

독립성분분석을 이용한 국부기저영상 기반 동작인식

Motion Recognitions Based on Local Basis Images Using Independent Component Analysis

조용현
Yong-Hyun Cho

대구가톨릭대학교 공과대학 컴퓨터정보통신공학부
School of Computer and Information Comm. Eng., Catholic Univ. of Daegu
e-mail : yhcho@cu.ac.kr

요 약

본 논문에서는 중심이동과 국부기저영상을 이용한 동작인식 기법을 제안하였다. 여기서 중심이동은 1차 모멘트 평형에 기반을 둔 것으로 위치나 크기 변화에 강건한 동작영상을 얻기 위함이고, 국부기저영상의 추출은 독립성분분석 기법에 기반을 둔 것으로 각 동작들마다에 포함된 통계적으로 독립인 동작특징들의 집합을 얻기 위함이다. 특히 국부기저영상을 빠르게 추출하기 위해 뉴턴(Newton)법의 고정점 알고리즘에 기반을 둔 독립성분분석을 이용하였다. 제안된 기법을 240*215 픽셀의 160(1명*10종류*16동작)개 동물표현의 수화 동작영상을 대상으로 city-block, Euclidean, 그리고 negative angle의 척도들을 분류척도로 이용하여 실험하였다. 실험결과, 제안된 기법은 국부기저영상을 이용한 방법과 중심이동을 거치지 않는 국부기저영상을 이용하는 기법보다 각각 우수한 인식성능이 있음을 확인하였다.

키워드 : 독립성분분석, 고정점 알고리즘, 중심이동, 동작인식, 국부기저영상

ABSTRACT

This paper presents a human motion recognition method using both centroid shift and local basis images. The centroid shift based on 1st moment balance technique is applied to get the robust motion images against position or size changes, the extraction of local basis images based on independent component analysis(ICA) is also applied to find a set of statistically independent motion features, which is included in each motions. Especially, ICA of fixed-point(FP) algorithm based on Newton method is used for being quick to extract a local basis images of motions. The proposed method has been applied to the problem for recognizing the 160(1 person * 10 animals * 16 motions) sign language motion images of 240*215 pixels. The 3 distances such as city-block, Euclidean, negative angle are used as measures when match the probe images to the nearest gallery images. The experimental results show that the proposed method has a superior recognition performances(speed, rate) than the method using local eigen images and the method using local basis images without centroid shift respectively.

Key Words : Independent component analysis, Fixed-point algorithm, Centroid shift, Motion recognition, Local basis image

1. 서 론

인간의 동작인식은 컴퓨터 비전 등의 분야에서 중요한 요소기술이다. 특히 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[1-8]. 하지만 동작은 시작하고 끝나는 시점을 정확하게 구분하지 못함과 동일한 사람이 동일한 동작을 취하더라도 모양, 시간, 궤적이 변하기 때문에 생기는 시공간적 다

양성의 특징을 가지고 있어 인식에는 많은 제약이 있다 [1-4].

최근 손동작으로 이루어지는 몸짓인식이 인간과 컴퓨터의 상호작용으로 이용되는 한 가지 수단으로 대두되고 있다. 특히 컴퓨터를 이용한 인간의 몸짓인식은 장애인들로 하여금 컴퓨터로의 접근성을 용이하게 하고 있다. 이러한 인식에는 카메라와 같은 비전장치를 이용한 인간 친화적이면서도 구현이 간단한 비전기반의 시스템들이 주로 이용되고 있다[1-3]. 하지만 여기에도 카메라조절, 영상분할, 특징추출, 그리고 식별 등과 같은 해결해야할 어려운 문제들이 있다.

인간과 컴퓨터를 연결하기 위한 시각 인터페이스 기술로 입력 측에서는 정밀한 센싱에 대한 연구가 주로 진행 중이며, 주로 얼굴 및 제스처 인식에 대한 연구가 활발히 진행되고 있다[1-8]. 얼굴인식 관련기술로 MIT media lab에서

접수일자 : 2008년 7월 7일

완료일자 : 2008년 8월 26일

본 연구는 2007년 대구가톨릭대학교 연구년에 의한 것임

는 얼굴특징을 추출하여 정규화한 후 고유치(eigenvalue)의 변화를 통해 얼굴을 인식하는 엔진을 개발하고 있다[7,8]. 또한 eTrue나 Visionics 등의 기업체에서는 동영상에서 얼굴영역 분리 및 인증 기술과 국부적인 특징분석에 의한 인식 알고리즘을 개발하고 있다. 특히 미국의 DARPA에서는 20년 이상 영상이해 프로그램을 진행하고 있으며, 영상인식 및 이동물체 추적에서 상당부분 실용화에 가까운 기술을 확보하고 있다[5].

국내에서도 연구소와 대학을 중심으로 시각 인터페이스를 위한 입체영상 및 신호처리 기술에 대한 연구들을 진행하고 있다. 국내에서는 주로 입체영상 기술에 대한 연구나 양안 및 단안식 입체 영상 신호처리 관련연구 및 홀로그램에 관한 기술들이 진행되고 있다. 또한 손동작 인식을 위한 사용자 인터페이스나 신경망 기반의 알고리즘 등에 대한 연구들이 활발하게 이루어지고 있다[4-6].

한편 수화는 청각 장애인이 의사전달을 위해 사용하는 동작언어이다[4-6]. 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[1,2].

