

Facial Actions 과 애니메이션 원리에 기반한 로봇의 얼굴 제스처 생성

Generation of Robot Facial Gestures based on Facial Actions and Animation Principles

박정우, 김우현, 이원형, 이희승, 정명진*
(Jeong Woo Park¹, Woo Hyun Kim¹, Won Hyong Lee¹, Hui Sung Lee², and Myung Jin Chung^{1,*})

¹Department of Electrical Engineering, KAIST

²Hundai Motors

Abstract: This paper proposes a method to generate diverse robot facial expressions and facial gestures in order to help long-term HRI. First, nine basic dynamics for diverse robot facial expressions are determined based on the dynamics of human facial expressions and principles of animation for even identical emotions. In the second stage, facial actions are added to express facial gestures such as sniffing or wailing loudly corresponding to sadness, laughing aloud or smiling corresponding to happiness, etc. To evaluate the effectiveness of our approach, we compared the facial expressions of the developed robot when the proposed method is used or not. The results of the survey showed that the proposed method can help robots generate more realistic facial expressions.

Keywords: facial gestures, realistic facial expressions, human-robot interaction

I. 서론

국내외 여러 연구 기관에서 서비스 로봇이 세계 시장에서 차지하는 비중은 앞으로 더 커질 것으로 예상된다. 사람의 역할을 대신하는 서비스 로봇의 성공을 위한 가장 핵심적인 기술은(HRI: Human-Robot Interaction)이 될 것이다[1]. HRI 기술은 크게 세 가지 (인지, 인지 및 감정 처리, 표현) 기술로 구분할 수 있다.

인식은 여러 센서들을 통해 로봇에 주어지는 외부 자극을 받아들이고 주변 상황 정보를 알 수 있도록 해주는 기술이다. 마이크를 이용한 화자 인식, 카메라를 이용한 얼굴 인식 등이 그 예이다. 인지 및 감정 처리는 로봇이 해야 할 동작의 종류나 감정 상태를 로봇의 내/외부 정보를 이용하여 학습이나 추론 과정을 통해 결정하는 것이다. 표현은 인지 및 감정 처리에서 결정된 행동의 종류나 감정 상태를 얼굴 표정, 제스처, 휘, 소리 등을 통해 나타내는 기술이다.

위에서 언급한 기술 모두가 HRI를 위해 중요하지만, 표현 기술은 사용자가 로봇의 행동이나 표현을 보고 feed-back을 주면서 로봇과 상호작용한다는 측면에서 매우 중요하다고 할 수 있다. 특히, 교육 환경에서 사람과 상호 작용이 가능한 로봇들이 아이들에게 잘 받아들여지는 경향이 있으나 오랜 시간 동안 아이들과 상호 작용 하기 위해서는 다양한 표현을 로봇이 만들 수 있어야 한다고 알려져 있다[35]. 이러한 관점

에서 HRI 을 위해 로봇의 표현 생성에 관한 연구가 필요하다고 본다. 본 논문에서는 보다 사실적으로 로봇의 표정을 생성하는 기법을 제안하여 장시간 동안 사람이 로봇과 상호 작용할 수 있도록 기여하고 사람들로 하여금 로봇이 살아있다는 느낌을 받을 수 있기를 기대한다.

이후의 논문 구성은 다음과 같다. II 장에서는 로봇의 얼굴 표정 생성에 관한 이전 연구들을 소개하고, III 장에서는 본 논문의 기반이 되는 LAESM (Linear Affect-Expression Space Model)에 대해 간단히 설명할 것이다. IV 장에서는 로봇의 표정 사이의 변화를 자연스럽게 연결 시킬 수 있을 뿐만 아니라 하나의 감정에 대해서 다양한 표정 변화를 보여줄 수 있는 방법을 제안한다. V 장에서는 facial actions에 기반하여 얼굴 제스처를 만드는 방법에 대해 소개한다. VI 장에서는 실험결과 및 제안한 방법을 검증하고, VII 장에서 결론 및 후속 과제에 대해 언급할 것이다.

II. 이전 연구

지금까지 로봇의 감정 생성과 표현에 관해 많은 연구들이 진행되었고, 이들 연구에서 다양한 방법들이 가상의 로봇 얼굴이나 혹은 실제 로봇에 적용되었다. 가상의 얼굴 로봇은 주로 avatar 혹은 character 형태로 많이 제작되었고, 대부분 FAP (Facial Animation Parameter)에 기반하여 표정을 생성한다 [2-6]. 애니메이션에 흔히 볼 수 있듯이, 가상의 캐릭터들의 표정은 매우 자연스럽게 실제 로봇에 비해 정교한 표정들을 훨씬 잘 만들어낸다. 반면, 실제 로봇은 FACS (Facial Action Coding System)에 기반하여 표정을 만들며 크게 기계 형[8-12], 마스크트 형[13-16], 안드로이드 형[17-20]으로 구분할 수 있다. 동일한 표현 알고리즘이 적용된다고 가정할 때, 안드로이드 형의 경우는 로봇의 외형이 uncanny valley에 빠질 위험이 크고,

* Corresponding Author

Manuscript received February 15, 2014 / revised March 15, 2014 / accepted March 30, 2014

박정우, 김우현, 이원형, 정명진: KAIST 전기 및 전자공학과
(pjw@rr.kaist.ac.kr/ishrain@rr.kaist.ac.kr/leestation@rr.kaist.ac.kr/mjchung@ee.kaist.ac.kr)

이희승: 현대자동차(huisung.lee@kaist.ac.kr)

※ 본 논문은 지식경제부 산업융합원천기술개발사업에 의하여 연구되었음(N02120248).

기계 형의 로봇은 사람과 많이 다른 형태라 *uncanny valley* 차트의 가장 왼쪽 끝에 위치해서 사람들에게 친근감을 주도록 디자인하기가 어렵다[22]. *Uncanny valley*의 개념을 처음 제안한 Mori의 최근 인터뷰에 따르면 로봇의 외형은 정확히 사람의 것과 같을 필요는 없고 *uncanny valley*를 뛰어넘는 형태의 디자인을 만들기 보다는 첫 번째 꼭대기에 위치할 수 있을 정도의 외형을 가진 로봇이면 충분히 사용자들에게 친근감을 줄 수 있다고 하였다[22,23]. 첫 번째 꼭대기에 해당하는 로봇들이 대부분 마스크트 형의 로봇이다.

지금까지 많은 연구자들이 로봇의 얼굴 표정을 더 자연스럽게 사람의 표정 변화와 유사하게 만들 수 있는 방법을 제안하기 위해 노력하였다. 초기에는 Ekman이 제안한 기본 6가지 표정(화남, 놀람, 기쁨, 슬픔, 역겨움, 공포)을 생성하는 것이 목표였다. 그 이후에는 표정의 개수를 더 하거나 감정의 강도를 표현하기도 하였다. 두 개의 감정이 섞인 복합 표정 생성 혹은 제스처나 다른 표현 수단(LED, 음향, 뿔, 더듬이 등)과의 동기화 문제에 대한 연구도 진행되었다[10,12,15,21,24]. 또 다른 연구에서는 정적인 상태에서의 로봇에 생동감을 부여하기 위해 눈동자와 목 움직임, 눈 깜빡임에 대한 결과도 보여 준다[25].

