

일반논문 (Regular Paper)

방송공학회논문지 제22권 제6호, 2017년 11월 (JBE Vol. 22, No. 6, November 2017)

<https://doi.org/10.5909/JBE.2017.22.6.774>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

## 카메라 비전 센서를 활용하는 실시간 사람 점유 검출

길 중 인<sup>a)</sup>, 김 만 배<sup>a)†</sup>

### Real-time People Occupancy Detection by Camera Vision Sensor

Jong In Gil<sup>a)</sup> and Manbae Kim<sup>a)†</sup>

#### 요 약

빌딩, 집에 설치되어 있는 점유센서는 사람이 없으면 소등하고, 반대이면 점등한다. 현재는 주요 센서로 PIR(pyroelectric infra-red)이 널리 사용되고 있다. 최근에 비전 카메라 센서를 이용하여 사람 점유를 검출하는 연구가 진행되고 있다. 카메라 센서는 정지된 사람을 검출할 수 없는 PIR의 단점을 극복할 수 있는 장점이 있다. 또한 카메라 센서는 사람의 행위 분석, 사람 트래킹 등 PIR이 제공할 수 없는 기능을 가지기 때문에 향후 가격 대비 성능이 만족되면 PIR을 대체할 것으로 기대된다. 본 논문에서는 PIR 센서의 단점을 극복하기 위해서 카메라를 이용한 점유센서 기법을 제안한다. 제안 방법은 트래킹, 인식, 검출의 3가지 단계로 구성되어 점유검출의 효율성을 높인다. 실시간 처리도 중요한 성능이므로 처리 속도가 향상되도록 설계되었다. 비디오 프레임은 USB로 15fps로 입력되는 데, 본 소프트웨어는 14.5fps로 처리한다. 점유 성능 검증에서는 82%의 정확도를 얻었다.

#### Abstract

Occupancy sensors installed in buildings and households turn off the light if the space is vacant. Currently PIR (pyroelectric infra-red) motion sensors have been utilized. Recently, the researches using camera sensors have been carried out in order to overcome the demerit of PIR that can not detect static people. If the tradeoff of cost and performance is satisfied, the camera sensors are expected to replace the current PIRs. In this paper, we propose vision sensor-based occupancy detection being composed of tracking, recognition and detection. Our software is designed to meet the real-time processing. In experiments, 14.5fps is achieved at 15fps USB input. Also, the detection accuracy reached 82.0%.

Keyword : Occupancy sensor, Camera sensor, Tracking, Recognition, Human Detection

---

a) 강원대학교 컴퓨터정보통신공학과(Dept. of Computer and Communications Engineering, Kangwon National University)

† Corresponding Author : 김만배(Manbae Kim)

E-mail: manbae@kangwon.ac.kr

Tel: +82-33-250-6395

ORCID: <http://orcid.org/0000-0002-4702-8276>

※ 이 논문은 2017년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. 2017R1D1A3B03028806).

· Manuscript received July 25, 2017; Revised September 11, 2017; Accepted September 11, 2017.

## I. 서론

현재 많은 건물에는 사람이 존재하면 점등하고(light on), 사람이 없으면 소등하는(light off) 점유센서(occupancy sensor, motion sensor)가 설치되어 있다. 전기소비를 줄이기 위해 설치된 것으로 대부분의 점유센서는 현재 PIR(pyroelectric infra-red)을 사용하고 있다<sup>[1,2]</sup>. PIR은 사람의 신체에서 발생하는 열(thermal temperature)을 감지하고, 열의 변화량을 계산하여 모션을 측정한다. 따라서 사람이 정지 상태에 있을 때에는 검출 성능이 낮아진다. 이 센서는 1) 정지된 사람의 검출이 불가능하고, 2) 뜨거운 공기의 흐름 또는 햇빛 복사(radiation)에 민감하고, 3) 사람이 아닌 생물체와 사람을 구분하지 못하는 등의 단점이 있다.

최근에 PIR 대신에 카메라 비전센서를 활용하려는 연구가 진행되고 있다<sup>[3-6]</sup>. 카메라 센서는 위에서 언급한 PIR의 단점을 극복할 수 있고, 부가적으로 사람 추적, 사람 명수 파악, 사람 행위 등 다양한 지능 정보의 획득이 가능하다.

Benezeth 등은 CAPTHOM 프로젝트의 일환으로 비전 센서를 활용하여 사람 점유 및 행위를 분석하는 연구를 수행하였다<sup>[3]</sup>. 이를 위해 얼굴 검출 및 트래킹 기술을 활용하였다. 측면에 카메라를 고정하였고, 최대 2명의 사람에 대해서 실험하였다. Han 등은 비전센서와 PIR를 이용하여 사람검출 연구를 수행하였다<sup>[4]</sup>. Nakashima 등은 카메라 센서를 활용하여 움직이는 사람을 검출하는 방법을 제안하였다<sup>[5]</sup>. 한 명에 대한 실험이어서, 여러 사람이 있는 경우에는 적용의 어려움이 있다. Amin 등은 PIR과 비전센서를 이용하여 사람의 명수를 구하는 연구를 진행하였다<sup>[6]</sup>. 많은 연구에서는 PIR과 비전센서를 동시에 사용하여 성능을 높이지만, 단점은 2개 이상의 이중 센서를 사용하는 것이다. [3, 5]에서는 카메라만 사용하여 점유센서를 제안하지만, 사람의 가려짐이 발생할 때에 사람 검출의 성능이 저하되고, 특히 장시간 정지된 사람에 대한 처리 방법이 부족하다.

본 논문에서는 실내 공간에서 사람이 입실하면 점등하고, 모든 사람이 퇴실하면 소등되는 점유센서 기법을 제안한다. 비전센서는 천장에서 수직으로 바닥(ground floor)을 촬영하는 오버헤드(overhead) 카메라를 이용한다. 공간에는 많은 사람들의 입실, 퇴실이 반복되고, 의자, 소파, 책상 등 다양한 가구들이 공존하고 있다. 또한 사람은 장시간 정

적 상태에 있을 수 있기 때문에, 움직임이 없는 사람을 계속 추적(tracking)할 수 있어야 한다. 또한 비휴먼 객체들은 트래킹을 중단하거나, 제거되어야 한다. 기본적으로 사람의 명수 계산(people counting)을 하고 명수가 0이 되면 소등한다.

