15 October 2001

ELSEVIER

Optics Communications 198 (2001) 71–81

OPTICS COMMUNICATIONS

www.elsevier.com/locate/optcom

Artificial neural networks for noisy image super-resolution

Harold Szu, Ivica Kopriva *

Digital Media RF Laboratory, Department of Electrical and Computer Engineering, Room 308, George Washington University, 725 23rd Street NW, Washington, DC 20052, USA

Received 4 June 2001; accepted 15 August 2001

Abstract

Noisy incoherent objects, which are too close to be remotely separated by optically imaging beyond the Rayleigh diffraction limit, might be resolved by employing the artificial neural network (ANN) smart pixel post-processing and its mathematical framework, independent component analysis (ICA). It is shown that ICA ANN approach to super-resolution based on information maximization principle could be seen as a part of the general approach called space-bandwidth product adaptation method. Our success is perhaps due to the blind source separation smart-pixel detectors behind the imaging lens (inverse adaptation), while the Rayleigh diffraction limit remains valid for a single instance of the deterministic imaging systems' realization. The blindness is due to the unknown objects, and the unpredictable propagation effect on the net imaging point spread function. Such a software/firmware enhancement of imaging system may have a profound implication to the designs of the new (third) generation imaging systems as well as other non-optical imaging systems. © 2001 Elsevier Science B.V. All rights reserved.

PACS: 42.79.S; 42.30; 84.35; 42.30.K; 42.25.F

Keywords: Independent component analysis; Optical diffraction; Fourier optics; Unsupervised neural networks; Image processing

1. Introduction

Fourier optics [1–4] shows point sources at the object plane spreading in the far field image plane which sets the Rayleigh resolution criterion: the finest structure that a system can resolve is given by:

$$\delta_{X_{\mathrm{RES}}} = 1.22\lambda f_{\#} = 1.22\frac{\lambda f}{B} \tag{1}$$

where λ is the wavelength, f is the focal length of the lens, and B is the system aperture size. Resolution beyond the classical diffraction limit may be possible by means of bandwidth extrapolation super-resolution methods without noise [2] or space-bandwidth (SW) adaptation process using Wigner distribution based on a priori information [5–7]. The importance of a priori knowledge in super-resolving systems was reported in the early papers about this subject [8–12]. We earlier reported applications of the statistical inversion in reticle-based optical trackers [14–16]. We now extend the one component imaging equation to a set of several components called vector $\boldsymbol{x}$ showing that the noisy super-resolution can be obtained statistically by employing multiple detectors and an ANN/ICA (artificial neural network, ANN;

* Corresponding author. Tel.: +1-202-994-0880; fax: +1-202-994-0223.

E-mail address: ikopriva@seas.gwu.edu (I. Kopriva).

PII: S0030-4018(01)01497-3

independent component analysis, ICA) post-processing. We can recover the unknown source signals vector $\boldsymbol{s}$, assuming that its components are statistically independent, from linear image data vector equation $\boldsymbol{x} = \underline{\boldsymbol{h}}\boldsymbol{s}$ using the software/firmware (e.g. FPGA) post-processing de-mixing ANN weight matrix $\underline{\boldsymbol{w}}$ [17–33]. We insist that components of the recovered vector $\hat{\boldsymbol{s}}$ be as statistically independent as possible giving: $\hat{\boldsymbol{s}} = \underline{\boldsymbol{w}}\boldsymbol{x} \Rightarrow \langle \hat{\boldsymbol{s}}\hat{\boldsymbol{s}}^{\mathrm{T}} \rangle = \langle \underline{\boldsymbol{w}}\boldsymbol{x}(\boldsymbol{x}\underline{\boldsymbol{w}})^{\mathrm{T}} \rangle = \boldsymbol{w}\underline{\boldsymbol{h}}\langle \boldsymbol{s}\boldsymbol{s}^{\mathrm{T}} \rangle \underline{\boldsymbol{h}}^{\mathrm{T}}\boldsymbol{w}^{\mathrm{T}} = \underline{\boldsymbol{d}}$, where the superscript T is the matrix transpose operation, $\underline{\boldsymbol{d}}$ is diagonal matrix (it is identity matrix if the source signals have unit variance) and the de-correlation of sources $\langle \boldsymbol{s}\boldsymbol{s}^{\mathrm{T}} \rangle = \underline{\boldsymbol{d}}$ is the consequence of the incoherence light source assumption. Then the net system transfer $\underline{\boldsymbol{w}\boldsymbol{h}}$ must be the identity matrix $\boldsymbol{I}$ implying that ANN discovered using the ANN gradient ascent search or learning algorithms [17–33], the internal synaptic representation $\underline{\boldsymbol{w}} = \underline{\boldsymbol{h}}^{-1}$ of the external world $\underline{\boldsymbol{h}}$ achieving both: the super-resolution beyond diffraction limit and blind arbitrary de-noising.

2. Problem formulation

We begin with a brief review of the imaging system notation [2, p. 127], and then contrast it with the ICA notation, Fig. 1. If we assume that object illumination is perfectly incoherent i.e. the phasor amplitudes across the object plane are statistically independent then image intensity is obtained as a convolution of the intensity impulse response (p.s.f.) or Green's function with the ideal point object intensity as Eqs. (6)–(15) in Ref. [2]:

$$I_i(u,v,t) = \kappa \int\int_{-\infty}^{\infty} \left| h(u-\tilde{\xi}, v-\tilde{\eta}) \right|^2 I_g(\tilde{\xi},\tilde{\eta},t)\,\mathrm{d}\tilde{\xi}\,\mathrm{d}\tilde{\eta} \tag{2}$$

where κ is some real constant, $(\tilde{\xi},\tilde{\eta})$ are object coordinates normalized on the image plane, (u,v) are image coordinates, $I_g(\tilde{\xi},\tilde{\eta};t)$ is time-varying intensity representation of the object and $h(u,v)$ is the Fraunhofer diffraction pattern (i.e. the far field approximation of the Huygens–Fresnel principle) of the exit pupil $P(x,y)$:

$$h(u,v) = \frac{A}{\lambda z_i} \int\int_{-\infty}^{\infty} P(x,y) \exp\left\{ -\mathrm{j}\frac{2\pi}{\lambda z_i}(ux+vy) \right\} \mathrm{d}x\,\mathrm{d}y \tag{3}$$

