Interactive 3D Reconstruction: New Opportunities for Getting CAD-ready Models

Julius Schöning

Institute of Cognitive Science
University of Osnabrück, Germany
`juschoening@uos.de`

Abstract

A multitude of image-based 3D reconstruction and modeling techniques exist, which have achieved significant success in recent years. However, these techniques still lack certain abilities. For example, current 3D reconstruction techniques cannot decompose an object into its individual subparts. Thus, a printed model will consist of one single monolithic piece, which does not allow composing or decomposing parts, does not allow movable or flexible parts, and does not allow manufacturing the model from multiple different materials like wood, metal, or plastic. I reviewed the work in the research area of 3D reconstruction and provide an analysis of neglected research objectives and current drawbacks. Furthermore, I propose a mock-up of an interactive tool as a guideline for future research which describes a possible architecture, user interfaces, and processing pipeline, to overcome existing drawbacks of 3D reconstruction techniques.

1998 ACM Subject Classification I.4.5 Reconstruction

Keywords and phrases 3D Reconstruction, 3D Modeling, Interactive Reconstruction, CAD-ready

Digital Object Identifier 10.4230/OASIcs.ICCSW.2015.54

1 Introduction

The reconstruction of 3D objects and scenes from image and video footage captured in 2D (monocular) or 3D (stereo) is an active research topic in the area of computer vision. Many approaches reconstruct objects or scenes with automatic algorithms from image sequences [28, 20, 22, 23, 34, 27]. Even today, 3D reconstruction is a difficult task due to the numerous irregularities of objects and scenes. Thus, many automatic algorithms fail to create a 3D model when, e.g., only few images are available [14], the images are not taken in a controlled environment, the object contains textureless surfaces, or parts of objects are occluded. In addition, when a 3D model is created by an automatic or semi-automatic method [15, 2, 8, 17, 12], the re-transformation via a 3D printer or CNC-machine ends up with a solid monolithic block of, e.g., plastic. This is because most reconstruction algorithms focus on the surfaces or the skeleton tracking of an object and not on its subparts, despite it being possible to manufacture more than solid blocks with today's techniques of CNC-machines and 3D printers out of these models.

When people look at objects or scenes they can directly extract the visible geometry, as well as the hidden geometry. In addition, humans are capable to do this extraction even without the necessity for stereo information and can identify materials and the degree of freedom of individual subparts. But an even more basic decision a person can make without great difficulty is whether a part is fixed or movable. I want to expand the computational power of today's computer vision techniques with the conceptional background knowledge provided by the user who, since birth, acquired real world knowledge and became a domain

2015 Imperial College Computing Student Workshop (ICCSW 2015).
Editors: Claudia Schulz and Daniel Liew; pp. 54–61
OpenAccess Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

expert. Currently, this knowledge cannot be formalized into computer algorithms and it is unlikely that this situation will change in the foreseeable future. The idea of uniting the strengths of computer vision with the user's domain knowledge and his or her decision-making ability is, however, not new in this research area but is rarely found. I therefore provide a literature review of 3D reconstruction and modeling techniques and analyze weaknesses of existing methods. I provide a discussion of how an interactive process could overcome existing weaknesses and propose a possible working process for the interactive creation of non-monolithic CAD-ready 3D models with flexible subparts out of defined material. Such CAD-models allow manufacturing each subpart on CNC-machines or 3D printers.

2 State of the Art Review

For 3D reconstruction techniques three main research areas are observable: Fully automatic reconstruction, interactive reconstruction, and supervised reconstruction. All of them work "either-or" with monocular or stereo images or video footage. The 3D reconstruction techniques discussed in the following do not necessarily focus on reconstructing scenes or objects for modeling purposes only. Instead other applications like, e.g., robot navigation, might be considered.

