

# 멀티 모달 학습을 사용한 시각적 특징 기반의 운전자 졸음 감지

최형탁\*, 백문기\*, 강재식\*, 이규철<sup>o</sup>

## Driver Drowsiness Detection Based on Visual-Feature Using Multi-Modal Learning

Hyung-tak Choi\*, Moon-ki Back\*, Jae-sik Kang\*, Kyu-chul Lee<sup>o</sup>

### 요약

운전자라는 제한적인 상황에서 졸음을 판단하기 위한 시각적 기반 졸음 감지 방법이 있다. 운전자의 눈, 머리 움직임 등의 특징들을 추출하여 졸음을 판단하는데 졸음 발생 시 나타나는 변화가 상이하며 다른 개체이기 때문에 판단 기준으로 두 가지 이상의 특징을 사용함에 있어 편향 현상이 발생할 수도 있으며 데이터 누락 시 졸음 판단이 어려울 수도 있다. 본 논문에서는 시각적 특징 데이터에 구애받지 않기 위하여 DBM(Deep Boltzmann Machine)을 사용하는 멀티 모달 학습 방법을 제안한다. 일반적으로 사용하는 시각적 기반 특징들 중 두 가지(눈, 입)를 사용하여 졸음 발생 시 변화를 관찰한다. 먼저 멀티 모달 DBM을 사용하여 두 특징 데이터를 학습한 후 RNN(Recurrent Neural Network)을 사용하여 시간적 변화를 고려한다.

**Key Words** : Multimodal-learning, Drowsy detection, Deep Boltzmann Machine, Restricted Boltzmann Machine, visual feature

### ABSTRACT

There is a vision-based drowsiness detection method for judging drowsiness in a limited situation of driver. The driver's eye and head movements are used to determine the drowsiness. In the case of drowsiness, the change in drowsiness is different. Since the drowsiness is different from other drowsiness, This may be difficult. In this paper, we propose a multimodal learning method using DBM(Deep Boltzmann Machine) in order not to be affected by visual feature data. Two of the commonly used visual-based features (eye, mouth) are used to observe changes in drowsiness. First, we study two feature data using multimodal DBM and then consider the temporal change using RNN (Recurrent Neural Network).

※ 본 연구는 국토교통부 국토교통기술촉진연구사업의 연구비지원 (17CTAP-C132686-01)에 의해 수행되었습니다.

• First Author : (ORCID:0000-0002-1698-7917)Chungnam National University Department of Computer Science & Engineering, poo4967@gmail.com, 학생회원

o Corresponding Author : (ORCID:0000-0003-0857-807X)Chungnam National University Department of Computer Science & Engineering, klee@cnu.ac.kr, 정회원

\* Chungnam National University Department of Computer Science & Engineering, zmmzment@gmail.com, steelroadk@gmail.com

논문번호 : 201806-C-093-SE, Received December 15, 2017; Revised June 18, 2018; Accepted June 28, 2018

## I. 서론

운전자가 졸음에 빠진다는 것은 매우 심각한 문제이며 이로 인한 교통사고는 매년 발생하고 있다. 졸음이 발생한다는 것은 운전자 입장에서는 한 순간이지만 차량은 운전자가 통제하지 않은 상태가 지속되어 매우 위험하고 큰 사고로 직결되어 추가적인 인명 피해로 이어질 수 있다. 실제로 운전자가 약 5초 정도를 졸음에 빠질 경우 시속 100km 인 차량은 약 140m를 운전자가 없는 상태로 이동한 것과 같다. 이러한 졸음 운전의 위험성이 논의되면서 졸음 감지 시스템 개발을 위한 연구들이 많이 이루어졌으며 다양한 졸음 판단 방법과 기준들이 존재한다. 이 때 시각적 기준은 사진을 통해 얻은 특징들을 사용하여 졸음을 판단하기 때문에 제한적인 상황에서 가장 사용하기 적절하다. 사용되는 특징들은 주로 눈, 입, 목의 움직임 등이 있으며 졸음 발생 시 나타나는 변화가 나타나는지 지속적인 관찰을 하여 사용자의 상태를 분류한다.

단일 특징을 관찰하는 졸음 판단 시스템은 운전자의 이미지를 획득하여 특징을 추출한 후 PCA, LDA 방법을 사용하거나 이미지 처리 방법 등 다양한 방법으로 추출한 특징의 유효성을 검사하여 정확도를 높인다. 즉, 추출된 특징의 데이터만 유효하고 졸음에서 나타나는 변화만 관찰되면 높은 정확도를 기대할 수 있다. 하지만 단일 데이터가 졸음에서 나타나는 변화는 한정적이므로 운전자의 상태를 졸음으로 분류하기에는 근거가 부족하다.

이를 보완하기 위해 추가적인 특징을 사용하여 두 가지 이상의 특징을 통해 운전자의 상태를 분류하는 방법이 있다. 눈 뿐만 아니라 추가적인 머리 움직임을 관찰하여 운전자의 상태를 졸음으로 분류하기에 부족한 근거를 보완한다.

