

## 얼굴과 음성 정보를 이용한 바이모달 사용자 인식 시스템 설계 및 구현

김명훈\*, 이지근\*\*, 소인미\*\*\*, 정성태\*\*\*\*

### Design and Implementation of a Bimodal User Recognition System using Face and Audio

Myung-Hun Kim\*, Chi-Geun Lee\*\*, In-Mi So\*\*\*, Sung-Tae Jung\*\*\*\*

#### 요 약

최근 들어 바이모달 인식에 관한 연구가 활발히 진행되고 있다. 본 논문에서는 음성 정보와 얼굴정보를 이용하여 바이모달 시스템을 구현하였다. 얼굴인식은 얼굴 검출과 얼굴 인식 두 부분으로 나누어서 실험을 하였다. 얼굴 검출 단계에서는 AdaBoost를 이용하여 얼굴 후보 영역을 검출 한 뒤 PCA를 통해 특징 벡터 계수를 줄였다. PCA를 통해 추출된 특징 벡터를 객체 분류 기법인 SVM을 이용하여 얼굴을 검출 및 인식하였다. 음성인식은 MFCC를 이용하여 음성 특징 추출을 하였으며, HMM을 이용하여 음성인식을 하였다. 인식결과, 단일 인식을 사용하는 것보다 얼굴과 음성을 같이 사용하였을 때 인식률의 향상을 가져왔고, 잡음 환경에서는 더욱 높은 성능을 나타냈었다.

#### Abstract

Recently, study of Bimodal recognition has become very active. In this paper, we propose a Bimodal user recognition system that uses face information and audio information. Face recognition consists of face detection step and face recognition step. Face detection uses AdaBoost to find face candidate area. After finding face candidates, PCA feature extraction is applied to decrease the dimension of feature vector. And then, SVM classifiers are used to detect and recognize face. Audio recognition uses MFCC for audio feature extraction and HMM is used for audio recognition. Experimental results show that the Bimodal recognition can improve the user recognition rate much more than audio only recognition, especially in the presence of noise.

▶ Keyword : Audio Recognition, Face Recognition, AdaBoost, HMM, PCA, SVM, Bimodal

• 제1저자 : 김명훈

• 접수일 : 2005.10.12, 심사완료일 : 2005.11.10

\* 원광대학교 컴퓨터 공학과 석사과정, \*\* 원광대학교 컴퓨터 공학과 박사과정

\*\*\* 원광대학교 컴퓨터 공학과 박사과정, \*\*\*\* 원광대학교 전기전자및정보공학부 부교수

※ 이 논문은 2005년도 교육인적자원부 지방연구중심대학육성사업 헬스케어기술개발사업단의 지원에 의하여 연구되었음

## I. 서론

최근 인체의 인식에 대한 연구가 활발히 진행되고 있다. 인식 기술에는 지문, 홍채, 얼굴, 음성 등 각각 독립적으로 수행되어 왔다. 이러한 방식은 각 생체 정보들이 가지는 문제점으로 인해 인식률의 저하를 가져온다. 따라서 이러한 단점을 보완하기 위해 복수개의 생체 정보를 이용하는 멀티모달 생체 인식에 대한 연구가 진행되고 있다 [1][2].

본 논문에서는 이 중에서 거부감이 없는 인식방법인 얼굴인식과 음성인식의 합성 방법을 사용하였다. 얼굴인식 방법으로는 얼굴 영역의 파라미터를 구하기 위해 주성분 분석을 사용하였고, 학습과 인식 방법으로는 SVM(Support Vector Machine)을 사용하였다. 또한, 음성인식을 하기 위해 HMM(Hidden Markov Model)을 사용하였다. 합성 방법으로는 각각 독립적으로 얼굴과 음성을 인식 한 다음 인식 결과를 합성하는 방법을 사용하였다.[3]

실험 결과 잡음에 의한 음성의 인식을 저하해 얼굴 인식과 합성함으로써, 전체적인 인식을 향상을 볼 수 있다. (그림 1)은 본 논문의 시스템 구조이다.

