Reliable Scalar-Visual Event-Detection in Wireless Visual Sensor Networks

Alexandra Czarlinska and Deepa Kundur
{czlinska, deepa}@ece.tamu.edu
Department of Electrical and Computer Engineering, Texas A&M University

***Abstract*—In this work we consider an *event-driven* Wireless Visual Sensor Network (WVSN) where each camera node transmits a frame to the cluster-head *only if* an event of interest was captured in the frame for energy and bandwidth conservation. Specifically, we consider the scenario where each camera node receives decision support from an independent but possibly attacked (and hence error-prone) scalar-sensor regarding the presence or absence of an event. We study the overall detection performance achieved by various techniques that utilize the scalar and image-based decisions. We conclude that in image sequences involving extraneous lighting and background changes (such as in the case of outdoor surveillance), the combination techniques generally achieve a lower total probability of error.**

I. Introduction

Wireless Visual Sensor Networks (WVSNs) equipped with cameras are envisioned for a variety of innovative applications ranging from surveillance and intelligent infrastructure monitoring to wildlife data collection [1]. To realize this vision, WVSN research must overcome the challenges associated with the increased processing, transmission energy and bandwidth required for image data compared with scalar data (such as temperature readings) [2]. One interesting approach to addressing these challenges is to exploit the collaboration potential between camera nodes and scalar sensor nodes, especially to detect events of interest occurring in the environment [3].

In this work we examine an *event-driven* WVSN where camera nodes transmit frames to a cluster-head (or base station) *only if* the frames contain "relevant events" (as will be discussed). The motivation for such selective transmission based on local decisions stems from energy and bandwidth considerations. Ideally, each camera node should make a correct decision regarding a frame. Specifically, we wish to achieve a high probability of event detection P_D, otherwise important frames will not be transmitted. Simultaneously we wish to achieve a low probability of false alarm P_{FA}, otherwise non-event frames will be wastefully transmitted. Since both probabilities are important and competing (lowering P_{FA} lowers P_D), we aim to maximize P_D for a fixed chosen P_{FA} which leads to a total probability of error P_{err}.

To achieve the local event-detection goal we may rely on a variety of image processing algorithms that can be applied to the image frames. Indeed a large body of research exists detailing image change and shot change detection in off-line and real-time settings [4]. Many of the techniques are not particularly suitable for WVSNs because of their complexity or assumptions about the availability and characteristics of event and non-event statistics. In cases of limited node energy we would ideally wish to retain a low probability of error despite the selection of a simpler detection algorithm.

In this work we consider a scenario where each WVSN camera node receives detection assistance from an independent scalar-sensor node as depicted in Figure 1. The specific type of scalar-sensor is not considered but rather the interaction is examined in terms of the resulting probabilities of error. Specifically, we consider the case where the scalar-sensors may suffer from errors due to the presence of malicious attacking nodes (shown as A in Figure 1) [5], [6]. We consider three different methods of incorporating the detection decisions of the scalar-sensors with the decisions available from the image processing and study their performance in terms of the probability of error on several test sequences. The focus and contribution of this paper is thus a study of the detection performance achieved by *combining* the scalar-sensor decisions with local image processing. We do not specifically focus on the development or improvement of image processing algorithms for WVSNs. Rather we compare the performance of three methods for utilizing the decisions and discuss conditions under which they perform best.

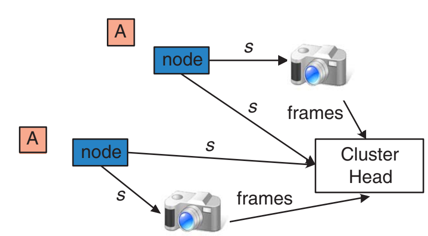

Fig. 1. A camera node receives a decision s from a scalar-sensor. Attacking nodes A may cause errors in s. The cluster-head also receives the s data.

