# Optimal Decision Fusion For Sensor Network Applications

Marco F. Duarte *IEEE Member*, and Yu Hen Hu *IEEE Fellow*

**Abstract**

Energy conservation is a major concern of large scale sensor network systems when deployed for long-term monitoring purposes. In such a wireless sensor network, information obtained by individual sensors must be aggregated via wireless communication channels, and hence will consume a considerable amount of energy. This problem can be mitigated using "data fusion" where data sampled at individual sensors is processed locally to calculate a local statistic. These locally processed results then will be transmitted to a "fusion center" via wireless channel to deduce a global statistic. There are two types of data fusion modes: decision fusion where a decision from small discrete set of possible decisions is transmitted from each sensor node; and value fusion where a real number is transmitted from each sensor node. Numerous decision fusion methods have been proposed in the past. Yet, few have claimed optimality in terms of maximizing the probability of correct decision of the fused decision. In this paper, we present a simple, yet practical proof to show how to derive an optimal decision rule for a family of decision fusion problems where sufficient amount of training data samples are available. This proof is general in the sense that it is not tied to any specific existing decision fusion algorithms. Yet, it can be used to access the optimality of any existing decision fusion algorithm for particular set of problems.

**Index Terms**

decision fusion, data fusion, information fusion, sensor network, data aggregation, Bayesian inference, classifier fusion, stack generalization, mixture of experts, classifier ensemble

## I. Introduction

The emergence of small, low-power devices that integrate micro-sensing and actuation with on-board processing and wireless communication capabilities stimulates great interests in wireless distributed sensor networks [1], [2], [3]. The sensor networks often perform tasks such as detection, classification, localization and tracking of one or more targets in the sensor field. The sensors are typically battery-powered and have limited wireless communication bandwidth. Therefore, efficient collaborative signal processing algorithms that consume less energy for computation and communication are needed for these applications [4].

In this paper, an important collaborative signal-processing task, data fusion of sensor signals, will be discussed. In literature, there are two types of data fusion methods: *decision fusion* versus *value fusion*. The value fusion combines *continuous*, *real-valued* samples or estimates reported from individual sensors, while decision fusion combines a *finite, countable set of integer-valued* decisions from sensors. In reality, real numbers have to be quantized into discrete values before they can be transmitted digitally. Hence the distinct between these two fusion methods may be somewhat blurred. In this paper, however, we adopt the narrower definition the term *decision fusion* to refer to detection or classification applications. In this family of applications, each sensor will make a local decision among often a handful of possible choices. As such, *the total number of different combinations of decisions made by all sensors will be enumerable*.

In [5], optimal data fusion is presented under the constraint of a fixed $k$ out of $n$ weighted threshold fusion architecture. In [6], a hierarchical model is used and Bayesian Gibbs sampling method is used to design the fusion rule.

Data fusion has also been studied in the context of combining multiple classifiers. In [7], three types of classifier combination methods, namely, *averaged Bayes classifiers*, *voting principals*, and *Dempster-Shafer fuzzy combinations* have been reviewed. Some experiments have been conducted but no conclusive comparison results are available. In [8], the accuracy of individual classifiers are estimated, and classifiers are selected dynamically

The authors are with the University of Wisconsin - Madison, Department of Electrical and Computer Engineering, Madison, WI 53706.
This project is supported by DARPA under grant no. F 30602-00-2-0555

based on which classifier will yield best performance in specific local region. Ji and Ma [9] proposed to use a structure consisting of randomly generated linear local classifiers with a voting fusion mechanism to perform pattern classification tasks. Petrakos et al. [10] discussed the effect of correlations between classifiers and their impacts on fusion performance.

The paper is organized as follows. Section II presents the theoretical framework for this problem. Section III specifies our approach for Optimal Decision Fusion. Section IV details the application of the Optimal Decision Fusion Algorithm in the Collaborative Signal Processing Sensor Networks area. Section V contains the experiments we performed with our method and their results. Finally, section VI presents some conclusions to the paper.

