

3차원 깊이 정보 기반의 감시카메라 영상 분석

*이수빈 **서용덕

서강대학교 영상대학원 미디어공학과

*subin@sogang.ac.kr

Image Analysis for Surveillance Camera Based on 3D Depth Map

*Lee, Subin **Seo, Yongduek

Department of Media Technology, Sogang University

요약

본 논문은 3차원 깊이 정보를 이용하여 감시카메라에서 움직이는 사람을 검출하고 추적하는 방법을 제안한다. 제안하는 방법은 GMM(Gaussian mixture model)을 이용하여 배경과 움직이는 사람을 분리한 후, 분리된 영역을 CCL(connected-component labeling)을 통하여 각각 블랍(blob) 단위로 나누고 그 블랍을 추적한다. 그 중 블랍 단위로 나누는 데 있어 두 블랍이 합쳐진 경우, 3차원 깊이 정보를 이용하여 두 블랍을 분리하는 방법을 제안한다. 실험을 통하여 제안하는 방법의 결과를 보인다.

1. 서론

감시카메라는 건물의 내부와 외부, 도로와 같은 공공시설, 지하철, 엘리베이터 등 사람이 다니는 대부분의 곳에 설치되어 있다. 우리는 감시카메라를 기반으로 이러한 장소들에 보안 시스템을 구축하고 있다. 초기 감시카메라를 기반으로 하는 보안 시스템은 단순히 영상을 확인하는 시스템이었다. 하지만 사람이 일일이 영상을 확인하여 직접 감시하는 데에는 많은 제약이 따른다. 이에 영상 분석을 통하여 움직이는 사람 또는 물체를 자동으로 검출하고 추적하며, 행동의 움직임을 분석하는 보안 시스템으로 진화해가고 있다. 현재 대부분의 영상 분석은 2차원 영상을 기반으로 이루어지고 있다[1]. 하지만 2차원 정보만으로는 사람들이 겹쳐져 하나의 사람으로 인식하는 것 등 해결할 수 없는 문제들이 종종 발생한다.

본 논문에서는 스테레오 카메라를 이용하여 3차원 깊이 정보를 구하고, 그 정보를 이용하여 움직이는 사람을 검출하고 추적하는 감시카메라 보안 시스템을 제안한다. 제안하는 방법은 크게 영상에서 배경과 전경을 분리하는 것과 연속된 영상에서 움직이는 영역을 추적하는 것으로 나눌 수 있다.

배경과 전경(움직이는 사람 또는 물체)을 분리하는 방법에는 여러 가지가 있다. 그중 가장 기본적인 방법은 이전 프레임과 현재 프레임의 차이를 계산하여 정해진 범위에 속하는 영역을 배경 혹은 움직이는 물체로 인지하는 것이다. 하지만 이 방법은 범위에 매우 민감하다. 두 번째는 n 장의 프레임의 평균으로 배경을 구하는 것이고, 세 번째는 Mixture of Gaussians 방법으로 K 개의 배경 분포 모델을 가지고 배경을 분리하는 것이다. 이 방법의 경우 모델의 수와 gaussian 초기화 방법, 업데이트 시간에 따라 결과가 달라진다. 네 번째는 Kernel Density Estimators 방법으로 n 개의 가장 최근 픽셀 값의 히스토그램으로 주어진 배경 확률 밀도 함수 값의 범위에 따라 배경인지 아닌지 결정하는

방법이다. 그 외에도 Mean-shift based estimation, Sequential Kernel Density Approximation, Eigenbackgrounds 방법 등이 있다[2].

추적은 연속된 영상에서 움직이는 물체들의 위치를 찾는 것으로 크게 타겟의 묘사(target representation)와 위치추정(localization), 필터링(filtering)과 데이터 연관(data association) 두 부분으로 나눌 수 있다. 먼저 타겟의 묘사와 위치추정은 물체 내부의 세그멘테이션(segmentation)을 통한 블랍 추적(blob tracking), 측정값의 유사도 최대화를 기반으로 위치추정을 반복하는 커널 기반 추적(kernel-based tracking), 물체의 경계를 검출하는 윤곽 추적(contour tracking) 등의 방법이 있다[3-5]. 필터링과 데이터 연관은 칼만 필터(Kalman filter), 파티클 필터(particle filter)의 방법이 있다[6,7].

