

# 다중 센서 융합 알고리즘을 이용한 운전자의 감정 및 주의력 인식 기술 개발

## Development of Driver's Emotion and Attention Recognition System using Multi-modal Sensor Fusion Algorithm

한철훈 · 심귀보\*

Cheol-Hun Han and Kwee-Bo Sim\*

중앙대학교 전자전기공학부

### 요 약

최근 자동차 산업 및 기술이 발전함에 따라 기계적인 부분에서 서비스적인 부분으로 관심이 점점 바뀌고 있는 추세이다. 이와 같은 추세에 발맞추어 운전자에게 보다 안정적이며 편리한 운전 환경을 조성하기 위한 방법으로 감정 및 인지 인식에 대한 관심이 점점 높아지고 있다. 감정 및 주의력을 인식하는 것은 감정공학 기술로서 이 기술은 1980년대 후반부터 얼굴, 음성, 제스처를 통해 인간의 감정을 분석하고 이를 통해 인간 친화적인 서비스를 제공하기 위한 기술로 연구되어 왔다. 이와 같은 기술을 자동차 기술에 접목시키고 운전자의 안정적인 주행을 돕고 운전자의 감정 및 인지 상황에 따른 다양한 서비스를 제공할 수 있다. 또한 Real-Time으로 운전자의 제스처를 인식하여 졸음운전이나 부주의에 의한 사고를 사전에 예방하고 보다 안전한 운전을 돕는 서비스가 필요시 되고 있다. 본 논문은 운전자가 안전 운전을 하기 위해 생체-행동 신호를 이용하여 감정 및 졸음, 주의력의 신호를 추출하여 일정한 형태의 데이터베이스로 구축하고, 구축된 데이터를 이용하여 운전자의 감정 및 졸음, 주의력의 특징 점들을 검출하여, 그 결과 값을 Multi-Modal 방법을 통해 융합함으로써 운전자의 감정 및 주의력 상태를 인식할 수 있는 시스템을 개발하는데 목표를 두고 있다.

### Abstract

As the automobile industry and technologies are developed, driver's tend to more concern about service matters than mechanical matters. For this reason, interests about recognition of human knowledge and emotion to make safe and convenient driving environment for driver are increasing more and more. recognition of human knowledge and emotion are emotion engineering technology which has been studied since the late 1980s to provide people with human-friendly services. Emotion engineering technology analyzes people's emotion through their faces, voices and gestures, so if we use this technology for automobile, we can supply drivers with various kinds of service for each driver's situation and help them drive safely. Furthermore, we can prevent accidents which are caused by careless driving or dozing off while driving by recognizing driver's gestures. the purpose of this paper is to develop a system which can recognize states of driver's emotion and attention for safe driving. First of all, we detect a signals of driver's emotion by using bio-motion signals, sleepiness and attention, and then we build several types of databases. by analyzing this databases, we find some special features about drivers' emotion, sleepiness and attention, and fuse the results through Multi-Modal method so that it is possible to develop the system.

Key Words : Multi-Modal Sensor Fusion, 감정인식, 안전운전시스템, 부주의 운전, PCA, AdaBoost, 제스처 인식

### 1. 서 론

자동차의 수가 점점 증가함에 따라 교통사고도 그 만큼 증가하고 있다. 교통사고의 주요 원인 중 하나가 운전자의 감정을 제어하지 못하는 데에 있는데 이와 같은 경우를 방지하기 위해 현재 운전자의 감정 상태를 인식하기 위한 연구가 일본, 미국, 스위스 등 선진국과 우리나라에서도 활발

히 이루어지고 있다[1].

스위스의 취리히에 본사를 둔 RInspeed사에 만든 'Senso'라는 이름을 가진 자동차는 바이오투드 자료를 이용하여 사람의 마음 상태에 적응하도록 설계되었다. 운전자가 난폭운전을 하는 것이 감지되면 차량 내부를 마음을 진정시키는 블루 패턴의 색상으로 만들고 긴장을 풀어주는 음악이나 향기를 내뿜어 준다. 또 운전자가 졸고 있다고 감지하면 의자를 흔들거나 내부를 옐로나 오렌지 패턴으로 바꾸고 내부에 감귤 향을 뿌려준다. 일본 Toyota와 Sony는 감정을 가진 자동차를 개발해서 'Pod'라고 명명했다. Pod는 자동차의 주행거리(성능)보다는 운전자의 무드를 주로 담당하며, 신체 신호를 모니터하고, 긴장하고 있는지 또는 화난 상태인지를 감지하여 비정상적인 운전을 할 경우에는 경고해 준다. 운전자가 너무 급격히 제동하거나 흔들리게 운전

접수일자 : 2008년 10월 1일

완료일자 : 2008년 12월 3일

\* 교신 저자

이 논문은 2007년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임(KRF-2007-313-D00493). 연구비지원에 감사드립니다.

하면 LED로 만들어진 '얼굴'이 붉게 변한다. Pod는 운전자의 감정을 감지하고 자신의 상태를 표현할 수 있도록 전자 개성을 가지고 있다. 국내에서도 삼성자동차가 졸음운전 방지 장치 및 방법에 대해서 특허를 발표했는데, 그 내용은 음성과 영상 데이터를 이용하여 운전자의 졸음 상태를 인식 판단하여 운전자가 졸음 상태일 경우 가청음을 출력하여 운전자의 졸음운전을 예방하는 기술이다. 현대자동차에서는 지문인식과 음성 신호에 대해서 운전자들의 감정 및 운전습관 등을 데이터베이스로 구축하여 운전자의 감정을 인식하여 난폭운전 및 졸음운전을 예방하는 기술에 대해서 특허를 발표했다. 이렇게 기존의 연구에서는 단일 센서에 의한 운전자의 감정을 인식하는 연구가 주류를 이루고 있다[2-4]. 예를 들어 감정인식에서 보편적으로 많이 사용되는 음성과 얼굴 영상만을 가지고 운전자의 감정을 인식한다면 주의력이나 졸음 같은 인지적인 면을 정확히 인식하기는 힘들다. 따라서 운전자의 감정 및 주의력을 동시에 인식하기 위해서는 다양한 생체-행동 신호를 고려하여 실험이 이루어져야 하는데 지금까지는 이에 대한 연구가 전무하다. 본 논문에서는 다중 센서 융합 기술을 이용하여 이러한 문제를 해결하고자 한다.

