통계학적 학습을 이용한 머리와 어깨선의 위치 찾기

정회원 권 무 식*

Localizing Head and Shoulder Line Using Statistical Learning

Musik Kwon* ***Regular Member***

요 약

영상에서 사람의 머리위치를 찾는 문제에 있어서 어깨선 정보를 이용하는 것은 아주 유용하다. 영상에서 머리 외곽선과 어깨선의 형태는 일정한 변형을 유지하면서 같이 움직이므로 이를 ASM(Active Shape Model) 기법을 사용해서 통계적으로 모델링 할 수 있다. 그러나 ASM 모델은 국부적인 에지나 그래디언트에 의존하므로 배경 에지나 클러터 성분에 민감하다. 한편 AAM(Active Appearance Model) 모델은 텍스쳐 등을 이용하지만, 사람의 피부색, 머리색깔, 옷 색깔등의 차이로 인해서 통계적인 학습방법을 쓰기가 어렵고, 전체 비디오에서 외모(Appearance)가 시간적으로 변한다. 따라서, 본 논문에서는 외모(Apperance) 모델을 변화에 따라 바꾸는 대신, 영상의 각 화소를 머리, 어깨, 배경으로 구분하는 분별적 외모 모델(discriminative appearance)를 사용한다. 실험을 통해서 제안된 방법이 기존의 기법에 비해서 포즈변화와 가려짐, 조명의 변화 등에 강인함을 보여준다. 또한 제안된 기법은 실시간으로 작동하는 장점 또한 가진다.

Key Words : ASM(Active Shape Model), Discriminative Appearance, Ω-shape.

ABSTRACT

Associating the shoulder line with head location of the human body is useful in verifying, localizing and tracking persons in an image. Since the head line and the shoulder line, what we call Ω-shape, move together in a consistent way within a limited range of deformation, we can build a statistical shape model using Active Shape Model (ASM). However, when the conventional ASM is applied to Ω-shape fitting, it is very sensitive to background edges and clutter because it relies only on the local edge or gradient. Even though appearance is a good alternative feature for matching the target object to image, it is difficult to learn the appearance of the Ω-shape because of the significant difference between people's skin, hair and clothes, and because appearance does not remain the same throughout the entire video. Therefore, instead of learning appearance or updating appearance as it changes, we model the discriminative appearance where each pixel is classified into head, torso and background classes, and update the classifier to obtain the appropriate discriminative appearance in the current frame. Accordingly, we make use of two features in fitting Ω-shape, edge gradient which is used for localization, and discriminative appearance which contributes to stability of the tracker. The simulation results show that the proposed method is very robust to pose change, occlusion, and illumination change in tracking the head and shoulder line of people. Another advantage is that the proposed method operates in real time.

※ 이 논문은 학술진흥재단의 해외 Post-doc. 연수지원에 의하여 연구되었으며, 카네기멜론대학교의 T.Kanade 교수와 공동연구됨.
* 삼성전자 정보통신총괄 (musik.kwon@samsung.com)
논문번호 : KICS2006-09-378, 접수일자 : 2006년 9월 11일, 최종논문접수일자일자 : 2007년 1월 3일