2.1 Fully automatic reconstruction

Based on large image collections fully automatic 3D reconstruction methods have been shown to work well [25, 15]. The photo explorer by Snavely et al. [27] uses an unstructured image collection of a scene, e.g., acquired from the Internet, and converts them to a 3D model which can be exported. Instead of an image collection Pollefeys et al. [23] use a monocular hand-held camera to build visual models of scenes for fusing real and virtual scenes. 3D modeling of indoor environments for robot navigation is done by RGB-D mapping using a depth camera [10]. Another famous approach, known as *KinectFusion* project [18, 16], uses a RGB-D camera as hand scanner. The reconstruction of objects on consumer mobile phones which are moved around the object of interest is described in [28]. A method shown by Pan et al. [20] for probabilistic feature-based online rapid model acquisition reconstructs freely rotated objects in front of a static video camera in near real-time. Due to the fact that the system guides the user with respect to the manipulation, i.e., rotation of the object, this method might not entirely fit in the category of fully automatic methods. However, like some other methods in this category it still requires some interaction to be initialized before reconstruction.

The handling of delicate structures, textureless surfaces, hidden boundaries, illumination, specularity, or even dynamic or moving objects, like in natural recordings, are not taken into account by these automatic reconstruction techniques. High quality automatic reconstruction usually depends on a predefined recording environment with, e.g., special or fixed lighting conditions, a rotating object with a fixed camera or a rotating camera with a fixed object, and special camera systems like ToF cameras [30, 7], stereo cameras [13], IR cameras [9] or laser sensors [29, 1].

2.2 Interactive semi-automatic reconstruction

Kowdle et al. [15] noticed the above mentioned issues of fully automatic reconstruction for object creation and came up with a semi-automatic approach. In this approach they put the user in the loop of computational reconstruction. This interactive approach yields a reconstruction of a monolithic high quality object or scene from an image collection. Interactive approaches for reconstructing man-made architecture [2, 8, 17, 12] are based on the presence of common geometric primitives. Usually such approaches use a computational pre-processing step that extracts edges or point clouds to create an initial 3D model. In the interactive processing step the user iteratively refines the model by editing the polygon based model or sketching new polygons. This kind of 3D drawing for reconstruction purposes is also mentioned by Hengel et al. [32, 31].

With interactive 3D reconstruction techniques, models can be reconstructed from a single image [6]. In general, the potential of semi-automatic approaches is shown in many computer vision systems and computer vision related disciplines. To point out a few examples where interactive concepts led to a significant improvement: interactive segmentation of images [19, 4, 24], the integration of domain knowledge in a visual recognition system [5], or the inter-active training of a video recognition system [11].

2.3 Supervised or active learning reconstruction

Supervised and active learning as a field of machine learning improves many computer vision issues, e.g., annotation, object categorization and video categorization. Nowadays, supervised or active learning are not yet used in interactive 3D reconstruction. Using active learning, 3D reconstruction improves itself by asking the user for specific information. Using this specific information a supervised learner can be trained. As a result the supervised learner is able to process common reconstruction task like reconstructing walls or entire rooms after a few iteration of active learning. These learning loops improve an interactive reconstruction process and reduce the interaction tasks of the user. A plausible reason why supervised or active learning are not used in 3D reconstruction yet, might be the high dimensional solution space, to which the learning algorithms must be adapted.

2.4 Discussion

What all activities have in common is that reconstructed 3D objects or scenes are monolithic and not interpreted in terms of parts, subparts and more deeper details. Depending on the scope, monolithic results are sufficient for tasks like indoor navigation [21, 10], reconstruction of urban outdoor environments [22] or reconstruction of human body as a deformable collection of objects [3], but not quite sufficient in creating detailed models with composable subparts for, e.g., replicas. Current reconstructed 3D models are only a kind of a solid monolithic geometry without subparts. As consequence these 3D models cannot be applied to many use cases like simulation tasks with these models might not be meaningful because the interrelationships between the subparts of the object are missing.

3 Interactive 3D reconstruction

Are the existing techniques sufficient to build 3D replica from images? Can these techniques be applied by a non-expert user? Is special hardware required? Does the user need a rotating plate, laser, stereo camera or even a main frame computer? The answer to these questions is mostly *No*. These and other questions, together with the discussed considerations, are the starting-point for our proposed research guideline. A major question is: “How can I get models of real world objects in such a way that I will be able to translate them back to real world replicas?”. In the following an overview of the relevant research questions concerning this problem is given.

