Virtual Fitting by Single-shot Body Shape Estimation

Masahiro Sekine*[1]   Kaoru Sugita[1]   Frank Perbet[2]
Björn Stenger[2]   Masashi Nishiyama[1]

[1] Corporate Research & Development Center, Toshiba Corporation, Japan
[2] Toshiba Research Europe Ltd., Cambridge Research Laboratory, UK

Abstract

We propose a novel virtual fitting system for seamlessly adjusting 2D clothing images to users by inferring their 3D body shape models from single-shot depth images. To increase fitting accuracy between virtual clothes and the body of the user, the system transforms and overlays the clothing image onto the body image in real time. Observations indicate that the system attains high fitting accuracy when overlaying clothing images captured from a person with a body shape similar to that of the user. We therefore develop a method for rapidly acquiring body shape models and selecting suitable clothing images based on shape similarity. We show a number of fitting results, and evaluate the fitting accuracy of both our method and an existing method without the consideration of the body shape.

Keywords: virtual fitting, body shape estimation

1. Introduction

Virtual fitting systems using a large-sized display bring about novel shopping services [1, 2] and are a focus of expectations in the retail market. Users can virtually try on clothes while standing in front of such a system as if they were standing in front of a full-length mirror. In order to realize the experience, a system capable of a quick response overlays clothing images to life-sized body images of users.
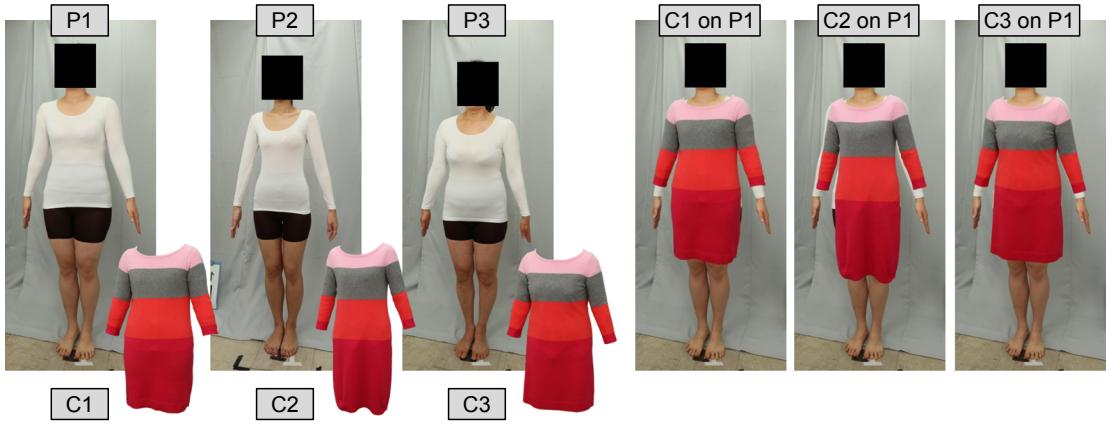
For users to feel as if they were wearing real clothes, two issues need to be considered: how to realize high-quality modeling of clothes and how to accurately overlay clothing images onto various body images. For selling clothes in the retail market, realization of high-quality clothing images is the top priority. We use photographed 2D images to model clothes because 2D images obtain better quality than 3D computer graphics (CG) in a real-time process. Furthermore, we need to consider how to treat the relationship between body and clothes. The diversity of body shapes greatly influences the appearance of clothes. For example, Figure 1 shows the difference in the appearance of clothes depending on body shapes of users. It is evident that the appearance of clothes, e.g., the width around the waist and the length of the hem, differs depending on the difference in body shape, even though the users adopt the same pose. Thus, we focus on how to treat the relationship between body shape and appearance of clothes for the virtual fitting system.

