반복적 확장 갈만 필터를 이용한
얼굴의 3차원 움직임량 추정

박 강 령*・김 재 회**

요 악

 컴퓨터 시각 인식 방법을 이용하여 얼굴의 3차원 움직임량을 추정하고자 하는 연구는 가상 현실 환경에서 얼굴 움직임에 의한 3차원 그래픽 화면의 품질을 높이기 위해 필수적으로 요구되는 기술로서 최근 활발히 연구되고 있다. 기존의 3차원 움직임량을 추정하고자 하는 연구들은 대부분 확장 갈만 필터(extended kalman filter)를 이용하였으나, 이러한 방법은 필터의 초기값을 정확하게 설정해야 하는 제약 요소를 갖고 있으며, 또한 얼굴의 회전 방향 변환에 있어서 대칭하지 못하는 경우 역시 종종 발생한다. 본 논문에서는 이러한 문제점을 해결하기 위하여 확장 갈만 필터의 형태변화 밀도변화 갈만 필터를 이용하여 얼굴의 3차원 움직임량을 추정하였다. 반복적 확장 갈만 필터에서는 확장 갈만 필터에서 계산되어 나오는 추정 오차 공간은 연속적이며 미리 정해진 임계치보다 커지는 경우, 현재 얼굴의 움직임량을 새로운 추정치로 향상시켜야 하는 것으로 판단하여 확장 갈만 필터에서 사용하는 회전 및 이동 속도, 그리고 회전 및 이동 각도를 변환함으로서 얼굴의 움직임량을 정확하게 추정할 수 있도록 하는 방법이다. 실험 결과 반복적 확장 갈만 필터를 사용하였을 경우에 얼굴의 균일한 회전 방향 변환에도 얼굴의 3차원 움직임량을 정확하게 추정할 수 있음을 알 수 있었다.

3-D Facial Motion Estimation Using Iterative Extended Kalman Filter

Kang-Ryoun Park*・Jai-Hie Kim**

ABSTRACT

Accurate 3-D facial motion estimation using computer vision is required for 3-D view control in desktop VR system or simulator and gaze detection on a monitor, etc and it has recently been researched vigorously. Most previous researches use a EKF (Extended Kalman Filter) for 3-D facial motion estimation, but it has restriction of choosing accurate initial filter parameters. In addition, the EKF cannot estimate 3-D facial motion when a user changes his face direction abruptly. To overcome such problems, we use IEKF (Iterative Extended Kalman Filter) transformed from EKF. IEKF is that when the estimated error covariance value calculated form IEKF exceeds in a predetermined threshold, it adjusts the initial parameters(rotational and translational velocities or accelerations) adaptively so that the 3-D facial motion can be estimated exactly. As experimental results, IEKF can even estimate 3-D facial motion accurately in case of abrupt facial directional change.

1. 서 론

얼굴의 3차원 움직임량을 정확하게 추정하는 연구들은 컴퓨터 비전과 휴먼 컴퓨터 인터페이스 등에서 많은 응용 범위를 가지고 최근 활발히 연구가 진행되고 있다. 그 대표적인 예로는 화상의 시스템에서 사용자의 얼굴 움직임에 따라 화상의 카메라를 자동으로 포커싱(focusing)하는 시스템, 그리고 공장 자동화 시스템에서 관리자의 양손을 입력 수단으로 사용하지 못하는 상황에서 사용자의 얼굴 움직임에 의해 반복적 확장 갈만 필터를 하는 시스템, 가장 현실 환경에서의 얼굴 움직임에 의한 3차원 그래픽 화면의 품질, 시뮬레이터에서의 훈련자 얼굴 움직임에 의한 화면 조정, 모니터상의 시선 위치 판단 및 model-based image coding 등이 있다[1]. 기존에 사용된 얼굴의 3차원 움직임 추정 알고리즘으로는 신경망(Neural Network), 광류 분석법(optical flow method) 및 확장 갈만 필터(Extended Kalman Filter)들이 있는데, 이 가운데 신경망과 광류분석법은 처리시간이 오래 걸리고 움직임 추정 에러 크다는 단점 때문에 확장 갈만 필터를 많이 이용하고 있다[2-7, 11, 13]. 그러나 이러한 확장 갈만 필터 방법 역시 필터의 초기값에 따라 얼굴 움직임을 제대로 추정하지 못하고 발생하는 경우가 종종 발생하며, 또한 얼굴의 회전 방향 변환에 이에 대처하지 못하는 경우가 종종 발생한다. 본 논문에서는 이러한 문제점을 해결하기 위하여 확장 갈만 필터의 형태변화 반복적 확장 갈만 필터를 이용하여 얼굴의 3차원 움직임량을 추정하였다.
2. 염금 특징점 추출

