

# Body Trunk Shape Estimation from Silhouettes by Using Homologous Human Body Model

Shunta Saito<sup>\*a</sup>, Makiko Kochi<sup>b</sup>, Masaaki Mochimaru<sup>b</sup>, Yoshimitsu Aoki<sup>a</sup>

<sup>a</sup>Keio University, Yokohama, Kanagawa, Japan;

<sup>b</sup>National Institute of Advanced Industrial Science and Technology, Odaiba, Tokyo, Japan

## Abstract

This paper attempts to estimate a 3D shape of human torso from 2 pictures which are front and side silhouette of a subject. Our method assumes that input pictures have accurate contours of a subject's body. And regions of torsos are extracted manually from input pictures. We use a deformable body model for estimation. This model is generated by principal component analysis to a homologous model database which is made of whole body scanned data. A result shape is computed by optimizing some deformation parameters of the model with comparing model's silhouettes and input silhouettes.

**Keywords:** 3D shape estimation, silhouette, homologous model, PCA

## 1. Introduction

Designing of industrial products often uses measurements of human body part sizes. Recently, the useful design for any users, which is called "universal design", is attracting attention from many product makers. Designers are required to consider the usability in many scenes of product use by various users. Mochimaru et al.[1] developed Dhaiba, a virtual human model to simulate of scene of product use by various users. It uses statistical processing results of many human body shapes. To analyze body shape database statistically, each body shape data is represented as a homologous model which is the statistical friendly format for a human body shape. Homologous models consist of the same number of the same topology, and each data point is defined based on the anatomical homology. However, collecting many human body shape data as homologous models requires many full-body scan using a big and expensive equipment. Therefore, to collect more numerous human body shape data, it is important to establish an easier alternative technology to measure a human body shape and represent the result as a homologous model.

Then, if a human body shape can be acquired by easy method, it isn't required to measure some sizes of user's body parts directly by using particular measurement instruments for making custom-made products, and it will be easier to make universal design products. These needs raise expectations for an easier alternative technology for human body shape measurement.

Thus, we propose an approach to estimate a human body shape from only 2 pictures: front and side silhouette of a subject. Using the deformable model consisted of homologous models database, we attempts to solve the human body trunk shape estimation problem as some optimization problems of deformation parameters.

## 2. Related Works

There are some related works addressing 3D shape reconstruction of human body parts. Methods using range scanners[2] or projectors and cameras[3][4] can reconstruct the object's 3D shape with accuracy. However, these methods require a dedicated hardware set and specialistic preparation (ex. projector-camera calibration), so that they cannot be said easy for personal use. On the other hand there are researches of model-based shape estimation which use previous knowledge of the target object. Suzuki et al[5] proposed a method using Active Shape Model (ASM)[6] for human face reconstruction from some silhouette images. ASM is a technic to generate a deformable model by principal component analysis (PCA) to a database. The database is consisted of the target object's shape data which are represented in the same topology and as point clouds which have unified number of points. Therefore, ASM is appropriate to deal with the shapes without a topological difference among individual shape like human body part. However, although It is difficult to shut up countenance completely when a camera captures the subject, the method in [5] treats the shape estimation problem as the problem only from shape variation and model configuration, so that it cannot consider individual shape deformation caused by facial expressions. This is an issue especially in the case of body trunk shape estimation. Shape deformation caused by facial expressions is deformation

\* [ssaito@aoki-medialab.org](mailto:ssaito@aoki-medialab.org), +81-45-566-1796, [www.aoki-medialab.org](http://www.aoki-medialab.org)

of facial muscle between skin and bone. Therefore, when we think of body trunk shape estimation, normally, the deformation caused by tension of muscle under skin. However, the modeling of muscle under skin is exceedingly difficult. Furthermore, there is another problem in the case of body trunk shape estimation. When a camera captures the subject's face, it is easy to standardize the bone posture to ignore shape deformation caused by postural change. However, in the case of body trunk, it is not ignorable cause of shape deformation. Namely, we treat the problem of the human body trunk shape estimation as a conjuncture of 3 problems: (1) shape estimation by using ASM; (2) model configuration identification; (3) body posture estimation. Then, we use 2 silhouette images captured from front and side of human body.

