

동적 모델을 이용한 얼굴 영상에서의 관심 영역 추출

이형일⁰ 김경환
서강대학교 전자공학과
(regnius, gkim)@sogang.ac.kr

Region-of-Interest Detection from a Facial Image Using Active Model

Hyungil Lee⁰ Gyeonghwan Kim
Dept. of Electronic Engineering, Sogang University

요약

본 논문에서는 얼굴 인식 시스템에서 정면 얼굴 영상의 관심 영역을 추출하는 효율적인 방법을 소개한다. 얼굴 인식 시스템은 얼굴 요소의 특징을 이용하여 자동으로 얼굴을 구별하는 시스템이며, 얼굴 요소로는 눈, 코, 입과 눈썹을 주로 사용한다. 본 논문에서는 동적 모델을 이용하여 눈과 입을 관심 영역으로 하여 이 영역을 세 단계로 나누어 추출한다. 첫 번째로 전체 얼굴 모델을 이용하여 similarity 변환을 적용하여 얼굴의 대략적인 위치를 찾는다. 두 번째 단계에서는 얼굴 근처에서 각각의 눈, 입 모델을 비선형 변환을 적용하여 정확한 눈과 입을 찾는다. 최종 단계에서는 이렇게 맞춘 모델로부터 전체 모델을 변형시킨 후에 변형 전과 후의 적합성을 판단하여 최종 위치를 정한다. 제안한 알고리즘을 130명의 영상에 대하여 적용한 결과 눈과 입을 정확하게 추출한 경우는 120명이고, 입을 정확히 추출한 경우는 119명이었다. 본 논문에서 제안하는 관심 영역 추출 방법은 일반적인 모델 방법에 특정 목적에 적합한 모델을 혼합한 방법으로 일반적인 모델만을 적용한 방법과 프로젝션 분석 등의 특정 목적만을 위한 방법보다 좋은 결과를 얻을 수 있었다.

1 서론

얼굴에 관련된 연구는 크게 특징 추출에 관한 연구[2, 3, 5, 6], 영상을 해석하여 사람을 식별하는 연구[1]와 표정을 알아내는 연구[8]로 나눌 수 있다. 얼굴 특징 추출에 관한 연구는 전체 얼굴 추출과 얼굴의 주요소 추출로 나누어 생각할 수 있다. 이러한 얼굴 특징 추출은 일반적으로 얼굴 인식과 표정 인식의 전단계의 과정이다. 따라서, 얼굴 특징 추출은 얼굴에 관련된 연구에서 중요한 부분을 차지한다.

프로젝션 분석을 이용한 방법[5]은 x축과 y축 프로젝션을 사용하여 각 요소의 특성을 해석하여 영역 분할을 하는 방법이다. 이 방법은 배경이 단순해야 하며, 회전에 대하여 민감한 특성을 가지고 있다. 따라서 이 방법은 이러한 조건이 만족되어야만 좋은 결과를 도출 한다. 가변 템플릿(deformable template)을 이용한 방법은 similarity 변환을 이용하여 윤곽선을 추출하는 방법으로, 결과 영상이 일정한 모양을 가지는 단점이 있다. 영상의 에너지를 이용한 방법[6]은 윤곽선을 최소 에너지 상태로 정의하고, 에너지 함수는 스플라인(spline) 곡선을 사용하고 있다. 이 방법은 윤곽선의 초기 위치로부터 에너지를 최소화하는 과정을 통하여 동적 윤곽 모델(active contour model; snakes)[10]이 물체의 윤곽선으로 수렴하도록 한다. 이 방법은 형태가 변하는 물체에 대하여 능동적으로 추출할 수 있는 장점이 있으나, 이미지의 최소화 방향을 찾을 수 있는 초기 함수가 주어져야 하며, snake point 주위의 잡음에 민감할 수 있는 문제점이 있다. 이를 보완한 방법이 ASM(active shape model)[3]이다. 이는 추출하고자 하는 모양을 PDM(point distribution model)이라고 불리는 모델로 정의하여 이 모

델과 비슷한 형태를 가지는 물체를 에너지 함수를 이용하여 영상에서 찾는 방법이다. 이 방법은 학습된 모델의 변화량에 근거하여 모델이 능동적으로 변하는 장점이 있지만, 초기 위치에 대한 문제는 여전히 남아있다. 따라서 본 논문에서는 동적 모델을 이용하여 전역적인 검색으로 초기 위치를 찾으며 지역적 검색으로 정확한 관심 영역을 추출하는 방법을 제안한다.

