

Active Shape 모델과 Gaussian Mixture 모델을 이용한 입술 인식 (Lip Recognition Using Active Shape Model and Gaussian Mixture Model)

장 경 식 [†] 이 임 건 ^{**}
(Kyung-Shik Jang) (Im-geun Lee)

요 약 이 논문은 입술의 형태를 효과적으로 인식하는 방법을 제안하였다. 입술은 PDM(Point Distribution Model)을 기반으로 점들의 집합으로 표현하였다. 주성분 분석법을 적용하여 입술 모델을 구하고 모델에서 사용하는 형태계수의 분포를 GMM(Gaussian Mixture Model)을 이용하여 구하였다. 이 과정에서 계수를 정하기 위하여 EM(Expectation Maximization) 알고리즘을 사용하였다. 입술 경계선 모델은 입술을 구성하는 각 점과 주변 영역에서의 화소값 변화를 이용하여 구성하였으며 입술 탐색시 사용되었다. 여러 영상을 대상으로 실험한 결과 좋은 결과를 얻었다.

키워드 : 입술 인식, 주성분 분석법, EM 알고리즘

Abstract In this paper, we propose an efficient method for recognizing human lips. Based on Point Distribution Model, a lip shape is represented as a set of points. We calculate a lip model and the distribution of shape parameters using Principle Component Analysis and Gaussian mixture, respectively. The Expectation Maximization algorithm is used to determine the maximum likelihood parameter of Gaussian mixture. The lip contour model is derived by using the gray value changes at each point and in regions around the point and used to search the lip shape in a image. The experiments have been performed for many images, and show very encouraging result.

Key words : lip recognition, principle component analysis, expectation maximization algorithm

1. 서 론

최근에 시간에 따른 눈동자나 입술의 움직임에 대한 정보를 여러 분야에 응용하려는 연구가 활발히 진행되고 있으며, 이 중에서도 입 모양을 인식하거나 추적하는 연구는 많은 관심을 끌고 있다[1,2]. 입술의 정확한 형태 및 변화하는 모양 추출 결과는 그래픽 애니메이션(animation) 분야에서 화자의 입 모양을 쉽게 실시간으로 생성할 수 있고 음성 데이터와 입 모양 변화를 같이 사용하여 음성을 인식하는 경우에 인식을 개선할 수 있음이 알려져 있기 때문에 이에 대한 연구가 많은 관심을 끌고 있다[3, 4]. 또한, 말할 때 변화하는 입 모

양을 각 개인을 식별하는 방법으로 사용하는 연구가 진행되고 있다[5].

Mlrhosseinl[1]은 입술을 구성하는 에지(edge)의 각도, 방향, 곡률(curvature) 등 입의 형태를 반영하는 조건들을 만족하는 에지들을 찾음으로써 입술의 경계선을 추출하였다. Oliver[2]는 색 정보에 대한 입술 모델을 설정하고 입의 형태적 특성을 반영하는 특징들과 시간에 따른 입술의 형태 변화를 HMM(Hidden Markov Model)을 이용하여 모델링하고 인식하였다. Yang[3]은 입술 색에 대한 적응적 Gaussian mixture model을 구성하고 입력에 따라 모델을 변화하여 입술 영역을 추출하였다. Wark[6]은 에지 정보와 색 정보를 이용하여 위 입술과 아래 입술의 경계선을 각각 4차 방정식과 2차 방정식으로 근사화하여 입술의 경계선을 추출하였다.

