

화상처리에 의한 한국어수화인식시스템 개발을 위한 인식 방법

김태수, 전중창
위덕대학교 정보통신공학부

A Method of Development of Korean-Sign Language Recognition System Based on Image Processing

Tae-Soo Kim, Joong-Chang Chun,

Faculty of Information and Communication Eng., Uiduk University

요 약

화상처리에 의한 수화인식은 손의 움직임에 대한 추적을 통한 그 궤적의 정보를 이용하여 주로 처리하여 왔다. 본 논문에서는 궤적 정보만으로 정확히 인식할 수 없는 수화 단어에 대하여 국소 특징 인식 기법을 통하여 보다 정확한 인식을 행한다. 제안한 방법에 의하여 95%이상의 인식 결과를 얻을 수 있었다.

1. 서론

사람의 의사표현 수단으로서 손짓은 그 표현의 다양성으로 인하여 손짓만으로도 언어를 이루고 있다. 일반 건청인에게 있어서도 대화를 하면서 의사표시를 보다 효과적으로 하기 위하여 손짓이나 표정을 흔히 사용하기도 한다. 이와는 대조적으로 농아인들에게 있어서는 손짓만으로 이루어진 언어 즉, 수화가 중요한 의사전달 체계이며, 특히 선천적인 농아인의 경우 높은 문맹률로 인하여 수화만이 유일한 의사전달 수단이 되기도 한다. 국내의 인구 중 35만 농아자와 그 가족 및 친지를 포함하여 100만인으로 추산하고 있다. 또한 병원이나 관공서 등에서 농아자가 개인적인 문제를 다룰 때는 수화 통역자를 통하지 않고 직접 해야 하는 경우가 많으며 긴급시 의사전달, 컴퓨터를 통한 수화교육 및 장기적으로 수화에 의한 전화통화 등의 이유로 인해 수화인식의 필요성이 높아지고 있다 [1-4]. 수화인식에 관한 연구는 대개 데이터글러브(Data Glove)를 사용하였다. 데이터 글러브를 사용한 방법은 글러브에 손가락마다 센서를 부착하여 손동작의 3차원 공간상의 정보를 실시간으로 입력받아 사용

하므로 비교적 용이하게 손짓의 분석과 인식에 이용되고 있다. 그러나 이러한 센서의 이용은 농아인에게 심적인 부담과 부자유스러움을 부여하며 특히 글러브의 크기가 사람에 따라 일치하지 않는 점등이 문제로 대두되고 있다. 이에 반하여 카메라를 이용한 영상처리 기법에 의한 수화의 인식은 농아인에게 있어서 친화적이며 현실적인 지적 인터페이스의 역할을 할 수 있으리라 생각된다. 이러한 화상처리 기법으로는 2차원 영상에 의한 수화의 글로벌 특징만으로 인식을 행하는 후지이 기법이 있다. 그리고 Stamer 등은 HMM을 이용하여 인식을 행하였다[5]. 그러나, 이들의 방법은 정지단어와 양손 수화를 대상으로 하지 않았고, 학습시킨 패턴을 동화상으로부터 탐색적으로 추출하는 데는 효율이 나쁘다[5][6].

일반적인 손의 움직임과 제스처를 실시간으로 인식하는 경우에 촬영한 화상으로부터 화상 전체 또는 팔의 움직임에 대한 특징만을 추출함으로써 인식이 가능해진다. 그러나 수화의 경우는 손 모양, 손의 움직임, 손의 위치에 의해서 많은 단어를 표현하고 있다. 그러므로, 몸 전체의 움직임이 서로 유사한 동작이 다

수 존재하기 때문에 전체적인 특징의 추출만으로는 수화단어를 명확하게 구별할 수 없게된다. 따라서 수화를 인식하는 경우에 손의 위치 및 손의 움직임과 같은 글로벌 특징(Global Features)과 손 모양, 손 방향과 같은 국소 특징(Local Features)의 두 가지 측면에서 특징을 추출하여 인식해야 하는 필요성이 대두된다. 현재까지 동화상으로부터 수화를 인식하기 위한 다양한 방법들이 제시되어 왔다. 이러한 첫 번째 시도로 Matsuo 등은 글로벌 특징과 손 형태를 파악한 국소 특징을 이용한 인식 방법을 사용하였으나, 하드웨어적으로 처리를 하여 인식기가 복잡하고 인식 알고리즘의 효율이 저하된다[7].

본 논문에서는 글로벌 특징만으로 특징지어지는 유사한 의미를 갖는 수화 단어가 복수 개 존재할 경우, 이들 수화 단어에 대하여 보다 정확한 수화단어의 인식을 위하여 2차원 손 모양 추출 및 패턴매칭에 의한 국소 특징에 따른 새로운 수화단어 인식 방법을 제안한다.

