

입술의 형태 모델과 Down Hill 탐색 방법을 이용한 입술 인식

이임건[†] · 장경식^{††}

요 약

본 논문은 입술형태를 인식하기 위한 방법을 제안하였다. 입술은 GLDM(Gray Level Distribution Model)을 기반으로 표현하였으며 입술인식은 모델에 대한 입력영상의 정확도에 대한 목적함수를 최적화하는 문제로 단순화하였다. 최적화를 위해 다운힐 심플렉스(Down Hill Simplex) 알고리즘을 이용하였으며 지역 최소점으로 수렴하는 문제를 해결하기 위한 새로운 방법을 제안하였다. 제안한 방법으로 기존의 능동적 형태 모델(ASM : Active Shape Model)에서 찾지 못하던 입술의 윤곽을 찾아낼 수 있음을 보였다.

Lip Recognition using Lip Shape Model and Down Hill Search Method

Imgeun Lee[†] and Kyung Shik Jang^{††}

ABSTRACT

In this paper, we propose a novel method for lip recognition. Lip model is built based on the concatenated gray level distribution model, and the recognition problem is simplified as the minimization problem of matching object function. The Down Hill Simplex Algorithm is used for the minimization with the proposed novel method for setting initial condition, which can refrain iteration from converging to local minima. The proposed algorithm shows extracting lip shape from the test image where Active Shape Model fails.

Key words: 입술 인식, 능동적 형태 모델(ASM : Active Shape Model), Down Hill Simplex

1. 서 론

근래 들어 얼굴 내에서 눈과 입의 위치뿐만 아니라 시간에 따른 이들의 형태변화에 대한 정보를 추출하여 응용하려는 연구가 활발히 진행되고 있으며, 이들 중에서도 입 모양을 찾거나 추적하는 연구는 최근에 많은 관심을 끌고 있다[1,2]. 입 모양 추출 결과는 여러 분야에서 응용 가능하다. 음성 인식 분야에서 음성 데이터와 화자의 입 모양 변화를 같이 사용하면 많은 개선을 얻을 수 있음이 알려져 있기 때문에 이에 대한 연구가 많은 관심을 끌고 있다[3,4]. 또한 얼굴에서 입술은 가장 변화가 심한 부분이기 때문에 입술의 변화를 인식하면 얼굴 영상의 전송시 많은 압축률을 얻을 수 있다. 그래픽 애니메이션(animation) 분야에서 말하는 사람의 변화하는 입 모양을 정확히 추출하면 화자의 입 모양을 보다 쉽게 실시간으로 생성할 수 있다. 또한, 사람이 말할 때 변화하는 입 모양은 각 개인마다 다르기 때문에 최근에는 각 개인을 식별하는 방법에 대한 연구가 진행되고 있다[5].

입술형태를 인식하는 연구로서 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. Wark[6]는 에지 정보와 색 정보를

접수일 : 2003년 1월 3일, 완료일 : 2003년 6월 20일

[†] 정희원, 동의대학교 영화영상공학과 전임강사

^{††} 동의대학교 멀티미디어공학과 조교수

이용하여 윗입술과 아랫입술의 경계선을 각각 4차 방정식과 2차 방정식으로 근사화하여 입술의 경계선을 추출하였다. Zhang[7]은 입술의 양쪽 끝점을 찾고 이들 점을 연결하는 선과 입 부분에서 찾은 에지, 교차하는 점의 수 등에 대한 규칙을 사용하여 열린 입과 닫힌 입을 구분하고 에지 강도와 입술에서 붉은색 성분 차이 및 분산을 이용하여 입술을 인식하였다. Basu[8]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[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. Luettin[12,13]과 Cootes[5]는 능동적 형태 모델을 이용하여 입술을 인식하였다. PDM(Point Distribution Model) 방법을 사용하여 입술을 표현하고 주성분 분석법(PCA : Principle Component Analysis)을 적용하여 입술 형태 모델을 구성하였으며 입술을 구성하는 각 점에서 profile을 추출하여 입술 경계선 모델을 구성하였다. 각 점의 인근 영역을 탐색하여 형태계수(shape parameter)와 자세계수(pose parameter)를 변화시키며 입술을 인식하였으며 좋은 결과를 얻었다.

색 정보를 이용하는 방법[2,8]은 입술 색의 개인별 차이, 조명의 영향을 많이 받으며, 에지 및 투영을 이용하는 방법은 입술 주변에 많은 에지가 존재하기 때문에 입술의 정확한 경계선을 찾기 어렵다. snake를 이용하는 방법[9,10]의 경우 계수설정 및 수렴성(convergence)의 기준 등이 잘못되는 경우 오류를 발생시킬 수 있다. 또한, Delams의 방법[9]에서는 입술 바깥쪽 경계선의 인식률은 높았지만 입을 벌린 경우에 안쪽 경계선의 인식률은 낮다. 이에 반해 Luettin의 방법[12]은 입술을 벌린 경우에도 안쪽 경계선의 추출에 높은 인식률을 보였으나 입술의 탐색 시 많은 시간이 소요되며 초기 위치의 선택에 따라 인식율의 차이가 많다.

