

얼굴 특징영역상의 광류를 이용한 표정인식 (Recognition of Human Facial Expressions using Optical Flow of Feature Regions)

이 미 애 * 박 기 수 **
(Mi-Ae Lee) (Ki-Soo Park)

요 약 표정인식 연구는 맨·머신 인터페이스 개발, 개인 식별, 가상모델에 의한 표정복원 등 응용가치의 무한한 가능성과 함께 다양한 분야에서 연구되고 있다. 본 논문에서는 인간의 기본정서 중 행복, 분노, 놀람, 슬픔에 대한 4가지 표정을 얼굴의 강체 움직임이 없는 얼굴동영상으로부터 간단히 표정인식 할 수 있는 방법을 제안한다. 먼저, 얼굴 및 표정을 결정하는 요소들과 각 요소의 특징영역들을 색상, 크기 그리고 위치정보를 이용하여 자동으로 검출한다. 다음으로 Gradient Method를 이용하여 추정된 광류 값으로 특징영역들에 대한 방향패턴을 결정한 후, 본 연구가 제안한 방향모델을 이용하여 방향패턴에 대한 매칭을 행한다. 각 정서를 대표하는 방향모델과의 패턴 매칭에서 그 조합 값이 최소를 나타내는 부분이 가장 유사한 정서임을 판단하고 표정인식을 행한다. 마지막으로 실험을 통하여 본 논문의 유효성을 확인한다.

키워드 : 표정인식, 특징영역, 광류, 방향모델, 방향패턴

Abstract Facial expression recognition technology that has potentialities for applying various fields is applying on the man-machine interface development, human identification test, and restoration of facial expression by virtual model etc. Using sequential facial images, this study proposes a simpler method for detecting human facial expressions such as happiness, anger, surprise, and sadness. Moreover the proposed method can detect the facial expressions in the conditions of the sequential facial images which is not rigid motion. We identify the determinant face and elements of facial expressions and then estimates the feature regions of the elements by using information about color, size, and position. In the next step, the direction patterns of feature regions of each element are determined by using optical flows estimated gradient methods. Using the direction model proposed by this study, we match each direction patterns. The method identifies a facial expression based on the least minimum score of combination values between direction model and pattern matching for presenting each facial expression. In the experiments, this study verifies the validity of the proposed methods.

Key words : Expression Recognition, Feature Region, Optical Flow, Direction Model, Direction Pattern

1. 서 론

얼굴은 언어를 제외한 인간의 신체에 있어 내적 심리 상태를 가장 잘 대변하는 부분이다[1,2]. 따라서 상대방의 표정을 보고 그 사람의 정서나 감정 상태를 인식하고자 하는 표정인식 연구는, 최근 컴퓨터의 성능 및 영상처리 기술의 발전과 더불어 컴퓨터와 사용자와의 자연스런 인터페이스 개발[3,4], 개인 식별[5], 가상모델에

의한 표정복원[6] 등 응용가치의 무한한 가능성과 함께 여러 분야에서 연구되고 있다.

표정인식 연구는 크게 두 가지로 나눌 수 있는데, 첫째는 정지얼굴영상의 표정을 인식하는 것으로 템플릿 매칭(Template Matching)[7], 특징치(Feature)[8], 3D 컴퓨터그래픽 얼굴 형상모델을 이용한 방법[9] 등이 제안되었다. 정지얼굴영상을 이용할 경우, 동영상에 비해 처리량은 적으나, 얼굴의 비강체적 변형이 큰 경우 매칭이나 정합에 의한 인식이 힘들고, 얼굴 표정영상을 자동으로 지정하기 어렵다는 단점이 있다. 둘째는 동영상에서 사용자의 얼굴변화를 분석하여 표정을 인식하는 연구로, 표정의 변화를 연속적으로 처리할 수 있어 실시간

* 정 회 원 : 국립한밭대학교 BK21사업단 교수
malee@hanbat.ac.kr

** 정 회 원 : 고신대학교 정보미디어학부 교수
pkisoo@kosis.ac.kr

논문접수 : 2004년 12월 14일
심사완료 : 2005년 4월 27일

맨 머신 인터페이스에 유용하다. 대표적 알고리즘으로는 인공신경망을 이용한 학습을 통해 인식하는 방법[10,11]과 얼굴 전체의 움직임 패턴에 주목하여 생성시킨 모션 필드의 데이터베이스와 이용자의 모션필드에 부합되는 3차원 모델과의 상관관계에 의해 인식하는 방법[12], 표정을 결정하는 요소들의 광류를 추정한 뒤, 얼굴 특징점들의 움직임으로부터 Ekman 등의 AU[13]를 이용해 인식하는 방법[14] 등이 있다. 그러나 방대한 처리량과 함께, 인공신경망을 위한 학습과 데이터베이스 구축을 위해 대량의 샘플들이 요구된다. 또한 얼굴영역과 표정을 결정하는 요소들의 자동 검출 및 특징점들의 자동 선택이 용이하지 않고, 비강체적 얼굴의 특성 상 개인차의 영향에 민감하다는 단점이 있다.

본 논문에서는 이러한 문제점을 개선하기 위해 얼굴의 강제적 움직임은 없으나 배경, 얼굴 크기를 고정하지 않은 얼굴동영상으로부터 표정을 결정하는 특징영역들을 자동으로 설정하고, 특징영역들의 움직임 방향을 이용해 표정인식할 수 있는 방법을 제안한다. 이를 위해 대역적 색상정보와 히스토그램을 이용한 피부색상 정보, 얼굴 크기 및 위치 정보를 이용하여 자동으로 얼굴 및 눈썹, 눈, 입술 요소들을 검출하고, 각 요소들의 특징영역을 설정한다. 다음으로 얼굴영역의 광류를 Gradient Method[15]를 이용해 추정한 후, 광류 u, v 값으로 각 특징영역들의 표정별 방향모델(Direction Model)을 구축한다. 움직임 방향을 학습한 방향패턴(Direction Model Pattern)과 테스트영상의 방향패턴(Test Images Pattern)과의 매칭을 행한 후, 매칭의 조합 값이 최소를 나타내는 부분이 가장 유사한 정서임을 판단하고 표정인식을 행한다.

