Human Gait Classification Based on Hidden Markov Models

Dorthe Meyer*

Universität Erlangen–Nürnberg
Lehrstuhl für Mustererkennung (Informatik 5)
Martenstr. 3, D-91058 Erlangen, Germany
Email: {demeyer}@informatik.uni-erlangen.de

Abstract

This paper describes a system for automatic gait analysis. In most clinical systems markers are used to determine the trajectories. We use a system for object recognition without segmentation to track body parts. From these trajectories periodic features are extracted. Another method to determine feature vectors is based on the optical flow computed by monotony operators. Both methods do not presume any markers. They are used here to produce sequences of feature vectors. These sequences of feature vectors are regarded as random variables. They are used to train hidden Markov models for different kinds of gait. The models will be used for gait classification.

1 Introduction

Application of gait analysis can be found in several fields, for example medical diagnosis, physical therapy and sports. It is used to receive information about gait disorders of patients with knee or hip pain, or tumors. It is also possible to control cycles of motion for rehabilitation or training.

To analyze human gait mostly motion parameters like angular acceleration, velocities and displacements of different parts of the body are used. Especially the legs, the excursions of the hip and knee seem to be important.

In most medical examination systems the trajectories are determined by markers which are attached to several points of the body. There are several problems using markers. As the information you get is that of the skin surface and not of the joint, it might be necessary to determine for example several points around the wrist. Another problem is the shifting of the skin surface when the person is moving, which causes variations of the marker positions. Patients also may feel obstructed walking with stickers all over their body.

The evaluation in clinics is mostly done by the doctor using his experience. Our aim is to develop a system which classifies the gait automatically without the use of markers.

One example for motion analysis using markers is given in [11]. He attaches LEDs to the body, tracks them and computes the trajectories. The periodicy of the motion is used for evaluation by matching the curvature of one period of the trajectory with the model trajectory which consists of one period.

Several models of the human body are used for the localization of persons or body parts. [10] uses a model of the human body consisting of 14 cylinders with elliptic cross sections. He matches the lines of the image with the contours of the projected model. Hidden contours of the model are removed. [2] generates a 3–D model of the human body consisting of tapered super–quadrics.

[5] compares the static segmentation with a segmentation using motion information. He describes the limbs as ribbons which are found by region tracking. [3] detects different body parts by an iterative approach using multiple views. Starting with a single deformable model, this is segmented into two parts if the model does not fit the following frame.

There also exist approaches which use just the local motion information for classification. [1] computes a binary motion image and a motion history image to describe the history of motion

*The author is member of the center of excellence 3D image analysis and synthesis sponsored by the Deutsche Forschungsgemeinschaft.

in a blurred sequence. [7] computes local motion statistics in xyt–cells. The feature vector consists of the summed normal flow in each cell. The classification of periodic action is done by a 3–D template match. [6] computes features from the optical flow field. The difference of the phase of these features in periodic actions are used for recognizing people by their characteristic gait.

[9] uses the gray level values of rows and columns in an image sequence or in difference images to extract features for gesture recognition. He uses hidden Markov models and a neural net for classification.

The approach presented in this contribution does not presume any markers. Two methods for extracting features are described. One is based on trajectories of different body parts, the other on the optical flow field computed by monotony operators. We distinguish different kinds of gait for recognition. The successful results using statistical methods like hidden Markov models in speech recognition suggest their use also for motion analysis. For the classification step from every image of a sequence a feature vector is extracted. This vector is considered to be the output of a hidden Markov model.

The paper is structured as follows. In section 2 we introduce the approach and describe the use of hidden Markov models for gait classification. We give an overview on the system. Section 3 describes the two ways of extracting features. In section 4 we present experiments, and we give an outlook on future work in section 5.

2 An approach for hidden Markov models applied to gait classification

Our aim is to classify different kinds of gait like walking, running, hopping and limping from image sequences. The different classes are denoted as Ω_κ, $\kappa = \{1, \ldots, 4\}$. The person is moving from the left to the right, but the action is also containing some periodic aspects. The period duration T covers one step with the right leg and one with the left leg. To describe the action completely, the duration T and the step width are important, but their determination is not considered here.

