Development of 3D Human Body Model

**Vasiliy M. Konstantinov, Vladimir L. Rozaliev, Yulia A. Orlova
and Alla V. Zaboleeva-Zotova**

Abstract This article describes a software system for scanning and three-dimensional reconstruction of the human body using the two MS Kinect devices. The article describes the main stages of the conversion of partial frames of depth in the general polygon mesh model of the human body. Also, we propose and describe a method for constructing the surface in the "empty zones". We show that using two Kinect devices our scanning system can reconstruct the human body and build the surface in "empty zones". The main idea of our reconstructing approach is to divide the whole surface of human body in two parts: the front part and the back part. Each of that two parts is dividing in three parts with overlaps: the upper part (head, part of the chest, arms), middle part and lower part. Scans of each part aligned to each other with ICP algorithm and stored in two separate point clouds, which represents the front and back part of human body. Finally, two scans reconstructed and "empty zones" between them build with our algorithm based on Bezier curve model. In conclusion represent the final model of human bode with some parameters.

V.M. Konstantinov · V.L. Rozaliev (✉) · Y.A. Orlova
Volgograd State Technical University, Volgograd, Russia
e-mail: vladimir.rozaliev@gmail.com

V.M. Konstantinov
e-mail: konstantinovr1@gmail.com

Y.A. Orlova
e-mail: yulia.orlova@gmail.com

A.V. Zaboleeva-Zotova
Russian Foundation for Basic Research, Moscow, Russia
e-mail: zabzot@gmail.com

1 Introduction

Three-dimensional models of real world objects, in particular models of a human body are applied in many areas, such as applications of the added and virtual reality, video games, in the reengineering process. Such models are created by expensive specialized hardware and software applications.

In this work we offered and described the method of receiving three-dimensional model of a human body with use two MS Kinect devices without additional equipment. The Kinect device is much cheaper than specialized scanners and provides depth stream which can be transformed to a set of three-dimensional points, and then to a polygonal grid of human body.

Now there are a large number of reconstruction systems of complex objects using the Kinect device. However, the reconstruction of the human body imposes a number of additional restrictions on the configuration of the scanning system. So to provide acceptable details level of the surface must be obtained separately the original surface of the scanned body [1, 2]. Some systems use a rotating mechanism to rotate the scanned person. In other multiple devices, which scanning certain areas [3]. Third, these approaches are combined [4].

Another problem is that not all parts of the body surface into the view of the camera. These blank areas are called "empty zones" [5, 6]—areas where the surface is not determined. This problem might be solved in different ways, but most often just an empty area being completed manually in the package of 3D graphics [7].

2 Description of the System Workflow

As entrance data the system accepts two streams from two depth cameras of a Kinect sensor. The system processes each scan (R'_k) in streams which represents the image with the permission 640×480 pixels. Each pixel describes distance to a site of the scanned surface. This distance is expressed in millimeters.

Before processing of depth it is necessary to carry out their filtration. It is a problem it is connected with that these depths returned by the device rather strongly noise. These noise have casual additive character statistically independent of a signal. In this case it is necessary to remove noise from shots of depth and in too time to keep the clearness of borders. It is caused by that borders on a shot of depth express transition between surface levels. Smoothing of borders is inadmissible and will lead to loss of information. Use of the bilateral filter allows to smooth the image with preservation of a clear boundary [8].

