

# 데이터 퓨전을 이용한 얼굴영상 인식 및 인증에 관한 연구

## 2D Face Image Recognition and Authentication Based on Data Fusion

박성원 · 권지웅\* · 최진영\*

Sung Won Park, Ji Woong Kwon and Jin Young Choi

한국 IBM 소프트웨어연구소

\*서울대학교 전기컴퓨터공학부, 자동화시스템공동연구소

### 요 약

얼굴인식은 이미지의 밝은 변동(표정, 조명, 얼굴의 방향 등)으로 인해 한 가지 인식 방법으로는 높은 인식률을 얻기 어렵다. 이러한 어려움을 해결하기 위해, 여러 가지 정보를 융합시키는 데이터 퓨전 방법이 연구되었다. 기존의 데이터 퓨전 방법은 보조적인 생체 정보(지문, 음성 등)를 융합하여 얼굴인식기를 보조하는 방식을 취하였다. 이 논문에서는 보조적인 생체 정보를 사용하지 않고, 기존의 얼굴인식방법을 통해 얻어지는 상호보완적인 정보를 융합하여 사용하였다. 개별적인 얼굴인식기의 정보를 융합하기 위해, 전체적으로는 Dempster-Shafer의 퓨전이론에 근거하면서, 핵심이 되는 질량함수를 새로운 방식으로 재정의한 퓨전모델을 제안하였다. 제안된 퓨전모델을 사용하여 개별적인 얼굴인식기의 정보를 융합한 결과, 보조적인 생체정보 없이, 개별적인 얼굴인식기보다 나은 인식률을 얻을 수 있었다.

### Abstract

Because face images have many variations(expression, illumination, orientation of face, etc), there has been no popular method which has high recognition rate. To solve this difficulty, data fusion that fuses various information has been studied. But previous research for data fusion fused additional biological information(fingerprint, voice, etc) with face image. In this paper, cooperative results from several face image recognition modules are fused without using additional biological information. To fuse results from individual face image recognition modules, we use re-defined mass function based on Dempster-Shafer's fusion theory. Experimental results from fusing several face recognition modules are presented, to show that proposed fusion model has better performance than single face recognition module without using additional biological information.

**Key Words** : 얼굴인식, 데이터 퓨전, Dempster-Shafer 이론, 질량함수

### 1. 서 론

얼굴인식[2]은 현재 가장 활발하게 연구가 진행되고 있는 주제 중 하나로서 다양한 방식이 제안되었지만, 그 중 실용화할 만큼 충분히 높은 인식률을 보이는 한 가지 방법이 없다. 얼굴영상은 조명의 밝기 및 얼굴의 방향과 표정에 따라 변이가 컸 2-D 얼굴영상인식은 전형적인 ill-posed problem이다. 따라서 얼굴영상인식은 데이터 퓨전[7,12], 즉 각기 다른 장단점을 가진 두 개 이상의 독립된 시스템들로부터 산출된 결정들을 융합하여 더 나은 결정을 획득하는 이론을 이용해볼 만한 주제이다. 현재의 얼굴영상인식에 관한 연구는 가장 성능이 좋다고 평가되는 은닉마코브모델[15,16,17]이나 eigenface [20] 이론 등을 단독으로 적용하여 인식률을 개선시키려는 시도가 일반적이며, 데이터 퓨전 등을 이용하여 다

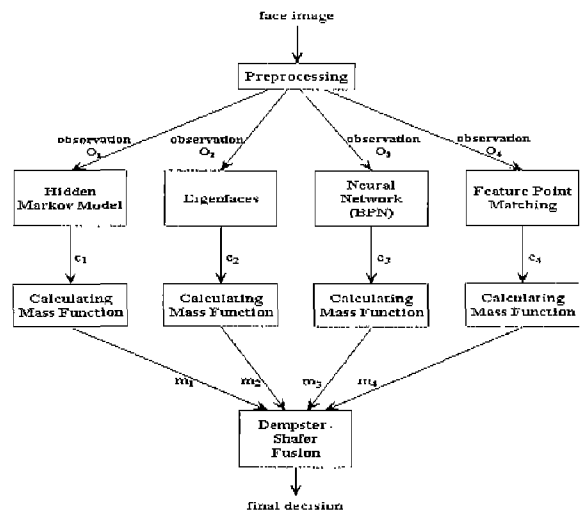


그림 1. 4개의 모듈을 갖는 제안된 퓨전 시스템  
Fig 1. Proposed data fusion system having 4 modules

접수일자 : 2001년 05월 19일

완료일자 : 2001년 07월 31일

감사의 글 : 본 연구는 두뇌한국21사업과 뇌과학연구사업의 지원에 의해 이루어졌습니다.

