \qquad (4)$$

$$P(d\,|\,I) = 1 \bigg/ \left(1 + e^{\frac{-(d-\mu_I)}{b_I}}\right) \qquad (5)$$

$$b_m = \sqrt{3\sigma^2/\pi^2} \;\; m \in i, c \qquad (6)$$

These functions conform to our requirements that they take into account knowledge of the variation in d, that they are well distributed and that they are guaranteed to produce outputs between zero and one.
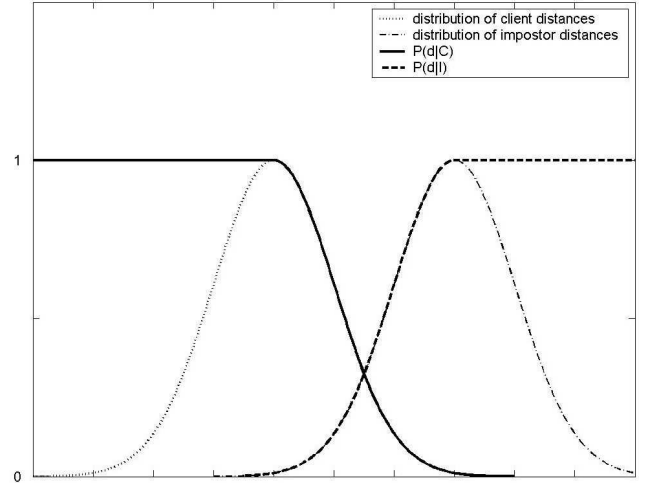

Fig. 2 – Intra and Inter-Class Distributions and Likelihoods

2.4 Bayesian Classification

From our estimates of the likelihood, the posterior probability of a subject being a client, $P(C/d)$, can be calculated directly. To achieve this we use Bayes' rule (7), with the assumption that the prior probabilities of a client or an impostor are equal, Eq. (8), and calculating $P(d)$ using Eq. (9):

$$P(C\,|\,d) = \frac{P(d\,|\,C)P(C)}{P(d)} \qquad (7)$$

$$P(C) = P(I) \qquad (8)$$

$$P(d) = P(C)P(d\,|\,C) + P(I)P(d\,|\,I) \qquad (9)$$

Eq. (7) then simplifies to:

$$P(C\,|\,d) = \frac{P(d\,|\,C)}{P(d\,|\,C) + P(d\,|\,I)} \qquad (10)$$

Having calculated the posterior probability, a suitable decision threshold, t, can be implemented for the verification task. Hence if $P(C/d)$ is greater than t we

accept the assertion that the subject is a client, otherwise we reject them as an impostor. The value of t may be adjusted to achieve the desired trade-off between false accept and false reject rates.

2.5 Data fusion rules

Having obtained the posterior probability for our static and dynamic gait vectors we may now combine them using weighted product and sum decision rules [4, 5]. The rules are given by Eqs. (11, 12), where $P(C|d_i)$ is the posterior probability from the ith classifier and R is the number of classifiers to be fused.

$$\text{Sum rule:} \quad P(C \mid d_1,...,d_R) = \sum_i^R w_i P(C \mid d_i) \quad (11)$$

Product rule:

$$P(C \mid d_1,...,d_R) = \sqrt[R]{\prod_i^R P(C \mid d_i)^{w_i R}} \quad (12)$$

The optimal weights, w_i, for weighted fusion are determined by the additive error, E_{add} from each classifier [18] and may be found using Eq. (13):

$$w_i = \left(\sum_k^R \frac{1}{E_{add}^k} \right)^{-1} \frac{1}{E_{add}^i} \quad (13)$$

Due to the difficulties in calculating the value of E_{add} we instead approximate this value with the equal error rate of each classifier.

3 Methodology

Using the Southampton HiD database [16] consisting of 1,079 sequences from 115 subjects walking to the left (such as the example image shown in Fig. 3) we were able to construct training, gallery, client and impostor sets; these sets were converted to both dynamic and static feature vectors. The training set consisted of 145 sequences of 15 subjects that could be used to estimate the intra and inter-class mean and variance; the gallery consisted of single sequences from 100 subjects; the client set consisted of 834 sequences each matched to a subject in the gallery set; the impostor set consisted of 834 sequences where the sequences were not matched to a subject in the gallery.

