Neurocomputing 151 (2015) 626–631

Contents lists available at ScienceDirect

Neurocomputing

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

Estimation of human body shape and cloth field in front of a kinect

Ming Zeng [a], Liujuan Cao [b,*], Huailin Dong [a], Kunhui Lin [a], Meihong Wang [a], Jing Tong [c]

[a] Software School of Xiamen University, Xiamen, China
[b] Department of Computer Science, Xiamen University, Xiamen, China
[c] College of IOT Engineering, Hohai University, Changzhou, China

ARTICLE INFO

Article history:
Received 13 November 2013
Received in revised form
6 June 2014
Accepted 7 June 2014
Available online 11 November 2014

Keywords:
RGBD data
Non-rigid deformation
Human body estimation
Cloth field

ABSTRACT

This paper describes an easy-to-use system to estimate the shape of a human body and his/her clothes. The system uses a Kinect to capture the human's RGB and depth information from different views. Using the depth data, a non-rigid deformation method is devised to compensate motions between different views, thus to align and complete the dressed shape. Given the reconstructed dressed shape, the skin regions are recognized by a skin classifier from the RGB images, and these skin regions are taken as a tight constraints for the body estimation. Subsequently, the body shape is estimated from the skin regions of the dressed shape by leveraging a statistical model of human body. After the body estimation, the body shape is non-rigidly deformed to fit the dressed shape, so as to extract the cloth field of the dressed shape. We demonstrate our system and the therein algorithms by several experiments. The results show the effectiveness of the proposed method.

1. Introduction

Estimation of human body is an important topic in computer graphics and computer vision. It has wide applications such as virtual try-on [1], shape reconstruction [2], shape based image editing [3], to name a few. Since it plays a central role in such wide applications, the human body estimation has been a hot topic in research communities for recent years.

To obtain the model of human bodies, some works acquire the color or depth data of naked human bodies (usually in tight clothes) and then reconstruct the shapes from the acquired data, such as [4,2]. However, it is not convenient to require users to show their bare body in front of the sensor. To alleviate this, some researchers seek to estimate the hidden body under the dressed shape. For example, Balan et al. [5] and Hasler et al. [6] use an images set to estimate the human body, and Hasler et al. [7] explore the human body from a dressed mesh, which provides much more geometry constraints than images for the estimation. As the commodity RGBD sensors, say Microsoft Kinect [8], begin to be prevalent, many applications need an easy-to-use system to estimate the human body based on this kind of sensors. For instance, the virtual try-on systems usually require to estimate the shape of the user's body, so as to "wear" clothes for the user. To this end, we aim at devising an system based on RGBD data to estimate the human body conveniently.

In our system, we first reconstruct the full dressed shape (with clothes). The dressed model provides much strong geometry constraints for body estimation than the single view geometry. Then the skin regions are recognized from color images and the corresponding mesh regions of the exposed body are used as a tight bound of the bare body. Given the dressed shape and the skin regions, we estimate the naked shape in a subspace of the human body. At the end, using the estimate naked body and the dressed shape, the system extracts the cloth field by comparing their corresponding vertices. The cloth field can be used to build cloth database for further research.

In summary, this paper makes a systematic contribution which integrates two novel algorithms. It introduces an easy-to-use pipeline on Kinect to estimate 3D human bodies. The first ingredient algorithm is an easy-to-operate method to reconstruct human shape (with clothes) using a Kinect, and the second algorithm is an deformation based method to extract cloth field of the human.

2. Related work

Shape reconstruction: To build the 3D model of a human, different views of the human should be captured. Image based methods reconstruct the shape from images in multiple views. These images are obtained from cameras around the human, say the light stage [9]. Other methods capture the depth map (i.e. the

* Corresponding author.
E-mail addresses: mingzeng85@gmail.com (M. Zeng), caoliujuan@gmail.com (L. Cao).

http://dx.doi.org/10.1016/j.neucom.2014.06.087
0925-2312/

partial mesh) of the human, and align these partial data together. KinectFusion [10] and it variants [11] integrate and reconstruct the shape as the Kinect moves around the object, but they do not consider the deformable shapes. To reconstruct the deformable model, Chang and Zwicker [12] proposed a reduced deformable model to account for the shape deformation. Tong et al. [1] leverage a statistical model to estimate the human motion. Recently, Zeng et al. [13] proposed a non-rigid deformation method under the as-rigid-as-possible assumption. Li et al. also proposed a non-rigid modeling system which further considers the texture consistency. This paper follows the path of Tong [1], which leverages a statistical model of human body to estimate the slight human motion, and then completes the whole dressed human shape.

