

# 할선법과 모멘트의 고정점 알고리즘 독립성분분석에 의한 영상분리

조용현, 민성재<sup>o</sup>  
대구가톨릭대학교 컴퓨터정보통신공학부  
e-mail:yhcho@cataegu.ac.kr

## Image Separation Using Fixed-Point ICA of Secant Method and Moment

Yong-Hyun Cho, Seong-Jae Min<sup>o</sup>  
School of Computer and Information Communications. Eng.,  
Catholic Univ. of Daegu

### 요 약

본 연구에서는 할선법과 모멘트의 고정점 알고리즘 독립성분분석을 이용하여 혼합영상을 분리하는 기법을 제안하였다. 여기서 할선법은 독립성분 상호간의 정보를 최소화하기 위한 목적함수의 최적화 과정에서 요구되는 1차 미분에 따른 계산을 간략화하기 위함이고, 모멘트는 최적화 과정에서 발생하는 발진을 억제하여 보다 빠른 학습을 위함이다. 제안된 기법을 512x512의 픽셀을 가지는 8개의 영상을 대상으로 임의의 혼합행렬에 따라 발생하는 혼합영상의 분리에 적용한 결과, 할선법에 기초한 고정점 알고리즘보다 우수한 분리율과 빠른 분리속도가 있음을 확인하였다.

### 1. 서론

독립성분분석(independent component analysis : ICA)은 데이터의 차원을 감소시키는 선형변환 기법 중 가장 널리 이용되는 주요성분분석(principal component analysis : PCA)의 일반화 기법이다. PCA는 공분산(covariance)을 이용하는 가우스(Gaussian) 분포의 입력 데이터에 적당한 2차원 다차원 방법이다<sup>1-6)</sup>. 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<sup>3-6)</sup>. 발표된 신경망을 이용한 ICA 방법으로는 고정점(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<sup>4,5)</sup>. 하지만 이들 방법들에서도 신경망을 이용하나 신경망 자체가 가지는 학습에 따른 제약들은 그대로 내포하고 있어 그에 따른 연구가 절실하다.

본 연구에서는 기존의 신경망이 가지는 학습시간과 학습률 사이의 절충관계를 효과적으로 해결할 수 있는 조합형 알고리즘을 제시한다. 이를 위해

negentropy를 최대화하는 과정에서 기울기 이용법의 그 계산을 간략화 하기 위해 할선법(Secant method)<sup>7)</sup>을 이용하며, 또한 기울기의 변화에 따른 경신알고리즘에서 발진을 줄여 더욱 빠른 학습능력의 독립성분분석을 위해 모멘트를 이용하였다. 제안된 기법을 512x512의 픽셀을 가지는 8개의 영상을 대상으로 임의의 혼합행렬에 따라 발생하는 혼합영상의 분리에 적용하여 그 타당성과 성능을 확인하고자 한다.

### 2. 할선법과 모멘트의 고정점 알고리즘 독립성분분석

ICA는  $m$ 개의 입력신호  $s$ 로부터 선형적으로 혼합된  $n$ 개의 신호  $x$ 가 알려져 있을 때, 혼합된 신호로부터 역으로  $m$ 개의 독립인 입력신호를 찾는 기법이다. 이때 혼합신호와 입력신호와의 관계는 일반적으로 다음 식(1)과 같다.

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

로 표현된다. 여기서 잡음  $n$ 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. 이상의 관계를 요약하면 그림 1과 같다. 그림에서 보면 혼합기(mixer)의 출력신호인 혼합

신호는 알 수 있으나 혼합기의 특성인  $\mathbf{A}$ 와 입력신호인 원 신호  $\mathbf{s}$ 는 알려져 있지 않다. 결과적으로 입력신호와 출력신호가 일치하도록 하는 즉,  $\mathbf{y} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{A}\mathbf{s}$ 에서  $\mathbf{W}=\mathbf{A}^{-1}$ 의 관계를 만족도록 하는 것이다.

