

## 색상 군집화를 이용한 입술탐지 알고리즘

정 종 면\*

# A Lip Detection Algorithm Using Color Clustering

Jongmyeon Jeong\*

### 요 약

본 논문에서는 색상 군집화를 이용한 입술탐지 알고리즘을 제안한다. RGB 색상 모델로 주어진 입력영상에서 AdaBoost 알고리즘을 이용하여 얼굴영역을 추출한 후, 얼굴영역을 Lab 컬러 모델로 변환한다. Lab 컬러 모델에서 a 성분은 입술과 유사한 색상을 잘 표현할 수 있는 반면 b 성분은 입술의 보색을 표현할 수 있기 때문에 Lab 컬러로 표현된 얼굴영역에서 a와 b 성분을 기준으로 최단 이웃(nearest neighbour) 군집화 알고리즘을 이용하여 피부 영역을 분리한 후, K-means 색상 군집화를 통해 입술 후보 영역을 추출하고, 마지막으로 기하학적 특징을 이용하여 최종적인 입술영역을 탐지하였다. 실험 결과는 제안된 방법이 강건하게 입술을 탐지함을 보인다.

▶ Keywords : 입술탐지, Lab 색상 모델, 최단 이웃 군집화, K-means 군집화

### Abstract

In this paper, we propose a robust lip detection algorithm using color clustering. At first, we adopt AdaBoost algorithm to extract facial region and convert facial region into Lab color space. Because a and b components in Lab color space are known as that they could well express lip color and its complementary color, we use a and b component as the features for color clustering. The nearest neighbour clustering algorithm is applied to separate the skin region from the facial region and K-Means color clustering is applied to extract lip-candidate region. Then geometric characteristics are used to extract final lip region. The proposed algorithm can detect lip region robustly which has been shown by experimental results.

▶ Keywords : Lip detection, Lab color model, the nearest neighbour clustering, K-means clustering

---

•제1저자 : 정종면 •교신저자 : 정종면

•투고일 : 2014. 1. 21. 심사일 : 2014. 2. 3. 게재확정일 : 2014. 2. 18.

\* 목포해양대학교 해양컴퓨터공학과(Dept. of Computer Engineering, Mokpo National Maritime University)

## I. 서론

최근 인간 컴퓨터 상호작용(Human-Computer Interaction: HCI)의 중요성이 커지고 있다. 인간 컴퓨터 상호작용을 위한 영상처리 기술은 카메라를 통해 입력되는 영상으로부터 인체를 탐지하는 것이 선행되어야 한다. 인체탐지 기술은 지문, 홍채, 얼굴, 목소리 등 사람의 인체 정보를 또는 인간의 움직임 등 행동정보를 이용하여 사람을 인지하는 기술을 말하는데, 영상 정보를 통한 인간 컴퓨터 상호작용 기술은 게임을 비롯한 엔터테인먼트, 스마트 기기, 지능형 홈을 비롯한 다양한 분야에서 응용될 수 있다.

입술탐지는 다양한 응용 분야에서 시각 정보를 활용한 직접적인 서비스 또는 간접적인 서비스를 가능하게 하기 때문에 최근에도 활발한 연구가 이루어지고 있다. 인간 컴퓨터 상호작용, 잡음이 심한 환경에서의 음성, 화자 인식 등에서 영상-시각(audio-visual) 융합 기술로서 입술 탐지 기술은 효과적으로 적용될 수 있다. 입술 탐지를 위한 기존의 방법들은 영상에 기반한 방법(image based method), 모델에 기반한 방법(model based method), 하이브리드 방법(hybrid method)으로 나눌 수 있다.

