On the Detection of Footsteps Based on Acoustic and Seismic Sensing

Satish G. Iyengar, Pramod K. Varshney
Department of EECS, 335 Link Hall, Syracuse University, Syracuse, NY 13244
Email: {siyengar, varshney}@syr.edu

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

Abstract—In this work, we present a copula-based framework for integrating signals of different but statistically correlated modalities for binary hypothesis testing problems. Specifically, we consider the problem of detecting the presence of a human using footstep signals from seismic and acoustic sensors. An approach based on canonical correlation analysis and copula theory is employed to establish a likelihood ratio test. Experimental results based on real data are presented.

Index Terms—Canonical Correlation Analysis, Copula theory, Multisensor systems, Sensor fusion, footstep detection

I. INTRODUCTION

HUMAN activity detection plays an important role in surveillance systems. Multiple sensors may be deployed at different locations of a region of interest to monitor for intruders or trespassers. Data from each of these sensors is then fused (according to a pre-designed fusion rule) to arrive at a final decision regarding the presence or absence of any human activity. Video is perhaps the most common mode of sensing that one can envision for zone monitoring and protection systems. However, in some scenarios (e.g. military applications), the deployed sensors may be required to possess long sensing range capabilities and a direct line-of-sight (LOS) may also be unavailable. Unattended Ground Sensor (UGS) network of acoustic and seismic sensors is usually employed in such scenarios [1], [2]. Both acoustic and seismic sensors also have an added advantage of being passive, inexpensive and easy to install. Several studies have looked at footstep signal detection using sensors of single modality. For example, Succi et al. in [3] and [4], consider the problem of footstep signal detection and personnel tracking using seismic sensors. In this paper, we pose the footstep detection problem as a binary hypotheses test and consider a framework of joint processing of the acoustic and seismic information. Human detection using acoustic and seismic footstep signals has also been addressed in [5] where the authors describe methods to enhance the footstep signature in the acquired noisy signal to aid a listener.

A multisensor surveillance system not only increases the spatio-temporal coverage but also enhances system performance [6]. Further, a sensor network with each (or a subgroup of) sensor(s) acquiring data of different modalities (e.g. an audio-video sensor network) can be highly attractive if one is able to fuse and efficiently exploit the synergy between such disparate sources of information. Fusion of sensors’ data can be achieved at three different levels: (i) data level, (ii) feature level and (iii) decision level. In (i) and (ii), sensor observations and relevant features extracted from sensor observations are fused respectively, to arrive at a final decision. This is equivalent to the centralized processing of sensors’ data. In (iii), each sensor makes a local decision regarding the presence of the target based on its observations and only these local decisions are fused to arrive at a final decision. One problem common to both strategies (i) and (ii) is that it is not easy to obtain the complex dependence structure between data of different modalities. For example, consider an acoustic sensor and a video camera monitoring a region. Both the sensors provide information pertaining to the same event but in different ‘domains’. Presence of a target may result in an increase in both the acoustic energy and the pixel intensities of images acquired by the video camera. However, it is difficult to define a parametric form for the acoustic-video dependence structure accurately. As a result, often one has to resort to rigorous training of the system or is forced to assume joint Gaussianity. Several authors have used canonical

![Diagram](https://example.com/diagram.png)
correlation analysis (CCA) for multimodal signal processing. For example, Elad et al. in [7] develop an algorithm to locate pixels in images (sampled video signal) that correspond to the acquired audio signal. Tekalp et al. in [8] use CCA to extract audio and lip information for speaker identification. In this paper, a novel method for fusing heterogeneous data is presented which makes use of copulas in conjunction with CCA. Though we have used acoustic and seismic sensors, the method is general and can be extended to data obtained from other sensing modalities as well.

The paper is organized as follows. In Section II, we formulate the problem and describe the proposed fusion framework for general multimodal signals. Section III addresses the footstep detection problem using acoustic and seismic sensors using the test derived in Section II. Performance results based on the real data are described in Section IV. Finally we draw conclusions and discuss future research directions in Section V.

