

# Detecting Abnormal Gait

Christian Bauckhage and John K. Tsotsos  
Centre for Vision Research  
York University  
Toronto, ON

Frank E. Bunn  
StressCam Operations & Systems Ltd.  
Toronto, ON

## Abstract

*Analyzing human gait has become popular in computer vision. So far, however, contributions to this topic almost exclusively considered the problem of person identification. In this paper, we will view gait analysis from a different angle and shall examine its use as a means to deduce the physical condition of people. Understanding the detection of unusual movement patterns as a two class problem leads to the idea of using support vector machines for classification. We will thus present a homeomorphism between 2D lattices and binary shapes that provides a robust vector space embedding of body silhouettes. Experimental results will underline that feature vectors obtained from this scheme are well suited to detect abnormal gait. Wavering, faltering, and falling can be detected reliably across individuals without tracking or recognizing limbs or body parts.*

## 1. Introduction

In times of increased interest in biometrics for authentication and access control, it comes with little surprise to see that gait analysis is an increasingly active area of computer vision research. Indeed, dealing with the problem of person identification, human gait is appealing for two reasons: in contrast to fingerprint or retina scans, gait can be analyzed from a distance but is yet as significant as the former. The fact that even though no two body movements are ever the same gait is a unique personal characteristic was first noted by Bernstein in the 1920s. Medical studies from the 1960s systematically corroborated this observation [13] and psychological experiments conducted in the 1970s revealed that humans effortlessly recognize people by the way they walk [5].

The spectrum of methods that have been proposed in vision based gait analysis is vast (see [7, 12, 16] for detailed surveys). However, it is noticeable that although early contributions to the problem did not extract human silhouettes from image sequences [14] the majority of recent work relies on shape analysis. Little and Boyd [11] derive the shape of motion from optical flow and compute characteristic fea-

tures therefrom. Bobick and Johnson [2] subdivide the binary silhouette of walking people into body parts and measure different relations among limbs. While Collins et al. [4] propose template matching of body silhouettes as a baseline algorithm in gait recognition, Sarkar et al. [16] favor similarity detection based on temporal shape statistics. Tolliver and Collins [17] store normalized silhouettes as vertices of similarity graphs and recover significant parts of the walking cycle from computing the eigenvectors of the Laplacian matrix of these graphs. BenAbdelkader et al. [1] compute self-similarity plots from silhouettes and apply subspace methods to recognize individuals therefrom. Other authors base their work on a statistical theory of shape developed by Kendall [10]: Boyd [3], Veeraraghavan et al. [18] and Wang et al. [19] apply Procrustean distances to compare and classify human silhouettes.

Apart from the popularity of shape based approaches, the papers considered in this rough survey reveal that gait analysis is almost exclusively being applied in identification tasks. However, gait can disclose more than identity. Human subjects easily recognize different types of motion (e.g. walking, dancing, wavering, ...) when asked to interpret moving light displays (obtained from filming subjects wearing light bulbs on their joints) or sequences of body silhouettes [9]. Consider, for example, Fig. 1. It shows two series of shapes that were extracted from videos of walking people. While there is nothing unusual in the upper series which shows a subject moving away from the camera, the subject in the lower series obviously has difficulties walking. The movements captured in this sequence suggest the person is in a dubious condition.

As there are numerous applications for the detection of *abnormal* gait, it seems worthwhile to explore techniques that can accomplish this. The work presented in the following is a first step in this direction. First, we will describe a robust shape encoding scheme that allows characterizing types of movements across individuals. Then, we will present experimental results obtained from support vector classification of gait. A summary and an outlook will close this contribution.