일반적으로 손동작이나 표정 및 얼굴 등의 인식에서 특징추출은 인식성능을 결정짓는 매우 중요한 파라미터로 특징기반 방법과 영상기반 방법이 이용되고 있다[7-9]. 특징기반 방법은 특징과 관련된 정보들을 이용하는 것으로 지역적 특징추출, 2차원의 템플릿을 이용하는 템플릿 패턴정합, 특성을 벡터화한 Gabor 필터를 이용한 방법들이 있다[8]. 이들 방법들에서는 차원감소나 수행시간의 절약, 포즈나 크기 등에 강건한 특성이 있으나 인식에 필요한 정보손실이나 템플릿 구성에 따른 공간 소모, 변형의 정도에 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)[9-11]과 독립성분분석(independent component analysis : ICA)을 이용한 기법들이 주로 이용되고 있다[12,13]. PCA는 영상 속에 포함된 진폭을 고려한 2차원적 속성을 이용하나 ICA는 이를 일반화한 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 유도하는데 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 cumulant나 moment를 구하는 등 복잡한 계산이 요구되어 계산시간이 오래 걸린다. 이러한 제약을 해결하기 위해서 최근 고정점(fixed-point : FP)이나 기울기를 이용하는 알고리즘들이 제안되었다[12,13].

본 논문에서는 중심이동과 차원감소의 전처리 및 국부기저영상을 이용한 동작인식 기법을 제안하였다. 여기서 전

처리는 1차 모멘트 평형[14]에 기반을 둔 것으로 중심이동과 차원감소에 따른 위치나 크기 변화에 강건한 동작영상을 얻기 위함이다. 국부기저영상의 빠른 추출을 위해 뉴턴(Newton)법[15]에 기초한 FP-ICA를 이용하였으며, 특히 국부기저영상의 추출은 각 동작들마다에 포함된 통계적으로 독립인 동작특징들을 얻기 위함이다. 제안된 기법을 240*215 픽셀의 160(1명*10동물*16동작)개 동물표현의 수화동작영상을 대상으로 city-block, Euclidean, 그리고 negative angle의 척도들을 분류적으로 이용하여 실험하여 그 결과를 국부고유영상을 이용한 방법과 중심이동을 거치지 않는 국부기저영상을 이용하는 기법의 결과와 각각 비교·고찰하였다.

2. 전처리와 국부기저영상의 추출

인식의 성능과 효율을 높이기 위해 수행하는 전처리에는 기하학적정보들에 대한 정규화 기법과 화소들에 직접적인 효과를 갖는 필터링 기법들이 있다[10,11]. 일반적으로 동작영상은 배경과 의미 있는 형상(Shape)으로 구분되며, 형상에 비해 배경의 변화가 비교적 적을 경우에 원래보다 작은 크기의 유효영상을 추출할 수 있다. 이렇게 추출된 동작영상을 이용하면 보다 빠르고 정확하게 인식이 가능하다.

여기에서는 1차 모멘트 평형[14]를 이용한 중심을 계산하고 이동시켜 유효영상을 추출한다. 이는 기하학적 정보를 이용한 전처리로 배경을 제거함으로써 차원의 감소를 얻을 수 있다. 또한 전처리된 동작영상을 대상으로 독립적인 국부기저영상을 빠르게 추출하기 위해 뉴턴법[15]에 기초한 고정점 알고리즘의 독립성분분석을 이용한다.

2.1 중심이동과 차원감소의 전처리

입력되는 손짓이나 몸짓의 동작영상을 X축 과/또는 Y축으로 평행 이동시켜 이루어지는 중심이동은 1차 모멘트 평형[14]에 의해 중심을 계산함으로써 가능하다. 여기서 1차 모멘트를 구하는 과정은 먼저 원 영상 OrgImage로부터 차 영상 DeltaImage를 생성하여야 한다. 이때 n번째 좌표의 차 영상 $\Delta Image(n) = P_i(n) - P_i(n-1)$ 로 계산되며, Y행 화소좌표로부터 X열 화소좌표 방향으로 수행한다. 그림 1은 입력되는 m*n 픽셀의 동작영상과 계산된 차 영상 DeltaImage를 각각 나타낸 것이다.

$P_1(0)$...	$P_1(n-1)$	$P_1(n)$	0	$P_1(n)-P_1(n-1)$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
$P_i(0)$...	$P_i(n-1)$	$P_i(n)$	0	$P_i(n)-P_i(n-1)$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
$P_m(0)$...	$P_m(n-1)$	$P_m(n)$	0	$P_m(n)-P_m(n-1)$

그림 1. m*n 픽셀의 원 영상과 차 영상

Fig. 1. Original and difference image of m*n pixels

다음은 계산된 차 영상으로부터 유효한 영상의 화소만을 생성하기 위하여 논리 필터링을 수행한다. 이때 유효한 변화의 판별을 위한 값은 차 영상에 대한 화소의 변화평균값을 구하여 이용한다. 논리필터링은 차 영상의 값이 판별값보다 크거나 같으면 유효한 화소로 1의 값을 갖도록 하며, 그렇지 않으면 0의 값을 갖도록 함으로써 균일한 형상의 영

상을 생성한다. 한편 그림 2는 논리 필터링된 동작영상을 대상으로 1차 모멘트 평형을 구하기 위한 임의의 동작영상 단면을 나타낸 것이다.

