

Blind Separation of Spatio-temporal Data Sources

Hilit Unger and Yehoshua Y. Zeevi

Department of Electrical Engineering, Technion – Israel Institute of Technology
Haifa 32000, Israel

hilitg@tx.technion.ac.il, zeevi@ee.technion.ac.il

Abstract. ICA and similar techniques have been previously applied to either one-dimensional signals or still images. We consider the problem of blind separation of dynamic sources, i.e. functions of both time and two spatial variables. We extend the Sparse ICA (SPICA) approach and apply it to a sliding data cube, defined by the two dimensions of the visual scene and the extent in time over which the mixing problem can be considered to be stationary and linear. This framework and formalism are applied to two special problems encountered in two different fields: The first deals with separation of dynamic reflections from a desired moving visual scene, without having any *a priori* knowledge on the structure of the images and/or their statistics. The second problem concerns blind separation of ‘neural cliques’ from the background firing activity of a neural network. The approach is generic in that it is applicable to any linearly mixed dynamic sources.

1 Introduction

Most of the research devoted to the problem of Blind Source Separation (BSS) has been concerned with either one-dimensional functions of time or static images (for references, see [1]). Yet, many physical systems generate linear mixtures of dynamic data sets. In biomedical applications, for example, those encountered in functional MRI, one is interested in the dynamic activity of specific loci of the brain. Another application concerns video sequences acquired through a semireflective medium and thereby contaminated by superimposed reflections. The video captures the dynamics of events. Since most real-world scenarios are dynamic, it is desirable to extend the BSS techniques to functions of both time and space.

Our first application deals with separation of dynamic images, such as video signals. In such applications it is desirable to eliminate reflections superimposed on a dynamic scene recorded through a glass windshield of a moving vehicle, or eliminate the reflections of the sun superimposed on the image of the visual environment observed through the cockpit of an airplane. The video sequence acquired in such cases can be represented as a three-dimensional (volumetric) cube, in which spatial images are stacked along a third axis (Fig. 1).

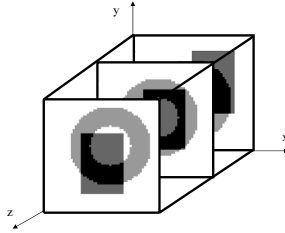


Fig. 1. Video sequence considered as a volumetric (cubic) data set. Shown is a data set comprised of three consecutive frames obtained from the sequence. Note the relative movement of the objects.

The second application presented here is concerned with the recording and analysis of biological neural networks, where there is a concerted effort to decipher the simultaneous messages signaled by the spatio-temporal firing patterns typical of the firing activity of massively connected neural networks. This application motivated our current study.

2 Sparse ICA (SPICA)

In Blind Source Separation an N -channel sensor signal x_i is generated by M unknown scalar source signals s_i , linearly mixed together by an unknown constant $N \times M$ mixing matrix A . In matrix notation, the N -dimensional vector of mixtures, X , is equal to the product of the $N \times M$ mixing matrix by the M -dimensional sources vector, S :

$$X = A \cdot S. \quad (1)$$

Under the assumption that the sources are statistically independent, the BSS method yields an estimate of \tilde{A} , the unknown mixing matrix, without prior knowledge of the sources and/or the mixing process. The sources are recovered (up to permutation and scale) by using an inverse of the estimated mixing matrix, provided it exists:

$$\tilde{S} = \tilde{A}^{-1} \cdot X. \quad (2)$$

It has been shown that when sources are sparse, they can be easily recovered from their linear mixtures using simple geometrical methods [2],[5]. This is based on the observation that whenever sources are sparse, there is a high probability that each data point in each mixture will result from the contribution of only one source. If we plot the N -dimensional scatter plot wherein each axis represents one of the mixtures, a co-linear cluster emerges with a specific orientation for each source. It can be shown that the coordinates of the vectors representing the centroids of these clusters correspond to the columns of the mixing matrix A . The simplest way to estimate the mixing matrix is to calculate the orientations of the clusters and select the optimal M angles from the histogram of angles. Another algorithm projects the data points onto a hemisphere, then uses clustering

(such as Fuzzy C-means) in order to recover the orientations. Another related maximum-likelihood-based approach is the well-known Infomax [3].

