

# Gait Recognition Using HMMs and Dual Discriminative Observations for Sub-dynamics Analysis

Nikolaos V. Boulgouris, *Senior Member, IEEE*, and Xiayi Huang

## Abstract

We propose a new gait recognition method that combines holistic and model-based features. Both types of features are extracted automatically from gait silhouette sequences and their combination takes place by means of a pair of HMMs. In the proposed system, holistic features are initially used for capturing general gait dynamics while, subsequently, model-based features are deployed for capturing more detailed sub-dynamics by refining upon the preceding general dynamics. Furthermore, holistic and model-based features are suitably processed in order to improve the discriminatory capacity of the final system. Experimental results show that the proposed method exhibits performance advantages in comparison to popular existing methods.

*keywords: gait, recognition, surveillance, biometrics*

## I. INTRODUCTION

The appeal of gait recognition [1] as biometric trait originates from its unobtrusiveness and its possibility of operation using image sequences that are of lower resolution than those required by other visual biometrics. On the other hand, gait is not as reliable as other biometric traits

Copyright (c) 2013 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to [pubs-permissions@ieee.org](mailto:pubs-permissions@ieee.org).

This work was supported in part by the European Commission under Contract FP7-215372 ACTIBIO.

N. V. Boulgouris is with the Department of Electronic and Computer Engineering, Brunel University, London, United Kingdom, ([nikolaos.boulgourisbrunel.ac.uk](mailto:nikolaos.boulgourisbrunel.ac.uk)).

X. Huang was with King's College, London, United Kingdom.

and, therefore, the development of efficient gait recognition systems still remains a formidable challenge and an area of active research.

Gait recognition methods can be broadly classified into two categories, i.e., those that use holistic features and others that use model-based features. Holistic methods [2], [3], [4] presume the availability of sequences of binary silhouettes representing walking individuals. Other such methods assume the availability of further information, e.g., depth [5]. Holistic methods have much less stringent requirements than model-based methods due to the fact that the accurate extraction of holistic features is usually simpler and easier. For example, some holistic methods use temporal silhouette averaging [6] while others (e.g., the method in [3]) use alternative templates calculated directly from the original silhouettes. The classification methods used in holistic approaches can be generic or tailored to the features used [7], [8], [9].

Model-based methods [10], [11], [12] assume a model for the human body and extract model parameters that are subsequently used in the recognition process. Such methods would generally be applicable only in cases the quality of gait video sequences is relatively good. Unlike holistic feature extraction methods, the application of which is simple, model-based methods rely on the availability of suitable algorithms for the construction of a human model and the efficient subsequent exploitation of model parameters in recognition. Attempts have been made towards the construction and exploitation of human models for gait recognition. In [11], the importance of each body component of a walking subject was investigated using the pre-labelled silhouettes from [10]. In addition, methodologies have been proposed for the efficient combination of results obtained using labelled body components. In [12], silhouettes were formed using a Layered Deformable Model (LDM) before the recognition process is applied.

Both holistic and model-based<sup>1</sup> approaches have advantages and disadvantages and have been used in the literature on a case by case basis, depending on the specific requirements of their applications scenarios. In this work, we exploit the advantages of both approaches by combining them using two Hidden Markov Models. A past work that used Hidden Markov Models was presented in [13]. That work did not use person specific HMMs but, instead, it deployed a population HMM model, which was used for the alignment of the frames of two

<sup>1</sup>In this section, the term “model” can refer to either a human body model or a Hidden Markov model. The meaning of the term should be easy to deduce from the context. In the rest of the paper, however, “model” will refer to HMM while features related to the human body model will be referred to as “labelled component features”.

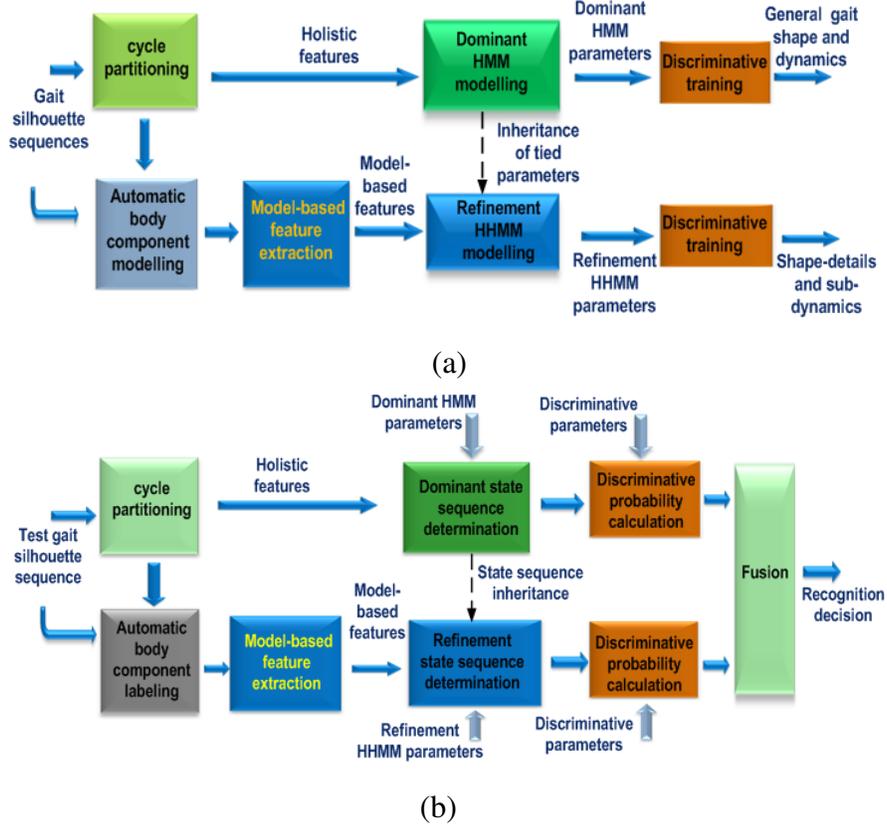


Fig. 1. Block diagram of the proposed algorithm: (a) training, (b) testing.

gait sequences rather than for recognition. A more recent work based on population HMMs for gait recognition was presented in [14]. In [15], a gait-based gender classification scheme was proposed based on the integration of shape appearance and temporal dynamics of both genders into a sequential model called mixed conditional random field (MCRF). In analogy to maximum likelihood decision used in HMMs, several classification strategies on the MCRF were assessed and discussed. Further, an HMM-based method for gait recognition was proposed in [16]. That method proposes an incremental framework based on optimal flow, dynamics learning, and recognition. The natural way of learning is simulated using a HMM, which represents the gait dynamics of a single subject and evolves incrementally from a population model that reflects the average motion process of human gait.

Other past works that have relied on HMMs for gait recognition were the methods in [17] and [18]. The features used by these methods are holistic and the associated HMMs are based on exemplars. Their classification approach is conventional, with the difference that, in [18], a

factorial HMM is used. Unlike these methods, the present work uses both types of features and makes the following contributions:

- An efficient method is introduced for the automatic labelling of body components in gait silhouettes and the extraction of body-related features from them.
- A novel method is proposed for gait recognition based on the combination of holistic and labelled body component features. Two distinct HMM models are used: a *dominant* (conventional) HMM that generates only holistic observations and a *refinement* Hierarchical HMM that generates only labelled component features that constitute detailed refinements of the holistic observations. In this way, the present system not only exploits holistically the general appearance of walking (as captured by raw binary silhouettes) but is also able to focus on details of walking style that are evident by processing human-model based features. In case labelled component features are not available, the system can operate efficiently using only the dominant HMM.
- A novel approach is proposed for improving the discriminatory capacity of the HMM-based scheme. Using this approach, the version of the features that is used for the estimation of HMM parameters is different from that used for recognition. Specifically, the features used for recognition are discriminative versions of the original features.

The proposed algorithm is experimentally assessed and its efficiency is shown in comparison to other popular methods.

The rest of this paper is organized as follows: Section II presents an overview of the proposed system. Sections III and IV present the automatic body-component labelling and the extraction of features respectively. Section V details the HMM-based modelling for gait recognition, using the features of the previous section. Section VI describes the proposed classification framework for gait recognition. Experimental results are presented in Section VII and, finally, conclusions are drawn in Section VIII.

## II. PROPOSED GAIT RECOGNITION SYSTEM

The first step in gait recognition is to detect the walking individual based on the originally available gait sequence of frames. Considering that human detection constitutes a separate research area by itself, in most gait recognition research it is presumed that silhouettes have been extracted from gait sequences using known methods for background subtraction [19]. For

this reason, some gait databases only provide silhouettes. Therefore, as usual in gait recognition research, we assume that binary silhouettes are available and we focus our efforts on improving recognition performance by efficient feature extraction and HMM-based gait modelling.

