2017-12

Automating Active Stereo Vision Calibration Process with Cobots

Mohamed, A

10.1016/j.ifacol.2017.12.030
IFAC-PapersOnLine

Automating Active Stereo Vision Calibration Process with Cobots

Abdulla Mohamed*, Phil F. Culverhouse*, Ricardo De Azambuja*, Angelo Cangelosi*, and Chenguang Yang**

** Centre for Robotics & Neural Systems, Plymouth University*
UK (e-mail: abdulla.mohammad@plymouth.ac.uk).
*** Zienkiewicz Centre for Computational Engineering, Swansea University, UK.*

Abstract: Collaborative robots help the academia and industry to accelerate the work by introducing a new concept of cooperation between human and robot. In this paper, a calibration process for an active stereo vision rig has been automated to accelerate the task and improve the quality of the calibration. As illustrated in this paper by using Baxter Robot, the calibration process has been done faster by three times in comparison to the manual calibration that depends on the human. The quality of the calibration was improved by 120% when the Baxter robot was used.

Keywords: Robotic systems, collaborative robot, active stereo vision, calibration

1. INTRODUCTION

Nature is the mother of creation. Engineers may consider the nature as one of the best sources of innovation as well as inspiration; where the inspiration is mainly gained from the creatures' ability to be altered according to nature and the surrounding environment. Usually, creatures sense the surrounding environment using five different sense organs ears, eyes, nose, skin, and tongue. Vision is the sense that provides 80% of information surrounding creature (Chapman 1998). Here, computer vision has an intense research where it is employed in many applications such as a self-driving car to identify traffic signs, lines, and depth perception using two cameras (for example sees: Das & Ahuja 1995; Szeliski 2011; Kuang et al. 2012; Dankers & Zelinsky 2004).

There are many forms of active stereo vision, or a system that changes the geometry of the camera's setup dynamically, such as pan and tilt of the stereo camera or pan and tilt each camera individually, variable baseline, and focal length (Fig. 1); controlling the angle of each camera to dynamically extend the field of view, improving object tracking and fixed the view of both cameras on the interesting point. While controlling the baseline improves the depth measuring, controlling the focal length helps to enhance the focusing. Selecting the right parameters is critical when designing a stereo vision system.

Several characteristics in an active stereo vision system can increase its performance comparing to the orthogonal or fixed stereo vision ones; where the active stereo vision system narrows the correspondence process to focus on the interest object in the scene by increasing the overlapping between the left and right images. The vergence angle, or the movement of both cameras in opposite directions, simplifies the measuring process by keeping the fixation point on the object. The fixation point is where both focal axes get intersect on the interesting point in the scene. This fixation tracks the object if the object moves or the system moves. Another characteristic of active stereo vision is the variable baseline (the distance between the origin of the cameras), where the depth (distance from the camera centroid to the object) is proportional to the baseline.

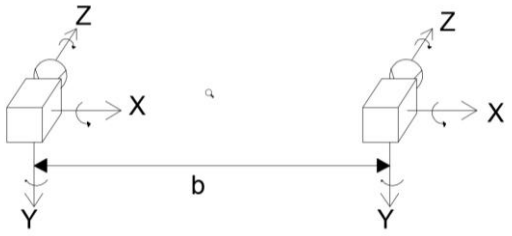

Fig. 1: Active stereo vision: the camera can rotate around the axis, and the distance between the cameras is variable.

Regarding the vergence angle, it is clear that the disparity map quality can be affected by changing the angle. The authors in (Krotkov et al. 1990) investigated to explore the relationship between the vergence angle and the quality of the disparity map. The system used in the experiment had two Degrees of Freedom (DOF) where the cameras pan independently; the baseline was fixed at 13cm. The experiment measured the distance from the camera to the object by changing the vergence angle. The correspondence method was used based on feature matching by edge detection. The experiment presented an error of 5% in measuring the depth at a distance of 3 meters. Although the result shows 5% of error difference in vergence angle, there is no relation between the quality of the disparity and the vergence angle.

