Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Iwer Petersen

Survey: Real-time 3D model reconstruction

Fakultät Technik und Informatik
Studiendepartment Informatik

Faculty of Engineering and Computer Science
Department of Computer Science

Iwer Petersen

Survey: Real-time 3D model reconstruction

Submitted at: March 10, 2014

Iwer Petersen

Title of the paper
Survey: Real-time 3D model reconstruction

Keywords
video projection mapping, real-time 3d reconstruction

Abstract
In this survey 3D scene generation from real-life objects is discussed for the purpose of achieving a mapped video projection onto moving and deforming objects.
Iwer Petersen

Thema der Arbeit
Echtzeit-rekonstruktion von dynamischen 3D Objekten: Eine Übersicht

Stichworte
Videoprojektions-Mapping, Echtzeit 3D Rekonstruktion

Kurzzusammenfassung
In dieser Übersicht werden Echtzeit 3D Modellierungsverfahren untersucht um Videoprojektions-Mapping auf sich bewegende und verformende Objekt zu ermöglichen.

Contents

List of Figures

1 Introduction

1.1 Video Projection Mapping

Video Projection Mapping is an visual art form originating from visual-jockey circles. From the technical side it can be closely compared to the field of augmented reality. The basic idea is to project virtual textures onto real objects in a way that it appears as if the lit object is a video screen itself. Therefore the projection has to be limited to fit the object precisely.

In earlier day this was done by applying aligned black mask to the texture, or by rendering the texture onto an aligned 2D polygon. The aligned of the mask or polygon was mostly done manually using the projector in its final position by "re-drawing" the object in projector space.

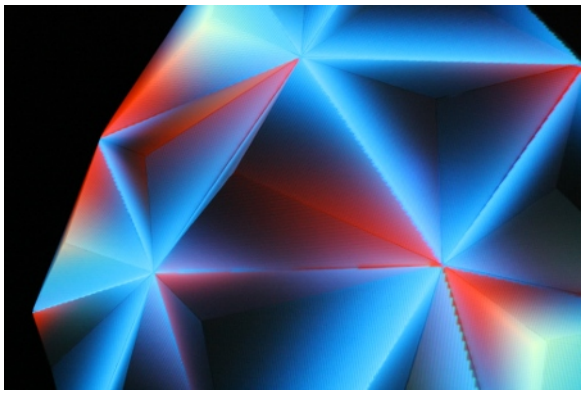
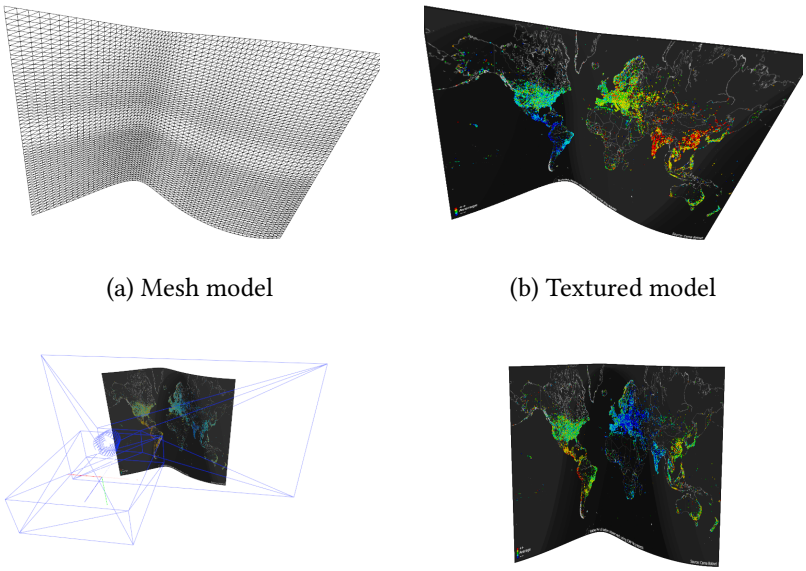


Figure 1.1: 3D Video Projection Mapping: Projection of Textured 3D model overlaying Object (CRYSTALCHOIR by Andrea Sztojánovits, Gábor Borosi)

But video projection mapping did not stay in two dimensional image space, but moved on to the third dimension. This more sophisticated method incorporated 3D modelling and rendering technology to create more advanced optical effects in the projected scenes like virtual lighting (see figure 1.1) and shadow casting.

