Classification of Gait Types Based on the Duty-factor

Fihl, Preben; Moeslund, Thomas B.

Published in:
IEEE Conference on Advanced Video and Signal Based Surveillance, 2007. AVSS 2007

DOI (link to publication from Publisher):
10.1109/AVSS.2007.4425330

Publication date:
2007

Document Version
Publisher's PDF, also known as Version of record

CLASSIFICATION OF GAIT TYPES BASED ON THE DUTY-FACTOR

Preben Fihl and Thomas B. Moeslund
Laboratory of Computer Vision and Media Technology
Aalborg University, Denmark
{pfa,tbm}@cvmt.dk

Abstract

This paper deals with classification of human gait types based on the notion that different gait types are in fact different types of locomotion, i.e., running is not simply walking done faster. We present the ***duty-factor****, which is a descriptor based on this notion. The duty-factor is independent on the speed of the human, the cameras setup etc. and hence a robust descriptor for gait classification. The duty-factor is basically a matter of measuring the ground support of the feet with respect to the stride. We estimate this by comparing the incoming silhouettes to a database of silhouettes with known ground support. Silhouettes are extracted using the Codebook method and represented using Shape Contexts. The matching with database silhouettes is done using the Hungarian method. While manually estimated duty-factors show a clear classification the presented system contains misclassifications due to silhouette noise and ambiguities in the database silhouettes.*

1. Introduction

The constant reduction in the price of surveillance cameras *and* the notion that more surveillance equals more safety have lead to a very large number of surveillance cameras mounted in both public and private spaces. Since it is unrealistic to have personnel watching and analyzing the extreme amount of video being produced, a massive amount of research (both public and private) is currently ongoing towards automating the analysis of surveillance video [11]. An application domain with a huge potential is the ability to automatic annotate videos in terms of e.g., detecting the presence of humans [3, 19], their ID [21], and their activities [14, 16].

One of the most basic but also most important type of activities to be recognized is gait[1]. And humans in surveillance videos spend a significant portion of their time doing gait. Different gait types can be recognized by the speed of an individual given knowledge of the surveillance context (frame rate, camera calibration, environment lay-out, etc.) and the motion pattern of the human, e.g., locomotion parallel to the camera plan. However, given a video sequence with unknown context and unknown motion pattern computer vision systems cannot simply infer the gait type from the speed of the individual. But a human observer can, in general, easily do this, so some other descriptor independent of the speed must exist.

1.1. The Duty-Factor

When a human wants to move fast he/she will run. Running is not simply walking done fast and the different types of gaits are in fact different actions. This is true for vertebrates in general. For example, birds and bats have two distinct flying actions and horses have three different types of gaits. Which action to apply to obtain a certain speed is determined by minimizing some physiological property and physiological research has shown that the optimum action changes discontinuously with changing speed. For example, turtles seem to optimize with respect to muscle power, horses and humans with respect to oxygen consumption and other animals by minimizing metabolic power. [1]

From a computer vision point of view the question is now if *one* (recognizable) descriptor exist, which can represent the continuum of gait, i.e., the different types. For bipedal locomotion, in general, the *duty-factor* can do exactly this. The duty-factor is defined as ”the fraction of the duration of a stride for which each foot remains on the ground” [2]. To illustrate the power of this descriptor we have manually estimated the duty-factor in 148 gait sequences containing humans walking, jogging, or running, see figure 1. These sequences come from 5 different sources and contain many different individuals entering and exiting at different angles. Some not even following a straight line. In figure 2 example images from the sequences are shown.

Figure 1 shows a very clear separation between walking and jogging/running which is in accordance with the fact that those types of gait are in fact different ways of moving. Jogging and running however, cannot be separated as clearly and there is a gradual transition from one gait type to the other. In fact, the classification of jogging and running

[1] By gait is meant bipedal locomotion like walking, jogging and running. In the human identification literature gait is often referred to as a style of walking whereas we adopt the more general definition of gait.

