A new strategy for improving vision based tracking accuracy based on utilization of camera calibration information

Hamza Alzarok, Simon Fletcher, Andrew.P.Longstaff
Centre for Precision Technologies, School of Computing and Engineering, University of Huddersfield, Queensgate, Huddersfield HD1 3DH, England
Email: Hamza.Alzarok@hud.ac.uk

***Abstract*— Camera calibration is one of the essential components of a vision based tracking system where the objective is to extract three dimensional information from a set of two dimensional frames. The information extracted from the calibration process is significant for examining the accuracy of the vision sensor, and thus further for estimating its effectiveness as a tracking system in real applications. This paper introduces another use for this information in which the proper location of the camera can be predicted. Anew mathematical formula based on utilizing the extracted calibration information was used for finding the optimum location for the camera, which provides the best detection accuracy. Moreover, the calibration information was also used for selecting the proper image Denoising filter. The results obtained proved the validity of the proposed formula in finding the desired camera location where the smallest detection errors can be produced. Also, results showed that the proper selection of the filter parameters led to a considerable enhancement in the overall accuracy of the camera, reducing the overall detection error by 0.2 *mm*.**

Keywords*— *Camera calibration, Visual tracking, Industrial robotics

I. INTRODUCTION

Precise camera calibration is an essential element in computer vision tasks (for example in robotics). With the rising demands of obtaining high precision measurements, it has also gained a lot of attention from researchers in this subject [1]. The objective of the camera calibration in visual tracking applications is to retrieve 3D information from a set of 2D images. The error resulting from the camera calibration was considered as one of main factors that affect the accuracy of the tracking system such as in seam welding tracking process[2]. There are three types of information that can be provided by the calibration process which can be listed as follow [3]:1) Intrinsic parameters, 2) Extrinsic parameters, and 3) Lens distortion coefficients. In order to perform the calibration process, the camera should observe an object whose geometrical structure is known with good precision [1]. The calibration techniques can be generally classified based on the existence of the object into two categories: photogrammetric calibration and self (or auto) calibration [4].

Photogrammetric calibration: In this kind of techniques, the use of a calibration object is necessary, the photogrammetric calibration can be further classified according to the dimensions of the object employed for the calibration process:

3D reference object-based calibration: the calibration technique in this category is performed by observing an object usually combining two or three planes orthogonal to each other (see Fig. 1 (right)), on some occasions, a plane with precisely known translation is used. The 3D geometry of the object has to be known prior to the calibration process [4]. Whilst this technique can be performed efficiently[5], it is often impractical due to the requirement to use an expensive calibration apparatus, and a complex setup [4]. Therefore, accurate planar objects are usually preferred over 3D targets because of the simplicity of its setup and use [1].

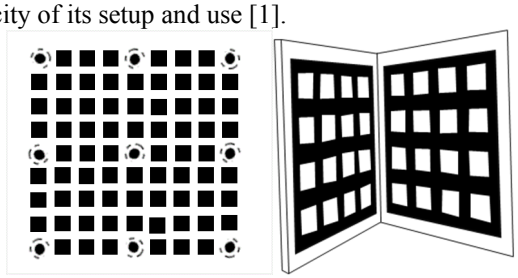

Fig. 1. 2D and 3D calibration apparatus

2D plane based calibration: the calibration object consists of a planar pattern (see Fig. 1 (left)), and the camera is used to observe the object at different orientations [6, 7]. Unlike Tsay's technique [8], the motion of the plane is not required to be known. The advantage of this technique is in the ease with which one can make your own calibration pattern and also in the simplicity of the calibration setup[4].

1D object based calibration: the calibration object consists of collinear points, one of the first examples for using 1D targets as calibration objects was introduced by Zhang's work[4], where a 1D stick with three or more points rotating around a

constant point, this technique should be useful for calibrating multiple cameras.

Self-calibration: It can also be called auto-calibration or 0D object based calibration, in this kind of approaches, there is no object used in the calibration process, therefore, this technique can also be named a 0D approach because only correspondences between the images are required. Due to that there is no calibration object in this approach, a large number of camera parameters is necessary to be estimated, and that makes the mathematical problem more difficult [4].

