

광류와 표정 HMM에 의한 동영상으로부터의 실시간 얼굴표정 인식[☆]

Realtime Facial Expression Recognition from Video Sequences Using Optical Flow and Expression HMM

전 준 철*
Junchul Chun

신 기 한**
Gihan Shin

요 약

비전기반 인간컴퓨터 상호작용은 컴퓨터와 인간의 상호소통을 자연스럽게 제공하는 측면에서 과학과 산업분야에서 주목받는 연구 분야이다. 그러한 측면에서 얼굴표정인식에 의한 인간의 심리적 상태를 추론하는 기술은 중요한 이슈이다. 본 연구에서는 감성인식 HMM 모델과 광류에 기반한 얼굴 움직임 추적 방법을 이용하여 동영상으로부터 얼굴표정을 인식하는 새로운 방법을 제시하였다. 특히, 기존의 감성상태 변환을 설명하는 HMM 모델은 특정 표정상태 간의 전환 시 항상 중립 상태를 거치도록 설계되어 있다. 그러나 본 연구에서는 기존의 표정상태 전환 모델에 중간상태를 거치는 과정 없이 특정 표정 상태간의 변화가 가능한 확장된 HMM 모델을 제시한다. 동영상으로부터 얼굴의 특성정보를 추출하기 위하여 템플릿 매칭과 광류방법을 적용하였다. 광류에 의해 추적된 얼굴의 표정특성 정보는 얼굴표정인식을 위한 HMM의 매개변수 정보로 사용된다. 실험을 통하여 제안된 얼굴표정인식 방법이 실시간 얼굴 표정인식에 효과적임을 입증하였다.

Abstract

Vision-based Human computer interaction is an emerging field of science and industry to provide natural way to communicate with human and computer. In that sense, inferring the emotional state of the person based on the facial expression recognition is an important issue. In this paper, we present a novel approach to recognize facial expression from a sequence of input images using emotional specific HMM (Hidden Markov Model) and facial motion tracking based on optical flow. Conventionally, in the HMM which consists of basic emotional states, it is considered natural that transitions between emotions are imposed to pass through neutral state. However, in this work we propose an enhanced transition framework model which consists of transitions between each emotional state without passing through neutral state in addition to a traditional transition model. For the localization of facial features from video sequence we exploit template matching and optical flow. The facial feature displacements traced by the optical flow are used for input parameters to HMM for facial expression recognition. From the experiment, we can prove that the proposed framework can effectively recognize the facial expression in real time.

☞ Keywords : HMM(Hidden Markov Model), optical flow, facial expression recognition, facial motion tracking
은닉 마코프 모델, 광류, 얼굴표정 인식, 얼굴 움직임 추적

1. 서 론

얼굴 표정은 사람의 감정 뿐 만 아니라 마음의 상태, 사회적 상호작용, 생리학적 신호 등과 같은 다양한 정보를 반영한다. 최근 컴퓨터 비전, 컴퓨터 그래픽스 그리고 HCI 분야에서 실시간 얼굴 표정 인식 및 분석 기술이 주요 관심사로 부각되고 있다. 특히 비전 기반 실시간 얼굴 모션 추적

* 종신회원 : 경기대학교 컴퓨터과학과 교수
jcchun@kgu.ac.kr

** 준 회 원 : (주) 핸디소프트 연구원
dalgundal@naver.com

[2009/01/27 투고 - 2009/01/30 심사 - 2009/02/23 심사완료]

☆ 본 연구는 2007학년도 경기대학교 학술연구비(일반연구과제) 지원에 의해서 수행되었음.

및 표정 인식 기술은 효과적인 인간 컴퓨터 상호작용을 위한 매력적인 입력 모드로 관심을 모으고 있다. 이처럼 얼굴 표정 인식에 관련된 연구의 중요성이 증대되고 있는 이유는 컴퓨터 성능의 향상에 따라 저렴한 비용으로 빠른 처리가 가능할 뿐 만 아니라 얼굴 검출, 얼굴 추적, 얼굴 인식 등과 같은 영역에서의 연구와 밀접한 연관성이 존재하여 상호보완적으로 연구 수행을 향상시키기 때문이다.

얼굴 표정의 동적 변화를 실시간으로 분석하기 위해서는 얼굴 동작의 시간적 변이를 효과적으로 반영할 수 있는 최적의 표정 정보의 추출과 실시간 추적이 필요하며 특정 표정 사이의 변화를 능동적으로 설명할 수 있는 표정변화 모델에 기반한 얼굴 표정의 해석 방법이 필요하다. 그러나 실시간 분석을 위한 얼굴 표정 변화의 추적은 조명 및 배경의 변화에 따라 매우 어려운 작업으로 인식되고 있다.

본 논문에서는 얼굴의 주요한 최소의 특징점의 움직임을 실시간으로 추적하고 특징점들의 변위를 확장된 표정 HMM 모델에 적용하여 얼굴 표정인식을 수행하는 시스템을 제안하였다. 제안된 시스템은 얼굴 영역 검출, 얼굴 특징점 추적, 그리고 얼굴 표정 인식의 세단계로 구성된다. 얼굴 영역 검출 단계는 얼굴 표정 인식에 있어서 가장 선행되어야 하는 단계로 얼굴의 후보 영역을 검출한 뒤 검출된 후보 영역들 중에서 정확한 얼굴 영역을 검출하는 과정으로 구성된다. 얼굴 후보 영역을 검출하기 위하여 피부색을 표현하기에 적합한 스킨컬러모델을 이용하여 후보 영역을 검출한다. 검출된 후보영역에서 주성분분석과 템플릿 매칭을 이용하여 최적의 얼굴 영역을 검출한다. 얼굴 후보 영역과의 비교를 위한 얼굴 템플릿 영상에 대하여 주성분 분석을 적용하여 고차원의 데이터를 저차원의 데이터로 축소하여 비교하므로 연산량을 감소시켜 실시간 처리가 가능하도록 하였다.

얼굴표정 변화에 따른 주요 특징점의 추적을 위하여 광류(Optical Flow) 알고리즘을 적용하였

다. 검출된 얼굴 영역으로부터 눈썹, 눈, 입의 위치를 결정하고 광류 알고리즘을 이용하여 매 프레임마다 특징점의 움직임을 추적한다. 이때 획득된 움직임 벡터는 표정 인식을 위한 매개변수로 사용된다.

최종 단계인 얼굴 표정 인식은 위해 은닉 마르코프 모델(HMM)을 이용하여 수행한다. 기존의 표정 인식 방법에 사용된 모델보다 다양한 표정간의 변화 패턴을 인식할 수 있는 확장된 모델을 통해 추출된 표정 인식을 수행한다.

2. 관련 연구

1971년 Ekman과 Friesen[1]에 의해 행복, 슬픔, 공포, 혐오, 놀람, 화남의 주요 감정표현 방법이 제안된 이후 얼굴표정 인식에 관한 다양한 연구가 진행되고 있다[2,3,4,5,6,7,8]. 얼굴표정인식은 정지영상 또는 동영상을 이용하여 진행되며, 동영상을 사용하는 경우 얼굴 표정의 변화폭이 큰 관계로 일반적인 패턴인식 방법에 비하여 비용이 많이 소용된다.

일반적으로 얼굴표정 인식 시스템은 얼굴 특징 벡터의 측정단계와 얼굴표정의 인식의 두 단계로 구성된다. 얼굴 표정은 눈썹, 눈, 코, 입, 피부와 같은 얼굴 요소들의 변형이 일어날 때 발생하는 얼굴 근육의 수축에 의하여 나타나며, 얼굴 표정의 강도는 얼굴 특징의 기하학적 변화 또는 근육 표현의 밀도에 따라서 결정된다. 이러한 얼굴의 특징 추출 방법으로는 크게 변형기반 방법과 모션기반 방법이 있다.

변형기반 방법은 영상기반과 모델기반 방법이 있으며, 전자의 방법은 얼굴 영상의 전체적인 변형을 분석하거나, 얼굴의 주요 특징 부분으로부터의 변화를 추출하는 방법으로 정확한 추출을 위하여 얼굴과 배경을 분리하는 처리가 필요하며, 빛의 변화에 민감한 특성이 있으므로 이러한 빛의 영향을 최소화하는 처리과정이 필요하다. 모델기반 얼굴 특징 추출 방법은 얼굴의 형태 또는 텍

스처 정보를 이용하여 얼굴의 변형을 추출하는 방법이다. 대표적으로 Active Shape Model(ASM)은 추출하고자 하는 객체에 대하여 정점으로 구성된 점 분포모델(Point Distribution Model)을 이용하여 얼굴변화를 분석하며, Active Appearance Models(AAM)은 얼굴의 형태와 그에 대응되는 텍스처 정보를 이용하여 얼굴 영상을 해석하는 방법이다[9].

