

스플라인 보간법을 이용한 3차원 수화 애니메이션의 생성

○李 鍾雨*, 吳 芝英*, 金 商雲*, 青木 由直**

* 明知大學校 컴퓨터工學科
** 北海道大學大学院工學研究科

Generation of 3D Sign Language Animation Using Spline Interpolation

○Jong-Woo Lee*, Ji-Young Oh*, Sang-Woon Kim*, Yoshinao Aoki**

* Dept. of Computer Engineering, Myongji University
** Graduate School of Engineering, Hokkaido University
kimsw@wh.myongji.ac.kr

Abstract

We have implemented a sign language animation system using 2D and 3D models. From the previous studies, we find out that both models have several limitations based on the linear interpolation and fixed number of frames, and they result in incorrectness of actions and unnatural movements. To solve the problems, in this paper, we propose a sign language animation system using spline interpolation method and variable number of frames. Experimental results show that the proposed method could generate animation more correctly and rapidly than previous methods.

1. 서 론

멀티미디어 통신시대에서 수화통신은 언어의 벽을 넘을 수 있는 유용한 통신 수단이 될 수 있다. 또한 인터넷은 사회 구성원간의 중요한 통신 수단으로서, 이를 더욱 용이하게 사용할 수 있도록 다양한 미디어를 개발하고 있다. 그러나, 농아자나 고령자 등 정보취약자를 위한 정보 매체는 아직까지 그 개발이 미흡한 실정이다. 이에 저자들은 수화통신으로 서로 다른 언어권의 사용자들이 의사소통을 할 수 있고, 농아자들이 인터넷에서 수화를 사용할 수 있도록 하여 이들의 사회참여를 유도하는 비언어 통신 방법을 연구하여 왔다[1].

지금까지 구현한 시스템에서는 문장을 입력하여 수화 파라미터로 번역한 후 그래픽 모델을 이용하여 수화 애니메이션을 생성하였다. 시스템 연구의 초기 단계에서는 2차원 모델을 이용하였다. 2차원 모델은 사양이 낮은 플랫폼에서도 빠른 속도로 수화 동영상을 합성할 수 있는 장점을 가지고 있다. 반면에, 보는 위

치가 정해져 있는 2차원 모델의 한계로, 일부 수화 단어는 알아보기가 매우 힘들었다. 2차원 모델의 이러한 한계를 극복하기 위해 3차원 모델[2]을 이용하여, 자유로운 방향에서 수화 동작을 볼 수 있도록 했다.

또한, 수화 동영상을 합성하기 위해 이 시스템에서는 기본적으로 키 프레임(key frame) 애니메이션 방식을 사용했다. 키 프레임 애니메이션 방식은 숙련된 애니메이션이터가 키 프레임만을 준비하고, 나머지 프레임들은 보간 공식을 이용하여 채우는 방식이다. 그러나 기존의 시스템에서는 키 프레임 사이의 프레임 수를 모든 동작에 대해 동일하게 적용했다. 따라서 움직임이 작은 동작이 움직임이 큰 동작에 비해 상대적으로 많은 프레임이 생겨나기 때문에 속도가 느려질 뿐만 아니라 부자연스러운 동작을 연출하는 역효과를 유발했다. 따라서, 이 논문에서는 키 프레임 사이에 채워야 할 프레임 수를 전 수화 동작과 다음 수화 동작 사이의 파라미터 차이에 따라 변화시키는 가변 프레임 방식(variable number of frames)을 제안한다.

그리고, 기존의 방식에서는 키 프레임 사이를 보간하기 위해 선형 보간(linear interpolation) 방식을 사용했다. 선형 보간 방식은 각 키 프레임 사이를 직선으로 연결하기 때문에, 곡선 동작이 주를 이루는 수화 동작에 있어서는 움직임이 어색하거나, 동작을 정확히 표현하기 어렵다는 단점이 있다.

이에 이 논문에서는 3차원 모델의 부자연스러운 동작을 개선하기 위해 3차 스플라인(cubic spline) 보간 방식을 사용한다.

