Decision Level Fusion: An Event Driven Approach

Siddharth Roheda
Department of Electrical and Computer Engineering
North Carolina State University
Raleigh, NC, USA
sroheda@ncsu.edu

Hamid Krim
Department of Electrical and Computer Engineering
North Carolina State University
Raleigh, NC, USA
ahk@ncsu.edu

Zhi-Quan Luo
Department of Electrical and Computer Engineering
University of Minnesota
Minneapolis, MN, USA
luozq@ece.umn.edu

Tianfu Wu
Department of Electrical and Computer Engineering
North Carolina State University
Raleigh, NC, USA
tianfu_wu@ncsu.edu

Abstract—This paper presents a technique that combines the occurrence of certain events, as observed by different sensors, in order to detect and classify objects. This technique explores the extent of dependence between features being observed by the sensors, and generates more informed probability distributions over the events. Provided some additional information about the features of the object, this fusion technique can outperform other existing decision level fusion approaches that may not take into account the relationship between different features.

Index Terms—Sensor Fusion, Decision Level Fusion, Event based Classification, Coupling

I. INTRODUCTION

One often resorts to fusing data from more than one sensors to successfully carry out target detection and classification. Data fusion has, over the years been recognized to distinguish three levels of fusion, namely data level, feature level, and decision level. Data level fusion generally processes raw data generated by each sensor, and performs the fusion of information according to some criterion before proceeding to inference. Feature level fusion, on the other hand, first glean features from raw data (e.g., transformed data) from diverse sensors, to subsequently coherently merge them for inference. In decision level sensor fusion, each sensor reaches an individual decision, prior to optimal combination of the decisions to yield a more informed decision, e.g. target classification.

In this paper, we present on the decision level fusion strategy, and propose a principled approach to fusion for improved inference performance. A classification decision is reached by cataloguing sets of events along with their probabilistic characterization for each sensor, and following a joint probabilistic and coherent evaluation of these events. These events are formalized for each sensor according to its description as an algebraic definition of these feature events.

II. BACKGROUND AND RELATED WORK

As noted earlier, sensor fusion has long been of interest, albeit with limited theoretical success particularly when heterogeneous data is present and for which a unified and systematic approach has remained elusive. An introduction and comprehensive survey to the area of fusion is provided in [1], [2]. Classical techniques in decision level fusion include Bayesian Inference [3] and Dempster-Shafer fusion [4]. Bayesian Inference has received a lot of attention and has shown success if prior knowledge about sensor reports is available and given. On the other hand, Dempster-Shafer Fusion was proposed to specifically lift such a restriction on information prior, at a cost of a substantial increase in computational complexity. Some applications of Dempster-Shafer Fusion can be found in [5], where, LIDAR data is combined with Multispectral imagery, and in [6], where, multi-sensor information like vibration, sound, pressure, and temperature is fused to detect engine faults. Furthermore, [7] provides a detailed comparison of Bayesian Inference and Dempster-Shafer Theory. In [8], a two-stage approach to fusion was proposed, involving knowledge-modeling, which learns from past behavior of the classifiers whose results are to be fused, and operation stage, that combines output of these classifiers based on knowledge learned in the first stage. More recent work proposes a sensor model based approach [9], where the sensor network is modeled as either that of similar or dissimilar sensors. Similar Sensor Fusion or Convex Quadratic
Fusion [9]–[11] is used when all the sensors explore the same characteristics/features of the target (for example, a set of 5 radars, looking at the same target), while Dissimilar Sensor Fusion or Analytic Center Fusion [9], [11] is alternatively used when sensors explore different characteristics/features of the target (for example, a radar and optical sensor looking at the same target). These assumptions turn out to be too restrictive, in that some sensors, albeit dissimilar, may have some common features, while offering additional features to enrich an object/target characterization. Our goal is to explore such a case, and demonstrate that a systematic and principled approach maybe designed, and the overall solution is improved on account of this enhancement.

