Sensor Selection for Active Information Fusion

Yongmian Zhang and Qiang Ji
Department of Electrical, Computer and Systems Engineering
Rensselaer Polytechnic Institute
Troy, NY 12180
jiq@rpi.edu

Abstract

Active information fusion is to selectively choose the sensors so that the information gain can compensate the cost spent in information gathering. However, determining the most informative and cost-effective sensors requires an evaluation of all possible sensor combinations, which is computationally intractable, particularly, when information-theoretic criterion is used. This paper presents a methodology to actively select a sensor subset with the best tradeoff between information gain and sensor cost by exploiting the synergy among sensors. Our approach includes two aspects: a method for efficient mutual information computation and a graph-theoretic approach to reduce search space. The approach can reduce the time complexity significantly in searching for a near optimal sensor subset.

Introduction

There has been a great deal of interest in the development of systems capable of using many different sources of sensory information (Waltz & Llinas 1990). In many applications, e.g., battlefield situation assessment, a number of information sources can be generated, but they are often constrained by limited time and resources. The more sensors we use, the more information we can obtain. On the other hand, every act of information gathering incurs the cost of utilizing those sensors, e.g., computational cost, operation cost, etc. In order to efficiently provide information to a decision-maker, it is important to avoid unnecessary or unproductive sensor actions. Thus, we must actively select a subset of sensors that are the most informative yet cost-effective. An important issue is how to determine a subset of sensors that is worth to be instantiated at particular stage of information gathering. There are numerous applications of sensor selection including computer vision (Paletta & Pinz 2000; Denzler & Brown 2002), control systems (Miller & Runggaldier 1997; Logothetis & Isaksson 1999) and sensor networks (Zhao, Shin, & Reich 2002; Ertin, Fisher, & Potter 2003), etc.

The strategies of sensor selection can be broadly classified into two categories: search-based approach and decision-theoretic approach. The search-based approach regards sensor selection as a search problem to find the best solution among all possible sensor combinations. Kalandros et al. (Kalandros, Pao, & Ho 1999) proposed a super-heuristics method, which begins with a base sensor combination and then generates alternative solution via random perturbations to the initial combinations. Fassinut-Mombot et al. (Fassinut-Mombot & Choquel 2004) proposed an entropy adaptive aggregation approach. After heuristically obtaining the initial subset, they iteratively aggregate and disaggregate the current subset until it converges. However, the search-based approach is computationally expensive due to the combinatorial search space. The decision-theoretic approach regards sensor selection as a decision-making problem. Kristensen (Kristensen 1997) treats the problem of choosing proper sensing actions as decision-making, and a decision tree is used to find the best sensor action at each step. Castanon (Castanon 1997) formulates the problem of dynamical scheduling of sensor for multiple object classification as a partially observed Markov decision process. However, it suffers from combinatorial explosion for a problem even in moderate size. Krishnamurthy (Krishnamurthy 2002) used dynamic programming to find an optimal sensor in a Hidden Markov model; while the approach is feasible only for the problems with a small number of sensors.

This paper focuses on the sensor selection problem with information theoretic approach, where the selection criterion is defined as a mixture of both expected information gain and cost. However, there are two difficulties to use this criterion. First, the computation of higher order mutual information (information gain) generally requires time exponential in the number of sensors to compute information gain exactly. Second, selecting k sensors out of n sensors is also a NP-hard problem. These difficulties are the impediments for real time application. For a fusion system containing many sensors, it is practically infeasible to evaluate all sensor subsets. To avoid the computational intractability of exact computation of information gain, myopic approaches are often used (Oliver & Horvitz 2003). The myopic procedure assumes that the decision maker will act after observing only one sensor. However, we should consider the fact that at each time the decision maker may observe multiple sensors before acting. So, Hecherman et al. (Heckerman, Horvitz, & Middleton 1993) presented an approximate nonmyopic computation for value of information by exploiting the statistical properties of large samples; while the approach is limited to binary hypothesis.

In this paper, we present a practical solution to circumvent computational difficulties in sensor selection when the information-theoretic criterion is used. Towards the first difficulty, we utilize the sensor pairwise information to infer the synergy among multiple sensors through exploiting the properties of mutual information. We circumvent the computational difficulty in computing higher order mutual information by efficiently computing their least upper bound instead. Towards the second difficulty, we prune the sensor synergy graph so that many weak sensor combinations are eliminated while preserving the most promising ones, and therefore the search space can be significantly reduced.