In this paper, we propose a novel virtual fitting system for seamlessly adjusting 2D clothing image to users by inferring their 3D body shape models from single-shot depth images. The proposed system exploits a method of overlaying clothing images captured from a person with a body shape similar to that of a user. We rapidly acquire a body shape model of a user and selects suitable clothing images based on a shape similarity. We prepare a clothing database that contains multiple data for pairing the body shape model and the clothing image. The database represents the relationship of appearances of clothes according to body shape variations. We search for body shape models similar to the model estimated for the user. This allows a selection of a clothing image appropriate to a body shape of a user. Furthermore, we develop a method of transforming the selected clothing image by calculating a scale and a position to conform to the shoulder line of the user. We demonstrate that our method considering the relationship between body shape and appearance of clothes improves fitting accuracy compared with an existing method that does not consider the relationship.

* masahiro5.sekine@toshiba.co.jp

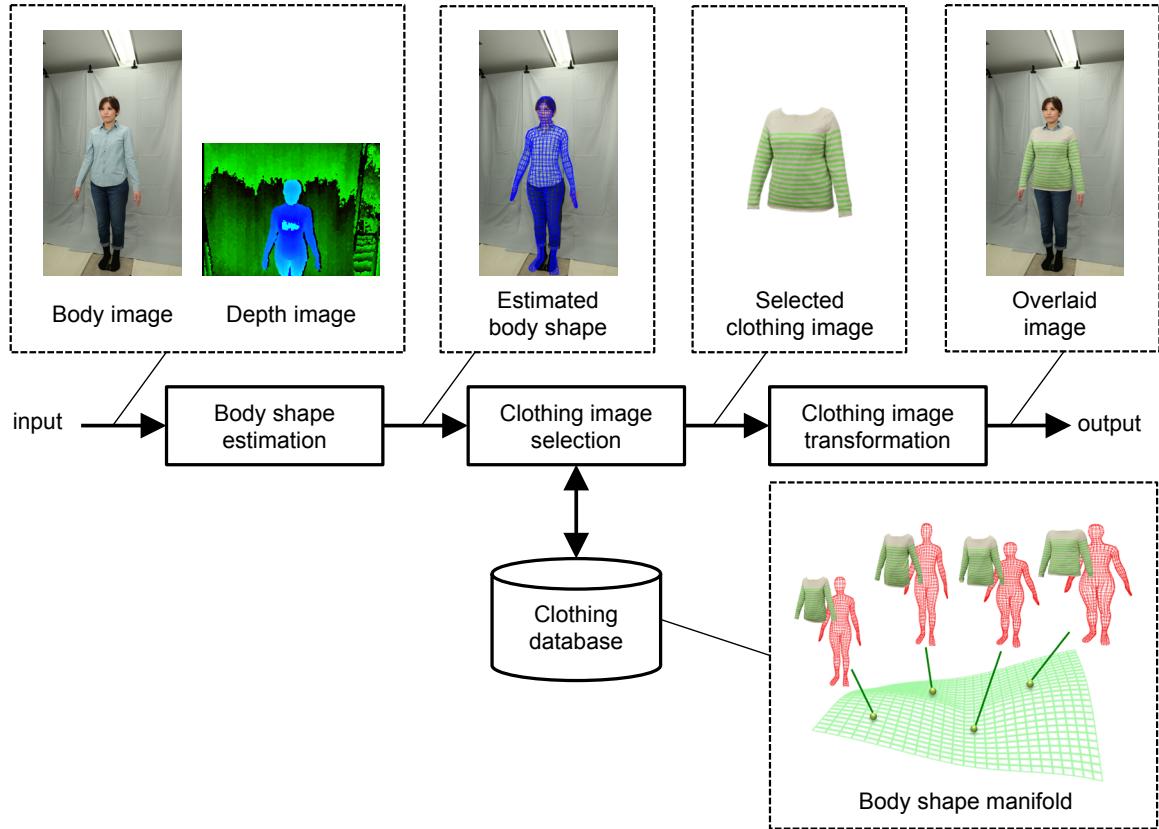*Fig. 1. Difference in appearance of clothes depending on body shape.*

Fig. 2. Overlay examples of clothing images corresponding to own / dissimilar / similar body shape.