영상에 기반한 방법은 RGB, HSI, YCvCr 등의 칼라 픽셀 정보를 직접 이용하여 입술을 탐지하는 것으로서 계산량이 상대적으로 적은 반면 조명의 변화 등에 상대적으로 취약한 문제가 있다[1]-[5]. 한편 모델에 기반한 방법은 입술의 형태에 대한 사전 정보를 바탕으로 입술을 탐지하는 방법이다. 이 방법은 매우 강건하게 입술을 탐지할 수 있지만 일반적으로 영상에 기반한 방법보다 계산량이 방대하다는 문제가 있다[6][7]. 하이브리드 방법은 영상 기반 방법과 모델 기반 방법을 결합하여 입술을 탐지하는 방법이다. 영상 기반 방법은 계산량은 적지만 조명 등에 취약하다는 문제가 있고 모델에 기반한 방법은 강건하지만 계산량이 많기 때문에 하이브리드 방법에서는 입술 후보 영역을 탐지하기 위하여 영상 기반 방법을 사용한 후, 정확한 입술을 얻기 위해 모델 기반 방법을 사용한다[8][9].

Chin 등은 워터셰드 분할(watershed segmentation)을 이용하여 입술을 탐지하는 방법을 제안하였는데, 분할의 성능에 민감하게 반응하기 때문에 정확한 입술 추출이 어렵다[10]. Wang 등은 Haar-like Feature와 SVM system을 통한 입술탐지 기법을 제안하였는데, 실시간 처리가 어렵다는 문제가 있다[11]. Huang 등은 최소 사각형을 기준으로 타원 경계와 2계 도함수에 의해 입술 이미지를 얻었다. 하지만 카

메라 운동 등 배경의 움직임에 취약하며, 기하학적 특징만을 이용하기 때문에 정확한 입술 특징 추출에 어려움이 있다[12]. Pingxian 등은 얼굴과 코에 대한 입술의 상대적인 위치를 사용하는 방법을 제안하였는데, 이 방법도 기하학적 특징만을 사용하며, 초기 위치 설정의 오류가 입술 전체영역에 영향을 미친다[13]. Skodras 등은 컬러공간과 입술 포인트의 각도를 통해 입술탐지를 하는 방법을 제안하였다. 이 방법은 입술색상의 특징과 기하학적 특징 두 가지를 사용하기 때문에 비교적 정확한 탐지 성능을 보이지만 입술 포인트의 각도는 입술모양 변화에 민감하기 때문에 입술영역에 따라 결과 차이가 크다는 단점을 가지고 있다[14]. 본 논문에서는 Skodras 등의 방법을 개선하여, 입술모양 변화에 강건한 입술탐지 알고리즘을 제안한다.

본 논문의 구성은 다음과 같다. 먼저 2.1에서 AdaBoost 알고리즘을 통한 얼굴탐지에 대해 기술한다. 2.2에서는 Lab 컬러시스템을 이용한 피부영역 추출 방법을 설명한다. 2.3에서 색상군집화에 의한 입술영역 결정을 설명한 후 3장에서 실험 결과를 보이며, 4장에서 결론을 맺는다.

## II. 본론

본 논문에서는 먼저 주어진 입력 영상으로부터 AdaBoost 알고리즘[15]을 이용하여 얼굴 영역을 탐지한 후, 얼굴 영역

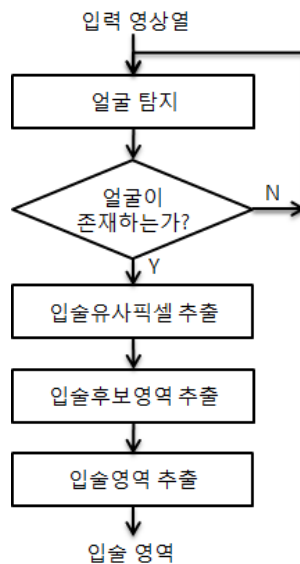


그림 1. 제안된 방법 흐름도  
Fig. 1. Flowchart of the proposed method

으로부터 색상에 대한 최단 이웃 군집화 알고리즘(nearest neighbour clustering algorithm)을 이용하여 입술 유사 픽셀을 추출하고, 추출된 입술 유사 픽셀을 K-means 군집화 알고리즘을 이용하여 입술 후보 영역을 얻은 다음, 기하학적인 검증 과정을 거쳐 입술 영역을 탐지한다. 그림 1은 제안된 입술탐지방법의 전체적인 흐름도를 보이고 있다.

