# Frontal-view gait recognition by intra- and inter-frame rectangle size distribution

Olivier Barnich, and Marc Van Droogenbroeck

O. Barnich@ulg.ac.be, M.VanDroogenbroeck@ulg.ac.be
Tel: 32 (0)4 366 26 93 (secr. +32 (0)4 366 26 91), Fax: +32 (0)4 366 29 89
Laboratory for Signal and Image Exploitation (IntelSig), University of Liège
Institut Montefiore, Grande Traverse 10, Sart Tilman, B-4000 Liège, Belgium

**Abstract**

Current trends seem to accredit gait as a sensible biometric feature for human identification, at least in a multimodal system. In addition to being a robust feature, gait is hard to fake and requires no cooperation from the user. As in many video systems, the recognition confidence relies on the angle of view of the camera and on the lightening conditions, inducing a sensitivity to operational conditions that one may wish to lower.

In this paper we present an efficient approach capable of recognizing people in frontal-view video sequences. The approach uses an intra-frame description of silhouettes which consists of a set of rectangles that will fit into any closed silhouette. A dynamic, inter-frame, dimension is then added by aggregating the size distributions of these rectangles over multiple successive frames. For each new frame, the inter-frame gait signature is updated and used to estimate the identity of the person detected in the scene. Finally, in order to smooth the decision on the identity, a majority vote is applied to previous results. In the final part of this article, we provide experimental results and discuss the accuracy of the classification for our own database of 21 known persons, and for a public database of 25 persons.

## 1 Introduction

The number of video-surveillance cameras has increased dramatically over the last few years. It has therefore become unrealistic to process manually or even visually the gigantic amount of information gathered by surveillance cameras, which explains why the automation of real-time visual surveillance tasks is currently one of the most active topics in computer vision. Visual surveillance has a wide spectrum of promising applications, including control of access to certain areas, human identification, crowd flux statistics, detection of anomalous behaviors, etc [12]. This paper focuses on one of these tasks, i.e. automatic human identification.

Automatic human identification can be achieved through a variety of biometrics using different kinds of sensors: fingerprint readers, iris scanners, microphones for voice recognition, and video cameras. One advantage of video

cameras is that they are not intrusive; also subjects can be filmed without their cooperation. Face recognition through the use of a video camera is a widely used biometric, although its efficiency is conditioned by the need for a relatively constrained image of the person's face. Unconstrained face recognition is possible (see [28]) but is almost useless for strong identification in practice. Asking a person to cooperate can also be an issue; not everyone is going to help the system. Gait recognition is therefore a viable alternative; in this case, it is neither necessary to restrict the field of view to constrained environment, nor to ask for cooperation. Gait recognition is not (yet?) as effective as the best face recognition algorithm but, acting as a complementary form of identification, it might reinforce a decision made in a multi-modal biometric system.

Gait as a biometric is quite a recent topic for discussion, which has gained in popularity since its introduction in [22]. Its robustness against poor imaging conditions makes it applicable to a wide range of real-world scenarios. Images can be acquired from a great distance, even in changing illumination conditions (i.e. outdoor, as shown in [18]). Furthermore, absolutely no kind of cooperation from the subjects is required. Gait is also difficult (if not impossible?) to fake. Yet, gait recognition techniques are still not accurate enough to use gait as the sole biometric of a real surveillance system. These recognition techniques are better used to reinforce a decision in a multi-modal biometric system (see [20, 21, 29]).

Gait recognition techniques are usually classified in two categories: model-based and holistic/silhouette approaches [3].

Model-based approaches make use of explicit gait models whose parameters are to be estimated by processing sequences of images, hereafter referred to as image frames or frames. The identification is performed entirely on the basis of the estimated values of the explicit gait model. Model-based approaches are generally scale and view invariant, as long as the parameters estimation is feasible given the imaging configuration. This is a major advantage, since training conditions are likely to differ from conditions of practical use. On the other hand, these methods often need high definition images in order to work properly. They also exhibit a significantly higher computational cost. Techniques in this category include modeling the thighs as a pair of thick lines, as in [7], modeling the silhouette of a walking person as a group of seven ellipses as in [10], or modeling the legs as two penduli joined in series, as in [27].

Holistic approaches do not assume any explicit model for the walking human. They extract information directly from the gait image sequences. Gait signatures are, for example derived from time series of binary silhouettes extracted from the original sequence with a background subtraction algorithm. This brings a suitable invariance to color, texture or illumination conditions (assuming that the used background subtraction algorithm is robust). A simple approach that uses areas of raw (re-sized) silhouettes as a gait signature is described in [8]. The contours of silhouettes have also been used, either directly [26] or through their Fourier descriptors [19]. An angular transform of the silhouette is proposed in [4]. This is said to be more robust than the raw contour descriptions. In [2], the gait dynamics are captured using principal components analysis of self-similarity plots. Feature vectors derived from the binary silhouettes can also be used to train HMM's, as in [15].

Other authors have used horizontal and vertical projections of the silhouettes [14]. In [17], time series of horizontal and vertical projections of silhouettes

(a) Lateral view (b) Frontal view

Figure 1: Lateral and frontal views of a walker.

are treated as *frieze* patterns. The framework of frieze patterns leads the authors to estimate the viewing direction of the walking humans and to align gait sequences from similar viewpoints both spatially and over time. The identification is then performed using cross-correlation and nearest neighbor classification between frieze patterns. In [16], a similar algorithm is used to compare frieze patterns of frame differences between a key silhouette and a series of successive silhouettes. The method is claimed to be more robust to silhouette differences between the training and test sets.

Nearly all silhouette-based approaches are designed to deal with image frames captured from the side of a person (see Figure 1(a)). While it is reasonable to assume that the lateral view captures an appropriate amount of gait and walking information, it is not easy to capture these image frames in practical scenarios. In order to obtain a sufficiently long sequence of images of a person walking *(i.e.* containing several gait cycles), cameras need to be put at a long distance. This hinders recognition, since small silhouettes are hard to discriminate. In hallways (see the example in Figure 1(b)), frames are rarely captured from the side, but from the front or the back of the walker (see Figure 1(a)). Front-view cameras, as opposed to lateral-view cameras, capture longer sequences of walkers, which results in more gait cycles. However front-view cameras are thought to be less efficient for gait recognition as they capture geometric and scale transformations of the silhouettes. But the human capacity to recognize people using only a frontal view of their walking silhouettes tends to prove that a frontal view contains enough information to perform automatic recognition. This is confirmed by Soriano et al. [24]. In an article in which gait signatures are derived from series of Freeman encoding of the re-sized silhouette shape, these authors showed that frontal view gait recognition is possible [24].

