中央研究院
資訊科學研究所
Institute of Information Science, Academia Sinica • Taipei, Taiwan, ROC

TR-IIS-12-007

A Framework for Fusion of Symbiotic Human Sensor and Physical Sensor Data

J.W. S. Liu, E. T.-H. Chu, and P. H. Tsai

Dec. 31, 2012 || Technical Report No. TR-IIS-12-007
http://www.iis.sinica.edu.tw/page/library/TechReport/tr2012/tr12.html

Institute of Information Science, Academia Sinica
Technical Report TR-IIS-12-007

A Framework for Fusion of Symbiotic Human Sensor and Physical Sensor Data

J. W. S. Liu, E. T.-H. Chu, and P. H. Tsai

ABSTRACT

Past experiences tell us that a disaster warning and response system can improve its surveillance coverage of the threatened area and situation awareness by supplementing in-situ and remote sensor data with human sensor data captured and sent by people in the area. This paper is concerned with fusion and processing methods with which the system can make use of human sensor data and physical sensor data synergistically to speed up the decision process and improve the quality of its decision. We formulate the problem in a statistical detection and estimation framework. Within this framework, value fusion and decision fusion of human sensor data and physical sensor data can be treated in a coherent way.

Keywords: Crowdsourcing, Multiple sensor fusion, Statistical detection and estimation.

J. W. S. Liu is affiliated with Institute of Information Science and Center for Information Technology Innovation, Academia Sinica, Taiwan. Her email address is janeliu@iis.sinica.edu.tw

E. T. H. Chou is affiliated with Computer Science Department, National Yunlin University of Science and Technology, Taiwan. His email address is edwardchu@yuntech.edu.tw

P. H. Tsai is affiliated with Institute of Manufacturing Information System, National Cheng Kung University, Taiwan. Her email address is phtsai@mail.ncku.edu.tw

1 INTRODUCTION

Recently, the tremendous growth in usages of smart mobile devices equipped with cameras, temperature and vibration sensors, etc. and social networking services have enabled a wide spectrum of applications and services to be more pervasive, location-aware and context-aware than feasible even a few short years ago. Examples of such applications and services, crowdsourcing projects worldwide, approaches to integrating social sensing with pervasive services, and overview of the state of the art and remaining challenges can be found in [1-9].

It comes with no surprise that people with smart mobile devices and social networking services are playing an increasingly more essential role in disaster preparedness and response. Experiences from recent disasters (including the devastating oil spills, wildfires and floods worldwide) tell us that in-situ and remote *physical sensors* deployed by disaster surveillance, early warning and rapid response systems often cannot provide the system with adequate data for situation assessment purposes. When this happens, crowdsourcing human sensor data can be an effective solution. By a *human sensor*, we mean a person armed with one or more smart mobile devices and social networking services. By *human sensor data*, we mean observation (and measurement) data captured and contributed by human sensors.

The platform CROSS (a CROwdsourcing Support system for disaster Surveillance) [10] was built to support the exploitation of human sensor data and physical sensor data synergistically for disaster surveillance and decision support purposes. When physical sensor coverage is inadequate, the system starts a *crowdsourcing data collection* (*CDC*) process by broadcasting a call for participation to a crowd of human sensors. During the process, participating human sensors make observation(s) at and around locations as requested by the system and send the data thus captured back to the system. The process completes when the system has acquired sufficient data about the threatened area to give it situation awareness and support its decisions and operations.

The interactions and collaborations between the system and participants can be either crowd-driven or system-driven [11], and CROSS supports both types of strategies. When the system uses a *crowd-driven strategy*, it either does nothing other than collecting and processing reports from participants, relying solely on mobility and interactions of individual participants for coverage of the threatened area, or from time to time provides them as feedback with updates of the current condition of the threatened area based data collected and processed at the time. Working in this way, the system is similar to many crowdsourced sensing systems and applications (e.g., [1-7]). These applications have demonstrated the effectiveness of crowd-driven social and participatory sensing from massive crowd for a variety of usages, including generation of fine-grain maps of weather radar, noise level, air quality, snow depth, radiation level, traffic and road conditions, litters in parks, and so on.

For the purpose of collecting data to supplement physical sensor data prior to or during an emergency, the crowd-driven approach is not ideal, however: Oftentimes, the system should use as few participants as needed for each CDC process for many reasons including availability and costs of qualified participants. Without well-planned routes for participants to follow during a process, some locations may be visited by more participants than necessary while other locations are visited by too few. Consequently, the *response time* (i.e., the length of time from the start to the end of the process) of the process may be prolonged. In cases of emergencies such as wildfires and floods, the system also needs to help participants stay away from dangerous locations. *System-driven strategies* were motivated by these considerations. In this case, the system provides each participant with an exploration tour within his/her assigned region for him/her to follow during the current process and issues directives as needed to alter the tour. CROSS provides emergency managers with tools for selecting participants from human sensors who responded to its call for participation, assigning selected participants to explore regions of the threatened area and planning for each participant a tour.

This paper focuses on data fusion and processing methods that can help the system to determine the amount of collected data (hence the time required to collect the data) needed for the system to acquire situation awareness. The underlying problem addressed here, called *symbiotic data fusion and processing* (*SDFP*) problem for short, is how to use human sensor data and physical sensor data synergistically to speed up the decision process and/or improve the quality of the decision. Related problems include how to use physical sensor data to help assess the credibility of human sensor data and how to use human sensor data for discovery of erroneous and failed physical sensors. Problems of assessing the credibility and discovering truth of information reported by participatory sensors have been the focal points of intense efforts in many research areas, including machine learning and data mining. We will compare their approaches with the approach presented here in the section on related work.

For many likely scenarios, the system aims to detect from sensor data the occurrences of events and phenomena that warrant its actions. This is why we emphasize here statistical detection formulations [12-16] and solutions of the SDFP problem, rather than estimations of parameters that define the state of the area of interest. In terms of objective, our problem resembles the problems of improving the coverage of physical sensors such as the one studied by Xing, Tan, *et al*. [12]: Like them, our system also wants to improve its coverage, except that our system uses human sensor data to reduce the limitations of physical sensor coverage. Wang, Abdelzaher, *et al*. [17] applied the *expectation–maximization* (*EM*) algorithm [18] to find maximum likelihood estimates that quantify the correctness of binary-valued human sensor data from unknown crowd. Our system can also use the EM algorithm to process data from unknown human sensors for which the statistical model contains unknown parameters.

The work described here makes two contributions: First, our work is among the first, if not the first, to characterize and treat data from both physical sensors and human sensors used for surveillance and monitoring purposes in a coherent way. Our realistic, yet formal model of surveillance systems containing physical in-situ sensors and mobile human sensors enables us to build solutions of the problems in fusing physical and human sensor data on the rigorous foundation of stochastic detection and estimation theory. For many real-life scenarios, the solutions provide the system with not only quantitative assessment of its decision quality but also control over tradeoffs between conflicting quality criteria. The second contribution is the design of a fusion unit for processing data collected from all sensors and determining when to initiate and when to terminate CDC processes dynamically depending on whether the sensor coverage is sufficiently good. The design of this essential component is built on the solutions presented here. We will make the fusion unit a part of CROSS and thus make CROSS a full-function crowdsourcing support platform.

Following this introduction, Section 2 presents our assumptions on the surveillance system and its physical sensors. It also presents models of disaster threatened areas, physical sensors and participants of the CDC process. Section 3 presents the design and implementation of a central fusion unit for processing sensor data and making decisions. We use it to explain how various fusion and statistical decision techniques may be used to help the system manage CDC processes, specifically, how the system decides when it has collected sufficient data and hence can terminate the current CDC process. Section 4 presents statistical detection and estimation formulations of the SDFP problem, variants of which are what the system needs to solve. Section 5 discusses related work. Section 6 summarizes the paper and discusses future work.

