Scanning 3D Full Human Bodies Using Kinects

Jing Tong, Jin Zhou, Ligang Liu, Zhigeng Pan, and Hao Yan

Abstract—Depth camera such as Microsoft Kinect, is much cheaper than conventional 3D scanning devices, and thus it can be acquired for everyday users easily. However, the depth data captured by Kinect over a certain distance is of extreme low quality. In this paper, we present a novel scanning system for capturing 3D full human body models by using multiple Kinects. To avoid the interference phenomena, we use two Kinects to capture the upper part and lower part of a human body respectively without overlapping region. A third Kinect is used to capture the middle part of the human body from the opposite direction. We propose a practical approach for registering the various body parts of different views under non-rigid deformation. First, a rough mesh template is constructed and used to deform successive frames pairwisely. Second, global alignment is performed to distribute errors in the deformation space, which can solve the loop closure problem efficiently. Misalignment caused by complex occlusion can also be handled reasonably by our global alignment algorithm. The experimental results have shown the efficiency and applicability of our system. Our system obtains impressive results in a few minutes with low price devices, thus is practically useful for generating personalized avatars for everyday users. Our system has been used for 3D human animation and virtual try on, and can further facilitate a range of home-oriented virtual reality (VR) applications.

Index Terms—3D Body Scanning, global non-rigid registration, Microsoft Kinect.

1 Introduction

Many computer graphics applications, such as animation, computer games, human computer interaction and virtual reality, require realistic 3D models of human bodies. Using 3D scanning technologies, such as structured light or laser scan, detailed human models could be created [1]. However, these devices are very expensive and often require expert knowledge for the operation. Moreover, it is difficult for people to stay rigid during the capturing process. Image-based method is another solution for human modeling. The state-of-the-art multi-view method can get very impressive results [2]. But this kind of methods is computationally expensive, and they have problems when there are sparse textures or complex occlusions among different views [3].

As a new kind of devices, depth cameras such as Microsoft Kinect [4] have attracted much attention in the community recently. Compared with conventional 3D scanners, they are able to capture depth and image data at video rate and have little consideration of the light and texture condition. Kinect is compact, low-price, and as easy to use as a video camera, which can be acquired by general users.

Some researchers have tried to use Kinects as 3D scanners. For example, [5] used RGB images along with per-pixel depth information to build dense 3D maps of indoor environments. However, the main problem is that Kinect has a comparably low X/Y resolution and depth accuracy for 3D scanning. To address this issue, [6] described a method to improve the data's quality by depth super resolution. Using a Kinect and commodity graphics hardware, [7] presented a system for accurate real-time mapping of complex and arbitrary indoor scenes. However, most of these work only used Kinect to scan rigid objects.

- *Jing Tong is with State Key Laboratory of CAD&CG at Zhejiang University; College of Computer & Information Engineering at HoHai University, E-mail: tongjing.cn@gmail.com.*
- *Jin Zhou is with Institute of Applied Mathematics and Engineering Computations at Hangzhou Dianzi University, E-mail: zhoujin10@gmail.com.*
- *Ligang Liu is with Department of Mathematics at Zhejiang University, E-mail: ligangliu@zju.edu.cn.*
- *Zhigeng Pan is with Digital Media and HCI Research Center at Hangzhou Normal University, E-mail: zhigengpan@gmail.com.*
- *Hao Yan is with State Key Laboratory of CAD&CG at Zhejiang University, E-mail: yhao880514@gmail.com.*

Manuscript received 15 September 2011; accepted 3 January 2012; posted online 4 March 2012; mailed on 27 February 2012.
For information on obtaining reprints of this article, please send email to: tvcg@computer.org.

There are a couple of works that used Kinect to capture human faces. [8] presented an algorithm for computing a personalized avatar from a single color image and its corresponding depth map. [9] further captured and tracked the facial expression dynamics of the users in real-time and map them to a digital character.

To scan a full human shape, Kinect should be put around 3 meters away from the body, and the resolution is very low, as shown in Figure 1(a). Little geometry information is captured in the depth map. Though using the information of multiple frames to enhance the final resolution [6], the result is still not acceptable.

Recently, [10] estimated the body shape using SCAPE model [11] by image silhouettes and depth data from one single Kinect. However, due to the limited subspace of parametric model, personalized detailed shapes, such as faces, hairstyles, and dresses cannot be reconstructed by using this method. Furthermore, it takes approximately 65 minutes to optimize, which seems too slow for some practical applications in VR.

In this paper, we present a system to scan 3D full human body shapes using multiple Kinects. Each Kinect scan different part of the human body so that they can be put close to the body to obtain higher quality of data. However, there are two challenges in designing this kind of scanning system.

First, the data quality of the overlapping region between two Kinects is reduced due to the interference between them. Shutters can be used to allow different Kinects to capture data at different time. However, it reduces the frame rates and exposure time, which also reduces data quality. To avoid the interference issue, we use two Kinects to capture the upper part and the lower part of a human body from one direction, respectively. There is no overlapping region between these two parts. We use a third Kinect to capture the middle part of the human body from the opposite direction. A person is standing in-between the Kinects and turns around with the help of a turntable.

The other challenge is that the body can hardly be kept still during the capturing process. Thus non-rigid registration of the captured data is required. However, it is a difficult job due to the low quality of the data and the complex occlusion. We propose a two-phase method to deal with this challenge. In the first phase, pairwise non-rigid deformation is performed on the geometry field based on a rough template. In the second phase, a global alignment with loop closure constraint is used on the deformation field.

Our scanning system is easy to be built, as shown in Figure 2 as well as the accompanying video. The experimental results have shown that our system captures impressive 3D human shapes with personalized detailed shapes such as hairstyles and dress wrinkles. With the total cost of about $600, our system is much cheaper than the conventional

3D scanners, and can be used for lots of home-oriented VR applications.

The contributions of this paper are summarized as follows:

- A full body scanning system using multiple Microsoft Kinects is built;
- A non-rigid registration method with a rough template is proposed;
- A global non-rigid alignment method which deals with occlusions and meets the loop closure constraint is proposed.

