



# Independent Component Analysis in Multimedia Modeling

Jan Larsen



Intelligent Signal Processing Group  
Informatics and Mathematical Modelling  
Technical University of Denmark

www.imm.dtu.dk  
jl@imm.dtu.dk



## Motivation

- modeling of multimedia and multimodal data becomes increasingly important with the digitalization of the world
- signal detection/selection and pattern recognition are key components for new content based and context sensitive standards such as XML and MPEG7/MPEG21
- demonstrate the potential of ICA and BSS for modeling and understanding of multimedia data



## Special multimedia signal processing session at ICA2003

1. *Independent component analysis in multimedia modeling*  
Jan Larsen, Lars Kai Hansen, Thomas Kolenda, Finn Årup Nielsen
2. *Learning the semantics of multimedia content with application to Web image retrieval and classification*  
Alexei Vinokourov, David Hardoon and John Shawe-Taylor
3. *Review of ICA and HOS methods for retrieval of natural sounds and sound effects*  
Shlomo Dubnov and Adiel Ben-Shalom
4. *Audio/visual independent components*  
Paris Smaradgis and Michael Casey
5. *A tentative typology of audio source separation tasks*  
Emmanuel Vincent, Cédric Févotte, Laurent Benaroya, Rémi Gribonval



## Multimedia Modeling

from single medium (text/image/audio) to fusion of more media — heterogeneous multimodal fusion

### The utility of ICA/BSS

- ICA/BSS are learning-based models, hence relevant for “intelligent multimedia processing”, e.g., content based extraction and retrieval
- the independence assumption is amenable for interpretation and well-aligned with human perception of natural sound and images – cognitive component analysis
- ICA can be viewed as data projection onto latent subspace, i.e., feature extraction e.g., useful for data security and multimedia standards





## Outline

- ICA/BSS multimedia analysis
- multimedia applications of ICA/BSS
- learning algorithms
- example of content extraction from combined text and images
  - the ICA assumption for text and images
- example of chat room topic spotting
- conclusions and future challenges



## ICA/BSS multimedia analysis

Blind Source Separation based on

- spatial separation (direction of arrival)
- spectral separation (Wiener filter)
- time-frequency separation (masking)
- statistical independence

Linear mixing generative ICA model

$$\begin{array}{ccccccc} \mathbf{x} & = & \mathbf{A} & \cdot & \mathbf{s} & + & \boldsymbol{\varepsilon} \\ \begin{array}{c} P \text{ dim.} \\ \text{feature} \end{array} & & \begin{array}{c} P \times K \\ \text{mixing mtx.} \end{array} & & \begin{array}{c} K \text{ dim.} \\ \text{source} \end{array} & & \begin{array}{c} P \text{ dim.} \\ \text{noise} \end{array} \end{array}$$



## The ICA model in a multimedia perspective

- one medium – combined media
- model assumptions
- learning algorithms
  - maximum likelihood optimization
  - optimization of contrast functions, e.g., higher-order cumulants
  - kernel methods
  - Bayesian learning



## Single medium

- ICA is an extension of PCA and supplement to non-negative matrix factorization, independent and hierarchical factor analysis, ICA mixture model, projection pursuit, generative topographic mapping
- ICA is used for detection, extraction and explanation of hidden causes
- multimedia analysis results in interpretation of mixing matrix and sources

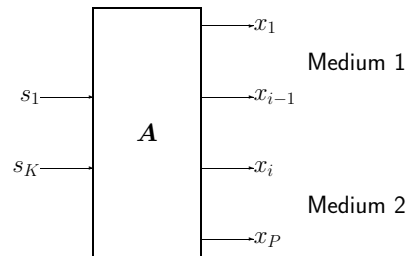
**Example:** If  $\mathbf{x}$  grey values of image pixels,  $\mathbf{a}_k$  is an eigenimage and  $s_k$  is the strength of the source.





## Combined media

- ICA establishes a common latent space of combined media
- can be viewed as a supervised method for establishing relations between media
- extension of e.g., partial least squares, canonical correlation analysis



## Model assumptions

- noise
  - no noise in image analysis
  - complex noise model in audio
- over- ( $\#sensors > \#sources$ ) and under-determined cases
  - overdetermined in image and text
  - underdetermined in audio
- convolutive
  - acoustical response in audio
  - filtering of image sequences
- constraints
  - incorporation of prior knowledge, e.g., non-negativity, spatial separation constraints
  - common and individual modality latent spaces in multimedia cases



## Outline

- ICA/BSS multimedia analysis
- multimedia applications of ICA/BSS
- learning algorithms
- example of content extraction from combined text and images
  - the ICA assumption for text and images
- example of chat room topic spotting
- conclusions and future challenges



## Multimedia applications of ICA/BSS

- most reported multimedia applications using ICA/BSS can also be tackled by other models
- main attraction is that ICA provides unsupervised grouping well-aligned with human perception



