

PAPER • OPEN ACCESS

Adaptive Feedback in Local Coordinates for Real-time Vision-Based Motion Control Over Long Distances

To cite this article: M M Aref *et al* 2018 *IOP Conf. Ser.: Mater. Sci. Eng.* **320** 012009

View the [article online](#) for updates and enhancements.



IES Ltd. develops the Virtual Environment (VE), the world-leading building simulation software which enables clients to design innovative buildings while minimising the impact on the environment. The VE is the only tool which allows designers to simulate the full performance of their design.

The successful candidate will join a team developing state-of-the art code for advanced building and district physics simulation. The team employs mathematical modelling techniques to analyse heat transfer mechanisms, air conditioning, renewable energy systems, natural ventilation, lighting, thermal comfort, energy consumption, carbon emissions and climate, and assess building performance against regulatory codes and standards in different countries.

careers@iesve.com

Adaptive Feedback in Local Coordinates for Real-time Vision-Based Motion Control Over Long Distances

M M Aref¹, P Astola², J Vihonen³, I Tabus², R Ghabcheloo¹,
and J Mattila¹

¹Laboratory of Automation and Hydraulics Engineering (AUT), Tampere University of Technology, FI-33101, Tampere, Finland

²Laboratory of Signal Processing (SGN), Tampere University of Technology, FI-33101, Tampere, Finland

³Cargotec Corporation, Tampere, Finland

Corresponding author: m.aref@ieee.org

Abstract. We studied the differences in noise-effects, depth-correlated behavior of sensors, and errors caused by mapping between coordinate systems in robotic applications of machine vision. In particular, the highly range-dependent noise densities for semi-unknown object detection were considered. An equation is proposed to adapt estimation rules to dramatic changes of noise over longer distances. This algorithm also benefits the smooth feedback of wheels to overcome variable latencies of visual perception feedback. Experimental evaluation of the integrated system is presented with/without the algorithm to highlight its effectiveness.

1. Introduction

In service robotics, the majority of tasks are based on robot intervention with known or unknown objects in indoor environments where global positioning systems are not accessible. Service robots benefit from using visual information to deal with objects their neighborhood. As reviewed in Walter *et al.* [1], state estimation and navigation for mobile platforms are commonly done using either global or local coordinates. Generally speaking, measurements using global coordinates (e.g. absolute position) are coarse and low frequency. Measurements using local coordinates, in which wheel odometry is available, are known to be smoother, fast and with a high resolution. In our case, the estimation from the environment is available in 30-40Hz while the dead-reckoning updates in 200Hz. To obtain smooth and globally valid feedback, fusion of these two types of measurements is a necessity but it is challenging and prone to drift [2], especially in heading estimation, which can affect coordinate mappings [3]. Some methods are based on stop, detect, look ahead and move [4]. However, our aim is to provide the object pose as a real-time feedback without the need for frequent replanning. Research that concerns itself with real-time detection [5] pays more attention to the precision of the image and less attention to the actual coordinates of the robot. For manipulation tasks, the repeatability of detection plays a vital role because the system might have physical interactions with the environment. Moreover, a path-follower controller with exact solutions needs a smooth and low-latency localization.

The object detection (ODT) results in this paper are comparable to Ha *et al.* [6]. However, in contrast with their conclusion, our method is capable of reaching smoother and more precise positioning of the robot with respect to the object without involving any image-based visual servoing in longer distances.



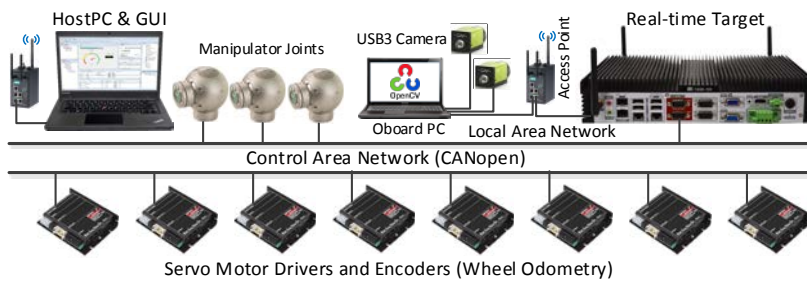


Figure 1. Control system's hardware architecture.



