INDEPENDENT COMPONENT ANALYSIS IN MULTIMEDIA MODELING

Jan Larsen, Lars Kai Hansen, Thomas Kolenda and Finn Årup Nielsen

Informatics and Mathematical Modelling, Technical University of Denmark
Richard Petersens Plads, Building 321, DK-2800 Kongens Lyngby, Denmark
email: {jl,lkh,thko,fn}@imm.dtu.dk, web: isp.imm.dtu.dk

ABSTRACT

Modeling of multimedia and multimodal data becomes increasingly important with the digitalization of the world. The objective of this paper is to demonstrate the potential of independent component analysis and blind sources separation methods for modeling and understanding of multimedia data, which largely refers to text, images/video, audio and combinations of such data. We review a number of applications within single and combined media with the hope that this might provide inspiration for further research in this area. Finally, we provide a detailed presentation of our own recent work on modeling combined text/image data for the purpose of cross-media retrieval.

1. INTRODUCTION

Processing of multimedia data has received increased attention during the last decade. While many research contributions in multimedia processing deal with issues related to a single medium, the even more challenging research topic is the fusion of more media that may be viewed as the fusion of highly heterogeneous multimodal data. The objectives of multimedia research are multiple, which can be exemplified by the areas covered in IEEE Transactions on Multimedia. The topics covered span content extraction and retrieval, human-machine interfaces, human perception, database technologies, data encryption and security, system integration and standards. The fact that independent component analysis (ICA) and blind source separation (BSS) methods can be viewed as statistical models, which are fitted to data using machine learning and adaptive signal processing tools, make them particularly useful for "intelligent multimedia processing" [1, 2] such as extraction of semantic representations, content based extraction, recognition and filtering applications. Moreover, recent research have shown that the independence assumption, separated sources and mixing matrices are amenable for interpretation and well-aligned with human perception in different media [3, 4, 5, 6, 7, 8, 9, 10].

Since ICA also can be viewed as performing projection of data onto latent subspaces, there is a potential in other multimedia areas e.g., in data security and in multimedia standards with the advent of the advanced MPEG standards [11], which enables content and context sensitive tools.

In section 2 the ICA/BSS model is presented in a multimedia perspective. Section 3 reviews the application of ICA/BBS for single and combined media within image/video, audio and text processing. An extended review of using ICA for content extraction from combined text and image data is presented in section 4. Finally, section 5 states conclusions and probe future challenges.

2. ICA/BSS MUTIMEDIA ANALYSIS

Blind source separation can be achieved with a number of assumptions on the mixed signals. Spatial separation, for example, is based on differences in direction of arrival, spectral separation, as in the Wiener filter, assumes that the sources have little spectral overlap, and finally independency based separation as in ICA.

Most currently used ICA algorithms assume linear mixing, i.e., the measurement is a linear combination of the source signals, as shown by the generative linear ICA model:

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\boldsymbol{x}$ is a P-dimensional column feature vector, $\boldsymbol{A} = [\boldsymbol{a}_1, \cdots, \boldsymbol{a}_K]$ is the $P \times K$ mixing matrix, $\boldsymbol{s} = [s_1, \cdots, s_K]^\top$, are independent sources, and $\boldsymbol{\varepsilon}$ is additive noise. The assumptions of the linear ICA model are further discussed in Section 4.1.

If the feature vector consists of data from a single medium the ICA model could be seen as an unsupervised model, which generalizes the classical principal component analysis model (PCA) [12] by assuming independent and non-Gaussian source distributions. The unsupervised ICA model is used for detection, extraction and explanation of a number of independent hidden causes. The result of the multimedia analysis is typically to provide an interpretation of the mixing matrix and the sources. If for instance $\boldsymbol{x}$ represents a grey-valued image reshaped into a vector, the columns of the mixing matrix, $\boldsymbol{a}_k$, will represent "eigenimages" associated with a particular source, and s_k will determine the strength [13]. If the number of components $K < P$ then each image $\boldsymbol{x}$ will be sparsely represented by a few "eigenimages". In other image applications $\boldsymbol{x}$ will represent features extracted from image or image sequences.

The literature provides a number of models related to ICA which has potential for multimedia. This includes: non-negative matrix factorization [14], which assume positive mixing matrix and sources; independent and hierarchical factor analysis [15, 16] are frameworks which generalize classical factor analysis, PCA and ICA; the ICA mixture model [17], which uses a probabilistic mixture of ICA models; topographic ordering of partly dependent sources [18]; projection pursuit [19]; and hierarchical generative topographic mapping [20].

When $\boldsymbol{x}$ contains data from more than one medium, the ICA establish a common latent source space for the media [10, 21] and can be viewed as a method for supervised learning of relations between the involved media. This is a generalization of classical methods such as partial least squares, canonical correlation and canonical variate analysis [12, 22]. Recent extensions of the simple ICA model, which operates from multimodal data includes the kernel canonical analysis [23] and the work by Lukic [24] that

identifies common latent spaces, as well as latent spaces individual to the modality.

In a multimedia context different assumptions of the ICA model Eq. (1) are to be considered. This includes the noiseless case, the overdetermined case ($P > K$) in e.g., image and text analysis, the underdetermined case ($P < K$) and convolutive mixing, e.g., in processing of audio signal [25].

This paper will not provide a detailed discussion of the rich literature on ICA learning algorithms, which includes: maximum likelihood optimization [26, 27, 28, 29, 30]; optimization of contrast functions from higher-order cumulants [31]; kernel methods [32]; and Bayesian learning [33, 34]. For a general discussion of different ICA models and estimation methods the following principal references [25, 35, 36, 37] are recommended.

3. MULTIMEDIA APPLICATIONS OF ICA/BSS

Most applications mentioned in this section can be treated by alternative methods, only ICA/BSS methods will be discussed. The main attraction of ICA is that it provides unsupervised grouping of data that has shown to be well-aligned with manual grouping [3, 4, 5, 6, 7, 8, 9, 10] in different media.

Medium	Topic	Reference
Image/ Video	natural scenes, feature extraction, noise reduction	[3, 6, 38, 39, 40, 41, 42, 43]
	watermark detection	[44, 45]
	content based retrieval	[10, 46, 47, 48, 49]
Multimodal brain data	EEG, MEG, fMRI	[5, 50]
Audio	general	[51, 52, 53]
	auditory perception	[4, 7]
	source separation, scene analysis	[54, 55, 56, 57, 58]
Text	document filtering, retrieval	[8, 9, 13, 59, 60]
Combined media	document content and inter-connectivity	[61, 62, 63]
	cross-language document retrieval	[23]
	combined text/image content extraction	[10, 49]
	audio-visual segmentation	[21, 64, 65, 66, 67]

Table 1. Overview of ICA multimedia applications.

3.1. Images/Video

The literature is rich on contributions which use ICA for analysis of image and video, however, we will merely present a potpourri of a number of applications.

3.1.1. Natural Scenes, Feature Extraction and Noise Reduction

[3, 6, 39] consider fundamental properties of natural scenes and demonstrates that the application of ICA provide Gabor-like [68] localized and oriented spatial filters, which resembles receptive fields in visual cortex. Thus edges are the independent components of natural scenes.

