Model Based Extraction of Articulated Objects in Image Sequences for Gait Analysis

Dorthe Meyer*, Joachim Denzler and Heinrich Niemann
Universität Erlangen–Nürnberg
Lehrstuhl für Mustererkennung (Informatik 5)
Martensstr. 3, D–91058 Erlangen, Germany
email: {demeyer, denzler}@informatik.uni-erlangen.de

Abstract

This paper describes an approach to the extraction of articulated objects which will be used for gait analysis. In most medical applications markers are used to determine trajectories of different body parts. This approach works without any markers. Monotony operators which compute the displacement vector field are used to initialize a contour based tracking algorithm — called active rays — for several body parts which are important for gait analysis. The contours of different parts of the human body are extracted and tracked. These parts are approached by simple 3D geometric objects (blocks), which 3D position and motion are estimated for the each image of the image sequence. Then, the trajectories of the moving parts represented by the 3D blocks can be determined and used for classification of different gait disorders.

1 Introduction

Application of gait analysis can be found in the fields of 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 we have to know the motion parameters like angular acceleration, velocities and displacements of different parts of the body. Especially the legs, the excursions of the hip and knee are important.

In most medical examination systems the trajectories are determined by markers which are attached to several points of the body. Another possibility is the use of adaptive templates, but there has to be a special feature to track, too. There are several problems using markers. To get the information of the joint instead of the skin surface, it might be necessary to determine for example several points around the wrist. Another problem is that the skin surface is shifting, when the person is moving, so that the positions of the markers vary. Patients also may feel obstructed walking with stickers all over their body.

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

One example for motion analysis using markers is given in [12]. 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 a model trajectory.

There are also several other approaches for motion tracking which may not be used in clinics. [11] 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. [5] generates a 3D model of the human body consisting of tapered super–quadrics. He uses several orthogonal views to track humans in action.

[8] compares the static segmentation with a segmentation using motion information. He describes the limbs as ribbons which are found by region tracking. [6] 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 exist approaches which do not use any segmentation, but just the local motion information. [1] computes a binary motion image and a motion history image to describe the history of motion in the sequence. [10] computes local motion statistics in xyt–cells. The feature vector consists of the summed normal flow in each cell. Different kinds of activity are recognized by this way, but it is not possible to detect details, which are needed for a medical diagnosis. [9] 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.

The approach presented in this contribution does not presume any markers. Different parts of the human body are tracked independently by a contour based tracking algorithm [3]. To automatically localize different body parts, we apply monotony operators to the first two images of the image sequence in order to compute the displacement vector field. Then, each body part is extracted and tracked by its contour. In order to extract the 3D motion and position of the body part, we estimated the shape and position of simple 3D blocks, by matching the 2D contour extracted in the image with the contour of the projection of the 3D block onto the image plane. By estimating these positions in the whole sequence we get the trajectories of the different body parts.

The paper is structured as follows. In section 2 we summarize the whole approach and explain the three steps: automatic detection of the body parts, and initialization of the contour based tracking algorithm, the tracking algorithm itself, and the estimation of 3D motion of the body parts. In section 3 we present experiments, and we give an outlook on future work in section 4.

2 A Three Step Model Based Approach

In this paper a three step model based approach is presented. Neither special markers on the human body nor a manual selection of "good" features for tracking are necessary. The approach makes use of knowledge about the human body (the relative sizes of body parts and the relations between the parts) and the motion information based on optical flow computation. The steps of the approach are shown in Fig. 1.

2.1 Optical flow based initialization of the segmentation

The initialization of the segmentation is done automatically using information on the optical flow. The method is based on monotony operators which is described in [7].

The monotony operator computes so called blobs in the first two images of the sequence. These blobs represent local minima and maxima of the gray value in the filtered image. Their position in two successive frames is used to compute the displacement vector field. The matching of the blobs is done hierarchically, different bandpass filters are needed to get different sizes of blobs and therefore lengths of vectors. The resulting displacement vector field of an image of a walking person can be seen in Fig. 2.

There is no smoothing affecting the directions of neighbored vectors. So different parts of the human body moving in different directions can be distinguished. This is important to get initialization points for legs and arms which are partly covered by the other leg or the trunk moving in another direction.

Single vectors without any other of the same direction around them are eliminated. Groups of vectors describing the same direction are assumed to belong to the same body part. The center of these vector groups are determined and used as the first initialization point. As we have knowledge about the relative position of the body parts in the image, we can determine which vector group belong to which part.