In [13], the gait template of a walking human is computed by averaging the corresponding binary silhouettes. The classification is then achieved using a nearest neighbor technique. The authors use the MoBo database [11] from the CMU to compare the classification results obtained by their method with

sequences captured from different viewpoints. The best single viewpoint results are obtained using the frontal view. But better classification scores are achieved by combining the frontal view with the lateral view.

This paper presents a gait recognition algorithm capable of recognizing persons from image frames captured in real-time with surveillance cameras located in hallways. Unlike many techniques in the literature which process complete gait sequences, our algorithm identifies a previously known person as soon as it obtains a complete gait cycle, which accounts for about 1 second or 25 frames. Requirements for our method are that (1) low image resolution (like $640 \times 480$) suffice, (2) walkers can wander at quite a long distance from the cameras, and (3) the algorithm should run in real time on any computer.

For noisy surveillance video frames, a precise detection of moving objects and their contours is difficult. In order to achieve a better resilience to noise, we chose a surfacic representation of the silhouettes in terms of a descriptor called "Cover by Rectangles", introduced in [1]. This descriptor provides a piecewise surfacic description of silhouettes which, unlike horizontal and vertical projections, is reversible and therefore does not induce any information loss. In addition, covers by rectangles limits the effect of noise to a local neighborhood as noise will impact locally on the description of the silhouette, in contrast with global surfacic measures. Section 2 derives a new silhouette representation based on the cover by rectangles approach. This representation serves to characterize gait silhouettes for each frame separately; we therefore call this an intra-frame descriptor. Section 2 also explains how we consider temporal and dynamic information by introducing inter-frame dependencies in order to derive a gait signature. We describe the complete gait identification algorithm in Section 3. Experimental results and an evaluation of our method are presented in Section 4. We show that gait recognition is possible, efficient, and achievable in real time, even for front-view video frames.

# 2 A surfacic gait representation

In order to identify a walking person, a time series of his silhouettes is extracted from the raw video frames, at a rate of one silhouette per frame. For each frame, the silhouette is converted into a set of features, which are used to update a gait signature. The gait signature is fed into a classifier which will output the class label corresponding to a particular person. Hereafter we present the intra-frame description of a silhouette.

## 2.1 Cover by rectangles of a binary silhouette

The cover by rectangles, proposed in [1], is a morphological descriptor. Consider a binary silhouette $S$. The cover by rectangles, denoted $C(S)$, is defined as the union of all the largest rectangles that can fit inside of $S$ (see Figure 2 for an example). This union is unique and the cover $C(S)$ has the following useful properties: (1) the elements of the set overlap each other, introducing redundancy (i.e. robustness), (2) each element (rectangle) of $C(S)$ covers at least one pixel that belongs to no other rectangle, and (3) when displayed in the frame, the union of all rectangles reconstructs $S$ so that no information is ever lost.

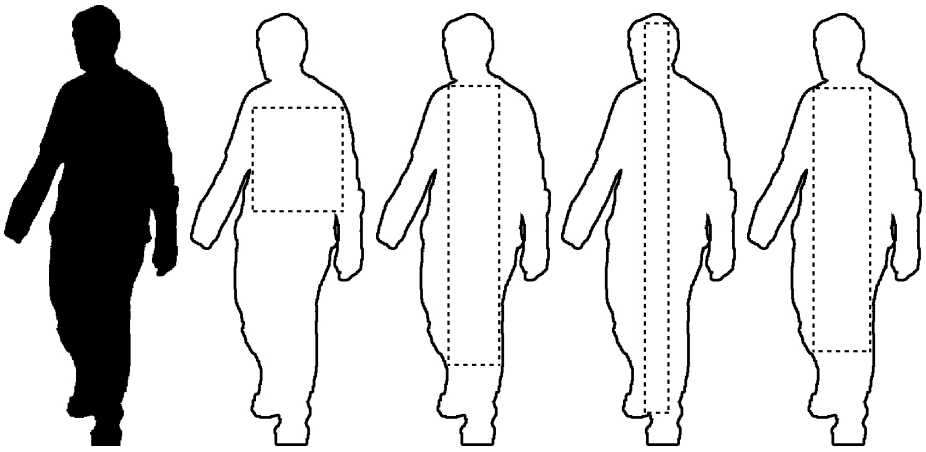


Figure 2: The cover by rectangles $C(S)$ is the union of all the largest rectangles that can be wedged inside of the silhouette.

Other morphological surfacic descriptors, such as the morphological skeleton [23], have been developed to represent shapes. However, since they provide an isotropic description of the silhouettes through, for example, the union of open balls included in $S$, they are unsuited for the description of gait. Moreover, it is important to ensure that a local modification of the silhouette does not lead to a global change in its description. Figure 3 compares the effect of a slight modification of the shape in the case of the skeleton and features (widths or heights) derived from the rectangles of $C(S)$. In Section 4.2, we show that a gait signature based on the cover by rectangles of the silhouettes of a walking human is robust and allows the correct identification of people from noisy silhouettes (see Figure 6) through a set of experiments.

## 2.2 Rectangle size probability distributions

The number of largest wedged rectangles that will fit inside a binary silhouette can be very high (more than a thousand). It is thus impractical to use all the rectangles directly as a set of features. In order to find a more compact representation, we can operate on one of the size distribution densities, as shown in Figure 4. These distributions offer different but suitable interpretations of a silhouette. For example, the largest number of rectangles containing a given pixel is to be found inside the torso (Figure 4(b)), and the tallest rectangles pass through both the legs and the head (Figure 4(d)).

As can be seen, much of the information resides in the distributions of the normalized sizes (width or height). These distributions can be estimated as a discrete histogram whose bins correspond to the ratios of rectangles that fall within given size intervals.

