

베이지안 네트워크에서 기하학적 관계를 이용한 얼굴 동작 인식

윤영지*, 정유선*, 신보경*, 김혜민*, 박동석*, 박호식**, 배철수**, 이상동*

조선대학교* 관동대학교**

sdna@chosun.ac.kr

Facial Behavior Recognition Using Geometric Relations of Bayesian Network

Youn Youngji*, Jeoung YouSun*, Shin BoKyoung*, Kim HyeMin*, Park Dongsuk*, Park Hosik**, Bae Cheolsoo**, Ra Sangdong*

.Chosun Univ*. Kwandong Univ**

요약

얼굴 동작을 효과적으로 인식하는 방법을 제안하고자 한다. 얼굴 동작은 얼굴 표정, 얼굴 자세, 시선, 주름 같은 얼굴 특징이나 얼굴 행동 등으로 표출될 수 있다. 이러한 표출된 정보들은 얼굴 동작이 다양하고 명확하지 않아 연구 진행에 많은 어려움이 있다. 그러므로, 본 논문에서는 얼굴 동작을 묘사하는 FACS를 기반으로 하여 시각적 관찰에 의해 주요한 얼굴 동작을 표현하고, 베이지안 네트워크를 통하여 여러 정보를 분석 융합하여 얼굴 행동을 추론할 수 있도록 하였다. 베이지안 네트워크의 방향성 추론으로 시각 정보를 선택할 수 있고, 관측된 현상을 토대로 방향성 추론 하여 얼굴 동작의 신뢰 전파를 통하여 분류 인식한다.

I. 서론

인간의 감성을 자동으로 전달하고자 하는 필요성으로 많은 연구가 수행되고 있다. 하지만 현실 세계에서 이를 적용하기에는 2가지 문제점을 가지고 있다. 하나는 얼굴 동작이 다양하고 복잡하다는 것이고, 다른 하나는 얼굴 동작의 표출 방식이 매우 다양하다는 것이다. 즉, 얼굴 동작은 동공의 움직임, 주름의 발생 등 기하학적 변형으로 나타난다. 그러나 하나의 시각 정보만으로 이러한 변화를 효율적으로 획득할 수 없다.

특정 응용분야에서는 관심 얼굴 동작을 몇 개의 동작으로 한정 지을 수 있으므로 특정 분야로 한정할 경우 얼굴 표정에서 다양함으로 기인되는 애매함을 배제할 수 있다. 게다가 얼굴 특징과 같은 시각 정보들은 대다수 효율적으로 검출할 수 있다. 또한, 효율적인 얼굴 특징들을 선택적으로 분석 융합하면 보다 정확한 인식을 할 수 있을것이다.

II. 얼굴 특징 추적

2.1. 적외선 기반 눈 검출 및 추적 시스템

눈의 동공은 얼굴에서 주요한 특징을 나타낸다. 동공은 눈의 움직임 뿐만 아니라 얼굴과 시선의 위치 정보 까지 제공한다. 그러나, 다양한 조명 조건과 여러 얼굴 방향에 대해 건설하게 동공을 검출 및 추적하는 것이 쉽지 않다. 본 논문에서는 IBM[1]에 의해 개발된 것과 유사한 적외선 기반의 눈 검출기를 개발하여 실용적인 검출 및 추적 시스템을 구축하였다.

얼굴 특징 추적 대다수의 얼굴 특징들은 얼굴 특징점의 이동 윤곽으로 표현할 수 있다. 그러므로 얼굴 특징점을 건설하게 검출 및 추적해야 한다. 그림 2에 표시한 지점과 같이 22개의 특징점과 이동 윤곽으로 얼굴 동작 특징을 나타내었다. Wiskott가 제안한 방법과 유사한 능동적 추적 시스템을 구축하여 각 특징점을 Gabor 웨이블 커널을 이용하여 식 (1)과 같이 특징지었다.

$$\Psi(k, \vec{x}) = \frac{k^2}{\sigma^2} e^{\frac{k^2 \vec{x}^2}{2\sigma^2}} (e^{ik\vec{x}} - e^{-\frac{\sigma^2}{2}}) \quad (1)$$

여기서 σ 는 128×128 영상에서 π 로 설정하였다. Gabor 커널의 조합은 3개의 공간 주파수($\pi/2, \pi/4, \pi/8$)와 0°에서 150°사이에서 30°간격으로 6개의 방향으로 구성하였다. 각 화소 pixel(x)에 대한 18개의 복소 형태의 Gabor 계수의 조합 $\Omega(x)$ 를 Gabor 커널의 컨볼루션으로 얻을 수 있었다.

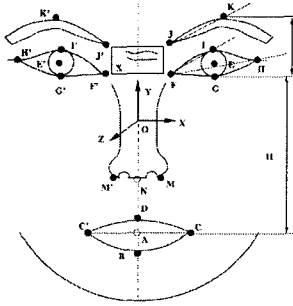


그림 2. 얼굴 특징점과 주름의 기하학적 관계.
Fig. 2. The geometric relations of facial feature points and furrows.

