

국부기저영상을 이용한 수화영상 인식

조용현, 홍성준, 이화주*

*대구가톨릭대학교 컴퓨터정보통신공학부

e-mail:{yhcho,sjishong,hj7989}@cu.ac.kr

Sign Language Images Recognition Using Local Basis Images

Yong-Hyun Cho, Seong-Jun Hong, Hwa-Ju Lee**

School of Computer and Information Comm. Eng., Catholic Univ. of Daegu

요 약

본 논문에서는 각 개인의 동작영상에 대한 국부고유공간에 바탕을 둔 기저영상을 이용한 효율적인 수화영상 인식 기법을 제안하였다. 여기서 국부고유공간의 추출은 주요성분분석을 이용한 것으로 동작영상의 국소특징을 더욱 더 잘 반영하기 위함이고, 기저영상의 추출은 독립성분분석을 이용한 것으로 수화영상 내에 포함된 고차원의 독립적인 특징들을 반영하여 보다 개선된 인식성능을 얻기 위함이다. 제안된 기법을 240*215 픽셀의 80(1명*5동물*16동작)개 동물을 표현하는 수화동작을 대상으로 Euclidean의 분류척도를 이용하여 실험한 결과, 단순 국부고유공간을 이용한 방법에 비해 우수한 인식성능이 있음을 확인하였다.

1. 서론

동작인식은 휴먼로봇 인터페이스 분야의 기본적인 상호작용 중에 하나이며, 음성인식이나 얼굴 및 표정 인식과 같이 여러 분야에 응용 될 수 있다. 따라서 사람과 로봇과의 원활한 의사소통을 위해 시각을 기반으로 사람의 동작의도를 파악하려는 연구들이 활발하게 진행되고 있다 [1-6].

일반적으로 손동작이나 표정 및 얼굴 등의 인식에서 특징추출은 인식성능을 결정짓는 매우 중요한 파라미터로 특징기반 방법과 영상기반 방법이 이용되고 있다[1-4]. 특징기반 방법은 특징과 관련된 정보들을 이용하는 것으로 지역적 특징추출 및 특성을 벡터화한 Gabor 필터를 이용한 방법들이 있다[6]. 이들 방법들에서는 차원감소나 수행시간의 절약, 포즈나 크기 등에 강건한 특성이 있으나 인식에 필요한 정보손실에 따라 인식성능이 크게 달라지는 제약이 있다. 영상기반 방법은 영상 전체를 대상으로 특징을 계산하는 방법으로 holistic analysis에 바탕을 둔 기법들이 주로 이용되고 있으며, 주요성분분석(principal component analysis : PCA)[6]과 독립성분분석(independent component analysis : ICA)을 이용한 기법들이 주로 이용되고 있다[7]. PCA는 영상 속에 포함된 진폭을 고려한 2차원적 속성을 이용하나 ICA는 이를 일반화 한 기법으로 고차원의 속성을 이용한다. 이 방법들은 중복신호 제거의 효과가 있어 뇌의 시각 기능을 잘 반영하며 타 방법에 비해 높은 인식률을 얻을 수 있는 장점을 갖고 있다. 특히 PCA가 신호의 차원 자체를 줄여 조밀한 코딩을 유

도하는데 비해 ICA는 출력 신호의 값을 드물게 나타내는 방법으로 중복 신호를 제거하며 이는 뇌의 기능 모델에 좀 더 관련이 있다. 하지만 ICA에서는 데이터의 통계적 독립성을 판정하기 위해 고차원의 cumulant 계산이 요구되어 시간이 오래 걸린다. 이러한 제약을 해결하기 위해서 최근 고정점 (fixed-point : FP)이나 기울기를 이용하는 알고리즘들이 제안되었다[7].