## II. Problem Formulation

### A. Decision Fusion Framework

We assume a data fusion architecture that consists of a *fusion center* and $K$ distributed sensors. These sensors observe a common feature vector $x$. Then, each of these sensors will make a decision that maps $x$ to a set of $N$ class labels $\mathbf{C} = \{C_1, C_2, \cdots, C_N\}$. We may conveniently map them into $N$ integers such that $C_n = n$ for $n = 1, 2, \cdots, N$. Later, we use the set membership notation $x \in C_n$ to denote that the label of a feature vector $x$ is $C_n$.

Each sensor will evaluate the feature vector $x$ and make a decision $d(x)$ among the different classes. The sensors may use the same type of decision rules or different types of decision rules. We only assume that the decision rule running on a sensor remains the same during training and testing phases.

Depending on the specific signal modality, and sensor locations, each sensor may only use a portion of $x$ to reach its decision. For example, an acoustic sensor is capable of utilizing only acoustic signal to make a decision, while a seismic sensor can only use seismic data. Two sensors located at different locations may receive different acoustic vectors due to different distances from the (unknown) target location, or additive noise.

The feature vector $x$ can be regarded as a concatenated vector at the fusion center that combines all the feature vectors used by every individual sensors. With this notation, individual sensors then can make their own local decisions (classifications) based only on a subset of elements in this common feature vector $x$. Thus, one may assume the local decision made by individual sensors are based on the common feature vector $x$. The fact that individual sensors use only a portion of this feature vector can be regarded as a variation of the decision rules implemented on individual sensors.

We further assume that each of the $K$ sensors will forward its local decision $d_k \in \mathbf{C}$ to the fusion center. Thus, at the fusion center, it will receive a $K \times 1$ decision vector $\mathbf{d} = \begin{bmatrix} d_1 & d_2 & \cdots & d_K \end{bmatrix}$. With these notations, we may define *value fusion* and *decision fusion* as follows:

*Value fusion*: The fusion center uses the combined feature vector $x$ to reach a decision.

*Decision fusion*: The fusion center uses the decision vector $\mathbf{d}$ to reach a decision.

In this paper, we will focus on the case of decision fusion. That is, we want to develop an optimal decision fusion rule. The criterion of optimality will be defined after we review the statistical pattern classification theory.

### B. Statistical Pattern Classification: Decision Rules

In a statistical pattern classification framework, the feature vector $x$ is assumed to be a sample drawn from a probability distribution (feature space) with probability density function *p.d.f.* $p(x)$. Due to some natural assignment process that is beyond our control, each of the sample $x$ is assigned to a unique class label $C_n$. The probability that $x$ belongs to a specific class $C_n$, denoted by $P(x \in C_n) = P(C_n)$, is known as the *a priori probability* [1] . The *likelihood* that a sample will assume a specific feature value $x$ *given* that this is drawn from a particular class $C_n$ is denoted by the conditional probability $p(x|C_n) = p(x|x \in C_n)$. Using Bayes rule, the probability that a particular sample belongs to class $C_n$, *given* that the sample assumes the value of $x$, is denoted by the *a posteriori* probability

$$p(x \in C_n|x) = p(C_n|x) = \frac{p(x|C_n)P(C_n)}{p(x)} \tag{1}$$

[1]As a convention, we use lower case *p* to denote *p.d.f.* while using upper case *P* to represent probability.

In the rest of this paper, we will use a *decision rule* and a *pattern classifier* interchangeably. As we said earlier, a decision rule is a mapping from the feature vector $x$ to a an element in the class label set **C**. That is, $d(x) \in \mathbf{C}$. If $x \in C_n$ and $d(x) = C_n$, then a correct classification (decision) is made. Otherwise, we say it is a misclassification (incorrect decision).

The statistical *maximum a posteriori* (MAP) classifier stipulates that in order to maximize the probability of a correct classification (and hence minimize the probability of misclassification), the decision rule must choose the class label among all $N$ classes that yields maximum a posteriori probability. In other words,