Active Information Fusion

We assume that the underlying fusion process is a dynamic Bayesian network (DBN) to account for temporal changes of the real-world as well as the sensor reliability, as shown in Fig. 1. The root node of such a network contain the hypothesis variable whose states correspond to multiple hypotheses about the state of the environment. Sensors occupy the lowest level nodes without any children. Evidences are gathered through sensors. Conditional probabilities between information variables and sensors quantify the reliability of sensor measurements, and the sensor reliability may change over time. In general, a network will have a number of intermediate nodes that are interrelated by cause and effect.

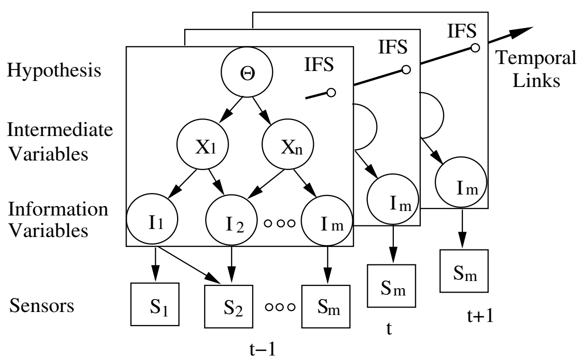


Figure 1: An information fusion system with dynamic Bayesian networks, where Θ, X, I and S denote hypothesis variable, intermediate variables, information variables and sensors, respectively.

Active information fusion is to selectively choose the sensors so that the information gain can compensate the cost spent in information gathering. A utility function used as a selection criterion consists of two components: expected information gain and sensor activation cost, and these two components are mutually utility independent (Keeney & Raiffa 1993). Therefore, we simply use a general multilinear utility function as

$$U(u_1, u_2) = (k_1 u_1 + 1)(k_2 u_2 + 1), \tag{1}$$

where k_1 and k_2 are the preference parameters and $k_1 + k_2 = 1$; u_1 is the information gain (to be discussed later), and $u_2 = 1 - C(\mathbf{S})$ is the cost saving. The activation cost for sensors $\mathbf{S}$, denoted as $C(\mathbf{S})$, needs to be normalized.

From the viewpoint of information theory (Cover & Thomas 1991), the mutual information $I(\Theta; \mathbf{S})$ between hypothesis variable Θ and sensors $\mathbf{S}$ measures the expected information gain. Considering the process at a time instant t, $I(\Theta; \mathbf{S})$ for a subset of sensors $\mathbf{S} = \{S_1, \cdots, S_n\}$ may be written as (t is dropped for notational clarity)

$$\begin{aligned} I(\Theta; \mathbf{S}) &= H(\Theta) - H(\Theta|\mathbf{S}) \\ &= \sum_{\Theta, S_1, \cdots, S_n} \left\{ P(\theta, s_1, \cdots, s_n) \log \frac{P(\theta|s_1, \cdots, s_n)}{P(\theta)} \right\}, \end{aligned} \tag{2}$$

where the joint probability $P(\theta, s_1, \cdots, s_n)$ and conditional probability $P(\theta|s_1, \cdots, s_n)$ at time t can be directly obtained through DBN inference by considering the state of temporal variables at time $t-1$ and current observations at time t; $H(\Theta)$ and $H(\Theta|\mathbf{S})$ are the expected entropy before and after instantiating sensors $\mathbf{S}$.

Sensor Selection

Eq. (2) provides a selection criterion in identifying the uncertainty reduction capability given a sensor set $\mathbf{S}$. Just considering one time slice and all binary valued sensors (the best case), the time complexity to find the best sensor subset with Eq. (2) is

$$T(n) = \sum_{i=1}^{n} \binom{i}{n} 2^i i C_I, \tag{3}$$

where n is the total number of sensors and C_I is the BN inference time by instantiating one sensor. We can see from Eq. (3) that, it is impractical to simply implement this criterion even for the best case because it generally requires time exponential in the number of summations to compute mutual information exactly. In the following subsections, we develop a methodology to address this computational difficulty.