수의 독립된 얼굴인식 시스템을 융합하려는 시도는 드물었다. 데이터 퓨전을 직용한 연구사례의 대부분은 지문인식 및 음성인식 등 얼굴인식보다 인식률이 높은 생체인식법과 얼굴인식을 결합[3,19]하여 전반적인 개인인식성능을 향상시킨 것일 뿐, 얼굴인식 자체의 성능을 향상시킨 것은 아니다.

이와 달리, 이 논문에서는 가장 일반적으로 사용되는 데이터 퓨전이론의 하나인 Dempster-Shafer 퓨전이론[6,18]을 이용하여 얼굴영상인식 시스템의 성능을 향상시킬 수 있음을 보인다. 또한 본 논문에서는 개별 모듈의 신뢰도를 반영하는 질량함수(mass function) 퓨전 규칙을 제안하고 향상된 인식결과를 제시함으로써 Dempster-Shafer 퓨전이론을 확장하였다.

개별 모듈로는 은닉마코브모델, eigenfaces, 오류역전파신경망[5,11,13], 특징점추출법[4,8,9,10,14]의 4가지 얼굴영상인식 모듈을 사용하였다(그림 1). 이 4개의 모듈은 얼굴영상의 각각 다른 종류의 변이에 강인하다는 특징을 가지고 있으므로 데이터 퓨전을 적용함으로써 더 향상된 성능을 얻을 수 있음을 보인다. 이 논문에서는 개별 모듈의 성능향상 보다는 퓨전시스템과 개별 모듈의 인식률을 비교해보는 것을 목표로 하였다.

$$\text{여기서 } k = \sum_{B_1 \cap B_2 \cap \dots \cap B_n = \emptyset} m_1(B_1)m_2(B_2)\dots m_n(B_n)$$

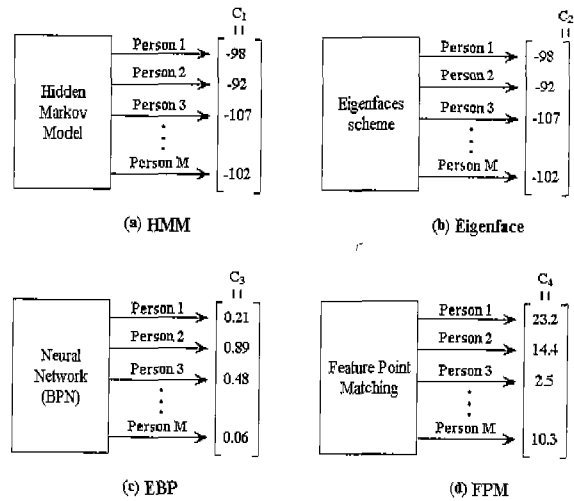


그림 2. 각 모듈별 출력값  
Fig 2. Outputs from each module

## 2. Overview : Dempster-Shafer 퓨전이론에 관한 이론적 배경

Dempster-Shafer 퓨전이론[1,2,7]은 다수의 모듈에서 산출된 서로 독립된 질량함수(mass function)가 정의되어있을 때[6,18], 퓨전시스템의 질량함수를 구하는 이론이다. 주어진 입력패턴이 각 클래스에 속한다고 볼 수 있는 신뢰도(belief)를 클래스별로 정량화한 것이다. 각각의 모듈은 클래스 별로 서로 다른 신뢰도 및 질량함수값을 가지게 된다.

서로 배제적인 명제들(mutually exclusive propositions)로 이루어진 전체 집합 $\Theta$ 이라 하고,  $2\Theta$ 를  $\Theta$ 의 powerset으로써 모든 가능한 명제들의 조합을 포함한다고 하면 질량함수  $m$ 은 아래와 같은 성질을 갖는다.

$$m: 2^\Theta \rightarrow [0, 1], \quad m(\emptyset) = 0, \\ 0 \leq m(A) \leq 1 \text{ and } \sum_{A \in \Theta} m(A) = 1$$

위의 4가지 성질만 만족하도록 질량함수를 정의하면, 퓨전이 가능하기 때문에, 질량함수를 정의하는 방법은 여러 가지가 가능하고, 따라서 질량함수의 정의방법은 전체 퓨전시스템의 성능을 결정하는 중요한 요소이다. 이 논문에서 사용된 질량함수 정의방법은 4장에서 소개된다.

