Khan, A., Aragon-Camarasa, G., Sun, L., and Siebert, J. P. (2016) On the Calibration of Active Binocular and RGBD Vision Systems for Dual-Arm Robots. In: IEEE International Conference on Robotics and Biomimetics, Qingdao, China, 03-07 Dec 2016.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

http://eprints.gla.ac.uk/131297/

Deposited on: 10 November 2016

On the Calibration of Active Binocular and RGBD Vision Systems for Dual-Arm Robots

Aamir Khan[a], Gerardo Aragon-Camarasa[a], Li Sun[b], J. Paul Siebert[a]

Abstract — This paper describes a camera and hand-eye calibration methodology for integrating an active binocular robot head within a dual-arm robot. For this purpose, we derive the forward kinematic model of our active robot head and describe our methodology for calibrating and integrating our robot head. This rigid calibration provides a closed-form hand-to-eye solution. We then present an approach for updating dynamically camera external parameters for optimal 3D reconstruction that are the foundation for robotic tasks such as grasping and manipulating rigid and deformable objects. We show from experimental results that our robot head achieves an overall sub millimetre accuracy of less than 0.3 millimetres while recovering the 3D structure of a scene. In addition, we report a comparative study between current RGBD cameras and our active stereo head within two dual-arm robotic testbeds that demonstrates the accuracy and portability of our proposed methodology.

I. INTRODUCTION

RGBD camera sensors (i.e. Kinect-like cameras) have had an impact in robotics and robot vision research as they have provided a low-cost, ready-to-use and off-the-shelf sensor to accommodate different robotic configurations, settings and tasks. However, the accuracy of RGBD camera sensors varies according to the distance between the object to be imaged and the sensor [1]. That is, RGBD sensors limit the perceptual capabilities of robots since they provide low-resolution depth maps and usually suffer from image noise. Likewise, their rigid configuration does not allow the robot to adjust the cameras' physical configuration in order to image objects at different distances from the camera with high accuracy. These limitations make characterising the 3D structure of a given object challenging.

Photogrammetric vision systems provide the required accuracy but at a high cost. To mitigate these costs without compromising accuracy, we have designed an active binocular robot head system with off-the-shelf components for an industrial dual-arm Yakasawa robot (Figure 1) under the FP7 CloPeMa project[1]. This robot head is capable of changing its gaze under computer control. Our robot head has been used successfully for clothes perception and manipulation research [2], [3], [4] because of its ability to provide high resolution imaging for 3D mapping and range sensing. Due to the ability of the robot head to target different parts of a scene, it is required to maintain accurate calibration of its intrinsic and extrinsic parameters with respect to the robot's reference frame.

[a] School of Computing Science, University of Glasgow, G12 8QQ, UK
`gerardo.aragoncamarasa@glasgow.ac.uk`
[b] Intelligent Robotics Lab, University of Birmingham, B15 2TT, UK
[1] `http://www.clopema.eu/`

Fig. 1: Top: Our active binocular robot head. Bottom: Dual-arm Yakasawa robot with our robot head integrated.

In this paper, we describe the methodology we used to calibrate our active robot head and describe our developed solution to dynamically update cameras' extrinsic parameters in order to achieve geometric compatibility with respect to robot's kinematic chain. Finally, we provide a comparative evaluation between our robot head and RGBD cameras while integrating them within a robot's kinematic frame.

II. BACKGROUND

For camera systems under dynamic actuation, either in a monocular or stereo configurations, researches have proposed rigid, continuous self-calibration and a combination of both. A summary of current approaches is provided below.

Rigid calibration methods consist of estimating intrinsic and extrinsic camera parameters, and the mechanical relationships between their actuation platforms and camera reference frames ([5], [6], [7]). Optimisation routines such as bundle adjustment [8], are applied to reduce back projection errors from 2D to 3D measurements and, consequently, arrive to a close-form and stable solution within a defined camera parameter space. However, there exists two main limitations for rigid approaches. On the one hand, mechanical

wear and tear is not considered and the solution obtained depends on the quality of the mechanical parts at the moment the calibration was carried out. On the other, the distance and orientation between actuator joints have to be precisely measured. Hence, errors induced by the measuring device and mechanical backlash during their operation are not taken into account.

