

FACS와 AAM을 이용한 Bayesian Network 기반 얼굴 표정 인식 시스템 개발

Development of Facial Expression Recognition System based on Bayesian Network using FACS and AAM

고광은 · 심귀보*

Kwang-Eun Ko and Kwee-Bo Sim*

중앙대학교 전자전기공학부

요 약

얼굴 표정은 사람의 감정을 전달하는 핵심 메커니즘으로 이를 적절하게 활용할 경우 Robotics의 HRI(Human Robot Interface)와 같은 Human Computer Interaction에서 큰 역할을 수행할 수 있다. 이는 HCI(Human Computing Interface)에서 사용자의 감정 상태에 대응되는 다양한 반응을 유도할 수 있으며, 이를 통해 사람의 감정을 통해 로봇과 같은 서비스 에이전트가 사용자에게 제공할 적절한 서비스를 추천할 수 있도록 하는 핵심요소가 된다. 본 논문에서는 얼굴표정에서의 감정표현을 인식하기 위한 방법으로 FACS(Facial Action Coding System)와 AAM(Active Appearance Model)을 이용한 특징 추출과 Bayesian Network 기반 표정 추론 기법이 융합된 얼굴표정 인식 시스템의 개발에 대한 내용을 제시한다.

Abstract

As a key mechanism of the human emotion interaction, Facial Expression is a powerful tools in HRI(Human Robot Interface) such as Human Computer Interface. By using a facial expression, we can bring out various reaction correspond to emotional state of user in HCI(Human Computer Interaction). Also it can infer that suitable services to supply user from service agents such as intelligent robot. In this article, We addresses the issue of expressive face modeling using an advanced active appearance model for facial emotion recognition. We consider the six universal emotional categories that are defined by Ekman. In human face, emotions are most widely represented with eyes and mouth expression. If we want to recognize the human's emotion from this facial image, we need to extract feature points such as Action Unit(AU) of Ekman. Active Appearance Model (AAM) is one of the commonly used methods for facial feature extraction and it can be applied to construct AU. Regarding the traditional AAM depends on the setting of the initial parameters of the model and this paper introduces a facial emotion recognizing method based on which is combined Advanced AAM with Bayesian Network. Firstly, we obtain the reconstructive parameters of the new gray-scale image by sample-based learning and use them to reconstruct the shape and texture of the new image and calculate the initial parameters of the AAM by the reconstructed facial model. Then reduce the distance error between the model and the target contour by adjusting the parameters of the model. Finally get the model which is matched with the facial feature outline after several iterations and use them to recognize the facial emotion by using Bayesian Network.

Key Words : Facial Expression Recognition, AAM, FACS, Bayesian Network

1. 서 론

HRI와 같은 HCI 분야에서 인간과 로봇 간의 인터랙션이 명확하게 수행될 수 있도록 보편적으로 공감되는 해법이 없다는 점은 HCI를 한 단계 진일보시키고자하는 많은 연구진

과 연구자들에게 있어 중요한 이슈이다. 그 중에서도 인간과 로봇 간의 감정적 소통은 HCI 분야의 큰 부분을 차지하고 있다. 기계가 인간의 감정을 정확히 판단하는 것은 대단히 어려우며 이를 수행하기 위하여 기존의 많은 연구자들이 다양한 방법을 통해 연구를 진행한 바 있다. 감정이 HCI에 적용되기 위하여 기계가 인식할 수 있는 감정에 대한 개념과 이를 인식하기 위한 센서 인터페이스, 인식결과를 판단할 수 있는 결정방법(Decision Method) 등이 요구된다.

Paul Ekman은 사람의 표정은 외형, 성별, 나이 등의 요소에 관계없이 감정에 대하여 공통적 반응이 나타난다는 전제에서 해부학적 기준을 통해 사용자의 얼굴 표정을 6가지의 기본 정서로 분류하였다[1]. Ekamn이 제시한 기본 6정서로 [분노, 슬픔, 행복, 두려움, 놀람, 혐오]를 구분하였으며 다른 감정은 이 기본 6정서의 조합으로 표현하였다. 본 논문에서는 이 구분에 따라 표정에서 나타나는 감정을 분류할

접수일자 : 2009년 4월 6일

완료일자 : 2009년 6월 13일

* 교신 저자

"본 논문은 본 학회 2009년도 춘계학술대회에서 선정된 우수논문입니다."

