지적 아바타 통신을 위한 얼굴 표정의 실시간 생성에 관한 검토

*李 容厚, *金 商雲, **青木 由直
*明知大学校 컴퓨터工学科, **北海道大学大学院工学研究科
전화 : 031-330-6222 / 핸드폰 : 016-9896-5678

An Empirical Study on Real-time Generation of Facial Expressions in Avatar Communications

*Dept. of Computer Science & Engineering, Myongji University
** Graduate School of Engineering, Hokkaido University
E-mail : yongwho@mju.ac.kr

Abstract

As a means of overcoming the linguistic barrier in the Internet cyberspace, recently a couple of studies on intelligent avatar communications between avatars of different languages such as Japanese-Korean have been performed. In this paper we measure the generation time of facial expressions on different avatar models in order to consider avata models to be available in real-time system. We also provide a short overview about DTD (Document Type Definition) to deliver the facial and gesture animation parameters between avatars as an XML data.

1. 서론

수화 제스처는 인터넷 사이버공간에서 개인을 상징하는 아바타(Avatar)의 통신 보조 수단으로도 사용될 수 있다. 특히 얼굴 표정을 가진 수화 영상은 단지 팔과 손의 움직임만을 보여주는 영상에 비해 이해하기가 쉽다. 따라서 3차원 아바타에 표정을 생성하기 위한 연구가 진행되고있으며, FACS (Facial Action Coding System)의 AU (Action Unit)를 주로 사용하고 있다.

논문 [1]에서는 구조 혹은 모양이 다른 아바타 모델에 얼굴 표정을 생성하기 위해서 볼, 턱의 영역을 추출하는 방법을 제안하였고, 이렇게 찾아낸 볼, 턱을 가지고 포물형 편미분식(Parabolic Partial Differential Equations)과 Runge-Kutta법을 이용하여 보다 자연스러운 표정을 생성할 수 있는 방법을 제안하였다.

본 논문에서는 아바타 모델의 폴리곤 수가 너무 많은 경우에, 자연스러운 표정 생성은 가능하지만 실-시간 처리가 어렵다는 점에 주목하여, 폴리곤의 개수에 따른 시간측정을 계산하는 등 실-시간 시스템에서 사용 가능한 아바타 모델에 대한 검토 결과를 보고한다. 또한, 그림 1.1처럼, 서로 다른 모델간에 표정 전달을 가능하도록 하기 위해서, XML 데이터를 이용하여 전달할 수 있도록 하는 DTD(Document Type Definition)을 제안한다.

그림 1.1. 다양한 Avatar들 간에 XML을 이용한 시스템

2. 관련 연구

Ekman과 Friesen[2]은 눈으로 식별 가능한 모든 얼굴의 변화를 표현할 수 있는 시스템을 만들었다. FACS(Facial Action Coding System)라고 하는 이 시스템은 얼굴 표정의 변화를 일으키는 기본적인 AU(Action Unit)들의 나열로 얼굴 표정을 기술한다. 특정 근육의 움직임은 하나 이상의 AU을 변화시키게 되므로, AU과 개개의 근육과의 상관관계는 절대적이

라고 할 수 없다. FACS에는 얼굴의 표정을 기술하는 46개의 AU들과 head orientation과 gaze를 표현하는 12개의 AU들을 합쳐서 총 58개의 AU들이 정의되어 있다.

Essa와 Pentland[3]은 기존의 FACS가 사람의 판단에 기반하는 heuristic 한 점을 지향하고자 얼굴의 근육구조와 피부의 물리학적 특정을 반영한 3차원 모델을 이용하여 muscle actuation을 직접 계산해 내고 이 값을 사용하는 새로운 표정 기술 시스템인 FACS+를 개발하고, 이를 이용해 표정을 분석하고 인식하였다.