Self-calibration methods of PTU camera systems have been extensively researched for the last 20 years and is still an active research area. Nevertheless, a general and generic solution has yet to be devised. For instance, self-calibration approaches have constrained the solution to one degree of freedom for each camera [9], or individual yaw movements per camera plus a joint neck movement for the head [10]. Adding degrees of freedom therefore results in lower precision of the reconstructed scene [11]. To overcome constrains on the kinematic structure of the systems, hybrid approaches consist of initialising calibration parameters using rigid solutions and, then, update intrinsic and extrinsic camera parameters while interacting with the environment. These approaches are based on particle filtering techniques to update camera and external orientation of the cameras with respect to a world reference frame [12]. As our robot also handles highly-deformable objects, these approaches fall short in terms of accuracy and precision performance since they attain an overall precision within 0.5cm and 1cm [13].

In order to overcome the above limitations, we propose a hybrid approach that handles dynamic content in the observed scene. Our approach consists of simple yet robust techniques. We must note that we do not attempt to solve the general problem of self-calibration but provide a robust and general purpose approach to handle both deformable and non-deformable objects. We have therefore assumed that our vision system does not change its focus and we have adjusted the focal length, aperture and shutter settings for each camera to obtain a desired focus range.

III. Materials

A. RGBD Cameras

1) ASUS Xtion Pro: The ASUS Xtion Pro is an RGB-D camera that uses depth sensing technology. It is based on structured light and has a depth range from 0.8 to 3.5 m, a 3D point cloud resolution of 640×480, an RGB image resolution of 1280×1024, frame rate of 30 fps and a latency of 1.5 frames. The depth error of this sensor decreases according to the increasing distance between the sensor and the object. After calibrating the sensor, it achieves a depth accuracy at best of 4.7 mm at a distance of 0.96 m and drops to 38.6mm at a distance of 3.6m [14].

2) Kinect V2: Kinect v2 is also an RGB-D camera from Microsoft based on time of flight technology. It has a depth range of 0.5 to 4.5 m, 3D resolution of 512×424, RGB resolution of 1920×1080, frame rate of 30 fps and latency of 20 ms. Similar to the ASUS Xtion Pro, the depth error of the Kinect V2 sensor decreases according to the distance between the sensor and the object. It is evident that the distance measurements delivered by this sensor are much

more precise than the ASUS Xtion Pro. However, the Kinect V2 reconstruction error is approximately 20mm at a distance of 3 m [15].

B. Active Binocular Robot Head

For our active binocular robot head, we employed relatively inexpensive and commercially available components in order to allow us to capture high-quality 3D depth maps and dense point clouds for deformable object recognition [3] and manipulation [2]. Hence, our robot head comprises two off-the-shelf Nikon DSLR cameras (D5100) that capture 16 MegaPixels images (MP) every 400ms. Each camera is mounted on two degrees of freedom pan and tilt platforms (PTU-D46). Cameras are rigidly separated by a pre-defined baseline for optimal stereo capturing. We interface our active robot head to an Intel Core i7-3930K computer with 32GB of RAM running Ubuntu and ROS.

Fig. 1 depicts the robot head as mounted on our dual-arm robot. The active robot head's visual capabilities include: autonomous gaze control and visual search based on SIFT features [16]. To achieve real-time performance, we have implemented a GPU variant of SIFT [17]. GPU-SIFT features are used for verging the cameras and for tracking features in the scene for our dynamic calibration routine (Section IV-D). Likewise, we use a GPU version of our stereo matcher [18] to compute horizontal and vertical disparities of two images captured by our robot head. Thus, our robot head can produce 16MP depth maps and point clouds in 0.2 fps.

C. Robots

In this paper, we evaluate the developed calibration methods over two different robotic testbeds. The first testbed is a dual-arm industrial robot manufactured by YASKAWA Motoman, as shown in Fig. 1. This robotic testbed comprises two MA1400 manipulators and a custom made turntable. Each arm has 6 DOF and features 4 kg maximal load weight, 1.4 meters of maximal reaching distance, and ± 0.08 mm accuracy. The dual-arm robot is powered and controlled by the DX100 controller. Our active binocular robot head is rigidly mounted on the turntable and in between both robot arms.

The second testbed is the Rethink Baxter robot (Figure 2). Baxter is a humanoid, anthropomorphic robot with two seven degree-of-freedom arms and state-of-the-art sensing technologies. Its key purpose is to be able to operate continuously within humans environments and run for longer periods of time. Baxter's positional accuracy is $\pm 5mm$ with a maximal reaching distance of 1.2 meters.