$$d(x) = \arg\max_n P(C_n|x) \tag{2}$$

For each feature vector $x$, the probability of correct classification is

$$P_c = \int_x \sum_{n=1}^{N} P(d(x) = C_n|x \in C_n)P(C_n)p(x)dx \tag{3}$$

In most practical applications, a set of *training* data samples are available. It is assumed that this training set consists of independently drawn samples from the feature space. In this training set, the class label of the $i^{th}$ training feature vector $x(i)$ is specified. In such a situation, the probability of classifying a sample $x \in C_n$ is into a class $C_m(\neq C_n)$ can be denoted as $P(x, C_m|C_n) = P(d(x) = C_m|x \in C_n)P(C_n)$. Since there are $N \times N$ such entries, they can be arranged into a *confusion matrix* $\mathbf{M}(x)$ such that the $(n, m)$ entry of $\mathbf{M}(x)$ is $P_e(x, C_m|C_n)$. Averaging over the entire training set, the performance of a particular classifier $d$ can be characterized with the corresponding confusion matrix **M**:

$$\mathbf{M} = \sum_i \mathbf{M}(x(i)) \tag{4}$$

where the summation is taken over the indices $i$ represents the entire training data set.

## III. Optimal Decision Fusion

Given the above notations and definitions, we now turn our attention to the optimal decision fusion method. As defined earlier, the decision fusion is itself a decision rule $\ell(\mathbf{d}(x)) \in \mathbf{C}$ that maps a feature vector $\mathbf{d}(x) = \begin{bmatrix} d_1 & d_2 & \cdots & d_K \end{bmatrix}$ to one of the class label. A critical observation is that the sample space of the decision fusion classifier is finite and countable. In particular, since each decision $d_k$ takes only $N$ possible values, the combined vector $\mathbf{d}(x)$ can have at most $N^K$ different combinations. For example, in a region-detection problem, each sensor in the same geographical region reports whether it detects a target based on its own sensor reading. Hence $N = 2$. If there are 3 sensors within the same region all reporting to the same fusion center, then there are at most $2^3 = 8$ different combinations of their outputs

For each feature vector $x$ in the training data set (or equivalently, the feature space), the set of $K$ sensors will yield a unique decision vector $\mathbf{d}(x)$. Therefore, the set of $N^K$ decision vectors obtained using the set of $K$ decision rules $\{d_k; 1 \leq k \leq K\}$ uniquely partition the feature space into $N^K$ disjoint regions, denoted by $\{r_m; 1 \leq m \leq N^K\}$. As such, the probability of correct classification of the fusion classifier is the sum the probability of correct classification over each region $r_m$. If a classifier maximizes its probabiity of correct classification in every $r_m$, it maximizes the overall probability of correct classification. Hence it suffices to focus on developing an optimal classifier for each of the individual region $r_m$.

Denote the unique decision vector that every $x \in r_m$ maps to as $\mathbf{d}(m)$. The MAP principle stipulates that

$$\ell(\mathbf{d}(m)) = C_{n^*} \text{ if } P(C_{n^*}|\mathbf{d}(m)) > P(C_n|\mathbf{d}(m))$$

for $n \neq n^*$. Using Bayes rule, if $P(x \in r_m) \neq 0$, then

$$\begin{aligned} P(C_n|\mathbf{d}(m)) &= \frac{P(\mathbf{d} = \mathbf{d}(m)|x \in C_n) \cdot P(x \in C_n)}{P(x \in r_m)} \\ &= \frac{P(\mathbf{d} = \mathbf{d}(m); x \in C_n)}{P(x \in r_m)} \end{aligned} \tag{5}$$

If the feature space consists of a countable set of points, the last expression can be replaced by

$$P(C_n|\mathbf{d}(m)) = \frac{|\{x|x \in r_m \cap C_n\}|}{|\{x|x \in r_m\}|} \tag{6}$$

where $|A|$ is the cardinal number of a set $A$. Clearly that the MAP classification label $C_{n^*}$ for $r_m$ can be defined as:

$$n^* = \begin{cases} \arg\max_n \int_{x \in r_m \cap C_n} p(x)dx & \text{continuous feature space;} \\ \arg\max_n |\{x|x \in r_m \cap C_n\}| & \text{discrete feature space.} \end{cases} \tag{7}$$

Clearly, in a discrete feature space, the class label of $\mathbf{d}(m)$ should be assigned according to a majority vote of class labels among all $x \in r_m$.

