
에너지 최소화 기반 능동형태 모델을 이용한 입술 윤곽선 추출

장 경 식*

Lip Contour Extraction Using Active Shape Model Based on Energy Minimization

Kyung-Shik Jang*

이 논문은 2004년도 동의대학교 교내학술연구 지원사업 연구비(과제번호 2004AA165)를 지원받았음

요 약

이 논문에서는 능동형태 모델을 개선하여 입술의 형태를 효과적으로 추출하는 방법을 제안하였다. 입술의 형태 변형은 능동형태 모델에 기반을 둔 통계적 형태변형 모델을 사용하여 표현하였다. 능동형태 모델에서 각 점은 지역적인 정보인 프로파일을 기반으로 독립적으로 이동하기 때문에 많은 오류가 발생할 수 있다. 전역적인 정보를 사용하기 위하여 이 논문에서는 능동윤곽선 모델에서 사용하는 것과 유사한 에너지 함수를 정의하고 전체 에너지가 최소화되는 위치로 점들이 이동하게 하였다. Tulip 1 데이터베이스에 있는 입술 영상을 대상으로 실험한 결과, 제안한 방법이 기존 방법보다 실제 형태에 가깝게 입술을 추출하였다.

ABSTRACT

In this paper, we propose an improved Active Shape Model for extracting lip contour. Lip deformation is modeled by a statistically deformable model based Active Shape Model. Because each point is moved independently using local profile information in Active Shape Model, many error may happen. To use a global information, we define an energy function similar to an energy function in Active Contour Model, and points are moved to positions at which the total energy is minimized. The experiments have been performed for many lip images of Tulip 1 database, and show that our method extracts lip shape than a traditional ASM more exactly.

키워드

능동형태 모델, 능동윤곽선 모델, 입술 인식, 주성분 분석법

1. 서 론

최근에 눈동자나 입술의 움직임에 대한 정보를 여러 분야에 응용하려는 연구가 활발히 진행되고 있으며, 이 중에서도 입 모양을 인식하거나 추적하는 연구는 많은 관

심을 끌고 있다. 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 인식률 개선을 얻을 수 있다[1, 2].

입술형태를 추출하는 연구로서 Oliver[3]는 입술색 모델을 설정하고 입의 형태적 특성을 반영하는 특징들과 시간에 따른 입술의 형태 변화를 HMM(Hidden Markov Model)을 이용하여 모델링하고 입술 윤곽선을 추출하였다. Wark[4]는 에지 정보와 색 정보를 이용하여 위와 아래 입술의 윤곽선을 각각 4차 방정식과 2차 방정식으로 근사화하여 입술의 윤곽선을 추출하였다. Basu[5]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[6]와 Lievin[7]은 색 정보와 MRF(Markov Random Process)에 기반하여 입술 영역을 찾고 snake를 이용하여 입술의 윤곽선을 추출하였다. Luettin[8]과 Iain[9]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 입술을 추출하였다.

많은 연구들이 입술이 다물어진 상태에서 입술의 바깥쪽 윤곽선만을 찾았으며 이는 입을 벌린 경우 치아와 혀로 인하여 안쪽 입술의 윤곽선을 찾는 것이 어렵기 때문이다. Delmas의 방법[6]에서는 입술 바깥쪽 윤곽선의 추출 성공률은 높았지만 입을 벌린 경우에 안쪽 윤곽선의 추출 성공률은 낮았다. 이에 반해 능동적 형태모델을 이용한 방법[8, 9]은 입을 벌린 경우에도 안쪽 윤곽선의 추출에 높은 추출 성공률을 보였다. 그러나 능동형태 모델을 이용하는 방법은 각 점의 인근 영역을 탐색하여 자신의 프로파일 모델과 비슷한 점을 찾는 과정에서 각 점이 독립적으로 이동하기 때문에 탐색결과가 실제 입술의 형태와 많은 차이가 나는 결과가 발생한다. 이후 과정에서 형태 계수의 분포를 사용하여 가장 유사한 입술 형태로 보정을 수행하지만 오류 원인의 대부분은 각 점이 독립적으로 이동하기 때문에 발생한다.

