

논문 2007-44CI-4-5

유비쿼터스 센서 네트워크 환경을 위한 다중 생체인식 시스템

(Multi-Modal Biometrics System for Ubiquitous Sensor Network Environment)

노 진 수*, 이 강 현**

(Jin Soo NOH and Kang Hyeon RHEE)

요 약

본 논문에서는 무선 오디오와 영상 인터페이스를 이용한 스위치 제어 및 인식 등의 다양한 유비쿼터스 센서 네트워크 응용 서비스를 지원하기 위한 음성과 얼굴인식 시스템을 구현하였다. 제안된 시스템은 하드웨어 부분으로 무선 오디오 및 이미지 센서, 심리응용모델을 이용한 음성인식과 주성분 분석법(PCA: Principal Components Analysis)을 이용한 얼굴인식 알고리즘, 그리고 LDPC(Low Density Parity Check)로 구성되어 있다. 제안된 음성과 얼굴인식 시스템은 센서의 효율적인 에너지 사용을 위하여 HOST PC에 삽입된다. 그리고 인식 시스템의 정확도를 개선하기 위하여 전방향 에러 정정 시스템을 구현하였다. 또한, 무선 채널 잡음의 효과적인 제거와 정정을 위하여 테스트환경과 시뮬레이션 계수를 최적화 하였다. 결과적으로 사람 음성과 음성센서의 거리가 1.5m 이하일 경우에 FAR과 FRR을 각각 0.126%, 7.5%를 얻었고, 얼굴인식 알고리즘을 2회로 제한하였을 경우, GAR과 FAR을 각각 98.5%와 0.036%를 획득하였다.

Abstract

In this paper, we implement the speech & face recognition system to support various ubiquitous sensor network application services such as switch control, authentication, etc. using wireless audio and image interface. The proposed system is consist of the H/W with audio and image sensor and S/W such as speech recognition algorithm using psychoacoustic model, face recognition algorithm using PCA (Principal Components Analysis) and LDPC (Low Density Parity Check). The proposed speech and face recognition systems are inserted in a HOST PC to use the sensor energy effectively. And improve the accuracy of speech and face recognition, we implement a FEC (Forward Error Correction) system. Also, we optimized the simulation coefficient and test environment to effectively remove the wireless channel noises and correcting wireless channel errors. As a result, when the distance that between audio sensor and the source of voice is less than 1.5m, FAR and FRR are 0.126% and 7.5% respectively. The face recognition algorithm step is limited 2 times, GAR and FAR are 98.5% and 0.036%.

Keywords : Biometrics, Voice and Face Recognition, Ubiquitous, Sensor Network

I. 서 론

정보화시대에 있어서 엄청난 정보량이 컴퓨터에 의해 대량 생산되고, 여기에 정보통신 기술의 비약적인 발전은 기존의 계산기로서의 컴퓨터가 아닌 정보단말기로의 컴퓨

터로 발전하여 더욱더 우리의 생활에 밀접한 영향을 주고 있다. 이런 기술의 진보는 유비쿼터스 컴퓨팅이라는 새로운 정보통신 혁명을 야기하게 되었고, 이런 사회 발전의 흐름과 끊임없이 환경을 인간 친화적으로 바꾸고 싶어 하는 인간의 욕구와 맞물려 무선 센서 네트워크 (WSN: Wireless Sensor Network)의 필요성이 제기되고 있다.

무선 센서 네트워크란 센서가 달려 있어 수집이 가능하고 수집된 정보를 가공할 수 있는 프로세서가 달려 있으며 이를 전송할 수 있는 무선 송수신기를 갖춘 소형장

* 학생회원, ** 평생회원, 조선대학교 전자공학과
(Dept. of Electronic Engineering, Chosun University)

※ 본 논문은 조선대학교 2007년도 교내연구비의 지원으로 수행되었음.

접수일자: 2007년5월27일, 수정완료일: 2007년6월21일

치, 즉, 센서 노드로 구성된 네트워크를 의미하며, 기존의 네트워크와 다르게 의사소통의 수단이 아니라 환경에 대한 정보를 수집하는 것을 그 목적으로 한다. 이에 따라, 인간, 사물 그리고 컴퓨터가 유기적으로 연계되어 다양하고 편리한 서비스를 제공해 주는 유비쿼터스 컴퓨팅 환경에서, 외부 환경의 감지와 제어 기능을 수행하는 센서 네트워크 기술이 최근 활발히 연구되고 있다^[1,2].

무선통신기술과 전자디바이스기술의 발전으로 말미암아 저가격, 저전력, 다기능 센서 노드로 구성된 무선 센서 네트워크에 대한 관심이 급격히 고조되고 있다^[3]. 무선 센서 네트워크는 기존에 구축된 센서네트워크를 무선네트워크로 대체 하는 기술로서, 각 센서노드는 센서에 의한 수집, 수집된 데이터의 처리, 멀티 흡을 통한 네트워킹 등의 기능을 가지고 있으며, 이는 기존의 전통적 의미의 센서에서 정보처리 능력의 향상을 의미한다^[4]. 이러한 정보처리 능력의 향상에 힘입어 WSN^[5]의 응용 분야는 균대정보의 수집과 추적, 환경 모니터링, 환자 감시 그리고 스마트 환경 등의 분야에 적용되어지고 있다^[6~10].