반면 모션기반 방법으로는 밀집광류(Dense Optical Flow), 특징점 추적(Feature Point Tracking), 차영상(Difference Image) 방법, 그리고 마커추적(Marker Tracking) 등이 있다. 밀집광류는 얼굴전체 또는 얼굴 특징 부분의 화소값의 변화를 추출하여 얼굴의 움직임을 분석하는 방법으로 잡음이나 조명변화에 영향을 많이 받으며 많은 미세한 움직임에 민감한 경향이 있다. 특징점 추적은 눈, 코, 입 등 주요 얼굴 특징들의 움직임을 추출하는 방법으로 주요 특징들 외의 다른 특징들은 고려하지 않는다. 차영상 방법은 기본 얼굴과 각 표정 얼굴 사이의 차 영상으로부터 움직임을 획득하며 정확한 추출을 위하여 얼굴 영상의 정규화 과정이 필요하다. 마커추적은 얼굴에 마커를 부착하여 얼굴 표정의 움직임에 따라 발생하는 마커의 위치 변화를 추적하여 움직임 데이터를 획득하는 방법이다.

얼굴 표정 인식 방법은 크게 공간적 분석법(Spatial Method)과 시공간적 분석법(Spatio-Temporal)으로 분류할 수 있다. 공간적 분석법은 영상의 분석을 통해서 표정을 분류하는 방법으로 얼굴 영상에 대하여 주성분 분석, 독립요소 분석, 가버 웨이블릿 필터와 같은 방법을 이용하여 영상을 분류한다. 공간적 분석법은 영상의 분석 시 처리 속도가 빠르다는 장점이 있으나, 얼굴의 변화가 클 경우 매칭이나 정합에 대한 인식이 어렵다는 단점이 있다.

시공간적 분석법은 시간의 변화에 따른 얼굴 표정을 해석하는 방법으로 신경망, 은닉마코프 모델 등의 방법이 있다. 은닉마코프 모델은 얼굴 행

동 패턴을 분석하여 표정을 인식하는데 유용하게 사용된다[10,11]. 시공간적 모션 에너지 템플릿 방법은 얼굴의 움직임을 2차원의 움직임 공간으로 표현하여 두 템플릿 사이의 유클리디언 거리를 계산하여 얼굴 표정을 추출한다[12]. 시공간적 분석법은 얼굴 표정 변화를 연속적으로 처리할 수 있으나 학습이나 데이터베이스 구축에 드는 비용이 높다는 단점이 있다.

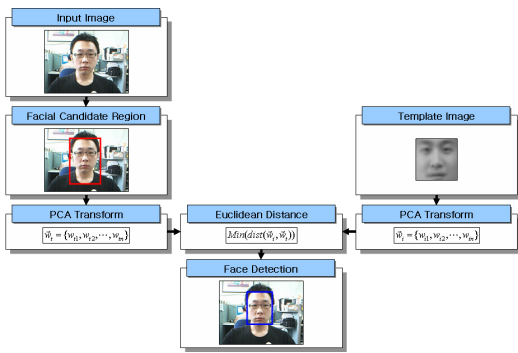
3. 제안시스템 구조

본 논문에서는 실시간으로 동영상으로부터 얼굴의 특징점을 추출하고 광류 알고리즘을 이용하여 최소한의 얼굴 특징점을 추적하고 이들의 움직임 벡터와 확장된 표정 은닉 마르코프 모델을 이용하여 얼굴 표정을 인식하는 방법을 제안하였다. 제안된 얼굴인식 시스템은 크게 얼굴 영역 검출, 얼굴 특징점 추적, 얼굴 표정 인식으로 구성된다. 카메라로부터 입력된 영상에 대하여 전처리 과정을 거친 후 스킨컬러모델을 이용하여 얼굴 후보 영역을 획득한다. 검출된 후보 영역에서 주성분분석과 템플릿 매칭을 통해 정확한 얼굴 영역을 검출한다. 얼굴 특징점 추적 단계에서는 전 단계에서 검출된 얼굴 영역에 대하여 광류 알고리즘을 이용하여 얼굴의 눈썹, 눈, 입의 특징점들의 움직임 벡터를 획득한다. 이와 같이 검출된 얼굴 특징점의 움직임 벡터와 은닉 마르코프 모델을 이용하여 얼굴 표정의 인식을 수행한다.

3.1 얼굴영역 검출

얼굴영역 검출은 전처리 과정, 얼굴 후보 영역 검출, 얼굴 영역 검출의 세 단계로 구성된다. 입력된 영상에 대하여 최대-최소 정규화 방법과 참조 화이트를 적용하여 외부조명변화로부터의 영향을 개선하고 정확한 얼굴 영역을 찾을 수 있도록 처리한다. 정규화 된 영상에 대하여 HT(Hue-Tint) 컬러모델[13]을 적용하여 얼굴 후보영역을 검출하고 미디언 필터와 모폴로지 연산을 통해 잡음을 제

거한다. 최종적인 얼굴 영역을 검출하기 위하여 여러 장의 훈련 데이터로부터 80×80 의 얼굴 템플릿 영상을 생성하고, 얼굴 후보 영역과 템플릿 영상에 대하여 주성분분석을 이용하여 저차원의 가중치벡터로 변환하고 이들을 비교하여 얼굴 영역을 검출한다. 이처럼 검출해야할 영역을 단계적으로 축소하여 처리량을 감소시키므로 빠른 얼굴 검출이 가능하다. 그림 1은 얼굴영역 검출과정을 보여주고 있다.



(그림 1) 얼굴영역 검출 과정

정확한 얼굴 영역을 검출하기 위하여 입력 영상에 대하여 최대-최소 정규화 방법과 참조 화이트를 통한 전처리 과정을 거쳐 영상을 개선시킨 후 컬러모델을 이용하여 얼굴 후보 영역을 검출한 후 미디언 필터와 모폴로지 연산을 적용하여 잡음을 제거한다. 최대-최소 정규화 방법은 주변 조명의 영향을 최소화하기 위한 방법으로 영상에서의 밝기값의 범위를 새롭게 정의된 범위로 변경시켜준다.

전처리된 입력영상으로부터의 피부영역 검출은 비선형 스킨컬러 모델인 HT 스킨 컬러모델을 이용하여 검출한다. 검출된 얼굴 후보영역에 대하여 주성분분석과 템플릿 매칭을 통해 정확한 얼굴 영역을 검출한다. 얼굴 후보 영역에 대하여 주성분분석을 적용하여 특징 벡터로 변환하여 템플릿의 특징 벡터와 유사도를 측정하여 얼굴 영역을 검출한다.

주성분분석은 고유값을 내림차순으로 정렬하고 저 순위의 주성분들을 제거하여 고차원의 데이터를 저차원의 데이터로 축소하는 방법으로, 주어진 얼굴 영상의 훈련 데이터와 각 영상의 차 벡터로 구성된 공분산 행렬의 고유벡터를 고유공간에 투영시킨 영상이 고유 얼굴이며, 이 고유값에 해당하는 고유벡터일수록 얼굴과의 유사성이 높다. $N \times N$ 의 2차원 얼굴 영상에 대하여 $N^2 \times 1$ 의 얼굴 벡터로 변환한 M 개의 얼굴 벡터 ($\Gamma_1, \Gamma_2, \dots, \Gamma_M$)에 대한 평균 영상 벡터는 다음과 같다.

$$\Psi = \frac{1}{M} \sum_{k=1}^M \Gamma_k \quad (1)$$

여기서 Ψ 는 $N^2 \times 1$ 벡터인 평균 영상이고 Γ_k 는 훈련 영상이며, M 은 훈련 영상의 개수이다. 각 훈련 영상 벡터에 대한 평균 영상 벡터와의 차 벡터를 계산하고 이에 대한 공분산 행렬은 (식 2)에 의하여 구할 수 있다.

$$C = \frac{1}{M} \sum_{i=1}^M \phi_i \phi_i^T = AA^T \quad (2)$$

위의 식에서 M 개의 고유값 λ_i 와 그에 대응되는 고유벡터 μ_i 는 다음과 같이 정의된다.

$$C\mu_i = \lambda_i\mu_i \quad (3)$$

이때 (식 1)과 (식 2)에 의해 다음의 (식 4)를 유도한다.

$$A^T A \mu_i = \lambda_i \mu_i \quad (4)$$

(식 4)로부터 (식 5)를 최종적으로 유도 할 수 있다.

$$AA^T A\mu_i = \lambda_i A\mu_i \quad (5)$$

이때 AA^T 의 고유벡터는 $A\mu_i$ 가 되고 고유값은 λ_i 이다. 그러므로 고유벡터 μ_k 는 다음과 같이 정의할 수 있다.

$$\mu_k = A\mu_i \quad (6)$$

얼굴 영상 Γ_i 를 얼굴 공간상에 투영했을 때, 전체 고유 얼굴 μ_i^T 에 대한 얼굴 영상의 특징 벡터는 다음과 같다.

$$\omega_{ik} = \mu_k^T(\Gamma_i - \Psi), k = 1, 2, \dots, M \quad (7)$$

전체 영상의 특징 벡터 $\Omega_i^T = [\omega_{i1}, \omega_{i2}, \dots, \omega_{iM}]$ 는 훈련 영상에 대한 고유 얼굴 공간상의 위치를 나타낸다. 가중치 벡터 ω_{ik} 를 이용하여 템플릿 매칭을 수행한다. 템플릿 영상은 다양한 얼굴 훈련 데이터들의 평균 영상을 이용하여 획득하며 템플릿의 크기는 80×80 로 정규화 하였다.



