

국부적 요소 모델을 이용한 얼굴 특징점 추출

Face Landmark Detection Using Local Component Model

김대환¹, 전승선¹, 오두식¹, 조성원¹, 김재민¹, 김상훈², 정선태²

¹ 서울시 마포구 홍익대학교 지능정보처리 연구실
E-mail: kdh@hongik.ac.kr

² 서울시 동작구 숭실대학교 정보통신 전자공학부
E-mail: {hosuwa,cst}@ssu.ac.kr

요 약

객체의 특징점을 추출할 때, 일반적으로 모델 기반 접근을 사용한다. 본 논문에서는 이러한 모델 기반 특징점 추출 알고리즘으로 PCA를 근간으로 하는 Active Appearance Model을 이용하는 데, 기존의 AAM 알고리즘은 모든 특징점을 하나의 군집으로 기준하여 PCA를 수행하지만 본 논문에서는 이것을 각 주요 부위별 학습 모델로 분리하여 수행한다. 그리고 이러한 모델에서 특징점을 찾을 때, 발산하는 문제에 빠지지 않기 위한 방법을 제시한다. 제시한 방법의 모델을 이용하여 실험 할 경우의 결과와 이를 통한 개별 모델의 특성에 대하여 파악한 결과를 제시한다.

Key Words : Face Feature Detection, Active Appearance Model, PCA, Component Modeling

1. 서 론

영상에서 객체를 추출하고, 특징점을 검출하는 방법은 여러 가지가 있다. 동영상 신호는 시간에 따라 각 프레임간의 상관관계를 이용하거나 한 프레임 내부에서 특징을 파악하여 응용되는데, 날씨 모델링이나 실시간 감시 등에 쓰이고 있다. 최근 카메라를 장착한 첨단 자동차에서도 활용되고 있다. 정지 영상을 분석하여 객체를 추출하는 방법은 1개의 영상을 입력 받아 그것을 분석하여 사용하는 인증 시스템, 보안 시스템에서 사용되고 있다. 동영상의 경우 실시간 처리 및 저장에 따른 비용과 연산 속도 문제가 있지만 정지영상은 동영상 처리보다는 빠르고 비용이 적게 든다는 장점을 가지고 있기 때문에 널리 쓰이고 있다.

정지 영상을 분석하여 특징점을 찾는 방법으로는 입력된 영상에서 실시간으로 특징점을 찾아내는 방법과 모델을 만들어 놓은 후, 그 모델에 근거하여 특징점을 찾아내는 방법이 있다. 모델 근거 특징점 검출의 경우 통계적 확률을 이용하게 되는데, 이는 좀더 정확하고 빠르게 찾기 위한 방법 중 하나로, 온라인에서의 부하를 오프라인으로 전이함으로써 온라인 연산량을 줄여줄 수 있다.

이러한 모델 학습 방식은 일반적으로 영상의 특징점을 미리 파악 해 놓은 뒤, 이 특징점들

의 지형적, 주파수적 특성을 사용하여 정합해 나간다. 이 때, 차원 축소와 정보의 효율화를 위해 PCA(Principal Component Analysis)[1] 학습을 주로 사용한다. 그리고 특징점과 PCA를 결합한 AAM(Active Appearance Model)[2,3,4]을 사용하는데, 이 때 모델을 구성할 때 모델에 사용되는 특징점은 하나의 PCA로서 사용되기 때문에 이들이 유기적으로 연결되어 있다. 하지만 이러한 모델의 경우 일부 특징점이 맞지 않는 경우가 발생할 수 있다. 모델의 일정 부분 이상의 에러가 더 이상 없을 경우 최적점으로 간주[5]하여 멈추기 때문에 모든 특징점을 한꺼번에 정합할 수 있는 확률이 낮아진다는 것이다.

본 논문에서는 이를 피하기 위하여 전체적인 모델을 대략적으로 구성한 것과, 각 부위별 모델을 정교하게 구성해 놓는 것을 교체 사용함으로써, 좀 더 정확한 최적점 결과에 다다른 방식을 제안한다.

2. 기존 방법

본 논문은 AAM을 이용한 얼굴 특징점 추출을 기반으로 한다. 따라서 기존의 얼굴의 다양한 특징점을 찾는 방법에 대하여 서술한다. AAM은 PCA의 특성을 대부분 가지고 있다. 이러한 학습 모델에 대한 구성 방법과 그 특

정, 그리고 기존 방식인 전체 모델에 대한 한계를 찾는다. 그리고 이를 해결하기 위하여 본 논문에서는 어떠한 방식으로 바뀌어야 할지 제시한다.

