Error Analysis and Design Considerations for Stereo Vision Systems Used to Analyze Animal Behavior

Gordon Towne [1], Diane H. Theriault [1], Zheng Wu [1], Nathan Fuller [2],
Thomas H. Kunz[2], and Margrit Betke[1]

[1] Image and Video Computing Group, Department of Computer Science, Boston University

[2] Center for Ecology and Conservation Biology, Department of Biology, Boston University

Abstract

This paper presents an analysis of patterns of error in the triangulation of 3D points from stereo camera systems that are used in field work to study the behavior of bats in flight. A measure of the error present in a 3D reconstruction is proposed. A method for empirically testing the performance of a particular stereo camera configuration through a software simulation is presented. Randomly generated 3D calibration points are projected onto the image planes of simulated cameras, and autocalibration is performed using the direct linear transform (DLT) method. The accuracy of the computed 3D reconstruction is determined by computing the proposed measure of geometric error over a grid of reference scene points. A series of experiments are performed with this simulator to evaluate the accuracy of 3D reconstruction with various camera placements and under differing levels of noise. Results of these experiments are used to motivate suggestions for the design and calibration procedure of stereo vision systems used in ecology field work.

1. Introduction

A multi-camera tracking system [7, 8] has been developed that has been used by field biologists to monitor the behavior of wildlife. The system is targeted to track bats and birds within multi-camera views and analyze the flight behavior of large groups. Flight trajectories of individuals of these groups are reconstructed in three dimensions. The flight trajectory data computed by the system can be used to enhance understanding of the group behavior of animal populations, with applications to behavioral ecology, biological engineering, and conservation. Our system has been used to reveal flight characteristics and group structure of a colony of Brazilian free-tailed bats (*Tadarida brasiliensis*) [6].

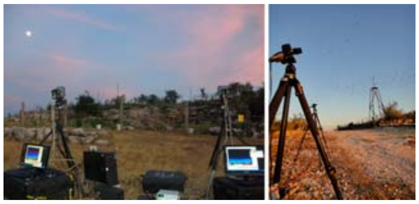

Figure 1: Sample stereo setups of cameras recording multi-view data of bat colonies in Texas.

Obtaining the flight trajectory data requires the implementation of algorithms capable of tracking flying bats or birds given calibrated multi-camera views of groups possibly containing hundreds or thousands of such individuals. Multi-view, multi-target tracking is a difficult problem in general as data association must be performed across time and across all camera views [7]. Since many targets must be tracked simultaneously within dense groups, occlusions between individuals occur frequently. The similar appearance of individual bats or birds within these groups also limits the extent to which appearance information can be leveraged to perform data association.

The approach by Wu et al. [7] uses deferred-logic, computing an initial solution to the spatio-temporal association problem at each time step, and subsequently improving the solution by considering the stability of associations across multiple time steps. The success of this method relies on accurate triangulation of 3D points in order to perform the data association. The methods proposed in this paper for estimating the reconstruction error in a multi-camera setup will enable more accurate trajectories to be generated by allowing reasonable estimates for the error in point reconstruction to be incorporated into the tracking algorithms. We will also provide strategies to position cameras in such a way as to minimize this error in future data collection. Common camera calibration methods [2] are applicable to small calibration volumes that are close to the cameras and do not easily scale to wild-animal monitoring where a camera setup has a baseline of several meters and where the targets of interest are hundreds of meters away. Here, a small numerical issue or noise during calibration results in significant reconstruction errors. In this paper, we adopt a wand-based self-calibration method that is easy to implement in the field. We analyze its performance with different camera configurations.

2 Geometric Error Measure

We propose the following measure of the geometric error present in a reconstructed scene. Assume a set of reference scene points with known 3D coordinates,

$S = \{s_1, s_2, ...s_n\}$, is given. For each $s_i \in S$, let $N(s_i)$ be those points immediately neighboring s_i in the reference scene. In our simulation experiment, reference points form an evenly spaced 3D grid and the neighboring points are the six closest neighbors on the grid. Each of the reference points $s_i \in S$ has a corresponding point $r_i \in R$, where r_i is determined by reconstructing the position of s_i from its imaged coordinates. We define the geometric error for a point $r_i \in R$ to be

$$E_i = 1/|N(s_i)| \sum_{s_j \in N(s_i)} |\|s_i - s_j\| - \|r_i - r_j\||. \quad (1)$$

