

# Mean Shift와 변위지도를 결합한 카메라 이동환경에서의 다수 인체 추적

홍수연<sup>0</sup>, 변혜란

연세대학교 컴퓨터학과

{syhong<sup>0</sup>, hrbyun}@cs.yonsei.ac.kr

Multiple Human Tracking using Mean Shift and Disparity map with an Active Camera

SooYoun Hong<sup>0</sup>, Hyeran Byun

Department of Computer Science, Yonsei University

## 요 약

본 논문은 스테레오 카메라를 이용한 이동 카메라 환경에서 다수의 사람을 검출하여 검출된 사람을 추적하는 방법을 제안한다. 카메라가 이동하게 되면 카메라의 움직임과 검출 대상이 되는 사람의 움직임이 동시에 발생하기 때문에 카메라 움직임을 변환 모델을 사용하여 보정하고, 독립적인 움직임을 추출하여 사람을 검출하였다. 추적은 검출된 사람 영역의 컬러 분포에 기반하여 평균 이동(Mean Shift) 알고리즘을 적용하였다. 평균 이동 알고리즘은 빠르고 안정적인 성능으로 실시간 추적에 적합하다. 그러나 객체의 컬러 정보만으로는 배경과 컬러 분포가 유사한 객체의 경우 추적에 실패할 수 있는 단점이 있다. 이점을 보완하기 위하여 본 논문에서는 변위 지도(Disparity map)를 결합하여 객체와 배경을 분리하는 깊이 마스크를 생성하였다. 변위 지도를 사용하여 다수의 사람이 등장 할 경우 발생하는 가려짐, 경침 등 다양한 실내 환경에서 발생하는 문제도 해결하였다. 본 논문에서 제안하는 알고리즘은 다양한 데이터에 대해서 실험한 결과 정확한 검출과 추적에 우수한 성능을 확인할 수 있었다.

Keywords: 이동 카메라 환경, 사람 검출, 사람 추적, 평균 이동, 변위 지도

## 1. 서 론

사람을 검출하고 추적하는 기술은 이동로봇, 감시 시스템, 인간과 로봇간의 상호작용 등 많은 응용시스템에서 연구되고 있는 분야 중 하나이다. 현재의 비전 기술을 이용하여 실시간으로 객체를 검출하고, 추적하는 일은 어려운 일임에도 불구하고, 컴퓨터 성능의 발달로 인해 영상처리 기법의 발전과 더불어 이에 관한 연구가 활발히 진행되고 있다. 본 논문에서는 이동로봇 환경에서 다수의 사람을 검출, 추적하는 방법을 제안한다. 인간과 로봇이 상호작용을 하기 위해서 해결되어야 할 중요한 부분 중 하나는 사람의 위치와 상태를 인식하는 것이다. 이 분야의 궁극적인 목적은 주변 환경 변화에 영향을 받지 않고 사람의 위치를 파악, 추적하는 기술을 개발하는 것이다.

사람을 검출하고 추적하는 방법은 크게 2가지 환경으로 구분할 수 있다. 고정된 카메라 환경에서 움직이는 객체를 검출하는 방법과 이동 카메라 환경에서 움직이는 객체를 검출, 추적하는 것이다. 전자의 경우에는 대부분의 비디오 감시시스템과 같이 고정된 배경으로부터 전경을 분리하여 움직이는 객체를 검출하고 검출된 객체가 움직이는 사람인지 비강체(non-rigid) 대상인지를 판단하는 방법이 많이 사용되고 있다. W4시스템에서는 [1] 각 화소(pixel)의 최소 밝기값과 최대 밝기값, 표준편차를 이용하여 저장되어 있는 고정된 배경을 몇 프레임 동안 학습시켜 화소 단위로 배경을 모델링하여 객체를 검출하여 실루엣을 추출하고 추적하는 방법을 제안하였다. 그러나 위의 방법은 고정된 배경을 획득할 수

없는 이동로봇 환경에는 적합하지 않다.

이동 카메라 환경에서 움직이는 객체를 검출하고 추적하는 방법으로는 Yilmaz 등은 [2] 인위적으로 주어진 초기 칸투어(contour) 정보를 이용하여 객체의 칸투어를 추적하는 level set 방법을 제안하였다. 그러나 제안된 방법은 계산량이 많아 실시간 적용에 어려운 문제가 있다. Davis 등은 [3] 이동카메라 환경에서 컨덴세이션(CONDENSATION: Conditional Density Propagation) 알고리즘을 이용하여 보행자를 추적하였다. 그러나 제안된 방법은 보행자라는 한정된 환경만을 고려하였고 컨덴세이션 알고리즘의 사용으로 실시간 추적에 어려움이 있다.