The object to be video mapped is reconstructed as 3D mesh model (1.2a) which can be textured with the desired graphics (1.2b). This 3D object can be rendered from any viewpoint in 3D space. For a precise mapping of the projection onto the object not only the viewpoint in

3D space has to be chosen correctly. Also the intrinsic calibration of the virtual camera, which projects the 3D model to a 2D image, has to match the one of the projector closely (1.2c). The projector then basically applies the inverse projection to the rendered 2D image and overlays the real object with its 3D model (1.2d).



(a) Mesh model

(b) Textured model

(c) Virtual camera placement

(d) Rendered projector view

Figure 1.2: 3D video projection mapping principle

Traditionally objects to be video mapped are static objects like sculptures or buildings which are lit by projectors in fixed locations. The idea behind this survey is to expand the capability of such a system to be able to map 3D video projection onto people or other deforming object like kinetic sculptures.

First attempts have been made to move on to dynamic objects by synchronously moving the 3D model by transforming the 3D scene and the real object by actuating it with stepper motors (see panGenerator collective [1], White Kanga [2]). Another methods uses marker-based tracking to find the objects position using marker-based tracking and resemble it in virtual 3D space (see Kato and Billinghurst [3], Yapo et al. [4]). With advanced 3D tracking methods it is possible to spare the visible markers (see Perez et al. [5], Petersen [6]).

However all those methods only work with rigid, non-deforming objects which are moving in space. This is a much simpler task because the 3D model of the object is constant, and can possibly be created in advance. To take into account that the object can deform over time, the model has to be adapted constantly to the changes in the real scene.

1.2 What is needed?

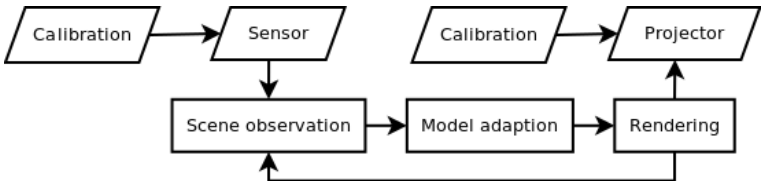


Figure 1.3: Basic workflow for dynamic 3d video projection mapping

As stated in the previous section a dynamic mesh representation is needed which is constantly adapted to the changes in the scene. This can be seen as repetitive 3D scanning the scene. The real scene has to be observed with some kind of sensors which are typically cameras or other optical sensors. According to the observation the 3D model must be adapted to match the changes in the scene. Then the changed 3D scene can be rendered and send to a projector. Figure 1.3 depicts this basic operating principle. Key is that the this repetitive process can be executed at interactive frame-rates to minimize the offset between the projection and the real object.

Of course there is another task to handle when it comes to render a texture onto the created mesh. Beside the raw 3D mesh model, a two-dimensional texture space has to be calculated to map a texture to the adapted mesh. Therefore every mesh vertex has to be added a 2D texture coordinate which defines the mapping of a 2D texture to the 3D mesh. Because in an evolving mesh vertices can emerge and disappear, this texture space has to be updated also in the model adaption step. This aspect is also a challenging task but is out of the scope of this survey.

This paper examines the current research state in terms of realtime 3D reconstruction methods. In the following, several methods for 3D reconstruction in general and real-time capable methods in particular will be reviewed for the described application.