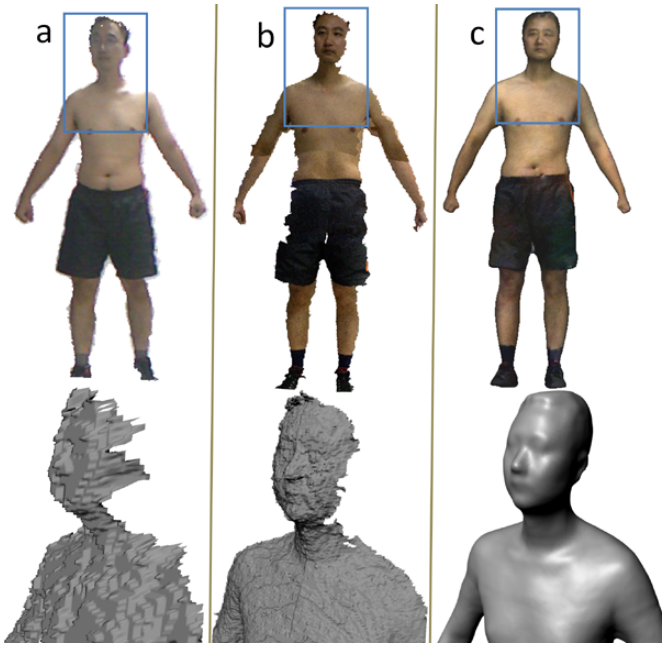

Fig. 1. (a) The raw data of capturing a full human body with one single Kinect has much low quality as the Kinect has to be put far from the body. (b) The raw data captured using our system has higher quality as multiple Kinects are used to capture different parts of the body at a closer distance. (c) The reconstructed human model created using our method.

2 RELATED WORK

2.1 Scanning Systems

Acquiring 3D geometric content from real world is an essential task for many applications in computer graphics. Unfortunately, even for static scenes, there is no low-priced off-the-shelf system, which can provide good quality, high resolution distance information in real time [12]. Scanning devices based on structured light or laser scan can capture human body with much high quality. However, these devices are expensive and often require special knowledge to operate. For example, the price of Cyberware Whole Body Color 3D Scanner is about $240,000 [13].

Being a newly developed distance measuring hardware, the depth camera technology opens a new epoch for 3D content acquisition. There are two main approaches employed in depth camera technologies currently. The first one is based on the time-of-flight (ToF) principle [12], measuring time delay between transmissions of a light pulse. Because of the specific chip needed for active lighting pulse, the price of Swiss Ranger 4000 is about $8,000. The second approach is based on light coding, projecting a known infrared pattern onto the scene and determining depth based on the pattern's deformation. Such device only needs standard infrared CMOS imager, so the cost is much lower than the ToF device. A most popular one is the Microsoft Kinect sensor [4], which is at a price of only $150. In this paper, we have designed a scanning system for automatically capturing 3D full human bodies by using 3 Kinects, which can be purchased by everyday users for home-oriented VR applications.

2.2 Non-rigid Registration

As human bodies' non-rigid deformation in the scanning process, we need to register the scanned data. There have been three main types of methods for non-rigid registration in the literature.

Registration without a template. This kind of methods often requires high quality scan data, and often needs small changes in temporal coherence. For example, [14] puts all scans into a 4D space-time surface and uses kinematic properties to track points and register multiple frames. [15] registers two point clouds that undergo approximate isometric deformations.

Registration with a template. Such kind of methods often needs to acquire a relative accurate template, and then uses the template to fit each scan. Some works use markers to fit the template [1], and the others utilize global optimization [16]. In [16], a smooth template using static scanner is generated and is then registered to each of the input scan using a non-linear, adaptive deformation model [17].

Registration with a semi-template. Accurate template can hardly be obtained in many cases. However, rough template, such as the skeleton model of articulated object, can be generally utilized. [18] manually segments the first frame, then identifies and tracks the rigid components of each frame, while accumulating the geometric information. [19] presents an articulated global registration algorithm as the optimization of both the alignment of range scans and the articulated structure of the model.

The first type of methods often requires high quality input data, and is computationally expensive; the second one needs an accurate template, which is hard to fulfil for many applications, especially for deforming objects. Our method uses a rough template which is constructed by the first data frame. Based on the feature correspondences returned in the corresponding color maps, the template is deformed [17] and thus it can drive the points accordingly. Thanks to the deformation model [17], the points in different frames can then be approximately aligned.

2.3 Global Alignment

Global alignment, especially the loop closure problem, is well-known in rigid scanning [20, 21]. A brute-force solution is to bring all scans into the Iterated Closest Point (ICP) [22] iteration loop. However, this often requires solving awfully large systems of equations. Another greedy solution is to align each new scan to all previous scans [23]. However, it cannot spread out errors of previous scans, and is not guaranteed to achieve consistent cycle.

We agree with the idea of creating a graph of pairwise alignments between scans [24, 25]. First, pairwise rigid alignment is computed in the geometry level. Global error distribution then operates on an upper level, where errors are measured in terms of the relative rotations and translations of pairwise alignments. The graph methods can simultaneously minimize the errors of all views rapidly, and do not need all scan in memory. This makes it especially practical for models with lots of scans.

Global alignment has been less studied in non-rigid registration. [26] warps two consecutive frames by feature point correspondences and Laplacian coordinates constraints [27]. To align all frames simultaneously, a global matrix system lists all constraints as Diagonal submatrices is solved, which is similar to the brute-force solution. [18] maintains the accumulated 3D point clouds of previous frames, and registers the points of a new input frame to the accumulated points, which suffers the same problem of the greedy solution.

Another problem of non-rigid registration is occlusion. In most previous papers, occluded parts are often predicted by their temporal or spatial neighbors [19, 26]. The interpolation based methods are hard to get correct registration due to complex occlusions. In our method, misalignment caused by complex occlusion can also be handled reasonably in the global alignment procedure.

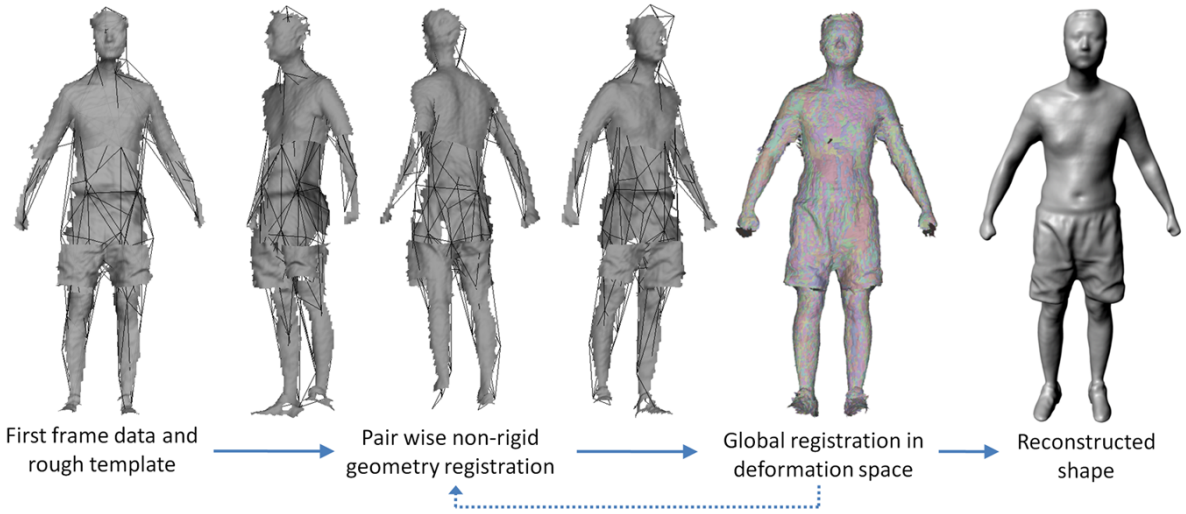

Fig. 3. Overview of our reconstruction algorithm.