[38] extends this ideas for extracting features from stereo and color images and again Gabor-like spatial filters are found.

[40] uses maximum autocorrelation factors, which is identical to the Molgedey-Schuster ICA algorithm [69], for decomposition of remote sensing images and biological shape analysis.

[41] uses ICA to factorize histograms of joint feature vectors for the purpose of object detection and localization in cluttered scenes of non-rigid objects. It is demonstrated that application of ICA provides improved detection performance.

[43] presents two ICA approaches for detecting facial components in the images contained in a video sequence. The aim is to map the detected facial components, such as eyes and mouth, to a 3D-wireframe model to be used for facial animation.

The sparse code shrinkage algorithm [42] combines ideas of ICA with wavelet shrinkage to obtain a completely data driven technique for noise reduction in images, which have shown to outperform standard approaches such as Wiener filters.

3.1.2. Watermark Detection

Watermark detection and extraction is important for authentication of multimedia material especially when distributed over the Internet. [44, 45] deploys ICA for watermark detection and extraction which have shown robust to several important image processing attacks.

3.1.3. Content Based Image Retrieval

Content based image retrieval is a highly challenging aspect of multimedia analysis [70]. The task is hard because of the limited understanding of the relations between basic image features and abstract content descriptions. It is simply complicated to describe content in terms of intensity, edges, and texture. Therefore most current image retrieval systems, say on search engines like Google and FAST Multimedia Search, are based on analysis of an image and adjacent text on web page of the image.

Among the first commercial content based image retrieval systems worth mentioning are IBM's QBIC system [71], the VIR Image Engine from Virage, Inc. [72], and Visual RetrievalWare product by Excalibur Technologies [73]. These systems as well as the research prototypes mentioned in the reviews [74, 70] aim at using primitive image features for retrieval. However, the most widely used image searches are primarily based on image associated keywords and adjacent text. If we want to perform more advanced searches it is necessary to invoke context sensitive text based approaches, i.e., invoke statistical tools like the vector space approach known as latent semantic indexing (LSI) [75, 46, 47, 48] and ICA extensions [10, 49].

3.2. Multimodal Brain Data

ICA is an effective technique for removing artifacts and separating sources of multimodal brain signals such as electroencephalographic (EEG) and magnetoencephalographic (MEG) with application to brain research and for medical diagnosis and treatment [5]. A similar approach is proving useful for analyzing functional magnetic resonance brain imaging (fMRI) data [50].

3.3. Audio

The instantaneous mixing model is often insufficient for audio processing since audio/sound signals are convolved with response of the acoustic environment in which they propagate. The relevant model is thus often a convolutive mixture model, $\boldsymbol{x}(n) =$

$\sum_i \boldsymbol{A}(i)\boldsymbol{s}(n-i)$, where $\boldsymbol{A}(i)$ are matrix filter coefficients and n is the time index. [51] provides an overview of research topics in blind separation of convolutive mixed signals, concentrating on audio signals and related methods. In general, it is advantageous to use additional information e.g., specific speech signal priors [76], or by combining statistical independence with geometric source location (beam forming) [52]. Another challenge arise from the fact that the number of microphones is less than the number of audio sources calls for algorithms to cope with the underdetermined case, see e.g., [29, 30]. It is also important to consider a framework evaluation of such algorithms, which is considered in [53] for blind audio source separation tasks: extraction of sound sources for listening purposes; and identification of mixing matrix and sound sources for the purpose of classification and description.

3.3.1. Auditory Perception

[4, 7] analyze natural sound signals, e.g., highly non-Gaussian signals from radio stations broadcasting speech and classical music. The resulting mixing vectors are quite wavelet-like with a located regular time-frequency structure similar to that of the human auditory system.

3.3.2. Source Separation and Scene Analysis

[54] uses multiple-cause neural networks, which are related to ICA, for musical instrument separation. The signals are considered as a composition of hidden causes using Saund's multiple cause model [77].

[55] pursues automatic music transcription with the purpose of identifying instruments and notes and written transcription of these. The mixing matrix contains the spectral shape of each note and sources are different notes. The notion of sparse coding (most sources will be zero) and the relation to Saund's multiple cause model [77] is discussed.

[56] compares Computational Auditory Scene Analysis (CASA) and BSS models for speech separation. Whereas CASA is based on simple human auditory features, the BSS is data driven and based on statistical independence only. Subband processing is carried out for both CASA and BSS and it is suggested to combine these approaches.

[57] considers similarity in sound effects which is required for musical authoring and search by content in MPEG-7 applications [11]. Several approaches including, higher order spectra and ICA-based features with temporal hidden Markov modeling are reviewed and evaluated in the context of multimedia sound retrieval.

[58] proposes to use independent subspace analysis (ISA) for separating individual audio sources from a single-channel mixture. ISA is an extension of classical ICA, which can handle single-channel case by projecting onto a high-dimensional manifold. Non-stationary source signals are handled by using dynamic components. Further, the paper introduces the ixegram, which measures mutual similarity of components in an audio segment. Clustering the ixegram provides the source subspaces and time trajectories, which is exemplified by separating speech and music sources.

3.4. Text

Texts are usually handled in the so-called "vector space model" (VSM) and Latent Semantic Analysis (LSA), see Section 4.3.1. ICA applications include [8, 13, 59, 60] and aim at discovering independent topics in document collections. Source separation on dynamically evolving textual data appears in [9] with a Molgedey-Schuster analysis [69] performed on the CNN.com Internet chat-room text. Similar data is considered in [78] using a hidden Markov model, and complexity pursuit is considered in [79, 80].

3.5. Combined media

3.5.1. Content and inter-connectivity of documents

Jointly modeling the content and inter-connectivity between document enables e.g., meta-analysis of large document collection or better search tools. In [81] documents are clustered based on co-authorship. [82] performs co-citation analysis and clusters cited authors, which subsequently are mapped onto on a low-dimensional space, e.g., by multi-dimensional scaling [22]. [61, 62] combine co-citation analysis with words from the title of the documents.

[63] models collection of webpages from the term-document matrix (see Section 4.3.1) and their inter-connectivity assembled in an inlink-document matrix. Joint Probabilistic Latent Semantic Analysis (PLSA) is used to identify a common latent space for predicting terms and links. PLSA is related to performing ICA or PCA (LSI) on the combined set of term-document and inlink-document matrices. A possible application is intelligent web crawling, which starts from words, predicts the latent variable, then predicts link, and so on.

3.5.2. Cross-Language Document Retrieval

[23] uses kernel canonical correlation analysis, which is related to kernel ICA [32], to find a common latent/semantic space from multi-language documents (French/English). Comparison with LSI shows improved retrieval performance.

3.5.3. Audio-Visual Segmentation

The problem of audio-visual segmentation is addressed in [64]. Effective audio and visual features for characterizing scene content is presented and related to the MPEG-7 standard [11]. Further algorithms for segmentation/classification, and some testbed systems for video archiving and retrieval, are reviewed.

[65, 66] use an ICA related mutual information analysis of joint audio-visual features to identify a common subspace.

[67] extracts an speech signal from other acoustical signals by exploiting its coherence with the speaker's lip movements. The audio-visual coherence is learned by using a joint statistical model of visual and spectral audio input and separation can be achieved by maximum likelihood.

