



MPEG-4 SNHC에서 얼굴과 몸체의 Animation에 관한 기술 동향

호요성 광주과학기술원 정보통신공학과 교수
안정환 광주과학기술원 정보통신공학과 박사과정

1. 서론

최근 오디오/비디오 부호화 기술과 컴퓨터 그래픽 기술이 점차 융합되어, 카메라로 입력된 자연 영상과 컴퓨터에 의해 생성된 합성 영상이 혼합된 복합 영상에 대한 영상 서비스의 요구가 증대되고 있다. 또한 인터넷상에서의 멀티미디어 서비스가 보편화됨에 따라, 삼차원 컴퓨터 그래픽 정보를 네트워크를 통해 송수신하는 공동 설계 작업, 원격 게임, 원격 교육 등과 같은 대화형 그래픽 응용 서비스가 점차 증가하고 있는 추세이다.

이에 대응하는 새로운 표준방식의 요구에 따라 현재 MPEG-4라 불리는 새로운 Aural/Visual 표준화 작업이 진행 중에 있는데, 특히 MPEG-4의 SNHC (Synthetic/Natural Hybrid Coding)는 기존의 A/V 부호화에서 탈피하여 좀더 현실감 있는 삼차원 영상과 가상 현실 등 자연적이거나 인공적으로 만든 Aural/Visual 정보의 융합에 목적을 두고 있다. 현재 이것의 응용분야로는 가상 원격 회의, 원격 쇼핑, 원격 진료,

원격 교육, 가상 스튜디오 등 다양하다. MPEG-4 SNHC의 궁극적인 목표는 다중 사용자가 네트워크를 통한 3D 가상 환경에서 소리와 시각으로 서로 통신을 하며, 가상 환경에 있는 멀티미디어 정보를 가지고 서로 상호 작용을 하는 것이다. 즉 사용자가 이 가상 환경에서 그래픽 적인 표현 방법에 의해 하나의 일부분으로 융화가 되어 마치 사람처럼 표현되어야 한다. 이러한 3차원 가상 환경에서 사용자는 가상의 얼굴과 몸으로 만든 움직임은 아바타(Avatar)로 표현되어 가상 환경의 일부분이 된다. 각 사용자는 가상 환경에서 그들의 움직임, 얼굴의 표현, 행동들을 서로 볼 수가 있고 다양한 형태의 물체가 그와 같은 가상 환경 안에 있게 된다. 위에서 언급한 여러 가지 3차원 가상환경 응용분야에 사용자를 대신하는 아바타를 구현하기 위해서는 Facial and Body Animation 등의 중요한 기술이 집약되어야 한다. 본 고에서는 MPEG-4 SNHC의 Face and Body Animation에 대한 기술 동향을 살펴해보도록 한다.

2. Face Animation

SNHC에서는 인공적으로 얼굴을 표현하기 위해 Face object를 정의하였는데, Face object는 렌더링을 위해 3-D polygon mesh의 형태로 얼굴을 만들어 낸다. 얼굴의 모양, 피부와 표정은 일반적으로 FDP(Facial Definition Parameter)나 FAP(Facial Animation Parameter) 들을 포함하는 비트열에 의하여 조정된다. 즉 FDP와 FAP은 얼굴의 모양과 피부 등을 정의하고, 얼굴의 표정과 감정, 말을 하는 등의 움직임 등을 정의하도록 만들어져 있다. 처음에 얼굴을 구성할 때는 무표정한 일반적인 얼굴을 기본으로 한다. 또한 비트열로부터 즉시 얼굴의 움직임을 제공할 수 있는 FAP들을 수신할 수 있는 능력을 가진다. 즉 표정, 말소리 등에 따라서 반응할 수 있다. 만약 FDP들이 수신되면 그것들은 일반적인 얼굴로부터 얼굴 모양이나 선택적으로 피부색에 의하여 결정되는 특정한 얼굴로 변환하는데 사용되어야 한다. 선택적으로 완전한 얼굴 모델은 얼굴 노드에 삽입을 위한 영상으로서 FDP 집합을 통하여 전송되기도 한다.