The proposed system is based on a *dominant* and a *refinement* HMM. The block diagram of the proposed system is shown in Fig. 1 and consists of a training and a testing module. In training, based on a set of scale-normalized gait silhouettes, body components are labelled using an automatic algorithm and three features are extracted from each labelled component in each frame. In order to train the dominant HMM, the original silhouettes themselves are used, while the refinement Hierarchical HMM is trained subsequently using the labelled component features. It should be noted that, as the general stages of gait can be far better modelled using the holistic silhouette feature, the upper level of the refinement HMM inherits its state transition probabilities from the dominant model. This resembles an architecture involving *tied states* [20], with the upper-level states in the refinement HHMM being tied to the states of the dominant HMM. Furthermore, a discriminative transform is applied and discrimination parameters are extracted in order to be used in the recognition process.

In the recognition stage, frames in a test gait sequence are automatically labelled and features are extracted from labelled components. The original silhouettes along with their associated labelled component features are assumed to be observations that are generated by their respective HMMs. Specifically, silhouettes are assumed to be observations generated by the dominant HMM while labelled component features corresponding to these silhouettes are assumed to be observations generated by the refinement HMM. In addition, the observations themselves (silhouettes and labelled component features), are different forms of the same raw data and their meaning is dual, depending on the HMM through which they are observed. The proposed algorithm will be described in detail in the following sections.

### III. AUTOMATIC BODY-COMPONENT LABELLING

The refinement HMM of our system relies on the availability of information about distinct body components in each silhouette. There are a few existing methods performing this task [12], [21]. The Layered Deformable Model (LDM) introduced in [12] is the most comprehensive approach for reconstructing body components based on both automatically-segmented silhouettes and manually-labelled silhouettes. However, the drawbacks of the LDM method are that it involves a

large number (22) of parameters while good performance can be achieved only when manually-labelled silhouettes are available. In this work, we propose a method for automatically labelling components in silhouettes.

Initially, we segment a gait cycle out of a gait silhouette sequence using the method proposed in [22], i.e., by constructing a signal representing the number of pixels in each silhouette, filtering the signal by taking into account its autocorrelation, and locating the “peaks” in the filtered signal. The silhouettes between the first and the third “peak” constitute one gait cycle. The subject in the silhouettes corresponding to those three “peaks” is in stride posture, and those silhouettes are used as the *key silhouettes* for the extraction of body component features.

Subsequently, assuming that the walking subject is facing the left side of the scene<sup>2</sup> the right arm and the right thigh of the subject are severely occluded. Therefore, in our body model, we only consider six body components - *head*, *torso*, *left arm*, *left thigh*, *left leg*, and *right leg*. In order to label the above six body parts, we extract seven parameters (see fig. 2) - torso width ( $l_{to}$ ), left upper arm angle ( $\theta_{ua}$ ), left upper arm length ( $l_{ua}$ ), left forearm angle ( $\theta_{fa}$ ), left forearm length ( $l_{fa}$ ), left thigh length ( $l_{th}$ ), and left thigh angle ( $\theta_{th}$ ). The above parameters are calculated using the steps detailed in Table I by taking into account human anatomy [23].

Taking the above seven parameters into account, a body block mask, as shown in Fig. 3(a), is constructed for each frame. In particular, with the knowledge of the torso width, a rectangular block for the torso area is formed while a trapeziform block for the left thigh is constructed based on the knowledge of  $l_{to}$ ,  $h_k$ ,  $l_{th}$  and  $\theta_{th}$ . If  $(x_1, y_1)$ ,  $(x_2, y_2)$ ,  $(x_3, y_3)$  and  $(x_4, y_4)$  denote the coordinates of the top-left, bottom-left, top-right and bottom-right corners of the trapeziform block respectively, the area indicating the left thigh is easily labeled as shown in Fig. 3(b).

Finally, by applying the block mask on the original gait silhouette, we can obtain the labelled silhouette with six automatically-labelled body components. In Fig. 4, some examples of the automatically labelled silhouettes are shown. As seen, the six body components are correctly labelled in all frames. Interestingly, even when a bag is carried, it is correctly excluded from the torso area, which is due to the accurate calculation of the torso’s width based on the average silhouette. It is also worth pointing out that, although the torso is occluded by the left arm in

<sup>2</sup>we take the mirror sequence in case it does not.

TABLE I  
STEPS FOR THE LABELLING OF BODY COMPONENTS

Step 1: Based on anatomy [23], the heights of a human's neck, shoulder, elbow, pelvis, and knee are taken as in Fig. 2(a).

Step 2: In order to calculate the width of the torso, we calculate the average silhouette,  $I_{avg}$ , of the subject. Within the upper torso area of the average silhouette, we locate the leftmost  $x_l$  and rightmost  $x_r$  column of that includes pixels with intensity larger than  $\alpha \cdot 255$ , i.e., satisfying  $p\{(x, y) | I_{avg}(x, y) > \alpha \cdot 255, x < x_l \text{ or } x > x_r\} = 0$ . In our experiments we set  $\alpha = 0.9$ . The **width of the torso** is calculated as  $l_{to} = x_r - x_l$ . This step is graphically depicted in Fig. 2(b). Furthermore, the width of the arm ( $h_a$ ) and the knee ( $h_k$ ) are set to equal a fixed proportion of the width of the torso, i.e.,  $h_a = \alpha_{at}l_{to}$  and  $h_k = \alpha_{kt}l_{to}$ . In our work we set  $\alpha_{at} = 0.35$  and  $\alpha_{kt} = 0.75$  based on experimentation.

Step 3: At shoulder level, the shoulder point of the left arm ( $x_s, y_s$ ) on that row is located by applying an offset equal to  $l_{to}/4$  from the silhouette centre. Similarly, the elbow point of the left arm ( $x_e, y_e$ ) is located at elbow level. Besides, the left-hand point ( $x_h, y_h$ ) is detected as the left-most point between elbow level and hand level. The locations of the above points are shown in Fig. 2(c) and (d).

Step 4: The **lengths of the left upper arm** and **the left forearm** can be calculated as  $l_{ua} = \sqrt{(x_e - x_s)^2 + (y_e - y_s)^2}$  and  $l_{fa} = \sqrt{(x_h - x_e)^2 + (y_h - y_e)^2}$ , respectively. The corresponding **angles** are calculated as  $\theta_{ua} = \arctan \frac{x_e - x_s}{y_e - y_s}$  and  $\theta_{fa} = \arctan \frac{x_h - x_e}{y_h - y_e}$ , respectively.

Step 5: By applying steps 3 and 4 on the three key silhouettes we obtain three sets of  $l_{ua}$ ,  $l_{fa}$ ,  $\theta_{ua}$  and  $\theta_{fa}$ . In order to get a robust estimate of these parameters, we take the mean value of those three  $l_{ua}(l_{fa})$  as the length of the upper arm (forearm) in all frames. The parameters  $\theta_{ua}$  and  $\theta_{fa}$  in the rest of the silhouettes of the gait cycle are calculated using linear interpolation.

Step 6: Using the same approach that was followed in steps 3 - 5, we locate the knee point and calculate **the length**  $l_{th}$  and **the angle**  $\theta_{th}$  of the left thigh in the three key silhouettes, and subsequently in all silhouettes of the gait cycle.

most frames, using the torso block information in the body block mask, we can easily retrieve the full torso area without any occlusion.

#### IV. FEATURE EXTRACTION FROM LABELLED SILHOUETTES

Feature extraction is a very important process in a gait recognition system. In this work, we investigate three main features that are suitable for capturing discriminatory gait information from sequences of automatically-labelled silhouettes. The three features that we extract are the *component area*, the *component centre*, and the *component orientation*. In this way, it is expected that differences in walking style will be reflected in at least one of the above features. The extraction of the features is detailed below.

1) *Component area:* The first feature that we consider is the area  $A_l$ , of each body component  $\mathcal{C}_l$ , i.e., the number of pixels  $\mathbf{x}$  in  $\mathcal{C}_l$ ,  $l = 1, \dots, L$ . These component areas are calculated for

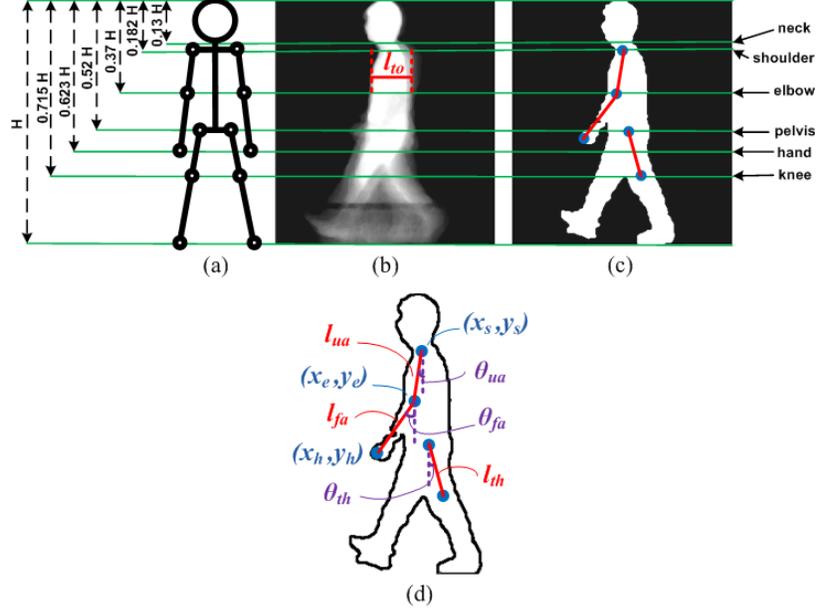


Fig. 2. The construction of the body model.