Zhang[7]은 입 부분에서 찾은 에지의 강도와 위, 아래 입술에서 붉은색 성분 차이 및 분산을 이용하여 입

[†] 정 회 원 : 동의대학교 멀티미디어공학과 교수
ksjang@dongeui.ac.kr

^{**} 정 회 원 : 동의대학교 영화영상공학과 교수
iglee@dongeui.ac.kr

논문접수 : 2002년 9월 27일

심사완료 : 2003년 2월 13일

술을 인식하였다. Basu[8]는 얼굴과 입술을 구성하는 각 화소들의 색 정보와 공간적 배치의 연관성을 이용한 특징을 선정하여 입술 영역을 찾고 칼만 필터(Kalman filter)와 입술 색 모델을 사용하여 입술을 추출하였다. Delmas[9]와 Lievin[10, 11]은 색 정보와 MRF(Markov Random Process)에 기반한 Bayesian segmentation을 이용하여 입술 영역을 찾고 snake를 이용하여 입술의 경계선을 추출하였다. Luetttin[12, 13]과 Cootes[5]는 ASM(Active Shape Model)을 이용하여 입술을 인식하였다. PDM(Point Distribution Model) 방법을 사용하여 입술을 표현하고 PCA(Principle Component Analysis)를 적용하여 입술의 모델을 구성하였으며 입술을 구성하는 각 점에서 profile을 추출하고 모델링하였다. 각 점의 인근 영역을 탐색하여 형태계수(shape parameter)와 자세계수(pose parameter)를 변화시키며 입술을 인식하였으며 좋은 결과를 얻었다.

색 정보를 이용하여 입술을 찾는 방법의 경우에 입술 색의 개인별 차이, 조명의 영향을 많이 받으며, 예지 및 투영을 이용하는 방법에서는 입술과 주변의 피부와의 경계가 뚜렷하지 않고 다양한 예지가 존재하기 때문에 입 모양의 정확한 경계선을 찾기 어렵다. snake를 이용하는 방법[9, 10]의 경우 초기화와 계수설정 및 수렴성(convergence)의 기준 등이 잘못되는 경우 오류를 발생시킬 수 있다. 또한, 많은 연구들이 입술이 다물어진 상태에서 입술의 바깥쪽 경계선을 찾았으며 이는 입을 벌린 경우 입 안쪽의 이와 혀로 인하여 입술 안쪽의 경계선을 찾는 것이 어렵기 때문이다. Delams의 방법[9]에서는 바깥쪽 경계선의 인식률은 높았지만 입을 벌린 경우에 이와 혀로 인하여 안쪽 경계선의 인식률은 낮았다. 이에 반해 Luetttin의 방법[12]은 입술의 바깥쪽은 물론 입술을 벌린 경우에도 안쪽 경계선의 추출에 높은 인식률을 보였다. 그러나 이 방법은 형태 계수의 분포를 하나의 가우시안 분포로 가정하기 때문에 입술의 평균 모양에서 많이 벗어난 형태를 갖는 입술의 추출에는 실패하는 경우가 많았다.

본 논문은 입술의 형태를 효과적으로 추출하는 방법을 제안하였다. 입술의 형태는 PDM을 기반으로 점들의 집합으로 표현하였고 학습 영상을 통계적으로 분석하는 주성분 분석법을 이용하여 입술 모델을 구하였다. 기존의 방법에서는 입술 모델에서 사용하는 형태계수를 하나의 Gaussian 분포로 가정하지만 입술 형태 변화가 심한 경우에는 표현하지 못하기 때문에 이 논문에서는 GMM(Gaussian Mixture Model)을 이용하여 모델링하였다. 이 과정에서 사용되는 계수들을 정하기 위하여

EM(Expectation Maximization) 알고리즘을 사용하였으며 가우시안의 개수를 효과적으로 정하기 위하여 k-means 군집 알고리즘과 MDL(Minimum Description Length) 방법을 사용하였다. 여러 입술 영상에 대해 실험한 결과 기존의 방법으로 인식하지 못하는 패턴의 입술을 정확히 인식하였으며 전체적으로 인식률의 향상을 얻었다.

2. 입술 분석

2.1 입술 형태 모델링

PDM[14]을 이용하여 입술의 형태를 표현하였다. 입술과 주변 피부와의 경계선을 따라 그림 1의 (a)와 같이 바깥쪽 입술선 위에 21개, 안쪽 입술선 위에 20개의 점을 위치시켰다. 이때 각 점은 입술의 정해진 위치에 있게 하였다.