Ⅰ. 서 론

영상에서 외곽선 모양이 통계적으로 변하는 물체의 세부적인 위치를 확인하고 이를 추적하는 방법으로 널리 사용되는 방법이 ASM (Active Shape Model) 모델[1]과 AAM (Active Appearance Model) 모델[2]이다. ASM 모델이 확률적으로 변화하는 형태에 랜드마크를 정합시키는 방법으로 국부적인 에지의 그래디언트를 이용하는 것에 반해서, AAM 모델은 그래디언트 성분 뿐 아니라 텍스쳐 성분의 정합도 동시에 수행한다. ASM 기법은 물체의 외곽선 정보, 즉 랜드마크 주변의 그래디언트 성분이 크다는 사실을 이용하기 때문에, 일반 외부환경에서 촬영된 영상에서 그래디언트 성분이 비교적 큰 배경의 에지나 클러터 성분이 있는 경우, 잡음에 민감한 문제가 발생한다. 따라서, 원하는 물체에 수렴하지 못하는 경우가 발생하거나 수렴영역이 아주 제한적이다. 한편, AAM 모델은 형상정보와 텍스쳐 정보를 동시에 포함하는 외모(appearance)의 학습을 수행하는데, 일반적으로 전체 비디오를 통해서 외모가 일정하지 않은 문제점이 있다. 또 다양한 사람의 피부색, 조명 등에 맞춰서 학습하기가 쉽지 않거나, 각 피부색이나 조명의 상황 별로 따로 학습시킨 후 테스트 영상이 주어졌을 때, 어떤 종류의 학습상황에 부합한지를 설정하는 과정이 따로 요구되기도 한다. 또한, 사람의 옷의 텍스쳐 같은 것은 너무 다양하기 때문에 학습에 의한 기대효과를 얻기가 거의 불가능하므로 물체의 모델이 사람의 얼굴 내부 영역으로 제한되게 된다. 그러나, 많은 연구결과에 의하면 대상물체가 그 주위의 환경과 관련성이 있을 경우 그 상관성을 이용하면 잡음이나 클러터 등으로 생성되는 아웃라이어를 줄이고, 최종적인 위치 추적 결과도 향상된다는 사실을 알 수 있다[3-5]. 따라서, 주위의 물체가 추적하고자 하는 물체와 유기적인 관계로 연결되어 있고, 통계적인 범위 하에서 변화한다면 이를 통합하여 학습시키는 방법을 찾는 과정이 요구된다고 할 수 있다. 본 논문에서는 주어진 영상에서 사람의 얼굴과 머리의 위치를 찾는 문제를 다루고자 한다. 주위의 관련성을 이용하기 위해서 어깨선을 포함하여 이용하는데, 어깨선의 모양과 위치가 머리의 위치과 관련해서 통계적으로 제한된 범위 내에서 일관성을 가지고 같이 움직이기 때문이다. 머리와 상반신 정보를 동시에 사용했었던 비슷한 다음과 같은 연구들이 있었다. 먼저, Isard 와 Blake[6] 등은 손으로 그린 상반신 템플릿을 콘덴세이션 (Condensation) 방법에 사용하였다. 스플라인 (Spline) 매개변수를 변화시키는 것을 허용했지만, 사람의 움직임이 영상에서 수평적이고 모델에 사용한 동일한 사람을 찾는 방법에 사용되었다. Patil[7] 등은 얼굴 검출을 하는데 있어서 어깨선 정보를 이용해서 얼굴 검출에서 잘못되어 나오는 결과를 줄였지만, 모델이 너무 단순해서 얼굴과 어깨선을 정확히 추출하지 못했다. Lee 등은[8] 얼굴과 어깨선의 윤곽선을 K-means 클러스터링 기법으로 나누고 에지와 정합시켜서 3차원 포즈를 찾기 위한 후보로서만 이용하였다. 본 논문에서는 머리의 외곽선과 연결되는 어깨선을 모양을 하나의 연결되는 물체로 간주하고 이를 통합하여 학습시킨다. 이러한 형태의 모양을 그리스 문자 오메가(Ω)의 형상과 유사한 관계로 오메가-형상(Ω-shape)으로 명명하기로 한다. 제안하는 오메가-형상에서는 형상정보에 대해서는 학습을 수행하는데, 기존의 AAM 모델과 달리 텍스쳐 영역에 대해서는 이를 학습하지 않고, 각 부분의 텍스쳐에 대한 통계적인 특성을 컬러공간상의 분포로서 유지하며, 이에 기반하여 각 화소가 머리, 상반신, 주위배경 중 어느 영역에 속하는지 화소별로 분류한 정보를 이용한다. 또한, 비디오 시퀀스의 경우 매 프레임마다 변화에 맞춰서 각 부분의 텍스쳐의 통계적인 특성에 대해서 업데이트 한다. 이것을 기존의 학습에 의한 외모(appearance)과 비교해서 분별적 외모(discriminative appearance)라고 명명한다. 분별적 외모(discriminative appearance)의 개념은 Ramanan 등[4]이 도입하였는데, 이 방법은 원래 선형적 리그레션(linear regession)기법을 이용해서 팔, 다리 영역의 텍스쳐 정보를 컬러 공간 상에 특정 영역으로 모델링 하고, 입력되는 각 화소에 대해서 주어진 컬러값이 그 영역에 해당하는 경우에 대해서 팔, 다리의 영역으로 분별한 기법이다. 그런데, 여기서 사용된 컬러 모델은 아주 단순하며 비디오에서 프레임의 변화에 맞춰서 업데이트하기가 어렵기 때문에, 본 논문에서는 이를 개선시켜 베이시안 클래시파이어 (Bayesian classifier)를 설계하여 이를 기준으로 주위 배경의 텍스쳐 성분과 오메가-형상 내의 각 부분의 텍스쳐 성분을 나누어 구분한다. 특히, 분별적 외모를 생성하는 클래시파이어는 원하는 물체 내부의 텍스쳐 특성을 추출할 때 있어서 주위의 배경에 대한 상대적인 특성을 고려해 주기 때문에 근처 배경과 대상물과 컬러가 어느 정도 서로 비슷한 상황에서도 분별력이 뛰어난 결과를 보인다. 이 과정과 유사한 방법으로 목표물과 근처배경에 대한

비슷한 온라인 구별 기법이 제안되었는데, 그 기법은 온라인으로 약 클래시파이어 (weak classifier)의 계수를 업데이트 하는 방법이다[9]. 그러나, 그 방법에서는 목표물의 형상에 대한 정보가 이용되지 않고, 작은 박스와 이를 포함하는 큰 박스의 형태로 물체와 주위배경을 구분하고 이를 온라인으로 추적하는 방법이므로 다양한 물체의 형태에 맞게 설계하는 것이 어렵다. 결국, 본 논문은 얼굴부분을 찾는 과정에 있어서 어깨선을 이용함으로서 얼굴 검출이나

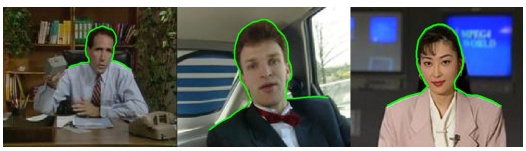


그림 1. 오메가-형상(Ω-shape) 트레이닝 영상의 예.

추적이 어려운 가려진 부분에서도 윤곽선을 찾는데 도움을 제공한다. 한편, 학습된 형상의 랜드마크를 정합하는데 있어서 그래디언트 성분 뿐 아니라, 분별적 외모 모델이 제공해 주는 각 영역의 경계가 되는 부분에 대한 정보를 동시에 활용해 줌으로서 잡음에 비교적 강건한 정합 특성을 보여준다. 따라서, 형상학습의 프레임워크하에서 기존의 텍스쳐 정보를 학습하는 방법에서 통계적인 분포를 업데이트 하는 방법으로 개선함으로서 기존의 AAM 방법이 학습하기가 어려운 점을 해소하고 있다. 동시에 물체 내부뿐 아니라 주위 배경을 고려한 텍스쳐 분포가 고려되는 장점이 있다.

