Experiments comparing precision of stereo-vision approaches for control of an industrial manipulator

John-David Yoder, Jeffrey West, Eric Baumgartner, Mathias Perrollaz, Michael Seelinger, Matthew Robinson

Experiments comparing precision of stereo-vision approaches for control of an industrial manipulator

Authors

John-David Yoder, Jeffrey West, Eric Baumgartner, Ohio Northern University, Ada, OH, USA
Mathias Perrollaz, Inria Rhône-Alpes, Saint Ismier, France.
Michael Seelinger, Yoder Software Inc., South Bend, IN, USA.
Matthew Robinson, Jet Propulsion Laboratory, Pasadena, CA, USA

Motivation, Problem Statement, Related Work

Despite years of research in the area of robotics, the vast majority of industrial robots are still used in "teach-repeat" mode. This requires that the workpiece be in exactly the same position and orientation every time. In many high-volume robotics applications, this is not a problem, since the parts are likely to be fixtured anyway. However, in small to medium lot applications, this can be a significant limitation. The motivation for this project was a corporation who wanted to explore the use of visual control of a manipulator to allow for automated teaching of robot tasks for parts that are run in small lot sizes.

Since the 1970s, researchers have been proposing ways to use vision in order to solve this problem. While the purpose of this paper is not to provide a complete review of vision-based robotic manipulation, an excellent overview of early work is found in [1]. There has been success in the application of such technologies, especially in 2-D and 2.5-D problems [2]. Despite the fact that the theory for the solution of this problem has been established, there are limited implementations of full 3-D applications. Basically, the reasons for this come down to "the devil is in the details" and, in real 3-D applications, the accuracy, robustness, and cost-effectiveness of vision-based systems have been insufficient to justify widespread use.

A variety of efforts have been put forth to solve the problem of vision-based robotic manipulation. Visual Servoing offers promise by "closing the loop" in the image plane [3]. However, in many assembly tasks, the robot end effector obscures the target when it gets close to task completion. Several methods have been proposed to get past this limitation, such as outfitting both the robot and the workpiece with fiducials that remain visible [4], or "eye-in-hand" in cases where it is possible to place the camera(s) in the robot end effector. Another limitation of visual servoing is that in general it provides the least accuracy in the direction of the focal axis. Camera Space Manipulation [5][6] offers an alternative methodology and has been shown to allow completion of high-precision 3-D tasks. Unfortunately, this method does not allow standard stereo cameras to be used, as it requires widely separated and highly vergent cameras and utilizes the simplified orthographic model (though later work created accuracy similar to the pinhole model) [7].

The Mars Exploration Rovers used stereo vision with calibration for placing instruments on rock and soil targets [8], but this showed some limitations in accuracy [9]. Recently, in space robotics applications, two additional approaches have been offered, HIPS [9] and AGATE [10][11]. Both of these approaches have shown promise on space-related platforms to produce high-precision, vision-based manipulation using stereo cameras, with an application of instrument placement. These papers also reported large numbers of experiments. In particular, [9] showed the improved accuracy using HIPS compared to precalibrated stereo on NASA testbeds. However, these testbeds typically have inaccurate kinematics (backlash, inconsistent zero offsets, lack of rigidity, etc.).

Technical Approach

In response to a request from a local manufacturer, several of the authors were involved in a project to use computer vision to automate the teaching of a robot task for parts with small lot sizes. The details of that task are found in [12]. While the corporation was pleased with the results of the project, the accuracy obtained was not sufficient for all of their parts. As such, the authors have begun a detailed investigation of the accuracy obtained, and how that compares to other approaches. It should be noted that the industry-focus of this project has influenced the technical approach.

The goal of this paper is to answer the question of whether techniques such as AGATE can create significant improvement in accuracy when applied to an industrial robot by using local data to modify the parameters of the camera-robot model. Also, while AGATE was developed for mobile manipulation (as it can control the mobile base and the manipulator relative to a visual target), this paper will use AGATE techniques for a fixed-base manipulator. The paper further includes the first direct comparison of AGATE with HIPS. Future work will compare these methods directly to visual servoing and traditional camera calibration techniques.

