# On Sample-Based Implementation of Non-Smooth Decision Fusion Functions

Nageswara S. V. Rao
Center for Engineering Science Advanced Research
Computer Science and Mathematics Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831-6355
*Proceedings of IEEE/SICE/RSJ International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2001.

**Abstract**

*A number of optimal fusion functions have been derived in the literature for multiple detection systems based on a complete knowledge of the detector distributions. In several practical systems, however, only measurements are available. A general result was recently shown that any fusion function with a suitable Lipschitz property derived under the complete knowledge of the distributions can be converted into a measurement-based one. While this result subsumes the well-known cases of independent and correlated detectors, it is not applicable to discontinuous fusion rules which often arise in practice. In this paper, we show that any fusion function with bounded variation can be converted into a measurement-based one with a somewhat weaker guarantee . These fusion functions subsume Lipschitz as well as several discontinuous fusion functions. In particular we show that given a sufficiently large sample, the measurement-based fusion function performs almost as well as the optimal one with an arbitrarily specified confidence.*

## 1 Introduction

The decision fusion problems deal with combining the decisions taken by the individual detectors of a multiple detector system [14]. Usually in the literature, the distributions of various detectors are assumed to be given, and an optimal fuser is derived under a certain criterion. Typically, the fusion rule is in the form of a Bayesian rule or Neyman-Pearson test, which can be derived both in the case of independent and correlated detectors. In practical applications, the information about the required probability distributions is based on the experience with the system, and it is possible to derive the distributions from the first principles. Often, the empirical data generated by the system during experimentation or operation is used in estimating the detector distributions. In practice, the optimal fuser is approximated by utilizing the estimated probabilities. Analytical justification for such approaches could be in terms of asymptotic results which show that as the sample size approaches infinity, the approximated fusion function performs as good as the optimal one [7]. Stronger results based on finite sample sizes were derived in [8], which are valid for Lipschitz continuous fusion functions and not valid for non-smooth fusion functions in general. In this paper, we extend these results to fairly general non-smooth fusion functions with somewhat weaker finite sample guarantees. In particular, we show that the usual method of estimating empirical means and using them in place of probabilities to obtain the approximation to fusion rules can result in large errors. Our method is based on a different strategy, namely, the empirical minimization [13], and yields provably better performance results compared to the usual method.

We consider a parallel suite of $N$ detectors and a fusion center [3] as in Fig. 1. Each detector $D_i$, for $i = 1, 2, \ldots, N$ makes a decision $u_i \in \{H_0, H_1\}$, and the fusion center receives $u = (u_1, u_2, \ldots, u_N)$ and outputs either $f(u) = H_0$ or $f(u) = H_1$ by suitably using the information $u$. The problem of designing the fusion functions to minize certain cost has been extensively studied [14]. Let $C_{ij}$, $i = 0, 1$, $j = 0, 1$ represent the cost of outputting $H_i$, when $H_j$ is true. Then the average-cost or Bayes risk is given by

$$C(f) = \sum_{i=0}^{1} \sum_{j=0}^{1} C_{i,j} \mathbf{P}(f(u) = H_i | H_j) \pi_j, \qquad (1.1)$$

where $\pi_i$ is the *a priori* probability of $H_i$, for $i =$

$0, 1$. The average-cost criterion is optimized by the likelihood ratio test [5] given by

$$T(u) = \frac{\mathbf{P}(u|H_1)}{\mathbf{P}(u|H_0)} > \frac{\pi_0(C_{10} - C_{00})}{\pi_1(C_{01} - C_{11})}. \qquad \text{(T.1)}$$

The decision of fusion center is $H_1$ if the above test evaluates to true and is $H_0$ otherwise. If the underlying probabilities are available in a convenient form, then $T(u)$ can be computed at given $u$. One of the most studied formulations of this problem deals with the case where $u_i$'s are independent, in which case $T(u)$ takes a simple form in terms of products [3]; $T(u)$ satisfies the Lipschitz property [8] in this case.