2 SCENARIOS, ASSUMPTIONS AND MODELS

Figure 1 shows four representative scenarios. We use them to support our assumptions and motivate our models and problem formulation. Only part (a) of the figure shows physical sensors: They are surveillance cameras. Physical sensors in other parts are omitted in order to keep the figure simple. Small circles in the figure represent locations where human sensor data are needed. It is convenient to think that there is a virtual sensor at each of these locations: During a CDC process, each *virtual sensor* provides the system with human sensor data sent by participants in a neighborhood of specified size around the sensor.

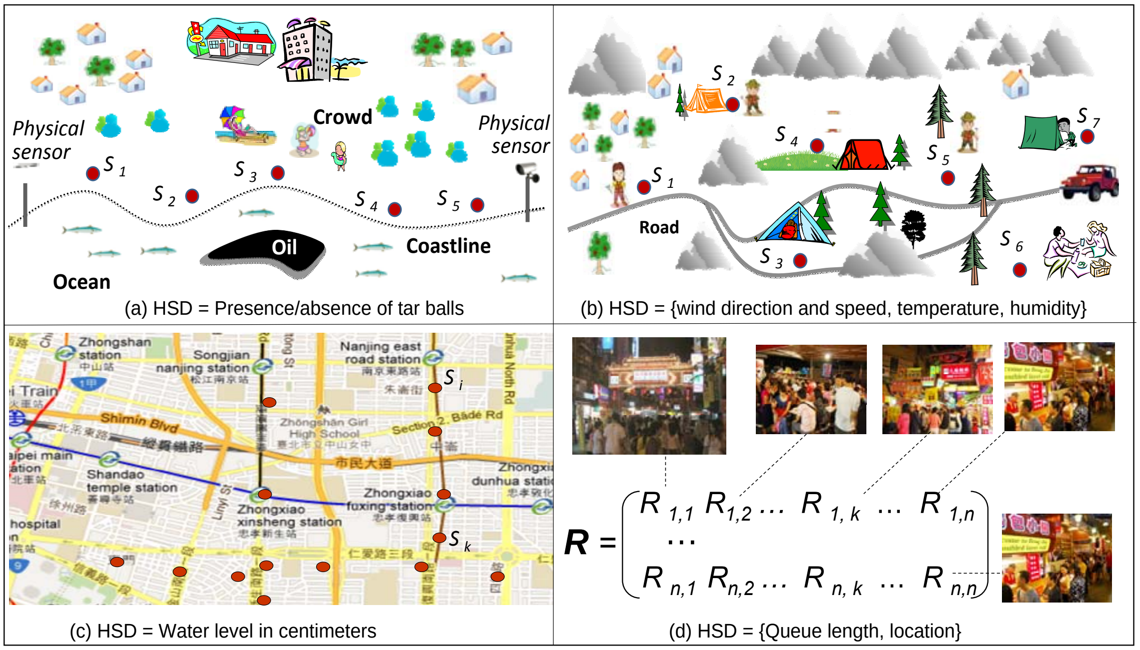

Figure 1 Representative scenarios

Except for where it is stated otherwise, we use the term (α, β)-coverage [12] to state the desired quality of sensor coverage precisely: In a process to detect a specified condition, the system is said to make a *false alarm* when it declares the condition to be true when the condition is not true. It is said to have detected the condition when it declares the condition to be true when the condition is indeed true. We say that the system has achieved (α, β)-coverage when its false alarm probability is no larger than the *threshold false alarm rate* α ($0 \leq \alpha < 0.5$) and its detection probability is at least equal to the *threshold detection rate* β ($0.5 < \alpha \leq 1$). For given threshold rates α and β chosen by the system, sensor coverage is said to be *sufficiently good* when the system can achieve (α, β)-coverage of the threatened area by processing available data in some way(s). The system starts a CDC process when available physical sensor data must be supplemented by human sensor data to give it sufficiently good sensor coverage.

2.1 Representative Scenarios

The oil spill and wildfire scenarios in parts (a) and (b) of Figure 1 are from [10, 11]. Similar to the litters-in-park case study presented in [17], each human sensor in the oil spill scenario reports a binary value indicating the presence or absence of tar balls at his/her locations. Based on their reports and inputs from physical sensors (e.g., surveillance cameras) nearby, the system decides whether the section of the beach monitored by the sensors is threatened by oil spill and preventive clean up operations should be launched.

In the park and street surveillance scenarios shown in (b) and (c), each human sensor is asked to measure and report value(s) of some environment parameters. (Examples given by the figure are wind speed and direction, temperature and humidity at his/her campsite not far from a wildfire or depth of water on roadway(s) in front of him/her during a downpour). In these cases, human sensor data have arbitrary values. The system makes its determination of whether the campsite is threatened by the wildfire or whether the road will likely to become flooded and impassable, and hence should be closed immediately, based on data from physical sensors and human sensors.

In the scenario illustrated by (d), human sensors are asked to report observed queue lengths in front of food stands in a popular night market. The system generates and displays estimated waiting times of the stands based on reported values and head counts from cameras at entrances of the market. We include this scenario as one of the case studies because we can use it to demonstrate how domain-specific enhancements (in this case, queuing analysis based on the routing matrix $\boldsymbol{R}$ that models the movements of customers in the market) can be incorporated with a general technique. This scenario occurs week after week and hence can give us a convenient, real-life setting for evaluation of ours techniques.

2.2 Models of System, Threatened Area and Sensors

Figure 2 shows the key elements of a symbiotic surveillance sensor system used by a disaster warning system that enhances the quality of its sensor coverage by crowdsourcing. As noted in [10, 11], ideally, the system would have a sufficient number of physical sensors of the right types at the right locations to give it sufficiently good coverage of the threatened area.

