

Current progress in multiple image blind demixing algorithms

Harold Szu, ONR Code 313 *Arlington VA*
Szuh@onr.navy.mil

Abstract

Imagery edges occur naturally in human visual systems as a consequence of redundancy reduction towards “sparse & orthogonality feature maps,” which have been recently derived from the maximum entropy information-theoretical first principle of artificial neural networks. After a brief math review of such an Independent Component Analysis (ICA) or Blind Source Separation (BSS) of edge maps, we explore the de-mixing condition for more than two imagery objects recognizable by an intelligent pair of cameras with memory in a time-multiplex fashion. .

We assume that during the first look the system has enough time to have two cameras pointed, triangulated, for instance, focused on the girl_G standing in the shadow of a tree_T hidden with birds_B. Mathematically these objects are mixed in two cameras with 2x2-matrix $[A_0]$ such as: $(x_1, x_2)^T = [A_0](G, T + \epsilon B)^T$ without knowing birds hidden quietly in the tree indicated by a small order of magnitude $O(\epsilon)$. Suddenly the birds begin to sing. By design, two-camera system has two separate pointing & focusing & memory capabilities. Since as a quick response, only one of the dominate camera having the de-mixed girl image G can point away and focus at the birds hidden in the tree with a second look defined mathematically by $(y_1, \text{blurred } y_2)^T = [A_1](T, B)^T$. By design, the new image y_1 of birds & tree is acquired by the dominating camera with a new focus, whereas the other has a blurred y_2 and is therefore replenished by the previous de-mixed image $(T + \epsilon B)$, provided that each camera processes its own memory update rule. We consider channel communication application that we can efficiently mix four images using matrices $[A_0]$ and $[A_1]$ to send through two channels.

1. Principle of Orthogonality and Sparseness

The visual cortical feature detectors might be the end result of a Redundancy Reduction Process (RRP) [1, 2], in which the activation of each feature detector is supported to be as statistically independent from the others as possible. Such as ‘factorial code (of joint probability density)’ potentially involves independencies of all orders, but most studies [3, 4] have used only the second-order statistics required for de-correlating the outputs of a set of feature detectors. Field [5] has observed that the early-unsupervised learning algorithms are mainly based on the second-order statistics.

Current understanding is that the need of high order statistics such as the 4th order cumulant called Kurtosis may be captured completely by the information-Theoretical approach of maximum mutual information entropy underlying the Independent Component Analysis (ICA). From the knowledge representation point of view, the more efficient and robust representation, the better. Two principles are the keys to achieve efficient representation: *orthogonality* and *sparseness* in the hits frequency of feature detectors leading to unique identification. For example, an edge-map with one's over zero background is clearly sparse, local, and almost orthogonal. Biological evidence is the Hubel-Wiesel oriented edge-map [6] in the several octave scale of edges [7], similar to 2-D oriented Gabor Logon (information unit similar to a windowed FT or a WT without affine parameterization. The landmark accomplishment of ICA is to obtain, by unsupervised learning algorithm, the edge-map as image feature \vec{a} , shown by Helsinki researchers using fourth order statistics of \vec{u} -- Kurtosis $K(\vec{u})$, and derived from information-theoretical first principle of ICA by Bell & Sejnowski [8,9,10]. The key to a *good* representation is to strike the balance among conflicting requirements, such as efficiency and sensitivity, redundancy and robustness, noise tolerance and misclassification. Orthogonality is obviously the desirable characteristic in an efficient representation. For example, if there exists many features sets that are closely matched to particular signals, then obviously the features set where every features are orthogonal to one another would require less terms to represent those signals. Orthogonal representation can be attained through the use of transformation, such as Fourier Transforms (FT) and/or by statistical methods, such as Principal Component Analysis (PCA), which is shown here to be a special case of the Independent Component Analysis (ICA). It is important to point out that the methods mentioned above provide orthogonal feature set in the global sense: To further reduce the redundancy of the representation, more restricted local requirement must be imposed. This prompts the next refinement in the design of feature set: sparseness. This characteristic can be achieved using transformation such as Wavelet Transforms (WT), or Adaptive Wavelet Transforms (AWT), or by statistical means such as ICA.

From algorithmic point of view, by passing signal through FT we convert the data into frequency domain.

PCA maps the data into linear combination of “principal components”. Since the image signal is usually scanned from 2-D into 1-D vector, stationary image exhibits coincidental frequency. Further generalization of FT involves WT an ultimately AWT, while PCA is generalized into ICA. We found that continuous AWT and discrete ICA are roughly isomorphic. While AWT is useful for supervised feature extraction, ICA is for unsupervised. While AWT is continuous and invertible, ICA is not.