이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구입니다(No. R01-2008-000-20658-0). 연구비지원에 감사드립니다.

수 있도록 하는 감정인식 시스템을 제안한다.

사람이 사람의 감정을 인식할 때 한 채널의 정보만 활용하기보다 다양한 채널을 활용하여 감정을 인식하므로 실제 감정인식 수행에서는 Multi Modal로 센서 융합 기반의 인식 방법이 요구되지만 이를 수행하기 위하여 개발 센서 기반의 인식 시에도 그 성공률이 높아야 한다는 전제가 필요하다. 이를 충족시키기 위해 본 논문에서는 얼굴 영상에서 감정 특징을 검출 하고 이를 기반으로 감정 상태를 추론하는 방법으로 Active Appearance Model과 Facial Action Coding System을 이용한 감정 특징 검출 기법과 Bayesian Network를 이용한 감정 상태 추론 기법을 제시하고자 한다. 그리고 이 결과가 지능로봇의 HRI로서 활용되기 위한 방법 또한 제시한다.

2. 관련 기술 동향

2.1 감정 특징 검출 기법

사람의 얼굴이 표정으로 감정을 표현할 때에는 기준이 되는 특징영역들이 존재한다. 그렇지만 실제 사용 환경에서는 얼굴 외형의 차이, 데이터베이스 이미지 품질, 외부 노이즈 등의 이유로 정확히 특징만 검출하기가 어렵다[2]. 사람의 얼굴에서 감정적인 특징을 포함하고 있는 영역은 눈, 눈썹, 입 꼬리, 입술 주변부 등으로 주로 눈썹과 입 꼬리 형태에 초점을 맞추어 감정 내포 영역을 설정한다. 기존의 특징 검출과 관련되어 연구되었던 사례로 Eye Blinking Detection, Eye Location Detection, Segmentation of Face area and feature detection, etc [3]과 같은 경우가 있기도 하다. 감정 특징 검출은 크게 두 가지 방법으로 구분되는데 첫 번째로 Holistic 방법으로 전체적인 얼굴 이미지에서의 픽셀 값의 Intensity를 모델링하거나 표현함으로써 특징을 검출하는 방식이다. 두 번째 방법은 기하학적 접근 방식으로 얼굴에서의 특징 점의 기하학적 배열의 위치를 탐색하며 특징을 검출하는 방식이다.

본 논문에서는 이러한 두 가지 특징검출 기법이 융합된 Active Appearance Model을 사용하여 얼굴 표정 상에 존재하는 감정 특징 영역을 기하학적 모델로 구축하여 특징 검출을 수행하고자 한다.

2.2 Active Appearance Model

Active Appearance Model은 영상에서의 특징 검출을 위해 Shape Model과 Texture Model을 구축하는 통계학적인 접근 기법으로 컴퓨터 비전 응용에서 이미지에서의 객체 움직임 추적을 위하여 많이 쓰이는 기법이다. Cootes에 의하여 소개되었으며, Shape 모델과 Texture 모델을 동시에 사용하므로, 이전 버전이라고 할 수 있는 ASM(Active Shape Model)에 비해 외란이나 얼굴 움직임 등에 강인한 특성을 보인다 [4].

AAM에서 쓰이는 Shape는 임의의 Object의 위치, 크기, 회전 효과가 제거된 형태의 기하학적인 정보이다. 또한 Texture는 특징추출, 혹은 추적하고자 하는 목표 이미지에서의 픽셀 Intensity이다. 이 두 가지 정보에 대한 모델을 구성함으로써 얼굴 표정 인식을 위한 특징 검출이 더 좋은 성능을 보임을 이미 여러 논문에서 증명된 바 있다 [4].

n 개의 특징점을 가지는 Shape를 벡터 형식으로 표현한다면 다음과 같다.

