Recognizing Human Gait Types

Chapter · March 2010

DOI: 10.5772/9293 · Source: InTech

Thomas B. Moeslund
Aalborg University

11

Recognizing Human Gait Types

Preben Fihl and Thomas B. Moeslund
Aalborg University
Denmark

1. Introduction

Everyday people will observe, analyze, and interpret the motion and actions of the people surrounding them. This is a source of very valuable information about not only what each person is doing but also things like their intentions, their attitude towards the observer, situations they perceive as dangerous or interesting, etc. In open spaces where people are moving around the type of motion will be an important cue for a lot of this information. More specifically, the human gait has been actively investigated in different research areas for this reason. Within psychology the expression and perception of emotions through styles of motions has been investigated by e.g. (Montepare et al., 1987) and the question of how people infer intention from actions has been studied within neuroscience by e.g. (Blakemore & Decety, 2001). In biomechanics (Alexander, 2002) describes how the choice of different gait types (walking and running) are based on the minimizing of energy consumption in muscles and (Whittle, 2001) describes how gait analysis can be used within clinical diagnostics to diagnose a number of diseases.

A lot of the information that is embedded in the gait can be extracted by simply observing a person. Systems that operate around people can benefit greatly from such observations. This fact has driven much research within robotics and computer vision to focus on analysis of human gait with a number of different applications as the aim.

Robots that move and work around humans will be very dependant on their ability to observe people and to interact with humans in an efficient way and the ability to recognize basic human activities is furthermore necessary. Methods for recognizing human gestures to enable natural human-robot interaction has been presented in e.g. (Yang et al., 2006; Meisner et al., 2009; Waldherr et al., 2000). Natural human-robot interaction also requires the robot to behave in a way that is in accordance with the social rules of humans. A method for adapting the robot behavior according to the motion of people is presented in (Svenstrup et al., 2009). Since the human gait is a very distinctive type of motion it can be used in many contexts to detect the presence of people, e.g. from surveillance cameras (Cutler & Davis, 2000; Ran et al., 2007; Viola et al., 2005). Gait as a biometric measure has also received much attention because it is non-intrusive (Collins et al., 2002; Liu & Sarkar, 2006; Veeraraghavan et al., 2005; Wang et al., 2004; Yam et al., 2002). Finally, there has been considerable interest in the computer vision community in the classification of gait types or, more generally, of different types of human action (Blank et al., 2005; Dollár et al., 2005; Schüldt et al., 2004). The research in human action recognition is applicable in a number of areas besides human-

robot interaction, e.g. in advanced user interfaces, annotation of video data, intelligent vehicles, and automatic surveillance.
An interesting and challenging area of human gait type recognition is motion in open spaces like town squares, courtyards, or train stations where one of the main human activities is that of gait, i.e. people are walking, jogging, or running. The movements of people in such spaces are however rarely constrained so seen from a camera this will result in challenges like changing direction of motion, significant changes in the scale of people, varying speeds of motion, and often also dynamics backgrounds. This chapter will show how to build a gait type classification system that can handle a number of the challenges that a real life scenario imposes on such a gait classification system. i.e. a general system which is *invariant* to camera frame rate and calibration, view point, moving speeds, scale change, and non-linear paths of motion.
Much research concerned with gait attempts to extract features related to the person specific style of gait whereas this work is concerned with the three general types of gait (walking, jogging and running) and it is therefore more related to the action recognition research than the research on the use of gait in personal identification.
Systems that are invariant to one or more of the factors listed above have been presented in the literature, but so far none has considered all these factors simultaneously. (Masoud & Papanikolopoulos, 2003) presents good results on classification of different types of human motion but the system is limited to motion parallel to the image plane. (Robertson & Reid, 2005) describes a method for behavior understanding by combining actions into human behavior. The method handles rather unconstrained scenes but uses the moving speed of people to classify the action being performed. The moving speed cannot be used for gait-type classification. A person jogging along could easily be moving slower than another person walking fast and human observers distinguishing jogging from running do typically not use the speed as a feature. Furthermore, estimation of speed would require scene knowledge that is not always accessible. (Blank et al., 2005) uses space-time shapes to recognize actions independently of speed. The method is robust to different viewpoints but cannot cope with non-linear paths created by changes in direction of movement. Other state-of-the-art approaches are mentioned in section 8 along with a comparison of results.
Current approaches to action classification and gait-type classification consider two or three distinct gait classes, e.g. (Masoud & Papanikolopoulos, 2003; Robertson & Reid, 2005) who consider walking and running, or (Blank et al., 2005; Dollár et al., 2005; Schüldt et al., 2004) who consider walking, jogging, and running. However, this distinct classification is not always possible, not even to human observers, and we therefore extend the gait analysis with a more appropriate gait continuum description. Considering gait as a continuum seems intuitive correct for jogging and running, and including walking in such a continuum makes it possible to apply a single descriptor for the whole range of gait types. In this chapter we present a formal description of a gait continuum based on a visual recognizable physical feature instead of e.g. a mixture of probabilities of walking, jogging, and running.

1.1 Gait type description based on the Duty-factor

The work presented in this chapter describe the major gait types in a unified gait continuum using the *duty-factor* which is a well established property of gait adopted from the biomechanics literature (Alexander, 2002). To enhance the precision in estimation of the duty-factor we use an effective gait type classifier to reduce the solution space and then

calculate the duty-factor within this subspace. The following section will elaborate and motivate our approach.
A current trend in computer vision approaches that deal with analysis of human movement is to use massive amounts of training data, which means spending a lot of time on extracting and annotating the data and temporally aligning the training sequences. To circumvent these problems an alternative approach can be applied in which computer graphics models are used to generate training data. The advantages of this are very fast training plus the ability to easily generate training data from new viewpoints by changing the camera angle.
In classifying gait types it is not necessary to record a person's exact pose, and silhouettes are therefore sufficient as inputs. Silhouette based methods have been used with success in the area of human identification by gait (Collins et al., 2002; Liu & Sarkar, 2006; Wang et al., 2004). The goal in human identification is to extract features that describe the personal variation in gait patterns. The features used are often chosen so that they are invariant to the walking speed and in (Yam et al., 2002) the same set of features even describe the personal variation in gait patterns of people no matter whether they are walking or running. Inspired by the ability of the silhouette based approaches to describe details in gait, we propose a similar method. Our goal is however quite different from human identification since we want to allow personal variation and describe the different gait types through the duty-factor.
A silhouette based approach does not need a completely realistic looking computer graphics model as long as the shape is correct and the 3D rendering software Poser[1], which has a build-in Walk Designer, can be used to animate human gaits.
To sum up, our approach offers the following three main contributions.

1. The methods applied are chosen and developed to allow for classification in an unconstrained environment. This results in a system that is invariant to more factors than other approaches, i.e. invariant in regard to camera frame rate and calibration, viewpoint, moving speeds, scale change, and non-linear paths of motion.
2. The use of the computer graphics model decouples the training set completely from the test set. Usually methods are tested on data similar to the training set, whereas we train on computer-generated images and test on video data from several different data sets. This is a more challenging task and it makes the system more independent of the type of input data and therefore increases the applicability of the system.
3. The gait continuum is based on a well-established physical property of gait. The duty-factor allows us to describe the whole range of gait types with a single parameter and to extract information that is not dependant on the partially subjective notion of jogging and running.

The remainder of this chapter will first give a thorough introduction of the duty-factor and show its descriptive power. Next, the gait classification frame work will be described in detail. The framework is shown in Fig. 1. The human silhouette is first extracted (section 3) and represented efficiently (section 4). We then compare the silhouette with computer graphics silhouettes (section 6) from a database (section 5). The results of the comparison are calculated for an entire sequence and the gait type and duty-factor of that sequence is

[1] Poser version 6.0.3.140 was used for this work. Currently distributed by Smith Micro Software, Inc.

extracted (section 7). Results are presented in section 8 and section 9 contains a discussion of these results. Sections 10 to 12 present solutions to some of the additional challenges that arise when the gait classification system is applied in an online system with multiple cameras, real-time demands, and maintenance of silhouette quality over long time. Section 13 concludes the chapter.

