IMTC 2003 - Instrumentation and Measurement
Technology Conference
Vail, CO, USA, 20-22 May 2003

Sensor Fusion Using Dempster-Shafer Theory II: Static Weighting and Kalman Filter-like Dynamic Weighting

Huadong Wu[1], Mel Siegel [2(contact author)], Sevim Ablay[3]

[1,2]Robotics Institute, Carnegie Mellon University
Pittsburgh PA 15213 USA, phone: +1 412 268 8742
[3]Applications Research Lab, Motorola Labs
1301 E. Algonquin Road, Schaumburg IL 60196 USA, phone: +1 847 576 6179
e-mail: {[1]whd, [2]mws}@cmu.edu, [3]Sevim.Ablay@motorola.com
URL: [1]http://cs.cmu.edu/~whd, [2]http://www.cs.cmu.edu/~mws

Abstract ***- Context sensing for context-aware HCI challenges traditional sensor fusion methods with its requirements for (1) adaptability to a constantly changing sensor suite and (2) sensing quality commensurate with human perception. We build this paper on two IMTC2002 papers, where the Dempster-Shafer "theory of evidence" was shown to be a practical approach to implementing the sensor fusion system architecture. The implementation example involved fusing video and audio sensors to find and track a meeting participant's focus-of-attention. An extended Dempster-Shafer approach, incorporating weights representative of sensor precision, was newly suggested. In the present paper we examine the weighting mechanism in more detail; especially as the key point of this paper, we further extend the weighting idea by allowing the sensor-reliability-based weights to change over time. We will show that our novel idea – in a manner resembling Kalman filtering remnance effects that allow the weights to evolve in response to the evolution of dynamic factors – can improve sensor fusion accuracy as well as better handle the evolving environments in which the system operates.***

Keywords: **sensor fusion, context-aware computing, human-computer interaction, Dempster-Shafer theory, Kalman filtering**

I. INTRODUCTION

We build this paper on two previous papers, Sensor Fusion for Context Understanding [1] and Sensor Fusion Using Dempster-Shafer Theory [2], presented at IMTC2002. Inasmuch as this paper is a continuation and expansion of the second of these two papers, we need refer back to them in order to make the background work clear.

The goal of "context aware computing" – which, as a practical matter, is more-or-less synonymous with "context aware human computer interaction" – is for computers to understand environmental context, and thereby to more accurately interpret noisy and ambiguous inputs received from the humans with whom they interact. To achieve this end requires meeting two challenges (1) how to represent a concept as anthropocentric as "context" in a computer, and (2) how to design and deploy the sensors and sensor fusion systems that will populate the representation's slots.

Of course, one always approaches a hard problem by initially simplifying it, e.g., by beginning with concrete parameters that can be adequately defined and sensed such as, e.g., location, which is not difficult, especially if the humans involved are cooperative. Enough effort has already gone into this problem that reasonably reliable solutions – say, in the 80% range – have been demonstrated and documented. Sensor fusion combining a handful of sensors of this order of reliability can be expected, a priori, to provide system accuracy approaching arbitrarily close to 100% – how close obviously depending on the size of the hand.

But real life is yet a little harder, because (1) in real life it is unlikely that there will be available as many even "pretty good sensors" as would be needed to achieve the desired high level of system performance, and (2) in real life the number of available sensors, and the reliability of each sensor's reporting, will vary unpredictably, and in many situations practically unknowably on the requisite timescale.

To make it easier to deal with these sticky problems, in these early papers we consider only simple situations wherein we can assume that (1) the context information

---

[1] Huadong Wu is the recipient of a Motorola Partnerships in Research Grant

is represented by discrete symbols or numbers, (2) the mapping from sensor output to context representation is unambiguous, and (3) the sensors are "smart" enough to report not just data, but also meaningful estimates of its reliability, i.e., honest measures of self-confidence. Fortunately, we have available several historical data sets that record information streams from multiple sensors of differing modality, and the corresponding ground truth, thus enabling us to simulate this idealized scenario. With this simulation, we can start measuring, analyzing, comparing, and contrasting the performance of all conceivable sensor fusion architectures and implementations. Later we will further verify our findings regarding sensor fusion methods by using artificially generated data whose actual probability distributions we can know.