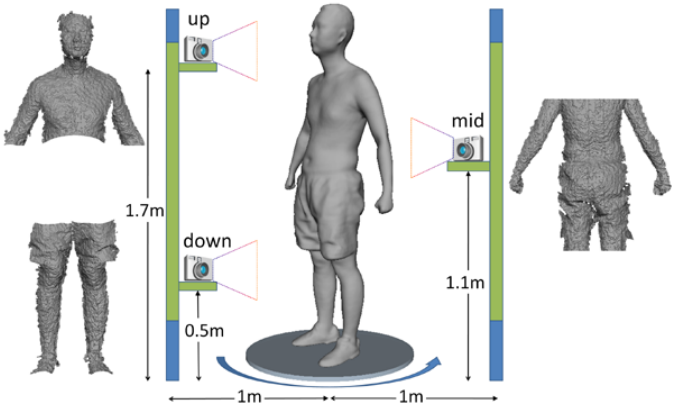

Fig. 2. The setup of our system. Three Kinects are used to capture different parts of a human body at a close distance without interference. The captured data of a single frame from three sensors is also illustrated.

3 SYSTEM SETUP

The setup of our scanning system is illustrated in Figure 2. To avoid the interference problem, two Kinects are used to capture the upper part and the lower part of a human body respectively, without overlapping region, from one direction. A third Kinect is used to capture the middle part of the human body from the opposite direction. The distance between two sets of Kinects are about 2 meters. A turntable is put in between them.

While scanning, the person stands on the turntable and rotates 360 degrees in about 30 seconds. Please see the accompanying video. Each Kinect acquires 1280×1024 color images and 640×480 depth images at 15 frames per second individually. 3D coordinates can be automatically generated using OpenNI [28]. Using [28], the depth and color image are also well calibrated. The captured data from three sensors are synchronized and calibrated automatically [29].

A depth and color threshold method is used to roughly segment the foreground human data from the background. Then, sliver faces followed by vertices not referenced by any face are deleted. As the body is only 1 meter away from the Kinects, the quality of depth values, compared with that in Figure 1(a), has been improved greatly, as shown in Figure 2. Laplacian smooth [27] is performed to reduce the noise. To reduce the computation cost, simplified mesh with $1/10$ vertices is used in our experiment, as shown in Figure 3.

With three Kinects ($450), two pillars which fix the Kinects ($30) and one turntable ($120), the total cost of our scanning system is about $600, which is much cheaper than the conventional 3D scanning devices. Thus the system can be utilized for many home-oriented applications.

4 RECONSTRUCTION APPROACH

Denote $D_i = \{M_i, I_i\}, i = 1, \ldots, n$ as the captured data, where n is the number of the captured frames, M_i is the merged mesh and I_i is the corresponding image of the i-th frame respectively.

4.1 Overview

Generally the body can hardly be kept rigid when the person stands on the turntable during the scanning process. For example, the arms and head may be moving and the chest is deformed due to breathing.

We propose a practical approach for registering multiple frames of noisy partial data of human body under non-rigid deformation. Figure 3 shows an overview of our system. A rough template is constructed by the depth data of the first frame. Then the template is used to deform the geometry of successive frames pairwisely. Global registration is then performed to distribute errors in the deformation space, where problems of cycle consistency and occlusion are handled. The pairwise registration and global registration iterates until convergence. Then, every frame is deformed to the first frame driven by the templates. Finally, reconstructed model is generated using Poisson reconstruction method [30].

4.2 Template Generation

Unlike the work of [16], an accurate template is unavailable in our system. We use the method proposed in [31] to construct an estimated body shape as the template mesh T_1 from the first frame. Due to the data noise and influence of dresses, the resulting template can only approximate the geometry of the body shape, as shown in Figure 4. It is impossible to use this template to register each frame by geometry fitting. We will show that it is enough to use this template to track the pairwise deformation of successive frames. Denote $T_1 = \{v_1^k\}, k = 1, ..., K$ where K is the number of the nodes of T_1. To reduce the computation cost, we simplify T_1 so that about 50-60 nodes are used in our system.

4.3 Pairwise Geometry Registration

Suppose $M_i, i = 1, ..., n$ forming a cycle. $f_{i,j}$ denotes the registration that can deform mesh M_i to register with mesh M_j. In this section, we introduce how to find the pairwise registration $f_{1,2}, f_{2,3}, ..., f_{n-1,n}, f_{n,1}$.

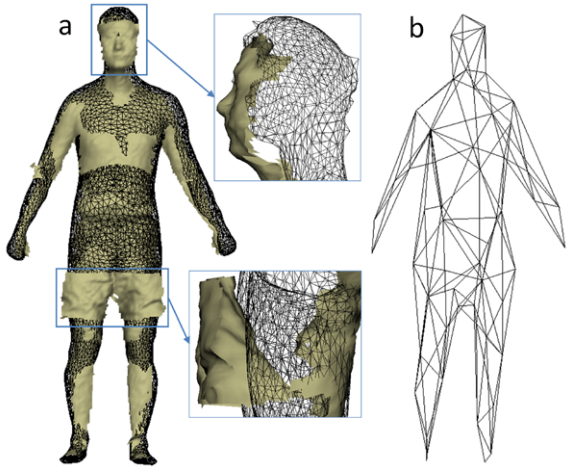

Fig. 4. An estimated body shape is constructed from the first frame as the template mesh (a). To reduce the computation cost, the template mesh is simplified to 50-60 nodes (b).

Deformation model. Our deformation model is based on [17]. Suppose we have two meshes M_i and M_j, and the template mesh at frame i is T_i. Denote $f_{i,j} = \{(R_i^k, t_i^k) | v_i^k \in T_i\}$, where v_i^k is a node of T_i, 3×3 matrix R_i^k and 3×1 vector t_i^k specify the local deformation induced by v_i^k. Specifically, for a point p of M_i near v_i^k, its destiny $\tilde{p}$ can be calculated as $\tilde{p} = R_i^k(p - v_i^k) + v_i^k + t_i^k$.