개별 명제들의 조합  $A$ 에 대한 정량화된 총 신뢰도를 할당할 값인 신뢰도 함수  $bel$ 은 아래와 같이 정의된다.

$$bel(A) = \sum_{X \subseteq A} m(X), \text{ for all } A \subseteq \Theta \quad (2.1)$$

그리고,  $n$ 개의 모듈로부터 독립적으로 산출된 질량함수값  $m_1, m_2, \dots, m_n$ 을 퓨전한 퓨전시스템 전체의 질량함수값은 다음과 같다.

$$m(A) = \frac{\sum_{B_1 \cap \dots \cap B_n = A} m_1(B_1)m_2(B_2)\dots m_n(B_n)}{1 - k} \quad (2.2) \\ = \frac{\sum_{B_1 \cap \dots \cap B_n = A} m_1(B_1)m_2(B_2)\dots m_n(B_n)}{\sum_{B_1 \cap \dots \cap B_n \neq \emptyset} m_1(B_1)m_2(B_2)\dots m_n(B_n)}$$

## 3. 질량함수

각각의 모듈의 출력값  $c_i = [c_{i1}, c_{i2}, \dots, c_{iM}]^T$  ( $i = 1, \dots, M$ 은 클래스의 수)는 각기 다른 범위를 갖고, 크기에 따른 신뢰도도 각기 다르다(그림 2). 따라서, 각각의 모듈의 출력값이  $[0, 1]$ 의 범위에 놓이고, 크기에 따른 신뢰도의 의미를 갖게 하기 위해,  $c_i$ 를 아래와 같은 규칙에 의해 선형변환(linear transformation)시킨  $c_i' = [c_{i1}', c_{i2}', \dots, c_{iM}']^T$ 를 구한다.

- I) 모든  $c_i'$ 는  $[0, 1]$ 사이 존재한다.
- II)  $c_{ij}$ 의 가장 큰 값은  $c_{ij}'=1$ 로 변환한다.
- III)  $c_{ij}$ 의 가장 작은 값은  $c_{ij}'=0$ 로 변환한다.
- IV) 해당 클래스의 신뢰도가 높을수록 큰 값으로 변환한다.

이때,  $i$ 번째 모듈의 질량함수  $m_i(A)$ 는 다음과 같이 정의된다[9].

$$m_i(A) = \begin{cases} \frac{c_{ij}'}{n} & \text{if } A = \{C_j\} \text{ for } j \in KL \\ 1 - \sum_{j \in KL} m_i(\{C_j\}) & \text{if } A = \Theta - \bigcup_{j \in KL} \{C_j\} \\ 0 & \text{otherwise} \end{cases}$$

여기서,  $KL$ 은  $c_i'$ 의  $k$ 개의 가장 큰 요소들의 색인(index)의 집합을 의미한다.

이렇게 각 모듈에 대한 질량함수를 구하면, 식(2.2)에 의하여 각 클래스에 해당되는 퓨전 시스템의 질량함수  $m(\{C_j\})$  ( $j=1, \dots, M$ ,  $M$ 은 클래스 개수)를 구할 수 있다. 주어진 입력패턴이 클래스  $C_j$ 에 속할 신뢰도는 식(2.1)에 의하여  $bel(\{C_j\})=m(\{C_j\})$ 로 구해진다.

## 4. Dempster-shafer 이론의 확장

본 논문에서는 퓨전 시스템에 사용된 개별 모듈의 인식률에 따라 질량함수를 다르게 재정의하여 보다 신뢰성있는 모

들의 결정이 전체 퓨전 시스템의 결정에 더 많이 반영되도록 한다면, 더 효율적인 퓨전 시스템이 될 것이라는 가정하였다. 즉, 인식률이 높은 모듈의 경우 질량함수의 값을 클래스에 따라 차이를 크게 주어, 신뢰도가 높은 클래스 한 개 또는 두 개에 대해서만 상대적으로 큰 질량함수 값을 갖도록 한다. 반대로 인식률이 낮은 모듈의 경우는 각각의 클래스에 할당된 질량함수의 차이가 작아지도록 변환한다.