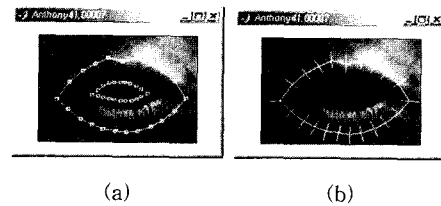


그림 1 입술 모델링과 특징 추출

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

입술은 식 (1)과 같이 41개 점들의 좌표로 구성되는 벡터로 표현하였다. 학습 영상에 있는 입술의 크기 및 형태가 다양하기 때문에 이들을 대표하는 하나의 모델을 구성하기 위하여 동일 크기, 방향을 갖고 무게중심의 좌표가 동일하도록 이들을 스케일(scale), 평행이동(translation), 회전(rotation)시키는 선형 기하학적 변환(linear geometrical transform)을 통하여 정규화하였다 [5, 14]. 정규화된 벡터들의 평균 벡터를 구하고 식 (2)에 따라 공분산 행렬(covariance matrix)을 구한 후 공분산 행렬의 고유 벡터(eigen vector)와 고유 값(eigen value)을 구하였다. 이 식에서 P_i 는 고유 벡터로서 $P_i^T P_i = E$ 를 만족하는 직교행렬(orthogonal matrix)이고 λ_i 는 고유값이다. 입술 모델은 식 (3)에 의하여 표현된다. 이 식에서 P 는 고유값을 크기 순으로 정렬하고 가장 큰 값을 갖는 t 개의 고유값에 대응하는 고유벡터를 취하여 구성하였다. t 는 $\sum_{i=1}^t \lambda_i \geq V \sum_{i=1}^b \lambda_i$ 식에 의하여 결정하며 V 는 상수이다. b 는 고유 벡터들에 대한 가중치

들로 구성되는 벡터로서 형태 계수(shape parameter)이며, 식 (4)에 의하여 결정된다. 그림 2는 모델에 의하여 생성한 입술 모양으로 (b)는 평균 입술 모양이고 (a), (c)는 각각 $-2\sqrt{\lambda_1}$, $2\sqrt{\lambda_1}$ 만큼 변형된 입술 모양이다.

$$\sum P_i = \lambda_i P_i \quad \text{where,} \quad (2)$$

$$\Sigma = \frac{1}{m} \sum_{i=1}^m (X_i - \bar{X})(X_i - \bar{X})^T$$

$$X = \bar{X} + P \cdot b \quad \text{where,} \quad (3)$$

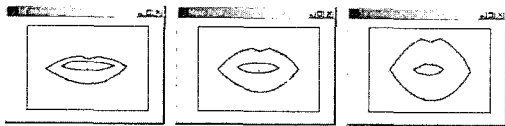
$$P = \begin{bmatrix} P_1 & P_2 & P_3 & \dots & P_i \end{bmatrix},$$

$$\lambda_i \geq \lambda_{i+1}$$

$$b = (b_1, b_2, \dots, b)^T$$

$$b = P^T (X - \bar{X}) \quad \text{where,} \quad (4)$$

$$-3\sqrt{\lambda_i} \leq b_i \leq 3\sqrt{\lambda_i}$$



(a) (b) (c)
그림 2 평균 입술과 변형된 입술

2.2 화소값 분포를 이용한 입술선 특징 선정

각 점에 인접한 영역의 화소값 분포를 이용하여 입술 경계선에 대한 특징을 설정하였다. 그림 1 의 (b)와 같이 학습 영상에서 입술을 구성하는 각 점을 지나며 입술 경계선에 수직인 방향으로 점의 양쪽으로 n 개씩과 자신을 합하여 (2n+1)개 점을 선택하고 그 점들의 화소값들로 구성되는 벡터 g를 구하였다. 화소값을 직접 사용하는 경우에 조명 등의 밝기 변화에 따른 영향을 많이 받기 때문에 이를 최소화하기 위하여 식 (5)에 따라 정규화하고 특징 벡터(g')를 구하였다. 학습에 사용된 입술의 동일 위치에 있는 점에서의 특징들에 대한 평균 벡터(\bar{g})와 공분산 행렬(Sg)을 구하였으며, 이는 입술 탐색과정에서 사용된다.