We consider formulation of [8] wherein the probabilities needed to evaluate the tests of the form (T.1) are unknown, but a independently and identically distributed (iid) *sample* is available in the form of $(u^1, H^1), (u^2, H^2), \ldots, (u^l, H^l)$, where $u^i \in \{H_0, H_1\}^N$ is the $i$th example and $H^i \in \{H_0, H_1\}$ is the corresponding correct hypothesis. Only a *finite sample* is given here as opposed to the formulae for the underlying probabilities required to implement Bayesian rule or Neyman-Pearson test often expressed in the form of test in Eq. (T.1). As a result only an approximate implementation of the required test is possible in general.

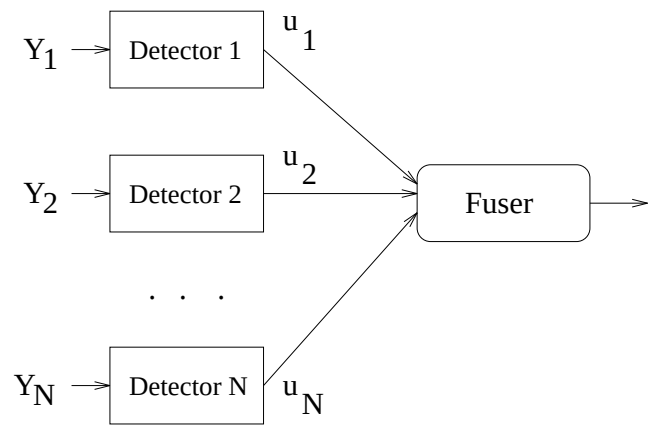
The fusion rule for decision problems is often expressed in terms of the probabilities, $p = (p_1, p_2, \ldots, p_n)$, and the data, $u = (u_1, u_2, \ldots, u_N)$, in the form

$$R(p, u) > 0, \qquad \text{(T.2)}$$

where the decision is $H_1$ if the inequality is true and is $H_0$ otherwise. For simplicity of notation, we also represent the function $R(p, .)$ simply by $R(p)$. Here $R(p, u)$ minimizes certain cost functional $C$ such that $C(R(p, .)) \leq C(f)$ for any function $f$. In the above example, the test $T(u)$ minimizes the Bayesian risk in Eq (1.1). If the underlying probabilities are known, $R(p, u)$ for given $u$ can be explicitly evaluated. If only a sample is known, then an estimator $\hat{p}$ of $p$ based on the sample $(u^1, H^1), (u^2, H^2), \ldots, (u^l, H^l)$ is employed [8]. Then, the *empirical implementation* $R(\hat{p}, u)$ is used in place of $R(p, u)$. The resultant performance of $R(\hat{p}, u)$ depends on the properties of $R$ and closeness of $\hat{p}$ to $p$. The function $R(p, u)$ is assumed to be Lipschitz in [8] with respect to $p$, i. e. these exists a positive constant $L$ such that

$$|R(p + \Delta p, u) - R(p, u)| < L||\Delta p||$$

for all $\Delta p$, $u$, where $||\Delta p||$ denotes the Euclidean norm of $\Delta p$ in $\Re^n$. Let $\bar{p}$ be the empirical mean of $p$ based




on the sample. Given a sample of size

$$l = \left\lceil \frac{r^2 n L^2}{2(R(\bar{p}, u))^2} \ln(2n/\lambda) \right\rceil$$

for any $r > 1$, $R(\bar{p}, .)$ is shown to achieve the optimal cost $C(R(p, u))$ with probability $1 - \lambda$. Fusion rules for several systems satisfy the required Lipschitz condition, including, independent detectors for the Bayesian [2] and Neyman-Pearson formulation [12], and also non-independent detectors formulated in terms of correlation coefficients [4].