An unsupervised learning strategy of Artificial Neural Network (ANN) is to change the weight matrix $[W]$ of ANN to sieve or so-to-speak squeeze in parallel all the useful information from the time series of input vector $\mathbf{x}(t)$ until the output vector $\mathbf{u}(t) = [W] \mathbf{x}(t)$ contains no more useful information at maximum entropy $H(\mathbf{u})$ shown in Fig.1. In other words, hopefully all the good stuff is already kept in the memory weight matrix $[W]$. This strategy is different to the supervised learning, because one can not assume any specific output goal for the input, except the most natural one “garbage-output for any useful information-input” by the strict definition of no supervision. Such an intelligent learning may be described by the motto of neurocomputers-“data-in & garbage-out” as opposed to the usual motto-“garbage-in & garbage-out” in traditional & non-intelligent computers. This new paradigm is useful for solving a statistically matrix inversion $[A]^{-1}$ which mathematically underlies the Independent Component Analyses (ICA) as follows: $\mathbf{u}(t) = [W] \mathbf{x}(t) = [W] [A] \mathbf{s}(t)$, where t stands for both time of signal or the scanning order of pixels. If the learning of weight matrix $[W]$ can achieves the maximum entropy $H(\mathbf{u})$ of the output \mathbf{u} or the linear slope portion of the maximum entropy sigmoidal neuron output $H(\mathbf{y}) = H(\sigma(\mathbf{u})) = H(\mathbf{u})$ which implies that all nth moments of the ANN output components $\mathbf{u} = \{u_1, u_2\}$ of two sensor neurons are independent in terms of the normalized statistical histograms $\rho(\mathbf{u})$ defined as: $\langle u^n \rangle = \int u^n \rho(u) du$. Specifically, the whitening of the second moment of the output shows: $\langle \mathbf{u}(t) \mathbf{u}^T(t) \rangle = [W][A] \langle \mathbf{s}(t) \mathbf{s}^T(t) \rangle [A]^T [W]^T = [I]$ This Oja’s sphering is equivalent to $[W] = [A]^{-1}$ provide that statistical decorrelation of sources $\langle \mathbf{s}(t) \mathbf{s}^T(t) \rangle = [I]$ is true (if not, pre-whitening filter $[Wz] = \langle \mathbf{x}(t) \mathbf{x}(t)^T \rangle^{-1/2}$ is often used by Bell-Sejnowski *et. al.* The fourth cumulant, the Kurtosis $K(\mathbf{u})$, is often used by Helsinki’s group to seek the statistical matrix inversion. $K(\mathbf{u}) = \langle u^4 \rangle - 3 (\langle u^2 \rangle)^2$ in terms of a single weight vector update: $d\mathbf{w}/dt = dK/d\mathbf{w}$. The other weight vectors are found by the projection pursuits.

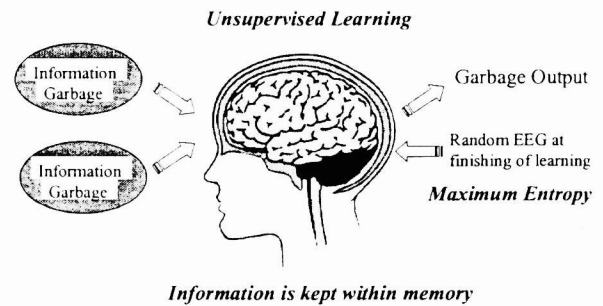


Fig.1 An unsupervised learning strategy of Artificial Neural Network (ANN)

Such an unsupervised learning methodology has been given in solving the statistically Blind Source Separation (BSS), as first introduced by J. Herault and C. Jutten, $\Delta[W] = g(\mathbf{x}) \tanh(\mathbf{u}^T)$ in terms of some ad hoc odd functions, in the first Snowbird ANN Conference in 1986. Both ICA subsequently coined by P. Como and BSS further elaborated by Herault & Jutten were appeared in Signal Processing Journal in 1991. Oja *et. al* [11] elaborated the nonlinear PCA learning, because neuron output $\mathbf{v} = \tanh(\mathbf{u}) = \mathbf{u} - 2/3 \mathbf{u}^3 + \dots$ has a similar Taylor expansion as $dK(\mathbf{u})/d\mathbf{w}$.

Recently, Bell and Sejnowski [8-10] at Salk Institute have derived from the first principle of unsupervised ICA the sparse edge map, a typical characteristics in early vision as discovered in cat’s eyes by the Hubel and Wiesel [5,12]. This may be considered one of major milestones from the information-theoretical viewpoint. The first principle of ICA may have several forms, e.g. absolute entropy versus mutual entropy, Neg-entropy-- the distance from the normality, Edgeworth versus Gram-Charlier expansions (of pdf in terms of moments) which are related to the maximum Shannon entropy $H(\mathbf{u})$. The essential portion related to the change of weight matrix is equivalent in achieving the redundancy reduction toward independent components which gives rise naturally to a sparse orthogonal edge map (unfortunately only at one wavelet resolution). The ANN unsupervised learning changes the ANN weight matrix to sieve or squeeze anything useful (higher order correlation information) from the input sequence until the outputs are left with (nothing but maximum entropy) redundant garbage or noise. This strategy is on the contrary to the supervised one because in a truly unsupervised learning we cannot assume any output goal but the garbage-output for information-input. Amari [13,14] *et. al* have further contributed to the speedup of learning by suggesting a natural gradient descent, rather than the original entropy gradient involving a non-local weight matrix inversion.

