Blind Separation of Superimposed Shifted Images Using Parameterized Joint Diagonalization

Efrat Be’ery and Arie Yeredor, Senior Member, IEEE

Abstract—We consider the blind separation of source images from linear mixtures thereof, involving different relative spatial shifts of the sources in each mixture. Such mixtures can be caused, e.g., by the presence of a semi-reflective medium (such as a window glass) across a photographed scene, due to slight movements of the medium (or of the sources) between snapshots. Classical separation approaches assume either a static mixture model or a fully convolutive mixture model, which are, respectively, either under- or over-parameterized for this problem. In this paper, we develop a specially parameterized scheme for approximate joint diagonalization of estimated spectrum matrices, aimed at estimating the succinct set of mixture parameters: the static (gain) coefficients and the shift values. The estimated parameters are, in turn, used for convenient frequency-domain separation. As we demonstrate using both synthetic mixtures and real-life photographs, the advantage of the ability to incorporate spatial shifts is twofold: Not only does it enable separation when such shifts are present, but it also warrants deliberate introduction of such shifts as a simple source of added diversity whenever the static mixing coefficients form a singular matrix—thereby enabling separation in otherwise inseparable scenes.

Index Terms—Approximate joint diagonalization, blind source separation (BSS), image reflections, image separation, spatial shifts.

I. INTRODUCTION

QUITE commonly in photography, when scenes are recorded through window glass under poorly balanced lighting conditions, a reflection of the room interior appears (linearly) superimposed over the outside scene. Separating such reflections from the scene using a single image (snapshot) is obviously an ill-posed problem. However, when (at least) two such snapshots are available, each taken under slightly different conditions, it is conceivable that successful (blind) separation can be attained by proper exploitation of the diversity in the different snapshots. The diversity sources can be divided into two categories according to the type of the images mixtures: static mixtures or convolutive mixtures.

In static mixtures, only the (relative) intensities of the source images change from one snapshot to another. The static model can be regarded as an extension of the classic (static) 1-D blind source separation (BSS) problem for 2-D signals

\[ x[m, n] = A \cdot s[m, n] \]  

(1)

where at each pixel \([m, n]\), \(s[m, n]\) \(\triangleq [s_1[m, n], s_2[m, n], \ldots, s_L[m, n]]^T\) is the \(L \times 1\) vector of source images, \(x[m, n]\) \(\triangleq [x_1[m, n], x_2[m, n], \ldots, x_P[m, n]]^T\) is the \(P \times 1\) mixtures vector and \(A\) is the \(P \times L\) mixing matrix, containing the linear mixing coefficients. Intensity attenuation may be achieved by changing the lighting conditions between snapshots (e.g., [3]). However, uniform modification of the lighting conditions over different areas of the source images may often be practically unfeasible, which might undermine the static (shift-invariable) mixing model (1). Thus, an alternative way for tampering with the source images’ relative intensities (without changing the lighting conditions) was considered in [4]–[6], where the polarization conditions were changed between snapshots. By introducing a linear polarizer in front of the camera, the relative coefficients of the mixed images may be altered. This optical setup is usually more true to the static model in (1) than a change in illumination conditions. However, even in this approach, the linear model is not fully accurate if the polarizer angle is changed by mechanical rotation (see [7]).

In convolutive mixtures, convolutions of the source images with different (unknown) filters are assumed to occur prior to the mixing. In many 2-D convolutive BSS (CBSS) algorithms, the estimation of the mixing filters (as well as the subsequent separation) is executed in the frequency domain, where the convolutive mixing transforms to different static mixtures at each frequency, forming models similar to (1). Convolutive image mixtures occur, for example, when light rays reflected from the objects go through optical devices (spatial filters), such as lenses. In this case, the optical transfer function is usually called a blur kernel. Sometimes, the convolutive image mixtures result from defocus, i.e., one (or more) of the source images is out of focus (blurred) in one (or more) of the mixtures. In this case, the change in focus conditions may be exploited as an alternative, or additional source of diversity (e.g., [8] and [9]; see also [10]).

In this paper, we introduce another possible source of (convolutive) diversity in the form of spatial-shifts. Different relative spatial shifts of the sources between snapshots are the inevitable outcome of possible movement of the reflective medium...
or of the recorded objects from one snapshot to another. Moreover, such relative shifts can be introduced deliberately (e.g., by slightly slanting the window between snapshots), so as to impose the desired diversity. The accommodation of spatial shifts allows the exploitation of this further diversity between snapshots, and, thus, enables us to attain improved separation, especially when the static mixing coefficients alone give rise to ill-conditioned mixing. Although considered convolutive in general, the mixing in our model is far more succinctly parameterized than a general convolutive model, since only a small number of unknown shift-parameters is added to the unknown static mixing coefficients. Yet, our simple model of relative spatial shifts often provides an accurate description of the mixtures.

Naturally, for the model to be exact, the two scenes must remain unchanged between snapshots (besides the possible relative shifts and uniform intensity changes). Usually, this would imply that the scenes should either contain only still, rigid and nondeformable objects with no movements in the direction towards/from the camera, or, alternatively, that the time between snapshots should be sufficiently short, such that any internal displacement is negligible with respect to the transversal spatial shift of the entire scene. Note, however, that these basic requirements are also shared by other (aforementioned) separation scenarios. If no intensity changes are present between snapshots (we would refer to this condition as a “singular mixing matrix” in the sequel), then to still enable separation, the relative shifts of the sources between the snapshots must be different. For example, in the common case of two-snapshots-two-sources, at least one source must appear shifted between snapshots, and the other source, if shifted as well, must be shifted in a different direction.

To the best of our knowledge, the problem of blind separation of image mixtures consisting of pure unknown relative spatial-shifts in addition to unknown scalar mixing coefficients has not been considered explicitly (to date) in open literature. Separation of dynamic images from a static background over several frames (an underdetermined problem) has been considered in [11]. Separation of two layers from multiple snapshots, taken from various known camera positions and involving transparencies and reflections was considered by Tsin et al. [12] and also by Szeliski et al. [13]. The problem of separating 1-D time-domain mixtures with different relative time-delays has been considered in [14], where a separation approach based on specially parameterized approximate joint diagonalization (AJD) was proposed. The AJD approach exploited the extended “alternating-columns—diagonal-centers” (AC-DC) algorithm [15]. In this paper, we address a model similar to the 1-D model considered in [14], but extend the approach of [14] to accommodate 2-D image signals, with the 1-D time-delays substituted by 2-D spatial-shifts.

This paper is structured as follows. Following the problem formulation in Section II, we outline an AJD-based solution for estimating the mixtures’ parameters in Section III. We describe the frequency-domain separation procedure (estimation of the source images) in Section IV. The overall computational complexity is evaluated and discussed in Section V. The suggested 2-D AC-DC algorithm was applied to synthetic mixtures, as well as to real-life photographs of reflective scenes. The results of the separation are presented in Section VI, where we also discuss the applicability of our method for separating overlapping video sequences. Concluding remarks are summarized in Section VII.

The following notation is used. Boldface uppercase letters denote matrices, boldface lowercase letters denote column vectors, and standard lowercase letters denote scalars. The superscripts \((\cdot)^T\), \((\cdot)^H\), \((\cdot)^*\), and \((\cdot)^{-1}\) denote the transpose, conjugate transpose, conjugate and inverse, respectively; \(\text{vec}\{\cdot\}\) denotes the concatenation of a matrix’ columns into one vector; \(\text{diag}\{\cdot\}\) denotes the vector of diagonal values when operating on a matrix, or a diagonal matrix with the specified diagonal values when operating on a vector; \(\|\cdot\|_F\) denotes the Frobenius norm, \(\text{Tr}\{\cdot\}\) denotes the trace; \(\otimes\) and \(\circ\) denote Kronecker’s product and Hadamard’s (element-wise) product, respectively. Parentheses/brackets are used to enclose continuous/discrete-space indices, respectively. The symbol \(f\) (not \(j\)) is used to denote \(\sqrt{-1}\), so as to avoid conflict with the simple index \(j\).

II. PROBLEM FORMULATION

We consider the following model of continuous 2-D sources and sensors (to be later transformed into a discrete model):

\[
x_p(u, v) = \sum_{l=1}^{L} a_{pl} s_l(u - \mathbf{d}_{pl}^u, v - \mathbf{d}_{pl}^v) \quad p = 1, 2, \ldots P
\]

(2)

where \(s_l(u, v)\) are the 2-D sources, \(x_p(u, v)\) are the observations (mixtures), \(a_{pl}\) are the mixing coefficients and \(\mathbf{d}_{pl}^u, \mathbf{d}_{pl}^v\) are the shifts in \(u, v\) directions of source image \(l\) in mixture \(p\) with respect to its position in mixture 1. Without loss of generality, we use as a “working assumption” zero shifts of the source images in the first mixture, i.e., \(\mathbf{d}_{1l}^u = \mathbf{d}_{1l}^v = 0\) for \(l = 1, 2, \ldots L\). Note that this is equivalent to determining (without loss of generality) that the origin of the scene’s coordinate system coincides with the position of each source’s origin in the first mixture. In this continuous model, \(u, v, \mathbf{d}_{pl}^u, \mathbf{d}_{pl}^v\) are all measured in some physical length units, e.g., microns. The shifts are assumed to have occurred prior to the sampling (digital imaging) process and are, therefore, not necessarily an integer multiple of the pixels’ physical dimensions. We generally assume that \(L \leq P\), but we shall mainly concentrate on the case \(L = P\).