IV. Methodology

To obtain metrically accurate depth maps under dynamic camera motion, we have adopted an hybrid approach. That is, we first employ rigid camera calibration routines to obtain intrinsic and extrinsic camera parameters and then we find the rigid Euclidean transformations between the robot and calibration target and the robot head and robot (Section IV-B and IV-C, respectively). Dynamic, on-line, calibration in

Fig. 2: Rethink Baxter robotic holdin the calibration target. In our experiments, we placed the RGBD camera (Kinect V2 in the figure) at the top of the robot's head.
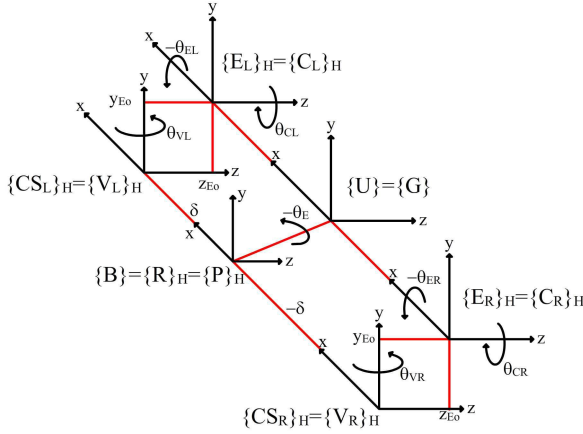

Fig. 3: Coordinate frames in the HOME position of the CLoPeMa robot head.

turn consists of tracking known 3D positions from stable interest point features observed from previous camera poses (Section IV-D). Tracked features are then used to update camera extrinsic parameters accordingly.

A. Forward Kinematics Derivation

In order to find the geometric transformations to integrate our robot head with the robot, we firstly need to deduce the forward kinematic model of the active binocular robot head. Figure 3 depicts the coordinates frames in the "home" position of the robot head (H subscript in the figure). We use Sharkey et. al. [19] notation to express our forward kinematic model. Table I defines the coordinate frames. We assumed that the world reference frame, $\{W\}$, lies on the base, $\{B\}$, of the robot head and $y_{CS_0} = z_{CS_0} = \theta_C = 0$. Sharkey et. al. [19] established that if the cameras are targeting the same point X in $\{W\}$, then both open kinematic chains for each camera will be closed around X. Therefore, the forward kinematics for the open kinematic chain from the base to the

TABLE I: Coordinate frames definition.

{W}	World
{B}	Base
{R}	Roll (Not considered)
{P}	Pan (Not considered)
{CS$_{R/L}$}	Camera Seperation (left / right) Offsets : $(0, y_{CS_0}, z_{CS_0})$; and from {B} is given by $+ / - \delta$
{E$_{R/L}$}	Independently Elevated (left / right)
{V$_{L/R}$}	Vergence (left / right) Offsets wrt to {E} : $(0, y_{V_0}, z_{V_0})$
{C$_{L/R}$}	Cyclotorsion (left / right) (Not considered)
{O$_{L/R}$}	Optical (left / right) Offsets wrt {C} or {V} (for systems without {C}) : (α, β, γ) and (u, v, w)
{T}	Target object. For simplicity, both centres represent a single point in {W} Offsets wrt $z_{OL/OR}$: d_{TL} and d_{TR}

target position is found as:

$$
{}^{B}X = H_T^W \times \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^T =
$$
$$
\begin{bmatrix} s\delta + d_T Cos(\theta_E) Sin(\theta_V) + z_{E_o} Sin(\theta_V) \\ d_T Sin(\theta_E) + y_{E_0} \\ d_T Cos(\theta_E) Cos(\theta_V) + z_{E_o} Cos(\theta_V) \\ 1 \end{bmatrix} \quad (1)
$$

where ${}^{B}X$ is the point w.r.t. the base of the robot head, H_T^W is the transformation between world and target reference frames, s indicates if the camera separation is for the left, $s = 1$, or right, $s = -1$, camera. Thus, Equation 1 defines both forward kinematic chain for the left and right camera.

In practice, the above forward kinematic chain is not closed as we assume that the tilt reference frame is aligned to the optical image plane. In our robot head, it is required to find the geometric transformation from the tilt reference frame to the principal point of the camera's image plane, ${}^{E}T_C$ ($\{E\}$ and $\{C\}$ as defined in Table I, respectively). To find H_C^E, we first need to carry out camera calibration to find the intrinsic geometry of the camera and hence the principal point of the camera's image plane as described on what follows.

