Research on 3D Surface Reconstruction by Deformable Model-based Methods

Jiahui Wang

2005

Contents

Figure index

Table index

Chapter 1

Introduction

This work pertains to the three-dimensional (3D) reconstruction of object shape by deformable model-based methods.

In the past and even presently, the deformable models are widely used in the research and commercial applications. In this chapter, the motivation of object shape reconstruction and existing deformable model techniques are described. At the end, the methods proposed in this work are described shortly.

1.1 Motivation of 3D Shape Reconstruction

Humans are primarily visual creatures, vision is our most powerful sense. Not all animals depend on their eyes, as we do, for 99% or more of the information received about our surroundings[65]. The vision sense also enables us to interact intelligently with the environment, all without direct physical contact. Through it we learn the position and identities of object and the relationships between them, and we are at a considerable disadvantage if we are deprived of this sense. Because the vision is our irreplaceable ability and computer has become the most indispensable instrument in our life, it is no wonder that attempts have been made to give computer a sense of vision almost since the time that digital computers first became generally available[40]. This is always called computer vision.

Computer vision is the enterprise of automating and integrating a wide range of progresses and representations used for vision perception. The input to the computer vision system may be an image, or series of images, but the output must be something quite different, which relates the visual input to previously existing models of the world[64]. There is a large representational gap between the input images and the output models, which explain, describe, or abstract the image information. The image-to-model correspondence is a bridge of the gap by alignment. In the alignment approach, the representations of the object (images) and the models (2D or 3D) maintain their pictorial nature[76].

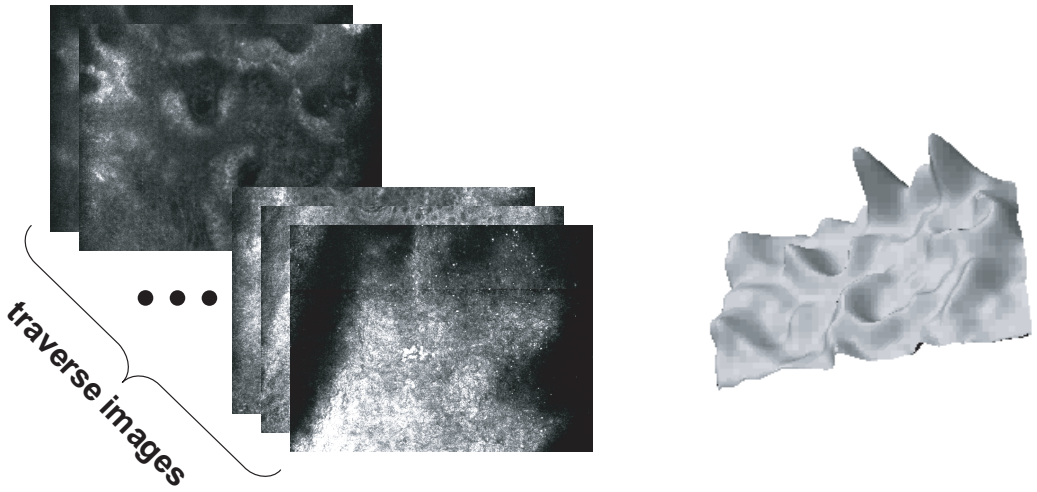
After compensating for the transformations separating the images and the appropriate model, corresponding parts of the two representations will be in close register. The ability to perform image-to-model correspondence can help us to identify ambiguous subparts and allows us to direct our attention to particular parts of the model. For instance, the top part of Figure 1.1 shows isolated parts of a puppet. The parts on their own are difficult to recognize (they are, in fact, the eye and ear in the face), but they become recognizable when the entire figure is recognized as a puppet. The effect will be more distinct for 2D images

Figure 1.1: The parts in the top image are difficult to recognize in isolation. In the bottom image, the recognition of the same parts is based on their correspondence with the appropriate parts in a puppet.

to 3D model. Figure 1.2 (a) shows confocal microscopy traverses of skin inner tissue. To observe the structure of skin inner tissue surface, the 2D confocal microscopy images are not sufficient even for experts. However, the skin structure becomes clear, when a 3D visual model (Figure 1.2 (b)) is reconstructed by the means of registering the corresponding image features and 3D model.

The correspondence can be used as a separate stage that drives the alignment process. A correspondence is first established between selected features in the images and model representations. Based on this match, the alignment transformation is derived and applied to bring the two representations into register. However, unexpected accidents may corrupt the transformation.

For instance, multiple camera images of a foot model are illustrated in Figure 1.3 (a). Because the object is occluded by obstructs, gaps in the images are occurred. A 3D model that reconstructed by image-based volumetric intersection algorithm is shown in Figure 1.3



(a) (b)

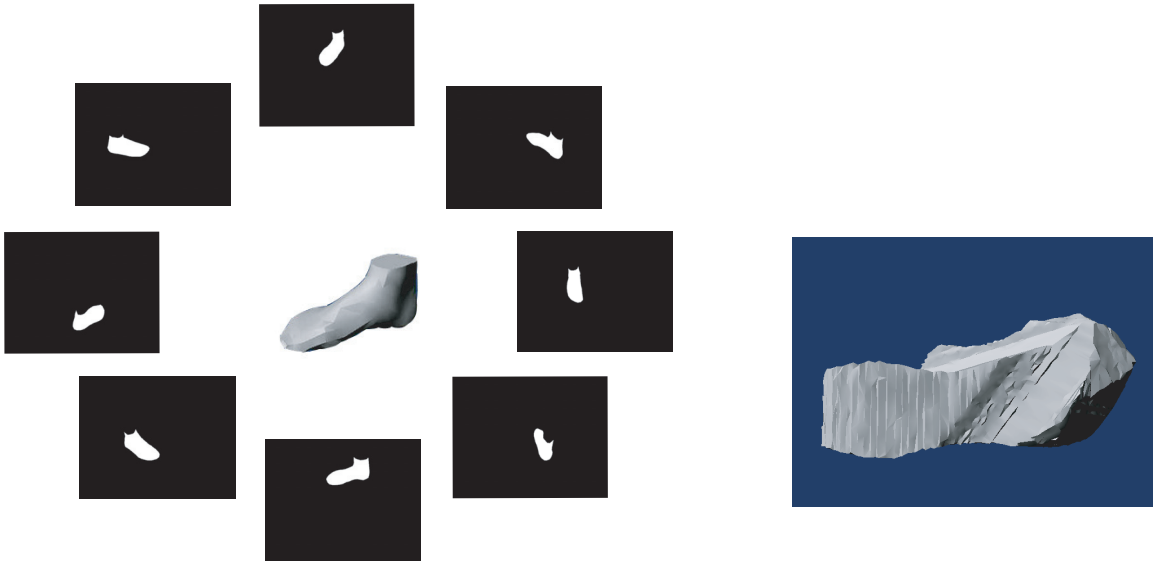
Figure 1.2: Confocal microscopy traverses ((a)) are widely used to observe skin inner tissue. 3D model ((b)) makes the observation clear.

(b). It is difficult to recognize the foot model due to the effect of image gaps.

To generate satisfied result, the revision must be considered from the beginning level. It is obvious that segmentation is crucial step in the succession of operations preceding 3D shape reconstruction, where its results are used to select the objects of interest among the various objects usually present in given image data. The most naive approach to segmentation is based solely on voxel characteristics (e.g. intensity) and is called point-based segmentation. Intensity thresholding belongs to this category. Such an approach however, does not take into account spatial information, such as proximity and connectivity. To take advantage of this kind of information, one has to use edge-based segmentation technique, described an object in terms of its bounding surface. Because there is no prior criterion, the partition of useful and unexpected edge elements is very difficult. Model-based approaches, i.e. deformable model methods give promising solution to address the image segmentation difficulty.

1.2 Overview of Deformable Model Techniques

The deformable models have wide applications in pattern recognition and computer vision, including image/video database retrieval[97], object recognition and identification[37], restoration[3], and object tracking[7].



(a) (b)

Figure 1.3: Multiple view images of a foot model ((a)). Image-based method generated unsatisfied 3D model ((b)) due to the effect of occlusion.

A deformable model can be characterized as a model, deforms the shape to match objects in given images. The deformations are under implicit or explicit optimal constraints. The model is active in the sense that it can adopt itself to fit the given data. Deformable model is a useful instrument because of its flexibility and its ability to both impose geometrical constraints on the shape and to integrate local image evidences.

Although the term deformable models first appeared in the work by Terzopoulos and his collaborators in the late eighties[38][72][73][74], the idea of deforming a template for extracting image features from image data back much farther, to the work of Fischler and Elschlager's spring-loaded templates[28] and Widrows' rubber mask technique[92]. Similar ideas have also been used in the work by Blake and Zisserman[10], Grenander et al.[31], and Miller et al.[50]. The popularity of deformable models is largely due to the seminal paper snakes: Active Contours Models by Kass, Witkin, and Terzopoulos[38]. Since its publication, deformable models have grown to be one of the most active and successful research areas in image processing and computer vision. Various names, such as snakes, active contours or surfaces, balloons, and deformable contours or surfaces, have been used in the literature to refer to deformable models.

In[36], Jain et al. partition the work on deformable model methods into two classes:

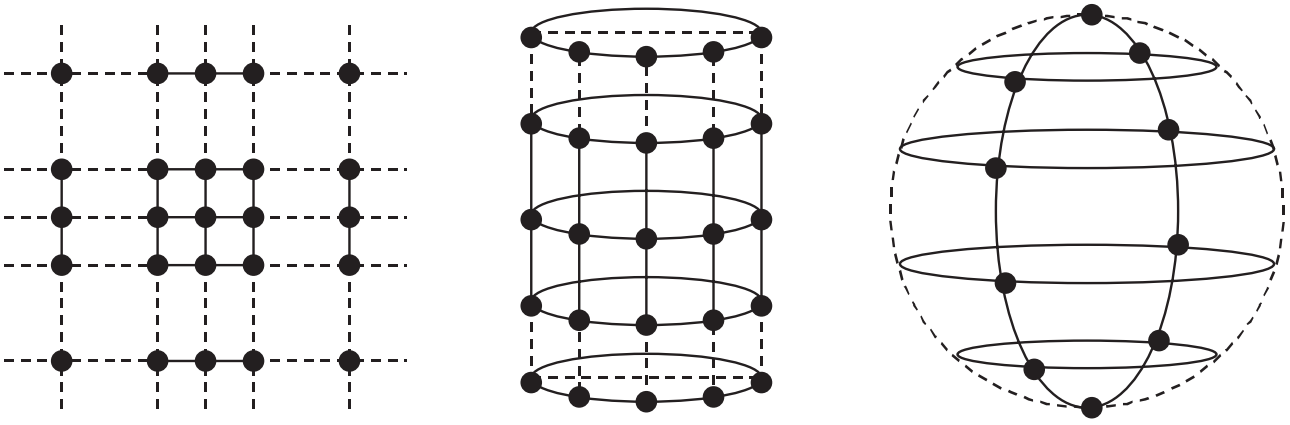
free-form deformable model method (FDMM) and parametric deformable model method (PDMM). The FDMMs can represent any arbitrary shape as long as some constraints like continuity, smoothness are satisfied. The other class, PDMM, encodes a specific shape and its variation where the shape can be characterized by a parametric formula and its deformation modes.

1.2.1 Free-form Deformable Model

FDMM always starts from an initial model. The initial model assumes very little structure about the object shape except for some regularization constraints like continuity and/or smoothness of the boundary. Such a free-form model can be deformed to match salient image features like lines, edges and surfaces using potential fields (energy functions) produced by those features. Since there is no global structure template, it can represent any arbitrary shape as long as the regularization requirements are satisfied. Kass, Witkin and Terzopoulos[38][74] introduced one of the earliest and most popular free-form deformable model: the Active Contour Model (ACM). An ACM is a geometrical curve (2D) or surface (3D), which approximates image contours through energy minimization. It behaves like an elastic rope that wriggles towards the contour or that slides down the potential hill. The internal energy keeps the shape and ensures the spatial and temporal continuity, while the image based potential force (image energy) at different scales to broaden its attraction range. Distinct weaknesses of this approach, however, is remained:

1. The regularized initial model decreases the flexibility of model;
2. Open form surface cannot be handled;
3. If the initial model is not initialized near the desirable object boundaries, the exploration of object boundaries will be failing.

In fact, the first and the second issues are correlated. The topologies of both 2D and 3D initial models are generally regularized, e.g. orthogonal lattice (Figure 1.4). Control points are set on the lattice. The deformation of model is implemented by adjusting the positions of the control points. The orthogonal and parallel are implied conditions for the optimization of initial model. The model surface smoothness is evaluated by discrete derivatives of control points' coordinate. However, the 3D spatial curvature at the contour of an open form surface is difficult to be estimated. Thus open form surface cannot be handled. Duan and Qin pro-



(a) (b) (c)

Figure 1.4: Topologies of initial models of FDMMs in 2D ((a)) and 3D ((b), (c)). Control points are created on the orthogonal lattice.

posed "Manifold Flow"[24] that affords an initial open patch (triangle or rectangle) to enlarge itself and flow directly over the object boundary through the expansion of its bounding contour towards along the surface tangent. Manifold Flow can grow on arbitrary boundary open or closed. The disadvantages of "Manifold Flow" is that the result may be affected largely by the position and orientation of the initial patch. Moreover, there is no established criterion for how to determine the size of initial open patch.

The third issue is also very important and has not solved completely. Although the image energy can broaden attraction range for the model, its effect is not enough. Spring force[38] is defined to be proportional to the distance between a point $\boldsymbol{X}$ on the model and a user-specified point $\boldsymbol{P}$. Spring forces act to pull the model toward the $\boldsymbol{P}$. The further away the model is from $\boldsymbol{P}$, the stronger the pulling force. The point $\boldsymbol{X}$ is selected by the means of finding the closest point on the model to $\boldsymbol{P}$ around a local neighborhood of $\boldsymbol{P}$.

Cohen proposed to increase the attraction range, using a pressure force together with the image energy. The pressure force can either inflate or deflate the model, hence, it removes the requirement to initialize the model near the desired object boundaries[17]. A disadvantage in using pressure forces is that they may cause the deformable model to cross itself and form loops[70]. Another approach for extending attraction range is to define the potential energy function using a distance map as proposed by Cohen and Cohen[18]. The value of the distance map at each pixel is obtained by the means of calculating the distance be-

tween the pixel and the closest boundary point, based either on Euclidean distance[23] or Chamfer distance[12]. Defining the potential energy function based on the distance map, one can obtain a potential force field that has a large attraction range. The distance potential force, however, can cause difficulties when deforming a contour or surface into boundary concavities[93]. To address this issue, Xu and Prince employed a vector diffusion equation that diffuses the gradient of an edge map in regions distant from the boundary, yielding a different force field called the gradient vector flow (GVF) field[93][94]. The amount of diffusion adapts according to the strength of edges to avoid distorting object boundaries. However, both the distance map and GVF are relying on reliable edge information in image. The initial model may be pushed to converse direction due to image noise. In this work, novel approaches are proposed to address these issues.

1.2.2 Parametric Deformable Model

The key difference between FDMM and PDMM is the usage of prior knowledge. PDMM incorporates additional prior knowledge into the models. Use of prior knowledge in a deformable model can lead to more robust and accurate results. This is especially true in applications where a particular structure that requires delineation has similar shape across a large number of subjects. Incorporation of prior knowledge requires a training step that involves manual interaction to accumulate information on the variability of the object shape being delineated. This information is then used to constrain the actual deformation of the contour or surface to extract shapes consistent with the training data. The active shape model (ASM), suggested by Cootes et al.[19][20][34][45] is now the most popular parametric deformable model method.

The standard deformation of ASM is implemented by adjusting both pose and shape variations to search the desirable location in given images. Here the shape variations are derived from prior knowledge by principal component analysis (PCA). Only deformations that produce shapes similarity to the prior knowledge are allowed. The image search stops when changes in both the pose and shape are insignificant. A limitation of the ASM is that it only typically search around the current position along profile normal to the edge element and does not take advantage of all the available information across the given images. Thus the ASM is not robust while the reliability of edge elements is doubtful.

Many extensions have been proposed to improve the performance of ASM. Duta and

Sonka[25] applied the ASM to segment subcortical structures from MR brain images. Wang and Staib[86] incorporated an additional smoothness prior into the ASM to allow the generation of more flexible shape instances.

Edwards, Cootes, and Taylor proposed an extension to the ASM, called active appearance model (AAM)[21][22][26]. Besides shape mode variations, AAM also employs PCA-based gray-level mode variations to incorporate all the information in parameters. Because the objects represented by AAM are more specific than those represented by ASM, in many applications, AAM can lead to more robust results than ASM, whereas AAM cannot directly handle cases well outside of the prior knowledge, e.g. occlusions. Moreover, in 3D space, to compute gray-level mode variations, the prior knowledge instances should be normalized in identical 3D intensity space[52][53], hence the time-consuming establishment of 3D intensity space must be performed. The most fatal issue is that AAM is at completely disadvantage, when the prior knowledge instances do not contain intensity information.

1.3 Proposed Methods

As effective image segmentation approaches, in this work FDMM and PDMM are used to address the 3D object shape reconstruction issues.

FDMM-based 3D reconstruction methods are proposed to deal with open form object surface, such as skin inner tissue surface. FDMM aims to impose smoothness on the initial model and encourage movement towards boundary elements in given images. However, the open form surface reconstruction is difficult for this technique due to the boundary condition. To address the open form surface issue, the initial model is partitioned into lateral surface initial model and central initial model. The lateral surface initial model corresponds to the contour of initial model. The lateral surface initial model is used to extracted object surface in lateral surface. Then the extracted object surface in lateral surface is recomposed to the central initial model as constant. This strategy is called active open surface model (AOSM). However, the partition of the initial model may cause uncontinuity of the reconstructed surface.

To overcome this issue, a triangle patches based initial model is introduced. The definition of the energy function particularly the internal energy is modified correspondingly, hence the processing can be more convenient. A novel external constraint is also proposed, so that the attraction range of the model is broadened. The proposed method is named energy-modified

active contour model (EACM).

In experiments, the 3D models of skin inner tissue surface are reconstructed by proposed methods from volumetric confocal microscopy images. The results are validated in comparison with human experts' observation.

EACM is also performed in another application-reconstruction of 3D foot model from multiple camera images. Thus, this approach is shown as well as available for reconstruction of closed object surface. However, because the proposed method does not consider the global anatomical structure of object, distinct shape characteristics are removed. A PPDM-based method is proposed to address this difficulty.

The PDMM-based method is similar to ASM in spirit. The pose and shape of initial model are described by a group of parameters, including scale, rotation, translate and weighting parameters of shape mode variations. Instead of analyzing gray-level mode variations from prior knowledge, the intensity information of corresponding pixel in given images is fused to assess the propriety of current model by the means of projecting the model to multiple camera images. In the foot shape reconstruction experiment, the PDMM-based method generated more accurate 3D model than the FDMM-based method and conventional volumetric intersection method. The PDMM-based method is named modified active shape model (MASM).

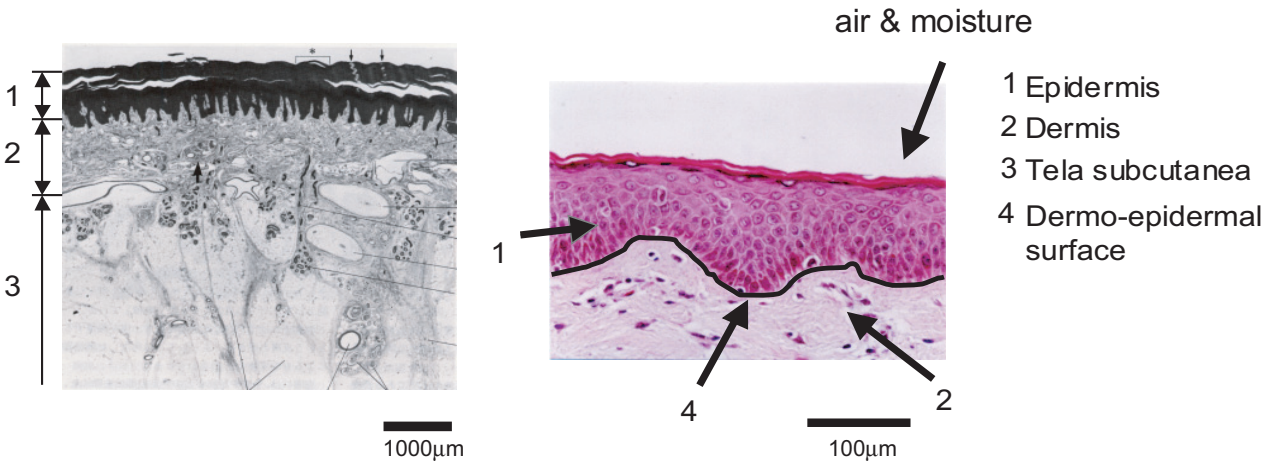
1.4 Contents

Chapter 2 describes the modified conventional FDMM-AOSM, so that it can handle the reconstruction of open form surface. The experiment on skin inner tissue surface reconstruction is introduced in the latter of this chapter. The general-purpose FDMM-based method-EACM is described in chapter 3. This method makes modification on both initial model and the definition of energy function. The proposed method is used to deal with both skin inner tissue surface and foot shape reconstruction. Chapter 4 explains the PDMM-based method and experiments of foot shape reconstruction by the proposed PDMM-based method (MASM). Chapter 5 gives the conclusions of this work and discussion of future works. The generation of triangle patches for 3D model surface is depicted in Appendix A. The Levenberg-Marquardt optimization algorithm is described in Appendix B. Appendix D describes a novel occlusion assessment method for multiple camera system. The Rosenbrock's optimization algorithm and the downhill simplex method are described in Appendix C and Appendix E respectively.

Chapter 2

Free-form Deformable Model-based Open Form Surface Reconstruction Method

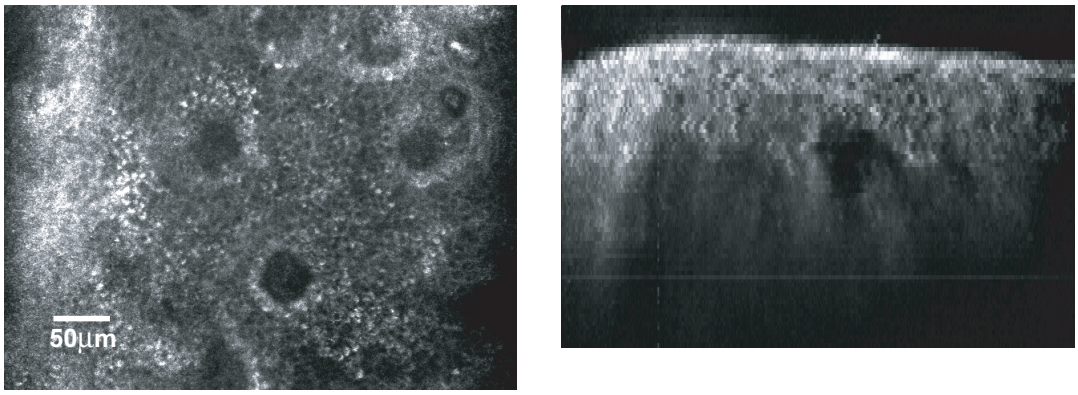
Attempting to reconstruct arbitrary shape structures, e.g. human skin inner tissue surface (Figure 2.1), in 3D can be a complicated task. Figure 2.2 illustrates confocal microscopy traverse and synthesized vertical section of skin inner tissue. To reconstruct the bottom surface of highlight region, i.e. dermo-epidermal surface, a common approach uses second intensity derivative. The zero crossing corresponds to the edge element. In Figure 2.3 the extracted edge elements are plotted in white. Because confocal microscopy images contain noise, unexpected edge elements are occurred frequently. Moreover, not only the dermo-epidermal surface, but also the upper epidermal surface (the upper surface of the highlight region) is extracted simultaneously.



(a) (b)

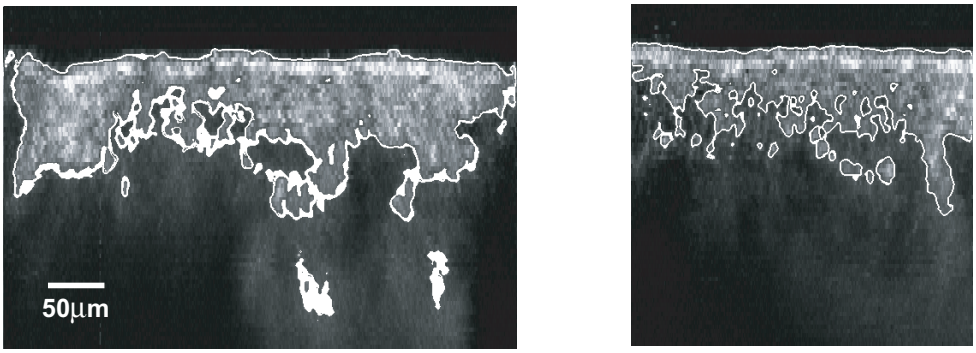
Figure 2.1: (a) Structure of human skin inner tissue[57]. (b) The specimen of face skin inner tissue.

Deformable model methods start from initial models similar to the object. Only the necessary image features such as edge elements are adopted. However, the skin structure is depending on the environment, sex and aging, thus it doesn't have stable shape. For instance, some researchers reported the skin thickness of face is decreased with age[43], however the increasing data is also reported[30]. As that described in last chapter, the FDMM can handle the arbitrary object shape reconstruction without prior shape constraint, thus in this research FDMM-based method is proposed for the shape reconstruction of skin inner tissue surface.



(a) (b)

Figure 2.2: The confocal microscopy traverse of skin inner tissue ((a)) and synthesized vertical section ((b)). The highlight region is corresponding to epidermis. The bottom surface of epidermis is called dermo-epidermal surface.



(a) (b)

Figure 2.3: The zero crossings of second intensity derivative are plotted in white. Unexpected edge elements are caused by image noise.

2.1 Principle of Free-form Deformable Model Method

In this section the conventional FDMM is described briefly. The so-called FDMM is a large category. Active contour model (ACM) method is the most popular FDMM. ACM is a 2D energy minimization curve. It deforms to match salient image features like lines and edges, using energy function produced by those features. The value of energy function is getting small when the model is approaching the boundary of object. To search the position that minimizes the energy function, the model and object will be registered. Such an ACM

assumes very little structure about the object shape except for some regular constraints like continuity and smoothness of the boundary. Unexpected edge elements are ignored by these constraints.

Let $C(q) : [0,1] \rightarrow \mathbf{R}^2$ be 2D space curve and let $I : [0,a] \times [0,b] \rightarrow \mathbf{R}^+$ be a given image, in which the objects boundaries are expected to be detected. The conventional ACM approach associates the curve C with an energy function given by

$$\boldsymbol{E} = \sum_{i=0}^{n} E(\boldsymbol{v}_i) = \sum_{i=0}^{n} [E_{int}(\boldsymbol{v}_i) + E_{image}(\boldsymbol{v}_i) + E_{con}(\boldsymbol{v}_i)] \tag{2.1}$$

where E_{int}, E_{image}, and E_{con} are referred as internal energy, image energy, and external constraint, respectively. $\boldsymbol{v}_i$ denotes a node on curve C. It is called control point. The control points are created symmetrically along the initial model at regular intervals (Figure 2.4).

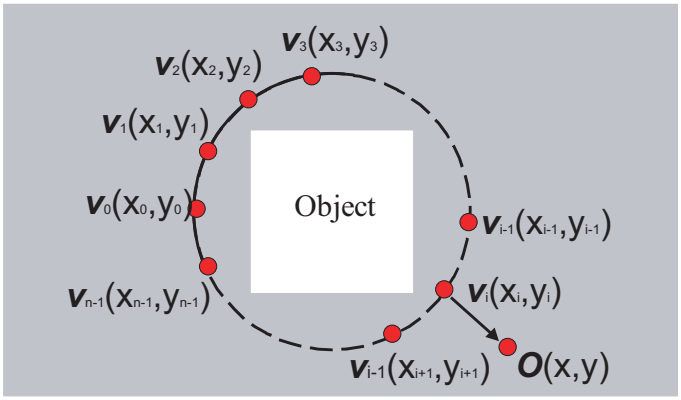


Figure 2.4: The control points are created symmetrically along the initial model at regular intervals. The minimization of $\boldsymbol{O}\boldsymbol{v}_i$ makes the model approach object boundary.

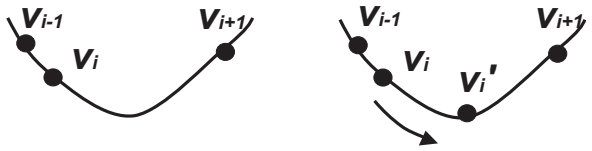
The curve around $\boldsymbol{v}_i$ being a controlled continuity spline with the first order derivative membrane term in (2.2) favoring control points to become closer to one another (Figure 2.5) and the second order derivative thin-plate term favoring control points to become equidistant (Figure 2.6).

$$\begin{aligned} E_{int}(\boldsymbol{v}_i) &= \alpha_i X'(\boldsymbol{v}_i)^2 + \beta_i X''(\boldsymbol{v}_i)^2 \\ &= \alpha_i |\frac{\partial \boldsymbol{v}_i}{\partial x} + \frac{\partial \boldsymbol{v}_i}{\partial y}|^2 + \beta_i |\frac{\partial^2 \boldsymbol{v}_i}{\partial x^2} + \frac{\partial^2 \boldsymbol{v}_i}{\partial y^2}|^2 \end{aligned} \tag{2.2}$$

where α_i and β_i are weighting parameters. The effect of first and second derivative terms are combined to make the curve smooth and not bent too much (Figure 2.7).

(a) (b)

Figure 2.5: The minimization of first order derivative makes distant control points ((a)) become close to one another ((b)).



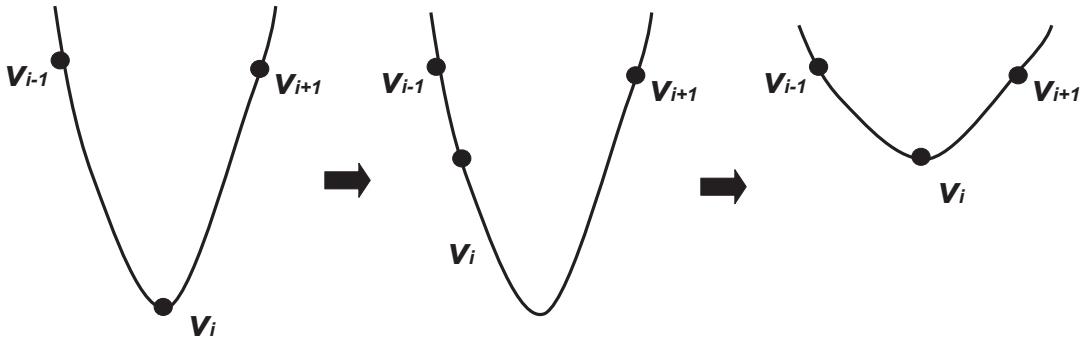
(a) (b)

Figure 2.6: The minimization of second order derivative favors the unequal neighboring control points ((a)) become equidistant ((b)).

The image energy is derived from image features of object. It includes three elements, which represent the attraction to lines, edges and terminations ((2.3)) respectively. The total image energy can be expressed as a weighted combination of these three elements. By adjusting the weighting parameters, a wide range of model behaviors can be created.

$$E_{image}(\boldsymbol{v}_i) = w_{line}E_{line}(\boldsymbol{v}_i) + w_{edge}E_{edge}(\boldsymbol{v}_i) + w_{term}E_{term}(\boldsymbol{v}_i) \tag{2.3}$$

The external constraint is an optional element for energy function. When the initial model



(a) (b) (c)

Figure 2.7: To integrate the close ((a)→ (b)) and equidistant ((b)→ (c)) effects, the internal energy offers satisfied smoothness.

is created apart from the object boundary, the external constraint may push the initial model towards the target. It represents a force of a spring connected between a point on the contour and manually specified points in the image plane.

$$E_{con}(\mathbf{v}_i) = -k[(x_i - x)^2 + (y_i - y)^2] \tag{2.4}$$

where k is a positive real constant, $\mathbf{v}_i(x_i, y_i)$ and $O(x, y)$ is defined as Figure 2.4. Since $-k \leq 0$, the external constraint is a counterforce.

Because image energy and external constraint are describing the influence from external environment on initial model, they are combined called external energy function.

