A PARAMETRIC COPULA BASED FRAMEWORK FOR MULTIMODAL SIGNAL PROCESSING

Satish G. Iyengar, Pramod K. Varshney[*]

Department of EECS
Syracuse University, NY 13244
Email: {siyengar, varshney}@syr.edu

Thyagaraju Damarla

U.S. Army Research Lab
Adelphi, MD 20783 - 1197
Email: rdamarla@arl.army.mil

ABSTRACT

We present a framework for the joint processing of multimodal data such as audio-video data streams. We first consider the problem of estimating the joint distribution of statistically dependent multimodal random variables. We discuss the issues involved and provide a copula based solution. Application of this approach to solve a multisensor fusion problem for the detection of a random event is also discussed.

Index Terms— Copula theory, Multisensor fusion, Hypothesis testing, Kullback-Leibler distance, Statistical dependence

1. INTRODUCTION

Statistical signal processing tasks such as detection, estimation and tracking always require complete specification of the joint probability distribution of the observed samples. However, in many cases, the derivation of the joint probability density function (PDF) becomes mathematically intractable. In problems such as multimodal signal processing, random variables associated with each modality may follow probability distributions that are different from one another. This is due to several *physical* differences such as in their dimensionality, support and sampling rate. Moreover, in most applications, the signals share a common source and thus may exhibit statistical dependency. Consider, for example, an acoustic sensor and a video camera monitoring a region for trespassers. Presence of a target may result in an increase in both the acoustic energy and the pixel intensities of the images acquired by the video camera. Both sensors provide information about the same event (and hence are statistically dependent) but in different *'domains'*. In this case, it is highly likely that the acoustic features and the pixel intensities will not follow the same probability distribution. We are thus faced with two challenges when modeling the joint distribution of random variables corresponding to multimodal data,

- How do we characterize the inter-modal dependence structure?
- How do we model the joint distribution between statistically dependent multimodal measurements when the underlying marginals follow disparate distributions?

We discuss parametric modeling of multimodal data in Section 2. We show here how copula functions provide an approach to model the joint multimodal statistics. We formulate a binary hypotheses testing problem in Section 3 and present a multisensor detection example in Section 4.

[*]This work was supported by the US Army Research Office under grant W911NF-06-1-0250. We thank Lav Varshney for technical discussions.

2. MODELING

The acoustic-video sensor example given above motivates the following general definition for heterogeneous random vectors that would arise with multimodal signals.

Definition 1 *A random vector* $\mathbf{Z} = \{Z_n\}_{n=1}^N$ *governing the joint statistics of an N-variate data set can be termed as heterogeneous if the marginals* Z_n *are non-identically distributed.*

In the following, we assume that the marginal PDFs, $f_{Z_n}(z_n)$, are known and the goal is to construct the joint PDF $\mathbf{f_Z(z)}$ of the multimodal random vector $\mathbf{Z}$. Further, the variables Z_n exhibit statistical dependence so that $\mathbf{f_Z(z)} \neq \prod_{n=1}^N f_{Z_n}(z_n)$.

Characterizing multivariate statistical dependence is one of the most widely researched topics and has always been a difficult problem [1]. The most commonly used bivariate measure, the Pearson's correlation ρ captures the linear relationship between variables and is a weak measure of dependence when dealing with non-Gaussian random variables. Two random variables X and Y are said to be uncorrelated if the covariance, $\Sigma_{X,Y} = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ is zero ($\rho = 0$). Statistical independence has a stricter requirement in that X and Y can be called independent only if their joint density can be factored as the product of the marginals. In general, a zero correlation does not guarantee independence (except when the variables are jointly Gaussian).

The problem is further compounded when dealing with a multimodal random vector such as $\mathbf{Z}$ with complex inter-modal interactions between the component variables Z_n that follow disparate probability distributions. Thus, the derivation of multimodal joint PDF becomes difficult and one is often forced to assume multivariate Gaussian or inter-modal independence to construct a tractable statistical model. A multivariate Gaussian model would necessitate the marginals to be Gaussian and thus would fail to utilize the knowledge of the given marginal PDFs. Assuming statistical independence neglects inter-modal dependence thus leading to suboptimal solutions.

