ENGINEERING AMBIENT VISUAL SENSORS

G.M.P. O'HARE[*], M.J. O'GRADY, R. JURDAK, C. MULDOON

*Adaptive Information Cluster, School of Computer Science & Informatics,
University College Dublin (UCD), Belfield, Dublin, Ireland*

NOEL E. O'CONNOR, CIARÁN Ó CONAIRE, PHILIP KELLY

*Adaptive Information Cluster, School of Electronic Engineering, Dublin City University
(DCU), Dublin 9, Ireland*

Visual sensors are an indispensable prerequisite for those AmI environments that require a surveillance component. One practical issue concerns maximizing the operational longevity of such sensors as the operational lifetime of an AmI environment itself is dependent on that of its constituent components. In this paper, the intelligent agent paradigm is considered as a basis for managing a *camera collective* such that the conflicting demands of power usage optimization and system performance are reconciled.

1. Introduction

Increasingly there is a demand for the deployment of low cost visual systems which, while fulfilling their surveillance or monitoring functions, remain unobtrusive. Ubiquitous visual sensing demands a low cost base and in this paper, we describe a low cost alternative to the high cost high precision camera infrastructure that is typically advocated. It is our contention that the use of a networked federation of low cost camera units, which can opportunistically collaborate in a shared monitoring task, can offer cost savings both from a monetary and a power usage perspective, yet yield a quality of performance that is comparable to that of their expensive counterparts. In a typical distributed camera infrastructure, the camera units are envisaged as being autonomous in nature and equipped with both a communication and processing capability. They are thus able to participate in shared tasks and to opportunistically make inferences as to how best to contribute to the shared goal. Such units are most likely to be devoid of a dedicated power source. Thus issues pertaining to

reductions in transmission and processing of sensed data, while maintaining an adequate service, are paramount to maximizing the longevity of the network.

2. Related Research

Surveillance systems have been categorized into generations and at present, Third Generation Surveillance Systems (3GSSs)[1] are predominant. Some of the research challenges identified with this generation concerns the respective merits of distributed and centralized intelligence as well as multi-camera surveillance techniques[2]. In response to this, a number of researchers have investigated intelligent agents as an underlying software framework by which to support networks of spatially diverse cameras. Camera units that incorporate an onboard processing unit may be referred to as *intelligent* or *smart[3]*. However agent researchers would consider such terms to imply the adoption of Distributed Artificial Intelligence (DAI) techniques. The authors advocate *camera collectives* whereby smart cameras collaborate to achieve shared goals.

Various attributes of the intelligent agent paradigm have been harnessed in surveillance research. Remagnino et al[4] consider agents as providing an architecture for managing the various components in a camera network, resulting in a scalable and maintainable design and implementation. Their modular and autonomous characteristics are perceived as making them a particularly apt solution. In contrast, Bramberger et al[5] focus on the mobility characteristic and envisage agents as being assigned individual surveillance tasks which they can accomplish by migrating between cameras. Patricio et al harness the social aspect of agents in their framework[6]. Uniquely, their reasoning model is based on the Belief-Desire-Intention (BDI) model. Though the issue of power management has been broadly considered[7], the use of agents for prolonging the operational lifespan of visual sensor networks has as yet not been explored. And it is this that motivates this paper.

3. Experimental Design

In order to test the efficacy of our agent-based approach, we characterize a typical scenario of sensing people as they move through a communal physical space. In many such applications the surveillance systems seeks to assist in the identification of individuals. The key to achieving this is acquiring quality and prolonged camera exposure to the face of persons that move through the space.

A camera network (Figure 1), consisting of 10 static wireless internet cameras, was deployed in an office environment of dimensions 7m x 15m. The resolution of each camera was 320 x 240 pixels, and wireless communication

was realized via IEEE 802.11g. Cameras were positioned at heights ranging from 0.85m to 1.70m, with the lens oriented at head height for optimal person viewing. This setup gives almost 100% coverage of the test environment.
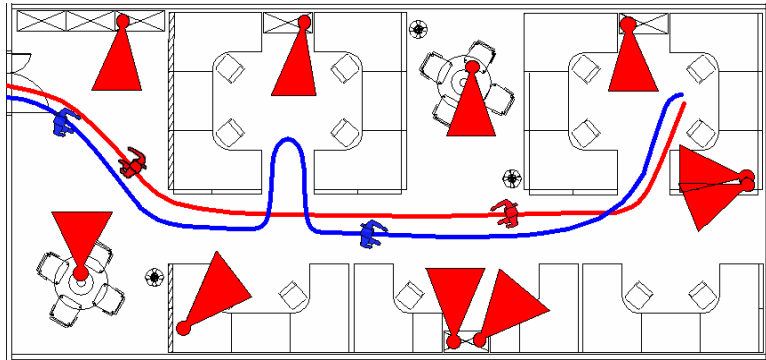

Figure 1. Bird's Eye view of the experimental configuration

3.1. *Data Capture Process*

Using a multi-threaded application, running on a standard workstation, each of the 10 cameras was polled continuously, achieving a capture frame-rate of approximately 4 frames per second. The image data was stored locally, time-stamped and subsequently processed off-line using three different analysis modules. At all stages, the assumption is made that, at most, one person is present in each video frame.

