Sensor Fusion Using Dempster-Shafer Theory

Article · March 2003
Source: CiteSeer

2 authors:

Huadong Wu
Carnegie Mellon University
15 PUBLICATIONS **547** CITATIONS

SEE PROFILE

Mel Siegel
Carnegie Mellon University
160 PUBLICATIONS **2,455** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project ProbeSight View project

Project Visual System on Depth Perception and Stereoscopic Vision View project

Sensor Fusion Using Dempster-Shafer Theory

Huadong Wu[1*], Mel Siegel[2(contact author)], Rainer Stiefelhagen[3], Jie Yang[4]

[1,2]Robotics Institute, Carnegie Mellon University

5000 Forbes Ave., Pittsburgh, PA 15213, [2]phone: (412)-268-8742;

[3]Interactive Systems Laboratories, University of Karlsruhe, Germany

[4]Interactive Systems Laboratories, Carnegie Mellon University, PA,

E-mail: [1]whd@cmu.edu, [2]mws@cmu.edu, [3]stiefel@ira.uka.de, [4]yang+@cs.cmu.edu,

ABSTRACT

Context-sensing for context-aware HCI challenges the traditional sensor fusion methods with dynamic sensor configuration and measurement requirements commensurate with human perception. The Dempster-Shafer theory of evidence has uncertainty management and inference mechanisms analogous to our human reasoning process. Our Sensor Fusion for Context-aware Computing Project aims to build a generalizable sensor fusion architecture in a systematic way. This naturally leads us to choose the Dempster-Shafer approach as our first sensor fusion implementation algorithm. This paper discusses the relationship between Dempster-Shafer theory and the classical Bayesian method, describes our sensor fusion research work using Dempster-Shafer theory in comparison with the weighted sum of probability method. The experimental approach is to track a user's focus of attention from multiple cues. Our experiments show promising, thought-provoking results encouraging further research.

Keywords: intelligent sensors, sensor fusion, context-aware, Dempster-Shafer theory

1. INTRODUCTION

The ultimate goal of context-aware computing is to have computers understand our real physical world. We envisage "smart environments", where human-computer-interactions (HCI) feel natural, as if we were communicating with human assistants or service personnel. This seems an impossible task today, as it faces a two-fold challenge (1) how properly to represent our colorful world, with its abstract concepts and unpredictable human feelings, in a computer understandable way, and (2) how to design and deploy sensors that can sense all the clues and content and map them into the context representation.

Researchers in the context-aware computing community are now challenging this seemingly insurmountable task, starting with simple context information, such as a user's location. The argument we offer is that (1) although a user's situational context may include complex information, we can always try to decompose the hard-to-explain complex information contents into simpler and less abstract information pieces; (2) among various relevant information sources about a user's activities and intentions and the current environment, some pieces of information (such as location and user identification) have already been demonstrated (see background research of [1][2]) to be both useful and not too difficult to implement.

A user's context content may include various aspects of relevant information; meanwhile, different sensors have different measurement targets, different resolutions and accuracies, and different data rates and formats. Thus, the mapping from sensors' output to context information can be extremely complicated. Generalized solution do not exist, and systematic research to asymptotically approach it has hardly begun yet. Our research aims to push forward in this direction. Our current work mainly deals with intermediate and higher level symbolic or modality fusion vs. lower-level signal processing.

In this paper, we deal with simplified situations where we assume that the context information can be represented by discrete symbols or numbers, and the mapping from sensor output data to the context representation data structure is well defined. Specifically, our starting point is that the sensors we will use can generate context information fragments that comply with our context representation specifications. They report both context information and the corresponding confidence estimations. However, sensor outputs often have overlaps and conflicts, sensors are highly distributed, and sensor configuration is very dynamic (sensors come and go, sensors' performance varies all the time). Our goal is to build a system framework that manages

---

information overlap and resolves conflicts. This system provides generalizable architectural support that facilitates sensor fusion, i.e., the sensor-to-context mapping process.

Our approach to achieve this goal is to use layered and modularized system design, in which the sensed context information is separated from the sensors' realization, and the sensor fusion process is analogous to the human perception and reasoning processes. Using the Dempster-Shafer theory of evidence algorithm as our baseline sensor fusion approach reflects this analogy.

