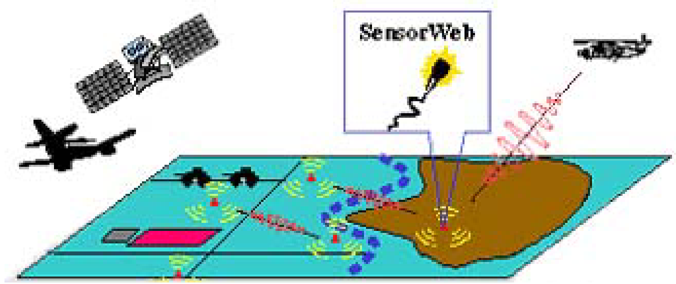


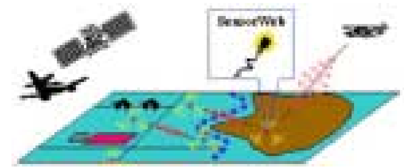
Fusion of Heterogenous Sensors in Uncertain Environments

John W. Fisher and Mujdat Cetin

Massachusetts Institute of Technology

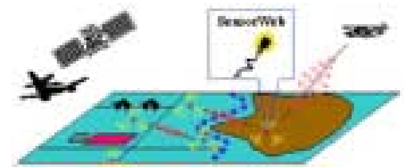
SensorWeb MURI Review Meeting

June 18, 2001

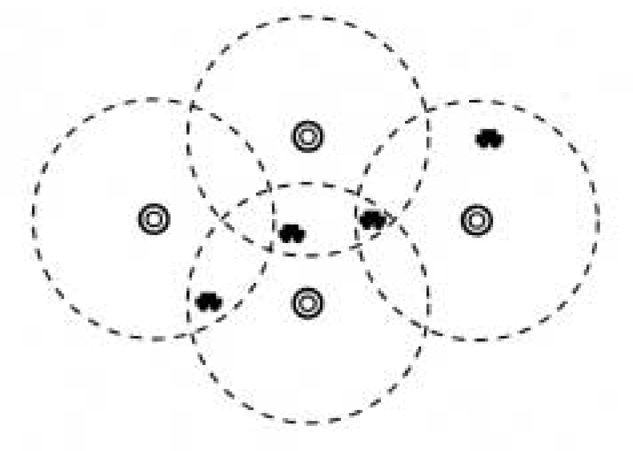


Outline

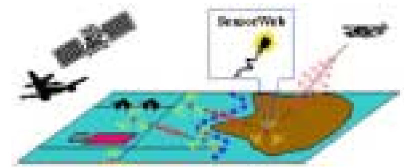
- Information Theoretic Sensor Fusion
  - John Fisher
- A Variational Approach to Array Processing Accommodating Sensor Location Uncertainties
  - Müjdat Çetin



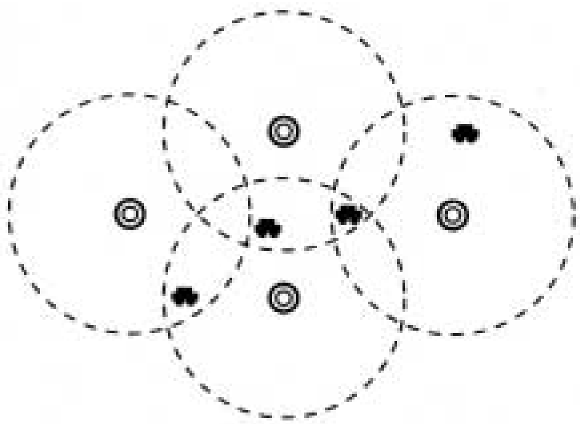
Information Theoretic Sensor Fusion



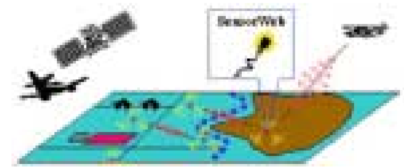
- Heterogenous sensors contain complementary information.
- Information from one sensor can be used to disambiguate mixed signals from another.
- Signal-level fusion faces challenges, including
  - A lack of accurate joint statistical models
  - high-dimensionality
  - mixed sampling rates



Information Theoretic Sensor Fusion



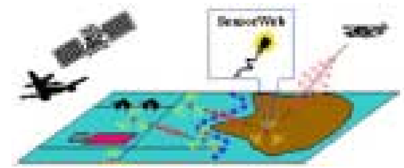
- How do we relate signals from heterogenous sensors to each other?
  - Complex temporal dependency within and between signals and modalities
  - Complex joint statistical properties
  - High dimensionality
- Can we learn and/or exploit structure in the overlapping field of regard of such sensors?
  - Recovering relative geometry



An approach for signal level fusion

Using principles from information theory and nonparametric statistics we

- project high dimensional data onto a maximally informative, low-dimensional subspace.
- model the complex stochastic relationships between the signals using a nonparametric density estimator in the subspace.
- use learned densities to process across signal modalities.



Why invoke information theory?

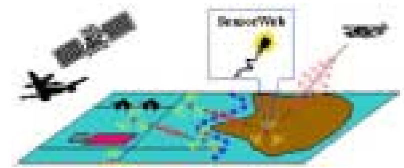
Log Likelihood vs. Nonparametric Entropy

- Given N samples $\{x_j\}$ drawn from some p(x)
- Likelihood under some parameterized model:

$$\log L = \sum_j \log p_\theta(X = x_j) \rightarrow -N(H(p) + D(p \,\|\, p_\theta))$$

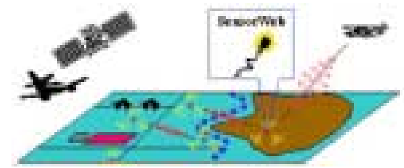
- Nonparametric Entropy using WLL estimator

$$\hat{H} = -\frac{1}{N}\sum_j \log \hat{p}(X = x_j) \rightarrow H(p) + D(p \,\|\, \hat{p})$$

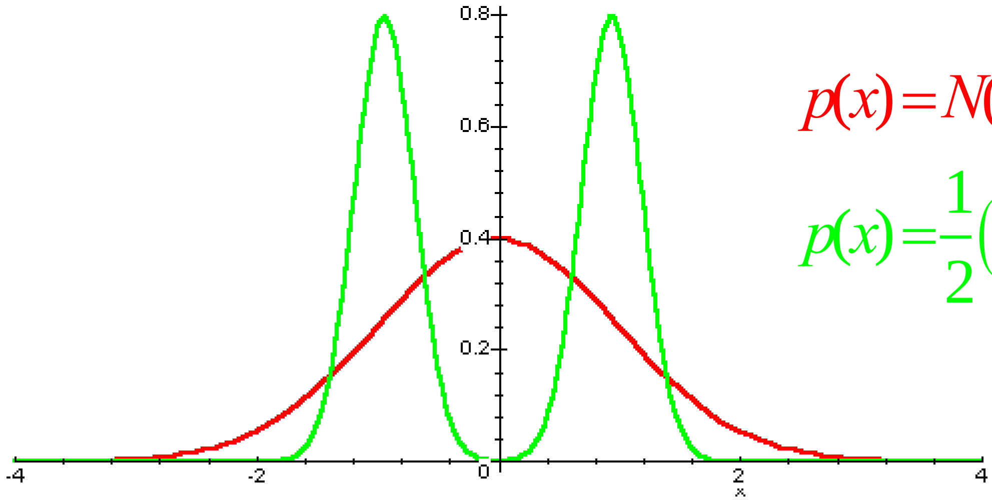


Differential entropy vs. moments?

- Densities are a complete uncertainty model
- Moments summarize the uncertainty in terms of the "spread" of a density about a central point.
  - Appropriate for uni-modal densities.
- Entropy summarizes the uncertainty in terms of the compactness (volume) of the density.
  - Appropriate for densities with complex structure (e.g. multi-modal)
  - The notion of volume is formally defined in terms of "typicality", that is entropy is related to the volume of the "typical" set.



Gaussian vs. bi-modal gaussian mixture

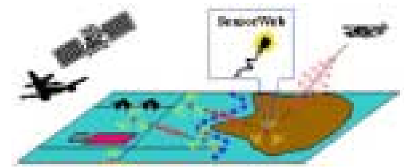


$$p(x) = N(0,1)$$

$$p(x) = \frac{1}{2}\left(N(-m,v) + N(m,v)\right)$$

$$m^2 + v^2 = 1 \quad \Rightarrow \quad \mu = 0, \sigma = 1$$

The variance is the same for both densities, but the entropy of the bi-modal density is lower.



Fano’s inequality

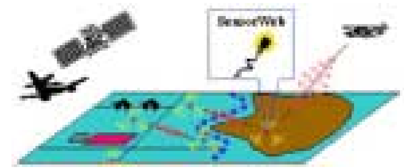
$$P(\Theta) \xrightarrow{\theta_i} p(Y \mid \Theta = \theta_i) \xrightarrow{y} \hat{\Theta}(y) \longrightarrow \hat{\theta}$$

$$P(\hat{\Theta}(Y) \neq \Theta) \geq \frac{H(\Theta) - I(\Theta, Y) - 1}{\log(N_\Theta - 1)}$$

Fano’s “Equality”

$$P(\hat{\Theta}(Y) \neq \Theta) = \frac{H(\Theta) - I(\Theta, Y) - H(E \mid Y)}{H(\Theta \mid E = 1, Y)}$$

$$P(\hat{\Theta}(Y) \neq \Theta) = \frac{H(\Theta \mid Y) - H(E \mid Y)}{H(\Theta \mid E = 1, Y)}$$



where…

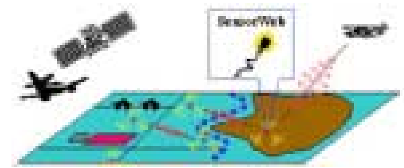
- Mutual information quantifies the reduction in uncertainty (on average) about one random variable achieved by observing another.

$$
\begin{aligned}
I(\theta, y) &= H(\theta) + h(y) - h(\theta, y) \\
&= H(\theta) - H(\theta \mid y) \\
&= h(y) - h(y \mid \theta)
\end{aligned}
$$

- The entropy terms depend on the whether the random variable is discrete or continuous.

$$H(z) = -\sum_i \log\left(p_Z\left(z_i\right)\right) p_Z\left(z_i\right) \quad \text{, z discrete}$$

$$h(z) = -\int_{\Omega_z} \log\left(p_Z\left(z\right)\right) p_Z\left(z\right) dz \quad \text{, z continuous}$$



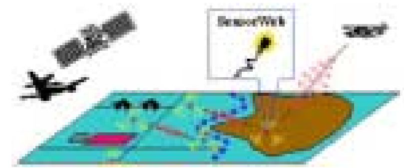
MI as a Criterion for Learning/Adaptation

Challenges

- MI as criterion for adaptation is an integral function of a probability density (and so is the approximation).
- In general we aren't given the density, only samples.

Learning Approach

- Use Parzen Density estimator
- Exploit the property that the Uniform density is the max entropy density for finite support.



