Proceedings of the 1998 IEEE
International Conference on Robotics & Automation
Leuven, Belgium • May 1998

Integrating Dependent Sensory Data

Albert C.S. Chung, Helen C. Shen,
Department of Computer Science
The Hong Kong University of Science & Technology
Clear Water Bay, Kowloon
Hong Kong.

Abstract

In sensory data fusion and integration consideration, sensor independence is a common assumption. In this paper, we demonstrated the impact of including dependent information in sensory data combination process. The team consensus approach based on information entropy can improve the measurement accuracy remarkably. The major benefits of the approach are (a) the simple linear combination of the weighted initial local estimates for each sensor; and (b) the low order bivariate likelihood functions which can be represented easily. A comparison of the team consensus approach with the Bayesian approach is presented.

1 Introduction

Sensory Data Dependence gives an additional piece of information about the interactions between the sensor observations. Consideration of dependence in the process of data combination is expected to give a higher quality of information. Hence, the information-theoretic entropy, which is a common tool to measure the randomness of a given data set, is used to measure and describe the nature of interactions among sensory data. If the interaction between two sensors can reduce the uncertainty level, then the observation of one sensor is positively relevant to another sensor. Otherwise, it is considered as negatively relevant. This situation should be carefully handled in the combination process.

It is well known that combining sensory data has two major advantages: redundancy and complementarity [1]. *Redundancy* means not only is the sensory data duplicated, the correlation among sensors are also positive in terms of estimation errors. Positive error correlation implies that when the estimation error of one of the redundant sensors increases, the estimation errors of other redundant sensors also increase and vice versa. *Complementarity* means not only does each sensory data have an unique part of the observation domain, the correlation among sensors are also negative in terms of estimation errors. Negative error correlation implies that when the estimation error of one of the complementary sensors decreases, the estimation errors of other complementary sensors increase and vice versa. Hence, estimating error correlation gives a new and alternative definition to the sensor type.

In this paper, we begin by introducing entropy as a measure of uncertainty among the data set, X_i, observed by sensor i. The initial local estimates u_i of each sensor, based on the sensor observations, likelihood functions and entropy values, are derived and then combined by the Markovian decision process cooperatively to form a team of dependent sensors. When two sensors are detected to be negatively relevant, they are then re-set to be independent to maintain the uncertainty level. The proposed approach is demonstrated by a team of two negative correlated sensors, namely a sonar sensor and a b/w CCD camera.

2 Team Consensus Approach with Dependence

Suppose that there are m individual sensors observing a random variable $\theta(\in \Theta)$. Individual and joint posterior distributions are given by

$$p_i(\theta|x_i) \propto \pi(\theta) \times l(x_i|\theta)$$

$$p_{ij}(\theta|x_i, x_j) \propto \pi(\theta) \times l(x_i, x_j|\theta) \quad (1)$$

where random variables x_i and x_j are the observations of sensors i and j about θ, $\pi(\theta)$ is a common prior distribution, and $l(x_i|\theta)$ and $l(x_i, x_j|\theta)$ are the univariate and bivariate likelihood function given θ.

2.1 Entropy Measure

Entropy, which was introduced by Shannon [2] in 1948, has long been used to measure the probabilistic uncertainty of a random variable. Its value is directly proportional to the degree of uncertainty (or randomness) of the measured variable; the smaller the uncertainty, the smaller the entropy.

Self-Entropy, which [3] measures how uncertain a sensor is about its own observation x_i, is defined as

$$h_{i|i}(x_i) = -\sum_{\theta \in \Theta} p_i(\theta|x_i) \log p_i(\theta|x_i) \qquad (2)$$

Conditional Entropy, which [3] measures how uncertain sensor i is about the joint observations x_i and x_j given that observation of sensor x_j is unknown, is defined as

$$h_{i|j}(x_i) = -\sum_{x_j \in X_j} p(x_j|x_i) \sum_{\theta \in \Theta} p_{ij}(\theta|x_i, x_j) \log p_{ij}(\theta|x_i, x_j) \qquad (3)$$

where $p(x_j|x_i)$ is the conditional distribution of x_j given x_i. It shows that given x_i, $h_{i|j}$ is simply the expected value of self-entropy of their joint observations. When the observation of sensor j is explicitly known, Equation (3) reduces to

$$h_{i|j}(x_i, x_j) = -\sum_{\theta \in \Theta} p_{ij}(\theta|x_i, x_j) \log p_{ij}(\theta|x_i, x_j)$$