2. CONTEXT SENSING APPROACH

2.1. Sensor fusion architecture

The sensor fusion system architecture is illustrated in Figure 1. Such a system usually has one central context data repository for each major entity (e.g., a user, a conference room, etc.) to collect all the relevant context information about that entity. A context data repository usually collects several aspects of context information, each aspect of context information has its own "Sensor Fusion Mediator" to collect that kind of context information. Finally, a Sensor Fusion Mediator is responsible for collecting and monitoring the status of its corresponding sensors.
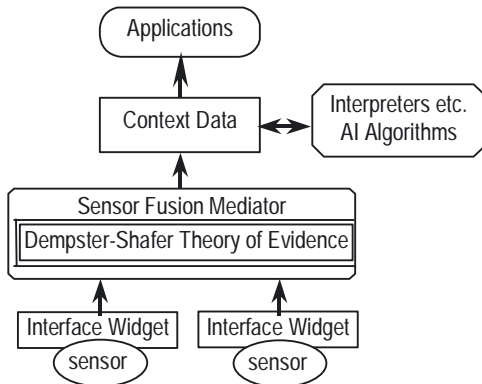

Figure 1. System architecture for sensor fusion of context-aware computing

The interaction between sensors and the system is through each sensor's "Interface Widget". An Interface Widget collects the sensor raw output data and translates it into context information templates defined in the context representation data structure. All the sensed context information will include a pair of numbers that indicate this sensed message's confidence interval; and for the items with dynamic content, there will be a time stamp to indicate when this information was updated. For example, an individual speaker user-identification sensor might decide that the current speaker is user-A with confidence interval of [0.5,

0.7], or else be user-B with confidence interval of [0.3, 0.5]. It will then report through its Interface Widget in the format:

```
Context.Location[room-NSH-A417].People = {
  {name=user-A, confidence=[0.5,0.7],
   proximity=inside, background=Background[user-A],
   ..., time=update-time}, (OR)
  {name=user-B, confidence=[0.3,0.5],
   proximity=inside, background=Background[user-B],
   ..., time=update-time}, ...}
```

2.2. Dempster-Shafer sensor fusion algorithm

The Bayesian theory is the canonical method for statistical inference problems. The Dempster-Shafer decision theory is considered a generalized Bayesian theory. It allows distributing support for proposition (e.g., this is user A) not only to a proposition itself but also to the union of propositions that include it (e.g., "this is likely either user A or user B"). In a Dempster-Shafer reasoning system, *all* possible mutually exclusive context facts (or events) of the same kind are enumerated in "the frame of discernment Θ". For example, if we know that there is person in an instrumented room, and we want to recognize whether s/he is the already-registered user A, user B, or somebody else, then our "frame of discernment" about this person is:

$$\Theta = \{A, \ B, \ \{A, B\}, \{\text{somebody else}\}\} \qquad \text{EQ.1}$$

meaning s/he is "user-A", "user-B", "either user-A or user-B", or "neither user-A nor user-B, must be somebody else"[1].

Each sensor, sensor S_i for example, will contribute its observation by assigning its beliefs over Θ. This assignment function is called the "probability mass function" of the sensor S_i, denoted by m_i. So, according to sensor S_i's observation, the probability that "the detected person is user A" is indicated by a "confidence interval"

$$[\text{Belief}_i(A), \text{Plausibility}_i(A)] \qquad \text{EQ.2}$$

The lower bound of the confidence interval is the belief confidence, which accounts for all evidence E_k that supports the given proposition "user A":

$$\text{Belief}_i(A) = \sum_{E_k \subseteq A} m_i(E_k). \qquad \text{EQ.3}$$

The upper bound of the confidence interval is the plausibility confidence, which accounts for all the observations that do not rule out the given proposition:

$$\text{Plausibility}_i(A) = 1 - \sum_{E_k \cap A = \phi} m_i(E_k). \qquad \text{EQ.4}$$

---

[1] This may be a "null" item, set to zero, to mean this situation is out of the question. We discuss this in the context of normalization later.