Pairwise registration. For two successive frames M_i and M_{i+1}, corresponding feature points can be obtained by optical flow [32] in the corresponding images (see Figure 5). Particularly, we use [33] to find the feature correspondences between M_n and M_1. Following [17], we compute (R_i^k, t_i^k) by solving

$$\min_{(R_i^k, t_i^k)} (E_{\text{reg}} + w_{\text{rot}} E_{\text{rot}} + w_{\text{con}} E_{\text{con}})$$

Suppose v_i^j, v_i^k are two neighboring nodes of M_i, $E_{\text{reg}} = \sum_k \sum_{j\in N(k)} \left\| R_i^k(v_i^j - v_i^k) + v_i^k + t_i^k - (v_i^j + t_i^j) \right\|$ ensures the smoothness of the neighboring deformation. Let c_1, c_2, c_3 be the 3×1 column vectors of 3×3 matrix R, $\text{Rot}(R) = (c_1 \cdot c_2)^2 + (c_1 \cdot c_3)^2 + (c_2 \cdot c_3)^2 + (c_1 \cdot c_1 - 1)^2 + (c_2 \cdot c_2 - 1)^2 + (c_3 \cdot c_3 - 1)^2$. $E_{\text{rot}} = \sum_k \text{Rot}(R_i^k)$ is used to specify the affine transformation to be rotation. Suppose v_k^l and $v_{k+1}^{\text{index}(l)}$ are corresponding feature points of two successive frames, $\tilde{v}_k^l$ is the deformed point of v_k^l. $E_{\text{con}} = \sum_l \left\| \tilde{v}_k^l - v_{k+1}^{\text{index}(l)} \right\|$ is used to match the feature points constraints. The energy is minimized using standard Gauss-Newton algorithm as described in [17].

Projection to the first frame. After the pairwise registration, we can find that it is over determined. Since only $n-1$ pairwise deformation is needed to recover all the relative position of all frames. Here, to deform M_j back to the first frame, we simply let $\tilde{M}_j = f_{2,1}(...(f_{j-1,j-2}(f_{j,j-1}(M_j))))$, where $f_{j,j-1}$ is the inverse transformation of $f_{j-1,j}$. The result is shown in Figure 6(a). In the head area, the first and last frames do not match due to error accumulation of pairwise registration. The arm area has more serious problem of misalignment which is caused by complex occlusions. So we need a global deformation registration to solve these issues.

4.4 Global Deformation Registration

Similar to the problem arisen in rigid registration [25], the desired pairwise deformation $\hat{f}_{1,2}, \hat{f}_{2,3}, ..., \hat{f}_{n-1,n}, \hat{f}_{n,1}$ should meet the following two conditions:

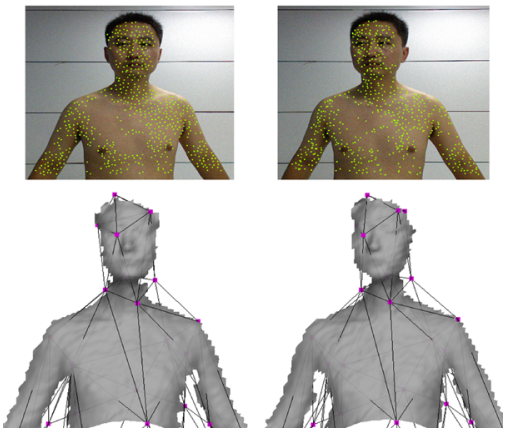

Fig. 5. Corresponding feature points of two successive frames and the deformed templates.

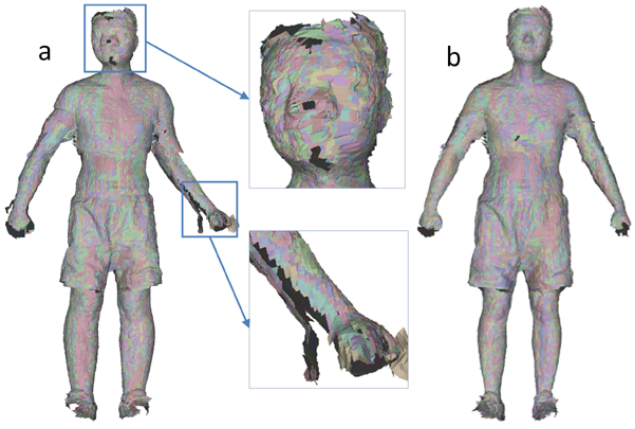

Fig. 6. (a) After pairwise registration, the first and last frame does not well match due to error accumulation, as shown in the nose part. More serious problem occurs by misalignment of complex occlusions, as shown in the hand part. (b) Global deformation registration is used to deal with these problems.

1. It is cycle consistent, that is, the composition of any deformation around a cycle should be identity:

$$\forall i, \hat{f}_{i,i+1}\hat{f}_{i+1,i+2}...\hat{f}_{n-1,n}\hat{f}_{n,1}\hat{f}_{1,2}...\hat{f}_{i-1,i} = I \quad (1)$$

2. The original pairwise deformation is relatively correct, so we should minimize the weighted squared error of the new and old deformation:

$$\min \sum w_{i,j}^2 \left\| \hat{f}_{i,j} - f_{i,j} \right\|^2 \quad (2)$$

where the weight $w_{i,j}$ is the confidence of each pairwise deformation $f_{i,j}$. Here, we set $w_{i,j} = 1/\text{Dist}(f_{i,j}(M_i), M_j)$, where $\text{Dist}(f_{i,j}(M_i), M_j)$ is the average distance of corresponding point pairs of $f_{i,j}(M_i)$ and M_j.

To satisfy these conditions, let's check a node v_i of template T_i. In the neighbor of v_i, the local deformation can be approximated as a rotation $r_{i,i+1}$ and translation $t_{i,i+1}$. Finding optimal $\hat{f}_{i,i+1}$ is equivalent to finding $\hat{r}_{i,i+1}, \hat{t}_{i,i+1}$ for each node.