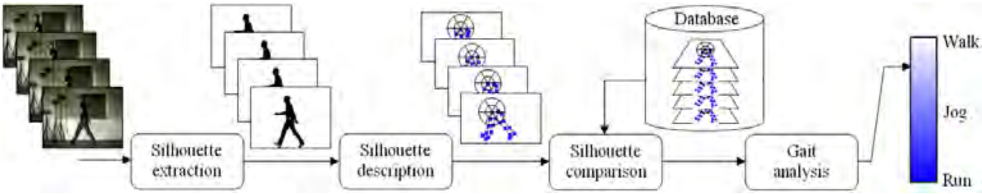

Fig. 1. An overview of the approach. The main contributions of the method presented here are the computer generated silhouette database, the gait analysis resulting in a gait continuum, and the ability to handle unconstrained environments achieved by the methods applied throughout the system. The gait analysis is further detailed in Fig. 7.

2. 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. 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. Furthermore, physiological research has shown that the optimal action changes discontinuously with changing speed. (Alexander, 1989)

From a computer vision point of view the question is now if *one* (recognizable) descriptor exist, which can represent the continuum of gait. 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*" (Alexander, 2002). Fig. 2. illustrates the duty-factor in a walk cycle and a run cycle.

To illustrate the power of this descriptor we have manually estimated the duty-factor in 138 video sequences containing humans walking, jogging, or running, see Fig. 3. These sequences come from 4 different sources and contain many different individuals entering and exiting at different angles. Some not even following a straight line (see example frames in Fig. 10).

Fig. 3. 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 is dependent on the observer when considering movements in the transition phase and there exists no clear definition of what separates jogging from running.

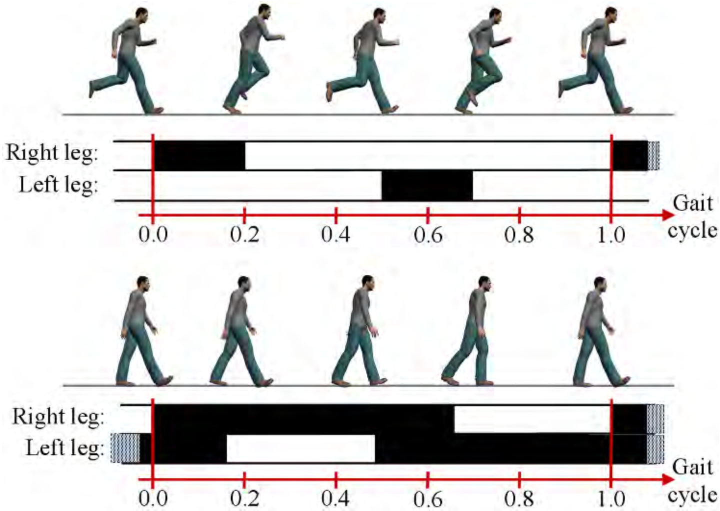

Fig. 2. Illustration of the duty-factor. The duration of a gait cycle where each foot is on the ground is marked with the black areas. The duty-factor for the depicted run cycle (top) is 0.2 and 0.65 for the depicted walk cycle (bottom).

This problem is apparent in the classification of the sequences used in Fig.3. 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]. (Patron & Reid, 2007) reports classification results from 300 video sequences of people walking, jogging, and running. The sequences are classified by several people resulting in classification rates of 100% for walking, 98% for jogging, and only 81% for running, which illustrates the inherent difficulty in distinguishing the two gait types.
With these results in mind we will not attempt to do a traditional classification of walking, jogging, and running which in reality has doubtful ground truth data. Rather, we will use the duty-factor to describe jogging and running as a continuum. This explicitly handles the ambiguity of jogging and running since a video sequence that some people will classify as jogging and other people will classify as running simply map to a point on the continuum described by the duty-factor. This point will not have a precise interpretation in terms of jogging and running but the duty-factor will be precise.
As stated earlier walking and jogging/running are two different ways of moving. However, to get a unified description for all types of gait that are usually performed by people in open spaces we also apply the duty-factor to walking and get a single descriptor for the whole gait continuum.

---

[2] The problem of ambiguous classification will be clear when watching for example video sequences from the KTH data set (Schüldt et al., 2004), e.g. person 4 jogging in scenario 2 versus person 2 running in scenario 2.

Even though jogging and running are considered as one gait type in the context of the duty-factor they still have a visual distinction to some extend. This visual distinction is used with some success in the current approaches which classify gait into walking, jogging, and running. We acknowledge the results obtained by this type of approaches and we also propose a new method to classify gait into walking, jogging, and running. In our approach however, this is only an intermediate step to optimize the estimation of the duty-factor which we believe to be the best way of describing gait.

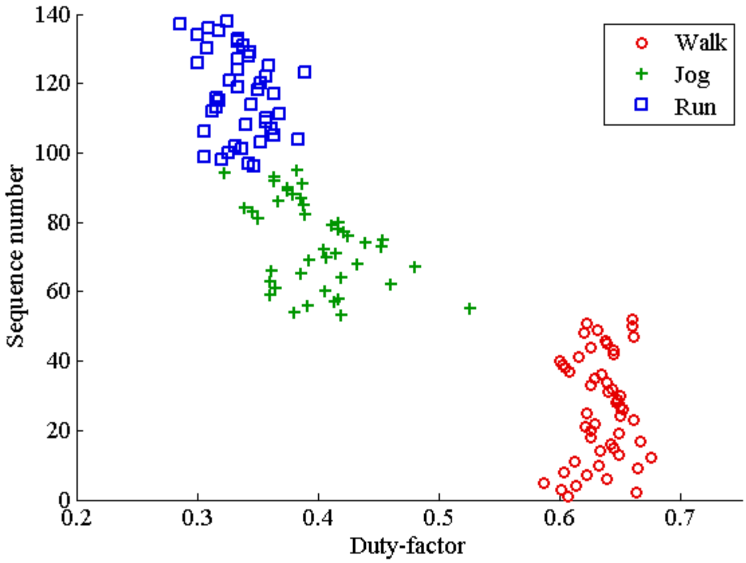

Fig. 3. The manually annotated duty-factor and gait type for 138 different sequences. Note that the sole purpose of the y-axis is to spread out the data.

3. Silhouette extraction

The first step in the gait analysis framework is to extract silhouettes from the incoming video sequences. For this purpose we do foreground segmentation using the Codebook background subtraction method as described in (Fihl et al., 2006) and (Kim et al., 2005). This method has been shown to be robust in handling both foreground camouflage and shadows. This is achieved by separating intensity and chromaticity in the background model. Moreover, the background model is multi modal and multi layered which allows it to model moving backgrounds such as tree branches and objects that become part of the background after staying stationary for a period of time. To maintain good background subtraction quality over time it is essential to update the background model and (Fihl et al., 2006) describes two different update mechanisms to handle rapid and gradual changes respectively. By using this robust background subtraction method we can use a diverse set of input sequences from both indoor and outdoor scenes.

4. Silhouette description

When a person is moving around in an unconstrained scene his or her arms will not necessarily swing in a typical "gait" manner; the person may be making other gestures, such as waving, or he/she might be carrying an object. To circumvent the variability and complexity of such scenarios we choose to classify the gait solely on the silhouette of the legs. Furthermore, (Liu et al., 2004) shows that identification of people on the basis of gait, using the silhouette of legs alone, works just as well as identification based on the silhouette of the entire body.