The data we use are sequences of images of a person moving as it can be seen in Figure 1. We

Figure 1: Example of a gray level image, frame 011 of an image sequence.

will extract one observation vector from two succeeding frames, so $N+1$ images will lead to N feature vectors. The observed feature vector in the n–th frame is denoted by $\boldsymbol{o}_n$, so the whole observed sequence will be described by a random variable $\boldsymbol{O} = \langle \boldsymbol{o}_1, \ldots, \boldsymbol{o}_N \rangle$. The dimension of the vector $\boldsymbol{o}_n$ is determined by the number of extracted features. This random variable will describe in general more than one period.

The classification is done by hidden Markov models (HMMs). These have been used in speech recognition successfully. We use discrete HMMs. The output vectors $\boldsymbol{o}_n$ are quantized before training the HMMs and testing the sequences.

The used HMMs $(\pi, \boldsymbol{A}, \boldsymbol{B})$ consist of I states $\boldsymbol{S} = \{S_1, \ldots, S_I\}$. In the training phase the initial state probability π, the state transition probability matrix $\boldsymbol{A}$ and the output probability $\boldsymbol{B}$ are computed. We consider HMMs of degree one, the actual state just depends on the preceding state.

We expect several characteristics of the HMMs. A large training set should induce a uniform initial state probability as there is no assumption for the person to start in the first frame. The training is expected to end up in a cyclic left–to–right model, as the state transitions do not go backward in time, but include a periodic motion.

For every kind of gait one HMM is trained. In the classification phase the probabilities for each HMM to generate the observation sequence is computed and maximized, which means $\operatorname{argmax}_\kappa p(\boldsymbol{O}|(\pi, \boldsymbol{A}, \boldsymbol{B})_\kappa)$ has to be found.