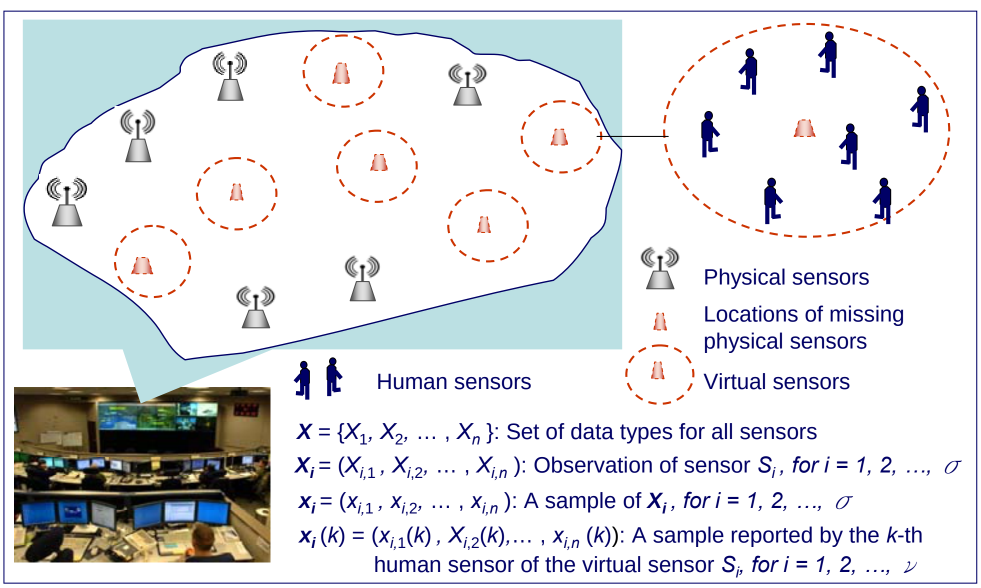

Figure 2 Elements of a symbiotic surveillance sensor system

We confine our attention to the case where the physical sensors at all locations are *functionally identical:* Each of them provides the system with n types of data. (As an example, in the scenario in Figure 1(b), the sensor in every weather station in a national park provides 4 types of data: wind speed, wind direction, temperature and humidity.) We denote the set of data types by $\boldsymbol{X}=\{X_1, X_2, \ldots, X_n\}$ and view a *observation* $\boldsymbol{x_i} = (x_{i,1}, x_{i,2}, \ldots, x_{i,n})$ made by sensor S_i as a *sample value* (or *sample*) of the random variable $\boldsymbol{X_i} = (X_{i,1}, X_{i,2}, \ldots, X_{i,n})$, an instance of the data type $\boldsymbol{X}$. We assume here that the component random variables $X_{i,j}$ and $X_{i,k}$, for $j, k = 1, 2, \ldots n$ and $j \neq k$, are statistically independent for all sensors S_i.

For reasons including deployment costs, damages and poor operating conditions, v physical sensors are missing or broken. When existing physical sensors cannot provide the system with sufficiently good coverage, it starts a CDC process in order to acquire human sensor data on conditions around the location of each missing sensor. We assume that the system knows the identity and location of each missing physical sensor and solicit from human sensors the same data type $\boldsymbol{X}$ as that of the physical sensors. In short, the system has a *virtual sensor* at the location of each missing physical sensor. We use $S_1, S_2, \ldots, S_v$ to denote the virtual sensors, and denote existing physical sensors by $S_{v+1}, S_{v+2}, \ldots, S_\sigma$ when the threatened area should be covered by a total of σ sensors.

Without loss of generality, we assume that an observation $\boldsymbol{x}_i$ is made by every physical sensor S_i, for $i = v+1, v+2, \ldots, \sigma$, immediately before each CDC process starts. The value $\boldsymbol{x}_i$ is the sum of a vector of noise-free observed values plus an additive noise $\boldsymbol{\Theta}_i = (\Theta_{i,1}, \Theta_{i,2}, \ldots, \Theta_{i,n})$. The noise $\Theta_{i,j}$, for $i = v+1, v+2, \ldots, \sigma,$ in the j-th observed values of all physical sensors are statistically independent, identically distributed. We let $A_j(t)$ denote the distribution function of $\Theta_{i,j}$ (i.e., the probability of $\Theta_{i,j} \leq t$) where t is from a scenario-specific set of values.

Let $\boldsymbol{x}_i(k) = (x_{i,1}(k), x_{i,2}(k), \ldots, x_{i,n}(k))$ denote the (*human sensor*) *sample* reported by a k-th participant from a neighborhood around the virtual sensor S_i during the current CDC process. The system computes for each virtual sensor S_i, for $i = 1, 2, \ldots, v$, the observation $\boldsymbol{x}_i$ and distribution function of $\boldsymbol{X_i}$ of that virtual sensor from the human sensor samples $\boldsymbol{x}_i(k)$ reported by participants around S_i and distribution functions of the samples.

2.3 Models of Participants

The distribution functions of human sensor samples clearly depend on the participants who sent them. Previously, we have considered two types of participants: ideal and motivated ones, called type-I and type-M, respectively [10, 11]. A *type-I human sensor* is likely to be a government disaster responder or a volunteered responder. The person may have been trained. He/she is at least experienced as a human sensor. The system can rely on him/her to generate "ground truth" against which data from other participants can be measured. Upon request by the system, a type-I participant can visit one or more suspicious physical sensors and calibrate and fix them. Hereafter, we will not consider type-I participants, but their presence gives us reason to assume that during a CDC process, all existing physical sensors are functional.

Again, a *type-M participant* is a motivated individual. He/she may be a registered volunteer, a person affected by the disaster, and so on. The system knows him/her and knows that he/she will not lie and will make observations independently of other participants. However, the sensor data collected and reported by him/her may not be accurate. Consequently, the sample values reported by the k-th type-M human sensor at S_i contains an additive error component denoted by $\boldsymbol{\Theta}_i(k) = (\Theta_{i,1}(k), \Theta_{i,2}(k),\ldots, \Theta_{i,n}(k))$. We ignore factors such as technical problems, mass panic, etc. and assume that errors of participants are statistically independent and identically distributed. Moreover, virtual sensors being functionally identical, $\Theta_{i,j}(k)$ is a random variable with distribution function $B_j(t)$ for all virtual sensors and all participants.

3 FUSION AND DECISION PROCEDURE

Except for the presence of human sensor data which need to be treated differently from physical sensor data in some cases, our fusion problem is essentially that of a fusion center in a distributed multiple sensor system. The problem of centralized fusion for multiple physical sensors has been treated extensively since the late 1980's. The *centralized decision fusion* (*CDF*) *procedure* described by the pseudo code in Figure 3 makes use of some of the principles, approaches and methods from literatures (e.g., [14-21]). We will present in the next section specifics on some of the techniques used by the procedure. Here, we use the CDF procedure to provide a context for statements of key assumptions and design rationales and explain some of the work a fusion center needs to do to fuse and process all sensor data during a CDC process. The prototype CROSS fusion unit is structured as the procedure. Hereafter, we also refer to the unit as the system when there is no need to be specific.

3.1 Assumptions

The version of CDF procedure in Figure 3 makes several assumptions. An implicit assumption is that the central fusion unit using the procedure is a decision module: Its mission is to decide whether an object is present, or a phenomenon has occurred, or a specified condition is true and so on. The system takes action according to the decision. Specifically, the procedure is presented in terms of binary hypothesis testing and uses (α, β)-coverage as the quality measure. They can be easily replaced by other commonly used methods (e.g., maximum *a posterior* (MAP) and maximum likelihood (ML) decisions and multiple hypothesis testing), and quality criteria (e.g., probability of error and Bayesian costs). A fusion unit in a general-purpose disaster surveillance system that uses diverse physical sensors to monitor diverse disaster conditions will need to provide a library of these fusion methods, including codes that implement rules for multiple-hypothesis testing to minimize decision error and other optimization criteria.

Centralized Decision Fusion Procedure:

// Check whether coverage of existing physical sensors is sufficiently good

1 do decision fusion for physical sensors using NP hypothesis testing:
   a. for $i = v+1, v+2, \ldots \sigma$, acquire local decision u_i, threshold false alarm rate f_i and detection probability d_i from sensor S_i;
   b. generate likelihood ratio from u_i, f_i and d_i for all $i = v+1, \ldots \sigma$;
   c. for an overall threshold false alarm probability $F = \alpha$, compute decision thresholds $\eta_p \geq \eta^*_p$;
   c. compute overall decision U and detection probably D;

2. if $D \geq \beta$, go to take action according to decision U; // Coverage sufficiently good; CDF procedure ends.

// Start a CDC process

3. broadcast Call-For-Participation; wait for responses;
4. from responded human sensors, select participants and allocate them to v virtual sensors;
5. wait for M_i or more samples $x_i(k)$, for $k = 1, 2, \ldots M_i$ …sent by human sensors from v_i, for all $i = 1, 2, \ldots v$;

// Do fusions and then check whether coverage of all sensors is sufficiently good.