$$g = (g_1, g_2, \dots, g_{(2n+1)})^T \quad (5)$$

$$g' = \frac{g'}{\sum g_i} \quad \text{where,}$$

$$g'_i = g_{i+2} - g_i \quad (i=1, \dots, 2n-1)$$

$$dg' = g' - \bar{g}'$$

$$S_g = \frac{1}{m} \sum_{i=1}^m dg'_i dg_i'^T \quad \text{where,} \quad (6)$$

m은 학습 영상의 수

3. Gaussian mixture를 이용한 형태 계수 모델 설정

3.1 Gaussian mixture 구성

입술 모델을 표현하는 식 (3)을 이용하여 입술 모양을 생성하려면 학습 영상에서 얻어진 형태 계수 b의 분포 내에서 b값을 선택하여야 한다. 기존의 방법에서는 b의 분포를 하나의 가우시안 분포로 가정하지만 이 논문에서는 Gaussian mixture를 이용하여 b의 분포를 표현하였다. 이 방법은 주어진 데이터를 몇 개의 군집(cluster)으로 나누고 각 군집에 대한 가우시안 분포인 mixture component를 구한 후 이들을 선형적으로 결합하여 하나의 가우시안 분포로 표현하는 방법이다. k개의 mixture component를 갖는 Gaussian mixture의 확률 밀도 함수는 식 (7)과 같이 표현된다.

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

$$\sum_{i=1}^k p(i) = 1, \quad p(i) \geq 0$$

이 식에서 x는 특징 벡터이고 θ 는 $\{p(i), \theta_i\}_{i=1}^k$ 이며, θ_i 는 계수 벡터(parameter vector)로서 $\{\mu_i, \Sigma_i\}_{i=1}^k$ 이다. p(i)는 mixture 계수로서 데이터가 i번째 mixture component에 의해 발생할 사전확률(prior probability)을 나타낸다. g는 다변량 가우시안 분포 함수이다.

Gaussian mixture에 대한 최대 우도 계수(maximum likelihood parameter)를 결정하기 위하여 EM(Expectation Maximization) 알고리즘[15]을 사용하였다. EM 알고리즘은 주어진 계수들을 사용하여 평균 우도 함수를 구하는 E-step 단계와 데이터들에 대한 최대 우도 계수의 예측 값을 계산하는 M-step 단계로 구성된다. 데이터 D = $\{x_1, x_2, \dots, x_n\}$ 에 대한 k개의 mixture component를 갖는 Gaussian mixture를 얻기 위해 다음과 같은 단계를 반복 수행한다.

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

단계 2. (E-step) 각 mixture component에 대하여 식 (8)의 Pi(x)를 구한다.

단계 3. (M-step) 각 계수의 값을 식 (9)에 따라 구한다.

단계 4. 식 (10)의 우도 함수 값이 수렴하면 중지하고 그렇지 않으면 단계 2로 간다.

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

$$b(i) = \frac{\sum_{j=1}^n P_i(x_j)}{n},$$

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

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

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

3.2 모델 선택

Gaussian mixture를 사용하는 경우에 우선적으로 고려할 사항은 mixture component의 수를 결정하는 모델 선택 문제이다. 많은 mixture component를 사용하면 분포함수는 정확하게 구할 수 있지만 효율성 문제 때문에 mixture component의 수를 결정하는 것이 중요하다. mixture component 수를 결정하기 위하여 이 논문에서는 MDL(Minimum Description Length) 방법[15]을 사용하였다. Gaussian mixture에 대한 Description Length는 식 (11)과 같이 표현된다.

$$DL = NPn - \sum_{j=1}^n \log_2 G(b_j | \theta) + kF\left(\frac{n^2 + 3n + 2}{2}\right) \quad (11)$$

이 식에서 N은 데이터 수, P는 계수의 정확도이다. 이 논문에서는 b에서 b1~b4 계수만을 데이터로 하여 k-means 군집 알고리즘에 의하여 한 군집부터 15개의 군집으로 나누었다. 각각의 경우에서 Gaussian mixture와 DL 값을 구하고 가장 작은 값을 갖는 경우인 mixture component가 4개인 Gaussian mixture를 구성하였다.

