DETERMINISTIC BLIND SOURCE SEPARATION FOR SPACE VARIANT IMAGING

Harold Szu and Ivica Kopriva
Digital Media RF Laboratory
Department of Electrical and Computer Engineering
George Washington University
725 23rd Street NW, Washington DC 20052, USA
e-mail: ikopriva@seas.gwu.edu; SZUH@onr.navy.mil

Abstract

We consider a deterministic approach to the noise free blind image separation and deconvolution problem with positivity constraints. This is necessary because in some real world applications (telescope images in astronomy, remotely sensed images, etc.) the pixel values correspond to intensities and must be positive. Also mixing matrix itself must be positive if it for example represents point spread function of an imaging system in astronomy or spectral reflectance matrix in remote sensing. In related papers the blind source separation (BSS) problem with positivity constraints is being solved by using probabilistic approach assuming independence between the sources that requires use of all the pixel data. Implicit assumption of this approach is that unknown mixing matrix is space invariant. Here we propose solution that is deterministic and solves the problem on the pixel by pixel basis. Consequently, algorithm is capable to solve the space variant problems. This is accomplished by minimizing the 2^{nd} law of thermodynamics based contrast function called Helmholtz free energy. Formulation of our algorithm is equivalent to the MaxEnt formulation of the supervised separation problem with essential difference that the mixing matrix is unknown in our case. We demonstrate the algorithm capability to perfectly recover images from the synthetic noise free linear mixture of two images.

1.0 Introduction

In real world applications such as telescope images in astronomy or remotely sensed images the pixel values correspond to intensities and must be positive, [1-3,11,12,14]. Also mixing matrix itself must be positive if it for example represents point spread function of an imaging system, [19], or spectral reflectance matrix in remote sensing, [1-3]. As it has already been noticed in [12], the standard BSS approaches, [5-8], do not take into account these positivity constraints and that can lead to reconstructed images that have areas of negative intensity, [12]. The so-called non-negative ICA methods that explicitly take into account these positivity constraints are described in [9-14]. Like other ICA methods they are probabilistic methods and rely on the priors for the source pixels to be mixture of Laplacians with high probability for positive values around zero and zero probability for the negative values. As it has been proven in [9] the non-vanishing pdf in the positive neighborhood around zero is the condition for non-negative ICA. These probabilistic assumptions implicitly assume that unknown mixing matrix is space invariant. Here we propose solution that is deterministic and solves the problem on the pixel by pixel basis. Hence, we may assume unknown mixing matrix to be space variant which is important property for detection of small objects on the sub-pixel level that happens both in remote sensing, [2], and thermography based early cancer detection, [24]. In our approach we also do not insist the unknown sources to be statistically independent. Instead, in accordance with the principle of maximum entropy we generate for the source vector on the single pixel level the most probable outcome or distribution with maximal entropy under given constraints. This is accomplished by minimizing the 2^{nd} law of thermodynamics based contrast function called Helmholtz free energy. We remarque here that physics inspired approach to ICA has already been reported in [19] where analogy between information theoretic learning based on the Renyi's entropy and classical mechanics has been established. Formulation of our algorithm is equivalent to the MaxEnt solution of the problem (1), [16-17], with an essential difference that mixing matrix is unknown in our case. The solution is found by employing combinatorial optimization method to find possibly global minimum of the error energy function, [20, 21] and MaxEnt like algorithm to find the distribution of the source vector with the maximal entropy under given constraints. Derivation of the algorithm is given in Section 2. We demonstrate the algorithm capability to perfectly recover images from the synthetic noise free linear mixture of two images in Section 3 while conclusion is given in Section 4.

2.0 A Deterministic BSS Algorithm

The BSS problem is defined with:

$$\vec{X}_{(p,q)} = [A]_{(p,q)}\vec{S}_{(p,q)} \qquad (1)$$

where $\vec{X}$ and $\vec{S}$ are n and m dimensional column vectors of integers representing measured data and unknown sources respectively with $m \leq n$ and $[A]$ being nxm unknown mixing matrix. The subscript (p,q) denotes spatial coordinates i.e. the BSS problem is formulated on the pixel by pixel basis. Note that such formulation allows the mixing matrix $[A]_{(p,q)}$ to be spatially variant. We shall drop (*p,q*) subscript in the subsequent derivations in order to simplify notation. Because both $\vec{X}$ and $\vec{S}$ have the physical interpretation of intensity the positivity constraint is imposed on them:

$$\begin{aligned} x_i &\geq 0 \quad i = 1,..., n \\ s_i &\geq 0 \quad i = 1,..., m \end{aligned} \qquad (2a)$$