염금의 3차원 음직임량을 추정하기 위하여, 본 논문에서는 2차원 카메라 영상에서 관측된 염금 특징점(양 눈, 코, 입)의 위치 변화도를 이용하였다. 염금 영상에서 염금 특징점의 위치를 추출하기 위하여 이 논문에서는 먼저 염금 영역을 검출한 후 추출된 염금 영역내의 제한된 범위 내에서 양 눈과 코 및 입의 양 근무당을 추출하였다.

21 염금 영역의 추출

이 논문에서는 (그림 1)과 같이 시간적으로 연속된 두 영상간의 차영상 정보와 샌드볼 정보를 이용하여 염금 영역을 검출한다. 염금의 샌드 영상 처리부에서는 입력된 염금의 샌드 영상 정보에 대한 RGB 신호를 YIQ 모델로 변환함으로써 염금의 샌드 정보에 만족한 1636 사건(110~150)을 바탕으로 염금 영역을 검출한다.

(그림 1) 차영상과 샌드 정보를 이용한 염금 영역 검출

22 눈, 코 및 입의 양 근무당 추출 및 음직임 추적

염금 영역을 추출한 후, 추출된 염금 영역에서 조도의 변화에 따른 영역을 줄기 위하여 히스토그램 평활화(histogram equalization) 및 이진화(binaryization) 과정을 통해 이진 영상으로 변환하였다. 염금 영역에 대한 이진 영상이 구해진 후, (그림 2-2)처럼 염금 내 눈의 위치에 대한 사전지식을 이용하여 이진 영상의 제한된 범위 내에서 수직, 수평 히스토그램의 피크 위치(peak position)를 파악함으로써 눈의 위치를 정확하게 추출할 수 있었다. 양 눈이 검출된 후 (그림 2-b)처럼 입의 위치에 대한 존재 가능 범위를 설정한 후, 이 영역에 대한 이진화 및 수직 방향 히스토그램으로 입신의 수직 위치를 먼저 추출하였다.

추출된 입신의 수직 위치로부터 입의 양 근무당을 추출하기 위해 입신에 대한 수평 히스토그램을 구하여 이전의 수평 히스토그램에 근거하여 양 근무당의 위치를 결정하였다. 코구멍 역시 눈동자와 같은 방법으로 추출하였으며, 두 코구멍의 근접도가 큰 경우로 양 코구멍을 포함한 영역에 대한 수평 히스토그램을 통해 두 개의 피크 위치를 추출함으로써 양 코구멍의 수평 위치를 파악하였다. 눈 위치 추출시 안경 사용자의 경우, 안경에서의 반사되어 눈동자 흐름소가 많이 소실될 수 있다. 이러한 문제점을 해결하기 위하여 눈 영역의 이진화 시 p-tile method를 사용하여 흐름소의 개수를 일정하게 유지함으로써 흐름소 소실 문제가 해결되었다. (그림 3)은 본 논문의 방법을 사용하여 검출된 사용자의 양 눈 영역 및 눈, 입의 양 근무당 위치를 나타낸 것이다. 초기 영상에서의 염금 특징점 추출 방법과는 달리 이후 연속 영상에서는 약 야간이나 염금 영역을 추출하지 않고 이전에 추출된 특징점 부근을 탐색하는 방법을 이용하여 특징점의 위치를 추적하였다. 이때 이전 꼬 장의 영상에서 특징점의 음직임 정보를 바탕으로 현재 특징점의 위치를 예측하는 알고리즘을 사용하였다. 예측된 특징점 부근에 정해진 크기의 탐색 영역을 설정하여 그 탐색 영역내의 히스토그램 분석을 통해 특정점들의 위치를 추적하였다. 이때 전체 영상 크기는 320×240 pixel이고, 이중 염금 영역의 폭급 크기는 144×208 pixel, 눈 영역의 폭급 크기는 8×4 pixel, 코구멍의 폭급 크기는 4×4 pixel, 그리고 입의 폭급의 폭급 크기는 3×3 pixel이다. 여기서 외부 조명은 급격히 변하지 않는다고 가정하였다. 실험 결과, 염금 특징점의 실제 위치(사람이 눈으로 보고 직접 추출된 특정점의 위치)와 본 논문의 방법에 의해 추출된 특정점 위치 사이의 최소 자승 에러(RMS error)는 <표 1>과 같다.

