Blind Separation of Time/Position Varying Mixtures

Ran Kaftory and Yehoshua Y. Zeevi

***Abstract*— We address the challenging open problem of blindly separating time/position varying mixtures, and attempt to separate the sources from such mixtures without having prior information about the sources or the mixing system. Unlike studies concerning instantaneous or convolutive mixtures, we assume that the mixing system (medium) is varying in time/position. Attempts to solve this problem have mostly utilized, so far, online algorithms based on tracking the mixing system by methods previously developed for the instantaneous or convolutive mixtures. In contrast with these attempts, we develop a unified approach in the form of staged sparse component analysis (SSCA). Accordingly, we assume that the sources are either sparse or can be "sparsified." In the first stage, we estimate the filters of the mixing system, based on the scatter plot of the sparse mixtures' data, using a proper clustering and curve/surface fitting. In the second stage, the mixing system is inverted, yielding the estimated sources. We use the SSCA approach for solving three types of mixtures: time/position varying instantaneous mixtures, single-path mixtures, and multipath mixtures. Real-life scenarios and simulated mixtures are used to demonstrate the performance of our approach.**

***Index Terms*— Blind source separation (BSS), sparse component analysis (SCA), time/osition varying mixing/unmixing.**

I. Introduction

EXTENSIVE research has been devoted over the last two decades to the subject of blind source separation (BSS), especially in the context of independent component analysis (ICA). The research mainly focused on the stationary, instantaneous, and convolutive theoretical aspects of the problem and on practical applications. In the relevant models, the sources are attenuated by a fixed factor, and/or filtered by a fixed (in time/position) filter prior to being mixed. Two important categories of the approaches to solving the BSS problem of stationary mixtures are ICA and sparse component analysis (SCA). ICA assumes that the sources are statistically independent and, therefore, utilizes separation cost functions based on the maximization of nonGaussianity [1], negentropy [2], maximum likelihood [3], minimization of the mutual information [4], [5], diagonalization of the cumulant tensor [6], nonlinear decorrelation [7], and second order statistics [8]. Blind separation using SCA assumes that the sources are sparse or can be sparsified. The sources need not be statistically independent. This approach lends itself to a geometric interpretation of the mixing coefficients, whereby the mixing matrix entries can be retrieved from the scatter plot of the sparsified mixtures [9].

In most real-life scenarios, the mixing system is not constant as is in the case of instantaneous or convolutive model. It is varying as a function of time or position. For example, the attenuation of signals/images varies over time/position thus creating time/position varying instantaneous mixtures. The delay/shift or reverberation/blurring of a signal/image may also vary over time/position, creating a time/position varying single/multipath mixtures. Only few studies address this generalized BSS problem. Most of them use the ICA approach and assume a slow varying mixing system, thus, enabling the use of an adaptive version of the algorithms developed for the stationary cases.

In this paper, we extend and generalize the BSS problem and provide a unified approach to blind separation of certain classes of time/position varying mixtures that have not been dealt with so far. To this end, we present a framework of staged SCA (SSCA).

II. Problem Formulation

We consider the case of separating N_s sources from N_z linear time/position varying mixtures, where the number of the sources and mixtures $N_s = N_z$ is small. We use the vector notation for the independent variables $\underline{\xi} = (\xi_1, \ldots, \xi_{N_\xi})$, $\underline{\xi}' = (\xi'_1, \ldots, \xi'_{N_\xi})$, where N_ξ is the number of independent variables. Denoting $Z = [z_1(\underline{\xi}), \ldots, z_{N_z}(\underline{\xi})]$ as a vector of the observed mixtures, and $S = [s_1(\underline{\xi}), \ldots, s_{N_s}(\underline{\xi})]$ as a vector of the sources, the observed mixtures are generated from the sources by the linear time varying transformation

$$Z = H \star S + \eta \tag{1}$$

where

$$H = \begin{bmatrix} h_{11}(\underline{\xi}, \underline{\xi}') & \cdots & h_{1N_s}(\underline{\xi}, \underline{\xi}') \\ \vdots & h_{ij}(\underline{\xi}, \underline{\xi}') & \vdots \\ h_{N_z 1}(\underline{\xi}, \underline{\xi}') & \cdots & h_{N_z N_s}(\underline{\xi}, \underline{\xi}') \end{bmatrix}$$

is a mixing matrix of filters, η is some unknown noise, and the symbol $\star$ is defined as follows.

Definition 1: The symbol $\star$ denotes an integral operator with multidimensional kernel function acting on a single or multivariable function. This operator is used for representing the linear filtering of a signal with a time/position varying filter and composition of such time/position varying filters as follows.

Manuscript received July 14, 2010; revised December 19, 2010; accepted March 8, 2012. Date of publication May 1, 2012; date of current version December 20, 2012. This work was supported in part by the Ollendorff Minerva Center for Vision and Image Sciences. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Oscar C. Au.

The authors are with the Department of Electrical Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel (e-mail: kaftoryr@tx.technion.ac.il; zeevi@ee.technion.ac.il).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2012.2197005

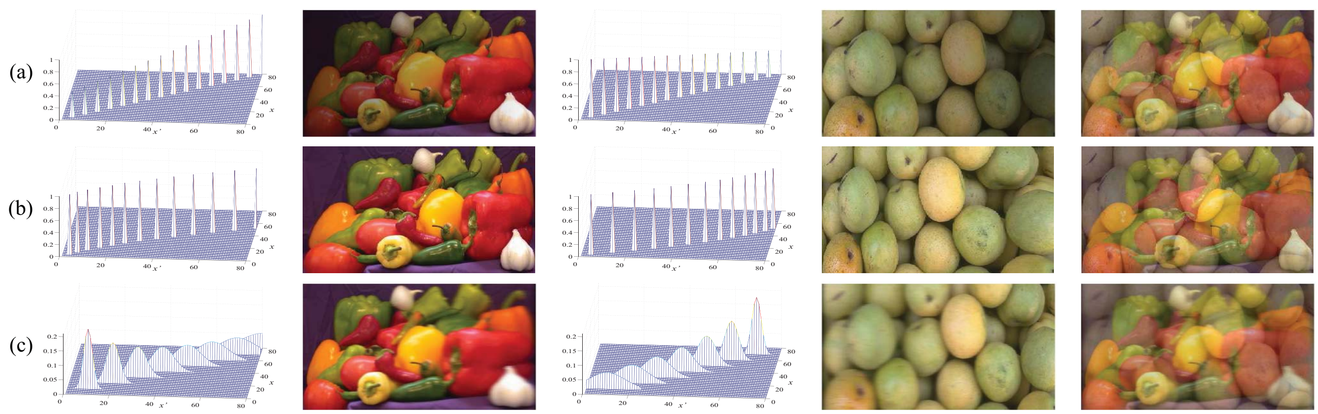


Fig. 1. Types of position varying image mixtures, where the system varies only along the x-axis. (a) Instantaneous case. (b) Single-path case. (c) Multipath case. From left to right: the filter $h_{11}(x, x')$, result of filtering one of the sources with this filter, the filter $h_{12}(x, x')$, result of filtering the other source with this filter, the mixture. Observe how the following properties of the filtered image changes along the x-axis: intensity for the instantaneous case, scale for the single-path case, and blur for the multipath case.

1) Filtering with a filter ($\star : \mathbb{R}^{2N_\xi} \star \mathbb{R}^{N_\xi} \to \mathbb{R}^{N_\xi}$)

$$h \star s \equiv \int_{-\infty}^{\infty} h(\underline{\xi}, \underline{\xi}')s(\underline{\xi}')d\underline{\xi}'. \quad (2)$$

2) Composition of filters ($\star : \mathbb{R}^{2N_\xi} \star \mathbb{R}^{2N_\xi} \to \mathbb{R}^{2N_\xi}$)

$$g \star h \equiv \int_{-\infty}^{\infty} g(\underline{\xi}, \underline{\xi}'')h(\underline{\xi}'', \underline{\xi}')d\underline{\xi}'' \quad (3)$$

where $g(\underline{\xi}, \underline{\xi}')$ and $h(\underline{\xi}, \underline{\xi}')$ are time/position varying filters used for filtering N_ξ-dimensional signals.

Similar to the definition of [10], we distinguish between three types of time/position varying mixtures.

1) *Time/Position Varying Instantaneous Mixtures:* In this case, only the attenuation of the signals varies over time/position. We also assume that the signals arrive at the sensors instantaneously (in the case of images, without position shifts) and that they do not reverberate, nor do they have multiple reflections. The filters of the mixing system are given explicitly as

$$h_{ij}(\underline{\xi}, \underline{\xi}') = a_{ij}(\underline{\xi})\delta(\underline{\xi} - \underline{\xi}') \quad (4)$$

where δ is the Dirac delta, and $a_{ij}(\underline{\xi})$ is the time/position varying attenuation of the jth source with respect to the ith sensor. Using the definition of $h_{ij}(\underline{\xi}, \underline{\xi}')$, (1) produces the following results for the generation of the time/position varying instantaneous mixtures:

$$z_i(\underline{\xi}) = \sum_j a_{ij}(\underline{\xi})s_j(\underline{\xi}) + \eta(\underline{\xi}). \quad (5)$$

2) *Time/Position Varying Single-Path Mixtures:* In this case, the attenuation and the delay/position shift of the signals/images change over time/position. This creates the doppler effect or the zooming/streching of images. We assume in this case, as in the instantaneous case, the absence of reverberations/multipath/blurring. The filters of the mixing system are given explicitly as

$$h_{ij}(\underline{\xi}, \underline{\xi}') = a_{ij}(\underline{\xi})\delta(d_{ij}(\underline{\xi}) - \underline{\xi}') \quad (6)$$

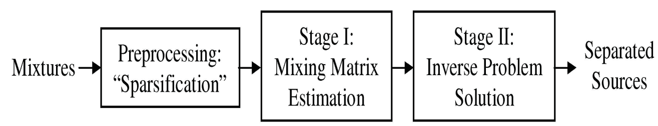


Fig. 2. SSCA method.

where $d_{ij}(\underline{\xi})$ is the time/position varying delay/position shift of the jth source with respect to the ith sensor. Using the definition of $h_{ij}(\underline{\xi}, \underline{\xi}')$, (1) yields the following results for the generation of the time/position varying single-path mixtures:

$$z_i(\underline{\xi}) = \sum_j a_{ij}(\underline{\xi})s_j(d_{ij}(\underline{\xi})) + \eta(\underline{\xi}). \quad (7)$$

3) *Time/Position Varying Multipath Mixtures:* In this case, the attenuation, delay/position shift, and reverberation/multipath/blur may change over time where the function of h_{ij} is arbitrary.

Fig. 1 provides a graphical (visual) interpretation of the above three types of position varying mixtures. Two image sources are mixed using different filters, $h_{ij}(x, x')$ which vary from left to right. In the instantaneous case, only the attenuation of each image varies. In the single-path mixture, varying the position shift varies the scale of the images. In the case of multipath mixtures, we use a variable low-pass filter, which blurs the image differently from left to right.

III. SSCA

Without loss of generality, we consider a system of two mixtures, $z_1(\underline{\xi}), z_2(\underline{\xi})$, of two sources, $s_1(\underline{\xi}), s_2(\underline{\xi})$ and adopt the SSCA approach for solving the time/position varying BSS problem as outlined in Fig. 2 and in the sequel.

A. Preprocessing: Sparsification

Sparse signals are defined as those signals whose value differs significantly from zero only in a few instances of time/position. The probability density function (PDF) of such signals is modeled as an exponential [9] (also known as the generalized Gaussian). The joint distribution of uncorrelated sparse signals is also approximately exponential. Therefore, if one source is active in/at a certain instance/position, the probability that any one of the other sources is active in the same instance/position is very small. For signals which are not naturally sparse, a proper sparsification transformation which is invariant to the mixing matrix can be applied. Sparsification is defined as follows.

Definition 2: Sparsification of a signal $s(\underline{\xi})$ is the process of applying a transformation $\top$ to the nonsparse signal $s(\underline{\xi})$ that yields a sparse output $\top[s(\underline{\xi})]$.

We define invariance of the sparsification transformation to the mixing matrix as follows.

Definition 3: A sparsification transformation $\top$ is invariant under mixing by the matrix, H, if

$$\top[H \star S] = H \star \top[S]. \tag{8}$$

The required preprocessing, according to the SSCA approach, implements an invariant sparsification transformation on both sides of (1), yielding the following separation problem:

$$\top[Z] = H \star \top[S]. \tag{9}$$

B. First Stage of the Separation Process: Mixing Matrix Estimation

In the noiseless case, all the time/position instances wherein the first sparsified source is active and the other is not-active satisfy the relations

$$\top[z_1] = h_{11} \star \top[s_1], \qquad \top[z_2] = h_{21} \star \top[s_1]. \tag{10}$$

We define the new position-varying filter, g_1 such that

$$\top[z_2] = g_1 \star \top[z_1]. \tag{11}$$

Equations (10) and (11) yield

$$h_{21} = g_1 \star h_{11}. \tag{12}$$

Similarly, for all the time/position instances in which the second sparsified source is active and the first is not, we define g_2 as $\top[z_2] = g_2 \star \top[z_1]$ and obtain $h_{22} = g_2 \star h_{12}$.