In several important cases, however, the required test does not the Lipschitz condition, in which case the specific method of [8] can perform very poorly (see Section III). For example, consider a monitoring area divided into two non-overlapping regions monitored by two detectors each of which declares a detection if its a posteriori probability is above a threshold. In detecting a single object, since monitoring regions are disjoint, an appropriate fusion rule is $R(p_1, p_2) = (p_1 > t) \bigvee (p_2 > t)$, which is not Lipschitz in $p$, since it is discontinuous at $(t, t)$. More generally, the fusion function involving Boolean combinations of decision taken by individual detectors are not Lipschitz. Such functions are captured by functions of bounded variation (Section II). In this paper, we generalize the results of [8] to non-smooth fusion functions. We show that given sufficiently large sample

$$l = \frac{256 M^2}{\epsilon^2} \left[ 4n \ln\left(\frac{8eM}{\epsilon}\right) + \ln(16/\delta) \right],$$

we have

$$\mathbf{P}\left[C(R(\hat{p})) - C(R(p)) > \epsilon\right] < \delta$$

irrespective of the detector distributions. This condition means that the performance of $R(\hat{p}, u)$ is within $\epsilon$ of the optimal with probability $1-\delta$. Here $\epsilon$ and $\delta$ are called the *precision* and the *confidence*, respectively.

We describe the fusion functions of bounded variation in Section 2. We show our main result in Section 3.

## 2 Non-Smooth Fusion Functions

Consider a one-dimensional function $h : [-A, A] \mapsto \Re$. For $A < \infty$, a set of points $P = \{x_0, x_1, \ldots, x_n\}$ such that $-A = x_0 < x_1 < \ldots < x_n = A$ is called a *partition* of $[-A, A]$. The collection of all possible partitions of $[-A, A]$ is denoted by $\mathcal{P}[-A, A]$. A function $g : [-A, A] \mapsto \Re$ is of *bounded variation*, if there exists $B$ such that for any partition $P = \{x_0, x_1, \ldots, x_n\}$, we have $\sum(P) = \sum_{k=1}^{n} |f(x_k) - f(x_{k-1})| \leq B$. A multivariate function $g : [-A, A]^d \mapsto \Re$ is of bounded variation if it is so in each of its input variable for every value of the other input variables.

The following are useful facts about the functions of bounded variation: (i) not all continuous functions are of bounded variation, e.g. $g(x) = x\cos(\pi/(2x))$ for $x \neq 0$ and $g(0) = 0$; (ii) differentiable functions on compact domains are of bounded variation; and (iii) absolutely continuous functions, which include Lipschitz functions, are of bounded variation.

For the rest of the section, we describe some preliminaries needed for the proof in the next section. We utilize the pseudo-dimension [1], which is described as follows. Let $G$ be a set of functions mapping from a domain $X$ to $\Re$ and suppose that $S = \{x_1, x_2, \ldots, x_m\} \subseteq X$. Then $S$ is *pseudo-shattered* by $F$ if there are real numbers $r_1$, $r_2$, ..., $r_m$ such that for each $b \in \{0, 1\}^m$ there is a function $g_0$ in $\mathcal{G}$ with $\mathrm{sgn}(f_b(x_i) - r_i) = b_i$ for $1 \leq i \leq m$. Then $\mathcal{G}$ has the *pseudo-dimension* $d$ if $d$ is the maximum cardinality of a subset $S$ of $X$ that is pseudo-shattered by $\mathcal{G}$. If no such maximum exists, we say that $\mathcal{G}$ has infinite pseudo-dimension. The pseudo-dimension of $\mathcal{G}$ is denoted $\mathrm{Pdim}(\mathcal{G})$. Pseudo-dimensions are known for several classes such as linear spaces, and sigmoid neural networks.