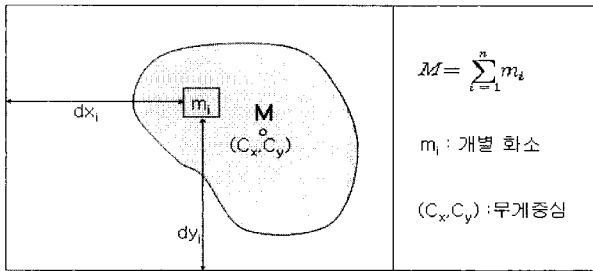


그림 2. 1차 모멘트 평형을 위한 동작영상 단면
 Fig. 2. Motion image area for 1st moment balance

여기서 1차 모멘트 평형은 각 개별 화소에 대한 모멘트 값의 합이 영상 전체의 중심거리와 영상 전체의 화소 값의 곱과 동일함을 의미하며[14], 다음의 식 (1)과 같이 표현된다.

$$M \times C_x = \sum_{i=1}^n (m_i \times dx_i) \quad M \times C_y = \sum_{i=1}^n (m_i \times dy_i) \quad (1)$$

여기서 \$m_i\$는 \$i\$번째 개별화소의 논리값으로 유효한 화소일 경우에는 1, 그렇지 않으면 0이며, \$M\$은 유효영상에 대한 개별화소의 논리값 합이다. 또한 동작영상의 무게중심으로 나타나는 유효영상의 중심좌표 \$(C_x, C_y)\$는 식 (2)와 같이 계산된다.

$$C_x = \frac{\sum_{i=1}^n dx_i}{M} \quad C_y = \frac{\sum_{i=1}^n dy_i}{M} \quad (2)$$

따라서 1차 모멘트 평형으로부터 얻어진 유효영상의 중심좌표 \$(C_x, C_y)\$를 기준으로 보다 작은 영상을 선택하여 입력영상으로 이용한다. 즉, 임의로 입력되는 원 동작영상을 중심좌표로 이동시켜 유효한 동작영상을 선택하면 인식에 불필요한 배경을 배제시킬 수 있어 차원이 감소된 영상을 얻을 수 있다. 이렇게 하면 동작영상의 특징인 기저영상을 추출하는데 이용되는 FP-ICA의 입력데이터를 줄일 수 있어 인식성능을 개선시킬 수 있다.

2.2 독립적인 국부기저영상의 추출

동작은 연속적으로 변하는 포즈의 집합이다. \$M_0 * N_0 (=N)\$ 픽셀을 가지는 \$P\$개 연속적으로 표본화된 정규화된 동작영상 \$x_p(p=1, 2, \dots, P)\$에 대해서 살펴보자. 만약 한 사람을 대상으로 \$H\$개의 동작을 가진다면, 동작 \$h\$에 대한 영상은 \$x_p^h(h=1, 2, \dots, H)\$로 나타낼 수 있다. 각각의 동작에 대해서 영상의 스트림은 \$P\$개씩 표본화되며, 이들 \$N*PH\$개 동작영상들로 구성된 단일행렬 \$X\$는 다음의 식 (3)과 같다.

$$X = (x_1^1, x_2^1, \dots, x_P^1, x_1^2, x_2^2, \dots, x_P^2, \dots, x_1^H, x_2^H, \dots, x_P^H) \quad (3)$$

식 (3)의 전체 동작영상 행렬을 이용한 공분산(covariance) 행렬 \$C\$는 \$C = XX^T\$로 정의되며, 여기서 \$T\$는 정치행렬을 나타낸다. 행렬 \$C\$의 고유벡터 \$e_j\$에 의한 고유치 \$\lambda_j\$는 고유방정식 \$Ce_j = \lambda_j e_j\$을 이용하여 계산될 수 있다. 이때 \$C\$의 모든 고유벡터에 의해서 정의되는 \$N\$차원 공간은

PCA에 의해서 감소될 수 있다. 결과로 만약 \$K(1 \le K \ll N)\$개의 고유치와 고유벡터가 선택된다면, 고유벡터에 의해 정의되는 \$K\$차원의 공간은 고유공간이다. 또한 영상 \$x_p^h\$는 다음의 식 (4)에 의해 고유공간 내에서 고유영상 \$g_p^h\$로 투영된다.

$$g_p^h = (e_1, e_2, \dots, e_K)^T x_p^h \quad (4)$$

동작 \$h\$는 고유공간 내에서 \$P\$개의 점들로 나타나며, 전체 \$H\$개의 다중 동작에 대해서 고유치와 고유벡터를 계산함으로써 전역고유공간을 이용한다.

한편 \$N*P\$개 동작영상들로 구성된 행렬 \$X^h\$는 다음의 식 (5)과 같이 나타낼 수 있다.

$$X^h = (x_1^h, x_2^h, \dots, x_P^h) \quad (5)$$

여기서 특별한 포즈의 집합으로 구성된 동작행렬 \$X^h\$에 대한 고유공간이 생성될 수 있으며, 각각의 포즈들은 생성된 고유공간 내에서 표현될 수 있다. 따라서 각 동작영상 행렬을 이용한 공분산행렬 \$C^h\$는 \$C^h = X^h X^{hT}\$로 정의되며, 행렬 \$C^h\$의 고유벡터 \$e^h\$에 의한 고유치 \$\lambda^h\$는 고유방정식 \$C^h e^h = \lambda^h e^h\$을 이용하여 계산될 수 있다. 여기서도 \$C^h\$의 모든 공간은 PCA에 의해서 감소될 수 있으며, 고유벡터에 의해 정의되는 \$K\$차원의 공간은 고유공간이다. 또한 영상 \$x_p^h\$는 다음의 식 (6)에 의해 고유공간 내에서 고유영상 \$g_p^h\$로 투영된다.

$$g_p^h = (e_1^h, e_2^h, \dots, e_K^h)^T x_p^h \quad (6)$$

이렇게 \$h\$개의 동작영상 각각에 대한 고유치와 고유벡터를 계산함으로써 국부고유공간을 얻을 수 있다. 본 논문에서는 전체 동작영상에 대한 전역고유공간을 이용한 것이 아니라 각각의 동작영상에 대한 국부고유영상을 추출하여 ICA의 입력으로 이용하였다. 이렇게 하면 동작영상의 국부적인 특징을 더욱 더 잘 추출할 수 있다. 하지만 고유벡터로부터 추출된 고유영상은 동작의 2차원적 특징들을 반영한 것으로 고차원의 독립적인 국부기저특징들을 이용하면 보다 개선된 인식성능을 얻을 수 있을 것이다. 따라서 영상 내에 포함된 고차원의 독립특징을 추출하기 위해서 ICA를 이용한다[12,13].

