인지과학, 제22권 제4호
Korean Journal of Cognitive Science
2011, Vol. 22, No. 4, 347～364.

시각 자극 복원을 위한 비음수 행렬 분해 기반의 영상 기저 자동 추출*

조 성 식 | 박 영 묘 | 이 성 환†
고려대학교 컴퓨터학과 | 고려대학교 바이오정보학 협동과정 | 고려대학교 뇌공학과

본 논문에서는 fMRI를 사용하여 뇌신경 반응을 측정한 후, 자극으로 주어진 10×10 크기의 이진 영상을 사전 정보 없이 복원하기 위해 비음수 행렬 분해를 이용한 자동화된 영상 기저 추출 방법을 제안한다. 영상 기저란 영상을 표현하는 기본 단위로, 기존 연구에서는 사전에 정의된 1×1, 2×1, 1×2, 2×2의 크기를 갖는 총 361개의 영상 기저에 반응하는 뇌 신호를 분석하여 기저 영상으로 복원하고, 모든 기저에 대한 복원 결과를 선형 결합하여 최종 복원 영상을 획득하였다. 사람이 사전에 정의한 영상 기저를 필요로 하는 기존 연구와는 달리, 본 연구에서는 비음수 행렬 분해를 기반으로 학습 데이터로 주어진 이진 영상을 가장 잘 표현하는 영상 기저를 자동 추출하였다. 자동으로 추출된 영상 기저를 사용하여 이진 영상을 복원한 결과, 기존 연구 방법보다 개선된 복원 정확도를 보였다.

주제어 : 시각 자극 복원, 시각 정보 처리, 영상 기저 추출, 비음수 행렬 분해

---

* 본 연구는 2008년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업 연구임(KRF-2008-313-D00938).

† 교신저자: 이성환, 고려대학교 뇌공학과
E-mail: swlee@image.korea.ac.kr

서 론

최근 들어 사람에게 영상, 소리 등을 자극으로 제시하고, 이에 대한 뇌신경 반응을 fMRI로 측정하여 분석하는 연구가 활발히 진행되고 있다. 특히, 영상 자극에 대한 뇌신경 반응 분석 연구는 뇌의 시각 정보 처리의 원리를 규명하는 연구와 자극 영상의 카테고리를 뇌 신호만을 이용하여 인식하는 마음 읽기 연구로 진행되어 왔다[1]. 이러한 마음 읽기 연구의 발전된 형태는 시각 자극의 카테고리만을 인식하는 것에서 더욱 확장된 자극 영상 자체를 복원하는 연구이다(그림 1).

Miyawaki 등[13, 14]은 이진 영상이 자극으로 주어지는 동안 획득한 fMRI 신호를 분석하여 실제로 주어진 자극 영상으로 복원하는 연구를 수행하였다. 위의 연구는 10×10의 이진 영상 자극에 대한 fMRI를 이용하여 본래의 자극 영상을 복원하기 위해서, 사전에 1×1(100개), 2×1(90개), 1×2(90개), 2×2(81개) 크기를 갖는 361개의 영상 기저를 정의 하여 사용하였다. 영상 기저란 영상을 표현하는 기본적인 단위로 Olshausen과 Field 등[16]의 연구에서는 이러한 영상 기저들의 조합으로 영상을 표현 가능함을 보여주었다. Miyawaki 등[13, 14]은 이러한 특성을 이용하여 361개의 영상 기저에 대응되는 fMRI 영상에 대한 복원 모델을 만들고 이를 결합하여 최종적인 복원 결과를 획득하였으나, 사전에 정의된 영상 기저를 필요로 하는 단점이 있다.

본 연구에서는, 기존 방법의 단점을 극복하기 위해 비음수 행렬 분해를 기반으로 학습 데이터로 주어진 이진 영상을 가장 잘 표현하는 영상 기저를 자동 추출하였다. 자동으로 추출된 영상 기저를 사용하여 이진 영상을 복원한 결과 기존 연구 방법보다 개선된 복원 정확도를 보였다.

관련 연구

현재 전 세계의 다양한 연구 그룹에서 fMRI 신호를 분석하여 뇌 질환을 진단하거나 주어진 자극 등에 관한 뇌신경 반응을 분석 및 예측하는 연구가 진행되고 있다. 이 중에서, 시각 자극에 대한 fMRI 신호 분석 연구는 크게 주어진 시각 자극에 대한 카테고리를 인식하는 연구와 주어진 시각 자극 영상 자체를 복원하는 연구로

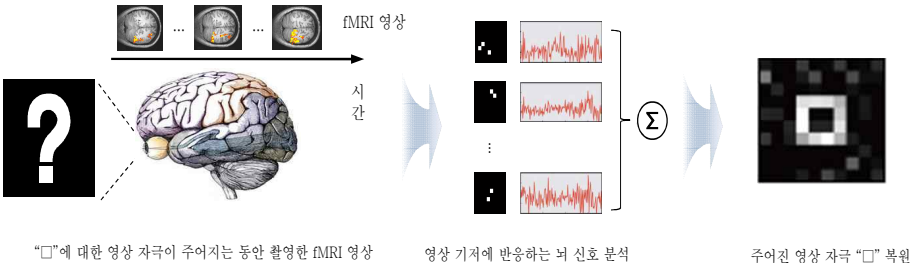

그림 1. fMRI 기반의 시각 자극 복원 개요

나누어진다. Kamitani 등[5, 6]은 시각 자극에 대한 fMRI 신호를 분석하여 영상을 구성하는 기본적인 속성 중 하나인 에지의 방향성을 자동으로 인식하는 연구를 수행하였다. Haxby 등[3, 4]과 Cox 등[1]은 시각 자극에 대한 fMRI 신호로부터 주어진 시각 자극을 사과, 배, 집, 자전거 등과 같은 카테고리로 자동 분류하는 방법에 대한 연구를 수행하였다. Haxby 등은 사람의 얼굴, 집, 의자, 등의 8개의 시각 자극이 보여지는 동안 획득한 fMRI 신호들 간의 상관관계 분석을 통해 각각의 객체가 시각/의미적으로 다른 뇌반응을 보임을 증명하였고, Cox 등은 서포트 벡터 머신(Support Vector Machine)을 기반으로 10개의 객체 카테고리를 fMRI 뇌신호만으로 분류하는 방법을 제안하였다.

