

실내 환경에서 시선 방향을 고려한 사람 행동 인식

공 병 용, 정 도 준, 김 향 준
 경북대학교 컴퓨터공학

Human Behavior Recognition based on Gaze Direction in Office Environment

Byung Yong Kong, Do Joon Jung, Hang Joon Kim
 Department of Computer Engineering, Kyungpook National University
 {bykong,djjung,kimhj}@ailab.knu.ac.kr

Abstract - 본 논문에서는 실내의 고정된 단일 컬러 카메라에서 획득된 비디오 스트림으로부터 사람의 행동을 인식하기 위한 시스템을 제안한다. 제안된 시스템은 사람의 시공간적 상태 변화와 사람의 시선 방향을 이용하여 규칙기반으로 행동을 인식한다. 사람의 의미 있는 상태변화를 이벤트로, 이벤트의 시퀀스 즉, 사람의 행동을 시나리오로 정의하였다. 따라서 입력 비디오 스트림에서 사람의 상태변화로 이벤트를 검출하고, 검출된 이벤트의 시퀀스로 사람의 행동을 인식한다. 사람의 시선은 얼굴과 머리 영역의 색 정보를 이용한 시선 방향 추정 방법으로 찾지며, 사람의 상태 변화는 사람의 위치와 키 등을 이용하여 검출된다. 본 시스템은 실내 환경에서 획득한 비디오에서 실험하였으며, 실험결과 시선 방향에 의해 서로 다른 행동을 구분하여 인식할 수 있었다.

1. 서 론

지능형 감시 시스템은 유비쿼터스 환경을 위한 요소로 많은 관심을 받고 있다. 지능형 감시 시스템 분야에서 사람의 행동을 인식하는 연구는 활발하게 진행되어 왔다[1][2][3][4][5]. 하지만 다양하고 복잡한 사람의 행동을 정확하게 인식하기에는 어려운 것이 사실이다. 사람의 시선 방향은 행동에 많은 영향을 미친다. 같은 위치에서도 시선 방향에 따라 다른 행동으로 비춰질 수 있기 때문에 미리 정의된 시나리오에 의해 사람의 행동을 인식하는 기존의 방법에 사람의 시선방향을 고려하는 추가 이벤트를 추가 하였다[4]. 본 논문에서는 사람 행동인식을 위해 크게 두 가지 정보를 사용한다. 첫째, 영상내의 움직이는 오브젝트(사람)의 시선 방향, 위치, 크기와 같은 상태 변화에 따른 이벤트들과 둘째, 미리 정의된 시나리오와 감시 환경의 문맥 정보가 사용된다. 시나리오는 컴퓨터 사용과 같은 각종 기기의 사용 여부와 사람의 출입과 같은 실내 환경에서 발생 가능한 사람의 행동을 표현한다. 사람의 시선 방향은 영상내 머리 영역의 픽셀 정보를 이용한 학습된 결과에 의하여 8개의 방향으로 표현 된다[6]. 본 논문에서는 이러한 시선 방향 정보를 좀 더 상세하고 다양한 행동을 인식하기 위해 사용한다. 사람의 행동을 표현하는 시나리오는 행동과 연관이 있는 여러 이벤트와 제약 조건들로 이루어지며 해당 시나리오에 존재하는 모든 이벤트가 발생하고, 제약 조건들을 만족할 때 인식 된다[4]. 실험에서는 총 8개의 이벤트와 4개의 시나리오를 사용하였다.

2. 지각 및 문맥 정보

2.1 지각 정보

본 장에서는 사람의 행동을 인식 하기위해 사람 영역의 추출 및 추적에 위한 영상처리 기술과 시선 방향을 인식하는 방법에 대하여 설명 한다.

2.1.1 사람 영역 추출 및 추적

본 논문에서는 영상내의 움직이는 영역을 사람 영역이라고 가정하였으며, 움직임을 검출하기 위해 배경 제거 기법을 사용하였다. 배경은 CrCb 색 공간에서 현재 영상과 배경 영상의 각 픽셀의 차이에 의해 제거하였다. 배경 영상 제거에 의해 생성된 이전 영상에 발생하는 잡음은 팽창 연산과 침식 연산을 이용한 모폴로지 필터를 통해 제거되어진다. 최종적으로 사람 영역은 움직임을 감싸는 사각형으로 표현된다. 영상에서 현재 프레임에 존재하는 사람은 이전 프레임에 존재하는 사람들 중 하나와 항상 일치 하게 된다. 추적 하고자 하는 이전 프레임의 사람 위치로부터 현재 프레임의 사람 중 일정 경계값 내에 존재하며 그 위치가 가장 가까운 사람을 찾음으로써 이전 프레임과 현재 프레임의 사람 사이의 추적이 이루어진다.