(a) Person walking away from the camera



(b) Person wavering towards the camera

**Figure 1. Silhouettes extracted from videos of walking people. Apparently, the information contained in sequences of shapes allows to distinguish between *normal* and *abnormal* gait.**

## 2. Robust Shape Encoding

Our approach to detecting abnormal gait from video was guided by the following considerations: Figure 1 indicates that shape dynamics provide a strong cue to distinguish normal from abnormal gait. Therefore, it appears natural to follow the general trend identified in the introduction and to base gait classification on shape analysis. In its most rudimentary form, the classification task can be treated as a two class problem and as such might be solved using support vector machines. Applying support vector machines, in turn, requires a vector space embedding of the silhouettes we are concerned with. In contrast to the identification task, abnormal gait detection should abstract from personal traits. The required vector space embedding should thus capture general aspects of body silhouettes. Moreover, ideally it would be insensitive to noisy boundaries and would not require much computational effort.

The shape encoding scheme presented in this section meets all these criteria.

Given binary images as shown in Fig. 1 which result from a suitable segmentation process, we understand a shape  $\mathcal{S}$  to be a set of  $L$  pixels,  $\mathcal{S} = \{\mathbf{p}_k \in \mathbb{R}^2 \mid k = 1, \dots, L\}$ . Figure 2 visualizes the following simple procedure which computes an  $m \times n$  array of *boxes* that can be thought of as a coarser representation of a shape:

1. compute the bounding box  $\mathcal{B}(\mathcal{S})$  of a pixel set  $\mathcal{S}$  (s. Fig. 2(a))

2. subdivide it into  $n$  vertical slices; (s. Fig. 2(b))
3. compute the bounding box  $\mathcal{B}(\mathcal{S}_j)$  of each resulting pixel set  $\mathcal{S}_j$  where  $j = 1, \dots, n$ ; (s. Fig. 2(c))
4. subdivide each  $\mathcal{B}(\mathcal{S}_j)$  into  $m$  horizontal slices; (s. Fig. 2(d))
5. compute the bounding box  $\mathcal{B}(\mathcal{S}_{ij})$  of each resulting pixel set  $\mathcal{S}_{ij}$  where  $i = 1, \dots, m$ ; (s. Fig. 2(e))

Obviously, this procedure is linear in  $mn$ . Neither does it require the computation of interpixel relations nor iterative maximization (or minimization). Its average complexity therefore amounts to  $O(mn\hat{p})$  where  $\hat{p}$  denotes the average number of pixels in a box  $\mathcal{B}(\mathcal{S}_{ij})$ .

Moreover, as indicated in Fig. 3, each box  $\mathcal{B}(\mathcal{S}_{ij})$  can be understood as a generalized pixel of height  $h_{ij}$  and width  $w_{ij}$  at location  $\mathbf{x}_{ij}$ . The storage requirement of a coarse shape representation is therefore a mere  $4mn$ . For small values of  $m$  and  $n$  bounding box splitting will thus yield a fast and storage efficient abstraction of shapes. Figure 3 also exemplifies that already moderate array dimensions  $m \times n$  can produce fairly accurate representations. This subjective impression is objectified by the figures in Table 1. It summarizes measurements obtained from 1178 silhouettes of an average size of 18878 pixels that were extracted from 6 sequences of walking people. Given the height  $h$  and width  $w$  of a shape’s initial bounding box, the array dimension  $m$