본 논문에서는 위와 같은 고정 카메라 환경의 문제점과 실시간 응용등의 문제점을 해결하기 위하여 이동카메라 환경에서 다수의 사람을 검출, 추적하는 방법을 제안한다.

## 2. 사람 검출

로봇의 시각 인터페이스를 이용한 자동화된 객체 검출 시스템은 카메라가 이동한다는 특징을 가지고 있다. 카메라가 이동하게 되면 카메라의 움직임과 검출 대상이 되는 사람의 움직임이 동시에 발생하기 때문에 문제가 발생한다. 이러한 문제점을 해결하기 위하여 본 논문에서는 카메라의 움직임을 보정하는 방법을 적용한다. 카메라의 움직임을 보정하게 되면 사람의 움직임 정보만을 프레임의 차이로써 알 수 있다. 이렇게 추출된 사람의 대략적인 위치정보와 변위 지도를 결합하여 사람을 검출하는 알고리즘을 개발한다.

2.1. 카메라 움직임 보정

카메라 움직임을 보정하기 위하여 먼저 연속적인 두영상에서의 대응점을 찾아야 한다. 본 논문에서는 특징이 되는 점들을 추출하기 위하여 KLT (Kanade-Lucas-Tomasi) 특징 추적 알고리즘[4]을 사용하였다. 연속적인 두 영상이 주어졌을 때 이전 영상에서 특징점을 추출하고, KLT 알고리즘을 사용하여 다음 영상에서 이전영상에서 추출된 특징점과 대응되는 대응점을 찾는다.

연속적인 두 영상의 대응점이 결정되면, 카메라의 움직임을 찾기 위하여 변환 모델(transformation model)을 사용한다. 본 논문에서는 카메라의 상, 하, 좌, 우의 움직임과 로봇의 전진, 후진 움직임까지 보정할 수 있는 바이리니어 모델(bilinear model)을 사용하였다.

2.2. 사람 영역 추출

카메라 움직임이 보정되면 보정된 영상과 현재 영상을 비교하여 최종적으로 사람 영역을 추출한다. 보정된 영상과 현재 영상과의 차 영상을 구하면 사람의 움직임을 정확하게 추출 할 수 있다. 본 논문에서는 변위 지도를 추가로 결합하여 최종적으로 사람 영역을 검출하였다.

3. 사람 추적

추적 알고리즘은 객체의 컬러분포에 기반한 방법을 적용하였다. 먼저 추적하고자 하는 대상의 컬러 분포를 히스토그램으로 추정하고 이로부터 객체의 컬러분포 중 가장 빈번한 컬러 정보를 얻어내었다. 이 정보를 다음 프레임에서 객체의 후보 영역(Candidate target)을 찾는 특징 정보로 사용하였다. 평균 이동 알고리즘[5]을 사용하여 확률 분포상에서 객체와 가장 유사한 후보 영역(Nearest mode)을 찾아내었다. 하지만 이러한 정보만으로는 다수의 사람을 추적할 때 빈번히 발생하는 가려짐이나 겹침 등의 문제를 해결할 수 없고 배경과 컬러 분포가 유사한 객체의 경우 추적에 실패할 수 있는 단점이 있다. 이러한 단점을 해결하기 위하여 본 논문에서는 스테레오 정보인 변위 지도를 결합하여 객체와 배경을 분리하는 깊이 마스크를 생성하였다. 그림 1은 전체 추적 흐름도이다.

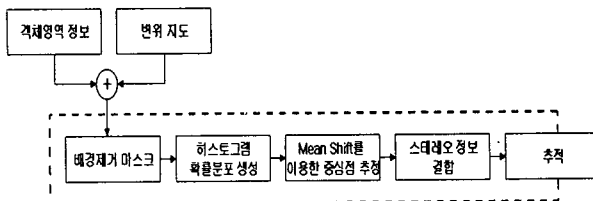


그림 1. 추적 흐름도

3.1. 사람의 컬러 분포 추출

히스토그램 기반 추적은 대부분의 추적 시스템에서 널리 사용되는 방법이다. 또한 효율적이며 빠른 실행 속도로 실시간 추적에 적합하다. 본 논문에서는 조명에 대한 영향을 줄이기 위해 HSV 컬러 모델을 사용하였다. 사람의 추적을 위해서는 사람에 대한 정

확한 컬러분포가 필요하다. 본 논문에서는 배경과 분리된 사람 영역만의 컬러 분포를 추출하기 위하여 2차원 HS 히스토그램과 변위 지도를 결합하여 배경제거 마스크를 생성 하였다. 그림 2 에 배경제거 마스크 생성과정을 나타내었다.