이 논문에서는 능동형태 모델을 개선하여 입술의 윤곽선을 효과적으로 추출하는 방법을 제안하였다. 입술의 형태 변형은 능동형태 모델에 기반을 둔 통계적 형태변형 모델을 사용하여 표현하였다. 각 점의 프로파일 모델을 기반으로 각 점이 새로운 점으로 이동하는 과정에서 능동 윤곽선 모델(Active Contour Model)에서 사용하는 외부 에너지와 내부 에너지 개념을 도입하고 전체 에너지가 최소화되는 위치로 점들이 이동하게 하였다. 여러 입술 영상

에 대해 실험한 결과 전체적으로 추출물의 향상을 얻었다.

II. 능동윤곽선 모델

능동윤곽선 모델은 에너지 함수로 정의되고 에너지 함수가 최소화되는 방향으로 점들이 반복적으로 움직이며 객체의 윤곽을 찾아내는 방법이다. $i=0, \dots, N-1$ 일때 $V_i=(x_i, y_i)$ 가 능동 윤곽선 모델의 i 번째 점이라고 하면, x_i, y_i 는 i 번째 점의 x, y 좌표를 나타낸다. 능동윤곽선 모델의 에너지 함수는 식 (1)과 같이 정의된다[10].

$$E_{snake} = \int_0^1 E_{snake}(v(s)) ds \quad (1)$$

$$= \int_0^1 (E_{int}(v(s)) + E_{image}(v(s)) + E_{con}(v(s))) ds$$

$E_{int}(v(s))$ 는 능동윤곽선 모델의 내부 에너지(internal energy)로서 연속성(continuity)과 부드러움(smoothness)에 의해 계산이 되는 값으로 윤곽선 모양을 결정하고 식 (2)와 같이 표현된다. 식 (2)의 첫 항은 윤곽선에 장력을 부여하고 점들 간의 거리를 균일하게 하는 에너지로 식 (3)과 같이 점들 간의 거리 차로 근사화된다. 두 번째 항은 곡선의 휨 정도를 나타내는 변화율(curvature)로서 윤곽선을 변화율이 작은 쪽으로 끌어당기는 역할을 하며 식 (4)와 같이 근사화된다.

$E_{image}(v(s))$ 는 능동윤곽선 모델의 외부 에너지(external energy)로서 영상 내에서 물체의 특징을 얼마나 잘 반영하는가에 의해 결정되는 값으로 능동윤곽선 모델을 물체의 경계선으로 이동하게 하는 역할을 한다. 보통 식 (5)과 같이 영상의 에지 정보를 사용한다. $E_{con}(v(s))$ 는 외부의 힘으로서 보통 물체에 대한 지식에 의하여 결정된다. 전체 에너지가 최소화된 능동윤곽선 모델은 물체에 대한 지식 및 특징을 잘 표현하면서 부드럽고 연속적인 곡선의 모양을 가지게 된다.

$$E_{int}(v(s)) = \alpha(s)|v_s(s)|^2 + \beta(s)|v_{ss}(s)|^2 \quad (2)$$

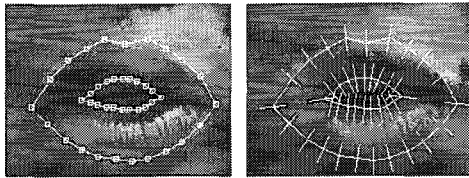
$$v_s(s) \approx |v_i - v_{i-1}| \quad (3)$$

$$v_{ss}(s) \approx |v_{i-1} - 2v_i + v_{i+1}| \quad (4)$$

$$E_{image}(v(s)) = \gamma(s) E_{edge}(v(s)) \quad (5)$$

III. 능동형태 모델

능동형태 모델은 능동윤곽선 모델로부터 개선되어 발전된 알고리즘으로 학습 데이터들의 상관관계를 이용하여, 임의의 새로운 이미지의 특성을 반복적인 과정에서 찾는 방법이다. 능동윤곽선 모델의 경우에 경계가 확연히 구분되는 영상에서 윤곽을 찾아내는 문제에는 유용하지만 표준적인 모델에 기반을 두고 변형하는 방법이 아니기 때문에 물체의 윤곽선이 뚜렷하지 않거나 잡음이 많은 경우에는 물체의 윤곽선을 찾는 데 한계가 있다.