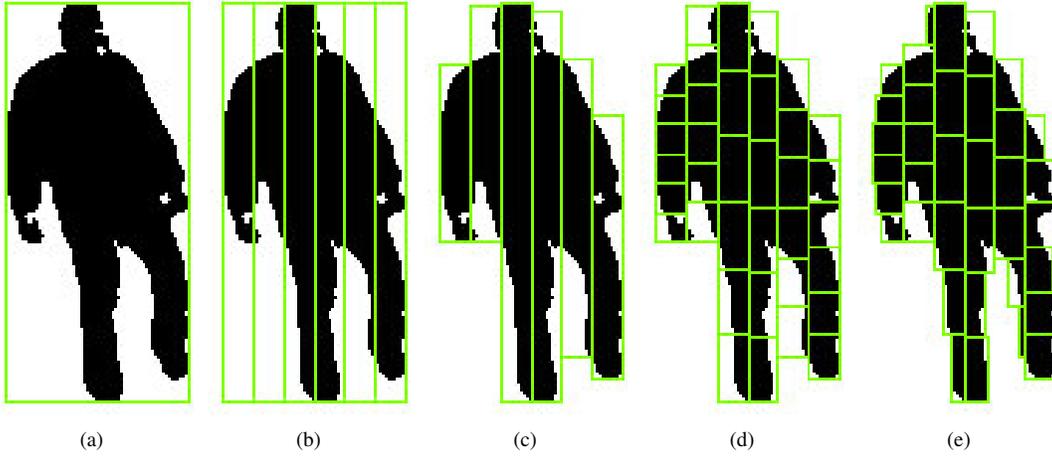


Figure 2. Example of using bounding box splitting to map a  $6 \times 6$  lattice onto a shape.

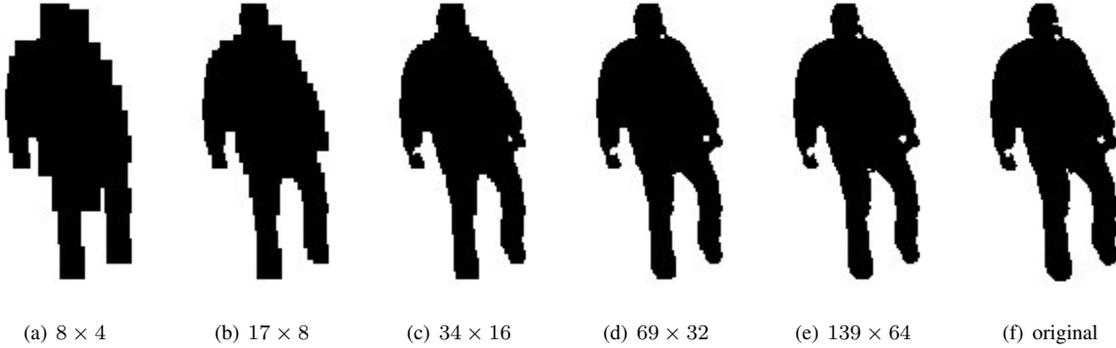


Figure 3. Approximation of a binary shape by means of box arrays of increasing dimensions  $m \times n$ .

was computed as a function of  $n$ :

$$m(n) = \left\lfloor \frac{h}{w} n \right\rfloor \quad (1)$$

where  $\lfloor x \rfloor$  indicates rounding  $x \in \mathbb{R}$  to the nearest lower integer, i.e.  $\lfloor x \rfloor = \sup\{y \in \mathbb{N} \mid y \leq x\}$ . The table lists the average compression rate, the normalized Hamming distance  $D = d_H(\mathcal{S}, \mathbf{B})/L$  between a shape  $\mathcal{S}$  and its coarse representation as a box array  $\mathbf{B}$  as well as the variance of these distances. While the compressions rate decreases slowly for growing values of  $n$ , the normalized reconstruction error decreases quickly. At a compression rate of about 93%, it already drops below 4%.

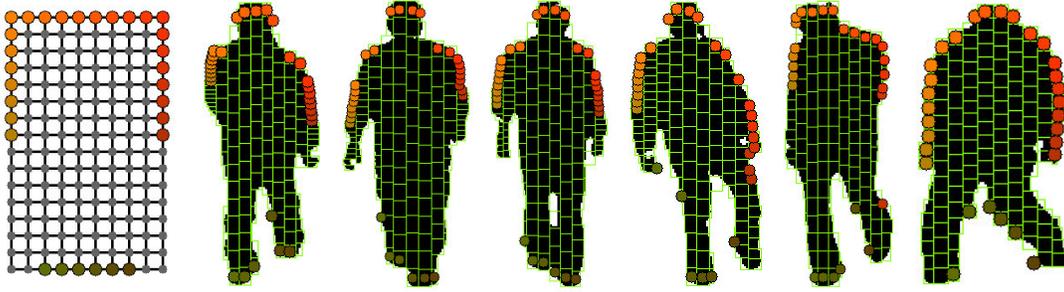
Therefore, even box arrays of small sizes provide descriptions that capture the essential properties of a silhouette. The runtime required for bounding box splitting will thus be far from threatening real time constraints that are important in most video based applications.