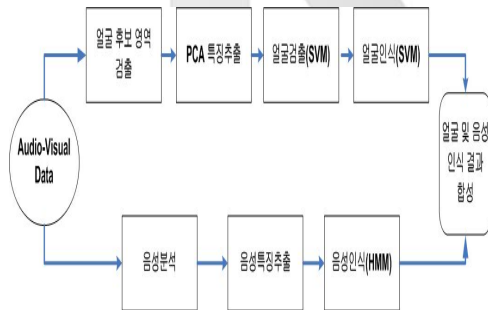


그림 1. 바이모달 시스템 구조  
Fig 1. Bimodal System Structure

## II. 얼굴 검출 및 인식

본 논문에서는 얼굴을 검출하기 위해 이미지 내, 간단한 특징정보를 추출하고, Cascade 구조의 AdaBoost 방법을 이용하여 얼굴 후보 영역을 검출하였다. 또한 얼굴을 검증하기 위하여 주성분 분석을 통하여 얼굴 특징을 축소하여, SVM을 통하여 얼굴을 검증하였다. (그림 2)에는 본 논문에서 제안된 얼굴 검출 시스템의 흐름도가 나타나 있다.

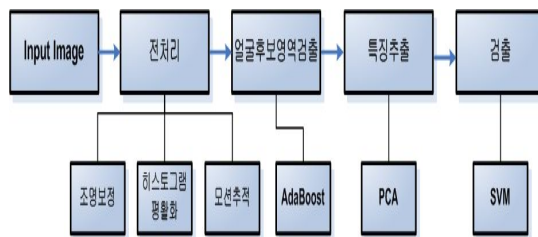


그림 2. 얼굴검출 시스템  
Fig 2. Face detection system

### 2.1 얼굴 검출 및 검증

#### 2.1.1 얼굴 후보 영역 검출

##### A. 전처리 단계

본 논문에서는 전처리 단계로 조명 보정, 히스토그램 평활화, 모션 추적 과정을 적용하였다. 조명 보정은 빛의 각도에 따른 그림자를 줄이기 위해 사용한 것이며, 히스토그램 평활화는 명암 값이 일정한 분포를 갖기 위한 방법이다. 모션추적은 얼굴의 움직임이 발생한 경우에 얼굴 검출을 수행하기 위해 사용하였다.

##### B. Harr-like 특징과 인테그랄 이미지

얼굴 후보 영역은 간단한 특징의 집합으로부터 검출될 수 있다. (그림 4)는 얼굴 영역 내 특징을 보여주고 있다.

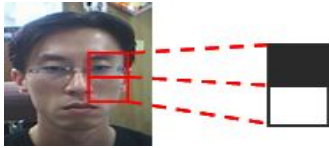


그림 3. 얼굴 영역 내 특징  
Fig 3. Face area feature

본 논문에서는 얼굴 검출을 위한 특징으로 Papageoriou et al 에 의해 제안된 간단하면서도 연산이 빠른 Harr-like 특징을 사용하였다.[4] Harr-like 특징은 인티그럴 이미지를 이용하여 빠르게 연산할 수 있다. 입력 영상에 윈도우를 이동시키면서 특징을 추출하며, 각 모형에서 얻어지는 특징의 수는 윈도우의 크기 및 그 수에 따라 달라진다.

C. 얼굴 후보 영역 검출

본 논문에서는 얼굴 후보 영역을 검출하기 위해 AdaBoost 다. AdaBoost 알고리즘은 Freund와 Schapire에 의해 소개되었다.[5] AdaBoost 알고리즘은 여러 개의 약한 분류기의 선형적인 결합을 통하여 최종적으로 높은 검출 성능을 가지는 강한 분류기를 생성하여 분류하는 기법이다. (그림 4)는 AdaBoost에 의해 강한 분류기가 생성되는 것을 보여주고 있다.

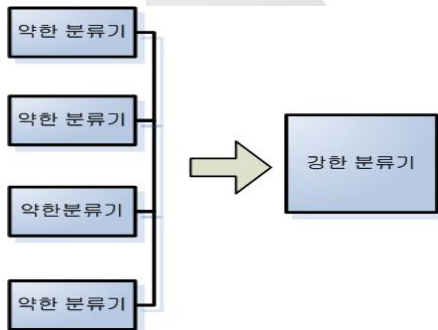


그림 4. 강한 분류기 생성  
Fig4. Create to strong classification

Cascade 구조는 계산 시간을 줄이고 얼굴 검출 결과를 향상되게 한다. 이 구조는 여러 개의 스테이지를 나누고 스테이지별 약한 분류기의 수를 다르게 하여 수행하는 방법이다.[6] 처음 스테이지 분류에서 대부분의 배경(비얼굴)을 제거되는 기능을 가지고 있어 단일 스테이지를 적용하는 방법보다 훌륭한 성능을 보였다. (그림 5)는 얼굴 검출을 위

한 Cascade 구조를 보여주고 있다. 본 논문에서는 얼굴 검출을 위해 총 24개의 Cascade 구조를 사용하였다.