현재 센서 네트워크는 하드웨어와 소프트웨어 플랫폼의 개발과 응용분야의 발굴을 위하여 학계와 산업체에서 많은 연구가 이루어지고 있으며, 대표적인 센서 네트워크 연구 그룹으로는 센서노드용 하드웨어인 MICA^[11]와 운영체계인 TinyOS^[12]를 개발한 버클리 대학과 상업용 응용 시스템에 필요한 연구를 수행하는 Intel^[13]을 들 수 있다.

본 논문에서는 음성과 영상 신호를 수집할 수 있는 무선 센서를 구현하였으며, 구현된 센서를 통하여 수집된 신호를 사용하여 다중 생체인식을 할 수 있는 알고리즘을 제안하였다. 설계된 무선 생체인식을 위하여 심리음향 모델^[14]을 이용한 음성인식과 주성분 분석법(PCA: Principal Components Analysis)을 이용한 얼굴인식 알고리즘을 제안하였고, 무선 채널 잡음에 강인성을 가지기 위하여 LDPC^[15] 모듈을 설계하여 삽입하였다.

본 논문에서는 제안된 생체인식 시스템의 성능 평가를 위해 FAR과 FRR을 측정하였으며, LDPC 모듈의 무선 환경에서 발생되어지는 외부 잡음에 대한 강인성을 측정하였다. 이를 위해 Ⅱ장에서는 본 논문에서 생체인식 알고리즘으로 사용된 심리음향 모델, PCA, 그리고 무선채널에서 잡음에 강인성을 가지는 LDPC에 대하여 알아보겠다. Ⅲ장에서는 생체인식 센서 및 제안 알고리즘의 H/W와 S/W적인 구현 방법을 기술한다. 그리고 Ⅳ장에서 제안된 알고리즘의 성능 측정 및 결과 검토를 하고 마지막 Ⅴ장에서 결론과 향후 연구방향을 제시한다.

II. 이론적 배경

2.1 LDPC

LDPC(low density parity check) 부호는 최근에 가장 주목 받는 오류정정부호로 1960년대 초 Gallager에 의해 제안된 부호로, 패리티 검사 행렬에 0이 아닌 원소의 수가 부호의 길이에 비해 현저히 적게 존재하는 부호로 정의되며 샐론(Shannon) 한계에 가장 근접하는 오류정정부호로서, 터보부호와 더불어 제4세대 이동통신시스템에 활용될 수 있는 매우 우수한 오류정정부호로 평가되고 있다.

식 (1)은 LDPC 부호화 과정을 나타낸다.

$$\begin{aligned} H &= (A_p^{-1} \cdot A) \text{mod}2 = [I \ A_2] \\ G &= \left(\frac{A_2}{I} \right) \\ c &= (G \cdot m) \text{mod}2 \end{aligned} \quad (1)$$

A, H : 패리티체크 행렬(parity check matrix)

A_p^{-1} : 역피봇 행렬(inverse pivot matrix)

G : 생성 행렬(generator matrix)

m : 전송 메시지(transmission message)

c : 부호어(code word)

패리티 체크 행렬 A 와 H 를 이용하여 생성 행렬 G 를 만든 다음 G 와 메시지 m 을 사용하여 부호어 c 가 생성 되어진다.

식 (2)는 LDPC 복호화 과정을 나타낸다.

$$q_n(x) = \alpha P(c_n = x | r_n) \prod_{m \in n} P(z_m = 0 | c_n = x, r) \quad (2)$$

$q_n(x)$: 가상 사후 확률

(pseudo posterior probability)

$\alpha P(c_n = x | r_n)$: 내부 확률

(intrinsic probability)

$\prod_{m \in n} P(z_m = 0 | c_n = x, r)$: 외부 확률

(extrinsic probability)

z_m : 패리티 체크 비트(parity check bits)

c_n : 부호어(code word)

n, m : 행렬 인덱스(row, column index)

r : 전체 수신된 데이터(total received data)

전송 채널을 통과한 코드워드 c 는 잡음 및 여러 성분이 첨가되어 r 의 형태로 수신되고 수신된 신호는 식 (2)에서 채널의 사후확률 계산을 통하여 복호된다.

2.2 심리음향 모델(Psychoacoustic Model)

심리음향 모델^[16-19]은 무선 음성인식 센서로부터 전송되어오는 오디오 데이터를 입력받아 본 논문에서 제안된 음성 인식 알고리즘에 필요한 정보를 제공해주는 역할을 하다. 사람의 귀가 인지할 수 있는 신호 대역은 20Hz에서 20KHz이며, 이를 절대 가청 주파수(Absolute threshold)라고 하며, 아래 식 (3)과 같이 표현되고, 그림 1과 같은 값을 가진다.

$$ATH(f) = (3.64(f/1000)^{-0.8} - (6.5e^{-0.6(f/1000-3.3)^2}) + (10^{-3}(f/1000)^4)) \quad (3)$$