Furthermore, note that this simple method relying on basic computational geometry realizes a homeomorphism between 2D shapes and 2D lattices: boxes  $\mathcal{B}(\mathcal{S}_{ij})$  below other boxes will always have lesser lattice coordinates  $i$ , boxes left of other boxes will always come along with lesser lattice coordinates  $j$  and vice versa (s. Fig. 4). Due to this topology preserving nature of the box array representation of shapes, a consistent vector space embedding of shapes is straightforward.

If the vector  $\mathbf{v}$  denotes the location of the bottom left corner of the initial bounding box of  $\mathcal{S}$ ,  $w$  and  $h$  denote its width and height and the vector  $\mathbf{u}^{ij}$  denotes the center of box  $\mathcal{B}(\mathcal{S}_{ij})$ , then the coordinates

$$\boldsymbol{\mu}^{ij} = \begin{pmatrix} (u_x^{ij} - v_x)/w \\ (u_y^{ij} - v_y)/h \end{pmatrix} \quad (2)$$

provide a scale invariant representation of  $\mathcal{S}$ . Sampling  $k$  points of an  $m \times n$  lattice therefore allows to represent  $\mathcal{S}$  as



**Figure 4. A sample of  $k = 30$  points on a  $16 \times 10$  lattice and examples of how it is mapped onto different silhouettes.**

n	compression rate	$D$	$\sigma^2$
4	99.5 %	0.127	0.00136
8	98.3 %	0.065	0.00030
16	93.1 %	0.037	0.00005
32	81.3 %	0.026	0.00002
64	72.6 %	0.005	0.00001

**Table 1. Average compression rates, normalized Hamming distances  $D$  and variances of these distances obtained from box arrays of dimensions  $m \times n$ . The number of columns  $n$  was chosen as indicated; the corresponding number of rows  $m$  resulted from Eq. 1.**

a vector

$$\mathbf{r} = [\mu_x^{i(1)j(1)}, \mu_y^{i(1)j(1)}, \dots, \mu_x^{i(k)j(k)}, \mu_y^{i(k)j(k)}] \in \mathbb{R}^{2k}$$

where  $i(\alpha) < i(\beta)$  if  $\alpha < \beta$  and likewise for the index  $j$ .

Note that while this embedding is scale invariant, it is not invariant against rotations. However, since shapes of walking people usually appear to be upright, rotation invariance is not of primary concern for our application. On the contrary, exceptional feature vectors that result from mapping a lattice onto a somewhat rotated silhouette will be a strong indicator of abnormal gait.

### 3. Features for Gait Classification

Choosing array dimensions  $m \times n$  that cope with the requirements listed at the beginning of section 2 seems an arbitrary task. Obviously, too fine a grid would be more sensitive against distorted shapes or individual traits than a coarser one. A meshing that is too coarse, however, might not capture essential silhouette dynamics. As a compromise between these extremes, a grid size of  $16 \times 10$  was chosen for our experiments.

In fact, there is another justification for this choice which we borrowed from a field rather loosely related to computer vision: in fine arts, the ratio  $3 : 5 : 8$  is often considered to be a pleasing measure of the relative sizes of head, torso and legs of the human body. The three shapes on the left of Fig. 4 illustrate that this rule of thumb indeed is reasonable for upright silhouettes. Given this esthetic basis for choosing  $m = 16$  rows, our choice for the number of columns followed a similar path. Observe that  $3:5$  and  $5:8$  are approximations of the golden ratio  $\phi = \frac{1}{2}(1 + \sqrt{5}) = 1.618\dots$ . So, with geometric number theory already manifest in our considerations, we chose  $n = 10$  to approximate a golden rectangle – a rectangle with a side ratio of  $1 : \phi$  and which is generally considered esthetically pleasing.