한편 수화는 청각 장애인 의사전달을 위해 사용하는 동작언어이다[3-5]. 수화는 손의 움직임과 형태로 표현가능하며, 어휘를 표현하기 위한 수화단어와 자음, 모음, 숫자, 알파벳 등을 표현하기 위한 지화로 구성되어 있다. 이러한 수화동작의 인식을 위해서는 의미 없는 손동작과 의미 있는 수화단어를 구분해 인식하는 것이 필요하다. 일반적으로 수화는 손의 동작과 형태 정보를 가지고 정의되기 때문에 손의 모양과 움직임을 동시에 고려해 특징으로 사용한다. 또한 수화인식에서 가장 중요한 문제는 의미 있는 동작과 의미 없는 동작을 구분하는 것이며, 특히 의미 없는 동작의 모델링을 위해서는 무한대의 학습데이터가 필요하기 때문에 일반적인 방법으로 해결이 불가능하다[3,4].

본 논문에서는 각 개인의 동작영상에 대한 국부고유공간에 바탕을 둔 기저영상을 이용한 효율적인 수화영상 인식 기법을 제안하였다. 여기서 국부고유공간의 추출은 주요성분분석을 이용한 것으로 동작영상의 국소특징을 더욱 더 잘 반영하기 위함이고, 기저영상의 추출은 독립성분분석을 이용한 것으로 수화영상 내에 포함된 고차원의 독립적인 특징들을 반영하여 보다 개선된 인식성능을 얻기 위

합이다. 제안된 기법을 240*215 픽셀의 80(1명*5동물*16 동작)개 동물을 표현하는 수화동작을 대상으로 Euclidean의 분류적으로도 이용하여 실험하였으며, 그 결과를 국부고유영상을 이용한 기법의 결과와 각각 비교·고찰하였다.

2. 국부고유공간 기반 기저영상의 추출

동작은 연속적으로 변하는 포즈의 집합이다. $M_0 \times N_0$ (=N) 픽셀을 가지는 P개 연속적으로 표본화된 정규화 동작영상 \mathbf{x}_p ($p=1, 2, \dots, P$)에 대해서 살펴보자. 만약 한 사람을 대상으로 H개의 동작을 가진다면, 동작 h에 대한 영상은 \mathbf{x}_p^h ($h=1, 2, \dots, H$)로 나타낼 수 있다. 각각의 동작에 대해서 영상의 스트림은 P개씩 표본화 되며, 이들 $N \times PH$ 개 동작영상들로 구성된 단일행렬 \mathbf{X} 는 다음의 식 (1)과 같다.

$$\mathbf{X} = (\mathbf{x}_1^1, \mathbf{x}_2^1, \dots, \mathbf{x}_P^1, \mathbf{x}_1^2, \mathbf{x}_2^2, \dots, \mathbf{x}_P^2, \dots, \mathbf{x}_1^H, \mathbf{x}_2^H, \dots, \mathbf{x}_P^H) \quad (1)$$

식 (1)의 전체 동작영상 행렬을 이용한 공분산 행렬 $\mathbf{C} = \mathbf{X}\mathbf{X}^T$ 로 정의되며, 여기서 T는 정치행렬을 나타낸다. 행렬 \mathbf{C} 의 고유벡터 \mathbf{e}_j 에 의한 고유치 λ_j 는 고유방정식 $\mathbf{C}\mathbf{e}_j = \lambda_j \mathbf{e}_j$ 을 이용하여 계산될 수 있다. 이때 \mathbf{C} 의 모든 고유벡터에 의해서 정의되는 N차원 공간은 PCA에 의해서 감소될 수 있다[7]. 결과로 만약 K ($1 \leq K \ll N$)개의 고유치와 고유벡터가 선택된다면, 고유벡터에 의해 정의되는 K 차원의 공간은 고유공간이다. 또한 영상 \mathbf{x}_p^h 는 다음의 식 (2)에 의해 고유공간 내에서 고유영상 \mathbf{g}_p^h 로 투영된다.

$$\mathbf{g}_p^h = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_K)^T \mathbf{x}_p^h \quad (2)$$

동작 h는 고유공간 내에서 P개의 점들로 나타나며, 전체 H개의 다중동작에 대해서 고유치와 고유벡터를 계산함으로써 전역고유공간을 이용한다.