Ⅱ. 특징 추출

본 논문에서 형상 학습을 위해서 그림 1과 같은 200여개의 영상에 대해서 학습한다. 이와 같이 학습된 형상의 정합을 위해서 두 가지 특징을 사용하는데, 하나는 그래디언트이고 다른 하나는 사람의 머리, 어깨, 머리근처의 배경의 컬러 히스토그램(color histogram)에 기반한 분별적 외모(discriminative appearance) 모델이다.

2.1 에지 그래디언트(Edge gradient)

주어진 선분, L 상에서 에지 그래디언트 포텐셜 (ϕ^e)을 정의한다. 이를 위해 먼저 주어진 영상의 화소, p 에서 Sobel 연산자를 적용하여 크기와 방향을 갖는 출력벡터 $g(p)$ 를 구하고, 다음과 같이 에지 그래디언트 에너지를 정의한다.

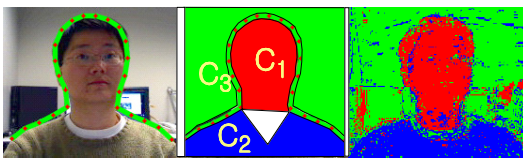
$$E^e(L) = \int_{p(\in L)} g(p) \bullet L / \|L\| \, dp \,. \qquad (1)$$

이것으로부터 선분 L 에 대한 에지 포텐셜 ϕ^e 은 다음과 같이 표현한다.

$$\phi^e(L) = 1 - 1/\sqrt{2\pi}\sigma_e \exp(-E(L)/\sigma_e^{\,2}), \quad E^e(L) \in [0,1]. \qquad (2)$$

2.2 분별적 외모 (discriminative appearance, DA)

외모(Appearance)는 영상 정합에 많이 이용되는 특징인데 포즈나 조명변화 등에 의해서 큰 변화를 겪는다. 또한 사람의 피부색과 옷의 색 등의 변화가



(a) (b) (c)

그림 2. (a) 오메가-형상(Ω-shape). (b) 주어진 Ω 형상에 대한 분별적 외모, D^{Ω}. (c) 추정된 분별적 외모, $\tilde{D}$.

심해서 통계적인 방법으로 학습하기도 쉽지 않다. 따라서 본 논문에서는 각 화소를 머리, 어깨, 배경으로 나누는 분별적 외모(discriminative appearance, DA) 모델에 대해서 제안한다. 그림 2(a)에서와 같이 일단, 오메가-형상이 추정되면 주어진 Ω 형상에 대해서 그림 2(b)와 같이 각 부분을 머리, 어깨, 배경으로 나눌 수 있게 되는데, 이를 Ω에 대한 분별적 외모 D^{Ω}라고 한다. 여기서, c_1, c_2, c_3는 각각 머리, 어깨, 배경 클래스를 나타낸다. 그림 2(b)의 흰색부분은 경계에 가까워서 고려하지 않는 부분이다. D^{Ω}는 주어진 Ω 형상의 위치에 따른 해당화소 p의 위치에 의해서 결정되는데, 그림 2(b)와 같이 $p \in c_i$ 이면 $D^{\Omega}(p) = i$ 로 정의하자.

2.2.1 베이시언 클래스 분류기(Bayesian Classifier)

D^{Ω}가 Ω에 대한 화소 p의 상대적 위치에 의해서 구해지는 분별적 외모라면, Ω를 추정하는데 있어서 Ω의 위치를 알 수가 없기 때문에, 화소 p에서의 컬러 정보를 이용한 분별적 외모에 대한 추정이 요구된다. 그림 2(c)에서와 같이 각 화소에서 추정된 분별적 외모의 인스턴스를 $\tilde{D}$라고 두자. 그러면 각 클래스를 구별하는 클래스 분류기를 다음과 같이 설

계한다.

$$\tilde{D}(p) = i \text{ if } pr(c_i \mid v(p)) > pr(c_j \mid v(p)) \text{ for all } j \neq i . \quad (3)$$

위 식에서 포스테리어(posterior)는 베이시언 규칙에 의해서 나타내진다. 즉, $pr(c_i \mid v) = \frac{pr(v \mid c_i) pr(c_i)}{pr(v)}$ 와 같이 나타낸다. 본 논문에서는 프라이어(prior) $pr(c_i)$ 의 초기치를 각 클래스 영역의 크기에 비례하도록 결정했는데, 추후 클리시파이어의 오차를 줄이는 방향으로 업데이트 한다. 라이크리후드 (likelihood) $pr(v \mid c_i)$는 컬러공간에서의 히스토그램(smoothed histogram)으로 구해진다. 그러면, 각 클래스, c_i 에대해서 히스토그램 $h(v \mid c_i)$을 어떻게 구하는지 설명

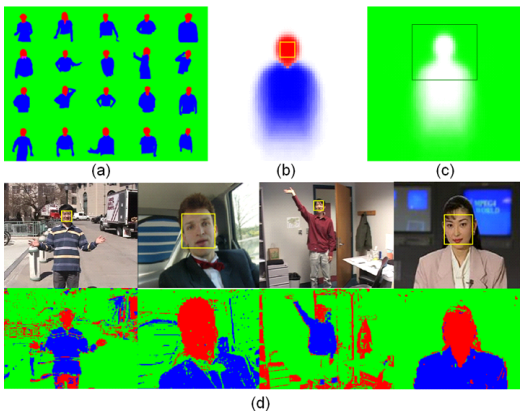


그림 3.(a) 정규화한 레이블된 영상. (b) 얼굴위치로부터 머리와 상반신 위치의 확률 분포. (c) 배경의 위치분포. (d) 영상에 적용된 분별적 외모(DA)의 예.

