

Energy-Aware High Resolution Image Acquisition via Heterogeneous Image Sensors

Hyungjin Kim, *Student Member, IEEE*, Mohammad Rahimi, *Member, IEEE*, Dong-U Lee, *Member, IEEE*, Deborah Estrin, *Fellow, IEEE*, and John D. Villasenor, *Senior Member, IEEE*

Abstract—We present an energy-efficient high resolution image acquisition approach based on a two-tiered system comprising low-cost, low-power, non-actuated, extremely resource-constrained stereo image sensor platforms and more capable but more power-consumptive high resolution imaging platforms with actuation capability. The resource constrained platforms are used to compute 3-D object location and subsequently to compute appropriate pan/tilt/zoom settings for the high resolution imaging platforms. The high resolution imaging platforms with actuation capability acquire high resolution images which can be utilized for various recognition purposes. We present a design methodology and system architecture, and evaluate latency and energy tradeoffs in the system. Experimental results show that use of the two-tiered network significantly reduces energy consumption of high resolution image acquisition versus a single-tiered network with minimum loss in detection capability.

Index Terms—Distributed detection, finite wordlength effects, fixed point arithmetic, image sensors.

I. INTRODUCTION AND RELATED WORK

IN this paper, we address the 3-D localization problem in the specific context of a two-tiered hierarchical network containing 1) a lower tier comprising multiple low-power image sensor platforms with severe limits on image resolution, image quality, processing and communications, and 2) an upper tier comprising more capable platforms but more energy-consumptive high resolution imaging devices. We further constrain the problem by considering environments in which objects of interest appear only sporadically, but when they do appear, it is of paramount importance that they should be detected and located. Environmental monitoring, security, various defense-related scenarios, and many other applications fit this profile.

In the system considered here, one or more stereo pairs of low-power sensors are configured to continuously monitor an

environment, identify the presence of objects, compute their location in 3-D space, and then direct the high resolution actuated sensor(s) to acquire and process zoomed images of the appropriate location, as shown in Fig. 1. There are many applications that map to this tiered architecture. For instance, higher resolution images can be used for vehicle classification and license plate matching, while lower resolution images could be used for the simpler task of vehicle detection. Another example is face detection and recognition that can be triggered by human detection sensors.

There are important advantages of using stereo imaging in the lower-tier of the system. Stereo sensors provide the ability to locate objects precisely in 3-D space and hence accurately redirect the actuated camera(s) toward the objects. In addition, object location information can be continuously recorded by lower-tiered stereo nodes for more accurate tracking of the objects as well as for advanced object motion-based analysis and recognition in post-processing stages [1], [2].

Continuous monitoring of the environment has a cost in energy. To minimize this cost, the system architecture includes a sequential task flow which consists of a primary and secondary lower-tiered sensor nodes. The primary sensor node continuously monitors the environment for the objects and only wakes up the secondary node to send the 2-D location of the object (not the image) when it detects the objects in its field of view. Furthermore, the secondary node performs a platform-optimized 3-D localization that restricts computations to fixed-point arithmetic considering both the precision needed in 3-D computation and the native word-length of the sensor node processor. As described further below, this approach reduces the energy consumption of the 3-D computation phase by approximately a factor of four.

The main contribution of this paper is a design approach to enable detection and high resolution image acquisition of objects via a heterogeneous network that utilizes stereo nodes at lower-tier and actuated camera(s) at upper-tier in the network. Through simulation and experimentation using commercially available Cyclops imagers [3] and high performance actuated Canon cameras, we show that use of the heterogeneous architecture enables very good detection performance compared to the single-tiered system, while minimizing the overall latency and energy cost of the system. Another contribution of this paper is the application of precision-optimized arithmetic to 3-D localization algorithm in the lower-tier platforms. We characterize the stereo performance of constrained sensor node and illustrate the significant energy reduction in 3-D localization of the objects using our processor-optimized approach.

Manuscript received October 30, 2007; revised April 28, 2008. Current version published September 17, 2008. This work was supported in part by the Office of Naval Research under Contract N00014-06-1-0253 and in part by the National Science Foundation under Grants CCR-0120778 and CCF-0541453. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Wayne Wolf.

H. Kim and J. D. Villasenor are with the Electrical Engineering Department, University of California, Los Angeles, CA 90095-1594 USA (e-mail: hjkimnov@ee.ucla.edu; villa@icsl.ucla.edu).

M. Rahimi is with the Center for Embedded Networked Sensing, University of California, Los Angeles, CA 90095-1594 USA (e-mail: mhr@cens.ucla.edu).

D. Lee was with the Electrical Engineering Department, University of California, Los Angeles, CA 90095-1594 USA. He is now with Mojix, Inc., Los Angeles, CA 90025 USA (e-mail: dongu@mojix.com).

Deborah Estrin is with the Computer Science Department, University of California, Los Angeles, CA 90095-1594 USA (e-mail: destrin@cs.ucla.edu).

Digital Object Identifier 10.1109/JSTSP.2008.925985

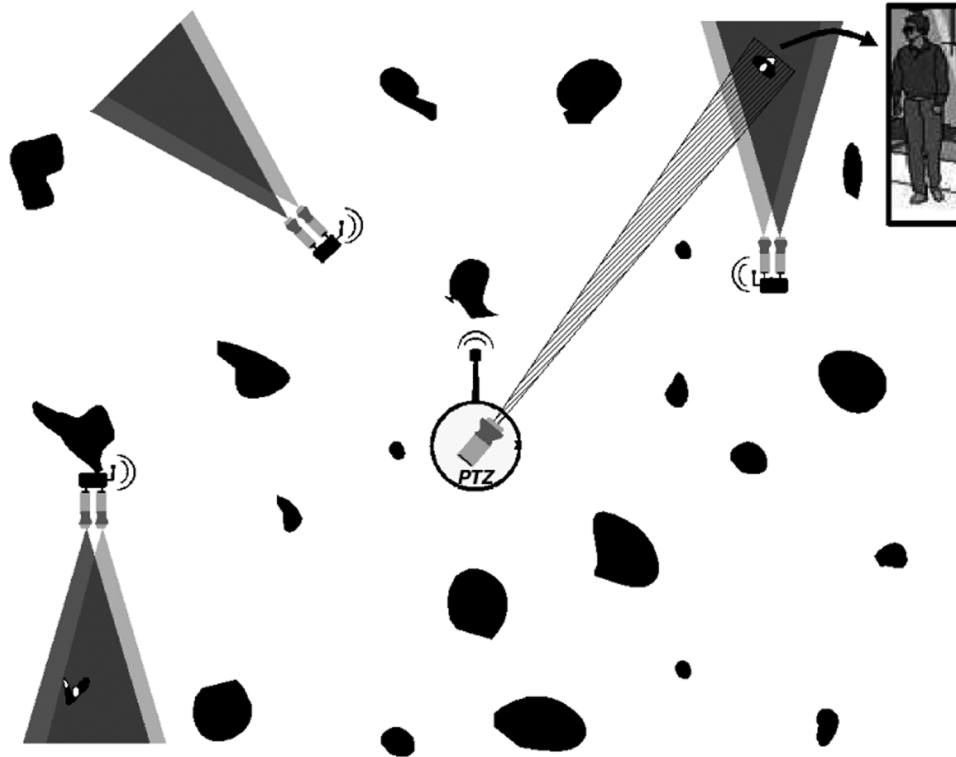


Fig. 1. Diagram of system setup with stereo pairs of low-power camera sensors and a high resolution pan-tilt-zoom (PTZ) actuated camera.

There is significant prior research in using actuated cameras for object detection and classification. In particular, Kansal *et al.* [4] demonstrate the capability of a single-tiered actuated camera system for achieving high resolution images. In this system, the actuated camera provides extended detection coverage in the fully zoomed-out state as well as high resolution images in the fully zoomed-in state. The advantage of such a single-tiered actuated system is reduced system cost due to reduced number of system components. However, potential disadvantages include latency due to the delay of scanning the environment and additional power consumption due to continuous use of high power actuated camera(s). Hence, applications with stringent latency or energy requirement can benefit from a heterogeneous architecture.

Another closely related research is the Mesheye project [5] which utilizes an approach with two classes of image sensors. The Mesheye platform consists of a low resolution stereo vision system which continuously determines the position, range, and size of moving objects in its field of view, and a color module which can be triggered to acquire a high resolution image. The approach in [5] is based on a high resolution camera and computationally-capable processor in each node. By contrast, in the system described here most of the nodes are equipped with a very low power (and computationally less powerful) processor, and actuation is used to allow a single high resolution camera to cover a large surveillance area.