하지만 각각의 특징들은 졸음 발생 시 나타나는 변화가 달라 추가적인 특징을 사용한다면 기존 시스템에 도입할 추가적인 알고리즘을 필요로 하며 이는 확장성의 저해 요인으로 작용한다. 또한 추가적인 특징들을 사용하는 시스템을 설계하여 졸음 판단의 근거를 향상시키지만 그와 동시에 추출해야 할 특징과 사용 알고리즘 증가로 인한 트레이드오프(Trade-off) 현상이 발생한다. 그리고 추가적인 알고리즘을 사용하여 졸음 판단 시스템을 설계하였을 때 두 가지 이상의 특징이 함께 적용되어 졸음 판단의 근거로 사용되지 않고 편향 현상의 문제가 발생될 수 있다. 예를 들어 운전자의 졸음 발생 시 눈과 목의 움직임으로 졸음 판단이 이루어지지 않고 눈을 감는 것만으로 졸음 판단이

계속 이루어지는 것이다. 만약 졸음 감지 시스템이 두 특징을 모두 입력 받아야 졸음을 판단하는 경우라면 편향 현상이 발생하지 않고 값의 누락이 발생하여 졸음 판단이 이루어지지 않는 더 심각한 문제가 발생한다.

위와 같은 문제를 해결하기 위해 본 연구에서는 운전자의 이미지로부터 추출한 특징들을 멀티 모달 학습을 사용하여 추가적인 알고리즘을 사용하지 않고 졸음을 판단하는 시스템 설계를 제안한다. 실제 세계의 정보는 여러 입력 채널을 통해 제공되며 인간은 여러 감각을 통하여 이를 인지한다<sup>[1]</sup>. 기계 학습 분야에서는 서로 다른 종류의 정보를 전달하여 유용한 정보를 발견하는 모델을 만들기 위한 학습이 필요했고 각각의 정보마다 모델을 만들기에는 시간과 비용이 들며 모델을 합치는 것 또한 추가적인 방법이 필요하다. 이러한 이유로 도입된 학습이 멀티 모달 학습이며 멀티 모달 설정에서 데이터는 여러 입력 모달로 구성되며 각 모달은 다른 종류의 표현과 구조를 가지고 있다. 즉, 각각의 특징들은 이미지로써 같은 구조의 데이터지만 졸음이라는 특정 조건에서 서로 다른 변화를 보이기 때문에 각각의 특징은 졸음을 표현하는 서로 다른 모달리티이다.

본문에서는 졸음 감지를 위해 시각적 특징을 추출하는 관련 연구에 대해 설명한다. 이어서, 이 논문이 제안하는 멀티 모달 네트워크 기반의 시각적 특징을 이용한 졸음 감지 시스템을 순차적으로 설명한다. 마지막으로 관련 연구에서 졸음을 감지한 정확도와 단일 특징을 이용한 졸음 감지 정확도, 본 논문에서 제시하는 멀티 모달 네트워크를 이용한 졸음 감지 정확도를 비교를 한 후 논문을 마친다.

## II. 관련 연구

운전자의 졸음을 판단하기 위한 기준을 생리학 데이터, 운전 행동 및 패턴 데이터, 운전자의 시각적 특징 데이터로 총 세 가지로 분류할 수 있다<sup>[2]</sup>. 가장 정확한 기준은 생리학적 특징인 심박수 및 뇌파를 측정하여 활용하는 것이며 이러한 생리학적 기준은 사용자의 졸음 상태를 세분화하여 분류할 수 있다. 하지만 운전자의 신체에 부착해야 되고 장비에 의존하기 때문에 운전자 상태에 사용하기에는 부적합하다. 이러한 이유로 운전자의 졸음을 판단하기 위해 운전자의 시각적 특징을 사용한다. 운전자의 시각적 데이터는 카메라를 부착하여 수집하기 때문에 운전 환경에 용이하다. 부착된 카메라를 통하여 수집된 이미지는 전처리 과정을 통하여 특징들을 추출한 후 추출

된 특징들의 변화를 관찰하여 운전자의 상태를 파악하는 구조를 가지며 이러한 구조 속에서 다양한 기법들을 사용한다.

먼저 운전자의 졸음 탐지에 필요한 얼굴과 눈을 추적한다<sup>2)</sup>. AdaBoost(adaptive boosting), blob detection을 사용하여 얼굴 영역과 눈 영역을 제외한 나머지 영역은 제거한다. 이후 눈의 유효성 검사는 SVM(Support Vector Machine) 분류기를 사용한다. SVM은 눈의 개폐 유무와 같이 이진 값을 판단할 때 사용하기 좋다. 이때 눈이 맞는지 아닌지를 판단하는 기반은 PCA(Principal Component Analysis)와 LDA(Linear Discriminant Analysis)를 사용한다. PCA는 이미지들의 주성분 벡터를 추출하며 LDA는 이미지들을 클래스별로 분류하는데 사용된다. 이렇게 유효성 검사를 통해 추출한 눈은 PCA와 LDA 그리고 다른 방법을 통한 총 28개의 특징(왼쪽, 오른쪽 눈 각각 14개 씩)을 추출한다. PCA와 LDA를 통해 양쪽 눈마다 12개의 특징을 추출한다. 나머지 2개의 특징은 소벨 윤곽선 검출(Sobel Filtering)을 통하여 얻은 이진 윤곽선 이미지를 수직 투영 그래프를 통해 그래프의 최소 점과 밀집 점이 특징이다. 이후 눈의 상태를 눈의 감긴 정도, 눈이 감긴 지속시간, 눈의 깜박임을 PERCLOS(percentage of eye closure), ECD(eye closure duration), FEC(frequency of eye closure)라고 정의한다. 눈의 감긴 정도는 눈꺼풀의 닫힌 정도가 아닌 운전자의 눈 개폐 유무를 0과 1로 표현하며 눈이 감겼을 때를 0이라고 한다. 어느 시간만큼을 클러스터로 지정하였을 때 0의 개수를 눈의 감긴 정도라고 한다. 이렇게 눈의 상태를 통해 운전자의 졸음을 결정하며 광범위한 운전자의 졸음을 감지할 수 있다.

