

# 칼만 필터와 특징 정보를 이용한 손 움직임 추정 개선

서보경<sup>o</sup>, 이장희, 유석인

서울대학교 컴퓨터공학부 인공지능연구실

[bkyung@ailab.snu.ac.kr](mailto:bkyung@ailab.snu.ac.kr), [whizkid@ailab.snu.ac.kr](mailto:whizkid@ailab.snu.ac.kr), [siyoo@ailab.snu.ac.kr](mailto:siyoo@ailab.snu.ac.kr)

## Robust Hand Tracking Using Kalman Filter and Feature Point

Bokyung Seo<sup>o</sup>, Janghee Lee, Suk I.Yoo

Artificial Intelligence Lab, Seoul National University

### 요 약

컴퓨터와 인간과의 상호작용에서 다양한 형태의 인터페이스에 대한 요구가 날로 커지고 있다. 그 가운데 실생활에서도 사물을 지칭하거나 의사소통의 수단으로 사용되는 손과 관련한 인터페이스에 대한 연구가 주목 받고 있다. 기존의 대부분의 연구들은 손을 입력 받으면 영상을 기반으로 손의 중심점을 찾아 그것의 위치를 인식하였는데 이는 물체에 의해 손이 가려진 것과 같이 잘못된 영상을 입력 받았을 때 원하는 결과를 얻지 못하는 상황을 야기할 수 있다. 본 논문은 이러한 점을 보완하기 위하여 손의 중심점을 찾을 때 방해 받는 물체에 덜 민감하게 반응하도록 칼만 필터를 적용하여 문제점을 개선할 수 있도록 하였다. 또한 결과의 정확도를 높일 수 있도록 손가락 끝점을 추출하여 칼만 필터의 매개변수에 반영시켜주었다. 그 결과 예기치 못한 상황이 발생했을 때에도 이것에 덜 민감하게 반응하면서 손의 위치를 비교적 정확하게 측정할 수 있었으며 시스템의 과정이 간단하여 실시간으로 응용하기에 적합한 것을 알 수 있었다.

### 1. 서 론

사용자가 컴퓨터와의 상호작용이 기존의 것보다는 더 빠르고, 쉬운 형태로 제공되기를 원함에 따라 새로운 형태의 인터페이스에 대한 요구가 날로 커지고 있다. 지금까지 인간과 컴퓨터 사이에서 의사소통의 대부분은 마우스와 키보드가 담당해왔다. 하지만 이런 도구의 도움 없이도 음성이나 홍채, 손동작 등을 통해 컴퓨터와 사용자간에 쉽고 자연스러운 의사소통이 이루어질 수 있다. 이와 같은 이유로 음성, 홍채, 얼굴표정, 손동작 등을 이용한 새로운 인터페이스에 대한 연구가 매우 활발하게 진행되고 있다[1,2].

특히 손을 이용한 인터페이스는 실생활에서 손이 사물을 지칭하거나 손동작과 같은 것을 통해 다양한 의사를 표현할 수 있다는 이유로 더 주목 받고 있다. 이 새로운 인터페이스가 잘 동작하기 위해서는 기본적으로 손을 잘 인식해야 한다. 하지만 지금까지 손을 잘 인식해낸 연구들의 대부분은 제한된 환경 속에서 이루어져왔다. 예를 들면 손의 움직임을 추적하는 연구의 경우, 손이 움직일 때 방해가 되는 물체가 존재하지 않는다는 것을 가정한다. 하지만 실생활에서는 매우 다양한 변수가 존재하기 때문에 손의 움직임이 어떤 물체에 의해 방해 받을 수 있다.

본 논문에서는 손의 움직임이 예기치 못한 방해 받는 상황이 발생할 시에도 손의 위치와 움직이는 방향을 비교적 정확하게 추정할 수 있는 모델을

제안하고자 한다.

기존의 대부분의 모델들은 손의 위치를 추적할 때 이진영상에서 손을 흰색 영역이라 가정하면 그 영역의 중심을 손의 중심점으로 여기는 방식이었는데 이는 손에 대한 잘못된 영상을 입력으로 받을 경우 원하는 결과를 얻지 못하는 상황을 발생시킬 수 있다. 제안된 모델은 손의 중심점을 찾는 데 방해 받는 물체에 덜 민감하도록 칼만 필터를 적용하여 문제점을 개선할 수 있도록 하였다. 또한 결과의 정확도를 높일 수 있도록 손가락 끝 점을 추출하여 칼만 필터의 매개변수에 반영시켜주었다.

