

조합형 Fixed Point 알고리즘을 이용한 독립성분분석의 성능개선

민성재*, 박용수, 조용현
대구가톨릭대학교 공과대학 컴퓨터정보통신공학부
e-mail:yhcho@cuth.cataegu.ac.kr

Performance Improvement of Independent Component Analysis Using Hybrid Fixed Point Algorithm

Seong-Jae Min*, Yong-Soo Park, Yong-Hyun Cho
School of Computer and Information Comm. Eng.,
Catholic Univ. of Daegu

요약

본 연구에서는 Newton 기법과 모멘트에 기초를 둔 fixed point 알고리즘의 신경망 기반 독립성분분석기법을 제안하였다. 여기서 Newton 기법은 함수의 접선에 기초를 둔 해를 구하는 방법으로 역혼합행렬의 빠른 경신을 위함이고, 모멘트는 접선을 구하는 과정에서 함수의 기울기변화 계산으로 발생하는 발진을 줄여 좀 더 빠른 학습을 위함이다. 제안된 기법을 256x256 픽셀(pixel)의 12개 지문영상으로부터 임의의 혼합행렬에 따라 발생하는 영상들을 각각 대상으로 시뮬레이션 한 결과, 기존의 fixed point 알고리즘에 의한 결과보다 우수한 분리성능과 빠른 학습속도가 있음을 확인하였다.

1. 서론

일반적으로 통계학, 신호처리, 그리고 신경망 등에서 주된 연구는 적당한 변환에 의해서 알맞은 데이터의 표현을 찾는 것이다^[1,2]. 그것은 데이터의 해석을 용이하게 하는 것으로 데이터의 후속적 해석을 위해서이다. 이때 선형변환을 이용하면 문제를 계산적으로나 개념적으로 좀 더 간단하게 만들 수 있으며, 결과의 해석도 용이하게 할 수 있다. 여러 가지 방법과 원리들이 적당한 선형변환을 찾기 위해서 개발되었으나, 그 중에서도 주요성분분석(principal component analysis : PCA)^[1-3]이 가장 널리 이용되고 있다. 이는 데이터의 차원을 감소시키는 선형변환 기법으로 공분산(covariance)을 이용하는 2차원 방법으로 가우스(Gaussian) 분포의 입력 데이터에 대해서는 적당한 것이다. 하지만 일반적으로 통신이나 신호 및 영상 처리 등에서의 입력 데이터에는 비가우스 분포로 이러한 경우에도 효과적으로 적용될 수 있는 선형변환 방법의 연구가 요구된다.

최근 PCA의 제약을 해결하기 위한 독립성분분석(independent component analysis : ICA)이 제안되었으며, 이는 데이터의 통계적 성질을 고려한 선형변환 기법으로 데이터의 분포정보를 이용하는 다차원 방법이다^[4-6]. 하지만 알려진 ICA 알고리즘들에서는 데이터의 통계적 독립성을 판정하기 위해 확률밀도함수와 역행렬을 구한다든지 고차원의 cumulant나 moment를 구하는 등 복잡한 계산이 요구된다.

따라서 복잡한 계산에 따른 많은 계산시간이 요구되는 수치적 기법에 의한 ICA의 대안으로 간단한 신경망을 이용하여 해결하려는 연구가 활발하게 진행되고 있다. 발표된 신경망을 이용한 ICA 방법으로는 fixed point(FP) 알고리즘, recursive least square 알고리즘, 확률적 기울기를 이용하는 Bell & Sejnowski(BS) 알고리즘, 자연기울기(natural gradient)를 이용한 natural gradient 알고리즘 등이 제안되었다^[5,6]. 하지만 이들 방법들에서도 신경망을 이용하나 신경망 자체가 가지는 학습에 따른 제약들은 그대로 내포하고 있어 그에 따른 연구가 절실하다.

본 연구에서는 빠른 실행시간을 가지는 새로운 학습알고리즘의 독립성분분석 기법을 제안한다. 제안된 기법은 엔트로피 최적화로부터 유도된 기법으로 그 계산을 간략화 하기 위해 Newton 기법을 이용하며, 기울기의 변화에 따른 경신알고리즘에서 발진을 줄여 더욱 빠른 학습시간을 가지도록 모멘트를 이용한 FP 알고리즘이다. 제안된 기법을 256x256픽셀(pixel)의 12개 지문영상으로부터 임의의 혼합행렬에 따라 발생하는 영상들을 각각 대상으로 시뮬레이션하여 성능을 확인하였으며, 기존의 fixed point 알고리즘에 의한 결과와 비교 검토하였다.