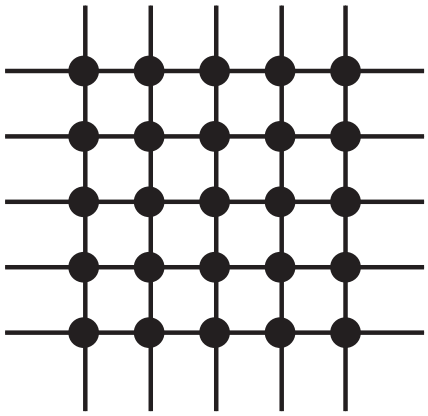


Figure 2.8: In 3D the initial model is extended from a curve to a surface. The deformation is performed, adjusting the positions of control points.

2.2 3D Active Contour Model

After the conventional ACM is reported, it has grown to be one of the most active and successful image segmentation and contour extraction techniques. Since then it was extended to be implemented in 2.5D[48][71] and 3D medical applications[75], where ACM was constrained to encourage axial symmetry and is evolving under the forces determined from a 2D image or a pair of 2D images. In [18], Cohen et al. proposed full 3D solution in true 3D data set.

The main enhancement and modification for extending the conventional ACM to 3D is on the initial model and energy function.

In 3D the initial model is extended to a surface. The surface is controlled by lattice control

points (Figure 2.8). For lattice control points' regular relationships, the internal energy can be calculated conveniently.

The energy function of 3D ACM also includes the same three components: (E_{int}, E_{image} and E_{con}). First, the internal energy is transformed as:

$$\begin{aligned} E_{int}(\boldsymbol{v}_i) &= \alpha_i X'(\boldsymbol{v}_i)^2 + \beta_i X''(\boldsymbol{v}_i)^2 \\ &= \alpha_i |\frac{\partial \boldsymbol{v}_i}{\partial x} + \frac{\partial \boldsymbol{v}_i}{\partial y} + \frac{\partial \boldsymbol{v}_i}{\partial z}|^2 + \beta_i |\frac{\partial^2 \boldsymbol{v}_i}{\partial x^2} + \frac{\partial^2 \boldsymbol{v}_i}{\partial y^2} + \frac{\partial^2 \boldsymbol{v}_i}{\partial z^2}|^2 \end{aligned} \tag{2.5}$$

where $X'(\boldsymbol{v}_i)$ and $X''(\boldsymbol{v}_i)$ are discrete 3D first and second order derivatives, respectively, α_i and β_i determine the relative weights given to these two terms.

The 3D extension of the image energy is derived from 3D image features of object, whereas the basic definition is not changed from (2.3).

Finally, the external constraint is transformed as

$$E_{con}(v_i) = -k[(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2] \tag{2.6}$$

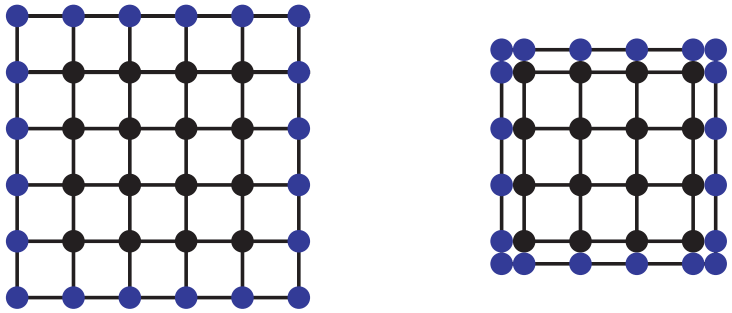
The 3D external constraint also represents a force of a spring connected between a point $\boldsymbol{v}_i(x_i, y_i, z_i)$ on the initial model and manually specified points $O(x, y, z)$ in 3D given image space.

2.3 Open Surface Strategy-Active Open Surface Model

According to the forementioned explanation, the initial models of both 2D and 3D ACM are in closed form. Thus, only closed object contour or surface can be handled. However, in many clinical situations, the processing is concerning open form surface, such as the skin inner tissue surface.

To cope with open form surface, the initial model will be created as a open plane (Figure 2.9 (a)). The surface smoothness of this open plane is also offered by minimizing the discrete derivatives-based internal energy. However, for the contour control points (blue points), the second order discrete derivative cannot be estimated. If the second order derivative is invalidated, the effect of first order derivative may drag the contour control points and their neighbor control points stuck together (Figure 2.9 (b)).

Although Duan and Qin gave a solution called "Manifold Flow" for this issue (Chapter 1), disadvantages are remained. First, during the image search, only the boundary contour of model is active and allowed to move. Second, the initial open patch to enlarge itself and



(a) (b)

Figure 2.9: To handle open form surface, the initial model is created as a open plane ((a)). If the second order derivative is invalidated, the effect of first order derivative may drag the contour control points and their neighbor control points stuck together ((b)).

flow directly over the object boundary through the expansion of its bounding contour towards along the surface tangent, thus the concave surface cannot be recovered. Moreover, there is no established criterion for how to decide the size of initial open patch.

2.3.1 The Modification on Initial Model

To overcome the difficulty on open form initial model, the most straightforward strategy is to create particular boundary condition for the open surface[77][78][79][82][83].

Concretely, the initial model is partitioned into two parts: lateral surface initial model and central initial model (Figure 2.10). The application is implemented in volumetric data space. For instance, the confocal microscopy volumetric data is illustrated in Figure 2.11. If the volumetric data is regarded as a hexahedron, its lateral surface can be considered a 2D continuous plane. To show this plane in an intuitional way, the lateral surface is unfolded, nevertheless the left and right edges are joined actually. To detect the dermo-epidermal surface the lateral surface initial model, i.e. a closed space curve, is set at the lateral surface (Figure 2.12). Because the initial model is a closed curve and the processing is implemented in a 2D surface, the object boundary in the lateral surface can be detected by conventional ACM (2D). The detected dermo-epidermal surface is plotted in Figure 2.13 in white. The detected object surface is then recomposed with the unprocessed central initial model as constant. Because the boundary condition is offered by the detected lateral surface object contour, the energy function of central initial model can be computed as conventional 3D ACM.

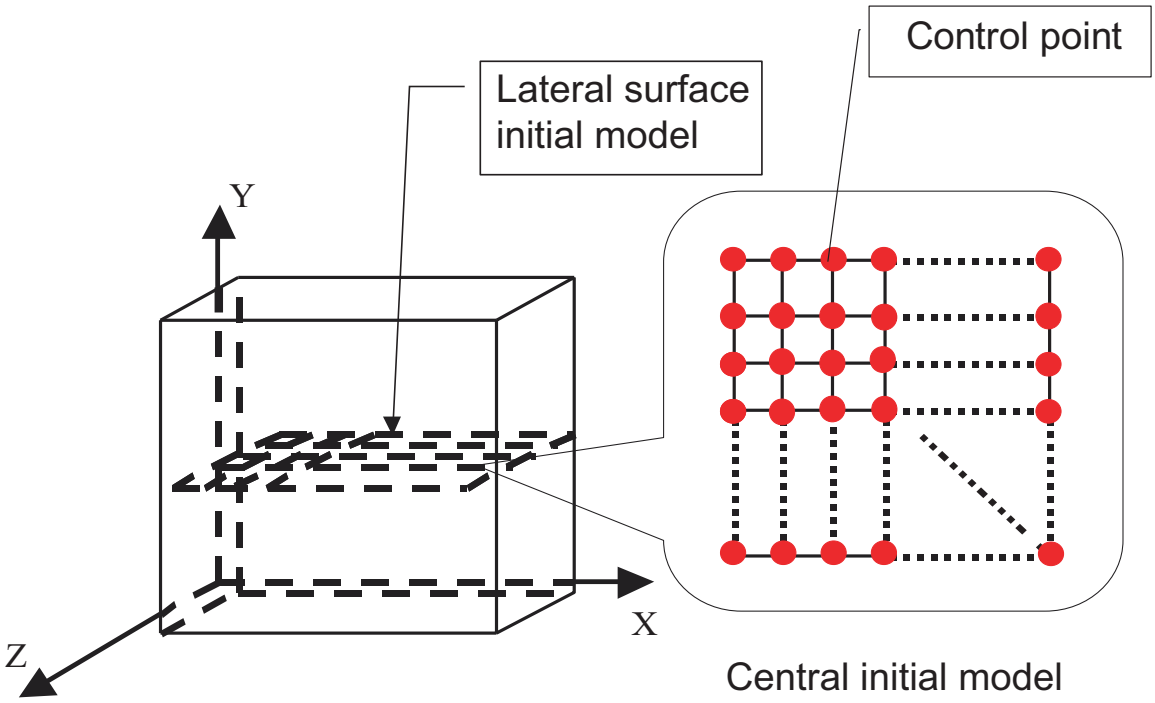


Figure 2.10: Initial models of active open surface model is partitioned into lateral surface initial model and central initial model.

Because this modification offered an open surface extension of conventional ACM, this strategy is called active open surface model (AOSM).

2.3.2 Image Energy with Gaussian Function

In many clinical situations, the shape reconstruction due to detect the edge elements from given gray-level image data, such as the skin inner tissue surface reconstruction. Thus, the E_{edge} component of image energy is kept. The image energy can be computed in many modes. For instance,

$$E_{edge}^{(1)}(\boldsymbol{v}_i) = I(\boldsymbol{v}_i) \tag{2.7}$$

$$E_{edge}^{(2)}(\boldsymbol{v}_i) = |\nabla I(\boldsymbol{v}_i)| \tag{2.8}$$

$$E_{edge}^{(3)}(\boldsymbol{v}_i) = (\nabla^2 G * I(\boldsymbol{v}_i))^2 \tag{2.9}$$

where, $I(\boldsymbol{v}_i)$ is the intensity of control point $\boldsymbol{v}_i$ in 3D voxel space. The definition of image energy is generally determined by applications. In this work, because the confocal microscopy

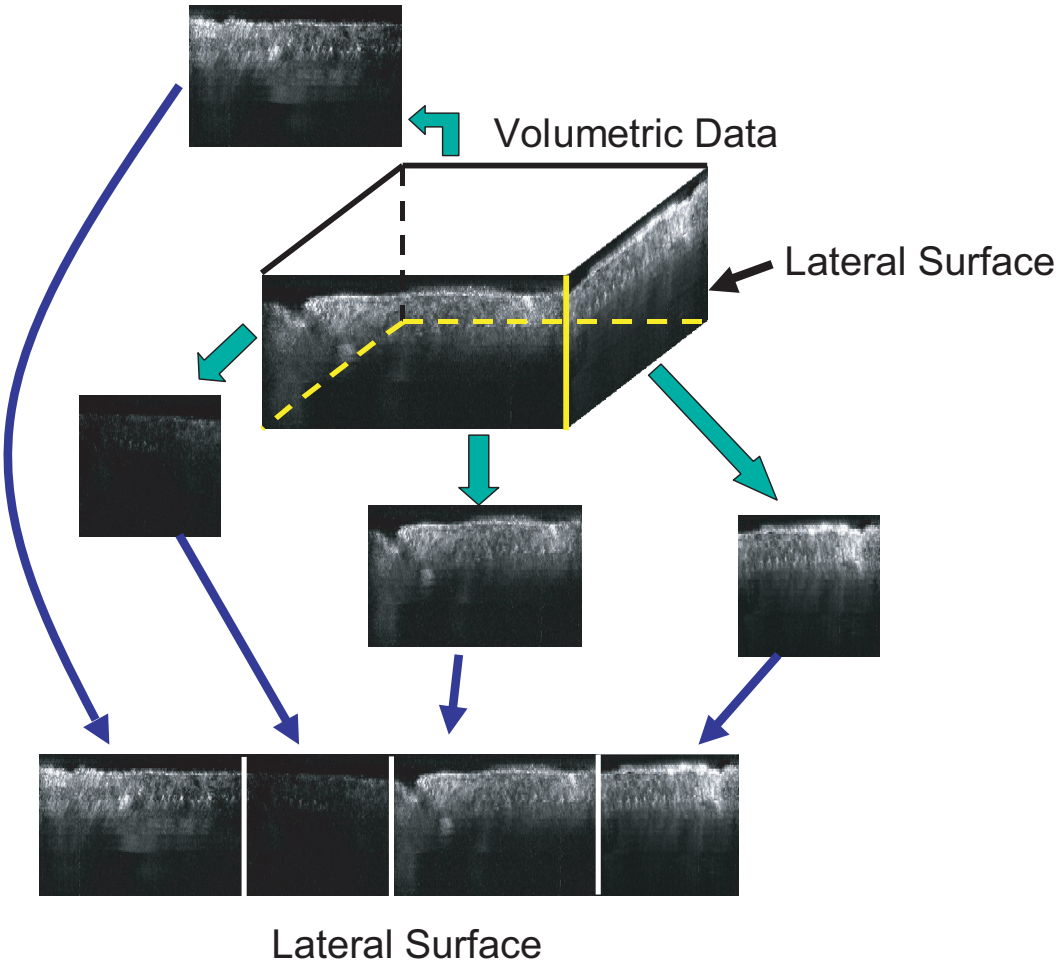


Figure 2.11: Unfolding the lateral surface of volumetric data.

images contain strong noise generally, the antinoise Gaussian function-based image energy ((2.9)) is adopted.

$$E_{image} = w_{edge}E_{edge} = -w_{edge}(\nabla^2 G_\sigma * I)^2$$
$$= w_{image}([\frac{1}{\sqrt{(2\pi)^3}\sigma^5}e^{-(r^2/2\sigma^2)}(3-\frac{r^2}{\sigma^2})] * I)^2 \qquad (2.10)$$

where G_σ is a Gaussian function with standard deviation σ. ∇ is the gradient operator. $\nabla^2 G_\sigma$ is so-called LoG filter[11][32][49][55]. Because of Gaussian function's smooth capability, the image data with strong noise can be removed effectively.

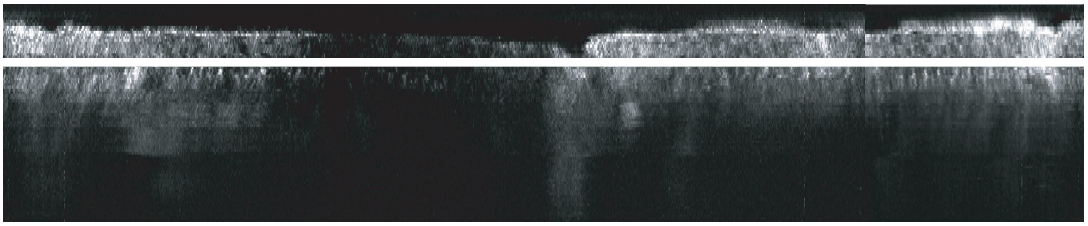


Figure 2.12: Lateral surface initial model in the lateral surface.

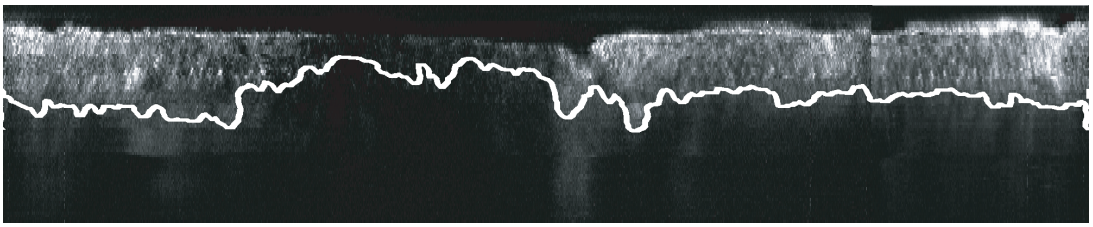


Figure 2.13: Detected dermo-epidermal surface in the lateral surface.

2.3.3 External Constraint

To reconstruct the 3D model of skin inner tissue surface, the external constraint is usually necessary. For instance, to reconstruct the dermo-epidermal surface from confocal microscopy images, although the exact position of dermo-epidermal surface is not known before the processing, it can be presumed that the dermo-epidermal surface is corresponding to the bottom of the high intensity region (Figure 2.2 (b)). Thus, initial model is created at the bottom half of the volumetric data. Then a reference plane is specified under the initial model. The distance to the reference plane for each control point is to be the external constraint (Figure 2.14). To specify minus weighting parameters for the external constraint, the initial model will approach the object surface by minimizing the external constraint.

$$\begin{aligned} E &= \sum_i [E_{int}(\boldsymbol{v}_i) + E_{ext}(\boldsymbol{v}_i)] \\ &= \sum_i \{w_{int}[\alpha_i X'(\boldsymbol{v}_i)^2 + \beta_i X''(\boldsymbol{v}_i)^2] + w_{image}|\nabla^2 G * I(\boldsymbol{v}_i)|^2 + w_{ext}(\boldsymbol{v}_i - \boldsymbol{s}_i)\} \end{aligned} \tag{2.11}$$

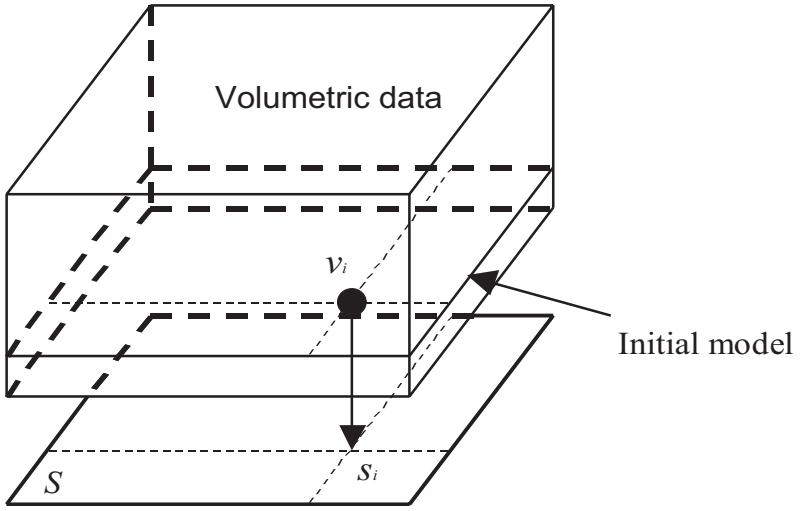


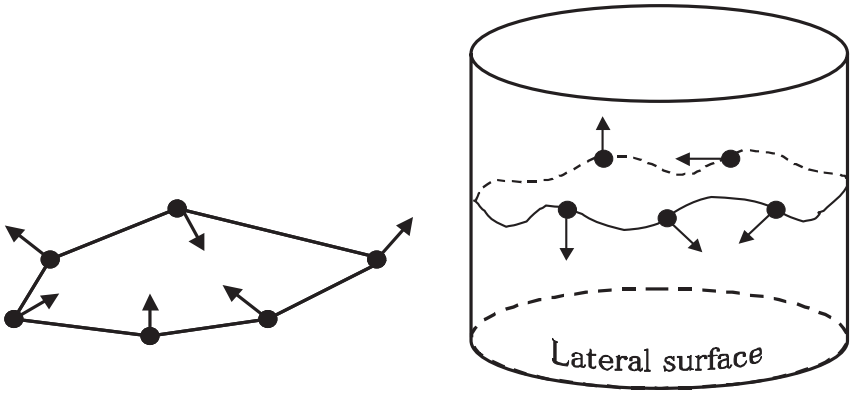
Figure 2.14: The initial model is created at the bottom half of the volumetric data. To push the initial model towards the object, a reference plane S is specified under the initial model. The distance from any control point $\boldsymbol{v}_i$ to corresponding point $\boldsymbol{s}_i$ on the reference plane is to be the external constraint.

2.3.4 Numerical Implementation

The total energy function is described as (2.11). Because there is strong noisy in confocal microscopy images and the shape of skin inner tissue surface is unstable, the automatic decision of weighting parameters is almost impossible. However, since the 3D shape reconstruction from confocal microscopy images is quite valuable, in the experiments the weighting parameters are determined by users' experience.

Because the initial model of AOSM is partitioned into lateral surface initial model and central initial model, the energy-minimizing model cannot be described in a single formula. To search desirable position for initial model, discrete dynamic programming[2][8] is applied. The initial model of AOSM is divided into stages by control points. In the dynamic programming framework, energy-minimization is enforced on the control points by changing their positions in a definite field. Since the convergence of this algorithm is guaranteed, the resulting solution is optimized.

The control points are adjusted by searching the 26 neighbor field in three-dimensions to minimizing the energy function (Figure 2.15 (a)), whereas there are only 8 candidate positions for control points on lateral initial model (Figure 2.15 (b)). The movements of the control



(a) (b)

Figure 2.15: The control points are adjusted in 26 neighbor field ((a)), whereas there are only 8 candidate positions for control points in lateral initial model ((b)).

points on lateral initial model are restricted in such a 2D surface-lateral surface of volumetric data.

2.4 Experiment on Skin Inner Tissue Surface Reconstruction

Because the skin inner tissue is invisible to human eyes and common optical cameras directly, the acquisition of skin inner tissue images will be a difficult work. In recent years, with the development of confocal microscope technology, confocal microscope images have become an important instrument for in-vivo observation of the skin inner tissue[15][41][42][66]. For instance, researchers can observe the changes in the dermo-epidermal surface before and after applying cosmetic without harming human subjects. This technology has thus received great attentions from cosmetics developers. However, to observe the structure of skin inner tissue surface, the 2D confocal microscopy images are not sufficient even for experts. In this work, 3D model of skin inner tissue surface is established from confocal microscopy images, using the proposed AOSM method.

2.4.1 Confocal Laser Scanning Microscope

Before explaining the experimental results, a simple introduction of the skin image acquisition instrument-Confocal Laser Scanning Microscope (CLSM) is described in this section.

2.4.1.1 Light Confocal Microscope

The introduction of laser-scanning confocal microscope generated as instant an excitement among biologists[59]. With this microscope, one can slice incredibly clean thin optical sections out of thick specimens; view specimens in planes tilted to, and even running parallel to, the line of sight; penetrate deep light-scattering tissues; gain impressive three-dimensional views at very high resolution; obtain differential interference or phase-contrast images in exact register with confocal fluorescence images; and improve the precision of micro-photometry.

Marvin Minsky applied for a patent in 1957 for a microscope that used a stage-scanning confocal optical system. Not only was the conception farsighted, but his insight into the potential application and significance of confocal microscopy was nothing short of remarkable[59][51].

In Minsky's embodiment of the confocal microscope, the conventional microscope condenser is replaced by a lens identical objective lens. The field of illumination is limited by a pinhole, positioned on the microscope axis. A reduced image of this pinhole is projected onto the specimen by the "condenser". The field of view is also restricted by a second (or exit) pinhole in the image plane placed confocally to the illuminated spot in the specimen and to the first pinhole (Figure 2.16). This is called trans-illuminating mode confocal microscope. Instead of trans-illuminating mode, epi-illumination mode makes a single objective lens serve as both the condenser and the objective lens (Figure 2.17).

2.4.1.2 Laser Illuminated Confocal Microscope

During the early 1970s, Egger and his co-workers at Yale University developed a laser-illuminated confocal microscope, in which the objective lens is oscillated in order to scan the beam over the specimen[27].

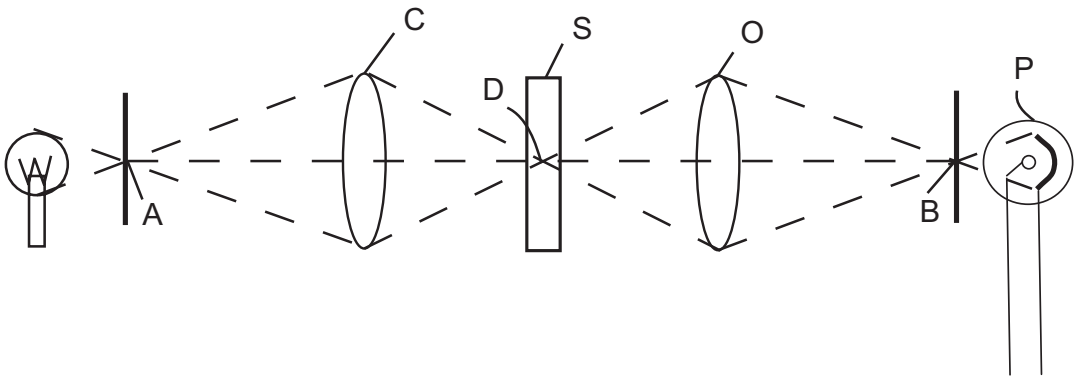


Figure 2.16: Optical path in trans-illuminating confocal microscope. The condenser lens (C) forms an image of the first pinhole (A) onto a confocal spot (D) in the specimen (S), and exit pinhole (B) are confocal points. The objective lens (O) forms an image of (D) onto the photocell (P) from second pinhole (B), which is confocal with (D) and (A).

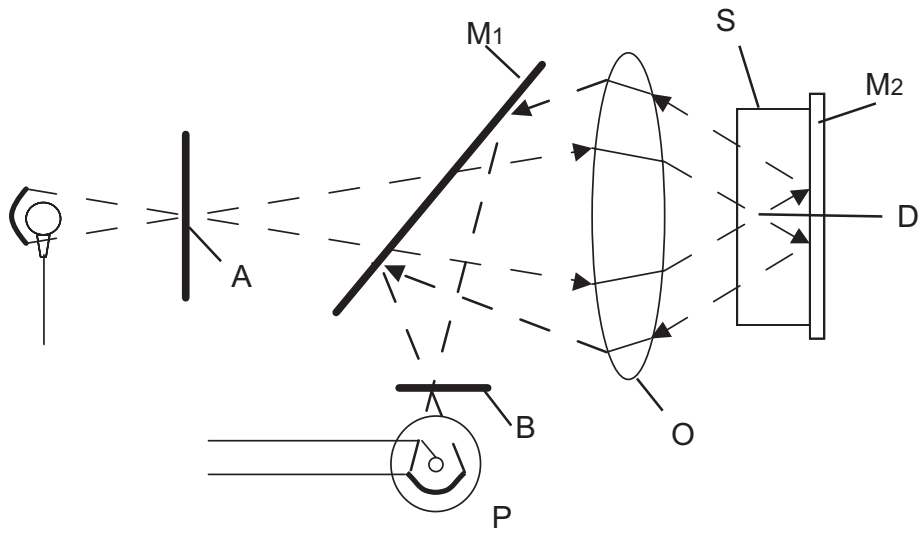


Figure 2.17: Optical path in epi-illnstead confocal microscope. The entrance pinhole (A), point (D) in the specimen (S), and exit pinhole (B) are confocal points. A partial mirror (M_1) transmits the illustrating beam and reflects the beam, which passed (D) and was reflected by the mirror (M_2), on which the specimen is lying. Only the reflected beam that passes point (D) focuses onto the detector pinhole and reaches the photocell (P). A single lens (O) replaces the condenser and objective lenses.

A few years later, Sheppard and Choudhury provided a thorough theoretical analysis on various modes of confocal and laser-scanning microscopy[67]. The following years, Sheppard et al.[68] and Wilson et al.[91] proposed an epi-illuminating confocal microscope of the stage-scanning type, equipped with a laser source and a photo multiplier tube (PMT) as the detector, using a novel specimen holder. The specimen holder, supported on four taut steel wires running parallel to the optical axis, allowed precise, z-axis positioning as well as fairly rapid, voice-coil-actuated scanning of the specimen in the xy-plane. Using this instrument, Sheppard et al. demonstrated the value of the confocal system particularly for examining integrated circuit chip. With stage-scanning confocal images, optical sections and profile images could be displayed on a slow-scan monitor over areas very much larger than can be contained within the field of view of any given objective lens by conventional microscopy.

The pioneering work described above was followed in several European laboratories by Brakenhoff et al.[13][14], Wijnaendts van Resandt et al.[89], and Carlsson *et al.*[16]. These investigators developed the stage-scanning confocal microscope further, verified the theory of confocal imaging and expanded its application into cell biology. In the meantime, video microscopy and digital image processing were also advancing at a rapid rate.

These circumstances culminated in the development of the confocal laser-scanning microscope (CLSM)[5][6] and publication of its biological applications by Carlsson et al.[16], Amos et al.[4] and White et al.[88]. The publications were followed shortly by introduction of laser-scanning confocal microscopes to the market by Sarastro, Bio-Rad, Olympus, Zeiss, and Lucid.

2.4.1.3 VIVASCOPE 1000 Laser Scanning Confocal Microscope

In this research, skin inner tissues are imaged by a commercially available CLSM (VIVASCOPE 1000, Lucid Inc., NY, USA) in Shiseido Ltd. Life Science Research Center. The system is illustrated in Figure 2.18.

The VivaScope 1000 is a CLSM capable of imaging living tissue at the cellular level. A non-invasive pathological examination capable of providing medical practitioners with information at the point and time of taking care more rapidly proceed down the path to diagnosis and treatment.

The VivaScopeR 1000 makes it possible to:

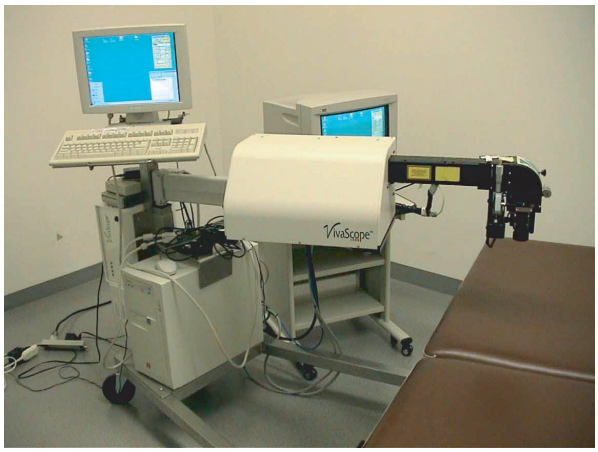


Figure 2.18: Vivascope1000 confocal microscopy system in Shiseido Ltd. Life Science Research Center.

- Support pathologic screening without invasion;
- Aid treatment with timely information;
- Receive results at the time and place of care;

Non-invasive treatment in dermatology and many other medical specialties has come a long way. But too often the assessment of a lesion still begins with an invasive procedure. Now there is a non-invasive solution for pathological visualization of living tissue. The VivaScopeR 1000 gives a ”window” into living tissue right in a clinical setting.

Using the VivaScope 1000’s imager, cellular and nuclear structures in living tissue one cell layer can be observed at a time. We can now produce images of thin virtual sections of skin or other exposed tissues without biopsy.

2.4.2 Experimental Results

A confocal microscopy traverse of human skin inner tissue is shown in Figure 2.19. It is assumed that this plane is paralleling to the XY coordinate plane. The images are acquired by CLSM with $\times 50$ objective lens. The image resolution is 640×480 pixels. Each pixel corresponds to a length of $0.703 \mu m$. The z-axis sampling distance is $3.32 \mu m$. A $640 \times 480 \times 400$ intensity voxel space (volumetric data) is established by linear interpolation. Figure 2.20 illustrates interpolated vertical sections in XZ-axis view and YZ-axis view. Comparing with the skin inner tissue specimen (Figure 2.1(b)), the high intensity region is corresponding to

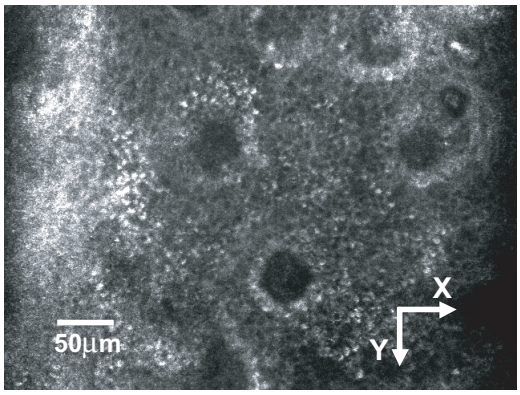
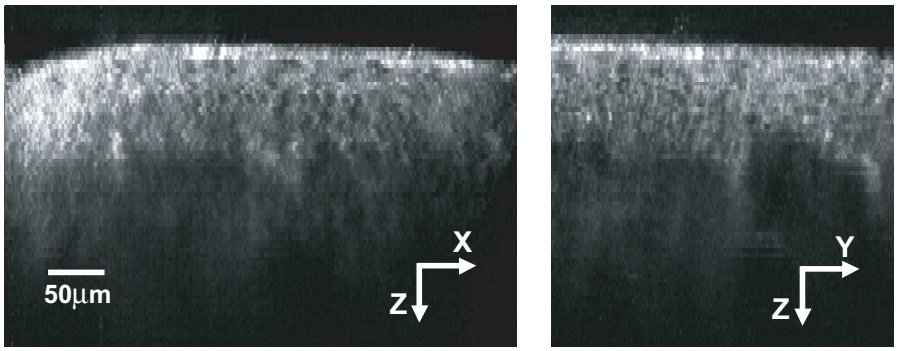


Figure 2.19: The confocal microscope traverse of skin inner tissue. This plane is treated as XY coordinate plane.