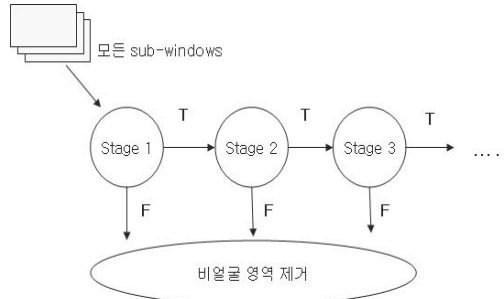


그림 5. 얼굴 검출을 위한 Cascade 구조  
Fig 5. Cascade structure for Face detection

2.12 얼굴 검증

본 논문에서는 얼굴 검증을 위해 이진 SVM 분류기를 이용하였다. SVM 분류를 하기 위해 얼굴 후보 영역의 특징 계수를 줄이는 방법으로 주성분 분석(Principal Component Analysis)을 사용하였다.[7] 주성분 분석을 통해 얼굴 후보 영역의 특징 벡터 계수를 줄였다. SVM 학습에서는 19x19 크기의 MIT 데이터베이스[8]의 얼굴 이미지에 PCA를 적용하여 특징 벡터로 사용하였고, 32x 32 크기의 정규화 된 BioID[9] 얼굴 데이터를 이용하여 검증하였다.

A. 주성분 분석(Principal Component Analysis)

PCA는 고차원 입력 벡터를 저차원 벡터로 표현하여 몇 개의 주성분 값으로 나타내어 주는 방법이다. PCA를 이용하여 얼굴 영역 이미지의 특징 파라미터를 추출하는 것은 전체 영상의 데이터를 데이터의 분산이 큰 몇 개의 고유얼굴(eigen faces)에 대한 축으로 선형 투사시켜 데이터의 차원을 줄이는 것이다. 여기에서 고유 얼굴은 원래의 얼굴 영역 이미지에 일치하는 공분산 행렬(covariance matrix)의 고유벡터(eigen vectors)를 나타낸다.

얼굴 이미지의 학습 집합(  $X$  )를  $M$  개의 벡터로 표현하면 식 1과 같고, 각 벡터는  $N$ 개의 픽셀을 갖는다. 즉, 벡터  $x_i$ 는 얼굴 이미지의 전체 픽셀  $N$ 개를 갖는 벡터를 나타낸다.

$$X = \{ x_1, x_2, \dots, x_M \} \dots\dots\dots (1)$$

M개의 얼굴 이미지의 평균벡터  $\Psi$ 는 식 2와 같고, 얼굴 이미지의 벡터 ( $x_i$ )와 얼굴 이미지의 평균벡터 ( $\Psi$ ) 사이의 차이는 식 3과 같이 구할 수 있다.

$$\Psi = \frac{1}{M} \left[ \sum_{i=1}^M x_i \right] \dots\dots\dots (2)$$

$$\Phi_i = x_i - \Psi, \quad i = 1, \dots, M \dots\dots\dots (3)$$

식 4를 이용하여 공분산 행렬 ( $C$ )을 구한다.

$$C = \frac{1}{M} \left[ \sum_{n=1}^M \Phi_n \Phi_n^T \right] \dots\dots\dots (4)$$

식 4에서 행렬 ( $C$ )의 고유값 ( $\lambda$ )과 고유벡터 ( $e$ )는 식 5의 방정식 해를 찾음으로써 구한다.

$$C e = \lambda e \quad (x \neq 0) \dots\dots\dots (5)$$

계산된 고유벡터들의 행렬을  $A$ 라고 하고, 다음 식 6을 통해서 얼굴 이미지 벡터  $x$ 의 주성분 값  $y$ 를 구한다.

$$y = A^T (x - \Psi) \dots\dots\dots (6)$$

식 6을 이용하면 원래 얼굴 이미지  $x$ 로 복원하는 식 7을 재구성할 수 있다.

$$x = A^T y + \Psi \dots\dots\dots (7)$$

본 논문에서는 행렬  $A$ 의 고유벡터들을 모두 사용하지 않고 상위 몇 개를 고유벡터로 사용하였다. 이를 다시 표현하면 식 8, 9와 같다.

$$y = A_K (x - \Psi) \dots\dots\dots (8)$$

$$x = A_K^T y + \Psi_K \dots\dots\dots (9)$$

본 실험에서는 고유 얼굴을 인식할 데이터의 각 프레임에 사영(projection)시켜 얼굴 특징 계수를 추출하였다.

