AUTOMATION AND COMPUTER SCIENCE DEPARTMENT

Eng. Daniel Moldovan

PhD. Thesis
(ABSTRACT)

3D MODELING TECHNIQUES FOR ULTRA-REALISTIC SYSTEMS

SUPERVISOR,
Prof.dr.eng.Sergiu Nedevschi
Abstract

The purpose of this thesis is to present my research contributions in the field of 3D modeling techniques suitable for ultra-realistic systems. In this research I employed both active and passive computer vision techniques for generating 3D models of the surrounding environment and of small objects, respectively.

Initially, I investigated a reconstruction method that belongs to passive vision that was using an omnidirectional imaging system. Assuming a single viewpoint and properly aligned catadioptric camera, I introduced a virtual omnidirectional pinhole camera model that is able to (1) display undistorted images on a cylindrical virtual screen and to (2) detect the correct location of the viewing point. Cylindrical images generated in this way can be used as input for an ultra-realistic environment that exhibits a cylindrical screen.

Once the correctness of the optical center's location was confirmed, I performed a rough 3D reconstruction of the surrounding environment by estimating the depth at which projections from two viewing points on a planar screen had the highest matching coefficient. The approach was based on the idea that color histograms of projections coming from two optical centers would match on the object's surface. The reconstructed 3D structure can then be used in order to represent the background of an ultra-realistic environment augmented with large size 3D displays.

In the second part of my work I analyzed a method of automatic reconstruction of the 3D model of a real object based on a sequence of range views provided by a scanner with an embedded camera. This method employs first the removal of the false 2D matches and then proceeds with the detection of 3D matches by using the one to one correspondences provided by the scanning device between the 3D points and their corresponding 2D image projections. Due to its ability to detect the matching points in a fast manner, this method is suited for applications characterized by a fast 3D reconstruction demand.

In order to present a real-life example of an ultra-realistic environment, final part of this thesis deals with the integration between the 3D reconstruction method and the visualization on an auto-stereoscopic display. A method to compute the maximum displayable resolution for the selected auto-stereoscopic display is also provided.

1. Motivation

Traditionally, most of the computer vision techniques have been based on 2D images and the subsequent analysis of purely 2D data. However, because it is recognized that 3D data offers a richer, less abstract and more useful source of information, an increasing body of research has been directed towards generating 3D models of both the surrounding environment and of the real objects by using computer vision techniques.

Integration of these 3D models would allow in the future the reproduction of an ultra-realistic environment based on natural 3D images that will make you feel as though you were actually immersed in a real environment. Such systems would contribute to the development of new modes of communications that will allow people in remote locations to have a genuine sense of being in the same place at the same time. For realizing such a high-presence environment using 3D images we need 3D displays to allow multiple viewers to simultaneously observe large 3D images displayed in the background and small objects displayed in the foreground. The location of these
displays with respect to the viewers would have to be variable in order to accommodate different types of scenarios.

The recent spread of broadband networks, advanced digital technologies and improved performance in image acquisition and display has made possible the manufacturing of high-definition displays and 3D imaging systems at a reasonable cost. In order to generate 3D models for such displays we need to employ various 3D modeling techniques with performances characterized by the following criteria: (1) resolution of the generated 3D model, (2) purpose of use (in which part of the scene will the 3D model be displayed: foreground or background) and (3) scalability (the modeling technique can be used for reconstructing of only small objects or also for wider objects like 3D maps of indoor environments or buildings).

An idealized scenario of the ultra-realistic environment we envision is represented by the integration of the real and the virtual world in a seamless way. In such an environment, physical persons will be able to interact with different kinds of objects displayed on table-top 3D displays as well as with people located far away that will be represented as floating 3D images. The background of such an ultra-realistic environment will be represented by a circular display that will cover an area that is wide enough to give the feeling of immersion into the far away environment.

In this research I introduce two 3D reconstruction methods for (1) map building and (2) object digitization that can be employed in the creation of such an ultra-realistic environment. As such, the 3D map building method is characterized by a low resolution 3D reconstruction that is suitable for indoor and outdoor maps reconstructions with applicability in background 3D displays. The object digitization method on the other hand is characterized by a high resolution of the 3D reconstruction being suitable for small objects digitization with applicability in foreground 3D displays.