ICA는 은닉신호의 분리(blind source separation : BSS)나 특징추출에 널리 이용된다[12,13]. 여기에서는 동작영상을 위한 통계적으로 독립인 기저영상을 추출하는데 이용한다. 그림 3은 선형 영상합성 모델을 나타낸 것이다.

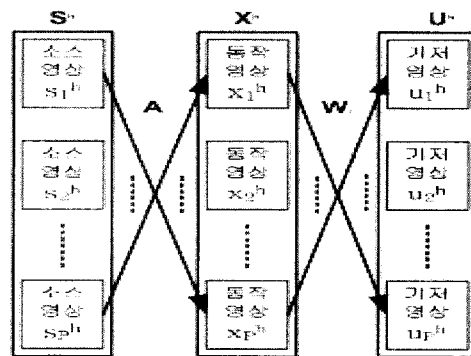


그림 3. 선형 영상합성 모델
 Fig. 3. Linear images synthesis model

여기서 동작에 대한 소스영상 s^h 는 영상들의 근거들 (causes)이며, 동작영상 x^h 는 관찰된 동작영상 패치들이고, u^h 는 독립된 출력으로 기저영상이다. 또한 A 와 W_1 각각은 혼합행렬과 역혼합행렬이며, A 의 행들은 기저벡터로 구성된 기저함수이다. 따라서 영상패치 x^h 로부터 ICA를 이용하여 $W_1(W_1 = A^{-1})$ 를 구하면 $W_1 * x^h = u^h$ 가 출력되고, 이는 동작영상의 독립기저영상이다. 이들 기저영상은 통계적 독립인 동작 특징으로 이는 각 특징 내에서 픽셀은 다른 특징 내의 픽셀로부터 예측할 수 없다는 것을 의미한다.

이때 영상 내에 존재하는 통계적으로 독립인 계수로 구성된 기저영상을 추출하기 위해 이용되는 ICA는 그림 4와 같이 구성된다. 여기서 ICA를 위한 입력행렬의 행은 다른 동작영상을 나타내고, 열은 다른 픽셀을 나타내는 데이터 행렬을 구성한다.

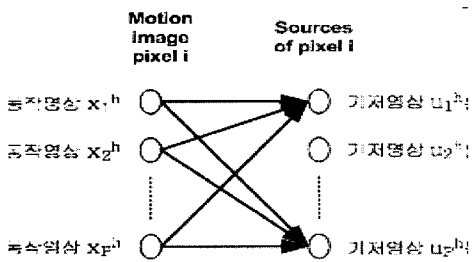


그림 4. 독립기저영상 추출을 위한 구조

Fig. 4. Topology for extracting independent basis images

그림 4에서 PCA에 의해 K개의 주요성분을 포함하는 고유영상 행렬 g_p^h 로에 기초를 두고 ICA를 수행하면 $W_1 * g_p^{hT} = u^h$ 이고, 다시 $g_p^{hT} = W_1^{-1} W_1 = A^{-1} u^h$ 로 구할 수 있다.

일반적으로 ICA를 좀 더 효과적으로 수행하기 위해서 혼합이나 영상패치를 대상으로 영 평균과 whitening의 과정을 각각 수행한다[12,13]. 이는 ICA를 수행하기 전에 미리 영상신호의 1차와 2차의 통계성을 고려한 정규화로 계산을 간략화하기 위함이다. 먼저, 영상의 영 평균은 영상벡터 x^h 에서 평균값 x^{h*} 를 빼 차로 $x^h = x^h - x^{h*}$ 이다. 또한 whitening은 영상벡터 x^h 의 공분산행렬이 단위행렬이 되도록 함으로써 얻어진다. 즉, $E\{x^h x^{hT}\} = I$ 가 되도록 하기 위해 PCA가 널리 사용되며, 이때 PCA whitening 행렬 V 는 $V = D^{(-1/2)} E^T$ 와 같다. 여기서 D 는 고유치 λ 로 구성되는 대각행렬이며, E 는 고유벡터 e 로 구성되는 직교행렬이다. 특히 PCA를 이용하면 신호의 차원인 주요성분의 개수를 조정할 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[13]. 특히 FP 알고리즘은 엔트로피 최적화 방법으로부터 유도되며, 지금까지 알려진 가장 빠른 학습속도의 기법으로 신호 내에 포함된 상호정보를 최소화함으로써 ICA의 해를 구하는 기법이다. 신호벡터 x^h 의 상관행렬 $E\{x^h x^{hT}\} = I$ 로 whitening 되어 있다고 가정할 때, 근사화된 반복기법의 역혼합행렬 W_1 를 구하는 뉴턴법은 다음 식 (7)과 같다.