2 동적 모델

PDM은 모양(shape)을 표현하는 방법 중 하나이며, 모양을 대표할 수 있는 경계의 point들과 학습되는 모델의 변화의 양으로 구성된다. 따라서 이것은 모델들의 경계선과 모양의 변화를 반영할 수 있다. 본 논문에서 사용된 모델은 그림 1과 같다.

3 영역 분할을 위한 알고리즘

얼굴 영상에서 얼굴의 주요소를 추출하기 위한 순서도는 그림 2와 같다. 입력 영상에 미리 만들어진 전체 얼굴 PDM을 적용하여 similarity 변환[9]을 이용한 탐색방법으로 가장 적합한 위치를 찾는다. 전역적으로 얼굴을 추출한 후 부분적으로 눈과 입을 찾는다. 눈(정확하게는 홍채)을 추출하기 위해서 만들어진 모델을 선형 변환을 이용하여 눈을 추출하며 좌우 눈의 위치를 고려하여 전체 모델을 변형시킨다. 변형 전과 후의 모델에서 각각 입을 찾는 알고리즘을 적용하여 적합한 것을 택한다. 위와 같은 방법으로 눈과 입의 위치와 윤곽선을 얻게 된다.

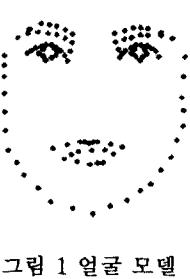


그림 1 얼굴 모델

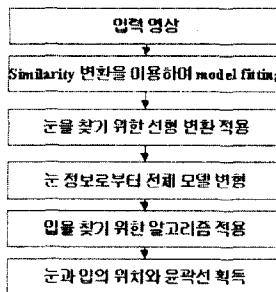


그림 2 영상에서 얼굴의 주요 요소를 찾기 위한 순서도

$$\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \\ \vdots & \vdots \end{pmatrix} \rightarrow \begin{pmatrix} M_{11} \\ M_{21} \\ \vdots \\ M_{n1} \\ \vdots \end{pmatrix} \quad (1)$$

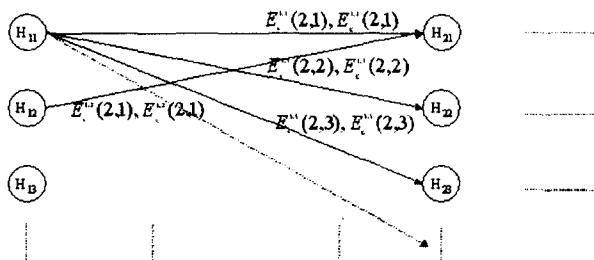


그림 3 H matrix

3.1 에너지 함수 정의

본 논문에서 사용한 에너지 함수는 식 (2)와 같고, 일반적인 snakes에서 사용된 곡률(curvature) 에너지(E_{curv})와 에지 에너지(E_{edge})를 사용하며 추가로 gray level의 상관성을 고려한 에너지($E_{\text{correlation}}$)를 사용한다.

$E_{\text{correlation}}$ 은 PDM을 만들 때 point주위의 값을 특징으로 하여 얻은 gray 값들의 샘플과 검색시의 PDM의 point주위의 값과의 상관 계수를 의미한다. E_{curv} 는 모델의 곡률값과 검색시 이동되는 곡률값과 차이의 절대값으로 하여 구한다. E_{edge} 는 소벨(Sobel) 연산자를 사용하여 얻은 에지 영상에서 현재 point와 다음 point를 연결한 선 위의 3×3 마스크의 평균 밝기 값으로 구한다. 위의 각 에너지들은 '0'과 '1' 사이의 값을 가진다. 식 (2)의 계수 α, β, γ 는 실험에 의하여 결정한다. PDM의 각 point들의 이웃하는 행렬을 식 (1)과 같이 벡터로 변환한다. 이 벡터는 그림 3의 열이 되는 것이다. 즉 $H_{1,1}, H_{1,2}, \dots$ 이다. 현재 point와 다음 point와의 연결이 그림 3의 행으로 나타난다. 이것을 H 행렬로 정의하며 식 (3)과 같이 구한다. 이 H 행렬의 최소값을 동적 프로그래밍(dynamic programming)[4]을 사용하여 찾으면 PDM의 point들의 에너지가 최소가 되는