2.2 Contour based segmentation and tracking of different parts of the human body

For contour segmentation and tracking a method called active rays is applied [4]. This algorithm works similar to active contours, except that all optimization problems are reduced from 2D to 1D. Starting with a

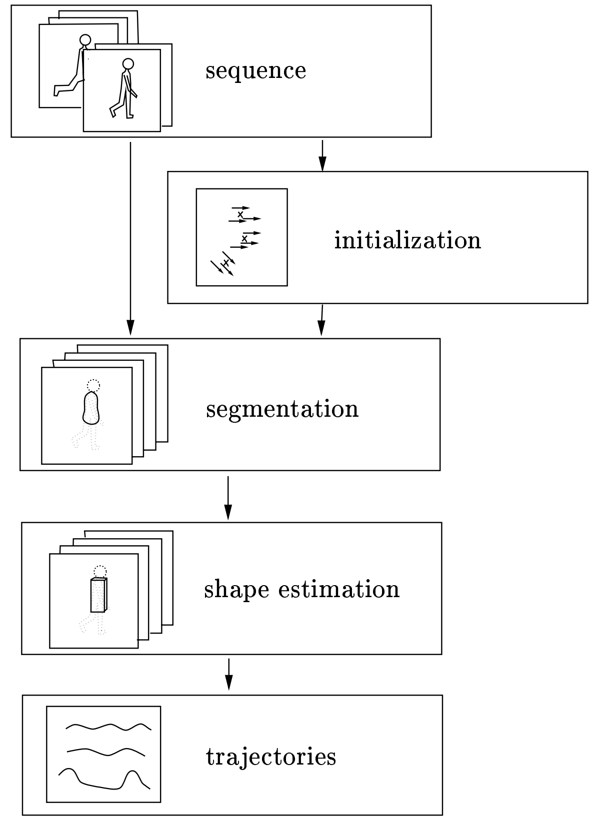

Fig. 1: Three steps of our approach for trajectories extraction: initialization, segmentation and shape estimation. As a result trajectories of the different body parts can be computed.

reference point inside the contour, we look for contour point candidates in certain directions from this reference point. For this, we sample the image along a straight line in this direction. For this 1D signal of gray values an energy is defined in a way, that we get low energies at positions of contour points on the 1D signal. By defining an internal energy similar to active contours, we combine neighboring contour points to get a contour, which is smooth and coherent in time. More details can be found in [4].

As a result of the initialization step, we get for each independently moving part of the human body the reference point, which is normally the center of gravity of the moving part. As already mentioned, this reference point is needed in order to extract the contour of the moving part by active rays, for example, the contour of the head.

The initialization step is only needed for the first image of the image sequence. For the consecutive frames, the computed reference point for image t is taken as initial reference point for image $t+1$. Thus, tracking of the moving parts of the human body is performed. For each image this results in six different contours (the head, the trunk, the legs, and the arms).

2.3 Model based 3D shape and position estimation of the different parts of the human body

In [2] a method is described, which estimates the shape of simple geometric objects in 3D based on an extracted contour in the 2D image plane. In addition, the position of the simple geometric object is also computed. With so called 3D bounding volumes, e.g. cylinder or polyhedron, coarse 3D knowledge about a moving object can be collected by tracking its 2D contour in the image plane. The idea is the following:

1. initially extract the 2D contour of the moving object by using active rays
2. estimate the parameters of the 3D bounding volume (i.e. the location in 3D and its shape), such that the projected contour $\langle \boldsymbol{v}_i \rangle_{1 \leq i \leq m}$ of the bounding volume best matches against the extracted contour $\langle \boldsymbol{v}'_j \rangle_{1 \leq j \leq n}$
3. take the parameters calculated by step 2 to update 3D knowledge about the motion and the shape of the object
4. calculate the object's position in 3D space and project the contour into the 2D image plane
5. take this 2D contour to initialize the contour extraction in the next image
6. do the contour extraction, i.e. the active rays will be attracted by edges in the image under the influence of its internal energy (details of this step can be found in [3])
7. go to step 2

This approach is transferred to the problem of estimating motion of different parts of the human body. During the previous segmentation step for each part of the body its contour is extracted and tracked. This 2D contour is taken to estimate the position of the part in 3D. For the shape of the part no estimation is necessary, because we have knowledge about the usual size of parts of the body. Starting with the head, we get constraints, which model parameters have to be chosen. The next part is the trunk, then the position of the arms and the legs are estimated.

Segmenting the whole sequence results in trajectories of the moving parts of the human body, which then can be postprocessed by a Kalman–Filter to eliminate errors. Based on the trajectories a classification into different classes of gait disorders is possible.

3 Experiments

Up to now, we have completed the steps initialization, segmentation, and shape estimation of Fig. 1. We expect to get quantitative results for the feature extraction and classification in the near future.

In Fig. 2 the displacement vector field computed from two successive frames (see Fig. 4) is shown. This sequence is taken from the University of California, Los Angeles.

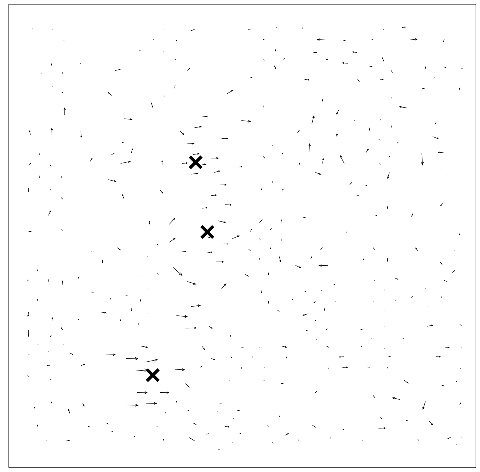

Fig. 2: Displacement vector field computed of the first two images and the initialization of the head, trunk and one leg (marked by crosses).