One of the remote sensing applications is to use the satellite, Landsat, to oversee the earth resource management, such as the deforestation of Amazon, the

major green lung of the Mother earth. Thus, to support the mandate of the World Bank loan and the taxation law of Brazil, Szu [15-19] has proposed in 1996 to supplement the infrequent on-site visits with real time surveillance technology based on maximum input entropy $H(s)$ of independent ground radiation sources, which turns out to be a different approach of the traditional maximum output entropy $H(u)$ to ICA. Thus, we shall briefly summarize $H(s)$ methodology herewith. First of all, such a remote sensing de-mixing technology can ascertain the percentage of tree being cut within the footprint of $30m^2$ of seven overlapped pixels based on seven spectral bands $x = \{x_1, x_2, \dots, x_7\}$ measured on the Landsat having six infrared bands ($1 \mu m$ to $10 \mu m$) and one visible band per each pixel. Although within the $30m^2$ resolution of single pixel the exact location of variety jungle canopy objects $s = \{s_1, s_2, \dots\}$ are unknown, but their mutual interaction must be in the short time scale to be statistically independent, i.e. having an instantaneous maximum entropy $H(s_1, s_2, \dots)$. The traditional methodology is based on the lookup table that each object s_i may be associated with the radiation distribution spectrum a_i which could be in practice calibrated in the laboratory. However, in a real time remote sensing observation the seasonal and diurnal variation $[A] = [a_1, a_2, \dots, a_7]$ are unknown except the fact that it satisfies mostly a linear law $x = [A] s$ per pixel. Szu and Hsu have demonstrated in 1997 a different ANN approach of ICA based on the assumption of "no extra input information (i.e. maximum entropy) except real positive photon sources s_i giving rise to real positive measurements x_i " We assume a constrained maximum source Shannon entropy $H(s) = -\sum_i s_i \log s_i$ carried by source neurons s_i with the help of dual space neurons called Lagrangian force λ_i neurons for the Hopfield-like energy function $\sum_{ij} \lambda_i ([A]_{ij} s_j - x_j)$ for the measurement data input and Helmholtz free energy $\lambda (\sum_i s_i - l)$ for the total percentage of unknown sources, which turns out to be a non-self-dual generalization of Hopfield ANN (if λ_i were s_i): $H(s) = -\sum_i s_i \log s_i - \sum_{ij} \lambda_i ([A]_{ij} s_j - x_j) - (\lambda - 1) (\sum_i s_i - l)$

(i) $\Delta[A] = dH/d[A] = \lambda_i s_j$ Hebbian learning between dual neurons in the inner product space

(ii) Setting the fixed point $dH/ds_j = -\log(s_j) - \sum_i \lambda_i [A]_{ij} - \lambda = 0$ and summing over s_j to evaluate Helmholtz free energy λ we derive the sigmoidal neuron output. $s_i = 1 / [1 + \sum_j \exp(\mu_j - \mu_i)]$; dual net sum $\mu_i = \sum_j \lambda_j [A]_{ij}$. The real world application is anticipated to have taken neighborhood pixel correlation into consideration.

Similarly, one of the limitation of Blind Source Separation by Unsupervised Learning is that the number of input cameras for the measurement data must be known ahead of time to be greater or equal to the contributing

sources, which are in principle unknown to begin with. On the other hand in a biological visual system, two-eye sensors are sufficient to separate more than two image sources. The difference shows that despite of the success the current unsupervised learning is till in the infancy and the power of smart memory of two eyes that can pay attention to different directions in order to check out several different image sources.

2. Sensors for 2 Sources at Constant $[A]$

Consider the case of 2 sensors for 2 sources, basic geometry can be shown in the following in equation (1), and the block diagram is plotted in Fig. 2.

$$\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = \begin{pmatrix} \cos(\theta_1) & \cos(\theta_2) \\ \sin(\theta_1) & \sin(\theta_2) \end{pmatrix} \begin{pmatrix} s_1(t) \\ s_2(t) \end{pmatrix} \quad (1)$$

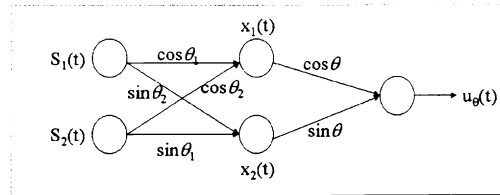


Fig. 2. Block diagram of two sensors for two sources

The measured data is written as in equation (2), and the learning weight, θ can be re-formatted as in equation (3). It can be shown that the output, $u_\theta(t)$ of Fig. 2 is theoretically invertible expressed in equation (4).