(a) (b)

Figure 2.20: Synthesized vertical sections of skin inner tissue. (a) XZ-axis view, (b) YZ-axis view.

epidermis. The low intensity region under epidermis is corresponding to dermis.

AOSM is applied to detect the dermo-epidermal surface from volumetric data for reconstructing the 3D model. As described above, the initial model of AOSM is divided into two parts: lateral surface initial model and central initial model. The dermo-epidermal surface in lateral surface of volumetric data is detected first. Second, the boundary of dermo-epidermal surface is recomposed to the central initial model. In the next, the central initial model is optimized, while the pre-detected contour is treat as constant. This processing flow is illustrated in Figure 2.21.

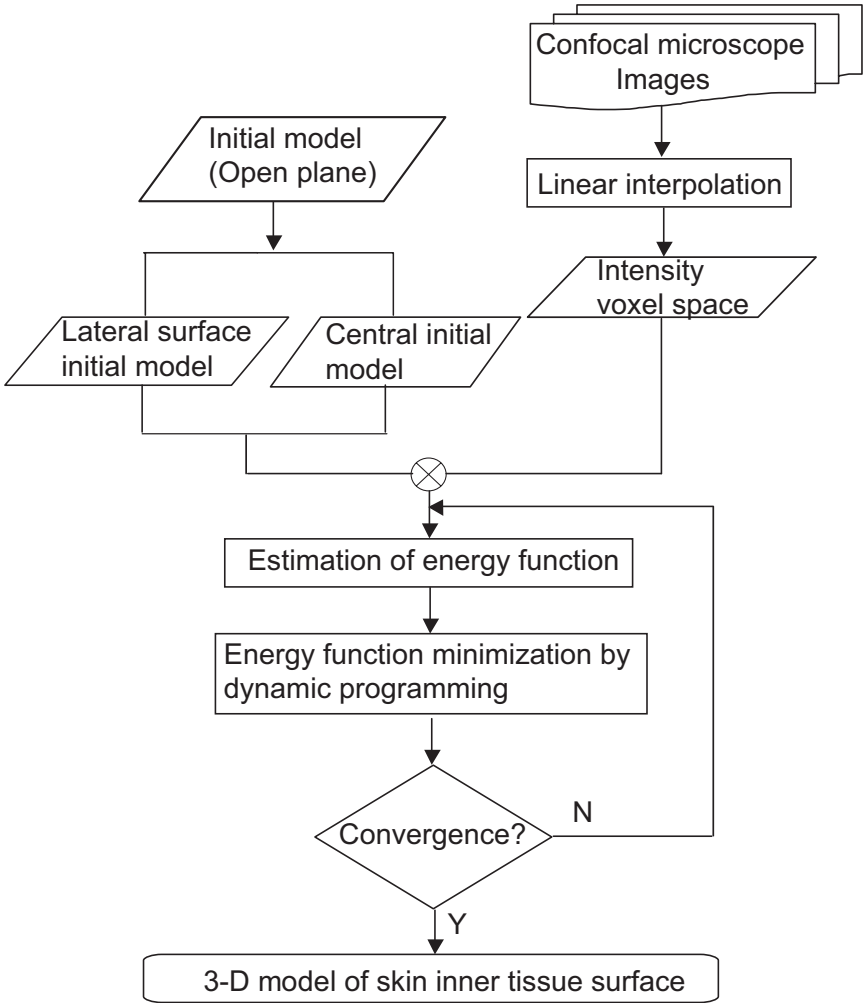
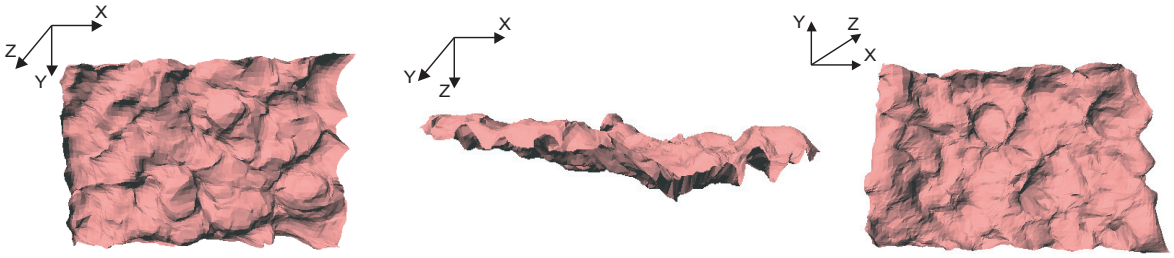


Figure 2.21: Flow chart of the AOSM-based skin inner tissue surface reconstruction system.

2.4.2.1 Reconstruction of Dermo-epidermal Surface

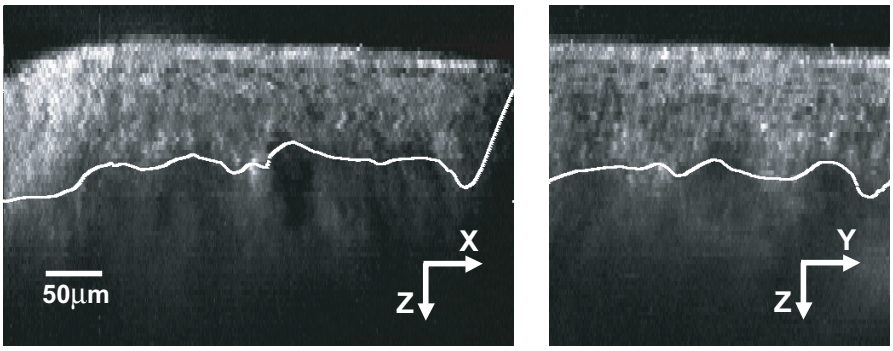
The reconstructed 3D model of dermo-epidermal surface is illustrated in Figure 2.22. The 3D model of dermo-epidermal surface is composed of triangle patches, which are established by 3 neighbor control points. The detail of establishing 3D model by control points is described in Appendix A.

To assess the register of 3D model and given image data, the positions of 3D model are depicts vertical sections of confocal microscopy images (Figure 2.23). The dermo-epidermal surface is plotted in white. It can be considered that the 3D model is almost registers the



(a) (b) (c)

Figure 2.22: Reconstructed 3D dermo-epidermal surface, displayed in top (a), side (b), down (c).



(a) (b)

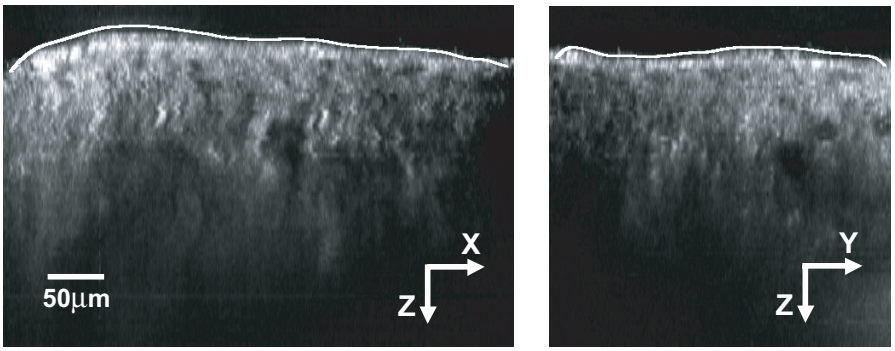
Figure 2.23: Reconstructed dermo-epidermal surface is plotted in white in vertical sections. (a) XZ-axis view, (b) YZ-axis view.

image of dermo-epidermal surface.

2.4.2.2 Upper Epidermal Surface Reconstruction and Quantitative Analysis of Epidermis Thickness

In [33][58][69], it has been shown that the epidermal thickness changes in response to aging, exposure to sunlight, humidity changes and so forth. Thus, the measurement of epidermal thickness is also a significant construction for cosmetics developers. A direct measurement of epidermal thickness can be achieved by calculating the average Euclidean distance of corresponding points on dermo-epidermal surface and upper epidermal surface.

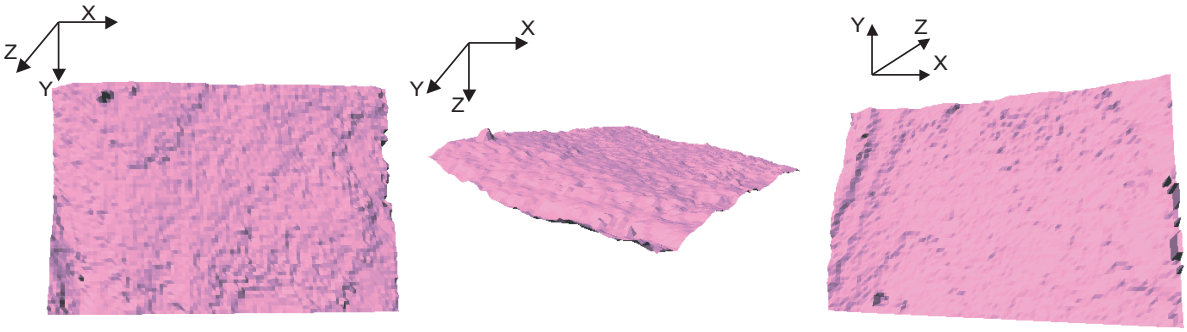
Figure 2.24 and Figure 2.25 illustrate the reconstructed upper epidermal surface in 2D and 3D by AOSM, respectively.



(a) (b)

Figure 2.24: Reconstructed 3D upper epidermal surface is plotted in white in vertical sections. (a) XZ-axis view, (b) YZ-axis view.

Since the dermo-epidermal surface has already been obtained, Figure 2.26 illustrates epidermis thickness distribution plot. The average thickness is about 110μm.



(a) (b) (c)

Figure 2.25: Reconstructed 3D upper epidermal surface, displayed in top (a), side (b), down (c).

2.5 Discussion

To validate the results further, the comparison with human experts' observation is performed. Figure 2.27 illustrates the detected dermo-epidermal surface in the same vertical section. The dermo-epidermal surface that detected by AOSM is plotted in white (Figure 2.27 (a)). The human experts' observed dermo-epidermal surface is plotted in yellow (Figure 2.27 (b)). It can be seen that the boundary near the end of the AOSM's result, particularly the right end is going up abruptly. As described above, the lateral surface initial model searched

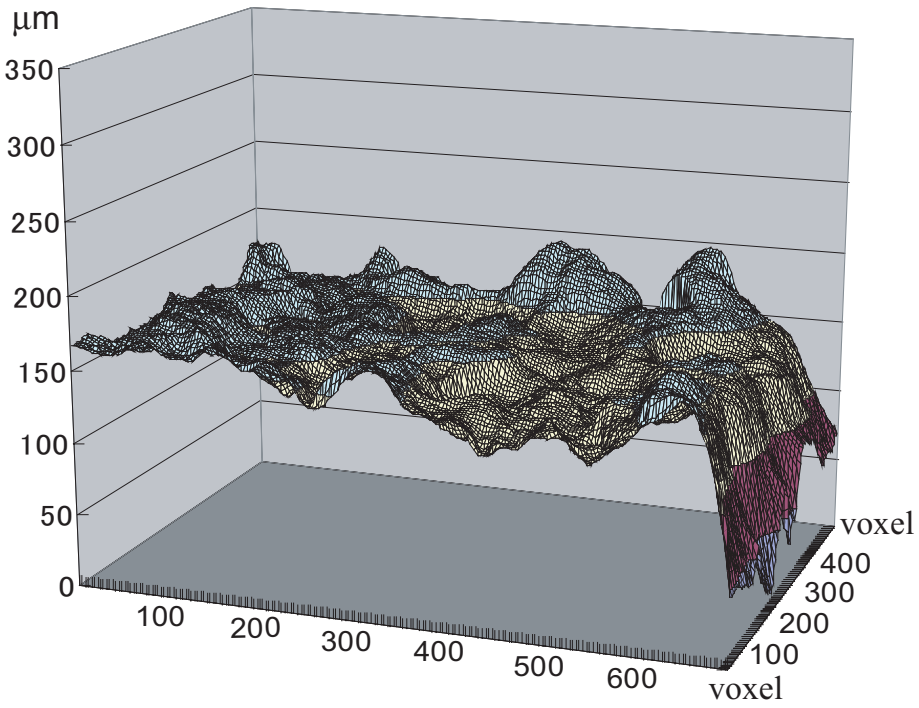
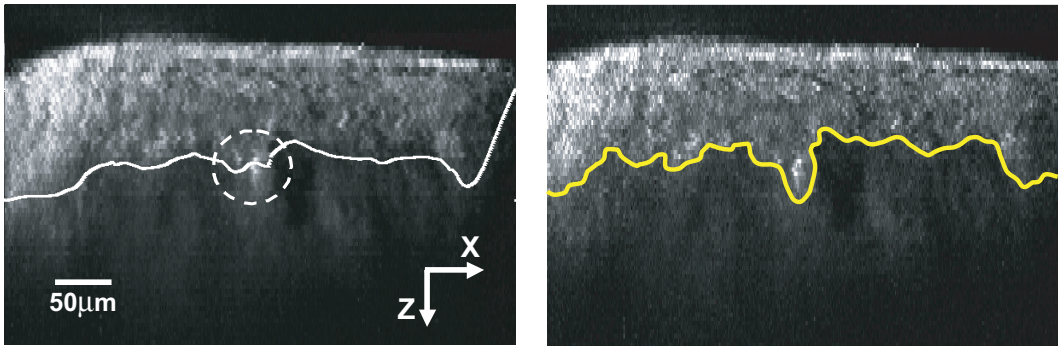


Figure 2.26: The distribution of epidermis thickness. The average thickness is about 110 μm.

the dermo-epidermal surface in lateral surface of volumetric data and the spatial position relation to central initial model is not concerned. Thus, when the contour part and the central part of the model are recomposed to make a integral surface model, the junction becomes unnatural.

Another distinct issue is concerning the shape of steepy edge element. The shape of the sur-



(a) (b)

Figure 2.27: Result comparison between AOSM ((a)) and human experts' observation ((b)).

face inside the broken circle should be more protuberant, whereas the AOSM gave a smooth result. To push the initial model approach object surface, the distance to a specified reference plane is generated to be the external constraint. However, the weak edge elements may be ignored by the identical external constraint, and the model will get over the object surface.

To overcome the enumerated issues, the modification on data structure of the model and definition of energy function are applied.

Chapter 3

Energy-modified Free-form Deformable Model Method

Although AOSM gives a reasonable solution for handling open form object surface, the losing in continuity at the junction of lateral surface object boundary and central part of object surface can be imaged. To address this disadvantage, an integral initial model is necessary. This will be achieved by modification on the energy function[84].

3.1 Modification on Internal Energy

The energy function of 3D ACM is composed of internal and external elements. The difficulty of open form surface is mainly concerning the internal energy. Because the boundary of open plane is not available for derivatives, the discrete derivatives based internal energy cannot be computed.

Generally, in an image the smooth edge corresponds to low frequency components, and the high frequency components correspond to unstable edges, e.g. noise. Here the image smooth technique can be used for reference. The image smooth makes use spatial smoothing filters to restrain the high frequency component in a gray-level image. The simplest smoothing approach is spatial averaging: A new image is constructed by calculating the average brightness in each small region of the images.

The coordinate of control point $\boldsymbol{v}$'s neighbors are added together and divided by the number of control points in neighborhood. In the next, the resulting coordinate $\boldsymbol{v}^{'}$ is used to update the coordinate of $\boldsymbol{v}$ ((3.1)).

$$\boldsymbol{v}' = \frac{1}{m}\sum_{i=1}^{m}\boldsymbol{v}_i \tag{3.1}$$

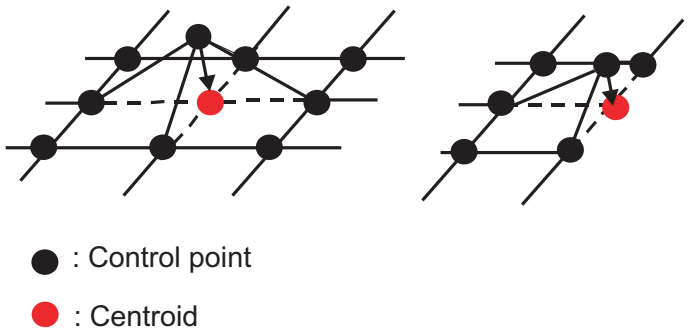
where i is the index of control points, m is the size of neighbor control points of $\boldsymbol{v}(x, y, z)$ (including $\boldsymbol{v}(x, y, z)$). Then the internal energy is defined as

$$E_{int}(\boldsymbol{v}) = w_{int}(\boldsymbol{v} - \boldsymbol{v}')^2 \tag{3.2}$$

By substituting (3.1),

$$
\begin{aligned}
&E_{int}(\boldsymbol{v}) = w_{int}(\boldsymbol{v} - \frac{1}{m}\sum_{i=1}^{m}\boldsymbol{v}_i)^2 \\
&= w_{int}[(x - \frac{1}{m}\sum_{i=1}^{m}x_i)^2 + (y - \frac{1}{m}\sum_{i=1}^{m}y_i)^2 + (z - \frac{1}{m}\sum_{i=1}^{m}z_i)^2] \\
&= \frac{1}{m}w_{int}[(mx - \sum_{i=1}^{m}x_i)^2 + (my - \sum_{i=1}^{m}y_i)^2 + (mz - \sum_{i=1}^{m}z_i)^2] \\
&= \frac{1}{m}w_{int}\{[(m-1)x - \sum_{i=1}^{m-1}x_i]^2 + [(m-1)y - \sum_{i=1}^{m-1}y_i]^2 + [(m-1)z - \sum_{i=1}^{m-1}z_i]^2\} \\
&= \frac{m-1}{m}w_{int}[(x - \frac{1}{m-1}\sum_{i=1}^{m-1}x_i)^2 + (y - \frac{1}{m-1}\sum_{i=1}^{m-1}y_i)^2 + (z - \frac{1}{m-1}\sum_{i=1}^{m-1}z_i)^2] \\
&= w_{int}[(x - \frac{1}{m-1}\sum_{i=1}^{m-1}x_i)^2 + (y - \frac{1}{m-1}\sum_{i=1}^{m-1}y_i)^2 + (z - \frac{1}{m-1}\sum_{i=1}^{m-1}z_i)^2] \qquad (3.3)
\end{aligned}
$$

where $w_{int} = \frac{m-1}{m}w_{int}$



(a) (b)

Figure 3.1: The surface smooth can be obtained by minimizing the distance between each control point and the centroid of its neighbor control points ((a)). The control points on the contour of the open plane can also be handled in this way ((b)).

Thus, the internal energy can be calculated by the sum of differences between $\boldsymbol{v}_i$ and its neighbor control points' centroid (Figure 3.1 (a)). Then the internal energy is defined as:

$$
\begin{aligned}
E_{int}(\boldsymbol{v}_i) &= w_{int}(\boldsymbol{v}_i - \boldsymbol{v}_i^g)^2 \\
&= w_{int}[(x_i - x_i^g)^2 + (y_i - y_i^g)^2 + (z_i - z_i^g)^2] \qquad (3.4)
\end{aligned}
$$

where (x_i^g, y_i^g, z_i^g) is the centroid of $\boldsymbol{v}_i$'s neighbor control points.

Furthermore, the control points on the contour of the open plane can also be handled (Figure 3.1 (b)) without any specific processing like the AOSM.

3.2 Triangle Patch Based Initial Model

The conventional FDMM creates the initial model with control points on orthogonal lattice. This favors the computation of discrete differential-based internal energy. However, in the last section, the centroid of neighbor control points is used for computing the internal energy, the orthogonal distribution of control points is not necessary.

In many clinical situations, particularly in the computer environment, the 3D surfaces are usually composed of triangle patches. Figure 3.2 illustrates a triangle patches made surface. The control points can be created on the vertices of the triangle patches, and the neighbor relation among control points are determined by the sides of triangle patches. Because there is no orthogonal constraint, this definition of control points may enhance the model's generality.

In this work, a triangle patches based initial model is adopted. The internal energy is computed easily using the triangle relationship between control points. For instance, in Figure 3.3, $\boldsymbol{P}_0$ is connecting the control points $\boldsymbol{P}_1, \boldsymbol{P}_2, ..., \boldsymbol{P}_6$. To minimizing the distance between $\boldsymbol{P}_0$ and $\boldsymbol{G}$, the centroid of $\boldsymbol{P}_1, \boldsymbol{P}_2, ..., \boldsymbol{P}_6$, the surface will become smooth and flat.

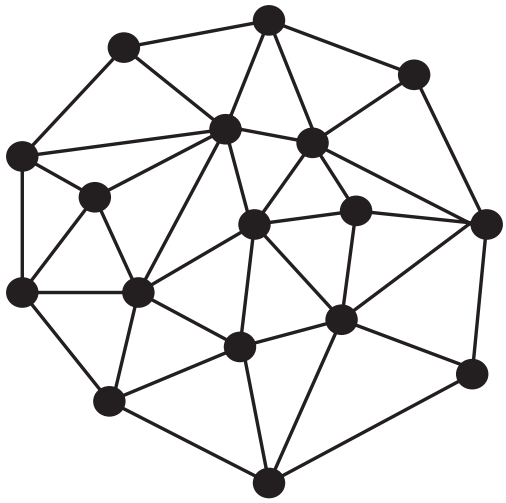


Figure 3.2: Triangle patches based surface.

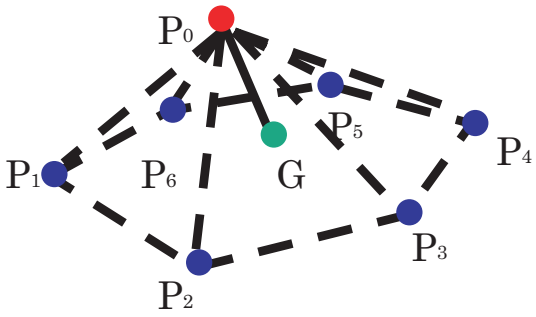


Figure 3.3: $\boldsymbol{P}_0$ is connecting the control points $\boldsymbol{P}_1, \boldsymbol{P}_2, ..., \boldsymbol{P}_6$. To minimize the distance between $\boldsymbol{P}_0$ and $\boldsymbol{G}$, the centroid of $\boldsymbol{P}_1, \boldsymbol{P}_2, ..., \boldsymbol{P}_6$, the surface will become smooth and flat.

3.3 Novel External Constraint

As that mentioned in section 2.3.3, the external constraint can push initial model towards expected boundary elements. However, the necessary of user-specified reference plane (or point) weakens the automation of this method. Although as described in Chapter 1, strategies of external energy, such as balloon, distance potential and gradient vector flow, have been exploited, they also have their own disadvantages. To integrate their opinion, a novel external constraint that makes the initial model has a large attraction range is proposed.

In skin inner tissue surface reconstruction application, to reconstruct the 3D model of dermo-epidermal surface from volumetric data, an open plane is created to be initial model. Although the exact position of dermo-epidermal surface is not known, it is corresponding to the bottom of the high intensity region in confocal microscopy images (Figure 2.20).

Differing from the conventional external constraint, when the initial model is created in the bottom of volumetric data, a “bottom-up” scan is implemented to search the closest edge element for each control point along the normal direction to initial model.

In (3.5) the value of the external constraint of each control point is the distance ($D(\boldsymbol{v}_i)$) between control point $\boldsymbol{v}_i$ and closest edge element $\boldsymbol{u}_i$ (Figure 3.4). However, the search may be failing, if there are boundary gaps along this direction. In the failure cases, the external constraint is set “0”, whereas the invalid point are still dynamic, for the effect of neighbor

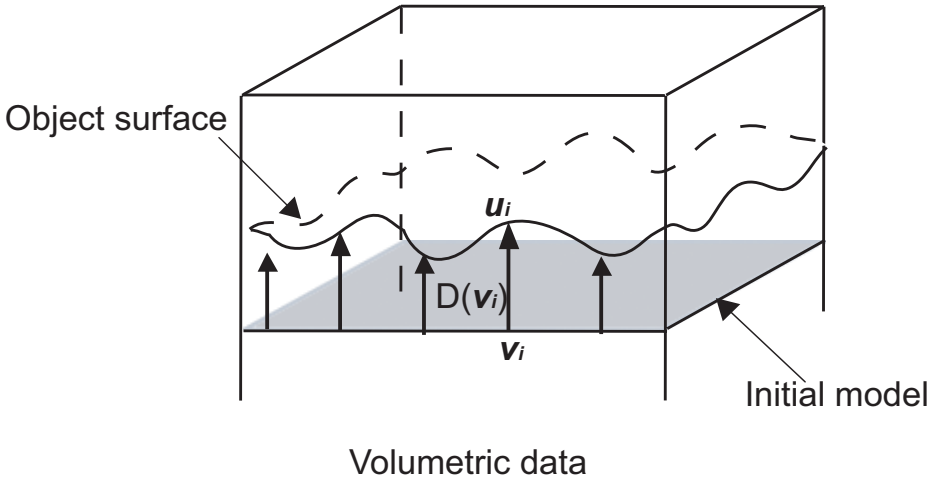


Figure 3.4: The initial model is created in the bottom of volumetric data and restrict the search direction as "bottom-up" along the normal direction to initial model. The minimization of the distance $D(v_i)$ between control point v_i and the closest boundary point u_i is the proposed external constraint.

control points.

$$D(\boldsymbol{v}_i) = w_{ext}(\boldsymbol{v}_i - \boldsymbol{u}_i) \tag{3.5}$$

Conversely, if the initial model is a closed surface, the search direction points to the geometry center of the initial model.

Including the external constraint $D(\boldsymbol{v}_i)$, the energy function is transformed as

$$\begin{aligned} E &= \sum_i [E_{int}(\boldsymbol{v}_i) + E_{ext}(\boldsymbol{v}_i)] \\ &= \sum_i [w_{int}(\boldsymbol{v}_i - \boldsymbol{v}_i^g)^2 + w_{image}|\nabla^2 G * I(\boldsymbol{v}_i)|^2 + w_{ext} D(\boldsymbol{v}_i)] \end{aligned} \tag{3.6}$$

In the experiments the weighting parameters of this energy function is also determined by users' experience like AOSM (section 2.3.4).

Since the proposed method in this chapter is similar to the ACM in spirit and the modifications are mostly concerning the energy function, the proposed method is called "energy modified active contour model" (EACM)

3.4 Numerical Implementation with Levenberg-Marquardt Algorithm

Various numerical implementations of deformable models have been reported. For instance, the finite difference method[38], greedy algorithm[90] and dynamic programming that is applied in ASOM. With these methods, only the knowledge of the functions at discrete points of a subdivision is known. Therefore, the distance between successive points must be made very small to achieve sufficient precision so as not to lose too much information. Because the initial model of EACM is an integral continuous formula, Levenberg-Marquardt algorithm (LM) is adopted to find the minimum of energy function.

LM is a kind of Quasi-Newton method. It is an integration of Steepest-Descent method (SD) and Gauss-Newton (GN) method. SD and GN are popular optimization methods. SD is the most straightforward method in optimization. By computing the gradient direction followed by a 1D search, SD iteratively approaches the minimum point of the object function in parameter space. Since only the first order derivative information is used, SD suffers from the slow convergence. However, it is relatively robust even the initial guess is far away from the true value.

GN goes one step further than SD: the second order derivative term is included to compute the update. GN converges faster than SD. The price to pay is the reduction in robustness, i.e. it is much more sensitive to initial guess than SD.

LM is a hybrid technique of SD and GN. A steering factor λ is introduced to switch between GN direction and SD direction. When $\lambda \rightarrow 0$, LM approaches SD method. The values of λ during the iterative process are chosen in the following way: at the beginning of the iterations, λ is set to a large value, so the LM manifests the robustness of SD and the initial guess can be chosen with less caution. In each iteration, if minimum is achieved, decrease λ by certain amount to speed up the convergence; otherwise, increase λ value to enlarge the searching area.

With LM, we work with continuous functions, whatever the size of the grid. Therefore, the function under consideration is known everywhere in the image data, independently of the chosen discretization. This yields a lower algorithmic complexity and better numerical stability in our applications. The detail of LM is described in Appendix B.

3.5 Experiments

Although EACM is used to address the skin inner tissue surface reconstruction, the modification on energy function and data structure of initial model broaden the application range of EACM, EACM is not only applied to establish 3D model of skin inner tissue surface, but also used for the foot shape recovery application. The input of foot shape recovery is multiple camera images. However, EACM needs volumetric data generally. In experiments, through the computation of normalized correlation between corresponding points of each voxel on multiple camera images, a 3D intensity space like the volumetric data of confocal microscopy images are established. The detail is described in section 3.5.2 with the experimental results.

3.5.1 Skin Inner Tissue Surface Reconstruction

First, the proposed EACM is implemented to the skin inner tissue surface reconstruction. The input is the same as that has described in section 2.4. The processing flow is illustrated in Figure 3.5.

3.5.1.1 Reconstruction of Dermo-epidermal Surface

Figure 3.6 depicts the dermo-epidermal surface detected by EACM. To compare it with Figure 2.23, it is obvious that EACM's result has better continuity.

The reconstructed 3D model of dermo-epidermal surface is illustrated in Figure 3.7. Because the initial model of EACM is always considered to be integral and it is adjusted in a systematic behavior, thus its reconstructed model is more coherent than the AOSM reconstructed 3D model (Figure 2.22).

3.5.1.2 Upper Epidermal Surface Reconstruction and Quantitative Analysis of Epidermis Thickness

Figure 3.8 and Figure 3.9 illustrate the reconstructed upper surface of epidermis in 2D and 3D respectively. Because the background around the upper epidermal surface is relatively simple, and the shape at the surface boundary does not change tempestuously, both the

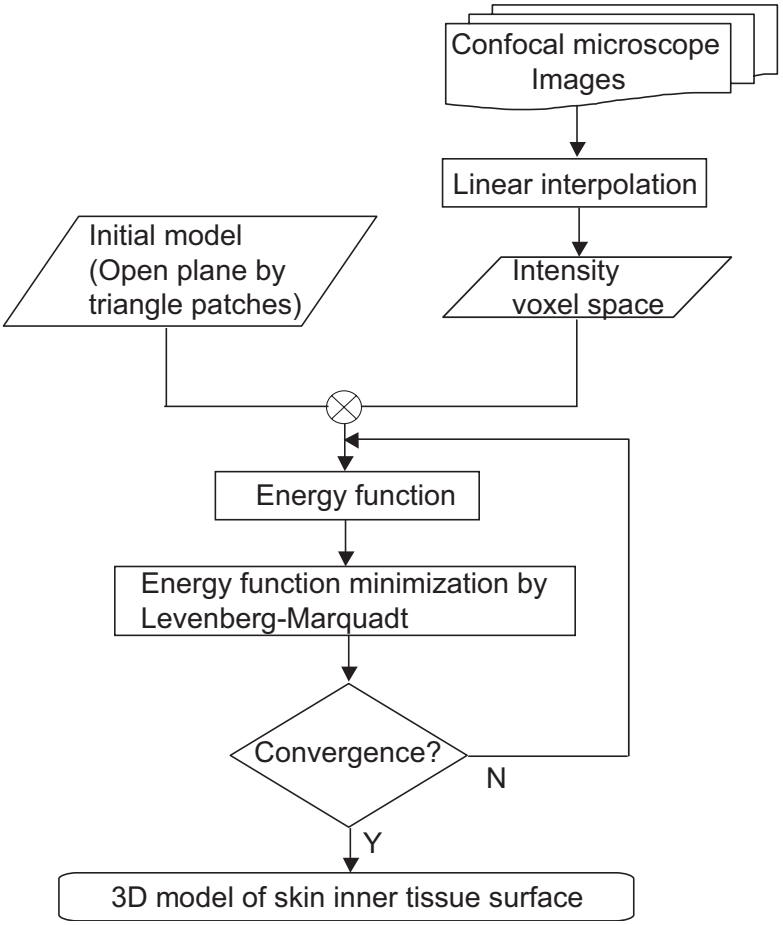


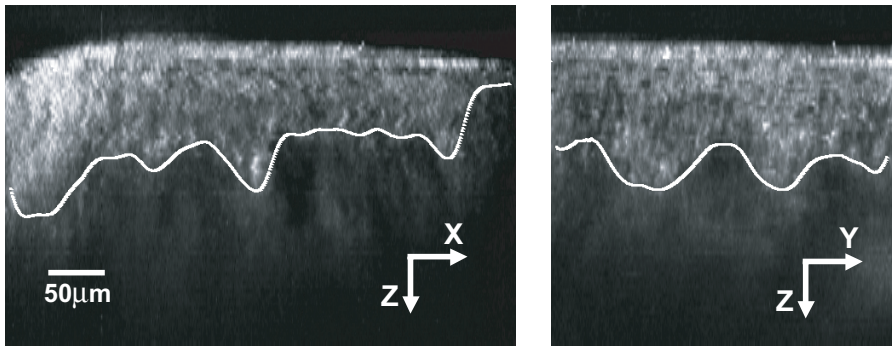
Figure 3.5: Flow chart of the EACM-based skin tissue intersection surface reconstruction system.