$ATH(f)$ 는 ATH 안의 dB 값을 나타내고, f 는 주파수 값이다. 이 주파수는 낮은 영역과 높은 영역에서 큰 값을 갖고 중간 대역(1KHz~5KHz)에서 낮은 값을 갖는다. 이것은 중간 대역에 음성의 중요한 정보가 들어 있다는 것과 일치한다. 위 곡선에서 가장 낮은 지점은 싸인 레벨의 ± 1 LSB와 사운드 압력이 동일하다. 사람의 귀는 충분히 높은 주파수의 순간적인 소리는 일정기간 유지를 하지 못한다. 이 영역에서 특정 주파수에 따라 듣는 특성이 달라지며, 신호크기가 일정 수준 이상에 이를 때까지 그 소리를 듣지 못한다. 심리음향모델은 신호에너지를 의해 마스킹되는 최대왜곡에너지를 계산하는데 이 에너지를 임계값(threshold)이라고 입력신호를 주파수와 위상으로 분석한다. 입력된 오디오 데이터를 받아들여 출력은 long블록과 short블록으로 구분된다. 출력의 long, short 안의 데이터 값들은 에너지 밀도, 채널대역폭, 채널수로

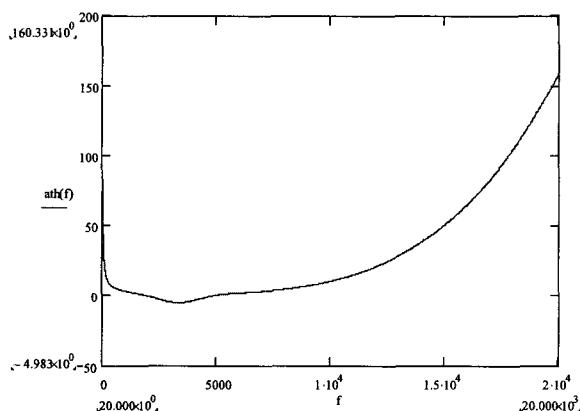


그림 1. 절대가청주파수

Fig. 1. Absolute threshold frequency.

표 1. 심리음향 모델(14단계)

Table 1. Psychoacoustic Model(14 steps).

순서	과정
1	입력 샘플 값 재구성
2	FFT 계산
3	예측된 값의 계산
4	비 예측 값의 측정
5	에너지와 각각 분학된 비 예측 값의 계산
6	에너지와 비 예측 값의 컨볼루션 실행
7	각각의 분학된 음조의 계산
8	각각의 분학된 SNR의 계산
9	각각의 분학된 Power Ratio의 계산
10	분학된 실제적인 에너지 쓰레스홀드의 계산
11	Quiet 상태에서의 쓰레스홀드와 pre-echo 조절
12	PE (Perceptual Entropy) 계산
13	블록 형태의 결정
14	각 SWB에서의 1/SMR의 계산

구성되어있다. 이 값들을 이용하여 음성인식을 진행한다. 전체 14단계는 표 1과 같다.

2.3 주성분 분석법 (PCA)

주성분 분석은 고차원 데이터로부터 데이터의 구조를 밝히거나, 데이터의 차원을 낮추는 데 많이 이용되는 다변량 통계 분석 방법이다. 이는 상관행렬 (correlation matrix)의 고유벡터(eigenvectors)를 찾아내는 문제로 행렬 연산으로 찾아내는 방법과 신경망 등을 사용하여 반복적으로 찾아내는 방법 등이 있다. 즉, 주어진 데이터를 분산이 최대가 되는 축으로 변환하는 것으로, 이 새로운 차원에서의 데이터의 벡터들을 주성분(principal components)이라고 한다. 이 때 분산이 작은 성분을 제거함으로써 데이터의 차원을 줄이는 동시에 데이터에 포함되어 있던 잡음을 제거할 수 있다. 데이터 행렬 X 의 차원을 k 로 낮추는 식은 (4)와 같다.

$$X \cdot V^k \quad (4)$$

여기서 V 는 X 의 상관행렬의 고유벡터에 해당하는 고유값(eigenvalue)의 내림차순으로 정렬한 행렬이고, k 는 이 중 k 개의 열을 사용하겠다는 의미이다.

III. 제안된 알고리즘과 회로 구현

본 논문에서는 무선 영상 획득 장치와 설계된 음성인식 센서 모듈을 사용하여 건물 출입자를 인식할 수 있는 생체인식 시스템을 제안하였다. 제안된 시스템의 전체 블록도는 그림 2와 같다.

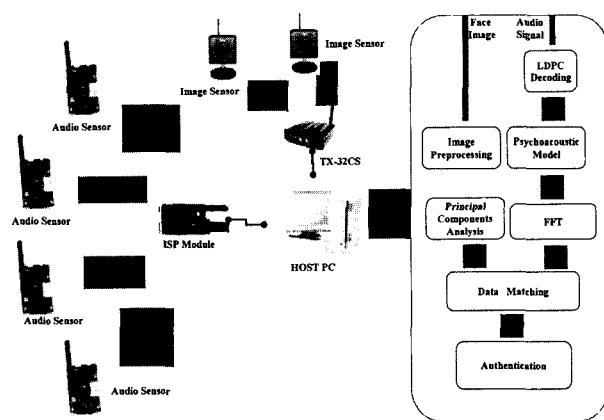


그림 2. 전체 블록도

Fig. 2. Total block diagram.

3.1 음성인식

센서를 통하여 수집된 생체 신호는 FFT를 사용하여 주파수 성분으로 변화 시킨 후 심리음향 모델을 적용시켜 음성신호를 인식하는 데 유효한 24 개의 목소리 특성값을 추출한다. 추출된 데이터는 HOST PC에 저장된 데이터베이스와의 상관관계를 검사하여 출입자를 인식하는 시스템으로 전체 시스템은 JAVA기반으로 설계되었다. 그림 3은 음성인식 시스템의 구현 모습이다.