2. Related works

We review related works on virtual fitting. At first, we review the methods to sense the person. An approach with a color or intensity image [3] and an approach with a depth image from a depth sensor [4] are known as approaches to sense the person. Our system uses an approach with a depth image from a depth sensor to prevent it from depending on the background in the virtual fitting scene. A depth sensor can acquire the joint positions on the skeleton of a user in real time. Consequently, the rough position of a user can be tracked by the joint positions in quick response [5]. Besides, the rough pose such as the twist angle of the upper body can be sensed by the joint positions. However, in the case of adjusting the scale and the position of a clothing image depending on the body shape of the user, the method only using the joint positions is not applicable. Furthermore, some methods of estimating the body shape from a depth image are proposed [6, 7]. These can get accurate estimated results, but need processing time of several minutes. In the virtual fitting scene, it is necessary to estimate the body shape rapidly.

Next, we review the methods to overlay the clothing image onto the body image. Approaches include that of using the 2D clothing image which be photographed by a camera [8, 9] and that of using the 3D CG model of clothes [10, 11]. The approach using the 2D clothing image has an advantage in it is easy to reproduce details such as the luster and the wrinkle. It realizes high-quality modeling of clothes. But on the other hand, there is a disadvantage in that it is hard to deform naturally depending on the body shape and the pose of the user. The approach using the 3D CG model has advantages in that it is easy to deform naturally depending on the body shape and pose of the user. But on the other hand, there is a disadvantage in that it is hard to reproduce details such as the luster and the wrinkle in real time. In our proposed system, we adopt the approach using the 2D clothing image in view of the ease of reproducing the appearance of real clothes for the shopping service in the retail market.

Fig. 3. Proposed process flow.

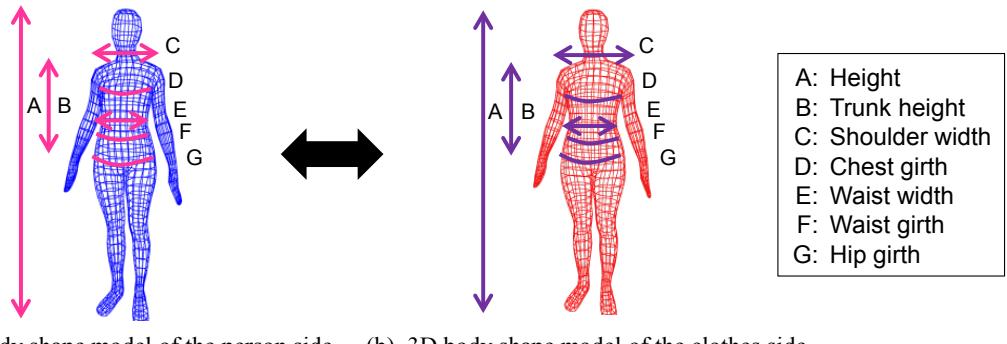## 3. Body shape sensing and fitting

3.1. Body shape manifold

   In this section, we discuss how to treat the relationship between body shape and appearance of clothes. We hypothesized that the system attains high fitting accuracy when overlaying clothing images captured from a person with a body shape similar to that of the user. First of all, we inspect our hypothesis by a prior experiment.

   We have conducted the experiment of manually overlaying clothing images corresponding to various body shapes, as Figure 2 depicts. The images of P1, P2, P3 are body of different persons and the images of C1, C2, C3 are clothes corresponding to P1, P2 and P3, respectively. P2 is the person whose body shape is dissimilar to that of P1. P3 is the person whose body shape is similar to that of P1. When we overlay the own clothing image onto the user (C1 on P1), it can naturally be fitted with high accuracy. In the case of overlay C2 on P1, the waist of the user protrudes and the hem length is too long. On the other hand, in the case of overlay C3 on P1, we can overlay the clothing image with relatively high accuracy because its body shape resembles the shape of own clothes. Based on the result of the experiment, we observed that consideration of body shape makes an important contribution to fitting accuracy.