운전자의 졸음 탐지에 필요한 얼굴과 입을 추적하며 Viola & Jones Algorithm을 사용하였다<sup>3)</sup>. 운전자의 입을 추출하여 추출된 입의 개폐 유무를 판단한다. 후에 열린 입은 SVM을 통하여 반 이상 열려 있는 입을 판단하여 하품의 유무를 판단한다. 이처럼 입의 개폐 유무만으로 하품을 판단하는 것이 아니라 입이 얼마나 벌어져 있는지 또한 판단하여 하품의 유무를 판단하여 피곤함을 결정한다.

이처럼 단일 특징만을 사용하여 판단된 졸음은 정확도가 높더라도 근거가 부족하다. 때문에 기존의 눈을 통하여 졸음을 판단함에 있어 추가적인 머리 움직임 사용하였다<sup>4)</sup>. 얼굴과 눈을 감지하기 위해 위 연구와 같은 알고리즘을 사용하였으며 총 6방향의 머리 움직임을 예측한다. 먼저 기준이 되는 머리 특징 점을 추출한다. 이전에서도 언급했듯이 움직임을 측정하기

위해서는 두 개 이상의 데이터가 존재하여야 하며 그 중 하나의 데이터는 기준이 되어야 한다. 다음으로 현재 운전자의 머리 위치를 추출한다. 이 두 데이터의 차이를 통하여 총 6가지 상태로 분류되며 그 상태는 오른쪽, 왼쪽, 뒤, 앞으로 젖힘이 있으며 오른쪽, 왼쪽 고개 돌림이 있다. 졸음 감지 시스템은 운전자의 눈만 감겨 있거나 또는 머리만 앞으로 젖혀 있다면 오렌지 상태(피곤한 상태)를 가진다. 운전자의 눈이 감겨있으며 머리가 앞이나 뒤로 젖혀있다면 상태가 레드(졸음 상태)가 되며 매우 위험한 상태임을 알림으로써 두 특징을 모두 사용하는 경우와 한 특징을 사용하는 경우가 나누어져 있다. 즉, 결정 트리를 통하여 추출한 특징 각 상태를 분류하는 것이다. 이러한 분류는 추출한 각 모달리티의 모든 특징을 고려했다고 보기에는 어려움이 있다<sup>5)</sup>. 추출한 특징들을 잘 융합하기 위하여 여러 가지 방법들이 제시되었으며 각 모달리티의 특징들을 모두 고려하는 방법들이 제시된다. 머신 러닝을 사용하여 각 모달리티를 잘 이해하는 모델을 생성한다면 여러 모달리티를 융합할 수 있다.

위와 같은 방법들은 운전자의 특징 데이터를 직접적인 방법으로 추출한다. 하지만 현재 딥 러닝의 발전으로 운전자의 얼굴 데이터를 CNN(Convolutional Neural Network)에 적용하여 졸음과 비졸음으로 분류하면 직접적인 특징 추출이 아닌 학습을 통한 운전자의 특징을 추출할 수 있다<sup>6)</sup>. 하지만 졸음이라는 현상은 시간을 가지고 나타나며 CNN에서는 시간적 변화를 포함한 특징을 추출하기에는 어려움이 있다.

### III. 본 론

졸음 판단의 근거를 강화하기 위해 눈, 입 등의 추가적인 특징을 사용하기 위해서는 각각 특징의 지속적인 변화를 관찰하여 졸음 상태를 판단해주는 시스템이 필요하다<sup>9)</sup>. 이때 시스템 설계에 따라 두 가지 특징이 모두 고려되어 졸음 판단 정확도가 향상되는지 알 수 없으며 값의 누락을 고려하지 않았다면 문제가 발생 할 수 있다. 이와 같은 문제를 해결하기 위하여 본 연구에서는 그림 1와 같은 시스템 설계를 제안한다. 운전자의 이미지로부터 추출한 특징은 눈과 입, 두 가지 특징을 사용한다. 본 연구에서 눈과 입의 특징을 사용한 이유는 졸음 상태는 순간의 판단이 아닌 지속적인 관찰을 통한 판단이기 때문이다. 운전자의 상태가 졸음 직전 피곤함으로 인해 잦은 하품을 한 후 입의 움직임은 관찰되지 않았다면 피곤한 상태 후 졸음 상태를 생각할 수 있다. 하지만 반대로 운전자가

