

능동 카메라를 이용한 이동물체 검출

김용진^o 이일병
연세대학교 컴퓨터과학과
{yjkim98^o, yblee}@csai.yonsei.ac.kr

Moving Object Detection using Single Active Camera

Yongjin Kim^o, Yillbyung Lee
Dept. of Computer Science, Yonsei University

요 약

능동 카메라에서 배경과 물체가 모두 움직이는 영상에서 이동물체를 검출하여 추적하기 위해 특징점을 추출하고 특징점을 이용해 영상 좌표계 변환 파라미터를 추정하여 카메라의 Ego-motion을 보정한다. 보정된 영상을 이용하여 움직이는 물체를 검출하고 잡음이 있는 관측영역에서 CONDENSATION 알고리즘을 이용하여 이동물체를 추정하는 실험을 수행한 내용의 논문이다.

1. 서 론

컴퓨터 비전과 로봇 응용 시스템에서 카메라를 통해 획득된 영상 정보를 이용하여 물체를 추적하는 연구가 최근 다양하게 진행되고 있다. 물체추적에 관한 연구는 주요 산업, 의료, 군사 분야에 널리 활용되고 있고, 특히 최근에는 Robotics 및 위성관련 연구에서 중요한 부분을 차지하고 있다. 그러나 영상 장치(Camera)를 통해 들어온 영상(video)에서 물체를 검출 및 인지, 추적 및 식별하는 과정을 실제적으로 구현하기 위해서는 많은 어려움이 있다. 우선적으로 영상장치 자체의 잡음과 왜곡 등이 있을 수 있으며, 주변 환경의 변화에 의한 비디오 프레임 이미지의 광도 변화에 의한 잡음, 비고정식 카메라의 경우 카메라의 모션으로 인한 Ego-motion으로 프레임 시퀀스에 apparent motion이 발생하여 추적하고자하는 물체를 쉽게 검출할 수 없다.

배경이 고정된 경우, 카메라에서 획득된 영상의 흐름에서 시간 $t-1$ 과 t 에서의 영상차를 이용하면 쉽게 물체를 검출할 수 있다. 하지만 배경이 고정되지 않고 움직이는 상황에서 이동하는 물체를 검출하는 것은 배경이 고정된 경우보다 좀 더 복잡하다. 배경의 움직임이 적고 이동물체의 움직임이 많은 경우라고 가정하면, 직관적으로 영상의 흐름에서 시간 $t-1$ 과 t 에서 영상차를 구하고 배경의 움직임과 이동 물체의 움직임 양(量)의 적절한 임계값 이상의 값은 물체로 간주하면 검출이 가능하다.

그러나, 변환(translation), 줌(zooming), 수직 및 수평 이동(tilting, panning), 회전(rotation) 등 선형 및 비선형적인 변형에 취약하여 단순 영상차 비교 방법은 좋은 성능을 보장할 수 없다. 본 논문에서는 팬틸트 (PAN-TILT) 카메라를 이용하여 획득된 영상(video)에서 움직이는 물체를 검출 및 추적을 위해, 카메라에서 입력되는 영상 흐름(sequence)에서 Geometric 변환 과정을 통해 카메라의 영상을 보정하여 이동 물체를 검출하고[1], 배경이 보정된 영상에서 이동물체를 검출하고 잡음이 있는 검출 영역에서 베이스 필터(Bayes filter)를 기반의 방법론으로 Factored Sampling을 적용한 CONDENSATION 알고리즘[2]을 이용하여 검출 및 추적하는 시스템을 제안한다.

