CCIT REPORT #553
August 2005

Efficient Blind Separation of Convolutive Image Mixtures

Sarit Shwartz, Yoav Y. Schechner and Michael Zibulevsky
Department of Electrical Engineering
Technion - Israel Institute of Technology.
Haifa 32000, Israel
`psarit@tx.technion.ac.il` , `yoav@ee.technion.ac.il`,
`mzib@ee.technion.ac.il`

Abstract

Convolutive mixtures of images are common in photography of semi-reflections. They also occur in microscopy and tomography. Their formation process involves focusing on an object layer, over which defocused layers are superimposed. It is possible to remove the mixture crosstalk (separate the layers) based on several images taken with different focus settings, if the defocus point spread function (PSF) is known. However, in general the PSF is unknown, hence there is need for a blind source separation (BSS) approach. We deal with cases in which the layers are independent, hence we seek to minimize the mutual information (MI) of the blindly estimated layers. However, direct MI optimization of convolutive image mixtures is very complex and suffers from local minima. In this paper we devise an efficient approach to solve these problems. Our method is convex, hence having a unique solution which is derived fast. It combines a parametric model of the PSF with factorization of convolution to multiplicative mixtures. Convolutive mixtures are factored into several small and simple problems by short time Fourier transform, where signal separation is efficiently done in each frequency channel by standard BSS tools. These tools, however, suffer from fundamental ambiguities. We overcome these ambiguities by exploiting the PSF parametric model. We apply our algorithm to semi-reflections, and demonstrate its performance in experiments.

Keywords: Blind source separation, Transparent layers, Independent Component Analysis, Sparsity.

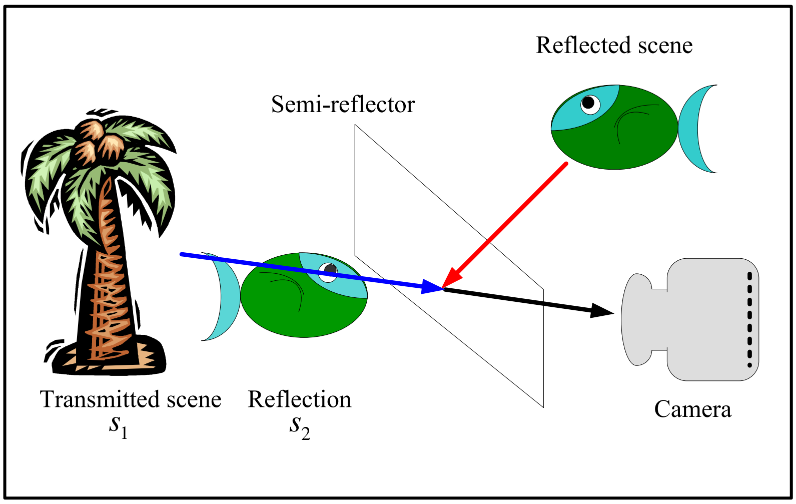


Figure 1: An image of a mixture of two layers, created by a semi-reflector.

1 Introduction

Convolutive image mixtures occur in various imaging modes, such as tomography, microscopy and photography. As an example, consider the setup depicted in Fig. 1. A photograph of a scene is taken through a glass window. The light that reaches the camera contains contributions from a transmitted scene, as well as from reflected objects. In general, the transmitted object and the reflected object are at different distances from the camera. Thus, if the camera is focused on the transmitted scene, the reflected object is defocus blurred (See Fig. 2), and vice versa.

Our goal is to decompose the mixed and blurred images and separate the transmitted scene from the reflected scene. This task is a simple problem if the blur kernel is known. However, typically the blur is unknown and therefore we need to perform blind source separation (BSS). In order to solve the BSS problem we need to use some prior. Typically, in case of semi-reflections, the objects are rather independent. A natural criterion for statistical dependency is mutual information (MI). Thus, MI is commonly used in BSS algorithms (see for example in Refs. [4, 15, 30, 37, 38] and references therein). Therefore, separation of such convolutive mixtures can be achieved by minimizing the MI of the estimated objects. Ref. [37] attempted this approach for semi-reflections. However, MI minimization was performed by exhaustive search. That optimization technique is too computationally expensive. Moreover, MI is not a convex function of the optimized parameters. Hence, there is no guaranty for convergence to a global minimum in the method presented in [37].

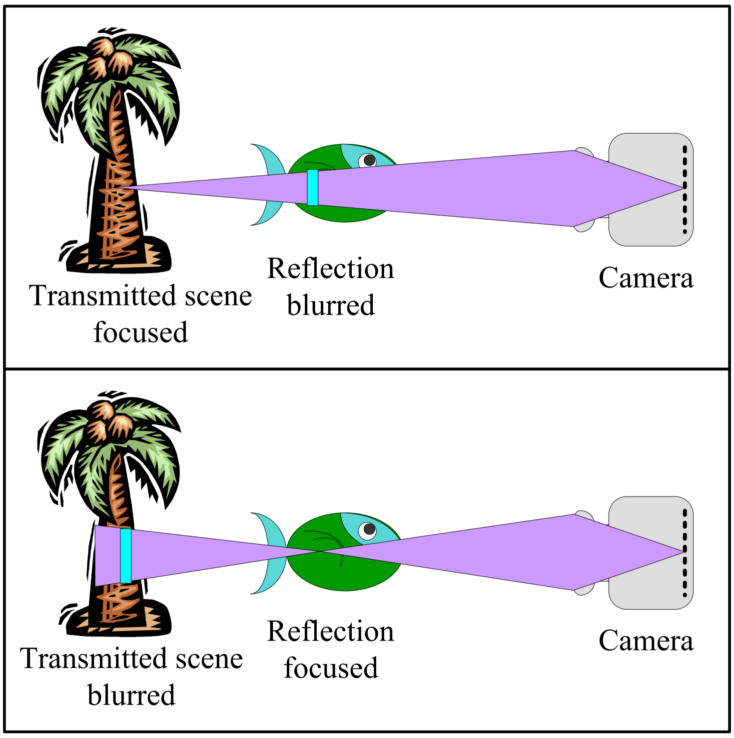


Figure 2: Image acquisition of the scene depicted in Fig 1: [Top] The camera is focused on the transmitted object while the reflection is blurred. [Bottom] The camera is focused on the reflection while the transmitted object is blurred.

To overcome the problems associated with convolutive source separation, frequency methods were introduced [9, 20, 26, 28, 39, 47] in the acoustic signals domain. The complexity is reduced by factorization of the large optimization problem into several small independent optimization problems. The convolutive problem is factored to several pointwise problems by applying a short-time-Fourier transform (STFT). Then, source separation is done for each of the STFT channels, by standard BSS tools. These algorithms inspired our work. However, it is important to note that these tools suffer from fundamental ambiguities, which may ruin the overall result, as we explain in this paper. It is worth mentioning that the STFT-based method had been attempted in images [18]. Ref. [18] used higher order cumulants as the separation criterion in each frequency channel, similarly to existing acoustic source separation algorithms. However, such criteria are not good for images and thus yield poor separation.

In this paper we devise an efficient approach to solve these problems. Our method is convex, hence having a unique solution which is derived fast. We use the factorization principle described above, but overcome its associated problems. This is achieved by exploiting a parametric model for the blur. The algorithm was applied and demonstrated in experiments of semi-reflections.

2 Related Studies

There are two branches of literature related to this work. One of them deals with separation of convolved and mixed signals in general (not necessarily images). The second is dedicated to the separation of semi-reflections. In this section, we discuss the related studies that were not mentioned in Sec. 1.

MI is a statistical measure. It requires the estimation of both marginal and joint entropies of the estimated sources. However, estimation of the joint entropy can be inaccurate, as detailed in Sec. 3.2. This can lead to poor source separation. For this reason, many existing BSS algorithms use alternative independence criteria, such as higher order cumulants [27, 32, 34, 45, 50, 53, 54, 55], which do not require entropy estimation. These criteria sometimes fail, as shown in [4, 43, 44]. Other algorithms assume the sources are i.i.d., or equivalently use *MI rate* [29, 49] as the statistical dependency criterion. This eliminates the need to estimate joint entropies. However, typically, the i.i.d. assumption is invalid. Thus, the estimated sources are *whitened.* This whitening effect corrupts the data severely both in acoustic and in imaging applications. An additional approach, related to MI, comprises the maximum likelihood /maximum *a-posteriori* (ML/MAP) algorithms [12, 21, 33]. ML/MAP optimization does not address the separation quality of the sources directly. In addition, ML/MAP optimization suffers from both high dimensionality and complexity: the number of optimization variables can reach millions.

As for semi-reflections, several algorithms have been developed to handle them. Some algorithms use motion cues [10, 11, 16, 17, 35, 42, 51, 56] or stereo [2, 41, 52] in order to separate the scene components. They require a sequence of images taken from different viewpoints. Another method uses a single image [22, 23]. However, the algorithm in [22] depends heavily on inputs from an oracle, requiring critical human interaction. The algorithm in [23] suffers from high non-linearity of the cost function and from local minima. This requires heuristic optimization, while convergence is not guaranteed.

Some algorithms require neither camera motion nor human interaction. They are based on polarization cues [6, 13, 38]. Nevertheless, there are scenarios in which polarization is ineffective. This occurs when the viewed objects, beside the semi-reflector, are specular, or when the optical axis of the camera is at a small angle relative to the semi-reflector normal. Moreover, polarization information can be distorted if the transparent medium has significant photo-elasticity [36]. This phenomenon is usually more sever in plastic media, and can cause errors in polarization-based scene analysis. Another recent algorithm exploit flash photography [1]. However, it requires dynamic range conditions of flash illumination that are difficult to obtain.

3 Problem Formulation

3.1 Blind Source Separation

Let $\{s_1, s_2\}$ be a set of independent sources. Each source is of the form $s_k = s_k(\mathbf{x})$, where $\mathbf{x} = (x, y)$ is a two dimensional (2D) spatial coordinate vector in the case of images. Let $\{m_1, m_2\}$ be a set of measured signals, each of which is a linear mixture of a convolved versions of the sources

$$m_i(\mathbf{x}) = (a_{i1} * s_1)(\mathbf{x}) + (a_{i2} * s_2)(\mathbf{x}) \quad , i = 1, 2 \; . \tag{1}$$

Here $*$ denotes convolution and $a_{ik}(\mathbf{x})$, $k = 1, 2$, are linear spatially invariant filters. These convolutions constitute a linear operator $\mathbf{A}$ that transforms $\{s_1, s_2\}$ to $\{m_1, m_2\}$.