To reduce the complexity of the problem, following [25], we first consider translation to be independent of rotation, and focus on calculating rotation $\hat{r}_{i,i+1}$. Let the matrix $E_{i,i+1}$ be the rotation such that

$$r_{1,2}r_{2,3}...r_{i,i+1}E_{i,i+1}r_{i+1,i+2}...r_{n-1,n}, r_{n,1} = I.$$

Let $\alpha_{j,j+1} = \frac{1}{w_{j,j+1}^2} \Big/ \sum_j \frac{1}{w_{j,j+1}^2}$, $E_{i,i+1}^{<\alpha>} = \exp\{\alpha \ln E_{i,i+1}\}$. It has

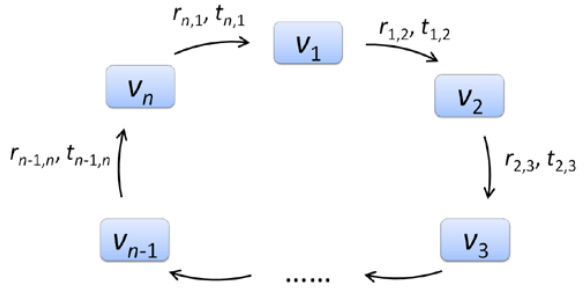

Fig. 7. A cycle of locally rigid pairwise registration. v_1 is registered to v_2, v_2 is registered to v_3, ... , v_n is registered to v_1.

been proven that $\hat{r}_{i,i+1} = r_{i,i+1}E_{i,i+1}^{<\alpha_{i,i+1}>}$ satisfies the cycle consistent constraint (1), and minimizes the energy (2) in the meaning of squared angular error [25].

After $\hat{r}_{i,i+1}$ is set, we distribute the accumulated translation error. $\hat{t}_{i,i+1}$ should also satisfy the cycle consistent constraint:

$$\hat{r}_{i,i+1}...\hat{r}_{i-2,i-1}\hat{t}_{i-1,i} + ... + \hat{r}_{i,i+1}\hat{t}_{i+1,i+2} + \hat{t}_{i,i+1} = 0,$$

and minimize the energy:

$$\sum w_{i,j}^2 \left\| \hat{t}_{i,j} - t_{i,j} \right\|^2.$$

This is a quadratic minimization problem with linear constraints, and it can be solved using Lagrange multipliers.

The weighted error distribution strategy takes advantage of both the pairwise registration and cycle consistency. To illustrate it, let's take a look at two extreme cases. Suppose $r_{i,i+1}, t_{i,i+1}$ deform the neighbor of v_i to the neighbor of v_{i+1} exactly, then we have

$$\hat{r}_{i,i+1} = r_{i,i+1}E_{i,i+1}^{<\alpha_{i,i+1}>} \approx r_{i,i+1}E_{i,i+1}^{<0>} = r_{i,i+1}.$$

In this case, the correct pairwise deformation is reserved in the resulting deformation.

Another extreme case often happens when there is complex occlusion, as shown in Figure 8. Due to large number of frames of occlusion, the left arm in frame i, j may have little intersection area, so the resulting $r_{i,j}, t_{i,j}$ can totally misalign the corresponding areas. However, since we have

$$\hat{r}_{i,j} = r_{i,j}E_{i,j}^{<\alpha_{i,j}>} \approx r_{i,j}E_{i,j}^{<1>} = r_{i,i-1}...r_{2,1}r_{1,n}r_{n,n-1}...r_{j+1,j},$$

reasonable registration can still be obtained using the cycle consistent constraint.

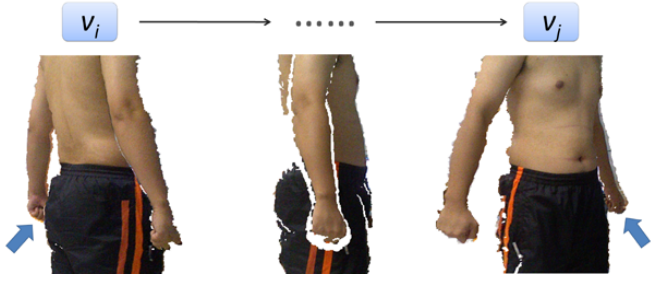

Fig. 8. Complex occlusion happens during the scanning process. In this example, left arm is occluded from frame $i+1$ to frame $j-1$.

5 RESULTS

Figure 9(a) shows the result of global rigid alignment. It can be seen the result is not so good, especially in the arm and head areas. Our

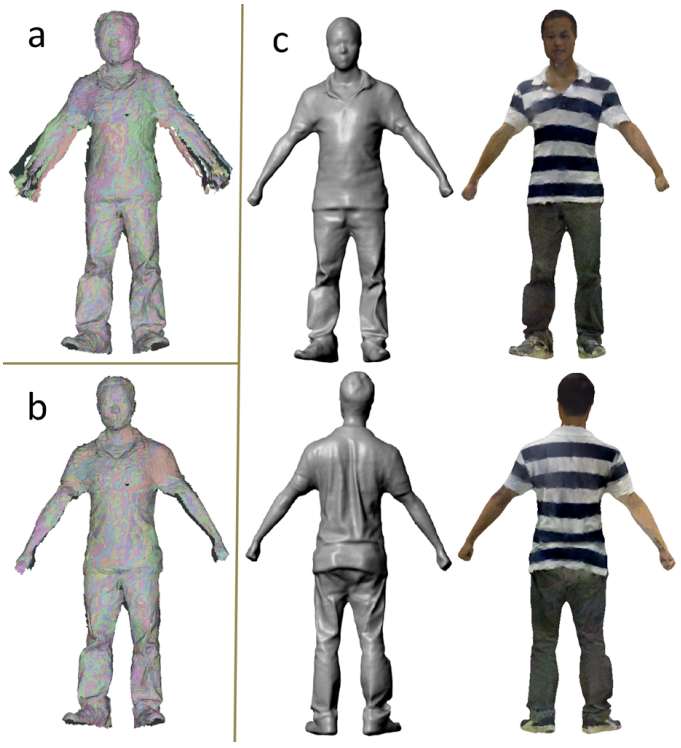

Fig. 9. The result of global rigid alignment (a) and our global non-rigid alignment algorithm (b); (c) the reconstructed model using our method. Note that the dress wrinkles are well captured in the model.

global non-rigid registration can get much better result (Figure 9(b)), and can further get very impressive model (Figure 9(c)).

To align two successive frames, Harris corners are found in the first image, and the corresponding pixels of the second image are found as the ICP closest point on meshes. Using these initial estimations, Lucas-Kanade optical flow [32] is used to find more accurate corresponding feature points. For areas with little texture information, closest point pairs are used directly. For each pairwise registration, about 2000 pairs of corresponding feature points are used.

Occlusions are inevitable in our experiment, especially in the arm area. In our experiment, one arm will be occluded for about $1/6$ of all captured frames. The interpolation based methods are hard to get correct registration, while reasonable results can still be obtained using the loop closure constraint at the global registration procedure.

