AN ACCUMULATIVE FUSION ARCHITECTURE FOR DISCRIMINATING PEOPLE AND VEHICLES USING ACOUSTIC AND SEISMIC SIGNALS

Kyunghun Lee*† *Benjamin S. Riggan*† *Shuvra S. Bhattacharyya**‡
*Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, USA
†U.S. Army Research Laboratory, Adelphi, MD, USA
‡Department of Pervasive Computing, Tampere University of Technology, Tampere, Finland
{leekh3, ssb}@umd.edu {benjamin.s.riggan.civ}@mail.mil

ABSTRACT

In this paper, we develop new multiclass classification algorithms for detecting people and vehicles by fusing data from a multimodal, unattended ground sensor node. The specific types of sensors that we apply in this work are acoustic and seismic sensors. We investigate two alternative approaches to multiclass classification in this context — the first is based on applying Dempster-Shafer Theory to perform score-level fusion, and the second involves the accumulation of local similarity evidences derived from a feature-level fusion model that combines both modalities. We experiment with the proposed algorithms using different datasets obtained from acoustic and seismic sensors in various outdoor environments, and evaluate the performance of the two algorithms in terms of receiver operating characteristic and classification accuracy. Our results demonstrate overall superiority of the proposed new feature-level fusion approach for multiclass discrimination among people, vehicles and noise.

Index Terms— Sensor fusion, multiclass classification, target detection, tracking.

1. INTRODUCTION

Detection and classification of people and vehicles in outdoor environments is important in various applications related to defense, border patrol, and surveillance. For example, such capabilities help to guard specific regions against enemy intrusion and attack, and to protect borders between countries. In these applications, acoustic and seismic sensors are frequently employed because of their power efficiency and reduced computational requirements compared to other sensing modalities, such as image-based sensing.

Signals from acoustic and seismic sensors have different spectral characteristics in the presence of people and vehicles. This diversity in sensor response provides potential for greater accuracy when signals from both modalities are fused as opposed to solutions that employ only acoustic or only seismic sensors. Voices of people typically generate acoustic signals in the range of 200–800 Hz, while footsteps of people generate seismic signals in the range of 1.9–2.79 Hz [1]. For vehicles, Altmann [2] analyzes spectral characteristics of sets of signals collected from acoustic and seismic sensors. This analysis reveals similarities and differences in signal characteristics between the two modalities along with their influences from factors that include the engine rotation rate, number of engine cylinders, vehicle speed, and track element length (for tracked vehicles).

In this paper, we investigate fusion algorithms for multiclass classification among people, vehicles, and noise (the absence of people or vehicles) using signals from acoustic and seismic sensors. We develop and comparatively evaluate two different multiclass algorithms, a score-level fusion algorithm that is based on Dempster-Shafer Theory (DST), and an accumulative algorithm that exploits feature-level fusion. Through an extensive experimental comparison, we demonstrate that our feature-level fusion algorithm achieves significantly better classification performance compared to the DST-based approach.

A distinguishing aspect of our work is our focus on fusion techniques for multiclass classification using both acoustic and seismic signals. This complements related prior work that has investigated binary classification using acoustic and seismic signal processing, but has not addressed multiclass classification problems (e.g., see [3]). Also, previous work on multiclass classifiers for people, vehicles, and noise (e.g., see [4, 5] has emphasized use of acoustic signals. In contrast

This paper has been accepted for publication in the Proceedings of the 2017 International Conference on Acoustics, Speech, and Signal Processing. The official/final version of the paper will be published on http://ieeexplore.ieee.org/Xplore.

to these works, this paper contributes fusion techniques for classification using both acoustic and seismic signals.

2. RELATED WORK

Various algorithms can be applied naturally to multiclass classification problems. These include k-nearest neighbor [6], decision trees [7,8], neural networks [9], and naive Bayes classifiers [10]. Other algorithms convert a multiclass problem into a set of binary classification problems, which are then solved using more powerful binary classifiers. The techniques that we develop belong to this second class of algorithms. We decompose our targeted multiclass classification problem into three binary classification problems — noise vs. person, noise vs. vehicle, and person vs. vehicle.

