

확장 칼만 필터를 이용한 얼굴의 3차원 움직임량 추정

한 승철, 박 강령, 김 재희
연세대학교 전자공학과

3-D Facial Motion Estimation using Extended Kalman Filter

Seung Chul Han, Gang Ryung Park, Jaihie Kim
Dept. of Electronics Engineering, Yonsei university

Abstract

In order to detect the user's gaze position on a monitor by computer vision, the accurate estimations of 3D positions and 3D motion of facial features are required. In this paper, we apply a EKF(Extended Kalman Filter) to estimate 3D motion estimates and assumes that its motion is "smooth" in the sense of being represented as constant velocity translational and rotational model. Rotational motion is defined about the origin of an face-centered coordinate system, while translational motion is defined about that of a camera centered coordinate system.

For the experiments, we use the 3D facial motion data generated by computer simulation. Experiment results show that the simulation data and the estimation results of EKF are similar.

1. 서론

사용자 얼굴의 움직임을 추적하는 연구들은 컴퓨터 비전과 휴먼 컴퓨터 인터페이스 등에서 많은 응용범위를 가지고 최근 활발히 연구가 진행되고 있다. 이 중 컴퓨터 비전 방법에 의해 사용자 응시위치를 파악하는 연구는 많은 응용 분야를 가지고 있는데 그 대표적인 예로 화상회의 시스템에서 사용자의 응시대상자에 따라 원격 카메라를 자동으로 포커싱하는 시스템, 그리

고 공장 자동화 시스템에서 관리자의 양손을 입력 수단으로 사용하지 못하는 상황에서 사용자의 응시 위치에 의해 프로세스 제어를 하는 시스템 등이 있다.

이와 같이 컴퓨터 비전 방법에 의해 모니터 상의 사용자 응시위치를 파악하기 위해서는 사용자의 얼굴 특징점(양눈, 양코웃음, 입의 양끝점등)의 정확한 3차원 위치를 알아야 하고 동시에 이들의 3차원 움직임을 정확하게 실시간으로 추정(estimate)할 수 있는 알고리즘이 필수적으로 요구된다.

기존에 사용된 3D 움직임 추정 알고리즘으로 신경망(Neural Network), 광류 분석법(optical flow methods) 및 확장 칼만 필터(Extended Kalman Filter) 등이 있는데, 이 중 신경망과 광류분석법은 처리시간이나 정확도 면에서 많이 떨어지는 단점 때문에 확장 칼만 필터를 많이 이용하고 있다[1][2][6][8]. 이에 따라 본 논문에서는 확장 칼만 필터를 이용하여 3차원 얼굴 움직임을 실시간으로 추정하는 연구를 수행하였다.

2차원 카메라 영상을 이용하여 얼굴 특징점의 3차원 정보를 얻어 낼 때 깊이 정보를 알기 어렵기 때문에 몇 가지 제약이 필요하다. 기존의 연구에서는 얼굴의 움직임은 갑작스러운 변화가 없이 부드럽게 움직인다는 제약을 두고 짧은 시간 동안의 얼굴 이동(translation)과 회전(rotation)은 등속도 운동으로 가정하여 문제를 풀어 나갔다[1][2]. 이러한 과정에서 반복적으로 이전 값을 이용해 현재 값을 추정하는 칼만 필터를 사용하였다. 그런데 3차원에서의 얼굴 움직임은 2차원 카메라 영상 정보와 비선형적인 관계에 있기 때문에 이산 칼만 필터를 사용하지 않고 확장 칼만 필터를 사용하였다[8][9]. 확장 칼만 필터는 비선형 관계를

선형화시킨 필터이기 때문에 자체적으로 에러 요인을 가지고 있다. 이러한 요인으로 초기에 확장 칼만 필터에 넣어주는 초기 얼굴 움직임 변수를 잘못 설정하면 값값에 수렴하지 못하고 발산을 하게 된다. 이러한 문제를 보완하기 위해 반복적 확장 칼만 필터(Iterated Extended Kalman Filter)를 사용하였다[6][7][10].

본 논문에서는 얼굴의 움직임 데이터를 컴퓨터 시뮬레이션을 통해 얻어내었다. 이 시뮬레이션 데이터를 바탕으로 여러 가지 변수를 바꾸어 가며 실험을 하였는데 이러한 실험을 통해 확장 칼만 필터를 이용한 얼굴 움직임 추정 방법의 문제점을 고찰하였다.