After one step of global alignment, all meshes are deformed, and pairwise deformation can be further calculated. The pairwise and global registration iterates when the average distance of all neighboring meshes is above 50% of the average edge length. In our experiment, the algorithm converges for about 1 to 2 iterations. Finally, 1/6 uniformly sampled frames are used to generate reconstructed model by Possion surface reconstruction [30]. Since the color image and depth image can be simultaneously captured and calibrated [28], the color information of deformed mesh is generated automatically. Using the method of [34], textured information can be generated for the reconstructed model.

Figure 10 shows more results of our algorithm. The average computing time of each step with an Intel Core i3 processor at 3.1 GHz is shown in Table 1. Six biometric measurements are calculated on the constructed human models and are compared with those measured on the corresponding real persons. The average error in centimeter is shown in Table 1. It is seen that our reconstructed models approximate the real persons well. Larger errors in the girth measures are caused by the dresses.

The accuracy of our biometric measurements is similar to the result shown in [10]. Unlike [10] which uses the subspace of SCAPE presen-

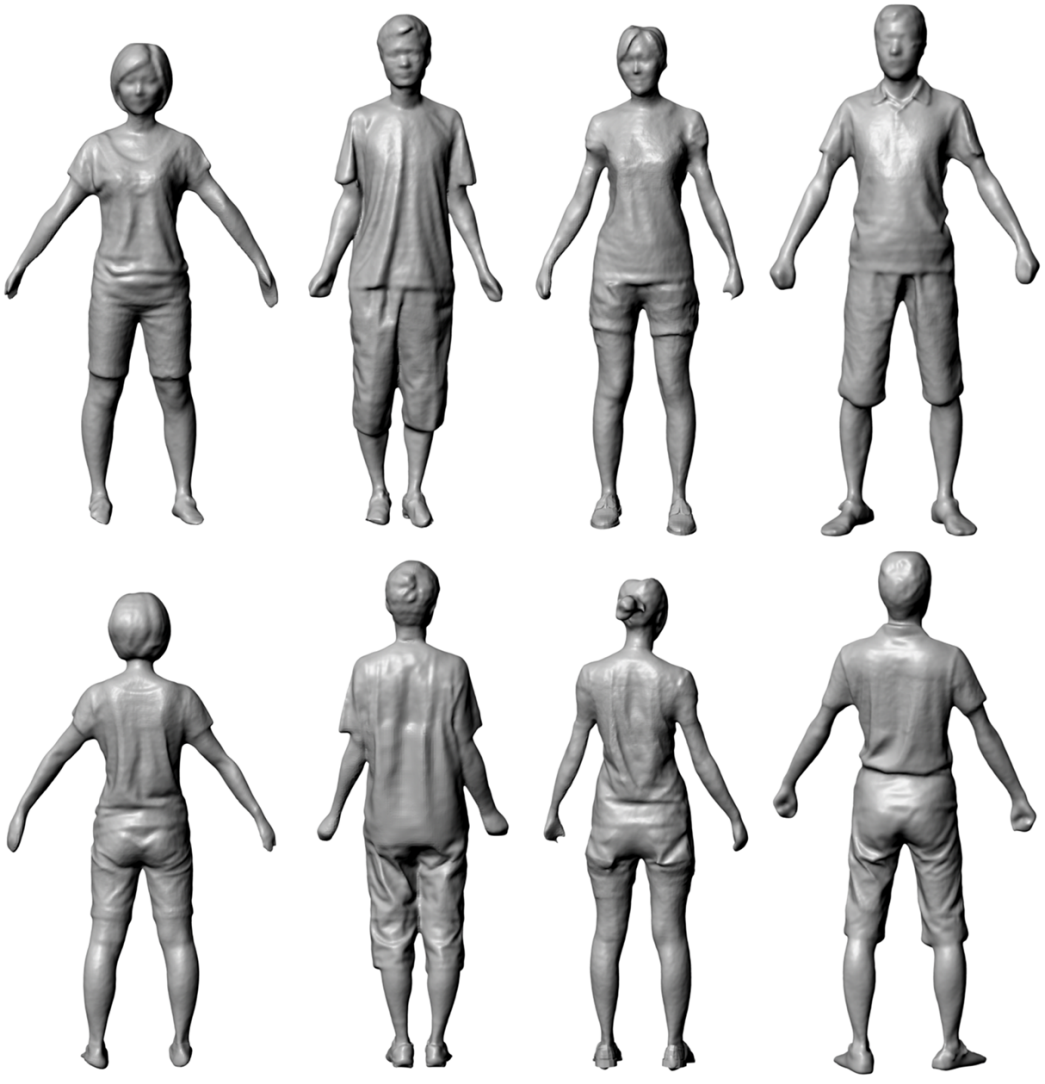

Fig. 10. Different 3D full human models generated using our system. Note that the geometric details such as faces, dresses and hairstyles are well captured in the 3D models.

Table 1. Average distance and computing time

Error of biometric measures (cm)	Neck to Hip Distance	Shoulder Width	Arm Length
	2.5	1.5	3.0
	Leg Length	Waist Girth	Hip Girth
	2.1	6.2	3.8
Computing time (min)	Data preprocessing	Registration	Reconstruction
	1.6	3.8	0.5

tation for human shapes and thus can only capture nearly-naked human bodies, our system can capture more personalized detailed shapes, such as asymmetric stomach, faces, clothes and even hairstyles. Moreover, it takes only about 6 minutes to reconstruct a human model in our method, while it takes approximately 65 minutes to reconstruct a human model by the method of [10].

5.1 Applications

As human models can be acquired quickly in our system, many applications in virtual reality can be developed as well. We have developed two applications in our system as follows. Please also see the accompanying video.

Virtual try on. Online shopping has grown exponentially during the past decade. More and more customers now turn to purchase their dresses online. A major problem related to the online shopping is that the customer can not try out a garment before purchasing. Though virtual try on applications are emerging on the web or in the department store, most of the existing methods are quite simplistic [35], where the body shape is expressed as 2D images or a few biometric measures. With the help of our low-price system, users can have their 3D body shapes constructed easily at home. The 3D garment model can be first roughly aliened using Laplacian surface editing [27] and further mapped to the body shape using physically-based cloth simulation [36], as shown in Figure 11.

Personalized avatar. The skeleton and skin weights of the reconstructed body mesh can be automatically extracted [37]. The model can be animated using motion capture data or online captured skeleton (Figure 12). The user's own personalized avatar will benefit many applications in video games, online shopping, human-computer interaction, etc., and provides favorable experience on VR applications.