To extract the silhouette of the legs we find the height of the silhouette of the entire person and use the bottom 50% as the leg silhouette. Without loss of generality this approach avoids errors from the swinging hands below the hips, although it may not be strictly correct from an anatomic point of view. To reduce noise along the contour we apply morphological operations to the silhouette. Some leg configurations cause holes in the silhouette, for example running seen from a non-side view in Fig. 5. (c). Such holes are descriptive for the silhouette and we include the contour of these holes in the silhouette description.

To allow recognition of gait types across different scales we use shape contexts and tangent orientations (Belongie et al., 2002) 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, see Fig. 4. With K bins in the log-polar histogram of the shape context we get an n x (K+1) matrix describing each silhouette. Scale invariance is achieved with shape contexts by normalizing the size of the histograms according to the mean distance between all point pairs on the contour. Specifically, the normalizing constant q used for the radial distances of the histograms is defined as follows:

$$q = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} | p_i - p_j | \quad (1)$$

where n is the number of points p sampled from the contour.

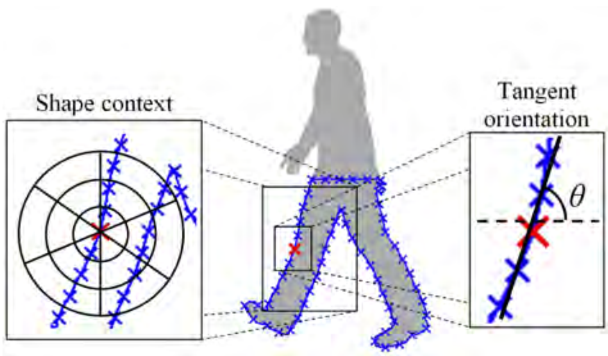

Fig. 4. Illustration of the silhouette description. The crosses illustrate the points sampled from the silhouette. Shape contexts and tangent orientations are used to describe the silhouette.

5. Silhouette database

To represent our training data we create a database of human silhouettes performing one cycle of each of the main gait types: walking, jogging, and running. To make our method invariant to changes in viewpoint we generate database silhouettes from three different camera angles. With 3D-rendering software this is an easy and very rapid process that does not require us to capture 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. The combination of the robust silhouette description and three camera angles enable the method to handle diverse moving directions and oblique viewing angles. Specifically, database silhouettes can be matched with silhouettes of people moving at angles of at least ±45 degrees with respect to the viewing direction. People moving around in open spaces will often change direction while in the camera's field of view (creating non-linear paths of motion), thus we cannot make assumptions about the direction of movement. To handle this variability each new input silhouette is matched to database silhouettes taken from all camera angles. Fig. 10, row 1 shows a sequence with a non-linear motion path where the first frame will match database silhouettes from a viewpoint of -30 degrees and the last frame will match database silhouettes from a viewpoint of 30 degrees. The silhouettes generated are represented as described in section 4. We generate T silhouettes of a gait cycle for each of the three gait types. This is repeated for the three viewpoints, i.e. $T \cdot 3 \cdot 3$ silhouettes in total. Fig. 5 shows examples of the generated silhouettes.

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.

To analyze the content of the database with respect to the ability to describe gait we created an Isomap embedding (Tenenbaum et al., 2000) of the shape context description of the silhouettes. Based on the cyclic nature of gait and the great resemblance between gait types we expect that gait information can be described by some low dimensional manifold. Fig. 6. shows the 2-dimensional embedding of our database with silhouettes described by shape contexts and tangent orientations and using the costs resulting from the Hungarian method (described in section 6) as distances between silhouettes.

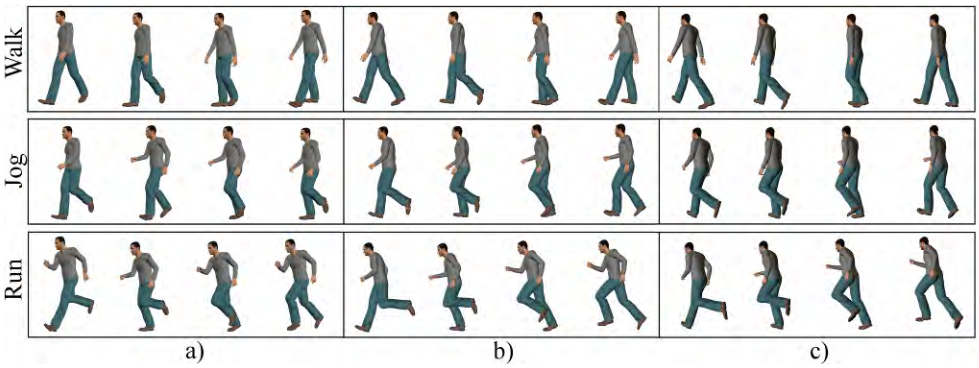

Fig. 5. Example of database silhouettes generated by 3D-rendering software. Silhouettes are generated from three viewpoints. a) and c) illustrate renderings from cameras rotated 30 degrees to each side. b) illustrates renderings from a direct side view.

According to figure 6 we can conclude that the first two intrinsic parameters of the database represent 1) the total distance between both feet and the ground and 2) the horizontal distance between the feet. This reasonable 2-dimensional representation of the database silhouettes shows that our description of the silhouettes and our silhouette comparison metric does capture the underlying manifold of gait silhouettes in a precise manner. Hence, gait type analysis based on our silhouette description and comparison seems promising

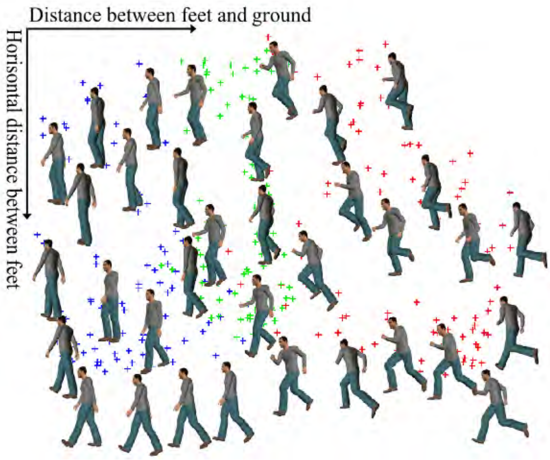

Fig. 6. Illustration of the ISOMAP embedding and a representative subset of the database silhouettes.

6. Silhouette comparison

To find the best match between an input silhouette and database silhouettes we follow the method of (Belongie et al., 2002). We calculate the cost of matching a sampled point on the input silhouette with a sampled 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 shape contexts 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:

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

The normalized shape contexts gives $c_{i,j} \in [0;1]$.
The difference in tangent orientation $\varphi_{i,j}$ between points p_i and p_j is normalized and added to $c_{i,j}$ ($\varphi_{i,j} \in [0;1]$). This gives the final cost $C_{i,j}$ of matching the two points:

$$C_{i,j} = a \cdot c_{i,j} + b \cdot \varphi_{i,j} \tag{3}$$

where a and b are weights. Experiments have shown that $\varphi_{i,j}$ effectively discriminates points that are quite dissimilar whereas $c_{i,j}$ expresses more detailed differences which should have a high impact on the final cost only when tangent orientations are alike. According to this observation we weight the difference in tangent orientation $\varphi_{i,j}$ higher than shape context distances $c_{i,j}$. Preliminary experiments show that the method is not too sensitive to the choice of these weights but a ratio of 1 to 3 yields good results, i.e. a=1 and b=3.
The costs of matching all point pairs between the two silhouettes are calculated. The Hungarian method (Papadimitriou & Steiglitz, 1998) is used to solve the square assignment problem of identifying which one-to-one mapping between the two point sets that minimizes the total cost. All point pairs are included in the cost minimization, i.e. the ordering of the points is not considered. This is because points sampled from a silhouette with holes will have a very different ordering compared to points sampled from a silhouette without holes but with similar leg configuration, see row three of Fig. 5. (c) (second and third image) for an example.
By finding the best one-to-one mapping between the input silhouette and each of the database silhouettes we can now identify the best match in the whole database as the database silhouette involving the lowest total cost.

