Measurement 125 (2018) 535–544

Contents lists available at ScienceDirect

Measurement

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

A hybrid method for skeleton extraction on Kinect sensor data: Combination of L_1-Median and Laplacian shrinking algorithms

Erdal Ozbay[a,*], Ahmet Cinar[a], Zafer Guler[b]

[a] Faculty of Engineering, Department of Computer Engineering, University of Firat, 23100 Elazig, Turkey
[b] Faculty of Technology, Department of Software Engineering, University of Firat, 23119 Elazig, Turkey

ARTICLE INFO

Keywords:
Point cloud
Curve skeleton
Kinect sensor
Object recognition

ABSTRACT

The distinction of three-dimensional objects is one of the main challenges in computer graphics and computer vision. Distinguishing and recognizing between objects and shapes which are frequently encountered in everyday life is an important problem. In this work, a robust curve skeleton extraction algorithm is introduced on point clouds data for 3D real objects. The curve skeleton of the 3D object is a discrete geometric and topological representation of 3D shapes and maps spatial relationship of the geometric parts according to the graphical structure. Skeleton structure is the integrated stage of an average point clouds data obtained from the existing point cloud. The presented algorithm works on the average metric values of the point clouds and compensates for some missing point clouds that can be found in point clouds generated from objects. The developed method uses a combination of L_1-Median and Laplacian shrinking algorithms. Moreover, a curve skeleton can be extracted on the partially deformed point cloud. Thus, curve skeleton becomes convenient to define and process objects used in the geometric modeling. The resulting skeletal structure provides a method of object recognition that can cope with objects having complex geometry.

1. Introduction

Object recognition has begun to be used widely in recent years, especially in the machine vision industry, inspection, manipulation and registration of objects areas. The vast majority of today's systems for object recognition are based on correlation-based template matching such as image recognition examples [1]. Accordingly, these systems are sensitively affected by illumination and occlusion of objects. Uncontrolled environment variables such as setting the ambient light ratio, shadows, haziness, and color shifts cause low-quality image production. In addition, partial replacement or removal of specific features and matching them to the database has a troublesome process [2].

The skeleton structure is a frame composed of a bone, cartilage or other rigid material containing or supporting the body of an object or a living body. Blum is the first to introduce his work on skeletal structure [3]. He used that skeletal structure to represent local symmetries of the shape. Studies on the structure of the notion skeleton have introduced together the processes of Medial Axis Transform (MAT) and Symmetry Axis Transform (SAT). MAT identifies the closest boundary point/points in an object for each object in the skeleton extractions. If it has two or more, closest boundary points, the internal point is included to the skeleton. A digital binary image provides region based shape features for skeleton extraction (Skeletonization). It is a commonly used preprocessing in Raster-to-Vector transformation or in pattern recognition. When this basic structure is examined it is seen that there are three main skeletonization techniques; determining the ridges in the distance map of the boundary points, calculating the Voronoi diagram generated by the boundary points, and the layer by layer erosion called thinning.

Another skeleton extraction technique was introduce by Dickinson [4,5]. He demonstrates the feasibility of a 3D shape matching, retrieval and comparison methodology based upon using the skeleton of 3D objects. According to the approach the skeleton graphics are directly calculated from the volumetric objects and are 1D approximation of the Medial surfaces.

In another study, El-Gaaly et al. published a paper on a probabilistic way to get skeletons from 3D point clouds [6]. According to this paper they have presented a probabilistic approach for estimating skeletal representations from 3D point clouds and decomposing them into intuitive component parts. The missing parts of the model due to the natural causes are predicted effectively using probabilistic estimation procedures.

On the other hand some applications in computer graphics have been carried out to recognize and detect complex 3D physical objects.

* Corresponding author.
E-mail addresses: erdalozbay@firat.edu.tr (E. Ozbay), acinar@firat.edu.tr (A. Cinar), zaferguler@firat.edu.tr (Z. Guler).

https://doi.org/10.1016/j.measurement.2018.05.029
Received 29 June 2017; Received in revised form 8 September 2017; Accepted 7 May 2018
Available online 09 May 2018
0263-2241/

Producing inferences over the data of models which represented the objects are among the popular methods in recent decades. One of the most popular among these methods is so as to work on skeletal data. Skeleton structures reflect the individuality of physical appearance of entities. The work done on the skeletal structure is valuable since many studies have shown that skeletal structures can reveal the type of organism and even the geography where it lives. Through this inferences the topological characteristics of the models are represented by the curve skeleton structure obtained from the point cloud of 3D models [7]. 3D point clouds come from a combination of numerical data (*x*, *y*, *z*) corresponding to the spatial locations of real objects and its color information. The curve skeleton structure to be obtained through point clouds can be again represented by a cluster of points. In order to obtain a skeleton from a topological point of view, the object is completed by combining articulation zones. The curve skeleton of point cloud represents a geometrically and topologically isolated three-dimensional shape of an object. The curve skeleton structure is sufficiently concise and impressive to represent 3D objects. Moreover, when the topological structure of the objects is efficiently represented in this way, surface details are avoided and an easier environment for data processing is provided [8]. As long as the skeleton is regularized, more natural results are produced closer to reality, and distinctive features to be used for identification and recognition of objects are emerging.