### 3. Proposed Method

Our purpose is to develop an easy method for body trunk shape estimation. Thus, inputs to our system are only 2 pictures. These are silhouettes of front and side of target body by using 2 cameras. We use the deformable model generated from homologous model database. This model is optimized to minimize the difference between model's and inputs' silhouettes. Figure 1 shows our method framework.

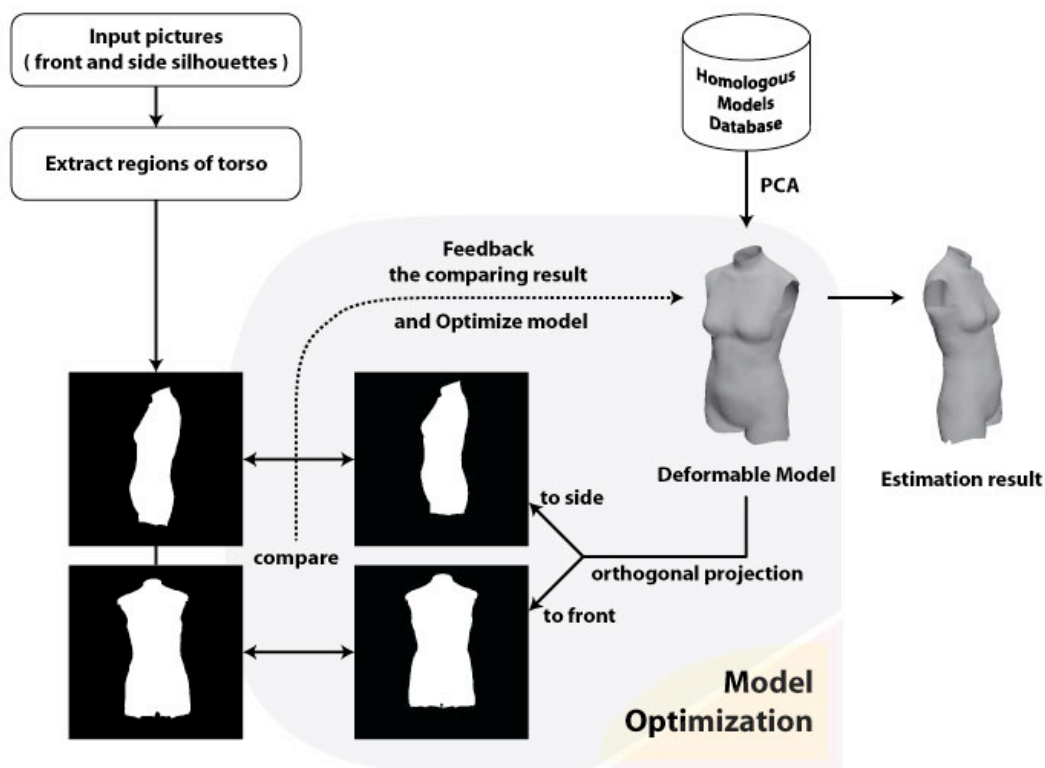


Fig. 1. Our system framework

#### 3.1 Input Silhouettes

To determine model configuration approximately, we assume that 2 cameras directions are at right angles to each other and the target body's front direction is parallel to the front camera's direction (see figure 2). Then, the body trunk region is extracted from a input full-body silhouette image. Then, the body trunk silhouette image is obtained. The body trunk region is defined with some anatomical feature points on human body as seen figure 3. In the optimizing stage (as will hereinafter be described in detail), contour is important to fit the model to inputs, so that, ideally, the input pictures have accurate contours of a subject's body. Therefore, the subject person is desirable to be naked when captured. However, it is just not realistic and the subject dressed in body fitting clothes as far as possible.

