Modelling the time-variant covariates for gait recognition

Galina V. Veres, Mark S.Nixon and John N. Carter

Department of Electronics and Computer Science
University of Southampton, Southampton, SO17 1BJ

Abstract. This paper deals with a problem of recognition by gait when time-dependent covariates are added, i.e. when 6 months have passed between recording of the gallery and the probe sets. We show how recognition rates fall significantly when data is captured between lengthy time intevals, for static and dynamic gait features. Under the assumption that it is possible to have some subjects from the probe for training and that similar subjects have similar changes in gait over time, a predictive model of changes in gait is suggested in this paper, which can improve the recognition capability. A small number of subjects were used for training and a much large number for classification and the probe contains the covariate data for a smaller number of subjects. Our new predictive model derives high recognition rates for different features which is a considerable improvement on recognition capability without this new approach.

1 Introduction

Recently much attention has been devoted to use of human gait patterns as biometric. Gait recognition aims to discriminate individuals by the way they walk and has the advantage of being non-invasive, hard to conceal, being readily captured without a walker's attention and is less likely to be obscured than other biometric features. Approaches to gait recognition can be broadly classified as being model-based and model-free. Model-based methods [3, 5, 10] model the human body structure and extract image features to map them into structural components of models or to derive motion trajectories of body parts. Model-free methods [1, 2, 4, 23] generally characterise the whole motion pattern of the human body by a compact representation regardless of the underlying structure. In this paper we employ the model-based (dynamic) method of Wagg and Nixon [22] and the model-free (static) method of Veres et al [20].

However, in these works only databases recorded over a short time interval were evaluated. Some studies over a more lengthy time interval were reported for face recognition. In [19] images of 240 distinct subjects were acquired under controlled conditions, over a period of ten weeks. They showed that there was not a clearly decreasing performance trend over a period of ten weeks and concluded that reduction in degradation is small enough as to be nearly flat over this time period. Other studies have shown that over a period of years, face recognition performance degrades linearly with time [16]. Some studies were done to show effects of ageing on face recognition [6, 7, 8]. In [6] a systematic method for modelling appearance variation due to ageing is presented. It was shown that ageing variation is specific to a given individual, it occurs slowly and it is affected significantly by other factors, such as the health, gender and the lifestyle of the individual. Taking this into consideration, reasonably accurate estimates of age can be made for unseen images. In [7, 8] face identification experiments are presented, where the age of individuals in the gallery is significantly different than the age of individuals in the probe. It was demonstrated that automatic age simulation techniques can be used for designing face recognition systems, robust to ageing variation. In this context, the perceived age of subjects in the gallery and probe is modified before the training and classification procedures, so that ageing variation is eliminated. O'Toole et al. [14, 15] use three-dimensional facial information for building a parametric $3D$ face model. They use a caricature algorithm in order to exaggerate or deemphasize distinctive $3D$ facial features; in the resulting caricatures, the perceived age is increased or decreased according to the exaggeration level, suggesting that $3D$ distinctive facial features are emphasized in older faces. Some recent efforts [9] were made to improve age estimation by devoting part of the classification procedure to choosing the most appropriate classifier for the subject/age range in question, so that more accurate age estimates can be obtained.

In this paper we consider a gait recognition problem when two databases (the gallery and probe) were recorded with a time interval of 6 months between the finish of recording the first database (gallery) and the start of recording the second database (probe), i.e. time-dependent covariates are added. Moreover, some extra covariates were added in the second database such as different shoes, clothes, carrying different bags. In real life the need to analyse such databases arises in security of access to a company or an embassy for example. It is possible to record people walking normally as a gallery, but later it will be necessary to recognize these people in different clothes, shoes, possibly carrying luggage and when time passes. It is shown that in this case correct classification rates fall significantly and recognition becomes unreliable. Similar results were obtained for the HumanID Gait Challenge Problem [17], where recognition fell from 82% to 6% after 6 months. Some other recent works reported a significant fall in recognition capability over a lengthy time interval [11, 12, 13]. Under the assumptions that we can have records of people walking normally from the probe and similar people have similar changes in gait, the predictive model of gait changes is suggested in this paper as way to increase CCRs when analysis is needed over time. The predictive model is based on available records both from the gallery and the probe and a prediction matrix is constructed for these subjects. Then prediction matrix is generalised for all subjects in the gallery and predicted gallery is obtained. The probe is analysed via the predicted gallery and CCRs are calculated. To show robustness of the suggested approach the predictive model was appllied both on static and dynamic feature sets. We show that CCRs can be increased by several times when using the new predictive model.