II. PROBLEM FORMULATION

Figure 1 shows the basic set up. Consider a binary hypotheses testing problem where the null hypothesis $H_0$ is tested against $H_1$. Let $X$ and $Y$ denote sensor observations (or corresponding feature vectors) of two different but correlated modalities. These observations (or features) are first linearly transformed to their respective CCA variates, $U$ and $V$. CCA maximizes the degree of linear correlation between the two modalities. We then design a copula-based log-likelihood ratio test (LLRT) using the derived canonical variates. The copula density function serves to "stitch" or couple the marginal distributions of the two modalities for the LLRT formulation. Below we describe CCA and Copula theory in brief.

A. Canonical Correlation Analysis

CCA is a multivariate statistical analysis method or tool to measure linear relationship between two multidimensional variables, (say) $X$, $Y \in \mathbb{R}^p$, $\mathbb{R}^q$ respectively [9], [10]. The method finds two basis vectors (also called canonical weights) $a_{p1}$ and $b_{q1}$, such that the variates $U = a^T X$ and $V = b^T Y$ are maximally correlated. One can use singular value decomposition (SVD) to obtain the canonical correlation matrix $K$.

Let $Z = \begin{bmatrix} X^T & Y^T \end{bmatrix}^T$. Assuming that $Z$ has a zero mean, its covariance matrix is given as

$$R_{zz} = E [ZZ^T] = \begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix}$$

(1)

To solve for the canonical correlations and variates, SVD of the coherence matrix $C_h$ is obtained where

$$C_h = R_{xx}^{1/2} R_{xy} R_{yy}^{1/2} = L K M^T$$

(2)

$L$ and $M$ are unitary matrices and $K$ is a diagonal matrix with each of its element being the canonical correlation corresponding to a particular linear combination. Also, the canonical variates $U$ and $V$ are given as

$$U = L^T R_{xx}^{1/2} X$$

(3)

$$V = M^T R_{yy}^{1/2} Y$$

(4)

In transforming the original variables to the CCA domain, often the constraint imposed is the normalization of the second order characteristics; i.e. the mean and variance of the $U$ and $V$ are zero and one respectively.

CCA helps us achieve two things: (a) Original data matrices $X$ and $Y$ are now mapped to a space where they have maximum linear correlation, (b) Dimensionality reduction can be achieved by discarding those variates for which the correlations are not significant.

B. Copula theory

Copulas are functions that couple multivariate joint distributions to their component marginal distribution functions [11], [12]. Sklar (1959) was the first to define copula functions which can be stated as below,

\textbf{Sklar’s Theorem}

The joint cumulative distribution function (CDF) $F(x_1, x_2, \ldots, x_n)$ of random variables $X_1, X_2, \ldots, X_n$ are joined by a copula function $C$ to their respective marginal distributions $F(x_1), F(x_2), \ldots, F(x_n)$ as

$$F(x_1, x_2, \ldots, x_n) = C (F(x_1), F(x_2), \ldots, F(x_n))$$

(5)

It is easy to see that $C(u_1, u_2, \ldots, u_n)$ is a distribution function with uniform marginals. Moreover, there always exists a unique copula $C$ for every continuous joint CDF. See [12] for more details. Hence the joint probability density function (PDF) is obtained by taking an $n^{th}$ order differentiation of (5) and is given as

$$f(x_1, \ldots, x_n) = \left( \prod_{i=1}^{n} f(x_i) \right) c_0(F(x_1), \ldots, F(x_n))$$

(6)

Thus, the dependence structure between the marginals is captured completely by the copula function and is separate from the choice of the marginals. For example, it is possible to construct a bivariate distribution $F(x_1, x_2)$ with a specified rank correlation (specified by $\theta$ in (6)) when the random variables $X_1$ and $X_2$ individually follow Gamma and Gaussian distributions respectively. This is well suited for joint processing of signals with different modalities and hence possibly different probability distributions. For the case when $X_1$ and $X_2$ are independent, $c_0(.) = 1$. Several copula functions have been defined especially in the econometrics and finance literature [11], [13]: the popular ones among them being multivariate normal copula, Student’s $t$ copula and copula functions from the Archimedean family.