For each possible proposition (e.g., user-A), Dempster-Shafer theory gives a rule of combining sensor S_i's observation m_i and sensor S_j's observation m_j:

$$(m_i \oplus m_j)(A) = \frac{\sum\limits_{E_k \cap E_{k'}=A} m_i(E_k)\,m_j(E_{k'})}{1 - \sum\limits_{E_k \cap E_{k'}=\phi} m_i(E_k)\,m_j(E_{k'})} . \qquad \text{EQ.5}$$

This combining rule can be generalized by iteration: if we treat m_j not as sensor S_j's observation, but rather as the already combined (using Dempster-Shafer combining rule) observation of sensor S_k and sensor S_l.

Compared with Bayesian theory, the Dempster-Shafer theory of evidence feels closer to our human perception and reasoning processes. Its capability to assign uncertainty or ignorance to propositions is a powerful tool for dealing with a large range of problems that otherwise would seem intractable.

2.3. Further Push: Weighted Dempster-Shafer Evidence Combination Rules

Implementation of equation EQ.5, the Dempster-Shafer combination rule, implies that we trust sensor S_i and S_j equally. This "equally trusting" approach can cause problems if our sensor fusion system is not properly designed. One often quoted example is that combining Dr. A's "disease I with 99% confidence and disease II with 1% confidence" judgment with Dr. B's "disease III with 99% confidence and disease II with 1% confidence" judgment will lead to a conclusion of "disease II", which common sense says is unlikely to be true.

The "equally trusting" approach in Dempster-Shafer evidence combination is suitable only for situations when both observations have the same accuracy estimates or in situations where their probability assignments over the frame of discernment can quantitatively reflect the ignorance going with their observations.

Because Dempster-Shafer theory are often used to deal with problems that the classical Bayesian method cannot deal with, e.g., due to lack of proved probability distribution model or due to unavailability of accurate mathematical analysis, in many systems using Dempster-Shafer theory, the "probability" numbers are in fact simply assigned by expert opinion [8].

In the process of building a generalizable sensor fusion architecture working with sensors of different accuracy, it is often difficult to require all sensor Widgets correctly to report their observation accuracy along with appropriate ignorance estimation. To approach this problem, we propose a new concept for a *"weighted Dempster-Shafer evidence combining rule"*. The basic idea is this: suppose we know how a sensor performs historically in similar situations; we

can then use the historically-estimated correctness rate as the reference to decide how much we trust the sensor's current estimation from its current observation.

Let m_i be sensor S_i's observation, and let w_i be its corresponding estimation correctness rate in history; let m_j be sensor S_j's observation, and let w_j be its corresponding estimation correctness rate in history. Then, our weighted Dempster-Shafer evidence combination rule — generalization of equation EQ.5 — becomes:

$$(m_i \oplus m_j)(A) = \frac{\sum\limits_{E_k \cap E_{k'}=A} [w_i m_i(E_k) \cdot w_j\, m_j(E_{k'})]}{1 - \sum\limits_{E_k \cap E_{k'}=\phi} [w_i m_i(E_k) \cdot w_j\, m_j(E_{k'})]} \qquad \text{EQ.6}$$

We choose the Dempster-Shafer theory of evidence algorithm as of our baseline for sensor fusion approach because of its flexibility and its intuitive reasoning style. We modify the evidence combination rule pragmatically to handle cases of sensors in which we hand unequal confidence. The following section describes a case study using the Dempster-Shafer approach in our sensor fusion architecture.

3. EXPERIMENTS AND DATA ANALYSIS

Our first attempt to implement the idea of a generalizable sensor fusion framework is through using software agents to simulate sensors' interaction with their Interface Widgets. The software-simulated video and sound sensors read in pre-recorded data and submit reports to the system Sensor Fusion Mediator as if they were generated in real time. The original data were used in [4], which reports observations of four conference-meeting participants' focus of attention.

3.1. Experiments of Tracking Meeting-Participants' Focus of Attention

As shown in Figure 2, an omni-directional camera set at the middle of a conference table was used to capture the activities of the four participants in a meeting. With the omni-camera's panoramic video image sequences, a skin-color-based face detector [6] was used to detect the face location and then the participant's head pose was estimated from the perspective user views via neural network algorithms. Along with the recording of panoramic video sequences, there was one microphone in front of each participant to record who was speaking at that moment.

