

멀티 모달 감정인식 시스템 기반 상황인식 서비스 추론 기술 개발

Development of Context Awareness and Service Reasoning Technique for Handicapped People

고광은 · 심귀보*

Kwang-Eun Ko and Kwee-Bo Sim*

중앙대학교 전자전기공학부

요 약

사람의 감정은 주관적인 인식 작용으로서 충동적인 성향을 띠고 있으며 무의식중의 사람의 욕구와 의도를 표현하고 있다. 이는 유비쿼터스 컴퓨팅 환경이나 지능형 로봇의 사용자가 처한 환경의 상황정보 중에서 사용자의 의도를 가장 많이 포함하고 있는 정보라고 할 수 있다. 이러한 사용자의 감정을 파악할 수 있는 지표는 사람의 얼굴 영상에서의 표정과 음성신호에서의 Spectrum 통계치 및 생체신호(근전위, 뇌파, 등)등이다. 본 논문에서는 감정인식 활용의 편의와 효율성 향상을 주목적으로 하여 사용자의 얼굴 영상과 음성을 이용한 감정인식에 대하여 개별 결과물만을 산출하고 그 인식률을 검토한다. 또한 임의의 상황에서의 인식률 향상을 위하여 영상과 음성의 특징을 기반으로 최적의 특징 정보들을 구별해 내고, 각각의 개별 감정 특징에 대한 융합을 시도하는 특징 융합 기반의 Multi-Modal 감정인식 기법을 구현한다. 최종적으로 감정인식 결과를 이용하여 유비쿼터스 컴퓨팅 환경에서 발생 가능한 상황 설정 시나리오와 베이지안 네트워크를 통해 유비쿼터스 컴퓨팅 서비스의 확률 추론 가능성을 제시하고자 한다.

Abstract

As a subjective recognition effect, human's emotion has impulsive characteristic and it expresses intentions and needs unconsciously. These are pregnant with information of the context about the ubiquitous computing environment or intelligent robot systems users. Such indicators which can aware the user's emotion are facial image, voice signal, biological signal spectrum and so on. In this paper, we generate the each result of facial and voice emotion recognition by using facial image and voice for the increasing convenience and efficiency of the emotion recognition. Also, we extract the feature which is the best fit information based on image and sound to upgrade emotion recognition rate and implement Multi-Modal Emotion recognition system based on feature fusion. Eventually, we propose the possibility of the ubiquitous computing service reasoning method based on Bayesian Network and ubiquitous context scenario in the ubiquitous computing environment by using result of emotion recognition

Key Words : Multi-Modal Emotion Recognition, Service Inference, Bayesian Network

1. 서 론

대표적인 유비쿼터스 기기로 지능형 로봇, 휴대폰 단말 시스템 등의 하드웨어에 적절한 Killer Application으로서 플랫폼, 소프트웨어 어플리케이션 등의 개발에 대한 수요가 점차로 증가하고 있는 추세이다. 이러한 어플리케이션은 사용자와 유비쿼터스 기기 사이의 상호 작용을 통한 서비스 제공 기술이 기반이 되어야 하며 이러한 상호교류 기술 중의 대표적인 사례 하나가 감정인식이다. 기존의 IT 분야의 산업에서 고객중심으로 서비스를 제공하기 위해서는 그 고객, 즉 사용자의 행동은 물론 감정, 기호 등을 종합적으로

파악하여 맞춤형 서비스를 제공하는 것이 중요하기 때문에 감정인식에 대한 연구의 필요성은 점차 증대되고 있다고 볼 수 있다[1].

인간의 감정은 특정 사물이나 행동에 대한 무의식중의 욕구를 내포하고 있으며 이것이 적절하게 이용되어 사용자가 유비쿼터스 기기와의 관계에서 원하는 서비스에 대한 정보를 잘 활용할 수 있다면 한층 진보된 형태의 서비스 제공이 가능하게 될 것이다. 이는 서비스 기기 스스로 능동적인 추론과 사용자와의 상호 교류가 가능하게 됨을 의미하며 차후 유비쿼터스 컴퓨팅 관련 산업의 지향점이라고 볼 수 있다. 이러한 감정인식과 서비스 추론의 연관성과 가능성을 고찰하기 위해서 본 논문에서는 유비쿼터스 서비스 기기들에게 사용자의 감정을 인식하는 인터페이스를 적용하고 인식결과를 통해 그에 따른 서비스를 추론할 수 있는 기술을 제시하고자 한다.

