

## A PARALLEL ARCHITECTURE USING DISCRETE WAVELET TRANSFORM FOR FAST ICA IMPLEMENTATION

Rong-bo Huang

Yiu-ming Cheung

Shi-ming Zhu

Department of Mathematics  
GuangDong Pharmaceutical College  
Guangzhou, P.R. China

Department of Computer Science  
Hong Kong Baptist University  
Hong Kong, SAR, P.R. China

Department of Mathematics  
Zhongshan University  
Guangzhou, P.R. China

### ABSTRACT

This paper utilizes a discrete wavelet transform to present a parallel architecture for independent component analysis (ICA), which is a hybrid system consisting of two sub-ICA processes. One process takes the high-frequency wavelet part of observations as its input, meanwhile the other process takes the low-frequency part. Their results are then merged to generate the final ICA results. Compared to the existing ICA algorithms, the proposed approach utilizes the full observation information, but the effective input length of the two parallel processes is halved. It therefore generally provides a new way for fast ICA implementation. In this paper, the experimental result has shown its success in extracting the independent components from a mixture.

### 1. INTRODUCTION

In the past decade, independent component analysis (ICA) has been extensively studied upon its attractive potential applications in medical signal processing [8], speech recognition [7, 12], signal and image processing [11], dimension reduction [6], and so forth. In literature, a classical definition of ICA is as follows: Suppose there are  $m$  independently and identically distributed non-Gaussian sources (also called *independent components interchangeably*) with at most one Gaussian source. All of them are statistically independent each other. The sources are sampled at discrete time  $t$ , denoted as  $y_t = [y_t^{(1)}, y_t^{(2)}, \dots, y_t^{(m)}]^T$ , and are instantaneously and linearly mixed by an unknown full-column matrix  $A$  with

$$x_t = Ay_t, 1 \leq t \leq N \quad (1)$$

where  $x_t = [x_t^{(1)}, x_t^{(2)}, \dots, x_t^{(d)}]^T$  is an observation at time step  $t$ . The ICA is to find out a de-mixing matrix  $W$  such

THE WORK DESCRIBED IN THIS PAPER WAS FULLY SUPPORTED BY A FACULTY RESEARCH GRANT OF HONG KONG BAPTIST UNIVERSITY WITH THE PROJECT CODE: FRG/01-02/11-24.

that

$$\hat{y}_t = Wx_t = WAy_t = P\Lambda y_t, \quad 1 \leq t \leq N \quad (2)$$

where  $\hat{y}_t = [\hat{y}_t^{(1)}, \hat{y}_t^{(2)}, \dots, \hat{y}_t^{(m)}]^T$  is an estimate of the sources  $y_t$ ,  $P$  is a permutation matrix, and  $\Lambda$  is a diagonal matrix. To achieve this goal, a variety of ICA algorithms have been developed, particularly from the information theoretic framework. For example, Information Maximization (INFOMAX) [1, 3] and Minimum Mutual Information (MMI) [2] both utilize a fixed nonlinearity function to perform ICA. Subsequently, they can separate either super-Gaussian or sub-Gaussian sources only, but not both. To circumvent this scenario, Xu et al. [4] presented Learned Parametric Mixture (LPM) based algorithm to carry out ICA. Many experiments have shown that the LPM can separate any combination of super-Gaussian and sub-Gaussian sources. However, the computation of LPM is quite tedious because many extra new parameters have to be learned together with  $W$ . To simplify the computation and speed up the performance convergent speed, Cheung and Xu [5] further developed an alternative approach, in which the nonlinearity separating function is estimated by a single polynomial term with an adjustable exponent. In general, the above-stated algorithms as well as other most algorithms are all to sequentially scan a long series of all observations so that their underlying information can be fully utilized in performing ICA. However, such a sequential scanning without invoking any parallel mechanism may limit the further improvement of ICA learning speed.

Recently, discrete wavelet transform (DWT) has been a popular tool in the areas of signal and image processing [9, 10]. One important characteristic of DWT is that not only a few large coefficients of its transform dominate the representations, but also it can decompose a signal at different scales and resolutions. Subsequently, each part can be processed in parallel. Hence, this paper utilizes DWT to present a parallel ICA (P-ICA) architecture, which is a hybrid system consisting of two sub-ICA processes. One process takes the high-frequency wavelet part of the observations as its input, meanwhile the other process takes the

low-frequency part. Their outputs are then merged to generate the final result. Compared to the existing ICA algorithms, the proposed approach utilizes the full observation information, but the effective input length of the two parallel processes is halved. It can therefore generally provide a new way for fast ICA implementation. In this paper, the experimental result has shown its success in extracting the independent components from a mixture.