그러나 이전 연구들의 변화를 자세히 들여다보면, 대부분 얼굴 표정 자체를 정교하게 표현하려고 노력했다기 보다는 다른 표현 수단을 추가하여 얼굴 표정을 보완하려는 측면이 많았다. 그래서 다른 입력 감정에 대해서뿐만 아니라 같은 입력 감정이 들어오더라도 로봇의 내/외부 상황에 따라 다양한 표정들이 생성 될 필요가 있다. 본 논문에서는 이러한 문제를 *dynamics*의 변화와 애니메이션 원리, *facial actions*을 이용하여 해결하려고 한다.

III. 사전 지식

먼저, 본 논문에서 새롭게 제안하는 방법의 기초가 되는 LAESM (Linear Affect-Expression Space Model)에 대해서 간략히 설명하고자 한다[14]. 그림 1은 LAESM을 이용하여 로봇의 얼굴 표정을 생성하는 과정을 보여준다. LAESM은 개발된 얼굴 로봇의 선형 독립인 표정들로 선형 표정 공간을 정의한 후, 선형 표정 공간의 기저들에 대응하는 감정 중에서 몇 가지 감정을 선택하여 선형 정서 공간의 기저로 사용한다. 그리고 나서 선형 정서 공간과 표정 공간 사이의 변환 관계가 식 (1)과 같이 만들어진다.

$$\begin{aligned} \mathbf{p}_{(D_p \times 1)} &= T_{2(D_p \times D_p)} T_{1(D_p \times D_e)} \mathbf{e}_{(D_e \times 1)}, \quad D_p \geq D_e \\ \mathbf{r}_{(D_p \times 1)} &= f(\mathbf{p}_{(D_p \times 1)}) \end{aligned} \quad (1)$$

여기서 벡터 \mathbf{e} 는 선형 정서 공간에서 감정의 위치를 나타내고, 벡터 \mathbf{p} 는 선형 표정 공간에서의 표정을 의미하며 전이 행렬 T_1 에 의해 정서 공간의 감정이 표정 공간으로 이동할 수 있다. 좌표 변환 행렬 T_2 는 정서 공간의 감정이 표정으로 변환될 때 발생하는 에너지를 줄이기 위해 표정을 재배치하는 역할을 수행한다. 실제 로봇에서의 표정은 \mathbf{p} 이며 함수 $f(\cdot)$ 에 의해 벡터 \mathbf{p} 가 실제 모터의 위치 값으로 변환된다. D_e 와 D_p 는 각각 선형 정서 공간과 표정 공간의 차원이다.

선형 정서-표정 공간의 관계를 이용하면 로봇의 표정에

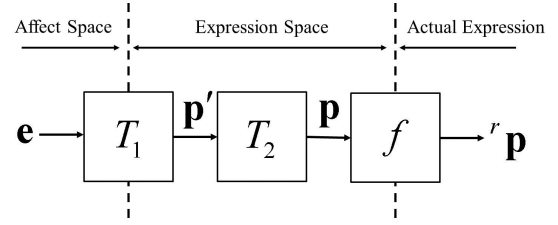


그림 1. Linear Affect-Expression Space Model [14].

Fig. 1. Linear Affect-Expression Space Model [14].

해당하는 감정의 정확한 좌표 값을 선형 정서 공간에서 결정할 수 있다. 또한, 정서-표정 공간의 관계를 결정하는 과정에서 정서 공간의 차원이 표정 공간의 차원보다 작기 때문에 모든 구동기를 직접 제어하지 않고 정서 공간의 기저들을 이용해서 효율적으로 로봇의 표정을 만들 수 있다. 마지막으로, LAESM은 Ekman이 주장한 기본 감정 이론을 따른다[26]. Ekman에 의하면, 사람에게는 구체적인 얼굴 표정과 밀접하게 연관되어 있는 기본 감정들이 존재한다는 것이다. 물론 이 주장이 완벽하게 증명된 것은 아니지만, 신경심리학자들로부터 많은 지지를 받고 있다는 점에서 의미가 있다. 이것은 LAESM에서 표정 공간의 기저인 표정에 해당하는 감정으로 정서 공간의 기저를 택한다는 것과 일맥상통한다.

IV. 동적 감정 모델

본 장에서는 LAESM을 기반으로 동적 감정 모델을 제안한다. 선형 정서 공간의 모든 점(감정)들은 표정 공간으로 옮겨질 수 있다. 그러나 LAESM만 사용되면 자연스러운 표정의 변화를 얻기가 어렵다. 감정이 변하게 되면, 이전 감정과 현재 감정 사이의 움직임 경로가 있어야 표정의 연속적인 변화를 얻을 수 있을 것이다. 그래서 2차의 미분 방정식을 이용하여 서로 다른 감정 사이의 경로를 생성하게 되면 선형 보간법에 비해 쉽게 다양한 경로를 만들 수 있다[9]. 2차 시스템의 두 가지 디자인 변수가 고유 주파수와 감쇠비이기 때문에, 이 두 변수만 조절함으로써 같은 입력 감정에 대해서도 자연스럽게 다양한 표정의 변화를 만들 수 있다. 또한, 동적 감정 모델은 사람의 얼굴 근육 조직이 *mass-spring-damper* 시스템으로 모델링 될 수 있다는 주장과도 부합한다[27,28].

1. 선형 정서 공간에서 동적 감정 모델

동적 감정 모델은 선형 정서 공간에 있는 점들의 움직임을 2차의 미분 방정식을 이용하여 식 (2)와 같이 정의한다[29].

$$\begin{aligned} M \cdot \ddot{\mathbf{e}} + C \cdot \dot{\mathbf{e}} + K \cdot \mathbf{e} &= \mathbf{s} \\ \mathbf{e} &= [e_1 \ e_2 \ \dots \ e_{D_e}]^T, \quad \mathbf{s} = [s_1 \ s_2 \ \dots \ s_{D_e}]^T \end{aligned} \quad (2)$$

벡터 \mathbf{e} 는 선형 정서 공간에 존재하는 현재 감정 벡터이고, \mathbf{s} 는 현재의 감정을 다음 위치로 움직이게 하는 입력 벡터이다. M, C, K 는 각각 *inertia, viscosity, elasticity*에 해당하는 대각 행렬이다. D_e 는 선형 정서 공간의 차원이다.

$$M \square \begin{bmatrix} m_1 & 0 & \dots & 0 \\ 0 & m_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & m_{D_e} \end{bmatrix}, \quad C \square \begin{bmatrix} c_1 & 0 & \dots & 0 \\ 0 & c_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & c_{D_e} \end{bmatrix}, \quad (3)$$

$$K \square \begin{bmatrix} k_1 & 0 & \dots & 0 \\ 0 & k_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & k_{D_e} \end{bmatrix}$$