Now if we have in the object plane two incoherent point optical sources placed at positions $(\tilde{\xi}_1,\tilde{\eta}_1)$ and $(\tilde{\xi}_2,\tilde{\eta}_2)$ and noise source placed at $(\tilde{\xi}_3,\tilde{\eta}_3)$ then:

$$\begin{aligned} I_{g_1}(\tilde{\xi}_1,\tilde{\eta}_1;t) &= I_{g_1}(t)\delta(\tilde{\xi}-\tilde{\xi}_1,\tilde{\eta}-\tilde{\eta}_1) \\ I_{g_2}(\tilde{\xi}_2,\tilde{\eta}_2;t) &= I_{g_2}(t)\delta(\tilde{\xi}-\tilde{\xi}_2,\tilde{\eta}-\tilde{\eta}_2) \\ I_n(\tilde{\xi}_3,\tilde{\eta}_3;t) &= I_n(t)\delta(\tilde{\xi}-\tilde{\xi}_3,\tilde{\eta}-\tilde{\eta}_3) \end{aligned} \tag{4}$$

Combining Eqs. (2) and (4) we obtain image intensity in the (u,v) plane as:

$$\begin{aligned} I_i(u,v;t) = {} & \\ & \kappa|h(u-\tilde{\xi}_1, v-\tilde{\eta}_1)|^2 I_{g_1}(t) + \kappa|h(u-\tilde{\xi}_2, v-\tilde{\eta}_2)|^2 \\ & \times I_{g_2}(t) + \kappa|h(u-\tilde{\xi}_3, v-\tilde{\eta}_3)|^2 I_n(t) \end{aligned} \tag{5}$$

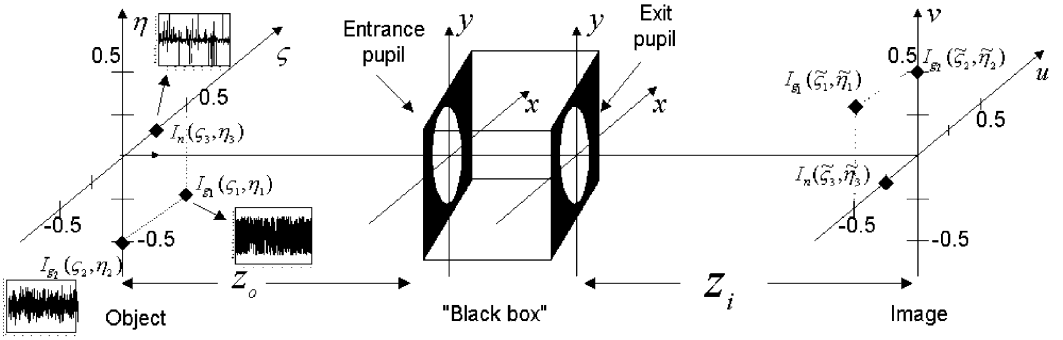

Fig. 1. Generalized model of an imaging system (taken from Ref. [1]).

Importantly, due to diffraction effects the point sources and the noise from the object plane are spread in the image plane. For an ideal imaging case of a circular aperture P with radius w we can compute the p.s.f. and obtain from Eq. (3), Fourier transform of a disk aperture, the weighted Bessel function of order one, the so-called Airy pattern for a uniform intensity at the origin and rippling after a stone having fallen through a water pond [1]:

$$I(r) = \left(\frac{A}{\lambda z_i}\right)^2 \left[2\frac{J_1(kwr/z_i)}{kwr/z_i}\right]^2 \tag{6}$$

where $k = 2\pi/\lambda$ is the wave number and $A = \pi r^2$ is the aperture area, J_1 is the Bessel function of the first kind and r is the radial coordinate. Based on Eq. (6) we rewrite Eq. (5) as:

$$I_i(u, v; t) = \kappa I(r_1) I_{g_1}(t) + \kappa I(r_2) I_{g_2}(t) + \kappa I(r_3) I_n(t)$$

$$|h(u - \tilde{\zeta}_i, v - \tilde{\eta}_i)|^2 = I(r_i)$$
$$r_i = \sqrt{(u - \tilde{\zeta}_i)^2 + (v - \tilde{\eta}_i)^2} \quad i = 1, 2, 3 \tag{7}$$

An imaging system operating at a 10 μm wavelength with an f-number of 2 has a 24.4 μm width of the diffraction blur. It means that diffraction sets the physical limit on increasing the system's resolution if better spatial resolution is required. For example a third generation thermal imaging systems requires 20 or 15 μm pixel size. We illustrate how the ANN smart pixel can exceed the Rayleigh criterion by putting two point sources placed at the relative coordinates $(0.5, -0.5)$ and $(0.0, -0.5)$ and a non-white chaff noise source placed at the relative coordinates $(0.25, 0.0)$ in the object (ζ, η) plane Fig. 1. So the relative distance between the sources is below the Rayleigh's diffraction limit, which is 1.22 in the dimensionless coordinates. We consider an array of smart pixels concentrated on three square pixel detectors with linear dimension 0.5 and each has a 50% filler factor and 50% active sensing area with centers at the positions $(-0.5, 0.5)$, $(0.0, 0.5)$ and $(-0.25, 0.0)$ in the image plane (u, v). This situation is illustrated in the contour diagram in Fig. 2. It shows overlapping in the image plain between the detectors' area and the intensity fluctuation of a spatially fixed point source with the center at $(-0.5, 0.5)$ in the image plane assuming for that particular point of time $I_{g_1}(t) = 1.0$. The gray circle shows area for which applies $I_{g_1}(u, v) \geqslant 0.1$. One can envision similar overlapping situations like Fig. 2 for the second point source with the center at $(0.0, 0.5)$ and for the noise source with the center at $(-0.25, 0.0)$ that however are not shown here due to the lack of space. It should be observed here that we do not exploit the knowledge of the source spatial distribution in blind discrimination method. We have intentionally taken the three sources placed at the relative distance below the Rayleigh's diffraction limit. Then problems caused by diffraction are the most dominant. We could use the measurement vector $\boldsymbol{x}$ with the total dimension $p = k \times l$ where k and l are dimensions of the CCD sensor. That would cause enormous processing difficulties due to the huge number of equations. On the other hand since the influence of diffraction is limited on the few neighboring pixels only it is wise to use the small kernels, 3×3 or 5×5 pixels, by means of which we can cover the whole image. Then it can happen either to have the same number of detectors and sources or to have more detectors then sources in which case we shall benefit from using more sensors than sources [32]. Intensities of the three smart pixels are the appropriate integrals of the intensity distribution

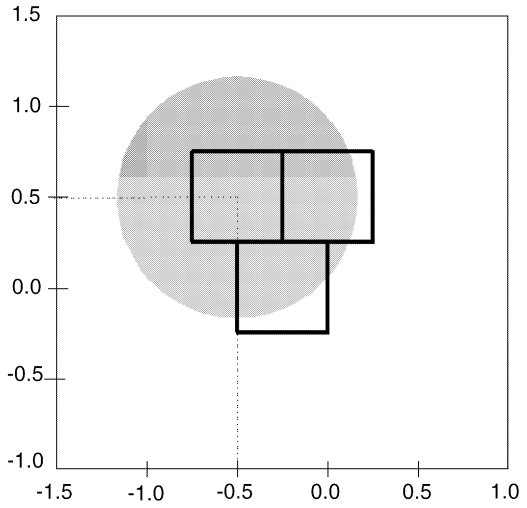

Fig. 2. A snap shot of the first source intensity spreading over imaging domain detectors of three smart pixels.