Denote $\{\hat{s}_1, \hat{s}_2\}$ as the set of the reconstructed sources. Reconstruction is done by applying a linear operator $\mathbf{W}$ on $\{m_1, m_2\}$. Each of the reconstructed sources is of the form

$$\hat{s}_i(\mathbf{x}) = (w_{i1} * m_1)(\mathbf{x}) + (w_{i2} * m_2)(\mathbf{x}) \quad , i = 1, 2 \; , \tag{2}$$

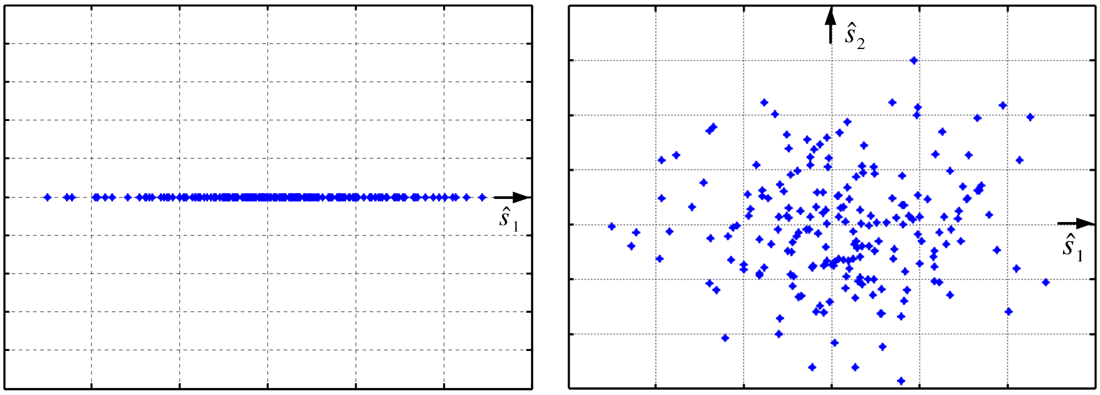
where $w_{ik}(\mathbf{x})$ are linear spatially invariant filters. All of the filter coefficients have continuous values. Thus, the estimated sources $\{\hat{s}_1, \hat{s}_2\}$ can have any continuous value.

Our goal is: given only the measured signals $\{m_1, m_2\}$, find a linear separation operator $\mathbf{W}$ that inverts the mixing process, thereby separating the sources. This separation task is an optimization problem. The separation criterion used is the independence of the estimated sources $\hat{s}_k$. Therefore, the mixing process is inverted by finding $\mathbf{W}$ that minimizes their MI.

3.2 MI of Image Mixtures

Two sources are statistically independent if their joint probability density function (PDF) is equal to the multiplication of their marginal PDFs:

$$p^{\text{indp}}_{\hat{s}_1,\hat{s}_2}(\hat{s}_1, \hat{s}_2) = p_{\hat{s}_1}(\hat{s}_1) p_{\hat{s}_2}(\hat{s}_2) \; . \tag{3}$$



(a) A 1D scatter plot of $\hat{s}_1$ (b) A 2D scatter plot of $\hat{s}_1$ and $\hat{s}_2$

Figure 3: 1D and 2D scatter plots of estimated sources. In the 1D plot the scattered dots are dense, enabling accurate estimation of marginal entropy. In the 2D plot the samples are sparsely scattered. This can lead to unreliable joint entropy estimation.

Thus, in order to assess the level of independence, we need to find the distance between $p^{\text{indp}}_{\hat{s}_1,\hat{s}_2}(\hat{s}_1, \hat{s}_2)$ and the joint PDF $p_{\hat{s}_1,\hat{s}_2}(\hat{s}_1, \hat{s}_2)$. This is the MI of the signals,

$$\mathcal{I}_{\hat{s}_1,\hat{s}_2} = \int\int p(\hat{s}_1, \hat{s}_2) \log \left[\frac{p(\hat{s}_1, \hat{s}_2)}{p(\hat{s}_1)p(\hat{s}_2)} \right] d\hat{s}_1\, d\hat{s}_2 \ . \tag{4}$$

The MI (4) can be rewritten using the marginal entropies $\mathcal{H}_{\hat{s}_k}$ and the joint entropy of the signals $\mathcal{H}_{\hat{s}_1,\hat{s}_2}$ (see for example [8]),

$$\mathcal{I}_{\hat{s}_1,\hat{s}_2} = \mathcal{H}_{\hat{s}_1} + \mathcal{H}_{\hat{s}_2} - \mathcal{H}_{\hat{s}_1,\hat{s}_2} \ . \tag{5}$$

MI is based on estimation of both joint and the marginal entropies of the images (Eq. 5). However, estimation of the joint entropy is problematic, as illustrated in Fig. 3. The figure presents two scatter plots of source samples, each having $N = 200$ samples. Fig. 3(a) is a one dimensional (1D) scatter plot of some source. Note that in this plot the scattered dots are dense, enabling accurate estimation of marginal entropy. Fig. 3(b) presents a scatter plot of one source vs. the other source. Here the scattered dots are sparse, and estimating the joint PDF and entropy from such samples becomes unreliable. Therefore, it is desirable to avoid joint entropy estimations.

Joint entropy estimation can be avoided if the mixtures are pointwise, rather than convolutive. In pointwise mixtures the sources are not blurred. The mixing filters are simple attenuation coefficients, $a_{ik} = a_{ik} * \delta(\mathbf{x})$. Therefore, the mixing process can be expressed as a simple mixing matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \ . \tag{6}$$

In this case, the separation operator $\mathbf{W}$ is a simple matrix, termed the separation matrix

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} . \tag{7}$$

In this case, the joint entropy can be written as (see for example Ref. [15])

$$\mathcal{H}_{\hat{s}_1,\hat{s}_2} = \log|\det(\mathbf{W})| + \mathcal{H}_{m_1,m_2} . \tag{8}$$

Substituting Eq. (8) into Eq. (5) yields

$$\mathcal{I}(\hat{s}_1, \hat{s}_2) = \mathcal{H}_{\hat{s}_1} + \mathcal{H}_{\hat{s}_2} - \log|\det(\mathbf{W})| - \mathcal{H}_{m_1,m_2} . \tag{9}$$

Note that $\mathcal{H}_{m_1,m_2}$ is independent of $\mathbf{W}$ and is constant for a given measurements set $\{m_1, m_2\}$. Therefore, it can be ignored in the MI optimization process. This method bypasses the problem involved in estimating $\mathcal{H}_{\hat{s}_1,\hat{s}_2}$ in the optimization.

It is desirable to do the same for the general convolutive case. However, if $\mathbf{W}$ is a convolutive operator, Eq. (9) does not hold. Moreover, the joint entropy of the reconstructed sources $\mathcal{H}_{\hat{s}_1,\hat{s}_2}$ can be expressed as $\mathcal{H}_{m_1,m_2}$ plus a function of $\mathbf{W}$, if and only if the reconstructed sources are constrained to be i.i.d. (See for example [29]). This assumption is rarely true. Existing algorithms using this assumption, or equivalently using the MI rate [29, 49] as the statistical dependency criterion, suffer from whitening. The separated sources are whitened, corrupting the estimation severely both in acoustic and in imaging applications. To bypass the joint entropy estimation, a different "trick" of factorization is presented and detailed in Sec. 4.

4 Efficient Separation of Convolutive Image Mixtures

Our aim is to use Eq. (9) in convolutive mixtures, despite the fact that it is valid only in pointwise mixtures. This is achieved by factoring the convolutive optimization problem into several smaller ones, which are apparently independent of each other. The inspiration for this approach comes from frequency domain algorithms developed for acoustic signals [9, 20, 26, 28, 39, 47]. A transformation to the images is performed, such that the *convolution is expressed as a set multiple pointwise problems.* This transformation is the STFT. Source separation is done in each of the STFT channels exploiting standard independent component analysis (ICA) tools. This enables high efficiency and a unique solution. Nevertheless, this approach has its own fundamental limitations. These inherent problems are discussed and solved in Secs. 5 and 6.

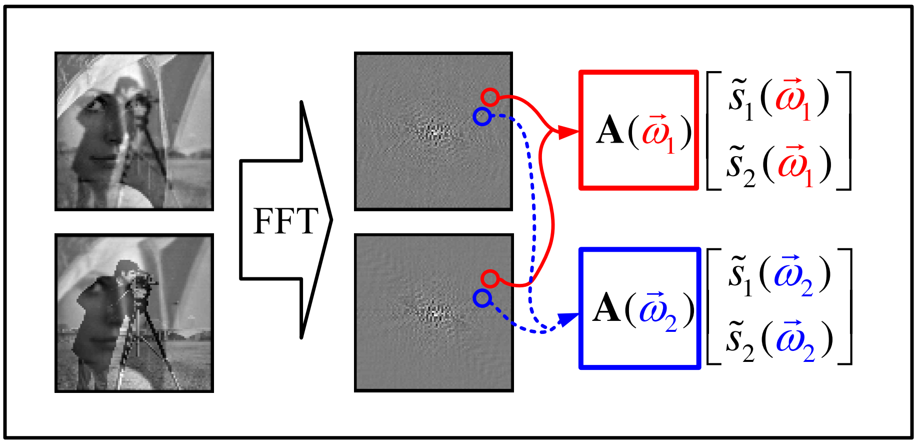


Figure 4: Factorization by Fourier transform of the measured signals. Each pixel in the Fourier domain corresponds to a different frequency. Each frequency has a different mixing matrix. Therefore, different frequency pixels are mixed by different processes.

4.1 Factorization by STFT

Applying the Fourier transform to Eq. (1) yields

$$m_i(\vec{\omega}) = a_{i1}(\vec{\omega})s_1(\vec{\omega}) + a_{i2}(\vec{\omega})s_2(\vec{\omega}), \quad i = 1, 2 \ , \tag{10}$$

where $\vec{\omega} = (\omega_x, \omega_y)$ is the frequency index vector. The convolution in the spatial domain becomes a multiplication in the frequency domain. In each frequency there is a “pointwise” mixture of the sources. Therefore, at each frequency a simple pointwise problem can be solved:

$$\hat{s}_i(\vec{\omega}) = w_{i1}(\vec{\omega})m_1(\vec{\omega}) + w_{i2}(\vec{\omega})m_2(\vec{\omega}), \quad i = 1, 2 \ . \tag{11}$$

Apparently, this is a simple problem that can exploit standard ICA tools, particularly Eq. (9). However, in the frequency domain the Fourier transform yields only a *single sample* per frequency channel, as illustrated in Fig. 4. Thus, no statistical ensemble is available for ICA (ICA is based on statistical measures that are estimated from data ensembles). Therefore, BSS is not possible in this case.

To obtain a data ensemble, rather than a single data point per frequency, *sub-band images* representing raw frames by several wide frequency bands are used. This is achieved by applying a STFT on Eq. (1),

$$m_i(\vec{\omega}, \mathbf{x}) = a_{i1}(\vec{\omega}, \mathbf{x})s_1(\vec{\omega}, \mathbf{x}) + a_{i2}(\vec{\omega}, \mathbf{x})s_2(\vec{\omega}, \mathbf{x}), \quad i = 1, 2 \ , \tag{12}$$

rather than a Fourier transform. Here $\vec{\omega}$ is the frequency index and $\mathbf{x}$ is the spatial index.

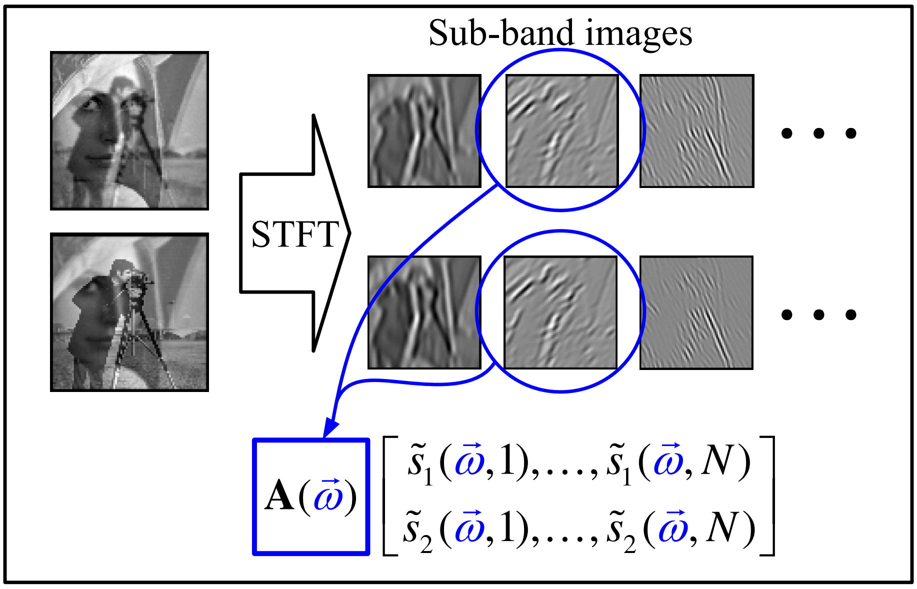


Figure 5: Short Time Fourier Transform of the measured images. Each frequency channel contains a sub-band image of the size of the raw frame. Therefore, per band, the convolutive mixture problem is transformed to a pointwise problem .

