Proceedings of IMESE 2004: ASME International Mechanical Engineering Congress, Anaheim, CA, Nov. 13-19, 2004

ACTIVE-VISION FOR THE AUTONOMOUS SURVEILLANCE OF DYNAMIC, MULTI-OBJECT ENVIRONMENTS

Ardevan Bakhtari, Michael D. Naish,† and Beno Benhabib

Computer Integrated Manufacturing Laboratory
Dept. of Mechanical and Industrial Engineering
University of Toronto
5 King's College Rd., Toronto, Ontario, Canada M5S 3G8
[bakhtar, beno]@mie.utoronto.ca

†Dept. of Mechanical and Materials Engineering
University of Western Ontario
London, Ontario, Canada
naish@eng.uwo.ca

ABSTRACT
This paper presents a novel method for the coordinated selection and positioning of groups of active-vision cameras for the autonomous surveillance of an *object-of-interest* as it travels through a multi-object workspace with an *a priori* unknown trajectory. Different approaches have been previously proposed to address the problem of sensor selection and control. However, these have primarily relied on off-line planning methods and only infrequently utilized on-line planning to compensate for unexpected variations in a target's trajectory. The method proposed in this paper, on the other hand, uses a real-time dispatching algorithm, which eliminates the need for any *a priori* knowledge of the target's trajectory, and, thus, is robust to unexpected variations in the environment. Experiments have shown that the use of dynamic sensors along with a dispatching algorithm can tangibly improve the performance of an active-surveillance system.

Keywords: *Active vision, surveillance, dispatching, sensor fusion.*

INTRODUCTION
In dynamic multi-object environments, an active sensing-system may provide autonomous surveillance of an-object-of-interest as it moves through the workspace. The environment may be cluttered with static and/or mobile objects that are not of interest (i.e., obstacles), which may prevent the viewing of the target for periods of time. Surveillance is defined here as the data-acquisition and analysis process for the recognition and/or parameter estimation of targets (i.e., features) on objects-of-interest.

On-line planning can be used to dynamically reconfigure an active sensing-system, based on the current and estimated future states of the environment to improve the performance of the surveillance system [1]. Sensing-system planning requires the dynamic selection of (1) the number and type of sensors to be utilized for data acquisition, and (2) their optimal position and orientation (pose).

On-line sensing-system reconfiguration has been commonly used for the surveillance of single-object environments. The system proposed in [2] handles moving objects by discretizing time and computing new viewing configurations for each time interval while attempting to minimize changes in sensor position from one time to the next. The system in [3] determines 2-D camera layouts using an off-line optimization method, where deviations from the planned approach are handled on-line using heuristics to adjust camera actions and temporal camera switching points. The systems proposed in [4] and [5] present approaches to automatically generate camera poses and optical settings based on task requirements and imaging constraints such as feature delectability, field of view, and resolution. The system proposed in [6] uses high-level operation goals, geometric goals, and uncertainty-reduction goals to generate task plans for the sensors. The approach presented in [7] automatically controls the vision system through influence diagrams based on Bayesian tracking methods.

The problem of on-line sensor planning in dynamic multi-object environments has not received the same degree of attention as the simpler single-object case. The system presented in [8] considers each sensor individually for placement in a dynamic-target, static-obstacle environment. By modeling the object and the surrounding environments in 3D polyhedrons, the algorithm determines constraints such as occlusions at each discretized time instant based on *a priori* known target trajectory. The algorithm of [9] encodes the *a priori* workspace knowledge as a discrete probability density function of objects locations and, then, generates camera poses using heuristic methods in order to maximize the probability of detecting the target. Since these systems rely on *a priori* knowledge, they are not considered to be robust to variations in the environment.

In this work, the principles of dispatching used for the operation of service vehicles [10] are used for effective real-time sensing-system reconfiguration in a dynamic multi-object environment. Real-time dispatching requires selecting an optimal subset of dynamic sensors to be used in a data-fusion process and maneuvering them in response to the motion of the object, while the unused sensors are positioned in anticipation of future service requirements. The goal is to provide information of increased quality for the task at hand while ensuring adequate response to future object maneuvers.

