

An Algorithm for Computing Extrinsic Camera Parameters for Accurate Stereo Photogrammetry of Clouds

Jiuxiang Hu, Anshuman Razdan, Joseph Zehnder

Abstract

In this paper, we present a technique for accurate stereo photogrammetry of clouds for observation and measurement of the formation of summer thunderstorms over elevated terrain. Two digital cameras are set on a ground baseline so that the straight-line distance between the two cameras and the highest terrain is about 16 km and the spacing between the center of cameras is about 0.6km. Stereo pair images are taken every 10 seconds and sent to a server via internet from the field. Due to large distances and other practical hurdles, we cannot accurately measure the location and orientation of the cameras. Further, the terrestrial landmarks chosen for calibration can not be precisely measured in the field. The discrepancy between true and observed (measured or computed) photogrammetry parameters gives rise to the *geometric error*. The main contributions of this paper are twofold. One, we develop and present theoretical analysis of the geometric error and the necessity to find an accurate solution for location and its impact on determining accurate orientation parameters in a large scale scene. Second, we have developed a coarse to fine iterative algorithm that minimizes the geometric error to find the best parameters (location and orientation) for each camera given approximate values of these parameters. Validation of our algorithm is from experimental results which show that terrestrial position of clouds matches with the composite reflectivity from the WSR-88D radar and the calculated heights agree with the tropopause heights in the sounding data.

I. INTRODUCTION

In meteorological terrestrial photogrammetry, stereo structures of clouds provides useful information for understanding the behavior of clouds [1]. For example, Warner and et.al. [2] examined the details of behavior in hailstorm by measurement of stereo photogrammetry of cumulous clouds. Rasmussen and et. al. [3] have measured the diameter location of the opaque debris clouds of tornadoes. The methods of stereo photogrammetry for such applications have been described by Kassander and Sims (1957) [4], Shaw (1969) [5], Warner and et. al. (1973) [2], and Rasmussen and et. al.(2003) [3] and others. One method [2] to obtain the true orientation and location of cameras involves each camera site surrounded by marker poles stationed at 10° intervals of azimuth, on a circle of radius 11 m centred on the camera pedestal. Although this method can measure the orientation and location of each camera precisely in its local coordinate system, it is very hard to obtain the orientation and location of all cameras in a global coordinate system when the cameras are located hundreds of meters apart. Handheld cameras were used for deducing cloud and tornado locations in field programs of Vertication of Rotation in Tornadoes Experiment [3]. Rasmussen assumed that the exact locations of all camera centers can be measured in the field. However, our theoretical result (see equation (14)) has shown that the accuracy of stereo reconstruction not only depends on 3D camera orientation (azimuth, elevation and roll angle of its principal axis) but also on the location of camera. This assumption is impractical in our application.

Motivation for this paper stems from our project designed to measure cumulous cloud development over Mt. Lemon in Santa Catalina Mountains near Tucson, Arizona, USA [6]. The surfaces of 3D clouds project differentially onto a pair of 2D image planes of two digital cameras separated by some distance. To examine the details of the cloud development in this environment, the digital cameras take a picture every 10 seconds during day time, then

Jiuxiang Hu is with the Imaging and 3D Data Extraction and Analysis (*I³DEA*) Lab, Division of Computing Studies, Arizona State University at Poly campus, Mesa AZ 85212. E-mail: hu.jiuxiang@asu.edu.

Anshuman Razdan is Director of Imaging and 3D Data Extraction and Analysis (*I³DEA*) Lab, and Associate Professor in Division of Computing Studies, Arizona State University at Poly campus, Mesa AZ 85212. Email: razdan@asu.edu

Joseph A. Zehnder is with Global Institute of Sustainability, Southwest Consortium for Environmental Research and Policy, and Professor in Department of Geography, Arizona State University, Tempe AZ 85287-3211. Email: zehnder@asu.edu.

transfer the images to a server via the internet. Once the data is collected the geophysicist on the team selected "good" days, where the convection was sufficiently isolated and the frames from the cameras can be matched with the composite reflectivity from the radar. The time lapse images allow us to determine the location and timing of the initial convection, time scales and the detailed vertical evolution of the initial turrets during the transition from shallow to deep convection.

In order to resolve the 3D structure of cloud, for each camera we first need to determine two sets of parameters. These are called the *intrinsic* and *extrinsic* parameters. It is relatively easy to obtain the internal or intrinsic camera parameters by a camera calibration process [7] while it is more difficult to get accurate *extrinsic* parameters of the cameras in the same coordinate system and therefore these must be computed. Location and orientation comprise the extrinsic parameters of a camera.

We assume that only gross GPS location information and azimuth, elevation and roll of optical axis of cameras are known. Therefore we must compute the best approximation to true values. This discrepancy creates the *geometric error* [8].

One of the most used algorithm to solve for extrinsic parameters is the Random Sample and Consensus (RANSAC) algorithm of Fischler and Bolles [9]. However, there are two shortcomings of RANSAC which prohibit its use in our application. One is that the RANSAC algorithm randomly samples seven corner matches from the pool of possible matches. These points are then used to construct a minimum estimate of the *essential matrix* using the epi-polar constraint. From the *essential matrix* we are then able to compute the rotation-translation parameters using the algorithm of Tsai[10]. This process is only assured to give valid solutions with the minimum number of points when the condition of the decomposed essential matrix from fulfills certain conditions which in practice cannot be guaranteed. The obvious solution is to sample more data, however the m^{th} power dependency of the RANSAC algorithm on the number of samples m becomes the bottleneck. Thus using anything but the minimum number of parameters is computationally expensive. Secondly, the RANSAC algorithm requires accurate location matching of the sample points. In our experiments with cloud images taken from a few kilometers away, the RANSAC algorithm failed to produce good results as exact pixel matching could not be guaranteed since each pixel occupies a spatially large area (few meters).