A fusion architecture for distinguishing between people and animals using different ultrasonic, seismic and passive infrared sensors is proposed in [3]. In this work, the decisions of different binary classifiers are fused to detect targets (people/animals), and to distinguish between the people and animal classes whenever a target is detected. Our work differs from this work in that we incorporate a feature-level fusion approach; we address multiclass classification among noise, people, and vehicle classes; and we employ acoustic and seismic sensor types.

Dempster-Shafer Theory (DST) [11] is a common approach used for late fusion, where information from multiple classifiers are combined to produce a single output. For example, Lee et al. [12] apply DST to integrate decisions of classification and detection, and demonstrate that this integration improves the performance of both classification and detection. Wu et al. [13] propose general methods for fusing the signals from multiple sensors to perform binary classification tasks.

Accumulative methods, like the Hough transform [14], have demonstrated excellent performance in a wide range of pattern recognition problems, including image registration [15] and biometrics [16]. The methods used in [15, 16] accumulate local similarity evidences (i.e. probabilities), which are provided by explicitly estimating the probability density function (pdf) over the feature space. The disadvantages of explicitly computing a pdf are efficiency and scalability.

As discussed in Section 1, the key distinguishing aspect of our work in this paper compared to related work in the literature is our joint consideration of seismic signal processing, acoustic signal processing, and multiclass classification for border patrol and related sensor network applications. Additionally, we propose an accumulative fusion framework where the pdf is learned implicitly through an SVM.

3. FUSION FRAMEWORK

In this section, we propose two fusion algorithms for multiclass classification using signals from an acoustic-seismic node. The first is an adaptation to score-level fusion using DST for multiclass classification. We view this approach as a baseline in our experiments to assess our second approach, which is the main fusion approach that is presented in this paper. This second approach involves the accumulation of similarity evidences derived from a local feature-level fusion model. We refer to this second approach as Accumulation of Local Feature-level Fusion Scores (ALFFS).

3.1. Cepstral Analysis and SVM Classification

For both acoustic and seismic signals in the baseline (DST-based) and ALFFS approaches, we employ cepstral analysis for feature extraction [17]. We extract cepstral coefficients using the feature extraction method described in [4, 18]. In cepstral analysis, DC components are removed, low order coefficients characterize the slow spectrum variation, and higher order coefficients characterize the fundamental frequency.

For each sensing modality, we select the first 50 cepstral coefficients for training and testing. We apply SVM classifiers [19, 20] with polynomial kernels for binary classification using the extracted features for each modality. The integration of multimodal features and SVM classifiers in the baseline and ALFFS fusion architectures is illustrated in Figure 1(a) and Figure 1(b), respectively. In Section 3.2 and Section 3.3, we elaborate on the design of these two alternative fusion architectures.

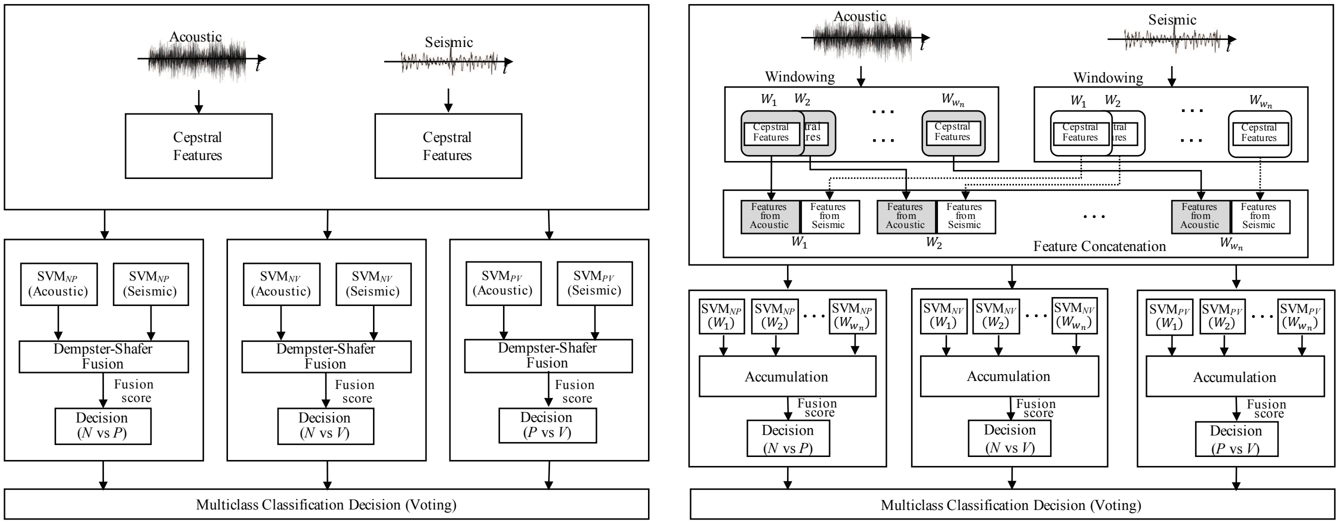
In Figure 1 and throughout the remainder of this paper, we abbreviate "noise", "person", and "vehicle" — the three available decision classes — by N, P, and V, respectively. We denote the set of all available decision classes as $\Delta = \{N, P, V\}$.

