A Centralized Omnidirectional Multi-Camera System with Peripherally-Guided Active Vision and Depth Perception

Nicholas D. Jankovic and Michael D. Naish

Dept. of Mechanical and Materials Engineering
The University of Western Ontario
London, Ontario, Canada

ndjankov@gmail.com, naish@eng.uwo.ca

***Abstract* - The growing popularity of omnidirectional vision technology has spawned numerous multi-camera designs that integrate various different camera types. This paper presents an omnidirectional vision system that combines a catadioptric camera, a fisheye camera and an active perspective camera. Aligning these cameras vertically provides a number of beneficial features, such as allowing simple peripherally-guided active vision, depth perception and a near spherical composite omnidirectional field of view. By having the active camera rotate around the outer perimeter, it can attain complete spherical access to the environment. The triangulation performance is evaluated experimentally using a target fixed to a long translation stage. Static positions of the target are estimated using a stereo pair that consists of one active perspective camera and one omnidirectional camera. Overall, the system provides sufficient accuracy to facilitate further surveillance research.**

Index Terms* - *Omnidirectional, modular, surveillance, active vision and multi-camera system.

I. Introduction

Omnidirectional vision technology has evolved rapidly in recent years, allowing it to find many new applications in the fields of surveillance and robot vision. The key advantage of omnidirectional cameras is their inherently large field of view (FOV), which can provide machines with greatly improved target detection or motion estimation abilities. However, in some situations, this large omnidirectional FOV may also become a disadvantage; for instance, when a detailed view of a target or obstacle is required. One viable solution is to integrate omnidirectional and perspective cameras into a multi-camera system, taking advantage of their complementary features. Multiple cameras can allow a machine to see with more human-like perception by inferring depth, while simultaneously acquiring detailed views and peripheral views.

Although many of existing multi-camera systems employ distributed solutions [1–3], this paper describes a centralized implementation that includes both omnidirectional and active perspective cameras. Several centralized configurations have also been reported in the literature. The system developed by Barreto and Araujo utilizes three horizontally aligned cameras, where one active pan/tilt camera is located on each side of a catadioptric camera [4]. Alternatively, the cameras can also be aligned vertically. Scotti *et al.* [5] present a surveillance system where a pan-tilt-zoom (PTZ) camera is situated directly below a catadioptric camera. Similarly, Wilhelm *et al.* [6] placed a binocular stereo head directly above a catadioptric camera. The latter two examples exhibit the greatest similarity to the presented system.

The objective of this paper is to present a unique omnidirectional multi-camera system and demonstrate its target triangulation capabilities. Section II provides a general description of the system, along with its specifications, while Section III outlines the experimental setup and the procedure used to evaluate the system's performance. Section IV describes the geometric model used to compute target depth and theoretical position errors. The results are then discussed in Section V, followed by the conclusions and future work in Section VI.

II. System Description

Even though similar multi-camera systems exist, there are a number of features that distinguish the system presented herein. Unlike the system developed by Barreto and Araujo [4], this system relies on a vertically aligned design to improve visibility. Unnecessary obstructions within an omnidirectional FOV are avoided by situating the active cameras either above or below other cameras. Aligning cameras in this manner provides each camera with complete 360° horizontal access to the environment. This stackable design is readily extendable to a modular framework, permitting many different system configurations to be created using only a few simple camera types [7]. In contrast, the designs presented in [5, 6] are intended for specific applications and do not openly address any potential for reconfiguration. The most notable benefit of a vertically aligned configuration is that it greatly simplifies peripherally guided active vision. Any active camera can pan directly to the target's azimuth angle relative to an omnidirectional image, without requiring any coordinate transformations. The centralized nature of the system limits the vertical separation between active and omnidirectional cameras, which reduces the necessary search space while scanning for the target. Specifically, the distance between a target's position in the omnidirectional image and the omnidirectional image centre can be used to guide the active camera's tilt axis to the approximate viewing angle of the target.