2. 얼굴의 3차원 움직임량을 추정하기 위해 사용하는 모델

본 논문에서 얼굴의 카메라 좌표계에 대한 움직임을 이동과 회전으로 분리하여 얼굴의 이동은 얼굴 좌표계 원점의 이동으로, 얼굴의 회전은 얼굴 축을 기준으로 한 회전으로 정의한다. 이러한 얼굴의 움직임을 정의하기 위해 3차원상의 얼굴 특징점을 카메라 중심 좌표계와 얼굴 중심 좌표계 두 부분으로 나누어 나타내었다. 이렇게 얼굴의 특징점을 두 부분으로 나타냄으로써 얼굴의 회전을 카메라를 기준으로 본 것이 아니라 얼굴 축을 기준으로 하여 일관성 있는 변수로 모델링할 수 있는 장점이 있다. 그림 1은 본 논문에서 사용하는 3차원 얼굴 좌표계와 카메라 좌표계를 나타낸 것이다.

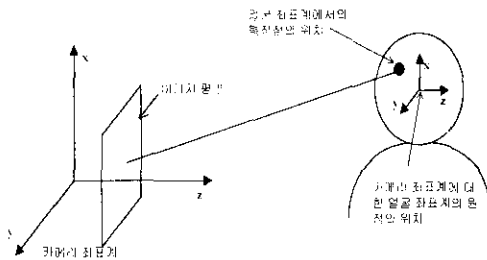


그림 1. 3차원 얼굴의 모델과 투영

2.1 카메라 모델

3차원 상에서의 얼굴의 특징점은 카메라의 이미지 평면에 원근 투영 변환(perspective transform)에 의해 맺히게 된다. 이 변환 관계는 다음과 같다.

$$X = f \cdot \frac{x}{z} \quad Y = f \cdot \frac{y}{z} \quad (1)$$

x, y, z 는 카메라 좌표계를 기준으로한 3차원 상에서의 얼굴의 특징점의 좌표이고, X 와 Y 는 그 점에 해당하는 카메라의 이미지 평면에 맺힌 좌표이다. f 는 카메라의 초점거리이다. 측정과 처리하는 과정에서 잡음이 섞이는 것을 감안하여야 하므로 위의 식은 다음과 같이 모델링할 수 있다.

$$X = f \cdot \frac{x}{z} + n_x \quad Y = f \cdot \frac{y}{z} + n_y \quad (2)$$

n_x 와 n_y 는 이미지 평면에서의 잡음으로 평균이 0인 독립적 가우스분포(independent zero mean Gaussian distribution)로 모델링할 수 있다.

2.2 얼굴의 3차원 움직임 모델

얼굴의 움직임을 이동과 회전으로 분리하기 위해 3차원상의 얼굴 특징점의 좌표를 카메라 중심 좌표계와 얼굴 중심 좌표계로 나누어 나타낸다. 얼굴의 이동은 카메라 좌표계를 기준으로 하여 얼굴 좌표계 원점의 이동으로 얼굴의 회전은 얼굴 좌표계에서 얼굴 축을 기준으로한 얼굴의 회전으로 정의한다. p_i 를 얼굴 좌표계에서 특징점의 좌표 벡터라고 하고 s 를 카메라 좌표계에서 얼굴 좌표계 원점의 좌표 벡터라 하면 관계식은 다음과 같다. i 는 특징점의 개수를 나타낸다.

$$P_i = s + R p_i \quad (3)$$

P_i 는 카메라 좌표계에서 얼굴 특징점의 좌표이고, R 은 3×3 회전 행렬이다. P_i 는 위와 같이 이동과 회전의 합으로 나타내어진다.

2.3 이동 움직임 모델(translational motion model)

얼굴의 이동은 앞서 가정대로 갑작스러운 변화가 없이 서서히 움직이는 것을 가정하므로 가속도를 무시하고 등속도 운동으로 간주하여 다음과 같이 모델링할 수 있다.

$$s = s_0 + v \quad (4)$$

s_0 는 카메라 좌표계에서 초기 얼굴 좌표계 원점의

좌표 벡터이고, \mathbf{v} 는 원점의 속도 벡터이다.