3.2. Baseline Fusion Architecture

In our baseline fusion architecture, we adapt score-level fusion with DST to perform multiclass classification. For each *distinct pair of decision classes* (*DPDC*), we employ DST fusion without weights as in [13]. Specifically, suppose that we have a pair of binary SVM classifiers $Z[\rho] = \{C_\alpha[\rho], C_\sigma[\rho]\}$ that discriminate between the two elements of a DPDC $\rho = \{X, Y\} \subset \Delta$ based on signals of type α and σ, where α and σ represent the acoustic and seismic sensing modalities, respectively. Then based on DST, the score $S(Z, A)$ associated with classifier pair Z and decision class $A \in \rho$ can be expressed as

$$S(Z,A) = \frac{Belief_A}{Belief_{\neg A}} = \frac{\sum\limits_{E_{\rho,\alpha} \cap E_{\rho,\sigma} = A} E_{\rho,\alpha} E_{\rho,\sigma}}{\sum\limits_{E_{\rho,\alpha} \cap E_{\rho,\sigma} = \neg A} E_{\rho,\alpha} E_{\rho,\sigma}}. \quad (1)$$

Here, $\neg A$ is the element of ρ other than A. Additionally, $E_{\rho,x}$ denotes the evidence associated with ρ that is derived from sensing modality $x \in \{\alpha, \sigma\}$. The value of $E_{\rho,x}$ for



(a) Fusion architecture using DST.

(b) Fusion architecture using ALFFS.

Fig. 1: Illustration of the baseline (a) and ALFFS (b) fusion architectures.

each modality x can be derived from the scores of the two associated SVM classifiers.

Equation 1 can be viewed as a standard DST-based approach to binary classification (for discrimination between A and $\neg A$) using SVM-based binary classifier subsystems. We extend this approach to multiclass classification by instantiating 3 different pairs of SVM classifiers $\{Z[\rho] \mid \rho \in \{\{N,P\},\{N,V\},\{P,V\}\}$, where each of these classifier pairs is connected to a fusion subsystem that operates based on Equation 1. The results from these 3 fusion subsystems are then combined using voting, as illustrated in Figure 1(a). Similar to [21, 22], the voting method chooses the class that is classified most frequently by the three SVMs.

3.3. ALFFS

Our ALFFS approach is motivated by the significant differences in spectral characteristics between acoustic and seismic signals. To systematically incorporate these different characteristics into the multiclass classification process, ALFFS applies concatenated features that are derived from both acoustic and seismic inputs.