From a formal point of view, let $\alpha$ be the cardinality of a cover by rectangles $C(S)$, *i.e.* $\alpha = \sharp\{C(S)\}$. We index the rectangles of $C(S)$ with a parameter $d$, so that $R_d$ $(d = 1, \ldots, \alpha)$ are the rectangles of $C(S)$. The width and height of $R_d$ are respectively denoted by $w_d$ and $h_d$; they are upper-bounded by $w^{max}$ and $h^{max}$: $\forall d$, $w_d \leq w^{max}$ and $h_d \leq h^{max}$. In order to build histograms, we

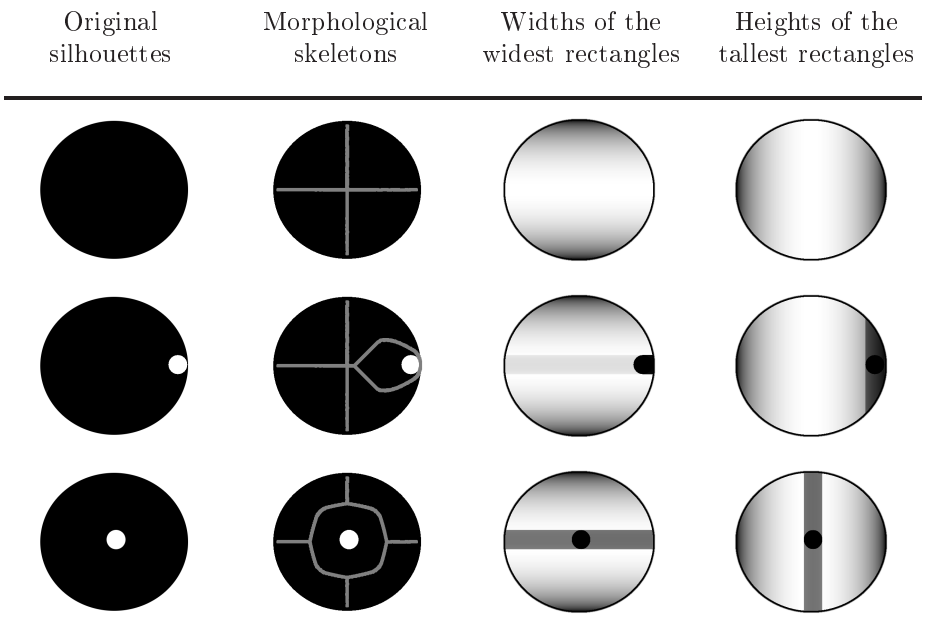



Figure 3: The first column shows three original images. The morphological skeletons (shown in gray in the second column) are modified by the presence of a small hole in the silhouette: a local perturbation leads to a global modification of the skeleton. The images the two right-hand columns represent the size distributions of the rectangles contained in $C(S)$. In these images, the gray level of pixels is proportional to the width (resp. height) of the widest (resp. tallest) rectangle comprising the given pixel.

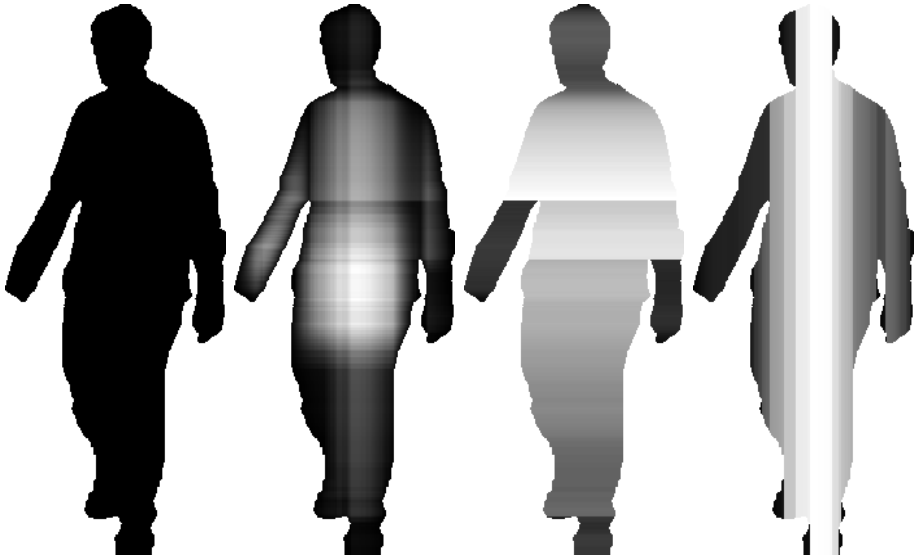



Figure 4: Illustration of several size distributions based on the description provided by the cover $C(S)$ of a binary silhouette $S$. A gray level of pixel $p$ in images (b), (c), and (d) displays respectively the density of rectangles, the width of the widest rectangle, and the height of the tallest rectangle where all these rectangles contain pixel $p$.

partition the widths and heights of the rectangles $R_d$ respectively into $M$ bins $B^W(i)$ and $N$ bins $B^H(j)$

$$B^W(i) = \left]i\frac{w^{max}}{M},\ (i+1)\frac{w^{max}}{M}\right] \quad (1)$$

$$B^H(j) = \left]j\frac{h^{max}}{N},\ (j+1)\frac{h^{max}}{N}\right] \quad (2)$$

where $i = 0, \ldots, M-1$ and $j = 0, \ldots, N-1$.

Following the above notations, we define the histogram $\text{hist}^W(i)$ of the normalized widths as

$$\text{hist}^W(i) = \frac{1}{\alpha}\sharp\left\{R_d |\, w_d \in B^W(i)\right\}, \quad (3)$$

the histogram of the normalized heights similarly as

$$\text{hist}^H(j) = \frac{1}{\alpha}\sharp\left\{R_d |\, h_d \in B^H(j)\right\}, \quad (4)$$

and the two-dimensional histogram $\text{hist}^{W\times H}(i,j)$ as

$$\text{hist}^{W\times H}(i,j) = \frac{1}{\alpha}\sharp\left\{R_d | w_d \in B^W(i),\, h_d \in B^H(j)\right\}. \quad (5)$$

Note that these histograms are normalized with respect to the largest rectangle of the cover of the silhouette. In a continuous space, they would be scale invariant. Such a normalization might seem counter-intuitive; much of the interpretation of the motion of a gait derives from the size of a silhouette, and it would not be good for frontal cameras to lose motion information. A finer analysis shows however that size information is still present in a normalized histogram. Indeed the cover of a scaled down version of a silhouette $S$ contains fewer rectangles ($\alpha$ is always lower than the number of contour points) than its original counterpart. Therefore the histograms have a distribution that adapts to both the shape and the size of a silhouette. In addition, if noise is added to the contour of the silhouette, it will modify the positions of the rectangles but not so much their size or number.

Of the three $\text{hist}^W(i)$, $\text{hist}^H(j)$, $\text{hist}^{W\times H}(i,j)$ histograms, the last one best describes $S$. However, its dimensionality is proportional to the product of the numbers of bins ($M \times N$), which is acceptable for an intra-frame description but might be too high for embedded systems if the features are to be fed into a classifier for inter-frame gait recognition. In order to solve this tractability issue, we introduce the composite histogram $\text{hist}^{W+H}(k)$ with $k = 0, \ldots, M+N-1$ defined as the strict concatenation of $\text{hist}^W(i)$ and $\text{hist}^H(j)$. $\text{hist}^{W+H}(k)$ has a dimensionality of $M+N$, and accounts for both the vertical and horizontal characteristics of the silhouette. Experiments detailed in Section 4 show that both $\text{hist}^{W\times H}(i,j)$ and $\text{hist}^{W+H}(k)$ are suitable descriptors.