2장에서는 배경과 전경을 분리하는 방법, 분리된 영역을 블랍 단위로 인식하고 3차원 정보를 이용하여 나누는 방법, 블랍을 추적하는 방법에 대해 설명한다. 3장에서는 실험 결과를 보인다.

2. 제안하는 방법

본 논문은 3차원 깊이 정보를 이용하여 감시카메라에서 움직이는 사람을 검출하고 추적하는 방법을 제안한다. 먼저 스테레오 영상이 입력으로 들어오면 먼저 3차원 깊이 정보(disparity value)를 계산한다. 스테레오 영상 중 왼쪽 영상을 이용하여 배경과 전경을 분리하고 모폴로지 연산을 통해 노이즈를 제거한다. 전경으로 분리된 영역들을 블랍 단위로 나누고, 작은 블랍은 삭제하고 인접한 블랍들은 합친다. 이후 3차원 깊이 정보를 이용하여 분리가 가능한 블랍은 두 개의 블랍으로 나눈다. 마지막으로 현재 검출된 블랍들과 이전 프레임에서 검출된 블랍들의 위치 비교를 통해 추적을 한다. 그림 1은 제안하는 방법의 순서도이다.

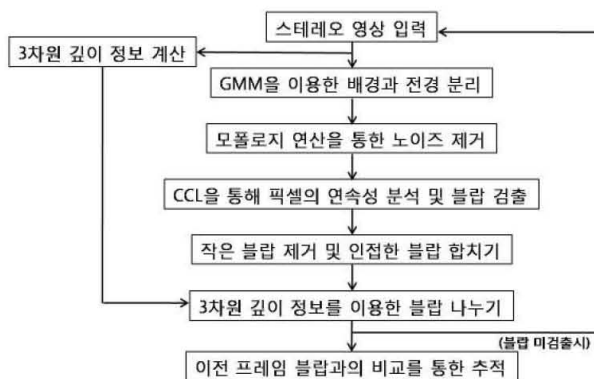


그림 1. 제안하는 방법의 순서도

영상으로부터 움직임을 찾는 것은 영상 분석의 첫 번째 과정이다. 배경과 전경을 분리하는 방법은 여러 가지가 있지만 본 논문에서는 Zoran[8, 9]이 제안한 GMM(Gaussian mixture model) 방법을 이용한다. GMM은 픽셀기반 배경 분리 방법으로 현재 픽셀에 대해 M개의 가우시안 분포를 구하고 그 값에 따라 배경인지 전경인지를 구별하는 것으로 식 (1)은 본 논문에서 사용한 M개의 구성요소(component)를 가진 GMM이다.

$$\hat{p}(\bar{x}|\mathcal{X}_T, BG+FG) = \sum_{m=1}^M \hat{\pi}_m \mathcal{N}(\bar{x}; \hat{\mu}_m, \hat{\sigma}_m^2 I) \quad (1)$$

$\hat{\mu}_1, \dots, \hat{\mu}_M$ 은 가우시안 구성요소(component)에서 평균의 추정치이고, $\hat{\sigma}_1, \dots, \hat{\sigma}_M$ 은 분산의 추정치이다. $\hat{\pi}_m$ 은 혼합 가중치(mixing weight)로 음수가 아니고, 합은 1이다.

그림 2는 GMM을 이용한 배경 분리 결과이다. 그림 2를 보면 노이즈처럼 보이는 수많은 흰색 점들과, 흰색 영역 내부에 빈 영역이 있는 것도 볼 수 있다. 노이즈를 제거하고, 내부가 비었거나 분리되어 있는 부분들을 연결시키기 위해 모폴로지 연산을 수행한다. 그림 3은 그림 2에 대해 모폴로지 연산을 수행한 결과이다. 그림 2에서 보이는 흰색 점들은 없어지고, 내부의 비어있던 영역은 채워진 것을 볼 수 있다.