Assuming that participant S's focus of attention is on another participant T, then S's head pose would pan towards T, with the pan angle following a Gaussian distribution. From previous statistical observation data, the pan-angle Gaussian model parameters were first calculated (prior probability given focus of attention is already known). Then,

given pan angle observation in real time (Pan_S) the posterior probability of S's focus of attention on T

$$P(\text{Foc}_S = T \mid \text{Pan}_S) \qquad \text{EQ.7}$$

was derived using Bayes formula [4]. For each participant S, the probabilities of S's focus of attention being on each of the other participants T_i were calculated.

For a system <u>with video input only</u>, the participant T_j with the biggest measured probability would be deemed to have S's focus of attention.

For all participants in the meeting, the information regarding how likely it was for person S to look at one of the others, T_i, based on who was/were speaking, (A, a vector that indicates four participants' talking status) was collected and tabulated. Then, not only regarding who was/were speaking at the current moment (A^t), but also regarding the status tracing back a short-time in history (A^{t-1}, A^{t-2}, …, A^{t-n}), this information was used to predicate participant S's focus of attention on one of the other participants T as:

$$(\text{Foc}_S = \text{T} \mid A^t, A^{t-1}, \ldots, A^{t-n}). \qquad \text{EQ.8}$$

For a system <u>with audio input only</u>, the participant T_i with the biggest probability number would be deemed to have S's focus of attention.
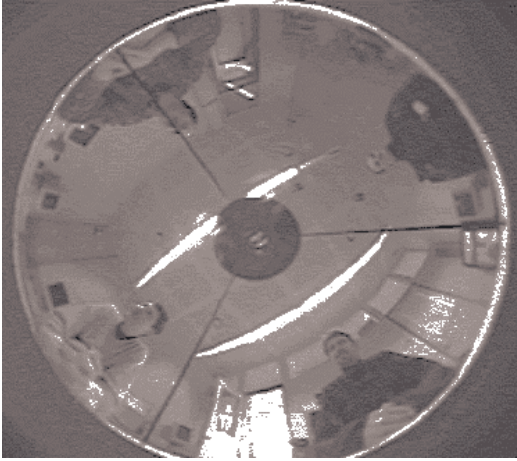

Figure 2. Settings of four users in meeting viewed from the omni camera set at the center of the table.

3.2. Sensor Fusion Case-Study: Theoretical Comparison between Dempster-Shafer and Bayesian Method

Given <u>both the video and audio</u> observation reports, the system Sensor Fusion Mediator's task is to intelligently combine the two inputs to generate a better focus of attention estimation. To obtain a baseline for comparison, let us first check how the classical Bayesian sensor fusion method would deal with this problem.

Let $P_{S\text{-prior}}(T_i)$ and $P_S(T_i)$ be the prior and posterior (given the omni camera's observation $\text{Pan}_S = \text{Pan}$ and the microphones' observation A^t, A^{t-1}, …, A^{t-n}) probability that

person S's focus of attention is likely on the person T_i respectively. Then the Bayesian sensor fusion inference method will conclude that:

$$P_S(T_i) = \frac{P_{S-prior}(T_i)P(\text{Pan}_S = \text{Pan}, A^t, A^{t-1}, \ldots, A^{t-n} \mid \text{Foc}_S = T_i)}{\sum_l P(\text{Pan}_S = \text{Pan}, A^t, A^{t-1}, \ldots, A^{t-n} \mid \text{Foc}_S = T_l)}$$

$$\text{EQ.9}$$

In the case that the camera and microphone sensor observations are independent conditionally on S's true focus-of-attention $\text{Foc}_S = T_i$:

$$P(\text{Pan}_S = \text{Pan}, A^t, A^{t-1}, \ldots, A^{t-n} \mid \text{Foc}_S = T_i)$$
$$= P(\text{Pan}_S = \text{Pan} \mid \text{Foc}_S = T_i) \cdot$$
$$P(A^t, A^{t-1}, \ldots, A^{t-n} \mid \text{Foc}_S = T_i)$$