입실하는 사람을 탐색하고, 퇴실할 때까지 추적하는 것을 기본 방법으로 사용한다. 배경 모델링과 전경추출에서는 배경의 지속적인 업데이트 및 다수의 전경객체를 찾는다. 사람과 객체의 움직임 때문에 모션은 사람 검출 및 추적에 적합한 데이터이다. 모션 정보로 운동 히스토리 영상(motion history image, MHI)을 활용하여 객체의 추적 성능을 높이고자 한다<sup>[7]</sup>.

활용 측면에선 PIR과 유사한 가격 대비 성능 경쟁력을 갖기 위해서는 카메라 센서 알고리즘은 실시간 처리가 가능해야 한다. 이를 만족하기 위해서 제안 기술은 실시간 처리가 가능하도록 설계되었다. 실험에서 서술되었지만, 15fps로 USB에 입력되는 RGB 비디오를 소프트웨어는 14.5fps로 처리되고 있다. 영상해상도는 720x480이다.

본 논문의 구성은 다음과 같다. 다음 장에서는 전체적인 제안방법을 설명하고, III장에서는 점유센서에서 필요한 전처리 및 정보 획득에 대해 설명한다. 이어서 IV, V, VI장에서는 점유센서를 구성하는 단계들을 자세히 서술한다. 실험 결과는 VII장에서 정리하고, 마지막으로 VIII장에서는 결론 및 향후 연구를 정리한다.

## II. 제안 방법

점유센서의 전체 흐름도는 그림 1에서 보여진다. 제안방법은 기본 정보(basic information) 생성, 트래킹(tracking), 인식(recognition), 및 검출(detection)의 4단계로 구성되고, 각 단계의 간략적인 내용은 다음과 같다.

기본정보 생성 단계에서는 전경(foreground) 및 배경(background)을 추출한다. 또한 연속 영상  $I_t$ 와  $I_{t-1}$ 의 차영상(difference image)을 구한다. 영상의 모션을 얻기 위해 MHI (motion history image)를 활용한다. 사람이 없을 때에 장면의 참조영상(reference image)을 저장해둔다. 참조영상과 현재영상의 차영상을 구한 후에, MHI의 가중치 합으로

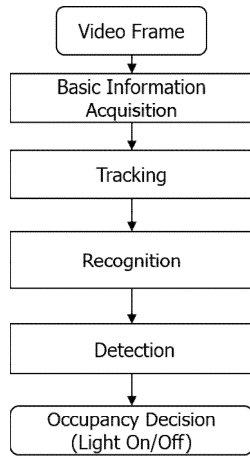


그림 1. 제안 방법의 전체 흐름도

Fig. 1. The overall block diagram of the proposed method

MHI 에너지 영상을 얻는다.

트래킹 단계에서는  $I_t$ 와  $I_{t-1}$ 을 비교하여 트랙커(tracker)의 새로운 추적점을 찾는데, 사각형 경계박스(bounding box) 기반으로 이동한 트랙커의 위치를 구한다. 경계박스 단위로 사람을 추적하기 때문에 사람의 영역이 넓으면 2개 이상의 경계박스가 발생한다. 이 방법은 파트기반모델(part based model)<sup>[8]</sup>에서 활용하는데 각 경계박스를 독립적으로 트래킹한다.

인식 단계는 경계박스중에서 의자 등의 사람이 아닌 경계박스 및 퇴실하는 사람을 제거한다. 트래킹 실패는 항상 발생하기 때문에 인식은 필수 기능이다.

검출 단계에서는 내부에 새로 입실하거나, 또는 내부에서 트래킹의 실패로 제거된 사람들을 재검출한다.

마지막으로 내부에 사람이 없으면 소등하게 된다. 반대로 1명이라도 입실상태이면 계속 점등한다.

### III. 기본 정보 생성

사람 점유를 판단하기 위해서 매 프레임마다 기본 정보들을 새로 획득하거나 갱신한다. 배경 모델링 및 전경 추출은 기본적으로 필요한 정보로 트래킹 등 다양한 단계에서 활용된다. 참조영상은 공간내 사람이 존재하지 않을 때에 저장되는 영상이다. 장기간 정지상태인 트랙커는 알고리즘

적으로 판단이 쉽지 않기 때문에 대안으로 참조영상을 활용한다.

트래킹은 보통 RGB 컬러를 활용하는데, 단순 컬러의 활용은 배경과 전경객체의 색의 유사성이 존재하면 트래킹 성능이 저하된다. 이를 보완하기 위해서 MHI 기반으로 트래킹을 수행한다.

전경객체를 제외한 배경을 구하는 배경모델링 방법들인 GMM(Gaussian Mixture Model)<sup>[9]</sup>, codebook<sup>[10]</sup> 등은 실시간처리에 적합하지 않다. 따라서 식 (1)의 방법을 이용한다<sup>[11]</sup>. 이 방법은 배경영상  $BG_{t-1}$ 의 픽셀값이 현재 영상  $I_t$ 보다 작으면 1을 더해주고, 반대이면 1을 감소한다.

$$BG_t^i = \begin{cases} BG_{t-1}^i + 1 & \text{if } BG_{t-1}^i < I_t^i \\ BG_{t-1}^i - 1 & \text{if } BG_{t-1}^i > I_t^i \end{cases} \quad (1)$$

여기서  $i$ 는 픽셀,  $t$ 는 프레임 번호이다.