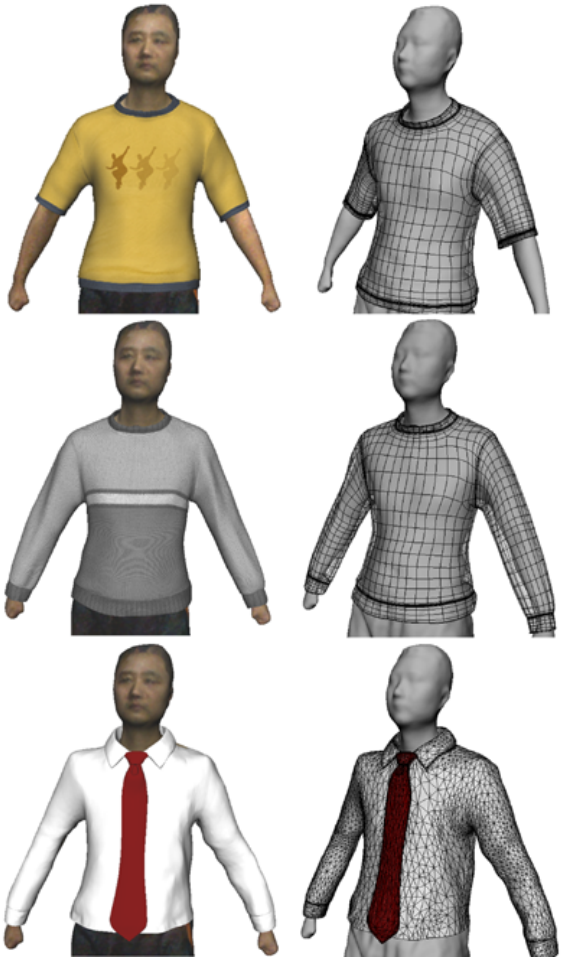

Fig. 11. Realistic virtual try on experience based on the reconstructed model. (Left) the try on results; (right) the corresponding meshes.

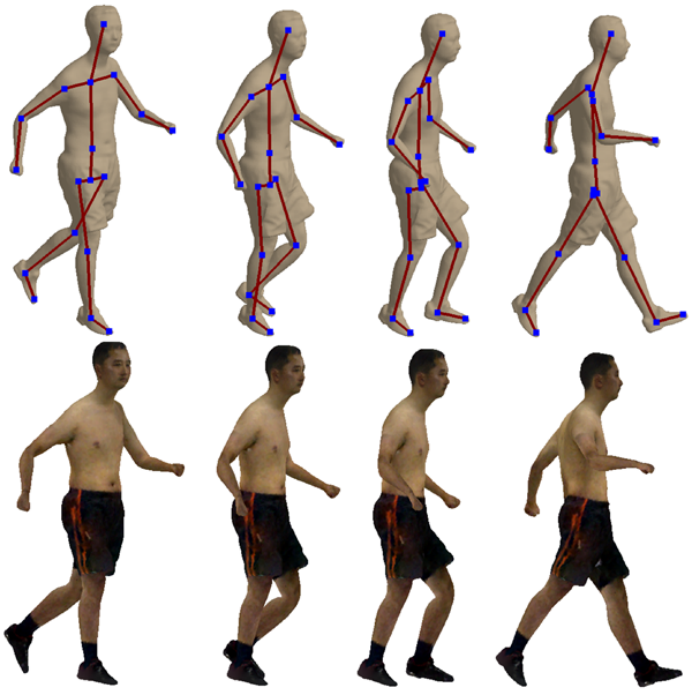

Fig. 12. Personalized avatar generated by our system. The motion of the human body is driven by a given skeleton motion sequence.

6 CONCLUSION

In this paper, a new system using three Microsoft Kinects is presented to scan 3D full human bodies easily. The proposed method can deal with non-rigid alignment with loop closure constraint and complex occlusions. A two-stage registration algorithm performs pairwise deformation on the geometry field firstly, then global alignment on the deformation field is adopted. Our algorithm is efficient and of memory efficiency. Our system can generate convincing 3D human bodies at a much low price and has good potential for home-oriented applications for everyday users.

The quality of the reconstructed models in our system is still poor for some specific applications due to low quality of depth data captured by Kinects. In the future, we plan to investigate more sophisticate de-noising and super-resolution approaches to improve the depth quality [6, 38] , as well as synthesizing fine-scale details in the resulting model [16]. We also plan to compare with results obtained with high precision scanning systems for a better evaluation. Though the problem of complex occlusions can be reasonably handled using global registration method, misalignments still occurred in our experiment. This will cause some unnatural bending in the arm areas, as shown in Figure 10. We will try to improve our registration algorithm to deal with this problem. It is also worthwhile to facilitating our system for more virtual reality applications.

ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers for their constructive comments. We thank Lei Yu, Jing Wu, Jingliang Wu, Xiaoyong Sheng, Xiaoguang Han, and Gang Liu for their help on the experiments. This work was supported jointly by the National Natural Science Foundation of China (61170318, 61070071, 60970076, 61173124), Doctoral Science Foundation of Hohai University(XZX/09B005-05) and Microsoft Research Asia.

REFERENCES

[1] Brett Allen, Brian Curless, and Zoran Popovic. The space of human body shapes: reconstruction and parameterization from range scans. *ACM Transactions on Graphics*, 22(3):587–594, 2003.

[2] Edilson de Aguiar, Carsten Stoll, Christian Theobalt, Naveed Ahmed, Hans-Peter Seidel, and Sebastian Thrun. Performance capture from sparse multi-view video. In *ACM SIGGRAPH 2008 papers*, SIGGRAPH '08, pages 98:1–98:10, New York, NY, USA, 2008. ACM.

[3] Young Min Kim, C. Theobalt, J. Diebel, J. Kosecka, B. Miscusik, and S. Thrun. Multi-view image and tof sensor fusion for dense 3d reconstruction. In *IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 1542–1549, 2009.

[4] Microsoft kinect. `http://www.xbox.com/kinect`, 2010.

[5] Peter Henry, Michael Krainin, Evan Herbst, Xiaofeng Ren, and Dieter Fox. Rgb-d mapping : Using depth cameras for dense 3d modeling of indoor environments. In *International Symposium on Experimental Robotics (ISER)*, 2010.

[6] Yan Cui and Didier Stricker. 3d shape scanning with a kinect. In *ACM SIGGRAPH 2011 Posters*, pages 57:1–57:1, New York, NY, USA, 2011. ACM.

[7] Richard A. Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J. Davison, Pushmeet Kohli, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*, pages 127–136, oct. 2011.

[8] Michael Zollhofer, Michael Martinek, Gnther Greiner, Marc Stamminger, and Jochen Sbmuth. Automatic reconstruction of personalized avatars from 3d face scans. *Computer Animation and Virtual Worlds*, 22(2-3):195–202, 2011.

