3DTV - PANORAMIC 3D MODEL ACQUISITION AND ITS 3D VISUALIZATION ON THE INTERACTIVE FOGSCREEN

Sven Fleck, Florian Busch, Peter Biber, Wolfgang Straßer WSI/GRIS, University of Tübingen Tübingen, Germany Ismo Rakkolainen

Stephen DiVerdi, Tobias Höllerer

FogScreen Inc. Espoo, Finland University of California Santa Barbara, USA

ABSTRACT

Future 3D Television critically relies on mechanisms for automatically *acquiring* and *visualizing* high quality 3D content of both indoor and outdoor scenes. The envisioned goal is that a photo-realistic 3D real-time rendering from the actual and potentially arbitrary viewpoint of the beholder who is watching 3DTV becomes possible. Such scenes include movie sets in studios, e.g., for talk shows, TV series and blockbuster movies, but also outdoor scenes, e.g., buildings in a neighborhood for a car chase or cultural heritage sites for a documentary. The goal of 3D model acquisition is to provide the 3D background models where potential 3D actors can be embedded. We present both the *3D acquisition* and semi-immersive *3D visualization* to give an impression how a future 3D Television system could be like.

Index Terms- TV, Stereo vision, Laser radar, Machine vision

1. INTRODUCTION – MODEL ACQUISITION FLOW AND VISUALIZATION

For 3DTV systems we present a platform with two distinct ways of model acquisition and one for visualization:

- Computer vision based 3D modeling: Graph cut based omnidirectional Stereo Vision
- 2. Laser range scanner based 3D modeling
- 3. 3D visualization on the interactive FogScreen

Our new platform comprises an 8 Mpixel omnidirectional camera C1 in conjunction with three laser scanners L1-L3 and no odometry is necessary. In our system, 3D models are acquired just by moving the platform around and recording all sensor data. The flexible mapping between sensors and 3D information (localization, geometry and texture) is illustrated in Fig. 1. To give an overview, both

	L1	L2	L3	C1		L1	L2	L3	C1
Loc.	X				Loc.	Х			
Geo.				X	Geo.		X	Х	
Text.				Х	Text.				Χ

Fig. 1. Left: Stereo vision flow, Right: Laser flow

the graph cut stereo pipeline and the laser scanner based modeling pipeline are illustrated in Fig. 2. The position and orientation of the platform is determined by probabilistic matching of laser scans L1 [1]. The stereo pipeline computes dense depth maps using pairs of panoramic images taken from different positions using a calibration



Fig. 2. Our 3D Modeling Architecture. Note that both the graph cut stereo flow and the laser scanner flow are included.

procedure relative to the platform. In contrast to classical multi camera based scene capturing techniques we require only one camera and sample the environment, which is much more inexpensive and is not limited in the size of the environment. The laser flow projects the data from L2 and L3 into space using the localization of L1. L2 and L3 together provide a full 360° vertical slice of the environment. The camera yields the texture for the 3D models.

This paper is organized as follows. Starting with the acquisition, we first describe the platform in section 2, the localization (section 3), the Stereo vision flow (section 4) and finally the laser modeling flow (section 5). Then we illustrate the 3D visualization of these models on the interactive FogScreen (section 6) before we conclude this paper.

2. OVERVIEW OF THE MOBILE SENSOR PLATFORM

We developed a mobile platform ("*The Wägele*"¹) that offers a very flexible sensor setup. We concentrate on the sensor aspect, the platform is moved around manually to acquire sensor data. However, it could also be mounted on a mobile robot to acquire environments

¹A Wägele - Swabian for a little cart

in a teleoperated way or on a pickup truck for outdoor acquisitions. In contrast to [2] our platform does not rely on odometry data and is thus a self-contained unit that aims at delivering high quality results. After a recording session the collected data is assembled to create a consistent 3D model in an automated offline processing step. First a 2D map of the scene is built and all scans of the localization scanner are matched to this map. After this step the position and orientation of the Wägele is computed at each time step. This data is then fed into the graph cut stereo pipeline or the laser pipeline.

