Analyzing gait with spatiotemporal surfaces

Sourabh A. Niyogi

Department of Electrical Engineering
and
MIT Media Laboratory
Cambridge, MA 02139
sourabh@media.mit.edu

Edward H. Adelson

MIT Media Laboratory
and
Department of Brain and Cognitive Sciences
Cambridge, MA 02139
adelson@media.mit.edu

Abstract

Human motions generate characteristic spatiotemporal patterns. We have developed a set of techniques for analyzing the patterns generated by people walking across the field of view. After change detection, the XYT pattern can be fit with a smooth spatiotemporal surface. This surface is approximately periodic, reflecting the periodicity of the gait. The surface can be expressed as a combination of a standard parameterized surface – the canonical walk – and a deviation surface that is specific to the individual walk.

1 Introduction

Walking humans are self-occluding, non-rigid articulated objects. As a result, optical flow and feature tracking typically have great difficulty analyzing the motion. However, if we examine an XYT volume of a set of walking people, some distinctive patterns reveal themselves, suggesting that alternate strategies may be useful.

Figure 1 shows a single frame of an image sequence with several people walking, and figure 2 shows the XYT volume of the sequence sliced at two heights. Uniform translation of the upper body results in tilted stripes; the moving legs generate braids. We have previously shown that one can fit XT snakes to these slices and thereby analyze the gait (Niyogi and Adelson 1994). This method avoids the need to identify and track feature points such as those required in analyzing in moving light displays.

Here we extend our previous work, treating the XYT pattern as a volume and fitting it with a spatiotemporal surface (c.f. Baker 1989, Bolles and Baker 1989). Our surface fitting routines are related to others that have been used in fitting 3-D data (Terzopolous et al 1988, Pentland and Sclaroff 1991, Cohen and Cohen 1990, McInerney and Terzopoulos 1993), but we treat the time axis somewhat differently than the spatial axes.

The surface fitting is reasonably easy because human walks tend to be similar in XYT. Thus we can define a canonical walk which can be used as an initial fit to a given observed walk. Figure 3(a) shows a surface traced out by a closed curve which outlines the upper body plus the left leg. Figure 3(b) shows a combination of surfaces that trace out both legs and upper body. The surface can be parameterized by spatial position and scale, temporal period and phase, and can be sheared in space-time according to the velocity of the walk. This surface was generated by combining data from several walkers using methods described below.

The surfaces can be used to control the parameters of a stick model (Hogg 1983, Rohr 1983, Niyogi and Adelson 1994). However, we can also use the surfaces themselves to directly accomplish various tasks, such as motion recognition and tracking.

2 Approach

Our approach is simple: First, we recover the parameters that define an individual's gait using some simple pattern analysis routines; these parameters control two canonical spatiotemporal surfaces which coarsely track the individual. Second, we deform these spatiotemporal surfaces to fit image data; this allows for accurate tracking of the individual.

2.1 Detection

To generate the initial fit we must estimate the parameters. We currently assume each human walks frontoparallel to the image plane in front of a fixed camera. Under these conditions, each walker generates a stripe in XT; slicing along this stripe allows us to use simple pattern analysis to recover gait parameters. If there are several walkers, as in the example shown, we estimate their several velocities with a Hough transform. We employ change detection to highlight moving objects; the background is recovered through median filtering and the difference between each frame of an image sequence and the background is squared and clipped to yield a new image sequence $C(x, y, t)$. A change-detected XT-slice, shown in figure 4(a), is transformed to Hough space, and the most popular straight lines are selected. The winners are shown in figure 4(b).

For each individual walker it is necessary to estimate the top and bottom heights as well as the period and phase of the walk. These parameters can be estimated by taking a vertical slice down the middle of the slanted walking pattern. Figure 5 shows a slice for a walker moving frontoparallel to the image plane: The walker maintains roughly the same size on the image plane; the legs are visible as they periodically cross the plane of the slice.

There are six parameters to be recovered, depicted in figure 6:

- x_i, x_f: the initial and final x position of the walker, at t_i and t_f. The walkers velocity is $v_x = \frac{x_f - x_i}{t_f - t_i}$.

- y^h, y^t: the bounding y coordinates of the walker, approximately the image y locations of the head and toe.

- T: The walker's period, in frames.

- ϕ: The phase of the gait, $(0 \leq \phi \leq 2\pi)$.