If the unknown mixing matrix has physical interpretation of the spectral reflectance matrix as in remote sensing, [1-3], or point spread function of the optical or non-optical imaging system, [19], than positivity constraints must be imposed on [*A*] too:

$$a_{ij} \geq 0 \; i = 1,..,n; \; j = 1,..,m \qquad (2c)$$

We shall rewrite (1) in a slightly different form:

$$\vec{X} = [A]N\vec{S}' \qquad (3)$$

where:

$$N = \sum_{i=1}^{m} s_i \qquad (4a)$$

$$s_i' = \frac{s_i}{N} \qquad (4b)$$

With (3) we have introduced unknown scaling factor *N* that helped us to assign to the components of the scaled source vector $\vec{S}'$ the meaning of probability because due to (4) they satisfy the constraint:

$$\sum_{i=1}^{m} s_i' = 1 \qquad (5)$$

We now formulate the Helmholtz free energy as a contrast function with probability constraint (5) taken into account explicitly as:

$$\begin{aligned} H([W], \vec{S}) &= U - T_0 S \\ &= \vec{m}^T \left[[W]\vec{X} - N\vec{S}' \right] + K_B T_0 N \sum_{i=1}^{m} s_i' \ln s_i' \\ &+ N(m_0 - K_B T_0)\left(\sum_{i=1}^{m} s_i' - 1 \right) \end{aligned} \qquad (6)$$

where Shannon entropy was approximated by:

$$S = -K_B T_0 \sum_{i=1}^{m} s_i' \ln s_i' \qquad (7)$$

where K_B represents Boltzmann's constant and T_0 represents temperature. They are introduced in (6) due to dimensionality reasons. We remarque here that our contrast function (6) is equivalent to the formulation of the solution of problem (1) using the principle of maximum entropy, [16-17], which gives as a solution distribution $p(\vec{S})$ with the maximal entropy under given macroscopic constraints $U = \left([W]\vec{X} - N\vec{S}' \right)$ defined by measured data $\vec{X}$:

$$\sum_{i=1}^{n} w_{ji} x_i = N s_j' \qquad (8)$$

where [*W*] is an *mxn* matrix and for the case *m*=*n* $[W] = [A]^{-1}$. Essential difference between our algorithm and classical maximum entropy solution [16-17] is that mixing matrix [*A*] and/or the unmixing matrix [*W*] are both unknowns. To solve the BSS imaging problem with the positivity constraints we formulate an algorithm as a combination of the global optimization algorithm, [20-23], that looks for the global minimum of the error energy function:

$$\left([W^*], N^* \right) = \arg\min \left([W]\vec{X} - N\vec{S}' \right)^T \left([W]\vec{X} - N\vec{S}' \right) \qquad (9)$$

and maximum entropy algorithm for finding the most probable distribution $p(\vec{S}) = \vec{S}'$ for a given doublet $\left([W^{(l)}], N^{(l)} \right)$ where *l* denotes iteration index in a solution of problem (9). We now formulate deterministic BSS algorithm in a form of four theorems.

Theorem 1. Equilibrium Partition Function of Helmholtz free energy yields McCullough-Pitt Sigmoid Threshold:

$$s_j' = \frac{1}{1 + \sum_{\substack{i=1 \\ i \neq j}}^{m} \exp\left(\frac{1}{K_B T_0} (m_i - m_j) \right)} = s(\vec{\mu}) \qquad (10)$$

Proof. We differentiated the Helmholtz free energy (6), set it to zero to solve for the source probability, and summed it to one to give the partition function:

$$\frac{\partial H(\vec{S})}{\partial s_j'} = -m_j N + N K_B T_0 \left(\ln s_j' \right) + N m_0 = 0$$

from which it follows:

$$s_j' = \exp\left(\frac{m_j - m_0}{K_B T_0} \right) \qquad (11)$$

Imposing the constraint of probability normalization condition $\sum_{j=1}^{m} s'_j = 1$, we obtained the partition function of the canonical ensemble in Statistical Mechanics:

$$\exp\left(\frac{m_0}{K_B T_0}\right) = \sum_{j=1}^{m} \exp\left(\frac{m_j}{K_B T_0} \right) \equiv Z \qquad (12)$$

eliminating the free energy m_0 in (11) we get the sigmoid threshold of the normalized source component s_j' (10). **Q.E.D**.