equation (7) over the three imaging detectors' active sensing areas. If we assume a uniform spatial responsivity we obtain for the first detector:

$$\begin{aligned} I_{\mathrm{D}_1}(t) = & \int_{\tilde{\xi}_1-L/2}^{\tilde{\xi}_1+L/2}\int_{\tilde{\eta}_1-L/2}^{\tilde{\eta}_1+L/2} |h(u-\tilde{\xi}_1, v-\tilde{\eta}_1)|^2 \,\mathrm{d}u\,\mathrm{d}v\, I_{g_1}(t) \\ & + \int_{\tilde{\xi}_1-L/2}^{\tilde{\xi}_1+L/2}\int_{\tilde{\eta}_1-L/2}^{\tilde{\eta}_1+L/2} |h(u-\tilde{\xi}_2, v-\tilde{\eta}_2)|^2 \,\mathrm{d}u\,\mathrm{d}v\, I_{g_2}(t) \\ & + \int_{\tilde{\xi}_1-L/2}^{\tilde{\xi}_1+L/2}\int_{\tilde{\eta}_1-L/2}^{\tilde{\eta}_1+L/2} |h(u-\tilde{\xi}_3, v-\tilde{\eta}_3)|^2 \,\mathrm{d}u\,\mathrm{d}v\, I_n(t) \end{aligned} \tag{8}$$

where $L = 0.5$ is the dimensionless size of the square detector. The real constant κ in Eq. (7) has been absorbed into intensities $I_{g_i}(t)$ in Eq. (8). Intensity expressions for detectors D_2 and D_3 differ only in the integration borders that are around $(\tilde{\xi}_2, \tilde{\eta}_2)$ and $(\tilde{\xi}_3, \tilde{\eta}_3)$ respectively. We can rewrite Eq. (8) in short notation:

$$\begin{aligned} I_{\mathrm{D}_1}(t) &= h_{11} I_{g_1}(t) + h_{12} I_{g_2}(t) + h_{13} I_n(t) \\ I_{\mathrm{D}_2}(t) &= h_{21} I_{g_1}(t) + h_{22} I_{g_2}(t) + h_{23} I_n(t) \\ I_{\mathrm{D}_3}(t) &= h_{31} I_{g_1}(t) + h_{32} I_{g_2}(t) + h_{33} I_n(t) \end{aligned} \tag{9}$$

or in matrix notation:

$$\boldsymbol{x} = \underline{\boldsymbol{h}}\boldsymbol{s} + \boldsymbol{n} \tag{10}$$

where h_{ij} are easily identified from Eq. (8) and their geometrical interpretations partially given in Fig. 2. The additive noise $\boldsymbol{n}$ has been added in matrix equation (10) in order to model the sensor noise if it cannot be assumed to be negligible.

3. Brief review of the independent component analysis theory

The ICA problem is described for a number of source signals coming from different sources and a number of receivers. Each receiver (antenna, microphone, photo-detector, etc.) receives a linear combination of these source signals $\boldsymbol{x} = \underline{\boldsymbol{h}}\boldsymbol{s} + \boldsymbol{n}$; $\boldsymbol{x}, \boldsymbol{s}, \boldsymbol{n} \in \mathbb{R}^N$, $\underline{\boldsymbol{h}} \in \mathbb{R}^{N\times N}$. Neither the structure of the linear combination (the mixing matrix $\underline{\boldsymbol{h}}$) nor the source signals (the vector $\boldsymbol{s}$) are known to the receivers. ICA succeeds to recover the unknown source signals $\boldsymbol{s}$ from the measurements of their linear mixtures $\boldsymbol{x}$ provided they are statistically independent, non-Gaussian (except may be one), as well as that mixing matrix $\underline{\boldsymbol{h}}$ is non-singular. Various kinds of the ICA algorithms recover vector of the unknown source signals $\boldsymbol{s}$ by means of the linear transformation $\hat{\boldsymbol{s}} = \underline{\boldsymbol{w}}\boldsymbol{x}$ minimizing or maximizing certain criterion $\Phi(\underline{\boldsymbol{w}})$ called contrast function [22,24,32] that ensures statistical independence between components of the vector $\hat{\boldsymbol{s}}$ i.e. $p(\hat{\boldsymbol{s}}) = \prod_{i=1}^{N} p_i(\hat{s}_i)$. One of the most popular batch algorithms, that will be used in the simulation experiment reported in Section 4, is the so-called JADE algorithm [24,32], that ensures source separation by joint diagonalization of the fourth-order cumulant matrices:

$$\min_{\underline{\boldsymbol{w}}} \Phi(\underline{\boldsymbol{w}}) = \min_{\underline{\boldsymbol{w}}} \sum_{ijkl\neq iikk} \left| \widehat{C}_4(z_i, z_j, z_k, z_l) \right|^2 \tag{11}$$

where $\widehat{C}_4(z_i, z_j, z_k, z_l)$ are sample estimates of the related fourth-order cross-cumulants i.e. [34,35]

$$\begin{aligned} \widehat{C}_4(z_i, z_j, z_k, z_l) = & \langle z_i z_j z_k z_l \rangle - \langle z_i z_j \rangle \langle z_k z_l \rangle \\ & - \langle z_i z_k \rangle \langle z_j z_l \rangle - \langle z_i z_l \rangle \langle z_j z_k \rangle \end{aligned} \tag{12}$$

In Eq. (11) the sum is over all the quadruples (i, j, k, l) of indices with $i \neq j$ so that for every i, j $(i \neq j)$ we have square matrix defined by $[k, l]$ pairs where $(k, l = 1, \ldots, N)$. Vector $\boldsymbol{z}$ in Eq. (11) represents whitened or standardized version of the measurement vector $\boldsymbol{x}$, Eq. (10), obtained as:

$$\boldsymbol{z} = \underline{\boldsymbol{v}}\boldsymbol{x} \tag{13}$$

such that $E\lfloor \boldsymbol{z}\boldsymbol{z}^{\mathrm{T}} \rfloor = \boldsymbol{I}$ and whitening matrix $\underline{\boldsymbol{v}}$ is obtained:

$$\underline{\boldsymbol{v}} = \underline{\boldsymbol{q}} \Lambda^{-1/2} \underline{\boldsymbol{q}}^{\mathrm{T}} \tag{14}$$

and Λ and $\underline{\boldsymbol{q}}$ are eigenvalue and eigenvector matrices obtained as a solution of the eigenvalue problem:

$$E[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}] = \underline{\boldsymbol{q}} \Lambda \underline{\boldsymbol{q}}^{\mathrm{T}} \tag{15}$$