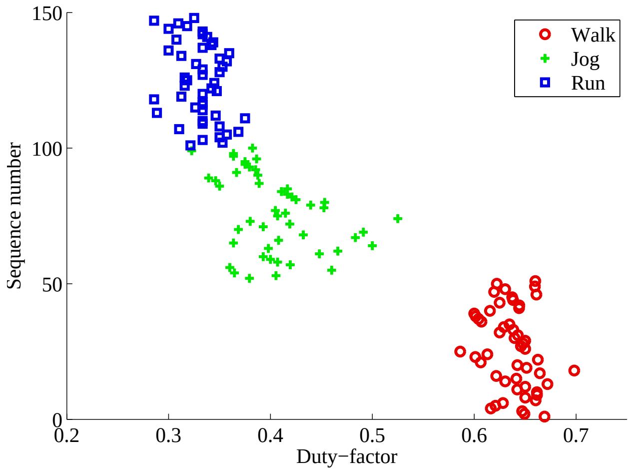

Figure 1: *The manually annotated duty-factor and gait type for 148 different sequences. Note that the sole purpose of the y-axis is to spread out the data.*

is dependent on the observer when considering movements in the transition face and there exists no clear definition of what separates jogging from running. This problem is apparent in the classification of the sequences used in figure 1. Each sequence is either classified by us or comes from a data set where it has been labeled by others. By having more people classify the same sequences it turns out that the classification of some sequences is ambiguous which illustrates the subjectivity in evaluation of jogging and running[2]. However, the duty-factor is still a good descriptor for dividing the continuum of the two gait types, and it may even be used to define an objective separation of the two.

1.2. Focus of the Paper

To automatically estimate the duty-factor we need a method to determine how much ground support a person has at a given time, i.e. does he/she have two feet, one foot or no feet in contact with the ground. The method needs to be able to handle an unconstrained camera setup without knowledge about the scene to fully exploit the potential of the duty-factor as a descriptor of different types of human gait.

A large number of computer vision-based methods have been proposed that use a human model to estimate the pose of a person or to recognize actions, e.g. [11, 14, 16]. However, to detect the duty-factor we do not need the exact pose of a person so we choose a model-free method based on silhouette data.

Silhouette based methods have been used with success in the area of human identification by gait [6, 10]. In such systems the goal is extraction of features that describe the personal variation in gait patterns often allowing variation in walking speed and in [20] also allowing people to run. Inspired by the ability of the silhouette based approaches to describe details in gait we propose such a method but since our goal is quite different from human identification we extract the duty-factor that allows personal variation but describe the different gait types.

Our approach is to extract silhouettes, make a compact representation of these and compare them to a database of computer generated silhouettes. For each of the database silhouettes we know the ground support and can hence infer the duty-factor. This yields the five blocks shown in figure 3. The following sections describe each block in detail.

[2]The problem of ambiguous classification will be clear when watching for example video sequences from the KTH data set (http://www.nada.kth.se/cvap/actions/) and the Weizmann data set (http://www.wisdom.weizmann.ac.il/ vision/SpaceTimeActions.html).