Even for the noisy case, it is possible to filter the time/position instances (usually by introducing some threshold) over which one of the sources can be regarded as active with respect to the not-active source, and where the noise is negligible.[1] The filters g_1 and g_2 are found by using a scatter plot of the sparse mixtures' data. A proper method of clustering these instances and curve/surface estimation is used to estimate g_1 and g_2 out of the these active/not-active samples of $\top[z_1]$ and $\top[z_2]$.

The first step of the SSCA is completed by substituting h_{21} and h_{22} with $g_1 \star h_{11}$ and $g_2 \star h_{12}$. This yields the following BSS problem:

$$\begin{aligned}\top[z_1] &= h_{11} \star \top[s_1] + h_{12} \star \top[s_2] + \eta_1\\ \top[z_2] &= (g_1 \star h_{11}) \star \top[s_1] + (g_2 \star h_{12}) \star \top[s_2] + \eta_2.\end{aligned} \tag{13}$$

[1]After the sparsification preprocessing, the SNR is improved.

Using the associativity property of the operator $\star$ yields

$$\begin{aligned}\top[z_1] &= h_{11} \star \top[s_1] + h_{12} \star \top[s_2] + \eta_1\\ \top[z_2] &= g_1 \star (h_{11} \star \top[s_1]) + g_2 \star (h_{12} \star \top[s_2]) + \eta_2.\end{aligned} \tag{14}$$

Denoting $s_1' \equiv h_{11} \star s_1$ and $s_2' \equiv h_{12} \star s_2$, and recalling the invariance property of the sparsification operator, (14) is rewritten as

$$\begin{aligned}\top[z_1] &= \top[s_1'] + \top[s_2'] + \eta_1\\ \top[z_2] &= g_1 \star \top[s_1'] + g_2 \star \top[s_2'] + \eta_2\end{aligned} \tag{15}$$

or in the context of the vector notation of (1)

$$\top[Z] = H' \star \top[S'] + \eta \tag{16}$$

where

$$H' = \begin{bmatrix} \boldsymbol{1} & \boldsymbol{1} \\ g_1 & g_2 \end{bmatrix}$$

and $\boldsymbol{1}$ is the identity filter (an ideal all-pass filter) which is defined as: $\boldsymbol{1} \equiv \delta(\underline{\xi} - \underline{\xi}')$.

We conclude this exposition with the analysis of a closely related BSS problem, wherein the filters g_1 and g_2 of the mixing matrix have already been estimated. The sources s_1' and s_2' are filtered versions of s_1 and s_2, filtered by the time/position varying filters h_{11} and h_{12}, respectively. It is proven in [11] that this is the upper bound on the quality of estimating the sources.

C. Second Stage of the Separation Process: Solving the Inverse Problem

The task of the second stage of the SSCA is to solve the inverse problem of $Z = H' \star S' + \eta$, and estimate the sources. A "naive" approach would be to directly invert the system

$$\hat{S}' = H'^{-1} \star Z \tag{17}$$

where

$$H'^{-1} \equiv (g_1 - g_2)^{-1} \star \begin{bmatrix} g_2 & -\boldsymbol{1} \\ -g_1 & \boldsymbol{1} \end{bmatrix}.$$

However, matrix inversion and the calculation of the determinant is feasible only for commutative rings, which is not the general case of sets of time/position varying filters. Even for a subset of time/position varying filters, which commute, direct inversion requires the inversion of the difference filter $g_1 - g_2$. This filter can be singular or badly conditioned, in which case, its inverse amplifies the noise and yields noisy estimation of the sources. To overcome this problem, we use a variational method which is effective in both cases of noncommutative mixing time/position varying filters and singular or badly conditioned mixing matrix. Instead of inverting the system which amplifies the noise, a cost function is defined

$$\mathbb{J} = \|Z - H' \star \hat{S}'\| + w\mathbf{R}(\hat{S}') \tag{18}$$

where $\mathbf{R}(\cdot)$ is a regularization operator, which suppresses the amplification of the noise[2] and w is a weighting parameter. The minimization of this cost function using the Euler–Lagrange equations yields an estimation of the separated sources.

[2]See Section VI-C for an example of a regularization operator, (52).

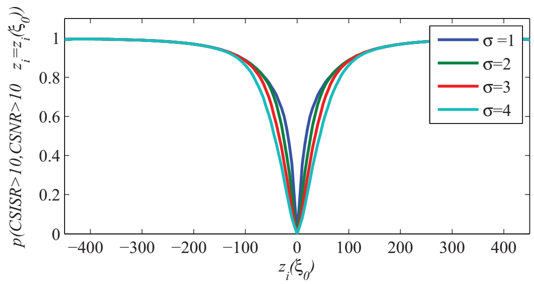


Fig. 3. Conditional probability that one source is active and the other is not (and the noise is negligible) given the thresholds th_1, th_2, and the observed mixture plotted for several values of σ given sparse signal distribution with $\mu = 1.5$ and $\nu = 2$. Observe how the probability is symmetric nondecreasing function of the absolute value of the observed mixtures.

IV. Conditions Which Enable the Use of the SSCA

In order to successfully apply the SSCA approach in separation of time/position varying mixtures, proper conditions must be fulfilled. In this section, we outline the conditions which concern the sparsity of the sources signals, signal-to-noise ratio, the filters of the mixing matrix, sparsification transformation, and the mixing matrix itself.

A. Conditions on the Noise and Signal Sources

The first step of the SSCA requires to filter time/position instances where one sparsified source is active while the others are not, and the noise is negligible. In the sequel, we show that for sparsified sources this probability increases as the absolute value of the observed mixture increases. For clarity, we show this for a mixture of two sources. We define the following criteria for the above instances.

Definition 4: A time/position instance $\underline{\xi}_0$ in which one sparsified source can be regarded as active and the other is not is defined as follows:

$$\left| \frac{h_{i1} \star \top[s_1(\underline{\xi}_0)]}{h_{i2} \star \top[s_2(\underline{\xi}_0)]} \right| > th_1 \tag{19}$$

where the ratio on the left-hand side is termed contributing-signal-to-interfering-signal ratio (CSISR) and th_1 is a subjective threshold, which depends on the robustness of the algorithm for estimating the filters of the mixing matrix as well as the filters themselves.

Definition 5: A time/position instance $\underline{\xi}_0$ in which the sparsified noise is negligible with respect to the mixture of the two sparsified sources is defined as

$$\left| \frac{h_{i1} \star \top[s_1(\underline{\xi}_0)] + h_{i2} \star \top[s_2(\underline{\xi}_0)]}{\top[\eta_i]} \right| > th_2 \tag{20}$$

where the ratio on the left-hand side is termed contributing signals-to-noise ratio (CSNR) and th_2 is a subjective threshold, which depends on the robustness of the algorithm for estimating the filters of the mixing matrix as well as the filters themselves.

The PDF of a sparsified source is modeled as an exponential [12], $\mathrm{pdf}(\top[s]) = ke^{-\mu|\top[s]|^{1/\nu}}$, where μ is a positive parameter, $\nu \geq 1$, and $k = (\int_{-\infty}^{\infty} e^{-\mu|\top[s]|^{1/\nu}})^{-1}$ is a normalization parameter [9].

We assume that the noise is not sparse even after applying the sparsification transformation. Therefore, we can assume that the PDF of the noise is Gaussian with zero mean: $\mathrm{pdf}(\top[\eta]) = ke^{-(\top[\eta]^2/2\sigma^2)}$, where σ is the standard deviation of the noise, and $k = (1/\sqrt{2\pi\sigma})$ is a normalization parameter.

Proposition 1: The conditional probability that one sparsified source is active and the other is not, where the noise is negligible in some time/position instance $\underline{\xi}_0$, given the observed sparsified mixture $\top[z_i(\underline{\xi}_0)]$, and the thresholds th_1 and th_2, is

$$\begin{aligned} &p\left(\mathrm{CSISR} > th_1, \mathrm{CSNR} > th_2 \mid \top[z_i] = \top[z_i(\underline{\xi}_0)]\right) \\ &= \frac{\int_{m_1}^{m_2}\int_{l_1}^{l_2} e^{-\frac{(\top[z_i(\underline{\xi}_0)]-\underline{\zeta})^2}{2\sigma^2} - |\underline{\zeta}-\underline{\phi}|^{1/\nu_{i1}}\mu_{i1}^T - |\underline{\phi}|^{1/\nu_{i2}}\mu_{i2}^T} d\underline{\phi}\, d\underline{\zeta}}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{(\top[z_i(\underline{\xi}_0)]-\underline{\zeta})^2}{2\sigma^2} - |\underline{\zeta}-\underline{\phi}|^{1/\nu_{i1}}\mu_{i1}^T - |\underline{\phi}|^{1/\nu_{i2}}\mu_{i2}^T} d\underline{\phi}\, d\underline{\zeta}} \\ &+ \frac{\int_{m_1}^{m_2}\int_{l_3}^{l_4} e^{-\frac{(\top[z_i(\underline{\xi}_0)]-\underline{\zeta})^2}{2\sigma^2} - |\underline{\zeta}-\underline{\phi}|^{1/\nu_{i1}}\mu_{i1}^T - |\underline{\phi}|^{1/\nu_{i2}}\mu_{i2}^T} d\underline{\phi}\, d\underline{\zeta}}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{(\top[z_i(\underline{\xi}_0)]-\underline{\zeta})^2}{2\sigma^2} - |\underline{\zeta}-\underline{\phi}|^{1/\nu_{i1}}\mu_{i1}^T - |\underline{\phi}|^{1/\nu_{i2}}\mu_{i2}^T} d\underline{\phi}\, d\underline{\zeta}} \end{aligned} \tag{21}$$

where T in μ_{ij}^T stands for the transpose of μ_{ij}. For $\underline{\zeta} \geq 0$

$$l_1 \equiv \frac{\underline{\zeta}}{1-th_1}, \quad l_2 \equiv \frac{\underline{\zeta}}{1+th_1}, \quad l_3 \equiv \frac{\underline{\zeta}th_1}{th_1+1}, \quad l_4 \equiv \frac{\underline{\zeta}th_1}{th_1-1}$$

and for $\underline{\zeta} < 0$, l_1 and l_2 are interchanged and so are l_3 and l_4.

If $z_i(\underline{\xi}_0) \geq 0$

$$m_1 = \frac{\top[z_i(\underline{\xi}_0)]th_2}{th_2+1}, \quad m_2 = \frac{\top[z_i(\underline{\xi}_0)]th_2}{th_2-1}$$

and, if $z_i(\underline{\xi}_0) < 0$, m_1 and m_2 are interchanged. (for the proof see [11].)

Fig. 3 depicts the conditional probability of sparsified signal distribution, with $\mu = 1.5$ and $\nu = 2$ plotted as a function of $\top[z_i(\underline{\xi}_0)]$ for several values of σ. It is observed that the probability is a monotonic nondecreasing function of $|\top[z_i(\underline{\xi}_0)]|$. This result is also valid for the case of a mixture of more than two sources. Therefore, as a rule of thumb, it is possible to use a proper threshold on the observed mixture such that the probability that only one source is active, and the noise is negligible, is high. In such instances, the estimation of the filters of the mixing matrix is possible. The number of such instances, which is required for the correct estimation of the filter is discussed in reference to the context of conditions required for the filters estimation.

B. Conditions on the Filters of the Mixing Matrix

The first step of the SSCA approach, requires the existence of g_i and its estimation based on samples of the observed measurements. Assuming that g_i exists, we can obtain only samples of the filter using a threshold over which we can assume a high probability for only one source to be active, and the noise to be negligible. These samples are distributed randomly and depend on both the source signals and the noise. If we assume that we know the filter g_i up to some finite number N_k of unknown parameters, and the function is injective, then we need at least the same number (N_k) of samples of that filter, in order to obtain the unknown parameters. If the filter is unknown, a Taylor approximation can be derived using some N_k-order polynomial function

$$\begin{aligned}\hat{g}_i(\xi_1,\ldots,\xi_{N_\xi},\xi'_1,\ldots,\xi'_{N_\xi}) \\ = \sum_{k_1=0}^{N_k}\cdots\sum_{k_{N_k}=0}^{N_k}\sum_{k'_1=0}^{N_k}\cdots\sum_{k'_{N_k}=0}^{N_k}\alpha_{k_1,\ldots,k_{N_k},k'_1,\ldots,k'_{N_k}}(\xi_1-\xi_{1_0})^{k_1} \\ \cdots(\xi_{N_\xi}-\xi_{N_{\xi_0}})^{k_{N_k}}(\xi'_1-\xi'_{1_0})^{k'_1}\cdots(\xi'_{N_\xi}-\xi'_{N_{\xi_0}})^{k'_{N_k}}. \quad (22)\end{aligned}$$

It takes $(N_k+1)^{2N_\xi}$ samples to calculate the unknowns α parameters by the Lagrangian interpolation formula, [13].