De-mixing matrix $\underline{\boldsymbol{w}}$ is obtained as the solution of the optimization problem [24,32]:

$$\underline{\boldsymbol{w}} = \arg\min \sum_i \mathrm{off}(\underline{\boldsymbol{w}}^{\mathrm{T}} \widehat{C}_4(z_i, z_j, z_k, z_l) \underline{\boldsymbol{w}}) \tag{16}$$

by using Jacobi method and $\mathrm{off}(\underline{\boldsymbol{a}})$ is measure for the off-diagonality of a matrix defined as:

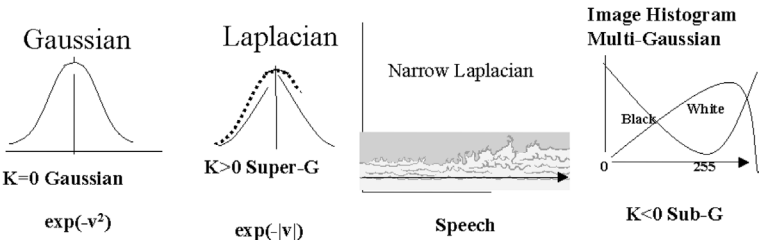

Fig. 3. Kurtosis illustration for different classes of signals.

$$\mathrm{off}(\underline{\boldsymbol{a}}) = \sum_{1 \leqslant i \neq j \leqslant N} |a_{ij}|^2 \tag{17}$$

So by pre-whitening and minimization of the square of the sample estimates of the fourth-order cross-cumulant matrices the second- and fourth-order statistical independence between $\hat{s}_i$ is obtained. This is how the JADE approximates statistical independence. The quality of how the sample estimates of the fourth-order cumulants of the data approximate the real cumulants influences directly the separation performance. Generally, more data points mean more reliable sample estimates and better separation performance. The advantage of criterion equations (11) and (16) over entropy based ICA method [23], is its distribution independence i.e. minimization of Eqs. (11) and (16) cancels fourth-order statistical dependence between components of $\hat{\boldsymbol{s}}$ regardless of their distributions. Additional property of the criteria (11) and (16) is their robustness relative to the additive noise $\boldsymbol{n}$ provided it is Gaussian. This is due to the fact that fourth-order cumulants are blind in relation to Gaussian processes i.e. the Gaussian processes have all the cumulants of the order higher than two equal to zero [34,35]. Algorithm reported in Ref. [23] is in trouble when components of the source vector $\boldsymbol{s}$ belong to both sub-Gaussian and super-Gaussian class of signals. These two classes of signals are distinguished by the value of the parameter called kurtosis defined for the zero mean signal $\hat{s}_i$ as:

$$\kappa(\hat{s}_i) = \frac{C_4(\hat{s}_i)}{C_2^2(\hat{s}_i)} = \frac{E\lfloor \hat{s}_i^4 \rfloor}{E^2\lfloor \hat{s}_i^2 \rfloor} - 3 \tag{18}$$

The sub-Gaussian processes have negative value of the kurtosis, the Gaussian processes have kurtosis equal to zero, while the super-Gaussian signals have positive value of the kurtosis. Typical examples are illustrated by Fig. 3. Examples of super-Gaussian signals are speech and music signals. Examples of super-Gaussian distributions are Laplacian and Cauchy distributions. Examples of sub-Gaussian signals are most of the communication signals and images, while example of the sub-Gaussian distribution is a uniform distribution. Since ICA/ANN based on Eqs. (11) and (16) is batch algorithm it is not suitable for the real time type of applications. If we need on-line i.e. adaptive ICA/ANN we can use the extended *Infomax* algorithm [25], that is also one of the well known ICA algorithms. The Infomax principle to source separation consists of maximizing information transfer through the system of general type (in this case optical system) (Fig. 4) [23]:

$$I(\boldsymbol{y}, \hat{\boldsymbol{s}}) = H(\boldsymbol{y}) - H(\boldsymbol{y}|\hat{\boldsymbol{s}}) \tag{19}$$

where $H(\boldsymbol{y})$ is the entropy of the output non-linearities (sigmoids) while $H(\boldsymbol{y}|\hat{\boldsymbol{s}})$ is the residual entropy that did not come from the input and it has the lowest possible value [23]. Then from Eq. (19) it follows that maximization of the information transfer is equivalent to the maximization of the entropy of the sigmoid outputs i.e.:

$$\max_{\underline{\boldsymbol{W}}} I(\boldsymbol{y}, \hat{\boldsymbol{s}}) = \max_{\underline{\boldsymbol{W}}} H(\boldsymbol{y}) \tag{20}$$

what is the reason why Infomax algorithm is also called the maximum entropy algorithm. Relations between marginal entropy $H(\hat{\boldsymbol{s}})$ and $H(\boldsymbol{y})$, joint entropy $H(\hat{\boldsymbol{s}}, \boldsymbol{y})$, conditional entropy $H(\hat{\boldsymbol{s}}|\boldsymbol{y})$ and $H(\boldsymbol{y}|\hat{\boldsymbol{s}})$ and the mutual information $I(\hat{\boldsymbol{s}}, \boldsymbol{y})$ are illustrated in Fig. 5. Full derivation of the learning equations based on criteria (19) and (20) can be found in Ref. [23]. As a final result maximization

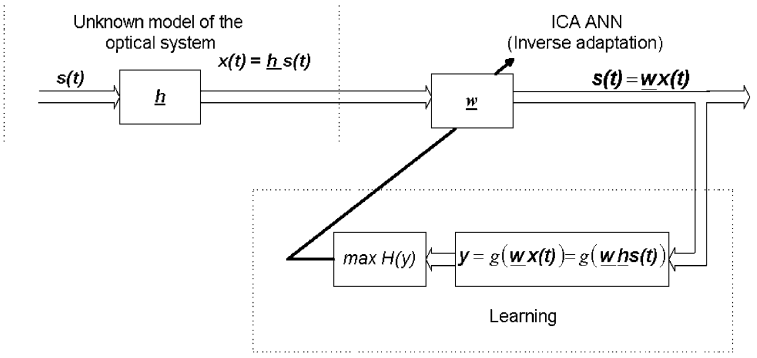

Fig. 4. Information maximization approach to ICA ANN.