하도록 하자. 정지 영상 혹은 첫 번째 프레임에서 사람이 영상에 수직방향에 가까운 포즈를 취하고 있다고 가정한다. 그리고 얼굴 검출을 통해서 주어진 얼굴의 위치 X와 얼굴의 크기에 정규화해서 사람의 얼굴, 어깨, 배경에 대한 대략적인 위치를 추정한다. 포즈변화와 사람간의 편차를 고려하기 위해서, X를 기준으로 각 위치에서 어느 클래스에 속할지에 대한 모델을 통계적으로 학습시킨다. 이를 위해서 400여 장의 서 있거나 앉아있거나 걸어가는 사람의 영상을 라벨링(labeling)하고 얼굴 위치에 대해서 정규화하였다. 이에 대한 예가 그림 3(a)에 나타나 있다. 각 위치에서의 확률값을 그림 3(a)와 3(b)와 같이 모델링 한다. 배경은 전체영상의 배경을 모두 사용하지 않고 그림 3(c)에서와 같이 머리 주위의 배경을 이용하는데 국부적인 특성을 더 잘 반영해 주기 때문이다. 각 클래스 c_i에 대해서 각 화소가 속할 확률을 $pr(p \in c_i \mid X)$로 나타내자. 그러면, 각 영역을 각 클래스에 속할 확률값에 비례해서 샘플링하고, 여기에 기반하여 각 클래스에 대한 $h(v \mid c_i)$를 생성한다. 이때, 사용되는 공간은 RGB 컬러공간 상에서 32x32x32로 양자화시킨 공간이며, 각 히스토그램을 구한후 메디언(Median) 필터를 사용하여 부드럽게 만들었다. 정지 영상 혹은 첫 번째 프레임에서는 프라이어(prior) $pr(c_i)$의 초기치를 각 클래스 영역의 크기에 비례하도록 결정하고, 결국 최종적인 베이시언 클래스 분류기를 생성시킨다. 그림 3(d)는 얼굴검출로부터 각 정지영상에 대해서 클래스 분류기를 통해서 생성한 추정된 분별적 외모의 추정치 $\tilde{D}$의 예를 보여준다.비디오 시퀀스에서 외모(appearance)는

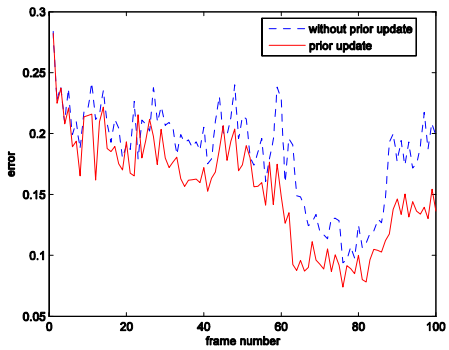


그림 4. 프라이어를 업데이트 했을 경우(적색실선)와 업데이트 하지 않았을 경우(청색점선) 추정된 오메가-형상과 실제형상의 오차비교.

이웃하는 프레임 사이에서 크게 변하지 않는다고 가정한다. 프레임 t에서 이전프레임의 c_i 영역에 해당하는 화소들을 샘플링함으로서 컬러 히스토그램 $h_{t-1}(v \mid c_i)$을 구한다. 그러면 현재 프레임에서의 라이크리후드(likelihood)는 다음과 같이 업데이트 한다.

$$pr_t(v \mid c_i) = (1 - \alpha_i) pr_{t-1}(v \mid c_i) + \alpha_i h_{t-1}(v \mid c_i) . \quad (4)$$

이때 α_i 값은 적당한 값으로 정해야 하는데, 너무 작으면 변화의 속도가 느려서 현재 상태의 변화를 따라가지 못하고 너무 큰 경우에 대해서는 안정성이 문제가 발생할 수 있다. 한편, 베이시안 클래시파이어의 성능을 향상시키기 위해서 이전 프레임 t에서 오메가-형상 D_t^{Ω}의 위치를 추정했다면 $\tilde{D}_t$ 상에서 발생하는 오차를 분석할 수 있다.

즉, 두 개의 차이를 계산해서 c_i에 속하는 화소들의 수가 실제로 존재해야 될 비율보다 γ만큼 적을 경우에 프라이어 $pr_{t+1}(c_i)$를 η만큼 더 증가시키고, 반대로 γ만큼 클 경우에 프라이어 $pr_{t+1}(c_i)$를 η만큼 더 감소시킴으로서 다음 프레임에서 오차가 줄어들도록 조절한다. 그림 4는 프라이어를 업데이트 했을 때 오메가-형상의 추정값과 실제값의 차이가 더 줄어들고 있음을 보여주고 있다.