단순히 시각 자극에 대한 카테고리만을 인식하는 연구에서 조금 발전된 형태로, 시각 자극이 주어졌을 경우에 대한 뇌신경 반응을 예측하는 연구를 Kay 등[7, 8]이 수행하였다. Kay 등은 fMRI 신호를 분석하여 객체의 카테고리를 인식하는 기존의 연구와는 차별되게 시각 영상이 있을 경우 그와 대응하는 뇌신경 반응을 예측하는 모델링 이론을 제안하였다. Kay 등의 방법은 볼펜 등과 같은 객체의 카테고리는 물론이고 빨간 볼펜, 파란 볼펜 등과 같은 고유한 객체에 대한 식별이 가능한 방법으로 기존의 Haxby 등과 Cox 등의 방법을 일반화시키고 발전시킨 형태의 연구이다. 이와 유사하게 T. Mitchell 등[12]은 60개의 단어 자극이 주어졌을 경우의 뇌신경 반응을 예측하는 모델링 이론을 제안하였다. 이 방법은 사람이 해당 단어 자극이 주어졌을 경우 활성화 되는 fMRI 영상을 예측하기 위하여 25개의 동사(see, hear, listen, taste, smell, eat, touch, rub, lift, manipulate, run, push, fill, move, ride, say, fear, open, approach, near, enter, drive, wear, break, clean) 자극에 대한 fMRI 영상의 선형

조합으로 표현하였다.

이와는 다른 방향으로 발전된 시각 자극 분석 연구로 시각 자극 영상 복원 연구가 최근 시작되고 있다. Naselaris 등[15]은 시각 자극에 대한 뇌신경 반응을 영상의 구조 정보 표현 모델과 의미론적 추론 모델로 동시에 모델링하는 방법을 제안하였다. 영상의 구조 정보 모델은 객체의 형태 등과 같은 시각 정보를 표현하며 의미론적 추론 모델은 앞의 연구에서 설명한 시각 자극의 카테고리 정보를 표현하였다. Naselaris 등은 제안한 모델을 이용하여 자연 영상 복원을 시도하였으나, 자극으로 주어진 영상을 똑같이 복원하는 것이 아닌 유사한 카테고리의 영상으로 복원하는 한계점을 지녔다. Miyawaki 등[13, 14]은 이진 영상으로 주어진 자극에 반응하는 시각 피질의 뇌신호를 fMRI로 측정하고, 이를 복원하는 연구를 수행하였다. 위의 연구는 10×10의 이진 영상 자극에 대한 fMRI를 이용하여 본래의 자극 영상을 복원하기 위해서, 사전에 1×1(100개), 2×1(90개), 1×2(90개), 2×2(81개) 크기를 갖는 361개의 영상 기저를 정의하였다. 이러한 영상 기저들의 조합으로 영상을 표현 가능한 특성을 이용하여 361개의 영상 기저에 대응되는 fMRI 영상에 대한 복원 모델을 만들고 이들의 복원 결과를 선형 결합하여 최종적인 복원 영상을 획득하였으나, 사전에 정의된 영상 기저를 필요로 하는 단점이 있었다.

본 논문에서는, 이러한 기존 연구의 단점을 극복하기 위해 비음수 행렬 분해를 기반으로 학습 데이터로 주어진 이진 영상을 가장 잘 표현하는 영상 기저를 자동 추출하고 이를 기반으로 이진 영상 자극을 복원하는 방법을 제안한다.

제안하는 방법

개요

그림 2는 본 논문에서 제안하는 비음수 행렬 분해(NMF: Non-negative Matrix Factorization)[10]로 추출한 기저를 이용하여 fMRI 데이터에서 자극 영상을 복원하는 과정에 대한 흐름도이다. 학습 단계에서는 비음수 행렬 분해를 이용해 시각 자극으로 주어진 랜덤 영상으로부터 공통적으로 가장 잘 나타나는 부분을 표현하는 기

