

연속DP와 칼만필터를 이용한 손동작의 추적 및 인식

문인혁, 금영광
연세대학교 보건과학대학 의공학과,
의용계측 및 재활공학 연구센터, 첨단의료기기 기술혁신센터

Tracking and Recognizing Hand Gestures using Kalman Filter and Continuous Dynamic Programming

Inhyuk Moon and Youngkwang Kum
Dept. of Medical Engineering, College of Health Science,
Research Institute for Medical Instruments and Rehabilitation Engineering,
Technology Innovation Center for Medical Instruments, Yonsei University
E-mail : ihmoon@dragon.yonsei.ac.kr

Abstract

This paper proposes a method to track hand gesture and to recognize the gesture pattern using Kalman filter and continuous dynamic programming (CDP). The positions of hands are predicted by Kalman filter, and corresponding pixels to the hands are extracted by skin color filter. The center of gravity of the hands is the same as the input pattern vector. The input gesture is then recognized by matching with the reference gesture patterns using CDP. From experimental results to recognize circle shape gesture and intention gestures such as "Come on" and "Bye-bye", we show the proposed method is feasible to the hand gesture-based human-computer interaction.

I. 서론

제스처는 장애인이나 고령자들에게 언어 이외의 의사 전달 수단으로써 사용될 수 있다. 특히 손 제스처는 청각장애자들에게 수화로서 언어로 사용되고 있다.

최근 컴퓨터의 성능이 발전함에 따라 영상을 이용한 손동작 인식방법이 활발히 연구되고 있다[1]. 이러한 연구의 특징으로는 특별한 장치를 인간에게 부착하지 않고도 제스처를 인식할 수 있는 장점을 갖는다.

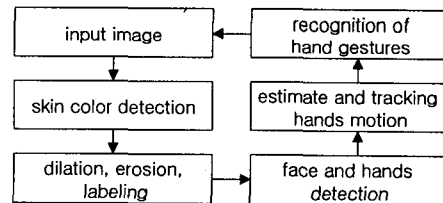


그림 3. 손동작 추적 및 인식 흐름도

본 논문은 단일 컬러 CCD카메라를 이용하여 입력된 컬러 영상에서 칼만 필터(Kalman filter)를 이용하여 손의 움직임을 예측하고, 색 정보로부터 손영역을 검출하는 방법을 제안한다(그림 1참조). 검출된 영역의 중심위치는 입력 제스처의 시계열 패턴이므로 연속 동적 계획법(continuous dynamic programming, CDP)을 이용한 입력 제스처를 인식하는 수법을 제안한다. 인식 가능한 제스처로는 원, 네모, 세모의 형태를 표현하는 동작, 상하좌우의 방향 지시 동작, 그리고 Come on, Bye-bye 등의 손짓 동작이다. 실험을 통하여 제안한 수법이 손동작의 추적과 인식에 유용함을 보인다.

II. 손동작 운동모델과 칼만 필터

2.1 손동작 운동모델

본 논문에서는 손의 운동은 3차원 공간에서 영상평면(image plane)에 평행한 병진 운동(translation motion)성분 만으로 이루어진다고 가정한다.

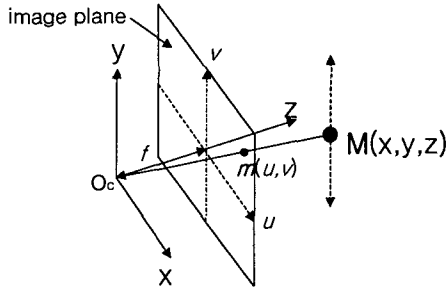


그림 4. 카메라의 정사영 모델

그림 4는 카메라의 정사영(perspective projection) 모델을 나타낸다. 3차원 공간에서의 손의 중심위치 $M = [X \ Y \ Z]^T$ 은 영상에서 $m = [u \ v]^T$ 의 위치에 관측된다. 이 정사영 관계는 다음과 같다.

$$u = f \frac{X}{Z}, \quad v = f \frac{Y}{Z} \quad (1)$$

여기서 f 는, 카메라의 초점(focal point) O_c 와 영상평면 간의 초점거리(focal length)를 나타낸다.