In visual tracking applications, there are two possible configurations to set up a vision system namely eye-in-hand and eye-to-hand. In the eye-in-hand camera configuration (ENH), the cameras are rigidly attached to the robot end-effector having the image plane for the camera parallel to the plane where the target is moving, which means that there is no a constant location for the camera whereas in the eye-to-hand configuration (E2H), the vision systems are fixed in the workspace[9]. The E2H camera can be located either above the workpiece [10-13] or at a certain distance from the robot arm and its workspace [14, 15]. As far as we are aware, no rule has been made previously in term of selecting the camera location, except for that camera should be located in a place where guarantees a good view for the tracked object during performing its task. This paper aims to present a new technique for selecting the proper location of the camera based on utilising the extrinsic calibration information, a new mathematical formula is proposed which predicts the best camera location that ensures the highest tracking accuracy. Moreover, the information obtained from the calibration process was also used for choosing the proper Denoising filter prior to be applied in the real time application.

II. CALIBRATION ALGORITHM

A. the principles

The camera calibration algorithm used in the experimental part of this paper was provided as a Matlab function within the camera calibration toolbox (version R2014b), it was basically built based on a calibration approach introduced by Zhang [7]. The technique is based on using planar patterns in 3D space for calibrating the camera, and lies between the photogrammetric calibration and self-calibration because of its use of 2D metric information instead of 3D or 1D. The technique was considered as a flexible technique due to two main reasons: 1) the calibration object can be easily obtained by designing your own checkerboard or by generating and printing it via Matlab and attaching it to a flat surface, and 2) the requirement of the calibration can be easily achieved which is to observe the pattern in at least two different orientations. It is worth mentioning the application used is only valid for checkerboard patterns. However, if a different type of pattern is used, designing a new code is required in order to detect the pattern points in the images. Another restriction for using the application is the FOV of the camera which should not be greater than 95 degrees. The calibration algorithm was built based on the assumption that the camera model is a pinhole model.

The principle of the calibration algorithm is to estimate the values of intrinsic and extrinsic parameters as well as the coefficients of the distortion. The calibration process via the algorithm can be summarized into two phases, the first phase is solving the intrinsic and extrinsic parameters via the closed-form solution, assuming that there is no lens distortion [7]. The next and last stage is the estimation of all camera parameters, involving the coefficients of the lens distortion using a nonlinear least square optimization technique called Levenberg–Marquardt (LM) algorithm [7, 16].

B. Evaluation of the calibration results

The calibration accuracy can be evaluated by examining the reprojection errors, extrinsic camera parameters or through seeing undistorted images (see the highlights in Fig. 2). It is recommended to use the three evaluation ways in order to obtain the best results from the calibration process. Also, the number of calibration patterns plays a role in the obtained calibration results. It is recommended from the Matlab function to use between 10 to 20 frames of the calibration pattern for obtaining satisfying results. The reprojection errors were determined by projecting the points of the checkerboard from world coordinates into image coordinates, then a comparison is made between the reprojected points and its equivalent detected points. As far as the author is aware, there is no suggestion what is an acceptable range for the reprojection errors resulting from the calibration process, although it has been mentioned within the camera calibration toolbox that the reprojection errors with less than 1 pixels are generally accepted.

Examination of the extrinsic parameters is another way for evaluating the calibration process, the plot of the 3D extrinsic parameters (as shown bottom right in Fig. 2) provides a centric view of the camera and the patterns which is beneficial in examining the relative positions of both the camera and the pattern in order to observe whether they match what was expected. The use of camera centric view is helpful in the evaluation of the calibration accuracy in the case where a stationary camera is used during the process. The pattern centric view is therefore only beneficial if the pattern was stationary.

Another way for evaluating the calibration accuracy is by observing the undistorted images. It is important to see if all images have straight lines which indicate to that the calibration is accurate, checking the undistorted images is necessary even if the calibration produces small reprojection errors. The reason behind viewing the undistorted image is that in some cases, the checkerboard pattern covers only a small percentage of the image leading to incorrect estimation of the distortion, even though the reprojection errors resulted from the calibration are small.

In this paper, the extrinsic camera parameters extracted from the calibration process were utilized in the selection of a location for the camera where the best detection performance can be achieved, a mathematical model was built and examined for estimating the appropriate location. Moreover, the extrinsic information was also used for selecting the suitable imaging filter.