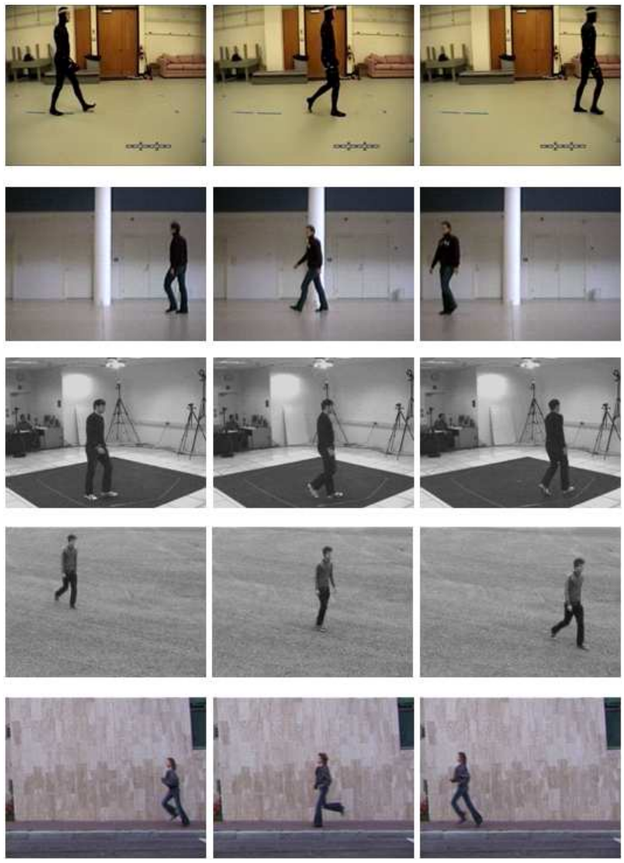

Figure 2: *Samples from 5 different data sets. First row: CMU Graphics Lab Motion Capture Database (http://mocap.cs.cmu.edu/). Second row: our own data set. Third row: the HumanEVA-I data set [17]. Fourth row: the KTH data set [15]. Last row: the Weizmann data set [5].*

2. Silhouette Database

To extract the duty-factor we create a database of human silhouettes performing the main gait types walk, jog, and

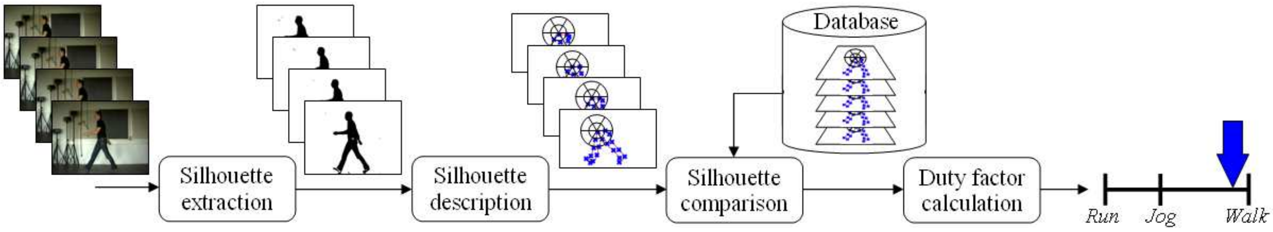

Figure 3: *An overview of the approach.*

run. The creation of a database that would cover all possible variations of the three gait types would be impractical (if not impossible) so we let the database cover only a single typical execution of each type. Furthermore, we create this typical execution by animating a computer graphics model of a human to perform a lifelike execution of the gait types[3] and thus avoiding extensive statistical analysis of humans performing the gait types. In this way we cover walk, jog, and run with 90 silhouettes in the database.

Each silhouette in the database is annotated with the number of feet in contact with the ground which is the basis of the duty-factor calculation.

We make the method invariant to some change in viewpoint by generating database silhouettes from three different camera angles. This is very easily done with the 3D rendering software and again, this does not require capturing of new real life data for statistical analysis. The database contains silhouettes of the human model seen from a side view and from cameras rotated 30 degrees to both sides, see figure 4. The three camera angles allow us to match database silhouettes with silhouettes of people moving at angles of at least ± 45 degrees with respect to the viewing direction. Each new input silhouette will be matched to database silhouettes from all camera angles and the method is therefore also invariant to changes in moving direction during a sequence, see figure 2 row two and three.

Figure 4: *The database contains silhouettes of the human model seen from three camera angles: the model seen from the side (middle) and seen from a camera rotated 30 degrees to each side.*

[3]The animation of the gait types are done in the 3D rendering software Poser 6[18] with the Walk Designer, see figure 4.

We do not store the silhouettes directly in the database but rather we extract the silhouettes of the legs and describe these with scale and translation invariant features which are then stored in the database. This process is described further in section 4.

