II.4 Data Fusion in IoT WSN

Assoc. Prof. Tham Chen Khong
Dept of ECE, NUS

Collaborative Signal Processing (CSP)

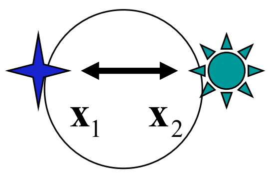
- In principle, more information about a phenomenon can be gathered from multiple measurements
  - Multiple sensing modalities (acoustic, seismic, etc.)
  - Multiple nodes
- Limited local information gathered by a single node necessitates CSP
  - Inconsistencies between measurements, such as due to malfunctioning nodes, can be resolved
- Variability in signal characteristics and environmental conditions necessitates CSP
  - Complementary information from multiple measurements can improve performance

Reference

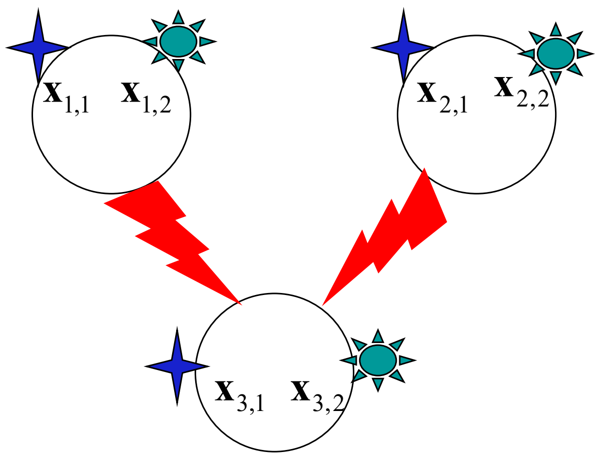
- [Brooks03] R. Brooks, P Ramanathan and A.K. Sayeed, “Distributed Target Classification and Tracking in Sensor Networks”, Proceedings of IEEE, vol. 91, no. 8, Aug 2003.

Categorization of CSP Algorithms Based on Communication Burden

- Intra-node collaboration
  - Multiple sensing modalities
    - E.g., combining acoustic and seismic measurements
  - No communication burden since collaboration is at a particular node
    - Higher computational burden at the node



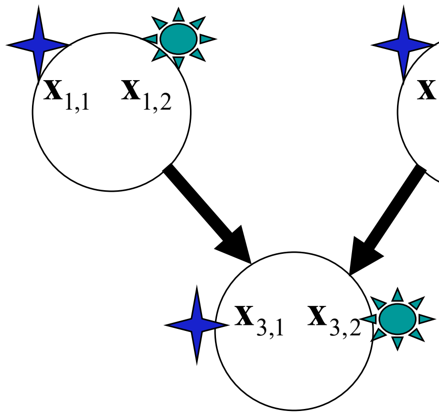
- Inter-node collaboration
  - Combining measurements at different nodes
  - Higher communication burden since data is exchanged between nodes
  - Higher computational burden at manager node



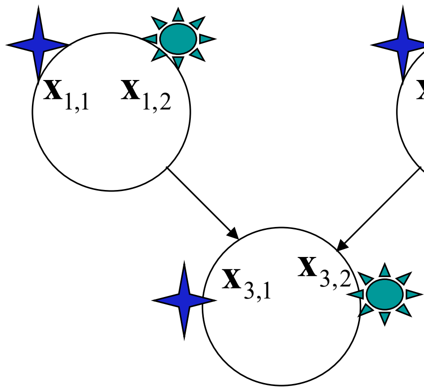
Manager node

Categorization of CSP Algorithms Based on Computational Burden

- Data fusion
  - Time series for different measurements are combined
  - Higher computational burden since higher dimensional data is jointly processed
  - Higher communication burden if different measurements from different nodes

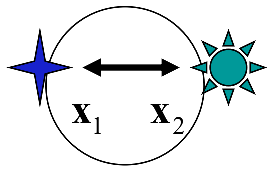


- Decision fusion
  - Decisions (hard or soft) based on different measurements are combined
  - Lower computational burden since lower dimensional data (decisions) is jointly processed
  - Higher communication burden if the component decisions are made at different nodes