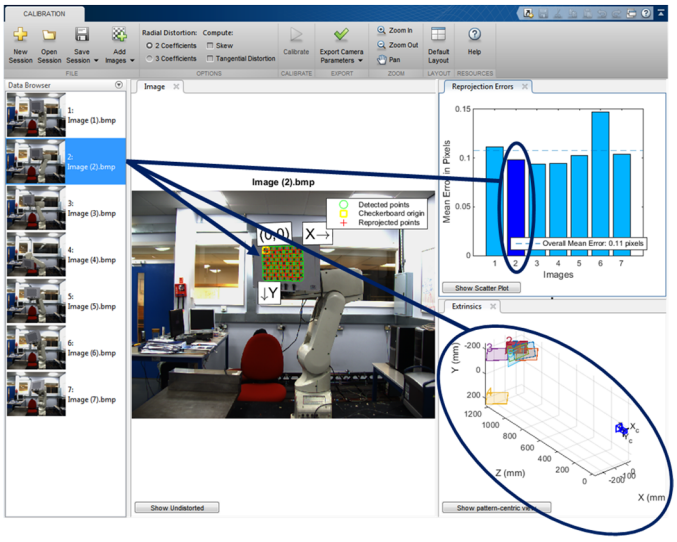

Fig. 2. Evaluation ways for the calibrations results

III. EXPERIMENTAL SETUP

A CCD camera (namely DFK Z12GP031 from the Imaging Source with resolution of 1.23 Megapixels and a pixel size of 4.84 µm^2) was positioned in an E2H configuration at approximately 150 cm from the workpiece plane. The vision sensor was used to capture frames at designated points within the working volume of a robot. A 6 DOF robot (namely Mitsubishi RV-1A) was programmed to perform a linear movement path, while a calibration board (a checker board with a square size of 15 mm) was kept fixed to the end-effector as shown in Fig. 3. The size of the checker board was 180 mm × 135 mm.

The experimental work can be divided into two parts: the first involves finding the best location of the camera based on the extrinsic camera information, and examining the prediction of the appropriate location via a mathematical model. The second part is utilising the calibration process for selecting the proper imaging filter that can considerably enhance the visual detection of the tracking system.

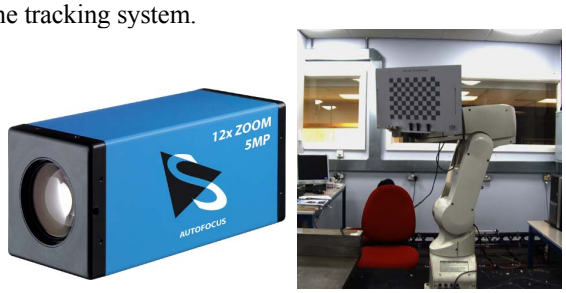

Fig. 3. Experimental setup (*right*), the vision sensor used (*left*)

A. Selection of the camera location

In this part, the robot is programmed to perform a simple task (pick and place) while its end-effector is holding the checkerboard (Fig. 3 (right)), the robot was commanded to stop at designated points, and the camera was used to capture frames for the checkerboard at each of these points. From the camera centric view (Fig. 4), it can be seen that each checkerboard pattern has four corners, the 3D coordinates of these corner points can be extracted through the pattern centric view from which the vertical displacement (Δy) was calculated for each pattern by using equation (1). This is then followed by calculating the vertical displacement value which is the summation of the vertical displacement for all patterns (see equation (2)).

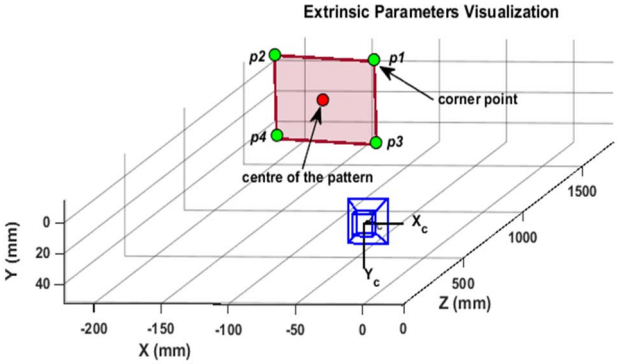

Fig. 4. Checkerboard pattern from the camera view plot

$$\Delta y = (y_{p1} + y_{p2}) - (y_{p3} + y_{p4}) \quad (1)$$

Where $y_{p1}, y_{p2}, \dots y_{p4}$: the y coordinates of the corner points for each checkerboard pattern

$$\Delta Y = \Delta y_1 + \Delta y_2 + \Delta y_3 + \cdots \Delta y_n \quad (2)$$

Where n: the number of calibration patterns

Essentially, if there is a balance between viewing the checkerboard form different sides, the results from equation 2 should be zero.