## 2.3 Gait as an inter-frame rectangle distribution

So far we have considered a single intra-frame silhouette, but a gait sequence is a temporal series of binary silhouettes. In order to capture the dynamics of a walking person we introduce an inter-frame dependency by defining a gait

signature based on the temporal series of the silhouettes $S$ of a walker. We assume that $t$ refers to the time of the current frame, and that hist$(i,\ j,\ t)$ is a histogram for $S$ at time $t$. We introduce two gait signatures, denoted $\mathcal{G}$, which consist of n-uples of $L$ consecutive histograms. We propose the following gait signature

$$\mathcal{G}^{W\times H}(i,j,t)=\Big\{\text{hist}^{W\times H}(i,\ j,\ t-(L-1)),...,\ \text{hist}^{W\times H}(i,\ j,\ t-1),\text{hist}^{W\times H}(i,\ j,\ t)\Big\}, \quad (6)$$

and a shortened version as

$$\mathcal{G}^{W+H}(k,t)=\Big\{\text{hist}^{W+H}(k,\ t-(L-1)),...,\ \text{hist}^{W+H}(k,\ t-1),\text{hist}^{W+H}(k,\ t)\Big\}. \quad (7)$$

# 3 Gait recognition algorithm

The gait recognition process is shown in Figure 5. For every frame of a gait sequence, it predicts the identity of the walking human. The algorithm consists of three steps, further detailed in this section:

1. extraction of a silhouette by a background subtraction technique at time $t$,
2. computation of a histogram at time $t$, which is used to update the gait signature, and
3. classification of a gait signature by a machine learning algorithm which outputs the identity of one of the persons known to the system.

## 3.1 Silhouette extraction

The quality and the changing nature of the illumination conditions encountered when using real surveillance cameras led us to adopt an advanced background subtraction technique which can deal with changing illumination, noisy sensors and cast shadows. This background technique was proposed by Zivkovic in [30]. It extends the widely used Mixture Of Gaussian algorithm ([25]) by selecting automatically and dynamically the optimal number of Gaussian distributions to use for each pixel. The result of this background extraction technique is illustrated in Figure 6. It can be seen that despite the use of an advanced background subtraction technique, the silhouette is not perfectly detected. Much of the gait recognition efficiency will therefore rely on the robustness of the gait signature.

## 3.2 Intra-frame silhouette description and gait signature by rectangle size distributions

In order to characterize a gait, we use one of the gait signatures introduced in Section 2.3. These are updated frame by frame, as soon as a silhouette histogram is computed at time $t$. Figure 7 displays a graphical representation of $\mathcal{G}^{W+H}(k,\ t)$ to show the quantity of information gathered in the signature.

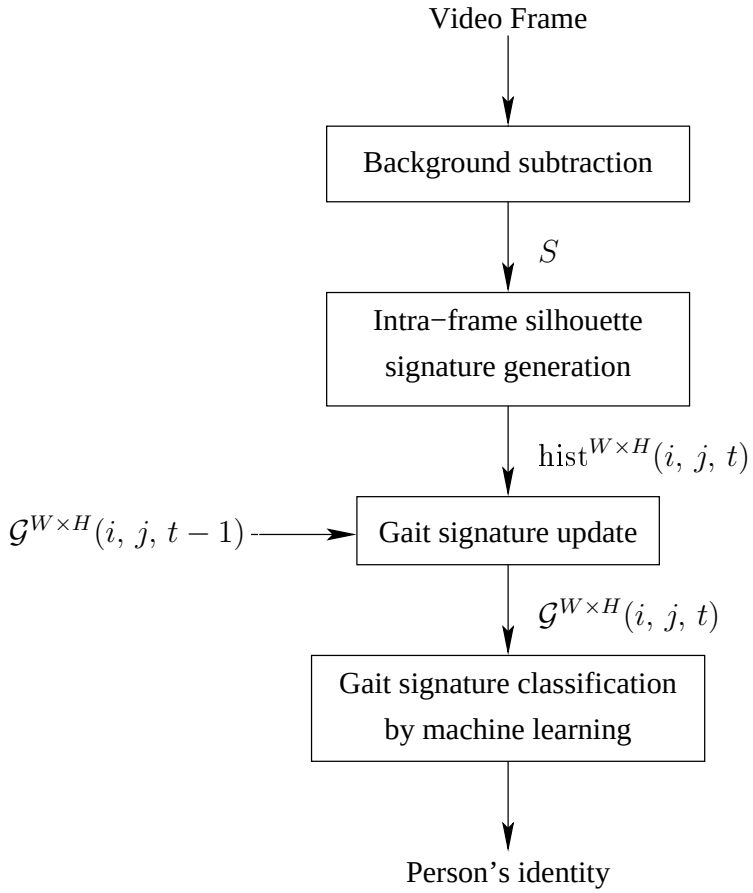



Figure 5: Steps of our gait recognition algorithm.

Figure 6: Example of binary silhouette extracted with the algorithm of Zivkovic, as described in [30].

Since we do not perform any kind of tracking, we restrict ourselves to only one person being present at a time in the field of view of the camera. The choice of using $\text{hist}^{W+H}()$ or $\text{hist}^{W\times H}()$ depends on the amount of training data available as the dimensionality of $\text{hist}^{W\times H}()$ is usually larger than the one of $\text{hist}^{W+H}()$.

It is important to note that our method comprises no gait cycle detection or normalization algorithm, unlike many techniques described in the literature (see [3]); our tests have proven that these techniques can be unnecessary.

## 3.3 Gait classification

The gait signature obtained at time $t$ is the feature set used for recognition. There is no special difficulty involved in mapping a gait signature to a class label, except that it must be fast, versatile, and accurate. Another criterion for the classifier is its ability to handle sets of features having high dimensionalities ($(M+N)\times L$ or even $M\times N\times L$ in our case). We chose a classifier, called extra-trees (for EXtremely RAndomized TREES) for its ability to handle features spaces of high dimensionality. Without going into detail, extra-trees is a kind of crossover between *bagging* [5] and *random forests* [6]. The goal of extra-trees is to reduce the variance by using a forest of independent trees instead of a single tree, and to reduce the bias by using a random selection of the thresholds at the splits of the trees (see [9] for a full description).

