

## Human Figure Reconstruction and Modeling from Single Image or Monocular Video Sequence

Fabio Remondino, Andreas Roditakis

*Institute of Geodesy and Photogrammetry*

*Swiss Federal Institute of Technology - ETH-Hoenggerberg, 8093 Zurich, Switzerland*

*e-mail: <fabio><roditak>@geod.baug.ethz.ch*

### Abstract

*In this contribution we firstly review some approaches to recover the 3D shape of human figures and the related movements. Then we present an improved and reliable version of an easy approach to recover 3D model of humans using just one frame or a monocular video sequence. A simplification of the camera model based on the collinearity condition is required, due to the absence of stereo view. This leads to a perspective or orthographic projection and requires additional information or constraints to solve the problem. The human figure is reconstructed in a skeleton form and visualized with the portable VRML language. Moreover a surface fitting is performed using a pre-defined 3D model obtained from laser scanner measurements.*

### 1. Introduction

In the last years the generation of 3D models of physical objects has become, due to the increasing improvements of computer power, a topic of interest for several researchers. Particular attention has also been paid on the reconstruction of realistic human models, which could be employed in a wide range of applications such as movies, medicine, surveillance, video games, virtual reality environments or ergonomics applications. A complete human model usually consists of the shape and the movement of the body. Some available systems consider the two modeling processes as separate even if they are very close. Most of the research activities in this area is focusing on the problem of tracking a moving human (human motion analysis) through an image sequence, using single camera, multiple views or special equipment for the data acquisition. Many techniques use a probabilistic approach, trying to fit a predefined 3D model to the image data. But less attention has been directed to the deterministic problem, i.e. find a reliable camera model to recover 3D information. Because a single frame or rotating monocular sequences do not allow the generation of 3D data using common techniques (e.g. perspective camera model), some assumptions have to be done in order to infer 3D measurements from 2D observation. In this contribution, after a short review of some techniques used to recover

the shape and the movements of a human, we present a method to find the poses and movements of a person imaged in a single uncalibrated image or moving in short a monocular video sequence digitized from old videotape.

#### 1.1 Related work on static shape reconstruction

Nowadays a classical approach commonly used to recover static human shapes relies on *3D scanners* [e.g. 5, 7, 32] (Figure 1): these sensors are quite expensive but simple to use and various software is available to model the 3D measurements. They work according to different technologies (laser light, structured light, time of flight, etc.) and provide for millions of points, often with related color information. They can scan a human body in less than 20 seconds, usually with multiple scans.



**Figure 1:** Example of body laser scanner. Vitus [32] (left) and Cyberware [7], with related 3D model (right).

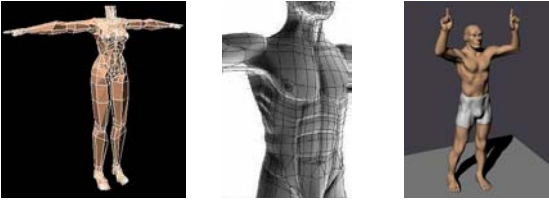
Other approaches used to recover the shape of human figures are *image-based* (Figure 2). They can use a sequence acquired with only one camera and multi-view geometry techniques [24] or a silhouette extraction approach [12, 18]. The former approach generates a 3D point clouds that can be modeled importing the measurements in reverse engineering software; the latter



**Figure 2:** 3D model of static human recovered from a single camera image sequence (left) [24]. The process to recover 3D model using a silhouette extraction approach (right) [12].

technique usually adjust a known 3D reference human model to the extracted image data. Other research activities tried to generate realistic 3D models of humans using only one image [e.g. 3, 17, 27]. They use anthropometric statistics and hypothesis on the human shape to recover pose and body measurements up to a scale parameter.

*Computer animation software* [e.g. 1, 20, 25] can instead produce a realistic 3D model of human subdividing and smoothing polygonal elements, without any measurements (Figure 3). These spline-based systems are mainly used for movies or video-games generation and the created virtual human is then animated using similar animation packages or with motion capture data.



**Figure 3:** Example of generated 3D human model using computer animation software (3Dmax [1], Lightwave [20], SculpLand [25]).

Our work is similar to [27], but additional changes and improvements are presented and discussed (section 3.1). The described method aims to find the poses of a human imaged in a single view or monocular videos with a simple and efficient way. The recovered human skeleton is generated only with a camera model and without pre-defined or superimposed 3D model.