$$\bar{x}(t) = \sum_{i=1}^2 \bar{a}_i s_i(t) = [A] \bar{s}(t) \quad (2)$$

$$\begin{aligned} u_\theta(t) &= \begin{bmatrix} \cos \theta & \sin \theta \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \bar{w}^T \bar{x}(t) \\ &= \sum_{i=1}^2 (\cos \theta \cos \theta_i + \sin \theta \sin \theta_i) s_i(t) \\ &= \sum_{i=1}^2 \cos(\theta - \theta_i) s_i(t) \end{aligned} \quad (3)$$

Knowing θ_i , (not blind) theoretically invertible at weights $\theta = \theta_i$:

$$\begin{bmatrix} u_{\theta_1} \\ u_{\theta_2} \end{bmatrix} = \begin{bmatrix} 1 & \cos(\theta_1 - \theta_2) \\ \cos(\theta_2 - \theta_1) & 1 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \quad (4)$$

However, in practice note that $K(u_{\theta_1}) = K(s_1) + \cos(\theta_1 - \theta_2)K(s_2)$ is still a variable because of two sources involvement. Unknown θ_i (blind) if it happens at the killing angles, $\theta = \theta_1 + \pi/2$ then Eq. 3 gives $\cos(\pi/2) = 0$ and $u_\theta(t) = \sin(\theta_2 - \theta_1) s_2(t)$. This happens at extremum Kurtosis $K(u_\theta(t)) = \sin(\theta_2 - \theta_1)^4 K(s_2(t))$ such a blind de-mixing

employs the killing-principle of elimination shown as follows.

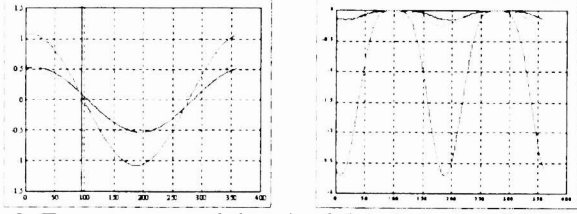


Fig. 3. Two sources and de-mixed data in space of the weight, $\theta \pm 90^\circ$ (left); Fig. 4. The analyses of Kurtosis (right), where one of the extremum is at $\theta = 99^\circ$.

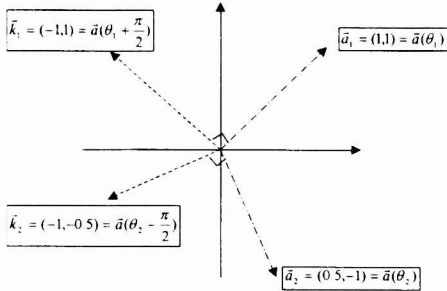


Fig. 5. Illustration of two mixed inputs and one de-mixed output by the killing weight vector $\mathbf{k}(\theta_1 - \pi/2)$

The simulation results of two sources and one mixed data is plotted in space of the weight, θ , shown in Fig. 3, and their Kurtosis is displayed in Fig. 4. The image results are illustrated in Fig. 5. Note that the extremum Kurtosis happens at source picture 1 (Yamakawa) vanishes and the next is source picture 2 (Szu) appears as shown in Fig. 5.

Maximum Output Entropy is shown equivalent to a batch mode operation of two eliminating (killing) vectors of Maximum Kurtosis as follows:

Given two arbitrary mixing vectors \bar{a}_1 at angle θ_1 and \bar{a}_2 at angle θ_2 in the 2 sensor 2-dimensional space.



$$[A] = \begin{pmatrix} 1 & 0.5 \\ 1 & -1 \end{pmatrix} = [\bar{a}_1 \quad \bar{a}_2] = [\bar{a}(\theta_1) \quad \bar{a}(\theta_2)]$$

$$[W][W_z] = [A]^{-1} = \begin{pmatrix} -1 & -0.5 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} \bar{k}_2 \\ \bar{k}_1 \end{pmatrix} = \begin{pmatrix} \bar{k}(\theta_2 - \frac{\pi}{2}) \\ \bar{k}(\theta_1 - \frac{\pi}{2}) \end{pmatrix} \quad (5)$$

$$[A][A]^{-1} = \begin{pmatrix} 1 & 0.5 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} -1 & -0.5 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} -1.5 & 0 \\ 0 & -1.5 \end{pmatrix} = -1.5[I]$$

If $(x,y) \rightarrow (y,x)$, in which θ is equal to $\theta + 3\pi/2$, then:

$$[A][A]^{-1} = \begin{pmatrix} 1 & 0.5 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} -0.5 & -1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 0 & -1.5 \\ -1.5 & 0 \end{pmatrix}$$

