재귀 신경망 기반 이벤트 영상의 엣지 추정

백승한, 박종일†
한양대학교
{hameli, jipark}@hanyang.ac.kr

Edge Estimation of Event Data Using Recurrent Neural Network

Seunghan Paek Jong-Il Park
Hanyang University

요 약

본 논문에서는 재귀 신경망을 통해 동적 비전 센서 (DVS: Dynamic Vision Sensor)의 출력에서 엣지를 추정하는 방법을 제안한다. 동적 비전 센서는 기존의 일반적인 카메라들과 달리 급격한 움직임이나 밝기 변화에 강인하게 동작한다. 그러나 동적 비전 센서에서 획득한 출력은 각각이 독립적이기 때문에 화소들의 상관관계를 이용한 알고리즘을 사용함에 어려움이 따른다. 제안하는 방법은 센서에서 획득한 출력을 일정한 시간단위로 분할하고 2차원 평면에 투영함으로써 출력의 정보량 및 상관관계를 향상시키고, 이를 재귀 신경망에 통과시켜 엣지 정보를 추정한다. 이 방법은 센서의 출력에 의해 형성된 패턴을 학습하여 엣지를 잘 추출하였으며, 기존의 컴퓨터 비전 알고리즘의 적용 및 시각 관성 측위 등의 분야에서 활용될 수 있다.

1. 서론

동적 비전 센서 (DVS: Dynamic Vision Sensor)는 사람의 실제 눈의 동작과 유사한 동작을 수행하기 위해 개발된 센서로, 기존의 프레임을 기반으로 동작하는 카메라들과 달리 장면에서 발생한 밝기 변화를 감지하고 이를 이벤트라는 형태로 기록하는 장비이다. 이렇게 획득된 이벤트는 동기적으로 장면을 획득하는 것 대신 장면의 밝기 변화의 정도에 따라 비동기적으로 동작하며, 이러한 특성으로 급격한 움직임이나 밝기 변화에 강인한 모습을 보인다. 그러나 각각의 이벤트는 독립적이며 하나의 이벤트가 갖는 정보량이 매우 적기 때문에 독립된 이벤트를 이용하여 의미 있는 정보를 획득하기에는 어려움이 있다. 이를 보완하기 위해 이벤트의 시공간적 (spatiotemporal) 특성을 고려한 이벤트 처리 방법들이 많이 연구되었다[1]. 그러나 이벤트는 비동기적이며 특정 시간 t 에 발생한 2차원 평면의 점 형태로 기록되기 때문에 해당 이벤트가 의미 있는 정보를 가지고 있는지 판별하는 것은 또다른 문제이다. 이러한 이유로 기존의 컴퓨터 비전 알고리즘을 적용하기 위해 일정한 단위로 이벤트를 누적하여 흑백 영상으로 복원하거나, 이벤트들의 상관관계를 이용하여 의미 있는 정보를 추출하기 위한 연구들이 많이 진행되었다[2, 3, 4].

본 논문에서는 센서와 이벤트의 비동기적 특성을 고려하여 의미 있는 정보 (feature)를 추출하는 방법을 제안한다. 동적 비전 센서는 사람의 눈을 모방하여 제작되었기 때문에, 사람이 사물을 인식하는 과정과 유사하게 사물의 윤곽선에 해당하는 부분에서 많은 밝기 변화가 발생한다[5]. 하지만 이는 일정 시간동안 발생한 이벤트를 2차원 평면에 누적하였을때 확인할 수

† 교신저자

있는 결과이며 독립적인 이벤트가 어떠한 연결성을 갖는지 확인하는 것은 힘들다.

이러한 이벤트간의 연결성을 추정하기 위해서는 각각의 이벤트를 독립적으로 사용하는 것이 아닌 일정 시간동안 발생한 연속적인 이벤트들의 상관관계를 이용하는 것이 필요하다. 이를 위해 재귀 신경망 (RNN: Recurrent Neural Network)을 통해 이벤트를 처리하기 위한 시도들이 이루어졌으며, 최근 연구에서 이벤트의 비동기적 특성을 고려함과 동시에 좋은 성능을 보여주었다[6].