6. do value fusion for virtual sensors: for each virtual sensor S_i, for $i = 1, 2, \ldots v$, do the following
   a. compute from $x_i(k)$, for $k = 1, 2, \ldots M_i$ …, test statistics;
   b. compute local decision u_i, false alarm rate f_i and detection probability d_i;
7. do decision fusion for all sensors using NP hypothesis testing
   a. compute likelihood ratio from u_i, f_i and d_i, for all $i = 1, 2, \ldots \sigma$;
   b. for threshold false alarm probability α, set detection thresholds $\eta \geq \eta^*$;
   c. compute over all decision U and detection probability D;
8. if D is less than β, send updated instruction to human sensors; goto step 5; // Continue to collect human sensor data.
9. terminate the current CDC process; go to take action according to decision U; // CDF procedure ends

Figure 3 Centralized Decision Fusion (CDF) Procedure

Sometimes, the system makes situation assessment and decision on the basis of not only sensor data but also other information. In that case, the system wants to get from the fusion unit estimates of some specified parameters, and the unit needs to solve an *estimation problem*. We leave the discussions on the estimation problem, as well as solutions for multiple-hypothesis testing, to a future paper.

Steps 1 and 2 in Figure 3 are based on an assumption stated earlier: Each time when a disaster warning system needs to acquire situation awareness and make a decision, it first checks whether it can make a decision of sufficiently good quality based on physical sensor data alone. It starts a CDC process only when it is not satisfied with the quality of the decision.

The CDF procedure uses both value fusion and decision fusion. In Step 6, the system does for each virtual server *value fusion* of the observations (samples) reported by human sensors from a neighborhood around the sensor: The system makes a local decision on whether the specified condition is true based on the sample values.

In contrast, the system does decision fusion in Step 1 and 7. The assumption is that value fusion is done by each physical sensor: Based on its own observation, the sensor makes a decision about the condition. The system acquires from the physical sensor S_i its local decision

u_i, false alarm probability f_i and detection probability d_i, not sensor's observation. The system generates the overall decision U and assesses the overall decision quality parameters F and D based on local decisions and false alarm and detection probabilities of physical sensors in Step 1 and of all sensors in Step 7. If after Step 7, the system is not satisfied with the quality of the overall decision, it continues to collect human sensor data, using additional participants and/or redirecting existing participants if necessary.

Finally, a restrictive assumption stated earlier is that observations of all sensors, physical and human, are statistically independent. This common assumption is made for sake of simplicity and cannot be easily removed. We will evaluate how decision quality degrades when observations and decisions of sensors are correlated via simulation as an important part of future work. We also leave the case of heterogeneous sensors to future work.

3.2 Rationales

The version of CDF procedure in Figure 3 uses the *N-P test* (i.e., Neyman-Pearson criterion for binary hypothesis test [19]) in decision fusion steps, and whenever applicable, also for value fusion. The next section will describe the test. A reason for using the test is that it does not require *a priori* probability of each hypothesis. This is important since in almost all cases considered here, the probability of whether a condition of interest is true is not known before observations are made. Another reason is that the N-P test is optimal (i.e., the most powerful test) in the sense that it maximizes the detection probability for a given acceptable false alarm probability. The test provides the system with control over the tradeoff between these conflicting quality measures. This is an also an important advantage for our application.

We focus here primarily on the case where the distributions $B_j\,(t)$'s of sample errors are known for all human sensors. In other words, the system has acquired reasonable good estimates of quality parameters of each human sensor, including an upper bound and a lower bound of his/her false alarm and detection probabilities, respectively. This assumption is valid most of time: As stated earlier, CROSS uses as human sensors type-M participants. Their quality as human sensors can be assessed during disaster preparedness through means such as volunteer registration and training. In this case, the N-P test can also be used in Step 6. The system uses unknown participants only when available type-M participants are insufficient. In that case, the distributions of the noise components in their samples are unknown. Other methods, including the EM algorithm [18], are warranted to iteratively estimate the model parameters and then make local decisions.

Having physical sensors perform value fusion locally is a good design choice for sensors that generate large volumes of raw data. Take high resolution cameras as an example. By deploying cameras along the coastline during a major oil spill, an early warning system aims to reliably detect tar balls on beach threatened by oil as illustrated by Figure 1(a). Rather than having the cameras send images periodically to the system to be processed there centrally, each camera has the capability of detecting conditions (including presence of tar balls) that warrant attention from the system. On a poor visibility day, when the system cannot achieve a high detection probability (say $\beta = 0.99$) for an acceptable false alarm probability (say $\alpha = 0.1$) based on the decisions of the cameras, it starts a CDC process during which participants are divided into groups and each group is responsible for finding tar balls around a virtual sensor located between the cameras. For this and similar scenarios, the decision fusion steps (i.e., steps 1 and 7) can be significantly simplified if all physical and virtual sensors work with the same threshold false alarm rate α. Then the system can conclude that the specified (α, β)-coverage is achieved, hence the CDC process can be terminated, when with the help of human sensor data collected so far when the detection probabilities achieved by three or more sensors are at least equal to β. The next section will justify this statement.

Oftentimes, physical sensors (e.g., sensors in weather stations and water level sensors on traffic light posts in scenarios shown parts (b) and (c) of Figure 1) produce only small amounts of data or have little or no processing power. In this case, centralized value fusion is the only alternative: The system does value fusion for physical sensors as it does for virtual sensors before decision fusion, or alternatively, value fuse sample values from all sensors together. Pure value fusion is known to have better performance (i.e., can achieve a lower overall false alarm probability and/or a higher detection probability) than decision fusion [15].

4 SYMBIOTIC DATA FUSION AND PROCESSING

This section presents variants of the statistical detection problem which the fusion unit must solve. To keep notations simple, we consider only the case of $n = 1$, that is, every sensor observes only one type of data. Extension to the case of $n > 1$ is straightforward because observations/decisions of different data types are statistical independent.

To state the detection problem for both decision fusion and value fusion, we let the M-dimensional random vector $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_M)$ represent the inputs from M independent sources. The fusion unit receives a sample $\boldsymbol{y} = (y_1, y_2, \ldots, y_M)$ of $\boldsymbol{Y}$ containing a sample value y_i of Y_i for every $i = 1, 2, \ldots, M$. In most cases of practical interest, y_i's are discrete valued.

In the context of the CDF procedure, M is equal to the number $\sigma - \nu$ of physical sensors and the number σ of all sensors in Steps 1 and 7, respectively. For these decision-fusion steps, y_i's are local decisions that are independently made by the sensors based on their observations. In Step 6, M is equal to the number of human sensors reporting from a virtual sensor. For each value-fusion step, y_i's are the samples (i.e., observed values) reported by the human sensors.

4.1 Binary Hypothesis Testing with Neyman-Pearson Criterion

In case of binary hypothesis testing, the fusion unit decides whether a hypothesis H_1 (e.g., tar balls on beach, campsite in danger of wildfire) is true or the alternative hypothesis H_0 (e.g., no tar balls on beach, no wildfire danger) is true based on the received sample $\boldsymbol{y}$. A *false alarm* occurs when the unit decides in favor of H_1 when in fact H_0 is true, and the unit successes in *detection* of $H_1$when it decides on H_1 when H_1 is indeed true.