Alternatively, we propose in this work, a copula based model to represent the inter-modal dependence structure. Copulas are functions that couple multivariate joint distributions to their component marginal distribution functions [2], [3]. The main advantage of the copula based approach is that it allows us to define inter-modal dependence irrespective of the underlying marginal distributions. One can thus construct joint distributions with arbitrary marginals and the desired dependence structure. This property is well suited especially for the joint processing of multimodal variables with different marginal distributions.

Sklar (1959) was the first to define copula functions.

Sklar's Theorem and its Implications

Theorem 1 *(Sklar's Theorem)*
Let $F_{\mathbf{Z}}(z_1, z_2, \cdots z_N)$ be the joint cumulative distribution function (CDF) with continuous marginals $F_{Z_1}(z_1)$, $F_{Z_2}(z_2)$, $\cdots$, $F_{Z_N}(z_N)$. Then there exists a copula function $C(\cdot)$ such that for all $z_1, z_2, \cdots, z_n$ in $[-\infty, \infty]$,

$$F_{\mathbf{Z}}(z_1, z_2, \cdots z_N) = C(F_{Z_1}(z_1), F_{Z_2}(z_2), \cdots, F_{Z_N}(z_N)) \quad (1)$$

For continuous marginals, $C(\cdot)$ is unique; otherwise $C(\cdot)$ is uniquely determined on $RanF_{Z_1} \times RanF_{Z_2} \cdots \times RanF_{Z_N}$ where $RanX$ denotes the range of X. Conversely, if $C(\cdot)$ is a copula and $F_{Z_1}(z_1)$, $F_{Z_2}(z_2)$, $\cdots$, $F_{Z_N}(z_N)$ are marginal CDFs then the function $F_{\mathbf{Z}}(\cdot)$ in (1) is a valid joint CDF with the marginals $F_{Z_1}(z_1)$, $F_{Z_2}(z_2)$, $\cdots$, $F_{Z_N}(z_N)$.

Note that the copula function $C(u_1, u_2, \cdots, u_N)$ is itself a CDF with uniform marginals as $u_n = F_{Z_n}(z_n) \sim \mathcal{U}(0, 1)$ (by probability integral transform).

The copula based joint PDF of N continuous heterogeneous random variables can now be obtained by taking an N^{th} order derivative of (1),

$$\begin{aligned} \mathbf{f}_{\mathbf{Z}}(\mathbf{z}) &= \left(\prod_{n=1}^{N} \mathbf{f}_{Z_n}(z_n)\right) \mathbf{c}(F_{Z_1}(z_1), \cdots, F_{Z_N}(z_N)) \\ &= \mathbf{f}_{\mathbf{Z}}^{\mathbf{c}}(\mathbf{z}) \end{aligned} \quad (2)$$

where $\mathbf{Z} = [Z_1, Z_2, \cdots, Z_N]$ and we use the *superscript* $'\mathbf{c}'$ to denote that $\mathbf{f}_{\mathbf{Z}}^{\mathbf{c}}(\mathbf{z})$ is the copula representation of $\mathbf{f}_{\mathbf{Z}}(\mathbf{z})$. Note that we need to know the true copula density function '$\mathbf{c}(\cdot)$' to have an exact representation as in (2). We emphasize here that any joint PDF with continuous marginals can be written in terms of a copula function as in (2) (due to Sklar's theorem). However, identifying the true copula is not a straightforward task. A common approach then is to select a copula function *a priori* and fit the given marginals and the desired dependence structure to derive the joint distribution.

