Available online at www.sciencedirect.com
ScienceDirect
Neurocomputing 71 (2008) 2164–2179
NEUROCOMPUTING
www.elsevier.com/locate/neucom

Blind separation of convolutive image mixtures

Sarit Shwartz*, Yoav Y. Schechner, Michael Zibulevsky

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

Available online 13 March 2008

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. We seek blind source separation (BSS) of such mixtures. However, achieving this by direct optimization of mutual information is very complex and suffers from local minima. Thus, we devise an efficient approach to solve these problems. While achieving high quality image separation, we take steps that make the problem significantly simpler than a direct formulation of convolutive image mixtures. These steps make the problem practically convex, yielding a unique global solution to which convergence can be fast. The convolutive BSS problem is converted into a set of instantaneous (pointwise) problems, using a short time Fourier transform (STFT). Standard BSS solutions for instantaneous problems suffer, however, from scale and permutation ambiguities. We overcome these ambiguities by exploiting a parametric model of the defocus point spread function. Moreover, we enhance the efficiency of the approach by exploiting the sparsity of the STFT representation as a prior. We apply our algorithm to semi-reflections, and demonstrate it in experiments.

Keywords: Transparent layers; Image statistics; Defocus blur; Independent component analysis; Blind source separation; Sparsity

1. Introduction

Typical blind source separation (BSS) methods seek separation when the mixing process is unknown. However, loose prior knowledge regarding the mixing process often exists, due to its physical origin. In particular, this process can be represented by a parametric form, rather than a trivial representation of raw numbers. For example, consider convolutive image mixtures caused by defocus blur. This blur can be parameterized, yet the parameters' values are unknown. Such mixtures occur in tomography and microscopy [23,34]. They also occur in semi-reflections [34], e.g., from a glass window (see Fig. 1). Here, a scene imaged behind the semi-reflector is superimposed on a reflected scene. The light that reaches the camera contains contributions from the transmitted scene, as well as from reflected objects. This mixture is linear, and thus pointwise BSS methods were applied to semi-reflections [1,8,13,21,35]. In general, however, 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. Defocus blur is a convolutive process. It is linear and can be approximated as space-invariant in narrow fields of view. Hence, a semi-reflection is generally a realization of convolutive mixtures,[1] rather than pointwise ones.

We claim that BSS can benefit from such a parametrization, as it makes the estimation more efficient while helping to alleviate ambiguities. In the case of semi-reflections, our goal is to decompose the mixed and blurred images into the separate scene layers. 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 such BSS problems (see for example in Refs. [6,15,29,34,35] and references therein). Therefore, separation of such convolutive mixtures can be

*Corresponding author.
E-mail addresses: sarityaki@gmail.com (S. Shwartz), yoav@ee.technion.ac.il (Y.Y. Schechner), mzib@ee.technion.ac.il (M. Zibulevsky).

[1]Ignoring the convolutive aspect is possible in special cases where the scene components are at similar distances from the camera, and the light is intense enough for the camera iris to be small. Some methods, which do not handle the convolutive aspect of the mixture, use a sequence of images taken from different viewpoints or sequences of moving layers [4,16,46,47,49]. Another method exploits flash photography [1].

0925-2312/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2007.07.042

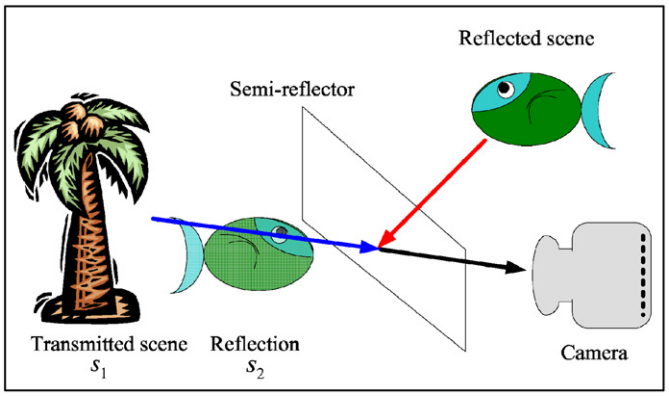

Fig. 1. An image of a mixture of two layers, created by a semi-reflector.

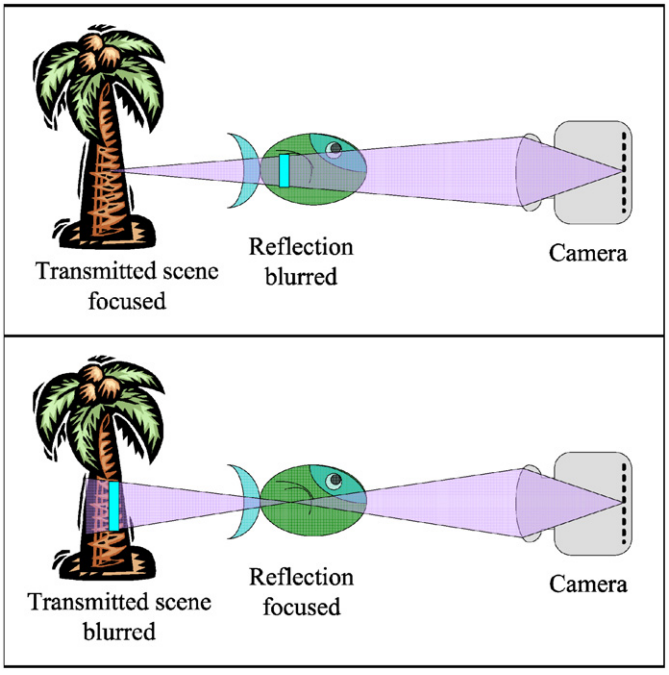

Fig. 2. Image acquisition of the scene depicted in Fig. 1: [Top] the camera is focused on the transmitted object while the reflection is blurred; [Bottom] the camera is focused on the reflection while the transmitted object is blurred.

achieved by minimizing the MI of the estimated objects. An attempt by Ref. [34] used exhaustive search, hence being computationally prohibitive. Moreover, MI is not a convex function of the optimized parameters. Hence, it may be complex to converge to a global minimum in the method presented in [34]. An additional problem in [34] is the need to estimate the joint entropy of the images, which is an inaccurate operation. Moreover, [34] is not scalable for more than two images. An alternative approach taken in Ref. [9] minimized higher order cumulants. That method suffers from a scale ambiguity: the sources are reconstructed up to an unknown filter. Moreover, the method's complexity increases fast with the support of the separation kernel.

The complexity of convolutive source separation has been reduced in the domain of acoustic signals, by using frequency methods [11,19,25,27,36,42]. There, BSS is decomposed into several small pointwise problems by applying a short time Fourier transform (STFT). Then, standard BSS tools are applied to each of the STFT channels. However, these tools suffer from fundamental ambiguities, which may ruin the overall separation quality, if applied as-is to the convolutive image mixtures (more details on these ambiguities are in Appendix A). Ref. [17] suggested that these ambiguities can be overcome by nonlinear operations in the image domain. However, this method encountered performance problems when simulated over natural images.

In this paper, we show that these problems can be solved by exploiting a parametric model for the unknown blur. Moreover, we use the sparsity of STFT coefficients to yield a practically unique solution, without a global search. We use the frequency-decomposition principle described above, but overcome its associated problems. The algorithm was applied and demonstrated successfully both in real experiments of semi-reflected true natural scenes as well as in simulations of such scenes.

2. Problem formulation

2.1. Source separation

Let $\{s_1,\ldots,s_K\}$ be a set of K independent sources. Each source is of the form $s_k = s_k(\mathbf{x})$, $k = 1,\ldots,K$, where $\mathbf{x} = (x, y)$ is a two-dimensional (2D) spatial coordinate vector in the case of images. Let $\{m_1,\ldots,m_K\}$ be a set of K measured signals, each of which is a linear mixture of a convolved version of the sources

$$m_i(\mathbf{x}) = a_{i1} * s_1(\mathbf{x}) + \cdots + a_{iK} * s_K(\mathbf{x}), \quad i = 1,\ldots,K. \tag{1}$$

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

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

$$\hat{s}_k(\mathbf{x}) = w_{k1} * m_1(\mathbf{x}) + \cdots + w_{kK} * m_K(\mathbf{x}), \quad k = 1,\ldots,K, \tag{2}$$

where $w_{ik}(\mathbf{x})$ are linear spatially invariant filters. We should note that even if the convolution kernels $a_{ik}(\mathbf{x})$ are perfectly known, the recovery by Eq. (2) may not be stable for some image components, as described in Appendix B. All of the filter coefficients w_{k1} have continuous values. Thus, the estimated sources $\{\hat{s}_1\}_{k=1}^{K}$ can have any continuous value.

Our goal is: given only the measured signals $\{m_1,\ldots,m_K\}$, find a linear separation operator $\mathbf{W}$ that

inverts the mixing process, thereby separating the sources.[2] This separation task is an optimization problem. The separation criterion used is the independence of the estimated sources $\hat{s}_k$. We thus seek $\mathbf{W}$ that minimizes the MI of the estimated sources.

2.2. MI of image mixtures

MI is expressed by using the marginal entropies $\mathscr{H}_{\hat{s}_k}$ and the joint entropy of the estimated sources $\mathscr{H}_{\hat{s}_1,\ldots,\hat{s}_K}$ as (see for example [10]),

$$\mathscr{I}_{\hat{s}_1,\ldots,\hat{s}_K} = \sum_{k=1}^{K} \mathscr{H}_{\hat{s}_k} - \mathscr{H}_{\hat{s}_1,\ldots,\hat{s}_K}. \tag{3}$$

Eq. (3) is non-negative. It equals zero if and only if the sources are statistically independent. However, estimation of the joint entropy $\mathscr{H}_{\hat{s}_1,\ldots,\hat{s}_K}$ may be unreliable [39]. It can be avoided if the mixtures are pointwise, rather than convolutive. In pointwise mixtures, the separation operator $\mathbf{W}$ is a simple matrix, termed the *separation matrix*. In this case, the MI can be expressed as (see for example Ref. [15])

$$\mathscr{I}_{\hat{s}_1,\ldots,\hat{s}_K} = -\mathscr{H}_{m_1,\ldots,m_K} - \log|\det(\mathbf{W})| + \sum_{k=1}^{K} \mathscr{H}_{\hat{s}_k}. \tag{4}$$

The term $\mathscr{H}_{m_1,\ldots,m_K}$ is the joint entropy of the measurements. Since it is a constant for a given measurements set and is independent of $\mathbf{W}$, it can be ignored in the MI optimization process. Thus, optimizing Eq. (4) bypasses the problem involved in explicit estimation of the joint entropies.

