Proceedings of the 2003 IEEE International Conference on Systems, Man and Cybernetics (SMC 2003)
Washington, D.C., Oct. 5-8, 2003

# A Multi-Sensor Surveillance System for Active-Vision Based Object Localization

**Ardevan Bakhtari, Maryam Eskandari, 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, meskanda, beno]@mie.utoronto.ca

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

***Abstract*** – *In this paper, the implementation of a novel surveillance system that incorporates multiple active-vision sensors controlled by a real-time dispatching algorithm is presented. The proposed system improves reliability and accuracy of target surveillance – tracking systems used for visual-servoing and other similar applications.*

*Experiments using a dispatched system have shown that the use of dynamic sensors can improve the performance of a surveillance system, primarily, due to the following factors: (i) decrease in the uncertainty associated with the object's estimated pose, (ii) increase in robustness of the system due to its ability to cope with a wider range of a priori unknown object trajectories, and (iii) increase in reliability through sensory fault tolerance.*

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

## 1 Introduction

Many autonomous tasks require sensory data to be collected in real time. The coordination and control of multiple mobile sensors provide an opportunity to significantly improve the quality and robustness of the acquired data as compared to information collected by a single-sensor and/or static systems. As sensor fusion may be used to combine information from multiple, coordinated sensors into a single representation, reducing the uncertainty of the measured data [3], sensing-system reconfiguration via effective dispatching can be used to increase sensor performance [5].

The principles of dispatching used for the operation of service vehicles (e.g., taxicabs and ambulances) can be similarly used in effective on-line sensing-system reconfiguration [7]. On-line dispatching requires selecting an optimal subset of dynamic sensors to be used in a data fusion process and manoeuvring them in response to the motion of the object. The goal is to provide information of increased quality for the task at hand while ensuring adequate response to future object manoeuvres.

The system proposed in [9] 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. Matsuyama et al. [4] determine 2-D camera layouts using an off-line optimization. Deviations from the planned approach are handled on-line using heuristics to adjust camera actions and temporal camera switching points. Such systems, however, rely heavily on *a priori* knowledge and, therefore, are not robust to variations in target trajectory. The dispatching algorithm presented herein utilizes, if available, *a priori* objective knowledge, however, can quickly adapt to unexpected trajectory variations. The specific focus of this paper is the implementation and evaluation of an active-vision target-tracking system that utilizes an online-dispatching algorithm developed in our laboratory.

## 2 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. In other words, a sensor fusion process does not need to combine the information from all of the sensors in the system. Instead, a subset of sensors ($k \leq n$, where $k$ is the subset size and $n$ is the total number of sensors) may be selected to survey the object at a particular demand instant. This group of sensors is referred to herein as a fusion subset.

In this context, the general sensor dispatching problem may be stated as: Given a set of sensors and a set of time-varying demand points (based on the predicted motion of a manoeuvring 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.

There are a number of different methods by which this sensor dispatching problem may be approached. These include optimization techniques, machine learning and heuristic approaches. One such heuristic sensor dispatching approach was detailed in [5], a brief review of which follows.

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. The rolling horizon serves as a basis for dispatching, allowing the dispatching methodology to balance the immediate need of servicing the first demand with the anticipated needs of future demands.

Dispatching is accomplished using two complementary strategies. A coordination strategy is used to specify which sensors will be used at each demand instant. The dispatcher must evaluate the fitness of each sensor for the immediate sensing demand. The pose of each sensor is specified through a positioning strategy. In this approach, the demands are considered sequentially, limited to the rolling horizon. An assignment search method considers only the first demand instant; the remaining demand instants on the rolling horizon are considered by a preassignment search method.

## 2.1 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 $k$) 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.

Assignment is triggered by an object entering the workspace or the completion of a previous service interval. Once assigned, the subset of sensors cannot be altered until the demand is serviced (i.e., data is acquired), completing the service interval. This contrasts with the desired poses of the assigned sensors which may be altered in real-time. The uncertainty of the predicted demand point locations at the time of assignment necessitates reevaluation over the service interval to ensure that the sensors are in the best possible positions. Pose adjustment is handled by a replanning method, Section 2.3.

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

1. Predict the object's pose, $D_j$, with respect to the world coordinate frame, at the demand instant, $t_j$.
2. For every sensor, $S_i$, $i = 1 \ldots n$:
   (a) Determine its best achievable pose with respect to $D_j$, and
   (b) Assess the corresponding (single sensor) performance metric, visibility of $D_j$, $v_s$, from the best achievable pose.
3. Rank all $S_i$ according to the achievable visibility, $v_s$, from highest to lowest.
4. Assign the top $k$ ranked sensors to $t_j$. (The desired pose of each assigned sensor is the best achievable pose determined in Step 2a).

