

# 형태계수의 Mixture Model을 이용한 입술 형태 표현과 입술 경계선 추출

장경식<sup>†</sup>, 이임건<sup>\*\*</sup>

## 요 약

본 논문은 입술의 경계선을 효과적으로 추출하는 방법을 제안하였다. 입술 형태는 PDM(Point Distribution Model)과 주성분 분석법을 이용하여 표현하고 입술 경계선은 GLDM(Gray Level Distribution Model)을 기반으로 표현하였다. 입술 경계선 추출은 모델에 대한 입력영상의 정확도에 대한 목적함수를 최적화하는 문제로 단순화하였으며, 최적화를 위해 다운힐 심플렉스(Down Hill Simplex) 알고리즘을 이용하였다. 탐색과정에서 지역 최소점으로 수렴하는 문제를 해결하기 위하여 입술 형태 모델의 형태계수를 GMM(Gaussian Mixture Model)을 이용하여 표현하였다. 형태계수에 대한 GMM을 이용하여 입술의 대략적인 형태를 찾고, 이때 사용된 mixture 성분을 이용하여 탐색과정에서 입술의 형태를 조정함으로써 지역 최소점에 수렴하여 입술의 정확한 위치를 찾지 못하는 문제점을 해결하였다. 여러 영상을 대상으로 실험하여 좋은 결과를 얻었다.

## Lip Shape Representation and Lip Boundary Detection Using Mixture Model of Shape

Kyung Shik Jang<sup>†</sup>, Imgeun Lee<sup>\*\*</sup>

## ABSTRACT

In this paper, we propose an efficient method for locating human lips. Based on Point Distribution Model and Principle Component Analysis, a lip shape model is built. Lip boundary model is represented based on the concatenated gray level distribution model. We calculate the distribution of shape parameters using Gaussian mixture. The problem to locate lip is simplified as the minimization problem of matching object function. The Down Hill Simplex Algorithm is used for the minimization with Gaussian Mixture for setting initial condition and refining estimate of lip shape parameter, which can refrain iteration from converging to local minima. The experiments have been performed for many images, and show very encouraging result.

**Key words:** Lip Localization(입술 추출), Shape Model(형태모델), Principle Component Analysis(주성분 분석법), Down Hill Simplex(다운힐 심플렉스)

※ 교신저자(Corresponding Author): 장경식, 주소: 부산광역시 부산진구 가야동 산 24번지(614-714), 전화: 051) 890-1988, FAX: 051)890-1619, E-mail: ksjang@deu.ac.kr  
접수일: 2003년 11월 28일, 완료일: 2004년 7월 19일

<sup>†</sup> 정회원, 동의대학교 멀티미디어공학과 조교수

<sup>\*\*</sup> 정회원, 동의대학교 영화영상공학과 전임강사

(E-mail: iglee.deu.ac.kr)

※ 이 논문은 2003년 동의대학교 교내학술연구 지원사업(과제번호 2003AA192)에 의하여 연구되었음.

## 1. 서 론

변화하는 입 모양을 추출하거나 추적하는 연구가 최근에 많은 관심을 끌고 있다. 말할 때의 입 모양 변화를 음성 데이터와 같이 사용하여 음성을 인식하면 인식률 개선을 얻을 수 있음이 알려져 있으며[1, 2], 그래픽 애니메이션(animation) 분야에서 화자의 입 모양을 쉽게 실시간으로 생성할 수 있다[3]. 또한,

변화하는 입 모양은 각 개인을 식별하는 방법으로 이용될 수 있다[4].

입술형태를 추출하는 연구로서 Mirhossein[5]은 입술을 구성하는 에지(edge)의 각도, 방향, 곡률(curvature) 등 입의 형태를 반영하는 조건들을 정하고 이를 만족하는 에지들을 찾음으로써 입술의 경계선을 추출하였다. Oliver[3]는 입술색 모델을 설정하고 입의 형태적 특성을 반영하는 특징들과 시간에 따른 입술의 형태 변화를 HMM(Hidden Markov Model)을 이용하여 모델링하고 추출하였다. Wark[6]는 에지 정보와 색 정보를 이용하여 위 입술과 아래 입술의 경계선을 각각 4차 방정식과 2차 방정식으로 근사화하여 입술의 경계선을 추출하였다. Basu[7]는 얼굴과 입술을 구성하는 화소들의 색과 공간적 배치의 연관성을 이용한 특징을 선정하여 입술 영역을 찾고 칼만 필터(Kalman filter)와 입술 색 모델을 사용하여 입술을 추출하였다. Delmas[8]와 Lievin[9, 10]은 색 정보와 MRF(Markov Random Process)에 기반하여 입술 영역을 찾고 snake를 이용하여 입술의 경계선을 추출하였다. Luetin[12]과 Iain[4]는 능동적 형태모델(Active Shape Model)을 이용하여 입술을 추출하였다. 입술형태를 결정하는 형태계수(shape parameter)와 스케일, 회전 등의 자세계수(pose parameter)를 변화시키며 입술을 구성하는 각 점의 프로파일(profile) 정보를 이용하여 다운힐 심플렉스(Down Hill Simplex) 탐색방법으로 입술을 추출하였다.