사용된 심리 음향 모델은 식 (3)을 사용하였으며 표 2는 고유 특성 값을 추출하기 위해 사용한 채널의 주파수 범위 값이다.

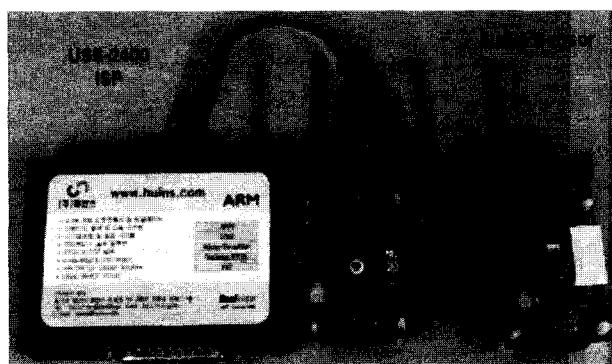


그림 3. 구현된 시스템 모듈

Fig. 3. Implemented system module.

표 2. 채널의 주파수 범위

Table 2. Frequency ranges of channels.

채널 번호	주파수 [Hz]	채널 번호	주파수 [Hz]	채널 번호	주파수 [Hz]
1	100	9	1,270	17	4,400
2	200	10	1,480	18	5,300
3	300	11	1,720	19	6,400
4	400	12	2,000	20	7,700
5	510	13	2,320	21	9,500
6	630	14	2,700	22	12,000
7	770	15	3,150	23	15,500
8	920	16	3,700	24	End

3.2 얼굴인식

본 논문에서 제안된 얼굴인식 알고리즘은 얼굴 정규화(face Normalization), 평균 얼굴(Average face) 계산, 얼굴의 공분산 행렬(Covariance Matrix) 계산, 고유얼굴(Eigen face) 추출, 그리고 검색 알고리즘(Search Algorithm)의 5단계로 이루어져 있다. 그림 4는 제안된 얼굴 인식 알고리즘이다.

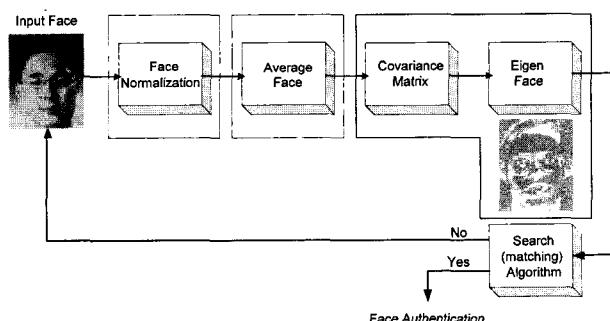


그림 4. 제안된 얼굴 인식 알고리즘

Fig. 4. The proposed algorithm of face identification.

1. 얼굴 정규화

이미지 정규화는 얼굴 인식 시스템에 필요한 얼굴과 배경부분을 명확하게 구분 시켜주며, 또한 외부 잡음에 따른 명암의 변화나 얼굴의 손상에 대하여 강인성을 부여해준다. 식 (5)는 표준편차를 이용한 정규화 과정을 나타내며 약 50개의 얼굴을 실험한 결과 본 논문에서는 추정평균은 100, 추정 표준편자는 80으로 설정하였다.

$$P = (I - \bar{I}) \times \frac{ustd}{std} + um \quad (5)$$

P = 정규화 얼굴(Normalized face)

I = 입력 얼굴(Input Data)



그림 5. 정규화된 얼굴

Fig. 5. Normalized face.



그림 6. 고유얼굴

Fig. 6. Eigen face.

$$\bar{I} = I \text{의 평균}(Average of I)$$

$$um = \text{추정 평균}(Assumption Average)$$

$$ustd = \text{추정 표준편차}(Assumption Standard Deviation)$$

$$std = I \text{의 표준편차}(Standard Deviation of I)$$

그림 5는 식 (5)가 적용된 정규화된 얼굴영상이다.

2. 고유얼굴

얼굴 영상의 각 픽셀이 얼굴공간(face space)상에서 구성된다고 할 때, 고유얼굴이란 얼굴공간을 구성하는 기저벡터에 해당하며 모든 인식 후보 얼굴의 공통적인 특징을 가장 잘 나타내는 기저벡터를 의미한다. 즉, 고유얼굴은 모든 인식 후보 얼굴영상의 평균얼굴 영상과 각 후보 얼굴영상의 차 벡터로 공분산 행렬의 고유벡터에 해당한다. 공분산 행렬의 고유치는 평균얼굴 영상에 대한 분산의 정도를 나타내기 때문에 가장 큰 고유치에 대응하는 고유벡터로 이루어진 고유얼굴이 가장 비슷한 얼굴이 되며, 고유치가 작아질수록 얼굴 특성이 작아진다. 식 (6)은 얼굴의 공분산 행렬을 나타낸다.

$$C = \frac{1}{M} \sum_{n=1}^M \Phi_n \Phi_n^T = AA^T, \quad (6)$$

$$A = \frac{1}{\sqrt{M}} [\Phi_1, \Phi_2, \Phi_3, \dots, \Phi]$$

M : 인식후보 얼굴 벡터 수(A number of the recognition candidate face vector)

식 (6)에서 Φ 는 각 인식후보 얼굴 벡터와 평균얼굴 벡터의 차 벡터이며 식 (7)로 나타내어지며, Ψ 는 평균

얼굴 벡터로 식 (8)과 같다.

$$\Phi_i = \Gamma_i - \Psi \quad (7)$$

$$\Psi = \frac{1}{M} \sum_{n=1}^M \Gamma_n \quad (8)$$

식 (7)과 (8)에서 사용된 Γ 는 인식후보 얼굴을 행벡터로 재배열한 벡터이다.