Figure 2 gives an overview of the different steps

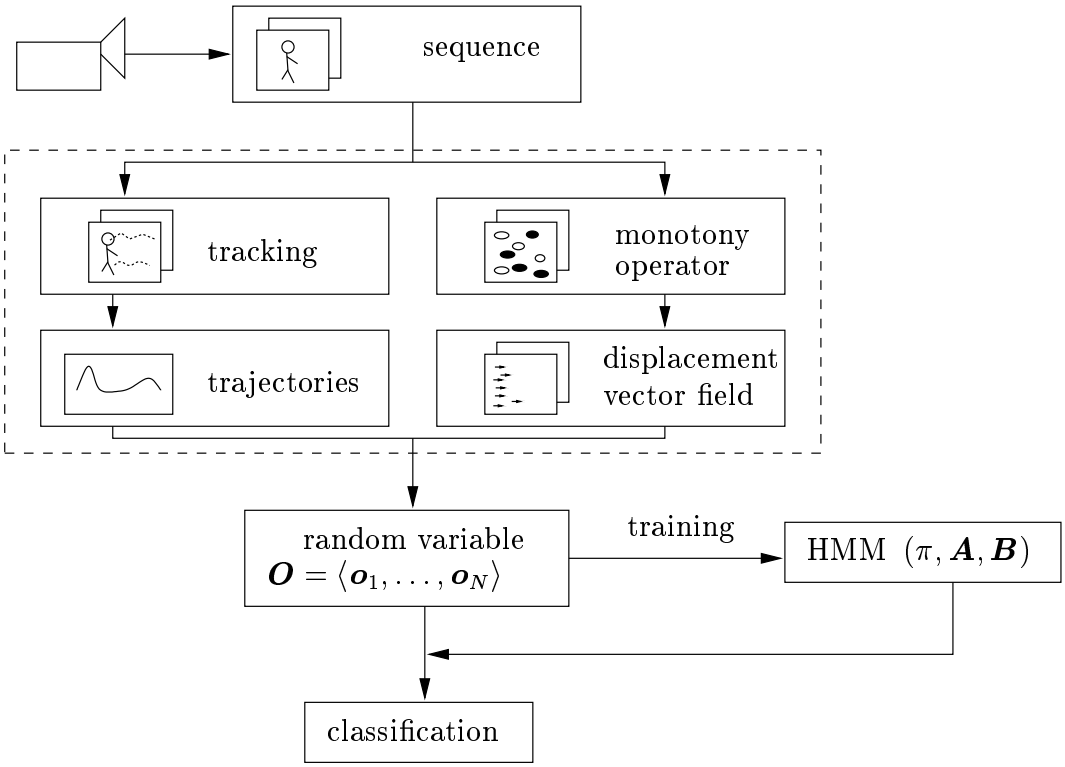

Figure 2: Overview of the different steps. The feature extraction is shown in the dashed box, the two methods of monotony operators and trajectories are used alternatively.

of feature extraction, tracking and classification. The feature extraction is described in section 3.

3 Features for gait recognition

We extract one observation vector from two succeeding frames of an image sequence. The features should be independent from the person and the frequency with which this person is walking. The feature vector is assumed to contain the motion information and be periodic in T. We consider two different methods to extract the vector. One is based on trajectories, the other one on optical flow.

3.1 Describing action by the motion of body parts

3.1.1 Tracking body parts in image sequences

Different body parts are localized and tracked in each image sequence. Important parts for describing human action are for example the head, the feet and the legs, as they contain most of the information how somebody moves.

We track three body parts, the head, the lower part of the right leg and the right foot. The statistical system which is used to train and localize objects is described in [8]. The training of the body parts from images in one sequence is done in an iterative way of localization and training. In the first images of the sequences the body parts are initialized as shown in Figure 3. The marked rectangles are used as object models for the first training and the localization in the second frame. The search for a body part in two successive frames is done locally as we consider smooth moving.

The features used for localization depend on the gray-level of the image, which are also represented by the clothing. This is still a problem, because for every person another model has to be trained. These problems may be solved using other features for detection in future.

In several sequences of the same person the body parts do not have to be trained for every sequence. The parts are searched in the whole

Figure 3: Initialization of the head, the foot and the leg in the first image of a sequence.

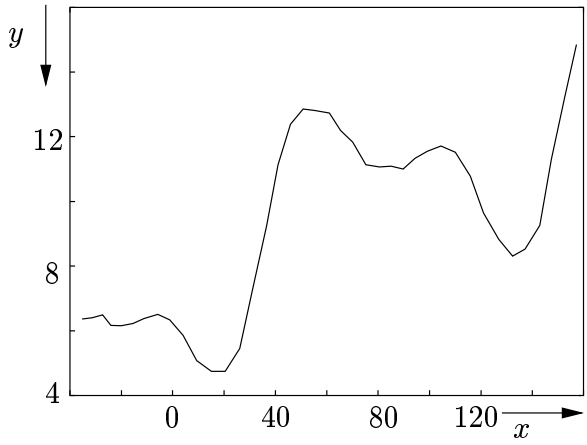

Figure 4: Trajectory of the head of a limping person. The axes show the position relative to a reference position which is determined in another sequence of the same person.

image in the first frame of the sequence. Afterwards localization is done locally.

The head is the easiest part to track. It preserves its shape and appearance over the whole image sequence and there are no problems caused by occlusion. The feet are more difficult to track, they are moving faster, in different directions, are partly occluded and change their shape.

There are also some problems because sometimes the feet are mixed up by the localization system. We solve this problem by using the information we get from the position of the head. If one foot is under the head (the same y–position), it will keep its trajectory, the search space is reduced. We also use a larger region covering the whole part of the leg under the knee (shinbone and foot). This region can be found easier. The foot or leg are then searched near that region which leads to results which are more stable.

Figure 4 shows the trajectory of the head we detected in an image sequence of a limping person, someone who was told to walk with a stiff knee. The (x, y)–position relative to a reference point is shown. The reference point $(0, 0)$ is the position of the head in the first frame of a sequence showing this person walking. The trajectory shows about three steps, two right ones and a left one.

