

모델 기반의 시선 방향 추정을 이용한 사람 행동 인식

(Human Activity Recognition using Model-based Gaze Direction Estimation)

정 도 준*, 윤 정 오**

(Do-Joon Jung and Jeong-Oh Yoon)

요 약 본 논문에서는 모델 기반으로 추정된 사람의 시선 방향을 이용하여 실내 환경에서 발생할 수 있는 사람의 행동을 인식하는 방법을 제안한다. 제안하는 방법은 크게 두 단계로 구성된다. 첫째, 행동 인식을 위한 사전 정보를 얻는 단계로 사람의 머리 영역을 검출하고 시선 방향을 추정한다. 사람의 머리 영역은 색상 정보와 모양 정보를 이용하여 검출하고, 시선 방향은 머리와 얼굴의 관계를 표현한 베이지안 네트워크 모델을 이용하여 추정한다. 둘째, 이벤트와 사람의 행동을 나타내는 시나리오를 인식하는 단계이다. 이벤트는 사람의 상태 변화로 인식하고, 시나리오는 이벤트들의 조합과 제약 사항을 이용하여 규칙 기반으로 인식한다. 본 논문에서는 시선 방향과 연관이 있는 4 가지의 시나리오를 정의하여 실험 한다. 실험을 통해 시선 방향 추정의 성능과 시선 방향이 고려된 상황에서의 행동 인식 성능을 보인다.

핵심주제어 : 사람 행동, 베이지안 네트워크 모델, 시선 방향, 이벤트, 시나리오

Abstract In this paper, we propose a method which recognizes human activity using model-based gaze direction estimation in an indoor environment. The method consists of two steps. First, we detect a head region and estimate its gaze direction as prior information in the human activity recognition. We use color and shape information for the detection of head region and use Bayesian Network model representing relationships between a head and a face for the estimation of gaze direction. Second, we recognize event and scenario describing the human activity. We use change of human state for the event recognition and use a rule-based method with combination of events and some constraints. We define 4 types of scenarios related to the gaze direction. We show performance of the gaze direction estimation and human activity recognition with results of experiments.

Key Words : Human Activity, Bayesian Network Model, Gaze Direction, Event, Scenario

1. 서 론

지난 수십여 년 전부터 공공장소를 비롯하여 병원,

은행, 주차장 등의 보안 취약 지역에서 감시 시스템이 많이 사용되어져 왔다. 최근 그 수요는 꾸준히 증가하는 추세이고, 사람의 행동을 인식하거나 시스템 스스로 상황을 판단하여 경보를 울리는 등의 지능형 감시 시스템에 대한 관심이 높아지고 있다[1, 2, 3].

감시 시스템은 특정 장소의 관찰을 용이하게 하기

* 국립과학수사연구원 문서영상과 영상분석실

** 경운대학교 정보통신공학과

위해 보안 담당자에게 실시간으로 영상을 보여주고 녹화한다. 보안 담당자가 감시 시스템의 모니터를 통해 여러 장소를 동시에 그리고 지속적으로 장시간 모니터링을 할 때 집중력과 업무효율이 현저히 저하될 수 있다. 이러한 문제점을 보완한 시스템으로는 대표적으로 IBM에서 개발한 스마트 감시 시스템(S3 : Smart Surveil-

lance System)이 있다[4]. S3에는 사물의 탐지, 추적과 분류 기술이 적용 되었으며, 디지털 영상의 자동 저장 및 효율적인 검색과 경보 기능을 제공 한다.

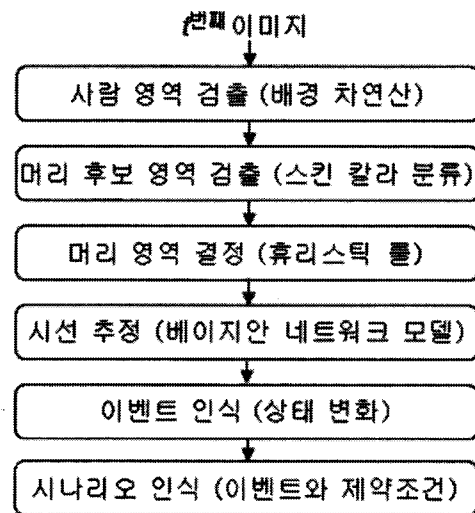
지능형 감시 카메라에 적용하기 위한 사람의 행동을 인식하는 연구에는 사람의 연속된 자세를 관찰하여 행동을 인식하는 방법, 미리 정의된 시나리오 모델을 사용하여 행동을 인식하는 방법[5, 6] 그리고 확률적 모델을 사용한 방법[1, 7]이 있다. 이들 대부분의 사람 행동 인식 방법은 사람 또는 자동차의 위치, 궤적, 속도와 같은 정보만을 이용하기 때문에 다양하고 복잡한 사람의 행동이나 상황을 검출하기에는 어려움이 있다.

시선 방향은 사람의 진행 방향과 행동의 목적에 많은 영향을 준다. 그러므로 위치나 크기, 속도와 같은 기본적인 정보 외에 사람의 시선 방향은 사람의 행동을 예측하는데 중요한 요소로 작용할 수 있다. 본 논문에서는 지능형 감시 시스템을 위한 사람 행동 인식에 있어서 사람의 시선방향을 고려함으로써 보다 상세한 사람 행동을 인식하는 방법을 제안한다.