제안하는 방법은 재귀 신경망을 통해 이벤트들의 시공간적 연결성을 고려하여 직선 형태의 특징(엣지)을 추출하였다. 이들은 장면을 인식하기 위한 고유한 정보로 사용되었으며 기존의 컴퓨터 비전 알고리즘에 적용하여 의미 있는 정보를 포함하고 있는지 평가하였다.

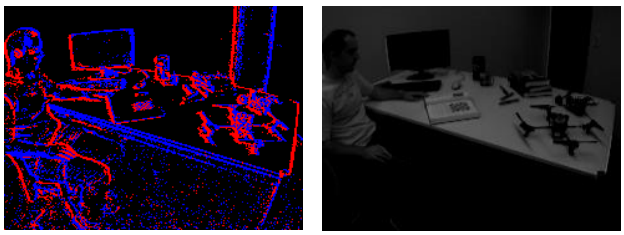


그림 1. 동적 비전 센서의 출력(좌)과 실제 프레임(우)의 비교[8]

2. 제안하는 방법

2.1 이벤트 데이터 처리

동적 비전 센서에서 관측된 이벤트 e 는 발생한 시간 t 와 발생한 위치 (x, y) 그리고 밝기의 변화의 극성 (polarity) p 를 {0,1}로 표현하며 다음과 같이 정의된다.

$$e = \{x, y, p, t\} \tag{1}$$

각각의 이벤트는 독립적이고 굉장히 적은 정보만을 포함하고 있다. 그렇기 때문에 이벤트들의 연관성을 찾기 위해서는 일정 구간의 연속적인 이벤트를 누적할 필요가 있다. 이는 n시간 동안 발생한 이벤트의 집합 $E = \{e_1, e_2, \cdots, e_k\}$와 발생 위치 $\mathbf{x}$에 대해 다음과 같이 나타낼 수 있다.

$$I(\mathbf{x}) = \sum_{e_i \in E} \delta(\mathbf{x} - \mathbf{x_i}) \tag{2}$$

이벤트 데이터는 밝기 변화가 적을 경우 적게 발생하지만 밝기 변화가 급격한 경우 굉장히 많은 수의 이벤트가 발생한다. 이러한 이유로 너무 짧은 간격으로 이벤트를 누적할 경우 너무 적은 이벤트가 포함될 수도 있으며, 너무 긴 간격으로 누적할 경우 너무 많은 이벤트가 포함될 가능성이 있다. 이러한 문제점을 보완하고자 밝기 변화의 증감에 따라 2개의 채널로 분리하였다. 이는 이벤트 고유의 특성을 반영함과 동시에 학습 데이터의 차원을 증가시키는 기능을 수행하였다.

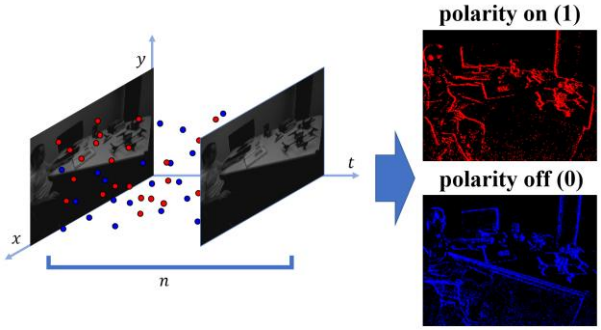


그림 2. n 시간 동안 발생한 이벤트 데이터와 극성에 따라 분리하여 획득한 이벤트 영상

2.2 재귀 신경망 구성

이벤트 데이터는 기본적으로 비동기적 특성을 갖지만 전후 이벤트 사이에 상관관계가 존재하기 때문에 이를 연속적인 데이터로 생각하는 것이 가능하다. 이러한 특성을 반영하기 위해 최근 재귀 신경망을 통해 이벤트 데이터를 처리하기 위한 시도들이 이루어졌고, 실제로 영상 복원 등의 분야에서 우수한 성능을 보여주었다[6, 7]. 본 논문에서는 앞서 언급한 선행 연구의 네트워크를 재구성하여 그림 3과 같이 구조의 네트워크를 구성하였다.