2.1 Facial Animation Parameter Set

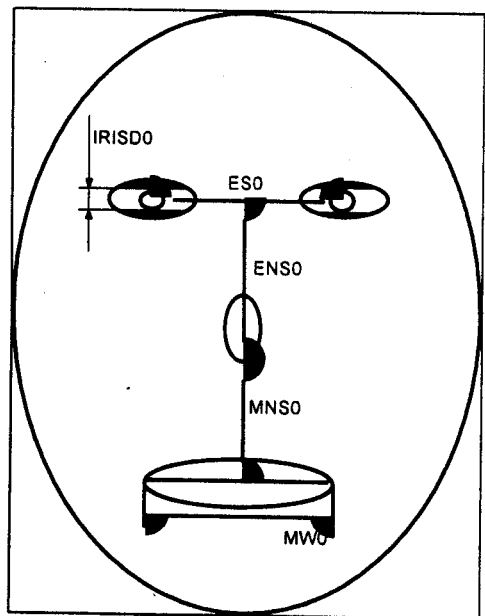
FAPs는 얼굴 움직임의 이론에 기반을 두고 근육의 움직임과 밀접한 관계가 있다. 그것들은 모든 기본적인 움직임들을 나타내므로 대부분의 실제 얼굴 표정을 다 표현할 수 있다. 만일 가능하지 않은 값들이 정의된다면 일반적인 사람들의 실제 움직임을 보이지 않고 마치 만화와 같은 특징을 보이게 된다.

FAPs은 2가지 상위 레벨 파라미터인 Viseme와 Expression을 갖는다. Viseme는 일정한 유한 소의 음 단위인 음소(Phoneme)와 상관관계가 있다. Viseme과라

메터는 Viseme 렌더링 (다른 파라미터의 관점에서 그것들을 표현하지 않고) 하고 Viseme렌더링은 계속 올바르게 유지하면서 다른 파라미터들의 결과를 더 향상시킨다. 확연하게 구분되는 Static viseme는 현재 표준안에 정의되어있고 추가되는 Viseme 파라미터는 다음에 계속 추가 될 것이다. 비슷하게 Expression 파라미터는 상위레벨의 얼굴 표정을 정의한다.

2.1.1 Facial Animation Parameter Units(FAPU)

움직임을 포함하여 모든 파라미터들은 FAPU(Facial Animation Parameter Units)로서 나타내진다. 이러한 단위들은 적당한 얼굴의 표정과 말의 발음 등을 만들기 위하여 일관된 방법으로 어떤 얼굴 모델에 FAPs를 해석하기 위하여 정의된 것들이다. FAPUs는 그림 1에 나와 있다. 그것들은 얼굴에서 특별하게 중요한 특징들 사이의 거리를 나타낸다.



[그림 1] 얼굴 움직임 파라미터 단위

IRISD0 = 홍채의 지름(위 아래의 눈꺼풀사이의 거리와 같다); IRISD = IRISD0/1024
 ES0 = 눈 사이의 거리; ES=ES0/1024
 ENS0 = 눈과 코 사이의 거리; ENS = ENS0/1024
 MNS0 = 입과 코 사이의 거리; MNS = MNS0/1024
 MW0 = 입의 폭; MW = MW0/1024
 AU = 각도 단위 = 10E-5 rad

2.1.2 무표정한 얼굴의 정의

처음에 시작함에 있어서 얼굴은 무표정한 위치에 있어야 한다. 즉 FAPU의 파라미터가 0의 값을 가질 때의 얼굴이다. 모든 FAP들은 무표정한 얼굴에 정의된 위치로부터 편차를 가지도록 표현되어야 한다. 무표정한 얼굴은 다음과 같이 정의된다.

- 좌표계는 우수좌표계; 머리 축은 좌표 축과 평행하다.
- 시선 방향은 Z-축 방향
- 모든 얼굴의 근육은 긴장되지 않고 풀려있어야 한다.
- 눈가풀은 눈의 홍채(흰자위)에 인접해 있어야 한다.
- 눈의 동공은 IRISD0의 1/3 이다.
- 입술은 다물고 있고 입술 선은 수평으로 되어 있으며 입술 끝의 높이는 같아야 한다.
- 입은 닫혀있고 윗니와 아랫니는 붙어있어야 한다.
- 혀는 바로 펴져 있고 혀의 끝부분은 윗니와 아랫니 사이의 경계에 닿아 있어야 한다.

2.1.3 FAP의 설명

복호기에서 FAP는 다음 3가지 중 하나이어야 한다.

1. 부호화기에 의해서 전송되어진 값들로 설정된다.
2. 부호화기에서 이전에 설정된 값을 유지한다.
3. 복호화기에 의해서 보간 된다.

