A Camera Based Interactive System for Human Body Modeling

S. Gopalsamy*
National Centre for Software Technology
Gulmohar cross road No. 9
Juhu, Mumbai - 400049, India.

Jin Gu† and Helen C. Shen‡
Department of Computer Science
Hong Kong University of Science and Technology
Clear water bay, Kowloon, Hong Kong.

Abstract

An interactive system is presented for three dimensional reconstruction of human models from multiple images taken from a human body image capturing setup. The reconstruction is based on two computer vision techniques – volumetric intersection by cross sections approach and stereo. The basic theory behind these techniques is outlined. The main components of the system – image capturing, image processing, camera calibration, 3D reconstruction and 3D surface representation – are explained and various user interaction tools are mentioned. In the system, different parts of a human body will be reconstructed separately. A stack of parallel cross sections, each of which is represented by a NURBS curve, will be computed for each part of the body by volumetric intersection. The concave parts of the body like chest of a female may not be captured by the cross sections. In those cases, the cross sections are modified by capturing the concave parts by stereo using structured light grid points. Then the cross sections are fine tuned by user interaction, wherever necessary, for smoothness and accuracy.

1 Introduction

Three dimensional reconstruction of real world objects in general and of human bodies in particular is an active area of research in the field of computer vision with many applications such as mannequin modeling for garment design in textile industry, computer games, and movie animation in entertainment field. There are two main categories of techniques for 3D data acquisition and reconstruction according to the sensing mechanism. The first category is active sensing technique like laser scanning [1], ultrasonic sensor and infrared range-finder where 3D information is recovered by emitting certain types of energy on the body. The second is passive sensing techniques where no extra energy is emitted on the body during the sensing process. The most typical example of this category is camera which captures the intensity images of the observed objects. The 3D model is reconstructed from the 2D images by computer vision techniques like stereopsis, shape-from-X (X can be shading, occluding contours, surface contours, etc.) and volumetric intersection.

Although active sensing techniques can usually provide high resolution and precision in the captured 3D data of the objects, it is more expensive and needs special operating setting. Whereas passive sensing is cheaper and can be used in less restricted environment, and also one can obtain acceptably accurate reconstruction in many cases. Moreover, using camera as passive sensor provides visual information like color and texture of the imaged surface. Combining both the reconstructed geometric information and the visual information, one can achieve a more complete model of the object.

In this paper we deal with camera based passive sensing techniques for the reconstruction of real human bodies. We combine two techniques – volumetric intersection and stereopsis for the reconstruction from images. For volumetric intersection, we use 2D contours of the images and reconstruct the object as a collection of cross sections [3, 8]. The cross sections may not capture some of the concave parts of the body like the Chest part of a female and the Hip. To overcome this, we use stereopsis to compute a concave part using structured light grid points on two images. The concave part is obtained as a 3D grid by stereopsis and then the cross sections of that part computed by volumetric intersection are modified by projecting them on to the grid. The cross sections are then fine tuned for smoothness and accuracy by interactive editing.

The rest of the paper is arranged as follows: Section 2 covers some of the theoretical aspects of the underlying techniques. Section 3 gives an outline of the reconstruction system. Section 4 deals with 2D and 3D interactive tools of the system. In section 5

*Formerly visiting scholar at Department of Computer Science, Hong Kong University of Science and Technology
†Research associate
‡Associate professor