A nonparametric approach for estimating g_i out of its samples would require the following condition.

Proposition 2: If $g_i(\underline{\xi},\underline{\xi}')$ is band-limited to a region Ω, it can be uniquely represented by nonuniformly distributed samples satisfying the Nyquist rate, on the average, as long as the samples are not on the zero-crossing contours of any member of the set of all functions band-limited to Ω (the proof is given in [14]).

In practice, the function g_i is usually known up to a finite number of parameters, using some prior knowledge regarding the physics of the mixing system. In other cases, it can be assumed that it is smooth and therefore band-limited to a small region Ω, which allows the use of a nonparametric function estimation by means of a low average sampling rate.

By definition, $g_1 = h_{21}\star h_{11}^{-1}$ and $g_2 = h_{22}\star h_{12}^{-1}$, therefore, the filters g_1 and g_2 exist only if h_{11}^{-1} and h_{12}^{-1} exist. We now provide the conditions for the existence of an inverse filter by using the frequency analysis of time/position varying systems.

Zadeh [15] was the first to present an approach suitable for the analysis of linear time-varying systems in the context of frequency analysis. His transform has rarely been used [16] but it is useful for our purposes. We consider the linear time/position varying system representation as the superposition integral: $z(\underline{\xi}) = h(\underline{\xi},\underline{\xi}')\star s(\underline{\xi})$, where $h(\underline{\xi},\underline{\xi}')$ is the system response at the instance $\underline{\xi}$ to the impulse $\delta(\underline{\xi}-\underline{\xi}')$. Zadeh defined the frequency response of the time/position varying system, to a unit impulse applied at $\underline{\xi}=\underline{\xi}_0$ as

$$\mathcal{H}_Z(\underline{\xi}_0,\underline{\omega}) = \int_{-\infty}^{\infty} h(\underline{\xi}_0,\underline{\xi}')e^{-j\underline{\omega}(\underline{\xi}_0-\underline{\xi}')^T}d\underline{\xi}'. \quad (23)$$

Zadeh's transform can be used for the synthesis of variable systems by using the system frequency response, as indicated by the following proposition.

Proposition 3: The output of the system $z(\underline{\xi}) = h(\underline{\xi},\underline{\xi}')\star s(\underline{\xi})$ can be calculated using the system frequency response

$$z(\underline{\xi}) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathcal{H}_Z(\underline{\xi},\underline{\omega})\mathcal{S}_F(\underline{\omega})e^{j\underline{\omega}\underline{\xi}^T}d\underline{\omega} \quad (24)$$

where $\mathcal{S}_F(\underline{\omega})$ is the Fourier transform of $s(\underline{\xi})$: $\mathcal{S}_F(\underline{\omega}) = \int_{-\infty}^{\infty}s(\underline{\xi})e^{-j\underline{\omega}\underline{\xi}^T}d\underline{\xi}$ (for the proof, see [11] or [17]).

It is important to find the frequency response of a composition of time/position varying filters. The following Theorem provides the relations between the frequency response of a system to the frequency response of its components.

Lemma 4: The Zadeh frequency response of the system: $f = g\star h$, is given by the Zadeh frequency responses, $\mathcal{G}_Z(\underline{\xi},\underline{\omega})$ and $\mathcal{H}_Z(\underline{\xi},\underline{\omega})$ as

$$\mathcal{F}_Z(\underline{\xi},\underline{\omega}) = \sum_{i=0}^{n}\frac{1}{i!}\frac{\partial^i\mathcal{G}_Z(\underline{\xi},\underline{\omega})}{\partial(j\underline{\omega})^i}\frac{\partial^i\mathcal{H}_Z(\underline{\xi},\underline{\omega})}{\partial\underline{\xi}^i} \quad (25)$$

where n is the order of the differential system representing the SISO system (for a proof see [18]).

A condition establishing relationship between the Zadeh frequency responses of the filter and its inverse is defined by the following theorem.

Theorem 5: The time/position varying filter $h(\underline{\xi},\underline{\xi}')$ has an inverse $g(\underline{\xi},\underline{\xi}')$ only if $\mathcal{G}_Z(\underline{\xi},\underline{\omega})$, which is the Zadeh frequency response of $g(\underline{\xi},\underline{\xi}')$, satisfies the equation

$$\sum_{i=0}^{n}\frac{\partial^i\mathcal{G}_Z(\underline{\xi},\underline{\omega})}{\partial(j\underline{\omega})^i}\frac{\partial^i\mathcal{H}_Z(\underline{\xi},\underline{\omega})}{\partial\underline{\xi}^i} = 1 \quad (26)$$

where n is the order of the differential equation (for the proof see [11]).

Therefore, the filters g_1 and g_2 exist only if h_{11}^{-1} and h_{12}^{-1} exist and fulfil (26), where h and g are substituted by h_{11} or h_{12} and h_{11}^{-1} or h_{12}^{-1}, respectively.

C. Conditions on the Sparsification Transformation

A sufficient condition for the existence of a sparsification transformation is its invariance to the mixing matrix. We provide a weaker explicit condition sufficient for a subset of sparsification transformations, which can be regarded as filtering with a time/position varying filter: $\top[H\star S] = \top\star(H\star S)$.

Recalling Definition 3 with this subset of transformations, the invariance to the mixing matrix means that $\top\star(H\star S) = H\star(\top\star S)$, or explicitly for a 2×2 mixing system

$$\top\star(h_{i1}\star s_1 + h_{i2}\star s_2) = h_{i1}\star(\top\star s_1) + h_{i2}\star(\top\star s_2). \quad (27)$$

This implies that a transformation which is invariant to the mixing matrix is commutative over $\star$.

Proposition 6: The operator $\star$ along with the set of $2N_\xi$-dimensional time/position varying filters perform a noncommutative operation: $g\star h \neq h\star g$, where g and h are time/position varying filters (for the proof see [11]).

Therefore, in general, by substituting g with $\top$ and h with h_{i1}, $\top\star h_{i1}\neq h_{i1}\star\top$. A similar noncommutative operation is obtained for h_{i2}. But, using the Zadeh transform, we can find a subset of transformations which do commute.

Corollary 7: The filters h and g commute, if their relative Zadeh frequency responses obey

$$\sum_{i=0}^{n}\frac{1}{i!}\frac{\partial^i\mathcal{G}_Z(\underline{\xi},\underline{\omega})}{\partial(j\underline{\omega})^i}\frac{\partial^i\mathcal{H}_Z(\underline{\xi},\underline{\omega})}{\partial\underline{\xi}^i} = \sum_{i=0}^{n}\frac{1}{i!}\frac{\partial^i\mathcal{H}_Z(\underline{\xi},\underline{\omega})}{\partial(j\underline{\omega})^i}\frac{\partial^i\mathcal{G}_Z(\underline{\xi},\underline{\omega})}{\partial\underline{\xi}^i}. \quad (28)$$

The proof follows directly from Lemma 4. This result implies that filters which have a fixed-in-$\underline{\xi}$ Zadeh frequency response, commute with similar filters [this is because $(\partial^i \mathcal{H}_Z(\underline{\xi}, \underline{\omega}))/(\partial \underline{\xi}^i) = 0$ and $(\partial^i \mathcal{G}_Z(\underline{\xi}, \underline{\omega}))/(\partial \underline{\xi}^i) = 0$]. Filters which have a fixed-in-$\underline{\omega}$ Zadeh frequency response, commute with similar filters (this is because $(\partial^i \mathcal{G}_Z(\underline{\xi}, \underline{\omega}))/(\partial (j\underline{\omega})^i) = 0$ and $(\partial^i \mathcal{H}_Z(\underline{\xi}, \underline{\omega}))/(\partial (j\underline{\omega})^i) = 0$).

Therefore, the filters $\top$ and h_{i1} or h_{i2} commute if the condition of (28) is fulfilled by substituting g and h with $\top$ and h_{i1} or h_{i2}, respectively.

D. Conditions on the Mixing Matrix

The second step of the SSCA requires the solution of the inverse problem of $Z = H \star S$. It can be accomplished by means of (17), by using the inverse of H. However, matrix algebra, which includes the calculation of the determinant and the inverse of a matrix, is defined only over a commutative ring. In our case, we show that in general, the set of $2N_{\xi}$-dimensional time/position varying filters together with addition and the operator $\star$ (which substitutes multiplication) constitute a noncommutative ring and, therefore, matrix algebra cannot be performed.

A ring of elements is defined with two binary operators: "addition" and "multiplication." It requires the elements to constitute a commutative group over "addition" and a monoid over "multiplication."

Theorem 8: The set of $2N_{\xi}$-dimensional time/position varying filters, and the two binary operators of addition and $\star$ (which serves as "multiplication") constitute a noncommutative ring (for the proof see [11]).

Nevertheless, it is possible to find a subset of $2N_{\xi}$-dimensional time/position varying filters which commute with each other, using Corollary 7. These filters are regarded as their centralizer, where the centralizer is defined as follows.

Definition 6: The centralizer of an element, h, of a monoid associated with the operator $\star$, is the set of all the elements g_i in that monoid which obey: $h \star g_i = g_i \star h$.

Therefore, matrix algebra can be performed for the subset of Filters, which constitute the centralizer of themselves. As a consequence, the inverse of a matrix can be calculated using the following corollary.

Corollary 9: If the elements of H, a matrix of $N_z - by - N_s$, $N_z = N_s$ time/position varying filters, belong to the centralizer of the matrix filters, and if H is invertible, then the inverse of H is given by

$$H^{-1} = \det(H)^{-1} \star \begin{bmatrix} c_{11} & \cdots & c_{1N_s} \\ \vdots & \ddots & \vdots \\ c_{N_z 1} & \cdots & c_{N_z N_s} \end{bmatrix} \tag{29}$$

where c_{ij} is the matrix cofactor and $\det(H)^{-1}$ is the inverse of the determinant $\det(H)$. The determinant of a matrix of time/position varying filters is calculated the same as in matrices of scalars where the operator $\star$ substitutes multiplication. The result of calculating the determinant is also a time/position varying filter and so is its inverse. (Corollary 9 is proved by Cramer's rule of linear algebra.)

As a result, a matrix of $N_z - by - N_z$ time/position varying filters, H, is invertible only if the determinant of H is invertible. Nevertheless, the existence of H^{-1} is a necessary condition for solving directly the inverse problem, but not sufficient. The matrix H can be ill-conditioned, meaning a small change in the input results in a large change of the output.

Looking at the inverse problem of the second stage in (16), we can observe that for the noisy case

$$\begin{aligned} \begin{bmatrix} \hat{s}'_1 \\ \hat{s}'_2 \end{bmatrix} &= (g_1 - g_2)^{-1} \star \begin{bmatrix} g_2 & -\mathbf{1} \\ -g_1 & \mathbf{1} \end{bmatrix} \star \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \\ &= \begin{bmatrix} s'_1 \\ s'_2 \end{bmatrix} + (g_1 - g_2)^{-1} \star \begin{bmatrix} g_2 & -\mathbf{1} \\ -g_1 & \mathbf{1} \end{bmatrix} \star \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} \\ &= \begin{bmatrix} s'_1 + (g_1 - g_2)^{-1} \star g_2 \star \eta_1 - (g_1 - g_2)^{-1} \star \eta_2 \\ s'_2 - (g_1 - g_2)^{-1} \star g_1 \star \eta_1 + (g_1 - g_2)^{-1} \star \eta_2 \end{bmatrix} \end{aligned} \tag{30}$$

where η_1 and η_2 denote the noise in the respective sensors.

Therefore, if the filter $(g_1 - g_2)^{-1}$ or any of the composed filters $(g_1 - g_2)^{-1} \star g_2$ or $(g_1 - g_2)^{-1} \star g_1$ amplify the noise, the error of estimation due to the noise is large.

We assume that η is a white Gaussian noise that results from a stationary stochastic process. Filtering such noise with a time/position varying filter results in a nonstationary process. Therefore, the error in estimating the sources can be small for some time/position instances or large for others. In the Zadeh transform domain, we can state the following proposition.

Proposition 10: The time/position varying power spectral density of the white Gaussian noise filtered by the time/position varying filter $g(\underline{\xi}, \underline{\xi}')$ is $\mathrm{NSD}|\mathcal{G}_Z(\underline{\xi}, \underline{\omega})|$, where NSD is the noise spectral density (the proof is given in [11]).