Sensor Synergy in Information Gain

Throughout this section, we assume that we have obtained $I(\Theta; S_i, S_j)$ and $I(\Theta; S_i)$ for all i and j, the mutual information of pairwise and singleton sensors, respectively. An efficient way to obtain those values can be found in (Liao, Zhang, & Ji 2004). We first define the synergy coefficient to characterize the synergy between two sensors, and then extend it to multiple sensors.

Definition 1 (Synergy Coefficient) *A measure of expected synergetic potential between two sensors S_i and S_j in reducing uncertainty of hypothesis Θ is defined as*

$$r_{ij} = \frac{I(\Theta; S_i, S_j) - \max(I(\Theta; S_i), I(\Theta; S_j))}{H(\Theta)}. \tag{4}$$

It can be easily proved that $r_{ij} \geq 0$; this follows that S_i and S_j taken together are more informative than they are when taken alone. The larger r_{ij} is, the more synergetic S_i and S_j are. Obviously, $r(\cdot, \cdot)$ is symmetrical in S_i and S_j and $r_{ij} = 0$ if $i = j$.

Definition 2 (Synergy Matrix) *Let a sensor set be* $\mathbf{S} = \{S_1, \cdots, S_n\}$, *the sensor synergy coefficient matrix is an* $n \times n$ *matrix defined as*

$$R = \left[\begin{array}{cccc} 0 & r_{12} & \cdots & r_{1n} \\ r_{21} & 0 & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & 0 \end{array} \right], \quad (5)$$

R is indeed an information measure of synergy among sensors. With a synergy matrix, naturally we can have its graphical representation.

Definition 3 (Synergy Graph) *Given a sensor synergy matrix, a graph* $G = (\mathbf{S}, \mathbf{E})$, *where* $\mathbf{S}$ *are nodes, representing the set of available sensors, and* $\mathbf{E}$ *are edges, representing the set of pairwise synergetic links weighted by synergy coefficients* r_{ij}, *is a sensor synergy graph.*

Fig. 2(a) shows an example of such a synergy graph. We use the synergy graph to infer the synergy among multiple sensors. To further discuss theoretical properties of $I(\Theta; \mathbf{S})$ for multiple sensors, we give the following definition.

Definition 4 (Synergy Chain) *Given a synergy graph* G, *if all sensors in a subset on* G *are serially linked, this subset of sensors is referred to as a sensor synergy chain.*

Definition 5 (Markov Synergy Chain) *Given a synergy chain with* n *sensors and* $n > 2$. *For all* $i = 1, \cdots, n-1$, *if* $r_{ij} > 0$ *for* $j = i+1$ *and* $r_{ij} = 0$ *for* $j \neq i+1$, *then the chain that describes the synergetic relationship among* $\{S_1, \cdots, S_n\}$ *is a Markov synergy chain.*

Fig. 2(b) shows the above definition. The Markov synergy chain represents an ideal synergy relation and does not exist in practice. But this does not prevent us from using it as a basis for estimating joint mutual information and for the graph-theoretic analysis of the synergy among sensors. With the above definitions, we give the following theorems.

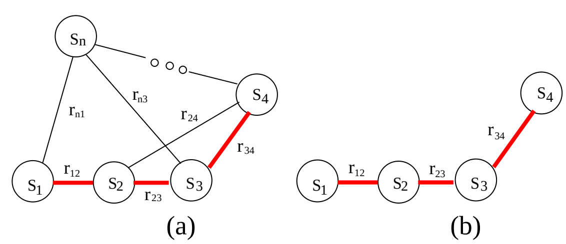


Figure 2: (a) A synergy chain $\{S_1, S_2, S_3, S_4\}$ (highlighted) on a synergy graph; (b) the corresponding Markov synergy chain

Theorem 1 (Markov Synergy Chain Rule) *Given a Markov synergy chain with a set of sensors* $\mathbf{S} = \{S_1, ..., S_n\}$. *For any* n, *the mutual information for a Markov synergy chain is*

$$I^M(\Theta; S_1, \cdots, S_n) = I(\Theta; S_1) + \sum_{i=1}^{n-1} (I(\Theta; S_i, S_{i+1}) - I(\Theta; S_i)). \quad (6)$$

Proof. See Appendix A.1 for the proof. ■

The significance of Theorem 1 is that it allows to efficiently compute joint mutual information as a sum of mutual information of pairwise sensors and singleton sensors.

