

영상 정렬 알고리즘을 이용한 팬틸트 카메라에서 움직이는 물체 탐지 기법

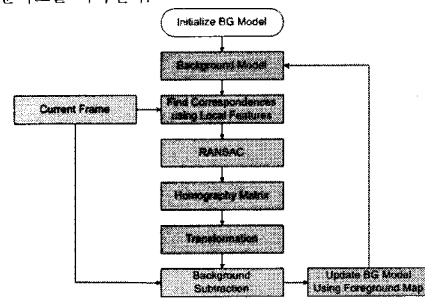
백영민, 최진영
 서울대학교 전기컴퓨터 공학부
 자동화시스템 공동연구소

Moving Object Detection in Pan-Tilt Camera using Image Alignment

Young Min Baek, Jin Young Choi
 School of Electrical Engineering Seoul National University
 Automation and System Research Institute (ASRI)

Abstract - 이동 물체 탐지(Object Detection) 기법은 대부분의 감시 시스템에서 가장 초기 단계로서, 이후에 물체 추적(Object Tracking) 및 물체 식별(Object Classification) 등의 기능 알고리즘에 입력으로 사용된다. 따라서 물체의 윤곽의 변화 없이 최대한 정교하게 이동 물체 영역 맵을 생성하는 것이 물체 탐지의 가장 중요한 요소가 된다. 카메라가 고정되어 있는 경우에는 현재 들어오는 영상에 대한 확률적 배경 모델을 생성할 수 있지만, 팬틸트 카메라와 같이 영상의 좌표가 변하는 환경에서는 배경 모델도 계속 변하기 때문에 기존의 배경 모델을 그대로 사용할 수 없다. 본 논문에서는 팬틸트 카메라와 같이 동적인 카메라에서 이동 물체 탐지를 위해, 국소 특징점(Local Feature)를 통해 카메라의 움직임을 판단하여 연속되는 영상간의 변환 행렬(Transformation Matrix)를 구하고 하고, 확률적 배경 모델링을 통한 이동 물체 탐지 기법을 제안한다. 자체 촬영한 이동 카메라 실험 영상을 통해서 이 알고리즘이 동적 배경에서도 매우 강인하게 동작하는 것을 검증하였다.

픽셀을 제외한 나머지 영역에 대해 다시 배경 학습을 통해 다음 영상에서 사용할 수 있도록 바꾼다. <그림 1>은 이동 카메라에서 물체 탐지를 위한 알고리즘 순서도를 나타낸다.



<그림 1> 알고리즘 순서도

1. 서 론

이동 물체 탐지 기법은 감시 시스템의 가장 초기 단계로서 이후 물체 추적이나 식별 단계를 수행하기 위해 매우 중요한 역할을 수행한다. 최근의 대부분의 고정 카메라에서 이동 물체 탐지(Moving Object Detection) 알고리즘은 고정되어 있는 배경을 가우시안 혼합 모델(Gaussian Mixture Model)[1]등을 이용해서 모델링하고, 입력 영상에서 배경 영역에 대한 소거법(Background Subtraction)이 주로 사용되었다. 하지만 팬틸트 카메라와 같이 배경 모델이 변하는 환경에서의 이동 물체 탐지는 매우 어려운 문제로서, 물체의 움직임과 더불어 카메라의 움직임도 동시에 추정해야 되는 문제를 갖고 있다. 따라서 팬틸트 카메라에서 이동 물체를 탐지하기 위해서 사전에 배경을 미리 팬과 틸트의 각도에 따라서 기억을 해두고, 카메라가 움직였을 때 해당 배경에서의 차를 구함으로써 물체 탐지 하는 방법이 제안[2]되었지만, 해당 알고리즘은 배경을 연속적으로 저장하는 것이 아니기 때문에 카메라가 이동 중에는 물체가 탐지 되지 않는 약점이 있다. 이러한 약점을 극복하기 위해서 모자이크(Mosaicing)기법을 이용해서 전체적인 배경을 생성하고, 전체 배경에서 카메라 현재 프레임의 위치를 추정하여 배경 영역을 제거해 주는 기법도 제안되었다. 하지만 커다란 배경을 갖기 위해서는 많은량의 저장공간이 필요하고 오차가 누적됨에 따라 시간이 흐를수록 오류를 발생할 확률이 높아진다.[3]