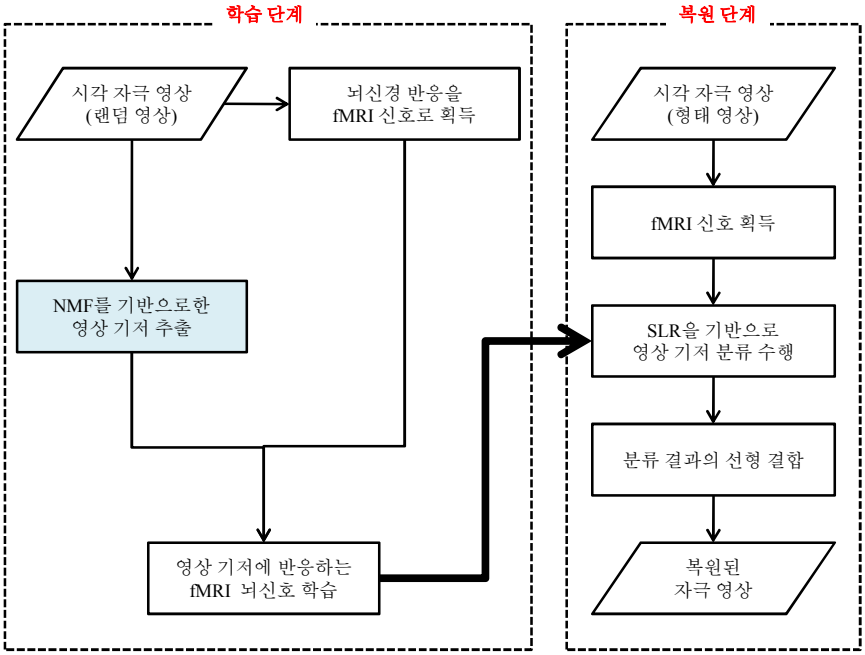

그림 2. 제안하는 기저 추출 방법을 통한 fMRI 데이터로부터의 시각 자극 영상 복원 흐름도

저를 추출한다. 시각 자극에 대한 뇌신경 반응을 기록한 fMRI 데이터로부터 SLR(Sparse Logistic Regression) 알고리즘[17]을 이용해 교사학습 방법으로 추출된 기저의 반응을 학습한다. 학습 결과로 fMRI에 대한 기저의 반응을 얻는 분류기를 얻게 되며, 이 분류기는 새로운 시각 자극에 대한 fMRI 데이터로부터 기저의 반응을 예측하고, 각 분류기의 결과를 선형 결합하여 최종적으로 복원된 영상을 얻는다. 기저의 추출, 학습 및 복원 결과에 대한 성능 평가를 위해 Miyawaki 등과 동일한 영상 및 fMRI 데이터를 사용하였다[13]. 영상 데이터는 10×10 패치의 이진 영상으로 구성되어 있으며, 랜덤 영상과 형태(기하학적 모양과 알파벳 모양) 영상의 두 분류로 나뉘어진다. fMRI 데이터는 사람이 자극으로 주어진 영상 데이터를 볼 때에 뇌의 여러 영역 중, 주 시각 영역 V1에 해당하는 시각 영역에서의 반응을 기록한 것이다(데이터에 대한 자세한 내용은 5절에서 설명 한다).

비음수 행렬 분해를 이용한 영상 기저 추출

NMF는 주어진 학습 데이터를 구성하는 지역적인 특징을 갖는 기저를 추출하는 특성이 있다. 이는 얼굴 데이터 집합으로부터 NMF를 이용해 눈, 코, 입 등의 형태를 갖는 기저를 추출 해낸 기존 연구 결과를 통해 증명되었다[9]. 본 논문에서는 NMF를 이용하여 학습 데이터로 주어진 이진 랜덤 영상으로부터 이들을 구성하는 지역적인 특징을 갖는 기저를 추출하였다. 이 기저는 다음 식과 같이 학습 데이터로 주어진 영상 집합을 음이 아닌 원소를 갖는 두 행렬로 분해함으로써 얻어질 수 있다.

$$V \cong WH \tag{1}$$

이 때, V는 n×m 크기의 학습 데이터 집합이며, n은 데이터의 특징 차원을, m은 데이터의 개수를 나타낸다. W는 $n \times r$ 크기의 V를 구성하는 기저의 집합이며, r은 기저의 개수를 의미한다. H는 $r \times m$ 크기의 행렬로, 학습 데이터에서 W의 열 벡터인 기저 영상이 발생하는 정도를 나타낸다. 본 논문에서 사용된 학습 데이터는 10×10 크기의 이진 영상이므로 n=100이다. W와 H에 대한 학습은 식 (2)와 같이 목적 함수를 구성하여, 이를 최소화 하는 W와 H를 구하는 것이다.

$$\min_{W,H} \quad f(W,H) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}(V_{ij} - (WH)_{ij})^2 \tag{2}$$

$$\text{단,} \quad W_{ai} \geq 0, H_{bj} \geq 0, \text{for } all\, a, b, i, j$$

여기서 a와 i, 그리고 j와 b는 각각 W와 H의 행과 열에 대한 첨자이다. 식 (2)에서 기술한 목적함수에 대한 조건을 만족시키기 위해 W와 H를 각각 반복적으로 갱신하여 V와 WH의 차이를 최소화 시키는 Multiplicative update 방법을 사용하며, 이를 위한 조건식은 식 (3), (4)와 같다.

$$W_{ai} \leftarrow \frac{W_{ai}}{\sum_j W_{aj}} \tag{3}$$

$$H_{a\mu} \leftarrow H_{a\mu} \sum_i W_{ai} \frac{V_{i\mu}}{(WH)_{i\mu}} \tag{4}$$

여기서 아래 첨자들은 각 행렬의 행과 열을 나타내는 첨자이다.