이 논문은 입술을 효과적으로 인식하는 방법을 제안하였다. 학습 영상을 통계적으로 분석하는 주성분 분석법을 이용하여 입술 모델을 구하였다. 능동적 형태 모델은 물체의 인식시 초기 위치의 설정에 따라 잘못된 결과를 얻을 수 있기 때문에 이 논문에서는

입술의 형태에 기반한 가중치 벡터를 이용하여 아랫입술과 윗입술 사이의 경계선을 찾고 이 경계선을 중심으로 위치계수를 결정하여 초기 위치의 정확성을 높였다. 다운힐 심플렉스 탐색 방법을 사용하는 경우 최적화 벡터의 차원이 높으면 지역 최소점(local minima)에 빠지는 경우가 많고 탐색 시간이 오래 걸리므로 초기 위치와 형태계수의 선택이 중요하다. 제안하는 방법은 입술을 구성하는 각 점에서의 화소 분포를 이용하여 대략적인 입술의 형태를 먼저 찾은 뒤 이를 초기값으로 선택하고 다운힐 심플렉스 탐색을 이용하였다. 또한 다운힐 심플렉스에서의 최적화 벡터의 편차를 정규화하여 알고리즘의 수렴성을 보장하였다. 제안하는 알고리즘은 다양한 입술 영상에 대해 우수한 인식율을 보였고 기존의 알고리즘에서 오인식하는 영상에 대해서도 입술의 내부와 외부의 윤곽선을 정확히 인식하였다.

2. 입술 형태 모델링

입술에 관한 대부분의 정보는 윤곽선에 포함되어 있으나 이를 계조도를 이용해 에지를 찾는 방법은 개별적인 화자들의 특성, 조명에 따라 오인식의 가능성성이 높고 인식 구분자를 설계하기가 매우 어렵다. 또한 특징 벡터의 차원이 높아지고 임여성분이 많이 남아 있어 실용적이 못하다. 모델 기반의 입술 인식 접근 방법은 윤곽의 특징을 낮은 차원의 벡터로 표현 가능하고 이동, 회전, 스케일등의 변화에 무관하게 모델을 구성할 수 있어 많이 이용된다.

인식 분야에서의 능동적 형태 모델은 학습 영상에서 입술 주변의 형태를 따라 점을 찍어 통계적 모델을 구성하는 것이다. 입술 형태에 따라 찍은 점들의 좌표로 구성되는 벡터를 x_s , x_s 의 평균을 \bar{x}_s 라고 하면 입술의 형태는 다음 식으로 표현될 수 있다.

$$\begin{aligned} x_s &= \bar{x}_s + P_s \cdot b_s \\ \text{where, } P_s &= \begin{bmatrix} P_1 & P_2 & P_3 & \cdots & P_t \end{bmatrix}, \quad \lambda_i \geq \lambda_{i+1} \\ b_s &= (b_1, b_2, \dots, b_t)^T \end{aligned} \quad (1)$$

여기서 P_s 는 x_s 의 공분산 행렬(covariance matrix)의 고유값(eigenvalue) λ_i 를 크기 순으로 정

렬하고 가장 큰 값을 갖는 t 개의 고유값에 대응하는 고유벡터(eigenvector)들로 구성되는 행렬을 의미하며 b_s 는 t개의 고유벡터들에 대한 가중치들로 구성되는 형태계수이다. 고유 벡터는 직교성이 있으므로 형태계수 b_s 는 다음 식으로 구할 수 있다.

$$b_s = P_s^T (x_s - \bar{x}_s) \quad (2)$$

본 논문에서는 그림 1에서와 같이 입술의 외부와 내부에 N개(외부 21개, 내부 20개)의 기준점을 찍고 이 점들로부터 직교방향으로 K개의 프로파일 정보를 얻어 연속으로 나열한 뒤 모델을 얻는다. 각 그림은 입술 경계를 따라 찍은 기준점과 기준점에서의 프로파일을 보인 것이다. M개의 학습 영상에서 i번째 영상의 j번째 기준점에서의 프로파일 벡터를 g_{ij} 라 하면 i번째 영상의 입술 모델에 대한 연속 프로파일 벡터는 다음과 같이 주어진다.

$$x_p = (g_{0i}, g_{1i}, \dots, g_{(N-1)i})^T \quad \text{where } 0 \leq i \leq M-1 \quad (3)$$