III. 얼굴 동작 인식

얼굴 동작은 전체 얼굴을 의미한다. 하나의 AU는 지역적 특징의 움직임을 나타낸다. 본 논문에서는 FACS(Facial Action Coding System)[2]에 따른 얼굴 동작 AU(Action Unit)와 관련하여 베이지안 네트워크 기반의 전체적인 얼굴 동작 인식모델을 제안한다.

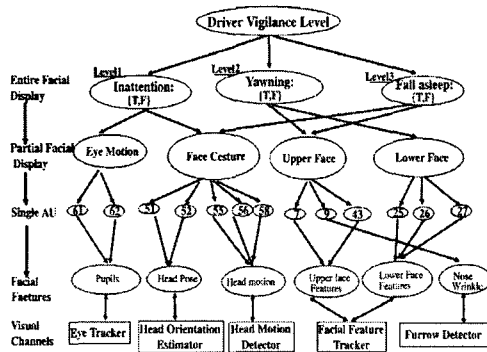


그림 3. 운전자 경계 상태 추정을 위한 BN 모델.
Fig. 3. BN model for driver vigilance estimation.

그림 3에서 나타난 것과 같이 베이지안 네트워크 모델에서는 시각 정보 감지를 포함한다. 시각 채널은 베이지안 네트워크의 한 부분은 아니지만 얼굴 특정 노드에 상응하는 정보를 획득하는데 사용되어진다. 시각 채널에서 양방향 연결은 정보를 제어한다. 그중 하나인 상향식 연결은 감지기에서 제공된 관측된 현상을 제공하고, 다른 하향식 연결은 베이지안 네트워크 추론으로 인하여 감지기를 동작시킨다. 상, 하향식 2가지 중 하나를 선택하기 위하여 베이지안 네트워크를 수동으로 데이터 학습을 하지만, 많은 학습 데이터를 요구하고, 학습된 경우라도 빠르게 동작하지 않는 경우도 발생한다. 그러므로 타협적으로 베이지안 네트워크를 별도로 초기 지정 할 수 있도록 하였다.

베이지안 네트워크 모델의 상층에서 간단한 Naive 베이지 분류기를 볼 수 있다. Naive 베이지 분류기는 가설 변수 C 와 상호 관계가 있는 배타적 상태 즉, 부주의, 하품, 잠듦 같은 c_1, c_2, c_3, \dots 등으로 구성된다.

부주의, 하품, 잠듦에 대응하는 속성 집합 A_1, A_2, A_3 와 가설 변수는 주 노드이고 각각의 속성은 하나의 중 노드에 대응한다. 식 (2)와 같이 $A_j = a_j$ 가 되는 c_i 상태의 확률을 찾는다.

$$\Pr(C = c_i | A_1 = a_1, A_2 = a_2, A_3 = a_3) \quad (2)$$

즉, 이 확률은 속성 변수가 $A_j = a_j$ 에서 가지는 값이 c_i 상태일 가능성을 나타낸다. 특별한 상태 c_i 의 확률이 최고가 될 때, 얼굴 동작은 c_i 를 포함한다고 할 수 있다.

표 1에 가설 변수 상태와 대응하는 속성 사이의 초기 조건 확률을 나타내었다. 이 수치는 이후 학습 절차에 의해 지속적으로 정제된다.

표 1. 초기 확률.
Table 2 Initial probabilities.

Entire display	Ia		Yw		Fa	
	T	F	T	F	T	F
Inattention	0.99	0.01	0.05	0.95	0.05	0.95
Yawning	0.05	0.95	0.99	0.00	0.05	0.95
Falling asleep	0.05	0.95	0.05	0.95	0.99	0.01

하부 층은 각각의 AU와 얼굴 특징과 특정한 이동 윤곽을 나타낸다. 예를 들면 AU7은 현재의 특징점 I, F, H, 와 G를 추적하여 각 $\angle IFH$ 와 $\angle HGF$ 의

변화를 측정할 얼굴 상부 특징이다.

초기화에서 중립 상태에서 모든 특징점은 추출되어지고, 관련 초기 매개 변수를 얻게된다. 이 초기 매개변수는 인식 과정에서 문턱치 값으로 사용되어진다.

IV. 실험 결과 및 고찰

제안된 방법을 운전자의 경계 상태 감시와 같은 특정한 분야에 응용하기 위한 일련의 실험을 진행하였다. 총 5명의 피검자를 대상으로 각각 실내에 앉은 상태로 각성 혹은 부주의 같은 낮은 피로 상태와 하품, 잠든 같은 높은 피로 상태를 나타내도록 요구하였고, 적외선 카메라 시스템을 이용하여 초당 20프레임 320×240 해상도로 얼굴 영상을 약 1분씩 총 5회 기록하여 실험하였다. 그러므로 같은 동작으로 모든 단계(부주의, 하품, 잠든)에 대하여 상향식 분류에 사전 확률을 설정한다.