## Multimedia applications of ICA/BSS

Medium	Topics	Refs.
Image/Video	natural scenes, feature extraction, noise reduction	8
	watermark detection	2
	content based retrieval	5
Multimodal brain data	EEG, MEG, fMRI	2
Audio	general	3
	auditory perception	2
	source separation, scene analysis	5
Text	document filtering, retrieval	5
Combined media	document content and inter-connectivity	3
	cross-language document retrieval	1
	combined text/image content extraction	2
	audio-visual segmentation	5



## Outline

- ICA/BSS multimedia analysis
- multimedia applications of ICA/BSS
- learning algorithms
- example of content extraction from combined text and images
  - the ICA assumption for text and images
- example of chat room topic spotting
- conclusions and future challenges



## Learning Algorithms

### Routes to ICA

- non-Gaussianity
- time-correlation
- non-stationarity
- constraints (mixing matrix, source priors)

### Two algorithms

- InfoMax
- Dynamic decorrelation



## Likelihood and minimum mutual information

$$\begin{aligned}
 KL(p^*(\mathbf{x}); p(\mathbf{x}; \mathbf{W})) &= KL(p^*(\hat{\mathbf{s}}); p(\hat{\mathbf{s}}; \mathbf{W})) \\
 &= KL\left(\prod_{i=1}^m p^*(\hat{s}_i); p(\hat{\mathbf{s}}; \mathbf{W})\right) \\
 \hat{\mathbf{s}} &= \mathbf{A}^{-1}\mathbf{x} = \mathbf{W}\mathbf{x}
 \end{aligned}$$

minimal MI between sources is optimal in likelihood sense

- approximate MI by kurtosis, skewness (Cardoso)
- expectation of fixed nonlinearities (Bell, Oja, Amari)





## Linear ICA model

Noise free, quadratic, linear ICA

$$\mathbf{x} = \mathbf{A}\mathbf{s}$$

Ambiguity

$\mathbf{A}$  and  $\mathbf{s}$  can be identified up to scaling (sign) and permutations

$$\mathbf{x} = \mathbf{A}\mathbf{s} = [\mathbf{A}(\mathbf{P}\mathbf{S})^{-1}] [\mathbf{P}\mathbf{S}\mathbf{s}]$$

■  $\mathbf{P}\mathbf{P}^\top = \mathbf{P}\mathbf{P}^{-1} = \mathbf{I}$  is a permutation matrix

■  $\mathbf{S}$  is a diagonal scaling matrix



## Likelihood

Generative model

$$p(\mathbf{x}|\mathbf{A}) = \int \delta(\mathbf{x} - \mathbf{A}\mathbf{s}) p_s(\mathbf{s}) d\mathbf{s}$$

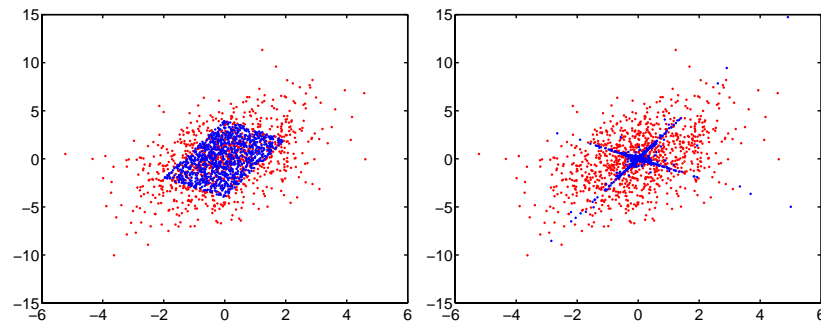
$p_s(\mathbf{s})$  is **prior** sources distribution which is assumed **non-Gaussian**

Data matrix  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]$

$$p(\mathbf{X}|\mathbf{A}) = \prod_{n=1}^N p(\mathbf{x}_n|\mathbf{A}) = \prod_{n=1}^N \int \delta(\mathbf{x}_n - \mathbf{A}\mathbf{s}_n) p_s(\mathbf{s}_n) d\mathbf{s}_n$$



## Why is non-Gaussianity required?



## Why is non-Gaussianity required?

Independent sources and scale ambiguity implies

$$E\{\mathbf{s}\mathbf{s}^\top\} = \mathbf{I}$$

A multi-variate Gaussian is fully specified by its covariance

$$\begin{aligned} E\{\mathbf{x}\mathbf{x}^\top\} &= E\{\mathbf{A}\mathbf{s}\mathbf{s}^\top\mathbf{A}^\top\} \\ &= \mathbf{A}\mathbf{A}^\top = (\mathbf{A}\mathbf{Q})(\mathbf{A}\mathbf{Q})^\top \end{aligned}$$

where  $\mathbf{Q}$  is a rotation matrix  $\mathbf{Q}\mathbf{Q}^\top = \mathbf{I}$