If the STFT window size is larger than the effective width of the blur kernel[1] (see for example [28]), then $a_{ik}(\vec{\omega},\mathbf{x}) \approx a_{ik}(\vec{\omega})$ and Eq. (12) becomes

$$m_i(\vec{\omega},\mathbf{x}) \approx a_{i1}(\vec{\omega})s_1(\vec{\omega},\mathbf{x}) + a_{i2}(\vec{\omega})s_2(\vec{\omega},\mathbf{x}), \quad i = 1,2\ . \tag{13}$$

Note that at each frequency channel $\vec{\omega}$, Eq. 13 expresses a pointwise mixture of sub-band images. That is, there are sufficient samples for estimating image statistics. This is illustrated in Fig. 5. At each frequency channel, the mixed sources can be separated by simple ICA optimization. Then, all the separated sources from all the frequency channels may be combined by inverse STFT.

To describe the ICA optimization, denote $\mathbf{W}(\vec{\omega})$ as the separation matrix at channel $\vec{\omega}$. In addition, denote $\mathcal{I}^{\vec{\omega}}(\hat{s}_1,\hat{s}_2)$ and $\mathcal{H}^{\vec{\omega}}_{\hat{s}_i}$ as the MI and marginal entropies of the estimated sources at channel $\vec{\omega}$, respectively. Then, similarly to Eq. (9), the MI of the estimated sources at each channel is given by

$$\mathcal{I}^{\vec{\omega}}(\hat{s}_1,\hat{s}_2) = \mathcal{H}^{\vec{\omega}}_{\hat{s}_1} + \mathcal{H}^{\vec{\omega}}_{\hat{s}_2} - \log|\det[\mathbf{W}(\vec{\omega})]| - \mathcal{H}^{\vec{\omega}}_{m_1,m_2}\ . \tag{14}$$

where $\mathcal{H}^{\vec{\omega}}_{m_1,m_2}$ is the joint entropy of the measured data at channel $\vec{\omega}$. Note that $\mathcal{H}^{\vec{\omega}}_{m_1,m_2}$ is independent of $\mathbf{W}(\vec{\omega})$ and is thus constant for a given measurements set $\{m_1(\vec{\omega},\mathbf{x}), m_2(\vec{\omega},\mathbf{x})\}$.

---

[1] A discussion regarding the STFT window width is given in App. A.

For this reason, it is ignored in the optimization process. Therefore, the optimization problem at each channel is:

$$\min_{\mathbf{w}(\vec{\omega})} \left\{ \left(\sum_{k=1}^{2} \hat{\mathcal{H}}_{\hat{s}_k}^{\vec{\omega}} \right) - \log |\det[\mathbf{W}(\vec{\omega})]| \right\} \ , \tag{15}$$

where $\hat{\mathcal{H}}_{\hat{s}_k}^{\vec{\omega}}$ is an estimator of the channel entropy of an estimated source. This is a significant expression. It means that MI minimization of a convolutive mixture can be performed while bypassing the numerical and complexity problems that stem from direct estimation of joint entropies. Hence the expected output is both more accurate and more efficient to obtain.

Solving the separation problem for each frequency channel independently looks like a simple problem. However, as explained in Sec. 5 this approach has some inherent problems. Therefore, it is recommended *not* to address the separation problem in each frequency channel independently, but rather use inter-channel knowledge transfer.

4.2 Estimation of Channel Entropies

In Sec. 4.1 it was shown that MI minimization can be simplified by factorization. In this section further reduction of computational complexity is achieved by adapting the formulation such that it becomes convex.

At each frequency channel we need to solve the optimization given by Eq. (15). Therefore, we need to estimate the marginal entropies of the sources at each channel. It is known from studies of image statistics (see for example [23, 46]) that sub-band images are *sparse signals*, i.e. most of their pixel values reside near the origin, hence their PDF has a peaky shape, as illustrated in Fig. 6. This PDF model may be exploited in order to estimate the channel entropy efficiently, while maintaining a reasonable accuracy.

A PDF model that is widely used in the literature to model sparse sub-band images is the generalized Laplacian (see for example [46])

$$p[\hat{s}_k(\vec{\omega})] = c(\rho) \exp\left[-|\hat{s}_k(\vec{\omega})|^{\rho}\right] \ , \text{ where } \quad 0 < \rho < 2 \ . \tag{16}$$

Here $c(\rho)$ is the normalization factor of $p[\hat{s}_k(\vec{\omega})]$. The sparsity of the represented signal is determined by the parameter ρ. The smaller ρ is, the narrower the PDF, representing a sparser signal. Examples of PDFs corresponding to different values of ρ are presented in Fig. 7. Note that this PDF model[2] assumes that the source has a unit variance. This

---

[2] The generalized Laplacian model can be extended to include also a scale parameter ϱ, by $p(\hat{s}_k) \sim \exp(-|\hat{s}_k/\varrho|^{\rho})$.

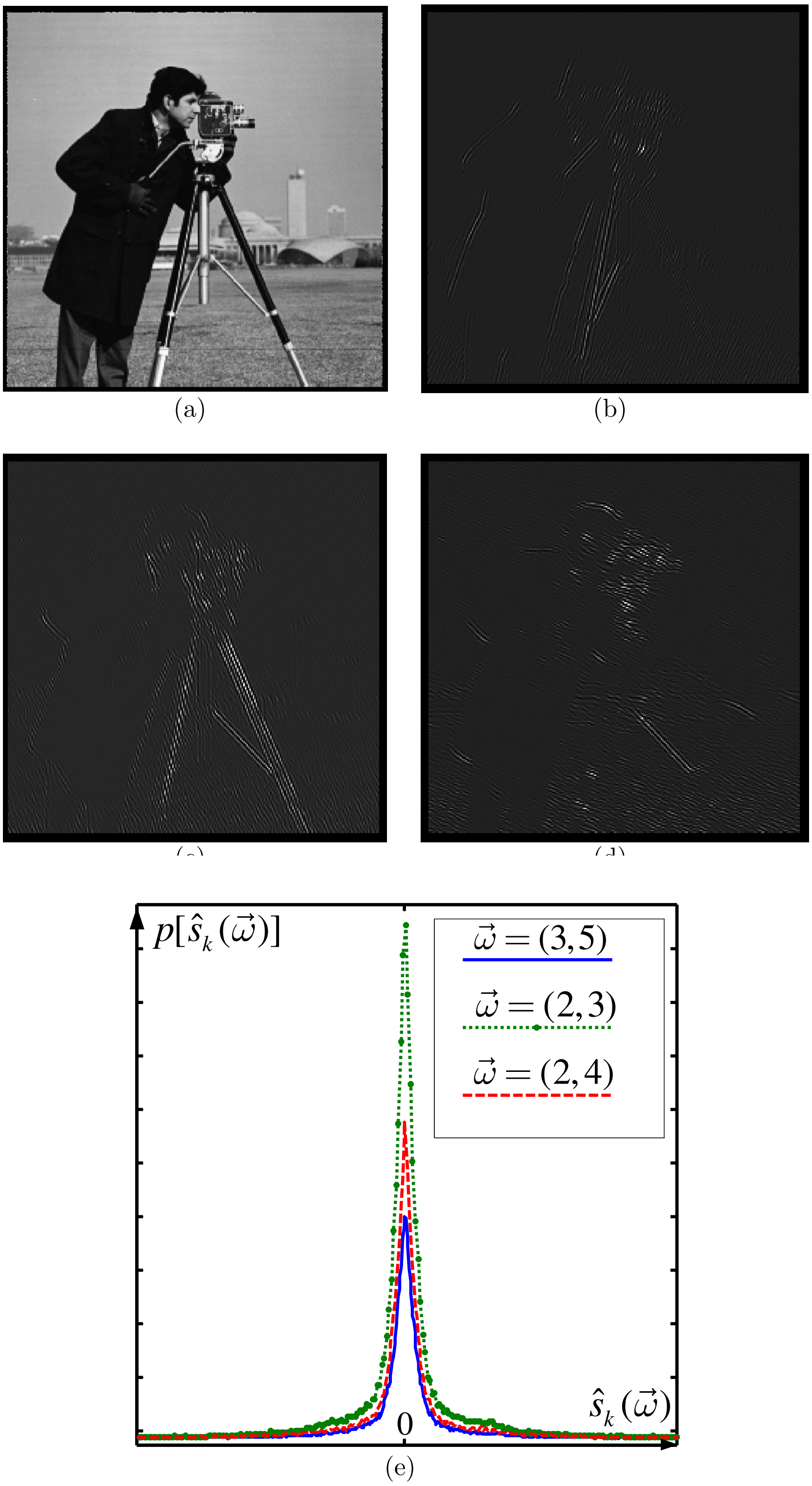

(a) (b) (c) (d) (e)

Figure 6: Sub-band images and their statistics. The STFT window width is 11×11 pixels (a) Cameraman image. (b) STFT channel $\vec{\omega} = [2, 3]$. (c) STFT channel $\vec{\omega} = [4, 4]$. (d) STFT channel $\vec{\omega} = [3, 5]$. (e) Histogram of STFT channels.

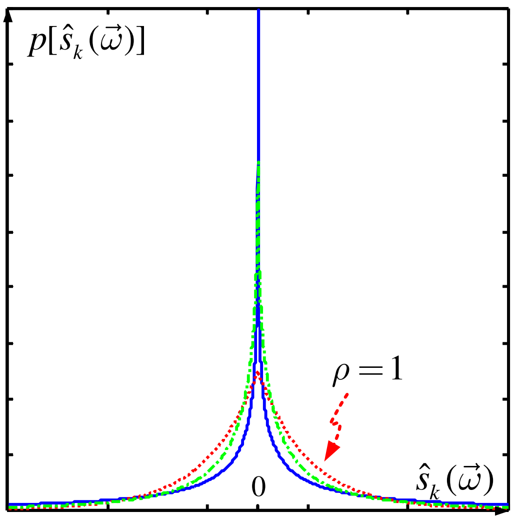


Figure 7: The generalized Laplacian model. Each plot corresponds to a different parameter ρ.

implicit normalization makes the optimization robust to some numerical issues.[3]

We now exploit this prior of image statistics to simplify the estimation of entropies in our optimization. Entropy is defined as (see for example [8])

$$\hat{\mathcal{H}}_{\hat{s}_k}^{\vec{\omega}} = E\left\{\log(p[\hat{s}_k(\vec{\omega})])\right\} \ , \tag{17}$$

where E is the expectation operator. Substituting Eq. (16) into Eq. (17) and replacing the expectation with empirical averaging, the estimated channel entropy is:

$$\hat{\mathcal{H}}_{\hat{s}_k}^{\vec{\omega}} = \frac{1}{N}\sum_{n=1}^{N} |\hat{s}_k(\vec{\omega}, n)|^{\rho} + C(\rho) \ . \tag{18}$$

Here $C(\rho) = \log[c(\rho)]$. Since $C(\rho)$ does not depend on $\hat{s}_k$, it can be ignored in the optimization process. Hence, the generalized Laplacian model yields a very simple expression (18) of entropy as a function of the variables. Moreover, the computational complexity of this entropy estimator is $\mathcal{O}(N)$.