Figure 2. iMoro: The test case.

Long distance detections are important for non-holonomic mobile bases and autonomous vehicles because they need some space for steering and positioning before approaching the target area regardless of whether they are dealing with known markers or semi-unknown geometrical primitives.

1.1 Problem Definition and Architecture

The main goal of this research is to position a non-holonomic mobile manipulator in a location defined by target object(s). Each target object is recognizable by its longitude axis and geometric centroid, which builds a tensor of position/orientation ${}^C T_M$ with respect to the camera frame. Since the result of ODT will be used as feedback for positioning tasks, it must be precise, smooth, well-behaved and available in real-time not only in terms of the image coordinates, but also in terms of the robot's coordinates and object's coordinates. This paper is a generalization of Aref *et al.* [7] for markerless detections.

Figure 1 illustrates the hardware architecture for the experiments and Figure 2 shows the test case robot with position gauge marker. Compared to Aref *et al.* [7], we use the ODT described in Sec.2.1 instead of ROS and ALVAR. This method benefits from the use of a stereo rig with two JAI-GO5000 USB3.0 cameras with a 0.16 meter baseline. Image grabbing and processing are done in C++ and OpenCV 2.4.9 developing environments. The sensor dimensions of the cameras are 2560x2048 pixels from which a cropped centered view of 1920x1080 pixels is used. Cameras operate at 35FPS, which was deemed as the shortest usable exposure time under the illumination conditions. These modules run on a Lenovo laptop with a Core i7 CPU. The latest results are transmitted to the *Real-Time Target* (RTT) on a LAN network. The algorithms described in Sec.2 are implemented on RTT.

2. Position Estimation for Manipulation in a Local Coordinate

The main aim of our project is to discover the potential of stereoscopic imaging for remote handling. The setup of stereo cameras was done according to the topology design, which takes view angle, stereoscopic precision, and lighting into account. Figure 3 describes the processing stages of the 3D Node module, i.e., the eye-in-hand vision system.

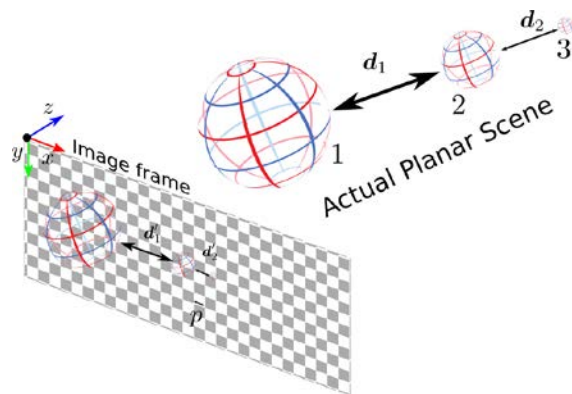


Figure 3. Estimation of distances on image frame can be comparable to pixel sizes if the objects appear in long distances.

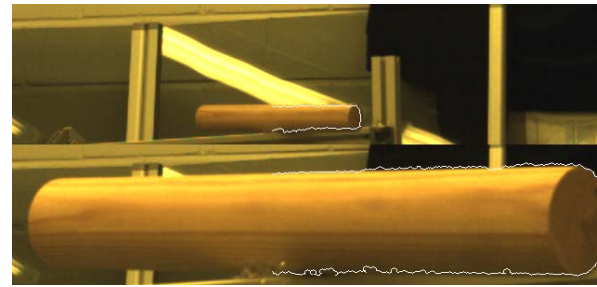


Figure 4. Detection sample from left camera with object at 3.5m and 1m. The detected object contour is illustrated partially with a white line.

2.1 Object Detection (ODT)

ODT follows that of Astola *et al.* [8] where a stereo camera system was used for object detection in 30-40Hz and pose estimation. In their research, object detection was based on color image thresholding in the hue channel and subsequent contour estimations. Object pose estimation performs stereo matching on the contour to derive the pose of the object in 3D coordinates. In Astola *et al.* [8], two versions of contour processing were studied – one using the convex hull and the other using the full contour. While the convex hull version performed faster computationally, the use of full contours with an additional filtering step provided superior results.