3 Sparse Decompositions

3.1 Overcomplete Representations

Natural images and image sequences are not typically sparse. In order to exploit the methods previously described, we have to apply a transformation that yields a sparse representation of the signals. It has been shown that for a wide range of natural images, smoothed derivative operators yield a good sparsification results [4]. However, an overcomplete representation obtained, for example, by the Wavelet Packet transform (WPT, proposed in [5]) matches better the specific structure of a given set of images and thereby yields better sparsification which, in turn, facilitates and improves the estimation of the mixing matrix.

3.2 WP Transform

According to the formalism of the Wavelet Packet transform, a signal is recursively decomposed into its approximation (L) and detail (H) subspaces.

In the case of 2D signals, using separable wavelets, the signal is decomposed into its approximation and vertical, horizontal and diagonal details sub-images. For 3-dimensional data cube, the signal is decomposed into 8 sub-volumes (Fig. 2).

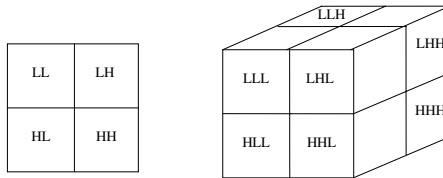


Fig. 2. WP decomposition. Left: 2D decomposition. Right: 3D decomposition.

We chose to use a separable transformation, for the sake of simplicity, by transforming rows first, then columns and then time (depth) axis. Nonseparable wavelets offer certain advantages, but are much more complex to deal with [6]. Their application in the context of sparsification is beyond the scope of this study.

3.3 Source Separation Using the WPT

After the mixture signals are decomposed into WP tree nodes using the WPT [5], a quality criterion is calculated for each node. The quality criteria should assign

high values for sparse nodes and lower values for less sparse nodes. Common choices for quality criteria are entropy or global distortion. The best node (or the top few nodes) is chosen and used as input data for the BSS algorithm.

Using the WPT has another advantage: because of downsampling in the process of the transform, the number of data points in each node is significantly smaller than the number of data points in the mixture signals, which speeds up the separation process.

4 BSS of Dynamic Reflections

Fig. 3 depicts an example of a typical scenario wherein a virtual (reflected) image is superimposed on a visual scene.

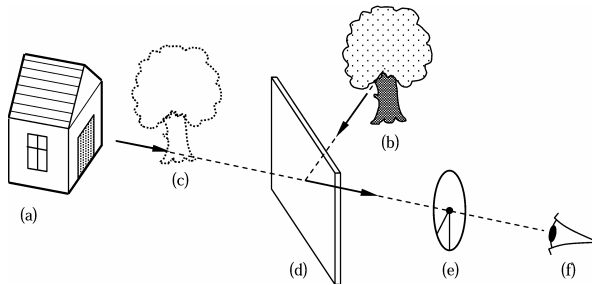


Fig. 3. A typical optical setup including a semireflecting windshield: (a) - object 1, (b) - object 2, (c) - virtual image of object 2, (d) - a semireflective lens, (e) - polarizer, (f) - camera. (adopted from [4]).

In the context of separation of reflections, the BSS problem usually reduces to the case of $M=2$ sources. The observed mixture is then given by

$$x(\xi_1, \xi_2, t) = a_{11}s_1(\xi_1, \xi_2, t) + a_{12}s_2(\xi_1, \xi_2, t), \quad (3)$$

where x , s_1 and s_2 are dynamic images, usually acquired as video sequences. It is assumed here that the dynamics of the image and of the superimposed reflections are limited to planar translation of rigid bodies. The more difficult problem of non-planar motion and rotation as well as non-rigid distortion are beyond the scope of this paper, and will be dealt with elsewhere. Likewise, the coefficients a_{11} and a_{12} are assumed to be constant, approximating spatial invariance and linear mixing [4].