실험결과 우리의 모델은 다음의 조건들을 만족시키는 것을 확인할 수 있었다. 1) 부분적인 방해에도 덜 민감하게 반응해 비교적 정확하게 측정할 수 있다. 2) 각 과정이 간단하고, 계산속도가 빠르므로 실시간으로 응용하기에 매우 적합하다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구에 대해서 알아보고, 3장에서는 영상에서 피부영역 검출, 해리스 코너 검출을 이용한 특징 정보 추출 및 칼만 필터 설계, 4장은 실험 및 결과를 보여주고, 마지막 장은 결론 및 향후 연구에 대해서 소개하도록 하겠다.

### 2. 관련 연구

#### 2.1. 피부영역 추출

일반적으로 손이나 얼굴과 같은 사물을 인식하기 위해서 피부색상을 이용하여 영상으로부터 분류해내는 연구들이 주로 이루어져 왔다[4,5,6,7,8,9,10,11].

피부영역 추출은 크게 세 가지 방법으로 나누어 볼 수 있다; 1)직접적인 모델; 2)매개변수가 없는 학습을 통한 모델; 3)매개변수가 있는 학습을 통한 모델.

직접적인 모델은 매우 단순하고, 직관적으로 피부색을 추출해낼 수 있다. 하지만 피부색상을 나타내는 적절한 컬러공간을 정의하는 일과 피부로 간주되는 색상범위의 임계값을 결정짓는 것이 어렵다. 최근에는 기계학습을 통해 이러한 조건들을 적절하게 찾아주는 모델이 제안되기도 했다.

두 번째로 매개변수 없는 학습을 통한 모델은 미리 학습을 한 후 분류를 수행하기 때문에 속도가 빠르고 영상의 컬러공간에 대해 독립적이다. 하지만 이것은 적절한 학습데이터와 저장 공간이 필요하고 분류하기 전 학습을 위해 상당히 긴 시간이 소요된다는 단점을 가지고 있다.

매개변수가 있는 학습을 통한 모델도 역시 매우 빠르다는 장점이 있다. 이것은 적은 매개변수로 표현이 가능하고, 저장 공간이 적게 소모된다. 하지만 학습할 때 매우 긴 시간이 소요되고 영상에서 피부의 분포가 모델의 성능에 미치는 영향이 매우 크다.

### 2.2. 특징 정보 추출

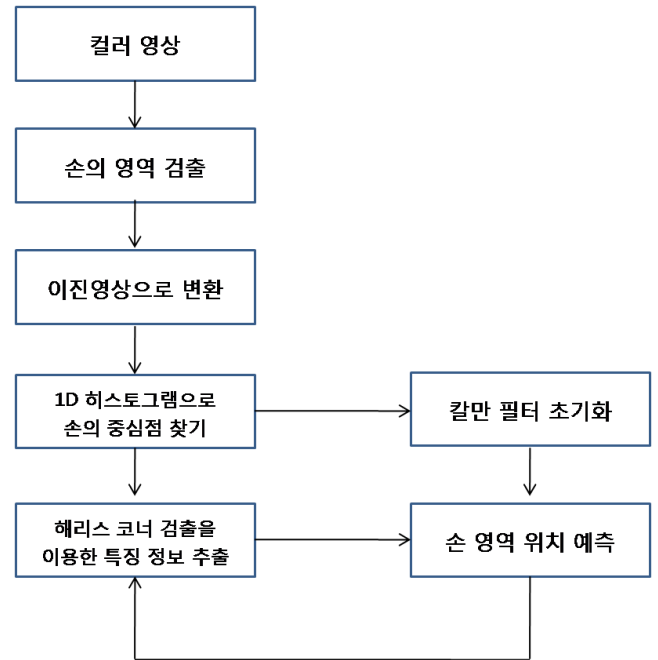
일반적으로 영상을 통해 특정 사물을 인식할 때 그것이 갖는 특징들을 주로 이용한다. 특징은 다음과 같이 크게 세 가지로 나누어 볼 수 있다; 1)일반적인 특징; 2)지역적인 특징; 3)고유한 특징.