운전자가 운전 중에 자의적 또는 타의적으로 여러 가지의 감정 및 정서 변화가 나타날 수가 있는데, 이들은 주로 행복(happy), 기쁨(pleasure), 분노(anger), 슬픔(sad), 놀람(surprise), 공포(fear), 혐오(disgust) 등으로 대표할 수 있으며, 동시에 신체 변화에 따른 졸음, 주의력 산만 등이 발생할 수 있다. 운전자는 각자 안전 운전을 위하여 운전 중에 일어나는 여러 가지 감정 및 정서에 대해서 대처하는 방법이 운전자 별로 각각 다르다. 그럼에도 불구하고 현재까지의 연구는 이러한 문제에 대한 연구가 체계적으로 이루어지지 않고 있는 실정이다. 대부분의 운전자(특히 초보 운전자나 노약자 등)들은 운전 중에 일어나는 많은 상황에 대해서 대처하기가 힘들며, 이런 경우 대형 사고로 이어지는 경우가 종종 발생한다. 따라서 운전 중에 나타나는 감정 및 인지 상황을 일반화시켜서 보다 운전자의 감정을 정확히 표현하고 이에 대해서 적절히 대처할 수 있는 연구가 필요하다.

운전자의 감정 및 인지를 인식하여 그 상황에 따라 운전자에게 보다 편안하게 운전에 집중 할 수 있도록 다양한 서비스를 제공할 수가 있다. 만약 운전자가 졸음운전을 하고 있을 경우 좀 시끄러운 음악을 틀거나 자동차 내부의 조명등을 깜빡이게 해서 졸음운전을 못하도록 하거나 운전자가 난폭운전을 하는 경우에는 차량 내부 조명을 마음을 진정시키는 기능이 있는 블루 패턴 바꾸고 차분한 음악을 들려줌으로써 운전자가 난폭운전을 못하도록 방지한다. 이외에도 감정 및 인지를 인식함으로써 운전자에게 제공해 줄 수 있는 기능은 많이 있을 것이다.

본 논문은 안전운전을 위한 보조시스템으로 음성과 영상 그리고 Eye tracker에 Multi-Modal Sensor Fusion Algorithm을 사용하여 운전자의 감정 및 부주의 운전을 인식하고 적절한 서비스를 제공하여 사전에 교통사고를 예방하고자 하는데 그 목표를 두고 있다. 다음 그림 1은 본 논문에서 제안한 전체적인 시스템의 구성을 보여준다.

본 논문의 구성은 다음과 같다. 2장에서는 음성 신호 및 얼굴 영상을 이용한 운전자의 감정 인식을 기술하며 3장에서는 영상과 아이트래커를 이용한 운전자의 부주의 운전 및 졸음운전 인식에 대해 설명한다. 그리고 4장에서 운전자의 감정 및 부주의 운전 인식 실험을 기술하고 5장에서 결론을 맺는다.

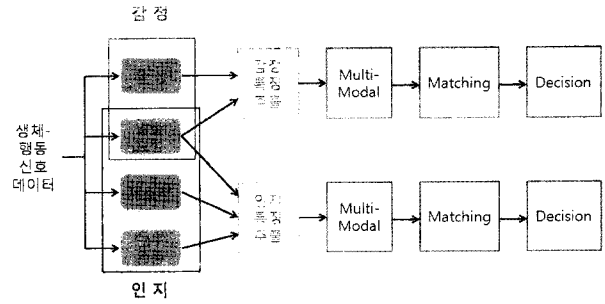


그림 1. 전체적인 시스템 구성도  
Fig. 1. Overall System Diagram

## 2. 음성 신호 및 얼굴 영상을 이용한 운전자의 감정 인식

### 2.1 음성신호를 이용한 감정 인식

감성 인식기는 두 부분으로 구성되어 있다. 첫 번째 부분은 음성으로 특징을 추출하는 부분이고 두 번째 부분은 그 특징들을 이용하는 패턴 인식 부분이다. 특징들은 피치의 통계치, 소리의 크기, 섹션개수 등이다. 피치 추출 방법으로는 가장 일반적인 방법들 중 하나인 autocorrelation approach를 사용했다. 피치 값은 0.1초마다 추출 되어졌고 그 값들의 평균이 Pitch mean으로 정의 했다. 그리고 분산 값 또한 동일한 데이터에서 얻어졌다. 소리의 크기는 magnitude estimation method에 의해서 구해졌고, 섹션 개수, Increasing Rate(IR), Crossing Rate(CR) 등은 우리의 이전 논문에서 사용한 방법으로 구했다[5].

본 논문에서는 마이크를 통해 입력된 음성신호에 대해서 6가지의 특징들을 Bayesian Learning(BL)에 적용하여 감정별 패턴을 분류하였다. Bayesian Learning은 사전확률을 이용하여 어떤 가설의 확률을 계산하는 방법이다. 그래서 본 논문에서는 400개의 음성 샘플들을 이용하여 각 감정과 특징들 간의 확률분포를 조사하여 사전확률을 계산하였다. 그리고 사용자의 확률 분포와의 유사정도를 파악하여 5가지 감정(평화, 기쁨, 슬픔, 놀람, 화남)으로 패턴을 분류하였다[6].

### 2.2 카메라 영상을 이용한 감정 인식

본 논문에서는 얼굴 영상을 이용하여 감정을 인식하기 위해서 피부색 추적 알고리즘과 Gray 형태 변환 방법을 이용하여 입, 눈과 눈썹들의 특징들을 추출하였다. 그리고 추출된 특징들은 다차원 특징 벡터로 구성되어 있어서 패턴을 분류하기에 용이하지 않다. 그래서 정보를 유지하면서 저차원으로 특징 벡터를 축소시키는 방법이 필요한데 본 논문에서는 이 방법으로 Principal Component Analysis(PCA)을 사용하였다.

