Reconstruction of Articulated Objects from Point Correspondences in a Single Uncalibrated Image

Camillo J. Taylor
GRASP Laboratory, CIS Department
University of Pennsylvania
Philadelphia, PA, 19104-6228

Abstract

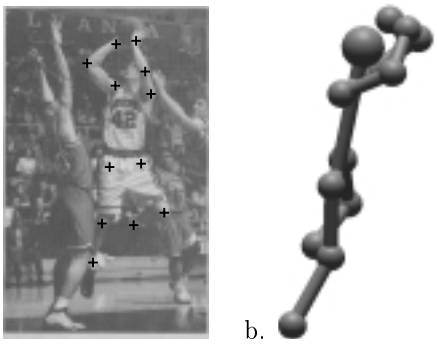
This paper investigates the problem of recovering information about the configuration of an articulated object, such as a human figure, from point correspondences in a single image. Unlike previous approaches, the proposed reconstruction method does not assume that the imagery was acquired with a calibrated camera. An analysis is presented which demonstrates that there are a family of solutions to this reconstruction problem parameterized by a single variable. A simple and effective algorithm is proposed for recovering the entire set of solutions by considering the foreshortening of the segments of the model in the image. Results obtained by applying this algorithm to real images are presented.

1 Introduction

This paper investigates the problem of recovering information about the configuration of an articulated object, such as a human figure, from point correspondences in a single image. The problem of deducing the pose of a human actor from image data has received a substantial amount of attention in the computer vision literature. This is, in part, due to the fact that solutions to this problem could be employed in such a wide range of applications. Most of the research in this area has focused on the problem of tracking a human actor through an image sequence, less attention has been directed to the problem of determining an individuals posture based on a single snapshot.

Figure 1a shows an image of an actor acquired from a newspaper clipping, while Figure 1b shows the simple stick figure model recovered from this image viewed from a novel vantage point. This model was obtained by locating the projections of the joints in the model on the image plane and applying the reconstruction algorithm described in section 2 to these measurements.

The analysis of this reconstruction problem presented in this paper is based on three assumptions.



a. b.

Figure 1: (a) represents an image containing a figure to be recovered the 12 crosses represent the estimated locations of the joints which are passed to the reconstruction procedure. (b) the recovered 3D model viewed from a novel vantage point.

Firstly, that the correspondence between the joints in the model and point features in the image is provided. Seondly, that the relative lengths of the segments in the model is known a'priori. Thirdly that the relationship between the positions of the joint features in space and their projections onto the image can be modeled effectively as a scaled orthographic projection and that the aspect ratio of the image is unity (if the aspect ratio is known the image can always be rescaled to unit aspect ratio).

The decision to employ a scaled orthographic projection model rather than assuming that the imagery was acquired with a calibrated perspective camera is one of the features that distinguishes this work from previous research such as [9]. This decision was motivated by the observation that most of the images that one might be interested in analyzing would be obtained from sources such as scanned photographs or Internet web sites where information about cameras calibration parameters would be unavailable but where a weak perspective model is appropriate.

The main technical contributions of this paper are described in Section 2 which presents a detailed analysis of the problem of recovering the pose of the model from the available image measurements. This analysis allows us to precisely characterize the set of solutions to this problem. A simple and effective algorithm for recovering all of the possible solutions is also presented.

In Section 3 the results obtained by applying the proposed technique to actual image data are presented. This section also presents the results of experiments that were designed to investigate the accuracy of the method by comparing the results to measurements obtained from independent sources.

1.1 Prior Work

Much of the early work in the computer vision literature on the problem of recovering the motion of articulated figures in images was inspired by Johansson's groundbreaking work on Moving Light Displays (MLDs) [7]. Rashid [13] describes a method which automatically segments the points in a MLD into body parts. The algorithm does not, however, attempt to provide an estimate of body posture.