To address these issues we have devised a new solution in the context of binocular vision and in next section we show how our algorithm can achieve the global minima for reducing the geometric error associated with computation of the extrinsic parameters. In section III, we show results of our algorithm on the data collected from the field.

II. BINOCULAR STEREO

A. Perspective Camera Model

We briefly present the basic pinhole camera model which is a combination of matrices with particular properties that represent the camera mapping (see [8] [11] for details). A CCD camera with a finite center is a mapping between the 3D world (object space) and a 2D image. That is, a point in the world coordinate frame \mathbf{E}^3 with coordinates $\mathbf{X} = (x, y, z)^T$ ¹ is mapped to the point on the image plane with coordinates $\mathbf{x} = (\hat{x}, \hat{y})^T$, where a line joining the point \mathbf{X} to the center of projection meets the image plane (see Fig. 1).

In this paper, the \mathbf{X} , \mathbf{Y} , and \mathbf{Z} axes are in UTM² coordinates along northward, upward and eastward directions³, respectively, and \hat{x} and \hat{y} are measured in pixels with top-left corner as image plane origin. In three steps, a 3D point can be mapped on the image [1]: (1) Transform the 3D points to the camera frame, (2) project the points on the image plane, and (3) map the points on the image plane image coordinates. If the world and image points are represented by homogenous vectors, i.e. $\mathbf{X} = (x, y, z, 1)^T$ and $\mathbf{x} = (\hat{x}, \hat{y}, 1)^T$, then a perspective camera model is expressed as a linear mapping between their homogenous coordinates as following:

$$\mathbf{x} \sim \mathbf{P}\mathbf{X} \quad (1)$$

If there are two points \mathbf{u} and \mathbf{v} in 3D and 4D, respectively, then $\mathbf{u} \sim \mathbf{v}$ means that there exists a nonzero scalar λ such that $\mathbf{u} = \lambda\mathbf{v}$. The camera projective matrix \mathbf{P} is a set of 3×4 homogeneous matrices, where the left hand

¹Transpose of the row vector (x, y, z) . In this paper, a bold-face symbol always represents a column vector or matrix.

²The Universal Transverse Mercator (UTM)

³In this application, depth of image faces southward of UTM coordinates and \hat{x} axis of image is along westward and \hat{y} along downward direction

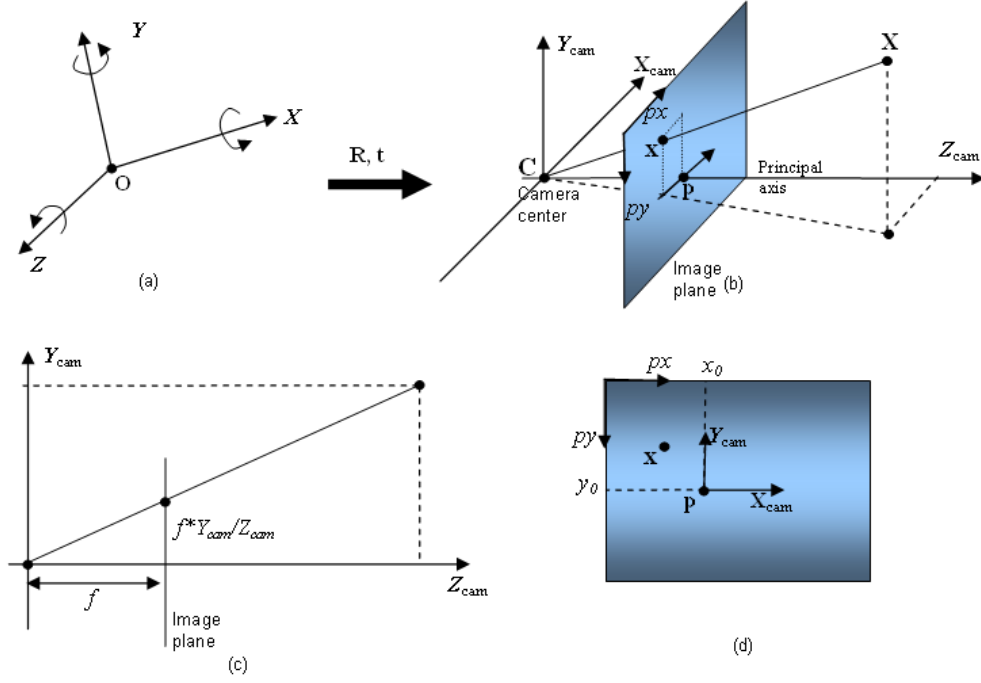


Fig. 1. Pinhole camera geometry and transformation between the world and camera coordinate frames. (a) The world coordinate system, (b) camera coordinate system, (c) mapping point onto the image plane, and (d) the image plane.

3×3 submatrix is non-singular. For a CCD camera the projective matrix \mathbf{P} can be calculated by

$$\mathbf{P} = \mathbf{K}\mathbf{R}[\mathbf{I}_3 | -\mathbf{C}] \quad (2)$$

where \mathbf{I}_3 is the 3×3 identity matrix, \mathbf{K} is called the *camera calibration matrix* and \mathbf{C} represents the coordinates of the camera center in the world coordinate frame. \mathbf{R} is a 3×3 rotation matrix representing the orientation of the camera coordinate frame. The calibration matrix of a CCD camera is

$$\mathbf{K} = \begin{pmatrix} a_x & 0 & \hat{x}_0 \\ 0 & a_y & \hat{y}_0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3)$$

where $a_x = fk_x$ and $a_y = fk_y$. f is the focal length and k_x and k_y are the number of pixels per unit distance in image coordinates in the x and y directions respectively. Similarly, $\mathbf{p} = (\hat{x}_0, \hat{y}_0)^T$ is the principal point in pixel dimensions. The parameters contained in the calibration matrix \mathbf{K} are called the *intrinsic parameters*.