위치다.

$$H_{i,j} = \alpha * \text{curvature energy}(E_i)$$

$$H_{i,j} = H_{i,i} + \beta * \text{edge energy}(E_i) + \gamma * \text{gray correlation energy}(E_i) \quad (3)$$

$$H_{i,j} = \sum_{k=1}^K H_{i,k} + \sum_{k=1}^K (\alpha E^{(i)}(j, k) + \beta E^{(i)}(j, k) + \gamma E^{(i)}(j, k))$$

3.2 비선형 변환

이 변환은 눈과 입에 적용될 특정한 방법으로 지역적인 템색을 하며, 본 논문에서는 총체적으로 비선형 변환이라 정의한다.

3.2.1 눈 (홍채)

홍채를 위한 기본적인 모델은 원이다. 이것을 찾기 위한 템색 방법은 similarity 변환을 그대로 사용한다. X와 Y축 크기 을을 변화시키면 타원, 사각형으로 모델이 변할 수 있다. 홍채를 찾기 위해서 회전은 고려하지 않아도 된다. 홍채는 에지 영상으로 찾을 수 있다. 눈을 찾으면 두 눈의 위치에 근거하여 전체 얼굴의 크기와 회전 정도를 변경하게 된다.

3.2.2 입

입 추출을 위한 초기 위치는 similarity 변환으로 찾은 위치로 한다. 경계 상자는 그림 1의 입 경계의 1.5배로 한다. 이 영역에서 선형 예측[7]으로 기울기와 절편을 구한다. 이 기울기와 절편에 맞추어 입을 위한 모델을 이동시킨 후 현재 기울기와 이전의 절편이 같을 때(혹은 그 값들의 차이가 일정한 값보다 작을 때)까지 위 과정을 반복한다. 또한, 눈에 근거한 모델 변경 전후의 위치 모두에서 위의 과정을 수행한다. 이 두 위치에서의 energy가 작은 쪽을 택하여 최종 위치로 한다.

4 실험 결과

얼굴의 주요소인 눈과 입 추출에 대한 실험을 하였다. 실험 환경은 Pentium III PC에서 하였으며, 사용된 영상은 Purdue 대학의 얼굴 영상을 가지고 실험하였다. 실험에서 사용한 similarity 변환에서 크기의 변화를 영상의 크기대비 얼굴의 크기가 모델 크기에서 $\pm 20\%$ 로 하였다. 실험에 사용된 영상은 정면을 보고, 눈을 뜬 상태의 영상이며 입은 다물고 있거나 약간 벌리고 있는 영상을 사용하였다. 조명 조건은 아주 밝거나 아주 흐리지 않은 적절한 상태의 영상이다. 또한, 배경이 복잡하지 않은 영상을 대상으로 하였다.

전역적인 검색으로 찾을 결과는 그림 4와 같고, 이를 이용하여 부분적인 모델을 적용하여 홍채를 찾은 결과는 그림 5이다. 이 때, 초기값은 앞 단의 결과를 이용한다. 입을 찾기 위한 알고리즘을 적용한 결과는 그림 6이다. 그림에서 보인 초기값은 모델 변형전의 위치이고, 결과 영상은 모델 변형전과 후의 결과 중 에너지가 최소가 되는 위치의 결과이다.

Purdue 대학의 얼굴 영상 자료를 가지고 입의 위치와 좌우 눈을 추출한 실험한 결과는 표 1과 같다. 남자 자료는 74개이고, 여자 자료가 56개다. 이 중에서 안경을 착용한 영상이 35개이고 미착용이 95개이다. 남녀에 따라서 결과가 크게 다르지는 않다. 그러나, 안경 착용 여부에 대해서는 표 2와 같이 큰 차이를 보인다.