## 2.2 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 that outlined in Section 2.1 for assignment and positioning; however, there are some differences. First, while the suitability of each sensor is considered for each demand, only those that have not been previously (pre)assigned may be preassigned to the demand under consideration. Second, the preassignment algorithm loops to consider additional demand instants until either all sensors have been preassigned or the end of the rolling horizon has been reached. This approach aims to service each demand with the sensors that can provide the highest quality data.

## 2.3 Initial System Configuration and Replanning

The quality of information provided through sensor dispatching is dependant on the initial pose of each sensor within the workspace (i.e., the pose of each sensor prior to object entry). The effect of the initial sensor placement becomes more pronounced as the speed of the sensors decreases relative to the speed of the object. In general, as sensor speed decreases, adequate performance requires the sensors to become more widely distributed in the workspace. If the object trajectory is known *a priori*, the sensing system can be reconfigured in an optimal manner. While beyond the scope of this paper, one approach for determining initial sensing-system configurations is outlined in [6].

As each sensor is assigned to a particular demand point, the desired pose of the sensor with respect to the demand point is also specified. This desired pose is used as input to the sensor motion controller. As the service interval elapses, the estimates of the demand point locations are continually updated as additional observations become available. If the demand point at the time of assignment lies outside of a specified confidence interval for the newly predicted demand point position then the demand point is replaced with the new prediction. Upon replacement, the desired pose of any sensor assigned to the demand is adjusted to reflect the change in demand position.

# 3 Implementation

An experimental setup was devised to evaluate the performance of the proposed dispatching algorithm. Although the algorithm can be used for a wide range of surveillance applications, the specific system detailed herein focuses on object tracking and localization: The system uses four mobile cameras to determine the position and orientation of a single target represented by a circular marker, maneuvering through a pre-defined workspace on a planar trajectory, Figure 1. A stationary overhead camera is utilized to obtain estimates of the gross target motion. These estimates are used to plan the motion of the four mobile cameras. The planar-motion cameras 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 a component list).

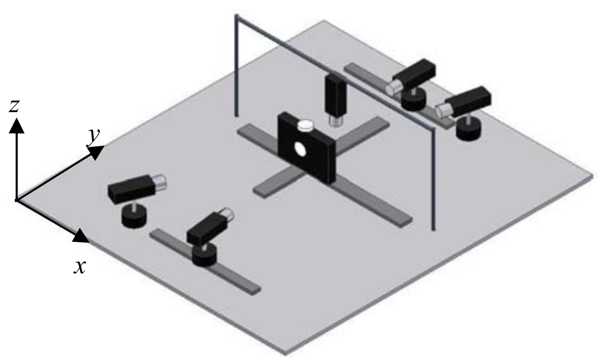



Figure 1. Physical system layout.

Table 1. Hardware specifications.

| Hardware | Characteristic |
|---|---|
| *Linear Stages* | *Range: 300 mm/500 mm*<br>*Positional Accuracy: 18 µm*<br>*Max velocity: 1.5 m/sec* |
| *Rotary Stages* | *Positional Accuracy: 12 arc sec*<br>*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 target's current position as well as a prediction of its future positions. 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) [2]. The KF is initialized with one of three motion models, $1^{st}$ order Gauss-Markov, constant jerk, or constant velocity depending on the expected target-motion trajectory. The acceleration and velocity of the target are assumed to be decoupled in the $x$ and $y$ directions and the acceleration in both directions is assumed to be subjected to Gaussian white noise. 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. The module uses the dispatching algorithm discussed in Section 2 to assess optimal sensor selection and positioning in accordance with the target's predicted location. Sensor positions are updated at the frequency of the prediction module, about 15 Hz.

***Motion Module:*** This module controls the motions of the cameras. The input to this module are the desired camera locations from the dispatching module, while the outputs are the motion commands (poses and velocity) 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. (In addition, this module independently controls motion of the target trajectory and records its actual pose during imaging for later assessment of the system performance.)

***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 (position and orientation) of the marker is estimated using an analytical solution developed earlier in our laboratory [8]. Through the knowledge of the pre-calibrated cameras' positions and bearings, as provided by the dispatching module, all estimates are determined in world coordinates. The data from cameras is then fused to decrease the uncertainty in the final estimate of target's pose.