일반적인 특징은 영상들이 갖는 공통적인 특징으로 밝기히스토그램, 공분산행렬, 주파수 도메인과 같은 것들을 예로 들 수 있다.

지역적인 특징은 모든 영상이 아닌 특정 영상만이 가질 수 있는 특징으로 테두리, 코너, 선, 곡선과 같은 것들을 예로 들 수 있다. 그 밖에 영상마다 다양한 특징들을 가질 수 있는데 이것들도 물체를 인식하는 하나의 특징으로 매우 유용하게 쓰일 수 있다.

### 3. 손의 위치 추적 시스템

본 논문에서 제안한 모델은 여러 단계로 나뉘어 살펴볼 수 있다. 각 단계는 그림 1과 같은 순서로 수행되며 이 장에서는 각 단계마다 어떠한 일을 하는 지에 대해 자세하게 살펴볼 것이다.



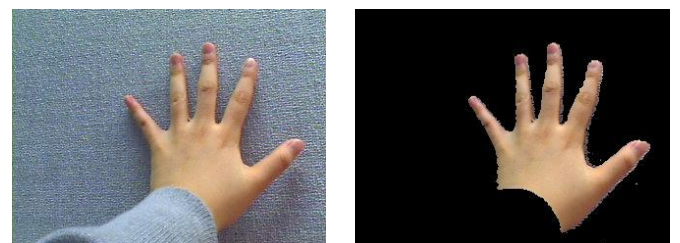
[그림 1] 전체 시스템의 흐름도

#### 3.1. 영상으로부터 손의 영역 검출

직접적인 모델을 이용하면 영상으로부터 매우 빠르고 간단하게 손의 영역을 찾을 수 있다. 직접적인 모델이란 RGB와 같은 컬러공간에서 피부로 간주되는 색상의 범위를 정의해놓은 것이다. 우리는 식(1)을 사용하여 손의 영역을 검출하였다[3].

$$\begin{aligned}
 &R > 95 \text{ and } G > 40 \text{ and } B > 20 \text{ and} \\
 &\text{Max}(R, G, B) - \text{min}(R, G, B) > 15 \text{ and} \\
 &|R-G| > 15 \text{ and } R > G \text{ and } R > B
 \end{aligned} \tag{1}$$

그림 2-(a)는 입력된 영상이고, 그림 2-(b)는 입력된 영상에 직접적인 모델을 적용하여 손의 영역만 추출한 결과이다.



(a) (b)

[그림 2] 영상으로부터 손의 영역 추출. (a) 원본영상 (b)직접적인 모델을 통해 추출된 손의 영역



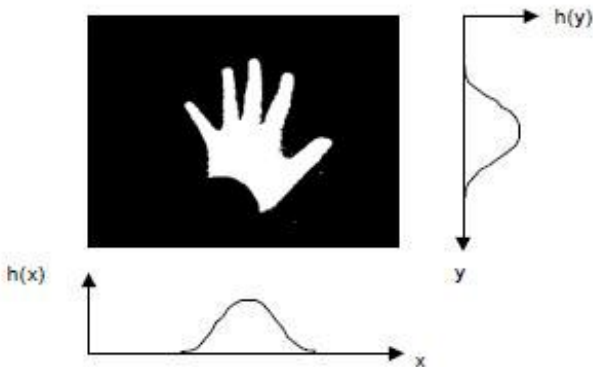
[그림 3] 검출된 손의 영역을 이진영상으로 표현

앞으로 각 단계에서는 손의 영역만이 필요할 것이다. 전체 시스템의 속도를 향상시키기 위하여 최종적으로 손의 영역만 검출한 영상에서 그림 3과 같이 손의 영역은 흰색, 배경영역은 검정색으로 값을 주어 영상을 이진화한다.