## 1.2 Related work on movements reconstruction

The main problem in body motion tracking and 3D reconstruction is the great number of degrees of freedom to be recovered. A recent overview of human motion analysis is presented in [22]. Existing and reliable commercial systems for capturing human motion typically involve the tracking of human's movements using *sensor-based hardware* [e.g. 2, 23, 30] (Figure 4, left). Other approaches instead rely on *2D monocular videos* of human as primary input [e.g. 13, 26]. They use computer vision techniques, image cues, background segmentation, blob statistics, prior knowledge about human motion, pre-defined articulated body models and often no camera model to recover motions and 3D information. *Multi-cameras approaches* [8, 10, 29] are instead employed to increase reliability, accuracy and avoid problems with self-occlusions. In [8] a photogrammetric approach is used to recover the poses/movements of the imaged human and the recovered 3D data are then modeled with a fitting process (Figure 4, right).



**Figure 4:** Motion capture system (left) [23]. Multi-cameras photogrammetric approach (right) [8].

## 2. The camera model

The assumed camera geometrical properties are those of the pinhole camera model. We consider a world coordinate system XYZ (object system) and a camera coordinate system  $xy$ . The position and orientation of the camera (with camera constant  $c$  and principal point  $x_0, y_0$ ) can be described by a vector  $\mathbf{X}_0$ , with coordinates  $(X_0, Y_0, Z_0)$ , indicating its center of projection and by a  $3 \times 3$  rotation matrix  $R$  indicating the orientation of the camera coordinate frame with respect to the world frame. Usually the object coordinates of a point  $P$  are recovered by means of the collinearity model. The collinearity condition states that a point in object space, its corresponding point in an image and the projective center of the camera lie on a straight line. Additional refinements of the collinearity model may be necessary to deal with geometric distortions and more complicated models for the lens and the corresponding image formation process could be used. If the point  $P$  is imaged in more than one frame with almost wide base line, its 3D coordinates in the world coordinate system, as well as the camera parameters, can be recovered e.g. with the bundle method. Although this method is very accurate, it requires that a point is imaged in at least two images and a good baseline between consecutive frames; therefore it is not really suitable to use it when a single image or (rotating) monocular video sequences are used. A simplification of collinearity equations leads to the perspective projection:

$$\begin{aligned} x &= -c \cdot \frac{X}{Z} \\ y &= -c \cdot \frac{Y}{Z} \end{aligned} \quad (1)$$

that relates the image measurements to the world coordinate system only through the camera constant  $c$ . As we are looking for the object coordinates  $X, Y, Z$  and we want to recover 3D information from a single view, we have the so called 'ill-posed' problem: for each point, we have two equations and three unknown coordinates, plus the camera constant. Therefore the system is underdetermined and some more assumptions need to be introduced.

### 3. The reconstruction algorithm

For man made objects (e.g. buildings), geometric constraints on the object (perpendicularity and orthogonality) and image invariant can be used to solve an ill-posed problem [28]. But in case of free form objects (e.g. the human body) these assumptions are not valid. Therefore equation (1) can be furthermore simplified, describing the relationship between the 3D object coordinates and the 2D image measurements with a scaled orthographic projection:

$$\begin{aligned} x &= s \cdot X \\ y &= s \cdot Y \end{aligned} \quad (2)$$

with a scale factor  $s = -c/Z$ .

The effect of orthographic projection is a simple scaling of the X and Y coordinates. Therefore, all points with the same X and Y coordinates will be projected in the same point onto the image. The scaled-orthographic model amounts to parallel projection, with a scaling added to mimic the effect that the image of an object shrinks with the distance. This camera model can be used if we consider the Z coordinate almost constant in the image or when the range of Z values of the object points (object's depth) is small compared to the distance between the camera and the object. In those cases the scale factor  $c/Z$  will remain almost constant and it is possible to find a value of  $s$  that best fits in equation (2) for all points involved. It also requires that not strong perspective effects are present in the image, even if with additional constraints also some of these images can be analyzed. This model simplifies the reconstruction problem and can produce reliable results for visualization purposes. Moreover it is not necessary to recover the absolute depth of the points with respect to the object coordinate system. Another advantage of this approach is that the camera constant is not required and this makes the algorithm suitable for all applications that deal with uncalibrated images. But, as it is generally an ill-posed problem, we still have an undetermined system, as the scale factor  $s$  cannot be determined only by means of equation (2) and one image.

If we suppose that the length  $L$  of a straight segment between two object points is known, we could use this relative information to solve the problem and reconstruct a model. The distance  $L$  can be expressed as:

$$L_{12}^2 = (X_1 - X_2)^2 + (Y_1 - Y_2)^2 + (Z_1 - Z_2)^2 \quad (3)$$

Then, combining equations (2) and (3), we end up with an expression for the relative depth between two points:

$$(Z_1 - Z_2)^2 = L_{12}^2 - [(x_1 - x_2)^2 + (y_1 - y_2)^2] / s^2 \quad (4)$$

So, if the scale parameter  $s$  is known, we can compute the relative depth between two points as a function of their distance  $L$  and image coordinates. Therefore the whole

reconstruction problem could be reduced to the problem of finding the best scale factor for a particular configuration of image points.