2.1.2 사람 시선 방향 인식

사람의 행동을 인식함에 있어 시선 방향은 사람의 의도와 목적을 파악하는데 중요한 단서가 될 수 있다. 이러한 이유로 시선 방향을 추정하는 다양한 방법이 연구되어 지고 있다. 본 논문에 사용되는 영상은 저해상도로써 고해상도 영상에서 눈과 같은 얼굴의 특징 점을 이용하여 시선 방향을 추정하는 방법은 적용하기 어려운 점이 있다[7]. 그러므로 사람의 얼굴과 머리 영역의 색 정보를 이용한 시선 방향 추정 방법을 사용하였다.

머리 영역 추출 일반적으로 영상에서의 사람 머리 부분은 사람 영역에서 윗부분에 존재하며 전체 크기에서 일정한 비율을 차지한다.[6] 또한 사람 얼굴의 컬러 분포는 색채 공간의 작은 영역에서 균집을 이루고,

2D-Gaussian 분포로 근사화 될 수 있다[6]. 따라서 스킨 컬러 모델은 색채 컬러 공간에서 얼굴 피부 색깔의 r, g 성분이 2D-Gaussian 분포를 따른다고 가정하고 근사화 시킨 모델이며, 아래 수식 (1)과 같이 표현된다.

$$r = \frac{R}{R+G+B}, g = \frac{G}{R+G+B} \tag{1}$$

$$N(m, \Sigma^2), \text{ where } m = (\bar{r}, \bar{g})$$

$$\bar{r} = \frac{1}{S} \sum_{i=1}^S r_n, \bar{g} = \frac{1}{S} \sum_{i=1}^S g_n \text{ and } \Sigma = \begin{bmatrix} \sigma_r^2 & \rho_{X,Y} \sigma_r \sigma_g \\ \rho_{X,Y} \sigma_r \sigma_g & \sigma_g^2 \end{bmatrix}$$

배경 제거로 얻어진 사람 영역에서 머리가 차지하는 일정한 비율로 머리 후보 영역을 지정하고 스킨 컬러 정보를 이용하여 후보 영역에서 보다 정확한 사람 머리 영역을 획득 하는 방법을 사용한다. 머리 후보 영역은 카메라의 위치에 따라 사람 영역에서 위로부터 1/5 ~ 1/7 에 위치하며, 머리 후보 영역에 존재하는 스킨 컬러의 검출에 의하여 보다 정확한 사람 머리 영역을 추출한다.

시선 방향 인식 시선 방향은 45° 간격으로 8방향 정의하였다. 우리는 8가지의 시선 방향을 분류하기 위하여 eigenspace 상의 각 영역의 가중치 벡터를 사용한다. n×n 이미지는 n² 차원의 이미지공간에서 한 점으로 표현되며, 이때 각각의 차원은 한 픽셀에 대응된다. 이런 이미지의 정보를 심각한 손실 없이 낮은 차원으로 나타내기 위해 PCA를 사용한다. M개의 시선 방향 학습 이미지들로부터 M'개의 고유 얼굴(E₁, E₂, E₃, ..., E_{M'})을 구하고, 특정 입력 이미지는 직교 공간에서 M'개의 고유 얼굴들의 선형 조합(linear combination)으로 표현 될 수 있다[9][10].

고유 얼굴을 구성하는 방법은 다음과 같다. M개의 학습 이미지는 고유공간을 구성하기 위해서 열벡터(Column vector)인 I₁, I₂, I₃, ..., I_M 으로 나타내고, 이들 입력 이미지의 평균 이미지인 A = 1/M ∑_{i=1}^M I_i 를 구한다. 각 학습 이미지와 평균 이미지와의 차이 벡터인 Φ_i = I_i - A들로 이루어진 새로운 벡터집합 Y = [Φ₁, Φ₂, ..., Φ_M]를 구한다. 얼굴 이미지의 분포를 최적으로 표현하는 M개의 직교(Orthogonal) 벡터 u_k를 구하기 위하여, 공분산 행렬(Covariance matrix)을 다음과 같이 구한다.

$$C = \frac{1}{M} \sum_{i=1}^M \Phi_i \Phi_i^T = YY^T \tag{2}$$

C 행렬의 크기는 N²×N²으로 계산상의 효율성을 위하여 수식 (3)과 같이 C'을 사용한다.

$$C' = \frac{1}{M} \sum_{i=1}^M \Phi_i^T \Phi_i = Y^T Y \tag{3}$$

C' 행렬로부터 구한 M개의 고유벡터(Eigenvector) u_k와 고유값(Eigenvalue) λ_k를 이용하여 u_k를 다음과 같이 구한다.

$$u_k = \frac{Y \times v_k}{\sqrt{\lambda_k}}, k = 1 \dots M \tag{4}$$

λ_k의 값에 따라 u_k 중에서 중요한 M'개를 선택하여 m개의 학습 이미지에 대해서 고유벡터의 선형 조합, W_m^{training} = [w₁, w₂, ..., w_{M'}]을 구한다.

$$w_k = u_k^T \Phi_i, k = 1 \dots M' \tag{5}$$

시선 방향을 추정하기 위하여 검출된 시선 방향 영역을 직교공간으로 투영하여 얻은 W^{input} 값을 학습 이미지들의 각 시선 클러스터 와 비교한 값이 가장 가까운 클러스터를 최종 시선 방향으로 추정한다.