II. Visual Event-Detection in WVSNs

An extensive body of research exists examining image change and shot change detection in image sequences [4]. Much of this research focuses on the processing, segmentation and classification of a stored movie or news-cast which is *already* available in its entirety. As such, it is often possible to obtain meaningful statistics via pre-processing of the frames. These statistics can then be used in detection and segmentation algorithms to distinguish between event and non-event frames. It is important to note that the definition of event and non-event

1-4244-1457-1/08/$25.00 © IEEE
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE CCNC 2008 proceedings.

is largely application dependent, as is the domain in which the processing is carried out (spatial-temporal or frequency-based such as DCT, JPEG or MPEG).

In contrast, a WVSN camera node collects *incoming* images containing unknown objects which may or may not enter into the frame at any time [7]. Since we do not know ahead of time what objects will be encountered and under what lighting and changing background conditions, it is not possible to assume that we know the statistics of an event frame. Rather we make the weaker assumption that an event frame is one where "significant" motion has occurred. Under certain deployment scenarios however, it may be reasonable to assume that we know the approximate statistics of a *non-event* frame (called the "null hypothesis" statistics). For instance, in this work we assume that a visual-node is deployed or activated during a non-event time. The visual-node is thus able to process the initial frames it collects and determine some approximate null-hypothesis statistics.

The specific statistics utilized by the visual-nodes are largely application and energy-use dependent [4]. The latter constraint is particularly worthy of mention in WVSNs since more sophisticated (and hence possibly more energy-intensive) processing might yield a lower total probability of detection error [7]. This translates into transmission energy savings since fewer irrelevant (non-event) frames are sent. However allocating more energy to visual event-detection (i.e. processing) may drain the nodes too quickly, limiting their lifetime. The trade-offs *between* various energy-allocation schemes for event-detection are a subject of ongoing study. In this work we focus on achieving a lower total probability of error given the use of scalar-sensors for a *fixed* visual event-detection algorithm.

We consider a relatively simple and general event-detection algorithm (i.e. the detection statistics employed are not tailored to the detection of any *specific* object). The choice of algorithm is based in part on processing simplicity and in part on observations regarding the real-world image sequences used in our testing as shown in Figures 2, 3, 4, and 5. The sequence of Figure 2 is an idealized indoor test where the lighting and background conditions do not change appreciably over time. The only significant change comes from the event of interest in the form of a test subject entering the camera's field of view. The dominant source of noise in this case is internal camera noise and flicker.

The sequence of Figure 3 shows outdoor parking-lot surveillance on a windy day, where the event of interest is the passing of an unidentified car. The detection task in this sequence is complicated by the presence of a nearby shrub which experiences significant swaying of its branches over time. Furthermore the background lighting changes visibly with cloud movement (between frames 3(a) and 3(b) for example). The sequence of Figure 4 also experiences changes due to swaying trees and variable light conditions. The event of interest is the appearance and movement of a test subject which temporarily disappears behind a tree in frames 4(c) and 4(e). Finally Figure 5 shows a truncation of sequences 2 and 3 where the camera's field of view now excludes the shrub and trees.

Statistical analysis of image sequences 2 and 3 (such as Levine's Test and the t-test) reveal that the mean and standard deviation are not reliable indicators of an event of interest occurring even after various DCT-domain filtering mechanisms are employed. This can be seen intuitively from the fact that the subjects of interest (person walking and car driving-by) do not occupy a much larger percent of a frame's pixels than the other randomly moving objects (shrub and trees). Hence the mean and variance of the frames *do* change based on the appearance of the subject, but these differences are not statistically distinguishable. In essence, the pixels corresponding to the person and car are getting dwarfed by the presence of many shrub and tree pixels which are also changing over time. Truncating the frames as shown in Figure 5 to exclude the vegetation does indeed improve the statistical difference between an event and non-event frame. However for the general WVSN deployment case (with cameras facing in various directions), we do not wish to select an event-detection technique which relies on the truncated assumption.