Several copula functions have been defined especially in the econometrics and finance literature (*e.g.* [4]); the popular ones among them being multivariate Gaussian copula, Student's *t* copula and copula functions from the Archimedean family. Given a copula density function $\mathbf{k}(\cdot)$ and the marginal distributions, the joint PDF estimate then has the form similar to (2),

$$\begin{aligned} \widehat{\mathbf{f}_{\mathbf{Z}}}(\mathbf{z}) &= \left(\prod_{n=1}^{N} \mathbf{f}_{Z_n}(z_n)\right) .\mathbf{k}(F_{Z_1}(z_1), \cdots, F_{Z_N}(z_N)) \\ &= \mathbf{f}_{\mathbf{Z}}^{\mathbf{k}}(\mathbf{z}) \end{aligned} \quad (3)$$

The dependence structure between the marginals is completely captured by the copula function and is separate from the choice of the marginals. Next, we describe the joint PDF construction using copulas.

2.1. Joint PDF Construction using Copulas

As an example, consider two random variables Z_1 and Z_2 associated with two sources of different modalities. Given a copula function $K(.)$, we wish to construct a copula based bivariate density function of the form as in (3). Table 1 lists some of the well known bivariate copulas. Each of these functions is parameterized by 'θ' that controls the '*amount of dependence*' between the two variables. Thus, it is required to estimate θ from the acquired bivariate observations. We describe how this is done using nonparametric dependence measures as this method is computationally efficient.

Table 1. Copula functions

Copula	$C(u_1, u_2)$	Kendall's τ
Gaussian	$\Phi_N[\Phi^{-1}(u_1), \Phi^{-1}(u_2); \theta]$	$\frac{2}{\pi}\arcsin(\theta)$
Clayton	$[u_1^{-\theta} + u_2^{-\theta} - 1]^{-\frac{1}{\theta}}$	$\frac{\theta}{\theta+2}$
Frank	$-\frac{1}{\theta}\log\left(1 + \frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{e^{-\theta}-1}\right)$	$1 - \frac{4}{\theta}\left[1 - \frac{1}{\theta}\int_0^{\theta} \frac{t}{e^t-1}dt\right]$
Gumbel	$\exp\left[-\left\{(-\log u_1)^{\theta} + (-\log u_2)^{\theta}\right\}^{1/\theta}\right]$	$1 - \frac{1}{\theta}$
Product	$u_1.u_2$	0

The copula dependence parameter θ can be expressed as a function of Kendall's τ, a nonparametric measure of association between two random variables [2]. Specifically, Kendall's τ measures the concordance between two random variables. Let $(z_1(i), z_2(i))$ and $(z_1(j), z_2(j))$ be two observations from a bivariate measurement vector $(\mathbf{Z_1}, \mathbf{Z_2})$ of continuous random variables. The observations are said to be concordant if $(z_1(i) - z_1(j))(z_2(i) - z_2(j)) > 0$ and discordant if $(z_1(i) - z_1(j))(z_2(i) - z_2(j)) < 0$.

The population version of Kendall's τ can be expressed in terms of $K(.)$ as

$$\tau_{Z_1,Z_2} = 4\int\int K(u_1, u_2; \theta) dK(u_1, u_2; \theta) - 1 \quad (4)$$

where $u_n = F_{Z_n}(z_n)$. Thus, for a given τ, the integral equation above can be used to solve for θ. Table 1 shows the relationship between τ and θ for some of the well-known copula functions.

When τ is unknown, θ can be obtained from the sample estimate $\hat{\tau}$. Given L $i.i.d$ measurements $(z_1(l), z_2(l))_l$ $(l = 1, 2, \cdots, L)$, the observations are rank ordered and $\hat{\tau}$ can be computed as

$$\hat{\tau}_{Z_1,Z_2} = \frac{c - d}{c + d} \quad (5)$$

where c and d are the number of concordant and discordant pairs respectively.

2.2. Joint PDF Construction assuming inter-modal independence

Joint PDF estimate assuming inter-modal statistical independence is given as the product of the marginals, *i.e.*.

$$\widehat{\mathbf{f}_{\mathbf{Z}}}(\mathbf{z}) = \prod_{n=1}^{N} \mathbf{f}_{Z_n}(z_n) = \mathbf{f}_{\mathbf{Z}}^{\mathbf{m}}(\mathbf{z}) \quad (6)$$

where we use superscript $'\mathbf{m}'$ to denote that the joint PDF estimate depends on the marginal independence assumption.