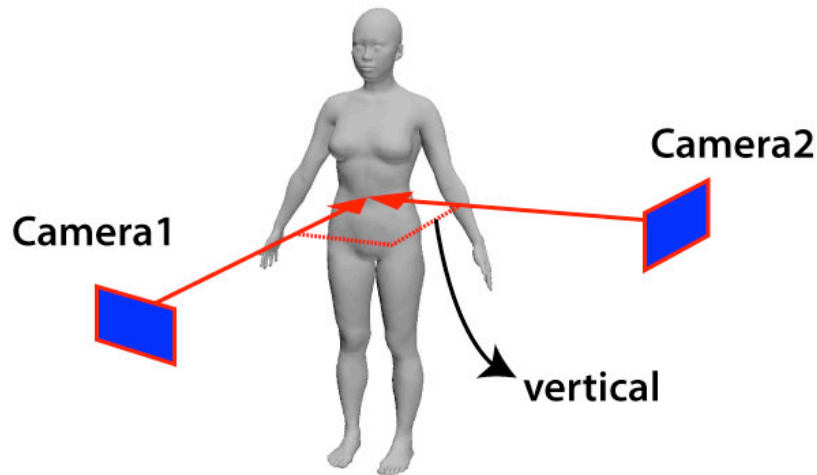


Fig. 2. Camera and subject configuration

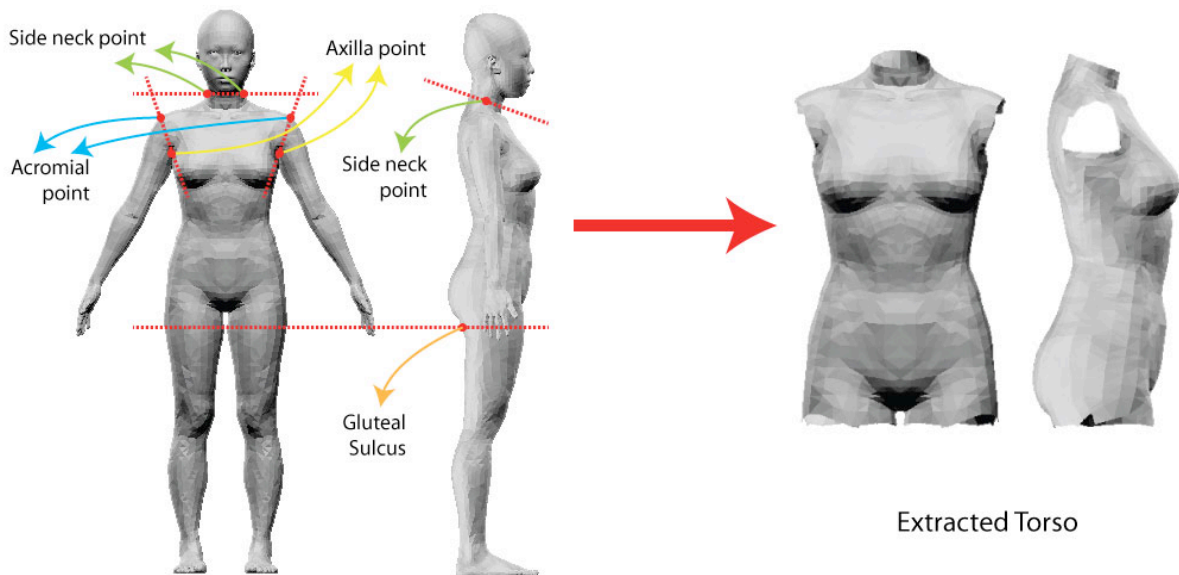


Fig. 3. Body trunk definition

### 3.2 Deformable Model

The deformable model is made synthetically from homologous human body trunk models. We preliminarily analyze homologous models database by PCA, and resulting subspace is the search space for the optimized shape deformation parameters. The subspace is consisted of eigenvectors of which cumulative contribution ratio is 95%.

#### 3.2.1. Homologous Model

Homologous models consist of the same number of the same topology, and each data point is defined based on the anatomical homology. The resulting eigenvectors of performed homologous models database represents high variance characters of body shape in the database.