PCA 알고리즘을 통해 고유 데이터 벡터를 구한 후 유클리디안 거리를 통해 학습 데이터와 입력 데이터간의 거리를 비교하여 그 거리가 최소가 되는 표정이 입력과 가장 유사한 표정이므로 그 학습데이터의 감정을 결과로 결정하게 된다[7]. 다음 그림 2는 본 실험에서 사용한 JAFFE 이미지이다.

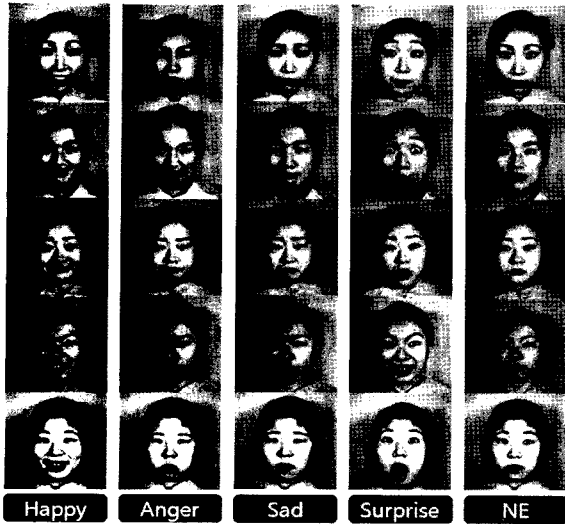


그림 2. 실험에 사용된 이미지  
Fig. 2. Image of used to Experiment

2.3 Multi-Modal Sensor Fusion을 이용한 감정인식

센서 융합 방법에는 일반적으로 결정 융합 방식과 특징 융합 방식이 있다. 결정 융합 방식은 독립적인 센서의 최종 출력 값을 사용하여 융합하는 것이며 특징 융합 방식은 각각의 매개체의 특징벡터들을 모두 고려하여 융합하는 방식을 말한다. 특징 융합 방법은 결정 융합 방법보다는 구현이 어렵지만 각각 매개체의 단점을 보완할 수 있다는 장점을 가지고 있다. 다음 그림 3은 이러한 특징 융합 방법의 일반적인 실행절차를 보여주고 있다.

본 논문에서는 이러한 특징 융합 방법을 실험하기 위해 음성 신호와 얼굴 영상에서 특징들을 추출하였는데 그 결과 각각 6가지와 5가지의 특징 벡터를 추출할 수 있었다. 하지만 11가지의 특징 벡터를 모두 고려하면 차원의 저주에 빠질 위험성이 크고 인식 속도가 느려지는 단점이 생길 수 있으므로 특징 선택 방법을 통해 우수한 특징들을 선택하게 된다. 이와 같은 특징 선택 방법으로는 여러 가지가 존재하지만, 본 논문에서는 Sequential Forward Selection(SFS) 방법을 이용하였다. SFS는 비어있는 집합에 순차적으로 특징들을 추가한 후 목적함수에 대입하여 그 결과가 가장 우수한 것들을 특징들로 선택하는 방법이다. 본 논문에서 사용된 목적함수는 식 (1)과 같다.

$$y = 3x_0 + x_1 + 4x_2 + 10x_3 - 5x_4 + 8x_5 + 7x_6 + 8x_7 - 10x_8 + 6.8x_9 + 7.3x_{10} - 5.2x_{11} \quad (1)$$

이 식에서  $y$ 는 목적함수 결과 값이고  $x_n$ 은 특징들의 종류를 나타낸다. 그리고 목적 함수 파라미터들을 다음과 같이 표현한 이유는 학습 데이터로부터 각각의 특징들을 추출하여 그 크기가 5번째 안에 있는 것들은 reward(+0.1)을 주고 그 이후에 있는 것들은 penalty(-0.1)를 주었다. 이와 같은 실험을 100번 반복한 결과 값이다. 이와 같이 특징들이 결정되면 이 값들을 인공 신경망 중 Back-Propagation(BP)로 학습하는 Multi Layer Perceptron(MLP)에 입력으로 설정하여 감정별 패턴을 분류하였다.

본 논문에서 사용한 BP 알고리즘은 출력층의 오차 신호를 역전파하여 은닉층과 출력층간의 연결 강도와 입력층과 은닉층간으로 연결 강도를 변경하는 학습 방법으로 다양한

분야에 그 응용 범위가 넓다.

다음 표 1은 본 논문에서 사용된 초기 파라미터 설정 값을 나타내고 있다. 표 1과 같은 초기 파라미터와 감정별 목표치 001~101, 초기 가중치는 -0.03 ~ 0.03로 랜덤하게 설정한 후 학습데이터들을 이용해 오차 범위보다 작아질 때까지 학습을 시킨다. 그리고 입력 데이터를 입력한 후 5가지 감정으로 패턴을 분류한다.

표 1. 신경망의 초기 파라미터 설정.

Table 1. Parameter setting of neural network

Parameter	Value
Hidden Units	13
Output Units	3
Learning Rate	0.005
Tolerance	0.1
Sigmoid Function	$1/1+e^{-3x}$

다음 그림 3은 음성 신호와 얼굴 영상을 이용한 특징 벡터를 구하고 이를 융합하는 특징 융합 방식을 표현한 구성도이다.

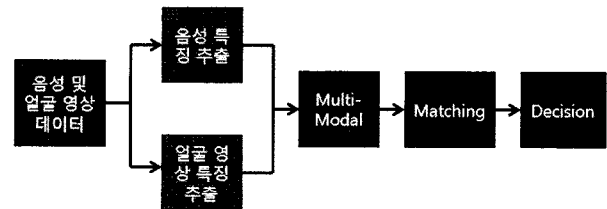


그림 3. 특징 융합 방법 구성도  
Fig. 3. The Feature Fusion Method

3. 영상과 아이트래커를 이용한 운전자의 부주의 운전 및 졸음운전 인식

3.1 영상을 이용한 졸음 및 부주의 운전 인식

사람의 피부색(flesh color)은 얼굴 검출에 있어서 훌륭한 특징점이 될 수 있다. 비록 사람마다 피부색이 다르지만 이러한 차이의 주된 이유는 색상보다는 밝기(intensity)의 차이에 있다는 것이 여러 논문을 통해 입증되었다[8].