3.1.1. *Processing Level 1: Person Detection*

The first few frames of the sequence are assumed to be devoid of people. The background model used is simply a background image, computed as an average of the first 20 frames. For each new image captured by the camera, the foreground pixels are detected using background subtraction. The Euclidian distance between each pixel and its corresponding pixel in the background image (in the RGB colour-space) is computed, creating a difference image. This difference image is then thresholded using the adaptive method of Rosin[8], resulting in a binary foreground image (Fig 2b). The foreground pixels are then projected onto the x-axis, giving a one-dimensional signal (Fig 2e). The peak of this signal relates to the most dominant moving object in the scene. The peak is selected from the smoothed signal and its width is determined by computing when the signal falls below 10% of the peak height. Left and right bounds of the

 person are marked as these points. The confidence of person detection is computed from the height of the foreground peak in the projection map using a manually tuned sigmoid function that assigns higher confidence to larger peak heights and a confidence close to zero to peaks that are caused by camera noise.

3.1.2. *Processing Level 2: Basic Face Detection*

The first step is to map the pixels from the current frame to skin likelihoods. RGB colour histogram models of skin and non-skin were trained using manually annotated skin and background images from the work of Sigal et al[9]. By computing the log-likelihood ratios for each pixel's skin probability to non-skin probability, a skin likelihood image is created (Fig 2c). This is then thresholded, using Rosin's algorithm (Fig 2d). After removing noise (using morphological processing) and pixels outside the person bounds, a connected component algorithm is employed to find the largest skin region. This region is designated the detected face (Fig 2f). Two confidence values are determined for this face using its width and position. These confidence values are then multiplied to give the final face confidence score.

3.1.3. *Processing Level 3: Viola Jones Face Detection*

The final type of processing invokes a complete state-of-the-art face detection algorithm on the input images. The images were converted to greyscale and input to the well-known Viola and Jones face detection algorithm[10].
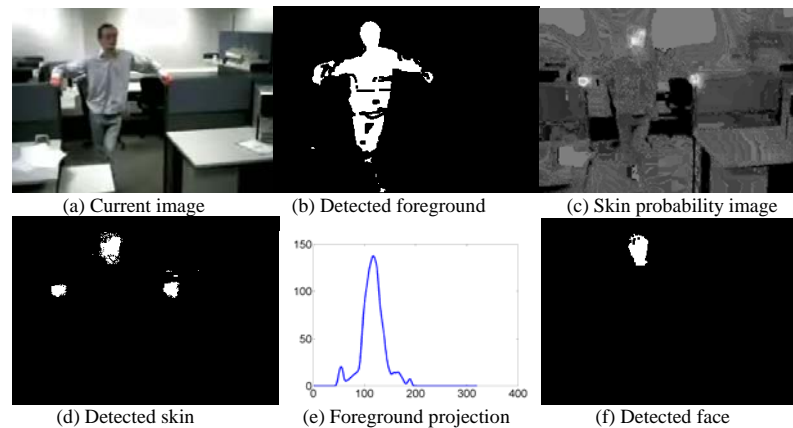

| (a) Current image | (b) Detected foreground | (c) Skin probability image |
| (d) Detected skin | (e) Foreground projection | (f) Detected face |

Figure 2: Key stages in the data capture process

4. Agent Configuration

Mirroring the physical camera configuration is a suite of intelligent agents, with each camera being monitored by a single agent. The agent framework used for the experiment was Agent Factory Micro Edition[11], a minimized footprint agent platform that subscribes to the BDI model of agency. This agent framework is a recent implementation of the classic Agent Factory framework[13], and has been successfully demonstrated in other AmI applications[14].

The 10 static wireless internet cameras that comprise the camera collective each produce a linear timestamped sequence of images. At any given instance in time 10 such images are available for potential transmission to a central network node to be woven into a single aggregated image sequence. Under optimum operating conditions the collective will be able to identify, for each given instance, a subset of camera agents that potentially can provide good quality facial frontage and these agents will thereafter bid to obtain the contract of supplying the best image.
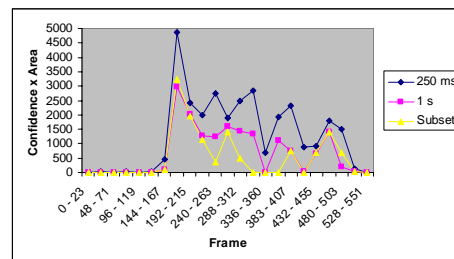
5. Experimental Results

Figure 3. Maximizing Coverage in terms of User Detection

Three experiments were performed to evaluate the smart camera framework. In each of these experiments three different sets of results were recorded. The first set represents the case when the auction is held four times a second, that is every 250ms (or equivalent to the frame rate) and using all agents. This represents the optimum in terms of confidence for user/face detection. The second set of results represents the case in which the auction is held once every second and, again, uses all agents. This set has a lower confidence, but fewer resources are consumed. The third set of results represents the case in which the auction is held once every second while also using a subset of agents. Agent choice is based upon an analysis of past behavior patterns. These results have a lower confidence than the second set, but the least resources are consumed. The results