AOSM and EACM give reasonable results. The epidermal thickness is also estimated by calculating the average Euclidean distance of corresponding points on dermo-epidermal surface and upper epidermal surface. The distribution of epidermis' thickness is illustrated in Figure 3.10. The average thickness of epidermis is about 91μm. According to experts' experience, EACM's result is more reliable than AOSM's (about 110μm).

3.5.1.3 Discussion

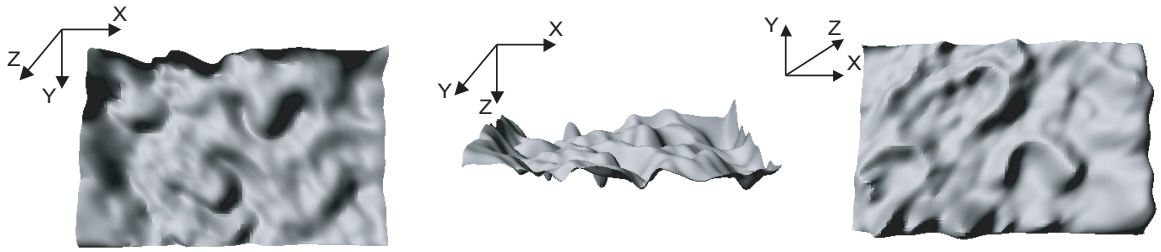
To validate the proposed method, comparison with human experts' observation and AOSM is implemented (Figure 3.11).

According to this comparison, EACM offers more reasonable result (Figure 3.11 (b)) to AOSM's result. However, there are still errors between EACM and human experts' obser-



(a) (b)

Figure 3.6: Reconstructed dermo-epidermal surface is plotted in white in vertical sections. (a) XZ-axis view, (b) YZ-axis view.



(a) (b) (c)

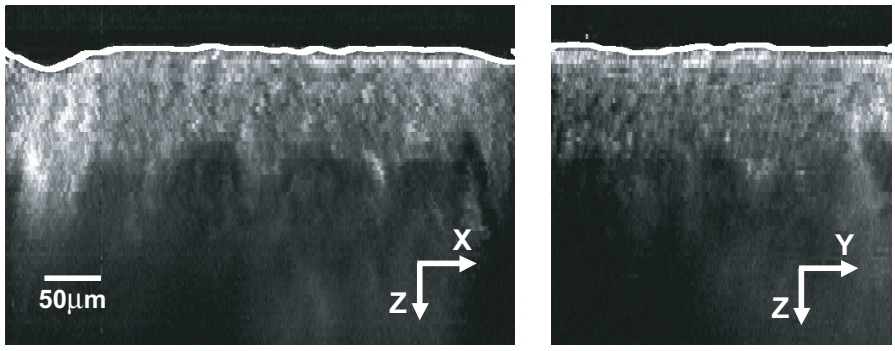
Figure 3.7: Reconstructed 3D dermo-epidermal surface, displayed in top (a), side (b), down (c).

vation. The errors are occurred mostly because of experts do not consider 3D context the volumetric data. Therefore, we believe that the surface obtained by EACM is even more reasonable than human experts' observation.

Table 3.1: The weighting parameters combinations of internal energy and image energy

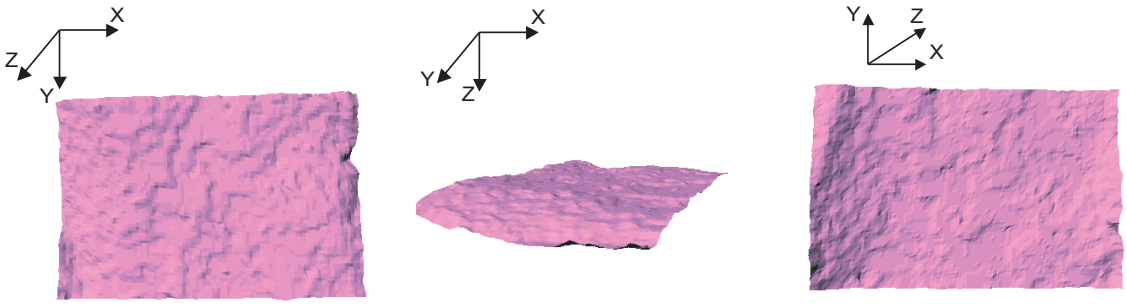
	w_{int}	w_{image}
(a)	0.8	0.2
(b)	0.6	0.4
(c)	0.4	0.6
(d)	0.2	0.8

Although the proposed method can be considered to be automatic, as described above the weighting parameters of the energy function are specified by experience. There are three



(a) (b)

Figure 3.8: Reconstructed upper epidermal surface is plotted in white in vertical sections. (a) XZ-axis view, (b) YZ-axis view.



(a) (b) (c)

Figure 3.9: Reconstructed 3D upper epidermal surface, displayed in top (a), side (b), down (c).

weighting parameters in energy function ((3.6)), whereas, the degree of freedom is two. In general the weighing parameter of external constraint is set "1". The weighting parameters of internal energy and image energy are free. An experiment is implemented to exam the possible combination of weighting parameters. Table 3.1 shows the examined items. Figure 3.12 illustrates the reconstructed 3D dermo-epidermal surface models that corresponds to Table 3.1. When the weighting parameter of internal energy is enlarged, the result is smooth and flat. Conversely, when the weighting parameter of image energy is enlarged, the undulation of surface shape is clear, whereas the surface is rough. These results are well in agreement with the theoretical consideration on the effect of changing weighting parameters. Depending on this investigation and experts' experience, the combination of $w_{int} = 0.6, w_{image} = 0.4$ is adopted in this experiment.

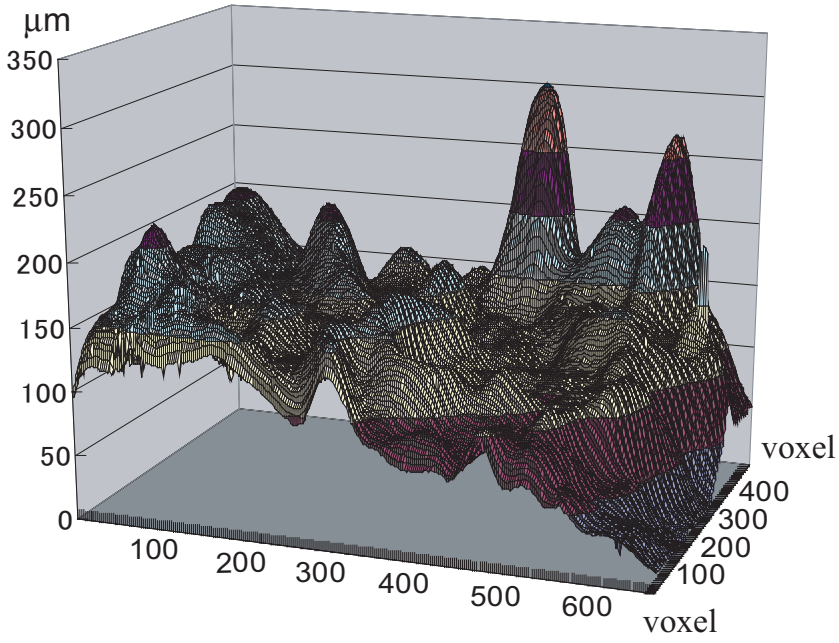
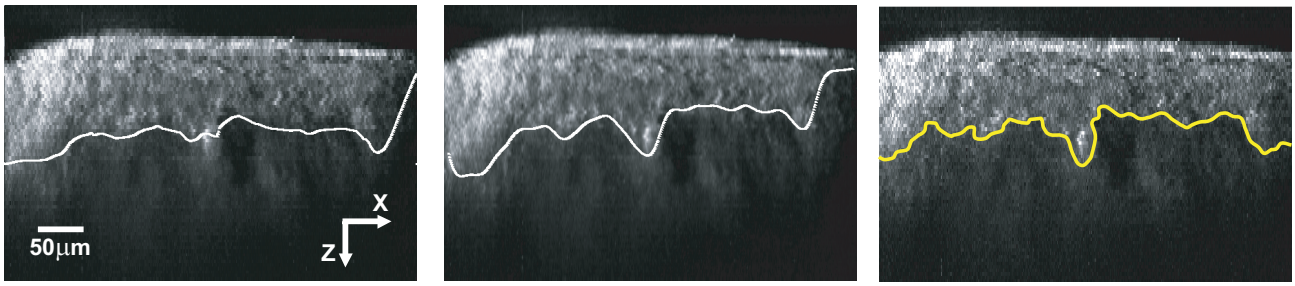


Figure 3.10: The distribution of epidermis thickness. The average thickness is about 91 μm.



(a) (b) (c)

Figure 3.11: Comparison among the AOSM ((a)), EACM ((b)) established dermo-epidermal surface and human experts' observation ((c)).

3.5.2 Human Foot Shape Reconstruction

In the experiment of skin inner tissue surface reconstruction, the open form surface reconstruction issue is well solved by EACM. Spontaneously people would wonder whether this approach is used to handle general closed object surface. Because the initial model of EACM is integral (not similar to the AOSM, which partitions the initial model into lateral and central parts.), the initial model of EACM can also be closed surface. In the next, the

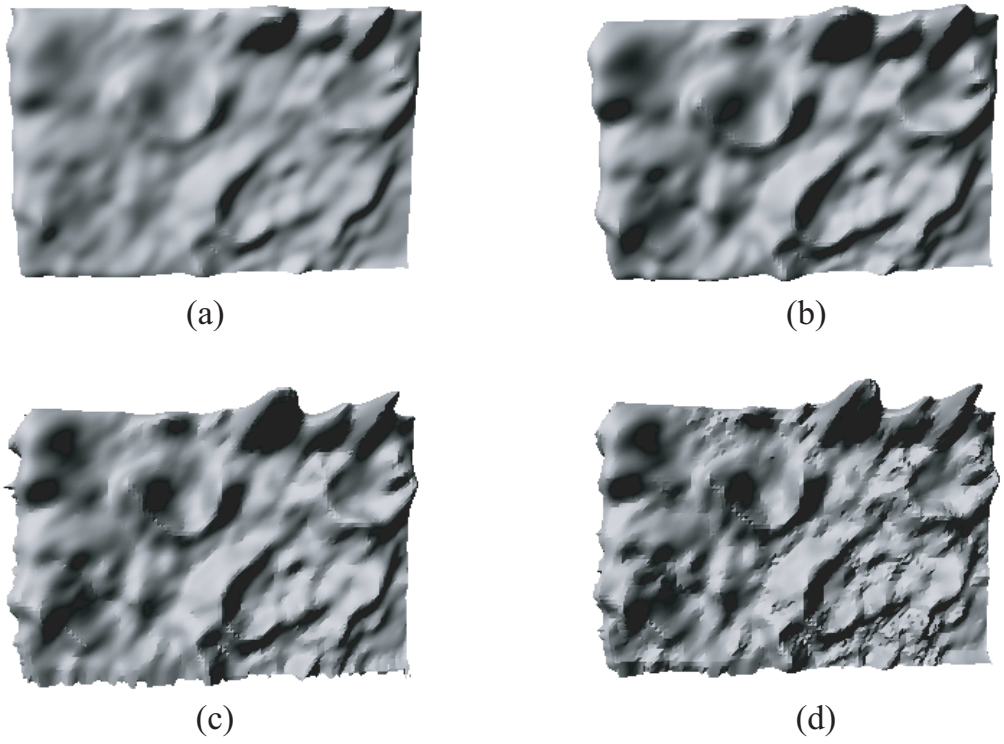


Figure 3.12: When the weighting parameter of internal energy is enlarged, the result is smooth and flat. Conversely, when the weighting parameter of image energy is enlarged, the undulation of surface shape is clear, whereas the surface is rough. In this experiment the combination of $w_{int} = 0.6, w_{image} = 0.4$ is adopted.

EACM is applied to deal with closed object surface- human foot. This application is pertains to anthropometry.

As a research subject relating ourselves nearly, anthropometry has been widely used in criminological, medical applications or selective trial of people[62]. Anthropometry also acts an important part in the design of clothes and shoes, which need to be fit to the human body very much. To create fitting shoes, the focus of shoes design is always on the measurement of biomechanical behavior of a lot of human data, e.g. images. However, to process such a large amount of data is an arduous task. Fortunately, modern computer makes it possible to do accurate measurement automatically in even 3D by computer vision and image processing techniques[39]. Now a day, some 3D foot scanners have been commercially available.

Although almost all these scanners can generate accurate 3D foot model, because of the

measurement space is fixed and the position constraint of cameras is strict, dynamic behavior of foot cannot be handled. However, feet are our motor organ, the dynamic behavior of them is the most important factor in practice. As the first step, to reconstruct foot shape in a relative free space is a significant work.

Triangulation-based structured light method, time-of-flight range finder and multiple view-based technique are general used object surface acquisition tools[56][96]. The precision of the measurement and the simplicity of the use have made structured light and time-of-flight range finder the most widely used systems for surface measurement[29]. However, depending on the size and resolution of the surface to measure, the acquisition time can range from seconds to half minute. This fact requires the object human body should be stable without motion during the measurement, so it is difficult to measure the foot shape in dynamically moving situation. Even the acquisition of still images is difficult, because making the foot retain a pose for half minute is almost impossible.

On the other hand, multiple views based method, e.g. volumetric intersection, makes use of multiple camera system to acquire a set of images around object. Silhouettes are extracted from these images, then combined to result in a 3D model[87][95]. However, because this method cannot recover concave surface, high accuracy is difficult to be achieved.

Although accurate result is difficult to be obtained, the image acquisition space of volumetric intersection method is relative free, and even dynamic behavior of object can be captured for multiple camera system's high acquisition speed. Thus, the multiple camera system is just according with our purpose. In order to enhance the accuracy of result, EACM is employed in this experiment. The EACM-based foot reconstruction method is implemented in an intensity voxel space. The value of voxel is estimated by calculating the normalized correlation of its correspondence pixels in multiple camera images.

Different from skin tissue, the shape of foot has a definite concept in our consciousness, and the measurement of foot shape is relatively easier. As the results of conventional foot shape measurement researches, databases of foot shape were established. In this experiment, a foot shape database is introduced as a prior knowledge to improve the performance of EACM[80][85]. The processing flow of foot shape reconstruction by EACM is illustrated in Figure 3.13.

The experiment is carried out in the computer simulative. Multiple camera system is created by Povray[61]. The term multiple camera images refer to images acquired from different positions in the space describing the same scene. As described above, this experiments is im-

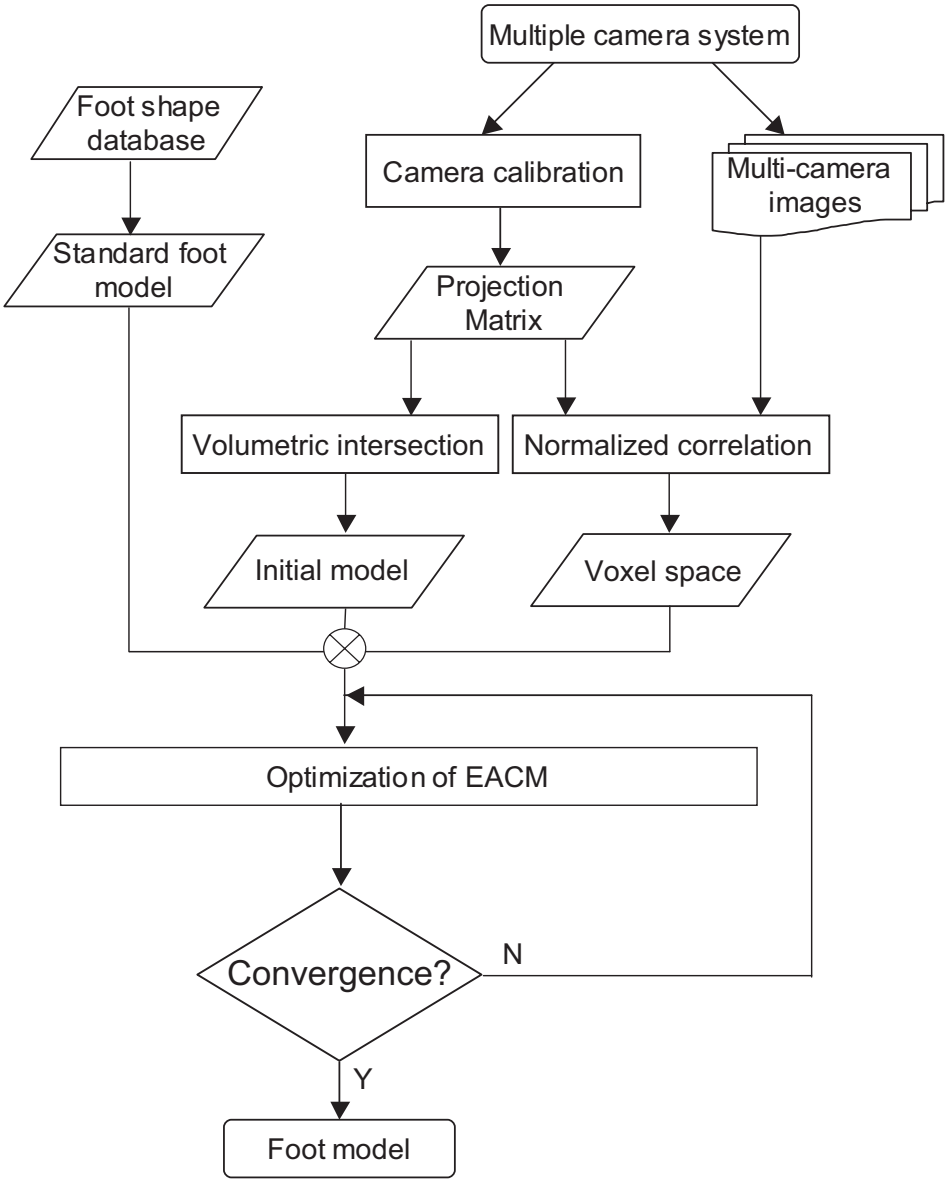


Figure 3.13: The flow chart of 3D foot shape reconstruction by EACM.

plemented in an intensity voxel space. To establish the normalized correlation based-intensity space, the camera geometry should be obtained first. Because this is a computer simulative experiment, the camera parameters are known, when the system is established.

From the next, a general multiple views-based shape reconstruction technique: volumetric intersection algorithm will be explained. Its result is used as the initial model of EACM.

3.5.2.1 Volumetric Intersection Method

To generate 3D models from multiple camera images, various techniques have been proposed, e.g. volumetric intersection method. The multiple camera images are processed to extract the silhouettes, then combined to result in a 3D model[44][87][95]. For illustration purpose, an example in 2D is given in Figure 3.14, in which the object is constructed by intersecting the 2D visual wedges from every view point.

Although easy to implement, volumetric intersection has its own limitations. Shapes estimated by volumetric intersection are coarse, if the distance between camera centers is too large. In Figure 3.14, there is distinct error in the top-right contour, for the distance between view points C_2 and C_3 is too large. This means that it is difficult to obtain detailed shapes, unless a large number of cameras are used. However, the concavities on an object cannot be recovered even a large number of cameras are adopted.

Figure 3.15 shows the multiple camera images of a foot that is rendered by Poveray. The corresponding silhouette images are illustrated in Figure 3.16 (3 pieces are illustrated). The reconstructed 3D model of object is computed volumetrically by carving away all voxels outside the silhouette area. A set of triangular polygons is generated from this voxel domain by Matching Cube algorithm[47]. The 3D model is shown in Figure 3.17. According to this result, the reconstructed 3D model is very coarse i.e. this model is only a rough approximation of the actual object shape. However, the position of reconstructed result is near the object. Thus, the volumetric intersection made model is used to be the initial model for EACM.

3.5.2.2 Normalized Correlation Based Intensity Voxel Space

In general, FDMM-based methods evaluate the adequacy of current model by surface features in image plane (2D) or volumetric data (3D). In this experiment an intensity voxel space is established by calculating normalized correlation of matching points of each voxel on multiple camera images. The relation between multiple camera images and voxel space can be described as follows: there is a object in this 3D voxel space, the object is projected to multiple camera images. Since the camera parameters of all the cameras in the simulative multiple camera system can be estimated easily, the projection matrix of each camera is established. Then the voxel on the object surface is described by extracted matching points of this voxel

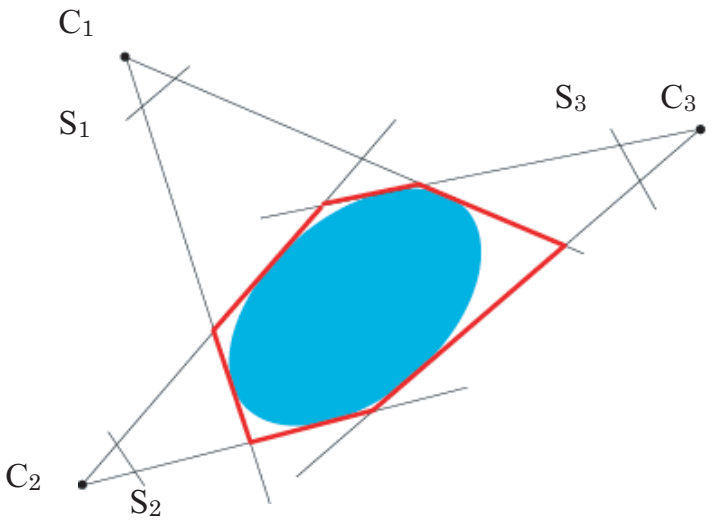


Figure 3.14: Diagram of reconstructing object surface with silhouettes S_k from view points C_k.

in each pair of the images (Figure 3.18). The normalized correlation may be a reasonable evaluation of reliability of voxels. Normalized correlation is calculated as C in (3.7).

$$
\begin{aligned}
C &= \frac{\sum_{i=-m}^{m}\sum_{j=-m}^{m} I_1'(u_{k1}v_{k1})I_2'(u_{k2}v_{k2})}{(2m+1)^2\sqrt{\sigma^2(I_1)\sigma^2(I_2)}} \\
I_k'(u_k,v_k) &= I_k(u_k+i,v_k+j) - \bar{I}_k(u_k,v_k) \\
\bar{I}(u_k,v_k) &= \sum_{i=-m}^{m}\sum_{j=-m}^{m}\frac{I_k(u_k+i,v_k+j)}{(2m+1)^2}
\end{aligned}
\tag{3.7}
$$

where I_k (k=1,2) is the intensity of the matching points, and $\bar{I}_k$ is the average of intensities of pixels in the window for correlation, which is sized $(2m+1)\times(2m+1)$. In this experiment, the practical size of match window is 15×15 voxels centering on (u_{k1}, v_{k1}) and (u_{k2}, v_{k2}). I_k' is the subtraction between I_k and $\bar{I}_k$ in the window. $\sigma(I_k)$ is the standard deviation of the intensity in the window.

In theory, since a voxel corresponds to all of the matching points in image plane, it seems to be reasonable to adopt the average of all correlation values between all the pairs of multiple camera images to be the evaluation of each voxel. However, the occlusion issue has to be cared. When a voxel is in occlusion, the occluded voxel corresponds to matching points in only a part of pairs. In such case, the other voxels must exist between that voxel and one of the cameras. Generally, the correlation value between correct matching points is higher

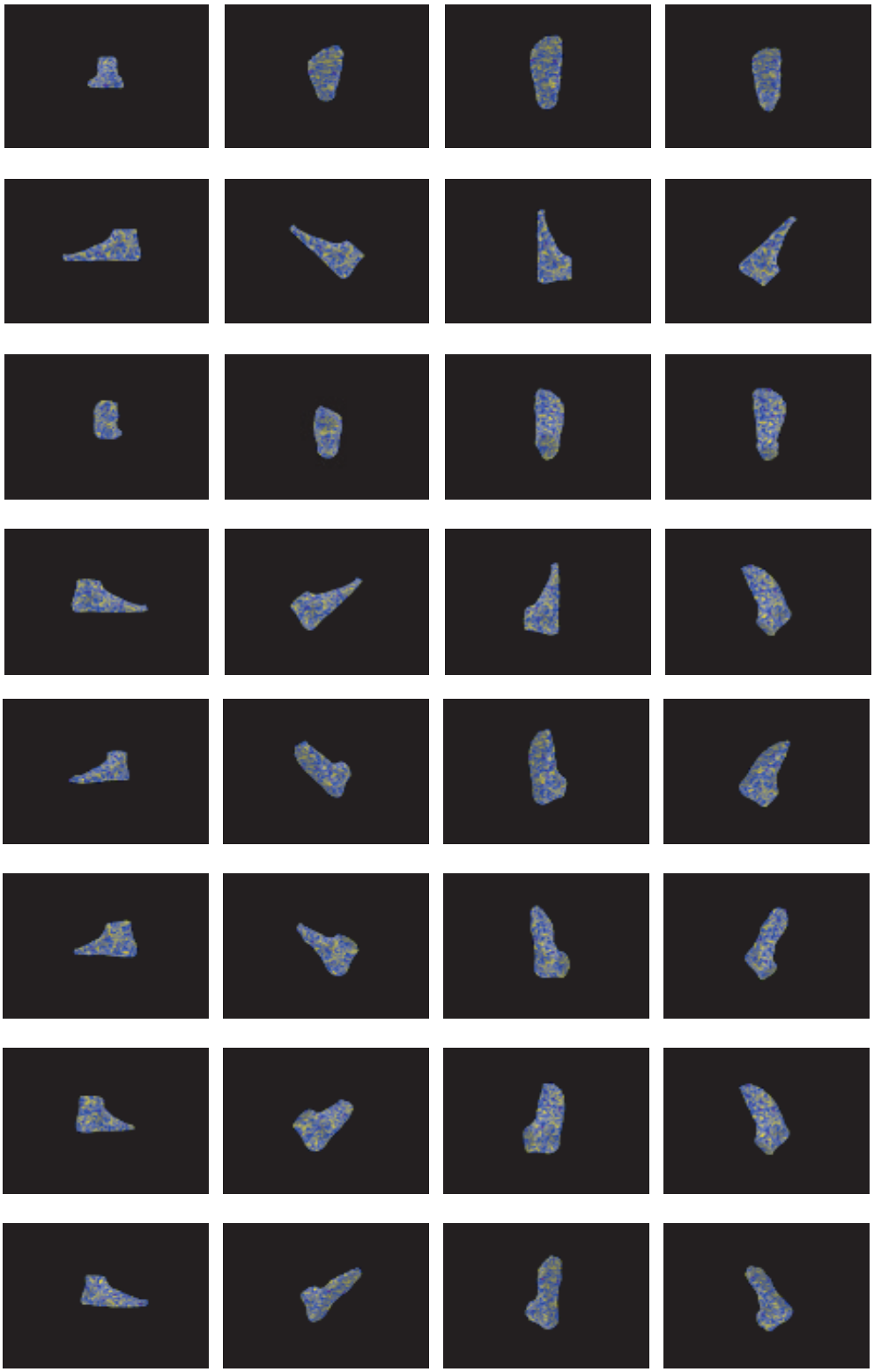


Figure 3.15: 32 multiple camera images of a CG foot model with random texture.

Figure 3.16: Silhouette images of the multiple CG foot model.

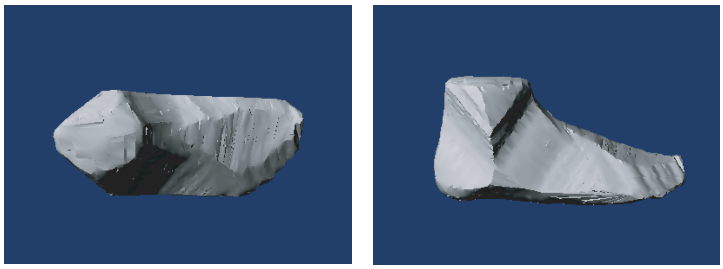


Figure 3.17: Reconstructed 3D model of the CG foot model by volumetric intersection.

than the correlation value between mismatching points. Thus, the maximum of all correspondent correlation values is adopted as the voxel value. Figure 3.19 illustrates some traverses of established volumetric data. It looks like the confocal microscopy images, whereas the explanations of the intensity are different. In the confocal microscopy volumetric data, the boundary of high intensity region and low intensity region corresponds to the object surface.

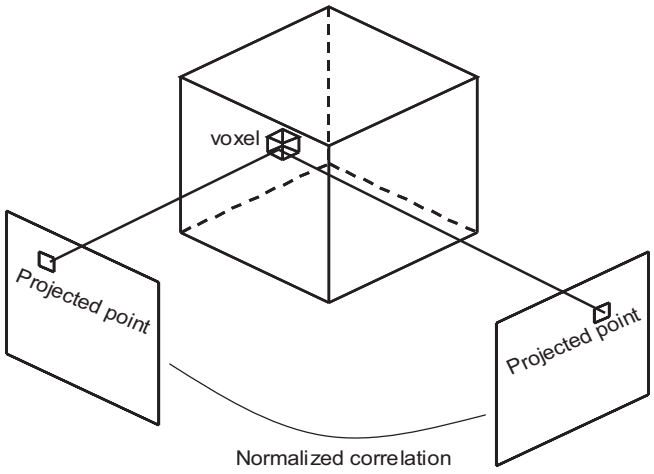


Figure 3.18: Normalized correlation between multiple camera image pair.

Thus, the object surface can be detected by searching the zero crossing of the second order intensity derivative. On the other hand, in the normalized correlation-based volumetric data, the high the voxel intensity is, the voxel is closer to the object surface. Thus, the intensity of each voxel is reasonable image energy.

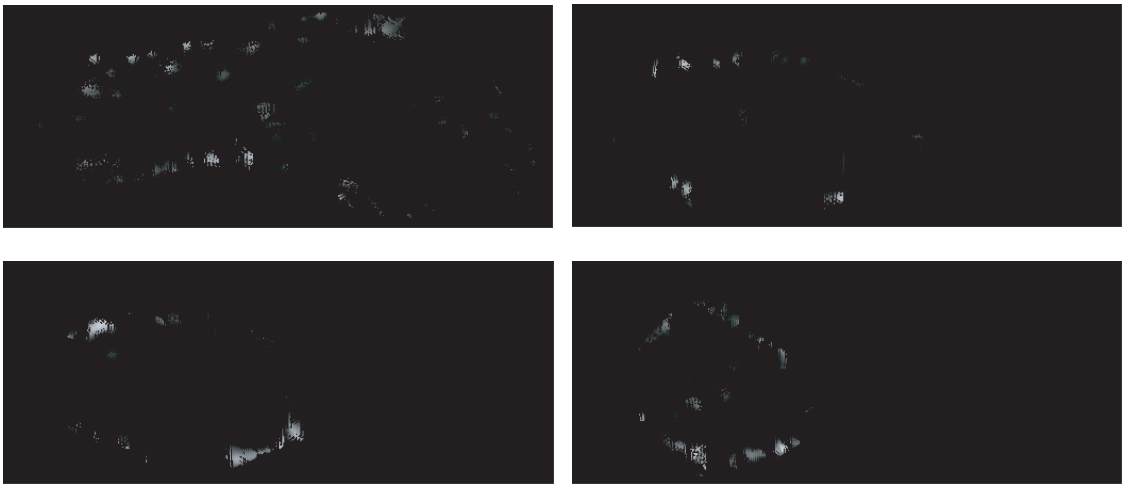


Figure 3.19: Traverses of normalized correlation based volumetric data. The high intensity voxels correspond to the surface of foot.

3.5.2.3 Result by EACM Method

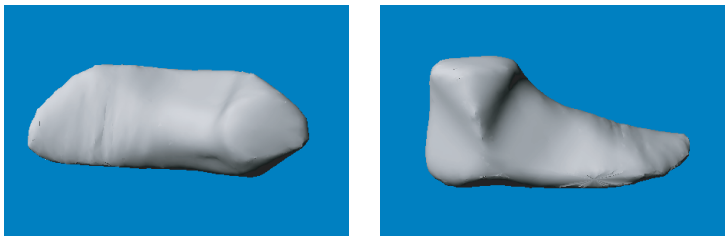


Figure 3.20: Reconstructed 3D model of the CG foot model by EACM.