Having decided the lattice dimensions, it remains to determine which lattice points to sample in order to provide a suitable vector representation for gait classification. Here, our choice resulted from inspecting exemplary series of shapes. It appeared that relations between head and shoulders and between the feet provide reliable cues to distinguish normal upright walking from any kind of wobbling. Figure 4 exemplifies this. The three shapes on the left correspond to instances of the gait cycle of three subjects walking normally. In each case, the head is well above the shoulders, shoulders and arms nearly enclose a right angle and the feet are not too far apart. The silhouettes on the right, in contrast, were extracted from video sequences showing abnormal gait. Here, the relation between head and shoulders as well as between the feet varies arbitrarily. Moreover, taking another look at Fig. 1 indicates that normal gait is also characterized by periodic motion of almost constant frequency of arms and feet whereas abnormal movements are aperiodic and random. A sample of lattice points that correspond to those parts of a silhouette which depict the head and shoulders and the arms and feet will thus provide valuable information for gait classification.

At this stage, the simple mapping between 2D lattices and shapes introduced above reveals its full potential. Due

to the mapping’s topology preserving properties, a sample of points with lattice coordinates on the upper and lower border of the lattice will usually, i.e. in the case of upright silhouettes, correspond to head, shoulders, arms, and feet. Figure 4 displays the  $k = 30$  lattice points we used in our experiments and where they are located on different silhouettes. Obviously, for normal gait this scheme allows to roughly keep track of limbs. Neither a feature tracking algorithm nor a recognition procedure have to be applied to infer the location of significant body parts.

Since gait is a phenomenon of inherent temporal nature, information about a single instance of a gait cycle might not be sufficient to determine whether normal or abnormal gait is being observed. In order to incorporate temporal context into the classification process, at each time step  $t$  we consider concatenated feature vectors

$$\mathbf{s}_t = \mathbf{r}_t \oplus \mathbf{r}_{t-1} \oplus \dots \oplus \mathbf{r}_{t-\Delta} \quad (3)$$

where for each  $t_l \in \{t, \dots, t - \Delta\}$  we have

$$\mathbf{r}_{t_l} = [\mu_x^{i(k)j(k)}(t_l), \mu_y^{i(k)j(k)}(t_l)], k = 1, \dots, 30 \quad (4)$$

A shape and its recent history will therefore be characterized by means of a high dimensional feature vector  $\mathbf{s}_t \in \mathbb{R}^{2k(\Delta+1)}$ . For the experiment described in the next section,  $\Delta$  was set to 1, 10 and 20, respectively, i.e. the the 60 dimensional shape descriptors of the current frame and its predecessors were combined into feature vectors  $\mathbf{s}_t$  of up to 1260 dimensions.

## 4. Experiments

In order to test the feasibility of our approach to abnormal gait detection, we experimented with a set of videos recorded in our lab. Seven subjects were asked to walk in a normal fashion as well as in an unusual way, e.g. as if they were suffering a balance deficiency such as dizziness. Each of the resulting sequences shows a single person walking in front of a homogeneous background (s. Fig. 5); the movement is either towards or away from the camera and the angle to the camera varies between 0 and about 45 degrees.

Motion segmentation from these sequences was done using a statistical background model. Subsequent filtering using a  $5 \times 5$  median followed by a connected component analysis provides shapes as seen throughout this paper.