Sensor Platform

The sensor acquisition platform is built using aluminium XC structural beams. This combines great stiffness with limited weight and allows for a very flexible setup. So it can be easily adapted to various sensor and scene requirements. Currently it comprises both an omnidirectional camera and three 2D laser range scanners in a self contained unit. It is mounted by just three screws on the basic cart for easy transportation and so composes "*The Wägele*" platform. Its sensors are calibrated relative to each other on this platform, which itself is very stiff and thus assumed static. See Fig. 3 for an overview. Fig. 4 shows our setup with the three laser scanners and the panoramic camera.



Fig. 3. Two setups of our mobile platform. Left: Two Laser scanners L1, L2 and one omnidirectional camera. Center & Right: Three laser scanners L1, L2, L3 and omnidirectional camera closely mounted together. L1: Laser Scanner (scanning horizontally) used for 2D mapping and self-localization. L2, L3: Laser Scanners (scanning vertically) for geometry acquisition of the "laser" model. They are offset, i.e., a slight overlap is achieved to get a full 360° scan field of view. C1: 8 Megapixel panoramic camera for texture acquisition.



Fig. 4. Wägele platform setup for both stereo & laser acquisition flow. Besides the camera C1 and the laser range scanner L1 for localization, two additional laser scanners L2, L3 are mounted.

Omnidirectional Camera

To achieve high quality images with a panoramic mirror a high spa-

tial resolution is necessary. We utilize a Canon EOS-20D SLR camera mounted vertically. We use the manual mode, where the highest possible aperture value of 32 is chosen for maximal depth of focus and the exposure time is set accordingly. On the camera a panoramic lens attachment from 0-360.com is mounted which offers a large field of view (115°), whereas 52.5° lie above the horizon and 62.5° lie below. The calibration is performed by fitting a cubic polynomial assuming a radialsymmetric mirror and using a calibration pattern.

Laser Scanner

Three 2D Sick LMS 200 laser range scanners are utilized, which are eye-safe and widely used in robotics and industry. They feature a 180° field-of-view, max. 0.25° angular resolution and a sample rate above 70 Hz, offering an accuracy of 10 mm, a statistical error of 5 mm and a systematic error of 15-40 mm. We use a RS422 (differential signaling) connection via a multi-RS422 to USB bridge to the host, running synchronously at 500kbps each. L1 is mounted horizontally to record range values . This data is used to build a two dimensional map and to localize the mobile platform with respect to this map. L2, L3 are mounted vertically in a way that no occlusion from the mounting occurs. Their data yields the geometric information for the "laser" 3D model. They slightly overlap and together offer a 360° field of view. Thus, the environment is recorded in slices (in contrast to 3D scanners).

3. MAP BUILDING & LOCALIZATION BY PROBABILISTIC SCAN MATCHING

We use a probabilistic scan matching approach to solve the simultaneous localization and mapping (SLAM) problem based on L1. It is split in the map building and the localization step relative to this map. This unit yields the pose information of the platform (and thus all sensors) during the acquisition of the 3D environment. For map building, a probability density function of a scan is computed (in 2D space) that is thought of having generated the scan. This density function is used as a generative model and (using a maximum likelihood formulation) an energy function is derived from this generative model. This energy function describes how probable another scan is, dependent on translation and rotation. This energy function is optimized quickly using Newton's algorithm. Gradient and Hessian are thereby calculated analytically. The environment is represented by a graph of spatial relations whereas the nodes represent the poses of the where the laser scans were recorded and the edges represent pairwise registrations of two scans. An error function over the graph is then optimized over all poses simultaneously (i.e., we have $3 \times nrScans$ free parameters). Details of our method can be found in in [1]. A part of the map's graph is illustrated in Fig. 5. After the map has been built it is used to localize all scans - and thus



Fig. 5. Part of the graph that the map consists of.

the platform. For this each scan is matched in turn to the map using the previous one as prior.

