WeebleVideo – Wide Angle Field-of-View Video Sensor Networks

Pushkar Sambhoos, Ahmad Bilal Hasan, Richard Han, Tom Lookabaugh, Jane Mulligan
Department of Computer Science, University of Colorado at Boulder
Contact authors: {sambhoos,ahmad.hasan,richard.han}@colorado.edu

ABSTRACT
Low cost networks of wireless sensors can be distributed to provide information about an environment. Even a network of sensors providing scalar measurements (for instance, of temperature) presents both formidable challenges in terms of integrating and interpreting measurements over space and time, and important opportunities in extended observations. Cameras are particularly powerful multi-dimensional sensors for dispersing in unknown environments for surveillance and tracking of activity. Understanding the spatial patterns of such activity requires the camera network to self-organize in terms of understanding relative positions of nodes. Cameras also pose problems for resource limited motes because of the high volumes of image data for local processing or transmission. We describe a self-righting or *weeble* node architecture for camera networks based on integrating a low cost camera into the Mica2 sensor node platform. The node uses a wide field of view lens (typically called a fish eye lens) which allows us to capture a very broad region around the node providing greater view overlap between the nodes and generally a larger frame for identifying and tracking activity.

Categories and Subject Descriptors
C.3 [**Special-Purpose and Application-Based Systems**] Real-time and embedded systems

General Terms
Algorithms, Design, Experimentation.

Keywords
sensor network, distributed cameras.

1. INTRODUCTION
Sensor networks have been an expanding field of research in the recent years. Adding the power of vision to nodes in wireless sensor networks (WSNs) provides a whole new dimension of sensing [1,2,3]. A wide array of applications can be built on a platform that offers the capability of vision on a sensor node. The objective of this paper is to introduce such a platform, which builds on the existing sensor node architecture offered by Mica-class mote platforms. We seek to adhere to the low cost vision of WSNs by designing an inexpensive node with a wide field of view lens to sense and monitor a wide area at minimal expense. Omnidirectional Vision techniques using wide angle lenses or hemispherical mirror cameras have proven effective in surveillance and localization tasks [4,5]. Our vision is that a large array of inexpensive wide angle video sensors can be scattered across terrain, can be built to be impact-resistant using inexpensive off-the-shelf technology, can self-correct for arbitrary orientation, and can effectively monitor a large surrounding neighborhood. We term this project WeebleVideo.

Aside from direct application to visual monitoring applications, the WeebleVideo project also presents a set of interesting research problems that could assist in advancing the field of sensor networking in general. The overall context is one of very cost- and size-limited nodes; we don't want to rely on Moore's law to rescue complex designs since we are at least as much interested in scaling designs to smaller nodes as in holding node size constant and scaling to more complexity. Interesting problems include:

Usefully Fusing Image Data: The nodes provide a collection of highly distorted images from different locations. We need to combine this data in a useful way. For example, we may want to perform Image-based Rendering [6,1], picking an arbitrary virtual location and orientation and fuse the image data from the nodes to form an estimate of an undistorted image we would see from that location There are issues of geometry, combining images to reduce noise, dealing with obstruction, and a variety of other image processing challenges. Moreover, we expect the data to be of limited resolution, low update rate, and noisy, so we need unusually robust fusion techniques.

Self- and Mutual-Location: The ability to make sense of the data will depend strongly on an accurate estimate of the relative location and orientation of nodes [7,8,1]. If we imagine that nodes are scattered rather than placed, nodes will need to be able to self-locate and mutually locate with high accuracy. We also require a reliable efficient mechanism for determining orientation in order to correct for tilt in the wide angle lense.

Severely Constrained Signal Processing: WSNs have very limited communication bandwidth, limited processing power, and limited memory available. We need to optimize node image processing, image compression, and communication operations carefully. For example, we would want to consider combining operations to reduce complexity, such as selective encryption in which only parts of a compressed image format are encrypted and the compression protocol itself provides remaining necessary privacy.

Wireless Multimedia Networking: Wireless node networks have to be very simple and currently do not embrace the capabilities of a typical LAN. Given the very limited communication bandwidth we envision, we will want to add the ability for nodes with particularly interesting local activity to preemptively secure bandwidth back to a base processing station. Adding limited quality-of-service style semantics to wireless node networks is intrinsically challenging.

Sensor Integration: Aside from the problem of integrating camera and image processing with the computing node, we can quickly envision interesting multi-sensor integrations. For example, we may want to integrate an accelerometer to detect motion of the sensor (for example, if it is accidentally or intentionally moved). Combining multiple sensors at a node brings up the problem of partitioning sensor fusion between the low complexity node and a higher capability but remote (across a limited bandwidth communication path) base station.