The simple algorithm we selected is based on difference images, similar to the technique found in the image change detection literature [8]. In essence, a difference image reveals all the pixels that have changed from the previous frame (containing both relevant and irrelevant changes such as the tree swaying). The Mean Squared Error (MSE) of the difference image is computed as the relevant statistic, and it is compared to a theoreticaly-obtained robust threshold T based on the chi-squared test (please refer to Appendix I for the details). In this technique we rely on the assumption that the standard deviation of the null hypothesis (non-event following deployment) can be estimated. Since the "alternative hypothesis" (event) statistics remain unknown, we employ a composite-hypothesis test. It is important to emphasize that a variety of other visual event-detection algorithms could have been selected for testing with the scalar-sensors. In this work we are interested in evaluating the *improvement* in detection performance of the WVSN due to the addition of the *scalar-sensors* rather than the merits of various image processing algorithms.

III. Reliability of Scalar Data

As explored extensively in the Sensor Network literature, there are many security attacks capable of rendering the data observed by the network unreliable [6]. For instance, we may consider the presence of a rival network with malicious nodes depicted as nodes A in Figure 1. These nodes may perform standard sensor network attacks or a new type of attack referred to as an actuation attack where rival nodes directly perturb the observations collected by the legitimate network from the environment [6]. In addition to these attacks, scalar sensors may occasionally malfunction due to component failure, miscalibration or harsh environmental conditions. Whether in the face of attack or failure, it is thus crucial to guarantee a level of scalar-data *reliability* if such data is to be

Fig. 2. Frame Seq. 1 with frames (a)-(f) from top left to bottom right: indoor test conditions with constant lighting and no background changes.

Fig. 3. Frame Seq. 2 with frames (a)-(f) from top left to bottom right: outdoor variable lighting due to clouds. Ex: The light intensity changes by 70% between frames (a) and (b). Additional background movement due to shrub.

Fig. 4. Frame Seq. 3 with frames (a)-(f) from top left to bottom right: changing outdoor light and background (swaying trees). The subject temporarily disappears behind a tree in frames (c) and (e).

Fig. 5. (a) Frame Seq. 4a) showing Seq. 2 modified to remove the shrub. (b) Frame Seq. 4b) showing Seq. 3 modified to remove the swaying trees.

used for event-detection in WVSNs. For the purposes of this work we assume that errors may be caused solely by a hostile attack. In [6] and [9] we showed how the *interaction* between a sensor network collecting scalar data and a malicious attacker may be modeled. As shown in Figure 1, we consider a sensor network where each node sends its scalar decision s to a cluster-head. Specifically, upon taking a measurement of the environment, each scalar sensor decides whether an event of interest is present (the "alternative hypothesis" $\mathcal{H}_1$) or absent (the "null hypothesis" $\mathcal{H}_0$), and sends this decision to the cluster-head. The cluster-head in turn uses information about the average number of events likely to occur during one time interval to decide whether an attack has occurred. We use the notation n to denote the number of nodes reporting to the cluster-head. The notation p is used to denote the probability that a node detects an event of interest, and q denotes the probability that a node is attacked. The effect of the attack is modeled as "flipping" a node's decision from $s = 1$ (event is present) to $s = 0$ (event is absent) or vice versa. This model abstracts away the specific type of scalar-sensor (motion or pressure) and focuses on the effect of changing the sensor's reported decision.

We may consider the sensor network and the attacking network as rivals, each vying for control of attack detection at the cluster-head. Based on such competition we are able to use the game theoretic concept of a Nash equilibrium to determine each network's best strategies. Indeed in [6] and [9] we showed that *given* that the cluster-head performs the check, the malicious network is generally *forced* to attack with a *small* probability q. This result is important because it can