## 2. THE PARALLEL ARCHITECTURE OF ICA USING DISCRETE WAVELET TRANSFORM

The parallel architecture of ICA using a discrete wavelet transform is shown in Figure 1. First, we utilize the DWT to decompose a series of observations, denoted as  $X$ , into two parts:  $HX$  and  $LX$ , respectively. The former is the high-frequency part of  $X$ , and the latter is the low-frequency one. We then extract the independent components from  $HX$  and  $LX$  via two ICA processes: ICA1 and ICA2, respectively. Eventually, their outputs are merged as the final ICA output via the reconstructor in Figure 1. In the following, we will present a theorem to elaborate the reason that P-ICA works.

**Theorem.** Given  $x = As$ , where  $x$  is a  $d$ -dimension vector,  $s$  is an  $m$  dimensional vector and  $A = (a_{ij})$  is  $d \times m$  matrix.  $\{\psi_{k,n} | k, n \in Z\}$  is an orthogonal wavelet basis, where we let  $Z$  be integral set for simplicity.  $\{c_{k,n}^j | k, n \in Z\}$  is the wavelet coefficient of  $s_j$  under the basis  $\{\psi_{k,n} | k, n \in Z\}$ . By denoting wavelet coefficients of vector  $s$  as  $\{C_{k,n} = [c_{k,n}^1, c_{k,n}^2, \dots, c_{k,n}^m]^T | k, n \in Z\}$ , the wavelet coefficients of vector  $x$  under the basis  $\{\psi_{k,n} | k, n \in Z\}$  are then given as follows:

$$\{b_{k,n} = Ac_{k,n} = [\sum_{j=1}^m a_{1j}c_{k,n}^j, \sum_{j=1}^m a_{2j}c_{k,n}^j, \dots, \sum_{j=1}^m a_{dj}c_{k,n}^j]\}. \quad (2)$$

**Proof:** Since

$$s_j = \sum_{k,n=-\infty}^{+\infty} c_{k,n} \psi_{k,n},$$

we then have

$$\begin{aligned} x_i &= \sum_j a_{ij} s_j \\ &= \sum_j a_{ij} \sum_{k,n=-\infty}^{+\infty} c_{k,n}^j \psi_{k,n} \\ &= \sum_{k,n=-\infty}^{+\infty} \sum_j a_{ij} c_{k,n}^j \psi_{k,n}, \end{aligned}$$

where  $i = 1, 2, \dots, d$ . Hence, the wavelet coefficient of  $x$  is:

$$\{b_{k,n} = Ac_{k,n} = [\sum_{j=1}^m a_{1j}c_{k,n}^j, \sum_{j=1}^m a_{2j}c_{k,n}^j, \dots, \sum_{j=1}^m a_{dj}c_{k,n}^j]\}$$

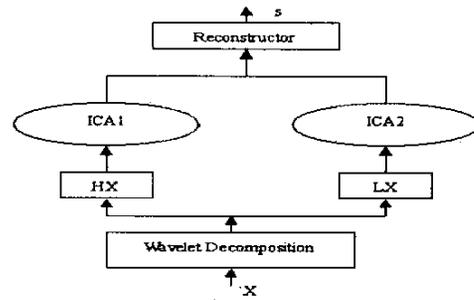


Figure 1: The architecture of P-ICA, which consists of two parallel sub-ICA processes: ICA1 and ICA2.

This theorem tells us that the mixing process from sources to observations is the same as the mixture from the wavelet coefficients of sources to those of observations. That is, we can perform ICA on the wavelet coefficients, instead of the original observations. Hereafter, we transform the observations to the wavelet representations, and whereby dividing them into two parts: high-frequency and low-frequency parts. After performing ICA on these two parts, we then merge their outputs to generate the final ICA results via the reconstructor as shown in the next section.

## 3. THE 2-LEVEL DECOMPOSING AND RECONSTRUCTING ALGORITHM OF WAVELET

Let  $\{V_k | k \in Z\}$  be a multi resolution analysis, and  $W_k$  in  $V_{k+1}$  be a complement space of  $V_k$ .  $L^2(R)$  can be therefore decomposed into direct sum of sub-space as follows:

$$L^2(R) = \sum_{k \in Z} W_k \quad (3)$$

There exists a unique decomposition for any  $f(x) \in L^2(R)$ :

$$f(x) = \sum_{k=-\infty}^{+\infty} g_k(x)$$

where  $g_k(x) \in W_k$ .