The use of multiple-tier camera networks in the sensor networks context was pioneered by the *SensEye* project [6], in which a three-tier camera network was used to explore the energy benefits and object localization success rate achievable through the use of the sensing hierarchy. *SensEye* consists of

a three-tier system comprising very constrained static camera nodes, high resolution static camera nodes, and high resolution actuated camera nodes. As a result, the actuated camera nodes are used principally to fill the gaps in the coverage of the second tier, providing additional coverage redundancy. In contrast to *SensEye*, which places emphasis on use of the hierarchy to fill any coverage gaps, the emphasis in the present work is on the integration of a lower tier of low resolution imagers with a higher tier of more capable platforms that, working together, enable objects to be found, tracked, and imaged in high resolution.

Another important earlier effort in this area is the work of Wolf *et al.*, who describe a distributed smart camera architecture that utilizes two different types of cameras that collaborate to obtain a high resolution image of a scene [7]. In this system, wide-angle cameras coordinate a set of close-up cameras, which in turn take detailed views of the scene, process parts of the scene, and combine them to generate a high resolution image. Both real-time and low-energy operation were considered by utilizing smart cameras [8] in the architecture.

Another important component of the present work is camera calibration, aspects of which have been addressed in various previous publications. In particular, Lymberopoulos *et al.* [9] presented methods for node localization and camera calibration using the shared field of view of camera pairs. Funiak *et al.* [10] examined automatic calibration of large camera networks using a distributed algorithm to track moving objects and Devarajan and Radke [11] demonstrated a distributed algorithm for automatic calibration of extrinsic parameters in a camera sensor network with no centralized processor. The mathematical framework in those approaches can be applied to calibrate the present

system, though we introduce processor-specific optimizations that significantly reduce the energy of these computations.

There have also been important related energy-aware sensing studies that use non-vision sensing modalities. Many of these techniques utilize a combination of low-power but less precise sensors with more accurate but high-power sensors for detection of events. For instance, Dutta *et al.* describe the design of a sensor network platform for detecting rare, random, and ephemeral events by utilizing multimodal (infrared, acoustic, and magnetic) sensors [12]. The authors describe a system in which power is duty-cycled or one sensor is used to trigger another, in order to minimize energy utilization. The system we describe in the present paper can be considered analogous to this approach in the sense that the primary sensor node in a lower tier triggers a secondary sensor node and the associated actuated camera. In addition, Benbasat *et al.* recently presented a formal framework for power-efficient selection of the combination of sensors in wearable sensor nodes based on a state detection scheme that utilizes a decision tree classifier which dynamically adjusts the activation and sampling rate of the sensors [13].

The remainder of this paper is organized as follows. Section II describes the system and related algorithms in detail Section III overviews system components and task flow Section IV discusses design considerations Section V shows simulation results on performance of the lower tier platform, experimental results for simplified two-tiered network, and simulation results on multiple lower-tier platforms with one high resolution actuated camera. Finally, Section VI summarizes the work.

II. METHODOLOGY

A. Low Power Stereo Pair

Determining 3-D location from multiple 2-D images is a well studied problem, and includes the two basic phases of 1) camera calibration involving intrinsic and extrinsic parameters and 2) the subsequent computations needed to compute actual 3-D location. In the context of energy-constrained sensor networks, the challenge lies in mapping these algorithms in the most energy efficient manner to resource-constrained sensors. This mapping can significantly impact the performance of the system. We briefly describe sensor calibration, 3-D location determination, and platform specific optimization. More details are available in [14] and [15].

1) *Camera Calibration:* Camera calibration aims to find a 3×3 camera calibration matrix K for each camera, and a 3×3 rotation matrix R and a 3×1 translation vector \mathbf{t} for camera pairs. K represents intrinsic information using the focal length f and a principal point (p_x, p_y) , where a principal point is the point where the line from the optical center of the camera and perpendicular to the image plane meets the image plane:

$$K = \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix}. \quad (1)$$

Various experimental procedures can be used for determining the elements of K involving, for example, the use of images of a

checkerboard taken from each camera. R and \mathbf{t} are extrinsic and are thus a function of the relative locations and orientations of camera pairs, and can be computed through the corresponding locations of imaged points in images acquired by the camera pairs. For a camera network consisting of camera $1, 2, \dots, i$, we assume that camera 1 is located at the origin of the world coordinate system and that the z -axis of the world coordinate system is coincident with the optical axis of camera 1. In combination, K_i, R_i , and \mathbf{t}_i determine the 3×4 camera projection matrix of camera i , P_i , which governs the relation between the homogeneous 2-D image coordinates of camera i , \mathbf{x}_i , and the homogeneous 3-D world coordinate \mathbf{X} through

$$\mathbf{x}_i = P_i \mathbf{X} = K_i [R_i | \mathbf{t}_i] \mathbf{X}. \quad (2)$$

For $i = 1$, R_1 is a 3×3 identity matrix and \mathbf{t}_1 is a 3×1 zero vector. The calibration process can be accomplished in a variety of ways, including the use of packages such as the MATLAB Camera Calibration Toolbox [16].

2) *3-D Location Determination:* There are several ways of performing 3-D location determination using image data from calibrated cameras. The most general approach is based on singular value decomposition (SVD), which uses 2-D coordinates from n calibrated cameras as input to a $2n \times 4$ SVD that gives the 3-D object location. An alternative approach is the closest point algorithm (CPA) [17] which provides 3-D location from image information from two calibrated cameras. Because CPA is simpler and thus better suited to energy-optimized implementation, it was chosen for this work.

CPA identifies the location midway between the closest approach of two lines extending from the optical center of each camera through the imaged object. The line from camera i , \mathbf{l}_i , can be expressed as $\mathbf{l}_i(s_i) = R_i^T(\mathbf{x}_{ni}s_i - \mathbf{t}_i)$ where R_i and \mathbf{t}_i is the rotation matrix and translation vector of camera i respectively, $\mathbf{x}_{ni} = K_i^{-1}\mathbf{x}_i$ is the normalized homogeneous 2-D coordinate of camera i , and s_i is a scalar. For \mathbf{l}_i and \mathbf{l}_j , s_i and s_j corresponding to the closest points are obtained through:

$$\mathbf{x}_{ni}^T R_i (\mathbf{l}_i(s_i) - \mathbf{l}_j(s_j)) = 0 \quad (3)$$

$$\mathbf{x}_{nj}^T R_j (\mathbf{l}_i(s_i) - \mathbf{l}_j(s_j)) = 0. \quad (4)$$

For s'_i and s'_j satisfying (3) and (4), the 3-D location can be given as the midpoint of $(\mathbf{l}_i(s'_i) + \mathbf{l}_j(s'_j))/2$. Matrix multiplications in these two equations can be computed through three inner products of the form:

$$R_i^T \mathbf{x}_{nj} = [\mathbf{r}_{i1}^T \mathbf{x}_{nj} \quad \mathbf{r}_{i2}^T \mathbf{x}_{nj} \quad \mathbf{r}_{i3}^T \mathbf{x}_{nj}]^T \quad (5)$$

where \mathbf{r}_{im}^T is the m th row of R_i^T . Equations (3) and (4) require eight inner product computations (24 multiplications and 16 additions) when i or j is 1, because R_1 is an identity matrix and \mathbf{t}_1 is a zero vector. Otherwise, 19 inner product computations (57 multiplications and 38 additions) are required. Because of the large number of inner product computations, in the discussion below we have chosen it as a target for energy-optimized implementation. Equations (3) and (4) also require computations distinct from the inner products (13 multiplications, 10 additions,

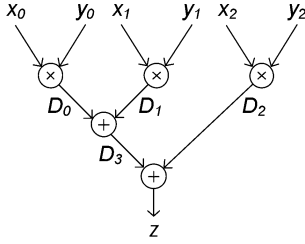


Fig. 2. Inner product operation: $z = x_0 y_0 + x_1 y_1 + x_2 y_2$.

and two divisions when i or j is 1, and 16 multiplications, 13 additions, and two divisions otherwise).

3) *Platform Specific Optimization*: Typically in systems with a dedicated floating-point unit, the data path is constant throughout the computation chain (usually in 32/64/80-bit floating-point). Emulating floating-point operations on fixed-point processors delivers high precision, but is extremely slow, especially on 8-bit processors such as the ATmega128L. Moreover, floating-point accuracy is rarely required in embedded environments, meaning valuable processor cycles and memory are wasted for computing overly precise results. For the present work we restrict computations to fixed-point arithmetic considering both the precision needed and the native word length of the processor using the methodology described in [15]. Where necessary multiword arithmetic is used, but again with the goal of minimizing the words used for each multiword arithmetic operation to minimize execution time while meeting precision targets.