3. Silhouette Extraction

We detect people moving in the scene by doing background subtraction. We use the Codebook background subtraction method as described in [7] and [8] which handles shadows and foreground camouflage by separating intensity and chromaticity in the background model. Furthermore, the background model is multi modal which allows it to model moving backgrounds like tree branches. To keep the background model consistent at all times [7] describe two different update mechanisms that handle rapid and gradual changes respectively. This robust background subtraction method allows us to use quite diverse video sequences from both indoor and outdoor scenarios as input.

4. Silhouette Description

When a person is moving around in a typical surveillance setup his or her arms will not necessarily swing in a typical "walk" manner but might be doing different gestures, e.g. hand waving, or the person might be carrying an object. To circumvent the variability and complexity of such scenarios we chose to analyse the gait and estimate the duty-factor solely on the silhouette of the legs. Furthermore, [9] shows that identification of people based on gait can perform equally well using either the silhouette of the legs or the silhouette of the whole person. The silhouette of the legs is defined as the bottom half of the silhouette[4].

To allow recognition of gait types across different scales we describe the leg silhouette in a scale and translation invariant manner. Recognition at different scales is required since we do not make assumptions about the distance between people and the camera, and since people way move

[4]A more anatomically correct division of the silhouette would be 55% for the legs and 45% for the upper body [13]. However, to avoid noise in the leg silhouette due to hands swinging below the 55%-line we chose 50% of the silhouette height as the division line without loss of generality.

closer to or further away from the camera when their path is not perpendicular to the viewing direction.

We use shape contexts [4] to describe the leg silhouettes. n points are sampled from the contour of the leg silhouette and for each point we determine the shape context and the tangent orientation at that point. With K bins in the log-polar histogram of the shape context we get an $n \times (K+1)$ matrix describing each silhouette. Scale invariance is achieved with shape contexts by normalizing the radial distances of the histogram by the mean distance between all point pairs on the contour.

5. Silhouette Comparison

With both database silhouettes and input silhouettes described by similar matrices the task is now to find the best match between input and database silhouettes. We follow the method of [4] and calculate the cost of matching a point on the input silhouette with a point on a database silhouette using the χ^2 test statistics. The cost of matching the shape contexts of point p_i on one silhouette and point p_j on the other silhouette is denoted $c_{i,j}$. The normalized histograms at points p_i and p_j are denoted $h_i(k)$ and $h_j(k)$ respectively with k as the bin number, $k = 1, 2, ..., K$. The χ^2 test statistics is given as

$$c_{i,j} = \frac{1}{2} \sum_{k=1}^{K} \frac{[h_i(k) - h_j(k)]^2}{h_i(k) + h_j(k)} \quad (1)$$

The difference in tangent orientation $\theta_{i,j}$ between points p_i and p_j is added to $c_{i,j}$ to give the final cost of matching the two points $C_{i,j}$ ($\theta_{i,j}$ is in the interval 0 to π).

The costs of matching all point pairs between the two silhouettes are calculated. Finding the one-to-one mapping between the two point sets that minimizes the total cost of matching is a square assignment problem and can be solved with the Hungarian method [12].

By finding the minimum cost of matching the input silhouette to each of the database silhouettes we can now identify the best match by taking the database silhouette with the lowest total cost.

6. Duty-Factor Calculation

As stated earlier the duty-factor is defined as the fraction of the duration of a stride for which each foot remains on the ground so we need to identify the duration of a stride and for how long each foot is in contact with the ground.

A stride is defined as one complete walk cycle and consists of two steps. A stride can be identified as the motion from a left foot takeoff (the foot leaves the ground) and until the next left foot takeoff. Accordingly a step can be identified as the motion from a left foot takeoff to the next right foot takeoff. Given this definition of a step it is natural to identify steps in the video sequence by use of the silhouette width. From a side view the silhouette width of a walking person will oscillate in a periodic manner with peaks corresponding to silhouettes with the feet furthest apart. The interval between two peaks will (to a close approximation) define one step [6]. This also holds for jogging and running and can furthermore be applied to situations with people moving diagonally with respect to the viewing direction. By extracting the silhouette width from each frame of a video sequence we can identify each step (peaks) and hence determine the mean duration of a stride t_s in that sequence.