Naked body shape estimation: Generally speaking, the naked body estimation needs assistance from a statistical model of human bodies, which provide sufficient shape priors for the estimation. Image based methods [6,14] take the silhouette of the human shape as the input information. For example, Balan and Black [5] estimate the 3D body shape of dressed person from silhouettes of multi-view images, combining constraints of different poses to recover the body. However, the silhouette is weak to provide enough geometry information. Mesh based methods (e.g. [7]) directly use the whole geometry mesh (with cloth) as the input, and estimate the naked shape under the cloth. These kind of methods are more robust than the image based methods, but usually need more complex acquisition setups.

3. Our method

3.1. System overview

The system requires the user to stand in front of a Kinect. The Kinect captures the RGB and depth data of the user. At the acquisition step, the system shows a human body with a standard pose on the screen and leave 10 s to allow the user to lay out the same pose with the displayed model. Then the user turns 90°, 180°, 270° in front of the Kinect to be captured from the back view and two side views. To alleviate the shape registration in the following steps, the user is required to keep the standard pose as same as possible. After the data acquisition, we adopt a non-rigid shape registration to register these four frames of rgbd data in a common coordinate. Since the RGBD data of side views only provides the "thickness" information of the body, after being used to align the frontal and back views, the side-view data is no longer needed, so we drop them in the following steps. Given the data from the frontal and the back view, we first utilize a skin detection and segmentation algorithm on the RGB image to pick out the skin region. The skin region serves as a tight constraint for the body estimation since it is not covered by clothes.

Given this RGBD data, the initial pose, and the skin constraint, we estimate the shape and pose parameters of a statistical human model (SCAPE [15]), which results in an estimated mesh $\mathbf{X}$ of the user's body. The statistical model guarantees the estimation lays in a plausible subspace of the human body. To account for the clothes, we take a non-rigid deformation scheme to deform the estimated mesh $\mathbf{X}$ to fit the captured depth data, leading to a dressed mesh $\mathbf{X}'$. At the final step, we subtract $\mathbf{X}$ from $\mathbf{X}'$ to obtain the vector field of the cloth $\mathbf{C}=\mathbf{X}'-\mathbf{X}$ which represents the amount of the dressed shape out-stemming from the naked shape.

3.2. Statistical model of human body

This section reviews the 3D full-body morphable model, which is the prerequisite of our method. A 3D full-body morphable model is a kind of 3D human shape controlled by sets of parameters. In our method, we adopt the SCAPE model [15] as our morphable model due to its simplicity. The SCAPE model determines a human shape by two sets of parameters: shape θ and pose β, and it is denoted by $\mathcal{S}(\theta,\beta)$. The shape parameters θ control the shape variations across different individuals, while the pose parameters β specify the shape deformation caused by changing pose. More specifically, the SCAPE model allows us to generate an individual body shape by giving θ, and with a pose by giving β.

SCAPE model should be learned from a database of human shape with different individuals and different poses. We follow Zhou et al. [3] to learn it from a public database [16]. In our case, $\theta\in R^{10}$ and $\beta\in R^{20}$, which cover well the human subspace spanned by the training data. We refer readers to [15,3] for more details about the definition and training of the SCAPE model.

3.3. Shape modeling

In this section, we present how to utilize the SCAPE model to reconstruct a human shape from depth data of four different views. In this stage, depth sensors capture scans of a human turning round before the sensors. During the capture, the human is asked to roughly keep a standard pose. Since the human need to turn round by himself, it is impossible to keep still. These inevitable pose differences between scans can be compensated by our algorithm.

Shape posing in subspace: As mentioned, for these depth data, we need to estimate a shape parameters θ and a pose of each scan, i.e. a global rigid transformation (R^i,t^i) and the local pose parameters β^i.

In particular, in the first scan D^1, we estimate the shape parameters θ and β^1 at the same time, and in the following scans, we fix the estimated $\theta^\star$ and only estimate β^i. For this task, we adopt a similar method to shape completion [15]. We optimize θ and β to minimize the marker point distance E_m to require the estimated shape match D^i:

$$E_m=\sum_{j\in marker}\|R_i\cdot\mathcal{S}(\theta,\beta)_j+t_i-D^i_j\|^2 \tag{1}$$

To minimize this objective function, an iterative fashion is used to optimize (R^i,t^i) and (θ,β) in turn. For the marker points, in the first scan, they can be initially chosen as joint locations from automatic skeleton detection [17]. An iterative closest point scheme is utilized to gradually add more marker points. For following scans, we take the previous result as initial value, and build the marker point correspondences by nearest neighbor searching.