The correct number of bits for each signal in the data path of an algorithm is derived analytically. The fixed-point optimization problem is divided into two parts: 1) range analysis for inspecting the dynamic ranges of signals and computing the minimal number of integer bits, and 2) precision analysis for allocating minimal number of fractional bits to each signal to meet a given precision requirement. Once the signal bit-widths are found, the corresponding C code fragment is generated.

Consider the 3-D inner product operation

$$z = (x_0, x_1, x_2) \cdot (y_0, y_1, y_2) \quad (6)$$

shown in Fig. 2 which is the common operation in the CPA algorithm as discussed in Section II.A. Instead of evaluating the expression in (6) directly, the output of each addition and multiplication operator is stored as a temporary signal ($D_{0..3}$) to allow quantization to take place. Let us assume ranges of $x_{0..2}$ lie in $[-1, 1)$ and the ranges of $y_{0..2}$ lie in $[-0.5, 0.5)$.

Table I shows the set of bit-widths obtained after the optimization process for the inner product operation using the example input ranges given above for the case where the output bit-width is $B_z = 16$ and the worst case error requirement is 2^{-4} . B is the bit-width, while IB and FB refer to the number of integer bits and fractional bits respectively. The shifts shown in the last column are required to perform quantization and binary point alignments.

The optimization procedure can of course be applied for any desired precision requirement, and allows identification of solutions that use the minimum computational resources (and thus

TABLE I
SIGNAL BIT-WIDTHS AFTER OPTIMIZATION FOR INNER PRODUCT
EXAMPLE FOR A WORST CASE ERROR REQUIREMENT OF 2^{-4}

Signal	B	IB	FB	\gg
$x_{0..2}$	8	1	7	0
$y_{0..2}$	8	0	8	0
D_0	8	0	8	7
D_1	8	0	8	7
D_2	16	1	15	0
D_3	16	1	15	-7
z	16	2	14	1

TABLE II
LATENCY AND ENERGY COMPARISONS OF INNER PRODUCT EVALUATION
BETWEEN FLOATING-POINT EMULATION AND CUSTOM PRECISION
FIXED-POINT ON AN ATMEGA128L PROCESSOR

Accuracy	Latency (cycles)	Energy (μ J)
32-bit floating-point	771	2.41
2^{-12} fixed-point	489	1.53
2^{-8} fixed-point	181	0.56
2^{-4} fixed-point	128	0.40

energy) to achieve the target precision. Table II shows the latency and energy required for inner product evaluation on an ATmega128L processor using floating-point emulation and the custom precision fixed-point methods described above. Energy in the table is obtained based on the MCU active power of 23 mW [3] and clock frequency of 7.37 MHz. An accuracy of 2^{-8} fixed-point is sufficient for 3-D location given the spatial resolution limits of 128 by 128 images and requires approximately four times less energy than the 32-bit floating-point computation. The example in Table I, which targeted a precision of 2^{-4} , gives an energy savings of approximately 80% over the 32-bit floating-point implementation. The other operations involved in 3-D location (in addition to the inner products) can be optimized in a similar manner, with similar energy reductions as described in more detail in Section V.A.

B. High Resolution Actuated Imaging Platform

The performance of high resolution image acquisition is strongly dependent on the accuracy of pan/tilt/zoom parameters computed in and sent from low-power stereo sensor nodes. The accuracy is based on relative orientations and locations among cameras which can be obtained through careful network calibration. Here, we briefly describe camera network calibration and actuation parameter setting for the actuated imaging platform.

1) *Network Calibration*: To successfully acquire high resolution images of the objects that have been detected by stereo sensor nodes, it is necessary to calibrate the intrinsic and extrinsic parameters of the actuated camera and the stereo sensors. The procedure includes deployment of the entire network including the actuated camera and stereo pairs followed by manual or automatic calibration of the actuated camera and sensor pairs. When a very small number of sensor platforms are deployed, calibration can be done manually. First, images are taken of a checkerboard or other known pattern for each stereo sensor pair and the actuated camera node. Next, the methodology described

in Section II-A is applied, which provides orientations and locations of each stereo sensor pair relative to the actuated camera. When many sensors need to be deployed, automatic calibration can be utilized. Algorithms of automatic calibration in camera network calibration can be found in [9]–[11].

2) *Actuation Parameter Setting*: To redirect the actuated camera toward the objects that are detected by stereo sensor nodes, we determine the pan, tilt, and zoom parameters of the actuated camera. The pan/tilt of the actuated camera is adjusted such that its center point is at the center of the detected object and the zoom parameter is set such that it includes the entire volume of the object. If we assume that the actuated camera is at (X_H, Y_H, Z_H) , an object location computed at a stereo pair is at (X, Y, Z) , and the camera points z -axis when pan and tilt angles are initialized as zero, then we can determine the pan/tilt/zoom parameters. First, the pan angle θ_p ($-\pi < \theta_{p,\min} \leq \theta_p \leq \theta_{p,\max} < \pi$) can be determined from:

$$\theta_p = \left\{ \begin{array}{ll} \tan^{-1} \frac{X-X_H}{Z-Z_H} & (Z > Z_H) \\ -\pi + \tan^{-1} \frac{X-X_H}{Z-Z_H} & (Z < Z_H, X < X_H) \\ \pi + \tan^{-1} \frac{X-X_H}{Z-Z_H} & (Z < Z_H, X > X_H) \end{array} \right\}. \quad (7)$$

Second, the tilt angle θ_t ($-(\pi/2) < \theta_{t,\min} \leq \theta_t \leq \theta_{t,\max} < (\pi/2)$) can be computed using

$$\theta_t = \tan^{-1} \frac{Y - Y_H}{\sqrt{(X - X_H)^2 + (Z - Z_H)^2}}. \quad (8)$$

Finally, if the resolution of the stereo pair sensors is $p_L \times p_L$, an object occupies $p_h \times p_w$ pixels in the stereo pairs, and the cameras in both tiers have equal angles of view (the actuated camera is zoomed-out), the maximum zoom factor of the actuated camera z is:

$$z = \frac{d_{HO}}{d_{LO}} \times \min \left(\frac{p_L}{p_h}, \frac{p_L}{p_w} \right) \quad (9)$$

where d_{HO} and d_{LO} are the distances between the actuated camera and sensor pairs to the object respectively. To accommodate for the object detection and 3-D localization error, we set the zoom factor to half the value of (9).

III. SYSTEM OVERVIEW

In this section we describe the stereo sensor pairs, the actuated camera, and the task flow in our system for object detection and high resolution image acquisition.

A. Stereo Sensor Node

Fig. 3 shows the Cyclops platform [3] that is used as the lower tier image sensing platform for this work. The main components include a camera module, a micro-controller (MCU), a complex programmable logic device (CPLD), 64 KB of external SRAM, and 512 KB of Flash. The camera can capture square images at a resolution of 128 by 128 with a field of view of approximately 43.5° vertically and horizontally. Outputs can be produced in grayscale, YCbCr, or RGB. The MCU is an ATmega128L which runs at 7.3728 MHz and 3.3V. The CPLD provides clocking for use in image capture, synchronization, and memory control. The average power consumption of Cyclops

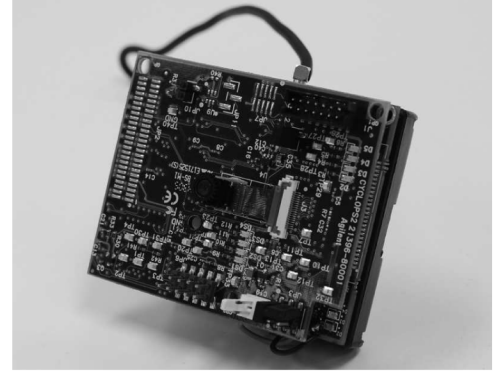


Fig. 3. Cyclops platform.



Fig. 4. Canon PTZ camera.

is approximately 40 mW [3]. Images are acquired to an intensity resolution of 8 bits for each color channel. Each Cyclops is coupled to a MICA2 mote that provides a wireless communications link at approximately 38 Kbits/s at ranges of up to approximately 1000 feet [18]. More detailed descriptions of the Cyclops platform and its use in sensor networks are provided in [3], [19]. While strictly speaking the Cyclops is distinct from the MICA2 mote, in the present paper we will use the term ‘‘Cyclops platform’’ to refer to the combination.