A. Mutual Information

Mutual Information (MI) has been an important and ubiquitous metric in many applications in data science, with a goal of revealing the synergy/differential between the dependence of two random variables and their lack thereof, by way of their associated probability density functions (PDF). Specifically, consider two random variables, X and Y, with a joint probability mass functions \( p(x, y) \) and marginal probability mass functions \( p(x) \) and \( p(y) \). The Mutual Information, \( I(X; Y) \), is the relative entropy between the joint distribution, \( p(x, y) \), and the product distribution, \( p(x)p(y) \) [12] as given by

\[
I(X; Y) = \sum_{x\in X} \sum_{y\in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}
\]  

This relationship may alternatively and perhaps more revealingly, be expressed in terms of the variables’ joint entropy, their marginal entropies and their mutual information [12],

\[
H(X, Y) = H(X) + H(Y) - I(X; Y)
\]

III. PROBLEM FORMULATION

We assume throughout a set of targets/objects, \( O = \{\alpha_1, \alpha_2, \ldots, \alpha_l\} \), whose detection and/or classification are of interest. Let the \( k^{th} \) feature observed by the \( l^{th} \) sensor be \( F_{k}^l \). Then, a set of mutually exclusive events, \( \Omega_k = \{a_{k,j}\}_{j=1,\ldots,K_k} \), may be defined for the feature \( F_{k}^l \). Here, \( a_{k,j} \) is the \( j^{th} \) event for \( F_{k}^l \) and is described as, \( a_{k,j} : F_{k}^l \in [u_j, v_j), \) with \( u_j \in \mathbb{R}^+ \) and \( v_j \in \mathbb{R}^+ \), and \( v_j > u_j \). The probability report for the \( k^{th} \) feature from the \( l^{th} \) sensor is then defined as,

\[
D_k^l = \{\Omega_k, \sigma_B(\Omega_k), P_k^l\}
\]

Where, \( \sigma_B(\Omega_k) \) is the Borel sigma algebra of \( \Omega_k \), and can be thought of as the set of all the possible events that can be described over the feature. \( P_k^l \) is the set of probabilities over the events in \( \sigma_B(\Omega_k) \).

Since, we define objects as a combination of certain events occurring over different features, we will be working in the product space,

\[
\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_k \times \Omega_{k+1} \times \Omega_{k+2} \times \cdots \times \Omega_{l-1} \times \Omega_k \times \Omega_{k+2} \times \cdots \times \Omega_{L}
\]

(4)

Where, \( K_l \) is the total number of features observed by the \( l^{th} \) sensor, and \( l = 1, \ldots, L \). Further, an object will be defined as some combination of events in this product space, \( \alpha_i \in \sigma_B(\Omega) \). Given the object definitions and the probability distributions over various features, we then want to find the fused probability report over the objects, \( D_f = \{O, P_f\} \).

IV. APPROACH

The sensor reports form a set \( \{D_k^l\}_{k=1,\ldots,K_k} \), that come from different sensors, and the target describing events may also be specified over different features. Further, the definitions of specific objects are the result of algebraic operations on the event space \( \sigma_B(\Omega) \), a Sigma-algebra on the product space, \( \Omega \), with associated probability measures as noted in Section III. Thus, we must determine probability distribution on \( \sigma_B(\Omega) \).

A. Determining the joint probability distribution

This discussion is in a 2-dimensional setting, in the interest of simplicity, but can be easily extended to an arbitrary number of reports. Let the reports \( D_k^l = \{\Omega_k, \sigma_B(\Omega_k), P_k^l\} \) and \( D_j^m = \{\Omega_j, \sigma_B(\Omega_j), P_j^m\} \) be defined over the \( k^{th} \) feature from the \( l^{th} \) sensor, and the \( j^{th} \) feature from the \( m^{th} \) sensor respectively. Then, we have the product space, \( \Omega = \Omega_k \times \Omega_j \), and we wish to determine a probability distribution on \( \sigma_B(\Omega) \).