$i$ 번째 개별 모듈이, 입력영상이 속할 가능성 순으로 선출한  $R$ 개의 후보 클래스 안에 실제로 올바른 클래스가 속할 확률을  $P_i(R)$ 로 정의한다.

$$P_i(R) = \sum_{j=1}^R p(j), \quad R=1, 2, \dots, 20$$

따라서  $P_i(R)$ 은 곧  $i$ 번째 개별 모듈의 성능과 동일하며, 개별 모듈의 성능 테스트를 통해 퓨전 이전에 미리 알 수 있는 값이다.

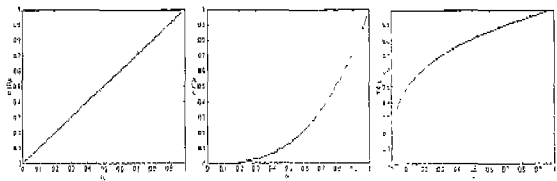
이제 기존의 질량함수  $m_i(A)$ 가 주어졌을 때 모듈의 성능에 따른 가중치를 반영한 새로운 질량함수  $m_i'(A)$ 를 재정의한다.

$$m_i'(A) = \frac{(am_i(A))^{\beta(P_i(R)-\gamma)}}{K} \quad (4.1)$$

여기서  $a > 1, \beta > 0, \gamma < P_i(R)$ ,

$$\text{그리고, } K = \sum_{A \in \Theta} (am_i(A))^{\beta(P_i(R)-\gamma)}$$

$a$ 의 역할은  $m_i(A)$ 의 값을  $[0, a]$  사이의 더 넓은 영역에 선형적으로 재할당하는 것이며,  $\gamma$ 의 역할은 성능이 높은 모듈, 즉  $P_i(R)$  값이 큰 모듈과 그렇지 않은 모듈을 구분하기 위한 경계값을 주는 것이다.  $P_i(R)$ 가 상대적으로 크면, 즉  $\beta(P_i(R)-\gamma)$ 가 1보다 크면, 기존의 질량함수값이 큰 몇 개의 클래스에만 더 큰 질량함수값을 할당하고,  $\beta(P_i(R)-\gamma)$ 가 1보다 작은 경우에는 기존의 질량함수값에 비해 재할당된 질량함수값의 격차를 더 작게 준다.



(a) the basic rule (b) in believable module (c) in unbelievable module

$x$ 축: 기존의 질량함수 값  $m_i(C_j)$      $y$ 축:  $(am_i(A))^{\beta(P_i(R)-\gamma)}$

그림 3. 재정의된 질량함수

Fig 3. re-defined mass function

그림 3은  $\alpha=10, \beta=10, \gamma=0.5$ 로 결정되었을 경우 기존의  $m_i(A)$ 값이 어떻게 재할당되는지를 보여주는 그림이다. 개별 모듈의 신뢰도를 반영하는  $P_i(R)$ 의 값이 큰 경우, 그림 3.(b)에 의해 기존의 질량함수값이 큰 몇 개의 클래스와 그렇지 않은 클래스간의 질량함수값의 격차가 커지게 됨을 알 수 있다. 따라서, 식 4.1에 의해  $K$ 의 값으로 정상화(normalization)를 시키면, 결과적으로 기존의 질량함수값이 큰 몇 개의 클래스는 더욱 큰 값을 가지게 된다. 마찬가지로,  $P_i(R)$ 의 값이 작은 경우, 그림 3.(c)에 의해 기존의 질량함수가 큰 클래스와 그렇지 않은 클래스간의 질량함수값의 상대적인 차이가 적어져, 결과적으로 기존의 질량함수값이

큰 몇 개의 클래스는 전보다 작은 값을 갖게 된다.

## 5. 실험결과

조영·표정·얼굴방향에 변이를 준  $120 \times 100$  픽셀의 흑백 얼굴영상 데이터베이스를 사용하였다. 총 20명에 대한 얼굴 영상을 이용하였으며, 1인당 30개의 얼굴영상을 취득하여, 그 중 20개는 학습을 위한 영상으로, 나머지 10개의 얼굴영상은 테스트를 위해 사용하였다.