네트워크의 구조는 기본적으로 A개의 인코더와 R개의 잔차 블록(Residual Block), A개의 디코더로 구성된 생성 모델의 형태를 가지며 인코더 내부에 재귀 신경망 (Convolution LSTM)이 위치하고 있다[9]. 마지막 층을 제외한 모든 층에는 ReLU 활성함수와 배치 정규화가 이루어졌으며 마지막 출력층에서만 Sigmoid 함수를 사용하였다.

제안하는 방법과 선행 연구와의 큰 차이점은 학습 과정에 사용한 입출력 데이터와 손실함수에 있다. 선행 연구에서는 입력 데이터로 극성을 고려한 단일 채널 이벤트 영상을 이용하였으며, 출력 데이터로 필터링 알고리즘을 통해 흑백 영상으로 복원한 것을 사용하였다[10]. 또한 장면의 재구축 및 시간 지속성에 대한 손실함수를 정의하여 사용함으로써 흑백 영상을 복원하였다.

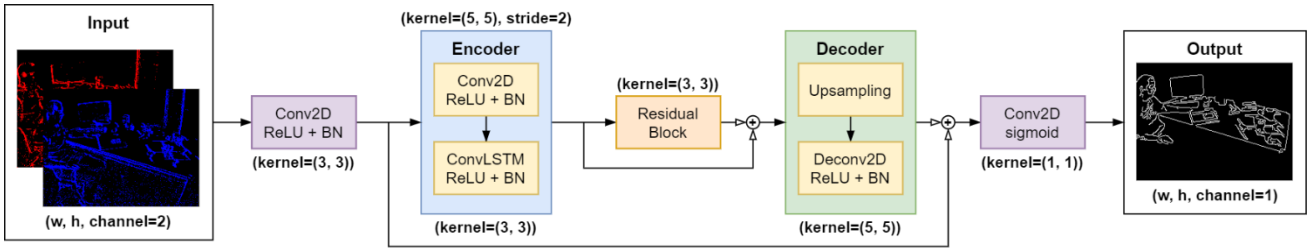


그림 3. 엣지 추정을 위한 전체 신경망의 구조 (A=1, R=1 인 경우의 네트워크 구조)

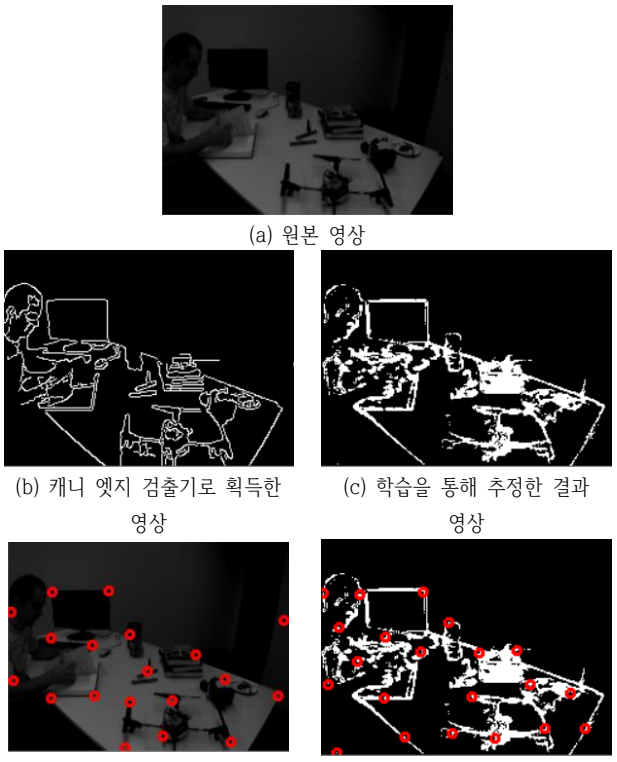
그러나 본 연구에서는 2-1에서 소개한 2채널 이벤트 영상을 입력으로 엣지가 추출된 이진 영상을 출력으로 하고 있으며, 출력 결과를 2차원 공간에서 각각을 이진 분류하는 문제로 계산함으로써 이진 교차 엔트로피 (Binary Cross Entropy)를 사용하였다. 하지만 이벤트 영상의 경우 대부분의 값이 0인 희소한 상태이기 때문에 이를 반영하기 위해서는 값이 존재하는 영역에 대해서만 손실함수를 계산할 필요가 있었다. 그래서 이를 고려하여 값이 있는 영역에 대해서만 손실함수를 계산하였다.