Let $\mathcal{G}$ be the class of functions from $Z$ to into $[0, M]$, where $M > 0$, and let $P$ be a probability measure on $Z$. Then $d_{L^1(P)}$ is the pseudo metric on $\mathcal{G}$ defined by

$$d_{L^1(P)}(g_1, g_2) = E(|g_1 - g_2|) = \int_Z |g_1(z) - g_2(z)| dP(z)$$

for all $g_1, g_2 \in \mathcal{G}$. The *covering number* $\mathcal{N}(\epsilon, \mathcal{G}, d_{L^1(P)})$ of a function class $\mathcal{G}$ is the smallest cardinality for a subclass $\mathcal{G}^* = \{g^*\}$ of $\mathcal{G}$ such that $d_{L^1(P)}(g, g^*) \leq \epsilon$, for each $g \in \mathcal{G}$.

## 3 Sample-Based Fusion Rules

We consider an empirical implementation of a general test $R(p, u)$ by computing $\hat{p}$ such that

$$C(R(\hat{p})) = \min_{p \in [0,1]^n} \frac{1}{l} C(R(p, u^i)) \tag{3.1}$$

based on the sample. Then $R(\hat{p}, u)$ is used in place of $R(p, u)$.

**Theorem 3.1** *Consider that the fusion function $R(p, u)$ is of bounded variation with respect to $p$, and $C \leq M$ is of bounded bounded variation. Given a training sample of size*

$$s = \frac{256M^2}{\epsilon^2}\left[4n \ln\left(\frac{8eM}{\epsilon}\right) + \ln(16/\delta)\right],$$

*we have*

$$\mathbf{P}[C(R(\hat{p})) - C(R(p)) > \epsilon] < \delta.$$

**Proof:** For simplicity of notation, let us denote $C(R(p, .))$ by $C(R(p, .))$. Here $CR(p)$ is of bounded variation with respect to $p$. Consider the function class

$$\mathcal{CR} = \{CR(q, .) : q \in [0, 1]^n\}.$$

Let $\widehat{CR}(q) = \frac{1}{l} C(R(p, u^i)$. By the result of Vapnik [13] (page 41), we have

$$\begin{aligned}&\mathbf{P}\left[CR(\hat{p}) - CR(p) > \epsilon\right] \\ &\quad \leq \mathbf{P}\left[\sup_{q \in [0,1]^n} |CR(q) - \widehat{CR}(q)| > \epsilon/2\right].\end{aligned}$$

To see this result, consider the condition

$$\mathbf{P}\left[\sup_{q \in [0,1]^n} |CR(q) - \widehat{CR}(q)| > \epsilon/2\right] < \delta$$

or equivalently

$$\mathbf{P}\left[\sup_{q \in [0,1]^n} |CR(q) - \widehat{CR}(q)| < \epsilon/2\right] > 1 - \delta.$$

Then, with probability $1 - \delta$, we have

$$CR(\hat{p}) \leq \widehat{CR}(\hat{p}) + \epsilon/2 \leq \widehat{CR}(p) + \epsilon/2 \leq CR(p) + \epsilon$$

where the first and third inequalities are due to the application of supremum bound for $\hat{p}$ and $p$, respectively, and the second inequality is due to the condition $\widehat{CR}(\hat{p}) \leq \widehat{CR}(p)$. As a result, we have

$$\mathbf{P}\left[CR(\hat{p}) - CR(p) > \epsilon\right] \leq \delta,$$

which shows the above Vapnik's result.

Now using Theorem 3 of Haussler [6], we obtain

$$\mathbf{P}\left[CR(\hat{p}) - CR(p) > \epsilon\right] \leq 2E\left[2\min(\mathcal{N}(\epsilon/2, \mathcal{CR}, d_{L^1}))\right] e^{\frac{-\epsilon^2 l}{256M^2}}.$$

