

얼굴 영역 추적을 통한 향상된 손 영역 추적에 관한 연구

손지수, 김동규, 이승호, 노용만
 카이스트 전기 및 전자공학부
 e-mail : {sze0621, d.kim, leesh09, ymro}@kaist.ac.kr

Improved Hand Region Tracking Using Face Region Tracking

Jisoo Son, Dongkyu Kim, Seung Ho Lee, Yong Man Ro
 School of Electrical Engineering, KAIST

요 약

손 영역 추적에서는 피부색이 가장 유용한 정보 중 하나이다. 그런데 손 영역과 얼굴 영역이 서로 겹치거나 가까이 있을 때 손 영역의 추적결과인 바운딩이 얼굴 영역까지 불필요하게 확장되는 문제점이 존재한다. 본 논문에서는 얼굴 영역 추적결과를 손 영역 추적에 사용한다. 구체적으로, 얼굴 영역 내에 손 영역의 바운딩이 침투하지 않도록 한다. 실험결과, 얼굴 영역 추적결과를 사용한 경우 그렇지 않은 경우에 비해 손 영역의 바운딩을 정확히 예측하였으며 초당 30~35 프레임의 빠른 계산속도를 유지하였다.

1. 서론

손 동작(gesture)은 풍부한 의미를 지닌 표현 방식이며 일상 생활에서 매우 자주 사용되고 있다[1]. 이러한 손 동작 인식 기술은 자연스럽게 직감적인 인간 컴퓨터 상호작용(Human-Computer Interface, HCI)을 가능케 하며, 수화 인식, 엔터테인먼트 등의 다양한 응용 분야가 있다.

손 동작 인식을 위해서는 높은 성능의 손 검출(detection) 또는 추적(tracking) 기술이 필요하다. 그러나 제한되지 않은 실질적인(real-world) 응용에서 손 검출 또는 추적을 정확히 수행하는 것은 어렵다[2]. 손은 기본적으로 관절(articular) 구조로, 형태가 정해져 있지 않으며 (deformable, non-rigid), 시점에 따라, 또는 물건을 잡고 있을 때 폐색(occlusion)이 발생하는 등 매우 가변적이기 때문이다.

최근 들어 제안되는 대부분의 손 추적 방법들은 RGB 영상과 깊이(depth) 정보를 함께 사용한다. 하지만 깊이 카메라는 현재까지는 일반적인 사용 환경에 널리 보급되어 있지 않으며, RGB 영상만 사용하는 손 추적에 비해 계산속도가 느려 실시간 처리에 적합하지 않다. 깊이 정보 없이 RGB 영상만으로 손 추적하는 알고리즘은 대부분 색상 정보를 사용하며 계산 속도가 빠른 것이 장점이다. 대표적인 기법들로는 Meanshift, Camshift[3], 칼만 필터, 파티클 필터[4] 등이 있다. 하지만 이러한 기법들은 추적하고자 하는 물체(object)와 비슷한 색상을 지닌 다른 물체가 배경에 있을 때 추적을 나타내는 바운딩 박스(bounding box)의 드리프트(drift) 현상이 발생한다. 특히 손과 비슷한 색상을 지니며 추적을 방해하는 물체로는 얼굴이 있다. 이 때 손과 얼굴의 위치가 가까울 때 추적을 나타내는 바운딩 박스가 불필요하게 얼굴까지 포함하며 확

장되는 문제가 있다.

본 논문에서는 앞서 언급한 드리프트 문제를 해결하기 위해 얼굴 영역 추적을 이용한 향상된 손 영역 추적방법을 제안한다. 제안하는 방법의 특징은 다음과 같다.

- 1) 제안방법은 추적을 통해 얼굴 영역을 먼저 예측함으로써 얼굴 영역과 손 영역의 위치가 겹치거나 서로 가까이 있을 때 불필요하게 손 영역의 바운딩이 얼굴 영역까지 확장되는 것을 방지할 수 있다.
- 2) 얼굴 영역 추적 시 발생하는 계산속도 저하를 최소화 하기 위해 낮은 차원의 효율적인 텍스처 특징을 사용하였다. 제안방법은 초당 30~35 프레임 정도의 속도를 유지하여 대규모의 실시간 처리 손동작 인식 응용에 적합하다.