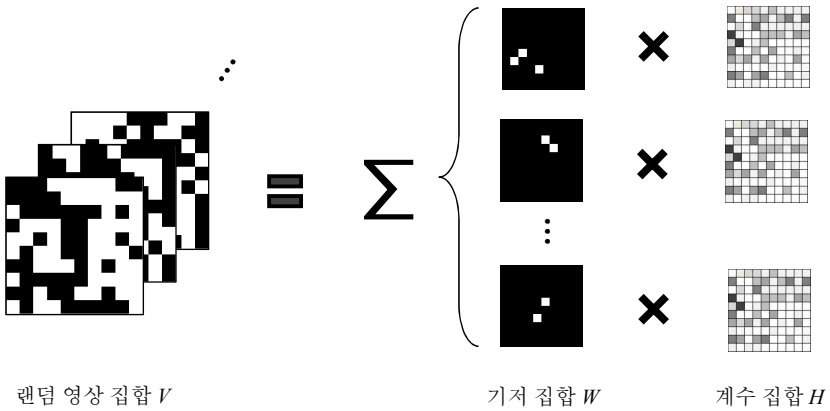

그림 3. NMF를 이용하여 랜덤 영상으로부터 기저를 추출하는 것의 도식화.학습에사용된 랜덤 영상은 추출된 기저 집합 W와 계수 집합 H의 선형 결합으로 표현 가능하다.

그림 3은 본 논문에서 사용한 랜덤 영상으로부터 NMF를 이용하여 기저를 추출하는 개념도를 나타낸다. 기저 추출을 위한 학습 데이터는 100×1 크기의 열 벡터로 변환된 10×10 크기의 랜덤 영상의 집합으로 구성된다. 여기에서 랜덤 영상이란, 10×10 크기의 이진 영상으로 영상의 각 픽셀 값을 무작위로 할당한 22 가지 종류의 영상을 말한다. 영상 자극 복원 실험에서 사용하는 특정한 형태의 이진 영상을 기저 추출을 위한 학습 데이터로 사용할 경우 사전 정보를 가지고 복원하게 되고, 다양한 형태의 영상 복원이 가능한 방법으로 확장 하는 것에 제한적이기 때문에 상기와 같은 형태의 랜덤 영상으로부터 기저를 추출하였다. 학습 결과, 랜덤 영상 집합 V를 이루는 데이터들의 지역적이고 공통적인 특성을 나타내는 100×1 크기로

표현된 기저의 열 벡터 집합 W와 V에 대한 W안의 기저에 대한 계수 값을 갖는 H를 얻는다. 상기의 Multiplicative update 학습법은 얼굴 영상에 대해선 눈, 코, 입과 같이 공통된 특징을 갖는 기저를 추출하였으나, 본 논문에서 사용하는 10×10 크기의 이진 영상에 적용한 결과, 영상으로부터 공통된 특징을 나타내는 기저를 얻지 못하였다. V에 대한 W와 H를 얻기 위한 Multiplicative update 학습법은 구현 과정이 간단하여 널리 사용되어 왔지만, V와 WH 사이의 거리를 근사화 하는 상기의 방법은 국부 최소점으로의 수렴을 보장해주지 못함을 보였다[2, 11]. 본 논문에서는 이와 같은 문제점을 보완하기 위해 식 (5)와 같은 Projected gradient 방법[10]을 사용하여 주어진 랜덤 학습 데이터 영상으로부터 W와 H의 값을 학습하였다.

$$(\widetilde{W}, \widetilde{H}) \equiv P\left[(\overline{W}, \overline{H}) - \alpha\left(\nabla_{w} f(\overline{W}, \overline{H}), \nabla_{H} f(\overline{W}, \overline{H})\right)\right] \tag{5}$$

이 때, α는 학습 속도를 결정하는 파라미터, $\nabla_{W} f()$와 $\nabla_{H} f()$는 각 W와 H에 대한 f의 도함수이며, 이 방법은 빠른 속도로 국부 최소로의 수렴이 보장된다. Multiplicative update 방법[9]은 학습영상에 대해 뚜렷하지 않은 형태의 기저를 얻는 반면, 본 논문에서 적용한 projected gradient 방법[10]으로 학습된 기저는 학습 데이터인 랜덤 영상으로부터 공통적으로 나타나는 부분을 뚜렷하게 표현하는 것을 확인 할 수 있었다. 본 논문에서는 NMF를 이용해 96개의 기저 영상을 추출 하였다. 이는 Miyawaki 등[13, 14]이 사용한 사전에 미리 정해진 1×1, 1×2, 2×1, 2×2 크기의 361개의 기저보다 적은 수이며, 기저의 형태 및 수를 자동으로 추출했다는 점에서 기존의 연구보다 큰 장점을 갖는다. 또한 기존의 방법은 많은 수의 기저로 인해 표현 범위가 중복되어 표현의 정확성이 떨어지지만, 본 논문에서 제안된 방법은 적은 수의 기저로 영상을 표현하기 때문에 이러한 문제를 해결 할 수 있다. 일반적으로 행렬 계수가 r인 비음수 행렬 분해에서는 $(n+r)r < nm$의 조건을 만족시키는 개수의 기저를 구하게 되는데, 본 논문에서는 최대치를 충족시키는 96개의 기저를 추출하였다.