That is, E_i is a measure of the discrepancy in the Euclidean distance between neighboring reference points s_i and s_j and their corresponding reconstructed points r_i and r_j. We then define the geometric error for the whole scene reconstruction to be

$$E = 1/n \sum_{i=1}^{n} E_i, \quad (2)$$

the mean of the error for each reconstructed reference point. This definition of the error provides a measure of the local spatial integrity of the reconstructed scene with respect to the reference scene. It is convenient in that its value is easily human-interpretable. When calculated over a set of evenly spaced reference points, the geometric error provides the expectation of the error in world coordinates over measured distances in the reconstructed scene. In contrast to the commonly-used error measure $\|s_i - r_i\|$, our geometric measure is translation and rotation invariant. It focuses on the overall spatial integrity. That is, if the reconstructed points and their corresponding reference points only differ in a global translation and rotation transformation, the reconstruction error could be very large but the geometric error will be small. The latter is a desired property because a rigid transformation does not affect many data analysis tasks, such as estimation of the speed of the bats and their relative distances.

3 Simulator Design

We developed a software system to test proposed camera setups and evaluate the expected accuracy of a resulting reconstruction. This system takes as input the intrinsic parameters and proposed configurations of two cameras and simulates the calibration workflow by

1. randomly generating 3D calibration points in a sphere of interest around a specified target point[1],
2. projecting the calibration points onto the image planes of the two virtual cameras,
3. using these points as input to the DLT method [2] to perform autocalibration of the camera setup,
4. evaluating the reconstruction accuracy by computing the geometric error over a set of evenly spaced reference points.

The reconstruction accuracy, E, is evaluated by calculating the geometric error between a set of evenly-spaced reference scene points that fill the view volume. Each of these reference points is projected onto the virtual cameras, and a corresponding reconstructed point is triangulated in the scene using the parameters computed in the autocalibration step.

The two simulated cameras use a pinhole model, where each camera is parametrized by its focal length f in millimeters, sensor size in millimeters, and resolution in pixels. Lens effects such as radial distortion and decentering are not taken into account

The simulator implements a method of autocalibration of a stereo camera configuration using the DLT method to solve for the intrinsic and extrinsic camera parameters at the same time [2]. The specific procedure for calibration involves imaging a wand of known length w at various locations in the scene. For each imaged location, the two endpoints p_1 and p_2 of the wand map to points $c_{1,L}$ and $c_{2,L}$, respectively, on the image plane of the "left" camera, and $c_{1,R}$ and $c_{2,R}$ on the image plane of the "right" camera. The corresponding points $c_{1,L}$ and $c_{1,R}$, and $c_{2,L}$ and $c_{2,R}$ are then localized at each imaged position from all viewpoints and given as input to the DLT method. This results in a reconstruction of the scene up to scale, which can be determined by scaling the reconstructed volume so that triangulated positions for the wand endpoints are consistent with known wand length w.

4 Experiments

To narrow the scope of the analysis to camera setups that resemble those commonly seen in practice, we designed experiments with the following limiting assumptions (which are are also representative of those made in the literature [1, 4, 5]):

1. Use two cameras ("left" and "right") with identical intrinsic parameters.
2. Position the cameras equal distance from a predefined target point in the scene.
3. Orient the cameras such that their optical axes intersect at the target point.

In the simulated experiments presented, a series of 3D calibration points were generated uniformly at random within a sphere of radius r around target scene point t, which defines the region of interest in the scene. These calibration points were used as input to the DLT method

[1] Throughout this paper, the *target* refers to the calibration object, not the scene points to be reconstructed during testing.

to calibrate the camera setup. For each of these calibration points, a small random perturbation whose standard deviation is 3 pixels was added to the projected location on the image plane of each of the cameras to simulate localization error present in real-world applications. After performing autocalibration of the simulated camera setup, the accuracy of the resulting reconstruction was determined by computing the geometric error, E_i, over a grid of uniformly spaced reference points that fill the overlapping fields of view of the two cameras.