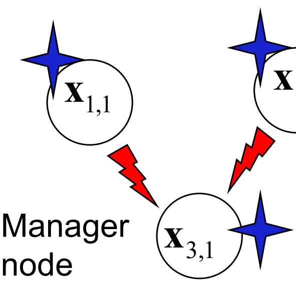


Various Forms of CSP

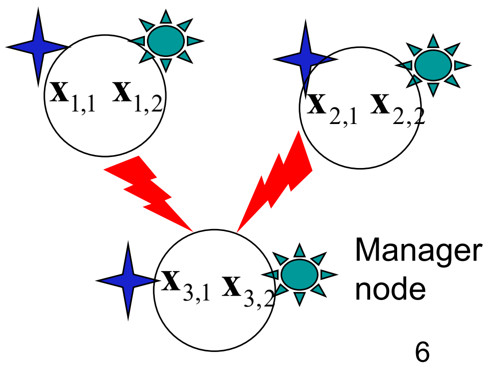
- Single Node, Multiple Modality (SN, MM)
  - Simplest form of CSP: no communication burden
    - Decision fusion
    - Data fusion (higher computational burden)



- Multiple Node, Single Modality (MN, SM)
  - Higher communication burden
    - Decision fusion
    - Data fusion (higher computational burden)



- Multiple Node, Multiple Modality (MN, MM)
  - Highest communication and computational burden
    - Decision fusion across modalities and nodes
    - Data fusion across modalities, decision fusion across nodes
    - Data fusion across modalities and nodes



Single Target Classification: Overview

- Single measurement classifiers
  - MAP/ML Gaussian classifiers
  - NN classifiers (benchmark)
  - Training and Performance Evaluation
  - Confusion matrices
- Multiple measurement classifiers
  - Data fusion (dependent measurements) (in depth)
  - Decision fusion (independent measurements) (not in depth)
- Different possibilities for CSP-based classification
  - Single node, multiple sensing modalities (SN, MM)
  - Multiples nodes, single sensing modality (MN, SM)
  - Multiple nodes, multiple sensing modalities (MN, MM)

The basic ideas illustrate general CSP principles in distributed decision making

Single Measurement Classifier

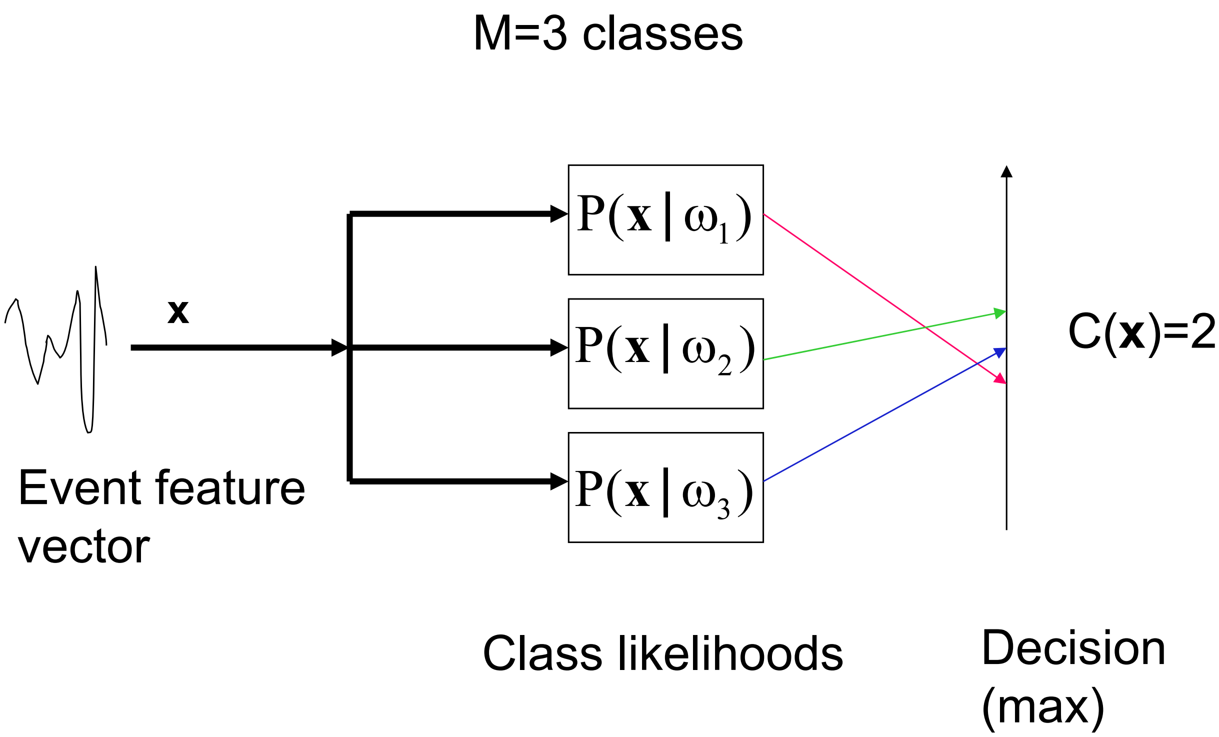
- M possible target classes: $\omega_m \in \Omega = \{m = 1, \cdots, M\}$
- $\mathbf{x}$: N-dim. (complex-valued) event feature vector
  - $\mathbf{x}$ belongs to m-th class with probability $P(\omega_m)$
- C: classifier assigns one of the classes to $\mathbf{x}$