2. 조합형 알고리즘의 독립성분분석기법

ICA는 m개의 입력신호 s로부터 선형적으로 혼합된 n개의 신호 x가 알려져 있을 때, 혼합된 신호로부터

역으로 m개의 독립인 입력신호를 찾는 기법이다. 하지만 입력신호들을 혼합하는데 이용된 혼합행렬 A 는 알려져 있지 않으며, 혼합과정에서 잡음 n 이 추가 될 수도 있다. 이때 혼합신호와 입력신호와의 관계는

$$x = As + n = \sum_{i=1}^m s(i)a(i) + n \quad (1)$$

로 정의된다. 여기서 n 은 보통 입력신호와 구별되지 않기 때문에 생략할 수도 있으며, $A=[a(1), a(2), \dots, a(m)]$ 으로 $a(i)$ 는 ICA의 basis vector이다. 결국 ICA는 알려진 혼합신호로부터 혼합행렬의 역행렬 A^{-1} 을 찾는 기법이다. 혼합행렬 A 와 역혼합행렬 W 에 대하여 상세히 살펴보면 다음 그림 1과 같은 구성도로 나타낼 수 있다. 여기서 $x = As$ 이고, $y = Wx$ 이다. 이때 ICA는 출력신호 y 의 각 요소들이 통계적으로 독립성을 유지해야 하며 직교성의 제약조건은 포함하지 않는다. 그림에서 보면 ICA는 혼합행렬과 일치하는 역혼합행렬을 찾는 과정에서 출력신호가 독립성을 가지도록 하는 기법이다. 따라서 ICA는 알려진 혼합신호 x 로부터 출력신호 y 를 찾는 기법으로 궁극적으로는 역혼합행렬 W 을 찾아서 원신호 s 의 근사값을 알아내는 것이다. 이때 영평균과 whitening된 혼합신호를 이용하면, 신호의 1차와 2차의 통계성에 따른 상관성이 이미 고려되어 좀 더 효과적으로 ICA 기법을 수행할 수 있다.

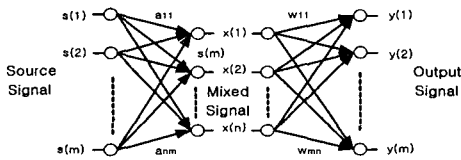


그림 1. 혼합행렬과 역혼합행렬의 상세 설명도

2.1 신호의 영평균과 Whitening

신호의 영평균은 데이터의 1차적 통계성을 고려한 정규화로 학습알고리즘을 간략하게 하는 기법이다^[4-6]. 이는 신호 데이터 벡터 x 에서 평균값 \bar{x} 를 뺀 차를 구함으로써 간단하게 구할 수 있다. 즉, 영평균된 새로운 데이터 벡터 x 는

$$x = x - \bar{x} \quad (2)$$

이다.

한편, 데이터의 2차적 통계성을 고려한 상호간의 의존성을 줄이기 위한 기법으로 whitening이 이용되고 있다^[3]. whitening은 데이터 벡터 x 의 공분산행렬이 단위행렬값을 갖도록 함으로써 구할 수 있다. 즉, $E\{xx^T\}=I$ 가 되도록 한다. whitening된 벡터의 성분들은 상호간의 상관성이 줄어들어 비상관성이 더 강한 독립조건을 위해 요구되는 전처리 기법이다. 이러한 whitening의 기법으로 PCA가 널리 사용된다. 일반적인 PCA whitening 행렬 V 는

$$V = D^{(-1/2)}E^T \quad (3)$$

이다. 여기서 D 는 고유값 λ 로 구성되는 대각행렬이며, E 는 고유벡터 c 로 구성되는 직교행렬이다. PCA whitening이 가지는 장점은 이미 잘 알려진 표준 소프트웨어를 이용할 수 있으며, 또 다른 장점으로는

원 신호, 즉 독립성분의 개수를 추정하는 편리한 수단이 되는 것이다.

