

# 한-일 수화 영상통신을 위한 3차원 모델

愼 晟孝\*, 金 商雲\*, 青木 由直\*\*

\* 明知大學校 컴퓨터工學科

\*\* 北海道大學 工學研究科

## 3D Model for Korean-Japanese Sign Language Image Communication

Seong-Hyo Shin\*, Sang-Woon Kim\*, Yoshinao Aoki\*\*

\* Dept. of Computer Engineering, Myongji Univ.

\*\* Graduate School of Engineering, Hokkaido Univ.

shshin@ce.myongji.ac.kr

### Abstract

In this paper we propose a method of representing emotional expressions and lip shapes for sign language communication using 3-dimensional model. At first we employ the Action Units(AU) of Facial Action Coding System(FACS) to display all shapes. Then we define 11 basic lip shapes and sounding times of each components in a syllable in order to synthesize the lip shapes more precisely for Korean characters. Experimental results show that the proposed method could be used efficiently for the sign language image communication between different languages.

### 1. 서론

수화는 청각장애자들과 그 주변 사람들에게 있어서 중요한 통신 수단이다. 따라서, 컴퓨터를 이용한 수화 교육 시스템에 대한 연구가 활발하다. 그러나, 대부분의 연구는 단지 팔과 손의 움직임만을 보여주는 데 그치고 있다[1],[2]. 최근 얼굴에 의한 감정 표현을 기존의 시스템에 추가하여 디스플레이하는 방법이 제안되었다[3]. 논문 [3]의 결과에 의하면, 얼굴 표정을 가진 수화 영상은 기존의 영상에 비해 이해하기 쉽다. 이에 수화가 의미하는 말을 입모양으로 표현하는 구화(口話)를 추가함으로써, 보는 사람이 보다 이해하기 쉽도록 하는 연구가 진행되었다[4]. 그러나, 위의 연구는 손을 위주로 하는 2차원 모델을 사용하기 때문에 애니메이션 속도는 빠르지만, 현실감이 떨어진다. 그리고, 화자의 정면만을 보여줄 수 있고, 다른 각도에서의 수화 동작을 관측하는 것은 불가능하다는 단점이 있다.

이러한 문제점을 해결하기 위하여 3차원 모델을 이용하는 방법이 사용되고 있으나[5], 많은 제어점의 사용으로 인하여 애니메이션하는 데 많은 시간이 소요된다.

따라서, 이 논문에서는 최소한의 제어점을 사용하여 3차원의 모델을 구성한 후, 손에 의한 수화 동작, 얼굴에 의한 감정 표현, 그리고 구화에 대한 입술 모양을 표현하는 방법을 제안한다.

한글에 대한 입모양을 구현하기 위하여 논문 [5]에서와 같이 자모를 발음방법에 의하여 여러 가지로 분류한 후, 전체 11가지(자음 : 3, 모음 : 8) 기본 입모양을 정의한다. 그리고, 한 음절에 대한 입모양을 합성하기 위하여, 네 가지 기본 입술형을 시간별로 합성한다. 이 때, 입모양을 보다 자연스럽게 정확하게 표현하기 위하여 각 음소의 발음 시간을 이용하여 각 음소사이의 입술 모양을 보간한다. 그리고, 얼굴 구성 요소들에 대한 움직임을 표현하기 위하여 FACS(Facial Action Coding System)의 AU(Action unit)들의 조합을 이용한다[4]. 또한 제안한 방법의 성능을 고찰하기 위하여 지능적인 영상 통신 기법을 이용한 수화 대화 시스템을 구현한다[6].

### 2. 수화 모델

그림 1은 기존에 사용하였던 2차원 수화 모델[4]과 이 논문에서 제안하는 3차원 수화 모델을 비교한 것이다. 그림 1의 (a)에 나타난 2차원의 모델은 전체 291개의 제어점을 사용하였고, 각 제어점 사이를 선으로 이어 수화 및 감정 표현에 필요한 손, 팔, 얼굴 등을 구성하였다. 선을 위주로만 나타내기 때문에 팔의 동작 및 얼굴의 감정을 현실적으로 나타내는 데에는 한계가 있다. 따라서, (b)와 같이 3차원의 모델을 이용함으로써, 위의 문제를 해결할 수 있다. (b)는 1071개의 제어점을 이용하였으며, 세 제어점을 이어 1840개의 삼각형을 만들었다. 기존의 연구[5]에서는 사진으로 실제 사람의 영상을 입력받아, (b)의 표준 모델을 사람의 영상에 맞추는 작업을 실행하였다. 이 작업은 실제 사람이 수화를 하는 것처럼 보이기 위한 것이나, 텍스처 매핑(texture mapping)등에 많은 시간이 소요되므로

이 논문에서는 이 작업을 배제하였다.