본 논문에서의 사람 행동 인식은 미리 정의된 시나리오를 통하여 인식하는 방법[5, 8]을 사용한다. 시나리오는 사람의 행동을 표현하며, 특정 행동과 연관이 있는 이벤트들과 제약 조건들로 이루어져 있다[5]. 본 논문에서는 실내 환경에서 발생 가능한 행동들 중 몇 가지를 사람 시선 방향과 관련된 이벤트들과 시나리오들로 표현하여 인식하며[8], 시선 방향 추정은 베이저안 네트워크를 이용한다[9].

그림 1은 본 논문에서 사용한 사람 행동 인식 방법의 흐름도이다. 입력 이미지에서 배경 영상과의 차연산으로 사람 영역(전경 영역)을 검출하고 검출된 사람 영역에서 스킨 칼라 픽셀들을 분류하여 머리 후보 영역을 검출한다. 머리 후보 영역 중에서 얼굴 영역 근처에 머리카락 영역이 존재하는지의 유무에 따라 머

리 영역을 결정하고 얼굴 영역과 머리 영역을 2개의 타원으로 표현한다. 2개의 타원의 위치 관계에 의해서 시선이 추정되고 추정된 시선은 이벤트 인식과 시나리오인식에 사용된다. 미리 정의된 장소, 사물의 위치에 대해 사람의 상태가 어떻게 변하는지에 따라 이벤트를 인식하고 인식된 이벤트와 시선 방향, 시간의 제약조건을 사용하여 시나리오를 인식하며 인식된 시나리오가 사람의 행동이 된다.



<그림 1> 사람 행동 인식 방법의 흐름도

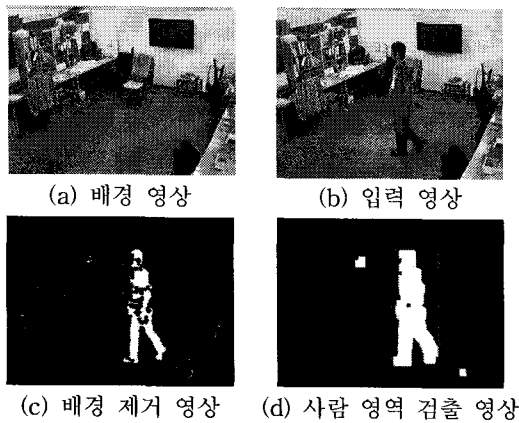
2. 시각 및 문맥 정보

본 장에서는 사람 영역의 검출을 위한 영상처리 기술과 행동 인식을 위해 사용된 실험 환경의 사전 정보, 그리고 사람 시선 방향을 추정하는 방법에 대하여 설명한다.

2.1 사람 영역 검출

본 논문에서는 영상내의 전경 물체를 사람이라고 가정하며, 전경 물체의 검출을 위해 배경 제거 기법을 사용한다. C_r , C_g 채널 공간에서 픽셀 값 차를 이용하여 배경을 제거한다. 배경이 제거된 이진 영상에서 발생하는 잡음(noise)은 형태 처리법(morphological analysis)으로 완화하며, 연결 요소 분석법(connected

component analysis)을 사용하여 전경 영역을 바운딩 박스로 표현한다. 바운딩 박스들 중 가로, 세로의 크기가 각각 25, 50 픽셀 이상인 박스만을 사람 영역으로 간주한다. 이 크기는 카메라에서 가장 멀리 떨어져 있는 사람이 앉아 있을 때도 사람 영역으로 검출될 수 있는 크기로서 실험에 의해 결정하였다. 바운딩 박스로 표현된 사람의 영역은 가로(width), 세로(height), 그리고 영상내의 위치(x, y) 속성을 가지며, 영상내의 위치는 박스 아래쪽의 중심을 나타낸다. 그림 2는 입력 영상으로부터 사람의 영역을 검출한 결과이다.



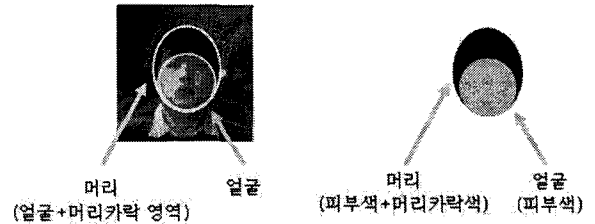
<그림 2> 사람 영역 검출

2.2 머리 모델링과 검출

사람의 시선 방향을 추정하기 위해서 사람 영역에서 머리 영역을 검출하고, 머리 영역과 얼굴 영역을 잘 감싸는 타원을 찾는다. 이렇게 찾아진 두 타원의 크기와 위치 정보를 보고 시선 방향을 추정한다. 이를 위해서 얼굴과 머리를 각각의 타원 모델로 나타내고 두 타원 모델의 파라미터를 찾아낸다. 사람의 얼굴 윤곽은 타원의 형태를 나타내기 때문에 얼굴을 타원 모델을 사용하여 용이하게 검출할 수 있다[10].

머리는 얼굴 영역과 머리카락 영역으로 구성된다. 그림 3과 같이 두 개의 타원이 겹쳐진 이중 타원이 본 논문에서 사용하는 타원 모델이다. 피부색 영역은 얼굴 타원으로 나타내어지고, 머리 타원은 피부색 영역 머리카락 색 영역을 더한 부분으로 나타내어진다. 모델은 $e_f = (x_f, y_f, w_f, h_f)$, $e_h = (x_h, y_h, w_h, h_h)$ 파라미터로 구성된 각각의 타원이다. (x, y) 는 타원의 중심

좌표이고, w 와 h 는 각각 타원의 폭과 높이이다. 그림 3은 시선 방향이 정면인 이미지를 모델링 한 예이다.