2.2 문맥 정보

카메라에 관측되는 환경의 정보 즉, 인식하고자 하는 사람 행동에 영향을 미치는 특정 장소나 사물에 대한 위치 또는 영역의 정보는 행동 인식에 중요한 요소이다. 의자, 컴퓨터, 출입구와 같은 장소와 사물에 대한 위치, 영역정보와 특징을 정의하여 사용하였다.

3 사람 행동 인식

본 논문에서는 행동인식을 위하여 사람의 특정 행동과 연관된 이벤트와 시공간적 제약조건으로 이루어진 시나리오 인식 방법을 사용하였다[4]. 기존의 시나리오 인식 방법에서 사람 시선 방향을 고려한 이벤트를 추가하여 모호한 사람 행동을 보다 상세하게 인식하는 방법을 제안한다. 시나리오는 사람의 특정 행동을 인식하기 위한 서로 연관이 있는 이벤트들과 시공간적 제약조건으로 이루어져 있다[4].

3.1 이벤트 인식

사람의 행동을 인식할 때 사람의 상태 즉, 위치나 키 그리고 시선 등은 중요한 단서가 될 수 있다. 이벤트는 이런 여러 상태들의 시공간적 변화를 나타낸다. 본 논문에서는 총 8가지의 이벤트를 정의하였다.(CROUCHES DOWN, STAND UP, LEAVES, ENTERS, MOVES CLOSE TO EQUIPMENT, MOVES AWAY FROM EQUIPMENT, SITS ON, LOOKS AT) 이벤트는 사람의 상태를 지속적으로 관측하여 시공간적 의미가 있는 다른 상태로 변화할 때 비로소 인식이 되어 진다.

3.2 시나리오 인식

시나리오는 사람의 행동을 표현한다. 시나리오는 앞에서 설명된 사람의 행동과 연관이 있는 이벤트들과 시공간적 제약조건들로 이루어져 있다. 이런 시나리오를 인식함으로써 사람의 행동을 인식되어진다. 그림 1은 컴퓨터를 사용하는 사람의 행동을 표현하는 시나리오의 예이다. 각각의 사람 행동을 표현한 시나리오의 모든 이벤트들이 발생하고 모든 제약 조건을 만족 될 때 시나리오에 해당하는 사람 행동이 인식된다.

Scenario
Name="Person uses a computer",
Events=(t1, moves close to(P1:Person, E1:Equipment)),
(t2, sits on(P1:Person, E1:Equipment)),
(t3, looks at(P1:Person, E2:Equipment))
not (t4, stands up(P1:Person))
Constraints = t1 <= t2, t4 <= t3 + normal using time(E2)
Conditions = function(E1,"chair"), function(E2, "computer"),
Success = alarm(P1, "using the", E1)

