MPEG-4 SNHC 표준을 따르는 3차원 애니메이션 시스템의 구현

윤 승 욱, 안 정 환, 전 정 희, 호 요 성
광주과학기술원 정보통신공학과
전화 : 062-970-2263 / 팩스 : 062-970-2247

Implementation of A 3-D Animation System Based on The MPEG-4 SNHC Standard

Seung-Uk Yoon, Jeong-Hwan Ahn, Jeonghee Jeon, Yo-Sung Ho
Kwangju Institute of Science and Technology (K-JIST)
E-mail : suyoon@kjist.ac.kr

Abstract

In this paper, we propose a 3-D animation system to track and analyze motion of the human object. The proposed system consists of two separate layers: motion analysis layer and 3-D model registration layer. Following the MPEG-4 SNHC standard, we generate object motion using body definition and animation parameters. In the implemented system, we acquire human motion data from a single camera and extract body definition parameters from arbitrary VRML human models.

I. 서론

최근에 컴퓨터 그래픽스 기술의 발달로 3차원 아바타 (avatar)와 3차원 애니메이션에 대한 관심이 고조되고 있다. 3차원 애니메이션 시스템은 크게 3차원 정보를 입력받는 부분과 이를 처리하여 화면에 보여주는 부분으로 나뉜다. 기존의 3차원 애니메이션 시스템은 데이터 입력을 위해 접촉 장치인 모션 캡쳐 장비나 데이터 글러브 등을 주로 사용해 왔고, 이를 처리하는 부분도 특정 하드웨어와 소프트웨어에 의존적이거나 개별적인 방법을 사용해 왔다. 하지만, 이런 입력 장치들은 고가이며, 사용자의 편의를 고려하지 않아 조작하기에도 복잡하다. 또한, 입력된 데이터를 처리하는 방식도 개발 환경에 따라 매우 다양하고, 일관된 기준이 없다. 따라서, 저가이면서 사용자의 편의를 고려하는 입력 장치를 이용하고, 입력 데이터를 처리할 때 일관된 기준을 제공할 수 있다면 이런 문제점을 크게 개선할 수 있다.

본 논문에서는 컴퓨터 비전 장비인 카메라를 이용하여 입력을 처리하고, 인체의 움직임을 MPEG-4 SNHC의 Face and Body Animation (FBA) 표준에 맞게 해석하기 위해 FBA의 Body Definition Parameter (BDP)와 Body Animation Parameter (BAP)를 분석하였다. 그리고 이를 바탕으로 움직임 매개변수를 생성함으로써 기존 3차원 애니메이션 시스템이 가지는 입력 장비의 문제점과 시스템의 호환성 문제를 해결하였다. 또한, 구체적으로 신체 일부인 머리의 움직임을 사용하여 시스템이 성공적으로 동작함을 확인하였다. 특히, 제안된 시스템은 이미 구축되어 시스템에 포함된 특정한 모델뿐만 아니라, 일반적인 Virtual Reality Modelling Language (VRML) 신체 모델이면 어느 것에든 적용할 수 있도록 구현되었다.

II. 3차원 애니메이션 시스템 구현

본 장에서는 제안된 시스템의 전체적인 구조와 각 계층별 구현 방법에 대해 살펴본다.

2.1 3차원 애니메이션 시스템의 기본 구조

제안된 시스템은 크게 두 계층으로 분류될 수 있다. 첫 번째 계층은 움직임을 분석하는 부분으로 카메라로부터 사람의 동작을 입력받아 특징을 추출하고, 이를 추적하여 움직임 매개 변수를 생성해 낸다. 두 번째 계층은 임의의 VRML 사람 모델을 입력으로 받아, 이를 분석하여 신체를 정의하는 매개변수를 생성한다. 그런 후에 움직임 매개변수와 신체 정의 매개변수를 처리하여 최종적으로 화면에 3차원 신체 모델의 움직임을 표현한다.

그림 1은 제안된 시스템의 전체적인 기능 구조를 그린 것이다.