<그림 3> 머리 모델

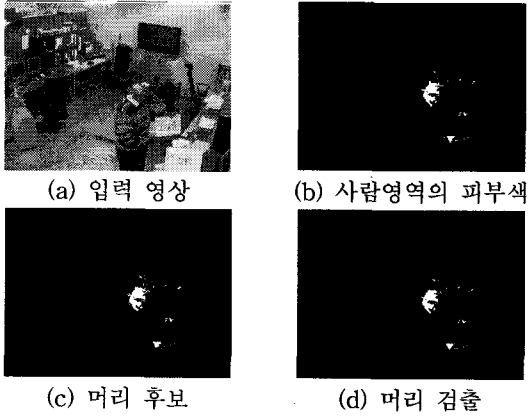
타원은 회전하지 않으며, 얼굴 타원 e_f 는 항상 머리 타원 e_h 의 앞에 위치하고, 머리 타원만 존재할 경우에는 얼굴 타원이 머리 타원에 가린 것으로 간주하여 시선 방향은 back으로 추정한다. 얼굴 타원은 다음 식과 같이 나타낼 수 있다.

$$\frac{(x-x_f)^2}{(w_f/2)^2} + \frac{(y-y_f)^2}{(h_f/2)^2} = 1 \quad (1)$$

타원은 가변의 비율을 가지며, 초기 타원의 크기는 초기 프레임으로부터 계산된다. 그림 3에서 보는 것처럼 얼굴 영역은 얼굴 색상 분포를 가지며, 머리카락 영역(머리에서 얼굴 영역을 뺀 부분)은 머리카락의 색상 분포를 가진다.

영상에서 머리와 얼굴을 잘 나타내는 타원을 찾기 위해 먼저 영상에서 사람 영역을 검출하고, 그 영역 내에서 머리 후보 영역을 검출하여야 한다. 머리 검출에서의 목적은 머리의 위치와 이미지의 비율을 결정하고, 그것을 잘 감싸는 타원을 나타내는 것이다. 영상에서 머리를 검출하기 위해, 먼저 얼굴 후보를 분류하고 나서 후보와 이웃하는 머리카락 영역을 검출하는 것이다. 이웃하는 영역에 머리카락이 존재한다면 실제 얼굴로 분류한다. 얼굴은 얼굴 영역을 모두 포함하는 안쪽 타원에 의해 표현되고, 머리는 머리 영역을 모두 포함하는 바깥쪽 타원에 의해 각각 표현된다. 하지만, 피부색만으로 얼굴을 검출하는 데는 한계가 있다. 이 점은 모양 정보를 사용하여 보완 할 수 있다. 머리 검출에 모양 정보를 추가하기 위해, 피부색으로

검출된 피부 영역에 연결 요소 분석법을 사용한다. 그 영역은 피부색 영역을 모두 포함하는 가장 작은 타원으로 나타내어진다. 타원안의 피부색 비율이 사전에 정의된 임계값보다 크면 피부색 영역은 얼굴 후보로 분류한다. 그림 4는 머리 검출 과정을 보여준다.



<그림 4> 머리 검출 과정

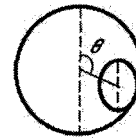
사람 피부의 컬러 분포는 색채 공간의 작은 영역에서 군집을 이루고, 2차원 가우시안 분포로 근사화 될 수 있다[11]. 따라서 색채 컬러 공간에서 얼굴 피부 색 깔의 C_r , C_g 성분이 2차원 가우시안 분포를 따른다고 가정하고 근사화 시킨 피부색 컬러 모델을 사용하여 얼굴 영역을 검출할 수 있다[7]. 얼굴 패치로부터 얻은 피부색의 채도는 C_r , C_g 공간에서 군집을 이루며 작은 타원 모양에 가깝다. 우리는 타원 함수로 나타낸 멤버십 함수로 얼굴색을 분류한다[9]. 머리카락 영역 검출도 피부색 영역 검출에서 사용한 방법과 동일한 방법으로 검출한다. 머리카락 패치 샘플을 사용하여 멤버십 함수의 파라미터를 결정하고 머리카락 영역을 분류한다.

2.3 시선 방향 추정

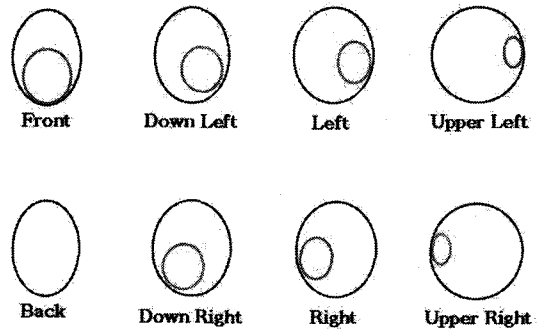
시선 방향 추정을 위해 베이지안 네트워크를 이용한다. 베이지안 네트워크에서 두 타원의 크기 비율과 타원 중심간의 각도로부터 시선 방향을 추정한다.