BG는 항상 현재영상으로 접근하기 때문에 사람이 정지하면 사람은 배경으로 흡수된다. 다양한 기법이 소개되어 왔지만 상기 문제는 해결이 어려운 문제이다. 이를 극복하기 위해서 다음의 참조영상을 활용한다

참조영상(reference image)은 공간에 사람이 없을 때에 저장되는 영상이다. 단순히 빈 공간일 때에 저장할 수 있지만, 시간축으로 조명의 변화가 지속적으로 발생하기 때문에 이 저장된 영상과 현재영상과의 차이가 크게 된다. 이를 해결하기 위해서 프레임마다 트랙커의 경계박스를 고려하여 갱신된다. N개의 트랙커가 있으면 N개의 경계박스가 존재한다. 경계박스 내부에 있는 픽셀은  $\alpha=1$ 을 가지고, 아니면 0을 가진다. 이를 이용하여 참조 영상 Ref는 다음 식에서 구해진다.

$$Ref^i = \alpha Ref^i + (1 - \alpha) BG^i \quad (2)$$

경계박스가 존재하지 않으면 배경영상 BG로 접근하고, 아니면 경계박스는 이전 참조영상 값을 유지한다. 즉,  $\alpha=1$ 이면 Ref는 변화가 없고,  $\alpha=0$ 이 되면 Ref=BG가 된다.

참조영상은 버려진(abandoned) 객체를 판별하기 위해서 사용되는데, Tian<sup>[12]</sup> 등은 그림 2에서 보는 것처럼 가방이 놓이면, 이전 영상을 배경영상으로 저장해둔다. 배경의 히

스토리 정보를 이용하여 정적영역을 버려진 객체 또는 제거된 객체로 결정한다.

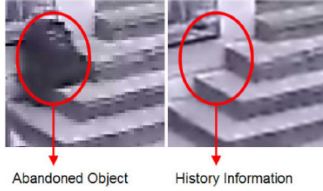


그림 2. 배경 히스토리 정보를 이용한 객체의 종류 판단<sup>[12]</sup>  
Fig. 2. Decision of object types using background history information

Bovick 등이 제안한 운동정보 표현식인 MHI<sup>[7]</sup>는 주로 사람의 행위 및 포즈를 인식하는 행위 검출 분야에서 활발하게 이용되어 왔다. MHI의 기본 정의는 다음과 같다. 식 (3)과 같이 연속영상  $I_t$ ,  $I_{t-1}$ 로부터 차영상  $\Delta_t$ 를 구한다.

$$\Delta_t^i = |I_t^i - I_{t-1}^i| \quad (3)$$

여기서  $i$ 는 픽셀이다.  $\Delta_t$ 로부터 MHI는 식 (4)와 같이 구한다.

$$MHI_t^i = \begin{cases} \tau & \text{if } \Delta_t^i > T \\ \text{MAX}(0, MHI_{t-1}^i - \delta) & \text{otherwise} \end{cases} \quad (4)$$

윗식에서  $\tau$ 는 사람의 행위가 지속되는 프레임의 개수이고,  $\delta$ 는 감소(decay) 파라메타이다.  $\Delta^i > T$ 이면 MHI 값은  $\tau$ 가 되고, 아니면  $\delta$ 만큼 감소하게 된다. 최근 움직임이 있는 픽셀은 큰 값을 가지는 특성을 가지고 있다.  $\delta$ 가 크면 픽셀의 MHI는 급격히 감소하게 된다.

실험에서는  $\tau=255$ 를 사용하는데, 픽셀의 변화가 발생하면 MHI=255이다. 반대이면  $\delta$ 만큼 삭감한다. 그림 3은 연속영상과 MHI의 예를 보여준다. 이 MHI는 트래킹에서 효과적으로 활용한다.

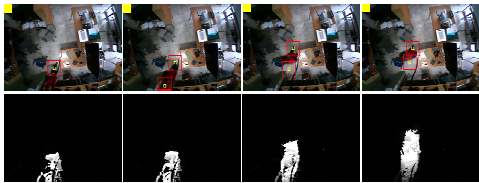


그림 3. 연속영상과 MHI  
Fig. 3. Subsequent images and their MHIs

참조영상과 현재영상과의 차영상을 얻은 후에, 이 차영상과 MHI의 가중치 합으로 MHI 에너지 MHIE(MHI Energy)를 얻는다.

$$MHIE^i = \omega MHI^i + (1 - \omega) |Ref^i - I^i| \quad (5)$$

여기서  $\omega \in [0, 1]$ 이다.

조명에 의해서 발생하는 그림자(shadow)는 영상 픽셀 값을 변화시키기 때문에 모션으로 판단하는 되는 경우가 발생한다. 이 영역은 사람이 없음에도 객체로 인식하게 된다. 즉, 그림자는 픽셀값의 변화로 모션 객체로 오판되는 (false alarm) 경우가 발생한다. 그림자는 노이즈와 달리 정상 데이터이지만 사람이 없는 곳에서 객체로 추출되는 확률이 높기 때문에 가능하면 그림자의 영향을 줄여야 한다<sup>[13]</sup>. 그림자는 주로 RGB를 사용하여 검출한다. 영상 RGB와 배경 RGB의 차이값 중에서 최대값을 구한다.

$$\rho^i = \text{MAX}(|R^i - BG_R^i|, |G^i - BG_G^i|, |B^i - BG_B^i|) \quad (6)$$

$\rho$  값이 임계치 T보다 작으면 증감 값  $w$ 는 +1, 아니면 -1을 가진다.

$$w^i = \begin{cases} +1 & \text{if } \rho^i < T \\ -1 & \text{otherwise} \end{cases} \quad (7)$$

실험에서는 T=5를 사용한다. 그림자의 강도는 조명에 의존하기 때문에 최적 T값을 구하는 것은 어렵다. 따라서 실험실에서 다양한 조명에서 성능을 검증한 결과 5가 만족스러운 결과를 주었다.