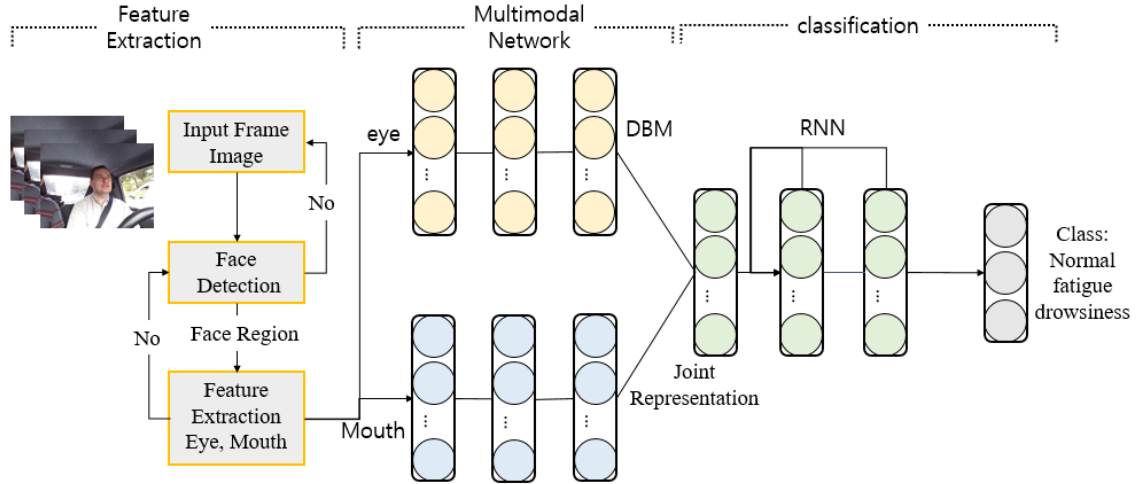


그림 1. 졸음 감지를 위한 시스템 구조도  
Fig. 1. System structure for detecting drowsiness

졸음을 극복하여 정상 상태 또한 가능할 수 있으며 이때 눈의 상태가 감겨져 있다면 확실한 졸음이라는 것을 판단할 수 있다.

### 3.1 특징 추출

수집된 운전자의 이미지로부터 전처리를 통하여 졸음 판단에 필요한 특징을 추출한다. 먼저 Viola & Jones Algorithm을 사용하여 얼굴 영역을 추출한다<sup>[7]</sup>. 이 기법은 수 백 가지의 하르(haar) 피쳐 조합을 이용하여 물체를 탐지하며 모든 사람들의 얼굴 특징이 뚜렷하며(눈과 눈썹은 확실한 음영을 가지고 있음) 유사하게 생겼기 때문에 얼굴 탐지에서 좋은 성능을 보인다. 이 방법은 2001년 제안된 이후로 많은 발전이 이루어져 왔고 얼굴 탐지뿐만 아니라 눈, 입 등 얼굴의 특징적인 부분을 탐지하는 좋은 알고리즘으로 발전해 왔다. 이러한 기법을 사용하여 얼굴 영역을 제외해 배경은 제거한다. 이후 얼굴 영역에서 하르 피쳐를 이용하여 adaboost를 통하여 데이터 구성에 필요한 눈과 입을 추출한다.

추출한 각각의 특징 데이터는 졸음 판단을 위한 각각의 모달리티로 구성되며 지속적인 관찰을 위한 데이터로 시퀀스한 성질을 가진다.

### 3.2 멀티 모달 네트워크

추출한 특징으로 구성된 각각의 모달리티는 서로 다른 형식을 가지기 때문에 두 모달리티 간의 이질성이 존재한다. 비록 같은 이미지 데이터로 구성되어 있지만 졸음을 표현하는데 있어 두 모달리티는 서로 다

른 변화를 가지기 때문이다. 이러한 문제를 해결하기 위하여 멀티 모달 네트워크에서는 두 모달리티를 재구성하는 생성 모델을 활용한다.

#### 3.2.1 데이터 특징 학습

두 모달리티를 재구성하는 생성 모델을 학습하기 위하여 본 연구에서는 DBM(Deep Boltzmann Machine)<sup>[8]</sup>을 사용한다. DBM은 깊게 쌓은 확률 모델로 RBM(Restricted Boltzmann Machine)으로 구성되어 있다. RBM을 구성하고 있는 입력 레이어의 노드들은  $\mathbf{h} \in \{0, 1\}^F$ , 히든 레이어 노드는  $\mathbf{v} \in \{0, 1\}^D$ , 두 레이어는 각 노드마다 서로 연결되어 있으며 같은 레이어안에 존재하는 노드는 연결되어 있지 않다. 이처럼 입력 레이어와 히든 레이어로 이루어져 있다. RBM의 목적은 입력 데이터를 재구성하는 비지도학습 모델이며 에너지 함수:  $E(\mathbf{v}, \mathbf{h}) = -\mathbf{b}^T \mathbf{v} - \mathbf{c}^T \mathbf{h} - \mathbf{v}^T \mathbf{W} \mathbf{h}$ 를 기반으로 정의