### 3.3.1 Majority vote policy on a sliding temporal window

Our gait recognition algorithm is synchronous: it provides the name for the person in the field of view whatever the time $t$ might be. This is less restrictive than many techniques described in the literature which have to process the *complete* gait sequence before producing a single class label. On the other hand, this guarantees no temporal consistency, and a new, possibly different, class label might be computed by the system for each new frame, on the basis

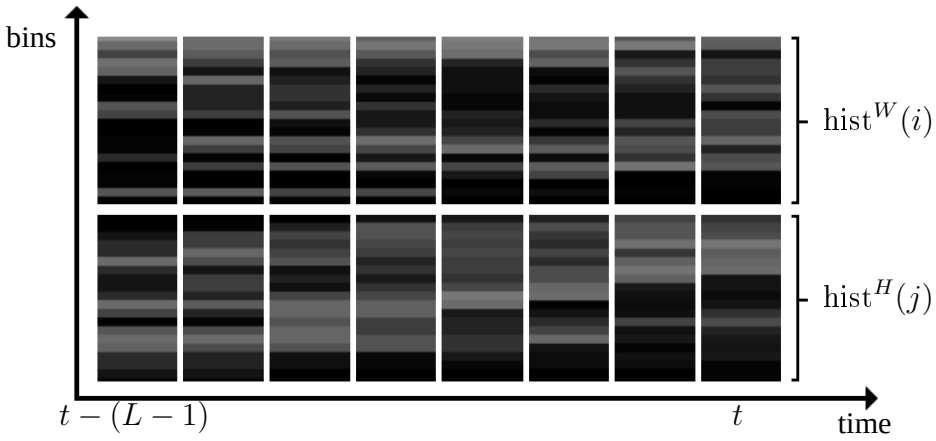



Figure 7: A graphical representation of $\mathcal{G}^{W+H}(k,\, t)$. All these displayed bin values are part of the feature set given to the gait classification algorithm.

of the previous $L$ frames. In order to smooth the result over time, we add a step that performs a majority vote on the previous $V$ class labels produced by the classifier. Since the gait signatures already account for the information contained in the previous $L$ frames, this brings a total delay of $L+V$ frames in achieving a reliable identification of a person once he has entered into the field of view of a camera.

# 4 Experimental results

In this section, we present results of multiple experiments that were run in real time on $640 \times 480$ pixels wide video sequences. Our algorithm can handle higher resolutions as well, but we haven't noticed any significant performance improvements when using higher resolutions.

Let us first determine appropriate values for all the parameters of the method. Afterward, we will present the precision of the classification on our database of 21 persons and then test our algorithm on a public database comprising videos of 25 persons.

We ran a first series of experiments on a dataset, hereafter called LAB5, which contains 4 sets of walking sequences for 5 persons. These sequences of the LAB5 data set were captured in our lab (see Figure 8) under strict and constant illumination. Videos were obtained from a consumer market webcam in order to get a realistic noise level and to ensure similar acquisition conditions to those of common situations. The goal of this set-up and this first series of videos was to determine appropriate values for the few parameters of our system.

The parameters to be refined were:

- which gait signature to use: either $\mathcal{G}^{W\times H}(i,\ j,\, t)$ or $\mathcal{G}^{W+H}(k,\, t)$,
- the numbers of bins $M$ and $N$,
- the number of frames $L$ aggregated in a single gait signature, and

Figure 8: Examples of frames of the LAB5 and LAB21 datasets captured in our lab.

- the length $V$ of the sliding temporal window used for the majority vote policy.

The decision to use $\mathcal{G}^{W\times H}(i,\ j,\ t)$ or $\mathcal{G}^{W+H}(k,\ t)$ depends on the amount of training data and memory available to the classification process. If all other parameters are kept unchanged, the use of $\mathcal{G}^{W\times H}(i,\ j,\ t)$ generally leads to better results. However, the dimensionality of the corresponding feature space is $M \times N \times L$ instead of $(M + N) \times L$. As a result, a larger amount of data is necessary to train the system and the resulting extra-trees model that has to be loaded into memory at run-time is significantly larger.

In order to determine $M$ and $N$, the numbers of bins, we tested values ranging from 2 to 40. It was observed that higher values of $M$ or $N$ (or both) generally leads to a better performance. However, the performance starts to be acceptable for 10 bins and then saturates with 20 bins and above. It is therefore recommended to use a value in the interval range [10, 20] for $M$ and $N$. Depending on the size of the training dataset and the dimensions of its silhouettes, the statistical significance of all the bins of the histograms needs to be taken into account. Indeed, from small training sets of small silhouettes, it is impossible to populate a large histogram with enough statistical significance. Consequently a value closer to 10 needs to be chosen. By contrast, larger training sets of larger silhouettes would incline us to take values of closer to 20.

A similar reasoning applies to the number of silhouettes $L$ aggregated in a gait signature: the higher, the better. Since the value of $L$ impacts on the reactivity of the system and no significant gain in performance is observed for values of $L$ larger than 20, taking $L = 20$ offers an appropriate compromise. Note that this parameter may be refined according the framerate of the cameras used. Typical cameras have a framerate of 25 images per second: $L = 20$ corresponds to a signature of about 1 second which roughly matches the length of a gait cycle. For slower framerates, $L$ has to be adapted.

The discussion regarding the appropriate value for $V$, the length of the sliding temporal window used for the majority vote policy, is again similar to the one regarding $L$. With $V$ at a high level, the results are better but the drawback is that this increases the number of frames needed to identify a person. From

a practical point of view, a majority vote regarding 10 consecutive frames is sufficient; it improves the performance of the system to a satisfactory level. If $L = 20$ and $V = 10$, the algorithm delays its answer for 30 frames, *i.e.* 1 second for commonly-used cameras.

## 4.1 Tests on a database of 21 persons

In order to estimate performance of our system, we used a second dataset, called LAB21, which was composed of 4 sets of laboratory sequences of 21 different subjects. All the classification tests were conducted by training the algorithm using 3 of the 4 sequences available for each subject and testing it on the left out one. We used the ratio of correctly classified gait signatures as a performance criterion. This ratio was computed for different numbers of frame per gait signature and for different histogram resolutions. For the sake of simplicity, we restricted ourselves to the case where $M = N$ , and disabled the majority vote on the previous $V$ frames (or to equivalently set $V$ to 1) in the first instance. This allowed us to assess the raw classification precision of the system, regardless of whether the majority vote improved the performance, as shown further on.