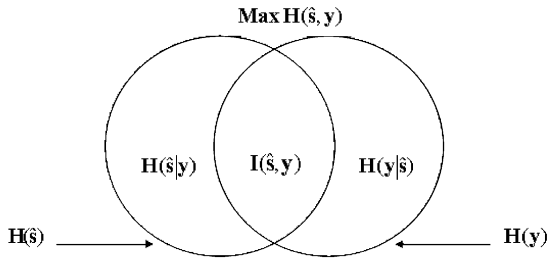

Fig. 5. Marginal entropy $H(\hat{s})$ and $H(\boldsymbol{y})$, joint entropy $H(\hat{s},\boldsymbol{y})$, conditional entropy $H(\hat{s}|\boldsymbol{y})$ and $H(\boldsymbol{y}|\hat{s})$ and the mutual information $I(\hat{s},\boldsymbol{y})$.

of Eq. (20) using the natural [21] or relative gradient [31] for faster convergence gives the following learning rule [21,23,25]:

$$\underline{\boldsymbol{w}}(\boldsymbol{k}+1) = \underline{\boldsymbol{w}}(\boldsymbol{k}) + \mu\lfloor \boldsymbol{I} - \varphi(\hat{\boldsymbol{s}})\hat{\boldsymbol{s}}^{\mathrm{T}}\rfloor \underline{\boldsymbol{w}}(\boldsymbol{k}) \tag{21}$$

where from both the information maximization and maximum likelihood maximization approach the following requirement is imposed on the non-linearity $\varphi(\circ)$ [28,29]:

$$\varphi(\hat{\boldsymbol{s}}) = -\frac{\mathrm{d}}{\mathrm{d}\hat{\boldsymbol{s}}}\log p(\hat{\boldsymbol{s}}) \tag{22}$$

and $p(\hat{\boldsymbol{s}})$ is the true p.d.f. of the source signals. At this point it can be seen how ICA ANN approach to the super-resolution depends on the a priori information related to the p.d.f.'s of the intensity distribution of the input signals. It was pointed out in Refs. [5–7] how a priori information is important in order to obtain optimal separation i.e. to attain optimal super-resolution. Based on Fig. 6 it

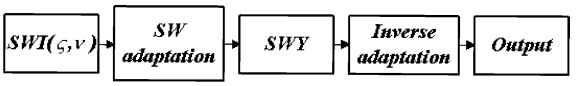

Fig. 6. General block diagram of the SW adaptation process, taken from Ref. [5].

can be seen that ICA ANN approach to the super-resolution is a part of the general approach called the SW product adaptation process [5–7]. What is in common to our ICA ANN approach to the super-resolution and SW product shape adaptation method [5,7] is that something has to be done in order to maximize information transfer through the system. According to Refs. [5–7] the problem of super-resolution is the problem of the adaptation of the shape of Wigner distribution chart of the input (SWI) such that it matches the shape of the system transfer distribution chart (SWY). Examples of such adaptations are given in Refs. [6,9,12]. In the SW product adaptation method certain transformations, that will ensure maximization of the information transfer, are possible if we have some a priori information about the signal that can be related to the object shape, temporally restricted signals, etc. [5–12]. In the information maximization based ICA we can attain maximum of the information transfer through the system (optimal signal recovery) if our non-linearities match the shape of the p.d.f. of the data (Eq. (22) and Fig. 4). The influence of the optical system is present in the form of the mixing matrix $\underline{\boldsymbol{h}}$ (Eqs. (9) and (10)). The *inverse adaptation* block in Fig. 6 is replaced by the de-mixing matrix $\underline{\boldsymbol{w}}$ such that in

ideal case it will be $\underline{\boldsymbol{w}} = \boldsymbol{h}^{-1}$. In the blind source separation scenario true p.d.f.'s are basically unknown. However, it was shown in Ref. [27] that learning rule Eqs. (21) and (22) will still be super-efficient provided that $p(\hat{\boldsymbol{s}})$ are even and $\varphi(\hat{\boldsymbol{s}})$ are odd. Then $\varphi(\hat{\boldsymbol{s}}) = \tanh(\hat{\boldsymbol{s}})$ is an appropriate choice for a class of super-Gaussian signals, while $\varphi(\hat{\boldsymbol{s}}) = 2\hat{\boldsymbol{s}} - \mathrm{sign}(\hat{\boldsymbol{s}})\hat{\boldsymbol{s}}^2$ is suitable for sub-Gaussian signals. The adaptive algorithm capable to cope with both sub-Gaussian and super-Gaussian sources simultaneously is given with [25]:

$$\Delta\underline{\boldsymbol{w}} \propto [\boldsymbol{I} - \boldsymbol{K}\tanh(\hat{\boldsymbol{s}})\hat{\boldsymbol{s}}^{\mathrm{T}} - \hat{\boldsymbol{s}}\hat{\boldsymbol{s}}^{\mathrm{T}}]\underline{\boldsymbol{w}} \times \begin{cases} k_i = 1 & \text{super-Gaussian} \\ k_i = -1 & \text{sub-Gaussian} \end{cases} \tag{23}$$

where k_i are elements of the N-dimensional diagonal matrix $\boldsymbol{K}$. Then k_i can be estimated from [25]:

$$k_i = \mathrm{sign}(E\{\mathrm{sech}^2(\hat{s}_i)\}E\{\hat{s}_i^2\} - E\{[\tanh(\hat{s}_i)]\hat{s}_i\}) \tag{24}$$

and $E(\mathrm{o})$ is the expectation operator. More details about foundations of the ICA theory can be found in Refs. [17–33].