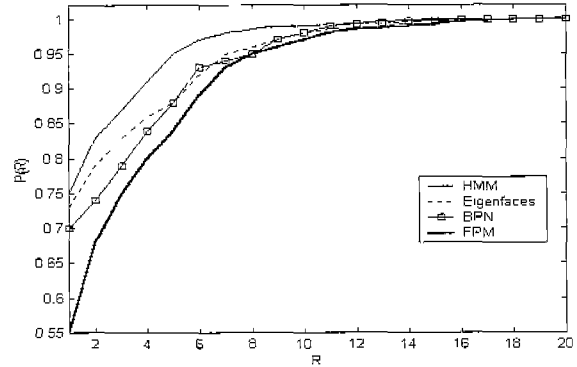


그림 4. 개별 모듈의 성능

Fig 4. Performance of each module

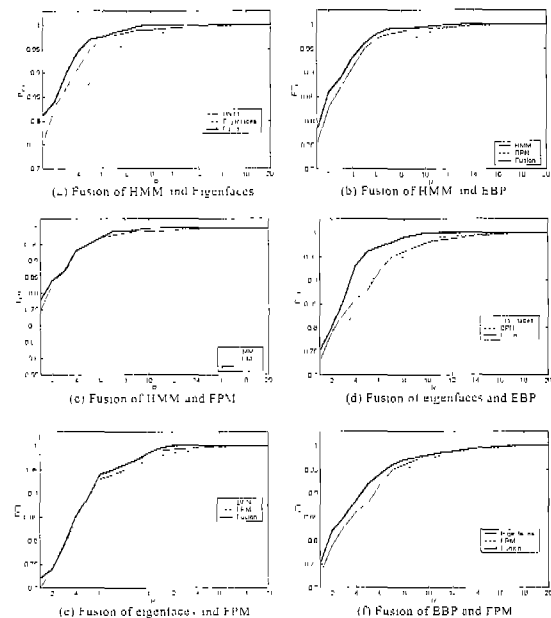


그림 5. 2개 모듈 퓨전 시스템

Fig 5. Fusion of 2 modules

### 5.1 개별 모듈의 인식률

개별 모듈의 인식률은 그림 4에서 나타나있다. HMM은 4개의 모듈 중 가장 뛰어난 성능을 보였는데, 특히 얼굴방향과 표정 및 조영의 강도에 대해서 강한 특성을 보였다. 두 번째로 인식률이 높은 모듈인 eigenfaces의 경우 정면얼굴에 대해서는 가장 높은 인식률을 보였지만 얼굴방향에 변이가

있는 영상에 대해서는 인식률이 급격히 떨어지는 단점을 보였다. 오류역전파신경망의 경우 학습에 사용된 것과 비슷한 영상을 테스트하면 비교적 큰 변이를 갖는 영상일지라도 잘 인식하였지만, 학습에 비슷한 영상이 없는 경우는 4개의 모듈 중 가장 인식률이 낮았다. 따라서 학습 데이터의 선정에 인식률이 매우 의존적인 결과를 보였다.

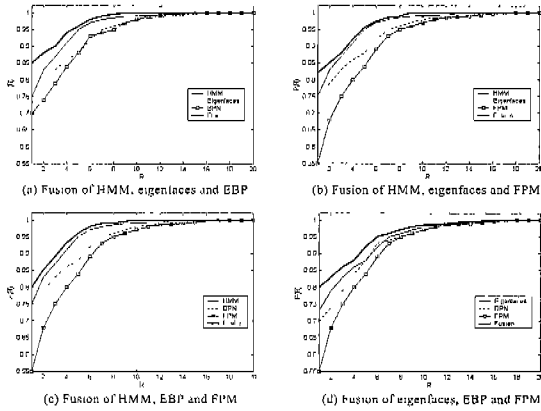


그림 6. 3개 모듈 퓨전 시스템  
Fig 6. Fusion of 3 modules

마지막으로 특징점 추출법은 가장 인식률이 낮았는데, 특히 인식률의 개인차가 매우 컸다. 정확한 3-D 머리 모델 구축에 실패한 개인의 경우 얼굴영상의 변이의 종류에 무관하게 인식률이 전반적으로 낮은 결과를 보이는 것으로 분석할 수 있다.

