

## 신경회로망을 이용한 2D 얼굴근육 파라미터의 자동인식

김동수\* 남기환\* 한준희\* 배철수\* 권오홍\* 나상동\*\*  
\*관동대학교 전자통신공학과 \*\*조선대학교 컴퓨터공학과  
전화 : (0396) 670-3411 / 팩스 : (0396) 670-3499

### Automatic Estimation of 2D Facial Muscle Parameter Using Neural Network

Dong su Kim\* Kee hwan Nam\* Jun hee Han\* Oh Heung Kwon\* Cheol soo Bae\*  
Sang-dong Ra\*\*  
\*Kwandong University \*\*Chosun University  
E-mail : baecs@kdccs.kwandong.ac.kr

#### Abstract

Muscle based face image synthesis is one of the most realistic approach to realize life-like agent in computer. Facial muscle model is composed of facial tissue elements and muscles. In this model, forces are calculated effecting facial tissue element by contraction of each muscle strength, so the combination of each muscle parameter decide a specific facial expression. Now each muscle parameter is decided on trial and error procedure comparing the sample photograph and generated image using our Muscle-Editor to generate a specific face image.

In this paper, we propose the strategy of automatic estimation of facial muscle parameters from 2D marker movement using neural network. This also 3D motion estimation from 2D point or flow information in captered image under restriction of physics based face model.

#### I. 서론

사람의 얼굴 인식은 얼굴 영상 입력을 통한 인증과

같은 보안이나 많은 얼굴 중에서 비슷한 사람의 검색 등에 활용되고 있으며, 또한 향후 로봇 공학에서 로봇이 임의 사람을 알아보는 부분에서 사용될 수 있다. 이렇게 인간과 유사한 인터페이스 형태는 최근 연구가 활발히 진행되고 있으며, 인간과 인간사이의 의사전달에서 말이 아닌 얼굴 표정으로 인간의 감성을 전달하는 이러한 필요성으로 얼굴인식 기술은 그 가치가 더해가고 있어 많은 연구가 수행되고 있다[1][2]. 본 연구에서는 얼굴인식 분야에서 얼굴근육움직임에 대해 다루고자 한다. 얼굴 근육 모델은 얼굴 조직 성분과 근육열(string)로 구성되어 있다. 이 모델에서 얼굴 조직 성분에 영향을 주는 힘은 각 근육의 수축강도로써 계산되어지고, 각 근육파라미터의 결합은 특정한 얼굴 이미지를 나타낸다. 기존의 연구에서는 수작업으로 포토사진과 합성 이미지를 비교하는 실험값과 예리값으로써 각 근육 파라미터를 결정하였으나, 본 연구에서는 얼굴 안면 이미지에서 지시점 움직임으로부터 얼굴 근육 파라미터의 자동인식 방법을 제안하고자 한다. 또한 연구의 목표는 합성하지 않은 본래의 느낌을 가지는 얼굴 표정과 확실한 근육 파라미터를 인식하는 것이다.

본 논문에서는 2차원 정보 즉, 정면 이미지 상에서의 지시점의 움직임과 광류(optica flow)등으로부터 열

갈 근육 파라미터를 자동으로 인식하는 2가지 방법을 제안한다. 첫 번째는 작은 원형 지시점은 표정을 지었을 때 얼굴의 변형량을 측정하기 위하여 피사체 얼굴에 붙여지고, 어떠한 명확한 표정과 중립적인 표정사이의 차이를 찾아낼 수 있다.

얼굴표정들의 여러 형태들로 학습된 신경회로망은 근육 파라미터상의 지시점 움직임을 변환할 수 있고, 3D 근육 파라미터로부터 2D 포인터 움직임을 이미지 합성처리의 역 매핑을 실행할 수 있다. 그래서 얼굴모델을 기본으로 한 캡처된 이미지 상의 2D 포인터 행적들로부터 3D 움직임을 인식하게 한다.