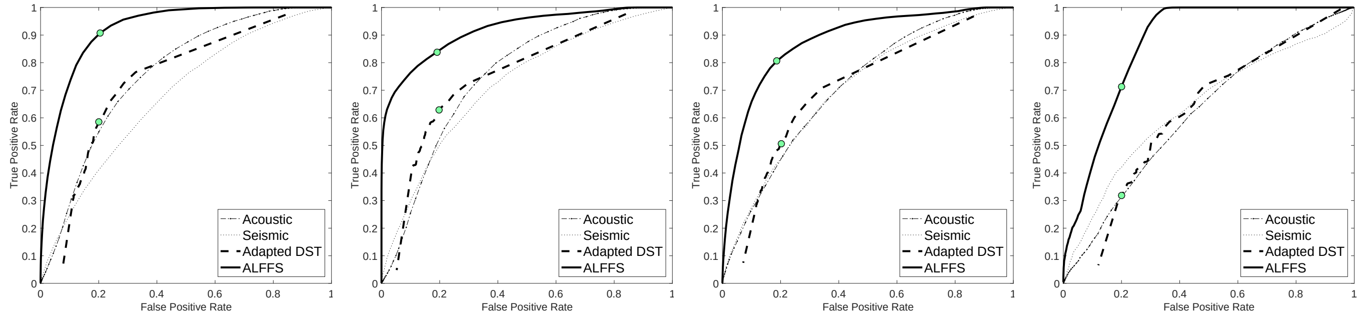
Algorithm 1 presents a pseudocode representation of the ALFFS approach. In the signal processing system represented in Algorithm 1, the subscripts α and σ are used to represent correspondence with acoustic and seismic signals, respectively, as in Section 3.2. The input to the system consists of data frames (segments of contiguous signal samples) Γ_α and Γ_σ, and two parameters w_n and w_r, which respectively specify the number of windows and the ratio of inter-window overlap that are to be employed when processing the input frames. The two signals Γ_α and Γ_σ are corresponding acoustic and seismic signals, meaning the two modalities observe the same activity.

Algorithm 1: A pseudocode representation of the ALFFS approach.

```
   Input  : $\Gamma_\alpha, \Gamma_\sigma, w_n, w_r$
   Output: $Class$
1  $D_\alpha(1), D_\alpha(2), \ldots, D_\alpha(w_n) \leftarrow Window(\Gamma_\alpha, w_n, w_r)$
2  $D_\sigma(1), D_\sigma(2), \ldots, D_\sigma(w_n) \leftarrow Window(\Gamma_\sigma, w_n, w_r)$
3  for $i = 1$ to $w_n$ do
4      $F_\alpha(i) \leftarrow f_{cepstral}(D_\alpha(i))$
5      $F_\sigma(i) \leftarrow f_{cepstral}(D_\sigma(i))$
6      $F_{concat}(i) \leftarrow f_{concat}(F_\alpha(i), F_\sigma(i))$
7  end
8  for $p \in \{\{N,P\},\{N,V\},\{P,V\}\}$ do
9      for $j = 1$ to $w_n$ do
10         $Score_p(j) \leftarrow SVM_p(F_{concat}(j))$
11     end
12     $\kappa(p) \leftarrow Score_p(1) + Score_p(2) + \ldots + Score_p(w_n)$
13     $R(p) \leftarrow f_{dec,p}(\kappa(p))$
14 end
15 $Class \leftarrow f_{voting}(R(\{N,P\}), R(\{N,V\}), R(\{P,V\}))$
```

In the first two steps of Algorithm 1, a windowing function $Window$ decomposes the input data frames into overlapping windows consisting of w_n samples each, where the ratio of overlap is determined by the parameter w_r. The function $f_{cepstral}$ is a function that returns cepstral features for a given window of signal samples. The concatenation of acoustic and seismic features for each window is performed by the function f_{concat}.

The outer `for` loop (line 8) iterates through all relevant DPDCs. For each DPDC p and window index j, the algo-



(a) Test ID #1.

(b) Test ID #2.

(c) Test ID #3.

(d) Test ID #4.

Fig. 2: ROC curves for multiclass classification.

rithm computes a binary classification score $Score_p(j)$ by applying an SVM classifier SVM_p that is trained specifically for DPDC p. Line 12 then accumulates all of the scores for the given DPDC p to provide a single composite score $\kappa(p)$ across all windows and both sensing modalities. This composite score is then thresholded by the decision function $f_{dec,p}$ to produce the decision $R(p)$ associated with DPDC p. In our experiments, we use a common threshold of 0 for all three decision functions $\{f_{dec,p}\}$.