Equation (4) also shows that, for a given scale parameter  $s$ , there are two possible solutions for the relative depth of the endpoints of each segment (because of the square root). This is caused by the fact that even if we select point 1 or point 2 to have the smaller Z coordinate, their (orthographic) projection on the image plane will have exactly the same coordinate. In order to have a real solution, the quantity in the right side of the equation (4) must be greater than zero, which leads to:

$$s \geq \frac{\sqrt{[(x_1 - x_2)^2 + (y_1 - y_2)^2]}}{L_{12}} \quad (5)$$

By applying inequality (5) to each segment with known length one can find the maximum scale parameter than can be used in equation (4) to calculate the relative depth between two segments endpoints. Because of the employed orthographic projection, we have to decide an arbitrary depth for the first point and then compute the second point depth relative to the first one. For the next point we use a previous calculated depth and equation (4) to compute its Z coordinate and so on in a segment-by-segment way. Due to the difference in the left side of equation (4), we have also to decide, for each segment, which one is closer to the camera. Then, knowing the scale factor, equation (2) can be used to calculate the X and Y coordinates of the measured image points.

The algorithm (Figure 5) can be applied to a wide range of "articulated" objects since there are no limitations on the configurations of the objects and only relative distances are required.

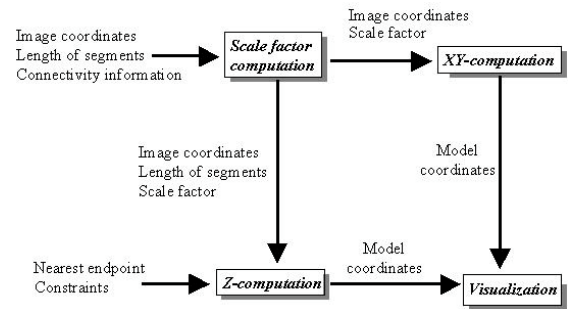


Figure 5: The data flow of the algorithm.

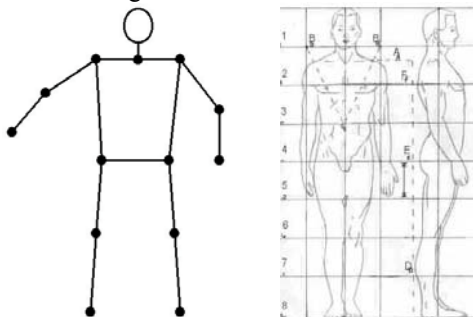
#### 3.1 Constraints and occlusions

The presented reconstruction algorithm proceeds on a segment-by-segment basis. As also mentioned in [27], the algorithm can not always model images that present strong perspective effects, as we are using an orthographic approach. In fact, in some results, because of measurement or assumption errors, a segment can get foreshortened or warped along one axis (Figure 7-A, first

model). But if a segment is almost parallel to the image plane or can be assumed to lie on a plane, then its two end points can be treated as being at the same depth. Therefore, imposing additional constraints, such as requiring that two points must have the same depth, this 'mistake' can be avoided and the resulting 3D model is more accurate and reliable (Figure 7-A, second model). Other constraints could be the perpendicularity of two segments or a closure constraint, imposing that the two points must coincide (e.g. Figure 7-B). Furthermore, measurements occlusions are handled selecting the most adequate point in the image and computing its 3D coordinate using a depth constraint (e.g. right knee in Figure 8-A or right shoulder in Figure 8-D).

### 3.2 The human body and its representation

The human skeleton system is treated as a series of jointed links (segments), which can be modeled as a rigid body. For the specific problem of recovering the poses of a human figure, we describe the body as a stick model consisting of a set of thirteen joints (plus the head) connected by thirteen segments (we consider the shoulder girdle as unique segment), as shown in Figure 6, left. The head is used to model some figure where it is inclined, as shown in Figure 7-B/C.



**Figure 6:** The skeleton with 13 joints and the head used for the reconstruction (left). The human body represented as an average of eight heads high (right).

The algorithm needs the knowledge of the relative lengths of the segments as opposed to absolute measurements (since the absolute scale of the figure is absorbed by the scale factor  $s$ ). The lengths of the segments can be obtained from anthropometric data (motion capture databases or literature). Two sets of lengths are used in our tests, leading almost to the same results. The first set of relative distances between human joints is derived from a motion capture database [4] (Table 1, central). The second set is more general and follows the studies performed by Leonardo Da Vinci and Michelangelo on the human figure [14, 31] (Table 1, right). It represents the human figure as an average of eight heads high (Figure 6, right). We added a coefficient  $i$  to these lengths to model the variation of the human size from the average

( $i=1$ ). Once the program has computed the 3D coordinates of the human joints, they are given to a procedure that uses VRML language to visualize the recovered model. All the joints are represented with spheres and they are joined together with cylinders or tapered ellipsoids.

