Journal of Manufacturing Systems 33 (2014) 678–689

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

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

Volumetric template fitting for human body reconstruction from incomplete data☆

Tsz-Ho Kwok, Kwok-Yun Yeung, Charlie C.L. Wang*

Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Received 8 March 2014
Received in revised form 10 May 2014
Accepted 19 May 2014
Available online 17 June 2014

Keywords:
Template fitting
Volumetric mesh
Incomplete data
Human body reconstruction
RGB-D camera

ABSTRACT

In this paper, we present a method for reconstructing 3D human body from incomplete data, which are point clouds captured by inexpensive RGB-D cameras. Making use of the volumetric mesh in a template, the fitting process is robust. This method produces high quality fitting results on incomplete data, which are hard to be offered by the surface fitting based methods. The method is formulated as an optimization procedure, so that the results of volumetric fitting rely on the quality of initial shape (i.e., the shape of template). In order to find a good initial shape, we develop a template selection algorithm to choose a template in an iterative manner by using the statistical models of human bodies. Experimental results show that our method can successfully reconstruct human body with good quality to be used in design and manufacturing applications.

1. Introduction

Three-dimensional models of human body are widely used in many applications of robotics, virtual reality, design and manufacturing. At present, the most standard and reliable method for reconstructing 3D human bodies is to use structured-light or laser-based scanners, which can produce accurate results of 3D scanning. However, these devices usually occupy large space and are very expensive (e.g., the Cyberware system [1] costs more than \$240,000). Moreover, the scanning procedure takes a long time (e.g., >10 s). Another common approach for human body modeling is the image-based algorithms (e.g., [2,3]); however, as lack of depth information, they always produce results with low accuracy.

Recently, RGB-D cameras (e.g., Microsoft Kinect at a price of ~\$100), have drawn much attention in the community of computer graphics, design and manufacturing. The cameras can provide both depth (D) and color (RGB) information. This kind of small and inexpensive device allows consumers being able to afford it out-of-laboratory. Therefore, many researchers (e.g., [4–6]) have started to use Kinect in their 3D human modeling applications. Unfortunately, these RGB-D cameras provide noisy information in low resolution and the accuracy of depth values drops tremendously when the distance between camera and subject is large (see the detail analysis given by Khoshelham [7]). In the recent work of Tong et al. [5], the cameras are placed very close to the subjects. Three Kinect sensors and one turntable platform are utilized to compensate for the narrow visible region. However, the major disadvantage of their method is that the subject has to stand statically on the platform about half a minute during the data collection.

Our work is motivated by building up a human body scanner with inexpensive RGB-D cameras. This scanner is "instant" because the step of data collection in our platform is just taking a camera shot. Instant data-collection is a very important feature for scanning impatient subjects (e.g., children). When Kinect sensors are used to capture a full human body, they must be placed at least three meters away from the subject. However, on the other aspect, we need to place the cameras as close as possible to obtain depth information in high accuracy. We also try to reduce the number of sensors in the system. This is because too many sensors will increase the complexity of hardware installation (e.g., calibration) and meanwhile lead to interference as mentioned by Maimone and Fuchs [8]. In our system, two Kinect sensors are installed to obtain the human shape. The relative position and orientation of the sensors would affect the result of reconstruction. In order to capture the major information of the subject's shape, the sensors are installed in the front and at the back of the subject to harvest 3D information of the major body (i.e., from neck to thigh) - see Fig. 1 for an illustration. The two sensors are calibrated by using a rectangular box which has some color marker on its planar faces (see Fig. 2). Each sensor should be able to view at least 4 markers. Note that the dimension

☆ This work was supported by the Hong Kong RGC/GRF Grant (CUHK/417508 and CUHK/417109) and the Direct Research Grant (CUHK/2050518).

* Corresponding author. Tel.: +852 39438052; fax: +852 26036002.
E-mail address: cwang@mae.cuhk.edu.hk (C.C.L. Wang).

http://dx.doi.org/10.1016/j.jmsy.2014.05.009
0278-6125/

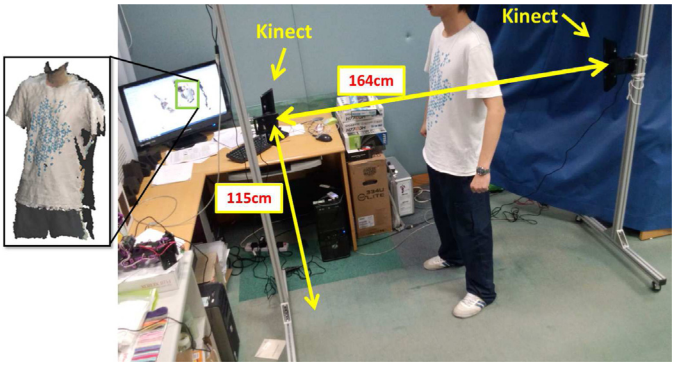

Fig. 1. Hardware setup of our scanning system: two Kinect sensors are placed to capture front and back of the major region (between neck and thigh) of a human body.