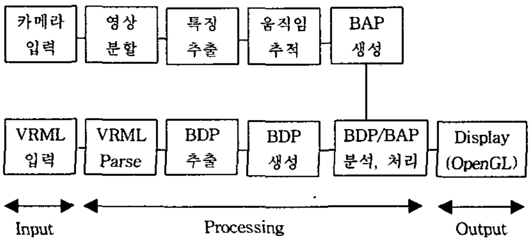


그림 1. 3차원 애니메이션 시스템의 기능 구조

2.2 움직임 분석 계층

(1) 입력 인터페이스

움직임 분석 계층에서는 우선 카메라로부터 신체 일부분의 움직임을 입력 받는다. 제안된 시스템에서는 저가의 USB PC 카메라를 입력 인터페이스로 이용하였다. 시스템이 카메라를 인식하고, 입력 영상의 규격을 확인한 후, 카메라를 초기화한다. 카메라 입력영상은 YUV 형식이며, 다음 단계에서의 영상 분할과 특징 추출을 위해 YUV 영상을 RGB 형태로 변환한다.

그림 2는 입력 인터페이스를 나타낸 것이다.

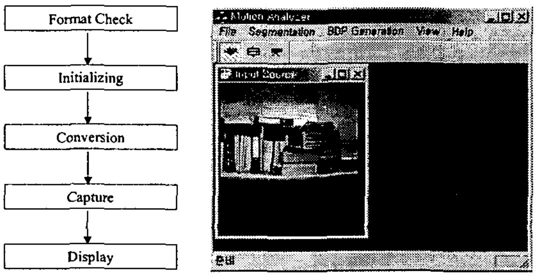


그림 2. 입력 인터페이스

(2) 영상 분할 (Background Subtraction)

입력 영상을 받은 후에는 움직이는 물체와 배경을 분리할 필요가 있다. 영상분할 방법은 여러 가지가 있지만, 본 논문에서는 감시 시스템 등에 주로 이용되는 간단한 자동분할 방식을 사용하였다. 즉, 배경화면과 새롭게 입력된 움직이는 물체 (foreground object)를 분리하기 위하여, 기준 시퀀스 (reference sequence)를 미리 저장하여 이의 통계적인 특성을 모델링하였다.

그림 3은 본 논문에서 사용된 영상 분할 과정을 나타낸 것이다.

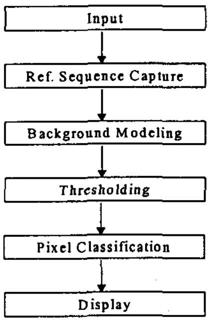


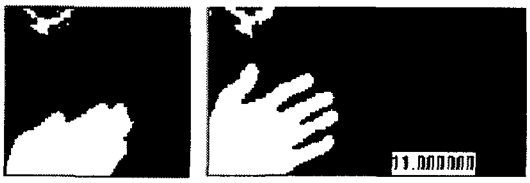
그림 3. 영상 분할 과정

$$\mathrm{sqrt}(\sigma_i) > \alpha * \mathrm{sqrt}(\sigma r) \quad (1)$$

여기서 σ는 $\sigma_R^2 + \sigma_G^2 + \sigma_B^2$, 즉, 화소의 채널별 분산의 합을 나타내며, σr은 기준 시퀀스의 분산, σi는 새로운 입력 영상의 분산을 표현한다. 식(1)에서 보는 바와 같이, 우선 기준 시퀀스에 속한 각 영상의 평균과 분산을 R, G, B 채널에 대해 계산한다. 이를 화소별로 시퀀스의 모든 영상에 대해 평균한 후 여기에 임계값 (threshold) α를 곱한다. 새로 입력된 영상 화소에 대해서도 채널의 평균 분산을 계산하여 이 값이 식(1)의 우변보다 클 경우, 이 화소는 움직이는 물체에 속하였다고 판단하고 이를 화면에 보여주게 된다. 각 화소가 움직이는 물체로 판단되면 흰색으로, 배경으로 판단되면 검은색을 사용하여 분할된 영상을 표시하였다.

그림 4는 분할된 결과를 나타낸다. 움직이는 물체가 배경으로부터 비교적 양호하게 분할된 것을 확인할 수 있다. 하지만, 영상분할 과정을 통해 얻고자 하는 물체는 신체의 일부인 손인데 비해, 위쪽에 흰색으로 표시된 부분처럼 원하지 않는 영역이 나타날 확률이 높다. 이는 예측하지 못한 반사 효과, 주변 광선의 변화, 갑작스러운 신체의 움직임, 그림자 등의 원인에 의

해 발생될 수 있다. 그림 4의 위쪽 부분은 손에 형광등 빛이 반사되어 움직이는 물체로 오판된 영역이다. 그림 4(b)에 표시된 숫자는 현재 프레임에서의 임계값을 나타낸다. 이와 같이 원하는 물체만을 정확히 분할하는데 문제점이 있기 때문에, 특징 추출 부분에서 이를 보완하기 위한 방법을 제안하였다.