3. 실험 및 학습

학습에는 Mueggler 등이 제안한 DAVIS 데이터 셋을 사용하였다 [8]. DAVIS는 동적 비전 센서와 기존의 프레임 기반 카메라를 결합한 장비로 이벤트와 프레임의 시간을 동기화하여 출력한다. 해당 데이터 셋은 (240 X 180)의 해상도를 가지며 일반 카메라로 촬영된 흑백영상과 이벤트 출력으로 구성된다.

해당 데이터 셋의 경우 흑백 영상과 이벤트의 시간이 동기화 되어있기 때문에 이를 이용하기 위해 이벤트를 300ms 단위로 분할하였으며 이를 2-1에서 언급한 방법을 이용하여 2채널 이벤트 영상을 획득하였다. 라벨 데이터의 생성을 위해서는 데이터셋의 흑백영상에서 캐니 엣지 검출기 (Canny Edge Detection)를 이용하여 엣지 영상을 획득하였다[11]. 획득된 엣지 영상은 일정한 간격으로 촬영되었기 때문에 비동기적인 이벤트 데이터와 함께 학습함에 어려움이 있다. 하지만 이벤트들이 직선의 형태를 띄는 경우에 대부분 흑백 영상에서 또한 직선으로 나타날 것이기 때문에 이는 고려하지 않았다.

최종적으로 300ms 단위로 잘린 2500여개의 데이터를 확보하였으며, 인코더와 디코더가 각각 2개 (A=2), 잔차 블록이 1개 (R=1)인 네트워크 구조를 설계하였다. 또한 배치 사이즈는 32로 설정하여 교차검증을 포함한 50회 학습을 진행하였다.



(a) 원본 영상

(b) 캐니 엣지 검출기로 획득한 영상

(c) 학습을 통해 추정한 결과 영상

(d) 원본 영상에 대한 특징점 추출 결과[12]

(e) 학습 결과에 대한 특징점 추출 결과[12]

그림 4. dynamic_6dof 데이터 셋의 학습 결과

4. 실험 결과 및 분석

학습 결과에 대한 정량적 평가를 위해 그림 5와 같이 학습을 통해 추정한 엣지 영상과 캐니 엣지 검출기로 얻은 영상 간의 신호대 잡음비 (PSNR: Peak Signal-to-Noise Ratio)와 구조적 유사도 (SSIM: Structural Similarity Index Measure)를 계산하였다.

실험 결과는 사용한 데이터의 복잡도에 따라 상이한 결과를 보여주었다. 우선 가장 간단한 “shapes_6dof” 데이터 셋의 경우 아주 안정적인 결과를 얻을 수 있었으며 추정된 엣지 영상 또한 아주 유사하였다. 하지만 데이터 셋의 복잡도가 올라감에 따라 PSNR과 SSIM의 수치가 낮아짐을 확인할 수 있었는데, 특히

이러한 현상은 영상의 대부분이 어두운 환경에서 촬영된 “hdr_boxes” 데이터 셋에서 확인할 수 있었다. 이는 동적 비전 센서에서 밝기 변화로 인지하였음에도 프레임 기반 카메라에서 포착되지 않아 발생한 것으로, 캐니 엣지 검출기를 통해 엣지 영상이 원활히 획득되지 않아 생긴 문제이다. 하지만 학습된 신경망이 어두운 영역의 이벤트를 정상적인 엣지로 검출한 것은 이벤트들의 상관관계를 잘 학습하였음을 알 수 있다.