FAP값은 초기값인 IFAP로 처음에 설정되고 IFAP집합은 만약 복호기에서 보간 되도록 지적된 특정한 Mask mode를 사용하지 않는다면 부분적인 Mask를 사용하여 유지되는 것으로 가정된다. 초기화되지 않은 FAP값들은 복호기에서 예측되어야 한다. 예를 들자면 만약 FAP Group 2(안쪽 입술)만이 사용된다면 FAP Group 8(바깥쪽 입술)은 사용되어서는 안된다. 바깥쪽 입술 점들은 부호화기에서 예측되어야 한다. 두 번째 예로서 FAP 복호화기는 대칭적인 값을 가지는 FAP 집합이 수신되면 오른쪽과 왼쪽이 대칭적이 되도록 조정해야 한다. (예를 들자면 왼쪽 눈이 움직이면 오른쪽 눈은 보간 되어 왼쪽 눈과 비슷한 요령으로 움직이게 해야 한다.)

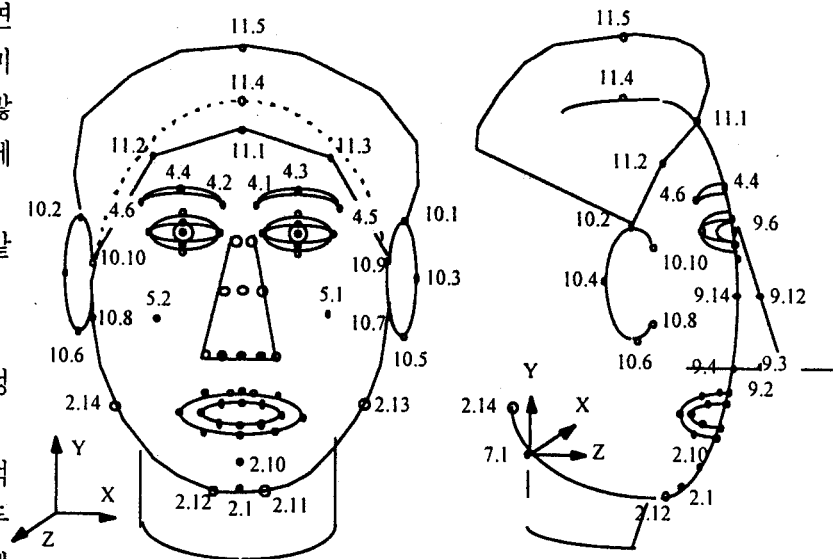
[표1] FAP Group

이름	FAP
visemes and expressions	2 FAPs
jaw, chin, inner lowerlip, cornerlips, midlip	16 FAPs
eyeballs, pupils, eyelids	4 FAPs
eyebrow	6 FAPs
checks	6 FAPs
tongue	10 FAPs
head toration	6 FAPs
outer rotation	3 FAPs
outer lip positions	5 FAPs
nose	5 FAPs
ears	7 FAPs

FAP 1과 2는 각각 연결고리와 표현방법을 의미하며 다른 FAP들보다 많은 의미를 가지므로 자세하게 규정하여야 한다. FAP의 그룹은 [표1]과 같다.

2.1.4 기본적인 표정의 정의

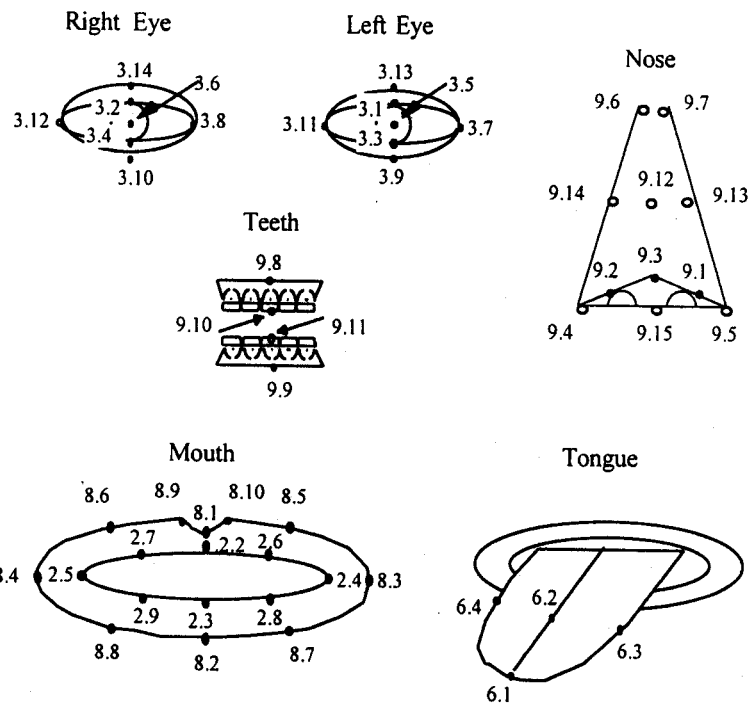
얼굴의 표정은 기본적으로 기쁨, 슬픔, 성냄, 두려움, 거슬림, 놀람 등 7개로 정의 된다.