웹캠(webcam)으로 입력 받은 동영상에 대해 실험을 수행한 결과, 추적된 얼굴 영역 정보를 손 영역 추적에 사용하는 경우 얼굴 영역 정보를 사용하지 않은 경우에 비해 손 영역을 정확히 예측하였다.

본 논문의 나머지 구성은 다음과 같다. 2 장에서는 제안하는 손 추적 방법에 대해 각 단계별로 자세히 설명한다. 3 장에서는 제안한 방법으로 실험을 수행한 결과를 보이고 4 장에서 결론을 맺는다.

2. 제안하는 손 영역 추적방법의 개요

본 논문에서 제안하는 손 영역 추적 방법에서는 Camshift 기반의 손 추적에서 손이 얼굴에 가까워질 때 역투영(backprojection)에서 파티클 필터로 얻은 얼굴 영역의 값을 줄임으로써 손의 위치를 추적할 수 있다. 여기서 역투영이란 영상 내의 각 지점들이 우리가 지정한 영역의 칼라 모델과 유사한 정도를 나타낸 것이다.

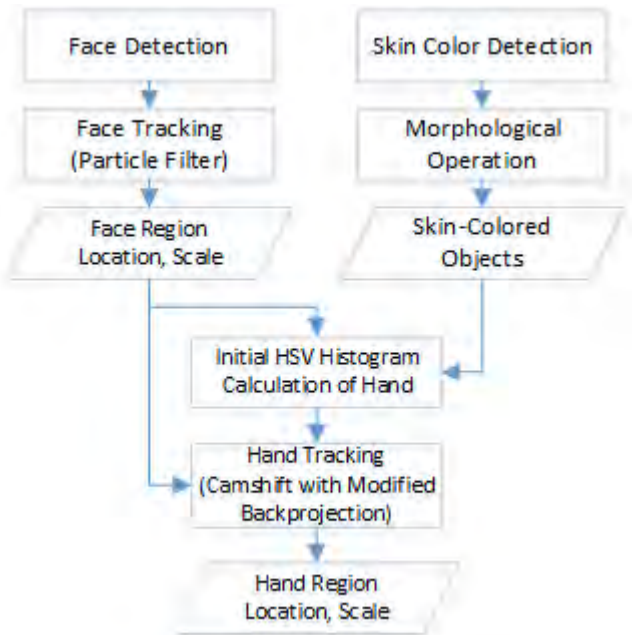


그림 1. 제안 방법 프레임워크

그림 1 은 제안 방법의 전체 프레임워크를 나타낸다. 프레임워크를 구성하는 각 모듈에 대한 설명은 다음과 같다.

(1) 얼굴 검출(Face Detection)

얼굴 추적 초기화를 위해 [5]에서 제안된 얼굴 검출기를 사용하였다. 제안 방법의 다른 모듈들의 동작을 위해서는 얼굴 검출이 선행되어야 한다.

(2) 얼굴 추적(Face Tracking)

얼굴 추적을 위해 파티클 필터를 사용하였다. 이때 각 파티클들의 가중치는 초기 상태의 RGB 히스토그램, 국부이진패턴(Local Binary Pattern, LBP[6]) 히스토그램과 각 상태의 히스토그램 간의 상관 계수(correlation coefficient)의 가중치 합으로 계산하였다.

(3) 피부색 영역 검출(Skin Color Detection)

피부색 영역을 검출하기 위해서 [7]에서 제안된 방법을 사용하였다. 프레임의 각 픽셀의 R, G, B 값들이 특정 조건들을 만족시키면 피부색 영역으로 판별한다.

(4) 모폴로지 연산(Morphological operation)

피부색 영역에서 얻은 이진 영상에서 닫기(closing), 침식(erosion), 팽창(dilation)의 모폴로지 연산을 거쳐, 배경에서 나타나는 작은 부분들은 제거한 피부색 영역들을 얻는다. 얼굴추적을 통해 얻은 얼굴 영역을 제외한 나머지 피부색 영역을 손 영역으로 보고 손 추적을 초기화 한다.