When determining the joint probabilities, it is important to account for the extent of dependence between the features: Completing independent features yield minimal mutual information, and the joint distribution with minimum mutual information should be selected; a high correlation between features, on the other hand, yields maximal mutual information, and the joint distribution with maximal mutual information should be selected. These are clearly the extreme cases of dependence, and do not address the partial correlation case. A good approximation to determine the joint probability between partially correlated features is to consider a convex combination of the joint probabilities maximizing and minimizing the mutual information. The joint probability for the occurrence of events, \( \alpha \in \sigma_B(\Omega_k) \) and \( \beta \in \sigma_B(\Omega_j) \), can then be determined as,

\[
P_{O}(\alpha, \beta) = \rho_P \max_{\alpha, \beta} I(\alpha, \beta) + (1 - \rho_P) \min I(\alpha, \beta),
\]

(5)

where, \( \rho \in [0, 1] \) is a pseudo-measure of extent of correlation between the features. \( \rho \approx 1 \) when the features are highly correlated, and \( \rho = 0 \) when features are independent of each other. \( \rho \) can be determined from the training data, by computing the correlation between the features using a measure like Pearson’s correlation or distance correlation [13].

It can readily be seen from Equation 1 that mutual information between two random variables is minimized when the joint probability distribution is selected as the multiplication of the marginal probabilities. Hence, we have,

\[
P_{\min} I(\alpha, \beta) = P_{\min}^{k}(\alpha) \ast P_{\min}^{m}(\beta).
\]

(6)

The problem of maximizing mutual information given the marginal probabilities requires some steps. For some random variables X and Y, conditioning on the marginal probability distributions of X and Y yields constant H(X) and H(Y). As may be seen from Equation 2, the maximization of Mutual Information between two random variables then becomes
equivalent to minimizing their Joint Entropy, which is known to be a concave function.

\[
P_{\text{\scriptsize 1MAX}}(m) = \min_{P_{\Omega}} \sum_{x \in \Omega_{1k}} \sum_{y \in \Omega_{1m}} -P_{\Omega}(x,y) \log P_{\Omega}(x,y)
\]

subject to:

\[
\sum_{x \in \Omega_{1k}} P_{\Omega}(x,y) = \xi_{j}, \quad P_{\Omega}(x,y) \geq 0.
\]

A greedy approach suggested for minimizing joint entropy given the marginal probabilities in [14] is used in our implementation. The main idea here is to keep large probability masses intact and not break them down into smaller chunks. The contribution of a probability mass toward the joint entropy only increases if it is divided into smaller chunks. That is, for \( p = a + b \), \(-p \log(p) \leq -a \log(a) - b \log(b)\), when \( 0 < p < 1 \) and \( a, b > 0 \). So, keeping the large probability masses from given marginal probabilities intact ensures that their contribution toward the joint entropy is minimized. As empirically demonstrated in [14], the minimal joint entropies are obtained to within 1 bit of the optimal values.

B. Determining Object Probabilities

Once the probability distribution over the product space, \( \Omega \), has been determined, the next step is to determine the object probabilities, \( P_j(o_1) \), in the fusion report, \( D_f = \{O, P_j\} \). Given the combination of events describing an object, it is pretty straightforward to determine the probability of the presence of an object using the general rules of probability. For an object, \( o \), which is to be described in terms of a combination between \( \alpha \in \Omega_{1k} \) and \( \beta \in \Omega_{1m} \), we can have either \( o = (\alpha \wedge \beta) \) or \( o = (\alpha \vee \beta) \). Where, \( \alpha \wedge \beta \) denotes the combination \( \alpha \) ’and’ \( \beta \), while \( \alpha \vee \beta \) denotes the combination \( \alpha \) ’or’ \( \beta \). Then, for the first case we have, \( P_f(o) = P_f(\alpha \wedge \beta) = P_{\Omega}(\alpha, \beta) \) and for the latter we have, \( P_f(o) = P_f(\alpha \vee \beta) = P_{\Omega}(\alpha) + P_{\Omega}(\beta) - P_{\Omega}(\alpha, \beta) \). This can be extended to any combination of more than two events.