This paper is an extension of our initial work reported in [11] for unoccluded single-object environments to cases with multiple static or dynamic obstacles. In [11], it was shown that the use of dynamic sensors over static ones can (i) significantly increase the viewing area through a reduced number of sensors and (ii) improve the performance of the surveillance system by decreasing uncertainty. Thus, in this paper, this comparison is not repeated in detail.

SENSOR DISPATCHING

In the context of object surveillance, sensor dispatching attempts to maximize the effectiveness of a sensing-system used to provide estimates of object parameters at predetermined times or positions along the object trajectory. Herein, it is assumed that the times at which the information is desired are fixed. These predetermined times are referred to as demand instants, t_j. The position of the object at a particular demand instant is a demand point, D_j. Without prior knowledge of the object trajectory, the demand point corresponding to each demand instant may be predicted from observations of the object motion. In general, the estimation of the demand point location changes (and its corresponding uncertainty diminishes) as the prediction accuracy improves over time; however, the demand instant remains constant.

If the sensing-system comprises multiple redundant sensors, a subset of these may be sufficient to satisfy the sensing requirements of a demand. Namely, a sensor-fusion process does not need to combine the information from all of the sensors in the system. Instead, a subset of sensors, herein referred to as a fusion subset, may be selected to survey the object at a particular demand instant, allowing other sensors to be configured in anticipation of future use.

In this context, the general sensor dispatching problem is stated as: Given a set of sensors and a set of time-varying demand points (based on the predicted motion of a maneuvering object at the corresponding set of demand instants), determine the subset of sensors (and their corresponding poses) that will sense the object at each demand instant while ensuring that the remaining sensors are adequately distributed throughout the workspace for possible future use.

To facilitate dispatching in real-time, a finite segment of the object trajectory (consisting of m demand instants) may be defined and the corresponding demand points estimated. This set of demands constitutes a rolling horizon that estimates the object motion a limited interval into the future. The period between two demand instants is the service interval—the amount of time available for planning and positioning the sensors before data acquisition must occur.

Dispatching is accomplished using two complementary strategies: A coordination strategy is used to specify which sensors will be used at each demand instant; and, a positioning strategy is used to determine the optimal pose of each sensor

COORDINATION STRATEGY

The coordination strategy consists of two complementary processes. An assignment method considers only the first demand instant, while the remaining demand instants on the rolling horizon are considered by a preassignment method.

ASSIGNMENT AND POSITIONING

Assignment considers the first demand point on the rolling horizon. The coordination strategy is implemented by performing a search that selects a subset of sensors (of size $\leq k$) to service the demand instant from the set of all sensors (total of n). In addition, the search method specifies a desired pose for each assigned sensor at the time of data acquisition, thereby implementing the positioning strategy.

The general approach to the assignment and positioning of sensors for a demand point can be summarized as follows:

1. Predict the target's pose, D_j, and the obstacles' poses, $O_{1j}, \ldots, O_{qj}$, at the demand instant, t_j.
2. For every sensor, S_i, $i = 1 \ldots n$:
   (a) Determine its best achievable unoccluded pose with respect to D_j, and
   (b) Assess the corresponding (single sensor) performance metric, visibility at t_j, v_s.
3. Rank all S_i according to the achievable visibility, v_s.
4. Assign the top k ranked sensors to t_j. Only unoccluded sensors are assigned to the demand instance.

PREASSIGNMENT AND PREPOSITIONING

The goal of preassignment and prepositioning is to position sensors in anticipation of future sensing requirements. Those sensors that have not been assigned to the most immediate demand instant (the first demand on the rolling horizon) may be preassigned to subsequent demands.