Hong[4]은 머리카락과 안경이 없는 얼굴 사진을 이용하여, 실-시간처리를 가능하게 하며, 전체론적 방법과 템플릿 기반 방법을 사용한다. 주요 기술은 라벨된 그래프를 사용한다. 즉, 얼굴사진에 표시된 그래프를 사용하며, 배열을 구성하는 그래프의 각각의 노드를 jet라고 칭한다. 두 가지 다르게 표시된 그래프를 GFK(General Face Knowledge) 사용하며, big GFK는 20개의 노드로 표시된 그래프이다. 그리고 small GFK는 입력 얼굴 이미지에서 추출한 얼굴 위치를 찾는다.

Huang[5]는 정면 사진 하나를 사용하며, 머리카락과 안경이 없는 얼굴 모델과 배경 변화가 없는 모델을 사용한다. 주요 기술은 PDM(Point Distribution Model)을 사용하고, PDM은 90개의 얼굴 특징 포인트를 정의한다. 또한 6개의 기본 감정을 15명 중국인의 90개 이미지들을 수작업으로 지역화시킨다.

Wang[6]은 정면 사진을 사용하며, 머리카락과 안경이 없는 얼굴을 이용한다. 첫 번째 프레임을 수작업으로 레벨링을 해 주어야 한다. 전체론적 방법과 분석적 방법을 혼용한 하이브리드 방법을 사용하며, 템플릿 기반 방법을 사용한다. 주요 기술은 라벨링된 그래프이다. 19개 얼굴 특징 점들인 FFPs(Facial Feature Points)를 사용하며, 19개 점 중에서 비어 있는 점(◎)인 7개를 제외한 12개 검은 점(●)은 얼굴 표정 인식을 위해 사용한다.

3. 시뮬레이션을 위한 얼굴 모델

실험용 얼굴모델은 3차원 얼굴 모델에 추가적인 요소를 더하여 새롭게 얼굴 모델을 만들었다. 기존의 472개와 4288개 폴리곤을 가지고 있는 3차원 얼굴 모델의 폴리곤 메쉬를 분할하여 메쉬 조밀도를 높였다.

그림 3.1처럼, 삼각 메쉬 분할은 메쉬를 이루는 세 포인트를 기준으로 서로 중간이 되는 점을 연결하여, 하나의 삼각 메쉬가 네 개의 삼각 메쉬로 분할이 된다.

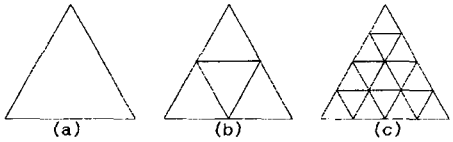

그림 3.1. 삼각 메쉬 분할. (a) 분할하지 않은 폴리곤, (b) 4배 분할한 폴리곤, (c) 16배 분할한 폴리곤

폴리곤 분할을 많이 할 수록 얼굴 피부 모델링의 섬세함과 자연스러움은 더욱 좋지만, 데이터 용량이 많아져서 에니메이션의 속도가 저하된다. 따라서 본 실험에서는 4배 분할까지만 고려한다.

4. 얼굴 표정의 생성 과정

얼굴 영상을 실시간으로 디스플레이하기 위해서는 가능한 적은 수의 제어 점으로 구성되는 모델 영상이 필요하다. 기존의 표정 생성 방법에서는 사진과 같은 사실적인 감정 표현을 위해 PAU6 (빰을 올린다)과 같이 얼굴 근육의 미세한 움직임을 표현하기 위해 많은 AU (PAU+CAU)의 조합이 필요하다. 그림 4.1은 AU 추출과 표정생성을 위한 절차를 나타낸다.

