

훈련 샘플 수집을 통한 온라인 학습 기반 사람 추적 방법

길종인 김만배

강원대학교 컴퓨터정보통신공학

{jigil, manbae}@kangwon.ac.kr

Online Learning based Human Tracking by Collecting Training Samples

Gil, Jong-in Kim, Manbae

Dept. of Computer and Communications Eng., Kangwon National University

요약

비디오로부터 객체를 검출하기 위해서는 오프라인에서 미리 객체를 검출할 수 있는 분류기가 학습되어있어야 한다. 이러한 분류기는 훈련에 사용된 훈련 집합에 매우 의존적이어서, 다양한 환경의 비디오 영상에 모두 적용할 수 있는 분류기의 설계는 불가능하다. 또한 분류기의 학습을 위해서는 상당히 많은 수의 훈련 집합이 필요하므로, 이는 신뢰도 높은 분류기 학습을 위한 높은 비용을 초래한다. 본 논문에서는 이러한 문제를 해결 할 수 있는 온라인 학습 기반 사람 추적 방법을 제안한다. 실험 영상으로부터 적절하게 훈련 집합을 수집함으로써 해당 실험 영상에 최적화된 분류기의 학습이 가능하며, 다양한 환경의 영상에 적용적으로 설계될 수 있다.

1. 서론

객체 추적은 보안 감시 시스템등의 분야에서 많은 응용을 가질수가 있는데, 기존의 객체 추적 방법은 컬러 히스토그램과 같은 특징을 이용한 템플릿 매칭[1]으로부터 시작되어 커널 기반 추적[2], 움직임 기반 추적[3] 등 다양한 연구가 수행되어왔다. 그러나 사람과 같은 객체는 움직이면서 그 외형이 변하거나 조명등으로 색이 변하므로 이를 해결하기 위해 객체 모델을 갱신하는 적응적 객체 추적 기법이 연구되고 있다. 그러나 이러한 추적 기법의 성공률이 항상 100%를 보장하는 것이 아니기 때문에, 객체 검출과 추적을 결합한 방법 또한 연구되고 있다. 그러나 검출 부분에서는 객체를 검출하기 위한 분류기를 오프라인에서 학습이 수행되어야 한다. 또한 분류기를 훈련하기 위해서는 많은 수의 훈련 집합이 필요하다. 그러므로 본 논문에서는 부류가 정해진 훈련 집합이 불필요한 온라인 학습 기반 추적 기법을 제안한다.

2. 훈련 집합 수집

제안하는 방법에서 객체 추적을 위한 초기화는 필요하지 않다. 단 첫 프레임에서 추적될 객체의 위치는 알려져 있다고 가정한다. 해당 위치로부터 특징 벡터를 추출한다. 본 논문에서는 LBP 기술자를 이용하여 특징을 추출하였다. 특징은 D차원 벡터이다. 즉, 특징 벡터 $\mathbf{x} \in \mathbb{R}^D$ 이다. 이제 하나의 긍정 특징 벡터가 특징 공간에 추가되었다. 이때 해당 객체의 위치를 (x_p, y_p) , 너비와 높이를 각각 w, h 라 할 때, 다음 식 (1)과 같이 해당 위치를 중심으로 경계상자를 설정할 수 있다.

$$BB_p = \left[x_p - \frac{w}{2}, y_p - \frac{h}{2} \right] \times \left[x_p + \frac{w}{2}, y_p + \frac{h}{2} \right] \quad (1)$$

상하좌우 네 방향에 대해 동일한 크기의 경계상자를 취한다. 즉, 네 위치의 좌표는 $(x_p \pm w, y_p \pm h)$ 가 된다. 네 위치에서 동일한 w, h 를 갖는 경계상자를 식 (1)과 같이 생성하고 이로부터 부정 특징 벡터를 추출한다. 이제 4개의 부정 특징 벡터가 특징 공간에 추가되었다. 다음 그림 1과 같이 추적될 사람은 장면에서 사라질 때까지 1개의 긍정 패치와 4개의 부정 패치를 항상 지니게 된다.



그림 1. 객체 경계 상자. (적색 : positive patch, 청색 : negative patch)

3. 분류기 학습

첫 프레임에서는 1개의 긍정 샘플, 4개의 부정 샘플이 존재한다. 프레임 번호를 t 라 하면, 프레임 t 에서는 t 개의 긍정 샘플, $(4 \times t)$ 개의 부정 샘플이 존재한다. 첫 프레임 $t = 1$ 에서 획득한 5개의 훈련 집합으로부터 분류기를 학습한다. 이때, 영상의 초반에는 훈련 집합의 수가 너무 적기 때문에 분류기로부터 신뢰도 높은 분류기의 생성이 어렵다. 따라서 충분한 수의 훈련 집합이 수집되기 전까지는 다른 방법을 사용해야 한다.

예를 들어 첫 번째 프레임으로부터 그림 2와 같이 5개의 샘플이 특징 공간에 추가되었다고 가정하자. 이때, 긍정 샘플, 부정 샘플, 후보 샘플 집합을 각각 P, N, C 라 하고, 각각의 샘플은 P_i, N_j, C_k 이다.

먼저 이전 프레임의 위치 (x_p, y_p) 를 중심으로 하여 탐색 영역을 설정한다. 기존의 검출 방법과 동일하게 슬라이딩 윈도우 기법을 이용하여 후보 특징을 획득한다. 이를 그림 2에서 △로 나타내었다.