**Table 1:** Two different sets of relative lengths of the segments used in the computation of the human model. MC = Motion Capture. L = Literature.

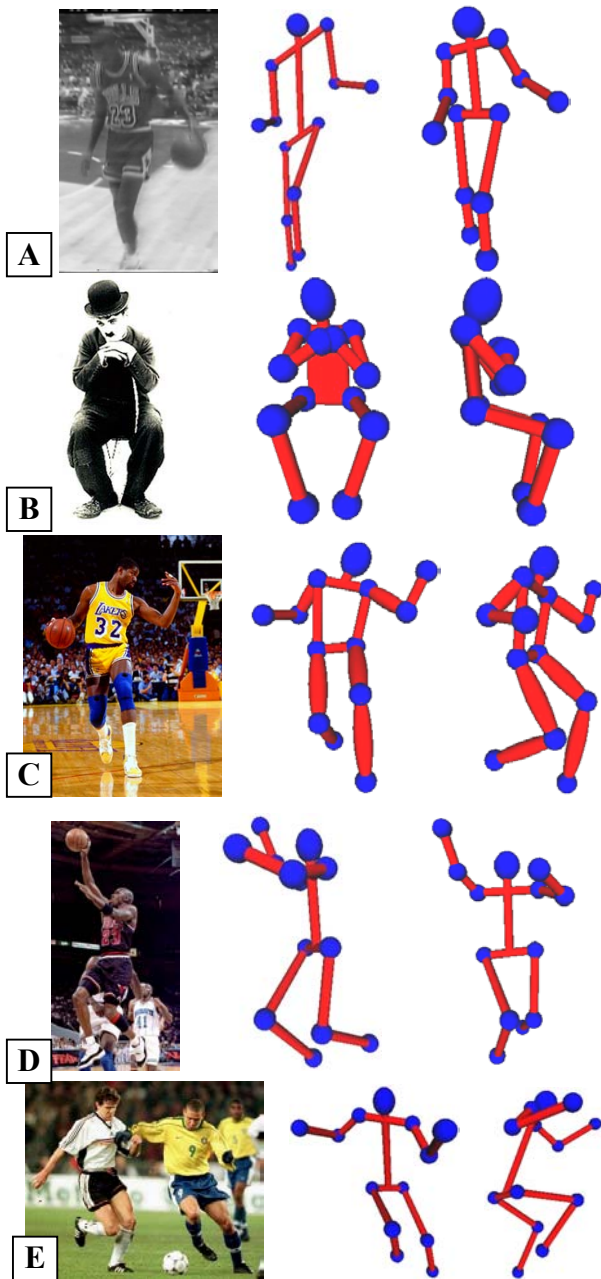
Segment	Relative Length (MC) [cm]	Relative Length (L) [unit]
Height	175	8 $i$
Lower arm	35	2 $i$
Upper Arm	25	1 ½ $i$
Neck-Head	25	1 ¼ $i$
Shoulder Girdle	44	2 $i$
Torso	53	2 ½ $i$
Pelvic Girdle	30	1 ½ $i$
Upper leg	46	2 $i$
Lower leg	52	2 $i$
Foot	22	1 $i$

### 4. Results from single images

In order to determine the accuracy, limitations and advantages of the method, a series of experiments are performed first on single images found on the Internet. Figure 7 shows some test images, where we measured the 13 joints and the center of the head (to model possible inclinations), as well as the recovered 3D skeleton looked from different viewpoints. The stretching or warping of the recovered models is strictly related to the assumption on the scale factor and on the used constraints. Figure 7-A shows a 3D model obtain with the simple orthographic projection (central image: it looks warped) and after the applied constraints (right model). We see how a depth constraint allow the generation of a 3D skeleton more realistic and accurate. The visualization program can represent the torso of the reconstructed human figure with a scaled cylinder or with a trapezoidal shape either with a unique segment between shoulder and pelvic girdle. While we can model the muscles and shape of upper arms and legs using tapered ellipsoids or simple cylinders.