The three decisions $\{R(x)\}$ are then operated on using a voting process, represented by the function f_{voting}, to produce the final multiclass classification result $Class$. We use the same voting process here as in the adapted DST approach of Section 3.2.

4. EXPERIMENTS

In this section, we present an experimental evaluation of the Adapted DST and ALFFS approaches, which were introduced in Section 3.2 and Section 3.3, respectively. In our evaluation, we employ 4 different datasets, which we refer to as Datasets #1–#4. Datasets #1–#3 were collected on Spesutie Island at the Aberdeen Proving Grounds in Maryland, USA during July 28–30, 2015. These three datasets were collected from different sensors installed in different locations and at different times of day. Further details about Datasets #1–#3 can be found in [23]. Dataset #4 was collected at the US Army Research Laboratory, Adelphi, Maryland, USA on September 16, 2013. Datasets #1–#3 were collected from soil, while Dataset #4 was collected from asphalt. Each dataset contains 1000 data frames, where each frame contains 6 seconds of acoustic and seismic data sampled at 4096Hz.

For training and testing, we used input data segments (IDSs) that each consist of 500 contiguous data frames from one of the four datasets. For training, we randomly extracted 50 different IDSs from Dataset #1 using the MATLAB `crossvalind` function. Similarly, for testing, we used `crossvalind` to extract 50 different IDS from each of the four available datasets. Thus, we employed 50 IDSs for training, and 200 IDSs for testing. We refer to the set of 50 IDS used for testing that we extracted from each Dataset #X as "Test ID #X". For ALFFS, we used $w_n = 50$ and $w_r = 0.4$.

Test ID	Acoustic	Seismic	DST	ALFFS
#1	64.5731	59.3908	66.2605	86.0842
#2	61.3988	57.8397	60.3166	73.8357
#3	56.9138	51.7074	57.2305	73.3988
#4	53.0541	67.2345	61.3026	76.9379
avg.	58.9850	59.0431	61.2776	77.5642

Table 1: Accuracy comparison (%).

To evaluate classification performance, we compared the Adapted DST and ALFFS approaches in terms of their measured ROC curves and accuracy levels. Among the different ways to compute ROC curves for multiclass problems, we employed the method discussed in [24], which is suitable for multiclass classifiers that are composed of binary classifiers. In this method, the multiclass ROC curve is computed by averaging the ROC curves across the corresponding set of pairwise (1-to-1) classifiers. Figure 2 and Table 1 show the measured ROC curves and accuracy levels, respectively. From these results, we see that the Adapted DST approach shows no significant performance improvement compared to the single-modality classifiers. In contrast, ALFFS exhibits significant improvements compared to the single-modality classifiers, as well as the Adapted DST approach. Specifically, ALFFS achives 0.9076, 0.8389, 0.8059, and 0.7120 true positive rate when operating at 0.2 false positive rate for Test ID #1-#4, respectively. Whereas, the baseline approach achieves 0.5858, 0.6278, 0.5070, and 0.3188 at the same false positive rate. Thus, ALFFS achieves fewer false alarms, even when only using a single seismic and single acoustic source. The results in Table 1 show that ALFFS achieves an absoulte improvement of 16.3% (relative improvement of 26.6%) in accuracy compared to the baseline fusion on average.

5. CONCLUSION

In this paper, we have introduced an algorithm, called Accumulation of Local Feature-level Fusion Scores (ALFFS), for multiclass classification among people, vehicles, and noise using a single unattended ground sensor node. ALFFS operates by extracting cepstral features, applying feature-level fusion, and applying a bank of support vector machines across sets of concatenated features that are extracted from overlapping windows of the multimodal input signals. We have also introduced an adaptation to our targeted multiclass classification problem of sensor fusion based on Dempster-Shafer Theory (DST). Through extensive experiments, we have demonstrated that ALFFS achieves an average of 16.3% (26.6%) absolute (relative) improvement over the adapted DST approach. Moreover, ALFFS achieves a significant reduction in the number of false alarms compared to the adapted DST approach (and the individual modalites).

6. REFERENCES

[1] T. Damarla, A. Mehmood, and J. Sabatier, "Detection of people and animals using non-imaging sensors," in *Proceedings of the International Conference on Information Fusion*, 2011, pp. 1–8.