We now show that

$$\mathcal{N}(\epsilon, \mathcal{CR}, d_{L^1(P)}) \leq 4\left(\frac{4eM}{\epsilon}\ln\frac{4eM}{\epsilon}\right)^{2n},$$

for any $P$, which yields the required sample size. Since $CR(.) = C(R(.))$ is of bounded variance, it can be represented as a sum of two monotone functions $CR = R_1 + R_2$. For $i = 1, 2$, let

$$\mathcal{R}_i = \{R_i(q, .) : q \in [0,1]^n\},$$

which is the class of functions obtained by composing a monotone function $R_i(.)$ with the identity function $I(.)$, i. e. $I(q) = q$. Since $q$ forms a linear space, by Theorem 11.3 of [1], we have

$$\mathrm{Pdim}(\mathcal{R}_i) = \mathrm{Pdim}(\{q\}) \leq \mathrm{Pdim}([0,1]^n) = n.$$

Then by using Theorem 6 of [6] we have

$$\mathcal{N}\left(\epsilon, \mathcal{R}_i, d_{L^1(P)}\right) \leq 2\left(\frac{2eM}{\epsilon}\ln\frac{2eM}{\epsilon}\right)^{n}$$

for any measure $P$. Since $CR = R_1 + R_2$ we obtain

$$\begin{aligned}\mathcal{N}\left(\epsilon, \mathcal{CR}, d_{L^1(P)}\right) &\leq \mathcal{N}(\epsilon/2, \mathcal{R}_1, d_{L^1(P)})\mathcal{N}(\epsilon/2, \mathcal{R}_2, d_{L^1(P)}) \\ &\leq 4\left(\frac{4eM}{\epsilon}\ln\frac{4eM}{\epsilon}\right)^{2n}\end{aligned}$$

for any $P$. The sample bound follows by using

$$\begin{aligned}\delta &= 2E\left[2\min(\mathcal{N}(\epsilon/2, \mathcal{CR}, d_{L^1}))\right] e^{\frac{-\epsilon^2 l}{256M^2}} \\ &\leq 16\left(\frac{8eM}{\epsilon}\ln\frac{8eM}{\epsilon}\right)^{2n} e^{\frac{-\epsilon^2 l}{256M^2}}\end{aligned}$$

and solving for $l$. □

Often empirical mean $\bar{p}$ of $p$ based on the sample is utilized in sample-based implementation of $R(.)$, i. e. $R(\bar{p}, u)$ used in place of $R(p, u)$ as in [8]. If $R(.)$ is discontinuous the performance of this method can be quite unsatisfactory. Consider that $R(.)$ is 1 in the interval $[p - \alpha, p + \alpha]$ and zero everywhere else. As $\alpha \to 0$, we have $R(\bar{p}, u) \neq R(p, u)$ and this method yields very high error. The main problem here is that the proximity of $\bar{p}$ to $p$ is not dictated by $R(.)$ but rather by the convergence of means to the expectation. Thus this method yields good results only when the variation of $R(p)$ is conducive with that in $p$, which is not necessarily the case if $R$ is discontinuous. In our method, on the other hand, $R(.)$ has a significant influence on $\hat{p}$ due to optimization in Eq (3.1). The required proximity of $\hat{p}$ to $p$ is enforced by $R$ and hence the performance of $R(\hat{p}, .)$ approaches that of $R(p, .)$.

It is interesting to note that in the simulation of systems with independence, the method based on $\bar{p}$ yields very good results; this is true because the independence implies Lipschitz $R(p, u)$. If independence is not satisfied, $R(.)$ could be discontinuous and the performance could be quite unsatisfactory. Thus the above example cautions against using the results based on independence in a system without this property.

The sample size in Theorem 1 is entirely distribution-free and can be precomputed by using $\epsilon$, $\delta$ and $M$. The sample size of [8] cannot be precomputed since it involves the term $R(\bar{p}, u)$. Nevertheless, the result of [8] essentially guarantees that $\epsilon = 0$ for finite sample; such result is not possible based on Theorem 1.

The finite sample result of Theorem 1 implies an asymptotic result of the type common in statistics literature. Using Borel-Cantelli Lemma Theorem 1 implies that $C(R(\hat{p})) \to C(R(p))$ as $l \to \infty$.

