Gait Feature Subset Selection by Mutual Information

Baofeng Guo and Mark. S. Nixon

Abstract— **Feature selection is an important pre-processing step for pattern recognition. It can discard irrelevant and redundant information that may not only affect a classifier's performance, but also tell against system's efficiency. Meanwhile, feature selection can help to identify the factors that most influence the recognition accuracy. The result can provide valuable clues to understand and reason what is the underlying distinctness among human gait-patterns. In this paper, we introduce a computationally-efficient solution to the problem of human gait feature selection. We show that feature selection based on mutual information can provide a realistic solution for high-dimensional human gait data. To assess the performance of the proposed approach, experiments are carried out based on a 73-dimensional model-based gait features set and a 64 by 64 pixels model-free gait symmetry map. The experimental results confirmed the effectiveness of the method, removing about 50% of the model-based features and 95% of the symmetry map's pixels without significant accuracy loss, which outperforms correlation and ANOVA based methods.**

I. INTRODUCTION

Human gait recognition is to identify a person through the pattern or style of walking. This technique becomes quite appealing when it is difficult to get other biometrics information at the resolution required [1]. The information contained in gait data, no matter the spatial-temporal measurements, such as step length, cycle time, etc., or the kinematic measurements, such as joint rotation and angles of the hip, knee , ankle, etc., all can reflect human body's distinct status and differ from each other to some extent. There is also evidence in medical and psychophysics studies, showing that human gait can be unique under certain circumstances [1], [2]. Therefore, since the 1990s gait recognition has received significant attention [3], [4].

Approaches to extract gait features can be roughly categorized as two classes: model-based and model-free methods. The model-based methods analyze the structures underlying the gait data, and extract measurable parameters, such as the kinematic values. Model-free methods generally consider the motion pattern of human body holistically, such as the silhouette pictures. However, no matter the model-based or the model-free methods, the extracted features could be quite high dimensional. This is partly for the reasons of opaqueness in the understanding gait's distinctness and to obtain as much useful information as possible. For example, the feature vectors in a model-based method [5], [6] have 73 components, let alone the model-free silhouette-based data

[7]–[9] that is usually a picture with the size such as 64 by 64 pixels.

Such high-dimensional feature (or data) spaces present several problems to gait recognition. First, it may affect the performance of many classification methods. Although some techniques, such as the k-nn method, may less suffer from this problem, it becomes severe for those probability-based algorithms, i.e., the so called 'curse of dimensionality' [10]. Second, due to the existence of many gait-related factors, e.g., health, age, body size, weight, speed, etc., plus the opaqueness in understanding the underlying recognition mechanism, many redundant features tend to be included to avoid any loss of useful information. For instance, we can argue that many model-based parameters are actually correlated to each other more or less, such as the hip and knee rotation. This is also true in the case of silhouette-based data, where it is known that the majority of pixels in the background area are useless. Finally, high-dimensional data always imposes requirements for storage space, computational load and communication bandwidth, which become vital factors in the actual deployment of a real-time system.

It is therefore advantageous to identify the important features for gait recognition, and then remove those features which convey little or redundant information. This can be implemented by feature selection [11], [12] or other dimensionality reduction techniques. Apart from meeting the computational challenges, interpretation of these identified key-features can also provide valuable clues or some insights, which are useful to understand what is the intrinsic mechanism regarding the gait's uniqueness. Previous researches in gait feature selection [2]–[4], [9], [13]–[16] have mainly considered the conventional statistical-tools, such as principal component analysis (PCA) and analysis of variance (ANOVA). However, there are still some remaining challenges for gait feature selection. PCA is a popular technique in dimensionality reduction, but this transform is not optimized for class separability. This means that the identified features may not coincide with the discriminatory information that recognition really requires. The PCA-based techniques also cannot directly pick up the important features, and have to resort to other extra strategies [9], which will further reduce the accuracy. The ANOVA technique can help to identify the features based on a null hypothesis test. But the difference between classes highlighted by ANOVA is mainly carried out by examining whether the population mean vectors are the same, so it lacks the explicit or definite relation with the recognition accuracy.