After this step, we obtain the estimated $\theta^\star$ and $(R^{i\star},t^{i\star})$, $\beta^{i\star}$.

Non-rigid registration to SCAPE: Given the estimated rigid transformation $(R^{i\star},t^{i\star})$ and human shape/pose $\mathcal{S}(\theta^\star,\beta^{i\star})$, together with the dense correspondence between estimated shape and scanned depth, we are ready to warp D_i to the data captured in the first frame.

Firstly, rigidly transform from D_i to $\overline{D_i}$ is performed by $T^{-1}(R^{i\star},t^{i\star})$, and then D^i is non-rigidly warped to D^i according to the warping field $\zeta^i:\mathcal{R}^3\to\mathcal{R}^3$. The warping field is defined by locally rigid transformation $\phi(R^i_j,t^i_j)$ of all vertices on the SCAPE model, and the $\phi(R^i_j,t^i_j)$ is calculated by normal and position of the jth vertex of $\mathcal{S}(\theta^\star,\beta^{i\star})$ and $\mathcal{S}(\theta^\star,\beta^{1\star})$. Here, we follow embedded deformation [18] to define the warping field ζ^i.

After warping all scans, we re-estimate the θ and β^1 according to the warped scans set: $\tilde{D}=\{\tilde{D}^i,i=1,2,\ldots,\#\text{ of scans}\}$, note that $\tilde{D}^1=D^1$. We minimize Eq. (1) again, but this time we find the nearest point in the scan set $\tilde{D}$ instead of a specified scan D^i. The optimal value is denoted as $\tilde{\theta}$ and $\tilde{\beta}$, and the optimal value determines the subspace shape $S_s=\mathcal{S}(\tilde{\theta},\tilde{\beta})$.

So far, we have non-rigidly registered the four scans into a common coordinate and the same pose.

3.4. Body estimation based on skin segmentation

The previous estimation of SCAPE is used to assist shape modeling from depth scans with different poses. However, the estimation is not the real shape of the naked body since it accounts for the clothes covered on the body. Given the modeled dressed shape, only tight constraints can be utilized—the skin regions. Therefore, we identify the skin region and impose tight constraints on these parts to re-estimate the parameters of the SCAPE. We take a Bayesian classifier to recognize the skin color [19]. Concretely, the color space is chosen to be YUV to better classify skin and non-skin color. The illumination component Y is dropped and only UV components are used. According to the Bayes rule, the skin classification is formulated as

$$P(s|c) = P(c|s)P(s)/P(c) \tag{2}$$

where $P(c)$ denotes the occurrence probability of a color c in the training set, $P(s)$ the prior probability of skin color in the training set, $P(c|s)$ the prior probability of a color c being a skin color. All these are trained from a set of images with human skin labeled manually.

When this classifier is used, each pixel is assigned a poster probability according to Eq. (3). With this probability, the pixels are classified into strong-skin ($>T_{max}$), weak-skin ($>T_{min}$), or non-skin ($<T_{min}$). The weak-skin pixel can be seen as a skin color if there is any strong-skin pixel neighboring to it. After the classification, a flood-in post-processing step is employed to fill holes on the skin regions.

After the skin segmentation, each vertex in the dressed mesh is labeled to skin vertex or non-skin vertex. For the skin vertices, it provide tight constraints for the SCAPE estimation. We re-formulate Eq. (1) as

$$E_{skin} = \sum_{j \in skin\ vertices} \|R_i \cdot \mathcal{S}(\theta,\beta)_j + t_i - D_j^i\|^2 \tag{3}$$

which requires the SCAPE model to fit the skin regions well, and we adopts the closest point scheme for the correspondence searching.

3.5. Cloth field estimation

Because the subspace shape S_s is a naked human shape, to generate dressing details, we need to deform S_s to fit the warped scans set $\tilde{D}$. We first subdivide S_s to present much more clothing features. Then we deform the subdivided S_s to fit $\tilde{D}$ by solving the following optimization problem:

$$\begin{aligned} \arg\min_{T_1+d_1 \ldots T_{|T|}+d_{|T|}} \quad & E_c + w_s \cdot E_s + w_I \cdot E_I \\ \text{s.t.} \quad & T_i v_k + d_i = T_j v_k + d_j,\ v_k \in vt(Tri_i \cap Tri_j). \end{aligned} \tag{4}$$

where, the parameters T_i and d_i are 3×3 affine transformation and 3×1 translation for ith triangle, respectively. Following the derivation in [20], T_i can be represented by original (v_1, v_2, v_3) and deformed $(\tilde{v}_1, \tilde{v}_2, \tilde{v}_3)$ positions of the triangle's vertices: $T_i = [v_2 - v_1\ v_3 - v_1\ v_4 - v_1]^{-1}[\tilde{v}_2 - \tilde{v}_1\ \tilde{v}_3 - \tilde{v}_1\ \tilde{v}_4 - \tilde{v}_1]$.