한편 $N \times P$ 개 동작영상들로 구성된 행렬 \mathbf{X}^h 는 다음의 식 (3)과 같이 나타낼 수 있다.

$$\mathbf{X}^h = (\mathbf{x}_1^h, \mathbf{x}_2^h, \dots, \mathbf{x}_P^h) \quad (3)$$

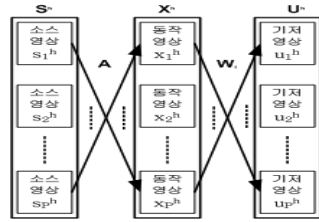
여기서 특별한 포즈의 집합으로 구성된 동작행렬 \mathbf{X}^h 에 대한 고유공간이 생성될 수 있으며, 각각의 포즈들은 생성된 고유공간 내에서 표현될 수 있다. 따라서 각 동작영상 행렬을 이용한 공분산행렬 \mathbf{C}^h 는 $\mathbf{C}^h = \mathbf{X}^h \mathbf{X}^{hT}$ 로 정의되며, 행렬 \mathbf{C}^h 의 고유벡터 \mathbf{e}^h 에 의한 고유치 λ^h 는 고유방정식 $\mathbf{C}^h \mathbf{e}^h = \lambda^h \mathbf{e}^h$ 을 이용하여 계산될 수 있다. 여기서도 \mathbf{C}^h 의 모든 공간은 PCA에 의해서 감소될 수 있으며, 고유벡터에 의해 정의되는 K 차원의 공간은 고유공간이다. 또한 영상 \mathbf{x}_p^h 는 다음의 식 (4)에 의해 고유공간 내에서 고유영상 \mathbf{g}_p^h 로 투영된다.

$$\mathbf{g}_p^h = (\mathbf{e}_1^h, \mathbf{e}_2^h, \dots, \mathbf{e}_K^h)^T \mathbf{x}_p^h \quad (4)$$

이렇게 h개의 동작영상 각각에 대한 고유치와 고유벡터를 계산함으로써 국부고유공간을 얻을 수 있다.

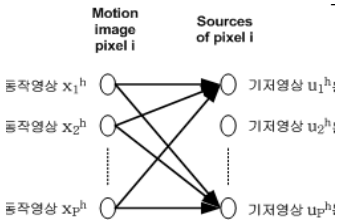
본 논문에서는 전체 동작영상에 대한 전역고유공간을 이용한 것이 아니라 각각의 동작영상에 대한 국부고유공간을 추출하여 ICA의 입력으로 이용하였다. 이렇게 하면 동작영상의 국부적인 특징을 더욱 더 잘 추출할 수 있다. 하지만 고유벡터로부터 추출된 고유공간은 동작의 2차원적 특징들을 반영한 것으로 고차원의 독립적인 국부기저 특징들을 이용하면 보다 개선된 인식성능을 얻을 수 있을 것이다. 따라서 영상 내에 포함된 고차원의 독립특징을 추출하기 위해서 ICA를 이용한다[7].

ICA는 은닉신호의 분리(blind source separation : BSS)나 특징추출에 널리 이용된다[7]. 여기에서는 동작영상을 위한 통계적으로 독립인 기저영상을 추출하는데 이용한다. (그림 1)은 선형 영상합성 모델을 나타낸 것이다.



(그림 1) 선형 영상합성 모델

여기서 동작에 대한 소스영상 \mathbf{s}^h 는 영상들의 근거들(causes)이며, 동작영상 \mathbf{x}^h 는 관찰된 동작영상 패치들이고, \mathbf{u}^h 는 독립된 출력으로 기저영상이다. 또한 \mathbf{A} 와 \mathbf{W}_I 각각은 혼합행렬과 역혼합행렬이며, \mathbf{A} 의 행들은 기저벡터로 구성된 기저합수이다. 따라서 영상패치 \mathbf{x}^h 로부터 ICA를 이용하여 $\mathbf{W}_I (\mathbf{W}_I = \mathbf{A}^{-1})$ 를 구하면 $\mathbf{W}_I * \mathbf{x}^h = \mathbf{u}^h$ 가 출력되고, 이는 동작영상의 독립기저영상이다. 이들 기저영상은 통계적 독립인 동작 특징으로 이는 각 특징 내에서 픽셀은 다른 특징내의 픽셀로부터 예측할 수 없다는 것을 의미한다. 이때 영상 내에 존재하는 통계적으로 독립인 계수로 구성된 기저영상을 추출하기 위해 이용되는 ICA는 (그림 2)와 같이 구성된다. 여기서 ICA를 위한 입력행렬의 행은 다른 동작영상을 나타내고, 열은 다른 픽셀을 나타내는 데이터 행렬을 구성한다.



