BLIND SEPARATION METHODS BASED ON CORRELATION FOR SPARSE POSSIBLY-CORRELATED IMAGES

Ines Meganem, Yannick Deville
LATT, Université de Toulouse, CNRS, 14 Av. E. Belin, 31400 Toulouse, France
imeganem@ast.obs-mip.fr, ydeville@ast.obs-mip.fr

Matthieu Puigt*
University for Information Technologies
Vanco Nikoleski bb, 6000 Ohrid, Macedonia
matthieu.puigt@uist.edu.mk

ABSTRACT
In this paper, we propose Blind Source Separation (BSS) methods for possibly-correlated images, based on a low sparsity assumption. To satisfy this sparsity condition, one of the versions of our methods applies a wavelet transform to the observed images before performing separation. Another version directly operates in the original spatial domain, when the sources are sparse enough in this domain. Both methods consist in finding, in the considered sparse representation domain, tiny zones where only one source is active. The column of the mixing matrix corresponding to this source is then estimated in this zone. We also propose extensions of these methods, with automated selection of adequate analysis parameters. Various tests show the good performance of these approaches (SIR improvement often higher than 40 dB).

Index Terms— Blind Source Separation, Sparse Component Analysis, image, wavelet, correlation.

1. INTRODUCTION
Blind Source Separation (BSS) methods consist in estimating a set of unknown source signals from a set of observed signals which are mixtures of these sources. There exist three classes of BSS methods [1]: Independent Component Analysis (ICA), Non-negative Matrix Factorization (NMF) and Sparse Component Analysis (SCA). Although the BSS field rapidly evolved these years, image separation has been much less explored than mono-dimensional signal separation. In order to directly apply mono-dimensional methods to two-dimensional signals, images are often rearranged as vectors. The image spatial structure is thus ignored. On the contrary, the approaches proposed in this paper are indeed intended for images, i.e. they take into account their two-dimensional structure. We derive them from the LI-TIFCORR [2] method, which was developed for mono-dimensional signals. Our new methods belong to the class of SCA approaches, but they only require very limited source sparsity. However, images may not be sparse enough in their original spatial representation. We therefore also propose a version of our methods, where we apply a sparsifying wavelet transform before performing separation. The structure of the considered data and the goal of BSS are detailed in the next section.

2. PROBLEM STATEMENT
We assume that we have N observations x_i (images) resulting from mixtures of N source images s_j. We here only consider the case of noiseless linear instantaneous mixtures. In the original spatial representation domain, the observations then read

\[ x_i(n, m) = \sum_{j=1}^{N} a_{ij} s_j(n, m), \quad i = 1 \ldots N, \]  

where (n, m) indexes a pixel. This yields in matrix form

\[ x = A s, \]  

where \( x = [x_1, \ldots, x_N]^T \) contains the observations for pixel location \( (n, m) \), \( A \in \mathbb{R}^{N \times N} \) is the mixing matrix containing the coefficients \( a_{ij} \), and \( s = [s_1, \ldots, s_N]^T \) contains the source values at \( (n, m) \).

BSS then consists in finding an estimate \( \hat{A} \) of \( A \) so as to determine an estimate \( y \) of the sources as follows

\[ y = \hat{A}^{-1} x = \hat{A}^{-1} A s. \]  

This can, however, be achieved only up to two kinds of indeterminacies. They respectively concern the scale factors and order with which the sources are estimated in the output \( y \). Deville and Puigt [2] therefore suggested to rewrite the source restoration problem as follows

\[ y = \hat{B}^{-1} x = \hat{B}^{-1} B s', \]  

with the following notations: \( s'_j = a_{i, \sigma(j)} s_{\sigma(j)} \) and \( b_{ij} = \frac{a_{i, \sigma(j)}}{a_{j, \sigma(j)}} \) with \( i, j = 1 \ldots N \). \( \sigma(.) \) is an arbitrary permutation of the indices \( j \). The \( s'_j \) and \( b_{ij} \) are the source signals and the mixing coefficients, after a permutation and a multiplication by a scale factor. We assume that all mixing coefficients \( a_{ij} \) are non-zero. BSS then consists in estimating \( B \). Our approaches are inspired from the following method.