사람의 머리는 머리(얼굴 영역 + 머리카락 영역)와 얼굴(얼굴 영역)로 구성되어 있다고 가정하였다. 사람의 시선 방향이 변함에 따라 각각 타원의 크기가 변하고, 그림 5와 같이 두 타원 중심 간의 각도도 변한

다. 예를 들어, 시선 방향이 back일 때 머리 타원에 의해 얼굴 타원이 모두 가리게 되고, 타원간의 중심 각도는 0도이며 같은 선상에 놓이게 된다. 타원 중심 간의 각도는 back에서 left, front, right 방향으로 회전함에 따라 증가하게 된다. 즉, 그림 5와 같이 얼굴 타원의 중심이 시계 방향으로 회전함에 따라 두 타원 중심 간의 각도는 증가한다. 시선 방향은 그림 6과 같이 8방향으로 양자화 한다. 얼굴과 머리 영역을 타원으로 표현한 두 타원으로부터 타원간의 크기 비율과 두 타원 중심 간의 각도를 알 수 있으며, 이 두 가지를 베이지안 네트워크의 사전 정보로 이용하여 시선 방향을 추정한다.

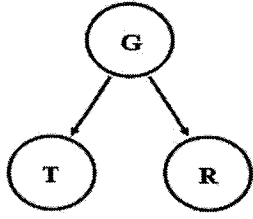


<그림 5> 두 타원 중심 간의 각도



<그림 6> 시선 방향에 따른 모델

베이지안 네트워크를 이용한 시선 방향의 추정은 머리의 방향에 따라 8방향(front, left, right, upper left, upper right, down left, down right, back)중 하나로 추정된다. 사람의 시선 방향을 추정하기 위한 베이지안 네트워크는 그림 7과 같이 구성 된다. G는 추정하고자 하는 시선 방향을 나타내는 은닉 노드(hidden node)이며, T와 R은 각각 두 타원 중심 간의 각도와 타원간의 크기 비율을 나타내는 관찰 노드(observation node)이다.



<그림 7> 시선 방향 추정을 위한 베이지안 네트워크

타원 중심 간의 각도(T)는 시선 방향(G)에 영향을 받고 타원 크기 비율에는 독립적이다. 반대로 타원 크기 비율은 시선 방향에만 의존하고 타원 중심 간의 각도에는 영향을 받지 않는다. 이 관계는 그림 7과 같이 은닉 노드와 관찰 노드 사이의 화살표에 의해 표현된다. 은닉 노드의 상태는 $G = \{\text{front, left, right, upper left, upper right, down left, down right, back}\}$ 로 정의 되고, 관찰 노드의 상태는 $T = \{0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ\}$, $R = \{0, 0.2, 0.4, 0.6, 0.8, 1\}$ 로 정의 된다. 그림 7에서 베이지안 네트워크의 결합 확률은 다음과 같이 조건 확률과 사전 확률로 분해된다.

$$P(G, T, R) = P(T, R|G)P(G) = P(T|G)P(R|G)P(G) \quad (2)$$

베이지안 네트워크를 이용한 시선 추정은 관찰 노드 T 와 R 이 주어졌을 때 은닉 노드 G 의 상태를 추정하는 것이다.

$$P(G|T, R) = \frac{P(G, T, R)}{P(T, R)} = \frac{P(G, T, R)}{\sum_{all G_n} P(G, T, R)} = \frac{P(T|G)P(R|G)P(G)}{\sum_{all G_n} P(T|G)P(R|G)P(G)} \quad (3)$$

여기에서 G_n 은 G 노드에서 상태 n 에 대한 각각의 값을 나타낸다. 시선 방향 추정을 위한 베이지안 네트워크의 조건 확률 테이블(conditional probability table)은 표 1, 2와 같다. 이 표는 2분간의 동영상으로부터 획득된 이미지들로부터 학습된 결과들이다. 표에서 각각의 열은 조건 상태를 나타내고, 행은 종속 상태를 나타낸다. 예를 들어 표 1에서 시선 방향이 front일 때 두 타원 중심 간의 각도

가 180° 일 확률은 0.84이고, 225° 일 확률은 0.08인 것이다. 표 2에서는 시선 방향이 right일 때 두 타원의 크기 비율이 0.4일 확률은 0.85이고, 0.6일 확률은 0.15이다. $P(G)$ 에 대한 사전 확률은 $P(\text{front}) = 0.125$, $P(\text{back}) = 0.125$ 와 같이 영상에서 각각의 시선 방향이 나타날 확률은 모두 동일하다고 가정한다. 어떤 영상에서 두 타원의 크기 비율과 중심 간의 각도를 알면 두 값을 각 항목에 맞게 양자화 한다. 이것으로부터 8방향에 대한 각각의 확률을 계산한다. 이 확률 값이 가장 크게 나타난 방향이 시선 방향으로 추정된다.

<표 1> 조건 확률 테이블($P(T|G)$)

시선방향	0°	45°	90°	135°	180°	225°	270°	315°
Front	0	0	0	0.08	0.84	0.08	0	0
Back	0	0	0	0.63	0.05	0.32	0	0
Left	0	0	0	0.81	0.19	0	0	0
Right	0	0	0	0	0.18	0.82	0	0
Upper Left	0	0	0	0.94	0.06	0	0	0
Upper Right	0	0	0	0	0.12	0.88	0	0
Down Left	0	0	0	0.95	0.05	0	0	0
Down Right	0	0	0	0	0.13	0.87	0	0