Theorem 2 Variation of the thermodynamic free energy value m_o with respect to virtual Lagrange constraint forces m_j gives displacement of virtual sources s'_j.

Proof. From (12) and (10) we verify that relation between the free energy m_0 and virtual forces m_j satisfy:

$$\frac{\partial m_0}{\partial m_j} = s_j' \qquad (13)$$

what is also condition that Lagrange multipliers must satisfy in the maximum entropy formalism [16-17]. **Q.E.D.**

Theorem 3. Virtual Lagrange forces m_j are changed in the direction of minimal displacement from data.

Proof. Using the perturbation theory, $\Delta m_j = \sum_{i=1}^{m} \frac{\partial m_j}{\partial s_i'} \Delta s_i'$ we derived the iterative update rule for the Lagrange multipliers with respect to data error:

$$m_j^{(k+1)} = m_j^{(k)} + \left(\frac{K_B T_0}{s_j'^{(k)}} + m_j^{(k)} \right) \left(\vec{w}_j^{(l)} \vec{X} - N^{(l)} s_j'^{(k)} \right) + \sum_{\substack{i=1 \\ i \neq j}}^{m} m_i^{(k)} \left(\vec{w}_i^{(l)} \vec{X} - N^{(l)} s_i'^{(k)} \right) \tag{14}$$

where k stands for iteration index related to the Lagrange multipliers learning rule and l stands for the iteration index related to the iterative solution of the optimization problem (9). In iterative MaxEnt algorithm free energy $m_0^{(k+1)}$ is computed from (12) and source probability $s_j'^{(k+1)}$ is computed from (11). **Q.E.D**.

Theorem 4. The unknown de-mixing matrix $[W]$ is determined at the minimum of the absolute value of the Helmholtz free energy.

Proof. Because the Shannon entropy S is a convex function that does not depend on the de-mixing matrix $[W]$ one could add it to Eq. (6), yielding:

$$\min_{[W]} \left| H + T_0 S \right| \cong \min_{[W]} \left| H \right| \cong \min_{[W]} \left| U \right| = \sqrt{\sum_{i=1}^{m} \left(m_i \left(\vec{w}_i \vec{X} - N s_i' \right) \right)^2} \tag{15}$$

which shows that minimisation of the data error energy $\left| U \right|$ is equivalent to the minimisation of the Helmholtz free energy $\left| H \right|$. The initial point of the unknown scaling constant $N = \sum_{j=1}^{m} s_j'$ was estimated from data vector $\vec{X}$ based on the triangle inequality $\left\| \vec{X} \right\|_2 \leq N \sum_{j=1}^{m} \left\| \vec{a}_j \right\|_2 s_j'$ and assuming unit column L_2 norm of the mixing matrix $[A]$ it becomes:

$$\left\| \vec{X} \right\|_2 \leq N \tag{16}$$

At some iteration l output doublet $\left([W]^{(l)}, N^{(l)} \right)$ is generated as an output of the optimization algorithm in an attempt to reach possibly global minimum of the estimation error energy (9)/(15). For a given doublet $\left([W]^{(l)}, N^{(l)} \right)$ the MaxEnt algorithm (10)-(14) computes the most probable solution for the source vector $\vec{S}^{(l)} = N^{(l)} \vec{S}'^{(l)}$. This represents feedback for the optimization algorithm that computes a new value of the estimation error energy and generates a new doublet $\left([W]^{(l+1)}, N^{(l+1)} \right)$. After each iteration l is completed we get a triplet $\left([W]^{(l)}, N^{(l)}, \vec{S}'^{(l)} \right)$. Algorithm accepts as a final solution the triplet $\left([W]^{*}, N^{*}, \vec{S}'^{*} \right)$ for which the estimation error energy (9)/(15) reaches a possibly global minimum. Because in error energy function data vector $\vec{X}$ is fixed only triplet $\left([W]^{*}, N^{*}, \vec{S}'^{*} \right)$, corresponding with given data model (3), will give a global minimum of the error energy function. This is because for a given $\left([W]^{*}, N^{*} \right)$ the MaxEnt algorithm (10)-(14) converges toward $\vec{S}'^{*} = \frac{\vec{S}^{*}}{N^{*}}$.