2. 수화인식시스템

본 본문에서 추구하는 수화인식시스템의 개념을 그림 1에 나타낸다. 그림 1에서와 같이 수화인식시스템은 스테레오카메라에 의한 글로벌 특징과 국소 특징의 양방을 추출해서 수화를 인식하게된다.

우선 카메라로 촬영한 수화의 화상으로부터 손의 영역만을 잘라내서 국소적인 정보와 글로벌 특징과 분리한다. 수화화상으로부터 손의 주변 화상을 잘라내기 위해서는 손에 흰색 장갑을 착용하고, 주위 배경을 단순화 시켜서 손 영역을 추출하며 움직임을 추적해간다.

특수한 경우를 제외하고는 수화인식시스템은 실내에 설치되는 경우만을 생각해도 되기 때문에 조명이나 촬영조건에 대한 변동은 적다고 할 수 있다.

우선 손의 위치, 손의 움직임에 대한 글로벌 특징을 구한다. 글로벌 특징을 구하기 위해서는 카메라에 의해 취득한 영상데이터로부터 손의 중심위치의 3차원 궤적을 위, 아래, 오른쪽, 왼쪽, 위, 아래의 방향과 45도 경사면에 대한 방향에 대하여 양자화하고, 시간방향에 대해서도 같은 방향의 동작이 계속되는 구간을 압축하여 코드화한다. 또한 손의 위치에 대해서도 몸의 부위에 대한 위치관계를 설정하여 놓고, 대상자별로 체격이 차이가 있기 때문에 정규화를 행하면 인식이 가능해진다.

손의 움직임을 추적하기 위해서, 칼만필터(Kalman

Filter)를 사용한다[9]. 칼만필터는 이미 보행자의 움직임의 파라미터 추정이나 인체로부터 얻어진 브로브(blob)의 움직임에 대한 파라미터를 추정하는데 적용하여 왔다[7].

본 시스템에서는 손의 움직이는 한 스텝 당 범위가 적으며, 또한 손의 움직이는 전체 범위가 상반신에 국한된다는 점을 고려하여 다음과 같이 화상평면상의 손의 중심점에 대하여 손의 위치와 가속도를 포함한 상태벡터 x 를 사용하여 모델을 설정한다.

$$x_k = P_{k-1}x_{k-1} + Q_{k-1}a_{k-1} \quad (1)$$

여기서, $P_{k-1} = \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix}$, $Q_{k-1} = \begin{pmatrix} T^2 \\ T \end{pmatrix}$, T 는 샘플링간격, a_{k-1} 은 시스템의 잡음이다.

다음은 식(1)에서 설정한 모델에서 시스템에서 발생하는 잡음 a_{k-1} 가 관측되는 화상에서의 잡음 성분과 서로 독립인 가우시안 잡음으로 가정하고, 시스템 프로세스를 단순화시킨 다음, 식(2)와 같은 추정모델을 가정한다.

$$y_k = Hx_{k-1} + w_{k-1} \quad (2)$$

여기서, $H = [1, 0]$ 이고, x_{k-1} 은 시각 $k-1$ 에서의 상태벡터, w_{k-1} 은 관측잡음이다.

y_k 는 시각 k 에서 손의 관측위치가 된다.

다음은 손 모양, 손의 방향과 같은 정보로부터 국소 특징을 구한다. 국소 특징을 구하기 위해서는 각 단어 별로 시작(Begin)과 끝(End)의 정보가 중요하기 때문에 이들 프레임 사이의 손 모양, 자세의 변화에 대한 추이 정보를 기술하여 둔다. 손이 이동하여 움직이는 중간에 있어서의 손 모양의 변화는 인간의 시각 특성상 별로 의미를 갖지 않아도 되기 때문에 변화되는 손 모양을 무시하고, 단어의 시작과 끝점에서의 손 영상에 대해서만 국소 특징을 구한다.

이렇게 글로벌 특징과 국소 특징을 각각 구하여 이들 특징을 이용하여 수화인식을 행한다. 즉, 농아인의 움직임에 대하여 글로벌 특징에 따라 인식을 행하고, 동일한 글로벌 특징을 갖는 단어를 구별하기 위해서 다음은 국소 특징을 이용한다. 물론 글로벌 특징만으로 단어의 구별이 이루어지는 경우에는 국소 특징은 생략한다.

다음은 CCD카메라로 대상자를 촬영하고 손의 동작을 검출한다. 손동작의 시작 위치와 동작코드를 검출하며, 동시에 동작영역을 분할하여 손 모양을 검출하

여 수화동작 정보가 기술되어 있는 DB와 대조하여 수화단어를 판정한다.