### *A. Performance Analysis*

The proposed optimal decision fusion (ODF) classifier has several distinct and significant properties:

1) The ODF is the optimal MAP classifier in the sense that it minimizes the probability of misclassification among all decision fusion classifier structures.
2) Unlike existing decision fusion methods, the ODF does *NOT* assume *independence* among local classifiers. In this sense, ODF is more general than all the existing works on decision fusion classifiers.
3) ODF is a non-parametric classifier in the sense that it does not impose a parametric statistical model on the decision vectors.

First, we establish the optimality of the proposed classifier.

*Theorem 1:* Given a fixed set of $K$ local $N$-class classifiers, the decision fusion classifier designed according to equation (7) is optimal in the sense that it yields minimum probability of mis-classification among all fusion classifiers that use $\{\mathbf{d}(m); 1 \leq m \leq N^K\}$ as feature vectors.

*Proof:* For each given feature vector $\mathbf{d}(m)$, the proposed decision fusion classifier assign a class label that maximizes the a posteriori probability. Since the entire set of feature vectors uniquely partition the original feature space into disjoint sets, maximizing the a posteriori probability of individual region will guarantee the a posteriori probability over the entire set of decision vectors is maximized.

Although the decision fusion classifier presented here is the optimal classifier among all classifiers that use decision vectors as feature vectors, it is *not* necessarily the optimal classifier of the original classification problem in the Bayesian sense.

*Theorem 2:* Denote $P_d$ to be the probability of mis-classification of a decision fusion classifier as described in theorem 1, and $P_e$ to be the probability of mis-classification of a Bayesian classifier that uses the original feature vector $x$. Then

$$P_d \geq P_e \tag{8}$$

*Proof:* The probability of correct classification of a Bayesian classifier is:

$$\begin{aligned} 1 - P_e &= \sum_{n=1}^{N} \int P(C_n)P(d(x) = C_n|x \in C_n)p(x)dx \\ &= \sum_{n=1}^{N} \int P(d(x) = C_n; x \in C_n)p(x)dx \\ &= \sum_{n=1}^{N} \int \sum_{m=1}^{M} P(d(x) = C_n; x \in C_n \cap r_m)p(x)dx \\ &= \sum_{m=1}^{M} \int \sum_{n=1}^{N} P(d(x) = C_n; x \in C_n \cap r_m)p(x)dx \\ &\geq \sum_{m=1}^{M} \int P(d(x) = C_{n^*}; x \in C_{n^*} \cap r_m)p(x)dx \\ &= 1 - P_d \end{aligned} \tag{9}$$

The last inequality follows from the decomposition of

$$\sum_{n=1}^{N} P(d(x) = C_n; x \in C_n \cap r_m) = \sum_{n=1, n \neq n^*}^{N} P(d(x) = C_n; x \in C_n \cap r_m) + P(d(x) = C_{n^*}; x \in C_{n^*} \cap r_m) \quad (10)$$

Intuitively, within each region $r_m$, the original Bayesian classifier allows each $x$ to be labeled (classified) into a class $n$ such that

$$P(C_n|x) > P(C_{n'}|x) \quad \text{for} \quad n' \neq n \quad (11)$$

But with decision fusion, **all** $x$ within $r_m$ must be assigned to a single label. Thus, this leads to the potential performance degradation.

*1) Comparison with Other Fusion Methods:* Let us use an example to illustrate the difference between the proposed ODF method and the voting method. Consider 3 local classifiers ($K = 3$) binary decision ($K = 2$) example. Consider a particular decision vector $\mathbf{d}(m) = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$. Based on majority voting fusion method, this will lead to a fused decision of 0. However, suppose the corresponding $r_m$ consists of 3 training samples with correct class labels 1, 0, 1 respectively. According to ODF, the class label of the decision vector 001 will be assigned to 1. As such, the ODF will out-perform the majority voting scheme within this particular region. Note that in this example, two local classifiers made two mistakes (the first and the third samples), and the third one made one mistake. However, minority opinion prevails with ODF.

## IV. Applications of Decision Fusion to Sensor Network Collaborative Signal Processing

In November 2001, a field experiment sponsored by the DARPA ITO SensIT project has been carried out at 29 Palms California, USA. Custom-made prototype sensor nodes are laid out along side a road. Each sensor node is equipped with acoustic, seismic, and polarized infrared sensors, a 16-bit micro-processor, and radio transceiver and modem. It is powered by external battery. During experiment, military vehicles such as AAV (amphibian assault vehicle), DW (dragon wagon) will pass through the road, and sensors will sample the corresponding multi-modality signature. The acoustic signal is sampled at 4960 Hz at 16-bit resolution. The set of data segments used in the experiment reported in the next section are taken from the acoustic signatures of a single AAV as well as a single dragon wagon vehicle travelling from east to west along the road during a time period of approximately 2 minutes each.