본 논문에서는 이러한 기존 알고리즘의 단점을 극복하기 위해서, 새로운 영상 정렬 기법을 이용한 이동 물체 탐지 기법을 제안한다. 먼저 영상의 변환 행렬을 구하기 위해서 SURF(Speeded-Up Robust Feature) 알고리즘[4]을 사용해서 연속되는 두 영상간의 대응점(Correspondences)을 구한다. 구해진 대응점에서 RANSAC을 이용해서 최적의 변환 행렬을 추정하고, 과거의 영상을 현재 영상과 동일한 좌표가 되도록 변환 해준다. 이러한 방법을 통해 두 영상의 겹치는 부분을 얻을 수 있는데, 겹치는 영역에 대해서만 특정한 규칙을 통해 배경을 학습하고, 겹치지 않는 새로 들어온 영역에 대해서는 모두 배경으로 가정하여 다음에 배경 학습에 이용하도록 한다.

알고리즘의 성능을 검증하기 위해서 자체적으로 촬영한 팬틸트 카메라의 영상을 이용하였다. 2장에서는 제안된 알고리즘의 순서도와 세부 내용에 대해 논하고, 3장에서 자체 실험한 영상을 이용해서 알고리즘의 성능을 보여 준다. 마지막으로 4장에서 결론을 내고, 추후 계획에 대해 논한다.

2. 본 론

2.1 팬틸트 카메라에서 물체 탐지 알고리즘

이동 카메라에서 물체 탐지하는 방법은 크게 두 부분: 1) 영상 정렬, 2) 배경 모델링으로 나누어진다. 영상 정렬 단계에서는 먼저 국소 특징점(Local Feature)를 이용해서 연속되는 영상간의 대응점(Correspondences)를 찾는다. 초기 대응점에는 많은 Outlier가 포함되어 있기 때문에, RANSAC을 이용해서 대응점에 포함되어 있는 Outlier를 제거하고, 변환 행렬 H 를 구한다. 변환 행렬은 모델링된 배경의 좌표계를 현재 영상의 좌표계로 변환하는 역할을 수행한다. 그 결과 현재 영상의 좌표계와 동일한 배경 모델을 얻을 수 있다.

동일한 좌표계의 배경 모델을 이용해서 배경 제거 기법(Background Subtraction)을 통해 전경맵(Foreground Map)을 생성한다. 전경으로 탐지된

2.2 영상 정렬(Image Alignment)

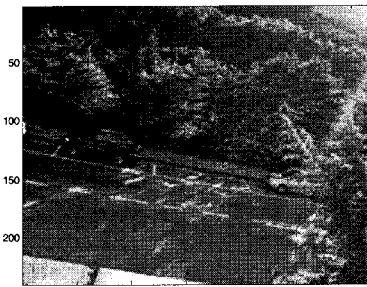
움직이는 카메라에서 물체가 실제 움직이는 영역을 검출하기 위해 중요한 점은 현재 카메라에서 들어오는 영상이 기존에 모델링한 배경 영상과 동일한 시점에 놓여져야 하는 것이다. 두 개의 영상을 동일한 좌표값을 갖도록 배치하는 기법을 영상정렬(Image Alignment)라고 한다. 영상 정렬을 통해서 좌표계가 다른 이전의 배경 모델을 이용해서 현재 영상에서 물체를 검출할 수 있다.