(그림 2) 독립기저영상 추출을 위한 구조

(그림 2)에서 PCA에 의해 K개의 주요성분을 포함하는 고유영상 행렬 \mathbf{g}_p^h 로에 기초를 두고 ICA를 수행하면 $\mathbf{W}_i^* \mathbf{g}_p^{hT} = \mathbf{u}^h$ 이고, 다시 $\mathbf{g}_p^{hT} = \mathbf{W}_i^{-1} \mathbf{W}_i = \mathbf{A}^{-1} \mathbf{u}^h$ 로 구할 수 있다.

일반적으로 ICA를 좀 더 효과적으로 수행하기 위해서 혼합신호나 영상패치를 대상으로 영 평균과 whitening의 과정을 각각 수행한다[7]. 이는 ICA를 수행하기 전에 미리 영상신호의 1차와 2차의 통계성을 고려한 정규화로 계산을 간략화하기 위함이다. 먼저, 영상의 영 평균은 영상벡터 \mathbf{x}^h 에서 평균값 \mathbf{x}^{ht} 를 뺀 차로 $\mathbf{x}^h = \mathbf{x}^h - \mathbf{x}^{ht}$ 이다. 또한 whitening은 영상벡터 \mathbf{x}^h 의 공분산행렬이 단위행렬이 되도록 함으로써 얻어진다. 즉, $E(\mathbf{x}^h \mathbf{x}^{hT}) = \mathbf{I}$ 가 되도록 하기 위해 PCA가 널리 사용되며, 이때 PCA whitening 행렬 \mathbf{V} 는 $\mathbf{V} = \mathbf{D}^{(-1/2)} \mathbf{E}^T$ 와 같다. 여기서 \mathbf{D} 는 고유치 λ 로 구성되는 대각행렬이며, \mathbf{E} 는 고유벡터 \mathbf{e} 로 구성되는 직교행렬이다. 특히 PCA를 이용하면 신호의 차원인 주요성분의 개수를 조절할 수 있어 ICA의 계산을 좀 더 간단하게 할 수 있다.

한편 FP 알고리즘은 신경망이 가지는 병렬성과 분산성, 그리고 더 작은 메모리 요구 등의 제약을 해결하기 위해 제안된 batch mode로 계산되는 ICA 기법이다[7]. 특히 FP 알고리즘은 엔트로피 최적화 방법으로부터 유도되며, 지금까지 알려진 가장 빠른 학습속도의 기법으로 신호 내에 포함된 상호정보를 최소화함으로써 ICA의 해를 구하는 기법이다. 신호벡터 \mathbf{x}^h 의 상관행렬 $E(\mathbf{x}^h \mathbf{x}^{hT}) = \mathbf{I}$ 로 whitening되어 있다고 가정할 때, 근사화된 반복기법의 역혼합행렬 \mathbf{W}_i 를 구하는 뉴우턴법은 다음 식 (5)와 같다.

$$\begin{aligned} \mathbf{W}_i^* &= \mathbf{W}_i - [E(\mathbf{x}^h \mathbf{f}'(\mathbf{W}_i^T \mathbf{x}^h)) - \beta \mathbf{W}_i] / [E(\mathbf{f}'(\mathbf{W}_i^T \mathbf{x}^h)) - \beta] \\ \mathbf{W}_i^* &= \mathbf{W}_i^* / \|\mathbf{W}_i^*\| \end{aligned} \quad (5)$$