HIPS and AGATE both make use of the CAHVOR camera model [13]. In both cases, the cameras can be calibrated using a standard checkerboard, but are typically calibrated by moving the manipulator through a series of poses in camera-space. In summary, a least-squares minimization is completed to compute the CAHVOR parameters based on data sets made up of 3-D robot positions and the 2-D image-plane appearance of the end effector in each camera. This minimization is completed after the manipulator moves through a pre-programmed set of positions designed to cover a significant portion of the robot's workspace and view of the camera(s). The CAHVOR parameters are then updated with additional measurements when they are available. Such measurements can be obtained while the end effector is approaching the target point. Addition of the localized samples results in a mapping between image space and physical space that is not globally accurate, but is very accurate relative to the manipulator in the region of the target point. This general approach has been shown to work well even in the presence of large kinematic errors [9][11]. Both HIPS and AGATE have been shown to provide the following advantages compared to traditional calibration methods: Robustness to poor kinematics, the ability to deal with changes to the internal and external parameters of the

camera, and the ability to achieve high precision manipulation relative to visually specified points. In addition, AGATE has been shown to be able to control mobile manipulators, and use more than two cameras [11].

Robinson [9] showed the advantage of HIPS (which finds the CAHVOR model parameters based on manipulator observations of fiducials on the manipulator) over pre-calibrated stereo, and the advantage of adding the local information – both in simulation and in a large number of experiments with a mobile manipulator. However, the manipulator in question was not as accurate as a typical industrial manipulator – it was less rigid and had significant kinematic errors (it was a prototype on a mobile manipulator testbed). The primary goal of this study is to see if a similar improvement takes place using an industrial manipulator, or whether the improved rigidity and kinematic accuracy negates any improvement due to on-the-fly CAHVOR model updates. For consistency with [9], we will call the version of the code that does not update the model during approach "Static AGATE" and the version that does update "Dynamic AGATE".

Experiments

Experiments were conducted with the 6-axis ABB IRB-140 robot shown in Figure 1. The vision sensor is a pair of PointGrey Flea2 cameras, delivering 640 x 480 RGB images at 15FPS. As shown in Figure 2, Cameras are placed on a common portable support with a baseline separation of 15cm, and are placed approximately 2.5m from the workspace. While the cameras are effectively parallel, no particular efforts were taken to ensure a proper rectified configuration.