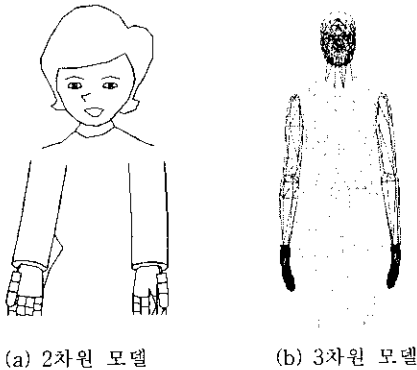


그림 1. 수화 모델의 비교

얼굴에 화자의 감정은 얼굴의 각 구성 요소들의 움직임을 보고 판단할 수 있다. 따라서, 얼굴을 모델링할 때에는 이러한 점을 충분히 감안하여, 신체의 다른 부분보다 자세하게 할 필요가 있다. 그림 2는 논문에서 사용하는 3차원 얼굴 모델을 나타낸 것이다. 특히, 눈썹, 눈, 입 등의 모양은 감정을 표현하는데 중요한 요소 중의 하나이다. 이를 고려하여 눈썹, 눈, 입부분에는 좀 더 많은 제어점을 두어 자세하게 표현하였다.



그림 2. 3차원 얼굴 모델

### 3. 얼굴에 의한 감정 표현 방법

그림 2의 3차원 얼굴 모델에 감정을 표현하기 위해서 이 논문에서는 FACS의 AU 조합 및 강도의 조절을 이용하여 6가지의 감정을 얼굴로 표현한다. “기쁨(Joy)”은 AU6 + AU12 + AU25를 결합하여 구현하며, 그림 3에 2차원 모델의 움직임과 비교하여 나타내었다. 하나의 AU는 얼굴의 일정부분을 움직이는 역할이며, 강도는 그 움직임의 정도를 나타낸다. 3차원 모델을 사용할 경우, 2차원 모델에서는 자세히 표현하지 못하는 부분까지 표현할 수 있다. 그림 3의 AU6은 “뺨을 올린다”이다. (a)의 2차원 모델은 단지 눈의 아랫부분이 올라가는 것으로만 구현하는데 비하여, (b)의 3차원 모델에서는 실제 뺨부분의 제어점 좌표를 이동시킴으로써 자세히 표현할 수 있다.

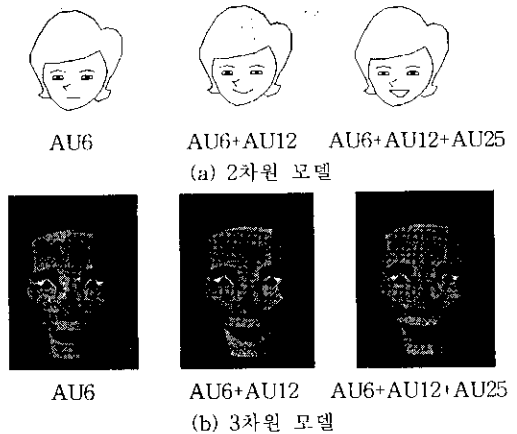


그림 3. “기쁘다(Joy)”에 대한 AU 구현 예

“슬픔(Sadness)”은 1 + 4 + 5, “혐오감(Disgust)”은 4 + 10 + 17, “화남(Anger)”은 4 + 5 + 7 + 24, “놀람(Surprise)”은 1 + 2 + 5 + 26, “두려움(Fear)”은 1 + 2 + 4 + 5 + 20 + 25로 구현한다.

### 4. 입모양 합성

#### 4.1 기본 입술형의 정의

한글은 14,364개의 음절을 가지고 있기 때문에 이 모든 음절에 대한 입모양을 정의하는 것은 불가능하다. 이 문제를 해결하기 위하여 음절단위로 입모양을 합성하는 것이 아니라 글자를 자모 단위의 음소로 나눈 후, 음소 단위로 입모양을 합성한다.