Q.E.D.

3.0 Simulation Results

We illustrate deterministic BSS algorithm on the 2D case of Eq.(3). We assume the mixing matrix $[A]$ has unit L_2 column norm and parameterize it in term of two angles as follows:

$$\begin{bmatrix} x_x \\ x_y \end{bmatrix} = N \begin{bmatrix} \cos q & \cos j \\ \sin q & \sin j \end{bmatrix} \begin{bmatrix} s_x' \\ s_y' \end{bmatrix} \tag{17}$$

Vector diagram representation of (17) is shown on Figure 1 where column vectors are $\vec{a} = [\cos q \;\; \sin q]^T$ and $\vec{b} = [\cos j \;\; \sin j]^T$.

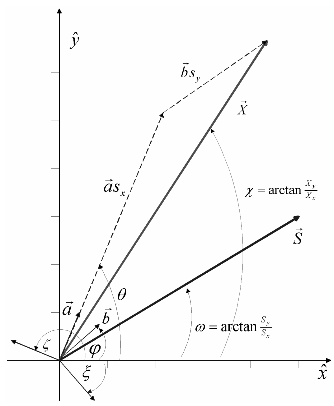

Figure 1. Vector diagram representation of the mixing model (3)/(17).

Due to the positivity constraints Eq.(2c) angles q and j are forced to lie in the first quadrant. Based on angle representation of the mixing matrix $[A]$ (17) the unmixing matrix $[W]$ can also be represented in term of two angles as follows:

$$[W] = \frac{1}{\sin(z - x)} \begin{bmatrix} \cos x & \sin x \\ \cos z & \sin z \end{bmatrix} = \begin{bmatrix} \vec{w}_1 \\ \vec{w}_2 \end{bmatrix} \tag{18}$$

where angles x, z are related to the angles q, j through:

$$x = j - \frac{p}{2}$$

$$z = q + \frac{p}{2} \quad (19)$$

If according to (17) we choose $q = 64^0, j = 45^0$, N=$8, s_1=5, s_2=3$ Eq.(17) becomes:

$$\begin{bmatrix} 4.3132 \\ 6.6153 \end{bmatrix} = 8x \begin{bmatrix} 0.4384 & 0.7071 \\ 0.8988 & 0.7071 \end{bmatrix} \begin{bmatrix} 5/8 \\ 3/8 \end{bmatrix} = 8x \begin{bmatrix} 0.4384 & 0.7071 \\ 0.8988 & 0.7071 \end{bmatrix} \begin{bmatrix} 0.6250 \\ 0.3750 \end{bmatrix} \quad (20)$$

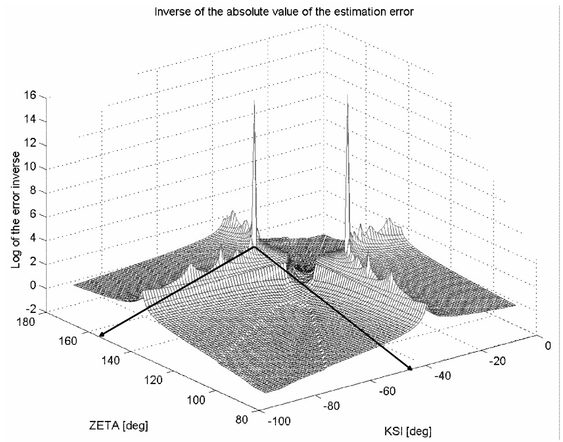

Figure 2. 2D plot of the log of inverse of (9)/(15) in the angle domain. Note two very sharp peaks that according to (19) give two solutions $x^1 = -45^0, z^1 = 154^0$ and $x^2 = -26^0, z^2 = 135^0$.