For how long each foot remains on the ground can be estimated by looking at the database silhouettes that have been matched to the sequence. Since each database silhouette is annotated with the number of feet supported on the ground the total ground support G of both feet for a video sequence is the sum of ground support of all the matched database silhouettes. However, with respect to ground support the database does contain few ambiguities as illustrated in figure 5 where a silhouette from jogging with no ground support (left) is similar to a silhouette from walk with both feet on the ground (right). To reduce errors resulting from such ambiguities we note that the ground support never changes between 0 and 2 feet in two consecutive frames and we can therefore filter out most of those occurrences.

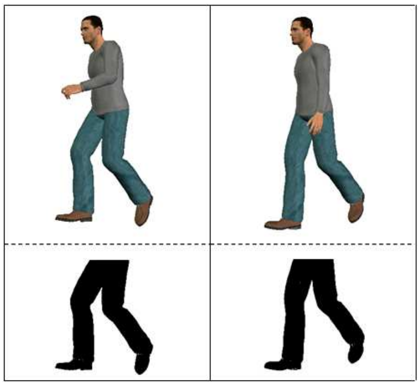

Figure 5: *Two similar silhouettes from the database with different ground support. The top row shows the pose of the human model and the bottom row is the database silhouettes. Left: silhouette from jogging with ground support=0. Right: silhouette from walking with ground support=2.*

To get the ground support for each foot we assume a normal moving pattern (not limping, dragging one leg, etc.) so the left and right foot have equal ground support and the mean ground support g for each foot during one stride is $\frac{G}{2 \cdot n_s}$, where n_s is the number of strides in the sequence. The duty-factor D is now given as $D = \frac{g}{t_s}$. In summary we have

$$\text{Duty-factor } D = \frac{G}{2 \cdot n_s \cdot t_s} \quad (2)$$

where G is the total ground support, n_s is the number of strides, and t_s is the mean duration of a stride in the sequence.

7. Results

To test our method we estimate the duty-factor automatically in 131 sequences from different data sets covering indoor and outdoor video, different moving directions with respect to the camera (covering ± 45 degrees from the viewing direction), different video resolutions, 22 different people, and varying silhouette heights (from 41 pixels to 454 pixels). For the silhouette description the number of sampled points n was 100 and the number of bins in the shape contexts K was 60. Figure 6 shows the estimated duty-factor from the test sequences. Figure 7 shows the confusion matrix when classifying the sequences based on the estimated duty-factor[5]. The classification results in 81.7% correct classifications.

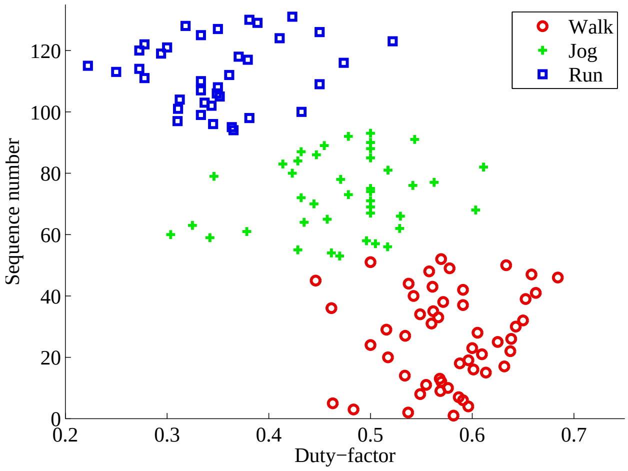

Figure 6: *The estimated duty-factor from 131 test sequences. The y-axis solely spreads out the data.*

	Walk	Jog	Run
Walk	44	8	0
Jog	5	31	5
Run	0	6	32

a) Number of occurrences

	Walk	Jog	Run
Walk	84.6	15.4	
Jog	12.2	75.6	12.2
Run		15.8	84.2

b) Occurrences in percent

Figure 7: *The results of the classification listed in confusion matrices. The rows represent the true classes and the columns show the classification.*

[5]The sequences of figure 6 are a subset of the manually annotated sequences of figure 1. Reliable foreground detection were not possible for some sequences due to an unsteady camera, strong shadows, and heavy foreground camouflage. Since backgorund subtraction is not the focus of this paper we left those sequences out.