It is desirable to do the same for the general convolutive case. However, if $\mathbf{W}$ is a convolutive operator, Eq. (4) does not hold. We note that similar expressions have been developed for convolutive mixtures (see for example [28]) assuming spatial whiteness. Nevertheless, algorithms based on this assumption, or equivalently using the MI *rate* [28,44] 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 Section 3.

An additional approach, related to MI, comprises the maximum likelihood/maximum *a posteriori* (ML/MAP) algorithms [12,20,32]. 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. Note that there are alternative independence criteria, such as higher order cumulants [26,31,33,40,45,48], which do not require entropy estimation. These are used in many existing BSS algorithms. However, these criteria sometimes fail, as shown in [6,38,39].

[2]This paper deals only with separation between distinct layers, and not with blind deconvolution of a single image.

3. Efficient separation of convolutive image mixtures

We may use Eq. (4) in convolutive mixtures, despite the fact that it is valid only in pointwise mixtures. This is achieved by decomposing the convolutive optimization problem into several smaller ones, which are apparently independent of each other. This approach is inspired by frequency domain algorithms developed for acoustic signals [11,19,25,27,36,42]. A transformation of the images is performed, such that the *convolution is expressed as a set of 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. Nevertheless, this approach has its own fundamental limitations, which are discussed and solved in Sections 4 and 5.

3.1. Factorization by STFT

Applying a Fourier transform to Eq. (1) yields

$$m_i(\vec{\omega}) = a_{i1}(\vec{\omega})s_1(\vec{\omega}) + \cdots + a_{iK}(\vec{\omega})s_K(\vec{\omega}), \quad i = 1,\ldots,K, \tag{5}$$

where $\vec{\omega} = (\omega_x, \omega_y)$ is the frequency index vector. The convolution in the spatial domain thus 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:

$$\hat{s}_k(\vec{\omega}) = w_{k1}(\vec{\omega})m_1(\vec{\omega}) + \cdots + w_{kK}(\vec{\omega})m_K(\vec{\omega}), \quad k = 1,\ldots,K. \tag{6}$$

Apparently, this is a simple problem that can exploit standard ICA tools, particularly Eq. (4). However, in the frequency domain, the Fourier transform yields only a *single sample* per frequency channel. Thus, no statistical ensemble is available for ICA. 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[3] on Eq. (1),

$$m_i(\vec{\omega},\mathbf{x}) \approx a_{i1}(\vec{\omega},\mathbf{x})s_1(\vec{\omega},\mathbf{x}) + \cdots + a_{iK}(\vec{\omega},\mathbf{x})s_K(\vec{\omega},\mathbf{x}), \quad i = 1,\ldots,K \tag{7}$$

rather than a Fourier transform. If the STFT window size is larger than the effective width of the blur kernel[4] (see for example [27]), then $a_{ik}(\vec{\omega},\mathbf{x}) \approx a_{ik}(\vec{\omega})$ and

[3]This operation is also termed as a *windowed Fourier transform*, which may be more appropriate for spatial coordinates as we use.

[4]A discussion regarding the STFT window width is given in Appendix C and in Section 9.

Eq. (7) becomes

$$m_i(\vec{\omega},\mathbf{x}) \approx a_{i1}(\vec{\omega})s_1(\vec{\omega},\mathbf{x}) + \cdots + a_{iK}(\vec{\omega})s_K(\vec{\omega},\mathbf{x}), \quad i=1,\ldots,K. \tag{8}$$

That is, there are sufficient samples for estimating image statistics. This is illustrated in Fig. 3. 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 $\mathscr{I}^{\vec{\omega}}(\{\hat{s}_k\}_{k=1}^K)$ and $\mathscr{H}^{\vec{\omega}}_{\hat{s}_k}$ as the MI and marginal entropies of the estimated sources at channel $\vec{\omega}$, respectively. Then, similarly to Eq. (4), the MI of the estimated sources at each channel is optimized by

$$\min_{\mathbf{w}(\vec{\omega})}\left\{-\log\left|\det[\mathbf{W}(\vec{\omega})]\right| + \sum_{k=1}^{K}\hat{\mathscr{H}}^{\vec{\omega}}_{\hat{s}_k}\right\}, \tag{9}$$

where $\hat{\mathscr{H}}^{\vec{\omega}}_{\hat{s}_k}$ is an estimator of the channel entropy of an estimated source. Hence, using this factorization, MI minimization of a convolutive mixture is expected to be performed while bypassing the numerical and complexity problems that stem from direct estimation of joint entropies.

Solving the separation problem for each frequency channel independently may appear simple. However, as we explain in Section 4 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.

3.2. Sparse separation in the STFT domain

Now, we exploit image statistics in order to enhance the efficiency of the optimization (9) in each frequency channel. At each channel, we need to estimate the marginal entropies of the sources. It is known from studies of image statistics (see for example [41]) that sub-band images are *sparse signals*, i.e., most of their pixel values reside near the origin. Hence, their probability density function (PDF) has a peaky shape, as illustrated in Fig. 4. We note that according to Ref. [51], sparsity of sources is a strong prior that can be exploited to achieve a very efficient separation.

A PDF model that is widely used in the literature to model sparse sub-band images is the generalized Laplacian (see for example [41])

$$p[\hat{s}_k(\vec{\omega})] = c(\rho)\exp[-|\hat{s}_k(\vec{\omega})|^{\rho}] \quad \text{where } 0<\rho<2. \tag{10}$$

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. Note that this PDF model[5] assumes that the source has a specific variance. This implicit normalization makes the optimization robust to some numerical pitfalls.[6]

We now exploit this prior of image statistics to simplify the estimation of entropies in our optimization. Entropy is defined as (see for example [10])

$$\hat{\mathscr{H}}^{\vec{\omega}}_{\hat{s}_k} = E\{-\log(p[\hat{s}_k(\vec{\omega})])\}, \tag{11}$$

where E is the expectation operator. Denote N as the number of pixels and n as the STFT shift (out of a total of N). Then, substituting Eq. (10) into Eq. (11) and replacing the expectation by empirical averaging, the estimated channel entropy is

$$\hat{\mathscr{H}}^{\vec{\omega}}_{\hat{s}_k} = \frac{1}{N}\sum_{n=1}^{N}|\hat{s}_k(\vec{\omega},n)|^{\rho} + C(\rho). \tag{12}$$

Here,

$$\hat{s}_k(\vec{\omega},n) = w_{k1}(\vec{\omega})m_1(\vec{\omega},n) + \cdots + w_{kK}(\vec{\omega})m_K(\vec{\omega},n), \quad k=1,\ldots,K \tag{13}$$

and $C(\rho)=\log[c(\rho)]$. 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 (12) of entropy as a function of the variables. Moreover, the computational complexity of this entropy estimator is $\mathcal{O}(N)$.

In principle, to achieve ultimate accuracy we may need to fit the PDF model (10) to the samples $\hat{s}_k(\vec{\omega})$, for estimating ρ, and then use this ρ in Eq. (12). However, we opt to *set* ρ in a way that makes the cost function convex. We can achieve convexity of the channel entropy (Eq. (12)) as a function of $\mathbf{W}(\vec{\omega})$ only if $\rho\geqslant 1$. However, the PDF of sub-band images is typically very sparse, i.e., $\rho<1$. The PDF representing the sparsest signal that yields a convex function in Eq. (12) corresponds to $\rho=1$, which is the value we choose to set (see also [3,30,51]). Substituting Eq. (12) in Eq. (9) yields the following MI

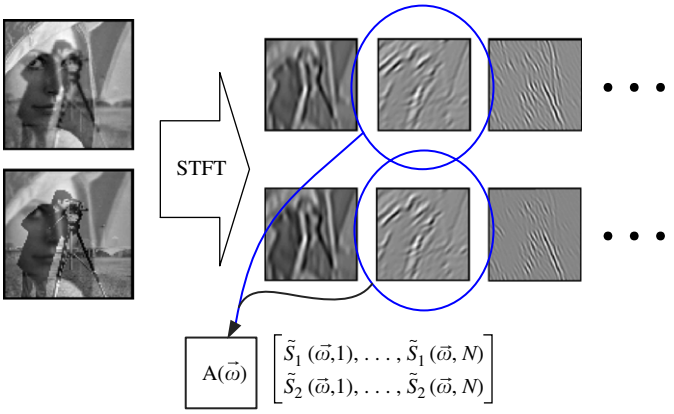

Fig. 3. Short time Fourier transform of the measured images. Each frequency channel contains a sub-band image of the size of the raw frame. Therefore, per band, the convolutive mixture problem is transformed to a pointwise problem.

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

[6]MI optimization can suffer from numerical instabilities due to a scale ambiguity, as explained in Appendix A.

