Creating Stereoscopic (3D) Video from a 2D Monocular Video Stream

Xiaokun Li[1], Roger Xu[1], Jin Zhou[2], and Baoxin Li[2]

[1] Intelligent Automation, Inc, Rockville, MD 20855
[2] Arizona State University, Tempe, AZ 85287

Abstract. It is a challenge to generate stereoscopic (3D) video through a single moving camera under widely varying conditions. We propose an efficient approach to create true stereoscopic video from a monocular video stream captured under various moving conditions. The approach contains three major steps. First, we apply Harris' corner detector to detect distinctive feature points from a pair of image frames selected from the incoming video captured by a moving camera. Second, according to the consecutive property of the video, a local-window search based algorithm is developed for fast and accurate feature correspondence between the two image frames. Third, a hierarchical image rectification technique is designed to guarantee the success in creating a true and visually-comfortable stereo image for each incoming image frame. Besides, a software-based video stabilization algorithm is also developed for improved stereo video generation performance. Extensive tests using real video collected under various situations were performed for performance evaluation of the proposed approach.

1 Introduction

In gaming and TV programs, 3D video effects are one of the most attractive features. 3D video techniques have also found wide civilian applications, such as medical operations, microscopy, scientific data display, and CAD/CAM. Military applications of 3D techniques include battlefield reconnaissance and surveillance. Conventional computer-based stereo vision techniques, although studied for many years, still have many limitations. For example, a conventional stereo vision system requires two identical cameras, a narrow baseline, fixed parameter settings and positions. It is only suitable for short-range scenes. However, in real world, camera motion is often nonstationary and viewpoints of the camera are different from time to time. Furthermore, parameter settings of the camera are sometimes unknown and variable during video capturing.

To generate stereo (3D) video captured by a moving camera under widely varying conditions presents a challenge. The main reason is that the camera is obliquely mounted on a platform when the platform moves non-linearly or when the camera parameters vary while the platform is moving. In recent years, many research efforts have been made on stereo generation with uncalibrated cameras. Fusiello *et al.* [1] developed a compact and efficient stereo generation algorithm via image rectification.

G. Bebis et al. (Eds.): ISVC 2007, Part I, LNCS 4841, pp. 258–267, 2007.

However, their approach assumes that the stereo rig is calibrated, which means the intrinsic parameters of the camera pair such as focal length, aspect ratio, and their relative position are already precisely known. Unfortunately, as mentioned earlier, the camera parameters are not readily available and the relative position between the two cameras is difficult to obtain or calibrate in practice. Loop and Zhang at Microsoft Research [2] developed one method to construct stereo images with uncalibrated cameras. Their method mainly relies on stereo matching and the residual distortion may result in poor visualization performance. The method proposed by Hartley & Zisserman [3], [4] for stereo generation from uncalibrated cameras is the most advanced one in the literature to our best knowledge. One important distinction of their method is that it is insensitive to unknown and variable camera parameter settings during image/video capture. However, the quality of the generated stereo images cannot be guaranteed. In some cases, the resulting stereo image may even be corrupted.

Although many efforts have been made on stereo generation with uncalibrated cameras, stereo video generation from a single moving camera is a fairly new research topic. In this paper, we focus on the following stereo vision problem: given two video frames acquired by a moving video camera at two different locations with different viewing angles, we create a stereo image frame based on the two video frames by means of feature extraction and correspondence, and image rectification. The resulting stereo frame gives a viewer a realistic 3D perception, and can be displayed on any type of display devices. Furthermore, for each video frame in the video stream, we can successfully construct a stereo image based on the current frame and a second frame with a predefined constant time delay very rapidly. As a result, stereo video can be generated. The biggest challenge in generation stereo video from the 2D video stream is how to generate a true and eye-comfortable stereo image for each video frame. As can be found in the later part of this paper, our proposed robust and efficient approach can successfully create stereo video in various situations.

2 Algorithm Description

The proposed approach consists of four algorithms executed in a sequential order. The first algorithm is software-based video stabilization. The second is feature extraction which provides an efficient way for image feature extraction by Harris' corner detection. The third is a local-window-based search method to find feature correspondence of input frame pair. The last algorithm is stereo video generation based on image rectification.

2.1 Software-Based Stabilization

When a camera is in motion during video acquisition, the acquired video may contain unexpected jitters that will certainly affect the image rectification process of stereo generation. Such scenario is illustrated in Fig. 1. To show how frame jitters may affect image rectification, we use the example in Fig. 2 (a) for illustration. Suppose