

Sparse Kernel Independent Component Analysis for Blind Source Separation

Asif Khan and Intaek Kim*

*Signal Processing Lab, Department of Communication Engineering
Myongji University, Yongin, 449-728, Korea*

(Received August 4, 2008 : accepted September 2, 2008)

We address the problem of Blind Source Separation (BSS) of superimposed signals in situations where one signal has constant or slowly varying intensities at some consecutive locations and at the corresponding locations the other signal has highly varying intensities. Independent Component Analysis (ICA) is a major technique for Blind Source Separation and the existing ICA algorithms fail to estimate the original intensities in the stated situation. We combine the advantages of existing sparse methods and Kernel ICA in our technique, by proposing wavelet packet based sparse decomposition of signals prior to the application of Kernel ICA. Simulations and experimental results illustrate the effectiveness and accuracy of the proposed approach. The approach is general in the way that it can be tailored and applied to a wide range of BSS problems concerning one-dimensional signals and images (two-dimensional signals).

Keywords : Blind Source Separation, Independent Component Analysis, Sparse Decomposition, Kernel ICA, Wavelet Packet

OCIS codes : (000.5490)Probability theory, stochastic processes, and statistics ; (070.5010)Pattern recognition and feature extraction ; (070.6020) Signal processing; (100.2000)Digital image processing

I. INTRODUCTION

In a typical BSS task, N mixtures are observed or received. Each of these available signals (or images), is assumed to be generated by a linear mixture of unknown sources, where the number of sources is usually assumed to be known to the observer. Thus, an N -dimensional vector of observed signals is generated by the product of an unknown $N \times M$ mixing matrix A and an M -dimensional vector of unknown source signals. The task is to estimate the mixing matrix and then recover the source signals. The Independent Component Analysis (ICA) model in matrix notation is given by:

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

Where X consists of mixed signals x_1, x_2, \dots, x_M and S consists of source signals s_1, s_2, \dots, s_N represented as row vectors and A is the mixing matrix. Under the as-

sumption that the sources are statistically independent, it is possible to recover sources by estimating the mixing matrix $A' \approx A$ and estimating the sources by its inversion:

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

It has been demonstrated by Zibulevsky et al [5], [10], that the assumption of sparseness is very powerful and can significantly improve the accuracy and the computational efficiency of Blind Source Separation methods. Sparsity means that only a small fraction of coefficients differ significantly from zero. Such a signal has a probability density function or distribution of values with a sharp peak at zero and fat tails. A standard sparse distribution is the Laplacian distribution. The advantage of a sparse signal representation is that the probability of two or more sources being simultaneously active is low. Thus, sparse representations lend themselves to good separability because most of the energy in a basis coefficient at any time instant belongs to a single source.

*Corresponding author: kit@mju.ac.kr

All signals can be sparsely represented, i.e. there exists a linear transformation T such that

$$d_i = Ts_i \quad (3)$$

is sparse. Application of the transformation to the mixtures in (1), due to the linearity of T , yields

$$Tm_i = T(a_{i1}s_1 + a_{i2}s_2) = a_{i1}d_1 + a_{i2}d_2 \quad (4)$$

Thus, the problem at hand is equivalent to separation of linearly mixed sparse sources. Different classes of signals require their specific optimal [10] sparse transformations. However a wide range of transforms can result in sparse transformation that permits good estimation of the mixing matrix. For example, in the case of natural images, the edge distribution is usually sparse. A sparse representation of a signal can often be achieved by a transformation into a Fourier, Gabor or wavelet basis.

Recently a completely new technique of Kernel Independent component analysis, based on novel kernel-based measures of dependence, was introduced by Francis R. Bach [6]. This method uses contrast functions based on canonical correlation in a Reproducing Kernel Hilbert Space (RKHS). The technique, as opposed to other techniques, takes full advantage of nonparametric flexibility of the kernel approaches making it flexible and more demanding computationally than current algorithms, and is also addressing the problem of outliers.

