A virtual try-on system in augmented reality using RGB-D cameras for footwear personalization

Article *in* Journal of Manufacturing Systems · June 2014

DOI: 10.1016/j.jmsy.2014.05.006

3 authors, including:

Chih-Hsing Chu
National Tsing Hua University
209 PUBLICATIONS **2,077** CITATIONS

SEE PROFILE

Journal of Manufacturing Systems xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

journal homepage: www.elsevier.com/locate/jmansys

A virtual try-on system in augmented reality using RGB-D cameras for footwear personalization

Yu-I Yang, Chih-Kai Yang, Chih-Hsing Chu*

Department of Industrial Engineering and Engineering Management, National Tsing-Hua University, Hsinchu, Taiwan

ARTICLE INFO

Article history:
Received 4 December 2013
Received in revised form 2 May 2014
Accepted 14 May 2014
Available online xxx

Keywords:
Mixed reality
Object tracking
Virtual try-on
Design evaluation

ABSTRACT

This paper presents a system for design evaluation of footwear using commercial depth-sensing technologies. In a mixed reality environment, the system allows users to virtually try on 3D shoe models in a live video stream. A two-stage object tracking algorithm was developed to correctly align shoe models to moving feet during the try-on process. Color markers on the user's foot enabled markerless tracking. Tracking was driven by an iterative closest point (ICP) algorithm that superimposed the captured depth data and predefined reference foot models. Test data showed that the two-stage approach resulted in increased positional accuracy compared with tracking using only surface registration. Trimming the reference model using the instant view angle increased the computational efficiency of the ICP algorithm. The proposed virtual try-on function is an effective tool for realizing human-centered design. This study also demonstrated a new application of RGB-D cameras to product design.

1. Introduction

Modern consumers seek personalized products and services in a consumer environment characterized by mass production [1]. In addition to functional requirements, designers must consider a product's emotional appeal, induced by product styling, as well as other affective attributes. The esthetic appeal of a product plays a crucial role in its success. Consumers tend to look for design elements that reflect their own tastes and allow them to differentiate themselves from other people. Evaluating whether and how much a design fits its users are critical in product customization and personalization [2]. This is particularly obvious in the apparel and fashion industries, in which designers must realize three essential dimensions of customization: fit, functionality, and esthetic design [3]. Recent progress in information and communications technology (ICT) has provided tools for realizing this challenging task.

Computer-aided design (CAD) has accelerated the product design process by automating the construction of product models, CAD also supports downstream manufacturing tasks, such as process planning and NC tool path generation. Most existing CAD tools were constructed by designers from an engineering perspective, and are not optimized for product users. Augmented reality (AR) and mixed reality (MR), in which virtual models generated based on computer graphics are superimposed over real objects and scenes, are considered more usable interaction technologies for product design [4,5]. This is particularly useful in evaluating the design of fashion products, such as apparel, footwear, and wearable items. In one study, most users found virtual try-on systems driven by personalized avatars to be useful, but expressed dissatisfaction that the created avatars were not sufficiently realistic or accurate [6]. Facial models used in a virtual hairstyle design program exhibited the same problem [7]. Moreover, markers and special patterns have been used in most applications to position virtual objects within real scenes [8]. The presence of large markers inevitably reduces the usability of those applications, because they are occlusions that lower visualization quality. Reducing the use of markers, or reducing the size of the markers, is advantageous in evaluating fashion product designs.

2. Related works

The ability to personally design products and instantly interact with the resulting designs is highly desirable for customers. The idea of design automation has been realized for free-form products using CAD techniques [9]. Apparels can be automatically re-constructed to accommodate the differences in individual body shape and size [10]. Implementing design personalization in an AR environment is also a promising approach to achieving this goal. Recent progress in depth-sensing technologies has enabled new applications in various industries. RGB-D cameras,

* Corresponding author. Tel.: +886 3 5742698.
E-mail address: chchu@ie.nthu.edu.tw (C.-H. Chu).

http://dx.doi.org/10.1016/j.jmsy.2014.05.006
0278-6125/

such as the Microsoft Kinect and the ASUS Xtion, have been used in robot planning [11], rehabilitation [12], and museum guidance [13]. Virtual try-on technology using RGB-D cameras has received attention because it enables users to see themselves wearing different clothes without physically changing clothes [14]. The Kinect's human pose estimation performance is adequate for real-time applications [15,16]. Most applications developed to virtually try on garments have used video streams to demonstrate the garments design [14,17,18]. By contrast, few studies have examined the virtual prototyping of footwear. Antonio et al. [18] developed a high-quality stereoscopic vision system that allows users to try on shoes from a large 3D database while looking at a "magic mirror", and reported that footwear customization using AR improved product quality and increased consumer satisfaction. Eisert et al. [19] used AR techniques to create a virtual mirror to visualize customized sports shoes in real time, a large display screen, in place of a mirror, showed the input of a camera capturing legs and shoes. A 3D motion tracker was developed to robustly estimate the 3D positions of both shoes based on silhouette information from a single camera view.

The previous studies [18,19] have demonstrated that the idea of virtual try-on is highly valuable in realizing personalized design. However, those applications adopted specialized equipment that is not only pricy, but also inaccessible to most users. To overcome this problem, this paper presents a prototype system for virtually trying on shoes that enables users to evaluate shoe styling and appearance in an economically viable way, using automatic object tracking based on data captured using Kinect. The system allows users to try on 3D shoes in a live video stream. Several tracking mechanisms were tested and compared regarding accuracy and efficiency. Small color markers placed on the instep yielded the initial foot position based on image segmentation techniques. Markerless tracking driven by an iterative closest point (ICP) algorithm was implemented to correctly position the model with respect to the moving foot during the try-on process. Quantitative analysis of the test results revealed that a two-stage tracking method exhibited superior performance in both tracking efficiency and visualization quality, and resulted in higher positional accuracy than tracking using only markers or the ICP algorithm, particularly when foot movements were fast. Various use scenarios demonstrate how the proposed system facilitates footwear design customization. This study provides a feasible approach to allow consumers to evaluate products interactive with humans, and as well as to engage consumers in the design process. This study also demonstrated a new application of RGB-D cameras to product design.