(표 1) 추출된 염금 영역 및 눈, 코, 입 특정점 위치

(그림 3) 추출된 염금 영역 및 눈, 코, 입 특정점 위치
3. 카메라 및 얼굴 좌표계

추출된 얼굴 특정점의 속성 정보를 이용하여 본 논문에서는 얼굴의 3차원 횡적임을 추정한다. 이때, 본 논문에서 카메라 좌표계에 대한 얼굴 횡적임을 이동(translation)과 회전(rotation)으로 분리하여 얼굴의 이동은 카메라는 좌표계 원점에 대한 얼굴 좌표계 원점으로 이동으로, 얼굴의 회전은 얼굴 축을 기준으로 한 회전으로 정의하였다. 이처럼 얼굴의 횡적임을 정의하기 위해서 3차원상의 얼굴 특정점을 카메라 중심 좌표계와 얼굴 중심 좌표계로 나누어 나타냈다. 이렇게 얼굴의 특정점을 2 부분으로 나 탐색으로써 얼굴의 회전을 카메라를 기준으로 본 것이 아니라 얼굴을 기준으로 하여 일관성 있는 이동으로 모델링 할 수 있는 장점이 있다. (그림 4)는 본 논문에서 사용하는 3차원 얼굴 좌표계와 카메라 좌표계를 나타낸 것이다. 이때 본 논문에서는 (그림 4)와 같이 초기에 카메라 좌표계에 대해 얼굴 좌표계의 y축의 원점은 서로 일치할 것으로 가정 했으며, 동시에 x축 역시 초기에만 서로 일치한 것으로 가정하였다. 이를 위하여 초기에 사용자가 수행으로 카메라 좌표계에 대해 얼굴 좌표계의 x, y축을 맞추어 주도록 하였다. 여기서 얼굴 좌표계를 블록으로 정의한 이유는 얼굴의 이동만 존재하는 경우 이를 카메라 좌표계에서 표현할 수 있지만, 얼굴의 회전의 경우 얼굴 좌표계에서 설정된 얼굴 축(x, y, z축)을 기준으로 이루어지는 카메라 좌표계에서의 표현하기 어렵기 때문이다.

\[ X = f \cdot \frac{x}{z}, \quad Y = f \cdot \frac{y}{z} \]  \hspace{1cm} (2)

\( n_z \)와 \( n_y \)는 이미지 평면에서의 추출 에러로서 본 논문에서는 평균이 0인 독립적 가우스 분포(independent zero mean Gaussian distribution)로 모델링한다.

3.2 얼굴의 3차원 횡적임 모델

전송한 바와 같이 본 논문에서는 얼굴의 횡적임을 이동과 회전으로 분리하기 위해 3차원 상에서 얼굴 횡적임의 좌표를 카메라 중심 좌표계와 얼굴 중심 좌표계로 나누어 나타내었다. 이때 얼굴의 이동은 카메라 좌표계를 기준으로 하여 얼굴 좌표계 원점의 이동으로, 얼굴의 회전은 얼굴 좌표계에서 얼굴 축을 기준으로 한 얼굴 특정점의 회전으로 정의하였다. \( P \)를 얼굴 좌표계에서 파악한 특정점의 위치 벡터 \((x, y, z)\)로 원록 높: \( i = 1 \), 오른쪽 높: \( i = 2 \), 왼쪽 높: \( i = 3 \), 오른쪽 높: \( i = 4 \), 왼쪽 높: \( i = 5 \)라 하고 \( s \)를 카메라 좌표계에 대한 얼굴 좌표계 원점의 위치 벡터라 하면 다음과 같은 관계식을 얻을 수 있다.

\[ P_i = s + R \cdot P \]  \hspace{1cm} (3)

\[ P_i \]는 카메라 좌표계에서 관측된 얼굴 특정점의 좌표이고, \( R \)는 3×3 회전 행렬이다.