4.1 Selection of Camera Parameters

An experiment was designed to empirically determine the optimal distance between cameras given a predetermined distance to the calibration object. The distance d from the camera to the object and the baseline b between cameras were varied, and multiple trial simulations performed for each combination. Identical cameras with focal length f=50 mm, diagonal sensor size of 35 mm and pixel resolution 1600×1200 pixels were used. In each trial, unique wand calibration points were generated in a region of interest around target point t with r=7 m. Autocalibration was performed using the wand-based self-calibration method. The resulting reconstruction was evaluated by computing the geometric error over a grid of reference scene points spaced 3 m apart. The trend shown in Fig. 2 indicates that as the distance d increases, the baseline b required to maintain accurate resolution increases nonlinearly within a range. The geometric error throughout the reconstructed scene, E, will be reduced as both b and d increase (Fig. 3).

4.2 Tolerance to Calibration Noise

An experiment was designed to determine the tolerance of the DLT autocalibration method to errors in the location of calibration points on image planes. It is common for the location of a calibration object to be manually annotated in images from multiple viewpoints, and thus there is a potential for errors to be introduced by labeling the location of a reference object at a small offset from its true projected position. Multiple simulations were run with identical camera setups (parameters as above) but increasing magnitude of noise in calibration point location. For each trial, a value e_p was fixed, and the location of each calibration point on both imaged planes was offset from its true location by e_p pixel-widths in a random direction. The value of e_p was tested on the range [0,20] pixels, and multiple simulations were run for each value of e_p. For each trial, the geometric error E in reconstruction resulting from the autocalibration was determined.

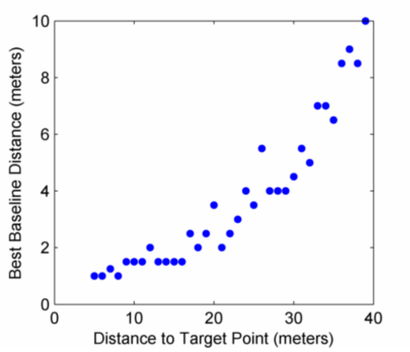

Figure 2: Best Baseline. Results of simulated experiments with varying quantities for baseline b and distance-to-target d. For each value of d, the baseline b that resulted in the lowest geometric error E in the reconstruction is shown.

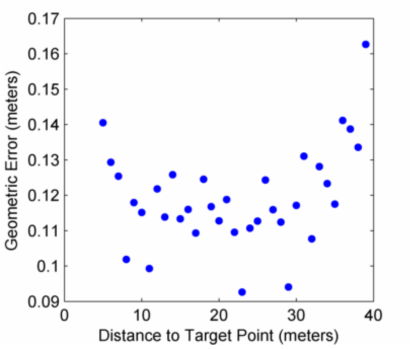

Figure 3: Error in Best Reconstruction. Results of simulated experiments with varying quantities for b and d. For each value of d, the lowest geometric error E observed for any baseline b is shown.

Our results indicate that the DLT calibration method is tolerant of consistent noise in the location of calibration points up to a 6 pixel offset, as the magnitude of the geometric error remains relatively constant through this range. With larger offsets, the calibration procedure often fails to converge on a configuration that models the simulated camera setup, and the geometric error of the resulting reconstruction increases.

4.3 Variation of Accuracy Within Volume

It is interesting to examine how the local error E_i varies spatially over the reconstructed volume. A particular camera setup might result in a scene reconstruction with high accuracy in a given region of interest, but low accuracy in other regions. This would result in a high geometric error E over the whole view volume, but could still provide useful 3D data for a specific application.

We generated two examples of the variation of accuracy over a reconstructed volume using two identical sets of camera parameters (Figure 4). The first simulation was run with parameters d =25 m and b =5 m , while the second was with d =35 m and b =8 m. Note that both are representative of the best camera configuration for a specific target distance from Figure 2.

Figure 4 shows the variation in geometric error spa-

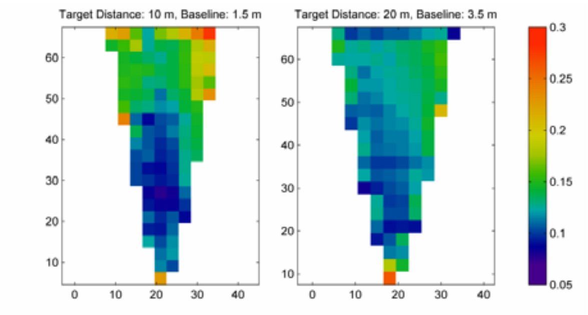

Figure 4: Error Pattern over Reconstructed Space. Geometric error E_i averaged over columns of the reference point grid where $d = 25$ m and $b = 5$ m (left) and $d = 35$ m and $b = 8$ m (right). Color represents magnitude of E_i. The error increases at the fringes of the scene and is slightly asymmetric due to the autocalibration algorithm arbitrarily selecting the right camera as a reference.

tially distributed throughout the reconstructed volume, where the error displayed is averaged over columns of the grid of reference points. It further illuminates the results shown in Figure 3 in that the camera setup with larger values for b and d results in not only greater overall reconstruction accuracy, but also more uniform accuracy over the reconstructed space.