$$\vec{x} = [x_{i1}, x_{i2}, \dots, x_{in}, y_{i1}, y_{i2}, \dots, y_{in}]^T \quad (1)$$

이미지의 특징점 영역을 Landmark로 표시하고 이 Landmark로 식 (1)의 해당되는 n 개의 특징점이 표현된다. N 개의 이미지에 대한 Shape 모델을 구하여 평균 Shape를 구하고 또 각 이미지에 대한 편차(Variability)를 표현한다면 아래와 같다.

$$dx_i = x_i - \bar{x}, \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (2)$$

PCA에 이 값을 적용하여 Shape와 Texture의 eigen 분산, 즉 특정 Shape와 Texture를 표현할 수 있는 선형 모델을 얻을 수 있다.

$$\vec{s} \approx \bar{\vec{s}} + \Phi \vec{b}_s, \vec{t} \approx \bar{\vec{t}} + \Phi \vec{b}_t \quad (3)$$

아래의 그림 1은 이와 같은 과정을 거쳐 생성된 Shape에 대한 예시이다.

이렇게 생성된 Shape와 Texture모델을 이용하여 사용자의 얼굴에서 나타나는 감정 영역을 특징으로 선택하여 그림 2와 같이 감정인식이 수행 가능하다.

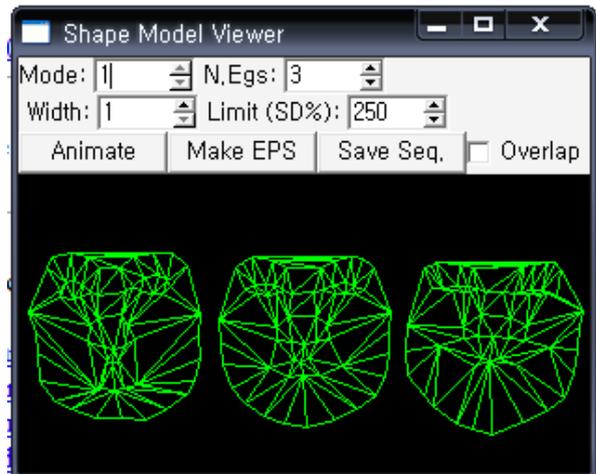


그림 1. Shape Model 생성
Fig. 1 Shape model generation

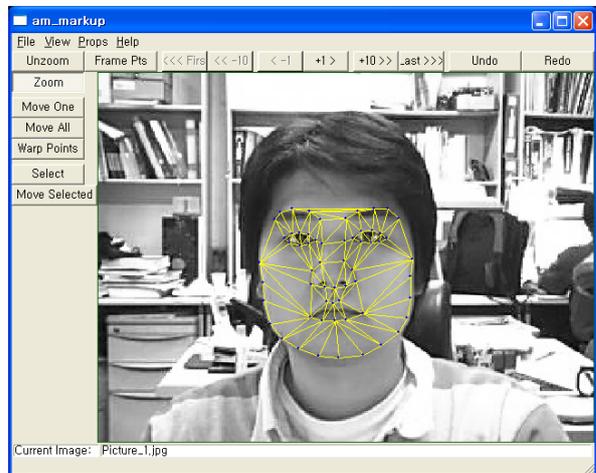


그림 2. AAM 모델 기반 얼굴 특징 탐색
Fig. 2. AAM model based face feature search

2.3 얼굴 표정 인식 과정

일반적으로 얼굴표정을 인식할 경우 아래의 그림 3과 같이 보편적인 수행 과정이 있다.

첫 번째로 영상 센서(카메라)를 통해 획득된 RAW 데이터를 Pre-Processing을 통해 화질 보정 및 노이즈 제거를 수행한다.

두 번째로 입력 영상에서 감정, 움직임 등의 특징점 검출을 수행한다.

세 번째로 검출된 특징점의 패턴분류 기반으로 주어진 영상의 상태를 분류한다.

네 번째로 Post-Processing을 통해 분류된 영상의 패턴을 다단계로 반복 분류한다.

최종적으로 선별된 결과를 인식 결과로써 도시함으로써 인식과정은 종료된다.