3. System framework

3.1. Object tracking for virtually trying on shoes

A design personalization system must allow consumers to quickly evaluate designs. The proposed virtual try-on system enables users to virtually wear shoe models and interactively evaluate designs in an MR environment. This function relies on automatic foot tracking in a video stream. Object tacking methods can be classified according to their use of physical markers. Most MR systems identify locations using markers in the real environment, which often contain special patterns that allow the camera position and orientation to be quickly calibrated in 3D space. However, one major limitation is that markers normally lie on planar objects, and do not offer satisfactory positioning accuracy or tracking robustness when lying on curved objects. Tracking using physical markers is thus poorly suited to virtually trying on shoes, because of the curvature of the human body. The integration of markerless tracking or other sensing technologies is thus necessary.

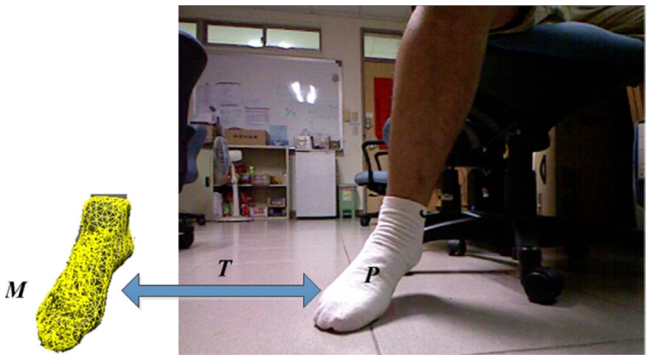

Fig. 1. Problem definition of virtually trying on shoes.

Markerless tracking requires automatic detection of objects based on color images. This topic has recently received much attention in computer vision and image processing; precisely identifying the location of a foot in a live video stream without use of markers is a challenging task. The main difficulty is that a human foot does not provide sufficient feature information that can be used by automated systems to identify it in images. Feet cannot always be distinguished from background objects or limbs, unless a person is wearing socks with predefined color or graphic patterns; thus, additional information is required to track feet in motion. Recent advances in depth-sensing technologies have provided effective solutions for tracking objects based on range data. This study implemented both marker and markerless tracking, and fully used color and depth images captured using an RGB-D camera; this hybrid approach provides a tradeoff between positional accuracy and computational efficiency in automatic object tracking.

3.2. Problem definition

This study involved tracking human foot motion based on color and depth images captured using Kinect. Precisely positioning a shoe model onto a human foot in real time is difficult, particularly because of the limited view angle of the depth camera, which allows only partial data to be captured for moving feet. Another difficulty arises from the need to determine whether a particular shoe comfortably fits a particular foot [20], allowances between the geometric shapes of feet and shoes are necessary to account for the free movement of the foot within the shoe.

Using data captured from foot motion to position foot models prevents these problems. A template foot model was adopted as a tracking reference because aligning similar shapes is more feasible than aligning two distinct shapes. The relative positions between the reference model and the shoe model to be displayed were defined prior to tracking. As shown in Fig. 1, the foot tracking task is described as follows:

$$\text{Min} \sum_{i=1}^{n} ||\boldsymbol{m}^* - \boldsymbol{p}_i \cdot \boldsymbol{T}|| \quad (1)$$

where the depth data **P** instantly captured using Kinect contains n points and **M** is the point cloud comprising the reference model. A best match Ω between **P** and **M** can be obtained using various techniques. The term $\boldsymbol{p}_i$ is a point in **P** and corresponds to $\boldsymbol{m}^*$ from **M** in Ω, $\boldsymbol{T}$ describes a 3D coordinate transformation matrix determined according to Ω, and is decomposed into a rotation matrix $\boldsymbol{R}$ and a translation matrix $\boldsymbol{t}$. It was necessary to identify **P** in a live image.

The foot tracking process flow is shown in Fig. 2. A two-stage tracking method was proposed to continuously identify the location of a foot based on the Kinect-captured video stream. The video

Please cite this article in press as: Yang Y-I, et al. A virtual try-on system in augmented reality using RGB-D cameras for footwear personalization. J Manuf Syst (2014), http://dx.doi.org/10.1016/j.jmsy.2014.05.006