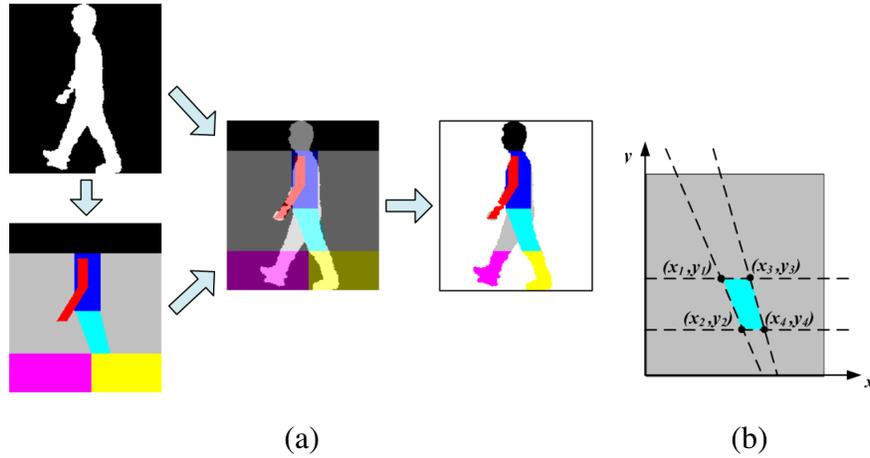


Fig. 3. (a) Labelling the body components based on the body block mask, (b) labelling left thigh using a trapeziform model.

each frame of a gait sequence, in order to obtain a feature vector, defined for the  $t$ th frame as:

$$\mathbf{f}_t^a = [A_{1t}, A_{2t}, \dots, A_{lt}, \dots, A_{Lt}]^T \quad (1)$$

where  $A_{lt}$  is the area for the  $l$ th component,  $l = 1, \dots, L$ , in the  $t$ th frame,  $t = 1, \dots, T$ . The superscript  $T$  denotes transposition.

2) *Component centre*: Due to the availability of labelling information for each pixel, it is possible to calculate the gravity centre  $\mathbf{g}_l$  of the  $l$ th body component. After calculating the



Fig. 4. Examples of automatically-labelled gait silhouettes.

gravity centres for all components, we calculate the vector distances  $\mathbf{v}_l = \mathbf{g}_l - \mathbf{g}$  between each one of the body component centres  $\mathbf{g}_l$  and the centre  $\mathbf{g}$  of the entire silhouette.

Subsequently, similar to Eq. (1), a feature vector,  $\mathbf{f}_t^v$ , is calculated based on the relative distance between the gravity centre of each body part to the centre of the whole body. This feature vector is defined for the  $t$ th frame as:

$$\mathbf{f}_t^v = [\mathbf{v}_{1t}, \mathbf{v}_{2t}, \dots, \mathbf{v}_{lt}, \dots, \mathbf{v}_{Lt}]^T \quad (2)$$

3) *Component orientation*: Although the features described above capture important discriminatory information, there is additional information that can further improve the accuracy of our gait representation. In order to capture the structure of the silhouettes in a more accurate way, we extract a third feature - the *orientation* of each body component. The calculation of the orientation feature takes place in two steps. We calculate the covariance of the pixel coordinates in each component followed by the application of Principal Component Analysis (PCA) [24] in order to find the principal orientation of each component. Similar to Eq. (1), a feature vector  $\mathbf{f}_t^o$  is calculated, defined for the  $t$ th frame as:

$$\mathbf{f}_t^o = [\mathbf{o}_{1t}, \mathbf{o}_{2t}, \dots, \mathbf{o}_{lt}, \dots, \mathbf{o}_{Lt}]^T \quad (3)$$

where  $l, t$  are component and frame indices respectively, with  $|\mathbf{o}_{lt}| = 1$ .

## V. GAIT REPRESENTATION USING A DOMINANT AND A REFINEMENT HMM

### A. Motivation

After having extracted the features presented in the previous sections, we use Hidden Markov Models for gait representation, training, and classification based on gait dynamics and sub-dynamics. In order to determine the HMM framework to be used, we consider that these two

types of gait dynamics are better captured by two different types of features since coarse holistic features cannot be used for determining detailed sub-dynamics while detailed features cannot be used for determining general gait dynamics. This necessitates that *first* general dynamics are modelled and *subsequently* detailed shape/sub-dynamics are modelled based on the previously modelled general gait dynamics. In this context, having a single HMM structure emitting one type of features would be inefficient while an HMM emitting both types of features would be highly unconventional and difficult to analyze. Instead, we use two Hidden Markov Models (fig. 5), each generating a different type of features (i.e., holistic or labelled component features):

- The first HMM, called the *dominant* HMM is a conventional HMM generating the holistic observations. This model is meant to represent the general shape and dynamics of gait and this is why it *relies exclusively on holistic observations*, which are most suitable for the purpose of modeling the walking style of a person *as a whole*.
- The second HMM, called the *refinement* HMM, is a two-level Hierarchical HMM [25] the lower level of which is generating detailed observations in the form of labelled body component features, extracted as described in the previous sections. The rationale for using a refinement HMM is that it can use the detailed labelled body component features in order to refine the general shape/dynamics that were captured previously through the dominant HMM. In this way, *further accuracy* is achieved by taking into account the detailed structure and motion of body components.

Several approaches for the actual implementation of the above HMMs can be conceived. One option would be to use semi-HMMs [26]. But in this case the number of model parameters would be very large and almost impossible to determine using the training gait data that are normally available. For this reason, we chose a simpler HMM modelling approach based on the two proposed HMMs shown in Fig 5.

The detailed labelled component features are formed by extracting (from each frame) six body components each of which is represented by the three features (area, centre, orientation) of the previous section. These more detailed features are compactly represented in a feature vector

$$\mathbf{f}_t = [\mathbf{f}_t^a \quad \mathbf{f}_t^v \quad \mathbf{f}_t^o]^T \quad (4)$$

which is a concatenation of the vectors representing each of the three features. Note that each of the three vectors in eq. (4) includes the values of the respective feature for each body component.

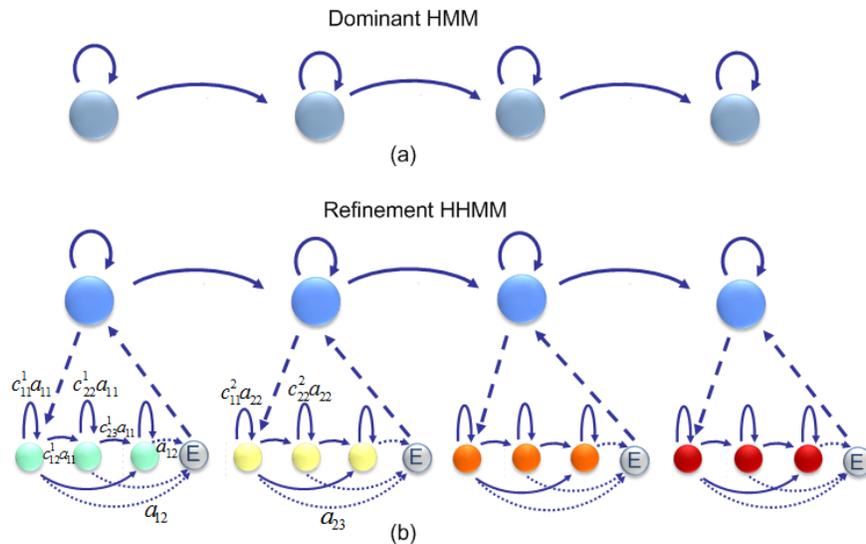


Fig. 5. The two HMMs used for gait recognition (four states instead of the actual five are shown). (a) The *dominant* HMM, which represents general shape and dynamics of gait using holistic features (observations). All states generate observations. (b) The *refinement* Hierarchical HMM that represents shape details and sub-dynamics of gait using detailed labelled component features. Note that only states at the lower level of the refinement HMM generate observations. The states labelled “E” in the lower level of the refinement HMM are “end” pseudo-states and do not generate observations. The “end” states indicate a return to the upper level and a transition to the next upper-level state. For this reason, transition to an “end” state (indicated using dotted arrows) practically means a transition to the first lower-level state of the next HMM.

These features do not exhibit the same changes throughout a gait cycle or within the same walking stage. For example, the value of the area feature is constant when the body part is visible but becomes less important when the body part is partly occluded.