### 3.2 손의 중심점 찾기

영상을 각각 x, y축으로 투영시키면 두 개의 일차원 히스토그램을 만들 수 있다. 다르게 표현하면 x축으로 투영시킨 일차원 히스토그램은  $x_i$  일 때 영상에서 x의 위치가 i인 모든 픽셀 값의 합을 값으로 갖는다. 그리하여 그림 4와 같이 두 개의 일차원 히스토그램을 통해 x, y 방향으로 손의 영역에 해당하는 픽셀들이 어떻게 분포되어있는가를 알 수 있다. 또한  $x_i$ 의 일차원 히스토그램의 값을 전체 히스토그램의 값을 합한 것으로 나누면 전체 손의 영역에서 i 번째 픽셀들이 차지하고 있는 비율을 구할 수 있다.

이처럼 x, y의 각각의 위치에서 전체 피부 영역으로부터 차지하는 비율을 알 수 있기 때문에 손의 중심점을 히스토그램 분포의 표본 평균이라고 생각하고 예측할 수 있다.



[그림 4] x와 y 방향의 피부 영역의 분포를 나타낸 일차원 히스토그램

히스토그램 분포를 이용한 손의 중심점은 다음과 같은 식(2)와 (3)을 통해 구할 수 있다.

$$X_c = \frac{\sum_i x_i \times h_x(i)}{\sum_i h_x(i)} \quad (2)$$

$$Y_c = \frac{\sum_i y_i \times h_y(i)}{\sum_i h_y(i)} \quad (3)$$

식(2)에서  $x_i$  는 x 방향으로 i 번째 위치를 의미하고 값은 i를 갖는다. 식(3)에서  $y_i$  도 역시 마찬가지로 y 방향으로 i번째 위치를 의미하고 값은 i를 갖는다.  $h_x(i)$  은 x축 방향으로 i번째 위치할 때 히스토그램의 값이고,  $h_y(i)$  은 y축 방향으로 i번째 위치할 때 히스토그램의 값이다.

### 3.3 특징 추출 및 칼만 필터 설계

3.2절에서 살펴본 손의 중심점을 찾는 방법은 입력영상으로부터 손의 영역을 검출해낸 영상을 기반으로 하기 때문에 잘못된 영상이 입력으로 주어지면 정확한 예측을 하기 힘들다. 잘못된 영상의 예로 손이 다른 물체에 의해 가려졌을 때를 들 수 있는데 실제 손은 움직이지 않아도 가려진 물체에 의해 잘못된 히스토그램을 형성하여 손의 중심점을 추정하는데 큰 오차를 발생시킬 수 있다.

우리 모델은 이러한 오차를 줄여주기 위해서 예측을 방해하는 요소가 나타나더라도 덜 예민하게 반응하면서 손의 위치를 찾아낼 수 있도록 칼만 필터를 적용하였다. 이 시스템은 1초에 20 프레임의 영상을 입력 받으므로 0.05초마다 손의 위치에 대한 정보를 가질 수 있다. 위치는 시간에 따라 변할 수 있지만 0.05초는 매우 순간적인 시간이기 때문에 영상과 영상 사이에서의 움직임의 변화는 크지 않을 것이라고 가정할 수 있다.

또한 더욱더 정확한 예측을 위해 손이 갖는 고유의 특징들을 이용할 수 있다. 손이 퍼져있는 상태에서 그것만이 갖고 특징은 손끝, 손가락 사이가 코너를 이룬다는 것이다. 이런 코너를 나타내는 점을 추출하기 위하여 해리스 코너 검출을 적용하여 손의 특징 정보를 추출하였다[12]. 검출된 점들은 특정 시간에 손끝, 손가락 사이의 위치로 손의 중심에서 일정 거리 이상의 점만 취하여 손가락 사이의 코너는 무시하였다. 최종적으로 얻은 점은 손의 끝점으로 1개에서 5개까지 가질 수 있고, 이 점들의 평균 이동 방향 및 속도가 손이 움직인 방향이나 속도와 유사할 것이라고 가정할 수 있다. 손가락 끝점들의 평균 이동 속도는 칼만 필터의 매개 변수 값으로 사용된다.