In this paper, we consider feature selection for gait recog-

nition from point of view of information theory. Like many separability measures, mutual information (MI) evaluates the statistical dependence between two random variables and so can be used to measure the utility of selected features to recognition. The advantages of the MI-based feature selection are its close relation with the desired Bayes error, a well-found theoretical framework and efficient implementation strategy [17], [18].

The remainder of this paper is organized as follows. We first briefly review two conventional metrics for feature selection in Section II. Then we introduce the MI-based feature selection in Section III. To assess the performance of these methods, we implement gait feature selection based on a 73-dimensional model-based feature set and a 64×64 pixels (model-free) symmetry map, respectively. Experiments are carried out to compare the MI-based method with the traditional correlation-based and ANOVA-based methods. The results are shown in Section IV. Finally, conclusions and discussions are given in Section V.

II. STATISTICAL METHODS FOR GAIT FEATURE SELECTION

The objective of feature selection is to find a subset of features, satisfying a certain criterion. In pattern recognition, a nature selection metric will be the classification accuracy or inversely the Bayes classification error. But direct minimization of the Bayes error cannot be analytically performed, so a wide range of alternative statistics that are easier to evaluate have been proposed. Two typical measures used in the gait feature selection are introduced as follows:

1) In statistics, Analysis of variance (ANalysis Of VAriance, ANOVA) is a general method for studying sampled-data relationships, in which the observed variance is partitioned into components due to different explanatory variables, such as between-class and within-class variance. 'One-way ANOVA' is the simplest case, and its purpose is to test for significant differences between classes' means. The F-test in ANOVA gives a statistic to indicate to what extent the samples of a variable are drawn from a distribution of the same means, i.e., no differences between means of different classes. It can be shown that this number, linked with the between- and within-class variance (similar to the Fisher's linear discriminant), determines whether the groups are actually different in the measured characteristic. This provides useful information for feature selection.

2) Correlation indicates the strength and direction of a linear relationship between two random variables. In other words, correlation refers to the departure of two variables from linear independence. Thus this metric is also useful to investigate how far a feature departs from the class label, and in a broad sense was used for feature selection.

The existing techniques lack explicit and definite relation with the classification accuracy or Bayes classification error, and using them for gait feature selection may deviate from the ideal solution. Hence, it is preferable to consider other metrics which are able to detect more general dependency in the data.

III. FEATURE SELECTION BASED ON MUTUAL INFORMATION

In information theory, mutual information measures the statistical dependence between two random variables and so can be used to evaluate the relative utility of each feature to classification [17]. Given the relationship between mutual information and Bayes classification error, the feature selected by mutual information analysis is approaching to a criterion by optimizing the Bayes error. Also, the implementation of mutual information needs relatively lower computational cost [18], which further makes it a convenient alternative metric.

A. Mutual information and entropy

The mutual information is a quantity that measures the mutual dependence of the two variables, and is defined as:

$$I(X,Y) = \int_Y \int_X p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)} \, dx\, dy \qquad (1)$$

where $p(x,y)$ is the joint probability density function of continual random variables X and Y, and $p(x)$ and $p(y)$ are the marginal probability density functions respectively. It is not difficult to find that mutual information is related to entropy as:

$$I(X,Y) \quad = \quad H(Y) - H(Y|X) \qquad (2)$$

given the Shannon entropy (discrete) defined as:

$$H(X) = -\sum_X p(x) \log p(x) \qquad (3)$$

The use of mutual information for feature selection can be intuitively justified by the following arguments: Let Y be variable standing for the class label (e.g., the subject's identity), and X a variable denoting the value for a gait feature. The entropy $H(Y)$ is known to be a measure of the amount of uncertainty about Y (i.e., the objective of recognition), while $H(Y|X)$ is the amount of uncertainty left in Y when knowing an observation X. From (2), it is seen that $I(X,Y)$ is the reduction in the uncertainty of class label Y by the knowledge or measurement obtained at gait feature X. Hence, mutual information can be interpreted as the amount of information that the measurement at gait feature X contains about the class label Y (see Venn diagram in Fig. 1). In other words, MI is capable to reflect the amount of information that a gait feature X contains about the class label Y. Since the variable defined by class label Y is the required classification result, the mutual information measures the capability (in the sense of information theory) of using this gait feature to predict the class label, i.e., the subject's identity.