[21] operates from a fused data set of audio/visual data from video streams and discover combined subspace components amenable for interpretation.

3.5.4. Combined Text/Image Content Extraction

The extraction of meaningful content from images and associated text have been addressed in [46, 47, 48] using the Vector Space Model, and [83] describes a more general probabilistic approach. [10, 49] uses ICA and kernel canonical correlation analysis, respectively. This is further exemplified in the next section.

4. EXAMPLE: CONTENT EXTRACTION FROM COMBINED TEXT AND IMAGES

It was previously argued that independent component analysis (ICA) is a valuable tool for unsupervised structuring of multimedia signals [13]. In particular, we have shown that the independent components of text databases have intuitive meanings beyond that found by LSI [8, 9, 13], and similarly that independent components of image sets correspond to intuitively meaningful groupings [8, 10, 13].

This section provides an extended review of [10], which explores independent component analysis of combined text and image data.

4.1. The ICA assumption

The combined image and text data we will analyzed using a linear ICA approach based on sources with sparse priors. Here will

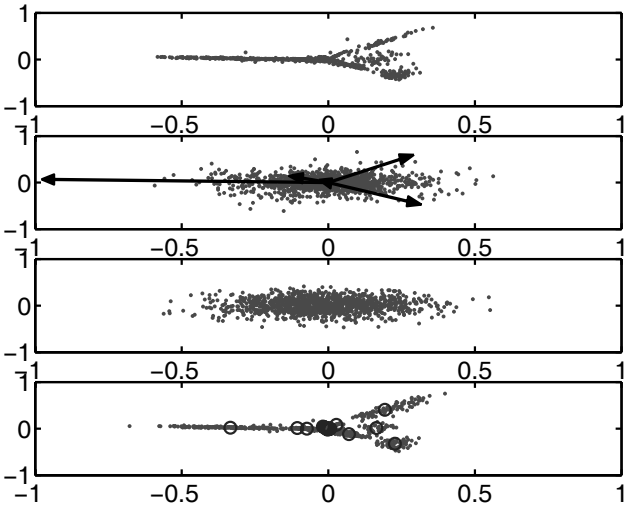


Fig. 1. The ICA assumption. The upper panel shows a scatterplot along two principal components of the text data discussed below. The second panel shows the results of a five component ICA estimated by Infomax [26, 27, 28]. The third panel shows samples from the normal distribution with the same mean and covariance matrix as in the first panel. Finally, the bottom panel shows a carefully optimized 15 component Gaussian mixture model. The centers are shown as open black circles overlayed on a sample from the mixture distribution.

briefly discuss the relevance of the ICA assumption. The upper panel of Fig. 1 shows scatter plots along two principal components of the real world text data discussed below. The "ray" like structure suggests that the data is a linear mixture of a few sparse independent sources. The second panel shows the results of a five component ICA on text data of the upper panel. The analysis is carried out with the Infomax algorithm [26, 27, 28], and produces five vectors (columns of the mixing matrix) indicated by arrows. The point cloud is a sample of the same size as in the upper panel, but drawn from the source density implicit in the Infomax algorithm, viz. $P(s_k) \propto 1/\cosh(s_k)$. Note that this density apparently is less sparse than the "posterior" density of the text data. In fact, if the estimated posterior source signals were projected on the mixing coefficient vectors, the original sample, shown in the first panel, would be recovered. In spite of the obviously inaccurate density estimate the mixing coefficients seem to be accurately determined. The third panel from above illustrate a sample from a Gaussian distribution with the same mean and covariance matrix as in the text data in the upper panel. Clearly the mean and covariance are not good statistics for this problem. The eigenvectors of the covariance matrix are axis parallel in the principal component plots and do not reveal any of the relevant structures in the upper panel. Finally, for comparison, the bottom panel shows a carefully optimized 15 component Gaussian mixture model. The BIC criterion was used to estimate the optimal number of components. The centers are shown as open black circles overlayed on a sample from mixture distribution. Clearly, the Gaussian mixture provides an accurate density model, however, it is also evident that it has no clue about the independent components that are clearly visible in the text sample. So from a exploratory point of view it is less interesting than the independent component model.

4.2. Modeling Framework

In order to perform content based retrieval of combined text and image data it is important to ensure that the media mutually support each other, i.e., that the independent components of the combined data do not dissociate. In this work adjacency is used to associate text and images, which is also the approach taken by the search engines. It is demonstrated that there is a synergistic effect, and that retrieval classification rates increase from combined media.

Consider a collection of web pages consisting of images and adjacent text from which we want to perform unsupervised modeling, i.e., clustering into meaningful groups and possibly also supervised classification into labeled classes. Let $\boldsymbol{z} = [\boldsymbol{z}_I; \boldsymbol{z}_T]$ be the column vector of image (I) and text (T) features. Unsupervised ICA provides a probability density $p(\boldsymbol{z})$ model from which we want to identify meaningful clusters. The objective of supervised modeling is the conditional class-feature probability, $p(y|\boldsymbol{z})$, where $y = \{1, 2, \cdots, C\}$ is the class label. We will show that a simple classifier can be obtained from the unsupervised ICA.

4.3. Feature Extraction

4.3.1. Text Features

The so-called bag-of-words approach is used to represent the text. This approach is mainly motivated by simplicity and proven utility, see e.g., [8, 9, 13, 75, 80, 84, 85], although more advanced statistical natural language processing techniques can be employed [86]. The text is represented as terms, which is one word or a small set of words that present a context. Each document, i.e., collection of terms adjacent to an image, is represented by a vector: the histogram of term occurrence, as proposed in the vector space model (VSM) [85]. The term vector is usually filtered by removing low and high frequency terms. Low frequency terms do not carry meaningful discriminative information. Similarly high frequency terms (also denoted stop-words) such as the and of are common to all documents. In this paper, the stop-words were manually constructed to form a list of 585 words. Moreover, stemming is performed by merging words with different endings, e.g., ing or s. The collection of all document histograms provides the *term-document*

matrix $\boldsymbol{Z}_T = [\boldsymbol{z}_T(1), \cdots, \boldsymbol{z}_T(N)]$, where N is the number of documents.

4.3.2. Image Features

The intention is to employ VSM on image features, and previous work [46, 47, 48] indicate that the VSM in combination with latent semantic indexing (LSI) is useful. Thus we seek to construct a *feature-image* matrix $\boldsymbol{Z}_I = [\boldsymbol{z}_I(1), \cdots, \boldsymbol{z}_I(N)]$. We suggest to use lowest level image features of the ISO/IEC MPEG-7 standard [11], which aims at setting the standards for multimedia content representation, understanding and encoding. The low level image features are color and texture which are implemented using hue, saturation and value (HSV) encoding [48], and Gabor filters [38, 68, 87], respectively. Experiments indicate that increased sensitivity to the overall shape, e.g., background is obtained by dividing each image into 4×4 array patches. Color and texture features are subsequently computed for each patch.