B. SVM (Support Vector Machine)

SVM은 Vapnik에 의해 제안된 방법으로 이진페턴 분류 문제를 해결하기 위한 것이다.[10] SVM은 각 범주의 대표적인 특성을 갖는 벡터를 이용하여 최적의 하이퍼분리면을 찾아서 분류하는 방법이다.

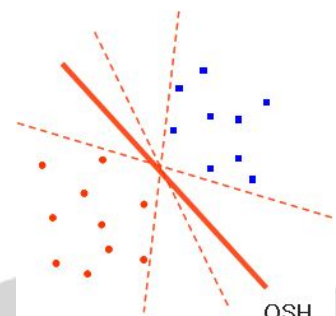


그림 6. 하이퍼분리면  
Fig 6. Hyperplane

(그림 6)에서 A범주와 B 범주를 구분하기 위한 하이퍼분리면은 무수히 많다. 그러나 두 범주간의 점들의 거리를 최대화 되도록 학습을 시키게 되면, OSH (Optimum Separation Hyperplane)은 유일한 해로 존재한다. 즉, 하이퍼 분리면을 벡터의 내적으로 표현하면 식 10과 같고  $W$ 는 하이퍼분리면과 직교하는 가중치 벡터이고,  $b$ 는 기준 값이다.

$$(W \cdot X) + b = 0$$

$$W \in R^n \quad b \in R \dots\dots\dots (10)$$

SVM의 학습 단계는 하이퍼분리면들 중에서 최적 하이퍼분리면을 찾는 과정이고, 최적의 하이퍼분리면은 마진  $2 / \|W\|$ 가 최대가 되도록 하며,  $\|W\|/2$ 을 최소화 하는 문제가 된다. SVM의 테스트 단계는 임의의 n차원 벡터의 입력 객체  $X_i \in R^n$ 을 두 개의 범주  $y \in \{+1, -1\}$

로 예측하는 과정이고, 학습을 통해 계산된 최적 하이퍼분리면에 응답하는 분류 결정 함수  $f(X)$ 는 식 11과 같다.

$$f(X) = \sin((W \cdot X_i) + b) \dots\dots\dots (11)$$

선형 분리가 가능한 경우에는 식 12와 같은 최적화 문제가 된다.

$$\begin{aligned} \max \frac{2}{\|W\|} \Rightarrow \min \frac{\|W\|}{2} \\ \text{subject to } y_i(X_i \cdot W + b) - 1 \geq 0 \quad \forall i \dots\dots\dots (12) \end{aligned}$$

최적화 문제에서 함수의 최소 또는 최대 문제를 라그랑지(Lagrange) 함수로 유도하면 식 13과 같은 문제가 된다.

$$\begin{aligned} \min L_p = -\frac{1}{2} \|W\|^2 - \sum_{i=1}^N \alpha_i [y_i(W \cdot X_i + b) - 1] \\ \text{subject to } \alpha_i \geq 0, i = 1, \dots, N \dots\dots\dots (13) \end{aligned}$$

식 13에 대해서 울프 쌍대문제(Wolfe dual problem)로 유도하고 KKT(Karush-Kuhn-Tucker)조건을 적용하면, 식 15를 제약조건으로 하고 식 14를 최대화하는 최적화 문제가 된다[9,10].

$$\max L_D = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j X_i \cdot X_j \dots\dots\dots (14)$$

$$\text{subject to } \sum_{i=1}^N \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C \dots\dots\dots (15)$$

식 14, 15에서 라그랑지 계수  $\alpha_i$ 는 2차 프로그래밍 문제(Quadratic Programming: QP)의 해[6,7]이며,

$$W = \sum_{i=1}^N \alpha_i y_i X_i \text{ 이고, } C \text{는 마진 내의 객체를 허용하기}$$

위한 페널티 값이다.  $0 < \alpha_i < C$ 이면 SV가 되고,  $\alpha_i = C$ 이면 마진에 위치하고,  $\alpha_i = 0$ 이면 그밖에 위치한다.