Consequently, amplification of the noise occurs whenever $|\mathcal{G}_Z(\underline{\xi}, \underline{\omega})| \geq 1$. The conditions for solving directly the inverse problem using the inverse of the mixing matrix H'^{-1} are therefore:

1) $|Z\{(g_1 - g_2)^{-1}\}| \leq 1$;
2) $|Z\{(g_1 - g_2)^{-1} \star g_1\}| \leq 1$;
3) $|Z\{(g_1 - g_2)^{-1} \star g_2\}| \leq 1$;

where $Z\{\cdot\}$ stands for the Zadeh frequency response. Whenever the conditions for solving directly the inverse problem are not fulfilled, the variational approach can be used.

V. Methods and Results for Blind Separation of Time/Position Varying Instantaneous Mixtures[3]

For this section, we assume that the filters of the mixing matrix have the form of (4). We also assume the following.

1) The model of the instantaneous filters of the mixing matrix is arbitrary, but known up to a finite number of parameters.
2) Apart from some instances, the zero-order Taylor approximation of the instantaneous filters around some instance in a small window, is sufficient to represent the filter over that window.
3) The sources are sparse or can be sparsely represented using an appropriate transform.

[3]Some of the methods and the results of this section were presented in [19].

A. Sparsification Using Wavelet Packets and Short Time Fourier Transform (STFT)

The Zadeh transform of the filters of the mixing matrix, h_{ij}, is denoted as follows:

$$\mathcal{H}_{Z_{ij}}(\underline{\xi}, \underline{\omega}) = \int_{-\infty}^{\infty} a_{ij}(\underline{\xi})\delta(\underline{\xi} - \underline{\xi}')e^{j\underline{\omega}(\underline{\xi}-\underline{\xi}')^T} d\underline{\xi}' = a_{ij}(\underline{\xi}). \quad (31)$$

Observe that the Zadeh frequency response of the filters of the mixing matrix is independent of $\underline{\omega}$. Therefore, using Corollary 7, every sparsification transformation which also has a Zadeh frequency response independent of $\underline{\omega}$, is an element of the centralizer. Elements of the centralizer commute with the filters of the mixing matrix, and are therefore invariant to the mixing matrix. Unfortunately, all the known linear sparsification transformations which can be written as a time/position varying filter extract, usually, the high frequencies of the signals. This implies that they have an $\underline{\omega}$-dependent frequency response and, therefore, are not members of the centralizer, nor are invariant to the mixing matrix.

Nevertheless, we can utilize the second assumption of this section and estimate the filters of the mixing matrix over a small window around some instances (where the assumption holds) as though they are fixed (in $\underline{\xi}$) filters. The Zadeh frequency response of such fixed filters is independent of $\underline{\xi}$. With reference to Corollary 7, if we assume that the filters of the mixing matrix obey $(\partial^i \mathcal{H}_Z(\underline{\xi}, \underline{\omega})/\partial \underline{\xi}^i) = 0$ then they are commutable with sparsification filters, which obey $(\partial^i \mathcal{G}_Z(\underline{\xi}, \underline{\omega})/\partial \underline{\xi}^i) = 0$ [which means that they are also fixed (in $\underline{\xi}$) filters]. Therefore, every sparsification transformation which can be written as a fixed (in $\underline{\xi}$) filter, is a member of the centralizer and, therefore, commutes and is invariant to the mixing matrix over these instances.

The STFT can be used for sparse representation of audio and other harmonic signals, whereas wavelet packets can be used for images [20]. The STFT window size or the support of the filter in the wavelet packets decomposition, can be chosen such that the assumption of a fixed (in $\underline{\xi}$) filter in a small window holds. This produces an invariant transform, since the STFT and the wavelet packet decomposition can be regarded as fixed (in $\underline{\xi}$) filters and in a small window the filters of the mixing matrix can also be regarded as fixed (in $\underline{\xi}$).

B. Parametric Estimation of the Filters of the Mixing Matrix

We assume for the rest of this section, that the mixtures are of 1-D signals. In the case of images, the columns are concatenated to form a single vector. If we group the instances, $\{\xi_l\}$, where one of the sources is active and the other is not, using some threshold, the values of g_i for these instances are according to (11)

$$g_i(\xi_l) \approx \frac{\top[z_i(\xi_l)]}{\top[z_1(\xi_l)]}. \quad (32)$$

As a result, the filter g also has the form of $g_i(\xi, \xi') = a_i(\xi)\delta(\xi - \xi')$. According to the first assumption, the filters of the mixing matrix are known up to a finite number of parameters, which means that: $a_i(\xi) \equiv a_i(\alpha_{i_k}; \xi)$, where α_{i_k} are the unknown parameters.

We wish to construct a unified framework for clustering time/position instances originated from the same sensor for estimating the α_{i_k} parameters. This unified framework is necessary since correct estimation of α_{i_k} solves the clustering problem and vice versa, correct clustering estimates the parameters α_{i_k}.

Suppose that we take N_m time/position instances $\xi_0, \ldots, \xi_{N_m-1}$ in which a signal was detected above the threshold. For each individual instance ξ_l, we define the ratio $r_i(\xi_l) \equiv \top[z_i(\xi_l)]/\top[z_1(\xi_l)]$. Since $r_1 \equiv 1$, the range of the index i for the rest of this section is limited to $i > 1$.

Clearly, if we identify and group the instances, $\{\xi_l\}$, where one of the sources is active and the other is not (meaning the other source and the noise are negligible), then we get according to (32) and the definition of r_i

$$r_i(\xi_l) \approx g_i(\xi_l) \equiv a_i(\alpha_{i_k}; \xi_l). \quad (33)$$

A maximum likelihood approach for estimating α_{i_k} would be to maximize

$$\arg\max_{\alpha_{i_k}} \mathbb{J}(\alpha_{i_k}) \equiv \arg\max_{\alpha_{i_k}} P\left(r_i(\xi_0), \ldots, r_i(\xi_{N_m-1}) \,|\, \alpha_{i_k}\right) \quad (34)$$

where $P\left(r_i(\xi_0), \ldots, r_i(\xi_{N_m-1}) \mid \alpha_{i_k}\right)$ stands for the probability of observing N_m instances of the ratio r_i given the α_{i_k} parameters.

Using Bayes' rule, we can calculate the conditional probability of (34) by

$$P\left(r_i(\xi_0), \ldots, r_i(\xi_{N_m-1}) \mid \alpha_{i_k}\right) = \frac{P\left(\alpha_{i_k} \mid r_i(\xi_0), \ldots, r_i(\xi_{N_m-1})\right) P\left(r_i(\xi_0), \ldots, r_i(\xi_{N_m-1})\right)}{P(\alpha_{i_k})}. \quad (35)$$

Since prior information regarding the distribution of α_{i_k} is not available, a uniform distribution is assumed. Omitting $P(\alpha_{i_k})$, which is in the same for every α_{i_k}, and omitting $P\left(r_i(\xi_0), \ldots, r_i(\xi_{N_m-1})\right)$, which does not depend on α_{i_k}, does not affect the maximization of (35) with respect to α_{i_k}. Therefore, estimating α_{i_k} by using the maximum likelihood approach corresponds to maximizing

$$\arg\max_{\alpha_{i_k}} \mathbb{J}(\alpha_{i_k}) \equiv \arg\max_{\alpha_{i_k}} P\left(\alpha_{i_k} \mid r_i(\xi_0), \ldots, r_i(\xi_{N_m-1})\right). \quad (36)$$

In order to evaluate the conditional probability $P\left(\alpha_{i_k} \mid r_i(\xi_0), \ldots, r_i(\xi_{N_m-1})\right)$, we first construct a density estimation to get the value of $a_i(\alpha_{i_k}; \xi_l)$ on a specific instance, given the measurements $\top[z_i(\xi_l)]$ and $\top[z_1(\xi_l)]$. It can be done using a kernel density estimation

$$\hat{f}(a_i, \xi \mid \top[z_i(\xi_l)], \top[z_1(\xi_l)]) = \frac{1}{N_m L_r L_\xi} K\left(\frac{a_i - r_i(\xi_l)}{h^r}, \frac{\xi - \xi_l}{h^\xi}\right) \quad (37)$$

where K is a multivariate kernel density estimator, and L_r, L_ξ are the kernel supports along the a and ξ axes, respectively.

Using the law of total probability, $\hat{f}(a_i, \xi)$ can be found by calculating

$$\hat{f}(a_i, \xi) = \sum_{l=0}^{N_m-1} \hat{f}(a_i, \xi \mid \top[z_i(\xi_l)], \top[z_1(\xi_l)]) \times P(\top[z_i(\xi_l)], \top[z_1(\xi_l)]). \quad (38)$$

We interpret the probability $P(\top[z_i(\xi_l)], \top[z_1(\xi_l)])$ as a measure of the correctness of calculating $a_i(\xi_l)$ using (32). If the noise is at least one order of magnitude smaller than the observed signals, the approximation of (32) holds. If the noise parameters can be estimated, for example in the case of normal distributed noise with a known variance σ^2, the probability of the noise being an order of magnitude smaller than the measurement $\top[z_i(\xi_l)]$ or $\top[z_1(\xi_l)]$, is[4]

$$P(\top[z_i(\xi_l)], \top[z_1(\xi_l)]) \equiv \min\left\{ \int_{-|l_i|}^{|l_i|} \frac{1}{\sigma\sqrt{20\pi}} e^{-\frac{\upsilon^2}{20\sigma^2}} d\upsilon,\ \int_{-|l_1|}^{|l_1|} \frac{1}{\sigma\sqrt{20\pi}} e^{-\frac{\upsilon^2}{20\sigma^2}} d\upsilon \right\} \quad (39)$$

where $l_i = \top[z_i(\xi_l)]$.

We want to evaluate the conditional probability of $a_i(\xi)$, being represented by α_{i_k} parameters, given the ratio of the measurements $r_i(\xi_0), \ldots, r_i(\xi_{N_m-1})$, i.e., (36).

It should be pointed out with reference to the definition of the density $\hat{f}(a_i, \xi)$, that all the points belonging to the group of instances $\{\xi_l\}$, in places where $r_i(\xi_l) \approx a_i(\xi_l)$, are samples of a single curve. This is due to the fact that $a_i(\alpha_{i_k}; \xi_l)$ is a function of ξ and for each value of ξ, a_i is single valued.

Therefore, we can calculate the probability of (36) by integrating along the above curve. This can be done by using a line integral over a scalar field, being the density $\hat{f}(a_i, \xi)$

$$\mathbb{J}(\alpha_{i_k}) \equiv P\left(\alpha_{i_k} \mid r_i(\xi_0), \ldots, r_i(\xi_{N_m-1})\right) = \int_{\xi_1}^{\xi_2} \hat{f}(a_i(\alpha_{i_k}; \xi), \xi)\sqrt{1 + [a_i'(\alpha_{i_k}; \xi)]^2} d\xi \quad (40)$$

where a_i' stands for the derivation with respect to ξ, and ξ_1, ξ_2 are the observation start and stop time/position instance, respectively.

The optimization can be executed by means of the Newton method as follows.

1) Take as an initial guess a vector of $\alpha_{i_k}^{(0)}$ parameters.
2) Use the vector $\alpha_{i_k}^{(m)}$ obtained from the previous step, and construct the gradient vector $\nabla\mathbb{J}(\alpha_{ij_k}^{(m)})$ and the Hessian matrix $\mathrm{Hess}_{\mathbb{J}}(\alpha_{i_k}^{(m)})$, using

$$\frac{\partial \mathbb{J}(\alpha_{i_k}^{(m)})}{\partial \alpha_{i_k}^{(m)}} = \frac{\partial P\left(\alpha_{i_k}^{(m)} \mid r_i(\xi_0), \ldots, r_i(\xi_{N_m-1})\right)}{\partial \alpha_{i_k}^{(m)}}$$
$$\frac{\partial^2 \mathbb{J}(\alpha_{i_k}^{(m)})}{\partial \alpha_{i_k}^{(m)} \partial \alpha_{i_j}^{(m)}} = \frac{\partial^2 P\left(\alpha_{i_k}^{(m)} \mid r_i(\xi_0), \ldots, r_i(\xi_{N_m-1})\right)}{\partial \alpha_{i_k}^{(m)} \partial \alpha_{i_p}^{(m)}}. \quad (41)$$

3) Update the estimated parameters

$$\alpha_{i_k}^{(m+1)} = \alpha_{i_k}^{(m)} - \mathrm{Hess}_{\mathbb{J}}^{-1}(\alpha_{i_k}^{(m)})\nabla\mathbb{J}(\alpha_{i_k}^{(m)}). \quad (42)$$

4) Repeat steps 2 and 3 until convergence.