The presented system is composed of three different cameras. Two omnidirectional camera types are employed, where one is a fisheye camera and the other is a catadioptric camera using a hyperbolic mirror. In this implementation, the

fisheye camera is pointed upwards to view the upper 183° "hemisphere" of the environment, whereas the catadioptric camera views the lower 208° "hemisphere." This effectively provides the system with a spherical peripheral FOV, which greatly reduces the number of possible "blind spot" areas where targets can hide. The active camera is designed in such a way that it can pan continuously (via a slip ring) around the perimeter of the system. It can also tilt vertically through a 180° range, providing it with spherical viewing access. This camera is located between the two other cameras so that unbiased triangulation estimations can be made using either of the two omnidirectional cameras. Fig. 1 shows an image of the prototype system and Fig. 2 illustrates the field of view for each camera.

Each colour camera operates at a resolution of 640×480 pixels and connects to a standard PC using a Firewire (IEEE 1394a) serial bus. The active camera axes are driven using stepper motors, which are controlled using an industrial PCI motion controller. The active camera is able to accelerate at rates up to 12.5 rev/s^2, reaching a top speed of 0.8 rev/s. This enables the system to potentially acquire a new target in 0.7 seconds when the new target is located on the opposite side of the system (i.e., 180° rotation relative to the current position).

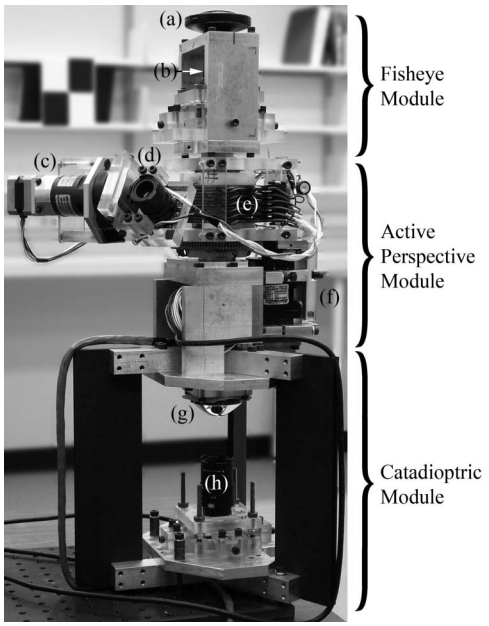


Fig. 1 Constructed prototype. a) Fisheye lens, b) 4mm camera, c) tilt motor system, d) 12mm camera, e) slip ring, f) pan motor system, g) hyperbolic mirror, h) 3.6mm camera.

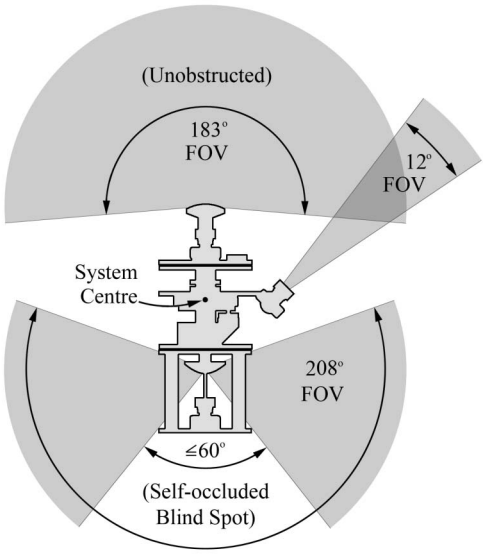


Fig. 2 Camera layout with accompanying fields of view.

III. Experimental Setup

For initial testing purposes, a simple algorithm has been developed to automatically track a single target. When a target is being monitored by one of the omnidirectional cameras, its motion within the image is modelled using a Kalman filter [8]. The predictions obtained from this filter are used to control the active camera motion, allowing quick interception of a moving target. Once the target appears within the active camera's FOV, a PID controller is used to keep the camera centred on the target.

