MacSeNet/SpaRTan Spring School on Sparse Representations and Compressed Sensing

Sparse Representations and Dictionary Learning for Source Separation, Localisation, and Tracking

Wenwu Wang

Reader in Signal Processing

Centre for Vision, Speech and Signal Processing
Department of Electronic Engineering
University of Surrey, Guildford
w.wang@surrey.ac.uk
http://personal.ee.surrey.ac.uk/Personal/W.Wang/

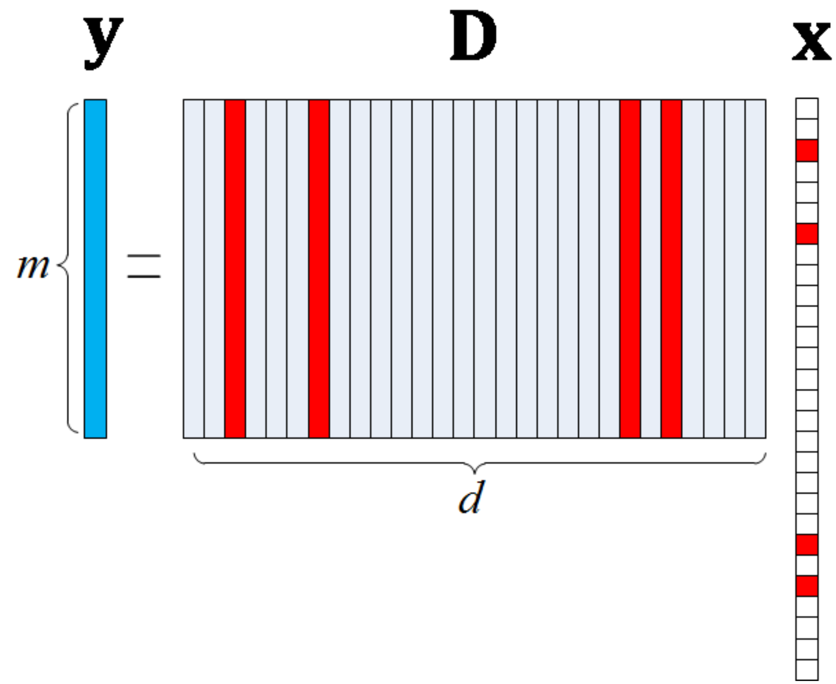
07/04/2016

Contents

- Dictionary Learning
  - Sparse synthesis model (SimCO algorithm)
  - Sparse analysis model (Analysis SimCO algorithm)
- Application Examples
  - Source separation
  - Signal denoising & despeckling
  - Beamforming
  - Multi-speaker tracking
- Future Work

Sparse Synthesis Model

UNIVERSITY OF SURREY



$$\mathbf{y} = \mathbf{D}\mathbf{x} \quad \text{s.t.} \quad \|\mathbf{x}\|_0 = s$$

$\|\cdot\|_0$ — ℓ_0 norm,

the number of non-zero entries

$\mathbf{D} \in \mathbb{R}^{m \times d}$ ---- dictionary

atoms --- columns of $\mathbf{D}$

($m < d$ overcomplete)

$\mathbf{y} \in \mathbb{R}^{m}$ ---- signal

$\mathbf{x} \in \mathbb{R}^{d}$ ---- representation

s ---- sparsity ($s < d$)

Synthesis Sparse Coding

UNIVERSITY OF SURREY

- Task:

Given $\mathbf{y}$ and $\mathbf{D}$, find the sparse representation $\mathbf{x}$

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \ \ \text{s.t.} \ \ \mathbf{y} = \mathbf{D}\mathbf{x}$$

- Existing algorithms:

(1) Greedy algorithms: OMP, SP

(2) Relaxation algorithms: BP

Y. Pati, R. Rezaiifar, and P. Krishnaprasad, “Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition,” in *Proc. 27th Asilomar Conf. Signals, Syst. and Comput.,* pp. 40-44, 1993.

W. Dai and O. Milenkovic, “Subspace pursuit for compressive sensing signal reconstruction,” *IEEE Trans. Inf. Theory*, vol. 55, pp. 2230-2249, 2009.

S. Chen and D. Donoho, “Basis pursuit,” in *Proc. 28th Asilomar Conf. Signals, Syst. and Comput*., vol. 1, pp. 41-44, 1994.

Synthesis Dictionary Learning (SDL)

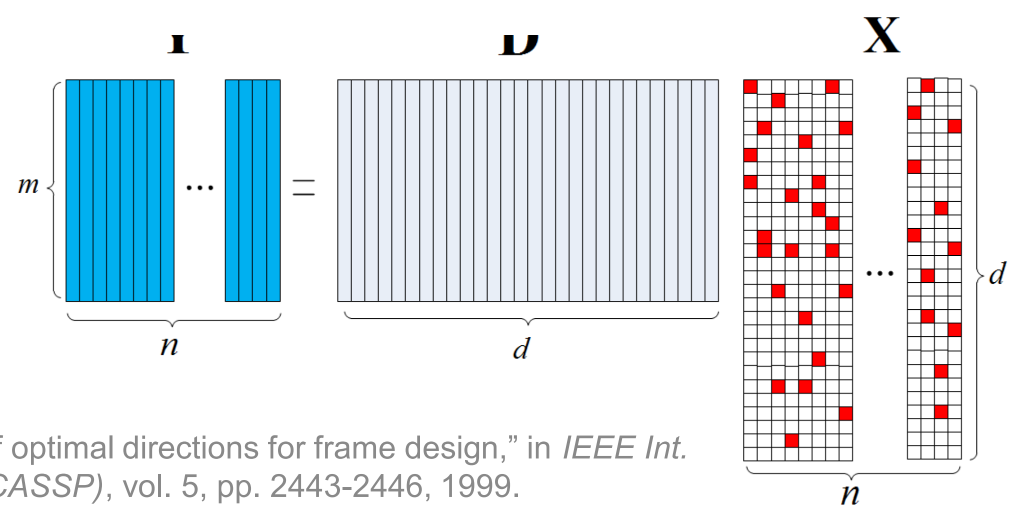
- Task:

Given a set of training signals $\{\mathbf{y}_i\}_{i=1}^{n}$, seek the dictionary $\mathbf{D}$ that leads to the best representation for each member in this set

$$\mathbf{D},\mathbf{X} \quad \sum_{i}$$

- Existing algorithms:

  MOD, K-SVD, SimCO



K. Engan, S. Aase, and J. Hakon Husoy, “Method of optimal directions for frame design,” in *IEEE Int. Conf. on Acoust., Speech, and Signal Processing (ICASSP)*, vol. 5, pp. 2443-2446, 1999.

M. Aharon, m. Elad, and A. Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representations,” *IEEE Trans. Signal Process*., vol. 54, no. 11, pp. 4311-4322, 2006.

SimCO – for synthesis dictionary learning

$$\inf_{\mathbf{D}\in\mathcal{D}} f(\mathbf{D}) = \inf_{\mathbf{D}\in\mathcal{D}} \underbrace{\inf_{\mathbf{X}\in\mathcal{X}(\Omega)} \|\mathbf{Y}-\mathbf{D}\mathbf{X}\|_F^2}_{f(\mathbf{D})}$$

$$\mathcal{D} = \{\mathbf{D}\in\mathbb{R}^{m\times d} : \left\|\mathbf{D}_{:,i}\right\|_2 = 1, i = 1,2,\ldots,d\}$$

$$\mathcal{X}(\Omega) = \left\{\mathbf{X}\in\mathbb{R}^{d\times n} : \mathrm{X}_{i,j} = 0, \forall i \notin \Omega\right\}$$ fixed sparsity pattern

Ω - sparsity pattern (indices of all the non-zeros in $\mathbf{X}$)

- **sparse coding: OMP → for a given $\mathbf{D}$, find $\mathbf{X}$**
- **dictionary learning:**

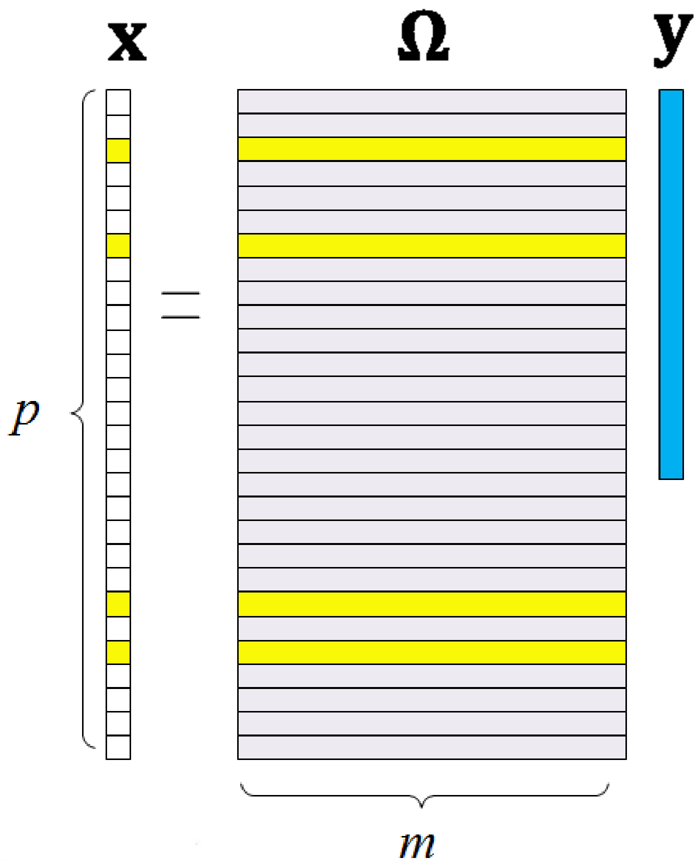
each column in D is one element in Stiefel manifold

(Stiefel manifold: $\mathcal{U}_{m,1} = \{\mathbf{u}\in\mathbb{R}^m : \mathbf{u}^T\mathbf{u} = 1\}$)

optimization on manifolds → D only contains unit ℓ_2-norm columns

W. Dai, T. Xu, and W. Wang, “Simultaneous codeword optimization (SimCO) for dictionary update and learning,” *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6340-6353, 2012.

Sparse Analysis Model



$$\mathbf{x} = \boldsymbol{\Omega}\mathbf{y} \quad \text{s.t.} \quad \|\mathbf{x}\|_0 = p - l$$

$\boldsymbol{\Omega} \in \mathbb{R}^{p \times m}$ --- analysis dictionary

atoms --- rows of $\boldsymbol{\Omega}$

($p > m$ overcomplete)

$\mathbf{y} \in \mathbb{R}^m$ ---- signal

$\mathbf{x} \in \mathbb{R}^p$ ---- representation

l ---- cosparsity

Analysis Pursuit

UNIVERSITY OF SURREY

- Task:

  Recover a signal $\mathbf{y}$ belonging to the analysis model from its measurements

  (1) recovery from noisy measurements:

$$\hat{\mathbf{y}} = \text{argmin}_{\mathbf{y}} \ \|\boldsymbol{\Omega}\mathbf{y}\|_0 \ \ \text{s.t.}\ \mathbf{z} = \mathbf{y} + \mathbf{v}$$

  (2) recovery from incomplete measurements with noise:

$$\hat{\mathbf{y}} = \text{argmin}_{\mathbf{y}} \ \|\boldsymbol{\Omega}\mathbf{y}\|_0 \ \ \text{s.t.}\ \mathbf{z} = \mathbf{M}\mathbf{y} + \mathbf{v}$$

- Existing algorithms: BG, OBG; GAP

R. Rubinstein, T. Peleg, and M. Elad, “Analysis K-SVD: A dictionary-learning algorithm for the analysis sparse model,” *IEEE Trans. Signal Process*., vol. 61, no. 3, pp. 661-677, 2013.