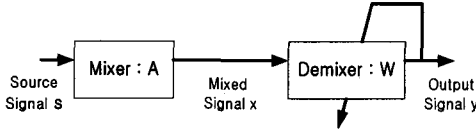


그림 1. ICA를 위한 구성도

일반적으로 혼합된 신호  $\mathbf{x}$ 를 영평균(zero-mean)과 whitening하면, 데이터의 1차와 2차의 통계성에 따른 상관성이 이미 고려되기 때문에 좀 더 효과적으로 ICA를 수행할 수 있다.<sup>[3-6]</sup>

한편 독립성분분석을 위한 고정점 알고리즘은 엔트로피 최적화로 신호 내에 포함된 상호정보를 고려하는 기법으로 그 수렴속도는 적어도 2차 함수정도로, 지금까지 알려진 기법 중 가장 빠른 학습속도를 가진다.<sup>[6,7]</sup> 일반적으로 변수들 상호간의 정보를 표현하는 상호정보는 비가우스성의 척도로 이용되는 negentropy로 정의될 수 있다.  $n$ 개의 랜덤변수 벡터  $\mathbf{y} = (y_1, \dots, y_n)^T$ 를 위한 negentropy  $J(\mathbf{y})$ 는

$$J(\mathbf{y}) = H(\mathbf{y}_{\text{gauss}}) - H(\mathbf{y}) \quad (4)$$

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

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

로 정의된다. 결국 ICA에서 독립성의 최대화를 위해 식 (5)로 정의되는 상호정보  $\mathbf{I}(y_1, y_2, \dots, y_n)$ 를 최소화하는 것은 negentropy를 최대화하는 것이다.

negentropy의 최대화는 식 (5)에서 우측 2번째 항의 최대화로 가능하다. 이를 위한 근사식으로 최대 엔트로피 원리를 이용하면

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

와 같이 정의된다. 여기서  $G$ 와  $c$ 는 각각 비2차(non-quadratic)함수와 상수이고,  $v$ 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 있으며, 독립변수들의 상호간 무관성을 고려한 하나의 독립성분  $y_i = \mathbf{W}^T \mathbf{x}$ 를 찾기 위한 목적함수  $J_G(\mathbf{W})$ 는 식 (6)에 따라

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

로 표현된다. 이를 전체  $n$ 개의 독립성분 모두로 확장하면

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

로 된다. 여기서  $\mathbf{w}_i$ 는 역혼합행렬  $\mathbf{W}$ 의 행벡터 중 하나이며, ICA 변환은  $\mathbf{s} = \mathbf{W}\mathbf{x}$ 로 얻어진다. 결국 식 (7)이나 (8)로 정의된 ICA를 위한 목적함수를 최대화하기 위한 방법이 요구된다.

FP 알고리즘은 batch mode로 계산되는 신경망 알고리즘이다. 데이터 벡터  $\mathbf{x}$ 가 상관행렬  $E\{\mathbf{x}\mathbf{x}^T\} = \mathbf{I}$ 로 whitening되어 있다고 가정할 때, 하나의 독립성분을 구하기 위해 식 (7)에서 FP 알고리즘을 유도하면 다음과 같다.  $J_G(\mathbf{W})$ 의 최대화는 반드시  $E\{G(\mathbf{W}^T \mathbf{x})\}$ 의 최적에서 얻어진다. 이때 Kuhn-Tucker 조건으로부터,  $E\{(\mathbf{W}^T \mathbf{x})^2\} = \|\mathbf{W}\|^2 = 1$  조건하에서  $E\{G(\mathbf{W}^T \mathbf{x})\}$ 의 최적은 다음 식

$$E\{\mathbf{x}g(\mathbf{W}^T \mathbf{x})\} - \beta \mathbf{W} = 0 \quad (9)$$

로 얻어진다. 여기서  $g(\cdot)$ 는 비2차 함수  $G$ 의 미분이고,  $\beta = E\{\mathbf{W}_0^T \mathbf{x} g(\mathbf{W}_0^T \mathbf{x})\}$ 인 상수이며,  $\mathbf{W}_0$ 는  $\mathbf{W}$ 의 최적값이다. 따라서 방정식 (9)를 풀기 위해 빠른 수렴속도를 가지는 할선법<sup>[7]</sup>을 도입한다.