Only a simple target is necessary for the purpose of this evaluation. Therefore, an orange ball that is approximately 20 cm in diameter is used. This target is large enough to be seen within the low-detail images of the omnidirectional cameras. Applying basic colour thresholding provides a straightforward means of detecting the target. Its centre in the image is estimated as the centroid of the orange blob of pixels.

In order to effectively evaluate the system's triangulation ability, the target's position must be precisely known. Fig. 3 shows a simple belt-driven translation stage, which has been constructed to provide acceptable target positioning. This stage has a maximum travel of 2.34 m and is driven by a stepper motor under open loop operation. The target can be positioned within 5 mm accuracy in the travel direction and within 1 mm in the orthogonal directions. Currently, the maximum travel velocity is 0.167 m/s. Placement of the translation stage relative to the system is accurate to within 10 mm in the horizontal plane and within 2 mm vertically. Fig. 4 schematically shows the system's position in relation to the translation stage, where the origin of the system's frame of reference is located at the centre of the active camera. The positioning accuracy of the target is sufficient for this experiment since it is several times greater than the system's triangulation accuracy.

IV. Triangulation Geometry

Employing a vertical design greatly simplifies the layout between the cameras by making them axisymmetric about the central axis. By enforcing a single effective viewpoint constraint, then each camera's position can be defined using basic geometry. The active camera is constrained to a circular path that is perpendicular to the central axis, where its location is defined by a vertical position (zero in the case of this system)

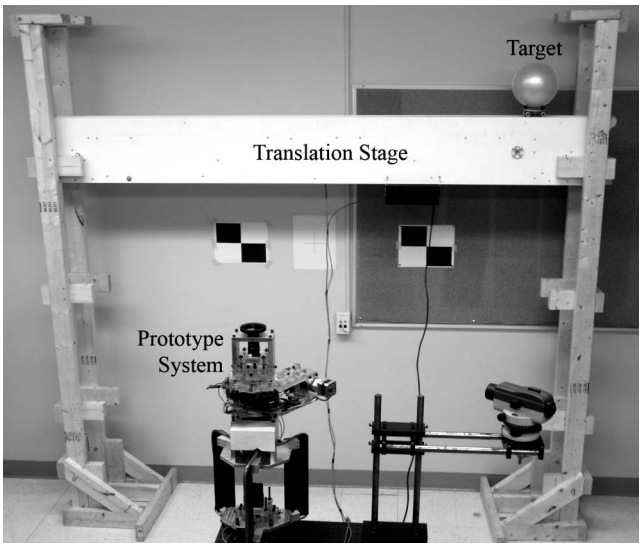


Fig. 3 Experimental setup.

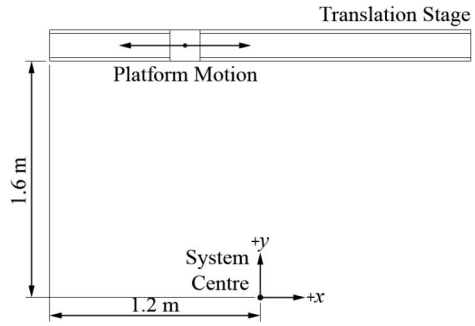


Fig. 4 Top view layout between the system and translation stage.

and a radial offset. A ray passing through the target can be defined by connecting a line between the centre of the target's image and the camera's centre of projection. When any two cameras have acquired the target, an estimate of the target's 3D position is obtained by computing the intersection between the two projected rays. Only the vertical angle of each ray and the relative camera geometry are necessary to determine the target's distance from the system. However, measurement errors usually prevent the projected rays from intersecting at the target's exact location in space. Fig. 5 illustrates the triangulation geometry, where an active and fisheye camera pair is used as an example.