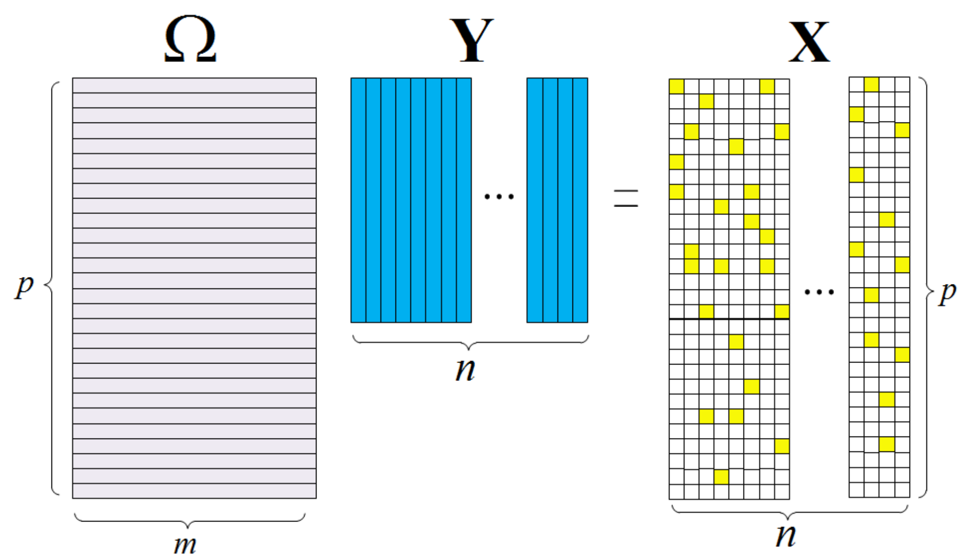
S. Nam, M. E. Davies, M. Elad, and R. Gribonval, “The cosparse analysis model and algorithms,” *Appl. Comput. Harm. Anal.*, vol. 34, no. 1, pp. 30-56, 2013.

Analysis Dictionary Learning (ADL)

- Task:

Given a set of training signals $\{\mathbf{y}_i\}_{i=1}^{n}$, seek the analysis dictionary $\boldsymbol{\Omega}$ so that the analysis representations of the signals can be as sparse as possible.

$$\min_{\boldsymbol{\Omega}} \sum_{i} \|\boldsymbol{\Omega}\mathbf{y}_i\|_0, 1 \le i \le n$$



Analysis Dictionary Learning (ADL)

- Existing algorithms:

(1) Analysis K-SVD:

high computational complexity

(2) AOL:

exclude the feasible dictionaries outside UNTF

(3) LOST:

less effective in reaching the pre-defined cosparsity

R. Rubinstein, T. Peleg, and M. Elad, “Analysis K-SVD: A dictionary-learning algorithm for the analysis sparse model,” *IEEE Trans. Signal Process.,* vol. 61, no. 3, pp. 661-677, 2013.

M. Yaghoobi, S. Nam, R. Gribonval, and M. Davies, “Constrained overcomplete analysis operator learning for cosparse signal modelling,” *IEEE Trans. Signal Process.,* vol. 61, no. 9, pp. 2341-2355, 2013.

S. Ravishankar and Y. Bresler, “Learning overcomplete sparsifying transforms for siangl processing,” in *IEEE Int. Conf. on Acoust., Speech, and Signal Processing (ICASSP)*, pp. 3088-3092, 2013.

Analysis SimCO Algorithm

UNIVERSITY OF SURREY

- cost function:

$$\min_{\mathbf{X},\mathbf{\Omega}} \|\mathbf{X} - \mathbf{\Omega}\mathbf{Y}\|_F^2$$

$$\text{s.t. } \forall i, \left\|\mathbf{X}_{:,i}\right\|_0 = p - l,$$

$$\forall j, \left\|\mathbf{\Omega}_{j,:}\right\|_2 = 1$$

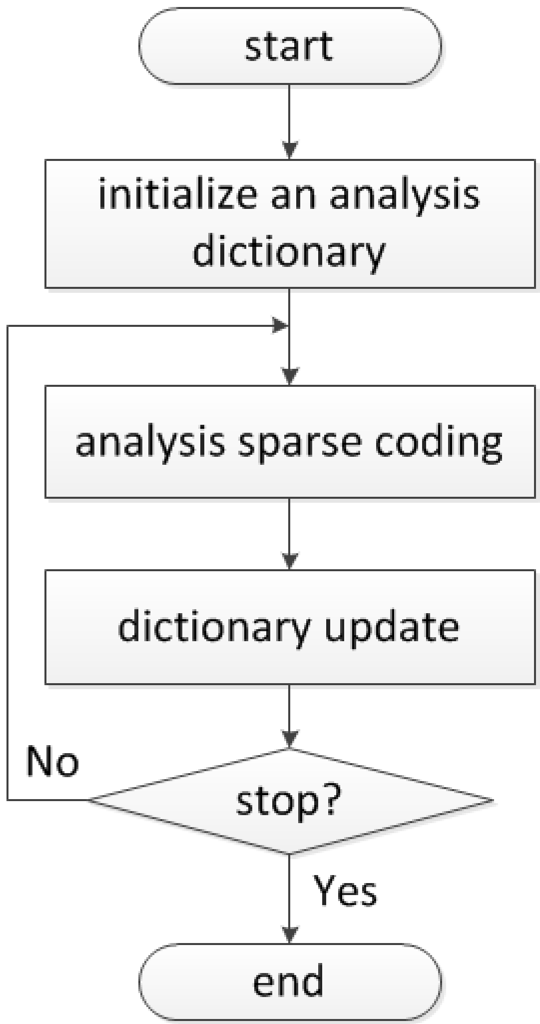
$\mathbf{\Omega}$ contains unit ℓ_2 - norm rows

- **two separate optimisation problems on $\mathbf{X}$ and $\mathbf{\Omega}$ respectively by keeping one fixed and changing the other.**

- **the transpose of each row in $\mathbf{\Omega}$ is one element in Stiefel manifold → modify the optimization framework of SimCO to update $\mathbf{\Omega}$**

W. Dai, T. Xu, and W. Wang, "Simultaneous codeword optimisation (SimCO) for dictionary update and learning", *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6340-6353, 2012.

Analysis SimCO framework



Input: $\mathbf{Y}, p, l$

Output: $\boldsymbol{\Omega} = \boldsymbol{\Omega}^{k+1}$

Initialization: $k = 0, \widehat{\boldsymbol{\Omega}} = \boldsymbol{\Omega}^{k}$

Main Iteration:

(1) $\mathbf{X}^{k} = \boldsymbol{\Omega}^{k}\mathbf{Y}$

(2) $\widehat{\mathbf{X}}^{k} = HT_{l}(\mathbf{X}^{k})$

(3) $\boldsymbol{\Omega}^{k+1} \leftarrow \boldsymbol{\Omega}^{k}$

(4) $k = k + 1$

(5) If the stopping criterion is satisfied, quit the iteration. Otherwise, apply another iteration.

Analysis SimCO – Dictionary Update

UNIVERSITY OF SURREY

$$\min_{\mathbf{\Omega}} f(\mathbf{\Omega}) = \min_{\mathbf{\Omega}} \|\mathbf{X} - \mathbf{\Omega}\mathbf{Y}\|_F^2 \quad \text{s.t.} \quad \forall j, \|\mathbf{\Omega}_{j,:}\|_2 = 1$$

- **Search Direction:** $\mathbf{H} = -\nabla f(\mathbf{\Omega}) = -\frac{\partial \|\mathbf{X}-\mathbf{\Omega}\mathbf{Y}\|_F^2}{\partial \mathbf{\Omega}} = 2\mathbf{X}\mathbf{Y}^T - 2\mathbf{\Omega}\mathbf{Y}\mathbf{Y}^T$

- **Line Search Path:**

$$\bar{\mathbf{h}}_j = \mathbf{h}_j - \mathbf{h}_j \mathbf{\Omega}_{j,:}^T \mathbf{\Omega}_{j,:}, \forall j \in \{1,2,\ldots,p\} \quad (\bar{\mathbf{h}}_j \mathbf{\Omega}_{j,:}^T = \mathbf{\Omega}_{j,:} \bar{\mathbf{h}}_j^T = 0)$$

$$\begin{cases} \mathbf{\Omega}_{j,:}(\mathrm{t}) = \mathbf{\Omega}_{j,:}, \text{if } \|\bar{\mathbf{h}}_j\|_2 = 0 \\ \mathbf{\Omega}_{j,:}(\mathrm{t}) = \mathbf{\Omega}_{j,:} \cos\left(\|\bar{\mathbf{h}}_j\|_2 t\right) + \left(\bar{\mathbf{h}}_j / \|\bar{\mathbf{h}}_j\|_2\right) \sin\left(\|\bar{\mathbf{h}}_j\|_2 t\right), \text{if } \|\bar{\mathbf{h}}_j\|_2 \neq 0 \end{cases}$$

- **Step Size: golden section search, find a proper step size** t

J. Dong, W. Wang, W. Dai, M. Plumbley, Z. Han, and J. A. Chambers, "Analysis SimCO algorithms for sparse analysis model based dictionary learning", *IEEE Transactions on Signal Processing*, vol. 64, no. 2, pp. 417 - 431, 2016.

Implementation

- Matlab toolbox of dictionary learning algorithms: SimCO
  - The toolbox contains implementation of multiple dictionary learning algorithms including our own algorithms primitive SimCO and regularised SimCO algorithms, as well as baseline algorithms including K-SVD, and MOD.
  - The toolbox has been made publicly available in compliance with EPSRC open access policy. Web address: http://personal.ee.surrey.ac.uk/Personal/W.Wang/codes/SimCO.html

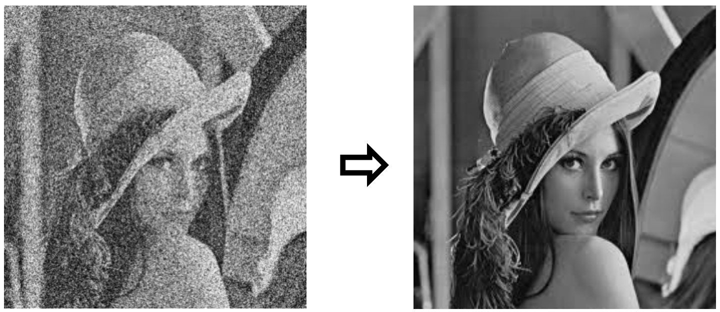
Implementation (cont.)