2.2.2 DA 경계 포텐셜

$\tilde{D}_i$를 구하고 나서 주어진 선분 L상에서 DA 경계 포텐셜 ϕ^b를 구한다. 이를 위해서 먼저 직사각형 형태의 영역 $S(L)$을 정의하는데, 직사각형의 너비를 선분 L의 길이와 똑같이 맞추고, 선분 L이 그 직사각형의 중심선에 위치하게 $S(L)$의 위치를 잡는다.

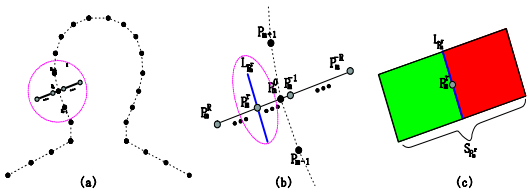


그림 5.(a) 오메가 형상의 랜드마크. (b) 각 랜드마크로부터 후보점과 선분의 지정. (c) 선분으로부터 영역을 정의.프라이어(prior) 또한 업데이트 한다

그러면 그림 5와 같이 선분 L이 $S(L)$을 양분하는데 각 부분이 대상물체의 안쪽 부분과 바깥부분이 된다. 따라서, 선분 L의 경계 에너지는 다음과 같이 표시된다.

$$E^b(L)=\int_{p(\in S(L))}\left\|D^{\Omega}-\tilde{D}\right\|/\|S(L)\|dp \quad (5)$$

각 화소 $p(\in S(L))$에서, $\tilde{D}(p)=D^{\Omega}(p)$이면 $\left\|D^{\Omega}(p)-\tilde{D}(p)\right\|=0$, 그렇지 않을 경우 $\left\|D^{\Omega}(p)-\tilde{D}(p)\right\|=1$로 정의한다. DA 경계 포텐셜 함수 ϕ^b는 다음과 같이 계산된다.

$$\phi^b(L)=1/(\sqrt{2\pi}\sigma_b)\exp(-E(S(L))/{\sigma_b}^2),\quad E^b(L)\in[0,1]. \quad (6)$$

이 포텐셜 값은 바깥부분과 안쪽 부분이 정확히 일치해야 가장 큰 값을 가지고 멀어질수록 값이 작아지게 된다.

Ⅲ. 분별적 외모를 갖는 ASM(Active Shape Model with the DA feature)

3.1. 형상 학습

형상학습은 기본적으로 전형적인 ASM기법에서의 학습과 비슷한 과정을 따른다. 먼저, 하나의 오메가-형상은 M개의 랜드마크 점들로 표현한다. 형상의 일관성을 유지하기 위해서 1번째 점과 24번째 점이 각각 오른쪽과 왼쪽 어깨 끝점으로 유지하고 4번째와 21번째 점을 머리외곽선과 어깨선의 코너점으로 설정한다. 여기서 M은 24로 설정하였다. 또한, 다른 점들은 서로 간격이 일정하게 유지되도록 최대한 맞추었다. 그리고, Procrustes 거리가 최소가 되도록 병진, 스케일, 회전 등을 조절하여 정규화를 수행한다. 그러면 오메가-형상 벡터는 다음과 같이 나타낼

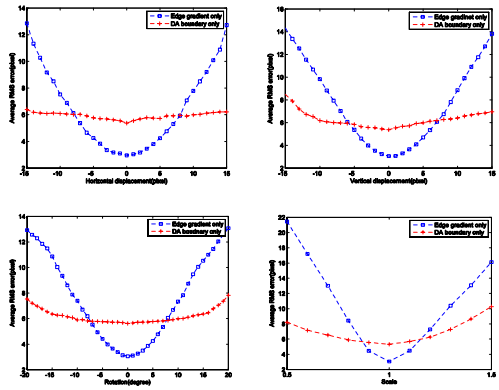


그림 6. 수평이동, 수직이동, 회전, 크기변화 등에 대해서 에지 그래디언트 포텐셜 $\phi^e(L_{p'_m})$과 DA 경계포텐셜 $\phi^b(L_{p'_m})$을 각각 따로 사용 했을 때 오차의 분포

수 있다.

$$\Omega=(p_1,p_2,\cdots,p_M), \quad (7)$$

이때, $p_m=(x_m,y_m)$이다. PCA(principle component analysis)를 적용하여 Ω가 평균형상 $\overline{\Omega}$와 M 개의 고유벡터 φ_m의 선형결합으로 표현한다.

$$\Omega=\overline{\Omega}+\sum_{m=1}^{2M}\lambda_m\varphi_m \quad (8)$$

이 표현에서 $\lambda=[\lambda_1,\cdots\lambda_M]^T$는 형상 파라메터이고, 형상을 학습시켜서 형상 파라메터 $[-k\sigma_m,k\sigma_{mi}]$가 학습된 허용 공간 내에 존재하도록 조건을 부여한다. 학습에 사용된 영상은 200개 이다.

3.2. 랜드마크 위치찾기

첫 번째 프레임에서 얼굴검출[10]을 통해서 최초 오메가-형상의 위치를 대략적으로 알고 있다고 가정한다. 랜드마크를 찾는 방법은 최초 오메가-형상의 외곽선에 수직한 방향으로 이루어진다. 즉, m-번째 랜드마크에 대해서 그림 5(a)와 같이 수직라인을 따라서 2R+1개의 후보점들이 샘플링 된다. 각각의 후보점들이 오메가 형상의 경계가 될 가능성이 가장 큰것인지 비교되고 가장 적합한 점이 새로운 랜드마크의 위치가 된다. 즉, m-번째 랜드마크의 r-번째 $(-R \le r \le R)$ 후보점을 $p^r{}_m$이라고 하고, 각 점에 대해서 오메가 형상의 경계가 되는 최적의 점을 구하기 위해서 (9)와 같이 랜드마크 포텐셜 함수를 정의하고 최대가 되는 점을 랜드마크로 정한다.