By definition a texture is a spatially extended pattern build by repeating similar units called texels. In the deployed Gabor filter bank [48], each filter output captures a specific texture frequency and direction. Fig. 2 shows the Gabor filter impulse responses. Each image patch is convoled with the Gabor filters and the total energies of the filtered outputs [88] are then considered as the texture feature. Since each image consist of 16 patches and the filter bank has 12 filters, there are a total of $16 \cdot 12 = 192$ texture features. The HSV color space [48] is believed to better link to human

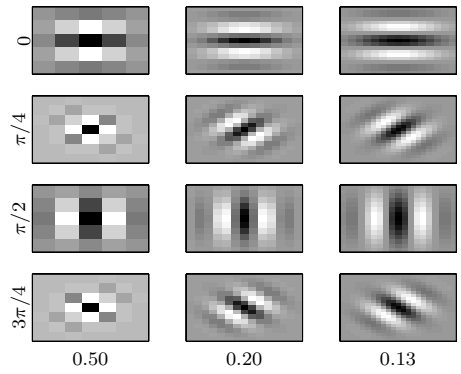


Fig. 2. Gabor filter bank used for texture feature extraction. Combining four directions $\theta = [0, \pi/4, \pi/2, 3\pi/4]$ and three texture frequencies $f = [0.50, 0.20, 0.13]$ gives a total of 12 filters in the bank.

color perception than standard RGB. The hue (H) can be interpreted as the dominant wavelength, the saturation (S) specifies the color saturation level (zero corresponds to gray tone image), and the value (V) specifies the lightness-darkness. Each color component is quantized into 16 levels, and each image patch is represented by the 3 HSV color histograms. This gives $48 = 16 \cdot 3$ features for each of the 16 patches, i.e., in total $48 \cdot 16 = 768$ dimensions color features.

4.3.3. Data Normalization

Various normalization schemes have been investigated, and compared on the basis of classification error. We found that projecting each sample feature onto a unit sphere performed best, thus also removing difference in image sizes and document lengths. This approach also outperformed normalizing to sum one (i.e., letting features representing term probabilities), which may seem more natural. For the P-dimensional raw feature vector $\boldsymbol{z} = [\boldsymbol{z}_T; \boldsymbol{z}_{IT}; \boldsymbol{z}_{IC}]$ we have,

$$\widetilde{\boldsymbol{z}} = \left[\frac{\boldsymbol{z}_T}{\|\boldsymbol{z}_T\|} ; \frac{\boldsymbol{z}_{IT}}{\|\boldsymbol{z}_{IT}\|} ; \frac{\boldsymbol{z}_{IC}}{\|\boldsymbol{z}_{IC}\|} \right]. \tag{2}$$

where T, IT and IC refer to text, texture and color, respectively. A further normalization was done by the overall variance of each feature modality to determine the input $\boldsymbol{x}$ of our model.

$$\boldsymbol{x} \quad = \quad \left[\frac{\widetilde{\boldsymbol{z}}_T}{\sigma_T} ; \frac{\widetilde{\boldsymbol{z}}_{IT}}{\sigma_{IT}} ; \frac{\widetilde{\boldsymbol{z}}_{IC}}{\sigma_{IC}} \right], \tag{3}$$

where $\sigma_T^2 = \frac{1}{P_T} \sum_i \frac{1}{N-1} \sum_k (\widetilde{z}_{iT}(k) - \mu_{iT}))^2$, where P_T is number of text features, and similarly for σ_{IT}^2 and σ_{IC}^2.

4.4. Unsupervised ICA Modeling

We consider the simple linear noiseless ICA model $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s}$ and assume $\pi^{-1}/\cosh(s_k)$ distributed super-Gaussian priors. Due to its robustness and simplicity we will use the Infomax algorithm [26, 27, 28] although, as mentioned in Section 2, the literature provides numerous ICA learning algorithm derived from other assumptions. As suggested in [8, 9, 10] latent semantic indexing (LSI) through PCA has demonstrated to be suitable for projecting onto a subspace. That is, the model is

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s} = \boldsymbol{U}\boldsymbol{\Phi}\boldsymbol{s}, \tag{4}$$

where $\boldsymbol{U}$ is the $P \times K$ matrix of K largest eigenvectors of the covariance of $\boldsymbol{x}$, and $\boldsymbol{\Phi}$ is the $K \times K$ mixing matrix. Quadratic ICA is thus performed in the subspace $\widetilde{\boldsymbol{x}} = \boldsymbol{U}^\top \boldsymbol{x}$. The ICA model is estimated from a training set $\boldsymbol{X} = \boldsymbol{x}(1), \cdots, \boldsymbol{x}(N)$ of N related images/text data samples to yield estimates $\widehat{\boldsymbol{U}}$, $\widehat{\boldsymbol{\Phi}}$.

The major advantage of combining ICA with LSI is that the sources are better aligned with meaningful content, which has been demonstrated for text documents in [8]. The different source components provide a meaningful segmentation of the feature space and mainly one source is active for a specific feature vector as demonstrated in Fig. 3. This enables an interpretation of the estimates sources as conditional component probabilities using a softmax normalization:

$$\widehat{p}(k|\boldsymbol{x}) = \frac{\exp(\widehat{s}_k)}{\sum_{k=1}^{K} \exp(\widehat{s}_k)}, \qquad \widehat{\boldsymbol{s}} = [\widehat{s}_1, \cdots \widehat{s}_K]^\top = \widehat{\boldsymbol{\Phi}}^{-1} \widehat{\boldsymbol{U}}^\top \boldsymbol{x}. \tag{5}$$

where $\widehat{p}(k|\boldsymbol{x})$ is the probability of component k given a particular observation $\boldsymbol{x}$ and the training set.

4.4.1. Component Interpretation

In order to interpret the individual components, the K'th column of $\widehat{\boldsymbol{U}}\widehat{\boldsymbol{\Phi}}$ will constitute text and image features associated with the K'th component/content segment. Since the textual features are term-histograms we can further display high occurrence terms – keywords – which in the experimental section are demonstrated to yield meaningful interpretation of the components. In particular, we rank the terms according to probability and terms which are above a certain threshold are reported as keywords. Similarly, high values of image features associated with a component provide a compact texture and color interpretation.

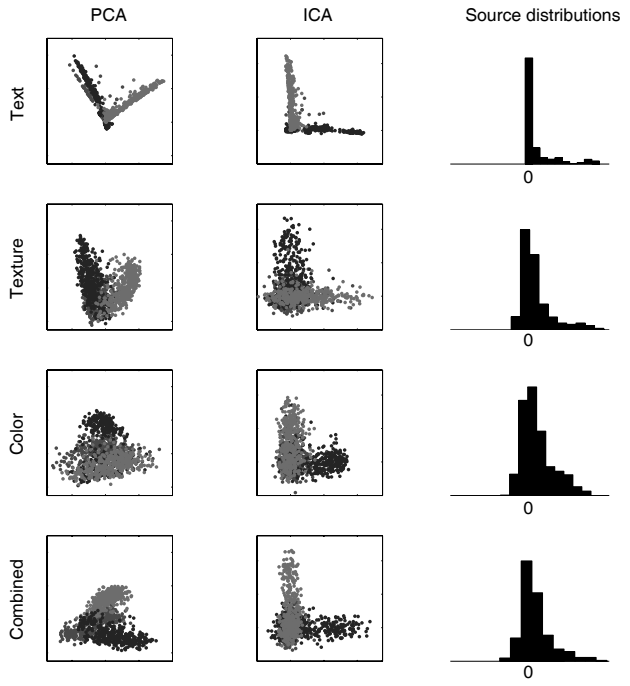


Fig. 3. Scatterplots of the text and image multimedia data, projected to a two-dimensional subspace found by PCA. Grey value of points corresponds to the three classes considered, see Fig. 4. The ray like structure strongly suggest an ICA interpretation, however, the relevance of this representation can only be determined by a subsequent inspection of the recovered source signals. As we will see in section 4.6, it turns out that there is an interesting alignment of the source signals and a manual labeling of the multimedia documents.