$$W_1^i = W_1 - [E\{x^h f(W_1^T x^h)\} - \beta W_1] / [E\{f'(W_1^T x^h)\} - \beta],$$

$$W_1^* = W_1^i / \|W_1^i\| \quad (7)$$

여기서 W_1^* 는 W_1 의 새롭게 경신된 값이고, $\beta =$

$E\{W_1^T x^h f(W_1^T x^h)\}$ 이다. 결국 식 (7)은 뉴턴법에 기초를 둔 ICA를 위한 FP 알고리즘이다. 또한 식 (7)의 상기 식 양쪽에 $\beta - E\{f'(W_1^T x^h)\}$ 를 곱해 구해지는 더욱 간단해진 뉴턴법의 FP 알고리즘은 다음의 식 (8)과 같다.

$$W_1^i = E\{x^h f(W_1^T x^h)\} - E\{f'(W_1^T x^h)\} W_1$$

$$W_1^* = W_1^i / \|W_1^i\| \quad (8)$$

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

따라서 중심이동과 차원감소의 전처리된 각 동작영상을 대상으로 공분산행렬을 이용하여 국부고유영상을 구하고, 이를 식 (8)의 뉴턴법에 기초한 FP 알고리즘으로 그림 4의 구조를 가진 ICA를 수행하면 통계적으로 독립인 국부기저영상을 추출할 수 있다. 따라서 전처리와 국부기저영상을 이용한 제안된 동작인식 과정을 요약하면 그림 5와 같다. 학습과정에서는 먼저, 학습과정으로 획득된 동작영상으로부터 1차 모멘트 평형을 이용하여 중심좌표를 계산하고, 동작영상을 중심이동 시킨 후, 차원이 감소된 유효영상을 검출하여 국부고유영상을 계산한다. 이렇게 전처리된 국부고유영상을 이용하여 FP-ICA를 수행하여 역혼합행렬 W_1 를 구한다. 마지막으로 학습된 역혼합행렬 W_1 를 이용하여 계산된 학습 동작영상의 국부기저영상을 저장한다. 한편 분류과정에서도 학습과정과 동일하게 시험 동작영상을 대상으로 유효한 국부고유영상을 획득한다. 다음은 이미 학습과정에서 얻어진 역혼합행렬 W_1 를 이용하여 시험 동작영상의 국부기저영상을 계산하고, 저장된 기저영상과 비교하여 동작을 인식하게 된다. 이때 분류의 척도로 city-block, Euclidean, negative angle의 거리를 각각 이용하였다.

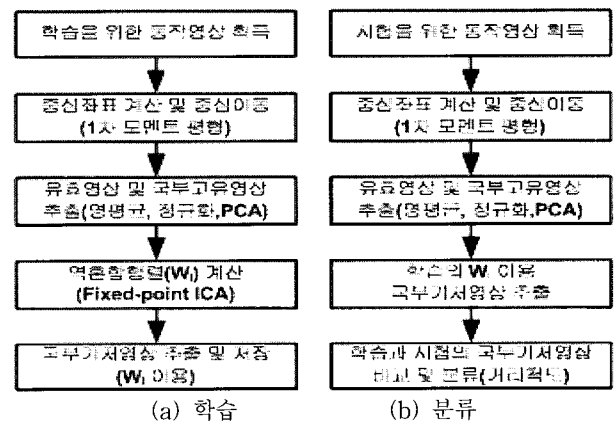


그림 5. 제안된 방법에 의한 인식과정

Fig. 5. Recognition process of proposed method

3. 실험 및 결과분석

중심이동과 차원감소의 전처리와 독립인 국부기저영상을 이용한 제안된 동작인식 방법의 성능을 평가하기 위해 240*215 픽셀의 160(1명*10동물*16동작)개 동물을 표현하는 수화동작을 대상으로 실험하였다. 여기서 10종류의 동물은 곰, 소, 사슴, 오리, 코끼리, 염소, 돼지, 토끼, 호랑이, 늑대이며, 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 city-block, Euclidean, 그리고 negative angle의 거리들을 이용하였다. 또한 특징추출에 의한 복원 성능을 평가하기 위해 원 영상의 픽셀값 x_i 와 복원된 영상의 픽셀값 y_i 사이의 절대평균합오차 $amse(absolute\ mean\ sum\ error)=(1/N)\sum |x_i - y_i|$ 를 정의하였으며, 여기서 N 은 수화영상의 전체 픽셀수이다. 실험은 펜티엄IV 3.0G 컴퓨터에서 Matlab 7.1로 구현하였다.

그림 6은 각각 10종류의 동물 중에서 5종류 동물에 대한 학습과 시험영상만을 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 변환된 수화영상이다.

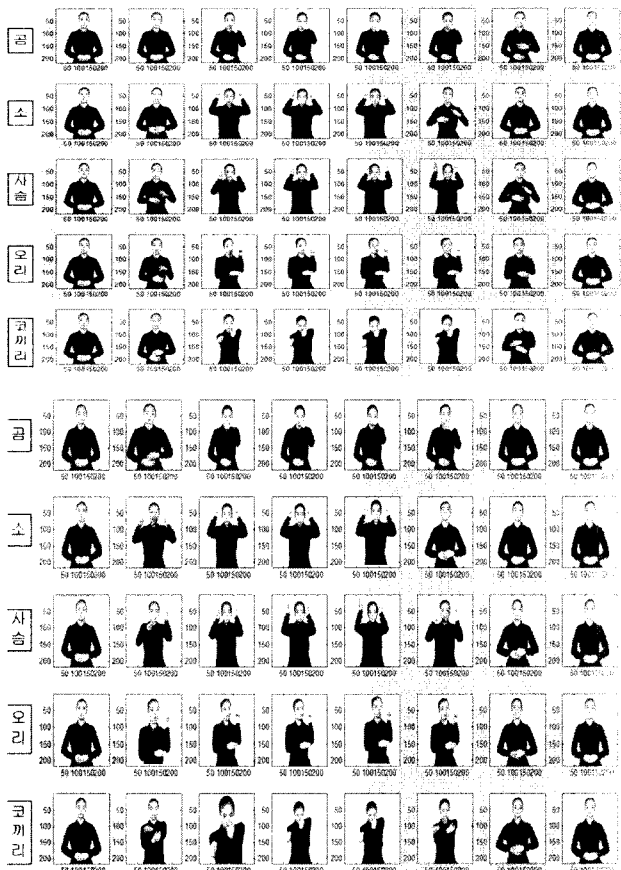
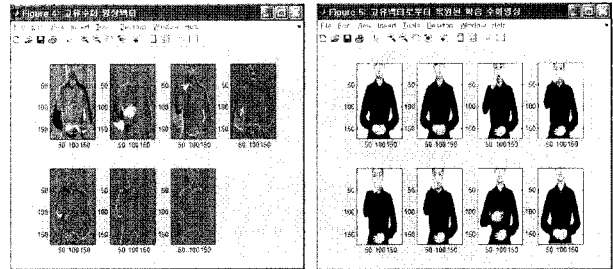