Section 2 describes the suggested prediction model for changes in gait over lengthy time interval. The methodology of constructing feature sets is presented in Section 3. Experimental results are presented and described in Section 4. Section 5 concludes this paper.

2 Prediction of Time-variant Covariates

The idea of our approach was inspired by work of Lanitis et al [8] where it was shown that reasonably accurate estimates of age can be made for unseen images. In case of recognition by gait over a lengthy time interval we assume that it is possible to predict the gallery over the given time interval and achieve good recognition results by analysing the probe via the predicted gallery. In this case the training set consists of a set of subjects from the gallery and the same set of subjects from the probe. The probe was recorded some time later after finishing recording the gallery. In the general case the predicted gallery can be defined as

$$\hat{\mathbf{G}} = \mathbf{G} + \mathbf{Q}, \tag{1}$$

where $\hat{\mathbf{G}}$ is the predicted gallery, $\mathbf{G}$ is the gallery and $\mathbf{Q}$ is a prediction matrix.

Let the gallery and the probe be divided into groups, where a number of groups corresponds to a number of subjects and each group represents feature vectors for a given subject. Let us consider at first the case when the number of groups (subjects) in the gallery equals the number of groups (subjects) in the probe and the groups (subjects) are the same. At first we sort the records of the subjects according to their groups and note the number of records per subject. Then the prediction matrix $\mathbf{Q}$ is constructed as follows. At first for each group in the probe and gallery the mean of the group is calculated

$$\bar{\mathbf{x}}_p^j = \frac{1}{n_p^j}\sum_{i=1}^{n_p^j}\mathbf{x}_{pi}^j \quad \text{and} \quad \bar{\mathbf{x}}_g^j = \frac{1}{n_g^j}\sum_{i=1}^{n_g^j}\mathbf{x}_{gi}^j, \tag{2}$$

where $\bar{\mathbf{x}}_p^j$ is the mean of group j in the probe, $\bar{\mathbf{x}}_g^j$ is the mean of group j in the gallery, $j = 1, \ldots, n_g$, where n_g is a number of groups, n_p^j and n_g^j is a number of records in the jth group of the probe and of the gallery, respectively, $\mathbf{x}_{pi}^j$ and $\mathbf{x}_{gi}^j$ are records for ith subject in jth group in the probe and gallery respectively.

Then the prediction matrix for each group is calculated as

$$\mathbf{Q}^j = \mathbf{e}(\bar{\mathbf{x}}_p^j - \bar{\mathbf{x}}_g^j), \quad j = 1, \ldots, n_g, \tag{3}$$

where $\mathbf{e}$ is a positive unit vector of $(n_g^j \times 1)$, and the final prediction matrix is

$$\mathbf{Q} = [\mathbf{Q}^1; \mathbf{Q}^2; \ldots; \mathbf{Q}^{n_g}]. \tag{4}$$

For the more general case when a number of groups in the gallery is not the same as a number of groups in the probe and/or subjects are not the same both in the gallery and in the probe, the prediction matrix is constructed as follows. Here we present a case when a number of groups in the gallery is more than a number of groups in the probe. Two assumptions are made in this case

1. Every subject in the probe exists in the gallery.
2. The gait of the similar subjects will change in a similar manner with time.

We are looking forward to gathering more data to provide a theoretical analysis or statistical observation to support the second assumption.