Here we consider the case of 2-level decomposing. Given scale function  $\phi(x)$  and wavelet function  $\psi(x)$ ,  $\{\phi(2^k x - j)\}, \{\psi(2^k x - j)\}$  are Rietz basis of space  $V_k$  and  $W_k$  respectively. The decomposition relation of  $\phi(x)$  and  $\psi(x)$  is given as follows.

$$\phi(2x - l) = \sum_{n=-\infty}^{+\infty} \{a_{l-2n}\phi(x - n) + b_{l-2n}\psi(x - n)\} \quad (4)$$

We obtain the decomposition series  $\{a_n\}, \{b_n\}$ . The decomposing algorithm is then given as follows:

$$c_n^1 = \sum_l a_{l-2n} c_l^2, \quad (5)$$

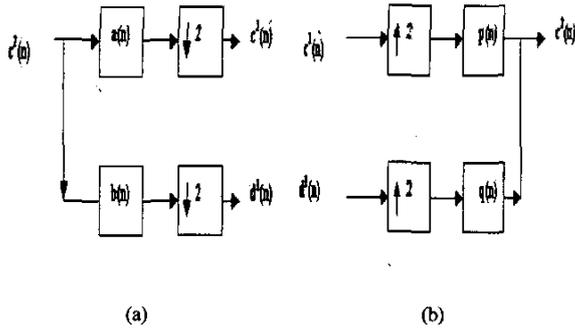


Figure 2: (a) The wavelet decomposing process, where  $\downarrow 2$  means to halve the sample size; (b) The wavelet reconstructing process, where  $\uparrow 2$  means to double sample size.

$$d_n^1 = \sum_t b_{t-2n} c_t^2. \quad (6)$$

The process of signal decomposition is shown in Figure 2(a). Subsequently, the original signal is decomposed into high-frequency and low-frequency parts.

The two-scale relation of  $\phi(x)$  and  $\psi(x)$  is given as follows.

$$\phi(x) = \sum_{n=-\infty}^{\infty} p_n \phi(2x - n), \quad (7)$$

$$\psi(x) = \sum_{n=-\infty}^{\infty} q_n \phi(2x - n). \quad (8)$$

The two-scale series  $\{p_n\}$ ,  $\{q_n\}$  are then obtained, whereby the reconstructor is to give:

$$c_n^2 = \sum_t (p_{n-2t} c_t^1 + q_{n-2t} d_t^2). \quad (9)$$

The reconstructing process is shown in Figure 2(b). Consequently, we can acquire  $s_j$ s via the inverse of DWT.

#### 4. THE ICA ALGORITHM

In the P-ICA, both of ICA1 and ICA2 can be realized by an existing ICA algorithm. Here, we adopt the Adaptive Polynomial Power Learning Estimation based ICA Algorithm (APPLE-ICA) [5] upon the fact that it can successfully separate any combination of sub-Gaussian and super-Gaussian sources with at most one Gaussian source. The APPLE-ICA algorithm utilizes a single polynomial term with its exponent parameter learned together with  $W$  towards maximizing the following cost function:

$$L(W) = Q(W, P)$$

$$\begin{aligned} &= \ln |\det(W)| + \sum_{j=1}^m \ln g_j(\hat{y}_t^{(j)}) \quad (10) \\ &= \ln |\det(W)| - \sum_{j=1}^m \frac{1}{p_j + 1} |\hat{y}_t^{(j)}|^{p_j+1} + C, \end{aligned}$$

where  $C$  is a constant term and  $P = \{p_1, p_2, \dots, p_m\}$ . The algorithm can be summarized as follows:

**Step 1.** Initialize  $W$  and a parameter

$$U = [u_1, u_2, \dots, u_m]^T$$

**Step 2.** Given an observed signal  $x_t$ , let:

$$\hat{y}_t = W x_t p_j = \lambda e_j^u, 1 \leq j \leq m$$

$$\phi_j(\hat{y}_t^{(j)}) = -\text{sign}(\hat{y}_t^{(j)}) |\hat{y}_t^{(j)}|_{p_j}^{p_j} \quad (11)$$

where  $\lambda$  is a positive constant.

**Step 3.** Update  $W$  and  $U$  by:

$$\begin{aligned} W^{new} &= W^{old} + \eta [I + \phi(\hat{y}_t) \hat{y}_t^T] W \\ u_j^{new} &= u_j^{old} + \eta \frac{\partial Q(W, P)}{\partial u_j} \end{aligned} \quad (12)$$

with

$$\frac{\partial Q(W, P)}{\partial u_j} = \frac{p_j |\hat{y}_t^{(j)}|^{p_j+1}}{p_j + 1} \left( \frac{1}{p_j + 1} - \ln |\hat{y}_t^{(j)}| \right), \quad (13)$$

The iterations of Step 2 and Step 3 are not stopped until both of  $W$  and  $P$  converge. For more details of the APPLE-ICA algorithm, interested readers can refer to the paper [5].