In \mathbf{E}^3 , rotations about the \mathbf{X} , \mathbf{Y} , and \mathbf{Z} axes are in the clockwise direction when looking towards the origin. These are:

$$\mathbf{R}_x(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha \\ 0 & -\sin \alpha & \cos \alpha \end{pmatrix} \quad (4)$$

$$\mathbf{R}_y(\beta) = \begin{pmatrix} \cos \beta & 0 & -\sin \beta \\ 0 & 1 & 0 \\ \sin \beta & 0 & \cos \beta \end{pmatrix} \quad (5)$$

$$\mathbf{R}_z(\gamma) = \begin{pmatrix} \cos \gamma & \sin \gamma & 0 \\ -\sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (6)$$

By Euler's rotation theorem, any rotation can be represented as a composition of rotations about the three axes. α , β and γ are explained below.

In our application, after translation of the origin of the world frame to the camera center, we first rotate the y axis clockwise through azimuth angle denoted by β such that the new z axis is along the azimuth of the principal axis. Since the camera is tilted upward at an elevation angle denoted by α , we next rotate the x axis clockwise through elevation angle so that the new y axis is along the principal axis of the camera. Generally, the camera is set up askew and so the lens horizon is not horizontal. We finally twist the new z axis through an angle of roll, denoted by γ , so that the new x and y axes are parallel to the horizontal side \hat{x} and vertical side \hat{y} of the image plane of the camera respectively. Now, we have the camera coordinate frame $(\mathbf{C}, \mathbf{X}_{cam}, \mathbf{Y}_{cam}, \mathbf{Z}_{cam})$ transformed from the world coordinate frame $(\mathbf{O}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$ (see Fig. 1). In this case, the rotation matrix in equation (2) is written as

$$\mathbf{R} = \mathbf{R}_y(\beta)\mathbf{R}_x(\alpha)\mathbf{R}_z(\gamma), \quad (7)$$

where $R_x(\alpha)$, $R_y(\beta)$ and $R_z(\gamma)$ are defined in equations (4), (5) and (6) respectively.

The position of the camera was measured with a GPS instrument [3]. The orientation was measured with a compass and an elevation scale, which is a weighted arm that hands against an angle scale. GPS accuracy can be as good as a few meters but it varies based on various factors, while the accuracy of orientation in elevation and azimuth is on the order of few degree at most. The accuracy is further compromised because the orientations of both the cameras must be measured or transformed to a common coordinate system. Field conditions with cameras being several hundred meters apart hinders accurate measurements. Further, since the principal axis and camera center are imaginary, and it is generally not possible to know the orientation of true north in the field, therefore it is impossible to measure the location, azimuth, elevation angle and roll of camera with enough accuracy to satisfy the requirements to reconstruct a large scale scene. Therefore, in order to determine earth-relative position and orientation of camera, it is necessary to locate six or more reference landmarks or points to determine the location of these. Reference landmarks should be easily identified in the field and in the image, and distributed evenly in the image. Examples are mountain peaks that are visible from both cameras. Due to large distances in the field, it was not practical to survey the location of landmarks. We use Google Earth (www.googleearth.com) for determining the location of these landmarks.

B. Geometric Error in a Large Scale Scene

Suppose that the homogenous world coordinates of $N \geq 6$ landmarks $\mathbf{X}_i = (x_i, y_i, z_i, 1)^T$ and their corresponding image points with homogenous image coordinates $\mathbf{x}_i = (\hat{x}_i, \hat{y}_i, 1)^T$ with $1 \leq i \leq N$ are known accurately by some means. We can then estimate the camera orientation (α , β and γ ; azimuth, elevation angle and roll angle) and location $\mathbf{C} = (x_0, y_0, z_0)^T$ in world coordinate system by minimizing the geometric error in the image [8]:

$$\epsilon = \min_{\mathbf{P}(\alpha, \beta, \gamma, x_0, y_0, z_0)} \frac{1}{N} \sum_{i=1}^N d^2(\mathbf{x}_i, \mathbf{P}\mathbf{X}_i), \quad (8)$$

where $\mathbf{P}\mathbf{X}_i$ is the exact image of \mathbf{X}_i under perspective matrix \mathbf{P} which depends on orientation (3 degrees of freedom) and location (3 degrees of freedom) of the camera. Therefore, the minimum of the geometric error given by equ.(8) has 6 degrees of freedom. This computation is time-consuming since each search involves several matrix multiplication operations. The geometric error ϵ is the root-mean-square (RMS) error in 2D Euclidean distance. In [8], Hartley and Zisserman recommended Levenberg-Marquardt iterative algorithm to find the minimum of geometric error. In [3], Rasmussen et al. made use of the Polak-Ribiere conjugate gradient method to minimize a cost function which is similar to equation (8) but has 4 only degrees of freedoms (focal length f , azimuth, elevation and roll angle). They assumed that the exact location of the camera center can be measured. However, in our case, finding exact location of centers of cameras that are hundreds and thousands of meter apart in the same coordinate frame was not possible. A few meter deviation of camera center location from the true center results in several degree deviation of its orientation. We show from our analysis in equation (14) that accurate location has significant impact in determining the orientation. Therefore, in solving for geometric error, we must find the global minima instead of a local minima.

In the rest of this subsection, we will discuss the geometric error generated from the deviation of camera from its true orientation and location and then present our coarse-to-fine algorithm to minimize the geometric error (8).