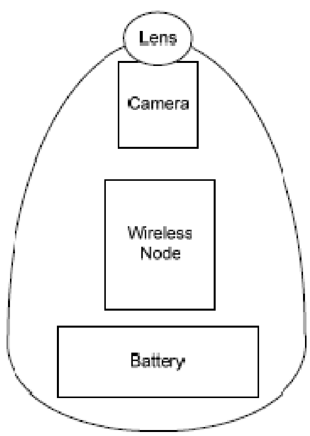

Figure 1 WeebleVideo Platform's Conceptual Design

The conceptual view of the WeebleVideo platform is shown in Figure 1. The idea behind this platform design is that the heavy battery in the base and the shape of the exterior causes the node to self-right on reasonably level ground much like Weebles toys. The fish-eye lens projects roughly a half sphere field of view onto the camera. We envision an inexpensive camera of the type currently used in low end disposable digital cameras and in camera-equipped cell phones.

2. SYSTEM DESIGN

As a test of feasibility, we have built the first part of the WeebleVideo platform in the form of a board that acts as an extension to the Mica2 platform [9] commonly used in sensor networks. It provides an easy way to add the capability of capturing images just by plugging a module onto the mote. The camera and the mote are interfaced using an intermediate board providing an interface to combine the two. The components of the platform are shown in Figure 2.

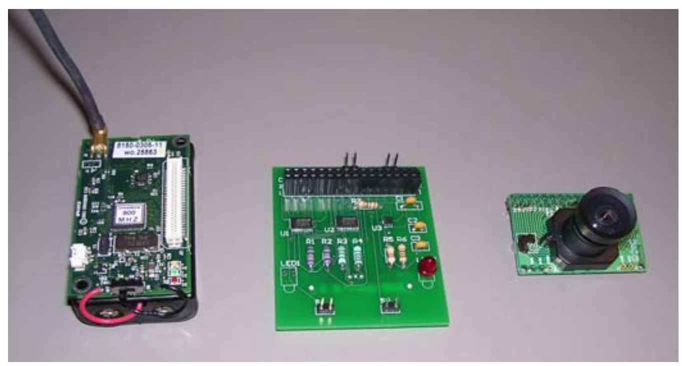

Figure 2. The WeebleVideo Components

All the modules shown in Figure 2 plug into each other to provide a compact integrated setup as shown in the Figure 3. The apparatus is powered by AA batteries and is capable of taking images at a maximum resolution of 640x480. The Mica2 platform incorporates a radio that is used to transmit images to other sensor nodes in the network. The platform uses the robust Mantis OS, which provides multi threading capability and a strong foundation for fast development of applications for the platform. The idea is to mount these devices in the form of Weebles and disperse them in the field to start capturing data. The nodes can then detect their neighbors and start monitoring the area for any activity. Many interesting questions arise over how to collect and communicate data, and to what extent it should be processed locally versus relaying the information back to the Base Station for more complex processing.

The original camera module is equipped with a lens providing a 45° FOV. A common technique for localization and surveillance in natural scenes is to use an "omnidirectional" camera which has either a hemispherical mirror or wide angle lens to allow viewing 360° around the sensor. To obtain a much wider view of the surroundings for weebles, the original lens was replaced with a fisheye lens offering a 150° FOV. The images shown in Figure 4 compare the capability of the two lenses. Both the images were taken from the same distance from the object. The fisheye offers much more information than the regular lens, but at the same time introduces distortion in the acquired image. The undistortion and processing of the acquired image poses interesting problems in the resource constrained environment with limited processing and memory capability available on the node. We are further considering improving the surrounding field of view by using three camera modules attached to the node, which increases the problems of data volume and processing.

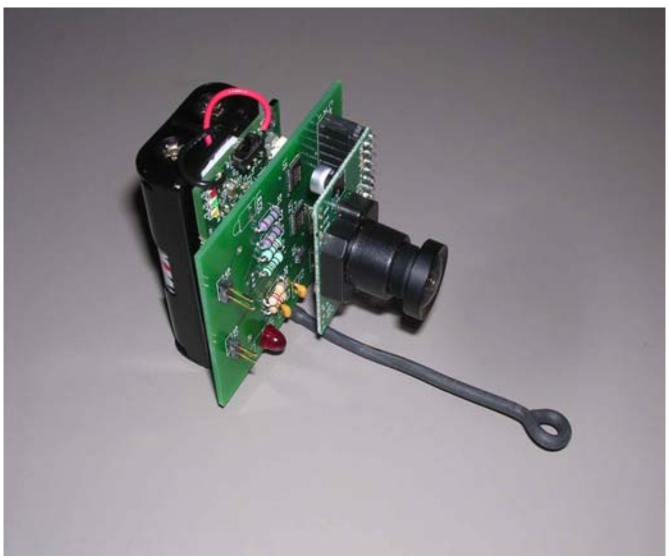

Figure 3. The WeebleVideo Board - Integrated

Figure 4. Images obtained from our camera's default narrow-angle lens (left) and the fisheye lens that we installed (right).