B. Actuated Platform

The actuated node is a Canon pan/tilt/zoom-capable (PTZ) camera (VC-C4R) which provides color images at a resolution 720 by 480 (Fig. 4). Its pan and tilt range are -170° to 170° and -90° to 10° , respectively, and its zoom range is $16\times$. The power consumption of Canon camera is approximately 12 W, which is over two orders of magnitude larger than the power consumption of the Cyclops.

C. Task Flow

Fig. 5 shows a high level diagram of the task flow assuming a system with two Cyclops platforms and a single upper tier high resolution imaging platform with actuation capability. The first step is camera calibration, which is described in more detail in Section II-A. The system then enters the location-finding loop. To minimize energy consumption, for object detection only one of the Cyclops platforms is on and the other Cyclops platform and the upper tier imaging platform are in a hibernation or sleep state. Simple frame differencing is used to determine the presence of an object, and when an object is detected, the second Cyclops platform is woken, an image pair is acquired, and a processor on one of the Cyclops platforms is used to compute the 3-D location and the associated pan, tilt, and zoom settings for actuation of the upper tier platform. The pan, tilt, and zoom

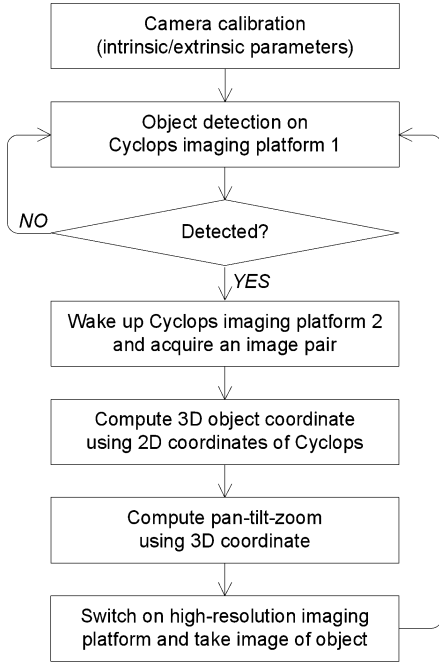


Fig. 5. Task flow of object localization.

information is then transmitted to the upper tier platform, which is then actuated accordingly and used to acquire a high resolution image of the object. Once the object disappears, the second Cyclops imaging platform and the high resolution imaging platform are placed again in hibernation mode (or powered off).

IV. DESIGN CONSIDERATIONS

This section considers some of the issues that are involved in the design of the proposed heterogeneous system. In particular, we look at the coverage of each tier in our system and its impact on the number of stereo pairs and high resolution image acquisition performance. We also investigate detection and image acquisition latency of the proposed system and compare it with the latency of a single-tiered system. Furthermore, we study the energy performance of the system.

A. Determination of the Number of Stereo Pairs

To determine the coverage area of each tier and the required density of the stereo pairs, it is important to take the application requirements for each tier into account. For instance, in the context of human detection and face identification, the required image resolution for each tier is different. In this case, the stereo pairs detect and localize human objects whereas the high resolution actuated camera acquires high resolution images of faces for use in an image identification routine. We assume the required image resolution and minimum size of the objects are known and we find the coverage area for each tier and the upper bound for the number of stereo pairs.

Assume that the minimum size of an object is $l \times l$ m², and that the number of pixels required in each tier is p_{th} and p_{hr} for stereo pairs and the actuated camera, respectively. Also, assume the image resolution for stereo pairs and actuated camera is $p_L \times p_L$ and $p_U \times p_U$, respectively. Furthermore, denote the field of view of the stereo pairs by α_L (radians) and the actuated camera

by $\alpha_{U_s}(z)$ where z is the zoom factor. The effective field of view of actuated camera is $\alpha_U = \min(\alpha_{U_s}(1) + \alpha_{U_{pan}}, 2\pi)$ where $\alpha_{U_{pan}}$ is the panning range and $\alpha_{U_s}(1)$ is the field of view when the actuated camera is fully zoomed out. If we assume $\alpha_{U_s}(z)$ can be approximated as $\alpha_{U_s}(1)/z$ (using $\tan(x) \approx x$ for $x \ll 1$), then the maximum radiuses of the area covered by stereo pairs and actuated camera are

$$r_L \approx \frac{p_L l}{p_{th} \alpha_L} \quad (10)$$

$$r_U \approx \frac{z_m p_U l}{p_{hr} \alpha_U(1)} \quad (11)$$

where z_m is the maximum zoom-in factor. The area covered by each platform is approximately $(\alpha_L r_L^2/2)$ and $(\alpha_U r_U^2/2)$, respectively. The upper bound on the number of stereo pairs needed to fill the coverage area of one actuated camera is

$$n_{Lm} = \left\lceil \frac{\alpha_U r_U^2}{\alpha_L r_L^2} \right\rceil. \quad (12)$$

For example, in the algorithm using a stereo pair performing human face detection in [20], r_L is around 2.1 m with $p_L = 128$, $l = 0.3$ m, $p_{th} = 24$, and $\alpha_L = 43.5^\circ$. Assuming that an image resolution for a face identification problem is 240×240 , r_U is around 12.6 m with $z_m = 16$, $p_U = 480$, $\alpha_U(1) = 43.5^\circ$, and $\alpha_U = 180^\circ$. In this case, $n_{Lm} = 150$. Smaller numbers of n_{Lm} can also be used at the cost of introducing more latency, but the overall detection performance will not be significantly reduced. Simulation results in Section V-C show that the use of around 20% of n_{Lm} provides detection success rate close to 100%.

B. Latency

Another concern in evaluating the proposed system is the latency in object detection and acquisition of high resolution images compared to a single-tiered system. Suppose that in a single-tiered system, the actuated camera can perform detection in the covered area in the fully zoomed-out mode ($(p_{hr}/z_m) \geq p_{th}$). Then a single-tiered actuated camera can cover the area by well-scheduled camera actuation, zooming in, and image acquisition only if an object is detected. For simplicity, we consider only panning here. The number of actuated positions (e.g., the number of times the camera needs to stop and acquire an image) to cover the coverage area is $n_s = \lceil (\alpha_U/\alpha_{U_s}(1)) \rceil$. If we assume that the overall latency consists of panning for covering the area, image acquisition, zooming-in, and zoomed-in image acquisition, and that objects randomly appear in the covered area of the system, then the overall expected latency of the single-tiered system is:

$$E(n_O(\alpha_{U_s}(1)/\omega_p + t_{ig}) + t_z(z_O)) + t_{ig} \quad (13)$$

where n_O is the number of stops from object appearance to detection, ω_p is the angular panning speed of the camera, and t_{ig} is the image acquisition and processing time. $t_z(\cdot)$ is the zooming delay of the actuated camera, and z_O is the zoom factor required to acquire a high resolution image of the object. Note that this is the best case in a single-tiered system and if the actuated camera can not cover the entire zoomed-in area while it is in zoomed-out state ($(p_{hr}/z_m) < p_{th}$), then there are additional steps needed

to cover the entire area depending on the zooming ratio required in the detection mode.

In contrast, for the two-tiered network described here, the latency of the system is due to stereo pair image acquisition and object detection, activation of the actuated camera, panning, zooming-in, and finally acquiring high resolution. In this case, the expected latency of the system is:

$$t_d + t_w + E(\max(\alpha_O/\omega_p, t_z(z_O))) + t_{ig} \quad (14)$$

where t_d denotes the object detection time of the stereo pairs, t_w is the wake-up time of the actuated camera (including stabilization and initialization), and α_O is the panning angle to the object.

For example, when a single-tiered network with the actuated camera utilized in this paper is used with the condition of $(p_{hr}/z_m) \geq p_{th}$, the average latency from object appearance to image acquisition with $\alpha_U = 360^\circ$ is around 11.8 s. In contrast, the average latency of the two-tiered network is around 5.5 s. The advantage of latency is mainly due to large α_U because the time consumed scanning the coverage area in the single-tiered network becomes dominant for large α_U . This difference can be larger if $(p_{hr}/z_m) < p_{th}$.