In order to calculate the geometric error resulting from the deviation of orientation and location of the camera, we simplify the analysis by applying a rigid transformation to place the world coordinate in the "true" camera

coordinate system. A rigid transformation does not change the geometric error (8). Note that $\mathbf{C} = (x_0, y_0, z_0)^T$ and $(\alpha, \beta, \gamma)^T$ represent the initial camera location and orientation prior to error correction. Let $\mathbf{C}_t = (x_t, y_t, z_t)^T$ and $(\alpha_t, \beta_t, \gamma_t)^T$ represent the true location and orientation, then $(x_0 - x_t, y_0 - y_t, z_0 - z_t)^T$ represents the location error vector and $(\alpha - \alpha_t, \beta - \beta_t, \gamma - \gamma_t)^T$ represents the orientation error vector. By above transformation, we have $(x_t, y_t, z_t)^T = (0, 0, 0)^T$ and $(\alpha_t, \beta_t, \gamma_t)^T = (0, 0, 0)^T$. Therefore, $(x_0, y_0, z_0)^T$ and $(\alpha, \beta, \gamma)^T$ now represents the deviation or error in location and orientation. We still denote the homogenous world coordinate of $N \geq 6$ landmarks as $\mathbf{X}_i = (x_i, y_i, z_i, 1)^T$ and their corresponding image points with homogenous image coordinates as $\mathbf{x}_i = (\hat{x}_i, \hat{y}_i, 1)^T$ in the rest of subsection. Our assumption is that the initial location and orientation (from measurement) are close to the true location and orientation. Hence, $x_0, y_0, z_0, \alpha, \beta$, and γ are small scalars.

For very small angles, it is useful to rewrite rotation matrices $\mathbf{R}_x(\alpha)$, $\mathbf{R}_y(\beta)$ and $\mathbf{R}_z(\gamma)$, defined by (4), (5), and (6) as the lowest order expansions:

$$\mathbf{R}_x(\alpha) = \mathbf{I}_3 + \alpha \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} + O^2(\alpha) \equiv \mathbf{I}_3 + \alpha \mathbf{J}_x + O^2(\alpha), \quad (9)$$

$$\mathbf{R}_y(\beta) = \mathbf{I}_3 + \beta \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} + O^2(\beta) \equiv \mathbf{I}_3 + \beta \mathbf{J}_y + O^2(\beta), \quad (10)$$

$$\mathbf{R}_z(\gamma) = \mathbf{I}_3 + \gamma \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + O^2(\gamma) \equiv \mathbf{I}_3 + \gamma \mathbf{J}_z + O^2(\gamma). \quad (11)$$

where the big- O symbols indicates the existence of additional terms of second or higher order involving α, β and γ . From (2), (7), (9), (10) and (11), we have

$$\begin{aligned} \mathbf{P}\mathbf{X}_i &= \mathbf{K}(\mathbf{I}_3 + \alpha \mathbf{J}_x + \beta \mathbf{J}_y + \gamma \mathbf{J}_z + O^2(\alpha, \beta, \gamma))([\mathbf{I}_3 | -\mathbf{C}])\mathbf{X}_i \\ &= \mathbf{K} \begin{pmatrix} 1 & \gamma & -\beta \\ -\gamma & 1 & \alpha \\ \beta & -\alpha & 1 \end{pmatrix} \begin{pmatrix} x_i - x_0 \\ y_i - y_0 \\ z_i - z_0 \end{pmatrix} + O^2(\alpha, \beta, \gamma) \\ &= \begin{pmatrix} a_x & 0 & \hat{x}_0 \\ 0 & a_y & \hat{y}_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_i^* + \gamma y_i^* - \beta z_i^* \\ -\gamma x_i^* + y_i^* + \alpha z_i^* \\ \beta x_i^* - \alpha y_i^* + z_i^* \end{pmatrix} + O^2(\alpha, \beta, \gamma) \\ &= \begin{pmatrix} a_x(x_i^* + \gamma y_i^* - \beta z_i^*) + \hat{x}_0(\beta x_i^* - \alpha y_i^* + z_i^*) \\ a_y(-\gamma x_i^* + y_i^* + \alpha z_i^*) + \hat{y}_0(\beta x_i^* - \alpha y_i^* + z_i^*) \\ \beta x_i^* - \alpha y_i^* + z_i^* \end{pmatrix} + O^2(\alpha, \beta, \gamma), \end{aligned} \quad (12)$$

where $x_i^* \equiv x_i - x_0$, $y_i^* \equiv y_i - y_0$ and $z_i^* \equiv z_i - z_0$. Due to the inaccurate estimation of camera's location and orientation, the perspective image of a reference landmark, \mathbf{X}_i , will shift away from its true image. From (12), the shift is given by:

$$\begin{aligned} d^2(\mathbf{x}_i, \mathbf{P}\mathbf{X}_i) &= \left(\hat{x}_i - \hat{x}_0 - \frac{a_x(x_i^* + \gamma y_i^* - \beta z_i^*)}{\beta x_i^* - \alpha y_i^* + z_i^*} \right)^2 \\ &\quad + \left(\hat{y}_i - \hat{y}_0 - \frac{a_y(-\gamma x_i^* + y_i^* + \alpha z_i^*)}{\beta x_i^* - \alpha y_i^* + z_i^*} \right)^2 + O^3(\alpha, \beta, \gamma) \\ &= \left(\frac{a_x x_i}{z_i} - \frac{a_x(x_i^* + \gamma y_i^* - \beta z_i^*)}{\beta x_i^* - \alpha y_i^* + z_i^*} \right)^2 + \left(\frac{a_y y_i}{z_i} - \frac{a_y(-\gamma x_i^* + y_i^* + \alpha z_i^*)}{\beta x_i^* - \alpha y_i^* + z_i^*} \right)^2 + O^3(\alpha, \beta, \gamma) \\ &= \frac{a_x^2}{z_i^4} \left((x_i^2 + z_i^2)\beta - x_i y_i \alpha - y_i z_i \gamma + z_i x_0 - x_i z_0 \right)^2 + \\ &\quad \frac{a_y^2}{z_i^4} \left(x_i y_i \beta - (y_i^2 + z_i^2)\alpha - x_i z_i \gamma + z_i y_0 - y_i z_0 \right)^2 + O^3 \left(\alpha, \beta, \gamma, \frac{x_0}{z_i}, \frac{y_0}{z_i}, \frac{z_0}{z_i} \right), \end{aligned} \quad (13)$$