The reconstructed foot model by EACM is shown in Figure 3.20. Although a smooth 3D model is obtained, the whole model surface is very stiff and gives a unnatural feeling. This is mostly because of the doubtful reliability of normalized correlation based voxel space. If the object surface has circular pattern texture, the normalized correlation will give error evaluation of the voxel. In this experiment the CG foot model has random pattern texture,

hence the circular pattern may occur. Moreover, the normalized correlation method implicitly assumes that the observed surface is parallel to the multiple image planes. However, this condition isn't always satisfied. Thus, miss matching will be caused. Furthermore, because there is no shape constraint, the shape features of object are getting vanished by the smooth term of energy function.

3.5.2.4 Refination with Standard Model

The forementioned difficulties are occurred mainly because the model is pushed towards the locations of the strongest local image features and not consider true anatomical boundary locations. This issue can be addressed by incorporating prior shape constraint derived from amount of shape instance about the same class of object.

Table 3.2: The detail of foot shape database

Size (U.S. size)	5	6	7	8	9	10
Male	-	17	56	68	48	23
Female	11	20	78	42	24	10

A human foot database (training set) includes 397 adults' right foot is adopted as a prior knowledge. The foot models are established by measuring the surface position information with a foot scanner (Figure 3.21). The scanner is developed by Digital Human Research Center of The National Institute of Advanced Industrial Science and Technology[35]. Table 3.2 shows the distribution of size and sex of foot models in the training set.

372 experts selected human foot's anatomical landmarks (sample points) are marked on each foot sample by special markers before the scanning. 3D foot shape models are established based on these sample points. Because the sample points of every foot model are corresponding to the same anatomical landmarks, the sample points are considered have been aligned.

A standard foot model is established by the means of calculating the mean position of each sample points ((3.8)).

$$\bar{\boldsymbol{v}}_i = \frac{1}{n}\sum_{j=1}^{n}\boldsymbol{v}_j \quad (i = 1 \cdots 372, n = 397) \tag{3.8}$$

Figure 3.21: The foot models in database are established by measuring the surface position information with a foot scanner[35].

where $\bar{\boldsymbol{v}}_i$ is a vertex of standard model, $\boldsymbol{v}_i$ is the coordinate vector of a sample point.

To minimize the Euclidean distance between each sample point $\boldsymbol{v}_i$ on foot model and the closest sample points $\boldsymbol{v}'(x_i, y_i, z_i)$ on the standard model, the foot model will be prevented from lapsing the basic foot shape. It is called standard model constraint. The standard model constraint is defined as

$$E_s(\boldsymbol{v}_i) = w_s w_v^i \left[(x_i - x_i^t)^2 + (y_i - y_i^t)^2 + (z_i - z_i^t)^2\right] \tag{3.9}$$

where w_v^i is determined by the reciprocal standard variation of each sample point of instances in training set. w_s is weighting parameter for the standard model constraint. In addition, the position of initial model and standard foot model are registered by superposing their geometric centroid. It is assumed that the sole of them are paralleling the XY plane of right hand coordinate system.

Including this constraint, the energy function of EACM becomes:

$$E = \sum_{i=1}^{n} \left[E_{int}(\boldsymbol{v}_i) + E_{ext}(\boldsymbol{v}_i) + E_s(\boldsymbol{v}_i)\right] \tag{3.10}$$

where E_{int} is the internal energy, E_{ext} is the external energy that composed of image energy and external constraint. For the detail of the external energy function, see (3.4) and (3.6)

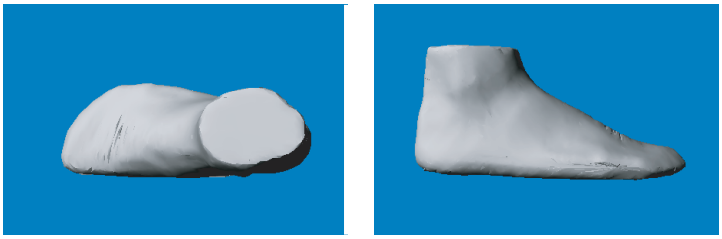


Figure 3.22: A reliable foot model is obtained, using the standard model constraint.

The reconstructed 3D model with standard model constraint is illustrated in Figure 3.22. Apparently, this foot model is more similar to the real foot in shape. Because the shape information of input CG foot model is known previously, the root-mean-square-error (RMSE) of result model surface and input foot model surface is calculated as a quantitative evaluation. The RMSE of volumetric intersection method is about 7.523 voxels. Since 1 voxel is equivalent to about 0.59mm the error of RMSE of volumetric intersection in "mm" unit is about 4.44mm. On the other hand, the RMSE of EACM with standard model constraint is 5.943×0.59=3.50mm.

3.5.2.5 Discussion

In section 3.5.1.3, investigation of energy function's weighting parameters is described. In the foot shape reconstruction experiment, a shape constraint is derived from foot shape database, we are very interested in investigating the reconstructed 3D models with different combinations of internal energy and the standard model constraint's weighting parameters. The detail of comparison is illustrated in Figure 3.23. From down to top, the weighting parameter of internal energy is becoming large and the model surface becomes smooth. On the other hand, from left to right the weight of standard model constraint is becoming large, the shape is getting similar to the standard foot model. The value under each model is the RMSE in voxel unit. This investigation illustrated that the optimal combination of energy function's weighting parameters can improve the performance of the reconstruction method.

Although the standard model provides normal criterion of object shape, it describes the common foot shape features only in a certain extend and cannot reflect the delicate difference between foot samples.

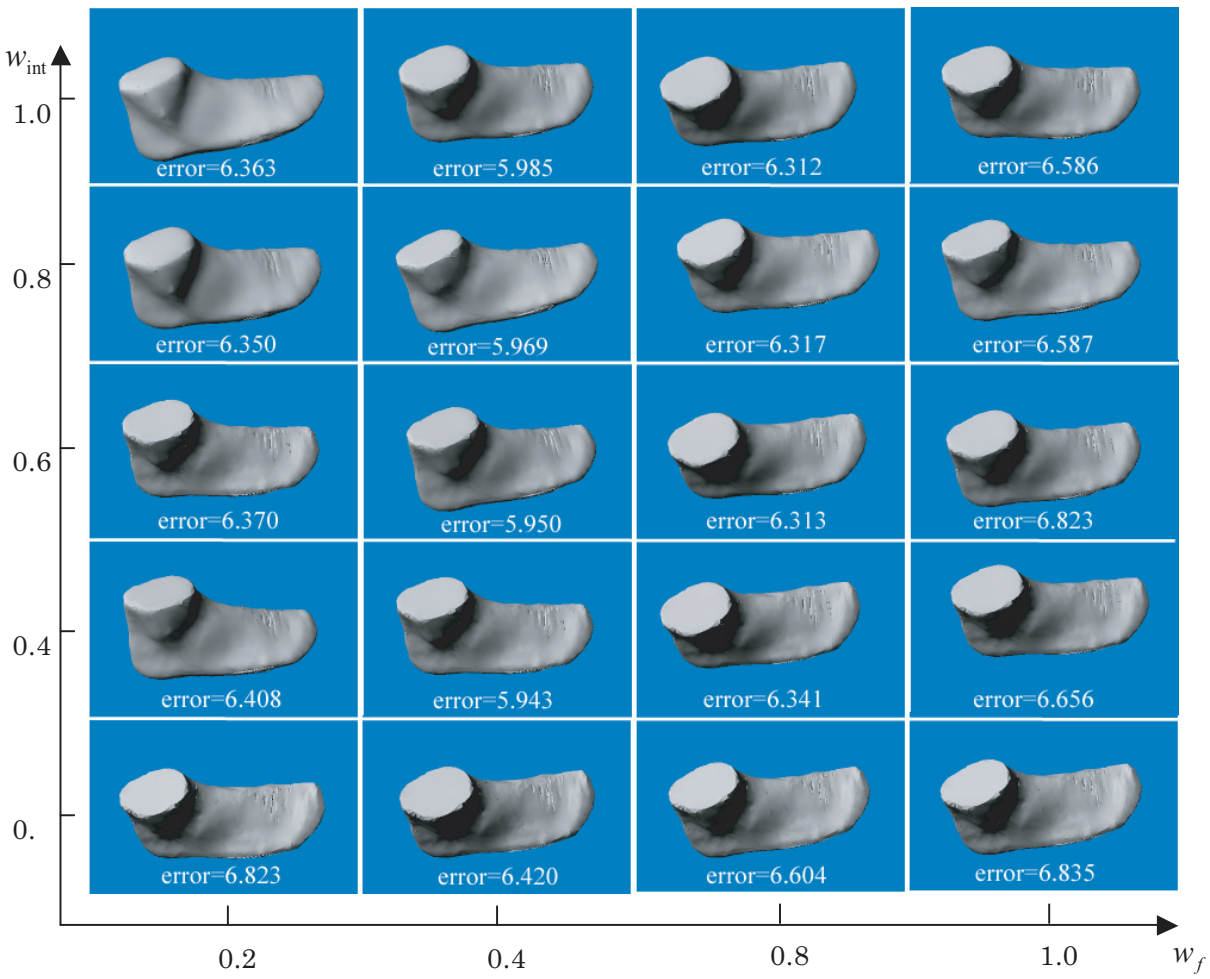


Figure 3.23: Reconstructed foot models under different combinations of weighting parameters of internal energy and standard model constraint.

Chapter 4

Parametric Deformable Model-based 3D Shape Reconstruction Method

4.1 Overview of Parametric Deformable Model Method

In the last chapter, EACM is used to detect the surface of human foot. To improve the shape similarity of foot, the foot model is refined with a standard model that derived from foot shape database. However, the accuracy of result is not satisfied. Although importance of each sample point of standard model is ranked by the variance of all the models in database, the shape generality of object is not always means the identity on the geometric size, whereas the comparability is more important conversely. For instance, the shape of ankle is an outstanding feature of foot. However, the height and size of the protuberance of ankle is various to each people. An ideal foot model must maintain the generality features and represent distinct features simultaneously. PDMM offers reasonable solution to this purpose.

Contrary to many other PDMM methods, active shape models (ASM) represents a general way of performing non-rigid object segmentation. Shape variation is extracted from prior knowledge (training set), using principal component analysis on point distribution models, rather than handcrafting a prior knowledge into the model. The optimal parameters can be obtained by the means of registering the initial model to distinct image features of object in given images.

During image search, the model is only allowed to deform to shapes similar to the one seen in the training set. Other deformable templates, such as ACM align to strong gradients for locating the object, regardless of their actual appearance in the image. However, sample points are not always placed on the strongest edge in the locality–they may represent a weaker secondary edge or some other image structure. ASM learns the typical appearance that along the perpendicular direction to the contour and builds a statistical model of geometric object structure to use for image search.

Although ASM is a powerful method, distinct limitation of the ASM is that its initial model does not consider gray-level variation of the object instance across images. Edwards, Cootes, and Taylor proposed active appearance models (AAM)[26] to apply gray-level mode variation of instance to overcome this limitation, whereas it causes other difficulties. AAM cannot handle cases well outside of the prior knowledge directly, e.g. occlusions. Moreover, in 3D space, to compute gray-level mode variations, the prior knowledge instances should be normalized in identical 3D intensity space[52], hence the time-consuming establishment of 3D intensity space must be performed. Moreover, the most fatal issue is that AAM is at

completely disadvantage, when the prior knowledge instances do not contain intensity information.

In this work, a PDMM-based object shape reconstruction method is proposed. The proposed method is similar to ASM in spirit. The proposed method is called modified active shape models (MASM). The initial model is described by pose and shape parameters. The evaluation of current model is implemented by the means of projecting the model to multiple view images without establishment of 3D intensity space. The reliability of each parameter is evaluated by assessing not only the registration of edge features, but also texture similarity of correspondence pixels among multiple views images[81].

From the next, the proposed MASM method is described in detail.

4.2 Initial Model

Similar to other deformable model methods, MASM starts from an initial model. The initial model is refined to fit the object surface. The initial model can be generated from a training set including m object models.

The training set is usually aligned manually when it was established. Each model is composed of n sample points. The sample points of each model describe the corresponding characteristic positions of object. Thus, a standard object model $\bar{\boldsymbol{v}}$ can be established, computing the average position of each sample point ((4.1)).

$$\bar{\boldsymbol{v}}_i = \frac{1}{m}\sum_{j=1}^{m}\boldsymbol{v}_j \quad (i = 1 \cdots n) \tag{4.1}$$

where i is the index of sample points, j is the index of models in the training set. In this work, the standard model is used as initial model.

4.3 Pose Parameters

After the initial model is obtained, position and orientation of the initial model should be registered to the given images.

The pose of a 3D model is defined similar to the camera model by scale, rotation and translation. It is assumed that the same 3D model with different poses in identical coordinate system are described as $\boldsymbol{V}_A$ and $\boldsymbol{V}_B$. The transform from $\boldsymbol{V}_A$ to $\boldsymbol{V}_B$ can be described as

$$\boldsymbol{V}_B = \boldsymbol{A}\boldsymbol{R}\boldsymbol{V}_A + \boldsymbol{T} \tag{4.2}$$

where

$$\boldsymbol{T} = \begin{pmatrix} \boldsymbol{T}_x \\ \boldsymbol{T}_y \\ \boldsymbol{T}_z \end{pmatrix}, \boldsymbol{A} = \begin{pmatrix} s & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & s \end{pmatrix} \text{ and}$$

$$\boldsymbol{R} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{pmatrix} \begin{pmatrix} \cos\theta_y & 0 & \sin\theta_y \\ 0 & 1 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y \end{pmatrix} \begin{pmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$\boldsymbol{A}$ is scale matrix by s, $\boldsymbol{R}$ describes rotation by $\theta = (\theta_x, \theta_y, \theta_z)$ and $\boldsymbol{T}$ is called translation vector. By adjusting the parameters $(s, \theta_x, \theta_y, \theta_z, T_x, T_y, T_z)$, the initial model can be refined to fit the object in position and orientation.

After the pose register is accomplished, the initial model will be close to the object. However, there may still differences between current model and the object in shape.

4.4 Shape Parameters

Principal component analysis (PCA) is an effective approach to describe the statistical relationship within a training set of objects. It can reduce the dimensionality of the data to something more manageable.

Each example in the training set is represented in a $3n$ dimensional space. Thus a set of m example shapes gives a group of n sample points in this $3n$ dimensional space. It is assumed that all the sample points lie within some region of the space, which is called the "Allowable Shape Domain" (ASD)[20]. Every sample point in this domain gives a set of sample points whose shape is broadly similar to that of those in the training set. Thus by moving about the ASD new shapes can be generated in a systematic way. However the n is more the cost of computation will go up distinctly. So we are anxious a good efficiency approach in low dimensional space. In order to achieve this goal, PCA is applied to the training set. For each shape in ASD, deviation from the standard model $\bar{\boldsymbol{v}}$ is calculated as:

$$d\boldsymbol{v}_i = \boldsymbol{v}_i - \bar{\boldsymbol{v}} \tag{4.3}$$

The $3n \times 3n$ covariance matrix $\boldsymbol{S}$ can be then calculated, using

$$\boldsymbol{S} = \frac{1}{m} \sum_{i}^{m} d\boldsymbol{v}_i d\boldsymbol{v}_i^T \tag{4.4}$$

The PCA gives the modes of variation of the shape. The variation modes are described by $\boldsymbol{p}_k$

(k=1, $\cdots$, $3n$), which the eigenvectors of covariance matrix $\boldsymbol{S}$ such that

$$\boldsymbol{S}\boldsymbol{p}_k = \lambda_k \boldsymbol{p}_k \tag{4.5}$$

where λ_k is the corresponding eigenvalue of $\boldsymbol{p}_k$, $\lambda_k > \lambda_{k+1}$.

It can be shown that the eigenvectors of the covariance matrix corresponding the largest eigenvalues describe the most significant modes of variations. Then the variations can be explained by a small number of modes. This means that the $3n$ dimensional space is approximated by a q dimensional space, where q is chosen so that the object shape can be described by relative small variables.

Therefore any samples in the ASD (i.e., any allowable shape) can be reached by adding a linear combination of the eigenvectors to the standard model $\bar{\boldsymbol{v}}$.

$$\boldsymbol{v} = \bar{\boldsymbol{v}} + \boldsymbol{P}\boldsymbol{B} \tag{4.6}$$

where $\boldsymbol{P} = (\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots \boldsymbol{p}_q)$ is the matrix of the first q eigenvectors, and $\boldsymbol{B} = (b_1, b_2, \cdots, b_q)$ is a vector of weighting parameters.

New samples of the shapes can be generated by varying the parameters b_k within suitable limits, so the new shape will be similar to those in the training set. The parameters are linearly independent, though there may be nonlinear dependencies still present. The limits for b_k are derived by examining the distributions of the parameter values required to generate the training set. Since the variance of b_k over the training set can be shown to be λ_k, suitable limits are typically of the order of

$$-3\sqrt{\lambda_k} \leq b_k \leq 3\sqrt{\lambda_k} \tag{4.7}$$

since most of the population lies within three standard deviations of the standard model.

4.5 Numerical Implementation

Appropriate pose and shape parameters are estimated with an iterative optimization procedure.

$$\begin{aligned} \boldsymbol{T} &\rightarrow \boldsymbol{T} + d\boldsymbol{T} \\ \theta &\rightarrow \theta + d\theta \\ s &\rightarrow s + ds \\ \boldsymbol{B} &\rightarrow \boldsymbol{B} + d\boldsymbol{B} \end{aligned} \tag{4.8}$$

Cost function is defined to evaluate the parameters. In the conventional ASM technique, the evaluation is implemented by checking the difference between model contour and edge elements in image data along the perpendicular direction. In general, a 2D or 3D intensity space is necessary. However, when the input is not volumetric data, such as multiple camera images, the issue will be complicated. Although as described in section 3.5.2.2 the 3D intensity space can be established by normalized correlation between image pairs, the precision is not satisfied. Thus, novel evaluation means is proposed by the means of projecting the model to multiple cameras images directly without establishing intensity space.

During the update step of the iterative process, if the difference of translation is too big, the refinement of the other parameters will be invalid. Consequently, it is assumed that the optimization of rotation parameters θ, scale parameter s and shape parameters $\boldsymbol{B}$ is started when the update ratio of the sample points is less than 40% at the beginning of the processing. The iteration will stop when the update ratio of the sample points of the last iteration and the current iteration is under a threshold. In our experiment, the threshold is always set to less than 1%.

To estimate optimal pose parameters, the silhouette images of multiple camera images are applied to be the assessment of the pose registration. It is assumed that all of the camera parameters have been obtained, hence the projection matrix $\boldsymbol{P}$ of each camera can be established easily. Then all the sample points on initial model are projected onto the silhouette images i.e. the world coordinates of sample points are transformed to a 2D coordinate system of image plane by (4.9).

$$\boldsymbol{v}_{i\_image} = \boldsymbol{P}\boldsymbol{v}_{i\_world} \tag{4.9}$$

where $\boldsymbol{v}_{i\_world}$ and $\boldsymbol{v}_{i\_image}$ are the world coordinate vector and image coordinate vector of sample points $\boldsymbol{v}_i$ respectively. In this means, 3D initial model and 2D multiple camera images is related. Because the intensity I_{ij} of silhouette pixels is always higher than the background pixels in a binary image, updating the pose parameters to maximize the cost function E_{pose} ((4.10)), the initial model may approache the object.

$$E_{pose}(\boldsymbol{A}, \theta, \boldsymbol{T}) = \sum_i \sum_j I_{ij} \tag{4.10}$$

where, i is the index of binary silhouette image of multiple camera images, j is the index of sample points.

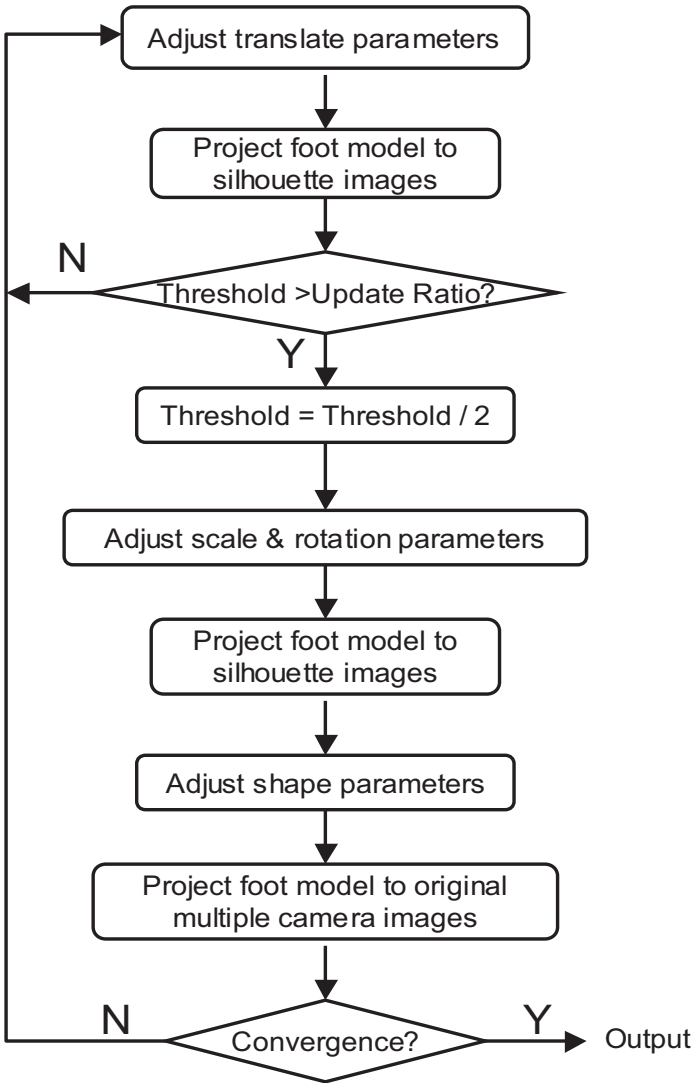


Figure 4.1: The flow chart of parameters optimization procedure.

On the other hand the cost function for shape parameters' optimization is also needed. The cost function is defined by the sum of absolute deviation of sample point's projection on multiple camera images.

$$E_{shape}(\boldsymbol{B}) = \sum_{i=1}^{f} \sum_{j=1}^{n} w_{ij} \left| I_{ij} - \bar{I}_j \right| \tag{4.11}$$

where i is the index of camera, j is the index of sample points. I_{ij} is the intensity of sample point's projection on image plane. w_{ij} is weight factor. $\bar{I}_j = \frac{1}{f} \sum_{i=1}^{f} I_{ij}$ is the average of I_{ij}.

Thus, the minimization of the cost function indicates the most credible position of sample points.

However, it is assumed that the background of the multiple camera image is black, i.e. the intensity of the background is "0". If the projection of sample points falls into the background field in all the multiple camera images, the cost function will be 0. However this is not a desirable result apparently. To overcome these issues, a distinctive condition is added to the cost function: the Euclidean distance between the contour sample points and the edge of object in the multiple camera images. By minimizing the distance, the sample points will be prevented from leaving the object images. The cost function is transformed as

$$E_{shape}(\boldsymbol{B}) = \sum_{i=1}^{f}\sum_{j=1}^{n}(w_{ij}\left|I_{ij} - \bar{I}_j\right| + w_{ij}^{'}D_{ij}) \tag{4.12}$$

where D_{ij} is the distance between the sample points whose projection is on the contour of the projection area of all the sample points (we call them "contour sample points") and the closest edge of foot in multiple camera images.

The optimization of the cost function is implemented with "Rosenbrock" algorithm[63] (For detail see Appendix C).

As other multiple camera-based modeling systems, the occlusion issue has to be cared. When a sample point is in occlusion, it is not visible in all of the multiple camera images. Thus, the occluded sample point will not do contribution in such invisible images. A novel sample points occlusion assessment method in Appendix D. Simultaneously, the estimation of contour sample points will also be explained.

To reflect the distance constraint and occlusion's effect on the process, we redefine the shape parameters cost function as

$$E_{shape}(\boldsymbol{B}) = \sum_{i=1}^{m}\sum_{j=1}^{n}(c_{ij}w_{ij}\left|I_{ij} - \bar{I}_j\right| + c_{ij}^{'}w_{ij}^{'}D_{ij}) \tag{4.13}$$

where

$$c_{ij} = \begin{cases} 0 & \text{occlusion} \\ 1 & \text{otherwise} \end{cases} \quad c_{ij}^{'} = \begin{cases} 1 & \text{counter sample points} \\ 0 & \text{otherwise} \end{cases}$$

4.6 Experiments and Results

4.6.1 Computer Graphics Data

The same CG simulative foot shape reconstruction experiment is carried out firstly.

Here the 3D model and multiple camera images are combined with the perspective projection matrix, and cost functions are defined to evaluate their registration. The total flow of MASM-based foot shape reconstruction is illustrated in Figure 4.2.

In the following experiments, the forementioned foot shape database (section 3.5.2) is adopted as the training set and the standard model derived from it is used to be initial model.

As described in section 4.4, the shape parameters of MASM are derived from the most significant modes of variations. Because the eigenvectors of the covariance matrix corresponding to the largest eigenvalues describe the most significant modes of variations, the contribution ratio of each mode for describing the model shape is calculated as

$$c_i = \frac{\lambda_i}{\lambda_T} \tag{4.14}$$

where λ_i is the eigenvalue of the covariance matrix, λ_T is the sum of all the eigenvalues. Furthermore, the cumulative contribution ratio ((4.15)) may help us to determine how many modes of variations should be adopted.

$$cum_i = \sum_{j=1}^{i} c_j \tag{4.15}$$

Table4.1 illustrates the contribution and cumulative contributions of shape mode variations. It can be found that more than 90% of the shape variance is explained by the first 12 modes of variations. On the other hand from the $13th$ mode the contribution ratios are less than 0.5%, thus the influence of them can be ignored. The first 12 modes are considered to be the principal components and the weights of these modes are the shape parameters in the experiments.

To give a visual explanation, examples of these shape variation modes are shown in Figure 4.3. The models are obtained by varying the first five shape parameters in turn. The first parameter varies the scale of foot model. The second parameter varies the height of ankle. The third parameter varies the shape of ankle section. The fourth and fifth parameters vary the length and curvature of the upside of foot.

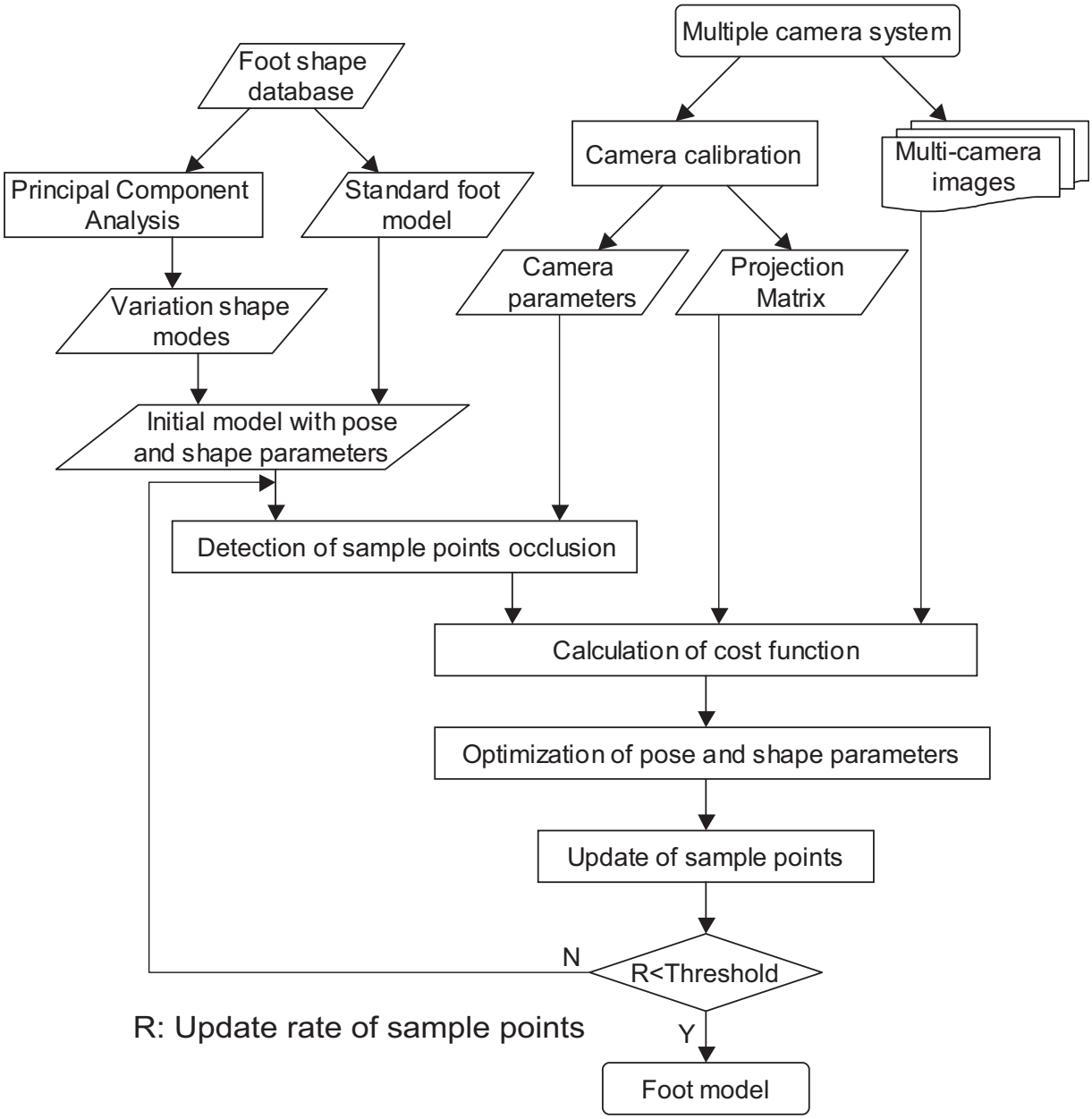


Figure 4.2: The flow chart of the foot shape reconstruction, using MASM.

Table 4.1: The contribution ratio and cumulative contribution ratio of variation modes

Eigenvalue	contribution ratio	cumulative contribution ratio
λ_1	64.58%	64.58%
λ_2	11.58%	76.16%
λ_3	4.69%	80.85%
λ_4	2.65%	83.50%
λ_5	1.89%	85.39%
λ_6	1.58%	86.97%
λ_7	1.35%	88.32%
λ_8	0.96%	89.28%
λ_9	0.88%	90.16%
λ_{10}	0.75%	90.91%
λ_{11}	0.59%	91.50%
λ_{12}	0.52%	92.02%
λ_{13}	0.48%	92.50%
λ_{14}	0.44%	92.94%
λ_{15}	0.41%	93.36%
λ_{16}	0.39%	93.75%
λ_{17}	0.35%	94.10%
λ_{18}	0.35%	94.45%
λ_{19}	0.29%	94.73%
λ_{20}	0.28%	95.01%
...	...	...

The result of MASM-based reconstruction method is illustrated in Figure 4.4. The point cloud in the images is the projection of sample points. Because the projection of sample points will be difficult to be recognized by the effect of texture on object surface, the sample points are projected to the silhouette images that are extracted by background subtraction. During the iterative (from top left to right bottom) projections of sample points are getting fitted the foot image well. Figure 4.5 is the reconstructed shape model displayed in 3D space.

Since volumetric intersection is a very popular surface reconstruction technique, we are

very interested in the comparison with it. Figure 3.17 has shown the result of volumetric intersection in 3D. It is obvious that MASM's result is smoother and more similar to the real human foot than volume intersection. The RMSE of MASM's result is 2.21mm.

Figure 4.6 shows the 2D instances of reconstructed foot surface of volumetric intersection method, EACM and MASM. The solid curve is the input foot surface contour. The highlight points indicate the position of sample points on reconstructed model. In Figure 4.6 (a), the result of volumetric intersection makes large errors to the object surface. The EACM-based method generates a smooth model surface, whereas errors are occurred near the corner of object surface (inside the dash line circle in Figure 4.6 (b)). On the other hand, MASM's result (Figure 4.6 (c)) gives the most reasonable shape to object.