The conditional entropy manifests profoundly the dependence between sensors i and j. It is used to capture the essence of observation relevance exchanged between sensor i and j. For example, if sensor j's observation is irrelevant to sensor i, the posterior distribution p_{ij} will be equal to p_i which makes sensor i's conditional entropy equal to its self entropy. In other words, sensor j's observation does not help sensor i to improve its state of uncertainty. Another example, if sensor j's observation is positively relevant to sensor i, then we expect that $h_{i|j}$ to be smaller than $h_{i|i}$, which means that this observation contributes to reducing the uncertainty of sensor i. Otherwise, the observation is negatively relevant, sensor i should at least maintain its uncertainty level. The properties of self entropy and conditional entropy are summarised as follows:

- $h_{i|j}$ is not necessarily equal to $h_{j|i}$,
- if the self-entropy $h_{i|i}$ is equal to the conditional entropy $h_{i|j}$, then observations are irrelevant (or independent); and
- if the self-entropy $h_{i|i}$ is larger than the conditional entropy $h_{i|j}$, then observations are positively relevant; and
- if the self-entropy $h_{i|i}$ is smaller than the conditional entropy $h_{i|j}$, then observations are negatively relevant.

2.2 Combining Sensory Data

Markov Chain has been used as a decision process to combine data because of two major reasons: (a) it is stated [5] that the consensus estimate is a linear combination of the weighted individual local estimates which greatly simplifies the computation process; and (b) weight (or transition probability) assigned by one sensor state to another sensor state is intuitively related only to the bivariate likelihood functions. Higher order functions are not necessary. Because of these reasons, the Markovian decision process is employed and briefly described as below. (see [4] for details)

Let u_i be an initial local estimate of sensor i based on some observations, the likelihood functions of x_i and x_j, and entropy values. Let $\vec{U}^0$ be an initial state vector of the initial individual expected estimates $(u_1, \ldots, u_m)^T$, T denoting transposition, and $\vec{U}^k$ be a state vector at k^{th} iteration. Let W be the transition matrix. Its nonnegative element w_{ij} is the weight (or transition probability) assigned by sensor state i to sensor state j and has the properties that $\sum_{j=1}^{m} w_{ij} = 1$ and $0 \leq w_{ij} \leq 1$. Let $\vec{\mathcal{K}}$ be a vector of stationary transition probabilities $(\kappa_1, \ldots, \kappa_m)^T$, where $\sum_{i=1}^{m} \kappa_i = 1$ and $0 \leq \kappa_i \leq 1$. Markov chain recursively combines and updates the individual sensor states of $\vec{U}^{k-1}$,

$$\vec{U}^k = W\vec{U}^{k-1} \quad k \geq 1$$

which is equivalent to

$$\vec{U}^k = W^k \vec{U}^0, \qquad (4)$$

and, as k trends to infinity, $\vec{U}^k$ converges to a consensus value u^*,

$$u^* = \vec{\mathcal{K}}^T \vec{U}^0 = \sum_{i=1}^{m} \kappa_i u_i. \qquad (5)$$

where

$$W^T \vec{\mathcal{K}} = \vec{\mathcal{K}}. \qquad (6)$$

It is observed that, from Equations (4,5), κ_i is large if and only if w_{ji} is large for $j = 1, \ldots, m$. Weights lying in the same column of the transition matrix, W, contribute positively to κ_i. This means sensor i will have greater inference on the consensus value u^* if and only if the weights (or transition probabilities) assigned to sensor i are large. Equation (5) reveals that the consensus estimate are the linear combination of the initial local estimates. Equation (6) can be viewed as the eigenvector problem with eigenvalue equal to one. Therefore, κ_i, for $i = 1, \ldots, m$, can easily be found by a variety of methods for solving eigenvector problems even though m is large.