As stated earlier, the *a priori* probabilities of the hypotheses are unknown typically. The fusion unit works with the conditional probability mass functions, which give the conditional probabilities of seeing $\boldsymbol{y}$ given H_1 or H_0 is true, respectively. These functions are known.

$$P(\boldsymbol{y} \mid H_1) = P(y_1, y_2, \ldots, y_M \mid H_1) \equiv \Pr[\boldsymbol{Y} = \mathbf{y} \mid H_1] = \Pi_{1 \le i \le M} P(y_i \mid H_1) \quad (1)$$
$$P(\boldsymbol{y} \mid H_0) = P(y_1, y_2, \ldots, y_M \mid H_0) \equiv \Pr[\boldsymbol{Y} = \mathbf{y} \mid H_0] = \Pi_{1 \le i \le M} P(y_i \mid H_0)$$

The last equality in each line follows from the fact that Y_i's are statistically independent.

The CDP procedure aims to maximize the probability of detection for a given threshold false alarm probability α. It uses the N-P test for reasons stated earlier. The test statistics is the *likelihood ratio* $L(\boldsymbol{y})$ defined in term of the conditional probabilities in (1) or equivalently, the log function of the likelihood ratio:

$$L(\boldsymbol{y}) = P(y_1, y_2, \ldots, y_M \mid H_1) / P(y_1, y_2, \ldots, y_M \mid H_0)$$
$$= \Pi_{1 \le i \le M} P(y_i \mid H_1) / P(y_i \mid H_0) \quad (2)$$

The N-P test has two commonly used decision rules, a deterministic rule and a randomized rule [14]. To state these rules, we let $\boldsymbol{\Lambda}$ denote the set of possible values of $L(\boldsymbol{y})$ for all observed values of $\boldsymbol{y}$, and let η and η^* be two adjacent values in $\boldsymbol{\Lambda}$ which are such that $\eta > \eta^*$ and

$$\Pr[L(\boldsymbol{y}) \ge \eta \mid H_0] \le \alpha \quad (3a)$$
$$\Pr[L(\boldsymbol{y}) \ge \eta^* \mid H_0] > \alpha \quad (3b)$$

Figure 4 illustrate the relationship between η and η^* as well as their relationship with other values in $\boldsymbol{\Lambda}$. The dotted and dashed curves are envelopes of conditional probabilities $\Pr[L(\boldsymbol{y}) \mid H_0]$ and $\Pr[L(\boldsymbol{y}) \mid H_1]$, respectively, for all values of $L(\boldsymbol{y})$.

Deterministic Rule: The *deterministic rule* uses η as the detection threshold and selects

$$H_1, \text{ if } L(\boldsymbol{y}) \ge \eta \quad (4)$$
$$H_0, \text{ if } L(\boldsymbol{y}) < \eta$$

The false alarm probability F and detection probability D achieved by rule (4) and detection threshold η are given by

$$F = \Pr[L(\boldsymbol{y}) \ge \eta \mid H_0] = \Sigma_{\Lambda(\boldsymbol{y}) \ge \eta} P(\boldsymbol{y} \mid H_0) \quad (5a)$$

$$D = \Pr[L(\boldsymbol{y}) \geq \eta \mid H_1] = \Sigma_{\Lambda(\boldsymbol{y}) \geq \eta} P(\boldsymbol{y} \mid H_1) \quad (5b)$$

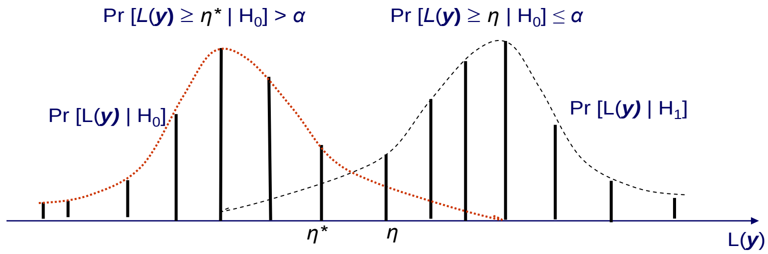

Figure 4 An illustrative example

Randomized Rule The randomized rule uses η and η^* as detection thresholds together with a random selection probability p which is the solution of the equation

$$p \Pr[L(\boldsymbol{y}) \geq \eta^* \mid H_0] + (1-p) \Pr[L(\boldsymbol{y}) \geq \eta \mid H_0] = \alpha \quad (6a)$$

The randomized rule selects

H_1, if $L(\boldsymbol{y}) \geq \eta$ (6b)

H_1 with probability p and H_0 with probability $1-p$, if $L(\boldsymbol{y}) = \eta^*$

H_0, if $L(\boldsymbol{y}) < \eta^*$

The false alarm probability achieved by rule (6) is α by definition of the rule. The detection probability is given by

$$D = p \Pr[L(\boldsymbol{y}) \geq \eta^* \mid H_1] + (1-p) \Pr[L(\boldsymbol{y}) \geq \eta \mid H_1] \quad (7)$$

Optimality It has been shown in [14] that rule (4) is optimal among all deterministic rules. It is also optimal among all rules when $\boldsymbol{y}$ is continuous valued. However, it is not optimal when $\boldsymbol{y}$ and hence $L(\boldsymbol{y})$ are discrete valued. The reason is that there may not be a threshold value in $\boldsymbol{\Lambda}$ for which the equality in (3a) holds. In that case, false alarm probability F achieved by rule (4) is less than α, and the detection probability D may not be the maximum possible under the constraint $F \leq \alpha$.

The randomized rule (6) is optimal: It maximizes the detection probability under the constraint of $F \leq \alpha$. Complex rules using more than two detection thresholds do not work better.

4.2 Binary Hypothesis Testing Based on Binary Valued Samples

The special case of hypothesis testing based on binary-valued samples is of practical importance. In the multiple sensor fusion problem treated in [15], local decisions of individual sensors and overall decision of the fusion center are all binary valued. This is assumed by the CDF procedure. In scenarios similar to the one shown in Figure 1(a), observations from human sensors are naturally binary valued. In other disaster scenarios, the system can also use binary-valued observations. Take scenarios in Figure 1 (b) – (d) as examples. Rather than the data types listed in the figure, the system can ask human sensors to send binary-valued reports indicating whether the condition "wind-is-from-direction-of-wildfire", "water-is-over-curb", or "queues-are-long" is true or not, respectively.

To state of the problem of fusing binary decisions formally, we let $y_i = 1$ when the sensor S_i decides in favor of H_1 and $y_i = 0$ if it chooses H_0. In addition to y_i, the fusion unit also knows the associated detection probability and false alarm probability

$$d_i = \Pr[y_i = 1 | H_1], \quad f_i = \Pr[y_i = 1 | H_0] \tag{8}$$

for all $i = 1, 2, \ldots, M$ sensors.

Similarly, we let $y_i = 1$ and $y_i = 0$ be the possible sample values of the random parameter Y_i reported by the i-th human sensor and say that H_1 is true if $Y_i = 1$ and H_0 is true if $Y_i = 0$. The conditional probabilities d_i and f_i defined in (8) can be used as quality measures of the i-th human sensor. At risk of abusing the terms, we call them detection and false alarm probabilities of the human sensor, respectively. It is easy to see that the problem of fusing binary-valued observations from human sensors is the same as the problem of binary decision fusion when the quality measures d_i and f_i are known for all human sensors.