To recover the above parameters, we employ the following steps:

- To recover x_i and x_f, all change detected XT-slices are collapsed into one XT-slice with $C_{hough}(x, t) = \sum_y C(x, y, t)$. Hough transforms allow us to find the $x(t)$ lines. Figure 4(b) shows an XT-slice with candidate $x(t)$ signals superimposed.

- To recover y^h and y^t, a slice $O(y, t)$ is obtained from a change detected image sequence $C(x, y, t)$ with $O(y, t) = C(x(t), y, t)$; finding the convex hull of this image yields the four parameters.

- To recover T and ϕ, the lower periodic pattern is cropped between $\frac{y^h + y^t}{2}$ and y^t from the slice $O(y, t)$ to obtain a periodic pattern $P(y, t)$. An autocorrelation sequence is built: $q(m) = \sum_{n,t} P(n, t) P(n + m, t)$ using FFT analysis (inverse Fourier transform of the magnitude of the Fourier transform). A peak is found in $q(m)$ between 1 Hz and 3 Hz. If there is no significant peak, the slice doesn't correspond to a walker, and processing stops. Peaks are found using simple thresholding. The walking period T is just twice the period of the slice. The phase ϕ is obtained by using the phase at the frequency $\frac{2\pi}{T}$.

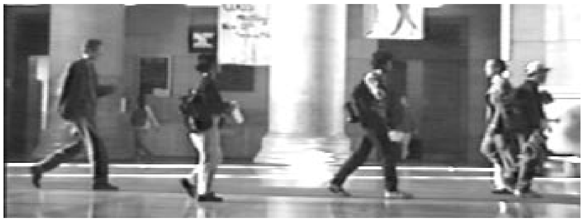

Figure 1: One frame from an image sequence with several walkers.

2.2 Modeling

The parameters derived above specify the canonical spatiotemporal surfaces that will be used to fit the walking pattern. There are two spatiotemporal surfaces: one corresponding to the top half of the body plus the left leg, another corresponding to the top half of the body plus the right leg. We perform separate analysis on the two spatiotemporal surfaces.

Following Terzopoulos and Metaxas (1991), each spatiotemporal surface $\mathbf{s}(u, v)$ is a closed tube in space-time whose intrinsic coordinates are (u, v). In our coordinate system, u is a parametric variable indexing the closed curve in XY which outlines the body plus one leg; the variable v traverses time $(t_i \leq v \leq t_f)$. The surface is implemented using a vector-valued parametric representation $\mathbf{s}(u, v) = [x(u, v), y(u, v), t(u, v)]^T$.

The spatiotemporal surface $\mathbf{s}(u, v)$ is composed of two components:

$$\mathbf{s}(u, v) = \mathbf{m}(u, v) + \mathbf{x}(u, v)$$

where $\mathbf{m}(u, v)$ is a canonical walk appropriately scaled and shifted, and $\mathbf{x}(u, v)$ is a displacement function which will be the deviation from the canonical walk.

The canonical walk $\mathbf{c}(u, v)$ is a periodic spatiotemporal surface of unit height and unit time. One period of this walk was obtained by bootstrapping from our previous work (Niyogi and Adelson 1994), in which four $x(y, t)$ spatiotemporal sheets were recovered for five walkers in 26 image sequences. We averaged across all spatiotemporal sheets to obtain four "canonical" sheets, and bridged the head and feet together with the addition of extra nodes. Arms are not present in these surfaces, but there is little reason why a more

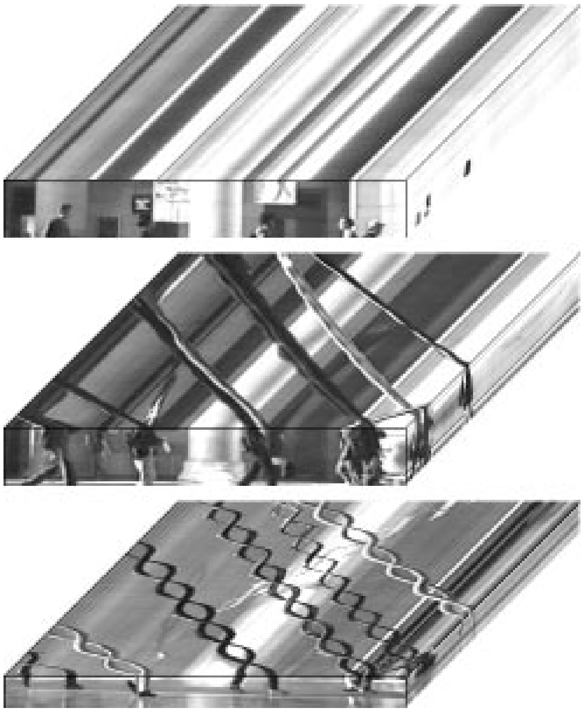

Figure 2: Slicing an XYT cube above the torso reveals stripes; slicing an XYT cube below the torso reveals braided patterns.