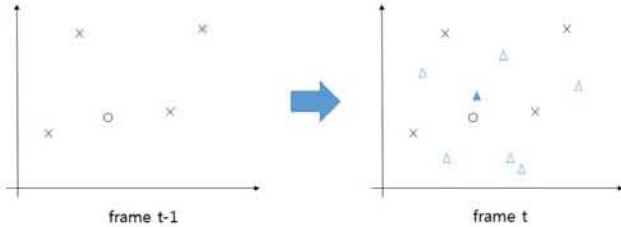


그림 2 2차원 특징 공간에서 후보 샘플들의 분류 결정 과정의 예. (X : 부정 샘플, O : 긍정 샘플, △ : 후보 샘플)

각 후보 특징 C_k 에 대해 분류집합 P, N 와의 거리를 L2-norm으로 측정한다. 이로부터 C_k 가 긍정에 속할 확률 $p(C_k)$ 를 식 (2)를 이용하여 계산한다. 이때, I, J, K 는 각각 긍정 샘플의 수, 부정 샘플의 수, 후보 샘플의 수를 나타낸다. 위 예에서 $K = 7$ 이다.

$$p(C_k) = \frac{\sum \|C_k - N_j\| \times I}{\sum \|C_k - P_i\| \times J} \quad (2)$$

K 개의 후보 특징에 대해 확률 $p(C_k)$ 를 측정하고 가장 큰 확률을 갖는 $p(C_k)$ 를 프레임 t 에서 객체의 위치로 결정한다. 그림 2에서 내부가 채워진 △가 객체의 위치로 결정될 것이다. 객체의 위치가 결정되면 다시 긍정 및 부정 샘플을 결정하고 이를 특징 공간에 추가한다.

4. 분류기 갱신

기존의 온라인 부스팅 알고리즘에서는 하나의 샘플이 추가되면 해당 샘플을 이용하여 분류기를 갱신하고 제거한다. 그러나 제안하는 방법은 새로운 긍정 및 부정 샘플을 무조건 특징 공간에 추가하고, 긍정 샘플의 수가 임계치 T_p 보다 많아지면, 특징 공간에 존재하는 샘플들 중 아웃라이어로 추측되는 샘플을 제거한다. 이때, 훈련 집합 수집에서는 긍정 샘플이 하나씩 증가하므로, 분류기 갱신 과정에서 하나의 긍정 샘플이 제거된다. 부정 샘플은 T_n 보다 증가한 개수만큼 제거되며, 보통 $T_n = 4 \times T_p$ 로 설정한다.

다음은 레이블링된 샘플에서 제거될 샘플을 찾는 것이다. 이때 긍정 샘플과 부정 샘플의 제거 방법이 다르다. 제거해야할 긍정 샘플은 식 (3), 제거해야할 부정 샘플은 식 (4)를 이용하여 추정한다. 즉, 식 (3),(4)에 따라 P_i, N_j 이 선택되고 제거될 것이다.

$$\hat{i} = \operatorname{argmax}_i \|P_i - \bar{P}\| \quad (3)$$

$$\hat{j} = \operatorname{argmin}_j \sum_i \|N_j - P_i\| \quad (4)$$

5. 실험 결과

제안 방법으로부터 획득한 실험 결과를 그림 3에서 보여준다. 실험을 위한 특징 기술자로는 비디오 초반의 사람 추적을 위해 컬러 히스토그램을 이용하였고, 분류기 학습을 위해서는 LBP[4]를 이용하였다. 분류기로는 AdaBoost[5]를 활용하였다. 그림 3에서 제안 방법의 실험 결과를 보여주고 있다. 비디오 초반의 추적 성능이 학습된 분류기의 신뢰도를 결정하기 때문에 상당히 중요하다.

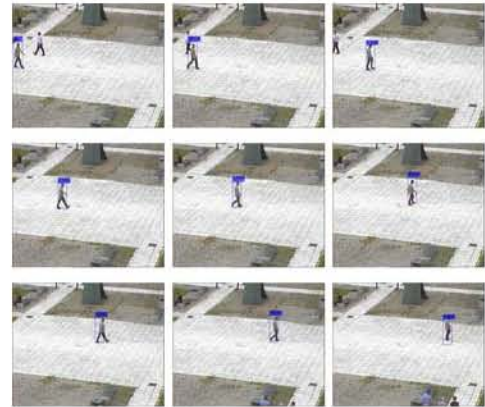


그림 3 단일 사람에 대한 추적 결과

6. 결론 및 향후 연구

본 논문에서는 초기에 훈련 샘플이 존재하지 않아도 실험 영상으로부터 직접 훈련 샘플을 수집하여 분류기를 학습하는 방법을 제안하였다. 제안하는 방법을 통해 검출을 통한 추적이 가능하도록 알고리즘을 확장하였다. 실험 영상으로부터 직접 긍정 및 부정 샘플을 추출하였으므로, 실험 영상에 적용적인 분류기가 학습될 수 있다는 장점이 있다. 본 논문에서 사용한 분류기 이외에도 여러 가지 기계학습 기법이 사용될 수 있으며 또한 실험 영상에 적합한 특징 기술자를 사용자가 선택하여 적용할 수 있다. 본 방법은 향후 다중 사람 추적을 위한 기법으로 확장될 것이다.

참고 문헌

- [1] R. Brunelli, *Template Matching Techniques in Computer Vision Theory and Practice*, Wiley, 2009.
- [2] D. Comaniciu, R. Visvanathan, and M. Peter, "Kernel-based Object Tracking," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(5), pp. 564-577, 2003.
- [3] B. Heisele, "Motion-based object detection and tracking in color image sequences," *Fourth Asian Conference on Computer Vision*, pp. 1024-1033, 2000.
- [4] T. Ojala, M. Pietikainen and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition* 29, pp. 51-59, 1996
- [5] P. Viola and M. Jones, "Robust Real-Time Object Detection," *Int'l Conf. on Computer Vision*, pp. 12-14, July 2001.