2. 텍스트의 수화 파라미터 변환

2.1 3차원 모델과 관절각 파라미터

모델 구조는 계층을 가진 트리 구조로 표현할 수 있다. 그림 1은 3차원 모델의 팔과 손가락에 대한 계층 구조[3]이다.

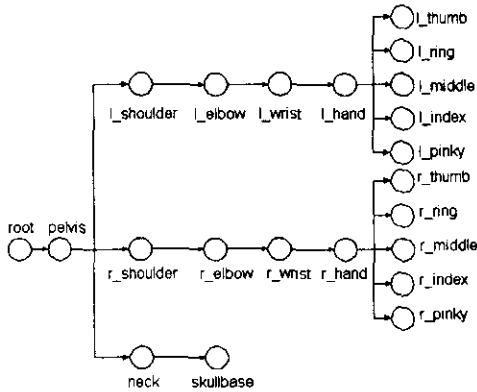


그림 1. 3차원 모델의 계층 구조

Fig 1. The hierarchical structure of 3D model

여기서 모델의 한 노드는 하나 이상의 근육과 한 개의 관절로 이루어진다. 그러므로, 각 노드는 근육에 대한 점 데이터와 관절의 위치, 회전각에 대한 정보를 가지게 된다.

수화 모델의 각 관절은 회전할 수 있는 축과 회전할 수 있는 범위가 정해져 있다. 이 때 관절이 회전할 수 있는 축의 수를 자유도라 하며, 각 관절의 자유도는 표 1과 같다.

표 1. 관절과 자유도

Table 1. Joints and DOF(Degree Of Freedom)

관절 명칭	DOF	회전축
shoulder	3	(x, y, z)
elbow	2	(x, y)
wrist	2	(x, z)
metacarpophalangeal(finger)	2	(x, z)
interphalangeal(finger)	1	(x)
carpometacarpal(thumb)	2	(x, z)

즉, 한쪽 팔에 7개의 자유도, Metacarpophalangeal(손가락 뿌리 관절)에 2개, Interphalangeal(손가락 마디 관절)에 각 1개씩, 한 손가락에 4개가 존재하며, 한 손에는 20 자유도(=4×5)가 존재한다. 그러므로, 양 쪽 팔에 대해 54(=27×2) 자유도가 존재한다.

2.2 텍스트의 수화 파라미터 변환

수화 애니메이션을 수행하기에 앞서 입력문을 수화 파라미터열로 변환하는 과정을 거친다. 그림 2는 한글 문장을 한국 또는 일본 수화 파라미터로 변환하는 과정이다. 한국어 문장이 입력되면 형태소 분석과 구문 분석 과정을 거쳐 대상 언어의 파라미터로 변환한다. 즉, 한국 수화로 번역할 경우에는 한국어 수화 사전에서 검색하고, 일본어 수화로 번역할 경우에는 한국어/일본어 테이블을 거쳐 일본 수화 사전을 검색하여, 일본 수화 파라미터를 찾아낸다. 반대로 일본어 문장이

입력된 경우에는 일본 수화 또는 한국 수화 파라미터를 검색한다.

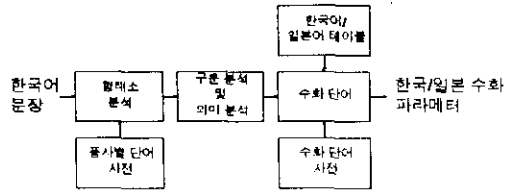


그림 2. 한글 문장의 수화 파라미터 변환 과정

Fig 2. Translation of Korean sentence into a series of sign language parameters

3. 수화 애니메이션의 생성

3.1 가변 프레임 방식

가변 프레임이란 키 프레임간의 프레임 수를 미리 고정시키지 않고 동작의 변화에 맞게 설정하는 방법이다. 동작 변화의 크기는 이전 동작 파라미터와 현 동작 사이의 파라미터의 변화를 이용하며, 식(1)과 같이 구한다.

$$F_i = MAX_j \left(\frac{|V_{i-1}^j - V_i^j|}{11 \times 180} \right) + 2, \quad (1 \leq j \leq 14) \quad (1)$$

여기서, i : 키 프레임 번호, j : 파라미터 번호,

V_i^j : i 번째 키 프레임의 j 번째 팔 파라미터,

F_i : i 번째 키 프레임과 $i-1$ 번째 키 프레임 사이의 프레임 수.