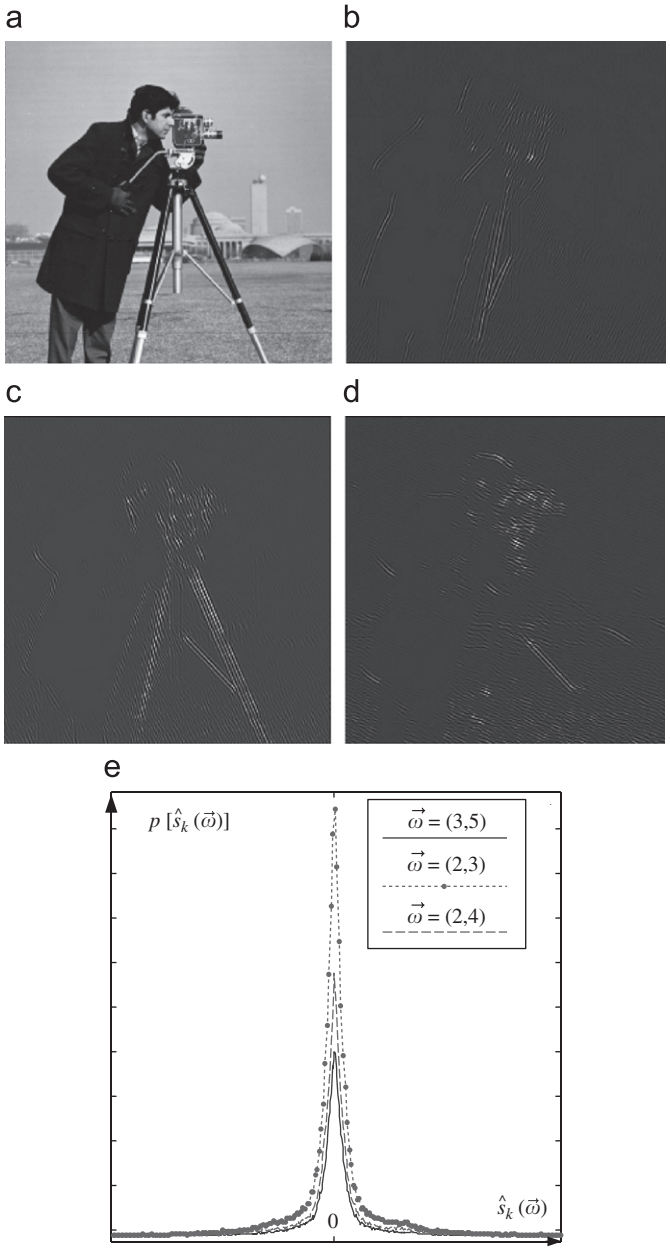

Fig. 4. Sub-band images and their statistics. The STFT window width is 11×11 pixels: (a) cameraman image; (b) STFT channel $\vec{\omega}=[2,3]$; (c) STFT channel $\vec{\omega}=[2,4]$; (d) STFT channel $\vec{\omega}=[3,5]$; (e) histogram of the STFT channels.

minimization formulation:

$$\min_{\mathbf{w}(\vec{\omega})}\left\{-\log|\det[\mathbf{W}(\vec{\omega})]|+(1/N)\sum_{k=1}^{K}\sum_{n=1}^{N}|\hat{s}_k(\vec{\omega},n)|\right\}, \qquad (14)$$

where $\hat{s}_k(\vec{\omega},n)$ is given in Eq. (13).

Eq. (14) 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 that is needed are simple summation of the absolute pixel values and a determinant of a small matrix. It is worth noting that the first term, $\log|\det[\mathbf{W}(\vec{\omega})]|$, is not necessarily convex in $\mathbf{W}$. However, practically the MI function (Eq. (14)) has a unique solution. Moreover, this term is very common in ICA optimizations [6,15,22,30] and practically does not affect the convexity. To minimize Eq. (14) we use the relative Newton optimization [50].

Eq. (14) 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 Section 4 and solve them in Section 5.

4. Inherent STFT problems

There are several inherent problems in the STFT factorization method. These problems can affect the separation quality. Some are ambiguities problems. They are shortly summarized here and thoroughly discussed in Appendix A. Another problem is non-uniform separation performance over channels, discussed subsequently.

There are two ambiguity problems. The permutation ambiguity implies that the separated sub-band images at each channel appear in a random permutation. Hence some sub-band images associated with the "first" estimated source may actually belong the "second" estimated source. As a result, when the channels are transformed back to the image domain using the inverse STFT, the reconstructed images can suffer from crosstalk, even though separation is achieved in each channel independently. See examples in Appendix A.

The scale ambiguity implies that the scale of each separated sub-band image at each STFT channel is arbitrary and unknown, leading to imbalance between frequency channels. As a result, 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 (see Appendix A).

4.1. Separation performance in different channels

The performance is frequency dependent, as illustrated in Fig. 5. Typically, there are a few frequency channels with good separation (Fig. 5(a)), a few channels with very poor separation (Fig. 5(c)) and the rest of the channels have mediocre separation quality (Fig. 5(b)). There are several reasons for this phenomenon, as discussed next.

There are several special channels for which the separation performance is expected to be poor. Suppose one of the sources in the scene has no energy in a certain channel. For example, without 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 describes the acquired sub-band images

$$\begin{aligned} m_1(\vec{\omega}^*,\mathbf{x}) &= a_{11}(\vec{\omega}^*)s_1(\vec{\omega}^*,\mathbf{x})+a_{12}(\vec{\omega}^*)s_2(\vec{\omega}^*,\mathbf{x}),\\ m_2(\vec{\omega}^*,\mathbf{x}) &= a_{21}(\vec{\omega}^*)s_1(\vec{\omega}^*,\mathbf{x})+a_{22}(\vec{\omega}^*)s_2(\vec{\omega}^*,\mathbf{x}) \end{aligned} \qquad (15)$$

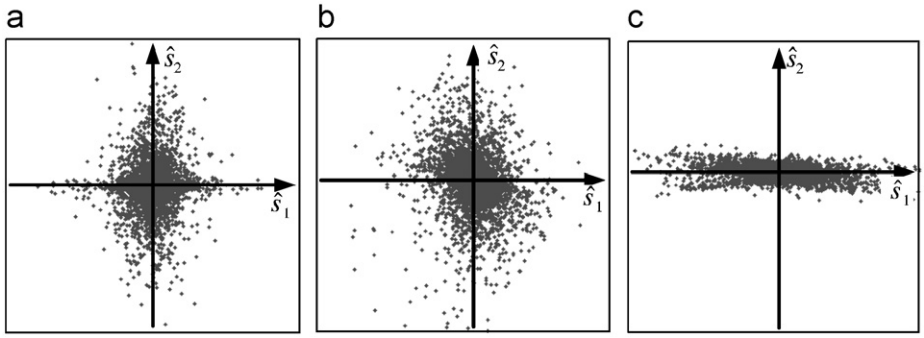

Fig. 5. Scatter plots of separation results in different channels: (a) good channel; (b) poor channel; (c) very poor channel.

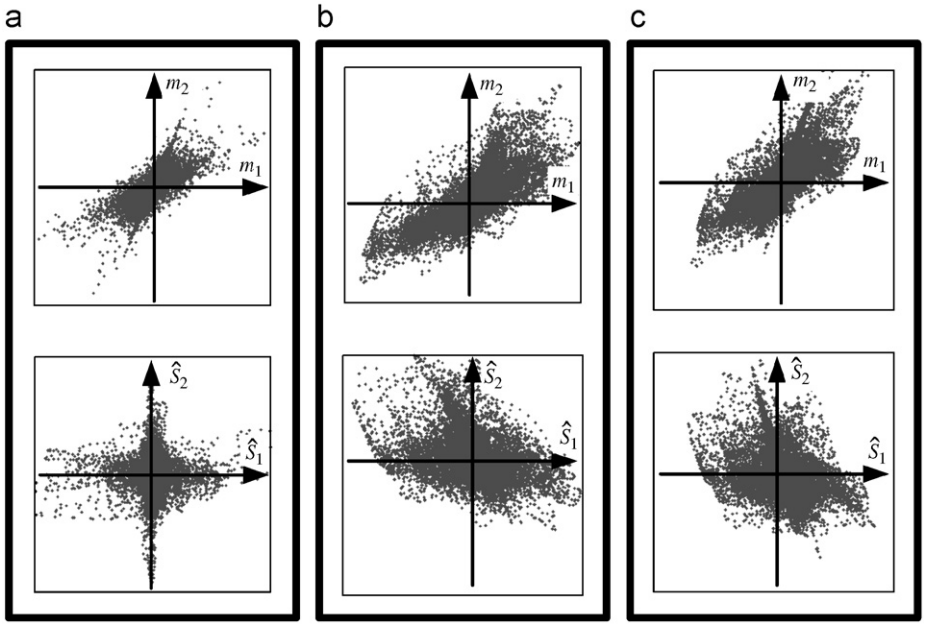

Fig. 6. Separation results for a fixed mixing matrix but different data: [Top] scatter plots of mixed signals; [Bottom] scatter plots of the estimated sources. The quality of the separation depends on the data: (a) good separation; (b) poor separation; (c) poor separation.

degenerates to

$$m_1(\vec{\omega}^*, \mathbf{x}) = a_{11}(\vec{\omega}^*)s_1(\vec{\omega}^*, \mathbf{x}),$$
$$m_2(\vec{\omega}^*, \mathbf{x}) = a_{21}(\vec{\omega}^*)s_1(\vec{\omega}^*, \mathbf{x}). \quad (16)$$

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.

In general, the quality of ICA optimization depends on the acquired data. This is illustrated in Fig. 6. 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 is extracted from different sub-band images. Fig. 6(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. 6(b) and (c) indicates poor separation (the PDF is not separable). In Section 5, we detail how to overcome this data dependency problem by using *a priori* knowledge about the blur kernel.

5. Inter-channel knowledge transfer

In Section 4.1, 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 Appendix A. In this section we bypass some of these problems by exploiting prior knowledge about the convolutive process.

5.1. Exploiting a parametric blur model

Blur caused by optical defocus can be parameterized [5,7,14,24,37,43]. As an example, consider a rough parametric model: a simple 2D Gaussian kernel with different widths in the x and y directions [34]. Denote $\vec{\xi}_{i,k} = [\xi_{i,k,x}, \xi_{i,k,y}]$ as the vector of the unknown blur parameters of the blur kernel of source k at image i and

$$G_{\vec{\xi}_{i,k}}(\vec{\omega}) = \exp[-\omega_x^2/(2\xi_{i,k,x}^2)]\exp[-\omega_y^2/(2\xi_{i,k,y}^2)] \tag{17}$$

as the filter which preserves light energy. In addition to defocus, let us incorporate attenuation $g_{i,k}$ of each source k into any mixture i.