We consider ICA as a restrictive case of BSS where only the even-determined or over-determined case is considered, and the mixing process is instantaneous. Investigations reveal that the problem of "independent component analysis" was actually first proposed and named by Herault and Jutten [1]. Linsker [2] proposed unsupervised learning rules that maximize the average mutual information between the inputs and outputs of an artificial neural network. Bell and Sejnowski [3] developed a BSS algorithm called BS-Infomax, which is similar in spirit to that of Linsker and uses an elegant stochastic gradient learning rule that was proposed by Amari et al. [4]. A two stage separation process based on sparse decomposition was proposed by Michael Zibulevsky and Barak A. Pearlmutter [5]. Francis R. Bach and Michael I. Jordan [6] introduced Kernel ICA. Alexander M. Bronstein and Michael M. Bronstein [7] used the Sparse ICA (SPICA) for removing reflections from images. In SPICA they used the Sparse BS Infomax ICA algorithm [3], [4]. Paul D. O'Grady compared various sparse and non-sparse methods [9].

Sparsity increases accuracy and Kernel ICA tries to handle outliers, near Gaussian data, and nonlinear transformations in sources separation problems. If the original source signals are such that one signal has constant or slowly varying intensities at some consecutive loca-

tions, and at the same locations the other signal has highly varying intensities, then all the blind source separation methods, including the above mentioned methods, fail to recover intensities at these specific locations from their mixtures. In this paper, we propose a two step technique for blind source separation which combines the advantages of both the methods. We first use wavelet packets as sparse decomposition of signals and obtain sparse representation of mixed (observed) signals. Then we apply Kernel ICA to estimate the original source signals, thus naming our method Sparse Kernel ICA. The method is more accurate and robust than all other ICA techniques for separating one-dimensional and two-dimensional signals, and show high resistance to outlier, near Gaussian data, and nonlinearity. In section II we explain our technique and in section III we provide the simulation results using synthetic one-dimensional and two-dimensional signals (images).

II. PROPOSED METHOD

The ability of sparse representations to efficiently represent useful statistical structure in a signal plays a key role in our source separation method. The method consists of two steps, first we obtain the optimal sparse representation of mixtures of signals and then as a second step we apply Kernel ICA to estimate the original signals. The optimality of the sparse transformation [5], [10] is the search for such a transformation that increases the proportion of those points contributed by only one of the sources. Richer representations, that over a wide range of natural signals lend themselves to relatively good sparse representations, such as the WPT, were proposed [5]. In order to obtain sparse representation of signals we use Wavelet Packet Transform (WPT). We apply the Kernel ICA algorithm on the selected nodes or coefficients of Wavelet Packet Transform to estimate the unmixing matrix. The separation results obtained after these two steps have high accuracy even in the presences of outliers. The block diagram representation of the proposed method is shown in Fig. 1.

Wavelet Packet Transform is a generalization of the structure of the Wavelet Transform to a full decomposition. Wavelet transform can decompose a signal

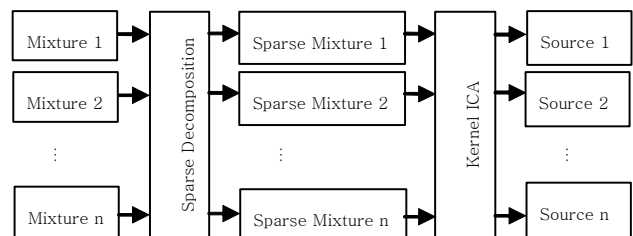


FIG. 1. Block Diagram of the proposed method.

into sub-bands with low frequency and sub-bands with high frequency. It can be seen in Fig. 2 that in wavelet packet decomposition both the approximation (represented by capital A in the figure) and detail (represented by capital D in the figure) at a certain level are further decomposed into the next level, which means the wavelet packet analysis can provide a more precise frequency resolution than the wavelet analysis. The essence of wavelet packet bases selection is to reflect more information with fewer coefficients in order to extract signal features.

We have to find quantitative sparseness criterion for selecting the best nodes. Given a vector y , we need to find such a function $q(y)$, which returns a large value if y is sparse or a small value if y is not sparse. The criterion we choose for the selection of best nodes is so-called L_0 norm, i.e. measure the number of vector coordinates which are higher than some threshold λ :

$$q^{-1}(y) = \frac{1}{n} \sum_{k=1}^n \mathbf{1}(x_k \geq \lambda) \tag{5}$$

where $\mathbf{1}$ is the indicator function and n represents the number of mixtures. A natural choice of the threshold would be $\lambda = \|y - Y\|^2$, where Y is the mean value of y .