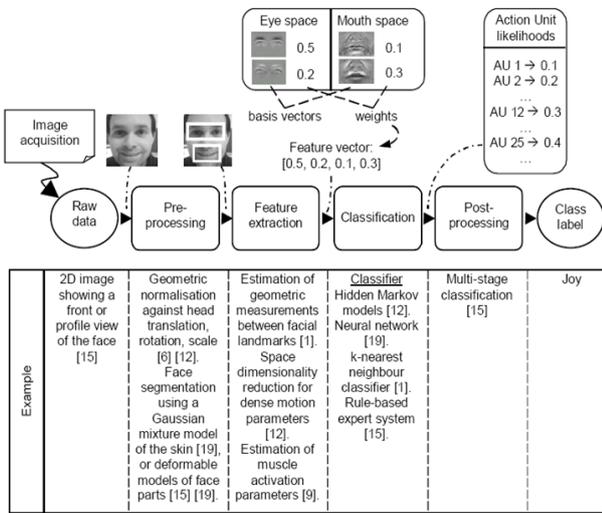


그림 3. 얼굴표정 인식과정
Fig. 3. Recognition process of facial expression

본 논문에서는 입력 영상을 Pre-Processing을 제외한 특징 검출, 표정의 패턴분류와 표정인식 단계에서 AAM과 Bayesian Network를 사용하여 얼굴 표정 인식을 수행하고자 한다.

2.4 베이저안 네트워크(Bayesian Network)

BN은 불확실한 상황을 확률로 표시하고, 복잡한 추론을 정량화된 노드의 관계로 단순화 시켜 노드의 연결 관계로 표현하는 방향성 비순환 그래프로, 미리 정의된 조건부확률 테이블(CPT)에 의해 확률 관계를 효율적으로 표현 가능하다[5][6].

BN를 구축하기 위한 구조 및 확률분포는 각각 전문가에 의해 설계되고, 전문가가 계산하거나, 주어진 데이터에 의해 학습된다. 학습 후 어떤 상황에 대한 증거가 관찰되면 이를 바탕으로 베이저안 추론 알고리즘을 통해 각 노드의 상태에 대한 확률이 계산된다.

BN의 네트워크 모델 G 는 네트워크 구조 B_s 와 파라미터 집합 θ 로 (G, θ) 쌍으로 정의 가능하다. $\theta = \{B_\phi, B_p\}$ 는 B_ϕ : 조건부 확률 분포, B_p : 초기 확률 분포로 구성된다. x_i 를 $\{x_i^1, x_i^2, \dots, x_i^n\}$ 의 p_i 개의 값을 갖는 BN G 의 i 번째 노드라 하고, π_i 를 $\{\pi_i^1, \pi_i^2, \dots, \pi_i^n\}$ 의 q_i 개의 상태 조합을 갖는 x_i

의 부모 집합이라고 하면, x_i 에서의 CPT값을 $P(x_i = x_i^j | \pi_i = \pi_i^j)$ 로 정의할 수 있다. 따라서 BN 모델링은 주어진 도메인을 가장 잘 표현하는 $\{B_s, B_\phi, B_p\}$ 를 찾는 것이다[10]. BN의 추론을 위하여 주어진 증거노드 E에 대한 추론 결과 R에 대한 확률을 구하면 다음과 같다.

$$P(R) = P(R | E) = \frac{P(E | R)P(R)}{P(E)} = \frac{P(R, E)}{P(E)} \quad (4)$$

조건부확률을 Chain Rule에 의해 표현하면,

$$P(x) = P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i | x_1, \dots, x_{i-1}) \quad (5)$$

그림 4에 주어진 예시 BN의 전체 노드의 결합 확률은 다음과 같다.