그림자 영상 Shadow는 다음 식처럼  $w$ 를 더해지고, 프레임이 입력되면 계속 갱신된다.

$$Shadow_t^i = Shadow_{t-1}^i + w^i \quad (8)$$

그림자가 아닌 픽셀의 명도값은 지속적으로 증가, 감소하지 않고, 랜덤하게 변화한다. 반대로 그림자는 픽셀값이 계속 낮아지기 때문에 (그림자는 컬러를 어둡게 함), 그림자 픽셀은 대부분  $|Shadow_t^i| > T$  조건을 만족한다. 좀 더



자세히 설명하며 정상 픽셀의 노이즈는 랜덤하기 때문에, 계속 더해지면 0으로 수렴하게 된다. 왜냐하면 확률적으로 +1 또는 -1 값을 동일 확률로  $w$ 에 더해주기 때문이다. 반면에 그림자 영역에서는 컬러값이 점진적으로 낮아진다. 따라서  $w$ 는 계속 -1 값만 더해진다.

#### IV. 트랙킹 단계

PIR 센서의 단점중의 하나는 사람 트랙킹의 어려움이다. 트랙킹은 이동하는 사람이 정지상태에 있을 때에도 계속 점유로 판단하기 위해서 필요하다. 트랙킹에는 커널 기반<sup>[14]</sup> 등이 주로 사용되고 있는데, 본 방법에서는 실시간 처리를 위해 MHIE 기반으로 트랙킹을 수행한다.

N개의 트랙커가 존재하면, 독립적으로 트랙킹을 수행한다. 따라서 1개의 트랙커 처리가 1 sec이면, N sec가 소비되게 된다. 특히 제안 방법의 단계들 중에서 처리 속도가 가장 높고, 따라서 속도 개선이 중요하다. 또한 성능을 저하시키는 트랙커의 표류(drifting) 및 중첩(occlusion)은 항상 발생한다.

##### 1. 정적 트랙커 조사

실내공간은 창문 커튼의 흔들림, 내부 외부 조명의 변화, 그림자의 영향을 많이 받는 환경이다. 따라서 정적 객체에 트랙킹 알고리즘을 적용하면 컬러의 변화로 추적점이 변화하게 된다. 따라서 먼저 트랙커가 정지상태인지를 조사하고, 정지이면 현재의 추적점을 유지한다. 연속영상  $I_t$ 와  $I_{t-1}$ 을 이용하여 트랙커  $n$ ,  $BB_n$ 의 차영상을 구하다. 다음에 픽셀의 절대 차이값이 T보다 작은 픽셀의 개수를 구한 후에, BB의 픽셀 개수로 나눈  $\rho$ 를 계산한다.  $\rho$ 가 0.04 이하이면 정적 트랙커로 판단한다. 이 경우  $BB_n$ 의 추적점  $TP_n$ 는 변경하지 않고, 이전 프레임  $I_{t-1}$ 에서의 값을 그대로 사용한다.

##### 2. 제안 트랙킹 방법

먼저 트랙킹의 탐색창(search window)을 설정한다. 탐색

창이 작으면 빠른 움직임의 추적이 어려워서, 트랙킹 표류(drifting)가 발생한다. 그 이유는 이동한 사람이 탐색창 외부에 위치되기 때문이다. 반대로 크면 처리속도의 저하가 발생한다. 탐색창의 크기와 추적 성능의 최적 관계를 구하는 것은 어려운 문제이다. 실내에서는 천장높이가 보통 3m 정도이므로, 영상에서 사람의 이동은 프레임간 상당한 거리로 이동하게 된다. 이러한 물리적 현상은 트랙커 표류의 원인을 제공한다.

트랙커의 경계박스에 대해 픽셀 단위로 전역 탐색을 하면 처리속도가 매우 높다. 속도를 줄이는 방법으로 탐색창을 8x8개의 블록(block)으로 분할하고 (그림 4(b)), 각 블록에는 식 (5)에서 구한 MHIE의 합  $E_k$ 를 저장한다 (그림 4(c)). k는 블록의 번호이다. 그림 4(c)는 그림 4(d)의 MHI 에너지의 관심 영역을 확대한 영상이다.

$$E_k = \sum_{i \in \text{block}(k)} \text{MHIE}_i \quad (9)$$

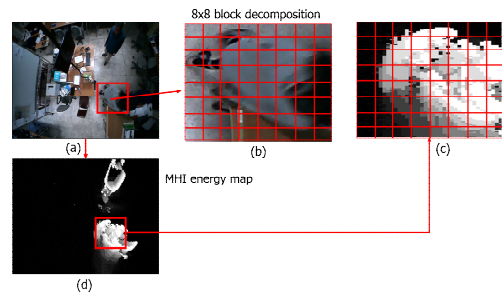


그림 4. 탐색창을 8x8 block으로 분할후, 블록에 픽셀의 MHI의 합을 저장함  
Fig. 4. Saving the sum of all pixel MHIs into a 8x8 block after decomposition of a search window

이 탐색창 내에서 3x3 블록의 MHIE 합이 가장 큰 블록을 찾고, 중심 블록이 트랙커  $n$ 의 추적점  $TP_n$ 이 된다. x, y 축으로 1블록씩 이동하기 때문에, 픽셀단위로 전역탐색을 수행하는 것보다는 처리 시간이 대폭 감소된다.

$$TP_n = \operatorname{argmax}_n \left( \sum_{k \in 3 \times 3 \text{ block}} E_k \right) \quad (10)$$

그림 5는 제안방법으로 트랙킹 결과 영상을 보여준다. 각 영상의 MHIE는 아래 부분에서 보여진다. 한 사람이 이동

하면 두 개의 트랙커가 발생하고, 계속 사람을 추적하다가 한 개는 계속 사람을 트랙하고 다른 한 개는 사라지는 예를 보여준다. 적색 경계박스는 모션 트랙커이고, 청색은 정지 트랙커이다. 이 파트 기반 모델은 객체당 1개의 트랙커를 할당하는 것보다는 트랙킹 실패율을 줄일 수 있다.