기존의 감정인식에 대한 연구는 단일 생체신호(음성, 얼굴표정, 제스처, 뇌파, 등)를 이용하는 방향으로 진행되어져 왔다. 그러나 단일신호만으로는 특정 감정에 대한 인식률은

접수일자 : 2008년 11월 1일

완료일자 : 2009년 1월 31일

* 교신 저자

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

높지만 풍부한 감정 상태를 인식하는 것에는 취약하다는 단점이 있다. 따라서 인식률의 향상을 위해서는 다양한 생체 신호 간의 융합을 통한 상호 약점을 보완하는 방식으로 감정인식이 가능하도록 하는 인식기법의 개발이 요구된다.

또한 유비쿼터스 컴퓨팅 환경에서의 서비스 추론과 관련하여 상황인식과 관련된 많은 연구들이 이미 진행되어져 왔으며, 이러한 사례들을 기반으로 본 논문에서는 매우 간단한 상황인지 시나리오를 기반으로 감정인식 결과가 하나의 상황정보(Context)로서 활용되어 서비스 추론에 있어 매우 중요한 매개 정보의 역할을 할 수 있음을 보이고자 한다.

본 논문의 내용은 2장에서 감정인식 및 서비스 추론과 관련된 국내, 외 기술동향에 대하여 논하고, 3장에서 영상, 음성을 개별적으로 이용한 감정인식 기술 구현에 대하여 논한다. 4장에서는 3장 결과물을 활용하여 Multi-Modal 특징 융합 기반 감정인식 기법에 대하여 논하고, 5장에서는 상황인지 시나리오 기반의 서비스 추론 시물레이션을 진행한다. 6장에서는 감정인식 시물레이션 프로그램과 서비스 추론 시물레이션을 통해 제안 기법을 고찰하고, 마지막 7장에서는 향후 과제와 보완점에 대하여 기술하였다.

2. 관련 기술 동향

현재의 감정인식 기술의 수준에서는 기존의 공학적 접근만으로는 도저히 해결할 수 없는 여러 가지 문제점들을 가지고 있다[1]. 감정에 대한 정량적인 지표나 수단이 제시되지 않았으며 인간의 사고, 활동의 재현에 의한 판단, 이해가 불가능하다.

영상과 음성을 기반으로 내포된 감정특징 추출을 위한 연구가 활발히 진행되어지고 있으며, 음성의 경우 Fukuda는 음성신호의 Pitch와 전력 포락선 검출을 통해 20개의 일본어 샘플에 대한 실험을 진행하였고, Silva는 음성신호의 Pitch와 HMM(Hidden Markov Model)을 이용하여 영어/스페인어에 대한 감정인식을 실험하였다[2].

영상을 이용한 방법의 경우, 정적영상과 동적영상을 기반으로 한 방법이 연구 진행 중에 있으며, 그 중 특징 값을 기반으로 표정 인식하는 방법의 경우 얼굴에서 감정정보를 포함하고 있는 주요 특징 부분(눈, 코, 입 등)의 위치를 검색, 모양과 데이터베이스의 기하학적 관계를 비교, 파악함으로써 인식을 수행하는 과정이다. 또 다른 방법으로 Heuristic 분석 기법은 PCA(Principle Component Analysis), LFA(Local Feature Analysis), LDA(Linear Discriminant Analysis), ICA(Independent Component Analysis)등이 연구 진행 중이다.

또한 음성과 영상의 개별적인 특징을 융합하여 인식하는 연구가 활발히 진행 중에 있으며 크게 개별 감정 인식 결과의 가중치 합산에 의한 결정융합 방식과, 특징 검출 단계에서 융합하여 새로운 특징벡터를 생성해내는 특징융합 방식으로 나뉜다. Mingli Song은 특징 융합 방법으로 Hidden Markov Model(HMM)을 이용하여 음성 신호와 얼굴 영상에 대한 감정 인식 실험을 하였으며 De Silva는 결정 융합 방법으로 퍼지 룰 베이스를 이용하여 음성 신호와 얼굴 영상에 대한 감정 인식 실험을 하였다. 그리고 Busso는 두 가지 방법에 대한 비교 설명하였다. 이 밖에 연구로 Hatice Gunes는 얼굴 영상과 제스처에 대해서 결정 융합 방법과 특성 공간 방법을 적용하여 감정 인식 실험을 하였다.