색 정보를 이용하는 방법[5,6]은 입술 색의 개인별 차이와 조명의 영향을 많이 받으며, snake를 이용하는 방법[8,9]의 경우 계수설정 및 수렴성(convergence)의 기준 등이 잘못되는 경우 오류를 발생시킬 수 있다. 많은 연구들이 입술이 다물어진 상태에서 입술의 바깥쪽 경계선만을 찾았으며 이는 입을 벌린 경우 치아와 혀로 인하여 입술 안쪽의 경계선을 찾는 것이 어렵기 때문이다. Delmas의 방법[8,10]에서는 입술 바깥쪽 경계선의 추출 성공률은 높았지만 입을 벌린 경우에 안쪽 경계선의 추출 성공률은 낮았다. 이에 반해 능동적 형태모델을 이용한 방법[12,13]은 입술을 벌린 경우에도 안쪽 경계선의 추출에 높은 추출 성공률을 보였다. 그러나 이 방법은 전체 학습 영상을 대표하는 하나의 입술 모델만을 사용하기 때문에 많은 형태 변화를 갖는 입술의 추출에는 부적당

하며 입술을 추출하는 탐색과정에서도 지역 최소점으로 수렴할 가능성이 많다. 이에 따라 다양한 값의 분포를 보이는 형태계수를 GMM(Gaussian Mixture Model)으로 표현하여 능동적 형태 모델에서 형태 표현을 개선하려는 연구와 능동적 형태모델을 기반으로 입술 탐색과정의 초기점 생성 방법을 개선시키려는 연구도 수행되었다[14,15].

본 논문은 입술의 경계선을 효과적으로 추출하는 방법을 제안하였다. 입술의 형태는 PDM을 기반으로 점들의 집합으로 표현하고 주성분 분석법을 이용하여 입술 모델을 구하였다. 능동적 형태 모델을 사용하는 기존 방법에서는 입술 모델에서 사용하는 형태계수를 하나의 가우시안 분포로 가정하지만 이 방법은 입술 형태 변화가 심한 경우에는 부적당하다. 예를 들면 형태계수의 분포가 두개의 피크를 갖는 경우에 이들 사이의 값은 무의미하며, 이로 인해 입술 탐색시 지역 최소점에 수렴할 가능성이 있다. 이 논문에서는 형태계수를 GMM을 이용하여 모델링하고 다운힐 심플렉스를 이용한 입술탐색과정에서 사용하였다. 형태계수에 대한 GMM을 이용하여 입술의 형태를 찾고 이 과정에서 사용된 GMM 성분을 이용하여 탐색공간을 조정함으로써 지역 최소점에 수렴하는 가능성을 줄였다. 여러 입술 영상에 대해 실험한 결과 기존의 방법으로 인식하지 못하는 패턴의 입술을 정확히 추출하였으며 전체적으로 추출률의 향상을 얻었다.

## 2. 입술 형태와 입술 경계선 모델링

### 2.1 입술 형태 모델링

PDM[16]을 이용하여 입술의 형태를 표현하였다. 입술과 주변 피부와의 경계선을 따라 그림 1의 (a)와 같이 바깥쪽과 안쪽 입술 경계선 위에 모두 41개의 기준 점을 위치시켰다. 입술은 식 (1)과 같이 기준 점들의 좌표로 구성되는 벡터로 표현하였다. 학습 영상에 있는 입술의 크기와 방향 등이 다양하기 때문에 동일 크기, 방향을 갖고 무게중심이 동일하도록 이들을 스케일(scale), 평행이동(translation), 회전(rotation)시키는 선형 기하학적 변환(linear geometrical transform)을 통하여 정규화하였다[5,16]. 정규화된 벡터를  $x$ ,  $x$ 의 평균을  $\bar{x}$ 라 하면 입술의 형태 모델은 식 (2)로 표현된다.

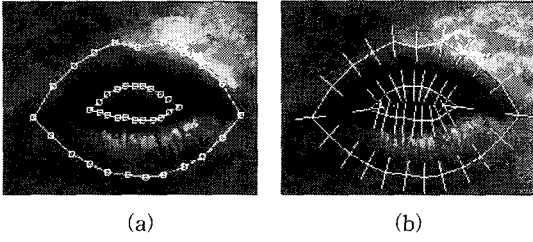


그림 1. 입술 형태와 경계선 모델링

$$X_i = (x_1, y_1, x_2, y_2, \dots, x_{41}, y_{41})^T \quad (1)$$

$$x = \bar{x} + P_s \cdot b_s$$

$$\text{where, } P_s = \begin{bmatrix} P_{s1} & P_{s2} & P_{s3} & \dots & P_{st} \end{bmatrix}, \quad \lambda_{si} \geq \lambda_{s(i+1)}$$