The approach used for preassignment and prepositioning is very similar to the approach of Assignment and Positioning method outlined in the previous section, however, there are some differences. First, while the suitability of each sensor is considered for each demand instant, only those that have not been previously assigned may be preassigned to the demand under consideration. Second, the preassignment algorithm iterates to consider additional demand instants until either all sensors have been preassigned or the end of the rolling horizon has been reached. If a sensor is not assigned by the end of the rolling horizon it is assigned to the first unoccluded demand instant. Sensors assigned to future demand instants, when possible, may also participate in the surveillance of the current demand point. This approach aims to provide the highest quality data by maximizing the effectiveness of the sensing system through utilization of all possible sensors.

POSITIONING STRATEGY

In a single-target, multi-object environment the positioning strategy performs sensor planning not only based on the trajectory of the object-of-interest (i.e., the target) but by considering also other objects that are not of interest but may act as occlusions. The first step in determining the best achievable pose is to determine the occluded regions in the workspace. To accomplish this, the pose of each object (target or obstacle) is predicted for all the demand instants. Next each object is modeled as a single geometric primitive (e.g. a cylinder in our work) rather than as a collection of 3D polyhedral in order to decrease computational complexity. Using the pose and size of each cylinder, representing an object, the algorithm determines the occluded regions of a sensor's workspace, Figure 1. Next, the algorithm determines the region of the workspace that the sensor can travel to before the object-of-interest reaches the demand instant, referred to herein as feasible region. This region is defined by the sensor's dynamic motion capabilities such as maximum velocity and acceleration. Lastly, the algorithm determines an optimal sensor pose that will yield maximum visibility, which is both feasible and unoccluded.

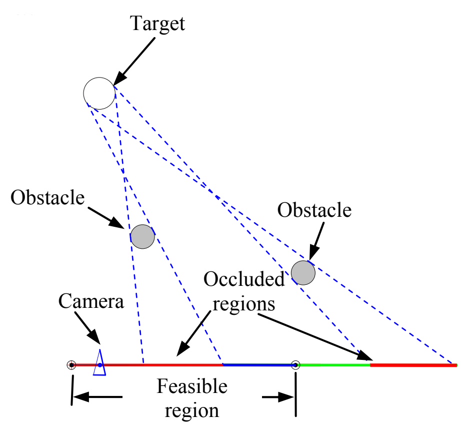


Figure 1. Occluded regions of a sensor's workspace,

IMPLEMENTATION

An experimental setup was devised to evaluate the performance of the proposed dispatching algorithm. This system uses four mobile cameras to determine the pose of a single target, represented by a circular marker, maneuvering through the workspace on a planar trajectory, Figure 2. Stationary obstacles positioned within the workspace at predefined locations act as obstacles. A stationary overhead camera is utilized to obtain estimates of the gross target motion and location of obstacles. These estimates are used to plan the motion of the cameras that have one or two degrees of freedom (dof) – all have one-dof rotational capability (pan), while two of the cameras can also translate linearly, (see Table 1 for component list).

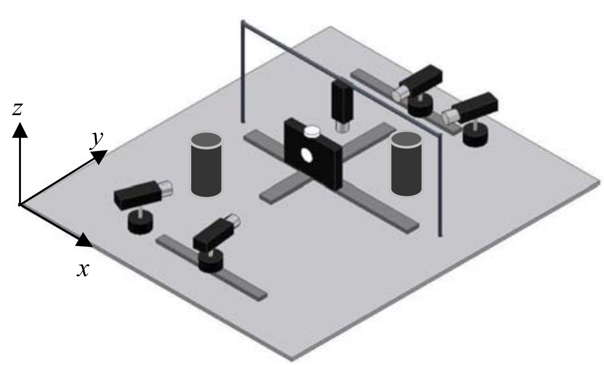


Figure 2. Experimental system layout.

Table 1. Hardware specifications.