손동작의 상태벡터 $x = [X \ Y \ X \ Y]^T$ 는 3차원 공간에서의 위치와 속도로서 정의하고, 손동작은 등속도로 운동한다고 가정한다. 그리고 손의 위치는 정사영으로 $y = [u \ v]^T$ 에 관측되어진다고 한다. 이 관계는 다음과 같이 손동작의 선형 상태방정식 (2)과 관측식 (3)으로 나타낼 수 있다.

$$x(t+1) = A x(t) + v_t \quad (2)$$

$$y(t) = C x(t) + w_t \quad (3)$$

여기서 v_t 와 w_t 는 공분산이 각각 V_t 와 W_t 인 손동작의 운동과 영상관측에 대한 평균 0인 백색잡음(white noise)이고, 상수 계수행렬 A 와 C 는 다음과 같다.

$$A = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad C = \frac{f}{Z} \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \end{bmatrix} \quad (4)$$

행렬 C 에서 초점거리 f 는 고정이고 손동작은 영상평면에 평행하게 이루어지기 때문에, 행렬 C 는 다음과 같이 나타낼 수 있다. 단, $\lambda = \frac{f}{Z}$ 이다.

$$C = \begin{bmatrix} \lambda & 0 & \lambda \Delta t & 0 \\ 0 & \lambda & 0 & \lambda \Delta t \end{bmatrix} \quad (5)$$

2.2 칼만 필터를 이용한 손동작 추적

임의의 시간 $t-1$ 에서의 상태 $x(t-1)$ 가 주어지면, 상태방정식 (2)와 관측식 (3)으로부터 상태 $\hat{x}(t)$ 와 관

측 $\hat{y}(t)$ 을 예측할 수 있다. 시간 t 에서 관측결과 $y(t)$ 가 얻어지면, 관측오차를 고려하여 예측상태 $\hat{x}(t)$ 로부터 시간 t 에서의 상태 $\hat{x}(t)$ 를 추정할 수 있다.

$$\hat{x}(t) = \hat{x}(t) + K_t [y(t) - C \hat{x}(t)] \quad (6)$$

여기서 K_t 는 칼만 이득(Kalman gain)이고, 예측공분산 P_t 과 관측잡음의 공분산 W_t 을 가지고 구할 수 있다.

$$K_t = P_t C^T [C P_t C^T + W_t]^{-1} \quad (7)$$

그리고, P_t 와 추정공분산 P_t 은 식 (8)(9)에 의해 계산된다.

$$P_t = E[x(t+1)[x(t+1)]^T] = A P_t A^T + V_t \quad (8)$$

$$P_t = P_t - K_t C P_t \quad (9)$$

식 (6)(9)의 추정된 상태 $\hat{x}(t)$ 와 공분산행렬 P_t 로부터 손의 3차원 위치를 확률적으로 알 수 있고, 식 (2)(8)에 적용하면 $\hat{x}(t)$ 와 P_t 을 얻을 수 있다. 따라서 현재의 손동작의 위치는 $\hat{x}(t)$ 로부터 구하고, 손동작 추적을 위한 탐색영역의 중심과 그 범위는 $\hat{x}(t)$ 와 예측공분산 P_t 로부터 계산 할 수 있다. 본 연구에서는 평균으로부터 $\pm 3\sigma$ 의 범위를 탐색영역으로 한다.

III. 손 검출을 위한 영상관측

3.1 피부색 영역검출

입력된 RGB칼라 영상을 다음과 같이 선형변환행렬을 이용하여 휘도성분이 분리된 YCbCr의 색좌표로 변환한다.

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (10)$$

여기서 Y 는 휘도 성분을, Cb 와 Cr 은 각각 파란색과 붉은색이 강조된 색좌표이다. 피부색은 Y 값이 변화하더라도 Cb , Cr 값은 크게 변화하지 않기 때문에[2], 피부색 문턱치(threshold value)는 피부색의 분포가 가우시안 분포를 따른다고 가정하여 확률적으로 구한다. 먼저, 입력영상의 일정 피부영역을 선택하고 YCbCr 색좌표로 변환한다. 그리고 선택영역의 Cb , Cr 의 평균 (μ_{Cb} , μ_{Cr})과 표준편차 (σ_{Cb} , σ_{Cr})를 구한다. 본 연구에서는 피부색 문턱치를 평균을 중심으로부터 $\pm 3\sigma$ 영역으로 정한다.

피부색 문턱치를 이용하여 설정된 탐색영역에서 피



그림 5. 입력된 컬러 영상과 손과 얼굴의 탐색영역



그림 6. 검출된 손과 얼굴 영역

부색을 가진 픽셀만을 검출할 수 있다. 하지만 고립점과 같은 잡음이 포함되므로, 수축(dilation) 및 확장(erosion) 기법을 적용하여 잡음을 제거한다. 그리고 라벨링(labeling) 기법으로 일정 크기 이상의 영역만을 가진 군집(cluster)만을 선택하고, 손과 얼굴에 해당하는 군집을 검출한다. 그림 6은 입력된 컬러 영상과 손과 얼굴의 탐색영역을 보이며, 그림 6은 탐색영역내에서 손과 얼굴에 해당하는 영역을 검출한 결과이다.