According to work presented in (Sahabi & Basu 1996), the disparity error was studied based on the verification of the vergence angle and the spatial resolution. Spatial resolution refers to the image having high resolution in the centre while the resolution drops by moving to the edge of the image. A single camera, with a focal length of 8.3mm, takes images at different two locations where the camera moved by 112mm. Objects were placed in front of the system with known distances. For the verge-angle experiment, it was found that there is no specific angle to reduce the disparity error based on the complete image. The result of this experiment was similar

to the result of the experiment done by Krotkov (Krotkov et al. 1990).

In its second experiment (Sahabi & Basu 1996), spatial resolution was used; where the resolution of the image decreases away from the centre of visual axes (as in the human eye). The results showed that when both cameras are focusing on the same point, the disparity error at that point becomes at the minimum range, which agrees with the theoretical result.

Another variable in active stereo vision is the baseline (Klarquist & Bovik 1997). The system consisted of a variable baseline and two cameras with a panning joint. In this work, a method was introduced to improve the quality of the depth map by using variable baseline. The process starts with a short baseline in order to simplify the matching process then the baseline is increased to explore the depth resolution of the scene. The new baseline is chosen based on the result of the previous baseline and the cycle repeats until a satisfactory resolution is reached. The experiment was run with different objects and different distance, and it found that the minimum distance is 50 cm to produce a fine resolution. The result of the experiment shows that the process produced a good depth map with smooth reconstruction, although no specific baseline details were published by the authors (Klarquist & Bovik 1997).

More work on the variable baseline was done by Nakabo (Nakabo et al. 2005) where an active stereo vision with a variable baseline and rotating angle for both cameras to pan and tilt independently was built. The work was used to uniform depth error by controlling the baseline and vergence angle during object tracking. The speed of the baseline travel is 4 m/s and the system run on image size 120x120x8bit resolution, with an image processing speed of 2ms (30FPS). The system tracks an object, estimates the distance and reconstructs the object only if the object is near to the platform. The matching process used in the experiment was Sum of Absolute Differences SAD with a window size of 5x5 pixels. The experiment was set to track an object that runs in a circle. Three experiments were carried out at fixed baseline 400mm, 800mm, and variable baseline. The result was compared and, the error generated by active baseline dropped by 30% about the fixed one. The result shows there is potential in producing a system that maintains a low depth error using a variable baseline.

Calibration of the active stereo vision is still an active research field due to the complexity in recalibrating the system during operation. Many works have tried to tackle the problem by implementing the fundamental matrix or the homography matrix to re-mapping between the views. These methods required to match the features in both images (Luong & Faugeras 1997; Bjorkman & Eklundh 2002; Szeliski 2011), they use a lot of powerful computations for matching the features between both images and currently lead to huge processing time.

The calibration process described in this paper is to calibrate an active stereo vision rig, where Baxter robot was used in this experiment to hold the checkerboard and to move it around. The calibration process for an active stereo vision is to acquire the parameters of the rig itself unlike what can be seen in the literature (Mišeikis et al. 2016; Quigley et al. 2010; Pradeep et al. 2014; Alexander et al. 2010). These references only describe how to calibrate the camera with an automated arm to get the position of the camera relative to the position of the arm.

Here in this paper, the collaborative robot Baxter holds the calibrated pattern and moves to different positions in order to collect points on the pattern instead of the traditional way of a human being holding the pattern. Baxter is a friendly robot to be controlled using the programming by demonstration methodology to identify the position of the pattern without requiring an intensive programming effort. Baxter is a safe robot to work in a busy lab, where the setup of the stereo vision rig in the lab occurs while many people are working and moving around it inside the lab. Another point is that Baxter can be moved to explore the 3D calibration space fully, something not guaranteed when a human moves the pattern.

The paper is organised as follow: the next section introduces the stereo vision calibration process, the experiment setup, and the result analysis. In the third section, the result and discussion are presented and, finally the conclusions and future development closing this work.

2. METHODOLOGY

The calibration of the stereo vision system is the step that requires finding the external and internal parameters of the system. The parameters found by the calibration process are used in image rectification where the epipolar lines are transformed to be parallel with the baseline to reduce the matching process into a 1-dimensional *search (Fig. 2) and* used in finding the depth of the object or reconstruction of the scene. In a fixed stereo vision system, the Zhang algorithm (Zhang 2000) is used to find the parameters of the system. In an active stereo vision, the parameters of the system, which are baseline, pan, and tilt angle, and focal length are changeable; as they need to be found again every time they change. Only the lens and camera characteristics are fixed.