유비쿼터스 컴퓨팅 환경에서의 서비스 추론 기술은 큰

범주에서 HCI(Human-Computer Interface)의 일부로 복미, 유럽 등지에서 다양한 연구가 진행 중이다. 대표적 사례로 미국의 MS에서는 USN(Ubiquitous Sensor Network)를 기반으로 이동성과 지능성을 주요 이슈로 하는 EasyLiving project를 진행하였으며, HP는 Real Web의 구현이라는 모토로 Cool Town project를 진행하였다. 유럽의 경우 HCI가 구현된 architecture개발에 중점을 두고 Smart Its, Paper++ project 등의 연구가 진행 중에 있다.

3. 영상과 음성의 개별특징 기반 감정인식 기술 구현

3.1 얼굴영상(표정) 특징 기반 감정인식

영상에서의 감정정보는 다양한 특징을 가지고 있으며 그 중 눈과 코, 입 주위를 감정특징 정보 포함영역으로 보고, 이 부분을 검출하기 위하여 다양한 이미지처리를 적용하며, 감정인식 이외에도 보안을 위한 얼굴 인식 등 다양한 영상처리 분야로도 적용이 가능하다[3].

영상을 기반으로 감정인식을 수행하기 위해서는 우선 획득된 얼굴 영상이미지에서 얼굴영역만 검출해야한다. 배경 이미지나 안경 등은 감정정보 포함 특징 검출에 있어 불필요한 노이즈이다. 본 논문에서는 그림 1에서와 같은 웹 카메라 입력영상에서 얼굴영역만을 검출하기 위하여 OpenCV의 Haar-Classifer를 사용하였다[3].

이렇게 추출된 얼굴영역 이미지에서 감정정보가 포함된 특징벡터를 추출해야한다. 얼굴영역에서 감정특징을 내포한 부분은 눈썹과 눈, 입으로서 이는 Histogram Equalization, Sobel Edge Detection, Harris Corner Detection의 영상처리를 통하여 검출 가능하다[4].

Histogram Equalization은 입력 영상을 Gray-Scale로 변환한 후, 불균등하게 분포할 수 있는 Contrast를 균일하게 보정하기 위한 전처리 과정으로 이를 통해 우리가 찾고자 하는 특징 영역이 더욱 선명해지도록 하는 기법이다. Sobel Edge 검출은 Gray-Scale로 변환된 이미지에서 영상 경계의 픽셀 값의 급격한 변동을 검출하는 알고리즘으로서 본 논문에서는 살색 영역과 확연하게 구분되는 눈, 눈썹, 입술 영역의 검출에 용이하다. 마지막 Harris Corner Detection은 코너밀도 집중지역을 검출하여 특징영역을 점으로 표현하는 방식으로 아래의 그림 2는 일련의 과정을 거쳐 검출된 최종적인 얼굴영역에서의 감정특징 값들을 표현하고 있다.

이렇게 추출된 얼굴영상 특징 벡터들은 다차원 특징벡터로 이루어진 데이터이며, 효율적인 감정인식 처리를 위하여 PCA를 통해 주성분은 유지한 채, 고차원데이터를 저차원데이터로 축소하여 적절한 특징벡터만을 포함한 Eigen-face를 생성해낸다[1]. 생성된 Eigen-face와 실험을 위하여 수집된 남녀 30인의 5가지 표정에 대한 Database를 비교하여 유클리디안 거리를 측정한 후 측정거리가 최소가 되는 Data가 사용자의 표정으로 나타나는 해당 감정이다.



그림 1. 얼굴영역 검출 및 특징 추출
Fig. 1 Detection and Feature Extraction of Face image



그림 2. 얼굴영상 기반 감정인식 시뮬레이터
Fig. 2. Simulator for emotion recognition based on face image

3.2 음성신호 특징 기반 감정인식

사용자의 음성은 감정정보를 포함하고 있는 대표적인 특징으로는 음성신호를 FFT에 적용하여 나오는 Acoustic feature에 해당하는 Pitch 통계치(최대, 평균), 소리의 크기, 섹션 개수, Increasing Rate(IR), Crossing Rate(CR)를 들 수 있다. 이러한 6가지 특징정보를 통해 감정을 인식하기 위한 방법은 그림 3과 같이 확률 기반의 Bayesian Learning을 이용하여 음성의 감정 패턴을 분류하였다.