Similar Sensors of Good Quality A surveillance system is likely to use *similar sensors*, i.e., the sensors are functionally identical, operating at the same threshold false alarm probability $f_i = f$ and achieving detection probability $d_i \geq d$ for all S_i. In this case, it suffices for the fusion unit to compute the test statistics from the number K of 1's among the M inputs y_i*'s.* The conditional probability mass functions of K conditional on H_0 and H_1 are the binomial distributions $B(M, f)$ and $B(M, d)$, respectively. Let k be the sample value of K observed by the fusion unit. The likelihood ratio is given by

$$L(k) = \Pr[K = k | H_1] / \Pr[K = k | H_0]$$
$$= d^k(1-d)^{M-k} / f^k(1-f)^{M-k} \tag{9}$$

By working with $\log L(k)$, the randomized rule of the N-P test simplifies to the following: Let t be an integer in $(0, M)$ which is such that

$$\Pr[k \geq t | H_0] \leq \alpha, \quad \Pr[k \geq t - 1 | H_0] > \alpha \tag{10a}$$

The simplified randomized rule is: select

H_1, if $k \geq t$ (10b)

H_1 with probability p and H_0 with probability $1 - p$, if $k = t - 1$

H_0, if $k < t - 1$

where the selection probability p is given by $p \Pr[k \geq t - 1 | H_0] + (1 - p) \Pr[k \geq t | H_0] = \alpha$.

The following theorem states that the fusion center can conclude that it can achieve the desired (α, β)-coverage if three or more sensors can achieve detection probability β or better with the threshold false alarm probability α:

Theorem 1 *In a system containing $M \geq 3$ similar sensors all of which operate with threshold false alarm probability α, (α, β)-coverage can be achieved using the randomized N-P-test when the detection probability of 3 or more sensors is equal to or higher than β.*

The theorem is based on the theorem and its proof in [15].

Similar Sensors of Poor Quality For our application, the acceptable false alarm rate is in order of 10% for some scenarios and much lower than 10 % for other scenarios. This quality criterion is not met by typical human sensors, and some physical sensors. It is well known that fusion center can achieve (α, β)-*coverage* using the randomized N-P-test even when the false alarm probability f of the individual sensors is larger than α [14]. A question of practical interest is how many sensors are required to get the overall false alarm probability F less than or equal to α. Similarly, we want to know the minimum number of similar sensors required to get an overall detection probability $D \geq \beta$ when their individual detection probability d is less β.

We can find answers to these questions from the expressions of the overall quality measures F and D in terms individual quality measures f and d of M similar sensors. To illustrate, suppose that the fusion center uses the deterministic rule with a single detection threshold t. Then, the overall F and D are given by

$$F = \sum_{t \leq k \leq M} C(M, k) f^k (1-f)^{M-k} \tag{11}$$
$$D = \sum_{t \leq k \leq M} C(M, k)\, d^k (1-d)^{M-k}$$

where $C(M, k)$ denotes the binomial coefficient $M!/k!(M-k)!$. Solving these equations for several likely values of F, D, f and d, we get Figure 5, which plots the minimum numbers of similar sensors required to achieve the overall (F, D)-coverage for several likely combinations of quality measures f and d of individual sensors.

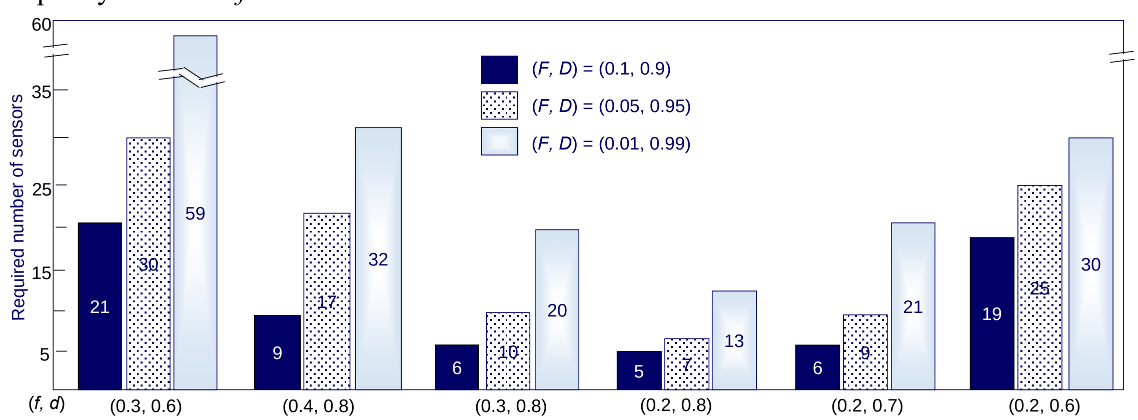

Figure 5 Required numbers of sensors for different f, d, F and D

The numbers of sensors shown Figure 5 are conservative estimates. The reason is that the deterministic rule is sub-optimal. For some of the combinations of f and d, the fusion center can reduce the required numbers of sensors by 1 or 2 from the numbers shown here by using the probabilistic decision rule. Take the case of trying to achieve (0.01, 0.99)-coverage using sensors with $f = 0.2$ and $d = 0.8$ as an example. The figure shows that desired coverage quality is met using 13 sensors. With detection threshold $t = 7$, the actual quality (0.007, 0.9929) exceeds the desired quality. The fusion center can reduce the required number of sensors to 12 by using the probabilistic rule: With detection thresholds 6 and 7 and selection probability $p = 0.8967$, it gets $F = 0.01$ and $D = 0.9945$.

The combinations of figures in Figure 5 tell us what intuition tells us all along: It is far better to use a relative small number of high quality participants (e.g., with $(f, d) = (0.2, 0.8)$) than a big crowd of participants of possibly poorer qualities. By better, we mean it takes the system less time to collect and processor human sensor data to reach a specified overall quality.

4.3 Binary Decisions Based on Arbitrary-Valued Observations

It is straightforward to apply N-P test rules to make optimum binary decisions based on arbitrary-valued observations when their joint distributions are known under both hypotheses. This is especially so when sample values are continuous, because the simple deterministic rule (4) is optimal. A special case of practical importance is when the data are jointly Gaussian under each hypothesis. This model has been widely used to characterize physical sensor data and has been treated extensively in literature, including [12-16, 20].

The problem of making decisions based on human sensor data of arbitrary values is made more complex by two factors. First, errors (e.g., additive noises) in human sensor observations are typically not Gaussian. Uniform distribution and some beta distribution are closer models, especially for data from a few (e.g., <5) human sensors. This makes evaluation of performance of optimum decision rules more complicated. The N-P test rules nevertheless can be applied.

Second, a more challenging complication is the fact that the system often does not know the values of noise-free components of observables from individual sensors and sometimes, not even their distributions. In other words, each of the M observables $Y_i = V_i + \Theta_i$ presented to the fusion unit is the sum of a random noise-free component V_i and a random additive noise Θ_i. The distribution of Θ_i is known for reasons stated earlier. The distribution of V_i may not be fully known under each of the hypothesis.

We would encounter this case in scenarios shown in Figure 1(b) and (c) when human sensors are asked to report wind direction with respect to the direction of wildfire or water depth instead of binary observations "wind-is-in-direction-of-fire" or "water-is-over-curb". –

Due to the effect of microclimate, wind direction, ambient temperature, etc. at each virtual sensor may differ significantly from that of surrounding area. During a downpour, water may accumulate on some road segments due to poor local drainage but not elsewhere. In both cases, the system may not even be able to compute the expected value of the "signal" V_i contained in the sample value from readings of surrounding physical sensors and virtual sensors.

We sometimes can formulate the problem of fusing such observations as a joint binary hypothesis testing and estimation problem treated in [20]: We are given the conditional distribution of the observables $\boldsymbol{Y}$ under each hypothesis $P(\boldsymbol{y} \mid H_0)$ and $P(\boldsymbol{y} \mid H_1, \boldsymbol{\Phi})$, where $\boldsymbol{\Phi}$ is a random parameter with a known probability density (or mass) function. The solution of the problem gives us a rule to decide in favor of H_0 or H_1, and if the decision is in favor of H_1, compute an estimate of $\boldsymbol{\Phi}$. The schemes described in [20] combine the N-P test for binary detection with Bayesian parameter estimation. Both schemes start from a detection step followed by parameter estimation. One scheme repeats the detection step after parameter estimation while the other scheme does not. By choosing whether to repeat the detection step, the system can trade off between the detection probability and estimation accuracy.