본 논문의 구성은 다음과 같다. 2장에서 전처리 단계로 얼굴동영상으로부터 얼굴영역 및 표정인식에 관련된 각 요소들을 검출하는 방법을 제시하고, 광류 추정 방법을 서술한다. 3장에서 표정인식에 이용되는 AU와 그에 관련한 특징영역들을 결정하고 표정인식 알고리즘을 제안한 뒤, 4장에서 실험을 통한 구체적인 예를 보이고, 본 연구의 유효성을 검증한다. 마지막으로 5장에서 결론을 맺는다.

2. 전처리

표정을 결정하는 얼굴근육들은 근육과 연결된 피부를 움직이고 주름과 골을 만드는 등 다양한 반응을 수반하지만, 얼굴동영상에서 이 모든 것을 포함하여 처리하기에는 어려움이 있다. 심리학에서는 표정인식에 대한 실험을 통해, 인간은 상대방의 심리상태를 눈 또는 입만으로는 약 70% 이상, 눈과 입을 동시에 참조할 경우 97% 이상 표정인식 할 수 있음을 밝혔다[16].

본 장에서는 동영상에 가지는 방대한 처리량 압축을 목표로, 얼굴동영상으로부터 얼굴영역 및 밝기에 민감하고 얼굴근육의 움직임을 가장 잘 반영하는 각 요소, 즉, 눈썹, 눈, 입을 검출하는 방법과 광류 u, v 를 Gradient Method로 추정한다. 이 때, 얼굴동영상 획득을 위한 촬영조건으로 배경이나 영상 상의 얼굴의 위치는 고정하지 않으나, 얼굴의 강제 움직임이 없고 안경을 착용하지 않은 정면 얼굴을 전제로 한다. 또한 일반 연구실에서 형광등 조명을 사용하였고 촬영 시 조명변화는 없게 하였다.

2.1 얼굴영역 및 요소 검출

얼굴동영상으로부터 얻어지는 컬러정보는 그림 1과 같이 RGB로 표현된다. RGB 컬러를 그대로 사용할 경우 밝기에 매우 민감하다. 밝기정보를 제외한 색상공간으로 피부색에 대한 히스토그램을 나타내면 그림 2의 (a)와 같이 그 분포가 집중되어 있음을 알 수 있다. 이것은 밝기정보를 분리한 색상공간에서는 피부색영역이 어느 정도 안정적으로 확보할 수 있음을 의미한다.



그림 1 얼굴동영상

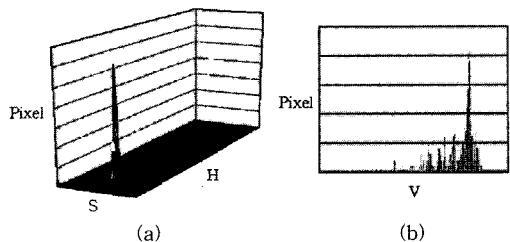


그림 2 피부색에 대한 HSV 히스토그램

본 논문에서는 사람의 시각체계와 비교적 유사하고 밝기정보를 용이하게 분리할 수 있는 HSV 색상공간을 이용하여, 두개의 색상정보(Hue, Saturation)와 밝기정보(Value)로 분리하였다[17]. 먼저 입력동영상으로부터 얼굴을 검출하기 위해, 대역적 피부 색상정보와 변환된 HSV 컬러 좌표계상의 히스토그램을 이용한 피부 색상정보로 얼굴후보영역을 지정한다. 피부색상에 대한 HS의 대역적 임계영역 값으로 화소들을 계수한 후, 그림 2의 (a)와 같은 3차원 히스토그램을 구한다. 그리고 (a)의 정점에 해당하는 화소들의 V값에 대해 2차원 히스토그램 (b)을 구한다. (a)와 (b)의 정점에 해당하는 화소값을 얼굴 대표 색상값으로 채택하여 그 이웃하는 화소

들과의 컬러값의 차가 문턱치($H = \pm 10, S = \pm 0.2$) 이하이면, 이웃 화소를 같은 영역으로 간주하여 영역을 확장하는 범람방식(Flood)으로 얼굴영역을 결정한다. 이때, 문턱치는 예비실험을 통해 최적값을 설정하였다.

피부색상 및 입술색상, 그리고 눈썹 색상정보의 대역적 임계영역은 웹상에서 얻은 얼굴영상(30매)과 직접 촬영한 얼굴영상(120매)의 최대치와 최소치를 적용한 임계값(표1)을 이용하였다. 본 방법은 고정된 피부색상 임계값을 사용하지 않음으로서 촬영 환경에 따라 변할 수 있는 임계값에 유연하게 대처할 수 있으며, 얼굴 크기, 위치에 상관없이 안정적으로 얼굴을 검출할 수 있다. 얼굴 후보영역이 결정되면 그 안에서 입술요소를 검출한다. 입술은 얼굴의 구조적 특성상, 주변이 피부색만 존재한다고 예상되므로, 유사색과의 병합을 배제할 수 있어 대역적 임계영역 값만으로 검출할 수 있다.

표 1 대역적 임계영역

대 상	피부 색상	입술 색상	눈썹색상
H 임계영역	$0 < H < 60$	$0 < H < 23$	$16 < H < 57$
S 임계영역	$0.13 < S < 0.68$	$0.25 < S < 1.0$	$0.18 < H < 0.47$

한편 얼굴색상과 유사한 배경이 영상에 존재할 경우, 검출된 얼굴후보영역이 얼굴임을 확정할 조건이 필요하다. 여기서는 그림 3의 얼굴의 모양정보와 입술의 위치정보를 참조하여 다음 세 가지를 판별조건으로 사용하였다. 첫째, 얼굴모양이 일반적으로 타원임을 고려할 때, $\text{얼굴횡축길이}(\overline{gh}) / \text{얼굴종축길이}(\overline{ac})$ 가 성립되어야 한다. 둘째, $\text{입술횡축길이}(\overline{mn}) / \text{얼굴횡축길이}(\overline{gh})$ 가 1:2.0~3.5 사이에 존재해야 한다. 셋째, 요소의 위치정보에 따라 입술영역이 $\text{얼굴종축길이}(\overline{ac})/2$ 의 아래영역 $\square ecdf$ 에 포함되어야 한다. 이 때, 얼굴 횡축길이는 귀 부분의 검출 유무에 따라 실제 값에 영향을 받을 수 있는데, 본 연구는 사전 실험을 통해 얼굴 횡축길이 측정에 가장 영향이 적은 곳이 입술 좌우 끝점의 연장부분과 얼굴 윤곽이 만나는 \overline{gh} 임을 확인하였다. 얼굴 후보영역이 위의 세 조건을 만족하면 피부색으로 결정된 얼굴의 종축 시종점(始終點)과 횡축 \overline{gh} 를 이용해 $\square acdb$ 가 얼굴영역임을 확정한다. 만약 비얼굴로 판별되면, 후보영역을 제외한 나머지 피부 색상영역을 이용하여 다시 후보영역을 결정하고 얼굴임을 판별한다. 연속적으로 3번까지 비얼굴로 확정되면 얼굴이 존재하지 않음으로 판단하고 알고리즘을 끝낸다.