할선법은 현재의 함수 값과 바로 이전의 함수 값만을 이용함으로써 해를 구할 수 있어 기존의 뉴턴(Newton) 기법에 기반을 둔 알고리즘에서처럼 별도의 1차 미분 계산이 요구되지 않는다. 따라서  $\beta$ 를  $\mathbf{W}_0$  대신에  $\mathbf{W}$ 의 현재값으로 나타낸 근사화된 반복방법의 할선법에 기초한 역혼합행렬  $\mathbf{W}$ 를 구하는 식은

$$\begin{aligned} f(\mathbf{W}^\#) &= [E\{\mathbf{x}g(\mathbf{W}^\# \mathbf{x})\} - \beta \mathbf{W}^\#] \\ f(\mathbf{W}) &= [E\{\mathbf{x}g(\mathbf{W}^T \mathbf{x})\} - \beta \mathbf{W}] \\ \mathbf{W}^* &= \mathbf{W} - f(\mathbf{W})\{(\mathbf{W} - \mathbf{W}^\#)^{-1}\}/\{f(\mathbf{W}) - f(\mathbf{W}^\#)\} \\ \mathbf{W}^* &= \mathbf{W} / \|\mathbf{W}\| \end{aligned} \quad (10)$$

와 같다. 여기서  $\mathbf{W}^\#$ 는  $\mathbf{W}$ 의 새롭게 경신된 값이며,  $\mathbf{W}^\#$ 은 바로 전에 계산된  $\mathbf{W}$ 의 값이고,  $\beta = E\{\mathbf{W}^T \mathbf{x}g(\mathbf{W}^T \mathbf{x})\}$ 이다. 결국 식 (10)은 할선법에 기초를 둔 ICA를 위한 FP 알고리즘이다. 여기서는 역혼합행렬  $\mathbf{W}$ 를 구하는데 별도의 1차 미분인 Jacobian matrix의 계산이 요구되지 않아 보다 빠른 계산시간을 가진다.

한편 식 (10)에서의 근사화값 계산에서  $\mathbf{W}$ 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  $\mathbf{W}$ 를 계산할 때 이전 변화의 속성을 반영하는 모멘트를 추가하였다. 이렇게 하면 할선법이 가지는 빠른 수렴속성을 그대로 살리면서도 최적

해로의 수렴에서 발생하는 발진을 억제시켜 더욱 빠른 속도의 독립성분분석이 가능하다. 제안된 할선법과 모멘트에 기초한 역혼합행렬  $W$ 의 경신식은 식 (10)의 3번째 식이 변형되어

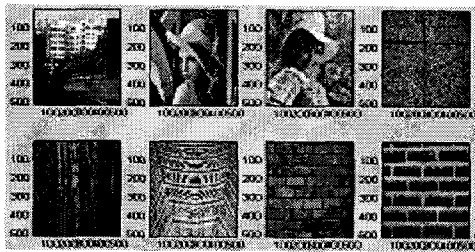
$$W^* = W - f(W) \{ (W - W^{\#}) / (f(W) - f(W^{\#})) \} + \alpha (W^* - W) \quad (11)$$

이 된다. 여기서  $\alpha$ 는 모멘트이다.

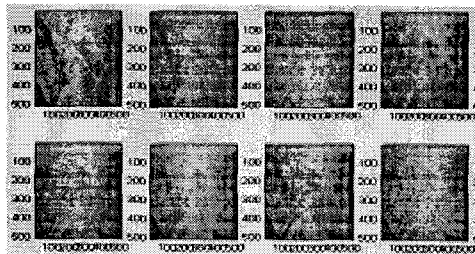
### 3. 시뮬레이션 및 결과분석

제안된 학습알고리즘의 신경망 기반 ICA의 타당성과 성능을 평가하기 위하여 512x512 픽셀 8개 영상들로부터 임의의 혼합행렬로부터 생성된 8개의 혼합영상을 대상으로 실험하였다. 영상의 분리성능을 평가하기 위해 각 데이터의 절대합오차(absolute sum error : ASE) =  $\frac{1}{N} \sum |x_i - y_i|$ 를 정의하여 원영상과 분리된 영상의 정합성을 비교하였다. 여기서  $x_i$ 와  $y_i$ 는 각각 원 신호와 복원된 신호의  $i$ 번째 값이며,  $N$ 은 픽셀의 총 수이다. 영상분리의 학습 종료조건으로는 역혼합행렬  $W$ 의 변화량  $\epsilon$ 이 0.0001 이하일 때나 하나의 영상을 분리하는데 소요되는 학습반복수가 100회 이상일 때로 하였다. 또한 비선형함수  $g(\cdot)$ 는  $(\cdot)^3$ 으로 하였다. 실험은 펜티엄IV-2.2G 컴퓨터 상에서 Matlab 5.1로 실행하였다.