그러므로 M개의 전체 학습 영상에 대한 입술 모델은 식 (3)과 같이 N개 점들에 대한 프로파일 정보를 갖는 $K \times N$ 개의 행과 M개의 열을 갖는 행렬로서 표현된다. 학습 영상의 연속 프로파일 행렬의 각 벡터에 대한 평균 벡터를 구하고 식 (4)에 따라 공분산 행렬을 구한 후 공분산 행렬의 고유벡터와 고유값을 구하였다. 이 식에서 P_p 는 고유 벡터로서

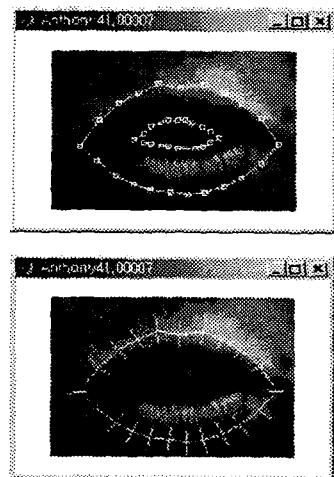


그림 1. 입술 모델링과 특징 추출

$P_p^T P_p = E$ 를 만족하는 직교행렬(orthogonal matrix)이고 λ_p 는 고유값이다. 입술 모델은 식 (5)에 의하여 표현된다. 이 식에서 P_p 는 고유값을 크기 순으로 정렬하고 가장 큰 값을 갖는 t 개의 고유값에 대응하는 고유벡터를 취하여 구성하였다. t는

$$\sum_{i=1}^t \lambda_i \geq V \sum_{i=1}^{K+1} \lambda_i \text{ 식에 의하여 결정하며 } V \text{는 상수이다. } b_p \text{는 고유 벡터들에 대한 가중치(weight)들로 구성되는 벡터로서 형태계수(shape parameter)이며, 식 (6)에 의하여 결정된다.}$$

$$\Sigma P_p = \lambda_p P_p \quad \text{where, } \Sigma = \frac{1}{M} \sum_{i=1}^M (X_p - \bar{X}_p)(X_p - \bar{X}_p)^T \quad (4)$$

$$X_p = \bar{X}_p + P_p \cdot b_p$$

$$\text{where, } P_p = \begin{bmatrix} P_{p_1} & P_{p_2} & P_{p_3} & \cdots & P_{p_t} \end{bmatrix}, \quad \lambda_p \geq \lambda_{p+1}, \\ b_p = (b_{p_1}, b_{p_2}, \dots, b_{p_t})^T \quad (5)$$

$$b_p = P_p^T (X_p - \bar{X}_p) \quad \text{where, } -3\sqrt{\lambda_p} \leq b_{p_i} \leq 3\sqrt{\lambda_p} \quad (6)$$

3. 목적함수를 이용한 입술 인식

본 논문에서의 입력영상의 입술은 2장에서 구성한 입술 모델에 대해 목적함수를 구하고 이를 최소화하여 인식한다. 목적함수의 최적해는 다운힐 심플렉스를 이용하였다. 다운힐 심플렉스는 계산량이 많으나 간단하게 구현할 수 있고 기울기 정보를 요구하지 않으므로 지역최소점으로 수렴하는 오류가 적다 [14]. 다운힐 심플렉스를 이용하기 위해서는 최적화해야 할 벡터의 차원보다 하나 많은 초기점을 설정하여 심플렉스를 만들고 이 심플렉스의 수축, 확장을 통해 최적값을 찾는다.

3.1 아래위 입술의 경계선 검출

능동적 형태 모델을 이용하여 탐색하는 경우 초기 위치에 영향을 받게 되므로 본 논문에서는 입술을 인식하는 첫 단계로서 두 입술 사이의 경계선을 찾고 그 위치를 중심으로 입술을 탐색하였다. 아랫입술과 윗입술의 경계 부분의 영상이 가장 어두우므로 영상의 각 열(column)에서 식 (7)을 이용하여 가장 어두운 화소들을 찾고 식 (8)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 K_1, K_2 는 상수이며, H_{Img}, W_{Img} 은 각각 입력 영상의 높이와 폭을 나타낸다. $Y_{LipCenterLine}$ 은 영상에서 아랫입술과 윗입술의 사이를 가로지르는 선에 대한 높이 좌표값을 나타낸다.

$$M(x) = \arg \min_{y \in H_{Img}} I(x, y) \quad (7)$$

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

$$Y_{LipCenterLine} = \arg \max_{y \in H_{Img}} L(y) \quad (8)$$

3.2 초기 입술 형태계수의 결정

두 입술 사이의 경계선 위치($Y_{LipCenterLine}$)가 정해지면 입술 모델을 그 위치로 평행 이동한 후 입술을 찾는 과정을 수행하였다. 이 과정은 모델을 구성하는 각 기준점의 프로파일을 이용하여 입술 경계선을 탐색하는 과정과 점들의 전체적인 모양을 허용범위 내로 조정하는 과정의 두 단계로 구성된다. 입력 영상에서 입술을 찾는 과정은 다음과 같다.