4. 입술 인식

4.1 두 입술 사이의 경계선 설정

능동적 형태 모델을 이용하여 탐색하는 경우에 초기 위치의 선정에 따라 잘못된 결과가 나올 수 있다. 이 논문에서는 입술을 인식하는 첫 단계로서 두 입술 사이의 경계선을 찾고 그 위치를 중심으로 입술을 탐색하였다. 영상의 각 열에서 식 (12)를 이용하여 가장 어두운 화소들을 찾고 식 (13)을 이용하여 입술 영상의 높이에 해당하는 크기를 갖는 벡터인 L(y)를 구하면 가장 큰 값을 갖는 위치에서 두 입술 사이의 경계선을 얻는다. 이 식들에서 I(x, y)는 (x, y) 위치에서의 화소값을 나타내고 K₁, K₂는 상수이며, H_{Lip}, W_{Lip}은 각각 영상의

높이와 폭을 나타낸다. 그림 3의 (a)는 L(y)를 나타낸 것이고 (b)는 두 입술 사이의 경계선을 나타낸 것이다.

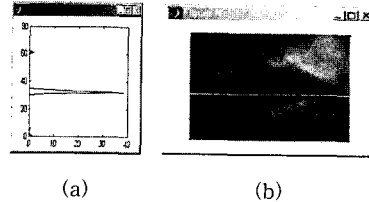


그림 3 입술선 검출

$$M(x) = \arg \min_{y \in H_{Lip}} I(x, y) \quad (12)$$

$$L(y) = \sum_{x \in W_{Lip}} \frac{\left(K_1 - \cosh \left(\frac{M(x) - \frac{H_{Lip}}{2}}{\frac{H_{Lip}}{K_2}} \right) \right)}{K_1}$$

$$Y_{LipCenterLine} = \arg \max_{y \in H_{Lip}} L(y) \quad (13)$$

4.2 입술 인식

두 입술 사이의 경계선 위치가 정해지면 입술 모델을 그 위치로 평행 이동한 후 입술을 찾는 과정을 수행하였다. 이 과정은 모델을 구성하는 각 점의 특징을 이용하여 입술 경계선을 탐색하는 과정과 점들의 전체적인 모양을 허용범위 내로 조정하는 과정의 두 단계로 구성된다. 입술 모델을 구성하는 각 점의 특징을 사용하여 점 주변의 정해진 영역들을 탐색하고 각 점은 자신이 갖고 있는 입술선 특징과 가장 비슷한 특징을 갖는 점으로 이동한다. 이 결과 각 점이 독립적으로 이동하기 때문에 전체적인 모양이 입술과 다른 형태를 갖게 되기 때문에 식 (3)의 입술 모델과 Gaussian mixture를 이용하여 적당한 형태계수를 찾음으로써 허용되는 형태로 조정하였다. 입력 영상에서 입술을 찾는 과정은 다음과 같다.

1. b는 0을, (θ, t, s) 는 초기값을 설정한다. t는 (t_x, t_y) 로서 평행 이동량, θ 는 회전각, s는 스케일 비를 나타낸다. 이 논문에서는 t_y 는 이전 단계에서 구한 두 입술 사이의 경계선 y 좌표를, t_x 는 영상 폭의 1/2을 사용하였다. 또한 θ 는 0, s는 임의의 값을 사용하였다.
2. 입술 모델을 표현하는 $X = \bar{X} + P \cdot b$ 식에 의하여 입술의 인스턴스를 생성하고 $T_{(\theta, t, s)}(X)$ 식에 의하여

스케일, 평행이동, 회전시킨다. $T_{(\theta, t, s)}(X)$ 는 평행 이동량 t , 회전각 θ , 스케일 비 s 를 사용하여 X 를 선형 기하학적 변환한다.