다음은 확정된 얼굴영역에서 눈 요소를 검출한다. 눈은 그레이 영상의 low-level에선 비슷한 패턴을 이룬다는 점과 양안의 대비성에 착안하여 한 쪽(본 논문에서

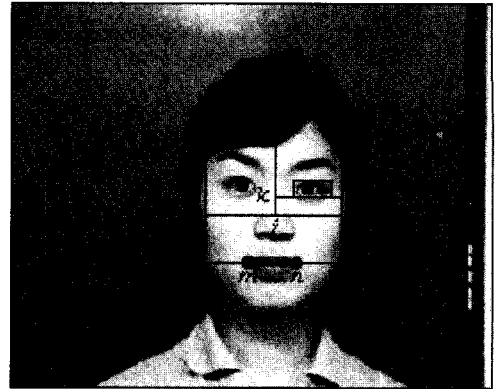


그림 3 얼굴의 모양정보와 요소의 위치 정보

는 왼쪽) 눈만 검출한다. 먼저 테스트영상 중, 얼굴 크기가 비슷한 10매의 눈을 조합하여 부분 템플릿을 구성한다. 그리고 얼굴위치정보 상, 얼굴면적 $\square acdb/2$ 의 위쪽인 $\square aefb$, 그리고 $\square aefb/2$ 의 왼쪽 영역에 위치함으로 그 영역에 주목하여 템플릿 매칭을 행한다. 이 때, 템플릿 크기는 입력영상의 얼굴크기에 따라 $\text{얼굴횡축길이}(\overline{gh})$: $\text{템플릿 횡축길이} = 4:1$ 을 적용하였다.

마지막으로 눈썹 요소를 검출하기 위해 눈썹의 위치정보와 색상정보를 이용한다. 일반적으로 눈썹은 얼굴의 종축 시점 i 와 확정된 눈의 종축 종점 k , 그리고 $\text{얼굴횡축길이}(\overline{gh})/2$ 로 이루어진 ($\square iklb$ - 눈영역) 안에 존재하며, 피부색상과 눈썹색상으로 나누어진다. $\square iklb$ 영역 안에서 입술요소 검출과 같은 방법으로 대역적 임계영역값으로 눈썹색상을 탐색한 뒤, 횡축과 종축의 시종점을 구하여 연결하였다.

본 방법은 배경의 유무, 얼굴 크기, 위치 등을 고정해야 하는 종래의 제약 조건을 완화시킴으로서 표정인식에 대한 작업환경을 좀 더 유연하게 대처할 수 있게 하였다.

2.2 광류 추정

얼굴동영상으로부터 얼굴과 표정을 결정짓는 요소들을 검출한 후, 표정변화에 따른 얼굴 근육의 움직임 추정을 위해 본 논문에서는 Gradient Method[15]를 이용하였다. 이 방법은 화소단위로 움직임을 추정할 수 있어 얼굴 근육의 움직임을 부분적으로 분석하는데 용이하다.

대상 물체의 농도값이 극소 시간 이동 후의 연속 두 영상 간에는 변화가 없다고 가정할 경우, 영상 상의 한 점 (x, y) 의 시각 t 에 있어서의 농도값을 $E(x, y, t)$ 라 할 때, 극소 시간 δt 이동 후의 $(x + \delta x, y + \delta y)$ 의 농도값은 같다고 볼 수 있다. 여기서 (x, y) 에 대한 광류를 $u(x, y), v(x, y)$ 라 두면, $\delta x = u\delta t, \delta y = v\delta t$ 가 성립하며 다

음 식 (1)을 얻을 수 있다.

$$E(x+u\delta t, y+v\delta t, t+\delta t) = E(x, y, t) \quad (1)$$

식 (1)의 좌변을 테일러 전개한 후, 고차 항을 무시하고 $\delta t \rightarrow 0$ 로 극한을 취하면 다음 식 (2)을 유도할 수 있다.

$$\frac{\partial E}{\partial x} \frac{dx}{dt} + \frac{\partial E}{\partial y} \frac{dy}{dt} + \frac{\partial E}{\partial t} = 0 \quad (2)$$

여기서 $u = \frac{dx}{dt}$, $v = \frac{dy}{dt}$, $E_x = \frac{\partial E}{\partial x}$, $E_y = \frac{\partial E}{\partial y}$,

$E_t = \frac{\partial E}{\partial t}$ 로 두면, 구속방정식 (3)을 구할 수 있다.

$$E_x u + E_y v + E_t = 0 \quad (3)$$

이 때, E_x , E_y , E_t 는 연속 두 영상 간의 농도변화량으로 영상 데이터로부터 직접 계산가능하나, 구속방정식 (3)에서 u, v 를 추정하기 위해서는 새로운 제약조건이 필요하다. 본 논문은 영상 상의 국소영역 내의 움직임은 일정하다는 가정 하에 최소자승법을 도입한 국소법을 이용하였다[18].

3. 표정인식

얼굴표정의 변화는 개인적 특성, 사회적 환경 또는 인종마다 그 표현의 정도가 조금씩 다르기 때문에 표정인식을 위한 정규화 작업은 용이하지 않다. 그럼에도 불구하고, 정도의 차이는 있으나 인종이나 서로 다른 문화권에서도 공통적으로 인식되는 보편성이 있음을 밝힌 Ekman 등은 인간의 정서를 행복, 놀람, 공포, 분노, 혐오, 슬픔의 6가지 표정으로 정의하였다. 그리고 얼굴근육을 바탕으로 하여 FACS(Facial Action Coding System)를 구성하고 얼굴의 움직임을 AU(Action Units)로 구분하였다[1,13].

