Journal of Intelligent and Robotic Systems **29:** 161–189, 2000.

Entropy-Based Markov Chains for Multisensor Fusion

A. C. S. CHUNG and H. C. SHEN
Department of Computer Science, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong; e-mail: helens@cs.ust.hk

(Received: 16 August 1999; in final form: 10 January 2000)

Abstract. This paper proposes an entropy based Markov chain (EMC) fusion technique and demonstrates its applications in multisensor fusion. Self-entropy and conditional entropy, which measure how uncertain a sensor is about its own observation and joint observations respectively, are adopted. We use Markov chain as an observation combination process because of two major reasons: (a) the consensus output is a linear combination of the weighted local observations; and (b) the weight is the transition probability assigned by one sensor to another sensor. Experimental results show that the proposed approach can reduce the measurement uncertainty by aggregating multiple observations. The major benefits of this approach are: (a) single observation distributions and joint observation distributions between any two sensors are represented in polynomial form; (b) the consensus output is the linear combination of the weighted observations; and (c) the approach suppresses noisy and unreliable observations in the combination process.

Key words: decision making, entropy, Markov chains, multisensor fusion, uncertainty.

1. Introduction

In recent years, much interest in multisensor fusion has been generated in the fields of robotics, computer vision, remote sensing and medical imaging because of the general belief that multiple observations can reduce uncertainty. Thus, multisensor fusion is a science in studying the methods for combining multiple sensory observations into a consensus output such that classification and estimation accuracy can be improved by reducing the measurement uncertainty. The consensus output can be either (a) one of the possible classes for the classification problem, or (b) a random variable for the parameter estimation problem. In general, multisensor fusion includes *modelling of belief*, *observation combination* and *decision rule* [4], as shown in Figure 1. Modelling of belief, such as probabilistic or fuzzy set representation, concerns how the degree of belief in the observations and output are modelled. Observation combination defines the way to aggregate the beliefs in multiple observations for an output. Decision rule chooses the consensus output in order to maximize the occurrence probability or possibility for the classification problem, or minimize the variance for the parameter estimation problem.

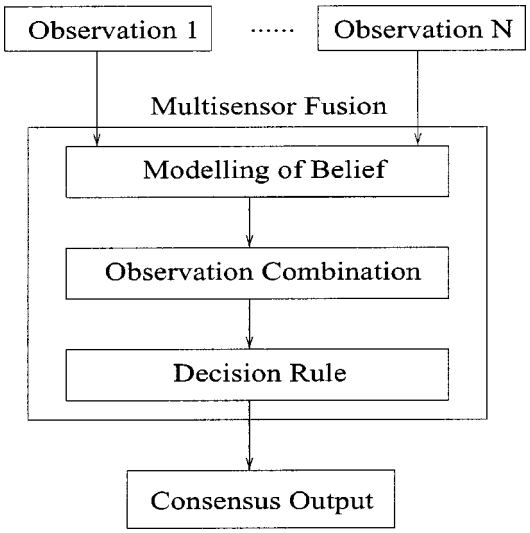


Figure 1. Components of multisensor fusion.

In multisensor fusion literature, there are numerous and context-dependent techniques. Among these techniques, certain properties are particularly important to the multisensor fusion performance. They are uncertainty reduction, observation interdependence, resource constraints and robustness.

Uncertainty reduction: this is a crucial property because it is absolutely discouraging if the combination of multiple observations brings about a consensus output of greater uncertainty. Reducing the uncertainty associated with the consensus output is the main goal of multisensor fusion. It is generally accepted that uncertainty is an unavoidable factor in sensory measurement. Uncertainty stems from low resolution of a sensor [3, 13], low representative competence of a feature [22], limited viewing angle or power of a visual sensor [27], and others.

Observation interdependence: two sensory observations can either be *dependent* or *independent*. Sensory observation dependence can help reduce system uncertainty. It is a concept for depicting the nature of interactions among a set of observations. Any two sensory observations are dependent if the values of one observation correlate with the values of another observation [8]; otherwise, they are independent. The features of observation dependence can be categorized into *independent, positively dependent* and *negatively dependent* on the basis of the change of the system uncertainty level [6, 7]. Thus, two sensor observations are independent if the level of uncertainty does not change after they are combined. Similarly, two sensor observations are defined to be positively dependent if the level of uncertainty is reduced after they are combined. As such, the negative analogue is equivalent to the negative dependence. By advocating the combination of positively dependent observations and avoiding the combination of negatively dependent observations, the system uncertainty level can be tremendously reduced.

As a result, the modelling of the positive and negative dependence becomes an integral part of the multisensor data fusion tasks and is of a particular interest in this paper. A good example [3] is the estimation of region occupancy probability. The authors show that the consideration of dependence among the observed regions and application of Bayesian fusion to integrate the information of three sensor readings can help to reduce the impact of false or bogus readings as compared with [13, p. 141], in which the author copes with the same problem and uses Bayesian fusion to update the grid points, but with the assumption of independence among them.

Resource constraints: most simple mobile robotic systems involve only two or more inputs. As the environment gets more complex, the number of inputs for performing area exploration and task accomplishment is significantly increased [14, p. 387]. System implementation resources, especially the size needed to represent the interrelations between the input sensory observations and the consensus output and computational time for observation combination, have all become key concerns in designing or selecting a multisensor fusion technique [16, p. 260]. A good multisensor fusion technique should show competence in (a) representing and updating the degree of belief about the multiple sensory observations effectively and (b) combining these observations efficiently even when the number of observations is large.

Robustness: this is another important factor in the sensory system, especially with a large scale system in which the possibility of having unreliable and noisy sensor is high. The rate of sensor failure is also high. A robust fusion technique should give stable performance when the sensory observations are deviated from the normal training data [18, p. 5; 2, p. 355]. The deviation may be caused by the sensor failure, the operation environment different from the training environment or sensor operation reliability, etc. These deviations are regarded as sensory input noise. A robust fusion is important because, if one of the sensors becomes unreliable or fails, the effect of the failure, which causes a large departure from the normal training data [25, p. 699], should not be amplified and affect significantly the performance of the whole system. When a sensor gives contaminated observation, its observation will become noisy and unreliable. As a result, a robust system should be capable of dealing with the noisy observations. Observation of a failed sensor can be gated or discarded automatically by a gating mechanism if we assume that the gating mechanism exists (see [24] for detail description of the sensor fault detection). However, this paper will not assume the existence of gating mechanism for failed sensors and will demonstrate a way to overcome the problem of unreliable observations.

This paper proposes an entropy-based Markov chain (EMC) fusion technique and demonstrate its applications in multisensor fusion. Experiments have been performed to demonstrate the proposed approach. The results show that our approach can reduce the measurement uncertainty by aggregating multiple observations. The major benefits of this approach are as follows:

(a) single observation distributions and joint observation distributions between any two sensors, which represent the interrelations between the sensory observations and the consensus output, are represented in polynomial form;
(b) the consensus output is a linear combination of the weighted observations, in which weights can be computed in polynomial time; and
(c) EMC is robust because it suppresses the noisy observation with high uncertainty level to minimize the contribution of the noisy and unreliable observation in the combination process.

The major contributions of this research are as follows:

(i) the dependent sensory observations are considered and the impact of considering observation interdependence is demonstrated in the experiments; and
(ii) the robustness of the EMC fusion technique is studied and compared with the Bayesian fusion.

2. Bayesian Fusion Technique

Bayesian fusion techniques has been used widely in a variety of application scenarios, including object detection [3, 13]; and mobile position estimation [20] in an unknown environment for mobile navigation; speech recognizing [31]; edge detection [23]; feature based object recognition [22] and its motion tracking [5]; wind shear avoidance [29] in aircraft guidance system; rain rate estimation [15] and area classification [27, 28] in geographical information systems; pixel classification of medical images [19]; pipe flaw detection [11]; and personal identity verification systems [21].

A fusion problem can be stated as follows: suppose that there is a common variable of interest $\theta \in \Theta$, which may be a discrete variable (i.e., pixel, object, person, an area, etc.) or a continuous variable (i.e., position, rain-rate, etc.), and there are m individual sensors. If the common variable θ is discrete, it should be assigned to one of the possible classes $\theta_1, \ldots, \theta_n \in \Theta$ on the basis of m sensory observations. If the common variable θ is continuous, its assigned value should be bounded by an interval $[\theta^+, \theta^-] \in \Theta$ on the basis of the m sensory observations. The sensory observations are represented by $\vec{x} = (x_1, \ldots, x_m) \in X^m$.

Modelling of belief. The degree of belief in the observations $\vec{x}$, given the observed variable θ, is represented by a function $l(\vec{x}|\theta)$ called the likelihood function to quantify the probability of occurrence of $\vec{x}$ given θ. $\pi(\theta)$ denotes the prior probability of the occurrence of the variable θ. This prior knowledge of the observed variable can be estimated initially by a uniform distribution if it is not known, and then updated during the experiment.