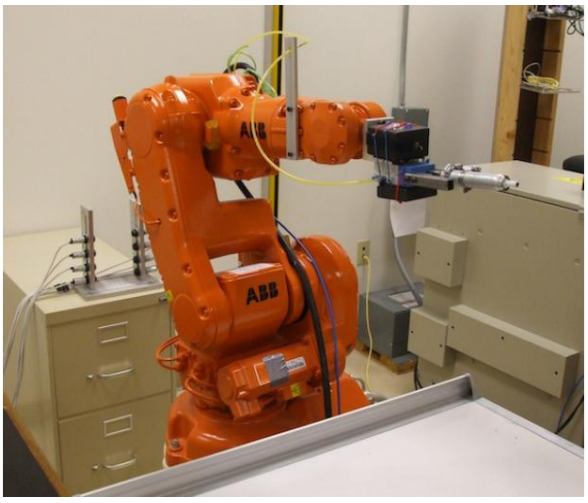

Figure 1: Experimental Setup

Figure 2: Camera System

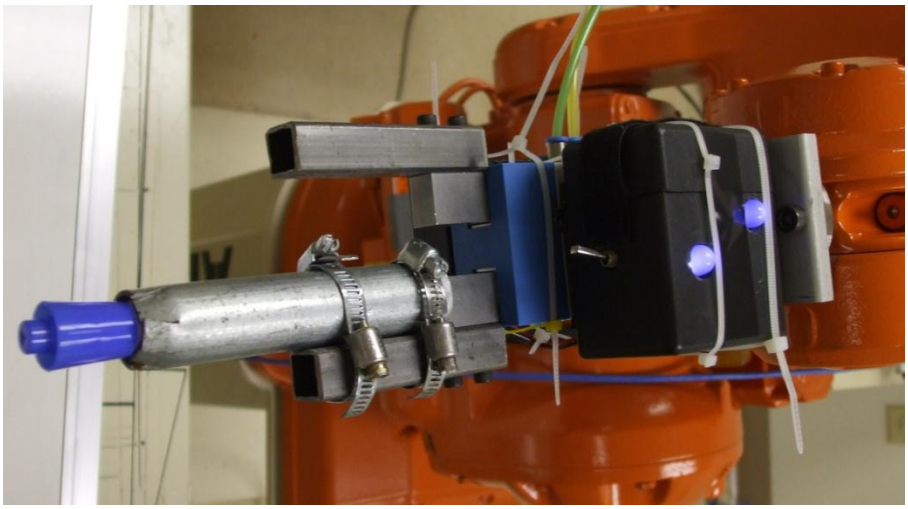

Figure 1: Close up of marker attachment

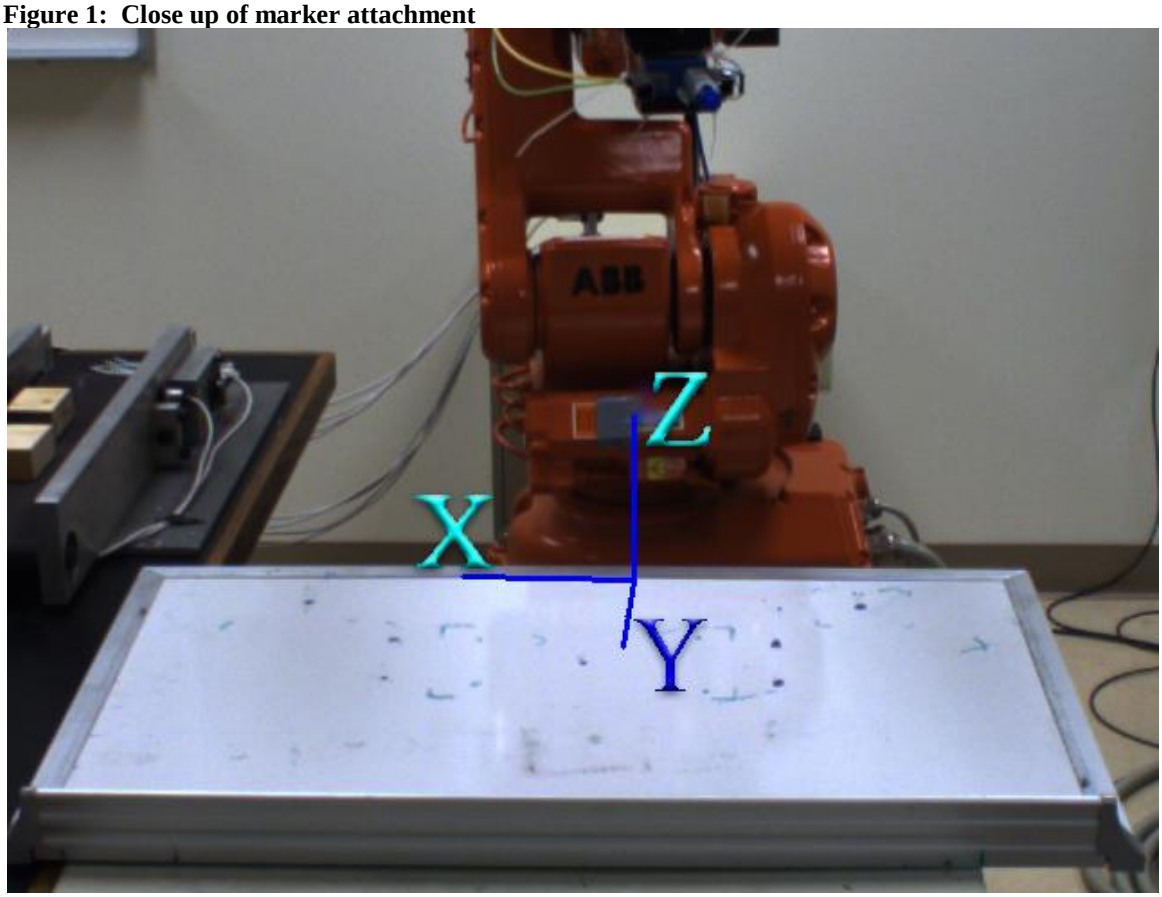

Figure 4: Robot Coordinate System

The communication between the robot and the PC is performed through the serial port, with a message-based protocol. The C++ program for perception, control and command runs on a desktop PC equipped with an INTEL XEON W3505 (2009), dual core 2.54GHz, with 4GB of RAM.

Central to HIPS or AGATE is the ability to locate the end effector of the robot as it moves through the workspace. These image-plane locations, along with the corresponding location of the robot, are used to update the CAHVOR camera parameters. To simplify this task, the robot's end-effector is equipped with a LED blinking system as shown in Figure 3. The blinking frequency is roughly a third of the frequency of the cameras to avoid possible aliasing. We chose a rather simple design of the LED lamp, which does not require synchronization between the emitter (LED) and the receiver (camera). This is possible because the arm can be static during the acquisition of a set of images. Once the images are acquired, two images, respectively representing the mean value and the standard deviation over time of the intensity of each pixel, are computed. The cyan component of both images is computed using the green and blue planes (because our LEDs produce cyan rather than blue light). Then connected components of high standard deviation are extracted. These regions are detected as LEDs if their mean value over the cyan image is also sufficiently high.

All control is done by the ABB's controller. A directly-connected serial connection is used to transfer commands and coordinates to the ABB. More details about the system, and the implementation of a precise manipulation task, can be found in [12].

A typical experiment consists of placing a fiducial in the workspace. For Dynamic AGATE, the robot is commanded to move half the distance to the target, and an additional measurement is made. If the results of this measurement are within normal bounds, this point is used to update the CAHVOR parameters. The process is repeated several times, until the robot is less than 4mm from the final goal, at which point it is moved to the final goal position. For Static AGATE, once the fiducial is found, the robot is moved directly to the target.