All mixing matrices  $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{Q}$  give same measurement covariance matrix





## Source estimation

Maximum a posteriori principle

$$\hat{\mathbf{s}} = \arg \max_{\mathbf{s}} p(\mathbf{s} | \mathbf{x}, \mathbf{A}) \propto$$

$$\begin{aligned} p(\mathbf{s} | \mathbf{x}, \mathbf{A}) &\propto p(\mathbf{x} | \mathbf{A}, \mathbf{s}) p_s(\mathbf{s}) \\ &= \delta(\mathbf{x} - \mathbf{A}\mathbf{s}) p_s(\mathbf{s}) \end{aligned}$$

$$\hat{\mathbf{s}} = \mathbf{A}^{-1} \mathbf{x} \text{ independent on } p_s(\mathbf{s})$$



## Choice of source distribution

Pham & Garat, 1997 studied asymptotic fluctuations due to mis-specified source distribution.

Under mild conditions asymptotic consistence is achieved

$$\boldsymbol{\delta} = \mathbf{I} - \hat{\mathbf{A}}^{-1} \mathbf{A}, \quad \text{vec}(\boldsymbol{\delta}) \sim \mathcal{N}(\mathbf{0}, \mathbf{C}/N)$$

Contamination

$$\mathbf{s} = \mathbf{A}^{-1} \mathbf{x}, \quad \hat{\mathbf{s}} = \hat{\mathbf{A}}^{-1} \mathbf{x}$$

$$\mathbf{A}^{-1} - \hat{\mathbf{A}}^{-1} = \boldsymbol{\delta} \mathbf{A}^{-1}$$

$$\mathbf{s} - \hat{\mathbf{s}} = \left( \mathbf{A}^{-1} - \hat{\mathbf{A}}^{-1} \right) \mathbf{x} = \boldsymbol{\delta} \mathbf{s}$$

$$\hat{\mathbf{s}} = \mathbf{s}(\mathbf{I} - \boldsymbol{\delta})$$



## Likelihood - continued

$$p(\mathbf{X} | \mathbf{A}) = \prod_{n=1}^N \int \delta(\mathbf{x}_n - \mathbf{A}\mathbf{s}_n) p_s(\mathbf{s}_n) d\mathbf{s}_n$$

integration by substitution,  $\mathbf{v}_n = \mathbf{A}\mathbf{s}_n$ ,  $d\mathbf{v}_n = |\det \mathbf{A}| d\mathbf{s}_n$

$$\begin{aligned} p(\mathbf{X} | \mathbf{A}) &= \prod_{n=1}^N \int \delta(\mathbf{x}_n - \mathbf{v}_n) p_s(\mathbf{A}^{-1} \mathbf{v}_n) \frac{1}{|\det \mathbf{A}|} d\mathbf{v}_n \\ &= \frac{1}{|\det \mathbf{A}|^N} \prod_{n=1}^N p_s(\mathbf{A}^{-1} \mathbf{x}_n) \end{aligned}$$

Maximum likelihood

$$\hat{\mathbf{A}} = \arg \max_{\mathbf{A}} p(\mathbf{X} | \mathbf{A}) = \arg \max_{\mathbf{A}} \log p(\mathbf{X} | \mathbf{A})$$



## Gradient ascent

$$\mathbf{A}^{(i+1)} = \mathbf{A}^{(i)} + \eta \cdot \frac{\partial p(\mathbf{X} | \mathbf{A}^{(i)})}{\partial \mathbf{A}}$$

where  $\eta$  is a step-size

Gradient of log-likelihood

$$\log p(\mathbf{X} | \mathbf{A}) = -N \log |\det \mathbf{A}| + \sum_{n=1}^N \log p_s(\mathbf{A}^{-1} \mathbf{x}_n)$$

$$\frac{\partial p(\mathbf{X} | \mathbf{A})}{\partial \mathbf{A}} = -N \mathbf{A}^{-\top} - \mathbf{A}^{-\top} \mathbf{A}^{-1} \sum_{n=1}^N \mathbf{x}_n \frac{\partial \log p_s(\mathbf{v})}{\partial \mathbf{v}^{\top}} \Big|_{\mathbf{v}=\mathbf{A}^{-1} \mathbf{x}_n}$$





## Natural gradient ascent

Gradient ascent is slow due to different length scales in gradient directions

- Natural gradient provides optimization in a non-orthonormal Riemannian space:  $\|d\theta\|^2 = d\theta^\top G(\theta) d\theta$ , where  $G(\theta)$  is the (local) metric tensor.
- Euclidian orthonormal space:  $G = I$