1. b 는 0을, (T, θ, s) 는 초기값을 설정한다. T 는 (t_x, t_y) 로서 평행 이동량, θ 는 회전각, s 는 스케일비를 나타낸다. 이 논문에서는 t_y 는 이전 단계에서 구한 두 입술 사이의 경계선 y 좌표를, t_x 는 영상 폭의 $1/2$ 을 사용하였다. 또한 θ 는 0, s 는 임의값을 사용하였다.

2. 입술 형태 모델을 표현하는 $x_s = \bar{x}_s + P_s \cdot b_s$ 식에 의하여 입술의 인스턴스를 생성하고 $T_{(T, \theta, s)}(x_s)$ 식에 의하여 스케일, 평행이동, 회전시킨다. $T_{(T, \theta, s)}(x_s)$ 는 평행 이동량 T , 회전각 θ , 스케일 비 s 를 사용하여 x_s 를 선형 기하학적 변환한다.

3. 인스턴스 각 기준점에서의 프로파일을 이용하여 입술의 경계선에 해당하는 새로운 점을 찾고 이들 점으로 구성되는 모양을 Y 로 한다. 식 (9)를 최소화하는 (T, θ, s) 를 찾고 이를 새로운 (T, θ, s) 로 정하며, $y = T_{(T, \theta, s)}^{-1}(Y)$ 식을 이용해 역변환 후 $y = \frac{y}{y \cdot x_s}$ 식에 의해 정규화한다.

4. $b_s = P_s^T (y - \bar{x}_s)$ 식을 이용해 b_s 를 구하고 식 (10)를 이용해 b_s 에 대한 조건을 조사하고 만족하지 않으면 b_s 값을 조정한다. 이때 D_{max} 는 상수이다.

5. b 와 (T, θ, s) 가 수렴하면 중지한다. 그렇지 않으면 단계 2로 간다.

$$E = |T_{(\theta, t, s)}(\bar{x}_s) - Y|^2 \quad (9)$$

$$D_m^2 = \sum_{i=1}^l \left(\frac{b_{s,i}^2}{\lambda_i} \right) \leq D_{max}^2, \quad b_s = b_s \cdot \frac{D_{max}}{D_m} \quad (10)$$

입력 영상에서 입술의 인스턴스를 구성하는 각 기준점의 주변 점들에서 프로파일을 추출하고 모델의 대응하는 기준점의 프로파일과 비교하여 가장 비슷한 프로파일을 갖는 위치로 각 기준점이 이동한다. 즉, 입술 모델을 구성하는 각 기준점의 프로파일을 구하는 방법과 같이 입술의 인스턴스를 구성하는 j 번째 기준점에서 현재 점을 중심으로 양쪽에서 $m (m > K)$ 개의 화소 값을 취하여 프로파일 X_{pj}' 를 구성하였다. 이후 X_{pj}' 에서 차례로 $(2K+1)$ 개 값을 취하여 x_{pj}' 를 구성하고 이들 x_{pj}' 중에서 모델의 j 번째 점에서의 평균 프로파일 \bar{X}_{pj} 와 가장 비슷한 부분을 찾았다. 이를 위하여 식 (11)의 마할라노비스 거리 (mahalanobis distance) 적도를 최소로 하는 x_{pj}' 를 선택하고 x_{pj}' 의 중간 부분에 해당하는 점을 인스턴스의 j 번째 기준점이 이동할 점으로 정하였다. S_{pj} 는 모델의 j 번째 기준점에서 프로파일들의 공분산 행렬이다.

$$f(x_{pj}') = (x_{pj}' - \bar{X}_{pj})^T S_{pj}^{-1} (x_{pj}' - \bar{X}_{pj}) \quad (11)$$

3.3 최적화 벡터와 심플렉스(Simplex)의 결정

인식을 위한 최적화 벡터는 능동적 형태 모델에서의 가중치 벡터와 입술의 자세에 관련되는 이동, 회전, 스케일링 파라미터를 결합하여 구성한다. 선형 이동, 회전, 스케일 요소를 각각 $\bar{T} = (t_x, t_y), \theta, s$ 라고 하면 최적화 벡터는 다음으로 주어진다.

$$(b_1, b_2, \dots, b_t, t_x, t_y, \theta, s)^T \quad (12)$$

2장에서 구한 연속 프로파일 모델에서 x_p 와 이의 예측치 \hat{x}_p 와의 오차 e 와 MSE(Mean Square

Error) E^2 는 다음 식으로 주어진다.

$$e = x_p - \hat{x}_p = (x_p - \hat{x}_p) - P_p b_t \quad (13)$$

$$E^2 = (x_p - \bar{x}_p)^T (x_p - \bar{x}_p) - b_p^T b_p \quad (14)$$

그리므로 입술을 인식하는 문제는 식 (12)의 최적화 벡터에 대해 목적함수 식 (14)을 최소화 시키는 문제로 단순화된다. 다운힐 심플렉스를 이용하여 목적함수의 최소값을 찾는 과정은 초기값의 선택에 따라 연산량의 차이가 많으므로 본 논문에서는 초기값을 결정하는 과정으로 능동적 형태 모델을 이용하였다. 능동적 형태 모델로 찾아진 형태 좌표들을 이용해 식 (12)의 최적화 벡터를 구한다.