7. Gait analysis

The gait analysis consists of two steps. First we do classification into one of the three gait types, i.e. walking, jogging, or running. Next we calculate the duty-factor D based on the silhouettes from the classified gait type. This is done to maximize the likelihood of a correct duty-factor estimation. Fig. 7. illustrates the steps involved in the gait type analysis. Note that the silhouette extraction, silhouette description, and silhouette comparison all process a single input frame at a time whereas the gait analysis is based on a sequence of input frames.
To get a robust classification of the gait type in the first step we combine three different types of information. We calculate an *action error* E for each action and two associated weights: *action likelihood* α and *temporal consistency* β. The following subsections describe the gait analysis in detail starting with the action error and the two associated weights followed by the duty-factor calculation.

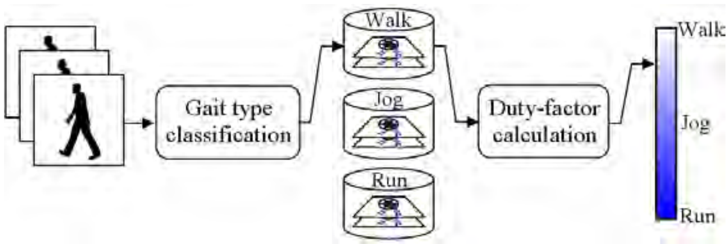

Fig. 7. An overview of the gait analysis. The figure shows the details of the block "Gait analysis" in Fig. 1. The output of the silhouette comparison is a set of database silhouettes matched to the input sequence. In the gait type classification these database silhouettes are classified as a gait type which defines a part of the database to be used for the duty-factor calculation.

7.1 Action Error

The output of the silhouette comparison is a set of distances between the input silhouette and each of the database silhouettes. These distances express the difference or error between two silhouettes. Fig. 8. illustrates the output of the silhouette comparison. The database silhouettes are divided into three groups corresponding to walking, jogging, and running, respectively. We accumulate the errors of the best matches within each group of database silhouettes. These accumulated errors constitute the *action error E* and corresponds to the difference between the action being performed in the input video and each of the three actions in the database, see Fig. 9.

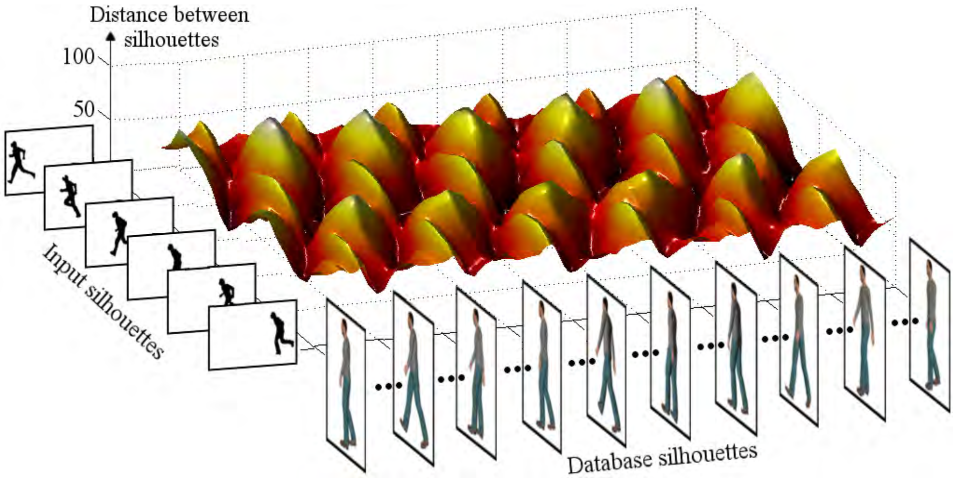

Fig. 8. Illustration of the silhouette comparison output. The distances between each input silhouette and the database silhouettes of each gait type are found (shown for walking only). 90 database silhouettes are used per gait type, i.e. T=30.

7.2 Action Likelihood

When silhouettes of people are extracted in difficult scenarios and at low resolutions the silhouettes can be noisy. This may result in large errors between the input silhouette and a database silhouette, even though the actual pose of the person is very similar to that of the database silhouette. At the same time, small errors may be found between noisy input silhouettes and database silhouettes with quite different body configurations (somewhat random matches). To minimize the effect of the latter inaccuracies we weight the action error by the likelihood of that action. The action likelihood of action a is given as the percentage of input silhouettes that match action a better than the other actions.

Since we use the minimum action error the actual weight applied is one minus the action likelihood:

$$\alpha_a = 1 - \frac{n_a}{N} \tag{4}$$

where n_a is the number of input silhouettes in a sequence with the best overall match to a silhouette from action a, and N is the total number of input silhouettes in that video sequence.

This weight will penalize actions that have only a few overall best matches, but with small errors, and will benefit actions that have many overall best matches, e.g. the running action in Fig. 9.

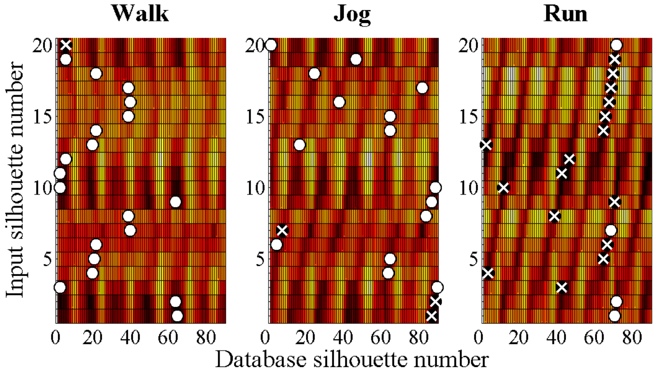

Fig. 9. The output of the silhouette comparison of Fig. 8. is shown in 2D for all gait types (dark colors illustrate small errors and bright colors illustrate large errors). For each input silhouette the best match among silhouettes of the same action is marked with a white dot and the best overall match is marked with a white cross. The shown example should be interpreted as follows: the silhouette in the first input frame is closest to walking silhouette number 64, to jogging silhouette number 86, and to running silhouette number 70. These distances are used when calculating the action error. When all database silhouettes are considered together, the first input silhouette is closest to jogging silhouette number 86. This is used in the calculation of the two weights.

7.3 Temporal Consistency

When considering only the overall best matches we can find sub-sequences of the input video where all the best matches are of the same action *and* in the right order with respect to a gait cycle. This is illustrated in Fig. 9. where the running action has great temporal consistency (silhouette numbers 14-19). The database silhouettes are ordered in accordance with a gait cycle. Hence, the straight line between the overall best matches for input silhouettes 14 to 19 shows that each new input silhouette matches the database silhouette that corresponds to the next body configuration of the running gait cycle.

Sub-sequences with correct temporal ordering of the overall best matches increase our confidence that the action identified is the true action. The temporal consistency describes the length of these sub-sequences. Again, since we use the minimum action error we apply one minus the temporal consistency as the weight β_a:

$$\beta_a = 1 - \frac{m_a}{N} \tag{5}$$

where m_a is the number of input silhouettes in a sequence in which the best overall match has correct temporal ordering within action a, and N is the total number of input silhouettes in that video sequence.

Our definition of temporal consistency is rather strict when you consider the great variation in input silhouettes caused by the unconstrained nature of the input. A strict definition of temporal consistency allows us to weight it more highly than action likelihood, i.e. we apply a scaling factor w to β to increase the importance of temporal consistency in relation to action likelihood:

$$\beta_a = 1 - w\frac{m_a}{N} \tag{6}$$

7.4 Gait-type classification

The final classifier for the gait type utilizes both the action likelihood and the temporal consistency as weights on the action error. This yields:

$$Action = \arg\min_a (E_a \cdot \alpha_a \cdot \beta_a) \tag{7}$$

where E_a is the action error, α_a is the action likelihood, β_a is the weighted temporal consistency.