In this paper, we use body trunk homologous models to make the deformable model. The body trunk model is extracted from full-body homologous model. Which part of body is defined as body trunk is same as figure 3. In the case of female, full-body model has 16490 vertexes and 87 joints. Then, the extracted body trunk model has 2437 vertexes and 4 joints.

To change posture of body trunk, an implanted joint has to be defined with 4 information: (1) position of the joint; (2-4) local xyz orthonormal bases of the local coordinate system of which the joint position is origin. This each joint information is implanted as a set of 3 points. Namely, the joint  $j$  is represented as a set of  $j_a, j_b, j_c$ . The configuration of these points is shown in figure 4. The joint position and 3 orthonormal bases (xyz) can be calculated from these 3 points. Functional joint center, x and y

orthonormal bases are defined as shown in figure 4. The z orthonormal basis is defined as a result of cross product  $\mathbf{dx} \times \mathbf{dy}$ .

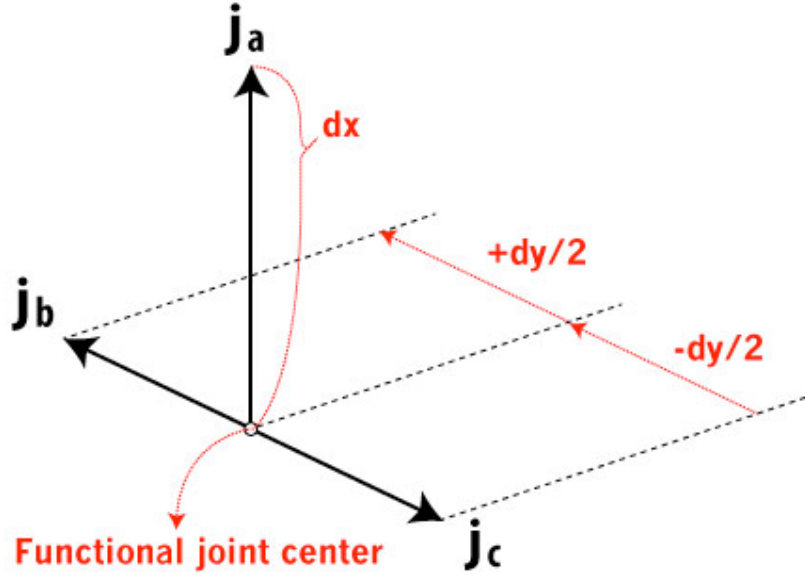


Fig. 4. Configuration of 3 points which mean joint information

Then, a homologous model consisted of  $N$  vertexes and  $M$  joints is represented as a vector  $\mathbf{S}$ . A homologous models database consisted of  $I$  homologous models is represented as a matrix  $\mathbf{D}$  and this matrix is treated as a data matrix for PCA.

$$\mathbf{S} = \left( \mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_n \quad \dots \quad \mathbf{v}_N \quad \mathbf{j}_1 \quad \mathbf{j}_2 \quad \dots \quad \mathbf{j}_m \quad \dots \quad \mathbf{j}_M \right) \quad (1)$$

$$\text{where } \mathbf{v}_n = \begin{pmatrix} v_{nx} & v_{ny} & v_{nz} \end{pmatrix}, \quad \mathbf{j}_m = \begin{pmatrix} j_{ma} & j_{mb} & j_{mc} \end{pmatrix} \quad (n=1,2,\dots,N; j=1,2,\dots,M).$$

$$\mathbf{D} = \left( \mathbf{S}_1 \quad \mathbf{S}_2 \quad \dots \quad \mathbf{S}_I \right)^t \quad (2)$$

where  $\mathbf{S}_i$  is the  $i$ th model vector in this database.

### 3.2.2. Shape Deformation

We perform PCA to the homologous models database  $\mathbf{D}$ , and obtain eigenvectors  $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_K$ . Here  $K$  is  $3(N+M)$  or  $I$ , whichever is greater and  $\mathbf{p}_k \in \mathbf{R}^3 (1 \leq k \leq K)$ .  $3(N+M)$  is the element count of a model vector  $\mathbf{S}$  in eq. (1). Then, a body trunk shape  $\mathbf{x}$  consisted of  $N+M$  points is represented as:

$$\mathbf{x} = \sum_{k=1}^K s_k \mathbf{p}_k \quad (3)$$

where  $s_k$  is the  $k$ th principal component score.