Thus, both joint PDF estimates (3) and (6), capture the given maginal densities. The copula based joint PDF estimate further captures Kendall's τ, the rank correlation between the variables.

We next formulate a binary hypotheses testing problem.

3. HYPOTHESES TESTING

A decision theory problem consists of deciding which of the hypotheses $\mathbf{H_0}, \cdots, \mathbf{H_k}$ is true based on the acquired observation vector of (say) L samples. An optimal test (in both the Neyman-Pearson (NP) and Bayesian sense) for a two hypotheses problem

($\mathbf{H_0}$ *vs.* $\mathbf{H_1}$) computes the log-likelihood ratio ($\mathbf{\Lambda}$) and decides in favor of $\mathbf{H_1}$ when the ratio is larger than a pre-defined threshold (η),

$$\Lambda(\mathbf{z}) = \log \frac{\mathbf{f_Z}(\mathbf{z}|\mathbf{H_1})}{\mathbf{f_Z}(\mathbf{z}|\mathbf{H_0})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{7}$$

$\mathbf{f_Z}(\mathbf{z}|\mathbf{H_i})$ is the joint PDF of the random observation vector $\mathbf{z} = [z_1, \cdots, z_N]^T \in \mathbb{R}^N$ under the hypothesis $\mathbf{H_i}$, (i = 0, 1) and includes all the statistics required to derive (7). In the NP set up, the threshold 'η' is selected to constrain the false alarm error probability, P_F to a value $\alpha < 1$ and at the same time minimize the probability of miss, P_M. The two error probabilities are given as

$$P_F = P(\Lambda > \eta|\mathbf{H_0}), \;\; P_M = P(\Lambda < \eta|\mathbf{H_1}) \tag{8}$$

Consider a binary hypotheses testing problem where

$$\begin{aligned} \mathbf{H_1} &: \;\; \mathbf{f_Z}(z_1, z_2, \cdots, z_N|\mathbf{H_1}) \\ \mathbf{H_0} &: \;\; \mathbf{f_Z}(z_1, z_2, \cdots, z_N|\mathbf{H_0}) = \prod_{n=1}^{N} \mathbf{f}_{Z_n}(z_n|\mathbf{H_0}) \end{aligned} \tag{9}$$

Thus, it is known that the random variables $Z_1, \cdots, Z_N$ are statistically independent under the hypothesis $\mathbf{H_0}$. However, the joint distribution $\mathbf{f_Z}(z_1, z_2, \cdots, z_N|\mathbf{H_1})$ under hypothesis $\mathbf{H_1}$ is unknown. We use copula theory to estimate $\mathbf{f_Z}(z_1, z_2, \cdots, z_N|\mathbf{H_1})$ and derive the log-likelihood ratio test.

3.1. Heterogeneous log-likelihood ratio test

We use (3) to estimate $\mathbf{f_Z}(z_1, z_2, \cdots, z_N|\mathbf{H_1})$ in terms of marginals and derive the copula based heterogeneous log-likelihood ratio test (HLRT) statistic as below,

$$\begin{aligned} \mathbf{\Lambda_k}(\mathbf{z}) &= \log \frac{\widehat{\mathbf{f_Z}}(\mathbf{z}|\mathbf{H_1})}{\mathbf{f_Z}(\mathbf{z}|\mathbf{H_0})} = \log \frac{\mathbf{f_Z^k}(\mathbf{z}|\mathbf{H_1})}{\mathbf{f_Z}(\mathbf{z}|\mathbf{H_0})} &(10)\\ &= \log \left(\prod_{n=1}^{N} \frac{\mathbf{f}_{Z_n}(z_n|\mathbf{H_1})}{\mathbf{f}_{Z_n}(z_n|\mathbf{H_0})} \right) + \\ &\quad \log \left[\mathbf{k}(F_{Z_1}^1(z_1), \cdots, F_{Z_N}^1(z_N)) \right] &(11) \end{aligned}$$

where the superscript 'i' in $F_{Z_n}^i(z_n)$ denotes the CDF of Z_n under hypothesis i.