V. EXPERIMENTS AND RESULTS

To substantiate the above approach along with the various steps, we select two different datasets.

A. Dataset 1

For the first one, we select two sensors, namely a Radar sensor and a telescopic optical sensor, for one (latter) of which we have acquired real data. Due to technical difficulty, our Radar measurements were never co-measured with the optical data. Both the sensors are ideally synchronized when observing a given target, which in our case, can be any object in outer space, such as satellite or space debris. For radar simulations we use MATLAB Simulink, and for telescopic image simulation we use some existing telescopic images collected by The Czech Technical University in Prague, and model the object distributions from those images to generate more images with required object characteristics. Each generated radar signal over one second is correlated with two telescopic images. Samples for objects with different velocities, cross-sections, ranges, and aspect-ratios are then generated. The radar signals are used to make decisions over velocity, range, and the cross-section, while the telescopic images are used to make decisions over the aspect-ratio, and displacement over time of an object in view.

1) Implementation Details: To proceed with evaluation, we generate the probability distribution over the features that we will use. Figure 2 shows a high level block diagram for implementation on the first dataset.

Let the received radar signal be, \( x(n) \), and its corresponding Fourier transform, \( X(k) = \sum_{n=0}^{N-1} x(n)e^{-i2\pi km/N} \) for each object, whose labels are distinguished with the velocity, range, and cross-section values of that object, \([v, r, cs]\). Using the training data and labels, SVM classifiers using the Radial Basis Function kernel are trained over the events of interest defined over \([v, r, cs]\), and used to determine a classification score over these events when a test sample is fed in, as shown in Figure 2. The classification probabilities are then approximated from the classification score using Platt’s technique [18]. These probabilities over the events of interest are the, \( P_k \), described earlier in Section III.

We have two telescopic images associated with 1-sec of radar return for the same object. Due to space limitations, we forego the details of target detection which is not central to our paper. The object of interest is initially detected by using target detection as discussed in [19]. Upon its detection, the probability distribution over the object’s aspect-ratio, and its displacement in the second image relative to its location in the first image is determined using the image flow technique discussed in [20].

2) Event and Object Definitions: For training and testing purposes, we define various events over the feature-sets from both the sensors. For the radar, as noted before, we use \([v, r, cs]\) and the events are defined as

\[
a_1^v: 0 \leq v \leq 10 \text{ m/s}, \quad a_2^v: 15 \text{ m/s} \leq v \leq 35 \text{ m/s},
\]

\[
a_1^r: 0 < r < 300 \text{ m}, \quad a_2^r: 300 \text{ m} < r,
\]

\[
a_1^{cs}: 0 < cs \leq 20 \text{ m}, \quad a_2^{cs}: 20 \text{ m} < cs \leq 50 \text{ m}^2.
\]

From the telescopic imaging sensor, the features, displacement and aspect ratio, \([d, AR]\) define the following events,

\[
a_1^d: 0 \leq d \leq 60 \text{ pixels}, \quad a_2^d: 90 \text{ pixels} \leq d \leq 210 \text{ pixels},
\]

\[
a_1^{AR}: 0 < AR \leq 1.5, \quad a_2^{AR}: 1.5 < AR.
\]

Furthermore, the objects for classification are defined in terms of these events as,

\[
o_1: \{a_1^v \land \{a_1^r \lor a_2^r\} \lor (a_2^v \land a_2^{cs})\}\}
\]

\[
o_2: \{a_1^v \land a_1^r \land a_2^r \land a_2^{cs} \land a_2^{AR}\}
\]

Given these events and object definitions, we wish to determine the fused report, \( D_f = \{P_f(o_1), P_f(o_2), P_f(\overline{o_1} \lor \overline{o_2})\} \). This can be thought of a classification problem with 3 classes, Class 1: Object 1, Class 2: Object 2, and Class 3: Neither Object 1 nor Object 2.
Fig. 1: ROC Curves for detection (Dataset 1) of (a): Class 1:Object 1, (b): Class 2:Object 2, and (c): Class 3:Neither Object 1 nor Object 2