Fig. 3 – Example Image from Gait Sequence

Using the procedure described in section 2.2 the training vectors were differenced from those of the same and different subjects to form 1,322 and 19,558 difference vectors respectively. These were then used to calculate the intra and inter-class means and variances for each modality. Posterior probabilities were then calculated using the methods set out above for the client and impostors sets for each modality. The modalities were then fused using weighted product and sum rules.

4 Results

The Equal Error Rates before fusion are 7.3% for the dynamic method, 15.5% for the first static method and 23.3% for the second static method. Using the definitions and theory of Daugman [6], since the error rate of the best classifier is less than half that of the other classifiers, both static methods are termed highly imbalanced with respect to the dynamic method. Because of this Daugman asserts that we should see no improvement when fusing these modalities if using decision fusion methods. Table 1 shows the results obtained from the experiments. In this we see that we see improvements over the unfused results in all cases. It is likely that Daugman's assertion does not hold for our experiments due to the increased flexibility of the score fusion methods we are using.

Using McNemar's Test [19] to evaluate the statistical significance of any improvement we find that the improvement over the dynamic method from adding either static modality using the weighted sum rule is significant at the 1% level; further we find that the improvement from fusing both static modalities over using the single best performing static modality is statistically significant at the 2% level. The slight improvement in verification performance by adding the second static modality to the combination of the dynamic and first static modalities is not statistically significant. Despite these improvements appearing to be small it must be remembered that the methods we are using are already highly effective and that for those results that are statistically significant it is unlikely that this improvement is due to feature space noise. Fig. 4 shows the Receiver Operator Characteristic curve for the single

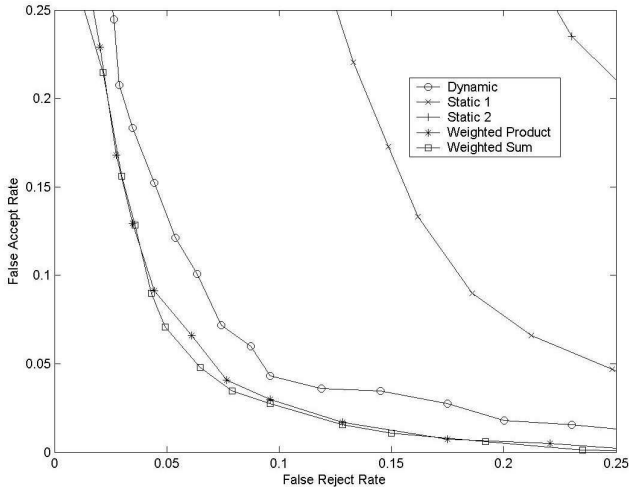modalities and the fusion of all modalities using the weighted product and sum rules.

Fig. 4 – Receiver Operator Characteristic Curves

Table 1 – Error Rates and Decidability Index

Method	Equal Error Rate (%)	Decidability Index (d')
Dynamic	7.3	2.90
Static 1	15.5	1.93
Static 2	23.2	1.75
Weighted Product (Dynamic & Static 1)	6.2	3.19
Weighted Product (Dynamic & Static2)	6.8	3.06
Weighted Product (Static 1 & Static 2)	14.5	2.10
Weighted Product (Dynamic & Static 1&2)	6.3	3.23
Weighted Sum (Dynamic & Static 1)	5.9	3.24
Weighted Sum (Dynamic & Static2)	6.7	3.11
Weighted Sum (Static 1 & Static 2)	13.6	2.16
Weighted Sum (Dynamic & Static 1&2)	5.8	3.33

It is also of interest to consider how the fusion process improves the separability of the clients and impostors; this can be measured by Daugman's decidability index [6] and is given by:

$$d' = \frac{|\mu_1 - \mu_2|}{\sqrt{\frac{1}{2}\left(\sigma_1^2 + \sigma_2^2\right)}} \qquad (14)$$

Where μ_1 and μ_2 are the mean values of the client and impostor posterior probabilities respectively, and σ^2_1 and σ^2_2 are the variances for the client and impostor posterior probabilities. The decidability concerns the area of overlap between the two distributions; if this area is large, decidability is low. By way of comparison the decidability index for an experiment with a classification rate of 99.2% on 252 examples was 3.43 [20]. The values for d' can be seen in Table 1 with an example distribution from the weighted sum fusion of all modalities shown in Fig. 5.