Towards Approximating Entropy (from Fisher '97)

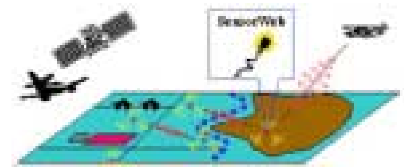
- definition of differential entropy

$$h(Y) = \int_{\Omega_Y} p(y) \log p(y)\, dy$$

- expand integrand as a 2nd order Taylor series about some density q(x).

$$p(y)\log p(y) \approx q(y)\log q(y) + \left(1 + \log q(y)\right)\left(p(y) - q(y)\right) + \frac{1}{2q(y)}\left(p(y) - q(y)\right)^2$$

- where q(x) is some density with "useful" properties

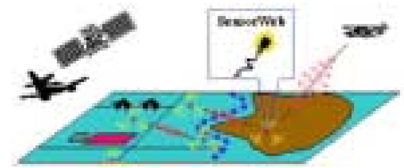


Approximating Differential Entropy

- Substitute approximation into integral and simplify

$$\begin{aligned}
\hat{H}(p) &= -\int_{\Omega_y} q(y)\log q(y) + \left(1+\log q(y)\right)\left(p(y)-q(y)\right) + \frac{1}{2q(y)}\left(p(y)-q(y)\right)^2 dy \\
&= -\int_{\Omega_y} p(y)\,dy + \int_{\Omega_y} q(y)\,dy - \int_{\Omega_y} p(y)\log q(y)\,dy - \int_{\Omega_y} \frac{1}{2q(y)}\left(p(y)-q(y)\right)^2 dy \\
&= -\int_{\Omega_y} p(y)\log q(y)\,dy - \int_{\Omega_y} \frac{1}{2q(y)}\left(p(y)-q(y)\right)^2 dy \\
&= H(p) + D\left(p\|q\right) - \int_{\Omega_y} \frac{1}{2q(y)}\left(p(y)-q(y)\right)^2 dy
\end{aligned}$$

- Consequently, maximizing this approximation to entropy is equivalent to minimizing the chi-squared distance between the density, p, and the expansion density, q.



Expansion about the Uniform Density...

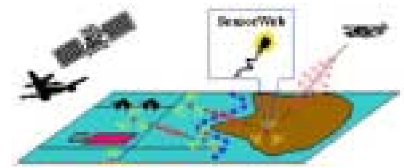
- When q(x) is the uniform density

$$H(p) + D(p \| p_u) = H(p_u) = \log V_{\Omega_Y}$$

is (trivially) true for all densities, p(x)

$$\hat{H}(p) = \log V_{\Omega_Y} - \int_{\Omega_y} \frac{V_{\Omega_Y}}{2} \left(p(y) - p_u(y)\right)^2 dy$$

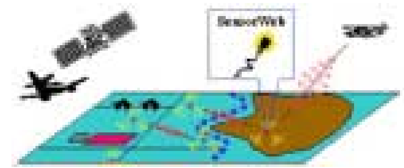
- Consequently, maximizing the approximation to entropy is equivalent to minimizing the ISE between the estimated density and the uniform density



Parzen Density Estimator

$$\hat{p}\left(y; S_y\right) = \frac{1}{N h_N} \sum_{i=1}^{N} \kappa\left(\frac{y - y_i}{h_N}\right)$$
$$= \frac{1}{N} \sum_{i=1}^{N} \kappa\left(y - y_i; h_N\right)$$

- Infers a density by convolving a kernel with the data.
- Broader L1 convergence properties than parametric approaches.
- Stone ’77 showed universal consistency.
- Does not outperform the parametric approach when the “right” parametric model is chosen.



Exact Evaluation of Integral Criterion Gradient

$$\hat{H}(\hat{p}) = \log V_{\Omega_Y} - \int_{\Omega_y} \frac{V_{\Omega_Y}}{2}\left(\hat{p}(y) - p_u(y)\right)^2 dy$$

$$= \log V_{\Omega_Y} - \int_{\Omega_y} \frac{V_{\Omega_Y}}{2}\left(\frac{1}{N}\sum_{i=1}^{N}\kappa(y - y_i; h_N) - p_u(y)\right)^2 dy$$

$$= \log V_{\Omega_Y} - \int_{\Omega_y} \frac{V_{\Omega_Y}}{2}\left(\frac{1}{N}\sum_{i=1}^{N}\kappa(y - g(x_i;\alpha); h_N) - p_u(y)\right)^2 dy$$

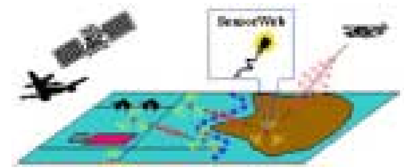
Gradient of approximation can be computed exactly by evaluation of N functions at N sample locations.

$$\frac{\partial}{\partial \alpha}\hat{H} = -\frac{V_{\Omega_Y}}{N}\sum_{i=1}^{N}\left[\varepsilon_i \frac{\partial}{\partial \alpha} g(x_i;\alpha)\right]$$

$$\varepsilon_i = f_r(y_i) - \frac{1}{N}\sum_{j \neq i}\kappa_a(y_i - y_j; h_N)$$

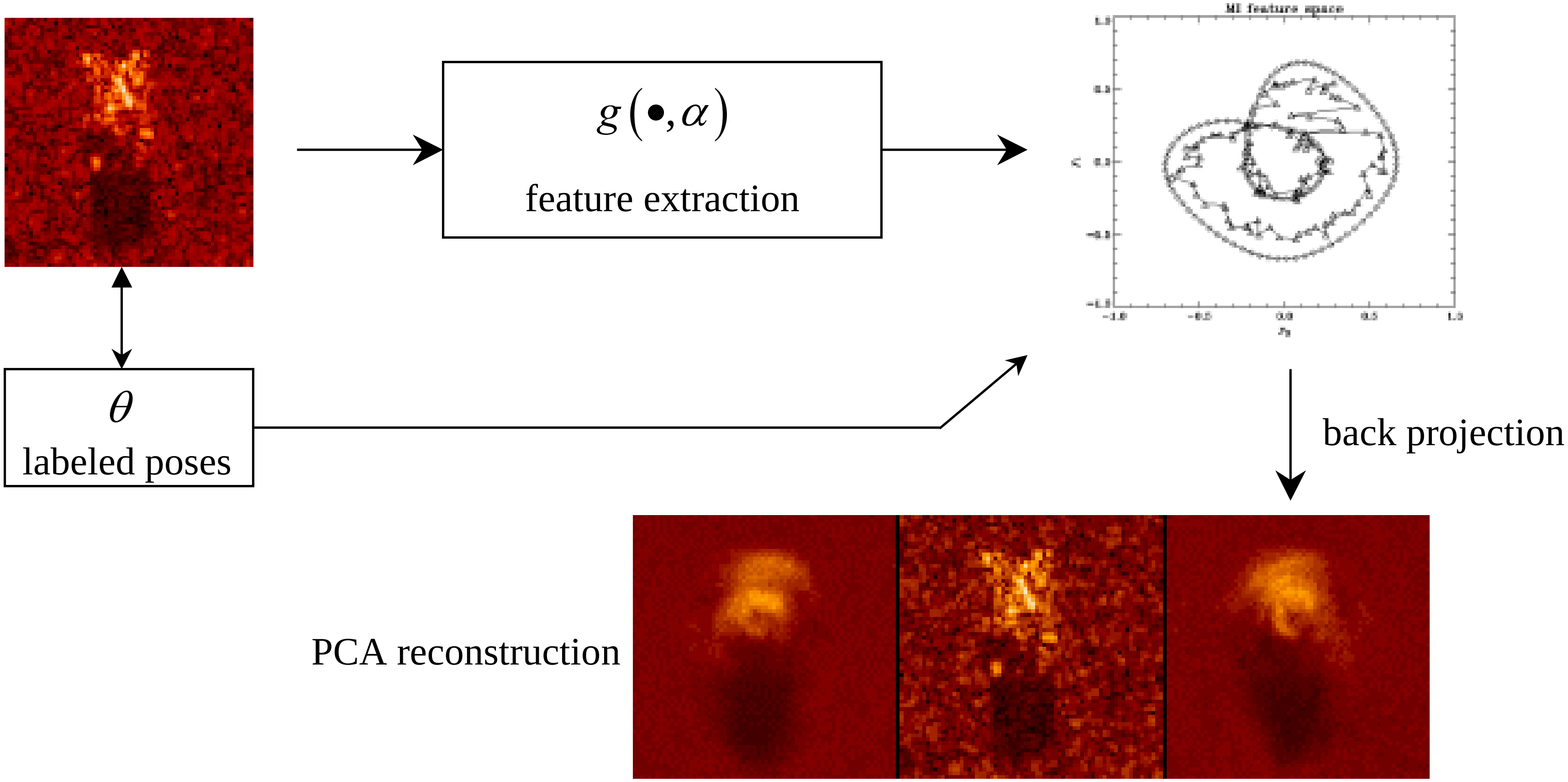
$$f_r(y) = p_u(y) * \kappa_z(y; h_N)$$

$$\kappa_a(y; h_N) = \kappa(y; h_N) * \kappa_z(y; h_N)$$



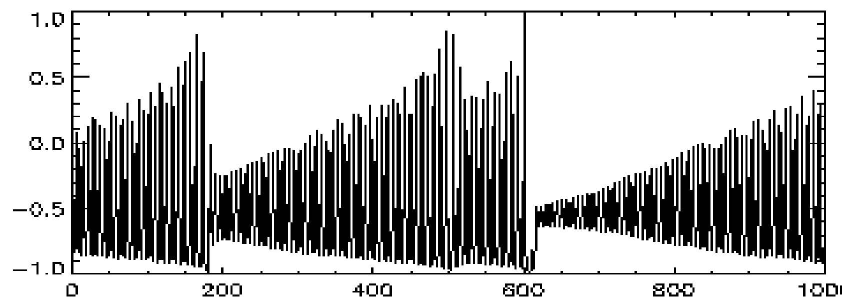
Information Preserving Transformations

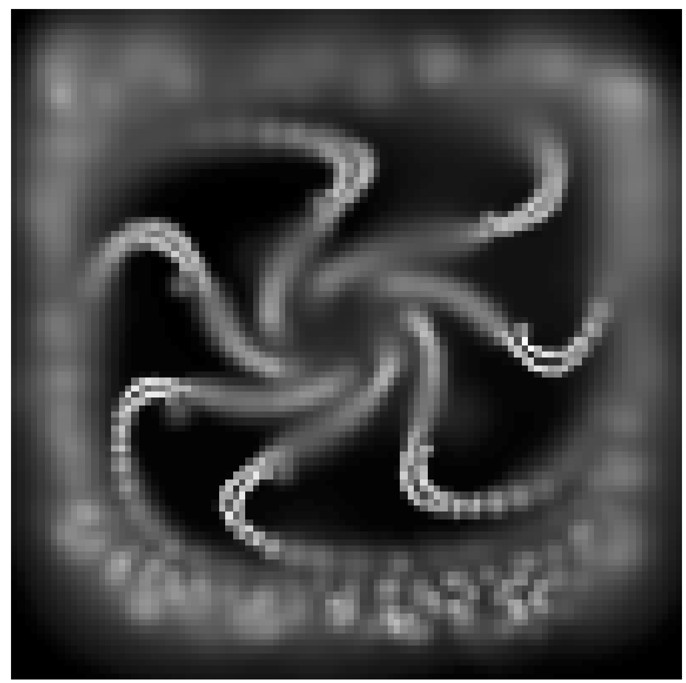
Adapt the mapping parameters, α, so as to maximize the information about the relevance parameter, θ.