$$b_s = (b_{s1}, b_{s2}, \dots, b_{st})^T \quad (2)$$

$$b_s = P_s^T(x - \bar{x}) \quad (3)$$

여기서  $P_s$ 는  $x$ 의 공분산 행렬의 고유값(eigenvalue)을 크기 순으로 정렬하고 가장 큰 값을 갖는  $t$ 개의 고유값에 대응하는 고유벡터(eigenvector)들로 구성되는 행렬이다.  $b_s$ 는  $t$ 개의 고유벡터들에 대한 가중치들로 구성되는 벡터로서 형태 계수(shape parameter)이며, 고유 벡터는 직교성이 있으므로 형태 계수는 식 (3)을 이용하여 구할 수 있다.  $t$ 는

$$\sum_{i=1}^t \lambda_{si} \geq 0.95 \sum_{i=1}^{40} \lambda_{si} \text{ 식에 의하여 결정하였다.}$$

### 2.2 입술 경계선 모델링

그림 1의 (b)와 같이 학습에 사용된  $i$ 번째 영상의  $j$ 번째 기준점에서 입술 경계선에 직교방향으로 길이  $N$ 의 프로파일(profile)  $g_{ij}$ 를 구하였다. 학습에 사용된 영상들에서  $j$ 번째 기준점의 프로파일들에 대한 평균벡터( $\bar{g}_j$ )와 공분산 행렬( $S_j$ )을 구하였다. 입력 영상에서 입술의 위치와 형태를 추출하는 과정은 입술 형태 모델을 이용하여 생성된 인스턴스(instance)를 구성하는 각 기준 점이 자신이 가진 프로파일과 가장 유사한 화소값 분포를 갖는 위치로 이동함으로써 수행된다. 즉, 인스턴스의  $j$ 번째 기준점이 있는 영상위치에서 길이  $K(K > N)$ 의 프로파일  $D_j$ 를 구하고  $D_j$ 에서 차례로  $N$ 개 값을 취하여  $d_j$ 를 구성한 후 이들 중에서  $j$ 번째 점에서의 평균 프로파

일  $\bar{g}_j$ 와 가장 비슷한 부분을 찾았다. 이를 위하여 식 (4)의 마할라노비스 거리(mahalanobis distance) 척도를 최소로 하는  $d_j$ 의 중간 부분을  $j$ 번째 기준점이 이동할 점으로 정하였다. 이 과정은 인스턴스를 구성하는 모든 점에 대해 수행된다.

$$Distance_j = (d_j - \bar{g}_j)^T S_j^{-1} (d_j - \bar{g}_j) \quad (4)$$

$i$ 번째 학습 영상의  $j$ 번째 기준점에서의 프로파일 벡터를  $g_{ij}$ 라 하면  $i$ 번째 영상의 입술 전체 경계선 모델인 연속 프로파일 벡터는 모든 기준 점들에서의 프로파일들을 연속으로 나열한 것으로 식 (5)와 같이 주어진다. 이 정보는 다운형 심플렉스를 이용한 탐색 과정에서 사용된다.

$$h_i = (g_{i1}, g_{i2}, \dots, g_{i41})^T \quad (5)$$

모든 학습 영상들에 대한 연속 프로파일 벡터들의 평균 벡터  $\bar{h}$ 를 구하고 공분산 행렬을 구한 후 공분산 행렬의 고유 벡터와 고유 값을 구하였다. 입술 전체 경계선 모델은 식 (6)에 의하여 표현된다. 이 식에서  $P_p$ 는 고유값  $\lambda_p$ 을 크기 순으로 정렬하고 가장 큰 값을 갖는  $t$ 개의 고유값에 대응하는 고유벡터를 취하여 구성하였다.

$$h = \bar{h} + P_p \cdot b_p$$

$$\text{where, } P_p = \begin{bmatrix} P_{p1} & P_{p2} & P_{p3} & \dots & P_{pt} \end{bmatrix}, \quad \lambda_{pi} \geq \lambda_{p(i+1)}$$

$$b_p = (b_{p1}, b_{p2}, \dots, b_{pt})^T \quad (6)$$

### 3. Gaussian Mixture를 이용한 형태 계수 모델링

이 논문에서는 GMM을 이용하여 입술 형태 모델의 형태계수  $b_s$ 의 분포를 표현하였다. 이 방법은 주어진 데이터를 몇 개의 군집(cluster)으로 나누고 각 군집에 대한 가우시안 분포인 mixture 성분을 구한 후 이들을 선형적으로 결합하여 하나의 가우시안 분포로 표현하는 방법이다.  $k$ 개의 mixture 성분을 갖는 GMM의 확률 밀도 함수는 식 (7)과 같이 표현된다.

$$G(x | \theta) = \sum_{i=1}^k p(i) g(x | \theta_i) \quad \text{where, } \sum_{i=1}^k p(i) = 1, p(i) \geq 0 \quad (7)$$