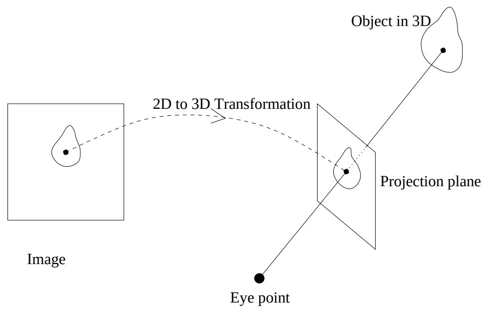

Figure 1: Camera calibration parameters

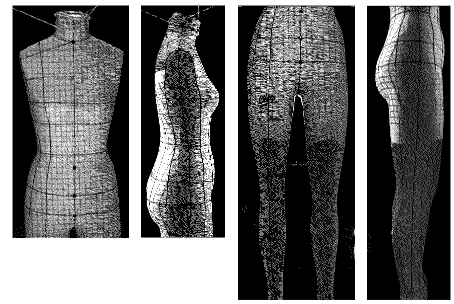

Figure 2: Images of a mannequin from four cameras

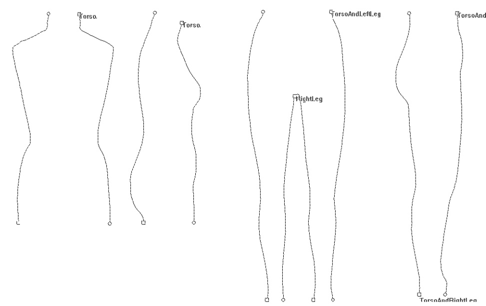

Figure 3: The four image contours and their segmentation

we present the results of our experimental reconstruction of a female mannequin and in section 6 we give some conclusions .

2 Basic theory

Our system is based on two computer vision techniques – volumetric intersection by cross sections and stereopsis – for 3D reconstruction from images obtained from calibrated CCD cameras. In this section we briefly describe some of the the theoretical aspects of these techniques. We will start with the definition of camera calibration which is needed for both the techniques.

Camera calibration: Calibration of the cameras used to capture images of the body means the following:

1. There is a 3D world coordinate system common for all the cameras.

2. For each camera, there is a *projection reference point* or *eye point* E and a *projection plane* Φ defined in world coordinates. The projection reference point will also be referred as *camera position.*

3. For each camera, there is a $2D \rightarrow 3D$ transformation T from image coordinates to world coordinates mapping the image into the projection plane satisfying the following condition: *If X is a point on the image corresponding to a point P on the object in 3D then $T(X)$ will be the intersection with the projection plane of the ray starting from the projection reference point E and passing through P.* See Figure 1.

2.1 Cross sections approach for volumetric intersection

In [3] the cross sections approach for volumetric intersection has been presented for 3D reconstruction of smooth single part objects. A smooth single part object has been defined as one which can be represented as a stack of parallel cross sections where each cross section is a smooth curve with only one connected component. A method has been given for the computation of required number of cross sections. Here, we extend that method to reconstruct the surface of a human body which can be considered as a multipart object consisting of various parts of the body – head, two hands, chest, torso, hip, and two legs. We can assume that each of these parts is a smooth single part object. Each part will be reconstructed independently. Some of the concepts of this method which are needed to understand our reconstruction system are explained below.

2.1.1 Image contours and segmentation:

The images of the human body taken by the CCD cameras are processed by standard image processing techniques to obtain the *image contours* which form the boundary of the object in the image. The image contours are then decomposed in to a number of continuous segments where each segment is either monotonically increasing or decreasing along y-axis in the image plane. The segments are named and associated to various parts of the body. See Figures 2, 3.

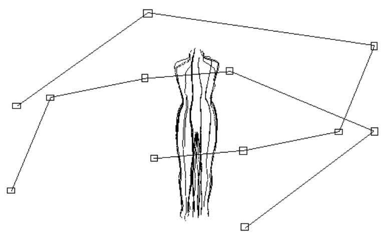

Figure 4: Projection plane contours and camera positions in 3D

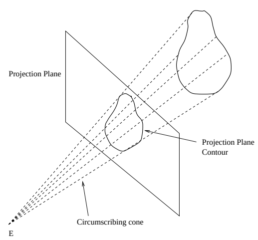

Figure 5: Circumscribing cone

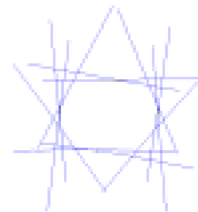

Figure 6: Cross section plane contours

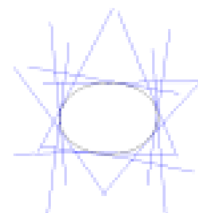

Figure 7: Cross section plane contours and a fitted cross section curve

2.1.2 Projection plane contours in 3D and Circumscribing cone

For each camera/ view, its image contours are transformed on to its projection plane by the image coordinates to world coordinates transformation matrix to obtain projection plane contours (Figure 4).