2.1 기존 알고리즘의 학습 모델 구성

기존 전체 얼굴 윤곽의 특징점의 개수를 v 개, 모델링에 사용되는 얼굴 모델 이미지의 개수는 M 개라고 할 때, 각 얼굴 윤곽 AAM의 모양(shape) X 는 v 개 특징점 좌표 세트로 정의된다.

$$X = (x_1, y_1, x_2, y_2, \dots, x_v, y_v)^t \quad (1)$$

M 개의 모델 이미지 모양의 특징점에 대해 정규화[6] 후, PCA하여, 평균 모양 S_0 , n 개의 모양 고유벡터 $S_i (i = 1, \dots, n)$ 을 구할 수 있으며, 따라서 모델 얼굴 또는 새로운 입력 이미지의 얼굴의 모양 X 는 다음과 같이 이들의 선형 결합으로 나타낼 수 있다.

$$X = S_0 + \sum_{i=1}^n S_i b_i = S_0 + P_s \vec{b}_s$$

$$(P_s = [S_1, \dots, S_n], \vec{b}_s = (b_1, \dots, b_n)^t) \quad (2)$$

또한, 명암도(texture)의 통계적 모델 구축을 위해 각 모델의 특징점들이 평균 모양 S_0 에 정합되도록 들로네 삼각법 알고리즘(DTA)[7]을 이용하여 모델 이미지를 워핑한다. 이후, 형태 정규화 된 모델 이미지로부터 그레이 레벨 정보를 샘플링 하여 $g_{im} (i = 1, \dots, M)$ 을 얻는다. 다시 조명 변화의 영향을 줄이기 위해, 다음과 같이 정규화 한다.

$$g = (g_{im} - \beta) / \alpha, \quad \alpha: scaling, \beta: offset \quad (3)$$

모양 벡터에서와 같이 명암도에 대해서도 PCA를 적용하여 각 모델 얼굴 또는 새로운 입력 이미지의 얼굴의 명암도 벡터 g 를 다음과 같이 평균 정규화 된 명암도 벡터 T_0 와 k 개의 고유 명암도 모드 벡터 $T_i (i = 1, \dots, k)$ 의 선형 모델을 구할 수 있다.

$$g = T_0 + \sum_{i=1}^k T_i c_i = T_0 + P_g \vec{b}_g$$

$$P_g = [T_1, \dots, T_k], \vec{b}_g = (c_1, \dots, c_k)^t \quad (4)$$

식 (2)와 (4)는 통합하여 다음과 같이 공통의 파라미터 벡터 \vec{c} 로 기술할 수 있다.

$$X = S_0 + Q_s \vec{c}$$

$$g = T_0 + Q_g \vec{c} \quad (5)$$

식 (5)와 같이 객체, 즉 얼굴의 모양과 명암도를 결합하여 모델링한 것을 결합 AAM이라 한다.

2.2 기존 알고리즘의 정합 과정

AAM 모델 (5)를 새로운 이미지에 정합하는 것, 즉 AAM 모델을 적용하여 새로운 이미지에서 얼굴 윤곽을 찾아내는 것은 최적화 문제이다. AAM에서는 (5)로 모델링된 얼굴 이미지의 명암도와 새롭게 입력된 얼굴 이미지의 명암도와 에러의 제곱을 최소화하도록 하는 모델 (5)의 파라미터 벡터 \vec{c} 를 찾고 이때의 파라미터 벡터 \vec{c} 를 가지고 (5)로 표현되는 얼굴 윤곽을 새로 입력된 이미지에서 찾고자 하는 얼굴 윤곽으로 판정한다.

에러 자승을 최소화 하는 파라미터 벡터 \vec{c} 를 어떻게 찾아 나아가느냐 하는 것과 파라미터 벡터를 최적화 절차 수행에 필요한 초기값 설정은 각 AAM 구현 알고리즘에 달려 있다. 초기값 설정은 보통 평균 형태 즉, $\vec{c} = 0$ 을 취한다.

3. 제안하는 방법

기존 방식의 경우에서와 같은 단일 정합은 모든 특징점이 PCA 당시의 편차 정도에 따라 결과는 이것에 한정되게 되어 원하는 부분에 미치지 못하고 정합이 종료되는 경우가 있다. 이러한 결점을 보완하기 위하여 각기 다른 부위에 부위별 모델을 구성하고 각 부위에 해당하는 모델을 다시 사용하여 정교화한다. 특히 정교화 부분에 있어서 발산을 방지하고자 주변 픽셀 정보를 이용하는 배경 명암도 제한 방법을 제시한다.