4.5. Probabilistic ICA Classification

Suppose that labels have been annotated to the data samples, i.e., we have a data set $\{\boldsymbol{x}(n), y(n)\}_{n=1}^{N}$ where $y(n) \in [1; C]$ are class labels. A simple probabilistic ICA classifier is then obtained as:

$$p(y|\boldsymbol{x}) = \sum_{k=1}^{K} p(y|k)p(k|\boldsymbol{x}), \tag{6}$$

where $p(k|\boldsymbol{x})$ is the conditional component probability estimated using ICA as in Eq. (5). Provided that the independent components have been estimated, the conditional class-component probabilities, $p(y|k)$ are easily estimated from data as the frequency of occurrence for specific component-class combination $k \in [1; K]$, $y \in [1; C]$, as shown by

$$\begin{aligned} \widehat{p}(y,k) &= \frac{1}{N}\sum_{n=1}^{N} \delta(\, y - y(n)\,) \cdot \delta(\, k - \arg\max_{\ell} \widehat{p}(\ell|\boldsymbol{x}(n))\,), \\ \widehat{p}(y|k) &= \frac{\widehat{p}(y,k)}{\sum_y \widehat{p}(y,k)}, \end{aligned} \tag{7}$$

where $\delta(a) = 1$ if $a = 0$, and 0 otherwise. The stagewise training of the probabilistic classifier is suboptimal, and all parameters in Eq. (6) could be estimated simultaneous, e.g., using the likelihood principle. However, the simple scheme provides a computational efficient extension of ICA to provide simple supervised classification. A more elaborate ICA mixture classifier, trained using a likelihood framework, and appropriate for multimodal data is presented in [89].

Fig. 4. Images examples from the categories Sports, Aviation and Paintball.

4.6. Experiments

The combined image and text database is obtained from the Internet by searching for images and downloading adjacent text. The adjacent text is defined as up to 150 words in one HTML paragraph tag <P> above or below the image, or within the row of a <TABLE> tag. For consistency, only jpeg images were retrieved, and we discarded images less than 72×72 pixels or pages without text. Three categories/classes of text/images were considered: Sport and Aviation and Paintball. The Sport and Aviation categories were retrieved from www.yahoo.com (17/04/2001) and the Paintball category from www.warpig.com (21/02/2002) starting from the directories and following links until depth 5.

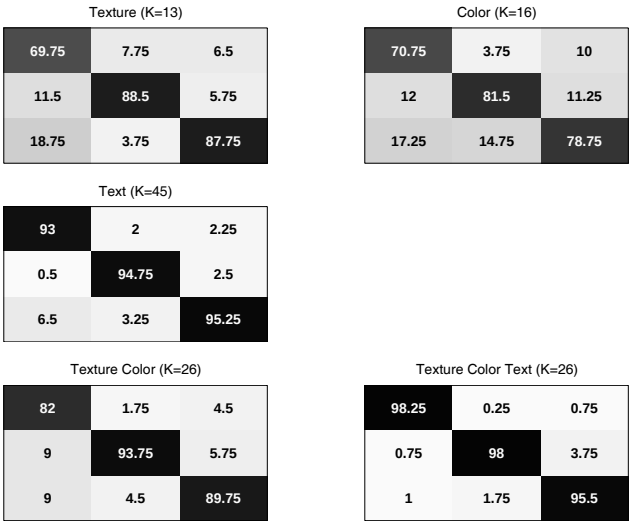
Category	Directory
Sports	recreation&sports → sports → pictures
Aviation	business&economy → transportation → aviation → pictures
Paintball	paintball → gallery → tournament

400 data from each category were downloaded resulting in a total of 1200 data sample, which were divided intro training and test sets of $3 \cdot 200$ samples each. Features were extracted as described above and resulted in 192 image texture features, 768 image color features, and 3591 text features (terms). In Fig. 4 examples of images from the categories are displayed.

4.6.1. ICA Classification

The test set classification confusion matrices obtained by using the probabilistic ICA classification scheme[1] described above are

[1]The source code of the deployed ML-ICA algorithm is available via the DTU:Toolbox [90].



Modality	Classification Error
Color	23.0%
Texture	18.0%
Texture/Color	11.5%
Text	5.7%
Combined (texture/color/text)	2.8%

Fig. 5. Optimal test classification confusion matrices (top) obtained by selecting the number of components, K, to minimize BIC criterion [9]. Rows and columns are estimated and correct classes, respectively, and the confusion reported in per cent sum to 100% column-wise. Rows/columns 1 through 3 correspond to Sports, Aviation and Paintball classes.

depicted in Fig. 5. ICA classification is done for single feature groups: texture, color, text, as well combinations texture-color and all features (texture/color/text). The number of components is selected using the BIC criterion [9] as shown in Fig. 6. Fig. 5 (bottom) further shows the order of importance of the different feature groups as expressed by the overall test classification error, and indicates the importance of extracting meaningful information. In this data set text features convey much more content information as compared to image features - both individually and in combination (texture-color). However, by combining all features the classification error is reduced approx. by a factor of 2 relative to using only text features. This indicates that the ICA classifier is able to exploit synergy among text and image features. For comparison we used the same classification framework with PCA, which resulted in classification errors of $40\% - 60\%$.

4.7. Image annotation application

An application of the suggested method is automatic annotation of text or keywords to new (test) images. In case we do not have available class labels we aim at assigning the image to a component by $\max_k p(k|\boldsymbol{x}_I)$. That is, we first need to estimate sources without knowledge of $\boldsymbol{x}_T$. It can be shown that the optimal source estimate $\max_{\boldsymbol{s}} p(\boldsymbol{s}|\boldsymbol{x}_I)$ is obtained for $\boldsymbol{x}_I = (\widehat{\boldsymbol{U}}\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{s}})_I = \widehat{\boldsymbol{U}}_I\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{s}}$ with definitions as in Eq. (4), and with $_I$ begin the columns corresponding to image features. That is, $\widehat{\boldsymbol{s}} = \widehat{\boldsymbol{\Phi}}^{-1}\widehat{\boldsymbol{U}}_I^{\top}\boldsymbol{x}_I$. If class labels are

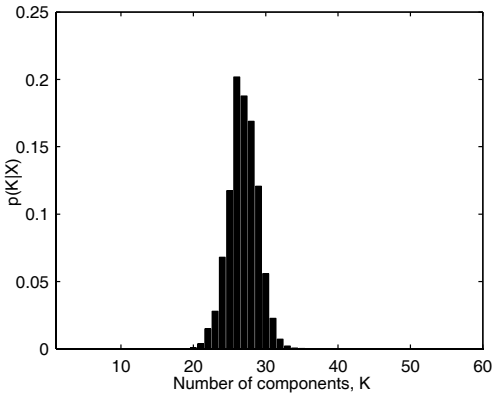


Fig. 6. Selection of components using BIC in the case of combined data (texture/color/text). In BIC, an asymptotic (large data set) estimate of $p(K|\boldsymbol{X})$ is computed, viz. the probability of the model having K components given the training data. The most probable model is obtained for $K = 26$ and the associated classification errors are reported in Fig. 5.