As an illustrative example, supposed that the human sensors in Figure 1(c) are asked to report water depth on a street in a small number of city blocks. $\boldsymbol{\Phi}$ is the amount of local rainfall or actual water depth. Its probability density function can be derived/estimated from data on measured or forecast rainfall of the surrounding area and historical records. H_0 is "no flooding danger" and H_1 is "flooding possible" and the action to be taken by the system depends on the estimate of $\boldsymbol{\Phi}$. We will apply and evaluate the schemes for this and other scenarios and report the performance data in a future paper.

5 RELATED WORK

In recent years, platforms such as Sahana and Ushahidi [22, 23] have been used worldwide to support crowdsourcing the collection and dissemination of crisis management information during and after major emergencies. In contrast, typical state-of-the-art disaster surveillance and warning systems do not incorporate crowdsourcing social reports as an integral part of their standard operation procedures. Except for disasters (e.g., [24-26]) that take days, even months, to develop, disaster surveillance and warning applications must be able to acquire situation awareness and made critical decisions within hours, even minutes: To do so, they must be able to process social reports in real-time automatically using relatively simple decision rules and extract from the reports information of good and quantifiable quality. The solutions presented in previous sections aim to meet these needs.

The problems in discovering, extracting, refining and validating the truth/information in social reports contributed by crowds have been addressed by many research communities. Numerous techniques and tools (e.g., [27-40]) based on a wide range of technologies (including machine learning, fuzzy systems, data mining, information retrieval and natural language processing) are now available. During 2010 Haiti and Chile earthquakes, some heuristic combinations of technologies (e.g., [28, 35]) were effectiveness for processing social reports from general crowds, but they require an enormous amount of human effort. Many other tools support semiautomatic processing of social reports in order to reduce human efforts. Such tools can deal with many complicating factors which do not arise in scenarios assumed here and hence, are ignored by our solutions. They still incur manual efforts and time too high to be acceptable for our application. A common shortcoming of most existing solutions is that they cannot provide the system with quantitative quality measures of the extracted information.

Rather than general use cases assumed by previous efforts, our work makes two restrictive but realistic assumptions: First and foremost is that a disaster warning system uses only participants who do not lie, make observations independently, and report observed data as requested. Second, the system acquires the statistical characteristics of noises in human sensor data (at least conservative bounds of false alarm and detection probabilities) as one of its preparedness phase tasks, just as that it knows statistical distributions of noises in physical sensor data. The abstract model of symbiotic sensors in surveillance systems shown in Figure 2 is based on these assumptions. The model treats data from human sensors used by the system to supplement data from in-situ physical sensors in a consistent way. Thus, it enables us to formulate the problem of fusing symbiotic data as stochastic detection problems and build solutions needed by the fusion unit on the rigorous foundation of classical stochastic detection and estimation, in particular, results on multi-sensor fusion [12-16, 18-21].

Our approach resembles the one taken by Wang, *et al*. [17, 41] who are among the first to apply statistical estimation and hypothesis testing techniques to processing social sensor data in order to discover and assess the truth carried by the data. Unlike our model, their models do not capture the symbiotic nature of sensors used by crowdsourcing enhanced disaster surveillance systems: Wang, *et al.* demonstrated via a case study that the EM algorithm [18] can out-perform the Bayesian interpretation scheme and Truth Finder [40, 41] for fusing binary-valued observations. For our application, a shortcoming of these schemes is that they do not give the fusion unit control over tradeoffs between quality measures (i.e., false alarm versus detection probabilities) that cannot be optimized at the same time. This is a reason that we treat the SDFP problem as a detection problem and use the N-P test whenever appropriate. The test is not only optimal when *a priori* probabilities are unknown but also simpler to implement than the EM algorithm. Comparing the performance figures in Figure 5 with simulation data on performance of EM algorithm [17], we note that the N-P test can achieve sufficiently good performance (e.g., $(F, D) = (0.01, 0.99)$) using a comparable number of poor-quality participants (e.g., with $(f, d) = (0.3, 0.6)$) as the number indicated by their data.

Recent studies (e.g., [3, 7-9]) on issues in integrating social sensing with pervasive and ubiquitous computing also consider fusion of data from mobile and ubiquitous sensors with data from social sensors. According to A. Rosi, *et al.* [8], the goals of these studies include extracting social information in order to enhance context-awareness of pervasive services and applications, exploiting social network tools and infrastructures to support some of the data organization and event notification functionalities of pervasive computing, and creating application-specific social-pervasive infrastructures for sensor integration. CROSS can be thought of as an application-specific infrastructure for integration of physical and human surveillance sensors. A difference between the SDFP problem and their fusion problems is that data provided to CROSS by physical and human sensors are of the same types, whereas they provide context-aware pervasive services and applications with data of complementary types.

6 SUMMARY AND FUTURE WORK

A crowdsourcing support system such as CROSS for disaster warning and response purposes not only provides mechanisms and tools for managing crowdsourcing human sensor data collection. It must also provide supports for fusion of data from multiple sensors. We described in the previous sections the work done by a central fusion unit to process and fuse inputs from physical surveillance sensors together with human sensor data collected from participants during a crowdsourcing process. It may use a combination of value fusion and decision fusion in ways exemplified by the CDF procedure, rather than relying solely on value fusion. The goal is to reach a decision of some specified quality or better on action(s) to be taken by the system, and to do so with the fewer human sensor reports, the better.

We have taken a statistical detection and estimation approach. By doing so, we are able exploit well established principles and techniques for fusion in multiple physical sensor systems and focus our attention on incorporating the fusion of human sensor data within a coherent framework with fusion for physical sensors. In the immediate future, we will first evaluate via numerical computations and simulation experiments the alternative solutions based on this approach, including the ones described in the previous sections, for the types of physical and human sensors that are likely to be used in different disaster scenarios, including the ones used in scenarios shown in Figure 1. We will add to CROSS a prototype fusion unit using the CDF procedure and an extensible library of fusion methods as a starting point. After thus enhancing the platform, we will use it for experimental evaluation of our techniques.

The solutions presented in the previous section are merely the tip of an iceberg of sensor fusion methods based stochastic detection and estimation theory. Thus far, our effort has been focused on binary hypothesis testing. A natural next step is to provide the fusion unit with code that applies the *maximum a posteriori* (*MAP*) rule for multiple-hypothesis testing to minimize probability of error and rules for computing parameter estimates according to specified optimization criteria.

A more urgent work, however, is to provide the system with the capability of making decisions and estimates based on data with incomplete models because the system does not know either the noise-free values of observables or the distributions of noises. The former arises in scenarios such as the ones in Figure 1(b) and (c) for reasons discussed in Section 4. We will exploit the optimum tests (e.g., combining N-P test and ML estimation) for joint detection and estimation proposed recently by Moustakides, *et.al*., [21] to build solutions for these scenarios. We have the latter case in scenarios when the system has no choice but to use unknown participants with unknown noise characteristics. Other methods, including the EM algorithm [18], are warranted to first estimate the model parameters and then make local decisions or estimations.

ACKNOWLEDGEMENT

This work was supported by the Taiwan Academia Sinica thematic project OpenISDM (Open Information Systems for Disaster Management).