When conducting triangulation, usually the target's depth is the quantity of greatest interest. The depth estimate of the target point can be determined as a function of the ray angles β and θ for each active and omnidirectional camera, respectively. The positive error is defined as the difference between the maximum and actual depth estimation, where the negative error relates to the minimum depth estimation in a similar manner. Ray angle variations are expressed as the angular uncertainties, $\delta\beta$ and $\delta\theta$. Uncertainties also apply to the relative camera positions (i.e., δZ and δR); however, these values are typically in the order of millimetres and have a negligible effect on the estimated error when the target distance is in the order of metres. As a result, the position uncertainty values can be excluded from consideration.

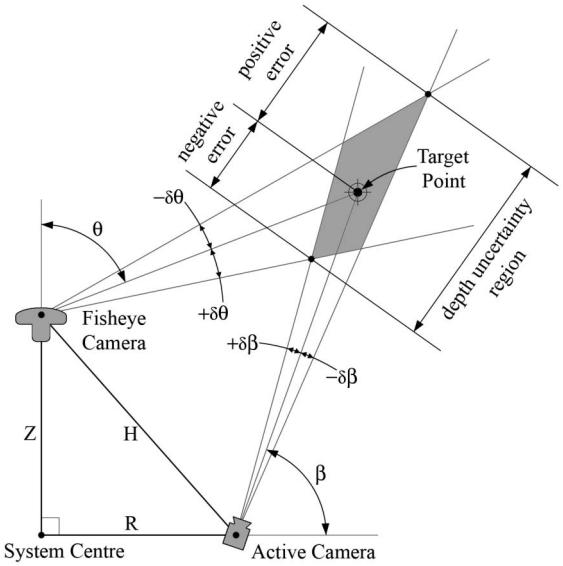


Fig. 5 Triangulation geometry with uncertainty.

In the case where angle θ is less than angle β, then (1) describes the distance D between the system centre and the target point using the cosine law

$$D = \sqrt{R^2 + L^2 - 2RL\cos(\pi - \beta)}\,, \tag{1}$$

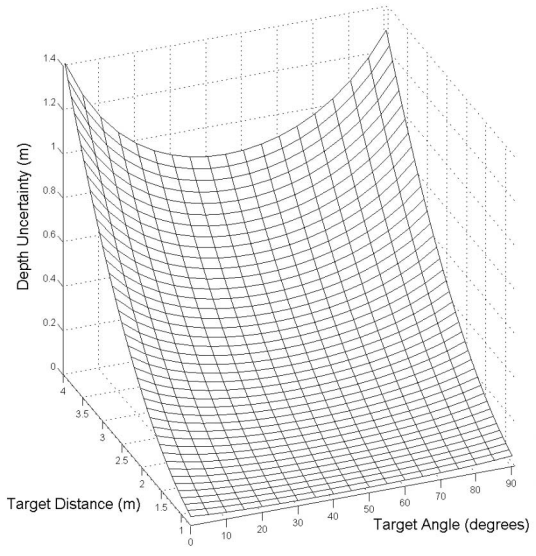
where L represents the distance between the active camera and the target point (2):

$$L = \frac{H\sin(\pi - \theta - \arctan(R/Z))}{\sin(\beta + \theta - \pi/2)}\,. \tag{2}$$

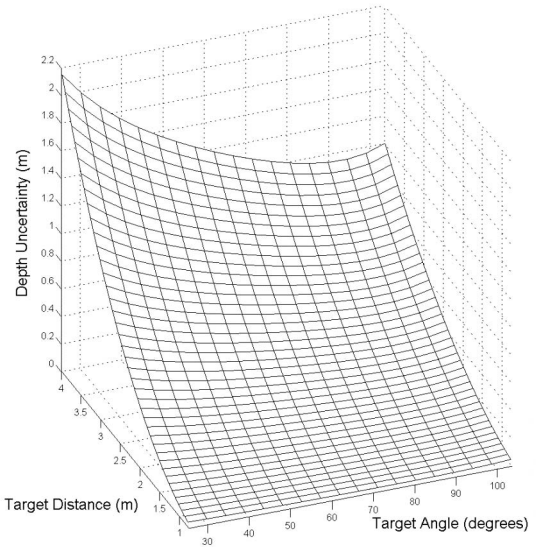
Similar equations also exist for the second case, where angle θ is greater than angle β. Two cases are necessary because the geometry collapses when β approaches zero in the first case and also when θ approaches zero in the second case. In order to compute the minimum possible target distance, θ in (1) and (2) is replaced with $\theta + \delta\theta$ and β is replaced with $\beta + \delta\beta$. Similarly, the maximum possible target distance is computed by replacing θ with $\theta - \delta\theta$ and β with $\beta - \delta\beta$. The overall depth uncertainty is the difference between the minimum and maximum distances.