2.2 조합형 Fixed Point 알고리즘

독립성분분석을 위한 Fixed Point 알고리즘은 신경망이 가지는 병렬성과 분산성, 그리고 더 작은 메모리 요구 등의 제약을 해결하기 위해 제안된 알고리즘이다^[5]. FP 알고리즘은 엔트로피 최적화 방법으로부터 유도된 기법으로 그 수렴속도가 적어도 2차 함수정도로 지금까지 알려진 기법 중 가장 빠른 학습속도를 가진다. 이는 신호 내에 포함된 상호정보를 고려한 것으로 그 척도는 독립성의 거리이다. 결국 상호정보를 최소화함으로써 ICA의 해를 구하는 기법이다.

상호정보는 랜덤변수 상호간의 독립성의 정보이론적 척도로 ICA 변환을 찾기 위한 판별로서 자연스럽게 이용될 수 있다. 일반적으로 변수들 상호간의 정보를 표현하는 상호정보는 비가우스성의 척도로 이용되는 negentropy로 정의된다. n 개의 랜덤변수 벡터 $y = (y_1, \dots, y_n)^T$ 를 위한 negentropy $J(y)$ 는

$$J(y) = H(y_{gauss}) - H(y) \quad (4)$$

이다. 여기서 $H(y)$ 는 랜덤변수 y 의 엔트로피(entropy)이며, y_{gauss} 는 가우스 랜덤벡터이다. 이때 negentropy $J(y)$ 를 이용한 상호무관한 n 개 랜덤변수 $y_i(i=1, \dots, n)$ 사이의 상호정보 $I(y_1, y_2, \dots, y_n)$ 는

$$I(y_1, y_2, \dots, y_n) = J(y) - \sum_{i=1}^n J(y_i) \quad (5)$$

로 정의된다. 결국 ICA에서 독립성을 최대화하기 위해 식 (5)로 정의되는 상호정보 $I(y_1, y_2, \dots, y_n)$ 를 최소화하는 것은 negentropy를 최대화하는 것이다. negentropy의 최대화는 식 (5)에서 우측 2번째 항의 최대화로 가능하다. 이를 위한 근사식으로 최대 엔트로피 원리를 이용하면

$$J(y_i) \approx c[E\{G(y_i)\} - E\{G(\nu)\}]^2 \quad (6)$$

와 같이 정의된다. 여기서 G 와 c 는 각각 비2차(non-quadratic)함수와 상수이고, ν 는 평균이 0이고 분산이 1인 표준화된 가우스 변수이며, 랜덤변수 y_i 도 평균이 0이고 분산이 1인 변수로 가정한다. negentropy의 근사화를 위한 식 (6)은 ICA 변환을 추정하는 목적함수로 이용될 수 있다. 따라서 ICA에서 독립성분의 상호간 무관성을 고려한 하나의 독립성분 $y_i = W^T x$ 를 찾기 위한 목적함수 $J_G(W)$ 는 식 (6)에 따라

$$J_G(W) = [E\{G(W^T x)\} - E\{G(\nu)\}]^2 \quad (7)$$

로 표현된다. 이를 전체 n 개의 독립성분 모두에 대해서 확장하면

$$\begin{aligned} & \text{maximize } \sum_{i=1}^n J_G(w_i) \quad \text{w.r.t. } w_i, i=1,2,\dots,n \\ & \text{subject to } E\{(w_k^T x)(w_j^T x)\} = \delta_{jk} \end{aligned} \quad (8)$$

로 된다. 여기서 w_i 는 역혼합행렬 W 의 행벡터 중

하나이며, ICA 변환은 $s = Wx$ 로 얻어진다. 결국 식 (7)이나 (8)로 정의된 ICA를 위한 목적함수를 최대화하기 위한 방법이 요구된다. 이러한 방법으로 가장 간단한 알고리즘은 기울기하강에 기초를 둔 기법들이다. 하지만 이들 알고리즘에서는 학습율의 설정에 따라서는 수렴이 느리거나 학습이 이루어지지 않는 제약들이 있다. 이러한 기울기하강에 기초한 알고리즘들의 제약들을 해결하기 위한 대안으로 FP 알고리즘이 제안되었다^[5].