[2] J. Altmann, "Acoustic and seismic signals of heavy military vehicles for co-operative verification," *Journal of Sound and Vibration*, vol. 273, no. 4–5, pp. 713–740, 2004.

[3] T. Damarla and L. M. Kaplan, "A fusion architecture for tracking a group of people using a distributed sensor network," in *Proceedings of the International Conference on Information Fusion*, 2013, pp. 1776–1783.

[4] H. Ben Salem, T. Damarla, K. Sudusinghe, W. Stechele, and S. S. Bhattacharyya, "Adaptive tracking of people and vehicles using mobile platforms," *EURASIP Journal on Advances in Signal Processing*, vol. 2016, no. 1, pp. 1–12, 2016.

[5] K. Lee, H. Ben Salem, T. Damarla, W. Stechele, and S. S. Bhattacharyya, "Prototyping real-time tracking systems on mobile devices," in *Proceedings of the ACM International Conference on Computing Frontiers*, Como, Italy, May 2016, pp. 301–308, Invited paper.

[6] S. D. Bay, "Combining nearest neighbor classifiers through multiple feature subsets," in *Proceedings of the International Conference on Machine Learning*, 1998, pp. 37–45.

[7] C. J. Stone R. A. Olshen L. Breiman, J. Friedman, *Classification and regression trees*, CRC press, 1984.

[8] J. R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.

[9] C. M. Bishop, *Neural networks for pattern recognition*, Oxford university press, 1995.

[10] I. Rish, "An empirical study of the naive Bayes classifier," in *IJCAI 2001 workshop on empirical methods in artificial intelligence*. IBM New York, 2001, vol. 3, pp. 41–46.

[11] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *The annals of mathematical statistics*, pp. 325–339, 1967.

[12] H. Lee, H. Kwon, R. M. Robinson, W. D. Nothwang, and A. M. Marathe, "Dynamic belief fusion for object detection," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2016, pp. 1–9.

[13] H. Wu, M. Siegel, R. Stiefelhagen, and J. Yang, "Sensor fusion using Dempster-Shafer theory [for context-aware HCI]," in *Instrumentation and Measurement Technology Conference, 2002. IMTC/2002. Proceedings of the 19th IEEE*, 2002, vol. 1, pp. 7–12 vol.1.

[14] D. H. Ballard, "Generalizing the Hough transform to detect arbitrary shapes," *Pattern Recognition*, vol. 13, no. 2, pp. 111–122, 1981.

[15] B. S. Riggan, W. E. Snyder, X. Wang, and J. Feng, "A human factors study of graphical passwords using biometrics," in *Proceedings of the German Conference on Pattern Recognition*, 2014, pp. 464–475.

[16] K. Krish, S. Heinrich, W. E. Snyder, H. Cakir, and S. Khorram, "Global registration of overlapping images using accumulative image features," *Pattern Recognition Letters*, vol. 31, no. 2, pp. 112–118, 2010.

[17] A. Martin, D. Charlet, and L. Mauuary, "Robust speech/non-speech detection using LDA applied to MFCC," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 2001, pp. 237–240.

[18] B. M. Smith, P. Chattopadhyay, A. Ray, S. Phoha, and T. Damarla, "Performance robustness of feature extraction for target detection & classification," in *2014 American Control Conference*, June 2014, pp. 3814–3819.

[19] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[20] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[21] S. Knerr, L. Personnaz, and G. Dreyfus, "Single-layer learning revisited: a stepwise procedure for building and training a neural network," in *Neurocomputing*, pp. 41–50. Springer, 1990.

[22] J. H. Friedman, "Another approach to polychotomous classification," Tech. Rep., Stanford University, October 1996.

[23] S. M. Nabritt, T. Damarla, and G. Chatters, "Personnel and vehicle data collection at aberdeen proving ground (APG) and its distribution for research," Tech. Rep., DTIC Document, 2015.

[24] D. J. Hand and R. J. Till, "A simple generalisation of the area under the ROC curve for multiple class classification problems," *Machine learning*, vol. 45, no. 2, pp. 171–186, 2001.