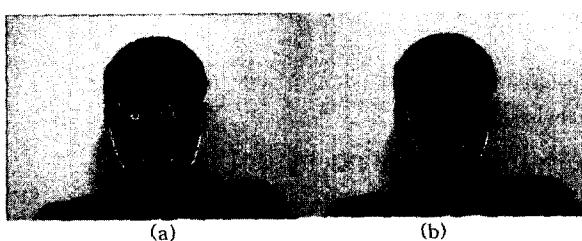


그림 4 Similarity 변환을 적용한 결과 (a) 초기 위치 (b) 결과



그림 5 눈을 찾기 위한 similarity 변환 적용한 결과 (a) 초기 위치 (b) 결과



그림 6 입을 찾기 위한 알고리즘을 적용한 결과 (a) 초기 위치 (b) 결과

그 이유는 안경에 반사되는 빛의 양이 많은 경우 홍채에서 애지 성분이 적어지고, gray level값이 커져서 잘못된 위치를 검색하게 된다. 입 추출 결과에서 에러가 발생하는 이유는 콧수염 때문이다. 콧수염이 길고 많으면 이곳을 입으로 판단하는 경우가 많이 생긴다.

영상에서 얼굴이 좌측 혹은 우측에 있다고 하더라도 위와 같은 성능을 보인다. 또한, 얼굴이 회전되어 있더라도 $\pm 30^\circ$ 이내에서는 위와 같은 성능을 보인다.

표 1 좌우 눈과 입의 윤곽선 추출 결과

	총 수	추출 수	에러 수	에러율(%)
좌측 눈	130	120	10	7.69
우측 눈	130	123	7	5.38
입	130	119	11	8.46

표 2 안경 착용 여부에 대한 눈 추출 결과 (착용 35, 미착용 95)

안경 착용 여부	좌측 눈		우측 눈	
	에러 수	에러율(%)	에러 수	에러율(%)
착용	9	25.71	3	8.57
미착용	1	1.05	4	4.21

5 결론

지금까지 얼굴 영상에서 관심 영역을 추출하기 위한 방법을 살펴보았다. 얼굴에서의 관심 영역 추출은 사람마다 그 특징이 다르기 때문에 어렵다. 본 논문에서는 이러한 변화를 고려하여 동적으로 변할 수 있는 모델을 이용하여 그 특징을 반영할 수 있는 추출 방법을 제안하였다.

기존의 방법인 가변 템플릿을 사용하는 경우 결과가 일정한 모양을 가지는 단점을 지역적으로 정밀한 부분을 찾으면서 모델을 변형시키는 방법을 사용하여 해결하였다. ASM은 전체적으로는 좋은 성능을 보이지만, 얼굴의 주요소가 원하는 결과(정확한 위치)와 부분적으로 떨어져 있는 단점이 있는데 제안한 방법은 전역적인 검색 뿐만 아니라 지역적인 검색을 통하여 보다 정확한 얼굴의 주요소를 추출하였다.

6 참고문헌

- [1] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15(10), pp. 1042-1052, 1993.
- [2] A. Yuille, P. Hallinan and D. S. Cohen, "Feature extraction from faces using deformable templates," *International Journal of Computer Vision*, 8(2), pp. 99-111, 1992.
- [3] T. F. Cootes and C. J. Taylor, "Active Shape Models - 'Smart Snakes,'" in *Proc. British Machine Vision Conference*, pp. 266-275, 1992.
- [4] D. K. Smith, *Dynamic Programming: a practical introduction*, Ellis Horwood, 1991.
- [5] G. Garcia and C. Vicente, "Face Detection on Still Images Using HIT Maps," *Third International Conference AVBPA 2001*, Sweden, pp. 102-107, June 2001.
- [6] T. Yokoyama, Y. Yagi and M. Yashida, "Facial Contour Extraction Model," *3rd. International Conference on Face & Gesture Recognition*, Nara, Japan, pp. 254-259, 1998.
- [7] H. Stark and J. W. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*, Prentice Hall, 1994.
- [8] I. A. Essa and A. P. Pentland, "Facial Expression Recognition using a Dynamic Model and Motion Energy," *IEEE ICCV' 95*, pp. 360-367, 1995.
- [9] I. L. Dryden and K. V. Mardia, *Statistical Shape Analysis*, John Wiley & Sons, 1998.
- [10] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," *International Journal of Computer Vision*, 1(4), pp. 321-331, 1987.