8. Discussion

Although the duty-factor is a suitable feature for classifying different types of gait the accuracy of the method at its current state results in some misclassification. The variation in gait from one person to the next causes difficulties when using a single prototype of each gait type. Furthermore, human observers that classify jog and run (including the people that have labeled the data sets) will also experience ambiguities. Aside from these circumstances the errors are caused by problems like poor silhouette segmentation and small silhouettes (heights below 65 pixels) especially in short sequences (eg. sequences containing only one step). The ambiguities of the database silhouettes does also result in some misclassifications especially in sequences where the moving direction is near ± 45 degrees. The lack of perfect separation between walk and jog which could be expected from the manually annotaded data (see figure 1) is mainly due to database ambiguities. An analysis of these ambiguities could allow us to handle a number of misclassification with improved recognition rate as a result.

We compare our result to similar approaches [5] and [15]. Note that parts of the data sets from these papers are included in our data set. [5] uses space-time shapes to classify 9 different human actions including walk and run. They achieve perfect recognition[6] in 81 video sequences of 9 different people from a side view, see figure 2 row five for an example. The running sequences do contain personal variation but no attempt is made to classify jogging movement. Comparing our result to [5] we would achieve an recognition rate of 96.5% on a larger data set including non-side views if we were to leave the jogging sequences out of consideration. In [15] 6 human actions are classified including walk, jog, and run using local support vector machines. Their data set contains both indoor and outdoor sequences with 25 different people moving at different directions (side view and $\pm 30 - 45$ degrees from the viewing direction), see figure 2 row four for an example. The data set also contains three hand actions (boxing, hand waving, and hand clapping) but for the results reported in [15] the gait actions are not confused with the three hand actions which makes our results comparable. [15] achieve recognition rates of 83.8%, 60.4%, and 54.9% recognizing walk, jog, and run respectively. Our method performs better for all gait types and especially for jog and run (see figure 7). [15] also states that personal variation of the gait types makes especially running and jogging similar in some cases causing errors in the classification.

To further explore the capabilities of the duty-factor we manually extracted the duty-factor for 5 persons doing 5 repetitions of each gait type (se figure 8). This shows that a better classification of gait types based on the duty-factor

[6]Their method misclassify 1 out of 549 space-time cubes.

could be expected in applications where the identity of the person is known or if some of the personal variation could by eliminated by choosing appropriate features for describing the movement.

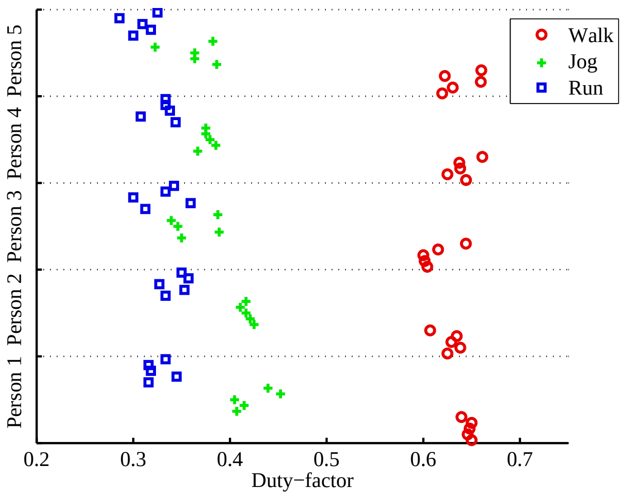

Figure 8: *The duty-factor of 5 different people performing each gait type 5 times.*

9. Conclusion

In this paper we have introduced the duty-factor as a new simple measure to characterize gait types such as walk, jog, and run. The characterization is independent of the context, e.g. frame rate and walking direction. In fact people can change moving direction during a scene as seen in figure 2. We also present an automatic system for classification of gait types based on the duty-factor. The recognition rates of the automatic system does not completely match the manual classification (figure 1) but further advancements of the method is believed to improve recognition rates. Currently we are working on issues like estimating the vertical distance from the ground to the feet to eliminate ambiguities like seen in figure 5 and to incorporate the inherent periodicity of the ground support in human motion.