3. STATUS AND APPLICATIONS

The immediate future work includes a variety of directions. First, we need to integrate the camera-board-mote apparatus into the physical self-righting form factor characteristic of the Weeble. Second, we intend to implement various image processing algorithms on the controller of the Mica2 platforms. The basic algorithms that would be implemented include fusion, as mentioned earlier and in the related work, as well as differencing and background subtraction using different background modeling techniques. A prerequisite to identifying a target is simple differencing of frames. However, simple differencing will often result in false positives due to a variety phenomena, e.g. a cloud shadow moving across a field, or trees swaying back and forth. In these cases, the human eye can easily ignore "unimportant" background changes, but automated algorithms face a difficult time accomplishing the same filtering. Thus, background subtraction remains a challenging research topic in the field of computer vision, and will be made even more challenging by on-board target tracking within our resource-constrained mote environment.

On the practical side, we wish to develop the platform and provide an interface to the developer with the image capturing and processing API.

We expect to deploy the WeebleVideo platform in a variety of different in situ applications.

Wildlife Movement Tracking

Wildlife monitoring is an area where the platform has many applications for tracking movements of animals, especially elusive wildlife. The nodes would be dispersed in an area where the target is to be tracked, and would cover a large area using wide FOV lenses. Each node would relay information about particular traces via the mesh network to the base station. Going a step further, microphones can be attached to the sensor nodes and can start monitoring an area for audio as well as visual observations of the animals allowing improved multimodal sensing.

Military area Monitoring

In military applications, these nodes would be scattered in an area randomly. The nodes would start communications with each other and form a topology map to monitor the area for any movement or unusual activity. When a particular node captures interesting data, it can send the data over the network to the base station to determine whether to raise an alert. Increasing the sophistication of this scenario, the node could actually ask its neighbors to start looking for activity near itself and link its information with that of other nodes in the area. This could provide valuable information in tracking the path of a target in the FOV of the nodes.

4. RELATED WORK

Many currently available platforms, e.g. Cyclops [3], offer a limited resolution on the images, which does not provide sufficiently high resolution, e.g. 64x64 for viable tracking and localization. With support for higher resolution, the Weeble node can capture much more data yielding more precise information. Our current camera component allows capturing of a region of interest (ROI) within a high resolution capture area, allowing the node to concentrate on a particular area of the image. This capability can be harnessed in a useful way to isolate important data at high resolution while limiting data volume by treating a small area in the entire field of view. Combined with on-board processing, less image data needs to be transmitted over the network.

The lenses available on similar sensor nodes with imaging capability usually offer only a small field of view limiting the area of the surroundings captured by the node. Wide field of view or omnidirectional cameras have been common in the literature for the past decade [4,5]. We have chosen fisheye lenses rather than mirror-based omnicams, because we anticipate they will be less fragile and more robust to scatter deployment. Most of the work on omnidirectional vision has not been directed at the resource-limited mote/sensor network scenario. These applications usually make use of full blown PC hardware requiring much higher computing power than sensor nodes.

Purushottam Kulkarni et al [2] suggest the use of multi tier camera sensors with heterogeneous capabilities. The WeebleVideo platform can be used in such a hierarchy at the top level where high resolution images are needed to capture more information. The Cyclops camera and the Weebles Platform complement each other in providing two levels of the hierarchical design in the multi-tier camera Sensor Networks.

The problem of observing activity from multiple camera nodes dropped in an unknown environment has been considered for at least a decade. From a Computer Vision perspective the challenges include computing the relative pose of nodes, identifying and tracking interesting targets and understanding the meaning and importance of observed activity. Typically background modeling techniques which maintain an estimate of the statistical variation at each pixel are used to monitor change in the scene over time, and identify regions of significant change or targets. Connected component techniques compute object centroids. The work of Stauffer and Grimson [10] described what they called a “forest of sensors”, conceptually similar to the nodes we propose, to passively learn patterns of activity by observing moving targets. They emphasize the need for relative pose calibration, construction of site models, target detection, classification and behaviour modeling.