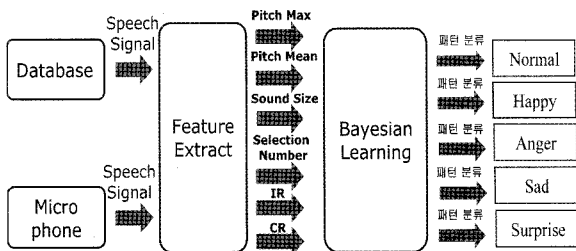


그림 3. 음성신호 기반 감정인식 Process
Fig. 3 Emotion recognition process based on voice signal

베이지 이론은 사전확률을 이용하여 임의의 사건의 확률을 계산하는 방법이다. 베이지 이론을 정확히 정의하기위해

서 몇 가지 표기법을 정의하겠다. 먼저 사건 h 가 발생할 초기 확률은 $P(h)$ 라고 한다. $P(h)$ 는 h 의 '사전확률(Prior probability)'라고 불리고 이것은 h 가 올바른 가설일 가능성에 대해 우리가 알고 있는 배경지식을 의미하는 것이다. 하지만 만약 우리가 아무런 사전지식도 없다면 각 후보 가설들에 대해 동일한 사전확률을 할당해도 된다. $P(D)$ 는 학습용 데이터인 D 가 관찰될 사전확률을 의미한다. 다음으로 가설 h 가 적용된 상태에서 D 가 관찰되는 확률은 $P(D|h)$ 로 나타낸다. 패턴 분류 문제에서 학습 데이터인 D 가 관찰되었을 때의 가설이 h 일 확률인 $P(h|D)$ 를 구하게 되면 패턴 분류문제가 해결 된다. 이때의 $P(h|D)$ 를 '사후확률(Posterior probability)'이라고 한다.

베이지 이론은 베이지 학습의 기반으로써 사전 확률이 $P(h)$, $P(D)$, $P(D|h)$ 로부터 사후확률을 구하는 방법을 제공한다.

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)} \quad (1)$$

위 식을 보면 사후확률인 $P(h|D)$ 는 $P(h)$ 가 증가하거나 $P(D|h)$ 가 증가하면 같이 증가한다.

학습을 하는 많은 경우에 먼저 후보 가설들의 집합 H 를 정의한다. 그리고 D 라는 데이터가 관찰되었을 때 가장 가능성 높은 가설 $h \in H$ 를 찾는다. 이때 가장 가능성이 높은 가설을 최대 사후확률(Maximum A Posteriori :MAP probability)이라고 한다. MAP 가설들은 베이지 이론을 사용하여 구하는데 좀 더 상세히 표현하면 다음과 같다.

$$\begin{aligned} h_{MAP} &\equiv \operatorname{argmax} P(h|D) \\ &= \operatorname{argmax} \frac{P(D|h)P(h)}{P(D)} \\ &= \operatorname{argmax} P(D|h)P(h) \end{aligned} \quad (2)$$

위 식에서 마지막 줄의 $P(D)$ 가 빠진 이유는 h 에 독립적인 상수이기 때문이다[Mitchell 97].

다음의 그림 4는 이와 관련된 일련의 과정에 의해 본 연구실에서 선행 연구, 개발된 음성신호 기반의 감정인식 시뮬레이터를 나타낸다.