## II. 근육파라미터

### 2.1 특징점 추출

지시점은 측정하기 위한 피사체 얼굴의 각 특징점과 얼굴표정모델에 붙여졌다. 특징점은 근육이 수축할 때 얼굴 모델의 격자점에서 큰 움직임을 주는 곳을 각 근육에 대해 선택하였으며, 만일 특징점이 이미 다른 근육으로 선택되어 졌다면 그 포인터는 두 번째로 큰 움직임을 특징점으로 선택되어진다. 몇몇의 포인터는 각 근육을 제어할 때 다른 포인터들에 대해서 독립적인 동작을 발견되어 덧붙여졌다. 그 결과 상안(上顔)영역에 16개의 특징점과 입 영역에 26개의 특징점으로 선택되어졌고, 그림 1과 같다.

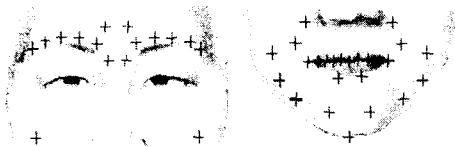


그림 1. 특징점  
Fig 1. Feature Points

### 2-2 학습 모형

학습 모형은 각 근육 개개의 수축과 근육들의 결합으로 구성되어 있다. 이 개개의 움직임 경우 가장 큰 힘과 중간정도 사이의 근육수축을 11단계로 양자화 하였으며, 이 결합의 경우 노여움, 혐오감, 공포, 행복감, 슬픔과 놀람으로 구성된 6가지 기본 얼굴표정을 만들고, 중간값과 각 표정들의 차이를 11단계로 양자화하였다.

그 학습 모형의 총 개수는 상안부 위에 143개 이고, 입 부근에서의 각 근육수축은 독립적으로 발생되지 않

아서 학습 패턴들의 조합들로 구성된다. 즉, “아”, “이”, “우”, “에”, “오”의 입술모양과 입을 다문모양이 모음에 관한 기본 입모양이다. 또한, 6개의 기본 표정은 상안 부근에 추가되고, 턱 회전을 특별히 도입하였다. 또한 11단계로 양자화 되어졌으며, 그 학습패턴의 수는 입 아래 부근에 대해 143개이다.

각 패턴은 근육패턴 벡터와 특징점의 움직임 벡터의 쌍으로 구성된다. 신경회로망은 역전파로 교육되어졌다.

## III. 평가

### 3-1 표정합성 얼굴모델의 정합

얼굴의 일반적인 형상모델로부터 대상인물의 형상모델을 얻는다. 얼굴 정합은 얼굴 전체의 크기 및 위치 조정, 얼굴 윤곽의 미세조정으로 이루어진다.

#### (1) 얼굴의 크기 및 위치 조정

얼굴의 크기 및 위치를 대략적으로 조정하기 위해 3차원 어파인 변환(Affine Transform)을 행한다. 3차원 어파인 변환에서는 얼굴의 크기 및 위치를 잘 반영하고 그림 (a)의 WFM(Wire Frame Model)의 ①, ②, ③, ④, ⑤의 점을 이용한다. 어파인 변환에서는 눈 안쪽의 특징점 2개의 중점을 이용하고 있기 때문 얼굴 영상의 4개 특징점의 깊이 좌표는 얼굴영상으로부터는 얻을 수 없기 때문에, 얼굴의 3차원 형상 모델의 깊이 좌표를 근간으로 해서 추정한다.

#### (2) 얼굴 윤곽 및 머리의 정합

어파인 변환은 선형변환이기 때문에, 얼굴의 크기 및 위치가 선형적으로 조정되어 대략적으로 정합된다. 그러나 개인에 따라 얼굴 윤곽과 머리의 윤곽이 일치하지 않기때문에 그림 (a)의 ⑧, ⑨의 점을 이용하여, 두발의 높이를 비례적으로 조절한다. 또한 ⑥, ⑦의 점을 이용하여 두발의 넓이를 비례적으로 조절한다.



(a) 정합 전 (b) 정합 후  
(a) Before Adjustment (b) After Adjustment  
그림 2. 모델정합의 예

Fig 2. A examples of model's adjustment

이와 같은 과정을 통해 모델영상과 대상을 정합시키게 되고, 정확한 지시점의 위치를 찾아 각 지시점의 움직임을 추정하게 된다.

3-2 공식적인 평가

실제의 인간 얼굴에 지시점들을 붙이고 임의의 표정을 지었을 때, 카메라에 의해 획득된 2D 영상으로부터 지시점의 움직임을 얻는다.