(5) 초기 손 영역 HSV 칼라 히스토그램 계산(Initial HSV Histogram Calculation of Hand)

모폴로지 연산에서 초기 손 영역을 찾아내면, 그 영역에 대해 HSV 칼라 히스토그램을 계산한다

(6) 수정된 역투영을 이용한 Camshift 기반 손 추적 (Camshift-based Hand Tracking using Modified Back-projection)

웹캠 입력 동영상의 매 프레임에 대해서 HSV 히스토그램 역투영을 계산한다.

2.1 입력된 비디오에서 얼굴 영역 추적

그림 2(가)에서 볼 수 있듯이 Camshift 역투영 영상을 관찰해보면 손과 얼굴 영역에서 높은 값을 갖는 것을 알 수 있다. 이 때 손과 얼굴이 가까우면 반복을 거치면서 손 영역과 더불어 얼굴 영역도 포함하게끔 바운딩 박스가 커지는 문제가 발생한다. 본 논문에서 제안하는 방법은 얼굴 추적으로 얻은 얼굴 영역의 위치와 크기 정보로 Camshift 의 역투영 영상에서 얼굴에 해당되는 영역 값을 줄임으로써 바운딩 박스가 손 영역 내에서 수렴하도록 유도한다. 본 논문에서 얼굴 추적에 사용한 파티클 필터는 얼굴 추적에서 많이 쓰이는 기법 중 하나이며, RGB 히스토그램과 LBP 의 조합을 통해 조명 변화, 폐색, 얼굴 포즈 변화에 강인한 것으로 알려져 있다[8].

파티클 필터는 유한 개의 가중 파티클을 가지고 반복을 거쳐 실제 사후 확률 분포를 근사한다[9]. 파티클 필터로 물체를 추적하는 기법은 이전 프레임에서의 관측으로 현재 프레임에서의 물체의 상태를 추정하는 것이다. 물체를 추적할 때 상태는 위치 좌표(x, y) 와 크기(scale)로 표현된다. 관측은 파티클들의 가중 합으로 결정되며, 가중치는 특징 비교에 따른 우도(likelihood)로 결정된다. 본 논문에서는 파티클 필터로 얼굴 추적을 할 때, 얼굴 검출기에 의해 나온 얼굴의 중심 좌표와 크기로 초기상태를 정의한다. 또한, 각 파티클은 이전 프레임에서 얻은 상태를 평균으로 하는 가우시안 분포를 따른다고 가정하여 생성한다. 이 때 각 상태에 대한 표준편차는 $\sigma_x, \sigma_y, \sigma_{scale}$ 로 정의된다. 마지막으로 관측모델은 초기 상태에서의 특징(feature)와 각 파티클에서 추출한 특징들 사이의 유사도를 측정함으로써 얻을 수 있다.

2.2 얼굴추적 결과를 이용한 향상된 손 영역 추적

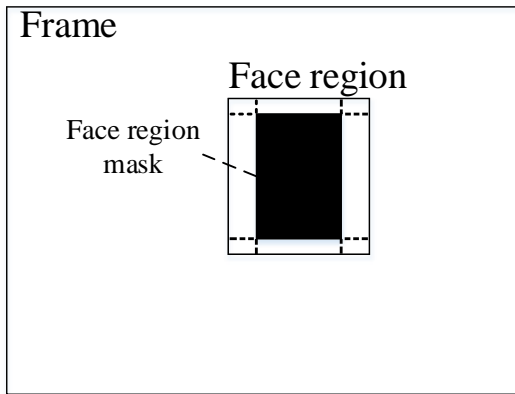
본 절에서는 2.1 절에서 얻은 얼굴 영역 정보를 이용하여 향상된 손 영역 추적방법에 대해 설명한다. Camshift 에서 추적을 수행할 때, 영상의 매 프레임에서 역투영을 구하게 된다. 역투영은 프레임의 각 픽셀들이 추적하고자 하는 물체의 색상 히스토그램에 얼마나 가까운지를 나타낸다. 그림 2(가)는 웹캠으로 입력 받은 동영상에서 Camshift 로 손을 추적할 때 그 프레임에서 구한 역투영 영상을 나타낸다. 그림에서 볼 수 있듯이 추적하고자 하는 손과 비슷한 색상을 지닌 얼굴 영역에서 큰 값을 갖는 것을 알 수 있다.