2.2 Facial Definition Parameter Set

FDP들은 특정한 얼굴에 대해서 복호기의 얼굴 모델을 고정시키거나 어떻게 그것에 생명력을 불어 넣을 것인지에 관한 얼굴 모델을 전송 받기 위하여 사용된다. FDP들은 일반적으로 압축된 FAP들의 열로써 한 단위당 한 번씩 전송된다. 그렇지만 만약 복호기가 FDP들을 수신하지 못하더라도, 기본의 얼굴과 FAPU들의 사용으로 FAP열들을 해석 할 수 있다. 이것은 방송이나 원격 회의에서 최소한의 작동을 보장해 주게 된다.

FDP의 집합은 BIFS Syntax에 의하여 규정된다. FDP 노드는 수신기에서 사용될 얼굴 모델을 정의한다. 다음의 두 가지 동



- Feature points affected by FAPs
- Other feature points

[그림 2] FDP의 특징점

작이 제공되어야 한다.

- 보정 정보(Calibration Information)가 수신되는데, 이것은 수신기에서 적절한 얼굴이 얼굴의 특징점과 추가적으로 3차원 Mesh나 피부색을 사용하여 형성되도록 해

준다.

- 얼굴 모델은 FAP들 각각에 대한 Animation 값들을 수신한다. 이 얼굴 모델은 수신기에서 적절한 얼굴 모델로 바뀐다.

3. Body Object

점차 하드웨어와 네트워크의 성능이 소프트웨어와 더불어 증가함에 따라 좀더 복잡한 A/V물체가 있는 인터페이스를 만들 수 있게 되어 가상 환경 하에서 사용자는 좀더 실감나는 상호작용을 위해 원래의 몸짓과 표현을 이용하여 서로 영향을 끼칠 수 있게 되었다. 다른 사람과 같은 물체를 공유하는 이러한 가상 환경 하에서 가상적인 사람 몸통(Virtual Human Body)의 상호 작용은 중요한 기여를 할 것이다.

그러나 가상적인 사람 몸통의 실 시간적인 3차원 표현은 아주 어려운 일이다. 대부분 이러한 몸을 나타내기 위해 골격, 피부와 옷 등으로 구성된 층이 있는 접근법을 사용한다. 기본적인 골격 층위에 3D 피부 층이 덧붙여지고 이 피부 층위에 3D 옷감 층이 다시 덧붙여져 마치 실제의 사람과 같은 가상의 사람을 만드는 것이다. 이렇게 만든 가상의 신체는 몸통의 움직임과 얼굴의 표정, 소리와 동기가 맞아야 한다. 따라서 이 가상의 사람 몸통은 위에서 언급한 Facial animation과 상호 유기적으로 모델링 되어야 한다. SNHC에서는 Facial animation과 마찬가지로 몸통 모델링과 Animation에 대한 파라미터도 정의하고 있다. 그러나 Body animation은 Facial animation과 달리 3차원적인 디바이스(Device)를 처리하여야 하고, 말소리에 대한 동기는 덜 필요한 대신, 가상 공간에서의 물체들과의 동기가 필요하다. 따라서 표준안은 몸통의 전체적인 스케일링(Scaling)과 상체와 하체사이의 비

율, 자세한 움직임 표현할 수 있도록 다양한 자유도와 몸통사이의 좌표축 등의 변화 등을 고려하고 있다.

몸통은 가상의 사람 신체 모델을 만든 것으로 렌더링을 위해 3D polygon mesh의 형태를 하고 있다. 이러한 몸통을 정의하는데 BDP(Body Definition Parameter)와 BAP(Body Animation Parameter)의 2가지 파라미터가 정의되어 있다. BDP는 미리 정해진 몸을 신체의 표면, 신체의 차원(Body Dimensions), 텍스처에 따라 사용자가 원하는 몸으로 변형시킨다. BAP는 신체 모델을 다시 설정하거나 계산할 필요없이 다른 신체 모델에 대해서도 몸의 자세, 움직임 등을 만들어 준다.