After getting the sparse representation of the mixture of signals, we use Kernel ICA [6] based on maximizing independence as minimizing correlation with kernel. Kernel based learning algorithms use the following non-linear mapping [7]:

$$\Phi : \mathfrak{R}^t \rightarrow F \quad x \rightarrow \Phi(x) \tag{6}$$

The data in the input space $x_1, x_2, \dots, x_M \in \mathfrak{R}^t$ is mapped to a potentially much higher dimensional feature space F . The kernel trick just needs to calculate the inner product of two vectors in F with a kernel function $k(\cdot, \cdot)$:

$$\langle \Phi(x_i), \Phi(x_j) \rangle = k(x_i, x_j) \tag{7}$$

Where $\langle \cdot, \cdot \rangle$ denotes an inner product. Kernel ICA assumes RKHS F with $k(x,y)$ and feature map $\Phi(x) = k(\cdot, x)$. Then the F-correlation is defined as the maximal

correlation between the two random variables $f(x)$ and $g(y)$, where f and g range over F :

$$\begin{aligned} \rho_F &= \max_{f,g \in F} \text{corr}(f(x), g(y)) \\ &= \max_{f,g \in F} \frac{\text{cov}(f(x), g(y))}{\sqrt{\{\text{var } f(x)\} \{\text{var } g(y)\}}} \end{aligned} \tag{8}$$

This means that $\rho_F = 0$ implies x and y are independent. Bach [6] defined contrast function as:

$$I_{\rho_F} = -\frac{1}{2} \log(1 - \rho_F) \tag{9}$$

This function is always non-negative and equal to zero if and only if the variables x and y are independent. In order to obtain a computationally tractable implementation of F -correlation, the reproducing property of RKHS is used to estimate the F -correlation.

For the selection of sparse transformation we follow the approach of P. Kisilev and M. Zibulevsky [10]. In the second step we apply Kernel ICA algorithm for blind separation thus naming our approach as Sparse Kernel ICA. The accuracy and effectiveness of our method are presented in the form of simulation results in the following section. The combination of WP based sparse transformation and Kernel ICA incredibly reduces the number of errors in the separation process.

III. RESULTS

We perform our experiments for one-dimensional and two-dimensional signals. In both cases, the selection of signals is such that they match the stated problem i.e. on some successive locations one signal has constant or slowly varying changes and at the same locations the other signal has highly varying values. For all of our experiments, we have used Wavelet Packet Transform with symlets wavelet (sym4).

3.1 Simulations for One-dimensional signals

We generate two uniformly distributed synthetic random signals, each of one hundred data points. Note the data points from 20 to 40; the first signal has constant value while the second signal has highly varying values. To obtain two linear mixtures of these source signals we created mixing matrix of size 2×2 having random values. These random values range from -1 to 1. Then we multiply this mixing matrix with a vector having the two random source signals and obtain two linearly mixed signals. The source signals and their random mixtures are shown in Fig. 3. After getting the mixtures we find their sparse representation using one-dimensional WPT and apply Kernel ICA on these WP coefficients. The estimated separation results by Kernel ICA [6],

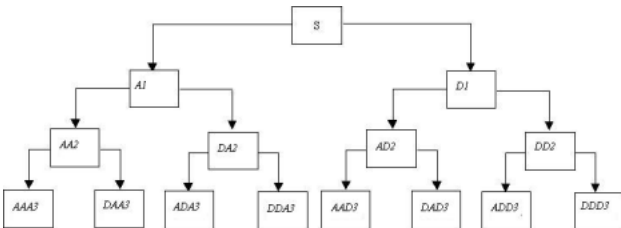


FIG. 2. 3-Level Wavelet Packet Decomposition.

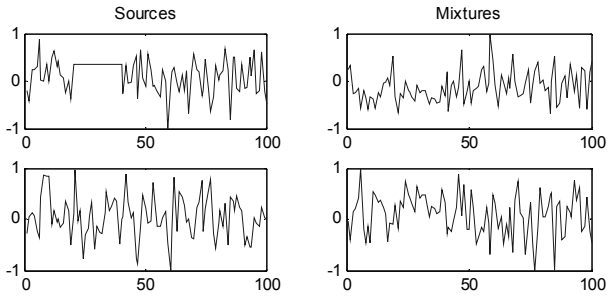


FIG. 3. Uniformly distributed two random signals (Left), two synthetic linear mixtures (Right).