In the phase of data acquisition of today's object recognition and detection applications, 3D scanners have begun to be widely used. 3D point clouds of objects through scanners such as Light Detection and Ranging (LIDAR) and Microsoft Kinect can easily be obtained for different purposes, especially because of their low cost and easy applicability [9]. A point cloud community is a collection of points that can be viewed from different angles and light conditions in 3D space and colors and textures can be applied to their surfaces [10]. Kinect sensor cameras need a library to generate and process point clouds of objects. This library is known as the Point Cloud Library (PCL) [11]. PCL is an open-source software library for transferring 3D data of real objects to spatial orientation. The PCL contains advanced algorithms such as Iterative Closest Points (ICP) and Random Sample Consensus (RANSAC). With the PCL library, these algorithms are examined to process the points through the metrics in the areas of object detection, recognition, modeling, and reconstruction. PCL and Open Natural Interaction (OpenNI) drivers are used in the developed open source software to production of point cloud of real objects using Kinect sensor camera. OpenNI drivers give us the Kinect's data as a set of three-dimensional points. When we achieve these in our program, we get an array of vectors. Vectors are store points with multiple coordinates in a single variable which access *x*, *y*, and *z* components [12]. PCL is a stand-alone and large-scale open source, BSD (Berkeley Software Development) licensed software library for 3D point cloud processing and 3D geometry processing. Major applications are as following; modeling [13], visualization [14], reconstruction [15], feature mapping [16], segmentation [8], detection and recognition [17] in the field of computer graphics.

In this paper, a robust metrical point cloud processing algorithm is presented on point cloud data of objects so as to create a directly curve skeleton without mesh structure. Whereas some studies have focused on 3D skeleton correction algorithm for 3D mesh models [18]. An approach inspired by Junjie et al. [19] has been applied to establish a natural skeleton structure and sustain the metrical processes properly in order to eliminate the occlusions and roughness that may occur during the acquisition of the point cloud of objects. This work extracts skeleton structure from the raw point cloud by applying only one iteration to the point cloud. The point cloud is topologically shrinked by one iteration reduction. This phase is becoming impossible in the curve skeleton extraction of point cloud structures with large missing data. After this phase, the L_1-Medial curve skeleton extraction algorithm is applied in a hybrid way to produce regular skeletal structure [20]. Thus, natural curve skeleton can be produced rapidly and efficiently even objects that have partially missing data.

Applied to Kinect data alone the full iteration Laplacian shrinking algorithm causes the curve skeleton data of some limbs to disappear in the last step and the curve skeleton in the center of the model to be relatively thick in the previous iteration. Similarly, the Medial skeleton extraction algorithm when app lied to Kinect data alone, causes the adjacent limbs to be included in the same skeletal structure and the skeleton of some protrusions to disappear.

The Medial skeleton extraction algorithm uses the *h* support radius value, which helps to determine the local neighborhood value during skeleton formation. If this value is larger than the skeleton distance of the different limbs, it causes the skeletal structure to be incomplete.

Since the Laplacian shrinking procedure is performed according to the neighborhood of the points, the skeletal limbs are clearly separated from each other after the first iteration of the shrinking step. The *h* support radius value to be used for the different skeletal structures separated from each other is raised to the minimum value that can be applied for the formation of the curve skeleton. The developed method has been observed to produce more natural results in a shorter time than both of the existing methods when applied to Kinect data alone.

2. Materials and methods

Low cost sensors are an alternative to expensive laser scanners for use in applications such as modeling, recognition, reconstructions, robotics, indoor mapping. One of the latest scanning technologies that can be developed at low cost range sensor of the consumer class is the Kinect sensor. Kinect has been originally designing as a source of natural interaction for the computer gaming environment. However, the characteristics of the 3D point cloud data generated by the Kinect sensor have attracted researchers working in these fields [21].

The Kinect sensor is an input device that combines IR camera, depth camera consisting of IR projector, standard RGB camera and microphones as shown in Fig. 1. Kinect sensor camera is able to capture depth and color images simultaneously at the resolution 640 × 480 pixels at a frame rate of about 30 fps. The combination of depth and color data is a 3D point cloud with about 300,000 points per each frame. [22]. The basic technique of the depth sensor is to emit an infrared light with an IR laser diode, and calculate the depth from the reflection of the light at different positions using a traditional IR-sensitive camera. The representation of scene consists of 3D point clouds, which can be computed from the depth maps provided by Kinect [12]. Proposed algorithm uses the subset of the point matches to calculate the parameters of the model.

The depth measurement principle, mathematical model and calibration parameters are explained below;

Fig. 2 shows the correspondence between the projected 3D points p_i for the camera image in p'_i position. The mathematical model includes fundamental 3 × 3 matrix *F*, which describe relations between each point *p* from the first image and corresponding point *p*′ in second image. In Eq. (1), *Fp* defines a line on which the corresponding point *p*′ must lie. The mathematical model, which satisfies most of the point matches, is used to determine good and false matches.

$$p'^{T} Fp = 0 \tag{1}$$

(p_i, p'_i) are given as a new pose correspondences. These correspondences create an equation system for estimating the camera exposure, associating the 3D coordinates of the points with their 2D image coordinates. Using this algorithm, the re-projection error (*d*) can be reduced to the minimum which is formulated as a non-linear least squares problem. The sum of the squared distances between the points projected to the 2D camera plane and the observed points is used to estimate camera exposure and known internal parameters [23].