4. Simulation results

It can be easily observed that vector version of Eq. (9), $\boldsymbol{x} = \underline{\boldsymbol{h}}\boldsymbol{s} + \boldsymbol{n}$, is basically the ICA problem, where components of the column vector $\boldsymbol{x} = [I_{\mathrm{D}_1}(t), I_{\mathrm{D}_2}(t), I_{\mathrm{D}_3}(t)]^{\mathrm{T}}$ are measured signals, components of the vector $\boldsymbol{s} = [I_{g_1}(t), I_{g_2}(t), I_n(t)]^{\mathrm{T}}$ are unknown source signals while h_{ij}, $i, j \in \{1, 2, 3\}$ are elements of the unknown mixing matrix $\underline{\boldsymbol{h}}$ and $\boldsymbol{n}$ is the additive noise vector that models the sensor noise. Here we see that noise is included in the model both as a part of the source signal vector to model diffraction noise and as a sensor noise in term of additive noise. It is known from ICA theory that in principle N sensors must be used in order to recover the N signals [17–32], although it has been shown in Ref. [33] how it is possible to recover several sources by using two sensors only. Here we shall use three detectors. As it was pointed out before we do not exploit here the knowledge of the source spatial position. We have intentionally chosen the case when the three sources are at the relative distance below the Rayleigh's diffraction limit. Now we recover signals emitted by two 'information' point sources as well as noisy signal emitted by the noise source. In doing so we obtain both: super-resolution beyond diffraction limit and de-noising. If we model noise as the Cauchy color noise then non-Gaussian nature of the noise is for ICA not limiting factor, as it is for some other noise canceling methods [36], but useful property. [1] By described ICA methodology we can recover unknown source signals using the software/firmware (FPGA) post-processing de-mixing matrix $\underline{\boldsymbol{w}}$ such that $\langle \underline{\boldsymbol{w}}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\underline{\boldsymbol{w}}^{\mathrm{T}}\rangle = \boldsymbol{s}$. Then, because of the three detectors imaging $\boldsymbol{x} = \underline{\boldsymbol{h}}\boldsymbol{s} + \boldsymbol{n}$ we have consequently obtained [17–19], $\underline{\boldsymbol{w}}\underline{\boldsymbol{h}}\langle \boldsymbol{s}\boldsymbol{s}^{\mathrm{T}}\rangle \underline{\boldsymbol{h}}^{\mathrm{T}}\underline{\boldsymbol{w}}^{\mathrm{T}} = \underline{\boldsymbol{d}}$, where $\underline{\boldsymbol{d}}$ is diagonal matrix (identity matrix if source signals have unit variance) and $\langle \boldsymbol{s}\boldsymbol{s}^{\mathrm{T}}\rangle = \underline{\boldsymbol{d}}$ is the consequence of the incoherence assumption. Beside to the non-Gaussianity assumption we shall additionally assume statistical independence of the point sources since radiation is emitted from the three physically separated sources. Consequently the net transfer function must equal to the identity matrix $\underline{\boldsymbol{w}}\underline{\boldsymbol{h}} = \boldsymbol{I}$ meaning that by post-processing we have found de-mixing matrix $\underline{\boldsymbol{w}} = \underline{\boldsymbol{h}}^{-1}$. Non-Gaussianity assumption is fulfilled if for example at least one point source has uniform intensity distribution and noise source has for example Cauchy distribution (Fig. 1) [26]. Mixing matrix non-singularity requirement is transformed into:

$$\begin{aligned} &\det \underline{\boldsymbol{h}} = h_{11}\Delta_{11} - h_{12}\Delta_{12} + h_{13}\Delta_{13} \neq 0 \\ &\Delta_{11} = h_{22}h_{33} - h_{23}h_{32} \\ &\Delta_{12} = h_{21}h_{33} - h_{31}h_{23} \\ &\Delta_{13} = h_{21}h_{32} - h_{31}h_{22} \end{aligned} \tag{25}$$

We can identify h_{ij} from their geometrical interpretations given by Fig. 2. that shows h_{11}, h_{21} and h_{31}. h_{ij} tells us how much the jth source is spread over ith detector sensing area. From Eq. (25) it can be seen that only first term will be product of the three most dominant components h_{11}, h_{22} and h_{33} while others will always contain products of at least two off-diagonal terms making

[1] This choice is physically justified since impulse noise can be modeled as Cauchy noise that is highly non-Gaussian process.

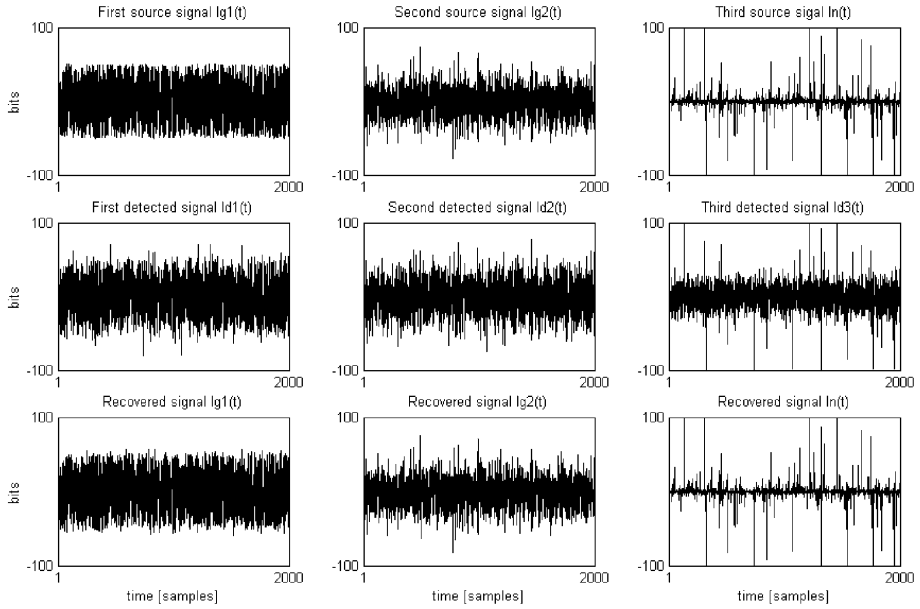

Fig. 7. ANN/ICA approach to super-resolution: first row is the intensity fluctuations in time of the fixed point source signals $\boldsymbol{s}^{\mathrm{T}}(t) = [I_{g_1}(\tilde{\xi}_1, \tilde{\eta}_1; t) I_{g_2}(\tilde{\xi}_2, \tilde{\eta}_2; t) I_n(\tilde{\xi}_3, \tilde{\eta}_3; t)]$; second row, detected signals $\boldsymbol{x}^{\mathrm{T}}(t) = \lfloor I_{\mathrm{D}_1}(t) I_{\mathrm{D}_2}(t) I_{\mathrm{D}_3}(t) \rfloor$; third row, recovered signals $\hat{\boldsymbol{s}}^{\mathrm{T}}(t) = \lfloor \hat{I}_{g_1}(\tilde{\xi}_1, \tilde{\eta}_1; t) \hat{I}_{g_2}(\tilde{\xi}_2, \tilde{\eta}_2; t) \hat{I}_n(\tilde{\xi}_3, \tilde{\eta}_3; t) \rfloor$.