The cone formed by the rays starting from the projection reference point of the camera and passing through the points of the projection plane contours is known as the *circumscribing cone* [5] corresponding to that camera. The object will lie inside the intersection of all the circumscribing cones corresponding to different cameras. This is the basis for all volumetric intersection methods [6, 7, 2, 8].

2.1.3 Cross section plane contours

These are the contours obtained by the intersection of a cross section plane with the circumscribing cones. For any cross section plane, these are actually the projections of the projection plane contours of different cameras on to the cross section plane with respect to the camera positions. (Figure 6). Note that, in general, there will be two contours corresponding to each camera. The body cross section obtained by intersecting the body with the cross section plane will lie inside the region bounded by all the cross section plane contours. This is the basis for cross sections approach [8, 3].

2.1.4 Cross section curves

Since we are interested only in the bounding surface of the body, we will consider only the bounding curves of the body cross sections which are called the cross section curves. For each part, a cross section curve will be a smooth closed curve with one connected component and it will lie inside the region bounded by all the cross section plane contours corresponding to that part Figure 7. Moreover the cross section curve will touch the two cross section plane contours corresponding to each camera. This is the basis of our algorithm for the computation of cross section curves.

2.1.5 Computation of cross section curves

In our system, we compute each of the cross section curves as a NURBS curve following the method given in [3]. An outline of the method is given below:

1. Consider one part of the body and take a cross section plane intersecting that part.
2. For each camera compute the two cross section plane contours corresponding to the part.
3. *Initial cross section curve*: To compute an initial curve, we can use only two cameras/views, like front and a side view, whose directions of projections are roughly perpendicular. The region bounded by the four cross section plane contours

of these two views will be a rough approximation of the body (part) section by the cross section plane. An approximate bounding circle of this region can be obtained by computing the four intersection points of these contours. This circle is converted to a NURBS curve and that curve is taken to be the initial cross section curve.

4. *Curve fitting:* For each camera/view, the curve is modified such that it touches the two cross section plane contours of that camera. This is done iteratively, taking one camera at a time. Note that the modification done for one camera may be affected when the curve is modified for another camera. Hence the iteration is repeated a few times and a satisfactory cross section curve is obtained.

2.2 Stereopsis

In stereopsis, the 3D coordinates of a point P seen by two cameras will be computed from its pixels values in the images from the two cameras using the camera calibration parameters as follows:

> For $i = 1, 2$, let (X_i, Y_i) be the pixel values of the point P in the image of the ith camera and let (x_i, y_i, z_i) be the point on the projection plane of that camera obtained by transforming (X_i, Y_i) by the image coordinates to world coordinates transformation matrix corresponding to that camera.
> Let (Ex_i, Ey_i, Ez_i) be the camera position (eye point) and consider the ray from (Ex_i, Ey_i, Ez_i) passing through (x_i, y_i, z_i) for $i = 1, 2$.
> Then the point of intersection of the two rays will be the 3D position of P.

In our system we use stereopsis to compute 3D coordinates of two sets of points, the anchor points and structured light grid points.

Anchor points:

These are some key points of the body like belly button, front neck, back neck, shoulder point, crotch, knee and ankle points. At the time of image capturing, these anchor points will be marked on the body so that they can be easily identified in the images. At the time of reconstruction, these points will be used to define the body parts and to align the coordinate system with a body-centric coordinate system.

Structured light grid points:

These points will be used to reconstruct non-convex parts of the body like chest part of a female and hip.