The (presampled, continuous-space) source signals (images) are assumed to be zero-mean, mutually uncorrelated, wide-sense stationary (WSS) processes with unknown spectra. Yet, it has to be stressed that the stationarity assumption is only used for simplifying the derivations, and is not instrumental for the resulting performance. The more important condition for successful separation in this context is that the sources be mutually uncorrelated. Luckily, this property is widely satisfied by images of independent sources, unlike the stationarity assumption, which is rarely satisfied in natural images.

The available data are samples of the continuous-space observations, \(x_p[m, n] \triangleq x_p(mW_u, nW_v)\), \(1 \leq m \leq M, 1 \leq n \leq N\), where \(W_u, W_v\) are the sampling intervals, typically the
III. ESTIMATION OF THE MIXING PARAMETERS

A. Formulation as a Joint Diagonalization Problem

The observations’ correlation functions are given by (3), shown at the bottom of the page, where $R_{pq}^{\xi}(\xi, \eta)$ denotes the correlation between the $p$th and $q$th mixed images at lags $(\xi, \eta)$, and $R_0^{\xi}(\xi, \eta)$ denotes the autocorrelation of the $\ell$th source image, such that the last equality is due to the statistical decorrelation between sources. Fourier-transforming (3), we obtain $S^{\ell}_{pq}(\tilde{\omega}, \tilde{\theta})$, the cross-spectrum between the $p$th and $q$th mixtures at (angular) frequencies $(\tilde{\omega}, \tilde{\theta})$, (4)

$$S^{\ell}_{pq}(\tilde{\omega}, \tilde{\theta}) = \sum_{l=1}^{L} a_{pl} a_{ql} e^{-j\omega(\tilde{\omega}_l - \tilde{\omega}_q)} e^{-j\theta(\tilde{\theta}_l - \tilde{\theta}_q)}$$

1 \leq p, q \leq P

where $S^{\ell}_{pq}(\tilde{\omega}, \tilde{\theta})$ is the $\ell$th source’s (unknown) spectrum. Now, assuming that the imaging (sampling) process does not introduce any aliasing, we obtain that the relation in (4) also holds for the spectra of the sampled images $x_p[m, n]$, with proper substitution of the “physical” angular frequencies $(\tilde{\omega}, \tilde{\theta})$ with unit-less angular frequencies $\omega \triangleq \tilde{\omega}_l - \tilde{\omega}_q$ and $\theta \triangleq \tilde{\theta}_l - \tilde{\theta}_q$, and of the “physical” shifts $(\tilde{\omega}_l, \tilde{\theta}_l)$ with their unit-less counterparts $\tilde{\omega}_l \triangleq \tilde{\omega}_l / \omega_l$, $\tilde{\omega}_q \triangleq \tilde{\omega}_q / \omega_l$ (as mentioned earlier, these are not assumed to be integers).

In addition to this substitution, (4) can also be expressed in matrix-form as

$$S_{\epsilon}(\omega, \theta) = B(\omega, \theta) S_\epsilon(\omega, \theta) B^H(\omega, \theta)$$

(5)

where $S_{\epsilon}(\omega, \theta)$ is a $P \times P$ matrix consisting of $S^{\ell}_{pq}(\omega, \theta)$ as its $(p, q)$th element, $S_\epsilon(\omega, \theta)$ is the sources’ joint spectra matrix, which is an $L \times L$ diagonal matrix (since the sources are uncorrelated) consisting of $S^{\ell}_{\epsilon}(\omega, \theta)$ as its $(\ell, \ell)$th elements, and $B(\omega, \theta)$ is the $P \times L$ matrix given by

$$B(\omega, \theta) \triangleq A \odot D^\epsilon(\omega; \Delta_\epsilon) \odot D^\epsilon(\theta; \Delta_\epsilon).$$

(6)

Here, $A$ is the constant matrix of mixing coefficients, whose $(p, \ell)$th element is $a_{pl}$, and the $P \times L$ matrices $D^\epsilon(\omega; \Delta_\epsilon)$, $D^\epsilon(\theta; \Delta_\epsilon)$ contain the exponential terms depending (respectively) on the $P \times L$ shifts-matrices $\Delta_\epsilon$, $\Delta_\epsilon$, which, in turn, contain all shifts $d_{pl}$, $d_{ql}$. More specifically, the $(p, \ell)$th elements of $D^\epsilon(\omega; \Delta_\epsilon)$, $D^\epsilon(\theta; \Delta_\epsilon)$ are given by

$$D_{p\ell}^\epsilon = e^{-j\omega d_{pl}} \quad D_{p\ell}^\epsilon = e^{-j\theta d_{ql}}$$

1 \leq p \leq P, \quad 1 \leq \ell \leq L.

(7)

The cross-spectral matrices $S_{\epsilon}(\omega, \theta)$ are unknown, but can be estimated from the available data, possibly by using the 2-D discrete-space Fourier transform (DSFT) of a truncated series of cross-correlations estimates (Blackman–Tukey’s method, e.g., [16]). Specifically, to estimate the $(p, q)$th element of $S_{\epsilon}(\omega, \theta)$, we may use

$$\hat{S}_{pq}^\epsilon(\omega, \theta) = \sum_{m=\tilde{M}}^{\tilde{M}} \sum_{n=\tilde{N}}^{\tilde{N}} \hat{\tilde{\tilde{F}}}_{pq}[m, n] e^{-j(\omega m + \theta n)}$$

(8)

where $\tilde{M}, \tilde{N}$ are (resp.) the truncation-window length in the $m$, $n$-directions, and

$$\hat{\tilde{F}}_{pq}[m, n] = \frac{1}{\tilde{M} \tilde{N}} \sum_{m=1}^{\tilde{M}} \sum_{n=1}^{\tilde{N}} x_p[m + m, n + n] x_q[m, n]$$

(9)

(under the convention that $x_p[m + m, n + n] = 0$ whenever the indices extend beyond the $[1 : \tilde{M}] \times [1 : \tilde{N}]$ support) are the (slightly biased) correlation estimates.

Note that if the sources are not stationary, the estimated “spectra” $\hat{S}_{pq}^\epsilon(\omega, \theta)$ are nearly meaningless. Nevertheless, if the zero correlation between sources is persistent, then the expression in (5) still holds (asymptotically), with the spectra substituted by spatial averages (over the entire images) of the location-dependent “local spectra.” Nevertheless, the diagonality of $S_{\epsilon}(\omega, \theta)$ is maintained, due to the persistent absence of correlation between sources.

When estimated values, rather than true values of $S_{\epsilon}(\omega, \theta)$ are used, (5) usually can no longer be satisfied simultaneously at all frequencies. However, once $S_{\epsilon}(\omega, \theta)$ is estimated at several frequencies-pairs (vectors)

$$R_{pq}^{\xi}(\xi, \eta) \triangleq E[x_p(u + \xi, v + \eta) x_q(u, v)]$$

$$= \sum_{k=1}^{L} a_{pk} a_{qk} E \left[ s_p(u - \tilde{d}_{pk}^\epsilon + \xi, v - \tilde{d}_{pq}^\epsilon + \eta) \cdot s_q(u - \tilde{d}_{pq}^\epsilon, v - \tilde{d}_{pq}^\epsilon) \right]$$

$$= \sum_{k=1}^{L} a_{pk} a_{qk} E \left[ s_p(u + \xi - \tilde{d}_{pq}^\epsilon, v + \eta - \tilde{d}_{pq}^\epsilon) \right] 1 \leq p, q \leq P$$

(3)
is an estimate of the unknown parameters of interest can be obtained by resorting to AJD, e.g., seeking to minimize the following least-squares (LS) criterion:

$$
\min_{A, \Delta_u, \Delta_v, \Gamma} C_{LS} \triangleq \sum_{i=0}^{I} \sum_{j=0}^{J} \left\| S_x(\omega_i, \theta_j) - B(\omega_i, \theta_j) S_0(\omega_i, \theta_j) B^H(\omega_i, \theta_j) \right\|_F^2
$$

(10)

where $\Gamma$ is an $I \times (I+1) \times (J+1)$ tensor (three-way array) containing the respective sources’ spectra $S_p(\omega_i, \theta_j) 1 \leq \ell \leq I; 0 \leq i \leq I; 0 \leq j \leq J$. Note that it is also possible to use a weighted LS criterion by introducing some positive weights $w_{i,j}$ into the sum; however, to simplify the exposition, we shall not pursue this possibility in here. The rest of this Section is concerned with efficient minimization of $C_{LS}$, leading to estimates of the mixing parameters (shifts and gains).