$$\text{EQ.10}$$

$$\sum_l P(\text{Pan}_S = \text{Pan} \mid \text{Foc}_S = T_l)P(A^t, A^{t-1}, \ldots, A^{t-n} \mid \text{Foc} = T_l) =$$
$$[\sum_j P(\text{Pan}_S = \text{Pan} \mid \text{Foc}_S = T_j)][\sum_k P(A^t, A^{t-1}, \ldots, A^{t-n} \mid \text{Foc} = T_k)]$$

$$\text{EQ.11}$$

The posterior probability EQ.8 can be rewritten as:

$$P_S(T_i) = P_{S-prior}(T_i) \cdot$$
$$\frac{P(\text{Pan}_S = \text{Pan} \mid \text{Foc}_S = T_i)}{\sum_j P(\text{Pan}_S = \text{Pan} \mid \text{Foc}_S = T_j)} \cdot$$
$$\frac{P(A^t, A^{t-1}, \ldots, A^{t-n} \mid \text{Foc}_S = T_i)}{\sum_k P(A^t, A^{t-1}, \ldots, A^{t-n} \mid \text{Foc}_S = T_k)}$$

$$\text{EQ.12}$$

Notice that the second and third item on the right side of equation EQ.12 are posterior probability of camera-only and microphones-only observation. Furthermore, comparing Dempster-Shafer evidence combination rule EQ.5 with this implementation, it can be concluded that the Dempster-Shafer theory of evidence actually reduces to Bayesian inference without a priori knowledge. This happens only under the conditions that (1) the sensors' observation are independent conditionally on the focus-of-attention grand truth; and (2) in the cases where the sensors' observation reports do not include non-zero probability assigned to any proposition union (such as $\{T_i, T_j\}$).

Careful examination will also lead to a very interesting discovery: the weights w_i and w_j in our weighted Dempster-Shafer evidence combination rule EQ.6 correspond conceptually to the prior probability $P_{S\text{-prior}}(T_i)$ in Bayesian rule EQ.12, which just confirms that, assuming the system behavior is time-invariant stochastic process, its historical probability distribution is a priori estimation in the Bayesian inference rule.

3.3. Experimental Data Analysis

In the baseline work [4], a weighted linear sum method was used to combine probabilities as:

$$P(\text{Foc}_S = T) = (1\text{-}\alpha)P(\text{Foc}_S = T \mid \text{Pan}_S) +$$
$$\alpha P(\text{Foc}_S = T \mid A^t, A^{t\text{-}1}, ..., A^{t\text{-}n}) \qquad \text{EQ.13}$$

Here four more experiments' data with the baseline method, the standard Dempster-Shafer method and the weighted Dempster-Shafer method sensor fusion results are shown in Table 1.

In the Table 1, the "valid frames" indicates the number of frames that both video and audio sensors have valid estimations regarding the person's focus-of-attention in the experiment. The "audio correct" and "video correct" columns show percentages of correct estimations with audio-only and video-only sensors' contribution respectively.

Table 1. Meeting participants' focus-of-attention estimation accuracy

	person	valid frames	audio correct	video correct	linear sum correct	DS correct	weighted DS correct
Experiment Set2	#0	1229	55.6%	70.5%	70.1%	70.0%	71.4%
	#1	1075	61.5%	66.2%	69.8%	70.0%	69.4%
	#2	1098	66.1%	78.3%	80.2%	80.8%	80.2%
	#3	991	68.8%	60.3%	65.6%	66.6%	70.0%
Experiment Set5	#0	768	73.8%	74.4%	76.8%	77.0%	77.0%
	#1	956	67.6%	68.5%	72.0%	72.3%	72.1%
	#2	1006	73.2%	83.0%	84.1%	84.2%	83.9%
	#3	929	53.3%	67.9%	75.7%	76.9%	73.2%
Experiment Set6	#0	799	59.1%	70.6%	71.2%	71.5%	71.0%
	#1	751	63.3%	84.6%	85.5%	85.8%	85.2%
	#2	827	57.4%	82.2%	83.3%	84.3%	83.4%
	#3	851	60.8%	80.6%	81.9%	82.3%	81.7%
Experiment Aufnahme2	#0	653	73.2%	84.2%	85.0%	85.0%	84.2%
	#1	653	57.3%	53.5%	54.2%	54.2%	54.5%
	#2	681	72.7%	65.6%	69.5%	69.3%	70.3%
	#6	435	85.3%	75.6%	78.2%	78.4%	79.8%
summary		13702	64.6%	72.8%	75.8%	75.4%	75.4%