7.5 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. Following this definition 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 gait 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 (see Fig. 2. for an illustration). 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 (Collins et al., 2002). 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 in silhouette width) 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 a sequence. We do not attempt to estimate ground contact directly in the input videos which would require assumptions about the ground plane and camera calibrations. For a system intended to work in unconstrained open scenes such requirements will be a limitation to the system. In stead of estimating the feet's ground contact in the input sequence we infer the ground contact from the database silhouettes that are matched to that sequence. Since each database silhouette is annotated with the number of feet supported on the ground this is a simple lookup in the database. The ground support estimation is based solely on silhouettes from the gait type found in the gait-type classification which maximize the likelihood of a correct estimate of the ground support.

The total ground support G of both feet for a video sequence is the sum of ground support of all the matched database silhouettes within the specific gait type.

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 $G/(2n_s)$, where n_s is the number of strides in the sequence. The duty-factor D is now given as $D=g/t_s$. In summary we have

$$Duty-factor\ D = \frac{G}{2 \cdot n_s \cdot t_s} \tag{8}$$

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.

The manual labeled data of Fig. 3. allows us to further enhance the precision of the duty-factor description. It can be seen from Fig. 3. that the duty-factor for running is in the interval [0.28;0.39] and jogging is in the interval [0.34;0.53]. This can not be guarantied to be true for all possible executions of running and jogging but the great diversity in the manually labeled data allows us to use these intervals in the duty-factor estimation. Since walking clearly separates from jogging and running and since no lower limit is needed for running we infer the following constraints on the duty factor of running and jogging:

$$\begin{aligned} D_{running} &\in [0;0.39] \\ D_{jogging} &\in [0.34;0.53] \end{aligned} \tag{9}$$

We apply these bounds as a post-processing step. If the duty-factor of a sequence lies outside one of the appropriate bounds then the duty-factor will be assigned the value of the exceeded bound.

8. Results

To emphasize the contributions of our two-step gait analysis we present results on both steps individually and on the gait continuum achieved by combining the two steps.

A number of recent papers have reported good results on the classification of gait types (often in the context of human action classification). To compare our method to these results and to show that the gait type classification is a solid base for the duty-factor calculation we have tested this first step of the gait analysis on its own. After this comparison we test the duty-factor description with respect to the ground truth data shown in Fig. 3., both on its own and in combination with the gait type classification.

The tests are conducted on a large and diverse data set. We have compiled 138 video sequences from 4 different data sets. The data sets cover indoor and outdoor video, different moving directions with respect to the camera (up to ±45 degrees from the viewing direction), non-linear paths, different camera elevations and tilt angles, different video resolutions, and varying silhouette heights (from 41 pixels to 454 pixels). Fig. 10. shows example frames from the input videos. Ground truth gait types were adopted from the data sets when available and manually assigned by us otherwise.

For the silhouette description the number of sampled points n was 100 and the number of bins in the shape contexts K was 60. 30 silhouettes were used for each gait cycle, i.e., T=30. The temporal consistency was weighted by a factor of four determined through quantitative experiments, i.e. w=4.

8.1 Gait-type classification

When testing only the first step of the gait analysis we achieve an overall recognition rate of 87.1%. Table 1 shows the classification results in a confusion matrix.

	Walk	Jog	Run
Walk	96.2	3.8	0.0
Jog	0.0	65.9	34.1
Run	0.0	2.6	97.4

Table 1. Confusion matrix for the gait type classification results.

The matching percentages in Table 1 cannot directly be compared to the results of others since we have included samples from different data sets to obtain more diversity. However, 87 of the sequences originate from the KTH data set (Schüldt et al., 2004) and a loose comparison is possible on this subset of our test sequences. In Table 2 we list the matching results of different methods working on the KTH data set.

Methods	Classification results in %			
	Total	Walk	Jog	Run
Kim & Cipolla (2009)*	92.3	99	90	88
Our method	92.0	100.0	80.6	96.3
Li et al. (2008)*	89.0	88	89	90
Laptev et al. (2008)*	89.3	99	89	80
Patron & Reid (2007)	84.3	98	79	76
Schüldt et al. (2004)	75.0	83.3	60.4	54.9

Table 2. Best reported classification results on the KTH data set. The matching results of our method are based on the 87 KTH sequences included in our test set. * indicate that the method work on all actions of the KTH data set.

The KTH data set remains one of the largest data sets of human actions in terms of number of test subjects, repetitions, and scenarios and many papers have been published with results on this data set, especially within the last two years. A number of different test setups have been used which makes a direct comparison impossible and we therefore merely list a few of the best results to show the general level of recognition rates. We acknowledge that the KTH data set contains three additional actions (boxing, hand waving, and hand clapping) and that some of the listed results include these. However, for the results reported in the literature the gait actions are in general not confused with the three hand actions. The results can therefore be taken as indicators of the ability of the methods to classify gait actions exclusively.

Another part of our test set is taken from the Weizmann data set (Blank et al., 2005). They classify nine different human actions including walking and running but not jogging. They achieve a near perfect recognition rate for running and walking and others also report 100% correct recognitions on this data set, e.g. (Patron et al., 2008). To compare our results to this we remove the jogging silhouettes from the database and leave out the jogging sequences

from the test set. In this walking/running classification we achieve an overall recognition rate of 98.9% which is slightly lower. Note however that the data sets we are testing on include sequences with varying moving directions where the results in (Blank et al., 2005) and (Patron et al., 2008) are based on side view sequences.
In summary, the recognition results of our gait-type classification provides a very good basis for the estimation of the duty-factor.

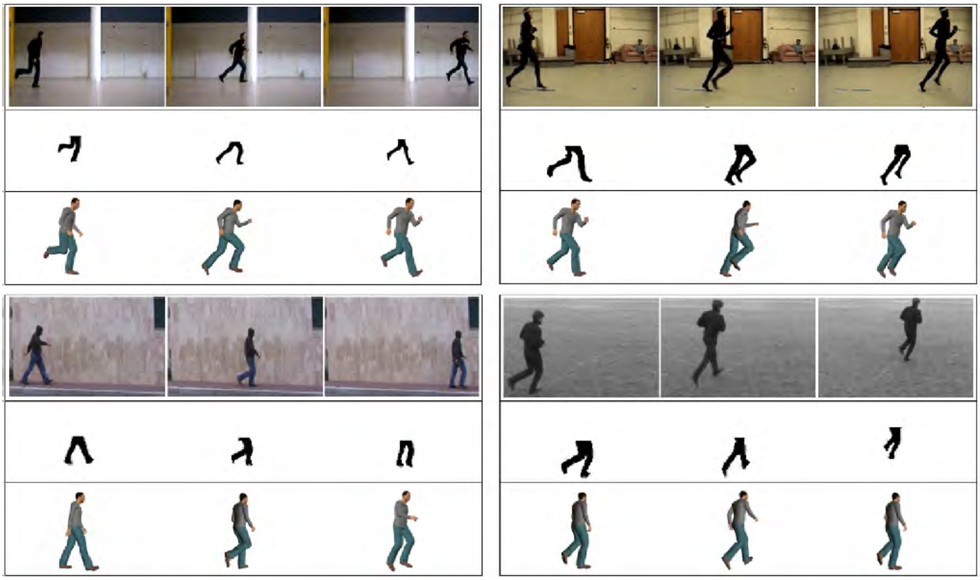

Fig. 10. Samples from the 4 different data sets used in the test together with the extracted silhouettes of the legs used in the database comparison, and the best matching silhouette from the database. Top left: data from our own data set. Bottom left: data from the Weizmann data set (Blank et al., 2005). Top right: data from the CMU data set obtained from mocap.cs.cmu.edu. The CMU database was created with funding from NSF EIA-0196217. Bottom right: data from the KTH data set (Schüldt et al., 2004).