Comparing with the statistical tools introduced in Section II, we can see that the mutual information has more clear and definite relation with the recognition objective. From the point of view of statistics, the mutual information measures more general dependency in the data, and therefore may lead to better capability to feature selection.
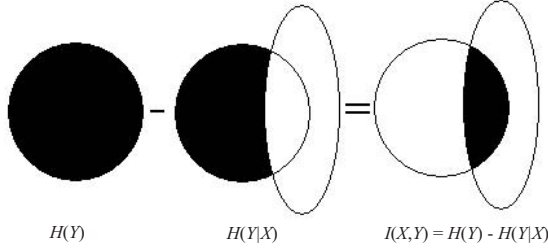
Fig. 1. Illustration of mutual information

B. Calculation of mutual information for feature selection

Feature selection based on MI can be described as follows: Given a set of original data vectors $\mathbf{x}'$ with M components or variables, and Y the corresponding output class label (e.g., the subject identity), find a subset variables $\mathbf{x} \subset \mathbf{x}'$ with N components ($N < M$) that maximizes MI $I(\mathbf{x}, Y)$, i.e., $J(\mathbf{x}^0) = \max_{\mathbf{x} \subset \mathbf{x}'} I(\mathbf{x}, Y)$. The number of ways of choosing N from the M features is $\binom{M}{N}$, which means that a huge number of MI evaluations might be needed.

Aiming at this problem, we propose a gradient ascent optimization strategy to maximize MI, described as follows. First, we show that a multi-dimensional MI can be decomposed into a series of one-dimensional MIs:

Let $\mathbf{x} = (X_1, X_2, \ldots, X_M)$ be a random vector representing the selected features $X_i, i = 1, 2, \ldots, M$, and Y the random variable corresponding to the class label. The mutual information between them can be written as:

$$I(\mathbf{x}, Y) = I((X_1, X_2, \ldots, X_M), Y) \tag{4}$$

If $\mathbf{x}$ only has two components, i.e., $\mathbf{x} = (X_1, X_2)$, (4) becomes:

$$
\begin{aligned}
I(\mathbf{x}, Y) &= I((X_1, X_2), Y) \\
&= H(X_1, X_2) - H(X_1, X_2|Y) \tag{5}
\end{aligned}
$$

From (2), we can derive the following two equations:

$$H(X_1, X_2) = H(X_1) + H(X_2) - I(X_1, X_2) \tag{6}$$

and

$$H(X_1, X_2|Y) = H(X_1|Y) + H(X_2|Y) - I(X_1, X_2|Y) \tag{7}$$

Substituting $H(X_1, X_2)$ and $H(X_1, X_2|Y)$ of (5) into (6) and (7), we get:

$$
\begin{aligned}
I(\mathbf{x}, Y) &= H(X_1, X_2) - H(X_1, X_2|Y) \\
&= H(X_1) + H(X_2) - I(X_1, X_2) - H(X_1|Y) \\
&\quad - H(X_2|Y) + I(X_1, X_2|Y) \\
&= \sum_{i=1,2} I(X_i, Y) - I(X_1, X_2) + I(X_1, X_2|Y)
\end{aligned}
\tag{8}
$$

Extending (8) to more than two components, we have the following equation:

$$
\begin{aligned}
I(\mathbf{x}, Y) &= \sum_i I(X_i, Y) - \sum_i \sum_{j \neq i} I(X_i, X_j) \\
&\quad + \sum_i \sum_{j \neq i} I(X_i, X_j|Y)
\end{aligned}
\tag{9}
$$

Based on equation (9), to maximize $I(\mathbf{x}, Y)$, $\mathbf{x} = (X_1, X_2, \ldots, X_M)$, the first variable can be chosen as:

$$X_1^0 = \max_i I(X_i, Y)$$

where X_k^0 represents the result of maximization at step k.