기저의 학습 및 영상 복원

SLR을 이용한 기저의 학습

SLR은 로지스틱 회귀분석 방법의 일종으로, 로지스틱 회귀분석이 단순히 있음/없음의 두 가지 값을 판별하는 것에 활용된다면, SLR은 로지스틱 회귀분석을 베이즈 추정의 범위로 확장시킨 기법이다. 특징 벡터의 차원 축소와 판별을 위한 모델(가중치) 추정이 동시에 실시되며, 뇌 신호와 같이 특징 벡터의 차원이 매우 높은 데이터의 경우에 유용하다. 본 논문에서는 SLR 알고리즘[17]을 사용하여 제안하는 방법으로 추출한 기저들에 대응하는 fMRI 신호에 대한 반응을 학습하였다. SLR은 교사 학습 방식의 분류기 학습 방법이다. 교사 학습을 위해서는 데이터와 해당 데이터가 속하는 클래스 레이블의 쌍으로 학습 데이터를 구성하게 되는데, 본 논문에선 클래스 레이블 값으로 그림 4와 같이 각 기저에 해당되는 시각 영상의 픽셀 값을 사용하며, 기저가 학습 영상으로부터 하나 이상의 대응되는 픽셀을 가질 경우에는 해당 픽셀 값들의 평균으로써 레이블 값을 계산한다. 이렇게 96개의 기저에 대응되는 입력 영상의 레이블 값과 영상에 대응되는 fMRI 데이터 쌍의 집합을 학습 데이터로 구성하여 학습 데이터가 분포된 공간에서 클래스 분류 성능을 최대

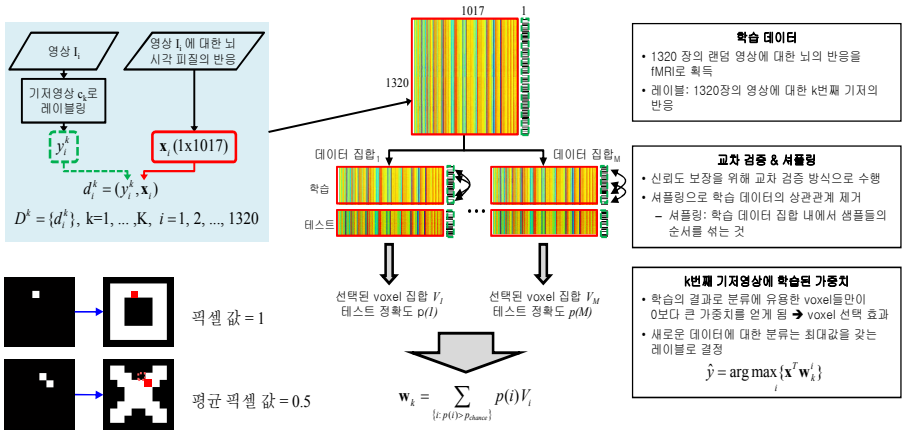

그림 4. 레이블 할당 및 교차 검증 과정 수행을 통해 가중치 계산

로 하는 분류식을 학습한다. fMRI와 같은 뇌신호 데이터의 경우, 학습에 사용되는 샘플의 수 보다 특징(복셀)의 수가 큰 경우가 많다. 본 논문에서도 학습에 사용되는 데이터는 1320개이고 데이터의 특징은 1017차원이므로 올바른 학습 결과를 얻기 어렵다. SLR은 학습 과정에서 분류에 유용한 특징들만을 선택하고, 분류에 큰 기여를 하지 않는 특징들에 대한 가중치를 0으로 할당하여, 주어진 데이터에서 유용한 특징(복셀)을 선택하는 특성이 있다. 본 논문에서는 96개의 기저를 학습 한 결과 평균적으로 89개의 복셀만이 분류에 유용한 복셀로 선택되었다.

영상 복원

그림 5는 입력 fMRI 데이터로부터의 복원 과정에 대한 흐름도이다. SLR을 통해 학습된 분류기로 입력 fMRI 데이터로부터 기저에 대한 반응을 분류한다. 이 결과로써 해당 기저 c_i에 대한 레이블 y_i와 그에 대한 확률 값 p_i를 얻게 된다. fMRI 데이터에 대한 n개의 기저의 분류 값과 확률 값을 선형 결합하여 최종적으로 자극으로 주어진 이진 영상에 대한 복원을 수행한다. n개의 기저에 대한 분류결과를

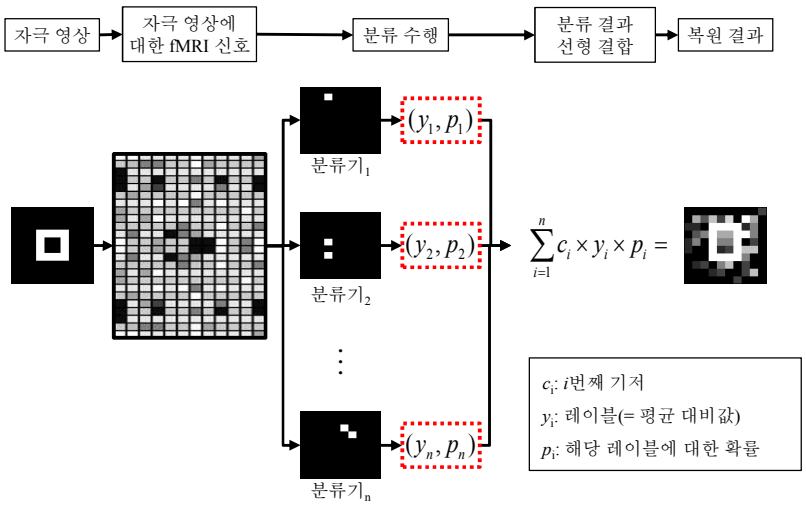

그림 5. fMRI 데이터로부터 시각 자극 영상을 복원하는 과정. 분류기$_1$ ... 분류기$_n$은 각 기저에 해당하는 뇌신호를 인식하기 위한 분류기를 의미한다.

통해 복원된 영상을 얻는 식은 다음과 같다.

$$\sum_{i=1}^{n} c_i \times y_i \times p_i = \hat{I}_k \tag{6}$$

여기서 n은 기저의 개수이며, $\hat{I}_k$는 k번째 fMRI 데이터로부터 복원된 영상이다.