MAP: $$C(\mathbf{x}) = m \quad \text{if} \quad P(\omega_m \mid \mathbf{x}) = \max_{j=1,\cdots,M} P(\omega_j \mid \mathbf{x})$$

Bayes rule: $$C(\mathbf{x}) = \arg\max_{j=1,\cdots,M} P(\mathbf{x} \mid \omega_j) P(\omega_j)$$

$$\boxed{\text{Equal priors (ML):} \quad C(\mathbf{x}) = \arg\max_{j=1,\cdots,M} P(\mathbf{x} \mid \omega_j)}$$

Single Measurement Classifier – Pictorially



Gaussian Classifiers

- Assume that for class j, **x** has a Gaussian distribution with mean vector $\boldsymbol{\mu}_j = \mathrm{E}_j[\mathbf{x}]$ and covariance matrix $\boldsymbol{\Sigma}_j = E_j[(\mathbf{x}-\boldsymbol{\mu}_j)(\mathbf{x}-\boldsymbol{\mu}_j)^T]$
  - $\mathrm{E}_j[\bullet]$ denotes ensemble average over class j
  - Superscript *T* denotes transpose
- Likelihood function for class j

$$\boxed{P(\mathbf{x}|\omega_j) = \frac{1}{\pi^N|\boldsymbol{\Sigma}_j|}\exp\left[-(\mathbf{x}-\boldsymbol{\mu}_j)^T\boldsymbol{\Sigma}_j^{-1}(\mathbf{x}-\boldsymbol{\mu}_j)\right]}$$

$$-\log P(\mathbf{x}|\omega_j) = \log|\boldsymbol{\Sigma}_j| + (\mathbf{x}-\boldsymbol{\mu}_j)^T\boldsymbol{\Sigma}_j^{-1}(\mathbf{x}-\boldsymbol{\mu}_j)$$

Training and Performance Assessment

- N^{Tr} training events available for each class

- 3-way cross validation – partition data into 3 sets (S_1, S_2, S_3) with equal number of events for each class

- Three sets of experiments:

Train	Test
S_1, S_2	S_3

Train	Test
S_1, S_3	S_2

Train	Test
S_2, S_3	S_1

Training and Testing

- In each experiment we have:
  - Training phase: estimate mean and covariance for each class from the two training data sets

$$\text{For} \quad \mathbf{x}_n \in \omega_j \qquad j = 1, \ldots, M$$

$$\hat{\boldsymbol{\mu}}_j = \frac{1}{N_0} \sum_{n=1}^{N_0} \mathbf{x}_n \qquad \hat{\boldsymbol{\Sigma}}_j = \frac{1}{N_0} \sum_{n=1}^{N_0} (\mathbf{x}_n - \boldsymbol{\mu}_j)(\mathbf{x}_n - \boldsymbol{\mu}_j)^T$$