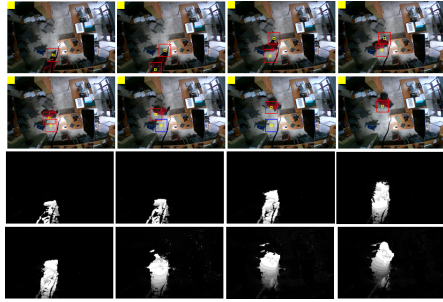


그림 5. MHI 에너지 기반 트랙킹의 결과 영상  
Fig. 5. Resulting tracking images obtained by the MHI energy-based tracking method

## 2. 가려진 사람 트랙킹

실제 두 명 이상이 중첩(overlap)되는 double tracker, multi-person tracker 경우가 종종 발생한다. 여러 사람이 겹치는 현상으로 가려짐 비율(occlusion level)이 증가하면 검출 확률이 대폭 감소하게 된다. continuous energy minimization problem, 객체 속도 및 위치의 활용, 다중 카메라의 활용, 모션과 컬러 특징 추출 등의 기법들이 제안되었지만, 검출 정확도는 50% 이하로 알려져 있다<sup>[15,16]</sup>. 그림 6은 가려짐이 발생하는 예를 보인다. 트랙커 내부로 타 트랙커가 침입되기 때문에, 혼합된 컬러, 모션 정보가 발생하게 된다.



그림 6. 가려짐이 발생하는 영상. 황색 경계박스는 중첩 영역임  
Fig. 6. Images with overlapped people. Overlapped regions are marked by a yellow bounding box

이를 해결하기 위해 제안방법은 그림 7에서 설명된다.  $I(t-1)$ 에서  $Tracker_B$ 는 다른 트랙커의 경계박스와 겹침이 없

지만 (그림 7(a)), 이동으로  $I(t)$ 에서  $Tracker_A$ 와 중첩이 발생한다 (그림 7(b)). 이때  $Tracker_B$ 를  $Tracker_A$ 와의 반대 방향으로 이동하여 중첩의 영향을 줄인다 (그림 7(c)).

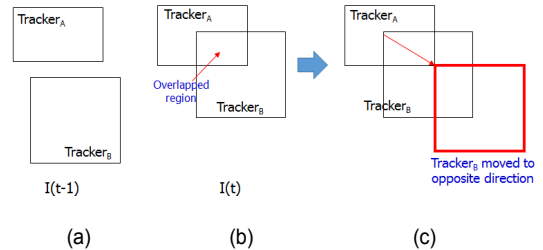


그림 7. 가려짐의 처리.  $I(t)$ 에서 overlap이 발생하면 (c)처럼 경계박스를 반대 방향으로 이동시킴

Fig. 7. Handling occlusion: When TrackerB is overlapped with TrackerA, TrackerB is forced to the opposite direction

트랙커 이동은 상대적 위치 정보를 필요로 한다. 가려진 트랙커의 위치를 구하기 위해서 4-bit region outcode를 이용한다<sup>[17]</sup>. 그림 8은 outcode를 보여주는데, 가운데 사각형 윈도우를 중심으로 점이 윈도우 내부에 있으면 0000이고 점의 위치에 따라 bit값이 1이 되는 기법이다. 트랙커의 새로운 추적점이 1001에 있고, 중첩이 발생하면, 가운데 경계박스의 반대 방향으로 이동시킨다. 이동 방향은 경계박스들의 좌표를 알고 있기 때문에 계산이 가능하다. 그림 9는 중첩 발생할 때에 얻어지는 트랙커들을 보여준다.

1001	1000	1010
0001	0000	0010
0101	0100	0110

그림 8. 4-bit 영역 아웃코드

Fig. 8. 4-bit region outcode

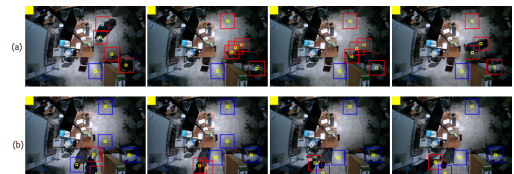


그림 9. 중첩이 발생하는 프레임에서 트랙킹하는 2개의 연속 영상  
Fig. 9. Two subsequent images showing trackers' movement under overlap

## V. 인식 단계

본 시스템에서는 트랙커 제거는 정지상태에서만 활성화 되도록 설계되었다. 이동하는 트랙커는 사람일 확률이 크고, 또한 움직이는 가구들은 정지상태가 되면 제거한다. 인식(Recognition) 단계의 주 목적은 정적 트랙커가 사람인지 아닌지를 판단하는 것이다. 움직이다가 정지 상태인 의자나, 그림자 등은 사람이 아니어서 제거해야 한다. 반면에 이동하는 사람, 오랜 정지 상태에 있는 사람은 계속 트랙킹되어야 한다. 또한 트랙커는 표류하면서 다른 물체 또는 배경을 따라가게 된다. 다른 예는 동일인에게 2개 이상의 트랙커가 할당되는 경우에 1개의 트랙커는 동일 사람을 추적하지만, 다른 트랙커는 표류되기도 한다.

MHI 에너지의 평균이  $T$ 보다 작으면 제거한다. 왜냐하면, 배경은 계속 현재영상으로 수렴하기 때문이다. 에너지의 평균은 다음 식에서 계산된다.

$$Energy_{BB} = \frac{1}{N_{BB}} \sum_{i \in BB} MHIE_i \quad (11)$$

여기서  $N_{BB}$ 는 트랙커 경계박스의 면적이다. 이 에너지값이 임계치  $T$ 보다 크면 트랙커는 유지하고, 아니면 제거한다.

$$BB^i = \begin{cases} Maintain & \text{if } Energy_{BB} > T \\ Reject & \text{otherwise} \end{cases} \quad (12)$$

실험에서는  $T=12$ 을 사용하였다. MHIE에는 참조영상과 MHI이 포함되는데, 시간축으로 MHI는 0으로 수렴하지만, 참조영상은 변화가 없다. 따라서  $T$ 보다 작으면 배경으로 판단하게 된다.