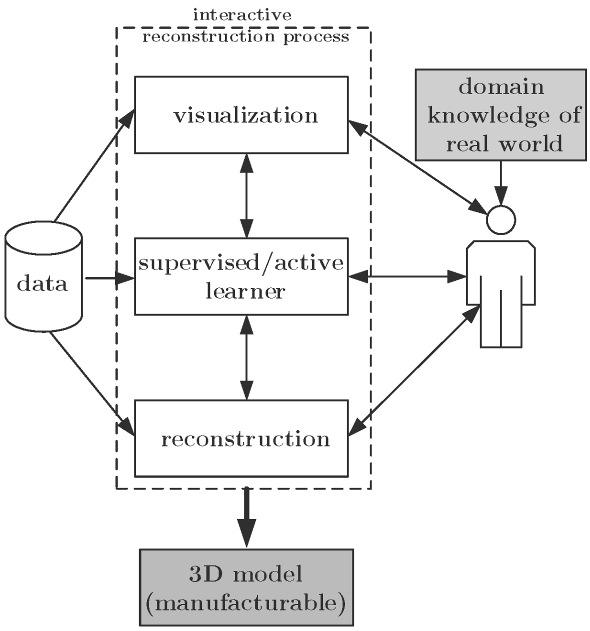


Figure 1 Interactive reconstruction architecture.

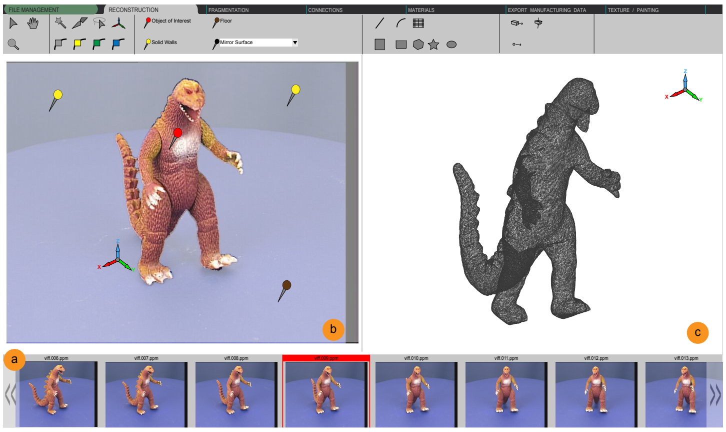
In an early stage of the architectural design, it became obvious that due to hidden information like material attributes, the fragmentation of objects or scenes, and the determination of degrees of freedom, deep real world knowledge must be present. The necessary real world knowledge cannot be explicitly integrated in existing algorithms due to the sheer complexity it would create, e.g., for material detection. Interactive modeling techniques are, however, able to integrate such complex domain knowledge iteratively.

A possible setup for such an interactive architecture is shown in Figure 1. The architecture consists of three main parts: i) the input data in the form of monocular or stereo images / video footage, as well as additional data like physical interrelationships – gravity, reflection etc.–, ii) the interactive reconstruction process and iii) the user as domain expert [26]. The interactive reconstruction process should join the computational power of today's computers with the conceptional knowledge of the user in order to solve at present computationally unfeasible issues. Thus, the computer remains the "work horse" of the process, while supervised and active learning algorithms translocate the load from the user to the computer, if possible.