특정한 조명에서 피부색은 얼굴의 회전이나 크기의 변화에 강인한 특성을 가진다. 이러한 이유로 얼굴 검출 시스템에서 피부색을 특징 점으로 많이 이용한다. 본 논문에서는 이러한 피부색을 모델링하는 방법으로  $Y C_r C_b$  컬러 모델을 사용하였다. 휘도는  $Y$ 로 기호화 되고 푸른 정보와 붉은 정보는  $C_r, C_b$ 로 기호화 된다. RGB에서  $Y C_r C_b$ 로의 변환은 다음과 같은 공식에 의해 계산된다.

$$\begin{aligned} Y &= 0.2990R + 0.58700G + 0.11400B \\ C_b &= -0.16874R - 0.33126G + 0.50000B \\ C_r &= 0.50000R - 0.41869G - 0.08131B \end{aligned} \quad (2)$$

다음의 그림 4에서 왼쪽은 카메라로부터 입력된 이미지이며, 오른쪽은 식 (2)에 의해 피부색 필터링을 거친 이미지

를 나타낸다.



그림 4. 피부색 필터링 이미지  
Fig. 4. Image of Flesh Filtering

영상을 이용하여 줄임과 부주의 운전을 인식하기 위해서는 전처리 과정으로 강인한 얼굴영역 검출이 필요하다. 본 논문에서는 Mean Shift Algorithm과 AdaBoost Algorithm을 사용하였다. Mean Shift Algorithm은 색상에 대한 확률 분포의 밀도를 구하여 AdaBoost의 검색창의 영역을 피부색의 중심영역으로 제한시켜 보다 빠른 검색을 할 수 있는 효율적인 알고리즘이다. 강인한 얼굴영역을 검출하고 추적하기 위하여 우리는 AdaBoost 알고리즘을 사용하였다[9]. 그림 5는 AdaBoost 알고리즘을 이용한 얼굴 검출의 결과를 나타낸다.

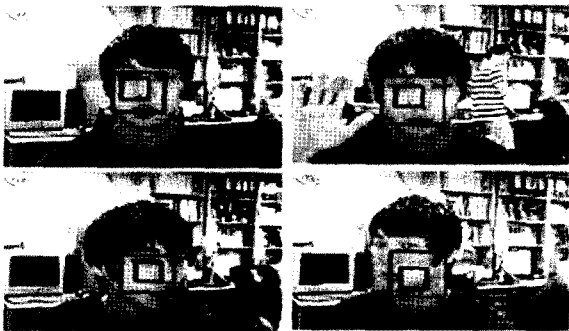


그림 5. AdaBoost 알고리즘을 이용한 얼굴 검출  
Fig. 5. Face Detection with AdaBoost Algorithm

AdaBoost Learning Algorithm은 다수의 특징 집합으로부터 작은 수의 핵심적인 특징을 선택하는 알고리즘으로서 연속적인 영상에서 적용하기에 적합한 장점을 가지고 있다 [10-11]. 얼굴의 이미지와 얼굴이 아닌 이미지를 가장 잘 판단할 수 있는 사각형 특징 점을 찾도록 설계된다. 각각의 특징 점에 대해서 판단의 에러를 최소화 할 수 있는 최적의 임계값을 결정하여야 한다. 약 분류기(weak classifier)  $h_j(x)$ 는 다음과 같이 나타낼 수 있다.

$$h_j(x) = \begin{cases} 1 & \text{if } p_j f_j(x) < p_j \theta_j \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

식 (3)에서  $f_j$ 는 특징점,  $\theta_j$ 는 임계 값 그리고 부등호의 방향을 표시하는 parity  $p_j$ 로 구성된다.

본 논문에서 임계값을 구하는 방법은 Haar like feature를 계산하고, 그 결과가 가우시안 분포라는 가정 하에 평균과 분산을 구한 후, 얼굴과 비 얼굴의 가우시안 분포가 만나는 점을 임계값으로 설정하였다.

사전 초기화 작업을 통해 분류기가 학습되면 입력된 영상에 대하여 분류기를 적용한다. 얼굴이 검출될 가능성이 많은 영역에서는 "1"을 얼굴이 아니라고 판단되는 영역에서는 "0"을 결과 값으로 나타낸다. 계층적인 구조를 가진 최종 학습되어 있는 분류기는 내부에 여러 단계의 분류기를 가진다.

분류기는 24x24크기의 Search Window를 이동하면서 내부에 각 단계에 포함되어 있는 분류기를 사용하여 모든 위치를 검토한다. 계층적 분류기를 사용함으로써 계산 시간의 효율성을 높일 수 있다.

앞에서 설명한 AdaBoost Algorithm을 이용하여 얼굴영역을 검출하고 그 검출된 얼굴 영역에서 눈과 눈썹의 특징점을 찾아내기 위해 이미 알고 있는 지정학적 위치와 색상 정보를 이용해 눈과 눈썹영역을 추출해낸다.

눈의 경우 얼굴 영역 내에서 무채색 색 정보를 많이 갖고 있는 점에 착안해 무채색영역을 추출해내 분리하여 하였으나, 흘러내린 머리, 코, 얼굴내의 그늘영역 등에 의해 왜곡이 심하였다. 이를 보완하기 위해 눈썹의 위치를 동시에 고려하였다.

다음의 그림 6과 같이 입력된 이미지에서 피부색 필터링을 거친 후 레이블링(Labeling)을 수행하여 얼굴 영역을 결정한다. 그리고 얼굴영역을 반전시켜(Inverting) 눈과 눈썹의 영역을 찾기 위하여 레이블링을 다시 수행하여 얻어진 이미지와 시작단계에서의 피부색 필터링을 수행한 이미지와 XOR 연산을 수행하면 원하는 눈과 눈썹의 영역을 얻을 수 있다. 이렇게 얻어진 눈과 눈썹의 영역에서 레이블링을 통해 각각의 좌표를 얻어 헤드 제스처 인식을 위한 데이터로 사용한다.