여기서 몸통은 일반적인 가상 사람의 몸을 갖고 미리 정해진 자세로 렌더링이 되어 있다. 그리고 전송되는 비트열로부터 BAPs를 받아 Body를 움직이게 한다. 만일 BDP를 전송 받으면 BDP파라미터가 정하는 특별한 신체로 변형이 된다. 처음에 Body는 다음과 같은 기본적인 자세를 취하고 있다. 발은 앞쪽을 향하고 있고, 두 팔은 손바닥이 안쪽을 향하게 몸의 옆쪽에 붙어 있다. 관절들의 움직임 범위는 제한이 없는데, 사람 몸의 실제적인 움직임부터 간단한 사람 형태의 모델을 사용하는 네트워크 게임까지 다양한 응용에 적합하도록 디자인되어야 한다.

3.1 Body Animation Parameter Set

BAPs는 다음과 같이 다양하게 분류할 수 있다.

- Global Positioning Domain Parameters : 이 파라미터는 신체에서 관찰할 수 있는 곳들의 전체적인 좌표 값들과 방위를 가리키는 값으로 쇄골, 어깨, 팔꿈치,

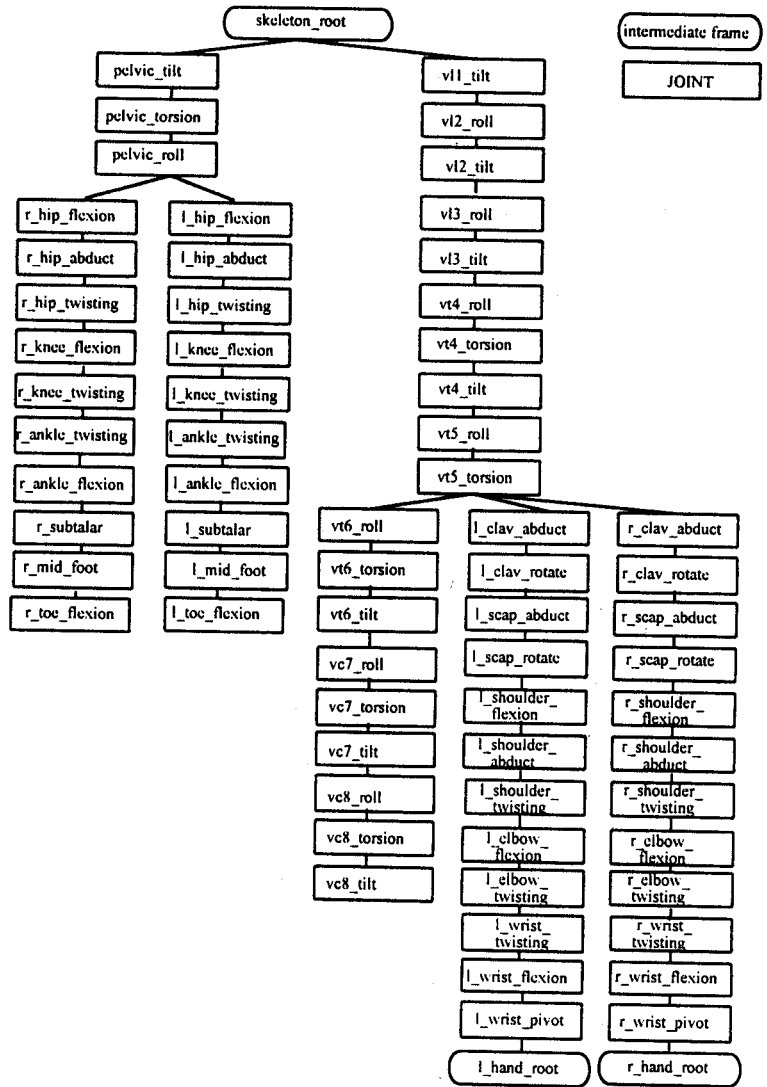
골반, 힙 등에서의 위치와 방향을 가진다.