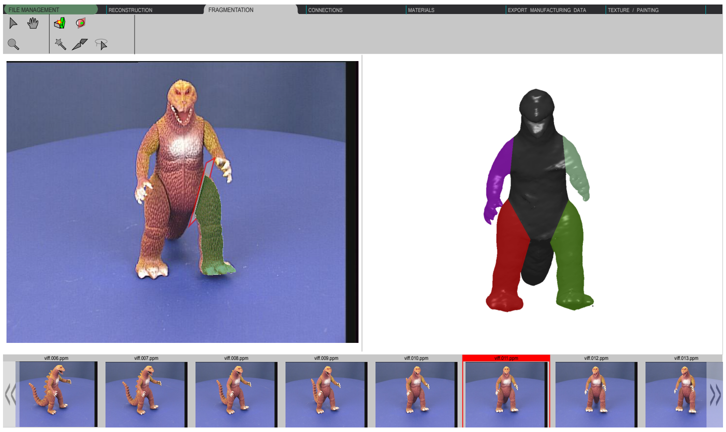
To achieve a better understanding of already existing techniques, algorithms, methods and processing pipeline, I designed a mock-up application, without underlying functionality, which is shown in Figure 2. For presentation purposes, the interactive reconstruction process is described using the Oxford "dinosaur" image collection [33], which already can be automatically reconstructed in high quality due to sufficient and overlapping images. However, all known reconstruction methods – automatic or semi-automatic – create only a monolithic or a skeleton model, not a model consisting of subparts. Throughout the following discussion, normal user interaction techniques like undo, redo, zoom can be assumed to be available.

3.1 Reconstruction and Fragmentation

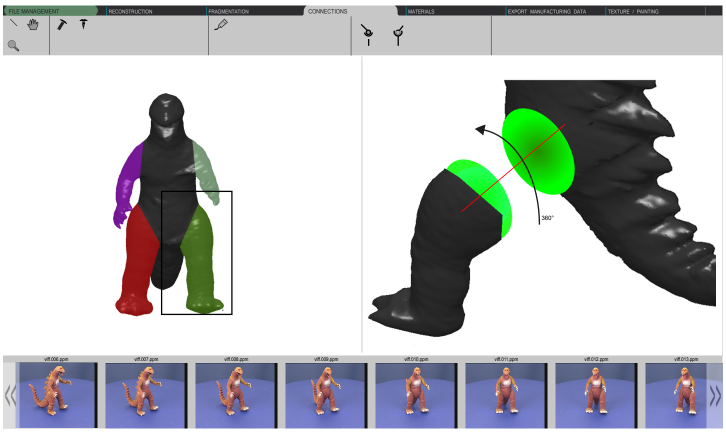
Based on an image or keyframe from a storyboard ⓐ the user may identify objects of interest with a marker pin. This triggers the calculation of a point cloud [25], which is then presented



a Reconstruction view, interactive creation of a 3D model out of monocular or stereo images or video footage.



b Fragmentation view, breaking the monolithic 3D model down into its parts.



c Connection view, adding detachable or permanent connections and theirs degrees of freedom.

Figure 2 Mock-up of interactive 3D Reconstructor – Several subfigures.

to the user. For fine trimming etc. common tools like scalpel, marker for borders and edges, and scribble tools can be used.

The shared view of the pre-processing ⓑ and the post-processing window ⓒ, as shown in Figure 1a, enable the user to directly modify the point cloud or the meshed point cloud. This allows the user to add missing or occluded information to the 3D model.

After the monolithic 3D model is reconstructed with a desired level of detail, the next step is to break down the model to its components or subparts until every subpart itself is monolithic, as shown in Figure 1b.

3.2 Connections and Materials

If all subparts have been identified the user has to model the connections between these parts. This step has to account for specific details like the type of connection, e.g., a ball joint, as well as specific information like extra allowance, rotation axis and maximal rotation angle, as illustrated in Figure 1c.

The assignment of materials to each subpart is, in general, straightforward. However, an automatic consistency check should be included to ensure the compatibility of connection types and materials.

3.3 Exporting manufacturing data

The last step in creating a replica is the re-transformation to the real world. This is done using 3D printers or CNC-machines. Since a variety of online services exist, the actual presence of such a machine is no requirement. After determining the dimensions for scaling the model, the export of the manufacturing data can be performed automatically without further user interaction. The replica itself is manufactured out of the assigned materials. In case of the dinosaur a painting or a texture would be necessary. How this manufacturing step can be performed without standardized hardware is, at the moment, difficult to assess. Of course, possible solutions like printing the texture to self-adhesive foil exist. Because painting and texture are really the last step of creating replicas and it is not necessary in lots of cases, this topic is not considered in detail yet. Next to the exporting manufacturing data, exports for CAD applications like simulation and modeling tools are provided. Using the export to CAD function provides engineers and architects with models, which they would call a model.