It can be seen from (11) that the copula function allows one to exactly factor out the role of cross modal dependence from the strategies employed by the individual modalities. This allows to quantify performance gains (if any) achieved due to inter-modal dependence.

3.2. Marginal log-likelihood ratio test

It is interesting to note the form of the test statistic in (11). The first term,

$$\mathbf{\Lambda_m}(\mathbf{z}) = \log \left(\prod_{n=1}^{N} \frac{\mathbf{f}_{Z_n}(z_n|\mathbf{H_1})}{\mathbf{f}_{Z_n}(z_n|\mathbf{H_0})} \right) \tag{12}$$

is the test obtained when dependence between the variables Z_1, Z_2, $\cdots$, Z_N is neglected. We call this test the marginal likelihood ratio test (MLRT). In problems where the derivation of the joint density becomes mathematically intractable, tests are usually employed assuming independence between variables conditioned on each hypothesis.

We next compare performances of HLRT and MLRT detectors.

3.3. Performance Analysis using Error Exponents

The asymptotic performance of a likelihood ratio test can be quantified using the KL distance, $D\left(\mathbf{f_Z}(\mathbf{z}|\mathbf{H_1})||\mathbf{f_Z}(\mathbf{z}|\mathbf{H_0})\right)$, between the PDFs underlying the two hypotheses. For 'L' *i.i.d.* N-variate measurements $\mathbf{Z}_l = [Z_1, Z_2 \cdots Z_N]_l$, $(l = 1, 2, \cdots, L)$, through Stein's Lemma [5], we have for a fixed value of $P_M = \beta$, $(0 < \beta < 1)$,

$$\lim_{L \to \infty} \frac{1}{L} \log P_F = -D\left(\mathbf{f_Z}(\mathbf{z}|\mathbf{H_1})||\mathbf{f_Z}(\mathbf{z}|\mathbf{H_0})\right) \tag{13}$$

The greater the value of $D\left(\mathbf{f_Z}(\mathbf{z}|\mathbf{H_1})||\mathbf{f_Z}(\mathbf{z}|\mathbf{H_0})\right)$, faster is the convergence of P_F to zero as $L \to \infty$. The KL distance is thus indicative of the performance of a log-likelihood ratio test. Further, it is additive for independent observations,

$$D\left(\mathbf{f_Z^m}(\mathbf{z}|\mathbf{H_1})||\mathbf{f_Z}(\mathbf{z}|\mathbf{H_0})\right) = \sum_{n=1}^{N} D\left(\mathbf{f}_{Z_n}(z_n|\mathbf{H_1})||\mathbf{f}_{Z_n}(z_n|\mathbf{H_0})\right) \tag{14}$$

where $D\left(\mathbf{f}_{Z_n}(z_n|\mathbf{H_1})||\mathbf{f}_{Z_n}(z_n|\mathbf{H_0})\right)$ is the KL distance for a single modality Z_n.

Assuming the knowledge of the true underlying copula $\mathbf{c}(\cdot)$, we have the following theorem to compare the asymptotic performances of HLRT and MLRT for the binary hypotheses testing problem formulated in (9).