Concerning our goal of distinguishing between normal and abnormal gait, many pattern recognition techniques would be applicable. We opted for support vector machines with radial basis kernel function because they are known to yield highly reliable class boundaries for two class problems. However, dealing with feature vectors of high dimensionality, training SVMs can be burdensome for it requires solving a quadratic optimization problem. We thus applied

$\Delta$	#frames normal	#frames abnormal	leave one out error
1	1359	1128	26.5%
10	1323	1101	11.6%
20	1283	1071	6.1%

**Table 2. Training parameters and results.**

$\Delta$	gait	#frames	accuracy
1	normal	1227	73%
	abnormal	1413	61%
10	normal	1195	77%
	abnormal	1382	70%
20	normal	1157	72%
	abnormal	1350	82%

**Table 3. Test parameters and results.**

the SVM<sup>light</sup> algorithm [8] that tackles this issue by decomposing the training process into a series of smaller tasks.

Table 2 summarizes the training parameters of three series of experiments. In each series, the same set of 7 videos showing 4 individuals was used for training. Four of these videos display normal walking, the remaining 3 are examples of abnormal gait. Because of the varying length of temporal context, the number of frames available for training is different in each experiment. Choosing a temporal context of  $\Delta = 20$ , for instance, implies that the first twenty frames of each training sequence have to be skipped from training. The figures in the last column of the table are an estimate of the quality of the resulting SVM. Performing a test on the training data produced the listed error rates. Since the error rates do not vanish, the feature space areas of normal and abnormal gait obviously overlap. However, if the the temporal context increases, i.e. if the dimension of the feature space grows, the overlap and thus the error rate decrease. This agrees with intuition and mathematical wisdom: deciding whether or not a person moves abnormal will be more reliable the longer the person is being observed. Also, in higher dimensional vector spaces data will be more sparse and there will be more space for partitioning.

Parameters and results of testing are shown in Table 3. Here, we considered 7 videos of 5 individuals, 4 of which show abnormal gait and 3 display usual walking behavior. Again, due to the different temporal context, the number of frames varies among the experiments. What is noticeable, is that while a growing temporal context considerably improves the detection of abnormal gait, its effect on nor-