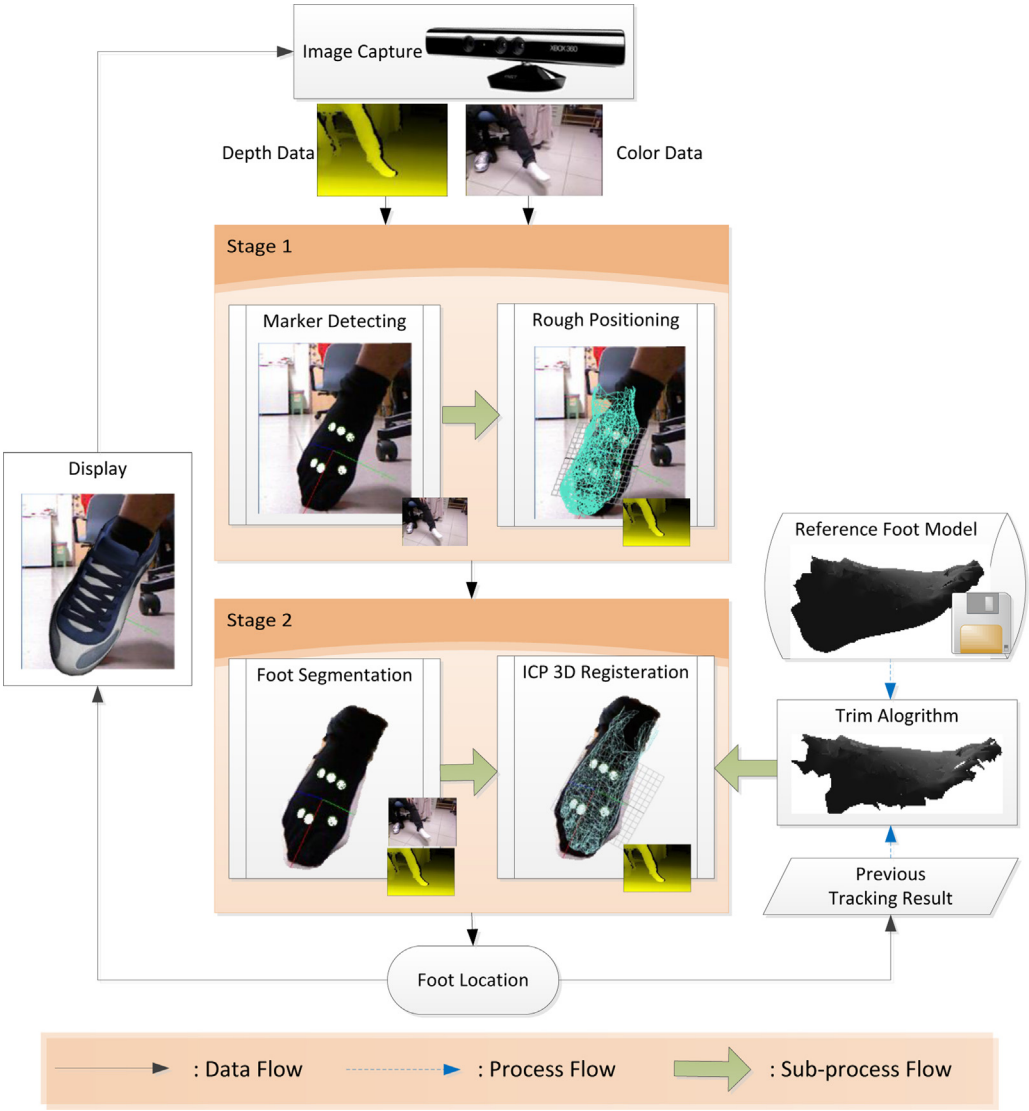

Fig. 2. Process flow of two-stage foot tracking.

stream consisted of live continuous color and depth images. The purpose of the first tracking stage was to quickly estimate the location of a foot using markers placed on the foot; these markers enabled only an approximation of the initial foot position. The pixel coordinates of the marker were identified in the color image, and the depth of each pixel was obtained from the depth image. In the second tracking stage, the precise foot location was computed using two techniques: image segmentation, which required both color and depth data, and ICP-based shape registration, which required only depth data. ICP-based tracking exhibited more satisfactory performance in the reference model after trimming with the depth view angle.

4. Methods

4.1. Marker detection

Color markers allowed quick determination of the initial foot position in 3D space. Six round markers, fluorescent green in color, were placed on the foot instep (see Fig. 3). The color image captured from the real scene was then converted into the YIQ space specified by the NTSC (National Television Standards Committee) [21]. The Y component represents the luma information, and I and Q represent the chrominance information. Recognizing markers in the traditional RGB space is problematic because of uncontrollable variation in light sources, density, and white balance; by contrast, identifying fluorescent green markers in the YIQ space is straightforward when

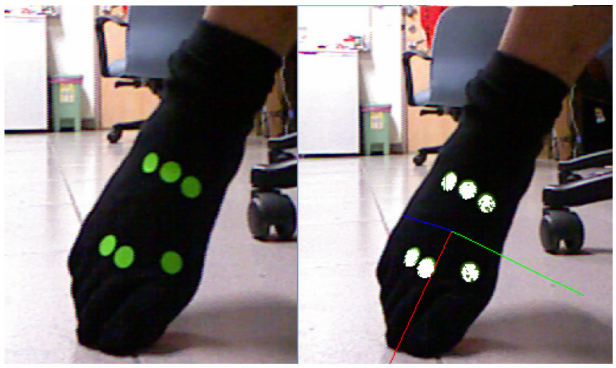

Fig. 3. Placement of color markers on the user's instep. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Please cite this article in press as: Yang Y-I, et al. A virtual try-on system in augmented reality using RGB-D cameras for footwear personalization. J Manuf Syst (2014), http://dx.doi.org/10.1016/j.jmsy.2014.05.006

using the corresponding I and Q values. The conversion between the YIQ and RGB spaces is written as:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2)$$

It was assumed that no other objects in the environment bore the same color as the fluorescent markers. The pixels of the markers were determined in the color image by examining the I and Q values.

4.2. Rough positioning

The depth sensor automatically established the correspondence relationship between the color and depth images, pixel by pixel. The correspondence relationship was used to determine the depth of each pixel identified by the color markers as being part of the foot, generating a set of 3D points $M = \{m_1, m_2 \ldots m_{n_m}\}$. The centroid of those points was denoted as $\bar{m}$. Using $\bar{m}$ as the origin, principle components analysis (PCA) [22] was performed on $M' = \{m_1 m, m_2 m \ldots m_{n_m} m\}$. The PCA result provided three orthogonal base vectors v_1, v_2, v_3 corresponding to three principal component values $\sigma_1^2, \sigma_2^2, \sigma_3^2$, with $\sigma_1^2 > \sigma_2^2 > \sigma_3^2$. The following conditions were imposed on those directions to guarantee unique solutions:

1. $v_1 \times v_2 = v_3$
2. v_1 is along the toe direction;
3. v_3 points to the depth camera.

The placement of color markers on the instep required a special geometric arrangement; the first three markers should be placed near the 1st–3rd cuneiform bones, and the remaining markers should be placed around the heads of the 1st–3rd metatarsal bones [23]. As shown in Fig. 3, such a geometric configuration produced the largest variance along the toe direction, corresponding to v_1 in the PCA result. The other two directions were determined according to the above three conditions. Arranging markers in this fashion enabled quick estimation of the foot position based on the images captured by Kinect. Although the estimation was approximate and did not include precise depth information, it was an adequate basis for a second, more precise tracking procedure.

4.3. Markerless tracking

Markerless tracking requires feature information other than physical markers to perform automatic object recognition and location estimation. The geometric information contained in the depth image provided useful features to perform markerless tracking. A reference foot model was used as a matching target to search for a region in the depth image that best matches the reference model geometrically; such recognition techniques are referred to as 3D shape or surface registration techniques. ICP [24] is an algorithm commonly employed to minimize the difference between two clouds of points, and has been successfully implemented in various real-time applications, such as geographical information systems, computer vision, and robot planning. In this study, the ICP algorithm was applied to optimally superimpose the reference model on the captured depth data. This method involves continually adjusting the position of a first point cloud through minimizing positional deviations with respect to the second point cloud.