Lee et al. [11] address the problem of determining relative pose of a set of sensor nodes by using planar geometric constraints on moving targets to robustly fit their behaviour to a planar model. Assuming a common groundplane for multiple nodes allows them to recover relative node and groundplane position. More recently Mantzel [7] describes a distributed approach to bundle adjustment for omnidirectional cameras with limited view overlap.

Aslam et al. [12] describe theoretical aspects of tracking algorithms in binary sensor networks where the Vision problem is assumed solved to the point of returning a moving toward/away binary token for each node. They show that to compute the location of the object they need an additional bit of proximity information. They demonstrate their algorithms using simulation. Whether image sensors could reliably provide the required bits is unknown.

Finally understanding the nature of observed target activity and whether it is usual or unusual is an open and active area of research [13] involving many and varied approaches. Generally these systems extend beyond modeling the environment and tracking moving targets to classifying the targets, describing or generalizing their behaviour and identifying particular people, animals or vehicles.

5. CONCLUSION

We have presented a vision for video sensor networks that exploits a platform containing a camera with a wide angle fisheye lens mounted in a Weeble-like self-righting platform. Our goal is to achieve inexpensive wide visual coverage of a sensor field. We have implemented a board that interfaces between a standard commercial camera with a fisheye lens and a mica2 mote. We believe that the WeebleVideo design poses an interesting paradigm and presents many new research challenges. We have outlined some of the promising directions for further research exploration.

Many challenges remain from a Computer Vision and activity recognition point of view. Many of the proposed ``forest of sensors" or ``sea of cameras" ideas in the literature have not been demonstrated on genuine resource-limited camera nodes with attendant real world sensor noise and errors. For example under what conditions is local image processing on the node preferable to global image processing at the base station. Similarly, can noisy features from low resolution images observed at low frame rates be effectively used for localization or tracking? How can we use communication and activation effectively to generate patterns of sensor activation for timely and efficient tracking?

6. REFERENCES

[1] Camillo J. Taylor and Babak Shirmohammadi, "Self localizing smart camera networks and their applications to 3d modeling," in *Proceedings Workshop on Distributed Smart Cameras (DSC 2006)*, Boulder, CO, Oct 2006.

[2] Purushottam Kulkarni, Deepak Ganesan, and Prashant Shenoy, "The case for multi-tier camera sensor networks," in *Proceedings of the international workshop on Network and operating systems support for digital audio and video (NOSSDAV)*, June 2005.

[3] Mohammad Rahimi, Rick Baer, Obimdinachi I. Iroezi, Juan C. Garcia, Jay Warrior, Deborah Estrin, and Mani Srivastava, "Cyclops: In situ image sensing and interpretation in wireless sensor networks," in *Proceedings of the 3rd international conference on Embedded networked sensor systems*, Nov. 2005.

[4] Kostas Daniilidis, "Page of omnidirectional vision," http://www.cis.upenn.edu/ kostas/omni.html.

[5] S.K. Nayar, "Catadioptric Omnidirectional Camera," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 1997, pp. 482–488.

[6] P. J. Narayanan and Takeo Kanade, "Virtual worlds using computer vision," in *Proc. of IEEE/ATR Workshop on Comp. Vis. for VR based Human Comm. (CVVRHC'98)*, Bombay, India, January 1998, pp. 2–13.

[7] William E. Mantzel, "Linear distributed localization of omni-directional camera networks," in *Proc. the 6th Workshop on Omnidirectional Vision, Camera Networks and non-classical Cameras*, 2005.

[8] Richard Hartley and Andrew Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2 edition, 2003.

[9] Crossbow Technology, ," http://www.xbow.com.

[10] Chris Stauffer amd W. Eric L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 22, no. 8, pp. 747–757, Aug. 2000.

[11] Lily Lee, Raquel Romano, and Gideon Stein, "Monitoring activities from multiple video streams: Establishing a common coordinate frame," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 22, no. 8, pp. 758–767, 2000.

[12] Javed Aslam, Zack Butler, Florin Constantin, Valentino Crespi, George Cybenko, and Daniela Rus, "Tracking a moving object with a binary sensor network," in *Proceedings of the 1st international conference on Embedded networked sensor systems*, Los Angeles, California, USA, 2003, pp. 150 – 161.

[13] Weiming Hu, Tieniu Tan, Liang Wang, and Steve Maybank, "A survey on visual surveillance of object motion behaviors," *IEEE Transactions on Systems, Man and Cybernetics– Part C: Applications and Reviews*, vol. 34, no. 3, pp. 334–352, Aug 2004.