2.3 Weight Assignment

This section deals with how appropriate weights are to be assigned based on self-entropy and conditional entropy. Consider a state transition from sensor state i to sensor state j with weight w_{ij}. If this transition is treated as an information flow, then sensor j will definitely gain information about sensor i. Sensor j can in turn compute its conditional entropy $h_{j|i}$ based on sensor i's observation. A greater weight should be assigned to this transition if the calculated conditional entropy $h_{j|i}$ is small. This implies that the weight (or transition probability) is inversely proportional to the conditional entropy. The same discussion applies to the self-entropy. The larger the self-entropy, the smaller the corresponding weight. If $h_{j|i}$ is smaller than $h_{i|i}$, then w_{ij} is larger than w_{ii}. This relationship is formulated as follows:

$$w_{ij} \propto \frac{1}{h^n_{j|i}} \quad for \quad i, j = 1, \ldots, m$$

where the weight assigned to sensor j by sensor i depends inversely on the conditional entropy of sensor j based on sensor i's observation, and n can be adjusted to reflect this dependence. The greater the n, the smaller the entropy and the larger the weight. It is then written in matrix form,

$$\begin{bmatrix} w_{11} & \cdots & w_{1m} \\ w_{21} & \cdots & w_{2m} \\ \vdots & \ddots & \vdots \\ w_{m1} & \cdots & w_{mm} \end{bmatrix} \propto \begin{bmatrix} \frac{1}{h^n_{1|1}} & \cdots & \frac{1}{h^n_{m|1}} \\ \frac{1}{h^n_{1|2}} & \cdots & \frac{1}{h^n_{m|2}} \\ \vdots & \ddots & \vdots \\ \frac{1}{h^n_{1|m}} & \cdots & \frac{1}{h^n_{m|m}} \end{bmatrix}.$$

Since $\sum_{j=1}^{m} w_{ij} = 1$, it follows that the weight is given by

$$w_{ij} = \frac{1}{h^n_{j|i} \sum_{k=1}^{m} \frac{1}{h^n_{k|i}}} \quad for \quad i, j = 1, \ldots, m. \tag{7}$$

2.4 Properties of Weights

It is observed that weight w_{ij} is a function of self-entropy and conditional-entropy, in which they are functions of univariate and bivariate likelihood functions respectively. It is worthwhile to note that only the univariate and bivariate likelihood distributions are needed throughout the decision process. However, in the Bayesian model, which will be discussed later, higher order distributions are necessary.

Much attention should be paid to the cases when $h_{i|j}$ diminishes to zero. For non-zero $h_{i|j}$, Equation (7) works perfectly. It should be noted that when sensor i is absolutely certain about its observation ($h_{i|i} = 0$) or joint observation ($h_{i|j} = 0$), the entire corresponding column in the transition matrix w_{ij} will then be set to one for $i = 1, \ldots, m$. When more than one sensor happens to be absolutely certain, normalization across the transition matrix row is necessary.

2.5 Local Estimate

Equation (5) shows that global consensus estimate is a weighted sum of local estimates. In turn, the local estimate is an estimation of sensor i about θ based on (a) the information about the joint observation x_i and x_j which is represented by a posterior distribution, $p_{ij}(\theta|x_i, x_j)$; and (b) the entropy of the joint observation x_i and x_j. It is given by

$$u_i = \sum_{j=1}^{m} w_{ij} \sum_{\theta \in \Theta} \theta p_{ij}(\theta|x_i, x_j),$$

for $i = 1, \ldots, m$, where w_{ij} is defined by Equation (7). It is noted that $p_{ij}(\theta|x_i, x_j)$ is set to $p_{ij}(\theta|x_i)$ when $h_{i|i} \geq h_{i|j}$ because the uncertainty level should at least be maintained in the case of negative relevance.

3 Experimental Results

This section demonstrates the proposed team consensus approach by considering a team of one b/w CCD camera and a sonar sensor. Their observations are represented by two random variables: x_1 and x_2 respectively. The quantity observed is the distance between sensors and object. The distance is represented by a random variable θ.

The aim of integration is to complement the weaknesses of sonar sensors and CCD cameras when they are estimating the object distance alone. In the experiment, the sonar sensor and the CCD camera are mounted on the gripper of a robot arm. Both sensors contribute to the decision process and finally reach a consensus, which is the estimated distance between the gripper and the object. The team consensus estimate u^* can then be fed into the robot controller for the next step of action.