Combination. Likelihood function and prior probability are combined by the Bayes' rule,

$$p(\theta|\vec{x}) \propto l(\vec{x}|\theta) \times \pi(\theta),$$

where $p(\theta|\vec{x})$ represents the posterior probability that the observed variable is θ given the observations $\vec{x}$. The assumption of observation independence is not always valid. If it is assumed in the combination process, the computation of posterior probability is simpler. The posterior probability with the assumption of independence is of the form

$$p(\theta|\vec{x}) \propto \frac{1}{C} \prod_{i=1}^{m} l(x_i|\theta),$$

where $l(x_i|\theta)$ represents the likelihood that the observation is x_i given θ, and C is a normalization constant. Moreover, there is a number of different methods proposed to evaluate the posterior probability, e.g. the median of the individual univariate posterior probabilities [21], etc.

Decision rule – Maximum a Posterior (MAP). If the observed variable is an object which is discrete, the observed object is assigned to a class $\theta^* \in \Theta$ on the basis of the following maximization function

$$\theta^* = \arg\max_{\theta_i \in \Theta} p(\theta_i|\vec{x}).$$

This function maximizes the posterior probability $p(\theta|\vec{x})$ and is commonly used in the classification problem.

Decision rule – Minimum variance. If the observed variable is a random variable, then the expected value of the random variable θ is given by:

$$E[\theta|\vec{x}] = \begin{cases} \sum_{\theta \in \Theta} \theta p(\theta|\vec{x}), & \text{if } \theta \text{ is discrete,} \\ \int_{\theta \in \Theta} \theta p(\theta|\vec{x}) d\theta, & \text{if } \theta \text{ is continuous.} \end{cases}$$

It is noted that $E[\theta|\vec{x}]$ has a property of minimum variance, which is shown in [8, p. 25; 10, p. 197].

Bayesian fusion has an obvious advantage of having a clear formulation for describing fully the interdependent relationships between a set of sensory observations $x_1, \ldots, x_m \in X$ and consensus output $\theta \in \Theta$ [8]. The interrelations between the consensus output and sensory observations are depicted by a joint distribution $p(\theta, x_1, \ldots, x_m)$, which tabulates the relative frequencies or the probabilities of the event occurrences in the discrete or continuous sample space $\Theta \times X^m$. The size of the table of joint distribution grows exponentially as the number of observations m increases. Assuming that all the sample spaces are discrete, that there are m observations $x_1, \ldots, x_m$ in the same sample space X, that the size of the sample space is denoted by $|X|$, and that the size of the output space is denoted by $|\Theta|$,

the total number of events included in the joint distribution table is $|\Theta| \times |X|^m$. Therefore, the exponential increase in the joint distribution table size is one of the disadvantages of Bayesian fusion if the number of interdependent observations is large. Of course, if all the observations are assumed to be independent, the joint distribution table of $p(\theta, x_1, \ldots, x_m)$ is no longer needed and will be broken into m bivariate distribution tables for θ and x_i. The required size is then tremendously reduced to $m \times |\Theta| \times |X|$. However, the assumption of observation independence is not always valid [8]. It represents the loss of information about the interactions among the dependent variables.

As the size of the joint distribution table increases exponentially, the time needed to compute the posterior probability also increases exponentially. In Bayesian fusion, the combination rule is

$$p(\theta|\vec{x}) \propto l(\vec{x}|\theta)\pi(\theta),$$

where the likelihood function is in the form

$$l(\vec{x}|\theta) = \frac{p(\vec{x}|\theta)}{\sum_{\theta\in\Theta} p(\vec{x}, \theta)} = \frac{p(\vec{x}, \theta)}{\sum_{\theta\in\Theta} p(\vec{x}, \theta) \sum_{\vec{x}\in X} p(\vec{x}, \theta)}.$$

Hence, the time required to compute the posterior probability $p(\theta|\vec{x})$ is at least of $O(|\Theta|+|X|^m)$ time, which is exponentially increased as the number of observations m increases.

To conclude, this section reviews the basic formulation of the Bayesian fusion technique, and shows its deficiency. Firstly, the size of the table of joint distribution in the Bayesian fusion grows exponentially as the number of observations m increases. Secondly, the time needed to compute the posterior probability in the Bayesian fusion increases exponentially. In an attempt to overcome the deficiency, the entropy based Markov chain fusion technique is proposed.

3. Entropy-based Markov Chains (EMC)

3.1. MODELLING OF BELIEF

For entropy-based Markov chain (EMC) fusion, the degree of belief is represented by a probabilistic model. As such, the degree of belief in each single observation x_i, given the observed variable θ, is represented by a function $l(x_i|\theta)$ called likelihood function to quantify the probability of x_i occurrence given θ. For a joint observation x_i and x_j, the degree of belief, given the observed variable θ, is represented by the likelihood function $l(x_i, x_j|\theta)$ to represent the joint probability of x_i and x_j occurrence given θ. $\pi(\theta)$ denotes the prior probability of the occurrence of the variable θ. It is the prior knowledge of the observed variable. It can be estimated initially by a uniform distribution if it is not known, and then updated during the experiment.

3.2. OBSERVATION COMBINATION

3.2.1. *Entropy Measure*

In Section 1, it is noted that the sensory observation dependence can reduce system uncertainty. This section continues the discussion by first introducing entropy for uncertainty measurement. Self-entropy and conditional entropy are then adopted from [1] to measure the uncertainty level of the single observations x_i or x_j, and the joint observation x_i and x_j, respectively. The difference between conditional entropy and self-entropy of a sensor shows the change of the uncertainty level. If the uncertainty level of the joint observation is lower than that of the single observation, it means that the two observations x_i and x_j are positively dependent and the combination of the two observations is beneficial to uncertainty reduction. Similarly, if the two observations are negatively dependent, the combination is harmful to the uncertainty reduction.

Shannon's entropy h which was introduced by Shannon [26] in 1948, has long been used to measure the probabilistic uncertainty of a random variable θ. If the probability function $p(\theta)$ is discrete, then the entropy is given by

$$h = -\sum_{\theta \in \Theta} p(\theta) \log p(\theta), \tag{1}$$

whereas, if the probabilistic function $p(\theta)$ is continuous, then the entropy is given by

$$h = -\int_{\theta \in \Theta} p(\theta) \log p(\theta)\, \mathrm{d}\theta. \tag{2}$$

An essential property of the entropy is that entropy h is directly proportional to the degree of uncertainty (or randomness) of the measured variable; the smaller the uncertainty, the smaller the entropy. Let us assume that the probability distribution $p(\theta)$ follows a normal distribution with a p.d.f. $f(\theta|\mu, \sigma^2)$. It can easily be shown that, by Equation (2),

$$h = -\int_{\theta \in \Theta} f \ln f \, \mathrm{d}\theta = \ln(\sqrt{2\pi \mathrm{e}\sigma^2}), \tag{3}$$

where

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

Hence, if the probability density function is Gaussian, Equation (3) reveals that the smaller the variance σ^2, the smaller the entropy. It is noted that the uniform distribution will have the maximum entropy value. It is clear now that the entropy is sensitive and directly proportional to the degree of uncertainty. This property is obvious especially when the probability density function is Gaussian. However, the Gaussian assumption is not always valid. Hence, the computation of the

uncertainty level mainly depends on the entropy, which is given by the Equations (1) and (2). Suppose that there are two observations x_i and x_j of sensors i and j about θ, their single observation posterior distributions and joint observation posterior distribution are in the form

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

and

$$p_{ij}(\theta|x_i, x_j) \propto \pi(\theta) \times l(x_i, x_j|\theta), \tag{4}$$

where $\pi(\theta)$ 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. To measure the uncertainty level of the single observation and the joint observation, self-entropy and conditional entropy are adopted from [1] and are given as follows:

Self-entropy, which measures how uncertain a sensor is about its own observation x_i, is defined as

$$h_{i|i}(x_i) = -\sum_{\theta\in\Theta} p_i(\theta|x_i) \log p_i(\theta|x_i). \tag{5}$$

Conditional entropy, which 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

$$h_{i|j}(x_i, x_j) = -\sum_{x_j\in X_j} p(x_j|x_i) \sum_{\theta\in\Theta} p_{ij}(\theta|x_i, x_j) \log p_{ij}(\theta|x_i, x_j), \tag{6}$$

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 (6) 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 self-entropy and conditional entropy are within the range of zero and one, i.e., $0 \leqslant h_{i|i} \leqslant 1$ and $0 \leqslant h_{i|j} \leqslant 1$, because they are the relative entropy and the base of the log function is the sample size of the output spaces $|\Theta|$. For self-entropy, let us assume that the sample size of the observation space X_i is $|\Theta|$. Therefore, Equation (5) can be rewritten as

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

The maximum entropy is attained when the distribution $p_i(\theta|x_i)$ is uniform and is equal to $1/|\Theta|$. Therefore, the maximum self-entropy is equal to one, i.e.,

$$h_{i|i}(x_i) = -\sum_{\theta\in\Theta} \frac{1}{|\Theta|} \frac{\log 1/|\Theta|}{\log|\Theta|}.$$

The minimum self-entropy is equal to zero when the distribution $p_i(\theta|x_i)$ is certain, i.e., all values are zero except one value is equal to one. As such, the self-entropy is bounded by zero and one. The same implies for the conditional entropy.