In principle, to achieve the ultimate accuracy we should fit the PDF model (16) to the samples $\hat{s}_k(\vec{\omega})$, to estimate ρ, and then use this ρ in Eq. (18). However, we opt to *set* ρ in a way that enhances computational efficiency. Optimization convergence to a global solution is greatly enhanced if the cost function (MI, in our case) is convex. We can achieve convexity of the channel entropy (Eq. 18) as a function of $\mathbf{W}(\vec{\omega})$ only if $\rho \geq 1$.

---

[3] MI optimization can suffer from numerical instabilities due to a scale ambiguity, as explained in Sec. 5.1.

However, the PDF of sub-band images is typically very sparse, i.e. $\rho < 1$. Therefore, we have a tradeoff between accuracy and convexity: for efficient optimization, we require a convex function, while for accuracy we require $\rho < 1$. The PDF representing the sparsest signal that yields a convex function in Eq. (18) corresponds to $\rho = 1$. Therefore, we choose to use $\rho = 1$.

Substituting Eq. (18) in Eq. (15) yields the following MI minimization formulation:

$$\min_{\mathbf{W}(\vec{\omega})} \mathcal{I}^{\vec{\omega}}_{\hat{s}_1,\hat{s}_2} = \min_{\mathbf{W}(\vec{\omega})} \left\{ \sum_{k=1}^{2} \frac{1}{N} \left[\sum_{n=1}^{N} |\hat{s}_k(\vec{\omega}, n)| \right] - \log|\det[\mathbf{W}(\vec{\omega})]| \right\} . \tag{19}$$

This is the core of our separation method. Note that this is a very simple cost function to calculate and optimize: for each frequency band (sub-band), all is needed is simple summation of the absolute pixel values and a determinant of a small matrix.

Eq. (19) may be solved for each frequency channel independently. Then, all the separated sub-band images can be combined to form the desired output. However, there are several inherent problems to dealing with each frequency channel independently. We discuss these problems in Sec. 5 and solve them in Sec. 6.

5 Inherent Problems

There are several inherent problems in the STFT factorization method. In Sec. 5.1 we discuss ambiguities inherent in MI optimization. These ambiguities can affect the reconstruction of the estimated image from the separated sub-band images. In Sec. 5.2 we discuss problems inherent to BSS. These problems can affect the separation quality in each of the frequency channels.

5.1 Ambiguities of MI Optimization

Optimization of MI possesses three ambiguities: permutation, sign and scale. We now detail these ambiguities and their influence on the separation performance.

5.1.1 Permutation Ambiguity

Minimization of MI results in sources that appear in an arbitrary order, since $\mathcal{I}_{\hat{s}_1,\hat{s}_2} = \mathcal{I}_{\hat{s}_2,\hat{s}_1}$. This permutation ambiguity implies that the separated sub-band images at each channel appear in a random permutation. Hence some sub-band images corresponding to the "first" estimated source may actually belong the "second" estimated source. As illustrated in Fig. 8, when the channels are transformed back to the image domain using the

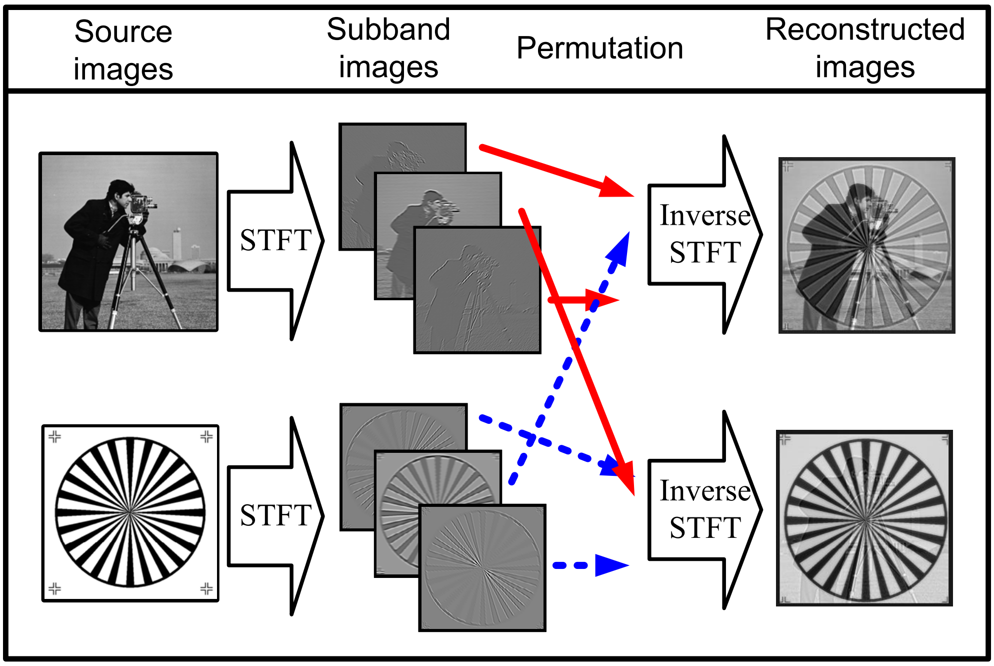


Figure 8: The effect of permutation ambiguity in reconstruction from sub-band images. The reconstructed images exhibit crosstalk, although separation exists in each of the channels.

inverse STFT, the reconstructed images can suffer from crosstalk. Even though source separation was achieved in each channel independently, distinct sub-band images from different sources are combined together in the reconstruction. If this problem is unsolved, factorization leads to an overall poor result.

5.1.2 Scale and Sign Ambiguities

Let $\hat{s}_1, \hat{s}_2$ be two statistically independent sources. Their joint PDF is thus separable:

$$p_{\hat{s}_1,\hat{s}_2}(\hat{s}_1, \hat{s}_2) = p_{\hat{s}_1}(\hat{s}_1) p_{\hat{s}_2}(\hat{s}_2) \; . \tag{20}$$

Therefore, the MI of $\hat{s}_1, \hat{s}_2$ is $\mathcal{I}_{\hat{s}_1,\hat{s}_2} = 0$ (see for example [8]). Denote $\bar{s}_1 = \iota_1 \hat{s}_1$ and $\bar{s}_2 = \iota_2 \hat{s}_2$, where ι_i are arbitrary constants. The joint PDF of $\bar{s}_1, \bar{s}_2$ is still separable and equals

$$p_{\bar{s}_1,\bar{s}_2}(\bar{s}_1, \bar{s}_2) = p_{\bar{s}_1}(\bar{s}_1) p_{\bar{s}_2}(\bar{s}_2) \; . \tag{21}$$

Therefore, the MI of $\bar{s}_1$ and $\bar{s}_2$ is zero as well. This means that minimization of MI results in sources that have an arbitrary intensity scale, per sub-band channel. The sign ambiguity is a special case of the scale ambiguity, for which $\iota = -1$.

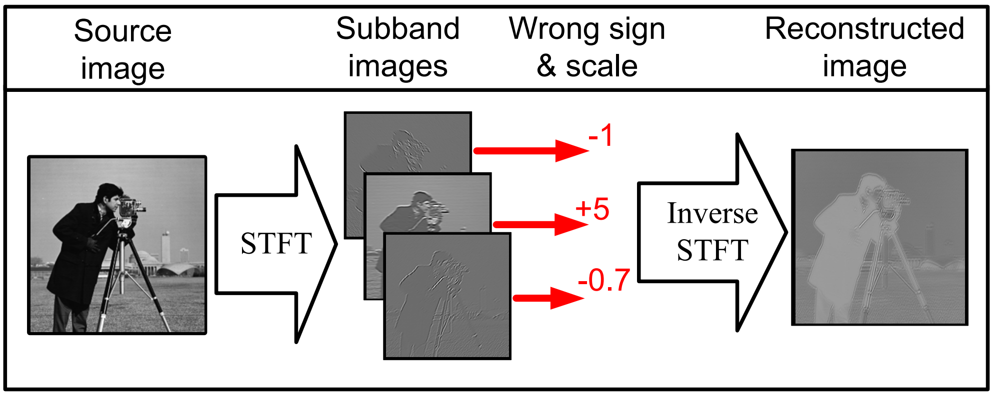


Figure 9: The effect of scale ambiguity in reconstruction from sub-band images. The reconstructed image appears unnatural, due to an unbalanced spectrum and opposite sign in some of the channels.

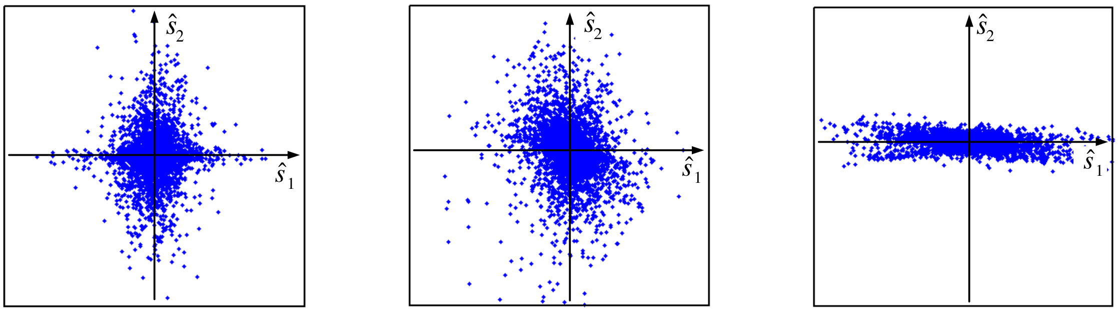
There are two problems that stem from this ambiguity. The first is numerical, and occurs at the channel-level. The second is more fundamental, and appears when all channels are combined. First, note that scale ambiguity implies that there are infinitely many solutions to a separation problem. This ambiguity may destabilize the optimization process. This numerical problem is avoided when using the PDF parameterization described in Sec. 4: the unit variance of the PDF model constrains the resulting sources to have unit energy at each band.

It is important to note, however, that while this method stabilizes the optimization at each channel, the fundamental ambiguity is *not solved*. The true scale of different channels is unknown, leading to imbalance between frequency channels. As illustrated in Fig. 9, when the estimated channels of a source are transformed back to the image domain using the inverse STFT, the reconstructed image can appear unnatural and suffer from artifacts. Hence, the true scale of each channel has to be estimated eventually, in order to make proper reconstruction based on multiple channels.

5.2 Separation Performance in Different Channels

In this section we discuss the separation performance in different channels. We use scatter plots in order to illustrate separation results in different scenarios. In order to make the paper self contained, a discussion about scatter plots in the context of sparse and independent signals is given in App. B.

The performance is frequency dependent, as illustrated in Fig. 10. Typically, there



(a) Good channel. (b) Poor channel. (c) Very poor channel.

Figure 10: Separation results in different channels.

are a few frequency channels with good separation [Fig. 10(a)], a few channels with very poor separation [Fig. 10(c)] and the rest of the channels have mediocre separation quality [Fig. 10(b)]. There are several reasons for this phenomenon. Some stem from fundamental limitations, as discussed in Sec. 5.2.1. Another reason is the dependency of separation optimization on the measured data. We discuss this problem in Sec. 5.2.2.