### 5. Explicit/Implicit surface fitting

Fitting a surface on a given set of points is a problem that has been approached in several ways in Computer Graphics literature. The main classification divides the methods that fit a surface onto a set of points into explicit and implicit, depending on the final mathematical representation of this surface. *Triangular meshes*, *volume grids* [6] and *parametric piecewise functions (NURBS)* [15] are explicit descriptions of the surface, while *soft* or *blobby objects*, also known as *metaballs* [8, 9] describe

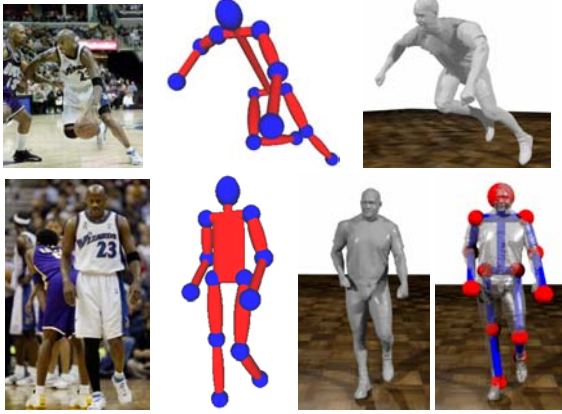


**Figure 7:** Results of the human reconstruction from single uncalibrated images. A: Compared results between simple orthographic model and after the applied constraints. B: The body represented with a scaled cylindrical torso (also note the inclined head). C: The torso represented with a trapezium and the arms/legs with tapered ellipsoids. D/E: Cylindrical representation for all the reconstructed segments.

the surface as the isosurface of a distance function. On one hand the explicit functions appear to be a popular representation in modeling software and are hardware supported, achieving rendering performances of millions of texture mapped polygons per second. Fitting such

surfaces on a set of measurements presents, though, the difficulty of finding the faces that are closest to a 3D point, and the disadvantage of non-differentiability of the distance function [16]. Implicit surfaces are more suitable for modeling soft objects as they have been used in modeling clouds [9] or soft tissue objects [16], but present difficulties in deformations and rendering.

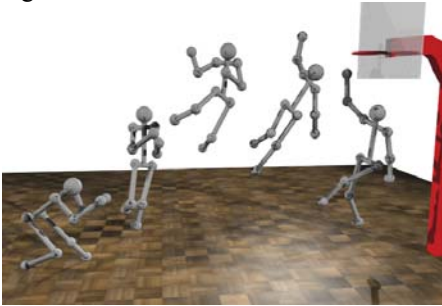
In our case, to improve the quality and realism of the recovered skeleton, we decided to fit a pre-defined model to our 3D data. We chose to use a polygonal mesh of a human body (obtained from laser data [7]) for the fitting process with the measured skeleton because of the simplicity of the deformation procedure. Using the modeling and animation features of Maya [21] animation software, we were able to fit the polygonal mesh to the skeleton in a semi-automatic manner and face the problems that arise from the discontinuity of the distance function by adjusting the influence weight of skeleton joints on the polygons. The initial 13 joints were used for the construction of the skeleton and few extra joints were inserted to avoid long ‘bones’ between them. The skeleton has a hierarchical structure with a root joint controlling the general position and orientation while the children joints adjust the rotation and translation of the body members to achieve a pose. For the movement of the skeleton there are two solutions available in Maya, called *Forward* and *Inverse Kinematics*. The first method requires the rotation and translation of all the joints, starting from the parent and ending to the last child joint, to achieve the final pose. The latter method requires that only the position and rotation of the desired pose, or *target locator*, is given from the user, and then the position of the intermediate joints is calculated automatically. In this case, the use of joint rotation constrains is essential in order to achieve a correct solution. In the present paper, we use inverse kinematics, because of the simplicity and automation of the procedure. The last step is binding the polygonal model with the skeleton. The software uses the joint location and skeleton hierarchy to decide which parts of the model ‘skin’ are affected. This procedure is called *skinning* [19] and for soft bodies, the influence of the joints along the border between two different segments is feathered with the neighboring ones to simulate the behavior of soft skin tissue in deformations. We had to specify how deep in the hierarchy the program had to search from the current joint, to calculate its influence to the polygonal skin. As we set a larger depth, the influence of one joint is diffused over a larger region, giving a softer deformation effect on the skin. In Figures 8 and 10 we present the results of fitting a polygonal model acquired with a Cyberware body scanner [7] to the poses recovered from single images and monocular sequences. In the second row of Figure 8 we also show how our recovered 3D model exactly fit the scaled laser model.



**Figure 8:** Results of fitting the polygonal skin model to the 3D skeleton extracted from single images.

## 6. Application to Video Sequences

The presented algorithm has been also extended to solve the reconstruction problem in many consecutive frames obtained from short video sequences. Therefore 3D poses and movements of a figure over a short interval of time are recovered (Figure 9) and then, by playing the reconstructed poses at an adequate frame rate, we can get a very realistic and observer independent reconstruction of the original motion.