자음들은 발음방법에 관찰하여 구개음, 연구개음, 목청소리를 하나의 기본 입모양으로 정의하고 혀끝소리, 입술소리를 각각 하나의 기본 입모양으로 정의한다[4].

모음에 대한 입모양 합성을 위하여 8개의 단모음을 기본 입모양으로 정의한다. 이중 모음은 두 단모음을 결합하여 구현함으로써 해결한다. 따라서, 자음에 대하여 3가지, 모음에 대하여 8가지의 기본 입모양을 정의한다.

#### 4.2 기본 입술형의 구현

4.1절에서 정의한 11가지 기본입술형을 디스플레이 하기 위하여 감정 표현에 사용하는 방법과 같이 FACS의 AU조합을 이용한다. 표 1은 11가지 기본 입모양에 대한 AU들의 조합과 그 강도를 나타낸다.

2차원 모델에 적용하는 AU조합과는 같지만 그 강도 및 구현에 있어서는 고려해야 할 사항들이 많다. 특히, 3차원적인 움직임을 잘 표현해야 한다. 그리고, 2차원 영상에서는 해당 제어점만 움직이면 되지만, 3차원은 그 주위의 점들의 움직임도 표현하여 실제감을 증가시킬 수 있다. 입술의 움직임을 예로 들어보면, 3차원 모델의 경우에는 입술의 제어점과 함께 양 볼의 제어점을 같이 움직여 실제 사람의 움직임과 유사하게 보여지게 할 수 있다. 그러나, 2차원 모델은 양 볼의 움직

임을 표현할 수가 없어 단지 입술의 선의 변화로만 움직임 표현한다.

표 1. 기본 입모양에 대한 AU조합 및 강도

기호	기본입술형	AU 번호				
		AU10	AU18	AU20	AU25	AU26
V1	ㅏ	1.0	0.0	1.0	0.0	1.5
V2	ㅑ	1.0	0.0	0.0	0.0	1.2
V3	ㅓ	1.5	1.0	0.0	0.0	1.0
V4	ㅕ	0.7	1.0	0.0	0.0	0.7
V5	ㅗ	0.3	0.0	0.5	0.0	1.2
V6	ㅛ	0.0	0.0	0.5	0.0	1.2
V7	ㅜ	0.0	0.0	0.7	0.0	0.3
V8	--	0.3	0.0	0.3	0.3	0.3
C1	입술소리	0.0	0.6	0.0	0.0	0.0
C2	혀끝소리	0.3	0.0	0.0	0.0	0.3
C3	그외소리	0.0	0.0	0.0	0.0	0.6

4.3 연속 발음의 구현

한글의 한 음절을 발음하기 위해서는 4.2절에 기술한 초, 중, 종성의 각 음소에 대한 기본 입모양을 시간별로 디스플레이하면 된다. 여기서, 각 음소별로 발음하는 시간을 고려해야 하며, 이 논문에서는 초싱 30%, 중성 40%, 종성 30%의 수준으로 하였다. 그리고, 각 음소별 기본 입술형을 연결할 때 보다 부드럽고 자연스럽게 하기 위하여 음소사이의 입술모양을 보간한다.

단어 "안녕하세요"의 발음은 C3→V1→V1→C2(안)→C3→V7→V2→C3(녕)→C3→V1→V1→C3(하)→C3→V6→V6→C3(세)→C3→V6→V2→C3(요)의 20개 기본 입술형을 시간에 따라 디스플레이하면 된다.

위의 각 기본 입술형을 순서대로 디스플레이하는 데 있어서 해당 기본 입술형이 디스플레이되는 시간을 고려해야 한다. 이 논문에서는 수화가 시작할 때 적절한 입모양을 설정하고 동시에 수화 동작과 입모양 합성이 종료되도록 동기화한다.

4.4 단어 사전의 구조

여러 가지 수화에 대한 정보는 수화의 손, 팔 동작 파라메타들과 얼굴 표정, 입모양 파라메타를 갖고 있는 단어 사전을 이용한다.

손, 팔의 움직임은 동작(action)의 수와 각 동작을 나타내는 54개의 파라메타로 이루어진다. 54개의 파라메타는 손과 팔에 있는 관절들의 회전각을 표시한다.