UNIVERSITY OF SURREY

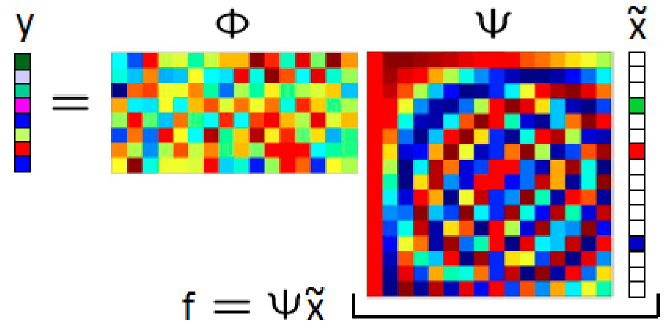
- Matlab toolbox of analysis dictionary learning algorithms: Analysis SimCO
  - The toolbox contains implementation of multiple dictionary learning algorithms including our own algorithms Analysis SimCO, Incoherent Analysis SimCO algorithms, as well as several baseline algorithms including Analysis K-SVD, LOST, GOAL, AOL, TK-SVD.
  - The toolbox has been made publicly available in compliance with EPSRC open access policy. Web address: http://dx.doi.org/10.15126/surreydata.00808101

Potential Applications

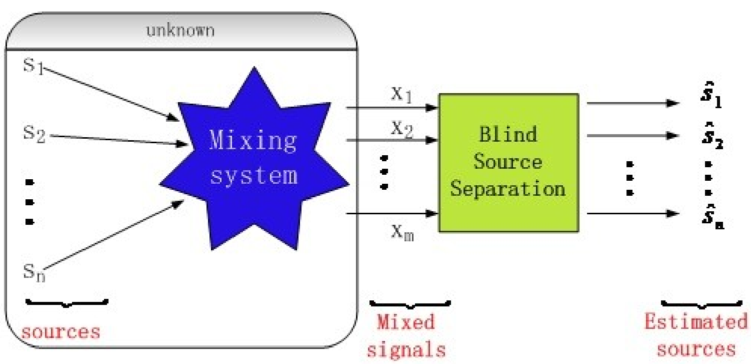
- Image denoising



- Compressed Sensing



- Blind Source Separation



- Image compression
- Inpainting
- Recognition
- Beamforming

…….

Selected Examples

- Signal denoising
- Source separation
- Beamforming
- Multi-speaker tracking

Denoising Examples

W. Dai, T. Xu, and W. Wang, “Simultaneous codeword optimization (SimCO) for dictionary update and learning,” *IEEE Trans. Signal Process*., vol. 60, no. 12, pp. 6340-6353, 2012.

Natural Image Denoising

Test images

Training images

PSNR Results

	$\sigma = 45$ (Input PSNR ~ 15 dB)			
Training data type	Type I		Type II	
co-sparsity l	40	80	40	80
ASimCO	25.73	24.24	22.44	24.52
IN-ASimCO	25.74	25.37	22.30	24.37
ASimCO-Random	25.54	25.71	22.57	22.71
ASimCO-IKSVD	22.22	22.53	22.17	22.37
AKSVD	22.17	—	22.18	—
LOST	22.17	22.39	22.17	22.27
TKSVD	22.17	22.19	22.18	23.11
(NA)AOL	23.54		22.18	
GOAL	23.85		22.19	

J. Dong, W. Wang, W. Dai, M. Plumbley, Z. Han, and J. A. Chambers, "Analysis SimCO algorithms for sparse analysis model based dictionary learning", *IEEE Transactions on Signal Processing*, vol. 64, no. 2, pp. 417 - 431, 2016.

Despeckling – Signal Model

Signal model:

$$\mathbf{w} = \mathbf{g} \circ \mathbf{u}$$

$$f_u(u) = \frac{L^L}{\Gamma(L)} u^{L-1} e^{-Lu} \qquad \Gamma(L) = (L-1)!$$

Transformed model:

$$\underbrace{\log \mathbf{w}}_{\mathbf{z}} = \underbrace{\log \mathbf{g}}_{\mathbf{y}} + \underbrace{\log \mathbf{u}}_{\mathbf{v}}.$$

Optimisation problem:

$$\mathbf{Y}^* = \arg\min_{\mathbf{Y}} \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\mathbf{Y}_{i,j} + e^{\mathbf{Z}_{i,j} - \mathbf{Y}_{i,j}} \right) + \lambda \left\| \mathbf{\Omega} \mathbf{Y} \right\|_1.$$

Despeckling – Signal Recovery

UNIVERSITY OF SURREY

Alternating direction method of multipliers (ADMM):

$$\arg\min_{\mathbf{Y}} \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\mathbf{Y}_{i,j} + e^{\mathbf{Z}_{i,j}-\mathbf{Y}_{i,j}}\right) + \lambda \|\mathbf{T}\|_1$$

$$\text{s. t. } \mathbf{T} = \mathbf{\Omega Y}$$

Augumented Lagrangian function of the above function:

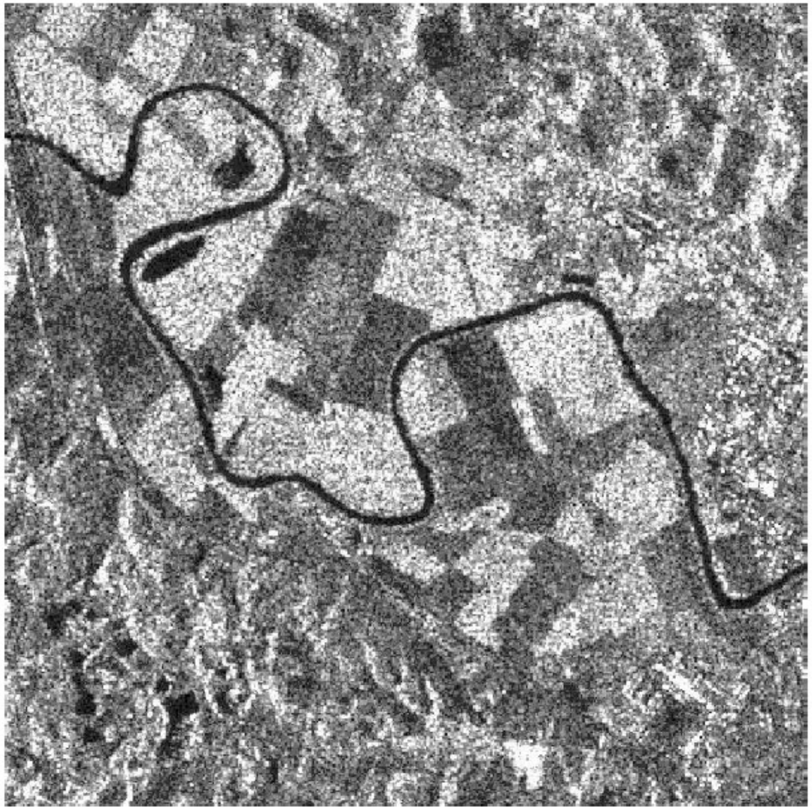
$$\sum_{i=1}^{m} \sum_{j=1}^{n} \left(\mathbf{Y}_{i,j} + e^{\mathbf{Z}_{i,j}-\mathbf{Y}_{i,j}}\right) + \lambda \|\mathbf{T}\|_1 + \gamma \langle \mathbf{B}, \mathbf{\Omega Y} - \mathbf{T} \rangle + \frac{\gamma}{2} \|\mathbf{\Omega Y} - \mathbf{T}\|_F^2$$

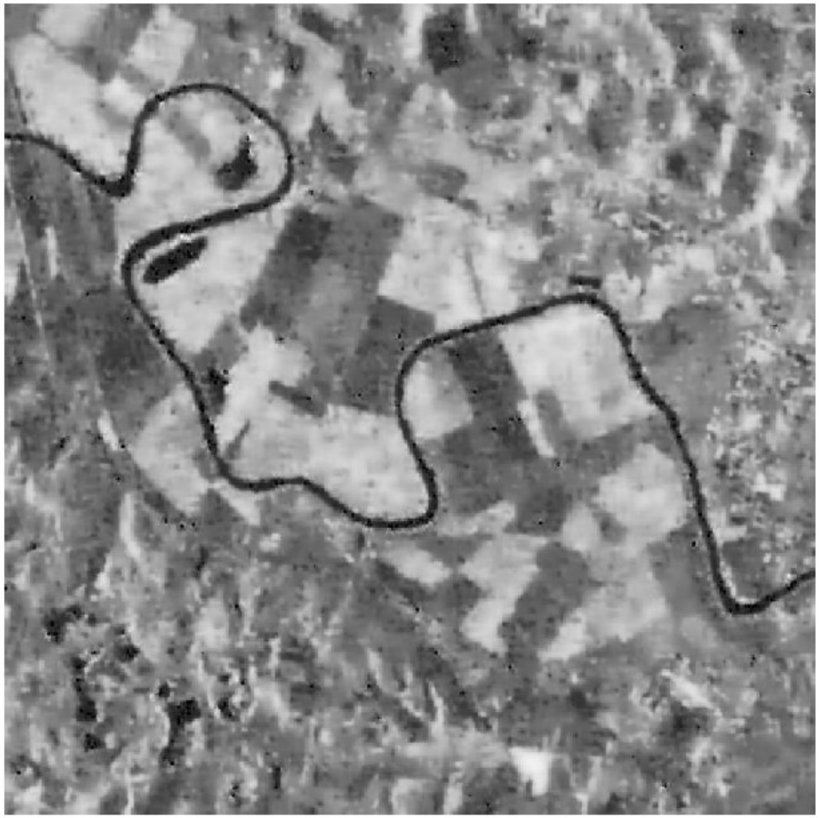
$$= \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\mathbf{Y}_{i,j} + e^{\mathbf{Z}_{i,j}-\mathbf{Y}_{i,j}}\right) + \lambda \|\mathbf{T}\|_1 + \frac{\gamma}{2} \|\mathbf{B} + \mathbf{\Omega Y} - \mathbf{T}\|_F^2 - \frac{\gamma}{2} \|\mathbf{B}\|_F^2$$

The ADMM algorithm iteratively updates each of the variables $\{\mathbf{Y}, \mathbf{T}, \mathbf{B}\}$ while keeping the rest fixed.

The restored log-image $\hat{\mathbf{y}}$ can be obtained by reshaping the solution $\mathbf{Y}^*$, and thus the denoised image $\hat{\mathbf{g}}$ is obtained by taking the exponential transform of $\hat{\mathbf{y}}$.