In this objective function, the correspondence term $E_c = \sum_{1\ldots|C|}\|v_i - v_i^\star\|^2$ requires that the deformed mesh fit $\tilde{D}$ regarding to correspondences $(v_i, v_i^\star)$. The smooth term $E_s = \sum_{i=1\ldots|T|}\sum_{j\in adj(i)}\|T_i - T_j\|_F^2$ ensures neighboring triangles with similar transformation. And the third term $E_I = \sum_{i=1\ldots|T|}\|T_i - I\|_F^2$ makes the mesh prefer less deformation.

The constraints in the optimization problem require that the shared vertex by two nearby triangles yield a same position under the two corresponding transformations, which intuitively means that the deformed mesh will not be split.

To solve the optimization problem, we adopt the non-rigid ICP scheme [13]. Specifically, we iteratively re-establish the valid closest correspondences and solve the therein objective function. Given the point correspondences, this optimization problem can be re-written into a vertex formulation (refer to [20]), and formulated into a linear system. For each iteration, we take a relaxed weighting strategy to determine the weights of energy terms. At the first iteration, we use $w_s = 1.0$, $w_I = 0.001$, and $w_c = 1.0$. As the iteration proceeds, w_c gradually increases with the speed $w_c^{new} = 2 \times w_c^{old}$ until $w_c \geq 100$. In our experiments, the procedure converges in less than 50 iterations.

The optimal $T_i^\star$ and $d_i^\star$ deform the subspace shape S_s to the clothed detailed shape S_d. After deforming the mesh in the SCAPE space into the dressed shape, we are able to obtain the cloth field by computing the differences between the SCAPE model and dressed shape.

4. Experiments

We conduct experiments to demonstrate the proposed method. A person dressed in a heavy coat is captured by a Kinect. The body is segmented from the background simply by a depth-value threshold. Fig. 1(a) shows the captured depth data (each vertex has color) of the frontal view of the person. Fig. 1(b) shows the result of the skin detection. Fig. 1(c–e) is registered shapes, which are seen from different viewpoints.

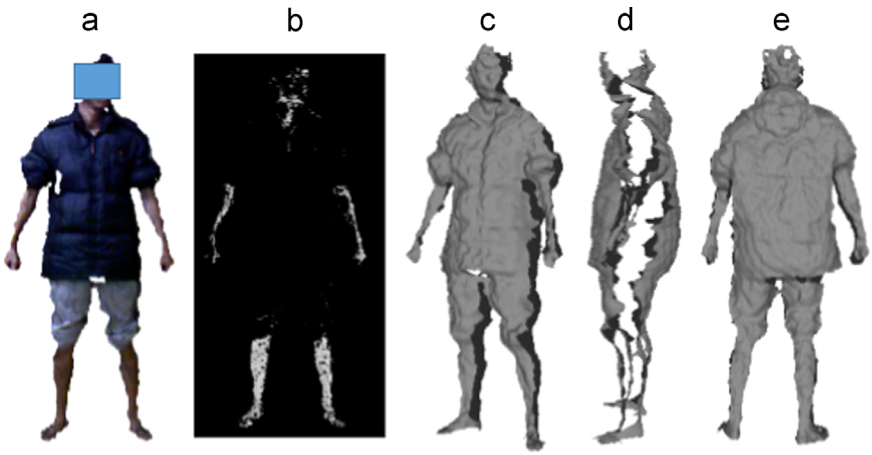


Fig. 1. The steps of skin detection and multiple-view registration. (a) The input depth data (with per-vertex color). (b) The detected color map, the white pixels indicate skin regions while black pixels indicate non-skin regions. (c–e) The registered geometry of frontal and back views (seen from different viewpoints). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

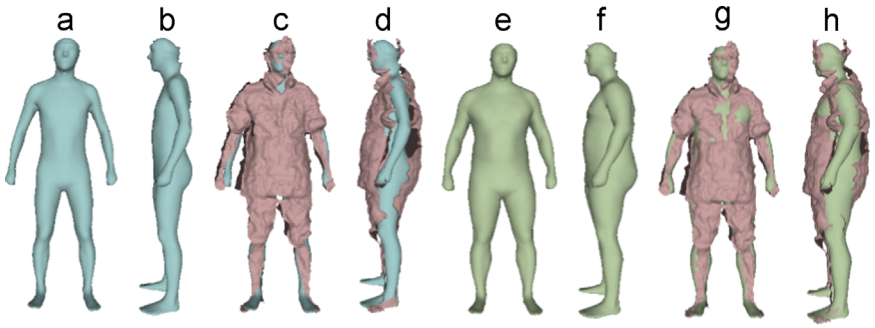


Fig. 2. The comparison between shape estimation with/without skin detection. (a–d) Results of our proposed method. (e–h) Results without skin detection.