Theorem 2 (Synergy Upper Bound) *For a synergy chain* $\mathbf{S} = \{S_1, \cdots, S_n\}$ *in a synergy graph* G, *its mutual information is upper-bounded by the mutual information of its corresponding Markov synergy chain, i.e.,*

$$I(\Theta; S_1, \cdots, S_n) \leq I^M(\Theta; S_1, \cdots, S_n). \quad (7)$$

Proof. See Appendix A.2 for the proof. ■

A synergy chain may correspond to multiple Markov synergy chains in a synergy graph, depending on sensor orders as can be seen in Eq. (6). Let

$$I^M_{min} = \arg\min_{\mathcal{S}} (I^M(\Theta; S_1, \cdots, S_n)), \quad (8)$$

where $\mathcal{S}$ denotes all Markov synergy chains of $\{S_1, \cdots, S_n\}$, then I^M_{min} is referred to as a least upper bound (LUB) of $I(\Theta; S_1, \cdots, S_n)$. For example, in Fig. 3, a synergy chain, $\mathbf{S} = \{S_1, S_2, S_3, S_4\}$, has multiple Markov synergy chains.

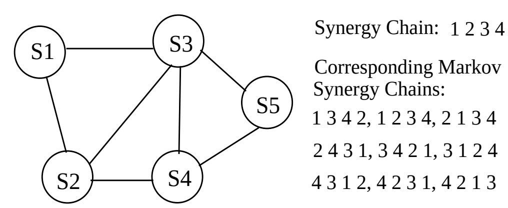


Figure 3: An illustration of synergy graph with 5 sensors.

Corollary 1 *The mutual information of a synergy chain is upper bounded the LUB of the chain, i.e.,*

$$I(\Theta; S_1, \cdots, S_n) \leq I^M_{min}. \quad (9)$$

The proof of Eq. (9) is the same as for Theorem 2. LUB is a much tighter upper bound as shown in an experimental example given in Fig. 4. For clarity, Fig. 4 only presents a part of sensor subsets from a subset space of 10 sensors. It can be seen that, the least upper bounds of $I(\Theta; \mathbf{S})$ closely follow the trend of their ground truth in the entire space of sensor subsets. Therefore, the least upper bound of $I(\Theta; \mathbf{S})$ provides measures of $I(\Theta; \mathbf{S})$ that can be used for evaluating an optimal sensor subset. Importantly, the least upper bounds of $I(\Theta; \mathbf{S})$ can be written simply in terms of the mutual information of pairwise and singleton sensors as given in Eq. (6). Therefore, the computational difficulty in computing higher order mutual information can be circumvented by replacing mutual information with their least upper bound.

Pruning Synergy Graph

The synergy graph is a completely-connected network since $r_{ij} \geq 0$. Some sensors are highly synergetic and some not. Intuitively, sensors that cause a very small reduction in uncertainty of hypotheses are those that give us the least additional information beyond what we would obtain from the

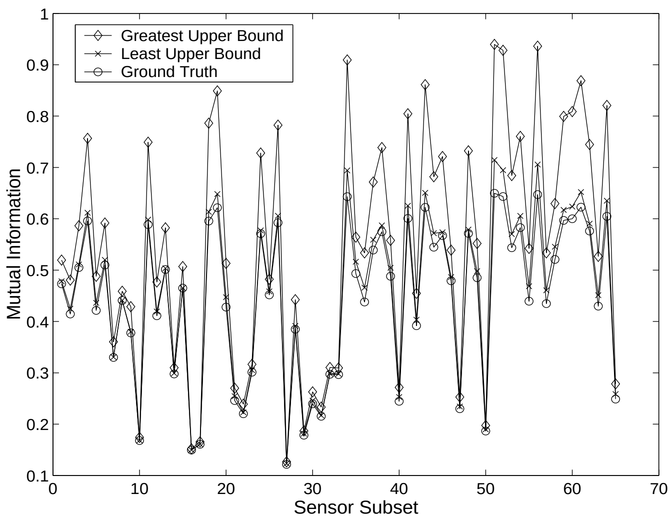
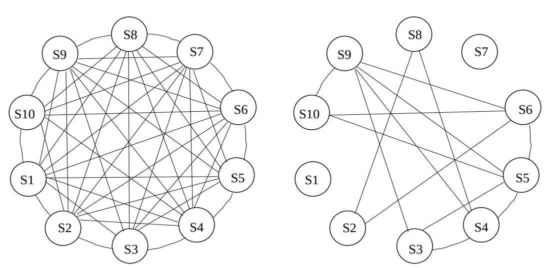