The next step is measuring the detection accuracy of the vision system, which is the difference between the position of the robot end-effector extracted from the calibrated images and the actual position measured by the reference system. A laser tracker (namely Faro ION) was used for measuring the actual position of the robot end-effector at designated points, one retro reflector (SMR 0.5) was mounted near to the robot end-effector.

In order to find the coordinates of the robot end-effector from the known coordinates of the checkerboard pattern, the centre point of the patterns (see Fig. 4) should be calculated first by averaging the four corner points for each pattern, then a homogeneous transformation defined by a 4×4 matrix, was performed by the applying rotation given by $R(\alpha, \beta, \gamma)$, then followed by a translation given by x_t, y_t, z_t. The result is:

$$T = \begin{pmatrix} \cos\alpha\cos\beta & \cos\alpha\sin\beta\sin\gamma - \sin\alpha\cos\gamma & \cos\alpha\sin\beta\cos\gamma + \sin\alpha\sin\gamma & x_t \\ \sin\alpha\cos\beta & \sin\alpha\sin\beta\sin\gamma + \cos\alpha\cos\gamma & \sin\alpha\sin\beta\cos\gamma - \cos\alpha\sin\gamma & y_t \\ -\sin\beta & \cos\beta\sin\gamma & \cos\beta\cos\gamma & z_t \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (3)$$

According to Euler 1, the homogeneous transformation matrix (HT) can be separated into six parameters $x_t, y_t, z_t, \alpha, \beta, \gamma$. It is worth mentioning that the calculation for those elements was automatically obtained by using a simplex optimisation technique.

After applying the transformation matrix, the results obtained will represent the measured positions of the end-

effector, the next step is to measure the positioning error which is the difference between readings obtained from the camera and those measured by the laser tracker. In order to obtain the end-effector positions from readings of the laser tracker, the transformation matrix (3×1) is necessary to be calculated. The parameters of this matrix represent the translations in 3 dimensions which can be denoted as x_t, $y_t$and z_t.

After obtaining the summation of the vertical displacement for the patterns (ΔY) and the overall detection error at the first camera location, the camera was located at different places, and the same previous steps followed with the old location will be also followed at the new locations. Fig. 5 shows the locations of the camera used during the experiment, and Table IV shows a summary of the obtained results. It can be noted that two smallest averaged displacements were measured in the locations 3 and 4. Therefore, the camera was relocated between the two mentioned locations, and both the averaged displacement and the overall detection error are measured, the results showed that a small detection error (1.25 mm) and smallest averaged displacement (0.25 mm) were obtained at the new location which can be called the optimum camera location.

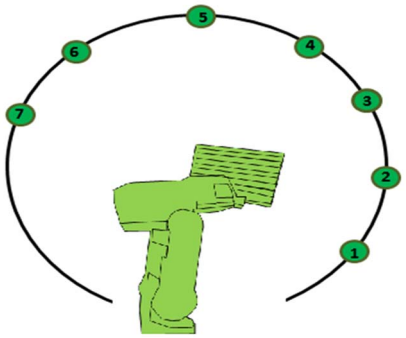

Fig. 5. Camera locations during the experiments.

TABLE IV. DETECTION ERRORS AND AVERAGED DISPLACEMENTS

locations	Overall error (mm)	Δy (mm)
LOC1	1.52	84.25
LOC2	0.93	46.25
LOC3	1.31	12.125
OPTIMUM	1.25	0.25
LOC4	2.79	-26.875
LOC5	3.71	-57.375
LOC6	3.43	-100

B. *Predictive mathematical model*

Let us consider the first location for the camera is a reference location, and has displacement of ($\Delta Y1 = 84.25$), and the next location is ($L2$) which has a displacement of ($\Delta Y2 = 46.25$). The distance between the two locations (d) is 53 cm

$$\therefore \frac{\Delta Y2 - \Delta Y1}{d} = \frac{46.25 - 84.25}{53} = -0.717$$

Again, let us again consider the first location for the camera is a reference location, and has displacement ($\Delta Y1$=84.25), and the next location is ($L4$) which has a displacement ($\Delta Y4$=-26.875). The distance between the two locations (d) is 150 cm.

$$\therefore \frac{\Delta Y5 - \Delta Y1}{d} = \frac{-26.875 - 84.25}{150} = -0.74$$

