H-013

Template Based Human Body Modeling using Kinect

Nima Baidar* and Kiyoharu Aizawa *†

*Dept. of Information and Communication Engineering, The University of Tokyo

†Interfaculty Initiative in Information Studies, The University of Tokyo

[nima,aizawa]@hal.t.u-tokyo.ac.jp

Abstract

Due to the cheaply available commercial depth cameras, various systems that use 3D scan can now be realized at natural environment. We propose a method of 3D human body shape modeling using Microsoft Kinect. However, due to the poor quality of raw depth data, we model the human body using SCAPE. SCAPE body model employs a low dimensional model of shape and pose dependent deformation that is learned from the database of range scans of human bodies. In our research, we try to estimate the parameters to model the human body using the Kinect depth information.

Introduction

3D model of human is required by various digital applications such as animation in movies and games, or motion analysis for medical purposes or even sport activities. This often requires using commercial 3D scanners which tend to be highly expensive, which often limits the 3D scanning in terms of its uses and place. Moreover, because it is almost impossible for a person to not move during scanning process, 3D modeling methods for rigid items often result in errors.

In this paper, we propose a data driven method to model a human body using the Microsoft Kinect. Our system is capable of performing a full body scan in a natural environment. We use the SCAPE model which is a 3D body model which accounts for changes in pose and variation in shape between humans. SCAPE model is learned from database of scans of various human bodies. Since the whole set of SCAPE database is not available, we use 3D scans provided by MPI Informatik[1] in order to learn the deformation model.

Related Works

Before the availability of commercial depth cameras, multiple synchronized cameras were often used to estimate the 3D model [2].[3] uses multiple but unsynchronized cameras. With the commercial 3D cameras being available, more researches [4][5] have used 3D cameras for 3D scanning of human bodies.[4] uses multiple Kinect cameras and the conventional alignment process to create a 3D model.The use of multiple camera makes the system costly and their system requires a standard setup for cameras.[5] uses silhouette as well as depth information to create the body shape. However, using a 2D silhouette can often cause ambiguities such as difficulty to differentiate between right and left leg from side -view.

Data Acquisition

We acquire the data using the Microsoft Kinect[7]. Kinect has a 640X480 resolution, each pixel containing both RGB and Depth values. However, some pixels do not have depth values. Kinect captures both RGB and depth information simultaneously at the frequency of 30 Hz. The depth sensor range for Kinect is 1.2-3.5 meters. For creating a 3D human model, first of all we require the subject to stand in a particular pose. Then we estimate the parameters for SCAPE model [5] based on the Kinect input. Then we deform the template mesh in order to create a 3D model close to the input depth.

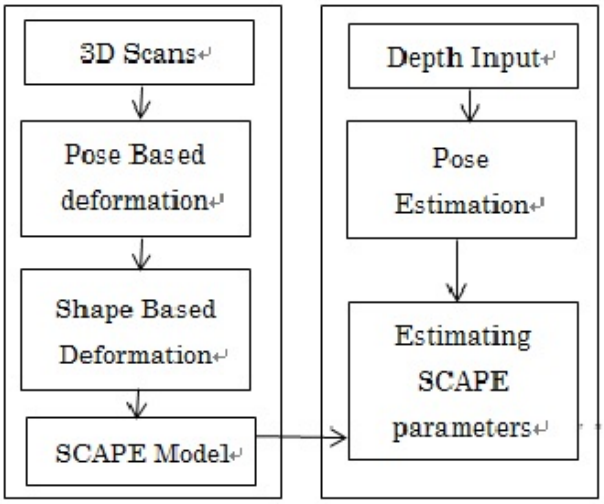

Figure 1: Algorithm Overview

SCAPE model

Our 3D surface estimation involves two main phases. The first phase consists of learning method for template deformation and the second phase consists of fitting the template to match the input from Kinect. We would like to talk about the first phase in detail below. Surface data are acquired using a whole-body scanner. The scanner captures range scans from four directions simultaneously and full body (triangular) meshes are constructed.

The database consists of two sets of data 1)pose set which contains scans of same person in 70 different poses and 2)shape set 37 different people in similar (but not identical)[6]. One of the meshes present in both the pose set and shape set is selected as a template mesh and rest of the meshes are called instance meshes. Each mesh has around 13000 triangles and 6000 vertices. The points(and therefore triangles) in template mesh and remaining instance meshes are mapped using non rigid mesh registration techniques.

Deformation Model

The objective is to build a parametric model which can recreate a 3D model based on few sparse data. The template mesh is used as reference mesh and transformations which morphs template mesh to other example meshes are estimated.Let (x1,x2,x3) be a triangle in a template mesh and (y1,y2,y3) be the corresponding triangle in the instance mesh. We define the two edges of a triangle starting at x_1 as $\Delta x = x_j - x_1, j = 2, 3$.

The deformation of triangle in template mesh to instance mesh is given by three transformations. 1.Non rigid transformation induced due to pose which is given by 3X3 matrix Q which is triangle specific in each mesh. 2.Rigid transformation induces due to pose. The deformed triangle is then rotated by rotation matrix R which is specific to a particular body part in the articulated skeleton. Hence, transformation induced due to pose can be written as

$$\Delta y = RSP\Delta x \tag{1}$$

Matrix R The rigid transformation matrix R (3X3)can be thought of as the relative transformation of the underlying skeleton of each body part between two meshes. The matrix can be derived using least squares fitting between two set of 3D points.