실험 결과 및 분석

실험 환경

본 논문에서는 Miyawaki 등[13, 14]의 연구에서 사용한 fMRI 데이터를 이용하여 제안한 알고리즘의 성능을 평가하였다. Miyawaki 등이 사용한 fMRI 획득 실험 과정은 랜덤 영상 세션과 형태 영상 세션으로 나누어지는데, 랜덤 영상 세션에서는 6 Hz로 깜박이는 체크보드로 이루어진 10×10 크기의 랜덤 패치가 피험자에게 시각 자극으로 제시가 된다. 각 자극은 6초간 화면에 표시되었다가 6초간 사라지는 것을 반복하며, 총 22개의 다른 종류의 랜덤 패턴이 제시가 된다. 형태 영상 세션에

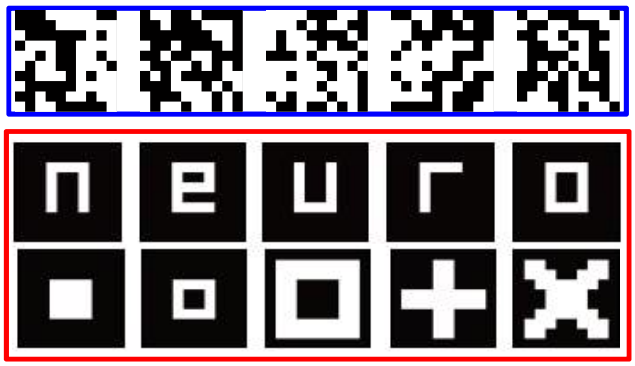

그림 6. 학습에 사용된 랜덤 영상(상)과 실제 복원에 사용된 10개의 기하학적 형태 영상 및 알파벳(하)의 예

서는, 역시 반짝거리는 체크보드로 이루어진 10×10 크기의 영상으로 구성이 되어 있으며, 5개의 기하학적 패턴과 6개의 알파벳 패턴이 2가지의 크기로 구성이 되어 있다. 형태 영상 세션은 12초간 피험자에게 보였다가 사라지는 것이 반복된다. 이 세션 동안 피험자는 화면의 중앙에 시선을 고정한 상태에서 자극을 관찰하게 된다. 랜덤 세션에서는 서로 다른 22의 종류의 형태의 랜덤 영상에 반응하는 뇌 신호를 20회로 나누어 촬영하였으며, 형태 세션에서는 서로 다른 10 종류의 형태 영상에 반응하는 뇌 신호를 12회에 나누어 촬영하였다(TR: 2000ms, TE: 300ms). 피험자의 뇌신호를 측정하기 위해 3.0 Tesla MAGNETOM Allegra fMRI 기기가 사용되었다. 획득된 fMRI 신호는 SPM2를 사용하여 slice-timing correction과 3D motion correction의 전처리를 하고 뇌의 각 영역이 분류된 상태로 1명의 피험자에 대한 데이터가 공개되어있다[18]. 그림 6은 피험자에게 제시되는 시각 자극의 예이다. 본 논문은 Miyawaki 등[13, 14]의 데이터에서 피험자 한 명으로부터 획득한 동일한 형태의 시각 자극과, 뇌의 시각피질 영역 중 V1 영역에서의 반응을 기록한 fMRI 신호를 사용하여 실험을 수행하였다.

실험 결과 및 분석

영상 기저에 대한 학습은 1320장(22×20×3)의 랜덤 영상 및 이에 대한 fMRI 데이터를 갖고 수행하였으며, 성능 평가를 위해 기하학적 형태 영상과 알파벳 영상으로 구성된 720장(12×10×6)의 이진 영상에 대한 fMRI 데이터를 사용하였다. 그림 7

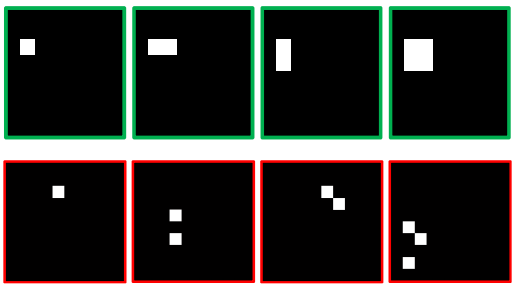

그림 7. Miyawaki 등[13, 14]이 제안한 사전에 정의된 기저(상)와 비음수 행렬 분해[15]로 추출한 기저(하)의 예

은 Miyawaki 등이 사용한 사전에 정의된 기저와 제안하는 방법으로 추출한 기저를 보여준다. 사전에 정의된 기저는 10×10 크기의 시각 자극 영상을 표현하기 위해 1×1, 1×2, 2×1, 2×2 4가지 크기의 기저에 대해 각각 100개, 90개, 90개, 81개씩 총 361개를 구성하였다. 제안하는 비음수 행렬분해 알고리즘으로 추출한 96개의 기저는 우선 기저의 수에 있어서 Miyawaki의 방법보다 크게 줄어든 것을 알 수 있다. 이는 NMF의 지역적으로 공통적인 특성을 표현하는 기저를 추출하는 특징으로 인한 것으로, 기존에 사용된 기저와 달리 1×1 크기의 단일 픽셀로 구성된 기저 외에도 다양한 형태의 공간적인 연결성을 갖는 기저가 추출되었다. 이는 영상을 표현하는 기저의 표현 범위가 중복되는 기존의 방법보다 더 효과적인 것을 알 수 있다. 비교 실험은 기하학적 형태와 알파벳 형태의 자극 영상에 대한 복원의 정확도로써 평가하였다. 추가적으로, 기저의 개수를 다양하게 추출함으로써 어떤 결과가 나오는지 비교해 보았다. 복원의 정확도는 두 영상간의 유사도를 잘 표현할 수 있는 식 (7)의 공간적 상관관계를 비교하였다(x_i: 추정할 영상의 픽셀, y_i: 추정된 영상의 픽셀, $\overline{x}$: x_i의 평균, $\overline{y}$: y_i의 평균, n: 총 픽셀의 개수). 비음수 행렬분해 알고리즘의 수식을 이용하여 96개의 기저를 추출하였지만, 개수를 조정해서 비교했을 때의 성능을 평가해보기 위해 기저의 수를 40개, 70개, 그리고 최적화된 개수인 96개로 조정하여 비교하였다.

