

# A RECURRENT ICA APPROACH TO A NOVEL BSS CONVOLUTIVE NONLINEAR PROBLEM

Daniele Vigliano, Raffaele Parisi and Aurelio Uncini

*Dipartimento INFOCOM, Università di Roma "La Sapienza" – Italy Via Eudossiana, 18, 00184 Roma – Italy, [daniele.vigliano@poste.it](mailto:daniele.vigliano@poste.it); [parisi@infocom.uniroma1.it](mailto:parisi@infocom.uniroma1.it); [aurel@ieee.org](mailto:aurel@ieee.org);*

**Abstract.** This paper introduces a Recurrent Flexible ICA approach to a novel blind sources separation problem in convolutive nonlinear environment. The proposed algorithm performs the separation after the convolutive mixing of post nonlinear convolutive mixtures. The recurrent neural network produces the separation by minimizing the output mutual information. Experimental results are described to show the effectiveness of the proposed technique.

**Key Words:** Blind Source Separation, Flexible ICA, Spline Adaptive function, Mutual Information, Recurrent networks, IIR filters.

## 1. Introduction

The first studies about Independent Component Analysis aimed at resolving the famous cocktail party problem first in instantaneous, then in reverberant environments. A critical issue is that linear mixing models are too unrealistic and “poor” in a lot of real situations; recently it starts to grow the interest in non linear convolutive separation.

Important theoretical results in nonlinear instantaneous ICA framework are in [Hyvarinen et al., 1999]. Several papers explore Post Nonlinear mixing problem (PNL) in instantaneous [Taleb, 2002] and in convolutive environment [Milani et al., 2002][Zade et al., 2002] but only few of them (see [Taleb et al., 1999][Hyvarinen et al., 1999]) deal with the issue of existence and uniqueness of the solution. Recent advances in BSS of nonlinear mixing models have been reviewed in [Jutten et al., 2003]. If there are particular request of performance, separation quality or strictness of the mixing environment, became important the pdf matching of the signals. The so called Flexible ICA algorithm performs the adaptive estimation of parameters bound to the pdf of signals, this lead to a better pdf matching; it improves the quality of separation and allows a faster learning.

Actually recent studies try to improve the severity of mixing models moving from single-block nonlinear structures (convolutive or at least instantaneous) to multi-block structures. In [Solazzi et al., 2004] sources are

recovered from a PNL mixing followed by an instantaneous mixing (so called PNL-L mixture); in [Vigliano et al., 2005] the mixing environment is composed by a PNL mixing block followed by a convolutive one (PNL-C mixture) moreover the issue of existence and uniqueness of the solution has been explored. Other improvements of the severity of mixing environment have been presented in [Vigliano et al., 2004] which explores how behave a FIR network in separating sources from the convolutive mixing of a convolutive post-nonlinear mixture (CPNL-C mixture).

Recent works start performing the separation using multilayer neural networks (see [Woo et al., 2004] for details).

This paper explores the performance of a full recurrent network in separating sources from the CPNL-C mixing environment (already introduced in [Vigliano et al., WIRN2004]); CPNL-C at this moment is the most general convolutive nonlinear environment in literature.

## 2. Separation in nonlinear convolutive environment

This section introduces BSS problem in nonlinear environment. Be  $\mathbf{x}[n]$  the  $N$  vector of mixed accessible signals and  $\mathbf{s}[n]$  the vector of hidden independent sources The expression of the hidden mixing model in closed form is:  $\mathbf{x}[n] = \mathcal{F}\{\mathbf{s}[n], \dots, \mathbf{s}[n-L]\}$ , in which  $\mathcal{F}\{\cdot\}$  is a convolutive nonlinear distorting function. The solution of the BSS problem can be expressed as:  $\mathbf{y}[n] = \mathcal{G}\{\mathbf{s}[n]\} = \mathcal{G} \circ \mathcal{F}\{\mathbf{s}[n]\}$ . In instantaneous environment ICA recovers the original sources up to some trivial acceptable non-uniqueness: outputs can be scaled and delayed version of flipped inputs.

In the more general convolutive nonlinear case, the issue of separating mixtures with the only constraint of output independence and no other a priori assumption is affected by a strong non uniqueness [Jutten et al., 2003], [Vigliano et al., 2005]. Given independent inputs, several well known examples show the existence of maps that can produce independent outputs even with non diagonal Jacobian matrix. Independence constraint alone is not strong enough to recover original sources from generic nonlinear mixing environments [Taleb, 2002].