**Figure 9:** All the recovered 3D models of the sequence of Figure 11 displayed together on a basketball field.

However, because the algorithm finds only relative distance in each frame, subsequent poses normally have no object coordinates in common. This has been solved assuming that a joint in one frame has the same coordinates as the same joint in the previous frame (plus a small translation vector). In the example presented in Figure 10, the image points were measured semi-automatically with a Least Squares Matching Algorithm [11] between consecutive frames. Due to the low quality of the images (digitized from an old videotape) and the small scale of the figure, it was not always possible to measure accurately the position of the joints in all the frames. Moreover, because of the fast movements of the player, a tracking algorithm could not achieve correct results. In case of good image quality, the corresponding

joints could be tracked automatically over the video sequence, as proposed in [8]. Nevertheless the recovered models show the reliability of the algorithm and its possible application in case of small perspective effect. In the last row of Figure 11 we also show the results of the fitting process between our recovered 3D skeleton and the laser scanner human body.

## 7. Extension to perspective projection

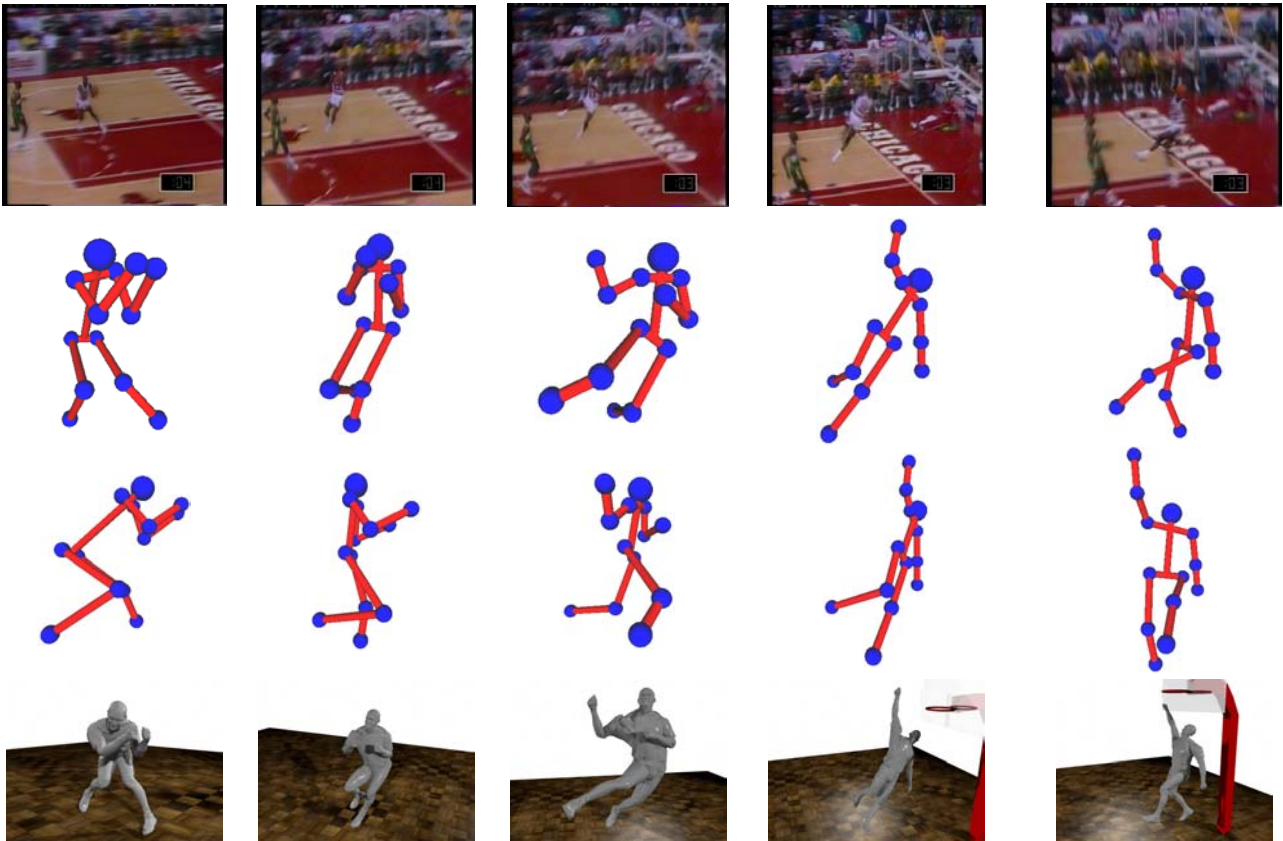
If the focal length of the camera is known, the algorithm could be modified to handle a more realistic projection model like the perspective camera model. Therefore, instead of using equation (2), we could apply equation (1) to describe the relationship between image and object coordinates. The problem with this approach is that now we need to determine the Z coordinate of the object points relative to the camera coordinate system and not, like in the original algorithm, their relative depth. Substituting equation (1) into (3) results in a quadratic equation with two unknowns,  $Z_1$  and  $Z_2$ :

$$L_{12}^2 = \left(\frac{x_1 Z_1 - x_2 Z_2}{c}\right)^2 + \left(\frac{y_1 Z_1 - y_2 Z_2}{c}\right)^2 + (Z_1 - Z_2)^2 \quad (6)$$