### 1. 얼굴탐지

주어진 입역 영상으로부터 얼굴을 탐지하기 위하여 본 논문에서는 잘 알려진 AdaBoost 알고리즘을 사용하였다[15]. AdaBoost 알고리즘은 분류 성능이 약한 분류기를 조합하여 분류 성능이 우수한 강한 분류기를 만들어 나가는 알고리즘으로써, 약한 분류기를 여러 단계로 증가시키면서 이전에 분류된 분류기에서 잘못 분류된 특징들에 대해 가중치를 높게 하여 다음분류기에 영향을 받도록 한다. 그림 2는 AdaBoost 알고리즘을 보이고 있다.

```

AdaBoost Algorithm
Begin Initialize  $D = \{x^1, y_1, \dots, x^n, y_n\}, k_{max}, w_1(i) = 1/n, i=1, \dots, n$ 
 $k \leftarrow 0$ 
do  $k \leftarrow k+1$ 
    train weak learner  $C_k$  using  $D$  sampled according to  $W_k(i)$ 
     $E_k \leftarrow$  training error of  $C_k$  measured on  $D$  using  $W_k(i)$ 
     $\alpha_k \leftarrow \frac{1}{2} \ln((1 - E_k) / E_k)$ 
     $W_{k+1}(i)$ 
 $\leftarrow \frac{W_k(i)}{Z_k} \times \begin{cases} e^{-\alpha_k} & \text{if } h_k(x^i) = y_i \text{ (correctly classified)} \\ e^{\alpha_k} & \text{if } h_k(x^i) \neq y_i \text{ (}\in\text{correctly classified)} \end{cases}$ 
until  $k = k_{max}$ 
return  $C_k$  and  $\alpha_k$  for  $k=1$  to  $k_{max}$ 
End
    
```

그림 2. AdaBoost 알고리즘  
Fig. 2. AdaBoost algorithm

AdaBoost 알고리즘을 이용하여 얼굴을 탐지하기 위하여 Viola 등은 Harr-like Feature를 이용하였는데, 빠른 연산을 위하여 입력 영상에 대한 적분 이미지(integral image)를 먼저 구한다. 그런 다음 주어진 영상을 미리 정의된 일정 크기(sub window)로 분할하여 영상 전체를 스캔하면서 그림 3에서 보이는 것과 같은 마스크를 이용하여 하나의 백색 영역과 흑색 영역의 평균 밝기를 구한 후, 그 차이가 미리 정의된 임계치(threshold)를 넘으면 그 마스크에 대한 영역을 특징으로 추출한다[16].

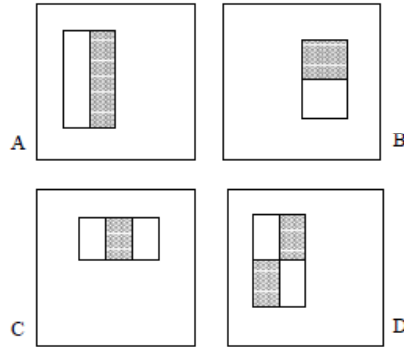


그림 3. Haar-like Feature 추출 마스크  
Fig. 3. An example of mask for Haar-like feature

그림 3에서 A부터 D 내부의 사각형은 서로 다른 Haar-like Feature를 나타낸다. 크기와 종류가 다른 Haar-like Feature들을 추출하고, 이를 AdaBoost 알고리즘에 적용하여 얼굴영역을 추출한다. AdaBoost는 구현이 매우 간단하고, 일반적인 학습 스키마를 이용하기 때문에 다양한 응용이 가능하다. 본 논문에서는 OpenCV에서 제공되고 있는 AdaBoost 라이브러리를 사용하여 정면의 얼굴을 탐지하였는데, 이 라이브러리는 97%의 탐지율을 보이는 것으로 알려져 있다. 그림 4는 얼굴탐지 결과를 보이고 있다.