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{(n-1)s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \qquad (7)$$

그림 8은 주어진 fMRI 데이터로부터 복원된 영상의 예를 나타낸다. 그림 9는 720장의 형태 영상에 대해 각각 기하학적 형태와 알파벳 형태에 대한 복원 정확도를 원래의 입력 영상간의 공간적 상관관계로 표현한 그래프이다. 특히, 제안한 방법으로 복원한 결과가 기하학적 형태를 복원하는 부분에 있어서 사전에 정의된 기저로 복원한 결과보다 뚜렷하게 뛰어남을 확인 할 수 있다.

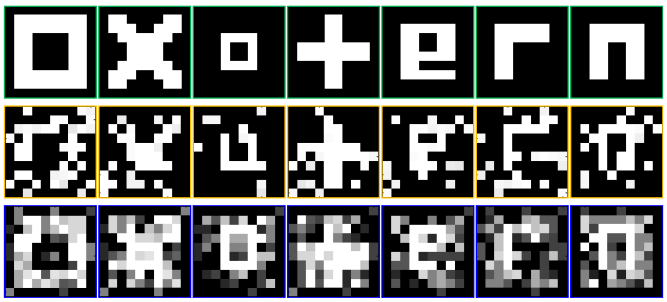

그림 8. 제시된 시각 영상(상)에 대한 복원 결과: Miyawaki 등의 방법(중), 본 논문에서 제안한 방법(하)

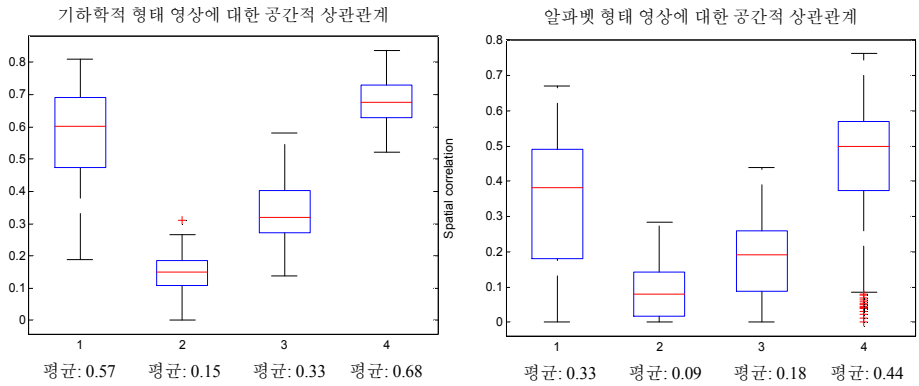

그림 9. 기하학적 형태의 영상(좌)과 알파벳 영상(우)에 대한 복원 결과 비교. 1은 Miyawaki 등의 결과, 2, 3, 4는 각각 본 논문에서 제안한 방법으로 기저를 40, 70, 96개를 추출했을 경우의 복원 성능 결과

결론 및 향후 연구 방향

본 논문에서는 시각 자극으로 주어진 10×10 크기의 이진 영상에 대한 피험자의 뇌의 반응을 fMRI를 사용하여 측정한 후, 사전 정보 없이 복원하기 위한 자동화된 영상 기저 추출 방법을 제안하였다. 기존 연구는 361개의 기저를 사전에 정의함으로써 기저 간 표현 범위가 중복되어 영상 복원률이 낮은 결과를 얻었다. 이에 반

하여 본 논문에서 제안한 방법은, 비음수 행렬 분해 알고리즘의 지역적인 특징을 갖는 기저를 추출하는 특성을 사용하여 주어진 이진 영상을 가장 잘 표현하는 96개의 영상 기저를 자동 추출하였다. 비교 실험을 통해 기존 방법보다 표현 범위가 중복되지 않는 적은 수의 기저로 복원 결과를 비교해 본 결과, 본 논문에서 제안한 방법이 기존의 방법보다 정확하게 시각 자극 영상을 복원하는 것을 확인할 수 있었다.

향후 연구로는, 이진 이미지의 간단한 기하학적인 패턴이나 알파벳 패턴을 시각 영상으로 사용하는 한계점을 극복 가능한 더욱 복잡한 형태를 갖는 자연 영상을 복원하기 위한 연구가 필요할 것이다. 시각 피질의 V1에서 나타나는 1차적인 반응뿐만 아니라, V2, V3, IT(Inferior Temporal gyrus)와 같이 고차원적인 정보를 처리하는 시각 정보의 처리 과정을 고려한다면, 본 논문에서 제안한 방법보다 정확한 복원이 가능하여 10×10 크기의 단순한 형태의 영상뿐만 아니라 자연 영상 등과 같이 훨씬 복잡한 형태의 영상을 복원 할 수 있을 것이라 기대된다.