The range of an event can be determined from the training data. The mean of the feature in question over the samples of the same class is computed, and a range of two times the standard deviation is taken on either side of the mean. Furthermore, the targets are defined as,

\[ o_1 \text{ (human target)} : \{a_1^w \wedge (a_1^s \vee a_1^v)\} \quad (13) \]
\[ o_2 \text{ (vehicular target)} : \{a_2^w \wedge a_2^s \wedge a_2^v\} \quad (14) \]

Given these events and object definitions, we wish to determine the fused report, \( D_f = \{P_f(o_1), P_f(o_2), P_f(\overline{o_1} \vee o_2)\} \), where, \( \{\overline{o_1} \vee o_2\} \) represents the no target case.

C. Performance Analysis

Table I, and II show the classification performance of different techniques (averaged over 10 runs of the technique) when implemented on dataset 1 and 2 respectively.

<table>
<thead>
<tr>
<th>Method</th>
<th>Average Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radar</td>
<td>89.47%</td>
</tr>
<tr>
<td>Telescopic Imaging</td>
<td>81.31%</td>
</tr>
<tr>
<td>Feature Concatenation</td>
<td>85.93%</td>
</tr>
<tr>
<td>Convex Quadratic Fusion (Similar Sensor Fusion)</td>
<td>86.07%</td>
</tr>
<tr>
<td>Analytic Center Fusion (Dissimilar Sensor Fusion)</td>
<td>88.61%</td>
</tr>
<tr>
<td>Dempster-Shafer Fusion</td>
<td>87.18%</td>
</tr>
<tr>
<td>Proposed Technique</td>
<td>90.36%</td>
</tr>
</tbody>
</table>

Classification accuracy is often not the best measure to quantify performance, particularly in cases where different classes have different numbers of samples, which is the case here. A better way to compare performance is to look at the Receiver Operating Characteristic (ROC) curves. Fig. 1 and 4 show the ROC curves for classification in case of datasets 1 and 2 respectively. It can be seen from the ROC curves...
TABLE II: Performance Comparison for the Second Dataset

<table>
<thead>
<tr>
<th>Method</th>
<th>Average Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seismic Sensor</td>
<td>85.41%</td>
</tr>
<tr>
<td>Acoustic Sensor</td>
<td>67.62%</td>
</tr>
<tr>
<td>Feature Concatenation</td>
<td>81.63%</td>
</tr>
<tr>
<td>Convex Quadratic Fusion (Similar Sensor Fusion)</td>
<td>86.09%</td>
</tr>
<tr>
<td>Analytic Center Fusion (Dissimilar Sensor Fusion)</td>
<td>89.96%</td>
</tr>
<tr>
<td>Dempster-Shafer Fusion</td>
<td>87.93%</td>
</tr>
<tr>
<td>Proposed Technique</td>
<td>92.04%</td>
</tr>
</tbody>
</table>

We propose a novel sensor fusion technique at the decision level that looks at targets as combinations of events over the features that describe it, while also considering the extent of correlation between different features. Experiments on various datasets showed that the proposed technique can outperform existing fusion techniques on the decision level.

VI. CONCLUSION

The ROC Curves for detection (second dataset) of (a): Class 1: Human Target, (b): Class 2: Vehicular Target, and (c): Class 3: No Target (for dataset 1) in Fig. 1 that other techniques show limited performance in correct classification of objects from class 2 due to the low number of samples for class 2 in comparison to those in class 1 and class 3. This causes the classifier to bias toward selecting class 1 or class 3, even when the sample is from class 2 in order to achieve high classification accuracy. But, our technique trains over occurrence of events rather than the object itself, hence does not face this issue. Improvement in performance is also seen for Dataset 2 (Fig. 4). In particular, detection of human targets is significantly improved, by taking ‘or’ between noise level event and weight event, which reduces misclassification due to noise of winds.

REFERENCES
