GRAPH-BASED MULTI-SENSOR FUSION FOR ACOUSTIC SIGNAL CLASSIFICATION

Umamahesh Srinivas†, Nasser M. Nasrabadi*, Vishal Monga†

†Department of Electrical Engineering, Pennsylvania State University, University Park, PA
*U.S. Army Research Laboratory, Adelphi, MD

Motivation

- Desire for a robust transient acoustic signal classifier
  - Real-world acquisition distortions
  - Signal non-stationarity
- **Opportunity:**
  - Availability of multiple measurements of the same acoustic event via co-located sensors
  - Inherent conditional correlations among such measurements
- **Challenge:** Principled data/feature fusion for robust classification

Contributions

Two-stage classification framework:

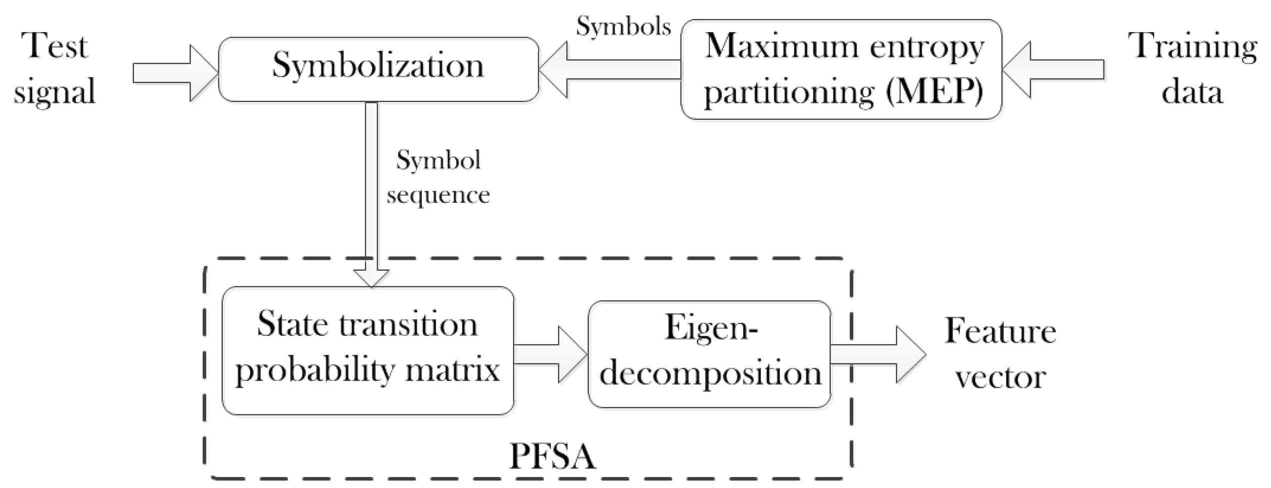
1. **Feature extraction**
   - Cepstral features
   - Symbolic dynamic filtering-based (SDF) features
2. **Classifier**
   - Probabilistic graphical models $\rightarrow$ boosting on discriminative trees

Feature representations

1. **Cepstral features:** Given signal $x[n]$,

$$P_c(\tau) = \left| \mathcal{F}\left(\log_{10}\left(|\mathcal{F}(x[n])|^2 \right) \right) \right|^2$$

   - Intuition: Capture the rate of change of information content across different frequency bands
   - Sensitive to noise
2. **Symbolic dynamic filtering-based features**
   - Motivation: Robustness exhibited in time-series applications such as anomaly detection [Rajagopalan and Ray, 2006]



Key steps:

- Amplitude quantization: $\mathcal{A} = \{a_1, a_2, \ldots, a_{|\mathcal{A}|}\}$
- Probabilistic finite state automata (PFSA)
  - Transition probabilities:

$$P(a_i|a_j) = \frac{N(a_j, a_i)}{\sum_{k=1}^{|\mathcal{A}|} N(a_j, a_k)}, \ \forall\, a_i, a_j \in \mathcal{A}$$