식 (7)과 식 (8)을 이용하여 입술을 가로지르는 경계선을 찾은 뒤 능동적 형태 모델로 얻어진 입술모양의 윤곽을 나타내는 최적화 벡터를 다음과 같이 표현하면 다운힐 심플렉스 탐색 방법에서의 초기 심플렉스는 아래 식으로 주어진다.

$$O^0 = (b_1, b_2, \dots, b_t, t_x, t_y, \theta, S)^T \quad (15)$$

$$O^j = O^0 + a^j e^j, \text{ where } 1 \leq j \leq N \quad (16)$$

이 식에서 e^j 는 $N=t+4$ 개의 원소를 갖는 단위 벡터이며 a^j 는 각 벡터의 방향 성분에 대한 가중치 상수이다. 식 (15)에서 최적화의 대상이 되는 벡터에는 형태계수와 이동, 회전, 스케일 요소를 모두 포함하고 있어 차원이 높으므로 초기값의 선정에 의한 수렴 시간의 차이가 많이 발생한다. 본 논문에서는 초기값을 능동적 형태 모델의 결과로 얻어진 값을 이용해 수렴 시간을 줄인다.

심플렉스 행렬 O 의 각 벡터는 입술의 형태를 결정짓는 정보들로 구성되어 있으며 능동적 형태 모델에서 처음 t 개의 고유값 벡터 b 와 이동변환 벡터 T , 회전 변환 인자 θ 그리고 스케일 요소 S 로 구성되어 있다. 이들 요소들의 값들은 입술의 형태에 미치는 영향이 모두 상이하므로 방향 성분에 대한 가중치 a^j 를 특성에 맞게 결정하지 않으면 다운힐 심플렉스는 최소값으로 수렴 못하는 경우가 많다. 특히 다운힐 심플렉스의 경우 가장 큰 오차를 만들어 내는 좌표점을 수축, 확장하는 과정을 통해 최소값에 수렴하므로 최적화 벡터의 요소들을 동일하게 다루면 지역 최소점에 빠지기 쉽다. 본 논문에서는 이를 해결하기 위해 최적화 벡터에서 고유값 요소 b 를 가우시안 분

포의 변화도의 90%를 초기값으로 결정하고 형태적 변환에 관련있는 T, θ, S 는 값의 90%를 오프셋 값으로 정해 이를 차감한 것을 초기값으로 정하였다. 차감된 오프셋 값은 최소값을 찾은 뒤 다시 더해진다. 가중치 벡터 a' 는 식 (17)로 주어지며 식에서 O_k^0 는 O^0 의 k 번째 원소이고 λ_k 는 k 번 째 고유값을 의미한다.

$$a' = \begin{cases} -O_k^0 + 0.9 \times 3\sqrt{\lambda_k} & \text{where } 1 \leq k \leq t \\ -O_k^0 + 0.1 \times O_k^0 & \text{where } t+1 \leq k \leq N \end{cases} \quad (17)$$

다운힐 심플렉스의 최소값 수렴은 초기점들 중에서 목적 함수의 값을 가장 크게 하는 점을 수축, 확장시키는 과정을 포함하는데 이 과정에서 입력 영상의 크기를 벗어나 프로파일 정보를 얻지 못하는 경우가 발생한다. 모델을 생성할 때에도 이러한 경계점 문제가 발생할 수 있는데 이 경우 모델의 생성과 인식 단계에서 동일한 방법을 사용한다. 본 논문에서는 경계에 있는 점의 픽셀 값을 연장하는 방법을 이용하였다.

4. 실험 및 고찰

실험에 사용한 입술 영상은 Tulip 1 데이터베이스 [15]에 있는 영상들을 사용하였다. 이들은 다양한 조명상태에서 총 12명(남자 9명과 여자 3명)이 'One', 'Two', 'Three', 'Four'의 4개 숫자를 두 번씩 발음할 때 입력된 연속적인 명암 영상(gray image)으로서 첫 번째 발음할 때의 영상들을 집합 1, 두 번째 발음 할 때의 영상들을 집합 2라고 지칭한다. 이들은 입술 영역만을 포함하고 있으며 다양한 입술 형태를 가지고 100×75 크기를 갖는다. 이 논문에서는 학습 영상으로 집합 1의 492개 중에서 200개의 입술 영상을, 실험 영상으로 집합 2의 464개 중 220개의 영상을 무작위로 추출하여 사용하였다.