후에 표준화시키고, 그 지시점의 움직임값은 신경회로망에 주어지고, 얼굴 이미지는 신경회로망 출력 파라미터로 얼굴근육모델을 사용하여 다시 발생시킨다. 그림 3은 예를 든 이미지를 보여준다. 근육 파라미터는 오직 2D 이미지로부터 결정되지만 3D 얼굴 이미지는 보다 나은 재생을 할 수 있다.

또한 예외적인 상황은 연구 모델의 어떤 근육 조합으로도 발생되지 않은 실험 샘플로 얼굴모양이 주어질 때 발생한다. 우리의 흐름근육모델은 아래 입술은 기준 위치에서 규정된 범위를 넘어서 움직일 수 없다. 왜냐하면 그 근육 움직임은 오직 수축하는 것으로 한정되어졌다. 이와 같은 문제를 해결하기 위하여 근육의 재배치와 확장영역을 보다 많이 포함하여 근육 모델을 개선하는 것이 필요하다.

3-3 객관적 평가

표 1은 순간 획득한 사진의 얼굴과 합성한 얼굴 사이의 오류를 보여준다. 첫 번째 종렬은 실험에서 적용한 42개 지시점 사용시의 평균적인 오류를 의미한다. 또한, 두 번째 종렬은 지시점 위치의 오류이지만, 이것은 인간의 민감도를 고려한 값이다. 먼저 일곱명의 사람은 실험과 오류처리 과정에서 근육 편집기를 사용해서 원본의 얼굴과 매우 유사한 인상으로 얼굴 표정을 만들었다.

표 1. Open Test에 대한 인식 오류  
Table 1. Estimated Error for Open Test

	지시점 위치오류		근육파라미터오류	
	pixels	$\sigma$	pixels	$\sigma$
노여움	9.79	5.01	3.41	2.37
혐오감	5.25	1.19	5.33	1.73
무서움	6.46	1.68	5.42	1.38
행복함	3.91	0.91	5.29	1.44
슬픔	6.91	1.75	4.51	1.47
놀람	6.48	1.48	6.38	1.59

그리고 이러한 얼굴들은 근육 파라미터로 평균화하고, 정확한 얼굴로 정의한다. 표준편차는 각각의 표정에 대해 계산되어지고, 인간표정 변화의 민감도를 의미한다. 두 번째 종렬은 표시점 위치의 오차이다. 예를 들어, 노여움의 경우, 오류는  $5.01\sigma$ 이고  $\sigma$ 는 3.14 pixel 이다. 세 번째 종렬은 근육 강도의 오류이다. 이 결과들은 합성 얼굴 표현에 대한 평가의 양적인 기준을 보여준다. 노여움의 경우가 가장 나쁜 것 중 하나이다.



a)원영상 b)평균영상 c)합성영상

그림 3. 놀람에 대한 공식적인 평가

Fig 3. Open Test for Surprise

VI. 지시점 생략

다음단계로, 움직임의 획득처리를 쉽게 하고 지시점의 처음 위치에 대한 강한 의존성을 해소 하기 위하여 지시점들은 생략되어야 한다. 본 논문에서 새로운 방법이 소개되어 진다. 이 방법은 얼굴 표정을 관찰하기 위해 광류(光流)를 이용한다. 흐름은 각 광채(flame)로 계산되고, 마지막으로 흐름은 각각의 흐름들이 합산되어진 것이며, 얼굴 Mask의 평균치가 된다. 그림 4(a)는 광류 계산 결과이고, 그림 4(b)는 평균의 Mask들을 보여준다.

얼굴표정은 신경회로망 출력 파라미터와 얼굴 근육 모델을 사용하여 다시 합성된다. 이 신경회로망은 처음 얼굴 이미지의 광류와 처음 얼굴 이미지와 상응하는 근육 파라미터 쌍으로 학습된다. 결과적으로 우리는 원래 이미지와 흡사한 모습의 이미지를 얻을 수 있다. 그림 4의 c),d)는 합성 이미지의 예이다.



(a) 광류 (b) 평균Mask들 (c) 획득 영상 (d) 합성영상

그림 4. 광류 합성

Fig 4. Synthesis by Optical Flow

더 자세한 표정의 특징은 지시점 행적방법 보다 광류를 이용한 방법에 의해서 획득할 수 있다. 이 인식이 앞으로의 과제이다.