비선형 SVM은 선형 분리가 불가능한 경우에 입력 공간을 분리하는 비선형 최적 결정면을 이용하기 위해 두 벡터

의 내적을 나타내기 위한 커널 함수로 입력 공간의 내적을 특징 공간의 내적으로 변형시켜 고차원으로 투영하여 계산을 용이하게 한다.

비선형 SVM 문제는 입력 데이터 공간을 커널 함수를 이용하여 보다 높은 특징 공간으로 맵핑하여 보다 높은 차원의 유클리드 공간으로 학습 데이터를 투영한다.

$$f(X_i) = \Phi(X_i) \cdot W + b \dots\dots\dots (16)$$

식 16은 하이퍼분리면의 일반식이고  $\Phi(X)$ 는  $X$ 의 기저함수로써 보다 높은 차원의 맵핑을 수행하는 함수이다. 식 17과 같이 두 벡터의 내적을 나타내기 위한 커널 함수  $K(X, X_i)$ 을 이용하여 투영시켜서 선형 분류문제로 해결한다.

$$K(X, X_i) = \Phi(X) \cdot \Phi(X_i) \dots\dots\dots (17)$$

맵핑함수를 사용한 라그랑지 듀얼 문제는 식 18, 19와 같다.

$$\max L_D = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \Phi(X_i) \cdot \Phi(X_j) \dots\dots\dots (18)$$

$$\text{subject to } \sum_{i=1}^N \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C \dots\dots\dots (19)$$

## 2.2 얼굴인식

본 논문에서는 검출된 얼굴영역을 인식하기 위해 멀티 SVM을 사용하였다. 즉, 모든 클래스에 대해 가능한 모든 쌍을 구성하여, 각각의 이진 SVM을 만든다. 각각의 이진 SVM 결과로 최대의 값을 선택하게 되고, 해당 클래스로 분류하게 된다. (그림 7)은 SVM을 이용한 멀티클래스 분류를 보여주고 있다.[11]

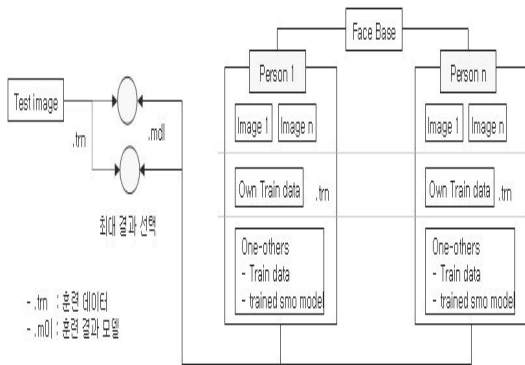


그림 7. SVM을 이용한 멀티클래스 분류  
Fig 7. Multiclass classification using SVM

### III. 음성인식

#### 3.1 음성 특징 추출

음성의 특징을 추출하는 방법은 화자의 음성에 대한 주파수 특성을 이용한다. 본 논문에서는 MFCC(Mel Frequency Cepstrum Coefficient)을 이용하여 특징을 추출했다. 이 방식은 인간의 청력이 일정 주파수 이상이 되면, 로그 형태로 되는 주파수 대역별 에너지 측정방법이다.

#### 3.2 음성 인식

음성 인식을 하기 위해 Hidden Markov Model(HMM)을 사용하였다. HMM은 다중 확률구조를 갖는 프로세스 모델에 적합하고 특정 사건의 의존 관계를 확률적으로 모델링할 수 있다. HMM은 Markov 체인에 은닉 상태를 추가하여 구성하게 된다.

HMM은 식 20과 같이 상태전이 확률, 관측 심볼의 확률분포, 초기 상태확률의 3가지 확률 파라미터로 정의된다.

$$\lambda = (A, B, \pi) \dots\dots\dots (20)$$

식 20에서 A는 상태전이 확률분포를 나타내는 상태전이 행렬이고 식 21과 같다.

$$A = [a_{ij}], \quad 1 \leq i \leq N \dots\dots\dots (21)$$

$a_{ij}$ 는 상태  $i$ 에서 상태  $j$ 로 전이될 확률이고  $N$ 은 상태 수이다. B는 관측 심볼의 발생확률 분포를 나타내며 다음 식 22와 같다.