The angular uncertainties $\delta\theta$ and $\delta\beta$ result from a nine-step calibration procedure [9] which aims to ensure the alignment of camera optics, find the intrinsic camera parameters, compute lens distortion parameters, determine the relative position and orientation of each camera and assess the accuracy and repeatability of the motion system used for the active module. The angular uncertainties are assumed to be the sum of uncertainties that contribute to the vertical angle. For the active camera, $\delta\beta$ is the combination of the perspective ray projection uncertainty and the tilt axis uncertainty; whereas, $\delta\theta$ is composed of the omnidirectional ray projection uncertainty and the calibrated orientation uncertainty.

The positive and negative errors are then used to define the theoretical upper and lower error bounds on the estimated distance of the target relative to the system. Fig. 6 (a) and (b) shows how the depth uncertainty region changes over each camera pair's working range. They also demonstrate how the estimation error is highly dependent on the target's position. The left and right horizontal axes, respectively, correspond to the target's distance from the system and the target's angle with the vertical system axis, where the vertical axis represents the size of the uncertainty region. Even though the depth uncertainty appears to be rather large, note that it is composed of both the positive and negative theoretical errors. These are actually the true estimates of the triangulation accuracy. The depth uncertainty region is used here to simplify presentation with a single surface, rather than showing two overlapping surfaces.



(a)



(b)

Fig. 6 Depth uncertainty for (a) active-fisheye camera pair and (b) active-catadioptric camera pair.

V. Results and Discussion

Two sets of static triangulation data have been collected, one for each camera pair. Fig. 7 (a) and (b) show the estimated position as determined by the active-fisheye and active-catadioptric camera pairs, respectively. In each experiment, multiple measurements are acquired, but only the mean values are shown. The error bars indicate the range of sampled data at each position. Also shown in Fig. 7 (a) and (b) are the actual paths taken by target as it moves from left to right. In order for each omnidirectional camera to have a clear view of the target, it is elevated 0.8 m above the system centre for the fisheye camera and positioned 0.4 m below the system centre for the catadioptric camera.

An alternative representation of the data is provided in Fig. 8 (a) and (b) for the active-fisheye and active-catadioptric camera pairs, respectively. In this case, the errors between the actual and estimated target positions are plotted against the target's depth relative to the system, permitting direct comparison with the theoretical error bounds discussed in Section IV. Error bars are included, which also correspond to the sampled range. For reference, the error bounds are computed at an angle of 75° for the active-fisheye pair and 85° for the active-catadioptric pair, so that comparison can be made with the uncertainty regions shown in Fig. 6 (a) and (b).

Observations made from the information presented in these two figures indicate that the active-catadioptric pair provides much more accurate position estimates than the active-fisheye pair. Both sets of data show some degree of offset relative to the actual position (Fig. 7) and relative to the zero mean error (Fig. 8), although this phenomenon is much more pronounced for the active-fisheye camera pair. The most likely cause for this is a minor misalignment in the fisheye camera's orientation that has not been fully accounted for during calibration. Recalibrating the omnidirectional camera orientations should remedy the offset problem. The variations in the active-fisheye position measurements are also greater than those observed for the active-catadioptric measurements. Since the target is located 0.8 m above the system (at a height of 2.2 m from the floor), it operates in close proximity to the overhead lighting. The resulting uneven and more intense lighting conditions negatively impact estimation of high target positions, observed by the active-fisheye pair. Specifically, the target's specular surface causes some problems when computing its centroid because simple thresholding is not able to detect the complete target, resulting in an offset from the true centroid. Employing a more robust target detection method and using an alternate target will certainly reduce variations in the position estimates.