The gallery and probe are divided into groups. The gallery is rearranged in such a way that the first n_g will be the groups which are in the probe, and the last n_{dg} groups are not in the probe. The number of groups in the gallery is $n_{gg} = n_g + n_{dg}$. Then the final prediction matrix is calculated as

$$\mathbf{Q} = [\mathbf{Q}^1; \mathbf{Q}^2; \ldots; \mathbf{Q}^{n_g}, \mathbf{Q}^{n_g+1}, \ldots, \mathbf{Q}^{n_{gg}}]. \quad (5)$$

The means of all groups in the probe and the gallery are calculated and prediction matrices for coincidental groups are taken as (3) in this case. The differences will be for groups in the gallery which do not exist in the probe. Further the calculation of the prediction matrices for such groups is presented. To be able to distinguish means of groups belonging only to the gallery from the means of the groups existing in the probe we will denote them as $\bar{\mathbf{x}}_g^{d_j}$, where $d_j = 1, \ldots, n_{dg}$. Then taking into consideration assumption 2, we can first compare the means of groups existing both in the probe and in the gallery with the means of groups existing only in the gallery. For each $\bar{\mathbf{x}}_g^{d_j}$ find the first nearest neighbour from $\bar{\mathbf{x}}_g^j$ by using formula

$$\text{find } k \text{ such as } \; k = j : \{\min_j |\bar{\mathbf{x}}_g^j - \bar{\mathbf{x}}_g^{d_j}|, \;\; j = 1, \ldots, n_g\}. \quad (6)$$

Then the predicted matrix for a given group is

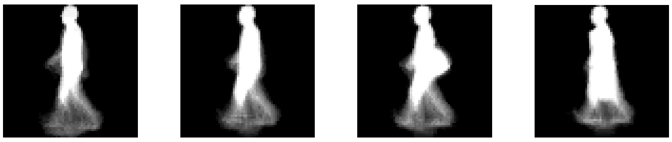
$$\mathbf{Q}^{n_g+d_j} = \mathbf{Q}^k + \mathbf{e}|\bar{\mathbf{x}}_g^k - \bar{\mathbf{x}}_g^{d_j}|, \quad (7)$$

where $\mathbf{e}$ is of $(n_g^k \times 1)$.

After the prediction matrix is obtained, the predicted gallery is calculated as (1) and the probe is classified via the predicted gallery. In some cases it is possible that the number of records per subject is much higher than the number of records used for calculation of prediction matrix. In this case the probe is divided into parts: one used for training and classification and second used only for classification. The suggested approach to predict changes of gait over lengthy time interval was applied to recognition by gait later in the paper.

3 Methodology

Two databases were analysed in the paper, both comprising indoor (studio) data since the purpose of this paper is to investigate a gait recognition problem. The first database, called the large database (LDB), consists of 115 subjects walking normally. The database arrangements are described elsewhere [18]. The LDB can be used to determine which image information remains unchanged for a subject in normal conditions and which changes significantly from subject to subject, i.e.



a) b) c) d)

Fig. 1. Average silhouette representing subject 46: a) LDB, b) SDB, walking normally, c) SDB, carrying a bag, d) SDB, wearing trench coat

it represents subject-dependent covariates. The small database (SDB) consists of a sample of 10 subjects from the LDB. Each subject was filmed wearing a variety of footwear, clothes and carrying various bags. They were also filmed walking at different speeds. In this case the covariate factor is more complicated than decribed in Section 2, since it does not only depend on the subjects, but on the other factors mentioned above. Each subject's data was captured during one continuous filming session. Examples of features from the LDB and the SDB is presented in Fig. 1. The figure shows subject 46's data recorded in the LDB and in the SDB walking normally, carrying a bag and wearing a trench coat.

The SDB is intended to investigate the robustness of biometric techniques to imagery of the same subject in various common conditions (carrying items, wearing different clothing or footwear). It worth noticing that sequence of filming data was LDB first and then SDB with approximately 6 months difference between recording LDB and SDB, i.e. time-dependent covariates are also added.

For brevity we shall not fully describe the extraction methods, complete explanations of the techniques used can be found in [20, 22]. These techniques yield two feature vectors for each database: the dynamic vector consists of 73 parameters describing joint rotation models of the limbs together with normalized information about the subject's speed and gait frequency; the static method consists of a 4096 dimensional vector derived from the subject's silhouette accumulated over a sequence of images.