3.2 손과 얼굴 영역의 중심위치 검출

검출된 손과 얼굴 영역의 중심(center of gravity) m_x, m_y 은 식 (11)(12)의 2차원 관성 모멘트(inertia moment)식을 이용하여 구한다.

$$m_x = \frac{1}{N} \sum_{i \in x} \sum_{j \in y} x_i p(x_i, y_j) \quad (11)$$

$$m_y = \frac{1}{N} \sum_{i \in x} \sum_{j \in y} y_j p(x_i, y_j) \quad (12)$$

여기서 $p(x_i, y_j)$ 는 영상의 (x_i, y_j) 좌표에서 피부색이 존재하면 1, 없으면 0을 의미한다. N 은 손 혹은 얼굴 영역을 구성하는 총 픽셀 개수이다. 그림 6은 손과 얼굴 영역의 중심을 검출한 결과를 나타낸다.

일단 손과 얼굴의 초기 탐색영역이 주어지면, 칼만 필터의 결과로부터 연속적으로 손과 얼굴의 탐색영역을 얻을 수 있다. 또한 앞 프레임의 추적결과로부터 탐색영역을 매 프레임마다 계산하기 때문에, 얼굴영역과 손 영역이 겹치는 경우에도 손의 추적이 가능하다.

IV. 연속 DP를 이용한 손동작의 인식

손동작 인식 방법으로 정합 경로가 제한된 연속 DP(continuous dynamic programming, CDP)를 사용한다. CDP는 매 프레임마다 참조동작패턴(reference gesture pattern)계열과 정합하여 패턴간의 유사도를 계산하기 때문에 입력동작을 구역화(segmentation)할 필요가 없다는 장점을 갖는다[3].

한 개의 모델 M 은 참조패턴으로서 T 프레임간 입력



그림 7. 중심 검출 결과

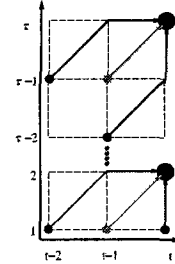


그림 8. 정합 경로가 제한된 CDP

된 영상으로부터 얻어진 N 차원의 특징벡터 m_r 의 계열로 나타낸다.

$$M = \{m_r | 1 \leq r \leq T\} \quad (13)$$

마찬가지로 연속적으로 입력된 영상으로부터 특징벡터 계열 u_t ($0 \leq t \leq \infty$)을 얻을 수 있다. 참조패턴과 입력패턴간의 정합(matching)의 유효성은 특징벡터간의 누적거리로서 계산한다. 시간 t 에 입력된 특징벡터와 참조패턴의 r 번째 특징벡터간의 거리 $d(t, r)$ 는 식 (14)와 같이 정의하고, 시간에 따라 거리를 누적한다.

$$d(t, r) = \frac{1}{N} \sum_{k=0}^N [(u_t(k) - m_r(k))^2] \quad (14)$$

시간 (t, r) 에서의 누적 거리를 $S(t, r)$ 라고 하면, $S(t, r)$ 은 r 에 따라 다음의 3가지 형태의 점화식으로 누적된다(그림 8 참조).

$$S(t, 1) = 3d(t, 1) \quad (15)$$

$$S(t, 2) = \min \begin{cases} S(t-2, 1) + 2d(t-1, 2) + d(t, 2) \\ S(t-1, 1) + 3d(t, 2) \\ S(t, 1) + 3d(t, 2) \end{cases} \quad (16)$$

$$S(t, r) = \min \begin{cases} S(t-2, r-1) + 2d(t-1, r) + d(t, r) \\ S(t-1, r-1) + 3d(t, r) \\ S(t-1, r-2) + 2d(t, r-1) + d(t, r) \end{cases} \quad (17)$$

단, 초기 누적거리는 $S(-1, r) = S(0, r) = \infty$ ($1 \leq r \leq T$)로 한다. 식 (17)은 $3 \leq r \leq T$ 구간에서의 갱신 점화식이다.

연속된 입력된 패턴에 대해 각 참조패턴과의 누적거리 $S(t, r)$ 를 구하고, 전체 참조 패턴에 대한 거리 $3T$ 로 정규화시킨 $A(t) = \frac{1}{3T} S(t, T)$ 를 계산한다. $A(t)$ 가 문턱치 이하가 되면, 입력패턴은 해당하는 참조패턴으로 인식된다.

V. 실험 및 결과

5.1 손동작 패턴

약 1m의 거리에 떨어져 손동작을 수행하고, 컬러 CCD 카메라로 320×240 크기의 영상을 입력하였다.

