Signal and image source separations by wavelet analysis

Ryuichi Ashino\textsuperscript{1}, Shusuke Kataoka\textsuperscript{2}, Takeshi Mandai\textsuperscript{3} and Akira Morimoto\textsuperscript{2}

\textsuperscript{1} Mathematical Sciences, Osaka Kyoiku University, Japan
\textsuperscript{2} Information Science, Osaka Kyoiku University, Japan
\textsuperscript{3} Department of Engineering Science, Osaka Electro-Communication University, Japan

E-mail: ashino@cc.osaka-kyoiku.ac.jp, j109604@ex.osaka-kyoiku.ac.jp, mandai@isc.osakac.ac.jp and morimoto@cc.osaka-kyoiku.ac.jp

Abstract. Blind Source Separation (BSS) is one of the hottest emerging areas in signal processing. The purpose of BSS is to separate and to estimate the original sources from the sensor array, without knowing the transmission channel characteristics. Besides methods based on independent component analysis which is one of the most powerful tools for BSS, several methods based on time-frequency analysis have been proposed. One of them is the quotient signal estimation method which can estimate the unknown number of sources. We present a new method using wavelet analysis and apply it to signal and image source separations.

1. Introduction

The human visual and auditory systems are sensors to measure the frequency of light and sound, respectively. By this reason, we humans understand the nature using time and frequency. Let us consider a music. By recording the music, we have a function $f(x), x \in \mathbb{R}$. The music can be reproduced in full if we use all the values of $f(x)$. On the other hand, we can also determine completely all properties of $f(x)$ from its Fourier transform, because the Fourier transform is one-to-one and we have the inverse Fourier transform. Either one of $f(x)$ and $\hat{f}(\xi)$ is sufficient as information representing the music. When we play a music with an instrument, say piano, we hit a certain key in a certain time according to a musical score, which is called a time-frequency information. Thus the human recognizes the nature based on time-frequency information.

![Figure 1](image-url)

\textbf{Figure 1.} Time-frequency information of a music — a musical score.

The \textit{cocktail party effect} \cite{5, 8} is a challenging problem in auditory perception. One of basic questions for the cocktail party problem is how to build a machine to solve the cocktail party problem in a satisfactory manner. This question corresponds to \textit{Blind Source Separation} (BSS) \cite{8}, which is one of the hottest emerging areas in signal processing \cite[Chapter 1]{6}. The purpose of
BSS is to separate and to estimate the original sources from the sensor array, without knowing the transmission channel characteristics. Certain a priori knowledge on the original sources is needed for this separation, and the original sources cannot be uniquely estimated. Besides methods based on independent component analysis [9] which is one of the most powerful tools for BSS, several methods based on time-frequency analysis have been proposed. One of them is the quotient signal estimation method [4, 10, 14, 15], which was refined in [1, 3] by using the analytic wavelet transform based on the analytic signal.

For image source separation, we combined the edge analysis by the wavelet transform and the quotient signal estimation method in [13]. As the monogenic signal is a two-dimensional generalization of the analytic signal [7], we proposed to use the monogenic wavelet transform for image source separation in [12]. In this lecture, we improve this method by introducing a new algorithm.

2. Blind separation problem of images

For the sake of simplicity, we will deal with only gray scale images having \( P \times Q \) pixels, which can be regarded as matrices in \( \mathbb{R}^{P \times Q} \). We will use the following notation:

\[
\mathbf{a}_j = (a_1, \ldots, a_n)^T.
\]

The full-rank matrix \( \mathbf{A} = (a_j,\ell) \in \mathbb{R}^{M \times N} \) with non-negative components is called the mixing matrix. Assume that there are \( N \)-tuple of unknown original images:

\[
\mathbf{s}_\ell = (s_{\ell[p,q]}) \in \mathbb{R}^{P \times Q}, \quad \ell = 1, \ldots, N,
\]

where the number \( N \) is also unknown. The mathematical model of observed images are

\[
x_j = \sum_{\ell=1}^{N} a_{j,\ell} s_\ell, \quad j = 1, \ldots, M,
\]

which can be represented as

\[
x_j[p,q] = \mathbf{A} \mathbf{s}_j[p,q].
\]