As described in detail in [1] and [2], our actual approach employs a layered and modularized architecture, isolating sensed context from sensor realization, and a Dempster-Shafer “theory of evidence” based sensor fusion algorithm whose formulating terminology imitates the terminology we conventionally attach to the human perception and reasoning processes. The modularization and architecture is described in [1]; and the works of applying the Dempster-Shafer algorithm to several of the historical information streams and comparing the results to an *ad hoc* weighted sum of probabilities algorithm are described in [2].

The conclusion of that paper [2] is that the Dempster-Shafer approach gives slightly better results quantitatively, but we argue that it provides a significant improvement in robustness, e.g., against data packet loss or catastrophic sensor failure, as well as a built-in, theoretically and intuitively justifiable mechanism for evaluating and reporting our confidence in the results as a function of the device and environmental conditions. The focus of the present paper is on reporting outcome realized when: (1) we incorporate, into the Dempster Shafer algorithm, weighting factors – actually this was introduced in a preliminary way in [2] – that give increased credence to sensors with better inherent reliability, e.g., higher precision, lower drift, built-in “soft failure” capacity, etc., and (2) we further incorporate, into the weighting factors, a new dynamic component – reminiscent of Kalman filtering – that organically evolves the weights. The sensor fusion mechanism is thus continuously calibrated according to the sensors’ recent performance whenever the ground truth is available.

II. CONTEXT SENSING APPROACH

A. Sensor fusion architecture

The sensor fusion system architecture, discussed in detail in [1], is reproduced in Fig. 1.