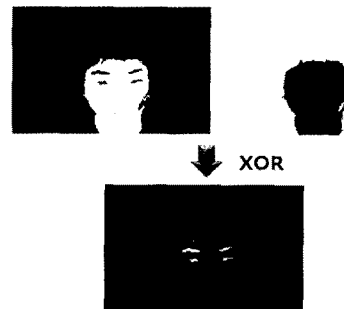


그림 6. Labeling과 XOR 연산을 이용한 눈과 눈썹 영역 검출

Fig. 6. Detection of Eye & brows with Labeling and XOR operation

그림 7은 입력 영상에서 눈과 눈썹의 좌표를 구하여 영상에 표현한 이미지를 보여준다. 사각형 박스는 전체 얼굴 영역의 좌표를 그리고 십자마크는 각각의 눈과 눈썹의 중심 좌표를 표시한다.



그림 7. 눈과 눈썹의 영역에 대한 좌표 데이터  
Fig. 7. Coordinates of eye and brows area

3.2 아이트랙커를 이용한 졸음 및 부주의 운전 인식

본 논문에서 제안한 시스템을 위해서 우리는 View-Point Eye Tracker를 사용하였다. Eye Tracker는 안경처럼 착용할 수 있는 고글형 타입으로 한쪽 눈의 아래 대각선 방향에 소형 적외선 카메라와 적외선 센서로 구성되며 추가로 전방에 시야를 획득하기 위한 소형 카메라가 부착되어 있다. 따라서 적외선 센서를 사용하여 동공(Pupil)을 인식하게 된다. 다음의 그림 8은 실험에 사용된 Eye Tracker의 착용 모습이다.

Eye Tracker를 사용하기 위해서는 사용자에게 따른 캘리브레이션(calibration)이 필요하다. 본 논문은 Eye Tracker의 Calibration의 방법으로 사용자의 인터페이스를 강조한 Glint Pupil Vector method를 사용하였다. GPV(Glint Pupil Vector)방법은 윈도우 전체화면의 영역에서 랜덤하게 사각형을 띄워주고 사용자는 머리가 고정된 상태에서 눈동자만 랜덤하게 나타나는 사각형을 응시하도록 하여 사각형과 눈동자의 이동 좌표를 대조하여 운동자의 움직임을 캘리브레이션하는 방식이다.



그림 8. View Point Eye Tracker의 착용 이미지  
Fig. 8. Image of wear View Point Eye Tracker

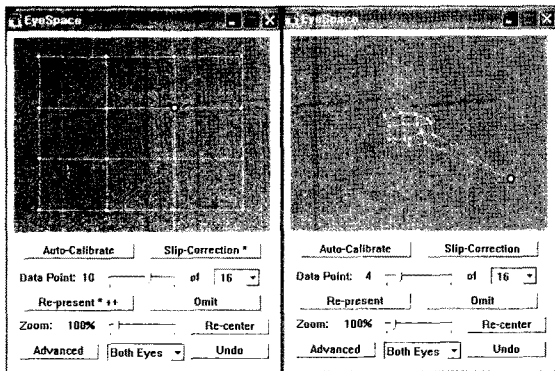


그림 9. Glint Pupil Vector를 이용한 캘리브레이션  
Fig. 9. Calibration using Glint Pupil Vector

위 그림 9에서 왼쪽의 Eye Space Plot은 초기의 균일분포의 영역을 나타내며 오른쪽의 Eye Space Plot은 GPV를 이용하여 캘리브레이션된 영역을 나타낸다. 오른쪽의 캘리브레이션된 Plot에서 오른쪽에 하나 튀는 부분은 눈이 깜박일 때 발생하는 캘리브레이션 포인트이다.

Eye Tracker를 이용하여 안구의 움직임을 추적하고 좌표를 실시간으로 얻어서 일정 시간 동안 안구가 움직임이 없거나 어느 한곳으로 시선이 편중되어있다면 부주의 운전으로 판단하고 눈의 깜박임의 주기가 불규칙하거나 감겨있을 경우 졸음운전이라고 판단할 수 있다.

Threshold Time을 적용하여 부주의 운전에 대한 판별을 같이 표현할 수 있다.

$$f(t) = \begin{cases} 1 & T_{th1} < |t| < T_{th2} \\ 0 & |t| < T_{th} \end{cases} \quad (4)$$

식 (4)에서 판별기(Discriminant)  $f(t)$ 는 임계 시간  $T_{th1}$ ,  $T_{th2}$  그리고 좌표의 움직임 시간(Motion Time)  $|t|$ 로 구성된다.

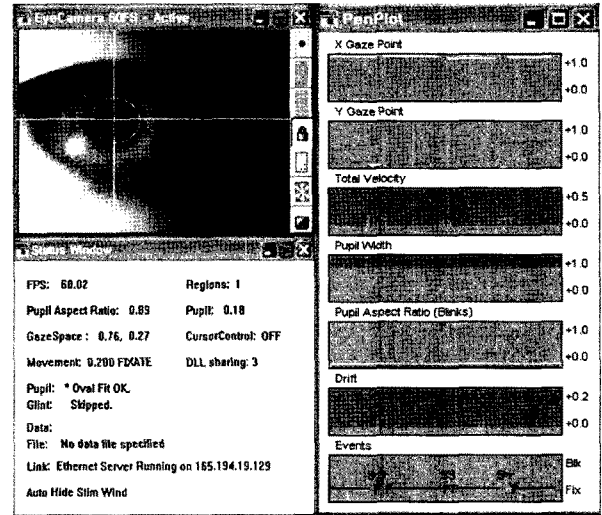


그림 10. 아이트랙커의 출력 데이터  
Figure 10. Output Data of Eye-Tracker

그림 10은 Eye Tracker를 사용하여 졸음 및 부주의 운전에 사용할 데이터들을 보여준다. 특히 출력 데이터 Events의 B는 눈의 깜박임(Blink)을 나타내며 F는 눈의 고정(Fixation), S는 안구의 순간적인 움직임(Saccade) 그리고 D는 느린 움직임(Drift)을 나타낸다.