FP 알고리즘은 알고리즘의 한 단계에 많은 수의 데이터들이 이용되는 batch mode로 계산을 수행하며, 신경망으로 수행될 수 있는 알고리즘이다. 데이터 벡터 x 의 상관행렬 $E(xx^T)=I$ 로 whitening되어 있다고 가정할 때, 하나의 독립성분을 위한 식 (7)에서 FP 알고리즘을 유도하면 다음과 같다. $J_G(W)$ 의 최대화는 반드시 $E(G(W^T x))$ 의 최적에서 얻어진다. 이때 Kuhn-Tucker 조건으로부터, $E((W^T x)^2) = \|W\|^2 = 1$ 조건하에서 $E(G(W^T x))$ 의 최적은 다음의 값

$$E(xg(W^T x)) - \beta W = 0 \quad (9)$$

로 얻어진다. 여기서 $g(\cdot)$ 는 비2차 함수 G 의 미분이고, $\beta = E(W_0^T xg(W_0^T x))$ 인 상수이며 W_0 는 W 의 최적값이다. 따라서 방정식 (9)를 풀기 위해 빠른 수렴속도를 가지는 Newton의 기법을 도입한다. 이를 위해 식의 좌변항을 $F(W)$ 로 나타내고, 그것의 Jacobian matrix $JF(W)$ 를 구하면 다음과 같다. 즉,

$$JF(W) = E(xx^T g'(W^T x)) - \beta I \quad (10)$$

이다. 한편 데이터 벡터 x 가 이미 whitening되어 있어 Newton의 기법에 따른 해를 구하는데 이용되는 $JF(W)$ 의 역행렬을 간단하게 계산하기 위해 식 (10)의 우변 첫째 항을 다음과 같이 근사화할 수 있다. 즉,

$$E(xx^T g'(W^T x)) = E(xx^T)E(g'(W^T x)) = E(g'(W^T x))I \quad (11)$$

이다. 식 (11)을 식 (10)에 대입하면 Jacobian matrix는 대각선 행렬이 되어 쉽게 역행렬을 계산할 수 있다. 따라서 β 를 W_0 대신에 W 의 현재값으로 나타내면 근사화된 반복기법의 역혼합행렬 W 를 구하는 Newton 기법은

$$\begin{aligned} W^* &= W - [E(xg(W^T x)) - \beta W] / [E(g'(W^T x)) - \beta], \\ W^* &= W^* / \|W^*\| \end{aligned} \quad (12)$$

와 같이 정의된다. 여기서 W^* 는 W 의 새롭게 경신된 값이고, $\beta = E(W^T xg(W^T x))$ 이다. 결국 식 (12)는 Newton 기법에 기초를 둔 ICA를 위한 FP 알고리즘이다. 또한 식 (12)의 첫째식 양쪽에 $\beta - E(g'(W^T x))$ 를 곱해 구해지는 더욱 간단한 Newton 기법의 FP 알고리즘은

$$W^* = E(xg(W^T x)) - E(g'(W^T x))W, W^* = W^* / \|W^*\| \quad (13)$$

으로 된다.

한편 식 (12)의 우측 두번째 분모항이나 식 (13)에서 우측 두번째 항의 계산을 보면 Newton 기법에서 접선을 구하기 위한 기울기로서 함수 $g(W^T x)$ 의 1차 미분이 요구된다. 하지만 함수의 미분에 따라 W 가

최적해로 수렴할 때 발진이 일어날 수도 있다. 따라서 수렴에 따른 발진을 막아 줄 수만 있다면 좀 더 빠르게 최적해로 수렴시킬 수 있을 것이다. 이를 위해 본 논문에서는 W 를 계산할 때 이전 변화의 속성을 반영하는 모멘트를 추가하였다. 이렇게 하면 Newton 기법이 가지는 빠른 수렴속성을 그대로 살리면서도 최적해로의 수렴에서 발생하는 발진을 억제시켜 더욱 빠른 속도의 독립성분분석이 가능하다. 제안된 모멘트를 이용한 역혼합행렬 W 의 최종계산식은 식 (13)을 변형한

$$\begin{aligned} W^* &= E(xg(W^T x)) - E(g'(W^T x))W + \alpha(W^* - W), \\ W^* &= W^* / \|W^*\| \end{aligned} \quad (14)$$

로 정의된다. 여기서 α 는 모멘트이다.

제안된 조합형 FP 알고리즘의 역혼합행렬 W 을 경신하는 과정을 요약하면 다음과 같다.

단계 1 : 초기 역혼합행렬 $W(0)$ 를 임의의 값 rand()로 설정한다.