Theorem 2 *The KL distance between the two competing hypotheses ($\mathbf{H_0}$ vs. $\mathbf{H_1}$) increases by a factor equal to multiinformation (under $\mathbf{H_1}$) when dependence between the variables is taken into account.*

$$\begin{aligned} D\left(\mathbf{f_Z}(\mathbf{z}|\mathbf{H_1})||\mathbf{f_Z}(\mathbf{z}|\mathbf{H_0})\right) &- D\left(\mathbf{f_Z^m}(\mathbf{z}|\mathbf{H_1})||\mathbf{f_Z}(\mathbf{z}|\mathbf{H_0})\right) \\ &= \underbrace{\mathcal{I}(Z_1; Z_2; \cdots; Z_N)}_{\geq \; 0} \end{aligned} \tag{15}$$

Multiinformation [6] is defined as

$$\mathcal{I}(Z_1; Z_2; \cdots; Z_N) = \int_{\mathbf{z}} \mathbf{f_Z}(\mathbf{z}) \log \left(\frac{\mathbf{f_Z}(\mathbf{z})}{\prod_{i=1}^{N} \mathbf{f}_{Z_i}(z_i)} \right) d\mathbf{z} \tag{16}$$

This is intuitively satisfying as the multiinformation $\mathcal{I}(\cdot)$ (which reduces to the well-known mutual information when $N = 2$) describes the complete nature of dependence between the variables. Our result, that the error exponent increases due to inter-modal dependence, agrees with Koval *et. al.* [7], where the authors use the logarithmic inequality $\left(\log(x) \geq 1 - \frac{1}{x}\right)$ to prove the result. Our approach is based on the copula representation of joint PDF and goes further to quantify the performance loss ($= \mathcal{I}(\cdot)$) when multimodal signals are statistically independent or when the dependence is deliberately neglected for simplicity.

In the next section, we present a multisensor detection example and compare detector receiver operating characteristics (ROCs) when 100 *i.i.d* sensor measurements are available in each decision window.

4. A MULTISENSOR DETECTION EXAMPLE

Consider a parallel network of sensors as shown in Fig. 1 where each of the deployed sensors may have different sensing capabilities. The sensors monitor a common region of interest and the goal is to design

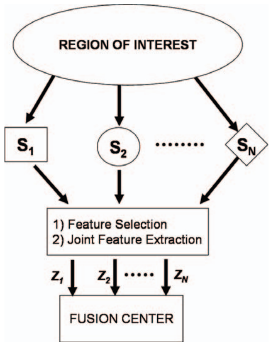

Fig. 1. A multisensor detection system with common region of interest (ROI). Different shapes for the sensors denote different sensing modalities.

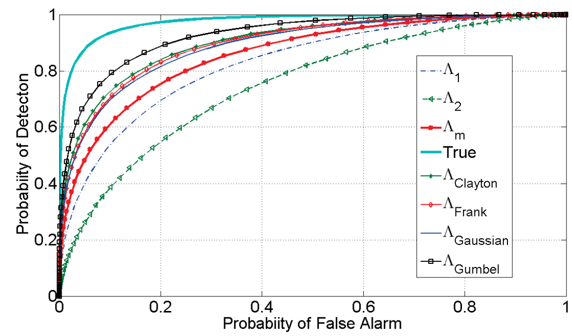

Fig. 2. Monte Carlo based ROCs for MLRT and HLRT constructed using different copula functions. We set $\beta_0 = \sigma_0^2 = 1$ and $\beta_1 = \sigma_1^2 = 1.1$.

an algorithm that can combine the acquired multimodal information for detecting the occurrence of a random event. We generate a synthetic multimodal data set and show simulation results for $N = 2$. In the future, we will apply this approach to some real datasets and evaluate its performance.

Copula models allow one to generate multivariate random vectors whose properties satisfy definition 1 [2]. We use the Student's *t* copula with degree of freedom (*d.o.f*) four and Kendall's τ equal to 0.1 to generate bivariate multimodal data under the hypothesis $\mathbf{H_1}$. The marginals follow gamma and Gaussian distributions under each hypothesis ($\mathbf{H_i}; i = 0, 1$),

$$\mathbf{f}_{Z_1}(z_1|\mathbf{H_i}) = \frac{e^{-\frac{z_1}{\beta_i}}}{\beta_i^{\alpha_i}\Gamma(\alpha_i)}.z_1^{\alpha_i - 1}, \;\; z_1 > 0 \text{ and } \alpha_i, \beta_i > 0 \quad (17)$$