여기서 \mathbf{W}_i^* 는 \mathbf{W}_i 의 새롭게 경신된 값이고, $\beta = E(\mathbf{W}_i^T \mathbf{x}^h \mathbf{f}'(\mathbf{W}_i^T \mathbf{x}^h))$ 이다. 결국 식 (5)는 뉴우턴법에 기초를 둔 ICA를 위한 FP 알고리즘이다. 또한 식 (5)의 상기 식 양쪽에 $\beta - E(\mathbf{f}'(\mathbf{W}_i^T \mathbf{x}^h))$ 를 곱해 구해지는 더욱 간단해진 뉴우턴법의 FP 알고리즘은 다음의 식 (6)과 같다.

$$\begin{aligned} \mathbf{W}_i^* &= E(\mathbf{x}^h \mathbf{f}'(\mathbf{W}_i^T \mathbf{x}^h)) - E(\mathbf{f}'(\mathbf{W}_i^T \mathbf{x}^h)) \mathbf{W}_i \\ \mathbf{W}_i^* &= \mathbf{W}_i^* / \|\mathbf{W}_i^*\| \end{aligned} \quad (6)$$

위의 경신식에서 $\mathbf{f}(\cdot)$ 는 비선형 함수이며 일반적으로 $(\cdot)^3$ 과 $\tanh(\cdot)$ 의 함수값을 가진다. 본 연구에서는 $\tanh(\cdot)$ 함수를 이용하였다.

따라서 각 동작영상을 대상으로 공분산행렬을 이용하여 국부고유공간을 구하고, 이를 식 (6)의 뉴우턴법에 기초한 FP 알고리즘으로 (그림 2)의 구조를 가진 ICA를 수행하면 통계적으로 독립인 국부기저영상을 추출할 수 있다. 따라서 제안된 인식과정의 학습과정에서는 먼저, 획득된 수화영상으로부터 국부고유영상을 계산한 후, 이를 이

용하여 FP-ICA를 수행하여 역혼합행렬 \mathbf{W}_i 를 구한다. 마지막으로 학습된 역혼합행렬 \mathbf{W}_i 를 이용하여 계산된 학습영상의 국부기저영상을 저장한다. 한편 분류과정에서도 학습과정과 동일하게 시험영상을 대상으로 국부고유영상을 계산한다. 다음은 이미 학습과정에서 얻어진 역혼합행렬 \mathbf{W}_i 를 이용하여 시험영상의 국부기저영상을 계산하고, 저장된 기저영상과 비교하여 수화를 인식하게 된다. 이때 분류의 척도로 Euclidean의 거리 척도를 이용하였다.

3. 실험 및 결과분석

독립인 국부기저영상을 이용한 제안된 동작인식 방법의 성능을 평가하기 위해 240*215 픽셀의 80(1명*5동물*16동작)개 동물을 표현하는 수화동작을 대상으로 실험하였다. 여기서 5종류의 동물은 곰, 소, 사슴, 오리, 코끼리이며, 이들 각각을 표현하기 위한 연속 수화동작으로 구성되었으나 0.1초 마다 표본화한 16개의 동작프레임을 실험에 이용하였다. 각 동물의 수화동작 프레임 16중에서 홀수 프레임 (1, 3, 5, ..., 15) 수화영상은 학습에, 나머지 짝수 프레임 (2, 3, 6, ..., 16) 수화영상은 시험에 이용하였다. 또한 분류를 위한 척도로는 Euclidean 거리를 이용하였으며, 실험은 펜티엄IV-3.0G 컴퓨터에서 Matlab 7.1로 구현하였다.

(그림 3)은 5종류 동물에 대한 학습과 시험 수화영상을 나타낸 것이다. 여기서 상단에서 하단으로 곰, 소, ..., 코끼리를 나타내고, 좌측에서 우측으로 각각 프레임 1부터 15, 또는 프레임 2부터 16까지를 나타낸다. 시험 수화영상에서 곰의 2와 소 및 오리의 10 프레임은 위치가, 그리고 오리의 4와 코끼리의 6번 프레임은 크기가 각각 변화된 수화영상이다.



