합성곱 신경망을 통한 온라인 객체 추적

길종인 김만배

강원대학교 컴퓨터정보통신공학과

jigil@kangwon.ac.kr, manbae@kangwon.ac.kr

Online object tracking via convolutional neural network

Jong in Gil and Manbae Kim

Computer and Communications Engineering, Kangwon National University

요약

본 논문에서는 부류가 정해진 훈련 집합이 불필요한 온라인 학습 기반 추적 기법을 제안한다. 추적기의 학습을 위해 합성곱 신경망(convolutional neural network: CNN)을 이용하였다. 추적영상으로부터 직접 훈련 샘플을 수집함으로써 분류기 학습을 위한 비용을 감소시킬 수 있었고, 목표 영상에 적응적인 객체 모델을 생성할 수 있다. 실험 결과를 통해 제안하는 방법이 우수한 성능을 보임을 입증하였다.

1. 서론

객체 추적은 보안 감시 시스템 등의 분야에서 많은 응용을 가질 수 있으므로 다양한 연구가 수행되어왔다. 그러나 사람과 같은 객체는 움직이면서 그 외형이 변하거나 조명 등으로 인해 색이 변하므로 이를 해결하기 위해 객체 모델을 갱신하는 적응적 객체 추적 기법이 연구되고 있다. 최근에는 객체 추적을 위해 객체 검출을 결합한 방법이 연구되고 있다. 이러한 방법을 검출에 의한 추적(Tracking-by-detection)이라 한다. 객체를 검출하기 위한 분류기를 학습하기 위해 여러 가지 기계학습법이 사용될 수 있다. 그러나 분류기는 오프라인에서 먼저 학습이 선행되어야 한다는 문제가 있고, 또한 분류기를 훈련하기 위해서는 많은 수의 훈련 집합이 필요하다. 이러한 많은 수의 훈련 집합을 생성하는 것은 큰 비용을 초래한다. 본 논문에서는 부류가 정해진 훈련 집합이 불필요한 온라인 학습 기반 추적 기법을 제안한다. 추적기의 학습을 위해 합성곱 신경망(convolutional neural network)을 이용하였다. 그림 1은 제안하는 방법의 전체 흐름도를 보여준다.

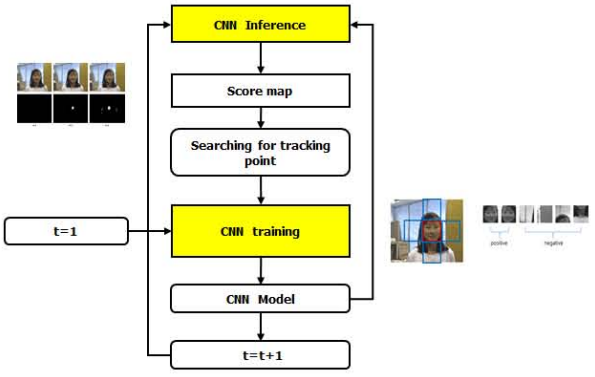

그림 1. 제안하는 방법의 전체 흐름도

2. 훈련 샘플 수집

본 논문에서 제안하는 방법은 영상으로부터 훈련 샘플을 직접 수집한다. 수집된 훈련 샘플은 분류기를 학습하는데 사용된다. 이러한 훈련 샘플 수집 방법은 많은 장점을 가질 수 있다. 먼저 사전에 준비된 훈련 샘플을 이용하여 분류기를 학습하는 경우, 훈련 샘플과 실험 영상의 차이(해상도, 화질 등)로 인해 충분한 성능을 내지 못할 가능성이 크다. 그러나 실험 영상에서 훈련 샘플을 직접 수집하게 되면 이러한 차이로부터 발생하는 성능의 차이를 극복할 수 있다. 또한, 검출기의 사전 훈련을 위해서는 많은 수의 훈련 샘플이 필요하다. 이렇게 많은 수의 훈련 샘플을 생성하는 것은 많은 비용을 초래한다. 따라서 목표 영상으로부터 직접 훈련 샘플을 수집함으로써 이러한 문제를 해결할 수 있다.

추적하고자 하는 객체의 위치를 (x,y), 객체의 높이와 너비를 (w,h)이라 할 때, 다음 식 (1)과 같이 해당 위치를 중심으로 바운딩 박스를 설정할 수 있다. 획득한 이미지 패치에 좌우 반전을 수행함으로써 두 개의 긍정 이미지 패치를 획득한다. 추적하고자 하는 객체의 위치로부터 상하좌우 네 방향에 대해 동일한 크기의 바운딩 박스를 취한다. 즉, 네 위치의 좌표는 $(x \pm w, y \pm h)$가 된다. 네 위치에서 동일한 너비와 높이를 갖는 바운딩 박스를 식(1)과 같이 생성하고, 이로부터 획득한 이미지 패치의 부류를 부정으로 설정한다. 이로써, 총 2개의 긍정 샘플과 4개의 부정 샘플을 획득 한다.

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

3. 합성곱 신경망의 구조

객체를 추적하기 위한 합성곱 신경망은 convolutional layer와 fully connected layer로 구성된다. 합성곱 신경망의 훈련을 위해 수집된 훈련 샘플은 32x32x3의 크기로 변경되어 입력된다. 활성함수로서

ReLU(Rectified Linear Unit), Sigmoid가 사용되었고, 출력 노드에서는 SoftMax가 사용되었다. 그림 2는 사용된 합성곱 신경망의 구조를 보여준다.