3.3 이동 횡적임 모델(model of translational motion)

이제 식 (3)에서 얼굴의 이동(\( s \))은 삼차원 시간 단위 \( (\Delta t) \)동안 등속도운동으로 가정할 수 있으므로 다음 식 (4)와 같이 동속도 운동 모델로 나타난다.

\[ s = s_a \cdot \Delta t + \frac{1}{2} s_a \cdot \Delta t^2 \]  \hspace{1cm} (4)

\( s_a \)는 카메라 좌표계에서 초기 얼굴 좌표계 원점의 좌표 벡터이고, \( s \)는 원점의 이동에 대한 가속도 벡터이다.

3.4 회전 횡적임 모델(model of rotational motion)

얼굴의 회전은 3차원 상에서 얼굴 좌표계를 기준으로 하
여 얼굴 특징점의 회전으로 정의하였다. 이 논문에서는 오일러 각도(euler angle)를 이용하여 시간에 따른 얼굴의 회전을 나타내었다. 식 (4)와 마찬가지로 얼굴의 회전 역시 좋은 단위 시간동안의 동가속도 운동으로 가정하여 회전을 모델링하였다.

4. 화장 간단 필터

본 논문에서는 앞서 기술한 카메라 투영 모델과 얼굴의 3차원 움직임 모델을 바탕으로 반복적 화장 간단 필터를 이용하여 얼굴의 3차원 움직임을 추정하였다. 반복적 화장 간단 필터의 출력점이 되는 화장 간단 필터를 먼저 살펴보기로 한다.

화장 간단 필터에 의해 산출된 움직임 추정방법은 예측(prediction), 측정(measurement) 및 보정(update)의 3단계를 반복하면서 얼굴의 산출된 움직임을 추정한다[11-13]. 즉, 이차원 카메라 영상에서 추출한 얼굴 특징점의 위치와 동가속도 운동 모델에 기반하여 얼굴의 산출된 이동(translation) 및 회전(rotation angle)을 구하는 것이다. 화장 간단 필터에 의해 산출된 움직임을 추정하기 위하여 본 논문에서는 식 (5)와 같이 18×1 크기의 상태 벡터(state vector)를 정의하였다.

\[ x(t) = (p(t), \phi(t), \chi(t), \omega(t), \alpha(t), \delta(t))^T \tag{5} \]

여기서 \( p(t) \)는 카메라 좌표계에 대한 얼굴 좌표계 원점의 \( X, Y \)축 방향의 사상원 이동(translation)을 나타내는 3x1 벡터이고, \( \phi(t) \)는 얼굴 좌표계에서 얼굴 특정점의 \( X, Y \)축 중심의 산출의 회전각도(rotation angle)를 나타내는 3x1 벡터이다. 또한, \( \chi(t), \omega(t) \)는 각각 산출된 이동 및 회전 속도를 나타내는 3x1 벡터들이고, \( \alpha(t), \delta(t) \)는 각각 산출된 이동 및 회전 각도를 나타내는 3x1 벡터들이다. 이 러한 상태 벡터는 3.3 및 3.4절에서 정의한 바와 같이 다음과 같은 동가속도 운동 모델에 기반한 상태 전이 행렬(state transition matrix), \( \Phi(t) \)에 의해 식 (6)과 같이 각각 단계의 산출된 움직임으로 예측된다.

\[ \hat{x}(t+1) = \Phi(t) x(t) + w(t) \tag{6} \]

또한, 3차원 상에서의 얼굴 특정점의 좌표를 2차원 이미지 투영하기는 과정에는 이차원 변환(affine transformation)과 원근 투영 변환(perspective projection transformation)이 함께 포함되어 있으므로, 선행적인 행렬 형태로 나타내지 못하고 식(8)과 같이 비 선형적 형태로 \( \alpha(t) \)의 형태로 나타낸다. 여기서 \( \alpha(t) \)는 제로 평균(zeromean) 및 분산값 \( U \)를 가지는 가우시안 잔차( gaussian noise)가 된다. 이러한 과정에서 측정 잔차가 싸게 되는데 본 논문에는 반복적 화장 간단 필터를 이용한 얼굴의 3차원 움직임 추정 31

\[ z_k = H(x_k) + n_k \tag{7} \]