C. Power

Another important aspect of the system is the energy consumption of a single-tiered system compared to that of the two-tiered system. Intuitively speaking, the two-tiered system is more energy efficient in environments where the objects appear sporadically. Thus, we place the actuated camera in the hibernate mode when it is not needed. We denote the power consumption of the actuated camera and stereo pairs by P_U and P_L respectively, and the number of stereo pairs by n_L . The time from wake-up to shutdown of the actuated camera is t_{Uon} and the average time between object appearances is t_{app} . The power consumption ratio of the two-tiered network to a single-tiered network with the actuated camera is $n_L P_L / P_U + t_{Uon} / t_{app}$, and is generally much less than one. When the platforms described above are used, $t_{Uon} = 4$ sec, $t_{app} = 60$ s, and $n_L = 10$, the ratio is around 0.27. This advantage is obtained at the expense of latency. When the actuated camera is always on, one can reduce latency at the cost of power due to n_L stereo pairs. For $n_L = 10$, the increase of power consumption is around 20%.

V. RESULTS

In this section, we examine the performance of the stereo pairs in terms of localization and energy consumption and evaluate the improvements due to the use of custom precision arithmetic described in Section II-A. We then study detection performance and energy consumption of a two-tier system comprising one actuated camera and two stereo pairs through real hardware experimentation. Finally, we explore the characteristics of a larger scale system consisting of one actuated camera and multiple stereo pairs through simulation to show the tradeoffs in detection performance, latency and system power consumption.

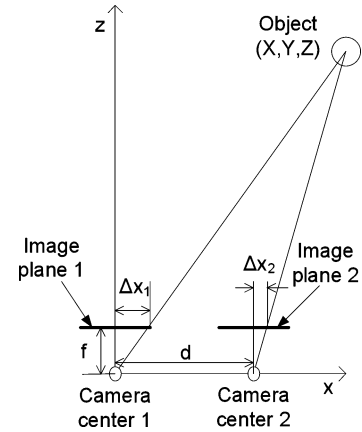


Fig. 6. Camera setup in simulation.

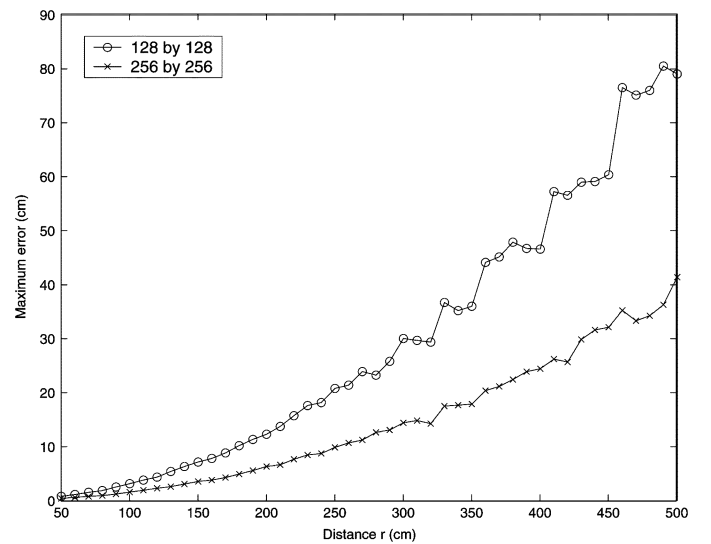


Fig. 7. Maximum object localization error as a function of object distance at different image resolutions with camera separation of 20 cm.

A. Simulation of Stereo Pair

We first consider the localization performance under simulation assuming that K , the intrinsic camera information, is known and that two cameras with focal length f are positioned on the x -axis, facing the z -axis, and separated by distance d as shown in Fig. 6. The image plane for each camera is positioned at z position f (corresponding to the focal length). The image of an object will appear in the image plane at positions (shown as Δx_1 and Δx_2 in the figure) corresponding to the point of intersection of the image planes with the lines connecting the camera centers to the object. Under this scenario, the primary source of localization error is the granularity imposed by the finite resolution of the digital image in combination with the parallax due to camera separation and positioning.

Fig. 7 shows the maximum localization error as a function of object distance r from camera 1 when the camera separation is 20 cm. The term “maximum” is used because for each distance value r , the errors associated with all possible object positions on a grid placed on a sphere centered on camera 1, with a radius r , and within the field of view of both cameras were computed. For each point on the sphere, the magnitude of the

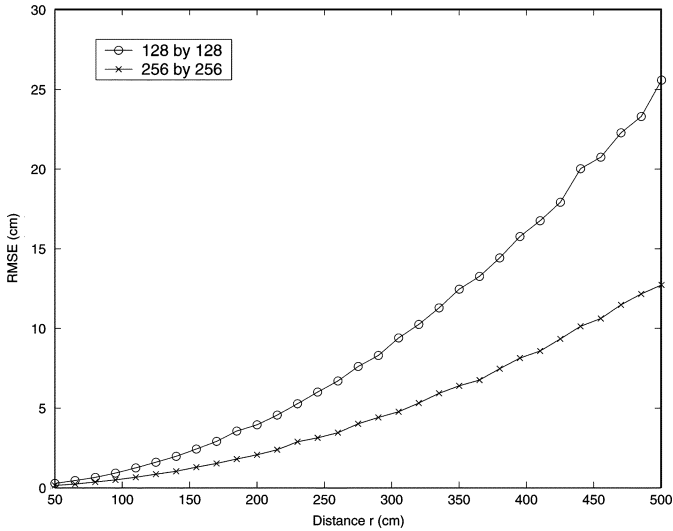


Fig. 8. RMSE of object localization as a function of object distance at different image resolutions with camera separation of 20 cm.

3-D error vector was computed. Finally, the maximum across all distance errors for that radius was selected for inclusion at the corresponding r position in Fig. 7. Two curves are shown in the figure, corresponding to image resolutions of 128×128 and 256×256 . The step-like increases evident in the figure, particularly for 128×128 resolution, arise due to the resolution granularity and can be explained with the aid of Fig. 6. Consider an object located at (X, Y, Z) which maps to x coordinates Δx_1 and Δx_2 . Then, using similar triangles:

$$d = \frac{Z}{f}(\Delta x_1 - \Delta x_2) \Rightarrow Z = \frac{df}{\Delta x_1 - \Delta x_2} \quad (15)$$

where d is the camera separation and f is the focal length. Because of the granularity of a digital image, $\Delta x_1 - \Delta x_2$ must be a multiple of a value related to pixel width, hence the possible Z values are

$$Z = c \frac{d}{n} \quad (16)$$

where constant c is a function of f and the image resolution and n is a positive integer. In the case of Cyclops platforms, $c = 161$ for a 128 by 128 image. The steplike increases in Fig. 7 at after $r = 400$ cm for the 128×128 curve are due to the change in n in (16) and occur when the image of the object in the two image planes is separated by a distance of under n pixel widths, and when in the worst case, the object appears at the left edge of one pixel in one image plane and the right edge of other pixel in the other image plane. The system will associate object locations with the centers of the pixels in which they appear, leading to a bigger increase as n decreases in the distance measurement. Clearly, using higher resolution images significantly reduces distance error so the effect of pixel granularity is particularly evident in the curve for 128×128 resolution, as shown in Fig. 7.

Fig. 8 illustrates root mean square error (RMSE) of localization as a function of object distance r from camera 1 when the camera separation is 20 cm with two different image resolutions. The RMSE values are significantly better than the worst case er-

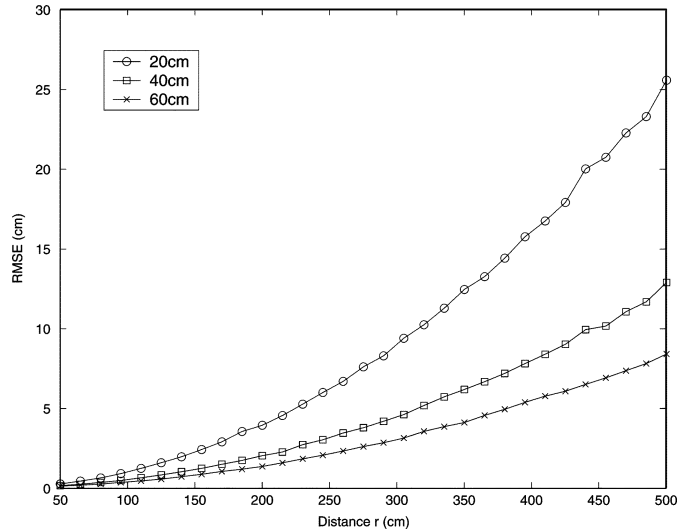


Fig. 9. RMSE of object localization as a function of object distance with different camera separations using 128×128 image.