Steepest ascent direction

$\nabla L = \arg \max_{d\theta} L(\theta + d\theta)$  subject to  $\|d\theta\|^2 = \epsilon^2$  sufficiently small

$$\nabla L(\theta) = G^{-1}(\theta) \frac{\partial L(\theta)}{\partial \theta}$$

If metric is the Fisher information matrix (Hessian matrix)  $G(\theta) = \partial L(\theta) / \partial \theta \partial \theta^\top$  then natural gradient is the Newton direction



## Natural gradient for InfoMax

$A$  is in the matrix space of nonsingular  $K \times K$  matrices, Amari, 1996

$$\nabla L(A) = A A^\top \frac{\partial L(A)}{\partial A}$$

Natural gradient ascent

$$A^{(i+1)} = A^{(i)} - \eta \left[ N A + \sum_{n=1}^N x_n \frac{\partial \log p_s(v)}{\partial v^\top} \bigg|_{v=A^{-1}x_n} \right]$$

Choice of source distribution

$$p_s(s) = \prod_{i=1}^K p_{s_i}(s_i) = \prod_{i=1}^K \frac{1}{\pi \cosh(s_i)}$$

corresponds to

$$\frac{\partial \log p_s(v)}{\partial v} = -\tanh(v)$$



## Overdetermined ICA

$$x = U \cdot \Phi \cdot s = A \cdot s$$

To stages: PCA succeeded by quadratic ICA

- $U$  is the  $P \times K$  matrix of  $K$  largest eigenvectors of the covariance of  $E\{xx^\top\} = U D U^\top$
- $\Phi$  is the  $K \times K$  mixing matrix
- Quadratic ICA is thus performed in the subspace  $y = U^\top x$
- No noise case and  $K$  sources PCA is optimal. In the noisy case PCA gives an approximate solution.



## Probabilistic Model Selection

Model Hypotheses

Consider a finite set of models enumerated by  $m = 0, 1, \dots, M$ .  $m = 0$  signifies the null-hypothesis corresponding to no non-trivial independent components





## Bayes optimal model selection

Bayes optimal decision rule (under 0/1 loss function) leads to the optimal model

$$m_{opt} = \arg \max_m p(m|\mathbf{X})$$

$$p(m|\mathbf{X}) = \frac{p(\mathbf{X}|m)P(m)}{\sum_{m=0}^M p(\mathbf{X}|m)P(m)}$$

is the probability of the model given data



## Probabilistic Model Selection

### Uniform Model Belief

In the case of equal model priors, i.e.,  $P(m) = 1/(M+1)$ , the model selection concerns computing the **evidence**  $p(\mathbf{X}|m)$

### Evidence

$$p(\mathbf{X}|m) = \int P(\mathbf{X}, \boldsymbol{\theta}|m) d\boldsymbol{\theta} = \int p(\mathbf{X}|\boldsymbol{\theta}, m)p(\boldsymbol{\theta}|m) d\boldsymbol{\theta}$$

- $\boldsymbol{\theta}$  are model parameters
- $p(\mathbf{X}|\boldsymbol{\theta}, m)$  is the likelihood
- $p(\boldsymbol{\theta}|m)$  is the prior which is normally assumed vague and normalizable



## Evidence approximation

Simpler than more involved methods like Laplace, variational Bayes and MCMC

### Normalized log-posterior

$$C(\boldsymbol{\theta}) = \frac{1}{N} (\log p(\mathbf{X}|\boldsymbol{\theta}, m) + \log p(\boldsymbol{\theta}|m))$$

and the maximum a posteriori (MAP) solution

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} C(\boldsymbol{\theta})$$



## Approximations

### Gaussian MAP approximation

$$C(\boldsymbol{\theta}) = C(\hat{\boldsymbol{\theta}}) - \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\mathbf{J}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^\top$$

$$\mathbf{J} = -\frac{\partial^2 C(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = O(1)$$

### Laplace Approximation

$$\begin{aligned} p(\mathbf{X}|m) &= \int \exp(NC(\boldsymbol{\theta})) d\boldsymbol{\theta} \\ &\approx \int \exp\left(NC(\hat{\boldsymbol{\theta}}) - \frac{N}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\mathbf{J}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^\top\right) d\boldsymbol{\theta} \\ &\approx p(\mathbf{X}|\hat{\boldsymbol{\theta}}, m) \cdot p(\hat{\boldsymbol{\theta}}|m) \cdot \left(\frac{2\pi}{N}\right)^{\frac{\dim(\boldsymbol{\theta})}{2}} \cdot |\mathbf{J}|^{-\frac{1}{2}} \end{aligned}$$