The conditional entropy manifests profoundly the inter-dependence between sensory observations x_i and x_j. For example, if observation x_j is irrelevant to observation x_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, observation x_j does not help sensor i to improve its uncertainty level. On the other hand, if observation x_j is positively relevant to observation x_i, $h_{i|j}$ is expected to be smaller than $h_{i|i}$, which means that this observation contributes to reducing the uncertainty of observation x_i. Otherwise, the observation is negatively relevant, sensor i should at least maintain its uncertainty level. The properties of self entropy and conditional entropy can be 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 $h_{i|j}$, observations are independent (or irrelevant); and
- if the self-entropy $h_{i|i}$ is larger than the conditional entropy $h_{i|j}$, observations are positively dependent (or relevant); and
- if the self-entropy $h_{i|i}$ is smaller than the conditional entropy $h_{i|j}$, observations are negatively dependent (or relevant).

3.2.2. *Markov Chains*

In multisensor fusion, a Markov chain can be regarded as an iterative process for updating the observations of each sensor, which consists of a number of sensor states and transition probabilities [17]. Information about the observations is exchanged from one state to another state during the iteration process. Based on the information obtained, the observations of each sensor are updated after each iteration and will eventually converge into a single consensus output [9]. This output represents the consensus of a pool of m sensor observations. Durrant-Whyte [12] also proposed exchanges of information among the dependent sensors. However, his model differs from our proposed one in that his model depends on the product of observed and prior information. Whereas, our proposed model depends on the linear combination of observed and prior (second order posterior probabilities) information. As shown in Figure 2, each sensor state $s_i \in S$ represents a sensory observation $x_i \in X$. In this paper, the transition probability is called the *weight* (w_{ij}) which is assigned by sensor state s_i to sensor state s_j. Weight w_{ij} represents the certainty of observation x_j after observation x_i is known. u_i^k represents the updated observation at the kth iteration.

Markov chain recursively combines and updates the observations of $\vec{U}^{k-1}$ by the following iterative process,

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

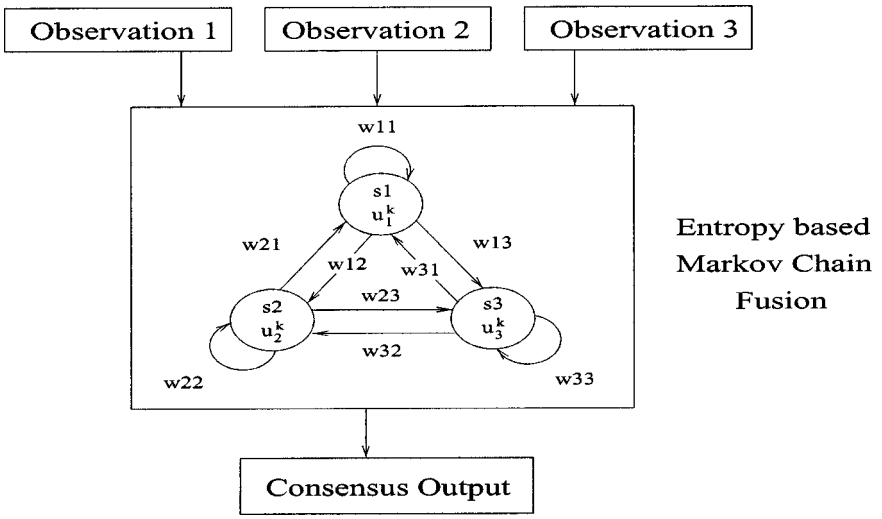


Figure 2. Entropy-based Markov chain (EMC) fusion for three observations.

which is equivalent to

$$\vec{U}^k = W^k \vec{U}^0, \tag{7}$$

where $\vec{U}^0 = (u_1^0, \ldots, u_m^0)^{\mathrm{T}}$ is the initial state vector of the initial local observations; $\vec{U}^k$ is the state vector at the kth iteration; and W is the transition matrix whose non-negative 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 $0 \leqslant w_{ij} \leqslant 1$.

As k approaches infinity, $\vec{U}^k$ converges to a consensus value u^*,

$$u^* = \vec{\mathcal{K}}^{\mathrm{T}} \vec{U}^0 = \sum_{i=1}^{m} \kappa_i u_i^0. \tag{8}$$

and

$$W^{\mathrm{T}} \vec{\mathcal{K}} = \vec{\mathcal{K}}, \tag{9}$$

where $\vec{\mathcal{K}} = (\kappa_1, \ldots, \kappa_m)^{\mathrm{T}}$ is the vector of stationary transition probabilities; $\sum_{i=1}^{m} \kappa_i = 1$; and $0 \leqslant \kappa_i \leqslant 1$. Equation (8) reveals that the consensus output is the linear combination of the initial local observations. From Equation (7), the convergence of the matrix W^k is reached if

- there exists a positive integer k such that every element in at least one column of the matrix W^k is positive [9], or
- all the recurrent states of the Markov chain communicate with each other (irreducible) and are positive recurrent and aperiodic [9; 17, p. 574].

Equation (9) can be viewed as the eigenvector problem $A \cdot \vec{x} = \lambda \vec{x}$ with eigenvalue λ 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. The vector of stationary transition probabilities $\vec{\mathcal{K}}$ can be found in polynomial time $O(m^3)$ [25].

It is observed that, from Equations (7) and (8), κ_i is large if and only if w_{ji} is large for $j = 1, \ldots, m$, that is,

$$\kappa_i \propto w_{ji}. \tag{10}$$

κ_i is proportional to w_{ji}. 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.

3.2.3. *Entropy-based 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 conditional entropy. The same argument applies to 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:

$$w_{ij} \propto \frac{1}{h_{j|i}^{\alpha_{ij}}} \quad \text{for } i, j = 1, \ldots, m,$$

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 x_i, and α_{ij} reflects this dependence. If the self-entropy $h_{j|j}$ is smaller than the conditional entropy $h_{j|i}$, implying negative dependence, $h_{j|i}$ will be set to $h_{j|j}$, and α_{ij} is set to one, i.e., $\alpha_{ij} = 1$. If the self-entropy $h_{j|j}$ is larger than the conditional entropy $h_{j|i}$, implying positive dependence, α_{ij} should be larger than one. The difference between $h_{j|j}$ and $h_{j|i}$ is then added to α_{ij}, i.e., $\alpha_{ij} = 1 + (h_{j|j} - h_{j|i})$. For computation simplicity, $(h_{j|j} - h_{j|i})$ can be set to the nearest one decimal place. It is obvious that, since $h_{j|j}$ and $h_{j|i}$ are bounded between zero and one, α_{ij} is bounded by one and two.

The greater the α_{ij}, the smaller the entropy (less uncertain) and the larger the weight. Therefore, α_{ij} describes the sensitivity of the entropy measure. It is then

written in matrix form,

$$\begin{bmatrix} w_{11} & \dots & w_{1m} \\ w_{21} & \dots & w_{2m} \\ \vdots & \ddots & \vdots \\ w_{m1} & \dots & w_{mm} \end{bmatrix} \propto \begin{bmatrix} \frac{1}{h_{1|1}^{\alpha_{11}}} & \cdots & \frac{1}{h_{m|1}^{\alpha_{m1}}} \\ \frac{1}{h_{1|2}^{\alpha_{12}}} & \cdots & \frac{1}{h_{m|2}^{\alpha_{m2}}} \\ \vdots & \ddots & \vdots \\ \frac{1}{h_{1|m}^{\alpha_{1m}}} & \cdots & \frac{1}{h_{m|m}^{\alpha_{mm}}} \end{bmatrix}.$$

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

$$w_{ij} = \frac{1}{h_{j|i}^{\alpha_{ji}} \sum_{k=1}^{m} (1/h_{k|i}^{\alpha_{ki}})} \quad \text{for } i, j = 1, \dots, m. \tag{11}$$

3.2.4. *Properties of Weights*

It is observed that weight w_{ij} is a function of self-entropy and conditional-entropy, which are functions of single (univariate) observation distribution $p(\theta, x_i)$ and joint (bivariate) observation distribution $p(\theta, x_i, x_j)$ between two sensors, respectively. It is worth noting that only the univariate and bivariate observation distributions are needed throughout the decision process. However, in the Bayesian model, which will be discussed later, higher order distributions are necessary.

Attention should be paid to the cases whereby $h_{i|j}$ diminishes to zero. For nonzero $h_{i|j}$, Equation (11) works perfectly. 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 $i = 1, \dots, m$. When more than one sensor happens to be absolutely certain, normalization across the transition matrix row is necessary. If the condition of convergence is not satisfied, the concensus output will be the weighted average of the initial local observations.

3.2.5. *Local Estimated Observation*

Equation (8) shows that global consensus output is a weighted sum of local observations. The local observation can be preprocessed before the iteration process starts. The local observation 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 estimated observation is given by

$$E\left[u_i^0(\theta)\right] = \sum_{j=1}^{m} w_{ij} \sum_{\theta \in \Theta} \theta p_{ij}(\theta|x_i, x_j), \tag{12}$$

for $i = 1, \ldots, m$, where w_{ij} is defined by Equation (11). It is noted that $p_{ij}(\theta|x_i, x_j)$ is set to $p_{ij}(\theta|x_i)$ when $h_{i|i} < h_{i|j}$ because the uncertainty level should at least be maintained in the case of negative relevance.