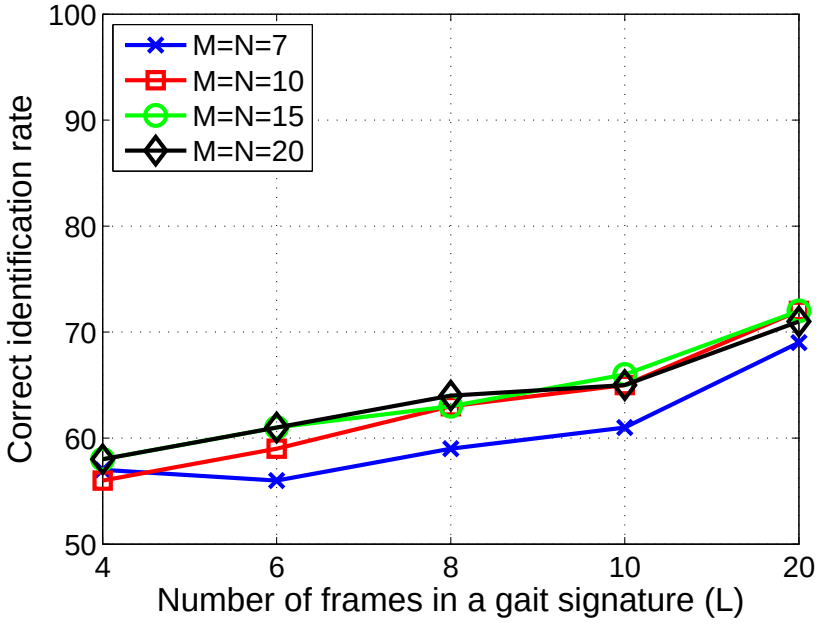
The results of the first series of tests are shown in Figure 9. The ratio of correctly classified gait signatures reached 74% for $\text{hist}^{W\times H}()$ and 72% for $\text{hist}^{W+H}()$. Both $\text{hist}^{W+H}()$ and $\text{hist}^{W\times H}()$ obtained the best results for a number of bins of 10 and a number of frames per gait signature (that is $L$) of 20. We also noticed that the performance of $\text{hist}^{W\times H}()$ was generally better than that of $\text{hist}^{W+H}()$, especially for small values of the parameters $M$, $N$, and $L$.

One could be misled by the relatively average examples of performance given by figures around the 75% mark. Remember that the examples of performance reflect all the synchronous decisions individually. Should a single class label be assigned to a test sequence as the average of the complete set of individual decisions, the performance ratio would overstep 95% of correctly classified gait sequences!

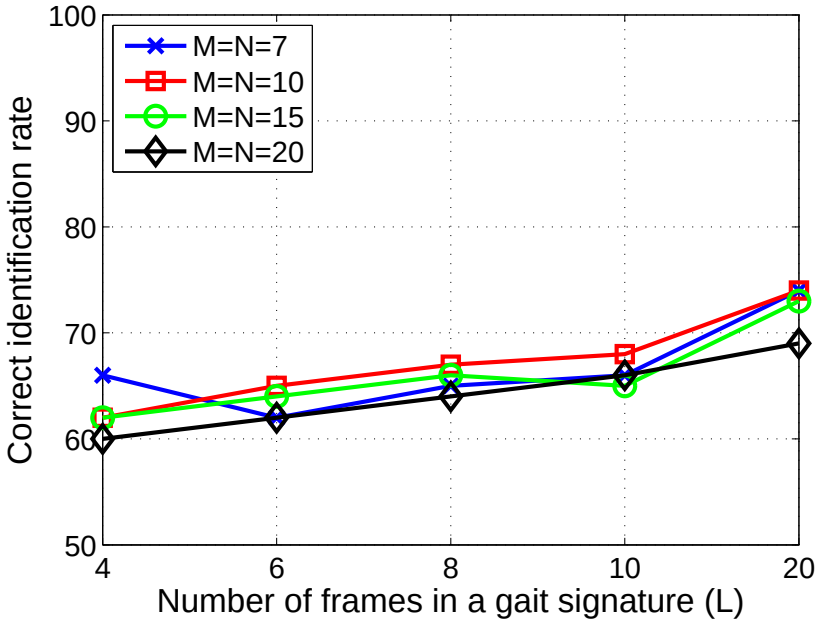
The second series of tests was limited to $\text{hist}^{W\times H}()$ in order to focus on the performance improvement brought about by the majority vote on the previous $V$ frames. The curves displayed in Figure 10 show that the use of the majority vote improves the performance of the system. For high values of $V$, the ratio of correct classifications peaks at 97%. In the same way as in the discussion on parameter $L$, we observe that an increase in the length of the majority vote time window improves precision. Interestingly, we also noticed that the choice of $M = N = 15$ outperformed the results of the choice of $M = N = 20$. This presumably originates from the small size of some silhouettes, which only contained a few wedged rectangles $\alpha$. If $\alpha$ is too small, which typically occurs when a person stands too far from the camera, it is impossible to estimate a histogram split into $20 \times 20$ bins with a good statistical significance; this poor estimation negatively impacts on performance.

## 4.2 Tests on frames acquired with surveillance cameras

The third data set used was named HW5. This consisted of frames captured with surveillance cameras located in hallways for five different persons and involving 3




(a) $\text{hist}^{W+H}(k)$




(b) $\text{hist}^{W\times H}(i,\, j)$

Figure 9: Performance of $\mathcal{G}^{W\times H}(i,\, j,\, t)$ on the LAB21 dataset with no majority vote policy (more precisely $V = 1$) using (a) $\text{hist}^{W+H}(k)$ and (b) $\text{hist}^{W\times H}(i,\, j)$.

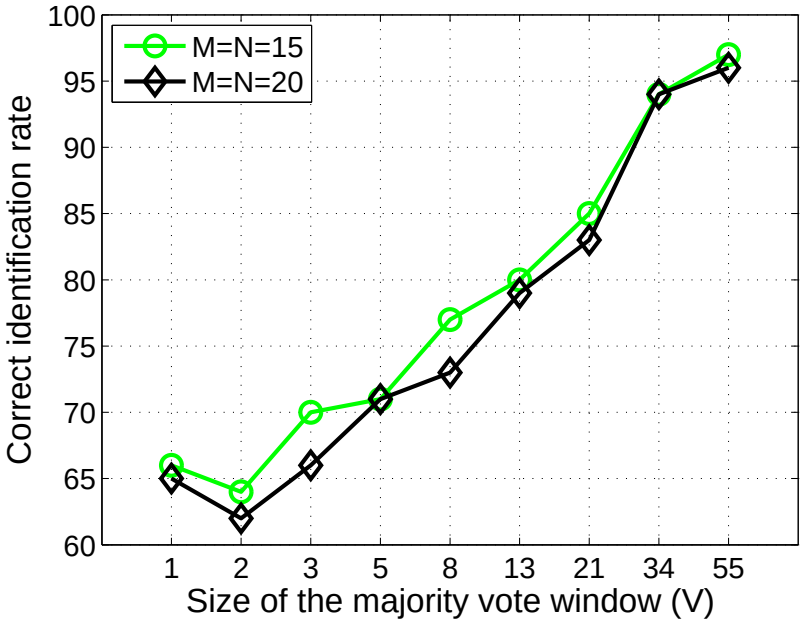



Figure 10: Performance of $\mathcal{G}^{W\times H}(i,\, j,\, t)$ on the LAB21 database using $\text{hist}^{W\times H}(i,\, j)$ for different lengths $V$ of the majority vote window ($L$ is set to 10).

sequences per person. In contrast with the previous sequences, the environment was totally unconstrained and some frames had a poor signal to noise ratio.