where we make use of the facts that: (i) \mathbf{x}_i is the true image of landmark \mathbf{X}_i , i.e., $\hat{x}_i = \frac{a_x x_i}{z_i} + \hat{x}_0$ and $\hat{y}_i = \frac{a_y y_i}{z_i} + \hat{y}_0$; (ii) the scenery of clouds is at a very large scale [2], and all landmarks selected are far away from the cameras. This means that the amount of deviation from the true location of the camera is much less than that of the coordinates of all landmarks, i.e., $x_0 \ll x_i$, $y_0 \ll y_i$ and $z_0 \ll z_i$, and so we have $x_i^* \approx x_i$, $y_i^* \approx y_i$ and $z_i^* \approx z_i$ in equ. (13) ($1 \leq i \leq N$). Therefore, the geometric error is dominated by the first two terms in the formula above, and total geometric error in (8) can be rewritten as:

$$\begin{aligned} \epsilon &= \min_{\alpha, \beta, \gamma, x_0, y_0, z_0} \frac{1}{N} \sum_{i=1}^N \frac{a_x^2}{z_i^4} \left((x_i^2 + z_i^2)\beta - x_i y_i \alpha - y_i z_i \gamma + z_i x_0 - x_i z_0 \right)^2 + \\ &\quad \frac{1}{N} \sum_{i=1}^N \frac{a_y^2}{z_i^4} \left(x_i y_i \beta - (y_i^2 + z_i^2)\alpha - x_i z_i \gamma + z_i y_0 - y_i z_0 \right)^2 \\ &\equiv \min_{\alpha, \beta, \gamma, x_0, y_0, z_0} E(\alpha, \beta, \gamma, x_0, y_0, z_0). \end{aligned} \quad (14)$$

The minimum of $E(\alpha, \beta, \gamma, x_0, y_0, z_0)$ can be solved by setting the first derivatives $\partial E/\partial\alpha$, $\partial E/\partial\beta$, $\partial E/\partial\gamma$, $\partial E/\partial x_0$, $\partial E/\partial y_0$, and $\partial E/\partial z_0$ to zero. By using Cramer's rule, there is only one minimum solution $(\alpha, \beta, \gamma)^T = (0, 0, 0)^T$, and $(x_0, y_0, z_0)^T = (0, 0, 0)^T$. Notice that (α, β, γ) and $(x_0, y_0, z_0)^T$ are differences of estimation from true value of orientation and location of a camera, respectively, but not orientation and location of the camera. Thus, the true location and orientation of camera is the only one minimum solution of the geometric error given in equation (8), if the initial estimate of the location and orientation of camera are close enough to the true values. In our case, they are small. Even a few meters deviation away from the true location will have a non-zero solution for orientation of the camera. Therefore, for a large scale scene, it cannot approach the solution for the true orientation. This is why we must consider orientation and location of camera together to minimize the geometric error in eq.(8) instead of considering orientation of camera only to minimize a cost function and assuming that we can obtain the exact location of camera as in [3].

C. Landmarks Survey

Our goal is to determine the azimuth, elevation and roll angles as well as the world coordinates of the two cameras in the same coordinate frame to minimize the geometric error (8). Obviously, the solution of eq.(8) is dependant on coordinates of the landmarks visible in the image(s). Unfortunately, it is very difficult to survey landmarks in the field due to lack of roads, and landmarks being far away. Therefore, we make use of geographic information system (GIS) to determine the location of landmarks. Selected landmarks satisfy two requirements. One is that they are visible from both cameras, and another is that they are able to be measured with some level of accuracy. All kinds of peaks are good candidates. But a image contains hundreds of such peaks and it is still not easy to select a good and sufficient set of landmarks to satisfy our requirement. Google Earth provides a virtual visualization tool which includes high-resolution aerial and satellite imagery and elevation data for the area where our field experiment was conducted. This allows a user to locate the landmarks with help of the 3D virtual terrain and texture views information with high-resolution satellite image (see Fig.II-C), and in turn provides accurate UTM coordinates. The shortcoming of this approach is that it is impossible to automatically match landmarks between camera image and Google Earth 3D terrain.

D. Camera Parameter Estimation By Coarse-To-Fine Algorithm

Based on the discussion above, we present our coarse-to-fine algorithm to compute the true position of a camera. Suppose that the initial orientation $(\alpha_0, \beta_0, \gamma_0)$ and location (x_0, y_0, z_0) of the camera are close to their true position, i.e., there are tolerances $\Delta\alpha$, $\Delta\beta$, and $\Delta\gamma$ for azimuth, elevation and roll angle of camera's principal axis such that the true azimuth is $\alpha \in (\alpha_0 - \Delta\alpha, \alpha_0 + \Delta\alpha)$, the true elevation angle is $\beta \in (\beta_0 - \Delta\beta, \beta_0 + \Delta\beta)$, and the roll angle is $\gamma \in (\gamma_0 - \Delta\gamma, \gamma_0 + \Delta\gamma)$ with a tolerance $(\Delta x, \Delta y, \Delta z)$ such that the true location of camera center is located in the cube $(x_0 \pm \Delta x, y_0 \pm \Delta y, z_0 \pm \Delta z)$. The amount of tolerances are determined by the accuracy of the equipment used. In practice, the tolerance are usually selected as two times as that of error of the equipment [3]. Let initial angle increment be θ and location increment τ . Our coarse-to-fine algorithm is as follows:

- Step 1: Set tolerances $\Delta x, \Delta y, \Delta z$ and $\Delta x, \Delta y, \Delta z$, increment θ , τ , and k_j ($j = 1, \dots, 6$).