3.1 제안하는 학습 모델

본 논문이 제시하는 학습 모델은 다음과 같은 과정을 거친다. 먼저 얼굴의 다양한 표정이거나 외부 조명에 관한 요소에 명암도가 유기적으로 균집되어 변하는 부분, 즉 눈썹, 눈, 코, 입을 각각의 구성 요소(component)로 정한다. 각 구성요소는 식(1)과 같은 형태로 각각 학습시킬 수 있다. 따라서 각 특징점은 $X_{cycbrow}$, X_{cyc} , X_{nose} , X_{mouth} , X_{chin} 등으로 나타낼 수 있고, 이를 각각 모델화 할 수 있다. 각 요소의 특성을 분석해 보면, 눈썹이나 입술의 경우 요소 주변 피부는 명암도 정보가 확연하게 다르다. 하지만 눈이나 코의 경우 명암도 변화가

크지 않기 때문에 경계값이 모호하다. 이 경우 특징점을 명암 경사도가 높은 부분으로 단순히 결정한 경우 정합 시 발산하는 문제가 생길 수 있다.

따라서 본 논문에서는 특징점과 특징점 주변의 형태를 조금 더 합리적으로 사용하는 확장 특징점 정보를 사용한다. 이러한 확장 특징점을 사용할 경우, 요소의 특징점 뿐만 아니라 알고있는 그 주변 정보를 제한 사항으로 활용함으로써 더욱 효율적으로 발산을 막는다는 것이다. 여기서 확장 특징점이라 함은 원하는 요소에 대한 특징점을 먼저 선정한 후, 그 특징점들의 노말 벡터 방향으로 몇 픽셀 바깥으로 새로운 보조 특징점 들을 선정한다. 이 때, 보조 특징점은 외부 배경의 인접 거리는 5pixel 정도로 제한하였다.

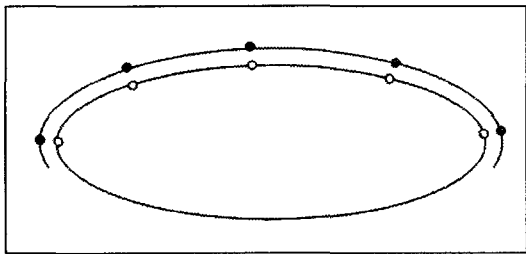


그림 1. 특징점(내부), 자동확장 특징점(외부)

그림1 에서와 같이 학습시 사용자가 특징점을 선별하여 그 특징점을 학습시키기 위하여 점을 찍으면(그림 1의 안쪽 점들), 그 점 정보를 이용하여 자동으로 외부 점(그림 1의 바깥쪽 점들)을 찍고, 이 점들을 이용하여 학습 모델을 구축하는 것이다.

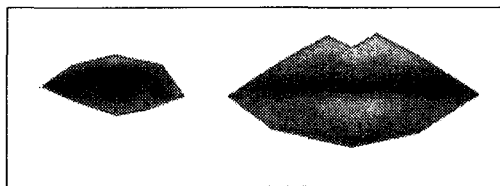


그림 2. 왼쪽 눈과 입 학습 결과 예제

그림 2는 실제 제안한 알고리즘으로 학습시킨 모델의 예제이다. 학습된 모델을 보면 학습된 각 요소에 주변 피부들도 약간 보인다. 이는 주변 정보들이 함께 학습 되어있음을 알 수 있다. 이렇게 학습된 모델의 경우 정합시에 학습된 특징점들을 모두 사용하여 안정적인 검출을 하게 되고, 검출 후에는 안쪽의 특징점들만을 사용하는 것이다.

마지막으로, 이 특징점들의 초기점을 주기 위한 전체 모델을 구성한다. 이 때, 전체 스케일을 이용하지 않은 저해상도 모델로 구축한

다. 저해상도에서 구축한 전체 얼굴의 학습 모델은 전체 모델보다 빨리 정합되고, 이 정합된 결과가 국소 부위 학습 모델의 초기점으로 부여되고, 여기에서 정교화 시키기 때문이다.

3.2 학습된 모델의 적용

구성된 각 요소 학습 모델은 기본적인 AAM 정합 과정을 따른다. 단, 제안하는 방법은 초기점이 얼굴 요소의 각 부위가 되기 때문에 이 초기점을 잘 정하는 것이 중요하다. 일반적으로 통계적 방식 의한 단일 초기점은 얼굴의 회전, 각 사람에 따라 민감하다는 단점이 있다. 이를 해결하기 위해 기존 알고리즘인 전체 AAM을 사용하는데, 시간을 줄이기 위해 저해상도 AAM으로 대략적인 초기점을 잡은 후, 그 지점에서 각 요소별 학습 모델을 적용하여 정교화 하는 것이다. 이는 저해상도에서 고해상도로 초기점을 넘겨주는 다중 스케일 방식과도 같은 접근 방식이다.

따라서, 본 논문에서는 이와 같은 방식으로 학습된 모델을 정합시키며, 최대한 빨리 최적점에 도달하기 위하여 저해상도 전체AAM을 사용하는 것을 제안하는 것이다.