B. Camera Calibration

To find intrinsic camera parameters, we used the OpenCV calibration routines[2] to sample the cameras' parameter spaces using different positions and orientations of the OpenCV's check-board target. We attached the calibration target to the robot gripper and used the dual-arm robot to automate the target's sampling process. Our robot head is converged to a fixed point in the robot's space and the robot head remains fixed on this point during camera calibration. We thus perform calibration for each camera in isolation by presenting the calibration target at different positions and orientations in order to sample the camera parameter space adequately.

For each sampled target position, OpenCV returns the cameras' pose w.r.t. the target. However, we found that by optimising these poses using sparse bundle adjustment[3], we were able to obtain a more accurate estimation of the camera's pose than OpenCV stereo calibration routines. Hence, we find the stereo geometric relationship between

[2]http://opencv.org
[3]https://sourceforge.net/projects/cvsba/

cameras by triangulating kinematic transformation across the calibration target as:

$$H_{O_R}^{O_L} = H_T^{O_L} \left(H_T^{O_R} \right)^{-1} \qquad (2)$$

C. Hand-Eye Calibration

To close the kinematic chain of the robot head and, consequently, the robot integrating the robot head, we require to find:

- The transformation between the robot gripper and the calibration target – $H_{gripper}^T$; and,
- The transformation from tilt (E) and the camera frame (O) for the left and right camera – H_O^E.

To find the above transformations, we implemented Tsai's hand-eye calibration [20] as a ROS node. Thus, for $H_{gripper}^T$, we feed the hand-eye calibration routine with the left camera poses and the forward kinematic chain of the robot for each sampled target position as described in Section IV-B. Note that the kinematic chain of the robot is defined using ROS Unified Robot Description Format (URDF) (http://wiki.ros.org/urdf) and we retrieve transformations using TF (http://wiki.ros.org/tf).

By obtaining the pose of the camera with respect to the robot base for both left and right cameras, we are now able to estimate H_O^E. The strategy adopted therefore consists of sampling random PTU pan and tilt positions while capturing images of the calibration target fixed. For each camera movement, we compute the forward kinematic chain for each camera using the kinematic model described in Section IV-A. Finally, we use our implemented ROS node for each camera in order to obtain the geometric transformation that relates $\{E\}$ and $\{C\}$.

We must note that we treat each actuated pan and tilt unit (PTU) separately during hand-eye calibration. We assume that the world reference frame lies on the base of each PTU of the robot head. If the cameras are targeting the same point in the world, then both open kinematic chains for each camera will be closed around this world point.

Tsai's hand-eye calibration algorithm provides us with an estimation of the kinematic transformation from the base of the robot to the gripper holding the calibration target and from the gripper to the calibration target. It is then possible to find the forward kinematic transformation from the robot's world reference frame to the left camera in the robot head's "HOME" position by triangulating transformations. For the right camera, we know the transformation of the stereo configuration; therefore, we can conclude that the kinematic chains for each camera will be closed around the robot base. The complete camera-hand eye calibration routine is depicted in Figure 4.

D. Dynamic Tracking of Extrinsic Camera Parameters

The estimated forward kinematic chain in the previous section holds true only when the cameras do not change their gazing point and remain in the "HOME" position. This is because mechanical information of the PTUs is inaccurate and, in consequence, Tsai's hand-eye calibration
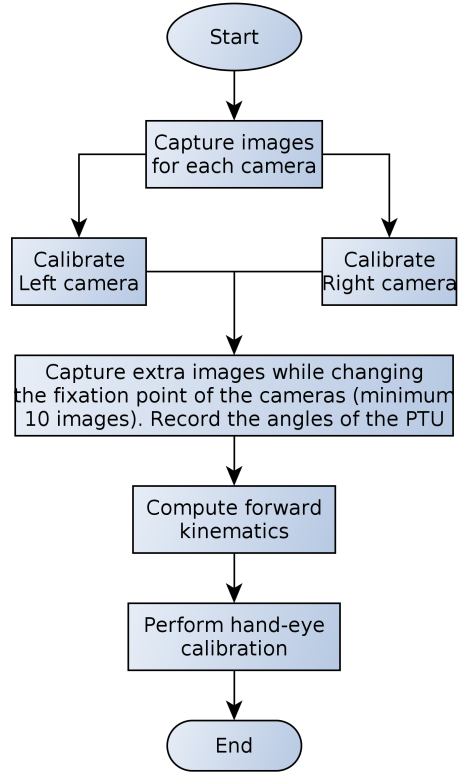