여기서 동작영역의 분할 및 글로벌 특징의 인식은 종래의 화상처리 기법을 그대로 적용하여 시스템을 구성한다. 그러나 국소 특징의 인식은 다음과 같은 알고리즘에 따라 행한다.

Step 1. 손 영역 화상의 분할 및 취득, Step 2. 특징 공간의 작성, Step 3. 손 화상의 패턴에 대한 특징 부호 할당, Step 4. 시작, 끝점의 손 모양 패턴과 DB를 대조하여 학습 및 인식 수행 순으로 이루어진다.

수화단어에 대한 인식이외에 자연수화에 대한 수화문장을 인식하기 위해서 수화단어와 단어 사이의 구별되는 시점을 알기 위해서 Segmentation기법을 적용한다. 일반적으로 손의 동작이 아주 작은 시점을 정확하게 포착하기가 쉽지 않다. 따라서, 본 연구에서는 Segmentation으로 손이 복부를 중심으로 아래에 위치하는 여부에 따른 정보와 양손이 동시에 접촉하는 경우 등을 포함한 정보를 중시하여 구분점으로 정한다. 물론 수화단어와 단어사이에 움직임이 멈추어지는 경우는 Segmentation으로 간주한다.

3. 국소 특징 인식

3.1 손 영역 추출

흰 장갑을 착용하여 촬영된 카메라로부터 들어온 화상으로부터 단어의 시작과 끝 영역을 잘라내어 국소 특징 인식을 위하여 임계값을 정하여 임계값 이하는 모두 흑색처리한다. 즉 손 이외의 모든 배경은 흑색으로 변환하여 2진화 한다. 변환된 흑백 화상으로부터 흰색 부분, 즉 손 영역을 중심으로 손 길이에 외접 사각형을 구하여, 그 크기로 잘라낸다. 다음은 이렇게 얻어진 화상으로부터 확대 축소 과정을 거쳐서 손의 화상을 정규화 한다. 손 영역 추출 과정을 그림 1에 나타낸다.

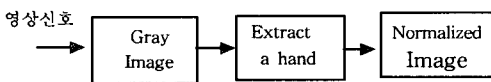


그림 1. 손 영역 추출 과정

3.2 손 모양 정의

손 모양은 손의 외형과 손의 방향에 의해서 정의된다. 손의 외형은 손가락의 펼침과 구부림에 의하여 표현된다. 기본적인 형태를 B로 표기하고, 각 손가락에 임지는 1, 집게손가락은 2, 중지손가락은 3, 약손가락은 4,

새끼손가락은 5로 번호를 부여한다. 즉, 집게손가락 하나를 펼쳐 보이는 외형을 B2로 표기한다.

다음은 손의 방향을 정할 때, 손가락의 방향은 F로 하고, 손바닥은 P로 둔다. 방향은 위쪽이 u(upper)과 같이 영어의 알파벳 첫째 문자를 따서 d, l, r, f, b로 표기한다. 즉, 수화단어 숫자 3의 경우 B234FuuuP로 표기하고, 5의 경우 B1FIP로 표기한다.

또한, 손 동작 코드로 HM1은 한 손으로 80×80의 정규화 크기로 하고, HM2는 양손(분리)로 80×80로, HM3는 양손(접촉)으로 100×100의 정규화 크기로 한다.

4. 실험 및 고찰

4.1 수화단어 실험

본 수화인식시스템에서는 최초로 글로벌 특징에 의해서 인식을 행하고, 글로벌 특징으로 분류된 동일한 특징을 가진 단어의 식별을 위해서 국소 특징 인식을 행하게 된다. 따라서, 본 실험에서는 글로벌 특징만으로 식별하기 힘든 단어에 대하여 식별이 어느 정도 가능한가를 검증할 필요가 있다.

본 시스템에 사용된 하드웨어로는 스테레오 카메라와 Matrox Meteor-II/MC4 화상처리 보드가 이용되었다. 본 실험에 주로 사용된 단어는 일상생활에 사용되는 표현이 간단한 약 100개의 단어가 선택되었다. 수화단어는 농아자를 대상으로 촬영하여 둔 Video Tape 및 표준수화사전을 사용하여 해당 단어의 동작을 관찰하여 글로벌 특징에 의한 단어 분류 방식에 준하여 손의 움직임이 글로벌 특징이 동일한 단어를 같은 범주로 분류하였다. 우선, 한 손 동작, 양손 동작으로 분류하였고, 양손 동작의 경우에 양손이 접촉하는 경우와 접촉하지 않는 경우로 분류하였다.