(a) (b)

그림 4. 분할된 손 영상과 주변 반사광의 영향

(3) 특징 추출 (Feature Extraction)

분할된 영상으로부터 움직임 추적을 위한 기본적인 정보를 얻는 과정이 특징 추출 단계이다. C.L. Huang [1]은 분할된 영상에 대해 모폴로지 연산을 적용하여 잡음을 제거하고, 이를 각각 X와 Y 방향으로 투사시킨 뒤 투사된 길이의 평균을 계산하여 중심점을 구했다. 또한 이를 기준으로 신체 각 부분의 비율을 정의하였다. 이 방법은 모폴로지 (morphology) 연산을 통해 잡음을 완벽히 제거할 수 있다는 전제로부터 출발했는데, 앞서 언급한 원인들로 인해 어려운 점이 있다. 따라서 본 논문에서는 윤곽선 (contour)을 이용하는 접근법을 채택하여 이 문제의 해결책을 제시하였다.

제안된 시스템은 우선 분할된 영상에서 모든 개체의 윤곽선을 검출한다. 윤곽선은 체인 코드를 이용하여 저장된다. 이렇게 저장된 윤곽선 정보를 이용하여 윤곽선 모멘트 (contour moment)를 계산하면, 식(2)에 의해 각 윤곽선의 면적을 구할 수 있다 [2].

$$V_{pq} = \iint_R x^p y^q \, dxdy \qquad (2)$$

이때 차수 (p;q)의 모멘트는 식(2)에 의해 표현되며, 임의의 윤곽선 영역 R의 면적은 p=q=0일 때 얻어진다. 기본적으로 분할된 신체 부분이 가장 많은 면적을 차지하므로 면적을 계산한 후, 특정 임계값 이상의 면적만을 선택하면 오판된 부분을 제거할 수 있다.

본 논문에서는 이렇게 선택된 신체 부분의 중심점을 결정하기 위해, 윤곽선 영역의 무게중심을 계산하였다. 무게중심은 식(3)을 통해 구할 수 있다.

$$x_c = M_{10}/M_{00},\ y_c = M_{01}/M_{00} \qquad (3)$$

여기서 M_{00}는 0차 모멘트이며, M_{10}, M_{01}은 1차 모멘트이다. 식(2)와 식(3)을 통해 계산된 면적과 무게 중심을 그림 5에 나타내었다.

Area = 3336.500000

Mass center (Centroid)

(x = 122.390029, y = 88.497827)

그림 5. 윤곽선의 면적과 무게 중심

(4) 움직임 추적과 BAP 생성

본 논문에서는 신체의 일부인 머리의 움직임을 추적하여 움직임 매개변수를 생성하였다. 우선, 목에는 관절 (joint)이 하나만 존재하므로 머리의 움직임은 지역적인 관절의 움직임이 아니라 전체적인 움직임 (global motion)으로 볼 수 있다. 따라서 식(4)을 통해 머리를 구성하는 화소 분포의 방향성을 계산함으로써 머리가 움직이는 방향과 각도를 얻는다 [3].

$$\theta = \frac{\arctan\left(\dfrac{2\left(\dfrac{M_{11}}{M_{00}} - x_c y_c\right)}{\left(\dfrac{M_{20}}{M_{00}} - x_c^2\right) - \left(\dfrac{M_{02}}{M_{00}} - y_c^2\right)} \right)}{2} \qquad (4)$$

여기서 M_{20}, M_{02}는 2차 모멘트이고, x_c와 y_c는 식(3)에서 계산된 무게중심이다.

이렇게 식(4)에 의해 계산된 각도를 이용하여 BAP의 매개변수를 생성한다. BAP의 실제 값은 식(5)를 통해 얻을 수 있으며, 움직이도록 설정된 관절은 BDP에 정의된 목 관절이다.

$$BAP = angle(deg) \times 100000/PI \qquad (5)$$

2.3 3차원 모델 등록 (Registration) 계층

MPEG-4 SNHC 그룹에서 제시한 FBA에 관한 표준안에서는 3차원 아바타를 얼굴(Face)과 몸통(Body)으로 구분하여, 이를 각각 Face Definition Parameter (FDP)와 BDP로 정의하고 있다.