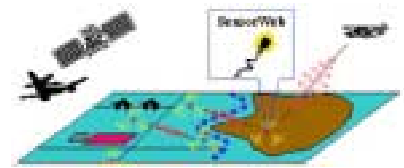


Complex temporal structure (Alex Ihler)

- Example from time-series modeling – pulsed laser data
- Learn a two dimensional statistic (which is a function of the past N samples) that has high mutual information with the next sample
- Low dimensionality does not necessarily equate to low complexity

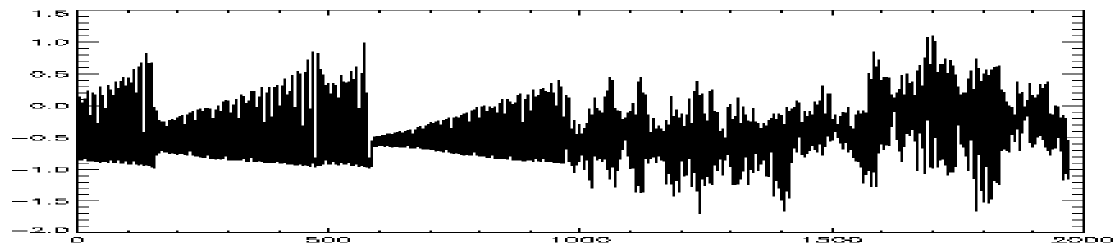




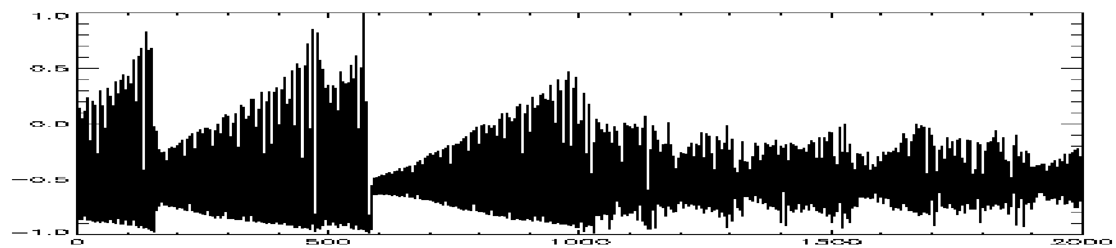


Synthesis Examples

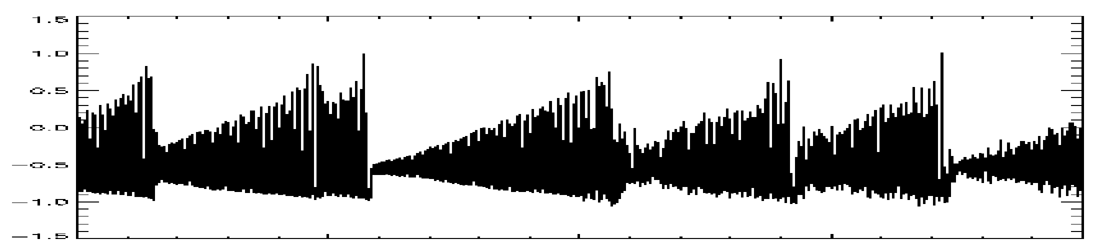
Gaussian assumption

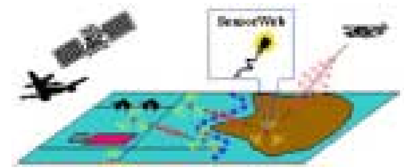


1D Learned Statistic

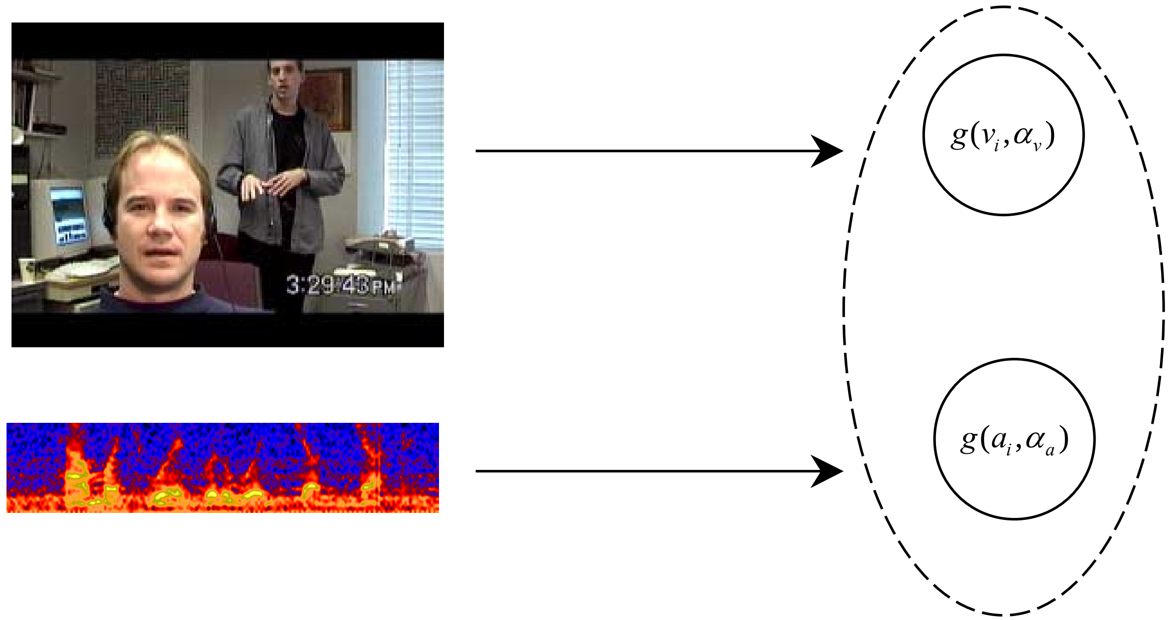


2D Learned Statistic





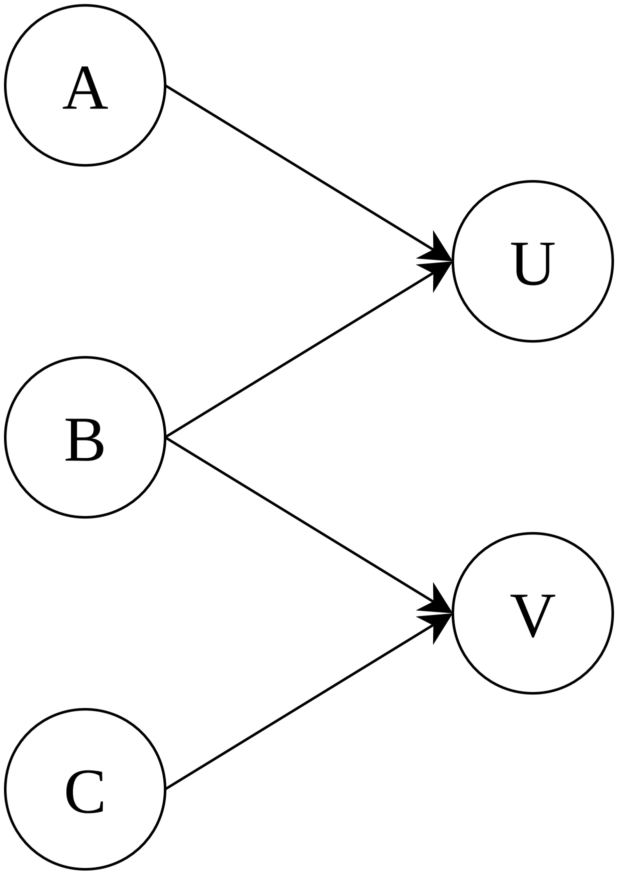
Audio/Video fusion using MI

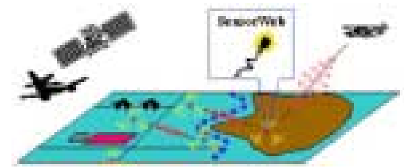


- Choose the mapping parameters such that the mutual information between the extracted features is maximized (i.e. project onto a maximally informative subspace.
- Why is this the “right” thing to do (or rather when)?

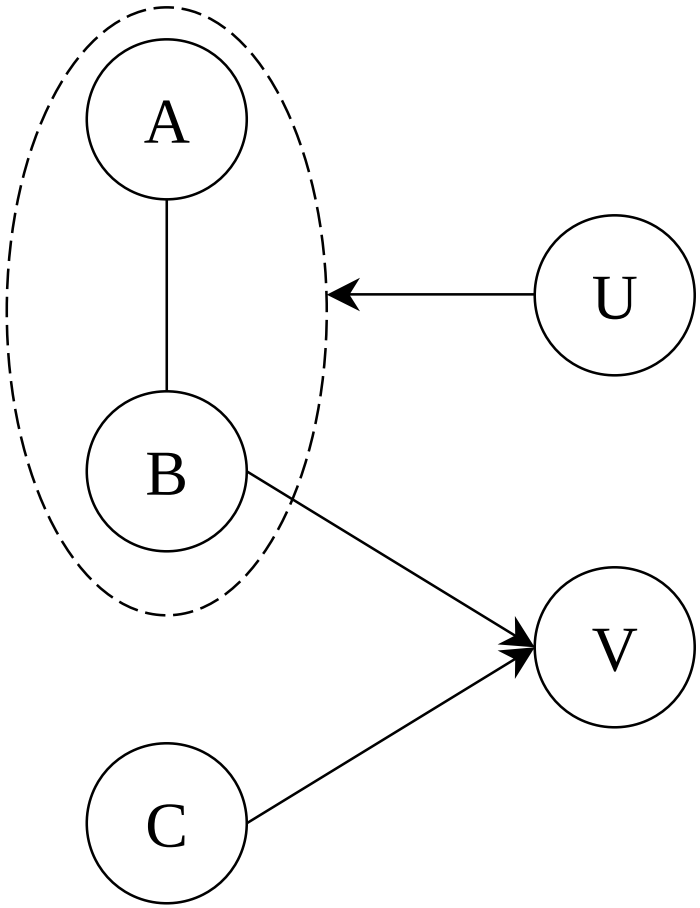
Independent Cause Model

$$p(A,B,C,U,V) = p(A)\,p(B)\,p(C)\,p(U|A,B)\,p(V|B,C)$$



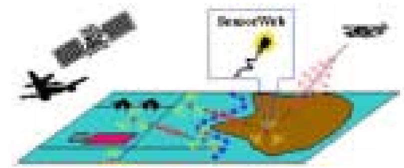


Induced dependency amongst causes



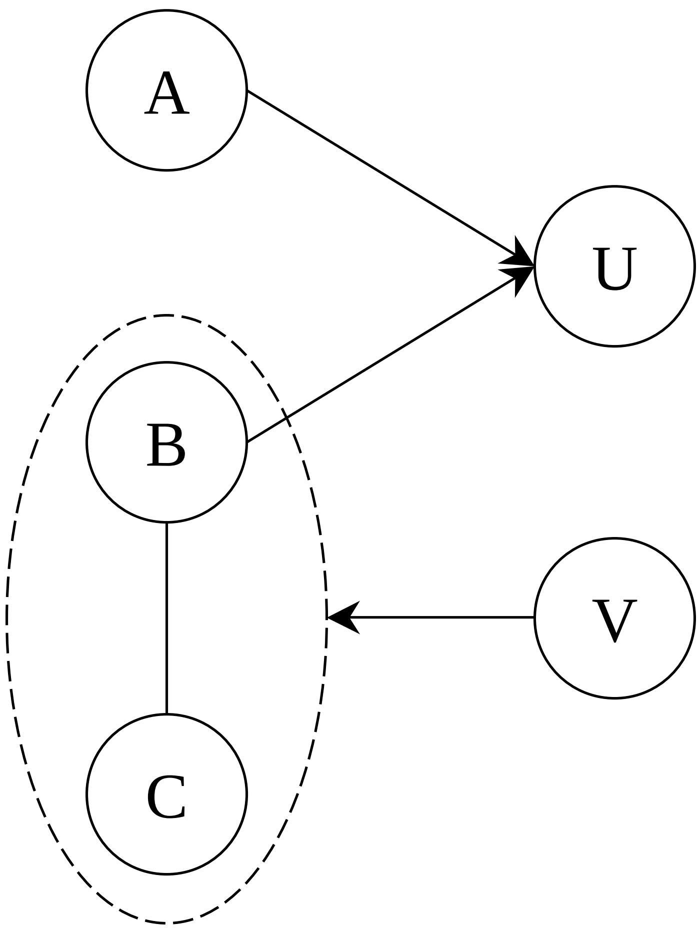
$$p(A,B,C,U,V) = p(A)\,p(B)\,p(C)\,p(U|A,B)\,p(V|B,C)$$