- 인스턴스 각 점에서의 특징을 이용하여 입술의 경계선에 해당하는 새로운 점을 찾는다. 이들 점으로 구성되는 모양을 Y 로 한다.
- 식 (14)를 최소화하는 (θ, t, s) 를 찾고 이를 새로운 (θ, t, s) 로 정한다. $y = T_{(\theta, t, s)}^{-1}(Y)$ 식을 이용해 역변환 후 $y = \frac{y}{y \cdot X}$ 식에 의해 정규화한다.
- $b = P^T(y - \bar{X})$ 식을 이용해 b 를 구하고 $b_a = b_1 \sim b_4$, $b_b = b_5 \sim b_6$ 의 두 부분으로 나눈다. b_a 를 Gaussian mixture $G(x)$ 에 적용하여 $G(b_a) > P_1$ 을 만족하는지 조사한다. 만족하지 않으면 식 (15)에 의하여 b_a 값을 조정한다. 식 (16)을 이용해 b_b 에 대한 조건을 조사하고 만족하지 않으면 b_b 값을 조정한다. 이때 P_1 와 D_{max} 는 상수이다.
- b 와 (θ, t, s) 가 수렴하면 중지한다. 그렇지 않으면 단계 2로 간다.

$$E = |T_{(\theta, t, s)}(\bar{X}) - Y|^2 \tag{14}$$

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

$$D_{ma} = (b_a - \mu_i)^T \Sigma_i^{-1} (b_a - \mu_i) \tag{15}$$

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

$$D_{mb} = \sum_{i=5}^6 \left(\frac{b_i^2}{\lambda_i} \right) \leq D_{max}, \quad b_b = b_b \cdot \frac{D_{max}}{D_{mb}} \tag{16}$$

입력 영상에서 입술의 인스턴스를 구성하는 각 점에 인접한 점들에서 특징을 추출하고 모델의 대응하는 점의 특징과 비교하여 가장 비슷한 특징을 갖는 위치로 각 점이 독립적으로 이동한다. 현재 입술의 인스턴스를 구성하는 각 점의 위치에서 화소값의 분포 G_s 를 구하였으며, 이 과정은 입술 모델을 구성하는 각 점의 특징을 구하는 과정과 같다. 이때 G_s 는 현재 점을 중심으로 양쪽에서 $m(m > n)$ 개의 화소 값을 취하여 구성하였다. 이후 G_s 에서 차례로 $(2n+1)$ 개 값을 취하여 g_s 를 구성하고 이들 g_s 중에서 \bar{g} 와 가장 비슷한 부분을 찾았다. 이를 위하여 식 (17)의 마할라노비스 거리(mahalanobis

distance) 척도를 최소로 하는 g_s 를 선택하고 g_s 의 중간 부분에 해당하는 점을 이동할 점으로 정하였다.

$$f(g_s) = (g_s - \bar{g})^T S_g^{-1} (g_s - \bar{g}) \tag{17}$$

5. 실험 및 고찰

실험을 위하여 Tulip 1 데이터베이스[16]에 있는 입술 영상들을 사용하였다. 이들은 다양한 조명상태에서 총 12명(남자 9명과 여자 3명)이 4개의 숫자를 두 번씩 발음할 때 입력된 연속적인 명암 영상(grey image)으로서 첫 번째 발음할 때의 영상들을 집합 1, 두 번째 발음할 때의 영상들을 집합 2 라고 지칭한다. 이들은 입술 영역만을 포함하고 있으며 다양한 입술 형태를 가지고 100*75 크기를 갖는다. 이 논문에서는 이들 영상에서 학습 영상으로 집합 1에서 200개의 입술 영상을 사용하였으며 실험 영상으로 집합 2에서 220개의 영상을 사용하였다. 펜티엄(Pentium) 4 PC(1.6 GHz)에서 실험하였으며 matlab으로 구현하였다.