Hardware	Characteristic
Linear Stages	*Range: 300 mm/500 mm* *Positional Accuracy: 18 μm* *Max velocity: 1.5 m/sec*
Rotary Stages	*Positional Accuracy: 12 arc sec* *Max velocity: 15 rev/sec*
Horizontal CMOS Cameras	*Resolution: 640x480 pixels*
Overhead CCD Camera	*Resolution: 640x480 pixels*

Overall, the surveillance system consists of four main modules, each with a distinct function: Prediction, dispatching, motion, and vision modules, which work in tandem to estimate the target's future location, reconfigure the active vision sensors, and optimally track the target.

***Prediction Module*:** The primary purpose of this module is to determine a rough estimate of the current position as well as a prediction of future positions of all objects in the workspace. The position of the target (a circular marker) is monitored continuously by the overhead (static) camera and the data fed to a recursive Kalman-Filter (KF) [12]. Data from the prediction module is used as input to the dispatching module.

Dispatching Module: This module has two primary functions: (i) selection of a subset of all cameras and their positioning in accordance to the current and future predicted target positions as received by the prediction module and (ii) directing the vision module to perform multi-camera imaging of the target location. The current camera locations (as received from the motion module) are also forwarded to the vision module to facilitate sensor fusion. Sensor positions are updated continuously at the frequency of the prediction module, about 15 Hz.

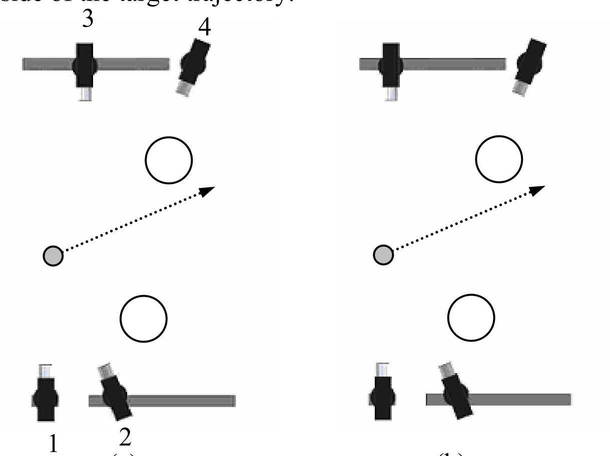
Motion Module: This module controls the motions of the cameras. The inputs to this module are the desired camera locations from the dispatching module, while the outputs are the motion commands (poses and velocities) to the linear- and rotary-stage controllers on which the cameras are mounted. This module also informs the dispatching module about the cameras' current locations and velocities.

***Vision Module**:* This module receives commands and information from the dispatching module to image the target and, subsequently, determine the target's pose in world coordinates. Each camera captures and processes images (of the circular marker) independently. The pose of the marker is estimated using an analytical solution developed earlier in our laboratory [13]. The data from cameras is, then, fused to decrease the uncertainty in the final estimate of target's pose.

AN EXPERIMENT

Numerous experiments were conducted in our laboratory, using the active vision set-up presented above, to illustrate the implementation of the proposed dispatching algorithm. These experiments served to evaluate the performance of the proposed surveillance system under changing objects trajectories and sensor dynamics. In general, the experiments verified that the performance of a surveillance system can be tangibly improved with the use of an effective dispatching strategy, primarily due to (i) increased robustness of the system (i.e., its ability to cope with *a priori* unknown target trajectories and presence of obstacles), (ii) decreased uncertainty associated with estimating the target's pose through sensor fusion, and (iii) increased reliability through sensory fault tolerance.

In the specific example discussed herein, the performances of two systems with varying motion capabilities (i.e. "fast" and "slow") are compared. In the fast system, the cameras are capable of 15 mm/s maximum velocity, 30 mm/s^2 acceleration/deceleration and 0.3 rad/s rotational velocity. In the slow system, the cameras have much more modest motion capabilities of 2.5 mm/s maximum velocity, 10mm/s^2 acceleration/ deceleration and 0.1 rad/s rotational velocity. The speed of the target in both systems is set to 5 mm/s, following the trajectory shown in Figure 3. The target is modeled as a 25 mm circle, while two 60 mm obstacles are positioned at each side of the target trajectory.