2.2.1 대응점(Correspondences) 검출

카메라 움직임 추정의 첫 단계로서, 변환 모델을 구하기 위해 연속되는 두 영상간의 대응점을 찾아야 한다. 일반적인 지점 추적(Point Tracking) 알고리즘의 경우, 칼라 모델을 이용한 Mean-Shift[5], Optical Flow를 기반으로 KLT[6]등이 많이 사용된다. 하지만 위의 알고리즘은 대응점 미스매칭으로 인한 변환 행렬의 드리프트(Drift)를 발생시킬 수 있으며, 연속되는 영상에서 한번 드리프트가 발생하면 이를 복구하기 위해 많은 시간이 흘러야 한다. 이러한 단점은 감시 시스템에 사용되는 물체 탐지 알고리즘에 적합하지 않다.

대응점 검출에 중요한 요소 중에 하나는 변화에 대한 강인성이다. 팬틸트 카메라는 2자유도를 갖지만 그에 따라 영상은 다양한 형태로 변화하기 때문에 이러한 변화에도 동일한 기술자(Descriptor)를 추출할 수 있어야 한다. 또한 초기의 대응점 후보에는 많은 Outlier가 포함될 수 있기 때문에, RANSAC 알고리즘을 이용해서 Inlier만을 추출할 때, 반복 회수를 줄이기 위해서 최대한 Inlier가 포함되는 개수를 높일 필요가 있다. 두 번째 중요한 요소는 연산 속도이다. 연산 속도는 실시간성이 요구되는 감시 시스템에서 매우 중요한 요소이므로 대응점 검출에서도 빠른 알고리즘이 요구된다.

따라서 본 연구에서는 대응점을 검출하는 알고리즘으로 Speeded-Up Robust Feature(SURF)[4]를 사용한다. SURF 알고리즘은 변화에 강인하면서도 비교적 빠른 연산 속도를 갖고 있다. 즉, 이전 영상에서의 특징점 x_i 와 현재 영상에 대응하는 점 x_i' 를 얻을 수 있다. <그림 2>은 SURF 알고리즘을 이용한 대응점 탐색 결과를 나타낸다.



<그림 2> SURF 알고리즘을 이용한 대응점 탐색

2.2.2 변환 행렬(Homography Matrix) 계산

영상의 자유도에 따라서 구하기 위한 최소한의 대응점의 개수가 달라진다. 가령 일반적으로 평면 영상 변환에 많이 사용되는 Perspective 변환 행렬을 구하기 위해서는 8개의 미지수 개수 보다 큰 9개의 제약 조건이 필요하며, 한 개의 대응점은 2개의 제약조건(Constraint)를 갖고 있기 때문에, $8/2=4$ 개의 대응점이 요구된다.[7]

일반적으로 영상에는 다수의 대응점이 존재하기 때문에, 여러 개의 제약 조건으로 변환 행렬을 구하기 위한 방법으로 Linear Direct Transformation(DLT)이 있다. DLT 알고리즘은 한 영상에서의 점 x_i 와 이에 대응하는 점 x'_i 를 알고 있는 경우 ($x_i \leftrightarrow x'_i$), 이에 대응하는 행렬 H 를 구할 수 있다. H 는 $x'_i = Hx_i$ 와같이 표현되며, 이 식은 $x'_i \times Hx_i = 0$ 를 만족해야 한다. 점 x_i 를 Homogeneous representation으로 표현하면,

$$Hx_i = \begin{pmatrix} h^1 T x_i \\ h^2 T x_i \\ h^3 T x_i \end{pmatrix} \text{ 와 } x'_i = (x'_i, y'_i, \omega'_i) \text{로 나타낼 수 있다. 위의 식을 통해 유}$$

도된 식을 h 에 대한 벡터형식으로 표현하면 식(1)과 같다.

$$\begin{bmatrix} 0^T & -\omega'_i x_i^T & y'_i x_i^T \\ \omega'_i x_i^T & 0^T & x'_i x_i^T \\ -y'_i x_i^T & x'_i x_i^T & 0^T \end{bmatrix} \begin{bmatrix} h^1 \\ h^2 \\ h^3 \end{bmatrix} = 0, \quad A_i h = 0 \quad (1)$$

여기서 마지막 행은 Dependent한 행(rank=2)이기 때문에 제거가능하며, 결국 한 대응점 별로 2x9의 Matrix가 생성된다. 위에서 언급 했듯이, Projective Matrix는 최소 4개의 대응점이 필요하기 때문에 다음과 같이 8x9의 Matrix를 얻을 수 있다.