공분산 행렬 C로부터 고유값 λ 와 이에 대응하는 고유벡터 μ 를 계산한다. 여기서, 고유값은 평균얼굴 영상에 대한 분산의 정도를 나타내며, 식 (9)를 이용하여 고유벡터 μ 를 계산할 수 있다. 고유벡터는 각 후보 얼굴의 기저 벡터가 되며 데이터베이스에 저장된다.

$$\mu_l = \frac{1}{M} \sum_{k=1}^M \lambda_{lk} \Phi_k \quad l=1,2, \dots, M \quad (9)$$

그림 6은 식 (9)를 통하여 계산된 고유얼굴이다.

3. 얼굴 인식

인식을 위한 얼굴이 입력되면 고유얼굴에 대한 사영을 취하여 그 성분값을 구한다. 사영을 취한다는 것은 내적을 의미하며 얼굴의 기저벡터 성분의 가중합으로 얼굴의 유사도를 판별 한다. 입력되는 얼굴영상에 대한 각 고유얼굴 성분값은 식 (10)과 같이 계산된다.

$$\omega_k = \mu_k^T (\Gamma - \Psi) \quad k=1,2,\dots,M \quad (10)$$

μ : 얼굴의 기저 벡터(basis vector of face)

Γ : 입력 얼굴 벡터(*input face vector*)

Ψ : 평균 얼굴 벡터(*average face vector*)

식 (10)을 통하여 계산된 가중치를 이용하여 입력 얼굴을 표현하는 고유얼굴 성분 벡터 ($\Omega = [w_1, w_2, w_3, \dots, w_M]$)를 구할 수 있다. Ω 가 구해지면,

$$\epsilon_k = |\Omega - \Omega_k|^2 \quad k=1,2,\dots,M \quad (11)$$

식 (11)과 같이 인식을 위한 얼굴과 후보 얼굴영상들의 가중치와 유클리디안 거리를 비교하여 그 거리가 최소가 되는 얼굴이 입력 얼굴과 가장 유사한 얼굴이 되므로 이 영상 후보를 인식 결과로 결정하는 것이다.

IV. 실험 결과

음성 및 영상의 무선 센서에서 수집된 데이터는 JAVA로 구현된 오실로스코프 상에서 확인 가능 하며 본 논문에서는 그림 7과 같이 실험실에 무선의 8개 오디오 센서와 2개의 영상 카메라를 배치하여 음성 및 얼굴인식 실험을 진행하였다.

수집된 데이터는 HOST PC로 전송이 되며 JAVA로 구현된 LDPC 디코더 블록을 통과한 후 오실로스코프 상에서 각각의 채널을 통해 들어오는 데이터를 확인 할 수 있으며, 최종적으로 JAVA 프로그램 상에서 음성과 얼굴 인식 시스템이 동작한다.

영상 센서는 이전영상과의 차이 값을 계산하여 물체의 움직임이 발생되면 영상을 수집하여 얼굴인식 알고리즘을 구동시킨다.

인식시스템의 성능평가를 위한 기준으로 사용되는

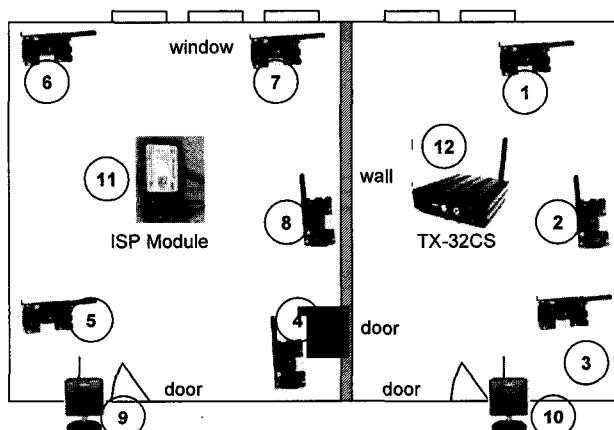


그림 7. 센서의 배치

Fig. 7. Deployment of sensor.