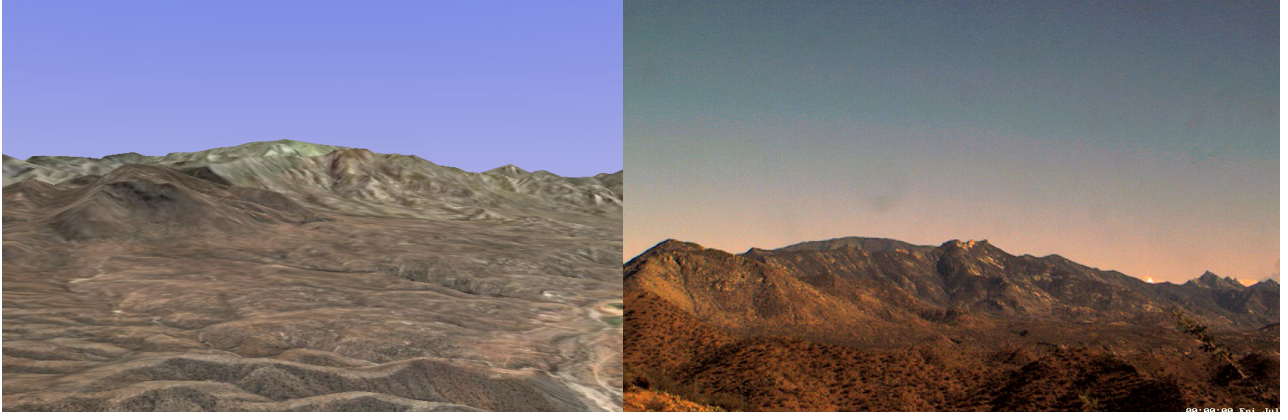


Fig. 2. Terrain views from Google Earth (left) and our camera snapshot (right).

- Step 2: Set $\alpha = \alpha_0 - \Delta\alpha + k_1\theta$, $\beta = \beta_0 - \Delta\beta + k_2\theta$, and $\gamma = \gamma_0 - \Delta\gamma + k_3\theta$. Compute the rotation matrix (7).
- Step 3: Set $x = x_0 - \Delta x + k_4\tau$, $y = y_0 - \Delta y + k_5\tau$, and $z = z_0 - \Delta z + k_6\tau$, and calculate perspective matrix (2).
- Step 4: Calculate the geometric error (8), and achieve the position of camera that minimizes the geometric error till now. Increase k_j ($j = 4, 5, 6$) by 1 till all parameters of the camera's orientation and location reach their maximum;
- Step 5: If the solution satisfies the requirement of accuracy, then stop, ; else replace the position of the camera $\alpha_0, \beta_0, \gamma_0, x_0, y_0, z_0$ with the minimum position, and set $\Delta\xi = \Delta\xi/C$ ($\xi = \alpha, \beta, \gamma, x, y, z$) and $\theta = \theta/C, \tau = \tau/C$, and go to Step 2.

where C is a constant larger than 1 . The purpose in setting $\Delta\xi = \Delta\xi/C$ ($\xi = \alpha, \beta, \gamma, x, y, z$)^T and $\theta = \theta/C, \tau = \tau/C$ is to obtain a higher precision. In this paper, we use $C = 5$, tolerance are 20° for orientation and 40m for location)

Not only does this algorithm assure that the minimum solution is not local but global and also computable. It is also more efficient than searching the whole space which costs significant computational time to reach the minimum solution.

E. 3D Reconstruction

Once we know the perspective matrices \mathbf{P}_1 and \mathbf{P}_2 of the two cameras and a set of given image correspondences \mathbf{x}_i^1 and \mathbf{x}_i^2 (which come from a set of unknown 3D points \mathbf{X}_i), the 3D positions can be solved by linear triangulation methods (see [8]). For each camera, we have a measurement $\mathbf{P}_1\mathbf{X}_i = \mathbf{x}_i^1$, $\mathbf{P}_2\mathbf{X}_i = \mathbf{x}_i^2$ and these equations can be combined into the form $\mathbf{A}\mathbf{X}_i = \mathbf{0}$, which is a linear equation in \mathbf{X}_i , with

$$\mathbf{A} = \begin{pmatrix} \hat{x}_i^1 \mathbf{p}_1^{3T} - \mathbf{p}_1^{1T} \\ \hat{y}_i^1 \mathbf{p}_1^{3T} - \mathbf{p}_1^{2T} \\ \hat{x}_i^2 \mathbf{p}_2^{3T} - \mathbf{p}_2^{1T} \\ \hat{y}_i^2 \mathbf{p}_2^{3T} - \mathbf{p}_2^{2T} \end{pmatrix}, \quad (15)$$

where \mathbf{p}_i^{jT} are the j^{th} row of \mathbf{P}_i . The two equations are included from each camera giving a total of four equations in four homogeneous unknown coordinates. This is a redundant set of equations because the rank is 3 and the solution is determined only up to a scale. The set of equations is solved using SVD method. This solution is not in the world coordinate system. We apply linear scaling factors obtained from sampling landmark points in real world coordinate system.