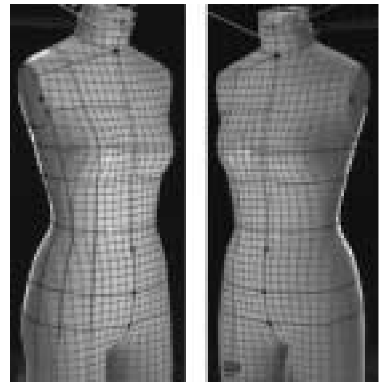

Figure 8: Structured light grid points

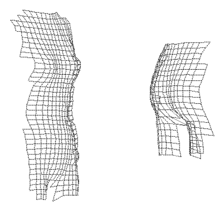

Figure 9: Two grids in the front and back of the mannequin obtained by stereopsis using structured light grid points.

At the time of image capturing, these points are displayed on the body by projecting structured light using slide projectors. See Figure 8. The cameras will be arranged in such a way that the structured light grid points will be seen by at least two cameras.

Correspondence for the structured light grid points in two images is done as follows: Each point corresponds to one node of the structured grid light. The grid is rectangular and each node is identified by a 2D index (i, j) with reference to an origin node. The origin node is chosen in such a way that it is easily identified in both the images. The grid point in an image corresponding to a node is assigned the index of that node. This can be easily done manually since the grid point corresponding to the origin is known. Hence the grid points in two images corresponding to one node will be assigned the same index which establishes the correspondence.

Once the correspondence is established, the 3D coordinates of the structured light grid points which are seen in two images are computed by stereopsis. Then adjacent points are joined to obtain a quadrilateral grid lying on the surface. See Figure 9. This grid will be used to modify the cross sections of the corresponding non-convex part.

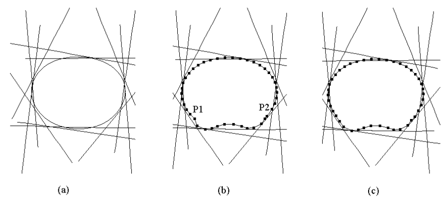

Figure 10: Modification of a cross section to fit a grid. (a) The cross section curve before modification; (b) the curve after projecting a portion of it on to the grid; and (c) the curve after a minor interactive editing.

2.3 Modification of a cross section to fit the grid

Let $\mathcal{S}$ be a cross section curve of a part corresponding to one cross section plane Ψ. Let Σ be a grid associated to that part obtained from structured light grid points. Then $\mathcal{S}$ can be modified to fit the grid as follows:

- *Step 1:* Intersect the grid Σ with the cross section plane and let L be the resulting intersection curve.
- *Step 2:* Let P_1 and P_2 be the points on $\mathcal{S}$ which are nearest to the end points of L.
- *Step 3:* Now we have to project the curve segment of $\mathcal{S}$ between P_1 and P_2 on to L. For this, we discretize the whole curve $\mathcal{S}$ converting it in to a piecewise linear curve and then move all the points in between P_1 and P_2 to their nearest points of L. This modified piecewise linear curve will be taken as the new cross section curve and it will fit the grid.

Note that the cross section curve may not be smooth after modification as above because the grid on the surface obtained from the structured light grid points may not be smooth and accurate. So we may have to do minor adjustments interactively to make it better. See Figure 10.

3 Human body reconstruction system

Our human body reconstruction system consists of the following five components:

1. Image capturing
2. Image processing
3. Camera calibration
4. 3D reconstruction
5. 3D surface representation

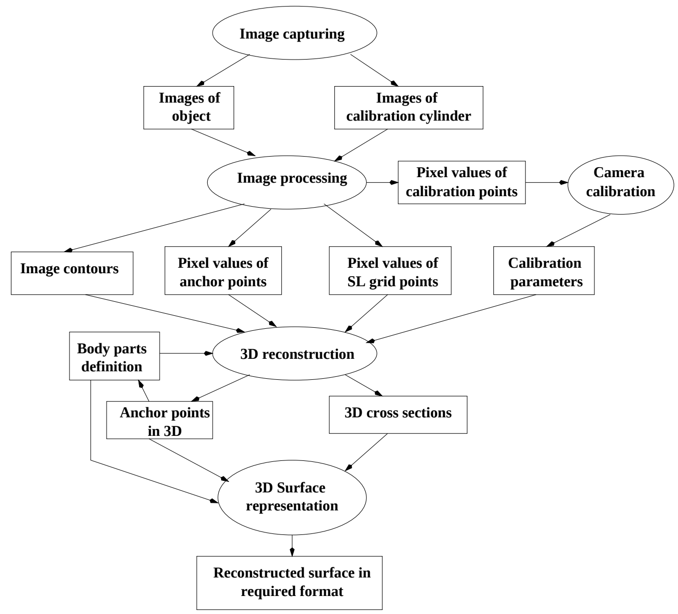

Figure 11: Human body reconstruction system – the five system components and the input/output to/from the components.

Below we give brief descriptions of these components. Figure 11 shows the input and output structures for these system components.

3.1 Image capturing

Image capturing setup: For the purpose of image capturing of real humans or mannequins, we use a *human body imaging setup* [4] which consists of a supporting framework of dimension 3 meters height and 1.5 meters radius. Eight CCD cameras are fixed on vertical frames in two horizontal levels to capture the images of the top and bottom parts from four views – front, left, back and right. Two levels of cameras are used in order to minimize the space occupied by the image capturing set up. Apart from these, four more cameras are fitted to capture images from front-left, front-right, back-left and back-right views. These four additional images will be used to compute structured light grid points.

Images for reconstruction: The person or the mannequin whose surface is to be reconstructed is stationed at the center of the frame work and images from the twelve cameras are captured. At the time of taking the images we make the following arrangements:

- The person/ mannequin is standing vertically facing the front camera.
- The anchor points are marked with good contrast colors.

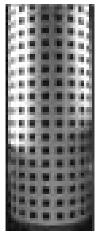

Figure 12: Calibration cylinder with grid pattern

- Structured light is projected on to specific non-convex parts.

Images for calibration: In our system we use a cylinder with a grid pattern (Figure 12) pasted on it as calibration target. After capturing the images of the human model, the calibration cylinder is placed at the center of the frame work and images from all the cameras are captured with out disturbing or modifying the camera setup.

3.2 Image processing

This part of the system takes the images of the object and the images of the calibration cylinder and produces the following output: Image contours, pixel values of anchor points in multiple views, pixel values of structured light grid points in multiple views, and pixel values of the corner points of the grid pattern in the images of the calibration cylinder.

3.3 Camera calibration

Camera calibration is done by Tsai's technique [9] by establishing 2D to 3D correspondence of the calibration points (corner points of the grid pattern) for each camera. The 3D coordinates are obtained by fixing a 3D coordinate system with respect to the cylinder.

3.4 3D reconstruction

This part takes the image contours, calibration parameters, pixel values of anchor points and structured light grid points from the image processing and camera calibration part and does the following:

1. Compute the three dimensional coordinates of the specified anchor points by stereopsis.

2. *Body parts definition:* Output/ display the 3D coordinates of the anchor points to the user and get from the user the start and end heights of various parts and also the required number of cross sections for each part.

3. *Cross section curves:* Compute the cross section curves for each part as NURBS curves by the method outlined in 2.1.5.

4. *Modification using structured light grid points:* The cross section curves obtained as above will in general be convex curves. So for non-convex parts the cross section curves will be modified by projection on to a grid obtained from structured light grid points as explained in 2.3.

5. *Interactive modification of the cross sections:* Some of the cross sections obtained as above may not be smooth or accurate. These can be modified interactively by picking and moving the control points of the curve.

The computed cross sections of all the parts of the body along with the body parts definition and the 3D anchor points will be the output.