   We therefore we develop a method of building a clothing database corresponding to the body shape manifold shown at lower right in Figure 3. There are multiple data for pairing the 3D body shape model and the clothing image in the database. We could overlay the clothing image with higher accuracy than is attainable with any existing method that uses only one clothing image.
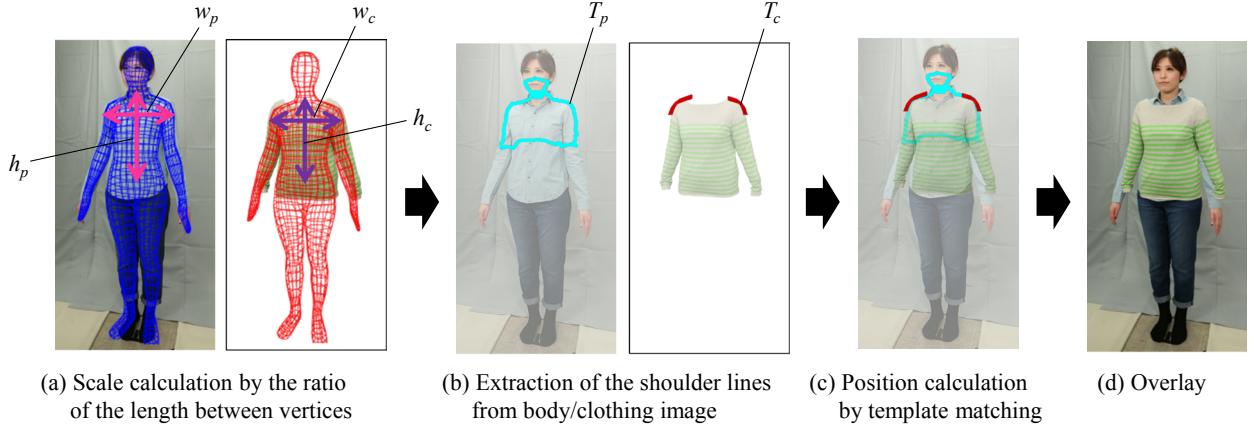
3.2. Process overview

   Figure 3 summarizes the proposed process flow. The input data in this processing are the body image and the depth image that are acquired synchronously. The first processing step is the single-shot body shape estimation (Section 3.3.1). This processing inputs a single-shot depth image and outputs the estimated 3D body shape model of the user. The second processing step is the clothing image selection (Section 3.3.2). This processing inputs the 3D body shape model and outputs a selected clothing image from the clothing database with the body

(a) 3D body shape model of the person side (b) 3D body shape model of the clothes side

A: Height
B: Trunk height
C: Shoulder width
D: Chest girth
E: Waist width
F: Waist girth
G: Hip girth

Fig. 4. Clue to search for the clothing image with similar body shape.

(a) Scale calculation by the ratio of the length between vertices (b) Extraction of the shoulder lines from body/clothing image (c) Position calculation by template matching (d) Overlay

Fig. 5. Scale and position calculation of the clothing images.