Assume that in each acquired image, one of the layers is focused,[7] i.e., $G_{\vec{\xi}_{k,k}} = 1$. Define $\mathbf{A}(\vec{\omega})$ as the mixing operator in frequency channel $\vec{\omega}$:

$$\mathbf{A}(\vec{\omega}) = \begin{bmatrix} 1 & g_{1,2}G_{\vec{\xi}_{1,2}}(\vec{\omega}) & \cdots & \cdots \\ g_{2,1}G_{\vec{\xi}_{2,1}}(\vec{\omega}) & 1 & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \cdots & \cdots & g_{K,K-1}G_{\vec{\xi}_{K,K-1}}(\vec{\omega}) & 1 \end{bmatrix}. \tag{18}$$

Thus, the separation matrix in each channel is parameterized by $\vec{\xi}_{i,k}$ and $g_{i,k}$ and is of the form

$$\mathbf{W}(\vec{\omega}) = [\mathbf{A}(\vec{\omega})]^{-1}. \tag{19}$$

An example is given in Section 5.2. Note that the parameters $\vec{\xi}_{i,k}$ and $g_{i,k}$ are the *same for all frequency channels*. Hence, there is a smaller 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, as we explain is Section 5.3, that we use three channels, $\vec{\omega}^1, \vec{\omega}^2, \vec{\omega}^3$, that yield the best separation results according to a ranking criterion. Define $\tilde{\mathbf{A}}(\vec{\omega}^1) = [\mathbf{W}(\vec{\omega}^1)]^{-1}$, and similarly $\tilde{\mathbf{A}}(\vec{\omega}^2) = [\mathbf{W}(\vec{\omega}^2)]^{-1}$ and $\tilde{\mathbf{A}}(\vec{\omega}^3) = [\mathbf{W}(\vec{\omega}^3)]^{-1}$. Let $\tilde{a}_{i,k}$ be the elements of $\tilde{\mathbf{A}}$. Then, for each blur kernel, we calculate the unknown blur parameters $\vec{\xi}_{i,k}$ and $g_{i,k}$ by solving the following set of equations:

$$\begin{cases} g_{i,k}G_{\vec{\xi}_{i,k}}(\vec{\omega}^1) = \tilde{a}_{i,k}(\vec{\omega}^1)/\tilde{a}_{i,i}(\vec{\omega}^1), \\ g_{i,k}G_{\vec{\xi}_{i,k}}(\vec{\omega}^2) = \tilde{a}_{i,k}(\vec{\omega}^2)/\tilde{a}_{i,i}(\vec{\omega}^2), \\ g_{i,k}G_{\vec{\xi}_{i,k}}(\vec{\omega}^3) = \tilde{a}_{i,k}(\vec{\omega}^3)/\tilde{a}_{i,i}(\vec{\omega}^3). \end{cases} \tag{20}$$

We solve this set to find the parameters $\vec{\xi}_{i,k}$ and $g_{i,k}$, thus deriving the blur and attenuation parameters based on those few selected channels.

[7] We stress that we seek layer *separation* rather than *deblurring*. Therefore, if source k is defocused in all the images, we denote the least defocused version of source k as the effective source we aim to reconstruct. Then, we denote $G_{\vec{\xi}_{i,k}}(\vec{\omega})$ as the relative defocus filter between the effective source and the defocused source at image i.

Now, we can use these parameters and Eq. (17) to calculate $g_{i,k}G_{\vec{\xi}_{i,k}}(\vec{\omega})$ for all the frequency channels. This directly yields the separation operator $\mathbf{W}$ for *all the frequency channels*. We invert the mixing process by using this $\mathbf{W}$. Moreover, knowing the parameters of $\mathbf{W}$ allows us to invert the convolutive mixture 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 by representing each blur kernel using parametric models other than Gaussian, requiring more parameters. This would require selection of additional channels, 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.

5.2. Separation of semi-reflections

In common applications such as semi-reflections [34] or widefield optical sectioning [23], no attenuation accompanies the change of focus. Hence, $g_{i,k} = 1$ for all i,k. For each signal, each source is affected only by two parameters in the Gaussian model. Thus, only two channels are needed to solve for the unknown $\vec{\xi}_{i,k}$. Moreover, in the special case of semi-reflections, we have only two sources. Therefore, the mixing operator and the separation operator are reduced to

$$\mathbf{A}(\vec{\omega}) = \begin{bmatrix} 1 & G_{\vec{\xi}_{1,2}}(\vec{\omega}) \\ G_{\vec{\xi}_{2,1}}(\vec{\omega}) & 1 \end{bmatrix},$$

$$\mathbf{W}(\vec{\omega}) = \begin{bmatrix} 1 & -G_{\vec{\xi}_{1,2}}(\vec{\omega}) \\ -G_{\vec{\xi}_{2,1}}(\vec{\omega}) & 1 \end{bmatrix}\{\det(\mathbf{A}(\vec{\omega}))\}^{-1}. \tag{21}$$

The equation system we need to solve in order to estimate $\vec{\xi}_{1,2}$ and $\vec{\xi}_{2,1}$ is

$$\begin{cases} -G_{\vec{\xi}_{1,2}}(\vec{\omega}^1) = w_{1,2}(\vec{\omega}^1)/w_{1,1}(\vec{\omega}^1), \\ -G_{\vec{\xi}_{1,2}}(\vec{\omega}^2) = w_{1,2}(\vec{\omega}^2)/w_{1,1}(\vec{\omega}^2), \\ -G_{\vec{\xi}_{2,1}}(\vec{\omega}^1) = w_{2,1}(\vec{\omega}^1)/w_{2,2}(\vec{\omega}^1), \\ -G_{\vec{\xi}_{2,1}}(\vec{\omega}^2) = w_{2,1}(\vec{\omega}^2)/w_{2,2}(\vec{\omega}^2). \end{cases} \tag{22}$$

Here, $w_{i,k}$ are the coefficients of matrix $\mathbf{W}(\vec{\omega})$ and $\vec{\omega}^1, \vec{\omega}^2$ are the best and second best channels according to the ranking we describe next. Then, the reconstruction of the separated layer is done in the frequency domain, which yields a better reconstruction quality than in the STFT domain. First, the acquired images and the reconstructed blurring kernels are transformed to the Fourier domain (Eq. (5)). Then, we solve Eq. (5) as

$$\hat{s}_1(\vec{\omega}) = \frac{m_1(\vec{\omega}) - a_{12}(\vec{\omega})m_2(\vec{\omega})}{\max[(1 - a_{12}(\vec{\omega})a_{21}(\vec{\omega})), \tau]}, \tag{23}$$

$$\hat{s}_2(\vec{\omega}) = \frac{-a_{21}(\vec{\omega})m_1(\vec{\omega}) + m_2(\vec{\omega})}{\max[(1 - a_{12}(\vec{\omega})a_{21}(\vec{\omega})), \tau]}, \quad (24)$$

where τ is a threshold. Note that in low frequency channels, $a_{12}(\vec{\omega})a_{21}(\vec{\omega}) \to 1$. Hence the denominator in Eqs. (23) and (24) is close to zero, i.e., the mixing is not invertible, as explained in Appendix B. Therefore, we use a threshold $\tau > 0$ in order to regularize this problem. Its value is user-defined: as it increases, the low frequency content and its noise decrease. Following Eqs. (23) and (24), we transform the reconstructed images from the Fourier domain to the image domain.

We stress the main benefit of this approach of parameter-based inter-channel knowledge transfer: the blur parameterization solves uniquely the permutation, scale and sign ambiguities of each individual channel. The sources' frequency channels 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.

5.3. Selecting a good channel

The parameter estimation method of Section 5.1 is based on the "best" frequency channels. This requires *ranking* of the channels. The ranking relies on a quality criterion for the separation (i.e., independence) of $\{\hat{s}_k\}$ at each channel $\vec{\omega}$ (see also [8,18]). The MI of these estimated sources is inconvenient: as seen in Eq. (4), it involves joint entropy estimation, which may not be reliable, given the small number of independent measurements at each band.

Thus, we opt for simpler ranking criteria. Recall that we exploit the sparsity of the sources. A scatter plot of two sparse independent signals has a cross shape. Specifically, in the $(\hat{s}_1, \hat{s}_2)$ plane, most of the samples should have small angles relative to the $\hat{s}_1$ and $\hat{s}_2$ axes. Therefore, we may define a function that penalizes for a non-cross shaped plot. A simple criterion is based on the $\mathscr{L}_1$ norm of the estimated signals, normalized by their $\mathscr{L}_2$ norm,

$$\chi^{\vec{\omega}}_{\mathscr{L}_1} = \sum_{k=1}^{2} \frac{\sum_{n=1}^{N} |\hat{s}_k(\vec{\omega}, n)|}{\sqrt{\sum_{n=1}^{N} [\hat{s}_k(\vec{\omega}, n)]^2}}. \quad (25)$$

This criterion increases as the samples in the scatter plot deviate from the $\hat{s}_1$ and $\hat{s}_2$ axes, and is reduced when each sample n has non-zero values exclusively in $\hat{s}_1$ or $\hat{s}_2$.

We have also considered a criterion $\chi^{\vec{\omega}}_{\text{angles}}$ that is explicitly based on estimation of the angles of the samples in the $(\hat{s}_1, \hat{s}_2)$ plane. For a cross shape, the angles should densely cluster around $\psi = 0^\circ, 90^\circ, 180^\circ$ and 270°. Hence, we define a function that penalizes for sample angles that are distant from these values of ψ. In this calculation, we ignore samples that are very close the origin of the $(\hat{s}_1, \hat{s}_2)$ plane, since their angles are relatively noisy. For a sample n, its squared Euclidian distance from the origin of the $(\hat{s}_1, \hat{s}_2)$ plane is $d^2(\vec{\omega}, n) \equiv [|\hat{s}_1(\vec{\omega}, n)|^2 + |\hat{s}_2(\vec{\omega}, n)|^2]$. Ignoring samples by the origin, the set of samples used in $\chi^{\vec{\omega}}_{\text{angles}}$ is

$$\Omega^{\vec{\omega}} = \{n : d^2(\vec{\omega}, n) > \zeta^2\}, \quad (26)$$

where ζ is a threshold. For each sample $n \in \Omega^{\vec{\omega}}$, let its angle in the $(\hat{s}_1, \hat{s}_2)$ plane be φ_n, relative to the $\hat{s}_1$ axis.