Since the reflected light is polarized, by using a linear polarizer, the relative weights of the two mixed video sequences can be varied to yield N different mixtures of the form:

$$x_n(\xi_1, \xi_2, t) = a_{n1}s_1(\xi_1, \xi_2, t) + a_{n2}s_2(\xi_1, \xi_2, t) : n = 1, \dots, N. \quad (4)$$

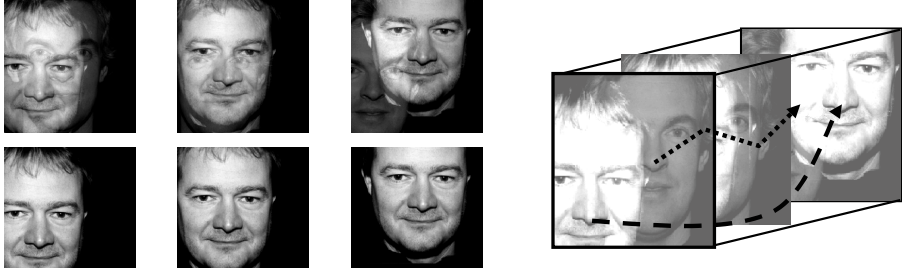


Fig. 4. Left frame of 6 images, simulation of blind separation of dynamic (moving) image from a superimposed reflection: frames from one mixture (up), and frames from one sequence of a recovered source (bottom). Right: data cube of one mixture. The arrows trace the trajectories of movements of the image and reflection, relatively to a stationary background.

Thus, we can use two or more video sequences obtained with different polarizations and separate objects and reflections. Simulation results are shown in Fig. 4.

5 BSS of Neural Cliques

In recent years, new optical [7] and electrical [8] imaging techniques for simultaneous recording of activity of populations of neurons in the brain tissue were developed. Whereas traditional methods for detection of action potentials in neurons were limited to a small number of neurons, it is now possible to record massive neural activity with spatial resolution of a single cell and a temporal resolution of a single action potential. It is therefore important to develop new techniques for the analysis of such activity.

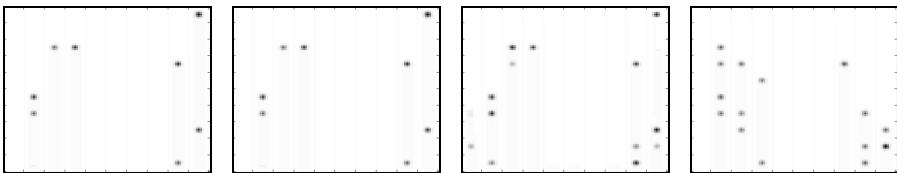


Fig. 5. Four states characterizing activity of an artificial neural network. The firing patterns depict functional phenomenon of localization.

The study of large populations of neurons enables to identify and analyze neural phenomena such as Synfire chains [9]: waves of synchronous neural activity that propagate over different areas of the biological neural network. It is believed that such separated activities represent processes related to higher level brain functions, e.g. percepts. Examining such spatio-temporal patterns of firing

neurons, or ‘neural cliques’, led us to the assumption that there are underlying sources that are mixed together into each observed firing pattern. To understand the concept of cliques in the context of spatio-temporal neural network activity, recall the representation of spatio-temporal data as a cubical data set (Fig. 1). Here each frame corresponds to a slice along time axis of duration Δt . A clique then corresponds to correlated pattern of activity of two or more such slices of duration $T > \Delta t$.

To provide some intuitive insight into the analysis of neural cliques by means of Blind Source Separation technique, we generate data using CSIM circuit-tool; a simulator for neural networks [11]. The network connectivity is randomized, and one input neuron excites a random subset of the network. The output discrete spiking activity is then converted into continuous analog signal which, in turn, is quantized for further computation. It is interesting to observe that such a random network that is not endowed with any spatial localization structure, exhibits functional localization such as depicted in Fig. 5.

We do not have prior knowledge of the number of sources, therefore we need to estimate it by using the PCA technique [10] or geometrical version of an ICA-type approach, that permits separation of a larger number of sources than the given number of mixtures [5].

We assume that each neural clique has a finite (yet unknown) duration and that the neural activity is quasi-stationary over time, i.e. mixing coefficients remain constant over the duration of the clique, but may vary over longer periods of time. The separation problem is still endowed with a large number of degrees of freedom: the duration of the examined mixtures, the starting frame of each observation and the number of observations considered. Choosing those parameters carefully is crucial in order to achieve meaningful results. The optimal parameters for this problem are yet to be studied.

Using our BSS approach, we then project the data deduced from slices onto a scatter plot wherein each axis represents activity in one mixture slice. Each point then represents the activity of a neuron at a specific time in the slice.