4 Application

The creation of replica is the main focus of this approach. This might enable an improvement in biological-inspired technologies, thus creating replica of plants and animals from images. Also an improvement for reverse engineering tasks, or an improvement for rapid prototyping out of mock-ups, or simply sharing sculpture with other artists is possible.

Under consideration of a 3D model with all parts and their physical coherence many applications might benefit. The reconstruction of scenes [21] can be used for robot navigation or for planning and executing relocations.

5 Conclusion

The literature review of 3D reconstruction and the discussion of our mock-up application point out the main weakness of current methods – the impossibility of creating non-monolithic 3D models. To overcome this weakness I started to implement the proposed semi-automatic and interactive architecture for image-based reconstruction as software prototype. I expect the identification of even more weaknesses of current 3D reconstruction and computer vision methods in the course of research conducted in the proposed directions. Finally I am optimistic an interactive 3D reconstruction tool could create CAD-ready models of real world objects which can be translated back to the real world with 3D printers and CNC-machines.

References

1 A. Aouina, M. Devy, and A. M. Hernandez. 3D modeling with a moving tilting laser sensor for indoor environments. *International Federation of Automatic Control World Congress*, 2014.

2 M. Arikan, M. Schwärzler, S. Flöry, M. Wimmer, and St. Maierhofer. O-Snap: Optimization-based snapping for modeling architecture. *ACM T GRAPHIC*, 32(1):1–15, 2013.

3 A. Barmpoutis. Tensor body: Real-time reconstruction of the human body and avatar synthesis from RGB-D. *IEEE Transactions on Cybernetics*, 43(5):1347–1356, 2013.

4 D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen. iCoseg: Interactive co-segmentation with intelligent scribble guidance. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010.

5 S. Branson, C. Wah, F. Schroff, B. Babenko, P. Welinder, P. Perona, and S. Belongie. Visual recognition with humans in the loop. In *European conference on Computer vision (ECCV)*, pages 438–451. Springer, 2010.

6 T. Chen, Z. Zhu, S.-M. Shamir, A.and Hu, and D. Cohen-Or. 3-Sweep. *ACM T GRAPHIC*, 32(6):1–10, 2013.

7 Y. Cui, S. Schuon, D. Chan, S. Thrun, and C. Theobalt. 3D shape scanning with a time-of-flight camera. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 1173–1180. IEEE, 2010.

8 P. E. Debevec, C. J. Taylor, and J. Malik. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. *Computer graphics and interactive techniques - SIGGRAPH*, 1996.

9 P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox. RGB-D mapping: Using kinect-style depth cameras for dense 3D modeling of indoor environments. *INT J ROBOT RES*, 31(5):647–663, 2012.

10 P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox. RGB-D mapping: Using depth cameras for dense 3D modeling of indoor environments. In *Experimental Robotics*, pages 477–491. Springer, 2014.

11 B. Höferlin, M. Höferlin, D. Weiskopf, and G. Heidemann. Scalable video visual analytics. *Information Visualization Journal*, 2013.

12 N. Kholgade, T. Simon, A. Efros, and Y. Sheikh. 3D object manipulation in a single photograph using stock 3D models. *ACM T GRAPHIC*, 33(4), 2014.

13 K. Konolige. Projected texture stereo. In *IEEE Robotics and Automation (ICRA)*, pages 148–155. IEEE, 2010.

14 A. Kowdle, Y.-J. Chang, D. Batra, and T. Chen. Scribble based interactive 3D reconstruction via scene co-segmentation. In *IEEE Image Processing (ICIP)*, pages 2577–2580. IEEE, 2011.