shape manifold. The last processing step is the clothing image transformation (Section 3.3.3). This processing inputs the body image and the depth image and the estimated 3D body shape model and the selected clothing image and outputs the overlaid image. The detailed algorithm of each processing step is shown below.
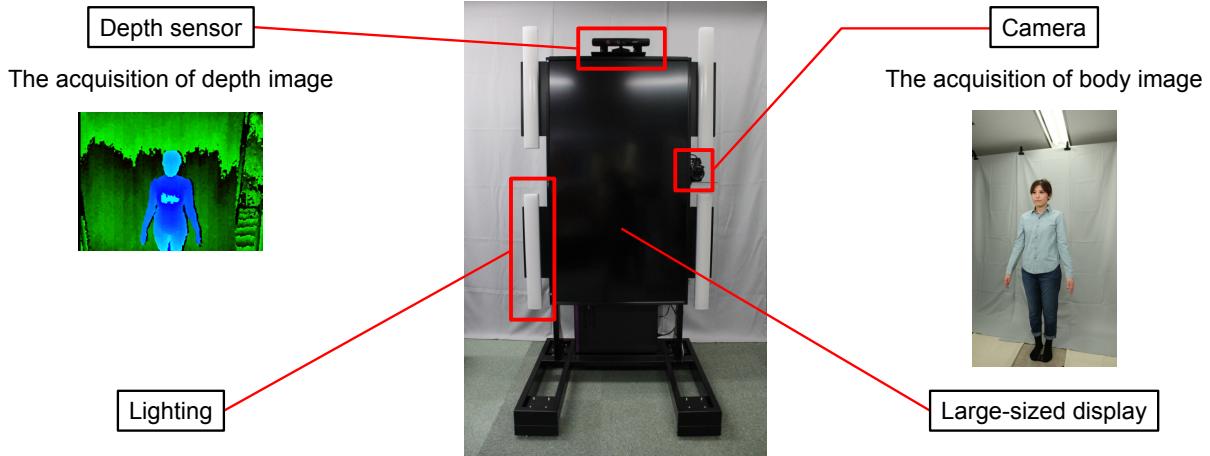
3.3. Detailed algorithm

3.3.1. Single-shot body shape estimation

First, we estimate the body shape of the user. We have developed a method for estimating 3D body shape from a single-shot depth image [12]. The method consists of a pre-training phase and a run-time estimation phase. At a pre-training phase, we generate large number of 3D body models which contain a shape factor and a pose factor of human body. Next, we generate pre-trained depth images linked to the 3D models. At a run-time estimation phase, we compare an input depth image with pre-trained depth images and determine a 3D body shape model matching the user. Finally, we apply a registration process by minimizing an energy function to optimize the selected model to the input depth image. The energy function is designed such that it is robust to clothing, leading to solutions that fit inside the input depth image, as a person fits inside their clothes. The estimation process of body shape only takes approximately one second.

3.3.2. Clothing image selection

Next, we select the clothing image that is suitable to fit. Therefore, we search for a similar body shape model from the body shape manifold included in the clothing database. We use characteristic sizes (A-G) in the body shape model shown in Figure 4 to search for a similar body shape model. The characteristic sizes A, D, F, G are quotation from the size designation of clothes determined in ISO3635-1981 [13]. The characteristic sizes B, C, E are what we select as sizes characterizing the size of 2D clothing image. The degree of similarity r between the 3D model of the person side and a 3D model of the clothes side is calculated as follows:

Fig. 6. System configuration.

$$r = \sum_i k_i |p_i - q_i| \quad , \tag{1}$$

where p_i is a characteristic size of the 3D model of the person side, q_i is a characteristic size of the 3D model of the clothes side, and k_i is a weight for each characteristic size. We calculate the degree of similarity between each model in this way, and the data that has minimal r is selected.

3.3.3. Clothing image transformation

Finally, we calculate the scale and the position to overlay the clothing image onto the body image. Figure 5 shows the method of the scale and the position calculation of the clothing images. To calculate the scale of the clothing image, we observe some vertices of the 3D body shape model. The horizontal scale s_w and the vertical scale s_h are calculated as follows:

$$s_w = \frac{w_p}{w_c} \quad , \quad s_h = \frac{h_p}{h_c} \quad , \tag{2}$$

where w_p and w_c are the shoulder width of body/clothing image, and h_p and h_c are the trunk height of body/clothing image. Furthermore, to calculate the position of the clothing image, we observe the extracted shoulder lines from body/clothing images. The shoulder line is one of the important factors in wearing clothes [14]. Based on such knowledge, we calculate the position by template matching between shoulder lines. The position **p** is calculated as follows:

$$\mathbf{p} = \arg\min_{T_p} \sum_{T_c} |l_p - l_c|^2 \quad , \tag{3}$$

where T_p and T_c are the extracted shoulder line images of body/clothing image, and l_p and l_c are the luminance of each pixel in image T_p and T_c. We overlay the clothing image by using the scale and the position calculated in this way.