TABLE I
OPTIMAL q VALUE FOR CLUSTER SIZE n AND PROBABILITY OF EVENT p

n	**Best q for $p = 0.5$**	**Best q for $p = 0.1$**
100	0.031	0.022
50	0.062	0.044
40	0.0780	0.055
30	0.103	0.074
20	0.152	0.110
10	0.990	0.205
5	0.505	0.405
3	0.655	0.655
2	0.999	0.999
1	0	0

be shown that if no check is performed at the cluster-head, the malicious nodes may employ *any* value of q. The optimal value of q selected by the malicious network depends on the specific value of n and p. Table I shows the optimal value of q for $p = 0.5$ and $p = 0.1$ (event of interest occurs with probability of 50% and 10% in any given frame respectively) for various n. We observe that with few exceptions, as the cluster-size *increases*, the optimal value of q *decreases*. Furthermore, a smaller value of p also tends to force a lower value of q. For a cluster-size of approximately 20 to 30 nodes, we can expect a probability of attack of $q \approx 0.1$ while other cases of n and p are shown in Table I and the general case is detailed in [6].

IV. VISUAL-SCALAR EVENT-DETECTION MODEL

In this work we consider an event-driven WVSN where camera nodes record frames and send these frames to a cluster-head *only if* an event of interest was captured. As shown in Figure 1, each camera node receives support from a scalar-sensor that makes an independent decision s based on its readings (such as temperature, pressure etc). Each scalar sensor has a probability q of being in error due to attack, while each camera has a probability q' of being in error (depending on the specific image detection algorithm employed).

Each camera node thus faces a scenario where for each captured frame it has access to an image processing-based decision IP and a scalar-sensor decision s. The probability q' that IP is in error may not be equal to the probability q that s is in error. The ultimate goal of the camera node is to maximize P_D for a fixed error P_{FA}. This leads to an overall P_{err} which is given by Eq. 1. We examine three techniques of incorporating the IP and s decisions. As shown in Table II, the first approach is to always trust the scalar sensor s (including when a disagreement between IP and s occurs). This approach is motivated by the results of Section III which state that cluster-head checking of the scalar data forces an attacker to use a $small$ value of q. The assumption of a small q may of course not hold if cluster-head checking is not performed (i.e. the attacker may use any q). Method 2 shown in Table II takes a more conservative approach by always marking a disputed frame as an "event" (i.e. a disputed frame will always be sent to the cluster-head). Finally in approach 3 we always trust the decision obtained via visual event-detection (image processing).

$$P_{err} = P_{FA} \cdot P(\mathcal{H}_0) + (1 - P_D) \cdot P(\mathcal{H}_1) \qquad (1)$$

TABLE II
METHODS

Method	Action
Method 1	Trust scalar-sensor s
Method 2	Always mark as "event" if disputed
Method 3	Trust Image Processing IP

TABLE III
VISUAL DETECTION BASED PERFORMANCE (IP)

Image Sequence	$\mathbf{P_D}$	$\mathbf{P_{FA}}$	$\mathbf{P_{err}}$
Seq. 1. Indoor walking	1.0	0.13	0.03
Seq. 4a. Outdoor car, no trees	0.98	0.17	0.11
Seq. 4b. Outdoor walking, no trees	0.50	0.03	0.17
Seq. 2. Outdoor car, with trees	0.87	0.26	0.20
Seq. 3. Outdoor walking, with trees	0.05	0.23	0.90

V. RESULTS AND DISCUSSION

The visual event-detection algorithm described in Section II and Appendix I was implemented in Matlab and tested on the image sequences shown in Figures 2, 3, 4 and 5. The detection performance of this algorithm is shown in Table III where the results are listed in order of degrading event-detection performance. In addition to showing the total probability of error P_{err} (given by Eq. 1), we also show the underlying probability of detection P_D (which ideally should be as close to 1 as possible) and the probability of false alarm P_{FA} (which ideally should be as close to 0 as possible). As expected, the image processing algorithm performs best on image sequences with the least amount of lighting and background changes or where the subject occupies a significant portion of the frame. For certain image sequences (such as Seq. 3), P_{err} is unacceptably high (0.9). The methods incorporating scalar-sensor data into the decision process (listed in Table II) were implemented using Matlab. The P_{err} results obtained for each image sequence are shown in Figures 6 and 7 where algo$_i$ refers to method i in Table II. We make the following two observations. First, incorporating and trusting the scalar sensor decisions as in Method 1 greatly reduces P_{err} for most sequences, as long as the scalar error q does not exceed the image processing error q'. This is especially true for the more difficult sequences (Seq. 2 and 3) containing lighting and background changes such that q' is large. Furthermore the condition for $q < q'$ may be met if q is due to a sensor network attack as outlined in Section III. Second, in the regime of larger sensor error q, Method 3 (IP) is superior. Method 2 is generally not optimal, but for certain sequences, it performs close to Method 1. Method 2 thus becomes important for cases where we do *not* know the relationship between q and q' (i.e. which of the two errors is smaller). For instance, though both sequences in Figure 6a) and 6b) were obtained outdoors in variable conditions, the image processing error q' varies dramatically between them.