Fig. 4: Flow diagram of the camera-hand eye calibration pipeline.

routine finds an optimal solution for the sampled poses. Thus, in order to maintain dynamic calibration, we compute the Euclidean transformation every time the cameras move by using known visual information from previous camera positions. Hence, we have to update the transformation from the world reference frame of the robot to the left camera and the left camera to the right camera. We achieve the latter by tracking previously observed and stable 3D points in the scene.

Tracking consists of stereo triangulating SIFT features coordinates from each camera. 3D projection of SIFT coordinates are expressed in terms of the world reference frame of the robot – these features are then used to initialised the dynamic updating routine. A PnP algorithm [21] is used on matched SIFT features between previous and current observations in order to recover the Euclidean transformation from 3D coordinates to the left camera. As outliers are likely to affect the performance of the PnP algorithm, we filter SIFT feature matches by adopting a RANSAC homography fitting strategy from the current image to the previous image. Finally, we compute the stereo relationship between the cameras by computing the missing link from the forward kinematic chains of each camera and the camera projection matrices are updated accordingly for optimal 3D reconstruction.

V. Experiments

A. Active Binocular Robot Head

The ability of the system to maintain dynamic calibration is expressed in terms of the accuracy of 3D reconstruction. The cameras, in this case, can be positioned in any angle range in the tilt and pan axes (within the hardware range limits of the PTU); therefore, the extrinsic properties of the stereo configuration need to be updated as explained in section IV-D. The objective of this experiment is thus to test the accuracy of our binocular robot head to recover and reconstruct an accurate point cloud despite any camera motion.

For this purpose, we employed an object model with known 3D structure. The object model used in these experiments is the check-board calibration target. This planar calibration target allow us to:

1) compare the reconstruction residuals between a calibrated stereo system and the updated extrinsic camera parameters; and,
2) verify the accuracy of the reconstructed geometry by measuring the RMSE between a known 3D (ground truth) measurement on the real object and the 3D reconstructed model.

To measure the accuracy of the system, we sample the calibration target over 10 different poses, by setting the robot head and robot in random position. We then measure the 3D euclidean distance from one point of the calibration target to next in line (Table II). The ground truth distance between squares on the check-board is 24mm. The RMSE between reconstructed points of the check-board while using the dynamic calibration parameters is therefore compared with respect to the ground truth. As observed in Table II, the reconstruction accuracy for X, Y and Z is less than 0.3 mm. We can therefore state that our binocular robot head in combination with the proposed rigid and dynamic calibration is capable of reconstructing the scene within sub millimetre accuracy. Likewise, Table III shows the mean and standard deviation of RMSE (in millimetres) values in Table II. From the results, we can therefore conclude that the system is able to recover the 3D geometry with sub-millimetre accuracy – our dual-arm robot has an accuracy of $\pm 0.08mm$. Hence, the estimated geometric transformations between robot head's and robot's frame are within optimal limits for the required accuracy for practical manipulation tasks.

B. Xtion Pro and Kinect V2

For consistency and completeness of our calibration approach, we repeated the experiments discussed above for two RGB-D sensors i.e ASUS Xtion Pro and Kinect V2, mounted on Baxter's head. Our aim is to test the accuracy of these RGB-D sensors while being integrated within the kinematic's chain of the robot.

We must note that we followed similar rigid calibration steps as for our robot head (Fig. 4). Hence, we calibrated both

TABLE II: Residual errors (in millimetres) between optimal stereo calibration and dynamic calibration of the extrinsic camera parameters. 3D coordinates are expressed with respect to the world reference frame of the robot head.

Pose #	X	Y	Z
1	0.17	-1.03E-02	-3.02E-02
2	1.50E-02	6.37E-02	-4.68E-02
3	2.04E-02	5.14E-02	3.76E-02
4	0.67	-6.69E-02	-0.201
5	0.79	-8.60E-02	-0.111
6	7.82E-02	1.85E-02	-3.47E-02
7	0.42	-3.67E-02	-0.119
8	6.95E-02	1.16E-02	-6.94E-03
9	0.38	-3.22E-02	-6.90E-02
10	0.32	-2.60E-02	1.50E-02
Mean	0.29	-0.011	-0.056
1 Std	0.27	0.025	0.06

TABLE III: RMSE (in millimetres) between estimated 3D reconstructed points and ground truth for the active binocular robot head.