2 3D Scanning: Static objects

In general optical sensors used for 3D reconstruction can be categorized by the need of a controlled lighting source. Casual cameras only rely on ambient lighting as illumination source and therefore stereoscopic vision systems are categorized as passive scanning method. System that use controlled illumination like laser scanners or structured light scanning systems are categorized as active scanning method (see Lanman and Taubin [7, chap. 1]). In this chapter some basic ideas of 3D reconstruction methods using active and passive scanning are presented.

2.1 Reconstruction using passive scanning

A popular passive scanning method uses multiple cameras to observe the scene, and feature descriptors to find corresponding points in the images. Those points are originating from somewhere on lines in 3D space. By intersecting the lines of corresponding points from different cameras the position of those points in 3D space can be reconstructed. Hartley et al. [8] describe this method in detail. For a precise calculation, it is necessary, that the cameras are calibrated and their intrinsic and extrinsic parameters are known. These can be acquired for example with Zhang's method (see [9]).

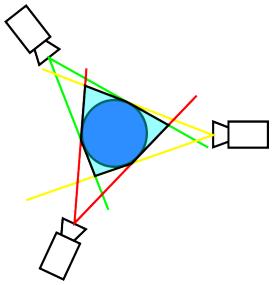


Figure 2.1: Topview: Intersecting 3D projections of extracted silhouettes

2.1.1 Reconstruction of Volumetric Voxelmodel

Cheung et al. [10] presented a 3D voxel reconstruction method using five cameras that constructs rough 3D models in realtime. The idea is to extract the moving objects silhouettes from the images of the cameras and to intersect the 3D projection cones of those silhouettes in 3D space to approximate the objects shape. In figure 2.1 three cameras observing a cylindrical object (blue). The reconstructed model (light blue) results from the intersecting 3D projections (red, green, yellow) of the extracted silhouettes. The system was created for analysis of human motion and did not need an exact mesh model.

2.1.2 Reconstruction with kinematic model

A model based approach of human body pose estimation is discussed by [11]. The approach relies on on a model of the human body, which consists of geometric primitives like cylinders and ellipsoids which are connected at skeletal joints. For each of the joints the freedom of movement is defined in a kinematic model. To reconstruct the movement of a human, the silhouette from one or more cameras is extracted and compared to a perspective projection of the shape model. With the kinematic model the joint poses can than be adapted to the changes in the scene.

2.2 Reconstruction using active scanning

Active scanning techniques involve controlled lighting of the scanned scene. Depending on the material of the scanned object active scanning is more sensitive. Active scanning methods often resemble a stereo camera system by replacing one of the cameras with a controlled light source. In that way the problem of estimating correspondences between the two camera images can be circumvented (see [7, chap. 1]).

2.2.1 Laser Scanner

The first Laser scanner where build in the 1970's and where projecting a single moving laser point onto the scene which was then recorded with a digital camera. The 3D shape is then Later this rather slow method was improved by sweeping a planar sheet of light over the scene, which shows in the camera image as a distorted line on the object (see figure 2.2). An inexpensive method of swept-plane scanning can be implemented by using the shadow of a stick as sweeping plane (see [7, chap. 4]).

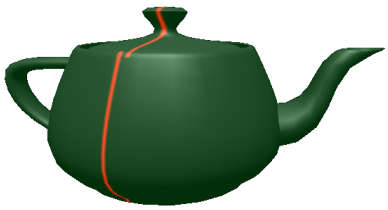


Figure 2.2: Illustration of a laser-scanned Teapot

2.2.2 Structured Light

With digital projectors the mechanical motion of the laser can be eliminated. But more than that, the capabilities of a projector for projecting arbitrary coloured patterns can be exploited to further improve the swept-plane scanning concept. A projector can project multiple lines simultaneously and reduce the scanning time needed for a laser to sweep the scene. By projecting a series of binary structured gray code images

(see figure 2.3) the resolution of the scan result can be drastically improved. Structured light

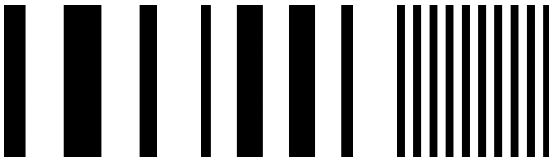


Figure 2.3: Binary coded structured light sequence

scanning has been intensively researched over the last years which led to very sophisticated algorithms (see [12]) and non-visible structured light scanning using infra-red (see section 2.2.3).