TABLE IV
TOTAL ERROR PROBABILITY P_{err} FOR $q = 0.1$

Image Sequence	**Meth.** 1	**Meth.** 2	**Meth.** 3
Seq. 1. Indoor walking	0.1515	0.0573	0.03
Seq. 4a. Outdoor car, no trees	0.1238	0.1531	0.11
Seq. 4b. Outdoor walking, no trees	0.11	0.09	0.17
Seq. 2. Outdoor car, with trees	0.12	0.20	0.20
Seq. 3. Outdoor walking, with trees	0.05	0.23	0.90

Finally, recalling the competitive network results discussed in Section III, we assume a cluster size n of approximately 20 to 30 nodes which subsequently guarantees an attack q restricted to $q \approx 0.1$. These results also apply to larger clusters since an increase in n *lowers* q [6]. Table IV summarizes the improvement in total error probability P_{err} obtained by utilizing the scalar sensors over the purely image processing (IP) case. It can be seen that in this regime of small q, Method 1 usually performs better than Method 2 but that the IP method is superior for the indoor sequence since $q' < q$ in this case.

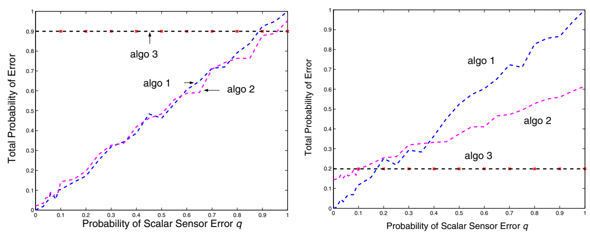

Fig. 6. Comparison of the P_{err} for the three methods (a) Seq. 3 (Outdoor walking with trees) (b) Seq. 2 (Outdoor car with trees)

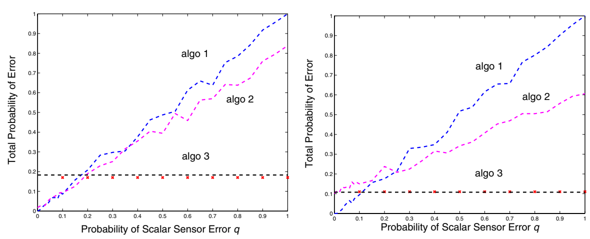

Fig. 7. Comparison of the P_{err} for the three methods (a) Seq. $4b$ (Outdoor walking no trees) (b) Seq. $4a$ (Outdoor car no trees)

VI. CONCLUSION

In this work we examine the detection performance of *event driven* Wireless Visual Sensor Networks (WVSNs) where image frames are transmitted back to a cluster-head *only if* the frame contains an event of interest. The decision regarding a frame can be made purely based on image processing techniques or it can be made using assistance from independent scalar-sensors which may themselves contain errors due to hostile attack. We show through simulation of various image sequences that combining the scalar-sensor decisions with the image processing yields a smaller total probability of error. This is especially true for outdoor surveillance frames containing extraneous changes where the scalar-sensor error due to attack is forced to be small through cluster-head checking.

