Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
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

[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].

*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).

0925-2312/$ - see front matter
doi:10.1016/j.neucom.2007.07.042

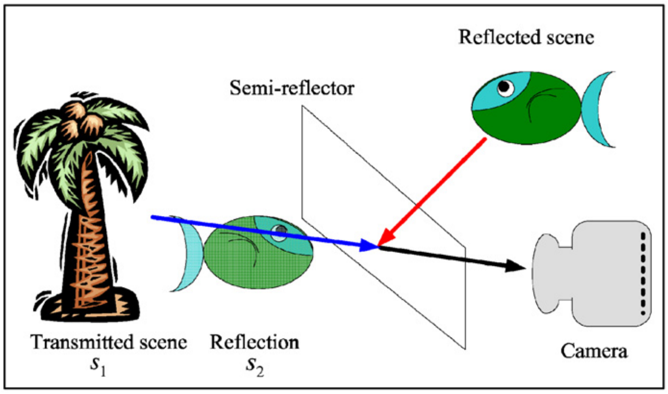

Fig. 1. An image of a mixture of two layers, created by a semi-reflector.

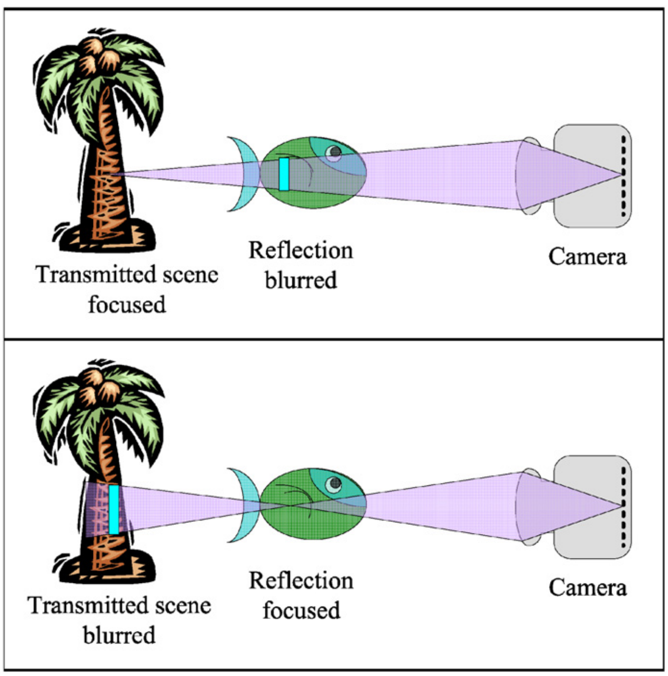

Fig. 2. Image acquisition of the scene depicted in Fig. 1: [Top] the camera is focused on the transmitted object while the reflection is blurred; [Bottom] the camera is focused on the reflection while the transmitted object is blurred.

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