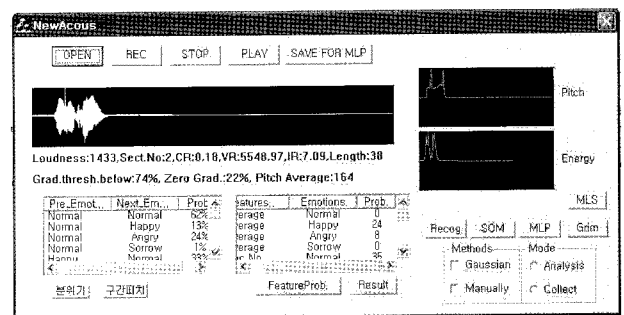


그림 4. 음성기반 감정인식 시뮬레이터
Fig. 4. Simulator for emotion recognition based voice signal

4. 멀티 모달(Multi-Modal) 특징융합 기반 감정인식 기술 구현

사람이 사람의 감정을 인식하는 과정에서 주로 여러 가

지 정보를 취합하여 그에 따라 타인의 감정을 유추한다. 따라서 한 가자의 감정정보만을 고려하여 감정을 판단하는 것은 그 결과에 대한 신뢰성에 문제가 있을 가능성이 크다. 본 논문에서는 이러한 감정인식 결과의 신뢰성과 정확성 향상을 위하여 3장에서 다룬 두 가지 감정인식 방법을 융합한 멀티 모달(Multi-Modal) 감정인식 시스템을 개발하고자 한다.

멀티 모달에 의한 감정인식은 크게 결정융합과 특징융합, 두 가지로 구분 지을 수 있다[6].

본 논문에서의 경우, 결정융합방식은 영상과 음성의 감정인식 결과를 단순 취합한 결과 도출 방법으로서 적용에 있어 용이하다는 장점이 있지만, 개별 감정인식 결과의 단순한 가중치 크기 합산에 따른 결과이므로 인식률과 성능에 있어 다소 부족하다.

반면, 특징융합은 영상과 음성에서의 특징 추출 단계에서 두 가지의 특징을 융합한 새로운 특징벡터를 생성, 이를 이용하여 감정인식을 수행하는 방법이다. 적용에 있어서 어려움이 많지만 그에 따른 성능과 인식률은 높은 편이다. 따라서 본 논문에서는 특징융합에 주로 초점을 맞추어 연구를 진행하였다.

4.1 특징 융합 기법

특징융합방식은 개략적으로 다음 그림 5와 같은 과정을 통해 이루어질 수 있다.

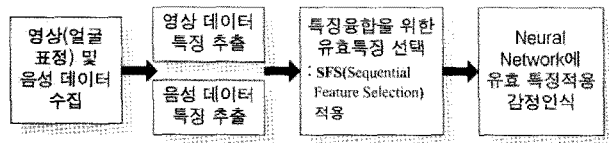


그림 5. 특징융합 Multi-Modal 감정인식 과정
Fig. 5 Emotion recognition process of feature fusion multi-modal

영상과 음성의 개별 특징 검출에 의한 추출 특징의 개수는 영상 5가지(얼굴 이미지의 양 눈, 눈썹 외곽선, 입 외곽선)와 음성 6가지(Pitch의 평균, 최대, Spectrum의 분산, Increasing Rate, Crossing Rate, 섹션개수)로 총 11가지이다. 다수의 특징을 적용하여 인식을 진행 할 경우, 인식과정의 계산 속도가 현저하게 저하되므로, 인식 속도와 효율성의 향상을 위하여 감정정보를 가장 많이 포함한 특징 정보만 선택해야 한다. 이러한 특징 선택은 비선형 최적화 방법으로 가장 기초적인 선택방법인 SFS(Sequential Feature Selection) 기법을 적용하고자 한다. SFS에 대한 상세 내용은 [7]을 참조한다.

SFS를 통해 선택된 특징의 패턴을 통해 가장 적절한 감정의 상태를 추론하기 위하여 패턴분류 및 인식에 뛰어난 Artificial Neural Network의 Back-Propagation으로 학습하는 Multi Layer Perceptron의 적용을 통한 특징별 감정 패턴분류 및 감정 패턴인식을 한다.

4.2 Artificial Neural Network를 이용한 감정별 패턴 분류 및 인식

본 논문에서는 그림 6과 같이 Back-Propagation으로 학습하는 Multi-Layer Perceptron(이하, MLP)를 사용한다. 한 개의 은닉층과 비선형 활성화 함수로 구성된 MLP에 대

한 일반적인 검증은 여러 연구자에 의해 이루어져 있다. 인식 패턴을 나타내는 입력 벡터는 입력층으로 들어간 뒤 은닉층으로 분배된 가중치에 곱해진 뒤 출력층을 통해 결과가 나온다.