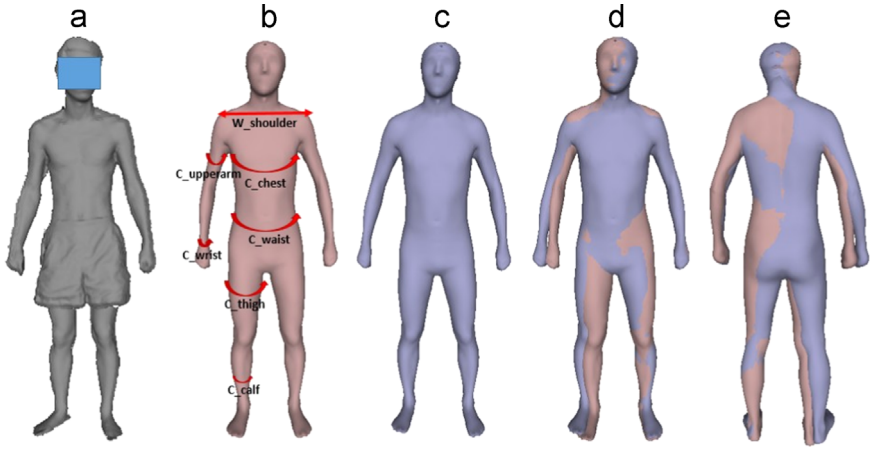


Fig. 3. The comparison between shape estimation with/without skin detection. (a) The scanned model. (b) The shape estimation from the model in (a). (c) The shape estimation from the same person but with cloth (Fig. 1(a)). (d and e) Two views of these two estimation results, and the two results are put together for ease of comparison.

Table 1
Shape parameters of bodies (unit: m).

Measurements	$W_{shoulder}$	$C_{upperarm}$	C_{wrist}	C_{chest}	C_{waist}	C_{thigh}	C_{calf}
Ground Truth	0.383	0.244	0.146	0.852	0.711	0.413	0.310
Dressed Est.	0.391	0.239	0.157	0.873	0.704	0.401	0.298

Comparison of with/without skin constraints: We compare the body shapes which are estimated with/without skin detection. With the skin detection, the non-skin regions do not influence the shape estimation, and the estimated shape is more reasonable. Fig. 2(a and b) (frontal and side views) is the estimated results only using constraints of skin-regions, where we see that it is consistent with the body shape of the person (Fig. 1(a)). From Fig. 2(c and d) it can be observed that the captured data almost covers the estimated shape, even leaving a substantial space on the clothed regions. Obviously, these space are the thickness of the clothes. In contrast, the estimated result without excluding non-skin regions is apt to account for the clothes as one part of the body. Therefore, estimated shape shown in Fig. 2(e and f) is much fatter than it should be (compared with Fig. 1(a) and Fig. 2(a and b)). We also see that it fits the captured data much more closely (Fig. 2(g and h)) than its counterpart (Fig. 2(c and d)). It is worth mentioning that the method to estimate body shape without skin detection used in the comparison is similar to that of [7] in spirit, both of them estimate the naked body shape in the SCAPE space without making a distinction between skin and non-skin regions, inevitably leading to overestimation of the body shape.

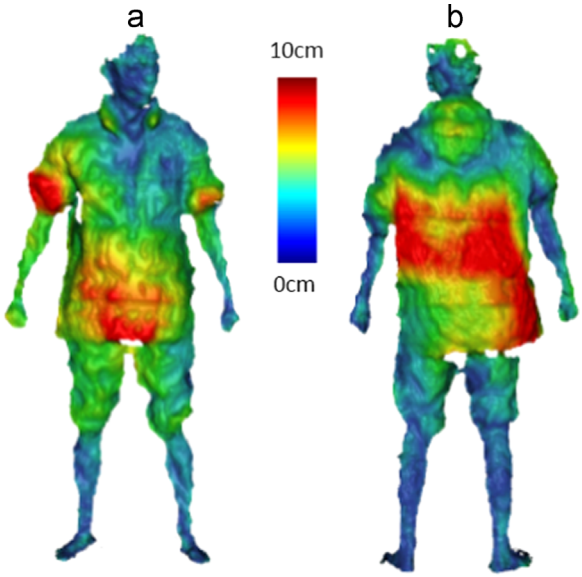


Fig. 4. The visualization of cloth field estimation. (a) The frontal view. (b) The back view. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