$$P(P, V, I, D) = P(P)P(V)P(I | P, V)P(D | I) \quad (6)$$

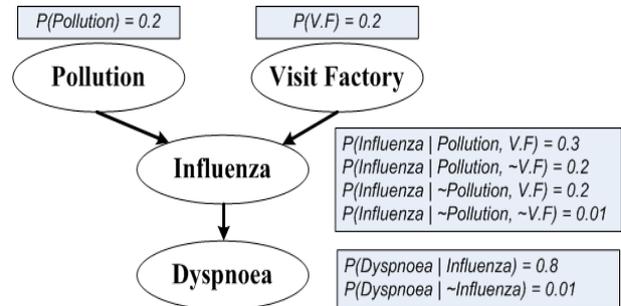


그림 4. 베이저안 네트워크 예시
Fig. 4. Bayesian network example

3. FACS+AAM에 기반한 얼굴 특징 검색 알고리즘 개발

3.1 FACS(Facial Action Coding System)

FACS는 Paul Ekman과 Wallace Friesen에 의해 1976년 제안된 얼굴 표정 분석 기법으로 해부학적 토대로 사람의 얼굴 근육을 분석, 46개의 Action Unit으로 얼굴 표정 특징 영역을 구분 지었다 [7].

각각의 Action Unit은 감정 표현 시 움직이는 얼굴 근육에 대한 모사로 이를 잘 활용할 경우 표정을 분석하는데 대단히 유용한 도구이다. FACS의 경우 다수의 얼굴 표정을 표현하고 있는 얼굴 이미지에 대하여 사람이 직접 수동으로 Action Unit을 비교, 분석하여 어떤 표정을 표현하고 있는 결정하는 일종의 Manual System이다. 그렇지만 이러한 FACS가 사람에 의해서가 아닌 자동으로 동작하는 Human Computer Interface 상에 적용된 연구사례는 현재까지 극히 드물다. 본 논문에서는 이 부분에 주안점을 두고 FACS를 자동화된 시스템에 적용하여 사람의 표정을 쉽게 구분할 수 있는 얼굴 표정 인식 시스템을 개발하고자 하는 것이다.



그림 5. FACS 기반 Action Unit 예시
Fig. 5. FACS based action unit example

3.2 FACS 기반 AAM 생성

AAM을 통해 감정인식을 위한 Facial Appearance 모델 생성 시 가장 문제가 되는 부분은 어떤 부위를 감정 특징 영역으로서 검출해야 신뢰성 있는 결과가 나오는지 하는 것이다. 이를 위하여 본 논문에서는 FACS에 기반한 AAM 생성을 수행하여 신뢰성 있는 결과가 도출될 수 있도록 하고자 한다. 즉, 앞서 언급한 바와 같이 표정 표현 시 가장 많이 움직이는 Action Unit의 영역을 감정 특징으로 선별, 이를 AAM을 통해 Appearance 모델로 생성하는 것이다.

4. 시뮬레이션 및 결과 고찰

본 논문에서 제안하는 얼굴 표정 인식 시스템의 시뮬레이션은 다음과 같은 과정을 거쳐 수행된다.

그림 6의 과정을 통해 FACS를 이용하여 선별된 Database의 얼굴영상 이미지들은 감정 정보를 가장 많이 내포한 Action Unit 영역을 120개의 특징점으로 지정하였다. 이는 수동으로 Landmark를 지정하는 방식으로 수행하였다. 지정된 120개의 특징점은 기 구축된 DB의 이미지 파일의 수에 따라 적절한 Appearance Model을 생성해내고, Action Unit 간의 상관관계 및 확률적 인과관계에 따라 구축된 Bayesian Network에 의하여 가장 높은 확률 값을 가지는 감정으로 판별하여 표정에서 나타나는 감정을 인식하게 된다.

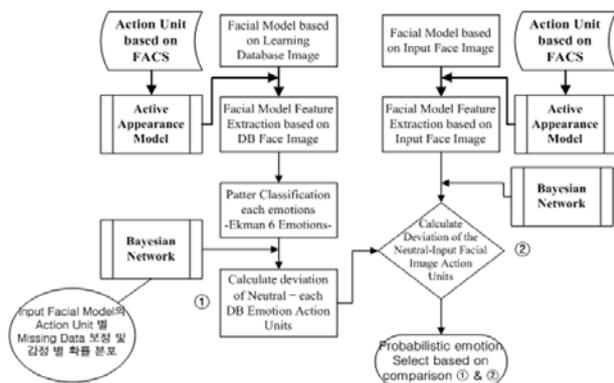


그림 6. 얼굴 표정 인식 시뮬레이션 과정
Fig. 6. Simulation process of facial expression recognition

그림 7에서와 같이 선별된 Landmark에 기반으로 입력 이미지에 대한 Shape Model과 Appearance Model이 생성되고, 아래의 표 1에 해당되는 Action Unit에 대응되는 이미지 영역의 값들을 인식하고자 하는 감정 특징으로 정의하였다.