[9] Thibaut Weise, Sofien Bouaziz, Hao Li, and Mark Pauly. Realtime performance-based facial animation. In *ACM SIGGRAPH 2011 papers*, SIGGRAPH '11, pages 77:1–77:10, New York, NY, USA, 2011. ACM.

[10] Alexander Weiss, David Hirshberg, and Michael J. Black. Home 3d body scans from noisy image and range data. In *13th International Conference on Computer Vision*, 2011.

[11] Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis. Scape: shape completion and animation of people. In *ACM SIGGRAPH 2005 Papers*, SIGGRAPH '05, pages 408–416, New York, NY, USA, 2005. ACM.

[12] Andreas Kolb, Erhardt Barth, Reinhard Koch, and Rasmus Larsen. Time-of-flight sensors in computer graphics. In M. Pauly and G. Greiner, editors, *Eurographics 2009 - State of the Art Reports*, pages 119–134. Eurographics, 2009.

[13] Cyberware. `http://www.cyberware.com/pricing/domesticPriceList.html`, 1999.

[14] Niloy J. Mitra, Simon Flöry, Maks Ovsjanikov, Natasha Gelfand, Leonidas Guibas, and Helmut Pottmann. Dynamic geometry registration. In *Proceedings of the fifth Eurographics symposium on Geometry processing*, pages 173–182, Aire-la-Ville, Switzerland, Switzerland, 2007. Eurographics Association.

[15] Qi-Xing Huang, Bart Adams, Martin Wicke, and Leonidas J. Guibas. Non-rigid registration under isometric deformations. *Computer Graphics Forum*, 27(5):1449–1457, 2008.

[16] Hao Li, Bart Adams, Leonidas J. Guibas, and Mark Pauly. Robust single-view geometry and motion reconstruction. In *ACM SIGGRAPH Asia 2009 papers*, SIGGRAPH Asia '09, pages 175:1–175:10, New York, NY, USA, 2009. ACM.

[17] Robert W. Sumner, Johannes Schmid, and Mark Pauly. Embedded deformation for shape manipulation. In *ACM SIGGRAPH 2007 papers*, SIGGRAPH '07, New York, NY, USA, 2007. ACM.

[18] Yuri Pekelny and Craig Gotsman. Articulated object reconstruction and markerless motion capture from depth video. *Computer Graphics Forum*, 27(2):399–408, 2008.

[19] Will Chang and Matthias Zwicker. Global registration of dynamic range scans for articulated model reconstruction. *ACM Transactions on Graphics*, 30:26:1–26:15, May 2011.

[20] E. Baffle, C. Matabosch, and J. Salvi. Overview of 3d registration techniques including loop minimization for the complete acquisition of large manufactured parts and complex environments. In *Eight International Conference on Quality Control by Artificial Vision*, volume 6356 of *Proceedings Of The Society Of Photo-Optical Instrumentation Engineers (SPIE)*, page 35605. SPIE-INT SOC OPTICAL ENGINEERING, 2007.

[21] Szymon Rusinkiewicz, Benedict Brown, and Michael Kazhdan. 3d scan matching and registration. `http://www.cs.princeton.edu/~bjbrown/iccv05_course/`, 2005. ICCV 2005 Short Course.

[22] Paul J. Besl and Neil D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:239–256, February 1992.

[23] T. Masuda, K. Sakaue, and N. Yokoya. Registration and integration of multiple range images for 3-d model construction. In *Proceedings of the 1996 International Conference on Pattern Recognition (ICPR '96)*, ICPR '96, pages 879–883, Washington, DC, USA, 1996. IEEE Computer Society.

[24] F. Lu and E. Milios. Globally consistent range scan alignment for environment mapping. *Auton. Robots*, 4:333–349, October 1997.

[25] Gregory C. Sharp, Sang W. Lee, and David K. Wehe. Multiview registration of 3d scenes by minimizing error between coordinate frames. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:1037–1050, 2004.

[26] Liao Miao, Zhang Qing, Wang Huamin, Yang Ruigang, and Gong Minglun. Modeling deformable objects from a single depth camera. In *IEEE 12th International Conference on Computer Vision*, pages 167–174, 2009.

[27] O. Sorkine, D. Cohen-Or, Y. Lipman, M. Alexa, C. Rossl, and H.-P. Seidel. Laplacian surface editing. *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, 71:175–184, 2004.

[28] Openni. `http://www.openni.org/`, 2011.

[29] Multi-camera self-calibration. `http://cmp.felk.cvut.cz/~svoboda/SelfCal/`, 2003.

[30] Michael Kazhdan, Matthew Bolitho, and Hugues Hoppe. Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing*, SGP '06, pages 61–70, Aire-la-Ville, Switzerland, Switzerland, 2006. Eurographics Association.

[31] Nils Hasler, Carsten Stoll, Bodo Rosenhahn, Thorsten Thormahlen, and Hans-Peter Seidel. Estimating body shape of dressed humans. *Computers & Graphics*, 33(3):211–216, 2009. IEEE International Conference on Shape Modelling and Applications 2009.

[32] Open source computer vision library (opencv). `http://opencv.willowgarage.com/wiki/`.

[33] Andriy Myronenko, Xubo Song, and . Carreira-Perpinn. Non-rigid point set registration: Coherent point drift. In *Advances in Neural Information Processing Systems 19 (2007)*, volume 19, pages 1009–1016. MIT Press, 2007.

[34] Ming Chuang, Linjie Luo, Benedict J. Brown, Szymon Rusinkiewicz, and Michael Kazhdan. Estimating the laplace-beltrami operator by restricting 3d functions. In *Proceedings of the Symposium on Geometry Processing*, SGP '09, pages 1475–1484, Aire-la-Ville, Switzerland, Switzerland, 2009. Eurographics Association.

[35] Nadia Magnenat-Thalmann, Etienne Lyard, Mustafa Kasap, and Pascal Volino. Adaptive body, motion and cloth. In *Motion in Games*, volume 5277 of *Lecture Notes in Computer Science*, pages 63–71. Springer Berlin / Heidelberg, 2008.

[36] David Baraff and Andrew Witkin. Large steps in cloth simulation. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '98, pages 43–54, New York, NY, USA, 1998. ACM.

[37] Ilya Baran and Jovan Popović. Automatic rigging and animation of 3d characters. In *ACM SIGGRAPH 2007 papers*, SIGGRAPH '07, New York, NY, USA, 2007. ACM.

[38] Jiejie Zhu, Liang Wang, Jizhou Gao, and Ruigang Yang. Spatial-temporal fusion for high accuracy depth maps using dynamic mrfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32:899–909, 2010.