By rewriting Equation (12), the local estimated observation can be viewed as an expected value of θ from a *local distribution* $u_i^0(\theta)$.

$$E\big[u_i^0(\theta)\big] = \sum_{\theta \in \Theta} \theta u_i^0(\theta), \tag{13}$$

where the local distribution is given by

$$u_i^0(\theta) = \sum_{j=1}^{m} w_{ij}\, p_{ij}(\theta|x_i, x_j). \tag{14}$$

Similarly, from Equation (8), consensus output u^* can be found by substituting the local observations u_i^0 by local estimated observations $E[u_i^0(\theta)]$ and is given by

$$u^* = \sum_{i=1}^{m} \kappa_i E\big[u_i^0(\theta)\big]. \tag{15}$$

3.3. DECISION RULE

For an observation assumed to be a random variable, Equation (8) gives the consensus decision. On the other hand, if the observed variable is an object which is discrete, the consensus decision is a classification issue, i.e., the observed object is assigned to a class $\theta^* \in \Theta$ such that θ^* has the highest posterior probability. If the local observations u_i^0 are the local posterior probabilities $p_i^0(\theta|x_i)$, the consensus decision is the consensus posterior probability $p^*(\theta|\vec{x})$, which is given by

$$p^*(\theta|\vec{x}) = \sum_{i=1}^{m} \kappa_i\, p_i^0(\theta|x_i),$$

where

$$\vec{\mathcal{K}} = (\kappa_1, \ldots, \kappa_m)^{\mathrm{T}}, \qquad \sum_{i=1}^{m} \kappa_i = 1, \quad 0 \leqslant \kappa_i \leqslant 1.$$

Therefore, the observed object is assigned to a class $\theta^* \in \Theta$ on the basis of the following maximization function

$$\theta^* = \arg\max_{\theta_i \in \Theta} p^*(\theta_i|\vec{x}).$$

This function maximizes the consensus posterior probability $p^*(\theta|\vec{x})$.

3.4. PROPERTIES OF EMC

Entropy based Markov chain (EMC) fusion exchanges sensory observations among the sensors and updates iteratively the local observations on the basis of the uncertainty levels of local observations and joint observations between two sensors. Self-entropy and conditional entropy represent the uncertainty levels of local observations and joint observations respectively. The consensus output is the convergent value of the iteration process. Suppose that there are m observations with the same sample space X. The consensus output space is denoted by Θ. The interdependence among the observations is modelled by single observation distribution $p(\theta, x_i)$ and joint observation distribution $p(\theta, x_i, x_j)$. Since there are m single observation distribution and C_2^m joint observation distributions, the size required to represent all the tables of distributions is

$$m \times |\Theta| \times |X| + C_2^m \times |\Theta| \times |X|^2.$$

Therefore, the size of the distribution tables increases polynomially when the number of observations increases. EMC combines observations by following the evaluation steps below:

Entropy. If each sensor knows all the sensory observations, self-entropy and conditional entropy will be in the form

$$\begin{aligned} h_{i|i}(x_i) &= -\sum_{\theta\in\Theta} p_i(\theta|x_i)\log p_i(\theta|x_i), \\ h_{i|j}(x_i, x_j) &= -\sum_{\theta\in\Theta} p_{ij}(\theta|x_i, x_j)\log p_{ij}(\theta|x_i, x_j), \quad \text{for } i, j = 1, \ldots, m, \end{aligned}$$

respectively. For m observations, there are m self-entropy values, $h_{i|i}(x_i)$ and $C_2^m/2$ conditional entropy values, $h_{i|j}(x_i, x_j)$. The time required to compute the posterior probabilities $p_i(\theta|x_i)$ and $p_{ij}(\theta|x_i, x_j)$ are of $O(|\Theta| + |X|)$ and $O(|\Theta| + |X|^2)$, respectively. The required time to compute self-entropy $h_{i|i}$ and conditional entropy $h_{i|j}$ are of $O(|\Theta|^2 + |\Theta||X|)$ and $O(|\Theta|^2 + |\Theta||X|^2)$, respectively. Therefore, the total time for evaluating all the entropy is of $O(m^2) \times O(|\Theta|^2 + |\Theta||X|^2)$.

Weight Matrix. From Equation (11), weight matrix W is evaluated by the following expression,

$$w_{ij} = \frac{1}{h_{j|i}^{\alpha_{ji}} \sum_{k=1}^{m} (1/h_{k|i}^{\alpha_{ki}})} \quad \text{for } i, j = 1, \ldots, m.$$

The time required to compute the weight matrix W is of $O(m^2)$.

Stationary Transition Probability. From Equation (9), the vector of stationary transition probabilities $\vec{\mathcal{K}} = (\kappa_1, \ldots, \kappa_m)^{\mathrm{T}}$ is computed as follows

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

This can be viewed as the eigenvector problem $A \cdot \vec{x} = \lambda \vec{x}$ with eigenvalue λ equal to one. Thus, the vector of stationary transition probabilities $\vec{\mathcal{K}}$ can be found in polynomial time $O(m^3)$ [25].

Linear Combination of the Weighted Observations. From Equation (8), the consensus output is evaluated as follows

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

Therefore, the combination process including entropy, weight matrix, stationary transition probability evaluations and linear combination of the weighted observations requires polynomial time.

Concerning the robustness of the entropy based Markov chain (EMC) fusion, let us assume that there are m observations. From Equation (8), the local observations are combined to get a consensus output u^*, that is,

$$u^* = \kappa_1 u_1^0 + \cdots + \kappa_i u_i^0 + \cdots + \kappa_m u_m^0.$$

The stationary probability κ_i reflects the weighting of sensor i's observation in the combination process. From Equation (10), i.e., $\kappa_i \propto w_{ji}$, it is clear that the stationary probability κ_i is directly proportional to the weights w_{ji} which is inversely proportional to the conditional entropy $h_{i|j}$. Thus, the stationary probability κ_i is also inversely proportional to the entropy $h_{i|j}$ for $j = 1, \ldots, m$, $\kappa_i \propto 1/h_{i|j}$.

If the local observation u_i^0 is replaced by the weighted sum of the local bivariate distributions, and the consensus output is replaced by the posterior distribution, as shown in Equation (14). The Equation (8) can be rewritten as

$$p^*(\theta|\vec{x}) = (\kappa_1, \ldots, \kappa_m) \begin{pmatrix} w_{11}p_1 + w_{12}p_{12} + \cdots + w_{1m}p_{1m} \\ w_{21}p_{12} + w_{22}p_2 + \cdots + w_{2m}p_{2m} \\ \vdots \\ w_{m1}p_{1m} + w_{m2}p_{2m} + \cdots + w_{mm}p_m \end{pmatrix}, \tag{16}$$

where $p_{ij} = p_{ij}(\theta|x_i, x_j)$ and $p_i = p_i(\theta|x_i)$.

Assuming that an observation of sensor i deviates from the normal training data, the uncertainty levels $h_{i|j}$ of the observation will increase accordingly. As such, the associated weights w_{ji} will be decreased. The stationary probability κ_i will be decreased to reflect the increment of the uncertainty level.

Therefore, the univariate and bivariate distributions p_i, p_{ij} and p_{ji} of the non-normal and noisy observation x_i in the combination process are assigned the smaller weights such that the contribution of the noisy observation is minimized. It is generally assumed that if an observation is abnormal or deviated from the normal, it will have a larger variance. Thus, EMC will suppress the univariate and bivariate distributions of the noisy observation with high uncertainty level to minimize the contribution of the noisy and unreliable observation.

To summarize, the entropy based Markov chain (EMC) fusion is different from the Bayesian fusion based on the fact that

(a) univariate observation distributions and bivariate observation distributions between any two sensors, which represent the interrelations between the sensory observations and the consensus output, are represented in polynomial form;
(b) the consensus output is the linear combination of the weighted local estimated observations, in which weights can be computed in polynomial time; and
(c) EMC is robust because it suppresses the noisy observation with high uncertainty level to minimize the contribution of the noisy and unreliable observation in the combination process.

4. Experiments

The proposed entropy based Markov chain fusion technique will be illustrated by two experiments:

(I) aggregation of three optical sensors with possible application in object recognition on a conveyor belt; and
(II) color mapping from the RGB space to the CMYK space with possible application in the printing industry.

For each experiment, three sets of results are obtained under three different conditions, namely,

(i) individual sensors without applying any fusion techniques;
(ii) applying the proposed entropy based Markov chain fusion technique and assuming sensors are independent; and
(iii) similar to (ii) but assuming sensors are dependent.

Condition (i) serves as one basis of comparison. Condition (ii) considers the common assumption made by other approaches, [13, 19, 22, 27]. Serving as a basis of comparison in terms of accuracy, the traditional Bayesian fusion technique is also implemented. Under the assumption of independence, the conditional entropy will be set to be the same as the self-entropy, i.e., $h_{i|j} = h_{i|i}$, i.e., observations made by sensor j will no longer affect observations made by sensor i. Thus, the weight matrix W, (Equation (11)), will be reduced to a matrix with identical columns. And, the posterior probability for the EMC fusion technique takes the form

$$p(\theta|\vec{x}) = \sum_{i=1}^{m} \kappa_i \, p(\theta|x_i),$$

where θ is the consensus output (decision), $\vec{x} = (x_1, \ldots, x_m)$ is the vector of decisions made by m sensors and $p(\theta|x_i)$ is the univariate posterior probability.