Figure 4: The least upper bound, greatest upper bound and ground truth of $I(\Theta, S)$, where the total number of sensor is 10. X-label 1-18: 5-sensor subsets; 19-34: 6-sensor subsets; 35-51: 7-sensor subsets; 52-66: 8-sensor subsets

other sensors. We prune the sensor synergy graph so that many weak sensor combinations are eliminated while preserving the most promising ones (the proof of this is beyond this paper), and the search space can be reduced significantly. We prune the synergy matrix of Eq. (5) by using

$$r_{ij} = \begin{cases} 1 & , \quad r(S_i, S_j) > \tau \\ 0 & , \quad otherwise, \end{cases} \tag{10}$$

where τ is a pruning threshold. Fig. 5 illustrates an example.



(a) (b)

Figure 5: (a) A completely-connected synergy graph; (b) a pruned synergy graph using the mean of r_{ij} as a threshold.

An Approximate Algorithm

Let $\mathbf{S}$ denote the current set of selected sensors and $\mathbf{S}'$ be a subset that forms a synergy chain in the synergy graph. Let $lub(\Theta; \mathbf{S})$ be the least upper bound of $I(\Theta; \mathbf{S})$ and $U(\cdot, \cdot)$ be a utility function as given in Eq. (1). n is the total number of available sensors. Our sensor selection algorithm based on the least upper bound of information gain is given in Table 1. Although the algorithm is greedy-like, the searching process in our approach is guided by a synergy graph so that the high quality of solution can be achieved.

```
SENSOR-SELECTION(n)
for each i, j, compute I(Θ; S_i) and I(Θ; S_i, S_j)
Construct a synergy graph G and prune it
Choose S_{i*}, S_{j*} such that U(I(Θ; S_i, S_j), C(S_i, S_j))
    is maximized for all i and j
S ← {S_{i*}, S_{j*}}
while |S| < n
    for each S', where |S'| = |S| + 1, and S' is
        a synergy chain on G and S ⊂ S'
        Find all Markov synergy chains of S'
        lub(Θ; S') ← arg min(I(Θ, S')), where
            I(Θ, S') is computed by Eq. (6), and min
            takes over all Markov synergy chains of S'
    S'* ← arg max(U(lub(Θ; S'), C(S'))), where
        max takes over all S'
    if U(lub(Θ; S'*), C(S'*)) > U(lub(Θ; S), C(S))
        S ← S'*
    else break
return S
```

Table 1: An approximate algorithm for sensor selection

Experiments

The ground truth of optimal sensor subset is obtained by a brute-force approach. We limit BN test model in 5 layers and up to 10 sensors (assuming binary valued) due to the exponential computation time behind the brute-force approach to get the ground truth. We construct 10 different BN models with 10 sensors, parameterized with 10 different conditional probability tables, yielding a total of 100 test models. The BN models with 7, 8 and 9 sensors are directly generated from the above models, e.g., 7-sensor models use S_1-S_7 from 10-sensor models.

If we only consider the information gain without the sensor cost, the algorithm can achieve almost 100% correctness in searching for optimal k out of n sensors from the test models. Our goal is to find the best subset of sensors among n sensors with a maximal utility. We assigned a cost for each individual sensor. The result averaged among 100 trials is given in Table 2, where the closeness is defined as a ratio of our solution to the ground truth. The result shows that the our solution is very close to the ground truth.

Table 2: The closeness of our solution to the ground truth

No. of Sensors	7	8	9	10
Closeness To Ground Truth	98.44%	98.23%	97.25%	98.11%

We evaluate the run time using a 2.0GHZ Pentium computer. For the convenience of comparison, the run time is measured on BN models with only two hidden layers including 7 hidden nodes. The models differ in the number of sensors from 7 to 10. The run time averaged among 10 trials is summarized in Table 3. The result shows that our approach can reduce the computation time significantly compared with the brute-force approach.