Essentially, the detailed lower-level observations, i.e., labelled component features, are obtained by taking a closer look at the holistic observations. In this sense, the observation of a gait sequence has a *dual* interpretation based on the level of detail at which observations are studied. The duality of gait observations means that each frame in a gait sequence can be conceived of as being generated by a state of the dominant HMM, with the associated probability determined *solely* based on holistic observations, while the detailed body component features extracted from each frame are assumed to have been generated (in a more detailed way) by a state at the lower level of the refinement HHMM. It should be noted that, since not all lower-level states need to be visited, all possible upper-level state sequences can lead to feasible lower-level state sequences. As in all HMM-based gait recognition research, left-to-right HMMs are used, a

choice which is aligned with the physiological process of walking. The advantage of the proposed approach is that it combines holistic and model-based gait representation and recognition within a common framework that is consistent with the process of walking.

### *B. Gait representation*

Consider a sequence of dual gait observations. The sequence of holistic observations, generated by the dominant HMM, is

$$\mathbf{H} = \mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3, \dots, \mathbf{h}_T$$

where  $\mathbf{h}_t, t = 1, \dots, T$  is a vector extracted holistically from each silhouette in the gait sequence. In this work, the binary silhouettes themselves, rearranged as vectors, are the holistic feature vectors used for each frame. The sequence of labelled component features (observations), generated by the states at the lower level of the refinement HHMM, is

$$\mathbf{F} = \mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \dots, \mathbf{f}_T$$

where  $\mathbf{f}_t, t = 1, \dots, T$  is a vector, defined as in eq. (4), that includes the body-component features of the frame at time  $t$ . The two main tasks associated with the above formulation for gait recognition are the determination of HMM parameters in both dominant and refinement HMMs (training) as well as the process for making recognition decisions (testing). In order to make the above tasks computationally tractable, we follow an exemplar-based approach, similar to that followed in [17], [13], [18]. This implies that each state that generates observations (in the dominant HMM or the lower level of the refinement HMM) is represented by an exemplar and the probability that an observation is generated by that state is a function of the distance between the observation and that state's exemplar. As will be explained in the next section, despite some limitations, this approach is suitable due to the high dimensionality of the input data and the limited availability of training samples. Since the probabilities of observations being generated by HMM states depend on the exemplars, the HMM parameters that need to be determined are the initial state probabilities, the state transition probabilities, and the exemplars. The way to determine these parameters is presented below. All notation that is used in our analysis below is summarized in Table II.

TABLE II  
NOTATION

Symbol	Dominant HMM
$T$	number of frames in gait sequence (half-cycle)
$N$	number of states
$n$	state index
$\pi_n$	initial state probability
$A$	state transition probability matrix
$\mathbf{e}_n$	exemplar for the $n$ th state
$\lambda$	dominant Hidden Markov Model parameters
$\mathbf{Q}$	state sequence
$q_t$	state variable at time $t$
$\mathbf{H}$	sequence of holistic gait observations
$\mathbf{h}_t$	holistic observation at time $t$
$a_{q_1 q_2}$	probability of transition from state $q_1$ to state $q_2$
$b_{q_t}(\mathbf{h}_t)$	probability of holistic observation $\mathbf{h}_t$ from state $q_t$ at time $t$
Refinement HHMM	
$N$	number of HMMs in the lower level
$M$	number of states in each lower HMM (excl. "end" states)
$m$	state index for the lower HMMs
$\zeta_{mn}$	initial state probability for $m$ th state of the $n$ th lower HMM
$C_n$	state transition probability matrix for the $n$ th lower HMM
$\mu_n$	lower HMM associated with the $n$ th upper state
$\mu$	$\{\mu_1, \mu_2, \dots, \mu_N\} \cup \{\pi, A\}$
$\mathbf{S}_n$	state sequence in the $n$ th lower HMM
$s_t^n$	state variable at time $t$ for the $n$ th lower HMM
$\mathbf{F}$	sequence of gait features based on a human model
$\mathbf{f}_t$	observation at time $t$ for lower HMM
$c_{s_1 s_2}^n$	probability of transition from state $s_{n_1}$ to state $s_{n_2}$ in the $n$ th HMM of the lower level
$\mathbf{u}_{m^n}$	exemplar for the $m$ th state of the $n$ th lower-level HMM
$g_{s_t}^n(\mathbf{f}_t)$	probability of observation $\mathbf{f}_t$ from state $s_t^n$

### C. Definition and training of the dominant HMM

The calculation of the parameters of the dominant HMM is straightforward and follows that in [17]. The parameters of the dominant HMM will be denoted  $\lambda = \{\pi, A, e\}$ , where  $\pi$  denotes

the *initial state probabilities*,  $A$  denotes the matrix of *state transition probabilities*, and  $e$  is the set of *state exemplars*. Since the order with which poses are assumed during walking is defined naturally, a left-to-right HMM is considered, the initial state probabilities of which are

$$\pi_n = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

where  $\pi_n$  is the initial state probability for the  $n$ th state,  $n = 1, \dots, N$ , of the dominant HMM.

The probability that the holistic observation  $\mathbf{h}_t$  at time  $t$  is generated by state  $q_t$  is denoted  $b_{q_t}(\mathbf{h}_t)$ . Since feature vectors (silhouettes) are of very high dimensionality, the calculation of the output density function would be very difficult using the limited number of available training samples. For this reason, the output probability distribution is calculated, similar to [17], [13], [18], based on the distance of  $\mathbf{h}_t$  from the state exemplar  $\mathbf{e}_{q_t}$

$$b_{q_t}(\mathbf{h}_t) = P(\mathbf{h}_t | \mathbf{e}_{q_t}) = V e^{-vD(\mathbf{h}_t, \mathbf{e}_{q_t})} \quad (6)$$

where  $V$  and  $v$  are normalization parameters. The parameters of the dominant HMM are calculated using expectation-maximization. Exemplars are calculated as in [17], in which the inner product distance was used for this calculation as it was found to be very efficient.

#### D. Definition and training of the refinement Hierarchical HMM

Once the dominant HMM has been trained using holistic observations, we proceed with the calculation of the parameters of the refinement Hierarchical HMM. In this case, observations (labelled component features) are generated by lower-level states of the refinement Hierarchical HMM. However, these observations are meant to capture gait sub-dynamics by further analyzing the general gait dynamics observed at the dominant HMM. For this reason, we assume that the initial state probabilities and the transition probabilities in the upper level of states of the refinement HHMM are inherited from the dominant HMM as they refer to general dynamics of gait, which are expressed best in the holistic features. This approach, which resembles *parameter tying* [20], acknowledges the fact that *general* gait dynamics should be the same regardless of whether they are observed in the dominant HMM or the upper level of the refinement HHMM. This simplifies parameter calculation and makes it more robust in our case, where *modelling general gait dynamics is more reliable based on holistic features rather than detailed labelled component features* (see caption of fig. 5).

Considering the above formulation, what remains to be determined in the refinement HHMM is the set of model parameters  $\mu_n$ ,  $n = 1, \dots, N$  of the HMMs comprising the lower level of the refinement HHMM. Henceforth, the parameters of the  $n$ th lower-level HMM will be denoted  $\mu_n = \{\zeta_n, C_n, u_n\}$ ,  $n = 1, \dots, N$ , while the ensemble of parameters of the refinement HMM will be denoted  $\mu = \{\mu_1, \mu_2, \dots, \mu_N\} \cup \{\pi, A\}$ . Note that  $\mu$  includes the upper-level initial state probabilities and transition probabilities, which have been inherited from the dominant HMM. It should also be noted that the last state in each lower-level HMM is an “end” state that does not generate observations. Such states essentially lead to the first lower-level state of the next HMM. Therefore, transition to the “end” state takes place whenever an upper-level transition takes place. As a result, the probabilities of transition to an “end” state are identical to the probability of an upper-level transition from the current upper-level state to the next. This is why these probabilities do not have to be calculated and the respective transitions are indicated with dotted line in fig. 5(b).

Since left-to-right models are used, the initial state probabilities for the lower level of the refinement HMM are

$$\zeta_{mn} = \begin{cases} 1 & \text{if } m = 1 \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

where  $\zeta_{mn}$ ,  $m = 1, \dots, M$  is the initial state probability for the  $m$ th state of the  $n$ th HMM at the lower level of the refinement HMM.

Similar to [17], [13], [18], the probability that the lower-level observation  $\mathbf{f}_t$  at  $t$  is generated by the lower-level state  $s_t^n$  is calculated using the distance of  $\mathbf{f}_t$  from the state’s exemplar  $\mathbf{u}_{s_t^n}$

$$g_{s_t^n}(\mathbf{f}_t) = P(\mathbf{f}_t | \mathbf{u}_{s_t^n}) = \tilde{V} e^{-\tilde{v} D_*(\mathbf{f}_t, \mathbf{u}_{s_t^n})} \quad (8)$$

where  $\tilde{V}$  and  $\tilde{v}$  are normalization parameters. The distance  $D_*$  in eq. (8) involves comparison between labelled component features and exemplars and has been applied using the method in [27].