그림 4.1에서 "3D Model"은 점과 선, 그리고 폴리곤 정보로 구성되어 있다. 각각의 점들이 연결되면 선이 되고, 또한 인접한 선들이 연결되면 폴리곤이 된다. 여기서 "Coordinate Information"은 "3D Model"에서 점에 대한 정보를 이용하여 위치적 정보를 파악할 수 있으며, "Color Information"은 폴리곤이 갖고 있는 색상 정보를 이용한다. 또한 "Polygon Information"은 특정한 어떤 폴리곤에서 인접한 폴리곤의 연관성이 있는지를 파악한다. 지금 까지의 "Coordinate", "Color", 및 "Polygon Information"을 이용하여 특정 영역을 찾아낼 수 있다. 그렇지만, 찾아낸 영역들이 눈썹, 눈꺼풀, 그리고 입술 영역인지를 알 수 없다. "AU, PAU, and their Area Extraction"에서는 지금 까지 추출된 영역들을 그룹핑한다. 그룹핑은 눈썹과 눈꺼풀, 입술 영역을 구별하는데 필요한 과정이다. 또한 찾아낸 AU 영역과 관련된 주변 영역을 추출해야 한다. "Check, Jaw, and Nose Extraction"에서는 턱과, 볼, 그리고 볼 영역을 추출한다. 마지막으로, "Expression Generation"에서 지금까지 찾아낸 영역을 가지고 턱과 볼에 대한 움직임대한 공식을 이용하여 표정을 생성한다.

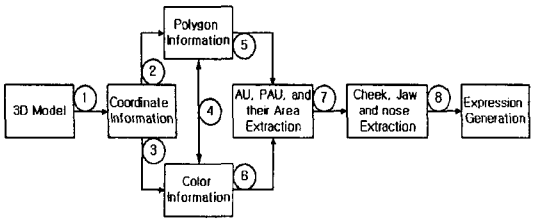

그림 4.1. AU 추출 및 얼굴 표정 생성 절차.

5. XML 데이터 생성을 위한 DTD 설계

DTD를 설계하기 위해서는 가장 일반적으로 사용되는 규칙중에 하나는 head와 body로 처음에 나누어야 한다. 그리고 body 부분에 전달하고자 하는 내용을 기술하게 된다. 최상위 엘리먼트를 FBML(Facial Body Markup Language)라고, 이것의 서브 엘리먼트는 fbml_head와 fbml_body로 나눈다. 그림 5.1은 fbml_body 부분의 서브 엘리먼트 구조이다.

그림 5.1. FBML을 위한 DTD 구조.

6. 실험

먼저 표정 에디터를 이용하여 다양한 종류의 감정과 강도를 편집하는 실험을 하였다.

표 6.1은 AU6번을 사용하지 않고 "Joy"표정을 생성한 표정은 그림 6.1이며, 코믹한 표정을 생성하기 위해서 AU6번을 사용하여 "Joy"표정을 생성한 것은 그림 6.2이다.

표 6.1. "Joy" 표정을 생성하기 위한 AU 강도 값.

표정	AU번호(강도 값)
Joy	AU-1(100%), 2(100%), 6(160), 12(156%)

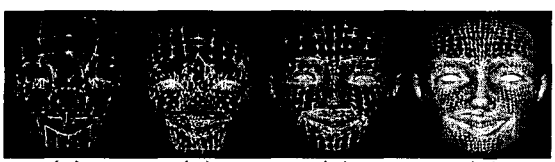

그림 6.1. AU6번을 사용하지 않은 Joy 표정 이며,
(a) 472개, (b)는 1888개, (c)는 4288개,
(d)는 17152개 폴리곤을 가진 모델

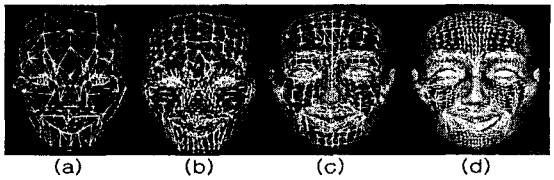

그림 6.2. AU6번을 사용한 Joy 표정 이며,
(a) 472개, (b)는 1888개, (c)는 4288개,
(d)는 17152개 폴리곤을 가진 모델

그림 6.3은 서로 다른 버텍스를 가진 모델에서 볼을 움직이는데 필요한 시간을 측정한 것이다. 여기서, 버텍스가 3500개 버텍스를 가진 모델에서 시간이 많이 필요한 것을 알 수 있다. 또한 실험치로 한쪽 볼을 40등분하여 포물형 편미분식과 대입하였다. 그러나 버텍스가 많은 모델인 경우, 대략 20등분해서 볼을 움직이는 방법으로 시간을 줄일 수 있었다.