4. 실험 결과 및 향후 과제

4.1 실험 환경

제안된 국부 요소 기반 얼굴 윤곽 검출 알고리즘을 평가하기 위해 본 논문에서는 IMM 얼굴 데이터베이스[8]를 이용하였다.

사용한 얼굴 데이터베이스인 IMM 얼굴 데이터베이스는 40명의 사람이 6개의 각기 다른 자세 또는 표정 또는 조명을 갖는 240 장의 이미지로 구성된다. 각 이미지는 640×480 해상도의 JPEG 포맷이다. IMM 얼굴 데이터베이스의 일부 샘플 이미지가 그림 3에 나타나 있다.



그림 3. IMM 얼굴 DB의 샘플 이미지들

모델의 구성과 정합 평가 실험을 하기 위하여 무작위 추출 학습 실험을 하였다. 40명중에 무작위로 4명을 추출하여 4×6개의 이미지로 제안된 알고리즘 학습 모델을 구성하였다. 이는 총 40명 전체의 10%에 해당하는 학습 모델로서 기존 방법의 한계를 보여주기 위하여 학습 영상의 수를 적게 정하였다.

갤러리 영상은 학습 모델에 사용된 영상을 포함하여 총 240장에 대하여 실험하였다.

4.2 실험 및 결과

실험의 성패는 미리 특징점들을 수동으로 찍어준 점과 실험을 통해 나온 결과의 각 특징점들의 좌표 거리가 어느 정도 근접하였는지 유클리디안 거리를 구하여 평균이 3픽셀 이내 일 경우에 성공[9]으로 간주 하였다.

표 1. 실험 결과

사용된 방법	입력 영상	성공	성공확률
기존 방식	240	218	90.8%
제안한 방식	240	226	94.1%

성공확률에 있어서는 약 3.6%의 성능 향상이 있었다. 실험에서 실패의 경우 학습 모델을 어떻게 구성하느냐에 따라 차이가 났는데, 일반적인 PCA의 특성상 학습 영상들이 큰 편차를 가질 때 가장 다양한 얼굴을 수용할 수 있다는 점에 착안하면 무작위로 뽑은 학습 모델이 비슷한 얼굴일 경우 전체적인 실패 확률이 높아졌다. 하지만 기존 방식의 문제점 이었던 발산의 경우 제안한 방식이 좀 더 안정적이고 발산을 차단하는 역할을 하였다는 것을 알 수 있다.

4.3 향후 과제

본 논문이 제시한 방법은 기존 방식보다 적은 얼굴 부위를 사용하면서도 각 특징점을 정교하게 찾을 수 있다는 장점이 있었다. 하지만 이 경우에 각 요소들이 특징점 바깥으로 발산하는 경우가 있기 때문에 이를 방지하기 위하여 확장 특징점을 사용하였다. 이 확장 특징법은 지정된 특징점으로 부터 노말 방향의 정보를 이용하는 것인데, 이러한 방법대신 다른 방법으로 특징점 발산을 방지할 수 있다면, 학습 모델에 대한 부하가 줄어들 수 있을 것이다. 즉 향후에는 강제적으로 외부요인을 학습 시키지 않고 정합 과정에서 제한을 두는 방법이 연구되어야 할 것이다.

참 고 문 헌

[1] F. Keinosuke, "Introduction to Statistical Pattern Recognition," Elsevier, 1990.
 [2] G. Edwards, C. J. Taylor, and T. F. Cootes, "Interpreting Face Images using Active Appearance Models," in Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition, pp. 300-30, 1998.
 [3] T. F. Cootes, G. J. Edwards, C. J.

Taylor, H. Burkhardt, and B. Neuman, "Active Appearance Models," in Proc. Eur. Conf. Computer Vision, vol. 2, pp. 484 - 498, 1998.

[4] T. F. Cootes, D. J. Edwards, and S. J. Taylor, "Active Appearance Models," IEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 6, pp. 681 - 685, Jun. 2001.

[5] Stephen Boyd and Lieven Vandenberghe, Convex Optimization, Cambridge University Press.

[6] J. C. Gower, "Generalized Procrustes Analysis," Psychometrika, 40:33--51, 1975.

[7] D. T. Lee and B. J. Schachter, "Two Algorithms for Constructing a Delaunay Triangulation," Int. J. Computer Information Sci. 9, pp.219-242, 1980.

[8] IMM face db, <http://www2.imm.dtu.dk/~aam/>

[9] O. Jesorsky, K. Kirchberg and R. Frischholz, "Robust Face Detection Using the Hausdorff Distance," In: J. Bigun, F. Smeraldi Eds. Lecture Notes in Computer Science 2091, Berlin: Springer, pp.90-95, 2001.