

신경회로망을 이용한 저가의 실시간 립싱크 애니메이션 시스템의 구현

*강이철^U **김철기 *김미숙 **차의영
*부산대학교 일반대학원 멀티미디어학과 *부산대학교 일반대학원 전자계산학과
(errorman, kck, sook99, eycha}@harmony.cs.pusan.ac.kr

Embodiment of Low-Cost Real Time Lip-Sync Animation System Using Neural Network

*Yi-Chul Kang^U **Cheol-Ki Kim *Mee-Sook Kim **Eui-Young Cha
*Inter Disciplinary Research Program of Multimedia, Pusan National University
**Dept. of Computer Science, Pusan National University

요 약

최근 인터넷 기술의 발달로 인한 실시간 인터넷 동영상 서비스 등 인터넷을 이용한 방송사업이 활성화 되어지고 이에 부가적으로 애니메이션이 감각적 서비스를 지원하고 있는 실정이나 고가의 모션캡처 시스템을 통한 캡처된 좌표를 적절한 보정을 한 후에 캐릭터를 움직이는 방법으로 이루어지고 있다. 이러한 모션캡처 시스템을 통한다면 시스템 및 장비 자체가 고가이고 또한 실시간으로 처리하기 위하여는 좌표값 보정 등이 필요하지만 본 논문에서 제안하는 좌표 추출 및 추적 기법을 이용하여 저가의 가정용 멀티미디어 오버레이 캡처보드와 CCD카메라를 통하여 영상을 캡처하고 캡처된 영상좌표와 실험용 GDI object를 링크시켜 실시간으로 사람의 입술의 움직임 모양대로 애니메이션이 립싱크되어서 움직여지는 것을 볼 수 있으며, 더 나아가서 외화의 우리말 더빙시 영상처리를 통한 정교한 화면 더빙 및 가상 캐릭터를 이용한 사이버 미팅까지 가능할 것이다.

1. 서론

최근 인터넷 기술의 발달로 인한 실시간 인터넷 동영상 서비스 등 인터넷을 이용한 방송사업이 활성화 되어지고 이에 부가적으로 애니메이션이 감각적 서비스를 지원하고 있는 실정이나 고가의 모션캡처 시스템을 통하여 캡처된 좌표를 적절한 보정을 한 후에 캐릭터를 움직이는 방법으로 이루어지고 있다. 이러한 모션캡처 시스템은 하이테크 멀티미디어 콘텐츠 개발에 있어서 가장 진보된 기술이며 가장 고가의 장비에 해당한다 [1]. 현재 시판되어지고 있는 모션 캡처시스템은 기계, 자기, 그리고 광학 방식이 있으며 기계 및 자기방식은 세밀한 움직임 추출이 불가능하고 시스템 조작상의 어려움이 있고, 광학방식은 세밀하고 유연한 움직임이 가능하며 조작이 간단한 장점이 있다. 최근에는 이러한 광학방식으로 교체하거나 새롭게 채택하고 있는 실정이다. 그러나 기존의 모션 캡처시스템으로 단지 애니메이션의 립싱크부분을 실시간으로 구현하기 위한 비용은 투자대 효과면에서 너무도 비현실적이다.

본 논문은 이러한 외국의 모션캡처 시스템의 일부인 립싱크 부분에서 국내에서 생산되어 손쉽게 구할 수 있

는 저가의 멀티미디어용 오버레이 캡처보드와 PC용 CCD 카메라(interlace 방식)를 이용하여 실시간으로 애니메이션의 립싱크를 구현할 수 있는 기법을 제안하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 실시간 특징점 추출 및 추적에 대한 관련 연구에 대해 살펴보고, 3장에서는 신경회로망을 이용하여 실시간 애니메이션 립싱크를 구현하기 위한 알고리즘을 제시한다. 4장에서는 실험 및 결과를 분석하고 5장에서 결론 및 향후과제를 설명한다.