On the other hand, the posterior probability for the Bayesian fusion technique is given by

$$p(\theta|\vec{x}) = \prod_{i=1}^{m} p(\theta|x_i).$$

Under the assumption of dependence, the posterior probability of the EMC fusion technique is given by Equation (16),

$$p(\theta|\vec{x}) = (\kappa_1, \ldots, \kappa_m) \begin{pmatrix} w_{11}p_1 + w_{12}p_{12} + \cdots + w_{1m}p_{1m} \\ w_{21}p_{12} + w_{22}p_2 + \cdots + w_{2m}p_{2m} \\ \vdots \\ w_{m1}p_{1m} + w_{m2}p_{2m} + \cdots + w_{mm}p_m \end{pmatrix}.$$

Whereas, the posterior probability for the Bayesian approach is given by

$$p(\theta|\vec{x}) \propto l(\vec{x}|\theta)\pi(\theta),$$

where the likelihood function is in the form

$$l(\vec{x}|\theta) = \frac{p(\vec{x}|\theta)}{\sum_{\theta\in\Theta} p(\vec{x},\theta)} = \frac{p(\vec{x},\theta)}{\sum_{\theta\in\Theta} p(\vec{x},\theta) \sum_{\vec{x}\in X} p(\vec{x},\theta)}.$$

To demonstrate the robustness of the proposed fusion technique, noise is also introduced into the test data in our experiments. For each experiment, $p(\theta|x)$ is computed from the training data and Guassian is assumed too. The following sections give the description of each experiment and the corresponding results obtained.

4.1. EXPERIMENT I – THREE OPTICAL SENSORS

This section shows the application of the proposed entropy based Markov chains fusion technique to aggregate the observations of three optical crossbeam sensors such that the object carried by the conveyer belt is classified. The observation of the crossbeam sensor is the length of the side view of the object parallel to the conveyer belt. The setup of the experiment is shown in Figures 3 and 4. The setup consists of a conveyer belt, a speed decoder, three pulse counters and three crossbeam sensors. In Figure 4, the conveyer belt carries the objects leftwards to pass through the optical crossbeam sensors. In this experiment, there are three crossbeam sensors and the angle between them is 30 degree. Crossbeam sensor [30] consists of a pair optical transmitter and receiver, and is simply an on and off optical device. As shown in Figure 5(a), when the object blocks the connection between T_2 and R_2, the crossbeam sensor sends a signal to switch on the pulse counter, x_2, which starts counting. The crossbeam sends a signal to switch off the pulse counter when there is no object between the transmitter and the receiver, as shown in Figure 5(b). D is

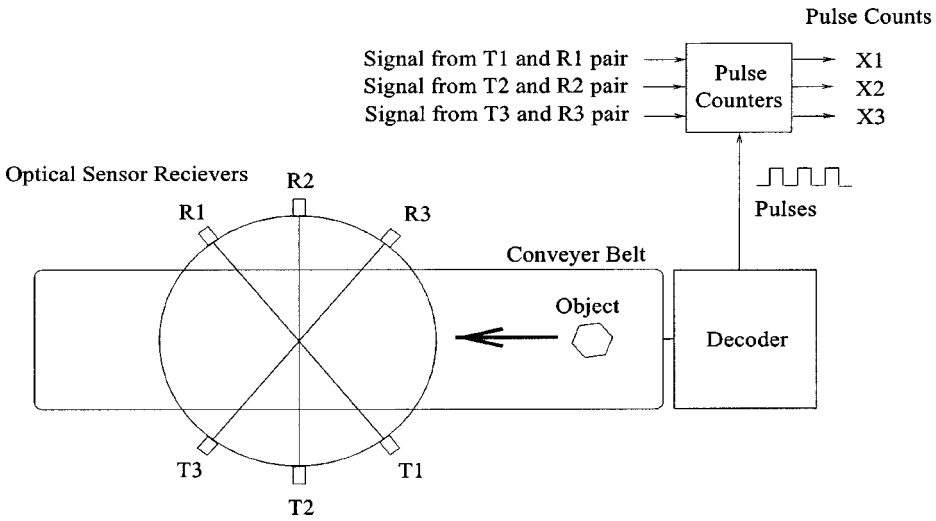


Figure 3. Diagram of the 3-optical crossbeam sensors.

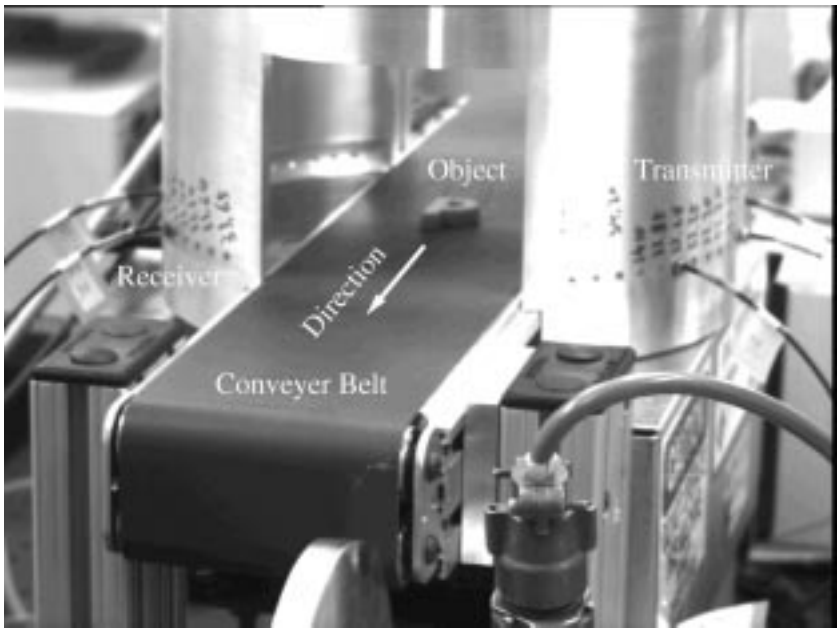


Figure 4. Physical setup of the crossbeam sensors and the conveyer belt.

the distance measured by the sensor and is directly proportional to the number of pulses recorded during the period when the connection between transmitter and receiver is blocked. The same applies to $T_1 - R_1$ and $T_3 - R_3$.

The decoder generates a pulse for each 0.01 mm movement of the conveyer belt. Therefore, the distance D between two extreme points 'viewed' by the crossbeam sensor can be computed by the multiplication of the number of counted pulses X_2

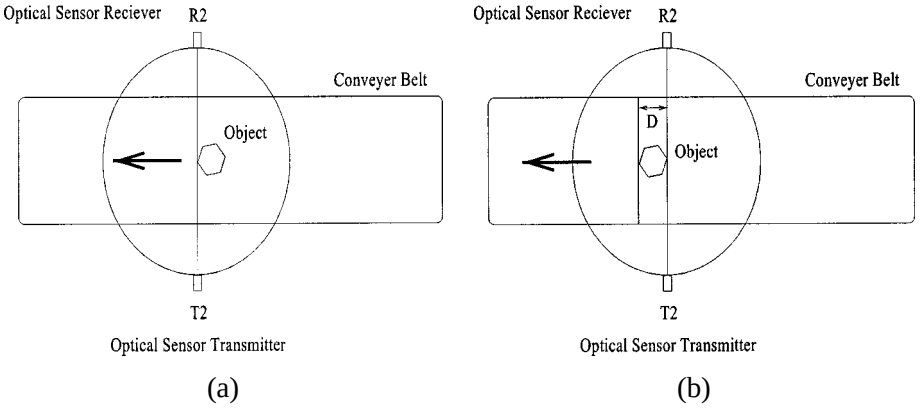


Figure 5. Detection process: (a) object blocks the connection; (b) object unblocks the connection.

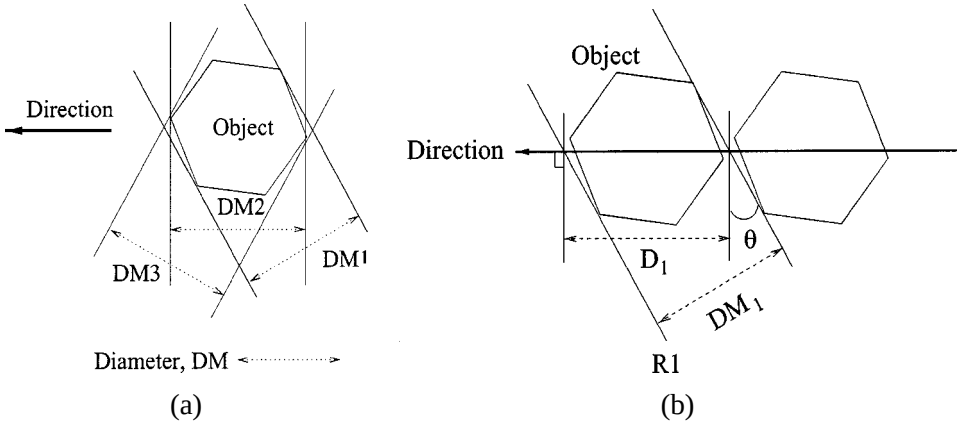


Figure 6. Computation of the diameter: (a) definition; (b) relation to pulse counter.