  - State transition matrix:

$$\mathcal{P} = \begin{bmatrix} P(a_1 \mid a_1) & \cdots & P(a_{|\mathcal{A}|} \mid a_1) \\ \vdots & \ddots & \vdots \\ P(a_1 \mid a_{|\mathcal{A}|}) & \cdots & P(a_{|\mathcal{A}|} \mid a_{|\mathcal{A}|}) \end{bmatrix}$$

  - Feature vector: eigenvector corresponding to unity eigenvalue of $\mathcal{P}$

Probabilistic graphical models: Background

- **Graph** $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ defined by a set of nodes $\mathcal{V} = \{1, \ldots, n\}$, and a set of edges $\mathcal{E} \subset \binom{\mathcal{V}}{2}$
- **Graphical model:** Random vector defined on a graph; nodes represent random variables, edges reveal conditional dependencies
- **Generative learning** [Chow and Liu, 1968]
  - Learn a single graph to minimize approximation error,

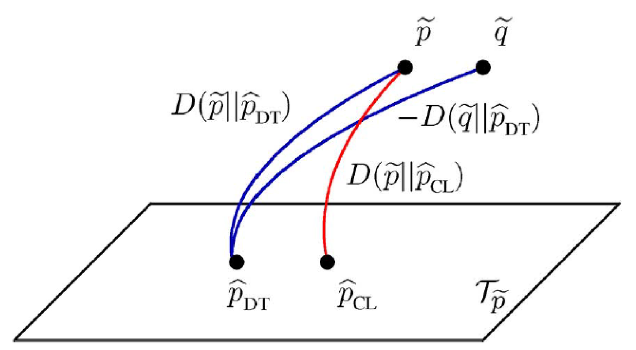
$$\text{Given } p, \text{ find } \hat{p} = \arg \min_{p_t \text{ is a tree}} D(p || p_t)$$

- **Discriminative learning** [Tan *et al.*, 2010]
  - Simultaneously learn a pair of graphs to minimize classification error

Tree-approximate J-divergence of $\hat{p}, \hat{q}$ w.r.t. p, q:

$$\widehat{J}(\hat{p}, \hat{q}; p, q) := \int_{\Omega \subset \mathcal{X}^n} (p(\mathbf{x}) - q(\mathbf{x})) \log \left(\frac{\hat{p}(\mathbf{x})}{\hat{q}(\mathbf{x})} \right) d\mathbf{x}$$

$$(\hat{p}, \hat{q}) = \arg \max_{\hat{p} \in \mathcal{T}_{\hat{p}}, \hat{q} \in \mathcal{T}_{\hat{q}}} \widehat{J}(\hat{p}, \hat{q}; p, q)$$