The initial points can be selected by using an approach similar to the application of the Hough transform in image processing [21]. A few remarks with reference to the implementation are in order.

1) The indefinite integral of (40) should be solved. In order to make this easier, we propose using the Epanechnikov kernel [22] as the kernel K in (37) which is used in evaluating $\hat{f}(a_i, \xi)$. The Epanechnikov kernel is defined as follows:

$$K(x, y) = \begin{cases} \frac{2}{\pi}(1 - x^2 - y^2), & (x^2 + y^2) \leq 1 \\ 0, & \text{otherwise.} \end{cases} \quad (43)$$

2) If $a_i(\alpha_{i_k})$ renders the indefinite integral of (40) to become unsolvable, an approximation for $a_i(\alpha_{i_k})$ should be used.
3) It is usually preferable to define $g_i'(\xi) \equiv \arctan(g_i(\xi))$ in order to eliminate the noise amplification accompanying the calculation of $\top[z_i(\xi)]/\top[z_1(\xi)]$. The definition of $r_i(\xi_l)$ to be inserted in (37) should be accordingly changed to $r_i(\xi_l) \equiv \arctan(\top[z_i(\xi_l)]/\top[z_1(\xi_l)])$. Subsequent to optimizing for the α_{i_k} parameters and finding $g_i'(\xi)$, $g_i(\xi)$ can be found using $g_i(\xi) = \tan(g_i'(\xi))$.

C. Source Estimation

As we have already shown, the Zadeh's frequency response of the elements of the mixing matrix is independent of ω and, therefore, these elements belong to their centralizer. This means that the elements of the matrix are commutative, permitting the application of the matrix algebra mentioned in Corollary 9. As long as the determinant $a_1(\xi) - a_2(\xi) \neq 0$,[5] the system is invertible and (17) can be used to estimate the sources.

D. System for Generating Position Varying Instantaneous Image Mixtures

In the process of imaging through a semi-reflector, such as a plain glass window, the reflected image is superimposed on the transmitted one. The left optical setup shown in Fig. 4 depicts an optical system assembled in our lab for generating position varying instantaneous mixture of images. The glass is used as a mixer of transmitted and reflected obtained images. The position varying mixing is achieved by changing the lighting condition of the transmitted image. First, mixture is acquired using room lighting whereas a second mixture is acquired using an extra nonuniform illumination.

[4]In the case where the noise parameters are unknown, we assume a uniform distribution for $(\top[z_i(\xi_l)], \top[z_1(\xi_l)])$.

[5]In instances, where $a_1(\xi) - a_2(\xi) \ll 1$ the system is singular or badly conditioned and the variational method should be used.

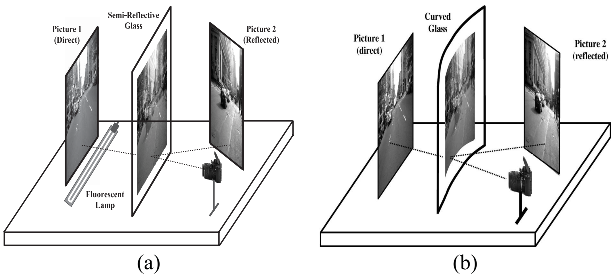


Fig. 4. Optical system generating (a) spatial varying instantaneous mixtures and (b) spatial varying single-path mixtures. Two pictures are positioned opposite to each other while a semi-reflective glass is mounted along the optical axis of one of them. In the case of instantaneous mixtures, the first mixture is generated using a uniform lighting. A second mixture is generated using a nonuniform illumination. In the case of single-path mixtures, the first and second mixtures are acquired from slightly different camera positions.

E. Results

We tested our approach on simulated and real mixtures. For the simulated mixtures, two mixtures were obtained from two audio signals. The signals were sampled at a rate of 32K samples/sec, and mixed using the following time varying instantaneous filters:

$$H = \begin{bmatrix} 1 & 1 \\ 1.1t + 0.6 & -1.7t + 2 \end{bmatrix}.$$

The filters change the attenuation of the signals linearly as a function of time. Random noise with normal distribution was added to the mixtures, yielding a peak-signal (mixture)-to-noise ratio (PSNR) of 20 dB. The mixtures were sparsified using STFT with window duration of $1/4$ of a second. The sparsity, measured by the Gini Index (which was found to be the best measure for sparsity, by [23]), increased from 0.28 to 0.51. For the parametric mixing system estimation, we estimated the probability density using the Epanechnikov radially symmetric kernel. The indefinite integral of (40) was solved and the intersections of the support of the kernel with the curve were found using MATLAB solver. The optimization scheme was implemented using MATLAB, and the algorithm was initialized by several starting points and converged to two local maxima. The system was directly inverted by calculating the inverse of the mixing matrix, H^{-1}. Fig. 5 depicts the signals of mixtures of two voices. On the top left are the spectrograms obtained by using a STFT with a window duration of $1/4$ second. Shown on the top right are the separated estimated sources. On the top middle, we can see the kernel estimated probability density. The audio mixtures and separated sources can be found at http://visl.technion.ac.il/∼kaftory/TIP. The CSISR of the top examples of Fig. 5 increased from 4 dB and −4 dB, for the two mixed sources, to 28 dB and 32 dB after the separation.

The setup depicted in Fig. 4 was used for generating real mixtures. The mixtures were acquired with a Nikon D100 digital camera controlled by a computer. The PSNR for the noisiest mixture, which was estimated by taking consecutive identical pictures, was 40 dB. We define $g_i'(\xi) \equiv \arctan(g_i(\xi))$ and assume that a second order polynomial is the estimated Taylor expansion of $g_i'(\xi)$. Since images are not naturally sparse, we used the Haar wavelet packet for sparsification. The sparsest node (node 1, 3) was chosen. This increased the Gini Index from 0.44 to 0.57. Again, for the parametric mixing system estimation, we estimated the probability density using the Epanechnikov radially symmetric kernel density estimation. We initiated our algorithm with several starting parameters and the algorithm converged to two local maxima. The system was inverted directly by calculating the inverse of the mixing matrix, H^{-1}.

Fig. 5(b) depicts results obtained in separation of reflection from a transmitted image. The mixtures were generated by the system shown in Fig. 4(a). The mixtures of the images are shown in Fig. 5(b.i). One mixture was acquired under normal room-lighting condition, and the other acquired with the addition of nonuniform lighting to the transmitted image. The estimated separated sources are shown in Fig. 5(b.iii). The mutual information of the mixtures is 0.3 and 0.15 for the separated sources. The kernel-estimated probability density, obtained by selecting pixels corresponding to wavelet packet coefficients exceeding the threshold, is shown on the middle.

VI. Blind Separation of Time/Position Varying Single-Path Mixtures[6]

Here, we assume that the filters of the mixing matrix have the form of (6), where there is no attenuation ($a_{ij}(\underline{\xi}) = 1$). In this section, we use as an example, two mixtures of a transmitted image and a phantom reflected from a curved semi-reflector, acquired from slightly different locations. When the semi-reflector is not flat, as is the case in a front windshield of a car or the canopy of a cockpit, the reflected image is usually distorted. Furthermore, the distortion of the reflection varies as a function of the viewing position [26]. We model these image mixtures in the context of a position varying single-path BSS and assume the following.

1) The model of the single-path mixing coefficient, $d_{ij}(\underline{\xi})$ in (6) is arbitrary but smooth.
2) The dependency of the reflection and the transmission coefficients on the viewing angle is negligible. Therefore, the image intensity does not change when the camera position varies.
3) The image sources contain edges or other features.

Considering two different mixtures and representing them in matrix form, the BSS problem can be written explicitly as

$$\begin{bmatrix} z_{\lambda_1} \\ z_{\lambda_2} \end{bmatrix} = \begin{bmatrix} h_{R_1} & h_{D_1} \\ h_{R_2} & h_{D_2} \end{bmatrix} \star \begin{bmatrix} R_\lambda \\ D_\lambda \end{bmatrix} + \eta \qquad (44)$$

where z_{λ_1} and z_{λ_2} are the mixtures observed from two different camera positions. h_R and h_D, defined by

$$h_R(\underline{\xi}, \underline{\xi}') \equiv \delta\left(\underline{\xi}' - M_R(\underline{\xi})\right) \quad h_D(\underline{\xi}, \underline{\xi}') \equiv \delta\left(\underline{\xi}' - M_D(\underline{\xi})\right) \qquad (45)$$

where M_R and M_D are mappings between the reflected image coordinate system and the transmitted image coordinate system to the camera system, respectively [24]. We explicitly define the filters g_i of (12) as

$$h_{R_2} = g_R \star h_{R_1}, \qquad h_{D_2} = g_D \star h_{D_1}. \qquad (46)$$

[6]Some of the methods and the results of this section were published in [24] and [25].

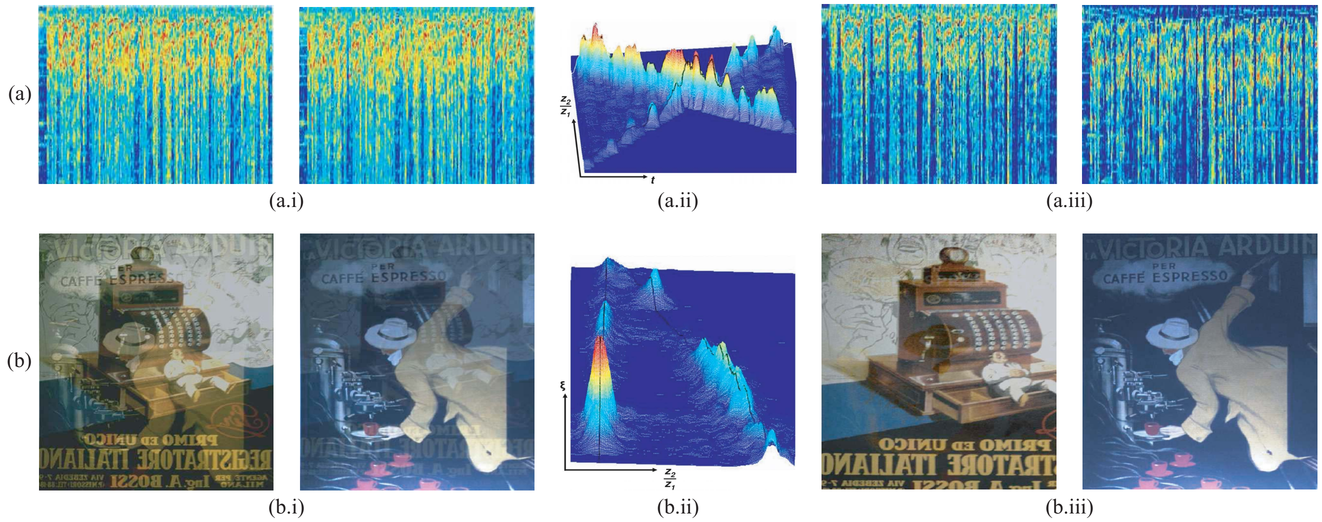


Fig. 5. (a) Two audio signals of a man's and a woman's voices are mixed. (a.i) Spectrograms of the two audio mixtures. (a.ii) Kernel-estimated probability density. The black lines indicate the estimated curves $g_i(\alpha_{i_k};\xi)$. (a.iii) Spectrograms of the estimated separated sources. (b) Mixtures produced by the setup depicted on Fig. 4(a). (b.i) Mixture of the images obtained in room lighting and with the addition of nonuniform lighting to the transmitted image. (b.ii) Kernel-estimated probability density. The black lines indicate the estimated curves $g_i'(\alpha_{i_k};\xi)$. (b.iii) Estimated image sources.

Solving for g_R and g_D yields

$$\begin{aligned} g_R &= \delta\left(\underline{\xi}' - M_{R_2}(\underline{\xi}) + M_{R_1}(\underline{\xi})\right) \\ g_D &= \delta\left(\underline{\xi}' - M_{D_2}(\underline{\xi}) + M_{D_1}(\underline{\xi})\right). \end{aligned} \tag{47}$$

A. Sparsification Using Scale Invariant Feature Transform (SIFT)

The Zadeh transform of one of the elements of the mixing matrix is

$$\begin{aligned} \mathcal{H}_{Z_R}(\underline{\xi},\underline{\omega}) &= \int_{-\infty}^{\infty} \delta\left(\underline{\xi}' - M_R(\underline{\xi})\right) e^{j\left(\underline{\omega}(\underline{\xi}-\underline{\xi}')^T\right)} d\underline{\xi} \\ &= e^{j\left(\underline{\omega}(\underline{\xi}-M_R(\underline{\xi}))\right)}. \end{aligned} \tag{48}$$

Since it is a function of both $\underline{\omega}$, and $\underline{\xi}$, finding its centralizer is a challenging task. None of the common sparsification transformations, which can be accounted for by conventional filters, belongs to the centralizer and is invariant to the mixing system. There exists, however, the SIFT [27], which cannot be accounted for by a filter, but is most suitable for our purpose of sparsification. It can be observed that the zero-order Taylor approximation of the filters of the mixing matrix, calculated around some point, is a translation of that point. The first-order scales the image about this point, while the second and higher order, produce nonlinear stretching of this image point. Therefore, the filters of the mixing matrix can be approximated by local affine transforms. The SIFT is invariant to local affine transforms, and can literally commute with the mixing matrix. The SIFT transforms an image into a large collection of local feature vectors, based on the appearance of an object of special points of interest (keypoints). Each feature vector is translation, rotation, and scale invariant, as well as invariant to illumination changes and 3-D projection. The feature vectors are highly distinctive and can, therefore, be regarded as a sparse representation, where only a small percentage of the pixels are chosen. Since the special points of interest (keypoints) are translation, rotation, and scale invariant, they can be identified and matched between two different images of the same scene.[7] Such matching points of interest are referred to as "matching keypoints."