The inversely r.m.s. matrix $[W_z]$ is used to normalize both the gain by the signal energy itself, i.e. in terms of covariance matrix, $[W_z] = \langle \bar{x}\bar{x}^T \rangle^{-1/2}$, and to de-correlate two

input data \bar{x} , so that such an energy-gain pre-conditioned input data becomes de-correlated: $\bar{x}' = [W_z]\bar{x}$ such that $\langle \bar{x}' \bar{x}'^T \rangle = [I]$. This fact makes the unsupervised learning

matrix $[W]$ much easier to be found, because the de-correlation condition has reduced the unknown weight matrix to an orthogonal transform matrix $[W(\theta)][W(\theta)]^T = [I]$ which has only one angle parameter in 2D. This is because $\langle \mathbf{u}\mathbf{u}^T \rangle = [W][W_z] \langle \mathbf{x}\mathbf{x}^T \rangle [W_z]^T [W]^T = [W][W_z]^T [W]^T = [I]$ by means of $[W_z] \langle \mathbf{x}\mathbf{x}^T \rangle [W_z]^T = [I]$ implies $[W][W]^T = [I]$ to be orthogonal.

3. Demo. in ICA: 3 sources with 2 sensors

Now we will demonstrate why two sensors can NOT de-mix 3-source-basic geometry in Eq (5) and the block diagram is plotted in Fig. 6. From Eq (6), it can be observed that it is theoretically non-invertible since the matrix is singular, i. e. the determinant of the matrix vanishes and the set of three linear equations is dependent.

$$u_{\phi_s}(t) = \sum_{i=1}^3 \cos(\phi_s - \theta_i) s_i(t) \quad (6)$$

Kurtosis maximum happens at one of unknown mixed angles, say $\theta_1 \pm \pi/2$, where one set of sources $s_i(t)$ vanishes from the mixed data $\bar{x}(t)$ and then the remainder two sources are linearly DEPENDENT and can NOT be obtained by the Bell-Sejnowski statistical matrix inversion algorithm.

$$\begin{bmatrix} u_{\theta_1} \\ u_{\theta_2} \\ u_{\theta_3} \end{bmatrix} = \begin{bmatrix} 1 & \cos(\theta_1 - \theta_2) & \cos(\theta_1 - \theta_3) \\ \cos(\theta_2 - \theta_1) & 1 & \cos(\theta_2 - \theta_3) \\ \cos(\theta_3 - \theta_1) & \cos(\theta_3 - \theta_2) & 1 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} = [C] \bar{s} \quad (7a)$$

$$[C] = \begin{bmatrix} 1 & \cos(\theta_1 - \theta_2) & \cos(\theta_1 - \theta_3) \\ \cos(\theta_2 - \theta_1) & 1 & \cos(\theta_2 - \theta_3) \\ \cos(\theta_3 - \theta_1) & \cos(\theta_3 - \theta_2) & 1 \end{bmatrix} \quad (7b)$$

$$= \begin{bmatrix} \bar{w}_{\theta_1} \\ \bar{w}_{\theta_2} \\ \bar{w}_{\theta_3} \end{bmatrix} \begin{pmatrix} \bar{a}_1 & \bar{a}_2 & \bar{a}_3 \end{pmatrix} = \text{singular, } \because \det[C] = 0.$$

e.g. analytical verifiable when $\theta_i = i\theta_0$

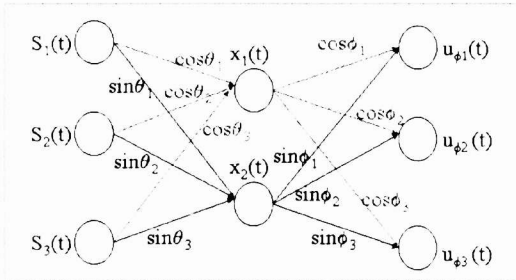


Fig. 6. Block diagram of two sensors and three sources

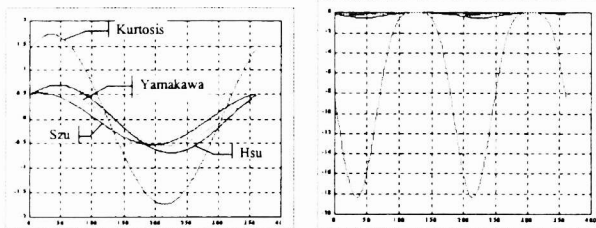


Fig. 7. Three sources and one mixed data in space of the weight, θ (left); Fig. 8. The analysis of Kurtosis (right)

Simulation results of three sources are plotted in space of the weight, θ in Fig. 7, and their Kurtosis is displayed in Fig. 8. The image results are illustrated in Fig. 9.



Fig. 9. Image illustration the de-mixing result that a set of three equations with two measurements.