2. My Approach

In general, computer vision techniques can be grouped in two main categories: Active Vision and Passive Vision. Active vision refers to those techniques that use a controlled source of structured energy emission such as a scanning laser source or projected pattern of light and a detector, such as camera. The nature of the interaction of the reflected source with the object surface is then used to recover 3D data. Passive vision techniques on the other hand refer to the measurement of visible radiation that is already present in the scene. Their objective is to acquire images of a scene from different viewpoints and to compute scene shape at every surface point.

Although active sensing can facilitate the computation of scene structure, active approaches are not always feasible, especially for modeling distant or fast-moving objects. However, the best active methods employing laser range scanning generally produce more accurate reconstructions than is possible using passive techniques. In order to exploit the strengths of both types of sensing, in my work I employed devices belonging to both passive and active vision.

I started the research by focusing on the background layer of such an ultra-realistic system and proposed two methods for generating (1) cylindrical images as well as (2) omnidirectional 3D maps to be displayed on the screen placed in the background.

Both methods employed the generation of a virtual omnidirectional camera by using an omnidirectional 2D image and a calibration pattern. Cylindrical images have been generated by
performing a back-projection process on a virtual screen and the omnidirectional 3D map has been built in a process that inspected the surrounding environment by checking the matching degree of projections on a planar surface from two optical centers.

After completing the first phase I focused my work onto the next target of generating digitized 3D models of small objects that could be later displayed on 3D displays located in the near vicinity of the user. For this task I employed a laser scanner that provided both range views and 2D images with a one-to-one correspondence between 3D points and the pixels. By employing 2D feature matching techniques I was able to automatically register the range views taken from unknown locations and to generate in the end a full view 3D model.

3D data being the ultimate source of information, an application in which a reconstructed 3D model is sent simultaneously for display on a large variety of 3D displays that exhibit various resolutions can be regarded as an ultimate display system. In the final part of my thesis, I presented an initial case study of the heuristic framework for adaptive visualization on 3D displays by integrating the scanning and reconstruction of a high-resolution 3D model with the visualization on gCubik – an autostereoscopic display characterized by a full-view, real-time 3D display albeit at a low resolution.

3. **Structure of the thesis**

This thesis deals with (1) omnidirectional cameras and (2) perspective camera / laser scanner combinations in order to recover the 3D structure of both the surrounding environment and the small objects. A real-life example of an ultra-realistic system that integrates 3D object reconstruction method with the visualization on an auto-stereoscopic display is also presented.

**Chapter 1** presents the characteristics of the 3D modeling techniques tackled in this research. Integration of these 3D models would allow in the future the reproduction of an ultra-realistic environment based on natural 3D images that will make you feel as though you were actually immersed in a real environment.

A short description about the computer vision techniques related to 3D reconstruction and omnidirectional vision is subsequently presented in **Chapter 2**. Initially I introduce some geometric considerations about camera models as well as perspective projection and camera calibration and then I continue with a presentation about omnidirectional vision sensors that includes also a description about omnidirectional image formation process.

In order to perform a successful 3D reconstruction, the required elements are a precise location of the viewing point and a distortion free image plane. In **Chapter 3** I tackle the problem of internal camera calibration and introduce the concept of a virtual camera that can be emulated on both perspective and omnidirectional physical cameras. Subsequently, I present a calibration method that estimates virtual camera's optical center and removes the geometrical distortions that are present in the virtual image plane.

The calibration method of the virtual camera was based on the fact that for perspective systems the incoming light rays are projected directly onto the photo-sensitive elements through a single point called the effective pinhole. In my work, I applied this method for both dioptic and catadioptric imaging systems that exhibit a single viewpoint.

With the help of a calibration pattern and thus avoiding complicated mathematical models, the optical center was detected by tracking the optical rays generated by pairs of 3D points that have
the same projection on the image plane. The calibration pattern was also used in order to generate normalized virtual screens, on which captured images are back projected, a process that eliminates the deformations.

Once I confirmed the correctness of the optical center's location, I proceeded to the rough 3D reconstruction of the depth to the surrounding objects by employing the previously calibrated virtual omnidirectional camera. Chapter 4 details my approach that was based on the idea that color histograms of projections coming from two optical centers would match on the surface of the object. In the experimentation I matched color histograms of projections on planar screens placed at an initial distance from one of the optical centers. This distance was then gradually incremented outward into the surrounding environment.