From the two previous examples, it can be noted that when we divide the difference in the summated vertical displacement value between two locations by the distance between them, the result is nominally a constant value which can be symbolized by (f). And therefore, the mathematical expression can be written as follow:

$$f = \frac{\Delta Yn - \Delta Yr}{d} \; or \; \Delta Yn = \Delta Yr + f \times d \qquad (4)$$

Where: ΔYn is the predicted displacement of the required location (mm)

ΔYr: The displacement of the reference location (mm)

f: Constant factor

d: The distance between the reference and required location (cm)

In order to predict the displacement of any camera location, the displacement of two locations at least is necessary to be known, so the factor (f) can be calculated. Table V shows the actual and predicted displacements at different camera locations, the first location of the camera was used as a reference location.

TABLE V. MEASURED AND PREDICTED DISPLACEMENT AT DIFFERENT LOCATIONS

Locations	ΔYr	d	f	Predicted ΔY	Measured ΔY
LOC2	84.25	53	-0.717	46.25	46.25
LOC3	84.25	114.5	-0.717	2.15566	12.125
OPTIMUM	84.25	121	-0.717	-2.50472	0.25
LOC4	84.25	150	-0.717	-23.2972	-26.875
LOC5	84.25	189	-0.717	-51.2594	-57.375

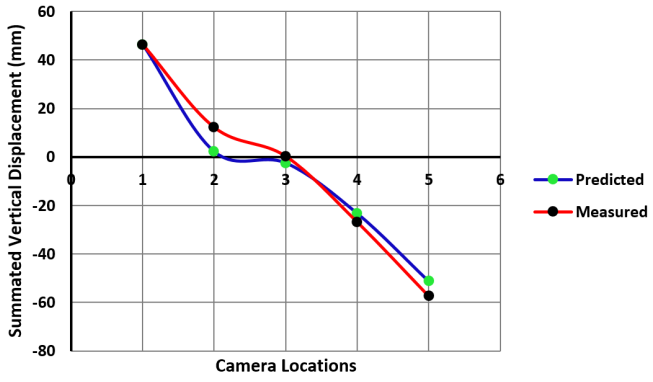

Fig. 6. Measured and predicted displacement at different locations.

In order to proof the validation of the proposed model, the experiments had been repeated with a different motion scenario

for the robot, in which means the calibration pattern was moved to different positions, the results can be shown in table below:

TABLE VI. MEASURED AND PREDICTED DISPLACEMENT AT DIFFERENT LOCATIONS

Locations	ΔYr	d	f	Predicted ΔY	Measured ΔY
lOC3	53.75	43.7	-0.88	15.294	11
lOC4	53.75	85.5	-0.88	-21.49	-21.125
LOC5	53.75	148	-0.88	-76.49	-78.625
OPTIMUM	53.75	62.5	-0.88	-1.25	0.375

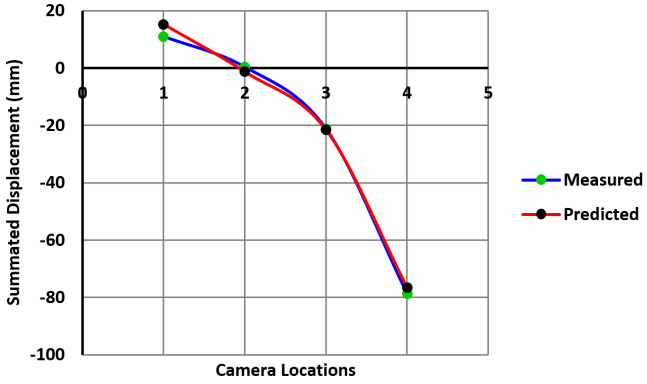

Fig. 7. Another examination for the validity of the predictive formula.

From Fig. 7, it can be seen how the predicted displacements are close to those measured from the calibration process, the accuracy of the prediction basically depends on how precisely the distance between the reference location and the desired location was measured.

C. Selection of the proper camera filter

In visual tracking applications, it is important that the captured images are clear and stable in order to obtain reliable information about the tracked object, the aim of this part is to select the proper averaging filter based on the calibration information. Similar to the procedures followed for measuring the detection error of the robot end-effector, the robot was programmed to move to different points. While holding the checkerboard, the camera was used to capture the pattern at each point after applying an averaging filter. Three different averaging filters were used in which differ in the number of averaged images, 4 images, 16 images, and 32 images.