The output of ODT is the object pose tensor ${}^C\mathbf{T}_M$, which is constructed by the pose estimation algorithm using the centroid and orientation of the detected object. In Astola *et al.* [8], the immediate goal of the ODT was to maintain high enough rate of detection and pose estimation, as the sensor fusion prefers high frequency outputs from ODT. As a consequence, virtually no filtering or image processing is done jointly on the acquired frames and pose estimates. This results in a fair amount of high frequency pose estimation noise originating from ODT. The accuracy of the pose estimation is heavily range dependent, as the object's dimensions in the image plane diminish as a function of distance. Additionally, for a fixed stereo baseline, the system's ability to accurately estimate object depth also decreases with distance. The distance-based non-linearities in pose estimation, especially in the heading, are accommodated in a novel sensor fusion approach.

2.2 Experimental Results on Open-loop

To compare open-loop performance we used a controller independent of the vision feedback, which smoothly drives the robot in a straight line. It accelerates in the beginning and then maintains a constant velocity. As shown in Figure 4, if both systems are given the same input data, the M-Estimator is capable of generating smoother feedback. The acceleration moment of the robot is magnified on the graph to highlight the fast and smooth reaction of the filter. As illustrated, while the M-Estimator has faster motion feedback, it remains smoother across longer distances.

2.3 Integration in feedback

Exploiting the results of ODT in a closed-loop control system requires smoothness, availability and consistency of data at each sample time. Initially, an ODT module does not have such characteristics. Therefore, in this section we propose an algorithm that can integrate the ODT result in a real-time controller loop. To highlight the differences, we compare the following subsystems at each step:

- Variation of Object Detection Algorithms: Ver.1 and 2 in Sec.2.1, Ver.2 in Sec.2.1, a marker-based AR tag detection method [3].
- Modifications of M-Estimator: Gain adaptation to measurement changes in Sec.2.4, and gain adaptation to depth in Sec.2.5.

2.4 M-Estimation for Real-time Feedback

For each component of cT_M , such as y , while we are mapping all the vectors into the object's local coordinate $\{W\}$, we also apply the following estimation equations to estimate z :

$$\frac{\partial}{\partial t} \begin{bmatrix} \hat{z} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{z} \\ \hat{b} \end{bmatrix} + w \begin{bmatrix} k_P \\ k_I \end{bmatrix} (y - \hat{z}) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} x \quad (1)$$

$$w = \exp\left(-\left(\frac{\|\hat{y} - z\|}{c}\right)^2\right) \quad (2)$$

where x is the corresponding velocity component from the interoceptive sensor, such as wheel odometry [7]. Moreover, w accommodates the varying noise statistics by introducing the so-called robust Welsch function. Although this filter is capable of generating a smooth and robust feedback for marker-based ODT, for marker-less ODT the parasitic disturbances are heavily range-dependent (as shown in Figures 4 and 5) and should be accommodated accordingly.

2.5 Adaptive Filtering

As shown in Figures 4 and 5, the method described in 2.4 is capable of rejecting noises and disturbances initially. However, across longer distances, the object's image size becomes too small (e.g. Figure 4) and detection characteristics are different, especially for the orientations. For the scene illustrated in Figure 3, assume that triangulation between the image frame and actual object places in a pin-hole camera model is supposed to be valid for objects 1, 2 and 3. Then, the ratio of d'_1/z_1 linearly affects the tangent of the object orientation. Practically, this deterministic calculation is not valid for object places such as 3, where the distance d_2 is comparable to real pixel wideness, p . Quantization errors simply can be equal to or bigger than p and signal to noise ratio is considerable. This

