Fast And Low Cost Automated Human 3D Modeling And Skeleton Generation

Samuel Cahyawiajya[1] and Iping Supriana[2]

**[1] Informatics Engineering, Institut Teknologi Bandung
Bandung, 40116, Indonesia**

**[2] Informatics Engineering, Institut Teknologi Bandung
Bandung, 40116, Indonesia**

Abstract

Virtual reality technology is a very popular research nowadays. One kind of virtual reality is a virtual fitting room where people can try their outfit virtually. Virtual fitting room can be done in many ways. One of it is done by using a 3D model of the user and attach some accessories to the 3D model. In virtual fitting room, we need a human 3D model which is representative and can move like a real human. In this paper, we propose a fast and low cost 3D modeling approach with skeleton embedding for human object by using depth and color image. The proposed solution is divided into 3 main phase, which is image preprocessing phase, model generation phase, and model transformation & rendering phase. Our implementation is done by using two Microsoft Kinect RGB and depth camera with 480x640 image resolution. The output quality of this solution is compared with ground truth

Keywords: *3D model, 3D reconstruction, skeleton, joint, transformation, Microsoft Kinect*

1. Introduction

Virtual fitting room is a technology which enable user to fitting clothes virtually. There are several method has known to create a virtual fitting room, one of those is using a 3D human captured model with capabilities to wear some predefined 3D accessories. In order to create a 3D human model, a 3D scanning and skeleton generation method is needed. The method of 3D scanning and skeleton generation is usually separated into two different task nowadays. So to create a human 3D model with skeleton, we need to generate the 3D model first and to then estimate the skeleton from the generated 3D model, and this task is usually time consuming.

In this paper, we propose a fast and low cost human 3D scanning with skeleton generation method. Our solution is done by using two Microsoft Kinect devices. Our solution is divided into 3 main phase, which is image preprocessing phase, model generation phase, and model transformation & rendering phase.

2. Related Works

(Sturm, Bylow, Kahl, & Cremers, 2013) have implemented upper human body scanning method using a single Microsoft Kinect with the scanned person seating on a swivel chair. The scanning process is done by capturing the person while the person is rotated on swivel chair. The 3D model then retrieved from captured imaged by registering the image into 3D space with ICP (Iterative Closest Pair) algorithm.

(Tong, Zhou, Liu, Pan, & Yan, 2012) have done their implementation by using three Microsoft Kinect devices. (Tong, Zhou, Liu, Pan, & Yan, 2012)also use ICP algorithm to do registration process. The registration step is divided into two steps which is pairwise registration between 2 successive frames and global deformation registration

In (Cui, Chang, Nöll, & Didier) implementation, captured images is rescaled by a super resolution algorithm. Resulting images are then mapped to 3D space by doing a global registration. To improve the quality of the resulting model from global registration, deformation model is done by using GMM (Gaussian Mixture Model) with ball joints constraint to ensures the rigid part stay connected while the deformation process. In (Zhang, Sun, Chi, & Sun, 2013) implementation, optimization of deformation model is done by using CPD (Coherent Point Drift) algorithm which to force GMM centroids to move coherently as a group to preserve the topological structure of the point set.

3. System Overview

Our solution is done by using 2 Microsoft Kinect devices which placed ± 2 meters in front and back of the captured object. In our solution, every human model is generated from 4 images, which is front color image, front depth image, back color image, and back depth image. We assume there is no occluded area from the captured image. The image acquisition setting of our solution is shown in Fig. 1. Our solution workflow is divided in three steps as shown in Fig. 2.

