A Framework for Calibrating and Benchmarking Computer Vision Algorithms in Surgical Robotics[1]

Rodney Dockter
Department of Mechanical Engineering,
University of Minnesota,
Minneapolis, MN 55455

Timothy M. Kowalewski
Department of Mechanical Engineering,
University of Minnesota,
Minneapolis, MN 55455

1 Background

The use of computer vision techniques in the medical field has been commonplace since the 1960's when image analysis was used for chromosome pairing [1]. The used of these techniques in recent years has grown to encompass endoscopy, tool inventory recording, and surgical tool tracking in vivo.

Given the potentially life altering stakes for these medical applications, a high degree of accuracy is required of any computer vision algorithm. One of the key elements in three-dimensional algorithms is accurately converting image values (in pixel space) into real world Cartesian coordinates. This conversion requires an accurate and straightforward calibration routine in order to identify the correct transformation. Once the calibration is complete it is also important for the developers to know the accuracy.

While there are existing calibration routines which follow a linear (pinhole) model [2,3], these methods do not work well for varied camera optics and do not handle image distortion well given the linear transformation matrix. Standard stereo correspondence methods were implemented by Dockter and Kowalewski [4] and were found insufficient in terms of speed and accuracy. Given the aforementioned needs, we present a method for calibration and benchmarking for stereo computer vision algorithms which utilizes basic printable objects and a standard machinists lathe and can accommodate more complex camera modeling. Both the calibration and benchmarking boards can be produced using 3D printing techniques and mounted on a machinists lathe thus making high accuracy calibration achievable using available means. This method is directly applicable to the high-distortion stereo optics common in surgical robotics.

1.1 Methods—Calibration. In a standard stereo camera configuration, two cameras are situated side by side with parallel optical axes. For a given object pixel location in each right and left camera, the depth of the object can be inferred by calculating the disparity (the pixel location difference from each channel). In order to know the correct mapping between disparity and depth, a calibration routine is used.

In order to accomplish this, a custom calibration board was designed with cross hatches arranged in a grid formation with 16.15 mm separation (Fig. 1).

A custom calibration program was written using OpenCV (Itseez, Russia) so that the mouse location in a given image could be recorded using a mouse callback function. This callback function allows the user to locate and click on a crosshatch and save that pixel position. The calibration program also uses a series of superimposed rectangle and circles in order to correctly orient the board before calibration begins. The board is placed in the visible field of each camera so that the board is perpendicular to the optical axis of each camera. Then, the distance from the board to the camera mount is measured to get a baseline Z value. This baseline Z value is entered into the calibration program.

Once the calibration board has been centered in each image for a given Z depth, each crosshatch in manually selected in each channel in order from the top left corner across the image. For a given crosshatch in the grid, the world (X_w,Y_w) of that hatch are entered into the calibration program. The user then selects that crosshatch in both camera images using the mouse click function. Each time a pair of hatches is selected from each channel, the separate (X_p,Y_p) pixel locations from each channel are recorded along with the world (X_w,Y_w) locations on the board. Both the pixel and world coordinates along with the indices, the computed disparity, and baseline Z are then recorded to a file for offline analysis. The calibration board is adjusted to a new baseline Z value in order to calibrate for a multitude of depth values.

In order to accurately determine the baseline Z values and maintain axial registration used in calibration, the calibration board is mounted on the tailstock of a lathe (Fig. 2). The range of depths used for a particular calibration was between 150 and 250 mm. This range of depths was partially chosen so as to include depths at which the smallest of the trajectory board paths was visible. The range of baseline depths depends on the camera optics.

In one calibration, the camera used was surgical endoscope taken from a da Vinci robot (Intuitive Surgical, Sunnyvale, CA). The endoscope is fed through the headstock of the lathe and held in place. This positioning maintains a fixed position for the camera throughout calibration.

1.2 Methods—Benchmarking. A benchmarking routine was designed to gauge the accuracy of 3D mapping for a given tracking algorithm. A board was developed using computer-aided design software from PTC Creo (PTC, Lansing, MI). This trajectory board contains three circumscribed paths located on the same horizontal plane.

Each path takes the shape of an arc connected by two straight lines (Fig. 3). Each path is recessed into the board so that the tool

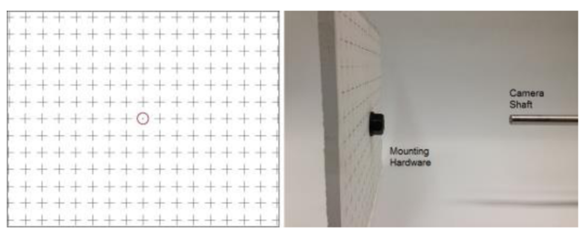

Fig. 1 Calibration board

Fig. 2 Calibration routine using lathe

[1]Accepted and presented at the Design of Medical Devices Conference (DMD2014), Minneapolis, MN, April 7–10, 2014.
DOI: 10.1115/1.4027047
Manuscript received February 21, 2014; final manuscript received March 3, 2014; published online July 21, 2014. Editor: Arthur G. Erdman.