A number of approaches have been proposed for tracking the configuration of articulated objects in video data. Bregler and Malik [2] describe an elegant scheme for updating the joint parameters of a stick figure model from the optical flow measurements in a video sequence by making use of the product of exponentials representation for a kinematic chain. Regh and Kanade [14] describe a scheme for tracking articulated objects with a large number of degrees of freedom such as a human hand. Pentland and Horowitz [12] present a method for estimating the motion of a human figure using a physics based model. In all of these systems, it is assumed that the initial configuration of the figure is known a'priori and the algorithm takes responsibility for updating the pose parameters in subsequent images. The reconstruction scheme described in this paper could be used to provide such an initial estimate.

Rohr [15] and Hogg [6] both describe impressive schemes for extracting the posture of pedestrians from video data. In this case, the assertion that the figure is walking provides a powerful constraint on the postures that can occur in the sequence. No such constraint is assumed in the method presented in this paper.

A number of researchers have described schemes for estimating the posture of a figure from image measurements under the assumption that the calibration parameters of the camera are completely known a'priori [17, 9, 11, 1]. This assumption makes it difficult to apply these techniques to real-world images which are usually obtained with uncalibrated camera systems.

A number of schemes have been proposed for detecting human figures in an image sequence without estimating their 3D configuration. Fleck, Forsyth and Bregler [4] describe an interesting scheme for locating naked human figures within an image. Wren, Azerbayejani, Darrell and Pentland [16] have proposed a robust algorithm for tracking a human figure in an image sequence without estimating the individuals posture. Iwasawa, Ebihara, Ohya and Morishima present a real-time method for localizing a human in thermal images. Morris and Rehg [10] propose a 'Scaled Prismatic Model' for tracking articulated objects in video data. In their tracking scheme, they maintain estimates for the position, orientation and length of each segment of the articulated object in the image. The reconstruction scheme described in this paper could be used to recover the posture of an actor based on the estimates for the positions of the limbs in the image provided provided by any of these systems.

Kakadiarsis [8] describes a system for recovering the posture of an actor using the imagery obtained from multiple cameras. Commercially available solutions for tracking human motion such as the OPTOTRAK system also rely on multiple calibrated cameras to recover measurements about the 3D configuration of a figure.

2 Analysis

This section presents an analysis of the problem of recovering the pose of the stick figure model shown in Figure 1b from point correspondences in a single image. The basic reconstruction scheme that will be proposed in this section can easily be extended to recover other articulated objects such as those shown in Figures 7 and 8. In the sequel it is assumed that the relationship between point features in the scene and their correspondents in the image can be modeled as a scaled orthographic projection. Under this assumption, the coordinates of a point in the scene, (X, Y, Z), can be related to the coordinates of its projection in the image, (u, v),through Equation 1. In this equation the parameter s denotes an unknown scale factor.

$$\begin{pmatrix} u \\ v \end{pmatrix} = s \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \qquad (1)$$

One of the first things to realize about this reconstruction problem is that there are more degrees of freedom in the model than there are image measurements. As simple as our stick figure model appears, it still possesses 25 independent degrees of freedom, 4

for each of the limbs, 8 for the torso and the unknown scale factor, s. This can be compared with the 24 independent image measurements that are obtained from the u,v coordinates of the 12 points on the image that correspond to the hands, feet, elbows, knees, shoulders and pelvis of the figure. It is apparent then that there will, in general, be multiple solutions for this reconstruction problem - that is multiple poses that could give rise to the same image measurements. If we were to employ a more sophisticated skeletal model involving more degrees of freedom this problem would only be exacerbated.

One way to approach this reconstruction problem would be to directly invert a complicated function which relates the parameters of the model to the observed image measurements. That is, the locations of the 12 features in the image can be expressed as a function of the joint angles and camera parameters and the goal of the reconstruction process would be to find values of these variables that produce the observed image measurements. This approach has been investigated by several researchers [17, 15]. The difficulty with this tack is that the function in question is highly non-linear which means that sophisticated iterative techniques are often required to find plausible solutions. Moreover, it can be difficult to precisely characterize the set of possible solutions in terms of the set of allowable parameter values.