4. Experiments

4.1. System overview

Figure 6 shows the system configuration used for the experiments. A depth sensor is placed at the upper part of a large-sized display to capture the depth image, and a camera is placed at the right side of a large-sized display in the direction where the image becomes vertically long to capture the body image. Because the depth sensor and the camera are far apart, the camera calibration is required to convert the coordinate system. Furthermore, the dedicated lighting is attached to a large-sized display. In these experiments, in order to keep the natural appearance of clothing images, we equate the lighting condition when we use the virtual fitting system with the lighting condition when we capture the clothing images.
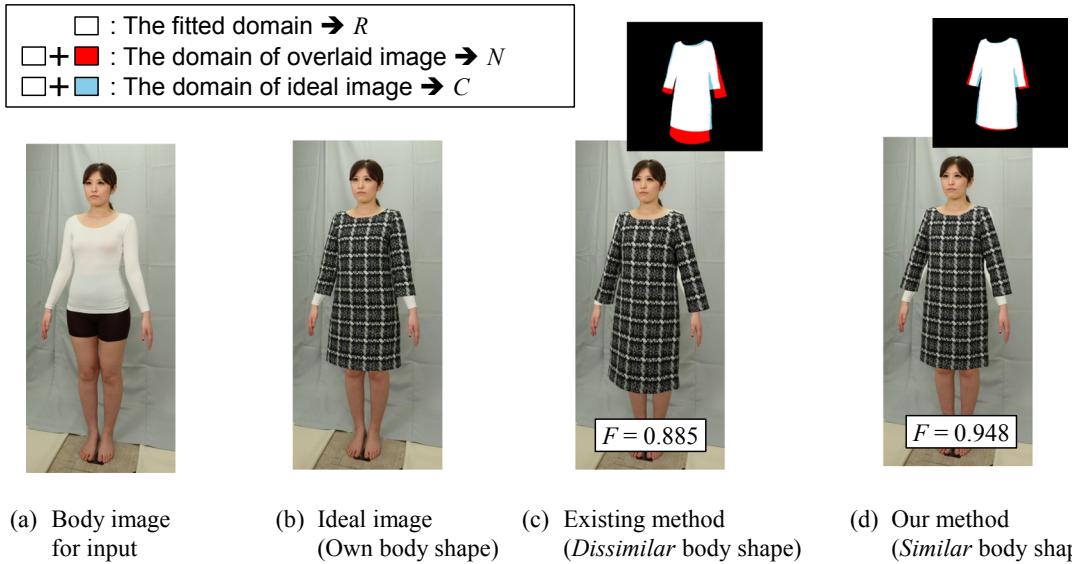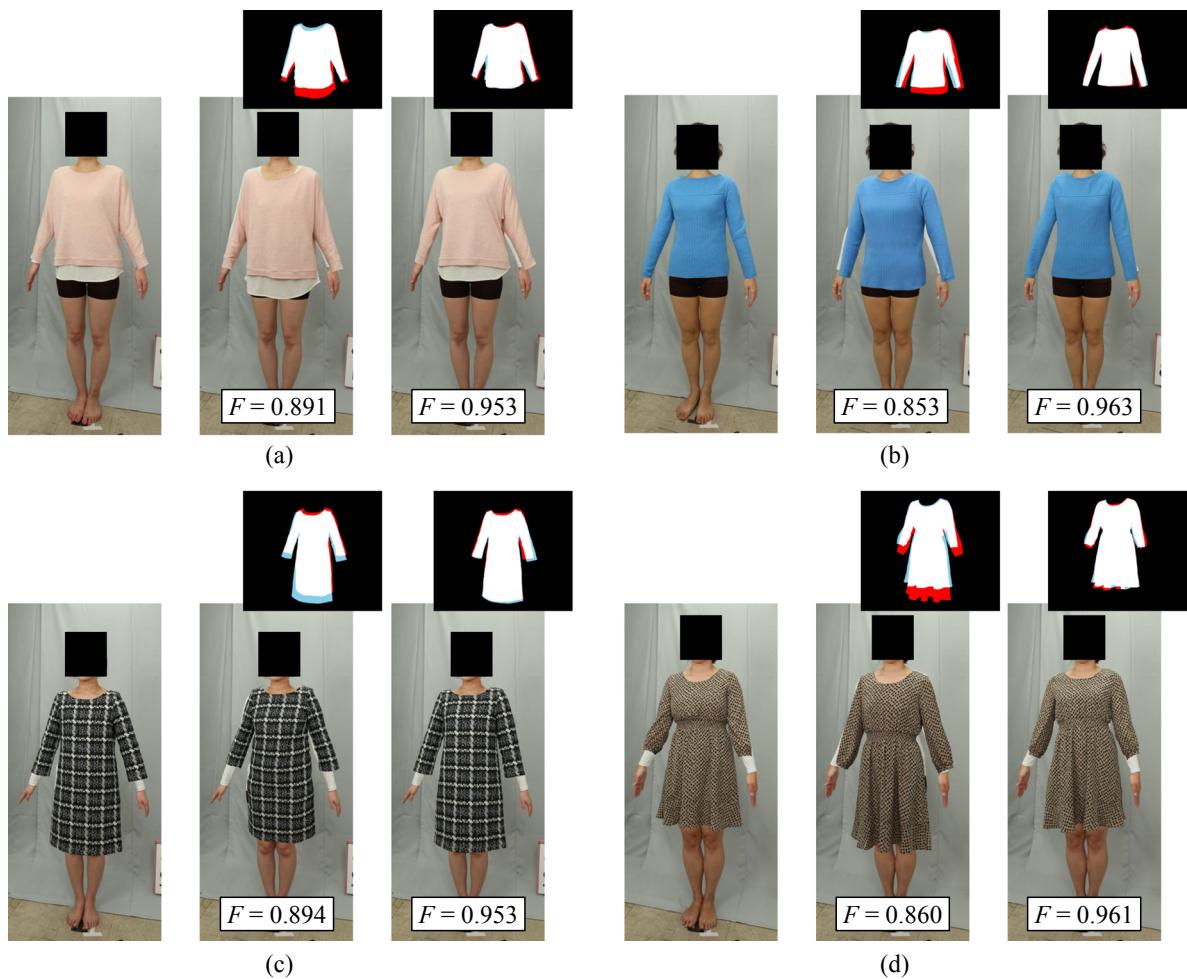
| : The fitted domain ➜ R |
| ☐+🟥 : The domain of overlaid image ➜ N |
| ☐+🟦 : The domain of ideal image ➜ C |