4. Piecewise Constant Mixing Matrix

Since N sources must require at least N measurements according to the determinant of N equations for linear independence proven similar to Eq. (6), we explore the conditions to circumvent the restriction. We seek after the HVS for clues, since two eyes can obviously perceive several more objects. On the other hand, we are aware that the HVS is quite complicate. We have (i) left-eye-versus-right-eye dominance or vice versus, (ii) attentive and pre-attentive vision, (iii) have with unbalanced focus with imperfect eyesight, and (iv) the most complicate of all we have endowed with a powerful cortex-17 associative memory. How can we delineate the realistic cause and effect (i-iv), so that we can design a simplest possible ANN to support a pair of smarter sensors?

An instantaneous refreshing memory is utilized with two-sensor-ANN Eq(3) Fig. 10, which must be able to

mechanically point independently at the scenery and demonstrated to be capable to de-mix several objects in a time multiplex fashion.

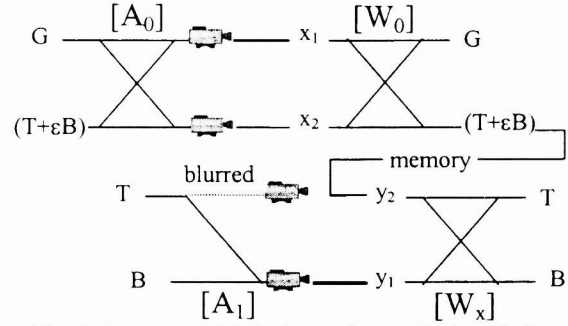


Fig. 10. 2 Cameras with independent pointing & focusing capability

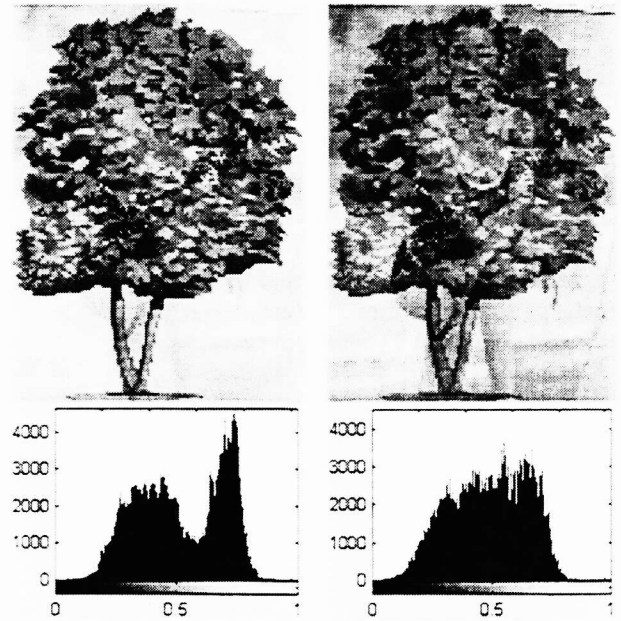


Fig. 11. 1st look (girl standing in the shadow of the tree hidden with birds)

We assume one sensor to be dominating over the other in the sense of giving the choice of 2 sensors, similar to HVS eyesight dominance (i) and (iii). We consider a scene that a beach girl, G , is standing in the shadow of a tree, T , hidden with two birds, B . These three objects are mixed in 2 cameras with 2×2 -matrix $[A_0]$ such as: $(x_1, x_2)^T = [A_0](G, T + \epsilon B)^T$ without the detection of birds hidden quietly in the tree indicated by a small order of magnitude $O(\epsilon)$. The first look pays attention at the person without being conscious of the birds hidden in the tree shown in Fig. 11. The unsupervised ANN has successfully found the two killing vectors to de-mixed the scene revealing a clear image of the

person from memory while the tree mixture is left on the other sub-dominate memory screen shown in Fig. 12. Suddenly, the birds begin to sing and that has prompted the smart sensor looking for birds hidden in the tree.