Define sets of samples associated with each of the four angles of the cross shape:

$$\Psi^{\vec{\omega}}_{\psi} = \{n : n \in \Omega^{\vec{\omega}} \text{ and } \varphi_n \in [\psi - 45^\circ, \psi + 45^\circ)\}. \quad (27)$$

Projecting the samples to the unit circle in the $(\hat{s}_1, \hat{s}_2)$ plane, the mean vector corresponding to each set $\Psi^{\vec{\omega}}_{\psi}$ is $\mathbf{v}^{\psi}(\vec{\omega}) \equiv [v^{\psi}_1(\vec{\omega}), v^{\psi}_2(\vec{\omega})]$, where

$$v^{\psi}_k(\vec{\omega}) = \sum_{n \in \Psi^{\vec{\omega}}_{\psi}} \frac{\hat{s}_k(\vec{\omega}, n)}{\mathrm{d}(\vec{\omega}, n)}. \quad (28)$$

The ideal angles ψ correspond to vectors $\mathbf{r}^{0^\circ} = [1, 0]$, $\mathbf{r}^{90^\circ} = [0, 1]$, $\mathbf{r}^{180^\circ} = [-1, 0]$ and $\mathbf{r}^{270^\circ} = [0, -1]$ on the unit circle in the $(\hat{s}_1, \hat{s}_2)$ plane. For sparse and independent signals, we expect $\mathbf{v}^{\psi}(\vec{\omega})$ to be close to $\mathbf{r}^{\psi}$, for each ψ. Hence,

$$\chi^{\vec{\omega}}_{\text{angles}} = \sum_{\psi=0^\circ}^{270^\circ} \|\mathbf{v}^{\psi}(\vec{\omega}) - \mathbf{r}^{\psi}\|^2. \quad (29)$$

Both $\chi^{\vec{\omega}}_{\mathscr{L}_1}$ and $\chi^{\vec{\omega}}_{\text{angles}}$ yielded consistent rankings of channels in our tests. Moreover, both criteria depend only on the sources estimated by ICA (in each frequency channel). These expressions determine which frequency channels yield the most separated sources. Thus, in our algorithm, we first perform ICA in all the frequency channels. We then calculate $\chi^{\vec{\omega}}_{\mathscr{L}_1}$ or $\chi^{\vec{\omega}}_{\text{angles}}$, thus ranking the channels. Then, we select the best channels as those that correspond to the smallest penalty function values. These channels are used in Section 5.1.

6. Algorithm summary

In this section, we summarize 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 Section 9).
2. To enable factorization, apply STFT on the input images.
3. For each frequency (sub-band) channel:
   - Find $\mathbf{W}(\vec{\omega})$ that minimizes $\hat{\mathscr{I}}^{\vec{\omega}}$ as in Eq. (14), say by the relative Newton method.
   - Calculate the ranking criterion $\chi^{\vec{\omega}}_{\mathscr{L}_1}$ or $\chi^{\vec{\omega}}_{\text{angles}}$.
4. Following step 3, select the channels that correspond to the lowest values of the ranking criterion. 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.

7. Direct BSS using a parametric blur model

Our algorithm (Section 6) has become more elaborate than a trick of factorization. Some overhead is caused by channels for which standard ICA tools yield an unreliable **W**. Thus, **W** is estimated based on other channels, where the estimation is reliable. This inter-channel knowledge transfer exploits a parametric blur model **p**. One may suggest a shortcut: estimating the PSF parameters directly, without ICA tools which optimize over **W**. In this section we discuss this alternative.

To recap, recall that for each frequency channel, we minimize Eq. (14) explicitly over **W**. Based on the estimated **W** in some channels, we derive the low dimensional vector of blur parameters **p**. Let us consider an alternative approach that skips the intermediate estimation of **W**. It would directly optimize over **p**:

$$\min_{\mathbf{p}} \hat{\mathscr{J}}^{\vec{\omega}}_{\hat{s}_1,\ldots,\hat{s}_K} = \min_{\mathbf{p}} \left\{ \sum_{k=1}^{K} \frac{1}{N} \sum_{n=1}^{N} |\hat{s}_k(\vec{\omega},n)| - \log|\det[\mathbf{W}(\vec{\omega},\mathbf{p})]| \right\}. \quad (30)$$

Hence, the elements of **W** would not be optimized directly in any channel, but be derived from the explicit optimization variables **p**, in contrast to Eq. (14).

Such an approach is problematic. The first term in Eq. (30) is a convex function in **W** but it is typically *not* a convex function in **p**. Even for simple models such as the Gaussian model (Eq. (17)), the sum over the absolute value of the samples is not a convex function in $\vec{\xi}_{i,k}$. Moreover, it is typically *multimodal*. This can lead to many local minima in the MI function. 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.

The second term in Eq. (30) is generally not convex in **p** nor **W**. However, from our experience and the experiences reported in [6,15,22,30] this term is practically convex in **W**, yielding a unique global solution to which convergence can be fast. For this reason, in contrast to Eq. (30), we do not optimize over the blur parameters **p**. Rather, we optimize over the separation matrix $\mathbf{W}(\vec{\omega})$ by Eq. (14), which is a practically convex optimization problem in each frequency channel.

8. Demonstrations

In this section we present results of a simulations and experiments we made. For the first simulation, we used two standard pictures (Fig. 7(a)) as the two scene layers. The blur kernel we used is a Gaussian with standard deviation 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 to the mixed images with a standard deviation of ~2.5 gray levels. The mixed and noisy images are presented in Fig. 7(b). The separation optimization was performed using 15×15 STFT channels. The reconstruction was done in the Fourier domain, and using $\tau \sim 0.1$.

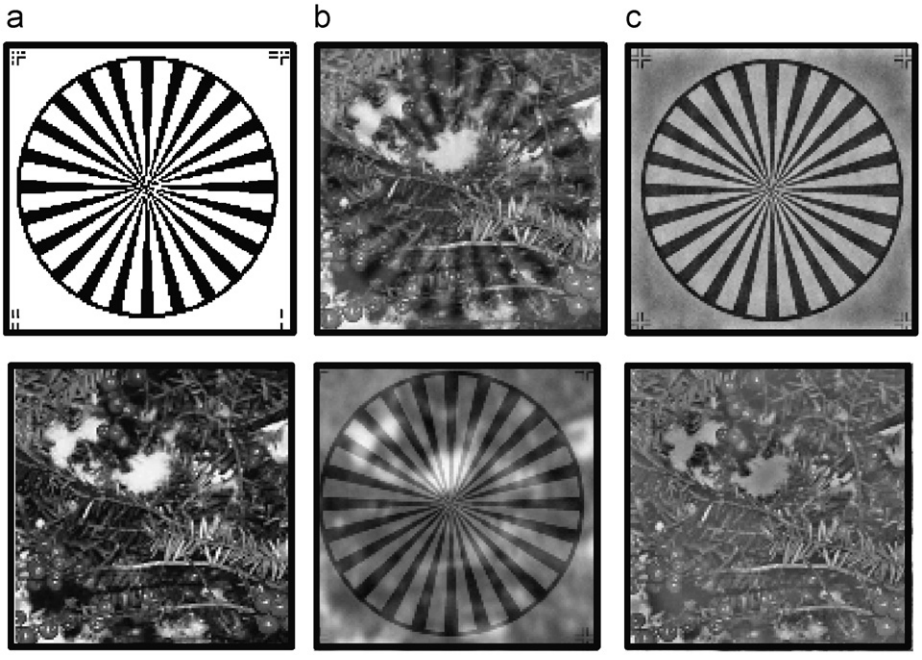

Fig. 7. Separation results: (a) original sources; (b) convolved and mixed sources; (c) reconstructed sources.

The separation results are presented in Fig. 7(c). The separation quality is excellent. There is no visible crosstalk between the images. However, as seen in Fig. 7(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 [34], and described in Appendix B: we cannot expect good separation and recovery in low spatial frequencies, even if the convolution kernels are perfectly derived. Therefore, we use regularization in the low frequencies in order to avoid instabilities.

In a second simulation, we used two natural images of size 122×162 pixels (Fig. 8a) as the two scene layers. The blur kernels we used are Gaussians with parameter vectors $\vec{\xi}_1 = [1, 2]$ and $\vec{\xi}_2 = [2, 1]$ pixels. In addition, we added an i.i.d. Gaussian noise with a standard deviation of $\sim$2.5 gray levels to the mixed images. The mixed, noisy images are presented in Fig. 8b. The separation results are presented in Fig. 8c. The separation quality appears good. There is no visible crosstalk between the images.

We have also quantitatively assessed the crosstalk reduction. For a fair quantitative comparison, we high-pass filtered the original sources and the mixtures, with the same filter used in the displayed reconstructed layers. The correlation coefficient of the mixed images is 0.96, but it is reduced to 0.4 by the algorithm. Apparently, we may conclude that the process successfully reduces crosstalk (here measured by correlation). However, this result should be viewed with caution: the resulting correlation is in fact lower than the correlation coefficient of the original sources, which is 0.6. We believe that one cause for this apparent hyper-performance lies in the fact that substantial crosstalk exists in lower frequencies, which are attenuated by the filters. Hence, change in the lower frequency components, which is an inevitable consequence of the recovery problem, disrupts quantitative assessment.

In the real experiments, we photographed semi-reflections using a digital camera having a linear radiometric response, Nikon D-100. In Fig. 9a an outdoor scene is reflected from the glass cover of a painting. We acquired two raw frames, where either scene layer is focused while the other is defocused. As mentioned in Appendix B, we cannot expect good reconstruction in the low frequencies. Therefore, for comparison, a high-pass version of the raw frames is given in Fig. 9b. These versions still exhibit significant crosstalk. The separation results are given in Fig. 9c. The crosstalk between the resulting layers is much reduced. Another experiment is presented in Fig. 10 where a shelf in our lab is reflected from the glass cover of another painting.

9. Discussion

We have presented an algorithm for separation of convolutive image mixtures. The algorithm exploits image statistics and *a priori* knowledge about the physical blur process. 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. We exploit image statistics and *a priori* knowledge about the physical blur process. Moreover, a parametric model for the blur kernel solves scale, sign and permutation ambiguities typical of ICA and STFT problems.

The computational complexity of the separation optimization is linear in the number of image layers K, the number of image pixels N and the number of STFT channels. In addition, the computational complexity of solving the parametric model (Section 5.1) is linear in the number of parameters of the blur model and proportional to $K(K-1)$. For example, in the case of semi-reflections, we use a model having two parameters, two image layers and the solution is based on a set of four equations (Eq. (22)).