A collaborative signal processing algorithm suite has been developed in the UW-Madison SensIT research team. The sensor network signal processing tasks include detection, localization, classification, and tracking. Sensors in the sensor field are group into regions according to their geographic proximity. Individual sensors report to a region manager where data fusion is performed. In this system, each individual sensor node will perform its own target detection and classification task using its own data samples. The detection is basically the classical constant false alarm rate (CFAR) [11] [12] detection algorithm. At a cycle of every 0.75 second, each sensor node that is activated will compute an energy reading from its three sensor channels. This new energy reading will be compared to a mean and variance of previous energy readings to determine if it is an outlier. If so, this energy reading may signal the presence of a target. Otherwise, this energy reading is assumed to be generated from the background noise process, and hence will be used to update the mean and variance of the energy readings. Upon a positive detection decision is made at a node, it will load a classification module to classify the type of vehicles. The details of detection and classification results will not be elaborated here for the sake of brevity. The detection and classification results as well as the mean and variance of the background noise energy sequence then will be transmitted to the region manager node. An important observation here is that using energy reading instead of raw waveform, only a few bytes will need to be transmitted for every 0.75 seconds. This represents a tremendous reduction of communication load via the wireless channel.

At the region manager node, a region-wide detection fusion and classification fusion will be performed based on results reported from individual sensor nodes within the region. The purpose of these fusion operation is to improve the quality of decisions. If region-wide target detection is confirmed, the energy based source localization algorithm then will be applied to locate the vehicle locations within the sensor field. The estimate vehicle locations then will be passed to a Kalman-filter based tracking algorithm to predict the new position of the vehicle and to facilitate other types of queries.

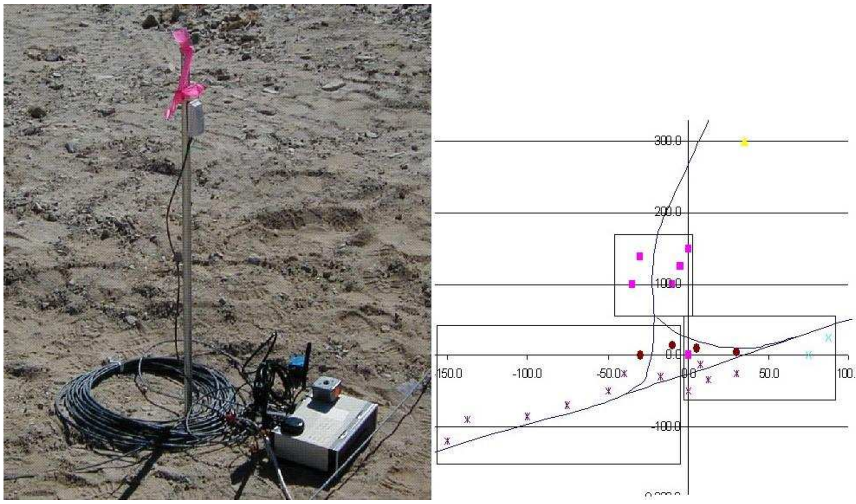


Fig. 1. (a) A Sensoria sensor node; (b) sensor layout in the sensor field

## V. Experiments

We run experiments on data collected from a sensor network. We assume that a number of sensor nodes are deployed in an outdoor sensor field. Each sensor node consists of an on-board computer, power source (battery), one or more sensors with different modalities, and wireless transceivers. Depicted in Figure 1(a) is a prototype sensor node used in the DARPA SensIT project, manufactured by Sensoria, Inc. With this sensor node, there are three sensing modalities: acoustic (microphone), seismic (geophone), and infrared (polarized IR sensor). The acoustic signal is sampled at 5 kHz at 12 bit resolution. The on-board computer is a 32-bit RISC processor running the Linux operating system.