The parameter calculation for the refinement lower-level HMMs is based on the observations  $\mathbf{F}$ . Since the initial state probabilities are as in eq. (7), the parameters that are to be determined are the state-transition probabilities as well as the exemplars for each of the HMMs at the lower level of the refinement HMM. In order to calculate these probabilities, we use the forward

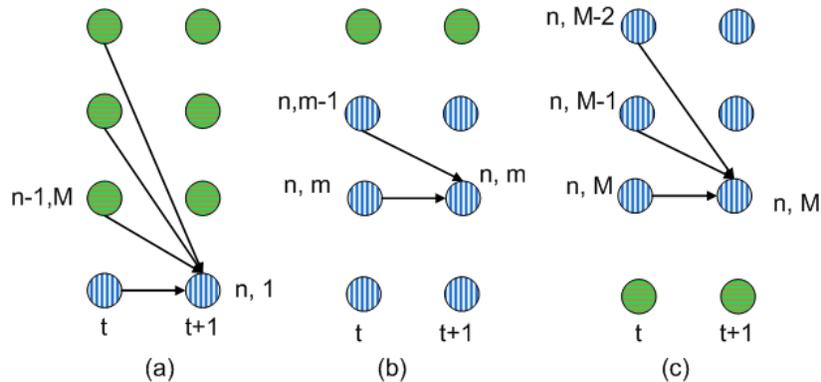


Fig. 6. Three cases during probability calculation in the refinement HMM. (a) transition to the first state of the  $n$ th HMM, (b) transition within the  $n$ th HMM, (c) transition to the last state of the  $n$ th HMM. Note that the “end” states of fig. 5 do not need to be shown here.

variable  $\alpha$  defined as

$$\alpha_t(n, m) = P(\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_t, s_t^n = m | \mu)$$

which represents the probability that the partial observation sequence  $\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_t$  is observed and the observation  $\mathbf{f}_t$ , at time  $t$ , is generated by the  $m$ th state of the  $n$ th lower-level HMM.

In order to calculate the forward variable for each time instant  $t$ , we consider that, at each given state in the lower level of the refinement HMM there are three cases at a specific time instant  $t$ . These are explained below:

- The state generating the observation at time  $t + 1$  is the first state of the  $n$ th lower-level HMM. This means that the transition was either from the same state of the  $n$ th HMM or from one of the states of the previous HMM (fig. 6(a)).
- The state generating the observation at time  $t + 1$  is the *same* or *one of the previous* states in the *same* HMM that generated the observation at  $t$  (fig. 6(b)).
- The state generating the observation at  $t + 1$  is the *last* state of the  $n$ th HMM (fig. 6(c)). Essentially, this is a subcase of the second case above.

Based on the above, the forward variable can be calculated recursively as follows (see fig. 7):

1) Initialization. Considering that  $\zeta_{1,1} = 1$ ,

$$\alpha_1(n, m) = \begin{cases} g_{s_1^n}(\mathbf{f}_1), & \text{if } n = 1, m = 1 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

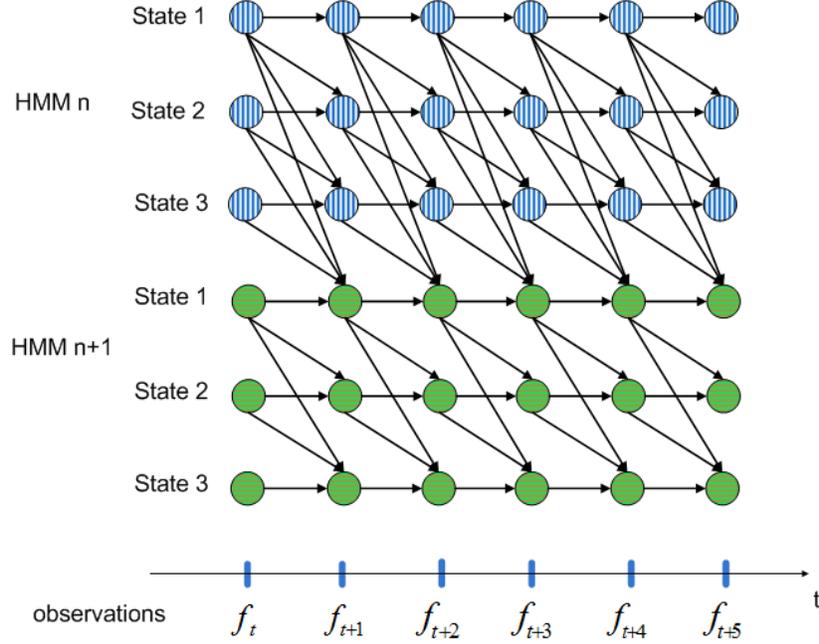


Fig. 7. Trellis for probability calculation. Arrows indicate allowable transitions between lower-level HMM states.

This is due to the fact that the first observation *has* to be generated by the first lower-level state of the first HMM.

2) Induction. Based on the three cases depicted in fig. 6, the forward variable at  $t + 1$  is

$$\alpha_{t+1}(n, m) = \begin{cases} \left( \alpha_t(n, m) a_{nn} c_{1,1}^n + \sum_{\tilde{m}=1}^M \alpha_t(n-1, \tilde{m}) a_{n-1,n} \right) g_{s_{t+1}^n}(\mathbf{f}_{t+1}) & \text{if } m = 1 \\ \left( \sum_{\tilde{m}=1}^m \alpha_t(n, \tilde{m}) c_{\tilde{m},m}^n \right) a_{nn} g_{s_{t+1}^n}(\mathbf{f}_{t+1}) & \text{if } m = 2, \dots, M \end{cases} \quad (10)$$

for  $t = 1, 2, \dots, T - 1$ , where  $a$  and  $c^n$  are, respectively, state transition probabilities for the upper and the  $n$ th lower-level model in the refinement HMM. It should be noted that  $c_{MM}^n = 1$  since transition from the last state of each lower-level HMM is produced only through a state-transition at the higher level.

Similarly, a backward variable can be defined as:

$$\beta_t(n, m) = P(\mathbf{f}_{t+1}, \mathbf{f}_{t+2}, \dots, \mathbf{f}_T | s_t^n = m, \mu)$$

which represents the probability of observing the partial observation sequence  $\mathbf{f}_{t+1}, \mathbf{f}_{t+2}, \dots, \mathbf{f}_T$  while at time  $t$  the upper-level state is  $n$  and the lower-level state is  $m$ . Using the forward

and backward variables, the transition probabilities are calculated through the use of variable  $\xi$ , calculated as

$$\begin{aligned}\xi_t^n(m, m') &= P(s_t^n = m, s_{t+1}^n = m' | \mathbf{F}, \mu) = \\ &= \frac{P(s_t^n = m, s_{t+1}^n = m', \mathbf{F} | \mu)}{P(\mathbf{F} | \mu)} = \\ &= \frac{\alpha_t(n, m) a_{nn} c_{mm'} g_{m'}^n(\mathbf{f}_{t+1}) \beta_{t+1}(n, m')}{P(\mathbf{F} | \mu)}.\end{aligned}\quad (11)$$

The total number of transitions, at time  $t$ , from the  $m$ th state of the  $n$ th HMM to the  $m'$ th state of the  $n$ th HMM is

$$\gamma_t^n(m) = \sum_{m'=1}^M \xi_t^n(m, m'), \quad n = 1, 2, \dots, N, \quad m = 1, 2, \dots, M \quad (12)$$

Using equations (11) and (12), the transition probability from state  $m$  to state  $m'$  within the  $n$ th HMM is naturally calculated as the ratio of the total number of transitions from state  $m$  to state  $m'$  over the total number of transitions from state  $m$ :

$$c_{mm'}^n = \frac{\sum_{t=1}^T \xi_t^n(m, m')}{\sum_{t=1}^T \gamma_t^n(m)}, \quad m = 1 \dots, M-1, \quad n = 1 \dots, N \quad (13)$$

It should be noted that transitions to “end” states are not counted in eqs (11), (12), and (13), as these are effectively transitions to the next HMM. The resultant transition probabilities  $c_{mm'}^n$  within the same HMM as well as the way they are combined with upper-level transition probabilities  $a_{nn'}$  are shown in fig. 5(b).

Once the state transition probabilities have been updated, each vector in the sequence of feature vectors is classified to the state that most likely generated it. Subsequently, the exemplars for each state are re-calculated as the averages of the feature vectors classified to that state. Using the state transition probabilities as well as the newly-updated exemplars, a new iteration of the preceding steps takes place, which further refines state transition probabilities (using eqs. (9)-(13)) and exemplars. A few iterations are sufficient for convergence.

Each gait cycle in a gait sequence consists of two half-cycles that may or may not be similar (gait is not always symmetric). In our system this means that two pairs of HMMs, a dominant/refinement HMM pair for each half-cycle, need to be built for each reference subject. As will be explained at the end of the next section, explicit labelling of the half-cycles is not needed during classification.