Then, we use  $s_k$  as a deformation parameter. How many parameters we use is determined by the accumulative contribution ratio of PCA. If the accumulation of 1st to  $k'$ th contribution ratio is over 95%, we use  $k'$  scores from 1st principal component score as the deformation parameters. Figure 5 shows the deformation of body trunk model by changing a deformation parameter.

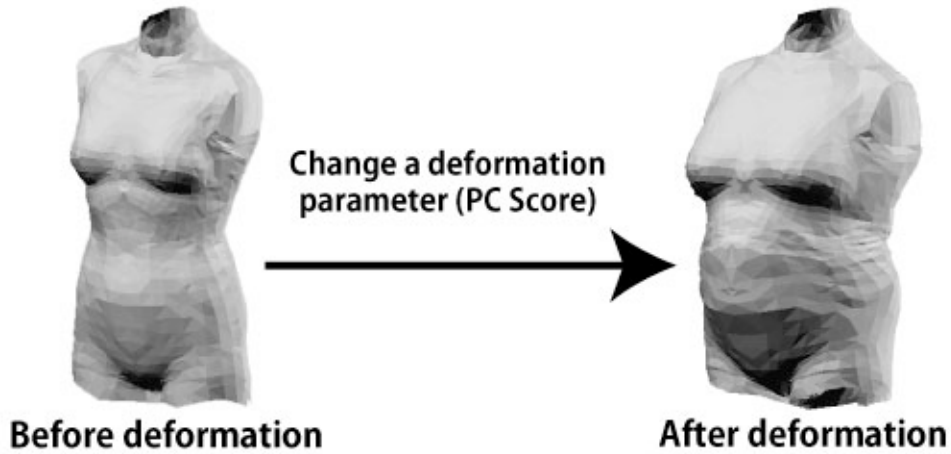


Fig. 5. A sample of shape deformation

### 3.2.3. Postural Change

Each model in the homologous body trunk models database has 4 joints inner skin meshes. These joints are automatically replaced when the shape deforms. Because PCA is performed to the joint implanted models and the result of shape deformation by eq. (3) has joint information. In the shape-deformed model, the implanted joints are positioned at the functional joint centers which are appropriate to the body after deformation. We defined the joint link structure as shown in figure 6.

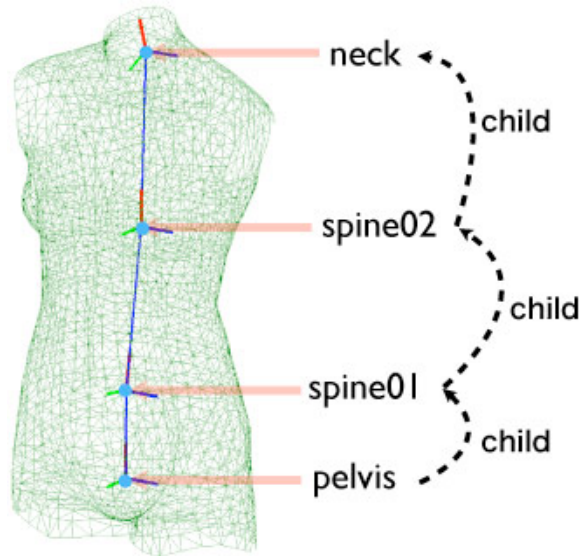


Fig. 6. The joint link structure

In the joint link structure, the parent and child connection is defined as shown in figure 6. Each joint has 3 DOF. The body trunk model is deformed by the following method based on each rotation angle of each joint.

Here  $\Sigma_w$  is the world coordinate system and  $\Sigma_j$  is the local coordinate system fixed on the joint  $j$ .