The sensor field (c.f. Figure 1(b)) is an area of approximately $900 \times 300$ meters in a California Marine training ground. The sensors, denoted by dots of different colors in Figure 1(b) are layout along side the road. The separation of adjacent sensors ranges from 20-40 meters. We partition the sensors into three geographically local *regions*. Sensors within each region will be able to communicate freely. One sensor within each region is designated as a *manager node*. The manager node will be given the authority to communicate with manager nodes of surrounding regions. This hierarchy of communication ensures that only local wireless traffic will be engaged, and hence contributes to the goal of energy conservation.

Military vehicles, including the *Assault Amphibian Vehicle* (AAV), the *Dragon Wagon* (DW), the *High Mobility Multipurpose Wheeled Vehicle* (HMMWV), and others are driving passing through the roads. The objective is to detect the vehicles when they pass through each region. The type of the passing vehicle then will be identified, and the accurate location of that vehicle will be estimated using an *energy-based localization algorithm*. In the following discussion, we will assume there is at most one vehicle in each region. During the experimentation in November 2001, multi-gigabyte data samples have been recorded and are used in this paper. We will call these data Sitex02 data set.

For each of the 0.75 second duration, the energy of the acoustic signal will be computed. This single energy reading then will be fed into a *constant false alarm rate* (CFAR) energy detector [4] to determine whether the current energy reading has a magnitude that exceeds a computed threshold. If so, a *node-detection* event will be declared for this duration. Otherwise, the energy reading is considered as contributions from the background noise.

Once a positive target-detection decision has been made, a pattern classifier using Maximum likelihood pattern classifier [4] is invoked. The acoustic signal is recording using a sampling frequency of 4960 Hz. We use a 50 dimensional feature vector based on the Fourier power spectrum of the corresponding acoustic time series within the 0.75-second duration. This feature is created by averaging by pairs the first 100 points of the 512-point FFT, which are then normalized; the resolution of the frequency spectrum sampling is 19.375 Hz due to the averaging.

From the classification decisions and the detection results from individual nodes, we will create a decision vector that will group the individual decisions based on the sensor to vehicle distance. This is needed because the number of nodes that participates in the fusion process varies with time, as nodes that do not show detection are excluded

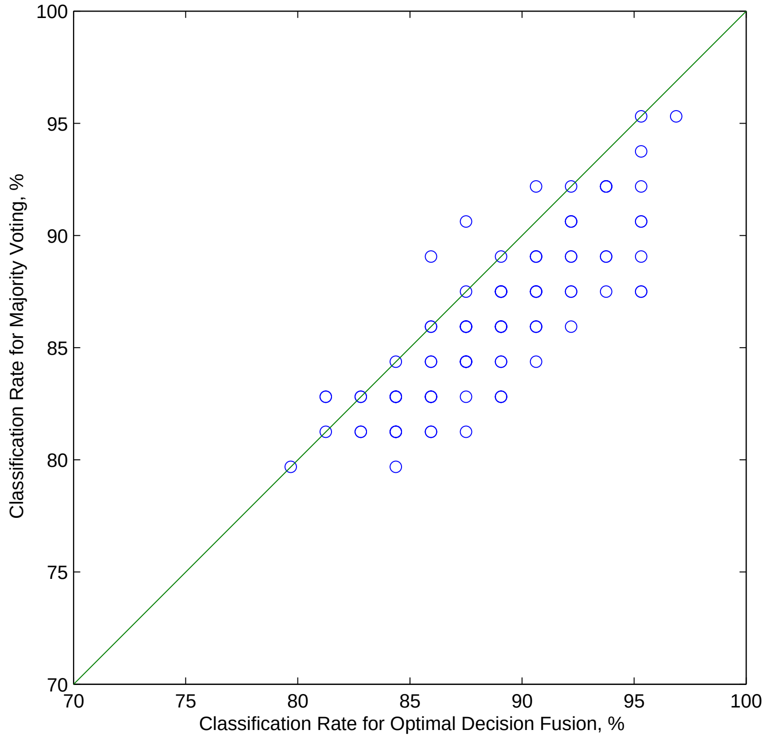


Fig. 2. Simulation results comparing ODF with majority voting method

from the fusion. We group the decisions of the nodes based on their locality: nodes within a certain distance interval from the vehicle are grouped and majority voting is performed among these grouped decisions to achieve a single decision result: this will be the decision of what we call a local pseudoclassifier. In this way, we obtain a decision vector that will contain the pseudoclassifier's decisions and we proceed to use our proposed method on these vectors. For this experiment, we used five local pseudoclassifiers with a range of 20 m for each one; thus, only nodes that are closer than 100 m to the vehicle are taken into account in the decision fusion at any given time. We compare the results of our proposed method with that of majority voting, as it is a widely accepted baseline method.