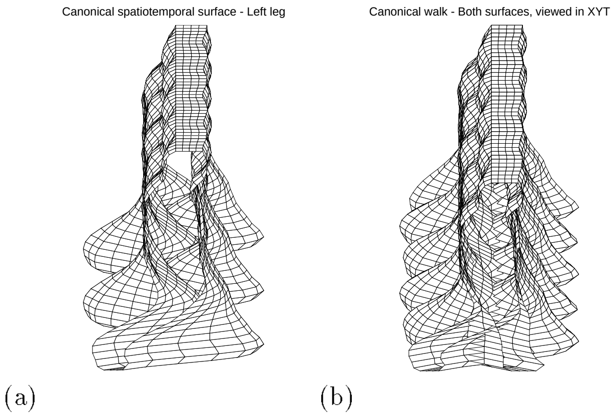
Canonical spatiotemporal surface - Left leg
Canonical walk - Both surfaces, viewed in XYT
(a)
(b)

Figure 3: (a) One of the two spatiotemporal surfaces that form a canonical walk. (b) Both surfaces of canonical walk $\mathbf{m}(u,v)$, with translation removed.

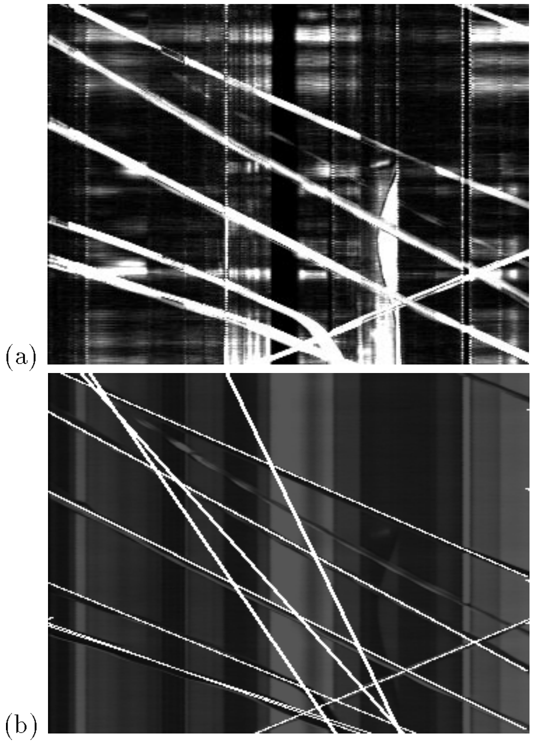

Figure 4: (a) Change detected XT slice, collapsed from all heights. (b) One XT slice, with potential walkers' translation superimposed in white.

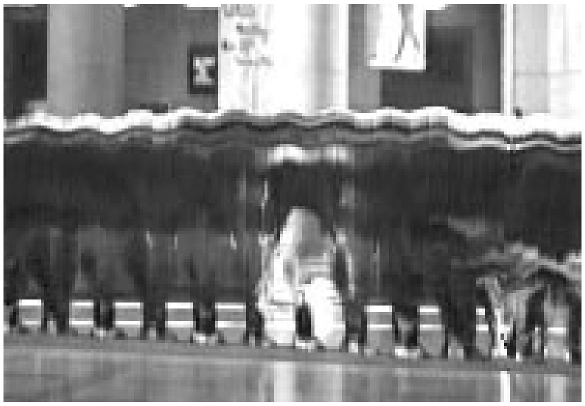

Figure 5: By slicing along the direction of walker translation, a periodic image is obtained. The white splotch in the center is a second walker walking in the opposite direction.

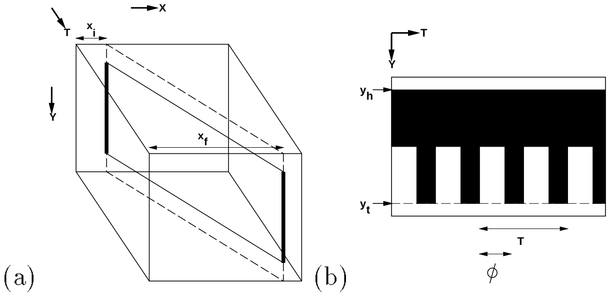

Figure 6: Diagram of the six model parameters to be recovered: (a) From a collapsed XT slice, Hough transforms allow recovery of x_i and x_f; (b) A model slice along the direction of translation, where black represent change-detected areas. Convex hull analysis on this slice yields y^h and y^t, and FFT analysis yields T and ϕ.

complicated surface description could not handle extra appendages.