$$B = [b_j(k)], \quad 1 \leq j \leq N, \quad 1 \leq k \leq M$$

$$b_j(k) = P(o_t = v_k | q_t = j) \dots\dots\dots (22)$$

$b_j(k)$ 는 현재 상태  $q_t$ 에서 관측 심볼 ( $o_t$ )이  $v_k$ 가 될 확률이고  $M$ 은 관측 심볼의 개수이다. 마지막으로  $\Pi$ 는 초기 상태 확률 분포이며, 식 23과 같다.

$$\Pi = [\pi_1, \pi_2, \pi_3, \dots, \pi_N] \dots\dots\dots (23)$$

HMM은 forward-backward 알고리즘, Viterbi 알고리즘, Baum-Welch 재추정 알고리즘을 이용하여 인식문제에 적용되어 진다. forward-backward 알고리즘은 관측열  $O = [O_1, O_2, O_3, \dots, O_T]$ 와 HMM  $\lambda = (A, B, \pi)$ 가 주어졌을 때, 관측 열을 발생시킬 확률  $P(O|\lambda)$ 를 계산하는 확률 계산법이다. Viterbi 알고리즘은 최적의 상태 열과 그 상태 열을 통한 확률을 구하는 데 사용된다.

HMM의 형태는 자신 및 모든 상태로 전이가 가능한 모델, 시간의 흐름에 따라 상태천이를 일정한 방향으로 진행하는 left-right model 등이 있다. 이 중 본 논문에서는 3-state left-right model을 사용하였다. 본 논문에서는 (그림 8)과 같이 left-to-right 모델을 사용했다.

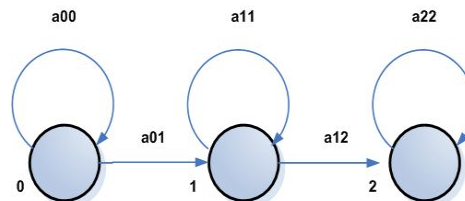


그림 8. 3-state left-to-right 모델  
Fig 8. 3-state left-to-right Model

### IV. 바이모달

바이모달이란 두 가지 정보를 결합한 것을 말한다. 본 논문에서는 두 가지 특징정보인 음성과 영상의 인식을 합성하였다.

영상 정보와 음성정보의 합성 방법으로는 크게 두 가지로 분류할 수 있다. 인식 전 얼굴의 파라미터와 음성의 파라미터를 합성하는 방법(early integration) [12]과 영상 정보와 음성 정보를 각각 인식한 후 가중치를 이용하여 인식 결과 값을 합성하는 방법(late integration)이 있다. 특징 추출 단계에서 합성하는 방법은 추출된 특징을 합성하여 새로운 특징을 생성하는 방법으로 이러한 방법은 두 개의 벡터의 값을 혼합하기 때문에 새로운 벡터의 인식을 위해 좋은 유사도 측정모델이 설계되어야 한다. 그러므로 특징벡터의 상관관계가 있는지의 여부도 확인하기 어렵고 인식률도 낮다는 단점이 있다. (그림 9)는 특징 추출 단계에서 합성하는 방법을 보여주고 있다.

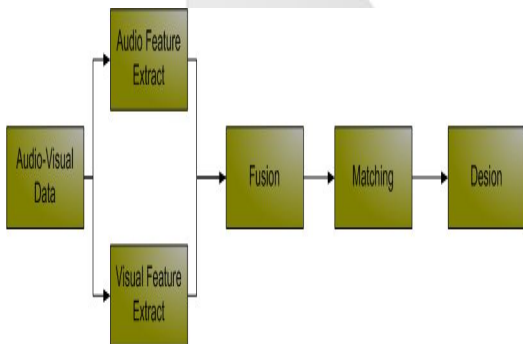


그림 9. 특징 추출 단계에서 합성하는 방법  
Fig 9. Early integration

본 논문에서는 각각의 인식률의 결과 값을 합성하는 방법을 사용하고 있다. 이 방법은 현재 인식의 합성방법 중 가장 좋은 성능을 나타내고 있다. (그림 10)은 본 논문에서 사용하고 있는 바이모달 구조이다.