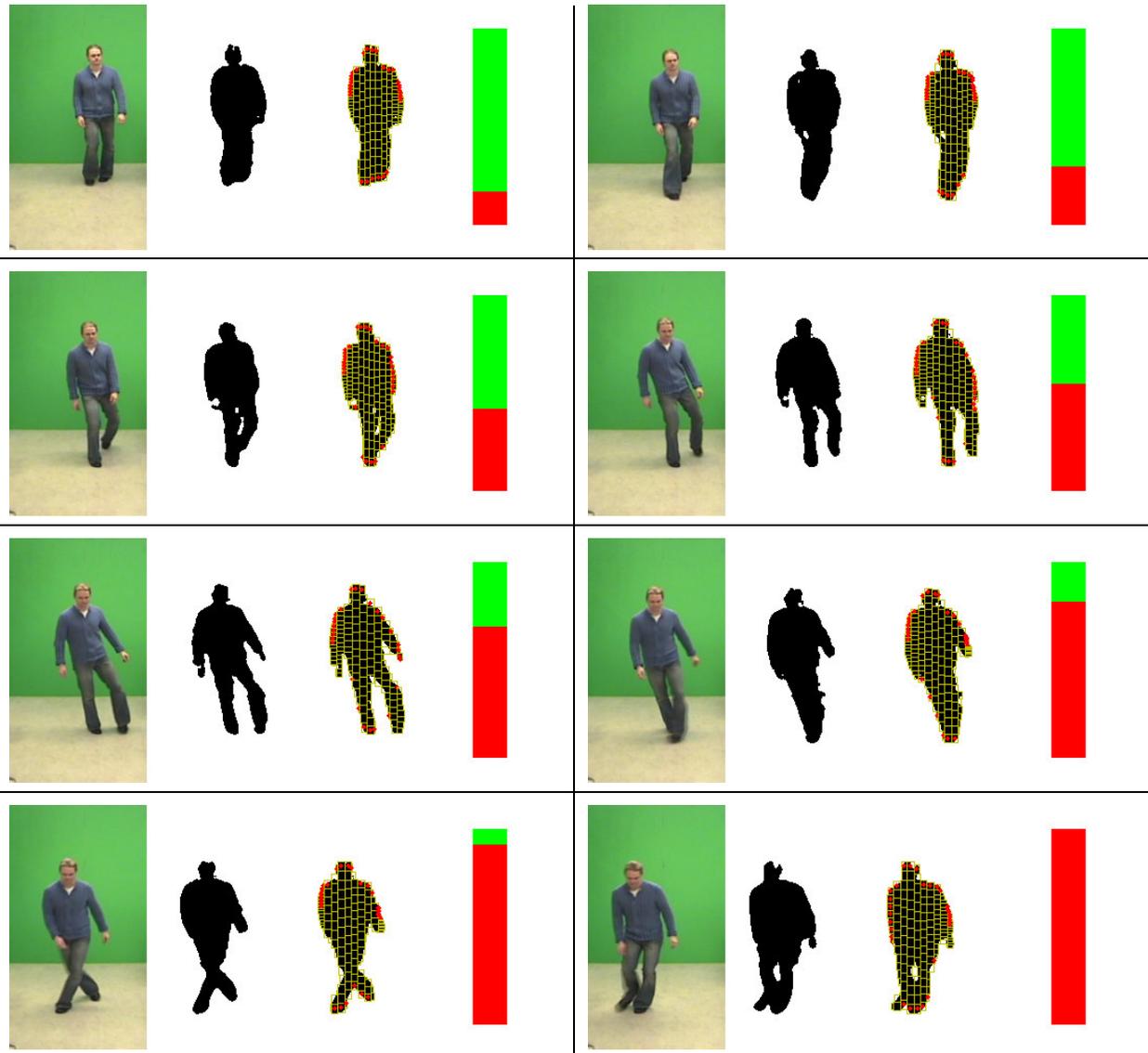


Figure 5. Example of abnormal gait detection over a period of 35 frames; the temporal context for classification was set to  $\Delta = 20$  frames. The status bar at the right of each panel indicates the percentage of gait instances that were classified *abnormal* during the last 40 frames. Thus at the beginning of this sequence, the subject's gait appeared to be fairly normal whereas at the end each of his last 40 movements has been classified as abnormal gait.

mal gait detection is inconsistent. Overall, especially for a temporal depth of  $\Delta = 20$ , the frame-wise recognition accuracy can be considered acceptable. However, the rate of false positives seems to be too high if we imagine real world applications of abnormal gait detection.

A solution for this problem is depicted in Fig. 5. The figure covers a period of 35 frames taken from a sequence of an abnormally walking subject. The status bar at the right of each panel indicates the percentage of frames that were clas-

sified *abnormal* during the last 40 frames. The more frames appear to be dubious the higher is the red bar. At the beginning of the part of the sequence that is shown here, only a few frames have been classified as abnormal gait. Corresponding to the wavering movement of the subject, however, the percentage of abnormal frames rises continuously throughout the sequence and reaches 100% in the panel at the bottom right.

The status bar thus illustrates the use of temporal context

on a higher level of abstraction. It can be seen as a temporal filter that acts on the results of frame-wise classification. If one or several frames of a sequence are misclassified, it will have little effect on the general tendency or confidence level that becomes apparent from temporal filtering.

For real world surveillance applications, it is thus possible to trigger an alarm once a certain level of abnormal gait has been detected over a given period of time.

## 5. Summary and Outlook

This paper considered automatic gait analysis as a means to deduce if an observed walking pattern appears to be normal or not. In contrast to the majority of gait analysis techniques known from literature, the problem dealt with in this paper hence requires a representation that abstracts from individual gait characteristics but allows the classification of gait across individuals. Addressing this requirement, we presented a homeomorphism between 2D lattices and shapes that enables a robust vector space embedding of silhouettes. Sampling suitable lattice points allows to roughly track the movement of limbs without that limbs have to be recognized.