F_i 는 V_{i-1} 과 V_i 사이에서 각도 변화가 가장 큰 파라미터를 기준으로 설정한다. 움직일 수 있는 각도의 최대 범위는 $\pm 180^\circ$ 이며, 이를 11등분으로 나눈다. 이때, 최소 프레임 수는 2로 한다.

그리고, 손가락의 각도 변화는 팔이나 어깨의 움직임에 비해 동작의 크기가 아주 작다, 따라서, 팔 동작 파라미터의 변화에 대해서만 식(1)을 적용한다. 팔 동작 파라미터는 표 1의 shoulder, elbow, wrist의 7개의 파라미터이며 양 팔에 대해서 모두 적용하므로 전체 14개가 된다.

3.2 3차 스플라인에 의한 보간

3차 스플라인 보간식은 각 키 프레임 사이의 애니메이션 경로를 식(2)와 같은 형태로 구하는 것을 말한다.

$$y_{ij} = ax_{ij}^3 + bx_{ij}^2 + cx_{ij} + d \quad \forall j, (1 \leq j \leq 54) \quad (2)$$

$$x_{ij} = \frac{k}{F_i} \quad k = 1, \dots, F_i \quad (3)$$

여기서, i : 키 프레임 번호, j : 파라미터 번호,

F_i : $i-1$ 번째 키 프레임과 i 번째 키 프레임 사이의 프레임 수.

y_{ij} 는 $i-1$ 번째 키 프레임과 i 번째 키 프레임 사이에서 j 번째 관절각 파라미터에 대한 3차 스플라인 보간식이다. 즉, 54개의 관절각 파라미터 각각에 대해 각 키 프레임 사이에 3차 스플라인 보간식이 생성되는 것이다. 여기서, y_{ij} 를 구하기 위해 식(4)를 이용하여 식(2)의 계수 a, b, c, d 를 구한다.

$$\begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot \\ 1 & 4 & 1 & 0 & \cdot & \cdot \\ 0 & 1 & 4 & 1 & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 0 & 1 & 4 & 1 \\ \cdot & \cdot & \cdot & \cdot & 0 & 1 \end{bmatrix} \begin{bmatrix} P_0 \\ P_1 \\ \cdot \\ \cdot \\ P_{m-1} \end{bmatrix} = \begin{bmatrix} P_0 \\ 3(P_2 - P_0) \\ \cdot \\ \cdot \\ 3(P_{m-1} - P_{m-3}) \\ P_{m-1} \end{bmatrix} \quad (4)$$

식(4)에 따르면, 키 프레임 파라미터 P_{i-1} 와 P_i 사이의 보간식 (2)를 구하기 위해, $P_0 \dots P_{m-1}$ 의 m 개의 키 프레임 파라미터를 필요로 하게 된다. 즉, 한 번에 전송되어온 키 프레임 파라미터의 개수가 5개일 경우 m 의 값은 5가 되며, 식(4)로부터 각 구간마다의 애니메이션 경로 식(2)을 $216 (= 4 \times 54)$ 개 구한다.

3.3 디스플레이 방법의 변화

가변 프레임 방식 및 스플라인 보간과 함께 애니메이션 과정을 바꾸어 디스플레이하는 속도를 향상시킬 수 있다. 기존의 방법은 그림 3(a)에 보인 바와 같이 애니메이션 중에 키 프레임 사이의 프레임을 보간하는 데 비해, 가변 프레임 방법에서는 (b)와 같이 애니메이션을 시작하기 전에 프레임 수를 결정하고, 키 프레임 사이에 들어갈 프레임을 모두 계산하여 프레임 큐에 저장한다. 그러므로, 애니메이션 모듈에서는 프레임 큐에 저장된 파라미터들을 순서대로 읽으면서 디스플레이하여, 애니메이션 속도를 향상시켰다.