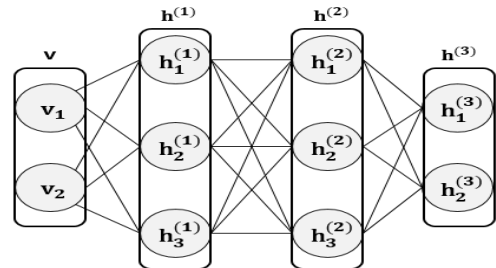


그림 2. DBM 그래픽 모델  
Fig. 2. DBM graphical model

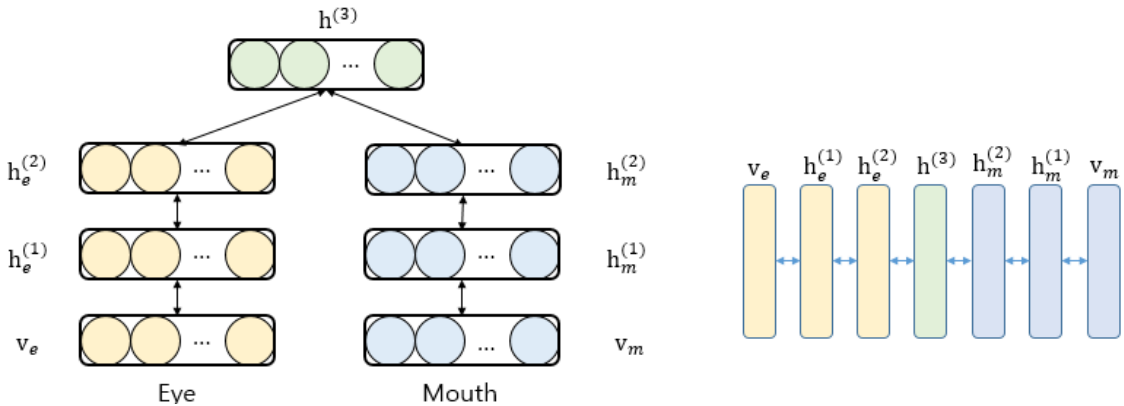


그림 3. multimodal DBM 그래픽 모델  
Fig. 3. multimodal DBM graphical model

된다.

입력 데이터가 주어지면 각각의 웨이트 값과 바이어스 값을 통하여 히든 노드 값을 추정한다. 후에 히든 노드 값을 역으로 재구성하여 입력 데이터와 재구성 데이터의 차이를 줄여가서 입력데이터와 가장 근사한 확률 분포를 가지게 되며 입력 데이터와 같은 확률 분포를 가졌을 때 히든 레이어는 입력 데이터를 가장 잘 이해하고 있는 모델이라고 정의 할 수 있다. 이렇게 생성된 모델은 입력 데이터를 재구성하여도 비슷한 확률 분포를 가지기 때문에 입력 데이터를 대신 할 수 있으며 결국 RBM은 입력 레이어와 히든 레이어의 결합 확률 분포로 나타낼 수 있으며  $v$ 는 입력 레이어,  $h$ 는 히든 레이어를 나타내며 다음과 같은 수식을 가진다.

$$P(v = \mathbf{v}, h = \mathbf{h}) = \frac{1}{Z} \exp(-E(\mathbf{v}, \mathbf{h})) \quad (1)$$

이러한 RBM을 깊게 쌓아 고차원적인 데이터를 재구성하기 위한 방법이 DBM이며 깊은 층을 가진 생성 모델이다. 기존의 2개의 층으로 이루어진 것보다 더 많은 층을 가지며 전체 레이어의 결합 확률 분포를 의미한다. 그래픽 모델에서 노드를 연결한 두 선은 결합 확률 분포를 의미하며 그림 2와 같이 RBM을 쌓아 구성한 모델이며 연속된 RBM이므로 전체레이어의 결합 확률 분포를 나타내며 이때  $\theta$ 값은 웨이트 값이다.

$$P(v, h^{(1)}, h^{(2)}, h^{(3)}) = \frac{1}{Z(\theta)} \exp(-E(v, h^{(1)}, h^{(2)}, h^{(3)}; \theta)) \quad (2)$$

데이터 재구성에 있어서 전체 모델의 변수를 수정하며 학습하기 때문에 더 추상적이고 복잡한 데이터를 이해할 수 있다.

### 3.2.2 융합 네트워크

두 모달리티를 DBM을 사용하여 재구성하였으면 마지막으로 재구성된 두 모달리티를 조인해야 한다. 두 데이터를 하나의 레이어로 표현하기 위해 본 연구에서는 그림 3와 같은 구조를 사용한다. 두 모달리티 눈과 귀를 각각  $v_e, v_m$  이라고 하며 입력 레이어는  $v = \{v_e, v_m\}$ 이며 기존 DBM을 표현하는 수식 (2)를 기반으로  $h^{(3)}$ 을 구할 수 있다. 두 모달리티 모두  $h^{(3)}$ 을 추정하는 조건이며  $P(h^{(3)} | v_e, v_m)$ 으로 나타낼 수 있다.