It is important to note an essential difference from the common frequency-domain AJD approach, which is often applied in the context of general convolutive mixtures. In the general approach, several spectral matrices (“target matrices”) are estimated at each frequency (e.g., from different segments of the images), and then AJD is applied separately to the respective set at each frequency. Such an approach is prone to suffer frequency-dependent permutation and scaling ambiguities inflicted by the solutions of each AJD at each frequency. Unless properly resolved, frequency-dependent ambiguities obviously give rise to severe distortions. However, in our proposed framework, a single AJD is applied to all frequencies together, with one “target matrix” at each frequency, relying on a specified frequency dependence of the diagonalizing matrices. This avoids the possibility of frequency-dependent ambiguities, since the same ambiguity would apply to all frequencies, resulting in an overall (single) permutation and scale ambiguity—which is quite acceptable.

Several AJD algorithms have been proposed in recent years; however, they all assume a constant diagonalization matrix $B$, rather than $B(\omega_i, \theta_j)$ which depends on the indices $(i,j)$. In [14] an extension of one particular AJD algorithm (AC-DC, [15]) was proposed to address the 1-D problem. In Section III-B, we propose further extension of the extended AC-DC, adapted to this 2-D minimization problem. For simplicity of the exposition we shall assume, from now on, that the number of mixtures equals the number of sources, namely $P = L$ (further simplification, to the case $P = L = 2$, will follow thereafter).

### B. AJD via Extended 2-D AC-DC

AC-DC [15] is an alternating-directions algorithm for minimizing a LS joint-diagonalization criterion such as $C_{LS}$ of (10). It is originally intended for finding a constant diagonalizing matrix $B$. In our case, the matrix $B$ is not constant; Nevertheless, it can be expressed [recall (6)] in terms of the three constant matrices, $A$, $\Delta_u$, and $\Delta_v$, containing the mixing parameters. It is then possible to minimize with respect to (w.r.t.) each column of $A$ and each pair of matching columns (columns with the same index) of $\Delta_u$ and $\Delta_v$, separately, thus alternating between minimizations w.r.t.:

- $\Gamma$ (in the “DC” phase);
- each column of $A$ (in the “AC-1” phase);
- each pair of matching columns of $\Delta_u$ and $\Delta_v$ (in the “AC-2” phase).

We shall now describe each phase in detail. To simplify notation, we shall omit the “hat” from above the estimated spectra $S_\hat{x}$ from now on.

1) “DC” Phase: In the DC phase, we wish to minimize $C_{LS}$ w.r.t. $\Gamma$, with $A$, $\Delta_u$, and $\Delta_v$ fixed. Since the $(i,j)$th column of $\Gamma$ is the diagonal of $S_\hat{x}(\omega_i, \theta_j)$, it participates only in the $(i,j)$th term of the sum in (10). Thus, the overall minimization can be decomposed into $(I + 1) \times (J + 1)$ distinct minimization problems, which are all linear in the unknown parameters. As such, each of these problems admits the well-known linear LS solution. Specifically, note that each $(i,j)$th term in the sum can be expressed as

$$
\|S_x(\omega_i, \theta_j) - B(\omega_i, \theta_j) S_0(\omega_i, \theta_j) B^H(\omega_i, \theta_j)\|_F^2 = [y_{ij} - H_{ij} \gamma_{ij} e_{ij} + H_{ij} \gamma_{ij} e_{ij}]
$$

where $\gamma_{ij}$ is the $(i,j)$th column of $\Gamma$, $y_{ij} \triangleq \text{vec}\{S_x(\omega_i, \theta_j)\}$, and

$$
H_{ij} \triangleq (B(\omega_i, \theta_j)^* \otimes 1) \odot (1 \otimes B(\omega_i, \theta_j))
$$

where $1$ denotes an $L \times 1$ vector of $1$-s. Note that this expression is sometimes also referred to as the Khatri–Rao product of $B$ and $B$. The well-known minimizer of each linear LS term in (11) is

$$
\gamma_{ij} = [H_{ij}^* H_{ij}]^{-1} H_{ij}^* y_{ij}
$$

(13)

Note that for simplifying the computational load, one may use the relations

$$
H_{ij}^* H_{ij} = (B(\omega_i, \theta_j)^* B(\omega_i, \theta_j))^* \otimes (B(\omega_i, \theta_j)^* B(\omega_i, \theta_j))
$$

(14a)

$$
H_{ij}^* y_{ij} = \text{diag}\left\{B(\omega_i, \theta_j)^* S_x(\omega_i, \theta_j) B(\omega_i, \theta_j)\right\}
$$

(14b)

with which the computation of $H_{ij}^* H_{ij}$ and of $H_{ij}^* y_{ij}$ requires only $O(L^2)$ multiplications (per frequency-point), for a total of $O(1JJ^2)$.

2) “AC-1” Phase: We now wish to minimize $C_{LS}$ w.r.t. $a$, the $\ell$th ($\ell = 1, 2, \ldots, L$) column of $A$, assuming the other columns, as well as $\Delta_u$, $\Delta_v$, and $\Gamma$, are fixed. To this end, let us define

$$
\bar{S}(\omega_i, \theta_j) \triangleq S_\hat{x}(\omega_i, \theta_j) - \sum_{k=1}^{L} S_k(\omega_i, \theta_j) b_k(\omega_i, \theta_j) b_k^H(\omega_i, \theta_j)
$$

(15)

where $b_k(\omega_i, \theta_j)$ is the $k$th column of $B(\omega_i, \theta_j)$. Note that since $S_k(\omega_i, \theta_j)$ are the implied estimates of the sources’ spectra, they are all real-valued [this can also be verified by observing that the expressions in (14a) and (14b) are real-valued, so all $\gamma_{ij}$ in (13), which contain all $S_k(\omega_i, \theta_j)$, are real-valued, as well].
We may, thus, express the LS criterion in (10) as (16), shown at the bottom of the page, where $C_{LS}$ is an independent constant. Observe now (from 6), that $b_{\ell}(\omega_i, \theta_j)$ can be written as $b_{\ell}(\omega_i, \theta_j) = \Lambda_{\ell}(\omega_i, \theta_j) a_{\ell}$ where

\[
\Lambda_{\ell}(\omega_i, \theta_j) \triangleq \text{diag} \left\{ e^{-j\omega f t_{\ell}}, \ldots, e^{-j\omega f t_{\ell}} \right\}. \tag{17}
\]

Consequently, $C_{LS}$ can be further simplified

\[
C_{LS} = C - 2a^T F a + d^T f
\]

where $F$ is the Hermitian matrix

\[
F \triangleq \sum_{i=0}^{I} \sum_{j=0}^{J} S_{\ell}^p(\omega_i, \theta_j) \Lambda_{\ell}(\omega_i, \theta_j) \bar{S}(\omega_i, \theta_j) \Lambda_{\ell}(\omega_i, \theta_j) \tag{20}
\]

and $f$ is a positive constant

\[
f \triangleq \sum_{i=0}^{I} \sum_{j=0}^{J} S_{\ell}^p(\omega_i, \theta_j). \tag{21}
\]

Differentiating (19) w.r.t. $a$ and equating to zero yields three potential solutions for the minimizing $a$: either $a = 0$ or

\[
a = \pm \sqrt{\frac{a^T F a}{f}}. \tag{22}
\]

Since $F$ is Hermitian, the argument of the square-root in (22) is real-valued for all real-valued $a$: however, it is not guaranteed to be positive. Thus, a real-valued square root does not always exist. Indeed, if $F$ is negative-definite, the argument of the square-root would always be negative, so the only possible (real-valued) solution is $a = 0$, and minimization of $C_{LS}$ w.r.t. $a$ is attained in this case by $a_{\ell} = 0$. Normally, however, $F$ is not negative-definite, and, therefore, the argument of the square-root can be made positive, at least with some $\alpha$. Then, it is easy to observe that any real-valued nonzero $a$ satisfying (22) is preferable to $a = 0$, since it attains a smaller value for $C_{LS}$ in (19).