표 1 실험 결과

	인식 성공	인식 실패						전체
		윗 입술		아래 입술		입술 폭		
		바깥쪽 경계선	안쪽 경계선 (1)	바깥쪽 경계선 (2)	안쪽 경계선 (3)	바깥쪽 폭 (4)	안쪽 폭 (5)	
영상 개수(개)	204	0	3	2	2	5	4	220
비율(%)	92.7	0	1.4	0.9	0.9	2.3	1.8	100

표 1은 실험 결과이다. 모든 경우에 인식결과의 위치가 입술의 실제 위치와 전혀 다른 결과는 없었으며, 오류인 경우는 실제 입술 경계선에서 조금씩 떨어지거나 전체적인 입술 모양이 다르게 인식된 경우가 발생하였다. 한 입술을 인식하는 시간은 평균적으로 8.2 초가 소요되었다. 이는 matlab으로 구현했을 때의 인식속도로써 인식속도 측면에서 많은 개선이 가능하다.

실험결과를 분석하기 위하여 인식한 입술의 경계선이 실제 위치와 일정 거리 이상 떨어진 경우를 오인식으로 판정하였다. 이때 거리의 기준으로 윗입술 높이의 1/2을 사용하였다. 윗입술의 바깥쪽 경계선에 대한 인식 오류는 없었으며, 아래 입술의 바깥쪽 경계선 인식에서 오류가 발생하였다. 이는 일반적으로 아래 입술과 인접 피부와의 경계가 윗입술과 인접 피부와의 경계보다 뚜렷하지 않기 때문이다. 오히려 입술의 양 끝점을 잘못 인식하여 인식 결과가 실제 입술보다 옆으로 긴 형태의 오

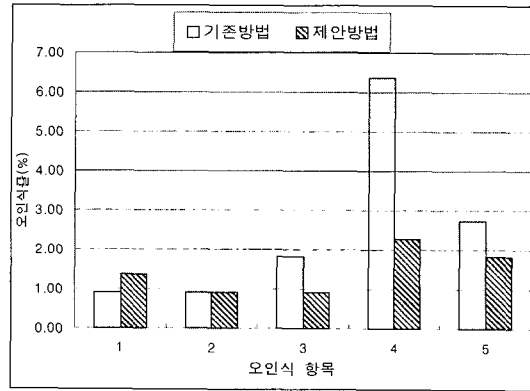
류가 많았다. 이는 입술 양 끝점에서 인접 피부와의 경계가 매우 불확실하기 때문이다. 윗입술의 안쪽 경계선과 아래 입술의 안쪽 입술 경계선에 대한 오류는 비슷하였다. 이는 입을 벌린 경우 입술 사이에 보이는 이로 인하여 여러 경계가 나타나기 때문이다. 그림 4는 인식 결과의 예이다. 입을 다문 경우, 입을 조금 벌린 경우와 많이 벌린 경우, 입술을 내민 경우, 콧수염이 있는 경우 등 여러 상황에서 입술을 정확히 인식하였다. 그림 5는 Luettin의 방법[13]과 본 논문에서 제안한 방법의 결과를 비교한 것이다. (a)는 본 논문에서 제안한 방법으로 인식한 결과이고 (b)는 기존의 방법으로 인식한 결과이다. 본 논문에서 제안한 방법이 보다 실제 형태에 가깝게 인식하였음을 알 수 있다.

그림 6은 오인식된 결과들이다. (a)는 아래 입술을 입 안쪽으로 넣은 상태로 아래 입술의 바깥쪽 경계선을 인식하지 못한 경우이고 (b)는 입술의 양 끝점을 인식하지 못한 경우이다. (c)는 벌린 입 사이로 보이는 이로 인하여 아래쪽 입술의 안쪽 경계선을 잘못 인식한 결과이다. 표 2는 Luettin의 방법을 구현하여 수행한 결과를 본 논문에서 제안한 방법과 오인식률을 비교한 것이다. 이 그래프에서 x축의 숫자는 표 1의 인식 실패의 각 항목에 해당된다. 전체적으로 인식률에서 많은 개선이 있음을 알 수 있다.