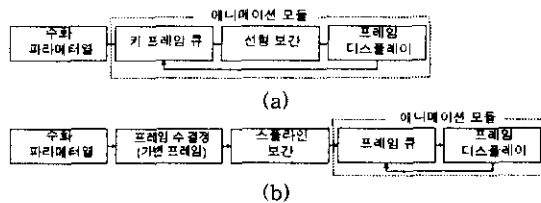


그림 3. 애니메이션 방법의 비교

- (a) 고정 프레임 방법
- (b) 가변 프레임 방법

Fig 3. Comparison of the two animation methods

- (a) Fixed number of frames
- (b) Variable number of frames

4. 실험 및 고찰

이 논문의 수화통신 시스템은 Pentium-233, 64MB RAM에서 Visual C++에 Open Inventor 라이브러리를 임포트시켜 구현하였다. 수화 애니메이션에 사용되는 3차원 인체모델은 "http://www.ywd.com"에서 VRML 파일 포맷으로 받은 것을 Open Inventor 2.0 파일 포맷으로 변환시켜 사용했다.

표 2는 키 프레임간 각도 변화에 대한 고정 프레임 방식과 가변 프레임 방식의 프레임 수와 각 프레임별 변화 각도를 비교한 것이다. 고정 프레임 방법에서는 모든 각도에 대해서 동일한 프레임 수를 적용하지만, 가변 프레임 방법에서는 각도 차에 따라 프레임의 수가 변화한다. 그리고, 한 프레임 당 움직이는 각도 변화를 비교해 보면, 고정 프레임의 경우 180° 일 때 25.7° 인데 비해 10° 에서는 1.4° 에 그치고 있어 그 차이가 심하다. 그러나, 가변 프레임의 경우에는 각각 13.8° 와 5° 로써 그 차이를 현저히 줄일 수 있다.

표 2. 키 프레임간 각도 변화에 따른 프레임당 변화량
Table 2. Variances per frame of different angles between key frames

키 프레임간 각도 변화	고정 프레임		가변 프레임	
	프레임수	변화각도	프레임수	변화각도
180°	7	25.7°	13	13.8°
135°	7	19.3°	10	13.5°
90°	7	12.9°	7	12.9°
30°	7	4.3°	3	10°
10°	7	1.4°	2	5°

표 3은 4개의 문장을 입력했을 때 고정 프레임 방법과 가변 프레임 방법에서 사용되는 프레임의 수를 비교한 것이다. 고정 프레임의 수는 표 3과 같이 모든 동작에 대해 동일한 프레임의 수를 할당하기 때문에 키 프레임 당 평균 7개가 소요된다. 그러나, 가변 프레임의 경우에는 평균 5.45개이다. 식(5)에 따라 시간 감소율 R 은 22.14% 정도임을 알 수 있다.

$$R = \frac{A-B}{A} \times 100 \approx 22.14(\%) \quad (5)$$

여기서, $A = 7 \times 0.3 = 2.1$

(고정 프레임 방식의 평균 디스플레이 시간)

$$B = 5.45 \times 0.3 = 1.64$$

(가변 프레임 방식의 평균 디스플레이 시간)

이와 같이, 고정 프레임에서 생기는 불필요한 프레임을 줄여 애니메이션 속도를 향상시킬 수 있었으며, 동작도 보다 자연스러운 것을 확인했다.

표 3. 문장별 프레임의 수 비교

Table 3. The comparison of frames in a sentence

단어	키프레임 입 수	고정 프레임	가변 프레임
나는 학교에 갑니다	7	49	48
선생님 안녕하세요	9	63	41
당신은 누구십니까	6	42	26
아침에 모두 기뻐하다	9	63	54
합 계	31	217	167
평균		7	5.45

그림 4의 (a)는 선형 보간식을 적용한 원 동작이며, (b)는 스플라인 보간식을 적용한 원 동작이다. 이 동작은 4개의 키 프레임을 가지고 있다. 그림 4의 (a)는 28 프레임이 소요되었고, (b)는 16 프레임이 소요되었다. 그림 3에서 보는 바와 같이 선형 보간식의 원 동작은 많은 프레임을 사용하여 다각형을 그리고, 스플라인 보간식은 비교적 적은 프레임으로 원 동작에 근사한 형태를 그리는 것을 알 수 있다. 선형 보간 공식에서 곡선 동작을 가지는 수화 동작을 자연스럽게 표현하기 위해서는 키 프레임을 6~8개 정도로 만들어 주어야 하지만, 스플라인 보간식을 이용하면, 3~4개의 키 프레임으로 보다 자연스러운 동작을 만들 수 있었다.