Table 3: The comparison of run time (seconds)

No. of Sensors	7	8	9	10
Brute-Force	63.87	355.05	2967.36	13560.54
Our Approach	1.020	1.099	1.209	1.430

Table 4: A case study of active sensor selection

Assessment Stage	Probability of Hypothesis	Blue Force Action Taken	Sensors Selected
1	P(Pas)=0.3333 P(Def)=0.3333 P(Off)=0.3333	Observe	S_2, S_5, S_6 S_7, S_9
2	P(Pas)=0.5170 P(Def)=0.3787 P(Off)=0.1043	Further Observe	S_8, S_9
3	P(Pas)=0.2590 P(Def)=0.5518 P(Off)=0.1892	Further Observe	S_5, S_6 S_7, S_9
4	P(Pas)=0.2164 P(Def)=0.6877 P(Off)=0.0959	Further Observe	S_5, S_7 S_8, S_9
5	P(Pas)=0.0213 P(Def)=0.8381 P(Off)=0.1406	Minor Offensive (Red Force May Change Intention)	

An Illustrative Application

We illustrate an application of multistage battlefield situation assessment to determine if the information gain can compensate the cost spent in information gathering. The scenario develops during a period of growing hostility between the Blue force and the Red force who poses a threat. A detail scenario can be seen in (Das 1999). The Blue force surveillance facilities include a number of offshore sensors, unmanned aerial vehicles (UAVs), surveillance helicopters (RAH66 Commanche), etc.. The Blue forces who are on duty in the restricted zone consists of 1) a Fremantle Class Patrol Boat (FCPB); 2) a Maritime Patrol Aircraft (MPA); 3) an Night Hawk Helicopter; 4) one F111 (Marritime Strike Aircraft). The Red forces include: 1) a major fleet unit; a Guided Missile Frigate (FFG); 2) one FCPB; 3) a communication ship. The Red force has two surface units armed with M386 and 110 SF Rocket Launcher that are ready to move to the locations where the Blue force is under their fire range.

A dynamic Bayesian network as shown in Fig. 6 is constructed to assess the situations for a battlefield scenario. The hypothesis node between two consecutive slices has temporal causality link and it is parameterized with a transition probability. There are 9 information sources available to supply information. The conditional probabilities and sensor costs are given subjectively. A Blue force commander has to select the sensors with the highest utility over time and take an appropriate action (Routine Action, Minor Offensive or Major Offensive), given a state of uncertainty about the hypothesis of the Red force intention (Passive, Defensive or Offensive). Table 4 presents a case study that sensors are actively selected over time based on the current state of hypothesis. P(Pas), P(Def) and P(Off) in the table denote the probability of Passive, Defensive and Offensive.

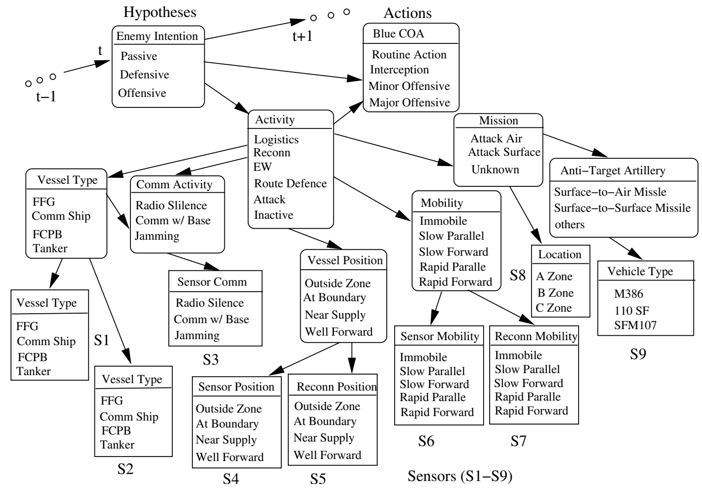


Figure 6: A dynamic Bayesian network model for the battlefield scenario. The BN structure remains unchange over time, only one slice si shown. S_1-S_9 are information sources (sensors). The part of structure adopted from (Das 1999).