AU를 활용한 종래의 연구를 살펴보면 크게 세 가지로 분류할 수 있다.

1) 2D, 3D로 제작된 CG 얼굴 형상모델과 얼굴의 정합을 통해 표정을 결정하는 특징점의 움직임을 검출하고 그 데이터를 AU와 비교 분석하여 표정을 인식하는 방법이 제안되었다. 그러나 이 경우, 비강체적 얼굴 특성 때문에 변형이 큰 경우 얼굴 특징을 추출하기 힘들고 정합에 시간이 많이 걸린다는 단점이 있다.

2) 얼굴 표정을 결정하는 특징영역을 지정하고 광류를 통해 움직임을 검출한 뒤, AU를 이용한 모션 특징과 표정과의 관계를 인공신경망을 이용한 학습을 통해 인식하는 방법이 제안되었다. 이 경우, 방대한 처리량과 함께 학습과 데이터베이스 구축에 필요한 대량의 샘플이 요구되는 단점이 있다.

3) 표정을 결정짓는 특징영역의 광류를 검출한 뒤, 그 래프를 이용한 움직임량의 패턴과 AU와의 대응으로 표정을 분석하는 방법이다. 그러나 이 경우, 얼굴 특징영

역의 자동결정이 어렵고, 움직임량의 크기에 있어 개인차의 영향에 민감하다는 단점이 있다.

이러한 문제점을 해결하기 위해 본 장에서는 인간의 기본 정서 중, 행복, 놀람, 분노, 슬픔의 4정서에 대응되는 AU를 살펴보고, 검출된 얼굴 각 요소들로부터 AU에 관련되는 특징영역들을 자동 설정하는 방법을 제안한다. 그리고 개인차의 영향을 최소화 하기위해 특징영역들에 대한 광류의 방향 값을 파라미터로 이용하여 표정인식하는 알고리즘을 제시한다.

3.1 표정인식에 관련된 AU와 특징영역 설정

Ekman 등은 얼굴 및 각 요소들의 움직임을 총 64가지의 AU로 분류하였다[19]. 또한 일반적으로 인간은 같은 정서 변화에 의한 얼굴의 움직임이 공통적인 얼굴근육을 이용한다는 점과, 근육에 따라 그룹화된 형태로 표현됨을 밝혔다. 종래의 표정복원과 표정인식 연구를 토대로 4가지 정서에 관련된 AU의 구성을 살펴보면 표 2와 같이 조합할 수 있다. 그리고 여기에 이용된 AU 리스트는 표 3에 나타내었다.

표 2 4가지 정서에 대한 AU의 조합

Emotion	AU No.
Happiness	5, 12, 26
Anger	4, 5, 7, 10, 25, 26
Surprise	1, 2, 5, 26
Sadness	1, 4, 15

표 3 표정 연구에 관련된 AU 리스트

AU No.	FACS Name
1	Inner Brow Raiser
2	Outer Brow Raiser
4	Brow Lowerer
5	Upper Lid Raiser
7	Lid Tightener
10	Upper Lip Raiser
12	Lip Corner Puller
15	Lip Corner Depressor
25	Lip Part
26	Jaw Drop

본 연구는 2.1에서 제안한 방법으로 요소들을 검출한 후, 표 2의 각 정서에 따른 AU의 조합을 바탕으로 특징영역들을 다음과 같이 자동 설정한다. 이 때, 다양한 얼굴 사이즈의 입력을 고려하여 특징영역의 크기는 검출된 요소 크기의 비율에 의해 결정된다.

(1) 눈썹 요소

눈썹은 내측을 올림(AU1), 외측을 올림(AU2), 전체를 내림(AU4)의 3가지 AU와 관련되며, 눈썹요소의 위치정보와 색상정보를 참조하여 특징영역을 규정할 수

있다. 먼저, 2.1에 의해 검출된 눈썹 영역에서(그림 4) 각 AU에 대응되는 정점을 구하면,

- AU1 : 내측 눈썹의 정점으로 횡축 시점과 종축의 첫 눈썹색상 화소를 만나는 곳.
- AU2 : 외측 눈썹 정점으로 횡축 종점과 종축의 첫 눈썹색상 화소를 만나는 곳.
- AU4 : 눈썹 전체 상하 움직임을 체크할 수 있는 정점으로 종축 시점과 횡축의 첫 눈썹색상 화소를 만나는 곳.

으로 규정할 수 있다. 다음은 각 정점에 근접한 주변 영역의 광류는 비슷한 값을 가진다고 생각되어지므로, 정점을 중심으로 특징영역을 설정하였다. 특징영역의 크기는 (눈썹 횡축길이/3)로 정하였다.

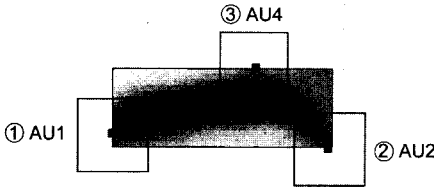


그림 4 AU와 관련된 눈썹 요소의 특징영역

(2) 눈 요소

눈은 눈꺼풀을 올림(AU5), 눈꺼풀을 긴장시킴(AU7)의 2가지 AU와 관련된다. 2.1에서 검출된 눈 요소(그림 5)의 위치정보를 참조하여 눈꺼풀의 움직임을 가장 잘 파악할 수 있는 (눈 횡축길이/2) 지점과 눈의 종축 시점을 정점으로 규정하고, 눈썹 특징영역에 사용된 크기를 이용하여 눈 특징영역을 설정하였다.

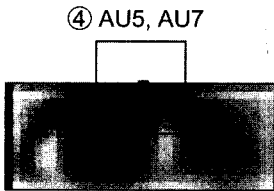


그림 5 AU와 관련된 눈 요소의 특징영역

(3) 입술 요소

입술은 윗입술을 올림(AU10), 입꼬리를 끌어올림(AU12), 입꼬리를 내림(AU15), 턱을 움직이지 않고 아랫입술을 내림(AU25), 턱을 내리면서 입술을 벌림(AU26)의 5가지이다. 각 AU들과 관련하여 규정지을 수 있는 특징영역은 입술요소의 모양 정보와 위치정보를 참조하여 설정할 수 있다. 2.1에 의해 검출된 입술 영역에서(그림 6) 각 AU에 대응되는 정점을 구하면,

- AU10 : 윗입술의 움직임을 현저히 파악할 수 있는 정점으로, (입술 횡축길이/2)지점과 입술의 종축 시

점이 만나는 곳.