단위 계단 함수(unit step function)가 식 (2)의 입력 s 로 사용될 경우, 식 (4)의 최종값 정리에 의해 감정 e_i 의 최종값이 1에서 $1/k_i$ 로 줄어들게 된다.

$$\begin{aligned} e_i(\infty) &= \lim_{s \rightarrow 0} s E_i(s) \\ &= \lim_{s \rightarrow 0} s \frac{1}{s m_i s^2 + c_i s + k_i}, \quad i = 1, 2, \dots, D_e \quad (4) \\ &= \frac{1}{k_i} \end{aligned}$$

그러므로 현재 감정을 원하는 목표 위치 $E \cdot e_g$ 로 이동하기 위해 식 (5)와 같이 미리 k_i 를 입력 s 에 곱해야 한다.

$$\begin{aligned} s &= K \cdot E \cdot e_g \\ e_g &= [e_{g_1} \quad e_{g_2} \quad \dots \quad e_{g_{D_g}}]^T, \quad 0 \leq e_{g_i} \leq 1 \quad (5) \\ E &= [\hat{e}_1 \quad \hat{e}_2 \quad \dots \quad \hat{e}_{D_g}] \end{aligned}$$

여기서 e_g 는 감정 평가 혹은 생성 과정에 의해 새롭게 생성된 감정 벡터이고, 그것의 성분들은 각 감정의 종류와 세기 정보를 가지고 있다. D_g 는 감정 생성 과정에서 구분할 수 있는 감정의 개수이다. E 는 예측된 감정 행렬(estimated emotion matrix)이고, 열 벡터 \hat{e}_i 는 정서 공간에서 감정의 위치를 나타낸다. 각 감정의 위치는 해당 감정에 대한 표정 값과 전이 행렬 T_i 의 의사 역행렬을 이용하여 구할 수 있다. 따라서, 현재 외부 입력으로부터 감정 정보를 뽑아낼 수 있다면 이전의 감정 상태와 현재의 감정 상태 사이의 이동 경로를 식 (2)를 통해 구할 수 있고, 식 (1)을 이용하여 실제 로봇의 얼굴 표정으로 만들 수 있다.

2. 2차 시스템의 계수 결정

2차 시스템의 계수를 어떻게 결정할 것인가는 매우 중요한 문제이다. 본 논문에서는 기본 9가지의 경우를 상승 시간(rise time)과 감쇠비(damping ratio)에 따라 정의하였고, 이 두 가지 변수는 사람 얼굴 표정의 변화 특성과 애니메이션의 과장 기법을 바탕으로 결정하였다. 하지만 고유 주파수(natural frequency)와 감쇠비가 2차 시스템의 특징을 결정하는 변수이므로 상승 시간과 감쇠비를 이용해서 고유 주파수를 구해야만 한다. 불행히도 고유 주파수는 식 (6)에서 알 수 있듯이 상승 시간과 감쇠비를 이용해서 정확히 계산할 수 없다[30]. 그래서 본 논문에서는 식 (6)을 이용하여 어렵잡아 구한 값을 바탕으로 2차 시스템의 응답특성을 구한 후 주어진 상승 시간에 부합하는 고유 주파수를 찾았다.

$$\begin{aligned} T_r &\approx \frac{2.16 \times \zeta + 0.6}{\omega_n} \\ \rightarrow \omega_n &\approx \frac{2.16 \times \zeta + 0.6}{T_r}, \quad 0.3 \leq \zeta \leq 0.8 \quad (6) \end{aligned}$$

계수 행렬 M, C, K 는 감정이 정서 공간에서 균일하게 분포한

표 1. 9가지 dynamics에 대한 계수 값.

Table 1. Coefficients of the second order system for nine dynamics.

$T_r \backslash \zeta$	1.0	0.3	0.1
0.125	(dynamics 1) $\omega_n = 35$ $C = 70$ $K = 1225$	(dynamics 4) $\omega_n = 15$ $C = 9$ $K = 225$	(dynamics 7) $\omega_n = 14$ $C = 2.80$ $K = 196$
0.5	(dynamics 2) $\omega_n = 7.80$ $C = 15.60$ $K = 60.84$	(dynamics 5) $\omega_n = 3.90$ $C = 3.12$ $K = 15.21$	(dynamics 8) $\omega_n = 3.40$ $C = 1.36$ $K = 11.56$
2.5	(dynamics 3) $\omega_n = 1.60$ $C = 3.20$ $K = 2.56$	(dynamics 6) $\omega_n = 0.75$ $C = 0.45$ $K = 0.56$	(dynamics 9) $\omega_n = 0.63$ $C = 0.126$ $K = 0.40$

질량을 가진 것($m_i = 1$)으로 가정하고 식 (7)을 이용하여 결정하였다.

$$\begin{aligned} \omega_{ni} &= \sqrt{\frac{k_i}{m_i}} \rightarrow k_i = \omega_{ni}^2 \\ \zeta_i &= \frac{c_i}{2\sqrt{m_i \cdot k_i}} \rightarrow c_i = 2 \cdot \zeta_i \cdot \sqrt{k_i}, \quad i = 1, 2, \dots, D_e \quad (7) \end{aligned}$$

위 과정을 통해 결정된 계수의 값은 표 1에 정리하였다. 상승 시간은 사람의 표정 변화를 고려하여 3가지로 나누었다 [31]. 먼저 미세 표정(micro-expression)은 사람의 표정 중에서 가장 빠르게 변하는 것이며, 0.25초 이내에 나타났다가 사라진다. 이것을 바탕으로 가장 빠른 상승 시간을 0.125초로 가정하였다. 사람의 대부분의 표정은 감정의 종류에 관계없이 1초 이내에 나타났다가 사라진다고 알려져 있다. 그래서 두 번째로 빠른 상승 시간을 0.5초로 정하였다. 마지막으로, 10초 이상 지속되는 사람의 표정은 거짓이며 5초 이상 지속되는 표정의 대부분도 거짓일 가능성이 높다는 사실에 의해 가장 느린 경우의 상승 시간을 2.5초로 가정하였다.

감쇠비를 결정하기 위해서 먼저 사람의 일반적인 표정의 과도응답(transient response)이 임계 감쇠(critical damping) 시스템이라 가정하였다. 이 때, 감쇠비는 1로 설정할 수 있다. 다음으로, 애니메이션에 나오는 캐릭터처럼 로봇의 얼굴 표정도 과장함으로써 강조의 효과를 얻기 위해 감쇠비가 0.3일 때와 0.1인 경우를 추가하였다[32]. 위와 같이 상승 시간과 감쇠비를 각각 3가지 경우로 나눈 것은 각각의 경우가 잘 구분되는 것과 로봇에서 구현할 때 효율성의 균형을 유지하기 위함이다.