Indeed, let us assume that a real-valued square-root $a$ exists in (22). Substituting back into (19), we observe that the minimization problem reduces into maximization w.r.t. $\alpha$ of $(\alpha^T F a)^2 = (\alpha^T \text{Real } F \alpha)^2$, subject to $\alpha^T \alpha = 1$. The desired solution is well-known to be obtained as the eigenvector of Real $\{F\}$ associated with the largest (positive) eigenvalue. Combining this solution with (19), we obtain

\[
a_{\ell} = \sqrt{\frac{\lambda_{\text{max}}}{f}} \cdot v_{\text{max}} \tag{23}
\]

where $\lambda_{\text{max}}$ and $v_{\text{max}}$ are (resp.) the largest eigenvalue of Real $\{F\}$ and the associated eigenvector.

For evaluating the computational load, we assume one "full sweep" (namely, minimization with respect to each of the $L$ columns) per iteration. At each frequency-point, the main computational burden is that of calculating $\bar{S}(\omega_i, \theta_j)$ [in (15)], which is $O(L^3)$. Apparently, it can be argued that since this is required for each column, the total load (per frequency-point) is $O(L^4)$, but noting that the elements $S_{\ell}^p(\omega_i, \theta_j) b_{\ell}(\omega_i, \theta_j) b_{\ell}^H(\omega_i, \theta_j)$ in (15) may be calculated just once (per iteration, per frequency-point) and then used for all of the $L$ columns, the load is indeed $O(L^3)$, for a total of $O(L^2)$ multiplications per iteration. Note that the eigenvalue decomposition of Real $\{F\}$ can be considered negligible here, since it requires $O(L^3)$ multiplications and is done only once for the entire frequency range (per column).

3) "AC-2" Phase: It is now desired to minimize $C_{LS}$ w.r.t. the $\ell$th ($\ell = 1, 2, \ldots, L$) columns of $\Delta_\ell$ and $\Delta_e$ denoted $d_{\ell}^e$ and
respectively, assuming the other columns, as well as $\mathbf{A}$ and $\mathbf{B}$, are fixed. Since the dependence of $C_{LS}$ in (19) on the shifts $d_1^f, d_2^f$ appears only through $\mathbf{F}$, we can rewrite (19) as

$$C_{LS} = \tilde{C} - 2a_1^f \mathbf{F} a_2^f$$

(24)

where $\tilde{C}$ is another constant. Therefore, minimization of $C_{LS}$ requires maximization of $a_1^f \mathbf{F} a_2^f$ w.r.t. $\left\{d_1^p, d_2^p\right\}_{p=1}^{L}$, which are all the shifts in the $t$th columns, excluding $d_1^f, d_2^f$, which were arbitrarily set to zero. More explicitly, we seek to maximize

$$\sum_{l=1}^{L} \sum_{q=1}^{L} a_1^p a_2^q \left[ \sum_{i=0}^{J} \sum_{j=0}^{J} G_{pq}^{(L)}(\omega_i, \theta_j) \right] e^{-j\left(f(d_1^p d_2^q - d_1^p d_2^q)\omega_i - f(d_1^p d_2^q - d_1^p d_2^q)\theta_j\right)} a_1^p a_2^q.$$  

(25)

Here, $G_{pq}^{(L)}$ denotes the $(p, q)$th element of the matrix

$$G^{(L)}(\omega_i, \theta_j) \triangleq S^L(\omega_i, \theta_j) \tilde{S}(\omega_i, \theta_j).$$

(26)

For evaluating the computationally, let us assume that the maximization involves searching a $2(L-1)$-dimensional grid of shift-values, where each of the unknown $d_1^p$ and $d_2^p$ (for $p = 1, 2, \ldots, L$ except $p = \ell$) takes $K_u$ and $K_v$ (resp.) potential values on a scalar grid. $K_u, K_v$ are determined by the user, considering the range of possible shifts in each direction and the desired search resolution. For a sweep of the $L$ columns of $\Delta_u$ and $\Delta_v$, this would require $O(IJL^2(K_uK_v)^{L-1})$ multiplications per iteration (including the negligible computational cost of $G^{(L)}(\omega_i, \theta_j)$), which may be rather large.

However, if we focus on the relatively simple (but quite common) case $L = 2$, we notice significant simplifications. First, we note that in this case only two elements (out of four) in the outer summation in (25) depend on the unknown shifts—the elements corresponding to $p \neq q$. Thus, for $\ell = 1$, we seek to maximize

$$a_{11021} \sum_{i=0}^{J} \sum_{j=0}^{J} G_{12}^{(1)}(\omega_i, \theta_j) e^{-f d_2^1 \omega_i - f d_2^1 \theta_j}$$

$$+ a_{11021} \sum_{i=0}^{J} \sum_{j=0}^{J} G_{21}^{(1)}(\omega_i, \theta_j) e^{+f d_2^1 \omega_i + f d_2^1 \theta_j}$$

(27)

w.r.t. $d_1^f$ and $d_2^f$. Observe now, that due to the conjugate-symmetric structure of $G^{(1)}(\omega, \theta)$ the two terms in this sum form a conjugate pair, and, therefore, the sum is given by twice the real-part of these terms. Thus, depending on the sign of $a_{11021}$, we either need to minimize or to maximize this real-part. A reasonable assumption for mixtures of images is that all the static mixture coefficients are non-negative. Consequently, the desired shifts are given by the maximization

$$d_2^1, d_2^1 = \max_{d_2^1, d_2^1} \text{Real} \{g_1(u, v)\}$$

(28)

where

$$g_1(u, v) \triangleq \sum_{i=0}^{I} \sum_{j=0}^{J} G_{22}^{(1)}(\omega_i, \theta_j) e^{f u \omega_i + f \theta_j}.$$  

(29)

As further simplification, we note that if the frequencies $\{\omega_i\}$ and $\{\theta_j\}$ are chosen as

$$\omega_i = \Omega, \quad i = 0, 1, \ldots, I$$

$$\theta_j = j\Theta, \quad j = 0, 1, \ldots, J$$

(30)

(with $\Omega$ and $\Theta$ some selected constants), then $g_1(u, v)$ is the 2-D inverse Fourier series (2-D IFS) of the sequence $G_{22}^{(1)}(\omega_i, \theta_j)$. Again, the search in each direction may be conducted over scalar grids of sizes $K_u, K_v$, giving rise to a 2-D search over a $K_uK_v$-sized grid.

For $\ell = 2$ (looking for $d_2^1, d_2^1$), we wish to maximize real $\{g_2(u, v)\}$ with $g_2(u, v) \triangleq \sum_{i=0}^{I} \sum_{j=0}^{J} G_{22}^{(2)}(\omega_i, \theta_j) e^{f u \omega_i + f \theta_j}$. This term can be similarly minimized by searching the 2-D IFS of $G_{22}^{(2)}$.

Thus, the overall computational load per iteration in the AC-2 phase for the $L = 2$ case is $O(IJKuK_v)$. This appears equivalent to the expression obtained above for general $L$ (substituting $L = 2$), yet due to the simplifications noted above, the implicit constant in the $O(\cdot)$ term is about four times smaller. Note further, that if $K_u \leq I$ and $K_v \leq J$, a computationally efficient 2-D fast Fourier transform (2-D FFT) can be used to compute the required 2-D DFS, thereby further reducing3 the $L = 2$ load to $O(IJ \log(IJ))$.

C. Initial Guess

For such alternating-directions algorithms, an “intelligent” initial guess is required, so as to avoid convergence to spurious minima. Luckily, a reasonable guess for the relative-shifts could be readily obtained from the data. From (3), we have

$$R_{p1}(\xi, \eta) = \sum_{l=1}^{L} \sum_{q=1}^{L} \sum_{i=0}^{J} \sum_{j=0}^{J} a_1^p a_2^q \left( \xi - d_2^1 \eta - d_2^1 \right)$$

$$+ \sum_{l=1}^{L} \sum_{q=1}^{L} \sum_{i=0}^{J} \sum_{j=0}^{J} a_1^p a_2^q \left( \xi - d_2^1 \eta - d_2^1 \right).$$

(31)

A basic property of an auto-correlation function, $R(\xi, \eta)$, is $R(0, 0) \geq |R(\xi, \eta)|, \forall \xi, \eta$. Hence, $R_{p1}^L(\xi - d_2^1 \eta - d_2^1)$ achieves its maximal value at $(\xi, \eta) = (d_2^1 \eta, d_2^1 \eta)$. Under reasonable conditions on the spatial-shifts, on the mixing coefficients and on the sources’ auto-correlation functions shape, the peaks related to the different sources are roughly distinguishable (see Fig. 1) to within accuracy which, although insufficient for separation, may be sufficient as an initial guess for the AJD. Consequently, rough estimates of the (unit-less) shifts pairs $d_2^u, d_2^v$ (for $1 \leq \ell \leq L$) can be extracted from the discrete estimate $R_{p1}^L[n, n]$ of $R_{p1}^L[n, n]$. Note, however, that

3Assuming $\log_2(IJ) < K_uK_v$.
D. Preprocessing Stage

Since the important condition of uncorrelated sources is generally better satisfied between edge-enhanced images than by their original counterparts, we attained considerably better estimation (hence, separation) when a linear shift-invariant (LSI) (zero-phase) edge-sharpening filter was applied to both mixtures as a preprocessing stage. Note that such LSI filtering is merely equivalent to the introduction of similar spectral shaping to both sources. The result simply implies some special weighting of the LS criterion (10) along the frequency-plane. Evidently, more weight is attributed in the AJD process to frequencies which correspond to dominant regions in the spectral shape of the filter.