3.1.2 Feature extraction from trajectories

The trajectories of the body parts contain the information we need for extracting the features. We just use the position of the parts. The rotation in the xy–plain is also given by the system of [8], but we do not consider it yet. We compute the displacements of body parts in x– and y–direction which are denoted v_x and v_y. They are derived from two succeeding frames:

$$v_x = \frac{x_{n+1} - x_n}{\Delta t}$$

$$v_y = \frac{y_{n+1} - y_n}{\Delta t}.$$

(x_n, y_n) is the position of a body part in the n–th frame. Δt denotes the frame rate of a sequence. The features are independent of sequences taken with different rates. There are two features per trajectory. Considering three body parts results in a six-dimensional vector.

Figure 5 shows one component of the vector extracted from the trajectory of Figure 4. It could be seen that the position of the head describes decreasing y–positions. This is caused by the person who is not moving exactly parallel to the image plain. In the feature vector this effect is removed, the values v_y appear periodic.

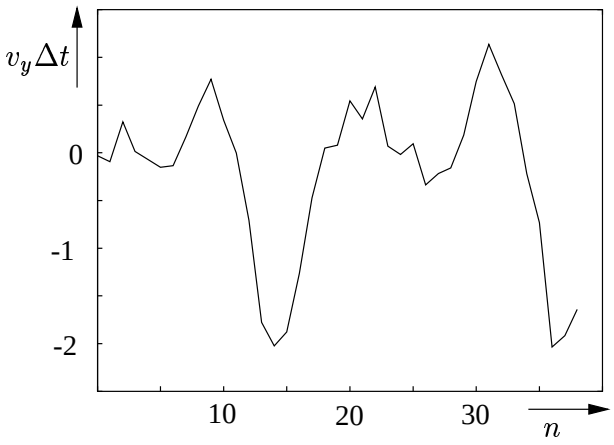

Figure 5: v_y of a head of a limping person computed from the trajectory in a sequence of 42 frames.

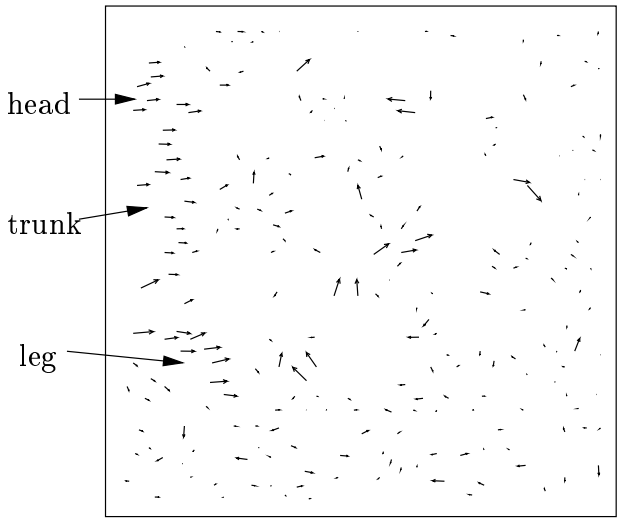

Figure 6: Displacement vector field computed by monotony operators.

3.2 Optical flow based motion recognition

Another possibility to receive features describing the motion information is the direct use of the optical flow field. This method does not depend directly on the clothing of the people walking. These features depend on the method computing the optical flow.

One requirement for the use of displacement vector fields is that different velocities of body parts should be distinguished. Small parts like the arms are moving in another direction than for example the leg or the trunk. So a system for determine optical flow assuming smooth motion is not describing the details of the real human action. One possibility to avoid such problems is the use of monotony operators.

3.2.1 Displacement vector field computed by monotony operators

The method of monotony operators is described in [4]. The monotony operator computes so called blobs in every image of a sequence. These blobs represent local minima and maxima of the gray value in the bandpass filtered image. Their position in two successive frames is used to compute the displacement vector field.

Different bandpass filters produce different sizes of blobs and therefore lengths of displacement vectors. Several filters are applied to the images in a hierarchic way. Larger displacements can be detected and details preserved.

An example of the displacement vector field is shown in Figure 6. The walking person can be seen on the left. The head, the trunk and one leg moving forward can be distinguished. The second leg does not move in this frame.