그림 10은 인식단계에서 사라지는 사람이 아닌 경계박스를 보여준다. 이 경계박스의 트랙커들은 트랙커 리스트에서 삭제된다.



그림 10. (a)에서 우측 상단의 경계박스가 사라짐. (b)에서는 우측 상단의 박스가 사라짐  
Fig. 10. Bounding boxes at the top-left are removed. Other boxes are maintained

## VI. 검출 단계

검출(detection) 단계에서는 새로 입실하는 사람을 탐색하고, 트랙킹 실패가 발생한 사람을 재검출한다. 빠른 검출을 위해서 영상을  $30 \times 30$  블록으로 분할한다. 각 블록에서 트랙커의 경계박스가 포함되어 있는 픽셀의 개수를 계산한 후에, 50% 이상이면 해당 블록의 이진 마스크 값을  $Mask=255$ 로 한다. 이 블록에서는 기존 경계박스가 존재하므로 트랙커를 탐색하지 않는다.

이진 마스크 영상에 최대치 필터를 적용하여 기존 경계박스에 가까운 곳에서는 발생하지 않도록 한다.  $Mask=0$ 인 블록을 찾고, 이 블록의 MHIE 합이 임계치 이상이면, 이 블록의 좌표를 저장해둔다. 이 과정을 거치면  $K$ 개의 블록이 얻어지면, 이중 가장 MHIE 합이 큰 1개의 블록을 찾아서 새로운 트랙커로 등록한다. 매 프레임마다 최대 1개의 트랙커가 생성된다.

그림 11은 검출단계에서 새로 생성되는 트랙커의 예를 보여준다. 그림 11(a)에서는 실내로 들어오는 사람을 검출하는 연속 영상이고, 그림 11(b)는 가운데 의자에 착석해 있는 사람이 트랙킹 실패로 사라졌다고, 다시 검출되는 예를 보여준다.

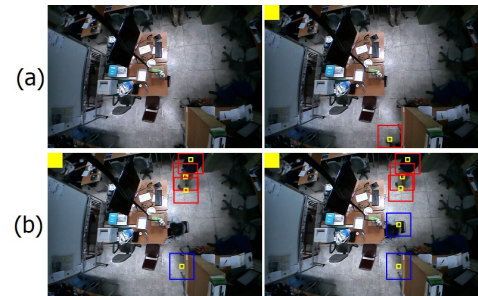


그림 11. 검출 단계에서 새로 생성되는 트랙커의 예  
Fig. 11. Examples of new trackers generated in detection part

## VII. 실험 결과

컴퓨터의 USB로 비디오는 프레임율 15fps로 입력되며, Visual C/C++ 기반 점유센서 소프트웨어는 14.5fps로 실시간 처리한다. 영상 해상도는  $720 \times 480$ 이고 RGB입력이다.

실험에서 사용한 컴퓨터의 사양은 CPU는 Intel core i7, 2.20GHz, 운영체제는 64bit Window 7, 4GB 메모리이다. 실험에서는 연구실에서 카메라를 천장에 수직으로 설치하였다. 천장높이는 2.7m이다.

제안 방법은 트래킹-인식-검출 단계로 구성되어 있는데, 각 단계별로 속도 개선을 하였다. 트래킹단계에서는 블록 단위로 이동을 추적할 수 있도록 해서 처리시간의 상당한 단축 효과를 얻었다. 인식단계에서는 트랙커 경계박스의 MHIE의 값으로 사람을 판별하게 하였다. 검출단계에서는 사람이 이동하면 발생하는 블록 MHIE값으로 입실하는 사람을 검출하고, 또한 트래킹 실패로 발생하는 사람들을 재검출하게 하여 전체적으로 처리시간의 감소를 얻었다. 트랙커의 개수가 증가해도 처리속도의 큰 저하는 발생하지 않는다.

실시간 처리는 성능 검증이 어렵기 때문에 프레임을 저장한 후에, 오프라인으로 검증을 하였다. 12개의 비디오 세트를 실험영상으로 사용하였고, 세트당 500~2,000장이다. 그림 12는 각 세트에서 한 장을 얻은 한 장의 영상을 보여준다. 촬영 시간은 오전, 오후, 밤에 촬영하여 다양한 조명 조건을 가지도록 하였다. 실내에 있는 사람은 1~4명으로 조정하였는데, 각 세트의 최대 명수는 표 1에서 보여진다. 특히 명수가 3이상인 4개의 세트에서는 가려짐이 많이 발생하도록 사람의 이동을 조절하였다. 가려짐의 비율은 50% 이상도 다수 발생한다.



그림 12. 12장의 비디오 세트에 있는 한 장의 영상들  
Fig. 12 One frame included in twelve video sets

점유센서의 객관적 성능 비교를 위한 실험영상은 오버헤드 카메라 영상을 포함해서 국내외적으로 제공되지 않고 있다. 따라서 본 실험에서는 실험실에서 촬영한 비디오를 이용하여 성능을 검증한다. [3]에서 일부 성능 결과를 보여주고 있지만, 최대 2명에 국한된 실험이라 비교가 어렵다. 점유센서의 성능을 검증하기 위해서 다음 3가지의 성능측정 방법을 이용한다; ① 사람 계수의 정확도, ② 혼돈 행렬 (confusion matrix), 및 ③ 점유 구간의 정확도.

### 1. 사람 계수의 정확도

사람 계수의 정확도를 분석하기 위해서 Mean Absolute Error (MAE)와 Mean Relative Error (MRE)를 사용한다.

$$MAE = \frac{1}{N} \sum_{k=1}^N |G^k - E^k| \quad (13)$$

$$MRE = \frac{1}{N} \sum_{k=1}^N \frac{|G^k - E^k|}{G^k} \quad (14)$$

여기서  $G_k$ 는 프레임  $k$ 의 실측(ground truth) 계수이고,  $E_k$ 는 측정 계수이다.  $N$ 은 프레임 수이다.