2. 본 론

물체추적 시스템에서 물체의 위치를 식별하기 위해 물체의 모델을 기반으로 한 인식 기반 추적 방법과 움직임 기반으로 하는 모션 기반 추적 방법이 있다. 물체의 모델을 기반으로 하여 추적하는 경우, 물체의 회전, 확대 및 축소 등의 변화를 인식하여 물체를 추적하여야 하므로 공간 및 시간의 복잡도가 증가하여 실시간 시스템에 적용하기 쉽지 않다[3][4]. 반면 움직임 기반 추적 방법은 영상내의 에너지(intensity)에 근거하여 추적하므로 추적하고자 하는 물체의 크기나 형태에 관계없이 상대적으로 덜 복잡한 시스템을 구현할 수 있으나, 잡음에 민감하여, 영상에 초래되는 전반적인 환경 변화에 적응하기 어렵다[5]. 확률적인 접근 방법인 베이스 필터

“본 연구는 산자부 뇌신경정보학 사업으로부터 지원을 받아 수행되었음.”

(Bayes filter)는 잡음이 있는 관측(Observation)으로부터 동적 시스템의 state를 추정할 수 있다[6]. 본 논문에서는 움직임 에너지에 기반으로 영상의 흐름에서 이미지 warping(Affine transform)을 이용하여 배경영상을 보정하고, 시차가 있는 영상의 차를 연산하고 차영상에서 입자 필터(Particle filter)를 이용하여 이동물체를 검출 및 추적한다.

2.1 베이즈 필터 (Bayes filter)

앞서 언급하였듯이 베이즈 필터는 잡음이 포함된 관측으로부터 동적 시스템의 state를 확률적으로 추정할 수 있다. 물체 추적을 위한 베이즈 필터에서 state는 추적하고자하는 물체의 위치를 나타낸다. 랜덤변수 x_t 는 시간 t 에서의 state를 나타내고, x_t 에 대한 확률분포는 $Bel(x_t)$ 로 나타내어진다. 시간 순으로 관측된 가용한 관측치를 $Z_t = \{z_1, z_2, \dots, z_t\}$ 라고 하면, $Bel(x_t)$ 는 아래와 같은 사후 밀도로 정의되어 진다.

$$Bel(x_t) = p(x_t | z_1, z_2, \dots, z_t) \quad ; \text{(식 1)}$$

식(1)을 간단히 표현하자면, 모든 가능한 위치 x 에 대하여 만약 센서에 의한 관측치가 z_1, z_2, \dots, z_t 와 같다면, 물체가 x 에 위치할 확률을 나타낸다. 시간의 흐름에 따라 관측치 Z_t 의 개수가 증가할수록 사후밀도 계산의 복잡도가 기하급수적으로 증가하여, 계산을 위해 Markov 가정을 필요로 한다. Markov 가정은 관측치는 추적 물체의 물리적인 현재 위치에 의해서 결정되어지고, 시간 t 에서의 물체의 위치는 이전의 state인 x_{t-1} 에 의해 결정되어 진다. 이전에 주어진 $p(x_{t-1} | Z_{t-1})$ 가 시간 $t-1$ 에서 존재한다고 가정하면, Z_{t-1} 의 관측치로 x_t 를 예측(prediction stage)할 수 있다. Chapman-Kolmogorov 공식에 의해서 시간 t 에서의 예측 밀도는 아래 (식 2)와 같이 나타낼 수 있다.

$$p(x_t | Z_{t-1}) = \int p(x_t | x_{t-1}) p(x_{t-1} | Z_{t-1}) dx_{t-1} \quad ; \text{(식 2)}$$

여기서 측정치 z_t 를 이용하여 update stage를 수행하면, 다음과 같은 식으로 나타내어진다.

$$\begin{aligned} p(x_t | Z_t) &= p(x_t | z_t, Z_{t-1}) \\ &= \frac{p(z_t | x_t, Z_{t-1}) p(x_t | Z_{t-1})}{p(z_t | Z_{t-1})} \\ &= \frac{p(z_t | x_t) p(x_t | Z_{t-1})}{p(z_t | Z_{t-1})} \\ &= k_t p(z_t | x_t) p(x_t | Z_{t-1}) \quad ; \text{(식 4)} \end{aligned}$$

$$\text{where, } p(z_t | Z_{t-1}) = \int p(z_t | x_t) p(x_t | Z_{t-1}) dx_t$$

여기서 $p(x_t | x_{t-1})$ 은 추적하는 물체의 동적특성을 나타내며, 위치를 추정하기 위한 조건부 확률로 motion model이다. 그리고 주어진 물체의 위치 x_t 로부터 관측치 z_t 의 조건부 확률은 $p(z_t | x_t)$ 로, 우도함수로 정의되고 perceptual model 이다.

