Model–Based 3D Scene Analysis from Stereoscopic Image Sequences

Abstract

An approach for the modelling of complex 3D scenes like outdoor street views from a sequence of stereoscopic image pairs is presented. Starting with conventional stereoscopic correspondence analysis a 3D model scene with true 3D geometry is generated. Not only the scene geometry but also surface texture is stored within the model. 3D camera motion can be estimated directly from the image sequence to track camera motion and to fuse measurements throughout the sequence into a common 3D model scene. From the textured 3D model realistic looking image sequences from arbitrary view points can be synthesized using 3D computer graphics methods. This approach can be used to synthesize new and intermediate views for stereoscopic television systems.

Introduction

For the reconstruction of complex 3D scene a series of problems need to be solved, especially the problems of 2D correspondence and 3D registration. To be able to triangulate the scene from a particular pair of images a calibration of the projection geometry and the relative orientation of the cameras to each other is needed. When using a binocular camera setup, the calibration can be performed once before the measurement and will remain constant throughout the measurement phase. Measurement of scene geometry relative to such a binocular camera system can then be obtained through triangulation of corresponding image points. During correspondence analysis one tries to uniquely identify the projections of a scene element onto the camera targets. In a complex natural scene parts of the scene may be occluded to the camera system so that the camera has to be moved throughout the scene. The scene is then analyzed from a sequence of image pairs. To register all measurements into a common scene coordinate system the 3D motion of the camera system is tracked and all measurements are included to build a true 3D model of the scene [1].

The presented approach addresses the problems stated above for building a true 3D model of a complex scene from a sequence of stereoscopic image pairs. Not only the scene geometry but also the surface texture is stored within the model. It is therefore possible to synthesize realistic looking image sequences from the model using 3D computer graphics methods [2].

Building a 3D model scene

A camera model as introduced by Yakimovsky and Cunningham [3] serves to calibrate the camera system. The calibration process is carried out once to determine relative camera geometry. The system is able to cope with arbitrary relative orientations but fast performance is achieved when the cameras are parallel to each other and the camera center is shifted parallel to the horizontal image plane coordinates only.

The correspondence problem is solved using an area based matching process. In areas where sufficient surface texture is available a small rectangular image patch is taken from one image and the corresponding image patch is searched for in the other image. Epipolar constraints limit the search space and the best correspondence is evaluated through cross correlation. The corresponding pairs are converted to relative depth measurements and recorded in a depth map. The confidence of the measurement, based on the maximum cross correlation coefficient and image texture measures, is stored in a certainty map for each corresponding point. Uncertain measurements and areas without corresponding points are interpolated and the resulting dense depth map is segmented into surfaces with similar surface orientation and depth. For the measurements a small base line is used to minimize perspective distortions between both views and to minimize areas occluded in one of the images. When using a small base line it is easier to find corresponding points but the depth triangulation is uncertain. This problem is eliminated when more views of the scene from a moving camera system become available. To track the 3D motion of the camera an algorithm to estimate 3D motion of given 3D objects directly from the image sequence is employed. Once the camera position is known the depth measurements from each image pair can be transformed from the relative image coordinate system into the global scene coordinate system and fused with the preceding measurements. The depth certainty increases when more image pairs are analyzed throughout the image sequence as the camera moves.

From the depth map a geometric model of the 3D scene is created. The 3D surface is constructed from small triangular image patches that are able to approximate arbitrarily shaped surfaces. The surface texture of each triangular patch is stored in a texture map. This feature allows to synthesize arbitrary new views of the scene [4] and to generate intermediate views for stereoscopic television systems.

References

This work was supported by the Research Institute of the German postal service TELEKOM.

[1] J. Aloimonos, D. Shulman: *Integration of Visual Modules*, Academic Press, San Diego, USA, 1989.

[2] R. Koch, C.–E. Liedtke, *3D–Modelling of Natural Scenes from Image Sequences for Stereoscopic Reconstruction*, 2nd Workshop on Stereoscopic Television, Darmstadt 1990.

[3] Yakimovski, Y., R. Cunningham, *A System for Extracting 3D Measurements from a Stereo Pair of TV Cameras*, Computer Graphics and Image Processing Vol 7, pp. 195–210, 1978.

[4] R. Koch, *Automatic Modelling of Natural Scenes for Generating Synthetic Movies*, Eurographics '90, Montreux, Schweiz, 1990.