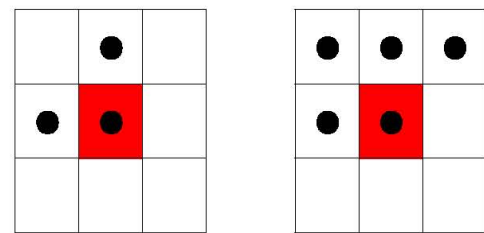


그림 2. GMM을 이용한 배경 분리 결과



그림 3. 그림 2의 모폴로지 연산 수행 결과

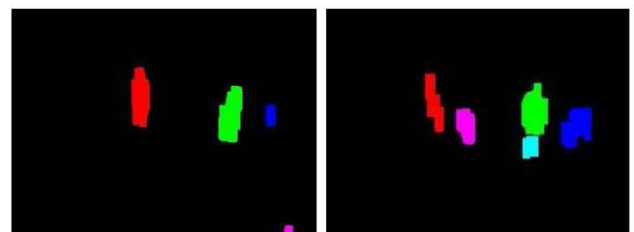
배경과 움직이는 영역(전경)을 분리하면 움직이는 영역을 각각의 블랍(blob)으로 검출하는 것이 두 번째 과정이다. 본 논문에서는 CCL(connected component labeling)을 이용하여 각 영역을 블랍 단위로 검출한다. CCL은 이진 영상에서 인접한 픽셀들은 연결하고, 떨어진 픽셀들은 나누어 각각의 영역을 하나의 블랍으로 검출하는 것이다[10]. CCL은 2단계 알고리즘(two pass algorithm)으로 나뉜다. 첫 번째 단계는 픽셀 값이 같음(equivalences)을 기록하고 임시 레이블(label)을 할당하는 것이고, 두 번째 단계는 동일한 클래스(equivalence class)의 레이블로 임시 레이블을 교체하는 것이다. 두 번째 단계에서 동일한 클래스로 연결하는 방법은 현재 픽셀의 북쪽과 서쪽(위쪽과 왼쪽) 픽셀과의 연결성을 보는 4방향 연결(4-connectivity) 방법과 현재 픽셀의 북쪽, 서쪽, 북서쪽, 북동쪽 픽셀과의 연결성을 보는 8방향 연결(8-connectivity) 방법이 있다. 그림 4는 4방향 연결과 8방향 연결 분석에 필요한 픽셀들의 위치를 나타낸 것으로 붉은색 픽셀이 현재 픽셀이고 검은 점은 점으로 표시한 것이 연결 성분 분석에 필요한 픽셀의 위치이다. 본 논문에서는 8방향 연결 성분을 가지고 레이블링(labeling)을 하였다.



(a) 4방향 연결 (b) 8방향 연결

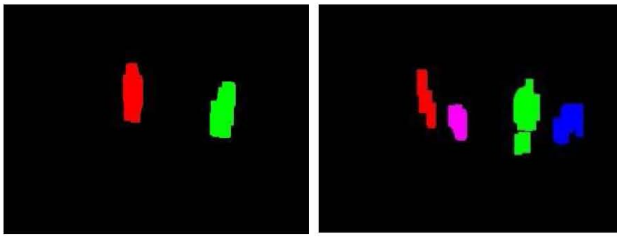
그림 4. 연결 분석에 필요한 픽셀들의 위치

그림 5는 레이블링 결과이다. 동일한 레이블을 가지는 블랍을 동일한 색상으로 표시하였다. 그림 5 (a)는 4개, (b)는 5개의 레이블로 나뉘는 것을 볼 수 있다. 이 때, 색상이 다른 레이블들은 각각 다른 블랍이다. 레이블링된 블랍들 중 크기가 작은 블랍들은 제거하고, 블랍들 사이의 거리와 disparity value가 일정 거리 이내에 있고, 그 값이 유사하다면 인접한 블랍이라 하고 두 블랍을 연결하는 과정을 수행한다. 그림 5 (a)에서 파란색, 자홍색 블랍의 경우 움직이는 사람으로 보기엔 작은 크기의 블랍임으로 제거한다. 또한 그림 5 (c)의 녹색과 청록색 블랍의 경우 일정 거리 내에 있으면서 disparity value도 유사하기 때문에 두 블랍은 하나의 블랍으로 생각한다. 그림 6 (a)는 블랍을 제거한 후의 결과이고, 그림 6 (b)는 블랍을 합친 후의 결과이다.