그림 4. AdaBoost 알고리즘을 이용한 얼굴 탐지 결과  
Fig. 4. Results of facial region detection using AdaBoost algorithm

### 2. 입술 유사 색상 픽셀 추출

II.1에서 얻은 얼굴영역으로부터 피부 영역을 추출하기 위하여 본 논문에서는 색상 군집화 알고리즘을 사용하였다. 이를 위하여 먼저 주어진 얼굴 영역의 색상 모델을 Lab 컬러 공간으로 변환시킨다. Lab 컬러 공간은 인간 감성에 접근하기 위하여 연구된 결과로 인간이 색체를 감지하는 노랑-파랑, 초록-빨강의 반대 색상에 기초하여 CIE에서 정의한 색상 공간이다[17]. Lab 컬러 공간에서 L은 인간 시각에서 빛 즉 명도와 매우 유사한 값을 보이는데, 0~100까지의 단계로 소수점 이하 단위도 표현한다. a와 b는 색상을 표현하는데 +a는 붉

은색, -a는 초록색 방향을 나타내며, +b는 노랑, -b는 파랑 색을 각각 나타낸다. 그림 5는 Lab 컬러 모델의 개념을 보이고 있다.

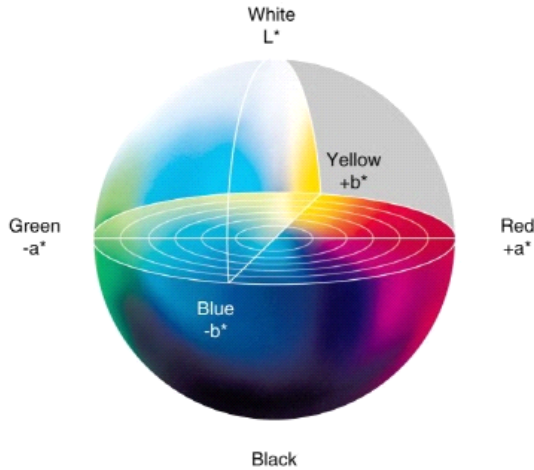


그림 5. Lab 컬러 모델 개념  
Fig. 5. Concept of Lab color model

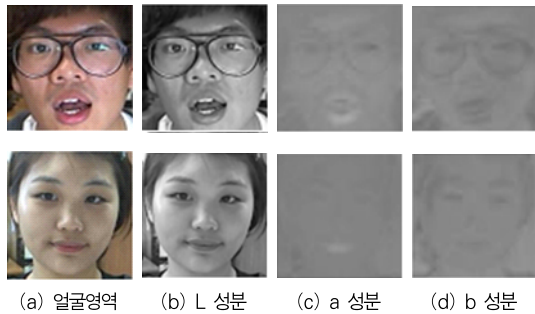


그림 6. Lab 컬러 모델의 특징  
Fig. 6. Characteristics of Lab color model

여러 색상 모델 중 피부 영역 추출을 위해 Lab 컬러 모델을 이용한 이유는 인종, 성별, 입술에 대한 화장 등에 관계없이 거의 대부분의 입술 영역은 붉은 색 성분이 많은 반면 초록색 값을 갖는 입술은 극히 드물기 때문이다. 즉 Lab 컬러 모델의 a 성분은 입술과 비 입술 영역을 다른 어떤 색상 모델보다 잘 구별할 수 있는 특징을 가지기 때문에 Lab은 다른 컬러 모델에 비해 입술을 탐지하는데 적합한 컬러모델이라고 할 수 있다. 그림 6은 RGB 컬러 모델로 표현된 얼굴 영역을 Lab 컬러 모델로 변환한 결과를 보이고 있는데, 그림 6(b)에서 보이는 바와 같이 입술 영역은 항상 다른 영역에 비해 상대적으로 높은 a 값을 갖는 것을 볼 수 있다[14].

본 논문에서는 참고문헌 [14]에서와 같이 Lab 색상 모델의 이런 특징을 이용하여 입술을 탐지하는데, 먼저 RGB 색상 모델로 주어진 얼굴영역을 Lab 색상 모델로 변환한 다음 얼굴 영역에서 가장 a성분이 높은 픽셀 10개와 가장 a성분이 낮은 픽셀 10개를 각각 선택하여, a 성분이 가장 높은 10 픽셀의 평균 a 값과 b값의 평균  $\mu_1 = (a1_\mu, b1_\mu)$ 을 구하여 이를 입술과 유사한 색상의 마커로 삼고, 또한 a 성분이 가장 낮은 10픽셀의 평균 a 값과 b값의 평균  $\mu_2 = (a2_\mu, b2_\mu)$ 을 구하여 입술 색상과 반대 색상의 마커로 삼는다.