5.2.1 Fundamental Limitations

There are several special channels for which the separation performance is expected to be poor. Let us focus on the low frequency channels. Separation in them is poor for two reasons. First, separation of convolutive image mixtures in the low frequency bands is inherently ill-conditioned, *even if the blur kernel is perfectly known* (see for example Refs. [24, 37]). This limitation is due to the finite aperture of the imaging system. Even the ideal BSS algorithm cannot achieve better results than those obtained when the blur operator is known. Hence, low spatial frequencies are not well separated and reconstructed.

In BSS, problems in low frequencies may be compounded by mutual dependence of sources. For example, in natural scenes the illumination tends to come from the top. Therefore, source scenes tend to be brighter at the top and darker at the bottom. These lighting similarities cause inter-image dependencies in low frequency channels. These dependencies violate the independence assumption upon which BSS depends.

Problems can also occur when one of the sources in the scene has no energy in a certain channel. For example, without the loss of generality, assume that $s_2(\vec{\omega}^*, \mathbf{x}) = 0, \forall \mathbf{x}$ in some frequency $\vec{\omega}^*$. The set of equations that describe the acquired sub-band images

$$\begin{aligned} m_1(\vec{\omega}^*, \mathbf{x}) &= a_{11}(\vec{\omega}^*) s_1(\vec{\omega}^*, \mathbf{x}) + a_{12}(\vec{\omega}^*) s_2(\vec{\omega}^*, \mathbf{x}) \\ m_2(\vec{\omega}^*, \mathbf{x}) &= a_{21}(\vec{\omega}^*) s_1(\vec{\omega}^*, \mathbf{x}) + a_{22}(\vec{\omega}^*) s_2(\vec{\omega}^*, \mathbf{x}) \end{aligned} \tag{22}$$

degenerates to

$$\begin{aligned} m_1(\vec{\omega}^*, \mathbf{x}) &= a_{11}(\vec{\omega}^*) s_1(\vec{\omega}^*, \mathbf{x}) \\ m_2(\vec{\omega}^*, \mathbf{x}) &= a_{21}(\vec{\omega}^*) s_1(\vec{\omega}^*, \mathbf{x}) \;. \end{aligned} \tag{23}$$

In this case, the rank of the equation system is 1. It is a singular set and it is not invertible. Note that it is not known *a-priori* that $s_2(\vec{\omega}^*, \mathbf{x}) = 0$ since s_1 and s_2 are unknown. The separation algorithm assumes that there are two independent sources with non-zero energy at each channel, which is not true in this case. Therefore, the separation results in this case are useless.

5.2.2 Dependency on the Data

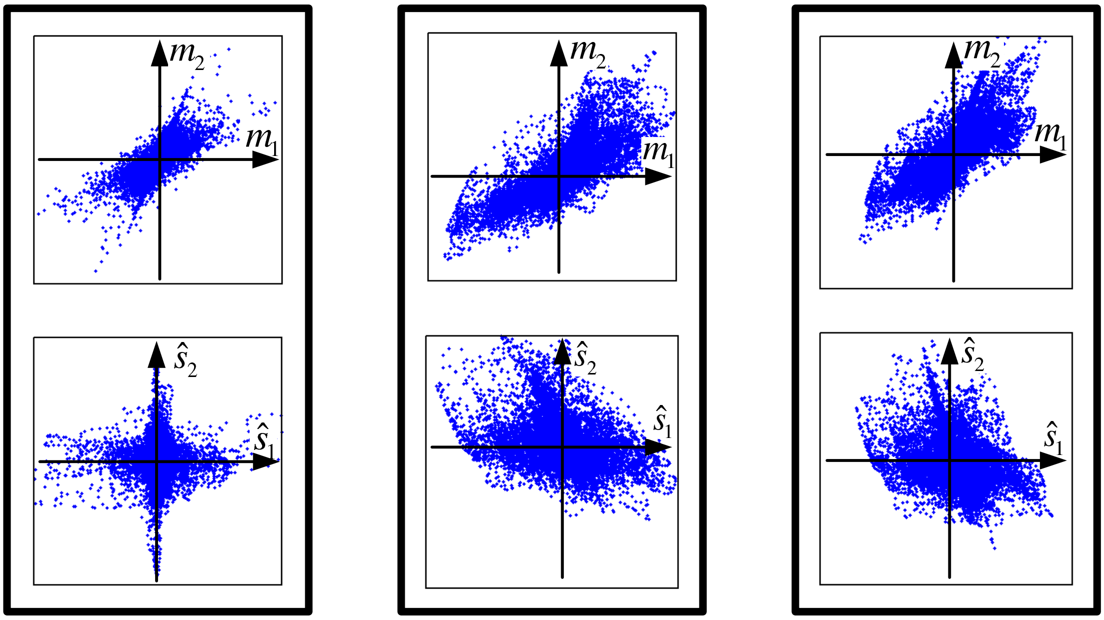
In general, the quality of ICA optimization depends on the acquired data. This is illustrated in Fig. 11. The results of three identical separation simulations of sparse sources are represented by scatter plots of the mixed sources (Top) and the estimated sources (Bottom). The mixing matrix in the three simulations is identical. The only difference between the three simulations is the source data, which belongs to different sub-band images. Fig. 11(a) presents a scatter plot that is interpreted as corresponding to independent estimated sources, i.e., good separation results, while the structure of the scatter plots in Figs. 11(b) and 11(c) indicates poor separation (See App. B). In Sec. 6, we detail how to overcome this data dependency problem by using a-priori knowledge about the blur kernel.

6 Inter-Channel Knowledge Transfer

In Sec. 5.2 we showed that the separation performance can be poor in several channels. Moreover, performance is affected by the acquired data. In addition, the separation performance can be hindered by permutation and scale ambiguities, as explained in Sec. 5.1. In this section we bypass some of these problems by exploiting prior knowledge about the convolutive process.

6.1 Exploiting a Parametric Blur Model

Blur is caused by optical defocus. Defocus blur can be parameterized [3, 5, 14, 25, 40, 48]. A rough parametric model is a simple 2D Gaussian kernel with different widths in the x and y directions [37]. In this simple case, the mixing matrix at each frequency channel is



(a) Good separation. (b) Poor separation. (c) Poor separation.

Figure 11: Separation results for identical mixing matrix but different data: [Top] Scatter plots of mixed signals. [Bottom] Scatter plots of the estimated sources. The quality of the separation depends on the data

of the form

$$\mathbf{A} = \begin{bmatrix} 1 & G_{\vec{\xi}_1}(\vec{\omega}) \\ G_{\vec{\xi}_2}(\vec{\omega}) & 1 \end{bmatrix} . \tag{24}$$

Here $\vec{\xi_i} = [\xi_{i,x}, \xi_{i,y}]$ is a vector of the blur parameters and

$$G_{\xi_i}(\vec{\omega}) = \exp\left(-\frac{\omega_x^2}{2\xi_{i,x}^2}\right) \exp\left(-\frac{\omega_y^2}{2\xi_{i,y}^2}\right) . \tag{25}$$

Thus, the separation matrix in each channel is parameterized by $\vec{\xi}_1, \vec{\xi}_2$ and is of the form

$$\mathbf{W} = \begin{bmatrix} 1 & -G_{\vec{\xi}_1}(\vec{\omega}) \\ -G_{\vec{\xi}_2}(\vec{\omega}) & 1 \end{bmatrix} [\det(\mathbf{A})]^{-1} . \tag{26}$$

It is important to note that the vectors $\vec{\xi}_1, \vec{\xi}_2$ are the *same for all frequency channels.* Hence, there is a very small number of actual unknown blur variables. On the other hand, there is a large number of frequency channels upon which the estimation of these variables can be based.

Suppose we choose two channels $\vec{\omega}^*$ and $\vec{\omega}^\#$, that yield good separation results.[4] Then,

---

[4] In Sec. 6.2 we detail how to choose these channels.

we can create the following set of equations

$$\left\{ \begin{array}{l} -G_{\vec{\xi}_1}(\vec{\omega}^*)[\det(\mathbf{A})]^{-1} = w_{1,2}(\vec{\omega}^*) \\ -G_{\vec{\xi}_2}(\vec{\omega}^*)[\det(\mathbf{A})]^{-1} = w_{2,1}(\vec{\omega}^*) \\ -G_{\vec{\xi}_1}(\vec{\omega}^{\#})[\det(\mathbf{A})]^{-1} = w_{1,2}(\vec{\omega}^{\#}) \\ -G_{\vec{\xi}_2}(\vec{\omega}^{\#})[\det(\mathbf{A})]^{-1} = w_{2,1}(\vec{\omega}^{\#}) \end{array} \right. , \tag{27}$$

where $w_{i,k}(\vec{\omega})$ is an element of the separation matrix $\mathbf{W}(\vec{\omega})$. This set has four equations with four unknown variables corresponding to the two vectors $\vec{\xi}_1, \vec{\xi}_2$. We solve this set to find the blur parameters $\vec{\xi}_1, \vec{\xi}_2$. We thus derive these blur parameters based on those few selected channels.

Now, we can use the blur parameters to calculate the separation operator $\mathbf{W}$ for *all the frequency channels*. Using this $\mathbf{W}$, we invert the mixing process. Moreover, knowing the parameters of $\mathbf{W}$ allows us to invert the convolutive blur process in the Fourier domain or in the image domain rather than the STFT domain. This yields better accuracy and computational efficiency.

It may be possible to achieve higher accuracy using more than two parameters to represent each blur kernel. In this case, more than two channels are needed in order to obtain enough equations to solve for all of the model parameters. In general, the minimal number of channels needed for parameter estimation equals the number of model parameters.

We stress that the blur parameterization solves uniquely the permutation, scale and sign ambiguities: the sources are not derived in a random order or with inter-channel imbalance, but in a way that must be consistent with the blur model, hence with the image formulation process. In addition, the data dependency problem is alleviated since the separation operator is estimated based on selected channels that perform well.

6.2 Selecting a Good Channel

In order to select a good channel, we need a quality criterion for the error in the estimated separation matrix. The criterion we use was developed in Ref. [19]. We present here a short summary.

Define the ideal separation matrix $\mathbf{W}^*$ at a certain channel as the true solution to the separation problem $\mathbf{W}^* = \mathbf{A}^{-1}$. It satisfies

$$\mathbf{W}^* = \arg\min_{\mathbf{W}} E[\mathcal{I}_{\mathbf{W}}(\hat{s})] \ . \tag{28}$$

The expectation in this equation is based on the true and unknown PDF of m. However, we acquire a finite amount of measurements N in each channel. Since we use empirical averaging, the solution resulting from the optimization is not $\mathbf{W}^*$. Define $\hat{\mathbf{W}}$ as the solution of the optimization problem given only the acquired data,

$$\hat{\mathbf{W}} = \arg\min_{\mathbf{W}} \left\{ -\sum_{k=1}^{2}\sum_{n=1}^{N} \log\{p_{\hat{s}_k}[\hat{s}_k(n)]\} - \log|\det(\mathbf{W})| \right\} . \tag{29}$$

Define the error between the ideal matrix and the estimated matrix as

$$\Delta\mathbf{W} \equiv \mathbf{W}^* - \hat{\mathbf{W}} . \tag{30}$$

The covariance matrix of this error is given by

$$E[(\Delta\mathbf{W})(\Delta\mathbf{W}^T)] . \tag{31}$$

The diagonal elements of this covariance matrix represent estimates for the square errors introduced to the reconstructed sources. Therefore, a simple quality criterion for the separation is

$$\sum_{k=1}^{2}\{E[\Delta\mathbf{W}\Delta\mathbf{W}^T]\}_{k,k} , \tag{32}$$

where $\{\}_{k,k}$ stands for a diagonal element of the covariance matrix.