We now turn to our problem of footstep detection using acoustic and seismic signals.

III. FOOTSTEP DETECTION

In Fig 1, let $Z_a$ and $Z_s$ denote $N$ samples of noisy acoustic and seismic footstep signal measurements. Note that in general, the data length of the two signals may not be equal due to the difference in sampling rates. However, it is straightforward to interpolate/resample one or both of them so that they are of equal lengths. From the collected $N$ samples, we estimate
time localized frequency content of the received signals using Short Time Fourier Transform (STFT) and obtain $X_{N\times P}$ and $Y_{N\times Q}$ where each row contains complex DFT coefficients for a particular time bin. We define the two hypotheses $H_0$ and $H_1$ to indicate the absence and presence of footsteps (and hence the target) respectively. Now,
\[ X_k(t), Y_j(t) = \begin{cases} W^a_k(t), W^a_j(t) & \text{if } H_0 \text{ true} \\ S^s_k(t) + W^s_k(t), S^s_j(t) + W^s_j(t) & \text{if } H_1 \text{ true} \end{cases} \] (7)

Here, $W^a_k$ and $S^s_k$ are $k^{th}$ DFT coefficients of acoustic noise and acoustic footstep signal respectively. Similarly, $W^s_j$ and $S^s_j$ are $j^{th}$ DFT coefficients of seismic noise and seismic footsteps signal respectively. The time bins are $t = 1, 2, \ldots, N$ and the DFT points are $k = 1, 2, \ldots, P$ and $j = 1, 2, \ldots, Q$ for acoustic and seismic signals respectively.

We assume a zero mean complex Gaussian PDF for the DFT coefficients of both acoustic and seismic signals. Several studies in the speech signal processing literature have applied the complex Gaussian pdf to characterize DFT coefficients of noisy speech (e.g. [14]). Moreover, the Gaussian model is motivated by the central limit theorem since the DFT coefficients are a weighted sum of random variables resulting from process samples. Recent studies have however suggested that the complex Laplacian and complex Gamma distributions are stronger candidates to explain DFT coefficients distribution [15].

With the complex Gaussian model,
\[ f(X_k(t)|H_0) = \frac{1}{\pi \alpha^2_k} \exp \left( -\frac{|X_k(t)|^2}{\alpha^2_k} \right) \] (8)
\[ f(X_k(t)|H_1) = \frac{1}{\pi \beta^2_k} \exp \left( -\frac{|X_k(t)|^2}{\beta^2_k} \right) \] (9)

where $\beta^2_k = \alpha^2_k + \sigma^2_k$, $\sigma^2_k$ being the variance for each frequency bin $k$ of noise-free footstep STFT. Thus, the LRT statistic for acoustic sensor alone can be derived as
\[ \log \Lambda_X = \sum_{k=1}^{K} \left( \frac{1}{\alpha^2_k} - \frac{1}{\beta^2_k} \right) \cdot \sum_{t=1}^{N} |X_k(t)|^2 \] (10)

assuming independence in time and frequency bins. In the above equation, we assume that $\alpha^2_k$ is known or we have access to the noise-only data for its estimation. This is a reasonable assumption provided the background noise does not exhibit high variability as it is always possible to perform noise characterization before system deployment. Also, $\beta^2_k$ in (10) is replaced by its maximum likelihood estimate (MLE) to obtain the generalized likelihood ratio test (GLRT) for the acoustic sensor.

Similarly, the test statistic $\log \Lambda_Y$ can be derived for the seismic sensor as well.