입모양은 단어의 음절의 수와 음절의 입모양에 대한 4개의 파라메타로 이루어진다. 4개의 파라메타는 음절을 구성하는 4가지 요소들의 기호이다. 그림 4는 4개의 동작을 갖는 "안녕하세요"에 대한 단어사전의 구조를 예로 보인 것이다.

안녕하세요		4
1st action	54 parameters	
2nd action	54 parameters	
3rd action	54 parameters	
4th action	54 parameters	
emotion	E1	
	5	
1st syllable	C3, V1, V1, C2	
2nd syllable	C3, V7, V2, C3	
3rd syllable	C3, V1, V1, C3	
4th syllable	C3, V6, V6, C3	
5th syllable	C3, V6, V2, C3	

그림 4. "안녕하세요"에 대한 단어 사전 구조의 예

5. 실험

제안한 방법의 성능을 평가하기 위하여, 기존에 연구되어졌던 수화 영상과 비교한다. 2차원 영상은 SUN과 PC에서 구현하였으며, 3차원 영상은 PC에서 Open Inventor 라이브러리를 이용하여 구현하였다.

기존에 연구되어졌던 수화 영상을 그림 5, 6, 7에 나타내었다. 전체 40개의 프레임 중에서 5프레임 간격으로 나타내었다. 그림 5는 "기쁘다(Joy)"라는 감정과 "안녕하세요"의 구화를 추가한 2차원 수화 영상이다. 수화의 동작과 함께 입모양도 "안녕하세요"라는 단어를 뜻하기 때문에 이해하기가 보다 쉽다. 그러나, 선으로만 표현되기 때문에 애니메이션 속도는 빠르나 현실감이 떨어진다.

그림 6은 3차원 모델을 이용하여 감정 표현과 구화를 표현한 수화 영상이다. 그림 5에 비해 현실감이 뛰어나다. 그림 7은 측면으로 본 수화 영상이다. 정면으로만 봐서 분석하기 어려운 수화 동작일 경우 그림 7과 같이 측면 또는 다른 각도에서 자유자재로 볼 수 있다는 장점이 있다.

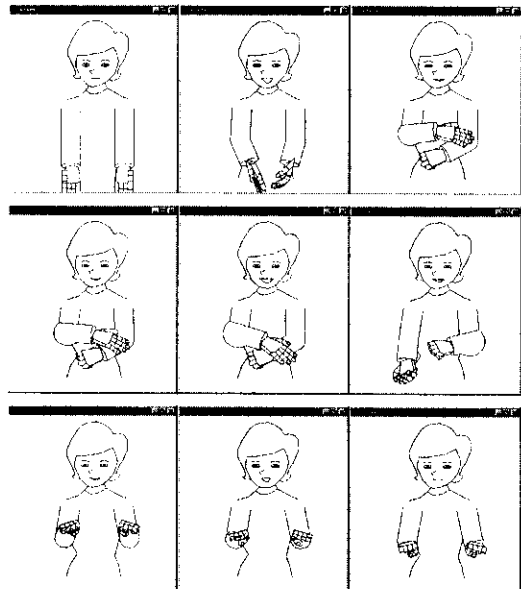


그림 5. 2차원 수화 영상[4]