Then, the second variable is chosen as:

$$X_2^0 = \max_{i \neq 1}[I(X_i, Y) - \sum_{i \neq 1} I(X_i, X_1^0) + \sum_{i \neq 1} I(X_i, X_1^0|Y)]$$

The remaining variables are chosen in the same way until the pre-specified number, N, of variables is reached:

$$
\begin{aligned}
X_n^0 &= \max_{i \neq j}[I(X_i, Y) - \sum_j \sum_{i \neq j} I(X_i, X_j^0) \\
&\quad + \sum_j \sum_{i \neq j} I(X_i, X_j^0|Y)]
\end{aligned}
$$

where $X_j^0, j = 1, 2, \cdots, n-1$ are the variables already selected.

The above strategy selects features sequentially, and so avoids the problem of combinatorial explosion. At each step, the next feature will be selected so as to maximize $I(\mathbf{x}, Y)$ incrementally. This is a similar idea to the gradient ascent or other hill-climbing algorithms.

Fig. 2. Illustration of one sequence in Southampton HiD gait database

IV. EXPERIMENTS

To assess the MI-based gait feature selection, an indoor Southampton HiD Gait database [4] (http://www.gait.ecs.soton.ac.uk/database) was used. This database consists of 2163 sequences from 116 subjects walking to both the left and right. Each subject was filmed from a fronto-parallel viewpoint at a resolution of 720×576 pixels, in controlled laboratory conditions. The sequence is recorded at a rate of 25 frames per second with approximately 90 frames per gait sequence. An example of this sequence is shown in Fig. 2.

A. Gait feature sets

Based on this dataset, two approaches are employed to obtain the gait features. The first approach is a model-based method [5], which extracts a 73-dimensional measurement vector for each gait sequence (see some examples in Fig. 3(a)), based on techniques such as adaptive model and deformable contours. The detailed feature components include 18 model parameters based on joint rotation models

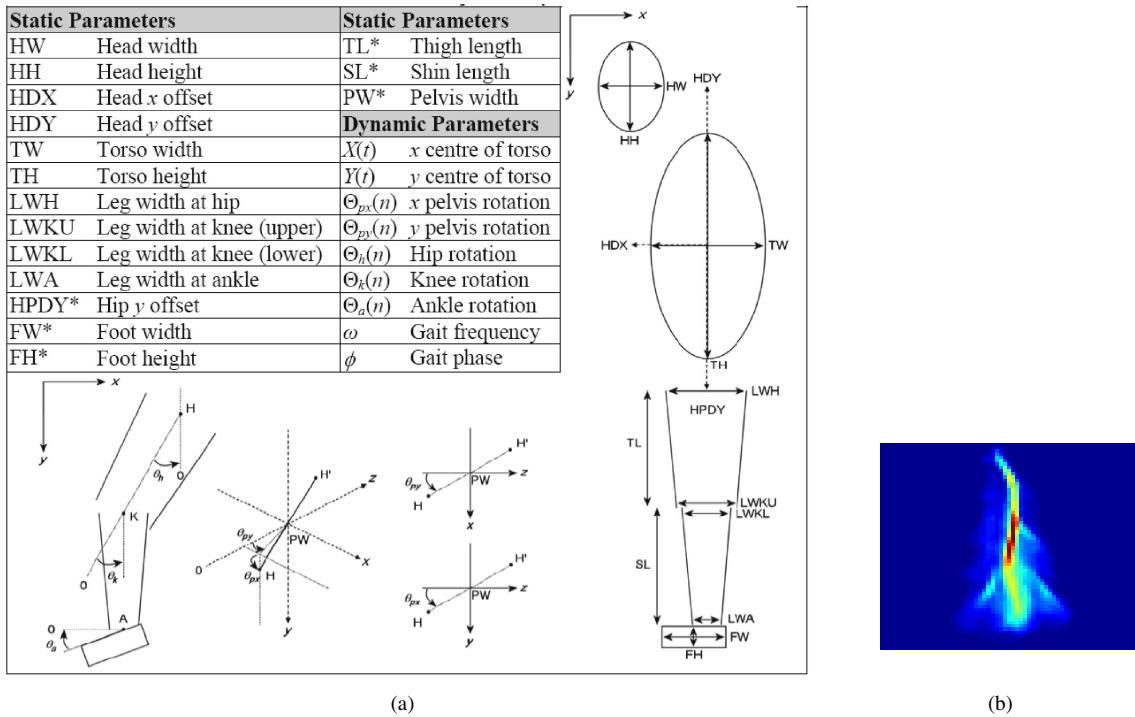| Static Parameters		Static Parameters	
HW	Head width	TL*	Thigh length
HH	Head height	SL*	Shin length
HDX	Head x offset	PW*	Pelvis width
HDY	Head y offset	**Dynamic Parameters**	
TW	Torso width	$X(t)$	x centre of torso
TH	Torso height	$Y(t)$	y centre of torso
LWH	Leg width at hip	$\Theta_{px}(n)$	x pelvis rotation
LWKU	Leg width at knee (upper)	$\Theta_{py}(n)$	y pelvis rotation
LWKL	Leg width at knee (lower)	$\Theta_h(n)$	Hip rotation
LWA	Leg width at ankle	$\Theta_k(n)$	Knee rotation
HPDY*	Hip y offset	$\Theta_a(n)$	Ankle rotation
FW*	Foot width	ω	Gait frequency
FH*	Foot height	ϕ	Gait phase