(a) (b)
그림 1. 입술 형태와 프로파일
Fig. 1. Lip shape and Lip Profile

입술의 형태는 PDM(Point Distribution Model)을 이용하여 표현한다[9]. 입술과 주변 피부와의 경계선을 따라 그림 1의 (a)와 같이 바깥쪽과 안쪽 입술 윤곽선 위에 모두 41개의 기준 점을 위치시켰다. 입술은 식 (6)과 같이 기준 점들의 좌표로 구성되는 벡터로 표현하였다. 학습 영상에 있는 입술의 크기와 방향 등이 다양하기 때문에 정규화하였다[11]. 정규화된 벡터를 x , x 의 평균을 \bar{x} 라 하면 입술의 형태 모델은 식 (7)로 표현된다.

$$X_i = (x_1, y_1, x_2, y_2, \dots, x_{41}, y_{41})^T \quad (6)$$

$$x = \bar{x} + P_s \cdot b_s \quad (7)$$

여기서 P_s 는 x 의 공분산 행렬의 고유값(eigenvalue)을 크기순으로 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)이다[11].

학습 영상에서 그림 1의 (b)와 같이 입술을 구성하는 j 번째 점을 지나며 입술 윤곽선에 수직인 방향으로 점의 양쪽으로 n 개씩과 자신을 합하여 $(2n+1)$ 개 점을 선택하고 식 (8)과 같이 그 점들의 화소값들로 구성되는 프로파일 벡터 g_j 를 구하였다. 조명 등의 밝기 변화에 따른 영향을 최소화하기 위하여 식 (9)에 따라 정규화한 프로파일 벡터(g_j^n)를 구하였다. 식 (10)에 따라 학습에 사용된 입술의 동일 위치에 있는 점에서의 정규화한 프로파일들에 대한 평균벡터(\bar{g}_j^n)와 공분산 행렬(S_j)을 구한다.

$$g_j = (g_{1j}, g_{2j}, \dots, g_{(2n+1)j})^T, \quad (8)$$

where, g_{kj} 는 화소값

$$g_j^n = \frac{g_j^j}{\sum |g_{kj}^j|}, \quad (9)$$

where, $g_{kj}^j = g_{(k+2)j} - g_{kj}$

$$S_j = \frac{1}{m} \sum_{i=1}^m (g_{ij}^n - \bar{g}_j^n) (g_{ij}^n - \bar{g}_j^n)^T \quad (10)$$

where, m 은 학습 영상의 수

$$D_j = (h_j^n - \bar{g}_j^n)^T S_j^{-1} (h_j^n - \bar{g}_j^n) \quad (11)$$

입력 영상에서 입술을 탐색하는 과정은 입술 형태 모델을 이용하여 생성된 인스턴스(instance)를 구성하는 각 점이 자신이 가진 프로파일 모델과 가장 유사한 프로파일을 갖는 영상 위의 점으로 이동함으로써 수행된다[11]. 이를 위하여 인스턴스의 j 번째 점이 있는 영상위치에서 정규화된 프로파일 H_j^n 을 구한다. 이 과정은 각 점의 프로파일 모델을 구하는 과정과 같으며, H_j^n 는 현재 점을 중심으로 양쪽에서 $2k+1(k>n)$ 개의 화소 값을 취하여 구성한다. 이후 H_j^n 에서 차례로 $2n+1$ 개 값을 취하여 h_j^n 를 구성하고 이들 중에서 j 번째 점에서의 프로파일 모델인 \bar{g}_j^n 와 가장

비슷한 부분을 찾기 위하여 식 (11)의 마할라노비스 거리 (mahalanobis distance) 척도를 이용한 정합도를 구한다. 이 과정은 인스턴스를 구성하는 모든 점에 대해 수행된다.

IV. 에너지 최소화 기반 능동형태 모델

능동형태 모델은 표준적인 모델에 기반을 두고 물체의 윤곽선을 찾는 방법이지만 각 점이 독립적으로 이동하는 반면에 능동윤곽선 모델은 표준적인 모델이 없지만 내부 에너지 등을 사용하여 각 점이 유기적으로 연관되어 이동하며 물체의 윤곽선을 추출한다.

능동형태 모델에서는 입술을 구성하는 각 점의 인근 영역을 탐색하여 각 점의 프로파일 모델과 비슷한 프로파일을 갖는 점을 찾는다. 이 과정에서 제약조건 없이 각 점이 독립적으로 이동하기 때문에 그림 2와 같이 실제 입술의 형태와 많은 차이가 나는 결과가 발생한다. 그림 2의 (b)는 (a) 상태에서 각 점이 독립적으로 이동한 결과를 나타낸다.

