A DATASET OF KINECT-BASED 3D SCANS

Alexandros Doumanoglou, Stylianos Asteriadis, Dimitrios S. Alexiadis,
Dimitrios Zarpalas, Petros Daras, Member, IEEE

Information Technologies Institute, Centre for Research and Technology Hellas,
6th km Charilaou Thermi, GR-57001, Thessaloniki, Greece
e-mail: {aldoum, stiast, dalexiad, zarpalas, daras}@iti.gr

ABSTRACT

Hereby, a new publicly available 3D reconstruction-oriented dataset is presented. It consists of multi-view range scans of small-sized objects using a turntable. Range scans were captured using a Microsoft Kinect sensor, as well as an accurate laser scanner (Vivid VI-700 Non-contact 3D Digitizer), whose reconstructions can serve as ground-truth data. The construction of this dataset was motivated by the lack of a relevant Kinect dataset, despite the fact that Kinect has attracted the attention of many researchers and home enthusiasts. Thus, the core idea behind the construction of this dataset, is to allow the validation of 3D surface reconstruction methodologies for point sets extracted using Kinect sensors. The dataset consists of multi-view range scans of 59 objects, along with the necessary calibration information that can be used for experimentation in the field of 3D reconstruction from Kinect depth data. Two well-known 3D reconstruction methods were selected and applied on the dataset, in order to demonstrate its applicability in the 3D reconstruction field, as well as the challenges that arise. Additionally, the appropriate 3D reconstruction evaluation methodology is presented. Finally, as the dataset comes in classes of similar objects, it can also be used for classification purposes, using the provided 2.5D/3D features.

Index Terms— 3D reconstruction dataset, Kinect Sensor, Vivid VI-700 Non-contact 3D Digitizer, Fourier-based 3D reconstruction, Poisson surface reconstruction

1. INTRODUCTION AND RELATED WORK

Since its release date (Nov. 2010), the Microsoft Kinect sensor has attracted the attention of many researchers and home enthusiasts, due to its ability to produce high-resolution depth maps in real-time and mainly due to its low price. Many Kinect-based applications have already appeared, including 3D reconstruction-based ones; however, there is still a lack of appropriate datasets of small-sized objects, with fine details, captured by a Kinect sensor (along with the ground truth), that can be used for experimentation with 3D reconstruction problems. Instead, there exist a few datasets dealing with 3D information of objects, usually obtained using laser scanners. One of the mostly used datasets for 3D reconstruction of small-sized objects is the MiddleBury Multi-View Stereo dataset, described in [1], where the authors captured two objects from many different positions, using a CCD camera, while the reference 3D models were captured using a Cyberware Model 15 laser stripe scanner. 200 individual scans were taken for each object. 32 objects, usually found in a household, were captured by stereo cameras in [2]. Ground truth regarding 3D data was acquired using a SICK LMS400 range scanner with 0.5 degree angular resolution. The objects were positioned on a turntable and scans were taken at 30° intervals. The authors in [3] have developed a dataset consisting of 6 outdoor complex surfaces. For acquiring 3D positions (which were further utilized as ground truth for multi-view reconstruction), they used a Zoller+Fröhlich IMAGER 5003 laser scanner. For tackling complexity of the surfaces, each surface was scanned from multiple positions.

In this paper, a dataset of small to medium-sized objects is proposed, captured using a Kinect sensor. We apply two well known surface reconstruction methodologies, in order to highlight the challenges imposed by the Kinect-based data, contrasting to a standard laser scanner. The dataset consists of 3D information of each object from a multitude of viewpoints, as well as calibration information, necessary for reconstruction, while the corresponding 3D data, as they were captured using a laser scanner are provided for reconstructing ground truth. Moreover, RGB color images and depth information of each view are provided as part of the dataset. Objects were grouped in classes of visual and contextual similarity, hence can be used for studies on 2.5D/3D feature extraction for classification purposes. It is expected that the proposed dataset can contribute to the improvement or development of novel algorithms targeting the challenges imposed by Microsoft Kinect's mechanism of acquiring three-dimensional information.

This work was supported by the EU funded project 3DLIVE, GA 318483

2. SETUP AND ACQUISITION

A total of 59 small-sized objects (toys) were scanned sequentially from multiple viewing angles using a turntable. Fig. 1 shows representative examples of the classes composing the dataset. Each class consists of objects belonging to the same type of object (17 land mammals, 6 dinosaurs, 11 sea mammals, 10 objects with humans, 2 guns, 2 bugs, 5 cars and 6 uncategorized objects); hence, the proposed dataset can be utilized for classification purposes, based on appearance and 3D features. The turntable was rotated at steps of 20° clockwise, in the case of the Kinect sensor and 40° counter clockwise, in the case of the Vivid VI-700. This resulted in a total of 18 views per object for the Kinect and 9 for the Vivid VI-700. The distance between the Kinect sensor and the rotation axis of the turntable varied from 67 to 70cm, while Vivid scanner was positioned from 85 to 92cm with regards to the rotation axis.