(a)   (b)

Fig. 3. (a) List of some geometric human body model parameters and their illustration [19]; (b) Example of a gait symmetry map.

for the hip, knee and ankle, and 28 parameters describing the subjects' speed, gait frequency, body proportions, etc.

The second approach is a model-free method based on analyzing the symmetry of human motion [8]. This method is reinforced from the psychologists' view that human gait is a symmetrical pattern of motion, and therefore suggest symmetry is suitable for gait recognition. The generalized symmetry operator was applied to locate features according to their symmetrical properties without relying on the borders of a shape or on general appearance. For each sequence, the obtained gait signature is a 64 by 64 pixels symmetry map, such as in Fig. 3(b).

B. Experimental results

Tests of recognition accuracy have been carried out to assess the proposed feature selection method. Currently-popular support vector machines (SVMs) [20], [21] were chosen as the classifiers in these experiments since they are less sensitive to the data's dimensionality. Although SVMs are used here, the proposed method is not limited to this particular classifier and other classification algorithms are also applicable.

Three feature selection methods, namely Pearson correlation [22], ANOVA [9] and the mutual information based methods are implemented respectively, and then tested on the above two feature sets to obtain the corresponding recognition rate on the reduced feature subsets.

To calculate the selection metrics, i.e., the correlation coefficients, one-way ANOVA's f-statistics and MI, half of the sequences from each subject were randomly chosen as the training set. This training set is assumed as the known data

samples, and then can be used for multiple purples, such as to learn the above selection metrics and validate the SVMs' parameters. The remaining 50% sequences are assumed as the unseen data to form the testing set on which the recognition accuracy was assessed. Since SVMs are inherently binary (two-class) classifiers, a 'one-against-one' scheme was used with subsequent majority voting to give a multi-class result. The kernel function used is an inhomogeneous polynomial, and the SVMs' parameters, i.e., polynomial order and the penalty parameter C are chosen by a validation procedure based on the training data.

The main objective of feature selection is to choose the most 'informative' or 'relevant' features, and achieve the possible highest recognition accuracy on the reduced feature set. The experiments were thus designed to assess the change of recognition accuracy as features are progressively selected (i.e., the cumulative recognition rate). The feature selection results for the model-based feature set are shown in Fig. 4, where the line marked with 'O' represents the results of the MI-based method, benchmarked by that of the correlation-based method (the line marked with '+') and the ANOVA-based method (the line marked with 'Δ'). Data points are at 2 features increments at each step. The random samplings to generate the testing set and training set were repeated 10 times to allow an estimate of the error inherent in this sampling process. The corresponding recognition rates with their error bars (i.e., CCR $\pm$ stand deviation) are also shown in Fig. 4.