3. SUMMARY OF LI-TIFCORR
LI-TIFCORR [2] is a SCA method for uncorrelated mono-dimensional signals. Unlike WDO (W-disjoint-orthogonality) methods, such as [3], LI-TIFCORR only needs a weak sparsity assumption: it only requires, for each source, one little zone where this source is isolated (i.e. where only this source has non-zero variance). Separation is then based on correlation parameters. The general structure of the method is as follows (see [2] for more details):

1. The pre-processing stage consists in deriving the Short-Time Fourier Transforms (STFT) of the observations, to sparsify them.
2. The detection stage finds the zones where one source is isolated,

*Most of this work was performed when Matthieu Puigt was with the Laboratoire d’Astrophysique de Toulouse-Tarbes (LATT).

1The following equations apply to a given index \( (n, m) \). We omit it for readability hereafter.
which are called single-source zones.
3. In each single-source zone, the identification stage estimates the
column of the mixing matrix corresponding to the isolated source.
4. The combination stage recovers the sources from the overall
estimate of the mixing matrix and the observations, according to (4).

In the next section, we extend this method to image separation, to
possibly-correlated sources and to more automated operation.

4. PROPOSED EXTENSIONS FOR IMAGE SEPARATION

4.1. Assumptions and definitions

The methods proposed hereafter are based on some assumptions and
definitions, which concern the source images considered in a given
representation domain. This domain can be the original spatial do-
main or a domain obtained after applying a sparsifying transform
(e.g. the wavelet transform). Let $S_i$ be the source images and $X_i$ the
observed images (for $i = 1 \ldots N$) in the chosen representation
domain. We divide this domain into small zones, that we denote
as “analysis zones”. These zones consist of adjacent points.
For the sake of simplicity, we consider square zones. We explore the
representation domain, using adjacent or overlapping zones.

To each analysis zone $\Omega$, we associate the following parameters

- The mean of the signal $X_i$ is
  \[ \bar{X}_i(\Omega) = \frac{1}{M} \sum_{(n,m)\in\Omega} X_i(n,m), \]  
  where $M$ is the size of zone $\Omega$ and $(n,m)$ now refers to one
  signal value in the considered representation domain.

- The covariance of two signals $X_i$ and $X_j$ is
  \[ C_{xixj}(\Omega) = \frac{1}{M} \sum_{(n,m)\in\Omega} [X_i(n,m) - \bar{X}_i(\Omega)] [X_j(n,m) - \bar{X}_j(\Omega)]^*. \]

- Their covariance coefficient is
  \[ c_{xixj}(\Omega) = \frac{C_{xixj}(\Omega)}{\sqrt{C_{xixi}(\Omega)C_{xjxj}(\Omega)}}. \]

- For all $i = 1 \ldots N$, we define $X_i(\Omega)$ and $S_i(\Omega)$ as the vectors
  of centered values of $X_i$ and $S_i$ respectively. This means
  that $X_i(\Omega)$ contains the values $X_i(n,m) - \bar{X}_i(\Omega)$, with
  $(n,m) \in \Omega$, and the same holds for $S_i(\Omega)$.

Definition 1 A source is said to be “isolated” in an analysis zone of
the representation domain if only this source has a non-zero variance
in this analysis zone. This zone is then called a single-source zone.

Definition 2 A source is said to be “accessible” in the representation
domain if there exist at least one analysis zone where it is
isolated.

Assumption 1 i) each source is accessible in the representation
domain and ii) there exist no analysis zones where all vectors $S_i(\Omega)$
are zero.

The images must therefore satisfy a sparsity assumption. If this
condition is not met in the original spatial domain, a sparsifying trans-
form is used to meet it, as detailed below. This transform plays the
same role as the TF transform for mono-dimensional signals (espe-
cially audio signals) used in the LI-TIFCORR method.