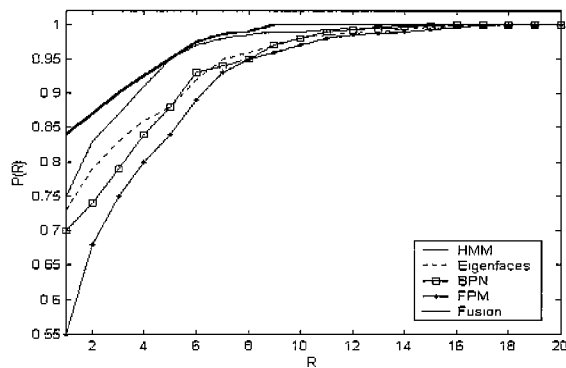


그림 7. 4개 모듈 퓨전 시스템  
Fig 7. Fusion of 4 modules

5.2. 퓨전 시스템의 성능

개별 모듈이 가장 가능성있는 후보로 뽑은 클래스가실제로 입력영상이 속한 클래스일 확률 P(1)는 은닉마코브모델이 0.75, eigenfaces가 0.73, 오류역전파신경망이 0.68, 특징점추출법이 0.5였다. 식(4.1)에서  $\alpha=10, \beta=10, \gamma=0.5$ 로 고정시켰을 때 전체 퓨전 시스템의 인식률이 가장 높았고,  $\beta(P_i(1)-\gamma)$ 를 계산한 값은 은닉마코브모델이 2.8, eigenfaces가 2.3, 오류역전파신경망이 1.8이었고 특징점추출법은 0.50으로, 특징점추출법만이 1.0 이하의 신뢰할 수 없는 모듈로 선택되었음을 알 수 있다.

퓨전 시스템은 어떤 모듈을 몇 개씩 퓨전하는지에 따라

성능이 달라짐을 알 수 있다. 총 4개의 모듈 중 2개, 3개, 4개의 모듈을 퓨전한 결과를 각각 그림 5, 6, 7에 나타내었다. 4개 중 어떤 모듈을 몇 개를 선택하여 퓨전하든, 모든 조합에서 퓨전 이전의 개별 모듈에 비해 퓨전 이후의 성능이 향상된 것을 알 수 있다. 퓨전에 사용된 모듈 수가 증가할수록 인식률은 전반적으로 높아졌다. 특히 모듈 수가 증가할수록 가장 낮은 인식률을 보였던 특징점 추출법이 전체 퓨전시스템에 미치는 악영향이 줄어들어 안정된 얼굴인식 시스템이 구축되었음을 알 수 있다.

전반적으로 좋은 결과를 산출한 3개 또는 4개의 모듈을 퓨전한 결과와 달리, 2개의 모듈만을 퓨전한 결과는 모듈 선택에 따라서 인식률의 편차가 컸다. 2개 모듈의 퓨전시스템 중에서는 은닉마코브모델과 eigenfaces의 퓨전이 가장 성능이 좋았다. 개별 모듈에 비해 가장 주목할만한 향상을 보인 것은 그림 5.(d)이다. 두 모듈을 퓨전한 경우, 성능이 가장 우수한 모듈인 은닉마코브모델을 퓨전 시스템에 추가한 경우보다는 낮은 인식률을 보였지만, 개별 모듈 각각의 성능에 비해 퓨전 시스템의 성능 향상의 폭이 가장 컸다. 이는 오류역전파신경망과 eigenfaces가 각각 강인한 얼굴영상 변이의 종류가 서로 달라서 상호보완적인 성능을 갖기 때문이다. 대체로 eigenfaces가 오류역전파신경망에 비해 인식률이 높지만, eigenfaces는 특히 얼굴방향의 변이에 강인하지 못한 단점을 가지고 있다. 이에 반해 오류역전파신경망은 학습 데이터 선정에 의해 얼굴방향의 변이에 강인하게 할 수 있기 때문에 두 모듈의 퓨전 시스템은 퓨전의 시너지 효과가 더욱 크다.

6. 결론 및 요약

이 논문에서는 Dempster-Shafer 퓨전이론에 기반하여 4개의 독립된 얼굴인식모듈을 융합한 퓨전시스템을 제안하였고, 2~4개의 모듈을 퓨전한 결과 가장 성능이 뛰어난 단일 모듈보다 더 나은 모듈을 얻었다. 대체로, 퓨전 시스템에 사용된 개별 모듈의 성능이 더 우수할 때 퓨전 시스템의 성능도 우수했지만, 퓨전시스템에 사용된 모듈의 수가 늘어날수록, 성능이 나쁜 모듈이 내린 잘못된 결정의 영향을 줄어 들었다. 또한 개별 모듈들이 서로 상이한 변이를 갖는 입력영상을 인식함에 있어서 강인한 경우 퓨전 시스템은 더욱 주목할 만한 성능향상을 보였다.