(그림 2) 얼굴 훈련 데이터의 템플릿 영상 생성

획득된 템플릿 영상에 대하여 주성분분석을 적용하여 가중치 벡터를 획득하며, 검출된 얼굴 후보 영역 내에서 템플릿 크기와 동일한 영역에 대하여 주성분분석을 이용하여 생성된 가중치 벡터와 템플릿에 대한 가중치 벡터간의 유클리디언 거리를 이용하여 유사도를 측정하여 최종 얼굴 영역을 검출한다.

$$MIN(Distance(\vec{w}_I, \vec{w}_T))$$

$$Distance(\vec{w}_I, \vec{w}_T) = \sqrt{\sum_{n=0}^{79} (w_{I_n} - w_{T_n})^2} \quad (8)$$

3.2 얼굴 특징점 검출

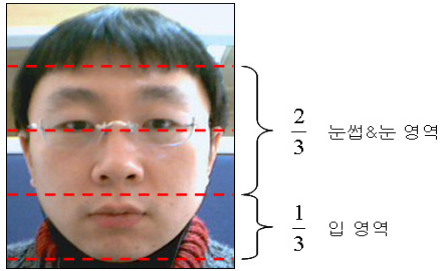
본 논문에서는 얼굴의 기하학적 정보를 이용하여 눈썹, 눈, 입의 특징점을 획득하고, 광류 알고리즘을 이용하여 특징점의 움직임을 추적한다. 얼굴의 특징점을 추적하는 과정은 얼굴 특징점 영역의 추출과 얼굴 특징점 추적의 두 단계로 구성된다. 얼굴의 특징점은 MPEG 4에 정의된 얼굴의 주요 특징점 가운데 얼굴표정 인식에 필요한 18개의 특징점을 추출하여 사용하며 그 내용은 표 1과 같다.

(표 1) 얼굴 특징점

구분	개수	위치(x, y)
눈썹	3개 ×2	$(eyebrow\ area_L, eyebrow\ area_H/2)$
		$((eyebrow\ area_L + eyebrow\ area_R)/2, (eyebrow\ area_H)/2)$
		$(eyebrow\ area_R, eyebrow\ area_H/2)$
눈	4개 ×2	$(eye\ area_L, eye\ area_H/2)$
		$((eye\ area_L + eye\ area_R)/2, eye\ area_T/2)$
		$(eye\ area_R, eye\ area_H/2)$
		$((eye\ area_L + eye\ area_R)/2, eye\ area_B/2)$
입	4개	$(mouth\ area_L, mouth\ area_H/2)$
		$((mouth\ area_L + mouth\ area_R)/2, mouth\ area_T/2)$
		$(mouth\ area_R, mouth\ area_H/2)$
		$((mouth\ area_L + mouth\ area_R)/2, mouth\ area_B/2)$
총	18개	

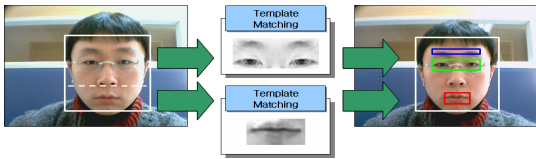
그림 3은 정확한 얼굴의 특징점을 검출하기 위하여 눈썹, 눈, 입의 얼굴에 대하여 어떠한 비율로 분포하는지를 나타낸다. 얼굴 영역을 3등분했을 때 눈썹과 눈은 2/3 영역에 분포하고 입은 나머지 1/3 영역에 분포하게 된다. 이와 같이 주어진 특징점의 위치 정보를 이용하여 각 특징점을 검출하

기 위한 후보 영역을 설정한다. 각 후보 영역에 대하여 눈썹과 눈이 포함된 템플릿과 입이 포함된 템플릿을 이용하여 템플릿 매칭을 수행한다.



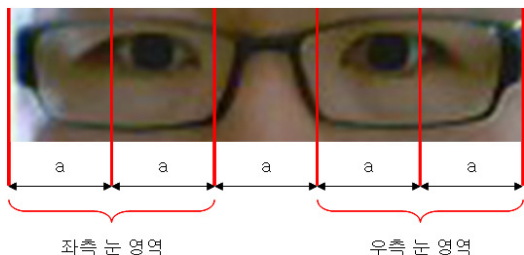
(그림 3) 얼굴 영역의 기하학적 정보에 의한 특징점 분포영역

검출된 후보 영역에 대하여 템플릿 매칭을 수행한 결과는 그림 4와 같다.



(그림 4) 특징점 후보 영역으로부터 눈썹과 눈, 입 영역 검출 결과

이때, 검출된 눈썹과 눈 영역에 대해서 좌·우측에 존재하는 눈썹과 눈 영역으로 나누어야 한다. 검출된 눈 후보 영역에 대하여 최종 눈 영역을 획득하기 위하여 그림 5와 같이 눈 후보 영역을 나눈다.

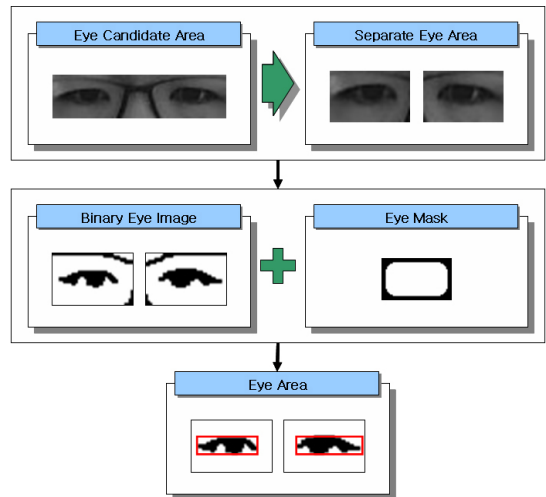


(그림 5) 눈 후보 영역으로부터 좌·우측 눈 영역 획득

눈동자 영역에 포함되는 픽셀의 밝기값이 가장 낮기 때문에 (식 9)를 이용하여 가장 어두운 화소값을 임계치로 하여 이진 영상으로 변환한다.

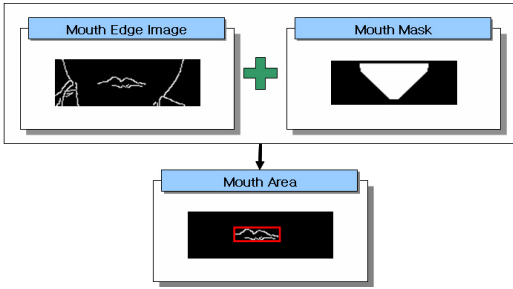
$$\begin{aligned} \text{basis pixel} &= \text{MIN}(\text{pixel in eye area}) \\ \text{eye pixel} &\leq \text{basis pixel} + \sigma \end{aligned} \quad (9)$$

또한 사용자가 안경을 사용할 경우 정확한 좌·우측 눈 영역의 검출이 어렵기 때문에 그림 6과 같이 마스크 영상을 이용하여 각 검출된 후보 영역에 대하여 AND 연산을 적용해 안경 정보를 제거할 수 있다. 안경 정보가 제거된 눈 영역에 대하여 히스토그램 분석을 통해 최종 눈 영역을 획득한다.



(그림 6) 눈 영역 획득 과정

정확한 입 영역의 획득은 캐니에지 알고리즘을 이용하여 검출된 에지영상의 히스토그램 분석을 수행하며, 입의 후보 영역에 대하여 에지를 검출하게 되면 배경 및 턱 선과 같은 에지 정보도 함께 획득된다. 눈 영역의 검출 과정과 마찬가지로 마스크 영상을 이용하여 입 영역을 제외한 나머지 부분의 에지 정보를 제거하여 입 영역을 획득한다.