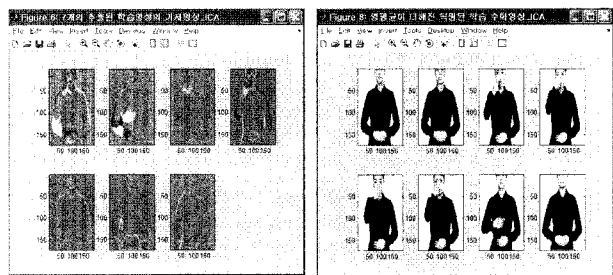
그림 6. 5가지 학습 및 시험 수화영상
Fig. 6. 5 probe and gallery sign language images

한편 그림 7은 전처리된 학습 곰 수화영상을 대상으로 가장 낮은 절대평균합오차를 가진 7개의 국부고유 및 국부기저 영상으로부터 복원된 8개의 수화영상들을 각각 나타낸 것이다. 여기에서는 각각의 원 수화영상으로부터 차원이 감소된 특징영상들을 추출함으로써 데이터의 차원을 감소시킨 압축효과를 갖는다. 육안으로 볼 때, 국부고유나 국부기저 영상 각각으로부터 복원된 학습 수화영상은 원 수화영상과 매우 일치함을 알 수 있다. 이때 복원된 수화영상과

학습 수화영상과의 전체오차는 각각 약 2.9033 과 $1.2312e-012$ 정도이며, 국부기저영상을 특징으로 이용할 때가 국부고유영상을 이용할 때보다도 훨씬 우수한 복원 성능이 있음을 알 수 있다.



(a)



(b)

그림 7. 곰 수화 학습영상의 7개 국부고유 및 복원영상(a)과 7개 국부기저 및 복원영상(b)

Fig. 7. 7 local eigen and reconstructed images(a) and 7 local basis and reconstructed images(b) of bear sign language probe images

표 1은 곰 수화영상 중에서 학습영상 8개를 대상으로 중심이동과 차원감소의 전처리 영상과 그렇지 않은 비전처리 영상을 대상으로 고유영상 및 기저영상의 개수에 따른 복원 성능을 절대평균합오차 $amse$ 로 나타낸 것이다. 표 1에서 보면, 2가지 모두에서 고유나 기저 영상의 수가 증가함에 따라 $amse$ 로 표시되는 복원성능은 개선된다 다시 저하됨을 알 수 있다. 이는 고유나 기저 영상인 특징벡터의 개수가 증가함에 따라 복원성능도 개선되지만 지나치게 증가하면 오히려 과학습을 일으켜 복원성능이 저하된 것으로 추측된다. 한편 전처리 영상은 그렇지 않은 영상에 비해 $amse$ 가 큼을 알 수 있다. 이는 전처리에 의해 유효영상의 일부가 배경과 함께 제거됨으로써 복원에 필요한 정보가 손실되기 때문이다. 또한 PCA에 의한 고유영상을 이용한 복원보다는 ICA에 의한 기저영상을 이용한 복원성능이 우수함을 알 수 있다. 이 또한 ICA가 영상 내에 포함된 고차원의 특징벡터를 추출하기 때문이다. 본 논문에서는 복원성능이 가장 우수한 7개일 때를 기준으로 실험하였다.

한편 그림 8은 위치가 이동된 그림 6의 곰 시험수화 영상 2번을 대상으로 논리 필터링한 후, 1차 모멘트에 의해서 계산된 중심좌표를 이용하여 중심이동 시킨 후, $240*215(51,600)$ 픽셀의 크기를 $192*172(33,024)$ 픽셀로 감소시켜 전처리한 예 영상이다. 여기서 전처리 영상은 원 영상보다 약 1.56배 정도 차원이 감소된 영상으로 원 수화영상에서 인식에 불필요한 배경을 제거하여 입력데이터의 차원을 감소시킴으로써 계산시간과 인식률의 성능을 개선할 수 있다.