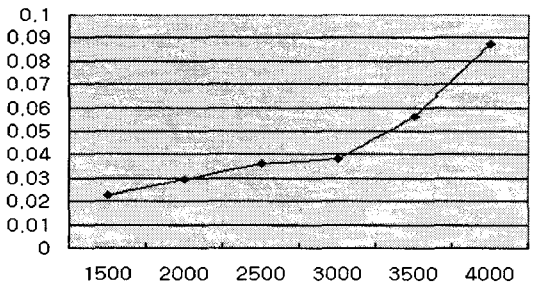

그림 6.3. 서로 다른 버텍스를 가진 모델에서 볼의 움직임 시간 측정

기존의 수작업으로 턱을 움직였을 경우 턱 기준 축을 중심으로 하지 않고, 턱을 내렸기 때문에 턱 전체가 내려오는 현상이 발생하였다. 그러나 Runge-Kutta법을 이용하면 턱의 움직임을 수작업으로 생성할 때 보다 자연스러운 움직임을 생성할 수 있었다.

그림 6.4는 다양한 모델에 Runge-Kutta법을 이용하여 시간을 측정하였다. 또한 Runge-Kutta법에는 4차 까지 있으므로, 1차에서 4차까지 나누어서 시간을 측정하였다. 모델마다 1차와 4차 사이의 결과를 비교한 결과, 뚜렷한 차이를 느낄 수 없었다. 실험에서 3500개 이상의 모델에서는 시간이 많이 소비되는 것을 볼 수 있다. 그러므로 버텍스가 많은 모델일 경우에는 4차에서 1차로 줄이는 것이 턱을 움직이는데 효율적이다. 또한 버텍스가 적은 모델인 경우에는 Runge-Kutta법에 4차까지 사용하여도 시간 소비는 적다.

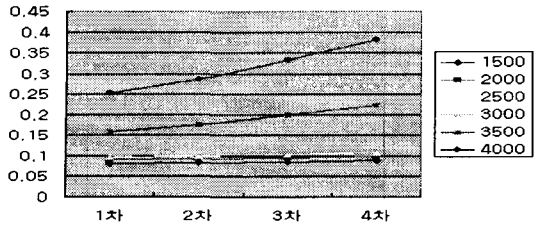

그림 6.4. 다양한 모델에 턱의 움직임 시간 측정

표 6.2에서 "모델을 로드하는 시간"은 그림 4.1에서 ①번을 수행한 시간을 의미하며, "그룹핑하는 시간"은 ⑦번까지 수행한 시간, "외각 부분을 계산 시간"은 ⑧번까지의 시간, "턱 부분 추출 시간"과 "볼 부분 추출 시간"은 ⑨번까지의 시간이며, 마지막으로 "표정 생성된 최종 시간"은 "Expression Generation"까지 수행한 결과이다.

표 6.2. 모델 종류에 따른 시간 측정 결과(단위:밀리초)

모델의 종류 / 표정 생성 과정	472개 폴리곤	1888개 폴리곤	4288개 폴리곤	17152개 폴리곤
모델을 로드하는 시간	66.59	264.26	967.29	18572.30
그룹핑하는 시간	84.11	429.37	1935.30	33857.23
외각 부분을 계산 시간	87.60	451.37	2102.41	82816.37
턱 부분 추출 시간	90.14	482.59	2146.31	83723.13
볼 부분 추출 시간	93.01	512.99	2237.03	85633.77
표정 생성된 최종 시간	96.01	535.69	2300.92	86944.92

그림 6.5에서 수행 시간의 대부분은 "그룹핑하는 시간"과 "외각 부분을 계산 시간"이 차지하고 있다. 그러므로 이 두 가지 과정의 수행 시간을 단축시킴으로써 시스템 성능을 높일 수 있다.