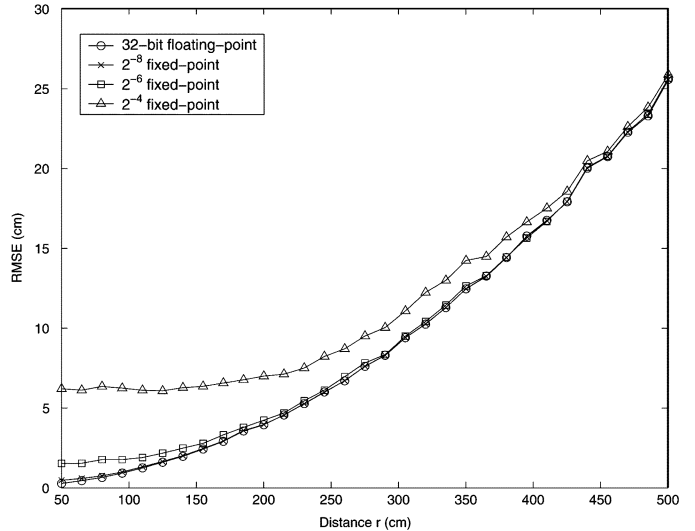


Fig. 10. RMSE of object localization as a function of object distance with different precisions using 128×128 image.

rors shown on Fig. 7 with some cases up to four times improvement in localization accuracy.

Additionally, we examined the impact of localization accuracy versus distance of the stereo pairs. Fig. 9 illustrates the RMSE of localization when the image resolution is fixed at 128×128 and camera separations of 20, 40, and 60 cm are used. In this case, the error is approximately inversely proportional to camera separation as expected from (16). However, the increased camera separation requires more careful installation and alignment, and bad alignment of stereo pair can reduce their common field of view.

Applying the optimization methodology described in Section II.A to the stereo pair, we can exploit the trade-off between localization accuracy and energy consumption. Fig. 10 shows the RMSE of object localization with different precisions when image resolution is 128×128 . This plot confirms the appropriateness of using 2^{-8} fixed-point arithmetic, which shows negligible accuracy degradation compared to 32-bit floating-point

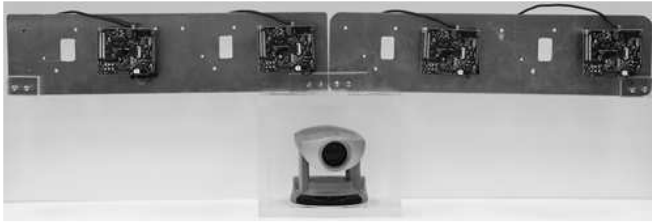


Fig. 11. Implementation of a two-tiered system using four Cyclops platforms and one high-resolution Canon PTZ camera.

case. The use of 2^{-4} would give additional energy reduction but the loss of localization performance is quite big when r is small. As with Fig. 10, error due to data truncation is dominant when r is small, otherwise error due to the pixel granularity is dominant.

Utilizing 2^{-8} fixed-point computation, the 3-D location computations consume 2232 clock cycles when camera 1 is used and 4404 cycles when camera 1 is not used. These translate to energy consumptions of $7.0 \mu\text{J}$ and $13.7 \mu\text{J}$, respectively. By contrast, if floating-point emulation is utilized, the clock cycle requirements become 9509 and 18671 cycles respectively, corresponding to $29.7 \mu\text{J}$ and $58.2 \mu\text{J}$, respectively. Hence, the custom precision methodology reduces the execution time and Cyclops platform energy consumption by approximately a factor of 4. The other operations such as computation of actuation parameters which are not treated in this paper can be optimized in a similar manner which will also lead to similar energy reductions.

B. Experimental Results

Fig. 11 depicts the experimental system comprising one Canon PTZ camera and four Cyclops platforms mounted on a metal structure at 20 cm separations. Internal and external parameters of Cyclops are determined by camera calibration (Section II-A). The system is designed to enable measurements from different camera pairs and to enable collaboration using more than two cameras. For object detection, simple frame differencing is used with the midpoint of the detected object chosen as 2-D image coordinate. To perform controlled experimentation with varying sized objects in different speeds, objects are generated using a 50-in plasma screen.

Next, a set of tests was performed using a pair of Cyclops platforms and the Canon PTZ camera. The screen displaying the objects was placed 120 cm from the plane containing the imaging platforms. We assume that human objects appear from three to seven meters from the Cyclops platforms with a walking speed from 0.8 m/s to 1.2 m/s. When displayed on the screen, the size and speed of the objects are appropriately scaled with the consideration of the numbers above. Additionally, we assume that only one object appears on the screen at a time.

Given the goal of obtaining high resolution images of as many of the objects as possible, the key issues are 1) the energy costs among various single-tiered and two-tiered approaches, and 2) performance differences among the approaches in terms of success in capturing high resolution images. From overall system energy standpoint, almost any reasonable two-tier approach based on selectively powering on the Canon PTZ

camera will be more efficient than a single-tiered approach because the power used by the Canon PTZ camera (12 W) exceeds the Cyclops power (40 mW) by such a large margin. However, these energy savings come at a cost in performance due to the time delays associated with wake-up of the actuated camera and, to a lesser extent, to the process of distributing the detection/imaging processes across multiple tiers.

In a single-tiered system consisting of only Canon PTZ cameras, frame grabbing requires 300 ms, object recognition can be performed in 100 ms [6], and actuation to enable the camera to zoom in and acquire a high resolution image takes approximately 300 ms (assuming a $90^\circ/\text{second}$ pan rate and an average pan angle of one half the effective field of view angle). Thus, the total time from object appearance to acquisition of the high resolution image is approximately 700 ms.

In the two-tier system there are more steps to be performed and the time requirements are correspondingly longer. In addition, the performance of the two-tier system depends on how the resources in the different tiers are utilized. From a detection performance standpoint, the most effective approach is to have one Cyclops maintained permanently in the on state. Measurements indicate that the “always on” Cyclops requires approximately one second to perform object detection for 128×128 resolution and to awake the second Cyclops and the actuated camera. The second Cyclops then uses approximately one more second to do its own object detection, compute the 3-D position, and transmit the appropriate pan/tilt/zoom instructions to the actuated camera. The actuated camera requires approximately 370 ms to wake up, 920 ms to stabilize, 1.3 s to initialize [6], and 300 ms for actuation as noted above. With the exception of the final actuation step, all of these steps can be performed without knowledge of the specific object location, so the actuated camera can be initiated at the same time the second Cyclops is woken. Thus, the total time from object appearance to high resolution image capture is around 3.9 s.

Table III gives the results for the cases of a single-tiered system using a single Canon PTZ camera which is always on with an object speed of 1.2 m/s, and a two-tier system based on an “always on” Cyclops that detects objects and then wakes a second Cyclops and the Canon PTZ camera with object speeds of 0.8 m/s, 1.0 m/s, and 1.2 m/s. Two different rates are considered, corresponding respectively to average periods between object appearance of 1 minute and 2 minutes. The table gives the percentage of objects that are successfully imaged using the actuated camera, which means that the height of an object is more than half of the height of the image resolution (240 pixels in our experiments because the resolution of the actuated camera is 640×480). The table also provides the average power for the two object appearance rates.

As expected, Table III confirms that two-tier operation requires significantly less average power than single-tiered operation. The average power of the two-tiered system mainly relies on the object appearance rate since it determines how often the Canon PTZ camera is turned on and at a rate of 2 objects/min power reduction is approximately a factor of seven. Notably, the two-tiered systems detected more than three quarters of the objects when the scaled speed is less than or equal to 1 m/s. The success rate mainly depends on object dwell times on the screen

TABLE III
SUCCESS RATE OF OBJECT IMAGE CAPTURE AND AVERAGE POWER COMPARISON BETWEEN A SINGLE-TIERED AND A TWO-TIERED SYSTEM. FOR THE SINGLE-TIERED SYSTEM, OBJECT SPEED IS 1.2 M/S (ROW 1). FOR THE TWO-TIERED SYSTEM, OBJECT SPEEDS ARE 1.2 M/S (ROW 2), 1.0 M/S (ROW 3), OR 0.8 M/S (ROW 4)

	System	Object speed (m/s)	1 object/min		2 objects/min	
			Success (%)	Average power(W)	Success (%)	Average power(W)
1	Single-tier	1.2	100	12	100	12
2	Two-tier	1.2	35.0	0.82	36.7	1.62
3		1.0	80.0	0.82	77.5	1.62
4		0.8	98.3	0.82	97.5	1.62