$$\mathbf{f}_{Z_2}(z_2|\mathbf{H_i}) = \frac{1}{\sqrt{2\pi\sigma_i^2}}.e^{-\frac{z_2^2}{2\sigma_i^2}}, \;\; -\infty \leq z_2 \leq \infty, \text{and } \sigma_i^2 > 0 \quad (18)$$

where $\beta_1 > \beta_0$ and $\sigma_1^2 > \sigma_0^2$. We set $\alpha_1 = \alpha_0$. While a bivariate Gaussian distribution with non-identical means and/or variances would satisfy Definition 1 and could have been used to generate the synthetic data set, we adopt the above model to emphasize that the proposed copula based methodology for multimodal signal processing does not restrict the marginals to the same parametric family of distributions.

Let Λ_1 and Λ_2 denote log-likelihood ratio tests corresponding to Z_1 and Z_2 respectively. We plot, in Fig. 2, the ROCs for Λ_1, Λ_2, $\mathbf{\Lambda_m}$ $(= \Lambda_1 + \Lambda_2)$ and $\mathbf{\Lambda_k}$ using 50,000 Monte Carlo trials. Assuming that the true joint PDF (used to generate data) is unknown we construct $\mathbf{\Lambda_k}$ using arbitrary copula functions (from table 1). As evident from Fig. 2, $\mathbf{\Lambda_m}$ performs better than each of the single sensor test statistic as expected. However, fusion using HLRT outperforms MLRT as the copula functions (even though misspecified) capture the rank correlation between the multimodal measurements and hence deviate from the erroneous inter-modal independence assumption.

5. CONCLUDING REMARKS

We have presented a copula based framework for multimodal signal processing. We show how copula functions allow us to model the joint PDF of multimodal measurements. We also present a multisensor detection example where the use of copula functions is advantageous than assuming inter-modal independence. We note that this may not be true in general. Derivation of a universal improvability condition, selection of the best copula and the incorporation of dependence measures other than the rank correlations will be addressed in the future.

6. REFERENCES

[1] D. Mari and S. Kotz, *Correlation and Dependence*. London: Imperial College Press, 2001.

[2] R. B. Nelsen, *An Introduction to Copulas*. New York: Springer-Verlag, 1999.

[3] D. Kurowicka and R. Cooke, *Uncertainty Analysis with High Dimensional Dependence Modeling*. West Sussex, England: John Wiley and Sons, Ltd, 2006.

[4] R. T. Clemen and T. Reilly, "Correlations and copulas for decision and risk analysis," *Management Science*, vol. 45, pp. 208–224, Feb. 1999.

[5] H. Chernoff, "Large-sample theory: Parametric case," *Ann. Math. Statist.*, vol. 27, pp. 1–22, Mar. 1956.

[6] M. Studen, Y, and J. Vejnarov, "The multi-information function as a tool for measuring stochastic dependence," in *Learning in Graphical Models*, pp. 261–298, Kluwer Academic Publishers, 1998.

[7] O. Koval, S. Voloshynovskiy, and T. Pun, "Analysis of multimodal binary detection systems based on dependent/independent modalities," in *IEEE 9th Workshop on Multimedia Signal Processing*, pp. 70–73, 2007.

[8] T. Cover and J. Thomas, *Elements of Information Theory*. Hoboken, New Jersey: John Wiley and Sons, Ltd, 2006.

[9] S. G. Iyengar, P. K. Varshney, and T. Damarla, "On the detection of footsteps using acoustic and seismic sensing," in *Proc. of 41st Annual Asilomar Conference on Signals,Systems and Computers*, (Pacific Grove, CA), Nov. 2007.

[10] J. Fisher III and T. Darrell, "Speaker association with signal-level audiovisual fusion," *IEEE Trans. Multimedia*, vol. 6 (3), pp. 406–413, 2004.

[11] T. Butz and J. Thiran, "From error probability to information theoritic (multi-modal) signal processing)," *Elsevier: Signal Processing*, vol. 85, pp. 875–902, May 2005.