  - Testing phase: Using $(\hat{\boldsymbol{\mu}}_j, \hat{\boldsymbol{\Sigma}}_j)$ estimated from the two training data sets, test the performance of the classifier on the third testing set

$$C(\mathbf{x}) = \arg \max_{j=1,\cdots,M} P(\mathbf{x} \mid \omega_j)$$

Confusion Matrix (multi-class)

ω_m \ $C(\mathbf{x})$	1	2	…	M
1	n_{11}	n_{12}		n_{1M}
2	n_{21}	n_{22}		n_{2M}
:			…	
M	n_{M1}	n_{M2}		n_{MM}

$$[\mathbf{CM}]_{ij} = n_{ij} = \text{number of events from } \omega_i \text{ classified as } \omega_j$$

Probability of Detection, Probability of False Alarm, Belief

- Probability of detection for class m

$$PD_m = \frac{n_{mm}}{\sum_{j=1}^{M} n_{mj}} \qquad \text{(m-th row)}$$

- Probability of false alarm for class m

$$PFA_m = \frac{\sum_{k=1, k\neq m}^{M} n_{km}}{\sum_{k=1, k\neq m}^{M} \sum_{j=1}^{M} n_{kj}}$$

- Prior belief in the classifier decisions (via training)

$$P(\mathbf{x} \in \omega_m \mid C(\mathbf{x}) = j) = \frac{n_{mj}}{\sum_{i=1}^{M} n_{ij}} \qquad \text{(j-th column)}$$

Binary Classification

Confusion Matrix for Binary Classification

	$\hat{P}$ (predicted)	$\hat{N}$ (predicted)	
P (actual)	TP	FN	Sensitivity/ Recall TP/(TP+FN)
N (actual)	FP	TN	Specificity TN/(TN+FP)
	Precision TP/(TP+FP)		Accuracy (TP+TN)/(TP+TN+FP+FN)

Benchmark: Nearest Neighbor (NN) Classifier

- S^{Tr} -- the set of all training event feature vectors $\mathbf{x}^{Tr}$ (containing all classes)

- $\mathbf{x}$ -- test event feature vector to be classified

$$C_{NN}(\mathbf{x}) = \text{class}\left(\arg \min_{\mathbf{x}^{Tr} \in S^{Tr}} \left\| \mathbf{x} - \mathbf{x}^{Tr} \right\|\right)$$

That is, find the *training feature vector* that is closest to the *test feature vector*. Assign the label of the closest training feature vector to the test event

Multiple Measurements

- K measurements (from a detected event)
  - Different nodes or sensing modalities

- $\mathbf{x}_k$ -- event feature vector for k-th measurement

- Classifier C assigns one of the M classes to the K event measurements $\{\mathbf{x}_1, \cdots, \mathbf{x}_K\}$

$$C(\mathbf{x}_1, \cdots, \mathbf{x}_K) = \arg \max_{j=1,\cdots,M} P(\omega_j \mid \mathbf{x}_1, \cdots, \mathbf{x}_K)$$

$$\boxed{\text{Equal priors (ML):} \quad C(\mathbf{x}_1, \cdots, \mathbf{x}_K) = \arg \max_{j=1,\cdots,M} P(\mathbf{x}_1, \cdots, \mathbf{x}_K \mid \omega_j)}$$

Data Fusion – Gaussian Classifier

- Assume that different measurements $(\{\mathbf{x}_1, \cdots, \mathbf{x}_K\})$ are jointly Gaussian and *correlated*

For $\omega_j, j = 1, \cdots, M$

the concatenated event feature vector (KN dim.)

is Gaussian with mean and covariance:

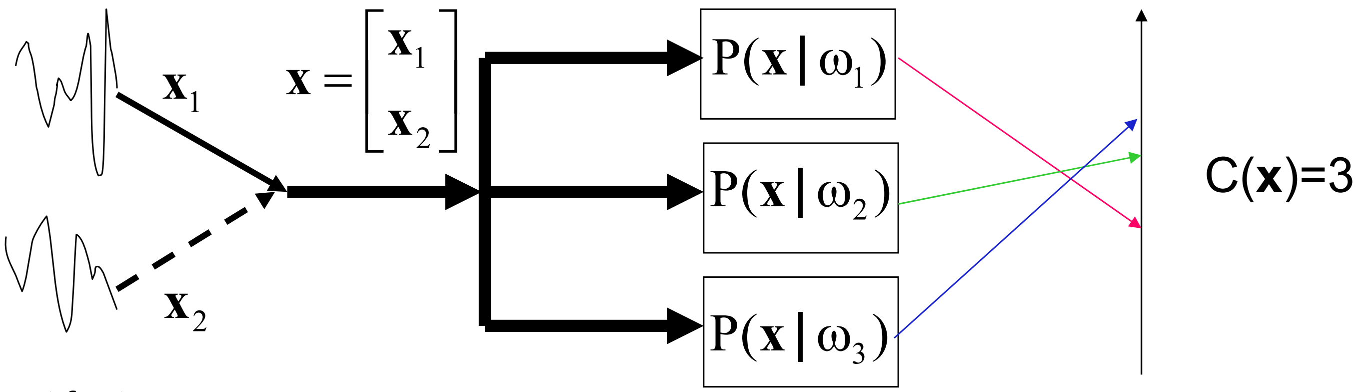
$$\mathbf{x}^c = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_K \end{bmatrix}$$

$$\boldsymbol{\mu}_j^c = E_j[\mathbf{x}^c] = \begin{bmatrix} \boldsymbol{\mu}_{j,1} \\ \vdots \\ \boldsymbol{\mu}_{j,K} \end{bmatrix} \quad \boldsymbol{\Sigma}_j^c = E_j[(\mathbf{x}^c - \boldsymbol{\mu}_j^c)(\mathbf{x}^c - \boldsymbol{\mu}_j^c)^T] = \begin{bmatrix} \boldsymbol{\Sigma}_{j,11}, \cdots, \boldsymbol{\Sigma}_{j,1K} \\ \vdots \\ \boldsymbol{\Sigma}_{j,K1}, \cdots, \boldsymbol{\Sigma}_{j,KK} \end{bmatrix}$$

$(\boldsymbol{\mu}_j^c, \boldsymbol{\Sigma}_j^c)$ characterize the j-th class and can be estimated from training data ➔ cross-validation, CM's, PD, PFA, belief

Multiple Measurement Classifier – Data Fusion

M=3 classes



Event feature vectors from 2 measurements

Concatenated event feature vector

Class likelihoods

Decision (max)

Higher computational burden

Data Fusion – NN Classifier

- Let S^{Tr} denote the set of all *concatenated* training event feature vectors $\mathbf{x}^{cTr}$ (containing all classes)

$$\mathbf{x}^{cTr} = \begin{bmatrix} \mathbf{x}_1^{Tr} \\ \vdots \\ \mathbf{x}_K^{Tr} \end{bmatrix} \qquad \text{(NK dimensional)}$$

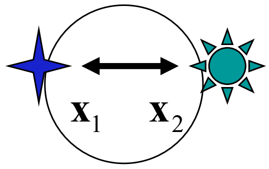
- Let $\mathbf{x}^c$ denote the concatenated test event feature vector to be classified

$$C_{NN}(\mathbf{x}_1, \cdots, \mathbf{x}_k) = \mathrm{class}\left(\arg \min_{\mathbf{x}^{cTr} \in S^{Tr}} \left\| \mathbf{x}^c - \mathbf{x}^{cTr} \right\| \right)$$