4 Experimental Results

In this section we present the results of an experimental assessment of the performance of the suggested predictive model in the tasks of recognizing people over lengthy time interval by gait, i.e when SDB was analysied via LDB S/L, and of predicting the future changes in gait. The system was tested on five different conditions. For the first we performed leave-one-out experiments using all our training silhouettes. The training set consists of 10 subjects belonging both to LDB and SDB, i.e. S_{10}/L_{10}. This experiment tested the accuracy of our approach in prediction of gait changes over time in subjects who had already provided training silhouettes. In the second experiment we are testing the predictive model for a case when the testing set was 10 subjects from the SDB walking normally but recorded 1 hour later in comparision to the training set, i.e. $S_{10}(1)/L_{10}$. This experiments shows how robust the predictive model is to extra small time intervals, which were not taking into consideration during training. For the third experiment we added 105 new subjects to the gallery and tested how our approach will work when many more subjects are available for recognition in gallery than were used in training, i.e. S_{10}/L_{all}. In fourth experiment the training set was kept the same, but at the classification stage non time-dependent covariates are added to the probe, i.e. S_{all}/L_{10}. This test showed how adding extra covariates which were not used in training will affect the performances of the predictive model. The last experiment investigates the performance of the predictive model when only 10 subjects from the gallery and the probe are available for training, but 105 subjects are added to the gallery and 12 different experiments affecting a gait of a subject are added to the probe at classification stage, i.e. S_{all}/L_{all}.

To show the robustness of the suggested approach the prediction was applied to both the static *stat* and dynamic *dyn* features of gait. The reduced datasets were used for representing both static and dynamic part of gait, since it was shown [20] that recognition rates are little changed if a subset of the features is used. Different dimensionality reduction techniques were chosen for dynamic and static parts of gait due to their different performance on different datasets. The reduction techniques were applied only to training sets (not on whole databases) and choice of reduction technique was made accordingly, i.e. the best from the point of view of the training set. Static feature set dimensionality reduction was achieved by one-way ANOVA (analysis of variance), and the backwards feature selection algorithm [21] was used for the dynamic feature set. In case of dimensionality reduction by ANOVA the features are selected which satisfy the condition $i : F_i > k_1(\max(F) - min(F))$, where F is the F-statistic, and k_1 is a coefficient. In our case $k_1 = 0.2$ was chosen for the best classification results. And by correct classification rate, CCR is understood a correct classification rate obtained by by comparison SDB (to mean probe) via LDB (gallery) if not mentioned otherwise. The CCR was calculated using Euclidean distance as it is the most popular distance metric in gait recognition literature and the 1-nearest neighbour rule. To show the influence of time and not a degradation of data quality on recognition, the CCR is calculated for each database separately, i.e. the LDB is analysed via LDB (L_{all}/L_{all}) and SDB via SDB (S_{all}/S_{all}). The results are presented in Table 1 and show acceptable CCRs for both dynamic and static feature sets. The CCR's are consistent with the size of the database and we are able to recognise people, that means that both sets of data are good.

CCRs for original and reduced datasets when SDB via LDB analysed are presented in Table 2, i.e. when LDB is considered as the gallery and SDB is considered as the probe. A subject in the probe can wear normal shoes, clothes and walk normally or can walk slower/faster than normal, wear different shoes,

	L_{all}/L_{all}	S_{all}/S_{all}
stat	98.47%	99.90%
dyn	72.32%	90.24%

Table 1. Analysis of databases without time-dependent covariates

raincoat or even carry a bag/rucksack. We try to match this subject in the probe to a subject in the gallery who walks normally, wears normal shoes and normal clothes (no raincoat) and does not carry any bags. Training is done only for 10 subjects walking normally in both databases. Analysis is done when all features are taken into consideration in the probe and the gallery and when a reduced set of features is considered. In Tables 2 it can be seen that as soon as time-dependent covariates are added to analysis the fall in CCR is very noticeable. In some case the recognition rate approaches chance. The significant reduction in feature space is achieved without significant loss in CCR, and in some cases the visible increase in CCR can be seen. At the same time the CCRs are very low even in the best case especially when all subjects and all experiments are considered and something should be done to improve CCR. One of the ways is to use the suggested predictive model. It was noticing that if for static features the small time interval in experiments S_{10}/L_{10} and $S_{10}(1)/L_{10}$ does not cause noticable change in the recognition cability, in case of dynamic features the difference in CRR is almost 17% whcih can cause the noticable reduction in CCR on the testing set in comparison with the training set. However CCRs are affected not only by time-dependent covariates but adding extra subjects in the gallery and by adding non time-dependent covariates to the probe. In this paper we try to remove time-dependent covariates, but the influence of non time-dependent covariates will remain since it was assumed that only some records/subjects are available for training.