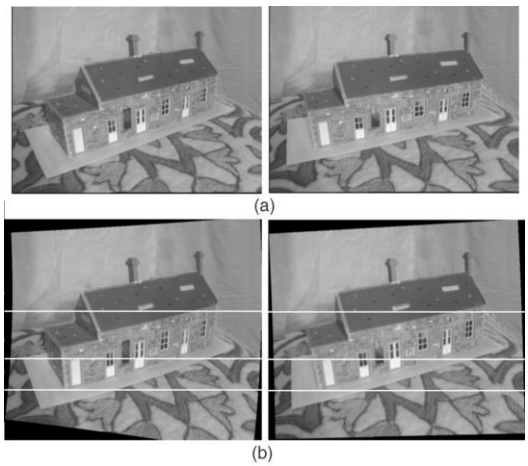

Fig. 2: Rectification image where epipolar lines become parallel with the baseline [source: (Lee et al. 2008)].

2.1 Stereo calibration

We start with a single camera model describing the pinhole camera system. This model is used as well to describe the cameras' CCD sensors used in this project. The centre of the camera is O, which identify the center of the Euclidean coordinate system. The image plane π is placed on Z axis and the distance between the origin and image plane is focal length f.

Suppose a point W with coordinates $[X\ Y\ Z]^T$ in front of the image plane. A projection point $w = [x\ y]^T$ on the image plane will be formed when we draw a line from W to the origin of the camera O. This creates a mapping from a 3D to a 2D space. Using a homogeneous coordinate to map between the points, we get eq. (1)

$$w = PW \tag{1}$$

Where $W = [X\ Y\ Z\ 1]^T$ and $w = [x\ y\ 1]^T$ became a homogenous vector. P is the camera projection matrix.

The camera projection matrix P contain the internal and external parameters eq. (2)

$$P = AR[R|t] \tag{2}$$

A is a 3x3 matrix describes the internal properties of the camera eq. (3) . Where α_x and α_y are the focal length in pixel in direction of x and y respectively. s is a skew parameter and in most new cameras is zero or close to zero (Xiao et al. 2010).

$$A = \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

R and t are the external parameters that refer to the transformation between the camera and world coordinate. Where R is a rotating matrix 3x3 and t is the translation vector 3.

The calibration process for a single camera depends on eq. (1) by providing the point coordinates of w and W that the image coordinate was found by applying corner detection and the points in world coordinate given by measuring the distance between the corners in the checkerboard. After finding these points, the camera projection matrix can be found algebraically. Zahoge (2000) presents a well-known algorithm that can be used to find P.

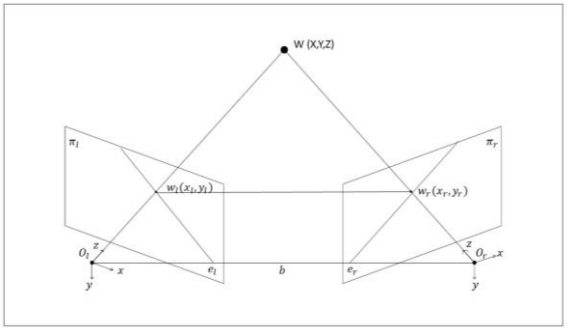

Fig. 3: Two-camera models. The model used describes the process of the stereo vision algorithm and depth measuring.

In a two-camera model, the same process of a single camera is applied. In this section, the parameters with a subscript l and r are used to refer to the left and right camera models respectively. *Fig.* ***3*** shows the model that is reviewed in this section. B is the baseline distance between the two origin cameras. Suppose that both cameras are looking to the same point in the world $[X\ Y\ Z]^T$, a point W will be projected on both image planes $w_l = [x_l\ y_l]$ and $w_r = [x_r\ y_r]$.

From the models, a plane is formed when O_l , W, *and* O_r are connected. This plane is called the epipolar plane. If w_l is known then w_r could be found by searching along a line $l_r = e_r \times w_r$. This line is called the epipolar line. From the epipolar line $l_r = e_r \times w_r = [e_r] \times w_r$ where $[e_r]$ is the cross product, and by mapping w_r to w_l this lead $w_r = H\, w_l$. H is the homography matrix 3x3 rank3, and describes the mapping between the two points. By combining both equations, we get $l_r = [e_r] \times H\, w_l = F\, w_l$ where $F = [e_r] \times H$ and it is called the fundamental matrix.