From Fig. 4, it is seen that the feature subset selected by the mutual information achieves a better recognition rate with fewer model parameter than for the correlation and ANOVA
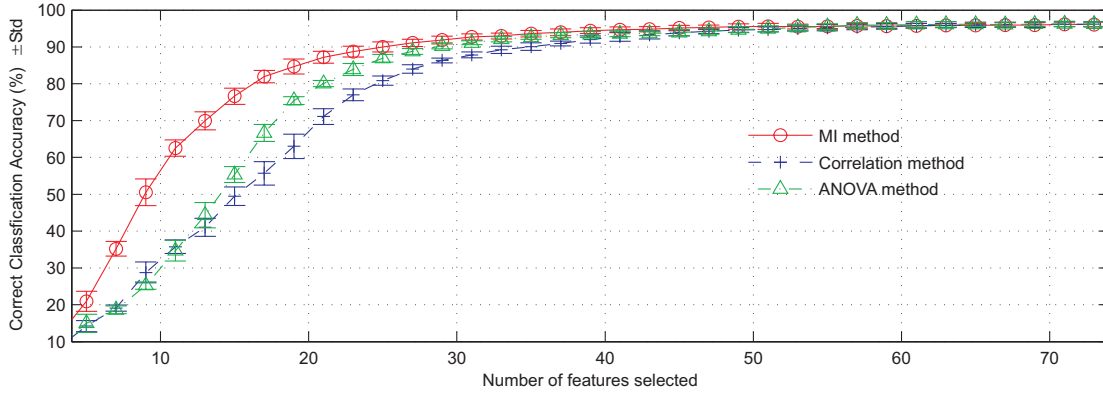
Fig. 4. Cumulative recognition rates as the model-based features are progressively selected, accuracy (%) ± standard deviation.

TABLE I
TOP 15 SELECTED MODEL-BASED PARAMETERS BY USING THE THREE METHODS.

Rank	MI method	Correlation method	ANOVA method
1	Ankle Width LWA	Leg Width 08	Leg Width 08
2	Leg Width 08	Leg Width 09	Leg Width 09
3	Hip Rotation Amplitude	Knee Width LWKU	Leg Width 04
4	Leg Width 01	Ankle Width LWA	Leg Width 05
5	Head dx HDX	Leg Width 07	Leg Width 07
6	Y Motion Amplitude Ay	Head Height HH	Leg Width 06
7	Torso Width TW	Knee Rotation angle(12)	Ankle Width LWA
8	Knee Width LWKU	Knee Rotation Mean	Knee Width LWKL
9	Knee Rotation angle(02)	Head dx HDX	Leg Width 03
10	Gait Frequency omega	Knee Width LWKL	Leg Width 01
11	Knee Rotation angle(11)	Knee Rotation angle(04)	Leg Width 00
12	Ankle Rotation angle(07)	Ankle Rotation angle(12)	Leg Width 02
13	Hip Rotation Mean	Knee Rotation angle(03)	Gait Frequency omega
14	Knee Rotation angle(13)	Leg Width 06	Y Motion Amplitude Ay
15	Leg Width 03	Ankle Rotation angle(06)	Head dx HDX

based methods. This is particularly significant when up to 35 features are selected. For example, the MI-based method achieves 90% correct recognition rate with 25 features, compared with 29 feature for the ANOVA-based method and 37 features for the correlation-based method. For the MI-based method, when more than 37 features (accounting for about 50% of all 73 features) are selected, the recognition rate then improves only slightly. This indicates that for this particular model-based feature set, about 50% parameters are almost redundant and can be removed without significant loss of recognition accuracy. When more and more features are selected, the differences between the three methods tend to be small. This is because at this stage, most of the essential features are already included. The results in Fig. 4 also suggest that the ANOVA-based method is generally better than the correlation-based method.

Table I further lists the names of the 15 top model-based features selected by the mutual information, the correlation and the ANOVA based methods, respectively. It is intuitively suggested that the features selected by the mutual information seem to contain the wider range of gait information.