Dataset	number of features	S_{10}/L_{10}	$S_{10}(1)/L_{10}$	S_{10}/L_{all}	S_{all}/L_{10}	S_{all}/L_{all}
stat	4096	62.10%	64.93%	43.84%	47.61%	22.54%
stat	174	70.32%	68.24%	42.47%	51.73%	20.74%
dyn	73	19.18%	22.27%	8.68%	16.41%	5.30%
dyn	34	28.77%	33.65%	20.09%	28.81%	13.50%

Table 2. CCRs for static and dynamic features before training

Results of applying the predictive model to the training set and later to the whole databases are presented in Table 3. When only the training set is considered (S_{10}/L_{10}), 99.54% CCR is achieved for static features and 90.41% CCR for dynamic features which is expected as the predictive model was built on these sets and these numbers verify the linear model suggested for prediction in the paper. However, there is a drop in CCRs when experiment $S_{10}(1)/L_{10}$ is considered, i.e. when the testing and the training set recorded with time interval 1 hour. In the case of static features the drop can be considered insignificant and we can say that this experiment is verified the predictive model for static features. In the case of dynamic features it is almost a 10% drop, which can be explained by the lower CCR due to differences between the training set and testing set, see Table 2. However, for both static and dynamic features, CCRs are much higher after training than they were before training for this experiment. Moreover, when more subjects are added to the gallery, the predictive model can cope with this situation quite well. Only insignificant reduction in CCRs are presented for experiment S_{10}/L_{all} for both the dynamic and static feature sets. Also CCRs are decreasing with adding extra covariates to the training sets (more subjects, different experiments or both), the final results are much better than before training. When all subjects and experiments are analysed the CCR of 65.44% is achieved for static features which is practically three times more than was before training. In case of dynamic features the increase in CCR is almost four times which is very significant.

Dataset	S_{10}/L_{10}	$S_{10}(1)/L_{10}$	S_{10}/L_{all}	S_{all}/L_{10}	S_{all}/L_{all}
stat	99.54%	94.79%	96.35%	76.80%	65.44%
dyn	90.41%	82.46%	88.21%	62.66%	50.24%

Table 3. CCR for different datasets after applying predictive model

Table 3 shows that the suggested prediction approach is able to correct time-dependent variances and even produces good results when the gallery is much bigger than training set. However, it is not able to take into account the non time-dependent covariates in the probe. Therefore some extra efforts are needed to improve CCRs in such cases, which can be different predictive models, different classifiers, or fusion algorithms. We did not use sophisticated classifiers in this paper so we can verify the suggested linear model for the given problem.

5 Conclusions

This paper deals with a problem of increasing correct classification rate when time-dependent covariates (6 months passed between the finish of recording the gallery and the start of recording the probe) together with some other covariates such as variety of footwear, clothes and carrying different bags are added to an analysed database for gait recognition. We have shown that CCRs are very low in this case. In this paper we suggest to use the prediction of gait over the given time interval. One assumption made for the predictive model is that similar subjects will have similar changes in gait over a lengthy interval. The predictive model is based on estimation of differences between the means of subjects in the gallery and the probe and incorporation of these differences in prediction of the feature vectors in the gallery over the given time interval. Then the predicted gallery is compared with the probe and cross-validation is done. The experimental results showed that good results can be achieved both on the training set and the testing set and when extra subjects are added to the gallery and/or extra covariates are added to the probe. The predictive model allows good estimation of gait even when extra subjects are added to the gallery. However, the results are less impressive (though very good in comparison with the results achieved without prediction) when extra covariates are added to the probe such as wearing different shoes/clothes, wearing bags and walking faster/slower. To increase the CCR still further in such cases different approaches can be used: better predictive model, different classifiers or fusion.

Acknowledgments

We gratefully acknowledge partial support by the Defence Technology Centre $8-10$ supported by General Dynamics.