them small relative to the first term implying that $\det \underline{\boldsymbol{h}}$ will be always positive. So the non-singularity requirement is in principle also satisfied. For described arrangement of detectors and sources, illustrated with Fig. 2, the mixing matrix is given with:

$$\underline{\boldsymbol{h}} = \begin{bmatrix} 0.8850 & 0.4902 & 0.4194 \\ 0.4902 & 0.8850 & 0.4194 \\ 0.4194 & 0.4194 & 0.8850 \end{bmatrix} \tag{26}$$

with $\det \underline{\boldsymbol{h}} = 0.3416$. This discussion shows that the requirements necessary for the ICA theory to work can in principle be fulfilled in order to achieve the so-called super-resolution i.e. resolution beyond the classical diffraction limit. So we have gone statistically beyond the resolution limit set by the deterministic diffraction theory. This concludes our statistical software/firmware approach to the deterministic super-resolution problem. The robustness to the noisy imaging environment is anticipated by the virtue of statistical approach similar to Wiener's regulation the noisy inverse filtering for a dual imaging system [13], but not yet experimentally demonstrated. We illustrate the exposed approach to super-resolution and denoising in Fig. 7 by application of the JADE algorithm [24,32] on the mixture of the three source signals whose intensities are changing in time according to:

$$\begin{aligned} I_{g_1}(t) &= 50\,[2\ \mathrm{rand}(t) - 1] \\ I_{g_2}(t) &= 20\ \mathrm{randn}(t) \\ I_n(t) &= \tan(\theta); \quad \theta = \pi\,[2\ \mathrm{rand}(t) - 1] \end{aligned} \tag{27}$$

where $\mathrm{rand}(t)$ stands for uniform and $\mathrm{randn}(t)$ for normal distribution. Signals are mixed with the matrix $\underline{\boldsymbol{h}}$ estimated from the given source–detectors arrangement and given by Eq. (26). Here we expressed intensity values in bits assuming intensity fluctuations (random or modulated) relative to some DC level (for 10-bit image DC level would be 512). That explains why relative intensities in Eq. (27) can be negative. Time evolution means we are processing sequence of images. The chosen example of signals Eq. (27) is the most difficult one from the ICA theory point of view since we have one sub-Gaussian signal: $I_{g_1}(t)$, one pure Gaussian signal: $I_{g_2}(t)$ and one super-Gaussian signal: $I_n(t)$ with the extremely high value of kurtosis (about

few thousands). With such difficult combination of signals we wanted to emphasize the power of the ICA methodology. First row in Fig. 7 represents source signals: $I_{g_1}(t)$, $I_{g_2}(t)$ and $I_n(t)$ from left to right respectively according to Eq. (27). Second row represents mixed signals $\boldsymbol{x} = \underline{\boldsymbol{h}}\boldsymbol{s} + \boldsymbol{n}$ according to Eqs. (9) and (10). It shows to us what, due to diffraction phenomenon, three detectors would see in time. In the scenario shown in Fig. 7 additive noise at -20 dB relative to the first source signal was generated, see Eq. (20). Third row represents de-mixed and de-noised signals: $\hat{I}_{g_1}(t)$, $\hat{I}_{g_2}(t)$ and $\hat{I}_n(t)$ again from left to right. We have used $T = 2000$ data samples to obtain the sample estimates of the related fourth-order cross-cumulants, Eq. (12). Diffraction effects are reduced significantly. Horizontal axis represents time in discrete samples and vertical axis is in bits relative to the DC level. In order to estimate the separation performance more reliably we have computed the signal to interference ratio (SIR) for each of the source signals according to:

$$\mathrm{SIR}(s_i) = 10\log 10\left(\frac{\mathrm{RMSE}(\hat{s}_i)}{\mathrm{RMSE}(x_i)}\right) \tag{28}$$

where

$$\mathrm{RMSE}(\hat{s}_i) = \sqrt{\frac{1}{T}\sum_{k=1}^{T}(s_i(k) - \hat{s}_i(k))^2}$$

$$\mathrm{RMSE}(x_i) = \sqrt{\frac{1}{T}\sum_{k=1}^{T}(s_i(k) - x_i(k))^2} \quad i = 1, 2, 3 \tag{29}$$

where T is the sample size here assumed to be $T = 2000$. The SIR parameter measures the quality of the signal recovery in term of the RMSE criteria of the recovered signal relative to the RMSE of the directly measured data. Fig. 8 shows SIR values in dB for different values of the additive noise level that is expressed relative to the first source signal level i.e. the horizontal axis of Fig. 8 is defined as:

$$\mathrm{SNR}_{[\mathrm{dB}]} = 10\log 10\left(\frac{\sigma_n}{\sigma_{s_1}}\right) \tag{30}$$

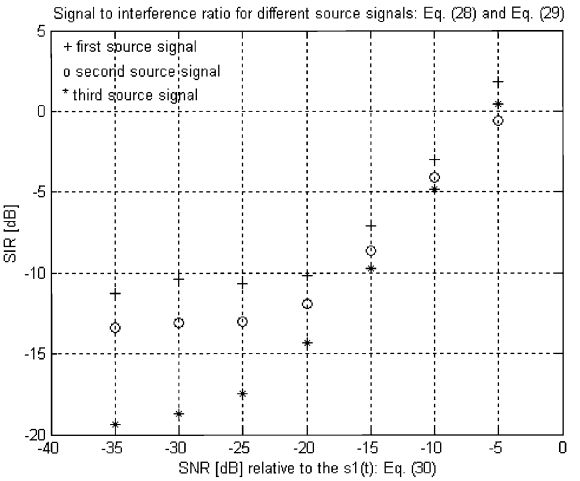

Fig. 8. Performance of the ANN/ICA approach to super-resolution in terms of SIR.

where $\sigma_{\boldsymbol{n}}$ and σ_{s_1} denote variance of the additive noise $\boldsymbol{n}$ and variance of the source signal s_1, respectively. We have denoted the noiseless case with -35 dB of SNR in order to be able to show it on the Fig. 8. The SIR values were estimated by averaging results over the 100 iterations. The sensor noise was modeled as Gaussian noise. We can see that if additive noise is neglected, for example sensors are cooled enough, the RMSE value after applying ICA on the measured data relative to the RMSE value of directly measured data (i.e. without post-processing) is 13.5 times smaller for the first source signal, 22 times smaller for the second source signal and even 86 times smaller for the third source signal. This gives to us illustration how much the diffraction effects could be reduced by means of the ICA post-processing. For SNR $= -5$ dB the SIR performance is approximately the same for directly measured and post-processed data.

5. Conclusion

The ANN statistical approach called ICA is employed to resolve two incoherent point sources with noise that are too close to be resolved under Rayleigh diffraction limit. It has been shown that ICA theory assumptions, statistical independence

and non-Gaussianity of the source signals and non-singularity of the mixing matrix, are fulfilled in principle in this scenario. Moreover mixing matrix elements are the point spread functions of the imaging system that can be identified in a blind way. It was also shown that ICA ANN approach to the super-resolution based on information maximization principle could be seen as a part of the general approach called SW product adaptation method. Simulation has demonstrated that by three detectors and ANN/ICA based post-processing not only super-resolution but also de-noising is attained. In the no additive noise scenario from 13 to even 86 times improvement after the ICA post-processing (depending on the assumed intensity distribution) is obtained relative to the directly measured data reducing diffraction effects significantly. We wish to propose the next step involving the costly but indispensable field tests, employing the smart pixels' detectors and the back-plane FPGA firmware or ASIC implementation of ANN/ICA algorithm for a rapid real-time post-processing, to demonstrate the noisy super-resolution and blind de-noise. We believe this third generation smart pixel imaging system will have an important impact on the designs of other, non-optical imaging systems.