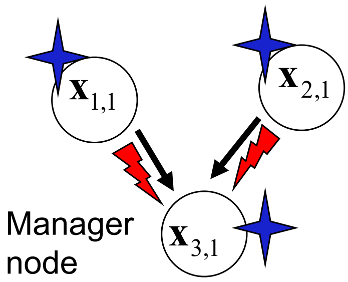
Forms of Data Fusion in CSP

K modalities, P nodes

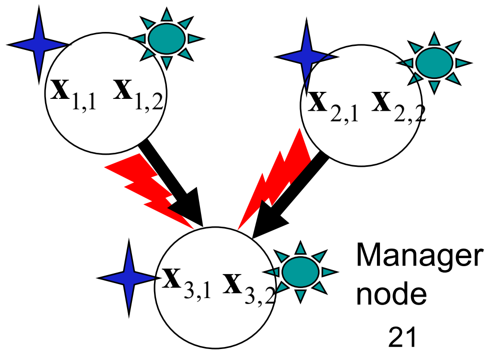
- Data fusion of multiple modalities (e.g, acoustic and seismic) at each node (SN, MM)
  - Higher comp. burden (NK dim. data)
  - No additional comm. burden



- Data fusion of a single modality at multiple nodes (MN, SM)
  - Higher computational burden at manager node (PN dim. data)
  - Higher communication burden due to transmission of N dim. data from different nodes to the manager node



- Data fusion of multiple modalities at multiple nodes (MN, MM)
  - Highest computational burden at manager node (NKP dim. data)
  - Highest communication burden due to transmission of KN dim. multi-modality data from different nodes to the manager node



Pros and Cons of Data Fusion

- Pros
  - Maximal exploitation of available information in multiple times series
  - Potentially the best performing classification scheme
- Cons
  - High computational burden
  - High communication burden if data fusion across nodes
  - Need larger amount of data for training
  - Inconsistencies between measurements could cause performance degradation (e.g. malfunctioning nodes)
- In contrast, **Decision Fusion**:
  - has lower computational and communication burden
  - however, different measurements have to be independent or uncorrelated (not covered here)

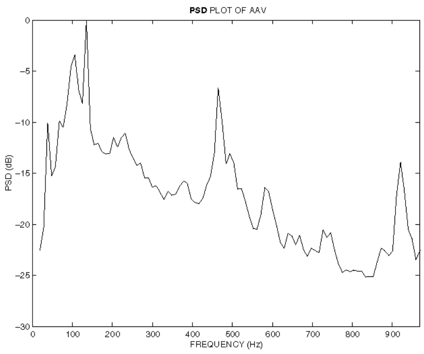
Experiments: Seismic Feature Characteristics

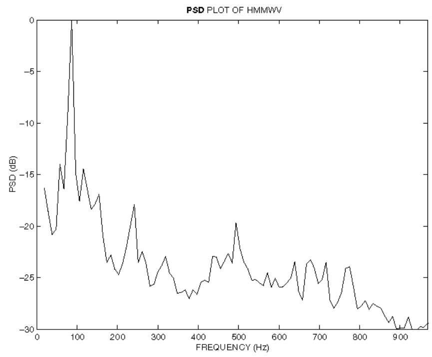
- Seismic signals
  - Sampling rate reduction from 4960 Hz to 512 Hz
  - 512-pt FFT of 512-sample (256-overlap) segments
    - 1 Hz resolution
  - The first 100 positive frequency FFT samples used (100 Hz)
  - 2-pt averaging of the 100 FFT samples yields the final N=50 dimensional FFT feature vectors
    - 2 Hz resolutions
  - About 10-50 feature vectors in each event depending on the vehicle
    - Event feature vector matrix **X** is 50x10 to 50x50
    - 50 dimensional mean event feature vectors **x**
- Complex or absolute value FFT features