Figure 2 shows log of the inverse of the error energy function (9)/(15) as a function of angles x, z for given values of $N=8, s_x=5, s_y=3$. Note two very sharp peaks that according to (19) correspond two solutions $x^1 = -45^0, z^1 = 154^0$ and $x^2 = -26^0, z^2 = 135^0$. Two solutions are consequence of the non-unique representation of data vector (17) i.e.:

$$\vec{X} = N\left(\vec{a}s'_x + \vec{b}s'_y\right) = N\left(\vec{b}s'_y + \vec{a}s'_x\right) \quad (21)$$

From the single pixel point of view this permutation is not a problem. From the space variant imaging point of view it could create problems because related components of two different source vectors corresponding with two different pixels could be assigned on two different images. However, from vector diagram on Figure 1 we can see that, depending on convention, for angles q and j it applies the following:

$$j \le c \le q \quad (22a)$$

or:

$$q \le c \le j \quad (22b)$$

where c is angle defined from data vector $c = \tan^{-1}\left(x_y / x_x\right)$. Then adopting convention that angle j is always greater than angle c and that q angle is always less than angle c this type of permutation indeterminacy can be resolved for the space variant case. In order to illustrate convergence of the MaxEnt algorithm (10)-(14) we use the same angles as in (20) and $N=256, s_x=255, s_y=1$ that gives:

$$\vec{X} = \begin{bmatrix} 0.4384 & 0.7071 \\ 0.8988 & 0.7071 \end{bmatrix} 256 \begin{bmatrix} 0.9961 \\ 0.0039 \end{bmatrix} \quad (23)$$

For an 8-bit image (23) represents the most challenging case from the convergence point of view because the true solutions $s_1'^* = 0.9961$ and $s_2'^* = 0.0039$ are at the greatest distance from the maximum entropy initial points $s_x'^{(0)} = s_y'^{(0)} = 0.5$. The true values of the scaling factor N and un-mixing matrix [*W*] were reported to the algorithm (10)-(14) based on the assumption that they were found by some global optimization method such as exhaustive search in the parameter domain as a deterministic method, [20], or by using some stochastic global optimisation methods such as simulated annealing, [21-23]. Figure 3 illustrates convergence of the MaxEnt algorithm (10)-(14) for example given by (23).

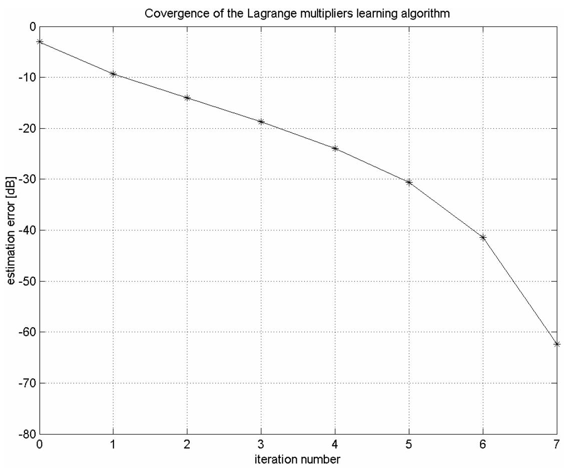

Figure 3. Convergence of the MaxEnt algorithm (10)-(15) for the case $s_1'^* = 0.9961$ given by example (23).

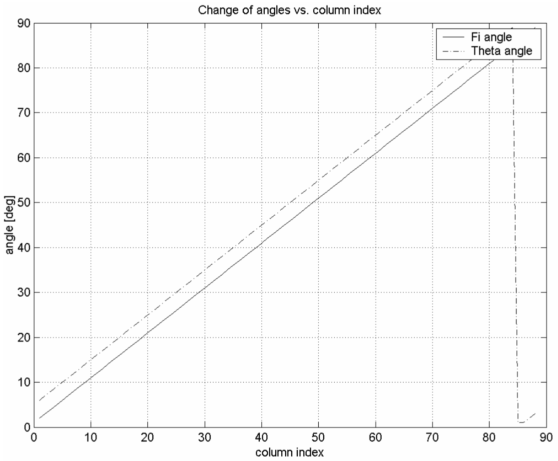

Figure 4. Change of the angles vs. column index . Solid line - j angle; dashed line - q angle.

We now mix two images by a mixing matrix that has been changed from pixel to pixel in order to simulate the space variant imaging problem. Angles q and j are changed column wise according to Figure 4 i.e. for every column index angles were changed for 1^0 and mutual distance between them was 6^0.