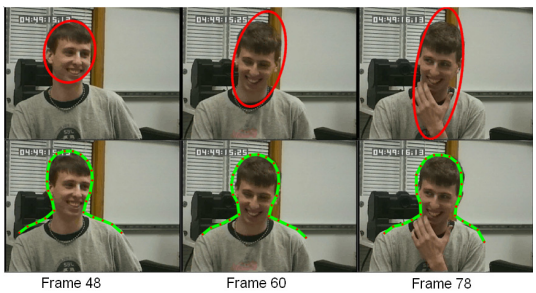


그림 7. Camshift[11]의 결과(첫째열)과 제안된 기법의 결과(둘째열)의 비교.

$$\phi(p^r{}_m) = \phi^e(L_{p^r{}_m}) \times \phi^b(L_{p^r{}_m}) \quad (9)$$

랜드마크 포텐셜함수는 앞에서 이미 정의된 에지 포텐셜과 DA 경계 포텐셜의 곱으로 정의했는데, 각각의 성분이 어떤 역할을 하는지 간단히 살펴보도록 하자. 이를 위해서 오메가-형상을 실제 참값 위치에서부터 의도적으로 옮긴 후, 두 성분의 곱이 아닌 각각의 성분으로 결정되는 오메가 형상이 어떻게 변하는지 살펴본다. 즉, 수평이동, 수직이동, 회전이동, 스케일 변화를 주고 이것을 초기위치로 선정하고, 각 성분에 의해서 결정되는 오메가-형상을 PCA를 통한 반복과정으로 수렴시키고, 이 수렴된 형상과 참값의 형상의 오차를 구한 결과가 그림 6에 도시되어 있다. 에지 그래디언트 포텐셜을 사용했을 경우에는 초기 위치가 참값 위치와 유사할 경우에 대해서는 잘 수렴하지만, 멀어질수록 오차가 크게 증가하는 경향을 보여준다. 반면에 DA 경계 포텐셜은 초기위치가 참값위치가 가까울 경우에도 정확하게 수렴하지는 않지만, 멀어지는 경우에 있어서도 참값의 범위를 크게 벗어나지 않는 경향을 보여준다. 이는 에지 그래디언트는 가까이 있는 강한 에지를 찾으려고 하는 경향이 강한 반면, DA 경계 포텐셜은 각 영역의 컬러분포의 특성에 부합하는 영역을 찾으려고 하는 경향이 강하기 때문이다. 즉, 에지 그래디언트 성분은 분명한 경계의 위치를 찾으려는데 기여하고, DA 포텐셜은 수렴되는 시스템의 안정성에 기여함을 알 수 있다.

Ⅳ. 실험결과

먼저 제안된 기법을 실시간 MeanShift 추적기의 하나인 CamShift 방법[11]과 비교하였다. CamShift 방법과 제안된 방법 모두 컬러 정보를 이용하고 있지만, 제안된 방법은 학습에 의한 형상 정보를 사용하기 때문에 그림 7에서와 같이 얼굴과 색깔이 유사한 손이 가까이 오는 경우에 대해서도 목표물의 모양을 유지하면서 추적을 하고 있다. 또한, 제안된 방법의

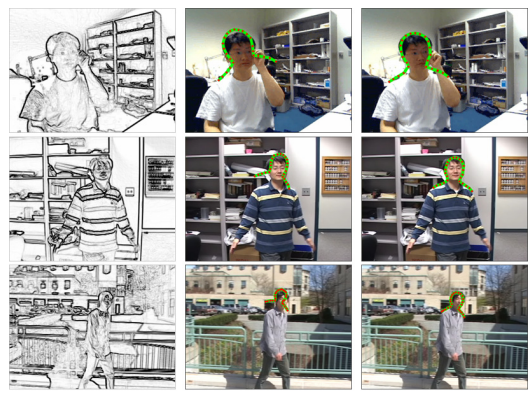


그림 8. 근처의 배경에지로 인해서 ASM 방법이 실패하기 시작하는 프레임에서의 결과와 이 때 제안된 방법의 결과.

성능을 평가를 위해서 ASM 기법과 비교하였는데, 오메가-형상에 적합하도록 수정하였다. 원래 사용되는 ASM[1] 방법에서는 얼굴 내부의 눈, 코,입 등의 특징점을 사용하고 있지만, 여기서의 목적은 얼굴이 돌아가거나 가려짐이 있는 경우에도 수행하는 것을 목적으로 하기 때문에 순수한 오메가-형상에 대해서만 학습을 시켰다. 테스트는 30개의 비디오 시퀀스에 대해서 수행하였는데, 카메라가 고정된 것도 있고, 움직이는 경우도 있다. 사람은 서있거나 걷거나 포즈를 바꾸는 경우도 있으며 프레임 수는 약

표 1. 테스트 비디오시퀀스에 대해서 제안된 방법과 수정된 ASM방법 및 Camshift 추적기와 성능비교.