Considering the vector field, we extracted those vectors which describe the same, the main direction. We determine the center of these vector group. This point is used to detect the persons trunk. A smaller frame is cut from the flow field. So we get a sequence of smaller images which contain only the walking person.

3.2.2 Feature extraction from the optical flow field

Features can be derived directly from the optical flow field. The displacements in x– and y–direction related to the frame rate are denoted $u(x, y)$ and $v(x, y)$.

There are different possibilities to derive features from these vectors. The mean of the displacement vectors is

$$\bar{u} = \frac{\sum u(x,y)}{\text{number of vectors}},$$

$$\bar{v} = \frac{\sum v(x,y)}{\text{number of vectors}}.$$

The center of gravity of the velocities is de-

noted as

$$y_{S,u} = \frac{\sum u(x,y) \cdot y}{\sum u(x,y)}$$

It should vary if for example the foot is moving fast and the trunk does not move as in some states of hopping. It may be more constant if the whole body moves forward slowly (limping).

It is also possible to consider the main direction which varies in the whole sequence periodically

$$\phi = \arctan \frac{\bar{v}}{\bar{u}}.$$

We did not use all of these features yet. Some others like variance require a very dense motion field.

4 Experiments

We present first qualitative experiments of classification. Our experiments are based on sequences like the one shown in Figure 1. We cut the images of these sequences to get square images of size 256×256 pixels as shown in Figure 3. There are sequences taken of 13 different people, each performing four kinds of motion: walking, running, limping and hopping. We have 21 sequences of walking and limping, 18 of people running and 17 hopping. We are aware of the fact that this is nor enough data to train a HMM nor to receive expressive results. We present a first qualitative result of the approach.

We tracked three body parts in the square images, the head, the right leg and the right foot. The head is the most stable one to track in all sequences. We used its trajectory for classification. This results in a two–dimensional feature vector, containing the v_x and v_y component.

Using features derived from the flow field, the mean $\bar{u}$ and $\bar{v}$ were computed. Computing the variance or the center of u or v seems not to be promising as the displacement vector field computed by monotony operators is not very dense. This method also leads to a two–dimensional feature vector. We performed experiments with the square images and with the original frames. The original ones consist of a larger number of frames per sequence to show the persons walking.

The four hidden Markov models were trained with a set of 18, 18, 16 and 15 sequences. There are 10 sequences left which are not included in the trainig data and only used for testing. We tested all sequences, but distinguished those included in the training data and not.

The quantization was performed for 15 clusters. HMMs with 5 states were trained. The results are shown in Table 1.

Both methods of feature extraction seem to work with the hidden Markov models. As the number of training and testing data is too small, it is not possible to say which method is better. In both cases there are also still more features which should be used. Especially the results of using longer sequences show that a larger training set is necessary, the data we used is just not enough.

The recognition rates for larger sequences are much lower than those of the smaller ones. One reason for the worse rates can be found in the larger image size. The person has the same size, but there is more noise in the image which will be included in the training data. The small images are those just covering the person. They are produced by cutting a small area where the motion is detected in the optical flow field.

Considering the HMMs and their probability, we realized that the initial state probability vector is not uniform. This is caused by the fact that we cut the images to the square size pixels out of a larger one. This was done by the way that the first image was the one with the right foot on the ground. This is the most stable position to initialize the body parts with a rectangular area which we need for the localization of body parts. Of course this makes the recognition a bit easier as the beginning state is the same in all sequences. The longer sequences start with frames with persons in different states.

More experiments with a larger number of states, 15 states for a HMM, were performed. The results can be seen in Table 2. The vector quantization is done with 15 and 30 clusters. The results are worse than the ones for 5 states only. One reason can be the insufficient training data as more states require more images, but it seems not to be necessary to have so many states to describe the action.