The most important issue for generic nonlinear problems is to ensure the presence of conditions granting, at least theoretically, the possibility to achieve the desired solution. In [Hyvarinen et al., 1999] authors proposed a constructive way (a Gram-Schmidt like method) to obtain solutions of the separation problem in an instantaneous nonlinear mixing environment; in order to grant the uniqueness of the solutions some constraints have been applied. In [Vigliano et al., 2005] authors introduced a theoretical proof of existence and uniqueness of the solution in a convolutive nonlinear

environment stricter than the Post Nonlinear one: the PNL-C; in that paper a general idea has been applied: adding “soft” constraints (a priori assumptions) to the problem can produce the uniqueness of the solution. Following such consideration, in this paper too, the a priori knowledge about the mixing model is exploited to design the recovery network: the so called “mirror” demixing model is used. Convolutional mixing environments add to the solution one more strong non-uniqueness: the filtering indeterminacies. Convolutional mixtures are separable but applying channel-by-channel filters to the independent recovered signals, outputs are still independent.

This indeterminacy may be unacceptable since it can strongly distort the sources. In any case after separation it is possible to equalize the outputs in order to obtain better results. According to these reasons filtering indeterminacy will no more considered in the rest of this paper.

### 3. The mixing-demixing architecture

This section explores the recovery of separated sources from nonlinear convolutional mixing; the a priori knowledge about the mixing model has been used to design the recovering network. The mixing environment is represented in figure 1.

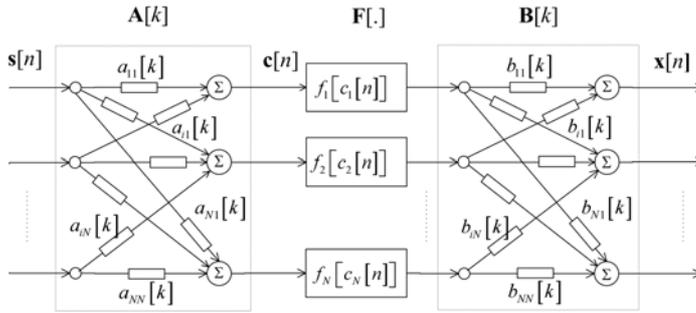


Figure 1. The block diagram of the convolutive nonlinear mixing model: the CPNL-C model.

In which  $\mathbf{A}[k]$  and  $\mathbf{B}[k]$  are  $N \times N$  FIR matrices with respectively  $L_a$  and  $L_b$  filter taps and  $\mathbf{F}[\mathbf{c}[n]] = [f_1[c_1[n]], f_2[c_2[n]], \dots, f_N[c_N[n]]]^T$  is the  $N \times 1$  vector of nonlinear distorting functions. The closed form for mixing model is:

$$\mathbf{x}[n] = \mathcal{F}[\mathbf{s}] = \mathbf{B}[n] * \mathbf{F}[\mathbf{A}[n] * \mathbf{s}[n]] \quad (1)$$

it is the so called CPNL-C mixing environment: convolutive mixing of a convolutive post-nonlinear mixture. Most of mixing environment already used in literature can be rewritten as particular case of the CPNL-C model. In order to grant the uniqueness of the solution, the recovering structure mirrors the mixing model but here convolutive blocks are realized by IIR architectures. The application of MIMO recurrent networks to the cocktail party problem in convolutive environment has been already exposed in several papers. The architecture presented in one of them [Choi et al., 1997] has been here adapted to this context and exploited to design  $\tilde{\mathbf{A}}$ ,  $\tilde{\mathbf{B}}$ . Figure 2 shows the recovering structure, with  $N=2$  for sake of simplicity.

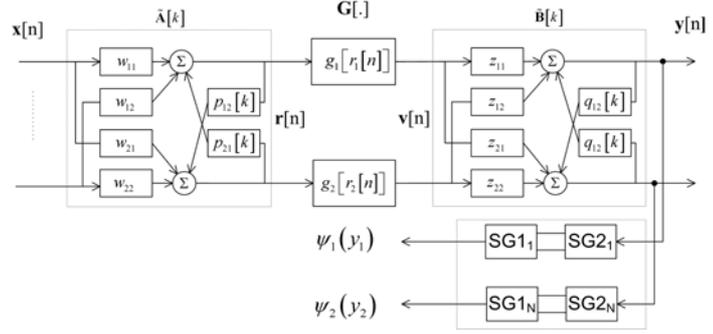


Figure 2. Recurrent network used for the nonlinear blind deconvolution and separation.