2.2.3 Pointclouds

While above methods all work on the cameras 2D images, recent developments more and more consider point clouds as common intermediate datatype for 3D scanning and further processing tasks. As 3D points can be generated from any multiview based scanning technique and more and more algorithmic solutions for three-dimensional problems are developed. Rusu [13] explains a whole universe of algorithms for 3D reconstruction, object detection and scene interpretation based on 3D point clouds.

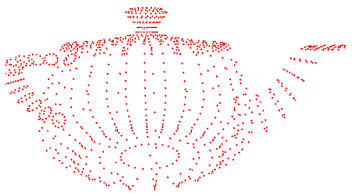


Figure 2.4: Point cloud representation of a Teapot

Depth Cameras

The launch of depth cameras for the gaming industry brought a revolutionary cheap 3D capable sensor to the market. Although intended for human pose detection for computer games, coloured 3D point clouds can easily be generated from the color and depth image (see figure 2.5) delivered from for example PrimeSense [14] based depth cameras.
These sensors basically resemble the structured-light idea with the distinctive feature, that they project a pattern with infra-red light, which is then picked up by an infra-red camera. In that way visible light is not interfering with the structured light scanning process. This feature is vital for the intended application, as it is explicitly wanted to artificially project possibly changing light onto the scanned object

Figure 2.5: Typical depth image. Depth information is encoded in intensity values (white: close, black: far)

which would most likely interfere with active scanning techniques using visual lighting.
Smisek et al. [15] did a geometrical analysis of the Microsoft Kinect depth camera and propose a calibration procedure to get the color and depth image aligned properly. With a calibrated Kinect one can generate point clouds from the depth images pixels by take a pixels x and y coordinates and map the color value of the pixel to the z coordinate. The color of the resulting 3D point is defined by the corresponding pixel on the color image.

2.3 Discussion

The research on 3D shape scanning has developed sophisticated methods to acquire a 3D model of a real-life object. Passive reconstruction methods seem to be more or less outdated while active reconstruction methods go in different directions depending on the purpose. While laser scanner are often chosen for industrial 3D reconstruction applications like autonomous vehicles, the appearance of cheap depth cameras enabled low budget 3D scanning for everyone.

This chapter focussed on 3D reconstruction in general. In the next chapter several approaches for real-time reconstruction will be proposed. All of those methods optimize ideas from this chapter to achieve real-time 3D reconstruction.

3 Real-time 3D reconstruction: Dynamic objects

Recent research projects have come up with innovative approaches to take 3D scanning to the next level by achieving real-time reconstruction. New sensors like depth cameras as well as proceedings in parallel computation led to faster 3D model acquisition time. Following examples reached the point that repetitive scanning at interactive frame rates is possible.

3.1 KinectFusion - Volumetric integration

Izadi et al. [16] gained a lot of attention with their work on KinectFusion at Microsoft Research. They presented a system, which can create a 3D model of the environment at interactive frame rates by waving a single Kinect depth camera around the scene. The system tracks the cameras position with respect to the captured scene and constantly integrates the depth images into a volumetric model which is represented by a signed distance function (see Curless and Levoy [17] for details). Mesh errors through sensor noise or occlusion are gradually corrected by averaging stored and measured distance function. The integration is processed on a GPU with the signed distance function residing in the graphics cards memory. Although this allows very fast computation, the memory requirements of the three-dimensional data structure grows rapidly with expanding scanning volume. Also the resulting mesh is not smoothly warping when the scene is changing, due to the averaging over time. When an object is moved in the scene during the scanning process, it more fades away from the old position while appearing at the new position.