Further the filtered depth shots (R_k, R_{k-1}) are processed for receiving a set of three-dimensional points (V_k) and normals (N_k) to them. As, resolution of the camera of depth isn't great it makes sense to scan a surface of a body of the person in parts. For combination of partial frames of a surface the variation of algorithm of
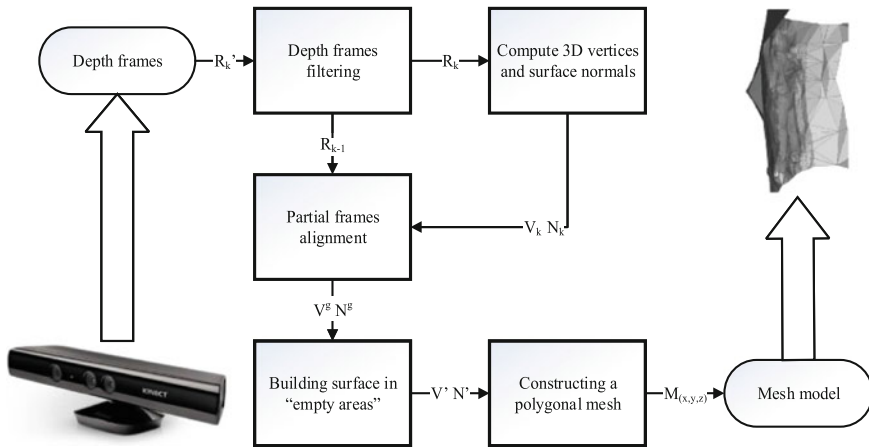
Fig. 1 Overall system workflow

ICP is used. At the exit the full frame of a surface (V^g, N^g) forward and back part of a body of the person turns out.

Further in view of the fact that there are surface sites which can't come into the view of any of depth cameras ("empty zones") it is necessary to complete a surface in these zones. The algorithm of creation of tops on the basis of square Bezier curves is for this purpose developed. At the exit of this stage the general surface (V', N') a human body united by the added tops between forward and a back surface will be received. After calculation of the general surface it is necessary to receive polygonal model $(M_{x,y,z})$. Is for this purpose used the special algorithm of reconstruction based on the solution of the equation of Poisson (Poisson Surface Reconstruction) [9].

As the output data the system provides the *.ply file which contains polygonal model of a surface of a body of the person. In Fig. 1 the scheme of the general functioning of system is submitted.

3 Description of a Creation Human Body Surface Model Method

Before start of system it is necessary to place MS Kinect devices and to allocate space for scanning of a scene. Two devices are located before the person and behind him, and sent to the opposite sides. The distance between two devices depends on growth of the scanned person. Height at which devices are located is equal to a half of growth of the scanned person.

During process of scanning of people has to be motionless precisely in the center (i.e. on identical removal from both devices) the virtual cube limited to the planes of

cameras of MS Kinect devices. Generally, the user can accept any motionless pose. However, thus no part of a lobby or back surface should block the review of the camera.

The viewing angle of the MS Kinect chamber is limited and makes 43° down. It means that the distance of capture of a surface to the utmost of the scanned person is limited. However if conditionally to divide a surface into 3 parts (top, average and lower) with rather big consecutive overlappings the distance can be reduced. It will increase the accuracy of measurement of distance to surface sites as now each partial shot will have permission 640 × 480 pixels, but not a frame of all surface.

After installation and control of system thus the preparatory stage is finished. The scheme of physical placement of system is submitted in Fig. 2.

Before processing of shots of a stream of depth it is necessary to carry out their filtration. In this work for a filtration the bilateral filter [10] is used. As the intensity of the color pixels is the distance to the surface. Thus, the formula of new value of depth (R_k) in concrete pixel (q) is expressed by a formula, where R'_k—value of distance in a concrete point of a shot of depth, u—the central point of level of a surface, $\mathcal{N}_\sigma(t) = \exp(-t^2\sigma^{-2})$—Gauss function, W_p—some weight coefficient.

$$R_k(q) = \frac{1}{W_p} \sum_{q \in \mathcal{U}} \mathcal{N}_{\sigma_s}(||u - q||_2) \mathcal{N}_{\sigma_r}(||R'_k(u) - R'_k(q)||_2) R'_k(q) \qquad (1)$$