The expression of the  $i$ -th output channel is:

$$\begin{aligned}
 y_j[n] &= \sum_{h=1}^N z_{jh} v_h[n] + \sum_{h=1}^N \sum_{\substack{k=1 \\ h \neq j}}^{L_q} q_{jh}[k] v_h[n-k] \\
 v_h[n] &= g_h[r_h[n]] \\
 r_h[n] &= \sum_{j=1}^N w_{jh} x_j[n] + \sum_{\substack{j=1 \\ h \neq j}}^N \sum_{k=1}^{L_p} p_{jh}[k] x_j[n-k]
 \end{aligned} \tag{2}$$

In which  $\mathbf{G}[\cdot]$  is the  $N \times 1$  vector of nonlinear compensating functions, one for each channel,  $\mathbf{W}$  and  $\mathbf{Z}$  are  $N \times N$  matrices,  $\mathbf{Q}[k]$  and  $\mathbf{P}[k]$  are FIR filter with  $L_p$  and  $L_q$  filter taps; the networks used in this paper have  $q_{ii}[k] \equiv 0$ ,  $p_{jj}[k] \equiv 0 \quad \forall i, j, k$ .

Introducing the knowledge about the particular kind of mixing model is the key to avoid the strict non uniqueness of the solution; such assumption limits the weakness of the output independence condition reducing the

cardinality of all possible independent output solutions; with this constraint the problem of recovery the original sources is not ill posed any more.

#### 4. Blind demixing algorithm and Network model

This section explores the blind demixing algorithm, the adaptive recurrent network and the network that performs the nonlinear function estimation. The blind algorithm performs an adaptive learning of the network parameters  $\Phi$  on the base of the output independence estimation.

The learning is realized minimizing the Mutual Information  $I\{\Phi, \mathbf{y}\}$  between outputs, with a steepest descent algorithm:  $\Phi(k+1) = \Phi(k) - \eta_{\Phi} [\partial I\{\Phi, \mathbf{y}\} / \partial \Phi]$ . The choice of a gradient based minimization procedure lead to terms like:

$$\frac{\partial}{\partial \Phi} \log [p_{y_i}(y_i)] = \frac{\partial p_{y_i}(y_i) / \partial y_i}{p_{y_i}(y_i)} \frac{\partial y_i}{\partial \Phi} = \psi_i(y_i) \frac{\partial y_i}{\partial \Phi} \quad (3)$$

in which  $\psi_i(y_i)$  are the so called Score Functions (SF). In this paper, Spline Neurons are used to perform the on-line estimation of both Score Functions and nonlinear compensating functions (for a detail about Spline Neurons see [Solazzi et al., 2004][Uncini et al., 1998]). The most attractive property of Spline Neurons, as function estimators, is the local learning: for each learning step only the four control points nearest to the training input sample are considered; no matter how many control points the Spline curve has.

For direct estimation of SF has been performed a MSE approach (see [Taleb et al., 1999] for details) but learning rules result still blind:

$$\frac{\partial \varepsilon}{\partial \mathbf{Q}_i^{\psi_j}} = \left[ \frac{1}{4} \mathbf{T}_u \mathbf{M} \mathbf{T}_u \mathbf{M} \mathbf{Q}_i^{\psi_j} + \frac{1}{\Delta} \dot{\mathbf{T}}_u \mathbf{M} \right] \quad (4)$$

in which  $\mathbf{M}$  is a matrix of coefficients,  $\mathbf{T}$  is the vector local abscissa and  $\Delta$  is the distance between the abscissas of adjacent control points.