이 식에서  $x$ 는 특징 벡터이고  $\theta$ 는  $\{p(i), \theta_i\}_{i=1}^k$ 이며,  $\theta_i$ 는 계수 벡터(parameter vector)로서  $\{\mu_i, \Sigma_i\}_{i=1}^k$ 이다.  $p(i)$ 는 mixture 계수로서 데이터가  $i$  번째 mixture 성분에 의해 발생할 사전확률(prior probability)을 나타낸다.  $g(\cdot)$ 는 다변량 가우시안 분포 함수이다. GMM에 대한 최대 우도계수(maximum likelihood parameter)를 결정하기 위하여 EM(Expectation Maximization) 알고리즘[17]을 사용하였다. 데이터  $D = \{x_1, x_2, \dots, x_n\}$ 에 대한  $k$ 개의 mixture 성분을 갖는 GMM을 얻기 위해 다음과 같은 단계를 반복 수행한다. 이 논문에서는 형태계수  $b_s$ 에서  $b_{s1} \sim b_{sn}$  만을 데이터로 하여 GMM을 구성하였다.

단계 1. 데이터를  $k$  개의 군집으로 나누고 각 군집에 대한 mixture 성분의 계수 ( $\mu_i, \Sigma_i, p(i)$ )를 초기화한다. 이때  $p(i)$ 는  $1/k$  을 사용한다.

단계 2. 각 mixture 성분에 대하여 식 (8)의  $P_i(x)$ 를 구한다.

단계 3. 각 계수의 값을 식 (9)에 따라 구한다.

단계 4. 식 (10)의 우도함수가 수렴하면 중지하고 그렇지 않으면 단계 2로 간다.

$$P_i(x) = \frac{p(i) g(x | \theta_i)}{\sum_{i=1}^k p(i) g(x | \theta_i)} \quad (8)$$

$$p(i) = \frac{\sum_{j=1}^n P_i(x_j)}{n}, \quad \mu_i = \frac{\sum_{j=1}^n P_i(x_j)x_j}{\sum_{j=1}^n P_i(x_j)} \quad (9)$$

$$\Sigma_i = \frac{\sum_{j=1}^n P_i(x_j)(x_j - \mu_i)(x_j - \mu_i)^T}{\sum_{j=1}^n P_i(x_j)}$$

$$L(\theta) = \sum_{j=1}^n \log G(x_j | \theta) \quad (10)$$

#### 4. Gaussian Mixture와 다운힐 심플렉스 탐색을 이용한 입술 추출

입술 경계선을 추출하는 과정은 각 기준점에서의 프로파일과 형태계수의 GMM을 이용하여 점들이 독립적으로 이동하여 입술 위치와 형태를 구하는 과정과 이때 사용된 mixture 성분과 입술 전체 경계선 모델을 이용하여 다운힐 심플렉스 탐색으로 최종 입술을 찾는 과정으로 구성된다.

#### 4.1 Gaussian Mixture를 이용한 입술 위치 및 형태 추출

입술 모델에 의해 생성된 인스턴스를 영상의 중심으로 평행 이동한 후 입술 추출 과정을 수행하였다. 이 과정은 모델을 구성하는 각 점의 프로파일을 이용하여 입술 경계선을 탐색하는 과정과 점들의 전체적인 모양을 허용범위 내로 조정하는 과정의 두 단계로 구성된다. 인스턴스를 구성하는 각 기준 점의 프로파일 정보를 사용하여 2.2절의 방법으로 각 기준 점 주변의 영역들을 탐색하고 각 점은 자신의 프로파일과 가장 비슷한 화소 분포를 갖는 위치로 이동한다. 이 결과 각 점이 독립적으로 이동하여 전체적인 모양이 전형적인 입술과 다른 형태를 갖게 되기 때문에 식 (2)의 입술 모델과 GMM을 이용하여 허용되는 형태 계수를 찾음으로써 입술 형태를 조정하였다. 입력 영상에서 입술을 찾는 과정은 다음과 같다.

$$b = \theta = 0; \quad t = (W/2, H/2); \quad s = 50;$$

while(1) {

$$x = \bar{x} + P_s \cdot b_s; \quad x = T_{(\theta, t, s)}(x);$$

search Y;

$E = |T_{(\theta, t, s)}(\bar{x}) - Y|^2$  를 최소화하는  $(\theta, t, s)$  를 찾는다.

$$y = T_{(\theta, t, s)}^{-1}(Y); \quad y = \frac{y}{y \cdot x};$$

$$b_s = P_s^T(y - \bar{x});$$

$$b_{sa} = (b_{s1} \sim b_{sn}); \quad b_{sb} = (b_{s(n+1)} \sim b_{st});$$

if  $G(b_{sa}) < T_G$

$$i = \arg \max_{i \in \{1, 2, \dots, k\}} g(b_{sa} : \theta_i)$$

$$D_{ma} = \sum_{k=1}^4 \left( \frac{b_{sk} - \mu_i}{\lambda_k} \right)^2$$

$$b_{sa} = (b_{sa} - \mu_i) \cdot \frac{D_{\max}}{D_{ma}} + \mu_i$$

end

$$\text{if } (D_{mb} = \sum_{k=5}^4 \left( \frac{b_{sk}^2}{\lambda_k} \right) \geq D_{\max})$$

$$b_{sb} = b_{sb} \cdot \frac{D_{\max}}{D_{mb}}$$

end

$$bs = (b_{s1}, b_{s2}, \dots, b_{st});$$

```

if( (bs 와 (θ, t, s) 가 수렴)OR
    (반복횟수 > MaxNumLoop))
    break;
end
}