## VI. CLASSIFICATION

Recognition when HMMs are used is usually achieved by determining the reference model that maximizes the probability of the observations. In our system, we use two HMMs and therefore we can calculate two probabilities, one for each type of observations. In this work, we fuse the two probabilities into a single similarity measure by forming their product. If  $z^*$  is the index of the recognized subject, then

$$z^* = \arg \max_z P(\mathbf{H}|\lambda^z)P(\mathbf{F}|\mu^z) \quad (14)$$

where  $\lambda^z$  are the dominant HMM parameters and  $\mu^z$  are the refinement HMM parameters for the  $z$ th subject in the reference database. The probability terms in the right-hand side of eq. (14) can be calculated using standard HMM theory. In the sequel, we discuss ways for the application of eq. (14) in our gait recognition framework. The present approach partly follows that in [15], where several classification methods were applied on mixed conditional random fields for gender classification.

### A. Recognition using original features

Given the HMM parameters in both the dominant and refinement HMMs of a reference subject, the probabilities that the associated observation (feature) sequences from a test subject are generated by the given reference pair of models is straightforward. Specifically, following the derivations in [20], the probability that observations  $\mathbf{H}$  are generated by the dominant HMM model  $\lambda$  is

$$P(\mathbf{H}|\lambda) = \sum_{all \mathbf{Q}} P(\mathbf{H}, \mathbf{Q}|\lambda) \quad (15)$$

where  $\mathbf{Q}$  is a sequence of states in the dominant HMM that can generate the observations. The probability in eq. (15) can be trivially calculated using the forward (or backward) algorithm in [20]. Similarly, the probability  $P(\mathbf{F}|\mu)$  in the refinement HMM is calculated like in a conventional two-layer Hierarchical HMM, in which only the terminal states (i.e., states in the lower level) are emitting observations. Based on the above calculation of  $P(\mathbf{H}|\lambda)$  and  $P(\mathbf{F}|\mu)$ , we define

$$P_1 = P(\mathbf{H}|\lambda)P(\mathbf{F}|\mu) \quad (16)$$

This probability is calculated as above in order to assess the performance of the combination of the independent application of the two features sequences. In this case, the index of the recognized subject will be determined using eq. (14).

In eq. (16), the holistic and non-holistic features are used independently and sub-dynamics (taken into account through  $P(\mathbf{F}|\mu)$ ) are determined using only the detailed labelled component features. Essentially, this means that “details” are determined without considering the “general picture”. However, the labelled component features are not meant to act alone but rather to refine the general gait dynamics, determined previously through the holistic features. In order to take this fact into account, we tested a scheme in which the refinement HHMM inherits its upper level state sequence from the dominant HMM. This leads to a modified probability

$$P_2 = P(\mathbf{H}|\lambda)P(\mathbf{F}|\hat{\mathbf{Q}}, \mu) \quad (17)$$

where  $\hat{\mathbf{Q}}$  is the most probable state sequence  $\hat{\mathbf{Q}}$ , calculated through the dominant HMM:

$$\hat{\mathbf{Q}} = \arg \max_{\mathbf{Q}} P(\mathbf{H}, \mathbf{Q}|\lambda) \quad (18)$$

Using eq. (17), the holistic observations contribute to the subsequent interpretation of the labelled component features and the determination of sub-dynamics. For this reason, eq. (17) yields roughly equal performance as eq. (16) and may be a good practical choice due to its marginally simpler implementation.

### *B. Recognition using discriminative probability calculation*

The original features extracted from the gait sequences, namely the silhouettes and the labelled component features, are suitable for modelling general gait dynamics and sub-dynamics by means of the dominant and refinement HMMs that generated these features. In the recognition stage, however, using discriminative versions of the original features can yield significant gains.

In order to highlight subtle discriminative differences in the dominant HMM, we apply Fisher’s linear discriminant analysis (LDA) [24] on the exemplars corresponding to each state (walking stage) in the Gallery (training) set. This means that, although the original exemplars are used when HMM parameters and probabilities are determined, for recognition purposes we project the test features and reference exemplars to a low-dimensional discriminative space and calculate

probabilities based on the projected features. In particular, a matrix  $\mathcal{W}_n$  is calculated for each state  $n$  in the dominant HMM, such that the Fisher's criterion is maximized:

$$J(\mathcal{W}_n) = \frac{|\mathcal{W}_n^T \cdot \mathcal{S}_{bn} \cdot \mathcal{W}_n|}{|\mathcal{W}_n^T \cdot \mathcal{S}_{wn} \cdot \mathcal{W}_n|} \quad (19)$$

where  $\mathcal{S}_{bn}$  is the between-class scatter matrix, and  $\mathcal{S}_{wn}$  is the within-class scatter matrix for the  $n$ th HMM state. In this way, discriminative features are produced and, instead of using eq. (6), a new discriminatory distance measure is defined as

$$\tilde{D}(\mathbf{h}_t, \mathbf{e}_n) = D(\tilde{\mathbf{h}}_t, \tilde{\mathbf{e}}_n) = D(\mathcal{W}^T \mathbf{h}_t, \mathcal{W}^T \mathbf{e}_n) \quad (20)$$

where  $\tilde{\mathbf{h}}$  and  $\tilde{\mathbf{e}}$  are the projected holistic features and exemplars. For the application of LDA, intra-class variation was obtained by using consecutive half-cycles as well as by applying the artificial rotations in [28].

A discriminative approach is also taken for the features (observations) generated by the refinement HHMM. However, in this case LDA cannot be applied due to the short length (six elements) of the vector describing each labelled component feature. Alternatively, for each state and feature in the lower-level of the refinement HHMM, the distance between test and reference labelled component features is based on feature values that have been scaled using the ratio of the intra-subject over inter-subject variance among gait sequences in the reference database. Variances are calculated per state and feature. This is equivalent to applying discriminative weighting in order to promote discrimination despite the fact that the feature vectors are short and have elements that represent different quantities.

Using the above strategy, the probability in eq. (16) can be modified in order to include discriminating features:

$$P_3 = P(\tilde{\mathbf{H}}|\lambda)P(\tilde{\mathbf{F}}|\mu) \quad (21)$$

where  $\tilde{\mathbf{H}}$  and  $\tilde{\mathbf{F}}$  are the sequences of discriminative features that are assumed to be emitted by the dominant HMM and the lower-level of the refinement HHMM respectively.

The discriminative transform was applied on exemplars from each model, i.e., the exemplars from the 1st state in all models were taken together in order to calculate a discriminative transform for the 1st state. Similarly, discriminative transforms were calculated for the 2nd, 3rd, 4th, and 5th and state. Frame-to-exemplar distances, as in eq. (20) are calculated based on discriminative versions of frames and exemplars. But since the exemplars represent specific walking poses, the

discriminative transform derived from these exemplars indicate the discriminative differences in that specific pose. This means that the discriminative transform is pose-dependent and, consequently, it will not be discriminative in case the frame is dissimilar to the pose represented by the respective state exemplar. Essentially, the discriminative transform is most efficient only when the frames are similar to the exemplars. For this reason, we propose calculating eq. (21) along the optimal path because this is when the observations (frames) are most relevant to the states (exemplars) from which they are assumed to have been generated. Therefore, we define a new metric as

$$P_4 = P(\tilde{\mathbf{H}}|\hat{\mathbf{Q}}, \lambda)P(\tilde{\mathbf{F}}|\hat{\mathbf{S}}, \hat{\mathbf{Q}}, \mu) \quad (22)$$

where  $\hat{\mathbf{Q}}$  is given in eq. (18) and the most probable lower-level state sequences in the refinement HHMM are calculated as

$$\hat{\mathbf{S}}_n = \underset{\mathbf{S}_n}{\text{arg max}} P(\mathbf{F}, \mathbf{S}_n|\hat{\mathbf{Q}}, \mu), \quad n = 1, \dots, N \quad (23)$$

where  $\hat{\mathbf{S}} = \{\hat{\mathbf{S}}_1, \hat{\mathbf{S}}_2, \dots, \hat{\mathbf{S}}_N\}$ , with  $\mathbf{S}_n$  being a state sequence corresponding to the  $n$ th HMM in the lower level of the refinement HHMM.<sup>3</sup> Metrics  $P_1, P_2, P_3, P_4$ , defined above, will be assessed in the experimental results section.

### C. Gait half-cycle fusion

A gait cycle consists of two half-cycles that have to be compared to the two reference half-cycle models that have been built for each reference subject. The calculation of probabilities  $P_1, P_2, P_3, P_4$ , defined earlier, is performed separately for each of two consecutive half-cycles in a test gait sequence. Specifically, each test half-cycle yields two probabilities, against each of the two half-cycle reference models of a reference subject. As a result, four probabilities are calculated from each test gait (full) cycle. In the classification module of our system, we consider the highest of the four probabilities as the most efficient measure of similarity between a test and a reference subject. This obviates the need to label explicitly the half-cycles according to which starts with the strike of the left or right heel.

<sup>3</sup>It should be noted that all possible optimal state sequences calculated in the dominant HMM (and inherited by the upper refinement model) lead to valid lower-level state sequences in the refinement model.

TABLE III  
EIGHT DIFFERENT SETS IN THE ACTIBIO DATABASE.