The network used to perform the separation is a cascade of blocks well described in literature and previously used to resolve more simple problems. Deriving the cost function  $I\{\Phi, \mathbf{y}\}$  with respect to the learning parameter  $\Phi$  results:

$$\begin{aligned} \frac{\partial I\{\Phi, \mathbf{y}[n]\}}{\partial \Phi} &\simeq \frac{\partial \mathfrak{S}\{\Phi, \mathbf{y}[n]\}}{\partial \Phi} = \\ &= -\frac{\partial}{\partial \Phi} \sum_{h=0}^M \left[ \log |\det \mathbf{Z}| + \log |\det \mathbf{W}| + \sum_{i=1}^N \log (g_i [v_i [n-h]]) \right. \\ &\quad \left. + \sum_{i=1}^N \log p_{y_i} (y_i [n-h]) \right] \end{aligned} \quad (5)$$

In (5) the expected value has been replaced by the instantaneous value. The learning rules for each parameter of the set  $\Phi = \{z_{ij}, w_{is}, p_{ml}[k], q_{nr}[h], \mathbf{Q}^g, \mathbf{Q}^\Psi\}$  are:

$$\partial \mathfrak{S} / \partial \mathbf{Z} = -\mathbf{Z}^{-T} - \Psi_y^T \mathbf{v}[n] \quad (6)$$

$$\partial \mathfrak{S} / \partial \mathbf{Q}_i^{g_j} = -\left[ \dot{\mathbf{T}}_u \mathbf{M} / \dot{\mathbf{T}}_u \mathbf{M} \mathbf{Q}_i^{g_j} + 1/2 \Psi_y (\mathbf{Z})_j \mathbf{T}_u \mathbf{M} \right] \quad (7)$$

$$\begin{aligned} \partial \mathfrak{S} / \partial \mathbf{W} &= -\mathbf{W}^{-T} - \left[ \ddot{g}_1(r_1) / \dot{g}_1(r_1) \cdots \ddot{g}_N(r_N) / \dot{g}_N(r_N) \right]^T \mathbf{x}[n] + \\ &\quad - \text{diag} \left[ \mathbf{Z}^T \Psi^y \right] \left[ \dot{g}_1(r_1) \cdots \dot{g}_N(r_N) \right]^T \mathbf{x}[n] \end{aligned} \quad (8)$$

$$\partial \mathfrak{S} / \partial q_{ij, i \neq j} [k] = -\Psi_{i^y} y_j [n-k] \quad (9)$$

$$\begin{aligned} \partial \mathfrak{S} / \partial p_{ij} [k] &= -\left[ \ddot{g}[n]_i(r_1) / \dot{g}_1(r_1) \cdots \ddot{g}_N(r_N) / \dot{g}_N(r_N) \right]^T \mathbf{r}[n-k] + \\ &\quad - \text{diag} \left[ \mathbf{Z}^T \Psi^y \right] \dot{\mathbf{v}}^T [n] \mathbf{r}[n-k] \end{aligned} \quad (10)$$

in which  $\mathbf{M}$  and  $\mathbf{T}$  have the same sense as in (4) and the operator  $\text{diag}[\mathbf{r}]$  transforms the vector  $\mathbf{r}$  in a diagonal squared matrix.

Recurrent networks have been used because they allow more compressed representation of models. If compared with a FIR architecture, the IIR network here proposed requires a reduced number of weights and then allows faster and more accurate learning.

## 5. Experimental results

This section collects separation results of two mixtures obtained applying the same mixing environment to different sets of independent sources. Although the algorithm should be able to perform the separation of  $N$ -channel mixtures, for proper visualization of results each set of sources is composed only by a couple of signals. The first set is composed by white signals (a gaussian noise and a uniform distributed signal), the second one by correlated signals (a male and a female voice speaking respectively “*Le donne i cavalier l’arme*” and “*Riperdo una seconda volta quegli esigui beni*”).