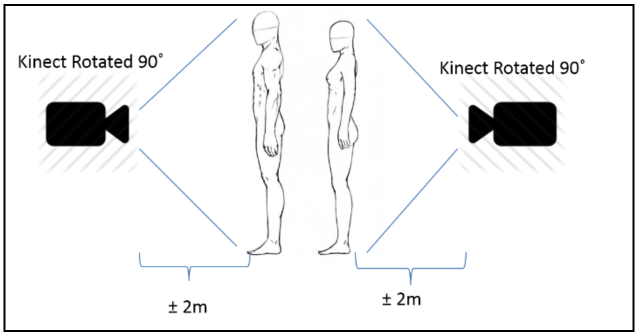

Fig. 1. Image acquisition setting

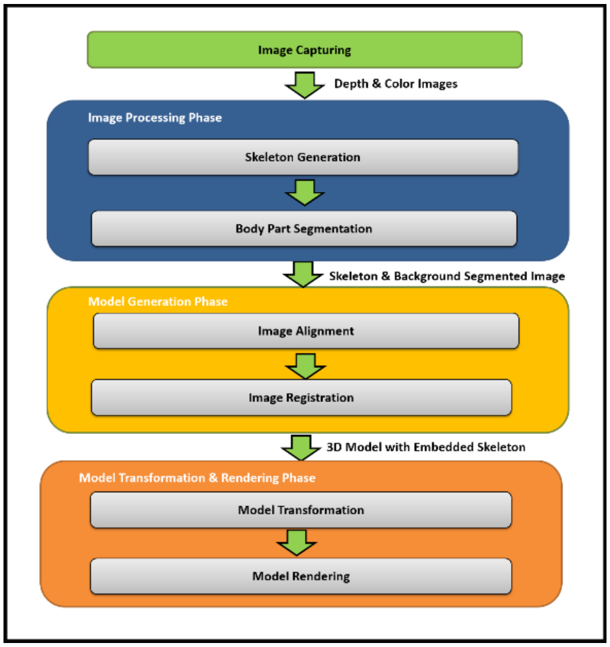

Fig. 2. Proposed solution workflow

3.1 Image Processing Phase

Our image processing phase consists of two processes, which is background segmentation, skeleton generation, and body part segmentation. Skeleton generation process is done on our previous implementation. Our skeleton generation process is divided into 4 steps, which are image preprocessing, face detection, distance transform, and anthropometry analysis. Our skeleton generation workflow is shown in Fig. 3.

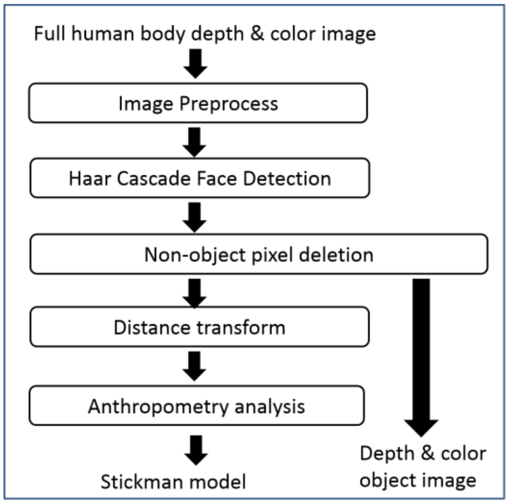

Fig. 3. Skeleton generation workflow

On image preprocess step, foreground detection is use to retrieve background removed images. Background removed images are then clipped to reduce computational cost on the next step. After image clipping, face detection is done on the clipped front color image to detect human face and then detect whole human object by using flood field from the detected face. To remove noise pixel (pixels that is not removed on foreground detection, but not a part of our detected human object), pixel deletion on non-human object is done and resulting depth & color image of the front and back Kinect device.

After pixel deletion, chessboard distance transform is performed on the depth image and from the distance transform result, we estimate the human joints location with anthropometry analysis calculation on (Cahyawijaya & Supriana, 2014). The detected joints then inserted into skeleton structure to be used in our 3D model. Our joints detection result is shown in Fig. 4. Our skeleton generation structure is shown in Fig. 5.

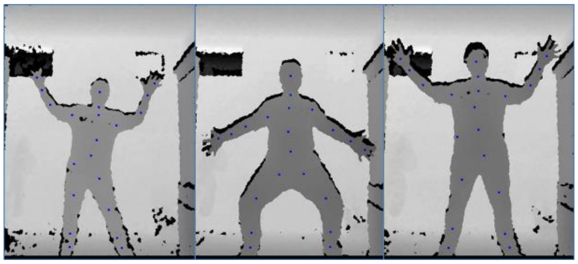

Fig. 4. Joint detection result

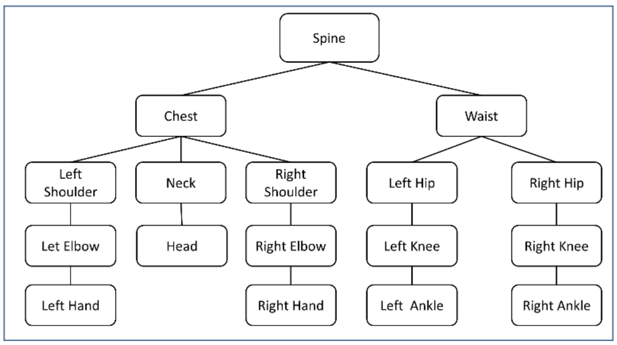

Fig. 5. Proposed skeleton structure

From the skeleton data and depth & color object images from skeleton generation step, we classify the body part of each pixels on the object images. This process is done to determine which vertex should be moved when a joint is moved. To classify each pixel on object image, first we calculate the intersection angle of each joint. The intersection angle of a joint is the mean of angle from the joint with the parent joint and the joint with the child joint. Our solution, classify human body into 17 body parts, which are head, upper body part, left upper arm, right upper arm, left lower arm, right lower arm, left hand, right hand, middle body part, lower body par, left thigh, right thigh, left calf, right calf, left foot, and right foot. Result of the body part segmentation process is shown in Fig. 6.