- AU12, AU15 : 입꼬리의 정점으로 입술의 횡축 종점과 종축의 첫 입술색상 화소가 만나는 곳.
- AU25, AU26 : 아랫입술의 움직임과 연계되는 정점으로, 입술 횡축길이/2)지점과 입술의 종축 종점이 만나는 곳.

으로 정하였다. 정점을 중심으로 한 특징영역의 크기는 (입술 횡축길이/3)로 설정하였다.

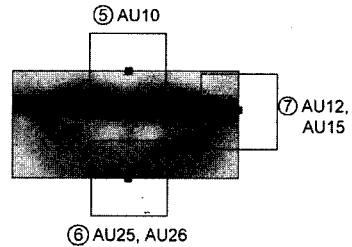


그림 6 AU와 관련된 입술 요소의 특징영역

3.2 표정인식 알고리즘

본 논문은 3.1을 통해 얻어진 특징영역들의 광류 값을 파라미터로 하여 각 정서에 대한 표정인식을 다음과 같은 알고리즘으로 행한다.

먼저, 어떤 특징영역을 F_k 라 하면(이 때, $k=1, \dots, 7$), 각 특징영역들의 광류는 식 (4)을 이용해 구할 수 있다.

$$F_{ku} = \frac{1}{\text{number of flow}} \sum u \tag{4}$$

$$F_{kv} = \frac{1}{\text{number of flow}} \sum v$$

다음은, 심리상태에 따른 표정변화와 AU의 연관성을 조합한 결과를 토대로 종래 표정복원에 적용된 AU(표 2) 및 얼굴 근육에 대한 해부학적 특성, 그리고 예비실험에서 학습데이터를 이용해 표 4와 같은 4가지 정서에 대한 각 특징영역의 방향모델(Direction Model)을 구축한다. 이 때, 표정의 강도, 즉 광류의 크기는 비강체적 얼굴의 특성상 피험자의 개인차에 크게 의존하게 되는데 AU의 방향에 바탕을 둔 본 알고리즘에서는 그 값을 제외함으로써, 개인차의 영향을 감소시킬 수 있다.

표정인식을 위해 얼굴표면의 움직임을 단순히 얼굴근육의 해부학적 구조로만 해석하기에는 내부구조의 복잡성과 근육간의 연계성 문제 때문에 어려움이 있다. 본 연구는 표정을 만드는데 기여도가 높은 얼굴근육들이 같은 위치와 같은 수축방향을 가질 경우 하나의 근육으로 분류된 그림 7[15]을 이용하였다. 여기서는 얼굴근육을 상하로 나누고 있으며, 눈썹요소는 해부학적 특성상 눈썹의 종축운동을 주관하는 근육들을 동시에 움직일 수 없는 저항관계에 있음을 나타내고 있다. 본 논문은

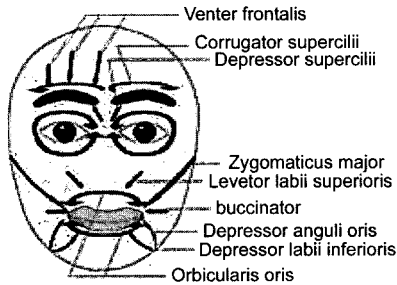


그림 7 표정변화에 관련한 얼굴근육

이러한 특성에 주목하고, 본 실험을 위해 준비한 표정 데이터 100set(1명*4표정⇒1set) 중 임의로 채택한 20set의 특징영역에 대한 세부적인 움직임은 예비실험을 통해 분석하였다. 그 결과, 눈썹 특징영역 ①, ②, ③에서 종축 움직임을 주관하는 광류 v 값이 같은 방향패턴(Direction Pattern)을 나타냄을 확인하였다. 즉, 눈썹 특징영역 ①의 움직임방향이 ↗, ↓, ↘을 나타내면 특징영역 ②, ③의 광류 v 값의 방향도 아래쪽으로 향함을 알 수 있었다. 이러한 움직임 패턴은 동일한 내적상태의 표정임에도 불구하고 행복과 같이 피험자의 개인차에 의해 눈썹의 움직임이 크게 위쪽 또는 아래쪽으로 나뉘는 경우, 그 판별에 용이하다.

눈요소의 특징영역 ④, 그리고 입술요소의 특징영역 중 ⑤, ⑥은 일반적으로 눈과 입의 개폐에 따른 종축 운동에 의존하므로 광류 v 값을 파라미터로 하여 방향패턴을 정하였으며, ⑦은 광류 u, v 값을 이용하여 방향패턴을 결정하였다. 광류 v 값을 파라미터로 하여 종축의 움직임만 참조하는 특징영역 ④, ⑤, ⑥과는 달리 ①, ②, ③, ⑦은 광류 u, v 의 기울기각을 이용하여 방향패턴을 정하였다. 즉, 그림 8의 (a)와 같이 u, v 에 의해 구해진 기울기각 θ 가 (b)의 8영역 중 어느 곳에 속하는지 계산한 뒤, 구해진 영역의 대표 방향을 그 방향패턴으로 결정한다.