(a) (b)

Figure 3. Initial sensor poses for (a) slow system and (b) fast system.

PERFORMANCE EVALUATION

System evaluation was carried out using a visibility metric as previously proposed in [11]. Target visibility by a sensor is calculated using the expected variance in the measurements that is a function of the sensor's Euclidean distance to the target and its bearing (i.e., the angle the camera's local axis makes with the normal of the target's surface). The system performance was also evaluated by determining the errors in the real-time estimation of the target's pose. Absolute error in position estimation, $E_{position}$, is the Euclidean distance between the true target position, (x_t,y_t,z_t), and the system's estimate of the target's position, (x_e,y_e,z_e):

$$E_{position} = \sqrt{(x_e - x_t)^2 + (y_e - y_t)^2 + (z_e - z_t)^2}\,. \qquad (2)$$

Similarly, the absolute error in surface normal estimation, $E_{orientation}$, is the angle between the true surface normal, $\mathbf{n}_t$, and the estimated surface normal, $\mathbf{n}_e$:

$$E_{orientation} = \cos^{-1}(\mathbf{n}_t . \mathbf{n}_e)\,. \qquad (3)$$

RESULTS

The pose of the moving target was estimated at six distinct locations (demand instants). The visibilities of each sensor at every demand instants for both fast and slow systems are given in Figures 4 and 5, respectively, and the fused visibility of both systems is shown in Figure 6. As can be noted, the visibilities of the target by the fast system are tangibly higher than that of the slow system. The corresponding absolute position errors associated are shown in Figure 7 and absolute errors in surface-normal estimations are given in Figure 8. Despite the presence of random noise in both systems, the data confirms the tangible improvement of system performance through the use of the fast system.

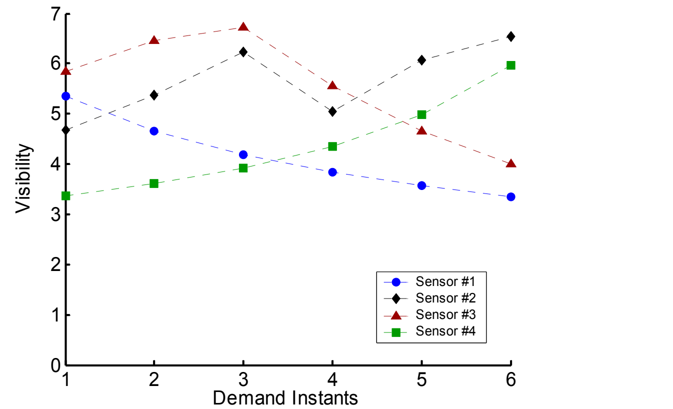


Figure 4. Observed visibilities of each sensor of the fast system in the presence of obstacles.

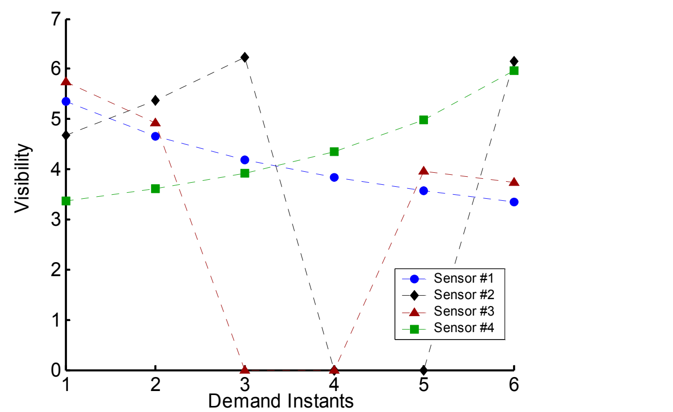


Figure 5. Observed visibilities of each sensor of the slow system in the presence of obstacles.