2. 관련 연구

기존의 컴퓨터 애니메이션 기법은 크게 동작제어 애니메이션 기법, 키프레임 애니메이션 기법, 그리고 그리고 모션 캡처 시스템으로 구분할 수 있다[2]. 이러한 기존의 애니메이션 기법은 모션 캡처 시스템을 제외하고는 인간의 수작업이 많이 필요로 하며 많은 시간과 자본이 요구되어진다. 그러나 모션캡처 시스템은 이러한 문제에 대한 해결책으로 실제 인간이나 움직임 대상체의 움직임을 컴퓨터에서 이용가능한 데이터 형식으로 변환하여 주는 역할을 한다. 특히 애니메이션의 립싱크 구현 뿐만 아니라 영화의 특수효과나 게임, 3차원 애니

메이션 등 고부가치 산업에서 많이 쓰여지고 있다. 그러나 모션 캡처 시스템 자체가 너무 고가인 단점이 있다. 모션 캡처 시스템을 통한 컴퓨터 애니메이션은 실시간으로 움직임을 포착하고 동작 시킬 수 있는 기술로써 시간적인 정확성과 논리적인 정확성을 동시에 만족하기 위해 실시간으로 립싱크 애니메이션을 구현하기는 고가의 하드웨어가 필수적으로 요구되어진다.

기존의 모션 캡처 시스템은 전신 혹은 특정부분의 움직임을 포착하기 위해 다수의 자기 센서를 부착하고 유, 무선으로 컴퓨터에 연결하여 움직임을 캡처하는 자기(Magnetic) 방식과 적외선 반사체를 움직임 관제에 부착하고 다수의 적외선 카메라를 이용하여 적외선에 반사되어지는 적외선 반사체를 여러 장의 2차원 이미지로 획득하여 3차원 공간상의 좌표로 움직임을 파악하는 광학(Optical)방식으로 크게 나눌 수 있다[1]. 이중 광학방식은 상당히 정밀한 위치 측정이 가능하여 얼굴표정 등을 캡처하는 것이 가능하다.

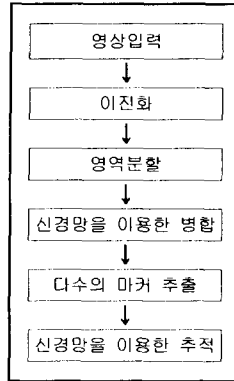
얼굴의 애니메이션 작업은 크게 표정의 애니메이션과 입술 모양의 변화(Lip Synchronization)를 중심으로 하는 대화 애니메이션 부분으로 구분할 수 있는데[1], 본 논문에서 주안점을 두는 대화 애니메이션은 가장 움직임의 변화가 심한 부분으로 부착되어진 마커들을 정확하고 빠르게 추출 및 추적해야 한다. 마커와 같은 특징점 추출을 위해 도입하는 기술로 영상의 영역화 기법이 대부분의 컴퓨터 시각시스템에서 사용되어지고 있으며 영역화를 통한 성과가 시스템 성능을 좌우하고 있다[7].

영역화 기법은 영역을 대표하는 특징 또는 특성에 기반하여 비슷한 화소들을 하나의 단위로 생각하여 동일한 성질을 가지는 영역들을 추출해 내는 영역기반 영역화 기법과 영상내부에서 경계선을 추출한 후 얻어진 경계선 기반의 영역화 기법으로 나눌수 있다[3][5]. 얼굴영상 획득후 마커의 경계 부분에 대한 기술이 중요하지 않으므로 영역에 기반한 영역화 기법이 보다 나은 효과를 낼 수 있다. 이러한 영역기반의 알고리즘은 개개의 화소로부터 출발하는 영역 확장(Region growing)기법은 단점이 있고, 전체 영상으로부터 출발하는 영역 분할 및 병합(split-and-merge)기법은 정확도면에서는 떨어지지만 계산량이 적다는 장점이 있다[3][5][6].

실시간으로 얼굴영상에서의 특징점인 마커를 추출하고 추적하기에는 노이즈에 강건하고 계산량이 적은 영역분할 및 병합기법이 적당한 방법이며 ITS분야 등에서 가장 많이 응용되어지고 있다[7].