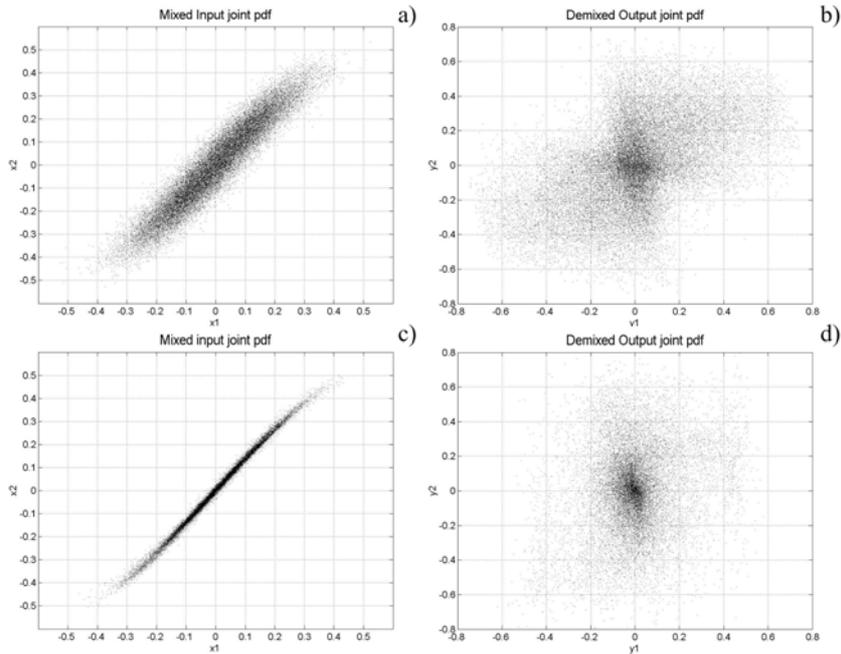


Figure 3. a-b) white signals: joint pdf of input mixture and joint pdf independent output; voices c-d) joint pdf of input mixture and joint pdf of independent output.

Figure 3 a-c) show the joint pdf of mixed signals and figure; figure 3 b-d) show the joint pdf of separated signals resulting after training, one note the typical plot of the joint pdf of separated signals. The demixing network has Spline Neurons ( $\mathbf{g}$  for distortion compensation and  $\Psi$  for Score Function estimation) with 53 control points and filters  $\mathbf{P}[k]$  and  $\mathbf{Q}[k]$  with 5 taps.

The applied nonlinear distortions for both input signals are:  $\mathbf{F}[f_1(p_1), f_2(p_2)] = [p_1 + 0.5p_1^3, 0.5p_2 + 0.8 * \tanh(5p_2)]$  (note that input signals are normalized in order to spread the nonlinear range of these

functions). Considering the notation of figure 1, the FIR matrices of the mixing environment are:  $\mathbf{A} = \begin{bmatrix} 0.8 + 0.4z^{-1} + 0.2z^{-2} & 0.6 + 0.3z^{-1} + 0.1z^{-2} \\ 0.5 + 0.3z^{-1} - 0.1z^{-2} & 0.8 - 0.4z^{-1} + 0.2z^{-2} \end{bmatrix}$ ,

and  $\mathbf{B} = \begin{bmatrix} 0.8 - 0.3z^{-1} + 0.06z^{-2} & 0.3 + 0.2z^{-1} - 0.06z^{-2} \\ -0.3 + 0.3z^{-1} + 0.11z^{-2} & 0.7 + 0.3z^{-1} - 0.1z^{-2} \end{bmatrix}$ .

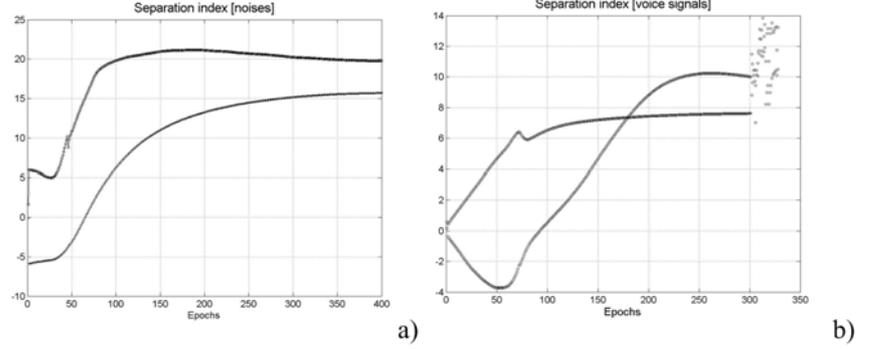


Figure 4. Separation index during the training; a) separation of white signals, b) separation of vocal signals

The so called “Separation Index”  $S_j$  (dB) introduced in [Shobben et al., 1999] gives a measure of how much channel  $j$ -th is separated from the others; here the Separation Index is evaluated for each channel.

$$S_j = 10 \log \left[ \frac{E \left\{ \left( y_{\sigma(j),j} \right)^2 \right\}}{E \left\{ \sum_{k \neq j} \left( y_{\sigma(j),k} \right)^2 \right\}} \right] \quad (11)$$