- Joint Angle Domain Parameter : 이 파라미터는 서로 다른 신체의 부분을 연결하는 관절 각들을 포함한다. 발가락, 발목, 엉덩이, 척추, 어깨, 쇄골, 팔꿈치, 손목 그리고 손가락 등이다. 회전각들은 반 시계방향을 양의 방향으로 한다. 다음에 정의된 위치에 있을 때 회전각을 0으로 한다. 회전 법선 벡터는 몸에 따라 움직인다. 그리고 원(Parent)신체 부분을 고려하여 고정된다. 즉 회전축은 몸이나 World좌표계에 따라 있는 것이 아니라 신체의 부분에 따라 움직인다.

- Hand and Finger Parameter : 손은 복잡한 움직임을 수행하는데 각 손에는 25개의 자유도를 가지고 있다. 그리고 손은 5개의 손가락을 가진 일반적인 구조를 가지고 있다고 가정된다.

- High Level Parameter : 이것은 하위 레벨 파라미터로 그것들을 묘사 할 필요 없는 상위 레벨의 표현이나 움직임을 정의하기 위해 사용된다. 상위 레벨 파라미터들과 그것들의 값은 아직 정의되어 있지 않다.

Body Animation은 위에 정의된 4가지의 파라미터와 함께 세부적인 DOF(Degrees of Freedom)을 정의 하고 있다. 이 DOF는 크게 다음과 같이 4부분으로 나뉘어 정의된다.



[그림 3] 간략한 신체구조를 가진 신체구조

- Lower body
- At the foot complex
- Upper body
- Hands

BAP 그룹은 [표3, 4]와 같다.

3.2 Body Definition Parameter의 정의

BDP는 수신단에서 사용자가 정의하는 특정한 몸으로 특성화하거나 어떻게 신체를 움직이게 할까라는 정보에 따라 다운로드하여 사용된다. BDPs는 일반적으로 사용자

[표 2] Global BAP 그룹

이름	BAP
head	6 BAPs
neck	6 BAPs
left arm1	12 BAPs
right arm1	12 BAPs
left arm2	12 BAPs
right arm2	12 BAPs
pelvis	6 BAPs
left leg1	12 BAPs
right leg1	12 BAPs
left leg2	12 BAPs
left leg2	12 BAPs

[표 3] Joint-based BAP 그룹

이름	BAP
pelvis	3 BAPs
left leg1	4 BAPs
right leg1	4 BAPs
left leg2	6 BAPs
right leg2	6 BAPs
simple lower spine	10 BAPs
simple neck	6 BAPs
head	3 BAPs
left arm1	5 BAPs
right arm1	5 BAPs
left arm2	7 BAPs
right arm2	7 BAPs
detailed spine	72 BAPs
global position	6 BAPs

가 어떤 응용 물을 이용하는 동안 한번만 전송이 된다. 그리고 압축된 BAPs의 비트 열이 뒤따른다. 그러나 만일 수신단이 BDPs를 전송 받지 않는다면 기본적인 모델을 사용하며 이것 역시 BAP를 해석 할 수 있다. 이것은 방송이나 화상회의에서 최소한의 작동을 보장해주게 된다.

현재 BDP 파라미터들은 아직 규정되어 있지 않다. 그러나 본질적으로 다음과 같은 것을 처리 할 수 있어야 한다.

가. 신체표면구조(Body Surface Geometry) 만일 텍스춰 사용이 된다면 텍스춰 축을 가지고 있어야 한다. 신체 표면 구조는 3D Mesh 압축 방법 등을 이용하여 전송되어 진다. 기하학적인 표면은 VRML포맷이 될 것이다.

나. 3차원 참조점(Reference Points) 신체의 요구된 차원은 신체 좌표 시스템에서 신체 경계표(Landmark)를 이용하여 정의될 수 있다.

다. 텍스춰 영상(Texture Images) 이것은 선택사양으로 각 표면에 대해 정의되어 진다.

라. 기하학의 부가 정보 (Attachment Information of the Geometry) 신체 표면의 부가 정보는 요구된 변형을 가진 신체 표면을 위치시키는데 사용되며 부가될 신체의 DOF의 관점에서 신체 표면의 국부위치와 방위도 초기화된다.

마. 나이, 성별 등은 선택사양이다.

다음 것들은 BDP에서 기본적으로 가정되어야 할 것들이다.

가. 사람 신체 모델을 초기화하기 위한 기본적인 상태

서있는 자세 : 이 자세는 발은 앞쪽 방향에 있고 두 팔은 손바닥이 안쪽을 보며 몸의 옆쪽에 놓여진 자세이다. 이 자세에서 모든 관절 각은 0의 값을 갖는다.