After processing the bilateral filter will receive a smoothed shot of depth. For transformation of a frame of depth to a cloud of three-dimensional points it is necessary to calculate the first step a matrix of a perspective projection of the
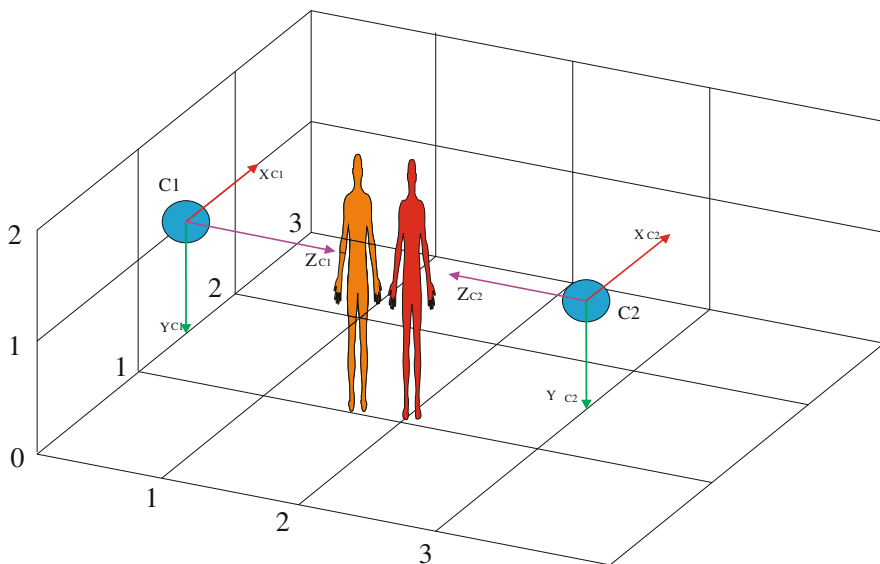

Fig. 2 Illustration of a system setup

camera (K). It can be received on the basis of an internal calibration matrix of the MS Kinect chamber.

Each pixel of a frame of depth represents a vector, where x—the coordinate of "pixel" on width of a shot, y—coordinate on height of a shot, z—value of depth.

$$Q = [x \quad y \quad z]^T \tag{2}$$

For receiving a set of three-dimensional tops (V_k) it is necessary to increase a vector (Q) by the return internal matrix for each pixel of a shot of depth.

$$V_k(q) = D_k(q)K^{-1} \tag{3}$$

The following step—calculation of a normal for each top. The vector of a normal (N_k) is calculated on the basis of coordinates of the next tops.

$$N_k(q) = (V_k(x+1, y) - V_k(x, y)) \times (V_k(x, y+1) - V_k(x, y)) \tag{4}$$

Forward and back surfaces have to be constructed by combination of partial frames. The top, average and lower frames have mutual overlappings therefore for their combination we will use algorithm of ICP (Iterative Closest Point) [11]. The top and average frame, then average and lower frames are consistently combined. On the last step the turned-out frames in a uniform surface are combined (forward and back). The scheme of overlappings in partial frames is submitted in Fig. 3.

The first step of algorithm consists in finding of coinciding tops in two partial frames. The current frame (R_k) and the previous frame are compared (R_{k-1}). If the matrix of transformation of the camera from the previous step (the previous partial
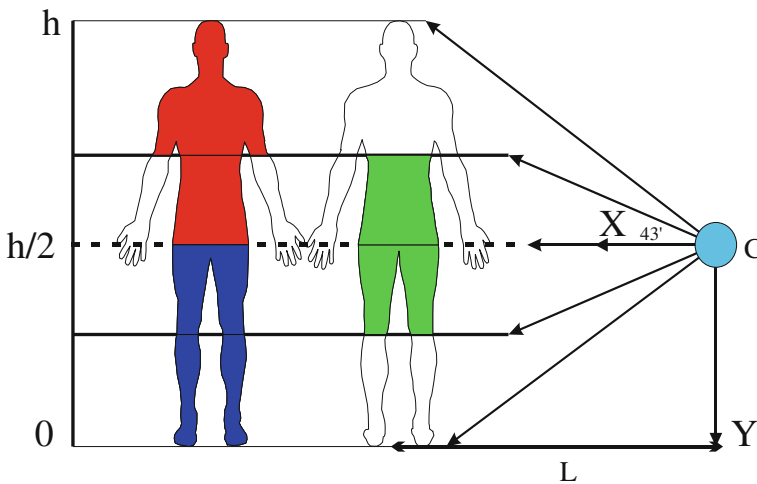