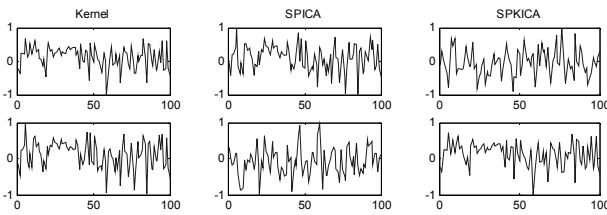


FIG. 4. Estimated sources by Kernel ICA, SPICA, and Sparse Kernel ICA (SPKICA).

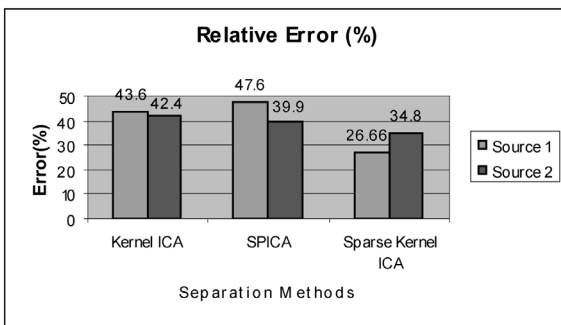


FIG. 5. Percent relative error in estimation of sources by Kernel ICA, SPICA, and Sparse Kernel ICA.

Sparse ICA (SPICA) [5], [8], and our Sparse Kernel ICA are presented in Fig. 4.

In order to find the errors in separation results, we normalized the original S_j and the estimated sources \tilde{S}_j by bringing their norm to one. The error was computed as

$$Error = \frac{\|\tilde{S}_j - S_j\|^2}{\|S_j\|^2} \cdot 100\% \quad (10)$$

Fig. 5 shows errors in estimation of source signals by the three methods. Sparse Kernel ICA recovered the original sources with minimum errors as compared to SPICA and Kernel ICA.

3.2 Simulations for Two-dimensional signals

We generated two synthetic linear mixtures of source images by using random values for coefficients of the

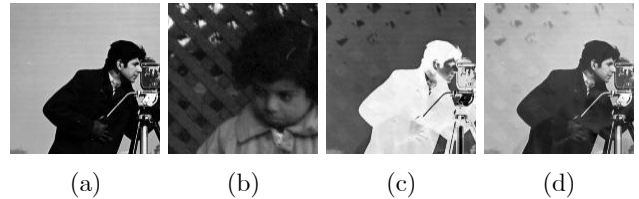


FIG. 6. (a) Source 1. (b) Source 2. (c) and (d) Synthetic mixtures.

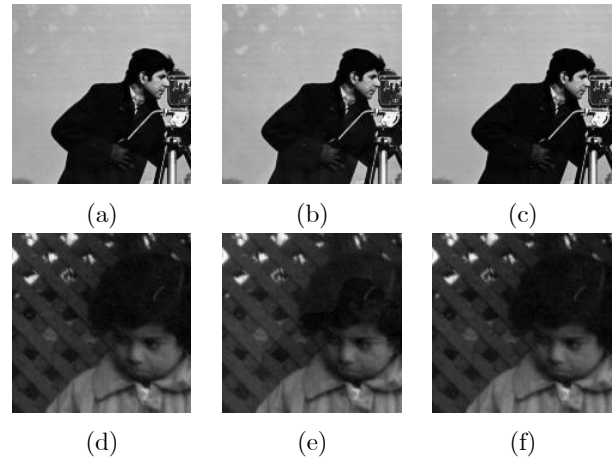


FIG. 7. (a) and (b) Separation results by SPICA. (c) and (d) Separation results by Kernel ICA. (e) and (f) Separation results by Sparse Kernel ICA.

mixing matrix. These sources and their mixtures are shown in Fig. 6. The source images were selected such that the top left corners of the images match with the problem addressed in this paper. Notice the top left corners of both the images, one has uniform intensities and the other has abrupt changes thus making the separation process difficult for SPICA and Kernel ICA. The results of applying SPICA [5], [8], Kernel ICA [6] and our proposed method are given in Fig. 7. Comparing the top left corners in (a), (b) and (c) of Fig. 7, the improvement by Sparse Kernel ICA is evident. SPICA estimated the two sources with 6.19% and 0.37% errors, Kernel ICA estimated with 9.35% and 2.83% errors, and our proposed method estimated them with 1.70% and 0.40% errors respectively. Total number of errors generated by our method is less than that of other two methods. Fig. 8 shows the total number of errors in estimated sources by each method.