3.1절에서의 식 (7), (8)로부터 아래의 입술 경계선을 찾는 과정을 (그림 2)에 보였다. 제안한 방법은 입술이 다물어져 있는 영상이나 입술이 벌어져 치아가 보이는 영상에서도 경계선을 잘 찾아낼 수 있다. (그림 2)의 (a), (c), (e)는 각 영상의 행에서 얻어진 $L(y)$ 의 계산값을 그래프로 보인 것이고 (b), (d), (f)는 $L(y)$ 를 이용하여 실제 입술영상에서 경계선을

나타낸 것이다. 그림 2의 각 영상에서의 $Y_{LipCenterLine}$ 의 선택 값은 각각 31, 30, 39이며 윗입술과 아랫입술의 경계선은 구해진 값에서 가로로 평행한 선분이 된다. 상수값 K1과 K2는 80과 10을 이용하였다.

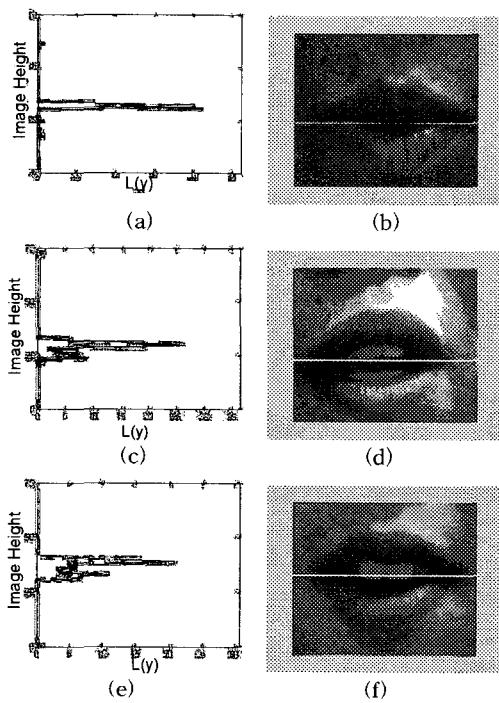
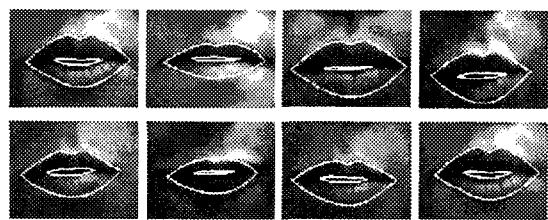


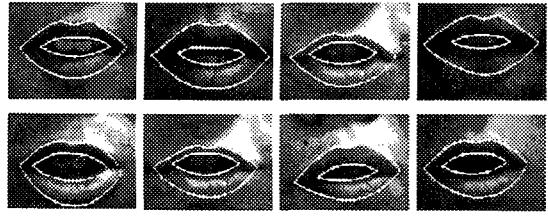
그림 2. 아래위 입술선 검출

다운힐 심플렉스의 목적함수의 연산 횟수는 5000번으로 제한하였으며 이 횟수 동안 수렴하지 못한 경우는 오인식으로 판정하였으나 연산 횟수의 한계를 넘어가는 경우는 발생하지 않았다. 목적함수의 오차 한계는 10^{-5} 으로 하였다. 인식결과 기준의 능동적 형태 모델을 사용한 경우 인식하지 못했던 영상에 대해서도 잘 인식하였다. 그림 3과 그림 4에 다양한 조명과 형태별로 제안한 알고리즘이 입술 경계를 인식하는 모습을 보였다.

실험결과를 분석하기 위하여 인식한 입술의 경계선이 실제 위치와 일정 거리 이상 떨어진 경우를 오인식으로 판정하였다. 실제 입술의 윤곽은 시각적인 인스펙션(inspection)을 사용하였으며 이때 거리의 기준으로 윗입술 높이의 1/2을 사용하였다[12,13]. 또한 인식율의 객관성을 높이기 위해 초기위치의 선택에 제한이 없는 연속 프로파일 모델 기반의 기준 알고리즘[12]을 구현하여 두 가지 방식의 결과를 비교



(a) 입술이 다물어져 있는 경우의 인식

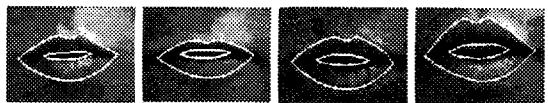


(b) 치아가 보이는 경우의 인식



(c) 혀가 보인 경우의 인식

그림 3. 입술의 형태별 인식 결과



(a) 조명이 전체적으로 고른 경우의 인식



(b) 조명이 부분적으로 밝은 경우의 인식



(c) 콧수염이 있는 경우의 인식

그림 4. 피부와 주변 조명 상태에 따른 인식 결과

하였다. 테스트 영상의 오인식 형태는 위아래 입술 각각의 바깥쪽과 안쪽 경계선에서의 오인식과 바깥쪽과 안쪽 입술의 폭을 잘못 인식하는 총 6가지로 분류하여 오인식을 분석하였다. 각 형태별 오인식 빈도를 표 1에 보였으며 표 1의 오인식 분류인 (1), (2), (3)이 두 가지 이상 동시에 발생하는 오인식인 경우는 폭의 오인식으로 분류하였다.