## Bayesian Information Criterion

Since  $\mathbf{J} = O(1)$ , the leading term for large  $N$  does not involve the often complicated Hessian, hence, the evidence is approximated as

$$p(\mathbf{X}|m) \approx p(\mathbf{X}|\hat{\boldsymbol{\theta}}, m) \cdot p(\hat{\boldsymbol{\theta}}|m) \cdot \left(\frac{2\pi}{N}\right)^{\frac{\dim(\boldsymbol{\theta})}{2}}$$

$$\dim(\boldsymbol{\theta}) = 1 + KN + K^2 + \frac{K(2P - K + 1)}{2} + P$$



## On the likelihood for overdetermined case

- signal space spanned by the first  $K$  eigenvectors has full covariance structure
- noise space  $\mathbf{e}$  spanned by the remaining  $P - K$  eigenvectors assumed to be isotropic with noise variance

$$\sigma_e^2 = (P - K)^{-1} \sum_{i=K+1}^N D_{ii}^2$$

- Assuming independence of signal and noise space we model

$$p(\mathbf{x}|\boldsymbol{\theta}, K) = p(\mathbf{y}|\boldsymbol{\Phi}, K)p(\mathbf{e}|\sigma_e^2)$$

$\mathbf{y} = \mathbf{U}^\top \mathbf{x}$  is the signal space in which ICA is performed



## Likelihood terms

### Noise space likelihood

$$p(\mathbf{E}|\sigma_e^2) = (2\pi\sigma_e^2)^{-N(P-K)/2} \cdot \exp\left(\frac{-N(P-K)}{2}\right)$$

### Signal space likelihood

$$p(\mathbf{Y}|\mathbf{A}, m) = p_s\left((\mathbf{U}^\top \mathbf{A})^{-1} \mathbf{Y}\right) \frac{1}{|\det(\mathbf{U}^\top \mathbf{A})|^N}$$



## Outline

- ICA/BSS multimedia analysis
- multimedia applications of ICA/BSS
- learning algorithms
- example of content extraction from combined text and images
  - the ICA assumption for text and images
- example of chat room topic spotting
- conclusions and future challenges





## Example of content extraction from combined text and images

- content based retrieval of combined text and image data from webpages
- mutual support between media are ensured by adjacency
- we want to demonstrate a synergistic effect
- components can be interpreted by providing key features (words,color,texture)



## Modeling Framework

$$\mathbf{z} = [\mathbf{z}_I; \mathbf{z}_T]$$

is the column vector of image ( $I$ ) and text ( $T$ ) features.

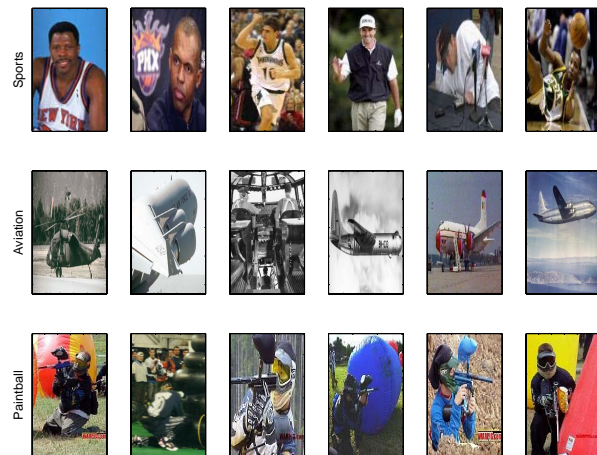
**Unsupervised** ICA provides a probability density  $p(\mathbf{z})$  model from which meaningful clusters are identified

**Supervised** modeling is the conditional class-feature probability,  $p(y|\mathbf{z})$ , where  $y = \{1, 2, \dots, C\}$  is the class label.

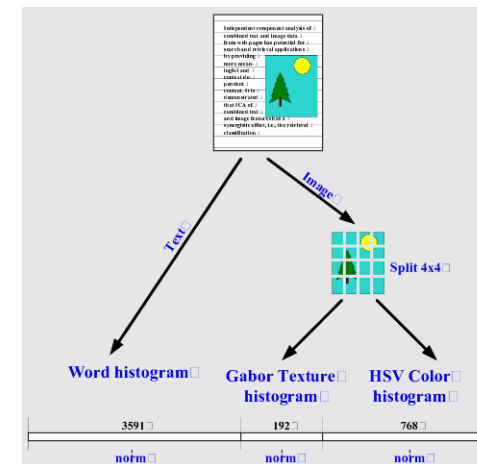
A simple classifier can be obtained from the unsupervised ICA



## Image examples from three classes



## Extracting features





## Text features

- collection of terms adjacent to an image is represented by a vector: the histogram of term occurrence
- filtered by removing low frequency words
- filtered by removing high frequency words - stop words. manually constructed to form a list of 585 words
- stemming by merging words with different endings, e.g., **ing** or **s**



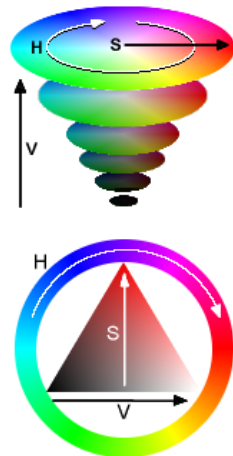
## Image features

- intention is to employ Vector Space Model on image features
- suggest low level image features of the ISO/IEC MPEG-7 standard
- image is divided into  $4 \times 4$  patches to increase sensitivity