Fig. 6.Body part segmentation result

3.2 Model Generation Phase

Model generation phase consists of two process which is image alignment and image registration. Image alignment is divided into two steps, which are position adjustment and size adjustment. Image alignment is needed to calculating object rotation between images, in our implementation, alignment is needed to adjust the position and size of objects in front and back image, so that the image size is uniform and the object position is perfectly align.

To adjust position, first we transform the depth of back image to be relative to the front Kinect device. After that we find the head top position of front and back image and use it as our reference point to translate the back image to be in the same position with the front image. After adjusting the position, we calculate height of front head top to crotch of the front and back image. Then we calculate the size ratio of front image and back image, and then resize the back image as much as the size ratio. Overview of our image alignment step is shown on Fig. 7.

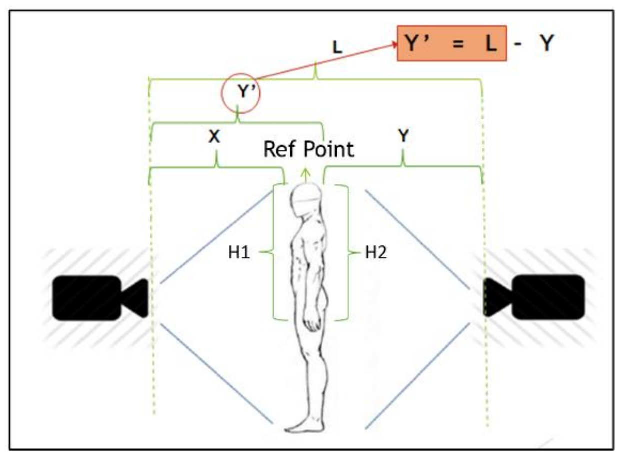

Fig. 7. Image alignment step

Image registration process maps the front image and back image into list of vertices in the 3D space. Because our solution only use image from 2 sides, in image registration process, we also need to estimate missing left side and right side part of the human 3D model.

In image registration process, we generate a vertex buffer and index buffers to render the mesh of the model in 3D. In this process we also create an index of each body part vertices to simplify body part transformation from the joint movement. The result from our image registration process is a 3D model with our own custom 3D model format as shown in Fig 8.

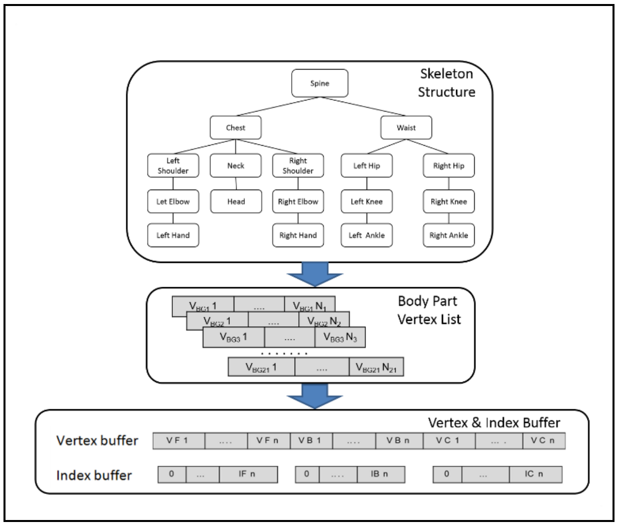

Fig. 8.Proposed 3D model structure

In image registering process, the vertex buffer and index buffers are generated by iterating every pixel on the front and back object image from top left to bottom right position. The vertex buffer contains all object pixels in the front image, all object pixels in the back image, and estimated vertex for left and right side of the object in sequential order.