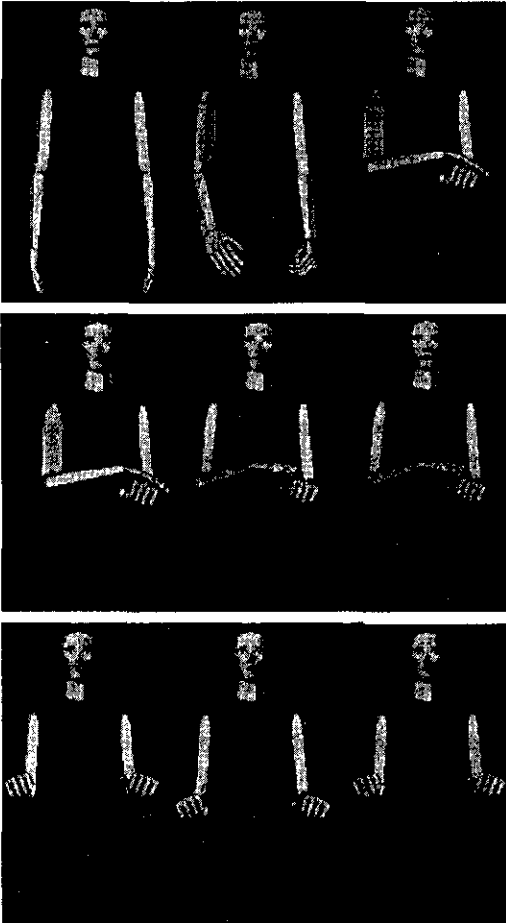


그림 6. 3차원 정면 수화 영상

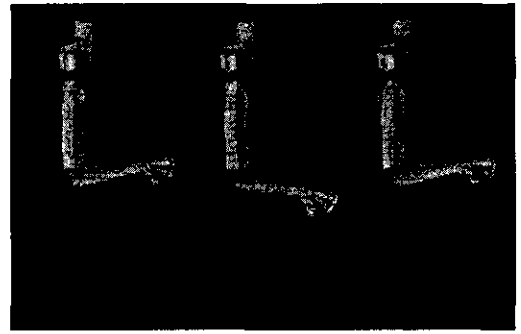
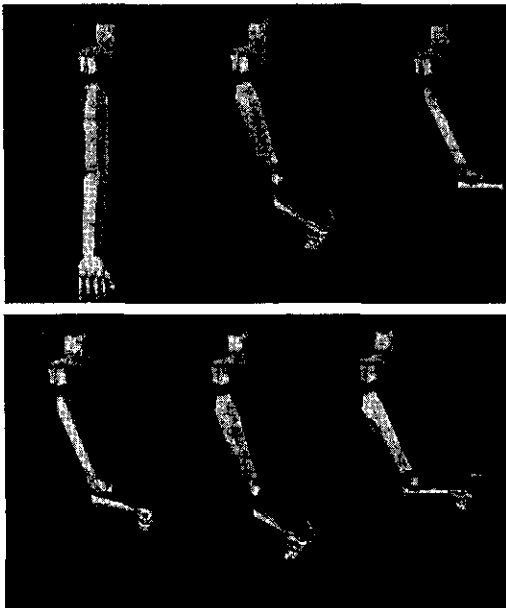


그림 7. 3차원 측면 수화 영상

## 6. 결론

이 논문에서는 3차원 모델을 이용하여 한국 수화 영상에 입술의 감정 표현과 입모양 합성을 추가하는 방법을 제안하였다. 3차원 모델을 이용함으로써 현실감 있게 수화 동작을 볼 수 있으며, 여러 각도에서 수화 동작을 분석할 수 있는 장점이 있다.

입모양을 보다 정확하고 자세하게 표현하기 위해서 자모음에 대해서 11가지의 기본 입술형을 정의하고, 발음시간을 이용하여 보간하였으며, FACS의 AU를 이용하여 얼굴 구성 요소의 움직임을 표현하였다.

실험 결과, 기존의 2차원 수화영상에 비해 보다 자연스럽고 이해하기 쉽다는 것을 알 수 있었다. 앞으로는 이종 언어간 수화 통신을 위한 수화 번역 시스템에 응용할 계획이다.

## 참고문헌

- [1] 徐, 棚橋, 坂本, 青木, “手話畫像通信のための手振り記述と單語辭典の構成法”, 電子情報通信學會論文誌, Vol. J76-A, No. 9, pp. 1332-1341, 1993. 9.
- [2] 吳芝英, 愼晟孝, 金商雲, 棚橋 眞, 青木 山直, “한글 텍스트의 수화 영상 변환 시스템”, 대한전자공학회 하계 학술대회 논문집, Vol. 20, No. 1, pp. 311-314, 1997.
- [3] 金商雲, 愼晟孝, 李今燦, 青木 山直, “手話畫像通信のための顔の感情表現”, 電子情報通信學會 秋季學術大會 論文集, D-11-81, 1997.
- [4] 愼晟孝, 金商雲, 青木 山直, “한글 수화 영상통신에서의 감정과 입모양 표현 방법”, 대한전자공학회 추계 학술대회 논문집, Vol. 20, No. 2, pp. 1217-1220, 1997.
- [5] 최운영, “상반신 3차원 모델을 이용한 수화 동영상 생성시스템”, 명지대학교 공학석사 학위 논문, 1996.
- [6] 吳芝英, 金商雲, 青木 山直, “지적 영상 통신에 의한 수화 채팅 시스템의 구현”, 대한전자공학회 추계 학술대회 논문집, Vol. 20, No. 2, pp. 1221-1224, 1997.