표 2. 관측 현상에 의한 얼굴 동작 분류.

Table 3. Facial behavior classification by observation evidence.

No.	Observation evidences					Classification results		
	ED	OE	HD	FT	FD	(Ia)	(Yw)	(Fa)
(a)	x	x	x	x	x	0.33	0.33	0.33
(b)	x	AU51	x	x	x	0.82	0.09	0.09
(c)	x	x	x	AU7 AU27	AU0	0.15	0.83	0.02
(d)	x	AU52	x	x	x	0.82	0.09	0.09
(e)	x	x	x	AU7 AU27	x	0.17	0.81	0.02
(f)	x	AU51	AU55	x	x	0.47	0.05	0.47

표 2에 각 감지기에서 관측된 현상과 분류 결과를 나타내었다. 표 2에서의 ED, OE, HD, FT와 FD는 각각 눈 검출(Eye Detector), 머리 방향 추정(Head orientation estimator), 머리 움직임 추정(Head motion detector), 얼굴 추적(Facial tracker)과 주름(Furrow detector)이다. 또한 각각 (Ia), (Yw)와 (Fa)는 부주의(Inattention), 하품(Yawning)과 잠든(Falling asleep)이다. "x"는 현상이 없는 것을 의미한다. 표 2의 (a)행은 피로와 관련하여 아무런 현상이 없을 때로, 각 피로 상태의 사전 확률은 베이지안 추론에 의해 동일하게 0.33이었다. 이것은 각성 상태이거나 일

굴에서 아무런 피로 상태가 감지 되지 않았다는 것을 의미한다. (b)행은 시각 채널 OE에서 AU51이 검출되었을 때이다. 부주의 상태에 대한 높은 사전 확률을 나타내었다. 얼굴 동작은 부주의의 한 상태로 분류되었다. (c)행은 시각 채널 FT와 FD에서 각각 AU7, AU27 과 AU0이 검출되었을 때이다. 분류 결과는 하품이다. (d)와 (e)행은 분류 결과는 각각 부주의와 하품이다. (f)행은 OE 와 HD 채널에서 각각 AU51과 AU55과 검출되었다. 부주의와 잠든이 사전 확률이 동일하게 0.47로 나타났다. 이 경우와 같이 얼굴 정보가 불분명 할 때는 피로 상태를 분류하기에는 어려움이 있었다.

실험결과 베이지안 네트워크를 특정한 응용분야에 적용시 얼굴 정보로도 좋은 동작 분류 결과를 나타내었다. 그러나 얼굴정보가 불분명할 때와 상태가 변화하는 과정에서는 정확한 분류가 되지 않아 향후 시간 정보를 이용하여 문제를 해결할 계획이다.

V. 결론

인간의 감성을 자동으로 전달하고자 하는 필요성으로 많은 연구가 수행되어지고 있으며 특정한 목적에 초점을 맞추게 된다면 관심 동작을 몇 개의 동작으로 한정 지을수 있다. 본 논문에서는 FACS를 기반으로 하여 시각적 관찰에 의해 주요한 얼굴 동작을 표현하였고, 시각 정보를 베이지안 네트워크를 통하여 분석 융합하여 얼굴 행동을 추론 하였다. 베이지안 네트워크의 하향식 추론으로 시각 정보를 선택할 수 있고, 관측된 현상을 토대로 상향식 추론 하여 얼굴 동작의 신뢰 전파를 통하여 얼굴 동작 결과를 나타내었다. 얼굴정보가 불분명할 때와 상태가 변화하는 과정에서는 정확한 분류가 되지 않아 향후 시간 정보를 이용하여 문제를 해결할 수 있을 것이다.

참 고 문 헌

- [1] C. Morimoto, M. Flickner, Real-time multiple face detection using active illumination, in: Proc. Fourth IEEE Internat. Conf. on Automatic Face and Gesture Recognition, 2000, pp. 8-13.
- [2] P. Ekman, W.V. Friesen, J.C. Hager, Facial Action Coding System (FACS): Manual, CD Rom, San

- Francisco, CA, 2002.
- [3] I. Cohen, N. Sebe, F. Cozman, M. Cirelo, T. Huang. Learning bayesian network classifiers for facial expression recognition using both labeled and unlabeled data. IEEE Conf. Computer Vision Pattern Recognition, 2003.
- [4] R. Kaliouby, P. Robinson. Real-time inference of complex mental states from facial expressions and head gestures, in: Proc. in CVPR Workshop on Real-Time Computer Vision for Human Computer Interaction, 2004. p. 154.
- [5] N. Oliver, E. Horvitz, A. Garg. Layered representations for human activity recognition, in: Proc. IEEE Internat. Conf. on Multimodal Interfaces, 2002, pp. 3-8.