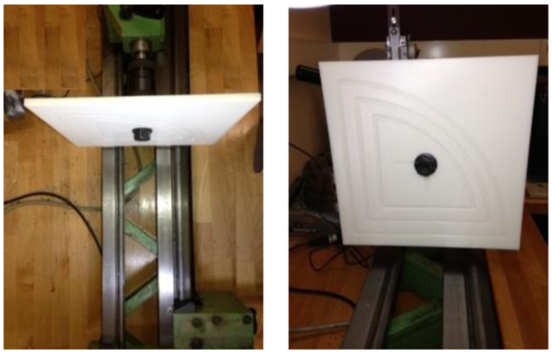

Fig. 3 Reference trajectory in lathe

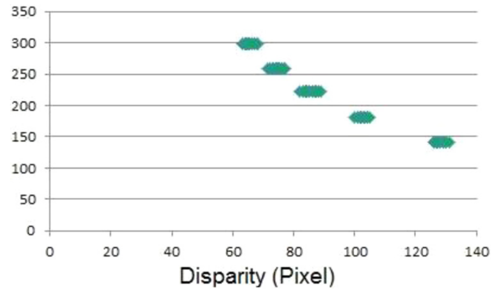

Fig. 4 Depth versus disparity plot

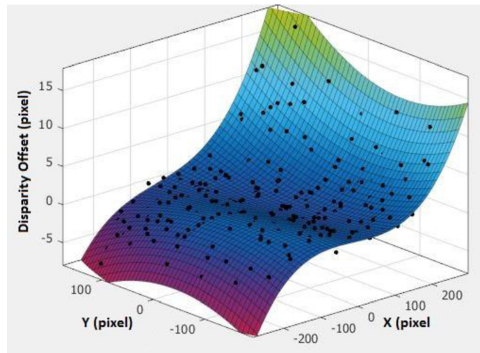

Fig. 5 Calibration data for radial distortion

tip can rest in the groove and repeatably follow the track. Since the board was 3D printed, the geometry of the board is known to within the tolerance of the 3D printer. For this board, a Stratasys Dimension 1200es 3D printer was used (Stratasys, Minneapolis, MN). This printer is capable of resolutions up to 0.254 mm.

The board is secured in the tailstock of the lathe much like the calibration process with the camera on the headstock. The surgical tool tip is then manually articulated through one of the tracing paths until a full lap around the path has been completed. After the application has been shut down, the location data from the data log are exported for analysis. For a given distance from the camera to the trajectory board, the distance is measured using the tools on the lathe thus offering high accuracy measurements for baseline depth Z_w.

Using reprojected data coordinates and artificially populated model coordinates, we calculate the error through a simple closest point search. The distance to the minimum point is then stored and the next data point is analyzed. Using this system we acquire the error between reprojected locations and the closest corresponding known trajectory point. This system can be used to a wide variety of tracking algorithms. There is potential for a slight inaccuracy using this method since the model points have finite spacing between one another (0.05 mm). In this case there is a possibility that a data point may actually lie on the trajectory but appear to have a slight error. However, this issue will only result in a small increase of the average reprojection error, and guarantee that the actual error is always less than or equal to the recorded value.

2 Results

The calibration routine used on the surgical endoscope provided promising results for mapping pixel coordinates into Cartesian space. The disparity was plotted against the corresponding baseline Z_w depth. The plots between pixel space and global space were used to develop a custom model for transforming disparity into depth. It was immediately obvious that the baseline Z value did not depend only on disparity, but also the (X_p,Y_p) pixel position in the image (Figs. 4 and 5). This is visible in the horizontal outliers for each grouping for a particular baseline Z. This additional dependence is due to the effects of radial distortion in the camera lens. A typical calibration routine required between 20 and 30 min to complete.

Both an exponential decay and a power function were fitted to this data. The best fit equation for the exponential decay had the form $Z_w = 634.34 * \exp(-0.012 * \text{disparity})$ with $R^2 = 0.9858$. However the power function equation had a correlation of $R^2 = 0.9946$ using the form $Z_w = 31{,}992 * \text{disp}^{-1.117}$. Upon visual inspection of the graph it is also clear that the power function better represents the relationship. In order to account for radial distortion in the lens, the model was expanded to encompass a third order polynomial with (X_p,Y_p) pixel position (and cross terms) used to normalize the disparity).

3 Conclusion

We have presented a method for the accurate calibration of cameras for use in computer vision algorithms for an array of medical device and surgical applications. We also present a framework for determining the accuracy of tracking algorithms in computer vision using a known trajectory board which can easily be 3D printed and implemented using a machinist's lathe.

Future work will include developing a robust, nonlinear, widely applicable model for transforming pixel location and disparity into Cartesian coordinates for stereo 3D cameras.

References

[1] Pun, T., Geric, G., and Ratib, O., 1994, "Image Analysis and Computer Vision in Medicine," Comput. Med. Imaging Graph., **18**(2), pp. 85–96.
[2] Zhang, Z., 2000, "A Flexible New Technique for Camera Calibration," IEEE Trans. Pattern Anal. Mach. Intell., **22**(11), pp. 1330–1334.
[3] Tsai, R., 1987, "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses," IEEE J. Rob. Autom., **3**(4), pp. 323–344.
[4] Dockter, R., and Kowalewski, T., 2013, "A Low-Cost Computer Vision Based Approach for Tracking Surgical Robotic Tools," ASME J. Med. Devices, **7**(3), p. 030917.