An approximation for the covariance matrix was developed in Ref. [31].

$$E[(\Delta\mathbf{W})(\Delta\mathbf{W}^T)] \approx H^{-1}E[\nabla\mathcal{I}_{\mathbf{W}^*}(\hat{s})\nabla\mathcal{I}_{\mathbf{W}^*}(\hat{s})^T](H^{-1})^T . \tag{33}$$

Here H is the Hessian of $\mathcal{I}_{\mathbf{W}^*}(\hat{s})$. The expectation is approximated by empirical averaging, and using $\hat{\mathbf{W}}$ instead of $\mathbf{W}^*$:

$$\begin{aligned} &E[\nabla\mathcal{I}_{\mathbf{W}^*}(\hat{s})\nabla\mathcal{I}_{\mathbf{W}^*}(\hat{s})^T] \approx \\ &\quad \sum_{k=1}^{2}\sum_{n=1}^{N} \Big(\log\{p_{\hat{s}_k}[\hat{s}_k(n)]\} + \log|\det(\hat{\mathbf{W}})|\Big)\Big(\log\{p_{\hat{s}_k}[\hat{s}_k(n)]\} + \log|\det(\hat{\mathbf{W}})|\Big) . \end{aligned} \tag{34}$$

The expression in Eq. (34) is a function of the estimated sources $\hat{s}_k$ and the estimated separation matrix $\hat{\mathbf{W}}$. Both $\hat{s}_k$ and $\hat{\mathbf{W}}$ are the outputs of ICA performed in each frequency channel. Hence, we have a closed-form expression that estimates the separation quality in each channel.

Thus, in our algorithm we first perform ICA in all the frequency channels. We then calculate Eq. (34) in each channel. Then, we choose the best channels as those with the smallest values of Eq. (34). These channels are used in Sec. 6

7 Algorithm Summary

In this section, we summaries the algorithm. The inputs to the algorithm are the mixtures of blurred images. The outputs of the algorithm are separated images.

1. Choose the STFT window width (see App. A).
2. To enable factorization, apply STFT on the input images.
3. For each frequency (sub-band) channel:
    - Find $\mathbf{W}(\vec{\omega})$ that minimizes $\hat{\mathcal{I}}^{\vec{\omega}}$ as in Eq. (19).
    - Calculate the estimated error of the separation (Eqs. 32 and 34).
4. Following step 3, select the channels with the lowest estimated errors. Based on the separation matrices $\mathbf{W}(\vec{\omega})$ of these channels, extract the vector of blur parameters $\mathbf{p}$.
5. Calculate the blur operator based on $\mathbf{p}$, and invert the mixing process in the Fourier domain or in the image domain.
6. If the separation is unsatisfactory, repeat stage 1 and subsequently the other stages.

8 Direct BSS Using a Parametric Blur Model

Our algorithm (Sec. 7) has become more complicated than a trick of factorization. The overhead is caused by channels for which standard ICA tools yield an unreliable $\mathbf{W}$. Thus, $\mathbf{W}$ is estimated based on other channels, where the estimation is reliable. This inter-channel knowledge transfer exploits a blur parametric model $\mathbf{p}$. One may suggest a shortcut: estimating the PSF parameters directly, without ICA tools which optimize over $\mathbf{W}$. In this section we discuss this alternative.

To recap, recall that for each frequency channel, we minimize Eq. (19) explicitly over $\mathbf{W}$. Based on the estimated $\mathbf{W}$ in some channels, we derive the low dimensional vector of blur parameters $\mathbf{p}$. Let us consider an alternative approach that skips the intermediate estimation of $\mathbf{W}$. It would directly optimize over $\mathbf{p}$:

$$\min_{\mathbf{p}} \hat{\mathcal{I}}^{\vec{\omega}}_{\hat{s}_1,\hat{s}_2} = \min_{\mathbf{p}} \left\{ \sum_{k=1}^{2} \frac{1}{N} \sum_{n=1}^{N} |\hat{s}_k(\vec{\omega}, n)| - \log |\det[\mathbf{W}(\mathbf{p})]| \right\} . \tag{35}$$

Hence, the elements of $\mathbf{W}$ would not be optimized directly in any channel, but be derived from the explicit optimization variables $\mathbf{p}$, in contrast to Eq. (19).

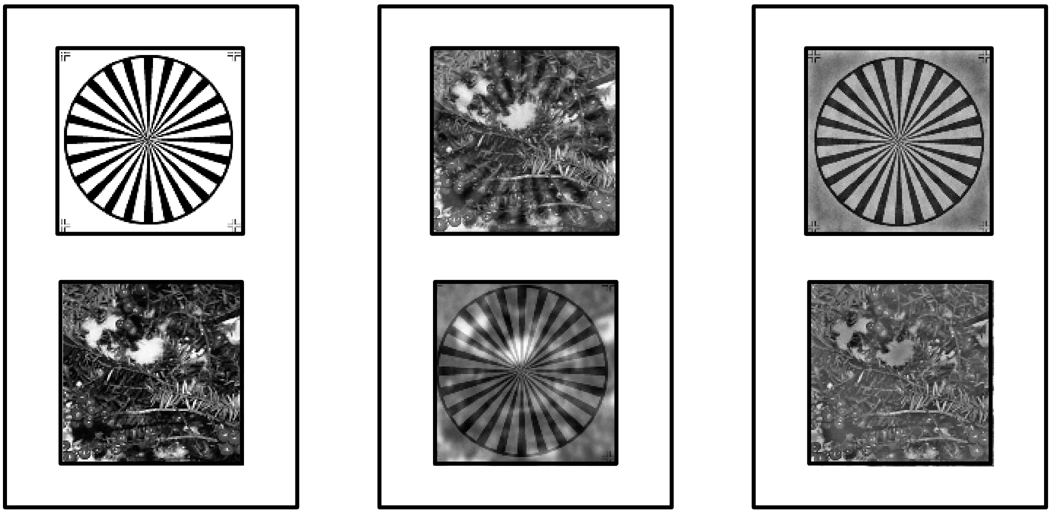
Such an approach is problematic. The MI is a convex function in $\mathbf{W}$ but it is typically *not* convex in $\mathbf{p}$. Even for simple models such as the Gaussian model (Eq. 25), the MI is not a convex function in $\vec{\xi}_1, \vec{\xi}_2$. Moreover, it is typically *multimodal.* The MI function can have many local minima. When using gradient-based optimization for multimodal functions, there is no guarantee of convergence to the global minimum. Therefore, a global optimization algorithm must be used for such an approach. In general, global optimization algorithms are less efficient than local gradient-based optimization algorithms. For this reason, in contrast to Eq. (35), we do not optimize over the blur parameters $\mathbf{p}$. Rather, we optimize over the separation matrix $\mathbf{W}(\vec{\omega})$ by Eq. (19), which is a *convex* optimization problem in each frequency channel.

9 Demonstrations

In this section we present results of a simulation and an experiment we made. For the simulation, we used two standard pictures [Fig. 12(a)] as the two scene layers. The blur kernel we used is a Gaussian with STD of $\xi_1 = 2, \xi_2 = 3$. We used the same blur for both layers. In addition, we added an i.i.d. Gaussian noise with a standard deviation of ~ 2.5 gray levels to the mixed images. The mixed, noisy images are presented in Fig. 12(b). The separation optimization was performed using a 15×15 STFT channels. The reconstruction was done in the Fourier domain.

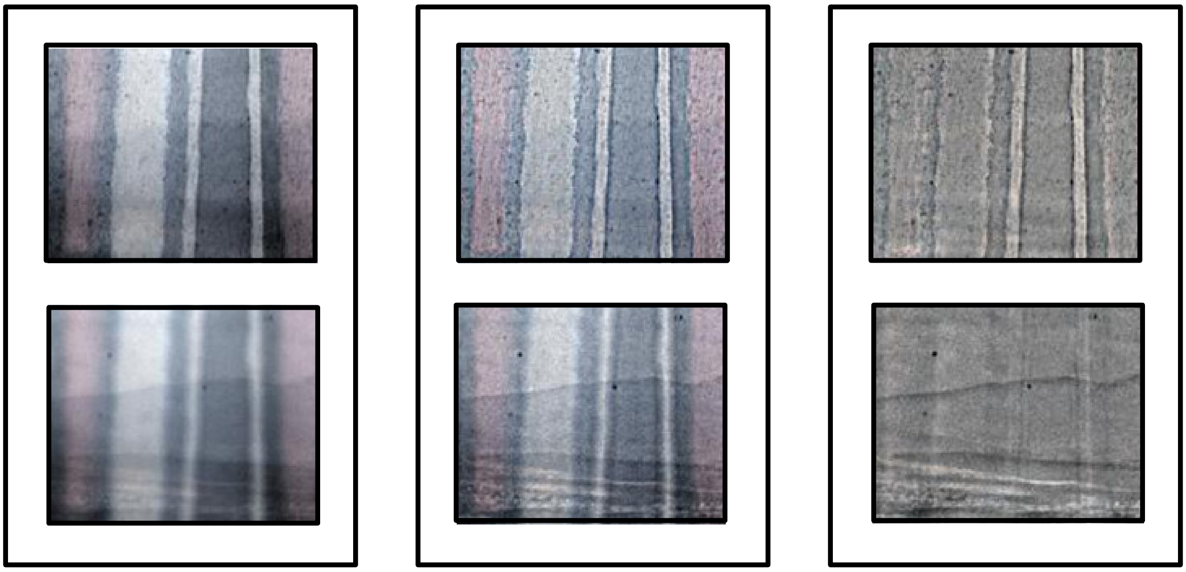
The separation results are presented in Fig. 12(c). The separation quality seems good. There is no visible crosstalk between the images. However, as seen in Fig. 12(c), the background (DC) value is incorrect and the contrast of the reconstructed images is lower than the contrast of the original images. This problem stems from the nature of the inversion problem, as shown in [37], and mentioned in Sec. 5.2.1: we cannot expect good separation and recovery in low spatial frequencies.

In addition, we present the results of an experiment we made. Two pictures of a glass covered painting were taken using a digital camera having a linear radiometric response, Nikon D-100. As shown in Fig. 13(a), an outdoor scene is reflected from the glass cover of the painting. At the top of Fig. 13(a) the transmitted painting is in focus and the reflection is blurred. At the bottom of Fig. 13(a) the reflected scene is in focus while the transmitted painting is blurred. As mentioned in Sec. 5.2.1, we cannot expect good reconstruction in the low frequencies. Therefore, for comparison, a high-pass version of the original images is given in Fig. 13(b). This version still exhibits significant crosstalk. The separation results are given in Fig. 13(c). The crosstalk between the resulting layers



(a) (b) (c)

Figure 12: Separation results:(a) Original sources. (b) Convolved and mixed sources. (c) Reconstructed sources.



(a) (b) (c)

Figure 13: Experimental results: (a) Acquired mixed images. (b) High-pass filtered version of the mixed images. (c) Reconstructed layers.

is much reduced.

10 Discussion

We have presented an algorithm for separation of convolutive image mixtures. We focused on separation of reflections. We used STFT in order to factor this complex problem into several simple pointwise problems. This factorization simplifies the optimization problem significantly. In addition, our algorithm exploits image statistics and *a-priori* knowledge about the physical blur process. This enhances the computational efficiency and numerical stability. In addition, the parametric model for the blur kernel solves the scale, sign and permutation ambiguities typical of ICA.