### 3.2.3 생성 모델

본 연구에서 사용하는 눈과 입 데이터는 줄음 변화에 있어 서로 다른 특징을 가지기 때문에 위와 같은 네트워크를 사용하여 각 데이터의 특징을 추출하여 융합한다. joint representation을 통해 생성된 레이어는 그림 4와 같이 이미지로 표시할 때 각 모달리티인 눈과 입의 형태를 찾아 볼 수는 없지만 생성된 모델을 통하여 재구성한 이미지를 확인하였을 때는 눈과 입의 모습을 확인할 수 있다. 즉, 생성된 모델이 재구성한 이미지를 확인하여 본 연구에서 제시하는 각 모달리티의 이질성을 제거하는 모델을 검증한다. 멀티 모달 네트워크에서는 각 모달리티의 특징을 잘 이해하는 모델이 생성되며 합쳐진 모습에서는 눈과 입을 확인할 수는 없지만 두 모달리티의 특징을 포함하고 있

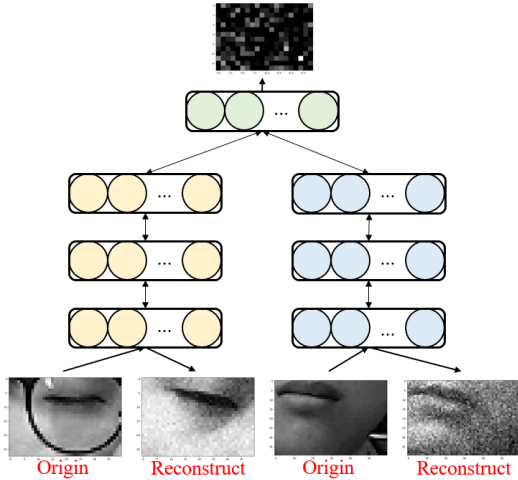


그림 4. 모델 검증  
Fig. 4. Model verification

는 데이터가 생성된다.

### 3.3 졸음 분류

졸음은 시간 축을 가지는 특징으로 지속적인 변화를 관찰하여 파악할 수 있다. 이러한 특성 때문에 본 연구에서는 운전자 상태 분류를 위하여 RNN (Recurrent Neural Network)을 사용한다. RNN 은 과거에 입력 받았던 데이터를 동시에 고려하는 딥 러닝 네트워크로 인간이 머릿속에 기억을 저장하고 있듯이 히든 레이어에서 기억을 가지고 있지만 모든 것을 기억할 수 없으므로 일정 상황에 맞게 과거에 입력 받았던 데이터는 사라진다. 현재 입력 데이터와 이전 입력 데이터를 모두 고려하는  $s^{(t)} = f(s^{(t-1)})$  수식을 가지며 이때  $s^{(t)}$  는 현재 상태를 의미한다. 이처럼 t 값을 고려하는 타임스텝 파라미터를 조정하여 일정 시간동안 입력 데이터의 지속적인 관찰이 가능하다. 데이터 전처리를 통해 특징 데이터 값은 세로축 y로 기존 데이터의 가로축과 세로축을 곱한 크기를 가진다. 그 다음 졸음 판단을 하기 위해 필요한 시간만큼을 x축으로 쌓는다. 이로 인해 x축은 시간 t축이 되며 t 값 하나가 이미지 데이터 하나를 표현한다. 그림 5와 같은 데이터 구성을 가지고 있어 시퀀스한 성질을 가지며 지속적인 관찰로 인해 졸음 판단을 가능하게 해준다. 멀티 모달 네트워크를 통하여 각 모달리티의 특징을 모두 포함하고 있는 데이터를 총 3가지 클래스로 분류하며 관련 연구<sup>[4]</sup>에서 제시한 운전자 상태를 참고하였다.

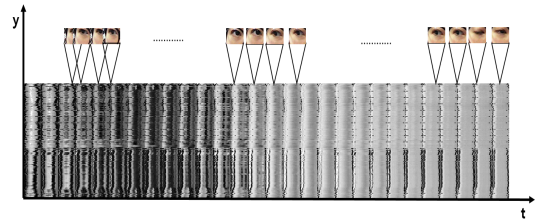


그림 5. 입력 데이터 구성  
Fig. 5. Input data configure

## IV. 실험

### 4.1 실험 내용

실험 데이터는 실내에 마련한 가상 운전 환경을 통하여 수집하였다. 비록 실제 운전 환경과 완전히 같다고 할 수는 없지만 졸음 데이터 특성상 수집이 어렵고 위험하기 때문에 실내 가상 환경 실험을 진행하였다. 실험을 통하여 수집한 시퀀스 이미지는 약 20만장으로 일반인 4명을 통하여 수집하였다. 눈과 입 특징을 추출하는 전처리 과정에서 특징 추출이 정상적으로 이루어지지 않는 경우(운전자가 존재하지 않는 경우 등)가 발생하고 이러한 이유로 실제 학습 데이터보다 추출된 특징 데이터가 더 적은 데이터 수를 가진다. 실험 참가자의 상태는 관련 연구<sup>[4]</sup>에서 레드, 오렌지, 그린 상태로 분류한 것과 같이 졸음, 피곤함, 정상 상태로 분류하였으며 실험에 앞서 자신의 상태를 데이터 수집 도중에 시간별로 파악하여 기록하도록 하였기 때문에 수집한 데이터는 실험 참가자의 상태 정보를 포함하고 있다. 테스트 데이터는 각 실험 참가자의 데이터를 일정부분 추출하여 구성하였다.

본 연구에서 제안하는 멀티 모달 학습의 성능 평가를 위하여 단일 모달리티를 사용한 졸음 판단 실험도 진행하였다. 단일 모달리티를 RNN을 통해 학습하여 생성된 모델에 데이터 셋을 입력하여 정확도를 측정 한 후 멀티 모달 학습을 통해 생성된 모델의 정확도와 비교할 것이다. 또한 관련 연구에서 졸음을 판단하기 위해 CNN을 사용한 결과 값<sup>[6]</sup>과 직접적인 결정 트리를 가지고 판단한 연구<sup>[4]</sup> 결과 값을 인용하여 비교할 것이다.