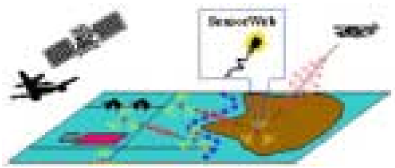
$$= p(U)\,p(A,B|U)\,p(C)\,p(V|B,C)$$



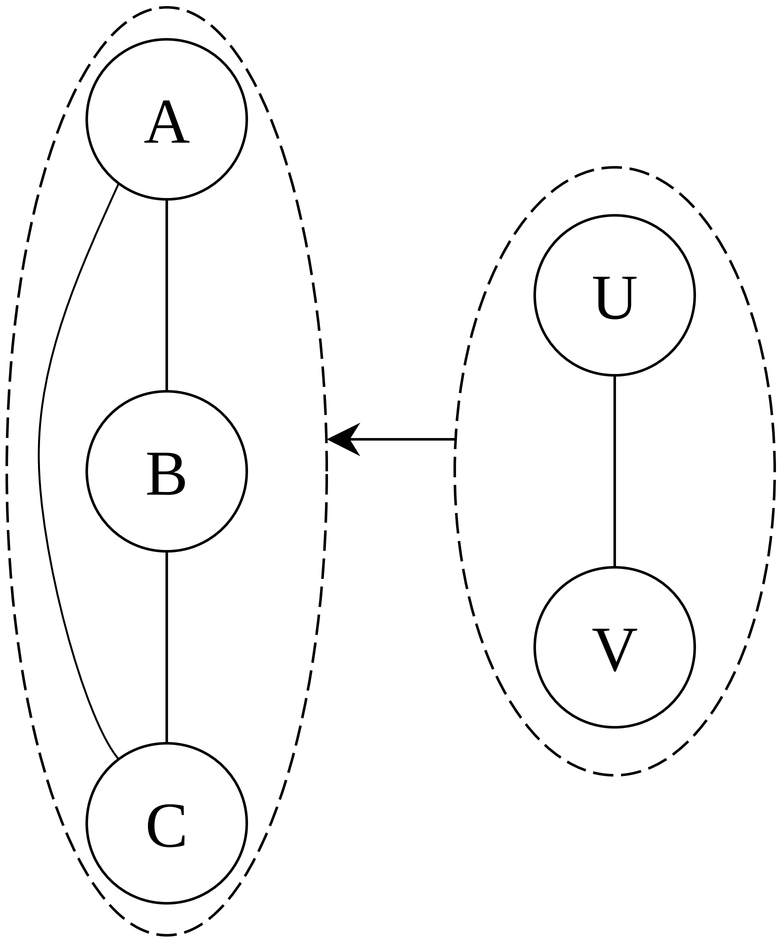
Induced dependency amongst causes

$$
\begin{aligned}
p(A,B,C,U,V) &= p(A)\,p(B)\,p(C)\,p(U|A,B)\,p(V|B,C) \\
&= p(U)\,p(A,B|U)\,p(C)\,p(V|B,C) \\
&= \color{red}{p(V)\,p(B,C|V)\,p(A)\,p(U|A,B)}
\end{aligned}
$$

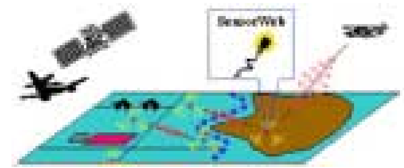




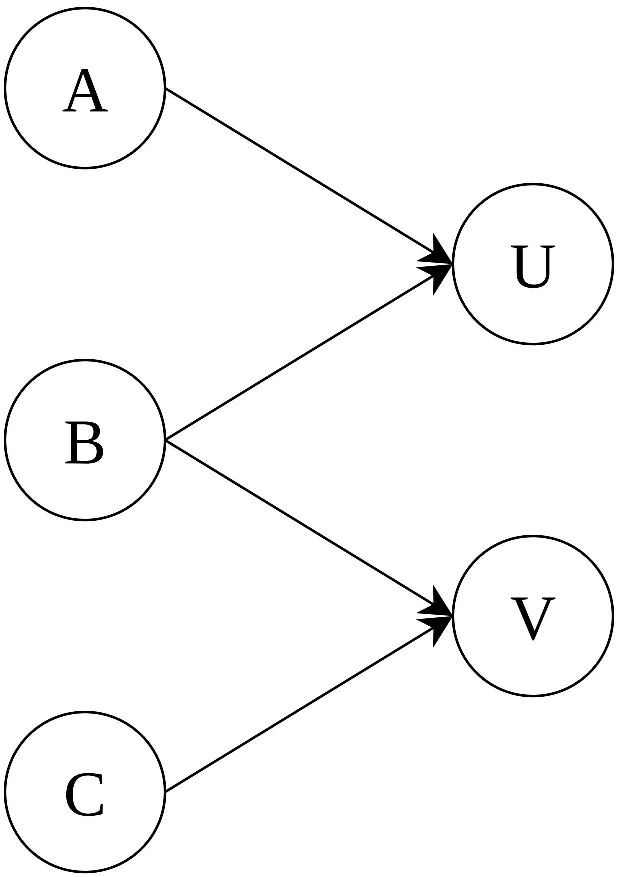
Joint observations increase complexity



$$
\begin{aligned}
p(A,B,C,U,V) &= p(A)\,p(B)\,p(C)\,p(U|A,B)\,p(V|B,C) \\
&= p(U)\,p(A,B|U)\,p(C)\,p(V|B,C) \\
&= p(V)\,p(B,C|V)\,p(A)\,p(U|A,B) \\
&= {\color{red} p(U,V)\,p(A,B,C|U,V)}
\end{aligned}
$$



Separating Functions

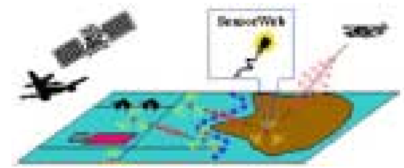


Suppose a separation of U and V *exists* such that:

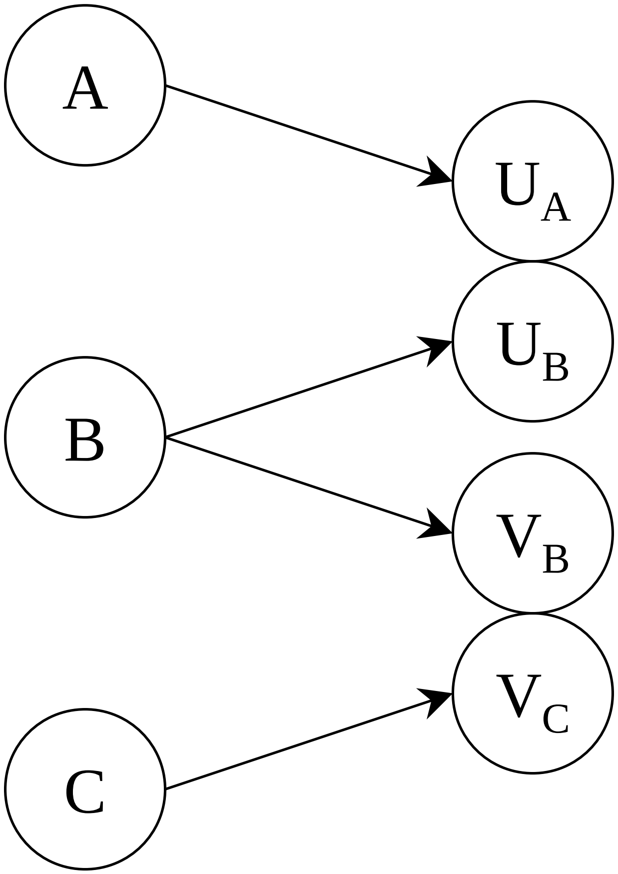
$$p(U, A, B) = p(A)\,p(B)\,p(U_A|A)\,p(U_B|B)$$

$$p(V, B, C) = p(B)\,p(C)\,p(V_B|B)\,p(V_C|C)$$

then…



Markov Property

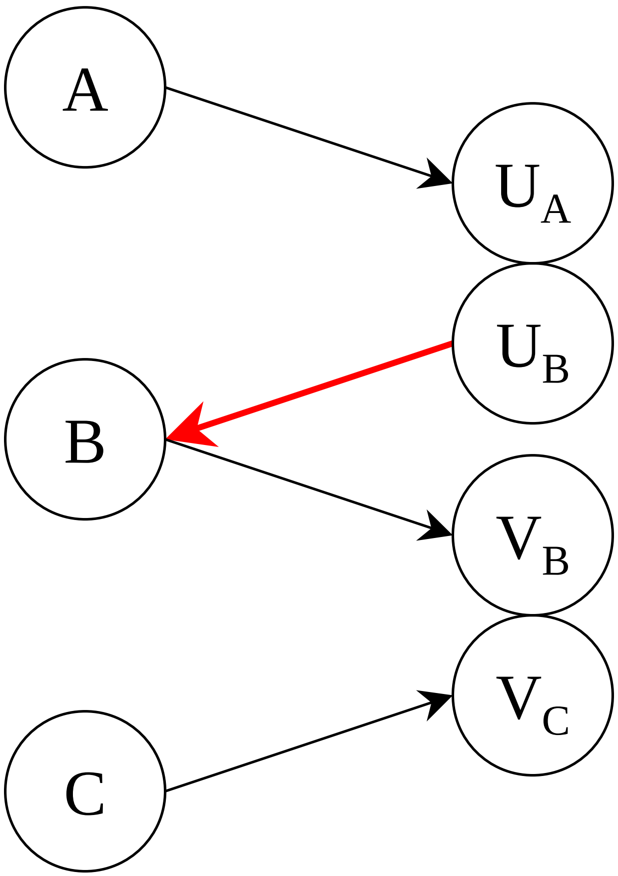


becomes

$$p(A,B,C) = p(A)\,p(B)\,p(C)\,p(U_A|A)\,p(U_B|B)\,p(V_B|B)\,p(V_C|C)$$

Bearing in mind that we still have the task of finding a separating function (or an approximate one).