We combined the AAV and DW data and randomly partition the data set into a training data set consisting of roughly 67% of data samples and a testing data set of the remaining samples. Then we use the proposed ODF method to develop a classifier and use that classifier to classify the testing data set. Meanwhile, we developed a majority vote fusion classifier and test it on the same testing data set. This experiment is repeated 100 times. Each time the classification rate of both the ODF and the voting methods are recorded. The classification rates of the voting versus the ODF methods are depicted in figure 2. It can be seen that only 4 out of 100 trials the majority vote method exceeds the ODF method. These situation occurs due to finite size of the training data set that can not accurately capture the characteristics of certain regions $r_m$.

## VI. Conclusions and Further Work

In this paper we have shown that using a simple algorithm, the optimal decision fusion method can be defined even without explicitly knowing the classification rates of the different sensors, or for dependent sensors in a region. We also have shown experimental results that holds our argument. It is important, however, to have enough features available during training so that the statistics for each one of the different regions based on the decision vector results are statistically defined.

In the future, we will report results of application of this method in other real-world problems, such as collaborative signal processing and handwritten character recognition. We will also report on any patterns observed for committees of multiple members.

## REFERENCES


[1] D. Estrin, Culler, D., Pister, K., and Sukhatme, G.,: Connecting the Physical World with Pervasive Networks, IEEE Pervasive Computing, vol. 1, Issue 1, 2002, pp. 59-69.

[2] D. Estrin, Girod, L., Pottie, G., and Srivastava, M.,: Instrumenting the world with wireless sensor networks, Proc ICASSP '2001, vol. 4, 2001, pp. 2033 -2036

[3] C. Savarese, Rabaey, J. M., and Reutel, J.: Localization in distributed Ad-hoc wireless sensor networks, Proc. ICASSP'2001, Salt Lake City, UT, (2001), 2037-2040

[4] D. Li, Wong, K.D., Hu, Y. H., Sayeed, A. M.,: Detection, classification, and tracking of targets. IEEE Signal Processing Magazine, **19**, (2002), 17–29

[5] Z. Chair, and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," IEEE Trans. AES, Vol. 22, no. 1, Jan. 1986, pp. 98-101.

[6] B. Chen, and P. K. Varshney, "A Bayesian sampling approach to decision fusion using hierarchical models", IEEE Trans. on Signal Processing, vo. 50, no. 8, Aug. 2002, pp. 1809-1818.

[7] L. Xu, A. Krzyzak, and C. Y. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition", IEEE Trans. Systems, Man, and Cybernetics, vol. 22, no. 3, 1992, pp. 418-435.

[8] K. Woods, W. P. Kegelmeyer Jr., and K. Bowyer, "Combination of multiple classifier using local accuracy estimates", IEEE Trans. Pattern analysis and machine intelligence, vol. 19, no. 4, april 1997, pp. 405-410.

[9] C. Ji, and S. Ma, "Combinations of weak classifiers", IEEE Trans. Neural Networks, vol.8 no. 1, 1997, pp. 32-42.

[10] M. Petrakos, J. A. Benediktsson, and I. Kanellopoulos, "The effect of classifier agreement on the accuracy of the combined classifier in decision level fusion," IEEE Trans. on Geoscience and Remote Sensing, vol. 39, no. 11, Nov. 2001, pp. 2539-2546.

[11] R. E. Schwartz,: Minimax CFAR detection in additive Guassian noise of unknown covariance, IEEE Transactions on Information Theory, vol. it-15, pp. 722-5, 1969.

[12] J. Dabkowski, Jr.,: CFAR operation in non-homogeneous and point clutter, Proceedings of the national electronics conference., pp. 611-615, 1970.