References

[1] J.W. Goodman, Fresnel and Fraunhofer diffraction, Introduction to Fourier Optics, second ed., McGraw-Hill, New York, 1996 (Chapter 4; pp. 63–90).
[2] J.W. Goodman, Frequency analysis of optical imaging systems, Introduction to Fourier Optics, second ed., McGraw-Hill, New York, 1996 (Chapter 6; pp. 126–165).
[3] M. Born, E. Wolf, Interference and diffraction with partially coherent light, Principles of Optics, sixth ed., Cambridge, 1980 (Chapter 10; pp. 491–554).
[4] J.W. Goodman, Coherence of optical waves, Statistical Optics, Wiley, New York, 1985 (Chapter 5; pp. 157–236).
[5] A.W. Lohman, D. Mendlovic, Space-bandwidth product adaptation and its application to superresolution: fundamentals, J. Opt. Soc. Am. A 14 (1997) 558–562.
[6] A.W. Lohman, D. Mendlovic, Space-bandwidth product adaptation and its application to superresolution: examples, J. Opt. Soc. Am. A 14 (1997) 563–567.
[7] A.W. Lohman, R.G. Dorsch, D. Mendlovic, Z. Zalevsky, C. Ferreira, Space-bandwith product of optical signals and systems, J. Opt. Soc. Am. A 13 (1996) 470–473.
[8] W. Lukosz, Optical systems with resolving powers exceeding the classical limit, J. Opt. Soc. Am. 56 (1966) 1463–1472.
[9] W. Lukosz, Optical systems with resolving powers exceeding the classical limit. II, J. Opt. Soc. Am. 57 (1967) 932–941.
[10] G. Toraldo di Francia, Resolving power and information, J. Opt. Soc. Am. 45 (1955) 497–501.
[11] G. Toraldo di Francia, Degrees of freedom of an image, J. Opt. Soc. Am. 59 (1969) 799–804.
[12] M.A. Grimm, A.W. Lohman, Superresolution image for one-dimensional objects, J. Opt. Soc. Am. 56 (1966) 1151–1156.
[13] H.H. Szu, Matched filter spectrum shaping for light efficiency, Appl. Opt. 24 (1985) 1426–1431.
[14] H.H. Szu, I. Kopriva, A. Persin, Independent component analysis approach to resolve the multi-source limitation of the nutating rising-sun reticle based optical trackers, Opt. Commun. 176 (2000) 77–89.
[15] I. Kopriva, A. Persin, Discrimination of optical sources by use of adaptive blind source separation theory, Appl. Opt. 38 (1999) 1115–1126.
[16] I. Kopriva, H.H. Szu, Blind Discrimination of the Coherent Optical Sources by Using Reticle Based Optical Trackers is a Nonlinear ICA Problem, in: P. Pajunen, J. Karhunen (Eds.), Proceedings of the 2nd International Workshop on Independent Component Analysis and Blind Signal Separation, Helsinki, Finland, 19–22 June 2000, pp. 51–56.
[17] H.H. Szu, Thermodynamics energy for both supervised and unsupervised learning neural nets at a constant temperature, Int. J. Neural Syst. 9 (1999) 175–186.
[18] H.H. Szu, Progresses in unsupervised artificial neural networks of blind image demixing, IEEE Ind. Elec. Soc. Newsl. 46 (1999) 7–12.
[19] H.H. Szu, ICA-an enabling technology for intelligent sensory processing, IEEE Circ. Syst. Newsl. (1999) 14–41.
[20] C. Jutten, J. Herault, Blind separation of sources, Part I: an adaptive algorithm based on neuromimetic architecture, Signal Process. 24 (1991) 1–10.
[21] S. Amari, Natural gradient works efficiently in learning, Neural Comput. 10 (1998) 251–276.
[22] P. Common, Independent component analysis, A new concept?, Signal Process. 36 (1994) 287–314.
[23] A.J. Bell, T.J. Sejnowski, An information-maximization approach to blind separation and blind deconvolution, Neural Comput. 7 (1995) 1129–1159.
[24] J.F. Cardoso, A. Souloumiac, Blind beamforming for non Gaussian signals, IEE-Proc. F 140 (1993) 362–370.
[25] T.W. Lee, M. Girolami, T.J. Sejnowski, Independent component analysis using an extended Infomax algorithm for mixed sub-Gaussian and super-Gaussian sources, Neural Comput. 11 (1999) 409–433.
[26] H.H. Szu, R. Hartley, Fast simulated annealing, Phys. Lett. A 122 (1987) 157–162.
[27] S. Amari, Superefficiency in blind source separation, IEEE Trans. Signal Process. 47 (1999) 936–944.

[28] D.T. Pham, Blind separation of mixtures of independent sources through a quasimaximum likelihood approach, IEEE Trans. Signal Process. 45 (1997) 1712–1725.
[29] J.F. Cardoso, Infomax and maximum likelihood for blind source separation, IEEE Signal. Process. Lett. 4 (1997) 112–114.
[30] S. Amari, A. Cihocki, H.H. Yang, Blind signal separation and extraction: neural and information theoretic approaches, in: S. Haykin (Ed.), Unsupervised Adaptive Filtering-Volume I- Blind Source Separation, Wiley, New York, 2000, pp. 63–138 (Chapter 3).
[31] J.F. Cardoso, B. Laheld, Equivariant adaptive source separation, IEEE Trans. Signal Process. 44 (1996) 3017–3030.
[32] J.F. Cardoso, High-order contrasts for independent component analysis, Neural Comput. 11 (1999) 157–192.
[33] A. Jourjine, S. Rickard, Ö. Yilmaz, Blind separation of disjoint orthogonal signals: demixing N sources from 2 mixtures, Proceedings of the ICASSP, Istanbul, Turkey, 5–9 June 2000, pp. 2985–2988.
[34] J.M. Mendel, Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications, Proc. IEEE 79 (1991) 278–305.
[35] P. McCullagh, Elementary theory of cumulants Tensor Methods in Statistics, Chapman and Hall, London, 1987, 1995 (Chapter 2; pp. 24–26).
[36] B. Widrow, J. Glover, J. McCool, J. Kaunitz, C. Williams, R. Hearn, J. Zeidler, E. Dong, R. Goodlin, Adaptive noise cancelling: principles and applications, Proc. IEEE 63 (1975) 1692–1716.