Results

To provide a measure of accuracy, the robot end effector was equipped with a standard whiteboard marker. The marker was placed in a holding device that provides approximately 1cm of compliance along the axis of the marker (shown in Figure 3). Circular fiducials were placed in the workspace, and the robot was instructed to approach the center of the fiducial and stop 5mm above the fiducial (above corresponds to the Z axis) of the base frame of the robot.

Figure 5: Marked fiducial

The jog mode of the robot is then used to move straight down to mark the fiducial. The vertical travel during the jog is used to measure the error in the Z direction. The distance from the mark to the center of the fiducial circle can be used to measure the error in the X-Y plane in the robot base frame, since the fiducials are placed on this plane. The X-Y-Z coordinate frame of the robot is shown in Figure 4. A sample of the marked fiducial is given in Figure 5 To provide a sense of scale, the outer diameter of the black circle is 76mm.

These tests were repeated 7 times for both the Static AGATE and Dynamic AGATE with the fiducials placed at varying locations in the workspace. One immediate finding was that Static AGATE did not consistently touch the fiducial (the marker often stopped above the fiducial). In these cases, the distance from the marker tip to the fiducial was measured and recorded as an error in the *z* direction.

Thirty tests were conducted using Static AGATE and Dynamic AGATE. The tests consisted of 20 different physical locations throughout a region of the workspace (ten of the positions are reached twice). Table I shows a summary of these measurements. In summary, the mean error with Dynamic AGATE was 11.0mm, while the mean error with Static AGATE was 14.6mm. A t-test showed this to be significant at the 95% confidence level (t=-2.169, significance =0.038). Dynamic AGATE showed a standard deviation of 7.8mm, reduced from 9.9mm with Static AGATE.