We create 3 index buffers with triangle list indexing for our model, the first one for mesh information of the vertices retrieved from front image, the second one for mesh information of the vertices retrieved from back image, and the third one for mesh information of the estimated side part of human model. In triangle list indexing, every three elements in index buffer refer to one triangle mesh. The front and back meshes are generated while iterating every pixel on each image with a set of rules as shown in Fig. 9

To estimate the side body part of the body, we create a border list from front and back images. After that we split the list per body part border. For each pixels on each body part border, we create a connection between the front and the back image by adding the border vertex into the vertex buffer and create an index buffer which elements refer to mesh between the front and the back image. The indexing process for the side body part is shown in Fig. 10. After the vertex buffer, index buffer,

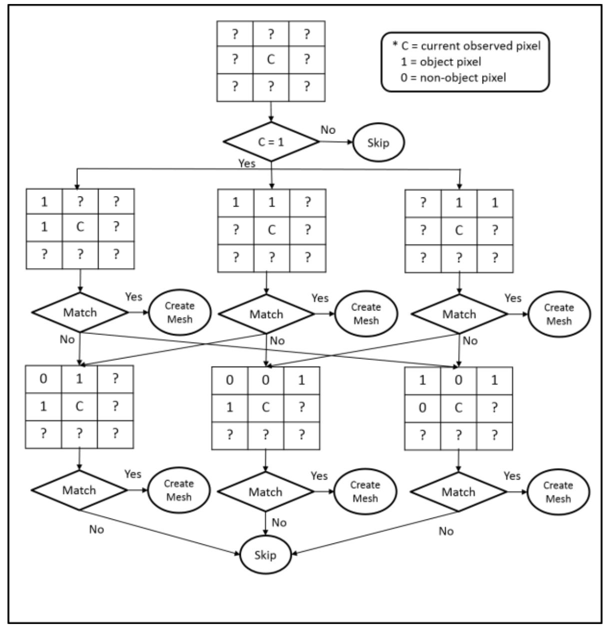

Fig. 9. Mesh generation rule for each observed pixel

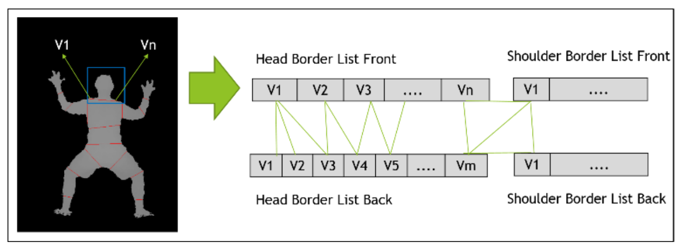

Fig. 10. Estimated side body indexing process

3.2 Model Generation Phase

Our generated model has a custom data structure, so custom model transformation & rendering phase is also needed to view and interact with the generated model. Transformation and rendering phase is divided into 2 steps, which are model transformation and model rendering. In model transformation step, we only define a rotation transformation. Rotation transformation consist of 3 value which is R_x, R_y, and R_z. If any joint is rotated, all child joints of the rotated joint is also rotated in the same degree as their parent rotated. The total transformation of a joint is a recursive function calculate from total transformation of the joint and the total transformation of the parent joint. The joint transformation calculation is shown as follow.

$$T_{(joint)} = R_z * R_y * R_x * S * T_{(parent\ joing)} \quad (1)$$

On the model rendering step, rendering is done per index buffer on the 3D model. We also control the near far clip on rendering step to handle depth buffer fighting problem. We also provided some simple animation for out 3D model such as walking and running. Our generated can have a different gesture depends on the captured object gesture, but to make the animation works perfectly, we need to set a default model gesture. So, we initialize the generated model gesture to our default gesture before animating or transforming the model. The default gesture is shown on Fig. 11.

Fig. 11. Default model gesture

4. Experimental Result

Our solution done is in C# by using Microsoft Kinect SDK to retrieve RGB and depth Image from Microsoft Kinect, AForge.NET library to perform face detection, and XNA framework to render and transform our resulting 3D model. Our solution is tested on Windows 7 Ultimate 64-bit operating system with Intel Core I5 2.67 GHz processor and 8 GB DDR3 RAM.

We evaluate our solution in three evaluation process which are calculating RMSE of our joint detection compared with ground truth, calculating accuracy of our body segmentation compared with ground truth, and conducting survey to some respondents about the 3D model shape, color, and movement quality. The result of our evaluation is shown in Fig. 12, Fig. 13, and Fig. 14.Our resulting 3D model and animation are shown on Fig. 15 and Fig 16.