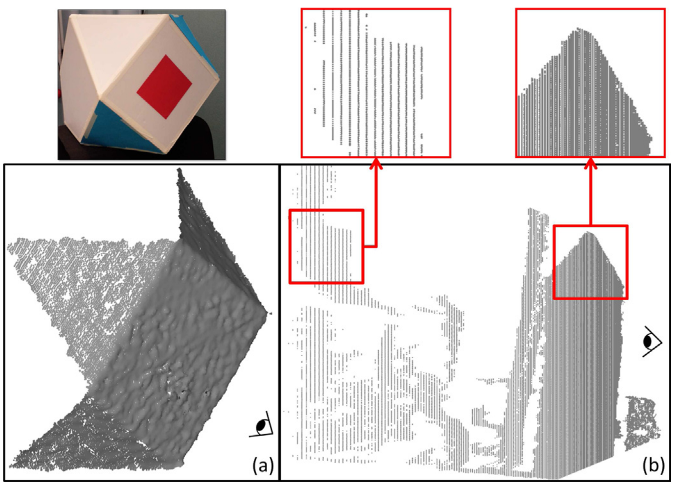

Fig. 2. The figures show the noisiness and accuracy of the depth image captured by RGB-D camera. The icon of eye shows the viewing direction of camera. The surface of the point cloud in (a) captured by the camera is very harsh, but the subject showed in photograph (upper) actually have a smooth surface. It is showed the data from RGB-D sensor is noisy. The figure (b) showing the RGB-D camera capturing an indoor environment. Two zoom views show the objects placed at around 120 cm (upper, right) and 260 cm (upper, left) away from the camera. It is clearly shown that the gap between each pixels increasing with respect to the depth value.

of the box and the coordinates of each marker are predefined and also known for the calibration algorithm. Based on the correspondences between marks extracted from the color images, the rigid transformation matrix of each sensor can be obtained. With the help of these transformation matrices, the RGB-D images captured by front and back sensors can be integrated together to represent the 3D point cloud of a subject. The 3D model of a subject will be reconstructed from this point cloud. Specifically, a volumetric template fitting based algorithm is developed for reconstructing human models from incomplete data. The template models are represented as volumetric meshes, which provide volumetric information to preserve the shape of human body robustly during the fitting procedure. Features of human body (e.g., the feature curves used in clothing industry) can be automatically extracted for the subject when the template fitting is completed.

For template fitting algorithms, a good template (as an initial guess) can always help to improve the quality of fitting. Taking advantage of statistical analysis in shape space, we propose an iterative selection algorithm in this paper to find a good template. A "good" template is defined as the one that are highly similar to the target shape. Overview of the proposed framework is shown in Fig. 3.

1.1. Related work

Literature review is taken in two categories including human body reconstruction and statistical models.

1.1.1. Human body reconstruction

There are many approaches in literature that focus on the human body reconstruction from point clouds generated by structured-light-based or laser-based scanners (Refs. [1,9]). However, such scanners are too expensive to be used out-of-laboratory. Therefore, we mainly focus on the camera-based approaches below.

Tong et al. [5] used three Kinect sensors to collect the 3D data from different parts of human body, and perform pair-wise non-rigid registration and global registration iteratively to combine the data sets collected in different time instances. A full human body can be reconstructed after the registrations. Weiss et al. [4] estimated body shape by fitting image silhouettes and depth data to SCAPE models [10]. The resultant model is a best match among all candidates instead of being generated from the data itself. The optimization step in their framework takes more than one hour. Wang et al. [6] used a single fixed 3D camera to scan a full body. They propose a part-based cylindrical representation for the human model, and estimate 3D shape of a human body from four key views extracted from a depth video sequence (captured by RGB-D camera). However, the data collection time for all above methods is too long for practical usage.

When the collected data is incomplete or with clothes, we need to estimate the underlying body shape. Hasler et al. [9] employed

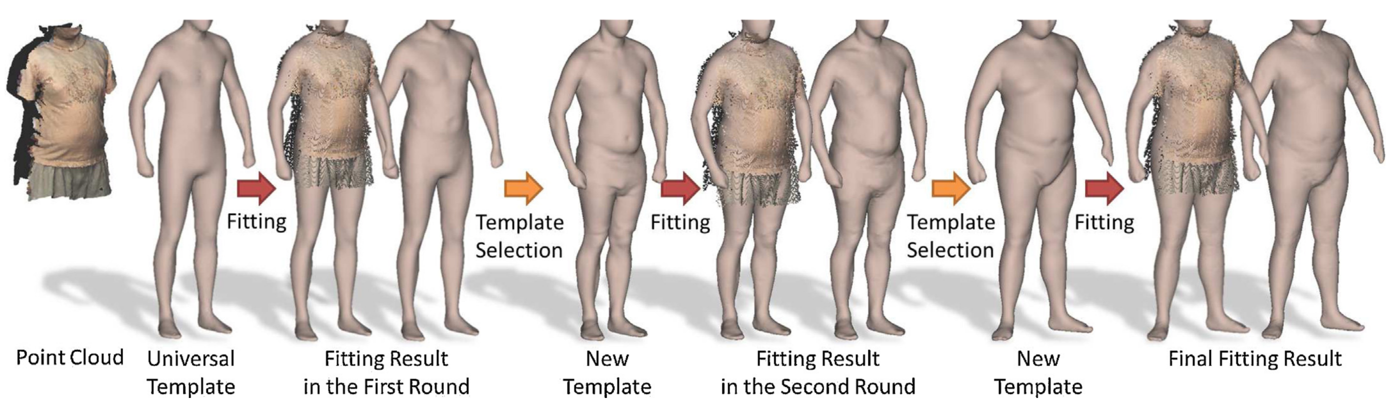

Fig. 3. Overview of our framework for human body reconstruction: by iteratively applying the interlaced selection and fitting algorithms developed in this paper, the 3D shape of a full human body can be reconstructed from the incomplete data captured by inexpensive RGB-D cameras.