칼만 필터는 크게 예측과 갱신 두 단계로 나뉘며 이 두 과정이 계속 반복되고, 아래의 식들을 통해 우리는

현재 좌표의 위치를 예측할 수 있다[13].

$$\hat{x}(t+1|t) = A(t)\hat{x}(t|t) \quad (4)$$

$$\hat{x}(t+1|t+1) = \hat{x}(t+1|t) + K(t+1)(y(t+1) - C\hat{x}(t+1|t)) \quad (5)$$

식(4)에서  $\hat{x}(0|0)$ 은 3.2절에서 살펴봤던 두 개의 일차원 히스토그램을 이용하여 얻은 중심점으로 초기화한다. 또한 식(5)에서  $K(t+1)$ 은 칼만 이득이라고 부르는데 식(6)를 이용하여 값을 얻을 수 있다.

$$K(t+1) = P(t+1|t)C(t)(P(t+1|t)C(t)+R)^{-1} \quad (6)$$

식(6)에서  $P$ 는 상태의 공분산을 나타내는데 이는 아래와 같은 수식으로 갱신된다.

$$P(t+1|t) = AP(t|t)A^T + GQG^T, P(0|0) = P_0 \quad (7)$$

$$P(t+1|t) = P(t+1|t) - K(t+1)CR(t+1|t) \quad (8)$$

위 식에서  $\hat{x}$ 는 현재 시점에서 예측된 손의 중심점을 의미한다.  $\hat{x}$ 는  $x$  좌표,  $x$  방향 속도,  $y$  좌표,  $y$  방향 속도 값을 갖는 일차원 벡터로 표현된다. 앞서 해리스 코너로 검출된 손끝들의 평균 이동 속도를  $\hat{x}$ 의  $x$ ,  $y$ 방향 속도에 매 영상마다 갱신을 해주면 보다 정확한 중심점을 예측할 수 있다.

#### 4. 실험 결과

본 논문에서 제안한 모델은 한 대의 웹 카메라를 이용하여 2.40GHz PC 환경에서 구현되었다. 매 초 300 \* 240의 크기의 영상을 20개씩 입력 받고, 손과 다른 색상을 띤 배경에서 실험을 진행하였다.

그림5는 실험을 통해 얻은 결과로 히스토그램만을 이용해 중심점을 찾은 경우와 칼만 필터를 적용하여 중심점을 예측한 경우를 비교한 것이다. 실제로 입력 받는 영상은 컬러영상이지만 손의 영역을 추출해내는 과정에서 이진영상으로 바꾸어주기 때문에 최종 결과도 이진영상으로 나타난다. 결과영상에서의 붉은색 점은 일차원 히스토그램만을 이용하여 찾은 손의 중심점이고, 푸른색 점은 히스토그램, 해리스 코너 검출 그리고 칼만 필터를 이용하여 예측한 손의 중심점이다.

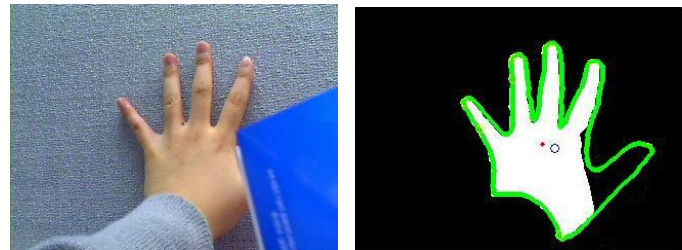
그림 5-(a)는 실험 초기의 영상으로 붉은색 점과 푸른색 점이 일치하고 있음을 알 수 있다. 하지만 그림5-(b)에서 책이 손의 일부를 가리면서 영상에 나타나고 그림5-(c), 그림5-(d)로 갈수록 실제로 손은 거의 제자리에 머무르고 있음에도 불구하고 물체의 방해로 인해 영상에서 손의 영역이 점점 더 가지고 있다. 이 때 히스토그램을 이용하여 찾은 붉은색 점의 경우에는 손의 영역이 책에 의해 일정 부분 가리게 되면서 중심점의 오차가 크게 발생하게 되는 것을 확인할 수 있다. 이와는 대조적으로 푸른색 점은 칼만 필터를 적용하여 비교적 정확하게 중심점을 예측하고



(a)



(b)



(c)



(d)