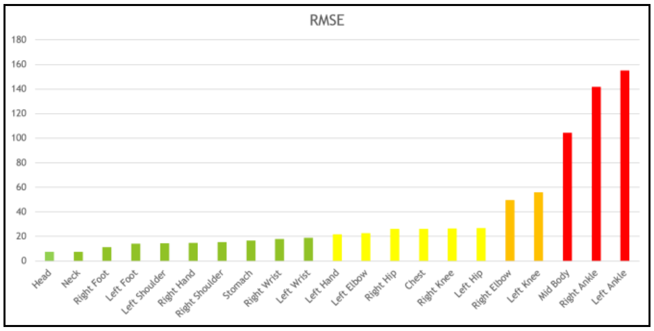

Fig. 12. RMSE of skeleton generation step compared with ground truth

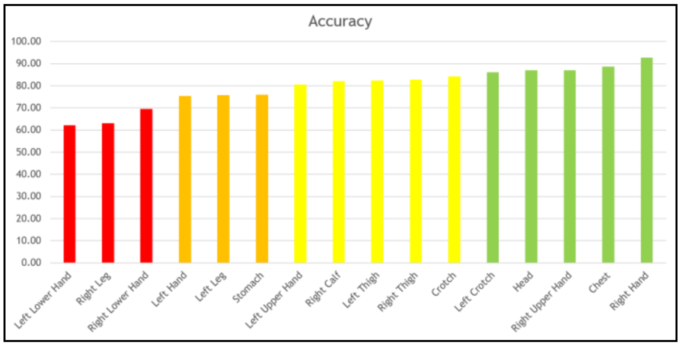

Fig. 13. Evaluation of body part segmentation step compared with ground truth

Respondents	Shape	Color	Movement
Respondent 1	6	7	9
Respondent 2	7	8	8
Respondent 3	8	7	7
Respondent 4	6	7	7
Respondent 5	6	7	6
Average	**6.6**	**7.2**	**7.4**

* In Scale of 1-10

Fig. 14. Evaluation of generated 3D model based on survey

Fig. 15. Result of the generated 3D model

Fig. 16. Generated 3D model walk animation

4. Conclusion

In this work, we have presented a new solution to create a fast and low cost human 3D scanner with embedded skeleton to support transformation on the generated 3D model. From our evaluation, we get our joint detection average RMSE of 37.861, our body part segmentation average accuracy of 79.71%, and our generated model shape quality 6.6 out of 10, color quality 7.2 out of 10, and movement quality 7.4 out of 10.

Although the result is not good enough, we would like to continue our work by improving this works in term of shape and movement quality of the model. To improve the shape quality, we need to find a better method to align and register the captured images. To improve the movement quality, we need to do a better skeleton generation approach and a better body part segmentation approach.

References

[1] Cahyawijaya, S., & Supriana, I. (2014). Automatic Human 3D Scanner Using Microsoft Kinect. *EECSI*.

[2] Cui, Y., Chang, W., Nöll, T., & Didier, S. (t.thn.). *KinectAvatar: Fully Automatic Body Capture Using a.* DFKI Augmented Vision.

[3] Sturm, J., Bylow, E., Kahl, F., & Cremers, D. (2013). CopyMe3D: Scanning and Printing Persons in 3D. *35th German Conference on Pattern Recognition (GCPR 2013).* Saarbrücken, Germany: Springer.

[4] Tong, J., Zhou, J., Liu, L., Pan, Z., & Yan, H. (2012). Scanning 3D Full Human Bodies Using Kinects. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*.

[5] Zhang, B., Sun, J., Chi, Z., & Sun, S. (2013). 3D Reconstruction of Patient-specific Femurs Using Coherent Point Drift . *TELKOMNIKA, Vol. 11*.

[6] Zhou, S., Fu, H., Liu, L., Cohen-Or, D., & Han, X. (2010). Parametric Reshaping of Human Bodies in Images. *ACM Transactions on Computer Graphics: Special Issue of ACM SIGGRAPH 2010.* SIGGRAPH.

Samuel Cahyawijya got his bachelor degree in Informatics Engineering Institut Teknologi Bandung in 2014. He is currently a founder of Pelita Cakrawala Inspirasi, a foundation that engage people to fight againts apathy and got a grant on one of most engaging national technopreneur event, Mandiri Young Technopreneur. He has been participated in Integrated Modular Avionic (IMA) research on Institut Teknologi Bandung from March 2014 until March 2015, supervised by Achmad Imam Kistijantoro, S.T,M.Sc,Ph.D. He has also been participating on computer graphics and computer vision research on Institut Teknologi Bandung since 2013, supervised by Prof. Dr. Ir. Iping Supriana.

Iping Supriana got his doctoral degree in 1986 and granted title of professor in 2009. One of his fenomenal research is Digital Mark Reader (DMR), an imaging system to automatically calculate score from test paper (called Lembar Jawab Komputer) with a speed of 12000 test paper image per minute which win the Asia Pacific ICT Award category Best Education & Training on 2004. He also got a national grant on Anugerah TIK Jawa Barat 2012. Currently, he served as a memeber of academic senate and a professor of computer graphics and computer vision in Institut Teknologi Bandung