8.2 Duty-factor

To test our duty-factor description we estimate it automatically in the test sequences. To show the effect of our combined gait analysis we first present results for the duty-factor estimated without the preceding gait-type classification to allow for a direct comparison.
Fig. 11. shows the resulting duty-factors when the gait type classification is not used to limit the database silhouettes to just one gait type. Fig. 12. shows the estimated duty-factors with our two-step gait analysis scheme. The estimate of the duty-factor is significantly improved by utilizing the classification results of the gait type classification. The mean error for the estimate is 0.050 with a standard deviation of 0.045.

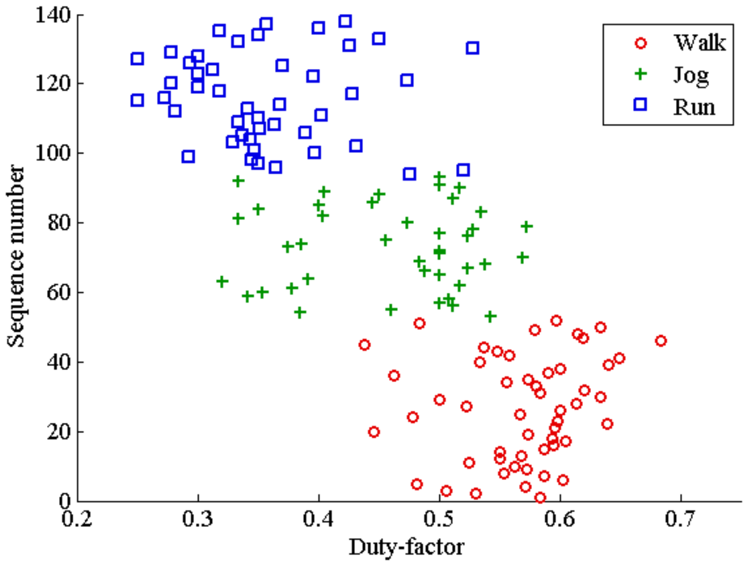

Fig. 11. The automatically estimated duty-factor from the 138 test sequences without the use of the gait type classification. The y-axis solely spreads out the data.

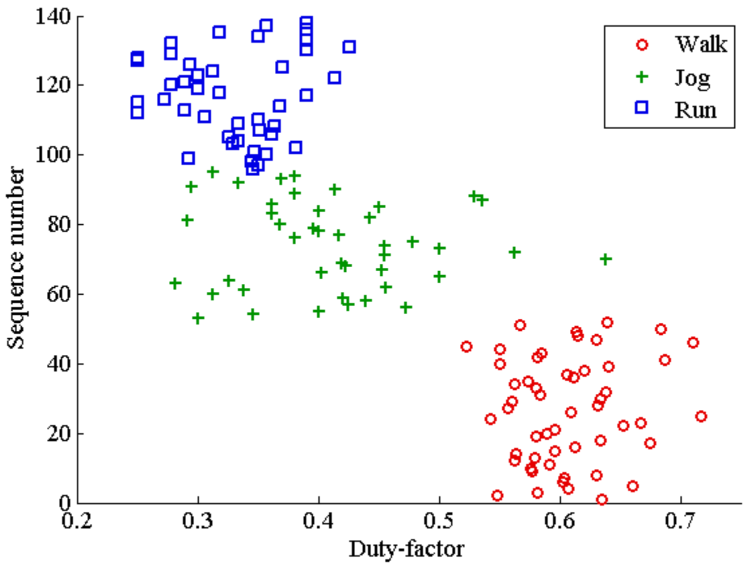

Fig. 12. The automatically estimated duty-factor from the 138 test sequences when the gait type classification has been used to limit the database to just one gait type. The y-axis solely spreads out the data.

9. Discussion

When comparing the results of the estimated duty-factor (Fig. 12.) with the ground truth data (Fig. 3.) it is clear that the overall tendency of the duty-factor is reproduced with the automatic estimation. The estimated duty-factor has greater variability mainly due to small inaccuracies in the silhouette matching. A precise estimate of the duty-factor requires a precise detection of when the foot actually touches the ground. However, this detection is difficult because silhouettes of the human model are quite similar just before and after the foot touches the ground. Inaccuracies in the segmentation of the silhouettes in the input video can make for additional ambiguity in the matching.

The difficulty in estimating the precise moment of ground contact leads to considerations on alternative measures of a gait continuum, e.g. the Froude number (Alexander, 1989) that is based on walking speed and the length of the legs. However, such measures requires information about camera calibration and the ground plane which is not always accessible with video from unconstrained environments. The processing steps involved in our system and the silhouette database all contributes to the overall goal of creating a system that is invariant to usual challenges in video from unconstrained scenes and a system that can be applied in diverse setups without requiring additional calibrations.

The misclassifications of the three-class classifier also affect the accuracy of the estimated duty-factor. The duty-factor of the four jogging sequences misclassified as walking disrupt the perfect separation of walking and jogging/running expected from the manually annotated data. All correctly classified sequences however maintain this perfect separation.

To test wether the presented gait classification framework provides the kind of invariance that is required for unconstrained scenes we have analyzed the classification errors in Table 1. This analysis shows no significant correlation between the classification errors and the camera viewpoint (pan and tilt), the size and quality of the silhouettes extracted, the image resolution, the linearity of the path, and the amount of scale change. Furthermore, we also evaluated the effect of the number of frames (number of gait cycles) in the sequences and found that our method classifies gait types correctly even when there are only a few cycles in the sequence. This analysis is detailed in Table 3 which shows the result of looking at a subset of the test sequences containing a specific video characteristic.

Video characteristic	Percentage of	Percentage of
Non-side view	43	41
Small silhouettes (1)	58	59
Low resolution images (2)	63	65
Non linear path	3	0
Significant scale change (3)	41	41
Less than 2 strides	43	41

Table 3. The table shows how different video characteristics effect the classification errors, e.g. 43% of the sequences have a non-side view and these sequences account for 41% of the errors. The results are based on 138 test sequences out of which 17 sequences were erroneously classified. Notes: (1): Mean silhouette height of less than 90 pixels. (2): Image resolution of 160x120 or smaller. (3): Scale change larger than 20% of the mean silhouette height during the sequence.

A number of the sequences in Table 3 have more than one of the listed characteristics (e.g. small silhouettes in low resolution images) so the error percentages are somewhat correlated. It should also be noted that the gait type classification results in only 17 errors which gives a relatively small number of sequences for this analysis. However, the number of errors in each subset corresponds directly to the number of sequences in that subset which is a strong indication that our method is indeed invariant to the main factors relevant for gait classification.
The majority of the errors in Table 1 occur simply because the gait type of jogging resembles that of running which supports the need for a gait continuum.

10. Multi Camera Setup

The system has been designed to be invariant towards the major challenges in a realistic real-world setup. Regarding invariance to view point, we have achieved this for gait classification of people moving at an angle of up to ±45 degrees with respect to the view direction. The single-view system can however easily be extended to a multi-view system with synchronized cameras which can allow for gait classification of people moving at completely arbitrary directions. A multi-view system must analyze the gait based on each stride rather than a complete video sequence since people may change both moving direction and type of gait during a sequence.
The direction of movement can be determined in each view by tracking the people and analyzing the tracking data. Tracking is done as described in (Fihl et al., 2006). If the direction of movement is outside the ±45 degree interval then that view can be excluded. The duration of a stride can be determined as described in section 2 from the view where the moving direction is closest to a direct side-view. The gait classification results of the remaining views can be combined into a multi-view classification system by extending equations 7 and 8 into the following and doing the calculations based on the last stride in stead of the whole sequence.

$$Action = \arg\min_a \left(\sum_V E_a \cdot \alpha_a \cdot \beta_a\right) \tag{10}$$

$$D = \frac{1}{n_V} \cdot \sum_V D_V \tag{11}$$

where V is the collection of views with acceptable moving directions, E_a is the action error, α_a is the action likelihood, β_a is the temporal consistency, D is the duty-factor, n_V is the number of views, and D_v is the duty-factor from view v.
Fig. 13. illustrates a two-camera setup where the gait classification is based on either one of the cameras or a combination of both cameras.