Markov Property

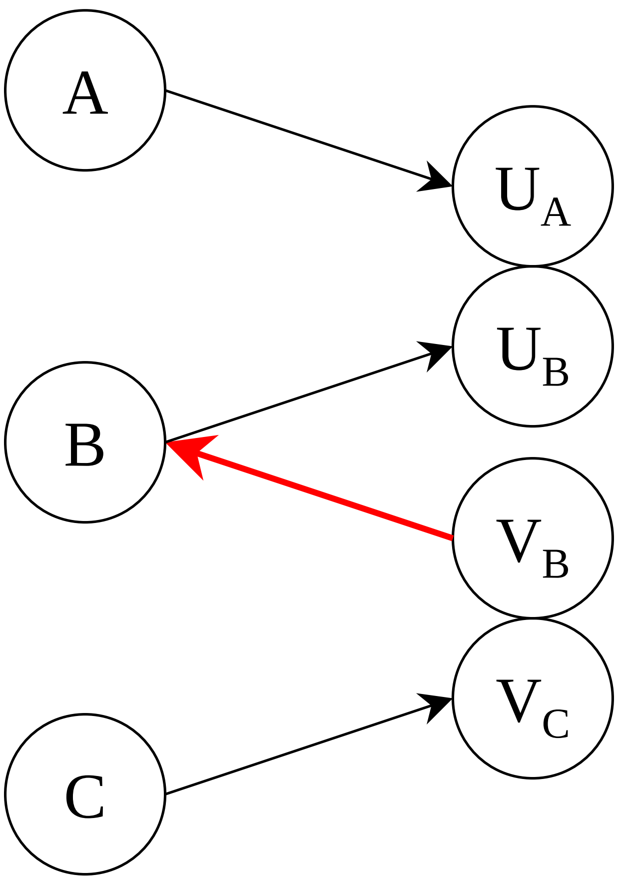


or

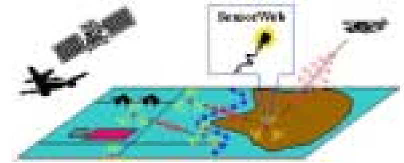
$$
\begin{aligned}
p(A,B,C) = {} & p(A)\,p(B)\,p(C) \\
& p(U_A|A)\,p(U_B|B)\,p(V_B|B)\,p(V_C|C) \\
= {} & p(U_B)\,p(B|U_B)\,p(V_B|B) \\
& p(A)\,p(C)\,p(U_A|A)\,p(V_C|C)
\end{aligned}
$$

Markov Property

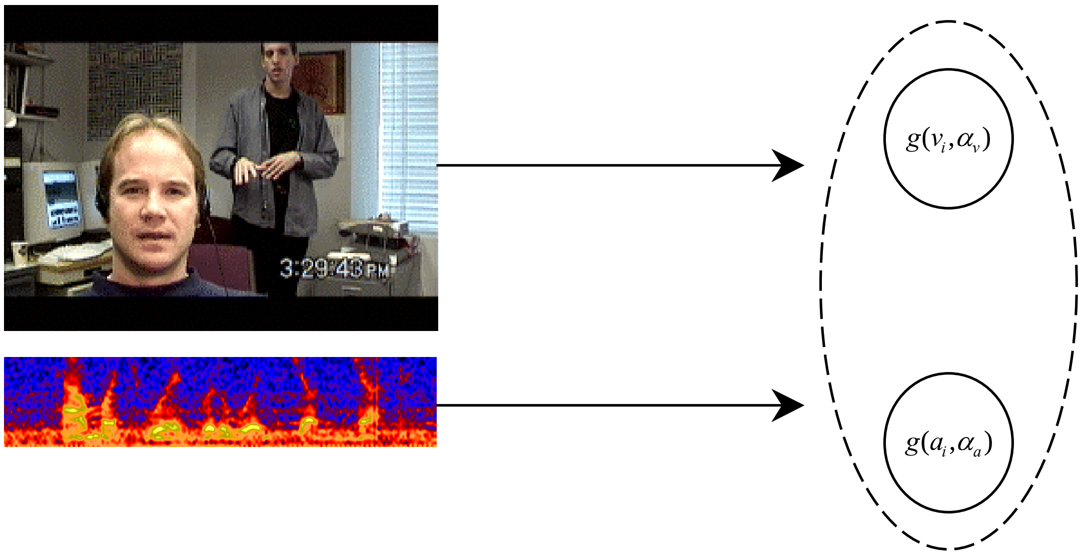
or



$$
\begin{aligned}
p(A,B,C) &= p(A)\,p(B)\,p(C) \\
&\quad p(U_A|A)\,p(U_B|B)\,p(V_B|B)\,p(V_C|C) \\
&= p(U_B)\,p(B|U_B)\,p(V_B|B) \\
&\quad p(A)\,p(C)\,p(U_A|A)\,p(V_C|C) \\
&= {\color{red} p(V_B)\,p(B|V_B)\,p(U_B|B)} \\
&\quad p(A)\,p(C)\,p(U_A|A)\,p(V_C|C)
\end{aligned}
$$



Audio/Video using MI



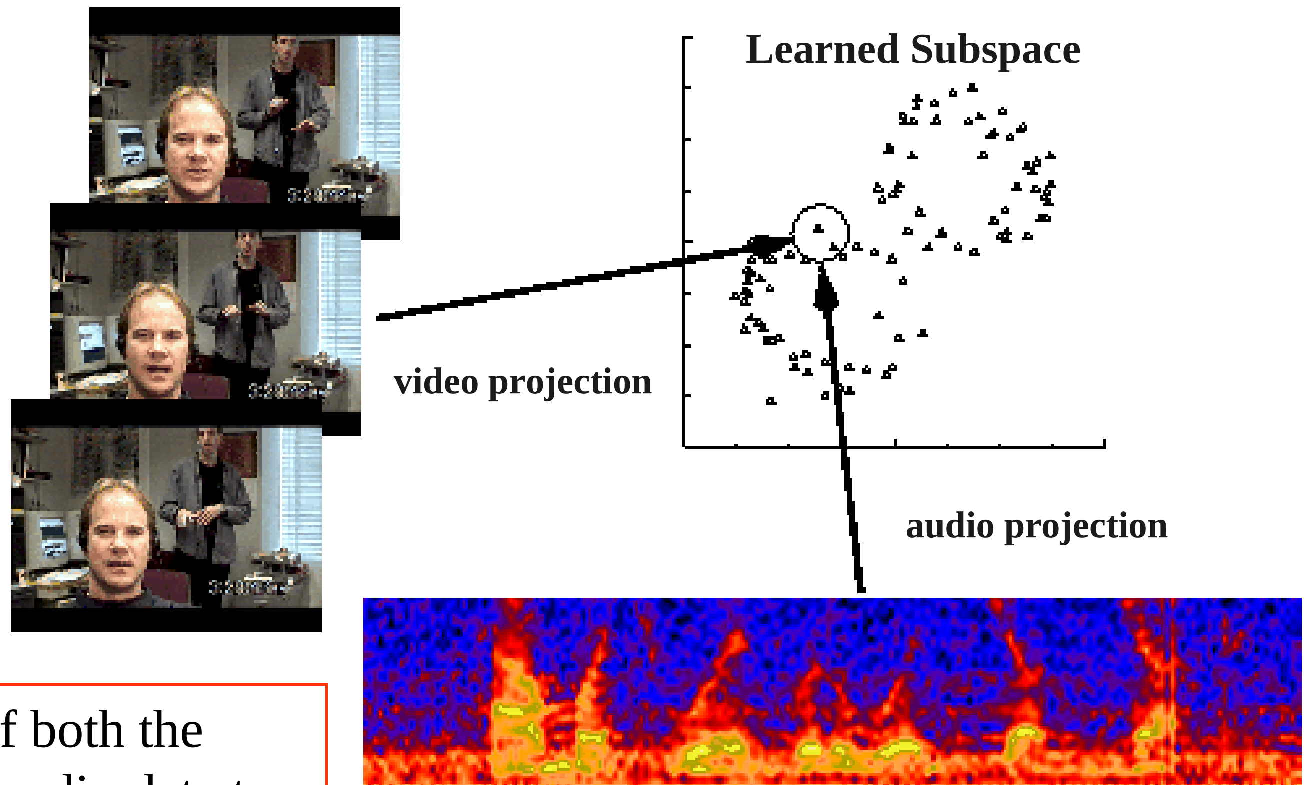
$$I\left(g\left(V,\alpha_v\right),g\left(U,\alpha_u\right)\right)\leq I\left(B,g\left(U,\alpha_u\right)\right)$$

$$I\left(g\left(V,\alpha_v\right),g\left(U,\alpha_u\right)\right)\leq I\left(B,g\left(V,\alpha_v\right)\right)$$

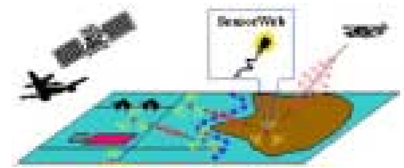
•By maximizing MI, we are summarizing the common information in the measurements, (i.e. which is related to their common cause).

•From the information theory perspective, the joint of the feature variables is a proxy for the “observable” part of their common cause.

Maximally Informative Subspace



Find a projection of both the video data and the audio data to a low-dimensional space such that MI is maximized.



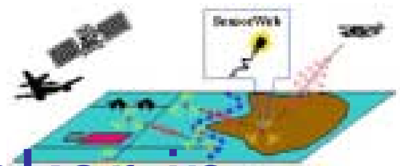
Learning the Subspace

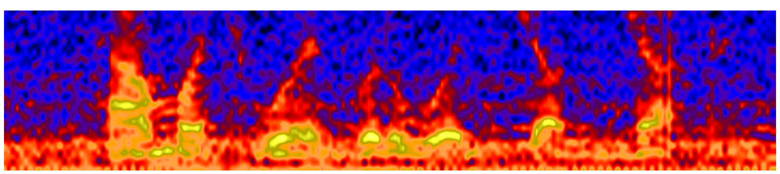
•The mapping parameters are chosen to maximize the mutual information in the low dimensional *output space*.

$$\left\{\hat{\alpha}_v, \hat{\alpha}_u\right\} = \arg\max_{\alpha_v, \alpha_a} I\left(f_v\left(V, \alpha_v\right), f_a\left(U, \alpha_u\right)\right)$$

•Video localization and audio filter design are inferred as a function of the learned weights.