[그림 5] 칼만 필터를 이용한 손의 위치 추적 결과

있다. 칼만 필터를 적용한 푸른색 점은 일차원 히스토그램으로부터 얻은 중심점을 초기 값으로 하고, 시간에 따라 변화하는  $x$ ,  $y$  좌표의 이동속도, 잡음을 고려하여 중심점을 예측한다. 여기서  $x$ ,  $y$ 좌표의 이동속도는 손가락 끝 점을 추출하여 손가락이 물체에 가려지는 경우도 있기 때문에 추출된 점들의 평균이동속도를 반영하였다. 평균이동속도는 실제 손의 이동속도의 값을 근사화시킨 값으로 오차를 갖고 있을 수 있다. 하지만 이 값은 칼만 필터에 적용되는 하나의 매개변수이고, 오차의 크기 또한 작기 때문에 중심점을 보다 더 정확하게 예측하는데 사용될 수 있다.

#### 5. 결론

본 논문에서는 예기치 못한 상황이 발생해도 그러한 변화에 덜 민감하게 반응하면서 손의 위치를 비교적

정확하게 추적할 수 있는 모델을 제안하였다. 이를 위해 칼만 필터를 적용하고, 결과의 정확도를 좀 더 높일 수 있도록 손가락 끝점을 추출하여 칼만 필터의 매개변수에 반영시켜주었다. 그 결과 손의 위치를 비교적 정확하게 예측할 수 있었으며 시스템의 과정이 간단하여 실시간으로 사용하기에 적합하다는 것을 확인할 수 있었다.

하지만 피부색과 유사한 환경에서 실험할 경우 손의 영역을 정확하게 추출해내는데 다소 어려움이 있었다.

이러한 문제점을 보완하기 위하여 실험배경에 크게 영향을 받지 않으면서 하나의 영상에 여러 개의 손이 나타났을 때에도 정확하게 손을 추적할 수 있는 모델에 대한 연구가 필요할 것으로 생각된다.

### Acknowledgement

이 연구를 위해 연구장비를 지원하고 공간을 제공한 서울대학교 컴퓨터연구소에 감사 드립니다.

### 참 고 문 헌

- [1] R. Feraud, O.J. Bernier, J.-E. Viallet, and M. Collobert, "A Fast and Accurate Face Detector Based on Neural Networks", IEEE Trans. Pattern Analysis and Machine Intelligence, 2001
- [2] T.S. Jebara and A. Pentland, "Parameterized Structure from Motion for 3D Adaptive Feedback Tracking of Faces", Proc. CVPR, 1997.
- [3] Y. Raja, S.J. McKenna, and S. Gong, "Tracking and Segmenting People in Varying Lighting Conditions Using Colour," Proc. Automatic Face and Gesture Recognition, 1998.
- [4] Son Lam Phung, Abdesselam Bouzerdoum and Douglas Chai, "Skin Segmentation Using Color And Edge Information", Proc. Int. Symposium on Signal Processing and its Applications, July 2003.
- [5] Vladimir Vezhnevets, vassili Sazonov and Alla Andreeva, "A Survey on Pixel- Based Skin Color Detection Techniques".
- [6] Hani, K. Almohair, Abd Rahman Ramli and Elsadig A.M, "Skin Detection in Luminance using Threshold Technique", International Journal of The Computer, the Internet and Management Jan-Apr 2007.
- [7] Tiago C, Filipe T and Hamid S, "Improved automatic skin detection in color images", In proc. VIth digital Image Computing: Techniques and Applications, Sunc, Sydney, 2003.
- [8] M.J. Jones and J.M. Rehg, "Statistical Color Models with Application to Skin Detection", Proc. CVPR, 1999.
- [9] J. Yang, W. Lu, and A. Waibel, "Skin-Color

Modeling and Adaptation", Proc. ACCV, 1998.

[10] J.-C. Terrillon, M.N. Shirazi, H. Fukamachi and S. Akamatsu, "Comparative Performance of Different Skin Chrominance Models and Chrominance Spaces for the Automatic detection of Human Faces in Color Images", Proc. Automatic Face and Gesture Recognition, 2000.

[11] K.S. Ravichandran and B. Ananathi, "Color Skin Segmentation Using K-Means Cluster", International Journal of Computational and Applied Mathematics, 2009.

[12] Konstantinos G. Derpanis, "The Harris Corner Detector", 2004.

[13] Greg Welch and Gary Bishop, "An Introduction to the Kalman Filter", 2006.