The specific fusion algorithm was chosen based on three criteria: The algorithm should require a minimum amount of *a priori* knowledge about the target's trajectory allowing system robustness to unexpected trajectory variations; data fusion must be carried out optimally, where uncertainties in all cameras are estimated and considered during the fusion process; and the fusion process should allow some degree of fault tolerance in the case of malfunction, where the invalid data must be identified and discarded.

The utilized data-fusion algorithm, developed by Brooks [1], uses range trees recursively to find an optimal region, where the true target position/orientation is located. Through off-line sensor modelling, (see Appendix), the uncertainty range of each camera is estimated and, subsequently, used to estimate a probability range, $R_n$, for a specific confidence level. Given multiple ranges (one per camera) the maximum intersection range, $R_e$, (if all cameras are functioning properly) is chosen as the optimal estimated region, whose mid-point is taken as the target's position/orientation:

$$R_e = R_1 \cap R_2 \cap \ldots \cap R_n . \quad (1)$$

In order to illustrate the data-fusion algorithm, let us consider the following example: Four cameras individually estimate target's $x$ position. Camera #1 returns an $x$ value of 1.5 m, where through off-line sensor modeling it is known that, with 99% confidence, the true target position is within ±0.2 m of the reading and, therefore, the range is defined to be 1.3 to 1.7 m. Individual ranges are obtained similarly for all camera readings, Figure 2.

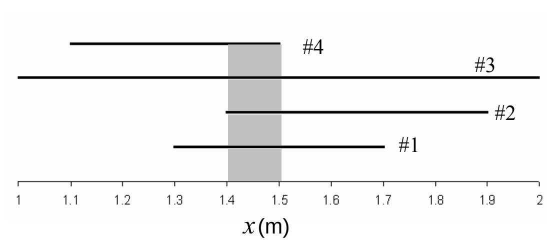



Figure 2. The intersection of four probability ranges.

The intersection of the ranges marked, by the grey region, Figure 2 ($1.4<x<1.5$), is taken as the range of true target's position, and the fused estimate is the midpoint of the region ($x$=1.45m). This example illustrates 1-D estimation with no fault tolerance allowed. However the data-fusion algorithm was employed for 3D target estimation and single fault tolerance.

# 4 An Experiment: Dispatching versus Non-Dispatching Surveillance

Numerous experiments were conducted in our laboratory, using the active vision set-up presented in Section 3, to illustrate the implementation of the proposed dispatching algorithm. The experiments verified that the performance of a surveillance system can be tangibly improved with the use of active sensors controlled by a dispatching algorithm. As noted in Section 2, the improvement is primarily due to (i) increased robustness of the system (i.e., its ability to cope with *a priori* unknown target trajectories), (ii) decreased uncertainty associated with estimating the target's pose, and (iii) increased reliability through sensory fault tolerance.

In the specific example discussed herein, the performances of two systems are compared; one system is equipped with four active-vision sensors (CMOS cameras) while the other system comprises four static sensors. Both systems track a moving target with *a priori* unknown trajectory. The dispatched system uses best three out of four cameras for each imaging instants, while the non-dispatched system images the target using all four cameras.

## 4.1 Trajectory Selection

The cameras for both the active and static systems were placed into optimal initial poses, with the expectation that the target would enter from one corner of the workspace and follow a straight-line trajectory to the opposite diagonal corner, Figure 3. The actual trajectory followed by the target, however, was intentionally chosen to be different than this expected trajectory. The prime objective of these experiments was to highlight the ability of the dispatching algorithm to deal with variances in target trajectories through replanning of sensor locations, while the static sensors are completely restricted to their initial configurations. The size of the tracking zone was limited by the viewing area of the non-dispatched system.

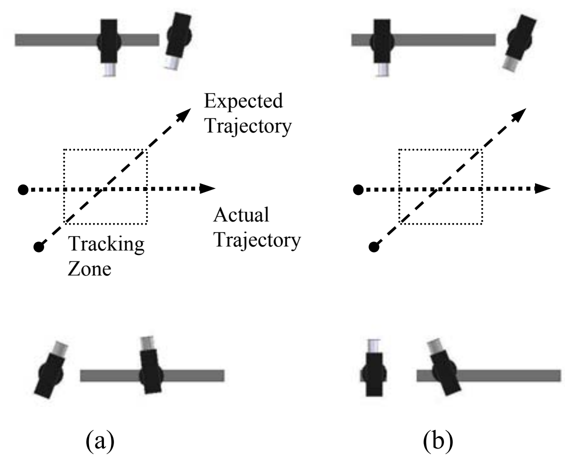



Figure 3. Optimal initial configuration for (a) the non-dispatched (b) the dispatched system and expected and actual target trajectories.

