Point-cloud Decomposition for Scene Analysis and Understanding

Sandro Spina[1]

Department of Computer Science, University of Malta
sandro.spina@um.edu.mt

Over the past decade digital photography has taken over traditional film based photography. The same can be said for video productions. A practice traditionally reserved only for the few has nowadays become commonplace. This has led to the creation of massive repositories of digital photographs and videos in various formats. Recently, another digital representation has started picking up, namely one that captures the geometry of real-world objects. In the latter, instead of using light sensors to store per pixel colour values of visible objects, depth sensors (and additional hardware) are used to record the distance (depth) to the visible objects in a scene. This depth information can be used to create virtual reconstructions of the objects and scenes captured. Various technologies have been proposed and successfully used to acquire this information, ranging from very expensive equipment (e.g. long range 3D scanners) to commodity hardware (e.g. Microsoft Kinect and Asus Xtion). A considerable amount of research has also looked into the extraction of accurate depth information from multi-view photographs of objects using specialised software (e.g. Microsoft Photo-Synth amongst many others). Recently, rapid advances in ubiqutous computing, has also brought to the masses the possibility of capturing the world around them in 3D using smartphones and tablets (e.g. http://structure.io/).

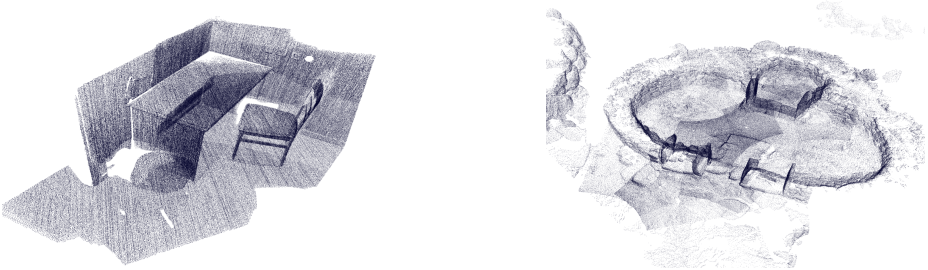

Fig. 1: Mnajdra (Scanned - Heritage Malta) and Office (Scanned - [2])

In a similar fashion to digital photography, the widespread availability of hardware capable of capturing 3D information is also leading to the creation of

massive repositories of 3D data sets. These data sets, *point-clouds*, minimally consist of 3D coordinate values representing surface positions of the scene acquired (scanned). A wide variety of scenes can be scanned with the quality of the data captured depending on the acquisition method used. The office room (figure 1a) indoor scene was acquired using commodity depth sensors (in this case Microsoft Kinect) resulting in a relatively noisy and incomplete scene. Figure 1b shows the point-cloud of a section of the smallest of three temples in the Mnajdra pre-historic site. The acquisition process was carried out using professional grade 3D-scanners. Depending on the task at hand, manipulation (post-processing) of these point-cloud data sets usually requires extensive expertise in the use of CAD and modelling software. In this work we propose automated mechanisms to alleviate some of these tasks. In particular, we address the problem of scene understanding where meaningful structures (e.g. walls) and objects (e.g. chairs) are automatically extracted from raw point-clouds representing a variety of scenes. Previous work in the area has produced solutions which target specific environments, thus leading to assumptions limiting the adaptability of these techniques to other scenarios. Recent examples include [2] and [1]. In our case, we first tackled the problem of identifying generic structures within scenes [3] by partitioning point-clouds into connected surface segments, then generalised the solution to introduce object extraction.

Our current solution is split in two phases. Initially a training phase is carried out to concisely describe individual objects (e.g. tables, chairs, sofas, aeroplanes, etc.). Each model is described in terms of a set of feature description graphs storing surface connectivity and contextual information. As is custom in these learning scenarios the representation used tries to minimise the distance between objects in the same class (e.g. different chair models) and maximise that between classes. In the second phase, the target scene is first decomposed via a segmentation process to produce a meaningful set partition of surfaces, then a Markov decision process is applied on these segments in order to enumerate valid solutions. The presentation outlines current research progress, objectives and future directions.

References

1. Lin, H., Gao, J., Zhou, Y., Lu, G., Ye, M., Zhang, C., Liu, L., Yang, R.: Semantic decomposition and reconstruction of residential scenes from lidar data. ACM Trans. Graph. 32(4), 66:1–66:10 (Jul 2013), http://doi.acm.org/10.1145/2461912.2461969
2. Nan, L., Xie, K., Sharf, A.: A search-classify approach for cluttered indoor scene understanding. ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2012) 31(6) (2012)
3. Spina, S., Debattista, K., Bugeja, K., Chalmers, A.: Point cloud segmentation for cultural heritage sites. In: Niccolucci, F., Dellepiane, M., Serna, S.P., Rushmeierand, H., Gool, L.V. (eds.) VAST11: The 12th International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage. vol. Full Papers, pp. 41–48. Eurographics Association, Eurographics Association, Prato, Italy (11/2011 2011), http://diglib.eg.org/EG/DL/WS/VAST/VAST11/041-048.pdf