위의 과정을 통하여 획득한 방향모델은 각 특징영역에 따라 하나 또는 그 이상의 방향패턴을 포함한다. 또한 방향 패턴의 종류도 빈도수에 따라 복선화살(Double Arrow, 예: ↔)과 단선화살(Single Arrow, 예: ↗)로

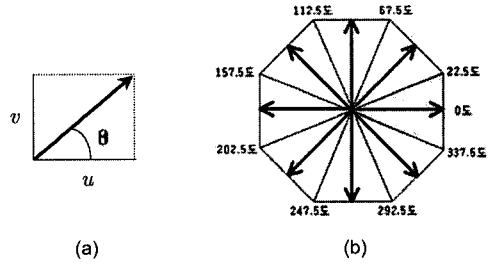


그림 8 기울기각을 이용한 방향패턴

구분하였다. 이는 예비실험을 통하여 얻어진 결과를 바탕으로 피험자들의 표정의 강도 및 개인차를 최대한으로 수용하기 위함이며, 특징영역별로 나타난 방향의 빈도수를 참조하여 복선화살로 표시한 방향패턴이 단선화살의 방향패턴보다 높은 빈도수를 가짐을 나타내었다. 앞에서 언급한 것처럼 눈썹은 종축운동을 주관하는 근육들을 동시에 움직일 수 없는 관계에 있으므로, 특히 행복의 경우, 특징영역 ①의 방향이 ↗ 또는 ↓을 나타내면 ②, ③의 방향패턴은 ↓으로 판별한다. 그 반대로 ↘, ↑, ↗ 셋 중 하나의 방향을 나타내면 ②, ③의 방향패턴은 ↑으로 판별한다. 그 외에는 주어진 방향모델의 패턴을 이용하면 된다.

마지막으로 위의 방법에 의해 구축된 방향모델을 이용하여 표정인식을 행하였다. 4가지 정서 중 하나를 표현한 임의의 테스트영상(Test Images)에 대해 각 특징영역의 광류를 추정한 후, 방향패턴을 계산하여 방향모델과의 매칭을 실시한다. 이 때, 분류방법에 있어 범주형 속성을 포함한 객체의 유사성 척도에 응용되는 이분형 변수[20]를 0, 1, 2 값으로 확장하고, 유사성 척도로 단순매칭(simple matching)을 이용하였다. 방향모델(표 4)에 있어서의 각 특징영역의 방향패턴을 Direction Model Pattern(DMP)이라 하고 테스트영상에서 얻어진 방향패턴을 Test Images Pattern(TIP)이라 할 경우, 식 (5)를 이용해 그 유사성을 평가한다.

$$sim(DMP_{E,k}, TIP_{EK}) = \begin{cases} 0 & (\text{Only, when the direction of double arrow is same.}) \\ 1 & (\text{Only, when the direction of single arrow is same.}) \\ 2 & (DMP_{E,k} \neq TIP_{EK}) \end{cases} \quad (5)$$

표 4 4가지 정서에 대한 방향모델

Feature region Emotion	Inner Brow (①)	Outer Brow (②)	Brow (③)	Upper Lid (④)	Upper Lip (⑤)	Lower Lip (⑥)	Lip Corner (⑦)
Happiness	↗↘↕↖↗	↑ ↓	↑ ↓	↑ ↓	↑	↑ ↓	↑ ↗
Anger	↓ ↘	← ↘	↓	↓	↑	↑	↑
Surprise	↑ ↘	↑	↑	↑	↓ ↑	↓	↘
Sadness	↘ ↓	← ↓ ↘	↓	↓	↑	↑	↑ ↘ ↓

이 때, $E_i(i=0, \sim, 3)$ 는 4정서 중 하나를 가리키며, k 는 특징영역을 나타내는 것으로 $k=1, \dots, 7$ 을 갖는다. DMP에 대한 TIP의 매칭이 이루어지고 난 후, 식 (6)을 통해 최소값을 갖는 $sim(DMP_{E_i}, TIP_E)$ 가 i 번째 정서임을 인식한다.

$$sim(DMP_{E_i}, TIP_E) =$$

$$\min_{i=0}^3 \sum_{k=1}^7 sim(DMP_{E_i, k}, TIP_{E, k}) \quad (6)$$

만약, 매칭의 결과가 두개 이상의 정서를 나타내는 동일한 최소값을 획득할 경우에는, 특징영역 중 움직임 변



그림 9 4정서에 대한 표정인식 성공 예

화가 가장 잘 표현되는 Lip Corner의 매칭값이 최소가 되는 정서를 우선 값으로 인식한다. 이 부분도 동일한 값을 가질 경우에는 최소값으로 나타난 모든 표정을 인식대상으로 한다.

4. 실험 결과

본 실험은 하드웨어가 Pentium IV 1.8GHz이고 OS가 Windows XP 환경인 컴퓨터에서 VC++6.0으로 구현되었다. 얼굴동영상 획득에 사용된 입력 장치는 CCD카메라(SONY DCR-PC101)로 해상도는 320×240이다. 피험자는 20~30세로 구성된 남녀 100명을 대상으로 하였으며, 촬영 전 미리 피험자들에게 4가지 정서에 대해 시사(示唆)하고 그에 따른 4가지 표정만 요구하였다. 그리고 촬영은 무표정에서 시작하여 하나의 정서에 따른 표정이 생성되는 순간까지를 One Scene으로 하였다. 촬영 장소는 일반 연구실로, 배경과 얼굴 위치는 제한을 두지 않았으나 촬영 시에는 얼굴의 강제 움직임이 없고 안경을 끼지 않은 정면얼굴로 하였다. 또한, 얼굴 크기는 고정하지 않았으나 표정 파악이 용이하도록 얼굴 횡축 넓이가 영상 횡축해상도/4 이상이 되도록 하였다. 본 실험에서는 얼굴 크기를 대략 두 종류(120×150 정도 80set, 90×120 정도 20set)로 나누어 실험하였다. 촬영 시 조명의 변화는 없도록 전제하였다. 실험 전, 영상 전체에 대한 광류를 추정하고 그 결과 머리의 움직임이 있는 데이터에 한해서는 재 촬영을 하는 방식으로 피험자 1명*4정서를 1set로 하여 총 100set의 데이터를 준비하였으며, 방향모델 구축에 사용된 학습데이터 20set를 제외한 80set(120×150 정도 64set, 90×120 정도 16set)로 실험을 행하였다.

본 논문의 제안방법으로 표정인식을 행하기 위해, 먼저 CCD 카메라를 통해 획득한 얼굴동영상 중 첫 번째 Frame(무표정 영상)에 대한 얼굴영역 및 요소들을 검출한 후 특징영역들을 지정한다. 이 때, 실험에서는 얼굴의 형태적 특성을 고려하여 횡축에 대해 좌우 각각 15픽셀씩 여유를 주어 얼굴을 검출하였다. 여기서 지정한 얼굴 및 특징영역들은 그 뒤에 연속되는 얼굴동영상에 대해서도 그대로 적용된다. 다음은 얼굴영역의 광류를 Gradient Method를 이용해 추정한 후, 광류 u, v 값으로 특징영역들에 대한 방향패턴을 결정하고, 본 논문이 제시한 방향모델의 방향패턴(Direction Model Pattern)과 테스트영상의 방향패턴(Test Images Pattern)과의 매칭을 행한 후, 매칭의 조합 값이 최소를 나타내는 부분이 가장 유사한 정서임을 판단하고 표정인식을 행한다.