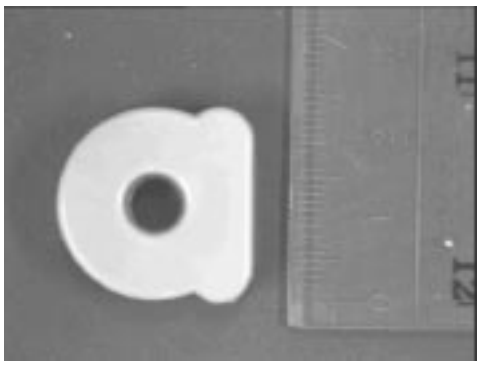
and distance (0.01 mm) per pulse. According to Wallack and Canny [30], the diameter is defined as the distance between the two extreme points given a fixed orientation, as shown in Figure 6(a). However, the diameter is not the distance measured D if the sensor is not perpendicular to the direction of the object movement. As shown in Figure 6(b), the diameter DM_i is then calculated by

$$DM_i = D_i \cos\theta_i \tag{17}$$

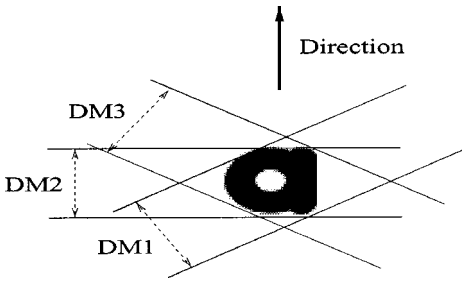
$$= X_i \times 0.01 \cos\theta_i . \tag{18}$$

Therefore, the diameter DM_i is directly proportional to the counted pulses, $DM_i \propto X_i$, for $i = 1, 2$ and 3. The counted pulse readings X_1, X_2, X_3 can then be used directly to classify the 'observed' object.

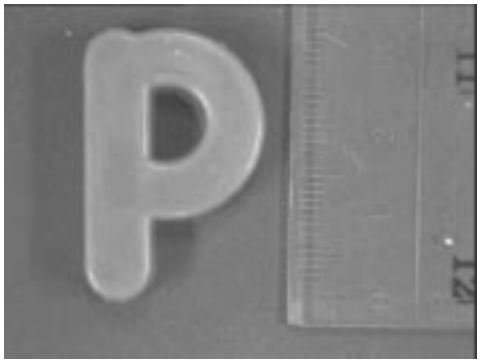
In this experiment, four objects are used. They are characters 'a', 'p', 'i' and 't', as shown in Figures 7(a), (c), (e) and (g), respectively. For these four objects, the diameters viewed by the different crossbeam sensors are shown in Figures 7(b),



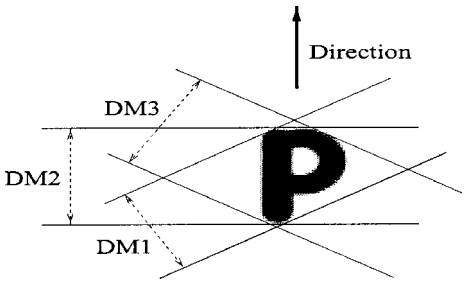
(a) Character ‘a’



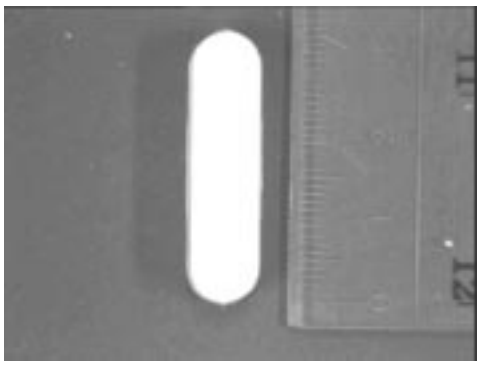
(b) Diameters of ‘a’



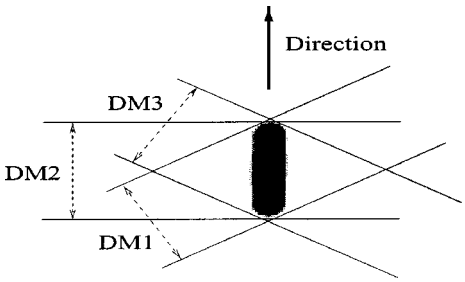
(c) Character ‘p’



(d) Diameters of ‘p’

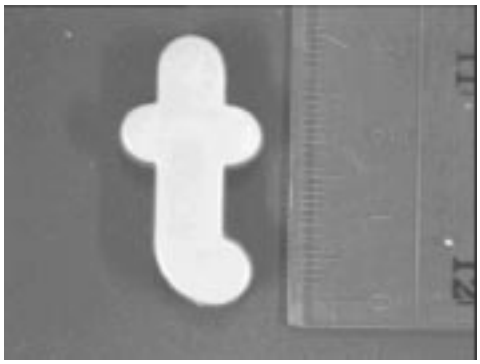


(e) Character ‘i’

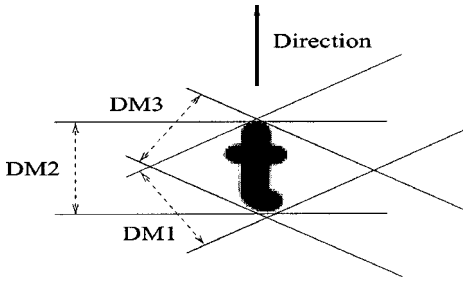


(f) Diameters of ‘i’

Figure 7. Characters used in the experiment and their corresponding diameters.



(g) Character 't'



(h) Diameters of 't'

Figure 7. (Continues).

(d), (f) and (h), respectively. As the object is being carried by the conveyer belt, the diameters recorded by the sensors are dependent on the object orientation. Intuitively, the chance of misclassification for characters 'a' and 'i' is relatively smaller than that for characters 'i' and 't'. Characters 'i' and 't' will be easier to be misclassified if there is only one optical sensor. It is believed that the addition of optical sensors can improve classification accuracy. Therefore, the purpose of the aggregating the observations of the three optical sensors is to reduce the uncertainty that is inherited by any individual sensor. It is obvious that the observations of the three optical sensors are dependent and complementary. Let us assume that the observed object is denoted by o and the sample space is denoted by $O = (o_1, o_2, o_3, o_4) = (a, p, i, t)$. Given a vector of sensory observations x_1, x_2, x_3, the posterior probability $p(o|x_1, x_2, x_3)$ that the object is either one of the possible objects in O will be computed by the proposed entropy based Markov chain fusion technique and Bayesian fusion. The observed object will then be assigned to either one of the possible objects with the highest posterior probability, that is,

$$o = \arg\max_{o_i} p(o_i|x_1, x_2, x_3).$$

4.1.1. *Results from Individual Sensors*

Each object was carried by the conveyer belt, and passed through the three optical crossbeam sensors 1000 times. Hence, there are in total 4000 sets of observations, each from three sensors, (x_1, x_2, x_3). From Table I, as expected, the classification error of each individual crossbeam sensor is large because the objects 'a' and 'p', 'i' and 't' have similar shapes and are quite difficult to be identified in one dimension. Sensor 2 has larger classification error, which can be explained as follows: As shown in Figure 3, the viewing angles of sensor 1 and 3 are smaller (30 degree) than that of the sensor 2 (90 degree from the line of object movement). From

Table I. Classification error of the individual crossbeam sensors

Sensor number	Classification error %
1	69.8
2	75.2
3	69.8

Equation (18), the relation between the diameter and number of counted pulses is

$$X_i = \frac{DM_i}{0.01 \cos \theta_i},$$

where X_i is the number of counted pulses, DM_i is the diameter viewed by sensor i and θ_i is the viewing angle of sensor i from the line of object movement. Therefore, when the viewing angle θ_i decreases, a small difference in diameter DM_i, e.g., character 'i' and 't', is magnified into a larger difference in term of the number of counted pulses X_i. Therefore, sensor 1 and 3 are more sensitive to the small difference in diameter than sensor 1, and hence give relatively smaller classification errors.

4.1.2. *Results from Fusion Techniques*

Two fusion techniques, the proposed EMC and the Bayesian, under two assumptions, independent and dependent were implemented for comparison. Table II shows

(i) error is reduced as compared to that given in Table I;
(ii) results from the EMC fusion technique in terms of classification error is comparable to that of the Bayesian approach; and
(iii) consideration of dependent relations improves further the classification errors.

Noise in terms of number of pulses N is added to each of the testing sensory observations individually, i.e., $(x_1 \pm N, x_2 \pm N, x_3 \pm N)$, where 10 values of N ranging from 10 to 100 were tested. In our experiments, we have found that the probability of the beam sensor giving pulses higher than 50 error pulses is quite low (approximately 1.19%). Therefore, we experimented up to 100. Figure 8 give the classification errors when noise at different levels are introduced. Note that in all cases, the deviation from those without noise is negligible, within $\pm 0.2\%$ for the EMC technique and a maximum of 0.4% for the Bayesian one.