이 논문에서는 능동형태 모델을 기반으로 능동윤곽선 모델의 에너지 개념을 도입하여 능동형태 모델을 구성하는 각 점이 유기적으로 연관되어 이동하게 하였다. 즉, 식 (12)의 에너지 함수를 정의하고 에너지 함수가 최소화되는 방향으로 능동형태 모델의 점들이 이동하며 객체의 윤곽을 찾아내게 하였다. 각 점이 이동한 후에는 형태계수를 이용하여 능동형태 모델의 표준 모델에 근사하도록 변형하는 과정을 수행한다.

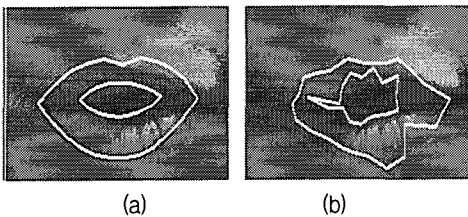


그림 2. 탐색 후 입술 형태
Fig. 2. Lip Shape after search

$$E = \int_0^1 E_{snake}(v(s)) ds \quad (12)$$

$$= \int_0^1 (E_{int}(v(s)) + E_{image}(v(s))) ds$$

$$E_{image}(v(s)) = w_{profile} E_{profile}(v(s)) \quad (13)$$

$$E_{int}(v(s)) = \alpha(s) |\Delta v(s)|^2 + \beta(s) |v_{ss}(s)|^2 \quad (14)$$

$$\Delta v(s) = \Delta v_i - \Delta v_{i-1} \quad (15)$$

외부 에너지로서 능동윤곽선 모델의 에지 강도를 사용하는 대신에 식 (13)과 같이 능동형태 모델의 프로파일 모델의 정합도를 사용하여 각 점이 물체의 경계선으로 이동하게 하는 역할을 한다. 식 (14)의 내부에너지는 두 부분으로 나누어진다. 식 (14)의 두 번째 항은 능동윤곽선 모델에서와 같이 곡선의 휨 정도를 나타내는 변화율 (curvature)로서 능동윤곽선 모델의 윤곽선을 변화율이 작은 쪽으로 끌어당기는 역할을 한다. 식 (14)의 첫 번째 항은 이웃점간 이동 변위를 최소화하는 부분으로 식 (15)와 같이 표현된다.

V. 실험 및 고찰

실험을 위하여 Tulip 1 데이터베이스[12]에 있는 입술 영상들을 사용하였다. 이들은 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 1에서 200개의 영상을, 실험 영상으로 집합 2에서 100개의 영상을 무작위로 추출하여 사용하였다

그림 3은 추출 결과의 예이다. 처음 4개는 1을 발음한 경우, 그 다음 4개는 2를 발음한 경우, 마지막 4개는 3을 발음한 경우에 연속적으로 나타나는 입모양의 윤곽선을 추출한 결과이다. 한 숫자를 발음하는 경우 약 8개의 영상이 존재하는데 하나씩 건너 뛴 영상을 나타내었다. 입을 다문 경우, 입을 조금 벌린 경우와 많이 벌린 경우, 입술을 내민 경우, 치아나 혀가 보이는 경우 등 여러 상황에서 입술 윤곽선을 정확히 추출하였다.

실험에 사용한 영상에 대해 수작업으로 기준이 되는 입술 윤곽선을 정하고 윤곽선 안의 화소가 실험결과로 얻어진 입술 윤곽선 안에 포함되는지를 비교하였다. 표 1은

기존 방법과 제안한 방법에 의한 입술 윤곽선 추출 결과이다. False Negative는 실제 입술에 속하는 화소가 추출결과에서 입술에 속하지 않는 화소로 분류된 평균 비율이며, False Positive는 실제 입술에 속하지 않는 화소가 추출결과에서 입술에 속하는 화소로 분류된 평균 비율이다.