VII. DETAILS OF VISUAL EVENT-DETECTOR

The detector we consider is based on the method proposed by Aach and Kaup [8] which we summarize here for reference. Let D_i be the i^{th} difference pixel between two frames. We take the entire difference image as one large block instead of using several smaller blocks to lessen the energy consumption. Thus the random variable of interest is given by

$$X = \sum_{i=1}^{n} D_i^2 = \sigma_j^2 \sum_{i=1}^{n} \frac{D_i^2}{\sigma_j^2} = \sigma_j^2 Y, \quad \text{for } j = 0 \text{ or } 1 \qquad (2)$$

where Y has distribution chi-squared with n degrees of freedom, and n is the total number of pixels in the difference frame. The hypothesis test can thus be summarized as:

$$\mathcal{H}_0 \ : \ \text{event (motion)}, X \sim \frac{1}{\sigma_0^2} f_{\chi^2,n}\left(\frac{x}{\sigma_0^2}\right) \qquad (3)$$

$$\mathcal{H}_1 \ : \ \text{no-event}, X \sim \frac{1}{\sigma_1^2} f_{\chi^2,n}\left(\frac{x}{\sigma_1^2}\right) \qquad (4)$$

where $f_{\chi^2,n}(x)$ is the probability density function (pdf) of the chi-squared distribution with n degrees of freedom. Hence the hypothesis test is given by for false alarm rate not exceeding α. $F_{\chi^2,n}$ is the cumulative distribution function of the chi-square distribution with n degrees of freedon, and $F^{-1}_{\chi^2,n}$ is its inverse.

$$x \underset{H_0}{\overset{H_1}{\gtrless}} \sigma_0^2 F^{-1}_{\chi^2,n}(1-\alpha) \qquad (5)$$

REFERENCES

[1] W.-C. Feng, J. Walpole, W.-C. Feng, and C. Pu, "Moving towards massively scalable video-based sensor networks," in *Workshop on New Visions for Large-Scale Networks: Research and Applications*, Washington, DC, March 2001, p. 385.
[2] K.-Y. Chow, K.-S. Lui, and E. Lam, "Efficient on-demand image transmission in visual sensor networks," *Eurasip Journal on Advances in Signal Processing*, vol. v 2007, p. 11 pages, 2007.
[3] A. Basharat, N. Catbas, and M. Shah, "A framework for intelligent sensor network with video camera for structural health monitoring of bridges," in *Proceedings. Third IEEE International Conference on Pervasive Computing And Communications Workshops, PerCom 2005 Workshops*, March 2005, pp. 385–9.
[4] R. Joyce and B. Liu, "Temporal segmentation of video using frame and histogram space," *IEEE Transactions on Multimedia*, vol. 8, no. 1, pp. 130–40, February 2006.
[5] L. Buttyan and J.-P. Hubaux, "Report on a working session on security in wireless ad hoc networks," *Mobile Computing and Communications Review*, vol. 6, no. 4, pp. 1–17, 2002.
[6] A. Czarlinska, W. Luh, and D. Kundur, "Attacks on sensing in hostile wireless sensor-actuator environments," in *IEEE Globecom*, Washington, DC, 26-30 November 2007, available: www.ece.tamu.edu/˜czlinska/Glob07/AttacksSenAct.pdf.
[7] K. Veeraraghavan, D. Peng, and H. Sharif, "Energy efficient multi-resolution visual surveillance on wireless sensor networks," in *IEEE International Conference on Electro Information Technology*, Lincoln, NE, 22-25 May 2005, p. 6 pp.
[8] R. Radke, S. A. O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: a systematic survey," *IEEE Transactions on Image Processing*, vol. 14, no. 3, pp. 294–307, March 2005.
[9] A. Czarlinska and D. Kundur, "Attack vs. failure detection in event-driven wireless visual sensor networks," in *ACM Multimedia & Security Workshop (MM&Sec'07)*, Dallas, TX, 20-21 September 2007, available: www.ece.tamu.edu/˜czlinska/ACM07/WVSN.pdf.