$$\begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{bmatrix} h = 0 \quad \rightarrow \quad Ah = 0 \quad (2)$$

하지만 n 개의 대응점(여기서 $n > 4$)이 선택되었다고 했을 때, 식(3)와 같은 최적화 문제를 통해 최적의 h 를 구할 수 있다.

$$h^* = \min_h \|Ah\| \quad (3)$$

여기서 h 는 9-vector형식으로 나타내며, h 를 다음과 같이 3x3 행렬로 재배열함으로써 우리가 원하는 변환 행렬 H 를 얻을 수 있다

2.2 배경 제거 및 모델링

영상 정렬 기법을 통해 현재 영상과 배경 영상의 좌표계를 정합한 후, 배경 영역 제거를 통해(Background Subtraction), 전경맵(Foreground Map)을 생성한다. 전경맵이 생성되었으면, 나머지 배경 부분에 대해서 학습을 함으로서 다음 영상에 대한 배경을 모델링한다.

2.2.1 배경 제거(Background Subtraction)

각 픽셀은 하나의 가우시안 분포로 배경 모델링되며, 따라서 두 개의 Parameter(μ, σ)를 갖는다. 어떠한 픽셀이 시간 t 에서 명암도 x_t 를 갖을 확률은 식(4)와 같이 추정된다.

$$P_r(x_t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_t - \mu)^2}{2\sigma^2}} \quad (4)$$

해당 확률에 일정한 임계치(τ)를 두어서, 식(5)와 같이 확률이 τ 보다 작은 경우 전경으로 인식한다.

$$\Psi^t(x_t) = \begin{cases} 1 & P_r(x_t) \leq \tau \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

Ψ^t 는 시간 t 에서 전경맵을 나타낸다. 최종적으로 계산된 전경맵을 통해 움직이는 물체의 위치를 추출할 수 있다.

2.2.2 배경 모델링

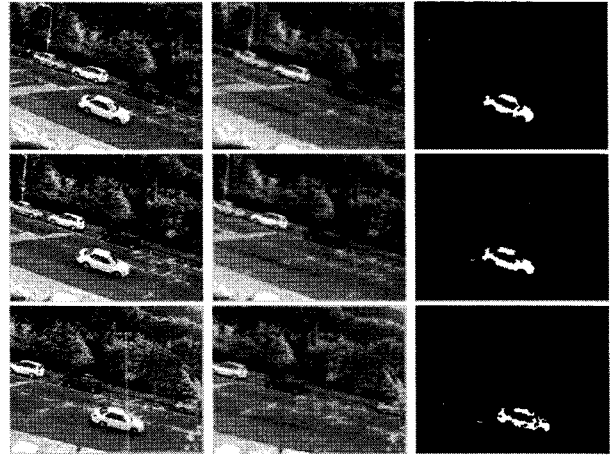
배경 제거 기법을 통해 전경맵을 얻었다면, 전경맵을 제외한 배경 부분에 대해서 모델을 새로 학습함을 통해 적응적인 배경 모델(Adaptive Background Model)을 생성할 수 있다. 배경 학습에서는 학습률 α 라는 Parameter를 둔다. 학습률은 얼마나 빨리 배경에 적응할 것인지에 대한 지표와 얼마나 오랫동안 배경을 저장할 것인지에 대한 Trade-off 관계에 있다. 실험을 통해 학습률을 구하고, 배경 영역에 대해 식(6), (7)과 같이 배경 학습을 수행한다.