Looking at the data in Table I, it is clear that most of the difference between the two is due to the XY-plane error. Comparing only this error, Dynamic AGATE resulted in a mean error (distance from the center of the cue) of 9.6mm compared to 13.5mm for Static AGATE. A t-test showed this to be significant at the 95% confidence level (t=-2.135, significance =0.041).

Using the data sets obtained during the calibration process and during the tests, a direct comparison was completed between AGATE and HIPS. Both methods were used to project the observed locations of the LEDs from camera-space into physical space. The methods use different approaches for this process. While full details are outside the scope of this paper (see [9] and [11]) HIPS find the 3-D point at the center of the shortest line between the projecting rays. AGATE finds the 3-D point that minimizes the image-plane error between the measured and projected points. Despite these differing approaches, the results were very similar. The average difference (as measured by the norm of the vector between the 3-D projections) was 0.021mm, and the standard deviation was 0.025mm. Given the scale of the other errors involved, the conclusion is that AGATE and HIPS would produce the same accuracy of control.

Understanding the Source of Positioning Errors

To draw any conclusions from the data presented, there must be an understanding of the sources of error that are included in the errors reported in Table I. The basic sources of error will be enumerated here:

1) Errors due to the limitations on the mechanical system. That is, a motor can only be commanded to within one encoder count – this limitation on the precision of control will manifest itself as an error in the final measurement.
2) Error due to inaccuracies in the models and the limitations of the algorithms.
3) Error in making the measurement.
4) Error in the detection of the target.
5) Error in the detection of the LEDs.
6) Error due to the kinematic modeling of the LEDs and the marker.
7) Error in ensuring that the same physical point is selected in both cameras (correspondence error).

While these are all sources of error, error #1 is very small for an industrial robot. ABB reports repeatability, not accuracy, of their robots. The repeatability of the ABB 140 used in these tests is ± 0.03mm [14], which is insignificant compared to all other errors.

Error #2 is related to the approach, and will be discussed later.

Error #3 has been minimized by physically marking the fiducial for later measurement. However, since the marker makes a point that has a diameter of over 1mm, it is likely that an error on the order of 1mm is reasonable for this item. A detailed examination of this could be undertaken by having multiple people measure the location of each mark, but this was not done for the experiments described here.

Using a fiducial limits the magnitude of error #4, the detection of the target. Multiple measurements were taken (though not a statistically significant number), and it was found that the standard deviation on the location of the center of the fiducial was approximately 0.2 pixels. While the volume of a voxel varies throughout the workspace, it is approximately 1.67x5.00x1.67mm (in X, Y, Z, respectively). Thus, simply due to the resolution, one should expect errors with a standard deviation of approximately 1mm. To mitigate the effects of this error, the fiducial location was found in ten consecutive images, and the average center location was used.

Regarding error #5, the LEDs are detected using image-differencing in order to reduce the likelihood of false positives. A center-of-gravity approach is used to find the center of each LED, and the two LED locations are combined into an "average" of their locations in order to reduce errors and eliminate the ambiguity between the two. Multiple tests were run and the standard deviation for the detection of the LEDs was found to be negligible.

Error #6 would always be present – the kinematics of any tool that is added on to the robot must be modeled and would always include error. Note that when used in teach-

repeat mode, this error would not cause any positioning errors. Because of the plan to use this in industry, the standard ABB functions were used to "teach" the ABB the location of the points of interest (the tip of the marker, and the point between the two LEDs). This involves taking the point being taught to the same point in physical space from four different directions. The ABB then completes a least-squares minimization to locate the point relative to the robot coordinate system. For the data set shown in Table I, the mean error for the marker was 0.73mm, while the mean error for the LED center position was 0.41mm.

Error #7, the correspondence error, is very important, and is related to the detection errors. Essentially this error is due to the fact that a given item (LED, fiducial, etc.) in X-Y-Z space will produce 4 values in camera space (assuming two cameras). If both cameras detect EXACTLY the same physical point, then there is no inconsistency between the camera-space measurements and the physical space measurements. However, this is impossible to do (as stated earlier, finding the center of the fiducial, for example, has a standard deviation of 0.2 pixels). An error of 1 pixel in the horizontal location of an item in one of the cameras, for instance, will create an error of over 20mm when it is projected back into physical space using the AGATE method. This was found to be identical using HIPS.