3. 실시간 애니메이션 립싱크 구현을 위한 알고리즘

실시간으로 마커의 위치를 추출하고 추적하기 위하여 사용되어지는 알고리즘의 전체 흐름도는 [그림 1]과 같다. 적외선 필터를 사용하지 않는 CCD카메라로부터 입력되어지는 아날로그 신호를 멀티미디어용 오버레이 캡처보드가 디지털 영상신호로 변환하여 얻어진 256 그레이 레벨(Gray Level)의 영상을 획득한 후 [식 1]과 같



[그림 1] 알고리즘 개요

이 Thresholding을 통한 이진영상을 구한다[5].

$$B(m, n) \approx \begin{cases} 1, & f(m, n) \leq T \\ 0, & f(m, n) > T \end{cases}$$

[식 1]

그리고, 입력영상을 균등한 크기의 블록으로 나누고 각 블록 내의 이진화 되어진 픽셀들의 합을 [식 2]와 같이 구하여 임계값인 t값을 조절함으로써 영역분할을 통하여 마커 외의 노이즈를 제거할 수 있다.

$$Block(i, j) = \begin{cases} 1 : \text{영역내의 픽셀의 수} > t \\ 0 : \text{영역내의 픽셀의 수} \leq t \end{cases}$$

[식 2]

단계 1: 실시간 입력 동영상에 대하여 식 1을 이용하여 이진영상을 구한다.

단계 2: 식 2를 이용하여 이진영상에 대한 각 블록의 블록값을 설정한다

단계 3: 신경회로망을 이용하여 1차 병합을 수행한다.

- (1) 새로운 블록 패턴 X가 주어지면, 최소거리의 승자 뉴런 j^* 를 선택하여 표시한다.
- (2) 경계 임계값 검사: 승자 뉴런과 입력 블록패턴의 거리가 경계 임계값 ρ 보다 적으면 같은 마커 영역으로 인지하고 통과

$$\|X - W_{j^*}\| < \rho$$

- (3) 경계 임계값 검사에서 실패하면(마커영역 밖에 있는 경우), 새로운 마커 뉴런을 만든다.

$$W_k = X$$

- (4) 경계 임계값 검사에서 통과되면, 승자 뉴런의 가중치를 조정한다.

$$W_{j^*}^{(new)} = \frac{X + W_{j^*}^{(old)} \|cluster_{j^*}^{(old)}\|}{\|cluster_{j^*}^{(old)}\| + 1}$$

단계 4: 병합 결과 이웃에 있는 마커들에 대한 상관관계를 이용한 병합

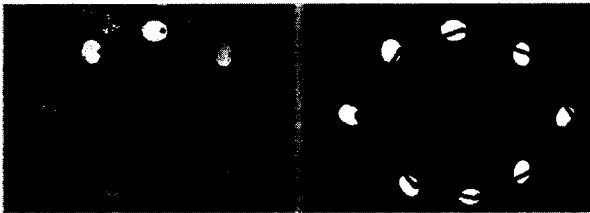
단계 5: 다중 마커의 추출(영역의 크기와 영역의 위치, 중심점 등)을 통한 마커의 추적.

신경회로망을 이용한 병합은 위의 ART2 알고리즘[4]을 응용하여 병합에 적용하고 마커의 크기가 일정함과 실험자 입의 최대 움직임 위치를 추정할 수 있으므로 일련의 사전 지식을 바탕으로 병합을 수행하여 다수의 마커를 추출한다. 이때 각 마커의 경계 임계값 설정시 마커간의 거리를 고려하여 설정한다. 또한 다수의 마커의 추적을 위하여 본 논문에서는 병합시 사용한 신경회로망 기법 중 자율학습이 가능한 ART2 알고리즘을 적용하여 추출시 구하여진 중심점과 좌표값이 변화할 때 마다

실험용 GDI Object가 이전 영상 위에 오버레이 되어 움직일 수 있도록 링크시킨다.

4. 실험 결과 및 분석

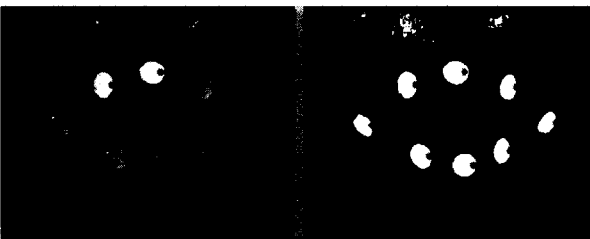
실험자는 저가의 멀티미디어용 오버레이 캡처보드를 펜티엄2-350MHz 프로세서가 장착된 컴퓨터에 설치하여 interlace 방식의 저가의 PC용 CCD카메라를 이용하였다. 이미지는 320 × 240 크기의 256 Gray 영상을 사용하였으며 초당 13프레임의 처리속도로 실험하였고 실험자의 입주위에 총 8개의 마커를 부착하여 실험에 임하였다.