Fig. 3 Illustration of a system setup

shot of depth) is known it is possible to project tops in global coordinates (V_{i-1}). As the Kinect chamber is fixed in one point and has only one degree of freedom (rotation round axis X) the matrix of transformation has an appearance

$$T_{i-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi_{i-1} & -\sin\varphi_{i-1} \\ 0 & \sin\varphi_{i-1} & \cos\varphi_{i-1} \end{bmatrix} \tag{5}$$

Thus the formula of a projection of tops to the card of depth (p) has an appearance

$$V_{i-1} = T_{i-1}^{-1} V_{i-1}^{g} \tag{6}$$

Further, if the projection contains tops (V_i), then it is necessary to calculate new values of a normal (n) for each top.

$$n = T_i N_i(p) \tag{7}$$

Coordinates of tops don't need to be recalculated as the camera in a new shot isn't displaced on spatial axes. If threshold value (ts) that is exceeded the current point will be recognized as the combined. If isn't present, then the point is considered new (from a frame of a new shot of depth).

$$n \times n_{i-1}^{g} < ts \tag{8}$$

Value of a threshold (ts) determines by the following formula

$$\operatorname{argmin} \sum\nolimits_u ||(TV_i(u) - V_{i-1}^{g}(u)) \times n_{i-1}^{g}(u)||^2 \tag{9}$$

At the exit of this stage back and forward surfaces in the form of clouds of three-dimensional points and normals to them will be received.

After paired combination of partial frames we will receive separate back and forward surfaces of a body. Between these surfaces will be there is a thin strip of "empty zones". These zones are formed because of inaccessibility to scanning in view of a physical arrangement of cameras.

The main idea of a method of filling the empty zones is represented in Figs. 4 and 5.

For finding of boundary tops threshold value of depth by means of the operator Sobel (3 × 3) is used [12]. The top (P_1) consistently gets out of a set of boundary tops for a forward surface (P_f). For each such top the next top (P_2) gets out of a set for a back surface (P_b). Further is under construction lines of a normal for both tops and the point of intersection (P_c) pays off. On the basis of the received tops the square Bezier curve on the following formula, where B(t)—auxiliary top, t—parameter [0, 1].
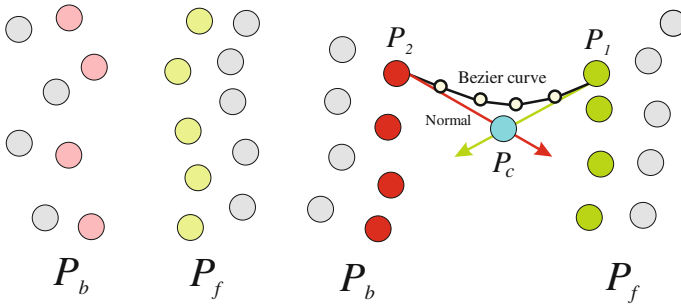
Fig. 4 Illustration of the method of constructing the surface in "empty areas"
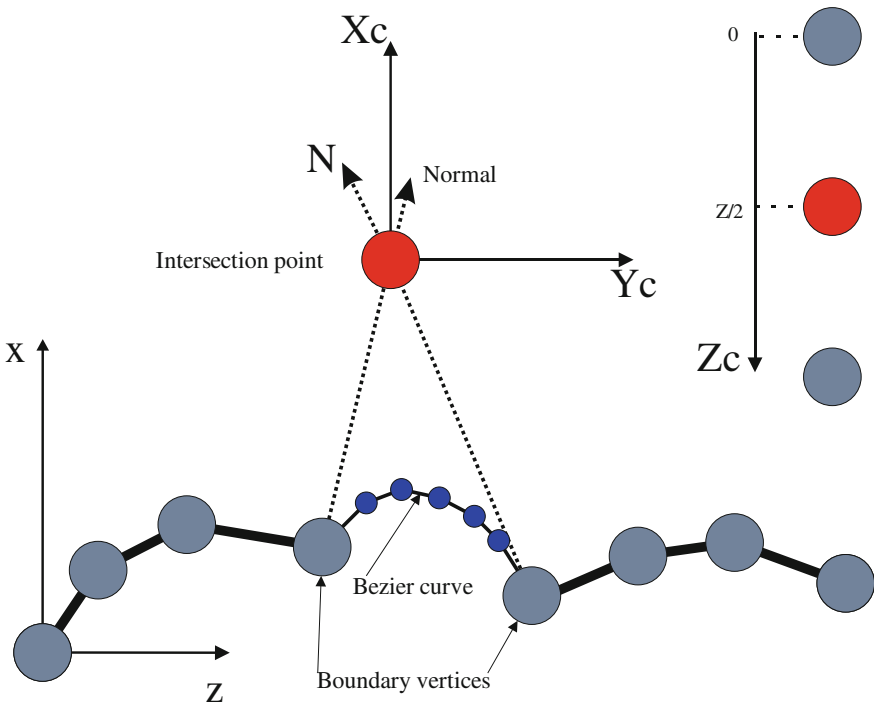