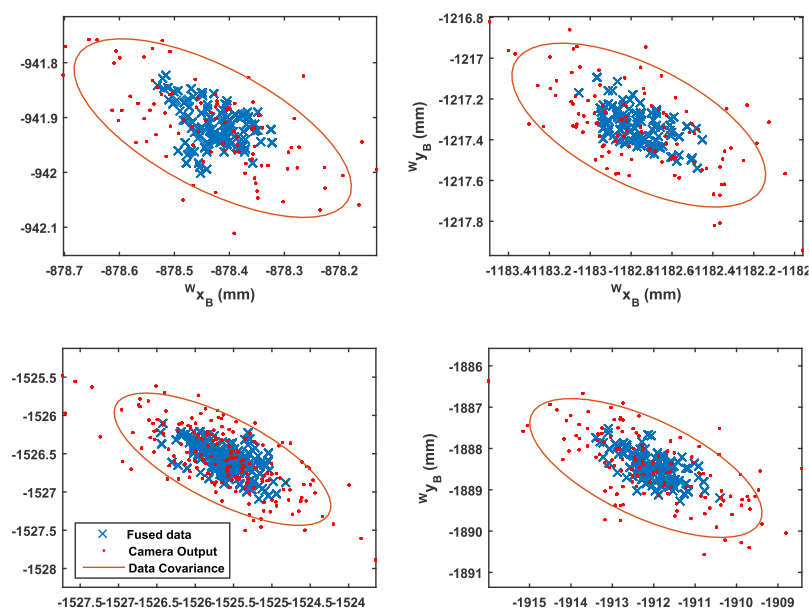


Figure 5. Raw and filtered output samples for marker-based detection in different distances. The numbers represent estimated position of the robot in target's coordinate frame.

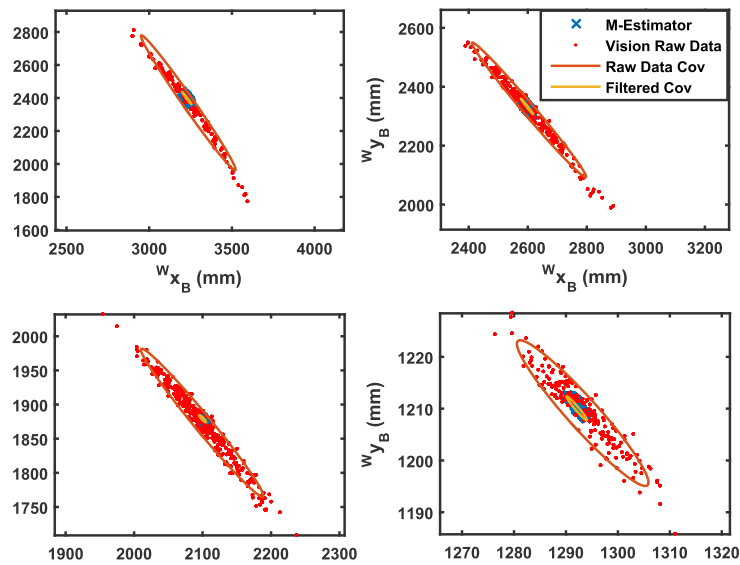


Figure 6. Results of similar tests to Figure 4 but for marker-less detection. Centroids of target objects rely on the origin of the coordinate system.

condition becomes worse if object 3 is placed too far from the adjusted focus range of the camera's optical system. This is the place where the pin-hole model is not necessarily correct, and d'_1/d_1 and d'_2/d_2 ratios do not exactly obey triangulation rules. Therefore, we consider correlation between the depth and accuracy of all estimated values.

As shown in Figure 5, noise amplitude in the first quarter of the path is comparable to the outliers of the last quarter. These changes result in larger values of $\|\hat{y} - z\|$ in (2). Therefore, w gain of (1), which is supposed to prevent integration of unexpected measurements in longer distances, prevents convergence to the final value, as shown in Figure 6. and the left graph of Figure 7. From the lessons we learned during the experimental studies on the system, the Welsch weight should be adaptive to the camera-object distance,

$$c = k_{w1} \exp(k_{w2}(\hat{d} - d_{min})) + k_{w3} \quad (3)$$

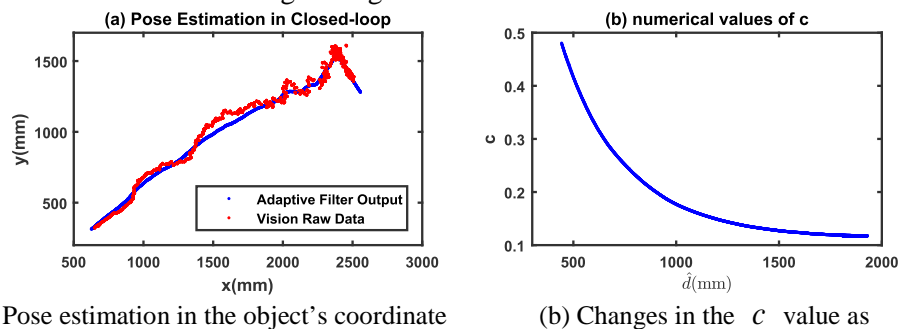
where \hat{d} is the estimated distance to the object and d_{min} is the minimum detection distance where we want value c to converge to $k_{w1} + k_{w3}$. Usually, this distance is the grasping point. $k_{w1}k_{w2}$ adjusts the rate of sensitivity of w to the amplitude of noises.

