Proceedings of the 1998 IEEE/RSJ
Intl. Conference on Intelligent Robots and Systems
Victoria, B.C., Canada • October 1998

Dependence in Sensory Data Combination

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

It is common to assume sensor independence in the sensory data fusion and integration. In [1] and [2], we have illustrated that the team consensus approach based on information entropy can remarkably improve the measurement accuracy. The major benefits of the approach are (a) the simple linear combination of the weighted initial expected estimates for each sensor; and (b) the low order bivariate likelihood functions which can be represented easily. In this paper, we demonstrate specifically both the positive and negative impacts of including dependent information in sensory data combination process; and show how the measurable consensus uncertainty level can be derived. A comparison of the team consensus approach with the Bayesian approach is presented.

1 Introduction

This paper reveals both the positive and negative impacts of sensory data dependence. It has been shown [1] that data dependence can offer additional information about the interactions between the sensor observations. Consideration of dependence in the process of data combination is generally 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 describe the nature and measure the degree of interactions among sensory data. Given two sensor observations, if the interaction between two sensors can reduce the overall uncertainty level, then one sensor is defined to be *positively dependent* on another sensor. Otherwise, it is regarded as *negatively dependent.* This situation should be measured and carefully handled in the combination process.

It is well known that combining sensory data has two major advantages: redundancy and complementarity [3]. *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 the 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 the 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 the other complementary sensors increase and vice versa. Hence, estimating error correlation gives a new and alternative definition to the sensor type.

This paper is an extension of [1] in which 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. The major difference, as compared with [1] and [2], is that the overall uncertainty level of the team consensus is measured before the data combination process is invoked. Sensors i and j are negatively dependent when the overall uncertainty level decreases after the dependent relationship is added. Therefore, 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 (TCA)

Suppose that there are m individual sensors observing a random variables $\theta(\in \Theta)$. Their 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 variable 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 [4] 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 [5] 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 [5] 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 and is not necessarily equal to $h_{j|i}$.

The conditional entropy manifests profoundly the dependence between sensors i and j. It is used [1] to capture the essence of observation relevance exchanged between sensor i and j. The properties of self entropy and conditional entropy are summarised as follows: Given observations of sensor i and j, (a) $h_{i|j}$ is not necessarily equal to $h_{j|i}$, (b) if the self-entropy $h_{i|i}$ is equal to the conditional entropy $h_{i|j}$, then observations are *irrelevant* (or independent). This implies sensor j's observation does not help sensor i to improve its state of uncertainty; and (c) if the self-entropy $h_{i|i}$ is larger than the conditional entropy $h_{i|j}$, then observations are *positively relevant*. This implies sensor j's observation contributes to reducing the uncertainty of sensor i; and (d) if the self-entropy $h_{i|i}$ is smaller than the conditional entropy $h_{i|j}$, then observations are *negatively relevant*. Sensor i should at least maintain its uncertainty level and ignore sensor j's observation.

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 [6] that the consensus estimate is a linear combination of the weighted local estimates which greatly simplifies the computation process; and (b) the 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 [2] 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 local estimates $(u_1, \ldots, u_m)^T$, T denoting transposition, and $\vec{U}^k$ be a state vector at the 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 $0 \le w_{ij} \le 1$. 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 $0 \le \kappa_i \le 1$. Markov chain recursively combines and updates the individual sensor states of $\vec{U}^{k-1}$,

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

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 estimate u^*,

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

where $W^T\vec{\mathcal{K}} = \vec{\mathcal{K}}$.

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 estimates are the linear combination of the initial local estimates. Moreover, $W^T\vec{\mathcal{K}} = \vec{\mathcal{K}}$ 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 adopted from [1] 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. Since $\sum_{j=1}^{m} w_{ij} = 1$, it follows that the weight is given by

$$w_{ij} = \frac{1}{h^n_{j|i} \sum_{k=1}^{m} \frac{1}{h^n_{k|i}}} \quad for \quad i,j = 1,\ldots,m. \tag{6}$$

2.4 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. The local estimate is given by

$$u_i = \sum_{j=1}^{m} w_{ij} \sum_{\theta \in \Theta} \theta p_{ij}(\theta|x_i, x_j), \tag{7}$$

for $i = 1,\ldots,m$, where w_{ij} is defined by Equation (6). 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.