4. STEREO VISION BASED 3D MODELING FLOW

A high quality, state-of-the-art stereo algorithm - namely the graph cut algorithm by Kolmogorov and Zabih - is used to calculate a disparity map for each panoramic image. Our implementation finds its roots in [2] that extends the work of Kolmogorov and Zabih [3] to omnidirectional vision. Our stereo matching pipeline works as follows: for each pixel in the first image the epipolar curve in the second image is created according to the epipolar geometry of our panoramic camera. These points representing the different disparities are used to construct a rectified window taking zoom into account. Then, an SSD error value for each disparity on this epipolar curve is computed and saved. The image that is being processed is compared both to the next and to the previous image. These steps provide the data needed by the graph cut algorithm. The key of graph cut is formulating the correspondence problem as an energy minimization problem. The minimization is done iteratively by transforming the energy minimization problem into several minimum cut problems done by an algorithm based on α -expansion moves. The resulting disparity map is converted into a point cloud and postprocessed as follows: the graph cut approach is by its nature restricted to discrete disparity steps. Disparity values are refined to subpixel accuracy by finding the optimum of a local quadratic model built using the original matching cost at the discrete disparity value and its adjacent disparity values. Regions around the epipoles (there are two epipoles in omnidirectional images) are removed because these typically provide too few constraints to extract reliable depth information. In a further step depth values that belong to the floor with high probability are corrected to be exactly on the floor. The epipole removal and the graph cut algorithm both mark some pixels as unknown or occluded. The distance values for these pixels are interpolated from the surrounding, known distances using linear interpolation along concentric circles.

The graph cut based stereo pipeline in action is shown in Fig. 6 where the results after each step are illustrated. First the winner takes all solution of the stereo matching then the graph cut results and afterwards the postprocessing steps (subdisparity refinement, epipole removal, floor correction and hole filling) are shown. Resulting 3D models are presented in section 7.



Fig. 6. Stereo processing using graph cut algorithm and postprocessing steps. (Although it runs on three images, only two source images are shown due to lack of space.) (1) Left source image. (2) Right source image. (3) Winner takes all solution of stereo matching. (4) Result of graph cut algorithm. (5) Postprocessing: Subdisparity refinement. (6) Postprocessing: Epipoles removal. (7) Postprocessing: Floor correction. (8) Postprocessing: Filling incomplete data. Represents the final disparity map after all postprocessing steps.

5. LASER SCANNER BASED 3D MODELING FLOW

Using the localization from L1, the absolute world coordinates of the scanner results of L2, L3 can be calculated. These give us a high quality representation of the surrounding geometry, which is then textured using the images from the omnidirectional camera C1. Compared to the stereo vision based flow, the model is more accurate due to the absolute measurement accuracy of the laser range scanners. Additionally, it works also in textureless environments. However, the spatial resolution is lower, as visual sensors like cameras tend to have higher resolutions than other sensors.

6. VISUALIZATION ON THE INTERACTIVE 3D FOGSCREEN

We visualize the acquired 3D models on a novel walk-through 3D display based on the 2D FogScreen mid-air projection screen [4]. The screen can also be made interactive, reacting to the touch of the viewers' hands in the air [5], as shown in Fig. 7. It is realized using also a Sick LMS200 laser scanner mounted so that its scanning plane is parallel to the FogScreen. This turns the passive projection



Fig. 7. Sensors of the Interactive FogScreen. Left: Sick LMS200 laser range scanner, Center: Calibration procedure Right: Interactive mouse in operation