If a segment is supposed to be almost parallel to the image plane, then the Z coordinates of the two endpoints can be assumed as being the same and the equation (6) can be solved for one of the Z values:

$$Z_1 = \frac{cL_{12}}{\sqrt{(x_1^2 - 2x_1x_2 + x_2^2 + y_1^2 - 2y_1y_2 + y_2^2)}} \quad (7)$$

Then, we can use the information about one endpoint to compute the depth of the next one in a segment and then apply equation (3) to determine their X and Y coordinates. On the other hand, if we can not assume the same depth for two endpoints, we have to assume for at least one point in the figure its Z coordinate and then reconstruct the other points based on this value. Another possibility is to express the whole system in dependency of one variable, say  $Z_1$ , and then search for a value of  $Z_1$  that results in the best posture of the figure. Although this approach should provide better results than the orthographic one, it requires the knowledge of the camera constant and of some lengths/depths of the object. It is not always possible from uncalibrated images or videos to recover the camera constant. But sometimes we can compute an approximation for it from imaged objects that have known dimensions, like field markers, buildings or smaller objects (DLT approach) or using straight and perpendicular lines (Vanishing Point approach) [24]. If this information is available, then a more sophisticated projection model than the orthographic projection can be applied.

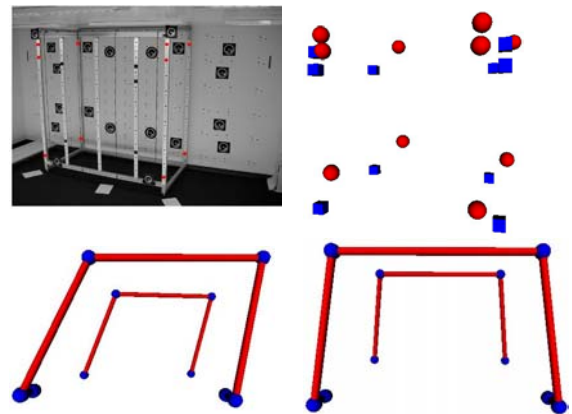


**Figure 10:** Results obtained from a short video sequence. The second and third rows show the reconstructed model viewed from two different viewpoints. The last row shows the results of the fitting process described in section 5.

A comparison between the two presented camera models is shown in Figure 11. An articulated object is imaged from one side and the 3D coordinates of some points are recovered. The camera constant is obtained as described in [24], with a vanishing point approach. In the bottom line of the figure the two reconstructed models are presented: on the left the result of the orthographic projection (plus depth constraints) shows how the object results warped. On the right the 3D model obtained with the perspective projection shows the improvement of the results, even if the perpendicularity of some segments is still not maintained.

## 8. Conclusion

In this work we presented the problem of recovering 3D models of humans from single images or monocular video sequences. The problem was solved with a scaled orthographic projection with some additional constraints inserted to control the ill-posed problem. Few experiments with a perspective camera model have also been done and presented.



**Figure 11:** A test image with perspective effect, with some points measured for the reconstruction (upper left). The points reconstructed with orthographic projection (blue cubes) and perspective model (red spheres) (upper right). The reconstructed 3D points connected with cylinders in case of orthographic model (lower left) and perspective (lower right).

The reconstruction and visualization algorithm was tested with several images and the presented results

show the reliability of our extended algorithm, also when some perspective effects are present. The fitting process with the laser data was performed to increase the visual quality of the results. And the accuracy of the reconstruction is demonstrated by the fitting results (Figure 8): indeed the laser model is just scaled and its segments are rotated to match our skeleton.

As future work we want to add the foot and the hands to the recovered skeleton and we will try to model muscles and shape of some segments using tapered cones. Moreover the perspective camera model will also be tested on human figures imaged in monocular sequences.

## Acknowledgments

The authors would like to thank Sidlei Nascimento da Silva for his useful contributions in the reconstruction algorithm.