11. Real Time Performance

The full potential of the gait analysis framework can only be achieved with real-time performance. Non-real-time processing can be applied for annotation of video data but for e.g. human-robot interaction, automated video surveillance, and intelligent vehicles real-time performance is necessary.

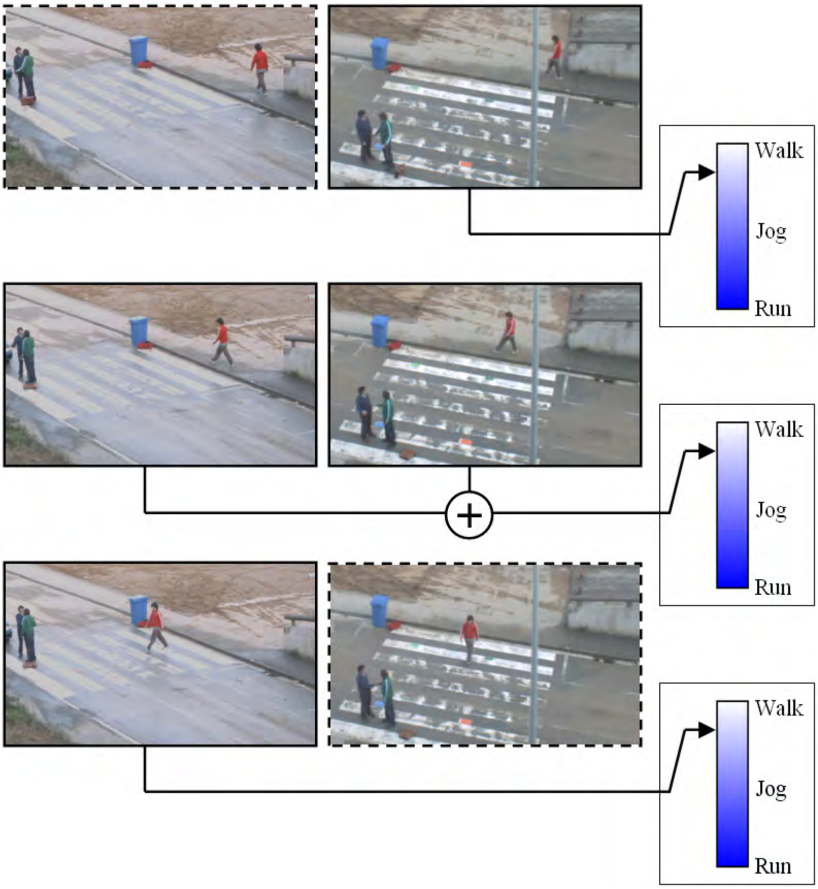

Fig. 13. A two-camera setup. The figure shows three sets of synchronized frames from two cameras. The multi-camera gait classification enables the system to do classification based on either one view (top and bottom frames) or a combination of both views (middle frame).

Real-time performance can be achieved with an optimized implementation and minor changes in the method. The extraction of the contour of the silhouettes is limited to the outermost contour. Disregarding the inner contours (see Fig. 14.) gave a decrease in processing time but also a small decrease in classification results due to the loss of details in some silhouettes.

Fig. 14. Left: the input silhouette. Middle: the outermost contour extracted in the real time system. Right: the contour extracted in the original system.

The most time consuming task of the gait classification is the matching of the input silhouette to the database silhouettes both represented in terms of Shape Contexts. By decreasing the number of points sampled around the contour from 100 points to 20 points and by decreasing the number of bins in the Shape Contexts from 60 to 40 the processing time is significantly improved while still maintaining most of the descriptive power of the method.
With these changes the gait classification system is running at 12-15 frames per second on a standard desktop computer with a 2GHz dual core processor and 2GB of RAM. This however also means a decrease in the classification power of the system. When looking at the gait type classification a recognition rate of 83.3% is achieved with the real-time setup compared to 87.1% with the original setup. The precision of the duty-factor estimation also decreases slightly. This decrease in recognition rate is considered to be acceptable compared to the increased applicability of a real-time system.

12. Online parameter tuning of segmentation

The silhouette extraction based on the Codebook background subtraction is a critical component in the system. Noise in the extracted silhouettes has a direct impact on the classification results. Illumination and weather conditions can change rapidly in unconstrained open spaces so to ensure the performance of the background subtraction in a system receiving live input directly from a camera we have developed a method for online tuning of the segmentation.
The performance of the Codebook background subtraction method is essentially controlled by three parameters; two controlling the allowed variation in illumination and one controlling the allowed variation in chromaticity. The method is designed to handle shadows so with a reasonable parameter setup the Codebook method will accept relatively large variations in illumination to account for shadows that are cast on the background. However, changes in lighting conditions in outdoor scenes also have an effect on the chromaticity level which is not directly modeled in the method. Because of this, the parameter that controls the allowed variation in chromaticity σ is the most important parameter to adjust online (i.e. fixed parameters for the illumination variation will handle changing lighting conditions well, whereas a fixed parameter for the chromaticity variation will not).
To find the optimal setting for σ at runtime we define a quality measure to evaluate a specific value of σ and by testing a small set of relevant values for each input frame we adjust σ by optimizing this quality measure.
The quality measure is based on the difference between the edges of the segmentation and the edges of the input image. An edge background model is acquired simultaneously with the Codebook background model which allows the system to classify detected edges in a new input frame as either foreground or background edges. The map of foreground edges has too much noise to be used for segmentation itself but works well when used to compare the quality of different foreground segmentations of the same frame. The quality score Q is defined as follows:

$$Q = \frac{\sum E_{fg} \cdot E_{seg}}{\sum E_{seg}} \tag{12}$$

where E_{fg} are the foreground edges and E_{seg} are the edges of the foreground mask from the background subtraction. So the quality score describes the fraction of edges from the foreground mask that corresponds to foreground edges from the input image.
The background subtraction is repeated a number of times on each input frame with varying values of σ and the quality score is calculated after each repetition. The segmentation that results in the highest quality score is used as the final segmentation. Fig. 15. and Fig. 16. show example images of this process.

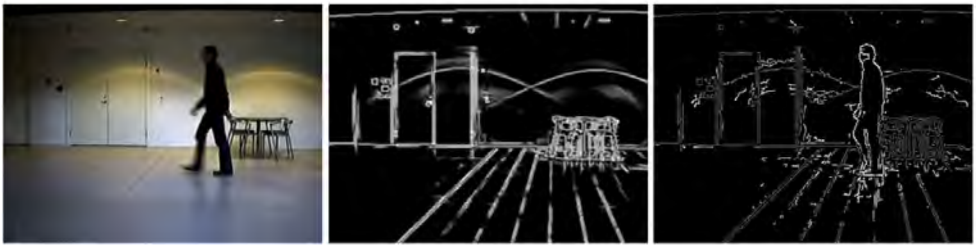

Fig. 15. Left: the input image. Middle: the background edge model. Right: the foreground edges.

Fig. 16. Three segmentation results with varying values of σ. Left: σ -value too low. Middle: optimal σ -value. Right: σ -value too high.

The repetitive segmentation of each frame slows the silhouette extraction of the gait classification system down but by only testing a few values of σ for each frame real time performance can still be achieved. The first frames of a new input sequence will be tested with up to 30 values of σ covering a large interval (typically [1:30]) to initialize the segmentation whereas later frames will be tested with only four to six values of σ in the range ±2 of the σ -value from the previous frame.