The speeds of the active sensors were chosen as $dx/dt$=30 mm/s and $d\alpha/dt$=1 rad/sec, while the speed of the target was set to 15 mm/sec. Through simulation it was determined that for active sensors, exceeding the aforementioned speeds does not significantly improve system performance.

## 4.2 Performance Evaluation

System evaluation was carried out using a visibility metric as previously proposed in [6], see Appendix. 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, $\mathbf{n}_e$:

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

## 4.3 Results

The pose of the moving target was estimated at ten distinct locations (demand instants). The visibilities of the demand instants for both dynamic and static systems are given in Figure 4. As can be noted, the visibility of the target by the dispatched system is tangibly higher than that by the non-dispatched system. As may also be noted, visibility of the static system is fairly constant over the entire target trajectory. This is due to two main factors: the constant camera bearings and the monotonic motion of the target. The corresponding absolute position errors associated are shown in Figure 5 for two repeated experiments under identical conditions. The absolute errors in surface-normal estimations are given in Figure 6. Despite the presence of random noise in both systems, the data confirms the tangible improvement of system performance through the use of a dispatching algorithm.

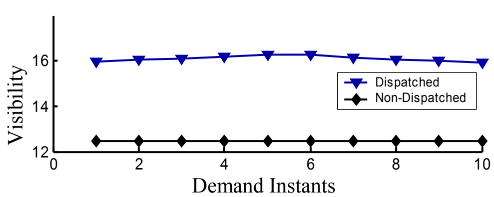



Figure 4. Observed visibility.

# 5 Conclusions

The implementation of a novel multi-sensor surveillance system has been presented in this paper for target localization in on-line applications. Dispatching is utilized to optimally select and position groups of sensors to track a moving target. Extensive experiments, some of which are presented herein, have shown that tangible improvements in accuracy and reliability can be obtained through the use of multi-active vision sensors controlled by the proposed dispatching algorithm.

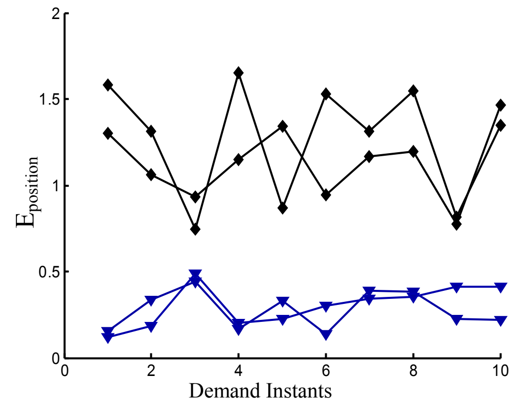



Figure 5. Absolute errors in target position estimates.

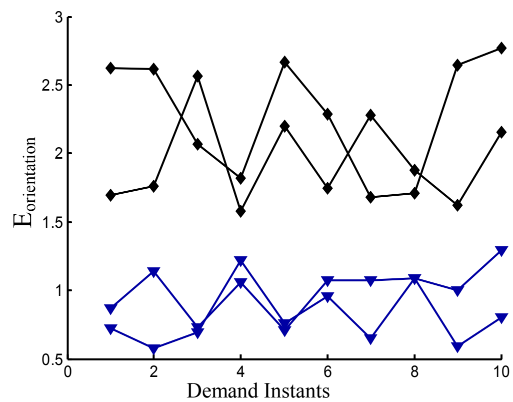



Figure 6. Absolute errors in target's surface normal estimates.

# Appendix

### A1. Sensor Modeling

Sensor modeling is an important part of optimal dispatching: The objective is to estimate a sensor's performance given a set of environmental conditions. The performance of each camera, in this work, is described by a visibility performance metric:

$$V_s = \frac{1}{\| R \|}, \qquad (A1)$$

where, ||R|| is the Euclidean norm of the covariance matrix associated with the parameter measurements. For the cameras used in our experiments, there are four variance measurements, three for target's position $(x,y,z)$ and one for orientation $(n_x,n_y,n_z)$. Through variation analysis it was determined that only two controlled parameters significantly affect the measurement variances: the Euclidean camera-to-target distance, $d$, and the camera's bearing, $\theta$, Figure A1.