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

Fig. 7. Annotation of 3 images not used for training the model. Keywords for I_3 are team names.

available, we can further assign class label by $\max_y p(y|\boldsymbol{x}_I)$. In both cases associated descriptive keyword can be generated as described in Section 4.4.1. An example of automatic image annotation is presented in Fig. 7.

5. CONCLUSIONS AND PERSPECTIVES

This paper demonstrated the potential of independent component analysis and blind sources separation methods for modeling, understanding and intelligent processing of multimedia data. The unique feature of ICA is that it provides unsupervised grouping of data which are amenable for interpretation and well-aligned with human perception.

A number of multimedia applications involving ICA/BBS and related method have been reviewed and are summarized in Table 1. The applications cover image/video, multimodal brain, audio, text, and combined media data. The potential of ICA/BSS is clear, however, there still exist a number of interesting open research issues, which are further discussed below.

We provided an extended review of our recent work on mod-

eling combined text/image data for the purpose of cross-media retrieval and web search. A more elaborate analysis of the assumptions of ICA emphasized the advantages of ICA for analysis of text and image data. Further, it was demonstrated that the synergy among text and image features leads to better classification performance, thus the common independent component space convey useful information related to the content of an image and adjacent text information. Finally, we provided an application example of automatic annotation of text to images using the suggested ICA framework.

ICA/BSS is an interesting object for future research. Better understanding of the concept of independent components of multimedia data, which seems to be well aligned with human perception, might provide increased utility. An incomplete list of future challenges includes:

- The construction of relevant and specific multimedia features for which linear ICA is the appropriate model.
- 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.
- Processing from mono binaural audio signals, and in general handling of underdetermined convolutive mixture models, e.g., by invoking more specific audio priors.
- Training and recall in large multimedia databases is a significant computational issue.
- Estimation and optimization beyond natural gradient based schemes has largely been ignored, hence, we anticipate a need for advanced active data subset selection methods, online learning algorithms, and adaptation to changing environment.
- Intelligent fusion of media types and the ability to use both labeled/unlabeled data.

6. ACKNOWLEDGMENTS

This work was partly funded by the Danish Research Councils through the Signal and Image Processing for Telemedicine (SITE) program.

7. REFERENCES

[1] L. Guan, S.-Y. Kung, and J.-N. Hwang, "Intelligent multimedia processing," in *Multimedia Image and Video Processing*, L. Guan, S.-Y. Kung, and J. Larsen, Eds., pp. 131–173. CRC Press, Sep. 2000.

[2] S.-Y. Kung and J.-N. Hwang, "Neural networks for intelligent multimedia processing," *Proceedings of the IEEE*, vol. 86, no. 6, pp. 1244–1272, 1998.

[3] A. Bell and T. Sejnowski, "Edges are the independent components of natural scenes," in *Advances in Neural Information Processing Systems*, M. C. Mozer, M. I. Jordan, and T. Petsche, Eds. 1997, vol. 9, pp. 831–837, The MIT Press.

[4] A. J. Bell and T. J. Sejnowski, "Learning the higher order structure of a natural sound," *Network: Computation in Neural Systems*, vol. 7, no. 2, pp. 261–267, 1996.

[5] T.-P. Jung, S. Makeig, M. J. McKeown, A. Bell, T.-W. Lee, and T. Sejnowski, "Imaging brain dynamics using independent component analysis," *IEEE Proceedings*, vol. 89, no. 7, pp. 1107–1122, 2001.

[6] J. H. van Hateren and D. L. Ruderman, "Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex," in *Proceedings of the Royal Society of London - B - Biological Sciences*, 1998, vol. 265, pp. 2315–2320.

[7] S. A. Abdallah and M. D. Plumbley, "If edges are the independent components of natural images, what are the independent components of natural sounds?," in *Proceedings of ICA'2001*, T.-W. Lee, T.-P. Jung, S. Makeig, and T. Sejnowski, Eds., San Diego, USA, December 2001, pp. 534–539.

[8] T. Kolenda, L.K. Hansen, and S. Sigurdsson, "Independent components in text," in *Advances in Independent Component Analysis*, M. Girolami, Ed., pp. 229–250. Springer-Verlag, 2000.

[9] T. Kolenda, L.K. Hansen, and J. Larsen, "Signal detection using ICA: Application to chat room topic spotting," in *Proceedings of ICA'2001*, T.-W. Lee, T.-P. Jung, S. Makeig, and T. Sejnowski, Eds., San Diego, USA, December 2001, pp. 540–545.

[10] T. Kolenda, L. K. Hansen, J. Larsen, and O. Winther, "Independent component analysis for understanding multimedia content," in *Proceedings of IEEE Workshop on Neural Networks for Signal Processing XII*, H. Bourlard, T. Adali, S. Bengio, J. Larsen, and S. Douglas, Eds., Matigny, Valais, Switzerland, Sep. 4-6 2002, pp. 757–766, IEEE Press.

[11] J.M. Martínez, "Overview of the MPEG-7 standard (version 5.0)," Tech. Rep., ISO, Coding of moving pictures and audio, 2001, http://mpeg.telecomitalialab.com/standards/mpeg-7/mpeg-7.htm.

[12] J. E. Jackson, *A User's Guide to Principal Components*, John Wiley & Sons, Inc., 1991.

[13] L. K. Hansen, J. Larsen, and T. Kolenda, "On independent component analysis for multimedia signals," in *Multimedia Image and Video Processing*, L. Guan, S.-Y. Kung, and J. Larsen, Eds., pp. 175–199. CRC Press, Sep. 2000.

[14] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[15] H. Attias, "Independent factor analysis," *Neural Computation*, vol. 11, no. 4, pp. 803–851, 1999.

[16] H. Attias, "Learning a hierarchical belief network of independent factor analysers," in *Advances in Neural Information Processing Systems*, M. S. Kearns, S. A. Solla, and D. A. Cohn, Eds., 1999, vol. 11, pp. 361–367.

[17] T.-W. Lee, M. S. Lewicki, and T. J. Sejnowski, "ICA mixture models for unsupervised classification of non-gaussian sources and automatic context switching in blind signal separation," *IEEE Transactions on Pattern Recognition and Machine Intelligence*, vol. 22, no. 10, pp. 1–12, October 2000.

[18] A. Hyvärinen, P. O. Hoyer, and M. Inki, "Topographic independent component analysis," *Neural Computation*, vol. 13, no. 7, pp. 1527–1558, 2001.

[19] P. J. Huber, "Projection pursuit," *The Annals of Statistics*, vol. 13, no. 2, pp. 435–475, 1985.