다음 그림은 남녀 피험자의 4정서 2set에 대한 실험 결과이다. 배경을 고정하지 않은 상태에서 얼굴, 요소 및 각 특징영역들을 자동으로 검출하고 있음을 확인할

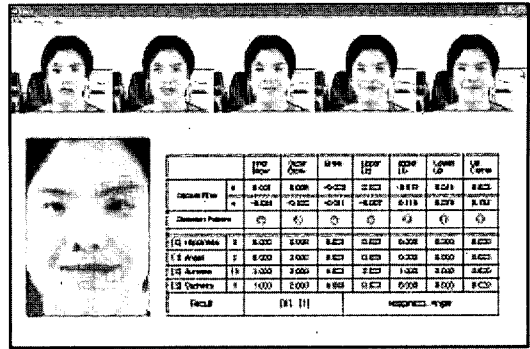


그림 10 두 표정이 인식된 경우

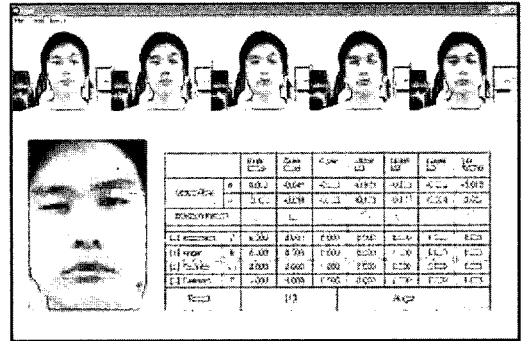


그림 11 인식이 실패한 경우1

수 있다. 이 때, 결과 영상(마지막 Frame)에서 요소들이 조금씩 벗어나 검출될 수 있는데, 이것은 첫 Frame에서 검출한 요소 및 특징영역들을 그대로 전 영상에 적용하고 있기 때문이다. 그러나 얼굴근육의 특성 상 표정 변화에 따른 얼굴의 움직임은 연속적으로 같은 방향 패턴을 가지며, 근접한 영역의 움직임 또한 비슷한 방향을 취한다고 볼 수 있다. 따라서 방향값을 파라미터로 하는 본 알고리즘에서는 첫 Frame의 정보 값으로 검출한 각 영역들을 One Scene 전체에 적용할 수 있다.

다음은 두개의 표정으로 인식된 결과를 나타내었다. 이 얼굴동영상의 경우, 눈썹의 움직임이 거의 없기 때문에 눈썹에 대한 방향 패턴이 제대로 구해지지 않았으며, 이로 인해 각 표정끼리의 패턴 특성이 감소되어 제대로 매칭이 이루어지지 않은 것으로 생각한다. 본 실험에서는 두 개 이상의 표정인식이 이루어진 경우, 성공률을 (100/인식개수)%로 계산하였다.

마지막으로 표정인식에 실패한 경우의 결과를 검증한다. 그림 11은 슬픔을 표현한 얼굴동영상에 대해 분노의 결과가 나타낸 예이다. 본 실험에서는 4가지 표정 중 슬픔이 가장 낮은 인식률을 나타내었다. 그 원인으로 생각되는 것은 슬픔의 움직임 패턴이 분노와 행복 사이에서

중첩되는 부분이 적지 않기 때문으로 여겨진다. 따라서 그림 11처럼 슬픔에 대한 얼굴표정을 잘 표출하지 못한 경우 오 인식이 발생하였다.

얼굴 해상도가 약 120×150 정도의 표정데이터 64set를 실험한 결과 행복, 분노, 놀람에 대한 인식률은 90% 이상, 슬픔에 대해서는 약 77%의 인식률을 획득하였다.

그림 12는 얼굴 크기를 다르게 한 경우의 예이다. 실험 결과, 영상 상의 얼굴 크기가 작아질수록 오 인식률이 증가하였으며, 그림 12 정도의 얼굴 크기(90×120)를 가진 표정데이터 16set를 실험한 결과, 표 5와 같은 인식률을 얻었다. 축소된 특징영역의 광류 값이 잡음 등의 영향을 민감하게 받은 것으로 추정된다.

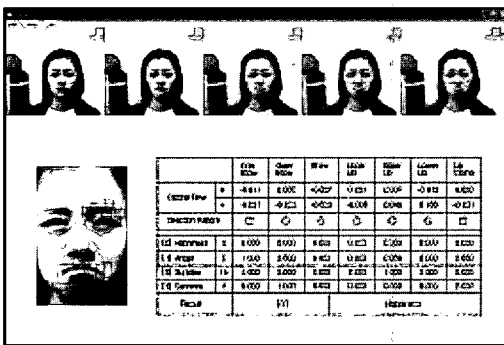


그림 12 인식이 실패한 경우2

표 5 표정인식 결과

Emotion	Recognition rate(%)	Face Size about 120×150 (64set)	Face Size about 90×120 (16set)	Average
Happiness		92.2	81.2	86.7
Anger		90.6	75.0	82.8
Surprise		100.0	93.7	96.9
Sadness		76.6	62.5	69.6

본 연구에서 대체로 낮은 인식률을 보인 슬픔에 대해서는 특징요소들의 추가 등으로 인식률을 높여 갈 계획이다. 또한 실험에서 사용된 One Scene의 Frame수는 10~30개로 Frame 분량에 따라 하나의 표정인식을 위한 처리 시간이 달라지는 문제점이 있다. 실시간 처리를 위해 앞으로 규격화된 표정데이터의 획득이 이루어질 수 있는 시스템 구축이 필요하다고 생각된다.

5. 결론

인간은 얼굴표정의 표현 강도가 상당한 개인성을 가지며 연령, 피부색, 주름, 음성 등 그 외 부차적 정보의 복합적인 조합에 의해 표정인식이 이루어진다고 볼 때,

얼굴의 움직임 정보만으로 인간의 표정인식 시스템을 구성하기란 상당히 열악한 면이 있다. 그럼에도 불구하고 심리학에서는 표정인식에 대한 실험을 통해, 인간은 상대방의 심리상태를 눈 또는 입만으로는 약 70% 이상, 눈과 입을 동시에 참조할 경우 97% 이상 표정인식 할 수 있음을 밝혔다.