대표적인 방법은 타인수락오류율 (FAR : False Acceptance Ratio)과 본인거부오류율 (FRR : False Rejection Ratio)이다. 본 논문에서는 타인수락오류율과 본인거부오류율로서 식 (12)를 사용하였으며 [1,2]와 비교 데이터를 얻기 위하여 GAR(Genuine Acceptance Rate)을 계산하였다.

$$\begin{aligned} FAR &= \frac{\gamma_a}{\gamma_b} & FRR &= \frac{\delta_a}{\delta_b} \\ GAR &= \frac{\delta_b - \delta_a}{\delta_b} = 1 - FRR \end{aligned} \quad (12)$$

γ_a : 타인이 본인으로 오인식된 횟수

γ_b : 본인대타인 매칭 횟수

δ_a : 본인이 타인으로 오인식된 횟수

δ_b : 본인대본인의 매칭 횟수

얼굴의 인식은 그림 8과 같이 얼굴 데이터베이스에 들어있는 고유 얼굴의 기저벡터와 얼굴획득 장치로부터 들어오는 입력얼굴의 기저벡터를 매칭시켜 유클리디안 거리가 일정 범위 값 안에 있는 얼굴을 입력된 얼굴과 동일한 얼굴로 인식한다.

그림 9는 324개의 얼굴을 본인 대 타인 매칭 시뮬레이션 한 결과 생성된 유클리디안 거리이다. 이 결과로

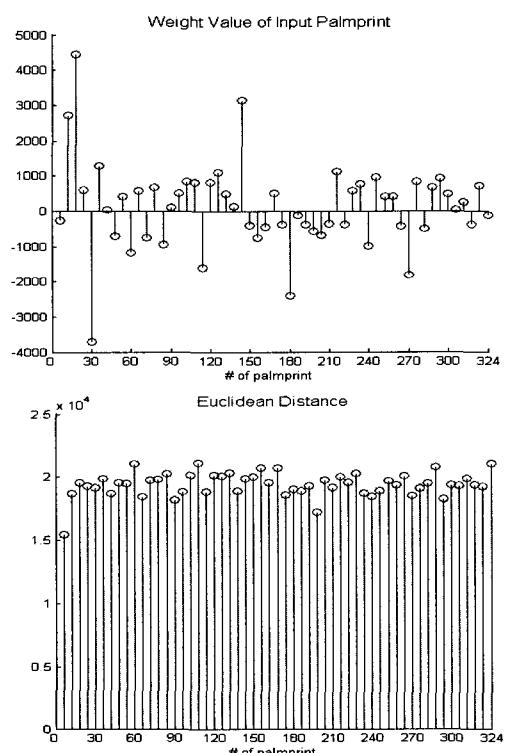


그림 8. 얼굴 인식 과정

Fig. 8 The process of the face identification.

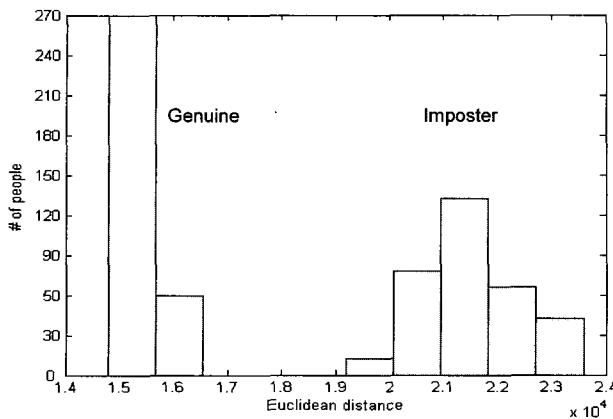


그림 9. 얼굴의 유클리디안 거리

Fig. 9. The euclidean distance of face.

표 3. 2차 인식으로 측정된 FAR과 GAR 성능

Table 3. The FAR and GAR performances of the second identification.

Critical Value	γ_a	FAR (γ_a/γ_b)[%]	δ_a	GAR ($1-\delta_a/\delta_b$)[%]
1.62×10^4	45	0.031	6	98.1
1.63×10^4	54	0.033	6	98.1
1.64×10^4	61	0.036	5	98.5
1.65×10^4	107	0.052	4	98.8

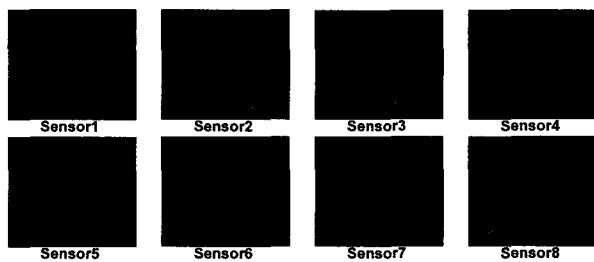


그림 10. 센서로부터 수집된 음성 신호

Fig. 10. The collected voice signal from sensors.

부터 얼굴인식 임계값(Critical value)을 1.62×10^4 부터 1.65×10^4 으로 설정하여 시뮬레이션을 하였다.

표 3은 재 인식에 따른 얼굴인식 시스템의 FAR과 GAR을 나타낸다.

표 3에서와 같이 2차 인식과정을 거쳐 나오는 데이터에서는 임계값이 1.64×10^4 일 때 FAR과 GAR이 각각 0.036[%], 98.5[%]를 얻었다.

음성 센서의 동작은 주변에서 발생하는 소리가 일정 레벨 이하일 때는 슬립 모드로 동작하다가 일정 레벨 이상의 소리 신호에 대해서만 신호를 수집하는 형태로 동작한다. 그림 10은 LDPC 디코딩 후 오실로스코프에서

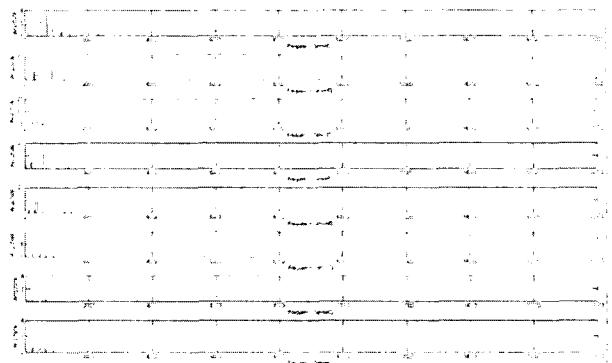


그림 11. 음성 신호로부터 추출된 값

Fig. 11. The extracted value from voice signal.