2.4 회전 움직임 모델(rotational motion model)

얼굴의 회전은 3차원 상에서 얼굴 좌표계를 기준으로 회전하는 것으로 정의한다. 이 논문에서는 쿼터니언(quaternion)을 사용하여 시간에 따른 얼굴의 회전을 나타내었다. 보통 물체의 회전을 나타낼 때 오일러 각도를 사용하는데 이것은 물리적으로 해석하기가 쉬운 반면 시간에 따른 연속적인 회전을 나타내는데 어려운 단점이 있다. 그러나 쿼터니언은 시간에 따른 연속적인 회전과 고차원 미분에서 시간의 변화에 대한 관계식을 유도하기가 용이하여 본 논문에서는 쿼터니언을 사용하여 얼굴의 회전을 나타내었다[13].

3. 확장 칼만 필터 식

본 논문에서는 앞에서 기술한 얼굴의 3차원 움직임 모델을 바탕으로 확장 칼만 필터를 이용하여 얼굴의 3차원 움직임량을 추정하였다.

3.1 시간 갱신 식

확장 칼만 필터에서 시간 갱신 식은 상태 벡터 ($\hat{\mathbf{x}}_k$)사이의 관계가 선형적이므로 프레임사이의 관계를 상대천이행렬(Φ_k)로 나타낼 수 있다.

$$\hat{\mathbf{x}}_{k+1} = \Phi_k \hat{\mathbf{x}}_k \quad (5)$$

$$\hat{\mathbf{x}}_k = [\mathbf{s}_k \ \mathbf{v}_k \ \mathbf{q}_k \ \boldsymbol{\omega}_k \ \mathbf{p}_{1,k} \ \mathbf{p}_{2,k} \ \mathbf{p}_{3,k}]^T \quad (6)$$

\mathbf{s}_k 는 카메라 좌표계에서 얼굴 좌표계 원점의 좌표 벡터이고, \mathbf{v}_k 는 카메라 좌표계에서 얼굴 좌표계 원점의 이동속도벡터이다. \mathbf{q}_k 는 쿼터니언이고, $\boldsymbol{\omega}_k$ 는 얼굴 좌표계에서 x, y, z축 각각의 회전 각속도를 나타내는 회전 각속도벡터이다. $\mathbf{p}_{1,k}$, $\mathbf{p}_{2,k}$, $\mathbf{p}_{3,k}$ 는 각각 얼굴 좌표계에서 특징점의 위치벡터이다.

3.2 측정 식

3차원 상에서의 얼굴 특징점의 좌표를 2차원 이미지도 투영시키는 과정은 비선형적인 관계이므로 행렬 형태로 나타내지 못하고 다음과 같이 함수($h(\cdot)$)의 형태로 나타낸다. 이러한 과정에서 측정잡음이 섞이게

되는데 본 논문에서는 잡음 변수를 독립적 평균이 0인 가우스 분포로 가정하였다.

$$\mathbf{z}_k = h(\mathbf{x}_k) + \mathbf{v}_k \quad (7)$$

3.3 확장 칼만 필터 식

위에서 기술한 식을 바탕으로 다음의 (8), (9), (10), (11), (12)와 같은 확장 칼만 필터식을 이용하여 얼굴의 3차원 움직임량을 추정한다.

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{z}_k - \hat{\mathbf{z}}_k^-) \quad (8)$$

$$\mathbf{K}_k = \mathbf{P}_k \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \quad (9)$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^- \quad (10)$$

$$\mathbf{P}_{k+1} = \Phi_k \mathbf{P}_k \Phi_k^T + \mathbf{Q}_k \quad (11)$$

\mathbf{P}_k 는 추정오차 공분산 행렬(estimation error covariance matrix)이고, \mathbf{R}_k , \mathbf{Q}_k 는 각각 측정(measurement) 잡음과 과정(process) 잡음이다.

\mathbf{H}_k 는 $h(\cdot)$ 함수를 상태 벡터에 관하여 편미분한 자코비안 행렬이다[3][4][5][11].

$$\mathbf{H}_k = \left[\frac{\partial h}{\partial \mathbf{x}} \right]_{\mathbf{x} = \hat{\mathbf{x}}_k} \quad (12)$$

4. 실험 및 결과

본 논문에서는 확장 칼만 필터의 성능을 시뮬레이션을 통해서 얻은 데이터를 이용하여 테스트하였다. 그림 2는 얼굴이 y축을 기준으로 회전하였을 때 시간에 따른 확장 칼만 필터 추정된 값과 시뮬레이션으로 생성한 실제값 비교하여 나타내었다.

그림 2에서 보면 에러가 있는 회전각이 확장 칼만 필터를 통과하여 시간이 변화함에 따라 실제 회전각에 가까워져 가는 것을 알 수 있다.