We also evaluated which kind of gaits are recognized best. The results are listed in Table 3. The results show that this depends on the features we used. Tracking the head, we received good rates for hopping persons, but poor ones for limping people. It seems to be understandable why hop-

	training set	only testing	sum
trajectories	74.6% (50)	70% (7)	74.0% (57)
flow (square)	76.1% (51)	80% (8)	76.6% (59)
flow (large)	55.2% (37)	70% (7)	57.1% (44)
flow (small)	61.2% (41)	50% (5)	59.7% (46)

Table 1: Recognition rates applying HMMs with 5 states.

	training set	only testing	sum
trajectories (15)	71.6% (48)	50% (5)	68.8% (53)
trajectories (30)	67.2% (45)	30% (3)	62.3% (48)
flow (square, 15)	83.6% (56)	60% (6)	80.1% (62)
flow (large, 15)	68.7% (46)	50% (5)	64.9% (50)

Table 2: Recognition rates for HMMs with 15 states with 15 and 30 clusters.

ping is recognized well. The head is describing a characteristic curve for hopping people, a steep, fast increase in the y–direction. Limping persons are not that easy to be recognized by the trajectory of the head, it may be too similar to others. The result may be better using more features, for example the trajectory of the foot.

Considering the results using the flow field, it is noticeable that the rates for recognizing hopping people is the best for the small images, but much worse for the larger sequences.

5 Future

In future the first topic to concentrate is to test the HMMs with a larger set of training and testing data. The data set used here is not large enough for a system based on statistics.

Other improvements are the use of more features. At least the other 2 body parts should be considered in the case of using trajectories. Computing the optical flow, the center and main direction of the flow are possible features, as the variance or the distribution of direction.

There are also still possible improvements in the step of feature extraction. For trajectories the system of tracking should be independent on the person and the clothing. We will work on these topics to get a more stable system with expressive results.

References

[1] J.W. Davis. Appearence-based motion recognition of human actions. Technical Report TR No. 387, M.I.T. Media Lab Perceptual Computing Group, Massachusetts, 1996.

[2] D.M. Gavrila and L.S. Davis. Tracking of humans in action: A 3D model-based approach. In *ARPA Image Understanding Workshop*, Palm Springs, 1996.

[3] I.A. Kakadiaris and D. Metaxas. 3D human body model acquisition from multiple views. In *Proceedings of the* 5^{th} *International Conference on Computer Vision (ICCV)*, pages 618–623, Boston, June 1995. IEEE Computer Society Press.

[4] D. Koller, K. Daniilidis, T. Thorhallson, and H. Nagel. Model–based object tracking in traffic scenes. In G. Sandini, editor, *Computer Vision - ECCV 92*, pages 437–452, Berlin, Heidelberg, New York, London, 1992. Lecture Notes in Computer Science.

[5] M.K. Leung. First sight: A human body outline labeling system. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(4):359–377, 1995.

[6] J.J. Little and J.E. Boyd. Describing motion for recognition. In *International Symposium on Computer Vision*, pages 235–240, Coral Gables, 1995.

[7] R. Polana and R. Nelson. Recognizing activities. In *IEEE Conference on Computer*

	walking	running	limping	hopping
trajectories	76.2% (16)	80.9% (17)	50.0% (9)	88.2% (15)
flow (square)	71.4% (15)	76.2% (16)	72.2% (13)	88.2% (15)
flow (large)	47.6% (10)	85.7% (18)	61.1% (11)	47.1% (8)

Table 3: Recognition rates for different classes using HMMs with 5 states.

Vision and Pattern Recognition, pages 815–818, Seattle, Washington, 1994.

[8] J. Pösl and H. Niemann. Statistical 3–D object localization without segmentation using wavelet analysis. In *Proceedings of the* 7^{th} *International Conference on Computer Analysis of Images and Patterns (CAIP)*, Kiel, Germany, September 1997, to appear.

[9] G. Rigoll, A. Kosmala, and M. Schuster. A new approach to video sequence recognition based on statistical methods. In *Proceedings of the International Conference on Image Processing (ICIP)*, volume 3, pages 839–842, Lausanne, Schweiz, September 1996. IEEE Computer Society Press.

[10] K. Rohr. Towards model-based recognition of human movements in image sequences. *Computer Vision Graphics and Image Processing*, 59(1):94–115, 1994.

[11] P.S. Tsai, M. Shah, K. Keiter, and T. Kasparis. Cyclic motion detection for motion based recognition. *Pattern Recognition*, 27(12):1591–1603, 1994.