식 (6)과 (7)로부터, 예측된 산출된 움직임(\( \hat{x}(t+1) \))은 식 (8)에 의해 보정되어, 정확한 산출된 움직임을 구하게 된다.

\[ \hat{x}(t+1) = \hat{x}(t+1)^- + K(t+1)(z(t+1) - H(\hat{x}(t+1)^-)) \tag{8} \]

여기서 \( K(t+1) \)는 칼만 가ین(Kalman gain)이며, 이때 칼만 가인은 다음 식 (9), (10), (11)과 같이 추정오차 공분산 행렬로부터 나타날 수 있다.

\[ K(t+1) = P(t+1)^- H_i^T (H_i P(t+1)^- H_i^T + R)^{-1} \tag{9} \]

\[ P_{t+1} = (I - K(t+1) H_i) P(t+1)^- \tag{10} \]

\[ P(t+1)^- = \Phi(t) D P(t) \Phi(t)^T + Q(t) \tag{11} \]

\( P(t) \)는 추정오차 공분산 행렬( estimation error covariance matrix)이고, \( R, Q(t) \)는 각각 측정(measurement) 및 과정( process) 정수다. \( H_i \)는 \( \cdot \cdot \cdot \)항수를 상태 벡터에 관하여 편의적 저코비안 행렬이다[7-9, 15].

\[ H_i = \begin{bmatrix} \frac{\partial H \alpha}{\partial \alpha} \\ \frac{-\partial H \delta}{\partial \delta} \end{bmatrix} \tag{12} \]

5. 반복적 화장 간단 필터

이러한 화장 간단 필터는 정확한 초기값을 사용하지 못하는 경우 필터가 수렴하지 못하고 발산하는 경우가 종종 발생한다. 동시에 식 (6)에서처럼 구간 동가속도 모델에 의해 예측으로 얼굴의 급격한 방향 변화가 생기는 경우 주기 적여하는 얼굴을 놓치버리는 경우가 종종 발생한다. 이러한 문제를 해결하기 위하여 본 논문에서는 반복적 화장 간단 필터법을 사용하였다.

반복적 화장 간단 필터(iterated extended kalman filter)는 이러한 화장 간단 필터가 가지고 있는 문제점을 보완하기 위해 (10)의 식에 의해 구해진 추정 오차 공분산 행렬 값이 정해진 임계치보다 크지 않는 경우, 이전 값에서 현재 값

\[ \hat{x}(t+1) = \hat{x}(t+1)^- + K(t+1) (z(t+1) - H(\hat{x}(t+1)^-)) \tag{13} \]
이때 $\tilde{z}_t$은 초기에 $\hat{z}(0)$으로 정하여 $t$의 값이 하나씩 증가하면서 $\hat{z}(t)$와 $\hat{z}(t+1)$의 차가 정한 값보다 작을 때까지 반복하여 칼만개인을 바꾸어 가며 얼마만큼 값을 찾아가는가. 여기서 얻은 $\hat{z}(t+1)$가 이전 값에서 예측한 현재 값($\tilde{z}_t$)이다. 이러한 과정은 확장 칼만 필터의 투프에서 칼만게인을 이용하여 이전 값과 측정 값을 이용하여 현재값을 추정할 때 포함되어 보다 정확한 현재 값을 추정한다. 즉, 식 (6), (7), (8)에 의해 계산된 칼만 게인을 그대로 이용하여 열굴의 3차원 움직임량을 추정하는 것이 아니라, 식 (5)의 회전 및 이동 속도, 그리고 회전 및 이동 각속도를 변형하여 추정 오차 공분산 행렬 값이 정해진 임계치 이내의 값이 되도록 만들어서 이에 계산된 칼만 게인으로부터 열굴의 움직임량을 추정한다.