In addition, the edge-enhancement preprocessing can be helpful for generating the initial guess using the cross-correlation between the edge-enhanced mixtures. The peaks exhibited by this cross-correlation function are more easily distinguishable than those exhibited by the cross-correlation of the raw images. To illustrate, we demonstrate typical cross-correlation functions with and without the preprocessing stage in Fig. 1. These plots were obtained with images taken from the first experiment, as described later on in Section VI-B.

We note in passing that our edge-enhancement preprocessing procedure might also be considered as a sparsifying transformation; see, e.g., [5] on the use of sparsity for separation.

The problem of finding “good” edge-sharpening filters is out of scope of this work. Hence, we did not put much effort in optimizing the edge-sharpening filter, and settled for a filter that was “good enough” for our purposes.

IV. SEPARATION OF THE SOURCES

Once all of the mixing parameters (the spatial-shifts, as well as the static mixing coefficients) have been estimated, they can be used to “reverse” the mixing operation and recover the source images. Generally, there are at least two possible approaches for separation: spatial-domain separation and frequency-domain separation.

A. Spatial-Domain Separation

Since the mixing model (2) was conveniently expressed in the spatial-domain, it might appear intuitively appealing to “undo” the mixing in the spatial-domain, as well. However, spatial-domain inversion of (2) would generally take a form which is considerably more complex than (2) itself. A recovered source cannot, in general, be expressed as a simple combination of the $P$ mixture-frames, in which each mixture-frame is (inversely) shifted.

Consider the simplest case of $L = 2$ sources/mixtures first. By “back-shifting” the second mixture by the estimated shifts of, say, the first source (namely, by $(-\tilde{\eta}_{21}, -\tilde{\eta}_{21})$), the position of the first source in the second mixture would be aligned with its position in the first mixture. Then, if the second, back-shifted mixture-frame is multiplied by $a_{11}/a_{22}$ and subtracted from the first mixture-frame, the first source is thereby eliminated from the result. However, this result now contains two superimposed, differently shifted versions of the second source, namely a filtered version thereof. In order to recover the second source, this filtering operation has to be inverted—which requires to apply an infinite impulse response (IIR) filter in the spatial domain. A similar operation can be used to recover the first source.

When the number of sources/mixtures is larger than 2, spatial-domain separation becomes even more complicated, since the sources have to first be eliminated one-by-one, creating “new” mixtures along the way. After such a “deflation” approach is applied, the remaining source would be recovered by applying the respective IIR filtering to the result, so as to invert the linear combinations of multiple-shifts thereof, incurred along the way.

An alternative approach for spatial-domain separation is taken by Szeliski et al. [13] in the case where the number of observations is much larger than the number of sources. The method relies on the statistical variation of the sources in the spatial-domain, and on the fact that pixel intensity values are non-negative. In their “Min-Composite” approach, Szeliski et al. propose to first “undo” the shifts of all mixtures such that one of the sources is aligned, and then take, in each pixel, the minimum intensity value (among all shifted mixtures) as an estimate of the desired source. Subsequent iterations with maximum-values are also considered.

Tsin et al. [12] use the “Min-Composite” approach as an initial estimate for an iterative procedure: Assume that a good estimate of the first source is available. It can then be aligned to fit its position in each mixture and eliminated (subtracted with proper gain), leaving estimates of only the second source in each mixture. These estimates are then shifted so as to align the replica of the second source, and averaged. The averaging effect refines the estimate of the second source, and the iterative algorithm
may proceed for successive refinement, switching roles between the two sources.

All of these spatial-domain approaches require the application of noninteger spatial shifts along the way. However, this can be usually attained (assuming sufficient sampling-rate) using relatively simple interpolation filters (even a two-taps linear interpolator may be sufficient). The drawback of these approaches in the context of our problem formulation (with an equal number of sources and mixtures) lies mainly with the subsequent processing. The rigorous approach mentioned earlier requires the application of IIR filtering. The Min-Composite approach, as well as the iterative approach by Tsin et al. rely on the availability of many more mixtures than sources for statistical stability. All of these approaches become considerably more involved when there are more than two sources.

On the other hand, when transformed to the frequency-domain, the mixing model becomes purely multiplicative in each frequency. The observation vector at each frequency is given by a matrix (which depends on the mixing parameters and on the frequency) times the source vector at that frequency. Therefore, straightforward, closed-form (noniterative) reconstruction of the sources can be attained simply by inverting this operation at each frequency.

With the exception of marginal end-effects, this frequency-domain approach can be easily shown to be equivalent to the (rigorous) spatial-domain approach. In fact, if null-margins, wider than the maximum shift-value, are added to the mixture-frames before transforming\(^5\) to the frequency-domain, exact equivalence can be attained, and yet, we find the frequency-domain approach both computationally and conceptually more simple than its spatial-domain counterpart.

### B. Frequency-Domain Separation

Using the (2-D) discrete Fourier transform (2-D DFT) of the observations, we obtain \(L\) 2-D series of size \(M \times N\)

\[
\hat{x}_p[r, t] = \sum_{m=1}^{M} \sum_{n=1}^{N} x_p[m, n] e^{-j2\pi((m-1)r/M)+(n-1)t/N)}
\]

\(p = 1 \ldots L\) \(r = 0 \ldots M - 1\) \(t = 0 \ldots N - 1\). \((32)\)

Denoting \(\hat{A}\) and \(\hat{A}_{kl}, \hat{A}_{k}\) the estimated mixing coefficients and (unit-less) spatial shifts, we construct matrices \(\hat{D}^v[r], \hat{D}^v[t]\) whose \(k, \ell\)th elements are given by

\[
\hat{D}^v[k, \ell] = \begin{cases} 
e^{-j2\pi k \ell r/M}, & 0 \leq r \leq M \frac{2}{M} \\ ne^{-j2\pi k \ell (r-M)/M}, & \frac{M}{2} < r \leq M - 1 \\ \end{cases}
\]

\(\ell = 1 \ldots L\). \((33)\)

The source images can now be separated (estimated) in the frequency domain using

\[
\hat{s}[r, t] = \left[ A \odot \hat{D}^v[r] \odot \hat{D}^v[t] \right]^{-1} \hat{x}[r, t]
\]

\(\triangleq \hat{B}^{-1}[r, t] \hat{x}[r, t]\) \((35)\)

where \(\hat{A}[r, t] \triangleq \{|x_1[r, t]|x_2[r, t]| \ldots |x_L[r, t]|\}^{T}\), and \(\hat{B}[r, t]\) is the estimate of matrix \(B(\omega, \theta)\) [see (6)] at the frequencies \(\omega = 2\pi r/M, \theta = 2\pi t/N\) (assuming the inverse exists). Transforming back to the space-domain, we obtain

\[
\hat{s}_p[m, n] = \sum_{r=0}^{M-1} \sum_{t=0}^{N-1} \hat{s}[r, t] e^{j2\pi((m-1)r/M)+(n-1)t/N)}
\]

\(m = 1 \ldots M\) \(n = 1 \ldots N\) \((36)\)

where \(\hat{s}[r, t]\) is the \(j\)th element of \(\hat{s}[r, t]\). The reconstructed images may possibly be subject to arbitrary (but frequency-independent!) permutations, scaling and shift, due to the inherent ambiguities of the blind separation problem.

This frequency-domain separation scheme implies exact inversion of the mixing (assuming exact estimates of the parameters) only for cyclic shifts (namely, for shifts in which any content shifted beyond the margins “returns” into the image from within the opposite margins). In reality, the shifts are obviously not cyclic, so even if the true parameters were used, our unmixing scheme would introduce some marginal end-effects. Nevertheless, these end-effects are rather tolerable when the frequency-domain mixing matrix is well-conditioned at all frequencies.

However, in the case of a singular (or nearly singular) static mixing matrix \(A\), the resulting frequency-domain mixing \(B[r, t]\) may be singular (or nearly singular) at several frequency-pairs\(^6\) (vectors). For example, for the case (later used in the experiments section) of the singular mixing matrix \(A = [0.5 \hspace{1cm} 0.5] [0.5 \hspace{1cm} 0.5]^{T}\) with shifts \(\Delta_u = [0 \hspace{1cm} 0 \hspace{1cm} -1.3 \hspace{1cm} 2.8] [0 \hspace{1cm} 2.1 \hspace{1cm} -1.2]^{T}\), the “singularity lines” in the frequency domain are illustrated in Fig. 2 (left): For the entire \(374 \times 374\) frequency-grid used in the experiment, we marked all grid-points \(\{|r, t|\}\) where the condition-number of the matrix \(B[r, t]\) is higher than 100, making it nearly singular.\(^7\) The corresponding “singularity stripes” are evident.