(그림 3) 5가지 학습 및 시험 수화영상

<표 1>은 고유영상을 7개로 하였을 때, PCA에 의해 추출된 국부고유영상을 이용한 방법과 뉴우턴법의 FP-ICA에 의해 추출된 국부기저영상을 이용한 방법 각각에 의한 결과이다. 여기서 학습영상과 시험영상 사이의 유사성을 측정하기 위해 Euclidean의 거리 척도를 이용하였다. 또한 평균오차는 올바르게 인식된 영상을 대상으로 학습 및 시험 수화영상에서 각각 고유영상 및 기저영상 사

이의 오차의 절대치 합의 평균을 나타낸 것이다. <표 1>에서 보면 2가지 방법에서 이미 학습 수화영상에 대한 인식은 3가지 척도 모두 100%의 인식을 보이거나, 시험 수화영상에 대해서는 국부고유영상의 경우 약 68.75%의 인식을 나타낸다. 하지만 제안된 방법에서는 약 78.75%의 인식을 나타낸다. 따라서 FP-ICA에 의한 국부기저영상을 이용하는 방법은 PCA에 의한 국부고유영상을 이용하는 방법에 비해 평균적으로 약 1.15배 정도 개선된 인식을 가진다. 또한 학습과 시험의 수화영상 각각에 대하여 국부기저영상은 국부고유영상에 비해 평균오차에서 평균적으로 훨씬 작은 값을 가진다. 이는 국부기저영상을 이용하는 방법이 국부고유영상을 이용하는 방법보다 분류과정에서 좀 더 정확하게 수화를 식별함을 보여 준 것이다.

<표 1> 국부고유 및 국부기저 영상에 대한 인식결과

입력영상	국부고유영상(PCA)		국부기저영상(FP-ICA)	
	인식률(%)	평균오차	인식률(%)	평균오차
학습영상	100	4.63e-012	100	0.59e-014
시험영상	68.75	1858.58	78.75	1.98

4. 결론

본 논문에서는 동작영상 각 개인의 국부고유공간에 바탕을 둔 기저영상을 이용한 효율적인 수화영상 인식 기법을 제안하였다. 여기서 국부고유공간의 추출은 주요성분 분석을 이용한 것으로 동작영상의 국소특징을 더욱 더 잘 반영하기 위함이고, 기저영상의 추출은 독립성분분석을 이용한 것으로 수화영상 내에 포함된 고차원의 독립 특징들을 반영하여 보다 개선된 인식성능을 얻기 위함이다.

제안된 기법을 240*215 픽셀의 80(1명*5동물*16동작)개 동물을 표현하는 수화동작을 대상으로 Euclidean의 분류 척도를 이용하여 실험한 결과, 단순 국부고유공간을 이용한 방법에 비해 우수한 인식성능이 있음을 확인하였다.

향후 제안된 방법을 좀 더 큰 규모의 다양한 인식문제에 적용하는 연구가 지속적으로 이루어져야 할 것이다. 이루어져야 할 것이다.

참고문헌

[1]Y. Wu and T. S. Huang, "Vision Based Gesture Recognition: A Review," *LNCS: Gesture-Based Communication in Human-Computer Interaction: International Gesture Workshop*, Vol. 1739, 2004.

[2] A. Mulder, "Hand Gestures for Hci," *Technical Report 96-1*, Simon Fraster University, 1996.

[3] 박호식, 배철수, "수화인식을 위한 얼굴과 손 추적알고리즘," *한국통신학회논문지*, 제 31권 11C호, 2006년 11월

[4] 김희곤, 오형철, 양윤모, "수화인식을 위한 특징량 추출하드웨어 설계," *대한전자공학회 추계학술대회 논문집*, 제 18권 2호, 1995년 12월

[5] 조재현, 업성용, "다중마커를 이용한 수화 인식기법," *서울여자대학교대학원 논문집*, 제 1998권 6호, 1998년

[6] M. H. Yang, D. Kriegman, and N. Ahuja, "Detecting Faces in Images: A Survey," *IEEE. Trans. on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 1, pp.64-58, Jan. 2002

[7] A. Hyvaerinen, J. Karhunen, and E. Oja, 'Independent Component Analysis', John Wiley & Sons, Inc., New York, 2001