Pose #	Mean	1 Std
1	0.10	0.077
2	0.09	0.034
3	0.12	0.067
4	0.31	0.072
5	0.05	0.060
6	0.12	0.067
7	0.22	0.077
8	0.22	0.077
9	0.05	0.057
10	0.34	0.069
Overall	*0.16*	*0.105*

RGBD sensors using available calibration routines[4]. We then obtain the geometric transformation from robot's gripper and calibration target using our hand-eye calibration routine. By having knowledge of this transformation, we then compute the RGBD camera's position w.r.t. the robot's kinematic chain. The accuracy of the RGBD cameras is shown in table IV for ASUS Xtion Pro and table V for Kinect V2.

From the errors reported in the table IV and table V, we can observe that the system is able to recover the 3D geometry with an overall RMSE of 12.4mm with ASUS Xtion Pro sensor and 11.6mm with Kinect V2 sensor. As observed, the larger accuracy errors are due to the intrinsic properties of the RGB-D sensors (ref. Section III-A) and the positional accuracy of Baxter's robot (ref. Section III-C). Hence, we can conclude that these RGBD cameras are able to recover the 3D geometry of objects and scenes within millimetres accuracy. Note that the latter statement depends on the positional accuracy of the robot test-bed and depth accuracies of RGB-D sensors.

---

[4]For the Xtion PRO, we used: http://wiki.ros.org/camera_calibration; whereas, for Kinect V2, https://github.com/code-iai/iai_kinect2/tree/master/kinect2_calibration

TABLE IV: RMSE (in millimetres) between estimated 3D reconstructed points and ground truth for ASUS Xtion Pro in Baxter.

Pose #	Mean	1 Std
1	13.63	9.35
2	14.04	5.18
3	10.55	5.24
4	11.85	6.15
5	14.63	6.66
6	14.88	5.85
7	15.54	8.56
8	8.83	4.40
9	9.38	3.63
10	11.59	6.40
Overall	*12.49*	*6.14*

TABLE V: RMSE (in millimetres) between estimated 3D reconstructed points and ground truth for Kinect V2 in Baxter.

Pose #	Mean	1 Std
1	13.68	7.401
2	13.42	4.24
3	12.84	4.70
4	12.88	5.58
5	8.09	3.45
6	11.43	5.09
7	9.08	3.97
8	8.44	2.16
9	9.85	3.30
10	16.55	8.38
Overall	*11.62*	*4.83*

VI. CONCLUSION AND FUTURE WORK

In this paper, we described our calibration methodology to calibrate our active binocular robot head system integrated within an industrial dual-arm Yakasawa robot and described a comparative experiment on the integration of RGB cameras into the Rethink Baxter robot. We also presented a simple, yet effective solution to dynamically update cameras' extrinsic parameters in order to achieve geometric compatibility with respect to the robot's kinematic chain. By comparing our binocular robot head with consumer RGBD cameras, we can conclude that our robot head provides an off-the-self depth sensing solution capable of reconstructing the observed 3D scene within sub millimetre accuracy. Likewise, we show that the implemented calibration routines in this paper can provide reliable results that allow reconstruction with relative 3-D errors of less than 0.3 millimetres. An open-source ROS package that implements our calibration methods can be found at `https://github.com/gerac83/glasgow_calibration`. A video demonstration of the calibration of our active binocular robot head can be accessed at: `https://youtu.be/9OYy9Q_bN2w`.

We are currently investigating how to maintain dynamic intrinsic camera calibration (i.e. change of focus) to enable our robot head to focus different depth of fields. We will also validate the extrinsic dynamic calibration in terms of drift errors and accuracy of the computed point clouds.

REFERENCES

[1] K. Khoshelham and S. O. Elberink, "Accuracy and resolution of kinect depth data for indoor mapping applications," *Sensors*, vol. 12, no. 2, p. 1437, 2012. [Online]. Available: http://www.mdpi.com/1424-8220/12/2/1437

[2] L. Sun, G. Aragon-Camarasa, S. Rogers, and J. P. Siebert, "Accurate garment surface analysis using an active stereo robot head with application to dual-arm flattening," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, May 2015, pp. 185–192.