Fig. 5 Building the square Bezier curve

$$B(t) = (1-t)^2 P_1 + 2t(1-t)P_c + t^2 P_2 \tag{10}$$

Thus, "empty zones" will be filled with a set of tops of B(t), as shown in Fig. 5.

The output will be obtained overall model of the human body in the form of three-dimensional vertices reciprocally clouds. The last stage—the construction of

the mesh-method, based on the solution of the Poisson equation. The output is a polygon mesh, which is described geometry of the human body in the format * .ply.

4 Results

On the basis of this method we have developed a program for the reconstruction of surface geometry of the human body. For the implementation of reconstruction used Poisson library PCL (Point Cloud Library). The method of combining the partial surface of the frame is realized on the basis of the library Kinect Fusion. The architecture developed by the program is shown in Fig. 6. In the nearest future development of the system provided for adding functionality to animate models using BVH skeleton.

Figure 7 in left part shows the partial frames depths, which will be processed by our system. On the right represents the reconstructed body model after filtering and alignment.
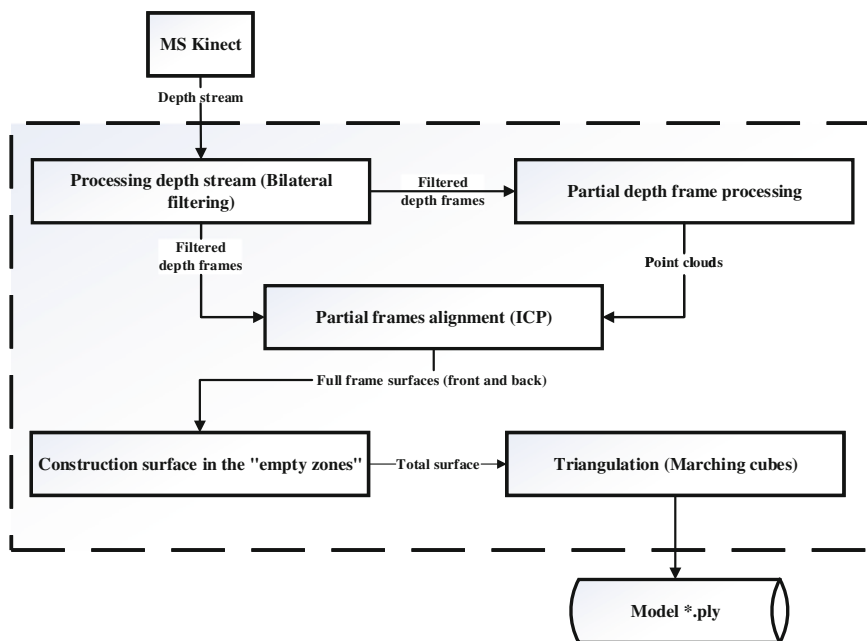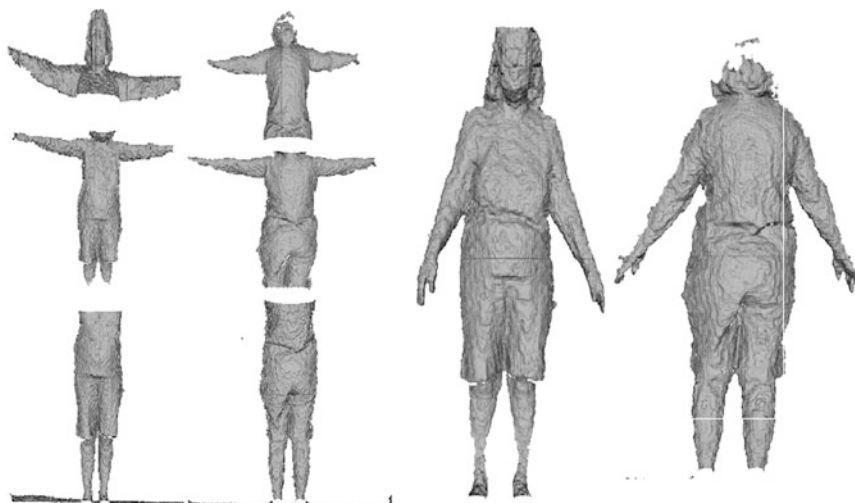