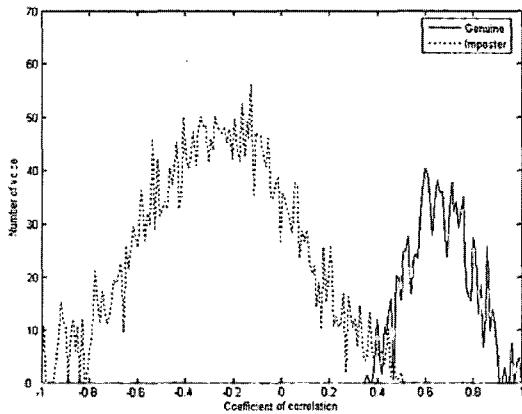


그림 12. 음성의 상관계수

Fig. 12. The coefficient of correlation of voice.

보여 지는 오디오 신호이며, 그림 11은 음성 인식과정을 보여주고 있다.

각각의 센서에서 수집된 음성신호로부터 계산된 24개의 특성값과 데이터베이스에 저장된 특성값의 유사도를 비교하여 본인 인식 여부를 결정한다. 유사도의 비교는 식 (13)의 상관계수를 사용하였다.

$$k = \frac{\frac{1}{n} \sum_{m=1}^n (a_m - \bar{a})(b_m - \bar{b})}{\sigma_a \sigma_b} \quad (13)$$

k : coefficient of correlation($-1 \leq k \leq 1$)

\bar{a}, \bar{b} : average

σ_a, σ_b : standard deviation

그림 12는 센서를 통하여 입력된 음성을 본인 대 타인 매칭 시뮬레이션 한 결과 생성된 상관계수이다. 이 결과로부터 음성인식 임계값(Critical value)을 0.45로 설정하여 실험을 진행하였다.

표 4는 임계값의 변화에 따른 음성 인식 시스템의

표 4. 임계값에 따른 FAR과 GAR 성능

Table 4. The FAR and GAR performances according to critical values.

Critical Value	γ_a	FAR[%]	δ_a	FRR[%]
0.37	11	0.048	5	12.5
0.42	27	0.117	8	20.0
0.45	32	0.139	11	27.5
0.48	62	0.270	14	35.0

표 5. 재인식에 따른 FAR과 FRR 성능

Table 5. The FAR and GAR performances according to the re-identification.

Critical Value	γ_a	FAR[%]	δ_a	FRR[%]
0.37	10	0.035	2	5.0
0.42	25	0.109	2	5.0
0.45	29	0.126	3	7.5
0.48	49	0.213	6	15.0

FAR과 FRR을 나타낸다.

표 3에서 전체적으로 FAR과 GAR이 낮은 성능을 얻었다. 그 이유는 임계값 범위 이하에 들어가는 데이터가 두개 이상 일 때에는 음성인식을 오인된 횟수로 처리했기 때문이다. 이와 같은 점을 개선하기 위하여 임계값 범위 안에 두개 이상의 값이 들어오면 다시 한전 음성을 획득하여 음성을 재인식하는 구조를 가지고 설계하였다. 표 5는 음성 인식 시스템의 재 인식을 거쳐 나오는 데이터의 FAR과 FRR 값이다.

표 4에서와 같이 재 인식과정을 거쳐 나오는 데이터에서는 임계값이 0.45 일 때 FAR과 FRR이 각각 0.126[%], 7.5[%]를 얻었다. 또한, 센서와 음성발원지의 거리에 따라 음성인식률에 차이가 발생됨을 확인 할 수 있으나 실험결과 음성발원지로부터 1.5미터 이내에 음성 인식 센서가 있을 경우에는 인식률에 큰 차이가 없으나 2미터 이상부터는 인식률이 급격하게 떨어짐을 확인 할 수 있었다.

V. 결 론

유비쿼터스 환경에서 휴면 인터페이스 기술 중 하나인 생체인식 기술의 필요성이 증가하고 있다. 이에 따라 본 논문에서는 무선 오디오와 영상 센서를 이용하여 출입자의 음성 및 얼굴인식을 할 수 있는 시스템을 설계하였다. 본 논문에서 설계한 무선 음성과 얼굴인식

센서의 채널잡음을 줄이기 위하여 설계한 LDPC 부/복호기로 구성되며, 소프트웨어는 심리음향모델, FFT, 그리고 데이터의 인식 알고리즘으로 구성되어 있다.

본 논문에서 구현된 시스템은 1.5 미터 이내의 거리에서 FAR 0.126%, FRR 7.5%의 성능 즉, 92.5%의 인식률을 가진다. 하지만 거리에 비례하여 성능이 급격히 떨어짐을 확인할 수 있었으며, 얼굴인식은 2차 인식과정을 거쳐 GAR이 98.5[%]의 높은 인식율을 얻었다.