그림 2. 배경제거 마스크 생성 과정

3.2. 히스토그램 역투영(Histogram back projection)

사람 영역의 2차원 HS 히스토그램 분포를 사용하면 사람 영역에서 가장 빈번히 나타나는 컬러 정보를 찾을 수 있다. 이 값을 이용해 영상에 히스토그램 정보를 역투영시키면 사람 영역만을 정확하게 추적할 수 있다[6]. 이때, 영상에 역투영된 이미지는 그림 3 에서와 같이 사람에 대한 정보 중 분포가 가장 빈번한 화소에 대한 확률 분포가 된다.

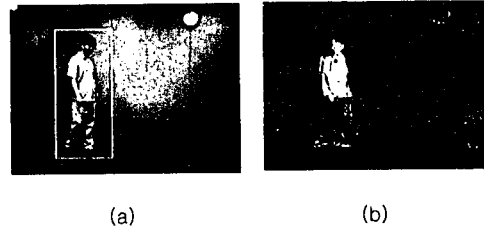


그림 3. (a) 원영상(b) 사람 정보 중 분포가 빈번한 화소에 대한 확률 분포 영상

3.3. 평균 이동(Mean Shift) 알고리즘

앞에서 얻어진 역투영 영상을 이용하면 사람 영역이 가지는 가장 큰 확률 분포 값의 중심을 찾을 수 있다. 본 논문에서는 특정 분포의 모드(mode)를 찾기 위해 확률분포의 기울기(gradient) 값을 반복적으로 찾아가는 평균 이동 알고리즘을[5] 사용하였다. 각 단계마다 반복적인 평균 이동 알고리즘의 적용으로 사람 영역의 평균  $y_0$ 에서 새로운 평균  $y_1$ 으로의 이동을 수식(1)과 같이 정의하였다. 기울기 벡터는  $y_1 - y_0$ 가 되고  $|y_1 - y_0|$ 이  $\epsilon$ 보다 작을 때까지 반복하여 사람 영역을 추적한다.

$$y_1 = \frac{\sum_{i=1}^n x_i w_i g \left( \left\| \frac{\hat{y}_0 - x_i}{h} \right\|^2 \right)}{\sum_{i=1}^n w_i g \left( \left\| \frac{\hat{y}_0 - x_i}{h} \right\|^2 \right)} \quad (1)$$

3.4. 스테레오 정보 결합

스테레오 카메라를 사용하면 왼쪽 영상과 오른쪽 영상의 특징 점을 추출해서 영상의 대응(correspondence)을 찾으면 대응들 간의 변위를 계산하여 변위로부터 3차원 위치 정보를 획득할 수 있다. 이렇게 얻어진 3차원 위치 정보를 결합하여 가려짐 문제를 해결하였다. 그림 4 에서와 같이 비슷한 확률분포를 갖는 두 객체간

의 가려짐이 발생하더라도 위치정보의 결합으로 가려짐 문제를 해결할 수 있다.

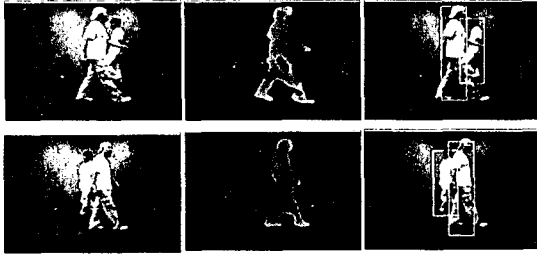


그림 4. (a) 원영상 (b) 변위 지도 (c) 추적 영상

4. 실험 및 결과

본 논문에서 제안한 방법을 Windows XP 에서 Visual C++ 6.0 을 이용하여 Pentium IV 2.4GHz 하드웨어 상에서 구현하였다. 실험데이터는 Videre 사의 STH-MDCS2 스테레오 카메라를 이용하여 실내환경에서 촬영하였다.

4.1. 가려짐 및 겹침에 대한 추적 결과 비교

본 논문에서 제안한 방법은 깊이 정보의 결합으로 가려짐 및 겹침에 대해서 평균 이동 추적 알고리즘 보다 좋은 성능을 나타내었다. 평균 이동 추적 알고리즘은 Comaniciu가 제안한 방법을 사용하여 실험하였다[5]. 그림 5에서 가려짐이 발생한 경우의 추적 결과를 제안한 방법과 평균 이동 추적 방법을 비교하여 나타내었다.