The results of this last series of experiments are shown in Figure 11. As expected, the precision of the classification suffered from the poor quality of the extracted silhouettes (remember the example of Figure 6). Nevertheless, thanks to the robustness of the proposed gait signature, the system still managed to identify correctly the persons in up to 81% of cases (one should compare this with the previous 97%). The 81% of correct classifications were obtained for a majority vote window of 55 frames, which corresponded to an identification delay of 2 seconds (or $L + V = 65$ frames).

### 4.3 Tests on the CMU MoBo database

To further evaluate the performance, our algorithm was tested on the publicly available MoBo database [11]. The MoBo database consists in video sequences of 25 subjects walking on a treadmill. Six calibrated and synchronized cameras were used to capture the subjects from six different viewpoints performing four different walking activities: slow walk, fast walk, incline walk, and walk with a ball. The database also comprises binary segmentation maps for each sequence. By using these segmentation maps, we are able to assess the performances of the features extraction and classification process exclusively (without any interference from the background subtraction algorithm).

To achieve a fair comparison with other techniques evaluated on the MoBo

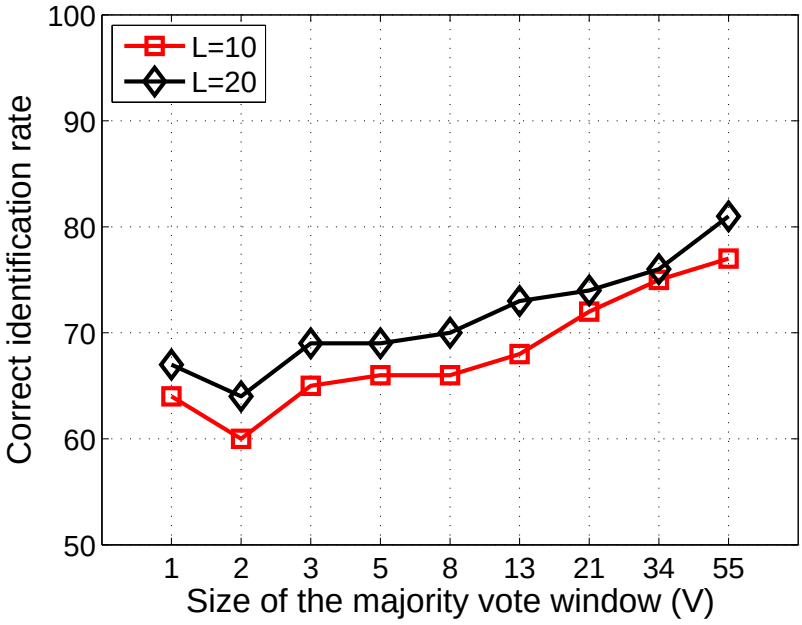



Figure 11: Performance on the HW5 dataset, which contained frames acquired with cameras located in hallways ($M$ and $N$ are set to 20).

| **Our algorithm** | **Slow** | **Fast** |
|---|---|---|
| $\mathcal{G}^{W\times H}(i,\, j,\, t)$ with $M = N = 10,\, L = 10$ | 100% | 100% |
| $\mathcal{G}^{W\times H}(i,\, j,\, t)$ with $M = N = 10,\, L = 20$ | 100% | 100% |
| $\mathcal{G}^{W\times H}(i,\, j,\, t)$ with $M = N = 20,\, L = 10$ | 100% | 100% |
| $\mathcal{G}^{W\times H}(i,\, j,\, t)$ with $M = N = 20,\, L = 20$ | 100% | 100% |

Table 1: Results obtained on non-overlapping parts of sequences from the same category of activity (training and testing sequences are both taken in the "slow walk" or "fast walk" subparts of the MoBo database).

database, we used exactly the same experimental set-up. For example, each complete walking sequence is given a unique class label; this is equivalent to setting $V$ to the total number of frames contained in the corresponding video sequence. Additionally, each sequence is divided in two non-overlapping parts of equal size. One part serves to train the algorithm, the other is used to evaluate it. We tested the method against the "slow walk" and the "fast walk" sequences separately. The results given in Table 1 show that the algorithm is able to successfully recognize every single person present in the database across the whole advised ranges of values of its parameters. For the sake of completeness, we also tested the method (with no adaptations) on the *lateral* sequences contained in the MoBo database using the same procedure. Interestingly, we observed identical scores (100% in all the cases). Future work will investigate the performance of our algorithm on lateral-view sequences.

Finally, we checked if the method was able to deal with greater discrepancies

| Comparison of two methods | **Slow/Fast** | **Fast/Slow** |
|---|---|---|
| Our algorithm: | | |
| - $\mathcal{G}^{W\times H}(i,\, j,\, t)$ with $M = N = 10,\, L = 10$ | 96% | 96% |
| - $\mathcal{G}^{W\times H}(i,\, j,\, t)$ with $M = N = 10,\, L = 20$ | 96% | 96% |
| - $\mathcal{G}^{W\times H}(i,\, j,\, t)$ with $M = N = 20,\, L = 10$ | 96% | 96% |
| - $\mathcal{G}^{W\times H}(i,\, j,\, t)$ with $M = N = 20,\, L = 20$ | 96% | 96% |
| Algorithm proposed in [13]: | | |
| - frontal view | $\emptyset$ | 88% |
| - 6 views | $\emptyset$ | 92% |
| - frontal and lateral views | $\emptyset$ | 96% |

Table 2: Results when training on one category of activity and testing on the other. Slow/Fast means that slow walking sequences were used for training while the tests were performed on fast walking sequences, and vice versa.

between training and test sequences on frontal views. Therefore our algorithm was trained on all the "slow walk" sequences and evaluated against all the "fast walk" sequences, and vice versa. From the results provided in Table 2, we see that the algorithm is able to successfully recognize persons even if the walking speed changes between the training and the testing steps. We also notice that our method outperforms that of [13] when using a single frontal camera; the best classification score presented in [13] was obtained by combining the frontal and the lateral views. In our case, sequences acquired with a single frontal camera suffice to produce the best recognition scores.