Even though dynamic target tracking is not the primary focus of this paper, some preliminary testing has been undertaken using a basic PID controller. The active camera is able to track the target along the translation stage while operating at the rated 0.167 m/s velocity. In fact, informal testing shows that the active camera can still track a target that is traveling at 0.37 m/s (assuming that it is also situated at a distance of 1.6 m). Mechanically, the active camera can move

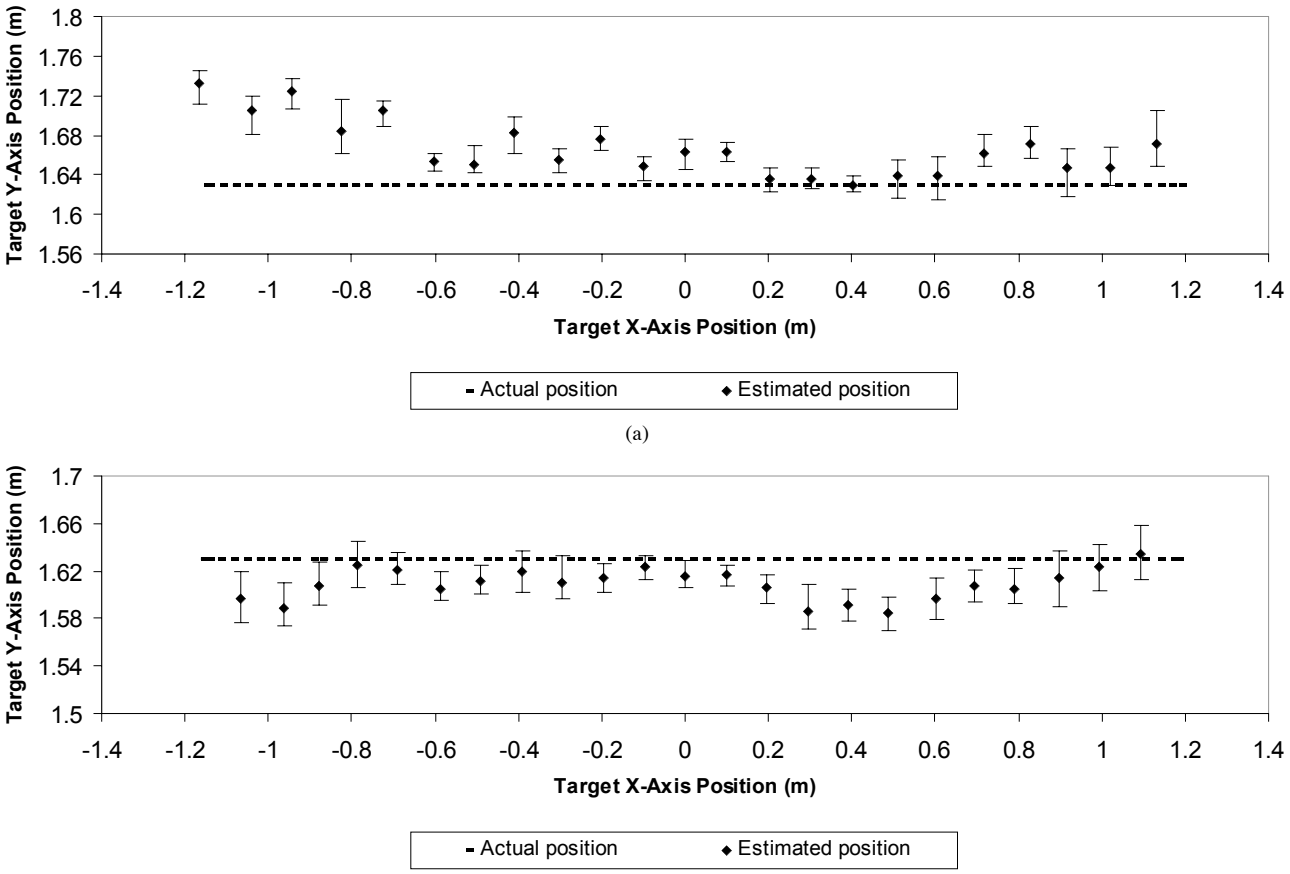


Fig. 7 Top view of the dynamic and static triangulation data. The target is located at a height of (a) 0.8 m above