13. Conclusion

The gait type of people that move around in open spaces is an important property to recognize in a number of applications, e.g. automated video surveillance and human-robot interaction. The classical description of gait as three distinct types is not always adequate and this chapter has presented a method for describing gait types with a gait continuum which effectively extends and unites the notion of running, jogging, and walking as the three gait types. The method is *not* based on statistical analysis of training data but rather on a general gait motion model synthesized using a computer graphics human model. This makes training (from different views) very easy and separates the training and test data completely. The method is designed to handle challenges that arise in an unconstrained scene and the method has been evaluated on different data sets containing all the important factors which such a method should be able to handle.The method performs well (both in its own right and in comparison to related methods) and it is concluded that the method can be characterized as an *invariant* method for gait description.

The method is further developed to allow video input from multiple cameras. The method can achieve real-time performance and a method for online adjustment of the background subtraction method ensures the quality of the silhouette extraction for scenes with rapid changing illumination conditions.

The quality of the foreground segmentation is important for the precision of the gait classification and duty-factor estimation. The segmentation quality could be improved in the future by extending the color based segmentation of the Codebook method with edge information directly in the segmentation process and furthermore including region based information. This would especially be an advantage in scenes with poor illumination or with video from low quality cameras.

The general motion model used to generate training data effectively represents the basic characteristics of the three gait types, i.e. the characteristics that are independent of person-specific variations. Gait may very well be the type of actions that are most easily described by a single prototypical execution but an interesting area for future work could be the extension of this approach to other actions like waving, boxing, and kicking.

The link between the duty-factor and the biomechanical properties of gait could also be an interesting area for future work. By applying the system in a more constrained setup it would possible to get camera calibrations and ground plane information that could increase the precision of the duty-factor estimation to a level were it may be used to analyze the performance of running athletes.

14. Acknowledgment

This work was supported by the EU project HERMES (FP6 IST-027110) and the BigBrother project (Danish Agency for Science, Technology, and Innovation, CVMT, 2007-2010).

15. References

Alexander, R. (1989). Optimization and Gaits in the Locomotion of Vertebrates, *Physiological Reviews* **69**(4): 1199 – 1227.

Alexander, R. (2002). Energetics and Optimization of Human Walking and Running: The 2000 Raymond Pearl Memorial Lecture, *American Journal of Human Biology* **14**(5): 641 – 648.

Belongie, S., Malik, J. & Puzicha, J. (2002). Shape Matching and Object Recognition Using Shape Contexts, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(4): 509–522.

Blakemore, S.-J. & Decety, J. (2001). From the Perception of Action to the Understanding of Intention, *Nature Reviews Neuroscience* **2**(8): 561–567.

Blank, M., Gorelick, L., Shechtman, E., Irani, M. & Basri, R. (2005). Actions as Space-Time Shapes, *ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision*, IEEE Computer Society, Washington, DC, USA, pp. 1395–1402.

Collins, R., Gross, R. & Shi, J. (2002). Silhouette-Based Human Identification from Body Shape and Gait, *FGR '02: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, IEEE Computer Society, Washington, DC, USA, pp. 351–356.

Cutler, R. & Davis, L. S. (2000). Robust Real-Time Periodic Motion Detection, Analysis, and Applications, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(8): 781–796.

Dollár, P., Rabaud, V., Cottrell, G. & Belongie, S. (2005). Behavior Recognition via Sparse Spatio-Temporal Features, *2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*.

Fihl, P., Corlin, R., Park, S., Moeslund, T. & Trivedi,M. (2006). Tracking of Individuals in Very Long Video Sequences, *International Symposium on Visual Computing*, Lake Tahoe, Nevada, USA.

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

Kim, T.-K. & Cipolla, R. (2009). Canonical Correlation Analysis of Video Volume Tensors for Action Categorization and Detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **31**(8): 1415–1428.

Laptev, I., Marszalek, M., Schmid, C. & Rozenfeld, B. (2008). Learning Realistic Human Actions from Movies, *CVPR 2008: IEEE Conference on Computer Vision and Pattern Recognition*, Alaska, USA.

Li, Z., Fu, Y., Huang, T. & Yan, S. (2008). Real-time Human Action Recognition by Luminance Field Trajectory Analysis, *MM '08: Proceeding of the 16th ACM international conference on Multimedia*, ACM, New York, NY, USA, pp. 671–676.

Liu, Z., Malave, L., Osuntugun, A., Sudhakar, P. & Sarkar, S. (2004). Towards Understanding the Limits of Gait Recognition, *International Symposium on Defense and Security*, Orlando, Florida, USA.

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

Masoud, O. & Papanikolopoulos, N. (2003). A Method for Human Action Recognition, *Image and Vision Computing* **21**(8): 729 – 743.

Meisner, E. M., Ábanovic, S., Isler, V., Caporeal, L. C. R. & Trinkle, J. (2009). ShadowPlay: a Generative Model for Nonverbal Human-robot Interaction, *HRI '09: Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*.

Montepare, J. M., Goldstein, S. B. & Clausen, A. (1987). The Identification of Emotions from Gait Information, *Journal of Nonverbal Behavior* **11**(1): 33–42.

Papadimitriou, C. & Steiglitz, K. (1998). *Combinatorial Optimization: Algorithms and Complexity*, Courier Dover Publications, Mineola, NY, USA.

Patron, A. & Reid, I. (2007). A Probabilistic Framework for Recognizing Similar Actions using Spatio-Temporal Features, *18th British Machine Vision Conference*.

Patron, A., Sommerlade, E. & Reid, I. (2008). Action recognition using shared motion parts, *Proceedings of the Eighth International Workshop on Visual Surveillance 2008*.

Ran, Y., Weiss, I., Zheng, Q. & Davis, L. S. (2007). Pedestrian Detection via Periodic Motion Analysis, *International Journal of Computer Vision* **71**(2): 143 – 160.

Robertson, N. & Reid, I. (2005). Behaviour Understanding in Video: A Combined Method, *10th IEEE International Conference on Computer Vision*, pp. 808–814.

Schüldt, C., Laptev, I. & Caputo, B. (2004). Recognizing Human Actions: a Local SVM Approach, *ICPR '04: Proceedings of the 17th International Conference on Pattern Recognition*, IEEE Computer Society, pp. 32–36.

Svenstrup, M., Tranberg, S., Andersen, H. & Bak, T. (2009). Pose Estimation and Adaptive Robot Behaviour for Human-Robot Interaction, *International Conference on Robotics and Automation*, Kobe, Japan.

Tenenbaum, J., de Silva, V. & Langford, J. (2000). A Global Geometric Framework for Nonlinear Dimensionality Reduction, *Science* **290**(5500): 2319 – 2323.

Veeraraghavan, A., Roy-Chowdhury, A. & Chellappa, R. (2005). Matching Shape Sequences in Video with Applications in HumanMovement Analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(12): 1896 – 1909.

Viola, P., Jones, M. J. & Snow, D. (2005). Detecting Pedestrians Using Patterns of Motion and Appearance, *International Journal of Computer Vision* **63**(2): 153 – 161.

Waldherr, S., Romero, R. & Thrun, S. (2000). A Gesture Based Interface for Human-Robot Interaction, *Autonomous Robots* **9**(2): 151–173.

Wang, L., Tan, T. N., Ning, H. Z. & Hu, W. M. (2004). Fusion of Static and Dynamic Body Biometrics for Gait Recognition, *IEEE Transactions on Circuits and Systems for Video Technology* **14**(2): 149–158.

Whittle, M.W. (2001). *Gait Analysis, an Introduction*, Butterworth-Heinemann Ltd.

Yam, C., Nixon, M. & Carter, J. (2002). On the Relationship of Human Walking and Running: Automatic Person Identification by Gait, *International Conference on Pattern Recognition*.

Yang, H.-D., Park, A.-Y.& Lee, S.-W. (2006). Human-Robot Interaction by Whole Body Gesture Spotting and Recognition, *International Conference on Pattern Recognition*.