When computing the parameters of the CAHVOR camera model, many of these errors are included. Recall that this computation is based on the image-plane location of the LEDs and the corresponding physical location of the robot. Thus, errors #1-3, and #5-7 are included (all errors other than the marker kinematics and the fiducial errors). In our experiments over multiple calibrations, these lead to a mean reconstruction error in the CAHVOR model of approximately 5mm. The reconstruction error is found by using the CAHVOR parameters to project the LED locations back into 3-D space and comparing those locations to the actual, recorded 3-D locations of the robot.

Main Experimental Insights

The main experimental insight is that Dynamic AGATE outperforms Static AGATE, even on this system with high-quality cameras and a very repeatable industrial manipulator. Furthermore, AGATE and HIPS produce effectively the same results.

It should be noted that the Dynamic AGATE discussed to this point in the paper typically resulted in six or seven model updates. Experiments were conducted to see if additional updates would further improve the accuracy. In order to do this, the approach was modified so that the robot moved one third of the way to the target each time rather than half way. This resulted in, on average, five additional updates per experiment. Another 30 experiments were conducted in this manner. However, no statistically significant change was observed. Thus it was concluded that this number of updates slowed the system without a significant advantage in accuracy.

The accuracy is not as high as that shown in previous work with AGATE and HIPS. Examining the sources of error described above, there are a variety of reasons for this

increased error. Error #1 is much smaller than other systems tested, and errors #2 and #3 are comparable. Error #6, however, is larger. In previous work with HIPS and AGATE, highly accurate kinematic models of the end effectors (specifically the relationship between the end effector and the fiducials) were used. It is more realistic for an industrial tool which will have to be changed regularly to have an operator use the tool-teaching algorithms built-in to the ABB. However, this introduces more error. Taking the mean of errors #1, #3, and #6 would predict a mean error of approximately 2mm.

Errors #4, #5, and #7, however, should be expected to be larger in the experiments described in this paper than in previous work using HIPS or AGATE. This is because the distance from the cameras to the workspace is much greater (2.5m in this case compared to approximately 0.5m). This leads to increased error in several ways. First, it increases the error in locating the target and the LEDs in the image plane. Second, it increases the size of a voxel. Thirdly, and most importantly, it increases the important of correspondence error. The importance of this effect was shown in [15]. But in that set of experiments, the error shows was only 1.5mm/pixel. In the experiments shown in this paper, the error was over 20mm/pixel. Given the magnitude of these errors, showing an overall error approximately 11mm is consistent with expectations.

It is also worth noting that a major advantage of this methodology is that it allows the placement of the stereo rig on a simple tripod. This tripod can be moved relative to the robot without impacting the performance of the system, since the system can re-estimate the CAHVOR camera model at any time. In industrial settings, the cameras would certainly be fixed – but this would provide the ability to adjust if the cameras are bumped or moved during operation or maintenance activities. The method furthermore does not require that the images are rectified, and the alignment of one camera relative to the other is arbitrary, as long as the workspace can be seen by both cameras.

Continuing Work

While the work to date has shown that Dynamic AGATE shows improvement over Static AGATE, and that AGATE and HIPS produce effectively the same result, much work remains to be done. In particular:

- Build a more accurate tool holder to improve the accuracy of measurements. It is unclear how much of the current error is due to the methodology, and how much is due to inaccuracy in the tool holder. As can be seen in Figure 3, the current setup is simply clamped on, in a manner that is neither accurate nor repeatable.
- Comparison with traditional CAHVOR stereo (using a checkerboard for calibration)
- The use of more than two cameras (which is directly supported in AGATE)
- Comparison with an "off the shelf" disparity-based stereo system.
- Comparison with visual servoing.