그림 2는 실험에 이용된 풍경화, 레나, 로즈, 바다, 나무무늬, 우산, 적벽돌, 그리고 벽돌타일의 8개 원영상과 임의의 혼합행렬로부터 생성된 8개의 복합영상을 각각 나타낸 것이다. 특히 그림 2(b)는 그림 2(a)로부터 8x8의 랜덤값을 가지는 혼합행렬  $A$ 로부터 생성된 복합영상들이다. 그림에서는 혼합된



(a) 원 영상



(b) 혼합영상

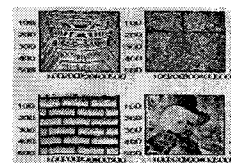
그림 2. 512x512 픽셀의 8개 독립영상

영상들 각각으로부터 원 영상들에 대한 정보들을 발견할 수 없으며, 실험에서는 혼합된 영상을 영평균과 whitening한 후 ICA의 입력신호로 이용하였다.

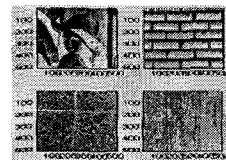
그림 3은 그림 2(b)의 복합영상들을 대상으로 각각 뉴우턴법, 할선법, 할선법과 모멘트의 FP 알고리즘 각각을 수행한 후 분리된 8개의 영상들 중에서 4개만을 보여준 것이다. 여기서도 모멘트  $\alpha = 0.002$ 로 하였다. 그림에서 보면 원 영상과 분리된 영상들 사이의 순서는 일치되지 않아 ICA가 영상의 분리순서는 고려하지 않음을 알 수 있다. 또한 레나나 로즈 영상의 경우 원 영상과 반대의 부호를 가진다. 이 결과 역시 ICA에서 독립성분을 추출할 때 순서뿐만 아니라 부호도 고려하지 않는 속성이 반영된 것이다.



(a) 뉴우턴법의 FP



(b) 할선법의 FP



(c) 할선법과 모멘트의 FP

그림 3. 분리된 8개의 영상 중 4개의 영상

한편 제안된 조합형 FP 알고리즘에서 모멘트  $\alpha$  값을 0에서 1까지 각각 10번씩 변화시키기에 따른 평균 학습반복수  $k$ 를 실험하였다. 여기서 모멘트  $\alpha = 0$ 은 기존의 할선법의 FP 알고리즘이다. 그 결과는 모멘트가 0일 때보다 0.07이상일 때가 상대적으로 많은 학습반복수를 보였으며, 모멘트의 값이 0.001에서 0.06사이일 때 상대적으로 개선된 학습성능을 보였다. 실험에서 모멘트의 값이 0.001에서 0.06사이일 때 평균 학습반복수는 약 92.4회 정도로 0일 때의 약 97.4회 보다 약 1.1배 정도 개선된 학습성능이 있음을 확인하였다.

표 1은 8개의 혼합된 영상에 대해 3가지 FP 알고리즘을 각각 10번씩 반복 실험하여 대응되는 원영상과 복원된 영상에 대한 픽셀 단위의 절대합오차 ASE, 학습반복수  $k$ , 그리고 분리에 소요된 CPU 시간  $t$  각각의 평균을 나타낸 것이다. 여기에서도 제안된 알고리즘의 모멘트는 0.002로 하였다. 실험에서 뉴우턴법의 알고리즘 경우, 영상의 분리를 위한 10회 시행 중 2회가 종료조건을 만족하지 못하였으나 다른 2가지 알고리즘은 시행 모두에서 영상의 분리가 이루어졌다. 이는 뉴우턴법의 알고리즘에서 기술

기인 도함수의 값이 0에 수렴될 경우 역혼합행렬의 경신과정에 0으로 나누어지는 결과가 발생되어 비현실적인 수렴현상으로 추정된다. 한편 표에서 뉴우턴법의 알고리즘 결과는 추가시행을 실시하여 학습의 종료조건을 만족한 경우에 대한 결과이다. 표 1의 10개 영상 중에서 각각 적벽돌, 벽돌타일, 철대문의 절대합오차들을 살펴보면, 다른 영상들에 비해 상대

선법은 뉴우턴법에 비해 약 학습반복수와 분리시간에서 각각 약 1.77배와 약 1.59배 정도 개선되었다. 따라서 제안된 알고리즘은 뉴우턴법의 알고리즘이나 할선법의 알고리즘보다 개선된 분리성과 분리속도가 있으며, 할선법의 알고리즘도 뉴우턴법의 알고리즘보다 우수한 분리성과 빠른 분리속도를 가짐을 알 수 있다.