표 1. 형태별 오인식 발생 결과

인식 성공		인식 실패						전체	
		윗입술		아랫입술		입술 폭			
		바깥쪽 경계선 (1)	안쪽 경계선 (2)	바깥쪽 경계선 (3)	안쪽 경계선 (4)	바깥쪽 폭 (4)	안쪽 폭 (5)		
능동적 형태모델	영상 개수(개)	187	0	2	2	6	15	8	220
	비율(%)	85.0	0	0.9	0.9	2.8	6.8	3.7	100
제안한 방법	영상 개수(개)	199	0	3	2	3	8	5	220
	비율(%)	90.5	0	1.4	0.9	1.4	3.6	2.3	100

윗입술의 바깥쪽 경계선에 대한 인식 오류는 발생하지 않았으나 입력 영상에서 아랫입술과 주변 피부와의 경계가 뚜렷하지 못한 경우 아랫입술의 바깥쪽에 빈도는 작지만 오인식이 발생하였다. 오인식은 주로 입술 안쪽 부분에서 발생하였으며 이는 치아와 혀가 보이는 경우 경계선에 대한 프로파일 값이 테스트 영상의 가우시안 분포와 차이가 많이 나는 경우이다. 그러나 치아와 혀의 영향은 능동적 형태 모델의 경우보다는 비교적 적음을 알 수 있다. 입술의 안쪽과 바깥쪽의 너비를 오인식 하는 경우도 비교적 많이 발생하였으며 이는 입술의 양 끝점에서 경계가 불분명하여 끝점을 정확히 찾지 못하는 경우이다. 너비 오인식은 입술 안쪽과 바깥쪽이 같은 빈도로 발생하였다.

제안한 인식 알고리즘에서 초기값은 능동적 형태 모델을 이용한 인식 결과를 선택하였으므로 능동적 형태 모델의 결과가 실제 입술 윤곽에서 너무 많은 오차가 나는 경우는 다운힐 심플렉스가 수렴하지 못하게 된다. 두 알고리즘이 모두 오인식을 하는 경우를 그림 5에 보였다.

그림 5에서 왼쪽의 그림은 능동적 형태 모델에서 초기의 자세값(pose parameter)의 오류가 심한 경우이며 중간 그림은 혀와 치아가 동시에 보여 경계선이 여러 개가 된 경우이다. 마지막 그림은 아래 위의 치아가 모두 드러난 경우로 인식을 못하고 목적함수가 지역 최소값으로 수렴한 경우이다.

능동적 형태 모델이 찾지 못한 입술의 경계를 제안한 알고리즘으로 정확하게 인식하는 예를 그림 6에 보였다. 입을 다문 경우, 입을 조금 벌린 경우와 많이 벌린 경우, 입술을 내민 경우, 콧수염이 있는

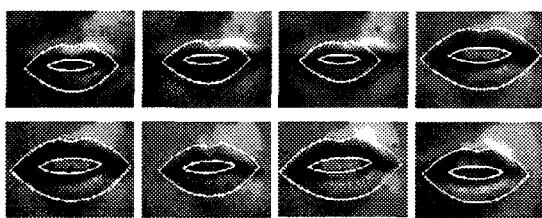


(a) 능동적 형태 모델

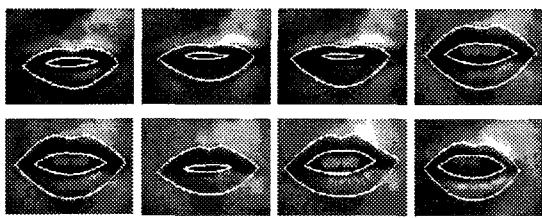


(b) 제안한 방법

그림 5. 능동적 형태 모델과 제안한 방식이 모두 오인식하는 예



(a) 능동적 형태 모델



(b) 제안한 방법

그림 6. 능동적 형태 모델이 오인식한 것을 제안한 방식으로 인식하는 예

경우 등 여러 상황에서 능동적 형태 모델보다 나은

인식 결과를 보인다.

