Reconstructing Depth from Spatiotemporal Curves

Rui Rodrigues — rpr@tom.di.uminho.pt
António Fernandes — af@di.uminho.pt
Universidade do Minho, Portugal

Kees van Overveld — overv@natlab.research.philips.com
Fabian Ernst — fabian.ernst@philips.com
Philips Research, Eindhoven

Abstract

We present a novel approach for 3D reconstruction based on multiple video frames taken from a static scene. Our solution emerges from the spatiotemporal analysis of video frames. The method is based on a best fitting scheme for spatiotemporal depth curves, which allows us to compute 3D world coordinates of the objects within the scene. As opposed to a large number of current methods, our technique deals with random camera movements in a transparent way, and even performs better in these cases than with pure translation. Robustness against occlusion and aliasing is inherent to the method as well.

1 Introduction and Context

Structure from Motion (SfM) is a computer vision field still in progress. It deals with the problem of recovering the 3D structure of a scene from different perspective projections (e.g. video frames taken with a moving camera) [1]. The range of applications of SfM techniques includes 3D-scene modelling, virtual view generation, 3D TV, image/video synthesis and autonomous navigation.

The availability of various perspective projections allows us to estimate depth (the distance to the camera) of objects by comparing the projections' relative displacement of such objects in different frames, and using knowledge about camera motion.

Several problems arise in this seemingly simple process. The first problem is related to camera information, which often is not readily available and has to be estimated as well. This is the camera calibration problem.

Second, it is not trivial to determine which parts of a set of digital images correspond to the same object or local feature in 3D space. This is known as the correspondence problem [2]. Some factors that contribute to the correspondence problem are image noise, periodic textures and the occlusion of objects.

A third problem arises due to numerical or geometric instability: the stability problem.

The techniques described in the extensive literature available on SfM range from block matching algorithms to stochastic techniques, texture-based to feature-based. Many of the concepts are inherited from motion estimation research.

A large number of techniques analyse the case of consecutive frame pairs or triplets (e.g. trilinear tensor), estimating motion (depth) for each pair or triplet, and integrating that estimated data overtime as a post-processing operation. These techniques face stability problems in the fusion of the estimated data.

The reader is referred to [1] for an overview and references on SfM methods.

In this paper we focus on fusion. One way to perform this fusion still at the stage of motion/depth estimation is to regard video data as 3D information, time being the third dimension [3]. In this context, a sequence with a given number of frames can be represented as a colour distribution on the spatiotemporal domain, resulting in a spatiotemporal volume – the VideoCube. If the camera motion between consecutive frames is relatively small, the similarity between frames allows identifying spatiotemporal curves and surfaces, corresponding to the temporal path of objects throughout the scene.

The shape of these spatiotemporal entities is related to both the camera motion and the position of identifiable points in world space.

We propose a SfM method which exploits the VideoCube assuming that camera parameters (both intrinsic and extrinsic) are known, or at least well estimated.

The method estimates the depth of a set of points chosen from the video images based on a set of depth candidates and a best-fit metric of the spatiotemporal curves corresponding to those candidates. It is assumed that the scene is static with little or no highlights.

The main features of our approach are:

-The ability to deal with arbitrary (including non-smooth) motion paths

-The potential to combine benefits of frame-to-frame coherence (meaning few occlusion differences on short time scale) with large baseline (meaning that due to the large total camera paths, geometric stability can be achieved). Furthermore, all frames can be used simultaneously.

- The correspondence problem is tackled with a stochastically stable matching technique

- Robustness to occlusion, noise and aliasing is inherent to the method

The paper is organised as follows: The VideoCube is introduced in section 2, along with a brief overview of the literature in the area of video spatiotemporal analysis, followed by the concept of spatiotemporal curves in Section 3. In section 4, a technique for estimating spatiotemporal curves is derived and later summarised in an algorithm. Section 5 contains the results obtained with a current implementation of the algorithm. Section 6 compares the technique with other SfM and spatiotemporal based techniques, showing the main differences. Finally, conclusions and future work close this paper in sections 7 and 8, respectively.