역투영 영상에서 Camshift 는 다음의 과정을 거쳐 추적을 수행한다.

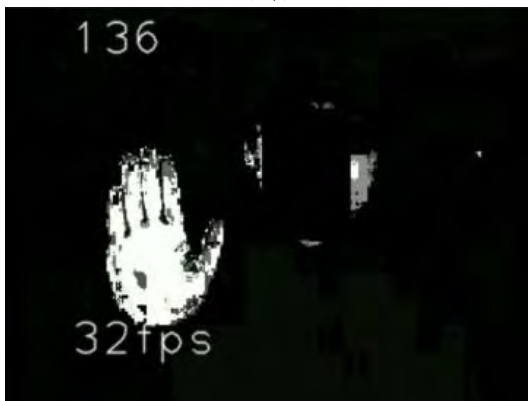
- (1) Meanshift 알고리즘을 수행한다.
- (2) Meanshift 가 수렴하면 윈도우 사이즈를 조절한다.
- (3) (1), (2)과정을 반복한다.



(가)



(나)



(다)

그림 2. (가) 기존 Camshift 의 역투영 영상, (나) 제안방법의 얼굴 영역 마스크(얼굴 영역 마스크의 중심점은 얼굴 영역의 중심점과 같고, 얼굴 영역과 얼굴 영역 마스크의 너비와 높이의 비율은 각각 0.6, 0.8로 정하였다), (다) 얼굴 영역이 마스크 처리된 제안 방법의 역투영 영상

이 때, 손이 얼굴 가까이에 있게 되면, 역투영 영상 내에서 손과 얼굴 영역이 합쳐져 하나로 나타나게 된다. 따라서 Camshift 는 반복을 거쳐 손과 얼굴 영역을 모두 포함하는 영역으로 수렴하게 된다. 이 문제를

해결하기 위해 파티클 필터를 이용하여 추적된 얼굴 영역에 해당하는 부분의 값을 작게 해주므로써 Camshift 과정 중에 손 영역에서 수렴하도록 유도했다. 그림 2(나)는 제안 방법에서 얼굴 영역에 마스크를 지정하여 역투영 영상에서 얼굴 영역의 값을 줄이는 방법을 나타내고 있다. 파티클 필터를 통해 얼굴 추적을 수행하고, 그로부터 얻은 얼굴 영역을 얻는다. 얼굴 영역 전체 대신, 실험적으로 정한 얼굴 영역 내 특정영역(그림 2(나)의 얼굴 영역 마스크(Face region mask))의 내부 픽셀값들을 0 으로 만든다. 그림 2(가)와 (다)는 각각 기존 Camshift 의 역투영 영상과 얼굴 영역이 마스크 처리된 제안 방법의 역투영 영상을 보여준다.

3. 실험결과

제안 방법의 효용성을 검증하기 위해 웹캠으로 입력 받은 동영상으로 수행하였다. 동영상 프레임 사이즈는 320 x 240 이며, 동영상 프레임은 총 940 개이다. 기존의 Camshift 만을 사용한 손 추적 성능과 본 논문의 제안 방법의 성능을 비교하기 위해 동일한 웹캠 동영상에 대해서 비교를 수행하였다.

Camshift 에서 HSV 히스토그램 계산 시, 색상(hue), 채도(saturation), 명도(value, brightness)에 대해 각각 16 개의 빈을 설정했고, 최대 반복(maximum iterations) 수는 10 으로 제한하였다. 파티클 필터에서 파티클의 개수는 400 개로 했으며, RGB 히스토그램을 계산할 때, Red, Green, Blue 각 채널에 대해 각각 5 개의 빈을 설정했다. 또한 LBP 는 uniform LBP[6]를 사용했으며, 얼굴 영역에서 위, 아래의 부분을 나누고, 다시 위 부분을 좌, 우 영역으로 나뉘, 총 3 부분에 대해 LBP 히스토그램을 구했다[10]. 이렇게 얻은 RGB 와 LBP 히스토그램에 대해 각각 초기 상태의 RGB 와 LBP 히스토그램과의 상관계수(correlation coefficient)를 계산하여 각 파티클의 가중치를 계산하였다.