5 Discussion

From our experiments, we can draw certain conclusions about best-practices for the configuration of stereo camera setups. The relationship shown in Figure 2 suggests that the best choice of camera baseline length increases nonlinearly with the distance from the cameras to the target scene point. It is also clear by the results shown in Figure 3 that the overall accuracy of 3D reconstruction increases as both baseline length and distance to target increase. These relationships can be leveraged to narrow the range of possible camera locations for a given application. In practice it is often the case that the distance between the imaging location and the region of interest is predetermined by the working environment. Thus the relationship in Figure 2 can be used to select an appropriate baseline length between cameras for a given application. However, it is important to note that the specific baseline length for optimal reconstruction may change with the internal camera parameters and reconstruction method. That is, for the particular set of assumptions outlined in the simulated experiments, Figure 2 provides the baseline length that is optimal for a specific distance to the scene target. Under different assumptions - eg, different focal lengths or different method of reconstruction - the specific value for the optimal baseline length may change.

The results (Sect. 4.2) indicate that a small amount of of noise in the location of calibration points can cause the DLT method to fail to converge on camera parameters that accurately model the true configuration. This highlights that care must be taken in localizing calibration points. This knowledge should be incorporated into the calibration workflow by selecting a physical calibration object with visibly distinguishable markers that map to as small a region on the image planes as possible while still being consistently visible in all views.

The patterns in Figure 4 suggest that reconstruction accuracy is consistently highest in the region of the image closest to where calibration data were obtained. Thus, in practice, care should be taken to obtain calibration points as close to the region of interest of the scene as possible. Since the accuracy of reconstruction deteriorates as the distance from the region of calibration data increases, it is also desirable to obtain calibration data from as broad a region of the scene as possible to avoid significant non-uniformity in accuracy of reconstructed points from different locations in the scene. If possible, it is desirable to obtain calibration data from points throughout the view volume, but it is particularly important to do so in the region where greatest reconstruction accuracy is required.

Acknowledgements. This material is based in part upon work supported by the NSF (IIS-0910908 and IIS-0855065) and ONR (024-205-1927-5). We would like to thank Ty Hedrick [3] and Stan Sclaroff for their assistance.

References

[1] S. D. Blostein and T. S. Huang. Error analysis in stereo determination of 3-d point positions. *IEEE Trans Pattern Anal Machine Intell*, 9(6):752–765, 1987.

[2] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.

[3] T. L. Hedrick. Software techniques for two- and three-dimensional kinematic measurements of biological and biomimetic systems. *Bioinspir Biomim*, 3(3):6 pp, 2008.

[4] J. Liu, Y. Zhang, and Z. Li. Selection of cameras setup geometry parameters in binocular stereovision. *IEEE Conf. on Robotics, Automation and Mechatronics*, 2006.

[5] G. Olague and R. Mohr. Optimal camera placement to obtain accurate 3d point positions. In *Proceedings of the Fourteenth International Conference on Pattern Recognition, Vol. 1*, pages 8–10, 1998.

[6] D. H. Theriault, Z. Wu, N. I. Hristov, S. M. Swartz, K. S. Breuer, T. H. Kunz, and M. Betke. Reconstruction and analysis of 3D trajectories of Brazilian free-tailed bats in flight. In *Workshop on Visual Observation and Analysis of Animal and Insect Behavior, Turkey*, 2010. 4 pp.

[7] Z. Wu, N. I. Hristov, T. L. Hedrick, T. H. Kunz, and M. Betke. Tracking a large number of objects from multiple views. In *International Conference on Computer Vision (ICCV)*, Kyoto, Japan, 2009. 8 pp.

[8] Z. Wu, N. I. Hristov, T. H. Kunz, and M. Betke. Tracking-Reconstruction or Reconstruction-Tracking? Comparison of two multiple hypothesis tracking approaches to interpret 3D object motion from several camera views. In *IEEE Workshop on Motion and Video Computing*, 2009.