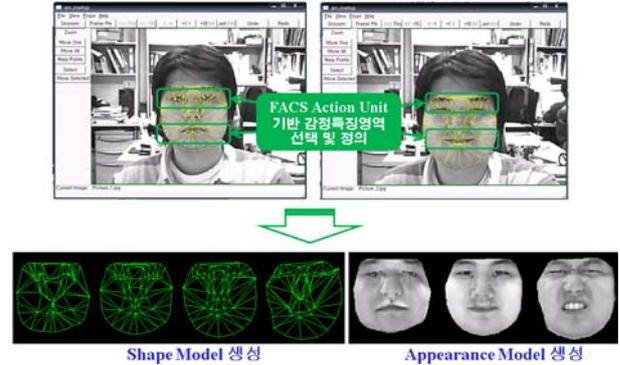


그림 7. Input Image에 대한 AAM 구축
Fig. 7. Structure of AAM for input image

표 1. Action Unit 정의
Table 1. Action Unit Definition

AU	Definition	AU	Definition
1	눈썹이 처짐	5	눈꺼풀이 약간 내려감
2	눈썹을 치켜뜸	6	입꼬리가 올라감
3	눈꺼풀이 약간 올라감	7	입꼬리가 내려감
4	눈꺼풀이 많이 올라감	8	입이 벌어짐

표 1을 통해 본 논문에서 인식하고자 하는 얼굴에서의 감정 포함 Action Unit을 정의하였다. AAM을 통해 검출된 각 AU 영역의 변화에 기반 Bayesian Network를 구성한 것은 아래의 그림 8과 같다.

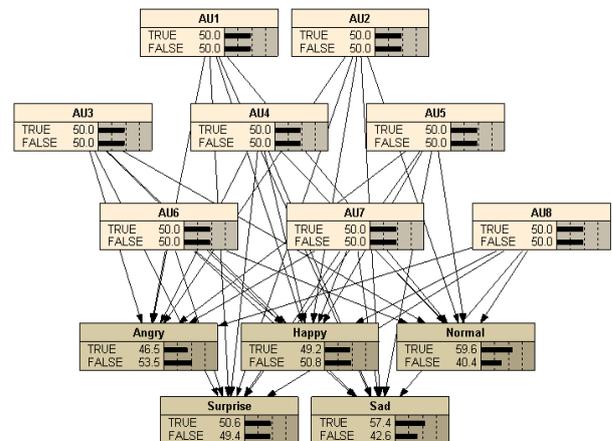


그림 8. FACS 기반 Bayesian Network Structure
Fig. 8. FACS based Bayesian Network Structure

그림 8에 주어진 Bayesian Network의 구조는 감정 표현 시 변화를 보이는 AU 간의 상관관계를 고려하여 임의로 구성된 형태이며, 네트워크의 각 노드 간의 확률 인과 관계를 표현하는 조건부확률테이블의 값들은 AU 움직임 또는 변화 여부에 따라 Decision 노드의 확률 값들이 변화하도록 8개의 Parent Node의 경우 True/False의 Deterministic 값을 지정하고, Child Node, 즉 Decision Node는 FACS 기반 얼굴 감정 표정 분석을 통해 각 AU의 움직임 간의 상관관계를 임의로 지정한 값을 대입하였다.