Given our recovered six parameters, we can control our canonical spatiotemporal surfaces as follows:

$$\mathbf{m}(u,v) = \mathbf{t}(u,v) + \mathbf{kc}(u, \frac{v - t_i}{T} + \frac{\phi}{2\pi})$$

$$\mathbf{k} = diag(y^t - y^h, y^t - y^h, 1)$$

$$\mathbf{t}(u,v) = [x_i + \frac{v - t_i}{t_f - t_i}(x_f - x_i), y^h, 0]^T$$

Note that we approximate the size variation by linear scaling in height only, and assume the walking speed and walking period are approximately constant. Figure 3(b,c) shows two views of canonical spatiotemporal surfaces $\mathbf{m}(u,v)$ for one walker, after scaling and translation. Figure 8 shows three typical frames from an image sequence tracked with a canonical walk.

To recover a more accurate spatiotemporal surface, the displacement surface $\mathbf{x}(u,v)$ is allowed to deform so as to fit the data while maintaining smoothness. The displacement function $\mathbf{x}(u,v)$ is initially set to zero.

We used a variation of the deformable surface model described in Terzopoulos and Metaxas (1991), and further developed in McInerney and Terzopoulos (1993). A brief summary of their work is presented below for completeness. They formulate a deformation energy for a surface $\mathbf{x}(u,v)$ as

$$\begin{aligned}\mathcal{E}_p(x) \quad = \quad & \int\int \alpha_{10}|\mathbf{x}_u|^2 + \alpha_{01}|\mathbf{x}_v|^2 + \\ & \beta_{20}|\mathbf{x}_{uu}|^2 + \beta_{11}|\mathbf{x}_{uv}|^2 + \beta_{02}|\mathbf{x}_{vv}|^2 du dv\end{aligned}$$

where α_{ij} and β_{ij} specify the elasticity of the material.

The $\mathbf{x}(u,v)$ mesh is tesselated into discrete nodal points, and approximated as a weighted sum of piecewise polynomial basis functions $\mathbf{N}_i$:

$$\mathbf{x}(u,v) = \sum_{i=1}^{n} \mathbf{N}_i(u,v)\mathbf{q}_i$$

where $\mathbf{q}_i$ is a vector of nodal variables associated with mesh node i:

$$\mathbf{q}_i = [\mathbf{x}_i^T, (\mathbf{x}_\xi)^T, (\mathbf{x}_\eta)^T, (\mathbf{x}_{\xi\eta})^T]^T$$

In our case, the $\mathbf{x}(u,v)$ mesh consists of C^1 continuous rectangular finite elements defined in dimensionless coordinates (ξ, η).

The above formulation results in equations of motion:

$$\mathbf{Cq}' + \mathbf{Kq} = \mathbf{f_q}(u,v)$$

with damping matrix $\mathbf{C}$, stiffness matrix $\mathbf{K}$, and $\mathbf{f_q}$ as nodal data forces. For our preliminary work, we used a constant stiffness matrix $\mathbf{K}$ for the entire mesh, formed by five smoothness parameters of the deformation energy α_{10}, α_{01}, β_{20}, β_{11}, and β_{02}. Each node in the displacement mesh is updated with:

$$\mathbf{q}^{(\tau+\Delta\tau)} = \mathbf{q}^{(\tau)} + \delta\tau \left(\mathbf{C}^{(\tau)}\right)^{-1} (\mathbf{f_q}^{(\tau)} - \mathbf{Kq}^{(\tau)})$$

To refine our spatiotemporal surface, each node of the spatiotemporal surface is attracted to spatiotemporal edges by applying forces from the change-detected image sequence $C(x,y,t)$. We compute two force image sequences by filtering $C(x,y,t)$:

$$F_x(x,y,t) = G_x^2(x,y) * C(x,y,t)$$

$$F_y(x,y,t) = G_y^2(x,y) * C(x,y,t)$$

where G_x^2 and G_y^2 are second derivatives of Gaussians in the x and y direction. For all nodes in the mesh with coordinates at (x,y,t) and normal $\hat{\mathbf{n}}$, a force $\mathbf{f}_{\mathbf{q}_i}$ is applied:

$$\mathbf{f}_{\mathbf{q}_i} = \hat{\mathbf{n}} \cdot [F_x(x,y,t), F_y(x,y,t), 0]^T$$

Nodes are constrained to move only in the x and y direction. Only the nodes in the displacement $\mathbf{x}(u,v)$ are modified.