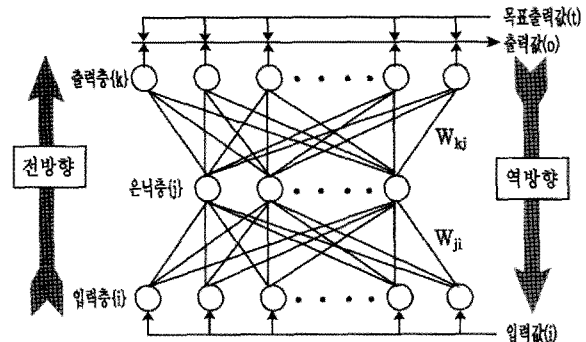


그림 6. Multi Layer Perceptron의 구조
Fig. 6. Structure of multi layer perception

ANN의 각 Artificial Neuron(이하, AN)은 가중된 입력들의 합을 취하고 그 결과를 비선형 활성화 함수로 적용하는 방식을 통해 동작한다. 이것을 수식으로 표현하면 다음과 같다.

$$out_i = f(x_i) = f\left(\sum_j W_{ij}out_j + \theta_i\right) \quad (3)$$

위 식 (1)에서 out_i 는 ANN을 구성하고 있는 임의의 계층에서의 i 번째 AN의 출력이고, out_j 는 직전 층의 j 번째 AN의 출력이다.

또한 입력 x_i 에 대한 활성화 함수로 사용되는 비선형 함수들 중 가장 많이 쓰이는 것은 Sigmoid 함수는 단조 증가 함수이고, 연속이며 0과 1에 접근하기 때문에 ANN에서 가장 많이 사용되는 활성화 함수이다.

$$f(x_i) = \frac{1}{1 + e^{-x_i/Q_0}} \quad (4)$$

Q_0 는 Sigmoid 함수의 기울기를 결정하는 매개변수로 $Q_0 \rightarrow 0$ 이면 활성화 함수는 계단함수로 근사한다.

이러한 조건 하에 구축된 ANN에서 어떤 입력 패턴을 적당한 분류로 사상하기 위하여 요구되는 정보는 가중치(weight, w)로 표현된다. 임의의 문제에 대한 적절한 w 는 초기에 주어지지 않지만 학습(Training)을 통하여 적절한 가중치를 찾게 될 수 있다. 이러한 과정을 거쳐 ANN의 학습 집합이 우수하고, 학습 알고리즘이 효과적이면 표본 학습 집합 외의 입력에 대한 분류 또한 잘 이루어진다는 일반성(Generalization)이 ANN을 사용하는 근거이다.

본 논문에서는 ANN의 가중치 학습을 위한 학습알고리즘으로서 Back-Propagation을 사용하였다. 관련 알고리즘에 대한 상세 설명은 [7]을 참조한다. ANN의 구축 및 학습을 통한 감정인식을 위하여 다음 표 1과 같이 Parameter를 설정하였다.

다음의 표 1과 같은 ANN의 초기 파라미터 설정과 감정별 목표치는 1~101, 초기 가중치는 -0.03~0.03으로 임의로 설정한 후 학습 데이터와의 비교를 통해 인식결과가 오차범위보다 작아질 때까지 학습을 반복한다. 그리고 영상, 음성 입력데이터를 입력한 후 5가지 감정으로 패턴 분류를

실시한다.

표 1. ANN를 위한 초기값
Table 1. Initial Parameters for ANN

Parameter	Values
Input Units	3~5
Hidden Units	11
Output Units	2
Learning rate	0.003
Tolerance	0.25
Sigmoid Function	$1/(1+e^{-3x})$

5. 상황인지 시나리오 기반 서비스 추론

유비쿼터스 컴퓨팅 환경에서 서비스 추론을 하기 위하여 중요한 연구 주제는 상황인지이다. 본 논문에서는 이에 대하여 개론적인 내용만을 제시하며, 실생활에서 가능한 표 2와 같은 시나리오를 제시하고 시나리오 상의 주어진 상황 정보와 감정인식 실험을 통해 인식된 멀티 모달 감정인식 결과를 융합하여 그에 따른 적절한 서비스를 추론하는 과정에 대하여 논한다.