Video Localization of Single Speaker in the Presence of Motion Distractors



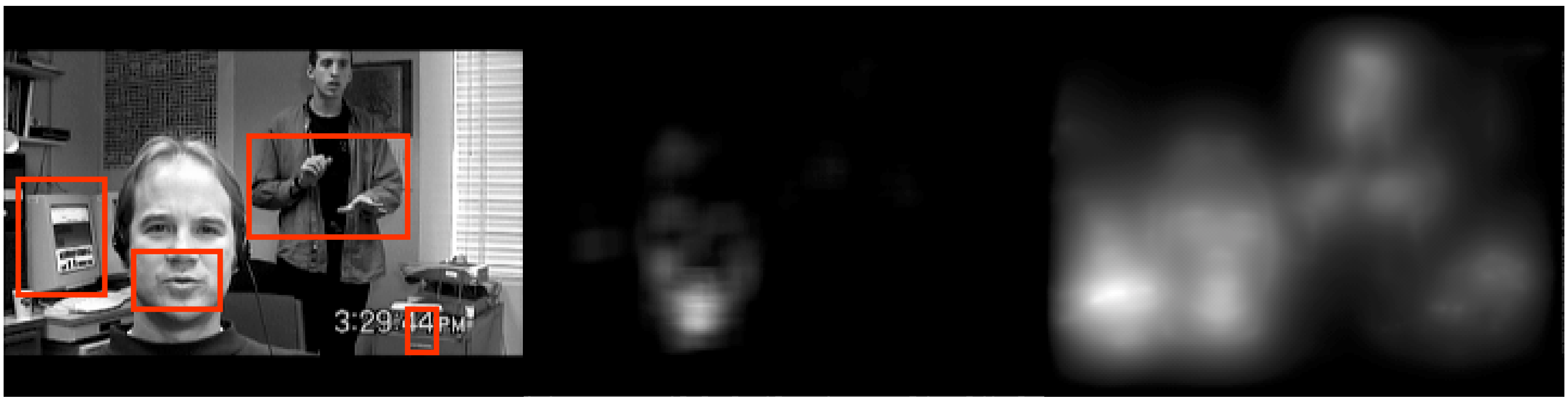


- Which pixels are "related" to the associated audio?
- Joint statistics of video and audio modalities are not well modeled by parametric forms.
- Slaney and Covell (NIPS '00) demonstrate that canonical correlations (a second-order statistical measure) do not successfully detect audio/video synchrony using spectral representations.
- Classical sensor fusion approaches are formulated as joint Bayesian estimation problems, which is equivalent to MI in the non-parametric case.

Detecting (change) motion is not enough

- Red squares indicate regions with large pixel variance
- Variance image of sequence at left
- Magnitude of MAX MI video projection shown at center
- Inspection of the learned video projection coefficients tells us which pixels are associated with the audio signal.

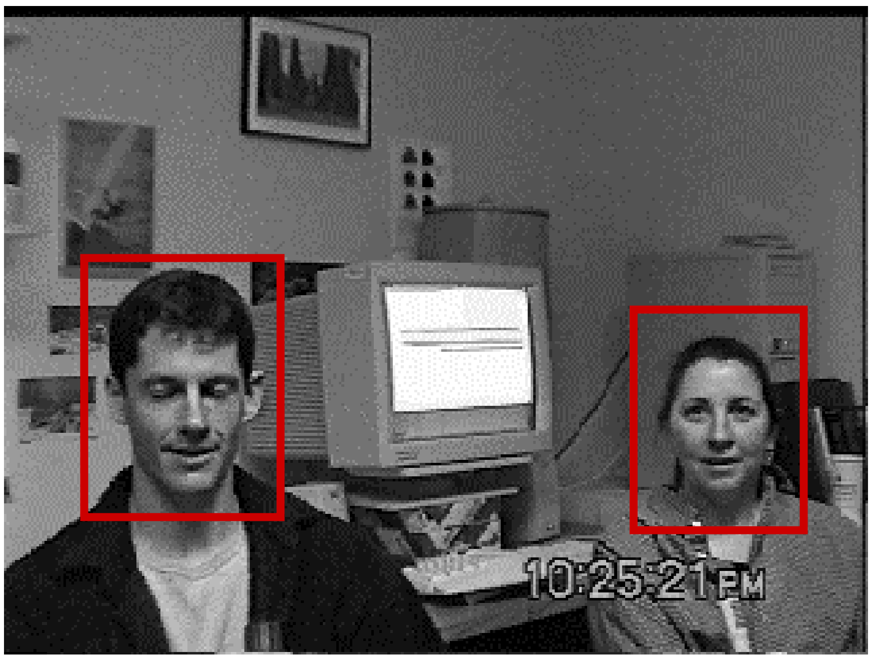
Representation: pixel vs. motion



•Similar result using an optic flow representation [Anandan '89] of motion in the video
•Fusion approach does not explicitly rely on how information is represented in data

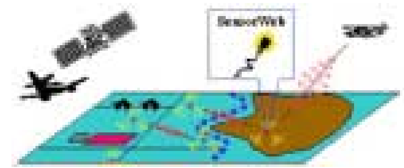
Video Localization (more examples)

Audio Enhancement



- Left channel
- Right channel
- Wiener (left)
- Wiener (right)
- MI (left)
- MI (right)

In this experiment, regions of the video are selected for enhancement (e.g. face detector, manually).

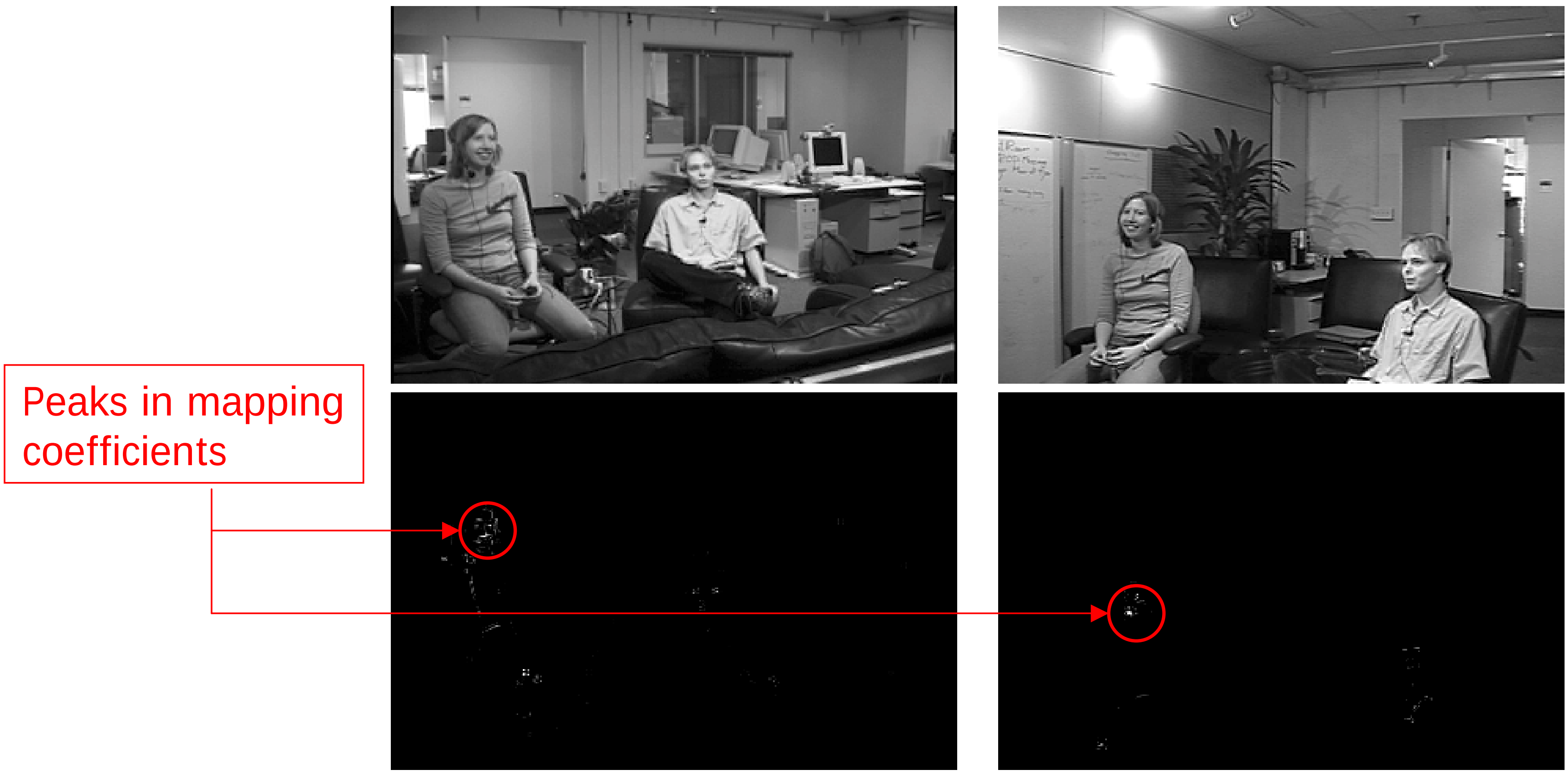


Wiener Filter Comparison

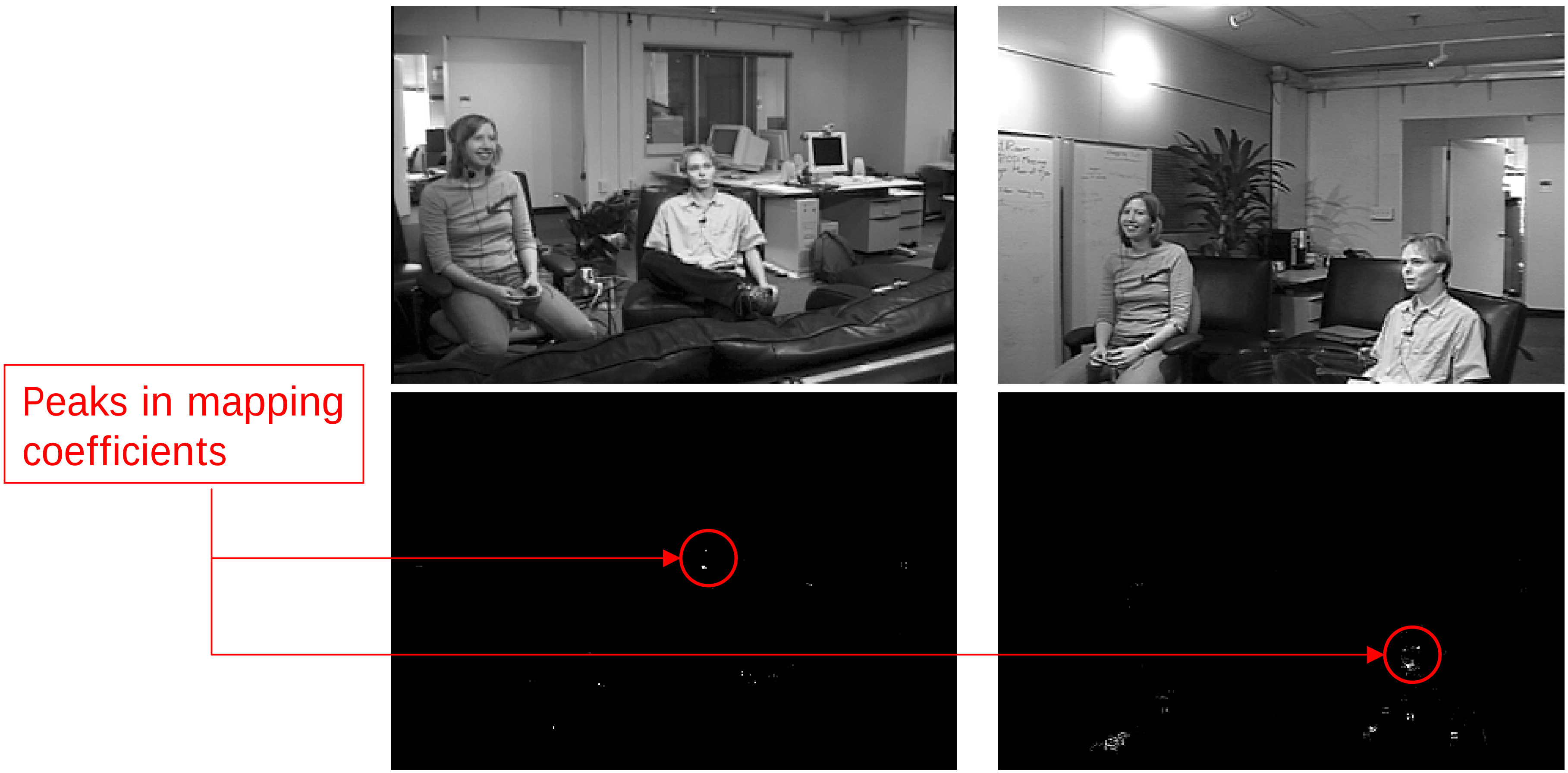
	Wiener filter	Pixel-Periodogram Representation	Optical Flow-Periodogram Representation
SPG (male voice)	10.43 dB	8.9 dB	9.2 dB
SPG (female voice)	10.5 dB	5.7 dB	5.6 dB