나. 좌표축의 설정 신체 좌표 시스템의 원점은 관절로 이어진 내부 점으로 Spine 원점이 된다.

좌표축의 방위는 x 좌표는 Humanoid의 왼쪽이고 y는 위쪽, z는 앞쪽이다.

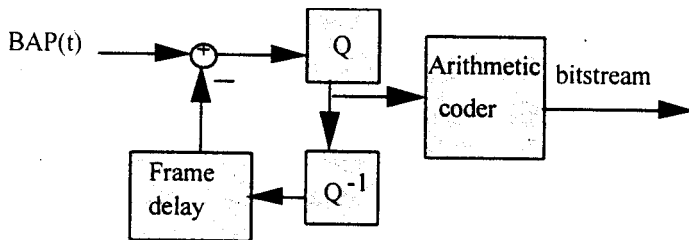
다. 신체 차원을 규정하기 위한 구경측정(Calibration)과 초기 파라미터의 설정

이것은 사용되는 가상 사람 신체 표면의 형태와는 독립적이다.

4. 비트열의 문법과 구문 의미(Bitstream Syntax and Semantic)

BAP와 FAP는 예측 부호화 방법(Predictive Coding Scheme)에 의해 양자화되고 부호화 된다. 현재 프레임에서 부호화되는 각각의 파라미터에 대해 이전 프레임에서 복호화된 값을 예측 값으로 사용한다. 그리고 예측 에러, 즉 현재 파라미터와 예측값 사이에 차이가 계산이 되고 산술부호화 된다. 이 예측 부호화는 누적되는 부호화 에러를 막는다.

BAP와 FAP는 각 움직임 파라미터들에 대해 서로 다른 정밀도를 요구한다. 따라서 다른 양자화 스텝크기가 적용된다. 비트 율은 BAP_QUANT와 FAP_QUANT라 불리는 양자화 스케일링 인자의 사용을 통해 양자화 스텝을 조절하여 제어된다. BAP_QUANT와 FAP_QUANT는 모든 BAPs에 균등하게 적용이 된다. 양자화 파라미터의 크기는 1부터 8사이이다.



[그림 4] BAP 예측 부호화

4.1 BAP Masking

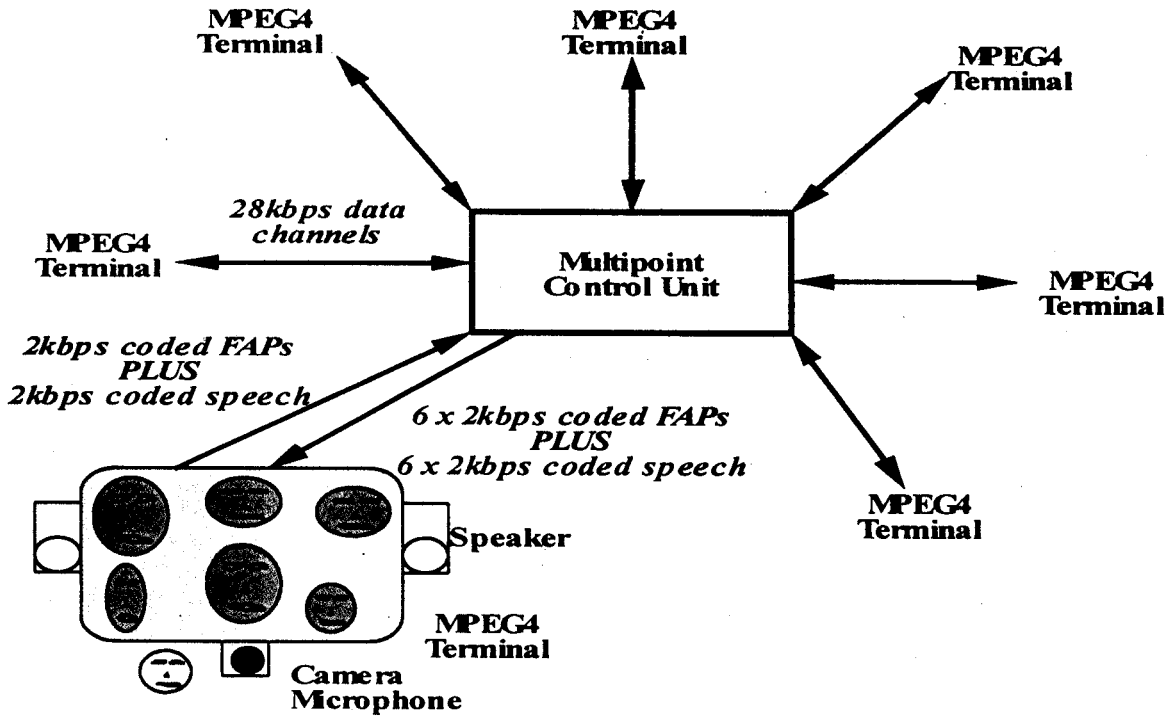
BAPs는 다양한 길이의 Mask에 의해서 차폐된다. Mask 형식은 다음과 같다. Mask에 앞서 선행되는 8비트는 어떤 BAP를 사용할 것인지 결정한다.