Matrix P Since the pose dependant data and template mesh belong to that of same person i.e. S matrix is identity matrix, and we have the rigid rotation R estimated from above step, we can solve for the non-rigid transformation matrix dependant of pose. However, the problem is underconstrained for any given triangle. Hence, a smoothness constrain which induces similar transformation in adjacent triangles is added. Specifically, the problem derives to the following equation.

$$P = \underset{P}{\operatorname{argmin}} \sum \|RP\Delta x - \Delta y\|^2 + w_s \sum_{k1,k2adj} \cdot \|P_{k1} - P_{k2}\|^2 \tag{2}$$

Subsequently, we learn a linear mapping from rigid transformation matrices R to non rigid pose based transformation matrices.i.e. $Q_\alpha(R)$
. **Matrix S** Using a similar method to that of learning P, non rigid shape based deformation matrices can be estimated using following equation.

$$S = \underset{S}{\operatorname{argmin}} \sum \|RSP\Delta x - \Delta y\|^2 + w_s \sum_{k1,k2adj} \cdot \|P_{k1} - P_{k2}\|^2 \tag{3}$$

Given the body shape deformations S between different subjects in the body shape set and the template mesh, a low dimensional linear model of the shape deformation is constructed using principal component analysis (PCA). Each S matrix is represented as a column vector and is approximated as $D_{U,\mu}(\beta) = U\beta + \mu$ where is the mean deformation, U are the eigenvectors given by PCA and is a vector of linear coefcients that characterizes a given shape.

Proposed Framework

Now we have explained how a template mesh can be deformed using the parameters rigid body part transformation R, and linear shape coefficients . Our objective now is to find these parameters from the Kinect depth input and reconstruct a new mesh. A new mesh y, not present in the training set,can be synthesized given the rigid rotations R and shape coefcients by solving

$$S = \underset{y}{\operatorname{argmin}} \sum \|RS_{U,\mu}(\beta)P_\alpha(R)\Delta x - \Delta y\|^2$$

In our method, we use the existing pose tracking algorithm in [8] .This algorithm provides the pose as shown in figure . We estimate the relative rigid transformation for each body part. Let $v1$ and $v2$ be the vectors representing a skeleton in the mesh and Kinect input respectively. Then, we calculate the rotation matrix $R_{[l]}$ required to transform $v1$ into $v2$ such that $v2 = R_{[l]}v1$.Now that we have the rigid matrix $R_{[l]}$, our next step is to find the matrix S. We estimate the joint rotation as $R_{[l1]}R_{[l2]}$ where $R_{[l1]}$ and $R_{[l2]}$ are rigid transformation/alignment of two adjacent joints.

Subsequently from the joint rotation, we calculate the exponential map $/deltarl_k$for the rotation matrix. The exponential map parameters are given by [Ma et al 2004].

$$\Delta r_i = \frac{\|\omega\|}{2sin\|\omega\|} \begin{bmatrix} r_{32} & r_{23} \\ r_{13} & r_{31} \\ r_{21} & r_{12} \end{bmatrix} \tag{4}$$

where $\|\omega\| = cos^{-1}\left(\frac{traceR_i - 1}{2}\right)$

Conclusion

We proposed a data driven method as a solution to create a 3D human model from noisy Kinect inputs. Untill now 3D scanning of human bodies had been an expensive procedure, and henceforth its applications had been limited. Realising an easy 3D body scanning method would increase usage. Moreover, using a deformable model has an advantage because of the noisy nature of the input data from Kinect. This would give us a smoother mesh compared to the 3D models created by aligning actual input depth. However, our current method is unable to recover the variation in human shape. In future, we plan to extend our work to include the variation induced due to change in shape. And, once a 3D model is created, we plan to extend it to continuos input of depth.

References

[1] A. Balan, L. Sigal, M. J. Black, J. Davis, and H. Haussecker. Detailed human shape and pose from images. CVPR, pp. 1– 8, 2007.

[2] N. Hasler, B. Rosenhahn, T. Thormahlen, M. Wand, J. Gall, and H.-P. Seidel. Markerless motion capture with unsynchronized moving cameras. CVPR, pp. 224–231, 2009.

[3] Jing Tong; Jin Zhou; Ligang Liu; Zhigeng Pan; Hao Yan; , "Scanning 3D Full Human Bodies Using Kinects," Visualization and Computer Graphics, IEEE Transactions on , vol.18, no.4, pp.643-650, April 2012

[4] Weiss, A.; Hirshberg, D.; Black, M.J.; , "Home 3D body scans from noisy image and range data," Computer Vision (ICCV), 2011 IEEE International Conference on , vol., no., pp.1951-1958, 6-13 Nov. 2011

[5] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers,and J. Davis. SCAPE: Shape completion and animation of people. ACM Trans. Graphics, 24(3):408–416, 2005.

[6] Microsoft Corp. http://www.xbox.com/kinect.

[7] OpenKinect project. http://openkinect.org.