본 논문에서 구현된 시스템은 대부분의 처리를 베이스 스테이션에 의존하는 문제점을 내포하고 있다. 따라서 베이스 스테이션에 오버헤드가 커지는 단점이 존재하며, 또한 센서와 음원과의 거리에 비례하여 성능이 급격히 떨어지는 문제점을 내포하고 있다. 앞으로는 이러한 문제점을 해결할 수 있는 방향의 추가적 연구가 이루어져야 할 것이다.

표절논문 인용주의

DOI 10.1007/11881223_110

<http://www.springerlink.com/content/rh0kl5p564033v1q/>은 참고문헌 [16]을 그리고
DOI 10.1007/11539902_132
<http://www.springerlink.com/content/agvy2rn0lbe5rjl/>은 참고문헌 [17,18,19]를 표절하였으므로 이를 인용 시, 주의를 요합니다.(원저자 일동 및 선임변호사)

References

- [1] Paramvir Bahl and Venkata Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," Proc. of IEEE INFOCOM, vol. 2, p.p. 775-784. March 2000.
- [2] Nissanka B., Priyantha, Anit Chakraborty, Hari Balakrishnan, "The cricket location-support system," Proc. of MOBICOM 2000, p.p.32-43, Boston, MA, Aug. 2000, ACM, ACM Press.
- [3] Ian D. Chakeres and Luke Klein-Berndt, "AODVjr, AODV Simplified," ACM SIGMOBILE Mobile Computing and Communications Review, vol. 6, no. 3, Jul. 2002, p.p.100-101.
- [4] E. H. Callaway, "Wireless Sensor Networks Architectures and Protocols," Auebach, 2003.
- [5] Ustart-2400, <http://www.huins.com>
- [6] R. Chellappa, G. Qian, and Q. Zheng, "Vehicle detection and tracking using acoustic and video sensors," in Proceedings of the International

- Conference on Acoustics, Speech and Signal Processing, Salt Lake City, Utah, May 2001.
- [7] H. Wang, D. Estrin, and L. Girod, "Preprocessing in a tiered sensor network for habitat monitoring," EURASIP JASP special issue of sensor networks, pp. 392-401, 2003.
- [8] Wen Hu et al. "The Design and Evaluation of a Hybrid Sensor Network For Cane-toad Monitoring,"
- [9] A. Taylor, G. Grigg, G. Watson, and H. McCallum, "Monitoring frog communities: an application of machine learning," in Proceedings of the 8th Innovative Applications of Artificial Intelligence Conference. Portland Oregon: AAAI, pp. 1564-1569, Aug. 1996.
- [10] S. Shukla, N. Bulusu, and S. Jha, "Cane-toad monitoring in kakadu national park using wireless sensor networks," in Proceedings of APAN, Cairns, Australia, July 2004.
- [11] MICA2, <http://www.xcross.com>
- [12] TinyOS, <http://webs.cs.berkeley.edu>.
- [13] Intel, <http://www.intel.com/research/exploratory/wireless-sensors.htm#sensornetwork>.
- [14] "ISO/IEC MPEG-2 Advanced Audio Coding 382(N-1)" -Presented at the 101st Convention 1996 November 8-11 Los Angeles, California, AN AUDIO ENGINEERING SOCIETY PREPRINT, 1996.
- [15] R.G. Gallager, "Low-density parity-check codes," IRE Trans. Inform. Theory, vol. IT-8, pp. 21-28, Jan. 1962.
- [16] 김병규, "MPEG-2 AAC 디코더 칩의 설계," 조선대학교 대학원 전자공학과 석사학위논문, 2000. 2.
- [17] 김정태, "오디오 워터마크 프로세서 설계," 조선대학교 대학원 전자공학과 석사학위논문, 2001. 2.
- [18] 이강현, "MPEG-II AAC의 MDCT/IMDCT를 위한 벡터프로세서 설계," 대한전자공학회 하계종합학술대회 논문집, pp.329-332, 1999
- [19] 김병규, 이강현, "MPEG-2 AAC의 MDCT/IMDCT를 위한 최적 알고리즘 개발," 대한전자공학회 하계종합학술대회 논문집, pp.538-541, 1999.

저 자 소 개



노 진 수(학생회원) -제1저자
2002년 조선대학교 전자공학과
공학사
2004년 조선대학교 대학원 전자공
학과 공학석사
2007년 조선대학교 대학원 전자공
학과 공학박사

<주관심분야 : UWB, 생체인식, 양자컴퓨팅>

23x30



이 강 현(평생회원)-교신저자
1979년, 1981년 조선대학교 전자공
학과 공학사 및 석사
1991년 아주대학교 대학원
공학박사
1977년~현재 조선대학교 교수
1991년, 1994년 미 스텐포드대
CRC 협동연구원.
1996년 호주시드니대 SEDAL 객원교수
2000년~현재 한국 멀티미디어 기술사 협회 이사
2002년 영국 런던대 객원 교수
2002년 대한전자공학회 멀티미디어연구회 전문
위원장
2003년 한국 인터넷 방송/TV 학회 부회장
2003년~현재 대한전자공학회 정보화위원회 이사
2005년~현재 :조선대학교 RIS 사업단장
<주관심분야 : 멀티미디어 시스템 설계, Ubiquitous convergence>