The method presented in this paper avoids these problems by taking a more direct approach to the reconstruction problem. The method proceeds by accounting for the foreshortening of each of the body segments in the model that is observed in the image and recovering the coordinates of the joints in the world coordinate system directly.

Consider, for example, Figure 2 which shows the projection of a line segment of known length, l, onto the image under scaled orthographic projection.

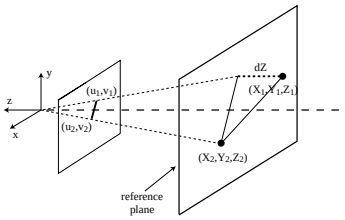


Figure 2: The projection of a line segment onto an image under scaled orthographic projection

In this case the projection of the two end points, (X_1, Y_1, Z_1) and (X_2, Y_2, Z_2), onto the image are represented by (u_1, v_1) and (u_2, v_2) respectively. If the scale factor of the projection model, s were known it would be a simple matter to compute the relative depth of the two endpoints, denoted by dZ in the figure, from the following equations.

$$\begin{aligned} l^2 &= (X_1 - X_2)^2 + (Y_1 - Y_2)^2 + (Z_1 - Z_2)^2 \\ (u_1 - u_2) &= s(X_1 - X_2) \\ (v_1 - v_2) &= s(Y_1 - Y_2) \\ dZ &= (Z_1 - Z_2) \\ \Rightarrow dZ &= \sqrt{l^2 - ((u_1 - u_2)^2 + (v_1 - v_2)^2)/s^2} \end{aligned} \quad (2)$$

In other words, this analysis allows us to compute the 3D configuration of the points in the scene as a function of the scale parameter s. For a given value of s two distinct solutions are still possible which reflect the fact that we can either choose point 1 or point 2 to have the smaller z coordinate. This ambiguity is similar to the ambiguities in positioning the segments of the model described by Lee and Chen [9] and Goncalves Bernardo Ursella and Perona [5].

Another way of viewing Equation 2 is that it places a lower bound on the scale factor s. Since dZ cannot be a complex number the quantity under the square root sign must be non-negative which leads to the following inequality:

$$s \geq \frac{l}{\sqrt{((u_1 - u_2)^2 + (v_1 - v_2)^2)}} \quad (3)$$

This analysis can easily be extended to the case of a jointed mechanism. Consider the simple kinematic chain, shown in Figure 3, consisting of 3 segments where the relative lengths of the segments are known.

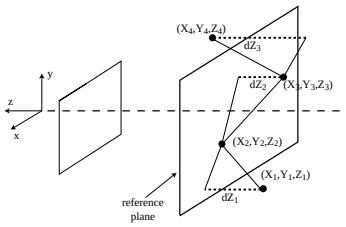


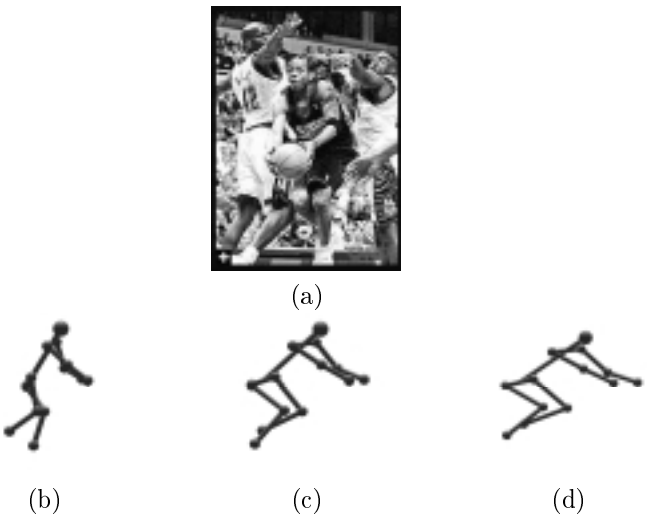
Figure 3: The projection of an articulated object onto an image under scaled orthographic projection

One can, as described previously, compute the relative depth of the endpoints of each segment as a function of the scale parameter s. Once the relative depths are known, it is a simple matter to choose one of the points in the figure as a reference and compute the Z coordinates of all the other points in the model relative

to that feature. Once again there will be two possible solutions associated with each of the segments which means that for every choice of s there are a total of 8 legal configurations that will project to the same image measurements.