단계 2 : $W = W(0) / \|W(0)\|$ 의 값을 계산한다.

단계 3 : $W_{old} = 0$ 로 설정한다.

단계 4 : $\|W - W_{old}\| > \epsilon$ 이고 $\|W + W_{old}\| > \epsilon$ 이면 다음 식에 따라 W 를 계산하고, 그렇지 않으면 과정을 종료한다.

(1) $W_{old} = W$ 를 계산한다.

(2) $W^* = E(xg(W^T x)) - E(g'(W^T x))W + \alpha(W^* - W)$ 를 계산한다.

(3) $W^* = W^* / \|W^*\|$ 를 계산한다.

위의 과정에서 벡터 x 는 whitening된 신호벡터이며, ϵ 은 1보다 적은 임의의 값으로 실험에서는 0.0001로 하였다. 특히 경신의 식에서 $g(\cdot)$ 는 비선형 함수이며 일반적으로 $(\cdot)^3$ 과 $\tanh(\cdot)$ 의 함수값을 가진다. 본 연구에서는 상대적으로 이들 함수에 대한 분리성능이 유사함을 확인하였고, 여기서는 $(\cdot)^3$ 함수를 적용한 결과를 보인다.

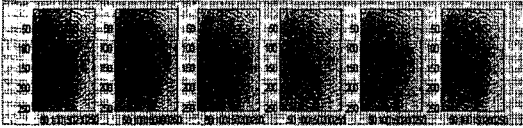
3. 시뮬레이션 결과 및 분석

제안된 ICA 알고리즘의 타당성과 성능을 평가하기 위해서 지문영상으로부터 임의의 혼합행렬에 따라 발생되는 혼합영상들을 각각 대상으로 시뮬레이션 하였다. 시뮬레이션에서는 256×256 픽셀을 가지는 12개의 독립된 지문영상을 대상으로 임의의 값을 가진 혼합행렬로부터 생성된 영상들을 대상으로 제안된 알고리즘을 수행하였다. 복원 성능을 평가하기 위해 정합도 A 를 정의하여 원 영상과 복원된 영상을 비교하였다. 이때 정합도는 $A = 100 - \frac{1}{256 \times 256} \left[\sum_{i=1}^{256} \sum_{j=1}^{256} |s_{ij} - y_{ij}| \right]$ 이다. 또한 시뮬레이션은 펜티엄 IV-1.5G 컴퓨터 상에서 Matlab 5.1로 실행하였다.

그림 2는 12개의 원 지문영상과 임의의 혼합행렬로부터 생성된 혼합영상 중에서 일부인 각각 6개의 영상들만을 나타낸 것이다. 그림에서 보면 혼합영상 각각에서는 원 지문영상을 발견할 수 없다. 실제에서는 혼합영상들은 알려져 있으나 원 영상들은 미지의 영상들이다. 시뮬레이션에서는 혼합된 영상들만 입력으로 하여 ICA를 수행한다.



(a) 원지문영상



(b) 혼합된 지문영상

그림 2. 256×256 픽셀의 12개 중 6개의 독립된 지문영상

그림 3은 그림 2(b)의 혼합영상들을 대상으로 제안된 기법의 FP 알고리즘에서 모멘트 α 값의 변화에 따른 학습반복수를 나타낸 것이다. 그림에서의 결과는 각각의 모멘트에 대하여 10번씩 학습과정을 반복하여 그 평균을 나타낸 것이다. 여기에서 $\alpha=0$ 은 기존의 FP 알고리즘의 경우이며, 0.001에서 0.05까지는 상대적으로 제안된 기법이 기존의 기법보다 빠른 학습이 됨을 알 수 있다. 12개의 혼합된 지문영상의 분리를 위한 학습반복수가 기존의 알고리즘에서는 약 93.6, $\alpha=0.015$ 에서 제안된 알고리즘은 약 89.4로 더 빠른 학습속도를 보였다. 이때 소요되는 CPU시간을 측정하면 기존의 FP 알고리즘은 93.6회 반복에 약 96.6초, 제안된 알고리즘은 89.4회 반복에 85.3초가 소요되어 학습시간에서도 약 1.4배정도 빠름을 알 수 있다.