Then,  ${}^wT_j$  is the homogeneous transformation matrix from  $\Sigma_w$  to  $\Sigma_j$ . The parent joint of  $j$  is  $j+1$  and the child joint of  $j$  is  $j-1$ . These are represented as follows:

$${}^wT_j = {}^wT_1 T_2 \dots T_j \quad (4)$$

$${}^{j-1}T_j = \text{Trans}(\mathbf{t}_0^j) \text{Rot}(\mathbf{r}_0^j) \text{Rot}(\mathbf{r}^j) \quad (5)$$

when  $\text{Trans}(t_x, t_y, t_z) \equiv \text{Trans}(\mathbf{t})$ ,  $\text{Rot}(z, r_z) \text{Rot}(y, r_y) \text{Rot}(x, r_x) \equiv \text{Rot}(\mathbf{r})$ .

Here  $\mathbf{t}_0^j, \mathbf{r}_0^j$  mean the parallel translation and rotation about each axis of xyz from  $\Sigma_{j-1}$  to  $\Sigma_j$  in the initial state. In the initial state, all joints are not rotated. Then,  $\mathbf{r}^j$  means the 3 dimensional vector. The elements of this vector represent rotation angles about each axis of xyz ( $r_x^j, r_y^j, r_z^j$ ) at the joint  $j$  after postural changes.

Here  ${}^j\mathbf{v}$  is the positional vector of the vertex  $v$  in  $\Sigma_j$ . Then, after postural changes by joint rotating, the position of  $v$  in  $\Sigma_w$  is represented as  ${}^w\mathbf{v}'$ :

$${}^w\mathbf{v}' = {}^wT_{j-1} \text{Trans}(\mathbf{t}_0^j) \text{Rot}(\mathbf{r}_0^j) \text{Rot}(\mathbf{r}^j) \prod_{k \in \text{child}(j)} \left\{ {}^jT_k \text{Rot}(\mathbf{r}^k) ({}^jT_k)^{-1} \right\} {}^j\mathbf{v} \equiv \mathbf{P}(j) {}^j\mathbf{v} \quad (6)$$

$$\text{when } \mathbf{r}^j = \begin{pmatrix} w_v^{jx} r_x^j & w_v^{jy} r_y^j & w_v^{jz} r_z^j \end{pmatrix}.$$

Here  $w_v^{jz} \in [0,1]$  is the vertex blending parameter, namely, the weight of rotation angle about z-axis at the joint  $j$  (same in x or y axis). In this paper, all weights of all joints in body trunk come from weights defined in Dhaiba. These are empirically defined by CG creators and anatomists. Examples of shape deformation results of postural changes are shown in figure 7.

