Exact Bias Correction and Covariance Estimation for Stereo Vision

Charles Freundlich[1], Michael Zavlanos[1] and Philippos Mordohai[2]
[1]Duke University. [2]Stevens Institute of Technology.

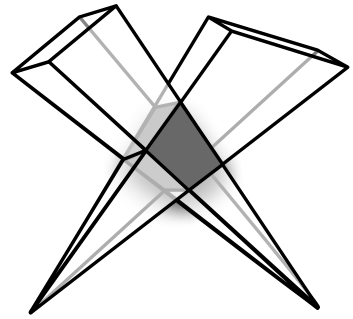
We present an approach for correcting the bias in 3D reconstruction of points imaged by a calibrated stereo rig. Our analysis is based on the observation that, due to quantization error, a 3D point reconstructed by triangulation essentially represents an entire region in space. The true location of the world point that generated the triangulated point could be anywhere in this region. We argue that the reconstructed point, if it is to represent this region in space without bias, should be located at the centroid of this region, which is not what has been done in the literature. We derive the exact geometry of these regions in space, which we call 3D cells, and we show how they can be viewed as uniform distributions of possible pre-images of the pair of corresponding pixels. By assuming a uniform distribution of points in 3D, as opposed to a uniform distribution of the projections of these 3D points on the images, we arrive at a fast and exact computation of the triangulation bias in each cell. In addition, we derive the exact covariance matrices of the 3D cells. We validate our approach in a variety of simulations ranging from 3D reconstruction to camera localization and relative motion estimation. In all cases, we are able to demonstrate a marked improvement compared to conventional techniques for small disparity values, for which bias is significant and the required corrections are large.

Pixelation, or quantization, error causes digital cameras to map distinct points from the continuous world to the same pixel in the image. In the case of stereo vision, sets of points in the world, which we hereafter refer to as *3D cells*, are mapped to discrete pairs of pixels in the two cameras. Since all points in a cell are indistinguishable after the projection, typically they are all mapped to the same 3D point after reconstruction. In classical stereo vision the reconstructed point is the intersection of two rays that pass through the two camera centers and the two pixel centers. It represents an entire region formed by the intersection of two infinitely long pyramids created by the camera centers and the pixels, shown in Fig. 1(a). The sides of each pyramid are formed by the planes defined by the camera centers and edges of the pixels. 3D cells are larger, more elongated and asymmetric the further away from the cameras they are. This means that points far from the cameras are often subject to large error.

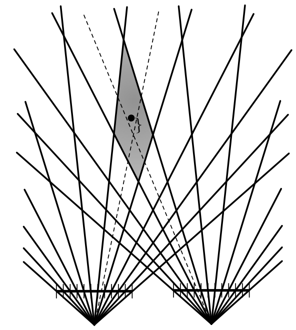
Assuming that points in the 3D cell are uniformly distributed in space, then the expected value of the reconstructed point is the first moment, or the centroid, of the 3D cell. Conventional stereo vision, however, does not use the centroid as the reconstructed point, causing systematic error. Several authors [1, 5, 7] have reported the bias in long range stereo vision, but to the best of our knowledge, the treatment and correction to the systematic error proposed here is novel. Our approach differs from previous work in that it is an exact and computationally simple solution that is able to remove the bias. An illustration of the proposed solution can be seen in Fig. 1(b) which shows the intersection of the rays (dashed lines) and the centroid of the 3D cell (dark dot). The distance between the two is the bias of conventional reconstruction. The bias can be computed as shown in the paper and the coordinates of the reconstructed point can be corrected.

After deriving the correction, we demonstrate that it completely removes the bias in three types of simulation. The first type of simulation examines the errors in 3D reconstruction by generating a set of 3D points, projecting them on the images and reconstructing them using the conventional and the proposed approach. In the second type of simulation, we localize a pair of cameras that observe a set of known 3D landmarks using Horn's absolute orientation algorithm [3]. In this case, the 3D coordinates of the landmarks are known, but the reconstructed coordinates of the observed landmarks are corrupted by quantization noise due to pixelation. In the third type of simulation, the relative motion of a camera pair is estimated based on two observations of an unknown set of 3D landmarks. The cameras take a pair of images of the landmarks, move to a new location and take a second pair of images. Motion is estimated by reconstructing the observed landmarks and applying Horn's algorithm on the two sets of 3D points. In all cases, bias is observed for conventional reconstruction while it is entirely removed after the proposed correction.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.



(a) Illustration of a 3D cell



(b) Cross-section of the cell

Figure 1: Illustrations of quantization error in stereo vision. (a) a 3D cell created by the intersection of the viewing cones (pyramids) emanating from corresponding pixels in the left and right camera. (b) Cross-section of the 3D cell. The dark dot is the centroid of the cell and should be used as the reconstructed point, instead of the intersection of the rays that is currently used.

The second contribution of this paper is a novel way to derive the covariance matrix of a reconstructed point by computing the second moments of a uniform distribution in the corresponding 3D cell. This estimate is exact and more accurate than common approximations that propagate uncertainty from the image plane to the reconstructed points under Gaussian assumptions [4]. Note that higher order moments of these distributions exist, but we ignore them in this work. We propose a test for the validity of our new covariance estimation method and show that it is indeed superior to conventional covariance propagation [4].

Our findings are directly applicable to stereo matching algorithms that treat disparity as a discrete variable, including most of the top performing methods participating in the evaluations hosted by Middlebury [6] and KITTI [2] that use Markov Random Fields or Semi-Global Matching for optimization. We leave the analysis of matching methods that treat disparity as a continuous variable by fitting planes to image patches or by relying on variational techniques, for future work.

[1] S. D. Blostein and T. S. Huang. Error analysis in stereo determination of 3-d point positions. *PAMI*, 9(6):752–766, 1987.

[2] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the KITTI vision benchmark suite. In *CVPR*, pages 3354–3361, 2012.

[3] B.K.P. Horn. Closed form solutions of absolute orientation using orthonormal matrices. *Journal of the Optical Society of America A*, 5(7): 1127–1135, 1987.

[4] L. H. Matthies and S. A. Shafer. Error modelling in stereo navigation. *IEEE Journal of Robotics and Automation*, 3(3):239–250, 1987.

[5] J. J. Rodriguez and J. K. Aggarwal. Stochastic analysis of stereo quantization error. *PAMI*, 12(5):467–470, 1990.

[6] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47(1-3):7–42, 2002.

[7] G. Sibley, L. Matthies, and G. Sukhatme. Bias reduction and filter convergence for long range stereo. In *Robotics Research*, volume 28, pages 285–294, 2007.