### 4.2 실험 결과

관련 연구<sup>[6]</sup>에서 졸음을 감지하는 정확도를 제시하기 때문에 본 연구의 정확도와 비교가 가능하다. 또한 본 연구와 다른 신경망 네트워크를 사용한<sup>[4]</sup> 연구는 적절한 비교가 가능하다. RNN을 이용하여 단일 특징 실험 결과는 그림 6과 같다. 각각의 데이터 셋에 대한 정확도

표 1. 데이터 정보  
Table 1. Dataset

extract feature	eye	mouth
training images	210,554 images	
feature images	361,964	180,982
test images	20,120 images	
feature images	360,42	180,21

를 보았을 때 두 모달리티를 모두 활용한 멀티 모달 학습이 더 높은 정확도를 보인다. 그에 반해 입만을 활용하여 생성된 모델은 정확도가 낮은 것을 확인할 수 있다. 실제로 입만을 통하여 줄음을 판단하는 것은 어려운 작업이다. 하지만 입과 눈을 함께 활용한다면 단순히 눈만 감았을 때 판단한 줄음보다는 잦은 하품 이후 눈을 감았을 때 더 설득력 있는 줄음 상태 판단이라고 할 수 있다.

입으로 파악할 수 있는 줄음 패턴은 한계가 존재하기 때문에 실제 다른 연구에서도 하품을 인지하는 특징으로 주로 사용한다. 마지막으로 눈과 입, 두 모달리티를 멀티 모달 네트워크에서 이질성을 제거하여 재구성한 하나의 입력 데이터를 생성한 후 RNN에 적용하여 줄음 판단을 위한 모델을 생성하였다. 그리고 생성된 모델에 같은 테스트 데이터를 입력해 주었을 때 단일 모달리티를 사용하였을 때 보다 높은 정확도를 보였다. 실제 입력 데이터는 시퀀스한 이미지이기 때문에 입의 변화와 눈의 변화를 모두 활용하면 단일 모달리티에서 파악하지 못했던 줄음 패턴을 찾을 수 있다. 또한 두 모달리티는 줄음 상태에서 확실하게 나타나는 변화(일반적으로 줄음 상태에 입의 움직임은 관찰되기 어려움, 눈의 닫힘 등)를 가지고 있고 멀티 모달 네트워크를 통하여 두 모달리티의 상관관계를 파악할 수 있는 데이터로 재구성하였기 때문에 높은 정

표 2. 실험 결과  
Table 2. Experiment result

workd	Accuracy
RNN(eye)	79.1
RNN(mouth)	67.2
<b>Multimodal</b>	<b>90.4</b>
TEYEB, lnes, et al. [4]	88
DWIVEDI, kartik, et al.[6]	88.57

확도를 가지는 결과 값을 얻을 수 있다.

이와 같은 결과 값의 객관적 판단을 위하여 두 특징(목의 움직임, 눈)을 사용한 시스템 결과 값과 CNN을 사용하여 생성된 모델의 줄음 감지 정확도를 표 2를 통하여 비교하였다. 관련연구 [6]은 두 가지 특징을 사용하지만 설계된 알고리즘으로 상태를 파악하기 때문에 두 특징을 모두 활용한다고 보기 어려우며 유연하지 못하다. 관련연구 [4]는 유연성있는 판단이 가능하지만 CNN 특성상 시간적인 변화를 고려했다고 보기 어렵다. 본 연구에서는 줄음이라는 시간적 변화를 가지고 나타나는 현상을 감지하기 위하여 RNN을 사용하였으며 두 특징을 모두 활용하기 위하여 멀티 모달 네트워크를 통하여 두 모달리티의 특징을 학습하였다. 그 결과 기존 연구의 정확도보다 향상된 정확도를 확인할 수 있으며 본 연구에서 사용하는 시스템은 추가적인 특징을 가져올 때 더 나은 확장성을 가질 수 있다.

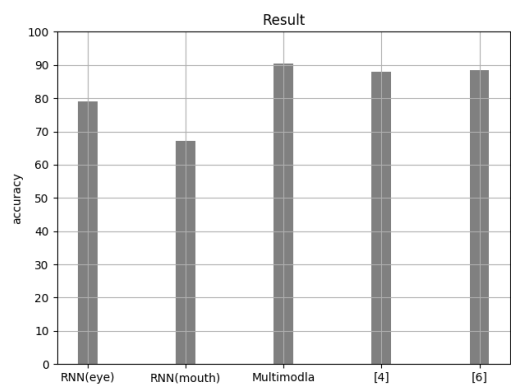


그림 6. 결과 값  
Fig. 6. Result

### V. 결론 및 향후 연구

기존의 시각 기반 줄음 판단 시스템에서는 이미지에

서 추출한 특징을 관찰하여 졸음 발생 시 나타나는 변화가 감지되면 졸음을 결정하였다. 이는 정해진 알고리즘을 통한 결정 트리이며 이때 시각적 특징들마다 나타나는 변화들은 모두 차이를 가지기 때문에 시스템마다 하나 이상의 특징을 관찰할 때 고려할 사항이 생기며 추가적인 시스템을 구현하여 두 개 이상의 특징을 통해 졸음을 판단할 때 문제점을 갖는다. 먼저 졸음 판단 기준의 편향 현상이며 두 개 이상의 특징을 고려하였을 때의 추가적인 알고리즘 도입이 필요한 확장성 문제이다. 본 연구에서는 이러한 시각적 특징의 활용성을 높이며 문제점을 해결하기 위해 멀티 모달 학습 방법을 제안하였다. 데이터 재구성을 통하여 서로 다른 형식의 데이터를 받을 수 있기 때문에 두 개 이상의 특징 데이터를 입력받을 수 있다.