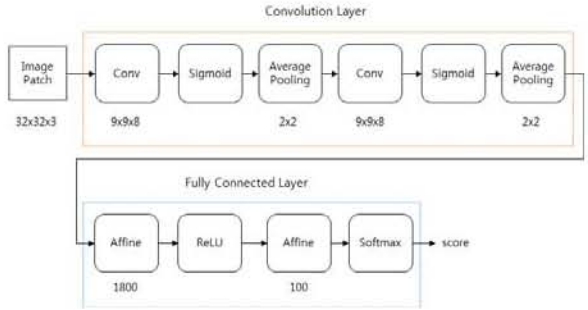

그림 2. 합성곱 신경망의 구조

4. 모델 학습

앞선 과정을 통해 획득한 훈련 샘플은 객체를 올바르게 추적하지 못하게 되었을 경우 신뢰도가 낮아지게 된다. 본 연구에서는 이러한 문제를 해결하기 위해 컬러 모델을 활용하였다. 첫 번째 프레임에서의 객체 위치는 신뢰할만하므로 해당 경계상자로부터 컬러 히스토그램을 측정한다. 다음 프레임이 입력되면 컬러 히스토그램을 역투영(back-projection)시켜 probability map, p를 획득한다. 합성곱 신경망의 손실함수는 cross-entropy를 활용하였는데, probabilyty map과 결합하여 새롭게 정의할 수 있다. 식 (2)에서 새로운 손실함수를 나타내고 있다.

$$L=\frac{1}{N}\sum_{n=1}^{N}\left(d_n p(x_n,y_n)+(1-d_n)(1-p(x_n,y_n))\right)\times \mathcal{J}(l_n,d_n) \quad (2)$$

$$\mathcal{J}(l_n,d_n)=\sum_{i=1}^{M}\left\{-d_{n,i}\ln(y_{n,i})-(1-d_{n,i})\ln(1-y_{n,i})\right\} \quad (3)$$

$$p(x_n,y_n)=\sum_{a=0}^{h_n}\sum_{b=0}^{w_n}p(x_n+a,y_n+b) \quad (4)$$

이때, i는 출력 노드의 인덱스, n은 해당 패치의 인덱스이며, x_n,y_n은 n번째 훈련 패치의 위치, w_n,h_n은 너비와 높이이고, d_n은 훈련 패치의 레이블(ground truth)이고, l_n은 CNN 모델의 출력이다. 식 (4)의 $p(\bullet)$은 역투영 영상으로부터 이미지 패치 내부의 모든 픽셀 값을 합산한 값이다. 이는 적분영상을 적용함으로써 쉽게 획득할 수 있다.

입력된 프레임으로부터 후보 이미지 패치를 획득하여 합성곱 신경망으로부터 분류를 수행한다. 출력된 score는 후보 이미지 패치가 긍정일 가능성을 나타낸다. 합성곱 신경망에서 출력 노드에 SoftMax를 적용하였으므로, 긍정 가능성과 부정 가능성의 합은 1이다. 모든 후보 이미지 패치에 대해 분류를 수행하여 얻은 score를 이용하여 score map을 생성한다. 일반적인 경우, 가장 높은 score를 갖는곳을 객체의 위치로 판단하지만, 본 논문에서 제안하는 방법은 적은 수의 레이어와 적은 수의 훈련 샘플만을 이용하였으므로 가장 높은 score를 선택하는 것은 신뢰할만 하지 않다. 이를 위해 무게중심(Center of gravity)을 측정하여 객체의 위치를 결정하였다. 무게중심은 다음 식(5), (6)를 이용하여 계산된다. 객체의 위치를 결정하였다면 그 위치로부터 다시 훈련 샘플을 수집하여 합성곱 신경망을 업데이트하는 과정을 반복하게 된다.

$$m_{pq}=\sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}i^p j^q f(i,j) \quad (5)$$

$$x_c=\frac{m_{10}}{m_{00}},\ y_c=\frac{m_{01}}{m_{00}} \quad (6)$$

5. 실험 결과

본 논문에서는 총 4개의 영상에 대해 실험을 수행하였다. 실험 영상은 [1-3]으로부터 획득하였으며, 수행한 실험 결과를 그림 3에서 보여주고 있다. 1, 2, 3행의 실험 영상에서는 사람의 얼굴을 추적하였고, 4번째 실험영상은 자전거를 타고 가는 사람을 추적하였다. 추적 결과를 노란색 바운딩 박스로 표시하였으며, 바운딩 박스 위의 숫자는 해당 바운딩 박스가 목표로 한 객체일 확률을 나타낸다. 실험결과로부터 목표 객체를 안정적으로 추적함을 확인할 수 있다.

그림 3. 추적 결과

6. 결론

본 논문에서는 합성곱 신경망을 이용하여 객체를 추적하는 방법을 제안하였다. 실험 영상에 적응적인 객체 추적을 위해 실험 영상으로부터 훈련 샘플을 직접 수집하였다. 합성곱 신경망의 구조에 따라 성능은 달라질 수 있으므로 이를 완화하기 위해 컬러 정보를 활용하였다. 컬러 모델로부터 획득한 정보를 손실 함수에 부합하여 객체 추적의 성능을 강화하였다. 하지만 적은 수의 샘플만을 활용하였으므로 객체 추적의 성능에는 한계가 존재한다. 추후 더 많은 훈련 샘플 정보를 획득하는 방법에 대한 연구가 필요하다.

감사의 글

2017년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. 2017R1D1A3B03028806)

참고 문헌 (References)

[1] http://vision.ucsd.edu/~bbabenko/miltrack.shtml
[2] Y. Wu, J. Lim and MH. Yang, "Online object tracking: A benchmark", IEEE Conf. on computer vision and pattern recognition, pp. 2411-2418, 2014.
[3] http://votchallenge.net/vot2013/dataset.html