# 5 Conclusions

Gait identification is currently an intensive topic for research. Most techniques described in the literature are based on lateral views of walking persons. It is known that lateral views contain appropriate information regarding the gait. However, using lateral views in indoor environments might be unfeasible, especially in hallways where a frontal view is almost inevitable.

This paper proposes a real-time *frontal-view* gait recognition system. A major contribution is introduced by defining a gait signature of a walking person. Successive binary silhouettes are extracted with a background subtraction algorithm. Each silhouette is then converted to an intra-frame histogram which compacts the width and height distributions of the set of all the rectangles that can be wedged inside the silhouette. Afterward, a given number $L$ of successive histograms is combined into a single spatio-temporal (inter-frame) gait signature. The identification of the persons is then computed by a classification of this signature by a machine learning algorithm called extra-trees. Finally, successive decisions are combined along several frames using a majority vote policy to determine the identity of the person currently present in the field of view of the camera.

Four series of experiments were conducted on different databases. The first series helped to determine the parameter values needed to optimize the performance of the overall system. The second series was intended to evaluate the precision of the classification for different ranges of values of the parame-

ters. It was shown that the ratio of correct classifications could reach 97% for a database of 21 persons. The third series of experiments served as a showcase for a practical scenario. Frames were captured with hallway surveillance cameras at our institute. Despite the noise and the unavoidable phenomena in such an unconstrained environment, the system was still able to identify the persons correctly in up to 81% of cases. Finally, we tested our algorithm on the publicly available MoBo database. Our method was able to successfully recognize the persons from video sequences taken in the MoBo database reaching a score as high as 96% to 100%, depending on the training and testing conditions.

**Acknowledgments**

The authors are grateful to Steve Frécinaux for his invaluable help in implementing and testing the algorithm.

# References

[1] O. Barnich, S. Jodogne, and M. Van Droogenbroeck. *Robust analysis of silhouettes by morphological size distributions*, volume 4179 of *Lecture Notes on Computer Science*, pages 734–745. Springer Verlag, 2006.

[2] C. BenAbdelkader, R. Cutler R., and L. Davis. Motion-based recognition of people in EigenGait space. *Proceedings of Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 267–272, May 2002.

[3] N.V. Boulgouris, D. Hatzinakos, and K.N. Plataniotis. Gait recognition: a challenging signal processing technology for biometric identification. *IEEE Signal Processing Magazine*, 22(6):78–90, 2005.

[4] N.V. Boulgouris, K.N. Plataniotis, and D. Hatzinakos. Gait recognition using linear time normalization. *Pattern Recognition*, 39(5):969–979, 2006.

[5] L. Breiman. Bagging predictors. *Machine Learning*, 26(2):123–140, 1996.

[6] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[7] D. Cunado, M. Nixon, and J. Carter. Automatic gait recognition via model-based evidence gathering. *Proceedings of IEEE Workshop on Automated ID Technologies (AutoID99)*, pages 27–30, 1999.

[8] J. Foster, M. Nixon, and A. Prügel-Bennett. Automatic gait recognition using area-based metrics. *Pattern Recognition Letters*, 24(14):2489–2497, 2003.

[9] P. Geurts, D. Ernst, and L. Wehenkel. Extremely randomized trees. *Machine Learning*, 36(1):3–42, 2006.

[10] W. Grimson. Gait analysis for recognition and classification. In *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition (FGR 02)*, pages 155–161, Washington, DC, USA, 2002. IEEE Computer Society.

[11] R. Gross and J. Shi. The CMU motion of body (MoBo) database. Technical Report CMU-RI-TR-01-18, Robotics Institute, Pittsburgh, PA, June 2001.

[12] W. Hu, T. Tan, L. Wang, and S. Maybank. A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man, and Cybernetics*, 34(3):334–352, August 2004.

[13] X. Huang and V. Boulgouris. Human gait recognition based on multiview gait sequences. *EURASIP Journal on Advances in Signal Processing*, 2008:8 pages, 2008.

[14] A. Kale, N. Cuntoor, B. Yegnanarayana, A. Rajagopalan, and R. Chellappa. Gait analysis for human identification. In *Proceedings of the International Conference on Audio-and Video-Based Person Authentication*, pages 706–714, Guildford, UK, 2003.

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

[16] S. Lee, Y. Liu, and R. Collins. Shape variation-based frieze pattern for robust gait recognition. *IEEE International Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007.

[17] Y. Liu, R. Collins, and Y. Tsin. Gait sequence analysis using frieze patterns. In *Proceedings of the 7th European Conference on Computer Vision-Part II*, pages 657–671, London, UK, 2002. Springer-Verlag.

[18] Z. Liu and S. Sarkar. Outdoor recognition at a distance by fusing gait and face. *Image Vision Computing*, 25(6):817–832, 2007.

[19] S. Mowbray and M. Nixon. Automatic gait recognition via fourier descriptors of deformable objects. In J. Kittler and M. Nixon, editors, *Audio Visual Biometric Person Authentication*, pages 566–573. Springer, 2003.

[20] M. Nixon, J. Carter, J. Shutler, and M. Grant. New advances in automatic gait recognition. *Elsevier Information Security Technical Report*, 7(4):23–35, 2002.

[21] M. Nixon, T. Tan, and R. Chellappa. *Human identification based on gait.* Springer, 2006.

[22] S. Niyogi and E. Adelson. Analyzing and recognizing walking figures in XYT. *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 469–474, June 1994.

[23] J. Serra. *Image analysis and mathematical morphology.* Academic Press, New York, 1982.

[24] M. Soriano, A. Araullo, and C. Saloma. Curve spreads: a biometric from front-view gait video. *Pattern Recognition Letters*, 25(14):1595–1602, 2004.

[25] C. Stauffer and E. Grimson. Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):747–757, 2000.

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

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

[28] S. Zhou, R. Chellappa, and W. Zhao. *Unconstrained Face Recognition.* Springer, 2006.

[29] X. Zhou and B. Bhanu. Feature fusion of side face and gait for video-based human identification. *Pattern Recognition*, 41(3):778–795, 2008.

[30] Z. Zivkovic. Improved adaptive gausian mixture model for background subtraction. In *Proceedings of the 17th International Conference on Pattern Recognition*, volume 2, pages 28–31, 2004.