In our experiment, a commercial sonar sensor is used. It has a limitation on the range of measurement from 0.49m to 12m, within 1 % of error. For objects closer than 0.49m, it gives readings with large error. Thus, we use a b/w CCD camera which is usually for object recognition rather than distance measure, to compensate the inadequacy of the sonar sensor. By considering the size of a small black circle placed on the object, the CCD camera can estimate the distance of the object observed. This is achieved by measuring the length of the diameter of the circle observed, i.e. the number of pixels along the diameter in the image. The closer the object, the larger the number of pixels in the image. The change in the size of the circle is significantly large as the gripper moves closer to the object. Whereas, the change in the size of the circle is small as the camera moves farther away from the object resulting in larger estimation error of the distance between object and gripper. In turn, this can be corrected by the sonar sensor's observation.

Let $E[\theta|x_i]$ and MSE_i be the individual expected distance and mean square error for CCD camera detection ($i = 1$) and sonar sensor detection ($i = 2$). They are given by

$$E[\theta|x_i] = \sum_{\theta\in\Theta} \theta p_i(\theta|x_i), \; and$$

$$MSE_i = \sum_{x_i} \sum_{\theta\in\Theta} (\theta - E[\theta|x_i])^2 p_i(\theta|x_i)$$

where θ is the true distance, x_1 is the number of pixels observed by the camera and x_2 is the observed distance by the sonar sensor.

If the errors of CCD camera and sonar sensor are measured by

$$e_1 = E[\theta|x_1] - \theta \; and \; e_2 = E[\theta|x_2] - \theta.$$

The correlation between e_1 and e_2 is derived by

$$COV(e_1, e_2) = \frac{\sum_{e_1}\sum_{e_2}(e_1 - \bar{e_1})(e_2 - \bar{e_2})p(e_1, e_2)}{\sigma_{e_1}\sigma_{e_2}}$$

where $\bar{e_1}$ and $\bar{e_2}$ are the error means; and $\sigma^2_{e_1}$ and $\sigma^2_{e_2}$ are the error variances. It is shown from the experiment that $COV(e_1, e_2)$ is negative. This reveals that the estimation errors are negatively correlated. Therefore, this team of two sensors compensate each other in the sense that for measuring distances less than 0.49m, the CCD camera can be expected to give better estimates and vice versa. We shall then estimate the consensus value u^* by the team consensus approach and finally compare with the Bayesian approach.

3.1 CCD Camera Detection

The actual distances between 28cm to 100cm were observed by the camera. Figure 1 gives the mean square error of the distances observed. As expected, this sensor gives better estimates for distances below 49cm.

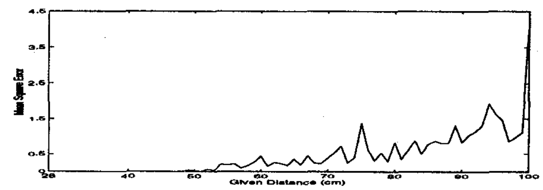

Figure 1: CCD Camera Detection

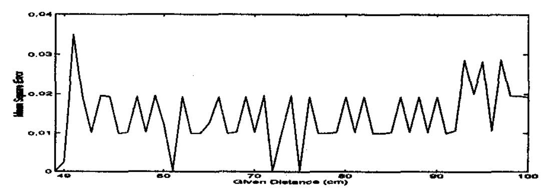

Figure 2: Sonar Sensor Detection

3.2 Sonar Sensor Detection

Distances between 28cm to 100cm were "observed" by the sonar sensor. Similar to the CCD camera. Figure 2 gives the mean square error of the distance observed. It should be noted that for distances ranging between 49cm to 100cm, very good estimates are observed, whereas error, which is very large and is not shown, increases dramatically when the distance falls below 49cm, the limitation of the sonar sensor given by the manufacturer of the sensor.

3.3 Team Consensus Approach assuming independence

Team consensus approach is first implemented assuming the independent relationship. (see [4] for details). The result is then compared with that of dependent relationship.