Using the decidability index to evaluate the various methods we see that they are in agreement with the Equal Error Rates (Fig. 6). It is apparent that the decidability index is roughly inversely proportional to the Equal Error Rate for our example methods; this is to be expected as the more separated the client and impostor sets are, the greater noise immunity one would expect the method to have.
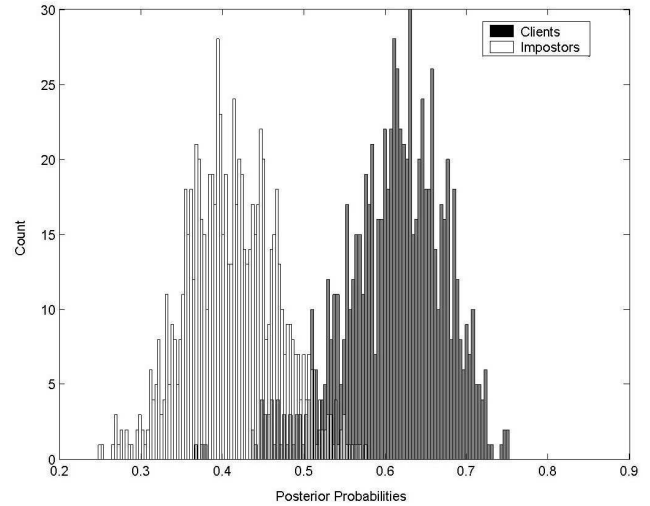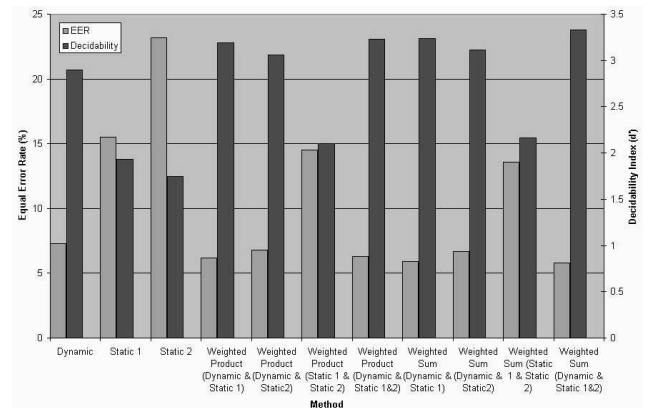

Fig. 5 – Distributions from the Weighted Sum Fusion

Fig. 6 – Equal Error Rate and Decidability Index

We also calculated the correlation between the various modalities using the methodology described in [21]. The correlation ρ_{nc} is given by:

$$\rho_{n_c} = \frac{nN_c^f}{N - N_c^t - N_c^f + nN_c^f} \qquad (15)$$

Where n is the number of classifiers under test, N is the total number of sequences (1,664), N_C^f is the number of sequences where all classifiers have an incorrect output at threshold C, and N_C^t is the number of sequences where all classifiers have a correct output at a threshold C. The paper proposes adding additional modalities in descending order of accuracy, calculating the correlation each time; only if the correlation is reduced is it acceptable to add this modality. Correlations for the various modalities are shown in Table 2 together with the percentage improvement in error rate over the best modality alone.

As we can see reduction in correlation does not necessarily lead to large improvements in recognition capability, this is most apparent in the case of all three modalities where the reduction in correlation is marked but the improvement in the EER is not statistically significant. We believe that there are a number of possible explanations for this occurrence: it may be that the fusion and weighting schemes we have tested are not correctly set to exploit this reduction in correlation; or that the sequences that we are testing are not sufficiently variable to be able to gain advantage from the reduced correlation.

Table 2 – Correlation and improvement in accuracy

Modalities	Correlation (ρ_{nc})	Improvement in EER (%)
Dynamic/ Static 1	0.20	19.1
Dynamic/ Static 2	0.16	8.2
Static 1/ Static 2	0.28	12.2
Dynamic/ Static 1/ Static 2	0.05	1.7

5   Conclusions

In this paper we have described a novel probabilistic framework for biometric recognition and data fusion; we show the framework applied to the fusion of various combinations of three gait modalities. Our results show that it is possible to gain improvements in classifier accuracy through probabilistic data fusion of multiple modalities, even when these modalities are highly imbalanced. We show also show that although in our experiments adding a third modality does not significantly improve verification accuracy, concurrent improvements in the decidability index and reduction in the correlation indicate that further scope for improvement exists either using a different fusion/weighting scheme or when using more variable data.