표 1. 뉴우턴법, 할선법, 제안된 고정점 알고리즘 각각에 의한 영상분리결과

입력 영상	뉴우턴법의 FP 알고리즘			할선법의 FP 알고리즘		
	절대합 오차, ASE	학습 반복수, k	분리 시간, t	절대합 오차, ASE	학습 반복수, k	분리 시간, t
풍경화	0.24	21.15	62.26	0.08	11.5	37.54
레나	0.26	10.42	30.67	0.1	12.31	40.18
로즈	0.33	30.08	88.55	0.09	7.47	24.38
바닥	0.3	7.57	21.08	0.23	8.54	27.88
나무무늬	0.42	13.14	38.38	0.29	9.08	29.64
우산	0.28	7.34	21.61	0.24	6.24	20.37
적벽돌	0.45	17.28	50.87	0.38	3.09	10.09
벽돌타일	0.47	9.2	27.08	0.41	7.5	24.48
계	2.75	116.18	340.5	1.82	65.73	214.36

입력 영상	제안된 FP 알고리즘		
	절대합 오차, ASE	학습 반복수, k	분리 시간, t
풍경화	0.04	9.33	32.19
레나	0.12	11.4	39.33
로즈	0.07	8.31	28.67
바닥	0.19	6.28	21.67
나무무늬	0.26	11.79	40.68
우산	0.18	3.84	13.25
적벽돌	0.35	2.45	8.45
벽돌타일	0.22	7.63	26.32
계	1.43	61.03	210.56

적으로 큰 값을 가진다. 이는 다른 영상들에 비해 이들 3가지 영상들은 다른 영상에 비해 독립성이 상대적으로 낮은 영상이며, 그에 따라서 분리성능도 상대적으로 떨어짐을 보여주는 것이다. 한편 8개 영상 모두에 대한 3가지 알고리즘의 각각의 절대합오차는 약 2.75, 약 1.82, 그리고 약 1.43으로 제안된 알고리즘은 뉴우턴법과 할선법에 비해 각각 약 1.92배와 약 1.27배 정도 개선된 분리성능이 있다. 또한 할선법은 뉴우턴법에 비해 약 1.51배 정도 개선된 성능이 있다. 학습반복수와 분리시간에서 보면, 뉴우턴법은 약 116.18과 약 340.5, 할선법은 약 65.73과 약 214.36, 그리고 제안된 알고리즘은 약 61.03과 약 210.56 정도이다. 먼저 학습반복수에서는 제안된 알고리즘이 뉴우턴법과 할선법에 비해 각각 약 1.9배와 약 1.08배 정도 개선되었으며, 분리시간에서는 각각 약 1.62배와 약 1.02배 정도 개선되었다. 또한 할

#### 4. 결 론

본 연구에서는 할선법과 모멘트에 기초한 조합형 고정점 알고리즘의 신경망 기반 독립성분분석기법을 제안하고, 이를 혼합영상의 분리와 영상의 특징추출에 적용하였다. 여기서 할선법은 negentropy 최대화 과정서 함수 값만을 이용함으로써 계산을 간략하게 하여 역혼합행렬의 경신속도를 빠르게 하기 위함이고, 모멘트는 학습과정에서 발생하는 발진을 억제하여 보다 빠른 학습속도를 얻기 위함이다.

제안된 기법을 512x512의 픽셀의 10개 영상을 대상으로 임의의 혼합행렬에 따라 발생하는 혼합영상의 분리에 적용한 결과, 제안된 알고리즘이 뉴우턴법에 기초한 기존의 알고리즘과 할선법만에 기초한 알고리즘보다 각각 우수한 분리성과 빠른 분리속도가 있음을 확인할 수 있다.

#### 참고문헌

- [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] P. Comon, "Independent Component Analysis A New Concept?", *Signal Processing*, vol.36, No.3, pp.287-314, Apr.1994
- [4] T. W. Lee, *Independent Component Analysis : Theory and Applications*, Kluwer Academic Pub., Boston, 1998
- [5] S. Roberts and R. Everson, *Independent Component Analysis : Principles and Practice*, Cambridge Univ. Press, 2001
- [6] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, Inc., 2001
- [7] K. Atkinson, *Elementary Numerical Analysis'*, John Wiley & Sons, Inc., New York, 1993