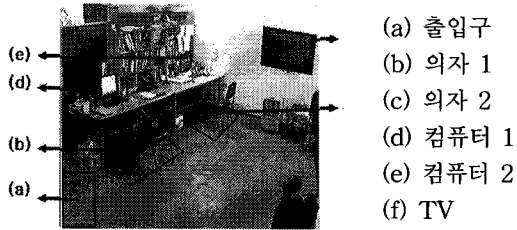
<표 2> 조건 확률 테이블($P(R|G)$)

시선방향	0	0.2	0.4	0.6	0.8	1
Front	0	0	0	0.34	0.66	0
Back	0.74	0.26	0	0	0	0
Left	0	0	0.83	0.17	0	0
Right	0	0	0.85	0.15	0	0
Upper Left	0	0.68	0.32	0	0	0
Upper Right	0	0.66	0.34	0	0	0
Down Left	0	0	0.06	0.94	0	0
Down Right	0	0	0.08	0.92	0	0

2.4 문맥 정보

사람의 행동 인식을 위해서는 사람의 위치, 크기 등의 정보 외에 관측되는 감시 환경에 대한 정보가 필요하다[5]. 출입구와 같은 특정한 장소나 컴퓨터, TV와 같이 사람 행동에 영향을 미치는 사물들에 대한

위치 및 영역에 대한 사전 정보는 실험을 통해 얻을 수 있다. 본 논문에서는 하나의 장소와 다섯 개의 사물에 대한 정보를 사용한다. 사용된 정보로는 사물의 유형과 영상 내에 위치하는 영역의 좌표, 그리고 사물에 해당하는 최소 사용 시간에 대한 정보이다. 그림 8은 장소와 사물의 영역을 나타낸다.



<그림 8> 정의된 장소와 사물의 영역

3. 사람 행동 인식

사람의 행동은 시·공간적인 움직임의 연속을 나타낸다. N. Rota 와 M. Thonnat은 행동 인식을 위하여 사람의 특정 행동과 연관된 이벤트들과 시·공간적 제약 조건으로 이루어진 시나리오를 인식하는 방법을 사용하였다[5, 6]. 하지만 사람의 위치, 키, 이동 속도와 같은 정보만을 이용하므로 보다 상세한 사람의 행동을 인식하기에는 어려움이 있다. 예를 들어 사람이 컴퓨터 앞에 있는 의자에 앉아 있는 상황에서 그 사람이 컴퓨터를 사용하고 있을 수도 있으며 옆의 사람과 이야기를 나누거나 혹은 TV를 보고 있을 수도 있기 때문이다. 그래서 사람의 시선 방향은 사람의 특정 행동의 의도나 목적을 알기 위한 중요한 정보가 될 수 있다. 본 논문에서는 N. Rota 와 M. Thonnat이 제안한 기존의 시나리오 인식 방법에서 사람 시선 방향을 고려한 이벤트를 추가하여 모호한 사람 행동을 보다 정확하게 인식하는 방법을 제안한다. 시나리오는 사람의 특정 행동을 인식하기 위한 서로 연관이 있는 이벤트들과 시·공간적 제약조건으로 이루어져 있다[5]. 이벤트는 사람의 행동과 관련된 여러 상태들의 의미 있는 변화를 나타낸다[5].

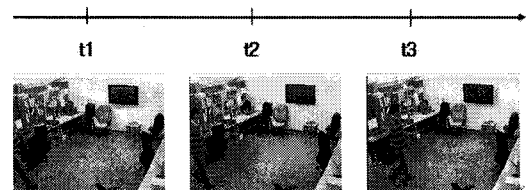
3.1 이벤트 인식

본 논문에서는 행동을 인식하기 위한 5개의 이벤트를 정의한다. 정의된 이벤트는 “STAND”, “SIT”, “ENTER”, “LEAVE”, “LOOK”이다. 이벤트는 사람의 상태를 지속적으로 관측하여 시·공간적 의미가 있는 다른 상태로 변화할 때 인식 된다. 예를 들어 사람 영역 박스의 높이가 임계값(105 pixels) 보다 큰 상태에서 작은 상태로 변화 될 때 “SIT” 이벤트가 발생된다. 반대로 임계값 보다 작은 상태에서 큰 상태로 변화하면 “STAND” 이벤트가 발생 된다. “ENTER”와 “LEAVE”는 사람의 위치가 특정 영역의 내부 또는 외부로 이동할 때 발생된다. 이벤트는 인식된 시간(t_i)과 이벤트의 주체인 사람(P_i) 그리고 연관된 장소(A_i) 혹은 사물(E_i)로 표현된다. 예로 “(t_i , enter(P_i , A_i))”는 t_i 의 시간에 P_i 사람이 A_i 영역에 들어가는 이벤트를 나타낸다.

3.2 시나리오 인식

시나리오에 속한 모든 이벤트들이 발생하고 모든 제약 조건이 만족 될 때 시나리오에 해당하는 사람 행동이 인식된다. 예를 들어 사람이 사무실에 들어오는 행동은 그림 9와 같은 세 개의 이벤트로 표현될 수 있다. 첫째, 출입구 영역에 사람이 들어온다: (t_1 , enter(P_1 , A_1)). 둘째, 시선은 사무실 내부를 향한다: (t_2 , look(P_1 , A_2)). 셋째, 출입구 영역에서 벗어난다: (t_3 , leave(P_1 , A_1)). 이 세 가지 이벤트가 발생하고 A_1 이 “출입구”, A_2 는 “실내방향”, $t_1 < t_2 < t_3$ 의 제약 조건을 만족하면 “Person 1 enters an Office” 라는 시나리오가 인식이 되게 된다. 본 논문에서 정의된 시나리오의 예는 표 3과 같다.