The "linear sum correct" column shows the focus-of-attention estimation correctness rate with the sensor fusion algorithm that uses weighted probability linear combination method of equation EQ.13, with parameter $\alpha = 0.5$. It can be seen that though not true in every case, overall, the linearly combined estimation accuracy is better than the results of using only one kind of sensors.

Using Dempster-Shafer theory of evidence algorithm (equation EQ.5) to implement sensor fusion, which is actually reduced to classical Bayesian method[2], the results is very close to that of linear combination method as shown in the "DS correct" column in Table 1.

Finally, the "weighted DS correct" column in Table 1 shows the estimation correctness rate resulting from using weighted Dempster-Shafer evidence combination rule EQ.6 to fuse the audio and video inputs.

4. CONCLUSION AND AND FUTURE RESEARCH

Table 1 summarizes that, for this meeting participants' focus-of-attention analysis application example, sensor fusion using either probability linear combination method, Dempster-Shafer theory of evidence combination method, or the weighted Dempster-Shafer theory of evidence combination method shows a noticeable overall estimation accuracy improvement over any single sensor modality, but all three sensor fusion methods have very similar performance.

Now notice that these sensor fusion performances are measured under the conditions where both microphones' "who have been talking" observation and camera's head-pose observation give valid focus-of-attention estimations. From the fact that valid frame number for each meeting participant varies substantially in one experimental data set, it seems that our generalizable sensor fusion architecture will achieve much higher performance improvement in real applications if our sensor fusion mediator is able to correctly detect the validity of the sensors' output in real time. Our preliminary evidence supports this capability.

In continuing research, we are experimenting with integrating a larger number of sensors in a dynamic configuration. As a theoretically-generalized Bayesian inference method, or as an advanced extension of the canonical Bayesian theory, we expect the Dempster-Shafer theory of evidence algorithm, especially our proposed implementation of weighted Dempster-Shafer evidence combination rule, to demonstrate performance enhancement with such situations.

5. REFERENCES

[1]   Huadong Wu, thesis proposal: "Supporting Sensor Fusion for Context-aware Computing", July 2001,
http://www.cs.cmu.edu/~whd/ContextAware/Proposal.pdf

---

[2] As described in the Section 0, here, the audio and video sensor observations are assumed conditionally independent and they only give non-union proposition probability distribution estimations.

[2]   Anind K. Dey, "Providing Architecture Support for Building Context-Aware Applications", PhD thesis, November 2000, Georgia Institute of Technology, http://www.cc.gatech.edu/fce/ctk/pubs/dey-thesis.pdf.

[3]   Enrique H. Ruspini, John D. Lowrance, Thomas M. Strat, "Understanding Evidence Reasoning", Technical Note 501, December 1990, Artificial Intelligent Center, Computer and Engineering Sciences Division, SRI International, Menlo Park, California 94025, USA

[4]   Rainer Stiefelhagen, Jie Yang, Alex Waibel, "Estimating Focus of Attention Based on Gaze and Sound", Proceedings of Workshop on Perceptive User Interfaces PUI 2001, Orlando, Florida, USA

[5]   Lawrence A. Klein, "Sensor and Data Fusion Concepts and Applications" (second edition), SPIE Optical Engineering Press, 1999, ISBN 0-8194-3231-8

[6]   Jie Yang and Alex Waibel, "A Real-Time Face Tracker", Proceedings of WACV, Page 142-147, 1996.

[7]   A. Bendjebbour, Y. Delignon, et al., "Multisensor Image Segmentation Using Dempster-Shafer Fusion in Markov Fields Context", IEEE Transaction on GeoScience and Remote Sensing, Volume 39 Issue 8, August 2001.

[8]   J. R. Boston, "A Signal Detection System Based on Dempster-Shafer Theory and Comparison to Fuzzy Detection", IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, Volume 30, Issue 1, February 2000.