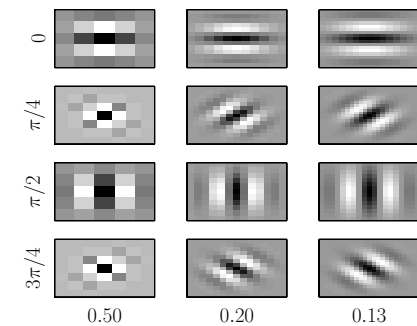
## Color

- Hue-Saturation-Value color space (better than RGB)
- Hue is dominant wavelength,
- color components quantized into 16 levels
- $48 = 16 \cdot 3$  features for each of the 16 patches, i.e.,  $48 \cdot 16 = 768$  features



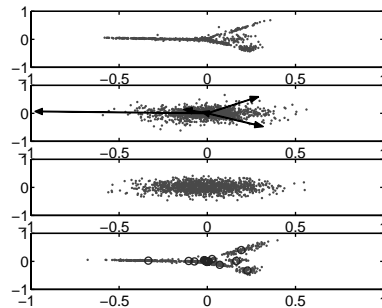
## Texture

- Gabor filter bank
- filter output captures a specific texture frequency and direction
- energy of filtered outputs are the texture feature
- total of 16 patches times 12 filters, 192 texture features





## The ICA Assumption



**Panel 1:** scatterplot of text data along two PC's show ray structure

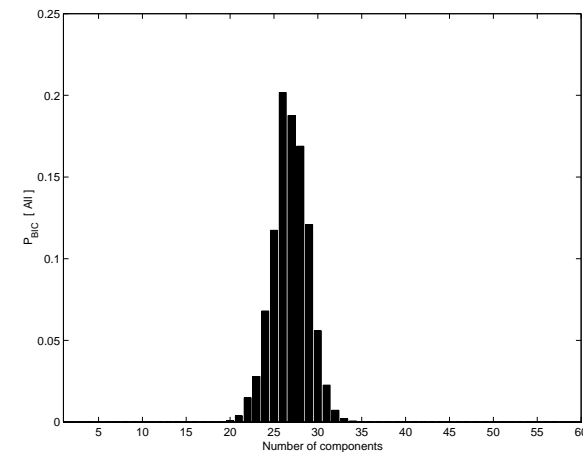
**Panel 2:** 5 component InfoMax and **prior** source scatterplot  $\propto 1/\cosh(s_k)$ . **posterior** source scatterplot resembles first panel

**Panel 3:** Gaussian with same mean and variance as data in first panel. Major axes parallel with PC axes - does not reveal the structure

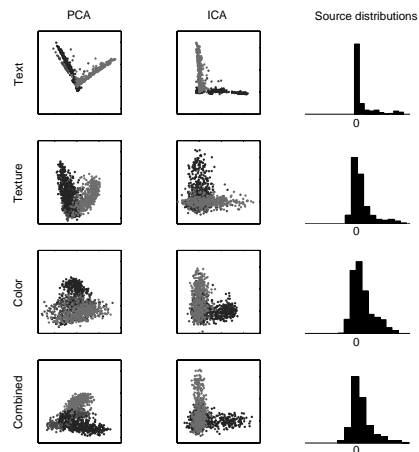
**Panel 4:** 15 component Gaussian mixture model. Good density estimator but does not reveal independent components



## Selection of components using BIC



## ICA modeling



## ICA classification

Interpretation as conditional component probabilities

$$\hat{p}(k|\mathbf{x}) = \frac{\exp(\hat{s}_k)}{\sum_{k=1}^K \exp(\hat{s}_k)}, \quad \hat{\mathbf{s}} = [\hat{s}_1, \dots, \hat{s}_K]^\top = \hat{\Phi}^{-1} \hat{\mathbf{U}}^\top \mathbf{x}$$

Posterior class probabilities

$$p(y|\mathbf{x}) = \sum_{k=1}^K p(y|k)p(k|\mathbf{x})$$

$p(y|k)$  are estimated from data as the frequency of occurrence for specific component-class combination  $k \in [1; K]$ ,  $y \in [1; C]$