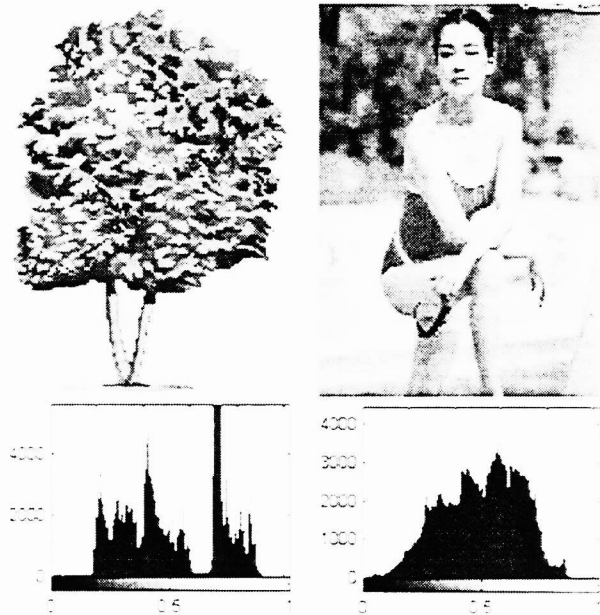


Fig. 12. De-mixed girl image and tree-bird image

Due to the mechanical difficulty of double focus requirement of simultaneously triangulation, only the dominant sensor is guided to quickly point, focus, and pick up the new input of bird-tree scene to replace the girl image. The sluggish focus response of the sub-dominate sensor acquire a not-quite-in-focus and blurred image of birds, which is smart enough to replace itself with the previously de-mixed memory of tree scene shown in Fig. 13. The new input together with early de-mixed impression y'_2 can be utilized to de-mix the birds and the tree. We show mathematically that two different mixture (x_1, x_2) due to matrix $[A_0]$ mixing sources (G, y'_2), which is a combination of T and B different from the second look y_2 defined as follows. A quick response calls for a second look (y_1 , blurred y_2) of matrix $[A_1]$ mixing sources (T B), of which the memorized $y'_2 = T + \epsilon B$ without the detection of birds hidden quietly in the tree indicated by a small order of magnitude $O(\epsilon)$ replaces the actual blurred y_2 . We can still yield a reasonable clear separation of the bird from the tree in Fig. 14. $x_1 = \cos(\theta_1)y'_2 + \cos(\theta_2)G$; $x_2 = \sin(\theta_1)y'_2 + \sin(\theta_2)G$; $y_1 = \cos(\theta_3)T + \cos(\theta_4)B$; blurred(y_2) = $\sin(\theta_3)T + \sin(\theta_4)B$ = memorized y'_2 , where $[A_0]$ of $\theta_1 = 36^\circ$ $\theta_2 = 60^\circ$; $[A_1]$ of $\theta_3 = 30^\circ$ $\theta_4 = 45^\circ$ but not yet point-in-focus & blurred y_2 is by design replaced by the refresh old memory $y'_2 = T + \epsilon B$.

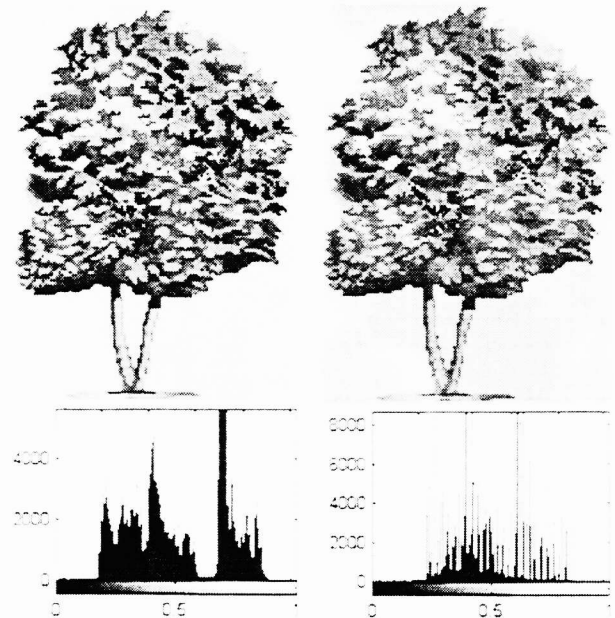


Fig. 13. 2nd look of the bird-tree scene and image remained in the memory

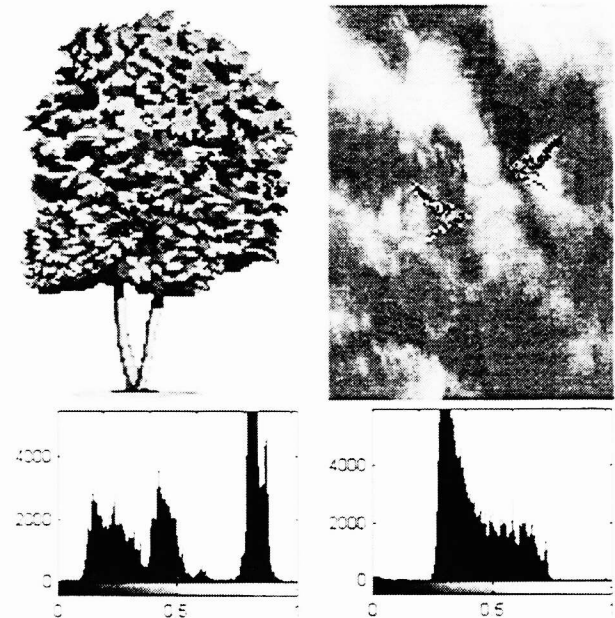


Fig. 14. Demix birds from a tree

Mathematically speaking, if we can add another sensor to Eq. (6), then $[A]_{3 \times 3} = [1 \ 0.5 \ -1; 1 \ -1 \ 0; 1 \ 0.5 \ 1]$ gives 3 mixture inputs shown in Fig. 16. A blind statistical matrix inversion gives the de-mixing results shown in Fig. 17. Compare with 2 sensors time multiplexing results, we find that the design of the smart sensor with memory gives us a reasonable well performance. Both cameras cannot point simultaneously and instantaneously at the scenery in a

perfect triangulation. The design of a smart pair of sensors allow only one dominates camera to focus at bird-tree scene for a new input and replenish the previously de-mixed memory of tree scene on the other sensor memory as one of the new input.