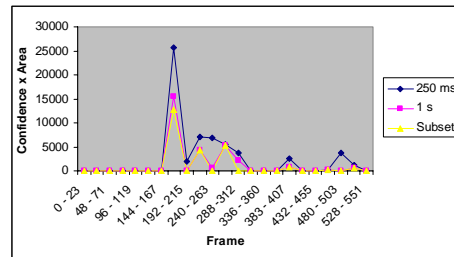 of the three experiments have been depicted graphically on Figures III, IV and V respectively.

Figure 4. Maximizing basic face detection

All three experiments represent the cases when the cameras are maximizing user detection, basic face detection, and front/side face detection respectively. The results are consistent in each case; lowering the auction frequency and the number of agents participating in the auction leads to a saving in resources but also to a drop in coverage. In all three experiments, including the scenarios in which all agents participate in the auction, there will be only a single agent transmitting frames. This leads to a loss of information regarding alternative views of the detected person.
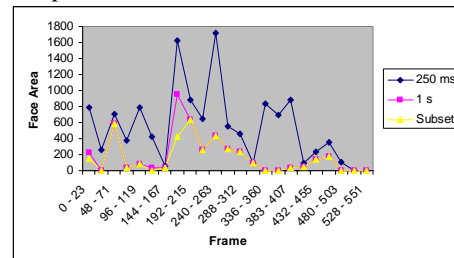

Figure 5. Maximizing face detection using both frontal and side face detection algorithms.

6. Conclusion

This paper has introduced the notion of a camera collective, a network of autonomous and intelligent low cost cameras suitable for a variety of ambient visual monitoring services. Our approach is characterized by the use of multi-agent systems as the delivery mechanism for collaborative visual processing. The approach offers the capability to participate in shared tasks and to opportunistically make inferences as to how best to contribute to the shared goal. This was exemplified through the scenario of sensing people as they move through a communal physical space. Experimental results confirm that

significant power saving can be obtained by reducing both auction frequency and those agents participating in the auction without compromising the shared surveillance task.

References

1. K. Marcenaro, F. Oberti, G. Foresti and C. Regazzoni, Distributed Architectures and Logical-Task Decomposition in Multimedia Surveillance Systems, *Proceedings of the IEEE*, 89(10):1419-1440 (2001)
2. M. Valera and D.A. Velastin, Intelligent Distributed Surveillance Systems: A Review, *IEE Proc. Vis. Image Signal Proces*s., 152(2):192-204 (2005)
3. W. Wolf, B. Ozer and T. Lv, Smart Cameras as Embedded Systems, *IEEE Computer*, 35(9):48-53 (2002)
4. P. Remagnino, A.I. Shihab and G.A. Jones, Distributed Intelligence for Multi-camera Visual Surveillance, *Pattern Recognition*, 37:675-689 (2004)
5. M. Bramberger, A. Doblander, A. Maier, B. Rinner and H. Schwabach, Distributed Embedded Smart Cameras for Surveillance Applications, *IEEE Computer*, 39(2):68-75 (2006)
6. M.A. Patricio, J. Carbo, O. Perez, J. Garcia and J.M. Molina, Multi-Agent Framework in Visual Sensor Networks, *EURASIP Journal on Advances in Signal Processing – Special issue on Visual Sensor Networks* (2007)
7. A. Maier, B. Rinner and H. Schwabach, A Hierarchical Approach for Energy-aware Distributed Embedded Intelligent Video Surveillance, *Proceedings of the 11th International Conference on Parallel and Distributed Systems (ICPADS'05)*, 2:12-16 (2005)
8. P.L. Rosin, Unimodal thresholding, *Pattern Recognition*, 34(11):2083–2096 (2001)
9. L. Sigal, S. Sclaroff, and V. Athitsos, Skin color-based video segmentation under time-varying illumination, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(7):862–877 (2004)
10. P. Viola and M.J. Jones, Rapid object detection using a boosted cascade of simple features, *IEEE Conference on Computer Vision,*511–518 (2001)
11. C. Muldoon, G. M. P. O'Hare, R. W. Collier, and M. J. O'Grady, Agent Factory Micro Edition: A Framework for Ambient Applications. In Intelligent Agents in Computing Systems, *Lecture Notes in Computer Science*, 3993:727-734 (2006)
12. G.M.P. O'Hare, Agent Factory: An Environment for the Fabrication of Distributed Artificial Systems, In: *Foundations of Distributed Artificial Intelligence (*Eds O'Hare & Jennings), Wiley Publishers (1996)
13. M.J. O`Grady and G.M.P. O`Hare, Just-In-Time Multimedia Distribution in a Mobile Computing Environment, *IEEE Multimedia*, 11(4):62-74 (2004)