주어진 얼굴 영역의 픽셀들을  $X = \{x_1, x_2, \dots, x_n\}$ 이라고 했을 때 얼굴 영역으로부터 입술 유사 색상 픽셀 추출은 다음과 같은 최단 이웃 군집화 알고리즘[18]을 이용하여 얼굴 영역 픽셀들을 입술과 유사한 색상과 그 반대 색상으로 분류한다.

$$\begin{aligned} \text{if } \|\mu_1 - x_i\| > \|\mu_2 - x_i\| & \quad \text{입술유사색상} \\ \text{Otherwise} & \end{aligned}$$

입술 유사색상 픽셀을 얻은 후, 얼굴 영역의 모든 픽셀들을 두 개의 마커를 기준으로 최단 이웃 군집화 알고리즘을 이용하여 얼굴 영역 픽셀들을  $\mu_1$ 에 가까운 입술과 유사한 색상과  $\mu_2$ 에 가까운 입술반대 색상으로 분류한다. 그림 7은 추출된 입술 유사 색상 픽셀을 보이고 있다.



그림 7. 입술 유사 색상 추출  
Fig. 7. Extraction of lip similar color

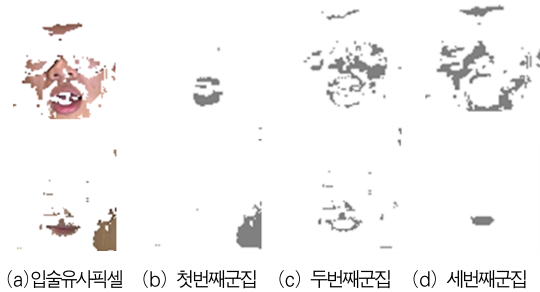


그림 8. 입술유사픽셀에 대한 K-means 군집화 결과  
Fig. 8. Result of K-means clustering for the lip-similar pixels

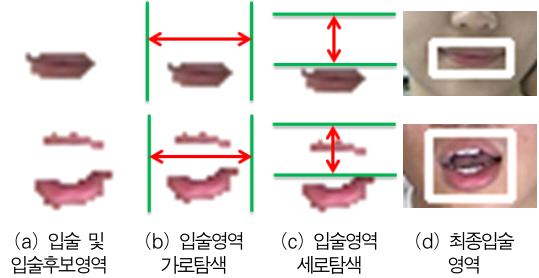


그림 9. 기하학적 특징을 이용한 입술 영역 추출  
Fig. 9. Lip region extraction using geometrical characteristics of the lip area

### 3. 입술영역 결정

II.2에서 생성된 입술 유사 픽셀은 입술영역 픽셀 중 푸른 색 성분을 갖는 픽셀들이 제거된 것으로서 입술과 얼굴을 구성하는 픽셀들을 포함한다. 입술 영역에 비해 얼굴영역은 빛의 반사 등의 원인으로 한 가지 색상을 갖지 못하는 경우가 많기 때문에 보다 정확한 입술 후보 영역 추출을 위하여 K-means 군집화를 수행한다. K-means 군집화 알고리즘은 다음과 같다[18].

- 1) 군집 수 K의 값을 결정하고, 각 군집의 초기 값을 설정한다.
- 2) 모든 데이터를 각 군집의 초기 값과 비교하여 거리가 가장 가까운 군집에 포함시킨다.
- 3) 각 군집에 포함된 데이터로 새로운 군집의 중심 값을 계산한다.
- 4) 군집 중심의 변화 값이 거의 없어질 때까지 2) 와 3)을 반복한다.

K-means 군집화를 위한 특징 벡터는 Lab 색상 모델상의 색상이며 K는 실험을 통해 3으로 하였는데, 군집화가 수행된 후 분류된 3개의 영역 중 평균 a 값이 가장 높은 영역을 입술 후보영역으로 결정하였다. 그림 8은 3-means 군집화 알고리즘을 이용하여 입술 유사픽셀을 세 개의 군집으로 나눈 결과를 보이고 있다.