(그림 7) 입 영역 획득 과정

3.3 광류를 이용한 얼굴 특징점 추적

다양한 얼굴 표정은 검출된 특징점의 변화에 의해 결정되며 이러한 특징점은 동영상에서 동적으로 변화하게 된다. 따라서 실시간 표정의 인식을 위해서 특징점의 위치를 효과적으로 추적하는 방법이 필요하며, 본 연구에서는 광류[14]에 의한 추적 방법을 적용하였다. 본 논문에서 적용된 광류 알고리즘은 연속된 두 영상에서의 화소의 이동 변위를 계산하고 나타내는 방법으로서 일반적으로 물체의 이동 또는 운동에 관한 사항을 측정하거나 인식하는데 적용될 수 있다. 본 논문에서는 Lucas-Kanade가 제안한 LK 광류 알고리즘을 적용하였다. LK 알고리즘은 영상의 특징이 되는 국부영역($n \times n$ 특징 윈도우) 안의 움직임 속도는 동일하다는 가정 하에, 영상의 밝기 값의 차이를 이용하여 움직임 정보인 모션 필드를 측정한다 [15]. 즉, LK 알고리즘은 특징 윈도우 영역인 국부영역에 대한 모션 필드를 계산하는 방법으로 다른 광류 알고리즘에 비하여 정확하고 잡음에 강한 장점을 가지고 있으며, 기본적인 특성은 다음과 같다.

영상에서의 화소를 $\mathbf{u}=(u,v)$ 라고 하고 시간 t 에서의 영상은 $I(\mathbf{u},t)$ 로 표현할 수 있으며, 시간 $t+1$ 에서 픽셀 \mathbf{u} 는 모션 벡터에 의해 $\mathbf{u}'=F(\mathbf{u},\mu)$ 으로 이동하게 된다. 여기서 μ 는 모션 벡터, $F(\mathbf{u},\mu)$ 는 모션 모델이라 한다. 만일 주변 조명이 일정하다고 가정한다면 영상에서의 화소 좌표는 이동하지만($\mathbf{u} \rightarrow \mathbf{u}'$), 화소의 밝기 값은 변하지

않기 때문에 (식 10)과 같이 나타낼 수 있다.

$$I(F(\mathbf{u},\mu),t+1)=I(\mathbf{u},t) \quad (10)$$

이 때, LK 알고리즘은 모션 벡터 μ 를 계산하기 위해 (식 11)과 같이 밝기 값 변화량의 제곱 값을 최소화함으로써 얻을 수 있다.

$$\min E(\mu)=\sum_{\mathbf{u} \in \Omega}(I(F(\mathbf{u},\mu),t+1)-I(\mathbf{u},t))^2 \quad (11)$$

위 식에서 Ω 는 측정하고자 하는 영상의 영역이며, 모션 벡터를 얻기 위해서는 영역 Ω 안에 존재하는 화소를 이용하여 계산한다. (식 12)는 일차 테일러 확장을 통해 식 11을 전개한 결과이다.

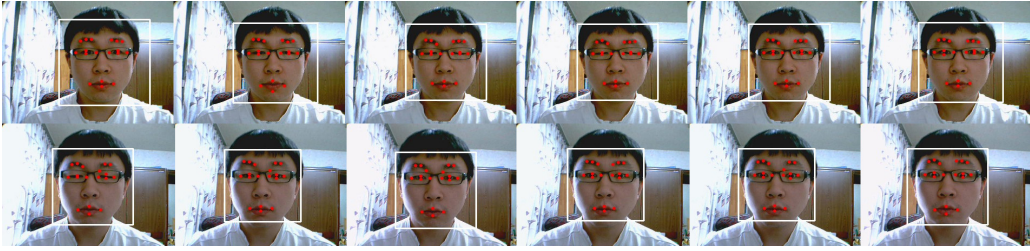
$$\mu=-\left(\sum_{\Omega}(\mathbf{I}_u F_\mu)^T(\mathbf{I}_u F_\mu)\right)^{-1}\sum_{\Omega}I_u(\mathbf{I}_u F_\mu)^T \quad (12)$$

이 때, \mathbf{I}_u 와 \mathbf{I}_t 는 각각 공간적 영상 변화량과 시간적 영상 변화량을 의미하며, F_μ 는 함수 F 에서 μ 에 대해 편미분한 것이다. 그림 8은 광류를 이용하여 얼굴의 특징점을 추적한 결과이다.

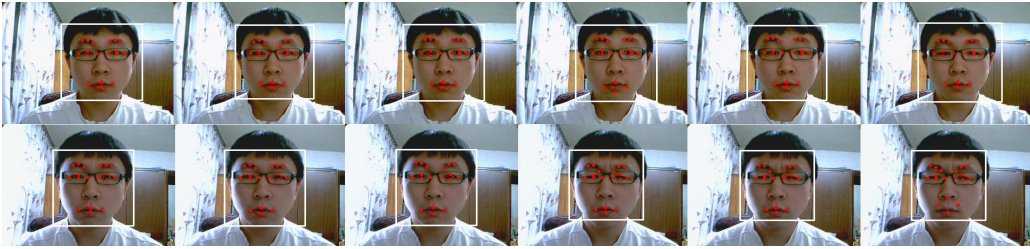
3.4 확장된 HMM을 이용한 얼굴인식 프레임 워크

얼굴 표정 인식 환경은 비이산적이고 시간의 연속성을 고려해야 하며, 다음 상태의 예측이 불가능하다는 특징을 가지고 있다. 또한, 얼굴 표정의 행동 패턴을 추출하는 것이 목적이므로, HMM을 이용한 얼굴 표정인식이 적절하다[16,17]. 기존의 얼굴 표정 인식을 위한 HMM 모델은 기본 얼굴로부터 각 최종 얼굴 표정 사이의 인식만을 수행하였으나, 본 논문에서는 각 최종 얼굴 표정 사이의 표정 패턴도 인식할 수 있도록 새로운 모델을 제안하였다.

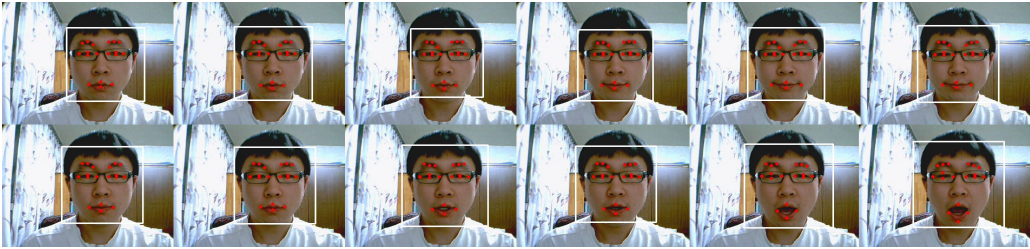
HMM은 각 상태간의 전이 확률을 가지는 유한 상태 기계로, 각 상태는 직접 관측이 불가능하며 각 상태들로부터 일정 확률로 관측할 수 있는 관측 심볼을 통하여 상태를 추정하는 방법이다 [18][19]. 본 논문에서 사용된 HMM은 행복, 화남,



(a) 특징점 추적 결과(눈)



(b) 특징점 추적 결과(눈썹)



(c) 특징점 추적 결과(입)

(그림 8) 얼굴의 특징점 추적 결과

놀람, 슬픔 그리고 기분(중립)상태 등 5개의 기본 감정 상태를 포함하는 모델로 구성되어 있다. 일련의 입력영상으로부터의 획득할 수 있는 특정 얼굴 표정은 얼굴 움직임의 시간적 변화에 의해 표현되기 때문에 각 표정의 특별한 형태에 대하여 훈련된 HMM 모델에 의하여 모델링 된다. 이러한 HMM 모델은 다음과 같이 정의된다.

$$\begin{aligned} \lambda &= (A, B, \pi) \\ a_{ij} &= P(q_{t+1} = S_j | q_t = S_i), \quad i \leq j \leq N \\ B &= b_j(O_t) = P(O_t | q_t = S_j), \quad i \leq j \leq N \\ \pi_j &= P(q_1 = S_j) \end{aligned} \quad (13)$$

이 때, A 는 상태 변환확률행렬, B 는 관측확률 분포, π 는 초기상태분포, 그리고 N 은 HMM에서

의 상태 수를 나타낸다. 관측치 O_t 는 얼굴표정 특성의 연속적인 모션을 나타내는 값으로 정의된다. 따라서 B 는 특정 시간 t 에서 관측벡터의 확률밀도함수로 표현되며, 다음과 같은 가우지언분포로 정의된다.