A speaker-identification sensor might decide, for example, that the current speaker is User-A with confidence interval of [0.5, 0.7], or s/he is User-B with confidence interval of [0.3, 0.5]. It might then report via its Interface Widget a database entry 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},
  {name=User-B, confidence=[0.3,0.5],
   proximity=inside, background=Background[User-B],
   …, time=update-time}, …}
```

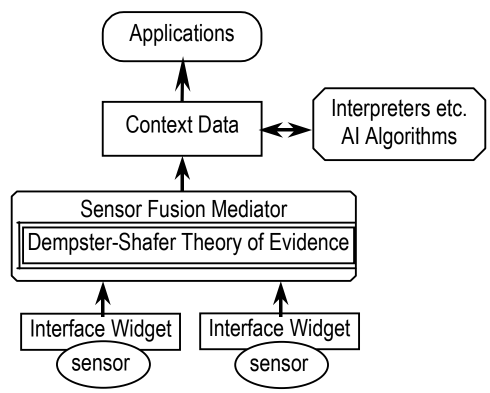

Fig. 1. System architecture for sensor fusion of context-aware computing

B. Dempster-Shafer sensor fusion algorithm

The Dempster-Shafer decision theory is essentially a generalized Bayesian statistical theory. Its new feature is that it allows distributing support for a proposition (e.g., "this is User-A") 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 the mutually exclusive context interpretations are enumerated in a "frame-of-discernment", denoted Θ. For example, if we know that there is a person in an instrumented room, and from the reality constraints this person can normally only be “User-A” or “User-B”. Now our task is to specify the user’s identity as one of the four possibilities described as:

$$\Theta = \{A,\ B,\ \{A,B\},\ \phi\}$$

meaning s/he is “User-A”, “User-B”, “either User-A or User-B” (which is actually an indication of ignorance), or “neither User-A nor User-B” (which is an indication of exceptional situation).

With the frame of discernment Θ defined, each sensor Si would contribute its observation by assigning its beliefs over Θ. This assignment function is called the “probability mass function” of Si, denoted mi.

So, according to Si’s observation, the probability that “the detected person is user-A” is indicated by a “confidence interval” whose lower bound is a "belief" and whose upper bound is a "plausibility":

$$[Belief_i(A),\ Plausibility_i(A)]$$

$Belief_i(A)$ is quantified by all pieces of evidence E_k that support proposition “User-A”:

$$Beliefi(A) = \sum_{E_k \subseteq A} m_i(E_k)$$

$Plausibility_i(A)$ is quantified by all pieces of evidence E_k that do not rule out proposition "User-A":

$$Plausibilityi(A) = 1 - \sum_{E_k \cap A = \phi} m_i(E_k)$$

For each proposition in Θ, e.g., "User-A", Dempster-Shafer theory gives a rule of combining sensor Si’s observation *m*i and sensor Sj’s observation *m*j:

$$(m_i \oplus m_j)(A) = \frac{\sum_{E_k \cap E_{k'} = A} m_i(E_k) m_j(E_{k'})}{1 - \sum_{E_k \cap E_{k'} = \phi} m_i(E_k) m_j(E_{k'})}$$

This rule can be chained straightforwardly if we view *m*j not as sensor *S*j’s observation, but instead as the previously combined observations of sensors *S*k and *S*l.

By associating "belief" with the lower end of a probability range and "plausibility" with its upper end, the Dempster-Shafer approach manages to capture the key features of the human perception-reasoning process. In contrast, the Bayesian approach, which is essentially a subset of the Dempster-Shafer approach, provides no mechanism for dealing quantitatively with the ranges of "belief" and "plausibility" that humans characteristically attach to their estimates of likelihood.

III. APPLICATION EXAMPLE: TRACKING MEETING PARTICIPANTS’ FOCUS-OF-ATTENTION FROM MULTIPLE CUES

A. Initial experiments

The experimental arrangement, discussed in detail in [2], is reproduced in Fig. 2. An omni-directional camera with face detection software provides one focus-of-attention sensor, microphones in front of each meeting participant provide another focus-of-attention sensor, and human examination of the videotape provides ground truth. Measures of confidence are derived from the relative strengths of the signals supporting in turn the hypothesis that each meeting participant is the instantaneous focus-of-attention.

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

The sensor fusion task is, given both the video and audio observation reports, optimally to combine the two inputs to generate a better focus-of-attention estimation, i.e., one with a higher and narrower confidence range. Given only one of the sensors, e.g., due to data packet loss in transmission, the system should revert gracefully to the estimation provided by the remaining working sensor.

B. Initial results