Assumption 2 Over each analysis zone $\Omega$ of the representation do-
main, the non-zero vectors $S_i(\Omega)$ are linearly independent (if there
exist at least two such vectors in this zone).

We therefore allow correlated sources. We thus extend the appli-
cability of the proposed methods beyond the case of non-correlated
sources which was considered in [2] for LI-TIFCORR.

4.2. First proposed methods: non-automated versions

The 2D methods that we propose follow the general structure of
LI-TICORR provided in Section 3. However, we split their detailed
description in more stages hereafter, for the sake of clarity.

1. The pre-processing stage consists in possibly applying a suited
sparsifying transform to the observations. If the source images
are sparse enough in the original spatial representation domain,
this pre-processing stage is not used (our resulting SCA method is
called SpaceCorr-2D). Otherwise, we propose to use 2D wavelets
in this sparsifying stage (our corresponding SCA method is called
WaveCorr-2D). We selected the wavelet transform because it is well
known for its sparsifying properties: it concentrates the signal en-
ergy in a few coefficients, which is the reason why it is heavily used
in many image processing applications and essentially in compres-
sion. In Section 5, we compare separation results obtained with and
without this wavelet transform.

2. The detection stage consists in finding the single-source analysis
zones, using the following property :

Property 1 A necessary and sufficient condition for a source to be
isolated in a zone $\Omega$ is
\[ |c_{xixi}(\Omega)| = 1, \quad \forall i, 2 \leq i \leq N. \]  

Proof of Property 1 : Using a linear sparsifying transform, (1) yields for each analysis
zone
\[ X_i(\Omega) = \sum_{j=1}^{N} a_{ij} S_j(\Omega), \quad i = 1 \ldots N. \]

The covariance coefficients defined in (7) can be expressed as
\[ c_{xixi}(\Omega) = \frac{\langle X_i(\Omega), X_i(\Omega) \rangle}{\|X_i(\Omega)\| \|\hat{X}_i(\Omega)\|}. \]

where the notations $\langle \ldots \rangle$ and $\|\|\|$ stand for the inner product
and vector norm. Applying the Cauchy-Schwarz inequality to (10)
then shows that
\[ |c_{xixi}(\Omega)| \leq 1, \quad \forall i, i = 1 \ldots N, \]

with equality if and only if $X_i(\Omega)$ and $\hat{X}_i(\Omega)$ are linearly
dependent. We are now going to demonstrate Property 1 in two steps:
we assumed above that at least one vector $S_j(\Omega)$ is non-zero:
i) If only one source, with index $j$, is active in a considered analysis
zone (i.e. if only $S_j(\Omega)$ is not equal to zero), then, since all mixing
coefficients $a_{ij}$ are assumed to be non-zero, (9) shows that all vec-
tors $X_i(\Omega)$, with $i = 1 \ldots N$, are non-zero and colinear. Therefore,
equality holds whatever $i$ in (11) and the detection condition (8)
was fulfilled.

ii) We now suppose that at least two vectors $S_j(\Omega)$ are non-zero. If
$X_i(\Omega)$ and $\hat{X}_i(\Omega)$ were linearly dependent for all $i$, with $i =
2 \ldots N$, then, due to Assumption 2, all the columns of the mixing
coefficients $a_{ij}$ are assumed to be non-zero, (9) shows that all vec-
tors $X_i(\Omega)$, with $i = 1 \ldots N$, are non-zero and colinear. Therefore,
equality holds whatever $i$ in (11) and the detection condition (8)
was fulfilled.

2To demonstrate that $A \iff B$, we demonstrate i) that $A \implies B$ and ii)
that not($A$) \implies not($B$).
matrix $A$, with indices equal to the indices $j$ of the non-zero vectors $S_j(\Omega)$, would be colinear. This is not true since $A$ is assumed to be invertible. Therefore, in this case, at least one pair of vectors $(X_1(\Omega), X_i(\Omega))$ does not consist of linearly dependent vectors, so that $|c_{x_1x_i}(\Omega)| < 1$ and condition (8) is not fulfilled. This completes the proof of Property 1.