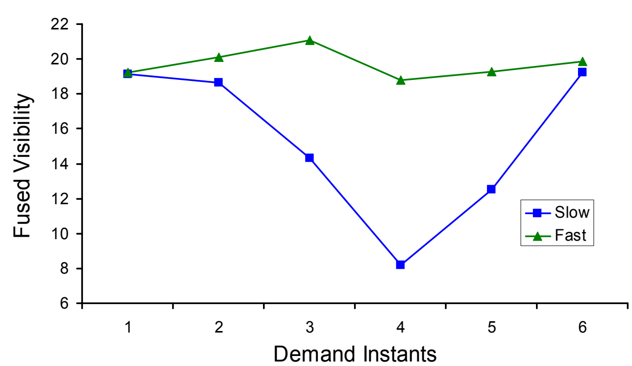


Figure 6. Fused Visibility of both systems.

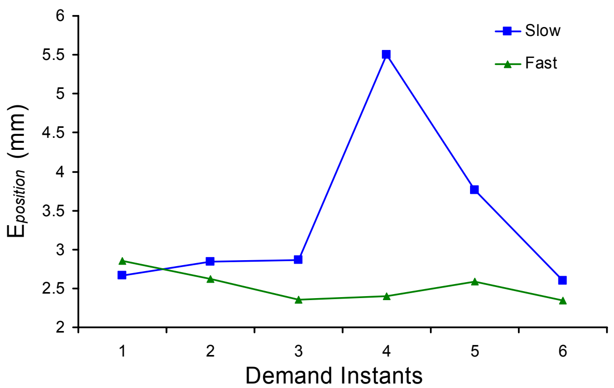


Figure 7. Absolute errors in target position estimates in the presence of obstacles.

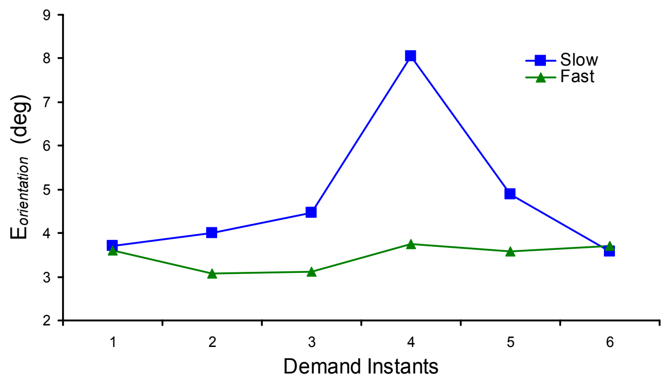


Figure 8. Absolute errors in target's surface normal estimates in the presence of obstacles.

DYNAMIC OBSTACLES – A SIMULATION

The experimental setup has, thus, far been used to test the dispatching approach for the surveillance of a dynamic-target in a static-obstacle environment. However, the presented approach is capable of dealing with the general surveillance problem in a dynamic multi-object environment. In order to illustrate the dispatching system's ability to cope with such environments, numerous simulated experiments were performed, an example of which is presented in this Section.

The following simulated experiment considered the same systems described above (i.e., fast and slow) plus an additional static system (all cameras restricted to their initial configuration). The target also follows the same trajectory as above, however, in these experiments multiple dynamic obstacles are utilized. The obstacles move along trajectories shown in Figure 9 with a 7 mm/s constant velocity. The visibilities of each camera for the fast, slow, and static systems are shown in Figures 10-12, respectively, and the fused visibility is shown in Figure 13.

CONCLUSIONS

A novel method is presented in this paper for the coordinated selection and positioning of groups of active-sensors for the autonomous surveillance of a single target in a multi-object dynamic environment. Experiments and simulations, some of which are presented herein, have shown that tangible improvements in performance can be obtained through the use of multiple active-vision sensors controlled by the proposed dispatching algorithm. Furthermore, it was also concluded that as the dynamic capabilities of the obstacles increase the dynamic capabilities of the sensors become more influential in the performance of the surveillance system.