screen into an immaterial 2D touch screen, and greatly extends the application possibilities. The FogScreen can also be extended into a 3D display and be made 3D interactive [6]. Our visualization system allows a single user to view objects floating in mid-air from any angle, and to reach and walk through them. Two individual, but coordinated, images are projected onto opposite sides of a thin film of dry fog without cross-talk, and the 3D tracking system allows the user to interact with the content on both sides of the screen. The permeability of the FogScreen is important not only for imaging and visualization, but also for a variety of new interface techniques. Our contribution is a new visualization method for 3D/virtual environments to create an exciting way to view 3D imagery. For these early experiments, we use a commercial computer vision-based 3D head tracking to provide motion parallax and correct perspective rendering for a single user. Multiple types of stereoscopic imagery can be used. Interestingly the FogScreen creates a more pronounced sense of depth than a traditional screen in our informal usability tests. The viewer can also cross the screen to see the content from the back. The result is a wall-sized, immaterial display that creates an engaging 3D visual of the acquired scene. If a full screen image of a virtual environment is displayed, it creates more of a portal effect, giving the impression of looking through a window into another world. Several screens could be put into a CAVE-like configuration, but one screen works as a "poor man's CAVE". The user can freely move in the viewing area and can look behind by moving to the other side of the screen. The head-tracked, two-sided FogScreen visualization setup seems to be good for viewing and browsing human-sized, natural environments like streets, halls, etc. Also, e.g., astronomy and other scientific visualizations might look engaging on it. Users can walk around the screen to see the full scene and could easily get a better idea of the entire 3D volume.

7. RESULTS

We will now outline some results to give an idea of the achieved visual quality of the 3D models. All models are also available for download from the project's website.

Stereo Vision based Acquisition Flow

Fig. 8 shows renderings of 3D models acquired using the graph cut based omnidirectional stereo vision flow.



Fig. 8. Stereo Flow: Top: Office room. Bottom: Outdoor parking lot scene in front of our institute.

Laser Scanner based Acquisition Flow

Fig. 9 illustrates some renderings of our laser scanner based 3D models: an outdoor car scene and the hallway of our institute.



Fig. 9. Laser Flow: Top: Outdoor car scene. Bottom: Renderings of our whole institute's hallway.

3D Visualization on the Interactive FogScreen

The interactive FogScreen system in action is shown in Fig. 10, displaying our 3D model.



Fig. 10. FogScreen Visualization: 3D Visualization of our 3D model on the Interactive FogScreen

8. CONCLUSION

We have have covered essential parts of future 3DTV systems -3D model acquisition and 3D visualization of these models.

On the *acquisition* side, two flows have been presented: graph cut based omnidirectional stereo vision flow and laser range scanner based flow. For this purpose, we have designed a simple yet highly flexible sensor platform – the Wägele. It can be transported easily and is capable of acquiring both indoor and outdoor scenes. In conjunction with the platform this enables hassle free acquisition of 3D scenes for 3DTV.

On the visualization side, we have presented a novel 3D interactive display for visualizing these 3D models – the 3D Interactive FogScreen. It is based on the patented FogScreen technology and is capable of displaying 3D models in a 3D fashion while still keeping the advantage of enabling walk-throughs – a big benefit also for amusement parks which is one area where such 3DTV devices will be seen in the near future.

9. ACKNOWLEDGMENTS

We would like to thank Michael Wand for providing the XRT rendering framework. This work is supported by EC within FP6 under Grant 511568 with the acronym 3DTV.

10. REFERENCES

- [1] Peter Biber, Sven Fleck, and Wolfgang Straßer, "A probabilistic framework for robust and accurate matching of point clouds," in 26th Pattern Recognition Symposium (DAGM 04), 2004.
- [2] Sven Fleck, Florian Busch, Peter Biber, Henrik Andreasson, and Wolfgang Strasser, "Omnidirectional 3d modeling on a mobile robot using graph cuts," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2005.
- [3] Vladimir Kolmogorov and Ramin Zabih, "Computing visual correspondence with occlusions using graph cuts," in *International Conference on Computer Vision (ICCV'01)*, 2001.
- [4] K. Palovuori and I. Rakkolainen, "Fogscreen.," in U.S. patent 6,819,487 B2. November 16, 2004, 2004.
- [5] Sven Fleck, Mika Piirto, and Ismo Rakkolainen, "Laserscanning for the interactive walk-through fogscreen," in 9th International Conference on (IV-HCI), 2005.
- [6] S. DiVerdi, I. Rakkolainen, T. Hoellerer, and A. Olwal, "A novel walk-through 3d display," in SPIE Electronic Imaging, Stereoscopic Displays and Applications XVII, 2006.