4.4. Model trimming

The limited view angle of Kinect is problematic in ICP-based tracking. Although the obtained depth image contained only partial

Fig. 4. The reference foot model.

information on foot geometry, the reference model approximated a closed volume, except the top portion, as shown in Fig. 4. This partial geometry information provided less feature information that could be used in 3D surface registration. Moreover, the ICP algorithm was not designed to accomplish partial-to-whole matching, and most ICP-based applications perform favorably when two point clouds to be registered cover the same (or almost the same) geometry [25]. This problem was overcome by dynamically trimming the reference model by using the instant camera view angle. This view angle can be determined using the OpenNI framework [26], an open source SDK used for development of applications based on RGB-D cameras. Mesh points in the reference model that could not be seen from the present view angle were thus removed. As shown in Fig. 5, the normal direction of those points was of an angle greater than 90° with respect to the camera view angle.

4.5. Foot segmentation

Kinect-captured depth images normally contain 640 by 480 pixels, the computational time required to precisely locate a foot region in an image this size using ICP alone is too lengthy to be feasible in real-time applications. A foot segmentation procedure (see Fig. 2) was proposed to solve this problem by roughly locating the foot region within the image. This procedure consisted of two steps. First, the foot region was separated from the background by applying the seeded region growing method [27], using the pixel corresponding to $\bar{m}$ as a seeding pixel on the region to be separated. The region of interest then expanded as pixels with a depth gradient smaller than a given threshold d_d were added; pixels that did not belong to the human body had a gradient value greater than the

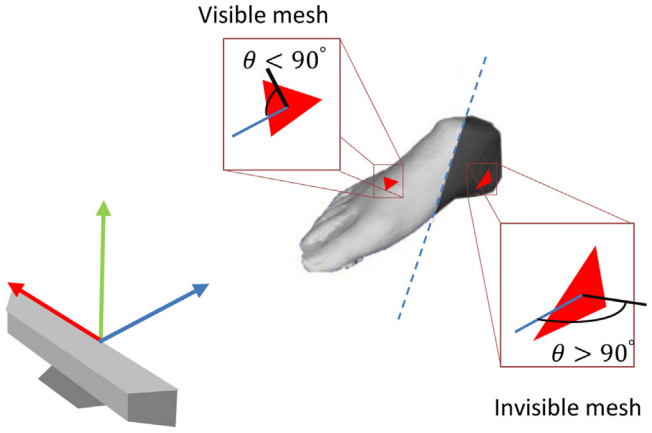

Fig. 5. Dynamic trimming of the reference foot model.

Please cite this article in press as: Yang Y-I, et al. A virtual try-on system in augmented reality using RGB-D cameras for footwear personalization. J Manuf Syst (2014), http://dx.doi.org/10.1016/j.jmsy.2014.05.006

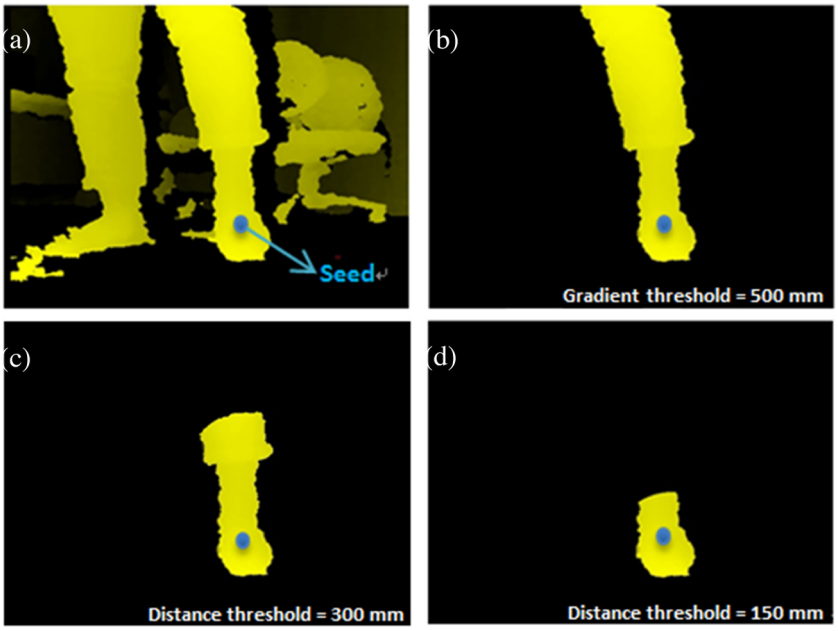

Fig. 6. Foot segmentation procedure.

threshold, and thus allowed the user's body to be extracted from the scene, as shown in Fig. 6(a) and (b).

Subsequently, the foot was located by controlling the region of interest, any pixel with a distance to the seeding pixel smaller than a given threshold d_f was included in the foot region. The size of the region was controlled with different threshold values, as shown in Fig. 6(c) and (d), thus allowing the user's foot to be segmented from the depth image.

4.6. ICP-based shape registration

The term $\mathbf{P^s}$ denotes the foot region segmented from the depth image and contains n points; $\mathbf{M^t}$ is the trimmed reference model and contains m points. The number of 3D points in these two models can differ. The term m is assumed to be greater than n. The ICP algorithm is described as follows:

Step 1: For each $\boldsymbol{p}_i$ in $\mathbf{P^s}$, the closest point $\boldsymbol{m}_i^*$ in $\mathbf{M^t}$ is found. The set of all closest points is denoted as $\mathbf{M}^*$.
Step 2: A rotation matrix $\boldsymbol{R}$ and a translation matrix $\boldsymbol{t}$ are calculated to minimize the summation of the mean square error (MSE) for each point pair generated from $\mathbf{P^s}$ and $\mathbf{M^t}$.
Step 3: The coordinate transformation $\boldsymbol{T}$ on $\mathbf{M^t}$ is performed.
Step 4: If the termination conditions are not satisfied, Step 1 is repeated; otherwise the process stops.