이것을 구현하기 위하여 본 논문에서는 Bayesian Network(이하 BN)를 이용한 확률적 추론 기법을 적용하여 사전확률에 근거한 서비스 추론결과를 도출하고자 한다. 주어진 상황정보와 인식된 감정에 대한 BN을 구성하기 위하여 구조학습과 조건부 확률 테이블에 대한 파라미터 학습이 필요하다. 이에 대한 상세 과정은 [8]을 참조하여 진행하였다.

표 2. 서비스 추론을 위한 상황인지 시나리오
Table 2. Scenario of context recognition for service reasoning

사용자 상황에 따른 감정 인식 수행	휴대폰으로 사용자가 거실에서 누군가와 영상 통화를 하고 있다.	
	통화 중 사용자의 표정과 음성신호를 통해 감정인식을 수행한다.	
감정 인식결과	우울함 (Negative Emotion)	즐거움 (Positive Emotion)
서비스 추론 결과	▶ 통화연결음, 벨소리를 슬픈 음악으로 변경. ▶ 휴대폰 바탕에 재미있는 그림, 만화를 띄움.	▶ 통화연결음, 벨소리를 밝은 음악으로 변경 ▶ 해당업무 및 할 일, 스케줄 안내 개시

6. 실험 및 결과 고찰

영상과 음성을 통해 검출된 특징에 기반으로 한 감정인식 시뮬레이터를 그림 5와 같이 구축하여 사용자의 감정 상태를 인식하였다. 반복적인 실험을 통한 결과물의 인식률은 영상만을 이용한 경우 58%, 음성만을 이용한 경우 65%, 멀티

모달에 의한 경우 70%로 성능향상이 있음을 알 수 있다.



그림 7. 멀티 모달 감정인식 시뮬레이터
Fig. 7. Simulator for emotion recognition based on multi-modal

아래의 표 3은 그림 6의 시뮬레이션 프로그램을 10명의 피험자가 10회씩 반복 실험을 통해 5가지 감정에 대한 결과를 정리한 표이다.

표 3. Multi-Modal 감정인식 결과
Table 3. Result of emotion recognition based on multi-modal

(단위 : %)

	평화	환난	슬픔	기쁨	놀람	평균
S1	70	80	60	60	50	64
S2	70	80	60	60	60	66
S3	80	80	60	60	60	68
S4	80	90	70	70	70	76
S5	70	80	60	70	60	68
S6	80	80	60	60	60	68
S7	70	90	60	60	60	68
S8	80	80	60	70	70	72
S9	70	80	60	70	60	68
S10	70	90	60	70	60	70
평균	74	83	61	65	61	68.8

이러한 감정인식 실험결과와 4장에서 제시한 상황인지 시나리오를 기반으로 간단한 베이지안 네트워크를 구성, 서비스 추론 시뮬레이션을 수행한 결과는 다음 그림 8과 같다. 베이지안 네트워크 구축은 Norsys Co.의 Netica Application을 이용하였다.

각각의 노드는 위의 표 2에서 언급된 4W1H(Where, When, Who, What, How)의 각 요소를 표현하고 있으며, 하나의 노드 안에는 노드를 표현하는 state들이 정의되어 있다. 노드 간의 인과관계 및 확률 관계는 임의의 주어진 상황을 가정하여 저자가 작성하였고, 베이지안 이론에 의하여 시나리오와 같이, ① “휴대폰으로 사용자가 거실에서 누군가와 영상 통화를 하고 있다.” ② 사용자의 감정인식을 수

행하고 그 결과는 우울함이다.]의 과정이 각 노드로 입력이 된다. 그 결과 서비스 노드에서 가장 높은 확률로 나타나는 state를 선택하게 되면 시나리오 상에서와 같은 형태의 서비스 추론이 가능하게 됨을 알 수 있다.