The fundamental matrix can be extended to have the camera projection matrix as shown in eq. (4), where P_l^+ is the pseudo invert of P_l. The fundamental matrix defines the internal and external parameters of the stereo vision system.

$$F = [e_r] \times P_r\, P_l^+ \tag{4}$$

In the stereo vision rig, the projection camera matrices are presented in eq. (5) and eq. (6) where R and t represent the rotation and translation between the left and the right origins. O_l is the origin of the rig.

$$P_l = [I\ |0] \tag{5}$$

$$P_r = [R\ |t] \tag{6}$$

Eq. (5) and eq. (6) are in normalise coordinate and we combine them we get eq. (7)

$$E = [t]_\times\, R = R[R^T t]_\times \tag{7}$$

Essential matrix describes the transformation between the left and right origin in the stereo vision system.

The calibration process used in stereo vision is the same where a checkboard is used as a reference of the points in world coordinate and image processing used to find the points in image coordinate. The process is initially done on each camera separately to find the projection camera matrix for each camera (i.e. the intrinsic parameters), and then it is used to calculate the essential matrix to find the external parameters between the cameras, i.e. the extrinsic ones.

2.2 Experiment setup

The platform was built using 3D printed parts and aluminium extruded tube as a rail. Two carriers were used to carry the two cameras and their motors; these carriers move horizontally driven by a stepper motor (Fig. 4). Integrated stepper motors with encoders are used to control the rotating angle of the cameras individually; then, the cameras are attached to these motors by a 3D printed bracket. The design of the bracket was chosen carefully to keep the rotating axis of the motor intersecting with the origin of the cameras.

The Robot Operating System (ROS) (Quigley et al. 2009) was chosen to control the stereo vision platform. The controller was designed to provide each motor with its own controlling interface (Fig. 5).

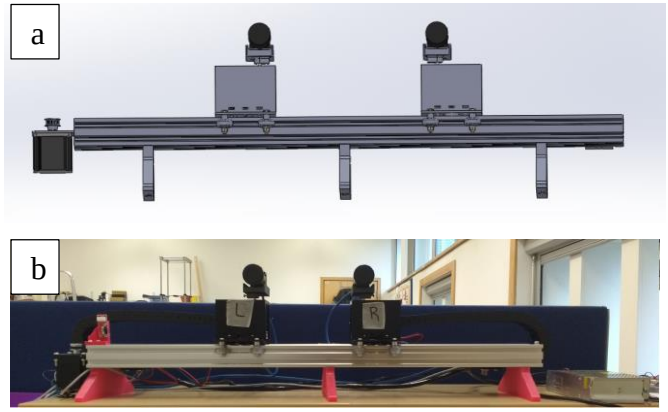

Fig. 4: (a) CAD model, (b) photo of active stereo vision rig used in the experiment

The nodes under the motor controller namespace (nodes 1 to 4) are the ones responsible for controlling the motors. For each motor, there is a micro-stepping controller to control the motor and read the value of the encoder attached to its shaft. Nodes 1 to 3 were designed to work as micro-stepping controllers and communicate with the motor master through a USB cable. The communication between the motor master (node 4) and the motor nodes (Nodes 1 to 3) were through topics established by ROS. Each node publishes an encoder position under the name of *angle_position* to the motor master and receives a topic to move the motor under the name of *move_motor*. The topic name was combined with the motor name e.g. for the left motor node 2, and move_motor topic results in the full name of the topic publish */left/move_motor*. Node 4 is the master node in the motor controller namespace where this node is responsible for doing the geometry calculation for the rig. This geometry published under topic */rig/transformation* to the node 8 to do further processing with images (i.e. calculating the depth map).

There are two modes to control the camera motors: servo mode and continuous rotating mode. In servo mode, the motors are controlled by sending angle values in degrees; while, in rotate mode, they are controlled by using angular speed in rpm. Node 1 baseline motor controller, is set either by giving the required baseline in millimetres or by controlling the speed of the carrier in m/s. All nodes controlling the motors provide a position feedback with an accuracy of ±0.05 degrees resulting in ±0.1 mm. The variety of control modes provides the flexibility necessary for studying the stereo vision configuration.