3.2 Multi-Kinect real-time reconstruction

Using multiple depth cameras it is possible to cover a scanning volume from multiple directions reducing the need for past frame data to reconstruct a full 3D model. Tong et al. [18] created a 3D scanning system for humans to acquire precise, static 3D models by installing three depth cameras in different heights around a turning platform. Although they still capture multiple frames while the human on the table is turning, they already fuse the data of three sensors for every frame. In that way it is possible to maximise the achieved resolution by getting closer to the sensor which improves the triangulation error. Alexiadis et al. [19] took a similar approach by placing multiple depth cameras around the scene to cover a certain volume from all directions. As they work towards a multi-user 3D environment with realistic representations of the users, dynamic mesh representations are vital. Both solution are realized using point cloud algorithms described in detail by Rusu [13].

3.3 Real-time structured light reconstruction

An advanced high speed structured light scanning method that work in real-time is presented by Ide and Sikora [20]. They assemble two scan unit which each consists of a Video Projector, a $500Hz$ high speed camera and two color cameras. The cameras are synchronized with the video signal which is sent to the projectors. In that way the projectors of the two scan units

can alternating project patterns which are captured by the high speed cameras and used for 3D reconstruction. The four color cameras then provide color information for reconstructed 3D model. This methods produces very high detailed meshes with about $5.2M$ vertices at about $10Hz$.

3.4 Modelbased Deformation Tracking

Jordt and Koch [21] effectively decouple the reconstruction of complex surface geometry details from the complexity of the deformation. Therefore they initially capture a detailed surface model and register a spline surface to this model. An optimization algorithm then adapts the control points of the spline surface to match the objects appearance in color and depth images. With this inverse mapping of a model to the measured data, this method show similarities to the method proposed in section 2.1.2.

3.5 Discussion

Depending of the desired target application current real-time reconstruction methods are specialized in different ways. With pure reconstruction of static scenes in mind, a method like 3.1 are perfectly suitable. However it does not provide a constantly time-varying mesh which would be needed for the goal pursued with this paper. While a method like 3.3 provide very precise 3D meshes in fast frame rates, it requires a more advanced and expensive hardware set-up. 3.4 reduces the complexity of the reconstruction by assuming the detailed surface of the reconstructed object roughly stay constant. However in case of video mapping onto people for example, the system to create should be able to reconstruct the wrinkles of the closing frame by frame as closely as possible. A method like 3.2 therefore seems to be the right direction to go as it makes use of inexpensive sensors, implements a one-frame reconstruction process which should provide most up-to-date mesh data for a close mapping of the real scene.

4 Conclusion

In this paper the state of the art of 3D mesh reconstruction in real-time was reviewed for the purpose of dynamic 3D video projection mapping. The requirements for the targeted

application led to the need of a 3D mesh created in real-time. To understand 3D model generation in general, the methods and techniques for 3D scanning where examined. With this basic understanding different recent approaches of real-time 3D reconstruction where analyzed.

Coming from 3D scanning technology, real-time 3D reconstruction is and has been an intensively researched topic in the fields of computer vision, computer graphic, virtual reality and multimedia. New sensors and algorithms as well as parallel computing are the key technologies that enable new methods for real-time 3D reconstruction. The proceedings of conferences like the international Conference on Computer Vision and ACM SIGGRAPH will therefore be subject of further research. Special attention will go to the results of the 3D human reconstruction and action recognition Grand Challenge during the ACM Multimedia '13, which promises to come up with novel ideas.

The next step in the authors work will be, to start to implement the targeted solution with respect to the pure 3D mesh reconstruction using multiple depth cameras and point cloud processing algorithms. Further the texture-mapping problem, which was omitted for this paper, will need to get closer attention.