In the baseline work that generated the historical data sets, the authors used a linearly weighted sum of probabilities to estimate combined probability via an *ad hoc* formula. The video-only focus-of-attention estimation accuracy was around 75±5%, the audio-only focus-of-attention estimation accuracy was around 65±5%; and the linear combination of these two increased the overall accuracy by about 2±1 percentage points.

For the same data streams, the Dempster-Shafer combination algorithm arguably shows a small overall estimation accuracy improvement over either single sensor modality. Preliminary experiments [2] with a weighted Dempster-Shafer algorithm arguably show an additional small overall estimation accuracy improvement.

Thus, as a practical matter, all three algorithms for sensor fusion perform similarly; the small gains, even if they are real, are insignificant in any practical sense. Nevertheless, imaginative examination of the tabulated experimental results (Table 1 in [2]) suggests that the result would be substantially improved if we could improve the measure of "self-confidence" provided by the individual sensors.

IV. WEIGHTED DEMPSTER-SHAFER ALGORITHM FURTHER INVESTIGATION

A. Weighting means non-democratic voting

The fundamental Dempster-Shafer combination rule implies that we trust any sensors S_i and S_j equally. Misplaced trust can produce counterintuitive outcomes, e.g., if two observers agree that there is an arbitrarily small possibility of X, but they agree on no other possibility, Dempster-Shafer will say X is the only possible conclusion. Nor is this scenario far-fetched, as in many Dempster-Shafer applications the frame-of-discernment, and the numerical values of "belief" and "plausibility", are essentially educated guesses supplied by human experts. The human has a tendency to hedge one's bet by assigning a small probability to an unlikely alternative conclusion, which expands the overall frame-of-discernment. It thus becomes easy for two experts' sub-frames-of-discernment to share only one outcome that both experts think is unlikely. The result is the described catastrophe: the algorithm concludes the small area of agreement is the only possible conclusion.

But in sensor systems we should be able to do better by quantitatively invoking technical knowledge about each sensor's *expected* performance (based on, e.g., the sensor manufacturer's specifications), ground-truth knowledge about each sensor's current *actual* performance (based on, e.g., current working status), and *historical* knowledge about the evolution of their performance as the sensors age (based on, e.g., a regular stream of occasional ground-truth observations).

This sort of differential trust can be accounted for by a simple modification to the Dempster-Shafer formula in which the observations mi are weighted by trust factors wi derived from the corresponding expectations, calibrations, and histories of the corresponding sensor S_i's performance. The weighting process is expressed formally by inserting the weights wi as factors multiplying the probability mass functions, i.e., the observations m_i:

$$(m_i \oplus m_j)(A) = \frac{\sum_{E_k \cap E_{k'} = A} [w_i m_i(E_k) \cdot w_j\, m_j(E_{k'})]}{1 - \sum_{E_k \cap E_{k'} = \phi} [w_i m_i(E_k) \cdot w_j\, m_j(E_{k'})]}$$

B. Dynamic weighting means the voting process is continously calibrated

When the ground truth is available, e.g. shortly after current measurements or from additional information channels, it can be used by making the weight factors w_i as functions of time t. In this approach $w_i(t)$ is reminiscent of Kalman filtering.

We believe that adding sensor-property based weighting, and particularly adding dynamic sensor-property based weighting, to the Dempster-Shafer framework is the major contribution of our work.

A simple but effective practical implementation is to define $w_i(t)$ (with backward-looking time step Δt) as:

$$w_i(t) = \sum_{n=0}^{\infty} c_i(t - n \cdot \Delta t) \cdot p^n$$

where the $c_i(t)$ is the function describing the correctness of the senor S_i's estimation at time t:

$$c_i(t) = \begin{cases} 0 & \text{correct estimation} \\ 1 & \text{incorrect estimation} \end{cases}$$

and the p is the “remnance factor”, which is in the range 0.0 to 1.0, deciding how rapidly past performance will be discounted.

V. NEW EXPERIMENTAL RESULTS

A. Using the pre-recorded data