The feature selection results for the model-free gait symmetry pictures are shown in Fig. 5. The line marked with 'O' are the results of the MI-based method, comparing with the correlation-based method (the line marked with '+') and the

ANOVA-based method (the line marked with '∆'). It is seen that for the case of the MI-based method, the recognition rate change little when more than 100 important symmetry pixels are selected, which compares to 150 important pixels selected by the ANOVA-based method. Moreover, the correlation-based method cannot reach the 95% accuracy rate that are achieved by both the MI and the ANOVA-based methods, after using about 200 pixels in the symmetry maps (these account for about 5% in all 64×64 pixles). These empirical findings are based on this specific dataset, so in real world applications noise on the symmetry maps should be taken into account.

From Fig. 5, it is seen that the cumulative recognition results on the model-free gait data are consistent with our previous findings in the model-based feature set (see Fig. 4), with the best recognition rate achieved by the MI-based method, followed by the ANOVA and the correlation based methods. These results confirmed the effectiveness of using MI as the feature selection metric to approximate the classification accuracy. On the other hand, the correlation coefficient can only measure the strength and direction of a linear relationship between two random variables. So it is weaker to be validly used to infer the prediction accuracy (i.e., no strong causal relationship between the variables), and the poorest results were obtained among the three methods.
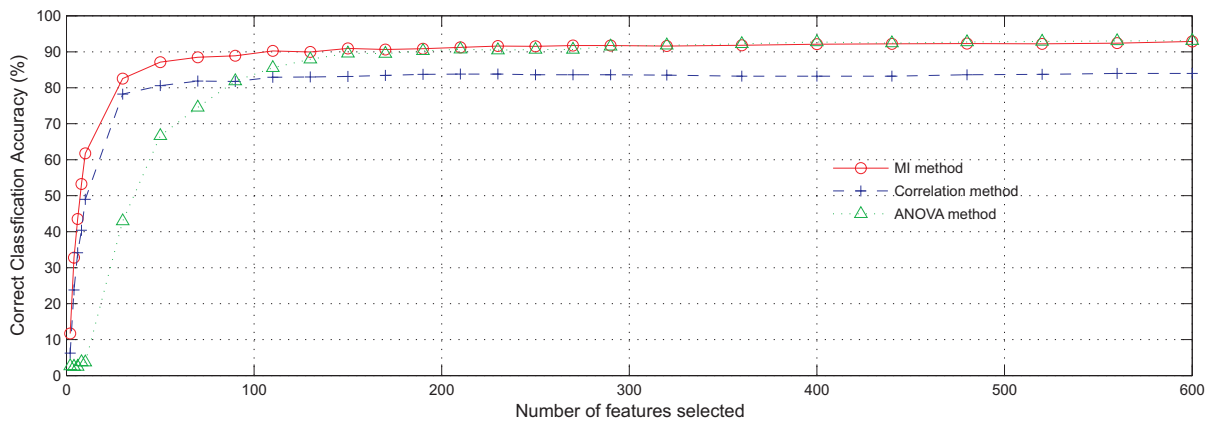
Fig. 5. Cumulative recognition rates as the model-free symmetry pixels are progressively selected.

Standing in the middle of the performances, the method based on the ANOVA was built on several assumptions, such as normality of the distributions and homogeneity of variances, which may reduce the robustness of the method if such assumptions did not hold.

V. CONCLUSIONS

In this paper, we presented a feature selection method for gait recognition based on mutual information analysis. This feature selection technique can effectively identify the most pertinent features for classification purposes, and improve system's efficiency accordingly. To assess the proposed method, experiments were carried out on a model-based gait feature set and a model-free symmetry gait data. The results showed that the MI-based method has good application capability in gait feature selection, with fewer features that can obtain good recognition accuracy. It also outperformed the correlation-based and the ANOVA-based methods. Further research is undergoing to explore the selected results to understand the underlying distinctness among human gaits.