[20] P. Tino and I. Nabney, "Hierarchical GTM: constructing localized non-linear projection manifolds in a principled way," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 639–656, 2002.

[21] P. Smaragdis and M. Casey, "Audio/visual independent components," in *Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation*, Nara, Japan, April 2003.

[22] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*, Academic Press Ltd., 1979.

[23] A. Vinokourov, J. Shawe-Taylor, and N. Cristianini, "Inferring a semantic representation of text via cross-language correlation analysis," in *Advances in Neural Information Processing Systems*, S. Becker, S. Thrun, and K. Obermayer, Eds. 2003, vol. 15, The MIT Press.

[24] A.S. Lukic, M.N. Wernick, L.K. Hansen, and S.C. Strother, "An ICA algorithm for analyzing multiple data sets," in *Proceedings of ICIP'02 IEEE Int. Conf. on Image Processing 2002*, Rochester, New York, September 2002, pp. 821–824.

[25] A. Cichocki and S.-i. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, John Wiley, Chichester, UK, April 2002.

[26] A. Cichocki, R. Unbehauen, and E. Rummert, "Robust learning algorithm for blind separation of signals," *Electronics Letters*, vol. 30, no. 17, pp. 1386–1387, August 1994.

[27] S.-i. Amari, A. Cichocki, and H.H. Yang, "A new learning algorithm for blind signal separation," in *Advances in Neural Information Processing Systems NIPS-1995*, Cambridge, MA, 1996, vol. 8, pp. 757–763, MIT Press.

[28] A. Bell and T.J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, 1995.

[29] S.-i. Amari, "Natural gradient learning for over- and under-complete bases in ICA," *Neural Computation*, vol. 11, pp. 1875–1883, 1999.

[30] M. S. Lewicki and T. J. Sejnowski, "Learning overcomplete representations," *Neural Computation*, vol. 12, pp. 337–365, 2000.

[31] J.-F. Cardoso, "The three easy routes to independent component analysis; contrasts and geometry," in *Proceedings of ICA'2001*, T.-W. Lee, T.-P. Jung, S. Makeig, and T. Sejnowski, Eds., San Diego, USA, December 2001, pp. 1–6.

[32] F. Bach and M. I. Jordan, "Kernel independent component analysis," *Journal of Machine Learning Research*, vol. 3, pp. 1–48, 2002.

[33] P. Højen-Sørensen, O. Winther, and L. K. Hansen, "Mean field approaches to independent component analysis," *Neural Computation*, , no. 4, pp. 889–918, 2002.

[34] H. Lappalainen, "Ensemble learning for independent component analysis," in *Proceedings of ICA'99*, Aussois, France, 1999, pp. 7–12.

[35] T.-W. Lee, *Independent Component Analysis: Theory and Applications*, Kluwer Academic Publishers, September 1998.

[36] T.-W. Lee, M. Girolami, A.J. Bell, and T.J. Sejnowski, "A unifying information-theoretic framework for independent component analysis," *Int. Journ. on Comp. and Math. with Appl.*, vol. 31, no. 11, pp. 1–21, March 2000.

[37] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley, New York, 2001.

[38] P. O. Hoyer and A. Hyvärinen, "Independent component analysis applied to feature extraction from colour and stereo images," *Computation in Neural Systems*, vol. 11, no. 3, pp. 191–210, 2000.

[39] J. Hurri and A. Hyvärinen, "Simple-cell-like receptive fields maximize temporal coherence in natural video," *Neural Computation*, 2003.

[40] R. Larsen, "Decomposition using maximum autocorrelation factors," *Journal of Chemometrics*, vol. 16, no. 8–10, pp. 427–435, 2002.

[41] X. Zhou, B. Moghaddam, and T. S. Huang, "ICA-based probabilistic local appearance models," in *Proceedings of International Conference on Image Processing (ICIP'01)*, October 2001, http://www.merl.com/papers/docs/TR2001-29.pdf.

[42] A. Hyvärinen, "Sparse code shrinkage: Denoising of nongaussian data by maximum likelihood estimation," *Neural Computation*, vol. 11, no. 7, pp. 1739–1768, 1999.

[43] K. Takaya and K.-Y. Choi, "Detection of facial components in a video sequence by independent component analysis," in *Proceedings of ICA'2001*, T.-W. Lee, T.-P. Jung, S. Makeig, and T. Sejnowski, Eds., San Diego, USA, December 2001, pp. 260–265.

[44] S. Noel and H. Szu, "Multimedia authenticity with independent-component watermarks," in *14th Annual International Symposium on Aerospace/Defense Sensing Simulation, and Controls*, Orlando, Florida, April 2000, http://www.isse.gmu.edu˜snoel/AeroSense%202000.pdf.

[45] F. Sattar D. Yu and K. K. Ma, "Watermark detection and extraction using independent component analysis method," *EURASIP Journal on Applied Signal Processing*, vol. 2002, no. 1, pp. 92–104, 2002.

[46] M. La Cascia, S. Sethi, and S. Sclaroff, "Combining textual and visual cues for content based image retrieval on the world wide web," in *IEEE Workshop on ContentBased Access of Image and Video Libraries*. 1998, pp. 24–28, IEEE Computer Society.

[47] Z. Pečenović, "Image retrieval using latent semantic indexing," M.S. thesis, AudioVisual Communications Lab, Ecole Polytechnique F'ed'erale de Lausanne, Switzerland, 1997.

[48] T. Westerveld, "Image retrieval: Content versus context," in *Proceedings Content Based Multimedia Information Access, RIAO 2000*, Paris, France, 2000, pp. 276–284.

[49] A. Vinokourov, D. Harddon, and J. Shawe-Taylor, "Learning the semantics of multimedia content with application to web image retrieval and classification," in *Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation*, Nara, Japan, April 2003.

[50] V. D. Calhoun, T. Adali, L. K. Hansen, J. Larsen, and J. J. Pekar, "ICA of functional MRI data: An overview," in *Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation*, Nara, Japan, April 2003.

[51] K. Torkkola, "Blind separation for audio signals – are we there yet," in *Proceedings of ICA'99*, Aussois, France, January 1999, pp. 239–244.

[52] L. C. Parra and C. V. Alvino, "Geometric source separation: Merging convolutive source separation with geometric beamforming," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 6, pp. 352–362, 2002.

[53] E. Vincent et al., "A tentative typology of audio source separation tasks," in *Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation*, Nara, Japan, April 2003.

[54] J. Klingseisen and M. D. Plumbley, "Towards musical instrument separation using multiple-cause neural networks," in *Proceedings of ICA'2000*, Helsinki, Finland, June 2000, pp. 447–452.

[55] M. D. Plumbley, S. A. Abdallah, J. P. Bello, M. E. Davies, G. Monti, and M. B. Sandler, "Automatic music transcription and audio source separation," *Cybernetics and Systems*, vol. 33, no. 6, pp. 603–627, 2002.

[56] F. Berthommier and S. Choi, "Evaluation of casa and bss models for subband cocktail party speech recognition," in *Proceedings of ICA'2001*, T.-W. Lee, T.-P. Jung, S. Makeig, and T. Sejnowski, Eds., San Diego, USA, December 2001, pp. 301–306.