3.5 3D surface representation

This is an application dependent part. Taking the cross sections and the body parts definitions as input, this part produces the surface in a format required by the application. The application we have addressed in our system is mannequin modeling for garment design. Here it is required that the reconstructed surface is represented as follows, giving both the geometrical and topological details.

1. *Topology:* Decomposition of the body into various natural parts.

2. *Geometry:* A continuous triangular mesh of the whole body.

3. *Feature lines:* Some specific sections of the body like chest line, waist line, etc.

4 User interaction

For 3D reconstruction of complex objects like the human body surface from images, it is very difficult and may not be possible to have a completely automatic system. There will be problems both in the 2D image processing part and the 3D reconstruction part. The problems in image processing are mainly due to different/insufficient contrast levels in various areas of the images which result in noises in automatically extracted contours and grid/corner points. There will also be missing points and contour segments. Apart from these, there are the known problems of correspondence for stereopsis and segmentation. Whereas in 3D reconstruction, one problem is inaccuracy in some areas due to small errors in calibration and image processing. The other is the inherent problem in 3D reconstruction from image contours which is that there may not be unique solution. We find that many

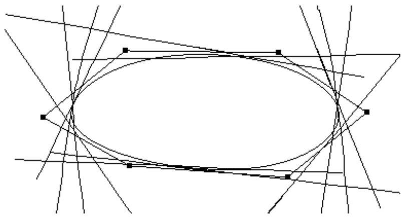

Figure 13: Editing a cross section

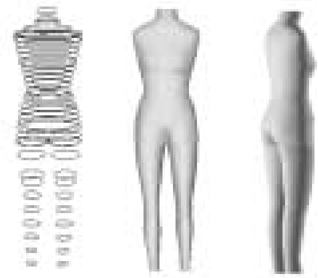

Figure 14: Final surface – cross sections and shaded views

of these problems can be solved effectively by user interaction. Here we mention some of the necessary user interaction tools provided in our system.

2D interaction: The system has the following tools for interactive image processing:

- Clean, that is fill with background color, a part of the image specified by rubber-band rectangle.
- Add/delete pixel.
- Add a piecewise linear curve in the image by inputting a sequence of points.
- Editing segments: rename, change orientation, join and split.
- Assign grid coordinates to structured light grid points.

3D interaction: Here the user can edit any specific cross section by picking and moving its control points. At the time of editing one can select to view only that cross section from top as shown in Figure 13. In this view we display the cross section curve along with its control points and control polygon. We also display the cross section plane contours derived from the images. The computed cross section is supposed to be inside the region bounded by these contours and be touching them.

5 Experimental results

In this section we present the results of our experimental reconstruction of a female mannequin using our system. Twelve images of the mannequin have been captured. Figures 2, 3 show four images of the mannequin and the extracted image contours with segmentation. At the time of image capturing two structured light grids – one in the front and the other in the back of the mannequin – were displayed to capture the concave portions of chest and hip.

For the purpose of reconstruction, the mannequin has been regarded as consisting of the seven parts – neck, shoulder, chest, torso, hip, left leg, and right leg. These body parts were demarcated using the anchor points back neck, front neck, shoulder points, bust point, lower bust point, belly button, crotch point, and ankle points whose 3D coordinates were obtained by stereopsis. Following the reconstruction procedure, cross sections were computed for each part. Also, two grids (Figure 9) on the surface were obtained using the structured light grid points. Then the cross sections of chest were modified using the grid on the front side and the cross sections of hip were modified using both the front and back grids. Finally the cross sections were fine tuned interactively for smoothness and accuracy of the surface. Figure 14 shows the final cross sections and the resulting surface in two views.

In order to verify the reconstruction, we projected the computed cross sections on to the projection planes of all the cameras and displayed them along with the respective projection plane contours. Figure 15 shows these for six cameras from which we can see that most of the cross sections fit well with the projection plane contours. In the Figure 15, (a), (c) and (e) correspond to upper level cameras and the projection plane contours are there only for the upper parts. Similarly, (b), (d) and (f) correspond to lower level cameras with the projection plane contours covering the lower parts. Note that in the overlapping part (Hip) the cross sections match the projection plane contours of both upper and lower level cameras even though they have been computed only using the lower level contours.