B. Mixing Matrix Estimation

Consider for example a pair of images, one of which is a locally-distorted version of the other. The difference in location of the matching keypoints, $\underline{\xi}_0 - \underline{\xi}_1$, for the x and y axes, can be used in the process of estimating a surface, which corresponds to the mapping of one image to the other. Suppose one observes in one of the mixtures, at some position $\underline{\xi}_0$, the existence of a SIFT point of interest (keypoint), which is similar[8] to a SIFT corresponding point found in the other mixture in the position $\underline{\xi}_1$. Suppose that the keypoint belongs to the transmitted image. $M_{D_1}(\underline{\xi}) - M_{D_2}(\underline{\xi})$ is the difference (measured in pixels) between the mapping of a point in the transmitted image to the first and second observed images. Therefore, for the matching keypoints, this value is $M_{D_1}(\underline{\xi}) - M_{D_2}(\underline{\xi}) = \underline{\xi}_0 - \underline{\xi}_1$. Recalling (47), the value of $g_D(\underline{\xi})$ is known for this keypoint. For matching keypoints which belong to the reflected image, the value of g_R is known. Therefore, by scatter-plotting the location difference of the matching keypoints, two surfaces can be estimated. One corresponds to g_D and the other to g_R. We can usually assume that the transmitted image is not distorted but may be slightly shifted between the mixtures, and solve the problem of clustering the keypoints to g_R or g_D. It is done by first grouping all the matching keypoints of which their relative offsets are consistent with a small, close to zero shift, and assigning them to the surfaces of g_D. The remaining keypoints are assigned

[7] See [27] for a complete illustration of the method.

[8] Refer to [27] for the matching criteria.

to g_R and used in estimating the surface of g_R by utilizing a surface fitting technique. If we have a priori information about the curvature of the semi-reflector, we can assume a parametric surface fitting model for the surface estimation. Otherwise, we can assume that the surface is smooth and apply a nonparametric surface fitting with smoothing constraints.

C. Variational Source Estimation

From the Zadeh's transform of the filters of the mixing matrix in (48), one may conclude that the elements of the mixing matrix do not commute. Therefore, a direct inversion of the mixing system is not feasible. We implement a variational framework for solving the problem. Recalling the BSS problem

$$\begin{aligned} z_{\lambda_1} &= R'_\lambda + D'_\lambda + \eta_1 \\ z_{\lambda_2} &= g_R \star R'_\lambda + g_D \star D'_\lambda + \eta_2 \end{aligned} \tag{49}$$

we filter both sides of the first equation with the position varying filter g_R and subtract it from the second equation. This yields the following equation, to be solved for D'_λ:

$$z_{\lambda_2} - g_R \star z_{\lambda_1} = (g_D - g_R) \star D'_\lambda + \eta \tag{50}$$

where we used the distributivity property of $\star$.

A variational approach to solving (50) is by minimizing the cost functional

$$\mathbb{J} = \| z_{\lambda_2} - g_R \star z_{\lambda_1} - (g_D - g_R) \star D'_\lambda \|^2 + w\mathbf{R}(D'). \tag{51}$$

The regularization operator, $\mathbf{R}(D')$, that we use for color images, is based on the Beltrami flow. Sochen *et al.*, [28], interpret an image as a manifold (surface) embedded in a high dimensional space, where x and y are two spatial coordinates of this space, and the intensity at each of the λ channels is represented as an additional dimension. Hence, a color image, such as the reflected or the transmitted image, is a manifold embedded in a 5-D space. From this viewpoint, image regularization can be interpreted as a process that minimizes the surface area of this manifold. The surface area of an image is measured by the Polyakov action [29]. It is given for the transmitted image by

$$\mathbf{R}(D') = \iint_{\Omega} \sqrt{\det(\mathbf{G}_\mathrm{D})}\,dx\,dy \tag{52}$$

where $\mathbf{G}_\mathrm{D}$ is a 2×2 matrix. Each of the matrix entries, depends on the spatial location. Their values are given [30] by

$$\begin{aligned} \mathbf{g}^\mathrm{D}_{11} &= 1 + \rho^2 \sum_\lambda \left(\frac{\partial D'_\lambda}{\partial x}\right)^2 \\ \mathbf{g}^\mathrm{D}_{12} &= \mathbf{g}^\mathrm{D}_{21} = \rho^2 \sum_\lambda \frac{\partial D'_\lambda}{\partial x}\frac{\partial D'_\lambda}{\partial y} \\ \mathbf{g}^\mathrm{D}_{22} &= 1 + \rho^2 \sum_\lambda \left(\frac{\partial D'_\lambda}{\partial y}\right)^2 \end{aligned} \tag{53}$$

where ρ is a parameter for scaling the intensity dimensions. It is useful not to use the Polyakov action, but to use a modified version (by multiplying it by a positive function) where its derivative is the Beltrami flow which is defined as

$$\text{Beltrami} \equiv \frac{1}{\sqrt{\det(\mathbf{G}_\mathrm{D})}} \mathcal{D}iv \left[\sqrt{\det(\mathbf{G}_\mathrm{D})}(\mathbf{G}_\mathrm{D})^{-1}\nabla D'_\lambda\right] \tag{54}$$

where $\nabla D'_\lambda$ is the spatial gradient of D'_λ and $\mathcal{D}iv$ is the divergence operator.

The properties of using the Beltrami flow for the restoration of color images are investigated thoroughly in [31] and [32]. Some of these properties are:

1) forcing the color channels to spatially align and therefore to suppress color distortions;
2) reducing the noise;
3) preserving the image edges.

Minimizing $\mathbb{J}$, is accomplished by reaching $d\mathbb{J}/dD'_\lambda = 0$

$$\begin{aligned} &\sum\nolimits_\lambda (g_D - g_R)^T \star \left(z_{\lambda_2} - g_R \star z_{\lambda_1} - (g_D - g_R) \star D'_\lambda\right) \\ &\quad + \frac{w}{\sqrt{\det(\mathbf{G}_\mathrm{D})}} \mathcal{D}iv \left[\sqrt{\det(\mathbf{G}_\mathrm{D})}(\mathbf{G}_\mathrm{D})^{-1}\nabla D'_\lambda\right] = 0 \end{aligned} \tag{55}$$

where $(g_D - g_R)^T$ is the position varying filter obtained by switching the place along the coordinates $\underline{\xi}$ and $\underline{\xi}'$. Equation (55) is a nonlinear partial differential equations (PDE). Rather than solving it directly, we use the fixed point lagged diffusive method, used in [33]. By means of this method, the PDE can be solved by lagging the nonlinear term of the Beltrami operator one iteration behind. Then, for example, by knowing ${D'}^{(t)}_\lambda$ in some iteration t, ${D'}^{(t+1)}_\lambda$ can be found by solving a linear PDE, while applying a conjugate gradient minimization scheme.

In operator form, we obtain the following operator definitions:

$$\begin{aligned} &\mathcal{L}\left({D'}^{(t)}_\lambda\right){D'}^{(t+1)}_\lambda \equiv (g_D - g_R)^T \star (g_D - g_R) \star {D'}^{(t+1)}_\lambda \\ &\quad - \frac{\gamma}{\sqrt{\det(\mathbf{G}_{\mathrm{D}^{(t)}})}} \mathcal{D}iv \left[\sqrt{\det(\mathbf{G}_{\mathrm{D}^{(t)}})}(\mathbf{G}_{\mathrm{D}^{(t)}})^{-1}\nabla {D'}^{(t+1)}_\lambda\right] \end{aligned} \tag{56}$$

and

$$\mathcal{K} \equiv (g_D - g_R)^T \star \left(z_{\lambda_2} - g_R \star z_{\lambda_1}\right). \tag{57}$$

Equation (55) can be written using the above operator definition as

$$\mathcal{L}({D'}^{(t)}_\lambda){D'}^{(t+1)}_\lambda - \mathcal{K} = 0. \tag{58}$$

A direct solution of these equations, as required by each minimization step, would be

$${D'}^{(t+1)}_\lambda = \mathcal{L}({D'}^{(t)}_\lambda)^{-1}\mathcal{K}. \tag{59}$$

Since it is difficult to find and invert $\mathcal{L}({D'}^{(t)}_\lambda)^{-1}$, the step $d{D'}^{(t)}_\lambda$ is introduced

$${D'}^{(t+1)}_\lambda = {D'}^{(t)}_\lambda + d{D'}^{(t)}_\lambda. \tag{60}$$

The step $d{D'}^{(t)}_\lambda$ can be found by solving

$$\mathcal{L}({D'}^{(t)}_\lambda)d{D'}^{(t)}_\lambda = \mathcal{K} - \mathcal{L}({D'}^{(t)}_\lambda){D'}^{(t)}_\lambda. \tag{61}$$

Equation (61) can be solved by using the conjugate gradient method.

A similar restoration process with the Beltrami operator, was used in [34] for the unmixing and restoration of color images taken through a scattering medium.

D. System for Generating Position Varying Single-Path Mixtures

Optical systems that yield position varying image mixtures are depicted in Fig. 4. Shown on the right-hand side is a curved semi-reflector used as a mixer between a transmitted image and a reflection. The position-varying single-path mixing is imposed by the curvature of the semi-reflector. Two different mixtures are acquired by moving the camera to slightly different positions, which results in approximately the same transmitted image, but different distortions of the reflected image.

E. Results

We tested our approach on simulated and real mixtures. For the simulation, we implemented a simulator based on ray-tracing and used it to simulate a superposition of a reflected and transmitted image. Two observation points were used to obtain two slightly different mixtures. The shape of the semi-reflector was chosen to be a sphere, thus, producing a reflection with a fisheye-lens effect. We also assumed that the semi-reflector is very thin, to avoid distortions of the transmitted image. Finally, Gaussian noise was added to the mixtures, yielding a PSNR of 25 dB. For the real mixtures, the setup depicted in Fig. 4 was used. Instead of a semi-reflector made out of glass, we curved a thin polycarbonate sheet. The mixtures were acquired with a Nikon D100 digital camera interfaced with a computer. The PSNR for the noisiest mixture, which was estimated by taking consecutive identical pictures, was 40 dB.

The mixtures were sparsified using a SIFT software.[9] Keypoint matching was performed using the same software with a threshold on the vector angles from the nearest to second nearest neighbor of 0.6 (see the software for the threshold details). We use the parametric surface fitting on the simulated mixtures and a nonparametric on the experiments, by first calculating the location difference of corresponding SIFT keypoints. We take advantage of the fact that the transmitted image is not distorted. Thus, the location difference between the corresponding SIFT keypoints is constant. We estimate this constant by finding the maximum of the histogram of the location differences. After removing all the features corresponding to this constant, a second degree polynomial surface fitting is applied.

Fig. 6 depicts the results. The mixtures of the images are shown on the left. The estimated surfaces obtained by the parametric surface fitting are shown in the middle. The separated, estimated, sources are shown on the right. The estimated sources are well separated.

The mutual information of the images, which was 2.7 and 1.3 for the mixtures in the simulation and the experiment, respectively, was reduced to 0.8 and 0.6, respectively, after the separation.

[9]SIFT MATLAB code can be downloaded from www.cs.ubc.ca/~lowe/keypoints.

VII. Blind Separation of Time Varying Multipath Mixtures

We limit ourselves to filters which are time-varying, attenuated-and-delayed, versions of constant channel distortions

$$h_{ij}(t,\tau) = a_{ij}(t) f_{ij}(d_{ij}(t) - \tau) \tag{62}$$

where a_{ij} is a time-varying attenuation, f_{ij} is a finite impulse response filter and d_{ij} is a time-varying delay function. In this section, we assume the following.