6. 실험 및 결과

실험환경으로는 Pentium-Pro 200MHz를 사용하여 모니터 위에 설치된 CCD카메라(CCN-2411A)와 비디오 capture board (DT3152)를 통해 사용자의 열굴 움직임을 읽고, 이때 입력된 영상으로부터 열굴의 3차원 움직임량을 추정하였다. 본 논문에서는 IEKF에 의한 움직임량(X, Y, Z translation 방, Y, Z rotation 방) 추정 성능을 기존의 EKF뿐만 아니라, 실제 가상 현실 시스템에서 많이 사용되는 위치 추적 장치인 Polhemus Fastrak과 비교 실험하였다. Polhemus Fastrak는 송신단과 수신단 사이의 전자기장 점합 방식에 의해 작동되며, 최대 60Hz의 sampling rate으로 3차원 움직임량을 나타내는 6가지 자유도 변수(X, Y, Z translation량, 3 회전각(yaw, pitch, roll))를 나타낸다. 다음 (그림 5)는 실험에 사용한 데이터의 예이다.

(그림 5) 열굴의 3차원 움직임량 추정을 위한 실험 데이터의 예
(그림 6)은 IEKF에 의한 방법과 EKF, 그리고 Polhemus
7. 얼굴의 3차원 움직임 추정을 이용한 3D view의 조정

본 논문에서는 이러한 얼굴의 3차원 움직임 추정을 이용하여 desktop VR환경에서 3차원 그래픽 view를 조정하는 시스템을 개발하였다. (그림 7)는 실제 시스템의 구성 환경이다. 3D 화면의 한정이 있는 부분을 보기에 위해 얼굴을 움직이는 경우, 얼굴의 3차원 움직임량을 계속 추적하여 추출된 움직임량을 바탕으로 3D view를 조정한다. 실험 결과 본 연구의 알고리즘을 사용하여 3D view를 조정하였을 경우, 기존의 mouse 장비 등을 사용하였을 경우보다 현실감 나는 가상 현실 환경을 제공함을 알 수 있었다. 연속적으로 입력되는 영상은 영상 처리(Win95환경의 Pentiun Pro 200MHz)에서 처리되며, 계산된 얼굴의 3차원 움직임 결과는 Unix/Window Socket을 이용하여 SGI Workstation으로 전송되어 3차원 view를 조정한다.

8. 결 론

얼굴의 3차원 움직임량을 정확하게 추적하는 연구들은 컴퓨터 비전과 휴먼 컴퓨터 인터페이스 등에서 많은 활용 범위를 가지고 최근 활발히 연구가 진행되고 있다. 기존에 사용된 얼굴의 3차원 움직임 추정 알고리즘으로는 신경망(Neural Network), 광학 분석법(optical flow methods) 및 확장 칼만 필터(Extended Kalman Filter)등이 있는데, 이중 신경망과 광학분석법은 차례로 시간이 경過할수록 안정적인 결과를 내어질수록 확장 칼만 필터를 많이 사용하 고 있다. 그러나 확장 칼만 필터 방법 역시 필터의 초기값에 따라 얼굴 움직임을 제대로 추정하지 못하고 발생하는 경우가 종종 발생하며, 얼굴의 위치 방향 변화시 이상에 대처하지 못하는 경우도 종종 발생한다. 본 논문에서는 이러한 문제점에 해결하기 위하여 확장 칼만 필터의 변형 형태인 반복적 확장 칼만 필터를 이용하여 얼굴의 3차원 움직임량을 정확하게 추정하였다. 향후 연구에서 얼굴 특징점의 추 축 정확도를 높이려면 보다 정확하게 얼굴의 3차원 움직임량을 추정할 수 있을 것으로 예상된다.


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


박강영

e-mail : parkgr@iciti.com
1994년 연세대학교 전자공학과(공학사)
1996년 연세대학교 전자공학과(공학석사)
2000년 연세대학교 전자공학과(공학박사)
2000년 ~현재 LG전자기술원 선행연구원 핀란드방자 : Biometric영상 처리, 컴퓨터vi-

김재희

e-mail : jhkim@bubbleyesnet.ac.kr
1979년 연세대학교 전자공학과 졸업
1982년~1984년 Case Western Reserve Univ.
전기공학과-공학석사, 공학 박사
1984년~현재 연세대학교 기계전자공학부 교수

핀란드방자 : Biometric영상 처리, 컴퓨터 vision, 패턴인식