As another measure of accuracy, we have compared our reconstruction with that obtained by laser scanned data. Figure 16 shows the cross section at bust level of the mannequin obtained by our method along with that obtained from laser scan. We can see that the two curves match fairly well.

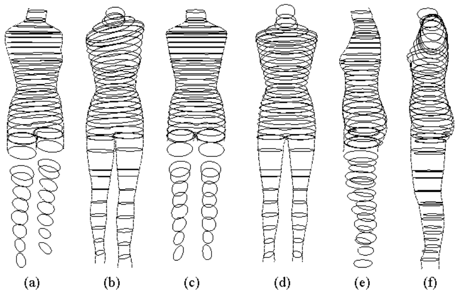

Figure 15: Projections of the cross section curves on to six projection planes and the corresponding projection plane contours

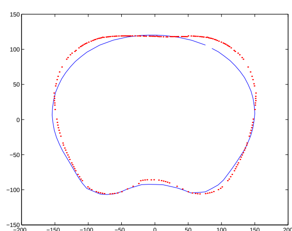

Figure 16: The cross section at bust level of the mannequin obtained by our method (solid line) along with that from laser scanned data (dotted line)

6 Conclusion

We have presented a 3D interactive reconstruction system for modeling of human body surface from multiple images taken from a human body image capturing setup. We have outlined the theoretical basis of the computer vision techniques used for reconstruction. We have highlighted the necessity of human interaction in the reconstruction process of such a complex object as the human body surface and have listed some of the interaction tools provided in our system. We have done experimental reconstruction of a female mannequin. Our experimental results show that it is possible to reconstruct most of the body parts including neck, chest, torso, hip and the legs quiet accurately and to combine them into a single continuous surface.

Acknowledgments

This research project is supported by grants from the Industry and Technology Development Council (grant no. AF/122/96), Research Grants Council of the Hong Kong Special Administrative Region, China (competitive earmarked research grant no. HKUST/754/96E), and UGC-UST Research Infrastructure (grant no. RI 95/96, EG18). The authors thank other members of the project – Dickson W. Tong, Allen Y. Tin, Yifan Liu, and Atkin S. Lui for their role in image capturing and image processing.

References

[1] Stephen Addleman. Whole-body 3d scanner and scan data report. In Richard N. Ellson and J. H. Hurre, editors, *Three-Dimensional Image Capture*, volume 3023, pages 2–5, 1997.

[2] A. Bottino, A. Laurentini, and P. Zuccone. Toward non-intrusive motion capture. In *Computer Vision-ACCV'98, Proceedings of the Third Asian Conference on Computer Vision, Hong Kong, China, 1998.*, pages 416–423, 1998.

[3] S. Gopalsamy and Helen C. Shen. Cross sections approach for 3d reconstruction from 2d silhouettes of multiple images. *Communicated for publication*, 1999.

[4] Jin Gu. *3D Reconstruction of Sculptured Objects*. PhD thesis, Department of Computer Science, Hong Kong University of Science and Technology, 1998.

[5] A. Laurentini. How many 2d silhouettes does it take to reconstruct a 3d object? *Computer Vision and Image Understanding*, pages 81–87, 1997.

[6] W.N Martin and J.K. Aggarwal. Volumetric description of objects from multiple views. *IEEE Transactions on Apttern Analysis and Machine Intelligence*, 5:150–158, 1983.

[7] H. Noborio, S. Fukuda, and S. Arimoto. Construction of the octree approximating a three-dimensional object by using multiple views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(6):769–82, 1988.

[8] P. Srinivasan, P. Liang, and S. Hackwood. Computational geometric methods in volumetric intersection for 3d reconstruction. *Pattern Recognition*, 23(8):843–57, 1990.

[9] R.Y. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323–44, 1987.