In camera controller namespace, two nodes (node 5 and 6) were designed to capture the image from the left and right cameras. The images are published through topics under *raw_image* images to the rectify image (node 7) that received the geometry of the platform from the motor master (mode 4) to rectify the image and publish these images under the name of *rec_image*. The *rec_image* received by the stereo vision master (node 8) to do the calculation of the depth map, and track object while communicate with the motor master (node 4) to change the geometry of the platform if required. The same approach was used in motor controller topics used in camera controller (e.g. for the left camera topics will be */stereo/left/raw_image*).

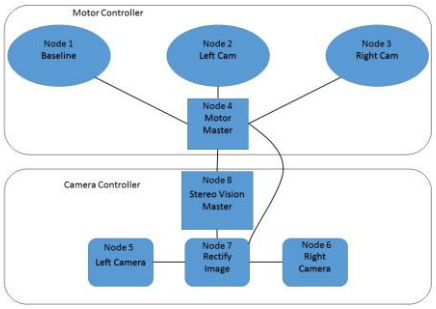

Fig. 5: ROS Controller diagram illustrates the communication between nodes.

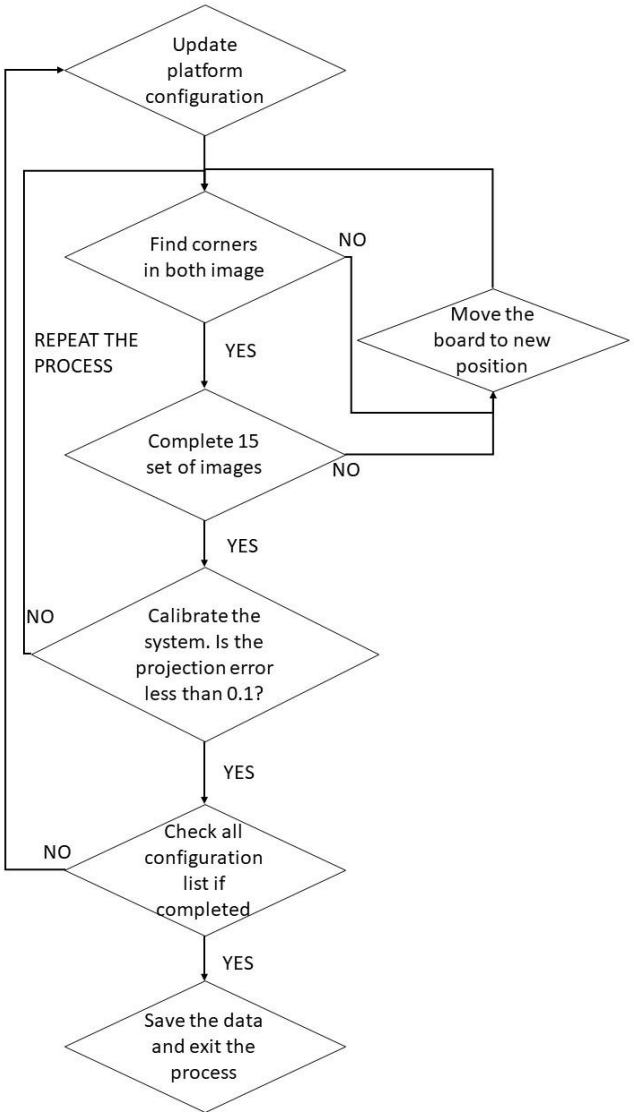

Fig. 6: Flow chart of the calibration process

A desktop computer was used to control the Baxter robot. A connection was set between the two PCs exchanging UDP packets. The UDP connection was used because there were two different versions of ROS (Indigo and Kinetic). The stereo vision platform was using ROS Kinetic and Baxter, ROS

Indigo. The UDP socket was connected to an ROS node on both systems and, when the rig completes the capturing of a photo of the checkerboard, it sends a string through UDP to Baxter's PC signal to move to new position. When the robot gets to the new position, it sends a string back to the platform to confirm the checkerboard is in a new position. The flow of the process is shown in Fig. 6.

The experiment setup is shown in Fig. 7, where the stereo vision platform was fixed to be in front of the robot. The origin of Baxter is laying on the Z axis of the platform and the distance between both origin was set to be 3 meters.