By rewriting Equation (7), the local estimate can be viewed as an expected value of θ from a *local distribution* $u_i(\theta)$.

$$u_i = \sum_{\theta \in \Theta} \theta u_i(\theta), \tag{8}$$

where the local distribution is given by

$$u_i(\theta) = \sum_{j=1}^{m} w_{ij} p_{ij}(\theta|x_i, x_j). \tag{9}$$

Similarly, from Equation (5), *consensus distribution* can be found by substituting the local estimates by local distributions and is given by

$$u^*(\theta) = \sum_{i=1}^{m} \kappa_i u_i(\theta). \tag{10}$$

2.5 Consensus Variance and Entropy

Consensus variance is a concept of variance for the consensus estimate. Consensus variance is defined literally in terms of consensus estimate u^* and consensus distribution $u^*(\theta)$ and is given by

$$(\sigma^*)^2 = \sum_{\theta \in \Theta} (u^* - \theta)^2 u^*(\theta). \tag{11}$$

Extending this concept, we then derive the consensus variance in terms of local estimates u_i, local variance $\sigma_i^2 = \sum_\theta (u_i - \theta)^2 u_i(\theta)$, and consensus estimate u^*.

$$(\sigma^*)^2 = \sum_{i=1}^{m} \kappa_i (\sigma_i^2 + u_i^2) - (u^*)^2. \tag{12}$$

Consensus Entropy is simply a concept of entropy for the consensus estimate and is crucial to the measurement and analysis of sensory data dependence From Equation (2), *consensus entropy* is defined as

$$Discrete: \quad h_c = -\sum_{\theta \in \Theta} u^*(\theta) \log u^*(\theta). \tag{13}$$

Suppose that the consensus distribution $u^*(\theta)$ follows a normal distribution with a p.d.f. $f(\theta|u^*, (\sigma^*)^2)$.

$$Continuous: \quad h_c = -\int_{-\infty}^{\infty} f \ln f d\theta = \ln(\sqrt{2\pi e}\sigma^*) \tag{14}$$

where

$$f(\theta|u^*, (\sigma^*)^2) = \frac{1}{\sqrt{2\pi}\sigma^*} exp(\frac{-(u^* - \theta)^2}{2(\sigma^*)^2}).$$

Given the consensus distribution is normal, Equation (14) expresses the relation of two important concepts: consensus variance and consensus entropy. Of course, a sensor is absolutely certain about its observation when consensus variance diminishes to zero.

Consensus entropy h_c has two roles to play: (a) to measure the uncertainty level of consensus estimate u^* and more importantly (b) to reflect effectively the impact of dependence between sensor i and j upon a sensor pool with $m(\geq 2)$ sensors. It differs from the self-entropy $h_{i|i}$ and conditional entropy $h_{i|j}$. The self-entropy and conditional entropies confine the uncertainty measurement to sensors i and j only. However, the consensus entropy h_c measures the uncertainty level of the whole sensor system with $m(\geq 2)$ sensors.

Based on the consensus entropy, we define two dependence concepts. Given observations x_i and x_j, if the consensus entropy h_c decreases after a dependent relationship between sensors i and j is added, sensors i and j are defined as *positively dependent* which is interpreted as beneficial to the system performance; if the consensus entropy h_c increases after a dependent relationship between sensors i and j is added, sensors i and j are defined as *negatively dependent*, which is viewed as an adverse effect to the system performance. Therefore, the negatively dependent relationship between sensor i and j is then re-set to be independent because the uncertainty level should be maintained.

3 Experimental Results

This section demonstrates the presented 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.

In [2] and [1], we demonstrated the formulation of consensus estimate of independent and dependent sensors. 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 aim of integration is to complement the weaknesses of sonar sensors and CCD cameras when they are estimating the object distance alone.