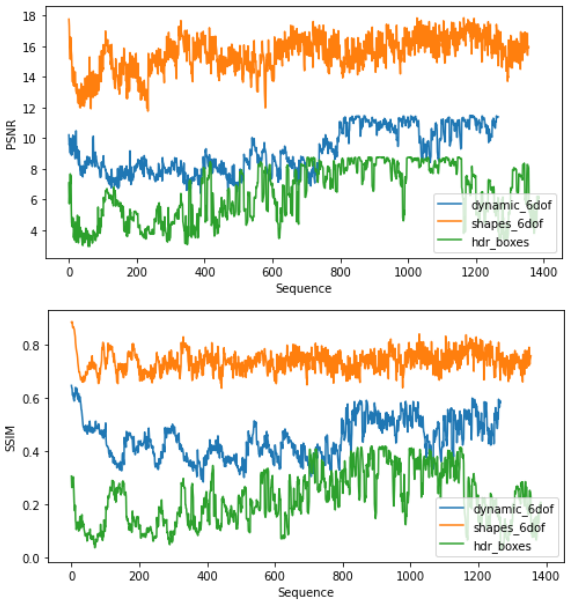


그림 5. 학습을 통해 얻은 결과와 캐니 엣지 검출기를 이용한 결과 간의 PSNR, SSIM 수치 비교

5. 결론

최근 2,3년간 동적 비전 센서의 출력에서 의미있는 정보를 획득하기 위한 연구는 굉장히 활발하게 이루어지고 있다. 각각의 이벤트가 갖는 정보의 양은 굉장히 적기 때문에 이들의 상관관계를 잘 활용하는 것이 중요하다. 제안한 방법은 이벤트 데이터에서 엣지 정보를 획득했을 뿐만 아니라 전반적인 이벤트 데이터가 압축된 효과를 얻을 수 있었다. 이는 동적 비전 센서의 출력에 다른 알고리즘을 적용함에 있어 연산량의 감소와 정확도 향상을 기대할 수 있다. 또한 동적 비전 센서의 특성상 밝기의 변화에 따라 처리해야할 이벤트 데이터 수가 급변하기 때문에 대체로 높은 하드웨어 성능을 요구한다. 하지만 제안한 방법을 적용함으로써 저사양 컴퓨터에서 이벤트 카메라를 동작하기 위한 전처리 기술로서 사용될 수 있을 것이다.

감사의 글

본 연구는 한국전자통신연구원 연구운영비지원사업의 일환으로 수행되었음. [21ZS1200, 인간 중심의 자율지능시스템 원천기술 연구]

참고문헌

[1] R. Benosman, C. Clercq, X. Lagorce, S-H Ieng, and C. Bartolozzi, “Event-Based Visual Flow”, IEEE Transactions on Neural Networks and Learning Systems (TNNLS), vol. 25, no. 2. 2014.

[2] C. Scheerlinck, N. Barnes, R. Mahony, "Continuous-time Intensity Estimation Using Event Cameras", Asian Conference on Computer Vision (ACCV), Perth, 2018.

[3] E. Mueggler, C. Bartolozzi, D. Scaramuzza, “Fast Event-based Corner Detection”, British Machine Vision Conference (BMVC) 2017.

[4] R. Li, D. Shi, Y. Zhang, K. Li, R. Li, “FA-Harris: A Fast and Asynchronous Corner Detector for Event Cameras”, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2019.

[5] C. L. Gentil, F. Tschopp, I. Alzugaray, T. Vidal-Calleja, R. Siegwart, and J. Nieto, “IDOL: A framework for IMU-DVS odometry using lines,” IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2020.

[6] H. Rebecq, R. Ranftl, V. Koltun, and D Scaramuzza, “High Speed and High Dynamic Range Video with an Event Camera”, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2020.

[7] P. R. Gantier Cadena, Y. Qian, C. Wang, M. Yang, “SPADE-E2VID: Spatially-Adaptive Denomalization for Event-Based Video Reconstruction”, IEEE Transactions on Image Processing 2021.

[8] E. Mueggler, H. Rebecq, G. Gallego, T. Delbruck, D. Scaramuzza, "The Event-Camera Dataset and Simulator: Event-based Data for Pose Estimation, Visual Odometry, and SLAM," International Journal of Robotics Research, Vol. 36, Issue 2, pages 142-149, 2017.

[9] X. Shi, Z. Chen, H. Wang, D. Yeung, W. Wong, and W. Woo, “Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting”, Conference on Neural Information Processing Systems (NIPS), 2015.

[10] H. Rebecq, D. Gehrig, D. Scaramuzza,”ESIM: an Open Event Camera Simulator”, Conference on Robotics Learning (CoRL), 2018.

[11] J. Canny, “A Computational Approach to Edge Detection”, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 1986

[12] J. Shi and C. Tomasi, Good features to in Computer Vision and Pattern Recognition, 1994. Proceedings CVPR 94., 1994 IEEE Computer Society Conference on. IEEE, 1994, pp. 593.600.