

Mosaic-Based 3D Scene Representation and Rendering

Zhigang Zhu

Department of Computer Science
City College of New York, New York, NY 10031
zhu@cs.cuny.edu

Allen R. Hanson

Department of Computer Science
University of Massachusetts Amherst, MA 01003
hanson@cs.umass.edu

Abstract - In this paper we address the problem of fusing images from many video cameras or a moving video camera. The captured images have obvious motion parallax, but they will be aligned and integrated into a few mosaics with a large field-of-view (FOV) that preserve 3D information. We have developed a geometric representation that can re-organize the original perspective images into a set of parallel projections with different oblique viewing angles. In addition to providing a wide field of view, mosaics with various oblique views well represent occlusion regions that cannot be seen in a usual nadir view. Stereo pair(s) can be formed from a pair of mosaics with different oblique viewing angles and thus image-based 3D viewing can be achieved. This representation can be used as both an advanced video interface and a pre-processing step for 3D reconstruction.

Keywords – stereo mosaics; image-based rendering; visual representation;

I. INTRODUCTION

A 2D panoramic mosaic of a 3D scene generated from video from a translating camera has the geometry of multiple viewpoints [1,2], but it only preserves information from a single viewing direction. Three-dimensional (3D) structure and surface information from other viewing directions of the original video is lost in such a representation. A digital elevation map (DEM) generated from aerial photometry consists of a sampled array of elevations (depths) for a number of ground positions at regularly spaced intervals [3]. It usually only has a nadir viewing direction, hence the surfaces from other viewing directions cannot be represented. However, in some applications such as surveillance and security inspection, a scene or an object (e.g. a vehicle) needs to be observed from many viewing directions to reveal anomalies hidden in unusual views. Stereo panoramas [4,5] have been presented as a mechanism for obtaining the “best” 3D information from an off-center rotating camera. In the case of a translating camera, various layered representations [6-8] have been proposed to represent both 3D information and occlusions, but such representations need 3D reconstructions.

This paper presents an approach for fusing images from many spatially distributed video cameras or a moving video camera into a few mosaiced images that preserve 3D information. In both cases, a virtual 2D array of cameras with FOV overlaps is formed to generate complete coverage of a scene (or an object). Note that many viewing directions are

already included in the original camera views. This property has been used in the X-slit mosaics with non-parallel rays [9] for image-based rendering. In this paper we propose a representation that can re-organize the original perspective images into a set of parallel projections with different oblique viewing angles (in both the x and the y directions of the 2D images). Mosaics with 2D oblique parallel projections are a unified representation for our previous work on parallel-perspective stereo mosaics [11,12] and multi-camera mosaics [13,14]. Such representations provide a wide field of view, optimal 3D information for stereo viewing and reconstruction, and the capability to represent occlusions. This paper is organized as follows. In addition to the description and discussions of the unified presentation in Sections II and III, we also present a general ray interpolation method in Section IV, and a mosaic-based 3D rendering method in Section V. Experimental results are given in Section VI for two important applications – aerial video surveillance and under vehicle inspection. Section VII is a brief summary.

II. 2D OBLIQUE PARALLEL PROJECTION

A normal perspective camera has a single viewpoint, which means all the light rays pass through a common nodal point. On the other hand, an orthogonal image with parallel projections in both the x and y directions has all the rays parallel to each other. Imagining that we have a sensor with parallel projections, we could turn the sensor to capture images with different *oblique* angles (including both nadir and oblique angles) in both the x and y directions. Thus we can create many pairs of parallel stereo images, each with two different oblique angles, and observe surfaces occluded in a nadir view.

Fig. 1 shows the parallel stereo in a 1D case, where two oblique angles β_1 and β_2 are chosen. The depth of a point can be calculated as

$$Z = \frac{B}{\tan \beta_2 - \tan \beta_1} \quad (1)$$

where β_1 is β_2 are the angles of the two viewing directions, respectively, and B is the adaptive baseline between the two viewpoints that is proportional to the displacement of the corresponding points in the stereo mosaics. It has been shown by others [10] and by us [11, 12] that parallel stereo is

superior to both conventional perspective stereo and to the recently developed multi-perspective stereo with concentric mosaics for 3D reconstruction (e.g., in [5]). The adaptive baseline inherent in the parallel-perspective geometry permits depth accuracy independent of absolute depth in theory [10,11] and as a linear function of depth in stereo mosaics generated from perspective image sequences [12]. In contrast, the depth error of perspective stereo and the concentric stereo is proportional to the square of depth.

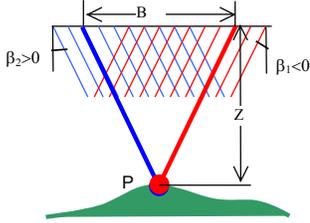


Fig. 1. Depth from parallel stereo with multiple viewpoints: 1D case.