It is shown from the experiment that the correlation $COV(e_1, e_2)$ between the errors of CCD camera e_1 and sensor e_2 is negative. This reveals that the estimation errors are *negatively correlated*. 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 $\sigma_{e_1}^2$ and $\sigma_{e_2}^2$ are the error variances. 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.

3.1 CCD Camera and Sonar Sensor Detection

Detection (28*cm* $\sim$ 100*cm*)	*MSE*
Sonar Sensor (28*cm* $\sim$ 48*cm*), MSE_2	36.666667
Sonar Sensor (49*cm* $\sim$ 100*cm*), MSE_2	0.014582
CCD Camera, MSE_1	0.466399

Table 1: Mean Square Error (MSE) of individual detection

The actual distances between 28cm to 100cm were observed by the camera as well as the sonar sensor. For sonar sensor, table 1 shows that for distances ranging between 49cm to 100cm, very good estimates are observed, whereas error, which is very large, increases dramatically when the distance falls below 49cm, the limitation of the sonar sensor given by the manufacturer of the sensor. Thus, we use a b/w CCD camera to compensate the inadequacy of the sonar sensor by considering the size of a small black circle placed on the object and by measuring the length of the diameter of the circle observed, i.e. the number of pixels along the diameter in the image. The change in the size of the circle is significantly large as the gripper moves closer to the object giving better estimates. 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.

3.2 Team Consensus Approach assuming independence

Team consensus approach is first implemented assuming the independent relationship [2]. The result is then compared with that of dependent relationship.

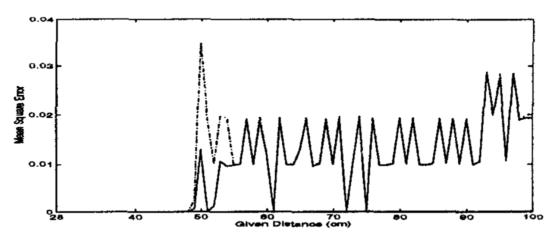

Figure 1: Team Consensus Approach assuming independence (Solid) vs Sonar Sensor Detection (Dashdot)

In Figure 1, the application of team consensus approach shows that, the mean square error is smaller than that provided by any single sensor especially at the 48cm to 55cm range. 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 [2]. 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 $max\{u_i\}$.

3.3 Team Consensus Approach assuming dependence

With the consideration of observation dependence, Figure 2 shows 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.

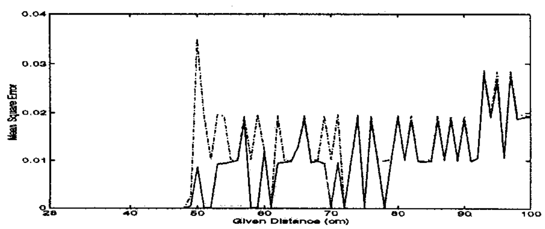

Figure 2: Team Consensus Approach with Dependence (Solid) vs Sonar Sensor Detection (Dashdot)

It is important to note that dependence can be either positive (decrease in consensus uncertainty) or negative (increase in consensus uncertainty). It is further illustrated by two real cases in Table 2. Given the observations of sonar and CCD camera, let $h_c(I)$ denote the consensus entropy assuming independence, $h_c(D)$ denote the consensus entropy assuming dependence and Δh_c denote the change in consensus entropy which is defined by $h_c(I) - h_c(D)$. Let ΔSE which is equal to $SE(I) - SE(D)$ be the corresponding change in square error.

Case I : Observations of sonar and CCD camera are *positively dependent* because the consensus uncertainty is reduced ($\Delta h_c > 0$). Moreover, the square error decreases ($\Delta SE > 0$) after the dependence has been introduced.

Case II : Observations of sonar and CCD camera are *negatively dependent* because the consensus uncertainty is enlarged ($\Delta h_c < 0$). Moreover, the square error increases ($\Delta SE < 0$) after the dependence has been introduced.