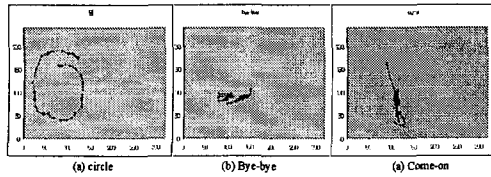


그림 9. 참조동작패턴: (a) circle (b) Bye-bye (c) Come on

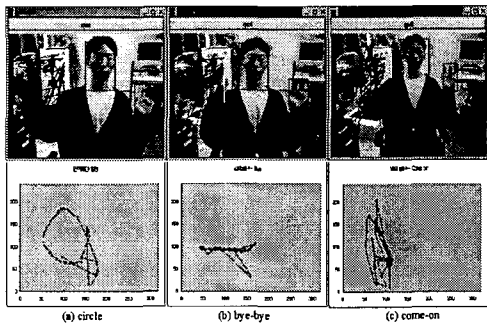


그림 10. 입력동작패턴: (a) circle (b) Bye-bye (c) Come on

본 연구에서는 얼굴과 양손에 대한 중심을 추적하지만, 오른손만이 동작하는 제스처를 대상으로 실험을 수행하였다. 본 실험에서 적용한 참조동작패턴으로는 원모양, "Bye-bye", 그리고 "Come on" 제스처이다(그림 10 참조).

그림 10은 연속 입력된 영상에서 CDP에 의해 손동작이 인식될 때까지 추적한 결과를 보여 준다. 각각 원, "Bye-bye", 그리고 "Come on"으로 정의된 손동작을 입력하였고, 그림 8에서 각 그래프는 추적된 좌표를 나타낸다.

5.2 손동작 인식결과

그림 10의 손동작중 한 동작이 입력되었을 때, CDP에 의해 매 프레임마다 참조동작패턴과 정합하였다.

그림 10(a)는 그림 10(b)의 "Bye-bye" 손동작 입력에 대해 각 참조동작패턴과 매 프레임 정합했을 때의 정규화된 누적거리 $A(t)$ 를 보여준다. 수평축은 프레임 수, 수직축은 $A(t)$ 이다. 이 결과로부터 "Bye-bye" 참조패턴과의 정규화 누적거리 $A(t)$ 가 최소임을 보여준다.

그림 10(b)는 그림 10(c)의 "Come on" 손동작 입력에 대한 정합결과이다. 입력동작패턴에 해당하는 "Come on" 참조패턴과의 정규화 누적거리가 최소인 것을 보여준다. 그러나 원모양의 손동작과 "Come on" 동작은 위아래로 손이 왕복하는 것이 비슷하기 때문에, 그 누적거리의 차가 작음을 알 수 있다.

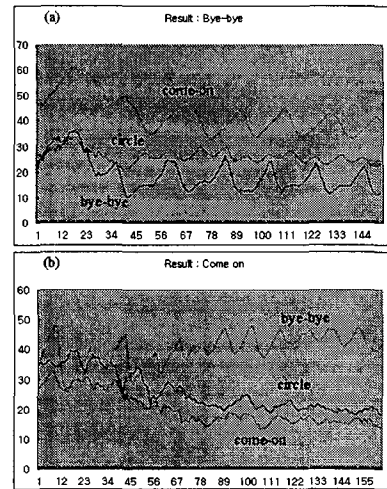


그림 11. CDP에 의한 인식결과: (a) Bye-bye (b) Come on

VI. 결론

본 논문에서 연속으로 입력된 손동작을 칼만 필터를 이용하여 연속적으로 추적하는 수법을 제안하였다. 또한 추적된 손의 위치를 CDP를 이용하여 참조동작패턴과 구역화없이 정합할 수 있는 수법을 제안하였다. 확률적 예측을 바탕으로 추적하고, 입력동작패턴에 대해 구역화없이 온라인으로 정합 및 인식이 가능하였다. 또한 320x240 크기의 동영상에 대해 5[frame/sec]의 처리속도를 보였고, 휴먼 컴퓨터 인터페이스에 적용 가능하다는 것을 보여 주었다.

현재는 손동작만을 인식하지만 손가락을 포함한 다양한 손의 동작을 인식하도록 확장한다면, 향후 수화 인식 시스템에 적용할 수 있을 것이다.

후기

본 연구는 과학기술부·한국과학재단 지정 연세대학교 의용계측 및 재활공학 연구센터의 지원에 의한 것입니다.

참고문헌

- [1] M. Yeasin and S. Chaudhuri, "Visual Understanding of Dynamic Hand Gestures," Pattern Recognition, Vol. 33, pp. 1805-1817, 2000.
- [2] R. Hsu, M. A. Mottaleb, and A. K. Jain, "Face Detection in Color Images," IEEE Trans. PAMI, Vol. 24, No. 5, pp. 696-706, 2002.
- [3] R. Oka, "Spotting Method Approach Towards Information Integration," Proc. of Real World Computing Symposium, pp. 175-182, 1997.