(1) VRML 파서 (Parser) 및 BDP 생성

제안된 시스템은 미리 입력되어 저장된 특정한 모델에 적용되는 것이 아니라, 임의의 VRML 신체 모델을 입력받을 수 있도록 설계되었다. 입력 가능한 모델은 관절과 세그먼트 (segment)로 구성된 사람 모델이다. 파일을 입력 받으면, 관절의 이름과 위치정보, 세그먼

트의 이름, 그리고 사람 모델을 구성하는 기하학 정보 (geometry data)와 연결성 정보 (connectivity data)를 추출한다. 그리고 추출된 정보를 바탕으로 BDP를 생성한다.

(2) BAP 파서 및 분석

신체의 움직임 추적을 통해 생성된 BAP가 입력되면 그림 6과 같은 과정을 거쳐 3차원 모델을 애니메이션 하는데 필요한 정보들을 추출한다. 우선 프레임 단위로 BAP가 입력되면 시스템 내부에는 BAP 관계 표가 불러진다. 이 표는 각종 세그먼트의 이름과 번호, 관절의 이름과 번호, BAP의 종류 등의 정보를 분석하여, 이들과 BAP와의 관련 정보를 나타낸다. 그리고, 매 프레임의 BAP를 읽을 때마다 이 표를 이용해 움직이도록 설정된 BAP를 찾고, BAP의 유형과 실제 움직이는 정도를 나타내는 값을 읽는다 [4].

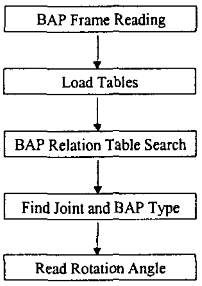


그림 6. BAP 파싱 (Parsing) 및 분석 과정

(3) 디스플레이 (Display)

마지막으로 이제까지 처리된 3차원 아바타의 정보와 움직임 정보를 이용하여 화면상에 모델의 움직임을 보여준다. 제안된 시스템에서는 OpenGL을 이용하여 3차원 아바타의 움직임을 표현하였다 [4].

III. 실험 결과

신체의 일부인 머리를 좌우로 움직이는 동작을 통해 본 논문에서 구현한 시스템의 동작을 검사하였다. 움직임 추적을 통해 만들어진 BAP는 90 프레임으로 고개를 좌우로 움직이는 동작을 기술한다.

그림 7에 실제 구현된 모델의 움직임을 나타내었다. 첫 번째 프레임은 정면을 바라보는 기본자세이다. 30번째 프레임은 고개를 왼쪽으로 돌린 상태이며, 90번째 프레임은 고개를 오른쪽으로 돌린 상태이다.

그림 7. 머리 움직임을 이용한 시스템 동작 확인

IV. 결론

본 논문에서는 한 대의 카메라를 이용하여 신체의 움직임 정보를 입력하여 처리하고, 신체의 움직임을 MPEG-4 SNHC의 FBA 표준에 맞게 해석하고 처리하였다. 또한, 이를 바탕으로 움직임 매개변수를 생성함으로서 기존 3차원 애니메이션 시스템이 가지는 입력장비의 문제점과 시스템의 호환성 문제를 해결하였다. 특히, 제안된 시스템은 이미 구축되어 시스템에 포함된 특정한 모델뿐만 아니라, 일반적인 VRML 신체 모델에 적용할 수 있도록 구현되었다. 추후에는 신체의 일부분이 아닌 전신으로 적용 범위를 확대하고, 사실적인 움직임을 위해 스테레오 비전을 이용해 3차원 데이터를 얻는 부분과의 연계에 대해 연구할 예정이다.

감사의 글

본 연구는 광주과학기술원 (K-JIST) 초고속광네트워크 연구센터 (UFON)를 통한 한국과학재단 우수연구센터 (ERC)와 교육부 두뇌한국21 (BK21) 정보기술 사업단의 지원에 의한 것입니다.

참고문헌

[1] C.L. Huang and C.C. Lin, "Model-Based Human Body Model Analysis for MPEG IV Video Encoder", Proceedings of ITCC, pp. 435-439, 2001.
[2] A.K. Jain, Fundamentals of Digital Image Processing, Prentice Hall, 1989.
[3] G.R. Bradski, "Computer Vision Face Tracking for Use in a Perceptual User Interface," Intel Technology Journal Quarter2, 1998.
[4] 안정환, 김성열, 허영, 한규서, 전병태, 호요성, "MPEG-4 SNHC 기반의 Body Animation 구현," 대한전자공학회 광주, 전남지부 2001년도 학술발표회 논문집, pp. 139-142, 2001.