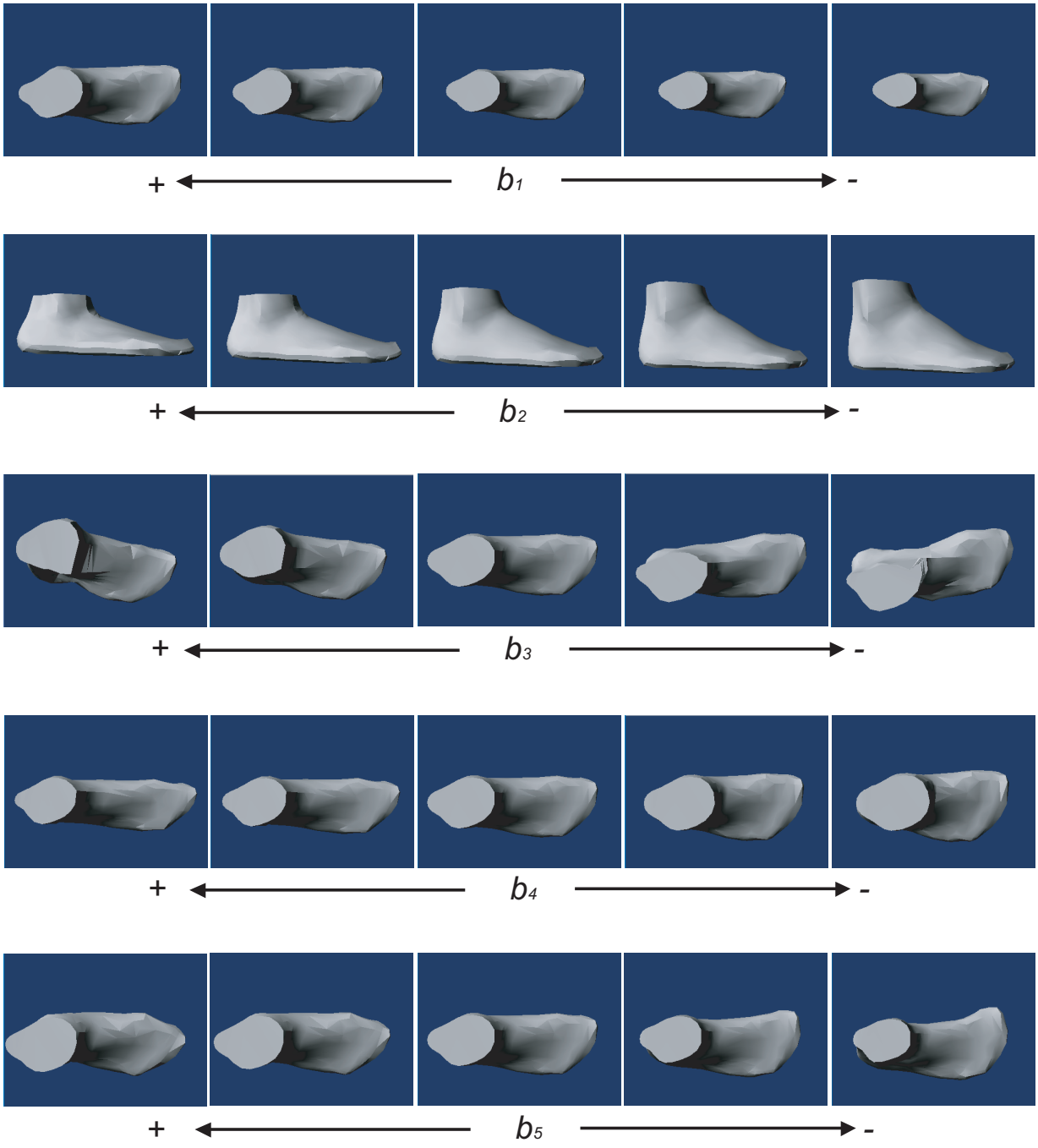


Figure 4.3: Effects of varying each of the first five shape parameters of the foot model individually.

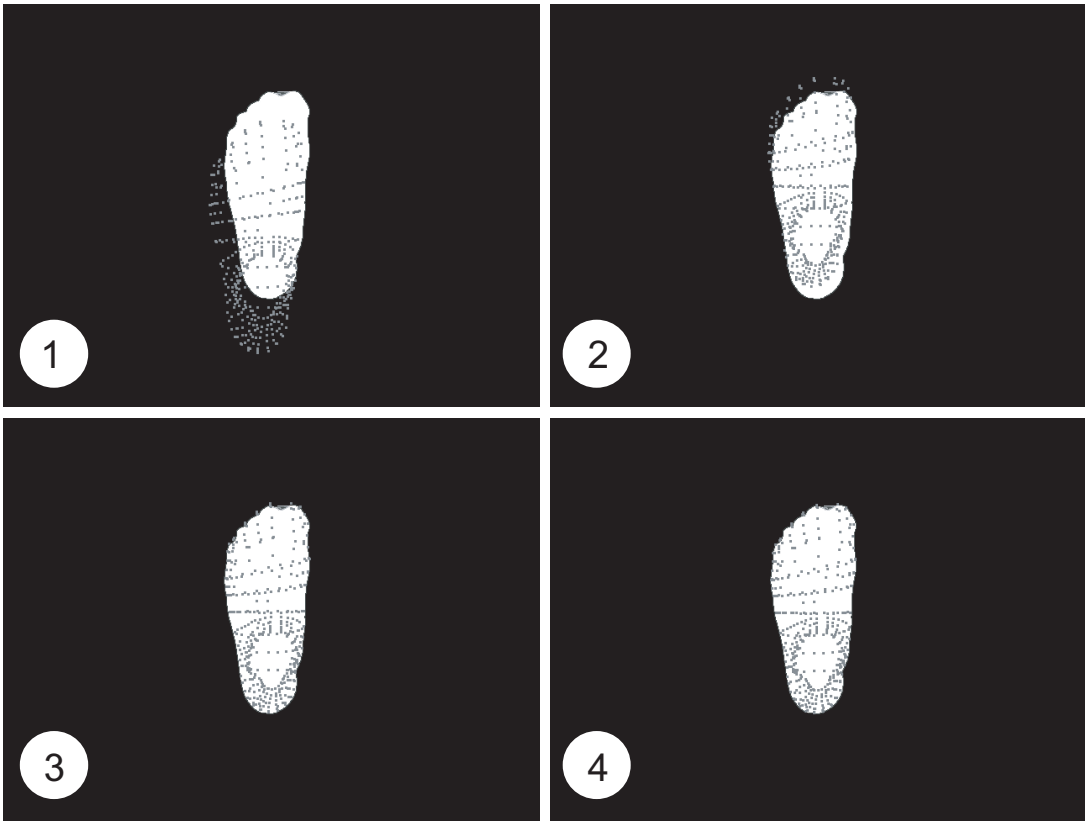


Figure 4.4: CG image of a foot model with the projection of sample points superimposed, during iterative process (from the top left to right bottom).

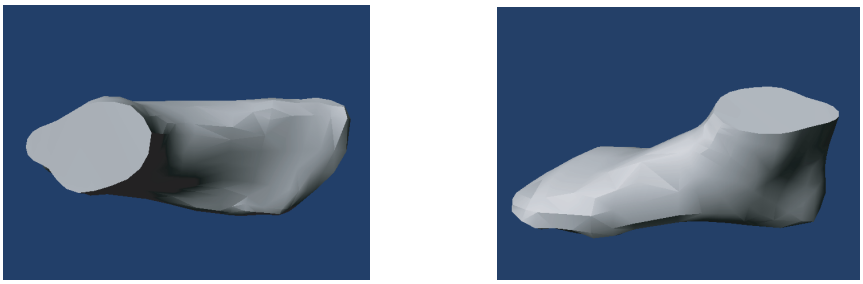


Figure 4.5: Reconstructed 3D model of CG foot model, using proposed method.

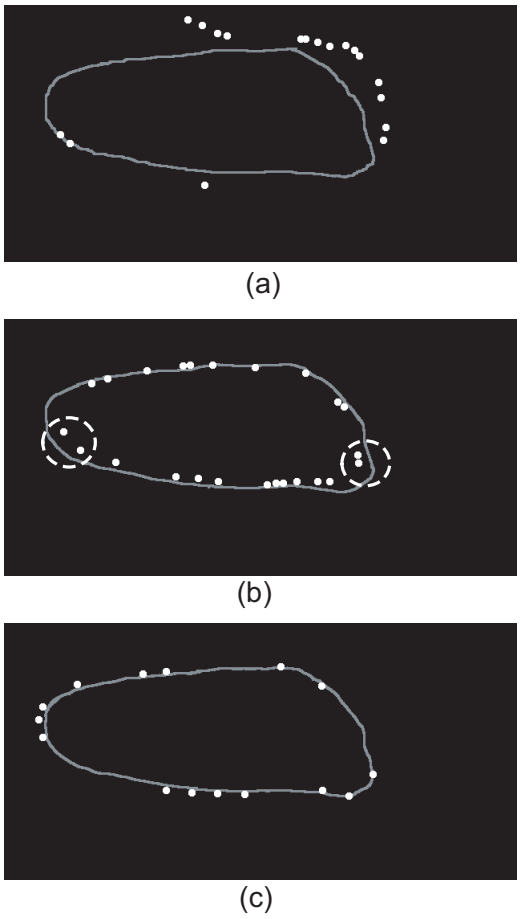


Figure 4.6: (a) The result of volumetric intersection makes large errors to the object surface. (b) The EACM provided a smooth model surface, whereas obvious errors are occurred near the corner part of object surface (inside the dash circle). (c) MASM generated the most reasonable surface model of object.

4.6.2 Real Data

Because the MASM-based foot shape reconstruction method achieves a good accuracy, it is also performed in real data experiments.

4.6.2.1 Multiple Camera Images Acquisition System

In the real data experiments, the multiple cameras are treated with standard video format (24 bit color or 8 bit black and white, 640×480 pixels) and frame rate (7.5 or 15 frames/sec). To record a scene with movement, the multiple images are acquired simultaneously. The precision of the synchronization of the multiple imaging devices plays an essential role for the accuracy potential of the measurement achieved using the images.

On the other hand, to record static scenes, the multiple camera images can be acquired at different times without a loss of accuracy. However, for applications involving recording people, the human body cannot be considered as a static object, because a person always moves slightly unconsciously due to, for example, breathing or muscle contraction. Therefore, for surface measurement of human body parts, it is always recommended to precisely synchronize the multiple cameras. Various methods can be used to acquire multiple camera images sequences. Although different camera systems have similar resolution and quantization, different levels of quality can be achieved depending on the system. They are listed below in order of decreasing accuracy potential:

- synchronized machine vision progressive scanning CCD cameras,
- synchronized machine vision interlaced scanning CCD cameras.

For the progressive scanning can be critical in viewing detail within motion images and modeling of dynamic object is a future work of our research, the multiple camera system is established with 8 Sony DFW-SX900 CCD cameras in this experiment.

4.6.2.2 System Calibration

System calibration is defined the simultaneous calibration of all the components involved in the acquisition system. Camera calibration refers to the determination of the intrinsic param-

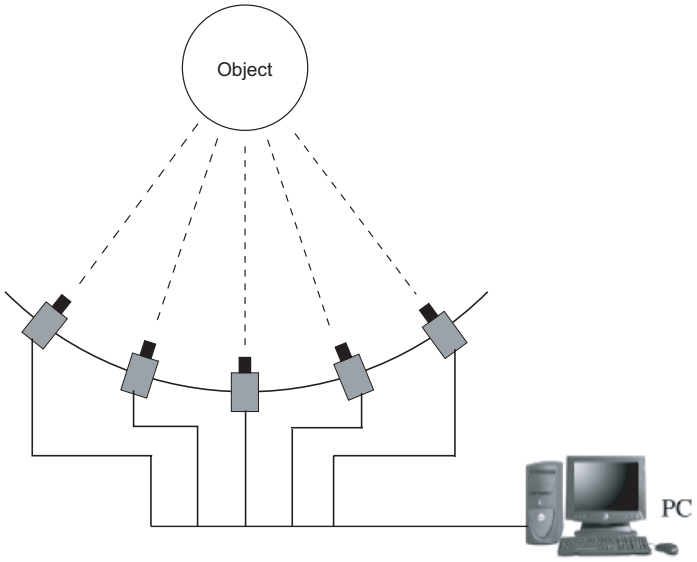


Figure 4.7: Progressive scan CCD camera multiple image acquisition system. Setup of the system, the cameras are synchronized together and connected to a PC.

eters describing the internal geometry of the individual imaging devices and other parameters (extrinsic parameters) modeling the systematic errors caused by the optical system and other sources and parameters that determine the parameters of exterior orientation to define the camera station and camera axis in the 3D space. A thorough determination of all the parameters is required for an accurate measurement. In the next sections, the mathematical model for the projection of the object space onto the digital image coordinate system is described first. The method used in this work to calibrate the multiple camera systems is then presented.

4.6.2.3 Mathematical Camera Model

In this section the mathematical model for the projection of a point in the world coordinate system is described with the transformation into image coordinates (Figure 4.8). This transformation can be divided into 4 steps.

Figure 4.8: The transform from world coordinate to image coordinate.

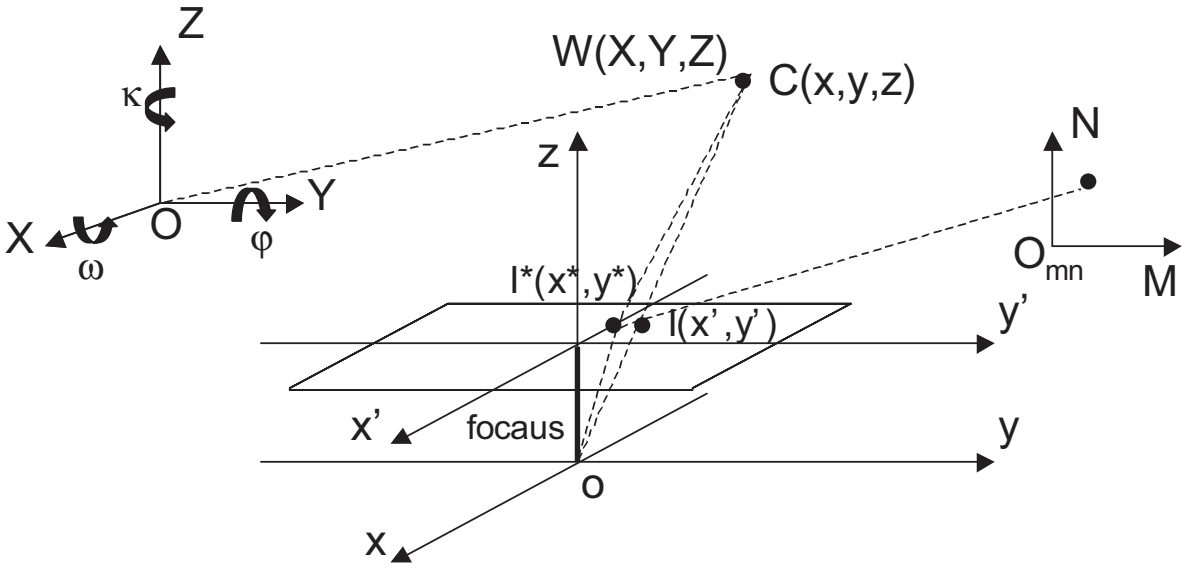


Figure 4.9: $W(X, Y, Z)$ in world coordinate is transformed to image coordinate (M, N), using the intrinsic and extrinsic camera parameters.

(1) A world coordinate vector $W(X, Y, Z)$ is projected to the camera coordinate as $C(x, y, z)$ by the following equation:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \boldsymbol{R} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \boldsymbol{T} \tag{4.16}$$

where $\boldsymbol{R}(\omega, \phi, \kappa)$ is 3×3 relation matrix that describes the rotation from world coordinate system to the camera coordinate system.

$$\begin{aligned} \boldsymbol{R} &= \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \\ &= \begin{bmatrix} \cos\omega\cos\kappa & -\cos\phi\sin\kappa & \sin\phi \\ \cos\omega\sin\kappa + \sin\omega\sin\phi\cos\kappa & \cos\omega\cos\kappa - \sin\omega\sin\phi\sin\kappa & -\sin\omega\cos\phi \\ \sin\omega\sin\kappa - \cos\omega\sin\phi\cos\kappa & \sin\omega\cos\kappa + \cos\omega\sin\phi\sin\kappa & \cos\omega\cos\phi \end{bmatrix} \end{aligned} \tag{4.17}$$

$\boldsymbol{T}$ is 3×1 translation matrix that describes the translation of origins from world coordinate system to the camera coordinate system.

$$\boldsymbol{T} = \begin{bmatrix} T_x & T_y & T_z \end{bmatrix}^T \tag{4.18}$$

(2) The transform from 3D camera coordinate system $C(x, y, z)$ to 2D image coordinate

system without distortions, denoted by $I(x', y')$, is described as:

$$x' = f\frac{x}{z} \quad y' = f\frac{y}{z} \tag{4.19}$$

where f is the focal length of lens.

(3) The coordinate vector $I^*(x^*, y^*)$ with distortion can be computed by

$$\begin{aligned} x^* &= x' - R_x \\ y^* &= y' - R_y \end{aligned} \tag{4.20}$$

where, R_x and R_y are symmetric radial and lens distortion respectively.

(4) Finally, the pixel coordinate vector in the computer can be obtained by the following transformation:

$$\begin{aligned} M &= \frac{x^* M_x}{\mu_x} + O_m \\ N &= \frac{-y^*}{\mu_y} \end{aligned} \tag{4.21}$$

where (O_m, O_n) is the coordinate of image center. μ_x, μ_y are the pixel size in x-axis and y-axis respectively.

In general, $\boldsymbol{R}$ and $\boldsymbol{T}$ are called extrinsic camera parameters, and the other parameters are called intrinsic camera parameters. If these parameters are obtained, an arbitrary point's pixel coordinate in computer memory can be computed from its world coordinate easily, and vice versa. The estimation of these camera parameters is called camera calibration. The transform from world coordinate to image coordinate is illustrated in Figure 4.9.

4.6.2.4 Camera System Calibration by Self-calibration

To orient and calibrate camera systems, various methods can be used. However, two characteristics of multiple camera systems have relevant importance for choosing adequate and appropriate calibration and orientation procedures: (a) the multiple cameras have usually either a fix position or they are displaced all together without changing their relative positions; (b) the multiple cameras have to be calibrated and oriented very often (e.g. at every acquisition sessions) because of the need of small adjustments (e.g. focus and iris, and even position and direction). For these reasons, a simultaneous calibration and orientation (system calibration) of the multiple cameras is more appropriate. These two problems are solved in this research work with a convenient self-calibration method.

This method only requires the cameras to observe a planar pattern shown at a few different orientations and a static cubic reference object and track a light maker in the calibration space[46]. The procedure is as follows:

- Estimate the intrinsic parameters by Zhengyou Zhang's flexible calibration method[98].
- Compute the initial estimate of the extrinsic parameters with Direct Linear Transform method[1] by using a cubic reference object.
- Track a distinct marker simultaneously from all the viewing points.
- Use the epipolar constraint provided by the marker positions in all the input views. The initial estimation of the extrinsic parameters are refined with the down-hill simplex optimization algorithm[54](For the detail of down-hill simplex see Appendix E).

■Estimation of the Intrinsic Parameters The calibration procedure of intrinsic parameters is as follows:

- Print a checker patter (Figure 4.10) and attach it to a planar surface;
- Take a few images of the model plane under different orientations by moving either the plane or the camera (Figure 4.11);
- Detect the feature points in the images;
- Estimate the intrinsic parameters and all the extrinsic parameters using the closed-form solution;
- Estimate the coefficients of the radial distortion by solving the linear least-squares;
- Refine all parameters by minimizing.

For more details see the reference[98].

■Estimation of the Initial Extrinsic Parameters with DLT Recording images using a camera is equivalent to mapping object point $O(X, Y, Z)$ in the world coordinate system to image point $I'(x', y')$ in the image plane. For digitization, this recorded image will be projected again to image $I(x, y)$ in the projection plane. However, for simplicity, it is possible to directly relate the projected image to the object. Object O is mapped directly to the projected image I by (4.22),(4.23).

$$\begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{4.22}$$

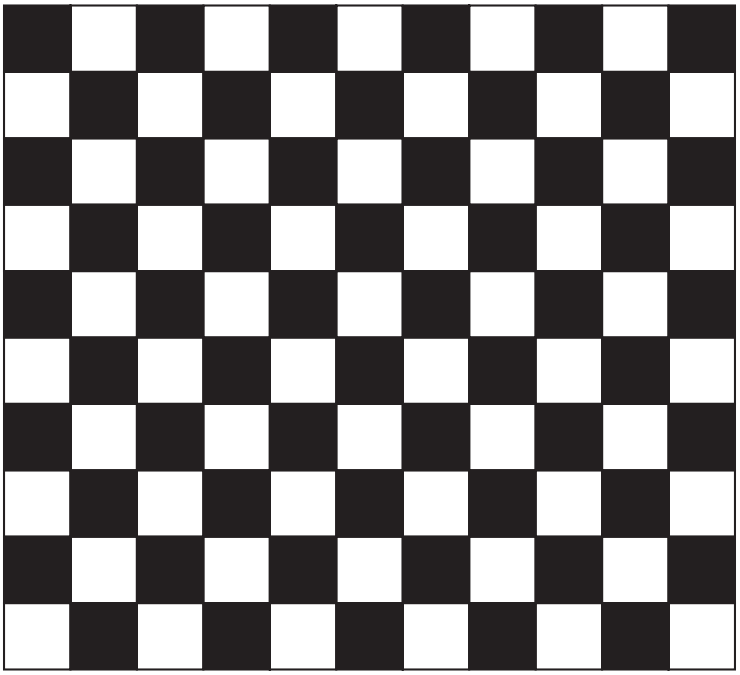


Figure 4.10: A plane with 11×10 checker pattern for estimating intrinsic camera parameters.

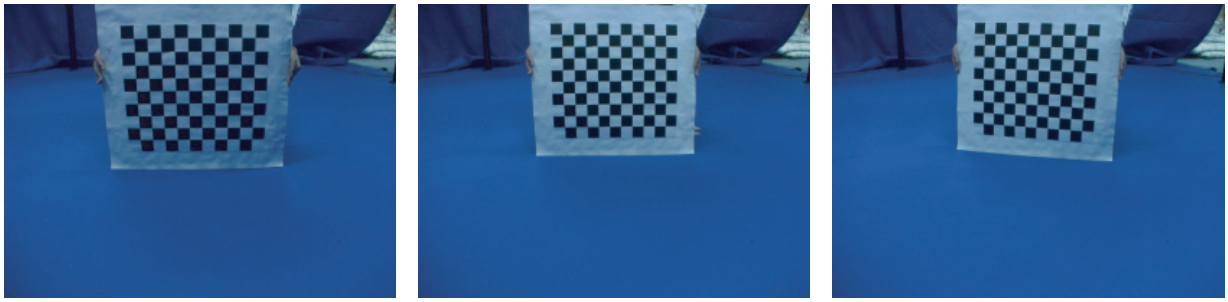


Figure 4.11: Take a few images of the checker pattern plane under different orientations.

Then

$$x = \frac{C_1}{C_4} \quad y = \frac{C_2}{C_4} \tag{4.23}$$

$\boldsymbol{P} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$ is called projection matrix. By expanding (4.23)

$$\begin{aligned} x &= \frac{a_{11}X + a_{12}Y + a_{13}Z + a_{14}}{a_{41}X + a_{42}Y + a_{43}Z + 1} \\ y &= \frac{a_{21}X + a_{22}Y + a_{23}Z + a_{24}}{a_{41}X + a_{42}Y + a_{43}Z + 1} \end{aligned} \tag{4.24}$$

If a group of control points whose x, y and z coordinates correspond image coordinates on image plane are already known. The projection matrix can be computed with (4.24). For the free degree is 11, the known control points should more than 6. The control points must not be co-planar. In other words, the control points must form a volume, the control volume. The control points are typically fixed to a calibration frame or control object. In this work, we use a cubic reference object (Figure 4.12) for compute the projection matrix.

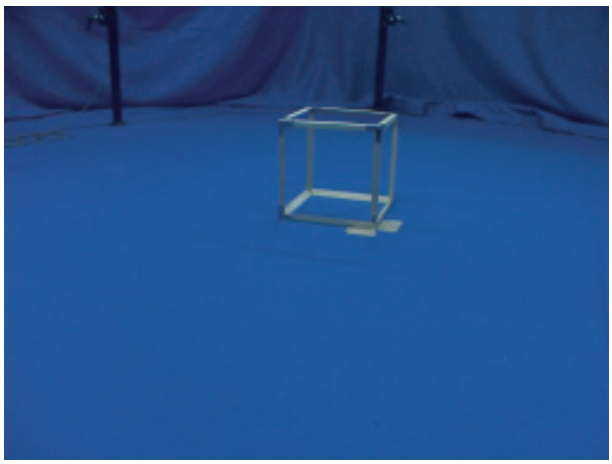


Figure 4.12: Compute the initial estimate of the extrinsic parameters with Direct Linear Transform method, using a cubic reference object.

Projection matrix $\boldsymbol{P}_{3\times4}$ can also be obtained by $\boldsymbol{P} = \boldsymbol{A}[\boldsymbol{R}|\boldsymbol{T}]$, and because the intrinsic parameters are known, the extrinsic parameters can be estimated from the projection matrix.

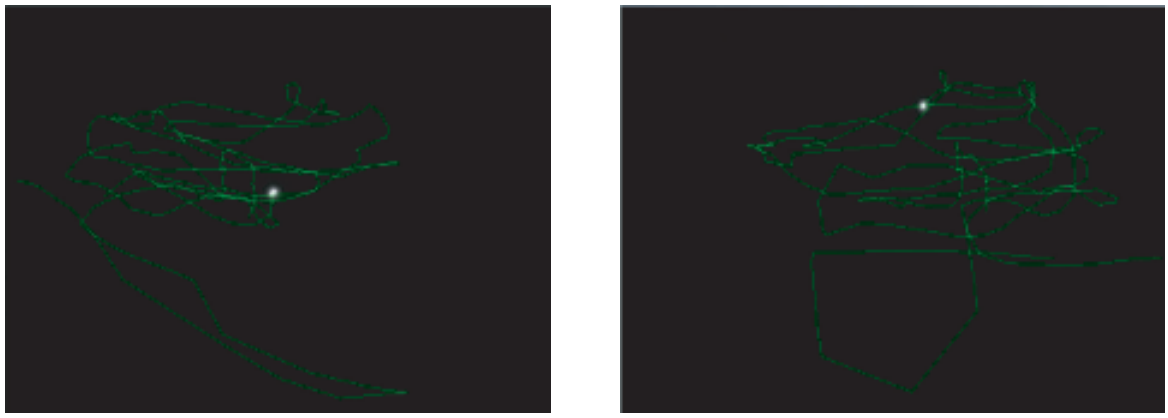


Figure 4.13: Track a distinct marker simultaneously from all the view points.

■Refination of the Extrinsic Parameters If $m(u,v)$ and $m'(u',v')$ are corresponding image points of a 3D object point in two cameras C' and C's image planes, then

$$\begin{bmatrix} u' & v' & 1 \end{bmatrix} \boldsymbol{F} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = 0 \tag{4.25}$$

i.e. $x'Fx = 0$, where F is the 3×3 fundamental matrix of maximum rank 2. (4.25) is also called epipolar constraint.

In general F can be estimated as follows:

$$\boldsymbol{F} = \boldsymbol{A'}^{-T}[\boldsymbol{T}]_x \boldsymbol{R} \boldsymbol{A}^{-1} \tag{4.26}$$

where $\boldsymbol{A'}$ and $\boldsymbol{A}$ are the intrinsic matrix of camera C' and C respectively. $\boldsymbol{T}$, $\boldsymbol{R}$ are respectively rotation matrix and translation matrix of camera C.

Because the camera parameters of each camera have been obtained by the method described above, an initial fundamental matrix can be estimated by (4.26).

To refine the parameters, particularly the extrinsic parameters, a highlight marker is tracked simultaneously in the calibration space (Figure 4.13). The marker is extracted from each multiple camera image. Then the initial extrinsic parameters are refined to satisfy the epipolar constraint. Down-hill simplex optimization algorithm is implemented for refining the camera parameters (Appendix E).

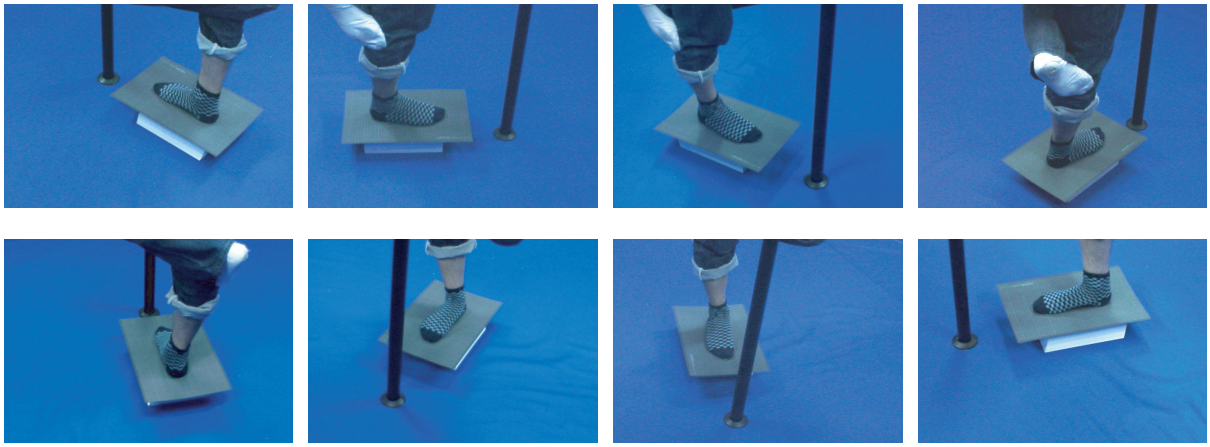


Figure 4.14: Multiple camera images of real human foot.

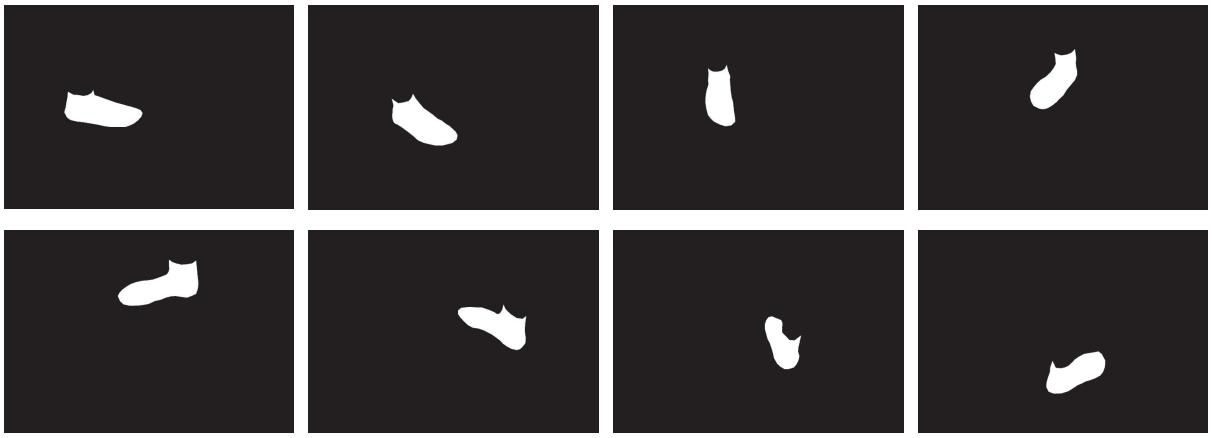


Figure 4.15: Multiple camera silhouette images of real human foot.

4.6.2.5 Experimental Results

In the real data experiments, to make the partition of foot images easily and decrease the light reflection of foot skin, object is put on socks. The socks can also add texture information for parameter optimization procedure. The multiple camera images are shown in Figure 4.14.