enter(P_1 , “출입구”) look(P_1 , “실내방향”) leave(P_1 , “출입구”)



<그림 9> 시나리오의 발생과정 예

<표 3> 시나리오 모델

구분	이벤트와 제약조건
시나리오 1	events : occur (t_1 , enter(P_1 :Person, A_1 :Area)), occur (t_2 , look(P_1 :Person, A_2 :Area)), occur (t_3 , leave(P_1 :Person, A_1 :Area)) constraints : $t_1 < t_2 < t_3$, function(A_1 , “출입구”), function(A_2 , “실내방향”)
시나리오 2	events : occur (t_1 , enter(P_1 :Person, A_1 :Area)), occur (t_2 , look(P_1 :Person, A_2 :Area)), occur (t_3 , leave(P_1 :Person, A_1 :Area)) constraints : $t_1 < t_3$, function(A_1 , “출입구”), function(A_2 , “실외방향”)
시나리오 3	events : occur (t_1 , sit(P_1 :Person, E_1 :Equipment)), occur (t_2 , look(P_1 :Person, E_2 :Equipment)), not occur (t_3 , look(P_1 :Person, E_3 :equipment)) constraints : $t_3 < t_1 + \text{normal using time}(E_2)$, function(E_1 , “의자”), function(E_2 , “TV”), function(E_3 , “not TV”)
시나리오 4	events : occur (t_1 , sit(P_1 :Person, E_1 :Equipment)), occur (t_2 , look(P_1 :Person, E_2 :Equipment)), not occur (t_3 , stand(P_1 :Person)) constraints : $t_3 < t_1 + \text{normal using time}(E_2)$, function(E_1 , “의자”), function(E_2 , “컴퓨터”)

표 3에서, 시나리오 1은 “Person enters an Office”, 시나리오 2는 “Person leaves an Office”, 시나리오 3은 “Person watches a TV”, 시나리오 4는 “Person uses a computer”를 각각 나타낸다. 제약 조건에서 “사용 시간”은 해당 사물과 연관된 행동을 하는데 걸리는 최소한의 시간을 의미한다. function은 사물 또는 장소가 해당 시나리오에 알맞은 것인가를 확인하는 함수이다.

4. 실험 환경 및 결과

본 연구의 성능 평가를 위해 실험실에서 웹캠으로 320×240 크기의 컬러 영상 시퀀스를 초당 15 프레임 씩 획득하여 실험을 수행하였다. 사람 머리 영역 검출 실험에서는 총 2,196장의 영상을 사용하였으며, 실험 결과 84%의 성공률을 보였다. 성공률은 검출 결과 영상을 보고 직접 판단하여 계산하였으며 성공적으로 검출된 머리 영역은 8 방향으로 분류되며 각 방향마

다 60개의 영상 총 480개의 머리 영역 영상이 시선 방향 추정을 위한 학습에 사용되었다. 학습은 수작업으로 이루어졌고, 얼굴영역과 머리영역을 잘 나타내는 타원 2개를 표현하여 두 타원의 중심 간의 각도와 타원간의 크기 비율을 계산하여 조건 확률 테이블(표 1, 2)을 작성 하였다.

표 4는 시선 방향을 추정된 실험 결과이다. 실험 결과 방향별로 약 60~75%, 전체 평균 69%의 성공률을 보였으나 일부 시선 방향은 낮은 인식율을 보였다. 사람이 카메라로부터 멀리 떨어져 있는 경우 얼굴 영역이 잘 못 찾아져서 인식이 안 되었고, 사람이 카메라 아래쪽 가까이 오는 경우에도 얼굴 영역이 잘 못 찾아지는 횟수가 많았다. 조명이 심하게 바뀌지 않도록 주의하였으나 해상도가 낮은 영상이므로 조명 변화에 영향을 받아 얼굴 영역과 머리 영역의 검출이 어려운 경우가 있었다. 처리 속도를 감안하여 해상도를 낮췄으나 향후 조금 더 높은 해상도에서는 나은 성능을 보일 수 있을 것이다.

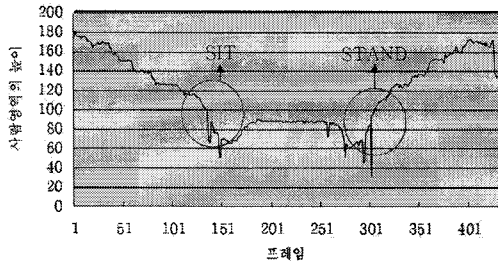
<표 4> 시선 방향 추정 결과

	0°	45°	90°	135°	180°	225°	270°	315°
0°	197	23	0	0	0	0	21	27
45°	21	193	10	0	0	2	15	4
90°	0	4	146	22	10	13	5	0
135°	5	11	45	195	21	6	22	3
180°	0	0	15	33	172	45	8	2
225°	0	0	3	9	38	165	23	0
270°	0	5	0	1	2	27	187	18
315°	39	27	12	0	5	8	29	161
합계	262	263	231	260	248	266	310	215
성공률 (%)	75.2	73.4	63.2	75.0	69.4	62.0	60.3	74.9

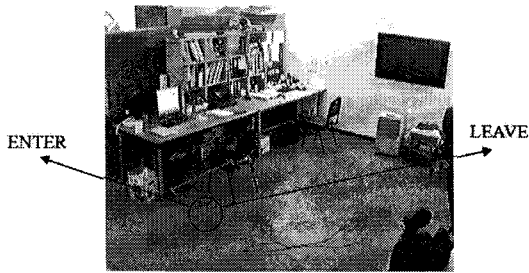
Windows XP, 2.8GHz, 메모리 2G의 시스템에서 MS Visual C++ 2008, OpenCV 라이브러리를 사용한 환경에서 처리 속도는 1 프레임당 평균 0.23 초로써 1초에 약 4 프레임 정도를 처리할 수 있는 성능을 보였다. 얼굴 영역 후보의 개수에 따라 얼굴 영역을 결정하는데 소요되는 시간에 차이가 있었다.