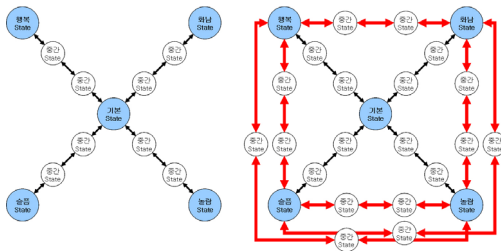
$$B = b_j(O_t) \sim (\mu_j, \Sigma_j), \quad i \leq j \leq N \quad (14)$$

이 때, μ_j 와 Σ_j 는 각각 평균 벡터와 공분산 행렬을 의미한다. HMM의 파라미터 λ 는 모 Baum-Welch의 방정식에 의해서 유도되며[13], 일련의 관측치 O_t 에 대하여 각 표정모델의 관측 확률 $P(O_t | \lambda_j)$ 는 forward-backward 방법[8]에 의

하여 계산된다. 결국 일련의 얼굴표정은 (식 15)와 같이 최대 확률을 나타내는 모델에 대응되는 감성상태로 분류된다.

$$c^* = \operatorname{argmax}[P(O|\lambda_c)], 1 \leq c \leq 5 \quad (15)$$

본 논문에서는 얼굴 표정 인식을 수행하기 위하여 새로이 확장된 은닉 마르코프 모델을 제안한다. 기존의 HMM을 이용한 얼굴 표정 인식에서는 각 얼굴표정의 변환에 있어서 중립얼굴을 통과하는 것을 전제로 하여, 중립 얼굴 표정과 각 최종 얼굴 표정 사이의 변화만을 정의하였으나, 본 논문에서는 기존의 모델에 각 최종 표정 사이에서 표정 변화를 추정하여 모델링하였다. 그림 9는 얼굴표정 인식을 위한 기존의 전형적인 모델과 제안된 모델을 각각 보여주고 있다.



(그림 9) 얼굴 표정 인식을 위한 기존의 HMM 모델(좌)과 제안된 HMM 모델(우)

기본 얼굴 표정과 각 최대 얼굴 표정 사이의 특징점들의 변위를 계산하여 훈련 데이터로 이용하였으며 웹 카메라로부터 입력된 영상으로부터 획득된 특징점들의 변위를 이용하여 표정 인식을 수행한다. 다음은 표정 인식을 위한 훈련데이터이다.

기존의 얼굴 표정 인식을 위한 훈련 데이터 생성 방법은 각 표정의 변화되는 동영상 데이터로부터 일정 프레임 간격으로 연속적인 데이터를 획득하였으나, 본 논문에서는 얼굴 표정 사이의 변화 패턴을 추출하기 위해 각 표정 사이의 보간법을 이용해 획득하여 동영상을 이용하지 않고 정지 영상만으로도 중간 데이터를 생성할 수 있다. 다음 그림 11은 보간에 의해 생성된 대표 표정 사이의 중간 데이터 이다.

4. 실험 결과

본 논문에서 제안한 얼굴 표정 인식 시스템의 실험 환경은 표 2와 같다. 실시간 영상을 입력받기 위해서 AlphaCam M2 웹 카메라를 사용하였으며, Windows XP 환경에서 실험하였다. 시스템은 Visual C++로 구현하였으며 윈도우 인터페이스를 위하여 MFC 라이브러리를 사용하였으며, 웹 카메라와의 연동을 위하여 OpenCV 라이브러리를 사용하였다.

(표 2) 실험 환경

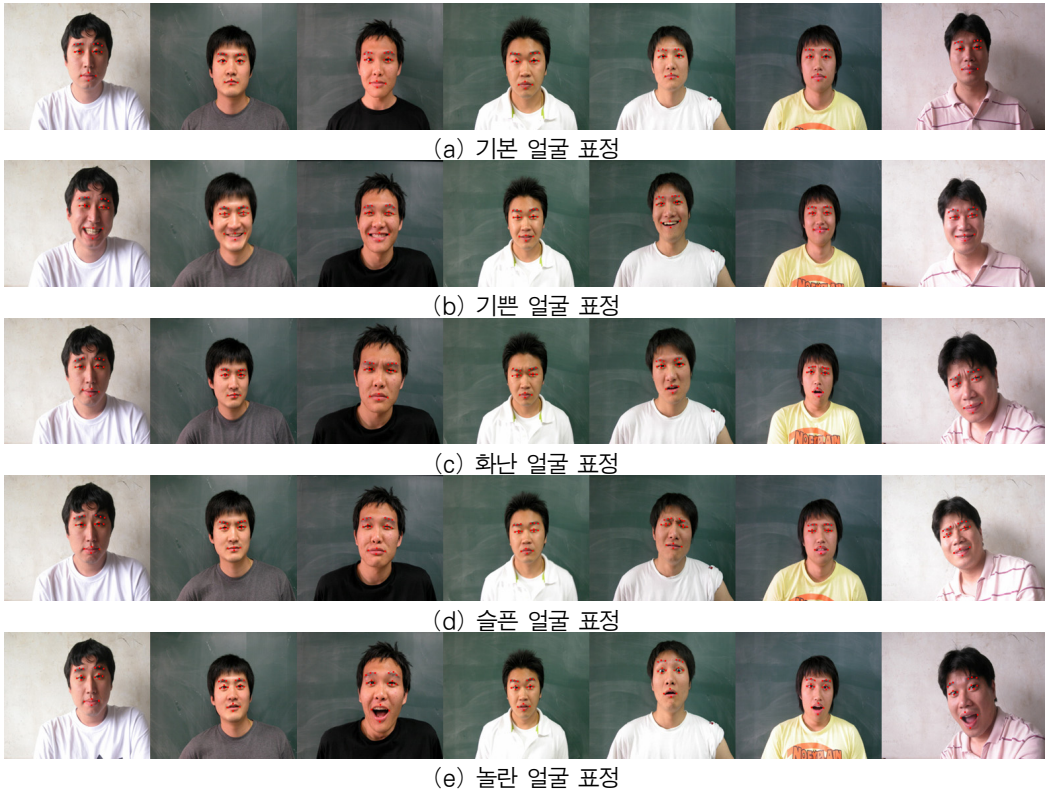
구분	세부 환경
CPU	Intel(R) Core(TM)2 CPU T5500 1.66GHz
RAM	2GB
Web Camera	AlphaCam M2
OS	Windows XP Service Pack 2
Language	Visual C++
Develop Tool	Microsoft Visual Studio 2005
Library	MFC, OpenCV

일반적으로 입력된 얼굴 영상에 안경이 존재할 경우 정확한 눈의 특징점을 검출하기가 어렵다. 본 논문에서는 검출된 눈 영역에 대하여 마스크 연산을 통해 안경을 제거하여 눈의 특징점을 추적한다. 그림 12에서 볼 수 있듯이 제안한 방법을 이용한 특징점 추적 결과가 더욱 정확함을 볼 수 있다.

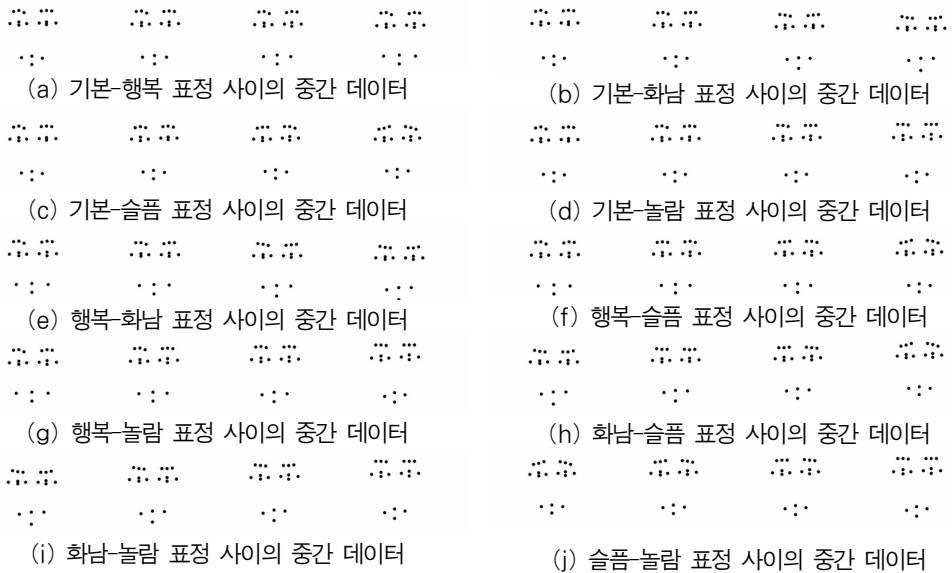
기존의 HMM을 이용한 얼굴 표정 인식은 중립 얼굴 표정과 각 최대 얼굴 표정 사이의 변화 패턴만을 인식하였다. 본 논문에서는 각 최대 표정 사이에서의 얼굴 표정 변화를 인식할 수 있도록 모델을 설계하여 적용하였으며, 입력 영상에서의 얼굴 표정과 각 최대 얼굴 표정간의 유클리디언 거리를 계산하여 표정의 유사도를 측정하였다. 두 개의 최대 표정을 나타내는 18개의 각 특징점에 대한 거리벡터 d 를 모션 패턴으로 정의하며, 두 개의 서로다른 표정 상태에서 서로 대응되는 2차원 좌표점의 위치를 각각 $e_i(x_k, y_k), e_j(x_k, y_k)$ 라고 할 때 d 는 다음과 같이 정의된다.