2.2 입자 필터 (Particle filter)

Particle filter[5]는 사전분포(prior distribution)와 사후 분포(posterior distribution)와 같은 베이저안 조건 확률에 기반을 두고 있다. sample 혹은 particle 이라고 불리는 입자와 important factor라고 불리는 가중치(weight)를 이용한 sampling 방법으로 사후 밀도를 추론할 수 있는 SMC(Sequential Monte Carlo) 알고리즘이다. 가중치가 부여된 N 개의 입자는 $\{x_t^{(i)}, w_t^{(i)}\}_{i=1}^N$ 으로 표현되며, $x_t^{(i)}$ 는 시간 t 에서의 i 번째 입자의 state를 나타내고, $w_t^{(i)}$ 는 입자 $x_t^{(i)}$ 의 가중치를 나타낸다. 그리고 가중치의 합은 1이다($\sum_{i=1}^N w_t^{(i)} = 1$). 사후밀도는 아래의 (식 5)와 같이 몬테카를로 근사 적분에 의해서 나타내어진다.

$$\begin{aligned} p(x_t | Z_t) &\approx \sum_{i=1}^N w_{t-1}^{(i)} \delta(x_t - x_{t-1}^{(i)}) \\ &= k_t p(z_t | x_t) \sum_{i=1}^N w_{t-1}^{(i)} p(x_t | x_{t-1}^{(i)}); \text{(식 5)} \end{aligned}$$

동적모델 $p(x_t | x_{t-1})$ 을 이용하여 사전 밀도(prior density)를 계산하고, 주어진 새로운 관측으로부터 사후 밀도(posterior density)의 값을 구한다.

$$p(x_{t-1}|Z_{t-1}) \xrightarrow{\text{dynamics}} p(x_t|Z_{t-1}) \xrightarrow{\text{measuremen}} p(x_t|Z_t) ; \text{(식 6)}$$

일반적으로 사후 밀도를 너무 복잡하고, closed form에서 쉽게 계산되지 않는다. 그래서 반복적인 sampling 방법을 사용하는데, 확률 밀도의 근사값을 추정하기 위해 factored sampling method[2]를 사용한다.

사후 밀도 $p(x_{t-1}|z_{t-1})$ 은 입자군 $\{x_{t-1}^{(i)}, w_{t-1}^{(i)}\}$ 에 의해 재귀적으로 근사되며, 입자들은 $w_{t-1}^{(i)} = p(z_{t-1}|x_{t-1}^{(i)})$ 와 같은 가중치는 갖는다. 입자필터 알고리즘은 3단계로 구성되는데, 샘플링(sampling), 예측(predicting), 측정(measuring)이다. 알고리즘은 샘플링과 예측 단계에서 사전 밀도와 측정단계의 사후 밀도 사이의 조건 확률 전파(conditional probability propagation)를 이용하여 사전, 사후 조건 밀도를 재귀적인 방법으로 계산한다.

2.3 실험

능동 카메라를 이용하여 이동물체를 검출하고 추적하기 위해 그림 1과 같이 시스템을 구성하였다. KLT 알고리즘을 이용하여 영상 프레임에서 특징점을 추출하고 특징점 대응을 통해 Ego-motion 보정을 위한 영상의 좌표 변환 파라미터를 추출한다. 추출된 파라미터를 이용해 이미지 변환(projective/bilinear transformation) 후, 영상차를 이용하여 움직이는 물체를 검출 및 추적한다.