References

1. A.F. Bobick, and A. Johnson Gait extraction and description by evidence-gathering *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 423-430, 2001
2. R. Collins, R. Gross, and J. Shi Silhouette-based human identification from body shape and gait *Proceedings of the International conference on Automatic Face and Gesture Recognition*, Washington,DC, 2002
3. D. Cunado, M.S. Nixon, and J. N. Carter Automatic extraction and description of human gait models for recognition purposes *Computer Vision and Image Understanding*, **90**(1): 1-41, 2003
4. P.S. Huang, C.J. Harris, and M. S. Nixon Recognizing humans by gait via parametric canonical space *Artificial Intiligence in Engineering*, **13**(4): 359-366, 1999
5. A. Kale, N. Cuntoor, and R. Chellapa A framework for activity-specific human recognition *Proceedings of IEEE International conference on Acoustics, Speech and Signal Processing*, Orlando, Fl,, 2002
6. A. Lanitis, C.J. Taylor and T.F. Cootes Modeling the process of ageing in face images, *Proceedings of the Seventh IEEE International Conference on Computer Vision*, **1**: 131-136 , 1999.
7. A. Lanitis and C.J. Taylor Towards automatic face identification robust to ageing variation, *Proceedings of Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 391-396 , 2000.
8. A. Lanitis, C.J. Taylor, and T. F. Cootes Toward Automatic Simulation of Aging Effects on Face Images *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**(4): 442-455, 2002
9. A. Lanitis, C. Draganova and C. Christodoulou Comparing different classifiers for automatic age estimation, *IEEE Transactions on Systems, Man, and Cybernetics-PartB: Cybernetics* , **34**(1): 621-628, 2004.
10. L. Lee, and W.E.L. Grimson Gait analysis for recognition and classification *Proceedings of the IEEE International Conference on Face and Gesture*, 155-161, 2002

11. Z. Liu, and S. Sarkar Simplest Representation Yet for Gait Recognition: Averaged Silhouette, *Proceedings of the 17th International Conference on Pattern Recognition* , **2**: 704-711, 2004.
12. Z. Liu, L. Malave and S. Sarkar Studies on Silhouette Quality and Gait Recognition, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* , **4**: 211-214, 2004.
13. Z. Liu, and S. Sarkar Effect of Silhouette Quality on Hard Problems in Gait Recognition, Accepted for future publication *IEEE Transactions on Systems, Man, and Cybernetics- PartB: Cybernetics* , 2005.
14. A.J. O'Toole, T. Vetter, H. Volz and E. Salter Three-dimensional caricatures of human heads: distinctiveness and the perception of age, *Perception*, **26**: 719-732, 1997.
15. A.J. O'Toole, T. Price, T. Vetter, J.C. Bartlett and V. Blanz 3D shape and 2D surface textures of human faces: The 'Role'of 'averages in atractiveness and age' , *Image and Vision Computing*, **18**(1): 9-20, 1999.
16. P. Phillips, D. Blackburn, M. Bone, P. Grother, R. Micheals and E. Tabassi Face recognition vendor test 2002 (FRVT 2002), *Available at http://www.frvt.org/FRVT2002/default.htm*, 2002.
17. S. Sarkar, P.J. Phillips, Z. Liu, I.R. Vega, P. Grother, and K.V. Bowyer The HumanID Gait Challenge Problem: Data sets, Performances, and Analysis *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**(2): 162-177, 2005.
18. J. D. Shutler, M. G. Grant, M. S. Nixon, and J. N. Carter On a large sequence-based human gait database *Proc. 4th International Conf. on Recent Advances in Soft Computing*, Nottingham, UK,66-72, 2002
19. D.A. Socolinsky and A. Selinger. Thermal face recognition over time , *Proceedings of Proceedings of the 17th International Conference on Pattern Recognition*, **4**: 187-190, 2004.
20. G.V. Veres, L. Gordon, J.N. Carter and M.S. Nixon What image information is important in silhouette-based gait recognition? *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, D.C., USA, **II**: 776-782 2004.
21. G.V. Veres, L. Gordon, J.N. Carter, M.S. Nixon. Feature extraction or feature selection in silhouette-based gait recognition? Submitted to *Pattern Recognition.*
22. D.K. Wagg and M.S. Nixon Automated markless extraction of walking people using deformable contour models *Computer Animation amd Virtual Worlds*, **15**(3): 399-406, 2004.
23. L. Wang , W.M. Hu and T.N. Tan A new attempt to gait-based human identification *IEEE Transactions on Circuits and Systems for Video Technology*, **14**(2): 149-158, 2002.