REFERENCES

[1] List of Crowdsourcing Projects, http://en.wikipedia.org/wiki/List_of_crowdsourcing_projects.

[2] Aggarwal, C. C. and T. Abdelzaher, "Integrating Sensors and Social Networks," in *Social Network Data Analysis,* Ed. by C. Aggarawal, Springer, 2011.

[3] Boulos, M. N. K., *et al.*, "Crowdsourcing, Citizen Sensing and Sensor Web Technologies for Public and Environment Health Surveillance and Crisis Management: Trend, OGC Standards and Application Examples," *International Journal on Health Geographics,* pp 1-67, 2011.

[4] Ferrell, J., "Crowdsourcing Snowstorm's Westward Model Shift," at http://www.accuweather.com/en/weather-blogs/weathermatrix/highres-models-shift-tonights-snow-west/58754 and NoiseTube, at http://noisetube.net/

[5] Saenz, A. "Japan's Nuclear Woes Give Rise to Crowd-Sourced Radiation Maps in Asia and US," at http://singularityhub.com/2011/03/24/, 2011

[6] Demirbas, M., *et al*., "Crowd-Sourced and Collaboration Using Twitter," *Proceedings of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks*, (WOWMOM), 2010.

[7] Breslin, J. G., *et al.,* "Integrating Social Networks and Sensor Networks," W3C Workshop on Future of Social Networking, January 2009.

[8] Rosi, A., *et al.*, "Social Sensors and Pervasive Services: Approaches and Perspectives," *Proc. of PERCOM*, March 2011.

[9] Beach, A., *et al.*, "Fusing Mobile, Sensor, and Social Data to Fully Enable Context-Aware Computing," *Proc. of the 12th ACM Workshop on Mobile Computing Systems and Application* (HOTMOBILE 2011), March 2011.

[10] Chu, E. T. H., Y. L. Chen, J. Y. Lin and J. W. S. Liu, "Crowdsourcing Support System for Disaster Surveillance and Response," *Proceedings of the 15th International Conference on Wireless Personal Multimedia Communication*, September 2012.

[11] Chu, E. T.-H., Y.-L. Chen, J. W. S. Liu and J. K. Zao, "Strategies for Crowdsourcing for Disaster Situation Information," *WIT Transactions on the Built Environment,* Vol.119, pp. 257-269, 2011.

[12] Xing, G., R. Tan, B., et al., "Data Fusion Improves the Coverage of Wireless Sensor Networks," *Proceedings of MibiCom*, 2009.

[13] Gallager, R. G.,*Stochastic Processes: Theory for Applications,* November 21, 2012, http://www.rle.mit.edu/rgallager/notes.htm,

[14] A. S. Willsky, G. W. Wornell, and J. H. Shapiro, "Stochastic processes, detection and estimation," MIT 6.432 Course Note, Fall 2003.

[15] Thomopoulos, S. C. A., R, Viswanathan, and D. C. Bougoulias, "Optimal decision fusion in multiple sensor systems," *IEEE Transactions on Aerospace and Electronic Systems*, Vol.23, No. 5, pp. 644 – 653, September 1987.

[16] Chair, Z. and P. K. Varshney, "Optimal Data Fusion in Multiple Sensor Detection Systems," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 22, No. 1, January 1986.

[17] Wang, D., H. Le, L. Kaplan, and T. Abdelzaher and "On Truth Discovery in Social Sensing: A Maximum Likelihood Estimation Approach," *Proceedings of the 11th*

international conference on Information Processing in Sensor Networks, 2012.

[18] Gupta, M. R. and Y. Chen, “Theory and use of EM algorithm,” Foundation and Trends in Signal Processing, Vol. 4, No. 2, 2010

[19] Scott, C., “Neyman-Pearson criterion,” h.ttp://cnx.org/content/m11548/latest/

[20] Moustakidis, G. V., “Joint detection and estimation: optimum tests and applications,” *IEEE Transactions on Information Theory,* Vol. 58, No. 7, July 2012.

[21] Nakamura, E. F., A. A. F. Loureiro, and A. C. Frery, “Information Fusion for Wireless Sensor Networks: Methods, Models and Classification,” ACM Computing Survey, Vol. 39, No. 3, August 2007.

[22] Ushahidi, an open source project for crowd sourcing crisis management, http://www.ushahidi.com/ and Mission4636, http://www.mission4636.org/

[23] Sahana, an open source disaster management system, http://www.sahanafoundation.org/

[24] Vieweg, S., *et al*, “Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness,” *Proceedings of the 28th international ACM Conference on Human factors in Computing Systems*, 2010.

[25]Deep Horizon Oil Spill, http://en.wikipedia.org/wiki/Deepwater_Horizon_oil_spill and http://www.google.com/crisisresponse/oilspill/

[26] 2009 California Wildfire, http://en.wikipedia.org/wiki/2009_California_wildfires

[27] Nincompoop, Statistical Tools for Extracting Ground Truth from Crowdsourced Data Sets, http://code.google.com/p/nincompoop/

[28] OpenCalais, A toolkit for Incorporating State-of-the-Art Semantic Functionality into Blog, Content Management System, Website or Application, http://www.opencalais.com/

[29] C. C. Aggarwal and T. Abedelzaher, “Social Sensing,” in *Managing and Mining Sensor Data*, Springer, 2013.

[30] G. Barbier, “Maximizing Benefits of Crowdsourced Data,” *Computational and Mathematical Organization Theory*, Vol. 18, No. 3, August 2012

[31]C. Wagner, “Extracting Meta-Measures from Data for Fuzzy Aggregation of Crowd Sourced Information,” *Proceedings of IEEE Conference on Fuzzy Systems*, June 2012.

[32] L. P. Cox, “Truth in Crowdsourcing,” *IEEE Journal on Security and Privacy*, Vol. 9, No. 3, September-October 2011.

[33] F. Gilbert, *et al.*, “YouProve, Authenticity and Fidelity in Mobile Sensing, ” *Proceedings of ACM Embedded and Networked Sensor Systems,* November 2011

[34] M. Lease, "On Quality Control and Machine Learning," *Proceedings of the 3rd Human Computation Workshop*, 2011.

[35] "Developing Swift River to Validate Crowd Sourcing," http://irevolution.wordpress.com/2009/04/10/developing-swift-river-to-validate-crowdsourcing/, April 2010.

[36] Ipeirotis, P. G., *et al*, "Quality Management on Amazon Mechanical Turk," *Proceedings of ACM SIGKDD Workshop on Human Computation,* 2010.

[37] Le, J., *et al,* "Ensuring Quality in Crowdsourced Search Relevance Evaluation: The Effects of Training Question Distribution, *Proceedings of the SIGIR 2010 Workshop on Crowdsourcing for Search Evaluatio*n, 2010.

[38] Zhu, D, and B. Carterette, "An Analysis of Assessor Behavior in Crowdsourced Preference Judgments," *Proceedings of the SIGIR 2010 Workshop on Crowdsourcing for Search Evaluation,* 2010.

[39] Hsueh, P.Y., *et al*, "Data Quality from Crowdsourcing: A study of Annotation Selection Criteria," *Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing*, 2009.

[40] X. Yin, H. Han, and P. S. Yu, "Truth Discovery with Multiple Conflicting Information Providers on the Web," *IEEE Trans. on Knowledge and Data Engineering,* June 2008.

[41] D. Wang, *et al*., "On Bayesian Interpretation of Fact-Finding in Information Networks," *Proceedings of International Conference on Information Fusion, 2011.*