위의 표에서 ω_n 은 고유 주파수이고 C, K 는 각각 대각성분이 모두 같은 대각 행렬이다.

V. FACIALACTIONS 과 애니메이션 원리에 기반한 얼굴 제스처 생성

1. Facial Actions

사람의 얼굴 표정에는 수천 가지의 종류가 있다. 그 중에서 대부분은 감정과 직접적으로 관련이 없는 것이며 communication signals, facial symbols, facial actions으로 구분할 수 있다[31].

- communication signals: 문법의 보충(물음표, 느낌표 등)이나 의사소통에 도움을 주기 위해 혹은 어떤 것을 설명하기 위한 얼굴의 움직임
- facial symbols: 윈크와 같은 것
- facial actions: 입술 깨물기, 볼에 바람 넣기, 흘쩍임 등

이 중에서 facial actions은 사람의 감정 상태를 반영하는 의식적이거나 무의식적인 움직임에 의해 나타난다. 일반적으로 사람들은 당황하거나 초조할 때 입술을 깨물곤 한다는 것으로부터 감정 상태를 드러낼 때 얼굴 표정뿐만 아니라 상황에 따라서 facial actions과 같이 다른 요소들도 사용한다는 것이다. 이처럼 사람에게서 관찰할 수 있는 것을 로봇에도 적용한다면 동적 감정 모델만으로는 만들 수 없었던 다양한 로봇의 표정들을 만들 수 있을 것이다. 본 논문에서는 흘쩍임, 대성통곡하는 듯한 표정과 흐뭇한 미소, 박장대소 등을 구현하였다.

2. 애니메이션 원리

애니메이션에 등장하는 캐릭터들의 움직임은 주로 디즈니의 12 가지 원리를 바탕으로 만들어진다[32]. 12 가지 원리 중에서 본 논문에서는 ‘secondary actions’ 과 ‘exaggeration’ 을 다룰 것이다. ‘secondary actions’ 는 어떤 사물이 행동을 취할 때 본래의 움직임에 수반하여 일어나는 움직임을 말한다. 이는 캐릭터가 의도하는 표현에 직접 영향을 주지 않지만 신뢰성을 증가시키는데 도움이 된다. 눈 깜빡임이나 호흡 과정에 일어나는 sinusoidal motion이 ‘secondary actions’ 의 좋은 예이다. ‘exaggeration’ 은 본래의 움직임이 내포하고 있는 의미를 강조하는 역할을 하며 캐릭터의 행동에 확실성을 부여하기 위해 주로 사용된다. 본 논문에서는 4장에서 동적 특성을 결정할 때도 고려되었고, 앞서 언급한 facial actions 도 ‘exaggeration’ 의 일부라고 볼 수 있다.

3. Facial actions 와 애니메이션 원리를 이용한 얼굴 제스처 생성

사람들은 로봇이 살아있는 생명체와 같다고 느낄 때 조금 더 로봇에 관심을 가지고 로봇과 교류하고자 할 가능성이 높을 것이다. 그래서 로봇의 디자인뿐만 아니라 로봇의 행동 표현 또한 사람과 유사할 필요가 있다. 이를 위해 본 장에서는 facial actions 과 애니메이션 원리를 적용하여 사람과 유사한 로봇의 얼굴 표정을 만들고자 한다.

그림 2는 제안하는 방법의 전체 구성도를 나타낸다. 동감정 모델에서는 2차의 미분 방정식을 사용하여 감정의 다양한 감정의 움직임을 만들어내고, LAESM은 감정의 움직임들을 표정으로 변환한다. Facial action generator $g_{fa}(\cdot)$ 와 secondary action generator $g_{sa}(\cdot)$ 는 얼굴 제스처를 생성하기 위해 추가 되었다. 이 두 블록은 사용자의 편의에 따라 식 (8)과 같이 활성화시킬 수 있다.

$$\begin{aligned} \bar{\mathbf{p}}(t) &= \mathbf{p}(t) + w_{fa} \cdot g_{fa}(\mathbf{p}(t)) + w_{sa} \cdot g_{sa}(\mathbf{p}(t), g_{fa}(\mathbf{p}(t))) \\ w_{fa}, w_{sa} &= 0 \text{ or } 1 \end{aligned} \quad (8)$$

Facial actions 은 각 구동기의 목표 위치를 원래 위치에서 표정의 경계 내에 위치한 다른 점으로 이동함으로써 구현하였다. 이 때, 목표 위치를 그림 3과 같이 원래의 위치와 수평선

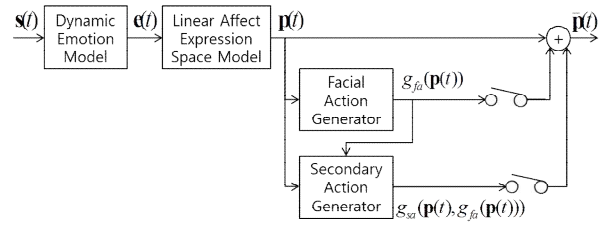


그림 2. 얼굴 제스처 생성을 위한 과정.
Fig. 2. Procedure of generating facial gestures.

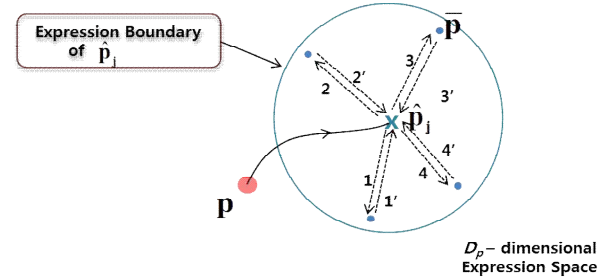


그림 3. Facial actions 의 적용.
Fig. 3. Implementation of facial actions.

위치를 주기적으로 바꿈으로써 다양한 얼굴 제스처를 만들 수 있다.

$$g_{fa}(\mathbf{p}(t)) = \begin{cases} 0, & n(T_o + T_m) \leq t < n(T_o + T_m) + T_o \\ \hat{\mathbf{p}}_j + b_j \cdot |\text{rand}(\cdot)| \cdot \frac{\mathbf{d}}{\|\mathbf{d}\|} - \mathbf{p}(t), & n(T_o + T_m) + T_o \leq t < (n+1)(T_o + T_m) \end{cases}$$

$$\mathbf{d} = [\text{rand}(\cdot)_{i,1}]_{D_p \times 1}, \quad -1 \leq \text{rand}(\cdot), \text{rand}(\cdot)_{i,1} \leq 1, \quad n = 0, 1, 2, \dots \quad (9)$$

T_o : duration of manifesting original expression

T_m : duration of manifesting modified expression

$\hat{\mathbf{p}}_j$: center of expression boundary corresponding to j -th target emotion $\hat{\mathbf{e}}_j$

b_j : radius of D_p -dimensional hyperspherical expression boundary for $\hat{\mathbf{p}}_j$

식 (9) 에서 $b_j \cdot |\text{rand}(\cdot)|$ 와 $\mathbf{d} / \|\mathbf{d}\|$ 는 각각 표정 경계의 중심에 있는 표정에서부터 떨어진 거리와 방향을 의미한다. 만약, 임의로 선택된 표정의 크기가 중심 표정에서 너무 멀리 떨어져 있으면 실제의 표정과 달라서 인식하기가 힘들 것이고 최악의 경우는 원래의 표정과 다른 표정으로 오 인식될 수도 있기 때문에 표정의 경계를 사용하여 원래 표정의 왜곡을 막기 위해 사용한다. 표정 경계에 대한 자세한 내용은 [33]에 설명되어 있다. T_o 는 원래의 표정을 지속하는 시간을 의미하고 T_m 은 표정 경계내의 임의의 표정을 유지하는 시간이다. $\hat{\mathbf{p}}_j$ 는 j 번째 감정에 해당하는 표정이며 표정 경계의 중심이다. b_j 는 j 번째 표정 경계의 반지름이다.