Comparison of estimation and ground truth: To validate the effectiveness of the proposed method, we compare our result with ground truth. We scan a naked person using KinectFusion [10] (Fig. 3(a)). Then we estimate the body in SCAPE space

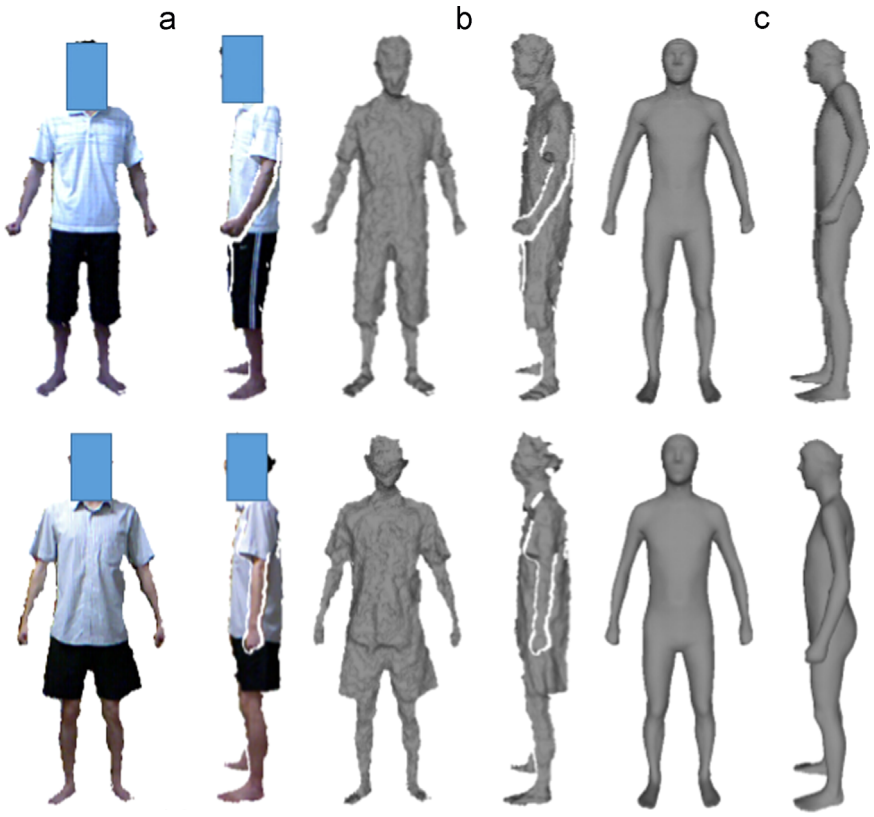


Fig. 5. More results. (a) The input RGB data. (b) The input depth data. (c) The estimation results. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

(Fig. 3(b)) from this naked model. For comparison, we use our method to estimate the naked body of the same person but with clothes (Fig. 3(c)). Fig. 3(d) shows the estimation result. We put these two models together, and it can be observed that these two results are very similar (Fig. 3(e) and (f)). To quantitatively compare these two results, we also measure some shape parameters (Fig. 3(b)) for these two models, respectively. These parameters include the width of the shoulder, the circumferences of upper arms, wrists, chests, waists, thighs, and calves. These measurements are listed in Table 1. From this table it can be found that the two bodies are very close in numerics.

Cloth field estimation: The cloth field extracted from the captured person is shown in Fig. 4. The cloth field is visualized according to deformation amount from the naked body. The heavier regions are specified by a warmer color, while the thinner regions are indicated by a cooler color.

More results: In this section, we show two more results. As in Fig. 5, each row shows results of an individual. For each row, column (a) is input RGB information, column (b) is input depth data, and column (c) is the body estimation using our method. In these two examples, the estimations are consistent with the body shape as seen from input data.

5. Conclusion and future work

In this paper we present an integrated system to estimate the human body using a single Kinect. The system captures and reconstructs the dressed human shape in a convenient way, and estimates the body in the subspace of the human body utilizing the shape constraints on the skin regions. The proposed system provides a simple yet practical solution to recover the human body, which is useful to the potential virtual try on application. We also extract the cloth field from the dressed shape and the body shape, which gives a feasible method to collect cloth data, and makes it possible to analyze properties of the clothes. Our experimental results show the feasibility and effectiveness of our system.