In the previous report [2], we showed the sensor fusion results using regular and our newly proposed weighted Dempster-Shafer methods, with the probability linear summation as the comparison baseline. With the same pre-recorded data, here we further calculated the sensor fusion results using the above described dynamic weighted Dempster-Shafer method. As an example, the remnance factor is chosen as 0.9, and the results are shown in Table 1.

It can be seen that, although the difference may not be very significant, the result of dynamically weighed Dempster-Shafer method is better than *any one of the* alternative method in *every experimental data set*.

B. Using new simulated data with known probability distribution

Analyzing the pre-recorded experimental data can provide us a good feeling regarding how well these sensor fusion methods work. However, to make our conclusions more clear and convincing, we would like to test the algorithm against data with known probability distribution.

Without losing generality, we imagine a typical focus-of-attention analysis scenario with a set of three sensors, and we use simulated sensor fusion data to compare sensor fusion methods.

Suppose some user's head would pan an angle according to a Gaussian distribution $N[-45°, \sigma_0]$, $N[0°, \sigma_{00}]$, or $N[45°, \sigma_0]$ when his/her focus-of-attention is on the left-side, the straight-forward, or the right-side meeting-participant respectively.

Table 1. Focus-of-attention analysis sensor fusion method comparison with the pre-recorded data

	User	valid frames	linear sum correct	DS correct	WDS correct	dynamic WDS correct
Experiment Set2	#0	1229	70.1%	70.0%	71.4%	74.9%
	#1	1075	69.8%	70.0%	69.4%	73.0/%
	#2	1098	80.2%	80.8%	80.2%	80.9%
	#3	991	65.6%	66.6%	70.0%	72.1%
Experiment Set5	#0	768	76.8%	77.0%	77.0%	80.1%
	#1	956	72.0%	72.3%	72.1%	77.0%
	#2	1006	84.1%	84.2%	83.9%	85.1%
	#3	929	75.7%	76.9%	73.2%	79.1%
Experiment Set6	#0	799	71.2%	71.5%	71.0%	74.5%
	#1	751	85.5%	85.8%	85.2%	86.2%
	#2	827	83.3%	84.3%	83.4%	83.8%
	#3	851	81.9%	82.3%	81.7%	82.8%
Experiment Aufnahme2	#0	653	85.0%	85.0%	84.2%	86.2%
	#1	653	54.2%	54.2%	54.5%	63.1%
	#2	681	69.5%	69.3%	70.3%	76.1%
	#6	435	78.2%	78.4%	79.8%	83.9%
summary		13702	75.8%	75.4%	75.4%	78.4%

Because of sensor Si's measurement noise which has Gaussian distribution $N[dft_i(t), \sigma_i]$ that is independent of the being measured angle, the Si observed pan angle would be:

$$N[-45° + dft_i(t), \sqrt{\sigma_i^2 + \sigma_0^2}\,],$$

$$N[dft_i(t), \sqrt{\sigma_i^2 + \sigma_{00}^2}\,],$$

or

$$N[45° + dft_i(t), \sqrt{\sigma_i^2 + \sigma_0^2}\,],$$

when the user's focus-of-attention is on the left-side, the straight-forward, or the right-side meeting-participant respectively.

The term $dft_i(t)$ here describes the sensor Si's drifting effects in measurement. But because the $dft_i(t)$ cannot be predicted, the sensor Si would reasonably infer the user's focus-of-attention as if there were no drift.

Using the observed pan angle (with $dft_i(t)$ being set to zero) distribution function, for user's head pan angle x, the probability density functions $pdf_L(t)$, $pdf_S(t)$, and $pdf_R(t)$ can be calculated. Next, the sensor Si's rational estimation regarding the user's focus-of-attention can be calculated with the relative probability density function values as:

$$\frac{1}{pdf_L(t) + pdf_S(t) + pdf_R(t)}\{pdf_L(t),\ pdf_S(t),\ pdf_R(t)\}$$

where the three numbers correspond to the probabilities that the user's focus-of-attention is on the left, straight, or the right person.

With this simulated sensing scheme, we can imagine a meeting scenario where the user has his/her focus-of-attention on the left-side, the one straight-across the table, and the right-side meeting participant randomly (with a probability distribution of 0.3, 0.4, and 0.3 respectively) for a random time length in the range of 5 to 15 seconds.

For each second, the user's real head pan angle is generated with σ_0=5° and σ_{00}=10°, and there are 3 sensors that will respectively generate 10 pan angle observations with σ_1=5°, σ_2=10°, and σ_3=20°.