The displacement vectors, corresponding to the different body parts, form clusters of vectors. In contrast to the head and the legs, the displacement vectors for the trunk are sparse, because there is not much gray value structure inside the clothing. This results in less blobs and therefore in less displacement vectors. Nevertheless an initialization point can be found.

The positions of three body parts, the head, the trunk and one leg, are marked by crosses in Fig. 2. The position of the second leg cannot be determined from these two images as it is not moving. This has to be done in one of the succeeding frames. There are also difficulties to initialize the position of the arms from this image, because they are occluded by the trunk. The right arm can be initialized later on when it is moving forward.

In Fig. 3, one example of a segmentation of the trunk by active rays is shown for image number 109. The reference point has been initialized by the position, shown in Fig. 2. Based on such segmentation we have estimated the bounding volumes for the head (24 contour points), trunk (72) and the two legs (24). In brackets the number of contour points for extraction are given.

In Fig. 4 four images of an image sequence are shown, overlayed with the results of the model estimation. In image 112, an error for the size of the right leg can be seen. This wrong model estimation is caused by an segmentation error of this leg. We can eliminate this error by using a Kalman filter to smooth the estimated parameters of the bounding volume.

The computation time for the model estimation is about 60 secs for one image. The segmentation can be done within the image frame rate (40 msecs) on an SGI Onyx with R10000 processor.

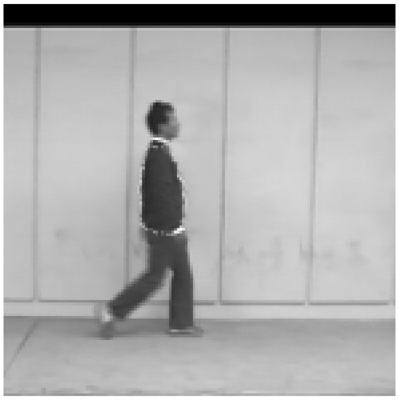

Fig. 3: Result for segmentation of the trunk by active rays for image 109 (72 contour points for representation).

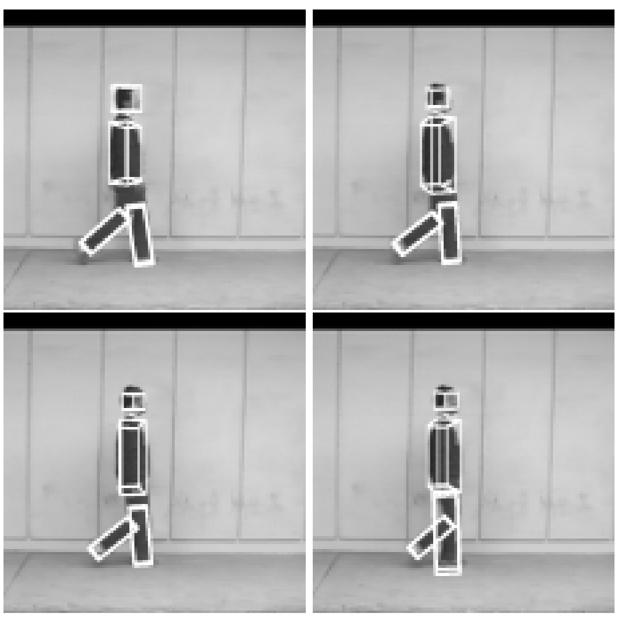

Fig. 4: Result for estimating the position of the body parts based on their 2D contours for images 109, 110, 111, 112 of the image sequence.

4 Future Work

In future work the motion of the different parts of the body will be estimated in order to get information about the trajectories. The motion has to be estimated at selected points like the joints. These points are determined by the geometric objects. The trajectories will be used for gait analysis. We well compute a feature vector for every image consisting of the displacements of the body parts. For each kind of gait, e.g. walking, running or limping, a HMM will be trained. An automatic classification system of gait will be developed.

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] J. Denzler and H. Niemann. 3D data driven prediction for active contour models based on geometric bounding volumes. *Pattern Recognition Letters*, 17(11):1171–1178, 1996.

[3] J. Denzler and H. Niemann. Active rays: A new approach to contour tracking. *International Journal of Computing and Information Technology*, 4(1):9–16, 1996.

[4] J. Denzler and H. Niemann. Active rays: A new approach to contour tracking. In *German-Slowenia Workshop on Speech and Image Processing*, pages 301–310, Ljubljana, 1996.

[5] 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.

[6] 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, Juni 1995. IEEE Computer Society Press.

[7] 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.

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

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

[10] R. Polana and R. Nelson. Recognizing activities. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 815–818, Seattle, Washington, 1994.

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

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