식 (9)를 사용하여 모든 구동기의 목표 위치를 원래의 목표 위치와 표정 경계 내에서 임의로 선택된 위치와 교대로 설정하여 흐느낌과 같은 의미 있는 얼굴 제스처를 만들 수 있다. 이 때, T_o 와 T_m 을 어떻게 결정하느냐에 따라 여러 가지 얼굴

제스처를 만들 수 있다. 슬픔과 관련된 사람의 호스김을 로봇에 구현하기 위해서는 T_o 을 수십 ms 로 설정하고 T_m 을 수 초 정도로 설정하면 된다. 반면, 같은 슬픔과 관련된 울부짖는 듯한 표정을 구현하기 위해서는 T_o 와 T_m 둘 다 수십 ms 로 설정하면 된다. 본 논문에서는 위의 가이드 라인을 참고하여 임의의 범위 내에서 T_o 와 T_m 을 매번 다르게 다르게 정하였다.

‘secondary actions’ 으로 눈 깜빡임 $g_b(\cdot)$ 과 호흡에 의한 sinusoidal motion $g_s(\cdot)$ 을 구현하였다. 식 (10)에서와 같이 두 기능은 사용자가 선택할 수 있도록 하였다.

$$g_{sa}(\mathbf{p}(t)) = w_b \cdot g_b(\mathbf{p}(t)) + g_s(\mathbf{p}(t)) + w_s \cdot g_s(\mathbf{p}(t))$$

$$w_b, w_s = 0 \text{ or } 1 \tag{10}$$

$$g_s(\mathbf{p}(t)) = \|\mathbf{p}(t)\| \cdot k \sin(\omega t), \quad k \propto \|\mathbf{p}(t)\|$$

눈 깜빡임의 경우, 로봇의 현재 표정과 상관없이 눈꺼풀에 해당하는 구동기만 직접 제어하면 된다. 눈 깜빡임의 주기는 약 4초로 정하였는데, 보통 성인의 눈 깜빡임 회수가 분당 14~17회로 알려져 있기 때문이다[34].

호흡에 의한 sinusoidal motion을 구현하기 위해서 사인함수를 사용하였다. 사인함수의 진폭은 현재 표정의 세기에 비례하며 현재 표정에서부터 아주 작은 변화를 주도록 k 값을 정하였다. 사인 함수의 주기는 약 3초로 정하였으며, 이는 건강한 성인의 경우 대략 1분에 18회 정도 호흡하기 때문이다. 더 자연스러운 ‘secondary actions’ 을 위해 주기는 하나의 값으로 고정하지 않고 위에서 정해진 주기를 바탕으로 일정 구간 내에서 다음 주기를 정하도록 하였다. 이로 인해, 때에 따라 빠르게 혹은 느리게 눈을 깜빡이게 된다. Sinusoidal motion도 마찬가지로이다.

VI. 실험 결과

1. 실험 환경

제안한 방법을 검증하기 위해 그림 4와 같이 새로운 얼굴 로봇을 제작하였다. 총 12 자유도를 가지며, 눈썹 4 자유도, 눈꺼풀 6 자유도, 입 2 자유도로 구성된다. 구동기는 로보티즈사의 다이내믹셀 RX-64, RX-24F, DX-117 세 가지 타입을 사용하였다.

2. Dynamics에 따른 표정 변화

앞서 VI 장에서 9가지의 기본 dynamics를 정하였다. 그림 5는 기쁨 표정에 대한 동적 감정 모델의 적용 결과이다. 그림 5(a)와 (b)의 경우 감쇠비는 같고 상승 시간이 다르기 때문에

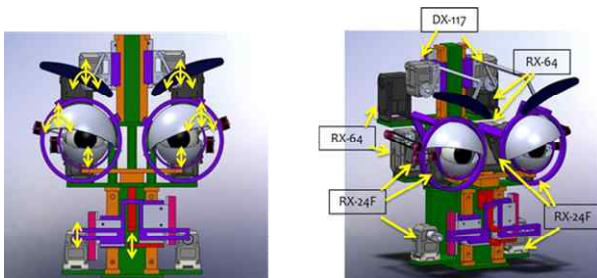


그림 4. 실험에 사용된 얼굴 로봇.
Fig. 4. Facial Robot for experiments.

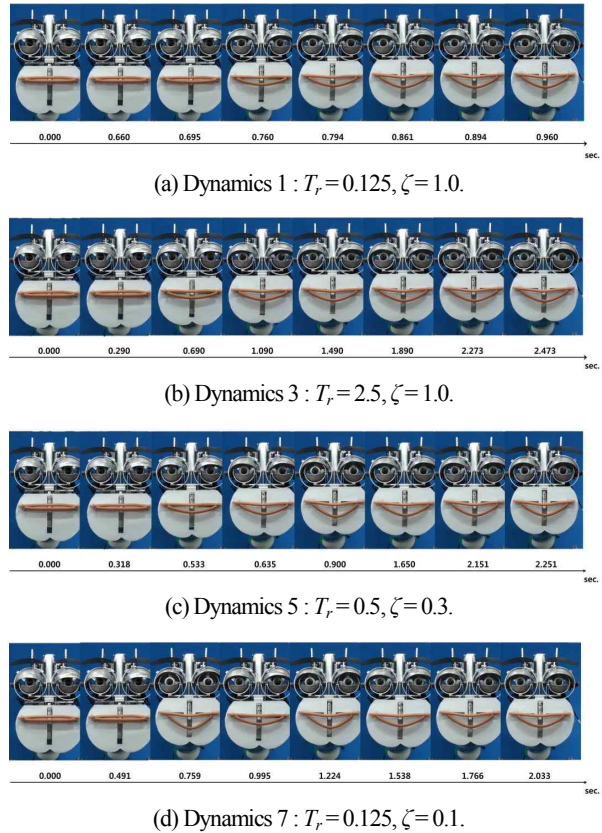


그림 5. 기쁨 표정에서 dynamics 변화에 따른 표정의 변화.
Fig. 5. Changes in happiness expressions according to the dynamics.