We iterate for three scales of Gaussian, from low-frequency to high-frequency. This has the effect of establishing large deformations quickly; smaller deformations refine the surface further. A coarse to fine approach in sampling the (u,v) mesh leads to faster convergence. Note that although we currently use operations on a change detected image sequence to apply forces on the mesh, a more sophisticated representation, perhaps based on spatiotemporally oriented filters, might be used instead.

Recovered spatiotemporal surfaces for a typical image sequence are shown in Figure 7. The surfaces are shown superimposed on the original image sequence in figure 9; compare to the coarse tracking of the canonical walk in figure 8. Note that by observing continuity in space time, the problem of occlusion is bypassed: self-occlusion of legs and occlusions of other walkers is not a significant problem.

While our current approach has only been demonstrated on frontoparallel walkers, clearly pose estimates could be found. The walker's pose can be obtained from the parameters that describe the shape of the spatiotemporal slice in figure 5. As pose varies, the appropriate spatiotemporal surface to be superimposed also changes. Depending on the estimate of walking direction, we can superimpose a different model surface estimate $\mathbf{m}(u, v)$. This would be similar in style to many of the view-based approaches currently employed in face recognition. We are currently extending our work in that direction.

3 Conclusion

Human motions generate characteristic spatiotemporal patterns, which can be fit with spatiotemporal surfaces. We have considered the case of humans walking frontoparallel to the camera. In this case, we are able to take advantage of certain regularities in gait patterns; human walks are periodic and tend to look similar between individuals. Thus we can establish a standard "canonical" walk that can be fit to individual walks by estimating a small number of parameters. Simple pattern analysis of spatiotemporal images allows us to estimate these parameters. We refine the initial estimate and deform the spatiotemporal surfaces to accurately track the individual.

Although we have only dealt with a restricted set of motions, it may be possible to generalize this approach to other motions. Any stereotyped motion pattern could be characterized with a canonical spatiotemporal surface; a dictionary of such surfaces could be used in recognition and tracking.

Acknowledgments

This work has been sponsored by the Television of Tomorrow Consortium. Tim McInerney kindly supplied code for the algorithm described in McInerney and Terzopoulos (1993), which proved invaluable. We thank John Wang and Stan Sclaroff provided helpful comments on the manuscript.

References

[1] Baker, H. H. "Building Surfaces of Evolution: The Weaving Wall," *IJCV*, 3:51-71, 1989.

[2] Bolles, R. C. and Baker, H. H. "Generalizing Epipolar-Plane Image Analysis on the Spatiotemporal Surface," *IJCV*, 3:33-49, 1989.

[3] Cohen, L. and Cohen, I. "A finite element method applied to the new active contour models and 3-D reconstruction from cross sections," *ICCV*, 1990.

[4] Hogg, D. "Model-based vision: A program to see a walking person," *Image and Vision Computing*, 1: 5-20, 1983.

[5] McInerney, T. and Terzopoulos, D. "A Finite Element Model for 3D Shape Reconstruction and Nonrigid Motion Tracking," *ICCV*, pp. 518-523, 1993.

[6] Niyogi, S. and Adelson, E. "Analyzing and Recognizing Walking Figures in XYT," *CVPR*, pp. 469-474, 1994.

[7] Rohr, K. "Towards Model-Based Recognition of Human Movements in Image Sequences," *CVGIP: IU*, 59(1): 94-115, 1994.

[8] Pentland, A. and Sclaroff, S. "Closed-Form Solutions for Physically-Based Shape Modeling and Recognition", *IEEE PAMI*, 13(7): 715-720, 1991.

[9] Terzopoulos, D. and Metaxas, D. "Dynamic 3D Models with Local and Global Deformations: Deformable Superquadrics," *IEEE PAMI*, 13(7): 703-714, 1991.

[10] Terzopoulos, D., Witkin, A. and Kass, M. "Constraints on Deformable Models: Recovering 3D Shape and Nonrigid Motion," *Artificial Intelligence*, 36: 91-123, 1988.