#### 5. EXPERIMENTAL RESULT

To investigate the performance of the proposed P-ICA, we used two sources: one is uniformly distributed, and the other is Gaussian distributed. The sample size  $N$  is set at 4,000. The observations were obtained by Eq.(1) with the mixing matrix:

$$A = \begin{pmatrix} 0.8 & 1.2 \\ -0.2 & 1.2 \\ 0.5 & 0.0 \end{pmatrix}.$$

After we scanned the learning data set around 150 times, the performance P-ICA was then converged. Figure 4 shows the final output of P-ICA, while sources and the mixtures are shown in Figure 3(a) and (b), respectively. It can be seen that the P-ICA has successfully recovered the wave-form of the original sources. That is, the original independent components have been successfully extracted from the mixture.

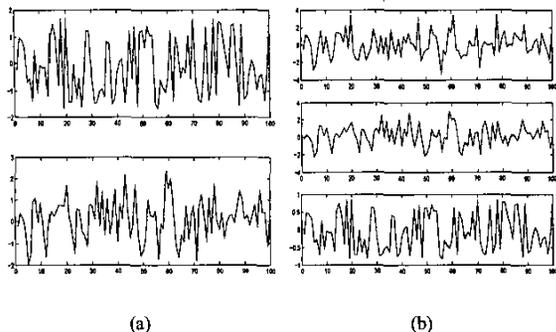


Figure 3: (a) The slide window of the source data, and (b) the slide window of the mixtures.

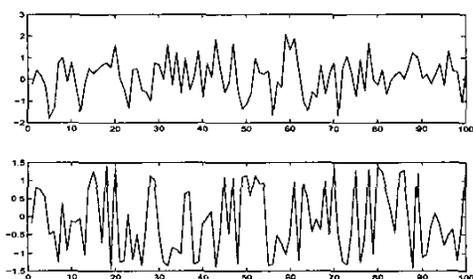


Figure 4: The slide window of the independent components recovered by P-ICA.

## 6. CONCLUDING REMARKS

We have presented a hybrid system consisting of two parallel ICA processes using a discrete wavelet transform. One ICA process extracts the independent components from the high-frequency part and the other from the low-frequency part. The reconstructor in P-ICA then merges their results, and finally generates the ICA outputs. Since this approach utilizes the full observation information, but the effective input length of the two parallel processes is halved. It therefore generally provides a new way for fast ICA implementation. In this paper, the experiment has demonstrated that P-ICA can successfully extract independent components from a mixture. In the future studies, we will further quantitatively investigate how much the ICA learning speed can be improved by such a parallel architecture.

## 7. REFERENCES

- [1] A.J. Bell and T.J. Sejnowski, "An Information-maximization Approaches to Blind Separation and Blind Deconvolution," *Neural Computation*, Vol. 7, pp. 1129-1159, 1995.
- [2] S.-I. Amari, A. Cichocki and H. Yang, "A New Learning Algorithm for Blind Source Separation," *Advances in Neural Information Processing Systems*, Vol. 8, MIT Press: Cambridge, MA, 1996.
- [3] J.P. Nadal and N.Parga, "Nonlinear Neurons in the Low-noise Limit: A Factorial Code Maximizes Information Transfer", *Network*, Vol. 5, pp. 565-581, 1994.
- [4] L. Xu, C.C. Cheung and S.I. Amari, "Learned Parametric Mixture Based ICA Algorithm," *Neurocomputing*, Vol. 22, pp. 69-80, 1998.
- [5] Y.M. Cheung and L. Xu, "A New Information-Theoretic Based ICA Algorithm for Blind Signal Separation," *International Journal of Computers and Applications*, Vol. 25, No. 2, pp. 106-110, 2003.
- [6] R.B. Huang, L.T. Law and Y.M. Cheung, "An Experimental Study: On Reducing RBF Input Dimension by ICA and PCA", *Proceedings of 1<sup>st</sup> International Conference on Machine Learning and Cybernetics 2002*, Vol. 4, pp. 1941-1946, 2002.
- [7] T.W. Lee, A.J. Bell and R. Ortlmeister, "Blind Source Separation of Real-world Signals", *Proceedings of 1997 IEEE International Conference on Neural Networks (IEEE-INNS IJCNN97)*, pp. 2129-2135, 1997.
- [8] T.W. Lee, M.S. Lewicki and T.J. Sejnowski, "Unsupervised Classification with Non-Gaussian Mixture Models Using ICA," *Advances in Neural Information Processing Systems 11*, MIT Press, 1998.
- [9] R. Nowak, "Optimal Signal Estimation Using Cross-validation," *IEEE signal Proceedings Letter*, Vol. 4, pp. 23-25, 1997.
- [10] S. Mallat, "A Wavelet Tour of Signal Processing," New York:Academic, 1998.
- [11] A. Hyvärinen and E. Oja, "Independent Component Analysis: Algorithms and Applications," *Neural Networks*, Vol. 13(4-5), pp. 411-430, 2000.
- [12] G.J. Jang, T.W. Lee and Y.H. Oh, "Learning Statistically Efficient Features for Speaker Recognition," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, Salt Lake City, Utah, May, 2001.