Combining shape representations derived from several frames into larger feature vectors provides temporal context for the classification task. Experimental results underline that gait classification using support vector machines yields satisfiable accuracy. Temporal filtering of the classification results further improves the reliability of the presented framework because it lessens the effect of sporadic misclassifications.

Currently, we are working on porting the presented approach to a real world scenario. This requires us to cope with two phenomena not considered in this paper. First, inhomogeneous, noisy, and non-static backgrounds are hampering shape segmentation. Preliminary tests where we applied background modeling techniques similar to the ones proposed in [6] have shown promising potential to overcome this problem. Second, our approach to abnormal gait detection will have to be adapted to situations with several walking people who will inevitably occlude each other. Here, we are experimenting with very robust tracking techniques as introduced in [15] so that several silhouettes can be extracted and analyzed simultaneously.

## References

- [1] C. BenAbdelkader, R. Cutler, and L. Davis. Gait recognition using image self-similarity. *EURASIP J. on Applied Signal Processing*, 4:572–585, 2004.
- [2] A. Bobick and A. Johnson. Gait recognition using static, activity-specific parameters. In *Proc. CVPR*, volume I, pages 423–430, 2001.
- [3] J. Boyd. Synchronization of oscillations for machine perception of gaits. *Computer Vision and Image Understanding*, 96(1):35–59, 2004.
- [4] R. Collins, R. Gross, and J. Shi. Silhouette-based human identification from body shape and gait. In *Proc. Int. Conf. on Automatic Face and Gesture Recognition*, pages 351–356, 2002.
- [5] J. Cutting and L. Kozlowski. Recognizing friends by their walk: Gait perception without familiarity cues. *Bull. of the Psychonomic Society*, 9(5):353–356, 1977.
- [6] A. Elgammal, D. Harwood, and L. Davis. Non-parametric model for background subtraction. In *Proc. ECCV*, volume 1842 of *LNCS*, pages 751–767. Springer, 2000.
- [7] D. Gavrilu. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, 73(1):82–98, 1999.
- [8] T. Joachims. Making large-scale svm learning practical. In B. Schölkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*, pages 169–184. MIT Press, 1999.
- [9] G. Johansson. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14(2):201–211, 1973.
- [10] D. Kendall. Shape manifolds, procrustean metrics and complex projective spaces. *Bulletin of the London Mathematical Society*, 16(1):81–121, 1984.
- [11] J. Little and J. Boyd. Recognizing people by their gait: The shape of motion. *Videre*, 1(2), 1998.
- [12] T. Moeslund and E. Granum. A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, 81(3):221–268, 2000.
- [13] M. Murray. Gait as a total pattern of movement. *American J. of Physical Medicine*, 46(1):290–332, 1967.
- [14] S. Niyogi and E. Adelson. Analyzing and recognizing walking figures in xyt. In *Proc. CVPR*, pages 469–474, 1994.
- [15] K. Okuma, A. Taleghani, N. de Freitas, J. Little, and D. Lowe. A boosted particle filter: Multitarget detection and tracking. In *Proc. ECCV*, volume 3021 of *LNCS*, pages 28–39. Springer, 2004.
- [16] S. Sarkar, P. Phillips, Z. Liu, I. Vega, P. Grother, and K. Bowyer. The human id gait challenge problem: Data sets, performance, and analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(2):162–177, 2005.
- [17] D. Tolliver and R. Collins. Gait shape estimation for identification. In *Proc. Int. Conf. on Audio and Video-Based Biometric Person Authentication*, volume 2688 of *LNCS*, pages 734–742. Springer, 2003.
- [18] A. Veeraraghavan, A. Chowdhury, and R. Chellappa. Role of shape and kinematics in human movement analysis. In *Proc. CVPR*, volume I, pages 730–737, 2004.
- [19] L. Wang, H. Ning, W. Hu, and T. Tan. Gait recognition based on procrustes shape analysis. In *Proc. ICIP*, volume III, pages 433–436, 2002.