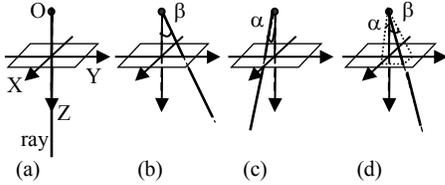


Fig. 2. Parallel projections with two oblique angles α and β (in the x and y directions).

We can make two extensions to this parallel stereo. First, we can select various oblique angles for constructing multiple parallel projections. By doing so we can observe various degrees of occlusions and can construct stereo pairs with different depth resolution via the selection of different pairs of oblique angles. Second, we can extend this 1D parallel projection to 2D (Fig. 2): we can obtain a mosaiced image that has a nadir view (Fig. 2a), oblique angle(s) only in one direction (Fig. 2b and c) or in both the x and the y directions (Fig. 2d).

III. 2D (VIRTUAL) ARRAY OF CAMERAS

It is impractical to use a single sensor to capture orthogonal images with full parallel projections in both x and y dimensions for a large scene, and with various oblique directions. However there are least three practical ways of generating images with oblique parallel projections using existing sensors: a 2D sensor array of many spatially distributed cameras (Fig. 3a), a “scanner” with a 1D array of cameras (Fig. 3b), and a single perspective camera that moves in 2D (Fig. 3c).

With a 2D array of many perspective cameras (Fig. 3a), we first assume that the optical axes of all the cameras point in the same directions (into the paper in Fig 3a), and the viewpoints of all cameras are on a single plane perpendicular to their optical axes. Then we can reorganize the perspective images into mosaiced images with any oblique viewing angles by extracting rays from the original perspective images with the

same viewing directions, one ray from each image. If the camera array is dense enough, then we can generate densely mosaiced images.

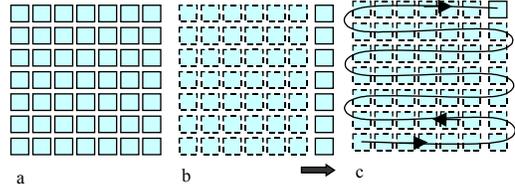


Fig. 3. Parallel mosaics from 2D bed of cameras. (a) 2D array; (b) 1D scan array and (c) a single scan camera.

If we only have a 1D linear array of perspective cameras (Fig. 3b), we can “scan” the camera array over the scene to synthesize a virtual 2D camera array. Then we can still generate stereo mosaic pairs with oblique parallel projections, given that we can accurately control the translation of the camera array. We have actually used this approach in an Under Vehicle Inspection System (UVIS) [13, 14, 18].

Even if we just use a single camera, we can still generate a 2D virtual bed of cameras by moving the camera in two dimensions, along a “2D scan” path shown in Fig. 3c. This is the case for aerial video mosaics [11, 12, 15, 17].

IV. PRISM: VIDEO MOSAICING ALGORITHM

In real applications, there are two challenging issues. The first problem is camera orientation estimation (calibration). In our previous study on an aerial video application, we used external orientation instruments, i.e., GPS, INS and a laser profiler, to ease the problem of camera orientation estimation [11, 12]. In this paper, we assume that the extrinsic and intrinsic camera parameters are known at each camera location. The second problem is to generate dense parallel mosaics with a sparse, uneven, camera array, and for a complicated 3D scene. To solve this problem, we have proposed a Parallel Ray Interpolation for Stereo Mosaics (PRISM) approach [11]. While the PRISM algorithm was originally designed to generate parallel-perspective stereo mosaics (parallel projection in one direction and perspective projection in the other), the core idea of *ray interpolation* can be used for generating a mosaic with full parallel projection at any oblique angle.

Fig. 4 shows how the PRISM works for 1D images. The 1D camera has two axes – the optical axis (Z) and the Y -axis. Given the known camera orientation at each camera location, one ray with a given oblique angle β can be chosen from the image at each camera location to contribute to the parallel mosaic with this oblique angle β . The oblique angle is defined against the direction perpendicular to the *mosaicing direction*, which is the dominant direction of the camera path (Fig. 4). But the problem is that the “mosaiced” image with only those existing rays will be sparse and uneven since the camera array are usually not regular and very dense. Therefore *interpolated* parallel rays between a pair of existing parallel rays (from two neighboring images) are generated by performing local

matching between these two images. The assumption is that we can find at least two images to generate the parallel ray. Such an interpolated ray is shown in Fig 4, where Ray I is interpolated from Image A and Image B.

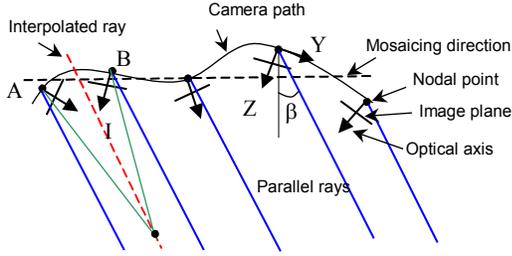


Fig. 4. Ray interpolation for parallel mosaicing from an camera array.