Set	Day	Rep.	Subjects	Condition
Gallery	1	1	28	Normal walking
Probe A	1	1	28	Carrying a bag
Probe B	1	1	27	Wearing a coat
Probe C	1	1	28	Wearing slippers or socks only
Probe D	1	1	28	Walking diagonally
Probe E	1	2	28	Normal walking
Probe F	2	1	27	Normal walking
Probe G	2	2	27	Stop for a while

## VII. EXPERIMENTAL RESULTS

For the application of our system, we used a dominant HMM with  $N = 5$  states and a refinement Hierarchical HMM comprising  $N$  HMMs with  $M = 3$  states each (fig. 5(b)). Note that “end” states do not count as real states. For the evaluation of recognition performance, Cumulative Match Scores (CMS) were calculated. In CMS, a Rank  $k$  score represents the percentage of probe (test) subjects whose corresponding gallery (reference) subject is within the top  $k$  matches.

We conducted a detailed evaluation of our methods using the ACTIBIO gait database (representative results on other databases also confirm our conclusions). The ACTIBIO database was recorded in the framework of the ACTIBIO project<sup>4</sup>. In this database, there are 28 subjects walking in an indoor environment. For each subject, there are up to eight gait sequences, captured on two recordings with the second recording taking place one month after the first. Two repetitions were captured in each day and several walking conditions were recorded in each repetition. A full description of each recorded condition is presented in Table III.

For each subject in our experiments, one of the eight sequences is used as a gallery (reference) sequence and the others are probe (test) sequences. Similar to [15], where several classification methods were tested on mixed conditional random fields for gender classification, we compared several classification approaches, based on eqs. (16), (17), (21), and (22). For the sake of comparison, the key recognition rates for all above experiments are summarized in Table IV.

<sup>4</sup><http://www.actibio.eu/>

The tabulated results indicate that substantial improvements over the simple fusion of holistic and model-based features (based on  $P_1$ ) are achieved using discriminative features (based on  $P_3, P_4$ ). In accordance with the discussion in Section VI.B, the results in Table IV confirm that classification based on  $P_4$  is more efficient than that using  $P_3$ .

Detailed CMS curves are shown in fig. 8. In figs. 8(a),(b) we present detailed results using holistic features only. In this case, no analysis of sub-dynamics took place. In figs. 8(c),(d), we present detailed results using our final scheme in which combination of holistic and model-based features has taken place. As seen, the best performing system is the one that combines analysis of sub-dynamics and discriminative features (bottom right corner of fig. 8).

Based on the above experiments, the following conclusions can be drawn:

- System performance based on the combination of holistic and labelled component features, presented in figs. 8(c) and 8(d), is superior to that achieved by the application of the holistic features alone, shown in figs. 8(a) and 8(b). This confirms that the combination of holistic and detailed (non-holistic) features in our system is able to capture gait sub-dynamics and improve recognition efficiency. It should be noted that the independent application of labelled component features does not yield competitive results. This implies that general gait dynamics (captured through holistic features) are essential for good recognition performance.
- The versions of the system that use discriminative features and state exemplars ( $P_4$  in Table IV, right column of Fig. 8) are superior to those based only on the original features (left column of fig. 1). The practical significance of this conclusion is that the application of discriminative techniques in HMM-based gait recognition systems can yield considerable performance improvements.

After assessing the efficiency of the proposed classification approaches within our HMM framework, we compare the best-performing of our algorithms with other methods. The two variants of our system that we compare are: a *Holistic (HL)* method based only on holistic features (silhouettes), i.e., using only the dominant HMM in fig. 5(a) (with classification based only on the first term on the right-hand side of eq. (22)), and another method involving both holistic and labelled component features, which is based on eq. (22) and is named *Discriminant HMM (DHM)*. Both are using discriminative features for similarity calculation. We compare our methods with the *Dynamic Time Warping (DTW)* approach [29], an HMM system (named

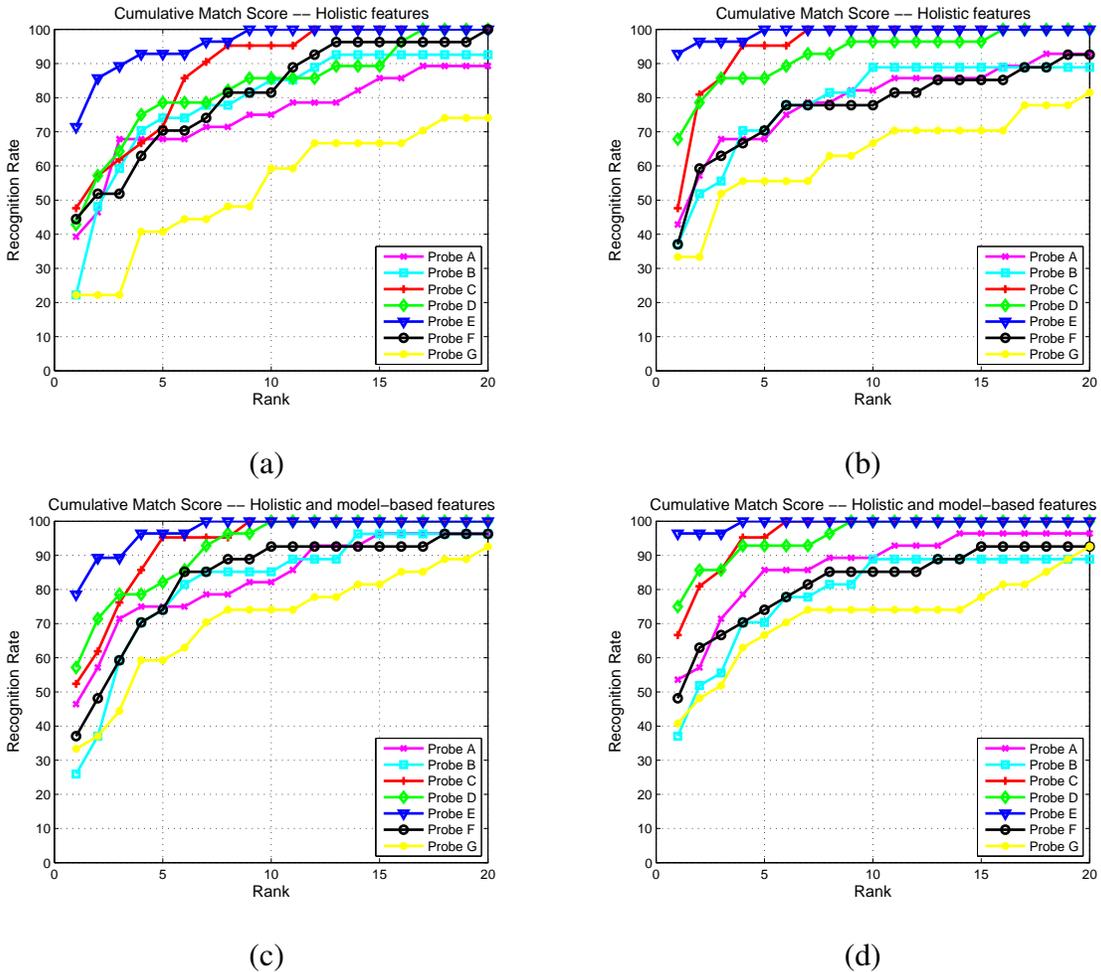


Fig. 8. Cumulative match scores for gait recognition. **Top row**: system based on holistic features only. (a) original features, (b) discriminative features (*HL* method). **Bottom row**: system based on the combination of holistic and labelled component features for analysis of sub-dynamics. (c) original features, (d) discriminative features (*DHM* method).

*S15*) based on a single layer with 15 states that emit holistic features (using our discriminative approach in classification), the model-based method *FGF*, the benchmark *Gait Energy Image (GEI)* method [6], the HMM-based method in [17] as well as the *Gait Entropy Image (GEnI)* method [3].

Our comparisons allow the assessment of the impact on system performance of a variety of characteristics, e.g., the number of HMM states, the use of labelled component features, the similarity calculation approach, and the use of discriminative features. Recognition results for our proposed *HL*, *DHM* methods in comparison to the *DTW*, *S15*, *FGF*, *GEI*, *HMM*, and *GEnI* are shown in Table V. The holistic method (*HL*), despite not using model-based features,

TABLE IV  
 RECOGNITION RATES ACHIEVED BY THE PROPOSED METHOD USING SIMILARITY MEASURES BASED ON EQUATIONS (16),  
 (17),(21), AND (22).

Probe	$P_1$		$P_2$		$P_3$		$P_4$	
	$R_1$	$R_5$	$R_1$	$R_5$	$R_1$	$R_5$	$R_1$	$R_5$
A	46	75	54	86	46	75	54	86
B	26	74	30	74	29	67	37	70
C	52	95	47	81	67	95	67	95
D	57	82	50	89	79	89	75	93
E	78	96	82	96	96	100	96	100
F	37	74	37	74	37	67	48	74
G	33	59	33	59	37	52	41	67

exhibits excellent performance. Once more, this indicates that the discriminative projection *per HMM state* constitutes a powerful tool for improving the performance of HMM-based systems in gait recognition. The comparison between *HL* and *S15* shows that using a higher number of states does not necessarily lead to increased ability to extract discriminatory information from gait sequences. The best performing version (*DHM*) of our proposed system improves on *HL*, providing further evidence that the additional deployment of labelled component features can improve the performance of holistic systems by analyzing gait sub-dynamics. Most importantly, the *DHM* invariably outperforms all other methods by achieving, in almost every rank, a higher average recognition rate than that achieved by the other methods in the comparison.