비디오번호	프레임수	가려짐	자세변화	조명변화	수정ASM 트래킹성공여부	Camshift 트랙킹 성공여부	제안방법 트래킹 성공여부
1	870	x	o	x	o	o	o
2	723	o	o	o	x	x	o
3	3327	o	o	x	x	x	o
4	872	x	o	x	o	x	o
5	1050	o	o	x	x	x	x
6	636	o	o	x	x	x	o
7	4670	o	o	x	x	x	o
8	692	x	o	x	o	x	o
9	456	x	o	x	x	x	o
10	514	o	x	x	x	x	o
11	455	x	o	x	o	x	o
12	452	x	o	x	o	x	o
13	360	o	o	o	x	x	o
14	781	x	o	o	x	x	o
15	636	x	o	x	x	x	x
16	873	o	o	x	x	x	o
17	750	x	o	x	x	x	o
18	902	o	o	x	x	x	o
19	964	o	o	x	x	x	o
20	1292	x	o	x	o	x	o
21	1146	o	o	x	x	x	o
22	1623	o	o	o	x	x	o
23	1175	x	o	o	x	x	o
24	1119	o	o	x	x	x	o
25	1444	o	o	x	x	x	o
26	1140	x	o	x	x	x	o
27	1053	o	o	o	x	x	o
28	1051	o	o	o	x	x	o
29	690	x	o	x	x	x	o
30	2792	o	o	o	x	x	o

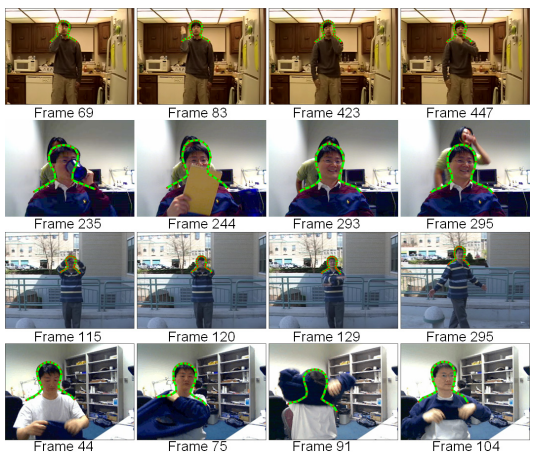


그림 9. 제안된 방법이 가려짐에 대해서 강건한 예.

수백개에서 수천개까지 다양한 경우에 대해 테스트를 수행하였다. 표 1에서 나타난 바와 같이 ASM 기법은 전체 30 비디오 중 6개의 경우에 대해서만 성공적으로 첫 프레임부터 마지막 프레임까지 목표를 추적하였고, Camshift는 한 개의 비디오에 대해서, 제안된 방법은 30개중 28개에 대해서 성공적으로 트랙킹을 수행하였다. 가려짐이 확연히 존재하는 상황은 17개의 비디오에서, 자세의 변화는 29개의 비디오에서, 그리고 두드러진 조명의 변화는 8개의 비디오에서 나타나 있다. 가려짐이 확연히 존재하거나, 두드러진 조명의 변화가 있는 경우에 있어서, Camshift와 수정된 ASM 기법은 하나도 추적에 성공하지 못했으며, 제안된 방법은 가려짐이 큰 비디오 1개와 자세변화가 큰 비디오 한 개의 경우에 끝가지 추적을 하지 못하였다. 그림 8에서는 기존의 ASM 기법이 오메가-형상을 따라가지 못하는 부분의 예를 보여주고 있는데, 근처의 배경에지 등으로 인해서 오메가-형상을 잘못 수렴시키는 상황을 보여주고 있다. 반면에 제안된 방법에서는 원래 오메가-형상을 제대로 따라가고 있는데, 이는 제안된 방법이 에지 포텐셜 뿐만 아니라, 영역에 기반하고 있는 분별적 외모 모델을 동시에 사용하기 때문이다. 그림 9에서는 제안된 방법이 부분적인 가려짐에 대해서 강건함을 보여주고 있다. 학습된 형상을 통해서 오메가 형상을 유지하려는 성향과 분별적 외모 모델이 결합해서 아주 심각한 가려짐이 아닌 경우에 대해서는 오메가 형상을 잘 추적하는 것을 보여준다. 그림 10에서는 제안된 방법이 포즈의 변화에 대해서 강건하게 추적하고 있는 예를 보여준다. 첫 번째 열의 비디오는 대상자가 외부환경 하에서 왼쪽으로 혹은

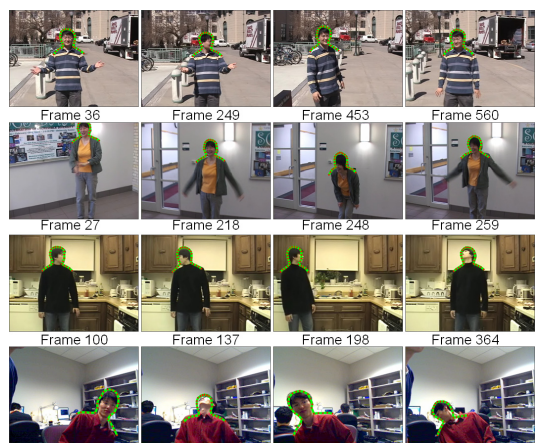


그림 10. 제안된 방법이 포즈의 변화에 대해서 강건한 예.