표 1. 고유 및 기저영상의 개수에 따른 복원성능

Table 1. Recall performance by changing the number of eigen and basis images

고유영상 수		고유영상의 수								기저영상 수							
		1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
절대 평균 합오차, amse	전처리	75.9264	53.5935	36.9231	25.1348	14.9675	8.3544	2.9033	38.6860	0.9984	0.5438	0.1074	0.0645	0.0031	0.0006	1.2312 e-012	0.1204
	비전처리	41.0377	25.1131	15.4942	10.9719	7.5255	5.1886	0.7483	10.1239	0.6513	0.2454	0.0413	0.0017	0.0007	0.0001	1.0123 e-017	0.0615

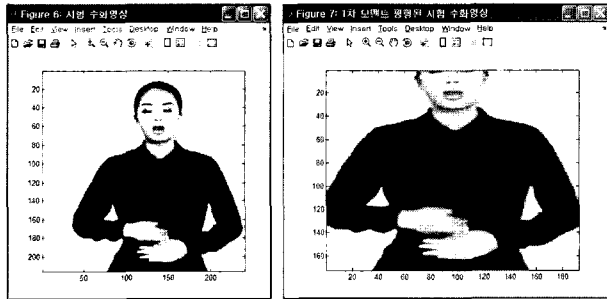


그림 8. 위치가 이동 및 전처리된 된 수화영상

Fig. 8. Position shifted and preprocessed sign language images

표 2는 고유영상과 기저영상을 각각 7개로 하였을 때, 비전처리된 학습 수화영상 80개와 시험 수화영상 80개를 각각 대상으로 거리척도에 따른 인식결과를 나타낸 것이다. 또한 PCA에 의해 추출된 국부고유영상을 이용한 방법과 뉴우턴법의 FP-ICA에 의해 추출된 국부기저영상을 이용한 방법 각각에 의한 결과이다. 여기서 학습영상과 시험영상 사이의 유사성을 측정하기 위해 city-block, Euclidean, negative angle의 3가지 거리척도를 이용하였다. 또한 평균 오차는 올바르게 인식된 영상을 대상으로 학습 및 시험 수화영상에서 각각 고유영상 및 기저영상 사이의 오차의 절대치의 평균을 나타낸 것이다. 표 2에서 보면 2가지 방법에서 이미 학습 수화영상에 대한 인식은 3가지 척도 모두 100%의 인식률을 보이나, 시험 수화영상에 대해서는 city-block, Euclidean, negative angle 각각에 대해 국부고유영상의 경우 약 72.5%, 약 68.75%, 77.5%의 인식률을 나타낸다. 하지만 제안된 방법에서는 각각 약 81.25%, 약 78.75%, 약 86.25%의 인식률을 나타낸다. 따라서 FP-ICA에 의한 국부기저영상을 이용하는 방법은 PCA에 의한 국부고유영상을 이용하는 방법에 비해 평균적으로 약 1.13배 정도 개선된 인식률을 가진다. 또한 학습과 시험의 수화영상 각각에 대하여 국부기저영상은 국부고유영상에 비해 각각 3가지 거리척도에 대한 평균오차에서 평균적으로 훨씬 작은 값을 가진다. 이는 국부기저영상을 이용하는 방법이 국부고유영상을 이용하는 방법보다 분류과정에서 좀 더 정확하게 수화를 식별함을 보여 준 것이다. 한편 80개의 시험 수화 중에서 위치, 각도, 크기, 그리고 위치와 각도와 크기가 동시에 변화된 수화의 경우에 2가지 방법 각각에 대해 오인식되는 결과가 발생하였다. 이는 수화영상의 기하학적 변화가 인식성능에 많은 영향을 미침을 보여 주는 것이다. 한편 negative angle의 거리척도에서는 학습 수화영상과 시험 수화영상이 완전히 일치되는 경우 -1.0의 출력 값을 갖는다. 또한 실험의 결과로부터 유사성 판별을 위한 거리척도 중에서 negative angle이 다른 2가지 거리척도에 비해

학습 및 시험 수화영상에 대한 인식률이 높음도 알 수 있다.

표 2. 국부고유 및 국부기저 영상에 대한 인식결과

Table 2. Recognition results of local eigen images and local basis image

거리척도	입력영상	국부고유영상 (PCA)		국부기저영상 (FP-ICA)	
		인식률 (%)	평균오차	인식률 (%)	평균오차
City-Block (L1-norm)	학습수화	100	6.64e-012	100	1.05e-014
	시험수화	72.5	4796.43	81.25	3.97
Euclidean (L2-norm)	학습수화	100	4.63e-012	100	0.59e-014
	시험수화	68.75	1858.58	78.75	1.98
Negative Angle (Cosine)	학습수화	100	-1.00	100	-1.00
	시험수화	77.5	-0.91	86.25	-0.96

한편 표 3은 역시 표 2의 실험과 동일한 조건에서 거리 척도에 따른 인식결과를 각각 나타낸 것이다. 중심이동과 차원감소의 전처리되지 않은 수화영상과 전처리된 수화영상을 대상으로 각각 뉴우턴법의 FP-ICA를 행한 결과이다. 여기에서도 영상 사이의 유사성을 측정하기 위한 척도로 city-block, Euclidean, negative angle의 3가지 거리척도를 이용하였으며, 평균오차 역시 올바르게 인식된 영상을 대상으로 나타낸 것이다. 표 3에서도 2가지 방법 모두에서 이미 학습된 수화에 대해서는 3가지 척도 모두에서 100%의 인식률을 보이나, 시험 수화에 대해서 전처리된 국부기저영상을 이용한 제안된 방법은 city-block, Euclidean, negative angle 각각에 대해 약 92.5%, 약 91.25%, 95.0%의 인식률을 나타낸다. 전처리의 제안된 방법은 비전처리의 방법에 비해 평균적으로 약 1.2배 정도 개선된 인식률을 가져 중심이동과 차원감소가 수화영상의 기하학적 변화에도 강건한 인식 성능이 있음을 확인할 수 있다. 또한 시험 수화영상에서 전처리 방법이 비전처리 방법에 비해 각각 3가지 거리척도에서 평균오차가 작아 좀 더 정확한 수화식별 성능이 있음을 알 수 있다. 따라서 비전처리의 방법에서 발생하는 위치, 각도, 크기, 3가지 동시 변화 등에 따른 오인식은 중심이동과 차원감소의 전처리에 의해 효과적으로 해결될 수 있음을 알 수 있다. 하지만 전처리의 제안된 방법에서도 다른 변화에 비해 각도의 변화에 상대적으로 인식성능은 더 민감함을 알 수 있다. 여기에서도 실험의 결과는 유사성 판별을 위한 거리척도 중에서 Negative angle이 다른 2가지 거리척도에 비해 높은 인식률이 있음을 알 수 있다.