본 실험은 일상 연구실의 조명하에서 촬영된 상반신 화상으로부터 640×480화소, 8비트 화상을 대상으로 하였다. 3명의 피험자로 하여금 한 단어에 대하여 다수의 샘플 화상을 취득, 시작과 끝 동작의 손 화상을 취득했다. 인식실험에서는 각 피험자의 손동작 및 손 모양을 촬영하여 앞 절에서 설명한 알고리즘에 따라서 한 손 화상, 양 손 화상으로 분할하여 실험을 행하였다. 실험에 사용된 동작의 수화 단어 예를 표 1에 나타낸다.

4.2 실험결과 및 고찰

인식실험에서 같은 피험자가 다른 수화단어를 HM1에 대하여 975 샘플, HM2에 대하여 574 샘플, 그리고

표 1. 손 분류에 따른 단어 예

손 동작 코드	손 분류	단어 예
HM1	한 손	<부, 모>, <형, 제>, <자, 매>, ...
HM2	양 손(분리)	직업, 기쁘다, 바쁘다, ...
HM3	양 손(접촉)	헤어지다, 떴다, 보통, 시작하다, ...

HM3에 대하여 890 샘플에 대하여 종래 인식 결과와 비교하기 위하여 실험을 실시하였고, 그 실험한 결과를 표 2에 나타낸다. 표 2에서와 같이 모든 수화단어에 대하여 95%이상의 높은 인식률을 보인다. 표 2에서와 같이 수화인식을 위해서는 종래의 글로벌 특징 인식만으로는 분별하기 어려운 단어에 대하여 국소 특징 인식 방법이 유효함을 알 수 있다. 또한 종래의 단순한 패턴 매칭에 의한 인식 결과 보다 양호한 특성임을 표 2에 보인다[7].

표 2. 제안한 패턴 매칭에 의한 인식 결과

손의 움직임	샘플수	정답수		인식률(%)	
		종래 방식	제안한 방식	종래 방식	제안한 방식
HM1	975	878	928	90	95
HM2	574	499	480	87	96
HM3	890	734	836	83	94

5. 결론

글로벌 특징만으로 정확히 인식할 수 없는 수화 단어에 대하여 2차원 손 모양 추출 및 패턴매칭에 의한 국소 특징 인식기를 통하여 인식을 행하였다. 인식 실험 결과 한 손만 사용한 경우는 물론 양손을 이용한 수화표현의 경우도 95%이상의 양호한 결과를 얻을 수 있었다. 본 논문에서는 얼굴과 손의 구별을 단순화하기 위해서 흰 장갑을 착용하였으므로, 보다 친화적인 수화를 위해서는 손과 얼굴을 구별하기 위한 칼라 처리가 필요하다고 본다. 따라서, 보다 정확한 인식을 위해서 칼라처리 및 3차원 해석이 금후의 과제이다.

[참고문헌]

- [1] S. Tamura, S. Kawasaki, "Recognition of Sign Language Motion Images", Pattern Recognition, Vol. 21, No.4, pp. 343-353, 1988
- [2] M.F. Costabile, C. Guerra and Pieroni, "Matching Shapes : A case study in time-varying images", Computer Vision Graphics Image Process. 29, pp.296-310, 1985
- [3] N. Hagiwara and M. Waki, "Recognition of the finger alphabet using projections", National Convention Record, Institute of Electronics and Communication Engineering of Japan, No. 1536, 1986
- [4] K. Imagawa, S. Lu, S. Igi and H. Matsuo, "Real-Time Tracking of Human Hands from a Sign-Language Image Sequence in Consideration of Disappearance by Skins Regions", Institute of Electronics and Communication Engineering of Japan, Vol. J81-D-II, No.8, pp.1787-1795, 1998
- [5] Starner, T., Pentlang, A., "Real-Time American Sign Language Recognition from Video using Hidden Markov Models", On Proc. International Symposium on Computer Vision, pp.265-270, 1995
- [6] Marcell, A., Kirsti, G., "Video-Based Sign Language Recognition Using Hidden Markov Models", Proceedings of International Gesture Workshop Bielefeld Germany 1997, pp.97-109, 1997
- [7] H. Matsuo, S. Lu, S. Igi, K. Imagawa, Y. Takata, Y. Nagashima, "A Japanese Sign Language recognition based on global features and localized characteristics in images captured using stereo camera", Human Interface, Vol. 3, No. 3, pp.135-144, 2001
- [8] S. Haykin, Kalman Filtering and Neural Networks, John Wiley & Sons, Inc., 2001

본 연구는 한국과학재단 목적기초연구 (과제번호: (R05-2001-000-01494-0(2002)))지원으로 수행되었음.