Based on this method, by incrementing also the pan angle, a rough depth map could be generated. Besides the applicability for 3D display, this map can be used also by a robot to decide the direction of movement based on detection of free space. As such, once the robot moves, it can record sequences of image frames that can be processed in order to offer more detailed 3D information.

Chapter 5 introduces an algorithm to align unregistered 3D range views taken from unknown locations. The aligning process uses positioning information computed from 2D images that have a one to one correspondence with the depth values in the range views. My method robustly detects pairs of corresponding 2D features that are subsequently used for extracting 3D correspondences from range data files. These 3D correspondences are then used for estimating relative locations of the range views. Due to its ability to detect the matching points in a fast manner, this method is suited for applications characterized by a fast 3D reconstruction demand.

Chapter 6 presents an initial case study of a heuristic framework for adaptive visualization on 3D displays. I exemplify an ultrarealistic system by performing an integration of my 3D reconstruction method with the visualization on an auto-stereoscopic display characterized by a full-view, real time, albeit at a low resolution 3D display. In this framework, a 3D model is to be generated once at a high resolution and subsequently - in order to be visualized on a wide range of displays with different resolutions - it will have to undergo a transformation that will reduce its resolution accommodating thus the 3D data to the low resolution 3D display.

Finally, the conclusions of this thesis, the contributions of the entire research presented in the dissertation and the future works are discussed in Chapter 7. The publications resulted from the research are also listed in here.

4. Main contributions

The contributions that this thesis is bringing to the Ultra-Realistic Communications technology developed at National Institute of Information and Communications Technology (NICT) in particular and to the computer vision community in general are listed below:

1. Introduction of a virtual omnidirectional pinhole camera model that is able to (1) display undistorted images on a cylindrical virtual screen and to (2) detect the correct location of the optical center. This method uses two screens as virtual image planes, and the optical center is estimated as the converging point of the optical rays passing through the corresponding point pairs on these planes. This method has the advantages that 1) it can be applied to any single viewpoint camera, and 2) it can remove any type of distortions.
2. 3D scene modeling of an indoor environment. The scene is reconstructed by employing previously calibrated omnidirectional camera and by matching projections from two optical centers on to a planar screen. We implemented an algorithm that would reconstruct the 3D structure based on the idea that color histograms of projections coming from two optical centers would match on the surface of the object. By tilting and changing the position of the projection screen both horizontally and into the surrounding environment, an omnidirectional depth map could be generated.

3. Automatic reconstruction of the 3D model of a real object based on a sequence of range views provided by a scanner with an embedded camera. This method employs first the removal of the false 2D matches and then detects 3D matches by using the one to one correspondences provided by the scanning device between the 3D points and their corresponding 2D image projections. These 3D correspondences are then used for estimating relative locations of range views. Due to its ability to detect the matching points in a fast manner, our method is suited for applications characterized by a 3D fast reconstruction demand.

4. Integration of 3D reconstruction method with the visualization on an auto-stereoscopic display. I introduce an application for automated 3D modeling and visualization that scans, registers and displays real objects on a hand-held 3D display in an online process. In the experiments, I employ the gCubik display characterised by a full-view, real-time 3D display albeit at a low resolution. A method to compute the resolution of the 3D model for which gCubik reaches its maximum resolution of display is also provided.

5. Published papers

The results obtained from this research work were published in a journal, international conferences and workshops. A patent application has also been submitted to the Patent Office of Japan. The list of the most representative publications is presented in the following:

Journal


Patent


International Conferences


The research activity resulted also in software tools for (1) generation of distortion free image planes, (2) optical rays tracing, (3) automatic range views registration and (4) integration between the 3D reconstruction method and the visualization on gCubik.

The work developed in this thesis has been partially funded by the Japanese government through the following projects:

- National Institute of Information and Communications Technology (NICT) project for research on *Universal Communications* from April 2006 to March 2011.

The financial support needed for completing the research about omnidirectional camera calibration and 3D modeling of the surrounding environment has been received mainly from the MEXT and Wakayama University as follows:

- A three years PhD scholarship given by MEXT from April 2002 to March 2005.
- A two years Research Assistant position at Wakayama University from April 2003 to March 2005.