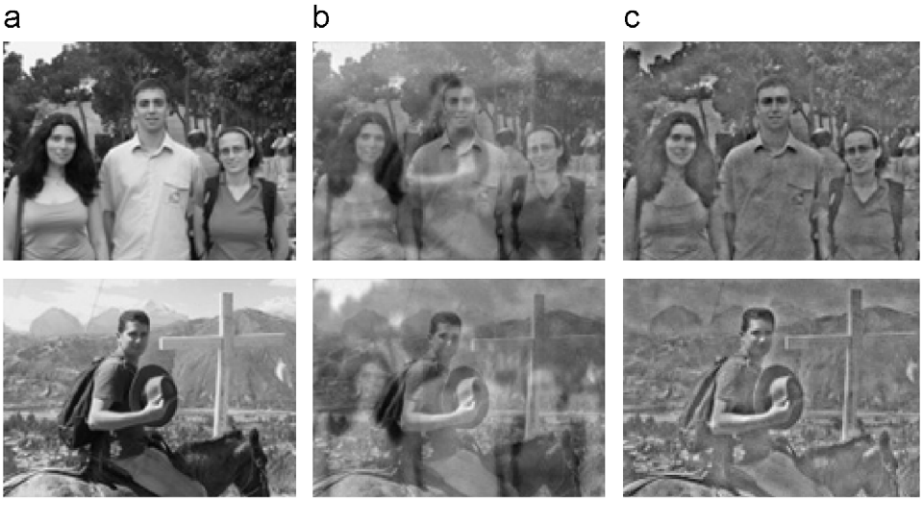

Fig. 8. Simulation results: (a) two original natural images; (b) the two convolved and mixed images; (c) reconstructed layers.

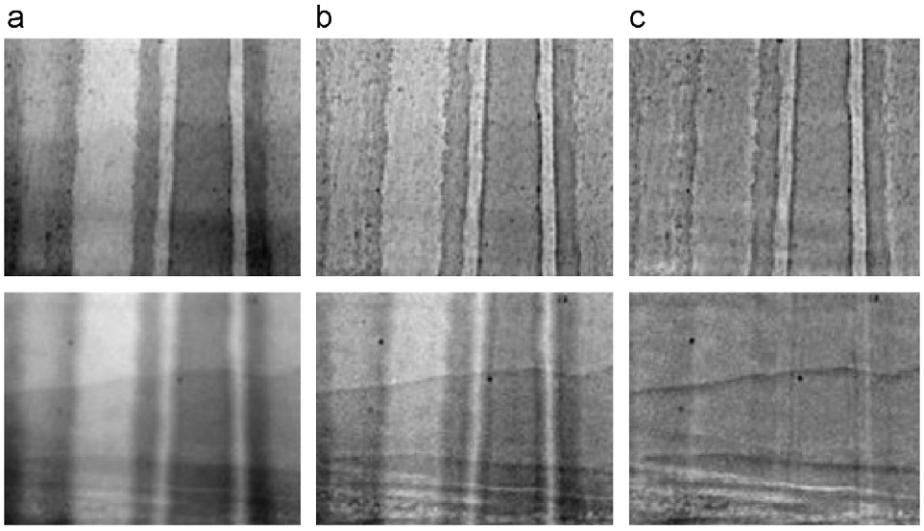

Fig. 9. Experimental results: (a) the acquired raw frames exhibit a convolutive mixture; (b) a high-pass filtered version of the raw frames; (c) reconstructed layers.

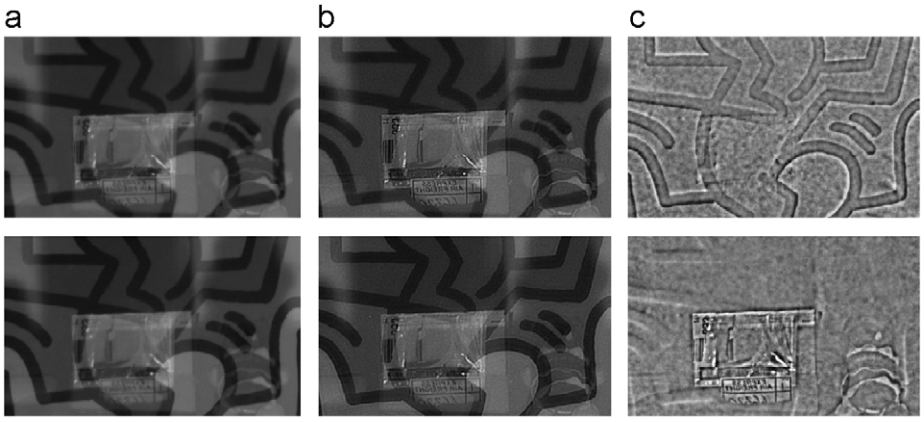

Fig. 10. Experimental results: (a) the acquired raw frames exhibit a convolutive mixture; (b) a high-pass filtered version of the raw frames; (c) reconstructed layers.

The convolutive image separation algorithm has a single parameter to tweak: the width of the STFT window. It can affect the separation results. As mentioned in Section 3.1, it must be wider than the effective width of the blur kernel. On the other hand, a very wide window can degrade the sparsity of the of the sub-band images.[8] We determine the STFT window width by trial and error (see Appendix C), but we believe this can be automated. For example, one may use multi-window STFT and choose the best window/channels using the criteria described in Section 5.3. This requires further research.

[8]If the STFT windows are too large, then most of them include spatial edges, which have a very wide spatial spectrum. Therefore, the STFT coefficients in all the frequency channels include large values. Since this phenomenon happens in most of the windows across the frame, $\hat{s}_k(\vec{\omega}, n)$ will not be close enough to zero for most of the samples. This contradicts the sparsity assumption. Hence, there is an optimal STFT window width for each data set.

In addition, we used source independence as our separation criterion. However, in applications such as optical sectioning microscopy, the source layers may have significant statistical dependency. Thus, an interesting extension to this research is to adapt this algorithm to deal with mutually dependent signals, and apply it to optical sectioning microscopy. Another possible direction is to consider nonlinear deconvolution methods, which are capable of overcoming some of the ill-conditioning described in Appendix B. This would be important in processing of color images, since color is highly affected by low frequency image components. Another possible extension is to use the principles of this paper with other types of PSFs, such as translation or blur due to motion [2]. In these cases, the PSF can often be parameterized. Hence, with the proper parameterization, the approach presented in this paper may have a general capacity.

Acknowledgments

This research has been supported by the "Dvorah" Fund of the Technion and by the Israeli Ministry of Science, Culture & Sports (Grant 3-3426). The research was carried out in the Ollendorff Minerva Center. Minerva is funded through the BMBF. Yoav Schechner is a Landau Fellow-supported by the Taub Foundation, and an Alon Fellow.

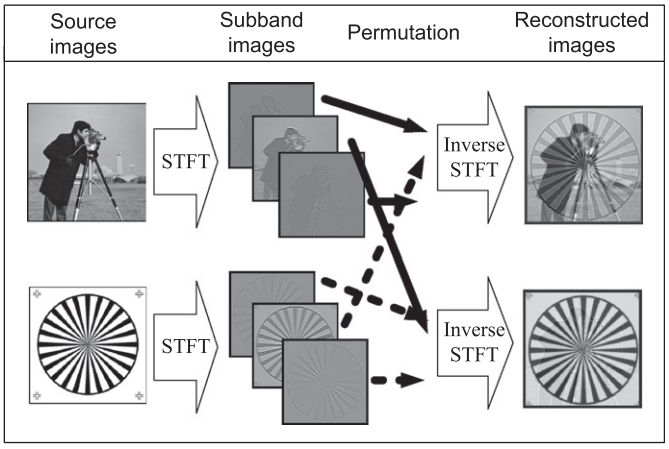

Fig. 11. The effect of permutation ambiguity in reconstruction from sub-band images. The reconstructed images exhibit crosstalk, although separation exists in each of the channels.

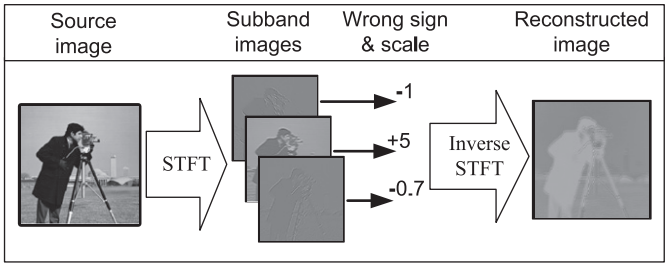

Fig. 12. The effect of scale ambiguity in reconstruction from sub-band images. The reconstructed image appears unnatural, due to an unbalanced spectrum and opposite sign in some of the channels.

Appendix A. Ambiguities

Optimization of MI involves permutation and scale ambiguities. We now detail these ambiguities and their influence on the separation performance. For simplicity, we demonstrate these ambiguities on an example of two layers. Nevertheless, the same explanation holds for the general case of K layers.

A.1. Permutation ambiguity

Minimization of MI results in sources that appear in an arbitrary order, since, for example, $\mathcal{I}_{\hat{s}_1,\hat{s}_2} = \mathcal{I}_{\hat{s}_2,\hat{s}_1}$. This permutation ambiguity implies that the separated sub-band images at each channel appear in a random permutation. Hence some sub-band images associated with the "first" estimated source may actually belong the "second" estimated source. As illustrated in Fig. 11, when the channels are transformed back to the image domain using the 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, frequency decomposition leads to an overall poor result. However, we counter this problem by constraining the solution to abide to a parametric form of the mixture, as detailed in Section 5.

A.2. Scale ambiguity

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

$$p_{\hat{s}_1,\hat{s}_2}(\hat{s}_1,\hat{s}_2) = p_{\hat{s}_1}(\hat{s}_1)p_{\hat{s}_2}(\hat{s}_2). \tag{31}$$

Therefore, the MI of $\hat{s}_1, \hat{s}_2$ is $\mathcal{I}_{\hat{s}_1,\hat{s}_2} = 0$ (see for example [10]). 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 $p_{\bar{s}_1,\bar{s}_2}(\bar{s}_1,\bar{s}_2) = p_{\bar{s}_1}(\bar{s}_1)p_{\bar{s}_2}(\bar{s}_2)$. 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 scale (which can also be negative) of values, per sub-band channel.