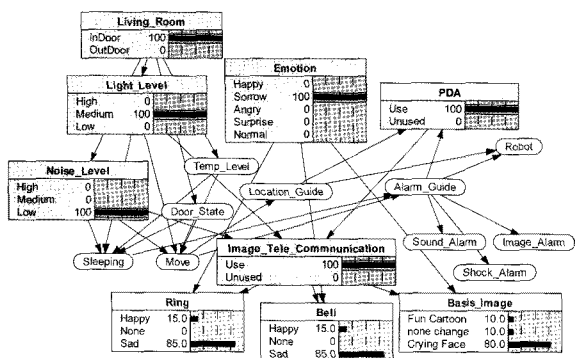


그림 8. 감정인식결과, 상황인지 시나리오에 기반 서비스 추론 모델

Fig. 8. Service reasoning model based on emotion recognition results and context recognition scenario

7. 결론 및 향후 과제

본 논문은 영상과 음성에 존재하는 감정 특징 정보를 활용하여, 멀티 모달 감정인식 기술을 구현하고, 이를 통해 적절하게 설정된 상황인지 시나리오를 기반으로 하는 서비스 추론 시뮬레이션을 진행하였다.

향후 특징융합에 있어서 더 나은 성과를 거두기 위해서는 좋은 성능의 유사도 측정모델의 개발이 요구되며 무엇보다 개별 감정인식 결과의 향상이 선행되어야 한다. 영상 기반 감정인식의 경우 단순한 이미지처리를 통해 데이터베이스와 입력 영상 간의 기하학적 차이점 분석에 중점을 두고 연구를 진행하였지만, 물리적인 차이점이나, 조명, 포즈 등의 외부적 요인에는 취약하다. 또한 음성의 경우 입력으로 들어오는 노이즈에 대한 필터링에 관한 연구가 적용되어 있지 않으므로 적절한 특징 검출에 신뢰성이 저하되는 경향이 있다. 차후 이러한 점을 보완하여 추가적인 연구를 진행할 예정이다. 또한, 특징 융합을 위하여 본 논문에서 제시한 Artificial Neural Network외에도 다른 학습 알고리즘의 적용을 통한 인식을 비교가 이루어져야 한다.

[주] 본 논문은 2008년도 한국지능시스템학회 추계학술대회에서 발표된 우수논문[9]의 개정된 논문임.

참 고 문 헌

[1] 신동일 “감정인식 기술 동향”, *주간기술동향 통권 1283호*, 2007. 2. 14.
 [2] 이대중, 이경아, 고현주, 전명근, “정지영상과 동영상의 융합모델에 의한 얼굴 감정인식”, *퍼지 및 지능시스템학회 논문지*, 2005, Vol. 15, No. 5, pp. 573-580
 [3] 주종태, 장인훈, 양현창, 심귀보 “다중 센서 융합

알고리즘을 이용한 감정인식 및 표현기법,” *제어·자동화·시스템 공학 논문지*, 제13권, 제8호, pp. 754-759, 2007. 08

[4] 정성환, 이문호, *OpenCV를 이용한 컴퓨터 비전 실무 프로그래밍*, 홍릉과학 출판사, 2007
 [5] 황선규, *영상처리 프로그래밍 by Visual C++*, 한빛미디어, 2007.
 [6] 김명훈, 이지근, 소인미 “얼굴과 음성정보를 이용한 바이모달 사용자 인식시스템 설계 및 구현,” *한국컴퓨터정보학회 논문지*, 제10권, 제5호, pp. 353-361, 2005.
 [7] Chang-Hyun Park, Kwee-Bo Sim “The Novel Feature Selection Method Based on Emotion Recognition System,” *LNCS*, Vol. 4115, 2006, pp. 731-740, Sept. 2006.
 [8] 고광은, 신동준, 심귀보 “장애인을 위한 상황인식 및 서비스 추론기술 개발,” *한국지능시스템학회 논문지*, 제18권, 제4호, pp. 512-517, 2008.
 [9] 고광은, 심귀보, “멀티 모달 감정인식 시스템 기반 상황인식 서비스 추론 기술 개발”, *2008년도 한국지능시스템학회 추계학술대회 학술발표 논문집*, 18권, 제2호, pp. 257~260, 2008. 10. 31~11. 1.

저 자 소개



고광은(Kwang-Eun Ko)
 2007년: 중앙대학교 전자전기공학부 공학사
 2007년~현재: 중앙대학교 대학원 전자전기공학부 석사과정

관심분야 : Multi-Agent Robotic Systems (MARS), Machine Learning Context Awareness



심귀보(Kwee-Bo Sim)
 1990년: The University of Tokyo 전자공학과 공학박사
 1991년~현재: 중앙대학교 전자전기공학부 교수

[제18권 6호 (2008년 12월호) 참조]

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