### V. 결론

본 논문은 2D 지시점의 움직임으로부터 3D 얼굴근육 파라미터를 자동적으로 인식하는 방법을 소개하였다. 모델이 2D의 대상얼굴과 정확하게 맞을 때 3D나 2D 파라미터로 전환할 수 있고, 표정 변화는 모음 발음이나 기본적인 표현의 결합에 한계가 있다. 일반적으로 수작업으로 모델과얼굴 이미지를 맞춘다. 그러나 실제 얼굴 영상에 모든 격자점의 위치는 일치할 수 없다. 물론 격자점의 경우 대상의 위치와 심하게 움직이고 있는 파라미터 전환은 민감한 반응을 하게 된다. 그러므로 실제 얼굴과 모델의 일치문제를 해결해야 한다.

추가로 입력되는 실제영상 모델이 안경 착용을 한 경우에 대해 고려한 평가에서 지시점을 이용한 방법은 움직임을 인식하는 포인터가 상안의 눈썹주위에 많이 선택되어 있기 때문에 안경태의 크기가 클 경우에는 눈썹부위를 가려지게 되어 눈썹주위의 포인터 움직임을 오인식하는 경우가 발생하게 되었고, 반대로 안경태가 작은 경우에는 인식에 영향을 주지 않았다. 또한 광류를 이용한 경우에도 영향을 주지 않았다. 하지만 이 결과에 대한 부분은 몇 개의 샘플로 아직 평가가 계속진행 중이기 때문에 완벽한 결과치는 아니다. 그림 5는 샘플이미지의 한 예이다.



그림 5. 샘플이미지  
Fig 5. Sample Image

현재 본 얼굴 근육모델은 오랜 계산시간이 필요하다. 그 결과 위의 방식은 실시간 처리가 불가능하게 되었고 모델의 표현한계를 볼 수 있었다. 그래서 새로운 근육 모델은 근육의 재배치와 물리적인 정의하에 실험하였다. 앞에서 지시점이 없는 광류 방법을 소개하였고 광류를 이용한 방법을 사용하면 더 섬세한 얼굴 표현을 획득할 수 있다. 이 인식은 다음 과제이다.

### 참고문헌

- [1] Yeencheng Lee, Demetri Terzopoulos and Keith Waters, "Constructing Physics-Based Facial Models of Individuals", *Graphics Interface '93*, pp.1-8(1993).
- [2] Demetri Terzopoulos and Keith Waters, "Analysis and Anatomical Models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.15, No.6, pp.569-579(1993).
- [3] Paul Ekman and Wallace V.Friesen, : *Facial Action Coding System*", Consulting Psychologists Press Inc.(1978).
- [4] F. Parke, "A Parameterized Model for Facial Animation", *IEEE Computer Graphics & Applications*, vol.2, No.9, pp.61-68, Nov.(1982).
- [5] J. P. Lewis and F. I. Parke, "Automated Lip-Synch and Speech Synthesis for Character Animation," *CHI+GI 1987 Conf. Proc.*, pp.143-147, Toronto, April(1987).
- [6] A. Lippman, "Semantic Bandwidth Compression : Speech-marker", *Picture Coding Symposium(PCS81)*, pp.29-30(1981).
- [7] H. Harashima, K.Aizawa and T. Satio, "Model-Based Analysis Synthesis Coding Videotelephone Image", *Trans. IEICE Japan*, E72, No.5, pp.452-459(1989).
- [8] K. Aizawa, H.Harashima and T.Satio, "Model-Based Analysis Synthesis Image Coding System for a Person's Face", *Imade Comm*, Vol.1, No.2, pp.139-152(1989).
- [9] 森島 繁生, 岡田 信一, 原島 傳, "知的インタフェースのための顔の表情合成法の検討", *日本電子情報通信學會論文誌*, J73-D-II, No.3, pp.351-359(1990).
- [10] 崔昌石, 原島 傳, 式部 幹, "顔の3次元モデルに基づく表情の記述と合成", *日本電子情報通信學會論文誌*, J73-A, No.7, pp.1270-1280(1990).