Fig. 6 Architecture

Fig. 7 Reconstructed woman body example

References

1. Huang, X.Q., Adams, B., Wicke, M., Guibas, L.J.: Non-rigid registration under isometric deformations. Comput. Graph. Forum **27**(5), 1449–1457 (2008)
2. Weiss, A., Hirshberg, D., Black, M.J.: Home 3D body scans from noisy image and range data. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1951–1958 (2011)
3. Newcombe, R.A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., et al.: KinectFusion: real-time dense surface mapping and tracking. In: 2011 10th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2011 pp. 127–136 (2011)
4. Rozaliev, V.L., Orlova, Y.A.: Recognizing and analyzing emotional expressions in movements. In: Isaias, P. (ed.) E-Learning Systems, Environments and Approaches. Theory and Implementation, Part II, § 9, pp. 117–131. Springer (2015)
5. Shum, H.P.H., Ho, E.S.L., Jiang, Y., Takagi, S.: Real-time posture reconstruction for Microsoft Kinect. IEEE Trans. Cybern. **43**(5), 1357–1369 (2013)
6. Li, H., Vouga, E., Gudym, A., Luo, L., Barron, J.T., et al.: 3D self-portraits. ACM Trans. Graph. **32**(6) (2013)
7. Sanna, A., Lamberti, F., Paravati, G., Rocha, F.D.: A Kinect-based interface to animate virtual characters. J. Multimodal User Interfaces **7**(4), 269–279 (2013)
8. Alekseev, A., Rozaliev, V., Orlova, Y.: Automatic coloring of grayscale images based on intelligent scene analysis. Pattern Recogn. Image Anal. (Advances in Mathematical Theory and Applications) **25**(1), 10–21 (2015)
9. Kazhdan, M., Bolitho, M., Hoppe, H.: Poisson surface reconstruction. In: Eurographics Symposium on Geometry Processing, pp. 61–70 (2006)

10. Cui, Y., Chang, W., Nöll, T., Stricker, D.: KinectAvatar: fully automatic body capture using a single kinect. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7729 LNCS (PART 2), pp. 133–147 (2013)
11. Lim, H., Lee, S.O., Lee, J.H., Sung, M.H., Cha, Y.W., et al.: Putting real-world objects into virtual world: fast automatic creation of animatable 3D models with a consumer depth camera. In: Proceedings—2012 International Symposium on Ubiquitous Virtual Reality, ISUVR 2012, pp. 38–41 (2012)
12. Chen, Y., Dang, G., Cheng, Z.Q., Xu, K.: Fast capture of personalized avatar using two Kinects. J. Manufact. Syst. **33**(1), 233–240 (2014)