III. EXPERIMENTAL RESULTS

A. Accuracy Analysis of 3D Reconstruction

In the large scale scene such as our application, the accuracy of stereo photogrammetry or 3D reconstruction is a significant issue, since a pixel in an image corresponds to a square of 16m side at a distance of 16km from the camera. We compute the error between the calculated UTM coordinates of 10 landmarks and their world UTM coordinates obtained from Google Earth. The accuracy of stereo photogrammetry depends on the estimation of the intrinsic and extrinsic parameter of cameras used, selection of matching pair of pixels from images, and measurement of UTM coordinates of landmarks for the estimation of extrinsic parameters. We assume that both measurement and selection have inherent errors, and assume that the errors follow Gaussian distributions with zero mean. Therefore, more landmarks sampled, more accurate estimation of extrinsic parameters and resulting 3d reconstruction.

As an example for one of the cameras used, the location and orientation were measured in the field. Its UTM coordinates are 3604495m , 514614m and 1177m respectively, and its azimuth, elevation angle and roll angles are 155°, 19° and 0°, respectively. By using coarse-to-fine algorithm with 19 landmarks, the UTM coordinates were calculated as 3604491m, 514610m and 1186m, respectively, and its azimuth, elevation angle and roll angle calculated as $161.3 \pm 0.05^\circ$, $5.2 \pm 0.05^\circ$ and $18.1 \pm 0.05^\circ$, respectively.

We selected additional 11 landmarks different from those used to calculate extrinsic parameters. We calculate the mean error (ME) between the calculated UTM coordinates of 11 landmarks and their world UTM coordinates obtained from Google Earth as follows:

$$ME(\xi) = \frac{1}{N} \sum_{i=1}^N |\xi_i - \bar{\xi}_i| \quad (16)$$

where N is the number of landmarks sampled, ξ_i (either x_i or y_i or z_i) denotes UTM coordinates of the i th landmarks observed from google earth and $\bar{\xi}_i$ (either \bar{x}_i or \bar{y}_i or \bar{z}_i) denotes UTM coordinates from computation. The mean errors in the values were 420m (northing), 105m (easting) and 27m (vertical). The anisotropy in the errors is due to the uncertainty in locating features on the camera images and matching them with the digital elevation map. The mountain ridges in the camera's field of view are oriented essentially in the north-south directions. Points in the vertical and east-west directions are on the tops of the ridge, while in the north-south direction they are along the ridge where locating a point on the DEM uniquely is more difficult. Still, the mean error is on the order of 20 pixels and is sufficient for the stereo analysis.

B. Stereo Reconstruction of Cloud

An example of cloud structure obtained with the technique described in the preceding section is shown in Figures 3 and 4. These correspond to a case of isolated convective development that occurred on July 26, 2005 over the Santa Catalina Mountains in southern Arizona. This date was chosen due to the ease of identification of cumulus over the Santa Catalinas and the ability to match elements on the digital images with those on the operational WSR-88D radar in Tucson, AZ. The radar is located about 60 km from peak over which the clouds develop. Composite reflectivity (i.e. the maximum reflectivity that occurs in the vertical column) at 1 km resolution was used and while this is coarse compared to the pixel resolution of the cameras, it provides us with the only validation of the stereo analysis scheme currently available.

Figure 3a,b shows the plan and cross sections of the points indicated on the image from the left hand camera (CC2). A three-dimensional perspective is shown in Figure 3d. Note that in Fig. 3b the horizontal axis is reversed, corresponding to a view looking toward the south, which is in the same direction as the view from the cameras. This time corresponds to 17:00 UTC (10:00 Mountain Standard time) and corresponds to the first available return from the Tucson radar. The cluster of points in the center coincides with the region of the highest reflectivity (30 - 40 dBz). There is a fair amount of noise in the radar data, especially in the low dBz range, so the correspondence is not exact, but the fit with the radar and the coherence of the cluster of cloud points to suggest that the algorithm performs well.

Another basis for comparison is the height of the cloud tops. By 17:48 UTC on July 26, 2006 there is evidence of a thunderstorm. The vertical temperature profile from the field sounding data indicates that the tropopause (a stable region where temperatures begin to increase with height) is located at the 16,000 m level, which coincides with the points at the top of the thunderstorm computed using our algorithm at 15,998 m. Figures 3 indicates that the stereo analysis technique described here captures both the horizontal and vertical structure of the cloud. Data