1) Apart from some instances, the zero-order Taylor approximation of the Zadeh transform of the multipath mixing coefficients, around some instances confined to a small window, is sufficient to represent the coefficients in that window.
2) The sources are either sparse in their native domain, or can be sparsely represented in the time-frequency domain.

We assume that these filters have finite support, with a maximum support of size $2L_M$. We can, therefore, find a window of size $2L_T$, $L_T \gg L_M$, around the time instance t, in which $h_{ij}(t, \tau < t - L_T) = 0$ and $h_{ij}(t, \tau > t + L_T) = 0$. Consequently, we can state that

$$h_{ij}(t,\tau) = h_{ij}(t,\tau)\mathrm{wind}(t-\tau) \tag{63}$$

where $\mathrm{wind}(t-\tau)$ denotes a unit window function centered at $\tau = t$ with support of $2L_T$.

The BSS problem can be written as

$$\begin{aligned} z_i(t) &= \sum_j \int_{-\infty}^{\infty} h_{ij}(t,\tau)\mathrm{wind}(t-\tau)s_j(\tau)d\tau + \eta \\ &= \sum_j \int_{-\infty}^{\infty} h_{ij}(t,\tau)s_j(t,\tau)d\tau + \eta \end{aligned} \tag{64}$$

where we define $s(t,\tau) \equiv \mathrm{wind}(t-\tau)s(\tau)$. We note that $s(t)$ can be calculated by $s(t) = \int_{-\infty}^{\infty} \delta(t-\tau)s(t,\tau)d\tau$.

Using Zadeh's frequency response, (64) can be written as

$$z_i(t) = \frac{1}{2\pi} \sum_j \int_{-\infty}^{\infty} \mathcal{H}_{Z_{ij}}(t,\omega)\mathcal{S}_{F_j}(t,\omega)e^{j\omega t}d\omega \tag{65}$$

where $\mathcal{H}_{Z_{ij}}(t,\omega)$ is the Zadeh frequency response of h_{ij}, and $\mathcal{S}_{F_j}(t,\omega)$ is the Fourier transform of $s_j(t,\tau)$ along the τ axis.

The BSS problem can be written in the Zadeh time-frequency domain as

$$\mathcal{Z}_{Z_i}(t,\omega) = \sum_j \mathcal{H}_{Z_{ij}}(t,\omega)\mathcal{S}_{F_j}(t,\omega) + \tilde{\eta} \tag{66}$$

where $\tilde{\eta}$ is the time-frequency representation of the Gaussian noise, assumed to be constant. We can regard $\mathcal{Z}_{Z_i}(t,\omega)$ as the frequency representation of the output of the mixing system given an input at the instance t using the filters $h_{ij}(t,\tau)$, with the addition of some noise constant. Clearly, in a similar way

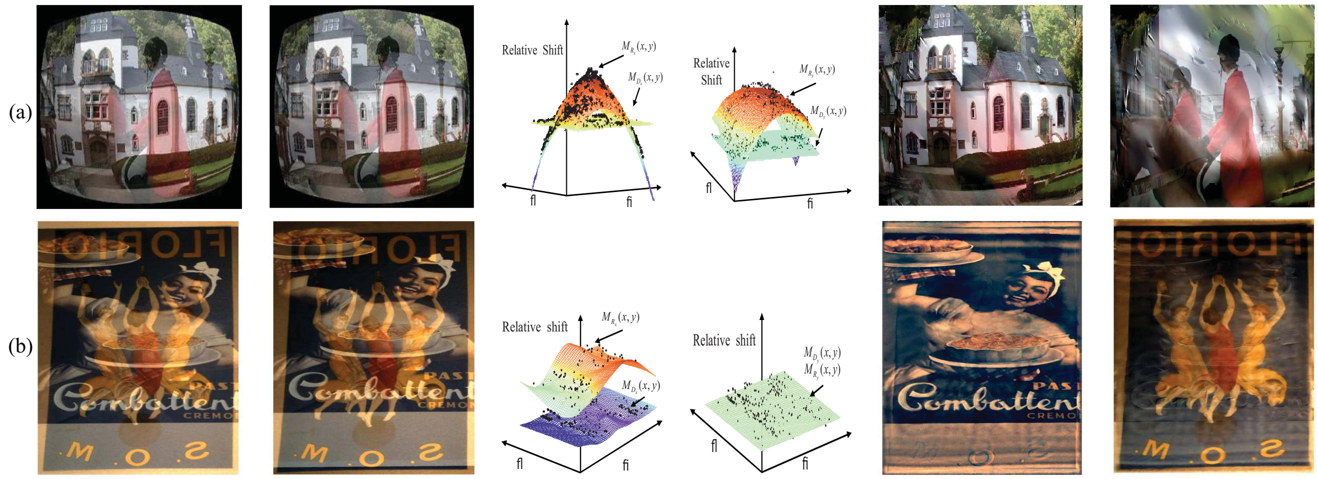


Fig. 6. (a) Simulated and (b) experimental single-path image mixtures obtained by a round semi-reflector. Left: Image mixtures. Middle: Estimated surfaces obtained by surface fitting of matching SIFT keypoints. Right: The estimated image sources.

to (65), $z_i(t)$ can be found using the inverse Zadeh transform of $\mathcal{Z}_{Z_i}(t,\omega)$

$$z_i(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathcal{Z}_{Z_i}(t,\omega)e^{j\omega t}d\omega. \tag{67}$$

The modified BSS problem can be written in the Zadeh domain as

$$\mathcal{Z}_{Z_i}(t,\omega) = \sum_j \mathcal{G}_{Z_{ij}}(t,\omega)\mathcal{S}'_{F_j}(t,\omega) + \tilde{\eta} \tag{68}$$

where $\mathcal{G}_{Z_{ij}}(t,\omega) = \mathcal{H}_{Z_{ij}}(t,\omega)\mathcal{H}^{-1}_{Z_{1j}}(t,\omega)$ and $\mathcal{S}'_{F_j}(t,\omega) = \mathcal{H}_{Z_{1j}}(t,\omega)\mathcal{S}_{F_j}(t,\omega)$, where a method for finding the sources $s'_1(t)$ and $s'_2(t)$ out of $\mathcal{S}'_{F_1}(t,\omega)$ and $\mathcal{S}'_{F_2}(t,\omega)$ is by calculating

$$s'(t) = \int_{-\infty}^{\infty} \mathcal{S}'_F(t,\omega)e^{j\omega t}d\omega.$$

A. Sparsification Using STFT

As stated earlier, $\mathcal{Z}_{Z_i}(t,\omega)$ is the time-frequency representation of the mixtures in the Zadeh transform domain. In a similar way to the STFT, and as one of the assumptions mentioned in the Introduction, $\mathcal{Z}_{Z_i}(t,\omega)$ is a sparse representation of audio and other harmonic signals.

Utilizing the assumption stated in the Introduction, we can find some windows of size $2L_L$, over which the zero-order Taylor approximation is sufficient to represent the Zadeh transform of the filters in these windows. We set the size of L_T such that $L_L > L_T + L_M$, where $2L_T$ is the size of the window used for calculating $s_j(t,\tau)$, and $2L_M$ is the maximum support of the filters of the mixing matrix.

It is shown in [11] that if the frequency content of the sources does not change over the same window where the zero-order Taylor approximation is sufficient to represent the Zadeh transform of the filters of the mixing matrix, the STFT with windows of size $2L_L$ of the observed mixtures $z_i(t)$, provides an estimation to $\mathcal{Z}_Z(t,\omega)$.

Since this estimation is sparse, the STFT under the assumptions can be regarded as a sparsification transformation.

B. Mixing Matrix Estimation

Sparse signals represented in the time-frequency space enable the estimation of the mixing system, since there are many frequencies and time instances, over which only one source is active. Therefore, $\mathcal{G}_{ij}(t,\omega)$ can be found from

$$\mathcal{G}_{Z_{ij}}(t,\omega) \approx \mathcal{Z}_{Z_i}(t,\omega)\mathcal{Z}_{Z_1}(t,\omega)^{-1}. \tag{69}$$

The Zadeh frequency responses of filters of the type used in this section, are

$$\mathcal{H}_{Z_{ij}}(t,\omega) = a_{ij}(t)\mathcal{F}_{F_{ij}}(\omega)e^{-j\omega(t-d_{ij}(t))} \tag{70}$$

where $\mathcal{F}_{F_{ij}}(\omega)$ is the Fourier transform of the fixed filter $f_{ij}(t-\tau)$.

The value of the filter $\mathcal{G}_{ij}(t,\omega)$ is therefore

$$\mathcal{G}_{Z_{ij}}(t,\omega) = \frac{a_{ij}(t)F_{F_{ij}}(\omega)e^{-j\omega(t-d_{ij}(t))}}{a_{1j}(t)F_{F_{1j}}(\omega)e^{-j\omega(t-d_{1j}(t))}}. \tag{71}$$

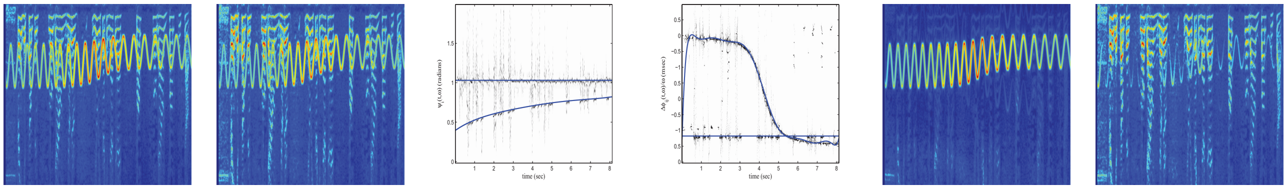
The filter $\mathcal{G}_{Z_{ij}}(t,\omega)$ is a complex function. Therefore, instead of finding it directly, we estimate its amplitude and phase separately. We define the angle $\psi_i(t,\omega)$ as follows:

$$\psi_i(t,\omega) = \tan^{-1}\left(\frac{\left|\mathcal{Z}_{Z_i}(t,\omega)\right|}{\left|\mathcal{Z}_{Z_1}(t,\omega)\right|}\right). \tag{72}$$

According to the assumptions related to properties of sources that are sparse in their time-frequency domain representation, there are time instances t_n and frequencies ω_m where only one source is active. We scatter-plot the points $\psi_1(t_n,\omega_m)$ and $\psi_2(t_n,\omega_m)$ for some fixed ω as a function of t. The points lie on one of the curves $\tilde{\psi}_{1j}(t,\omega)$ or $\tilde{\psi}_{2j}(t,\omega)$, where $\tilde{\psi}_{ij}(t,\omega)$ is defined as follows:

$$\tilde{\psi}_{ij}(t,\omega) = \tan^{-1}\left(\frac{a_{ij}(t)\left|\mathcal{F}_{F_{ij}}(\omega)\right|}{a_{1j}(t)\left|\mathcal{F}_{F_{1j}}(\omega)\right|}\right). \tag{73}$$

Note that the above curve is continuous, since $\tilde{\psi}_{ij}(t,\omega)$ does not depend on the value of the source j. Therefore, if we assume that $a_{ij}(t)$ is a continuous functions of t, $\tilde{\psi}_{ij}(t,\omega)$ is also a continuous function of t, bounded between 0 and $\pi/2$. We should note that we use the inverse tangent of the



(a) (b) (c)

Fig. 7. Audio signal of a man's voice is mixed with the sound of a siren generated according to a semi-realistic scenario. (a) Spectrograms of the two audio mixtures. (b) Amplitude and phase of the filters of the mixing matrix. (c) Spectrograms of the estimated separated sources.

ratio $a_{ij}(t)\left|F_{F_{ij}}(\omega)\right|/a_{1j}(t)\left|F_{F_{1j}}(\omega)\right|$ in order to suppress the amplification of the noise in the cases where the denominator is close to zero. We define the phase shift $\Delta\phi_i(t,\omega)$ as follows:

$$\Delta\phi_i(t,\omega) = \angle\mathcal{Z}_{Z_i}(t,\omega) - \angle\mathcal{Z}_{Z_1}(t,\omega) \tag{74}$$

where the symbol $\angle$ denotes the phase. We scatter-plot the points $\Delta\phi_1(t_n,\omega_m)$ and $\Delta\phi_2(t_n,\omega_m)$ on a graph for some fixed ω as a function of t. The points lie on one of the curves $\Delta\tilde{\phi}_{1j}(t,\omega)$ and $\Delta\tilde{\phi}_{2j}(t,\omega)$, where $\Delta\tilde{\phi}_{ij}(t,\omega)$ is defined as follows:

$$\Delta\tilde{\phi}_{ij}(t,\omega) = \angle\mathcal{F}_{F_{ij}}(\omega) - \angle\mathcal{F}_{F_{1j}}(\omega) + \omega(d_{ij}(t) - d_{1j}(t)). \tag{75}$$

In this case too, the above curve is continuous, since $\Delta\tilde{\phi}_{ij}(t,\omega)$ does not depend on the value of the source j. Therefore, if we assume that $d_{ij}(t)$ is a continuous function of t, $\Delta\tilde{\phi}_{ij}(t,\omega)$ is also a continuous function of t, bounded between 0 and 2π with a 2π fold.