Multi-Sensor Graphical Model (MSGM) framework

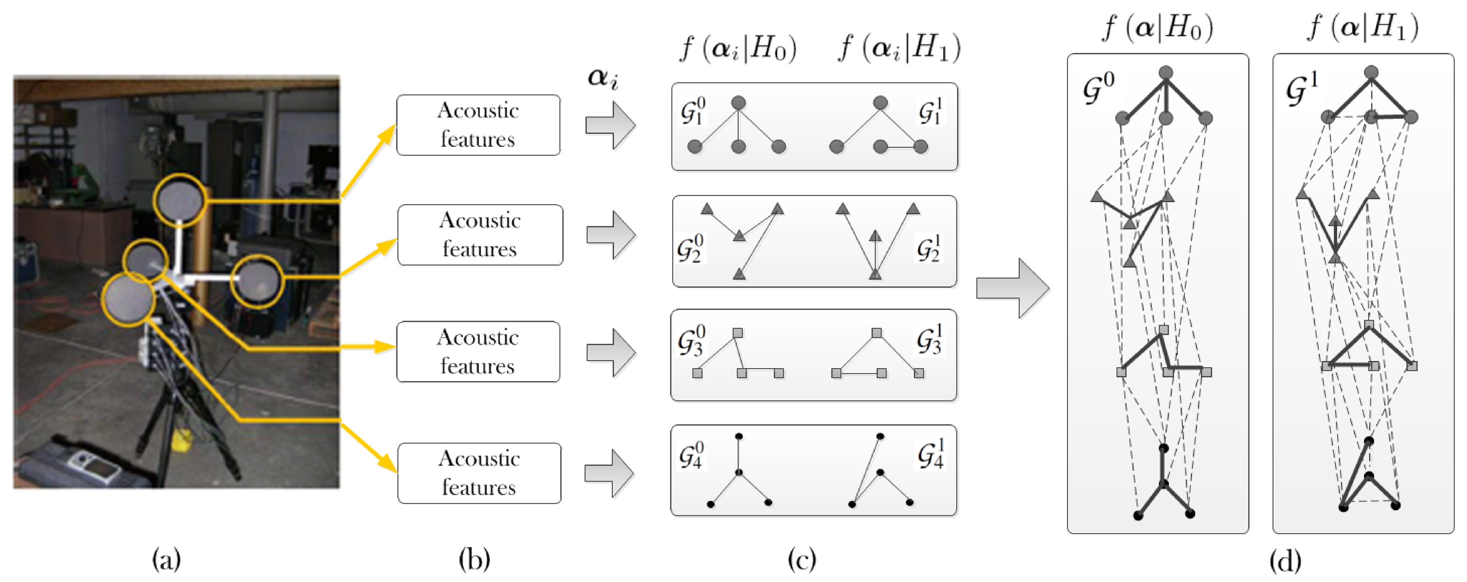


Figure: (a) Four co-located acoustic sensors. (b) Feature extraction. (c) Individual pairs of trees. (d) Thickened graphical models capturing conditional dependencies across feature sets.

Class label assignment:

$$\text{Class}(\mathbf{y}) = \begin{cases} \text{Hypothesis } H_0 & \text{if } \log\left(\frac{f(\alpha|H_0)}{f(\alpha|H_1)}\right) \geq 0 \\ \text{Hypothesis } H_1 & \text{if } \log\left(\frac{f(\alpha|H_0)}{f(\alpha|H_1)}\right) < 0 \end{cases}$$

Algorithm 1 MSGM (Steps 1-4 offline)

1: **Feature extraction (training):** Compute feature representations $\alpha_i, i = 1, \ldots, 4$, using correlated signals from the tetrahedral acoustic sensor array
2: **Initial disjoint graphs:** Discriminatively learn 4 pairs of N-node tree graphs $\mathcal{G}_i^0$ and $\mathcal{G}_i^1$ on $\{\alpha_i\}$, for $i = 1, \ldots, 4$, obtained from the training data
3: Separately concatenate nodes corresponding to the two classes, to generate initial graphs
4: **Boosting on disjoint graphs**: Iteratively thicken initial disjoint graphs via boosting to obtain final graphs $\mathcal{G}^0$ and $\mathcal{G}^1$

{**Online process**}

5: **Feature extraction (test):** Obtain feature representations $\alpha_i, i = 1, \ldots, 4$, from test signal
6: **Inference**: Classify based on output of the resulting classifier.

Experimental set-up

- Launch and impact of two types of artillery: mortar and rocket
- **Datasets:** CRAM04, CRAM05, CRAM06, Foreign
- Pre-processing to localize event segment in sensed signal
- **Comparison with:**
  1. SVM: support vector machine using RBF kernel - average of four channels classified separately
  2. CSVM: SVM on concatenated feature vectors
  3. J-SRC: joint dynamic sparsity approach [Zhang *et al.*, 2012]

Experiment: Launch vs. impact

Rocket: Training ratio, $r = 0.5$

Table: Cepstral features

Method	CRAM04	CRAM06	Foreign
SVM	0.7726	0.5845	0.8958
CSVM	0.7354	0.6063	0.9166
J-SRC	0.7911	0.6844	0.9140
MSGM	0.8032	0.7246	0.9199

Table: SDF features

Method	CRAM04	CRAM06	Foreign
SVM	0.7776	0.6079	0.8972
CSVM	0.7514	0.6221	0.9154
J-SRC	0.7962	0.6883	0.9166
MSGM	0.8066	0.7253	0.9221