Case	Δh_c	ΔSE
I	0.310409	0.000431
II	-0.067751	-0.001133

Table 2: Change in Consensus Entropy Δh_c and Change in Square Error ΔSE. Positive Dependence (Case I). Negative Dependence (Case II)

From these two cases, it is observed that dependence (or h_c) has a direct relationship with the square error (or SE). This observation can be used to improve the system square error by considering the change in consensus entropy Δh_c and by setting the sensor relationship to be (a) dependent when it is positive and (b) independent when it is negative. Table 3 shows that the system performance (MSE) has slightly been improved. This supports the expectation of greater improvement in a larger scale ($m > 2$) sensor system.

Detection ($28cm \sim 100cm$)	*MSE*
TCA (Independence)	0.014266
TCA (with Dependence)	0.008052
TCA (with h_c consideration)	0.008023

Table 3: Mean Square Error (MSE) of Team Consensus Approach (TCA)

3.4 Bayesian Approach

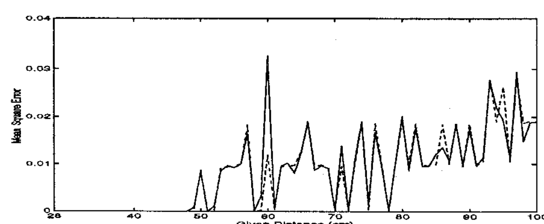

Figure 3: Bayesian Network Detection (Solid) vs Team Consensus Approach (Dashdot)

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,

$$E[\theta|x_1, x_2] = \sum_{\theta \in \Theta} \theta p_{ij}(\theta|x_1, x_2) \tag{15}$$

where, by Equation (1), $p_{ij}(\theta|x_1, x_2)$ in turn depends on the prior function, $\pi(\theta)$, and bivariate likelihood function, $l(x_i, x_j|\theta)$. Figure 3 reveals that the mean square error of the Bayesian approach shares the same error bound ($0 \sim 0.04cm$) with the team consensus approach. The MSE is 0.008409.

3.5 Summary of the results

Team consensus approach is established (a) to incorporate the uncertainties of a single observation and joint observations into the Markovian decision process and (b) to measure the consensus uncertainty such that the interdependence between sensors can be reflected effectively during the data combination process. The performance of the approach, as shown in Figure 2, when compared with the individual sensors in terms of mean square error, demonstrates its strength in improving measurement accuracy for a group of dependent sensors. It provides strong evidence to the generalization of the technique to a pool of m sensors. It also shows that the aggregation of dependent and negatively correlated sensors is constructive.

A distinct 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 3 gives satisfactory mean square error when it is compared with the Bayesian approach. It is shown that (See [1] for details) the Bayesian Network can be further viewed as a 'virtual sensor' to relieve the physical constraints and mathematical limitations of the 'sensors'. We have shown that the inclusion of the Bayesian sensor can improve the overall estimation accuracy.

4 Conclusion and Future Research

This paper presents the significant impact of including dependent information in sensory data combination process. The addition of positive dependent relationships, which leads to the decrease in the consensus uncertainty, 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 the measurable consensus uncertainty level and, as shown in [1], (a) the simple linear combination of the weighted initial individual expected estimates; and (b) the low order bivariate likelihood functions which can be easily represented. In terms of computation efficiency and data representation simplicity, the team consensus approach is attractive to implement.

Future research will be the application of team consensus approach with dependence in a larger scale ($m > 2$) sensor system.

Acknowledgement

The research is supported by the RGC Competitive Earmarked Research Grant, no. HKUST 754/96E, Hong Kong.

References

[1] Albert C. S. Chung, Helen C. Shen, "Integrating Dependent Sensory Data", *IEEE International Conference on Robotics and Automation*, May. 98, Leuven, Belgium, pp. 3546-3552, 1998.

[2] Albert C. S. Chung, Helen C. Shen, Otman A. Basir, "A Decentralized Approach to Sensory Data Integration", *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept. 97, Grenoble, France, pp. 1409-1414, 1997.

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

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

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

[6] Morris H. DeGroot, "Reaching a Consensus", *Journal of the American Statistical Association*, 69(345), March 74, pp. 118-121, 1974.