## VIII. CONCLUSIONS

We proposed a new gait recognition method that combines holistic and model-based features by means of a pair of HMMs. Holistic features were initially used for capturing general gait dynamics while, subsequently, model-based features were deployed for capturing more detailed sub-dynamics by refining upon the preceding general dynamics. Furthermore, holistic and model-based features were suitably processed with the purpose of enhancing the discriminatory capacity of the final system. Experimental results showed that the proposed method exhibits performance advantages in comparison to popular existing methods.

TABLE V

RECOGNITION RATES IN COMPARISON TO OTHER METHODS ON THE ACTIBIO DATABASE. TWO VARIANTS OF OUR PROPOSED SYSTEM ARE COMPARED: ONE (HL) BASED ONLY ON HOLISTIC FEATURES AND ANOTHER (DHM) BASED ON BOTH HOLISTIC AND MODEL-BASED FEATURES.

Probe	Rank 1 (%)								Rank 5 (%)							
	DTW	S15	FGF	GEI	HMM	GEnI	HL	DHM	DTW	S15	FGF	GEI	HMM	GEnI	HL	DHM
			[27]	[6]	[17]	[3]					[27]	[6]	[17]	[3]		
A	30	53	42	32	39	28	42	<b>54</b>	50	75	79	47	68	42	67	<b>86</b>
B	17	29	9	19	22	36	<b>37</b>	<b>37</b>	61	70	52	49	<b>74</b>	51	70	70
C	22	42	52	29	48	34	48	<b>67</b>	57	<b>95</b>	81	61	71	58	<b>95</b>	<b>95</b>
D	40	64	50	39	43	42	68	<b>75</b>	75	92	87	68	79	67	86	<b>93</b>
E	57	89	58	64	71	50	93	<b>96</b>	82	96	89	82	93	60	<b>100</b>	<b>100</b>
F	28	44	15	41	44	40	37	<b>48</b>	58	59	39	67	70	66	70	<b>74</b>
G	15	26	25	19	22	14	33	<b>41</b>	33	51	50	40	41	29	56	<b>67</b>

#### ACKNOWLEDGEMENT

The authors would also like to thank the anonymous reviewers for their feedback, which resulted in an improved manuscript.

#### REFERENCES

- [1] N.V. Boulgouris, D. Hatzinakos, and K.N. Plataniotis, "Gait recognition: a challenging signal processing technology for biometric identification," *IEEE Signal Processing Magazine*, vol. 22, pp. 78–90, Nov. 2005.
- [2] 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 Trans. Pattern Anal. and Machine Intell.*, vol. 27, no. 2, pp. 162–176, Feb. 2005.
- [3] K. Bashir, T. Xiang, and S. Gong, "Gait recognition without subject cooperation," *Pattern Recognition Letters*, vol. 31, no. 13, pp. 2052–2060, Oct. 2010.
- [4] M. Goffredo, J.N.Carter, and M.S. Nixon, "Front-view gait recognition," in *IEEE Second Int. Conf. on Biometrics: Theory, Applications and Systems*, Washington D.C., USA, Sep. 2008.
- [5] D. Ioannidis, D. Tzovaras, I. G. Damousis, S. Argyropoulos, and K. Moustakas, "Gait recognition using compact feature extraction transforms and depth information," *IEEE Trans. Information Forensics and Security*, vol. 2, no. 3, pp. 623–630, Sep. 2007.
- [6] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 28, no. 2, pp. 316–322, 2006.
- [7] I. Kotsia and I. Patras, "Support tucker machines," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, USA, June 2011, pp. 633–640.
- [8] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and gabor features for gait recognition," *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol. 29, no. 10, pp. 1700 – 1715, 2007.

- [9] D. Tao, X. Li, S. J. Maybank, and X. Wu, "Human carrying status in visual surveillance," in *Proc. CVPR*, 2006, vol. 2, pp. 1670–1677.
- [10] Z. Liu and S. Sarkar, "Effect of silhouette quality on hard problems in gait recognition," *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 35, no. 2, pp. 170–183, 2005.
- [11] N. V. Boulgouris and Z. X. Chi, "Human gait recognition based on matching of body components," *Pattern Recognition*, vol. 40, no. 6, pp. 1763–1770, 2007.
- [12] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "A full-body layered deformable model for automatic model-based gait recognition," *EURASIP Journal on Advances in Signal Processing*, 2008.
- [13] Z. Liu and S. Sarkar, "Improved gait recognition by gait dynamics normalization," *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol. 28, no. 6, pp. 863–876, 2006.
- [14] M. Hu, Y. Wang, Z. Zhang, and D. Zhang, "Multi-view multi-stance gait identification," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*, Sept. 2011, pp. 541–544.
- [15] M. Hu, Y. Wang, Z. Zhang, and D. Zhang, "Gait-based gender classification using mixed conditional random field," *IEEE Trans. Systems, Man, and Cybernetics, Part B*, vol. 41, no. 5, pp. 1429–1439, Oct. 2011.
- [16] M. Hu, Y. Wang, Z. Zhang, D. Zhang, and J. Little, "Incremental learning for video-based gait recognition with lbp flow," *IEEE Trans. Systems, Man, and Cybernetics, Part B*, vol. 43, no. 1, pp. 77–89, Feb. 2013.
- [17] A. Kale, A. Sundaresan, A. N. Rajagopalan, N. Cuntoor, A. K. Roy-Chowdhury, V. Krueger, and R. Chellappa, "Identification of humans using gait," *IEEE Trans. Image Processing*, vol. 13, no. 9, pp. 1163–1173, Sep. 2004.
- [18] C. Chen, J.-Liang, H.-Zhao, H.-Hu, and J.-Tian, "Factorial hmm and parallel hmm for gait recognition," *IEEE Trans. Systems, Man, and Cybernetics-Part C: applications and reviews*, vol. 39, pp. 114–123, 2009.
- [19] M. Piccardi, "Background subtraction techniques: a review," in *Proc. IEEE Int. Conf. Systems, Man and Cybernetics*, Oct. 2004, pp. 3099–3104.
- [20] L. Rabiner and B. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, Englewood Cliffs, NJ, 1993.
- [21] C. Panagiotakis and G. Tziritas, "Recognition and tracking of the members of a moving human body," *Articulated Motion and Deformable Objects*, pp. 86–98, 2004.
- [22] N.V. Boulgouris, K.N. Plataniotis, and D. Hatzinakos, "Gait recognition using dynamic time warping," in *Proc. IEEE Int. Symp. Multimedia Signal Processing*, Siena, Italy, Sep. 2004, pp. 263–266.
- [23] D. A. Winter, *The Biomechanics and Motor Control of Human Movement*, John Wiley & Sons, 2 edition, 1990.
- [24] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, Inc., 2001.
- [25] S. Fine, Y. Singer, and N. Tishby, "The hierarchical hidden markov model: Analysis and applications," *Machine Learning*, no. 32, pp. 41–62, 1998.
- [26] S.-Z. Yu, "Hidden semi-markov models," *Artificial intelligence*, vol. 174, pp. 215–243, 2010.
- [27] X. Huang and N. V. Boulgouris, "Model-based human gait recognition using fusion of features," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Taipei, Apr. 2009, pp. 1469–1472.
- [28] X. Huang and N. V. Boulgouris, "Gait recognition using linear discriminant analysis with artificial walking conditions," in *IEEE International Conference on Image Processing*, September 2010, pp. 2461–2464.
- [29] B.-H. Juang, "On the hidden markov model and dynamic time warping for speech recognition," *ATT Technical jnl*, vol. 63, pp. 1213–1243, 1984.

PLACE  
PHOTO  
HERE

**Nikolaos V. Boulgouris** (S'96 M'04 SM'09) is a Senior Lecturer with the Department of Electronic and Computer Engineering at Brunel University, U.K. From December 2004 to August 2010 he served as a Lecturer and a Senior Lecturer at King's College London, London, U.K. From 2003 to 2004, he was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering of the University of Toronto, Canada. He received the Ph.D. degree from the Electrical and Computer Engineering department of the University of Thessaloniki, Greece, in 2002.

Dr Boulgouris served as a guest co-editor for two journal special issues and was co-editor of the book *Biometrics: Theory, Methods, and Applications* that was published by Wiley - IEEE Press in 2009.

Dr. Boulgouris is an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING. He served as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS between 2007 and 2011.

PLACE  
PHOTO  
HERE

**Xiaxi Huang** received the BEng degree in Control Science and Engineering from Zhejiang University, China, in 2004, the MSc degree in Electrical and Electronic Engineering from University of Bath, Bath, U.K, in 2005, and the PhD degree in Electronic Engineering from King's College London, U.K, in 2010. His research interests include digital image/video processing and gait biometrics based on computer vision.