## 4 Conclusions

We considered multiple detection systems in the case when training examples are available, but no information is available about the probability of errors committed by the individual detectors. We showed a general result that any fusion function (derived for known distributions) of bounded variation can be implemented based on a training sample with an arbitrarily high precision and confidence. Our result subsumes the cases of non-independent and correlated detectors, and provides an analytical justification for using empirical approximations of the fusion rule derived for known distributions. More generically, we show that the empirical approximation must be carefully computed, and the usual method of using the empirical means in place of probabilities can result in high error if the fusion function has discontinuities. We believe specific properties of fusion rules can be used to obtain sharper results than those yielded by our general result. The computational aspects of our method is a topic for further investigation.

In this paper, we leveraged the fusion rules derived for known distributions to obtain fusion rules for sample-based formulation. This approach is useful since a large number of fusion rules have been derived

for various formulations under known distributions. If such fusion functions are not available, however, the problem might be addressed using empirical estimation methods [11, 10, 9] which directly operate on the data (bypassing any explicit estimation of probabilities). In such cases, one could derive fusers under known distributions and then use the method proposed in this paper. It would interesting to see the boundaries of performance of these two approaches. It appears that there will be cases more suited, in terms of performance and ease of implementation, to one method than the other.


## Acknowledgments

This research is sponsored by the Engineering Research Program of the Office of Basic Energy Sciences, U.S. Department of Energy, under Contract No. DE-AC05-00OR22725 with UT-Battelle, LLC, Office of Naval Research under order No. N00014-96-F-0415, and Ballistic Missile Defense Organization under MIPR No. 0100568954.


## References


[1] M. Anthony and P. L. Bartlett. *Neural Network Learning: Theoretical Foundations.* Canbridge University Press, 1999.

[2] Z. Chair and P.K. Varshney. Optimal data fusion in multiple sensor detection systems. *IEEE Trans. Aerospace Electronic Syst.*, 22(1):98–101, 1986.

[3] B. V. Dasarathy. *Decision Fusion.* IEEE Computer Society Press, Los Alamitos, California, 1994.

[4] E. Drakopoulos and C. C. Lee. Optimal multisensor fusion of correlated local decision. *IEEE Trans. Aerospace Electronics Syst.*, 27(4):593–605, 1991.

[5] W. A. Hashlamoun and P.K. Varshney. Further results on distributed Bayesian signal detection. *IEEE Trans. on Information Theory*, 39(5):1660–1661, 1993.

[6] D. Haussler. Decision theoretic generalizations of the PAC model for neural net and other learning applications. *Information and Computation*, 100:78–150, 1992.

[7] A. Naim and M. Kam. On-line estimation of probabilities for Bayesian distributed detection. *Automatica*, 30(4):633–642, 1994.

[8] N. S. V. Rao. Distributed decision fusion using empirical estimation. *IEEE Transactions on Aerospace and Electronic Systems*, 33(4):1106–1114, 1996.

[9] N. S. V. Rao. Multiple sensor fusion under unknown distributions. *Journal of Franklin Institute*, 336(2):285–299, 1999.

[10] N. S. V. Rao. Multisensor fusion under unknown distributions: Finite sample performance guarantees. In A. K. Hyder, editor, *Multisensor Fusion.* Kluwer Academic Pub., 2000.

[11] N. S. V. Rao and S. S. Iyengar. Distributed decision fusion under unknown distributions. *Optical Engineering*, 35(3):617–624, 1996.

[12] S. C. A. Thomopoulos, R. Viswanathan, and B. K. Bougoulias. Optimal and suboptimal distributed decision fusion. *IEEE Trans. Aerospace Electronics Syst.*, 25(5):761–765, 1989.

[13] V. Vapnik. *Estimation of Dependences Based on Empirical Data.* Springer-Verlag, New York, 1982.

[14] P. K. Varshney. *Distributed Detection and Data Fusion.* Springer-Verlag, 1997.