Mortar: Training ratio, $r = 0.5$

Table: Cepstral features

Method	CRAM04	CRAM05	CRAM06	Foreign
SVM	0.8480	0.8127	0.8590	0.8364
CSVM	0.8449	0.8280	0.7971	0.7799
J-SRC	0.8817	0.8712	0.8770	0.8133
MSGM	0.8939	0.8853	0.8879	0.8201

Table: SDF features

Method	CRAM04	CRAM05	CRAM06	Foreign
SVM	0.8603	0.8175	0.8623	0.8398
CSVM	0.8498	0.8361	0.8012	0.7846
J-SRC	0.8837	0.8793	0.8815	0.8161
MSGM	0.8996	0.8907	0.8892	0.8248

Experiment: Noise robustness of SDF features

- CRAM04: Mortar launch vs. impact
- Corrupt each signal with AWGN such that resulting SNR is 10 dB $\rightarrow$ representative of real-world noise during acquisition
- Training ratio, $r = 0.5$

Table: Cepstral features

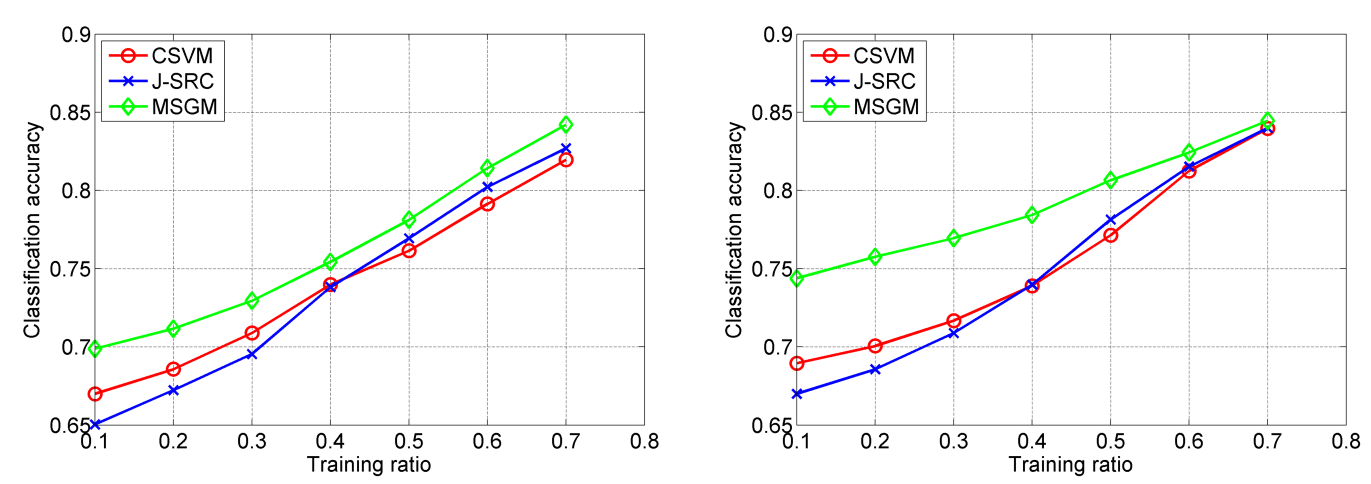
Method	CRAM04	CRAM05	CRAM06	Foreign
CSVM	0.7849	0.7618	0.7488	0.7226
J-SRC	0.8376	0.8292	0.8327	0.7640
MSGM	0.8592	0.8537	0.8501	0.7864

Table: SDF features

Method	CRAM04	CRAM05	CRAM06	Foreign
CSVM	0.8128	0.8054	0.7764	0.7557
J-SRC	0.8654	0.8511	0.8634	0.7989
MSGM	0.8772	0.8654	0.8621	0.8041

Experiment: Effect of training set size

- CRAM04 dataset: Rocket launch vs. impact



(a) Cepstral features (b) SDF features

Acknowledgment: Acoustics and EM Sensing Branch, U.S. Army Research Laboratory

IEEE ICASSP 2013

Email: uxs113@psu.edu, Web: http://signal.ee.psu.edu/