As a result, the inverse of the (nearly) singular \(B[r, t]\) may unduly enhance the mismatch between the cyclic shift and the physical shift models along these lines, resulting in a corresponding, severe “stripes effect” on the reconstructed image. We demonstrate this effect on one of the reconstructed images (in the same mixing example) in Fig. 2 (right).

It is important to note here that when the mixing-matrix is (nearly) singular, the spatial-domain reconstruction approach

\(^5\)Note that such margins should not be added before the parameters estimation stage, since they may introduce a bias in the estimation, towards a no-shifts condition.

\(^6\)In fact, for \(L > 2\), the matrix \(B[r, t]\) may be singular at some frequencies even if the mixing matrix \(A\) is regular.

\(^7\)Exact singularity (infinite condition number) occurs along straight lines in the continuous frequency-domain; However, points along the discrete grid might be “nearly” but not “exactly” singular, depending on their distance from these lines.
discussed earlier cannot offer a remedy to the problem. The singularity will be evident when trying to construct the IIR filter (see the previous subsection), which would not be stable in this case. This filter would have infinite gain along the same frequency-stripes, and would amplify the differences (present at the edges) between the true mixture and the presumed mixture. Even if null-margins are added to the mixture images before separation is applied, so as to avoid the cyclic-shifts, the resulting mixture would still be different, since the null margins cannot account for missing data of the true shifted sources, lost beyond the imaging frame. Moreover, we shall now show, that in our frequency-domain approach we would be able to offer a frequency-selective remedy, which cannot be applied in a purely spatial-domain approach.

In order to mitigate the “stripes effect,” we need to apply a more careful inversion (reconstruction) scheme. We propose to regard the cyclic mismatch as additive noise with some prescribed small variance (proportional to the extent of the shift), and apply a minimum mean square error (MMSE) inversion approach. In so doing, we shall employ several simplifying assumptions, in order to obtain a relatively simple result. This means that we might not end up with the true MMSE estimate. Nevertheless, MMSE estimation is not our prime goal here; we are merely seeking to mitigate the “stripes effect,” by regularizing the inversion at the relatively small number of problematic frequencies, with minimal effect on the separation at all other frequencies. This would be done automatically, without need for prior distinction between “problematic” and “benign” frequencies.

The “noisy” mixture at each frequency-bin \([r, t]\) is modeled as

\[
\hat{x}[r, t] = B[r, t] \hat{s}[r, t] + \hat{e}[r, t]
\]

where \(\hat{s}[r, t]\) are the respective “noise” components. Assuming, for simplicity, that these noise terms are uncorrelated between different frequency-bins, we may apply a separate MMSE estimate at each frequency, decoupled of all other frequencies. The MMSE estimator of \(\hat{s}[r, t]\) from \(\hat{x}[r, t]\) is well known to be given by

\[
\hat{s}[r, t] = C_{\hat{s}x}[r, t]^{-1} C_{\hat{x},s}[r, t] \hat{x}[r, t]
\]

where

\[
C_{\hat{s}x}[r, t] = C_{\hat{s}s}[r, t] B^H[r, t]
\]

is the covariance matrix between the sources-term \(s[r, t]\) and the mixtures-term \(\hat{x}[r, t]\), and

\[
C_{\hat{x},s}[r, t] = B[r, t] C_{\hat{s}s}[r, t] B^H[r, t] + C_{\hat{s}e}
\]

is the (auto-)covariance of the mixture-term \(\hat{x}[r, t]\). Here, \(C_{\hat{s}s}[r, t]\) and \(C_{\hat{s}e}\) denote the auto-covariances of the sources-term and of the noise-term, respectively. Since the sources are uncorrelated, \(C_{\hat{s}s}[r, t]\) is diagonal. We also assume that \(C_{\hat{s}e}[r, t]\) is diagonal. For further simplicity we assume that \(C_{\hat{s}s}[r, t]\) and \(C_{\hat{s}e}[r, t]\) do not depend on the frequency \([r, t]\). With all these simplifying assumptions, the MMSE estimate takes the form

\[
\hat{s}[r, t] = C_{\hat{s}s} B^H[r, t] \left( B[r, t] C_{\hat{s}s} B^H[r, t] + C_{\hat{s}e} \right)^{-1} \hat{x}[r, t].
\]

Our last simplifying assumption is that \(C_{\hat{s}s}\) has a constant diagonal. We also make a similar assumption regarding \(C_{\hat{s}e}\), with the exception that the \((1, 1)\) element of this matrix is null: This is due to the “working assumption” that the sources are unshifted in the first mixture, so that the cyclic-shift mismatch noise for \(x_i[r, t]\) has zero variance. Finally, using the estimated \(\hat{B}[r, t]\) instead of the true \(B[r, t]\), our reconstruction is formed as a regularized version of the direct inversion used in (33)

\[
\hat{s}[r, t] = \hat{B}^H[r, t] \left( \hat{B}[r, t] \hat{B}^H[r, t] + \mu I \right)^{-1} \hat{x}[r, t]
\]

where \(I\) denotes the identity matrix with the \((1, 1)\) element set to zero, and \(\mu\) is a small constant value, reflecting the presumed ratio between the energy of the “noise” (mismatch) and the energy of the source images. Obviously, when \(\mu\) approaches zero, (40) reduces to (33) if \(\hat{B}[r, t]\) is invertible.

As in the nonsingular case, we then apply the inverse transform in order to return to the space domain

\[
\hat{s}[m, n] = \text{IDFT} \{\hat{s}[r, t]\}.
\]

This approach reduces the “stripes effect” caused by the (nearly) singular frequency points (if any), yet has nearly no effect at nonsingular points, as long as \(\mu\) is small enough. See the improved experimental results (for the conditions of Fig. 2) in Section V.

For separation of color images, the same separation scheme may be applied to each color layer separately. The mixing parameters, which are assumed common to all color layers, are estimated just once, from the intensity signals only. Then, the separation process in each color layer uses the same estimated mixing parameters.

V. COMPUTATIONAL LOAD

Let us now summarize the overall computational load of the proposed separation scheme. We remind, for convenience, that \(L\) denotes the number of sources/mixtures; \(M, N\) denote the
image size; $I, J$ are the dimensions of the spectral frequency-grid; $K_u, K_v$ are the numbers of points along the horizontal and vertical grids (resp.) used for grid-based maximization in the AC-2 phase; and $\hat{M}, \hat{N}$ are the dimensions of truncated, estimated correlation sequence. In addition, we denote by $H_u, H_v$ the dimensions of the preprocessing edge-enhancement filter. We express the load in terms of orders of multiplications-counts as a function of these parameters. Note that this count excludes, e.g., additions and search operations, but they are roughly of either comparable or negligible complexity in each stage, so they hardly affect the order of total operations counts.

- The preprocessing filtering requires $O(MNH_uH_vL)$ multiplications (for a separable preprocessing filter this can be reduced to $O(MNH_u + H_v)L)$.
- The estimation of the correlation matrices requires $O(M\hat{N}\hat{M}L^2)$ multiplications.
- Transformation into the frequency domain for obtaining the spectral matrices requires $O(M\hat{N}IIL^2)$ multiplications; If a uniform frequency-grid is used, then this can be done using 2-D FFT in $O(IIL\log(IIL)^2)$, which may be significantly lower.
- The DC phase requires $O(IIL^3)$ multiplications per iteration (recall Section III-B-I for details).
- The AC-1 phase also requires $O(IIL^3)$ multiplications per iteration (recall Section III-B-II for details).
- The AC-2 phase requires $O(IIL^3(K_uK_v)^{L-1})$ multiplications per iteration (recall Section III-B-III for details).
- The separation in the frequency-domain requires two transformations of all images to/from the frequency-domain ($O(M\hat{N}\log(M\hat{N})L)$), in addition to the calculation of separation matrices and separation results at each frequency ($O(M\hat{N}L^2)$) for a total of $O(M\hat{N}L + \log(M\hat{N}))$ multiplications.