Note that we only require information about the relative lengths of the segments as opposed to absolute measurements since the absolute scale of the figure is absorbed by the scale factor s. A lower bound on this scale factor can be obtained by applying inequality 3 to each of the segments in the model to determine the minimum overall scale.

Figure 4 shows how the reconstruction obtained from the point correspondences in Figure 4a varies as a function of the scale parameter chosen, s. The larger the scale factor, the greater the degree of foreshortening associated with each segment in the model and the more distended the solution becomes along the z-axis.



(a) (b) (c) (d)

Figure 4: This figure indicates how the reconstructions computed from the point correspondences obtained from the image in (a) vary as a function of the scale parameter s. The reconstructions shown in (b), (c) and (d) correspond to scale factor values of 2.3569, 2.9461 and 3.5353 respectively.

Alternatively, one can view this as a method for characterizing the entire set of feasible solutions to this reconstruction problem in terms of a single parameter, s. If we were to represent the configuration of the figure in terms of joint angles, for example, it would be much more difficult to characterize the set of solutions in terms of these parameters.

It is important to note that for this model, enforcing the constraints on the relative lengths of the segments effectively enforces all of the kinematic constraints associated with the joints, except for joint angle limits. This means that the set of solutions that are obtained with this method will be identical to those that would be obtained with more complicated parameterizations.

A significant advantage of this approach to reconstruction is that it is very simple to implement requiring only a straightforward sequence of computations. The current implementation consists of a few lines of Matlab code. This is in contrast to other proposed approaches which involve more sophisticated techniques like constraint propagation, nonlinear optimization techniques or genetic algorithms. This reconstruction scheme can be applied to a wide range of articulated objects including hands, robotic manipulators and barnyard animals since there are no limitations on the configuration of the segments in the object.

2.1 Exploiting Additional Constraints

It is important to note that this reconstruction algorithm proceeds on a segment by segment basis and does not make any assumptions about the constraints imposed by the joints. If additional constraints are imposed, such as requiring that 2 segments in the structure be coplanar or that two segments be orthogonal then it may be possible to find a unique solution to the reconstruction problem by searching for a value of s where the resulting reconstruction would satisfy these constraints.

Consider for example the articulated object shown in Figure 7a . Since this object is a closed kinematic chain we can enforce the additional constraint that the first and last points in this chain must be coincident. This constraint can be expressed in the form of an objective function which represents the squared distance between the first and last points in the reconstruction. By plotting this function as a value of the scale parameter, s as shown in Figure 7b, it is possible to find the solution, or solutions which satisfy this constraint. Since this is a search over a single parameter it can be carried out quite efficiently using standard optimization techniques.

Figure 8a shows an articulated object where the 4 points of articulation are coplanar. One can obtain a measure of how coplanar the points in the reconstruction are by computing the sum of squared residuals to the best fit plane through the points. A graph of this residual against the scale parameter, s, is shown in Figure 8b. Once again a unique solution for the structure can be found by locating the minimum of this function. Similar objective functions can be constructed to reflect other constraints such as two segments being perpendicular or 3 points being colinear.

Segment	Relative Length
Forearm	14
Upper Arm	15
Shoulder Girdle	18
Foreleg	20
Thigh	19
Pelvic Girdle	14
Spine	24
Height	70

Table 1: Values used for the relative lengths of the segments in the human figure.

3 Experimental Results

3.1 Results with Real Images

The reconstruction method described in the previous section was applied to a number of images acquired from various sources including scanned photographs and web sites. In each of these images the 12 points which approximated the locations of the endpoints of the limbs namely, the hands, feet, shoulders, elbows, knees and pelvis were located manually. The user also specified for each of the 11 segments in the model which end was nearer to the observer, that is which end had the smaller Z coordinate.