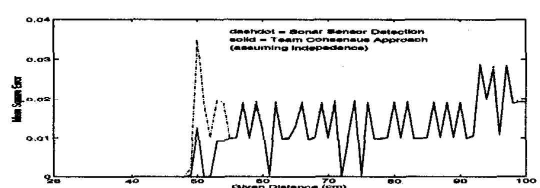

Figure 3: Team Consensus Approach assuming independence

In Figure 3, the application of team consensus approach shows that, the mean square error is smaller than that provided by any single sensor. This demonstrates that the team consensus approach can improve the measurement accuracy, as compared with the performance of each individual sensor.

However, the experiment gives an insight into the disadvantage of the approach with the assumption of independent relationship [4]. Since the consensus estimate is the linear combination of the individual estimates based on its own observation in which weights are constant and predetermined, consensus estimate is obviously bounded by $min\{u_i\}$ and $max\{u_i\}$.

3.4 Team Consensus Approach assuming dependence

With the consideration of observation dependence, Figure 4 and Table 1 show that the mean square error is further reduced when compared with the performances without the dependence relationship and each individual sensor. Performance is greatly improved in the mid-range ($49 \sim 73cm$) because sonar sensor's observation is corrected by the dependence relationship with the CCD camera.

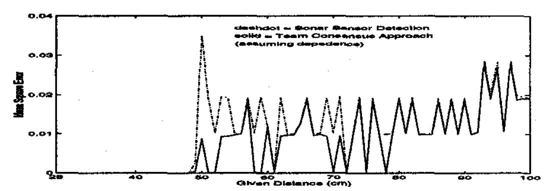

Figure 4: Team Consensus Approach with Dependence

Detection ($28cm \sim 100cm$)	*MSE*
CCD Camera, MSE_1	0.466399
Sonar Sensor ($28cm \sim 48cm$), MSE_2	36.666667
Sonar Sensor ($49cm \sim 100cm$), MSE_2	0.014582
TCA (Independence)	0.014266
TCA (With Dependence)	0.008052

Table 1: Mean Square Error (MSE) of individual detection and Team Consensus Approach (TCA)

3.5 Bayesian Approach

Bayesian approach has been used extensively in the area of data fusion. It relies heavily on the conditional posterior distributions among the random variables involved and Bayesian combination rule which is given by, for 2 sensors,

$$E[\theta|x_1, x_2] = \sum_{\theta \in \Theta} \theta p_{ij}(\theta|x_1, x_2) \tag{8}$$

where, by Equation (1), $p_{ij}(\theta|x_1, x_2)$ in turn depends on the prior function, $\pi(\theta)$, and bivariate likelihood function, $l(x_i, x_j|\theta)$.

3.6 Bayesian Network Alone

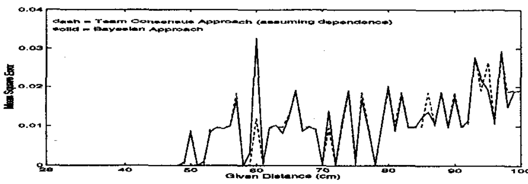

Figure 5: Bayesian Network Detection

Figure 5 reveals that the mean square error of the Bayesian approach shares the same error bound ($0 \sim 0.04cm$) with the team consensus approach.

3.7 Bayesian Network as the $(m+1)^{th}$ sensor

The discussion of the team consensus approach so far does not impose any form of restrictions on the nature of sensors. In the abstract level, sensor is just viewed as an estimator in which it can assign a probability distribution or assessment to reflect its state of information. It is possible to include Bayesian Network as a "logical sensor" in a pool of m sensors. The Bayesian expected distance and self-entropy are generalized from Equations (8,2) and given by

$$E[x_{m+1}] = E[\theta|x_1, \ldots, x_m], \quad and$$

$$h_{m+1|m+1}(x_{m+1}) = - \sum_{\theta \in \Theta} p_{m+1}(\theta|x_1, \ldots, x_m) \log p_{m+1}(\theta|x_1, \ldots, x_m).$$

Since Bayesian Network is treated as an independent "sensor", the conditional entropies are then equal to its corresponding self-entropy, i.e. $h_{m+1|i} = h_{m+1|m+1}$ and $h_{i|m+1} = h_{i|i}$.