[57] S. Dubnov and A. Ben-Shalom, "Review of ICA and hos methods for query by similarity of natural sounds and sound effects," in *Proceedings of ICA'2003*, Nara, Japan, April 2003.

[58] M. Casey and A. Westner, "Separation of mixed audio sources by independent subspace analysis," in *Proceedings of the International Computer Music Conference, ICMA*, Berlin, August 2000.

[59] C. L. Isbell, Jr. and P. Viola, "Restructuring sparse high dimensional data for effective retrieval," in *Advances in Neural Information Processing Systems*, M. S. Kearns, S. A. Solla, and D. A. Cohn, Eds., 199, vol. 11, pp. 480–486.

[60] E. Bingham, J. Kuusisto, and K. Lagus, "ICA and SOM in text documents analysis," in *ACM SIGIR 2002 International Conference on Research and Development in Information Retrieval*, Tampere, Finland, August 11-15 2002, ACM, pp. 361–362.

[61] F. Å. Nielsen and L. K. Hansen, "Author cocitation analysis of articles from "NeuroImage"," *NeuroImage*, vol. 13, no. 6, part 2, pp. S212, June 2001.

[62] F. Å. Nielsen, *Neuroinformatics in Functional Neuroimaging*, Ph.D. thesis, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark, 2001.

[63] D. Cohn and T. Hofmann, "The missing link - a probabilistic model of document content and hypertext connectivity," in *Advances in Neural Information Processing Systems*, T. K. Leen, T. G. Dietterich, and V. Tresp, Eds. 2001, vol. 13, pp. 430–436, MIT Press.

[64] Y. Wang, Z. Liu, and J.C. Huang, "Multimedia content analysis-using both audio and visual clues," *IEEE Signal Processing Magazine*, vol. 17, no. 6, pp. 12–36, 2000.

[65] T. Darrell, J. Fisher, P. Viola, and B. Freeman, "Audio-visual segmentation and the cocktail party effect," in *Proceedings International Conference Advances in Multimodal Interfaces (ICMI)*, October 2000, vol. 1948, pp. 32–40.

[66] J. W. Fisher III, T. Darrell, W.T. Freeman, and P. Viola, "Learning joint statistical models for audio-visual fusion and segregation," in *Advances in Neural Information Processing Systems*, T. K. Leen, T. G. Dietterich, and V. Tresp, Eds. 2001, vol. 13, MIT Press.

[67] D. Sodoyer, J.-L. Schwartz, L. Girin, J. Klinkisch, and C. Jutten, "Separation of audio-visual speech sources: A new approach exploiting the audio-visual coherence of speech stimuli," *EURASIP JASP*, vol. 1, pp. 1165–1173, 2002.

[68] B. MacLennan, "Gabor representations of spatiotemporal visual images," Tech. Rep. CS-91-144, Computer Science, Univ. of Tennessee, 1994.

[69] L. Molgedey and H. Schuster, "Separation of independent signals using time-delayed correlations," *Physical Review Letters*, vol. 72, no. 23, pp. 3634–3637, 1994.

[70] J. Eakins and M. Graham, "Content based image retrieval," Tech. Rep., University of Northumbria at Newcastle, 1999, http://www.unn.ac.uk/iidr/report.html.

[71] M. Flickner et al., "Query by image and video content: the qbic system," *IEEE Computer*, vol. 28, no. 9, pp. 23–32, 1995.

[72] A. Gupta et al., "The virage image search engine: an open framework for image management," *Storage and Retrieval for Image and Video Databases IV*, vol. Proceedings SPIE 2670, pp. 76–87, 1996.

[73] J. Feder, "Towards image content based retrieval for the world-wide web," *Advanced Imaging*, vol. 11, no. 1, pp. 26–29, 1996.

[74] J. Eakins, "Towards intelligent image retrieval," *Pattern Recognition*, vol. 35, pp. 3–14, 2002.

[75] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *J. Amer. Soc. for Inf. Science*, vol. 41, pp. 391–407, 1990.

[76] Gil-Jin Jang and Te-Won Lee, "A probabilistic approach to single channel blind signal separation," in *Advances in Neural Information Processing Systems*, S. Becker, S. Thrun, and K. Obermayer, Eds. 2003, vol. 15, The MIT Press.

[77] E. Saund, "A multiple cause mixture model for unsupervised learning," *Neural Computation*, vol. 7, no. 1, pp. 51–71, January 1995.

[78] A. Kabán and M. Girolami, "A dynamic probabilistic model to visualise topic evolution in text streams," *Journal of Intelligent Information Systems*, vol. 18, no. 2–3, pp. 107–125, March-May 2002.

[79] E. Bingham, A. Kabán, and M. Girolami, "Finding topics in dynamical text: application to chat line discussions," in *10th International World Wide Web Conference (WWW10)*, Hong Kong, May 1-5 2001, pp. 198–199.

[80] E. Bingham, "Topic identification in dynamical text by extracting minimum complexity time components," in *Proceedings of ICA'2001*, T.-W. Lee, T.-P. Jung, S. Makeig, and T. Sejnowski, Eds., San Diego, USA, December 2001, pp. 546–551.

[81] J. C. French and C. L. Villes, "Exploiting coauthorship to infer topicality in a digital library of computer science technical reports," Technical Report CS-96-20, Department of Computer Science, University of Virginia, December 1996.

[82] K. W. McCain, "Mapping authors in intellectual space: A technical overview," *Journal of the American Society for Information Science*, vol. 41, no. 6, pp. 433–443, September 1990.

[83] T. Westerveld, "Probabilistic multimedia retrieval," in *Proceedings of the 25th Annual International Conference on Research and Development in Information Retrieval (SIGIR 2002)*, 2002.

[84] K. Nigam, A.K. McCallum, S. Thrun, and T. Mitchell, "Text classification from labeled and unlabeled documents using em," *Machine Learning*, vol. 39, pp. 103–134, 2000.

[85] G. Salton, *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*, Addison-Wesley, 1989.

[86] C. D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*, MIT Press, Cambridge, Massachusetts, 1999.

[87] M. Pötzsch and M. Rinne, "Gabor wavelet transformation," Internet, 1996, http://www.neuroinformatik.ruhr-uni-bochum.de/ini/VDM/research/computerVision/imageProcessing/wavelets/gabor/contents.html.

[88] H. Knutsson and G. H. Granlund, "Texture analysis using two-dimensional quadrature filters," in *Proceedings IEEE Computer Society Workshop on Computer Architecture for Pattern Analysis and Image Database Management*, Pasadena, CA, 1983, pp. 206–213, IEEE Computer Soc. Press.

[89] T-W. Lee, M. S. Lewicki, and T. J. Sejnowski, "Unsupervised classification with non-gaussian mixture models using ICA," in *Advances in Neural Information Processing Systems*, M. S. Kearns, S. A. Solla, and D. A. Cohn, Eds., Cambridge MA, 1999, vol. 11, pp. 508–514, MIT Press.

[90] T. Kolenda et al., "DTU:Toolbox," Software, Informatics and Mathematical Modelling, Technical University of Denmark, 2002, http://isp.imm.dtu.dk/toolbox/.