Table II. Classification errors from fusion techniques

Fusion technique	Classification error % Independent	Classification error % Dependent
EMC	56.3	17.6
Bayesian	56.3	17.5

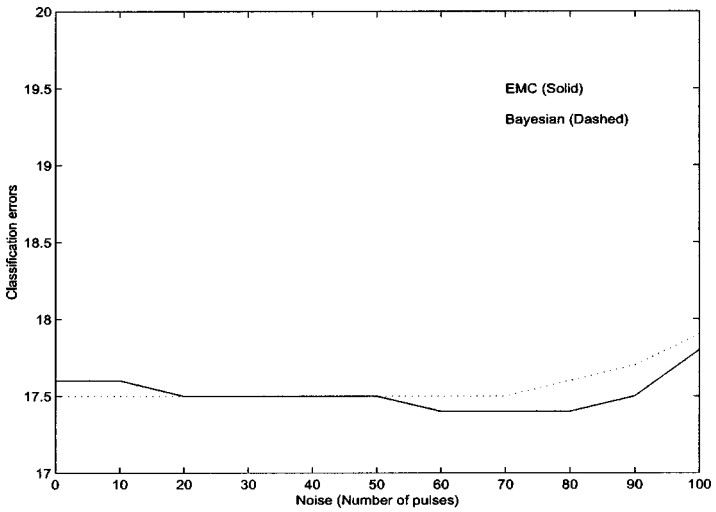


Figure 8. Classification errors (%) with noisy input data (dependent assumption).

4.2. EXPERIMENT II – COLOR MAPPING FROM RGB SPACE TO CMYK SPACE

The proposed entropy based Markov chain fusion technique is further demonstrated in color mapping applications. The experiment only concerns the color mapping from the RGB color space to the CMYK color space. RGB denotes the Red, Green and Blue primary colors while CMYK denotes Cyan, Yellow, Magenta, and blacK basic colors commonly used in the printing industry. Each CMYK representation is expected to be some combination of the three stimuli in the RBG representation. Their relationships can best be depicted by a Bayesian network, as shown in Figure 9. We shall assume that the values of stimuli C, M, Y and K are denoted by c, m, y, k and the sample space of each stimulus is denoted by $C = (c_1, \ldots, c_{12})$, $M = (m_1, \ldots, m_{12})$, $Y = (y_1, \ldots, y_{12})$ and $K = (k_1, \ldots, k_7)$. Similarly, the sample spaces of Red, Green and Blue in the RGB representation are denoted by $R = (r_1, \ldots, r_{256})$, $G = (g_1, \ldots, g_{256})$ and $B = (b_1, \ldots, b_{256})$, respectively. Suppose that probabilistic model is employed, given a RGB representation (r, g, b), the posterior probability that values of c, m, y, k are $c_i \in C$, $m_i \in M$, $y_i \in Y$, $k_i \in K$, respectively, is computed by aggregating the values of the three stimuli (RGB), that is $p(c, m, y, k|r, g, b)$. The

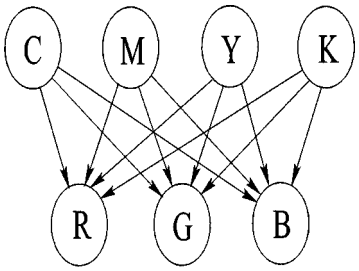


Figure 9. Bayesian network describing the interdependent relations between RGB and CMYK stimuli.

value of the combined posterior probability is then estimated by the proposed-entropy based Markov chain fusion technique. The stimuli (c, m, y, k) will then be assigned a value in one of the possible classes in (C, M, Y, K) such that the combined posterior probability is the highest, that is

$$(c, m, y, k) = \arg \max_{(c_i, m_i, y_i, k_i)} p(c_i, m_i, y_i, k_i | r, g, b).$$

The results will be compared with that obtained by the Bayesian fusion technique.

4.2.1. *Results from Individual Stimulus*

A slight variation in the value of stimulus might not be detected by human eye easily. In the CMYK representation, if the stimulus value changes within the range of 20, it still gives similar color appearance. Therefore, given a CMYK representation (C, M, Y, K), any CMYK representation (C_i, M_i, Y_i, K_i) will give similar color appearance and be treated as the 'matched' representation if the sum of differences between each pair of stimuli is less then 20, i.e.,

if $(|C - C_i| + |M - M_i| + |Y - Y_i| + |K - K_i|) \leqslant 20$,
then *SIMILAR COLOR APPEARENCE*.

Based on the given training data, which gives a table of mappings from CMYK to RGB, one RGB representation may match up to five CMYK representations. These multiple mappings cause the major matching errors. Therefore, in most mismatched cases, given the RGB representation, it gives more than one CMYK representations, which have equal and the highest posterior probabilities. In this situation, the system has to select one CMYK representations randomly among these representations and to check whether this chosen representation is matched.

The experiment of utilizing individual sensor (or stimulus) for color matching has shown that the matching error is expectedly large, as shown in Table III. It is because the color space of each single stimulus in RGB have 256 values. However, the color space of CMYK color representation has 12096 $(= 12 \times 12 \times 12 \times 7)$ values, where Cyan (C), Magenta (M), Yellow (Y) have twelve values, and Black (K)

Table III. Matching errors from individual stimulus

Stimulus	Matching Error
R	15.6%
G	15.6%
B	15.6%

Table IV. Matching errors from fusion techniques

Fusion technique	Matching error Independent	Matching error Dependent
Bayesian	1.26%	1.24%
EMC	1.26%	1.26%

has seven values. In other words, each individual stimulus in RGB representation may match up to 47 CMYK representations on the average.

4.2.2. *Results from Fusion Techniques*

Table IV shows the matching errors of selecting only one representation at a time. Again EMC and Bayesian techniques are comparable in terms of matching errors. Note that there is not much improvement when dependent relations are considered.

4.2.3. *Results with Noisy Data*

Fifteen levels of noise were introduced into the RGB input data. Figure 10 shows the results for the two fusion techniques under the independent assumption. Note that the performance of the Bayesian fusion technique (solid line) is very sensitive to noise as compared to the EMC (dashed line) technique. The matching errors from the Bayesian approach are over 95% whereas those from the EMC approach are between 17 and 21%. This is due to the multiplicative nature in computing the multivariate posterior probability which is given by

$$p(C, M, Y, K|R, G, B) = p(C, M, Y, K|R)\,p(C, M, Y, K|G)\,p(C, M, Y, K|B).$$

A small change in one of the univariate posterior probabilities will amplify the overall multivariate probability. On the other hand, the proposed EMC fusion technique is more robust. The performance is not affected by noise as much as the Bayesian approach. The EMC fusion technique computes the multivariate posterior probability based on weighted sum of other posterior probabilities, instead of

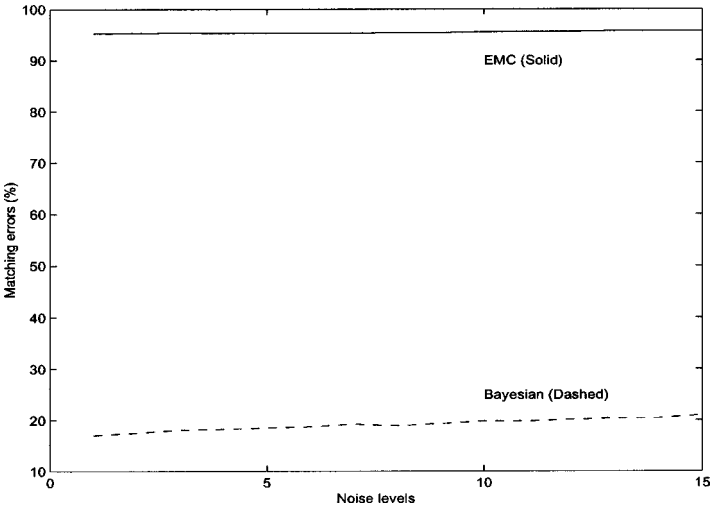


Figure 10. Matching errors with noisy input data (independent assumption).

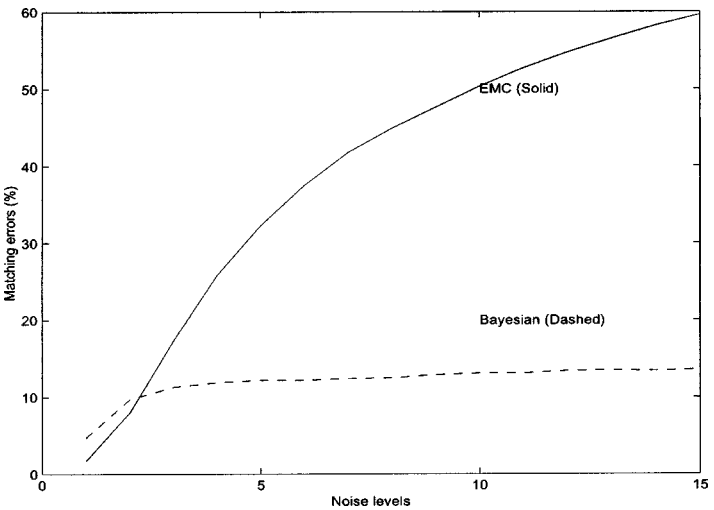


Figure 11. Matching errors with noisy input data (dependent assumption).

weighted product, i.e.,

$$p(C, M, Y, K|R, G, B) = \kappa_R p(C, M, Y, K|R) + \kappa_G p(C, M, Y, K|G) + \kappa_B p(C, M, Y, K|B).$$

Under the dependent assumption, the Bayesian approach is based on the likelihood function. And the EMC fusion technique is based on Equation (16). It is shown from Figure 11, the proposed EMC fusion technique (dashed line) is more robust than the Bayesian approach (solid line). The range of errors for the EMC approach is between 4.6 and 13.6% whereas that for the Bayesian approach is between 1.7 and 59.6%.