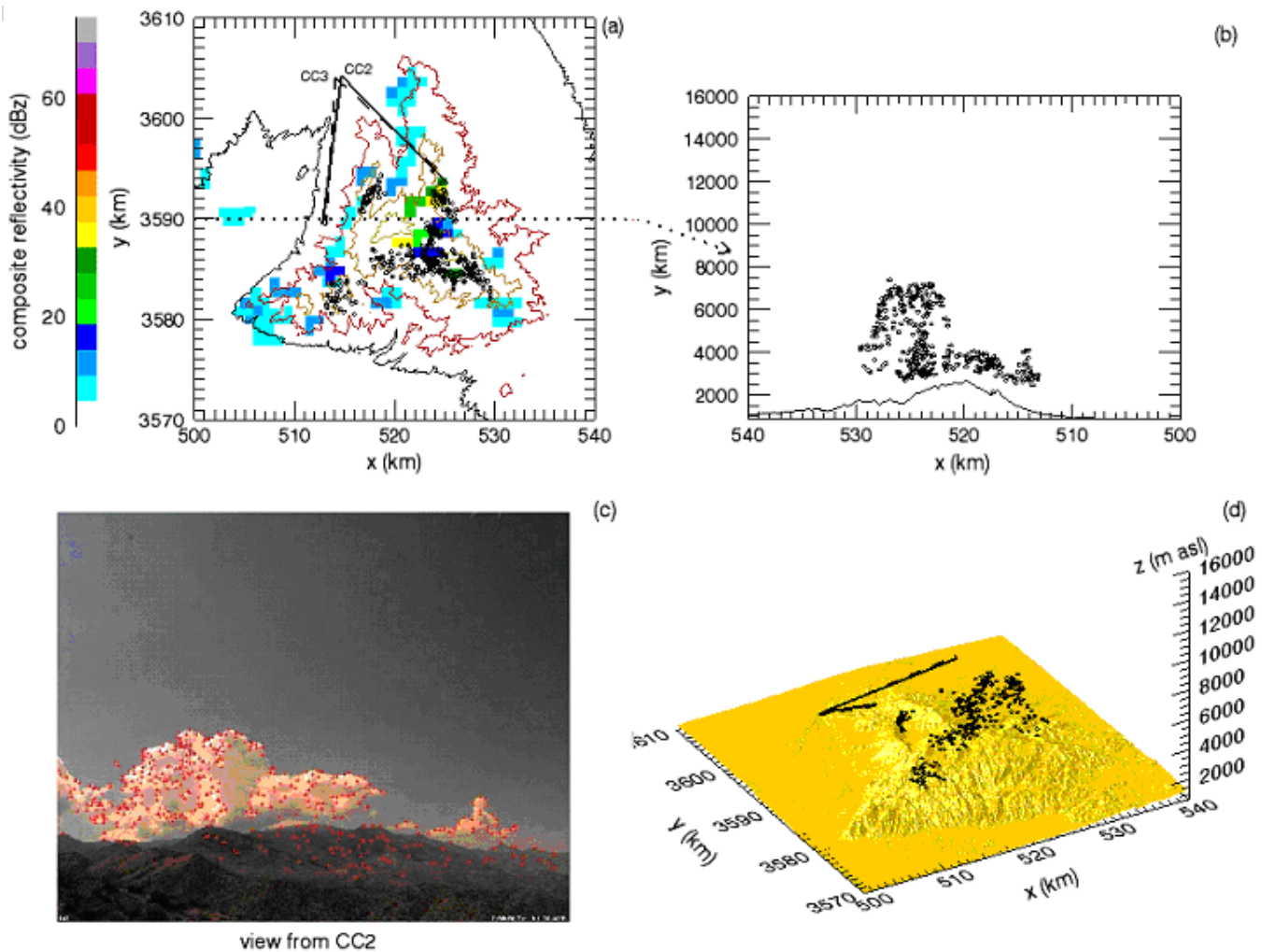


Fig. 3. Stereo reconstruction for July 26, 2005 at 17:00 UTC. (a) is the plan section of all points superimposed on the TUS composite reflectivity. Camera location and field of view are indicated. (b) is the cross section of all points with terrain cross section through dotted line in (a). (c) is the view from CC2 which looks toward the south. (d) is a 3-dimensional perspective of points on DEM of the topography.

for a more comprehensive validation of the scheme will be collected in summer 2006 using an aircraft with an airborne Doppler radar capable of 30 meter resolution. This is comparable with the pixel resolution of the cameras.

For detailed analysis and application to orographic convection the reader is referred to [1] and [6].

IV. CONCLUSION

We have developed and presented a theoretical basis for the analysis of geometric error for a large scale scene and the important role of accurate location in minimizing the geometric error. We have developed an algorithm that accurately deduces the orientation (including azimuth, elevation angle and roll angle of optical axis) and UTM location of cameras by starting from initial field measurement in the same world coordinate with survey of a sufficient landmarks (≥ 6). With the accurate orientation and location of cameras, accurate 3D positions of cumulus clouds in both images are determined by triangulation method. Experimental results show that terrestrial position of clouds matched with the composite reflectivity from the WSR-88D radar and the calculated heights agree with the tropopause height in the sounding. Stereo photogrammetric analysis allow us to determine the initial location of the convection, calculate vertical velocity of cloud tops for selected elements and the timing of the transition from shallow to deep convection.

Optical effects such as barrel distortion are not accommodated in our algorithm. Automated stereo matching for clouds in both images needs further work.

Acknowledgments: This work was supported by a grant from the National Science Foundation (#ATM-0352988).

REFERENCES

- [1] J. A. Zehnder, J. Hu, and A. Razdan, "A stereo photogrammetric technique applied to orographic convection," *Mon. Wea. Rev.*, Submitted, 2006.
- [2] C. Warner, J. Renick, M. W. Balshaw, , and R. H. Douglas, "Stereo photogrammetry of cumulonimbus clouds," *J. R. Met. Soc.*, vol. 99, pp. 105–115, 1973.
- [3] E. N. Rasmussen, R. Davises-Jones, and R. L. Holle, "Terrestrial photogrammetry of weather images acquired in uncontrolled circumstances," *J. Atmospheric and Oceanic Technology*, vol. 20, pp. 1790–1803, 2003.
- [4] A. R. Kassander and L. Sims, "Cloud photogrammetry with ground-located k-17 aerial cameras," *J. Met.*, vol. 9, pp. 43–49, 1957.
- [5] R. W. Shaw, "Rotating-lens stereo cloud photogrammetry," *Sci. Rep. MW-62, Stormy Weather Group, McGill University, Mootreal*, p. 16, 1969.
- [6] J. A. Zehnder, L. Zhang, D. Hansford, N. Selover, and M. Brown, "Using digital cloud photogrammetry to characterize the onset and transition from shallow to deep convection over orography," *Mon. Wea. Rev.*, vol. in press, 2006.
- [7] R. Mohan, G. Medioni, and R. Nevatia, "Stereo error detection correction and evaluation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 2, pp. 113–120, 1989.
- [8] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge University Press, (2nd ed) 2003.
- [9] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography," *Association and Computing Machine*, vol. 24, no. 6, pp. 381–395, 1981.
- [10] R. Y. Tsai and T. S. Huang, "Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 1, pp. 13–27, 1984.
- [11] M. Pollefeys, L. V. Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch, "Visual modeling with a hand-held camera," *International Journal of Computer Vision*, vol. 11, no. 3, pp. 207–232, 2004.