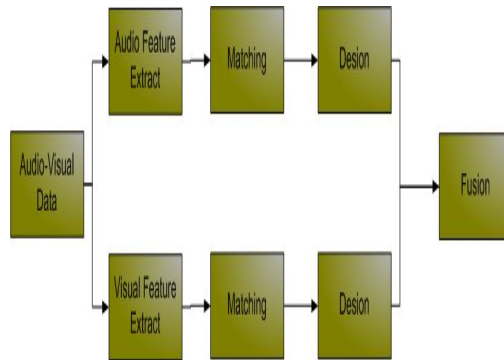


그림 10. 인식률의 결과 값을 합성  
Fig 10. Late integration

본 논문에서는 잡음에 따른 음성 인식률의 저하를 얼굴 인식률로 보상하였다. 여기서 잡음에 대한 음성 인식률을 보상하기 위해 얼굴 인식은 각각의 영상에 대해 인식된 결과에서 가장 많이 투표된 값을 선택하는 방법을 사용하였다.

### V. 실험 결과

#### 5.1 얼굴 검출 및 인식

얼굴 검출 및 인식 실험은 총 20명의 사람을 개인당 10장씩 영상을 추출하여 그 중 7장을 가지고 학습시켰다. 테스트는 개인당 3장씩 추출하여 테스트하였다.

본 논문에서 제안하는 얼굴 검출 방법은 <표 1>의 실험 결과와 같이 AdaBoost+PCA(60개)+SVM을 사용하여 얼굴 영역 검증을 수행하였다.

표 1 얼굴 검출 실험 결과  
Table 1. The result of Face detection

구분	AdaBoost + PCA(60) + SVM
검출률%	96.7
오검출률%	0.32
검출시간(ms)	15.1

또한 검출된 얼굴영역에서 PCA와 SVM을 통해 얼굴인식 실험을 하였다.

표 2 얼굴인식 실험결과  
Table 2. The result of Face recognition

구분	커널	SVM + PCA
	특징계수	60
	이미지 크기	32×32
FR	96.0	
FAR	1.1	
FFR	20	
평균 인식시간(ms)		27.5

### 5.2 음성 인식

본 논문에서는 음성인식 실험은 HTK(Hidden Markov Model Toolkit)을 이용하였다. 음성인식을 하기 위해 20명씩 10번 반복 발음하여 7번은 학습 데이터로 3번은 테스트 데이터로 이용하였다. 음성 데이터는 기본적으로 공간적 제약을 두지 않는 자연잡음이 들어 있는 상태에서 실험을 하였으며, 인위적인 잡음(Brown noise)을 점차 추가하면서 실험을 하였다.

표 3 음성인식 실험결과  
Table 3. The result of Audio recognition

잡음(Brown noise)	인식률(%)
noise-0.0	98.33
noise-0.5	95.00
noise-1.0	95.00
noise-1.5	85.00
noise-2.0	75.00
noise-2.5	43.33
noise-3.0	20.00

자연 잡음이 들어가 있는 상태로 인위적인 잡음을 추가하면 인식률이 감소됨을 알 수가 있다.

### 5.3 바이모달 시스템

본 실험에서는 바이모달 시스템은 음성과 얼굴 인식 결과를 합성하는 방법을 사용하였다. 음성의 잡음에 의한 얼

굴 보상방법으로는 인식된 각각의 영상의 결과에서 가장 많이 투표된 값을 선택하는 방법을 사용하였다. <표 4>는 얼굴과 음성 합성에 의한 방법은 단일 인식보다 인식률의 향상을 보여주고 있다.

표 4 바이모달 실험결과  
Table 4. The result of Bimodal

잡음(Brown noise)	바이모달 인식률(%)
noise-0.0	100%
noise-0.5	100%
noise-1.0	100%
noise-1.5	97.2%
noise-2.0	96.00%
noise-2.5	96.00%
noise-3.0	96.00%

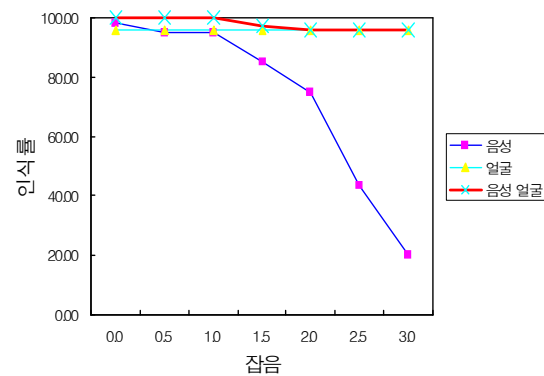


그림 11. 인식결과  
Fig 11. The result of recognitions

## VI. 결론 및 향후 과제

본 논문에서는 얼굴 정보와 음성 정보를 이용하여 바이모달 시스템을 구축하였다. 얼굴 인식 및 음성 인식을 각각 따로 추출 하였고, 각각 인식된 결과의 합성으로부터 바이