The total multiplications count naturally depends on the number of iterations required for convergence of the 2-D AC-DC algorithm (typically 15-20 in our experiments, see below). Denoting that number by $N_{it}$, the total count is

$$O(M\hat{N}L(H_uH_v + L + \log(M\hat{N}))) + O(M\hat{N}IIL^2) + O(N_{it}IIL^3(K_uK_v)^{L-1}) = (44)$$

Things simplify when we take the common case of $L = 2$. If we use a uniform frequency-grid and further assume that $H_u, H_v$ is of the same order as (or smaller than) $\log_2(M\hat{N})$, and that $K_u = I, K_v = J$ (thus, using a 2-D FFT in AC-2), the overall count reduces (and simplifies) into $O(M\hat{N}\log(M\hat{N}) + I(\hat{M}\hat{N} + N_{it}\log(IJ)))$.

Note that the iterative 2-D AC-DC estimation phase is not a major part in the overall computational load (assuming a reasonably small number of iterations). The estimation of the spectral matrices and the frequency-domain separation may be considerably heavier. Thus, other frequency-domain based separation schemes (e.g., [9]) would essentially be of comparable computational load.

Considerably cheaper methods for image separation certainly exist (e.g., [3], [6], and [17]). However, these methods do not account for relative spatial shifts and are, therefore, inapplicable in the context of our problem.

VI. SEPARATION RESULTS

In this section, we present image reflections separation results of the 2-D AC-DC algorithm. First, we present some results obtained with synthetic mixtures. Then, we demonstrate results obtained with real-life photographs containing reflections. Finally, we present another application for our 2-D AC-DC algorithm, in the form of separation of dissolved video sequences.

Although we used color images, all images in this section are presented (in print) in grayscale colors. Color versions of all images in this section can be found online in [18].

A. Creating Synthetic Mixtures

In order to be able to apply fractional shifts, the image mixtures were constructed in the frequency domain. First, the 2-D DFT of the source images was calculated

$$\hat{s}_r[l] = \sum_{m=1}^{M} \sum_{n=1}^{N} s_{r}[m,n] e^{-j2\pi(m-1)r/M+(n-1)t/N}$$

$$\ell = 1\ldots L \quad r = 0\ldots M-1 \quad t = 0\ldots N-1$$

Next, given the desired mixture parameters in the form of the $P \times L$ matrices $A, \Delta_u, \Delta_v$, we constructed two shifts-matrices $D^p[r], D^s[t]$ whose $p, t$th elements are given by

$$D^p[r] = \begin{cases} e^{-j2\pi p/r}M, & 0 \leq r \leq \frac{M}{2} \\ e^{-j2\pi p(r-M)/M}, & \frac{M}{2} < r \leq M-1 \end{cases}$$

$$D^s[t] = \begin{cases} e^{-j2\pi s(t-N)/N}, & 0 \leq t \leq \frac{N}{2} \\ e^{-j2\pi s(t-N)}M, & \frac{N}{2} < t \leq N-1 \end{cases}$$

Then, the vector $\hat{x}_r[l]$, a cyclic-shifted version of the mixtures, was created (in the frequency domain)

$$\hat{x}_r[l] = [A \odot D^p[r] \odot D^s[t]]\hat{s}_r[l]$$

where $\hat{s}_r[l] = [s_1[l], s_2[l], \ldots, s_L[l]]^T$ is the sources’ 2-D DFT vector. Next, we applied the inverse 2-D DFT to (44)

$$\hat{x}_r[l] = M \sum_{m=1}^{M} \sum_{n=1}^{N} \hat{s}_r[l] e^{j2\pi(m-1)r/M+(n-1)t/N}$$

Finally, the mixtures margins were removed to obtain linearly shifted versions of the mixtures

$$\hat{x}[m,n] = \hat{x}_r[m + P_m, n + P_n]$$

where the margins width $P_m, P_n$ satisfy the condition $P_m > \max_{p \in \{l\}} |k|_{p,l}$, $P_n > \max_{p \in \{l\}} |k|_{p,l}$. The frame-size of the observed images was then redefined accordingly

$$M := M - 2P_m, \quad N := N - 2P_n$$

As mentioned above, thanks to this frequency-domain technique, we were able to create fractional shifts, without using 2-D interpolation and 2-D decimation of the images. If the source images sampling rate complies with the Nyquist rate (as we
assumed), the proposed technique is equivalent to interpolation and shift of the images, followed by decimation. Note that source images reconstruction process (described in Section IV) is the (estimated) inverse procedure of the mixtures construction (excluding the margins removal step). The proposed procedure is a general mixing process for the \( L \) sources—\( P \) sensors case. In the simulations, we addressed the scenario of \( P = L = 2 \).

### B. Synthetic Mixtures—Results

We picked a set of two source images: an image of a fish and an image of chips (Fig. 3, left column). Both are color images, containing three color layers: red, green, and blue (RGB). Each of these layers was individually mixed, with identical linear coefficients and spatial-shifts. As mentioned earlier, in the estimation stage we first transformed the RGB mixture images into grayscale (intensity) images. Since this transformation is linear, the obtained grayscale mixture images admit the same linear-mixing and shifts model as the associated RGB components. We then subtracted the mean from each mixture (to comply with the zero-mean assumption) and applied the preprocessing edge-sharpening filter\(^8\) prior to estimation of the mixing parameters via the 2-D AC-DC algorithm. Using the estimated mixing parameters we then reconstructed each of the color layers separately.

We first demonstrate the performance with a nonsingular mixing matrix \( \mathbf{A} \). For this experiment, we used \( \mathbf{A} = \begin{bmatrix} 0.45 & 0.55 \\ 0.7 & 0.3 \end{bmatrix} \) and \( \Delta_{\mathbf{u}} = \begin{bmatrix} 0 & 0 \\ -1.3 & 2.8 \end{bmatrix} \Delta_{\mathbf{v}} = \begin{bmatrix} 0 & 0 \\ 2.1 & -1.2 \end{bmatrix} \).

We applied 2-D AC-DC to the mixture, extracting initial guesses for \( \Delta_{\mathbf{u}}, \Delta_{\mathbf{v}} \) from the estimated correlation (as described in Section III-C) and using the identity matrix as an initial guess for \( \mathbf{A} \). For the separation process we set \( \mu = 10^{-8} \) in (40). The sources and the mixture are presented in Fig. 3, and separation results are presented in Fig. 4.

Aside from some minor edge effects, both sets of color images are nicely reconstructed, with no visible cross-talk between the images. As explained before, the edge effects are the result of the noncircular shifts.

We also compared to the result of “clairvoyant” separation using the true (inverse) mixing matrix, but ignoring the spatial-shifts. The reconstructed images in Fig. 4 demonstrate the importance of shifts-estimation. Even when the true mixing matrix is known, the separation results are poor if the spatial-shifts are not accounted for.

In a second experiment, we simulated a shifts-only mixture scenario for the same source images, using a singular mixing matrix \( \mathbf{A} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} \), so as to demonstrate the attainable separability induced by the shifts in a case that would otherwise be inseparable. We used the same shifts matrices and the same initialization as in the first experiment. We stress out that the 2-D AC-DC algorithm used no a priori knowledge of this shifts-only scenario in estimating the mixtures (i.e., both the spatial-shifts and the coefficients matrix were estimated from the data, as in the case of a nonsingular coefficients matrix). In this experiment, we used \( \mu = 10^{-3} \). The mixture is presented in Fig. 3 and the reconstructed images are in Fig. 4.

As explained in Section IV, the inversion problem in the “singular” scenario is inherently ill-conditioned along “singularity lines” in the frequency domain. Direct inversion of these singularities enhances (along these lines) the difference between the implicitly assumed circular shifts and the actual noncircular shifts, and can severely impair the reconstruction, as demonstrated in Fig. 2. We, therefore, used the MMSE reconstruction approach (40). Note, however, that some residual artifacts are still visible in the respective separation in Fig. 4 (right), though considerably weaker than in Fig. 2. They are (mainly) caused by this inherent inversion problem, and are not due to inaccurate parameters estimation in the 2-D AC-DC algorithm.

In all these experiments the image sizes were \( M \times N = 374 \times 374 \), the correlation window size was \( M \times N = 25 \times 25 \), and the 2-D AC-DC frequency-grid was of size \( I \times J = 51 \times 51 \), equally spaced in \((-\pi, \pi) \times (-\pi, \pi)\). A breakdown of the empirical running-times (on Matlab 6.5, 2.99GHz PC) for these experiments is as follows:

- preprocessing time (edge-enhancement): 0.2 s;
- estimation of correlations and spectral matrices: 6.0 s;
- estimation of the mixing parameters (2-D AC-DC): around 20 iterations, each taking 0.13 s, for a total of 2.6 s;
- frequency-domain separation and image reconstruction (three color-layers): 3.3 s;

for a total of around 12.1 s. As noted earlier, the iterative 2-D AC-DC algorithm is not the most time-consuming phase of the entire scheme.