[3] L. Sun, S. Rogers, G. Aragon-Camarasa, and J. P. Siebert, "Recognising the clothing categories from free-configuration using gaussian-process-based interactive perception," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, May 2016, pp. 2464–2470.

[4] L. Sun, G. A. Camarasa, A. Khan, S. Rogers, and P. Siebert, "A precise method for cloth configuration parsing applied to single-arm flattening," *International Journal of Advanced Robotic Systems*, vol. 13, April 2016, cLOPEMA - 288553 (Clothes Perception and Manipulation). [Online]. Available: http://eprints.gla.ac.uk/117227/

[5] J. Neubert and N. J. Ferrier, "Robust active stereo calibration," in *Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on*, vol. 3, 2002, pp. 2525–2531.

[6] J. Salvi, X. Armangu, and J. Batlle, "A comparative review of camera calibrating methods with accuracy evaluation," *Pattern Recognition*, vol. 35, no. 7, pp. 1617 – 1635, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320301001261

[7] H. Kwon, J. Park, and A. C. Kak, "A new approach for active stereo camera calibration," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*, April 2007, pp. 3180–3185.

[8] Y. Furukawa and J. Ponce, "Accurate camera calibration from multi-view stereo and bundle adjustment," *International Journal of Computer Vision*, vol. 84, no. 3, pp. 257–268, 2009. [Online]. Available: http://dx.doi.org/10.1007/s11263-009-0232-2

[9] M. Sapienza, M. Hansard, and R. Horaud, "Real-time visuomotor update of an active binocular head," *Autonomous Robots*, vol. 34, no. 1-2, pp. 35–45, 2013.

[10] T. Dang, C. Hoffmann, and C. Stiller, "Continuous stereo self-calibration by camera parameter tracking," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1536–1550, July 2009.

[11] M. Brückner, F. Bajramovic, and J. Denzler, "Intrinsic and extrinsic active self-calibration of multi-camera systems," *Machine Vision and Applications*, vol. 25, no. 2, pp. 389–403, 2014. [Online]. Available: http://dx.doi.org/10.1007/s00138-013-0541-x

[12] G. R. Mueller and H. J. Wuensche, "Continuous extrinsic online calibration for stereo cameras," in *2016 IEEE Intelligent Vehicles Symposium (IV)*, June 2016, pp. 966–971.

[13] P. Hansen, H. Alismail, P. Rander, and B. Browning, "Online continuous stereo extrinsic parameter estimation," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, June 2012, pp. 1059–1066.

[14] B. Karan, "Calibration of kinect-type rgb-d sensors for robotic applications," *FME Transactions*, vol. 43, no. 1, pp. 47–54, 2015.

[15] D. Pagliari and L. Pinto, "Calibration of kinect for xbox one and comparison between the two generations of microsoft sensors," *Sensors*, vol. 15, no. 11, p. 27569, 2015. [Online]. Available: http://www.mdpi.com/1424-8220/15/11/27569

[16] A. Khan, G. Aragon-Camarasa, and J. P. Siebert, *A Portable Active Binocular Robot Vision Architecture for Scene Exploration*. Cham: Springer International Publishing, 2016, pp. 214–225. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-40379-3_22

[17] C. Wu, "SiftGPU: A GPU implementation of scale invariant feature transform (SIFT)," http://cs.unc.edu/ ccwu/siftgpu, 2007.

[18] P. Cockshott, S. Oehler, T. Xu, P. Siebert, and G. Aragon-Camarasa, "A parallel stereo vision algorithm," in *Many-Core Applications Research Community Symposium 2012*, 2012.

[19] P. Sharkey, D. Murray, and J. Heuring, "On the kinematics of robot heads," *Robotics and Automation, IEEE Transactions on*, vol. 13, no. 3, pp. 437 –442, jun 1997.

[20] R. Tsai and R. Lenz, "A new technique for fully autonomous and efficient 3d robotics hand/eye calibration," *Robotics and Automation, IEEE Transactions on*, vol. 5, no. 3, pp. 345 –358, jun 1989.

[21] L. Ferraz, X. Binefa, and F. Moreno-Noguer, "Very fast solution to the pnp problem with algebraic outlier rejection," in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 501–508.