(a)의 경우 기쁨 표정이 빨리 드러나지만 (b)의 경우가 천천히 나타남을 확인할 수 있다. 그림 5(a)와 (d)를 비교해보면, 감쇠비의 차이로 인해 (a)는 overshoot 없이 그대로 목표 표정에 도달하지만 (d)는 overshoot 이 발생하여 목표 표정으로 향하다가 원래 표정으로 돌아가는 듯한 표정이 나타난다. 그림 5(c)의 경우에는 overshoot 이 발생하지만 (d)에 비해서는 표정이 천천히 나타나고 overshoot 의 크기도 작은 특성을 보인다. 즉, 상승 시간과 감쇠비를 조절하여 같은 입력 감정에 대해서도 다양한 표정의 변화를 만들 수 있다.

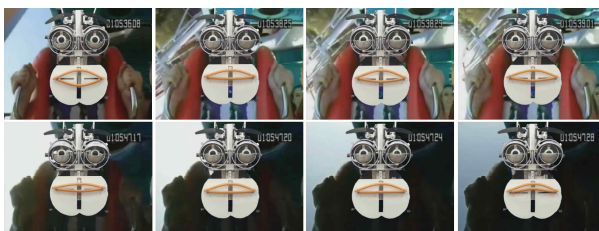
3. 설문 조사 결과

대부분의 이전 연구에서는 Ekman이 제안한 기본 6가지 표정 혹은 몇 개의 추가된 표정만 고려되었고, 정적인 표정만을 다루었다. 본 장에서는 생동감 있는 로봇의 얼굴 표정 생성을 위해 제안한 방법을 평가하기 위해 설문조사를 수행하였다.

먼저 설문조사에 앞서 합성된 영상을 만들었다. 다섯 종류의 영상이며, 한 가지 종류의 영상에 대해 이전 방법 (linear affect expressions space model 만 사용)과 제안한 방법 (dynamic emotion model, secondary actions, facial actions 추가)으로 만든 로봇의 표정 영상 두 개씩을 제작하여, 10개의 영상으로 구성되어 있다. 본 논문에서 제안한 방법으로 만든 표정의 경우, 사용된 dynamics 종류와 facial action 구현에 필요한 파라미터 값은 표 2에 요약하였다. 제안한 방법이 다양한 환경에서 그에 적합한 로봇의 표정을 생성하는데 효과적인지 검증하기 위해 다음 다섯 가지 상황을 가정하였다. 영상 1은 로봇이

표 2. 각 영상에 사용된 dynamics 종류 및 parameter 값.
Table 2. Types of the dynamics and parameter values for each video.

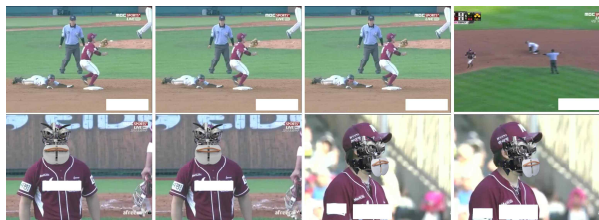
	Dynamics	Parameters for Facial Actions
Video 1	Dynamics1	T_m : 40ms ~ 160ms T_o : 29ms ~ 73ms
Video 2	Dynamics1 Dynamics2 Dynamics7	T_m : 40ms ~ 160ms T_o : 29ms ~ 73ms
Video 3	Dynamics2	T_m : 40ms ~ 160ms T_o : 29ms ~ 73ms T_o : 250ms ~ 625ms
Video 4	Dynamics2 Dynamics3 Dynamics8	T_m : 40ms ~ 160ms T_o : 29ms ~ 73ms T_o : 500ms ~ 1250ms
Video 5	Dynamics1 Dynamics2	T_m : 40ms ~ 160ms T_o : 29ms ~ 73ms T_o : 250ms ~ 625ms



(a) Video 1.



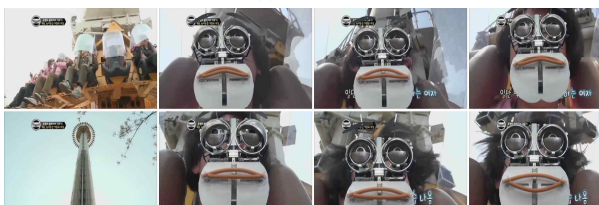
(b) Video 2.



(c) Video 3.



(d) Video 4.



(e) Video 5.

그림 6. 설문조사에 사용된 5가지 종류의 대표 영상.
Fig. 6. Representative images of videos.

플러코스터를 타는 것이고, 영상 2는 로봇이 우스꽝스러운 비디오를 보는 장면이다. 영상 3은 야구 경기에서 잘못 된 심판 판정에 대해 로봇이 항의 하는 장면이고, 영상 4는 로봇이 오디션에서 심사위원으로써 참가자를 평가하는 장면이다. 마지막으로 영상 5는 로봇이 자이로드롭을 타는 상황으로 가정하였다.

다섯 종류의 영상으로 나누게 된 이유는 설문 조사에 참여한 대상자들이 빠른 시간 내에 로봇이 왜 그런 표정을 짓는지를 쉽게 인지하도록 하기 위해서다. 그리고 상황에 따라 나타나게 되는 감정 표현의 종류 또한 다르기 때문에 여러 가지 감정을 다루기 위해서 다섯 종류의 영상으로 나누어 실험하였다. 영상 1에서는 주로 공포의 표정이 많이 나타나며 기쁨과 놀람의 표정도 조금 포함되어 있다. 영상 2에서는 기쁨의 표정이 많은 부분을 차지하고 놀람, 놀람과 슬픔이 섞인 표정도 있다. 영상 3은 화난 표정으로 구성되어 있다. 영상 4의 처음 부분은 슬픔, 놀람, 놀람과 기쁨이 섞인 표정이 나오며, 후반 부분에는 슬픔, 슬픔과 기쁨이 섞인 표정이 나온다. 영상 5는 두려움과 공포 및 두 감정이 섞인 표정으로 구성하였다. 그림 6은 설문조사에 사용된 영상의 일부 장면을 보여준다.

10개의 합성된 영상을 설문조사 대상자들에게 임의의 순서대로 보여주었다. 제작된 로봇을 본 적이 없는 참가자 20명을 대상으로 하였으며 20~30대 남녀(남자 15명, 여자 5명)로 구성되어 있다. 영상을 본 후, 참가자들은 “로봇의 얼굴 표정이 사실적으로 잘 표현되었는가?”에 대한 질문에 5-point Likert Scale (1 = strongly disagree, 2 = disagree, 3 = not sure, 4 = agree, 5 = strongly agree)로 구성된 설문지에 답을 하도록 하였다. 그림 7은 설문조사 결과를 정리한 것이다.

각 영상은 아래 웹사이트에서 확인할 수 있다.