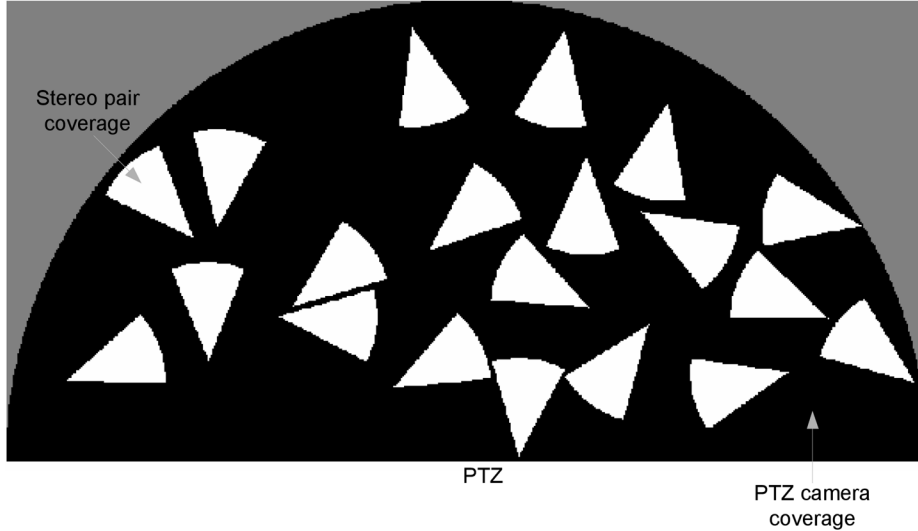


Fig. 12. Example of coverage of a single actuated camera (black) and 20 stereo pairs (white).

which is determined by object speed. This problem can be alleviated by reducing wake-up delay of the actuated camera.

C. Simulation of the Case of Multiple Stereo Pairs

Generally, the coverage of the actuated camera is larger than that of the stereo pair due to mobility and zooming, so multiple stereo pairs are required to cover the surveillance area of the actuated camera. We explore the tradeoffs in detection success rate, latency, and power consumption in this scenario.

Fig. 12 show an example of coverage of the actuated camera and multiple stereo pairs. In this simulation, we assume that object size is 30 cm by 30 cm, object detection threshold is 4 pixels, and image resolution and the field of view of stereo pairs are 128×128 and 43.5° , respectively. If an RMSE in 3-D localization equivalent to the object size is allowed, the coverage radius of the stereo pair is around 5.4 m. For the actuated camera, the image resolution is 640×480 , the field of view is 180° and the maximum zoom factor is 16, which gives a coverage radius of 25.3 m. Locations of stereo pairs are randomly chosen with two restrictions that 1) the coverage of each stereo pair is within that of the actuated camera and 2) there is no overlap among the coverage of stereo pairs. Fig. 12 shows the deployment example of 20 stereo pairs.

Fig. 13 shows success rates of object detection and high resolution image acquisition with various numbers of stereo pairs in the lower tier. In this simulation, we assume that objects appear

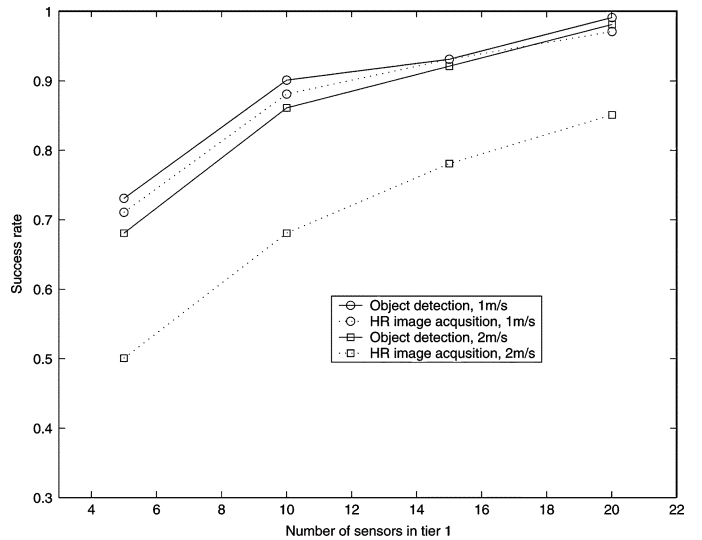


Fig. 13. Success rates of object detection and high resolution image acquisition with various numbers of stereo pairs in the lower tier.

from a random point on the left or right edge, proceed along a straight line, and disappear on the opposite edge of the coverage area. In this case, the rate of successful detection of objects rapidly improves with a larger number of stereo pairs. For instance, with five stereo pairs 70% of the objects are detected

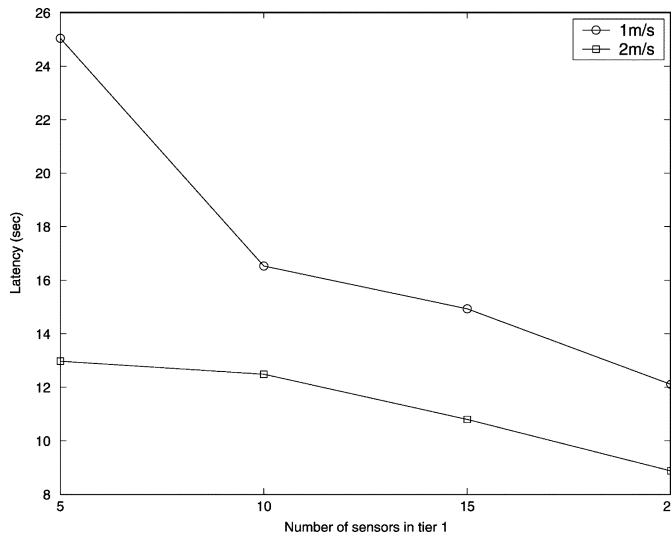


Fig. 14. Latency of high resolution image acquisition with various numbers of stereo pairs in the lower tier.

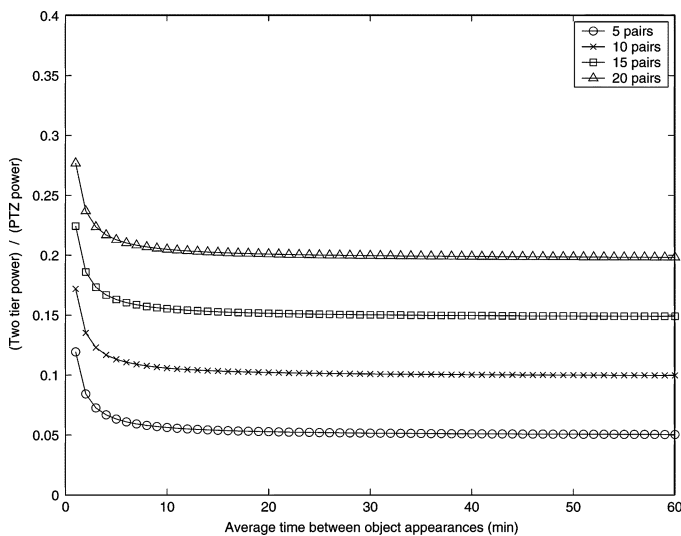


Fig. 15. System power consumption with various numbers of stereo pairs in the lower tier and object appearance rate.

and with 20 stereo pairs (which only covers 20% of the actuated camera coverage area) nearly all objects are detected. We attribute this to the fact that even with a small number of stereo pairs that partially cover the actuated camera area, there are few linear trajectories through the area that avoid all the stereo pairs. We also note that while the success rate of high resolution image acquisition is close to that of object detection when object speed is 1 m/s, it of course degrades with an increase in the speed of the objects due to the delay in wake-up of the actuated camera. This can be improved by utilizing a camera with reduced wake-up delay time.

Fig. 14 presents the latency of high resolution image acquisition with varying number of stereo pairs in the lower tier. The latency is defined as the time from the point that an object enters the coverage area of the actuated camera to the point that a high resolution image is acquired of the object.

Fig. 15 demonstrates the relative power consumption of a two-tier system (including both actuated and stereo nodes) compared to that of a single-tiered system with varying numbers of stereo pairs and object appearance rates. Interestingly, when the objects appear at 1 min intervals, the two-tiered system with 20 stereo pairs (with almost 100% detection rate as noted previously) provides power consumption reduction of approximately a factor of four. As the object appearance time increases, the gain in terms of power consumption increases and the curve becomes almost flat, since under these conditions the lower-tier dominates the power consumption of the two-tier system.