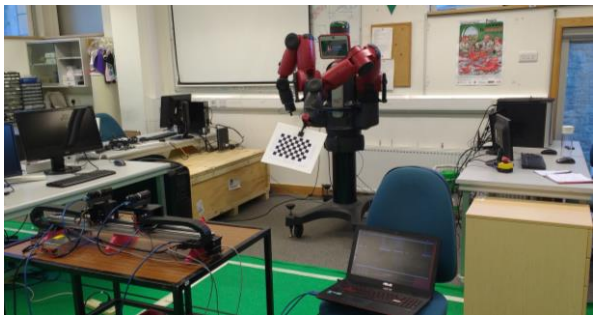

Fig. 7: Baxter holding the checkerboard while the rig works on the calibration (to the left bottom of figure).

2.3 Data

The data of this experiment describe the error generated during the calibration process. This error is given by the function *stereoCalibrate* in OpenCV library (Bradski & Kaehler 2008) where it returns the projection error of the points found in the views that describe how precise the parameters were acquired (Bradski & Kaehler 2008). This output is an important result to get better calibration parameters.

On the other hand, the calibration process for an active stereo vision system is time consuming where the calibration is required to be done multiple times under different configurations. Therefore, the time is measured in the experiment from the beginning of the calibration process until it completes a full set, which are 30 runs of calibration and each run has a different setup.

3. RESULT AND DISCUSSION

The experiment was to compare the speed and the quality of the calibration process of the active stereo rig where the calibration process runs for 30 times to calculate the geometrical dimensions of the rig under a different configuration. Where the left and right camera were systematically varied between ±16 deg while the baseline was fixed at 200mm. However, in the manual calibration, the checkerboard was carried by a human being (the first author of this work), and the rig took a picture every 2 seconds. These two seconds allowed the checkerboard to be moved to a new position; each calibration run has to have sixteen pictures to do the calibration process.

For the automated calibration, everything was set where the Baxter Robot was holding the checkerboard to move it, as explained in the experiment setup, and after the pictures were taken the platform updates itself to move to the new configuration.

The error generated by using the *stereoCalibrate* function for both the manual calibration process and the automated one is shown in Fig. 8. The error shown in the figure is the average error for a complete run. In addition, the error bar presents the Standard Deviation error of the 30 runs. The projection error in the automated process was smaller by 120% compared to the manual calibration, and this occurs due to the lack of precision of a human being to position the checkerboard and stand still during the time the picture is taken. The result shows that the margin error was dropped to be within ±0.17 pixel while the manual calibration margin error is ±0.38. The data statistically analysed using a t-test to show that the two means were not the same with $p_{value} = 3.35 \times 10^{-5}$.

Fig. 8: Projection Error generated by the stereoCalibrate function for the manual calibration process and automate calibration process.

The time consumed during the calibration process of both the manual calibration and the automated calibration is shown in *Fig. 9*. The result clearly shows using a robot to do the calibration process accelerate the speed of the calibration by three times where the manual calibration took 120 minutes to complete a full set, while, on the other hand, the automated process took only 45 minutes. The result shows that the automated process has the potential of doing more experiments in the stereo vision rig to get data that are more robust.

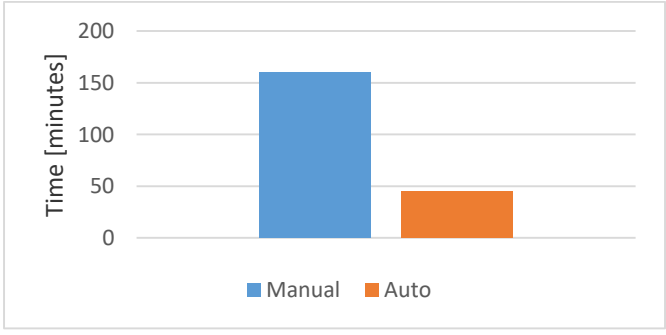

Fig. 9: The time took to complete a set of the calibration process.

The experiment compared the performance of a human being and the robot in a repeatable process where the robot managed to improve the quality of the calibration and drop the time of the calibration process by three times that the human took to complete one run. This large difference was due to the time spent moving between the calibration position and the PC's of the stereo vision platform to update the new configuration.