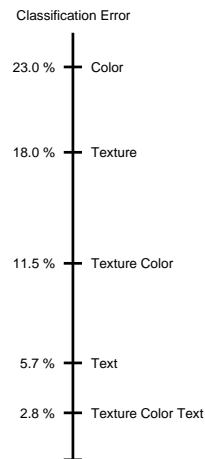
$$\hat{p}(y, k) = \frac{1}{N} \sum_{n=1}^N \delta(y - y(n)) \cdot \delta(k - \arg \max_{\ell} \hat{p}(\ell|\mathbf{x}(n)))$$

$$\hat{p}(y|k) = \frac{\hat{p}(y, k)}{\sum_y \hat{p}(y, k)}$$





## Performance



## Performance



## Component explanation

### ICA model

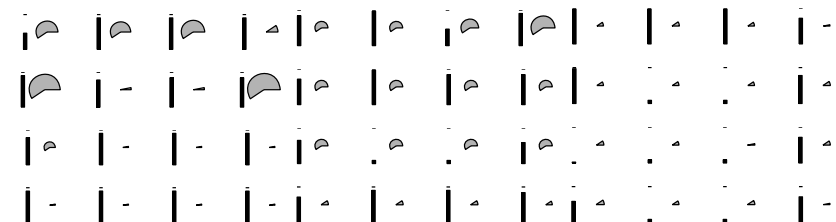
$$x = As = U\Phi s,$$

- $U$  is the  $P \times K$  matrix of  $K$  largest eigenvectors of the covariance of  $x$
- $\Phi$  is the  $K \times K$  mixing matrix. Quadratic ICA is thus performed in the subspace  $\tilde{x} = U^\top x$

$K$ 'th column of  $\hat{U}\hat{\Phi}$  are features of  $K$ 'th source. High values of the column provide a compact explanation



## Interpreting color features



Sports

Aviation

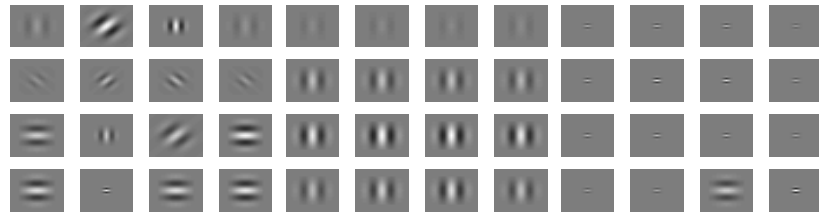
Paintball

Bar is Value, Pie diameter is Saturation, Pie angle is Hue





## Interpreting texture features



Sports

Aviation

Paintball



## Image Annotation

 $I_1$  $I_2$  $I_3$ 

Image	Label	Keywords
$I_1$	Sports	position college weight born lbs height guard
$I_2$	Aviation	na air convair wing
$I_3$	Paintball	check darkside force gog strike odt



## Outline

- ICA/BSS multimedia analysis
- multimedia applications of ICA/BSS
- learning algorithms
- example of content extraction from combined text and images
  - the ICA assumption for text and images
- example of chat room topic spotting
- conclusions and future challenges



## Chat analysis

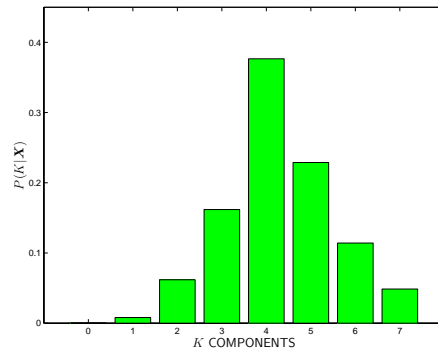
Kolenda, Hansen & Larsen, 2001

- retrospective analysis of CNN.com #CNN moderated chat channel
- topic spotting in unstructured chat text stream
- 8.5 hours on April 5th, 2000 was logged
- 4900 lines with 128 unique chatters participating
- preprocessing leaves  $T = 2498$  terms
- the stream was segmented into 300 characters long pseudo-documents windows
- windows were shifted 150 characters without breaking words
- $N = 1114$  pseudo-documents was generated

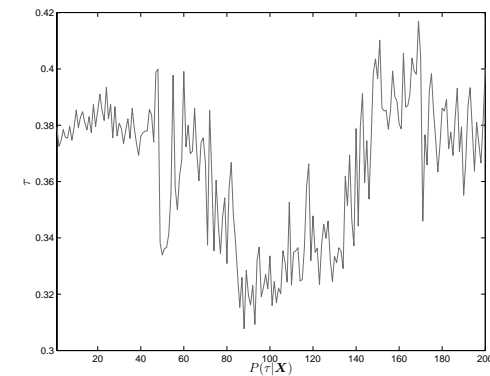




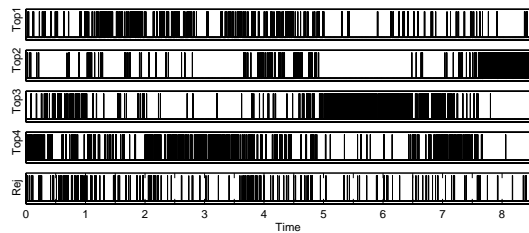
## Detecting the number of topics



## Selecting algorithm lag parameter



## Retrospective event detection



	keywords
topic 1	chat join pm cnn board message allpolitics visit check america
topic 2	gun show
topic 3	susan smith mother children kid life
topic 4	people census elian state clinton government thing year good father time