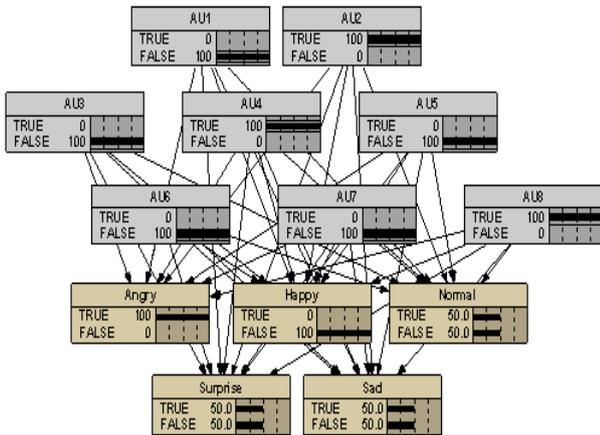


그림 9. 얼굴 표정 인식 결과 : 화(angry)
Fig. 9. Facial expression recognition result : angry

시뮬레이션으로서 그림 9에서 나타난 Bayesian Network에서 입력으로 들어온 이미지의 AU 2, 4가 Neutral Image에 대비 변화를 보일 경우 기 구축된 CPT에 의하여 Decision Node의 state의 확률 값들이 출력되게 된다. 그림 9의 경우 Angry node의 True state 값이 100을 표시함에 따라 [표정 인식 결과 = Angry]임을 알 수 있다. 이것은 본 시스템은 입력 이미지에서 눈썹을 치켜뜨고 눈꺼풀이 많이 올라간 경우 화난 얼굴이라고 정의한다는 것을 의미한다.

다음의 그림 10과 같은 경우도 동일 과정을 거쳐 얼굴 표정의 의미를 도출하고 있다.

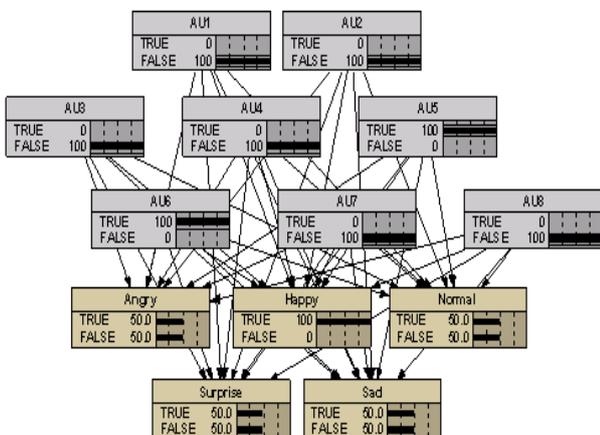


그림 10. 얼굴 표정 인식 결과 : 기쁨(happy)
Fig. 9. Facial expression recognition result : happy

그렇지만 본 시스템에는 몇 가지 해결해야하는 문제점이 존재하는데 대표적인 경우가 아래의 그림 11과 같은 경우이다.

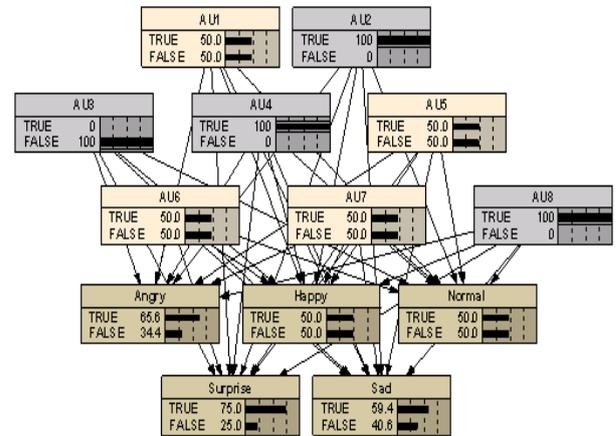


그림 11. 유사 표정 분류 시 AU 중첩되는 경우
Fig. 11. In case being piled one the another when classify the similar facial expression

각각의 얼굴 표정에 대응되어 움직이는 AU가 고유하지 못하다는 것이 원인이 되어 그림 11의 경우처럼 AU 2, 3, 4, 8이 움직일 경우, 정확히 특정 감정만을 표현한다고 시스템이 확실하게 인식할 수 없게 된다. Decision Node에서도 볼 수 있듯이 Angry와 Surprise, Sad Node의 True state 값이 각각 [Angry = 65.6%, Surprise = 75.0%, Sad = 59.4%]의 값을 가지게 되어 가장 큰 확률 값을 가지는 Surprise가 해당 얼굴 표정이라고 정의내릴 수 있다. 그렇지만 이는 그림 9, 그림 10의 경우와 다르게 정확히 100% 일치하는 상황이 아니고 개중 가장 높은 확률 값을 선택하는 경우에 불과하므로 이를 보완하여 좀 더 명확한 결정의 증거를 삼기 위한 추가적인 연구가 필요하다.