$$d = \{e_i(x_k) - e_j(x_k), e_i(y_k) - e_j(y_k)\}, \quad (16)$$

for all $i \neq j, k = 1 \dots 18$



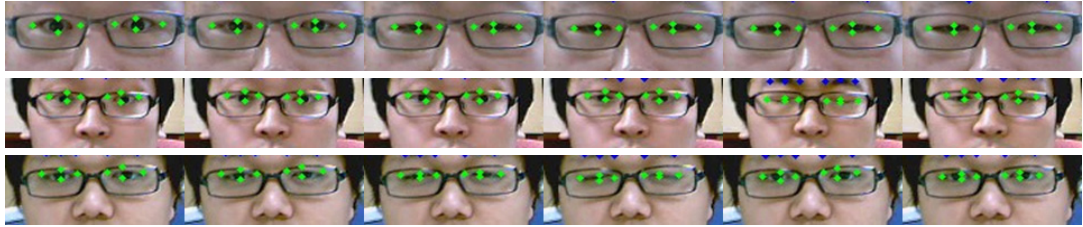
(그림 10) 표정 인식을 위한 훈련 데이터



(그림 11) 보간법을 이용한 각 표정 사이의 중간 데이터 생성 결과

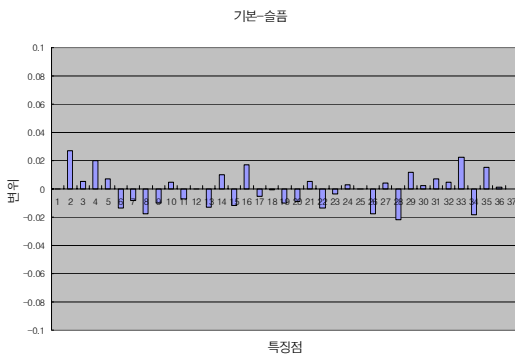


(a) 기존 방법에 의한 특징점 추출 및 추적

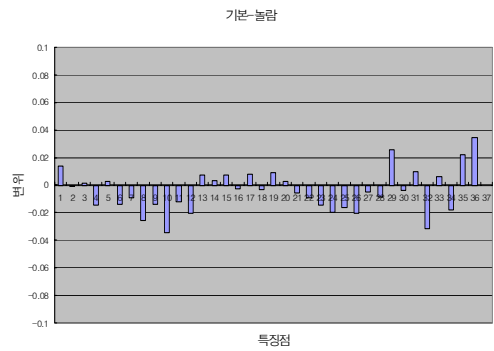


(b) 제안된 방법에 의한 특징점 추출 및 추적

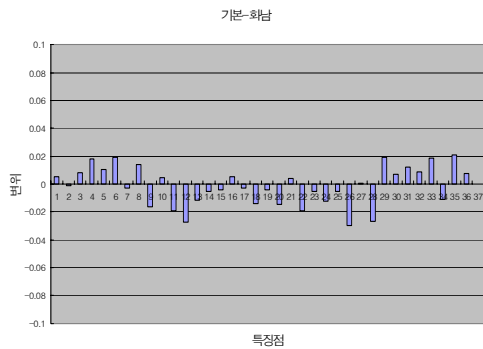
(그림 12) 얼굴 특징점 추적 결과 비교



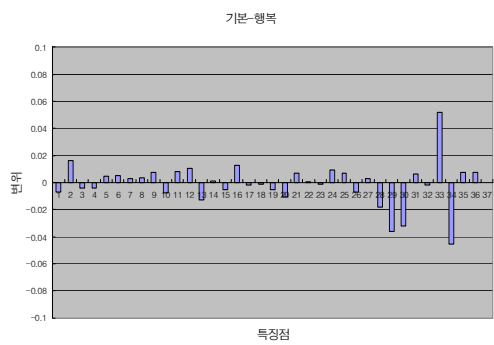
(a) 기본-슬픔 표정사이 변화패턴



(c) 기본-놀람 표정사이 변화패턴

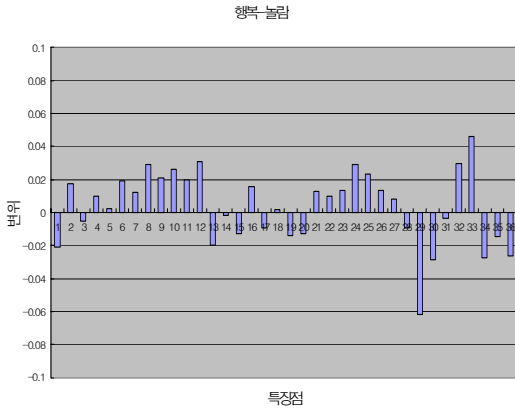


(b) 기본-화남 표정사이 변화패턴

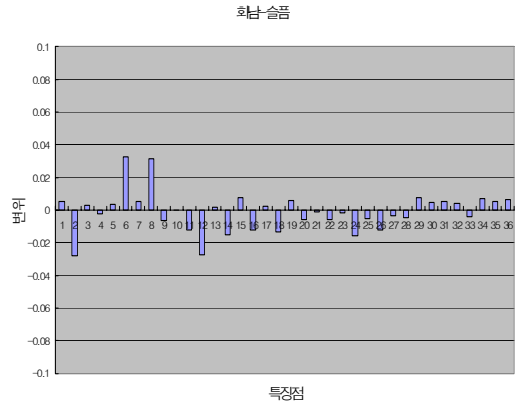


(d) 기본-행복 표정사이 변화패턴

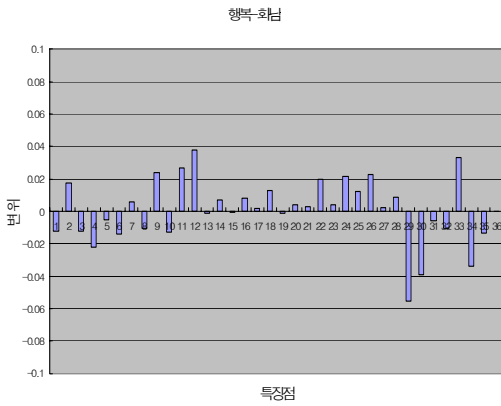
(그림 13) 기본 표정과 각 최대 표정 사이의 변화 패턴



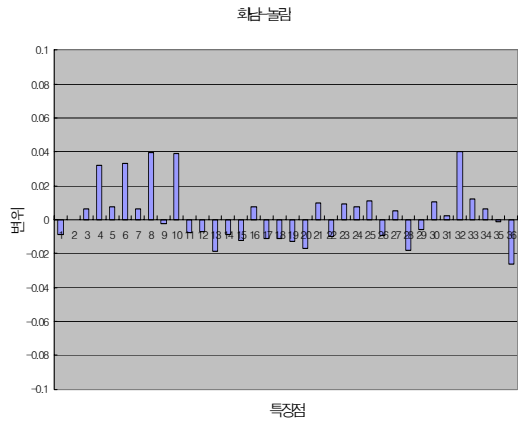
(a) 행복-놀람 표정사이 변화패턴



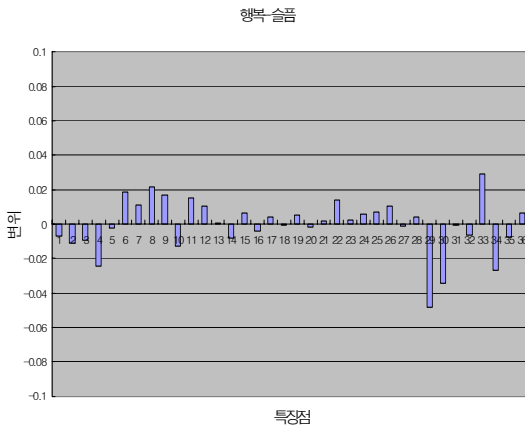
(d) 화남-슬픔 표정사이 변화 패턴