$F = 0.885$

$F = 0.948$

(a) Body image
for input

(b) Ideal image
(Own body shape)

(c) Existing method
(*Dissimilar* body shape)

(d) Our method
(*Similar* body shape)

Fig. 7. Comparison of fitting results.

$F = 0.891$

$F = 0.953$

$F = 0.853$

$F = 0.963$

(a)

(b)

$F = 0.894$

$F = 0.953$

$F = 0.860$

$F = 0.961$

(c)

(d)

Left: Ideal image (overlay own clothing image)
Center: Existing method (overlay the clothing image corresponding to dissimilar body shape)
Right: Our method (overlay the clothing image corresponding to similar body shape)
Upper: Each differential image

Fig. 8. Example of fitting results.

Fig. 9. Use scene of proposed virtual fitting system.

    (a) Before try-on   (b) After try-on

Fig. 10. Display example of the proposed virtual fitting system.

4.2. Evaluation of fitting accuracy

In system evaluations, we use the F-measure *F* as an index of fitting accuracy as follows:

$$F = \frac{2R}{N + C} \quad , \qquad (4)$$

where *R* is the number of pixels in the domain where the overlaid clothing image is equal to the ideal clothing image. *N* is the number of pixels in the domain of the overlaid clothing image. *C* is the number of pixels in the domain of the ideal clothing image.

We evaluated the fitting accuracy using a test set of 52 people and 10 clothing patterns. As in the case of an existing method, we overlay the clothing images except for the own clothing image. When we compared our method with an existing method, F-measure improved from 0.916 to 0.921. Notably, the lowest F-measure significantly improved from 0.769 to 0.842. Figure 7 shows the comparison of fitting results between the existing and our method. Because the clothing image corresponding to dissimilar body shape may be used in the existing method, F-measure may become small. On the other hand, F-measure becomes relatively big in our method. Figure 8 shows other examples of fitting results.

We can realize the virtual fitting system shown in Figure 9, 10 by using our method. This system can run the fitting process by our method in several seconds. As a result of user test, most of users were satisfied with our system. However, the users with the body shape not to be included in a clothing database were not satisfied. It is a problem to have to prepare the sufficient data which covers various body shapes.

5. Conclusion

In this paper, we proposed a novel virtual fitting system which adopted a method for rapidly acquiring body shape models and selecting suitable clothing images based on shape similarity. Furthermore, we presented a method of clothing image transformation focusing on the shoulder line. As a result, the fitting accuracy was

improved compared with an existing method. We collected the clothing images of more than 50 people per kind of clothes. But we have problems in that the photography cost is big and the content of the clothing database is big. We will consider how many clothing images with body shape manifold are necessary. Furthermore, in the future, we will consider the processing method for a person adopting a different pose and a clothing image with a different lighting condition.

References

[1] Virtual Fashion 3D, http://www.next-system.com/virtualfashion.html.

[2] 3D virtual fitting room, http://www.vismile.com.tw/en/main.html.

[3] R. Okada and B. Stenger, "A single camera motion capture system for human-computer interaction," *IEICE Tans. Inf. & Syst.*, vol.E91-D, no.7, pp. 1855-1862, 2008.

[4] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.1297-1304, 2011.

[5] Z. Zhou, B. Shu, S. Zhuo, X. Deng, P. Tan, and S. Lin, "Image-based clothes animation for virtual fitting," *Proceedings of ACM SIGGRAPH ASIA*, 2012.

[6] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis, "SCAPE: shape completion and animation of people," *Proceedings of ACM SIGGRAPH*, vol.24, no.3, pp.408-416, 2005.

[7] A. Weiss, D. Hirshberg, and M. Black, "Home 3D body scans from noisy image and range data," *International Conference on Computer Vision*, pp.1951-1958, 2011.

[8] J. Yoon, I. Lee, and H. Kang, "Image-based dress-up system," *Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication*, 2011.

[9] A. Hilsmann, P. Fechteler, and P. Eisert, "Pose space image based rendering," *Computer Graphics Forum*, vol.32, no.2, pp.265-274, 2013.

[10] P. Guan, L. Reiss, D. Hirshberg, A. Weiss, and M. Black, "DRAPE: Dressing any person," *ACM Transaction on Graphics*, vol.31, no.4, 2012.

[11] J. Li, J. Ye, Y. Wang, L. Bai, and G. Lu, "Fitting 3d garment models onto individual human models," *Computer and Graphics*, vol.34, no.6, pp.742-755, 2010.

[12] F. Perbet, S. Johnson, M. Pham, and B. Stenger, "Human body shape estimation using a multi-resolution manifold forest," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.668-675, 2014.

[13] ISO 3635-1981, "Size designation of clothes – Definitions and body measurement procedure," 1981.

[14] A. Fukai, "Fashion dictionary (in Japanese)," *Bunka publishing bureau*, 1999.