영상 1-1 : <http://youtu.be/5SRMPxA-uVM>

영상 1-2 : <http://youtu.be/Havqknfiaeo>

영상 2-1 : <http://youtu.be/ZWyllqnSmbS>

영상 2-2 : <http://youtu.be/Ufjb7HMajxY>

영상 3-1 : <http://youtu.be/roGK8aZEsyk>

영상 3-2 : <http://youtu.be/FQJWE8Skp5U>

영상 4-1 : http://youtu.be/iYyu_L3HpGg

영상 4-2 : http://youtu.be/yb_Oli3SU9o

영상 5-1 : <http://youtu.be/KXgH3ftNDpA>

영상 5-2 : http://youtu.be/D0U_xA19Smc

모든 영상에 대해서 제안한 방법을 이용하였을 때 로봇의 표정이 더 사실적으로 표현되었음(평균 값이 증가)을 알 수 있다. 이 결과가 의미 있는지를 알아보기 위해 t-test 를 수행하였고, 전체 영상에 대한 평균 값의 p-value 가 0.001보다 작음을 확인하였다. 각각의 영상에 대해서도 영상 1과 영상 5에 대해서는 p-value 가 0.001 보다 작고, 영상 2와 영상 4에 경우는 p-value 가 0.05 보다 작았다. 하지만, 영상 3의 p-value 는 0.05보다 컸다. 영상 3의 경우, 화난 표정 한 가지만을 포함하고, 기존 방법으로도 충분히 로봇이 처한 상황을 잘 표현할 수 있기에 유의미한 차이가 나지 않은 것이라 판단된다. 따라서 모든 경우에 대해 제안된 방법이 매우 효과적이라고

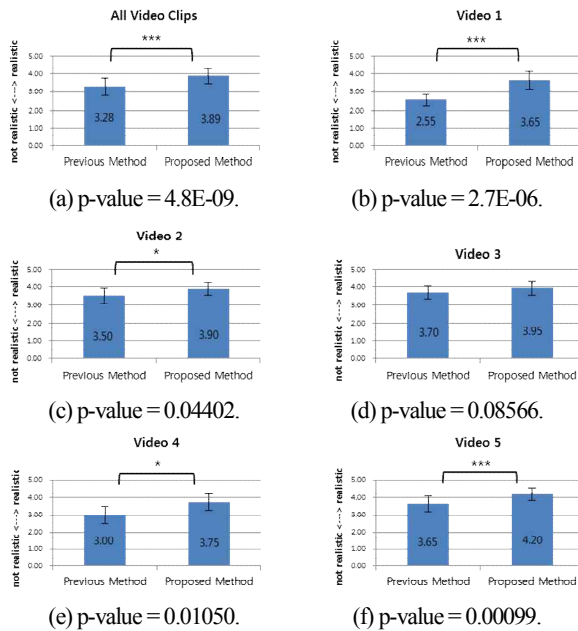


그림 7. 설문조사 결과.
Fig. 7. Results of the survey.

할 순 없지만, 대체적으로 본 논문에서 제안한 방법을 사용하였을 때 더 사실적으로 로봇의 상황에 적절한 표정을 만들 수 있다.

VII. 결론

본 논문에서는 같은 입력 감정에 대해서도 다양하고 더 사람과 유사한 로봇의 표정을 만드는 방법에 대해 제안하였다. Dynamics 의 변화를 이용하여 같은 목표 위치를 가지는 표정에 대해서도 다양한 형태로 표현하였으며, 눈 깜빡임과 호흡에 따른 sinusoidal motion과 같은 ‘secondary actions’ 을 추가하여 로봇에 생동감을 부여하였다. 그리고 facial action 을 이용하여 사람의 흐느낌이나 우는 표정, 박장대소 하는 것과 같이 기존의 로봇에서 볼 수 없었던 표정들을 구현하였다. 제안된 방법은 설문조사를 통하여 여러 상황에서 효과적임을 검증하였다. 그러므로 본 논문에서 제안 된 방법이 로봇의 감정 생성이나 인식 모델과 통합 된다면 매우 유용할 것이라 예상된다.

현재는 로봇이 다양한 상황을 스스로 인지하지 못하기 때문에, 추후에 상황에 적절한 dynamics를 결정하는 것과 facial actions 구현에 필요한 파라미터들을 로봇의 내/외부 환경을 고려하여 자동으로 결정하는 방법에 대한 연구도 필요하다.

REFERENCES

[1] P. Salvine, M. Nicolescu, and H. Ishiguro, “Benefits of human-robot interactions,” *IEEE Robotics & Automation Magazine*, vol. 18, no. 4, pp. 98-99, 2011.
 [2] J.-C. Martin, R. Niewiadomski, L. Devillers, S. Buisine, and C. Pelachaud, “Multimodal complex emotions: gesture expressivity and blended facial expressions,” *International Journal of Humanoid Robotics*, vo. 3, no. 3, pp. 269-291, 2006.
 [3] Z. Liu and Z. G. Pan, “An emotion model of 3D virtual characters in intelligent virtual environment,” *International Con*

f. Affective Computing and Intelligent Interaction (ACII 2005), pp. 629-636, 2005.
 [4] C. Becker-Asano and I. Wachsmuth, “Affective computing with primary and secondary emotions in a virtual human,” *Autonomous Agents and Multi-agent Systems*, vol. 20, no. 1, pp. 32-49, 2010.
 [5] E. Tanguy, P. J. Willis, and J. J. Bryson, “A dynamic emotion representation model within a facial animation system,” *International Journal of Humanoid Robotics*, vol. 3, no. 3, pp. 293-300, 2006.
 [6] T. Fukuda, M.-J. Jung, M. Nakashima, F. Arai, and Y. Hasegawa, “Facial expressive robotics head system for human-robot communication and its application in home environment,” *Proceedings of the IEEE*, vol. 92, no. 11, pp. 1851-1865, 2004.
 [7] P. Ekman, W. V. Friesen, and J. C. Hager, *Facial Action Coding System/The manual on CD Rom*, Research Nexus division of network information research corporation, Salt Lake City, UT, 2002.
 [8] C. L. Breazeal, *Designing Sociable Robots*, The MIT Press, Cambridge, Massachusetts, 2002.
 [9] H. Miwa, T. Okuchi, K. Itoh, H. Takanobu, and A. Takanishi, “A new mental model for humanoid robots for human friendly communication - Introduction of learning system, mood vector and second order equations of emotion -,” *IEEE Int. Conf. Robotics Automation (ICRA)*, pp. 3588-3593, 2003.
 [10] S. Sosnowski, A. Bittermann, K. Kuhlentz, and M. Buss, “Design and evaluation of emotion-display EDDIE,” *Int. Conf. Intelligent Robots and Systems (IROS)*, pp. 3113-3118, 2006.
 [11] K.-G. Oh, M.-S. Jang, and S.-J. Kim, “Automatic emotional expression of a face robot by using a reactive behavior decision model,” *Journal of Mechanical Science and Technology*, vol. 24, no. 3, pp. 769-774, 2010.
 [12] B.-S. Yoo, S.-H. Cho, and J.-H. Kim, “Fuzzy integral-based composite facial expression generation for a robotic head,” *IEEE Int. Conf. Fuzzy Systems (FUZZ)*, pp. 917-923, 2011.
 [13] A. J. N. van Breemen, “Animation engine for believable interactive user-interface robots,” *IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, pp. 2873-2878, 2004.
 [14] H. S. Lee, J. W. Park, and M. J. Chung, “A linear affect-expression space model and control points for mascot-type facial robots,” *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 863-873, 2007.
 [15] Y. Matsui, M. Kanoh, S. Kato, T. Nakamura, and H. Itoh, “A model for generating facial expressions using virtual emotion based on simple recurrent network,” *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 14, no. 5, pp. 453-463, 2010.
 [16] T. Riberio and A. Paiva, “The illusion of robotic life: Principles and practice of animation for robots,” *ACM/IEEE Int. Conf. Human-Robot Interaction (HRI)*, pp. 383-390, 2012.
 [17] T. Minato, M. Shimada, H. Ishiguro, and S. Itakura, “Development of an android robot for studying human-robot interaction,” *Int. Conf. Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (IEA/AIE)*, pp. 424-434, 2004.
 [18] S. Hitramatsu, T. Tsuji, and H. Kobayashi, “Development of the face robot SAYA for rich facial expressions,” *Int. Joint Conf. SICE-ICASE*, pp. 5423-5428, 2006.