The resolution of the Kinect depth-maps is 640×480, while the range-scans resolution is 200×200 in the case of the Vivid VI-700. Each object view is also accompanied by files corresponding to the 3-D coordinates of each point in the corresponding point cloud. These measurements were acquired by using standard OPENNI unprojection functions for the Microsoft Kinect sensor and VIVID SDK for the VI-700 Digitizer and are given in millimeters. Additionally, the necessary extrinsic calibration information is provided with the dataset, which can be used for the registration of the 3D data captured from the different views. More specifically, the rotation axis of the turntable is given with respect to the sensor's 3D coordinate system and in the form of a 3D point and an orientation vector. For the extraction of these parameters, a calibration pattern consisting of 2 planar surfaces meeting at the turntable's central axis was placed on the turntable and scanned both by Kinect and VI-700. Depth data of each planar surface were used in a plane fitting algorithm to obtain the 3D equations of the planar surfaces. Since they meet at the turntable's central axis, this axis was calculated using plane-by-plane intersection. The full dataset is available from the following location: `http://vcl.iti.gr/3d-scans/`.

3. 3D RECONSTRUCTION EXPERIMENTS

In order to validate the applicability of the dataset for 3D reconstruction, two well-known methodologies [4, 5] that show to be resilient to data noise were applied to the captured data. Both methods work with oriented point sets as input (i.e. vertices plus their oriented normals). Therefore, the normal on each vertex was calculated using Principal Component Analysis on neighborhoods of 18 points. The 3D point clouds generated for each view, as well as the corresponding sets of normals, were subsequently registered to a common coordinate system using the extrinsic calibration data.

Additionally, 3-D points corresponding to objects other than the scanned one (e.g. the turntable and the background) were excluded from the point set. The radius of the turntable used is approximately 250mm. We used this information along with the rotation axis position and direction to reject points that are outside the turntable surface when reconstructing the object's point cloud. Fig. 2 shows an object's point cloud as acquired from both sensors.

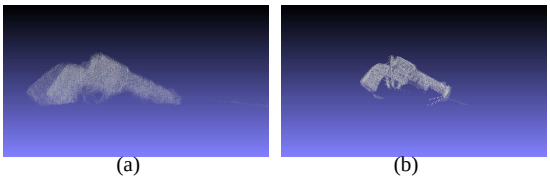


Fig. 2. Examples of registered point clouds taken from different views, using (a) Kinect sensor and (b)Vivid VI-700

3.1. Fourier-based 3D reconstruction

We applied the methodology described in [4] for extracting watertight surface reconstructions of each object in the dataset. Specifically, this technique computes a volumetric indicator function $\chi(x, y, z)$ of the solid model, which equals the unity inside the model and zero outside. The volumetric function is efficiently calculated in the 3D Fourier transform (FT) domain, making use of Stoke's theorem and exploiting only the information in the input oriented point set (vertex positions and normals). The indicator function in the original 3D domain is given by the inverse 3D FT and finally the model's surface is obtained as the extracted 0.5-level isosurface of the indicator function. Results of the reconstruction, for one example object, are presented in Fig. 3(c) and 3(d).

3.2. Poisson 3D reconstruction

Similarly to the previous approach, the Poisson reconstruction [5] method aims at computing an appropriate volumetric indicator function from the input oriented point set and extracting an isosurface of this function. The main idea behind the Poisson reconstruction method is that the gradient of the indicator function is actually a vector field that is non-zero only near the surface of the object, where it equals the surface normal. Therefore, the input oriented point samples constitute actually samples of the indicator function's gradient and the problem reduces to finding the scalar volumetric function $\chi(x, y, z)$ whose gradient equals the vector field $\vec{V}$ defined by the samples. Applying the divergence operator, the problem translates into computing the function $\chi(x, y, z)$ for which holds:

$$\Delta\chi \equiv \nabla \cdot \nabla\chi = \nabla \cdot \vec{V} \quad (1)$$

Additionally, since the accurate representation of the implicit indicator function is only necessary near the surface, the use