3.3 Mutli-Modal Sensor Fusion을 이용한 졸음 및 부주의 운전 인식

카메라와 아이트랙커를 이용하여 졸음운전 및 부주의 운전을 인식하기 위해 센서 퓨전을 하는 방법은 앞 절에서 제시한 특징 융합 방식(Feature Fusion method)을 사용하였다.

신경망(neural network) 이론의 응용분야는 다양하여 영상처리와 압축, 패턴인식, 로봇의 제어, 예측문제, 최적화문제 등 다양하게 적용되어지고 있다. 본 논문에서는 신경망 학습 알고리즘으로 역전파(BP) 알고리즘을 사용하였다. Multi-Layer Perceptron(MLP)에 입력 파라메타는 앞장에서 설명한 카메라를 이용한 체스처 데이터 5개와 Eye Tracker 5개의 출력 데이터를 사용한다. 얼굴영역은 눈과 눈썹에서부터 얼굴의 외곽 영역까지의 각각의 거리(Distance)를 이용하며 Eye Tracker는 눈동자의 움직임 좌표와 눈동자 외곽의 영역까지의 거리를 사용한다. MLP의 출력은 얼굴과 눈동자에 대해 상하좌우에 대한 이산 적으로 출력 값이 나오게 된다. 그림 11은 신경 회로망의 구조를 도시화하였다.

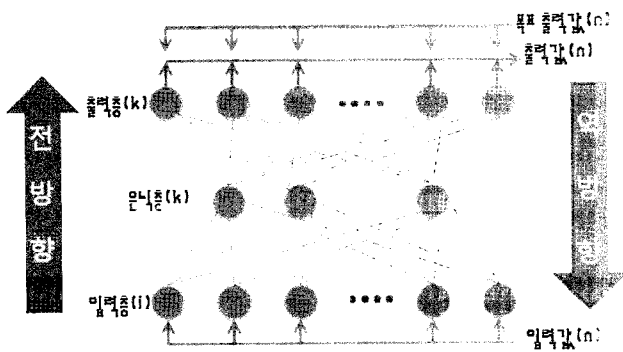


그림 11. 신경 회로망을 이용한 제스처 분류  
Fig. 11. Gesture classification with MLP

다음 표 2는 센서 융합방식(Neural Network)을 위한 초기 파라미터 설정 값들을 나타내고 있다. 여기에서 은닉층은 40개 층으로 구성하였으며 출력은 5개로 하여 운전자의 시선에 따른 상하좌우를 출력 값으로 얻을 수 있다. 이렇게 얻어진 출력은 식 (4)에 의해 임계 시간 외에 따른 줄음 및 부주의 운전을 인식할 수 있다.

표 2. 신경망의 초기 파라미터 설정.

Table 2. Parameter setting of neural network

Parameter	Value
Hidden Units	40
Output Units	5
Learning Rate	0.005
Tolerance	0.1
Sigmoid Function	$1/(1+e^{-3x})$

그림 12는 아이트래커와 카메라를 이용한 헤드 제스처의 센서 융합 방식을 나타낸다.

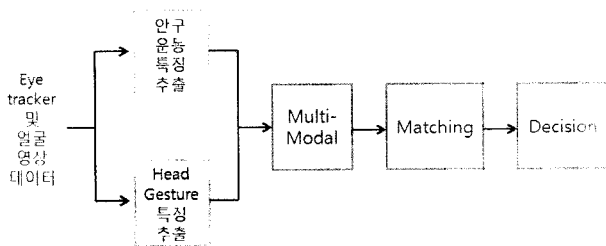


그림 12. 센서 융합 방법 구성도  
Fig. 12. Sensor Fusion Method

#### 4. 운전자의 감정 및 부주의 운전 인식 실험

##### 4.1 실험을 위한 DataBase 구축

본 논문에서는 음성 신호와 얼굴 영상을 이용하여 운전자의 감정 및 부주의운전을 인식하기 위해 동일한 환경에서 제스처와 감정별 Database를 구축한 후 실험을 하였다. 음성 신호의 경우 10명의 남성 대학생들(나이:25~31)에게 5가지 감정별로 40개의 음성 샘플을 얻었으며, 얼굴

영상의 경우는 10명의 남성 대학생들(나이:25~31)에게 5가지 감정별로 6개의 얼굴 영상을 촬영하여 얻은 샘플과 JAFFE 이미지를 이용하였다.

소리 및 영상의 크기가 피험자와 마이크 혹은 카메라의 거리에 따라 달라질 수 있으므로 그 거리를 30cm로 고정하였으며 녹음된 형태는 11kHz, 16bit, mono이고 촬영된 형태는 320×240 pixels, 8bit, bmp 이미지이다.

녹음된 문장들은 30개의 일상적이고 단순한 것들이었고 문장의 길이는 6~10음절로 제한하였다.

##### 4.2 음성과 영상을 이용한 감정 인식 실험

본 논문에서 음성 신호의 경우 단어의 의미상에 감정적인 특징을 잘 나타내는 15가지 단어를 정하여 실험이 이루어졌으며 그 단어들은 표 3과 같다.

표 3. 실험에서 사용된 감정별 단어들

Table 3. The emotional words is used in an experiment

Words	Emotion
그래, 알았어, 안녕	보통
하하하, 아싸, 안녕	기쁨
견디기 힘들어, 죽겠어, 안녕	슬픔
그러지 말랬지, 너 뭐야, 죽을래	화남
비명소리 (악~, 아~ 등)	놀람

마이크를 통해 입력된 음성 신호에 대해서 앞서 설명한 방법에 의해 의미적인 특징을 배제한 운율적인 특징만을 고려한 특징(피치의 최대치 및 평균치, 소리의 크기, 섹션 개수, IR, CR)들을 추출한다.

본 논문에서는 음성 신호에서 감정별 패턴을 분류하는 알고리즘으로 Bayesian Learning을 이용하였다. Bayesian Learning은 확률을 기반으로 하는 알고리즘으로써 먼저 구축된 Database를 통해 감정별 사전확률을 구하게 된다. 그리고 추출된 특징들을 입력 값으로 설정하고 알고리즘을 통해 사후확률을 구한 후 두 확률의 유사 정도를 통해 5가지 감정별(평화, 기쁨, 슬픔, 놀람, 화남) 패턴을 분류하였다.