For representing the effectiveness and accuracy of our proposed method we also compare Signal to Noise Ratio (SNR) of original sources and the difference of original sources and estimated sources. For the same source images and hundred different mixtures the average SNR values of SPICA, Kernel ICA and Sparse Kernel ICA are given in Table 1. For both of the estimated sources the SNR of Sparse Kernel ICA is better than that of other methods.

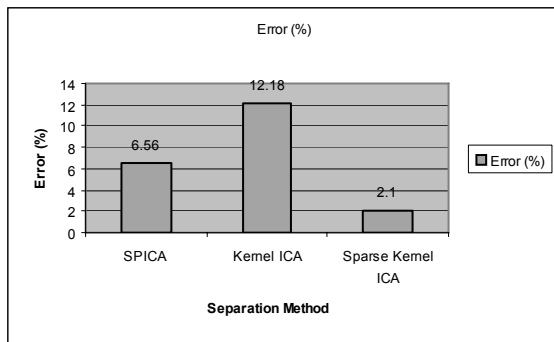


FIG. 8. Percent relative error of separation of mixed sources recovered by (1) SPICA, (2) Kernel ICA, (3) Sparse Kernel ICA.

TABLE. 1. Average Signal to Noise Ratio.

	SPICA	Kernel ICA	Sparse Kernel ICA
Source 1	23.30	19.90	31.12
Source 2	51.01	29.51	56.05

IV. CONCLUSION

We have presented a Blind Source Separation method for separation of linear mixtures of one-dimensional and two-dimensional signals (images). When the original signals are such that one has constant intensity values or uniformly varying changes in intensities for some successive points and the other signal has abrupt changes at the same points then the existing BSS methods fail to recover intensities at those points. We combine the technique of sparse decomposition and Kernel ICA and get accurate results even in the presence of above stated problem. After sparse decomposition using wavelet packets we have a small number of coefficients to work on for estimating an unmixing matrix, thus reducing the computational time of our method compared to that of Kernel ICA. In the paper we have presented experiments with mixtures of two signals but this can work with

mixtures of any number of signals provided the number of mixtures and number of original signals are known. The method is more appropriate for two-dimensional signals; because the problem of constant or slowly varying intensities arises two times, once in rows and then in columns while this phenomenon arises once in one-dimensional signals.

REFERENCES

- [1] J. Herault and C. Jutten, "Space or time adaptive signal processing by neural models," *In Proc AIP Conf on Neural Networks for Computing, American Institute of Physics*, pp. 206–211, 1986.
- [2] R. Linsker, "An application of the principle of maximum information preservation to linear systems," *Advances in neural information processing systems, Morgan Kaufmann, San Francisco, CA*, pp. 186–194, 1989.
- [3] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput*, vol. 7, pp. 1129–1159, 1995.
- [4] S. Amari, A. Cichocki, and H.H. Yang, "A new learning algorithm for blind signal separation," *Advances in neural information processing systems*, vol. 8, MIT Press, Cambridge, MA, 1996.
- [5] M. Zibulevsky and B.A. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural Comput*, vol. 13, pp. 863–882, 2001.
- [6] F. R. Bach and M. I. Jordan, "Kernel Independent Component Analysis," *Journal of Machine Learning Research*, vol. 3, pp. 1-48, 2002.
- [7] B. Scholkopf and A. Smola, "Learning with Kernels," (Cambridge, Mass, MIT Press, 2002)
- [8] Michael M. Broristeiti, Alexander M. Bronstein, Michael Zibirlevshy, and Yehoshzta Y. Zeevi, "Separation of Reflections via Sparse ICA," *International Journal of Imaging Systems and Technology*, vol. 15, no. 1, pp. 84-91, 2005.
- [9] Paul D. O'Grady, Barak A. Pearlmutter, and Scott T. Rickard, "Survey of Sparse and Non-Sparse Methods in Source Separation," *International Journal of Imaging Systems and Technology*, vol. 15, no. 1, pp. 18-33, 2005.
- [10] P. Kisilev, M. Zibulevsky, and Y.Y. Zeevi. "Multiscale framework for blind source separation," *JMLR*, vol. 4, pp. 1339-1364, 2004.