6. 결론

본 논문에서는 입술의 형태를 효과적으로 추출하는 방법을 제안하였다. Gaussian mixture model을 이용하여 입술 모델의 형태계수 분포를 모델링하였다. 이 과정에서 mixture component들의 계수를 정하기 위하여 EM 알고리즘을 사용하였으며 k-means 군집 알고리즘

표 2 오인식률 비교



과 MDL 방법을 사용하여 mixture component의 수를 정하였다. 여러 입술 영상에 대해 실험한 결과 기존의 방법으로 인식하지 못하는 패턴의 입술을 정확히 인식하였으며 전체적으로 인식률의 향상을 얻었다. 입술을 구성하는 점들의 특징을 이용하여 탐색하는 방법을 개선하면 보다 높은 인식률을 얻을 것으로 생각된다.

참고 문헌

[1] Mirhosseini A. R., H. Yan and K. M. Lam, "Adaptive Deformable Model for Mouse Boundary Detection", *Optical Engineering*, Vol. 37 No. 3(1998), pp. 869-875.
 [2] Oliver N., A. Pentland, "LAFTER: Lips and Face Real Time Tracker", *Proceedings of the 1997 Conf. on Computer Vision and Pattern Recognition*, (1997), pp. 123-129.

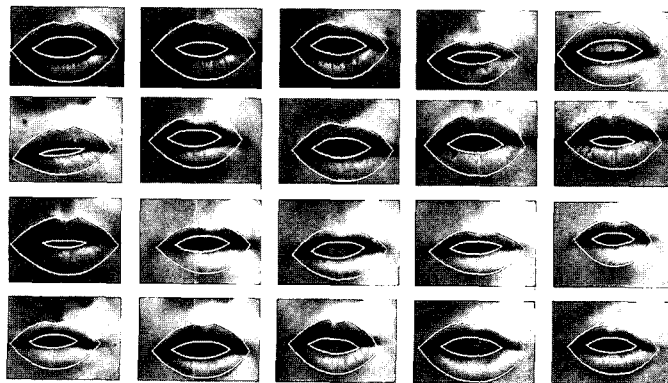


그림 4 인식 결과

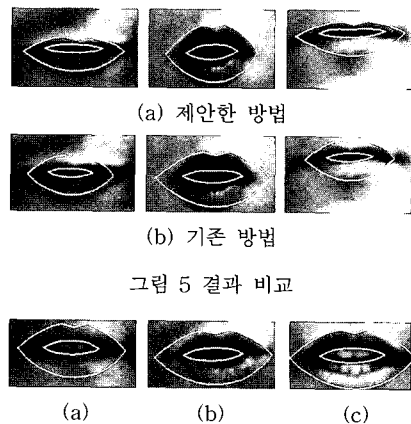


그림 6 오인식 결과

- [3] Yang J., R. Stiefelwagen, U. Meier and A. Waibel, "Real-time Face and Facial Feature Tracking and Application", *Proceedings of Auditory-Visual Speech Processing*, pp. 79-84, 1998.
- [4] Kaucic R., A. Blake, "Accurate, Real-Time, Unadorned Lip Tracking", *Proceedings of the 6th International Conf. on Computer Vision*, pp. 370-375, 1998.
- [5] Iain Matthews, 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.
- [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] L. Zhang, "Estimation of the mouth features using deformable templates", *IEEE International Conference on Image Processing*, Vol. III, pp. 328-331, 1997.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] Luettin, J, and Thacker, NA, "Speechreading using probabilistic models", *COMPUTER VISION AND IMAGE UNDERSTANDING*, vol. 65, pp. 163-178, 1997.
- [13] 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.
- [14] M. B. Stegmann, R. Fisker, "On Properties of Active Shape Models", *Informatics and Mathematical Modelling*, Technical University of Denmark, 2000.
- [15] Chad Carson, Serge Belongie, Hayit Greenspan, Jitendra Malik, "Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying", *IEEE Tans. on Pattern Recognition and Machine Analysis*, Vol. 24, No. 8, pp. 1026-1038, Aug. 2002.
- [16] 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(공학석사). 1996년 연세대학교 대학원 전자공학과 졸업(공학박사). 1994년~1998년 2월 대우전자 전략기술 제 1 연구소 선임연구원. 1998년~현재 동의대학교 멀티미디어공학과 조교수. 관심분야는 컴퓨터 비전, 인공지능 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. 관심분야는 영상처리, 컴퓨터 비전 등