모달 시스템을 구축하였다. 인식결과, 단일 인식보다는 얼굴과 음성 인식률을 합성함으로써, 인식을 향상이 됨을 알 수가 있었고, 잡음 환경에서 더욱 더 높은 성능을 나타냄을 알 수 있었다. 향후 각각의 인식률에 대한 보완이 더 필요하고, 단순한 합성이 아닌 효과적으로 합성하는 방법에 대한 연구가 필요하다.

## 참고문헌

- [1] M. N. Kannak, Q. Zhi, A. D. Cheok, and K. C. Chung "Audio-Visual Modeling For Bimodal Speech Recognition", 2001 IEEE International Conference, Vol. 1, 7-10 Oct. 2001 pp.181 - 186
- [2] S. Meng, and Y. Zhang "A Method Visual Speech Feature Area Localization", IEEE Neural networks and Signal Processing, Dec. 2003
- [3] N. A. Fox, R. Gross, P. D. Chalzal, J. F. Cohn, and R. B. Reilly, "Person Identification Using Automatic Integrated of Speech, and Face Experts", WBMA'03, Nov. 2003
- [4] R. Lienhart, and J. Maydt, "An extended set of Harr-like features for rapid object detection," Image Proceedings. 2002 International Conference on, vol. 1, 22-25, pp.I-900 - I-903, Sept. 2002
- [5] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," Journal of Computer and System Sciences, vol. 55, no. 1, pp. 119-139, 1997.
- [6] R. Lienhart, A. Kuranov, and V. Pisarevsky, "Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection," DAGM'03, 25th Pattern Recognition Symposium, Madgeburg, Germany, pp.297-304, Sep. 2003.
- [7] Jian Yang, and Jing-ju Yang, "Why can LDA be performed In PCA transformed space?", Patter Recognition 36, pp.563-566, 2003.
- [8] Center for Biological and Computational Learning at MIT, "MIT CBCL DATASETS," <http://cbcl.mit.edu/cbcl/software-datasets>, 2003.
- [9] HumanScan AG, "BioID Face Database", <http://www.humanscan.de/support/downloads/facedb.php>, 2003
- [10] V. Vapnik, "The Nature of Statistical Learning Theory," h Springer-verlag, New York, 1995.
- [11] 이호근, 김명훈, 이지근, 정성태 "SVM-SMO와 Pan-tilt 웹카메라를 이용한 실시간 얼굴 추적과 얼굴인식", 한국정보과학회, 제 31권 제 2호, 2004
- [12] C. C. Chibelushi, "Feature-Level Data Fusion for Bimodal Person Recognition", Image Processing and Its Application, Vol. 1 pp 399-403 1997

저 자 소개



김 명 훈

2004년 2월 원광대학교 전기전자및  
정보공학부(공학사)  
2004년 3월~현재 원광대학교 컴퓨터  
공학과 석사과정  
<관심분야> 영상인식 컴퓨터 비전,  
영상처리



이 지 근

2000년 2월 원광대학교 컴퓨터공학과  
(공학사)  
2002년 2월 원광대학교 컴퓨터 공  
학과(공학 석사)  
2002년 3월~현재 원광대학교 컴퓨  
터 공학과 박사과정  
<관심분야> 영상인식 영상인식,  
컴퓨터 비전, 영상처리



소 인 미

1994년 2월 원광대학교 컴퓨터공학과  
(공학사)  
2002년 8월 원광대학교 정보컴퓨터  
교육전공(교육학석사)  
2004년 3월~현재 원광대학교 컴퓨터  
공학과 박사과정  
<관심분야> 영상인식 컴퓨터 비전,  
영상처리



정 성 태

1987년 2월 서울대학교 컴퓨터공학과  
(공학사)  
1989년 2월 서울대학교 컴퓨터공학과  
(공학석사)  
1994년 8월 서울대학교 컴퓨터공학과  
(공학박사)  
1995년 3월 현재 원광대학교 전기  
전자및정보공학부 교수  
<관심분야> 영상인식, 영상처리, 컴  
퓨터 그래픽스