이벤트 중 “STAND”와 “SIT” 이벤트는 사람의 키의 변화를 통해 인식된다. 그리고 “ENTER”와 “LEAVE” 이벤트는 사람의 위치 변화에 의해 인식된다. 그림 10(a)는 “SIT”과 “STAND” 이벤트의 검출을

나타내고, 그림 10(b)는 사람의 이동한 궤적을 통해 의자 1의 영역으로의 “ENTER”, “LEAVE”이벤트가 검출 되는 상황을 나타낸다. 그림 11 부터 그림 14는 표 3에 정의된 시나리오들을 인식한 결과이다. 각 영상의 머리 영역 중심에서 밖으로 향하는 선은 시선 방향을 나타낸다.

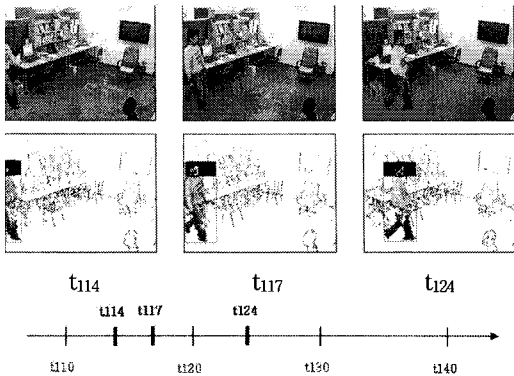


(a) 사람의 키 변화에 따른 이벤트 검출



(b) 사람의 위치 변화에 따른 이벤트 검출

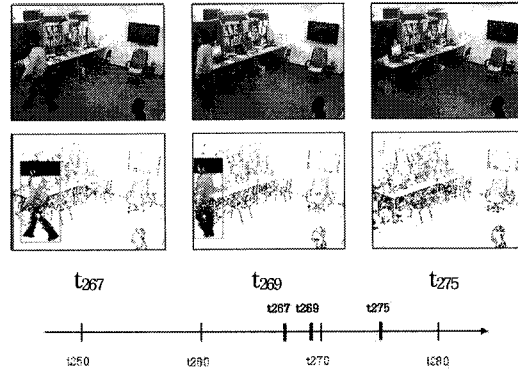
<그림 10> 이벤트 검출 상황



<그림 11> 시나리오 1의 인식 예

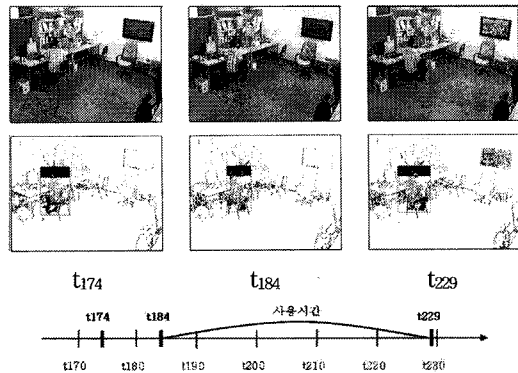
그림 11에서는 (t_{114} , enter (P_1 , “출입구”)), (t_{117} , look(P_1 , “실내방향”)), (t_{124} , leave(P_1 , “출입구”))의 이벤트가 발생하였으며, t_{124} 의 시점에서 시나리오 1이 발생하였다.

그림 12에서는 (t_{267} , look(P_1 , “실외방향”)), (t_{269} , enter(P_1 , “출입구”)), (t_{275} , leave(P_1 , “출입구”))가 순서대로 발생 하고, t_{275} 에서 시나리오 2가 인식되었다.



<그림 12> 시나리오 2의 인식 예

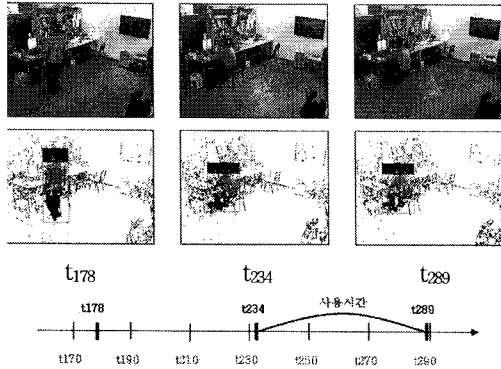
그림 13과 그림 14는 같은 의자에 앉아서 서로 다른 방향(TV와 컴퓨터)을 바라보는 상황이다. 사람이 컴퓨터 또는 의자 영역에 위치하거나 앉은 정보만을 가지고는 두 행동을 구분해 내기는 어렵지만, 제안된 방법을 통해 같은 자리에서 발생하는 서로 다른 행동을 구분할 수 있음을 보여준다. 하지만 시선 방향 인식에서 그림 13에서 사용시간은 TV를 바라본 뒤 다른 방향으로 시선이 변하지 않아야 하는 최소한의 시간을 나타내며 5초(75프레임)의 시간으로 정의하였다.