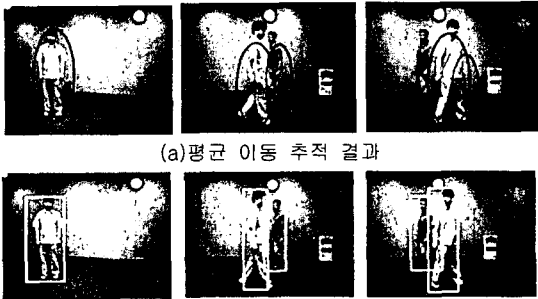


그림 5. 평균 이동 추적 결과와 제안한 추적 결과 비교

4.2. 다수에 대한 추적 결과

제안한 방법은 이동 카메라 환경에서 다양한 자세를 취하는 다수의 사람에 대한 추적에도 좋은 성능을 보였다. 이를 그림 6 에 나타내었다.

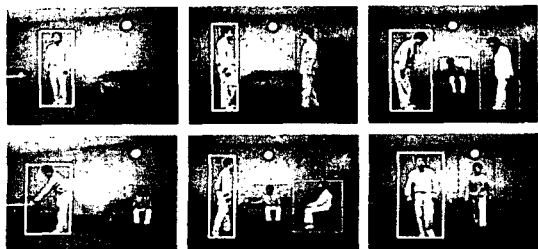


그림 6. 이동 카메라 환경에서 다수에 대한 추적 결과

4.3. 사람 추적 결과 평가

사람 추적 결과에 대한 평가는 바운딩 박스가 사람을 모두 포함하지 않는 경우, 바운딩 박스가 생성되지 않은 경우, 궤적이 올바르게 맞지 못한 경우, 같은 사람이 다른 사람으로 판단되는 경우 등 4 가지 조건을 오류로 판단하여 평가하였다. 제안한 방법의 우수성을 증명하기 위하여 추적에서 가장 보편적으로 사용되는 평균 이동 추적 기법과 Maggio와 Cavallaro가 제안한 평균 이동 알고리즘과 컨테이션 알고리즘을 결합한 하이브리드(Hybrid)추적 기법[7]을 비교하였다. 실험데이터는 5개의 테스트 데이터로 이루어져 있으며 모두 실내 환경에서 스테레오 카메라를 이용하여 이동카메라 환경에서 촬영한 데이터를 사용하였다. 실험데이터 A는 사람 한 명이 등장하여 여러 가지 자세를 취하는 데이터이다. 데이터 B는 두 명의 사람이 등장하고 서로 간의 가려짐이 발생하는 데이터이다. C와 D는 다수의 사람에 대한 추적 결과이다. 마지막으로 데이터 E는 다수의 사람이 다양한 자세를 취하며 부분적인 가려짐이 발생하는 데이터이다. 결과는 표1 과 같이 전체 95.09%의 추적률을 보였다. 사람 추적은 초당 최소 15fps에서 최대 20fps의 속도를 나타내었다.

표1. 사람 추적률

테스트 데이터	전체 프레임수	평균이동 추적기법	하이브리드 추적기법	제안한 방법
Test Set A	438	88.59%	93.16%	96.72%
Test Set B	383	85.64%	86.95%	96.21%
Test Set C	699	60.26%	85.70%	94.98%
Test Set D	699	81.84%	79.60%	95.05%
Test Set E	597	61.50%	86.05%	93.38%

참고 문헌

- [1] I.Haritaoglu, D.Harwood, and L.S.Davis, "W4: Real-Time Surveillance of People and Their Activities," *IEEE Transactions on PAMI*, Vol. 22, No. 8, pp. 809-830, 2000
- [2] A.Yilmaz, X.Li and M.Shah, "Object Contour Tracking Using Level Sets," *Asian Conference on Computer Vision*, 2004
- [3] L.Davis, V.Philomin and R.Duraiswami "Tracking humans from a moving platform," *15th Int. Conf. on Pattern Recognition*. Vol. 4, No. 4, pp. 171-78, 2000
- [4] C. Tomasi and T. Kanade, "Detection and tracking of point features," Carnegie Mellon University, Pittsburgh, PA, Tech. Rep. CMU-CS-91-132, 1991
- [5] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," *IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 2, pp. 1421-49, 2000
- [6] John G. Allen, Richard YD Xu, Jesse S. Jin "Object Tracking Using CamShift Algorithm and Multiple Quantized Feature Spaces," *Pan-Sydney Area Workshop on Visual Information Processing*, pp.3-7, 2003
- [7] E. Maggio and A. Cavallaro, "Hybrid particle filter and mean shift tracker with adaptive transition model", *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2005.