그림 7은 능동적 형태 모델에 의해 정상적으로 인식된 형태를 다시 제안한 방법을 적용하여 오히려 오인식을 한 예이다. 이 경우에 해당하는 영상은 대부분 발음을 위해 입술을 오므리는 경우 내부의 피부가 돌출되어 안쪽 입술의 형태가 바르지 못하고 치아를 포함한 내부 영역이 경계를 찾지 못할 정도로 어두운 영상들이다. 다운힐 심플렉스는 다수의 반복 연산을 수행하며 수렴 포인트를 찾아가나 이를 영상에 대해서는 수렴 위치를 찾지 못해 지역 최소점에 빠지는 경우들이었다. 결과 그림에서 보면 윗입술의 안쪽 경계가 수렴 위치를 찾지 못하고 윗입술 바깥경계로 다가가는 것을 볼 수 있다. 전체적으로 220개의 테스트 영상중에 능동적 형태 모델의 결과로 정상 인식된 입술을 다시 제안한 알고리즘으로 오인식 하는 경우는 5개였으며 능동적 형태 모델이 찾지 못한 입술 형태를 정상적으로 인식하는 경우가 17개였다.

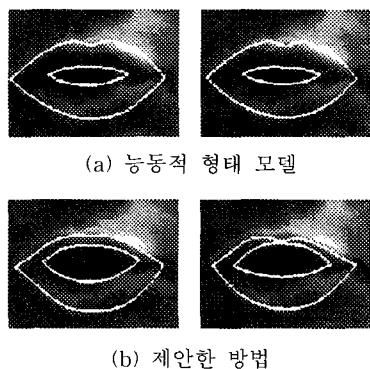


그림 7. 능동적 형태 모델이 인식한 것을 제안한 방식으로 오인식하는 예

5. 결 론

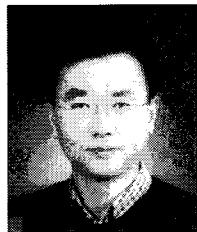
본 논문에서는 연속 프로파일 모델을 이용하여 입술 모델을 구하고 능동적 형태 모델과 최적화 알고리즘을 이용하여 입술을 인식하는 방법을 제안하였다. 최적화 알고리즘으로 이용한 다운힐 심플렉스 알고리즘은 입술인식을 위한 연산량이 많으므로 빠른 수렴을 위해서 능동적 형태 모델의 결과를 이용해서 초기값을 결정하였으며 보다 나은 능동적 형태 모델의 결과를 위해 입술의 형태에 기반한 가중치 벡터를 이용하여 초기 위치를 설정하고 입술의 경계선을 찾는 방법을 제안하였다. 또한 다운힐 심플렉스 알고리

즘에서는 초기 심플렉스의 구성에 관한 새로운 방법을 제안하여 지역 최소점으로의 수렴할 가능성을 낮췄다. 제안한 방법을 여러 입술 영상에 대해 실험한 결과 능동적 형태 모델로 인식하지 못하는 패턴의 입술을 정확히 인식하였으며 전체적으로 약 5% 이상의 인식률 향상을 얻었다.

참 고 문 헌

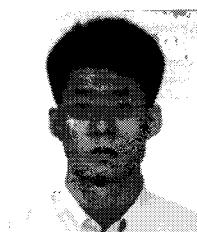
- [1] Mrhosseini A. R., H. Yan and K. M. Lam, "Adaptive Deformable Model for Mouth Boundary Detection", *Optical Engineering*, Vol. 37 No. 3(1998), pp. 869-875.
- [2] Oliver N., A. Pentland, "LAFTER: Lips and Face Real Time Tracker", *Proceedings of the 1997 Conf. on Computer Vision and Pattern Recognition*, (1997), pp. 123-129.
- [3] 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.
- [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 Bangham, Stephen Cox and Richard Marvey, "Extractoin of Visual Features for Lipreading", *IEEE Trans. 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] J. A. Nelder and R. Mead, "A Simplex Method for Function Minimization", *Computing J.* Vol. 7, No. 4, pp.308-313, 1965.
- [15] Movellan J. R., "Visual Speech Recognition with Stochastic Networks", *Advances in Neural Information Processing System*. Vol. 7, MIT Press Cambridge, 1995.



이 임 건

1991년 연세대학교 전자공학과
졸업(학사)
1993년 연세대학교 대학원 전자
공학과 졸업(공학석사)
1998년 연세대학교 대학원 전자
공학과 졸업(공학박사)
1998년~2001년 대우전자 디지털
TV 연구소 선임연구원
2002년~현재 동의대학교 영화영상공학과 전임강사
관심분야 : 영상처리, 컴퓨터 비전 등
E-mail : iglee@dongeui.ac.kr



장 경 식

1989년 연세대학교 전자공학과
졸업(학사)
1991년 연세대학교 대학원 전자
공학과 졸업(석사)
1996년 연세대학교 대학원 전자
공학과 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
E-mail : ksjang@dongeui.ac.kr

교 신 저 자

이 임 건 614-714 부산광역시 부산진구 가야동 산24
동의대학교 영화영상공학과