We have focused on separation of two image layers. However, there are applications in which there are more than two layer of mixed images. Nevertheless, MI can be generalized to K signals as the Kullback-Leibler Distance between the joint PDF and $p^{\text{indp}}_{\hat{s}_1,\ldots,\hat{s}_K}(\hat{s}_1,\ldots,\hat{s}_K)$:

$$\mathcal{I}_{\hat{s}_1,\ldots,\hat{s}_K} = \int \cdots \int p(\hat{s}_1,\ldots,\hat{s}_K) \log\left[\frac{p(\hat{s}_1,\ldots,\hat{s}_K)}{p(\hat{s}_1)\cdots p(\hat{s}_K)}\right] ds_1 \ldots, ds_K \,. \tag{36}$$

Eq. (36) can be rewritten as (see for example [8]).

$$\mathcal{I}_{\hat{s}_1,\ldots,\hat{s}_K} = \mathcal{H}_{\hat{s}_1} + \cdots + \mathcal{H}_{\hat{s}_K} - \mathcal{H}_{\hat{s}_1,\ldots,\hat{s}_K} \,. \tag{37}$$

This generalized MI preserves the property of being equal to zero if and only if the sources are statistically independent. Therefore, we can use the generalized MI as an indicator for statistical dependency. When $K > 2$ the dimensionality problem of estimating the joint entropy is even more severe. Therefore, the advantage gained by bypassing the joint entropy estimation is even more significant.

The convolutive image separation algorithm has a single parameter to tweak. As explained in App. A, the width of the STFT window needs to be determined, for obtaining good separation. This parameter depends on the acquired images. We used trial and error to determine it. This calls for further research into full automation of the process.

In addition,we used source independence as our separation criterion. However, in applications such as optical sectioning microscopy, the sources independence may be invalid. An interesting extension to this research is to adapt this algorithm to deal with mutually dependent signals, and apply it to optical sectioning microscopy.

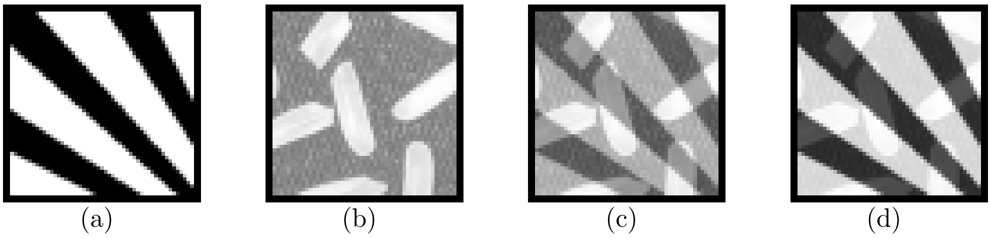

(a) (b) (c) (d)

Figure 14: Example scene. (a),(b) Image layers. (c),(d) Two image mixtures.

A Setting the Width of the STFT Window

The width of the STFT window determines the number of STFT channels. The wider the window, the more channels we have. How many STFT channels do we need? The answer is that the optimal window width depends on the input data itself. In particular, as mentioned in Section 4.1, the STFT window should be wider than the effective width of the point spread function (PSF) of the imaging systems. Otherwise, Eq. (13) is invalid.

Another consideration for the window width is the acquired scene. Typically, an increase of the number of channels reduces the energy per channel (sample values very close to zero). Therefore, each channel has a sparser data distribution. However, if one of the sources has almost no energy in a certain frequency channel, then the source's mixture in that channel cannot be inverted. This scenario is detailed in Sec. 5.2.1.

To illustrate this phenomenon, consider for example the scenario in Fig. 14. It contains two pointwise mixtures of images.[5] The mixed pictures were transformed to the short time fourier domain using several window widths. Scatter plots of a representative channel[6] for various window widths are presented in Fig. 15. Referring to App. B for interpretation of scatter plots, we conclude that in Fig. 15 a window width of 23×23 pixels appears to be best for this scene. A smaller window yields dense distributions, while a wider window almost eliminates one of the sources. On the other hand, a window of 23×23 pixels results in a clear "X" shape, which indicates two mixed sources having sparse PDFs.

The fact that an optimal window width exists is consistent with the observation in [7]. Ref. [7] showed that a transformation of images that yield a very sparse data, depends on the images themselves. In our work we limit ourselves to the STFT rather than to an arbitrary linear transformation. Nevertheless, the window width is a parameter of the STFT. Therefore, by changing the STFT window width we change the transformation applied to the images. Hence, by setting an optimal window width we obtain a sparser

---

[5] This implies that we do not have a lower bound on the width of the STFT window.

[6] For each window width, the representative channel is the channel for which the best separation results can be achieved.

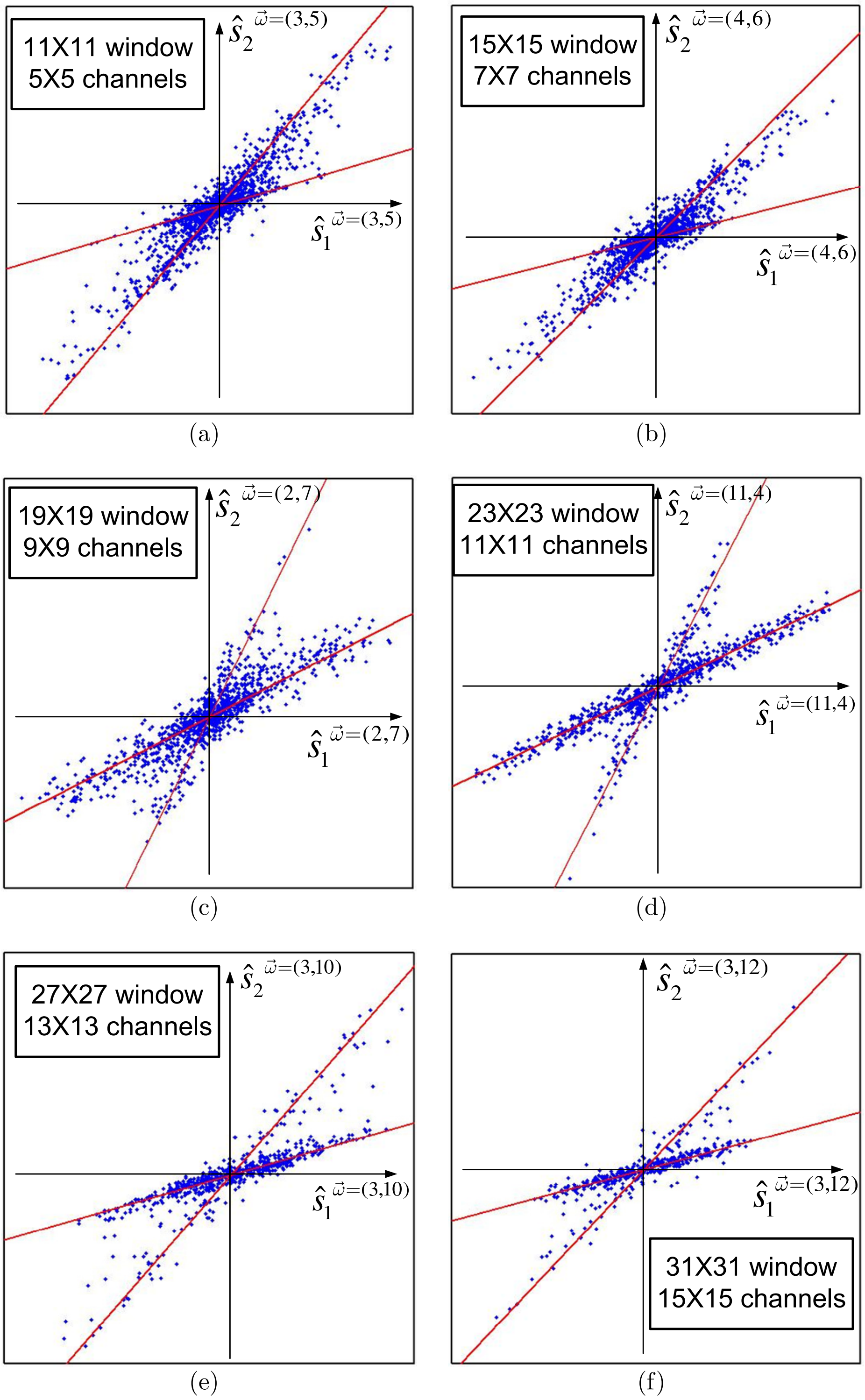
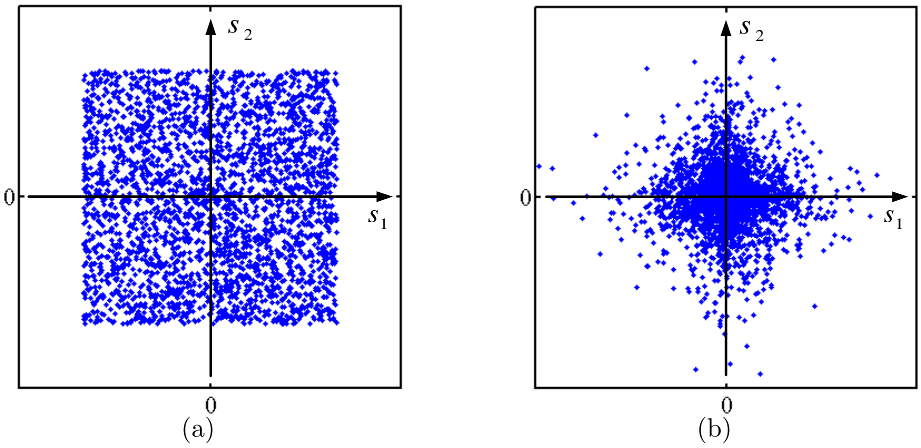


Figure 15: Scatter plots of representative STFT channels, based on different window widths. The lines correspond to the mixing matrix rows and indicate the expected location of the scattered points. A 23×23 pixels window yields the scatter plot that seems to be the best. In some of the other plots one of the sources has disappeared.



(a) (b)

Figure 16: Scatter plots of independent signals. (a) Dense signals. (b) For sparse signals, the scattered samples concentrate around a cross shape, aligned with the axes.

sub-band images.

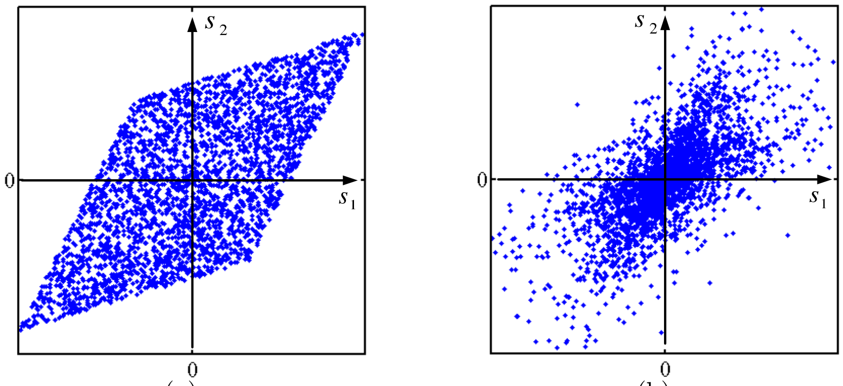
To conclude: the optimal STFT window width is a parameter that depends on the acquired scene. In this work, we defined the window width by trial and error. We performed the source separation with several different widths and chose the best result.

B Scatter Plot of Sparse and Independent Sources

This appendix describes the interpretation of scatter plots in the context of sparse and independent signals. While this is common knowledge in the literature, the explanation is given here to make the paper self-contained.

A scatter plot is very useful for revealing whether a source is sparse or dense. For example, Fig. 16(a) presents a scatter plot of two dense and independent signals, each having a uniform distribution. Note that the signals' samples are uniformly scattered. Fig. 16(b) presents a scatter plot of two sparse and independent signals, each having an exponential distribution. Note that the signals' samples are concentrated around a cross shape aligned with the axes, particularly around the origin. The samples are sparsely scattered as the distance from the origin increases.