There are still limitations in our system to be overcome in the future work. First, the current shape registration algorithm will fail when the deformation is large. A more robust way to this problem is to analyze the similarities of different views of shapes [21–23]. Second, we will try to design a more sophisticated method combining color and geometry information to improve the skin-region classifier's accuracy. Third, the current cloth extraction will fail when the user is in some complex clothes, since the topology of the body shape may be different from the dressed shape. This is still an open problem which needs further investigation.

Besides, estimating body shape from image is another promising research field. One avenue is reconstructing the shape from the self-captured multi-view images. A more challenging and interesting avenue is to estimate body from a single image. Although this is an under-constrained problem, there are several work trying to resolve this via introducing priors, e.g. [24,25]. An insight is to explore similar body images by searching from internet (might directly use methods or borrow ideas from image retrieval, e.g. [26–32]), thus to enrich the constraints for the body estimation.

Acknowledgments

We would like to thank the reviewers for their valuable comments. This work was partially supported by the following research funds: NSFC (Nos. 61402387, 61402388, and 61202284); the Key Technology R&D Program of Xiamen, Fujian (Nos. 3502Z20103001 and 3502Z20101002); the Leading Academic Discipline Program, "Project 211 (the 3rd phase)" of Xiamen University; the Fundamental

Research Funds for the Central Universities, Xiamen University (Nos. 2011121023, CXB2012012, CXB2012013, 201212G007 and CXB2013016); Shenzhen Key Laboratory for High Performance Data Mining with Shenzhen New Industry Development Fund (No. CXB201005250021); the Soft Science Research of Fujian Province of China (No. 2014R0090).