## References

1. 3D Studio Max: <http://www.3dmax.com> [June 2003]
2. Ascension: <http://www.ascension-tech.com/> [June 2003]
3. Barron, C., Kakadiaris, A.: Estimating Anthropometry and Pose from a single uncalibrated image. *Computer Vision and Image Understanding*, Vol. 81, 2001.
4. Biovision: <http://www.biovision.com> [June 2003]
5. BodySkanner: <http://www.scansuccess.com> [June 2003]
6. Curless, B., Levoy, M.: A volumetric method for building complex models from range images, Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, ACM Press, pp. 302 - 312, 1996.
7. Cyberware: <http://www.cyberware.com> [June 2003]
8. D'Apuzzo, N., Plankers, R., Fua, P., Gruen, A., Thalmann, D.: Modeling human bodies from video sequences. Videometrics VI, SPIE Proceedings, Vol. 3461, San Jose', USA, pp. 36-47, 1999.
9. Dobashi, Y., Kaneda, K., et. al.: A simple, efficient method for realistic animation of clouds, Proceedings of the 27th annual on Computer graphics and interactive techniques, ACM Press, pp. 19-28, 2000.
10. Gavril, D.M.: The Visual Analysis of Human Movement: A Survey. *Computer Vision and Image Understanding*, Vol. 73, No 1, pp. 82-98, 1999.
11. Gruen, A.: Adaptive least squares correlation: a powerful image matching technique. *South African Journal of Photogrammetry, Remote Sensing and Cartography*, 14(3), pp.175-187, 1985.
12. Hilton, A., Beresfors, D., Gentils, T., Smith, R., Sun, W., Illingworth, J.: Whole-body modeling of people from multiview images to populate virtual worlds. *The Visual Computer*, Vol. 16, pp. 411-436, Springer-Verlag, 2000.
13. Howe, N., Leventon, M., Freeman, W.: Bayesian reconstruction of 3D human motion from single-camera video. *Advances in Neural Information Processing System*, Vol. 12, pp. 820-826, MIT Press, 2000.
14. Human Figure Drawing Proportion: [www.mauigateway.com/~donjusko/human.htm](http://www.mauigateway.com/~donjusko/human.htm) [June 2003]
15. Krishnamurthy, V., Levoy, M.: Fitting smooth surfaces to dense polygon meshes, Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, ACM Press, New York, USA, pp. 313 - 324, 1996.
16. Ilic, S., Fua, P.: From explicit to implicit surfaces for visualization, animation and modeling, Proceedings of the international workshop on visualization and animation of reality based 3D models, on CD-Rom, Vulpera, Switzerland, 2003.
17. Lee, H.J., Chen, Z.: Determination of human body posture from a single view. *Computer Vision, Graphics, Image Process*, Vol. 30, pp. 148-168, 1985.
18. Lee, W., Gu J., Thalmann N: Generating animatable 3D virtual humans from photographs. *Eurographics*, Vol. 19, No. 3, 2000.
19. Learning Maya 2.5: Alias Wavefront, 1998.
20. Lightwave: <http://www.lightwave3d.com> [June 2003]
21. Maya: <http://www.aliaswavefront.com> [June 2003]
22. Moeslund, Granum: A Survey of Computer Vision-Based Human Motion Capture. *Computer Vision and Image Understanding*, Vol. 81(3), 2001.
23. Motion Analysis: <http://www.motionanalysis.com/> [June 2003]
24. Remondino, F: 3D reconstruction of static human body with a digital camera. Videometrics VII, SPIE Proc., Vol. 5013, pp. 38-45, 2003.
25. SculpLand: [www.sanyonet.ne.jp/~nakajima/SculpLand.html](http://www.sanyonet.ne.jp/~nakajima/SculpLand.html)
26. Sidenbladh, H., Black, M., Fleet, D.: Stochastic Tracking of 3D Human Figures Using 2D Image Motion. European Conference on Computer Vision, D. Vernon (Ed.), Springer Verlag, LNCS 1843, Dublin, Ireland, pp. 702-718, June 2000.
27. Taylor, C.T: Reconstruction of Articulated Objects from Point Correspondences in a Single Uncalibrated Image. *Computer Vision and Image Understanding*. Vol. 80, pp. 349-363.
28. Van den Heuvel, F.A.: 3D reconstruction from a single image using geometric constraints. *ISPRS Journal for Photogrammetry and Remote Sensing*, 53(6), pp. 354-368, 1998.
29. Vedula, S., Baker, S., et al: Three Dimensional Scene Flow. ICCV '99, Vol. 2, pp. 722-729.
30. Vicon, <http://www.vicon.com> [June 2003]
31. Visual Body Proportion: [www2.evansville.edu/drawinglab/body.html](http://www2.evansville.edu/drawinglab/body.html) [June 2003]
32. Vitus: <http://www.vitus.de> [June 2003]