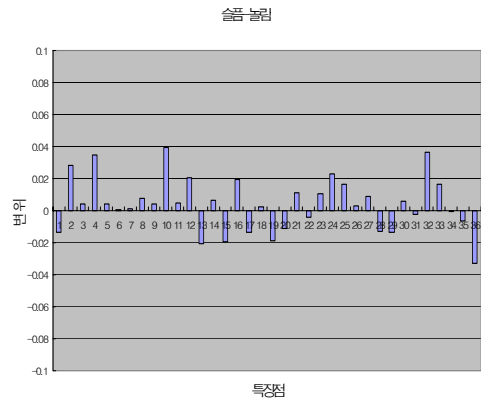
(b) 행복-화남 표정사이 변화 패턴



(e) 화남-놀람 표정사이 변화패턴



(c) 행복-슬픔 표정사이 변화패턴



(f) 슬픔-놀람 표정사이 변화 패턴

(그림 14) 각 최대 표정 사이의 변화 패턴

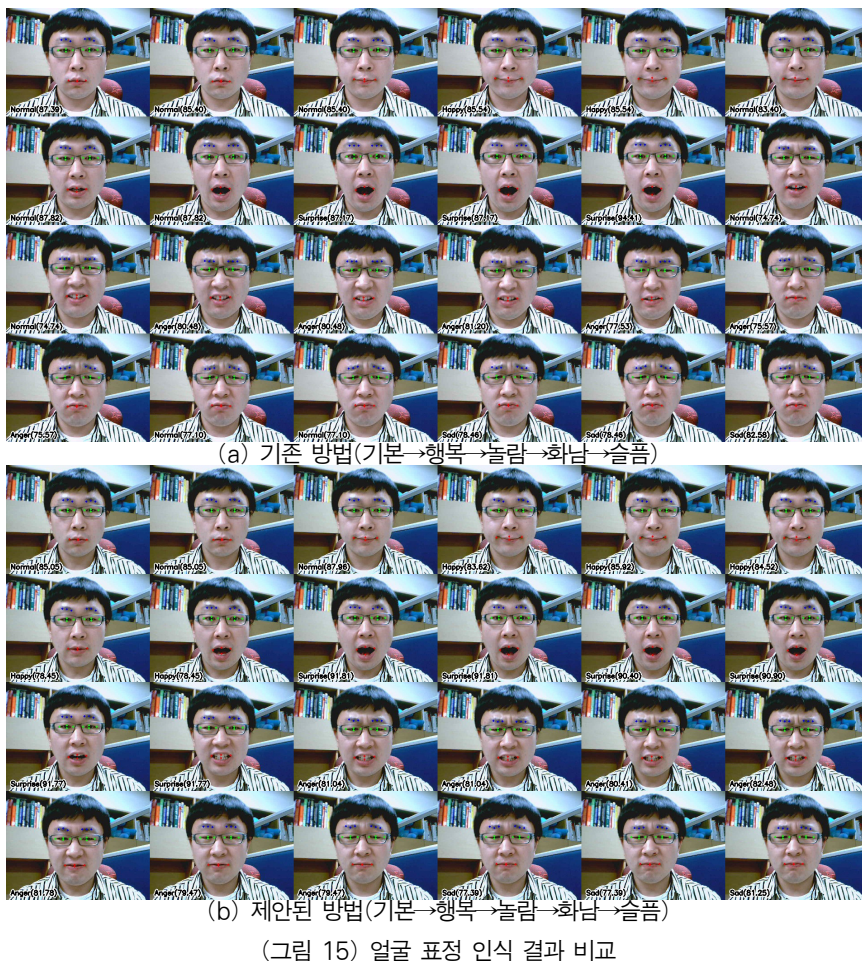
다음 그림 13은 기본표정과 대표 표정간의 변화 패턴을 그리고 그림 14는 각 대표표정간의 변화패턴을 18개의 특징점에 대하여 x 좌표값과 y 좌표값의 변화를 교대로 표시한 결과이다.

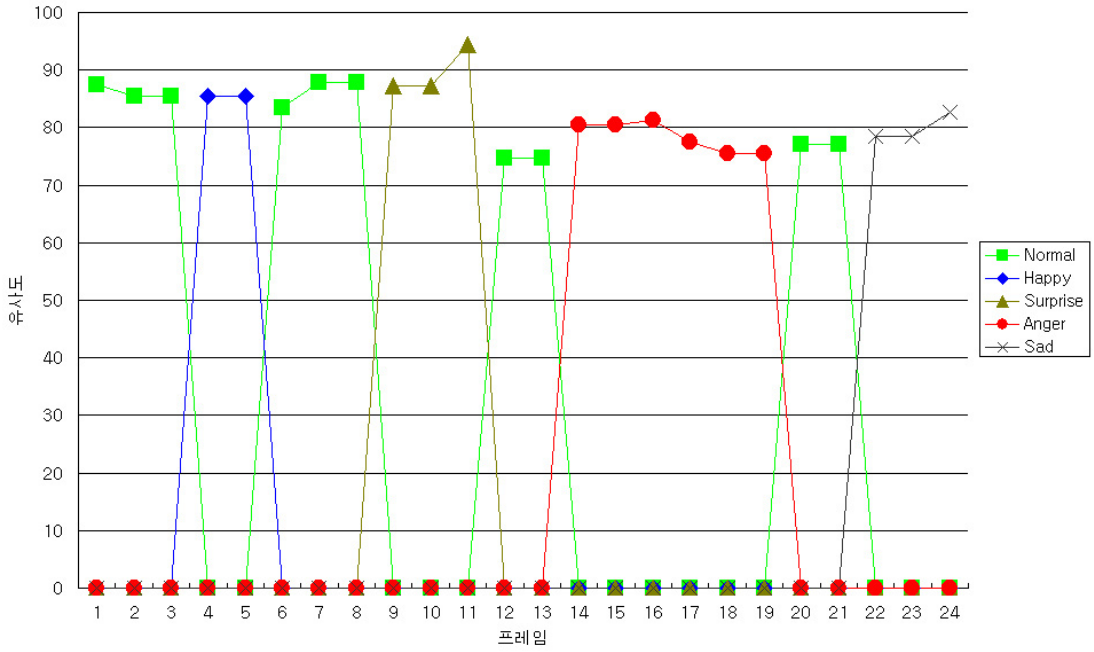
그림 15는 기존모델과 제안된 모델을 적용하여 동영상으로부터 입력되는 각 얼굴표정의 상태를 실시간으로 인식한 결과이며 그림 16에서 제시된 바와 같이 기존의 표정 인식 방법이 최대 표정사이에서 항상 기본(중립)표정을 지나야한다는 가정에서 실행되는데 비하여 본 논문에서 제안한 얼굴 표정 인식 방법이 중간단계의 표정을 반드시 거치는 과정 없이도 다양한 표정사이의 변화를 용이하게 인식할 수 있음을 알 수 있으며, 이는

동영상에서 일련의 표정변화가 사람의 감성 상태를 포괄적으로 포함한다는 측면에서 더 감성인식에 바람직하다고 판단된다.

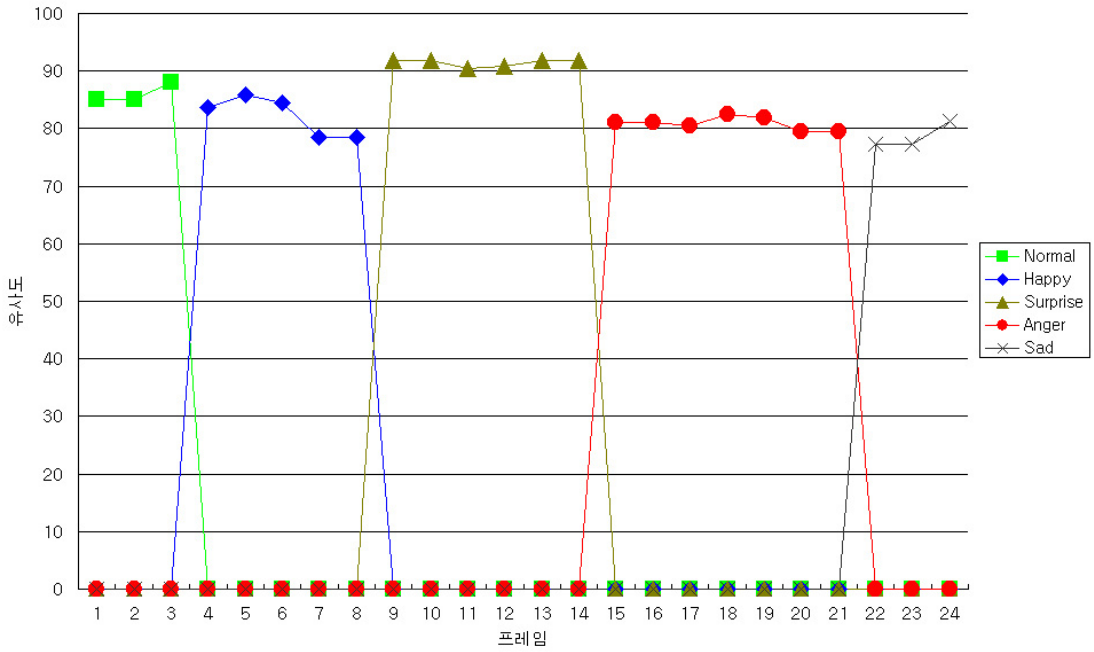
5. 결론

본 논문에서는 광류에 의한 얼굴특징 추적과 확장된 HMM 기반의 얼굴 표정 인식 시스템에 대하여 제안하였다. 카메라로부터 입력된 영상에 대하여 최대-최소 정규화 방법과 참조 화이트를 이용하여 조명의 영향을 완화한 개선된 입력 영상에 대하여 H-T 컬러모델과 모폴로지를 이용하여 얼굴의 후보 영역을 검출하였다.