5. 결 론

컴퓨터 비전 방법으로 모니터 상의 사용자 응시위치를 파악하기 위해서는 사용자의 얼굴 특징점의 정확한 3차원 위치와 이들의 3차원 움직임을 정확하게 실시간으로 추정하여야 하는데 본 논문에서는 확장 칼만 필터를 사용하여 연구를 수행하였다. 실험을 통해서

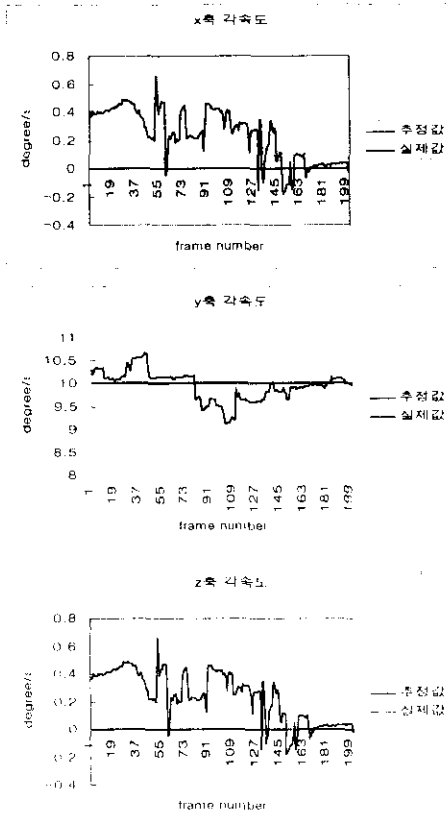


그림 2. 확장 칼만 필터를 사용하여 추정된 회전각속도

확장 칼만 필터 추정값과 시뮬레이션으로 생성한 실제 값을 비교 연구하였다. 확장 칼만 필터 알고리즘은 초당 10 프레임 이상의 이미지를 처리할 수 있었고, 상승 평균 평방근(root-mean-square) 에러가 0.5 degree/s로 좋은 성능을 나타내었다.

이 논문은 1998년도 한국 학술진흥 재단의 대학 부설연구소 연구비 지원에 의하여 연구되었음.

(연구 과제명: 가상 현실 시스템에서 얼굴/눈의 움직임에 의한 pointing에 관한 연구)

참고 문헌

[1] Ted J. Broida and Rama Chellappa, "Estimation of object motion parameters from noisy image," IEEE Trans. Pattern Analysis and Machine Intelligence, 8(1):90-99, Jan 1986

[2] Ted J. Broida and Rama Chellappa, "Estimating

the kinematics and structures of a rigid object from a sequence of monocular images," IEEE Trans. Pattern Analysis and Machine Intelligence, 8(1):90-99, Jan 1986

[3] Robert Grover Brown and Patric Y.C. Hwang, Introduction to Random Signal Analysis and Kalman Filtering, John Wiley & Sons, New York, 1997

[4] Simon Haykin, Adaptive filter theory, Prentice hall international editions 1996

[5] Mohinder S. Grewal Angus P. Andrews, Kalman Filtering Theory and practice, Prentice hall information and system science series 1993

[6] A. Azarbayeani. T. Starner. B. Horowitz. A. Pentland, "Visually Controlled Graphics," IEEE PAMI vol. 15, No. 6, pp. 602-605, June 1993

[7] Ali Azarbayejani and Alex P. Pentland, "Recursive Estimation of Motion, Structure, and Focal Length," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 17, No 6, June 1995

[8] A. Azarbayeani, B. Horowitz, and Alex P. Pentland, "Recursive Estimation of Structure and Motion using Relative Orientation Constraints," 1993 IEEE Conf. on CVPR IEEE Computer Society Press, pp.294-299, June 1993

[9] Ted J. Broida and Rama Chellappa, "Recursive estimation of 3D motion from a monocular image sequence," IEEE Trans. Aerospace and Electronics Systems, vol. 26 No. 4, pp 639-656, July 1990

[10] Maybeck, P. S., "Stochastic Models, Estimation, and Control, vol. 2, New York: Academic Press, 1922

[11] Gem-sun J. Young and Rama Chellappa, "3-D Motion Estimation Using a Sequence of Noisy Stereo Images: Models, Estimation, and Uniqueness Results

[12] Friedland, B., "Analysis of strapdown navigation using quaternions," IEEE Trans. on Aerospace and Electronic Systems, AES-14, p764-768, Sept. 1978