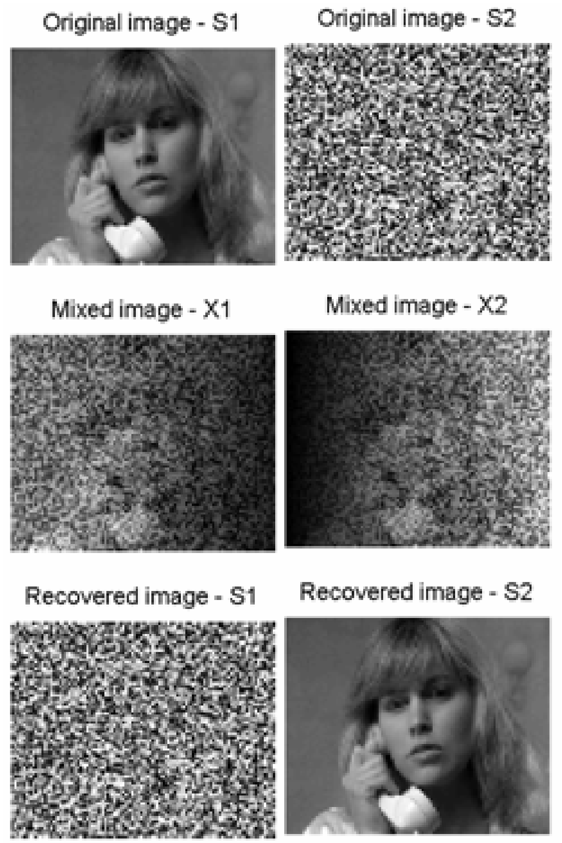

Figure 5. (Top row) source images; (middle row) space variant noise free mixture; (bottom row) error free recovery of the source images using deterministic BSS algorithm (10)-(16).

Figure 5 shows two source images (top row), two mixed images (middle row) and two separated images (bottom row) using deterministic BSS algorithm (10)-(16). Thanks to the fact that deterministic BSS algorithm solves the problem on the pixel-by-pixel basis the recovery was perfect although mixing matrix was space variant. We compare our result with two representative ICA methods that were applied on the same mixture shown on Figure 5. Separation results of the Bell-Sejnowski Infomax algorithm, [6], are shown on Figure 6 and results of the Fourth-Order cumulant based JADE algorithm, [8], are shown on Figure 7. Due to the space variant nature of the mixing matrix both algorithms fail to recover the original images.

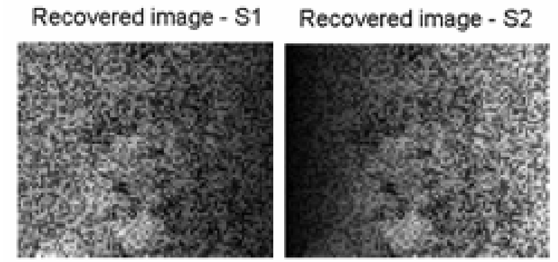

Figure 6. Recovered images using Infomax algorithm, [6] from the space variant mixture shown on Figure 5.

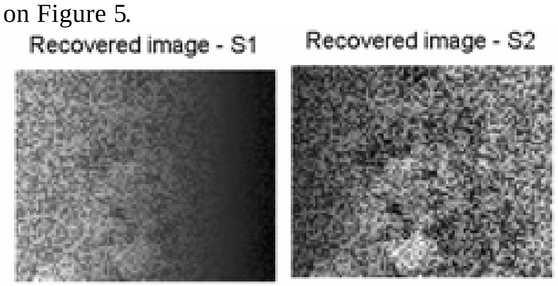

Figure 7. Recovered images using JADE algorithm, [8], from the space variant mixture shown on Figure 5.

4.0 Conclusion

Deterministic BSS algorithm capable of solving space variant imaging problem on the pixel-by-pixel basis has been presented. This is accomplished by two stage algorithm that combines minimization of the 2^{nd} law of thermodynamics based contrast function called Helmholtz free energy together with maximum entropy based algorithm that for each pixel computes the most probable value of the source vector under given macroscopic constraints defined by data vector. Performance of deterministic BSS algorithm has been demonstrated on the perfect recovery of images from the synthetic noise free space variant linear mixture of two images. Due to the space variant nature of the mixture ICA algorithms fail to recover unknown source images. The algorithm ability to recover source signals on the pixel-by-pixel basis is important for detection of small objects on the sub-pixel level what is the case in both remote sensing and thermography based early cancer detection.

5.0 References

1. H.H.Szu and I.Kopriva, "Comparison of the Lagrange Constrained Neural Network with Traditional ICA Methods," Proc. of the 2002 World Congress on Computational Intelligence-International Joint Conference on Neural Networks, pp. 466-471, Hawaii, USA, May 17-22, 2002.