표 1은 12개의 비디오의 최대 명수  $P_{max}$ , 그리고 MAE와 MRE를 보여준다. 실측 값은 매 프레임별 사람의 명수를

표 1. 사람 계수의 MAE와 MRE.  $P_{max}$ 는 최대 사람 명수임  
Table 1. MAE and MRE of people counting.  $P_{max}$ = maximum no. of people

Video	$P_{max}$	MAE	MRE(%)
V03	1	0.31	63.30
V04	1	0.54	59.46
V10	1	0.44	34.91
V12	1	0.30	73.68
V20	1	0.24	41.71
V21	3	0.80	43.13
V40	4	0.64	30.86
V41	4	0.66	30.84
V50	1	0.66	30.84
V51	2	0.45	39.85
V60	1	0.29	42.70
V71	3	1.51	69.97
Avg.		0.57	46.77



수작업으로 계산하였다. 평균 MAE는 0.57명, MRE는 46.77%이다. 본문에서 언급했듯이 하나의 객체당 다중 트랙커가 발생할 수 있는 상황에서 0.57은 만족스러운 결과로 판단된다.

실험 비디오중에서 Video40와 Video50의 사람 계수 결과는 그림 13에서 보여진다. x축은 프레임 번호이고, y축은 사람계수이다. 적색선은 실측된 사람계수이고, 흑색선은 측정된 사람계수이다. 파트기반 모델로 사람을 추적하기 때문에, 한명의 사람당 2~3개의 트랙커를 가질 수 있다 (표 1에서 Video40의 MRE=0.64, V50의 MRE=0.66). Video40의 프레임 80, 프레임 700에서는 사람계수가 0이므로 소등하게 된다.

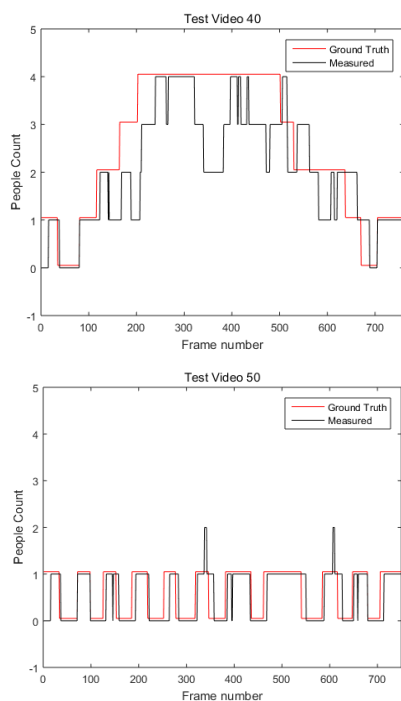


그림 13. 2개의 실험 비디오에서 얻어진 실측 및 측정된 사람 계수  
Fig. 13. People counting results on two videos with ground-truth and measurement

## 2. 혼돈 행렬

두 번째 실험 결과는 표 2, 3에서 보여진다. 12개의 비디오의 10,267 프레임에 대해 점유(occupied) 및 비점유(vacant)의 혼돈행렬을 측정한다. 이 행렬의 4가지 요소는 TP

(True Positive), TN(True Negative), FP(False Positive), FN(False Negative)이다. 10,267 프레임에서 TP, FN, FP, TN의 합은 각각 6,478, 1,136, 888, 2,248이다. 이 요소로부터 혼돈행렬은 표 3에서 보여진다. a는 TP, b는 FN, c는 FP, d는 TN이다. false alarm은 0.28, misdetection은 0.15이고, correct detection은 0.84, 0.71이다. 향후 성능개선을 위하여 오류율을 감소하는 것이 필요하다는 것을 알 수 있다.

표 2. 10,267 프레임의 검출 결과

Table 2. Detection output of 10,267 frames. (TP=True Positive, TN=True Negative, FP=False Positive (False Alarm), FN=False Negative)

Video	TP	FN	FP	TN
V03	90	14	22	92
V04	120	86	23	35
V10	670	16	201	104
V12	302	312	47	589
V20	500	88	78	233
V21	470	53	54	42
V40	661	15	22	58
V41	798	4	36	47
V50	338	77	49	286
V51	951	71	62	123
V60	1163	373	288	604
V71	415	27	6	35
Sum	6,478	1,136	888	2,248

표 3. 10,267 프레임의 점유 및 미점유 혼돈 행렬

Table 3. Confusion matrix of occupancy and vacancy of 10,267 frames

		Classification Result	
		occupied	vacant
Ground Truth	occupied	a/(a+b)	b/(a+b)
	vacant	c/(c+d)	d/(c+d)

		Classification Result	
		occupied	vacant
Ground Truth	occupied	0.845	0.155
	vacant	0.284	0.716

## 3. 점유 구간의 정확도

성능검증에서 중요한 것은 점등/소등의 정확도이다. 사람이 없는 구간에서는 소등되고, 반대로 사람이 있는 기간에는 계속 점등되면 성공, 한번이라도 꺼지면 실패한 경우를 분석한다. 2명이 실내에 있는데, 1명만 검출되어도 실제

시스템은 문제가 없다. 왜냐하면 사람계수가 0보다 크므로, 점등이 지속되기 때문이다. 이 구간 평가 결과는 표 4에서 보여진다. 이전 실험과 다른 것은 사람이 3명인데, 트랙커의 개수가 2이더라도 점등되기 때문에 성공으로 판단된다. 사람 미점유 구간은 총 58회이고, 소등은 49회, 점등은 9회가 발생하였다. 정확도는 84.5%이다. 점유구간은 총 89회이고, 점등은 72회, 소등은 17회가 발생하였다. 정확도는 80.9%이다. 표 5는 표 4의 혼동행렬을 보여준다.