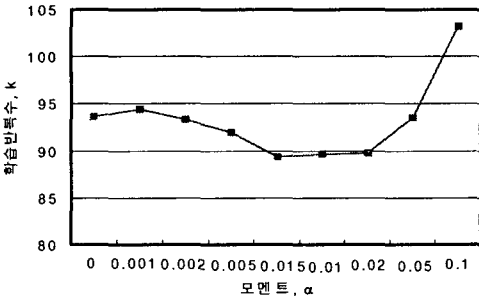


그림 3. 모멘트 α 의 변화에 따른 학습반복수

한편 표 1은 12개의 혼합된 지문영상으로부터 분리된 12개의 독립된 지문영상과 원 지문영상과의 정합도를 나타낸 것이다. 기존의 알고리즘에서는 지문영상 5가 12개 영상 중에서 99.92의 정합도로 가장 우수한 분리성능을 보이며, 제안된 알고리즘에서는 동일한 성능을 보인다. 한편 지문영상 6에서 제안된 알고리즘은 99.95의 정합도를 보여 가장 우수한 분리성능을 보이나 상대적으로 기존의 알고리즘에서는 99.51의 정합도로 그 분리성능이 떨어짐을 알 수 있다. 12개의 임의로 혼합된 지문영상들에서 분리된 원 지문영상들과의 전체 평균 정합도는 각각 기존의 알고리즘이 99.57이고 제안된 알고리즘은 99.85로 더욱 우수한 분리성능을 확인할 수 있다.

표 1. 원 지문영상과 복원된 지문영상의 정합도

분리된 지문	기존의 FP 알고리즘	제안된 FP 알고리즘
영상1	99.53	99.87
영상2	99.53	99.83
영상3	99.48	99.87
영상4	99.45	99.84
영상5	99.92	99.92
영상6	99.51	99.95
영상7	99.89	99.89
영상8	99.46	99.79
영상9	99.43	99.83
영상10	99.51	99.84
영상11	99.76	99.76
영상12	99.45	99.8
평균	99.57	99.85

따라서 모멘트를 이용한 제안된 조합형 FP 알고리즘은 기존의 FP 알고리즘에 비해 빠른 학습속도와 우수한 독립성분 분석 성능이 있음을 확인할 수 있다. 한편, 본 연구에서는 500개의 값을 가지는 4개의 인공적인 신호를 대상으로 실험한 결과 유사한 특성을 얻었으며, 두 개의 대상을 비교하면 문제의 규모가 클수록 그 학습속도와 분석성능의 차이는 증가함도 확인하였다. 결국 제안된 기법은 문제의 규모가 커질수록 우수한 특성이 있음도 알 수 있다.

4. 결 론

본 연구에서는 Newton 기법과 모멘트에 기초를 둔 fixed point 알고리즘의 신경망 기반 독립성분분석기법을 제안하였다. 여기서 Newton 기법은 함수의 접선에 기초를 둔 해를 구하는 방법으로 역혼합행렬의 빠른 경신을 위함이고, 모멘트는 접선을 구하는 과정에서 함수의 기울기변화 계산으로 발생하는 발진을 줄여 좀 더 빠른 학습을 위함이다.

12개의 256×256 지문영상들로부터 임의의 혼합행렬에 따라 발생하는 혼합영상들을 각각 대상으로 실험한 결과, 기존의 fixed point 알고리즘보다 우수한 추출성능과 빠른 추출시간이 있음을 확인하였다.

참 고 문 헌

- [1] K.I. Diamantaras and S. Y. Kung, Principal Component Neural Networks : Theory and Applications, Adaptive and learning Systems for Signal Processing, Communications, and Control, John Wiley & Sons, Inc., 1996
- [2] S. Haykin, Neural Networks : A Comprehensive Foundation, Prentice-Hall, 2ed, London, 1999
- [3] J. Karhunen and J. Joutsensalo, "Generation of Principal Component Analysis, Optimization Problems, and Neural Networks," Neural Networks, Vol. 8, No. 4, pp 549-562, 1995
- [4] P. Comon, "Independent Component Analysis A New Concept?," Signal Processing, vol.36, No.3, pp.287-314, Apr.1994
- [5] T.W. Lee, Independent Component Analysis : Theory and Applications, Kluwer Academic Pub., Boston, 1998
- [6] J. Karhunen, "Neural Approaches to Independent Component Analysis and Source Separation", "4th European Symp., Artificial Neural Network, ESANN96, Burges, Belgium, pp. 249-266, Apr. 1996