the system for the active-fisheye camera pair and (b) 0.4 m below the system for the active-catadioptric camera pair. Error bars represent the range of sampled data at each position.

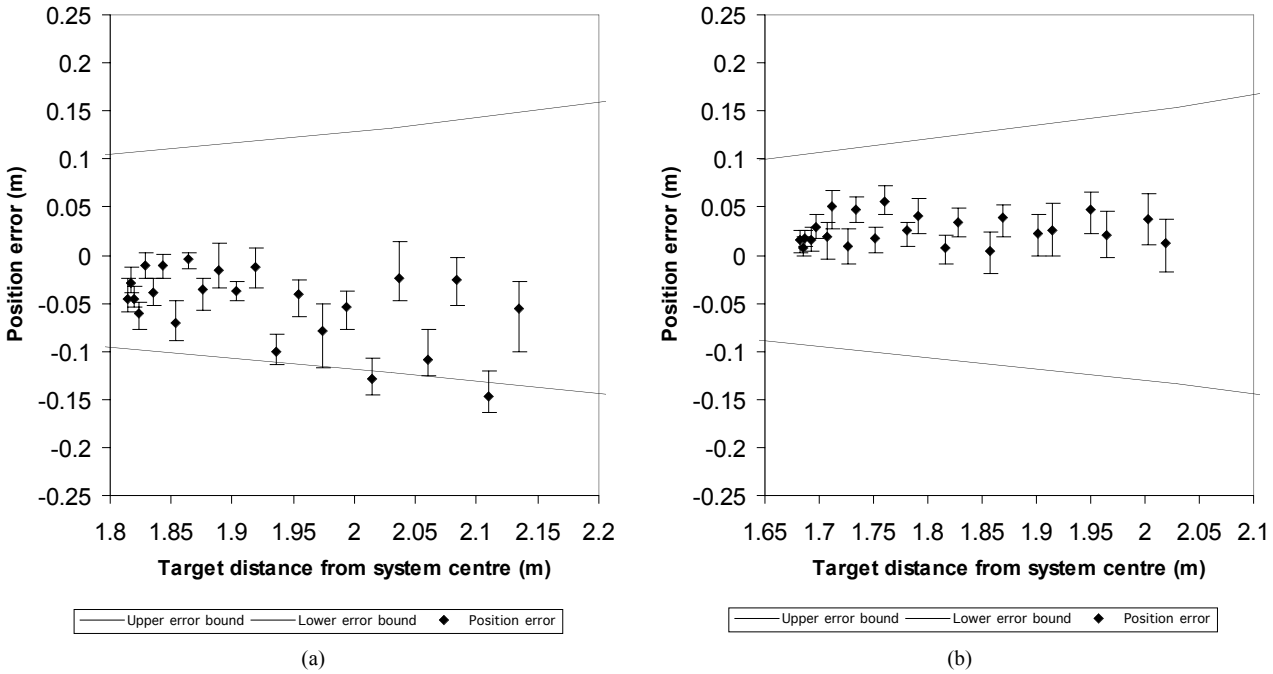


Fig. 8 Experimental triangulation errors relative to the calibrated uncertainty bounds. The mean error value from (a) the active-fisheye camera pair, across all of the samples, is -0.051 m; whereas its standard deviation is 0.041 m. Similarly, (b) the active-catadioptric camera pair has a mean value of 0.026 m with a standard deviation of 0.017 m. Error bars represent the range of sampled data at each position.