```

$T_G$ 와  $D_{max}$ 는 상수이며,  $W$ 와  $H$ 는 각각 영상의 폭과 높이이다.  $T_{(\theta, t, s)}(x)$ 는 평행 이동량  $t$ , 회전각  $\theta$ , 스케일 비  $s$ 를 사용하여  $x$ 를 선형 기하학적 변환한다. 입술 인스턴스  $x$ 의 각 기준점에서의 프로파일을 이용하여 입술의 경계선에 해당하는 새로운 점을 찾으며,  $Y$ 는 이들 점으로 구성되는 입술 모양이다.

#### 4.2 Gaussian Mixture와 다운힐 심플렉스 탐색을 이용한 입술 추출

입술 추출은 입술 전체 경계선 모델에 대한 목적 함수(cost function)를 구하고 입력 영상에 적용하여 이를 최소화함으로써 수행된다. 목적함수의 최적해는 다운힐 심플렉스를 이용하였다. 입술 경계선 모델을 이용하여 생성된 인스턴스의 기준 점들에서 얻어진 연속 프로파일 벡터  $x_p$ 와 이의 예측치  $\hat{x}_p$ 와의 오차에 대한 MSE(Mean Square Error)  $E^2$ 는 식 (11)으로 주어진다. 그러므로 입력영상에서 입술을 찾는 문제는 목적함수인 식 (11)을 최소화 시키는 문제로 단순화된다.

$$E^2 = (x_p - \bar{x}_p)^T (x_p - \bar{x}_p) - b_p^T b_p \quad (11)$$

최적화해야 할 벡터의 차원보다 하나 많은 초기점을 선정하여 초기 심플렉스를 만든다. 4.1절에서 수행하여 얻은 입술형태에 대한 형태계수, 선형 이동, 회전, 스케일 요소를 각각  $b_s, \bar{t} = (t_x, t_y), \theta, s$ 라고 하면 첫 번째 심플렉스 초기점은 식 (12)와 같다. 나머지 초기 심플렉스 초기점들은 식 (13)에 의해 정해진다. 식 (13)에서  $a^j$ 는 각 벡터의 방향 성분에 대한 가중치 상수들이 대각 성분을 이루는 대각 행렬이며,  $e^j$ 는 단위 벡터이다. 가중치 행렬  $a^j$ 의 대각 성분들은 식 (14)로 주어지며  $\lambda_{sk}$ 는  $b_{st} \sim b_{st}$ 에 대한 GMM의  $j$ 번째 성분에서  $b_{sk}$ 의 분산이며,  $\lambda_{sk}$ 는  $b_{st} \sim b_{st}$ 에서  $b_{sk}$ 의 분산이고  $O_k^0$ 는  $O^0$ 의  $k$ 번째 원소이다. 다운힐 심플렉스의 최소값 수렴은 초기점들 중에서 목적 함수

의 값을 가장 크게 하는 점을 수축, 확장시키는 과정이다.

$$O^0 = (b_{st}, b_{st}, \dots, b_{st}, t_x, t_y, \theta, S)^T \quad (12)$$

$$O^j = O^0 + a^j e^j, \text{ where } 1 \leq j \leq t+4 \quad (13)$$

$$a^j = \begin{cases} 0.5 \times 3\sqrt{\lambda_{kj}} & \text{where } 1 \leq k \leq 4 \\ 0.5 \times 3\sqrt{\lambda_k} & \text{where } 5 \leq k \leq t \\ 0.5 \times O_k^0 & \text{where } t+1 \leq k \leq t+4 \end{cases} \quad (14)$$

탐색과정에서 목적함수 값을 구하는 과정은 다음과 같다. 심플렉스를 구성하는 각 점에서의 형태계수, 선형 이동, 회전, 스케일 요소값을 식 (15)에 적용하여 입술의 인스턴스를 얻은 후 입술 인스턴스를 구성하는 각 기준점들의 위치에서 프로파일을 추출하여 연속 프로파일 벡터를 얻고, 입술 전체 경계선 모델과의 오류를 식 (11)을 이용하여 계산한다. 이 과정에서 4.1절에서 최종 사용한  $j$ 번째 mixture 성분의 형태계수 분포 공간만을 대상으로 하기 위하여 심플렉스를 구성하는 각 점에서의 형태계수 값이 식 (16)을 만족하는 경우에만 오류 계산을 수행하고 그렇지 않은 경우에는 최대 오류값을 할당한다.