Table V. Computational requirements

	Bayesian	EMC
Size of distribution table	$\|\Theta\| \cdot \|X\|^m$	$m\|\Theta\| \cdot \|X\| + C_2^m \|\Theta\| \cdot \|X\|^2$
Time required to compute probabilities	$O(\|\Theta\| + \|X\|^m)$	$O(\|\Theta\| + \|X\|) + O(\|\Theta\| + \|X\|^2)$
Time required to compute entropies	Not necessary	$O(m^2)O(\|\Theta\|^2 + \|\Theta\|\|X\|^2)$

4.3. SUMMARY OF THE RESULTS

Entropy based Markov chain fusion is established to incorporate the uncertainties of the univariate observation and bivariate observations into the Markov chains such that the interdependence between sensors can be reflected effectively during the data combination process. The performance of the approach, when compared with the individual sensors in terms of error, demonstrates its strength in improving

(a) the measurement accuracy for a group of positively dependent sensors,
(b) the classification errors for the color mapping and object identification, and
(c) the robustness.

It provides strong evidence to the generalization of the technique to a pool of m sensors. It also shows that the aggregation of positively dependent sensors is constructive.

An important advantage of the technique is its simplicity in terms of data structure and computation. No greater than second order of the joint observation distribution is necessary for m sensors by which it can greatly simplify the data structure to represent the interrelations among sensors, and accelerate the computation of sensory weights. Table V gives the summary of the theoretical computational requirements of the two approaches. Since the number of sensors used in our experiments is three, there is no significant difference between the two.

5. Conclusion

This paper proposes and demonstrates the entropy based Markov chain (EMC) fusion technique to aggregate multiple sensory observations into a consensus output. The significant impact of including dependent information in sensory observation combination process has been shown. As illustrated by the experimental results, the addition of dependent relationships is useful in the sense that EMC with dependence can remarkably improve the measurement accuracy, when compared with individual sensors. The major benefits of the approach are as follows:

(a) single observation distributions and joint observation distributions between any two sensors, which represent the interrelations between the sensory observations and the consensus output, are represented in polynomial form,
(b) the consensus output is the linear combination of the weighted observations, in which weights can be computed in polynomial time and
(c) EMC is robust because it suppresses the noisy observation with high uncertainty level to minimize the contribution of the unreliable and noisy observation in the combination process. The disadvantage of this approach is that, owing to the limited order of likelihood functions, dependence information may not be 'fully' represented as in the Bayesian fusion, 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, EMC is still attractive to implement.

Acknowledgement

The research is supported by the RGC Competitive Earmarked Research Grants, no. HKUST 618/94E and 754/96E, Hong Kong.

References

1. Basir, O. A. and Shen, H. C.: Modelling and fusing uncertain multi-sensory data, *J. Robotic Systems* **13**(2) (1996), 95–109.
2. Berger, J. O.: The robust Bayesian viewpoint, *The International Library of Critical Writings in Econometrics 7: Bayesian Inference*, Vol. 1, 1995, pp. 354–414.
3. Berler, A. and Shimony, S. E.: Bayes networks for sonar sensor fusion, in: *Proc. of the 13th Conf. on Uncertainties in Artificial Intelligence*, 1997, pp. 14–21.
4. Bloch, I. and Maitre, H.: Fusion of image information under imprecision, in: *Aggregation and Fusion of Imperfect Information*, 1998, pp. 189–213.
5. Buxton, H. and Gong, S. G.: Visual surveillance in a dynamic and uncertain world, *Artificial Intelligence* **78** (1995), 431–459.
6. Chung, A. C. S., Shen, H. C., and Basir, O. A.: A decentralized approach to sensory data integration, in: *Proc. of 1997 IEEE/RSJ Internat. Conf. on Intelligent Robots and Systems*, 1997, pp. 1409–1414.
7. Chung, A. C. S. and Shen, H. C.: Integrating dependent sensory data, in: *Proc. of 1998 IEEE Internat. Conf. on Robotics and Automation*, May 1998, pp. 3546–3552.
8. Clark, J. J. and Yuille, A. L.: *Data Fusion for Sensory Information Processing Systems*, Kluwer Academic, Dordrecht, 1990.
9. DeGroot, M. H.: Reaching a consensus, *J. Amer. Statist. Assoc.* **69**(345) (1974), 118–121.
10. DeGroot, M. H.: *Probability and Statistics*, McGraw-Hill, New York, 1986.
11. Dromigny, A. and Zhu, Y. M.: Improving the dynamic range of real-time X-ray imaging systems via Bayesian fusion, *J. Nondestructive Evaluation* **16**(3) (1997), 147–160.
12. Durrant-Whyte, H. F.: Sensor models and multisensor integration, *Internat. J. Robotics Res.* **7**(6) (1988), 97-113.
13. Elfes, A.: Multi-source spatial data fusion using Bayesian reasoning integration, in: *Data Fusion in Robotics and Machine Intelligence*, Academic Press, New York, 1992, pp. 137–163.

14. Goodridge, S. G., Kay, M. G., and Luo, R. C.: Multilayered fuzzy behavior fusion for real-time reactive control of systems with multiple sensors, *IEEE Trans. Industr. Electronics* **43**(3) (1996), 387–394.
15. Haddad, Z. S., Durden, S. L., and Im, E.: Bayesian fusion of trmm passive and active measurements, in: *IGARSS'97, 1997 Internat. Geoscience and Remote Sensing Symposium*, Remote Sensing – A Scientific Vision for Sustainable Development, Vol. 4, 1997, pp. 1642–1644.
16. Hall, D.: *Mathematical Techniques in Multisensor Data Fusion*, Artech Hourse, Boston, 1992.
17. Hillier, F. S. and Lieberman, G. J.: *Introduction to Operations Research*, McGraw-Hill, New York, 1990.
18. Huber, P. J.: *Robust Statistics*, Wiley, New York, 1981.
19. Hurn, M. A., Mardia, K. V., Hainsworth, T. J., Kirkbride, J., and Berry, E.: Bayesian fused classification of medical images, *IEEE Trans. Medical Imaging* **15**(6) (1996), 850–858.
20. Kang, D. H., Luo, R. C., Hashimoto, H., and Harashima, F.: Position estimation for mobile robot using sensor fusion, in: *Proc. of the 1994 Internat. Conf. on Multisensor Fusion and Integration for Intelligent Systems (MFI'94)*, October 1994, pp. 647–652.
21. Kittler, J., Matas, J., Jonsson, K., and Sanchez, M. U. R.: Combining evidence in personal identity verification systems, *Pattern Recognition Lett.* **18** (1997), 845–852.
22. Liu, J. and Chang, K. C.: Feature-based target recognition with a Bayesian network, *Optical Engineering* **35**(3) (1996), 701–707.
23. Nadabar, S. G. and Jain, A. K.: Fusion of range and intensity images on a connection machine (cm-2), *Pattern Recognition* **28**(1) (1995), 11–26.
24. Park, S. W. and Lee, C. S. G.: Fusion based sensor fault detection, in: *Proc. of 1993 IEEE Internat. Sympos. on Intelligent Control*, August 1993, pp. 156–161.
25. Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P.: *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge Univ. Press, 1992.
26. Shannon, C. E.: A mathematical theory of communication, *Bell Syst. Techn. J.* **27** (1948), 379–423.
27. Solberg, A. H. S., Jain, A. K., and Taxt, T.: Multisource classification of remotely sensed data: Fusion of landsat tm and sar images, *IEEE Trans. Geoscience Remote Sensing* **32**(4) (1994), 768–778.
28. Stassopoulou, A., Petrou, M., and Kittler, J.: Application of a bayesian network in a gis-based decision making system, *Internat. J. Geographical Inform. Sci.* **12**(1) (1998), 23–45.
29. Stratton, D. A. and Stengel, R.: Real-time decision aiding: Aircraft guidance for wing shear avoidance, *IEEE Trans. Aerospace Electronic Systems* **31**(1) (1995), 117–124.
30. Wallack, A. S. and Canny, J. F.: Generalized polyhedral object recognition and localization using crossbeam sensing, *The Internat. J. Robotics Res.* **16**(4) (1997), 473–496.
31. Wolff, G. J.: Sensory fusion: Integrating visual and auditory information for recognizing speech, in: *Proc. of the 1993 IEEE Internat. Conf. on Neural Networks*, Vol. 2, 1993, pp. 672–677.