In the pose parameter estimation, silhouette images are necessary. In the CG experiment, silhouette images are generated by background subtraction. However real data's background is more complex than CG data, hence, not only background subtraction, but also manual remove of the unwanted objects, e.g. the leg, is implemented. Although the manual "leg

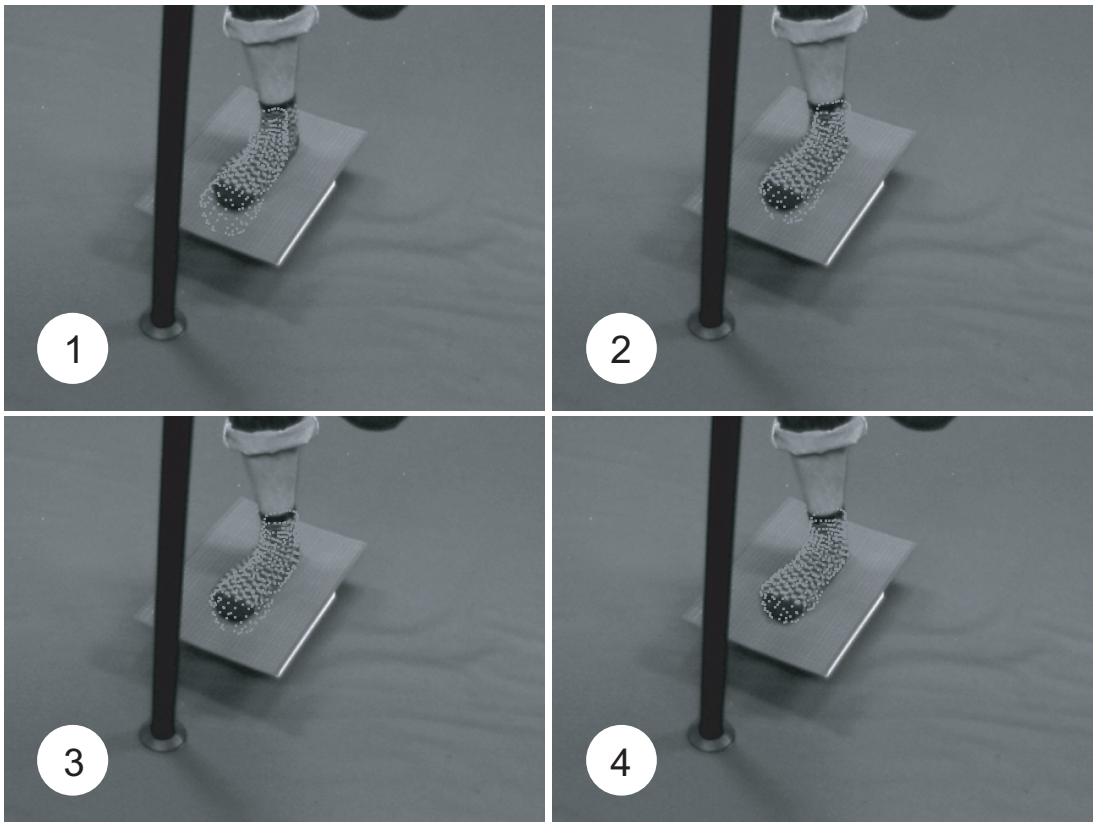


Figure 4.16: Multiple camera image of real human foot with the projection of sample points superimposed, during iterative process (from the top left to right bottom).

remove" work is complicated, for 8 multiple camera images, it is still acceptable. The silhouette images are shown in Figure 4.15. Although the silhouette of foot is obtained, gaps are occurred obviously. Because this problem happened after erasing the leg part, it is called "leg occlusion".

Although the image data of real foot is more complicated than CG data and the camera parameters are not as accurate as CG simulative experiment, the foot model is approaching the foot image during the iterative optimization procedure (Figure 4.16) and satisfied 3D model is established (Figure 4.17). To give a qualitative evaluation, volumetric intersection method is performed. The reconstructed 3D model is illustrated in Figure 4.18. It is obvious that volumetric intersection generated a coarse foot model. This happened principally because of the "leg occlusion" phenomenon. The volumetric intersection back-projects the voxels to multiple camera silhouette image in a cone space, then intersects all the cones to build a volume, which is guaranteed to contain the object. Because this approach relies on the

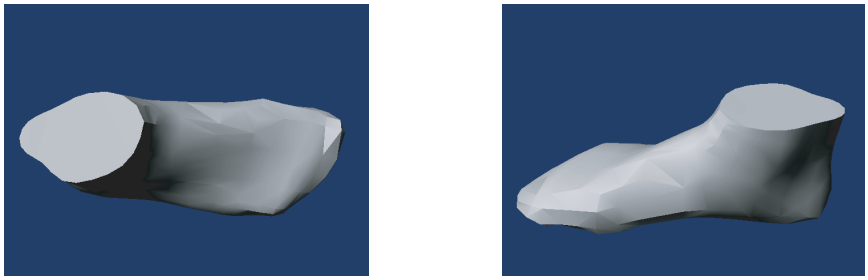


Figure 4.17: Reconstructed 3D model of real human foot, using proposed method.

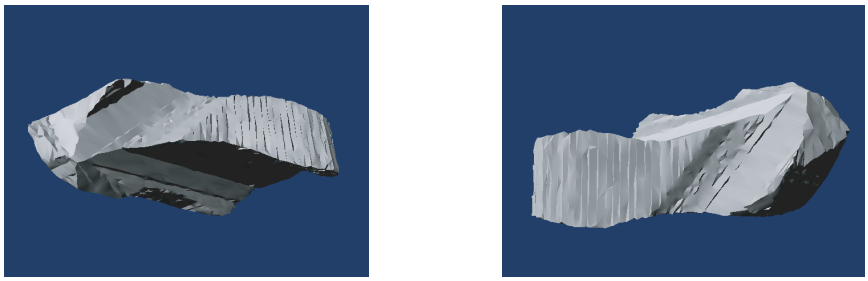


Figure 4.18: Reconstructed 3D model of real human foot, using volume intersection.

smallest crossing volume of all the volumetric cones, if there are gaps on silhouettes, the reconstructed model will be corrupted largely. Since MASM adopts training set to establish an standard foot shape as the initial model and statistical knowledge of foot shape is calculated by PCA, the effect of “leg remove” can be removed. Figure 4.17 shows that an integral foot model is reconstructed by MASM.

The qualitative evaluation described above validated that the proposed method is also available for real human foot reconstruction. However, to give more reasonable evaluation, quantitative evaluation is also expected.

A different point between CG data and real camera data is that the pre-evaluation answer is not known. Thus, the validation of the results is a difficult issue. In this work, an experiment concerning a plastic foot model, whose position is measured previously, was performed. The multiple camera images and corresponding silhouette images are shown in Figure 4.19 and Figure 4.20 respectively. The result of proposed iterative method is shown in Figure 4.21. The RMSE is about 2.46mm. This error is near the CG experiment.

Since there is no effect like “leg remove” for plastic foot model, the result of volumetric intersection method (Figure 4.22) is not as coarse as Figure 4.18. However, 8 cameras are too

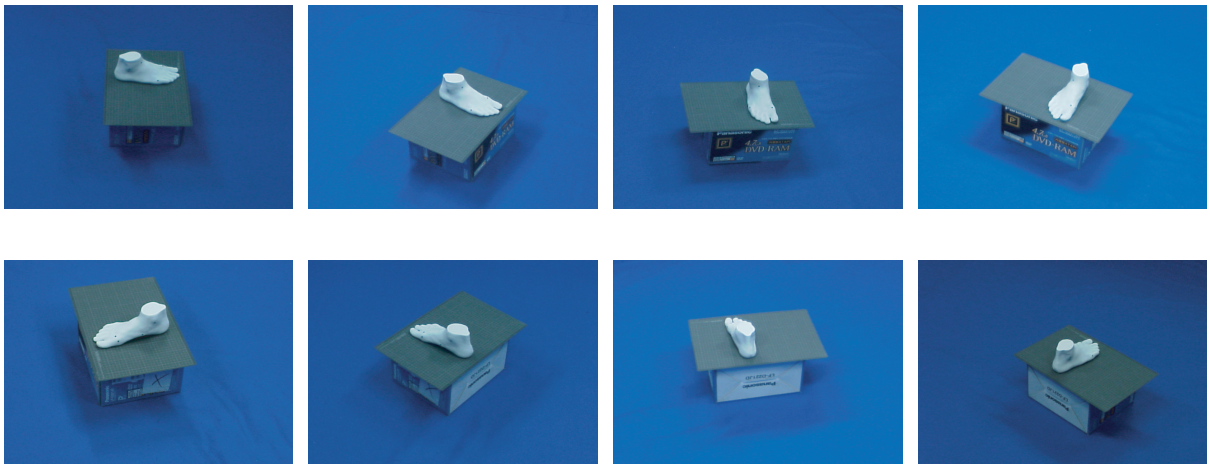


Figure 4.19: Multiple camera images of plastic foot model.

Figure 4.20: Multiple camera silhouette images of plastic foot model.

little for the volume intersection algorithm, the reconstructed 3D model is still too rough to be considered an acceptable approach. On the other hand, under the same condition (using 8 CCD cameras) a better 3D model is obtained by MASM (Figure 4.23).

Because there is no effect of “leg occlusion”, the plastic foot model experiment is simpler than the real human foot. However, since an integral initial model is adopted and the PCA-shape deformation constraints can prevent unexpected deformation, the “leg occlusion” is not an important factor for the proposed method. Furthermore, because the multiple camera images are acquired in the same image acquisition environment, the quantitative evaluation on plastic foot model experiment offers that the proposed method is available for real world

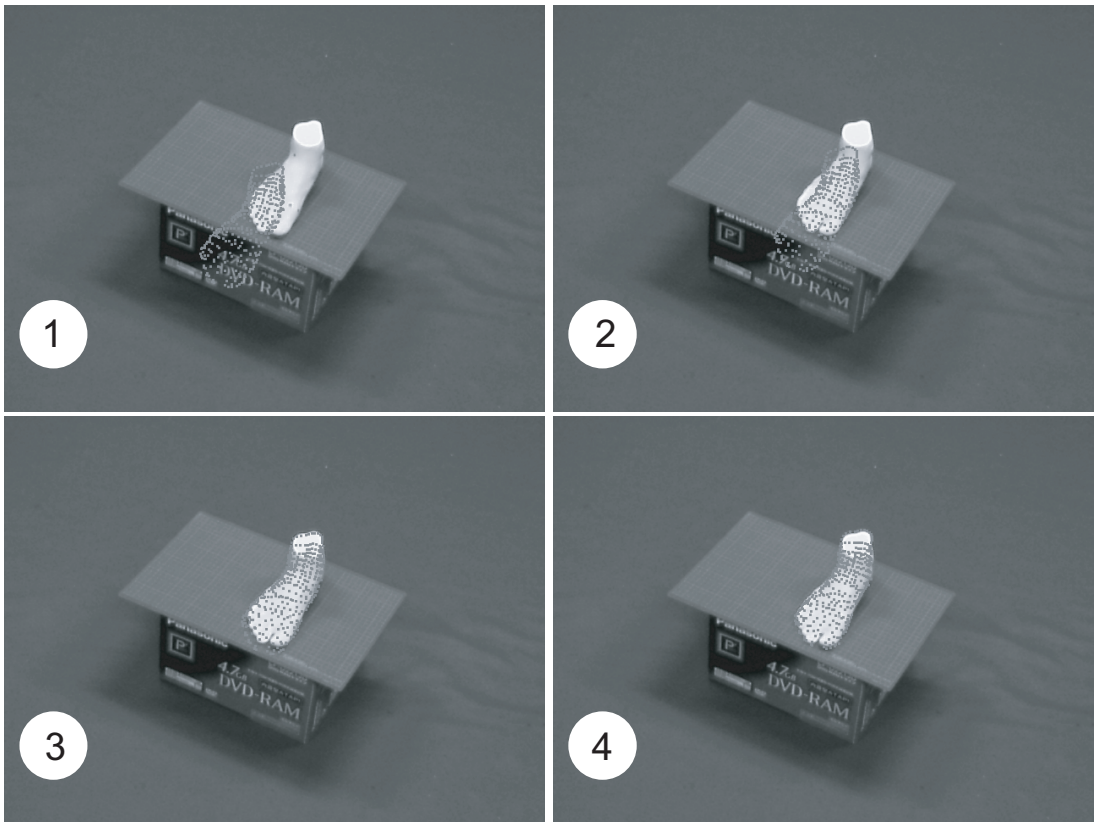


Figure 4.21: Real camera image of a plastic foot with the projection of sample points superimposed, during the iterative process (from the top left to right bottom).

object.

4.7 Discussion

PCA can reduce the dimensions of variations by effective variation modes, whereas the number of variation modes should be considered carefully. Although the cumulative contribution ratio is a useful assessment, other factors, e.g. time costing, the information that carried by each variation mode and so force, are also have to be concerned. To discover the effect of these factors, experiments with different number of variation modes are implemented. As that is shown in Table 4.2, when the cumulative contribution ratio is enlarged, the time cost is increasing and the error is decreasing. However, if the cumulative contribution ratio getting is more than 90%, the error's descending speed is getting slow, whereas the time costing is increasing quickly. This trend is more distinct when the cumulative contribution

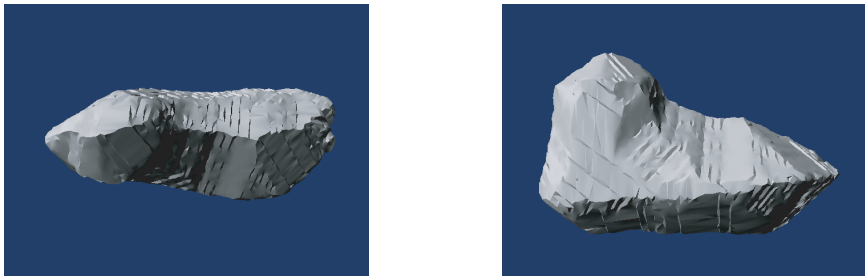


Figure 4.22: Reconstructed 3D model of plastic foot model, using volume intersection.

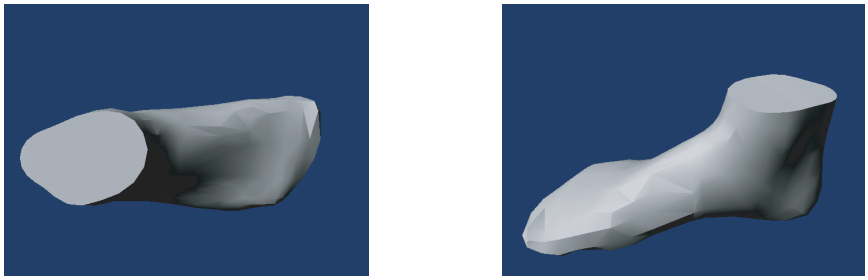


Figure 4.23: Reconstructed 3D model of plastic foot model, using proposed method.

ratio is more than 92%. Figure 4.24 illustrates Table 4.2 visually. Moreover, the contribution ration is less than 0.5% after the 13th variation mode (Table 4.1). Thus, the first 12 variation modes are adopted to be the shape parameters in the experiments.

Table 4.2: The investigation of the effect of cumulative contribution ratio on processing time cost and error

cumulative contribution ratio (number of variation modes)	time costing	Error (mm)
80.85% (3)	4’42”	3.38
85.39% (5)	5’30”	2.97
90.16% (9)	10’00”	2.72
92.02% (12)	14’22”	2.46
95.01% (20)	22’52”	2.45

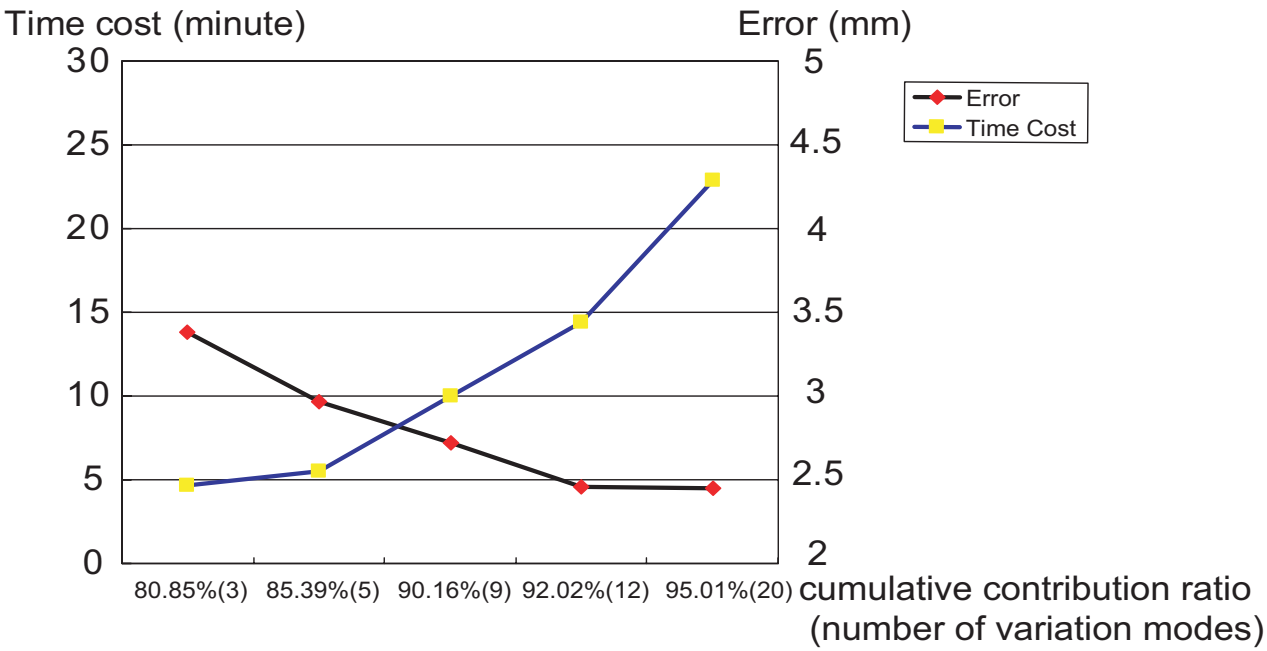


Figure 4.24: When the cumulative contribution ratio is enlarged, the time cost is increased and the error is decreased. If the cumulative contribution ratio is more than 90%, the error's descending speed is getting slow, whereas the time costing is increasing quickly. This trend is more distinct, while the cumulative contribution ratio is more than 92%.

Chapter 5

Conclusions, Discussion and Future Works

5.1 Summary of the Proposed Methods

5.1.1 Free-form Deformable Model-based 3D Shape Reconstruction Methods

The deformable model-based methods for reconstructing object shape in 3D are proposed in this paper. To handle open form arbitrary shape object such as the skin inner tissue surface, free-form deformable models-based methods AOSM (active open surface models) and EACM (energy modified active contour models) are proposed. The merits of these methods are summarized in the following:

1. The reconstruction of open form surface are addressed by proposed methods;
2. Triangle patch based initial model favors the EACM method to be a general approach for both open and closed form surface;
3. EACM offers novel external constraint, so that even the initial model is set far from the object surface, the model can also approach the desirable position;
4. Levenberg-Marquardt based numerical implementation favors EACM search the object surface with high convergence speed and global minimum.

5.1.2 Parametric Deformable Model-based 3D Shape Reconstruction Method

On the other hand, to generate high reliable reconstruction, a parametric deformable model-based method: MASM is proposed. This method makes use of prior knowledge (training set), so that the object shape can be recovered according to the common shape features of the same class of the object. The merits of this method is:

1. Unexpectable deformation are prevented by shape variation modes from training set;
2. Because standard model from the training set is used to be initial model, integral 3D model can be reconstructed even part of the object is occluded;
3. A novel parameter optimization procedure is proposed, so that the parameters can be optimized reliably.

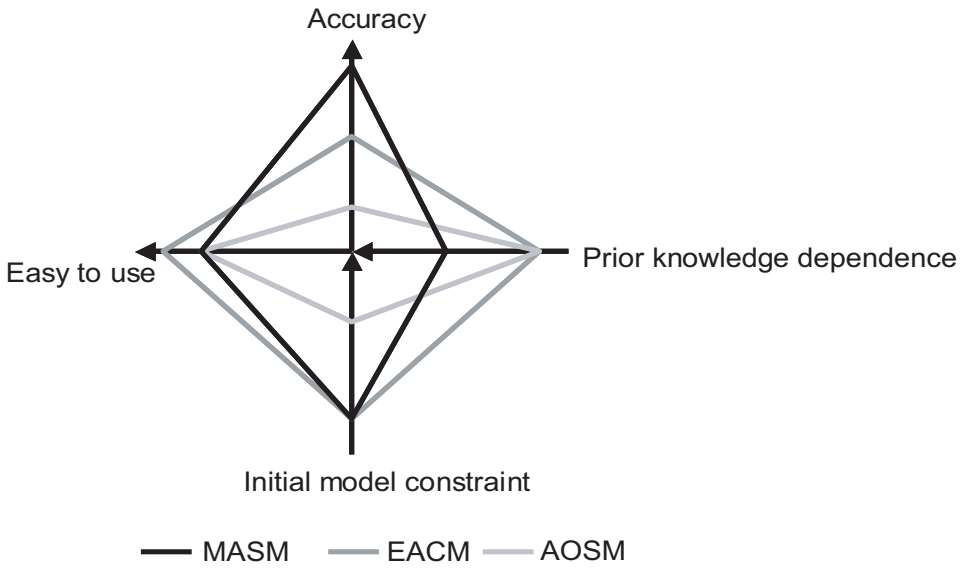


Figure 5.1: Benchmarking of the three proposed modeling techniques: AOSM, EACM and MASM on four aspects: accuracy, easy to use, initial model constraint and prior knowledge dependence. The farther from the center of the graph, the better it is (note the different directions of the axes).

Particularly, in the proposed method, the evaluation of model is implemented by projecting the 3D model to multiple camera images. Thus the gray-level features of object are efficiently applied to the optimization of shape model.

5.2 Comparison of the Capabilities

The proposed methods are implemented in some applications, such as reconstruction of skin inner tissue surface and human foot. The capabilities of AOSM, EACM and MASM are validated with these experiments.

To compare the proposed methods for object shape reconstruction, four aspects can be considered: accuracy, easy to use, initial model constraint and prior knowledge dependence. Figure 5.1 shows the four aspects for the proposed methods.

The accuracy of the 3D reconstruction techniques is one of the most important and significant aspects. The greatest accuracy potential is held by MASM, followed by EACM and AOSM. It is obvious that the prior knowledge increasing the stability of the shape reconstruction. In most applications, a training set, in which the instance describe the shape of the same

class of object, is introduced as prior knowledge. The PCA estimated shape mode variances gives promised result in the experiments of MASM.

The dependence of prior knowledge is a negative aspect to accuracy in a certain extent. Although the prior knowledge is efficient to improve the result accuracy, it will limit the object that can be handled by MASM. In some applications the shape of objects don't have distinctly common features, i.e. the shape variances of almost all the sample points are too large to align, such as the skin tissue intersection surface. The free-form deformable model based methods: AOSM and EACM are successful in coping with such object without prior knowledge.

The initial model constraint, in fact, is correlating very much to the easy to use aspect that will be described subsequently. Because the AOSM is an appropriative open form surface modeling processor, its initial model is always open space surface and partitioned into two parts, hence the processing becomes complicated. Conversely, the initial model of EACM and MASM are corresponding to the object shape, thus the EACM and MASM are more flexible.

The last investigating aspect is easy to use, which is concerning with the last two aspects. The MASM method implies different processing steps (system calibration of multiple camera system, pose matching, shape parameters optimization), whereas AOSM and EACM deliver the results through more simple procedure. However, the establishment of voxel space makes the pre-processing of AOSM and EACM complex. Particularly, the "easy to use" is decreased by the complicated initial model of AOSM.

5.3 Suggestions for Further Research

5.3.1 Accuracy Improvement of Human Skin Tissue Surface Reconstruction

Concerning the skin tissue intersection surface reconstruction system, the evaluation is a puzzling issue. In this work, the result surface is evaluated by the comparison with human experts' observation. However, this evaluation is obviously subjective and is not helpful for the result accuracy. To improve the accuracy, a skin inner tissue training set may be helpful. The training set is established with three uncorrelated conditions:

1. Age: samples with broad age range;

2. Time: the same sample is measured at different times;
3. Position: samples cover the different positions of human body, such as the skin tissue of leg, face and arm.

Global structure template can be extracted from this training set to constrain the model deformation. The parametric deformable model methods can also be implemented with this training set to generate high accuracy result. However, the alignment of samples in both measurement and analysis phases may be difficult.

5.3.2 Energy Evaluation of Free-form Deformable Model Methods

In FDMMs, energy functions are used to evaluate the current models. Generally, the evaluation is implemented in intensity space. Although normalized correlation can be used (section 3.5.2.2), the processing is complicated and the accuracy is doubtful. MASM gives constructive approach. The energy functions of FDMMs can also be evaluated by the means of projecting the model to multiple camera images. The pose of initial model and object can also be registered, using the pose parameters like MASM. However, the cost function of the pose registration has to be determined carefully. For instance, the texture correlation of control points between multiple camera images or the edge elements are reasonable assessment for the foot shape reconstruction application.

5.3.3 Data Acquisition and Motion Data Processing of Multiple Camera System

Concerning the multiple camera image acquisition systems, further research can be directed toward two different interesting goals: high quality or low cost. For establishing high quality acquisition system the number of cameras has to be extended, allowing a more complete imaging of the interested object (from the front, sideways and form the back). In the other direction, for demonstration and educational purposes, it would be very interesting to develop a (very) cheap multiple camera image acquisition system using, e.g. web cameras connected to a portable PC. A precise synchronization of the cameras is not possible and therefore high accuracy of the measurement cannot be achieved. Still, the great demonstration potential of a portable measurement system would be very attractive.

On the other hand, the requirement of motion data processing ability is getting important

today. Therefore how to develop the proposed method to cope with motion data will also be our future work. For instance, in the foot shape reconstruction system, the exploited method can be achieved, refining the initial model to fit the first frame of images sequence by MASM. Then sample points are tracked in rest frames. Thus the dynamic behavior can be described by the changing of sample points' position. However, low-importance shape variation modes of MASM for static 3D model may become pivotal elements for motion simulation. Thus, how to decide the necessary shape variation modes momentarily will be the most important issue in motion data processing.

Appendices

Appendix A Establishment of Triangle Patches by Control Points

In computer, 3D objects are always composed of triangle patches. Each triangle is established by two nearest neighbors and current control point $\boldsymbol{v}_1$ (Figure A.1). The triangle plane is determined by a 4×4 matrix

$$A = plane(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3) = \begin{pmatrix} x & y & z & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{pmatrix} \tag{A.1}$$

where $\boldsymbol{v}_i = (x_i, y_i, z_i)^T$

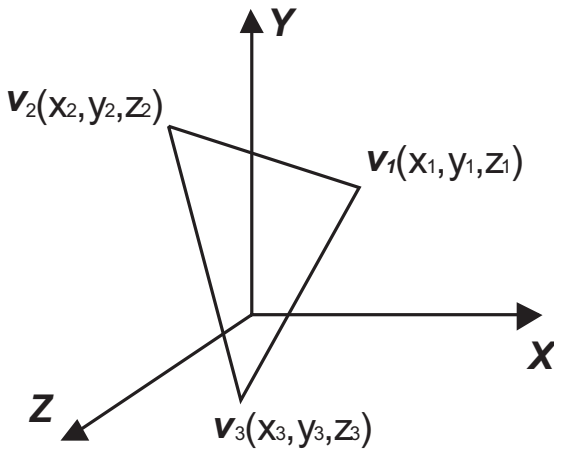


Figure A.1: A plane defined by three points (not in a line) in 3D space.

Then points inside the triangle can be estimated by exploring minimum to maximum of arbitrary two axes among X, Y and Z ((A.2)).

$$|A| = \begin{vmatrix} x & y & z & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{vmatrix} \Rightarrow \begin{vmatrix} x - x_1 & y - y_1 & z - z_1 \\ x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ x_3 - x_1 & y_3 - y_1 & z_3 - z_1 \end{vmatrix} = 0 \tag{A.2}$$

Appendix B Levenberg-Marquardt Iterative Optimization Technique

Levenberg-Marquardt[60] is a popular alternative to the Gauss-Newton method of finding the minimum of a function $F(\boldsymbol{x})$. The optimal variable vector $\boldsymbol{x}^*$ is estimated by iterative search ((B.1)) from the initial vector $\boldsymbol{x}$. It is assumed that $\boldsymbol{x}$ is a 3D vector, $\boldsymbol{x} = (x, y, z)$.

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k \tag{B.1}$$

where α_k is the search step and search direction is decided by vector $\boldsymbol{d}$.

$$\boldsymbol{d}_k = -\boldsymbol{H}_k \nabla F(\boldsymbol{x}_k) \tag{B.2}$$

where

$$\nabla F(\boldsymbol{x}_k) = \left(\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}, \frac{\partial F}{\partial z} \right) \tag{B.3}$$

Then $\boldsymbol{H}$ is transform matrix

$$\boldsymbol{H}_k = \nabla^2 F(\boldsymbol{x}_k) + \mu_k \boldsymbol{I} \tag{B.4}$$

where $\boldsymbol{I}$ is an unit matrix, μ is a positive real number. μ is approaching 0, while $\boldsymbol{x}$ is becoming optimal. Thus, μ is also a search step controller. $\nabla^2 F(\boldsymbol{x}_k)$ is called Hessen matrix whose expression is in (B.5)

$$\nabla^2 F(\boldsymbol{x}_k) = \begin{pmatrix} \frac{\partial^2 F}{\partial x^2} & \frac{\partial^2 F}{\partial x \partial y} & \frac{\partial^2 F}{\partial x \partial z} \\ \frac{\partial^2 F}{\partial y \partial x} & \frac{\partial^2 F}{\partial y^2} & \frac{\partial^2 F}{\partial y \partial z} \\ \frac{\partial^2 F}{\partial z \partial x} & \frac{\partial^2 F}{\partial y \partial z} & \frac{\partial^2 F}{\partial z^2} \end{pmatrix} \tag{B.5}$$

α in (B.1) is estimated by the following inequations, which are called Wolfe's conditions.

$$F(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k) - F(\boldsymbol{x}_k) \leq \sigma_1 \alpha (\nabla F(\boldsymbol{x}_k))^T \boldsymbol{d}_k \tag{B.6}$$

$$(F(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k) \boldsymbol{d}_k \geq \sigma_2 (\nabla F(\boldsymbol{x}_k))^T \boldsymbol{d}_k \tag{B.7}$$

Appendix C Rosenbrock Method

The Rosenbrock method is a 0th order search algorithm (it means it does not require any derivatives of the target function. Only simple evaluations of the objective function are used). However, it approximates a gradient search thus combining advantages of 0th order and 1st order strategies. It was published by Rosenbrock[63] in the 70th.

This method is particularly well suited when the objective function does not require a great deal of computing power.

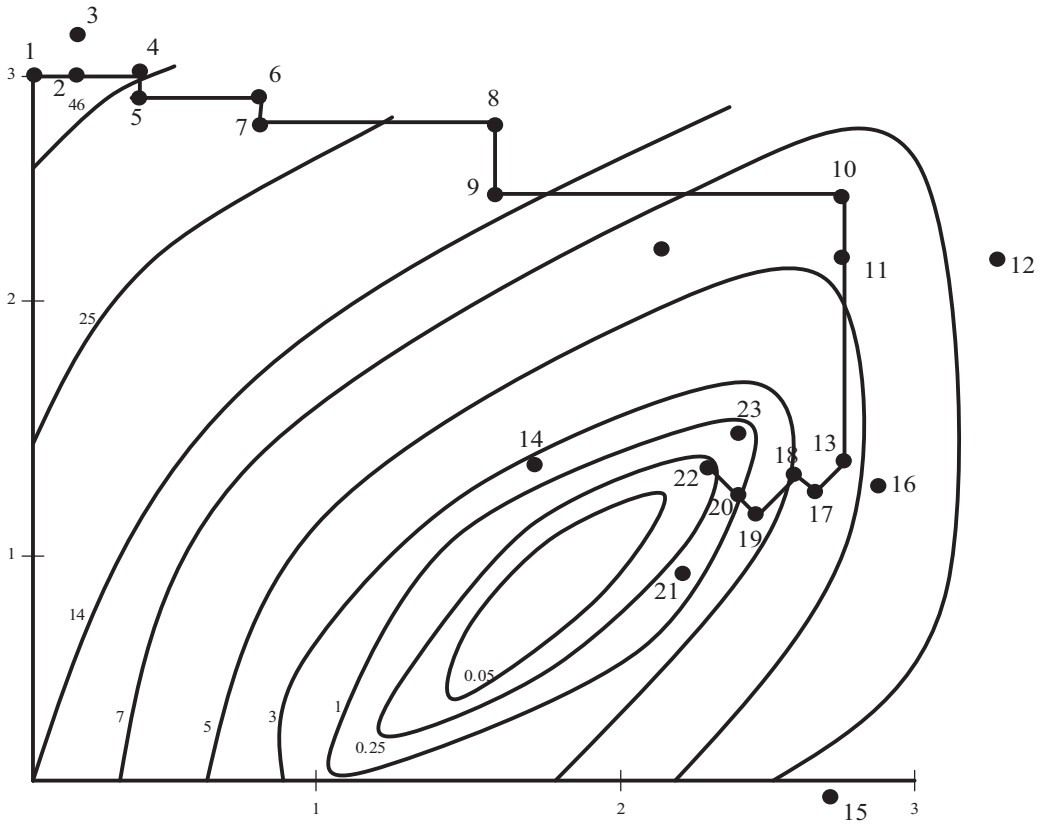


Figure C.1: The iterative processing of Rosenbrock algorithm.