Conclusion

It is computationally difficult to identify an optimal sensor subset with information-theoretic criterion in active information fusion. We presented a solution of finding a near optimal sensor subset by utilizing the sensor pairwise information to infer the synergy among sensors. The central thrust of this approach is to circumvent the computational difficulty in computing higher order mutual information by efficiently computing their least upper bound. Our approximate algorithm is greedy-like, but the searching process is guided by a synergy graph to achieve a near optimal solution, and to reduce the time complexity.

Acknowledgment

This material is based upon work supported by a grant from the US Army Research Office under Grant No. DAAD19-01-1-0402.

Appendix A: Proof of Theorems

A1. Proof of Theorem 1

Lemma 1 *(Chain Rule of Mutual Information) Let X, $Y_1, \cdots, Y_m$ are random variables, then*

$$I(X; Y_1, \cdots, Y_m) = I(X; Y_1) + \sum_{i=2}^{m} I(X; Y_i | Y_1, \cdots, Y_{i-1}). \quad (11)$$

Corollary 2 *Let X, Y, Z are random variables, then*

$$I(X; Z|Y) = I(X; Y, Z) - I(X; Y) \quad (12)$$

The proof of Lemma 1 and Corollary 2 is straightforward (Cover & Thomas 1991). We now prove Theorem 1.

Proof. Based on Lemma 1, we have

$$I(\Theta; S_1, \cdots, S_m) = I(\Theta; S_1) + I(\Theta; S_2|S_1) + I(\Theta; S_3|S_1, S_2) + \cdots + I(\Theta; S_m|S_1, \cdots, S_{m-1}) \quad (13)$$

From Definition 5 we have $I(\Theta; S_1|S_3) = 0$. Based on Venn diagram in Fig. 7, only the shaded region identifies $I(\Theta; S_3|S_1, S_2)$ while satisfying $I(\Theta; S_1|S_3) = 0$. While that region is also identified by $I(\Theta; S_3|S_2)$, i.e., $I(\Theta; S_3|S_1, S_2) = I(\Theta; S_3|S_2)$. Similarly, we have

$$\begin{aligned} &I(\Theta; S_4|S_2, S_3) = I(\Theta; S_4|S_3) \\ &\cdots\cdots\cdots \\ &I(\Theta; S_m|S_1, \cdots, S_{m-1}) = I(\Theta; S_m|S_{m-1}) \end{aligned} \tag{14}$$

Then, we immediately have

$$\begin{aligned} &I(\Theta; S_1, \cdots, S_m) \\ &\quad = I(\Theta; S_1) + \sum_{i=1}^{m-1} (I(\Theta; S_i, S_{i+1}) - I(\Theta; S_i)) \end{aligned} \tag{15}$$

The theorem is proved. ■

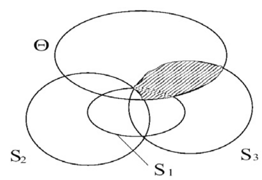


Figure 7: A Venn diagram represents the mutual information for 4 random variables.

A2. Proof of Theorem 2

Proof. Let us define $a = I(\Theta; S_1|S_2)$, $b = I(S_2; \Theta|S_1)$, $c = I(S_1; S_2|\Theta)$, $d = I(\Theta; S_1; S_2)$, $I(\Theta; S_3|S_2) = e$. With the symmetrical property of conditional mutual information, we have that $I(S_1; \Theta|S_2) = a$, $I(\Theta; S_2|S_1) = b$, $I(S_2; S_1|\Theta) = c$. We also have that $a, b, c, d, e \geq 0$. It can be easily verified that

$$I(\Theta; S_1) = I(\Theta; S_1|S_2) + I(\Theta; S_1; S_2) = a + d, \tag{16}$$

$$\begin{aligned} I(\Theta; S_1, S_2, S_3) = I(\Theta; S_1) + I(\Theta; S_2|S_1) \\ +I(\Theta; S_3|S_2) = a + d + b + e \end{aligned} \tag{17}$$

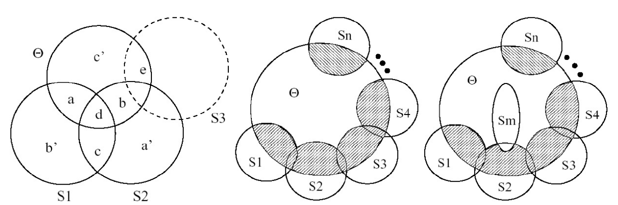


Figure 8: (a) A Venn diagram for a 3-sensor Markov synergy chain; (b) the shaded region identifies $I(\Theta; S_1, \cdots, S_n)$ of a Markov synergy chain; (c) the shaded region identifies $I(\Theta; S_1, \cdots, S_n)$ of a synergy chain