- [19] J.-H. Oh, D. Hanson, W.-S. Kim, I. Y. Han, J.-Y. Kim, and I.-W. Park, "Design of android type humanoid robot Albert HUBO," *IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, pp. 1423-1433, 2006.
- [20] D.-W. Lee, T.-G. Lee, B. R. So, M. Choi, E.-C. Shin, K. W. Yang, M.-H. Baek, H.-S. Kim, and H.-G. Lee, "Development of an android for emotional expression and human Interaction," *Proc. of Int. Federation of Automatic Control*, pp. 4336-4337, 2008.
- [21] H.-R. Kim and D.-S. Kwon, "Computational model of emotion generation for human-robot interaction based on the cognitive appraisal theory," *Journal of Intelligent & Robotic Systems*, vol. 60, no. 2, pp. 263-283, 2010.
- [22] M. Mori, K. F. MacDorman, and N. Kageki, "The uncanny valley," *IEEE Robotics & Automation Magazine*, vol. 19, no. 2, pp. 98-100, 2012.
- [23] N. Kageki, "An uncanny mind," *IEEE Robotics & Automation Magazine*, vol. 19, no. 2, pp. 112-108, 2012.
- [24] M. Zecca, N. Endo, S. Momoki, K. Itoh, and A. Takamishi, "Design of the humanoid robot KOBIAN - preliminary analysis of facial and whole body emotion expression capabilities-," *IEEE-RAS Int. Conf. Humanoid Robots*, pp. 487-492, 2008.
- [25] K.-G. Oh, C.-Y. Jung, M.-T. Choi, and S.-J. Kim, "Eye motion generation in a mobile service robot 'SILBOT II'," *IEEE Int. Conf. Advanced Robotics and its Social Impacts (ARSO)*, pp. 59-64, 2010.
- [26] P. Ekman, "An argument for basic emotions," *Cognition & Emotion*, vol. 6, no. 3-4, pp. 169-200, 1992.
- [27] K. Waters, and D. Terzopoulos, "A physical model of facial tissue and muscle articulation," *IEEE Int. Conf. Visualization in Biomedical Computing*, pp. 77-82, 1990.
- [28] D. Terzopoulos, K. Waters, "Physically-based facial modeling, analysis, and animation," *Journal of Visualization and Computer Animation*, vol. 1, no. 2, pp. 73-80, 1990.
- [29] H. S. Lee, J. W. Park, S. H. Jo, and M. J. Chung "A linear dynamic affect-expression model: facial expressions according to perceived emotions in mascot-type facial robots," *IEEE Int. Symposium on Robot and Human Interactive Communication (ROMAN)*, pp. 619-624, 2007.
- [30] R. C. Dorf and R. H. Bishop, *Modern Control Systems*, 9th Ed., Prentice Hall, Upper Saddle River, New Jersey, 2001.
- [31] J. Lasseter, "Principle of traditional animation applied to 3D computer animation," *Computer Graphics*, vol. 21, no. 4, pp. 35-44, 1987.
- [32] P. Ekman and Y. I. Yun, *Telling Lies*, (Korean Translation) Dongin Publishing, Seoul, Korea, 1997.
- [33] J. W. Park, H. S. Lee, S. H. Jo, M. G. Kim, and M. J. Chung, "Emotional boundaries for choosing modalities according to the intensity of emotion in a linear affect-expression space," *IEEE Int. Symposium on Robot and Human Interactive Communication (ROMAN)*, pp. 225-230, 2008.
- [34] J. S. Elston, F. C. Granje, and A. J. Lees, "The relationship between eye-winking tics, frequent eye-blinking and blepharospasm," *Journal of Neurology, Neurosurgery, and Psychiatry*, vol. 52, no. 4, pp. 477-480, 1989.
- [35] I. Leite, C. Marinoh, and A. Paiva, "Social robots for long-term interaction: A survey," *International Journal of Social Robot*, vol. 5, no. 2, pp. 291-308, 2013.



박정우

2005년 경북대학교 전자전기공학부 학사. 2007년 KAIST 전기및전자공학과 석사. 2007년~현재 KAIST 전기및전자공학과 박사과정 재학중. 관심분야는 HRI, Facial Robot, Machine Learning.



김우현

2007년 KAIST 전기및전자공학과 학사. 2009년 KAIST 전기및전자공학과 석사. 2009년~현재 KAIST 전기및전자공학과 박사과정 재학중. 관심분야는 HRI, artificial emotion, robot expression, and human gesture recognition.



이원형

2008년 KAIST 전기및전자공학과 학사. 2010년 KAIST 전기및전자공학과 석사. 2010년~현재 KAIST 전기및전자공학과 박사과정 재학중. 관심분야는 HRI, Robot's Emotion and Expressions, and Its Applications.



이희승

2000년 KAIST 전기및전자공학 공학 학사. 2002년 KAIST 전기및전자공학 공학 석사. 2008년 KAIST 전기및전자공학 공학박사. 2008년~2010년 삼성전자 책임연구원. 2010년~현재 현대자동차 책임연구원. 관심분야는 HRI/HCI, 임베디드 시스템, 센서융합.



정명진

1973년 서울대학교 공과대학 전기공학과 학사. 1977년 미시간대학교 전기공학과 석사. 1983년 미시간대학교 제어공학과 박사. 1983년~현재 KAIST 전기및전자공학과 교수. 관심분야는 sensor-based robot control and planning, HRI, and service robots for the disabled.