Moving Object Detection & tracking

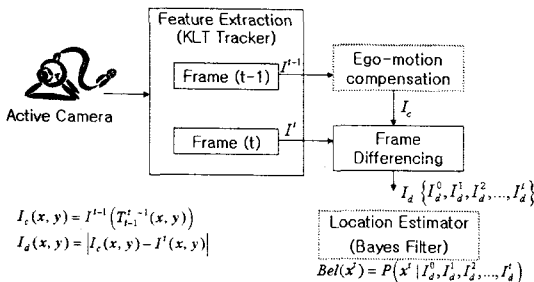


그림 1 이동물체 추적 시스템

2.3.1 특징점 추출

특징점 추출을 위해 사용된 Lucas-Kanade method는 KLT corner detector라고 알려져 있고, Kanade-Lucas-Tomasi(KLT) 추적 알고리즘[7]에서 사용되었다. 영상에

서 KLT corner detector에 의해 선택된 특징점은 엄밀히 말해 corner가 아니라 영상에서 작은 탐색 윈도우(small search window)를 이용하여 윈도우 내에서 두 직교 방향(x, y)에 대해 그래디언트(gradient)가 높은 값을 나타내는 점(point)로 나타낼 수 있다[8]

2.3.2 영상좌표계 변환

영상좌표계 변환을 위해 projective/bilinear model을 적용하였다. projective/bilinear model은 대부분의 카메라 움직임, 즉 변환(translation), 줌(zooming), 수직 및 수평이동(tilting, panning), 회전(rotation) 등의 선형적인 변형과 비선형적인 변형을 나타낼 수 있다[9].

projective model은 8개의 파라미터($m_i, i = 1, 2, \dots, 8$)를 이용하여, 일반적으로 아래와 같은 식으로 표현된다.

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} m_1 & m_2 & m_3 \\ m_4 & m_5 & m_6 \\ m_7 & m_8 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} ; \text{(식 7)}$$

where (x, y) : the original coordinates
 (x', y') : the transformed coordinates

위 (식 7)에서 transformed coordinates (x', y') 는 다음과 같이 식에 의해 구해진다.

$$x' = \frac{m_1x + m_2y + m_3}{m_7x + m_8y + 1} = \frac{u}{w} ,$$

$$y' = \frac{m_4x + m_5y + m_6}{m_7x + m_8y + 1} = \frac{v}{w} ;$$

연속된 영상 프레임 I^{t-1} 와 I^t 에서 서로 연관된 특징점의 집합을 $f^{t-1} = [f_x^{t-1}, f_y^{t-1}]^T$ 와 $f^t = [f_x^t, f_y^t]^T$ 라고 하고, bilinear model로 projective model을 근사하면 다음의 (식 8)와 같이 나타낼 수 있다.

$$f_x^t = q_1 f_x^{t-1} f_y^{t-1} + q_2 f_x^{t-1} + q_3 f_y^{t-1} + q_4 ,$$

$$f_y^t = q_5 f_x^{t-1} f_y^{t-1} + q_6 f_x^{t-1} + q_7 f_y^{t-1} + q_8 ; \text{(식 8)}$$

위 (식 8)에서 8개의 파라미터($q_i, i = 1, 2, \dots, 8$)은 카메라의 ego-motion 파라미터이다. (식 9)와 같이 cost function을 이용하여 최소제곱 최적화로 전체 영상에 적

용하기 위한 transformation model의 파라미터를 추정할 수 있다.

$$J = \frac{1}{2} \sum_{i=1}^N (f_i^{t-1} - T_{t-1}^t(f_i^{t-1}))^2 \quad ; \quad (\text{식 } 9)$$

where N : number of features,
T : transformation model

2.3.3 이동물체 검출 및 추적

(식 9)에서 추정된 transformation model의 파라미터 T를 이용하여 ego-motion이 보정된 영상 I_c 를 다음 (식 10)과 같이 구할 수 있다. 이동 물체는 (식 11)과 같이 I_c 와 I^t 프레임의 영상차에 의해 추출된다.

$$I_c(x, y) = I^{t-1}(T_{t-1}^t(x, y)) \quad ; \quad (\text{식 } 10)$$

$$I_d(x, y) = |I_c(x, y) - I^t(x, y)| \quad ; \quad (\text{식 } 11)$$

연속된 영상의 흐름에서 (식 11)에서 구해진 차영상을 이진영상으로 변환하고, CONDENSATION 알고리즘을 이용하여 차영상($I_d^0, I_d^1, \dots, I_d^t$)에서 일정 임계치 이상의 값을 갖는 픽셀에 해당되는 sample의 confidence 값을 설정하고 이동물체의 위치를 추정하기 위한 가중치를 갱신하여, 사후 확률 밀도($Bel(x^t) = p(x^t | I_d^0, I_d^1, \dots, I_d^t)$)를 추정할 수 있다.