만일 첫 비트가 Set되면 다음에 나오는

Global positioning domain(표 2 참조)파라미터가 사용된다. 그렇지 않은 면 Joint-angle(표 3 참조)에 기반을 둔 파라미터가 사용된다. 두 번째 비트가 Set되면 복잡한 손이 사용이 된다. 세 번째 비트가 Set 되면 상위 레벨 BAPs가 사용이 된다. 현재 세 번째 비트는 0이다. 네 번째 비트는 각 그룹들이 Global과 Joint BAPs에 사용되는지 결정한다. 만일 Set되면 적용이 된다. 첫 번째와 네 번째 비트가 Set되면 Mask는 11개의 추가적인 비트를 가진다. 각각의 비트는 일치하는 Global Positioning BAP Group의 전송을 나타낸다. 만일 처음 비트가 Set되고 네 번째 비트가 0이면, Mask는 38비트를 가진다. 여섯 번째 비트는 간단한 척추가 사용되는지 복잡한 척추가 사용되는지 나타낸다. 두 번째 비트와 다섯 번째 비트가 Set 되면 Mask는 손 BAPs의 그룹에 해당하는 전송을 나타내는 4비트를 추가적으로 가진다. 만일 일곱 번째 비트가 Set되고 다섯 번째 비트가 0이면, Mask는 추가적으로 50비트를 가진다. 현재 일곱 번째 비트는 0이다.

5. Face and Body Animation의 응용 분야

앞에서도 살펴보았듯이 MPEG-4 SNHC의 Face and Body Animation에서는 3차원적인 얼굴과 몸의 움직임을 표현하기 위해 다양한 변수를 정의하였다. 따라서 부호화기에서는 소형 카메라와 같은 감지기를 이용해 각 변수들의 값을 추출한 후, 이를 1Kbps이하의 낮은 비트율로 압축 부호화하여 전송한다. 복호화기에서는 이러한 변수의 값을 이용하여 미리 가지고 있던 인공의 얼굴과 몸의 모델에 적용하여 원래의 움직임을 재현한다.



[그림 5] Virtual meeting 응용

이것의 예로 그림 5와 같이 Virtual meeting을 생각할 수 있다. 현재 3D 가상 공간에 7명이 Virtual meeting을 하고 있다. 그림에서 보이는 것처럼 각 사람은 MCU (Multipoint Control Unit)로 연결이 되어 이 중 한 사람은 자신의 MCU를 통하여 부호화된 오디오와 FAP비트열을 모아서 전송하고 나머지 6명의 데이터를 받게 된다. 이 FAP는 각 개인의 터미널의 오디오와 비디오를 통하여 만들어진다는.

6. 결론

사용자들은 여러 사람들이 함께 참가하여 서로 상호 작용을 할 수 있는 보다 실제적인 가상 공간을 기대한다. 즉 기존 방식의 수동성에서 서로 대화하는 멀티미디어로의 발전을 요구하는 것이다. 이런 기대에 부응

하고자 MPEG-4 SNHC는 A/V 물체의 효과적인 통신을 위해 오디오, 비디오와 2D/3D 그래픽을 합성하는 부호화 및 가상 현실에서의 아바타의 구현 등에 대한 표준을 개발하고 있다. 따라서 우리나라에서도 국제적인 경쟁력을 확보하기 위해 Face and Body Animation에 대한 기술 분석과 응용 개발에 적극적으로 참여할 필요가 있다.

7. 참고문헌

- [1] ISO/IEC JTC1/SC29/WG11 N1730 : MPEG-4 Overview
- [2] ISO/IEC JTC1/SC29/WG11 N1820 : SNHC Verification Model 5.0
- [3] ISO/IEC JTC1/SC29/WG11 N1891 : SNHC Verification Model 6.0 **[KIS]**