Acknowledgements

This work is supported by the EU project HERMES (FP6 IST-027110).

References

[1] R. Alexander. Optimization and Gaits in the Locomotion of Vertebrates. *Physiological Reviews*, 69(4), 1989.

[2] R. Alexander. Energetics and Optimization of Human Walking and Running: The 2000 Raymond Pearl Memorial Lecture. *American Journal of Human Biology*, 14, 2002.

[3] B. Leibe and E. Seemann and B. Schiele. Pedestrian detection in crowded scenes. In *CVPR*, San Diego, Ca, USA, June 20-25 2005.

[4] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *PAMI*, 24(4), 2002.

[5] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri. Actions as space-time shapes. In *ICCV*, Washington, DC, USA, 2005.

[6] R. Collins, R. Gross, and J. Shi. Silhouette-based human identification from body shape and gait. In *Fifth IEEE Int. Conf. on Automatic Face and Gesture Recognition*, Washington, DC, May 20-21 2002.

[7] P. Fihl, R. Corlin, S. Park, T. Moeslund, and M. Trivedi. Tracking of Individuals in Very Long Video Sequences. In *Int. Symposium on Visual Computing*, Lake Tahoe, Nevada, USA, November 6-8 2006.

[8] K. Kim, T. Chalidabhongse, D. Harwood, and L. Davis. Real-time Foreground-Background Segmentation using Codebook Model. *Real-time Imaging*, 11(3):167–256, 2005.

[9] Z. Liu, L. Malave, A. Osuntugun, P. Sudhakar, and S. Sarkar. Towards understanding the limits of gait recognition. In *Int. Symposium on Defense and Security*, Orlando, Florida, USA, April 12-16 2004.

[10] Z. Liu and S. Sarkar. Improved gait recognition by gait dynamics normalization. *PAMI*, 28(6), 2006.

[11] T. Moeslund, A. Hilton, and V. Krüger. A Survey of Advances in Vision-Based Human Motion Capture and Analysis. *Journal of Computer Vision and Image Understanding*, 104(2-3), 2006.

[12] C. H. Papadimitriou and K. Steiglitz. *Combinatorial optimization: algorithms and complexity*. Courier Dover Publications, Mineola, NY, USA, 1998.

[13] S. Park and J. K. Aggarwal. Segmentation and tracking of interacting human body parts under occlusion and shadowing. In *MOTION '02: Proceedings of Workshop on Motion and Video Computing*, Washington, DC, USA, 2002.

[14] I. Rius, J. Varona, J. Gonzalez, and J. J. Villanueva. Action spaces for efficient bayesian tracking of human motion. In *ICPR*, pages 472–475, Hong Kong, 2006.

[15] C. Schüldt, I. Laptev, and B. Caputo. Recognizing human actions: a local SVM approach. In *ICPR*, Cambridge, U.K, 2004.

[16] Y. Sheikh, M. Sheikh, and M. Shah. Exploring the space of a human action. In *ICCV*, Washington, DC, USA, 2005.

[17] L. Sigal and M. J. Black. Humaneva: Synchronized video and motion capture dataset for evaluation of articulated human motion. In *Technical Report CS-06-08*, Brown University, Providence, USA, 2006.

[18] *http://www.e-frontier.com/go/poser/*, February 2007.

[19] P. Viola, M. Jones, and D. Snow. Detecting Pedestrians Using Patterns of Motion and Appearance. *International Journal of Computer Vision*, 63(2), 2005.

[20] C. Y. Yam, M. S. Nixon, and J. N. Carter. On the relationship of human walking and running: Automatic person identification by gait. In *ICPR*, Washington, DC, USA, 2002.

[21] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 35(4), 2003.