When complete, it is expected that this will provide a rich dataset of direct comparison of four different, previously-published approaches to the problem of visually-guided manipulation. The one approach not specifically examined is Camera-Space Manipulation. While this approach has shown very good results, it requires widely separated and highly vergent cameras, and therefore would require a completely different experimental setup than the other methods. This would make a direct comparison difficult.

Acknowledgment

The authors would like to thank American Trim Corporation for their support of this project. Furthermore, thanks are due to Amber Cool and Patrick Whitten, undergraduate students without whose assistance we could not have gathered sufficient data for this paper. Further thanks to Dr. Sami Khorbotly, who developed the hardware for the LED system.

References

[1] P. Corke, “Visual Control of Robot Manipulators, -- A review”. *Visual Servoing* (K. Hashimodo, ed.), *Robotics and Automated Systems*, pp 1-31. 1993

[2] M.K. Morel, “System helps prepare sintered metal parts” *Vision Systems Design*, Jan. 2004.

[3] S. Hutchinson, G. D. Hager, P.I. Corke, “A tutorial on visual servo control”, *IEEE Transactions on Robotics and Automation*, Volume 12, Issue 5, October 1996.

[4] B. Hammer, S. Koterba, J. Shi, R. Simmons, and S. Singh, “An autonomous mobile manipulator for assembly task”, *Autonomous Robots*, Volume 28, Number 1, 2010.

[5] S. Skaar, W. Brockman, and W. Jang, “Three-dimensional camera space manipulation,” *Int. Journal Robotic Research*, vol. 9, August 1990.

[6] E Gonzalez-Galvan, S. Skaar, U. Korde, W. Chen, “Application of a Precision-Enhancing Measure in 3D Rigid-Body Positioning Using Camera-Space Manipulation”, *Int. Journal Robotic Research.*, vol. 16, April 1997.

[7] E. J. Gonzalez-Galvan and S. B. Skaar, "Efficient camera-space manipulation using moments," *Proc. IEEE Intl. Conf. on Robotics and Automation*, pp. 3407-3412, 1996.

[8] E. T. Baumgartner, R. G. Bonitz, J. P. Melko, L. R. Shiraishi and P. C. Leger, “The Mars Exploration Rover Instrument Positioning System,” Proceedings of the 2005 IEEE Aerospace Conference, Big Sky, MT, March 2005.

[9] M. Robinson, E. Baumgartner, K. Nickels, and T. Litwin, “Hybrid image plane/stereo (HIPS) manipulation for robotic space applications,” *Autonomous Robots*, vol. 23, 2007.

[10] J.-D. Yoder and M. Seelinger, “Long-range autonomous instrument placement,” in Proc. *ISER*, Rio de Janeiroi, Brazil, 2006.

[11] M. Seelinger, J.-D. Yoder, and Eric Baumgartner, “Autonomous Go-And-Touch Exploration (AGATE)”, *Journal of Field Robotics*, Vol. 29, Issue 3, May/June 2012.

[12] M. Perrollaz, S. Khorbotly, A. Cool, J-D. Yoder and E. Baumgartner, “Teachless Teach-Repeat : Toward Vision-based Programming of Industrial Robots”, in *IEEE International Conference on Robotics and Automation*, 2012.

[13] Gennery, D. B. “Least-Squares Camera Calibration Including Lens Distortion and Automatic Editing of Calibration Points.” *Calibration and Orientation of Cameras in Computer Vision*, A. Grun and T. Huang Editors, Springer Series in Information Sciences, 34:123-136, 2001.

[14] ABB Corporation, “IRB 140”, [Online] http://www.abb.com/product/seitp327/7c4717912301eb02c1256efc00278a26.aspx?productLanguage=us&country=US, accessed 5/22/2012.

[15] E. T. Baumgartner, and N. A. Klymyshyn, “Sensitivity Analysis for a Remote Vision-Guided Robot Arm under Imprecise Supervisory Control,” Sensor Fusion and Distributed Robotic Agents, SPIE Proc. Vol. 2905, pp. 218-226, Boston, MA, October 1996.