오른쪽으로 움직이고 있으며 그림자도 나타난다. 두 번째 시퀀스는 대상자가 갑자기 몸을 앞쪽으로 굽힌다. 세 번째 시퀀스에서는 대상자가 몸을 움직이면서 동시에 고개를 많이 움직이며, 네 번째 시퀀스에서는 상반신을 전체적으로 크게 움직이는데 제안된 방법을 통해서 잘 추적하고 있음을 관찰할 수 있다. 그림 11에서는 비교적 두드러진 조명의 변화에 대해서도 강건하게 추적하고 있는 걸을 보여주는데, 같은 비디오에 대해서 AAM방법[2]의 추적결과를 보여주고 있다. 이때, AAM의 학습을 위해서 동일한 비디오에서 초기 100 프레임중 30 프레임을 샘플링하여 학습시켰음에도 불구하고 학습되지 않은 조명의 변화에 대해서 제대로 목표물을 추적하지 못함을 보여준다.제안된 알고리듬의 또 다른 장점의 하나는 실시간으로 동작하는 점이다. 얼굴의 크기가 약 30~40 화소의 크기의 영상에 대해서 펜티엄 4 CPU 2.8GHz에서 초당 50~70 프레임의 처리속도를 보였다. 그러나, 몇가지 알려진 알고리듬의 한계가 있다. 만일 포즈가 학습된 형태와 아주 다른 경우에 대해서 실제 오메가의 형상이 사람의 머리와 어깨선과 일치하지 않는 경우가 발생한다. 만일 포즈가 학습된 것과 아주 많이 차이가 나는 경우에 있어서 그림 12와 같이 사람의 실제 오메가 형상과 정합하지 않는 경우가 발생한다. 그러나, 비디오에서 많이 차이가 났던 포즈가 학습된 포즈와 유사한 형태로 돌아올 경우에 복원하는 것을 보여준다.

V. 결론 및 향후과제

(a) AAM 기법의 트래킹

(b) 제안된 기법의 트래킹

그림 11. 제안된 방법이 조명의 변화에 대해서 강건한 장면.

본 논문에서는 정지영상이나 비디오에서 오메가-형상을 찾는 방법을 제안하는데 머리와 어깨선의 연관성을 통해서 기존의 얼굴추적 기법만으로 잘 동작하지 않는 가려짐이나 포즈변화 같은 경우에 있어서 강건하게 동작하도록 설계할 수 있었다. 에지에 기반한 ASM 방법의 문제를 극복하기 위해 분별적 외모 모델을 사용하였다. 각 프레임에 맞게 변화되도록 온라인으로 업데이트 할 수 있는 베이시언 클래시파이어를 설계하여 머리, 어깨, 주위 배경부분을 잘 구분할 수 있었다. 그리고, 이 클래시파이어는 컬러 히스토그램에 기반하여 화소별로 결정이 이루어지므로 초당 약 70 프레임 정도 수행할 수 있는 실시간 처리가 가능하다. 향후 실제 환경에서 발생할 수 있는 더 많은 자세에 대해서 트래킹을 수행하는 것이 필요하다. 이를 위해서 다중의 학습모델을 개발하고 다중의 학습모델 중 각 경우에 맞도록 선택해 주는 알고리듬의 연구가 필요하다.

참 고 문 헌

〔1〕 T.F.Cootes, C.J.Taylor, D.H.Cooper and J. Graham. "Active shape models - their training and application." *Computer Vision and Image Understanding*, No.61, Vol.1, pp.38-59, 1995.

〔2〕 Matthews and S. Baker, "Active appearance models revisited," *Int. J. Computer Vision*, Vol.60, No.2, November, 2004, pp.135-164.

〔3〕 P.F.Felzenszwalb and D.P.Huttenlocher, "Efficient matching of pictorial structures", *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2000, Vol.2, pp.66-73.

〔4〕 D.Ramanan, D.A.Forsyth and A.Zisserman, "Strike a Pose: Tracking People by Finding Stylized Poses." *IEEE Conf. Computer Vision and Pattern Recognition*, 2005, pp.271-278.

〔5〕 A.Torralba, "Context priming for object detection", *Int. J. Computer Vision*, Vol.53, No.2, July, 2003, pp.153-167.

〔6〕 M.Isard and A.Blake, "Condensation: conditional density propagation for visual tracking", *International Journal of Computer Vision*, 29,1, pp.5-28, 1998.

〔7〕 R.Patil, P.E.Rybski, T.Kanade and M.M.Veloso, "People detection and tracking in high resolution panoramic video mosaic", *IEEE/RSJ Int. Conf.Intelligent Robots and Systems*, Vol.1, 2004, pp.1323-1328.

〔8〕 M.W.Lee and I.Cohen, “Human upper body pose estimation in static images”, *Proc. European Conf. Computer Vision* 2004, pp. 126-138.

〔9〕 S.Avidan, “Ensemble Tracking”, *IEEE Conf. Computer Vision and Pattern Recognition*, 2005, pp.494-501.

〔10〕 P.Viola and M.J.Jones, “Rapid object detection using a Boosted cascade of simple features”, *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2001, pp.511-518.

〔11〕 G.R.Bradski, “Computer vision face tracking for use in a perceptual user interface,” *Intel Technology Journal*, 2nd Quarter, 1998.

권 무 식(Musik Kwon) 정회원

1996년 8월 : 서울대학교 전기공학부 졸업

1999년 2월 : 서울대학교 전기공학부 석사

2004년 2월 : 서울대학교 전기 · 컴퓨터공학부 박사

2004년 3월 ~ 2006년 5월 : 카네기멜론대학교 로보틱스연구소 박사 후 과정

2006년 7월 ~ 현재 : 삼성전자 정보통신총괄

<관심분야> 물체인식, 3D그래픽스, 영상처리 등