표 4. 점유 및 미점유 구간에서의 소등, 점등 횟수. Acc=Accuracy의 단위는 %임

Table 4. Frequency of light on/off during occupancy and vacancy periods

No.	Vacancy frequency	Light Off	Light On	Occupancy frequency	Light On	Light Off
V03	4	4	0	3	3	0
V04	1	1	0	1	0	1
V10	8	5	3	9	9	0
V12	6	5	1	11	9	2
V20	5	5	0	9	7	2
V21	2	1	1	7	4	3
V40	2	1	1	2	2	0
V41	3	2	1	3	3	0
V50	9	9	0	13	12	1
V51	5	4	1	5	5	0
V60	11	11	0	23	16	7
V71	2	1	1	3	2	1
Sum	58	49	9	89	72	17
Acc.		84.5	15.5		80.9	19.1

표 5. 147개의 점유 구간의 혼동 행렬

Table 5. Confusion matrix of 147 occupancy periods

		Classification Result	
		occupied	vacant
Ground Truth	occupied	0.809	0.191
	vacant	0.155	0.845

## VIII. 결론 및 향후 연구

본 논문에서는 카메라 비전센서를 활용하는 점유센서 방법을 제안하였다. 기법은 기존정보를 바탕으로 트랙킹, 인식, 및 검출의 단계로 구성된다. 기존의 PIR 모션센서를 대

체할 수 있는 카메라 센서는 공간에서의 지능형 정보, 보안 등의 부가적 효과를 얻을 수 있다는 점에서 가치가 있다. 본 점유센서는 사람이 한명이라도 점유되면 점등하고, 아니면 소등하는 목적에 초점을 두었다. 개발된 소프트웨어는 실시간으로 처리되도록 설계되었다. 이를 위해서 트랙킹의 단순화, 중첩 트랙커 처리 등 새로운 기술들을 활용하였다. 82%의 정확도를 보여주는데 향후 정확도의 성능 개선이 필요하다. 또한 이동하는 가구 등과 사람을 정확히 분류하는 기술의 개선도 필요한데, 최근 많이 활용하는 있는 딥러닝인 인공지능망을 활용할 계획이다.

## 참 고 문 헌 (References)

- [1] P. Liu, S. Nguang, and A. Partridge, "Occupancy inference using pyroelectric infrared sensors through hidden markov model", *IEEE Sensors Journal*, 16(4), Feb. 2016.
- [2] F. Wahl, M. Milenkovic, and O. Amft, "A distributed PIR-based approach for estimating people count in office environments", *IEEE Conf. on Computational Science and Engineering*, 2012.
- [3] Y. Benezeth, H. Laurent, B. Emile, and C. Rosenberger, "Towards a sensor for detecting human presence and characterizing activity", *Energy and Buildings*, 43, 2011.
- [4] J. Han and B. Bhanu, "Fusion of color and infrared video for moving human detection", *Pattern Recognition*, 40, 2007.
- [5] S. Nakashima, Y. Kltazono, L. Zhang, and S. Serikawa, "Development of privacy-preserving sensor for person detection", *Procedia*, 2, 2010.
- [6] I. Amin, A. Taylor, F. Junejo, A. Al-Habaibeh, and R. Parkin, "Automated people-counting by using low-resolution infrared and visual cameras", *Measurement*, 41, 2008.
- [7] A. Bobick and J. Davis, "The recognition of human movement using temporal templates," *IEEE Trans. Pattern Recognition and Pattern Analysis*, Vol. 23, No. 3, Mar. 2001.
- [8] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models", *IEEE Trans. Pattern Recognition and Pattern Analysis*, Vol. 32, No. 9, Sep. 2010.
- [9] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," *In Proceedings Computer Vision and Pattern Recognition*, pp. 246 - 252, 1999.
- [10] J. Guo, Y. Liu, C. Hsia, M. Shih and C. Hsu, "Hierarchical method for foreground detection using code book model," *IEEE Trans. Circuit Sys. Video Tech.*, Vol. 21, No. 6, June 2011.
- [11] N. McFarlane and C. Schofield, "Segmentation and tracking of piglets in image," *Machine Vision and Applications*, 8:187-193, 1995.
- [12] T. Tian, R. Schmidt, H. Liu, A. Hampapur, and M. T. Sun, "Robust Detection of Abandoned and Removed objects in complex surveillance videos," *IEEE Trans. Sym. Man and Cybernetics - Part C: Applica-*

- tions and Reviews, Vol. 41, No. 5. pp. 565-577, Sep. 2011.
- [13] A. Sanin, C. Sanderson, and B. Lovell, "Shadow Detection: A Survey and Comparative Evaluation of Recent Methods", *Pattern Recognition*, Vol. 45, Issue 4, pp. 1684-1695, Apr. 2012.
- [14] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking", *IEEE Trans. Pattern Recognition and Pattern Analysis*, Vol. 25, No. 5, May 2003.
- [15] S. Tand, M. Andriluka, B. Schiele, "Detection and tracking of occluded people", *Int J. Computer Vision*, Nov. 2013.
- [16] H. Chandel and S. Vatta, "Occlusion Detection and Handling: Review", *Int. J. of Computer Applications*, Vol. 120, No. 10, June 2015.
- [17] J. Foley, *Computer Graphics: Principles Practice*, Addison-Wesley, 1994.

---

## 저 자 소 개

---



### 길 종 인

- 2010년 8월 : 강원대학교 컴퓨터정보통신공학과 학사
- 2012년 8월 : 강원대학교 컴퓨터정보통신공학과 석사
- 2012년 9월 ~ 현재 : 강원대학교 IT대학 컴퓨터정보통신공학과 박사과정
- 주관심분야 : 객체 트래킹, 딥러닝, 점유센서, 머신러닝



### 김 만 배

- 1983년 : 한양대학교 전자공학과 학사
- 1986년 : University of Washington, Seattle 전기공학과 공학석사
- 1992년 : University of Washington, Seattle 전기공학과 공학박사
- 1992년 ~ 1998년 : 삼성종합기술원 수석연구원
- 1998년 ~ 현재 : 강원대학교 컴퓨터정보통신공학과 교수
- 2016년 ~ 현재 : 강원대학교 정보통신연구소 소장
- ORCID : <http://orcid.org/0000-0002-4702-8276>
- 주관심분야 : 3D영상처리, 비전점유센서, 객체인식