A. CCA-Copula based likelihood ratio test

First, acoustic and seismic variates are obtained using CCA. To completely describe the linear relationship between the two variables, $s = \min(\text{rank}(X), \text{rank}(Y))$ such linear combinations can be used as below,
\[ U = |X|^2_{(N\times P)} \cdot A_{(P \times s)} \] (11)

and
\[ V = |Y|^2_{(N\times Q)} \cdot B_{(Q \times s)} \] (12)

The respective canonical correlation coefficients are
\[ r = [r_1, r_2, \cdots, r_s], \quad r_1 > r_2 > \cdots > r_s \]

where $r_m$ is the correlation between $m^{th}$ columns of $U$ and $V$. The presence of the footstep signal increases correlation between both modalities. Thus, canonical correlation coefficients under $H_1$ denoted by $r^1$ are greater than that under $H_0$ denoted by $r^0$. Figure 3 shows an example where all possible correlation coefficients are plotted under both hypotheses.

Now, given the copula function and the marginals $f(U|H_i)$ and $f(V|H_i)$ under each hypothesis (i = 0, 1), one can construct the LRT detector as,
\[ \log \left( f(U, V) \right) = \log \left( \frac{f(U|H_1)}{f(U|H_0)} \right) + \log \left( \frac{f(V|H_1)}{f(V|H_0)} \right) \]
\[ \log \left( \frac{f(U|H_1)}{f(U|H_0)} \right) + \log \left( \frac{f(V|H_1)}{f(V|H_0)} \right) \]
\[ = \log \Lambda_A + \log \Lambda_S + \log \Lambda_{A,S} \] (13)

It is interesting to note the form of (13). The first two terms are strategies employed by detectors of single modality. The third term accounts for the cross-modal dependence and correlation information.
Now, each column of \( U \) and \( V \) are a weighted sum of random variables (that are independent but not necessarily identically distributed) and we approximate them using the Gaussian distribution by invoking the Central Limit Theorem (CLT). Assuming independence between all the columns of the CCA-variates \( U \) and \( V \), \( \log \Lambda_A = 0 \) and \( \log \Lambda_S = 0 \) as the CCA normalizes the mean and variance to zero and one respectively under both hypotheses. Thus, the CCA-Copula test becomes solely a correlation based test and operates in a space where this inter-modal dependence aspect is maximized. A higher correlation indicates the presence of the signal. Thus,

\[
\log (T(U, V)) = \log \Lambda_{A,S} = \log \left( \frac{e(F(U), F(V)|H_1)}{e(F(U), F(V)|H_0)} \right)
\]

(14)

B. Construction of the Copula Function

In this work, we use a Gaussian copula [13] which can be given as

\[
c_N(\Phi(y_1), \Phi(y_2), \ldots \Phi(y_n)) = \frac{\phi(y_1, y_2, \ldots y_n; R)}{\phi(y_1) \phi(y_2) \ldots \phi(y_n)}
\]

(15)

where \( \Phi \) and \( \phi \) denote the standard normal distribution and density respectively and \( R \) is the correlation matrix. A multivariate density \( f(x_1, x_2, \ldots x_n) \) with arbitrary marginals \( F(x_1), \ldots F(x_n) \) can now be obtained by making the transformation \( Y_i = \Phi^{-1}[F(X_i)] \) for each \( y_i \).

Dependence between the variables \( x_1, x_2, \ldots x_n \) is characterized by a nonparametric measure of association such as Kendall’s \( \tau \). This is because linear correlation measures are not preserved when the variables undergo nonlinear transformations. \( \tau \), however, remains invariant under monotone one-to-one transformations (always true for continuous variables) and hence is more appropriate as a measure. Also, a simple relationship between \( \tau \) and linear correlation elements of \( R \) exists and is given as

\[
\rho_{ij} = \sin \left( \frac{\pi \tau_{ij}}{2} \right), \quad i, j = 1, \ldots n
\]

(16)

We assume that \( R \) under \( H_0 \) is known or we have access to the noise-only data for its estimation and construct the CCA-Copula test in (14) using the procedure described above.