<그림 13> 시나리오 3의 인식 예

그림 14에서 사용시간은 사람이 의자에 앉은 뒤 일어서는 행동을 하지 않는 최소한의 시간을 나타내며, 3초

(45프레임)의 시간으로 정의하였다.



<그림 14> 시나리오 4의 인식 예

5. 결 론

본 논문에서는 모델 기반으로 추정된 사람의 시선 방향을 이용하여 실내 환경에서 발생 할 수 있는 사람의 행동을 인식하는 방법을 제안하였다. 제안된 방법은 크게 두 단계로 구성된다. 첫째, 행동 인식을 위한 사전 정보를 얻는 단계로 사람의 머리 영역을 검출하고 시선 방향을 추정한다. 사람의 머리 영역은 피부색 컬러 모델과 두 타원 모델을 사용하여 검출하였고, 시선 방향은 두 타원 모델의 위치와 크기 관계를 표현한 베이지안 네트워크를 이용하여 추정하였다. 둘째, 이벤트와 사람의 행동을 나타내는 시나리오를 인식하는 단계이다. 본 논문에서는 시선 방향과 연관이 있는 4 가지의 시나리오를 정의하여 실험 하였다. 실험 결과 시선 방향을 고려하지 않은 상태에서는 정확한 행동을 검출할 수 없는 상황에서도 행동 인식이 가능하였다. 하지만 저 해상도에서 사람의 시선 방향을 인식하는 것은 어렵고 낮은 인식률로 행동을 인식하는데 어려움이 발생하였다. 하지만 정확한 시선 방향을 입력으로 한 경우에는 보다 높은 성능을 보였다. 사람의 시선 방향을 보다 정확하게 추정하기 얼굴을 나타내는 타원의 회전을 가능하도록 하여 고해상도 영상에서 사람들의 다양한 행동을 인식하는 연구를 계속 진행할 계획이다.

참 고 문 헌

- [1] P. Peursum, H. H. Bui, S. Venkatesh and G. West, "Robust Recognition and Segmentation of Human Actions Using HMMs with Missing Observation," EURASIP Journal on Applied Signal Processing, vol. 2005, no. 13, pp.2110-2126, 2005.
- [2] N. M. Robertson and I. D. Reid, "A general method for human activity in video," Computer Vision and Image Understanding, vol. 104, no. 2, pp.232-248, 2006.
- [3] 차경애, 권차욱, "영상정보와 환경정보를 이용한 실내 공간 모니터링 시스템 구현," 한국산업정보학회논문지, 제14권, 제1호, pp. 1-8, 2009.
- [4] <http://www.ibm.com> "IBM introduces advanced Digital video Surveillance Service."
- [5] N. Rota and M. Thonnat, "Video Sequence Interpretation for Visual Surveillance," Proceedings of the 3rd IEEE International Workshop on Visual Surveillance, pp.59-68, 2000.
- [6] V. T. Vu, F. Bremond, and M. Thonnat, "Automatic video interpretation: A recognition algorithm for temporal scenarios based on pre-compiled scenario models," Proceedings of the 3th International Conference On Computer Vision Systems, pp.44-53, 2003.
- [7] D. Ayers and R. Chellappa, "Scenario Recognition from Video Using a Hierarchy of Dynamic Belief Networks," Proceedings of International Conference on Pattern Recognition, pp.1835-1838, 2000.
- [8] 공병용, 정도준, 김항준, "실내 환경에서 시선 방향을 고려한 사람 행동 인식," 대한전기학회, 2007 CICS 정보 및 제어 학술대회 논문집, pp. 119-120, 2007.
- [9] D. J. Jung, K. S. Kwon, S. H. Park, J. B. Kim, and H. J. Kim, "Model-Based Gaze Direction Estimation in Office Environment," 4th IEEE International Symposium on Electronic Design,

Test and Applications, pp. 470-473, 2008.

- [10] 원철호, “타원 모델기반의 전처리 기법에 의한 얼굴 인식을 개선,” 한국산업정보학회논문지, 제13권, 제4호, pp. 56-63, 2008.
- [11] R. L. Hsu, M. Abdel-Mottaleb, and A. K. Jain, “Face detection in color images,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp.696-706, 2002.



정 도 준 (Do-Joon Jung)

- 정회원
- 2004년 2월: 경북대학교 컴퓨터 공학과 (공학석사)
- 2008년 8월: 경북대학교 컴퓨터 공학과 (공학박사)
- 2005년 3월 ~ 2011년11월 : 경운대학교 정보통신공학과 교수
- 2011년 11월 ~ 현재 : 국립과학수사연구원 문서영상과 영상분석실 공업연구사
- 관심분야: 컴퓨터비전, 패턴인식, 인공지능



윤 정 오 (Jeong-Oh Yoon)

- 정회원
- 1991년 2월: 경북대학교 전자공학과 (공학석사)
- 2002년 2월: 경북대학교 전자공학과 (공학박사)
- 1997년 3월 ~ 현재: 경운대학교 정보통신공학과 교수
- 관심분야: 멀티미디어통신, 영상처리, 정보보호

논문접수일 : 2011년 06월 09일
1차수정완료일 : 2011년 07월 25일
2차수정완료일 : 2011년 09월 01일
게재확정일 : 2011년 10월 30일