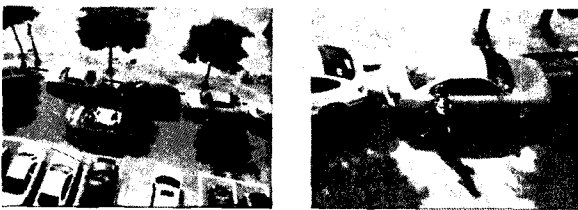


그림 2 실외환경에서 이동물체 검출 및 추적
(좌 : 차량, 우 : 보행자)

3. 결론 및 향후 과제

위 논문에서는 언급한 것과 같이 배경이 고정되지 않은 환경에서 영상차를 이용하여 이동하는 물체를 검출하기 위해서는 카메라의 ego-motion을 보정하여야 한다. ego-motion 보정을 위해 특징점을 추출하고, 특징점 정합을 통해 변환 파라미터를 산출하여 projective/bilinear

model을 사용하였다. 그리고 영상차를 이용하여 생성된 차영상에서 다소의 잡음이 발생하였다. Visual Clutter에서 원하는 물체추적을 가능하게 하는 CONDENSATION 알고리즘을 이용하여 움직이는 물체를 추정하고 추적하였다. 그러나 현재 본 논문의 실험에서 구현된 것은 하나의 물체에 대해서만 추적 가능하고, 다중 물체에 대하여 고려하지 않았다. 예를 들면, 영상에서 다종의 물체를 커다란 하나의 물체로 인식하거나, 가장 크고 많은 움직임 을 나타내는 하나의 물체만 추적하는 결과를 나타낸다. 앞으로 개선된 연구를 통해 영상에서 다종의 물체를 추적할 수 있는 알고리즘으로 확장할 계획이다.

참고 문헌

- [1] 김용진, 이일병, "움직이는 물체검출을 위한 영상 좌표계 변환에 관한 연구", 2006 한국컴퓨터종합학술대회, Vol.33 No.1(B) pp.322-324, 2006. 05.
- [2] M. Isard, A. Blake, "CONDENSATION-Conditional Density Propagation for visual tracking.", International Journal on Computer Vision, 29(1):5-28, 1998.
- [3] D.B. Gennery, "Tracking known 3-D object", Proceeding of AAAI 2nd, Nat. Conference on Artificial Intelligence, pp.13-17, 1982.
- [4] W. Burger, B. Bhanu, "Estimating 3-D Egomotion from perspective image sequences.", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.12, pp.1040-1058, 1990.
- [5] 이규원, 김진웅, "배경 학습과 Bayes 결정을 이용한 이동물체의 추적", 한국통신학회논문지 '99-1 Vol.24 No.11B pp.2138-2146, 1999.
- [6] D. Fox, J. Hightower, L. Liao, D. Schulz, and G. Borriello, "Bayesian filtering for location estimation.", IEEE Pervasive Computing, 2(3):24-33, July-September 2003.
- [7] J. Shi, C. Tomasi, "Good Features to Track", IEEE Conference on Computer Vision and Pattern Recognition, 1994.
- [8] A. Censi, A. Fusiello, V. Roberto, "Image Stabilization by feature tracking", the 10th International Conference on Image Analysis and Processing:665-667, 1999.
- [9] Yiwei Wang, Doherty, J.F., Van Dyck, R.E., "Moving object tracking in video", Applied Imagery Pattern Recognition Workshop, 2000. Proceedings, 29th:95-101, 2000.