$$\mu_t = (1 - \alpha)\mu_{t-1} + \alpha x_t \quad (6)$$

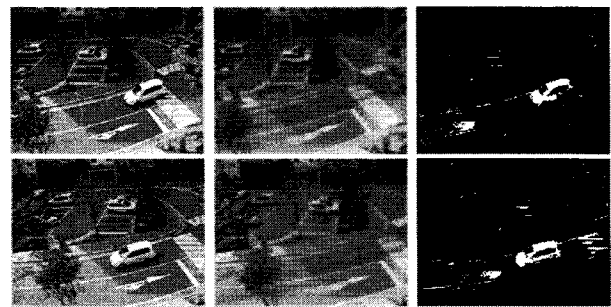
$$\sigma_t^2 = (1 - \alpha)\sigma_{t-1}^2 + \alpha(x_t - \mu_t)^2 \quad (7)$$

3. 실험결과

알고리즘의 실험을 위해서 서울대학교와 아파트 단지에서 촬영한 영상을 이용해서 실험하였다. 카메라가 지속적으로 움직이는 상황에서 물체 영역만을 탐지할 수 있는지를 확인하였다. <그림 3>은 영상의 시퀀스와 배경 모델, 전경맵을 보여준다. 왼쪽부터 현재영상, 배경모델, 물체 탐지 결과를 나타낸다.



(a) 서울대학교 실험결과



(b) 아파트 실험결과

<그림 3> 서울대학교와 아파트 동영상 실험 결과

위의 결과에서 모폴로지 연산(Morphological Operation)은 사용되지 않았다. 연산 속도는 MATLAB 구현으로 초당 3프레임 정도이다. 탐지 결과에서 발견되는 문제점으로, Perspective Model이 영상이 평면이라는 가정이 있어야 하지만, 카메라의 입력 영상은 3차원으로 맵핑한 것으로서 변환 행렬 하나로는 정확히 변환된 배경 모델을 구할 수 없다. 따라서 배경의 모서리 영역에서 약간의 노이즈가 발생하는 것을 확인할 수 있다. 하지만 전반적으로 물체영역이 잘 탐지되는 것을 확인할 수 있다.

4. 결 론

본 논문에서는 팬틸트 카메라에서 이동 물체를 탐지하는 알고리즘을 제안하였다. 국소 특징점을 이용해서 연속되는 영상의 변환 행렬을 구하고, 이 행렬을 이용해서 배경 모델을 현재 영상의 좌표계에 맞도록 변형할 수 있다. 현재 영상 좌표계에 맞도록 변형된 영상을 이용해서, 전경맵을 생성하고 배경 영역을 지속적으로 학습함으로써 배경 변화에 강인하도록 하였다. 실험 결과를 통해 알고리즘의 강인함을 알 수 있다. 몇 가지 발견된 문제점으로는 배경 모델의 개수가 하나로 정해져 있기 때문에 배경의 모서리 영역에서 전경으로 오인식 되는 경우가 종종 발생하며, 후백 배경영상의 한 계로 배경과 전경이 다른 색이라도 명암도가 비슷하면 배경으로 인식되는 문제가 발생한다. 추후 과제로 배경 영상이 있으면 픽셀 주변의 영역을 모두 고려해서 확률 모델을 생성하는 방법을 통해 이동 카메라에서 발생할 수 있는 모서리 효과를 제거하는 것이다.

[참 고 문 헌]

- [1] C. Stauffer and W.E.L. Grimson, "Adaptive background mixture models for real-time tracking" in Proc. of CVPR 1999, pp. 246-252.
- [2] F. Dellaert and R. T. Collins, "Fast image-based tracking by selective pixel integration" in Proc. of ICCV 1999.
- [3] N. I. Rao, H. Di and G. Y. Xu, "Panoramic Background Model under Free Moving Camera" in Proc. of FSKD 2006, vol. 1, pp. 639-643.
- [4] H. Bay, T. Tuytelaars and L. V. Gool, "SURF: Speeded-Up Robust Features," in Proc. of ECCV 2006.
- [5] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach towards Feature Space A Analysis" Trans. on PAMI, vol. 24, no. 5, pp. 603-619, May 2002.
- [6] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: a factorization method" Trans. on IJCV, vol. 9, no. 2, pp. 137-154, Nov 1992.
- [7] R. Hartley and A. Zisserman, Multiple View Geometry 2nd Edition. Cambridge, 2000.