참고 문헌

[1] 박성원, 데이터 퓨전을 이용한 얼굴영상 인식 및 인증에 관한 연구, 서울대학교 석사학위논문, 2000. 1.  
 [2] R. Chellappa, C. Willson, and S. Sitohey, Human and machine recognition of faces : A survey, *Proceedings of IEEE*, vol. 83, 1995.  
 [3] C. C. Chibelushi, F. Deravi, and J. S. Mason, Audio-visual person recognition: an evaluation of data fusion strategies, 1997 European Conference on Security and Detection, 1997.  
 [4] C. C. Chibelushi, J.S.D. Mason, and F. Deravi, Feature-level data fusion for bimodal person recognition, 1997 Sixth International Conference on Image Processing and Its Applications, vol. 1, pp. 399 403, 1997.

[5] C. L. Giles, A. C. Tsoi, A. D. Back, Face Recognition : A Convolutional Neural Network Approach, *IEEE trans. Neural Network*, vol. 8, no. 1, pp. 98-113, 1997.

[6] J. W. Guan and D. A. Bell, *Evidence Theory and its application*, vol. 1, North-Holland, 1991.

[7] D. L. Hall and J. Llinas, An introduction to multi-sensor data fusion, *Proceedings of the IEEE*, vol. 85 Issue. 1, pp. 6-23, 1997.

[8] X. Jia and M. S. Nixon, Extending the feature vector for automatic face recognition, *IEEE trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 12, pp. 1167-1176, 1995.

[9] K. M. Lam and H. Yan, Locating and Extracting the eye in human face images, *Pattern Recognition*, vol. 29, no. 5, pp. 771-779, 1996.

[10] K.-M. Lam and H. Yan, An analytic-to-holistic approach for face recognition based on a single frontal view, *IEEE trans. Pattern Analysis and Machine Intelligence*, vol. 20, No. 7, pp. 673-686, 1998.

[11] S. H. Lin, S. Y. Kung, Face recognition/detection by probabilistic decision-based neural network, *IEEE trans. Neural Network*, vol. 8, no. 1, pp. 114-140, 1997

[12] R. C. Luo and M. G. Kay, A tutorial on multisensor integration and fusion, Industrial Electronics Society, IECON '90., 16th Annual Conference of IEEE, vol. 1, pp. 707-722, 1990.

[13] A. R. Mirhosseini, H. Yan, Neural Networks for learning human facial feature from labeled graph models, Intelligent Information Systems, 1996., Australian and New Zealand Conference on, pp. 170-173, 1996.

[14] A. R. Mirhosseini, H. Yan, and K. M. Lam, Adaptive deformable model for mouth boundary detection, *Optical Eng.*, vol. 37, no. 3, pp. 869-875, 1998.

[15] A. V. Nefian and M. H. Hayes, Hidden Markov Model for face recognition, in ICASSP 98, vol. 5, pp. 2721-2724, 1998.

[16] L. R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE*, vol. 77, issue 2, pp. 257-286, 1989.

[17] F. Samaria and S. Young, HMM-based architecture for face identification, *Image and Vision Computing*, vol. 12, no. 8, pp. 537-543, 1994.

[18] G. G. Shafer, *Mathematical Theory of Evidence*. Princeton, NJ : Princeton Univ. Press, 1976.

[19] K. Takahashi, H. Yamasaki, "Audio-visual sensor fusion system for intelligent sound sensing," Multisensor Fusion and Integration for Intelligent Systems, 1994. *IEEE International Conference on MFI '94*, pp. 493-500, 1994.

[20] M. Turk and A. Pentland, Face recognition using eigenfaces, *Proceedings of International Conference on Pattern Recognition*, pp. 586-591, 1991.

## 저자 소개



### 박성원 (Sung Won Park)

1999년 : 서울대학교 전기공학부 졸업  
 2001년 : 동대학원 전기공학부 석사 졸업  
 2001년~현재 : 한국IBM 소프트웨어연구소 연구원

관심분야 : 패턴인식, 신경회로망, 머신비전



### 권지웅 (Ji Woong Kwon)

2000년 : 서울대학교 전기공학부 졸업  
 2000년~현재 : 동대학원 전기공학부 석사 과정

관심분야 : 패턴인식, 신경회로망, 머신비전, 유전알고리즘



### 최진영 (Jin Young Choi)

한국 퍼지 및 지능시스템학회 논문지 제7권 제4호 참조  
 현재 : 서울대학교 전기공학부 부교수