The relative lengths of the segments in the model were obtained from measurements of an "average" human body. Table 1 indicates the values that were used for the various segments. From these measurements, the minimum allowable scale factor was calculated by applying Equation 3 to all of the segments in the model. This value was then used to compute an estimate for the figures pose which is shown. In practice, the policy of choosing the minimum allowable value of the scale parameter as a default usually yields acceptable results since it reflects the fact that one or more of the segments in the model are typically quite close to perpendicular to the viewing direction and are, therefore, not significantly foreshortened. However, the software system allows the user to vary the scale parameter interactively to view the entire range of solutions. The actual process of constructing a 3D model from a photograph is quite straightforward and can be completed in a matter of minutes.

3.2 Comparison with OptoTrak Measurements

In order to determine the accuracy of the proposed method a series of experiments were carried out where the reconstructions obtained with this algorithm were compared with data obtained independently from an OptoTrak positioning system. In these experiments a human actor shown in Figure 6a was outfitted with

a.

b. c.

Figure 6: a) Image of an actor outfitted with OptoTrak markers on his hands, elbows and shoulders. b) Data points returned by OptoTrak sensors c) Results obtained by applying the proposed reconstruction algorithm.

6 OptoTrak markers, one on each hand, elbow and shoulder. The 3D positions of these markers were measured by the Optotrak device during a photo session where 11 pictures were taken. The scale factor, s, used in the reconstruction algorithm was obtained from the ratio of the height of the actor in the image in pixels to his relative height. Figures 6b and 6c show side by side comparisons of the 3D results obtained from the OptoTrak system and the results obtained from the reconstruction algorithm.

After computing the rotation translation and scale factor that best aligned the two data sets in each experiment, the average root mean squared distance between corresponding points was 3.76 cm while the median root mean square distance was 3.31 cm. In order to gauge the accuracy of the method one can compare these errors to the length of the forearm, 29cm, the length of the upper arm, 23cm, and the width of the shoulders, 47cm.

The disparity between the joint angles obtained from the optotrak data and the reconstruction algorithm was also analyzed. These joint angles were computed by considering the angles between the segments in the recovered models. The average disparity between the joint angles was 5.27 degrees and the median disparity was 3.81 degrees.

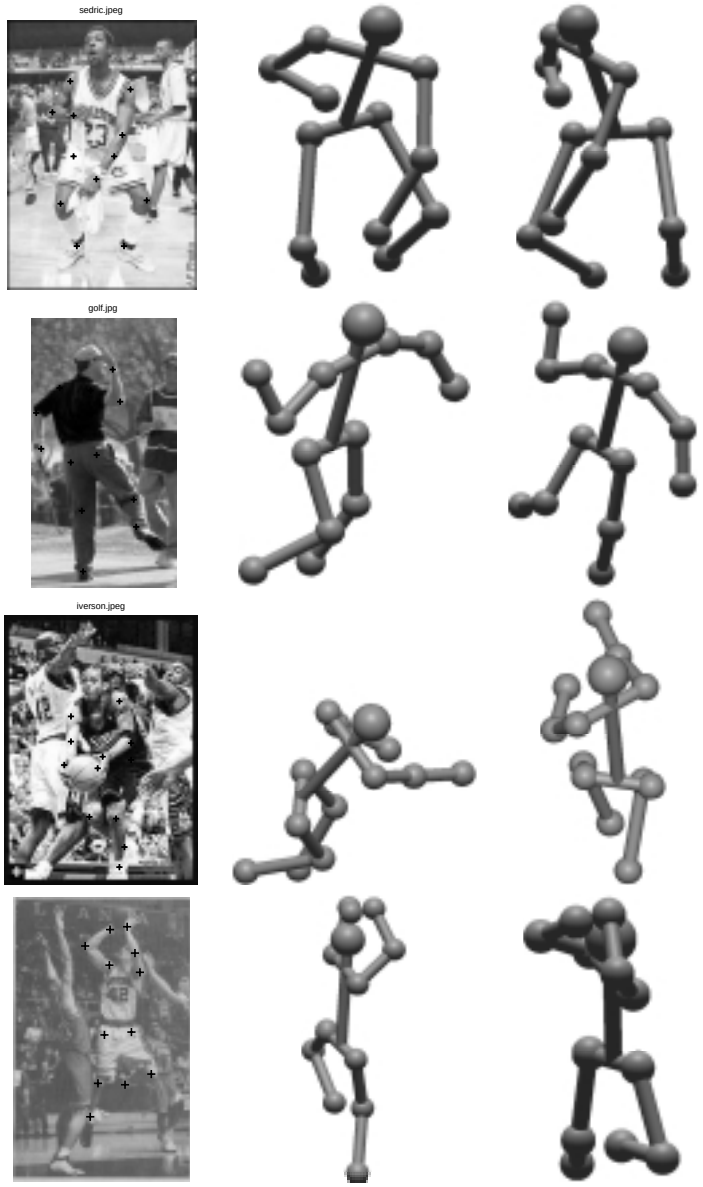
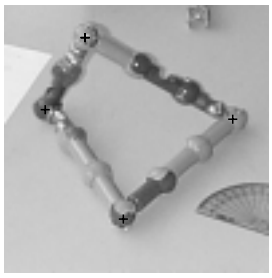