Bibliography

[1] panGenerator collective. Peacock. URL http://vimeo.com/49869407. last accessed: 20.07.2013.

[2] White Kanga. Modeling projection system. URL http://whitekanga.pl/. last accessed: 20.07.2013.

[3] H. Kato and M. Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality, 1999.(IWAR'99)*, pages 85–94, 1999.

[4] T.C. Yapo, Y. Sheng, J. Nasman, A. Dolce, E. Li, and B. Cutler. Dynamic projection environments for immersive visualization. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1–8, 2010.

[5] Guillermo Perez, German Hoffmann, Rodrigo Rivera, and Veronica Manduca. Mapinect. URL http://mapinect.wordpress.com/. last accessed: 20.07.2013.

[6] Iwer Petersen. Using object tracking for dynamic video projection mapping, 2012.

[7] Douglas Lanman and Gabriel Taubin. Build your own 3D scanner: 3D photography for beginners. In *ACM SIGGRAPH 2009 Courses*, SIGGRAPH '09, pages 8:1–8:94, New York, NY, USA, 2009. ACM. doi: 10.1145/1667239.1667247. URL http://doi.acm.org/10.1145/1667239.1667247.

[8] R. Hartley, A. Zisserman, and Inc ebrary. *Multiple view geometry in computer vision*, volume 2. Cambridge Univ Press, 2003.

[9] Z. Zhang. A flexible new technique for camera calibration. 22(11):1330–1334, 2000.

[10] G.K.M. Cheung, T. Kanade, J. Y Bouguet, and M. Holler. A real time system for robust 3D voxel reconstruction of human motions. In *IEEE Conference on Computer Vision and Pattern Recognition, 2000. Proceedings*, volume 2, pages 714–720 vol.2, 2000. doi: 10.1109/CVPR.2000.854944.

[11] R. Poppe. Vision-based human motion analysis: An overview. 108(1):4–18, 2007.

[12] S. Zhang. Recent progresses on real-time 3D shape measurement using digital fringe projection techniques. 48(2):149–158, 2010.

[13] Radu Bogdan Rusu. Semantic 3D object maps for everyday manipulation in human living environments. 2009.

[14] PrimeSense. Primesense. URL http://www.primesense.com/. last accessed: 20.07.2013.

[15] J. Smisek, M. Jancosek, and T. Pajdla. 3D with kinect. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 1154–1160, 2011. doi: 10.1109/ICCVW.2011.6130380.

[16] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, and Andrew Fitzgibbon. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, UIST '11, pages 559–568, Santa Barbara, California, USA, 2011. ACM. ISBN 978-1-4503-0716-1. doi: 10.1145/2047196.2047270. URL http://doi.acm.org/10.1145/2047196.2047270.

[17] B. Curless and M. Levoy. A volumetric method for building complex models from range images. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 303–312, 1996.

[18] J. Tong, J. Zhou, L. Liu, Z. Pan, and H. Yan. Scanning 3d full human bodies using kinects. 18(4):643–650, 2012.

[19] D.S. Alexiadis, D. Zarpalas, and P. Daras. Real-time, full 3-d reconstruction of moving foreground objects from multiple consumer depth cameras. 15(2):339–358, 2013. ISSN 1520-9210. doi: 10.1109/TMM.2012.2229264.

[20] K. Ide and T. Sikora. Real-time active multiview 3D reconstruction. In *2012 International Conference on Computer Vision in Remote Sensing (CVRS)*, pages 203–208, 2012. doi: 10.1109/CVRS.2012.6421261.

[21] Andreas Jordt and Reinhard Koch. Direct model-based tracking of 3D object deformations in depth and color video. 102(1-3):239–255, March 2013. ISSN 0920-

5691. doi: 10.1007/s11263-012-0572-1. URL http://dx.doi.org/10.1007/s11263-012-0572-1.