much faster, however the tracking speed is currently limited by the rate at which images can be processed, at about five frames per second. Dynamic test data exhibit similar distributions to the static data shown in Fig. 8., with lower triangulation accuracy. The mean error value and standard deviation for the active-fisheye pair are 0.044 m and 0.050 m, respectively, whereas the mean error value and standard deviation for the active-catadioptric camera pair are 0.034 m and 0.031 m. The increase in standard deviation for the dynamic triangulation case is caused primarily by unsynchronized data acquisition from both cameras and encoders. To help alleviate this problem, future versions of the prototype will utilize cameras that can be synchronized via an external trigger.

VI. Conclusions and Future Work

This paper describes the motivation for creating a centralized omnidirectional multi-camera system and highlights the many advantages that its design can provide. The most notable feature is the vertically aligned configuration, which permits simple peripherally guided active vision while providing depth perception between dissimilar camera types. Stacking cameras vertically also allows a nearly spherical peripheral FOV to be composed, while providing the active cameras unrestricted spherical access to the environment.

The triangulation performance of the system is evaluated by positioning a target on a long translation stage. This enables ground truth to be recorded, allowing comparisons between actual and estimated positions. Static triangulation tests are conducted using both active-omnidirectional camera pairs. The active-catadioptric camera pair exhibits better performance than the active-fisheye pair. Reduced accuracy is largely attributed to minor misalignment of the omnidirectional cameras, a specular target surface and uneven overhead lighting. Nevertheless, triangulation accuracy is still within acceptable limits at this stage of development. A number of improvements can be made to increase the system's triangulation accuracy, such as using a less specular target and employing a more robust target detection method.

Ultimately, software for tracking multiple targets will be implemented in order to proceed with more practical, real-world surveillance research. The goal of the next prototype is to construct truly modular and reconfigurable cameras and make them small enough to be mounted to a mobile robot platform, providing additional opportunities for investigation.

References

[1] M. Greiffenhagen, V. Ramesh, D. Comaniciu and H. Niemann, "Statistical modeling and performance characterization of a real-time dual camera surveillance system," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 335-342, 2000.

[2] D.R. Karuppia, Z. Zhu, P. Shenoy and E.M. Riseman, "A fault-tolerant distributed vision system architecture for object tracking in a smart room," in *Proceedings of the 2nd International Workshop on Computer Vision Systems*, Springer-Verlag: Berlin, pp. 201-219, 2001.

[3] M. Doi and Y. Aoki, "Real-time video surveillance system using omni-directional image sensor and controllable camera," in *Proceedings of the SPIE - Real-Time Imaging VII*, vol. 5012, pp. 1-9, 2003.

[4] J. Barreto and H. Araujo, "A general framework for the selection of world coordinate systems in perspective and catadioptric imaging applications," *International Journal of Computer Vision*, Kluwer: Netherlands, vol. 57, no. 1, pp. 23-47, 2004.

[5] G. Scotti, L. Marcenaro, C. Coelho, F. Selvaggi and C.S. Regazzoni, "Dual camera intelligent sensor for high definition 360 degrees surveillance," in *IEE Proceedings of Vision Image and Signal Process*, vol. 152, no. 2, pp. 250-257, 2005.

[6] T. Wilhelm, H.J. Bohme and H.M. Gross, "A multi-modal system for tracking and analyzing faces on a mobile robot," *Robotics and Autonomous Systems*, vol. 48, pp. 31-40, 2004.

[7] N.D. Jankovic and M.D. Naish, "Developing a modular active spherical vision system," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'05)*, Barcelona, Spain, pp. 1246–1251, April 2005.

[8] R. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, pp. 35–45, March 1960.

[9] N.D. Jankovic and M.D. Naish, "Calibrating an active omnidirectional vision system," in *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS'05)*, Edmonton, Canada, pp. 1355–1360, August 2005.