Figure 5: Results obtained by applying the reconstruction algorithm to images obtained from various sources. The first column of figures contains the original images. The crosses on these figures denote the estimated positions of the joints. The second and third columns show the 3D models recovered from these correspondences viewed from novel vantage points.

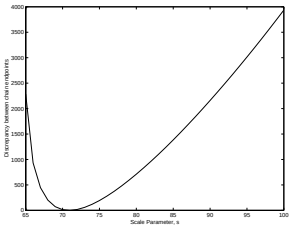
3.3 Exploiting Constraints

Figures 7a and 8a show images of articulated figures where additional constraints can be employed to uniquely determine the scale factor, s, as described in section 2.1. For the object in Figure 7a the closure constraint was used to determine a solution while for the object in Figure 8a the coplanarity condition was employed. Novel views of the reconstructed figures are shown in Figures 7c,d and 8c,d.

The accuracy of these reconstructions was measured by comparing the angles between sections that were computed from the reconstructed figure to those obtained by measuring the angles on the actual figure with a protractor. For the object in figure 7 the discrepancy between the measured and computed joint angles was always less than 1 degree while for Figure 8 the disparity was always less than two degrees



a.



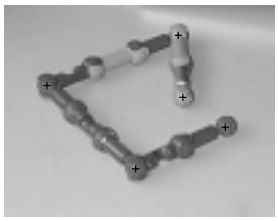
b.

c.

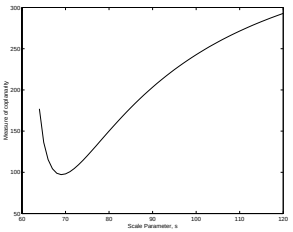
d.

Figure 7: a) Image of a closed kinematic chain b) Graph showing a plot of the disparity between endpoints of the chain in the reconstruction as a function of the scale parameter s. c) and d) Novel views of the reconstructed object

The accuracy of the proposed reconstruction algorithm will depend upon a number of factors including the accuracy with which the projections of the joints can be located in the image, the accuracy of the estimates for the relative length of the segments of the model and the accuracy of the estimate for the unknown scale parameter. These experimental results demonstrate that, on actual image data, the method produces results which are quite accurate and reliable.



a.



b.

c.

d.

Figure 8: a) Image of a kinematic chain where all of the articulation points are coplanar b) Graph showing a plot of the residual to the best fit plane through the articulation points in the reconstruction as a function of the scale parameter s. c) and d) Novel views of the reconstructed object

4 Future Work

The current implementation of the reconstruction system assumes that all of the joints can be located in the image. In practise, this cannot always be accomplished due to occlusions. In these situations it would be useful if the program were able to infer the likely position of occluded joints.