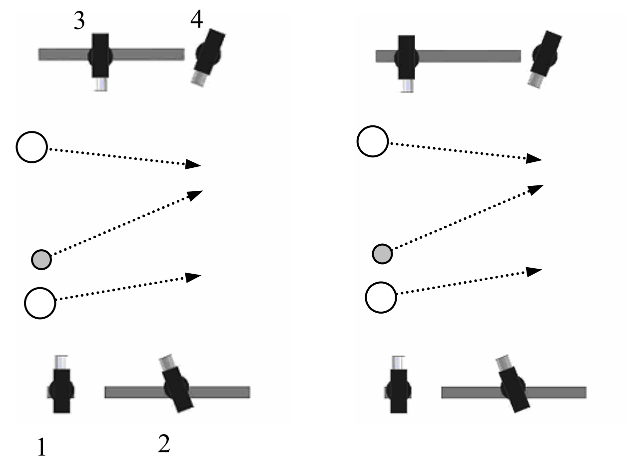


Figure 9. Initial sensor poses for (a) slow system and (b) fast system.

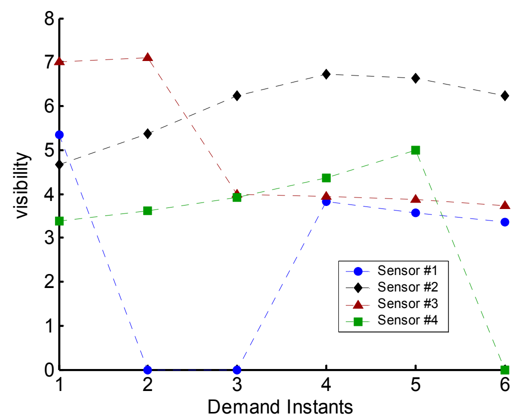


Figure 10. Observed visibilities of each sensor of the fast system in the presence of dynamic obstacles.

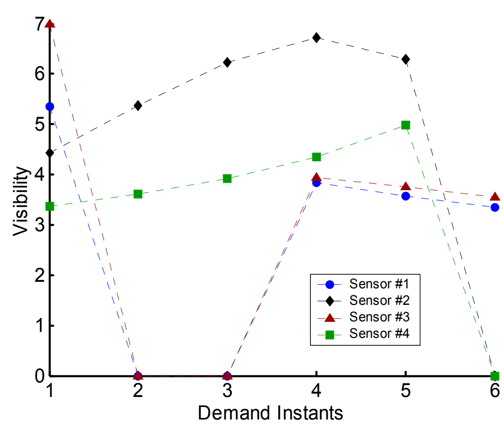


Figure 11. Observed visibilities of each sensor of the slow system in the presence of dynamic obstacles.

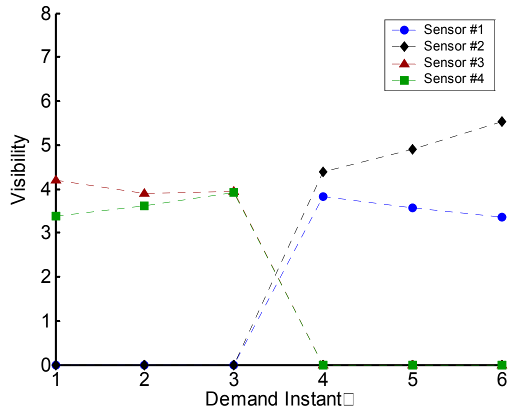


Figure 12. Observed visibilities of each sensor of the static system in the presence of dynamic obstacles.