$$x = \bar{x} + P_s \cdot b_s, \quad x = T_{(\theta, t, s)}(x) \quad (15)$$

$$\begin{aligned} -3\sqrt{\lambda_{kj}} \leq b_j \leq 3\sqrt{\lambda_{kj}} & \text{ when } 1 \leq j \leq 4 \\ -3\sqrt{\lambda_j} \leq b_j \leq 3\sqrt{\lambda_j} & \text{ when } 5 \leq j \leq t \end{aligned} \quad (16)$$

#### 5. 실험 및 고찰

Tulip 1 데이터베이스[18]에 있는 영상들을 사용하여 실험하였다. 이들은 다양한 조명상태에서 총 12명이 4개 숫자를 두 번씩 발음할 때 입력된 연속적인 명암 영상(gray image)으로서 첫 번째 발음할 때의 영상들을 집합 1, 두 번째 발음할 때의 영상들을 집합 2라고 지칭한다. 이들은 입술 영역만을 포함하고 있으며 다양한 입술 형태를 가진다. 이 논문에서는 학습 영상으로 집합 1의 492개 중에서 200개의 입술 영상을, 실험 영상으로 집합 2의 464개 중 220개의 영상을 무작위로 추출하여 사용하였다. 다운힐 심플렉스의 목적함수의 연산 횟수는 500번으로 제한하였다.

능동적 형태모델을 기반으로 다운힐 심플렉스 탐

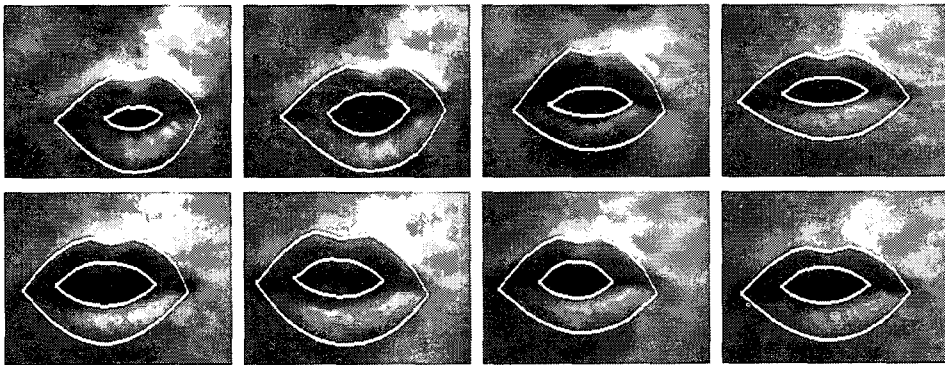
표 1. 실험 결과

	제안 방법	기존 방법
True Positive(%)	82.86	81.36
False Negative(%)	17.14	18.64
True Negative(%)	96.33	95.15
False Positive(%)	3.67	4.85
Error(%)	7.27	8.70

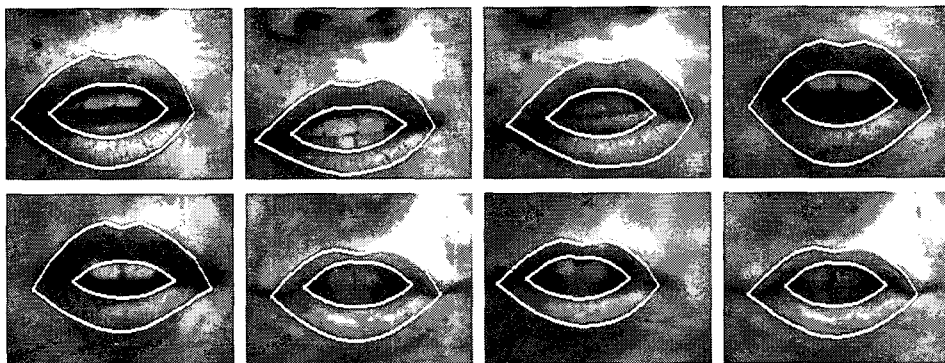
색을 이용하여 입술을 추출하는 기존 방법[12,13]과 본 논문에서 제안한 방법의 결과를 비교하였다. 입술 추출결과를 분석하기 위하여 이전의 연구에서는 직관에 근거한 방법을 사용하였다. 즉, 결과를 눈으로 보고 판정하여 추출 오류 및 성공을 분석하였다. 그러나 이 논문에서는 실험에 사용한 220개의 영상에 대해 수작업으로 기준이 되는 입술 경계선을 정하고



(a) 입술이 다붙어진 경우 추출결과



(b) 입술을 벌린 경우 추출결과



(c) 치아 및 허가 보이는 경우 추출결과

그림 2. 추출 결과 예

실험결과로 얻어진 입술 경계선과의 차이를 비교하는 객관적인 방법을 사용하였다. 표 1은 기존 방법과 제안한 방법에 의한 입술 추출 결과이다. True Positive는 실제 입술에 속하는 화소가 실험 결과 입술에 속하는 화소로 분류된 평균 비율이고 False Negative는 실제 입술에 속하는 화소가 실험 결과 입술에 속하지 않는 화소로 분류된 평균 비율이다. 또한, True Negative는 실제 입술에 속하지 않는 화소가 실험 결과 입술에 속하지 않는 화소로 분류된 평균 비율이고 False Positive는 실제 입술에 속하지 않는 화소가 실험 결과 입술에 속하는 화소로 분류된 평균 비율이다.