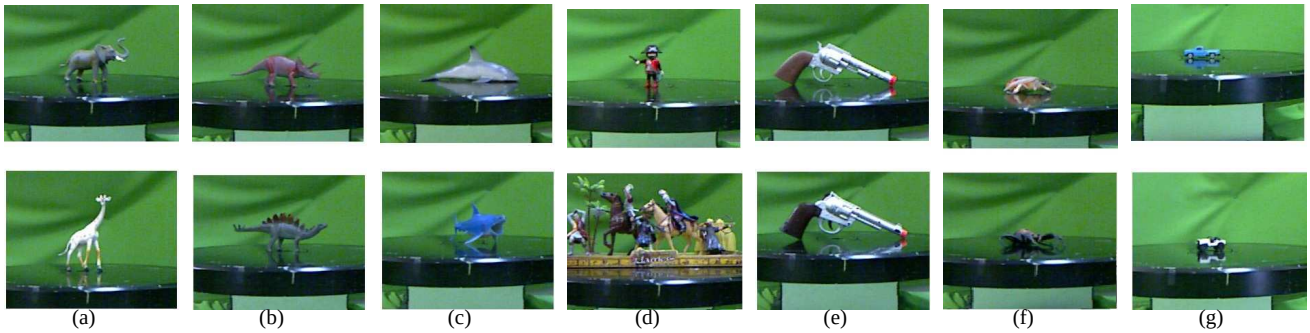


Fig. 1. Typical examples of different object categories: (a) Land Mammals, (b) Dinosaurs, (c) Sea Mammals, (d) Humans, (e) Guns, (f) Bugs, (g) Cars.

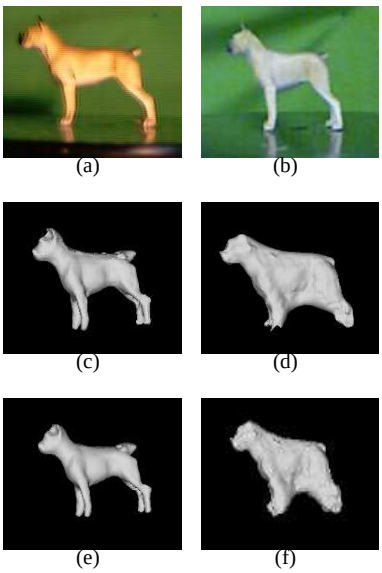


Fig. 3. RGB image of object captured with Vivid VI-700 (a) and Kinect (b). Reconstruction based on Fourier Coefficients of the Vivid VI-700 sensed point cloud (c) and the Kinect-based one (d). Poisson reconstruction of the Vivid VI-700 sensed point cloud (e) and the Kinect-based one (f).

of an adaptive octree-structure to represent the implicit function is possible, in order to solve efficiently the problem. We applied the method using an octree levels depth of 8. Results produced with Poisson reconstruction can be found in Fig. 3(e) and 3(f).

4. EVALUATION OF THE RECONSTRUCTION

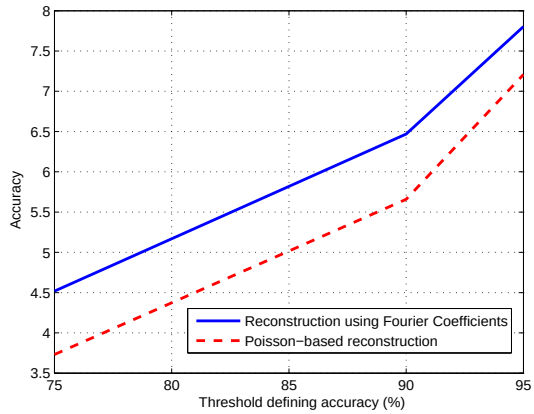
As can be seen from Fig. 3, Vivid VI-700-based range data, although of lower resolution and half as many views than the Kinect part of the dataset, provide much more accurate information for reconstructing the objects' surface. Furthermore, inspection of the whole dataset reveals that Poisson-based 3D reconstruction [5] has the ability to extract the 3D structure of our objects more effectively, yet at the cost of higher computational effort. Consequently, the Poisson-based reconstructed models from the Vivid VI-700 data were used as reference (ground truth), in each case. However, other (more recent or future) methodologies for 3D reconstruction may be even more accurate; here, the use of Poisson 3D Reconstruction provided with a satisfactory framework for evaluating Kinect-based reconstructions.

Since the relative positions of the Kinect and the Vivid VI-700 sensor are unknown for each object, an Iterative Closest Points (ICP) algorithm was applied to register the reconstructed model with the ground-truth one. This enables one to compare the two models effectively and perform a meaningful quantitative evaluation. The employed ICP algorithm is based on Delaunay tessellation of the 3D points for efficiently finding closest points [6].