그림 13은 음성신호를 이용하여 감정인식을 하기 위해 구성한 기본 UI 다이어그램이다.

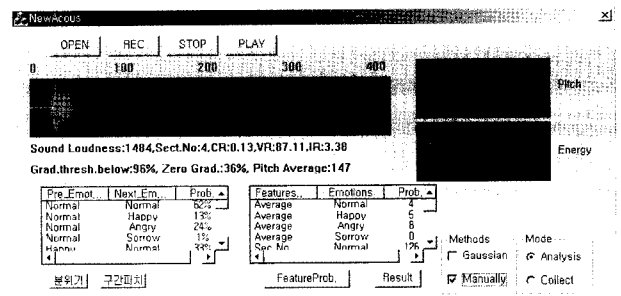


그림 13. 음성을 이용한 감정인식 UI  
Fig. 13. The emotion recognition UI with Voice signal

얼굴영상의 경우에는 10명의 피험자로부터 카메라를 통해 획득하며 실험 조건은 Database 구축 때와 동일하게 하였다. 피험자마다 동일한 감정에 대해 10번의 촬영을 하였으며 이를 바탕으로 50번의 감정 인식 실험이 이루어졌다. 이렇게 촬영된 얼굴영상에 대하여 피부색 추적 알고리즘을

통해 눈, 눈썹, 입에 대한 특징들을 추출하게 되는데, 피부색 축적 알고리즘은 피부색 영역만을 검출 한 후 2차원 평면에 피부색 픽셀만을 축적하는 방법을 말한다. 이렇게 추출된 특징들을 영상처리에 용이한 Gray 형태로 변환한 후 기본적인 히스토그램 평활화, Sobel 연산 등을 통해 영상처리를 하게 된다. 이렇게 추출된 특징들은 다차원 특징 벡터로 구성되어 있는데 이것을 고유의 정보를 유지하면서 저차원 벡터로 축소시키기 위해서 본 논문에서는 PCA 알고리즘을 사용하였다. 그리고 유클리디안 거리를 통해 5가지 감정별 패턴을 분류하였다. 그림 14는 PCA 알고리즘을 적용한 이미지이다.

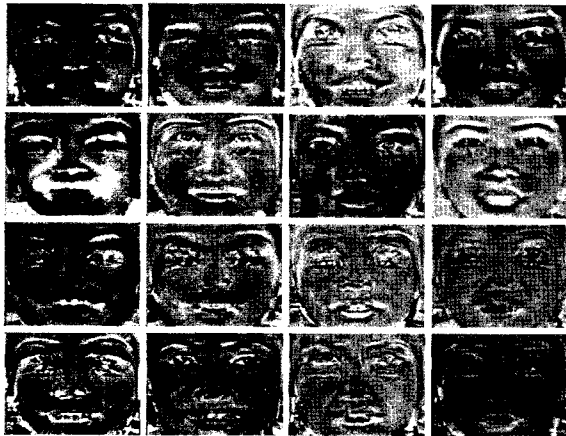


그림 14. PCA 알고리즘을 적용한 이미지  
Fig. 14. The facial image of used to PCA

표 4는 10명의 피험자를 실험한 결과이다. 실험 결과에서 볼 수 있듯이 특징 융합 방식을 사용한 센서 융합은 약 70%의 인식률을 보였다. 각각의 음성을 통한 인식률 약 66%와 얼굴 영상을 이용한 58%의 인식률보다 센서 융합을 통해 감정을 인식하는 것이 더 향상된 성능을 보였다. 하지만 학습 시 걸리는 시간이 오래 걸린다는 단점이 있으며 본 논문에서는 60만 번의 반복을 통해 오차를 허용범위에 도달할 수 있었다.

표 4. Muti-Modal 감정 인식률 (%)  
Table 4. The emotion recognition rate of facial image

	평화	화남	슬픔	기쁨	놀람	평균
S1	63	65	72	68	74	68.4
S2	75	72	68	72	68	71
S3	74	72	68	76	70	72
S4	72	65	72	68	70	69.4
S5	68	75	66	69	67	69
S6	78	74	70	64	72	71.6
S7	68	72	70	69	66	69
S8	72	70	65	72	66	69
S9	74	74	65	70	64	69.4
S10	73	85	58	65	78	71.8
평균	71.7	72.4	67.4	69.3	69.5	70.06

#### 4.3 영상과 아이트랙커를 이용한 부주의 인식 실험

본 논문에서 운전자의 졸음 및 부주의 운전 인식을 위한 센서로 BB2 스테레오 카메라와 View Point Eye Tracker를 사용하였다.

실험은 운전자가 취할 수 있는 제스처로 정면과 상하 좌우로 시선을 두는 행동을 재연하여 성능을 테스트하였다. 그림 15는 운전자의 헤드 제스처 인식 결과를 보여준다. 검출된 얼굴 전체 영역을 추적하고 눈과 눈썹의 위치를 검출한다. 그리고 얼굴 영역과 눈의 위치에 대한 거리(Distance)를 이용하여 상하좌우에 대한 헤드 제스처를 인식하였다.



그림 15. 운전자의 제스처 인식 실험 결과  
Fig 15. Experiment result of Gesture recognition

표 5는 카메라와 아이트랙커를 사용하여 운전자의 제스처 인식 실험을 한 결과이다. 실험은 5명의 피험자를 대상으로 상하좌우에 대해 각각 50번씩 반복하여 실험하였다. 실험 결과 눈과 눈썹의 검출율과 헤드 제스처의 인식률이 약 80%정도로 만족할 정도의 성능을 보였다. 하지만 눈동자의 위치를 추적하여 제스처를 인식하기 위해 Eye Tracker를 착용하고 실험을 하였을 경우 눈과 눈썹의 위치를 정확히 검출하지 못하여 헤드 제스처의 인식률이 약 67%로 다소 떨어지는 성능을 보였다.