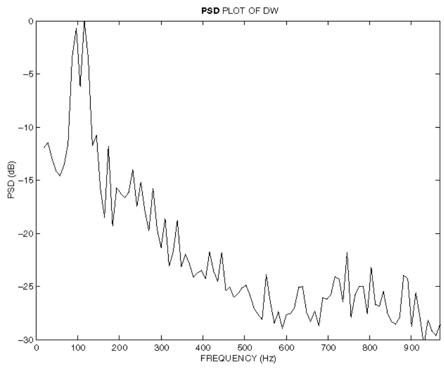
Class Descriptions

- Tracked vehicle class: AAV (Amphibious Assault Vehicle)
- Wheeled vehicle class: DW (Dragon Wagon) and HMWV (Humvee)
- Locomotion Class and Vehicle Class classification
- Approximately equal number of training and testing events for all classes
- 3-way cross validation for performance assessment

Representative Acoustic FFT Features







AAV – tracked
(Amphibious
Assault Vehicle)

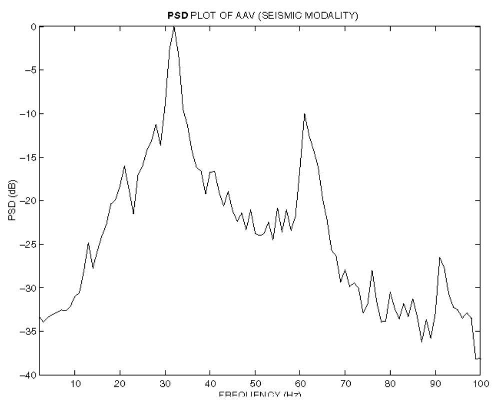
HMV – wheeled
(Humvee)

DW – wheeled
(Dragon Wagon)

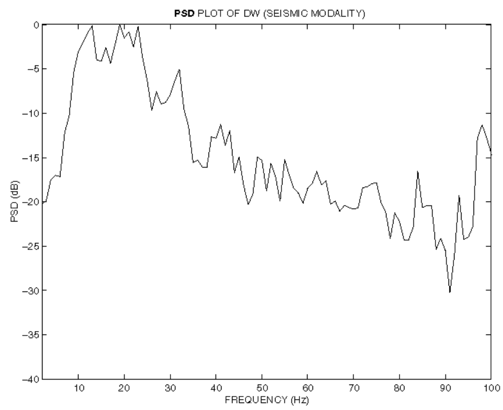
SITEX02 Data. DARPA SenseIT Program

Representative Seismic FFT Features

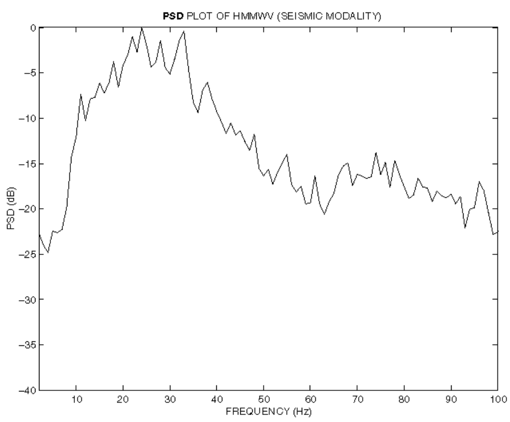
(a)



(b)



(c)



a) AAV (tracked)
b) DW (wheeled)
c) HMMWV (wheeled)

SITEX02 Data. DARPA SenseIT Program

Single Node Single Modality (SN, SM) – Locomotion Class

Absolute-value FFT acoustic features

Gaussian Classifier

ω_m \ C(**X**)	Wheeled	Tracked
Wheeled	109	11
Tracked	22	98

PD = 0.91, 0.82, Ave = 0.86

PFA = 0.18, 0.09

NN Classifier (benchmark)

ω_m \ C(**X**)	Wheeled	Tracked
Wheeled	102	18
Tracked	1	119

PD = 0.85, 0.99, Ave = 0.92

PFA = 0.01, 0.15

120 events for each class

Single Node Single Modality (SN, SM) – Vehicle Class

Absolute-value FFT acoustic features

Gaussian Classifier

ω_m \ C(**X**)	AAV	DW	HMV
AAV	53	5	2
DW	12	42	6
HMV	15	14	31

PD = 0.88, 0.70, 0.52, Ave = 0.70

PFA = 0.22, 0.16, 0.07

NN Classifier (benchmark)