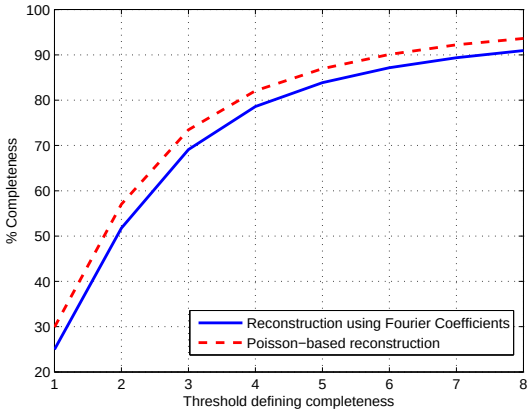
For quantitatively evaluating our results, both accuracy and completeness were measured [1]. Accuracy is defined as the distance d, for which T_{pnts} percent of the total points in the reconstructed model M have less or equal distance to their closest point of ground truth model G:

$$\begin{aligned} |M_d| &\geq T_{\text{pnts}} \times N \\ M_d &= \{\|M_i - G\| < d\}, M_i \in M \end{aligned} \tag{2}$$

Completeness measures how complete the reconstructed model is. It is defined as the percentage T of points of the ground truth model, whose distance to the reconstructed model is smaller than T_d. Fig. 4(a) shows the accuracy for three different thresholds (75-90%), while Fig. 4(b) shows completeness versus different criteria (thresholds), from 1mm to 8mm. It can be deduced that Poisson reconstruction outperforms the Fourier-based approach, as it is more robust to salient, high frequency information and can produce smoother reconstructions in areas of sparse samples. Results show that the proposed dataset is very challenging due to the objects' nature, as well as the inaccurate, noisy and sometimes incomplete Kinect data, especially for smaller objects. The objects



(a)



(b)

Fig. 4. Completeness and Accuracy of reconstructing kinect-based depth measurements, using Poisson-based and Fourier Coefficients-based reconstruction methods, for the whole dataset.

used present a large variety in terms of size and details, spanning from very small to larger ones, all having detailed structure. Taking a closer look at the completeness/accuracy results with respect to the size of each point cloud showed that both completeness and accuracy deteriorate for objects described with smaller point clouds. More precisely, analysis of variance between point clouds' sizes (as resulted after reconstructions using the Vivid VI-700 sensor) and accuracy, gave low significance values ($p < 0.1$) for thresholds $T_{\text{pnts}} \geq 90\%$, while significantly low p-values ($p << 0.05$) were given for point cloud size and completeness, for most thresholds.

5. DISCUSSION AND CONCLUSIONS

Current 3D datasets are usually aqcuired utilizing laser sensors, which are known to deliver quite accurate results and most reconstruction methods proposed in bibliography deal with such data. With the advent of the Kinect sensor, and its relatively low cost, the tendency and need for extracting 3D information based on it, has shifted towards more affordable solutions. The proposed dataset offers a kinect-based collection of depth scans of small to medium-sized objects, as well as the corresponding 3D data that can be used for reconstructing ground truth, using a laser sensor. The application of well-known techniques for reconstruction has highlighted the challenge imposed by the kinect-based collected data. The proposed dataset is accompanied by RGB image data, as the use of visual information for supporting 3D information extraction can play an important role for retrieving complex three-dimensional geometries. We encourage researchers in the field of 3D reconstruction to evaluate their methods on this dataset. The purpose would be to find out which types of methods can efficiently deal with the Kinect data and its shortcomings. A further criterion for the evaluation would be the execution time required for the reconstruction. Kinect is offering a high frame rate of depth images, and, thus, enables real time applications, which was out of topic with the laser scanners. Based on the aforementioned, Kinect is opening a new research field, real-time 3D reconstruction from imperfect-noisy data, and the offered dataset aims to provide the means for promoting such research.

6. REFERENCES

[1] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006, pp. 519–528.

[2] Jan. 2013. [Online]. Available: https://ias.cs.tum.edu/software/semantic-3d

[3] C. Strecha, W. von Hansen, L. J. V. Gool, P. Fua, and U. Thoennessen, "On benchmarking camera calibration and multi-view stereo for high resolution imagery," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[4] M. Kazhdan, "Reconstruction of solid models from oriented point sets," in *Proceedings of the 3rd Eurographics symposium on Geometry processing*, 2005.

[5] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson surface reconstruction," in *Proceedings of the 4th Eurographics symposium on Geometry processing*, 2006, pp. 61–70.

[6] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, "The quickhull algorithm for convex hulls," *ACM transactions on mathematical software*, vol. 22, no. 4, pp. 469–483, 1996.