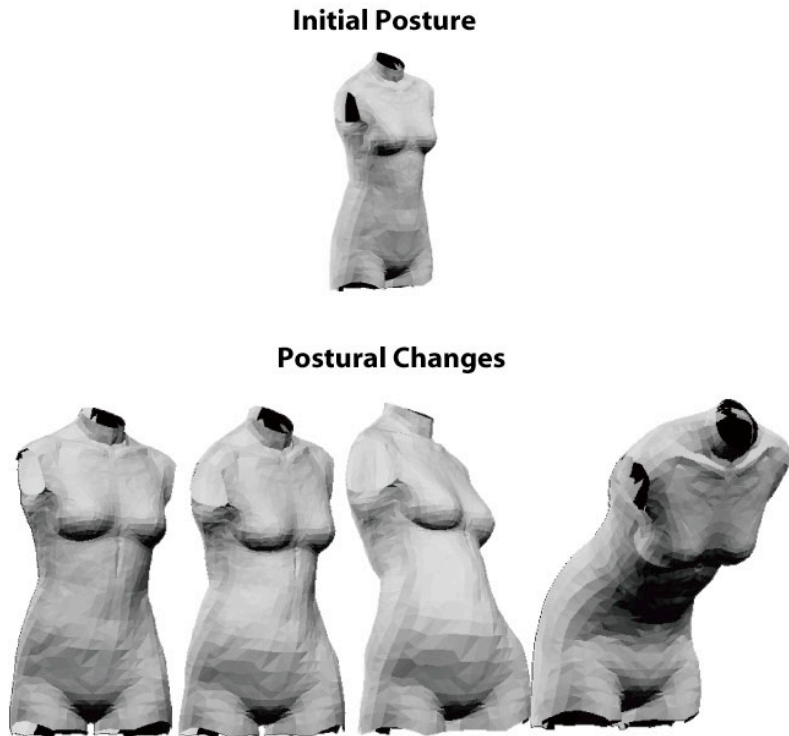


Fig. 7. Postural changes of a same body trunk model

### 3.3 Silhouette Comparison

The deformable model is optimized by referring the comparison result of silhouettes. To define the difference of silhouettes of the model and input, our system transform the model's silhouette to the added image of 2 distance-transformed images. The model's silhouette is obtained by orthogonal projection to front and side direction. Then, the example of the added image in the case of the model's side silhouette is shown in figure 8.

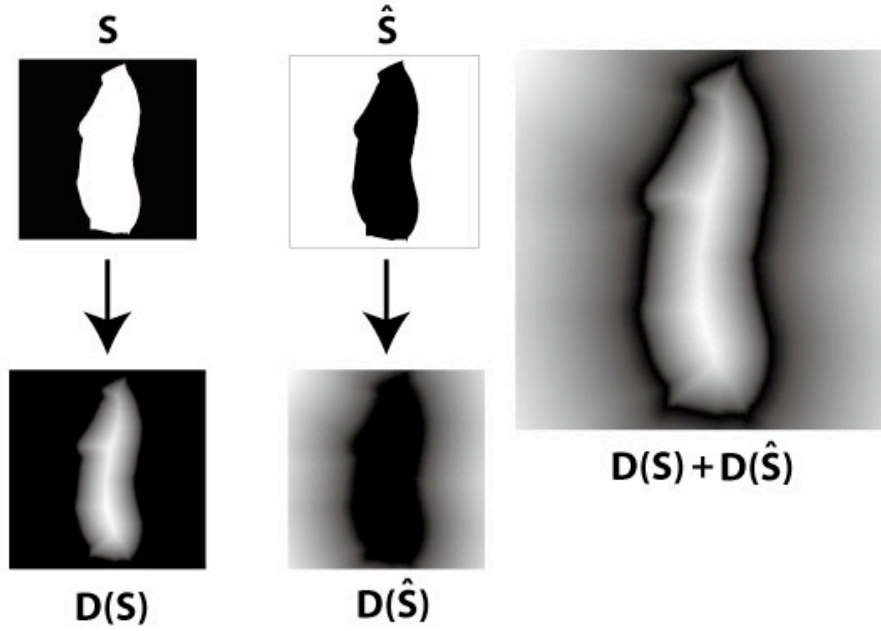


Fig. 8. The added image generating process in the case of the model's side silhouette

Here  $S, \hat{S}$  are the model's side silhouette and its reverse image.  $D(S), D(\hat{S})$  are these distance-transformed images and  $D(S) + D(\hat{S}) \equiv d$  is the added image of these. Then,  $L$  is the input silhouette which has only white or black pixels, and the difference between input silhouette and the model's silhouette  $E$ :

$$E = \iint_A d(u,v) du dv \quad (7)$$

where  $A = \{(u,v) | L(u,v) = \text{white}\}$ .

### 3.4 Model Fitting

The shape deformation is performed by  $k'$  parameters (see 3.2.2.) and these parameters are independent of each other. Then, the postural change is performed by 12 parameters because there are 4 joints and each joint has 3 DOF. Therefore,  $k' + 12$  functions are defined. Each function is represented as eq. (7), and has one of  $k'$  or 12 parameters as a variable. Our system optimizes these  $k' + 12$  parameters to minimize  $E$  by using Levenberg-Marquardt method.

## 4. Estimation Result

We show 2 estimation results in figure 9. These results were estimated from only silhouettes. At first, we made front and side silhouettes from a true value model by orthogonal projection. Then, we synthesize the deformable model from the database but the database didn't include the true value model. Therefore, the estimation was performed for obtaining the unknown 3D shape of body trunk.