Estimating the original images from the observed images is called blind separation. Once we know \( N \) and \( \mathbf{A} \), we can reconstruct original images \( s_j, \ell = 1, \ldots, N \) by solving the equation (1). Hence, in our case, the blind separation problem is reduced to the estimation of the unknown quantities \( N \) and \( \mathbf{A} \).

3. Riesz transform and monogenic signal

The \( n \)-dimensional Fourier transform of \( f(x) \in L^1(\mathbb{R}^n) \) is defined by

\[
\mathcal{F}[f](\xi) = \hat{f}(\xi) = \int_{\mathbb{R}^n} f(x) e^{-ix\cdot\xi} \, dx,
\]

where \( i = \sqrt{-1} \). The inverse Fourier transform of \( g(\xi) \in L^1(\mathbb{R}^n) \) is defined by

\[
\mathcal{F}^{-1}[g](x) = (2\pi)^{-n} \int_{\mathbb{R}^n} g(\xi) e^{ix\cdot\xi} \, d\xi.
\]

The Riesz transforms, denoted by \( \mathcal{R}_j, j = 1, \ldots, n \), are the Fourier multiplier operators \(-i D_j |D|\), that is,

\[
\mathcal{R}_j f(x) = \mathcal{F}^{-1} \left[ -i \frac{\xi_j}{|\xi|} \hat{f}(\xi) \right].
\]
The Riesz transforms $\mathcal{R}_j f$, $j = 1, \ldots, n$ of real-valued function $f$ are also real-valued. The masks $-\xi_1/|\xi|, -\xi_2/|\xi|$ of the two-dimensional Riesz transforms $\mathcal{R}_1$, $\mathcal{R}_2$ are pure imaginary. The imaginary parts $-\xi_1/|\xi|, -\xi_2/|\xi|$ of the masks are illustrated in Figure 2.

Let us define three fundamental unitary operators for time-frequency analysis. Denote by $T_b$, the translation operator by $b \in \mathbb{R}^n$, that is,

$$(T_b f)(x) = f(x - b),$$

by $D_a$, the dilation operator by $a > 0$, that is,

$$(D_a f)(x) = a^{-n/2} f(x/a),$$

and by $M_\omega$, the modulation operator by $\omega \in \mathbb{R}^n$, that is,

$$(M_\omega f)(x) = e^{i\omega \cdot x} f(x).$$

**Definition 1** The monogenic signal $\mathcal{M} f$ of $f \in L^2(\mathbb{R}^n)$ is defined by

$$\mathcal{M} f = (f, \mathcal{R}_1 f, \mathcal{R}_2 f, \ldots, \mathcal{R}_n f).$$

**4. Monogenic wavelet transform**

The **continuous wavelet transform** $W_\psi f(b,a)$ of $f \in L^2(\mathbb{R}^n)$ with respect to a window function $\psi$, called wavelet function, is defined by

$$W_\psi f(b,a) = \langle f, T_b D_a \psi \rangle = a^{-n/2} \int_{\mathbb{R}^n} f(x) \overline{\psi}(x - b/a) \, dx,$$

where $b \in \mathbb{R}^n$ and $a \in \mathbb{R}_+ = (0, \infty)$. The inversion formula, called the **inverse continuous wavelet transform**, $f(x) = \frac{1}{C_\psi} \int_{b \in \mathbb{R}^n} \int_{a \in \mathbb{R}_+} (W_\psi f)(b,a) T_b D_a \psi \frac{da}{a^{n+1}} \, db$ is obtained under the hypothesis that

$$\int_0^\infty \frac{|\hat{\psi}(\omega t)|^2}{t} \, dt = C_\psi < \infty, \quad \text{independent of } \omega,$$

which is called the admissibility condition. We have the following theorem:

**Theorem 2**

$$\mathcal{R}_j W_\psi = W_\psi \mathcal{R}_j = W_{-\mathcal{R}_j \psi}, \quad j = 1, \ldots, n.$$
The monogenic signal

\[ \mathcal{M}(W_{\psi}f) = ((W_{\psi}f), \mathcal{R}_1(W_{\psi}f), \ldots, \mathcal{R}_n(W_{\psi}f)) \]

of the continuous wavelet transform \( W_{\psi}f \) is called the monogenic wavelet transform of \( f \). By Theorem 2, we have the following Lemma 3.

**Lemma 3**

\[ \mathcal{M}(W_{\psi}f) = ((W_{\psi}f), (W_{-\mathcal{R}_1\psi}f), \ldots, (W_{-\mathcal{R}_n\psi}f)). \]

By Lemma 3, the monogenic wavelet transform of \( f \) can be regarded as the \((n + 1)\)-tuple of continuous wavelet transforms of \( f \) with \( n + 1 \) different wavelet functions \( \psi \) and \( -\mathcal{R}_j\psi \), \( j = 1, \ldots, n \). The vector valued function \((\psi, -\mathcal{R}_1\psi, \ldots, -\mathcal{R}_n\psi)\) is called the monogenic wavelet function generated from \( \psi \). Therefore, we can use the framework of the method for BSS using time-frequency information matrices made by several different wavelet functions proposed in [2].

5. Algorithms for image separation

To deal with images, we discretize the monogenic wavelet transform. The parameters \( b \) and \( a \) of the continuous wavelet transform \( W_{\psi}f(b,a) \) are called position and scale, respectively. Substitute \( b \) with \((p,q)\in\mathbb{N}^2\) and \( a \) with \( \alpha^r \), where \( \alpha > 1 \) is a properly chosen constant and \( r \in \mathbb{Z} \). Then, \((p,q)\) and \( r \) are discrete version of position and scale parameters instead of \( b \) and \( a \), respectively. For an image \( x = (x[p, q]) \in \mathbb{R}^{P \times Q} \), the spatial-scale information \( X \) of \( x \) is defined by

\[ X[p, q, r] = \mathcal{M}(W_{\psi}x)(p, q, \alpha^r) = ((W_{\psi}x)(p, q, \alpha^r), \mathcal{R}_1(W_{\psi}x)(p, q, \alpha^r), \mathcal{R}_2(W_{\psi}x)(p, q, \alpha^r)). \]

Denote by \( S_\ell \) and \( X_\ell \), the spatial-scale informations of the original images \( s_\ell \) and the observed images \( x_\ell \), respectively. By the linearity of the continuous wavelet transform and the Riesz transforms, (1) can be represented as

\[ X_\ell[p, q, r] = \sum_{\ell=1}^{N} a_{j, \ell} S_\ell[p, q, r], \quad j = 1, \ldots, M, \]

or

\[ X_\ell[p, q, r] = AS_\ell[p, q, r]. \]  

The following Algorithm 4 estimates the number \( N \) of sources and the mixing matrix \( A \).

**Algorithm 4**

(i) For each point \((p,q,r)\), compute the singular value decomposition (SVD) of \( X_\ell[p, q, r] = U \Sigma V^T \in \mathbb{R}^{M \times 3} \), where \( s_1, s_2, s_3 \) are singular values with \( s_1 \geq s_2 \geq s_3 \geq 0 \). When \( s_1 \gg s_2 \), record the first column \( u_1 \) of \( U \) if every element of \( u_1 \) is positive, record \( -u_1 \) else if every element of \( u_1 \) is negative.

(ii) Plot the histogram of recorded \( u_1 \) on the sphere \( S^{M-1} \) by the self-organizing map [11]. Note that \( ||u_1|| = 1 \), because \( u_1 \) is the first column of the orthogonal matrix \( U \).