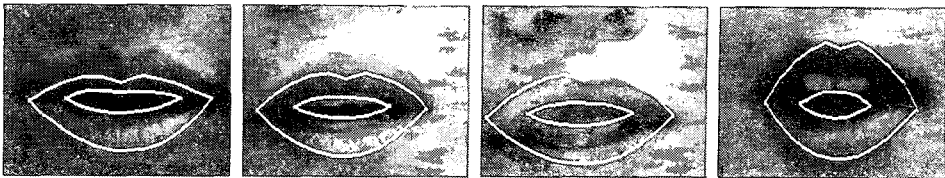
그림 2는 추출 결과의 예이다. 입을 다문 경우, 입을 조금 벌린 경우와 많이 벌린 경우, 입술을 내민 경우, 콧수염이 있는 경우, 치아나 혀가 보이는 경우 등 여러 상황에서 입술 경계선을 정확히 추출하였다. 그림 3은 기존 방법과 본 논문에서 제안한 방법의 결과를 비교한 것이다. (a)는 본 논문에서 제안한 방법으로 추출한 결과이고 (b)는 기존의 방법으로 추출한 결과이다. 본 논문에서 제안한 방법이 보다 실제

형태에 가깝게 추출하였음을 알 수 있다. 이는 기존 방법[13,15]의 경우 맹목적 탐색(blind search)을 수행함으로써 지역 최소점에 빠질 가능성이 많기 때문에 정확한 입술 모양을 찾지 못하는 경우가 많은데 비하여, 형태계수에 대한 GMM을 이용하여 입술의 형태를 찾고 이 과정에서 사용된 GMM 성분을 이용하여 탐색공간을 조정함으로써 얻어진 결과이다.

입술의 바깥쪽 경계선에 대한 추출 오류는 거의 없었으며, 대부분 입술의 안쪽 경계선에서 추출 오류가 발생하였다. 그림 4는 입술 경계선 추출이 잘못된 결과들의 예이다. (a)와 (b)는 위쪽 치아로 인한 경계선 부분을 안쪽 입술의 경계선으로 잘못 추출한 경우이고 (c)는 위 치아와 위 입술의 안쪽 경계선 부분을 아랫입술의 안쪽 경계선으로 잘못 추출한 경우이다. (d)는 혀와 치아의 경계선을 입술 경계선으로 잘못 추출한 결과이다. 대부분의 오류는 입을 벌린 경우 입술사이로 치아나 혀가 보이는 경우에 발생하였으나 본 논문에서 제안한 방법의 경우가 기존의 방법보다 오류의 빈도는 적었다. 이들 오류는 입술, 치아와 혀로 인하여 여러 경계가 나타나기 때문으로 판단된



(a) 제안한 방법



(b) 기존 방법

그림 3. 결과 비교



(a)

(b)

(c)

(d)

그림 4. 추출 오류

다. 그러나 이러한 모든 경우에 오류가 발생하지 않았는데 이는 영상에서 입술의 위치 및 크기, 입술에 나타나는 여러 경계(치아, 혀, 입술 사이의 경계)에서의 명도차이 등의 요인과 입술 전체 경계선 모델의 적절성 여부에 따른 것이라 판단된다. 특히, 이 논문에서는 입술 경계선에서의 프로파일을 이용한 전체 경계선 정보를 하나의 모델로 설정하여 사용하였으나 입을 벌린 경우, 다문 경우, 치아나 혀가 보이는 경우에 입술 전체 경계선 모델을 구성하는 프로파일 이 많은 변화를 보일 수 있다. 이에 따라 입술의 전체 경계선 모델을 설정하는 방법을 개선하여 보다 정확하게 입술 경계선을 추출하는 연구가 필요하다.

## 6. 결 론

본 논문은 입술의 경계선을 효과적으로 추출하는 방법을 제안하였다. 능동적 형태모델을 사용하는 기존 방법에서는 입술 모델에서 사용하는 형태계수를 하나의 가우시안 분포로 가정하지만 이 경우 다운힐 심플렉스 탐색을 수행할 때 지역 최소점에 수렴할 가능성이 있기 때문에 이 논문에서는 형태계수를 GMM으로 모델링하고 다운힐 심플렉스를 이용한 입술 탐색과정에서 사용하였다. 여러 입술 영상에 대해 실험한 결과 기존의 방법으로 추출하지 못하는 패턴의 입술을 정확히 추출하였으나 보다 다양한 형태의 입술을 정확하게 추출하기 위해서는 입술 경계선 모델을 보다 정확한 방법으로 구성하고 이를 이용하는 방법의 연구가 필요하다.