## 5. Conclusion

We proposed a method to estimate body trunk shape from 2 silhouette images. Seeing the results, the estimation of 3D shape became successful approximately. In this paper, the quantitative error assessment was not achieved. Therefore, we will perform the error assessment and the examination of processing time. These are key parts of future works.

Then, the input images of this system are front and side silhouettes and these are assumed as orthogonal images of the target shape. However, the actual inputs must be perspective projection images. Therefore, when the silhouettes are compared, the model has to be projected by perspective projection using intrinsic camera parameters of the device to capture input images. We have to make sure of whether the above silhouette comparison method is truly appropriate in the case of perspective projection. This is a future work, too.

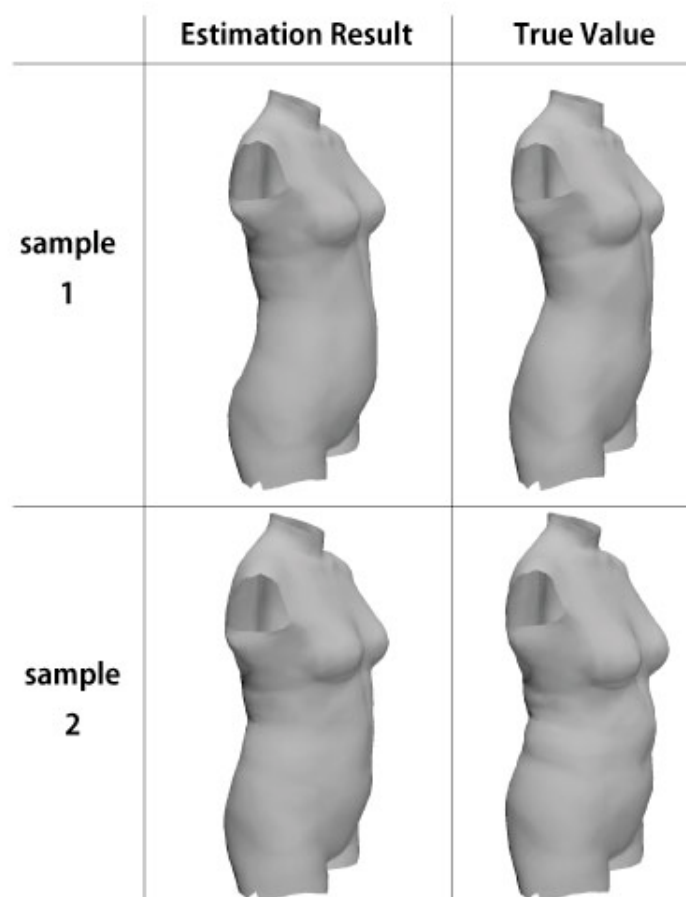


Fig. 9. Estimation results

## References

1. Mochimaru, M., Kouchi, M., Miyata, N., Tada, M., Yoshida, Y., Aoki, K., Kawachi, K., and Motomura, Y., "Dhaiba: Functional Human Models to Represent Variation of Shape, Motion and Subjective Assessment," Proceedings of SAE 2006 Digital Human Modeling for Design and Engineering Conference, 2006-01-2345, 2006.
2. M. A. Brunsman, H. A. Daanen and K. M. Robinette, "Optimal postures and positioning for human body scanning", Proceedings of International Conference on Recent Advances in 3-D Digital Imaging and Modeling (3DIM '97), pp. 266-273 (1997).
3. L. Zhang, N. Snavely, B. Curless and S. M. Seitz, "Spacetime faces: High Resolution Capture for Modeling and Animation", ACM Transaction on Graphics, 23, 3, pp. 548-558 (2004).
4. J. P. Siebert, S. J. Marshall: "Human body 3D imaging by speckle texture projection photogrammetry", Sensor Review, 20, 3, pp. 218-226 (2000).
5. Suzuki, S., Saito, H., Mochimaru, M., "Reconstruction of Facial Shape from Freehand Multi-viewpoint Snapshots", The journal of the Institute of Image Information and Television Engineers 63(4), 506-515, 2009-04-01.
6. T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, "Active Shape Models—Their Training and Application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38-59, Jan. 1995.