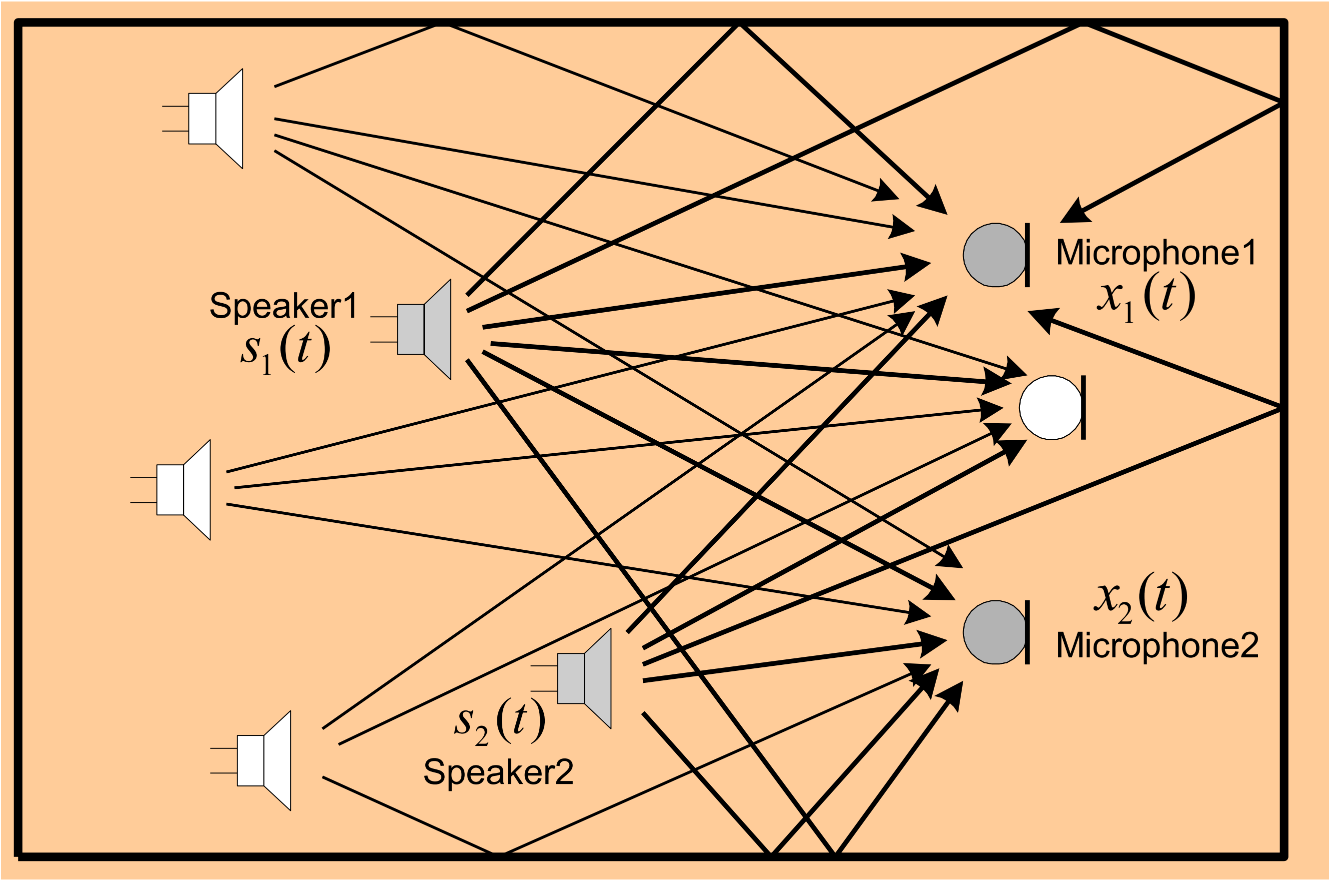
Despeckling – Real SAR Images



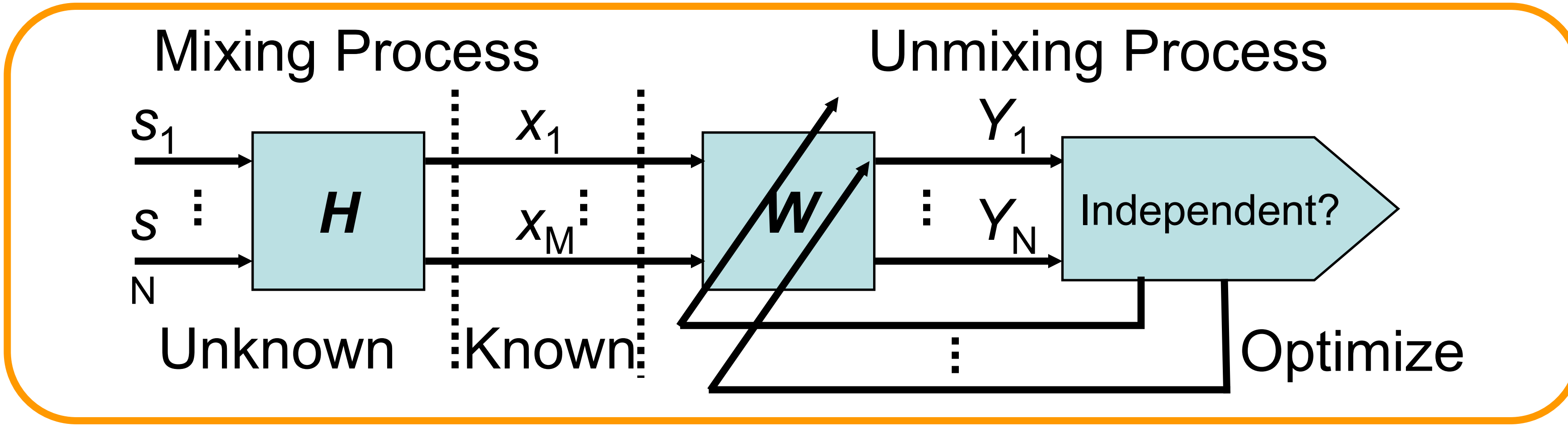


J. Dong, W. Wang, J. A. Chambers, "Removing speckle noise by analysis dictionary learning", in *Proc. IEEE Sensor Signal Processing for Defence* (SSPD 2015), Edinburgh, UK, September 9-10, 2015.

Source Separation: Cocktail party problem



Blind Source Separation & Independent Component Analysis



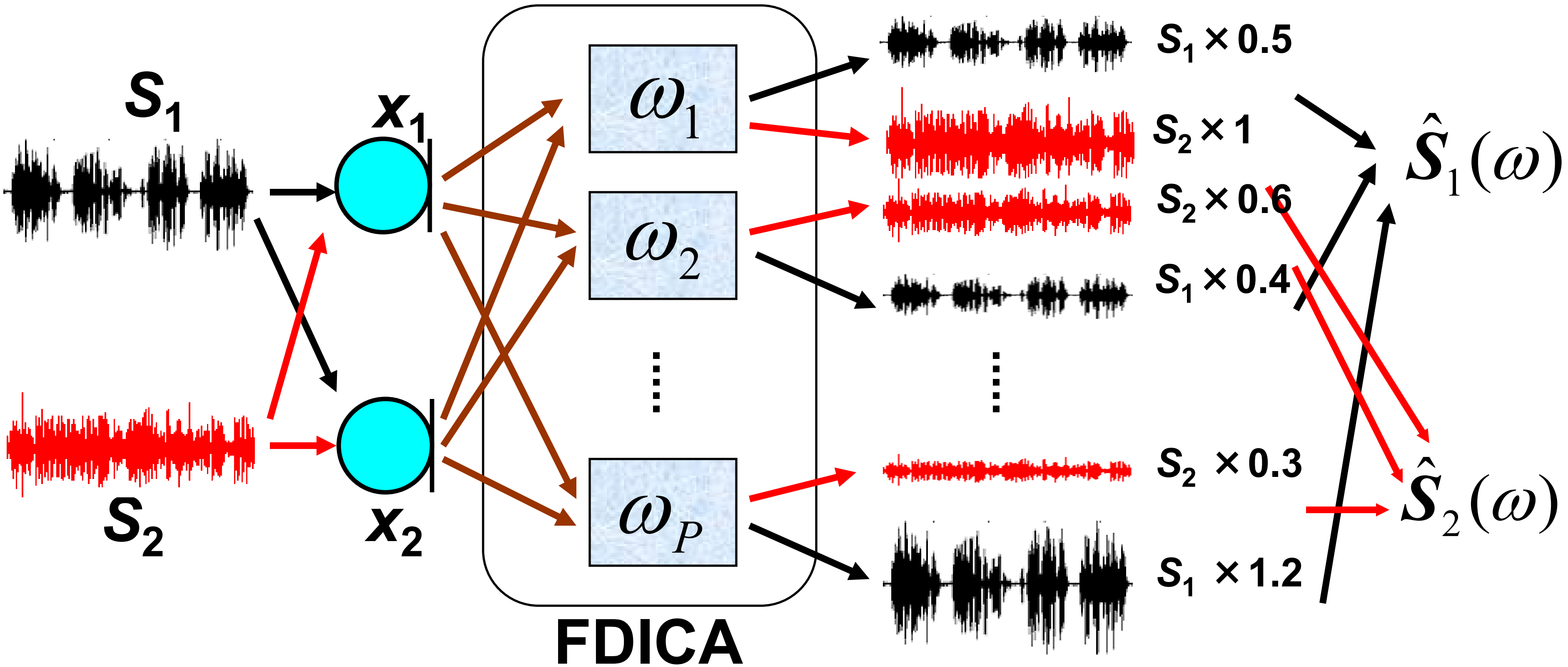
Mixing Model: $\mathbf{x} = \mathbf{Hs}$