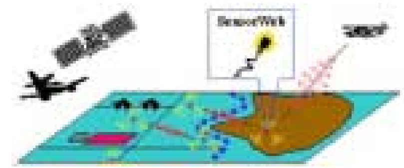
Acquiring correspondences

Acquiring correspondences



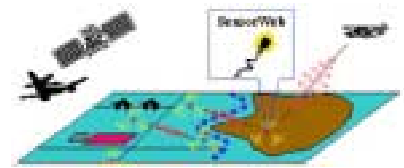
Acquiring correspondences



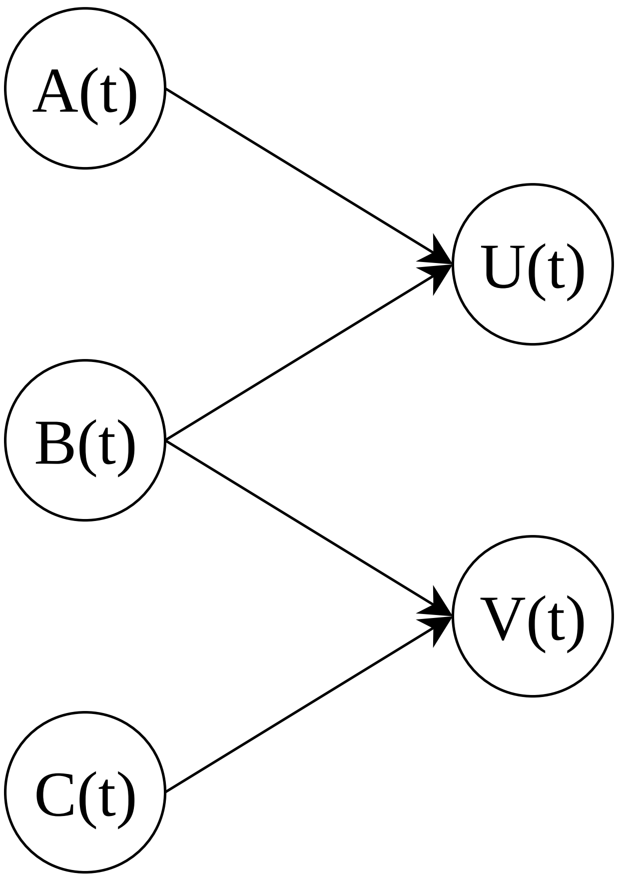


Extensions

- A basic algorithm has been developed
- Need to incorporate *multiple* independent causes (order estimation)
- Temporal dependency of joint measurements
- Testing on new data sources (e.g. audio, seismic, etc.)



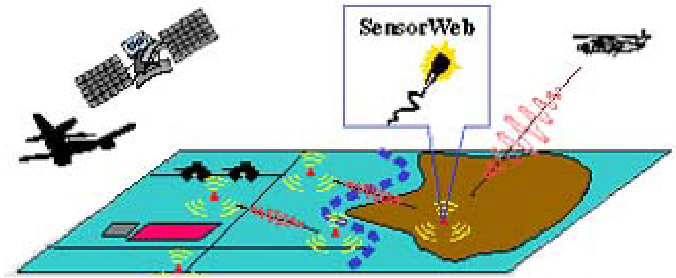
Exploiting array structure



- Two sensors observer mixture of three signals

$$U(t) = A(t) + B(t)$$
$$V(t) = G(B(t)) + C(t)$$

- G( ) is unknown and may be nonlinear
- Use knowledge of the mixing structure to separate signals



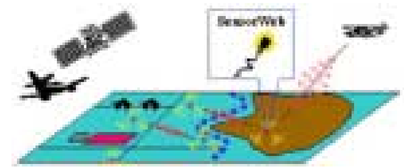
A Variational Approach to Array Processing Accommodating Sensor Location Uncertainties

Müjdat Çetin

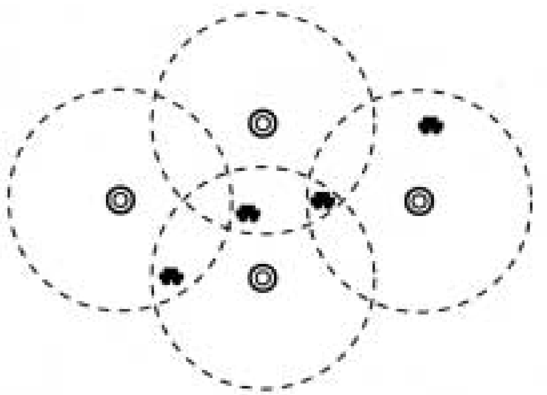
Stochastic Systems Group, M.I.T.

SensorWeb MURI Review Meeting

June 18, 2001

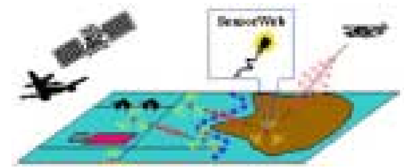


The Source Localization Problem



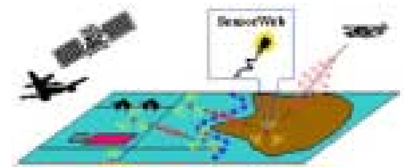
$$\mathbf{y}(t) = \mathbf{A}(\mathbf{r})\mathbf{s}(t) + \mathbf{w}(t)$$

- Find source location parameters based on data from multiple sensors
- Assumptions for a basic problem:
  - Unknown number of narrowband sources in near or far field
  - Omnidirectional sensors
  - Limited aperture size (→ limited Rayleigh resolution)
  - Sensor locations known only approximately



Variational Approach - Motivation

- View the problem as one of *imaging* a "source density" over the field of regard
  - Ill-posed inverse problem
  - Cast as an optimization problem and *regularize* by favoring fields with *concentrated densities*
  - Can include optimization over sensor locations
  - Analogous to auto-focusing and point-enhanced imaging in other array processing problems in which there are "phase defects" to be accommodated

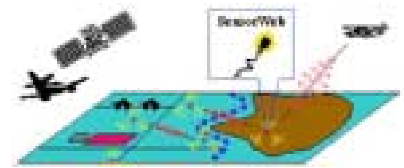


Variational Formulation

- Minimize the objective function:

$$J(\mathbf{s},\tilde{\mathbf{r}}) = \left\|\mathbf{y} - \mathbf{A}(\tilde{\mathbf{r}})\mathbf{s}\right\|_2^2 + \lambda\Psi(\mathbf{s})$$

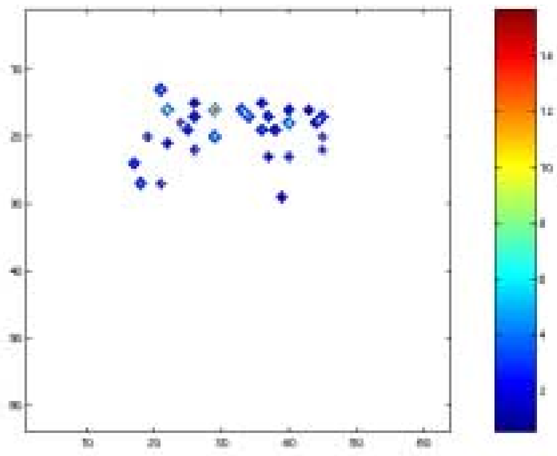
- $\Psi(\cdot)$: non-quadratic function, e.g. l_p norm $(p \leq 1)$
  - Preservation of strong features (source densities)
  - Preference of sparse source density field
  - Can resolve closely spaced radiating sources
- Sensor locations (boundedly) uncertain: $\left\|d(\mathbf{r},\tilde{\mathbf{r}})\right\|_\infty < \varepsilon$
  - Self-calibration capability important
- Potential use in other domains:
  - SAR imaging with unknown motion of the objects in the scene
  - Robust Passive Sonar in the littoral



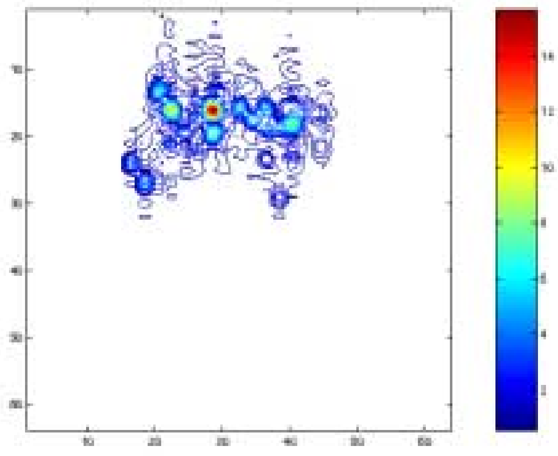
Application in SAR Imaging

- Superresolution Scatterer Localization (synthetic data)

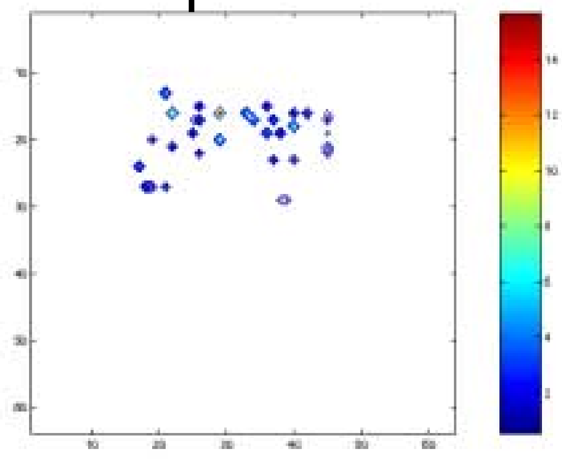
Ground truth



Conventional



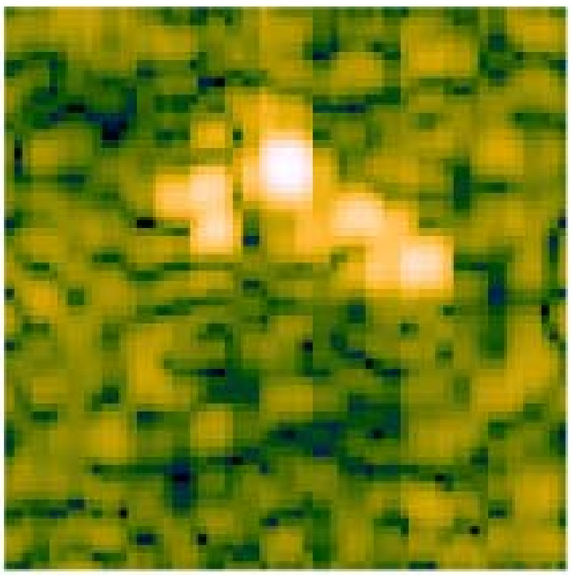
Proposed



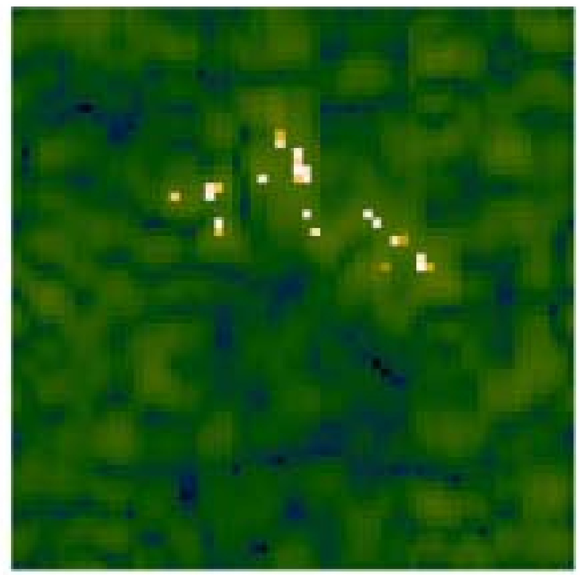
Application in SAR Imaging

- Superresolution Scatterer Localization (real data)