Investigating the mutual activity of two slices, one often observes that two slices that are selected within the duration of co-activation do not necessarily exhibit coincidence of spike activity. In fact, the spatio-temporal activity may be almost exclusively restricted to only one slice. Under these circumstances, the distribution of activity over the scatter plot will form either a vertical or a horizontal cluster (Fig. 6 left). To compare with, when the second slice is partially co-active in space and time, more clusters emerge over the scatter plot (Fig. 6 middle). These clusters should provide some insight into the structure of the network, and functionally are indicative of clique-type activity. The full meaning of such embodiment of co-activation has yet to be further studied. It should be observed though, that unlike the previous application of video data, here we face a non-linear phenomenon that limits the power of ICA-type techniques. Nevertheless, the formalism and approach of projecting the data onto a scatter plot is powerful in gaining some insight into the structure of non-linearly interacting sources (or cliques).

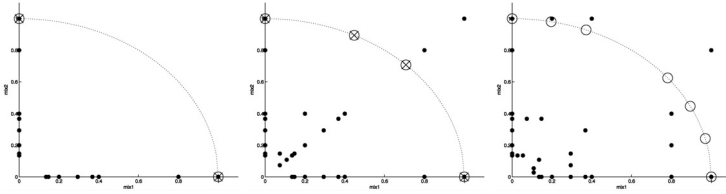


Fig. 6. Left, two slices with no co-activations. Middle, two slices with 2 emergent cliques. Cluster centers are marked with \times . Right, uncorrelated activity.

6 Conclusions

The extension of the Sparse ICA approach to three-dimensional problems broadens the range of ill-posed BSS problems that can be dealt with efficiently by providing relatively simple solutions to complex problems. Yet, the underlying assumptions of stationarity and linearity are not always met. More powerful results have yet to await the extension of these sparse ICA technique to the non-linear and non-stationary regime. This ambitious approach is under investigation.

The example of removal of reflections from a video sequence demonstrates that the sparse ICA approach is easily extended into the three-dimensional space and provides good results in the case of dynamic reflections.

Finding sources of neural activity in neural networks is a much more demanding and challenging task. Unlike the physics of separation of superimposed reflections, which can, to a good approximation, be considered linear, the neural cliques separation is necessarily non-linear, and most likely non-stationary. Nevertheless, as we have demonstrated here, the novel approach of using BSS techniques in isolation of the fingerprints of coherent neural activity from a neural network, can be instrumental in highlighting the functions of biological neural networks. It may be also instrumental in studies attempting to reverse engineer the structures of linear skeletons of such networks using spatio-temporal spiking activity.

Acknowledgement

Research supported in part by the Ollendorff Minerva center, by the HASSIP Research Network Program HPRN-CT-2002-00285, sponsored by the European Commission and by the Fund for Promotion of Research at the Technion.

References

1. Special issue on Independent Components Analysis. In: J. Machine Learning Research. Volume 4. (2003)
2. Zibulevsky, M., Pearlmutter, B.A.: Blind source separation by sparse decomposition in a signal dictionary. *Neural Comp.* **13** (2001) 863–882

3. Cardoso, J.: Infomax and maximum likelihood for blind source separation. *IEEE Signal Processing Letters* **4** (1997) 112–114
4. Bronstein, A., Bronstein, M., Zibulevsky, M., Zeevi, Y.Y.: Separation of reflections via sparse ICA. In: *ICIP03*. (2003) 313–316
5. Kisilev, P., Zibulevsky, M., Zeevi, Y.Y.: A multiscale framework for blind separation of linearly mixed signals. *J. Mach. Learn. Res.* **4** (2003) 1339–1363
6. Stanhill, D., Zeevi, Y.Y.: Two-dimensional orthogonal wavelets with vanishing moments. *IEEE Transactions on Signal Processing* **44** (1996) 2579–2590
7. Smetters, D., Majewska, A., Yuste, R.: Detecting action potentials in neuronal populations with calcium imaging. *Methods* **18** (1999) 215–221
8. Shahaf, G., Marom, S.: Learning in networks of cortical neurons. *J. of Neuroscience* **21** (2001) 8782–8788
9. Abeles, M.: *Corticonics, neural circuits of the cerebral cortex*. Cambridge University Press (1991)
10. Duda, R.O., Hart, P.E., Stork, D.G.: *Pattern Classification*. Wiley (2000)
11. <http://www.lsm.tugraz.at/csim/index.html>.