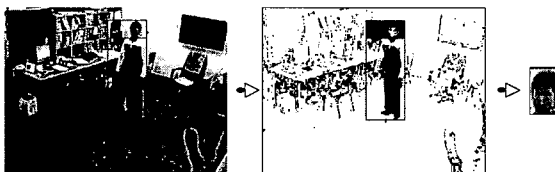
〈그림 1〉 시나리오 모델의 예

4. 실험 결과

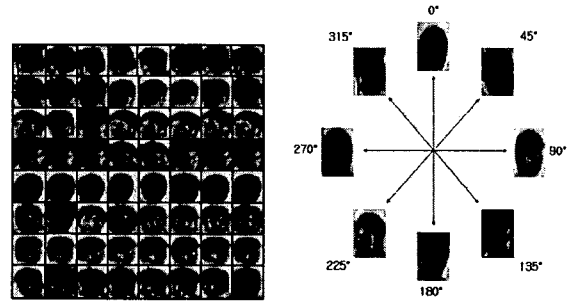
실험은 실제 환경에서 고정된 단일 카메라를 이용하였으며, 8개의 이벤트와 4개의 시나리오를 정의하여 실험하였다. 2장에서 설명된 방법에 의해 찾아진 사람영역과 머리 영역은 그림 2와 같으며 머리 영역의 검출 성공률은 약 83% 이다. 그림 3은 시선 방향을 인식 하기위해 사용된 학습 영상과 8방향의 시선 방향을 나타낸다. 총 240장의 이미지로 학습을 하였으며 입력된 정확한 머리영역에 대한 시선 방향은 약 70%의 인식률을 보였다. 그림 4는 이벤트 검출을 위한 사람의 키의 변화를 시간 흐름에 따라 표현된 그래프이며 키가 임계값 이하의 상태로 변하는 경우 'CROUCHES DOWN' 이벤트가 발생된다. 시나리오는 컴퓨터 사용, 사무실 출입, TV시청 4가지로 정의하여 실험 하였으며 정확한 시선 방향이 인식된 경우 76%의 인식률을 보였다. 동일한 위치에서 다른 행동의 예로 그림 5는 같은 위치에서 서로 다른 시선 방향을 향하는 상황으로 (a)에서는 시나리오가 발생하지 않으며 (b)에서는 컴퓨터를 사용하는 시나리오로 인식이 되게 된다.

5. 결 론

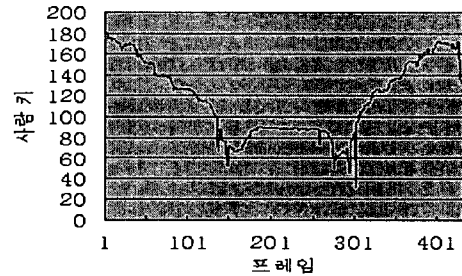
본 논문에서는 실내의 고정된 단일 칼라 카메라로부터 얻어진 비디오 스트림으로부터 사람의 행동을 인식하기 위한 방법을 제안하였다. 제안된 방법은 기존의 시나리오 인식에 의한 행동 인식 방법에서 시선 방향을 고려한 이벤트를 추가함으로써 보다 상세한 행동 인식을 가능케 하였다. 마지막으로 보다 정확한 행동인식을 위해 향상된 시선 방향을 추정 방법을 연구 중에 있으며, 장애물에 의한 폐색과 머리 영역 추적과 같은 전처리 영역에서도 보완이 필요하다.



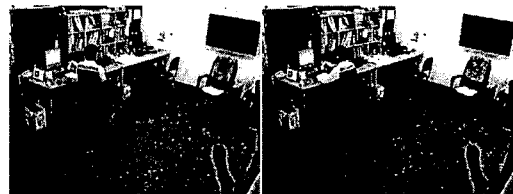
〈그림 2〉 사람 및 머리 영역 추출 결과



〈그림 3〉 시선 방향 과 학습 영상



〈그림 4〉 사람 키의 변화에 의한 이벤트 인식 예



(a) (b)
〈그림 5〉 시선 방향에 의한 다른 행동의 예

[참 고 문 헌]

- [1] D. Ayers and M. Shah, "Monitoring human behavior in an office environment", *In Computer Society Workshop on Interpretation of visual Motion*, 1998.
- [2] N.M. Robertson and I.D. Reid, "A general method for human activity in video", *Computer Vision and Image Understanding*, 104, 232-248, 2006.
- [3] Kellokumpu V, Pietikäinen M and Heikkilä, "Human activity recognition using sequences of postures.", *IAPR Conference on Machine Vision Applications*, 570-573, 2005.
- [4] N. Rota and M. Thonnat, "Video Sequence Interpretation for Visual Surveillance", *3rd IEEE International Workshop on Visual Surveillance*, VS'00, 59-67, 2000.
- [5] N. Ghanem, D. DeMenthon, D. Doermann and L. Davis, "Representation and Recognition of Events in Surveillance Video Using Petri Nets", *Second IEEE Workshop on Event Mining 2004, CVPR2004*, 112-112, 2004.
- [6] N.M. Robertson and I.D. Reid, "Estimating Gaze Direction from Low-Resolution Faces in Video", *Proceedings of the 9th European Conference on Computer Vision(ECCV)*, 2006.
- [7] Jian-Gang Wang, Eric Sung and Ronda Venkateswarlu, "Eye Gaze Estimation from a Single Image of One Eye", *IEEE International Conference on Computer Vision(ICCV)*, 2003.
- [8] J. Yang and A. Waibel, "A Real Time Face Tracker", *Applications of Computer Vision, Proceedings 3rd IEEE Workshop*, 142-147, 1996.
- [9] M. Turk and A. Pentland, "Face Recognition Using Eigenfaces", *IEEE Proceedings CVPR '91*, 586-591, 1991
- [10] A. X. Guan and H. H. Szu, "A local face statistics recognition methodology beyond ICA and/or PCA", *International Joint conference on Neural Network*, 1016-1021, 1999.