The objective function in the minimization process is the summation of the MSEs for all point pairs. The term $\mathbf{P^s}$ normally contains fewer points than does $\mathbf{M^t}$; the resolution (or the number of pixels) of the RGB-D camera was not particularly high, and the mesh density of the reference model before trimming could be freely adjusted. Subsequently, the rigid body transformation was estimated, represented as the transformation matrix $\mathbf{T}$ required to position $\mathbf{M^t}$ as close as possible to $\mathbf{P^s}$. Eq. (1) can be rewritten as:

$$\mathrm{Min}\sum_{i=1}^{n}||m_i^* - \{p_i \cdot R(\theta_x, \theta_y, \theta_z) + t(t_x, t_y, t_z)\}|| \quad (3)$$

where $\theta_x, \theta_y, \theta_z$ are rotation angles along the x, y, and z axes, respectively. The terms t_x, t_y, t_z indicate the translation components along the x, y, and z axes. Previous studies have proposed various methods to determine θ_x, θ_y, θ_z and t_x, t_y, t_z by using linear models. However, Eggert et al. [28] reported that solutions obtained using linear models tend to substantially deteriorate in the presence of noise. The signals sent back from Kinect inevitably contained errors; thus, a nonlinear model may perform more robustly in calculating the transformation matrix $\mathbf{T}$ than a linear one. The steepest descent algorithm [29] was applied to determine θ_x, θ_y, θ_z and t_x, t_y, t_z based on an approximation of the objective model using a quadratic model.

The ICP algorithm was terminated when (1) the number of iterations reaches a given limit l_t and (2) the MSE is lower than a given value e_t. The effectiveness of such an optimization-based approach largely depends on the quality of the initial solution. In each iteration of the algorithm, the foot location estimated according to the color marker tracking was used as a preliminary estimate.

5. Implementation results

A virtual try-on prototype system was implemented in an AR environment, ARToolKits [26]. ARToolKits is a software library used for building AR applications that involve the overlay of virtual imagery on the real world. Kinect was used as the RGB-D camera in the system, capturing a real scene and returning color and depth information in 640 by 480 pixels. The image generation rate was 30 frames per second (fps). The camera was installed 20 cm above the ground, and the try-on function was performed approximately 50–100 cm in front of the camera. All system functions were implemented using C++ on a personal computer with an Intel Pentium Dual-Core CPU T4400 at 2.20 GHz.

The tracking function in the prototype system was implemented using three different methods: color marker tracking, ICP-based tracking, and a two-stage approach using both. The technical capabilities and performance of these three methods in foot tracking from various perspectives in practice were thus compared. In the first test, a user was asked to slowly move and turn his foot in front of the depth camera. The movement lasted approximately 2 s, and was recorded at 30 fps, 60 frames in total were generated as the test data. Shoe models were positioned on the user's foot during the try-on process using the three methods. Fig. 7 depicts the overlay

Please cite this article in press as: Yang Y-I, et al. A virtual try-on system in augmented reality using RGB-D cameras for footwear personalization. J Manuf Syst (2014), http://dx.doi.org/10.1016/j.jmsy.2014.05.006

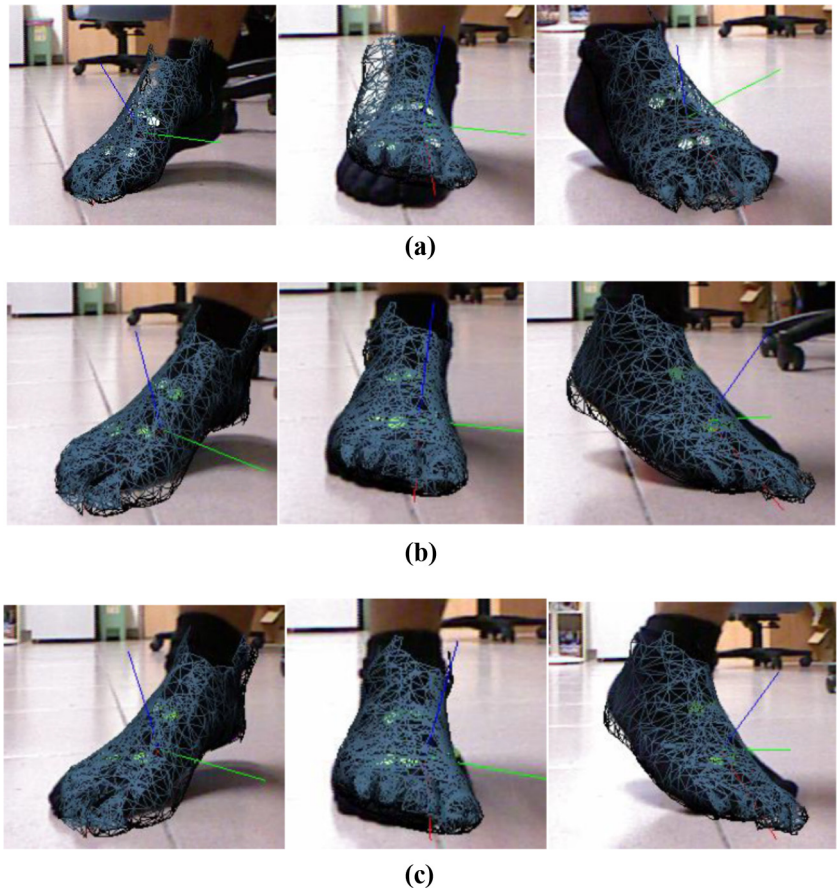

Fig. 7. Test results of three tracking methods (a) color marker (b) ICP algorithm, and (c) the two-stage method.

of the shoe model on the user's foot in various orientations. Marker tracking (first row) produced an unsatisfactory tracking result, and the misalignment between the virtual and real objects was observable. As expected, precisely identifying a curved object using a small amount of flat markers was inadequate. Both ICP-based tracking (second row) and the two-stage approach (third row) yielded more satisfactory results with smaller misalignments; in other words, both methods employing ICP resulted in a more correct alignment between the shoe and the foot.

The computational efficiency of the three methods was compared by estimating the maximal number of frames each produced per second. Each method was used to read all 60 test frames recorded previously, and different tracking mechanisms were then applied to sequentially process each frame; dividing the total number of frames by the accumulated processing time produced the average processing rate. Real-time applications normally require 30 fps to ensure high visualization quality. As shown in Table 1, tracking with color markers required the shortest computation time, because it enabled foot positions to be determined simply using the YIQ space color information; the tracking process did not require iterations. ICP-based tracking had the longest computational time, 20.2 fps. The two-stage method had a higher computational efficiency than using only ICP-based tracking, because the color markers rapidly produced an approximate foot position that ICP-based tracking could subsequently correct; using the two-stage method effectively reduced the number of iterations. The two-stage method had a computational time of 24.3 fps.