In addition, a scatter plot is very useful for revealing whether two sources are independent or not. While Fig. 16 plots independent signals, Fig. 17 displays scatter plots of signals that are linearly dependent. The latter were created by pointwise mixtures of independent sources. Note that the scattered samples in Fig. 17 create inseparable plots. In the case of sparse sources [Fig. 17(b)], the plot is manifested as a skewed "X". The skew angle depends on the coefficients of the mixture.



(a) (b)

Figure 17: Scatter plots of mutually dependent signals. (a) Linearly dependent dense signals. (b) Linearly dependent sparse signals. Compare to Fig. 16, and note the different plot appearance when signals become mutually dependent. For sparse signals, the scattered samples concentrate around a skewed "X" shape.

Acknowledgments

Yoav Schechner is a Landau Fellow - supported by the Taub Foundation, and an Alon Fellow. The work was supported by the "Dvorah" Fund of the Technion. The research was conducted in the Ollendorff Minerva Center in the Elect. Enc. Dept. at the Technion. Minerva is funded through the BMBF.

References

[1] A. Agrawal, R. Raskar, S.K. Nayar, and Y. Li. Removing photography artifacts using gradient projection and flash-exposure sampling. *ACM Transactions on Graphics*, 24(3):828–835, 2005.

[2] M. Borga and H. Knutsson. Estimating multiple depths in semi-transparent stereo images. In *Proc. Scandinavian Conf. Image analysis*, pages 127–134, 1999.

[3] M. Born and E. Wolf. *Principles of optics*. Pergamon, Oxford, 1975.

[4] R. Boscolo, H. Pan, and V.P. Roychowdhury. Non-parametric ICA. In *Proc. ICA2001*, pages 13–18.

[5] J. Braat, P. Dirksen, and A.J.E.M. Janssen. Assessment of an extended nijboer-zernike approach for the computation of optical point-spread functions. *J. Opt. Soc. America A*, 19(5):858–70, 2002.

[6] A. Bronstein, M. Bronstein, M. Zibulevsky, and Y.Y. Zeevi. Separation of semireflective layers using sparse ICA. In *Proc. IEEE ICIP*, pages 313–316, 2003.

[7] A.M. Bronstein, M.M. Bronstein, M. Zibulevsky, and Y.Y. Zeevi. Blind deconvolution of images using optimal sparse representations. *IEEE Trans. on Image Processing*, 14(6):726–736, 2005.

[8] T.M. Cover and J.A. Thomas. *Elements of information theory.* John Wiley and sons, NY, 1991.

[9] A. Dapena, C. Serviere, and L. Castedo. Separation of convolutive mixtures in the frequency-domain using only two frequency bins. In *Proc. IEEE Int. Conference on Acoustics Speech and Signal Processing*, volume 2, pages 1633–36, 1001.

[10] T. Darrell and E. Simoncelli. Nulling filters and the separation of transparent motions. In *Proc. IEEE Computer Vision and Pattern Recognition*, pages 738–739, 1993.

[11] T. Darrell and E.P. Simoncelli. Separation of transparent motion into layers using velocity-tuned mechanisms. In *Proc. European Conf. Computer Vision*, 2004.

[12] S.C. Douglas and X. Sun. Blind separation of acoustical mixtures without time-domain deconvolution or decorrelation. In *Proc. IEEE Int. Conference on Neural Networks for Signal Processing*, pages 323–32, 2001.

[13] H. Farid and E.H. Adelson. Separating reflections from images by use of independent components analysis. *J. Optical Soc. America*, 16(9):2136–2145, 1999.

[14] E. Hecht. *Optics.* Addison-Wesley, Mass., 1987.

[15] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent component analysis.* John Wiley and Sons, NY, 2001.

[16] M. Irani and S. Peleg. Motion analysis for image enhancement: Resolution, occlusion, and transparency. *J. Visual Communication and Image Representation*, 4(4):324–335, 1993.

[17] M. Irani, B. Rousso, and S. Peleg. Computing occluding and transparent motions. *Int. J. Computer Vision*, 12(1):5–16, 1994.

[18] W. Kasprzak and A. Okazaki. Blind deconvolution of timely-correlated sources by homomorphic filtering in Fourier space. In *Proc. ICA2003*, pages 1029–34.

[19] P. Kisilev, M. Zibulevsky, and Y.Y. Zeevi. Blind separation of mixed images using multiscale transforms. In *Proc. IEEE Int. Conference on Image Processing*, volume 1, pages 309–112, 2003.

[20] I. Kopriva, Z. Devcic, and H. Szu. An adaptive short-time frequency domain algorithm for blind separation of nonstationary convolved mixtures. In *Proc. IEEE Int. Conference on Neural Networks*, volume 1, pages 424 –429,, 2001.

[21] T.-W. Lee, A.J. Bell, and R. Orglmeister. Blind source separation of real world signals. In *Proc. IEEE Int. Conference on Neural Networks*, volume 4, pages 2129–34, 1997.

[22] A. Levin and Y. Weiss. User assisted separation of reflections from a single image using a sparsity prior. In *Proc. European Conf. Computer Vision*, pages 602–613, 2004.

[23] A. Levin, A. Zomet, and Y. Weiss. Separating reflections from a single image using local features. In *Proc. IEEE Computer Vision and Pattern Recognition*, volume 1, pages 306–313, 2004.

[24] F. Macias-Garza, A.C. Bovik, K.R. Diller, S.J. Aggarwal, and J.K. Aggarwal. The missing cone problem and low-pass distortion in optical serial sectioning microscopy. In *Proc. IEEE Acoustics Speech and Sig. Processing*, volume 2, pages 890–893, 1988.

[25] V.N. Mahajan. Zernike circle polynomials and optical aberrations of systems with circular pupils. *Applied-Optics*, 33(34):8121–4, 1994.

[26] N. Murata, S. Ikeda, and A. Ziehe. An approach to blind source separation based on temporal structure of speech signals. *Neurocomputing*, 41:1–24, 2001.

[27] M. Najar, M.A. Lagunas, and I. Bonet. Blind wideband source separation. In *Proc. IEEE Int. Conference on Acoustics Speech and Sig. Processing*, volume 4, pages 19–22, 1994.

[28] L. Parra and C. Spence. Convolutive blind separation of non-stationary sources. *IEEE Trans. on Speech and Audio Processing*, 8:320–327, 2000.

[29] D.T. Pham. Contrast functions for blind source separation and deconvolution of sources. In *Proc. ICA2001*, pages 37–42.

[30] D.T. Pham. Fast algorithm for estimating mutual information, entropies and score functions. In *Proc. ICA2003*, pages 17–22.

[31] D.T. Pham and P. Garrat. Blind separation of a mixture of independent sources through a quasi-maximum likelihood approach. *IEEE Trans. Sig. Proc.*, 45(7):1712–1725, 1997.

[32] K. Rahbar and J.P. Reilly. Blind source separation algorithm for MIMO convolutive mixtures. In *Proc. ICA*, pages 242–247, 2001.

[33] I. Sabala, A. Cichochi, and S. i. Amari. Relationships between instantaneous blind source separation and multichannel blind deconvolution. In *Proc. IEEE Int. Conference on Neural Networks and Computational Intelligence*, volume 1, pages 39–44, 1998.

[34] H. Sahlin and H. Broman. Separation of real-world signals. *Signal Processing*, 64(1):103–13, 1998.

[35] B. Sarel and M. Irani. Separating transparent layers through layer information exchange. In *Proc. European Conf. Computer Vision*, pages 328–341, 2004.

[36] Y. Y. Schechner and N. Karpel. Clear underwater vision. In *Proc. IEEE CVPR*, volume 1, pages 536–543, 2004.

[37] Y.Y. Schechner, N. Kiryati, and R. Basri. Separation of transparent layers using focus. *Int. J. Computer Vision*, 89:25–39, 2000.

[38] Y.Y. Schechner, J. Shamir, and N. Kiryati. Polarization and statistical analysis of scenes containing a semi-reflector. *J. Opt. Soc. America A*, 17:276–284, 2000.

[39] S. Shamsunder and G.B. Giannakis. Multichannel blind signal separation and reconstruction. *IEEE Trans. on Speech and Audio Processing*, 5(6):515–528, 1997.

[40] C.J.R. Sheppard and I.J. Cooper. Fresnel diffraction by a circular aperture with off-axis illumination and its use in deconvolution of microscope images. *J. Opt. Soc. America A*, 21(4):540–5, 2004.

[41] M. Shizawa. Direct estimation of multiple disparities for transparent multiple surfaces in binocular stereo. In *Proc. IEEE Int. Conf. Computer Vision*, pages 447–54, 1993.

[42] M. Shizawa and K. Mase. Principle of superposition: A common computational framework for analysis of multiple motion. *WVM*, 91:164–172.

[43] S. Shwartz, M. Zibulevsky, and Y.Y. Schechner. ICA using kernel entropy estimation with NlogN complexity. In *Proc. ICA2004*, pages 422–429.

[44] S. Shwartz, M. Zibulevsky, and Y.Y. Schechner. Fast kernel entropy estimation and optimization. *Sig. Processing, Special Issue on Information Theoretic Sig. Processing*, 85(5):1045–1058, 2005.

[45] C. Simon, Ph. Loubaton, C. Vignat, C. Jutten, and G. d'Urso. Blind source separation of convolutive mixtures by maximization of fourth-order cumulants: the non i.i.d. case. In *Proc. IEEE Int. Conference on Signals Systems and Computers*, volume 2, pages 1584 – 88, 1998.

[46] E.P. Simoncelli. Statistical models for images: Compression, restoration and synthesis. In *Proc. IEEE Asilomar Conf. Sig. Sys. and Computers*, pages 673–678, 1997.

[47] P. Smaragdis. Blind separation of convolved mixtures in the frequency domain. *Neurocomputing*, 22:21–34, 1998.

[48] P.A. Stokseth. Properties of a defocused optical system. *J. Opt. Soc. America A*, 59(10):1314–21, 1969.

[49] A. Taleb, J. Sole, and C. Jutten. Quasi-nonparametric blind inversion of wiener systems. *IEEE Transactions on Signal Processing*, 49(5):917–924, 2001.

[50] H.-L. N. Thi and C. Jutten. Blind source separation for convolutive mixtures. *Signal processing*, 45(3):209–229, 1995.

[51] J. Toro, F. Owens, and R. Medina. Using known motion fields for image separation in transparency. *Pattern Recognition Letters*, 24(1–3):597–605, 2003.

[52] Y. Tsin, S.B. Kang, and R. Szeliski. Stereo matching with reflections and translucency. In *Proc. IEEE Computer Vision and Pattern Recognition*, volume 1, pages 702–709, 2003.

[53] E. Weinstein, M. Feder, and A.V. Oppenheim. Multi-channel signal separation by decorrelation. *IEEE Trans. on Speech and Audio Processing*, 1(4):405–413, 1993.

[54] D. Yellin and E. Weinstein. Criteria for multichannel signal separation. *IEEE Trans. on Signal Processing*, 42(8):2158–2168, 1994.

[55] D. Yellin and E. Weinstein. Multichannel signal separation: methods and analysis. *IEEE Trans. on Signal Processing*, 44(1):106–118, 1996.

[56] W. Zhou and C. Kambhamettu. Separation of reflection component by Fourier decoupling. In *Proc. Asian Conf. Computer Vision*, pages 27–30, 2004.