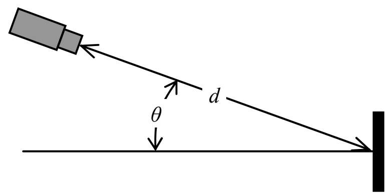



Figure A1. Camera's distance, $d$, and bearing, $\theta$.

### A2. Factorial Experiments for Variance Estimation

Two-factorial experiments were performed to determine the relationship between each measurement variance and the two controlled parameters, $d$ and $\theta$. As an example, results for two variances (estimation along the y-axis and surface normal) are shown in Figure A2. As can be noted, the distance and bearing of the camera significantly affect its performance in target localization. The improvement in variance is mainly due to the shape of the elliptical projection of the circular marker (target viewed better at around 20-40°) and noise in images has lower noise-to-signal ratios at close distances, (i.e., larger marker images).

In order to evaluate the complete system performance, the visibility of the fused estimates of data from multiple cameras is required. The visibility of k sensors, whose measurements are combined using sensor fusion, is defined as:

$$V_f = \frac{1}{\| P \|}, \tag{A2}$$

where, $P$ is the fused covariance matrix,

$$P = \left[ \sum_k R_i^{-1} \right]^{-1}. \tag{A3}$$

## References


[1] Brooks, Multi-Sensor Fusion, Fundamentals and Applications, Prentice Hall, New Jersey, 1998.

[2] 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.

[3] R. C. Luo and M. G. Kay, "Multisensor integration and fusion for intelligent machines and systems", in *Data Fusion in Robotics and Machine Intelligence*, Academic Press, San Diego, CA., pp. 7–136, 1992.

[4] 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.

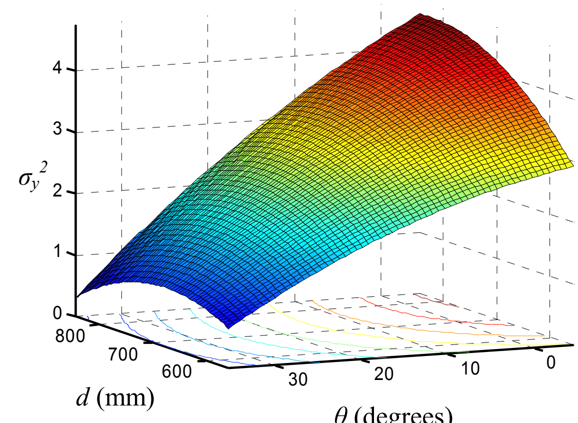
[5] M. D. Naish, E. A. Croft and B. Benhabib, "Coordinated dispatching of proximity sensors for the surveillance of maneuvering targets", *Journal of Robotics and Computer Integrated Manufacturing*, Vol. 19, Iss. 3, pp. 283–299, 2003.

[6] M. D. Naish, E. A. Croft and B. Benhabib, "Simulation-based sensing-system configuration for dynamic dispatching", IEEE, Int. Conf. on Systems, Man and Cybernetics, Tucson, AZ. Vol. 5, pp. 2964–2969, 2001.

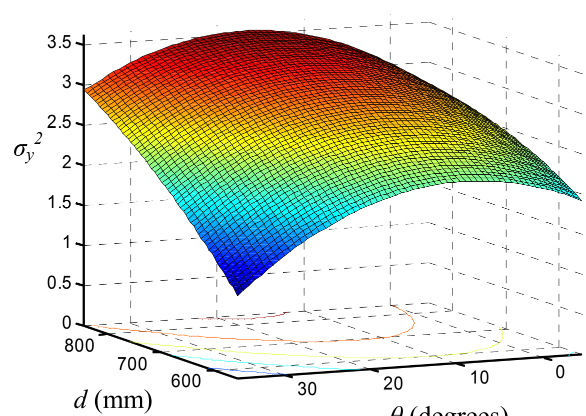
[7] H. N., Psaraftis, "Dynamic vehicle routing problems", in *Vehicle Routing: Methods and Studies*, Amsterdam, Netherlands, pp. 223-248, 1988.

[8] Safaee-Rad, R. Tchoukanov, I., Smith, K.C., and Benhabib, B., "Three-dimensional location estimation of circular features for machine vision", in *IEEE Transactions on Robotics and Automation,* Vol. 8, No. 5, pp. 624-640, 1992.

[9] 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.


(a)




(b)

Figure A2. Response surfaces of variances in (a) position and (b) orientation estimation.