Table 1
Comparison of computational efficiency for three tracking methods.

Method	Color markers	ICP only	Two-stage
Performance	>30 fps	20.2 fps	24.3 fps

The tracking accuracy was dependent on the processing rate of the tracking mechanism. This was easily reflected in tracking rapid foot movement. In a second test scenario, the user rapidly moved his foot toward the camera, approximately 100 cm in 0.5 s. Processing the foot motion using marker tracking did not result in a noticeable delay, but the method's positional accuracy was poor, as shown in Fig. 8(a). The position of the target was lost using the ICP-only method and the model deviated considerably from the user's foot. In contrast, the foot location was successfully and accurately recognized when using the two-stage method. Fig. 8 illustrates the tracking results of the three methods. The images are blurred because of the rapidity of the foot movement.

A series of 400 consecutive images were recorded to conduct a detailed performance analysis of the markerless tracking methods. Each image contained both color and depth information. In the tests, the tracking process was terminated when the number of iterations in the ICP algorithm exceeded 10. During the test, the user was asked to move his foot in various ways, such as rotating, swinging at an angle, and jumping. An accumulated tracking error E_{acc} was proposed to estimate and compare tracking accuracy between the two methods. E_{acc} was computed as

$$\sum_{i=1}^{n} \frac{1}{n} ||\boldsymbol{m}_i^* - \boldsymbol{p}_i|| \quad (4)$$

which is the average error of the best matching set generated by the ICP algorithm. Fig. 9 shows that the two-stage method produced a lower number of errors than ICP-only tracking did in almost all test

Please cite this article in press as: Yang Y-I, et al. A virtual try-on system in augmented reality using RGB-D cameras for footwear personalization. J Manuf Syst (2014), http://dx.doi.org/10.1016/j.jmsy.2014.05.006

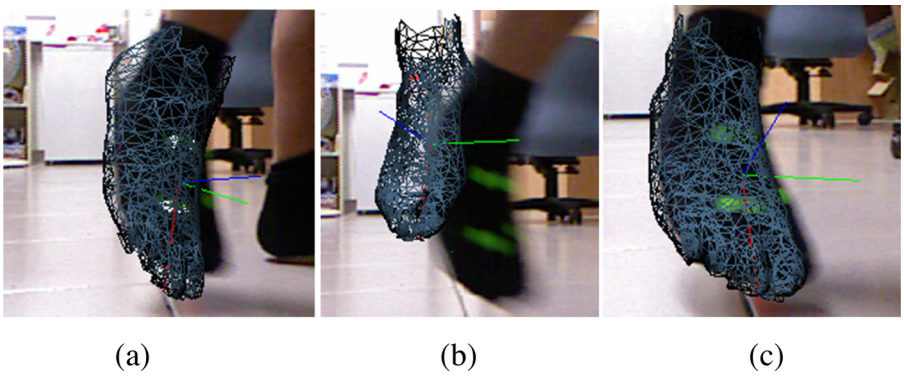

Fig. 8. Tracking results of three tracking methods in rapid foot movement.

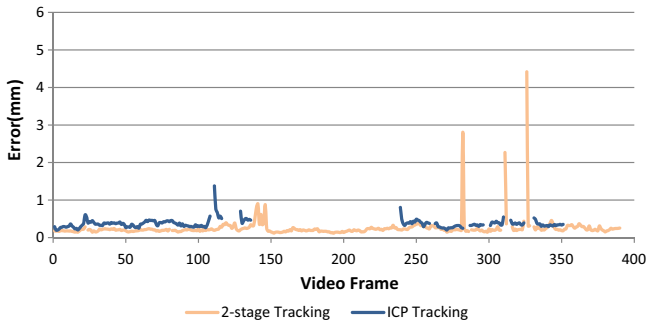

Fig. 9. Tracking results of the two ICP-based methods.

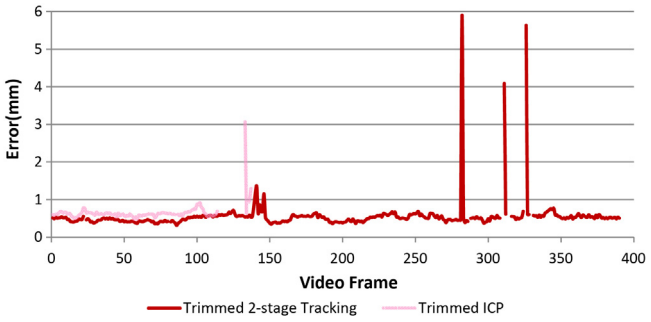

Fig. 11. Tracking performance when using two ICP methods for the complete model.

frames. The position of the user's foot was lost twice when using ICP-only tracking: between frames 135–240 and frames 350–400. No tracking errors are shown for these two periods because the corresponding numerical values were too high to be displayed in the figure.

The influence of model trimming on the performance of the trimmed two-stage method is shown in Fig. 10. ICP exhibited superior positional accuracy when using a trimmed model compared with using a complete model in all of the test frames. Loss of tracking did not occur in either case, because it was precluded by the use of color marker tracking. The main reason for the inferior performance when using the complete model was that the ICP algorithm converged more slowly with a greater number of points to be registered than it did with the trimmed model. Data points around the sole of the foot were removed from the trimmed model because they were invisible from the depth camera, these points did not provide useful information in shape registration, and could be removed without deteriorating the performance of foot tracking. The ICP algorithm converged more quickly and attained improved performance after those redundant points were removed.

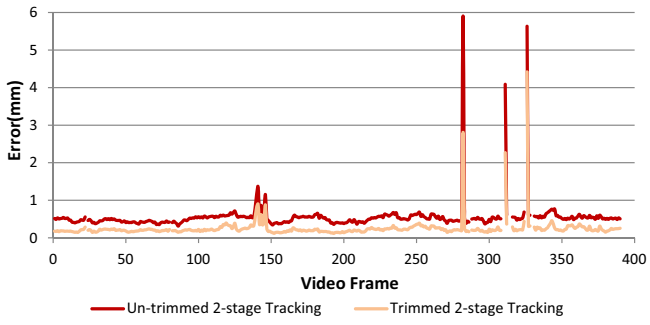

Fig. 10. Tracking performance for complete and trimmed models.