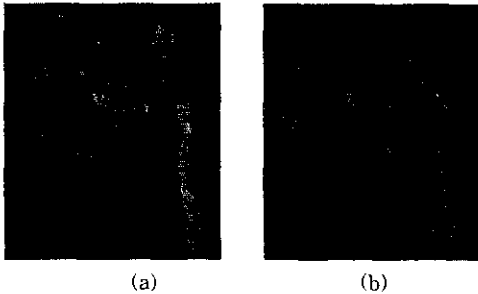
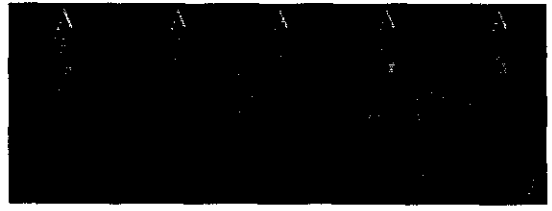


그림 4. "원" 동작의 비교
(a) 선형 보간식을 적용한 원 동작
(b) 스플라인 보간식을 적용한 원 동작
Fig 4. Comparison of "circle" motions
(a) Circular motion with linear interpolation
(b) Circular motion with spline interpolation

그림 5는 "나는 학교에 갑니다."에 대한 수화 애니메이션이다. 입력문 "나는 학교에 갑니다."는 수화 번역 단계에서 "나"+"학교"+"가다"의 수화 단어 열로 분석되고, 각 단어는 수화 파라미터로 변환된다. "나"에 대해 1개, "학교"에 대해 4개, "가다"에 대해 2개의 키 프레임 파라미터로 총 7개의 키 프레임이 생성된다. 시스템은 이들을 디스플레이 하는데 필요한 프레임 수와 애니메이션 경로를 계산하여 애니메이션을 수행한다.



[나] [학교] [가다]
그림 5. "나는 학교에 갑니다" 의 3차원 수화 동작
Fig 5. 3D animation of "I go to school."

5. 결론

이 논문에서는 수화 영상 통신에서 3차원 모델을 효과적으로 디스플레이하는 방법에 대해 연구하였다. 가변 프레임 방식과 3차 스플라인 보간식을 이용하여 수화 애니메이션을 좀더 자연스럽게, 실제와 유사하도록 구현하였다.

그러나, 수화 영상 통신을 위해 필요한 단어 사전의 구축이 미흡한 실정이다. 이를 확충하기 위해 수화 파라미터를 자동으로 생성하는 스크립트가 필요하다. 본 연구를 바탕으로, 역 운동학을 이용하는 스크립트를 개발하여 단어 사전을 확충하려 한다. 또한 사용자에게 친근감을 높일 수 있는 3차원 모델의 연구가 필요하다.

Acknowledgments

This work was partially supported by TAO(Telecommunications Advancement Organization of Japan) for the international collaborative research project on "International Collaborative Research on Communication Technique between Different Languages using Sign Language".

참고 문헌

[1] Ji-Young OH, Sang-Woon KIM, Shin TANAHASHI, Yoshinao AOKI, "A Sign Language Chatting System for Non-verbal Communication between Different Languages", Proceedings of ITC-CSCC'98, Sokcho, Korea, pp. 693 - 696, July 1998.
[2] 吳芝英, 金商雲, 靑木由直, "3차원 모델을 이용한 한-일 수화 영상통신 시스템의 구현", 대한전자공학회하계종합학술대회 논문집, vol. 21, no. 1, pp. 925-928, 1998. 6.
[3] "http://ece.uwaterloo.ca/~h-anim/spec.html"
[4] 이현찬, 채수원, 최영, 컴퓨터 그래픽스 및 영상 모델링, 시그마프레스, 1997.