In (11),  $y_{i,j}$  is the  $i$ -th output signal when only the  $j$ -th input signal is present while  $\sigma(j)$  is the output channel corresponding to the  $j$ -input. The trend of this index (Figure 4, a-b) confirms the growing of separation during the training for both tests. For each of two tests, Fig. 3 and Fig. 4 together show how the algorithm is successful in performing the separation of the output signals. In [Vigliano et al., 2004] a similar mixing environment was approached with a FIR-based architecture; the separation performances obtained with the recurrent algorithm here exploited result improved moreover while the FIR matrix required 15-tap filters, the recurrent structure required 5-tap filters. This leads to a significant gain in terms of computational effort. Even if the separation well behaves in demixing voices, it reaches the best results with white signals; the reason of this behaviour lie in the construction of cost function (5).

## 6. Conclusion

This paper explores a novel recurrent ICA approach to the BSS problem in the CPNL-C mixing environment. The use of mirror model confirms the existence and the uniqueness of the solution to the CPNL-C separation problem. Results assure good separation and good nonlinear compensation. The use of the recurrent architecture allows a more compact representation of the recovering model and grants a more accurate demixing.

The recovering network performs the on line estimation of both nonlinear compensating functions and Score Functions by Spline Neurons leading to a better matching and producing a more accurate learning.

## References

- Jutten, C., Karhunen, J., (2003), "Advances in Nonlinear Blind Sources Separation", 4th International Symposium on ICA and BSS (ICA2003), April 2003, Nara, Japan.
- Taleb, A., (2002), "A Generic Framework for Blind Sources Separation in Structured Nonlinear Models", In IEEE Trans. on signal processing, vol. 50. no 8 August 2002.
- Taleb, A., Jutten, C., (1999), "Sources Separation in post nonlinear mixtures", In IEEE Trans. on signal processing, vol. 47. no 10 August 1999.
- Hyvarinen, A., Pajunen, P., (1999), "Nonlinear Independent Component Analysis: Existence and Uniqueness Results", Neural Networks 12(2): 429-439, 1999.
- Solazzi, M., Uncini, A., (2004), "Spline Neural Networks for Blind Separation of Post-Nonlinear-Linear Mixtures", In IEEE Trans. on Circuits and Systems I Fundamental Theory and Applications, Vol. 51 , No. 4, pp 817 – 829, April 2004.
- Uncini, A., Vecci, L., Piazza, F., (1998), "Learning and approximation capabilities of adaptive Spline activation function neural network", In NN, Vol. 11, no. 2, pag. 259-270 March 1998.
- Milani, F., Solazzi, M., Uncini, A., (2002), "Blind Source Separation of convolutive nonlinear mixtures by flexible spline nonlinear functions", Proc. of IEEE ICASSP'02, Orlando, USA, May, 2002.
- Zade, M. B., Jutten, C., Najeb, K., (2001), "Blind Separating, Convolutive Post nonlinear Mixture", ICA 2001 In Proc. of the 3rd Workshop on Independent Component Analysis and Signal Separation (ICA2001), San Diego (California, USA), 2001, pp. 138–143.
- Shobben, D., Torkkola, K., Smaragdakis, P., (1999), "Evaluation of blind signal separation methods", In Proc. of ICA and BSS, Aussois, France, January 11-15, 1999.
- Vigliano D., Parisi R. and Uncini A., (2004), "A flexible approach to a novel BSS convolutive nonlinear problem: preliminary result", Proc. of "Italian Workshop on Neural Networks (WIRN04)", Perugia, Springer-Verlag Ed., Sept 2004.
- Vigliano, D., Parisi R., Uncini, A., (2005), An Information Theoretic Approach to a Novel Nonlinear Independent Component Analysis Paradigm, In Press On Elsevier Signal Processing Special Issue on Information Theoretic (2005).
- Choi, S., Cichocki A., (1997), "Adaptive Blind Separation of speech signals: Cocktail party problem", ICSP97, Seoul, Korea, 26-28, pp. 617-622, August 1997.
- Woo W.L., Khor L.C (2004), "Blind restoration of nonlinearly mixed signals using multilayer polynomial neural network", Vision, Image and Signal Processing, IEE Proceedings- , Volume: 151 , Issue: 1 , 5 Feb. 2004.