In future work we shall consider whether the weighting scheme used is optimal and how we may account for correlation when calculating the weights; we also wish to extend this work to other biometric modalities.

References

[1] G. V. Veres, L. Gordon, J. N. Carter, and M. S. Nixon, What image information is important in silhouette-based gait recognition?, *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR 2004)*, 2004, pp. 776-782.

[2] D. K. Wagg and M. S. Nixon, Automated markerless extraction of walking people using deformable contour models, *Computer Animation and Virtual Worlds*, vol. 15, pp. 399-406, 2004.

[3] D. K. Wagg and M. S. Nixon, On automated model-based extraction and analysis of gait, *Proc. 6th IEEE Int'l Conf. Automatic Face and Gesture Recognition.*, 2004, pp. 11-16.

[4] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, On combining classifiers, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 226-239, 1998.

[5] J. A. Benediktsson and P. H. Swain, Consensus theoretic classification methods, *IEEE Transactions on Systems, Man and Cybernetics*, vol. 22, pp. 688-704, 1992.

[6] J. Daugman, Biometric decision landscapes, TR482, University of Cambridge Computer Laboratory 1999.

[7] L. Wang, H. Ning, T. Tan, and W. Hu, Fusion of static and dynamic body biometrics for gait recognition, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, pp. 149-158, 2004.

[8] G. Shakhnarovich and T. Darrell, On probabilistic combination of face and gait cues for identification, *Proc. 5th IEEE Int'l Conf. Automatic Face and Gesture Recognition*, Washington, USA, 2002, pp. 169-174.

[9] N. Cuntoor, A. Kale, and R. Chellappa, Combining multiple evidences for gait recognition, *IEEE Int'l Conf. Accoustics, Speech, and Signal Processing (ICASSP 2003)*, Hong Kong, 2003, pp. 33-36.

[10] P. Langley, W. Iba, and K. Thompson, Analysis of Bayesian classifiers, *National Conf. Artificial Intelligence*, San Jose, USA, 1992, pp. 223-228.

[11] R. Chellappa, S. Zhou, and B. Li, Bayesian methods for face recognition from video, *Proc. Int'l Conf. Acoustics, Speech and Signal Processing (CASSP 02)*, Orlando, FL, USA, 2002, pp. 4068-71.

[12] L. Chengjun, A Bayesian discriminating features method for face detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 725-740, 2003.

[13] C. Liu and H. Wechsler, Probabilistic reasoning models for face recognition, *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, Santa Barbara, USA, 1998, pp. 827-32.

[14] B. Moghaddam, T. Jebara, and A. Pentland, Bayesian face recognition, *Pattern Recognition*, vol. 33, pp. 1771-1782, 2000.

[15] A. I. Bazin and M. S. Nixon, Facial Verification Using Probabilistic Methods, *Proc. British Machine Vision Association Workshop on Biometrics*, London, 2004.

[16] J. D. Shutler, M. G. Grant, M. S. Nixon, and J. N. Carter, On a Large Sequence-based Human Gait Database, *Proc. Recent Advances in Soft Computing*, pp. 66-71, 2002.

[17] D. K. Wagg and M. S. Nixon, Model-Based Gait Enrolment in Real-World Imagery, *Proc. Multimodal User Authentication*, 2003, pp. 189-195.

[18] F. Roli, G. Fumera, and J. Kittler, Fixed and trained combiners for fusion of imbalanced pattern classifiers, *Proc. 5th Int'l Conf. Information Fusion*, Annapolis, MD, USA, 2002, pp. 278-84.

[19] J. R. Beveridge, B. A. Draper, K. She, and G. H. Givens, Parametric and nonparametric methods for the statistical evaluation of human id algorithms, *Proc. 3rd Workshop on the Empirical Evaluation of Computer Vision Systems*, 2001.

[20] D. J. Hurley, M. S. Nixon, and J. N. Carter, Force field feature extraction for ear biometrics, *Computer Vision and Image Understanding*, vol. Accepted, 2004.

[21] K. Goebel, Y. Weizhong, and W. Cheetham, A method to calculate classifier correlation for decision fusion, *Information, Decision and Control*, Adelaide, SA, Australia, 2002, pp. 135-140.