According to the experiments, for Ver.1, there is also a significant change of the mean values while approaching the object. Thus, we can conclude that Ver.1 is not useful for the closed-loop control loop as it adds significant non-linear behavior to the position estimation in local coordinates. So, (1) is suitable for the ODTs similar to Ver.2 where the quantization errors do have a dramatic effect on the convergence rate. For these ODTs, substitution of (1) in (2) compensates the dependency of noise dynamics on the distance. As shown in Figure 7b, the adaptive M-Estimator is capable of following the camera output even after a considerable amount of outliers. These methods provide the robot with real-time feedback of position suitable for closed-loop position-based visual servoing as shown in Figure 8 by adaptation of estimation weights. Some of the tests are also demonstrated on the YouTube video¹.

¹ Object tracking on the go, by iMoro <https://www.youtube.com/watch?v=m4vsr1rvxUg>

3. Conclusions

In this paper, by combining our previous experience on machine vision application in non-holonomic steerable platforms for mobile manipulation, we modified the widely-applied principles of machine vision, and complementary and Kalman filtering, for pose estimation. On one hand, we demonstrated how modifications in ODT can reduce non-linearity of detection for unknown objects. On the other hand, by proposing an adaptive estimation method, we achieved nearly the same quality of mobile robot positioning compared to marker-based ODT in distances that had been considered too far for position-based motion control. It is important to note that long range ODT and position estimation is essential for non-holonomic robots to keep their maneuvers smooth without stopping and replanning. The results are experimentally demonstrated on a video for detection, estimation and positioning on the go. Extension of this work for higher degrees of freedom is a future work.



(a) Pose estimation in the object's coordinate

(b) Changes in the c value as

Figure 7. Path and fusion behavior is adapted by distance from the target because values are functions of range in (3) and the robot adjusts the weighting in (2) while approaching to the target.

4. References

- [1] Walter M R *et al.* 2015 A situationally aware voice-commandable robotic forklift working alongside people in unstructured outdoor environments *J. of Field Robotics* 32 pp 590–628
- [2] Teller S *et al.* 2010 A voice-commandable robotic forklift working alongside humans in minimally-prepared outdoor environments *IEEE Int. Conf. on Robotics and Automation* pp 526–33
- [3] Wang Y and Wang S 2017 A new directional-intent recognition method for walking training using an omnidirectional robot *J. of Intelligent & Robotic Systems* pp 1–16
- [4] Wu J, Yue X and Li W 2016 Integration of hardware and software designs for object grasping and transportation by a mobile robot with navigation guidance via a unique bearing-alignment mechanism *IEEE/ASME Trans. on Mechatronics* 21 pp 576–83
- [5] Jung B and Sukhatme G S 2010 Real-time motion tracking from a mobile robot *Int. J. of Social Robotics* 2 pp 63–78
- [6] Ha H U, Yoon H N, Kim Y K, Lee D H and Lee J M 2015 Object tracking of the mobile robot using the stereo camera *Int. Conf. on Intelligent Robotics and Applications* pp 322–330
- [7] Aref M M, Vihonen J, Ghabcheloo R and Mattila J 2018 On latencies and noise effects in vision-based control of mobile robots *Advances in Service and Industrial Robotics: Mechanism and Machine Science* 49 pp 191–9
- [8] Astola P, Aref M M, Vihonen J, Mattila J and Tabus I 2017 Object detection in robotic applications for real-time localization using semi-unknown objects *40th Int. Conf. on Telecommunications and Signal Processing*

5. Acknowledgements

This work is supported by the Academy of Finland under the grant numbers 286260 and 283171.