표 1. 실험 결과
Table. 1 Results

| | 제안 방법 | 기존 방법 |
|-------------------|-------|-------|
| False Negative(%) | 4.19 | 6.19 |
| False Positive(%) | 2.59 | 2.98 |
| Error(%) | 6.78 | 9.17 |

기존 방법과 본 논문에서 제안한 방법의 결과를 비교하면 본 논문에서 제안한 방법이 보다 실제 형태에 가깝게 추출하였다. 이는 기존 방법[8, 9]의 경우 각 점이 제약 조건 없이 독립적으로 이동하기 때문에 탐색과정에서 실제 입술의 형태와 많은 차이가 나는 결과가 발생하는데 반하여 본 논문에서 제안한 방법은 연속성과 부드러움에 기반하는 내부에너지를 적용함으로써 보다 실제 입술 형태와 비슷하게 탐색이 이루어지기 때문이다.

대부분의 오류는 입술의 위치가 중앙에서 벗어나거나 입을 벌려서 입술사이로 치아나 혀가 보이는 경우에 발생하였다. 입술 위치에 의한 오류는 탐색의 초기 위치를 간단하게 영상의 중앙으로 하기 때문에 발생하였으며, 치아나 혀가 보이는 경우에 발생하는 오류는 입술, 치아와 혀로 인하여 여러 경계가 나타나기 때문으로 판단된다.

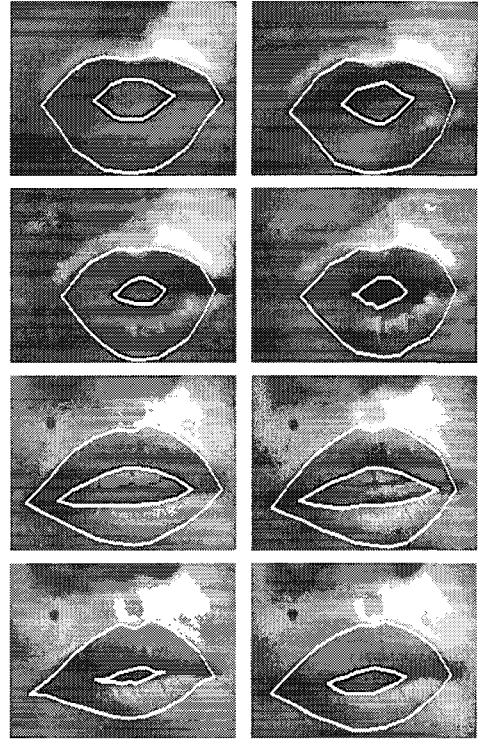


그림 3. 실험 결과
Fig. 3. Results

VI. 결 론

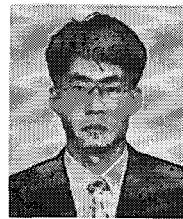
이 논문에서는 능동형태 모델을 개선하여 입술 형태를 효과적으로 추출하는 방법을 제안하였다. 능동형태 모델을 이용하여 물체를 찾는 과정에서 각 점은 독립적으로 탐색하기 때문에 많은 오류를 발생시킨다. 이 논문에서는 능동윤곽선 모델의 에너지 개념을 도입하여 능동형태 모델을 구성하는 각 점이 에너지 함수가 최소화되는 방향으로 유기적으로 연관되어 이동하며 객체의 윤곽을 찾게 하였다.

향후 연구과제로는 입술의 형태적 특징을 사용하여 초기 탐색 위치의 정확성을 높이고 각 점이 탐색하는 과정에서 형태에 관한 보다 전역적인 정보를 이용하는 방법에 대한 연구가 필요하다.

참고문헌

- [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] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] Luetttin 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.
- [9] 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.
- [10] Donna J. Williams and Mubarak Shah, "A Fast Algorithm for Active Contour and Curvature Estimation," CVGIP:Image Understanding, Vol. 55, No. 1, pp. 14-26, 1992.
- [11] M. B. Stegmann, R. Fisker, "On Properties of Active Shape Models," Informatics and Mathematical Modelling, Technical University of Denmark, 2000.
- [12] Movellan J. R., "Visual Speech Recognition with Stochastic Networks," Advances in Neural Information Processing System. Vol. 7, MIT Press Cambridge, 1995.

저자소개



장 경 식(Kyung-Shik Jang)

연세대학교 공학사 (전자공학)
연세대학교 공학석사 (전자공학)
연세대학교 공학박사 (전자공학과)

1998년 - 현재 동의대학교 멀티미디어공학과 부교수
※ 관심분야 : 컴퓨터 비전, 패턴 인식, 인공지능 등