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

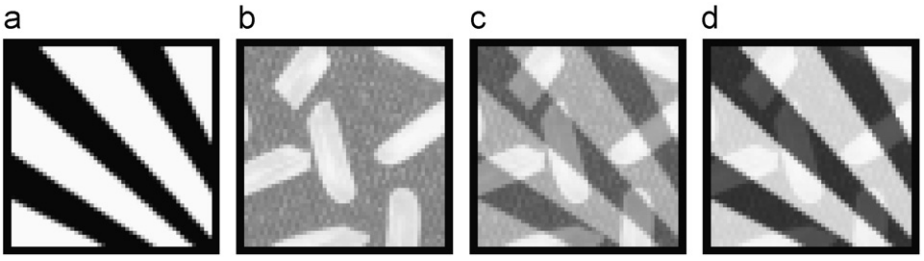

Fig. 13. Example scene: (a), (b) image layers; (c), (d) two image mixtures.

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 Section 3: the pre-set variance of the PDF model constrains the resulting sources to have fixed 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. 12, when the estimated channels of a source are transformed back to the

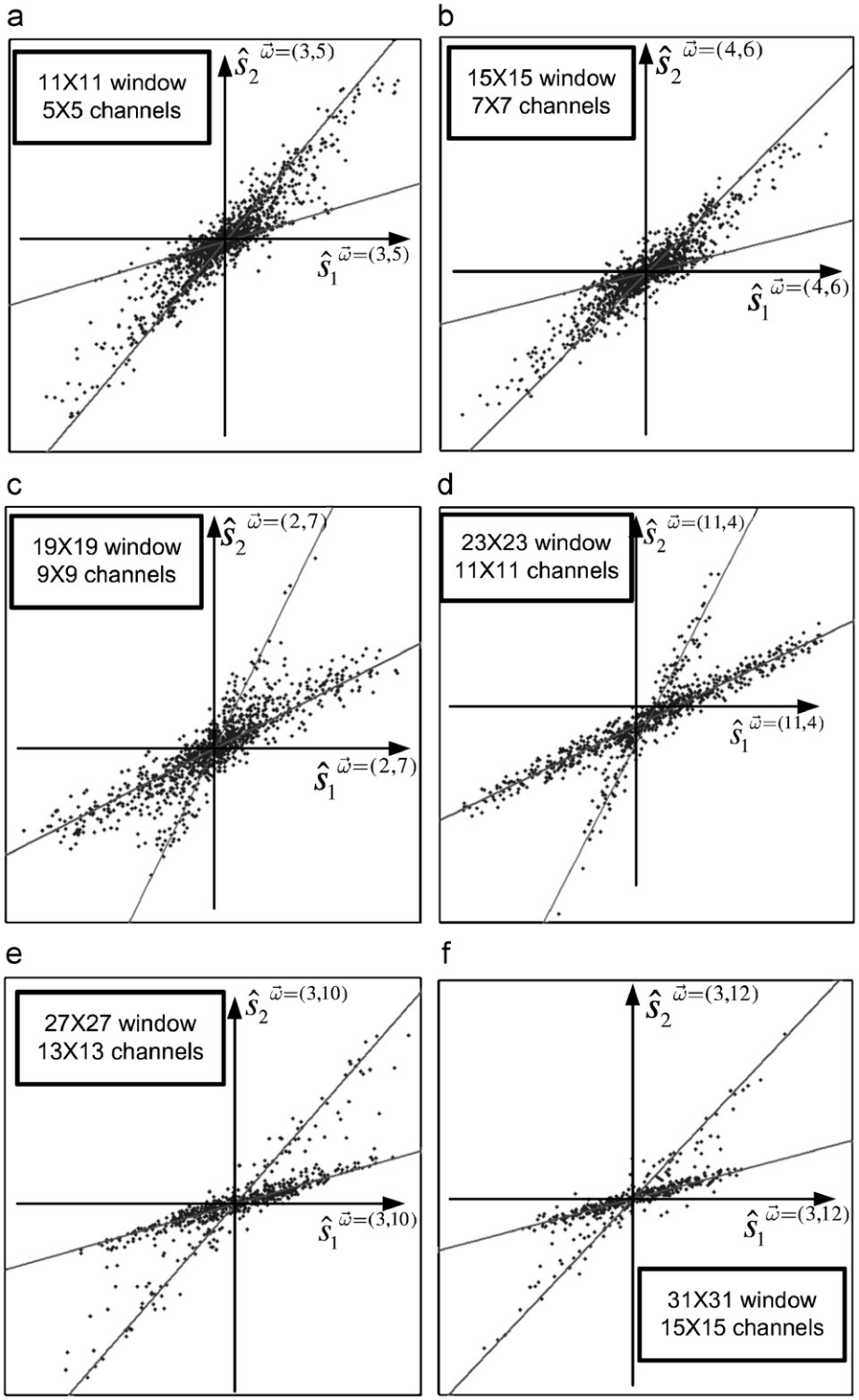

Fig. 14. Scatter plots of the STFT channels with different STFT window width. The red lines correspond to the mixing matrix rows and indicate the expected location of the scattered points: (a) 11 × 11 pixels window, 5 × 5 channels; (b) 15 × 15 pixels window, 7 × 7 channels; (c) 19 × 19 pixels window, 9 × 9 channels; (d) 23 × 23 pixels window, 11 × 11 channels; (e) 27 × 27 pixels window, 13 × 13 channels; (f) 31 × 31 pixels window, 15 × 15 channels.

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. This problem is solved by the method described in Section 5, since it makes the inter-channel scale consistent with a parametric model.

Appendix B. A fundamental inversion limitation

Irrespective of the *blind* aspect of the estimation, and thus of any BSS method, there is a fundamental limitation in separating a mixture of images convolved by defocus blur. It exists even if the operators **A** and **W** are perfectly known, skipping the need for blindly estimating them. This limitation exists in the low frequency bands.

Details are given in Refs. [23,34]. Here, we wish to convey an insight, by looking at the special case of two layers (sources), where either one is focused while the other is blurred by a symmetric defocus kernel. In this case, Eq. (1) can be written as

$$\begin{aligned} m_1(\mathbf{x}) &= s_1(\mathbf{x}) + (a_{12} * s_2)(\mathbf{x}), \\ m_2(\mathbf{x}) &= (a_{21} * s_1)(\mathbf{x}) + s_2(\mathbf{x}). \end{aligned} \tag{32}$$

In the frequency domain, Eq. (32) becomes

$$\begin{aligned} m_1(\vec{\omega}) &= s_1(\vec{\omega}) + a_{12}(\vec{\omega}) s_2(\vec{\omega}), \\ m_2(\vec{\omega}) &= a_{21}(\vec{\omega}) s_1(\vec{\omega}) + s_2(\vec{\omega}). \end{aligned} \tag{33}$$

Since a_{12} and a_{21} are symmetric defocus kernels, $\mathrm{Im}(a_{12}) = \mathrm{Im}(a_{21}) = 0$. Therefore,

$$\begin{aligned} \mathrm{Re}[m_1(\vec{\omega})] &= \mathrm{Re}[s_1(\vec{\omega})] + a_{12}(\vec{\omega})\mathrm{Re}[s_2(\vec{\omega})], \\ \mathrm{Re}[m_2(\vec{\omega})] &= a_{21}(\vec{\omega})\mathrm{Re}[s_1(\vec{\omega})] + \mathrm{Re}[s_2(\vec{\omega})]. \end{aligned} \tag{34}$$

Consider $\mathrm{Re}[s_2(\vec{\omega})]$ as a function of $\mathrm{Re}[s_1(\vec{\omega})]$, for a specific $\vec{\omega}$. In this case Eq. (34) can be visualized as a pair of straight lines in the $\{\mathrm{Re}[s_1(\vec{\omega}), \mathrm{Re}[s_2(\vec{\omega})]\}$ plane. The solution corresponds to the intersection of these lines. This intersection exists uniquely for $a_{21}(\vec{\omega})a_{21}(\vec{\omega}) \neq 1$. As the frequency decreases ($\vec{\omega} \to 0$), so $a_{12}, a_{21} \to 1$, since this is the nature of defocus blur. Hence at low frequencies the recovery is ill conditioned. Moreover, due to energy conservation, the average gray level (at $\vec{\omega} = 0$) is not affected by defocusing. Thus, $a_{21}(\vec{\omega}) = a_{21}(\vec{\omega}) = 1$ at $\vec{\omega} = 0$. This makes Eq. (34) ill posed, since there is no unique solution for $\{\mathrm{Re}[s_1(\vec{\omega}), \mathrm{Re}[s_2(\vec{\omega})]\}$.

Appendix C. Window size

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 3.1, the STFT window should be wider than the effective width of the PSF of the imaging systems. Otherwise, Eq. (8) is invalid.

Another consideration for the window width is the acquired scene. As mentioned in Section 3.2, we assume that each of the sub-band images has a sparse PDF. Otherwise, the generalized Laplacian model does not hold. However, a very wide window can degrade the sparsity of the of the sub-band images. If the STFT windows are too large, then most of them include spatial edges, which have a very wide spatial spectrum. Therefore, the STFT coefficients in all the frequency channels include large values. Since this phenomenon happens in most of the windows across the frame, $\hat{s}_k(\vec{\omega}, n)$ will not be close enough to zero for most of the samples. This weakens the sparsity assumption.

To illustrate this phenomenon, consider for example the scenario in Fig. 13. It contains two pointwise mixtures of images. The blur kernel is a delta function, and therefore the mixture in all the STFT channels is identical.[9] The mixed pictures were transformed to the short time Fourier domain using several window widths. Scatter plots of representative channels in various window widths are presented in Fig. 14. In this case, 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 wide results in a clear "X" shape, which indicates mixed sources having a sparse PDF.

References

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

[2] E. Be'ery, A. Yeredor, Blind separation of reflections with relative spatial shifts, in: Proceedings of the ICASSP, 2006, pp. 625–628.

[3] P. Bofill, M. Zibulevsky, Underdetermined blind source separation using sparse representations, Signal Process. 81 (11) (2001) 2353–2362.