ICP-only tracking exhibited inferior performance when using the complete foot model. As shown in Fig. 11, the position of the foot was completely lost starting from Frame 140, when the user began to move his foot rapidly, and was never relocated. One iteration of the ICP algorithm took longer to execute using the complete model compared with the trimmed one. In real-time applications, visualization quality requires an excess of 30 fps. Without the initial estimate provided by color marker tracking, ICP-only tracking failed to perform satisfactorily after a fixed number of iterations (10). The positional errors became excessive when the user moved

Fig. 12. Start of the virtual try-on system.

Please cite this article in press as: Yang Y-I, et al. A virtual try-on system in augmented reality using RGB-D cameras for footwear personalization. J Manuf Syst (2014), http://dx.doi.org/10.1016/j.jmsy.2014.05.006

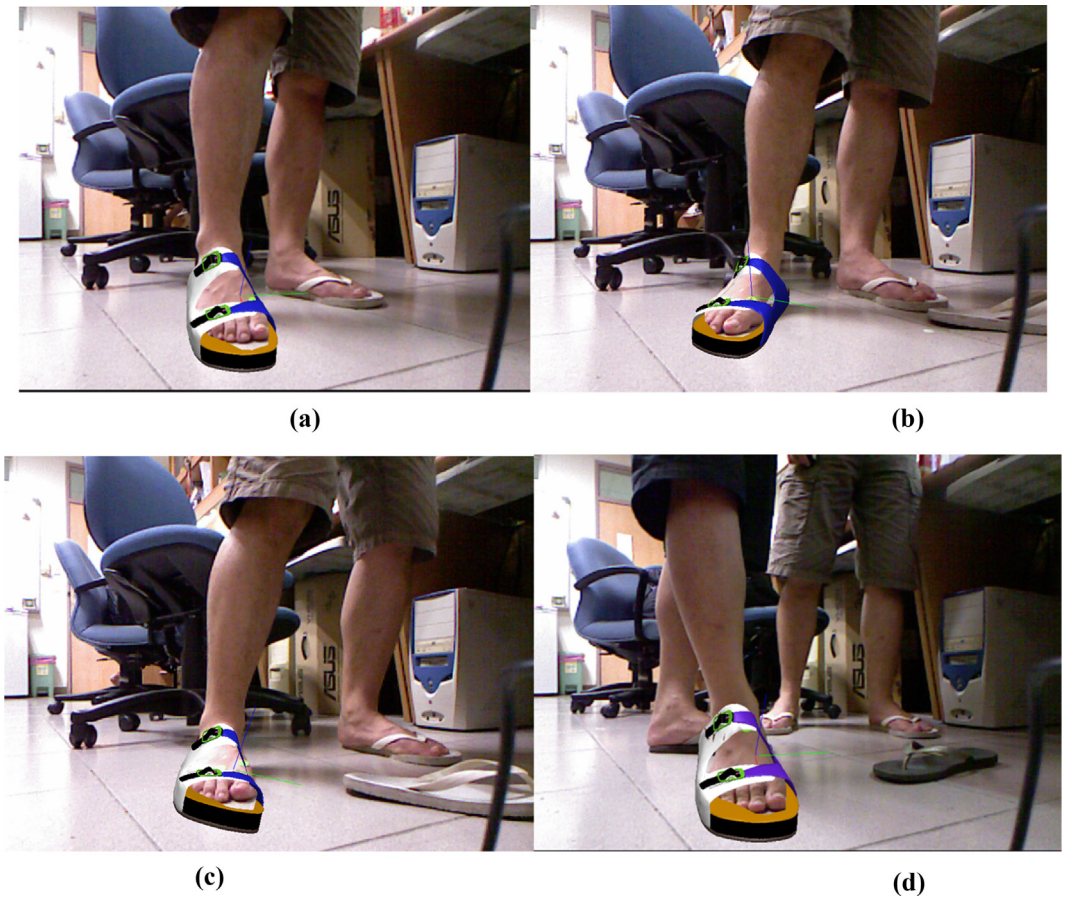

Fig. 13. (a)–(c) Trying on a saddle model in different poses and (d) trying on a larger saddle model.

Fig. 14. The try-on process in various foot movements.

Please cite this article in press as: Yang Y-I, et al. A virtual try-on system in augmented reality using RGB-D cameras for footwear personalization. J Manuf Syst (2014), http://dx.doi.org/10.1016/j.jmsy.2014.05.006

his foot rapidly, and the foot position could not be identified based on the images.

The following use scenarios demonstrate the usefulness of the proposed system in helping users evaluate footwear designs. The system prompts the user to select a shoe model to try on, and the try-on process begins when the user places the six color markers on his or her foot instep, as shown in Fig. 12. A saddle is automatically positioned to the user foot in the video stream after clicking the "Start Tracking" button. Fig. 13(a)–(c) shows images of the try-on process captured from various view angles. The shoe size can be adjusted, and the corresponding model change occurs immediately, as shown in Fig. 13(d). A previous study conducted by the authors [29] described the size adjustment mechanism; thus, its details are omitted here. The user can slowly move his or her foot in various poses, and the shoe model remains superimposed on the foot during the motion (see Fig. 14).