Conventional



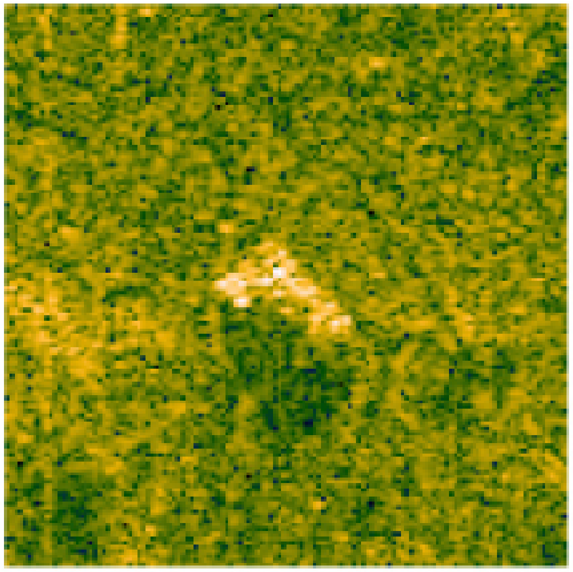
Proposed



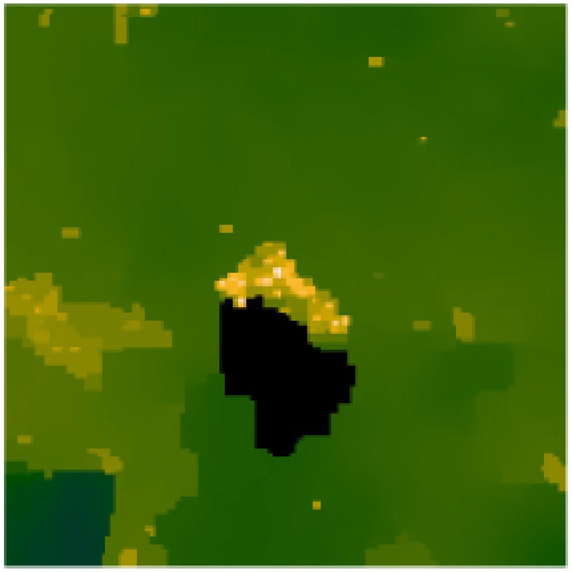
Application in SAR Imaging

- Region-Enhanced Imaging

Conventional

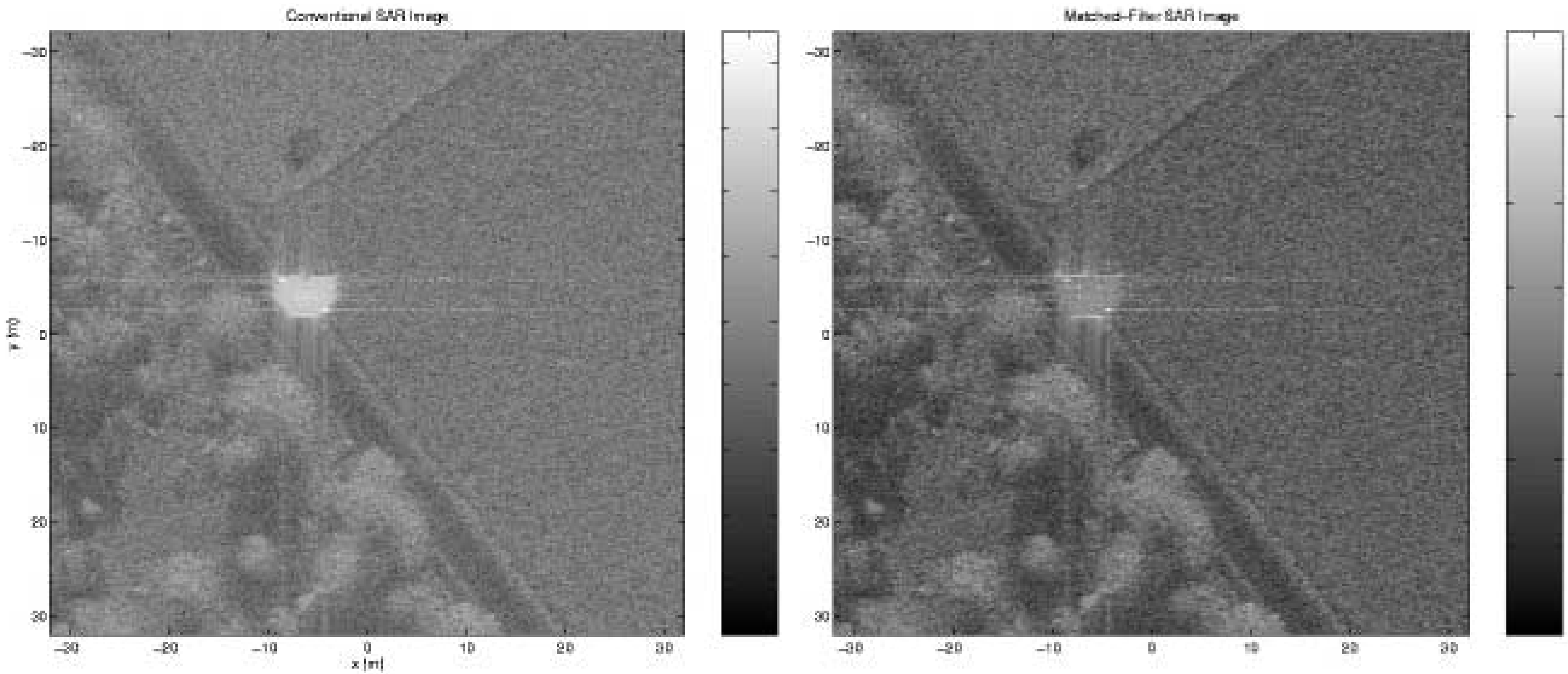


Proposed

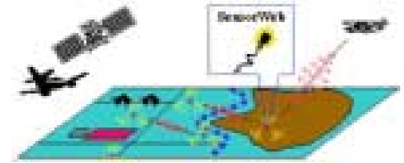


Moving Target Localization in SAR

Conventional Proposed

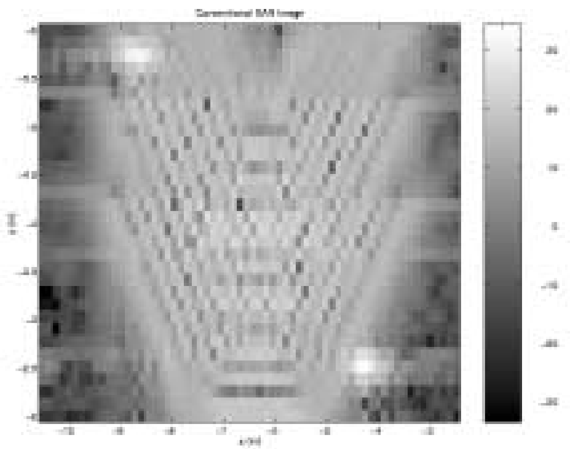


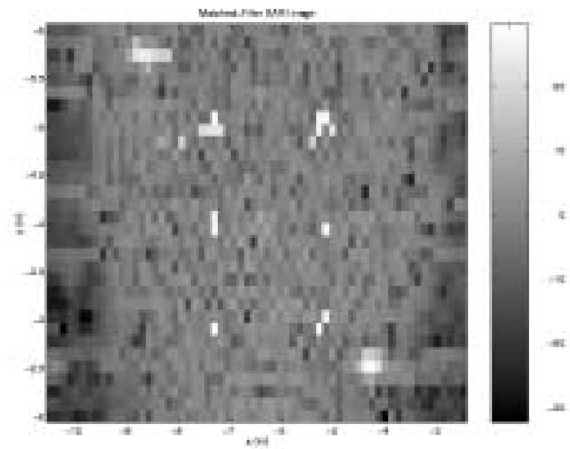
- Scene contains 6 moving and 2 stationary strong point scatterers



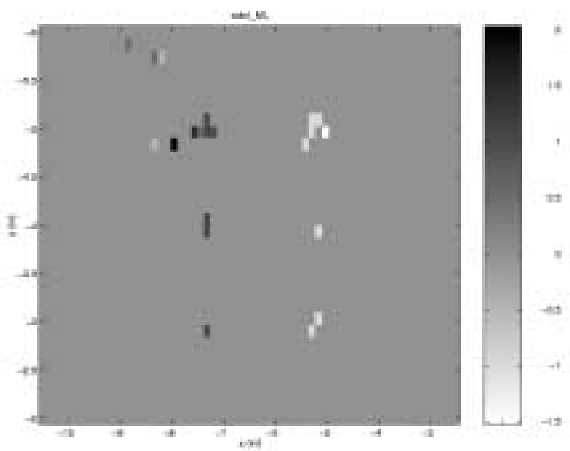
Moving Target Localization in SAR

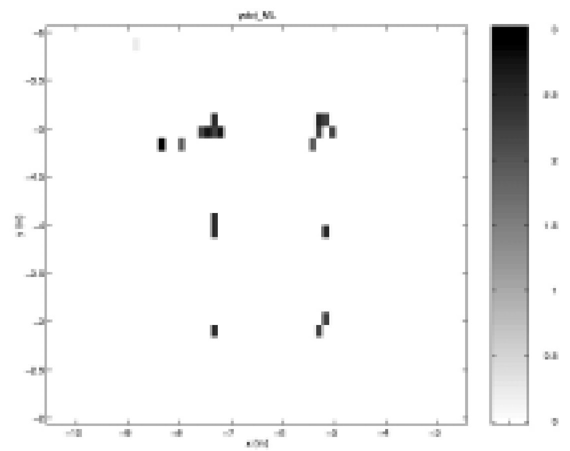
Detailed view

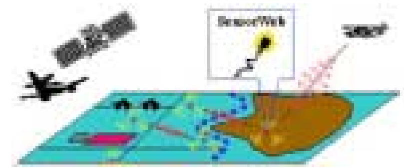




Velocity estimates







Summary and Extensions

- Proposed the development of a variational framework for passive source localization, *robust* to:
  - Limitations in data quality and quantity
  - Uncertainties in sensor locations
- Extensions:
  - <u>Sensors:</u> directional sensitivity, gain/phase uncertainties
  - <u>Signals:</u> structured broadband (e.g. harmonics), unstructured or uncertain broadband
  - <u>Medium:</u> attenuating, dispersive, reverberant