The current implementation also requires the user to specify which end of each segment in the model is closer to the observer. Other systems such as [9] and [17] are able to infer this information from constraints on the legal configuration of limbs in a human figure. It would not be difficult to incorporate a module which would automatically determine which of the 2048 possible solutions satisfy all the relevant joint angle constraints and present these as options to the user.

It would also be interesting to explore the application of these techniques to the problem of analyzing human motion in video imagery obtained with an uncalibrated camera since the solution method is very amenable to real-time implementation.

5 Conclusions

This paper investigates the problem of recovering information about the configuration of an articulated object, such as a human figure, from point correspondences in a single image. Unlike previous approaches, the proposed reconstruction method does not assume that the imagery was acquired with a calibrated cam-

era which means that it can be applied to images that are obtained from sources such as Internet web sites or scanned photographs.

An analysis has been presented which demonstrates that there are a family of solutions to this reconstruction problem parameterized by a single scalar variable, s. A simple and effective algorithm is proposed for recovering the entire set of solutions by considering the foreshortening of the segments of the model in the image. Further analysis indicates that when additional constraints on the structure are available, they can be exploited to obtain a unique solution to the reconstruction problem.

Experimental results on actual image data indicate that the method can be expected to produce accurate results on a variety of objects. As such, this scheme could be a useful tool for animators or others who are interested in non-invasive techniques for measuring the posture of articulated figures.

References

[1] B. Bascle, B. North, and A. Blake. Human articulated motion tracking. In *Workshop on Human Modelling*, July 1998.

[2] Christoph Bregler and Jitendra Malik. Tracking people with twists and exponential maps. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, 1998.

[3] C. Cedras and M. Shah. A survey of motion analysis from moving light displays. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 214–221, 1994.

[4] Margaret M. Fleck, David A. Forsyth, and Chris Bregler. Finding naked people. In *European Conference on Computer Vision*, pages 593–602, April 1996.

[5] Luis Goncalves, Enrico Di Bernardo, Enrico Ursella, and Pietro Perona. Monocular tracking of the human arm in 3d. In *ICCV*, pages 764–770, 1995.

[6] D. Hogg. A program to see a walking person. *Image and Vision Computing*, 5(20), 1983.

[7] Gunnar Johansson. Visual motion perception. *Scientific American*, 232(6):76–90, June 1975.

[8] Ioannis Kakadiarsis. *Motion-Based Part Segmentation, Shape and Motion Estimation of Complex Multi-Part Objects: Application to Human Body Tracking.* PhD thesis, University of Pennsylvania, 1996.

[9] Hsi-Jian Lee and Zen Chen. Determination of human body posture from a single view. *Comp. Vision, Graphics, and Image Proc.*, 30:148–168, 1985.

[10] Daniel D. Morris and James M. Rehg. Singularity analysis for articulated object tracking. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, 1998.

[11] Joseph O'Rourke and Norman I. Badler. Model-based image analysis of human motion using constraint propagation. *IEEE Trans. Pattern Anal. Machine Intell.*, 2(6):522, 1980.

[12] Alex Pentland and Bradley Horowitz. Recovery of nonrigid motion and structure. *IEEE Trans. Pattern Anal. Machine Intell.*, 13(7):730–742, July 1991.

[13] Richard F. Rashid. Towards a system for the interpretation of moving light displays. *IEEE Trans. Pattern Anal. Machine Intell.*, 2(6):574, 1980.

[14] James Regh and Takeo Kanade. Visual tracking of high dof articulated structures: an application to human hand tracking. In *European Conference on Computer Vision*, pages 35–46, 1994.

[15] K. Rohr. Towards model-based recognition of human movements in image sequences. *CVGIP: Image Understanding*, 59(1):94–115, January 1994.

[16] Christopher Wren, Ali Azarbayejani, Trevor Darrell, and Alex Pentland. Pfinder: Real-time tracking of the human body. *IEEE Trans. Pattern Anal. Machine Intell.*, 19(7):780–785, July 1997.

[17] Jianmin Zhao. *Moving Posture Reconstruction From Perspective Projections of Jointed Figure Motion.* PhD thesis, University of Pennsylvania, 1993.