2. H. H. Szu and C. Hsu, "Landsat spectral Unmixing à la superresolution of blind matrix inversion by constraint MaxEnt neural nets," *Proc. SPIE 3078*, pp.147-160, 1997.

3. H. Szu, "A Priori Maxent H(S) Independent Class Analysis (Ica) Vs. A Posteriori Maxent H(V) Ica," Proc. of the 3rd International Workshop on BSS and ICA, San Diego, USA, December 9-13, 2001, ed. Lee, Jung, Makeig, Sejnowski, pp. 80-89.

4. Szu, H. Thermodynamics energy for both supervised and unsupervised learning neural nets at a constant temperature. Int'l J. Neural Sys. **9**, 175-186 (1999).

5. S. Amari, A. Cihocki, H. H. Yang, "A new learning algorithm for blind signal separation," *Advances in Neural Information Processing Systems,* **8**, MIT Press, 757-763, 1996.

6. A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comp.* **7**, 1129-1159 , 1995.

7. A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, pp. 1483-1492, 1997.

8. J. F. Cardoso, A. Soulomniac, "Blind beamforming for non-Gaussian signals," Proc. IEE F, vol. 140, pp. 362-370, 1993.

9. M. Plubmley, "Conditions for Nonnegative Independent Component Analysis," IEEE Sig. Proc. Let., vol. 9, No. 6, June, 2002, pp.177-180.

10. M. Plumbley, "Adaptive Lateral Inhibition for Nonnegative ICA," ," Proc. of the 3rd International Workshop on BSS and ICA, San Diego, USA, December 9-13, 2001, ed. Lee, Jung, Makeig, Sejnowski, pp. 516-521.

11. J. W. Miskin, D. J. MacKay, "Application of Ensemble Learning ICA to Infra-Red Imaging," Proc. of the 2nd International Workshop on ICA and BSS, Helsinki, Finland, June, 2001, ed. P. Pajunen and J. Karhunen, pp. 399-404.

12. J. W. Miskin, D. J. MacKay, "Ensemble Learning for Blind Image Separation and Deconvolution," Advances in Independent Component Analysis, ed. M. Girolami, Springer, 2000, pp. 123-141.

13. D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, No. 21, pp.788-791, 1999.

14. L. Parra, C. Spence, P. Sajda, A. Ziehe, K-R. Müller, "Unmixing Hyperspectral Data," in *Advances in Neural Information Processing Systems (NIPS12),* S. A. Jolla, T. K. Leen and K-R. Müller (eds), MIT Press, 2000.

15. K. Huang, "Statistical Mechanics," J. Wiley, 1963.

16. E.T. Jaynes, "Information Theory and Statistical Mechanics," Phys. Rev., **106**, 620-630, 1957.

17. G. Deco and D. Obradovic, "Statistical Physics Theory of Supervised Learning and Generalization", Chapter 8 in "An Information-Theoretic Approach to Neural Computing", Springer, 1995, pp. 187-213.

18. J.C.Principe, and D.Xu, "Information-Theoretic Learning Using Reny's Quadratic Entropy," Proc. of the 1st International Workshop on ICA and BSS, Aussois, France, January 1999, ed. J.F. Cardoso, Ch. Jutten, Ph. Loubaton, pp. 407-412.

19. H. H. Szu, I. Kopriva, "Artificial Neural Networks for Noisy Image Super-resolution," Optics Communications, Vol. 198 (1-3) pp. 71-81, 2001.

20. R. Horst, H. Tuy, "Global Optimization - Deterministic Approaches," 3rd edition, Springer, 1996.

21. Szu H. and Hartley R., "Fast Simulated Annealing," *Physical Letters A*, **122**, No. 3, 1987, 157-162.

22. G. L. Bilbro, et.all, "Optimization by mean field annealing," *Advances in NIPS,* Morgan-Kaufmann, pp. 91-98, 1988.

23. G.L. Bilbro, "Fast Stochastic Global Optimization," *IEEE Tr. on Systems, Man and Cybernetics,* vol. 24, No. 4, pp. 684-689, 1994.

24. B. F. Jones, "A Reappraisal of the Use of Infrared Thermal Image Analysis in Medicine," *IEEE Trans. On Medical Imaging,* vol. 17, No. 6, pp. 1019-1027, 1998.