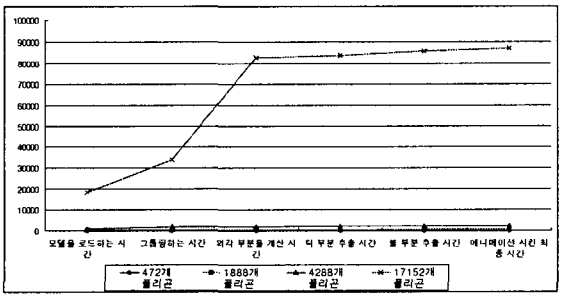

그림 6.5. 모델의 종류에 따른 시간 측정 결과 그래프

또한 다른 모델에 비해서 17152개 모델에 대한 수행 시간이 가장 길었다. 따라서 폴리곤의 수가 대략 10,000개를 넘으면 자연스러운 얼굴 표정 생성은 가능하지만, 실-시간 처리는 불가능함을 알수있고, 시스템에 용도에 따라서 융통성이 필요하다.

7. 결론

본 논문에서는 포물형 편미분식과 Runge-Kutta법을 사용한 얼굴 표정 생성법을 다양한 모델에 적용할 수 있음을 실험하였다. 실험 결과, 볼의 움직임의 경우, 폴리곤의 수를 줄이므로, 동작 시간을 줄일 수 있음을 보았고, 또한 턱의 움직임의 경우, 버텍스가 많은 모델에서는 Runge-Kutta법의 4차에서 1차로 줄이므로, 동작에 대한 시간을 볼 수 있었다.

이 논문에서는 속도와 자연스러움에 대한 임계치를 알아보고, 전체 표정 생성과정을 6분류로 나누어 시간을 측정하였다. 실험에서 폴리곤수가 10,000개를 넘는 경우, 실-시간 거리에서 사용이 어렵다는 것을 알았다.

또한 XML 데이터를 만들기 위해 DTD 문서를 제안하였으며, XML 문서를 이용함으로써 다양한 모델에 의사 전달이 가능하다. 그러나, 본 논문에서 현재 제시되고 있는 아바타 모델은 Texture를 지원하지 않으므로 실제 사진과 같은 표정을 나타내기 어렵다.

참고문헌

[1] 李容厚, 金商雲, 靑木山直, "인터넷 가상공간에서 지적 아바타 통신을 위한 코믹한 얼굴 표정의 생성법", 대한전자공학회논문지-S, vol. 40, no. 1, pp. 59-73,, 2002.

[2] P. Ekman and W.V. Friesen, Unmasking the Face, New Jersey: Prentice Hall, 1975

[3] I. Essa and A. Pentland, "Coding, Analysis Interpretation, Recognition of Facial Expressions," IEEE Trans, Pattern Analysis and Machine Intelli., vol, 19, no.7, pp.757-763, 1997.

[4] H. Hong, H. Neven, and C. von der Malsburg, "Online Facial Expression Recognition Based on Personalized alleries", Proc. Int'l Conf. Automatic Face and Gesture Recognition, pp. 354-359, 1998.

[5] C.L. Huang and Y.M. Huang, "Facial Expression Recognition Using Model-Based Feature Extraction and Action Parameters Classification",J. Visual Comm. and Image Representation, vol. 8, no. 3, pp. 278-290, 1997.

[6] M. Wang, Y. Iwai, and M. Yachida, "Expression Recognition from Time-Sequential Facial Images by Use of Expression Change Model",Proc. Int'l Conf. Automatic Face and Gesture Recognition, pp. 324-329, 1998.