IV. EXPERIMENTS AND RESULTS

We evaluate the performance of the acoustic, seismic and CCA-Copula detectors on real footstep signals collected from experiments carried out at the U.S. Army Research Laboratory. Experiments were conducted in the basement of a building and the sensors were placed along a long hallway. Several sensors of different modalities were used. In this work, we have used data obtained from three acoustic-seismic sensor pairs. Acoustic and Seismic data were collected at 16384 Hz and 8192 Hz respectively and are then resampled at an equal rate of 6 kHz. Each detector collects data for one second before making its decision. A Hamming window of length 20 ms with 50% overlap is used alongwith 128-point fast Fourier transform (FFT) to obtain the STFT matrices. Nineteen acoustic bands \( (P = 19, 187 - 1013 \text{ Hz}) \) and eleven seismic bands \( (Q = 11, 0 - 468 \text{ Hz}) \) are used for detection.

The first experiment was designed to characterize background noise. Sensors monitored the hallway in the absence of any human activity for 120 seconds. The data collected is a representative of background noise which is used to estimate \( \alpha_f^2 \) in (10) and \( e(F(U), F(V)|H_0) \) in (14). In the second experiment, sensors collected the footstep signals when one person walked from one point (say A) to the destination (say B) and back. Different styles of walking were considered. Here we show signal detection results when a person (a) walked at a normal pace (total duration of approximately 15 seconds), (b) walked at a relatively brisk pace (total duration of approximately 12 seconds) and (c) walked at a slow or soft pace from point A to B and back (total duration of approximately 77

---

Fig. 4. Detector performance for the normal walking style. A person walked from point A to point B and back in approximately 15 secs.

Fig. 5. Detector performance for a brisk walk. A person walked from point A to point B and back in approximately 12 secs.

Fig. 6. Detector performance for the stealthy walking style. A person walked from point A to point B and back in approximately 77 secs.
individual modalities. The CCA-Copula detector is based entirely on information from the first experiment to data under both hypotheses by randomizing the starting point of the 120 second noise length.

We show in Fig. 4-(6), ROC curves for different walking styles when the noise power in the acoustic sensor is twice of the noise collected from the experiments (reference condition); i.e. SNR = -3 dB re: SNR ref. The performance of the CCA-Copula detector is better than the detectors exploiting individual modalities. The CCA-Copula detector is based entirely on the acoustic-seismic linear dependence as the energy information gets normalized due to CCA. This correlation information is more robust especially in the low SNR situations. This is because, at low SNR values, the energy (variance) information in the individual modalities becomes unreliable when compared to the cross-modal dependence information and CCA helps in transforming the data to a domain where this dependence is maximized.

Another positive aspect of the CCA normalization is that the CCA-Copula detector becomes independent of the marginal PDFs as we can use empirical CDFs estimated from the data collected in the decision window in (13). This eliminates the need to make any model assumptions or use training data to estimate PDFs for the acoustic and seismic data.

V. CONCLUSION

In this work, we have shown that CCA and copula theory can be used together to derive parametric tests for joint processing of heterogeneous data. CCA transforms the data to a space where the linear correlation between the multimodal signals is the highest. Copula theory is then used to derive the joint distribution of the obtained heterogenous canonical variates. In our example where the tests are confined to the use of only second-order characteristics, the CCA-copula detector can be based on empirical CDFs. This eliminates the need to make model assumptions and the test is now parameterized only on the underlying copula function. Future work will focus on the selection of copula functions for the task at hand and the inclusion of other modalities.

ACKNOWLEDGEMENTS

This work was supported by Army Research Laboratory and was accomplished under Cooperative Agreement No. W911NF-07-2-0007. It was also supported in part by ARO grant W911NF-06-1-0250.

REFERENCES