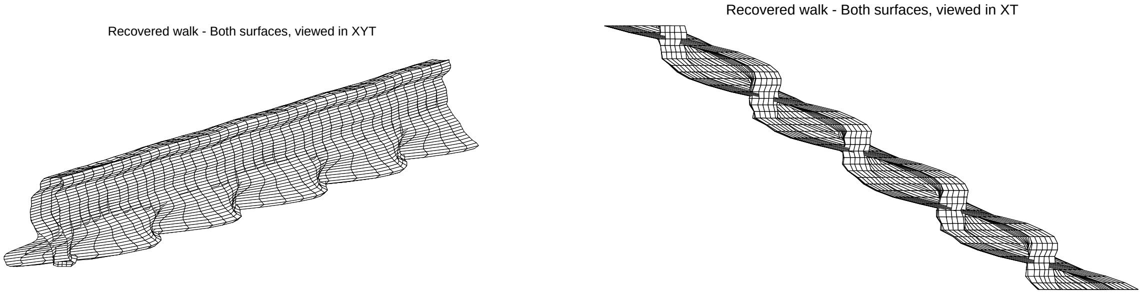
Recovered walk - Both surfaces, viewed in XYT
Recovered walk - Both surfaces, viewed in XT

Figure 7: Two views of both deformed spatiotemporal surfaces $\mathbf{s}(u,v)$.

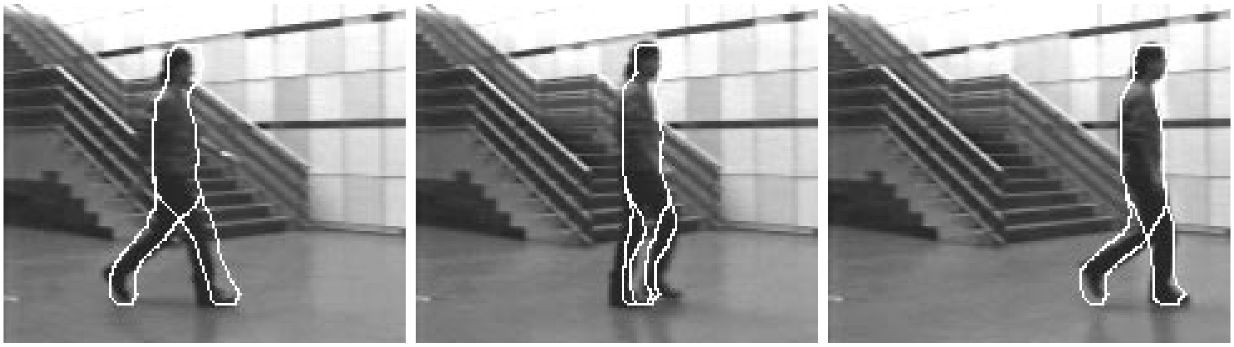

Figure 8: Three frames from an image sequence with both canonical spatiotemporal surfaces $\mathbf{m}(u,v)$ superimposed in white. The canonical spatiotemporal surface gives only a coarse fit to the walker.

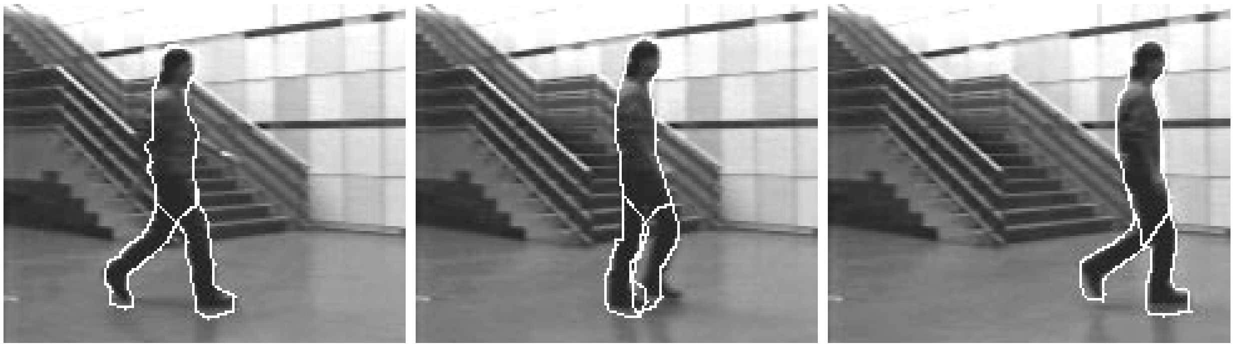

Figure 9: Three frames from an image sequence with both deformed spatiotemporal surfaces $\mathbf{s}(u,v)$ superimposed in white. The recovered surface tracks the walker much more accurately.