딥러닝을 통한 이미지 처리 분야에서 CNN이 높은 성능을 가지고 있지만 운전자의 얼굴을 입력해 주었을 때 졸음인지 아닌지를 판단해 주기에는 어려움이 있다. 졸음이라는 현상이 시간을 가지고 나타나는 시퀀스한 성질을 가지기 때문이다. 이러한 이유로 본 연구에서는 RNN을 사용하였으며 운전자의 상태를 세 가지로 분류하였다. 비록 지금은 위험성으로 인해 제한된 상황에서 수집된 데이터를 사용하였지만 향후 장거리 운전자의 데이터를 수집하여 실험을 진행할 계획이다. 또한 현재 얼굴에서만 추출한 특징들을 사용하였지만 생리화학적 특징, 운전 패턴 등 다양한 데이터를 수집하여 졸음 판단에 사용하면 현재 제안된 시스템의 용이한 확장성과 다양한 졸음 판단 기준들을 통하여 졸음 감지가 가능할 것이다.

## References

[1] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning." in *Proc. ICML-11*, pp. 689-696, 2011.

[2] J. Jo, S. J. Lee, K. R. Park, I. J. Kim, and J. Kim, "Detecting driver drowsiness using feature-level fusion and user-specific classification." *Expert Systems with Appl.*, vol. 41, no. 4, pp. 1139-1152, 2014.

[3] M. Saradadevi and P. Bajaj, "Driver fatigue detection using mouth and yawning analysis," *IJCSNS*, vol. 8, no. 6, pp. 183-188, Jun. 2008.

[4] I. Teyeb, O. Jemai, M. Zaied, and C. B. Amar, "A novel approach for drowsy driver detection using head posture estimation and

eyes recognition system based on wavelet network," *IEEE IISA 2014*, pp. 379-384, Jul. 2014.

[5] V. Radu, C. Tong, S. Bhattacharya, N. D. Lane, C. Mascolo, M. K. Marina, and F. Kawsar, "Multimodal deep learning for activity and context recognition," in *Proc. ACM on Interactive, Mob., Wearable and Ubiquitous Technol.*, vol. 1, no. 4, article 157, Nov. 2018.

[6] K. Dwivedi, K. Biswaranjan, and A. Sethi, "Drowsy driver detection using representation learning," in *2014 IEEE IACC*, pp. 995-999, Gurgaon, India, Feb. 2014.

[7] P. Viola, and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. CVPR 2001*, vol. 1, Dec. 2001.

[8] N. Srivastava and R. Salakhutdinov, "Multimodal learning with deep boltzmann machines," *NIPS'12*, vol. 2, pp. 2222-2230, Lake Tahoe, Nevada, Dec. 2012.

[9] H.-T. Choi, M.-K. Back, and K.-C. Lee, "Design of multi-modal learning based system to improve driver drowsiness detection based on visual-feature," in *Proc. KICS Fall Conf.*, pp. 391-392, 2017.

## 최형탁 (Hyung-tak Choi)



2017년 2월 : 충남대학교 컴퓨터공학과 졸업  
 2017년 3월~현재 : 충남대학교 컴퓨터공학과 석사과정  
 <관심분야> 기계 학습, 데이터 베이스시스템, 비전 컴퓨팅



**백 문 기 (Moon-ki Back)**



2013년 2월 : 충남대학교 전기  
정보통신공학과 졸업  
2015년 2월 : 충남대학교 컴퓨  
터공학과 석사  
2015년 3월~현재 : 충남대학교  
컴퓨터공학과 박사과정  
<관심분야> 기계 학습, 데이터  
베이스시스템, 딥 러닝, 이상 탐지

**강 재 식 (Jae-sik Kang)**



2015년 2월 : 충남대학교 컴퓨  
터공학과 졸업  
2018년 3월~현재 : 충남대학교  
컴퓨터공학과 석사과정  
<관심분야> 기계 학습, 데이터  
베이스시스템, 신호 처리

**이 규 철 (Kyu-chul Lee)**



1984년 2월 : 서울대학교 컴퓨  
터공학과 학사  
1986년 2월 : 서울대학교 컴퓨  
터공학과 석사  
1990년 2월 : 서울대학교 컴퓨  
터공학과 박사  
1990년 3월~현재 : 충남대학교  
컴퓨터공학과 교수

1994년 : 미국 IBM Almaden Research Center 초  
빙연구원

1995년~1996년 : 미국 Syracuse University, CASE  
Center 초빙 교수

2009년~현재 : 웹코리아 포럼 의장

2009년~현재 : 정보통신사업진흥회 포럼 의장

<관심분야> 데이터베이스, XML, 정보 통합, 사물  
인터넷, 기계 학습