5. 결론 및 향후 과제

본 논문에서는 얼굴 표정에서 나타나는 감정을 인식하기 위한 FACS 기반 AAM과 Bayesian Network를 이용한 얼굴 표정 인식 기법에 대하여 논하였다.

신뢰성 높은 얼굴 감정 특징 영역을 탐색하기 위하여 FACS를 이용한 AAM을 적용하였고, 이를 통해 생성되는 Appearance 모델은 Bayesian Network에 의하여 데이터베이스 구축에서 발생 가능한 Missing Data나 입력 데이터 변동에 강인한 추론이 이루어지게 된다.

향후 과제로는 FACS 기반 특징 검출 시 자동적으로 특징 영역을 지정할 수 있는 Automatic Landmarker Tool의 개발이 필요하며, 감정 상태 추론을 위한 베이지안 네트워크의 구조와 CPT에 대한 적절한 구조학습 및 파라미터 학습이 이루어질 수 있도록 추가 연구를 진행할 예정이다. 또한, 앞서 언급한 바와 같이 현재 수준의 시스템으로는 유사 표정, 즉 AU의 일부가 중첩되어 변화되는 감정들의 명확한 구분을 Decision Node의 최대 확률 값을 가지는 값을 선택하는 것으로 구분짓는 가장 단순한 형태에 불과하므로 추후 이에 대한 확률적인 명확한 구분이 가능하도록 추가 연구를

진행할 예정이다.

저 자 소 개

참 고 문 헌

[1] Ekman, P. & Friesen, W. V, "The repertoire of nonverbal behavior : *Categories, origins, usage, and coding. Semiotica*, 1, pp. 49-98

[2] Mark Weiser, "The Computer for 21st Century", *ACM SIGMOBILE Mobile Computing and Communications Review archive*, Vol 3, Issue 3, pp.3-11, July 1999

[3] Shi Yi-Bin, Zhang Jian-Ming, Tian Jian-Hua, Zhou Geng-Tao, "An improved facial feature localization method based on ASM," *Computer-Aided Industrial design and Conceptual design, 2006, CAIDCD '067th international conference on*

[4] Seiji Kobayasho and Shuji Hashimoto, Automated feature extraction of face image and its applications, in: *International workshop on Robot and Human Communication*, pp. 164-169

[5] T.F Cootes, G.J. Edwards and C.J. Taylor, "Active Appearance Model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.23, No.6, June 2001, pp. 681-685

[6] 박지형, "유비쿼터스 환경에서의 상황 인지 시스템 - 연구 활동 소개 도우미 -", *한국정밀공학회지*, 제 21권, 제11호 pp. 31-37, 2004. 11

[7] 황금성, 조성배, "베이지안 네트워크의 학습", *로봇공학회지*, 제3권, 제4호, pp.15-17, 2006. 10.

[8] P. Ekman and W. Friesen. *Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto, 1978.*



고광은(Kwang-Eun Ko)

2007년 : 중앙대학교전자전기공학부 공학사
2007년~현재 : 중앙대학교 대학원 전자전기공학부 석박사 통합과정

관심분야 : Multi-Agent Robotic Systems (MARS),
Machine Learning Context Awareness
Email : kke@wm.cau.ac.kr



심귀보(Kwee-Bo Sim)

1990년 : The University of Tokyo
전자공학과 공학박사

[제19권 2호 (2009년 4월호) 참조]

1991년~현재 : 중앙대학교 전자전기공학부 교수
2006년~2007년 : 한국 퍼지 및 지능시스템학회 회장

E-mail : kbsim@cau.ac.kr
Homepage URL : <http://alife.cau.ac.kr>