(a) 4개의 레이블 (b) 5개의 레이블

그림 5. 레이블링의 결과



(a) 작은 블랍을 제거한 결과 (b) 유사한 블랍을 합친 후의 결과
그림 6. 블랍 제거 및 합친 후의 결과



그림 7. 여러 사람이 하나의 블랍으로 검출되는 경우

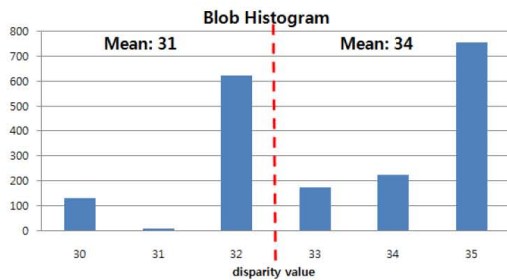


그림 8. 그림 7에서 검출된 블랍의 disparity value 히스토그램

그림 7은 여러 사람이 하나의 블랍으로 검출되는 경우를 보여준다. 이때에는 disparity value를 이용하여 분리가 가능한 블랍은 두 개의 블랍으로 나눈다. 그림 8은 그림 7에서 검출된 블랍의 disparity value의 히스토그램을 보여준다. 빈도수가 80개 이상인 disparity value에 대해 최소-최대 disparity value의 차가 6 이상이면 그 블랍은 disparity value의 중간 값을 기준으로 두 개의 블랍으로 나눈다. 그림 9는 그림 7에서 검출한 블랍을 두 개의 블랍으로 분리한 결과이다.

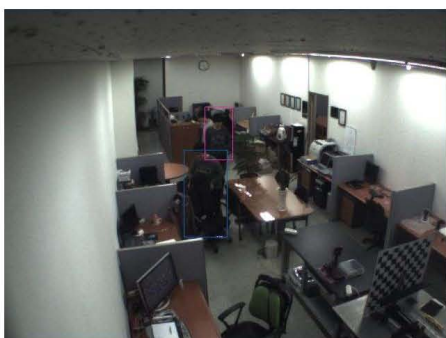
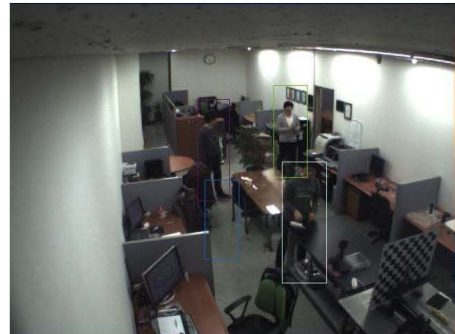


그림 9. 두 개의 블랍으로 분리된 결과

그림 10은 3차원 깊이 정보를 이용하여 블랍을 분리하는 다른 결과이다. 그림 10 (a)의 두 개의 블랍이 (b)에서 네 개의 블랍으로 분리된 것을 볼 수 있다.



(a) 여러 사람이 하나의 블랍으로 검출되는 경우



(b) 블랍 분리 결과

그림 10. 두 개의 블랍이 네 개의 블랍으로 분리된 결과

마지막으로 블랍이 검출되면 그 블랍들을 이전에 검출된 블랍들의 정보와 비교하여 추적한다. 먼저 이전에 검출된 블랍들과 현재 검출된 블랍들 사이의 거리를 모두 계산한다. 만약 이전에 검출된 블랍이 3개이고, 현재 검출된 블랍이 2개일 경우 모두 6가지의 블랍들 사이의 거리를 계산하게 된다. 이렇게 계산된 거리에 따라 블랍의 쌍(이전 프레임에서 검출된 블랍과 현재 프레임에서 검출된 블랍)을 작은 순서대로 정렬하고, 계산된 거리가 일정 거리이하($70/t$; t 는 실제 계산에 사용한 영상과 원본 영상과의 비율)인 경우에 추적중인 블랍이 있다고 하고, 이전에 검출된 블랍의 정보(색상 ID)를 현재 블랍의 정보로 유지하게 한다. 만약 현재 블랍 한 개가 이전 프레임의 두 개의 블랍과 일정 거리 이하의 값을 갖는다면, 두 블랍 중 더 짧은 거리에 있는 블랍으로 추적을 유지하게 한다. 반대의 경우로 이전 블랍 한 개에 대해 현재 블랍 두 개가 일정 거리이하에 존재한다고 하면 두 블랍중 더 짧은 거리에 있는 블랍을 선택한다. 또한 이전의 모든 블랍과 계산된 거리가 일정 거리 이상이라면 이전 프레임의 블랍과 유사한 블랍이 없다고 판단하여 새로운 정보(색상 ID)를 부여하고 새롭게 추적을 시작한다.