ω_m \ C(**X**)	AAV	DW	HMV
AAV	43	9	8
DW	0	49	11
HMV	1	13	46

PD = 0.72, 0.82, 0.77, Ave = 0.77

PFA = 0.01, 0.18, 0.16

60 events for each vehicle

Single Node Multiple Modality (SN, MM)
Data Fusion – Locomotion Class

Acoustic and seismic features

Gaussian Classifier

ω_m \ C(**X**)	Wheeled	Tracked
Wheeled	117	3
Tracked	25	95

PD = 0.97, 0.80, Ave = 0.88

PFA = 0.21, 0.02

NN Classifier (benchmark)

ω_m \ C(**X**)	Wheeled	Tracked
Wheeled	106	14
Tracked	4	116

PD = 0.88, 0.97, Ave = 0.92

PFA = 0.03, 0.12

120 events for each class

Single Node Multiple Modality (SN, MM)
Data Fusion – Vehicle Class

Acoustic and seismic features

Gaussian Classifier

ω_m \ C(**X**)	AAV	DW	HMV
AAV	59	0	1
DW	9	46	5
HMV	25	12	23

PD = 0.98, 0.77, 0.38, Ave = 0.71

PFA = 0.28, 0.10, 0.05

NN Classifier (benchmark)

ω_m \ C(**X**)	AAV	DW	HMV
AAV	43	6	11
DW	0	47	13
HMV	1	22	37

PD = 0.72, 0.78, 0.62, Ave = 0.71

PFA = 0.01, 0.23, 0.20

60 events for each vehicle

Comparison of Various Forms of CSP – Locomotion Class

Gaussian Classifier

(SN, SM)

ω_m \ C(**X**)	Wheeled	Tracked
Wheeled	109	11
Tracked	22	98

PD = 0.91, 0.82,
Ave = 0.86

PFA = 0.18, 0.09

(SN, MM) – Data Fusion

ω_m \ C(**X**)	Wheeled	Tracked
Wheeled	117	3
Tracked	25	95

PD = 0.97, 0.80,
Ave = 0.88

PFA = 0.21, 0.02

(SN, MM) – Dec. Fusion

ω_m \ C(**X**)	Wheeled	Tracked
Wheeled	110	10
Tracked	32	88

PD = 0.92, 0.73,
Ave = 0.83

PFA = 0.27, 0.08

Comparison of Various Forms of CSP – Vehicle Class

Gaussian Classifier

(SN, SM)

ω_m \ C(**X**)	AAV	DW	HMV
AAV	53	5	2
DW	12	42	6
HMV	15	14	31

PD = 0.88, 0.70, 0.52,
Ave = 0.70

PFA = 0.22, 0.16, 0.07

(SN, MM) – Data Fusion

ω_m \ C(**X**)	AAV	DW	HMV
AAV	59	0	1
DW	9	46	5
HMV	25	12	23

PD = 0.98, 0.77, 0.38,
Ave = 0.71

PFA = 0.28, 0.10, 0.05

(SN, MM) – Dec. Fusion

ω_m \ C(**X**)	AAV	DW	HMV
AAV	55	5	0
DW	8	44	8
HMV	20	13	27

PD = 0.92, 0.73, 0.45,
Ave = 0.70

PFA = 0.23, 0.15, 0.07

Inconsistencies between modalities are present

Challenges

- Uncertainty in temporal and spatial measurements critically affects estimation:
  - Uncertainty in node locations
  - Uncertainty in timing and synchronization
- Variability in signal characteristics:
  - Doppler shifts due to motion
  - Gear shifts, acceleration in vehicles
- Variability in environmental/sensor conditions:
  - Most algorithms exploit prior statistical information about sources
  - Observed statistical characteristics can vary markedly depending on environmental conditions, such as terrain, foliage, rain, wind etc.
  - Variability in sensor characteristics (e.g., gain calibration)
- A key challenge is to develop CSP algorithms that are robust to such uncertainty/variability in measurements and conditions

Questions?