De-mixing Model: $\mathbf{y} = \mathbf{Wx} = \mathbf{WHs} = \mathbf{PDs}$

Diagonal Scaling Matrix

Permutation Matrix

Frequency Domain BSS & Permutation Problem



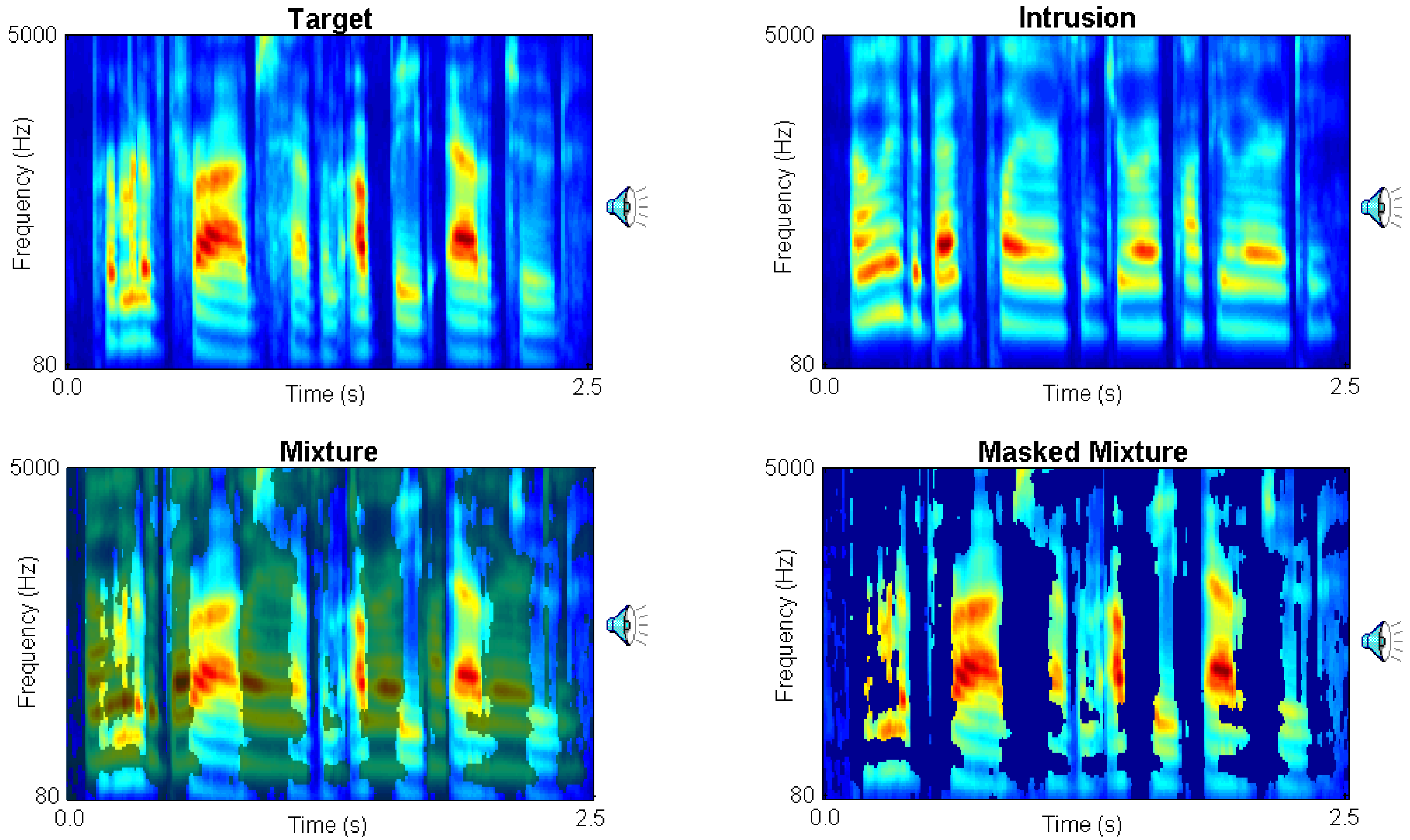
Solutions:

- Beamforming
- Spectral envelope correlation

Computational Auditory Scene Analysis

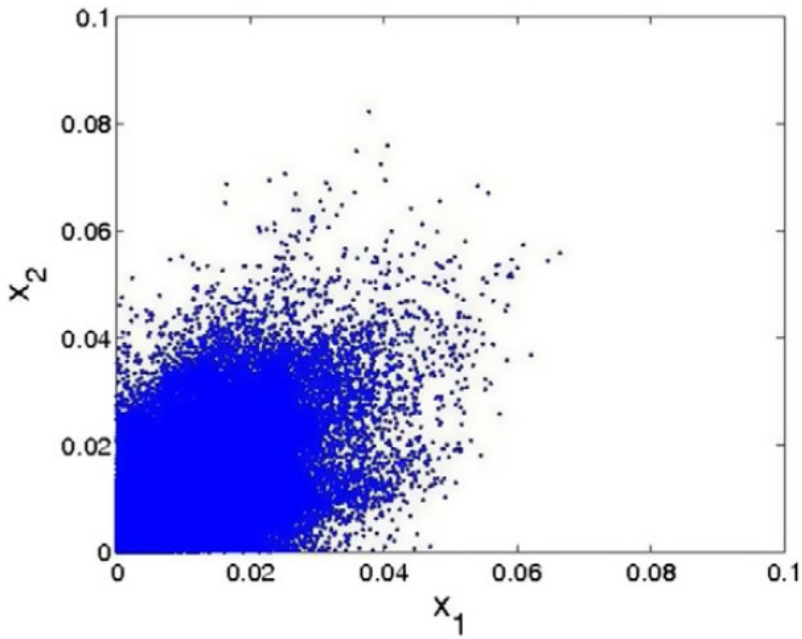
- Computational models for two conceptual processes of auditory scene analysis (ASA):
  - **Segmentation**. Decompose the acoustic mixture into sensory elements (segments)
  - **Grouping**. Combine segments into groups, so that segments in the same group likely originate from the same sound source

CASA – Time-Frequency Masking

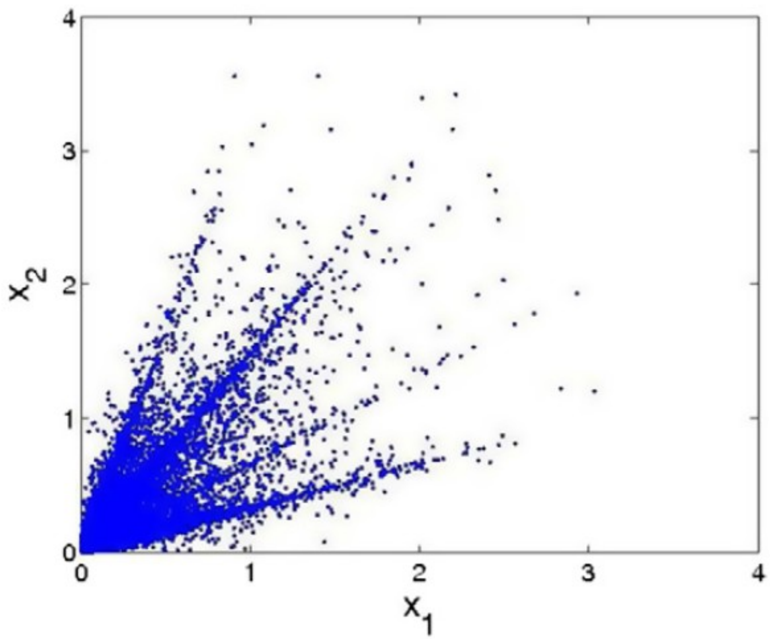


Demos due to Deliang Wang. Recent psychophysical tests show that the ideal binary mask results in dramatic speech intelligibility improvements (Brungart et al.’06; Li & Loizou’08)

Underdetermined Source Separation



Time domain



Time-frequency domain

$$\begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \end{pmatrix} \begin{pmatrix} \boldsymbol{s}_1 \\ \vdots \\ \boldsymbol{s}_4 \end{pmatrix}$$

Source Separation as a Sparse Recovery Problem

UNIVERSITY OF SURREY

Reformulation:

$$\underbrace{\begin{pmatrix} x_1(1) \\ \vdots \\ x_1(T) \\ \vdots \\ \vdots \\ x_M(1) \\ \vdots \\ x_M(T) \end{pmatrix}}_{\boldsymbol{b}} = \underbrace{\begin{pmatrix} \Lambda_{11} & \cdots & \Lambda_{1N} \\ \vdots & \ddots & \vdots \\ \Lambda_{N1} & \cdots & \Lambda_{MN} \end{pmatrix}}_{\boldsymbol{M}} \underbrace{\begin{pmatrix} s_1(1) \\ \vdots \\ s_1(T) \\ \vdots \\ \vdots \\ s_N(1) \\ \vdots \\ s_N(T) \end{pmatrix}}_{\boldsymbol{f}}$$

- The above problem can be interpreted as a signal recovery problem in compressed sensing, where $\boldsymbol{M}$ is a measurement matrix, and $\boldsymbol{b}$ is a compressed vector of samples in $\boldsymbol{f}$. Λ_{ij} is a diagonal matrix whose elements are all equal to a_{ij}.
- A sparse representation may be employed for $\boldsymbol{f}$, such as:

$$\boldsymbol{f} = \boldsymbol{\Phi c}$$

- $\boldsymbol{\Phi}$ is a transform dictionary, and $\boldsymbol{c}$ is the weighting coefficients corresponding to the dictionary atoms.

Source Separation as a Sparse Recovery Problem (cont.)

UNIVERSITY OF SURREY

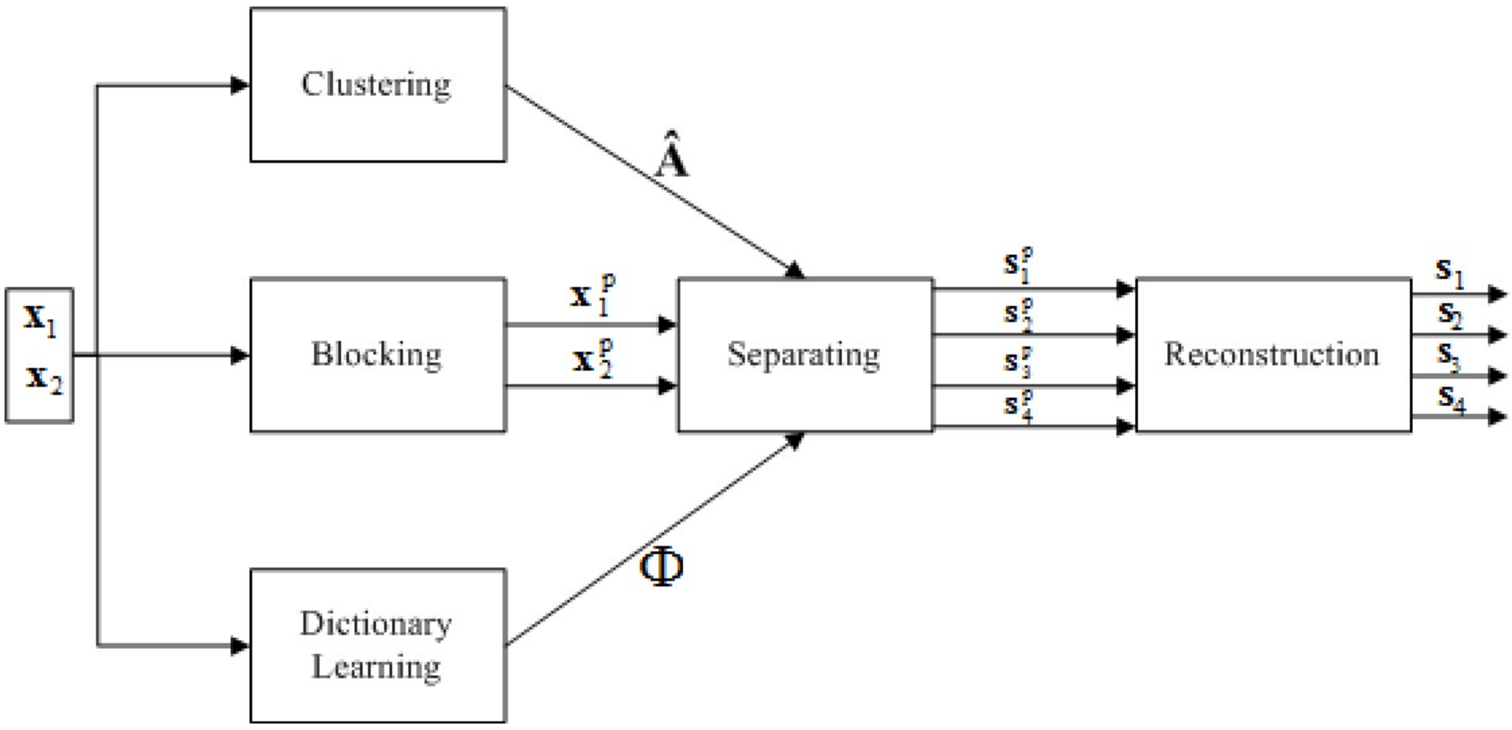
Reformulation:

$$\boldsymbol{b} = \overline{\boldsymbol{M}}\boldsymbol{c} \quad \text{and} \quad \overline{\boldsymbol{M}} = \boldsymbol{M}\boldsymbol{\Phi}$$

- According to compressed sensing, if $\overline{\boldsymbol{M}}$ satisfies the restricted isometry property (RIP), and also ***c*** is sparse, the signal ***f*** can be recovered from ***b*** using an optimisation process.
- This indicates that source estimation in the underdetermined problem can be achieved by computing ***c*** using signal recovery algorithms in compressed sensing, such as:
  - ✓ Basis pursuit (BP) (Chen et al., 1999)
  - ✓ Matching pursuit (MP) (Mallat and Zhang, 1993)
  - ✓ Orthogonal matching pursuit (OMP) (Pati et al., 1993)
  - ✓ L1 norm least squares algorithm (L1LS) (Kim et al., 2007)
  - ✓ Subspace pursuit (SP) (Dai et al., 2009)
  - ✓ …

Dictionary Learning for Underdetermined Source Separation

Separation system for the case of M = 2 and N =4:



Source Separation – Sound Demo

s1 s2 s3 s4

x1 x2

es1 es2 es3 es4

T. Xu, W. Wang, and W. Dai, Compressed sensing with adaptive dictionary learning for underdetermined blind speech separation, *Speech Communication*, vol. 55, pp. 432-450, 2013.