이렇게 결정된 입술 후보 영역은 윗입술, 아랫입술, 그리고 입술과 유사한 색상으로서 잡음으로 볼 수 있는 영역들을 포함하는데, 본 논문에서는 잡음을 제거하기 위하여 먼저 입술 후보 영역에 대한 모폴로지 열기 연산을 수행한다. 그런 다음 잡음이 제거된 영상에 단 한 개의 영역이 존재한다면 그 영역을 입술영역으로 결정한다. 그러나 입술의 모양은 사람의 표정에 따라 지속적으로 변화하기 때문에 입술영역은 경우에 따라 한 영역 또는 두 개의 영역으로 나타날 수 있으며 경우에 따라서는 잡음의 영향으로 세 개 이상의 입술후보 영역이 존재할 수 있다. 또한 입술이 두 개 이상의 영역으로 분리되어있는 경우 아랫 입술이 윗 입술에 비해 그 영역의 크기가 훨씬 크다. 이에 대한 고려를 위하여, 잡음이 제거된 입술 후보 영역이 두 개 이상 존재하는 경우에는 그 중 가장 큰 영역을 아랫입술영역이라고 가정한 다음 두 번째로 큰 영역을 윗입술 후보영역으로 둔다. 아랫입술의 가로길기와 윗입술 후보 영역의 가로길기를 비교하여 윗입술이 아랫입술의 가로길이보다 작으면서 아랫입술의 위쪽 방향으로 일정거리 이내에 존재할 경우에는 윗입술 후보영역과 아랫입술 영역 모두를 입술 영역으로 결정하고, 그렇지 않으면 아랫입술영역만 입술영역으로 결정한다. 제안된 방법은 입술의 기하학적인 특성을 고려함으로써 입술 영역을 정확하게 추출할 수 있다. 그림 9는 기하학적 특징을 이용하여 입술 영역을 추출하는 과정을 보이고 있는데, 그림 9(a)는 입술과 입술 후보 영역을 보이고 있고, 그림 9(b)는 입술 영역을 기준으로 입술 후보영역이 윗입술인지를 검증하기 위해 입술영역을 가로방향으로 탐색하는 그림을 보이고 있다. 그림 9(c)는 입술영역을 기준으로 세로방향으로 검색하는 그림을 보이고 있으며 그림 9(d)는 최종적으로 추출된 입술 영역을 보인다.





(a) 입력영상 (b) 얼굴영역 (c) 입술영역

그림 10. 실험결과

Fig. 10. Experimental Results

### III. 실험결과

제안한 알고리즘의 타당성을 보이기 위해 Intel Core i7 930 2.80GHz의 CPU와 4GB의 메모리를 갖는 PC에서 다양한 크기의 영상에 대해 테스트를 수행하였다. 테스트에 사용한 영상의 해상도는 최대 720x540, 최소 320x240 이다. 테스트 결과 본 논문에서 제안한 알고리즘은 강건하게 입술 영역을 탐지함을 확인하였다. 그림 10은 제안된 알고리즘으로 추출된 입술영역들의 예를 보이고 있다.

### IV. 결론

본 논문에서는 입력영상으로부터 얼굴을 탐지하고, 탐지된 얼굴영역을 통해 색상군집을 통한 입술탐지 알고리즘을 제안하였다. 제안된 알고리즘은 입력된 영상에서 AdaBoost 알고리즘을 이용하여 얼굴을 추출한 다음 Lab 색상 공간에서 색상 군집화 알고리즘을 이용하여 입술후보영역을 추출한 다음 입술의 기하학적인 특징을 이용하여 입술 영역을 결정하였다. 제안된 알고리즘에 대한 실험을 수행한 결과 입술탐지가 비교적 정확히 이루어짐을 확인하였다. 향후 제안된 입술탐지 알고리즘에 대한 정량적인 평가가 필요하다.