Rosenbrock method is an iterative optimization algorithm. In the first iteration, it is a simple 0th order search in the directions of the base vectors of an n-dimensional coordinate system. In the case of a success, which is an attempt yielding a new minimum value of the target function, the step width is increased, while in the case of a failure it is decreased and

the opposite direction will be tried (see points 1 to 16 in Figure C.1). Once a success has been found and exploited in each base direction, the coordinate system is rotated in order to make the first base vector point into the direction of the gradient (the points 13,16 and 17 are defining the new base). Now all step widths are initialized and the process is repeated using the rotated coordinate system (points 16 to 23).

The Rosenbrock algorithm has also been proved to always converge[9] (global convergence to a local optima assured). Initializing the step widths to rather big values enables the strategy to leave local optima and to go on with search for more global minima. It has turned out that this simple approach is more stable than many optimization algorithms and it requires much less calculations of the target function than higher order strategies.

Appendix D Occlusion Assessment of Multiple Camera System

Because the intensity of sample points' projection on multiple camera images is used to evaluate the parameters of MASM, the occlusion issue has to be cared. In Figure D.1 the solid line triangles represent the patches that faces the camera O and dash line triangles represent the patches on the reverse surface to the camera. The patches on the reverse surface are invisible, i.e. the vertices (sample points) of the invisible triangle patches are occluded. If the occluded sample point, for instance F is projected onto the image plane, the corresponding point H on the image plane don't represent F, but G. G is the intersection point of the sight line and a triangle patch. If H is still considered to be the projection of F, the intensity feature of G is used actually. Thus, errors are occurred imaginably.

In this work, to address the issue of occlusion, the spatial relationships between each sample point and all the triangle patches are investigated. First, the relationships are checked in 2D image plane. The relationships can be partitioned into four categories:

- Inside: A sample point's projection inside a triangle patch;
- Outside: A sample point's projection outside a triangle patch;
- Border: A sample point's projection is on the border of a triangle patch;
- Superposition: A sample point's projection superposes a vertex of a triangle patch.

Furthermore, if it is assumed that "Border" comes under "Inside" and "Superposition" comes under "Outside", there remains two relationship classes: Inside and Outside (Figure D.2). "Inside" and "Outside" can be simply determined by cosine theorem. In Figure D.3 sample point D is inside a triangle $\triangle ABC$, according to the cosine theorem,

$$\angle ADB = arccos(\frac{AD^2 + BD^2 - AB^2}{2AD \times BD}) \tag{D.1}$$

$$\angle ADC = arccos(\frac{AD^2 + CD^2 - AC^2}{2AD \times CD}) \tag{D.2}$$

$$\angle BDC = arccos(\frac{BD^2 + CD^2 - BC^2}{2CD \times BD}) \tag{D.3}$$

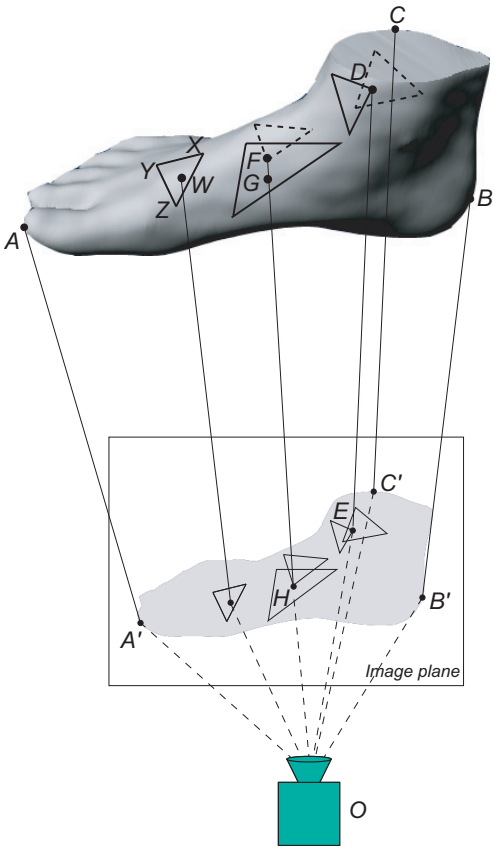


Figure D.1: From camera O, the occlusion is assessed by investigating the relationships of sample points and triangle patches.

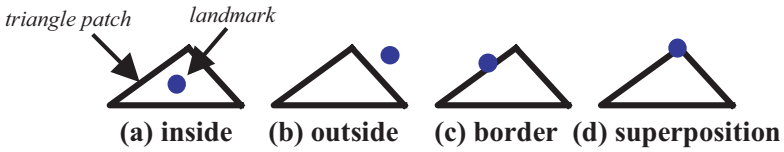


Figure D.2: Relationships of the projection of sample point and the projection of a triangle patch.

$\because$ D and $\triangle ABC$ are coplane

$\therefore \angle ADB + \angle ADC + \angle BDC = 360°$.

On the other hand, Sample point E is outside $\triangle ABC$,

$\because \angle AEC$ is obviously less than $180°$

$\because \angle AEB + \angle BEC = \angle AEC$

$\therefore \angle AEC + \angle AEB + \angle BEC < 360°$

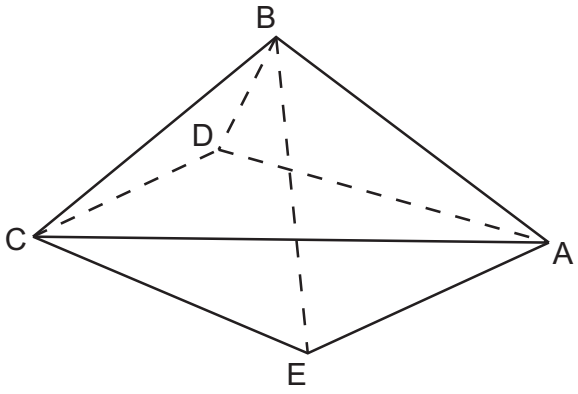


Figure D.3: The inside/outside relationship of a point and a plane in 2D.

Thus, the inside and outside relationships of sample points and triangle patches on image plane can be assessed by the algorithm of List A.1.

List A.1
if $\angle ADC + \angle ADB + \angle BDC < 360°$ Then
D is outside;
else
D is inside;

For "Outside", the sample point will not be occluded, but for "Inside", the problem is complex. In Figure D.1, the projection of occluded vertex F is inside a triangle patch's projection. On the other hand, although D's projection is also inside a triangle patches' projection, D is obviously not occluded. Thus, further investigation is needed.

A plane specified in three-point A, B, C (Figure D.4) form can be given in terms of the general equation (D.4) by

$$C_1x + C_2y + C_3z + C_4 = 0 \tag{D.4}$$

Curve l passes through the points $M_0(x_0, y_0, z_0)$, $M_1(x_1, y_1, z_1)$, intersects $\triangle ABC$ in a point $M_2(x_2, y_2, z_2)$, which can be determined by solving the four simultaneous equations (D.4)$\sim$ (D.7)

$$x = x_0 + (x_0 - x_1)t \tag{D.5}$$

$$y = y_0 + (y_0 - y_1)t \tag{D.6}$$

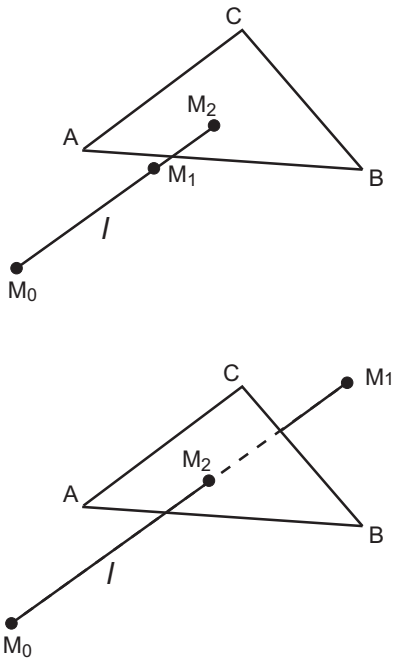


Figure D.4: If M_1 is not occluded by the plane $\triangle ABC$, $M_0M_1 < M_0M_2$. Otherwise $M_0M_1 > M_0M_2$.

$$z = z_0 + (z_0 - z_1)t \tag{D.7}$$

for x_2, y_2, z_2 and t. Then the Euclidean distance of M_0M_1 and M_0M_2 are given by (D.8) and (D.9) respectively.

$$M_0M_1 = \sqrt{(x_0 - x_1)^2 + (y_0 - y_1)^2 + (z_0 - z_1)^2} \tag{D.8}$$

$$M_0M_2 = \sqrt{(x_0 - x_2)^2 + (y_0 - y_2)^2 + (z_0 - z_2)^2} \tag{D.9}$$

If $M_0M_1 < M_0M_2$, M_1 is not occluded by $\triangle ABC$ (top of Figure D.4). Otherwise M_1 is occluded (bottom of Figure D.4). Therefore, in Figure D.1, to assess whether $W(x, y, z)$ is occluded by $\triangle XYZ$. The intersection point W' of OW and $\triangle XYZ$ is estimated. Then by comparing the length of OW and OW': if $OW > OW'$ then W is occluded, else W is not occluded

For attentively, if the sample points' projections are on the contour of all the sample points' projection area, these sample points' projection are inside no triangle patch's projection. This

kind of sample points are so called “contour sample points” (section 4.5). In Figure D.1 sample points A, B, and C are contour sample points.

Appendix E Downhill Simplex Method

The downhill simplex method (DSM) is due to Nelder and Mead[54]. The method requires only function evaluations, not derivatives. It is not very efficient in terms of the number of function evaluations that it requires. However the downhill simplex method may frequently be the best method to use. In case of many dimensions (more than 20) the function sometimes does not converge to the minimum but the simplex is constantly shrinking. The detail of simplex method will be described bellow.

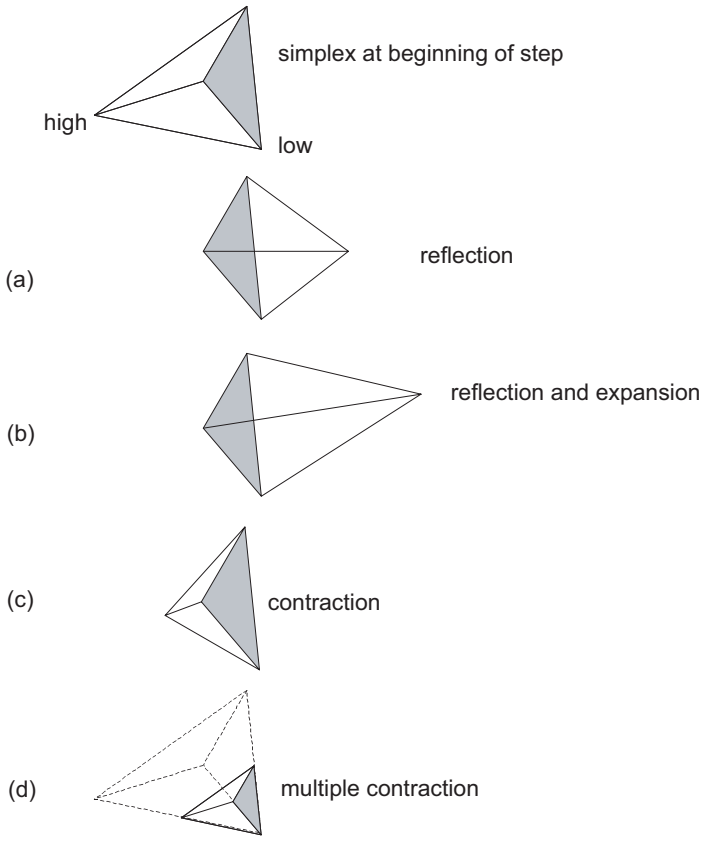


Figure E.1: Possible outcomes for a step in the downhill simplex method. The simplex at the beginning of the step, here a tetrahedron, is shown, top. The simplex at the end of the step can be any one of (a) a reflection away from the high point, (b) a reflection and expansion away from the high point, (c) a contraction along one dimension from the high point, or (d) a contraction along all dimensions towards the low point. An appropriate sequence of such steps will always converge to a minimum of the function.

A simplex is defined as a figure of N + 1 vertices in the N-dimensional search space (a tetrahedron in the 3 dimensional space). Each simplex defines a solution in the search space. The simplex can be expanded, contracted, and reflected. A contraction is $x = x_\sigma - \lambda$; an expansion is $x = x_\sigma + \lambda$; a reflection is $x = -x_\sigma$. There are of course several combinations of the above. The DSM takes a series of random steps as follows. First, it finds the point where the objective function is highest (high point) and lowest (low point). Then it reflects the simplex around the high point. If the solution is better, it tries an expansion in that direction, else if the solution is worse than the second-highest point it tries an intermediate point. If no improvement is found after a number of steps, the simplex is contracted, and started again. The idea of DSM is illustrated in Figure E.1 briefly.

An appropriate sequence of such steps will always lead to a minimum. Better results are obtained when large steps are tried.

Acknowledgements

I would like to thank Associate Prof. Hideo Saito at the Department of Information and Computer Science Faculty of Science and Technology Keio University for giving me the opportunity to work on interesting projects and for providing me with excellent research environment.

I thank Professor Shinji Ozawa who gave me many helpful advices from which I started my research work during my master course.

I thank Mr.Kuwahara, Mr. Yamashita and Dr.Takahashi from Shiseido research center in Yokohama for providing me with confocal microscope images. I thank Dr.Kimura, Dr.Mochimaru, Professor Kanade from National Institute of Advanceed Industrial Science and Technology Digital Human Laboratory in Tokyo for giving me kind advices for this work.

I thank Mr.Dong Han for his assistance in the experiments of my research work.

I thank all the members of Saito Laboratory of Keio University for their assistance in various technical fields and for their friendships.

Special thank are due to my wife Weiyi Zhao and my daughter Wanyun Wang. They give me strong inspirit from spirit and life.

I thank from heart my father, mother, sister, brother-in-law, and grandfather for their supports and love.

Finally, I thank my grandmother in heaven for her nurturance from my childhood.

August, 2005

Jiahui Wang

References

[1] Y. Abdel-Aziz and H. Karara, “Direct linear transformation from comparator coordinates into object space coordinates in close range photogrammetry,” In *Proc. ASP/UI Symposium on Close-Range Photogrammetry*, pp.1-18, 1971.

[2] A. Amini, T. Weymouth and R. Jain, “Using dynamic programming for solving variational problems in vision,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.12, no.9, pp.855-867, 1990.

[3] Y. Amit, U. Grenander and M. Piccioni, “Structural image restoration through deformable template,” Journal of American Statistical Association, vol.86, no.414, pp.376-387, 1991.

[4] W. Amos, J. White and M. Fordham “Use of confocal imaging in the study of biological structures,” Applied Optics, vol.26, pp.3239-3243, 1987.

[5] N. Aslund, K. Carlsson, A. Liljeborg and L. Majlof, “PHOIBOS, a microscope scanner designed for micro-fluorometric application, using laser induced fluorescence,” In *Proc. 3rd. Scandinavian Conference on Image Analysis, Student literature, Lund,* pp.338, 1983.

[6] N. Aslund, A. Liljeborg, P. Forsgren and S. Wahlsten, “Three dimensional digital microscopy using the PHOIBOS scanner,” Scanning, pp.227-235, 1987.

[7] B. Bascle and R. Deriche, “Region tracking through image sequences,” In *Proc. International Conference on Computer Vision*, pp.302-307, 1995.

[8] R. Bellman and S. Dreyfus, *Applied dynamic programming*, Princeton University Press, 1962.

[9] M. Bazaraa, H. Sherali and C. Shetty, *Nonlinear programming: Theory and algorithms*, John Wiley and Sons, 1993.

[10] A. Blake and A. Zisserman, *Visual reconstruction*, MIT Press, 1987.

[11] M. Bomans, K. Hohne, U. Tiede and M. Riemer, “3-D segmentation of MR images of the head for 3-D display,” IEEE Transactions on Medical Imaging, vol.9, no.2, pp.177-183 Jun,1990.

[12] G. Borgefors, “Distance transformations in arbitrary dimensions,” Computer Vision, Graphics and Image Processing, vol.27, pp.321-445, 1984.

[13] G. Brakenhoff, P. Blom and P. Barends, “Confocal scanning light microscopy with high aperture immersion lenses,” Journal of Microscopy, vol.117, pp.219-232, 1979.

[14] G. Brakenhoff, H. Voort, E. Spronsen, W. Linnemans and N. Namminga, “Three dimensional chromatin distribution in neuroblastoma nuclei shown by confocal scanning laser microscopy,” Nature, vol.317, pp.748-749, 1985.

[15] N. Caille, Y. Tardy and J. Meister, “Assessment of strain field in endothelial cells subjected to uniaxial deformation of their substrate,” Annals of Biomedical Engineering, vol.26, pp.409-416, 1998.

[16] K. Carlsson, P. Danielsson, R. Lenz, A. Liljeborg, L. Majlof and N.Aslund, “Three-dimensional microscopy using a confocal laser scanning microscope,” Optics Letters, vol.10, pp.53-55, 1985.

[17] L. Cohen, “On active contour models and balloons,” Computer Vision Graphics and Image Processing: Image Understanding, vol.53, no.2, pp.211-218, 1991.

[18] L. Cohen and I. Cohen. “Finite-element methods for active contour model and balloons for 2-D and 3-D images,” IEEE Transactions on Pattern Analysis and Machine Intelligence. vol.15, no.11, pp.1131-1147, 1993.

[19] T. Cootes, A. Hill, C. Taylor, and J. Haslam, “The use of active shape models for locating structures in medical images,” Image and Vision Computing, vol.12, no.6, pp.355-366, 1994.

[20] T. Cootes, C. Taylor, D. Cooper and J. Graham, “Active shape models-their training and application,” Computer Vision and Image Understanding, vol.61, no.1, pp.38-59, 1995.

[21] T. Cootes, G. J. Edwards, and C. Taylor, “Active appearance models,” In *Proc. European Conference on Computer Vision*, pp.484-498, 1998.

[22] T. Cootes, C. Beeston, G. Edwards, and C. Taylor, “A unified framework for atlas matching using active appearance models,” In *Proc. Information Processing in Medical Imaging*, pp.323-333, 1999.

[23] P. Danielsson, “Euclidean distance mapping,” Computer Graphics and Image Processing, vol.14, pp.227-248, 1980.

[24] Y. Duan and H. Qin, “2.5D active contour for surface reconstruction,” In *Proc. 8th International Workshop on Vision, Modeling and Visualization*, pp.431-439, 2003.

[25] N. Duta and M. Sonka, “Segmentation and interpretation of MR brain images: An improved active shape model,” IEEE Transactions on Medical Imaging, vol.17, no.6, pp.1049-1062, 1998.

[26] G. Edwards, C. Taylor, and T. Cootes, “Interpreting face images using active appearance models,” In *Proc. International Conference on Automatic Face Gesture Recognization*, pp.300-305, 1998.

[27] M. Egger, “The development of confocal microscopy,” Trends in Neurosciences, vol.12, pp.11, 1989.

[28] M. Fischler and R. Elschlager, “The representation and matching of pictorial structures,” IEEE Transactions on Computers, vol. 22, no. 1, pp. 67-92, 1973.

[29] D. Forsyth and J. Ponce, *Computer vision: A modern approach*, Prentice Hall, 2003.

[30] M. Gniadecka and G. Jemec, “Quantitative evaluation of chronological ageing and photoaging in vivo: Studies on skin echogenicity and thickness,” British Journal of Dermatology, vol.139, pp.815-821, 1998.

[31] U. Grenander, Y. Chow, and D. Keenan, *Hands: A pattern theoretic study of biological shapes*, Springer-Verlag, 1991.

[32] R. Haralick, “Digital step edges from zero crossing of second directional derivatives,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.PAMI-6, no.1, pp.58-68, 1984.

[33] T. Hayashi, K. Mimura and Y. Nishijima, “Changes in surface configuration of skin caused by aging and application of cosmetics. Three-dimensional analysis according to new system based on image analysis and Fourier transform,” Journal of Society Cosmetic Chemists Japan, vol.23, no.1, pp.43-55, 1989.

[34] A. Hill, A. Thornham, and C. Taylor, “Model-based interpretation of 3D medical images,” In *Proc. British Machine Vision Conference*, pp.339-348, 1993.

[35] http://www.dh.aist.go.jp/research/awaseru/infoot/index.html.en

[36] A. Jain, Y. Zhong and M-P. Dubuisson-Jolly, “Deformable template models: A review,” Signal Processing 71, pp.109-29. 1998.

[37] A. Jain and D. Zongker, “Representation and recognition of handwritten digits using deformable templates,” IEEE Transaction on Pattern Analysis and Machine Intelligence, vol.19, no.12, pp.1386-1390, 1997.

[38] M. Kass, A. Witkin and D. Terzopoulos, “Snakes: Active contour models,” International Journal of Computer Vision, vol.1, no.4, pp.321-331, 1988.

[39] M. Kimura, M. Mochimaru, M. Kauchi, T. Kanade and H. Saito, “Measurement of ball shape using stereo vision method,” In *Proc. 66th SICE-Pattern Measurement*, pp.1-5, 2004.

[40] B. Klause and P. Horn, *Robot vision*, 1986.

[41] M. Kuehn, M. Hausner, H. Bungartz, M. Wagner, P. Wilderer and S. Wuertz, “Automated confocal laser scanning microscopy and semiautomated image processing for analysis of biofilms,” Applied and Environmental Microbiology, vol.64, no.11, pp.4115-4127, 1998.

[42] M. Kyan, L. Guan, M. Arnison and J. Cogswee, “Feature extraction of chromosomes from 3-D confocal microscope images,” IEEE Transaction on Biomedical Engineering, vol.48, no.11, pp.1306-1318, 2001.

[43] C. Lasagni and S. Seidenari, “Echographic assessment of age-dependent variations of skin thickness. A sturdy on 162 subjects,” Skin Research and Technology, vol.1, pp.74-80, 1995.

[44] A. Laurentini, “The visual hull concept for silhouette based image understanding,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.16, no.2, pp.150-162, 1994.

[45] L. Lee, R. Romano and G. Stein, “Monitoring activities from multiple video streams: Establishing a common coordinate frame,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.22, no.8, pp.758-767, 2000.

[46] http://www.library.co.jp/Eng/index.html

[47] W. Lorensen and H. Cline, “Marching cubes: A high resolution 3D surface construction algorithm”, Computer Graphics, vol.21, no.4, pp.163-169, 1987.

[48] R. Maksimovic, S. Stankovic and D. Milovanovic, “Computed tomography image analyzer: 3D reconstruction and segmentation applying active contour models-‘snakes’,” International Journal of Medical Information, vol.58-59, pp.29-37, 2000.

[49] D. Marr and E. Hildreth, “Theory of edge detection,” Proceeding of Royal Society London, B, vol.207, pp.187-217, 1980.

[50] M. Miller, G. Christensen, Y. Amit, and U. Grenander, “Mathematical textbook of deformable neuroanatomies,” In *Proc. National Academy of Science*, vol.90, pp.11944-11948, 1993.

[51] M. Minsky, “Memoir on inventing the confocal scanning microscope,” Scanning, vol.10, pp.128-138. 1988.

[52] S. Mitchell, B. Lelieveldt, R. Geest, H. Bosch, J. Reiber, and M. Sonka, ”Multistage hybrid active appearance model matching: Segmentation of left and right ventricles in cardiac MR images,” IEEE Transactions on Medical Imaging, vol.20, no.5, pp.415-423, 2001.

[53] S. Mitchell, J. Bosch, B. Lelieveldt, R. Geest, J. Reiber and M. Sonka, “3-D active appearance models: Segmentation of cardiac MR and ultrasound images,” IEEE Transactions on Medical Imaging, vol.21, no.9, pp.1167-1178, 2002.

[54] J. Nelder and R. Mead, “A simplex method for function minimization,” Computer Journal, vol.7, no.4, pp.308-313, 1965.

[55] N. Nikolaidis and I. Pitas, *3-D image process algorithms*, A Wiley-Interscience Publication John Wiley and Sons, Inc. 2001.

[56] N. Nishida, S. Fukushima and M. Minoh, “A method of estimating human shape by fitting the standard human model to partial measured data,” In *Proc. Asian Conference on Computer Vision*, pp.276-281, 2000.

[57] K. Ogawa and F. Mizoguchi, *Histology*, Bunkodo, 1993.

[58] L. Olsen, H. Takiwaki and J. Serup, “High-frequency ultrasound characterization of normal skin. Skin thickness and echographic density of 22 anatomical sites,” Skin Research and Technology, vol.1, pp.81-85, 1995.

[59] J. Pawley, *Handbook of biological confocal microscope (SE)*, Plenum Press, 1995.

[60] A. Peressini, F. Sullivan and J. Uhl.Jr, *The mathematics of nonlinear programming*, Spring-Verlag, 1988.

[61] http://www.povray.org/

[62] J. Roebuck, Jr., *Anthropometric method: Designing to fit the human body*, Human Factors and Ergonomics Society, 1995.

[63] H. Rosenbrock, “An automatic method for finding the greatest or least value of a function,” The Computer Journal, vol.3, no.3, pp.174-184, 1960.

[64] A. Rosenfeld and A. Kak, *Digital picture processing*, Academic Press, New York, 1976.

[65] J. Russ, *The image processing handbook*, CRC Press, Boca Raton, 1994.

[66] M. Schrader and S. Hell, “Three-dimensional super-resolution with a 4pi-confocal microscope using image restoration,” Journal of Applied Physics, vol.84, no.8, pp.4033-4042, 1998.

[67] C. Sheppard and A. Choudhury, “Image formation in the scanning microscope,” Optical, vol.24, pp.1501, 1977.

[68] C. Sheppard, J. Gannaway, D. Walsh and T. Wilson, “Scanning optical microscope for the inspection of electronic devices,” In *Proc. Microcircuit Engineering Conference*, Cambridge, 1978.

[69] M. Takahashi, “Recent advances in skin bioengineering techniques,” Journal of Japanese Cosmetic Science Society, vol.23, no.4, pp,312-322, 1999.

[70] H. Tek and B. Kimia, “Volumetric segmentation of medical images by three-dimensional bubbles,” Computer Vision and Image Understand, vol.65, pp.246-258, 1997.

[71] D. Terzopoulos, A. Witkin and M. Kass, “ Symmetry-seeking models for 3D object reconstruction,” International Journal of Computer Vision, vol.1, no.3, pp.211-221, 1987.

[72] D. Terzopoulos, “On matching deformable models to images,” Technical Report 60, Schlumberger Palo Alto research, 1986. Reprinted in Topical Meeting on Machine Vision, Technical Digest Series, vol.12, pp.160-167, 1987.

[73] D. Terzopoulos and K. Fleischer, “deformable models,” The visual computer, vol.4, pp.306-331, 1988.

[74] D. Terzopoulos, A. Witkin, and M. Kass, “Constraints on deformable models: Recovering 3D shape and nonrigid motion,” Artificial Intelligence, vol.36, no.1, pp.91-123, 1988.

[75] D. Terzopoulos, “The computation of visible surface representations,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.PAMI-10, no.4, pp.417-438, 1988.

[76] S. Ullman. *High-level vision-object recognition and visual cognition*, MIT Press, 1996.

[77] J. Wang, H. Saito, S. Ozawa, T. Kuwahara, T. Yamashita and M. Takahashi, “3D visualization of skin inner tissue from confocal microscopic images,” In *Proc. International Conference on Quality Control by Artificial Vision*, 2003.

[78] J. Wang, H. Saito, S. Ozawa, T. Kuwahara, T. Yamashita and M. Takahashi, “Extraction of dermo-epidermal surface from 3-D volumetric images of human skin,” International Journal of Image and Graphics, vol.3, no.4, pp.589-608, 2003.

[79] J. Wang, H. Saito, S. Ozawa, T .Kuwahara, T. Yamashita and M. Takahashi, “Development of a new method for extracting surface with open form from volumetric images,” Journal of Computer Research and Development, vol.41, no.7, pp.1174-1181, 2004.

[80] J. Wang, H. Saito, M. Kimura, M. Mochimaru and K. Kanade, “Reconstruction of human foot from multiple camera images with 3D active contour models,” In *Proc. Asia International Symposium on Mechatronics*, pp.390-395, 2004.

[81] J. Wang, H. Saito, M. Kimura, M. Mochimaru and K. Kanade, “Shape reconstruction of human foot from multi-camera images based on PCA of human shape database,” In *Proc. International Conference on 3-D Imaging and Modeling*, 2005.

[82] J. Wang, H. Saito, S. Ozawa, T. Kuwahara, T. Yamashita and M. Takahashi, “Surface extraction of dermo-epidermal boundary from 3D volumetric images of human skin tissue,” Technical Report of IEICE. PRMU 2002-4, MI2002-20, May., 2002.

[83] J. Wang, H. Saito, S. Ozawa, T. Kuwahara, T. Yamashita and M. Takahashi, “ 3D visualization of skin inner tissue from confocal microscopic images,” In *Proc. 14th Visual Inspection Engineering Workshop*, pp.107-112, 2002.

[84] J. Wang, H. Saito, S. Ozawa, T. Kuwahara, T. Yamashita and M. Takahashi, “Surface extraction of skin inner tissue interface from 3D volumetric images of human skin via active contour model,” IEEJ Transactions on Electronics, Information and System, vol.125-C, no.5, pp.756-765, 2005.

[85] J. Wang, H. Saito, M. Kimura, M. Mochimaru and K. Kanade, “Reconstruction of human foot from multiple camera images,” In *Proc. 10th Symposium on Sensing via Image Information*, pp.317-322, 2004.

[86] Y. Wang and L. Staib, “Boundary finding with correspondence using statistical shape models,” In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 338-345, 1998.

[87] S. Weik, “A passive full body scanner using shape from silhouettes,” In *Proc. International Conference on Computer Vision*, pp.1750-1753, 2000.

[88] J. White, W. Amos and M. Fordham, “An evaluation of confocal versus conventional imaging of biological structures by fluorescence light microscopy,” Journal of Cell Biology, vol.105, pp.41-48, 1987.

[89] R. van Resandt Wijnaendts, H. Marsman, R. Kaplan, J. Davoust, E. Stelzer and R. Strickler, “Optical fluorescence microscopy in three dimensions: Microtomoscopy,” Journal of Microscopy, vol.138, pp.29-34, 1985.

[90] D. Williams and M. Shah, “ A fast algorithm for active contours and curvature estimation,” CVGIP: Image Understanding, vol.55, no.1, pp.14-26, 1992.

[91] T. Wilson, J. Gannaway and P. Johnson, “A scanning optical microscope for the inspec-

tion of semiconductor materials and devices," Journal of Microscopy, vol.118, pp.314-390, 1980.

[92] B. Widrow, "The rubber-mask technique," Pattern Recognition, vol.5, pp.175-211, 1973.

[93] C. Xu and J. Prince, "Snakes, shapes, and gradient vector flow," IEEE Transactions on Image Processing, vol.7, no.3, pp.359-369, 1998.

[94] C. Xu and J. Prince, "Generalized gradient vector flow external forces for active contours," Signal Processing-An International Journal, vol.71, no.2, pp.131-139, 1998.

[95] J. Zheng, "Acquiring 3-D models from sequences of contours," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.16, no.2, pp.163-178, 1994.

[96] L. Zhang, N. Snavely, B. Curless, and S. Seitz, "Spacetime faces: High-resolution capture for modeling and animation," In *Proc. ACM SIGGRAPH*, pp.548-558, 2004.

[97] Y. Zhong and A. Jain, "Object localization using color, texture and shape," In *Proc. Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, pp.279-294, 1997.

[98] Z. Zhang. "Flexible camera calibration by viewing a plane from unknown orientations." In *Proc. International Conference on Computer Vision*, pp.666-673, 1999.