Beamforming – Sparse Representation Formulation

- Extends the classic Bayesian approach to a sequential maximum a posterior (MAP) estimation of the signal over time.
- Sparsity constraint is enforced with a Laplacian like prior at each time step.
- An adaptive LASSO cost function is minimised at each time step *k* for *M* array sensors

$$\zeta_k(\mathbf{x}_k) = \frac{\|(\mathbf{y}_k - \mathbf{A}\mathbf{x}_k)\|_2^2}{\sigma^2} + \mu \sum_{m=1}^{M} w_{km} |x_{km}|$$

C. Mecklenbruker, P. Gerstoft, A. Panahi, M. Viberg, "Sequential Bayesian sparse signal reconstruction using array data," *IEEE Transactions on Signal Processing*, vol. 61, no. 24, pp. 6344 - 6354, 2013.

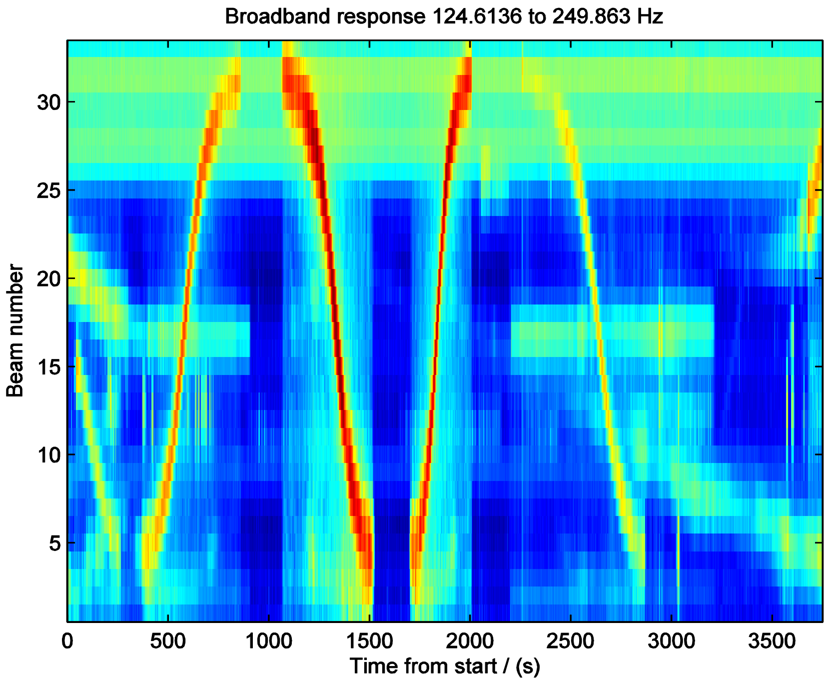
Beamforming – Portland03 Underwater Acoustic Dataset

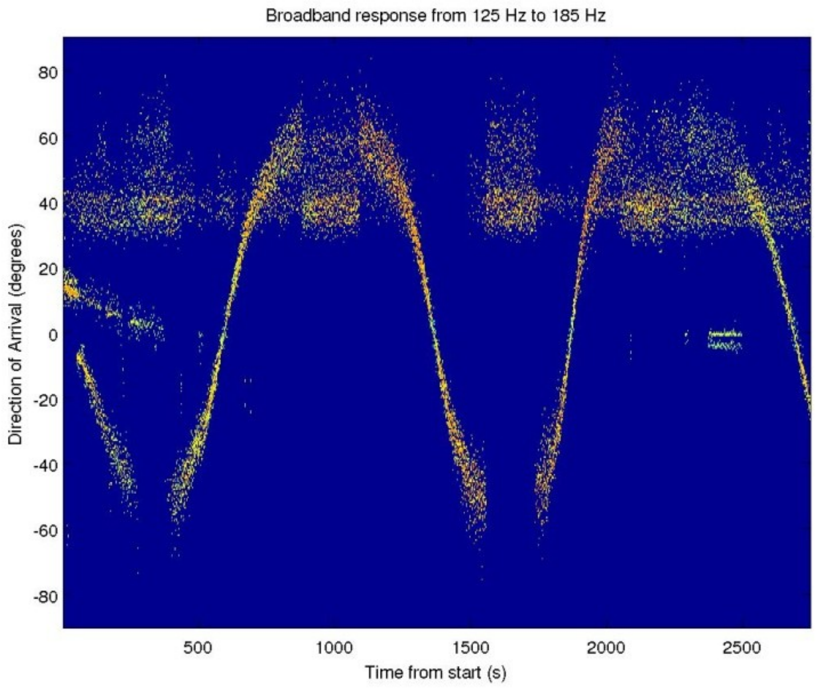
UNIVERSITY OF SURREY

- Data collected in 2003 in Portland harbour
- 31 element linear hydrophone array on the sea floor
- Single moving target: Sequence one "Beam-on" to the array, Sequence two "end-fire" to the array

Beamforming – Portland03 Underwater Acoustic Dataset

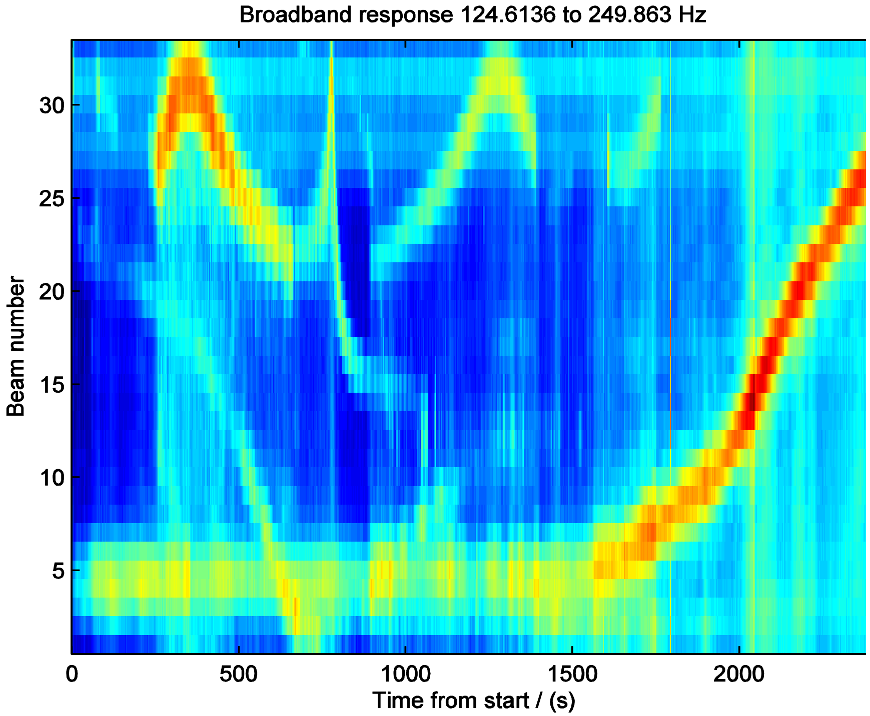
Sequence one – One target moving beam-on to the array

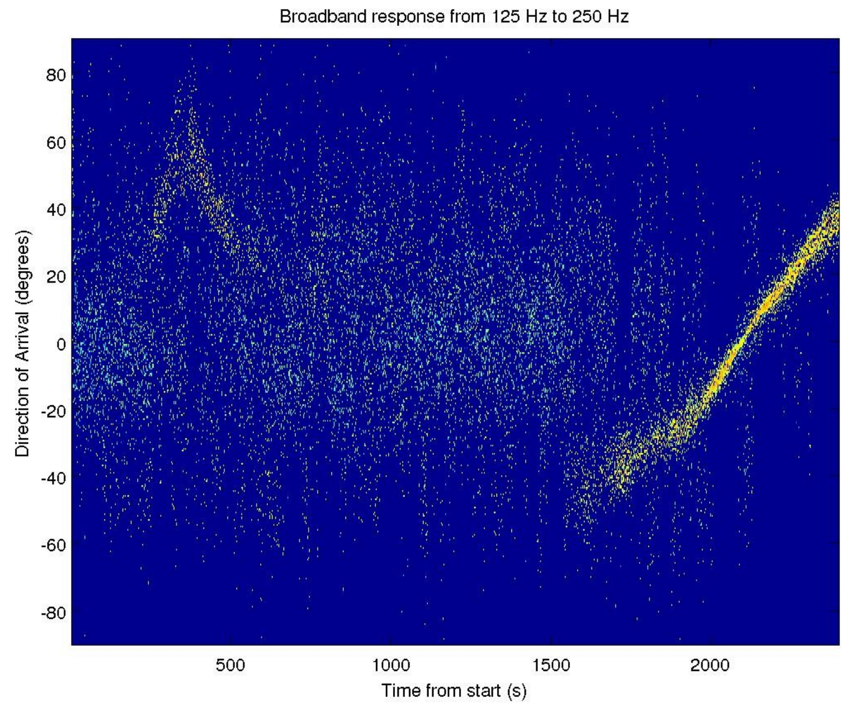




Beamforming – Portland03 Underwater Acoustic Dataset

Sequence two – One target moving end-fire to the array





M. Barnard and W. Wang, "Sequential Bayesian sparse reconstruction algorithms for underwater acoustic signal denoising" *Proc. IET Conference on Intelligent Signal Processing*, December, 2015.