표 3. 전처리에 따른 국부기저영상의 인식결과
Table 3. Recognition results of local basis images by preprocessing

거리척도	입력영상	비전처리 국부기저영상 (FP-ICA)		전처리 국부기저영상(중심 이동+FP-ICA)	
		인식률 (%)	평균오차	인식률 (%)	평균오차
City-Block (L1-norm)	학습수화	100	1.05e-014	100	8.94e-015
	시험수화	81.25	3.97	92.5	3.37
Euclidean (L2-norm)	학습수화	100	0.59e-014	100	4.58e-015
	시험수화	78.75	1.98	91.25	1.68
Negative Angle (Cosine)	학습수화	100	-1.00	100	-1.00
	시험수화	86.25	-0.96	95.0	-0.98

이상의 표 2와 표 3의 실험결과들에서 중심이동과 차원 감소의 전처리 과정을 수행하면 인식에 불필요한 정보를 제거할 수 있어 인식성능이 개선됨을 확인할 수 있다. 따라서 전처리의 제안된 방법은 학습 수화영상에 대해서는 100%, 새로운 시험 수화영상에 대해서는 최고 약 95.0% 정도의 높은 인식률을 가지며, 분류를 위한 유사성 측정에서 Negative angle의 거리척도를 이용하면 가장 정확하게 식별될 수 있음을 알 수 있다.

4. 결 론

본 논문에서는 중심이동과 차원감소의 전처리와 국부기저영상을 이용한 동작인식 기법을 제안하였다. 여기서 중심이동과 차원감소는 1차 모멘트 평행에 기반을 둔 것으로 위치나 크기 변화에 강건한 동작영상을 얻기 위함이고, 국부기저영상의 추출은 독립성분분석 기법에 기반을 둔 것으로 각 동작들마다에 포함된 통계적으로 독립인 동작특징들의 집합을 얻기 위함이다. 특히 국부기저영상을 빠르게 추출하기 위해 뉴우턴법의 고정점 알고리즘에 기반을 둔 독립성분분석을 이용하였다.

제안된 기법을 240*215 픽셀의 160(1명*10종류*16동작) 개 동물표현의 수화동작을 대상으로 city-block, Euclidean, 그리고 negative angle의 척도들을 분류척도로 이용하여 실험하였다. 실험결과, 전처리의 제안된 기법은 국부고유영상을 이용한 비전처리 방법에 비해 평균적으로 약 1.28배 정도, 국부기저영상을 이용한 비전처리 방법에 비해서는 약 1.2배 정도 인식률이 개선됨을 확인하였다. 또한 거리척도로 negative angle이 city-block이나 Euclidean을 이용하는 것보다 낮은 평균오차를 가져 상대적으로 우수한 식별성능이 있음도 알 수 있었다.

향후 제안된 방법을 좀 더 큰 규모의 인식문제와 다양한 분야에 적용하는 연구와 조명 등의 변화에도 견실한 인식성능을 가지는 전처리 기법에 대한 연구가 지속적으로 이루어져야 할 것이다.

참 고 문 헌

[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, pp. 103-115, 1999.

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

[3] T. Watanabe and M. Yachida, "Real-time Gesture Recognition Using Eigenspace from Multi-input Image Sequences," *IEEE Computer and System*, Vol. 30, No.13, pp. 810 - 821, 1999.

[4] 박호식, 배철수, "수화인식을 위한 얼굴과 손 추적 알고리즘," *한국통신학회논문지*, 제 31권 11C호, pp. 1071-1076, 2006년 11월

[5] 김희곤, 오형철, 양윤모, "수화인식을 위한 특징량 추출하드웨어 설계," *대한전자공학회 추계학술대회 논문집*, 제 18권 2호, pp. 729-732, 1995년 12월

[6] 조재현, 엄성용, "다중마크를 이용한 수화 인식 기법," *서울여자대학교대학원 논문집*, 제 1998권 6호, pp. 129-145, 1998년

[7] Y. L. Tian, T. Kanada, and J. F. Cohn, "Recognizing Action Units for Facial Expression Analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.23, No.2. pp. 1-19, Feb. 2001

[8] 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

[9] 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

[10] M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, Vol.3, No.1, pp.71-86, 1991

[11] N. Kambhatla and T. K. Leen, "Dimension Reduction by Local PCA," *Neural Computation* 9, pp. 1493-1516, 1997

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

[13] A. Hyvaerinen and E. Oja, "A Fast Fixed Point Algorithms for Independent Component Analysis", *Neural Computation*, 9(7), pp. 1483-1492, Oct.1997

[14] 김상철역, '재료역학' 정문출판사, 1992년 4월

[15] K. Atkinson, *'Elementary Numerical Analysis'*, John Wiley & Sons, Inc., New York, 1993

저 자 소 개

조용현(Yong-Hyun Cho)

2007년 제 17권 제 7호 참조