We use this property as follows. For each analysis zone we compute the mean $|c_{x_1x_i}(\Omega)|$ of $|c_{x_1x_i}(\Omega)|$ over all $i$, with $i = 2, \ldots, N$. We consider that the “best” single-source zones are the zones where the mean $|c_{x_1x_i}(\Omega)|$ is the highest.

3. The estimation stage consists in computing estimates of columns of $B$, using correlation parameters. Every single-source zone yields an estimate of a column of the matrix, using the left-hand term of the following formula, which holds when source $k$ is isolated

$$\frac{C_{x_1x_k}(\Omega)}{C_{x_1x_1}(\Omega)} = \frac{a_{ik}}{a_{1k}} = b_{im}, \quad i = 2 \ldots N. \quad (12)$$

with $m \neq k$ due to permutation. We may thus get several estimated columns of $b_{im}$ (with $b_{im} = 1$) for each actual column of $B$.

4. The identification stage consists in selecting the $N$ columns of matrix $B$ among all the columns computed in the previous stage. To this end, we order the single-source zones according to decreasing values of $|c_{x_1x_i}(\Omega)|$, i.e. decreasing quality. The first selected column is given by the first zone in this ordered list (i.e. the best zone). We then move along this list and keep a new column if it is sufficiently distant from the previous ones (i.e. if this distance is above a user-defined threshold).

5. The combination stage consists in multiplying the observations by the inverse of the mixing matrix to recover the source estimates, according to (4).

4.3. Further extensions: automated methods

The above methods include some user-defined parameters, such as the size of analysis zones and their percentage of overlap. This degrades their robustness and efficiency because, as shown in Section 5, their performance depends heavily on the values of these parameters and the best choice varies with the type and size of images.

We here propose automated methods (with or without sparsifying pre-processing) that tackle this problem in the case when the sources are assumed to be uncorrelated: in that case, if the outputs of our above methods yield well-separated sources, they are uncorrelated. On the contrary, if they are poorly separated, they are quite correlated, except in very specific cases (residual orthogonal mixing). Considering the case $N = 2$ hereafter for simplicity, the proposed automated methods therefore consist in checking separation performance by testing whether the covariance coefficient of the recovered sources is high. More precisely, they operate as follows:

- We run our above separation method (SpaceCorr-2D or WaveCorr-2D) for different analysis zone sizes and different percentages of overlap between analysis zones.
- For each run, we compute the covariance coefficient of the two separated images and we store the estimated columns of the mixing matrix.
- We then sort these covariance coefficients in increasing order. We only keep the estimated columns corresponding to the first $L$ coefficients (i.e. presumably the best separations).
- We apply clustering to these best columns. This clustering aims to regroup estimates corresponding to the same column of the mixing matrix. Indeed, sources are recovered up to a permutation indeterminacy, so columns can be estimated in (unknown) different orders from one run to another.
- For every cluster, we then choose a representative column which is kept as our final estimate. This works as follows:
  i) if the standard deviation of the cluster is low, we compute the median of the columns of the cluster,
  ii) if the standard deviation is high, we keep the column associated to the best single-source zone according to our criterion based on correlation (this yielded better test results).

This automated method gave really encouraging results as shown in the next section (although covariance coefficient values do not enable us to distinguish between the runs which yield very good separation).

5. EXPERIMENTAL RESULTS

We now present the performance of our methods obtained for image separation in tests performed with the following 6 couples of images (size: 512x512 pixels): (1) Lenna & Peppers, (2) Lenna & Tiffany, (3) aerial images, (4) texture images, (5) aerial images, (6) Lenna & Elaine.