The extension of this approach to 2D images is straightforward, and a region triangulation strategy similar to that in [11] can be applied here to deal with 2D cases. One important issue here is the selection of neighborhood images for ray interpolation. For example, with a 1D scan sequence of a single camera, it is hard to generate full parallel projection in the X direction, which is perpendicular to the motion of the camera, since the interpolated parallel rays far off the center of the images in the x direction have to use rays with rather different oblique angles in the original perspective images.

V. STEREO VIEWING AND 3D RECONSTRUCTION

Parallel mosaics with various oblique angles represent scenes from the corresponding viewing angles with parallel rays and with wide fields of view. There are two obvious applications of such representation. First, a human can perceive the 3D scene from a pair of mosaics with different oblique angles (e.g. using polarized glasses) without any 3D recovery. If we have mosaics with various oblique angles in both the x and the y direction, we can generate a virtual fly/walk-through by noting that (1) the translation in the xy plane can be simulated by shifting the current displayed mosaic pair, (2) the rotations around the X and the Y axes can be simulated by selecting different pairs of mosaics with different oblique angles, and (3) the rotation around the optical axis only needs to rotate the pair of mosaics in their image planes. The visual disparities can also be controlled by changing the angles between the two mosaics for stereo viewing.

Second, for 3D recovery, matches are only performed on a pair of mosaics, not on individual video frames. Stereo mosaic methods also solve the baseline versus field-of-view (FOV) dilemma efficiently by extending the FOV in the directions of mosaicing – in both the x and y directions. More importantly, the parallel stereo mosaics have optimal/adaptive baselines for all the points, which leads to uniform depth resolution in theory and linear depth resolution in practice. For 3D reconstruction, epipolar geometry is rather simple due to the full parallel projections in the mosaic pair.

VI. EXPERIMENTAL EXAMPLES

A. Video Mosaics from Aerial Video

In theory, with a camera undergoing an ideal 1D translation and a nadir view direction, two spatio-temporal images can be generated by extracting two columns of pixels at the front and rear edges of each frame in motion (Fig. 5). The mosaic images thus generated are *parallel-perspective*, with parallel projection in the motion direction and perspective projection in the other. In our aerial video environmental monitoring application, a single camera is mounted in a small aircraft undergoing 6 DOF motion, together with a GPS, INS and laser profiler to measure the moving camera locations and the distance to the terrain [11, 12]. So we generate seamless stereo parallel-perspective video mosaic strips from image sequences with a 1D scan path but with a rather general motion model, using the proposed parallel ray interpolation for stereo mosaicing (PRISM) approach [11]. In the PRISM approach for large-scale 3D scene modeling, the computation is efficiently distributed in three steps: camera pose estimation via the external measurement units, image mosaicing via ray interpolation, and 3D reconstruction from a pair of stereo mosaics [11, 12, 15].

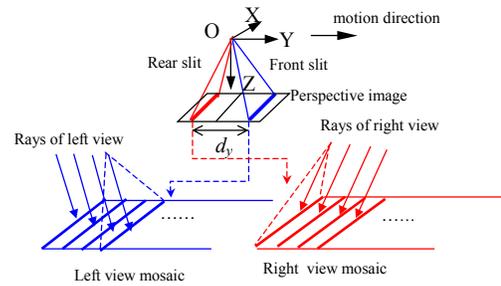


Fig. 5. Parallel-perspective stereo mosaics with a 1D camera scan path

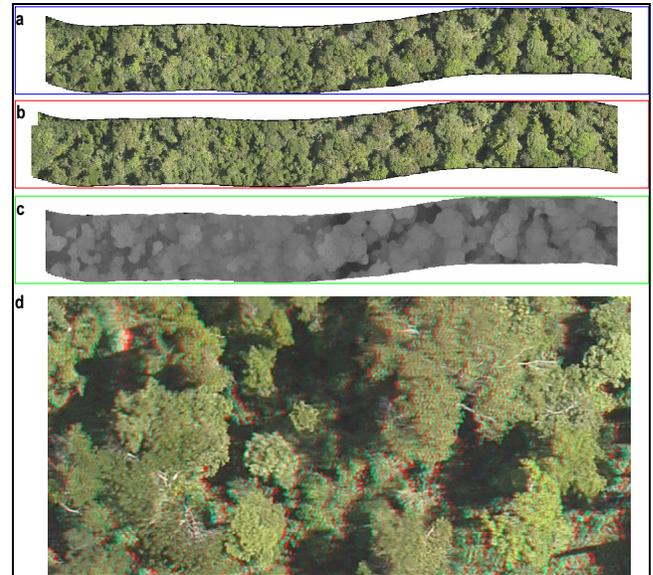


Fig. 6. Stereo mosaics and 3D reconstruction of a 166-frame telephoto video sequence.