[4] M. Borga, H. Knutsson, Estimating multiple depths in semi-transparent stereo images, in: Proceedings of the Scandinavian Conference on Image Analysis, 1999, pp. 127–134.

[5] M. Born, E. Wolf, Principles of Optics, Pergamon, Oxford, 1975.

[6] R. Boscolo, H. Pan, V.P. Roychowdhury, Non-parametric ICA, in: Proceedings of the ICA2001, 2001, pp. 13–18.

[7] J. Braat, P. Dirksen, A.J.E.M. Janssen, Assessment of an extended Nijboer–Zernike approach for the computation of optical point-spread functions, J. Opt. Soc. Am. A 19 (5) (2002) 858–870.

[8] A. Bronstein, M. Bronstein, M. Zibulevsky, Y.Y. Zeevi, Separation of semireflective layers using sparse ICA, in: Proceedings of the IEEE ICIP, 2003, pp. 313–316.

[9] M. Castella, J.-C. Pesquet, An iterative blind source separation method for convolutive mixtures of images, in: Proceedings of the ICA2004, 2004, pp. 922–929.

[10] T.M. Cover, J.A. Thomas, Elements of Information Theory, Wiley, New York, 1991.

[11] A. Dapena, C. Serviere, L. Castedo, Separation of convolutive mixtures in the frequency-domain using only two frequency bins, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 2, 2002, pp. 1633–1636.

[12] S.C. Douglas, X. Sun, Blind separation of acoustical mixtures without time-domain deconvolution or decorrelation, in: Proceedings

[9] This implies that there is no meaningful lower bound on the width of the STFT window.

of the IEEE International Conference on Neural Networks for Signal Processing, 2001, pp. 323–332.
[13] H. Farid, E.H. Adelson, Separating reflections from images by use of independent components analysis, J. Opt. Soc. Am. A 16 (9) (1999) 2136–2145.
[14] E. Hecht, Optics, Addison-Wesley, Reading, MA, 1987.
[15] A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis, Wiley, New York, 2001.
[16] M. Irani, B. Rousso, S. Peleg, Computing occluding and transparent motions, Int. J. Comput. Vision 12 (1) (1994) 5–16.
[17] W. Kasprzak, A. Okazaki, Blind deconvolution of timely-correlated sources by homomorphic filtering in Fourier space, in: Proceedings of the ICA2003, 2003, pp. 1029–1034.
[18] P. Kisilev, M. Zibulevsky, Y.Y. Zeevi, Multiscale framework for blind source separation, J. Mach. Learn. Res. 4 (7–8) (2004) 1339–1363.
[19] I. Kopriva, Z. Devcic, H. Szu, An adaptive short-time frequency domain algorithm for blind separation of nonstationary convolved mixtures, in: Proceedings of the IEEE International Conference on Neural Networks, vol. 1, 2001, pp. 424–429.
[20] T.-W. Lee, A.J. Bell, R. Orglmeister, Blind source separation of real world signals, in: Proceedings of the IEEE International Conference on Neural Networks, vol. 4, 1997, pp. 2129–2134.
[21] A. Levin, A. Zomet, Y. Weiss, Separating reflections from a single image using local features, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, vol. 1, 2004, pp. 306–313.
[22] Y. Lomnitz, A. Yeredor, A blind-ML scheme for blind source separation, in: Proceedings of the IEEE Workshop on Statistical Signal Processing, 2003, pp. 581–584.
[23] F. Macias-Garza, A.C. Bovik, K.R. Diller, S.J. Aggarwal, J.K. Aggarwal, The missing cone problem and low-pass distortion in optical serial sectioning microscopy, in: Proceedings of the IEEE Acoustics, Speech and Signal Processing, vol. 2, 1998, pp. 890–893.
[24] V.N. Mahajan, Zernike circle polynomials and optical aberrations of systems with circular pupils, Appl. Opt. 33 (34) (1994) 8121–8124.
[25] N. Murata, S. Ikeda, A. Ziehe, An approach to blind source separation based on temporal structure of speech signals, Neurocomputing 41 (2004) 1–24.
[26] M. Najar, M.A. Lagunas, I. Bonet, Blind wideband source separation, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 4, 1994, pp. 19–22.
[27] L. Parra, C. Spence, Convolutive blind separation of non-stationary sources, IEEE Trans. Speech Audio Process. 8 (2000) 320–327.
[28] D.T. Pham, Contrast functions for blind source separation and deconvolution of sources, in: Proceedings of the ICA2001, 2001, pp. 37–42.
[29] D.T. Pham, Fast algorithm for estimating mutual information, entropies and score functions, in: Proceedings of the ICA2003, 2003, pp. 17–22.
[30] D.T. Pham, P. Garrat, Blind separation of a mixture of independent sources through a quasi-maximum likelihood approach, IEEE Trans. Signal Process. 45 (7) (1997) 1712–1725.
[31] K. Rahbar, J.P. Reilly, Blind source separation algorithm for MIMO convolutive mixtures, in: Proceedings of the ICA, 2001, pp. 242–247.
[32] I. Sabala, A. Cichochi, S. Amari, Relationships between instantaneous blind source separation and multichannel blind deconvolution, in: Proceedings of the IEEE International Conference on Neural Networks and Computational Intelligence, vol. 1, 1998, pp. 39–44.
[33] H. Sahlin, H. Broman, Separation of real-world signals, Signal Process. 64 (1) (1998) 103–113.
[34] Y.Y. Schechner, N. Kiryati, R. Basri, Separation of transparent layers using focus, Int. J. Comput. Vision 89 (2000) 25–39.
[35] Y.Y. Schechner, J. Shamir, N. Kiryati, Polarization statistical analysis of scenes containing a semi-reflector, J. Opt. Soc. Am. A 17 (2000) 276–284.
[36] S. Shamsunder, G.B. Giannakis, Multichannel blind signal separation reconstruction, IEEE Trans. Speech Audio Process. 5 (6) (1997) 515–528.
[37] C.J.R. Sheppard, I.J. Cooper, Fresnel diffraction by a circular aperture with off-axis illumination and its use in deconvolution of microscope images, J. Opt. Soc. Am. A 21 (4) (2004) 540–545.
[38] S. Shwartz, M. Zibulevsky, Y.Y. Schechner, ICA using kernel entropy estimation with NlogN complexity, in: Proceedings of the ICA2004, 2004, pp. 422–429.
[39] S. Shwartz, M. Zibulevsky, Y.Y. Schechner, Fast kernel entropy estimation and optimization, Signal Process. 85 (5) (2005) 1045–1058 (special issue on Information Theoretic Signal Processing).
[40] C. Simon, Ph. Loubaton, C. Vignat, C. Jutten, G. d'Urso, Blind source separation of convolutive mixtures by maximization of fourth-order cumulants: the non i.i.d. case, in: Proceedings of the IEEE International Conference on Signals, Systems and Computers, vol. 2, 1998, pp. 1584–1588.
[41] E.P. Simoncelli, Statistical models for images: compression restoration and synthesis, in: Proceedings of the IEEE Asilomar Conference on Signals, Systems, and Computers, 1997, pp. 673–678.
[42] P. Smaragdis, Blind separation of convolved mixtures in the frequency domain, Neurocomputing 22 (1998) 21–34.
[43] P.A. Stokseth, Properties of a defocused optical system, J. Opt. Soc. Am. A 59 (10) (1969) 1314–1321.
[44] A. Taleb, J. Sole, C. Jutten, Quasi-nonparametric blind inversion of Wiener systems, IEEE Trans. Signal Process. 49 (5) (2001) 917–924.
[45] H.-L.N. Thi, C. Jutten, Blind source separation for convolutive mixtures, Signal Process. 45 (3) (1995) 209–229.
[46] J. Toro, F. Owens, R. Medina, Using known motion fields for image separation in transparency, Pattern Recognition Lett. 24 (1–3) (2003) 597–605.
[47] Y. Tsin, S.B. Kang, R. Szeliski, Stereo matching with reflections and translucency, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, vol. 1, 2003, pp. 702–709.
[48] D. Yellin, E. Weinstein, Multichannel signal separation: methods and analysis, IEEE Trans. Signal Process. 44 (1) (1996) 106–118.
[49] W. Zhou, C. Kambhamettu, Separation of reflection component by Fourier decoupling, in: Proceedings of the Asian Conference on Computer Vision, 2004, pp. 27–30.
[50] M. Zibulevsky, Blind source separation with relative Newton method, in: Proceedings of the ICA2003, 2003, pp. 897–902.
[51] M. Zibulevsky, B.A. Pearlmutter, Blind source separation by sparse decomposition in a signal dictionary, Neural Comput. 13 (4) (2001) 863–882.

Sarit Shwartz received her B.Sc. and M.Sc. degrees from the Department of Electrical Engineering at the Technion, Israel, in 2002 and 2006, respectively. She has since been working in the field of signal and image processing, focusing on image enhancement.

Yoav Y. Schechner received his B.A. and M.Sc. degrees in Physics and Ph.D. in Electrical Engineering from the Technion—Israel Institute of Technology in 1990, 1996, and 2000, respectively. During the years 2000–2002 Yoav was a research scientist at the Computer Science Department in Columbia University. Since 2002, he is a faculty member at the Department of Electrical Engineering of the Technion, where he heads the Hybrid

Imaging Lab. His research is focused on computer vision, the use of optics and physics in imaging and computer vision, and on multi-modal sensing. He was the recipient of the Wolf Foundation Award for Graduate Students in 1994, the Guttwirth Special Distinction Fellowship in 1995, the Israeli Ministry of Science (Eshkol) Distinction Fellowship and the Ollendorff Award in 1998, the Scwartz Foundation Award in 1999 and the Morin Fellowship in 2000–2002. He is now a Landau Fellow—supported by the Taub Foundation—and an Alon Fellow. He has received the Klein Research Award in 2006.

Michael Zibulevsky received the M.Sc. degree in Electrical Engineering from MIIT, Moscow, and the Ph.D. degree in Operation Research (Nonlinear Optimization) from the Technion—Israel Institute of Technology. Currently he is with the Department of Computer Science at the Technion. His area of interests includes nonlinear optimization and inverse problems in signal/image reconstruction (blind source separation, deconvolution, tomography, hyperspectral imaging, EEG/MEG).