그림 3 은 기존의 Camshift 만을 사용한 손 추적과 제안 방법의 손 추적 결과를 비교한다. 왼쪽 열(column)이 Camshift 만을 사용한 손 추적 결과를, 오른쪽 열이 제안 방법의 결과에 해당한다. 그림에서 빨간색 타원이 Camshift 로 추적한 손을, 초록색 박스는 제안 방법에서 파티클 필터로 추적한 얼굴을 나타낸다. 그림 3에서 볼 수 있듯이 기존 Camshift 기법보다 제안 방법이 불필요하게 손 영역에 대한 바운딩이 드리프트 되어 확장되는 것을 방지하는 것을 볼 수 있다.

그림의 각 프레임의 왼쪽 상단에는 프레임 인덱스, 왼쪽 하단에는 계산 속도 (frame per second, fps)가 나타나 있다. 제안 방법이 Camshift 와 파티클 필터를 같이 사용하지만, 거의 대부분의 프레임에서 30 fps 이상의 속도를 유지하였다. 따라서 실시간 처리 응용에 적합하다.

감사의 글

이 논문은 2014 년도 정부(교육부)의 재원으로, 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2011-0011383).

참고문헌

- [1] S. Mitra and T. Acharya, "Gesture Recognition: A Survey," *IEEE Trans. Systems, Man, and Cybernetics, Part C: Applications and Rev.*, Vol. 37, No. 3, pp. 311-324, May 2007.
- [2] A. Mittal, A. Zisserman, and P. H. S. Torr, "Hand detection using multiple proposals," in *British Machine Vision Conference*, 2011.
- [3] G. R. Bradski, "Computer vision face tracking as a component of a perceptual user interface," *Proc. IEEE Workshop Applications of Computer Vision*, pp. 214-219 1998.
- [4] M. S. Arulampalam, et al. "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Processing*, Vol. 50, No. 2, pp. 174-188, 2002.
- [5] P. Viola and M. J. Jones, "Robust real-time face detection," *Int'l J. Comp. Vis.*, Vol. 57, No. 2, pp. 137-154, 2004.
- [6] T. Ojala, et al. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 7, pp. 971-987, 2002.
- [7] Solina, Franc, et al. "Color-based face detection in the" 15 seconds of fame" art installation." (2003): 38-47.
- [8] Dongkyu Kim, Hyung-II Kim, Yong Man Ro, "A Comparative Study of Facial Features Robust to Varying Environment in Particle Filter-based Face Tracking," *Proceedings of Korea Multimedia Society*, Vol. 17, No. 2, 2014.
- [9] M. S. Arulampalam, et al. "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Processing*, Vol. 50, No. 2, pp. 174-188, 2002.
- [10] H.-I. Kim, et al. "Face detection for low power event detection in intelligent surveillance system," *IEEE International Conference on Digital Signal Processing (DSP)*, pp. 562-567, 2014.



그림 2. 얼굴추적 없이 Camshift 로 손 영역을 추적한 결과(왼쪽 열), 얼굴 영역 추적을 사용한 제안방법의 결과(오른쪽 열)

4. 결론

본 논문에서는 얼굴과 손 동작(hand-to-face gesture) 인식을 위해 향상된 손 영역 추적 방법을 제안하였다. 제안 방법은 Camshift 기반의 손 추적에서 손이 얼굴에 가까워질 때 역투영에서 파티클 필터로 얻은 얼굴 영역의 값을 줄임으로써 손 영역의 위치를 정확히 추적할 수 있다.

향후 연구는 본 논문의 제안 방법에 영상 분할(segmentation) 기법을 적용하여 손과 얼굴이 겹칠 때 더 세밀한 손의 모양과 위치를 인식하는 것이다. 이것은 얼굴과 손의 상대적인 위치관계와 손의 모양의 분석을 통한 사용자 인터페이스나 감성인식 연구 등에서 유용하게 활용될 수 있을 것이다.