6. Conclusion and future work

Modern consumers are attracted to products that reflect their tastes; mass customization has thus become an effective strategy in product development. Implementing this strategy requires active participation and prompt feedback from product users. Most tools used to design fashion products have been constructed from the perspective of designers, rather than that of product users. Such tools offer limited utility in engaging end users to actively participate in the design process. Allowing users to express design ideas and instantly interact with product prototypes is desirable. Recent progress in MR and RGB-D cameras has provided effective tools for realizing this goal. This paper presents a virtual try-on system for footwear, focused on evaluating shoe styling. This system involved an implementation of mixed reality, in which a shoe model is continuously positioned with respect to a user's foot in a video stream. The study developed three tracking methods for recognizing the foot location according to the color and depth of images captured using Kinect. Tracking using color markers placed on the foot instep offered rapid identification but poor accuracy; tracking using only 3D shape registration (i.e., using only the ICP algorithm) required a lengthy computation time, precluding its use in real-time applications. A two-stage tracking approach was proposed to maintain a favorable balance between tracking accuracy and efficiency by employing color markers to rapidly determine a preliminary estimation for subsequent ICP-based tracking. This reduced the number of iterations of the ICP algorithm, and maintained accuracy. Various foot movements were recorded in a series of continuing frames, and the performances of ICP-based tracking and the two-stage method were compared regarding positional errors. The results revealed that the two-stage method exhibited superior performance, and the ICP-based tracking method failed to provide accurate position recognition when the foot was moved rapidly. In addition, trimming the reference model using an instant camera angle accelerated the convergence rate in the ICP algorithm, reducing the number of data points used in shape registration. In conclusion, the two-stage tracking method using the trimmed reference model offered superior performance in the virtual try-on system. Several design scenarios demonstrated the effectiveness of the proposed system.

A virtual try-on system not only enables consumers to evaluate fashion products but also engages consumers during the design process. The objective of the mass customization of human-centered products can thus be realized. Future studies can extend this research by addressing the functional evaluation of shoe design (e.g., estimating and improving wear comfort) as well as simulating the deformation of shoe models subjected to various foot motions. The current system only uses a single reference foot model in the tracking process, which may not match well with all different users. Automatic re-sizing of the reference model based on foot size measurement may solve this problem.

References

[1] Fralix M. From mass production to mass customization. J Text Appar Technol Manage 2011;1(2):1–7.
[2] Berry C, Wang H, Hu SJ. Product architecting for personalization. J Manuf Syst 2013;32(3):404–11.
[3] Huang SH, Yang YI, Chu CH. Human-centric design personalization of 3D glasses frame in markerless augmented reality. Adv Eng Inform 2012;16:35–45.
[4] Tseng MM, Piller FT. The customer centric enterprise, advances in mass customization and personalization. Berlin: Springer; 2003.
[5] Fontana M, Rizzi C, Cugini U. 3D virtual apparel design for industrial applications. Comput Aided Des 2005;37(6):609–22.
[6] Gross C, Fuhrmann A, Luckas V, Encarnacão J. Virtual try-on: topics in realistic, individualized dressing in virtual reality. In: Proc. of the Virtual and Augmented Reality Status. 2004.
[7] Chin S, Kim KY. Facial configuration and BMI based personalized face and upper body modeling for customer-oriented wearable product design. Comput Ind 2010;61(6):559–75.
[8] Lu SC-Y, Shpitalni M, Gadh R. Virtual and augmented reality technologies for product realization. CIRP Ann 1999;48(2):471–95.
[9] Wang CCL, Hui KC, Tong KM. Volume parameterization for design automation of customized free-form products. IEEE Trans Automat Sci Eng 2007;4(1):11–21.
[10] Meng Y, Wang CCL, Jin X. Flexible shape control for automatic resizing of apparel products. Comput Aided Des 2012;44(1):68–76.
[11] El-Laithy RA, Huang J, Yeh M. Study on the use of Microsoft Kinect for robotics applications. In: Position Location and Navigation Symposium (PLANS), IEEE/ION. 2012. p. 1280–8.
[12] Chang YJ, Chen SF, Huang JD. A Kinect-based system for physical rehabilitation: a pilot study for young adults with motor disabilities. Res Dev Disabil 2011;32(6):2566–70.
[13] Wang CS, Chiang DJ, Wei YC. Intuitional 3D museum navigation system using Kinect. Lect Notes Electr Eng 2013;58:7–596.
[14] Hauswiesner S, Straka M, Reitmayr G. Virtual try-on through image-based rendering. IEEE Trans Vis Comput Graph 2013;19(9):1552–65.
[15] Tong J, Zhou J, Liu L, Pan Z, Yan H. Scanning 3D full human bodies using Kinects. IEEE Trans Vis Comput Graph 2012;18(April (4)):643–50.
[16] Li J, Ye J, Wang Y, Bai L, Lu G. Technical section: fitting 3D garment models onto individual human models. Comput Graph 2010;34:742–55.
[17] Meng Y, Mok PY, Jin X. Interactive virtual try-on clothing design systems. Comput Aided Des 2010;42(4):310–21.
[18] Antonio JM, Jose Luis SR, Faustino SP. Augmented and virtual reality techniques for footwear. Comput Ind 2013;64(9):1371–82.
[19] Luh YP, Wang JB, Chang JW, Chang YY, Chu CH. Augmented reality based mass customization of shoe design. J Intell Manuf 2013;24(5):905–17.
[20] Tsuyoshi N. Footwear fitting design. In: Emotional engineering. London: Springer-Verlag; 2011. p. 345–63.
[21] Buchsbaum WH. Color TV servicing. 3rd ed. Englewood Cliffs, NJ: Prentice Hall; 1975.
[22] Jolliffe IT. Principle component analysis. 2nd ed. New York: Springer; 2002.
[23] Platzer W. Color atlas of human anatomy. Locomotor system, vol. 1, 5th ed; 2004.
[24] Besl P, McKay N. A method for registration of 3-D shapes. IEEE Trans Pattern Anal Mach Intell 1992;14:239–56.
[25] Li Y, Gu P. Automatic localization and comparison for free-form surface inspection. J Manuf Syst 2006;25(4):251–68.
[26] http://www.hitl.washington.edu/artoolkit/
[27] Adams R, Bischof L. Seeded region growing. Pattern Anal Mach Intell 1994;16:641–7.
[28] Eggert DW, Lorusso A, Fisher RB. Estimating 3-D rigid body transformations: a comparison of four major algorithms. Mach Vis Appl 1994;9:272–90.
[29] Wang TY, Ke JC, Chang FM. Analysis of a discrete-time queue with server subject to vacations and breakdowns. J Ind Prod Eng 2013;30(1):54–66.

Please cite this article in press as: Yang Y-I, et al. A virtual try-on system in augmented reality using RGB-D cameras for footwear personalization. J Manuf Syst (2014), http://dx.doi.org/10.1016/j.jmsy.2014.05.006