## 참 고 문 헌

- [1] Yang J., R. Stiefelhagen, U. Meier and A. Waibel, "Real-time Face and Facial Feature Tracking and Application," *Proceedings of Auditory-Visual Speech Processing*, pp. 79-84, 1998.
- [2] Kaucic R., A. Blake, "Accurate, Real-Time, Unadorned Lip Tracking," *Proceedings of the 6th International Conf. on Computer Vision*, pp. 370-375, 1998.
- [3] Oliver N., A. Pentland, "LAFTER: Lips and Face Real Time Tracker," *Proceedings of the 1997 Conf. on Computer Vision and Pattern Recognition*, pp. 123-129, 1997.
- [4] Iain Mattheews, Timothy F. Cootes, J. Andrew Banghan, Stephen Cox and Richard Marvey, "Extraction of Visual Features for Lipreading," *IEEE Tans. on Pattern Recognition and Machine Analysis*, Vol. 24, No. 2, pp. 198-213, Feb., 2002.
- [5] Mirhossein A. R., H. Yan and K. M. Lam, "Adaptive Deformable Model for Mouse Boundary Detection," *Optical Engineering*, Vol. 37 No. 3, pp. 869-875, 1998.
- [6] Wark T., Sridharan and V. Chandran, "An Approach to Statistical Lip Modelling for Speaker Identification via Chromatic Feature Extraction," *Proceedings of the 14th International Conf. on Pattern Recognition*, Vol. 1, pp. 123-125, 1998.
- [7] Basu S., N. Oliver and A. Pentlan, "3D Modeling and Tracking of Human Lip Motions," *Proceedings of the 6th International Conf. on Computer Vision*, pp. 337-343, 1998.
- [8] Delmas P., Y. Coulon and V. Fristot, "Automatic Snakes for Robust Lip Boundaries Extraction," *IEEE International Conf. on Acoustics, Speech and Signal Processing*, Vol. 6, pp. 3069-3072, 1999.
- [9] Lievin M., F. Luthon, "Unsupervised Lip Segmentation under Natural Conditions," *IEEE International Conf. on Acoustics, Speech and Signal Processing*, Vol. 6, pp. 3065-3068, 1999.
- [10] Lievin M., P. Delmas, Y. Coulon, F. Luthon and V. Fristot, "Automatic Lip Tracking : Bayesian Segmentation and Active Contours in a Cooperative Scheme," *IEEE Conf. on Multimedia, Computing and System*, pp. 691-696, 1999.
- [11] M. Leivin, F. Luthon, "Lip Features Automatic Detection," *IEEE International Conference on Image Processing*, pp. 168-172, 1998.
- [12] Luettin J., N. A. Thacker and S. W. Beet,



“Locating and Tracking Facial Speech Features,” *Proceedings of The International Conf. on Pattern Recognition*, pp. 652-656, 1996.

- [13] J. Luettin, and N.A. Thacker, “Speechreading using probabilistic models,” *Computer Vision and Image Understanding*, Vol. 65, pp. 163-178, 1997.
- [14] 장경식, 이임건, “Active Shape 모델과 Gaussian Mixture 모델을 이용한 입술 인식”, 정보과학회 논문지, Vol 30, No. 5·6, pp. 454-460, 2003.
- [15] 이임건, 장경식, “입술의 형태 모델과 Down Hill 탐색 방법을 이용한 입술 인식”, 멀티미디어학회 논문지, Vol. 6, No. 6, pp. 968-976, 2003.
- [16] M. B. Stegmann, R. Fisker, “On Properties of Active Shape Models,” *Informatics and Mathematical Modelling*, Technical University of Denmark, 2000.
- [17] Morton Nadler and Eric P. Smith, *Pattern Recognition Engineering*, A Wiley-Interscience Publication, 1993.
- [18] Movellan J. R., “Visual Speech Recognition with Stochastic Networks,” *Advances in*

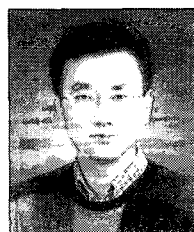
*Neural Information Processing System*. Vol. 7, MIT Press Cambridge, 1995.



**장 경 식**

1989년 연세대학교 전자공학과 졸업(학사)  
 1991년 연세대학교 대학원 전자공학과 졸업(석사)  
 1996년 연세대학교 대학원 전자공학과 졸업(박사)  
 1994년~1998년 2월 대우전자 전

략기술 제 1 연구소 선임연구원  
 1998년~현재 동의대학교 멀티미디어공학과 조교수  
 관심분야: 컴퓨터 비전, 인공지능, Digital Contents Watermarking 등



**이 임 건**

1991년 연세대학교 전자공학과 졸업(학사)  
 1993년 연세대학교 대학원 전자공학과 졸업(공학석사)  
 1998년 연세대학교 대학원 전자공학과 졸업(공학박사)  
 1998년~2001년 대우전자 디지털

TV 연구소 선임연구원  
 2002년~현재 동의대학교 영화영상공학과 전임강사  
 관심분야: 영상처리, 컴퓨터 비전 등