Since the measurement drifting effect is most difficult part to handle in real practices, our experiments study three seemly ad hoc situations but actually with the following considerations: (I). the sensor's drift cycles are relatively long compared with our experiment time: $dft_1(t)=5°\cdot sin(0.001\cdot t)$, $dft_2(t)=5°\cdot sin(0.0007\cdot t)$, and $dft_3(t)=5°\cdot sin(0.0003\cdot t)$ (the sensors' drift cycles are approximately 105, 150, and 345 minutes respectively); (II). The drift cycles are normal: $dft_1(t)=5°\cdot sin(0.01\cdot t)$, $dft_2(t)=5°\cdot sin(0.007\cdot t)$, and $dft_3(t)=5°\cdot sin(0.003\cdot t)$ (the sensors' drift cycles are approximately 10.5, 15, and 35 minutes respectively); and (III) the sensors' drift amplitudes are relative large compared with their built-in measurement noise: $dft_1(t)=10°\cdot sin(0.01\cdot t)$, $dft_2(t)=5°\cdot sin(0.007\cdot t)$, and $dft_3(t)=5°\cdot sin(0.003\cdot t)$.

With such three assumed sensor drift scenarios, we did 2 sets of experiments for each situation, simulated about an hour length meeting, and did sensor fusion method analysis. Again, as an example the remnance factor for the dynamically weighted Dempster-Shafer method is set as 0.9. The results are shown in Table 2.

In Table 2 fractions of the events that the user's focus of attention is correctly estimated are in percentage format, the columns specify sensors' drift scenario and experiment data sets, whereas the rows specify individual sensors' (Sensor S_1, S_2, and S_3) performance and the effectiveness of sensor fusion methods (Linear – probability linear combination, or averaging; DS – standard Dempster-Shafer method; wDS – weighted Dempster-Shafer method; and DSK – Dempster-Shafer method with Kalman filter-like dynamic weighting schemes).

Table 2. Sensor fusion method comparison using simulated sensory data

sensor, sensor fusion	Drift I		Drift II		Drift III	
	#1	#2	#1	#2	#1	#2
S_1	85.7%	87.0%	86.7%	85.0%	83.7%	82.5%
S_2	81.4%	82.3%	82.4%	81.1%	80.1%	77.0%
S_3	71.9%	72.7%	72.1%	70.7%	70.5%	69.2%
Linear	84.8%	86.3%	85.9%	84.3%	84.3%	81.6%
DS	84.6%	86.1%	85.7%	84.1%	84.4%	80.9%
wDS	84.9%	86.4%	86.0%	84.5%	84.6%	81.6%
DSK	86.3%	87.3%	87.3%	86.0%	86.6%	84.6%

The numbers in the table confirm our conclusion made in the previous paper [2], i.e., there is not much difference among effectiveness of sensor fusion methods of linear combination, standard Dempster-Shafer method, and the weighted Dempster-Shafer method, with the weighted Dempster-Shafer method doing a marginally better job. Also confirmed is that the Dempster-Shafer method with dynamic weighting scheme consistently outperforms all other alternative methods.

VI. CONCLUSION

From pre-recorded live experimental data analyses and from our artificially generated data analysis, we can tentatively conclude that: the four sensor fusion schemes, *ad hoc* linear weighting, standard Dempster-Shafer, Dempster-Shafer with static weights, and Dempster-Shafer with dynamic weights progressively show, at best, practically insignificant performance improvements. However, when the ground truth is available afterwards, it is better in practice to use the Dempster-Shafer with dynamic weights for sensor fusion scheme as it consistently outperforms alternative methods.

Our conclusion from previous work can also be safely repeated: the Dempster-Shafer method (most desirably with dynamic weights, or with static weights) is the preferred sensor fusion scheme for context-aware computing sensor fusion, because it resembles human users' inference processes and provides a great advantage or convenience to manage uncertainties.

Finally we have a speculation that whenever the dynamic weighted Dempster-Shafer can be used, it will practically be the best method to deal with sensors that intermittently work because the sensor fusion mechanism needs not be changed over different situations.

REFERENCES

[1]. H. Wu, M. Siegel, and S. Ablay, "Sensor Fusion for Context Understanding," presented at IEEE International Measurement Technology Conference (IMTC) 2002, Anchorage AK USA, 2002.

[2]. H. Wu, M. Siegel, R. Stiefelhagen, and J. Yang, "Sensor Fusion Using Dempster-Shafer Theory," presented at IEEE International Measurement Technology Conference (IMTC) 2002, Anchorage AK USA, 2002.

[3]. 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

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

[5]. Glenn Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, 1976.

[6]. *Advances in the Dempster-Shafer Theory of Evidence*, edited by Ronald R. Yager, Janusz Kacprzyk, and Mario Fedrizzi. Wiley, 1993.