### 참고문헌

- [1] A. Hulbert, T. Poggio, "Synthesizing a color algorithm from examples," *Science, New Series*, vol. 239, pp. 482-485, 1998.
- [2] U. Canzlerm, T. Dziurzyk, "Extraction of non manual features for video based sign language recognition", *Proc. of IAPR Workshop*, pp. 318-321, 2002.
- [3] X. Zhang, R. M. Mersereau, "Lip feature extraction toward an automatic speech reading system", *Proc. of IEEE Int. Conf. Image Processing*, vol. 3, pp. 226-229, 2000.
- [4] M. G. Song, T. T. Pham, J. Y. Kim and S. T. Hwang, "A Study on Lip Detection based on Eye Localization for Visual Speech Recognition in Mobile Environment", *Journal of KSCI*, Vol. 19, No. 4, pp. 478-484, 2009.
- [5] K. H. Lee, "Face Tracking Using Face Feature and Color Information", *Journal of KSCI*, Vol. 18, No. 11, pp. 167-174, 2013.
- [6] P. S. Aleksic, J. J. Williams, Z. Wu, A. K. Katsaggelos, "Audiovisual speech recognition using MPEG-4 compliant visual features", *EURASIP J. Appl. Signal Processing*, pp. 1213-1227, 2002.
- [7] R. Kaucic, B. Dalton, A. Blake, "Real-time lip tracking for audio-visual speech recognition applications", *Lecture Notes in Computer Science*, Vol. 1065, pp. 376-387, 1996.
- [8] S. Werda, W. Mahdi, A. Ben-Hamadou, "Colour and geometric based model for lip localization: application for lip-reading system", *14th International Conference on Image Analysis and Processing*, pp.9-14, 2007.
- [9] C. Bouvier, P.-Y. Coulon, X. Maldague, "Unsupervised lips segmentation based on ROI optimization and parametric model", *IEEE International Conference on Image Processing*, Vol. 4, pp. 301-304, 2007.
- [10] S. W. Chin and K. P. Sengs, "Improved

watershed lips detection and modified  $H^\infty$  tracking system based on lyapunov stability theory”, Int’l Conference on IHMSC, Vol. 2, pp. 355-358, 2009.

[11] L. Wang, X. Wang and J. Xus, “Lip detection and tracking using variance based Haar-like features and Kalman filter”, Int’l Conference on FCST, pp. 608-612, 2010.

[12] Y. H. Huang, B. C. Pan, S. L. Zheng, J. Pan and Y. Tangs, “Lip-reading detection and localization based on two stage ellipse fitting”, Int’l Conference on ICWAPR, Vol. 1, pp. 168-171, 2008.

[13] Y. Pingxian and G. Rongs, “Research on lip detection based on OpenCV”, Int’l Conference on TMEE, pp. 1465-1468, 2011.

[14] E. Skodras and N. Fakotakis, “An unconstrained method for lip detection in color images”, Int’l Conference on ICASSP, pp. 1013-1016, 2011.

[15] R. Duda, P. Hart, D. Stork, “*Pattern Classification*” 2nd Edition, pp. 474-482, A Wiley-Interscience Publication, 2000.

[16] P. Viola and M. Jones, “Rapid object using a boosted cascade of simple features”, IEEE Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. I.511 - I.518, 2001.

[17] A. K. Jain, “*Fundamentals of digital image processing*”, Prentice-Hall, pp. 66-71, 1989.

[18] J. T. Tou, R. C. Gonzalez, “*Pattern recognition principles*”, Addison-Wesley Publishing Company Inc., pp. 75-83, pp. 94-97, 1974.

## 저 자 소 개



### 정 종 면

1992 : 한양대학교 공학사  
 1994 : 한양대학교 공학석사  
 2001 : 한양대학교 공학박사  
 2001 - 2004 : 한국전자통신연구원  
 선임연구원  
 2008 - 2009 : 미국 The Ohio  
 State University Visiting  
 Scholar  
 2004 - 현 재 : 국립목포해양대학교  
 해양컴퓨터공학과 교수  
 관심분야 : 영상처리, 머신 비전,  
 디지털방송,  
 MPEG-2, 4, 7, 21 응용  
 Email : jmjeong@mmu.ac.kr