Therefore, for a Markov synergy chain $\{S_1, S_2, S_3\}$, $I(\Theta; S_1, S_2, S_3)$ can be identified by a Venn diagram as shown in Fig. 8(a). Continuing along the same thought above, then $I(\Theta; S_1, \cdots, S_n)$ can be represented by a Venn diagram of regions as shown in Fig. 8(b) for a Markov synergy chain $\{S_1, \cdots, S_n\}$. We assume a synergy chain has n sensors, and a sensor S_m, $m \notin \{1, \cdots, n\}$, interacts with this n-sensor synergy chain. The shaded Venn diagram of region identifies $I(\Theta; S_1, \cdots, S_n)$ as shown in Fig. 8(c). We can see that the shaded region of the Venn diagram in Fig. 8(c) is always less than or equal to that of its Markov synergy chain in Fig. 8(b). Thus, the theorem is proved. ■

References

Castanon, D. A. 1997. Approximate dynamic programming for sensor management. In *Proc. of the 36th IEEE Conf. on Decision and Control*, 1202–1207.

Cover, T. M., and Thomas, J. A. 1991. *Elements of Information Theory*. New York, NY, USA: John Wiley and Son.

Das, B. 1999. Representting uncertainties using bayesian networks. Technical Report DSTO-TR-0918, DSTO Electronics and Surveillance Research Laboratory, Salisbury South Australia 5108, Australia.

Denzler, J., and Brown, C. M. 2002. Information theoretic sensor data selection for active object recognition and state estimation. *IEEE Trans. Pattern Anal. Machine Intell.* 24(2):145–157.

Ertin, E.; Fisher, J.; and Potter, L. 2003. Maximum mutual information principle for dynamic sensor query. In *Proc. IPSN'03*.

Fassinut-Mombot, B., and Choquel, J.-B. 2004. A new probabilistic and entropy fusion approach for management of information sources. *Information Fusion* 5:35–47.

Heckerman, D.; Horvitz, E.; and Middleton, B. 1993. An approximate nonmyopic computation for value of information. *IEEE Trans. on PAMI* 15(3):292–298.

Kalandros, M.; Pao, L.; and Ho, Y.-C. 1999. Randomization and super-heuristics in choosing sensor sets for target tracking applications. 1803–1808.

Keeney, R. L., and Raiffa, H. 1993. *Decision with Multiple Objectives: Preferences and Value Tradeoffs*. Cambridge, United Kingdom: The University of Cambridge.

Krishnamurthy, V. 2002. Algorithms for optimal scheduling and management of Hidden Markov Model sensors. *IEEE Trans. on Signal Processing* 50(6):1382–1397.

Kristensen, S. 1997. Sensor planning with bayesian decision theory. *Robotics and Autonomous Systems* 19(3):273–286.

Liao, W.; Zhang, W.; and Ji, Q. 2004. A factor tree inference algorithm for bayesian networks and its application. In *the 16th IEEE Intl Conf. on Tools with Artificial Intelligence*.

Logothetis, A., and Isaksson, A. 1999. On sensor scheduling via information theoretic criteria. In *Proc. the American Control Conf.*, 2402–2406.

Miller, B. M., and Runggaldier, W. J. 1997. Optimization of observations: a stochastic control approach. *SIAM J. Control Optimal* 35(5):1030–1052.

Oliver, N., and Horvitz, E. 2003. Selective perception policies for guiding sensing and computation in multimodal systems: A comparative analysis. In *Proc. the 5th ICML*.

Paletta, L., and Pinz, A. 2000. Active object recognition by view integration and reinforcement learning. *Robotics and Autonomous System* (31):71–86.

Waltz, E., and Llinas, J. 1990. *Multisensor Data Fusion*. Boston, USA: Artech House.

Zhao, F.; Shin, J.; and Reich, J. 2002. Information-driven dynamic sensor collaboration for tracking application. *IEEE Signal Processing Magazine* 19(1):61–72.