MULTICAMERA BASED AUTOMATED ROTOSCOPING & DEPTH MAP ESTIMATION USING PATCHMATCH

Jonas Trottnow*, Simon Spielmann*, Volker Helzle*

*Institute of Animation - Filmakademie BW, Germany, research@filmakademie.de

Keywords: rotoscoping, depth map, Time-of-Flight camera, sensor fusion

Abstract

This short paper proposes a method for computing depth maps from multiple satellite cameras and fusing them to an enhanced one for the principal camera of a movie production. Additional precision is gathered from optionally adding a Time-of-Flight camera. The proposed algorithm is a generalization of the PatchMatch Stereo algorithm [1]. It calculates multiple stereo depth maps without the need of a common baseline and fuses them to one global depths map.

1 Introduction

In recent film productions it gets more and more valuable to have depth information along with the normal footage being captured. This depth information can be used for rotoscoping, stereo conversion and special effects. Up to now it is a time consuming manual process to split footage up into depth planes. With the proposed method the process of manually generating depth information is automated by using additional small cameras placed on the camera rig.

2 Setup

Surrounding the principle broadcasting camera small cameras (e.g. webcams) are placed, capturing the scene from slightly different viewpoints (see Fig.1). The cameras do not need to be placed on one baseline and the exact orientation and offset between them does not need to be known. To further enhance the result a Time-of-Flight camera is added to the rig. Time-of-Flight cameras are able to capture depth directly with a low resolution.

Fig.1: Custom camera rig mounted on an Arri-Alexa

3 Pre-processing

Initially a calibration grid has to be captured. From this calibration data all cameras are undistorted and the cameras fundamental matrix between the satellite cameras and the principal camera are estimated.

4 Modified PatchMatch Stereo

Through the estimated fundamental matrix the rotation between the cameras is eliminated, the simplified epipolar lines in one image are calculated for each pixel of the other image. This replaces the need of images being rectified to one baseline as being used in the original PatchMatch Stereo algorithm, which is not possible in our setup.

To calculate the disparity map between a satellite camera and the main camera the processing steps of the PatchMatch algorithm are modified to search on the epipolar line instead of a row-based matching approach.

PatchMatch Stereo splits the calculation up into different stages, starting with a completely random initial guess, followed by a spatial and view propagation and a plane refinement step as described in [1]. In contrast to other stereo algorithms PatchMatch calculates a plane for each pixel. In other words for each 3D pixel of the depth map a normal is estimated as well, which enhances the result especially when dealing with slanted and not fronto-parallel surfaces.

Depth maps are calculated for both cameras of a stereo pair. With a left-right consistency check, where the difference of depth values between matching pixels in the two frames is evaluated, a binary confidence image is generated in which trusted depth values and possibly inaccurate ones are separated. Based on this confidence image the different depth maps for the main camera are fused. As the offset between the satellite cameras and the main camera may vary, the disparity values between the depth maps are varying too. Based on the trusted pixels gathered from the confidence image a scaling is calculated to match the depth values.

$$Ax + y = B \qquad (1)$$

This is done by solving equation (1), where A is a vector of depth values from the first camera's depth map and B is the vector of corresponding depth values in the second camera. Only pixels that are trusted in both images are used here.

Untrusted depth values are then filled up with trusted ones from other depth maps. Still remaining untrusted depth values can be estimated with the use of the planes estimated by the PatchMatch algorithm through which not only interpolation but plane extrapolation takes place.

The Time-of-Flight camera can be used to map the disparities to real depth values and further enhance the smoothness of floating point depth values.

Acknowledgements

This work has been carried out in the EU project Dreamspace.

References

[1] Michael Bleyer, Christoph Rhemann, Carsten Rother. "PatchMatch – Stereo Matching with Slated Support Windows", *BMVC*, (2011).