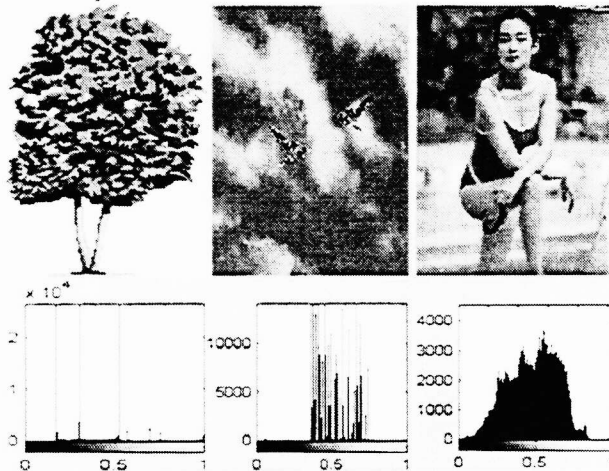


Fig. 15. The original images, T, B, G.



Fig. 16. 3 mixture inputs using 3 sensors



Fig. 17. 3 de-mix results by blind statistical matrix inversion

5. Summary

We have provided an elementary mathematical review of ICA limited by the measurement of two sensors. Then, we have explored several possible real world applications in terms of such artificial visual systems. Simulations of ICA are given to support designs of those real world applications based on two intelligent image/video sensors augmented

with independent pointing and focusing systems and some working memory.

Reference:

- [1] Atick J.J., "Could information theory provide an ecological theory of sensory processing?" v3, 213-251, *Network*, 92.
- [2] Barlow H.B., "Unsupervised learning", v 1, 295-311, *Neural Computation*, 89.
- [3] Linsker R., "Self-organization in perceptual network", vol. 21, pp. 105-117, *Computer*, 1988.
- [4] Miller K.D., "Correlation-based models of neural develop." 267-353, *Neuroscience & Conn. Theory*, Erlbaum, NJ, 88.
- [5] Field D. J., "What is the goal of sensory coding?", vol. 6, pp.599-601, *Neural Computation*, 1994
- [6] Hubel D.H., Wiesel T.N., "Receptive Fields & Func Arch. of Monkey Striate Cortex", *J. of Phy.* v195, 215-244, 68.
- [7] Hubel D.H., Wiesel T.N., "Uniformity of Monkey striate Cortex", *J. of Computation Neuron*, vol. 3 pp.61-72, 1992.
- [8] Bell A.J., Sejnowski T.J., "Edges are the Independent Components of Natural Scenes", v. 9, (*NIP 9*), 1996.
- [9] Bell A.J., Sejnowski T.J., "An Info-Maxm Approach to BSS and Blind Deconv", v7, 1129-1159, *Neural Comp*, 95
- [10] Bell A.J., Sejnowski T.J., "Learning high-order structure of a natural sound", v7, *Network*, Feb. 1996.
- [11] Oja E., Karhunen J., "Signal separation by nonlinear hebbian learning", *Proc of ICNN-95*, 417-421, 95.
- [12] Amari I., Cichocki A., Yang H., "A new learning algo. for BSS", *NIPS 1995*, v 8, 96, MIT press, Cambridge, Mass.
- [13] Cichocki A., Unbehauen R., Moszczynski L., Rummert E., "A new on-line adaptive learning algorithm for BSS", *proc. of ISANN*, pp. 406-411, Taiwan 1994.
- [14] H. Szu, C. Hsu, "Blind De-mixing with Unknown Sources", *ICNN*, v 4, pp 2518-2523, Houston, June, 1997.
- [15] Szu, Hsu, Yamakawa, "ICA using Wavelet Subband Orthogonality", *Proc. of SPIE* v3391, 180-193, Apr, 1998.
- [16] H. Szu, P. Cox, C. Hsu, "Statistical-mechanics demixing approach to select independent waves", in *Wavelet Appl. IV*, *Proc. of SPIE* 206-217vol. 3391, Orlando, April, 98'.
- [17] H. Szu, C. Hsu, Da-hong Xie "Continuous Speech Segmentation Determined by Blind Sources Separation", accepted in *Wavelet Application IV*, *Proc. of SPIE* pp 396-408, vol. 3391, Orlando, April, 1998.
- [18] H. Szu and C. Hsu, "Landsat Spectral Demixing à la Superresolution of Blind Matrix Inversion by Constraint MaxEnt Neural Nets," in *Wavelet Applications IV*, *Proc. SPIE*, 3078, 1997, 147-160.