Multi-Speaker Tracking

UNIVERSITY OF SURREY

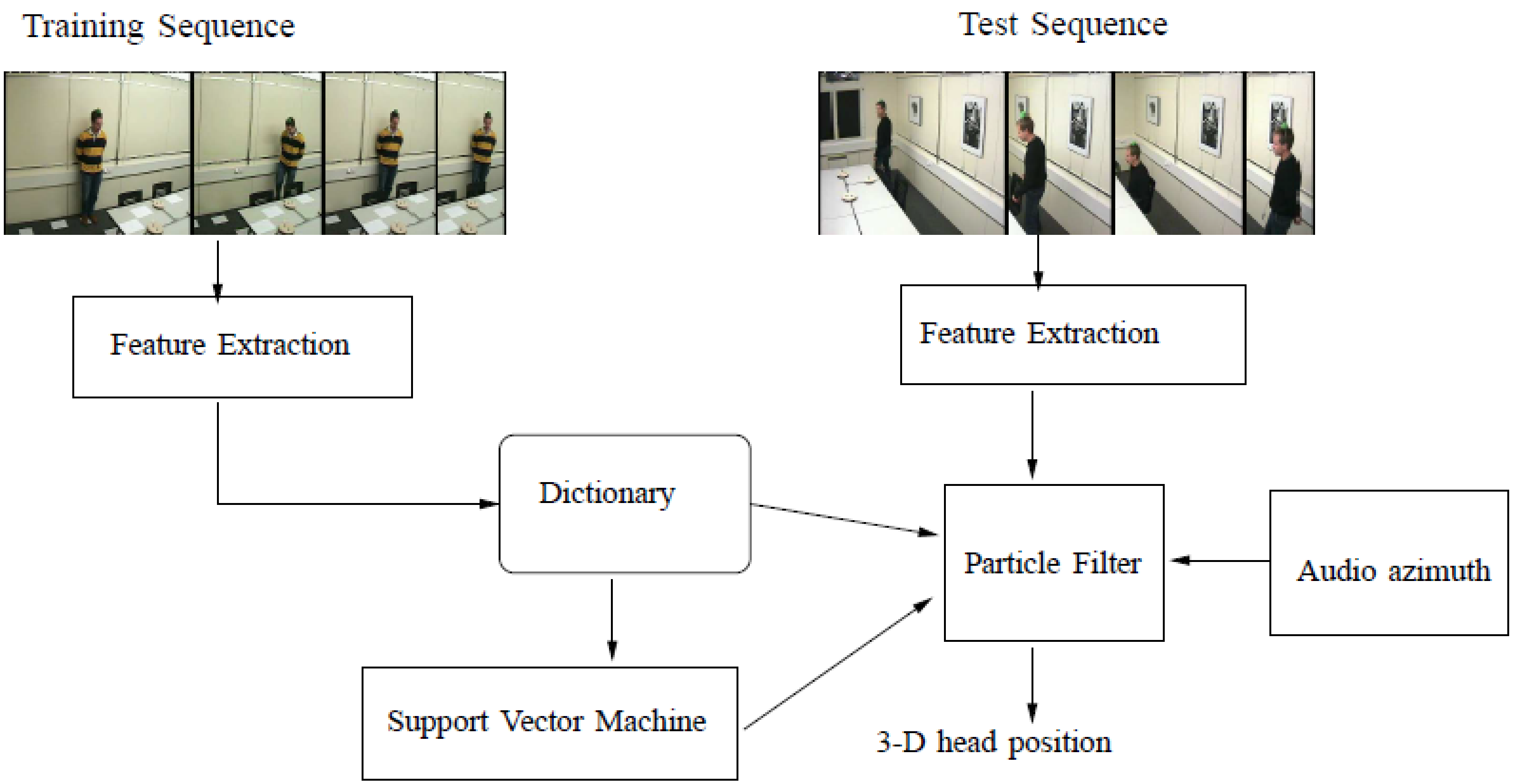
Challenges:

- Modelling the appearance of the moving speakers (or more broadly, moving objects) under different (office) environments with a variety of lighting conditions and camera resolutions.
- Dealling with occlusions when tracking multiple speakers.
- Dealing with the loss of visual trackers due to e.g. the lost view of the cameras.

Proposed solutions:

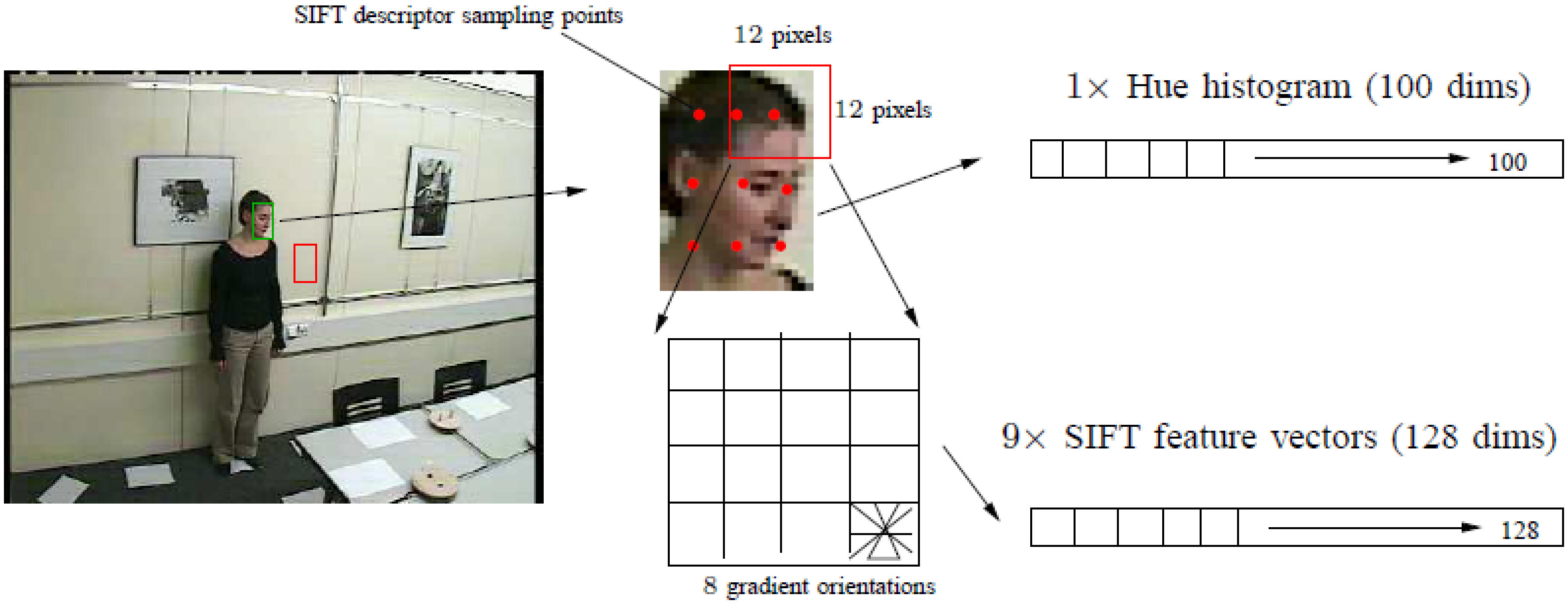
- Appearance modelling based on dictionary learning
- Incorporating identity models of speakers e.g. based on Gaussian mixture models (GMM) (not to discuss in this talk)
- Audio assisted re-initialisation of visual tracker (or re-booting of lost visual tracker)

Dictionary Learning based Method



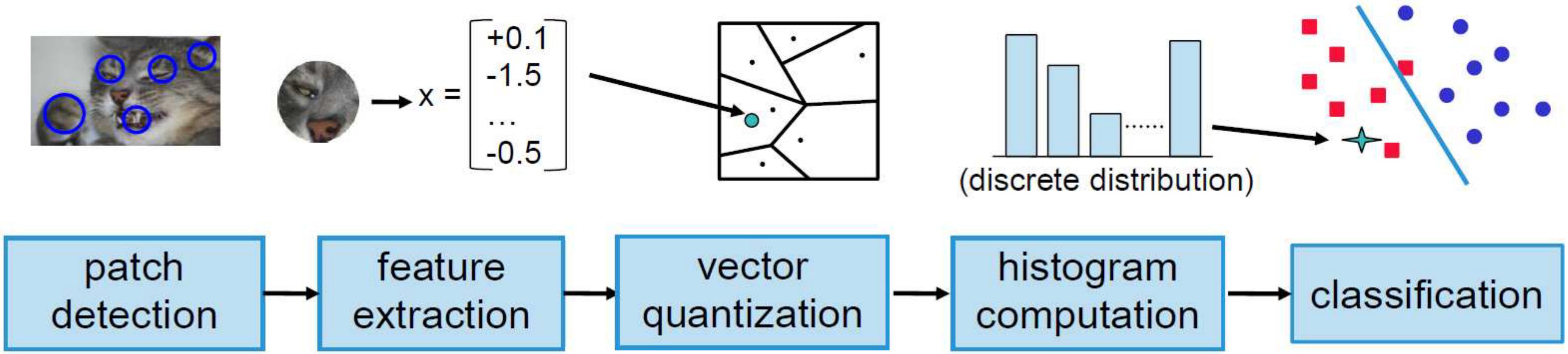
Overall system to generate the 3-D head position, showing training and testing (i.e. tracking) phases.

Feature Extraction



Extraction of features from image patches.

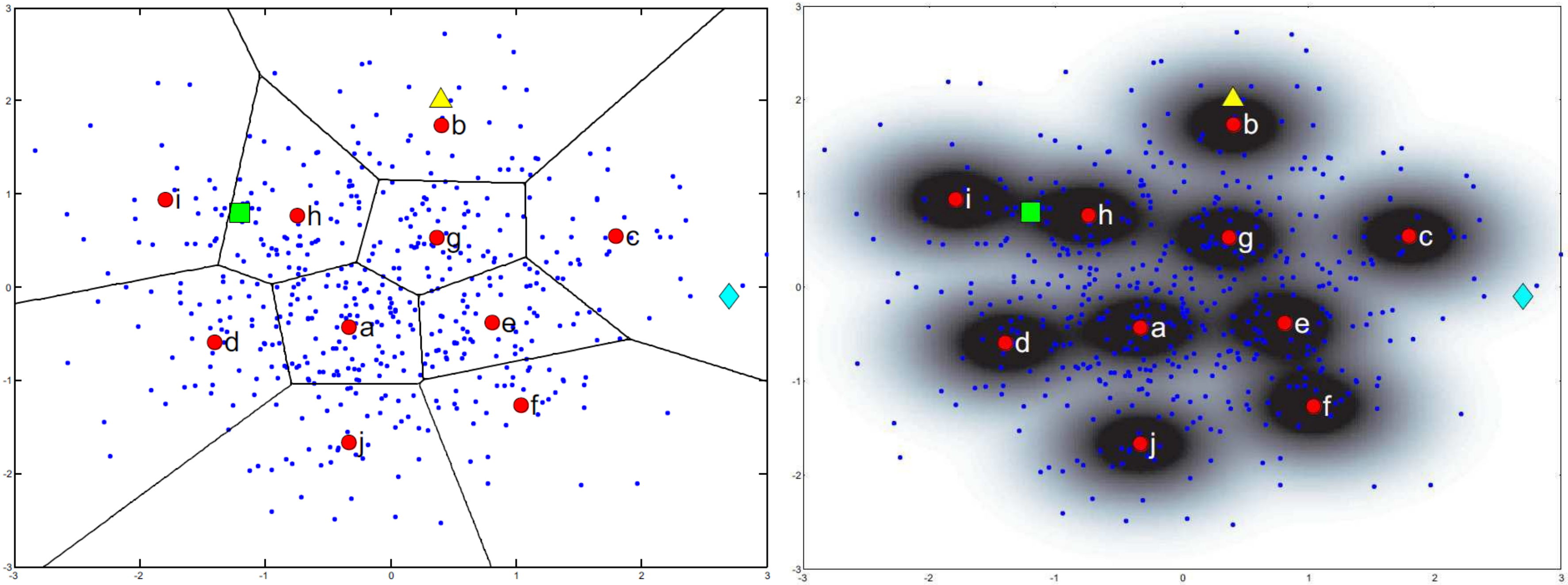
DL for Object Recognition



The dictionary learning pipeline for object recognition is shown above. Descriptors (i.e. features, such as SIFT) are clustered into a number of atoms using e.g. K-means. Each image patch is represented by a single histogram (coefficient vector) of cluster membership (i.e. atoms).