15 A. Kowdle, Y.-J. Chang, A. Gallagher, D. Batra, and T. Chen. Putting the user in the loop for image-based modeling. *INT J COMPUT VISION*, 108(1-2):30–48, 2014.

16 Microsoft Research. 3D surface reconstruction `http://research.microsoft.com/en-us/projects/surfacerecon/`, Nov 2014.

17 P. Musialski, P. Wonka, D. G. Aliaga, M. Wimmer, L. Gool, and W. Purgathofer. A survey of urban reconstruction. In *Computer Graphics Forum*, volume 32, pages 146–177. Wiley Online Library, 2013.
18 R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon. KinectFusion: Real-time dense surface mapping and tracking. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 127–136, 2011.
19 J. Ning, L. Zhang, D. Zhang, and C. Wu. Interactive image segmentation by maximal similarity based region merging. *Pattern Recognition*, 43(2):445–456, 2010.
20 Q. Pan, G. Reitmayr, and T. Drummond. ProFORMA: Probabilistic feature-based on-line rapid model acquisition. *British Machine Vision Conference (BMVC)*, 2009.
21 G. Pintore and E. Gobbetti. Effective mobile mapping of multi-room indoor structures. *The Visual Computer*, 30(6-8):707–716, 2014.
22 M. Pollefeys, D. Nistér, J.-M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S.-J. Kim, P. Merrell, and et al. Detailed real-time urban 3D reconstruction from video. *INT J COMPUT VISION*, 78(2-3):143–167, 2008.
23 M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch. Visual modeling with a hand-held camera. *INT J COMPUT VISION*, 59(3):207–232, 2004.
24 L. Quan, P. Tan, G. Zeng, L. Yuan, J. Wang, and S. B. Kang. Image-based plant modeling. In *ACM T GRAPHIC*, volume 25, pages 599–604. ACM, 2006.
25 J. Schöning and G. Heidemann. Evaluation of multi-view 3D reconstruction software. In *Computer Analysis of Images and Patterns (CAIP)*. In Press, Sep. 2015.
26 J. Schöning and G. Heidemann. Interactive 3D modeling - a survey-based perspective on interactive 3D reconstruction. In *Pattern Recognition Applications and Methods (ICPRAM)*, volume 2, pages 289–294. SCITEPRESS, 2015.
27 N. Snavely, St. M. Seitz, and R. Szeliski. Photo tourism: Exploring photo collections in 3D. *ACM T GRAPHIC*, 25(3):835, 2006.
28 P. Tanskanen, K. Kolev, L. Meier, F. Camposeco, O. Saurer, and M. Pollefeys. Live metric 3D reconstruction on mobile phones. In *IEEE Computer Vision (ICCV)*, pages 65–72. IEEE, 2013.
29 S. Thrun, W. Burgard, and D. Fox. A real-time algorithm for mobile robot mapping with applications to multi-robot and 3D mapping. In *IEEE Robotics and Automation (ICRA)*, volume 1, pages 321–328. IEEE, 2000.
30 M. van den Bergh and L. van Gool. Combining RGB and ToF cameras for real-time 3D hand gesture interaction. In *IEEE Workshop on Applications of Computer Vision (WACV)*, pages 66–72. IEEE, 2011.
31 A. van den Hengel, A. Dick, T. Thormählen, B. Ward, and P. H. S. Torr. VideoTrace:rapid interactive scene modelling from video. *ACM T GRAPHIC*, 26(3):86, 2007.
32 A. van den Hengel, R. Hill, B. Ward, and A. Dick. In situ image-based modeling. *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2009.
33 Visual Geometry Group, University of Oxford. Multi-view and oxford colleges building reconstruction - dinosaur `http://www.robots.ox.ac.uk/~vgg/data/data-mview.html`, Aug. 2014.
34 M. Zollhöfer, M. Nießner, S. Izadi, C. Rehmann, C. Zach, M. Fisher, C. Wu, A. Fitzgibbon, C. Loop, C. Theobalt, and M. Stamminger. Real-time non-rigid reconstruction using an RGB-D camera. *ACM T GRAPHIC*, 33(4):1–12, 2014.