표 5. Muti-Modal 제스처 인식률 (%)  
Table 5. The Gesture recognition rate

		상	하	좌	우	평균
Eye tacker 착용	S1	90	23	76	80	67.25
	S2	88	31	75	78	68.00
	S3	79	43	76	79	69.25
	S4	84	28	82	81	68.75
	S5	81	22	79	76	64.50
	평균	84.40	29.40	77.60	78.80	67.55
Eye tacker 미착용	S1	96	37	71	81	71.25
	S2	94	62	87	85	82.00
	S3	79	71	86	83	79.75
	S4	98	64	85	89	84.00
	S5	89	68	89	87	83.25
	평균	91.20	60.40	83.60	85.00	80.05

다음의 그림 16은 운전자의 제스처와 아이트랙커를 이용하여 졸음 및 부주의 운전을 인식하기 위한 인터페이스이다. 센서의 입력으로 카메라와 아이트랙커 그리고 마이크 입력을 선택할 수 있도록 하였으며 신경망 학습을 시키기 위한 독립 다이얼로그를 생성하여 학습 시켰다.

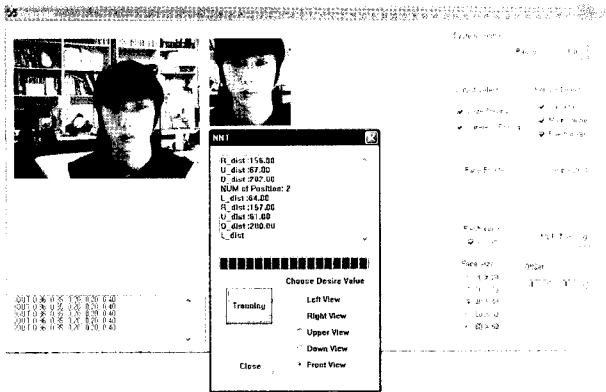


그림 16. 아이트랙커와 얼굴 영상을 이용한 졸음 및 부주의 운전 인식 실험 UI

Fig 16. Experiment UI for Doze and Careless Driving

### 5. 결 론

본 과제는 운전자가 안전 운전을 하기 위해 생체 행동 신호(얼굴 표정, 신체 움직임, 음성, 안구운동)를 이용하여 감정(기쁨, 화, 슬픔, 놀람, 공포) 및 졸음, 주의력의 신호를 추출하여 일정한 형태의 데이터베이스로 구축하고, 구축된 데이터를 이용하여 운전자의 감정 및 졸음, 주의력의 특징 점들을 검출하여, 그 결과 값을 Multi-Modal 방법을 통해 융합함으로써 운전자의 감정 및 주의력 상태를 인식할 수 있는 시스템을 개발하는 것이다.

실험 결과에서 볼 수 있듯이 본 논문에서 제안한 방법이 안전 운전을 위한 시스템에 적합하기에 대한 가능성을 충분히 보여주었다. 하지만 실질적으로 차량에 장착하여 실제 운전상황에 대한 운전자의 감정과 제스처에 인식률을 확인할 필요가 있다. 그리고 Eye Tracker의 경우 오히려 얼굴에 대한 영상을 외곡시켜 그 성능을 저하 시키는 결과를 가져왔다. 따라서 Eye Tracker를 직접 착용하는 방식보다는 원거리에서 카메라를 이용하여 눈동자를 추적하는 시스템을 도입한다면 좀 더 나은 성능을 발휘할 것이라 생각된다.

향후 연구는 운전자의 감정 및 졸음 그리고 부주의 운전에 대한 인식률을 좀 더 향상시킬 수 있는 알고리즘을 연구 개발하여야 할 것이며 본 논문에서 제안한 시스템에 출력에 안전운전을 위해 차량에서 제공할 수 있는 서비스 등에 대한 연구도 같이 진행 되어야 할 것이다.

### 참 고 문 헌

[1] 강동주, "교통사고 유형별 사고발생 특성 및 요인 분석," *한양대석사논문*, 1994.  
 [2] Lee C.M., Narayanan S.S. and Pieraccini. R., "Classifying emotions in human - machine spoken dialogs", *ICME'02*, vol. 1, pp. 737-740, 2002.  
 [3] New T.L., Wei F.S. and De Silva L.C., "Speech based emotion classification", *TENCON 2001*, vol. 1, pp. 297-301, 2001.  
 [4] J. Nicholson and K. Takahashi, R. Nakatsu, "Emotion recognition in speech using neural networks" *Proc. of ICONIP*, Vol. 2, 1996.

[5] D. Silval and P. C. Nag, "Bimodal emotion recognition", *Proc. of Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000*, pp. 332-335, 2000.  
 [6] C. H. Park and K. B. Sim, "Pattern Recognition Methods for Emotion Recognition with speech signal", *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 6, no. 2, pp. 150-154, 2006.  
 [7] Ho-Duck Kim, Hyun-Chang Yang, Chang-Hyun Park, and Kwee-Bo Sim, "Emotion Recognition Method of Facial Image using PCA ", *Journal of Korea Fuzzy Logic and Intelligent Systems Society(KFIS)*, vol.16, no.6, pp. 772-776, Dec. 2006.  
 [8] J. Yang and A. Waibel, "A real-time face tracker," *Proc. of Third Workshop on Applications of Computer Vision*, pp. 142-147, 1996  
 [9] P.Viola, M.J.Jones, "Robust Real-Time Face Detection," *International Journal of Computer Vision*, Vol.57, No.2, pp. 137-154, 2004  
 [10] R. Schapire and Y. Singer. "Improving boosting algorithms using confidence-rated predictions," 1999.  
 [11] Y. Freund and R. Schapire, "Experiments with a New Boosting Algorithm", *Proc. of ICML -96*, pp. 148-156, 1996.

### 저 자 소개



한철훈(Cheol-Hun Han)  
 2008년 : 중앙대학교 전자전기공학부  
 공학사  
 2008년~현재 : 중앙대학교 대학원  
 전자전기공학부 석사과정

관심분야 : Computer vision system, Emotion Recognition, Intelligent Robot 등



심귀보(Kwee-Bo Sim)  
 1990년 : The University of Tokyo  
 전자공학과 공학박사  
 1991년 ~ 현재 : 중앙대학교  
 전자전기공학부 교수  
 [제18권 5호 (2008년 10월호) 참조]

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