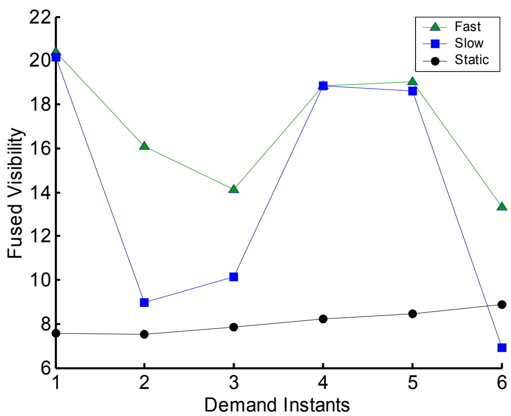


Figure 13. Fused Visibility of all three systems in the presence of dynamic obstacles.

ACKNOWLEDGEMENTS

This work has been supported by the Natural Science and Engineering Research Council of Canada (NSERC).

REFERENCES

[1] K. A. Tarabanis, P. K. Allen, and R. Y. Tsai, "*A Survey of Sensor Planning in Computer Vision*," IEEE Transactions on Robotics and Automation, Vol. 11, Iss. 1, pp. 86–104, Feb. 1995.

[2] S. Sakane, T. Sato and M. Kakikura, "*Model-Based Planning of Visual Sensors Using a Hand-Eye Action Simulator: HEAVEN*," Int. Conf. on Advanced Robotics, Versailles, France, pp. 163–174, 1987.

[3] T. Matsuyama, T. Wada, and S. Tokai, "*Active Image Capturing and Dynamic Scene Visualization by Cooperative Distributed Vision,*" in Advanced Multimedia Content Processing, Vol. 1554 of Lecture Notes in Computer Science, Berlin, pp. 252–288, 1999.

[4] C. Cowan and P. Kovesi, "*Automatic Sensor Placement from Vision Task Requirements*," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 10, Iss. 3, pp. 407–416, 1988.

[5] K. Tarabanis, R. Tsai, and P. Allen, "*Analytical Characterization of the Feature Detectability Constraints of Resolution, Focus, and Field of View for Vision Planning*," CVGIP: Image Understanding, Vol. 59, pp. 340–358, 1994.

[6] S. A. Hutchinson and A. C. Kak, "*Spar: A Planner that Satisfies Operational and Geometric Goals in Uncertain Environments*," AI Magazine, Vol. 11, Iss. 1, pp. 20–61, 1990.

[7] T. Levit, J. Agosta, and T. Binford, "*Bayesian Methods for Interpretation and Control in Multi-agent Vision Systems*," SPIE Applications of AI X: Machine Vision and Robotics, Orlando, FL, pp. 536–548, 1992.

[8] R.Y. Tsai and K.Tarabanis, "*Occlusion-Free Sensor Placement Planning*," In Machine Vision for Three Dimensional Scenes, H. Freeman, Ed. Orlando, Fl; Academic, 1990.

[9] Y. Ye and K. Tsotsos "*Sensor Planning for 3D Object Search*," Computer Vision and Image Understanding, Vol. 73, No. 2, pp. 145–168, 1999.

[10] H. N. Psaraftis, "*Dynamic Vehicle Routing Problems*," in Vehicle Routing: Methods and Studies, Amsterdam, Netherlands, pp. 223-248, 1988.

[11] M. D. Naish, E. A. Croft and B. Benhabib, "*Coordinated Dispatching of Proximity Sensors for the Surveillance of Maneuvering Targets*," Journal of Robotics and Computer Integrated Manufacturing, Vol. 19, Iss. 3, pp. 283–299, 2003.

[12] R.E. Kalman, "*A New Approach to Linear Filtering and Prediction Problems*," Transactions of the ASME, Journal of Basic Engineering, Vol. 82, Series D, pp. 35-45, 1961.

[13] Safaee-Rad, R. Tchoukanov, I., Smith, K.C., and Benhabib, B., "*Three-Dimensional Location Estimation of Circular Features for Machine Vision*," IEEE Transactions on Robotics and Automation, Vol. 8, No. 5, pp. 624-640, 1992.