4. CONCLUSION AND FUTURE WORKS

To conclude, this paper presents an upgrade in the calibration process of an active stereo vision rig that requires calculating the external parameters of the rig in order to evaluate the performance where many calibration runs should be done. The calibration process was upgraded to use a collaborative robot that holds the checkerboard and moves it around. Where one of the advantages of using Baxter was to move the checkerboard to exactly the field of view limits of the cameras when the two cameras verge to inside. The result of the experiment concluded that, by automating the calibration process, the time and the quality of the calibration were improved. The quality of the calibration, expressed here as the decrease of the projection error, was improved by 120% and the total time spent during the process was reduced by 300% when compared to the traditional manual system.

ACKNOWLEDGEMENT

This work was in part supported by the CAPES Foundation, Ministry of Education of Brazil (scholarship BEX 1084/13-5) and UK EPSRC project BABEL (EP/J004561/1 and EP/J00457X/1).

5. REFERENCES

Alexander, S., Daniel, K. & Wahl, F.M., 2010. The Basis of Control-Related Robotics Research - Open High-Rate Low-Level Control Architectures for Industrial Manipulators. In *Robotics (ISR), 2010 41st International Symposium on and 2010 6th German Conference on Robotics (ROBOTIK)*. [VDE Verlag].

Bjorkman, M. & Eklundh, J.-O., 2002. Real-time epipolar geometry estimation of binocular stereo heads. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(3), pp.425–432.

Bradski, G.R. & Kaehler, A., 2008. *Learning OpenCV : computer vision with the OpenCV library*, O'Reilly.

Chapman, R.F. (Reginald F., 1998. *The insects : structure and function*, Cambridge University Press.

Dankers, A. & Zelinsky, A., 2004. Digital Object Identifier. CeDAR: A real-world vision system Mechanism, control and visual processing. *Machine Vision and Applications*, 16, pp.47–58.

Das, S. & Ahuja, N., 1995. Performance analysis of stereo, vergence, and focus as depth cues for active vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(12), pp.1213–1219.

Klarquist, W. & Bovik, A., 1997. Adaptive variable baseline stereo for vergence control. *1997 Ieee International Conference on Robotics and Automation - Proceedings, Vols 1-4*, (April), pp.1952–1959.

Krotkov, E., Henriksen, K. & Kories, R., 1990. Stereo ranging with verging cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(12), pp.1200–1205.

Kuang, X. et al., 2012. Active vision during coordinated head/eye movements in a humanoid Robot. *IEEE Transactions on Robotics*, 28(6), pp.1423–1430.

Lee, Y., Toh, K.-A. & Lee, S., 2008. Stereo image rectification based on polar transformation. *Optical Engineering*, 47(8), p.87205.

Luong, Q.-T. & Faugeras, O.D., 1997. Self-Calibration of a Moving Camera from Point Correspondences and Fundamental Matrices. *International Journal of Computer Vision*, 22(3), pp.261–289.

Mišeikis, J. et al., 2016. Automatic Calibration of a Robot Manipulator and Multi 3D Camera System. In *2016 IEEE/SICE International Symposium on System Integration (SII)*. pp. 735–741.

Nakabo, Y. et al., 2005. Variable Baseline Stereo Tracking Vision System Using High-Speed Linear Slider. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*. IEEE, pp. 1567–1572.

Pradeep, V., Konolige, K. & Berger, E., 2014. Calibrating a Multi-arm Multi-sensor Robot: A Bundle Adjustment Approach. In Springer Berlin Heidelberg, pp. 211–225.

Quigley, M. et al., 2010. Low-cost accelerometers for robotic manipulator perception. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, pp. 6168–6174.

Quigley, M. et al., 2009. ROS: an open-source Robot Operating System. In *ICRA Workshop on Open Source Software*.

Sahabi, H. & Basu, A., 1996. Analysis of error in depth perception with vergence and spatially varying sensing. *Computer Vision and Image Understanding*, 63(3), pp.447–461.

Szeliski, R., 2011. *Computer vision : algorithms and applications*, Springer.

Xiao, Z. et al., 2010. A cross-target-based accurate calibration method of binocular stereo systems with large-scale field-of-view. *Measurement*, 43(6), pp.747–754.

Zhang, Z., 2000. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), pp.1330–1334.