3. 실험 결과

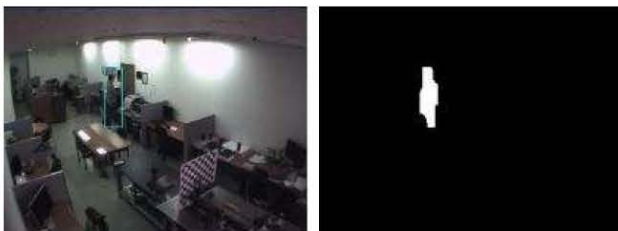
이번 장에서는 2장에서 설명한 제안하는 방법으로 실험한 결과를 보인다. 그림 11은 한 사람이 움직일 경우에 추적된 결과와 배경 분리의 결과이고, 그림 12은 세 사람이 움직일 경우에 추적된 결과와 배경 분리의 결과이다.



(a) frame 1300

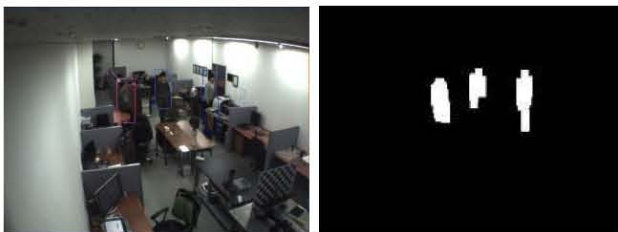


(b) frame 1320



(c) frame 1340

그림 11. 한 사람의 움직임에 대한 추적 및 배경 분리 결과



(a) frame 222



(b) frame 231



(c) frame 240

그림 12. 세 사람의 움직임에 대한 추적 및 배경 분리 결과

3. 결론

본 논문은 스테레오 카메라로부터 2D 영상 획득 및 공간의 3D 정보를 실시간으로 얻어 이를 바탕으로 공간 내에 존재하는 사람을 인지하고 추적하는 방법을 제안한다. 먼저 움직이는 사람과 배경을 분리하고, 분리된 영역에서 블랍의 크기 정보와 3차원 깊이 정보를 이용하여 실제 움직임으로 나타나는 블랍들을 검출하고 그 블랍들이 이전 프레임의 블랍들과 비교하여 추적한다. 제안하는 방법은 3차원 깊이 정보를 이용함으로써 2차원 영상 정보만을 이용하여 사람을 인지하고 추적하는 것보다 정확한 추적이 가능하다.

감사의 글

본 논문은 2010년 중소기업청 지원 기술혁신개발과제 "지능형 3D Active 감시 카메라 개발" 과제 지원으로 수행되었음.

4. 참고문헌

- [1] M. Valera and S.A. Velastine, "Intelligent distributed surveillance systems: a review," *IEE Proceedings Vision, Image and Signal Processing*, vol. 152, no. 2, pp. 192-204, 2005.
- [1] M. Piccardi, "Background subtraction techniques: a review," in *Proc. IEEE International Conference on SMC*, pp. 3099 - 310, 2004.
- [3] M. Isard and J. MacCormick, "BramBLE: A Bayesian Multiple-Blob Tracker," *IEEE International Conference on Computer Vision*, vol. 2, pp. 34-41, 2001.
- [4] Changjiang Yang, Ramani Duraiswami and Larry Davis, "Efficient Mean-Shift Tracking via a New Similarity Measure," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.176-183, 2005.
- [5] Yogesh Rathi, Namrata Vaswani, Allen Tannenbaum and Anthony Yezzi, "Particle Filtering for Geometric Active Contours with Application to Tracking Moving and Deforming Objects," *IEEE International Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 2-9, 2005.
- [6] R.E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35 - 45, 1960.
- [7] M. Sanjeev Arulampalam, Simon Maskell and Neil Gordon and Tim Clapp, "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174-188, 2002.
- [8] Zoran Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," *International Conference Pattern Recognition*, vol. 2, pp.28-31, 2004.
- [9] Zoran Zivkovic and Ferdinand van der Heijden, "Efficient Adaptive Density Estimation per Image Pixel for the Task of Background Subtraction," *Pattern Recognition Letters*, vol. 27, no. 7, pp.773-780, 2004.
- [10] http://en.wikipedia.org/wiki/Connected-component_labeling