[그림 2] 원영상(좌)과 이전영상위에 Link된 Overlay GDI object(우) <입을 벌리고 있을때(시간 t)의 실험 결과 영상>



[그림 3] 원영상(좌)과 이전영상위에 Link된 Overlay GDI object(우) <입을 다물고 있을 경우(시간 t-2)의 실험 결과 영상>



[그림 4] 일반적인 이진화 기법만을 사용했을시 Error Tracking 영상

실험 기법 결과	성공율	비고
다중 마커 추출 결과	98%	대부분 손실 없음
다중 마커 추적 결과	93%	

[표 1] 실험한 추출 및 추적기법 적용결과

구현 환경은 MS Windows98상에서 Visual C++ 6.0을 이용하여 구현하였으며 카메라 입력영상에서 1프레임을 캡처하여 이미지를 획득하는데 80ms가 소요되었다.

전술한 바와 같이 실시간으로 총 8개 마커의 추출 및 추적이 90% 이상이 가능하고 이를 통한 애니메이션의 립싱크가 가능함을 증명할 수 있었다.

5. 결론 및 향후과제

일반적으로 컴퓨터에서 영상을 Digitize하고 처리하는 멀티미디어용 오버레이 캡처보드와 모션캡처용 Frame Grabber는 겉보기에는 비슷한 기능을 가지고 있는 듯하지만 실제로 멀티미디어용 캡처보드는 영상을 캡처하고 데이터를 편집하기 위하여 낮은 해상도와 압축기술을 기반으로 설계된 반면 모션캡처용 Frame Grabber는 많은 양의 Data Stream을 효과적으로 처리할 수 있도록 영상의 질을 최우선으로 고려하여 설계되어 그 가격의 차이는 10배 이상 차이가 나는 결코 비교의 대상이 되지 않는다. 그러나 이러한 저가의 멀티미디어용 캡처보드를 사용하여 실시간으로 캡처한 영상을 본 연구자가 제안한 기법을 적용, 처리하여 애니메이션의 립싱크가 가능함을 우리는 실험을 통하여 확인 할 수 있었다. 실험 기법으로 구현한 애니메이션 립싱크시 입의 움직임이 무척 빠른 경우는 마커의 좌표 추적율이 저하되는 현상이 있었으나 이는 멀티미디어용 캡처보드의 성능상의 문제로 보아진다. 또한 본 실험은 마커의 수를 8개로 제한하고 고정된 환경 속에서 이뤄진 실험으로 앞으로 스테레오 비전을 이용하여 마커의 깊이 정보까지 끌어내어 보다 입체적인 애니메이션 립싱크가 가능하도록 하여야 겠다. 이를 응용하고 발전시킨다면 더 나아가시 외화의 우리말 더빙시 영상처리를 통해 정교히 더빙된 외화감상 및 가상 캐릭터를 이용한 실시간 사이버 미팅까지 가능할 것이다.

6. 참고 문헌

- [1] 김용순, 김영수, "3차원 캐릭터 애니메이션 기술 동향", 정보과학회지, 제17권, 제2호, p48-59, 02, 1999.
- [2] George Maestri, "Digital Character Animation", New Riders Publishing, 1997.
- [3] 서울대학교, "실시간 표적인식 및 추적기법연구", 자동제어 특화 연구센터 연구보고서, 1996.
- [4] S.Y. Kung, "Digital Neural Networks", Prentice Hall, pp. 78-85, 1993.
- [5] Rafael C. Gonzalez, Richard E. Woods, "Digital Image Processing", Addison Wesley, pp.413-481, 1993.
- [6] X. Wu, "Adaptive Split-and-Merge Segmentation Based on Piecewise Least-Square Approximation", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.15, No.8, pp.808-815, Aug. 1993.
- [7] 이철희, "신경회로망을 이용한 다중 표적의 추출과 추적", 부산대학교 이학석사학위논문, 02, 1999.