참고문헌

[1] D. Cox and L. Savoy, "Functional Magnetic Resonance Imaging (fMRI) "Brain Reading": Detecting and Classifying Distributed Patterns of fMRI Activity in Human Visual Cortex," *NeuroImage, Vol. 19*, No. 2, pp.261-270, June 2003.

[2] E. Gonzalez and Y. Zhang, "*Accelerating the Lee-Seung Algorithm for Nonnegative Matrix Factorization,"* Technical Report, Depart. of Computational and Applied Mathematics, Rice Univ., 2005.

[3] J. Haxby, M. Gobbini, M. Furey, A. Ishai, J. Schouten, and P. Pietrini, "Distributed and Overlapping Representations of Faces and Objects in Ventral Tem-poral Cortex," *Science, Vol 293,* No. 5539, pp.2425-2430, September 2001.

[4] J. Haxby, L. Ungerleider, V. Clark, J. Schouten, E. Hoffman, and A. Martin, "The Effect of Face Inversion on Activity in Human Neural Systems for Face and Object Perception," *Neuron, Vol. 22,* No. 1, pp.189-199, January 1999.

[5] Y. Kamitani and F. Tong, "Decoding Seen and Attended Motion Directions from Activity in the Human Visual Cortex," *Current Biology, Vol.16,* No.11, pp.1096-1102, June 2006.

[6] Y. Kamitani and F. Tong, "Decoding the Visual and Subjective Contents of the Human Brain," *Nature Neuroscience, Vol.8,* No.5, pp.679-685, May 2005.

[7] K. Kay and J. Gallant, "I Can See what You See," *Nature Neuroscience, Vol. 12,* No. 3, pp.245-246, March 2009.

[8] K. Kay, T. Naselaris, R. Prenger, and J. Gallant, "Identifying Natural Images from Human Brain Activity," *Nature, Vol. 452,* March 2008, pp.352-356.

[9] D. Lee and H. Seung, "Learning the Parts of Objects by Non-Negative Matrix Factorization," *Nature, Vol. 401,* No. 10, pp.788-791, October 1999.

[10] C.-J. Lin, "Projected Gradient Methods for Nonnegative Matrix Factorization," *Neural Computation, Vol. 19,* No. 10, pp.2756-2779, October 2007.

[11] C.-J. Lin, "On the Convergence of Multiplicative Update Algorithms for Nonnegative Matrix Factorization," *IEEE Transactions on Neural Networks, Vol. 18,* No. 6, pp.1589-1596, 2007.

[12] T. Mitchell, S. Shinkaraeva, A. Carlson, K. Chang, V. Malave, R. Mason, and M. Just, "Predicting Human Brain Activity Associated with the Meanings of Nouns," *Science, Vol 320,* No. 5880, pp. 1191-1195, May 2008.

[13] Y. Miyawaki, H. Uchida, O. Yamachita, M. Sato, Y. Morito, H. Tanabe, N. Sadato, and Y. Kamitani, "Visual Image Reconstruction from Human Brain Activity using a Combination of Multiscale Local Image Decoders," *Neuron, Vol. 60,* No. 5, pp.915-929, December 2008.

[14] Y. Miyawaki, H. Uchida, O. Yamachita, M. Sato, Y. Morito, H. Tanabe, N. Sadato, and Y. Kamitani, "Visual Image Reconstruction from Human Brain Activity: A modular decoding approach," *Journal of Physics : Conference Series, Vol. 197,* No. 1, pp.1-13, December 2008.

[15] T. Naselaris, R. Prenger, K. Kay, M. Oliver, and J.Gallant, "Bayesian Reconstruction of Natural Images from Human Brain Activity," *Neuron, Vol. 63,* No. 6, pp.902-915,

September 2009.

[16] B. Olshausen and D. Field, "Emergence of Simple-cell Receptive Field Properties by Learning a Sparse Code for Natural Images," *Nature, Vol. 381,* No. 13, pp.607-609, June 1996.

[17] O. Yamashita, M. Sato, T. Yoshioka, F. Tong, and Y. Kamitani, "Sparse Estimation Automatically Selects Voxels Relevant for the Decoding of fMRI Activity Patterns," *NeuroImage, Vol. 42,* No. 4, pp.1414-1429, October 2008.

[18] http://www.cns.atr.jp/dni/en/downloads/fmri-data-set-for-visual-image-reconstruction/

1 차원고접수 : 2011. 8. 1
2 차원고접수 : 2011. 11. 14
최종게재승인 : 2011. 12. 6

(*Abstract*)

Automatic Extraction of Image Bases Based on Non-Negative Matrix Factorization for Visual Stimuli Reconstruction

Sung-Sik Cho[1] Young-Myo Park[2] Seong-Whan Lee[3]

[1]Dept. of Computer Science and Engineering

[2]Interdisciplinary Graduate Program in Bioinformatics

[3]Dept. of Brain and Cognitive Engineering, Korea University

In this paper, we propose a automatic image bases extraction method for visual image reconstruction from brain activity using Non-negative Matrix Factorization (NMF). Image bases are basic elements to construct and present a visual image. Previous method used brain activity that evoked by predefined 361 image bases of four different sizes: 1×1, 2×1, 1×2, and 2×2. Then the visual stimuli were reconstructed by linear combination of all the results from these image bases. While the previous method used 361 predefined image bases, the proposed method automatically extracts image bases which represent the image data efficiently. From the experiments, we found that the proposed method reconstructs the visual stimuli better than the previous method.

Key words : Visual Stimuli Reconstruction, Visual Information Processing, Extraction of Image Bases, Non-Negative Matrix Factorization