VI. CONCLUSION

We have presented an energy-efficient object 3-D location approach based on a two-tiered hierarchical sensor network consisting of a lower-tier of resource-constrained imaging sensors using stereo imaging techniques and an upper-tier of higher resolution but more energy consumptive imagers. The 3-D location computations are performed on the lower tier sensor platforms with the aid of a custom precision design methodology that identifies the minimum bit-width representations and automatically produces C code appropriate to the native word length of the target processor. Once the 3-D location is computed, the appropriate pan/tilt/zoom commands are transmitted to an upper-tier sensor platform that then captures a high resolution image of the object. The precision-optimized computations result in an energy reduction by approximately a factor of four for the computations on the Cyclops platforms. Experimental results show that use of the two-tier network reduces power consumption both for the system as a whole and for the Cyclops platforms. The extent of power consumption reduction is a function of the object appearance statistics and the success rate, and latency of high resolution image acquisition are functions of the number of stereo sensor pairs and object speeds.

REFERENCES

- [1] F. Bashir, A. Khokhar, and D. Schonfeld, "Automatic object trajectory-based motion recognition using Gaussian mixture models," in *Proc. IEEE Int. Conf. Multimedia and Expo.*, 2005, pp. 1532–1535.
- [2] R. Rosales and S. Sclaroff, "3D trajectory recovery for tracking multiple objects and trajectory guided recognition of actions," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, 1999, pp. 117–123.
- [3] M. Rahimi, R. Baer, O. I. Iroezji, J. C. Garcia, J. Warrior, D. Estrin, and M. Srivastava, "Cyclops: In situ image sensing and interpretation in wireless sensor networks," in *Proc. 3rd Int. Conf. Embedded Networked Sensor Systems*, 2005, pp. 192–204.
- [4] A. Kansal, W. Kaiser, G. Pottie, M. Srivastava, and G. Sukhatme, "Virtual high-resolution for sensor networks," in *Proc. 4th Int. Conf. Embedded Networked Sensor Systems*, 2006, pp. 43–56.
- [5] S. Hengstler, D. Prashanth, S. Fong, and H. Aghajan, "MeshEye: A hybrid-resolution smart camera mote for applications in distributed intelligent surveillance," in *Proc. 6th Int. Conf. Information Processing in Sensor Networks*, 2007, pp. 360–369.
- [6] P. Kulkarni, D. Ganesan, P. Shenoy, and Q. Lu, "SensEye: A multi-tier camera sensor network," in *Proc. 13th Annu. ACM Int. Conf. Multimedia*, 2005, pp. 229–238.
- [7] W. Wolf, B. Ozer, and T. Lv, "Architectures for distributed smart cameras," in *Proc. Int. Conf. Multimedia and Expo*, 2003, vol. 2, pp. II-5–II-8.
- [8] W. Wolf, B. Ozer, and T. Lv, "Smart cameras as embedded systems," *IEEE Computer*, vol. 35, no. 9, pp. 48–53, Sept. 2002.
- [9] D. Lymberopoulos, A. Barton-Sweeny, and A. Savvides, *Sensor Localization and Camera Calibration Using Low Power Cameras*, ENALAB Tech. Rep. 090105, Sept. 2005.

- [10] S. Funiak, C. Guestrin, M. Paskin, and R. Sukthankar, "Distributed localization of networked cameras," in *Proc. 5th Int. Conf. Information Processing in Sensor Networks*, 2006, pp. 34–42.
- [11] D. Devarajan and R. J. Radke, "Distributed metric calibration of large camera networks," in *Proc. 1st Workshop on Broadband Advanced Sensor Networks*, 2004 [Online]. Available: <http://www.broad-nets.org/2004/basenets.html>
- [12] P. Dutta, M. Grimmer, A. Arora, S. Bibyk, and D. Culler, "Design of a wireless sensor network platform for detecting rare, random, and ephemeral events," in *Proc. 4th Int. Symp. Information Processing in Sensor Networks*, 2005, pp. 497–502.
- [13] A. Y. Benbasat and J. A. Paradiso, "Groggy wakeup-automated generation of power—Efficient detection hierarchies for wearable sensors," in *Proc. 4th Int. Workshop on Wearable and Implantable Body Sensor Networks*, 2007, pp. 59–64.
- [14] R. I. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2003.
- [15] D. Lee, H. Kim, S. Tu, M. Rahimi, D. Estrin, and J. D. Villasenor, "Energy-optimized image communication on resource-constrained sensor platforms," in *Proc. 6th Int. Conf. Information Processing in Sensor Networks*, 2007, pp. 216–225.
- [16] J.-Y. Bouguet, Camera Calibration Toolbox for Matlab 2006 [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc
- [17] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*. Upper Saddle River, NJ: Prentice-Hall, 2002.
- [18] MICA2 Datasheet Document Part Number: 6020-0042-07 Rev A 2000 [Online]. Available: http://www.xbow.com/Products/Product_pdf_files/Wireless_pdf/MICA2_Data sheet.pdf
- [19] M. Rahimi, S. Ahmadian, D. Zats, R. Baer, D. Estrin, M. Srivastava, and J. Warrior, "Demo abstract: Network of cyclops; Image inference and interpretation in sensor network," in *Proc. 3rd Int. Conf. Embedded Networked Sensor Systems*, 2005, pp. 310–310.
- [20] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004.



Hyungjin Kim (S'06) received the B.S. and M.S. degrees in electrical engineering from Seoul National University, Korea, in 1997 and 1999, respectively. He is currently a Ph.D. student in the Electrical Engineering Department, University of California, Los Angeles (UCLA). His research interests include MIMO communications, channel coding, sensor networks, compression and encryption algorithms, video/image processing and signal processing.



Mohammad Rahimi (S'04–M'06) received the M.S. degree in electrical engineering and the Ph.D. degree in computer science, both from the University of Southern California (USC), Los Angeles.

He is a Research Scientist at the Center for Embedded Networked Sensing (CENS), University of California, Los Angeles (UCLA). His research interests include wireless sensor networks, distributed camera networks and emerging applications of mobile phone technology. He has had multiple research collaboration with industrial partners to

make commercial and research products.



Dong-U Lee (A'07–M'07) received the B.Eng. degree in information systems engineering and the Ph.D. degree in computing, both from Imperial College London, U.K., in 2001 and 2004, respectively.

From 2005 to 2007, he was a postdoctoral researcher at the Electrical Engineering Department, University of California, Los Angeles (UCLA), where he developed high-performance hardware designs for wireless communications and mathematical function evaluations. He is now a Research Scientist at Mojix, Inc., Los Angeles, CA, where

he is specializing in hardware implementation aspects of RFID receivers. His research interests include computer arithmetic, communications, design automation, reconfigurable computing, and video image processing.



Deborah Estrin (S'78–M'80–SM'95–F'04) received the Ph.D. degree from the Massachusetts Institute of Technology, Cambridge.

She is a Professor of computer science, holds the Jon Postel Chair in Computer Networks, and is Founding Director of the National Science Foundation funded Center for Embedded Networked Sensing (CENS), University of California, Los Angeles. CENS' mission is to explore and develop innovative, end-to-end, distributed sensing systems, from ecosystems to human systems. Her earlier work

addressed Internet protocol design and scaling, in particular, inter-domain and multicast routing. Since the late 1990s her work has focused on multidisciplinary, experimental-systems research as applied to a variety of environmental monitoring challenges. Most recently this work includes participatory-sensing systems, at the personal and community level, leveraging the location, acoustic, image, and attached-sensor data streams increasingly available from mobile phones.

Dr. Estrin chaired a 1998 DARPA/ISAT study on sensor networks and a 2001 NRC study on Networked Embedded Computing which produced the report *Embedded Everywhere*. She served as a founding member of the National Ecological Observatory Network (NEON) Advisory board, and is currently a member of the NRC Computer Science and Telecommunications Board (CSTB), and TTI/Vanguard. Estrin was selected as the first ACM-W Athena Lecturer in 2006, was awarded the Anita Borg Institute's Women of Vision Award for Innovation in 2007, and was inducted as a member of the American Academy of Arts and Sciences in 2007.



John D. Villasenor (S'89–M'89–SM'97) received the B.S. degree in 1985 from the University of Virginia, Charlottesville, the M.S. and Ph.D. degrees in 1986 and 1989, respectively, from Stanford University, Stanford, CA, all in electrical engineering.

From 1990 to 1992, he was with the Radar Science and Engineering section of the Jet Propulsion Laboratory, Pasadena, CA, where he developed methods for imaging the earth from space. He joined the Electrical Engineering Department at the University of California, Los Angeles (UCLA) in 1992, and is currently

Professor. He served as Vice Chair of the Department from 1996 to 2002. At UCLA, his research efforts lie in communications, computing, imaging and video compression, and networking.