Soft Assignment for Dictionary Learning

UNIVERSITY OF SURREY



- Hard assignment: each descriptor contributes to only one histogram bin.
- Soft assignment: more than one descriptors can contribute to a histogram bin.

Soft Assignment for Dictionary Learning

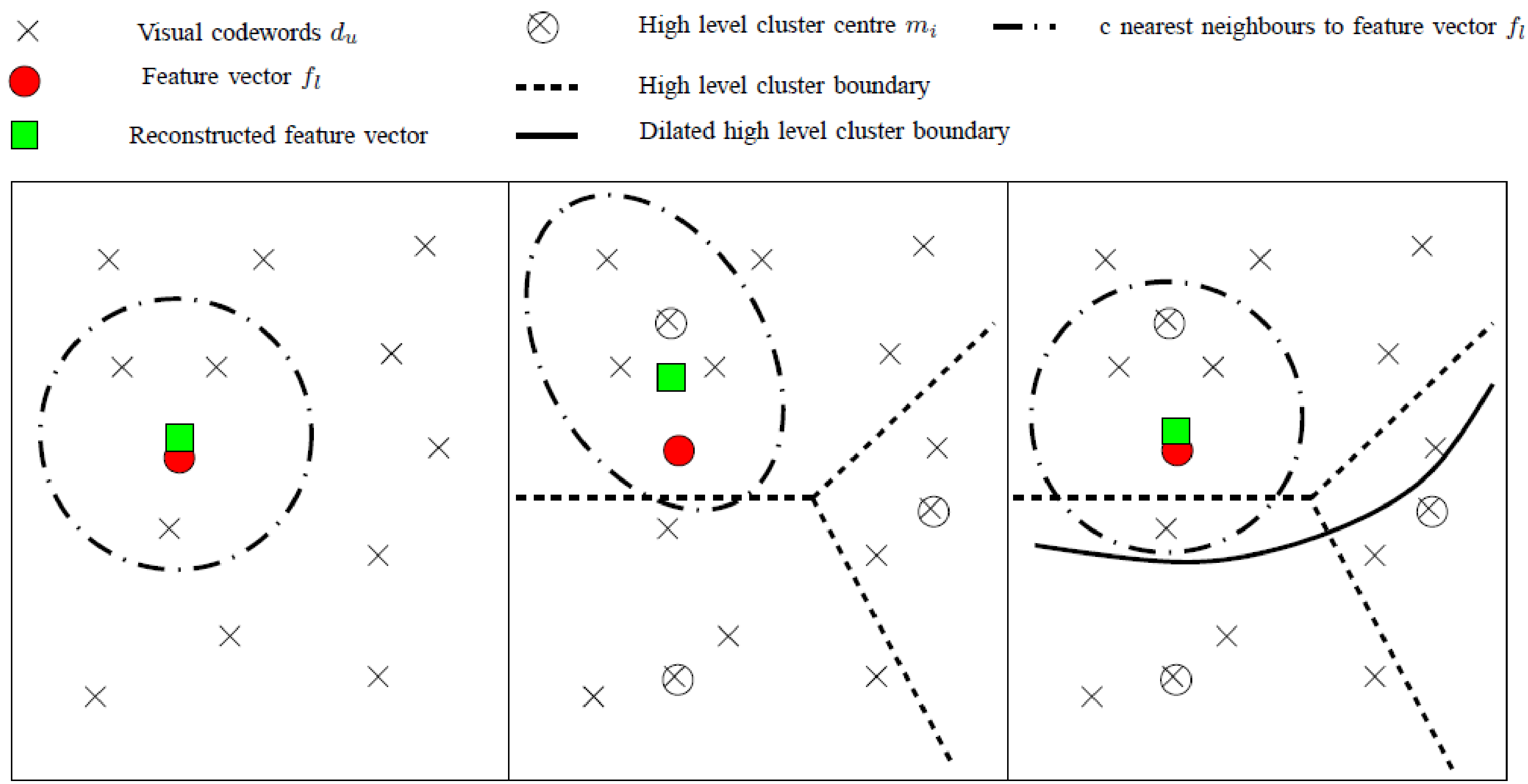
UNIVERSITY OF SURREY

$$C(w) = \frac{1}{I}\sum_{i=1}^{I}\frac{K_\sigma(D(w,r_i))}{\sum_{j=1}^{J}K_\sigma(D(w_j,r_i))}$$

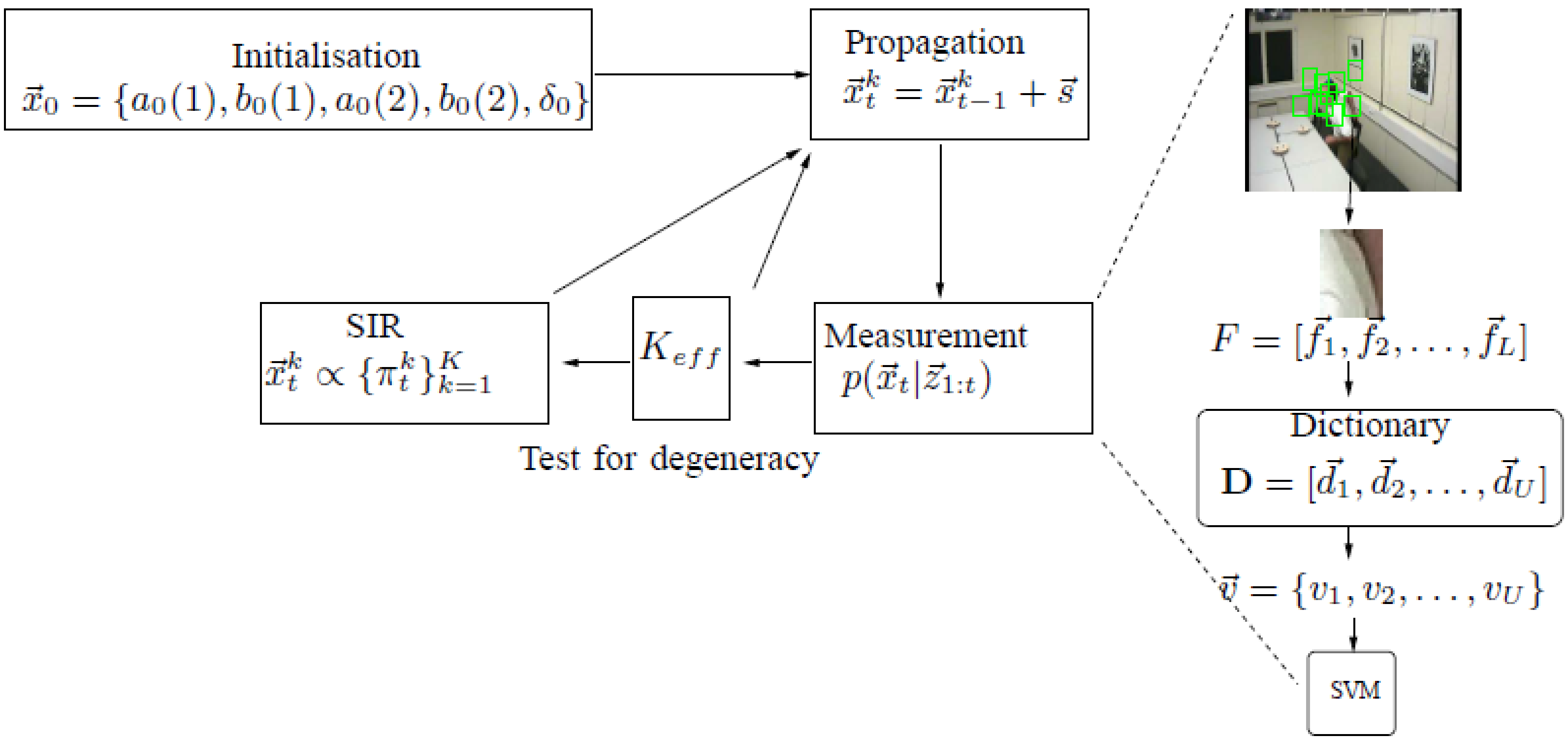
J :	is the number of atoms in the dictionary
I :	is the number of descriptors in the image
$D(w,r_i)$:	is the distance between atom w and the descriptors r_i .
K_σ :	is a Gaussian kernel with smoothing factor σ .
w :	is an atom in the dictionary.

This method has shown very good performance for object recognition in still images (Pascal VOC, ImageCLEF challenge) (van Gemert et al. 2010). The soft assignment technique can be further enhanced using a locality constraint approach.

Fast Hierarchical Nearest Neighbour Search



Particle Filter based Tracking Framework



M. Barnard, P.K. Koniusz, W. Wang, J. Kittler, S. M. Naqvi, and J.A. Chambers, "Robust Multi-Speaker Tracking via Dictionary Learning and Identity Modelling", *IEEE Transactions on Multimedia*, vol. 16, no. 3, pp. 864-880, 2014.

Demo

Future Work

- Exploit joint sparsity in both the array and source domains for source separation and beamforming
- Develop sparse polynomial dictionary learning and blind sparse deconvolution algorithms for reverberant source separation and beamforming
- Extend the sparse dictionary learning algorithm to multiplicative noise removal for sonar imaging
- Develop new sparse methods for large scale array beamforming and source separation
- Develop multivariate source models for source separation

Acknowledgement

- Internal collaborators: Miss Jing Dong, Dr Mark Barnard, Prof Mark Plumbley, Mrs Atiyeh Alinaghi, Dr Tao Xu (former student), Dr Qingju Liu, Dr Volkan Kilic (former student), Mr Jian Guan, Mr Alfredo Zermini, Mr Lucas Rencker, Dr Philip Jackson, Prof Josef Kittler, and Dr Saeid Sanei
- External collaborators: Prof Jonathon Chambers (Newcastle University), Prof John McWhirter (Cardiff University), Prof Ian Proudler (Loughborough University), Mr Zeliang Wang (Cardiff University), and Dr Wei Dai (Imperial College).
- Financial support: EPSRC & DSTL, UDRC in Signal Processing, and EC.