Currently, the algorithms were tested on the indoor Southampton HiD Gait database, which was filmed in controlled laboratory conditions. To obtain more general results for different surface types, times, clothings, etc., future research will be carried out to test the algorithm on other higher-variations datasets, such as the HumanID gait dataset [2].

REFERENCES

[1] A. Kale, A. Sundaresan, A. N. Rajagopalan, N. P. Cuntoor, A. K. Roy-Chowdhury, V. Krger, and R. Chellappa. Identification of humans using gait. *IEEE Transactions on Image Processing*, 13(9):1163–1173, 2004.

[2] Z. Liu and S. Sarkar. Improved Gait Recognition by Gait Dynamics Normalization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(6):863–876, 2006.

[3] S. Sarkar, P. J. Phillips, Z. Liu, I. R .Vega, P. Grother, and K. W. Bowyer. The HumanID Gait Challenge Problem: Data Sets, Performance, and Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(2):162–177, 2005.

[4] M. S. Nixon and J. N. Carter. Automatic recognition by gait. *Proceedings of the IEEE*, 94(11):2013–2024, 2006.

[5] D. K. Wagg and M. S. Nixon. Automated markerless extraction of walking people using deformable contour models. *Computer Animation and Virtual Worlds*, 15(3-4):399–406, 2004.

[6] C. Yam, M. Nixon, and J. Carter. Automated person recognition by walking and running via model-based approaches. *Pattern Recognition*, 37(5):1057–1072, 2004.

[7] L. Wang, T. Tan, H. Ning, and W. Hu. Silhouette analysis-based gait recognition for human identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1505–1518, 2003.

[8] J. Hayfron-Acquah, M. S. Nixon, and J. N. Carter. Automatic gait recognition by symmetry analysis. *Pattern Recognition Letters*, 24:2175–2183, 2003.

[9] G. Veres, L. Gordon, J. N. Carter, and M. Nixon. What image information is important in silhouette-based gait recognition. In *Proceedings of IEEE Computer Vision and Pattern Recognition conference*, pages –, Washington, USA, 2005.

[10] G. Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, 14(1):55–63, 1968.

[11] A. Jain and D. Zongker. Feature selection: Evaluation, application, and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2):153–158, 1997.

[12] A. Blum and P. Langley. Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97(1–2):245–271, 1997.

[13] B. Bhanu and J. Han. Human recognition on combining kinematic and stationary features. In *Proceedings of the Fourth International Conference on Audio Visual Biometric Person Authentication, AVBPA03*, pages 600–608, 2003.

[14] A. Veeraraghavan, R. Chellappa, and A. Roy Chowdhury. Role of shape and kinematics in human movement analysis. In *Proceedings IEEE Computer Vision and Pattern Recognition 2004, CVPR04*, volume 1, pages 730–737, 2004.

[15] Z. Liu., L. Malave, A. Osuntogun, P. Sudhakar, and S. Sarkar. Toward understanding the limits of gait recognition. *Proceedings of SPIE*, 5404:195–205, 2004.

[16] M. S. Nixon, T. N. Tan, and R. Chellappa. *Human Identification based on Gait*, chapter 5, pages 45–104. Springer, 2005.

[17] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information medical imaging. *IEEE Transactions on Medical Imaging*, 16(2):187–198, 1997.

[18] B. Guo, S. R. Gunn, R. I. Damper, and J.D. B. Nelson. Band selection for hyperspectral image classification using mutual information. *IEEE Geoscience and Remote Sensing Letters*, 3(4):522–526, 2006.

[19] D. Wagg. *Local and Global Models for Articulated Motion Analysis*. PhD thesis, School of Electronics and Computer Science, University of Southampton,UK, 2006.

[20] B. E. Boser, I. M. Guyon, and V. N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pages 144–152, Pittsburgh, PA, 1992.

[21] C. Cortes and V. N. Vapnik. Support-vector networks. *Machine Learning*, 20(3):1–25, 1995.

[22] H. Abdi. *in N.J. Salkind (ed.), Encyclopedia of Measurement and Statistics*, chapter Coefficients of correlation, alienation and determination. Thousand Oaks, CA: Sage, 2007.