본 논문에서는 동영상이 가지는 방대한 처리량 압축을 목표로, 인간의 기본정서 중 행복, 분노, 놀람, 슬픔에 대한 4가지 표정을 얼굴동영상으로부터 간단히 표정 인식할 수 있는 방법을 제안하였다. 먼저, 얼굴 및 얼굴 표정을 결정하는 눈썹, 눈, 입 요소들과 각 요소의 특징 영역들을 색상, 크기 및 위치정보를 이용하여 자동으로 검출하였다. 다음으로 얼굴영역의 광류를 Gradient Method를 이용해 추정된 후, 광류 u, v값으로 각 특징 영역들의 표정별 방향모델(Direction Model)을 구축하고, 움직임 방향을 학습한 방향패턴(Direction Model Pattern)과 테스트영상의 방향패턴(Test Images Pattern)과의 유사도를 이용하여 표정을 인식하였다. 실험 결과, 행복, 놀람에 대해선 86% 이상, 분노에 대해선 약 83%, 슬픔에 대해선 약 70%의 인식률을 얻었다.

본 논문에서는 촬영 시 배경, 얼굴 위치나 크기에 제약을 두지 않았으나, 정면얼굴로서 얼굴의 강제 움직임이 없는 것을 전제로 하고 있다. 향후 좀 더 자연스러운 맨 머신 인터페이스 환경을 구축하기 위해서는 사용자가 이용하기 용이한 촬영환경이 필요하다. 또한 인식률 향상을 위해 표정 표현이 가장 두드러지는 입술 형상의 추가적 매칭, 음성 정보를 조합하는 연구가 필요하다.

참고 문헌

- [1] P. Ekman, W. V. Friesen, Unmasking the Face, Prentice-Hall, Inc., Englewood Cliffs, 1975.
- [2] A. Mehrabian, J. A. Russell, An Approach to Environmental Psychology, MIT Press, 1974.
- [3] H. Kobayashi, F. Hara, "A basic study of man-machine interactive communication by using facial expressions," 9th Symposium on Human Interface, pp. 361-366, 1993.
- [4] O. Hasegawa, K. Sakaue and S. Hayamizu, "A Human-Like software robot which interactively learns and manages visual information in real world," IEICE, Vol. J82-D-II, No.10, pp. 1666-1674, 1999.
- [5] O. Yamaguchi, K. Fukui, "Smartface-A robust face recognition system under varying facial pose and expression," IEICE, Vol. J84-D-II, No.6, pp. 1045-1052, 2001.
- [6] K. Aizawa, H. Harashima and T. Saito, "A model-based analysis synthesis image coding scheme," IEICE, Vol. J72-D-I, No.3, pp. 200-207, 1989.

- [7] P. Vanger, R. Honlinger and H. Haken, "Application of synergistic in decoding facial expressions of emotion," Proceedings of IEEE International Workshop on Automatic Face and Gesture Recognizing, pp.44-49, 1995.
- [8] F. Hara, H. Kobayashi, "Use of face Robot for Human-Computer communication," IEEE Conference on SMC, pp.1515-1520, 1995.
- [9] C. Choi, H. Harashima and T. Takebe, "Analysis of facial expressions using Three-Dimensional facial model," IEICE, Vol. J74-D-II, No.6, pp. 766-777, 1991.
- [10] Y. Tian, T. Kanade and J. F. Cohn, "Recognizing actions units for facial expression analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No.2, pp. 97-115, 2001.
- [11] M. Rosenblum, Y. Yacoob and L. S. Davis, "Human expression recognition from motion using a radial basis function network architecture," IEEE Transactions on Neural networks, Vol. 7, No.5, pp. 1121-1138, 1996.
- [12] I. Essa, A. Pentland, "Facial expression recognition using image motion," Kluwer Academic Publishers, Computational Imaging and Vision Series, 1997.
- [13] P. Ekman, W. V. Friesen, Facial Action Coding System, Palo Alto, CA, Consulting Psychologists Press, 1978.
- [14] K. Mase, "Recognition of facial expression from optical flow," IEICE, Vol. E74, No.10, pp. 3474-3483, 1991.
- [15] B. K. P. Horn, B. B. Schunck, "Determining optical flow," Artificial Intelligence, Vol. 17, pp. 185-203, 1981.
- [16] H. Ohashi, G. Xu, S. Tsuji, "Recognition of facial expressions by neural network: A study on input data," IEICE, PRU-91-4, pp. 23-29, 1991.
- [17] 이미애, 박기수, "컬러정보와 부분 템플릿을 이용한 얼굴영역, 요소 및 회전각 검출", 정보처리학회 논문지, 제10-B권, 제4호, pp. 465-472, 2003.
- [18] J. K. Kearney, W. B. Thompson and L. Boley, "Optical flow estimation: An error analysis of Gradient-based methods with local optimization," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 9, No.2, pp. 229-244, 1987.
- [19] FACS-Facial Action Coding System(Ekman and Friesen), <http://www-2.cs.cmu.edu/~face/facs.htm>
- [20] N. Bratchell, "Cluster analysis," Chemometrics and Intelligent Laboratory systems, pp. 105-125, 1989.



이 미 애

1987년 부경대학교 응용수학과 졸업(학사). 1995년 일본고베대학원 시스템공학과(공학석사). 1998년 일본고베대학원 인공지능학과(공학박사). 2000년 한국과학기술원 연수연구원. 2001년~현재 한밭대학교 BK21사업단 계약교수. 관심분야는

컴퓨터비전, 패턴인식, 컴퓨터 그래픽 등



박 기 수

1988년 부경대학교 응용수학과 졸업(학사). 1996년 일본고베대학원 정보지능공학(공학석사). 1999년 일본고베대학원 인공지능학과(공학박사). 1999년~2000년 한남대학교 산업연구소 선임연구원. 2001년~현재 고신대학교 컴퓨터과학부 조교

수. 관심분야는 컴퓨터비전, 가상현실, 인공지능 등