## Chat analysis demo

<http://mole.imm.dtu.dk/webchat> <http://mole.imm.dtu.dk/webdemo/>

### CNN Chat Topics

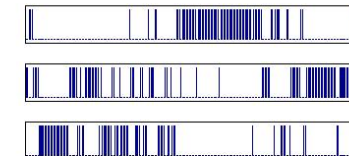
Channel: #News\_Cafe

Date: 29-Nov-2001  
Time frame: 14:25 - 16:20

Topic keywords:

- 1: enron  
1.00
- 2: tribunal military  
1.00 0.74
- 3: power ashcroft good people  
1.00 0.69 0.67 0.65

Topic activity times:



- By Thomas Kolenda copyrights 2001



## Outline

- ICA/BSS multimedia analysis
- multimedia applications of ICA/BSS
- learning algorithms
- example of content extraction from combined text and images
  - the ICA assumption for text and images
- example of chat room topic spotting
- **conclusions and future challenges**



## Conclusions

- independent component analysis and blind sources separation methods have potential for modeling, understanding and intelligent processing of multimedia data
- the unique feature of ICA is an unsupervised grouping of data which are amenable for interpretation and well-aligned with human perception
- combined text/image data for the purpose of cross-media retrieval and web search was presented
- synergy among text and image features leads to significant better classification performance
- retrospective dynamic topic spotting in chat rooms



## Probing the future

- relevant and specific multimedia features for which linear ICA is the appropriate model. Nonlinear ICA?
- representation issues in image/video, e.g., facial animation, motion parameters and active appearance models
- incorporation of natural language and semantic features in text processing
  - R.E. Madsen, J. Larsen & L.K. Hansen: "Part-of-Speech Enhanced Context Recognition," MLSP2004
- Cognitive component analysis
  - Hansen, L. K., Ahrendt, P., Larsen, J.: "Towards Cognitive Component Analysis, AKRR'05.



## Probing the future

- processing from mono/binaural audio signals: underdetermined convolutive mixture models, e.g., by invoking more specific audio priors
  1. — Olsson, R. K., Hansen, L. K.: "A harmonic excitation state-space approach to blind separation of speech, Advances in Neural Information Processing Systems," NIPS 2005.
    - R.K. Olsson, L.K. Hansen: "A harmonic excitation state-space approach to blind separation of speech," NIPS 2004.
    - R.K. Olsson, L.K. Hansen: "Probabilistic blind deconvolution of non-stationary sources," 12th ESPC, 2004.
    - R.K. Olsson, L.K. Hansen: "Estimating the number of sources in a noisy convolutive mixture using BIC," ICA 2004.
    - **non-stationary source AR(p)+sinusoidal**
    - **state space model and EM estimation**
  2. — Pedersen, M. S., Wang, D., Larsen, J., Kjems, U.: "Overcomplete Blind Source Separation by Combining ICA and Binary Time-Frequency Masking," MLSP'2005, 2005.
    - M.S. Pedersen, C.M. Nielsen: "Gradient Flow Convolutional Blind Source Separation," MLSP 2004.
    - **gradient flow for four close microphones**
    - **extension to convolutive**







3. — M.S. Pedersen, L.H. Hansen, U. Kjems, K.B. Ramussen: "Semi-blind Source Separation Using Head-Related Transfer Functions, ICASSP 2004."
  - **acoustical propagation constraint - HRTF**
  - **two microphones**
4. — Dyrholm, M., Makeig, S., Hansen, L. K.: "Model structure selection in convolutive mixtures, Neural Information Processing Systems," 2005.
  - M. Dyrholm, L.K. Hansen: "CICAAR: Convolutive Independent Component Analysis with an Auto-Regressive Inverse Model," IEEE Signal Processing Letters, 2003.
  - **can handle more sensors than sources**
  - **IIR estimation of the sources**
  - **example: speaker in office with loud background music, two microphone, distance 60cm.**  
**mic1.wav, mic2.wav, sep-music.wav, sep-speech.wav**
5. — T. Beierholm, B.D. Pedersen, O. Winther: "Low Complexity Bayesian Single Channel Source Separation," ICASSP'2004
  - **single channel separation using source priors**



## Probing the future

- **computational issue in training and recall in large multimedia databases**
- **estimation and optimization beyond natural gradient: advanced active data subset selection methods, online learning algorithms, and adaptation to changing environment**
- **intelligent fusion of media types, the ability to handle missing data and learning from combined labeled/unlabeled data**



## Acknowledgements

- **Thomas Kolenda**
- **Lars Kai Hansen**
- **Finn Årup Nielsen**
- **Ole Winther**