(a) 기존 방법(기본→행복→놀람→화남→슬픔)



(b) 제안된 방법(기본→행복→놀람→화남→슬픔)

(그림 16) 얼굴 표정 인식 결과 비교 그래프

보다 정확한 얼굴 영역을 검출하기 위하여 주 성분분석과 템플릿 매칭을 이용하여 얼굴 후보 영역과 템플릿 사이의 유클리디언 거리를 이용하여 얼굴 영역을 획득하였다. 검출된 얼굴 영역에 대하여 얼굴의 기하학적 정보를 이용하여 눈썹, 눈, 입의 특징점을 추출한다. 좌·우 눈썹에 대하여 각 3개의 특징점, 좌·우 눈에 대하여 4개의 특징점, 입 주변 4개의 특징점등 모두 18개의 주요 특징점에 대하여 LK 광류 알고리즘을 이용하여 움직임 실시간으로 추적하였다. 또한 훈련 데이터 생성 방법은 각 표정의 변화되는 동영상 데이터로부터 일정 프레임 간격으로 연속적인 데이터를 획득하여 훈련과정에 많은 시간을 소모하였으나, 본 논문에서는 얼굴 표정 사이의 변화 패턴을 추출하기 위해 각 표정 사이의 보간법을 이용해 획득하여 동영상을 이용하지 않고 정지 영상만으로도 중간 데이터를 생성하여 영상의 훈련단계에 필요한 비용을 최소화하였다. 이와 같이 추출된 얼굴 특징점들의 움직임 벡터정보들을 확장된 얼굴인식 HMM모델에 적용하여 얼굴 표정 인식을 수행하였다. 특히 기존의 얼굴 표정 모델은 기본 얼굴 표정과 각 최종 얼굴 표정 사이의 변화 패턴만을 인식하였으나, 본 연구에서는 각 최종 얼굴 표정 사이에서의 변화 패턴을 인식할 수 있도록 모델을 설계하고 적용하였으며 실험 결과 기존의 얼굴 표정 인식 모델에 비하여 다양한 표정 변화를 인식할 수 있음을 입증하였다.

향후 연구 방안으로는 다양한 얼굴포즈에 대한 얼굴 검출 방법 및 얼굴 특징점 추적 방법에 대한 연구가 필요하다. 또한 정확한 얼굴 표정 인식을 위하여 다양한 얼굴 표정 인식 방법들을 조합한 하이브리드 방법을 이용하여 인식률을 높이는 방안에 대한 연구가 필요하다.

참 고 문 헌

- [1] P. Ekman and W.V. Friesen: Facial Action Coding System (FACS), Consulting Psychologist Press, Inc., 1978.
- [2] B. Fasel, J. Luetttin, "Automatic Facial Expression Analysis", A Survey, Pattern Recognition, 36(1), pp. 259-275, 2003.
- [3] C.C Chien, Y.J Chang and Y.C Chen, "Facial Expression Analysis Under Various Head Poses" Proceedings of 3rd IEEE Pacific-Rim Conf. on Multimedia, pp. 16-18, 2002.
- [4] I. Cohen, N. Sebe, A. Garg, L. S. Chen, T. S. Huang, "Facial expression recognition from video sequences: temporal and static modeling", Computer Vision and Image Understanding, Vol. 91, pp. 160-187, 2003.
- [5] J.C Chun, O.R Kwon, P.Park: A Robust 3D Face Pose Estimation and Facial Expression Control for Vision-Based Animation, LNCS Vol. 4351, pp. 700-708, 2007.
- [6] J. Huang, V. Blanz, B. Heisele :Face Recognition with Support Vector Machines and 3D Head Models, International Workshop on Pattern Recognition with Support Vector Machines, Niagara Falls, Canada, pp. 334-341, 2002.
- [7] P. Michel, R. El Kaliouby: Real Time Facial Expression Recognition in Video using Support Vector Machines, ICMI, pp. 258-264, 2003.
- [8] K. Anderson and P. W. McOwan, "A Real-Time Automated System for The Recognition of Human Facial Expressions", IEEE Transactions on Systems, Man and Cybernetics, Part B vol. 36, no. 1, pp. 96-105, 2006.
- [9] M. Evgeny, P. Edmond C, "Tracking Facial Features with Occlusions", Journal of Zhejiang University SCIENCE A, vol. 7, No. 7, pp. 1282-1288, 2006.
- [10] Pardas and Bonafonte, M. Pardas, A. Bonafonte, "Facial Animation Parameters Extraction and Expression Recognition Using Hidden Markov Models", Signal Processing: Image

- Communication 17, pp. 675-688. 2002.
- [11] Y. Zhu, L. C. De Silva, C. C. Ko, "Using Moment Invariants and HMM in Facial Expression Recognition", Pattern Recognition Letters Archive Volume 23, 83-91, 2002.
- [12] Y. Zhu, L. C. De Silva, C. C. Ko, "A Solution for Facial Expression Representation, and Recognition", Signal Processing : Image Communication 17, 657~673, 2002.
- [13] K.P. Min, J.C Chun, G.R Park, "A Nonparametric Skin Color Model for Face Detection from Color Images", LNCS Vol. 3320, pp. 115-119, 2004.
- [14] M. Black, "Robust incremental optical flow," PhD thesis, Yale University, 1992.
- [15] J.L. Barron, and N.A. Thacker, "Tutorial: Computing 2D and 3D Optical Flow", Tina Memo No. 2004-012, 2005.
- [16] Philipp Michel, Rana El Kaliouby, "Real Time Facial Expression Recognition in Video using Support Vector Machines", ICMI 2003: 258-264, 2003.
- [17] Qiang Miao, Hong-Zhong Huang, Xianfeng Fan, "A Comparison Study of Support Vector Machines and Hidden Markov Models in Machinery Condition Monitoring", Journal of Mechanical Science and Technology Vol. 21, No.4, pp. 607-615, 2007.
- [18] Ara V. Nefian, Monson H, "Hidden Markov Models for Face Recognition", ICASSP98, pp. 2721-2724, 1998.
- [19] A.V. Nefian, M.H. Hayes, "Maximum Likelihood Training of The Embedded HMM for Face Detection and Recognition", Proc. of the IEEE International Conference on Image Processing, ICIP 2000, Vol. 1, pp. 10-13, 2000.

● 저 자 소 개 ●



전 준 철

1984년 중앙대학교 전자계산학과 졸업(학사)
 1986년 중앙대학교 대학원 전자계산학과 졸업(석사)
 1992년 The Univ. of Connecticut, 컴퓨터공학과 졸업(석사)
 1995년 The Univ. of Connecticut, 컴퓨터공학과 졸업(컴퓨터 그래픽스 전공 박사)
 1995년~현재 경기대학교 컴퓨터과학과 교수
 2001년~2002년 2월 Michigan State Univ. 컴퓨터공학과 패턴인식 연구실 방문 연구원
 2009년~현재 Univ. of Colorado, Boulder 컴퓨터공학과 방문 연구원
 관심분야 : 비전기반 HCI, 컴퓨터 애니메이션, 증강현실, 의료영상처리 등
 E-mail : jechun@kgu.ac.kr



신 기 한

2006년 경기대학교 전자계산학과 졸업(학사)
 2008년 경기대학교 대학원 전자계산학과 졸업(석사)
 2008년 3~현재 (주) 핸디소프트 연구원
 관심분야 : 비전기반 HCI, 컴퓨터그래픽스, 얼굴 표정 인식
 E-mail : dalgundal@naver.com