The performance achieved in each test is measured by the overall SIRI (Signal-to-Interference-Ratio Improvement) achieved by our BSS method. SIRI is defined as the ratio of the output and input SIRs of our system [2]. We provide SIRI values in dB. We use a symmetrical mixing matrix defined as

$$A = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}.$$

5.1. Influence of analysis zone size and overlap

We here present results for the non-automated version of WaveCorr-2D, applied to couple (1), for different sizes of analysis zones and different overlaps (for an overlap of 0.75 for example, zone $i + 1$ overlies 75% of each side of zone $i$). Table 1 shows the influence of both analysis zone parameters on the performance of our BSS method. In all the tables hereafter, "fail" means that the separation process fails to find an estimate for at least one source (because it does not find a zone where this source is isolated). Conversely, when it succeeds in finding them, separation quality may be poor, however: separation results are visually acceptable if SIRI is higher than almost 10 dB (Figure 1 shows an example of good separation).

Table 1: SIRI for various analysis zone parameters for couple (1).

<table>
<thead>
<tr>
<th>overlap</th>
<th>size of the analysis zones</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>X 40.9 56 38 36.5 fail</td>
</tr>
<tr>
<td>0.5</td>
<td>X 40.9 56 38 36.5 fail</td>
</tr>
<tr>
<td>0.75</td>
<td>X 37.8 45.3 43.5 36.9 fail</td>
</tr>
</tbody>
</table>

Table 1 shows that performance depends on the analysis zone parameters. These remarks are true for any representation domain. This shows the attractiveness of our automated methods, considered further in this section. The non-automated version already yields SIRIs higher than about 30 or even 40 dB in many cases. The runs performed with 2x2 pixel analysis zones yield very non-homogeneous results in Table 1. One could consider that it is not
safe to use so small zones, because each correlation parameter in (12) cannot be reliably estimated with so few points. Note, however, that the ratio of these parameters which defines a mixing matrix column in (12) is obtained with very high accuracy in any zone where a source is totally isolated, even when using one or very few points. But this accuracy then strongly decreases if this zone contains even a very small component from another source. For these reasons, we do not use little zones in Section 5.2. We also avoid very large zones (32x32 and 64x64): they do not yield good results because it is harder to find large single-source zones.

5.2. Influence of the representation domain

Tables 2 and 3 show separation performance of our non-automated methods without pre-processing (SpaceCorr-2D) and with pre-processing by wavelet transform (WaveCorr-2D). For each couple of images, we first provide the mean SIRI over 8x8 and 16x16 pixel analysis zones and over the same overlaps as in Table 1. We then provide the standard deviation, minimum and maximum of SIRI, and the success rate, which is equal to the percentage of cases when separation did not fail. These tables show that WaveCorr-2D here yields higher mean performance, except for couple (4). This couple of images is very hard to separate because of texture.

5.3. Performance of automated methods

We here present the performance of our automated method based on WaveCorr-2D. In Table 4, we compare for each couple of source images:
(a) the statistics (mean, std, min and max) of the SIRI obtained by running our non-automated WaveCorr-2D method separately for all cases (size and overlap of analysis zones) considered in Table 1.
(b) The results for a single run of our automated method, which automatically "gathers" all above cases for analysis zones (with $L = 6$). Table 4 shows that our automated method yields more homogeneous, and thus safer, results than the non-automated version. Its SIRI is at least higher than 30 dB and almost always higher than 40 dB. We here keep the small analysis zones (2x2 and 4x4) to take advantage of the very good results they can yield in some cases. We also keep large zones because, in a real application with various kinds of images, we cannot easily define in advance which zone sizes are too large.

6. CONCLUSION

In this paper, we proposed methods based on a sparsity assumption for separating images. In a version of these methods, we apply the wavelet transform to sparsify images. We also proposed automated methods that are more robust because they avoid dependency of performance on some parameters. This was confirmed by tests, which yield SIRI often higher than 40 dB. Several of our methods apply to correlated images, and we plan to test them with such images.

7. REFERENCES