It is observed that the values of the filters $\mathcal{G}_{Z_{ij}}$ can be found by

$$\mathcal{G}_{Z_{ij}}(t,\omega) = \tan\left(\tilde{\psi}_{ij}(t,\omega)\right)e^{j\Delta\tilde{\phi}_{ij}(t,\omega)}. \tag{76}$$

The objective is, therefore, to find the curves out of the cluttered points. This can be done either by using a parametric or nonparametric estimation using methods described in the previous sections.

C. Source Estimation

In the time-frequency domain, the filters $\mathcal{G}_{Z_{ij}}(t,\omega)$ of the above form, are commutative. Therefore, a direct inversion of the problem is feasible. In cases where $\mathcal{G}_{Z_{22}}(t,\omega) - \mathcal{G}_{Z_{21}}(t,\omega) \ll 1$, the system is singular or badly conditioned. A variational methods should then be used. Two variational methods–one in the combined time-frequency domain and the other in the time domain are presented in [11].

D. Results

We tested our approach on simulated audio signals in a semi-realistic acoustic scenario, in which a reporter is speaking to a two-microphone array, while an emergency vehicle is passing nearby. The functions $a_{ij}(t)$ and $d_{ij}(t)$ are calculated for this scenario as

$$a_{ij}(t) = \frac{1}{q_{ij}(t)}, \qquad d_{ij}(t) = t - \frac{q_{ij}(t)}{c} \tag{77}$$

where c is the speed of sound and q_{ij} is the distance between the jth source to the ith sensor. The signals recorded from the reporter and the emergency vehicle were sampled with anti-aliasing filter at the rate of $4K$ samples/second. We assumed that the microphones were close to each other; therefore, the time invariant atmospheric transfer function obeys $f_{11} = f_{21}$ and $f_{12} = f_{22}$. Finally, Gaussian noise was added to the mixtures, yielding a signal (mixture)-to-noise ratio of 20 dB. The value of $Z_i(t,\omega)$ was obtained using a windowed Fourier transform with a window of length 400 samples. This increased the Gini Index from 0.43 to 0.67.

We scatter-plot the points $\psi(t_n,\omega_m)$ for the time and frequencies, where the absolute values of the windowed Fourier transform of the mixtures are above the threshold. It is done for each ω as a function of t. Since according to the assumptions, this ratio should be the same for all ω, we take the maximum of the histogram and plot it as a function of time. We use the parametric curve fitting assuming a polynomial of various degrees, until a good approximation to the curves $\tilde{\psi}_{ij}(t,\omega)$ is achieved. We then scatter-plot the points $\Delta\phi(t_n,\omega_m)/\omega_m$ for the time and frequencies, where the absolute values of the windowed Fourier transform of the mixtures are above the threshold. These points are independent of ω. We use the parametric curve fitting, assuming a polynomial of various degrees, until a good approximation to the curves $\Delta\tilde{\psi}_{ij}(t,\omega)/\omega$ is achieved.

Fig. 7 depicts the results. The estimation of $\tilde{\psi}_{ij}(t,\omega)$ and $\Delta\tilde{\phi}_{ij}(t,\omega)/\omega$ using a polynomial fit are shown in Fig. 7(b). We use $\tilde{\psi}_{ij}(t,\omega)$ and $\Delta\tilde{\phi}_{ij}(t,\omega)$ to construct $\mathcal{G}_{Z_{ij}}(t,\omega)$ by

$$\mathcal{G}_{Z_{ij}}(t,\omega) = \tan\left(\tilde{\psi}_{ij}(t,\omega)\right)e^{j\Delta\tilde{\phi}_{ij}(t,\omega)}. \tag{78}$$

The system is inverted using the variational approach in the time-frequency domain. The spectrogram of the mixtures is depicted on Fig. 7(a), whereas the estimation of $\tilde{\psi}_{ij}(t,\omega)$ and $\Delta\tilde{\psi}_{ij}(t,\omega)/\omega$ are depicted in the middle and the estimated sources are depicted in Fig. 7(b). The mutual information of 1.6 of the mixtures, was reduced significantly to 0.7 for the separated sources.

VIII. Conclusion

We proved that in the general case, the filters of the mixing matrix do not commute and therefore inversion of the mixing system using matrix algebra is not feasible. However, we found in the Zadeh time-frequency domain, the conditions under which filters are commutative and, consequently, an inverse filter exists.

Instead of using an online form of instantaneous or convolutive BSS, we use the SSCA approach in a batch processing manner. We studied the conditions which allow the application of the SSCA, and showed that using an appropriate sparsification operator enables the estimation of the mixing system, as long as a proper threshold was imposed, so that the above-threshold instances are frequent enough and have a good signal-to-noise ratio. Estimation of the mixing system enables the system inversion, which can be done directly by matrix inversion. For cases where matrix algebra cannot be used, or where the system is badly conditioned, we developed a variational approach for solving the inverse problem.

The demonstration of our SSCA approach along with the methods for sparsification and parametric and nonparametric mixing matrix estimation showed the feasibility of using the SSCA for the solution of various time/position varying BSS problems. This was demonstrated on simulated and experimental data, obtained from three types of mixtures representing many real-life applications.

Acknowledgment

The authors thank J. Delpiano for uploading a program for calculating the mutual information to MATLAB Central, Research supported by the Ollendorff Minerva Center.

References

[1] A. Hyvarinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Comput.*, vol. 9, no. 7, pp. 1483–1492, 1997.

[2] A. Hyvarinen, "A family of fixed-point algorithms for independent component analysis," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, Apr. 1997, pp. 3917–3920.

[3] M. Gaeta and J. L. Lacoume, "Source separating without prior knowledg: The maximum likelihood solution," in *Proc. EUSIPCO*, 1990, pp. 621–624.

[4] P. Comon, "Independent componenet analysis, a new concept?" *Signal Process.*, vol. 36, no. 3, pp. 287–314, 1994.

[5] A. Bell and T. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, no. 6, pp. 1129–1159, 1995.

[6] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proc. F, Radar Signal Process.*, vol. 140, no. 6, pp. 362–370, Dec. 1993.

[7] C. Jutten and J. Herault, "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture," *Signal Process.*, vol. 24, no. 1, pp. 1–10, 1991.

[8] A. Belouchrani, K. Abed-Meraim, J. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[9] M. Zibulevsky, B. A. Pearlmutter, P. Bofill, and P. Kisilev, "Blind source separation by sparse decomposition," in *Independent Component Analysis: Principles and Practice*. Cambridge, U.K.: Cambridge Univ. Press, 2001.

[10] P. O'Grady, B. Pearlmutter, and S. Rickard, "Survey of sparse and non-sparse methods in source separation," *Int. J. Imag. Syst. Technol.*, vol. 15, no. 1, pp. 18–33, 2005.

[11] R. Kaftory, "Blind separation of time/position varying mixtures," Ph.D. dissertation, Dept. Electr. Eng., Technion - Israel Inst. Technology, Technion, Israel, 2009.

[12] B. Olshausen and D. Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1?" *Vis. Res.*, vol. 37, no. 23, pp. 3311–3325, 1997.

[13] J. P. Berrut and L. N. Trefethen, "Barycentric lagrange interpolation," *SIAM Rev.*, vol. 46, no. 3, pp. 501–517, 2004.

[14] F. Marvasti, "Random topics in nonuniform sampling," in *Nonuniform Sampling: Theory and Practice*. Norwell, MA: Kluwer, 2001, pp. 170–173.

[15] L. Zadeh, "Frequency analysis of variable networks," *Proc. IRE*, vol. 38, no. 3, pp. 291–299, Mar. 1950.

[16] H. Gafni and Y. Y. Zeevi, "A model for separation of spatial and temporal information," *Biol. Cybern.*, vol. 28, no. 2, pp. 73–82, 1977.

[17] R. Kaftory and Y. Y. Zeevi, "Blind separation of time/position varying mixtures," Dept. Electr. Eng., Technion - Israel Inst. Technology, Technion, Israel, CCIT Rep. 758, 2010.

[18] H. D'Angelo, *Linear Time-Varying Systems: Analysis and Synthesis*. Boston, MA: Allyn & Bacon, 1970.

[19] R. Kaftory and Y. Y. Zeevi, "Probabilistic geometric approach to blind separation of time-varying mixtures," in *Independent Component Analysis and Signal Separation* (Lecture Notes in Computer Science), vol. 4666. New York: Springer-Verlag, 2007, pp. 373–380.

[20] P. Kisilev, M. Zibulevsky, and Y. Y. Zeevi, "A multiscale framework for blind source separation," *J. Mach. Learn. Res.*, vol. 4, pp. 1339–1363, Jan. 2003.

[21] D. Ballard, "Generalizing the hough transform to detect arbitrary shapes," *Pattern Recognit.*, vol. 13, no. 2, pp. 111–122, 1981.

[22] D. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*. New York: Wiley, 1992, p. 139.

[23] N. Hurley and S. Rickard, "Comparing measures of sparsity," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4723–4741, Oct. 2009.

[24] R. Kaftory and Y. Y. Zeevi, "Blind separation of images obtained by spatially-varying mixing system," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2008, pp. 2604–2607.

[25] R. Kaftory and Y. Y. Zeevi, "Blind separation of position varying mixed images," in *Proc. IEEE Int. Conf. Image Process.*, Nov. 2009, pp. 3913–3916.

[26] M. Oren and S. K. Nayar, "A theory of specular surface geometry," in *Proc. IEEE 5th Int. Conf. Comput. Vis.*, Jun. 1995, pp. 740–747.

[27] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep. 1999, pp. 1150–1157.

[28] N. Sochen, R. Kimmel, and R. Malladi, "A general framework for low level vision," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 310–318, Mar. 1998.

[29] A. M. Polyakov, "Quantum geometry of bosonic strings," *Phys. Lett.*, vol. 103, no. 3, pp. 207–210, 1981.

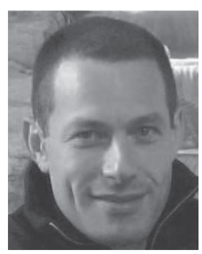
[30] R. Kimmel, R. Malladi, and N. Sochen, "Images as embedded maps and minimal surfaces: Movies, color, texture and volumetric medical images," *Int. J. Comput. Vis.*, vol. 39, no. 2, pp. 111–129, 2000.

[31] R. Kaftory, "Color image restoration using the Beltramy framework," M.S. thesis, Dept. Electr. Eng., Technion, Haifa, Israel, 2002.

[32] R. Kaftory, N. Sochen, and Y. Zeevi, "Variational blind deconvolution of multi-channel images," *Int. J. Imag. Syst. Technol.*, vol. 15, no. 1, pp. 56–63, 2005.

[33] C. Vogel and M. Oman, "Iterative methods for total variation denoising," *J. Sci. Comput.*, vol. 17, no. 1, pp. 227–238, 1996.

[34] R. Kaftory, Y. Y. Schechner, and Y. Y. Zeevi, "Variational distance-dependent image restoration," in *Proc. Conf. Comput. Vis. Pattern*, Jun. 2007, pp. 1–8.



Ran Kaftory received the B.Sc., M.Sc. and Ph.D. degrees in electrical engineering from the Technion-Israel Institute of Technology, Haifa, Israel, in 1996, 2001, and 2009, respectively.

He was the Signal and Image Processing Group Leader with the Electro Optics Research and Development Ltd., Haifa, from 2001 to 2005, where he was involved in research and development of military and homeland security systems. He is one of the founders of EyeCue Vision Technologies Ltd., Haifa, which is a company that develops computer vision technologies which are being integrated into toys and games, where he is currently the Chief Technical Officer of the company.

Yehoshua Y. Zeevi received the Ph.D. degree from the University of California, Berkeley.

He is the Director of the Ollendorff Center for Vision and Image Sciences, Technion-Israel Institute of Technology, Haifa, Israel. He was the Barbara and Norman Seiden Professor of computer sciences, and a member of the Technion Board of Governors and the Dean of the Faculty of Electrical Engineering. He was a Vinton Hayes Fellow at Harvard University, Cambridge, MA, a Visiting Professor with the Massachusetts Institute of Technology, Cambridge, Rutgers University, New Brunswick, NJ, and Columbia University, New York, NY, where he has been a regular visitor in the last ten years.

Dr. Zeevi is the Editor-in-Chief of the *Journal of Visual Communication and Image Representation*. He was one of the founders of iSight, UltraGuide, CByond, and of Cortica.