(iii) Estimate the number \( N \) of sources by the number of peaks by plotting the histogram.

(iv) Estimate the normalized columns of the mixing matrix \( A \) by the spherical coordinate of the peaks.
6. Conclusion
The monogenic wavelet transform of an $n$-dimensional function $f$ can be regarded as the $(n + 1)$-tuple of continuous wavelet transforms of $f$ with $n + 1$ different wavelet functions $\psi$ and $-R_j \psi$, $j = 1, \ldots, n$. By generalizing the time-frequency information matrix method to the $n$-dimensional case, we have proposed a novel method for blind image separation using the monogenic wavelet transform.

Acknowledgment
This work was partially supported by JSPS.KAKENHI (C)20540168, (C)20540193, (C)22540130.

References
[1] R. Ashino, C. A. Berenstein, K. Fujita, A. Morimoto, M. Morimoto, D. Napoletani, and Y. Takei, Mathematical background for a method on quotient signal decomposition, Appl. Anal., 86 (5), 577–609, 2007.
[2] R. Ashino, K. Fujita, T. Mandai, A. Morimoto, and K. Nishihara, Blind source separation using time-frequency information matrix given by several wavelet transforms, Information, 10 (5), 555–568, 2007.
[3] R. Ashino, T. Mandai, A. Morimoto, and F. Sasaki, Blind source separation of spatio-temporal mixed signals using time-frequency analysis, Appl. Anal., 88 (3), 425–456, 2009.
[4] R. Balan and J. Rosca, Statistical properties of STFT ratios for two channel systems and applications to blind source separation, in Proceedings of the 2nd International Workshop on Independent Component Analysis and Blind Signal Separation, Helsinki, Finland, 19–22, 2000.
[5] C. Cherry, Some experiments on the recognition of speech, with one and with two ears, Journal of the Acoustical Society of America, 25 (5), 975–979, 1953.
[6] A. Cichocki and S. Amari, Adaptive blind signal and image processing: Learning algorithms and applications, John Wiley & Sons, Chichester, West Sussex, England, 2002.
[7] M. Felsberg and G. Sommer, The monogenic signal, IEEE Trans. Signal Processing, 49 (12), 3136–3144, 2001.
[8] S. Haykin and Z. Chen, The cocktail party problem, Neural Computation, 17 (9), 1875–1902, 2005.
[9] A. Hyvärinen, J. Karhunen, E. Oja, Independent component analysis, John Wiley & Sons, New York, NY, 2001.
[10] A. Jourjine, S. Rickard, and O. Yilmaz, Blind separation of disjoint orthogonal signals: Demixing $N$ sources from $2$ mixtures, in Proceedings of the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing, 5, 2985–2988, 2000.
[11] T. Kohonen, Self-Organizing Maps, Second Edition, Springer-Verlag, Berlin, 1997.
[12] A. Morimoto, R. Ashino, T. Mandai, Image separation using the monogenic wavelet transform, in Proceedings of the 10th International Symposium on Communications and Information Technologies 2010 (ISITC 2010), 707–712, 2010.
[13] A. Morimoto, H. Koyama, D. Inoue, A. Oomichi, H. Nishimura, R. Ashino, and T. Mandai, Image Separation using Wavelet Analysis, Transactions of the Japan Society for Industrial and Applied Mathematics, 19 (3), 257–278, 2009, in Japanese.
[14] D. Napoletani, C. A. Berenstein, P. S. Krishnaprasad, and D. C. Struppa, Quotient signal estimation, in Progress in Mathematics, 238, Birkhäuser, Boston, MA, 151–162, 2005.
[15] O. Yilmaz and S. Rickard, Blind source separation of speech mixtures via time-frequency masking, IEEE Transactions on Signal Processing, 52, 1830–1847, 2004.