The HT matrix with a simplex optimisation technique was applied for obtaining the 3D position of the end-effector at the specified points, and a laser tracker was also used for obtaining the reference positions and for evaluating the detection accuracy of the vision sensor.

TABLE VII. AVERAGING FILTERS VERSUS 3D DETECTION ERRORS

Averaging filter	Ex (*mm*)	Ey (*mm*)	Ez (*mm*)
4	0.47	0.34	0.42
16	0.28	0.25	0.34
32	0.44	0.3	0.35

Table VII shows that using the averaging of 16 images enhanced the calibration accuracy compared with other averaging cases, and thus improved the detection for the 3D coordinates of the robot end-effector. The overall detection errors was reduced by 0.18 mm when the averaging of 16 images was applied instead of 4 averaged images. It can also be seen that using a higher number of averaged images does not always guarantee better detection performance. Clearly it may be possible to reduce the error further by trying a finer range of averaging levels but nevertheless the principle is clear and validated.

IV. CONCLUSIONS

This paper introduces a new idea for improving the positioning accuracy of the vision system based on the calibration information. The concept of the new idea is to utilise this information in the selection of the proper camera location, where the smallest detection error occurred. Also, a new mathematical formula was introduced for estimating the appropriate camera location, the validity of the proposed formula was examined with two different motion scenarios. Moreover, the proper optical filter can also be selected based on its influence on the accuracy of the calibration, the results showed significant enhancements in the detection accuracy of the vision system after applying the suitable averaging filter.

ACKNOWLEDGMENT

The authors gratefully acknowledge the UK's Engineering and Physical Sciences Research Council (EPSRC) funding of the EPSRC Centre for Innovative Manufacturing in Advanced Metrology (Grant Ref: EP/I033424/1) and the Libyan culture attaché in London.

REFERENCES

[1] P. Maric and V. Djalic, Improving Accuracy and Flexibility of Industrial Robots Using Computer Vision: INTECH Open Access Publisher, 2012.

[2] P. Xu, G. Xu, X. Tang, and S. Yao, "A visual seam tracking system for robotic arc welding," The International Journal of Advanced Manufacturing Technology, vol. 37, pp. 70-75, 2008.

[3] B. Bailey and A. Wolf, "Real Time 3D motion tracking for interactive computer simulations," vol. 3, p. 3.1, 2007.

[4] Z. Zhang, "Camera calibration with one-dimensional objects," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 26, pp. 892-899, 2004.

[5] O. Faugeras, Three-dimensional computer vision: a geometric viewpoint: MIT press, 1993.

[6] P. F. Sturm and S. J. Maybank, "On plane-based camera calibration: A general algorithm, singularities, applications," in Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on., 1999.

[7] Z. Zhang, "A flexible new technique for camera calibration," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 22, pp. 1330-1334, 2000.

[8] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," Robotics and Automation, IEEE Journal of, vol. 3, pp. 323-344, 1987.

[9] S. Hutchinson, G. D. Hager, and P. I. Corke, "A tutorial on visual servo control," Robotics and Automation, IEEE Transactions on, vol. 12, pp. 651-670, 1996.

[10] J. Borg, M. Mehrandezh, R. G. Fenton, and B. Benhabib, "Navigation-guidance-based robotic interception of moving objects in industrial settings," Journal of Intelligent and Robotic Systems, vol. 33, pp. 1-23, 2002.

[11] G. C. Buttazzo, B. Allotta, and F. P. Fanizza, "Mousebuster: A robot for real-time catching," Control Systems, IEEE, vol. 14, pp. 49-56, 1994.

[12] M.-C. Chien and A.-C. Huang, "FAT-based adaptive visual servoing for robots with time varying uncertainties," pp. 3700-3705.

[13] H. H. Lund, E. de Ves Cuenca, and J. Hallam, A simple real-time mobile robot tracking system: Citeseer, 1996.

[14] H. Wang, Y.-H. Liu, and W. Chen, "Visual tracking of robots in uncalibrated environments," Mechatronics, vol. 22, pp. 390-397, 2012.

[15] V. Gengenbach, H.-H. Nagel, M. Tonko, and K. Schafer, "Automatic dismantling integrating optical flow into a machine vision-controlled robot system," in Robotics and Automation, 1996. Proceedings., 1996 IEEE International Conference on, 1996, pp. 1320-1325.

[16] J. Heikkila and O. Silvén, "A four-step camera calibration procedure with implicit image correction," in Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on, 1997, pp. 1106-1112.