Detection ($28cm \sim 100cm$)	*MSE*
Bayesian Network Alone	0.008409
TCA with Bayesian Network	0.008011

Table 2: Mean Square Error (MSE) of Bayesian Network Alone and with Team Consensus Approach (TCA)

Figure 6 and Table 2 show that the team consensus approach and Bayesian Network can collaboratively perform detection to achieve a better estimation. A crucial issue, as stressed above, in including Bayesian Network as the $(m+1)^{th}$ sensor is the complexity of manipulating a higher order multivariate likelihood function, $l(\theta|x_1, \ldots, x_m)$.

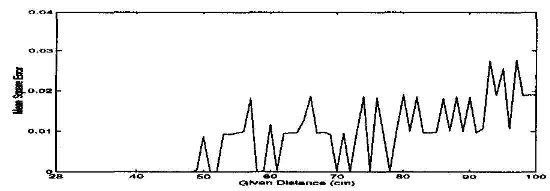

Figure 6: Team Consensus Approach with Bayesian Network as the $(m+1)^{th}$ sensor

3.8 Summary of the results

Team consensus approach is established to incorporate the uncertainties of a single observation and joint observations into the Markovian decision process such that the interdependence between sensors can be reflected effectively during the data combination process. The performance of the approach, as shown in Figure 4, when compared with the individual sensors in terms of mean square error, demonstrates its strength in improving measurement accuracy for a group of negatively correlated sensors. It provides strong evidence to the generalization of the technique to a pool of m sensors. It also shows that the aggregation of negatively correlated sensors is constructive.

An important advantage of the technique is its simplicity in terms of data structure and computation. A maximum of up to second order of likelihood function is necessary for m sensors by which it can greatly simplify the data structure to represent the interrelationships among sensors, and accelerate the computation of sensory weights.

The Bayesian approach is a general and optimal tool for all decision problems. The experiment shows that the team consensus approach illustrated in Figure 5 gives satisfactory mean square error when it is compared with the Bayesian approach. Moreover, we have included experimentally the Bayesian Network in the sensor team and viewed it as the third "virtual" sensor. The results, Figure 6 and Table 1, 2, show that the inclusion of the Bayesian sensor can improve the overall estimation accuracy. The main motivation to aggregate physical and logical sensors is that the physical constraints and mathematical limitations of the "sensors" can be relieved and compensated by appropriately choosing sensor and its model.

4 Conclusion and Future Research

This paper shows the significant impact of including dependent information in sensory data combination process. As proved by the experimental results, the addition of dependent relationships is useful in the sense that the team consensus approach with dependence can remarkably improve the measurement accuracy, when compared with individual sensors. The major benefits of the approach are, as stated above, (a) the simple linear combination of the weighted initial local estimates for each sensor; and (b) the low order bivariate likelihood functions which can be easily represented. The disadvantage is that, owing to the limited order of likelihood functions, dependence information may not be 'fully' represented as in the Bayesian Approach, in which data set interactions can be completely modelled by the higher order likelihood functions. However, in terms of computation efficiency and data representation simplicity, team consensus approach is still attractive to implement.

Future research will be (a) the application of team consensus approach with dependence in a larger scale $(m > 2)$ sensor system; and (b) the adaptive weight assignment because the weights should spontaneously reflect the environmental changes

Acknowledgement

The research is supported by the RGC Competitive Earmarked Research Grant, no. HKUST 618/94E, Hong Kong.

References

[1] Ren C. Luo and Michael G. Kay (1989) "Multisensor Integration and Fusion in Intelligent Systems", *IEEE Transactions on Systems, Man, and Cybernetics*, 19(5), pp. 901-931.

[2] C.E. Shannon (1948), "A Mathematical Theory of Communication", *Bell System Tech. J.*, 27, 379-423, 623-656.

[3] O.A. Basir and H.C. Shen (1996). "Modeling and Fusing Uncertain Multi-Sensory Data", *Journal of Robotics Systems*, 13(2), pp. 95-109.

[4] Albert C.S. Chung, Helen C. Shen, Otman A. Basir (1997) "A Decentralized Approach to Sensory Data Integration", *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept. 97, Grenoble, France, pp1409-1414.

[5] Morris H. DeGroot (1974) "Reaching a Consensus", *Journal of the American Statistical Association*, 69(345), March 74, pp118-121.