---

\(^8\)We used a 7 \( \times \) 7 filter, whose impulse response is proportional to \( h[n, m] = -\pi m \exp(-m^2 + n^2)/2), |n|, |m| \leq 3 \).
C. Noisy Synthetic Mixtures

To demonstrate the separation performance in the presence of additive noise, we also present some experiments with noisy synthetic mixtures. We applied noise to both “nonsingular” and “singular” mixtures, in a total of three experiments. White Gaussian noise was added\(^9\) independently to each color-layer of each mixture. The standard-deviation of the noise (out of 255 levels for each color layer) was 10 in the first experiment (with a “nonsingular” mixture), and 5 and 10 (respectively) in the second and third experiments (with a “singular” mixture).

The mixtures, as well as the respective separated images, are presented in Figs. 5 and 6. It is evident that although the noise is also present in the reconstructed images, the separation is nearly unaffected. It is to be noted that the separation scheme does not attempt to denoise the images—only separation is attempted, and successfully attained with these noise levels. For significantly higher noise levels, or for noise which is correlated between the snapshots, some preprocessing denoising may be required to enhance the performance, but this issue is beyond the scope of this work.

D. Real-World Conditions Results

To further demonstrate the performance of the algorithm, we used some real-world images reflections. We photographed (using a FUJIFILM F401, 4MPixels Digital Camera) a room interior (library) through a glass window. Outside the room we placed some object (teddy-bear, camera). The object’s reflection was superimposed on the interior scene. In order to obtain spatial-shifts, the window was slightly slanted from one snapshot to the other. The lighting conditions were not changed between snapshots (at least not deliberately).

The real-life images (in Fig. 7) contain not only the source images mixture, but also some natural “additive noise.” We refer to any diversion from the assumed mixture model (2) (linear with spatial shifts) as additive noise. Possible noise-sources are: thermal noise, quantization, saturation and more. The thermal noise may be considered as additive white Gaussian noise (AWGN), uncorrelated with the images. However, all other noise sources are not necessarily uncorrelated with the photographed images, or spatially white.

Thus, to mitigate any potential effect on the estimation accuracy, we incorporated some smoothing into the edge-sharpening preprocessing as follows: After subtraction of the mean we applied a Gaussian filter \((7 \times 7\) with \(\sigma = 1\), \texttt{fspecial(‘gaussian’,7,1)} in Matlab) for noise reduction and then a Laplacian filter \((3 \times 3\) with \(\alpha = 0.5\), \texttt{fspecial(‘laplacian’,0.5)} in Matlab) for the edge enhancement. We stress again, that the combined effect of these LSI filters is merely equivalent to introduction of frequency-dependent weights into the AJD process. To mitigate the noise effects on the separation we employed the MMSE reconstruction scheme with \(\mu = 0.007\) (we would have employed it anyway, since the true mixture is very likely to have been singular).

The initial guess for \(A\) was \(\hat{A}_0 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} 0.1 & -0.1 \\ -0.1 & 0.1 \end{bmatrix}\), reflecting the (reasonable) \textit{a priori} knowledge that there was most no change in lighting conditions between the two snapshots. The presumed slight differences in intensities are arbitrary and are not based on any prior knowledge. They are merely aimed at shifting the initial guess of \(A\) from singularity, since singularity in the first iterations sometimes slows down the convergence.

For each experiment (“teddy-bear”/“camera”) we used two snapshots. Image sizes were \(M \times N = 501 \times 351\) for the “teddy-bear” and \(M \times N = 371 \times 301\) for the “camera.” The correlation window size for both was \(\bar{M} \times \bar{N} = 25 \times 25\), and the 2-D AC-DC frequency-grid was of size \(I \times J = 51 \times 51\), equally spaced in \((-\pi, \pi) \times (-\pi, \pi)\). The reconstructed images are presented in Fig. 8. Although the separated images are noisier than those attained in the synthetic mixtures, the separation is easily seen to be perceptually correct. The running time for the entire scheme was roughly the same as with the synthetic images, about 13.0 s.

E. Separation of Dissolved Video Sequences

Sometimes in movies, when passing from one scene to another, a crossfade effect is used (i.e., one scene fades-out and the other fades-in). Our 2-D AC-DC is applicable also for separating this kind of dissolved video sequences, especially if there is panoramic camera movement in one (or both) of the dissolved scenes.

Each frame of the video sequence may be considered as a mixture. The \(p\)-th frame of the video sequence can be modeled as

\[
x_p = a_{p1}s_1 (u - \hat{d}_{p1}^{u}, v - \hat{d}_{p1}^{v}) + a_{p2}s_2 (u - \hat{d}_{p2}^{u}, v - \hat{d}_{p2}^{v})
\]

\(p = 1, \ldots, P\)
where \( P \) is the number of frames comprising the effect. The separation algorithm may be applied to any two frames of the movie. Hence, we can divide the movie into frame pairs, and separate each such pair, using the 2-D AC-DC algorithm.

Quite commonly, however, the crossfade effect is (nearly) linear; i.e., the fade-in/out coefficients \( a_{p1}, a_{p2} \) change linearly in time (that is, linearly in the frame index, \( p \)). If the camera movement is also (nearly) linear, we may use this \textit{a priori} knowledge to improve the separation results. Instead of dividing the frames into frame pairs, and applying the 2-D AC-DC algorithm to all of the frame pairs, we can pick a small number of frame pairs. Then, we may estimate the linear mixing coefficients and spatial shifts only for each of these selected frame pairs. Based on the estimated mixing parameters of the selected frame pairs, we can then estimate the mixing parameters of the remaining frame pairs, using linear interpolation/extrapolation. In fact, this may be also done using a single selected pair. Using this simplified procedure, we separated a synthetically generated dissolved video sequence.

A sequence of five representative frames (out of 50 frames comprising the entire sequence) of the mixed and separated sequence appears in Fig. 9. The full video sequences can be downloaded/viewed (in the form of an *.avi file using Cinepak compression) at [18].

	errorable of blind separation of image reflections. In most of the related BSS algorithms, some kind of diversity between the observed images is exploited in order to attain separation. Common diversity sources are polarization and change in lighting conditions. In this work, we introduced an additional (newly considered) source of diversity, in the form of relative spatial-shifts. Relative spatial-shifts of the sources between snapshots may be an unintentional outcome of movement of either the reflective medium or recorded objects. Alternatively, they may be introduced deliberately, e.g., by slightly slanting the window between snapshots (the tilt-angle must be kept small, so as not to deform the shifted scene by contraction due to projection).

The 2-D AC-DC algorithm is an integrated method to estimate both the mixing coefficients and the spatial-shifts. Since it exploits more than one source of diversity, the algorithm performs reasonably well even when either one of these diversity sources is deficient.

We formulated the problem of blind separation of shifted and linearly mixed images, as a specially parameterized AJD problem of 2-D spectra matrices. AJD was applied at all frequencies together, with one target-matrix at each frequency, where the diagonalizing matrices obey the model-prescribed frequency dependence. Thus, our approach is free of frequency-dependent permutation/scale ambiguities, sometimes encountered in other frequency-domain approaches (for general convolutive mixtures).

The underlying “working assumption” of our approach is that the sources are WSS and mutually uncorrelated. The stationarity assumption is usually quite unrealistic for images. Fortunately, however, for our approach, the only truly essential stationarity is with respect to the zero cross-correlations; stationarity with respect to the autocorrelations is not as important.

We demonstrated the successful performance of the proposed algorithm both with synthetic mixtures and with real-life reflec-
tion images. Although admittedly not suitable for the reflection removal problem in its most general form, we believe that our model and associated solution approach can successfully cover a variety of restrictive, yet realistic scenarios.

REFERENCES


Efrat Be’ery was born in Petach-Tikva, Israel, in 1981. She received the B.Sc. degree (magna cum laude) in electrical engineering and the M.Sc. (magna cum laude) degree from Tel-Aviv University, Tel-Aviv, Israel, in 2005 and 2007, respectively. Since 2006, she has been with POLYCOM, Inc., Petach-Tikva, as an Algorithm Engineer in the fields of video processing and audio processing.

Arie Yeredor (M’99–SM’02) was born in Haifa, Israel in 1963. He received the B.Sc. degree in electrical engineering (summa cum laude) and the Ph.D. degree from Tel-Aviv University (TAU), Tel-Aviv, Israel, in 1984 and 1997, respectively. He is currently a Senior Lecturer in the Department of Electrical Engineering—Systems, School of Electrical Engineering, TAU, where he teaches courses in statistical and digital signal processing. He also holds a consulting position with NICE Systems, Inc., Ra anana, Israel, in the fields of speech and audio processing, video processing, and emitter location algorithms. His research interests include estimation theory, statistical signal processing, and blind source separation. Dr. Yeredor has been awarded the yearly Best Lecturer of the Faculty of Engineering Award six times by TAU. He is an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—II: EXPRESS BRIEFS, and a member of the Signal Processing Society’s Signal Processing Theory and Methods Technical Committee.