References

[1] J. Tong, J. Zhou, L. Liu, Z. Pan, H. Yan, Scanning 3d full human bodies using kinects, IEEE Trans. Vis. Comput. Graph. 18 (4) (2012) 643–650.
[2] A. Weiss, D. Hirshberg, M.J. Black, Home 3d body scans from noisy image and range data, in: ICCV, 2011, pp. 1951–1958.
[3] S. Zhou, H. Fu, L. Liu, D. Cohen-Or, X. Han, Parametric reshaping of human bodies in images, ACM Trans. Comput. Graph.: Spec. Issue ACM SIGGRAPH 29 (4) (2010).
[4] B. Allen, B. Curless, Z. Popović, Articulated body deformation from range scan data, ACM Trans. Graph. 21 (3) (2002) 612–619.
[5] A.O. Balan, M.J. Black, The naked truth: estimating body shape under clothing, in: ECCV, vol. 2, 2008, pp. 15–29.
[6] N. Hasler, H. Ackermann, B. Rosenhahn, T. Thormählen, H.-P. Seidel, Multilinear pose and body shape estimation of dressed subjects from image sets, in: CVPR, 2010, pp. 1823–1830.
[7] N. Hasler, C. Stoll, B. Rosenhahn, T. Thormählen, H.-P. Seidel, Estimating body shape of dressed humans, Comput. Graph. 33 (3) (2009) 211–216.
[8] Microsoft, ⟨http://www.microsoft.com/⟩.
[9] E. de Aguiar, C. Stoll, C. Theobalt, N. Ahmed, H.-P. Seidel, S. Thrun, Performance capture from sparse multi-view video, ACM Trans. Graph. 27 (3) (2008).
[10] R.A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A.J. Davison, P. Kohli, J. Shotton, S. Hodges, A. Fitzgibbon, Kinectfusion: real-time dense surface mapping and tracking, in: ISMAR'11, 2011, pp. 127–136.
[11] M. Zeng, F. Zhao, J. Zheng, X. Liu, Octree-based fusion for realtime 3d reconstruction, Graph. Models 75 (3) (2013) 126–136.
[12] W. Chang, M. Zwicker, Global registration of dynamic range scans for articulated model reconstruction, ACM Trans. Graph. 30 (3) (2011) 1–15.
[13] M. Zeng, J. Zheng, X. Cheng, X. Liu, Templateless quasi-rigid shape modeling with implicit loop-closure, in: CVPR, 2013, pp. 145–152.
[14] X. Chen, Y. Guo, B. Zhou, Q. Zhao, Deformable model for estimating clothed and naked human shapes from a single image, Vis. Comput. 29 (11) (2013) 1187–1196.
[15] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, J. Davis, Scape: shape completion and animation of people, ACM Trans. Graph. 24 (3) (2005) 408–416.
[16] N. Hasler, C. Stoll, M. Sunkel, B. Rosenhahn, H.-P. Seidel, A statistical model of human pose and body shape, Comput. Graph. Forum 28 (2) (2009) 337–346.
[17] J. Shotton, A.W. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, A. Blake, Real-time human pose recognition in parts from single depth images, in: CVPR, 2011, pp. 1297–1304.
[18] R.W. Sumner, J. Schmid, M. Pauly, Embedded deformation for shape manipulation, ACM Trans. Graph. 26 (3) (2007) 80.
[19] A.A. Argyros, M.I.A. Lourakis, Real-time tracking of multiple skin-colored objects with a possibly moving camera, in: ECCV, vol. 3, 2004, pp. 368–379.
[20] R.W. Sumner, J. Popović, Deformation transfer for triangle meshes, ACM Trans. Graph. 23 (3) (2004) 399–405.
[21] Y. Gao, M. Wang, Z.-J. Zha, Q. Tian, Q. Dai, N. Zhang, Less is more: efficient 3-d object retrieval with query view selection, IEEE Trans. Multimed. 13 (5) (2011) 1007–1018.
[22] Y. Gao, M. Wang, D. Tao, R. Ji, Q. Dai, 3-d object retrieval and recognition with hypergraph analysis, IEEE Trans. Image Process. 21 (9) (2012) 4290–4303.
[23] K. Zhu, Y. Liu, A.G. Aboagye, H. Song, J. Gao, Similarity-based 3-d atmospheric nucleation data visualization and analysis, Tsinghua Sci. Technol. 18 (2) (2013).
[24] C. BenAbdelkader, Y. Yacoob, Statistical body height estimation from a single image, in: FG, 2008, pp. 1–7.
[25] P. Guan, A. Weiss, A.O. Balan, M.J. Black, Estimating human shape and pose from a single image, in: ICCV, 2009, pp. 1381–1388.
[26] R. Datta, D. Joshi, J. Li, J.Z. Wang, Image retrieval: ideas, influences, and trends of the new age, ACM Comput. Surv. 40 (2) (2008).
[27] R. Ji, X. Xie, H. Yao, W.-Y. Ma, Mining city landmarks from blogs by graph modeling, in: ACM Multimedia, 2009, pp. 105–114.
[28] R. Ji, H. Yao, X. Sun, B. Zhong, W. Gao, Towards semantic embedding in visual vocabulary, in: CVPR, 2010, pp. 918–925.
[29] R. Ji, L.-Y. Duan, J. Chen, H. Yao, J. Yuan, Y. Rui, W. Gao, Location discriminative vocabulary coding for mobile landmark search, Int. J. Comput. Vis. 96 (3) (2012) 290–314.
[30] R. Ji, H. Yao, W. Liu, X. Sun, Q. Tian, Task-dependent visual-codebook compression, IEEE Trans. Image Process. 21 (4) (2012) 2282–2293.
[31] R. Ji, L.-Y. Duan, J. Chen, L. Xie, H. Yao, W. Gao, Learning to distribute vocabulary indexing for scalable visual search, IEEE Trans. Multimed. 15 (1) (2013) 153–166.
[32] X. Shen, Z. Lin, J. Brandt, Y. Wu, Detecting and aligning faces by image retrieval, in: CVPR, 2013, pp. 3460–3467.

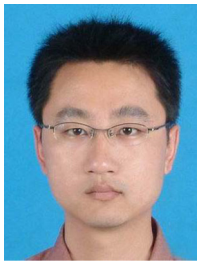
Ming Zeng is currently an assistant professor in the software school of Xiamen University, China. He obtained his Ph.D. degree in computer science from Zhejiang University in 2013. His main research interests include interactive computer vision, geometry acquisition, and 3D reconstruction.

Liujuan Cao is currently an assistant professor at the department of Computer Science, Xiamen University. Before that, she obtained her Ph.D. degree from Harbin Engineering University. Her research is in the field of multimedia analysis, geo-science and remote sensing, and computer vision. She has published extensively at CVPR, Neurocomputing, Signal Processing, ICIP, VCIP, etc.

Huailin Dong is a professor in the software school of Xiamen University, China. His main research interests include digital media processing, intelligent information processing, data mining, and software engineering.

Kunhui Lin is a professor in the software school of Xiamen University, China. His main research interests include multimedia, intelligent information processing, and cloud computing.

Meihong Wang is an assistant professor in the software school of Xiamen University, China. Her main research interests include artificial intelligence, database, and intelligent information processing.



Jing Tong is currently an assistant professor in the school of IOT engineering of Hohai University, China. He obtained his Ph.D. degree of computer science from Zhejiang University in 2013. His main research interests include computer graphics and computer vision.