JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 26 No. 4, pp. 11-19, April 2021
https://doi.org/10.9708/jksci.2021.26.04.011

Generative optical flow based abnormal object detection method using a spatio-temporal translation network

Hyunseok Lim*, Jeonghwan Gwak*

*Student, Dept. of Software, Korea National University of Transportation, Chungju, Korea
*Professor, Dept. of Software, Korea National University of Transportation, Chungju, Korea

[Abstract]

An abnormal object refers to a person, an object, or a mechanical device that performs abnormal and unusual behavior and needs observation or supervision. In order to detect this through artificial intelligence algorithm without continuous human intervention, a method of observing the specificity of temporal features using optical flow technique is widely used. In this study, an abnormal situation is identified by learning an algorithm that translates an input image frame to an optical flow image using a Generative Adversarial Network (GAN). In particular, we propose a technique that improves the pre-processing process to exclude unnecessary outliers and the post-processing process to increase the accuracy of identification in the test dataset after learning to improve the performance of the model's abnormal behavior identification. UCSD Pedestrian and UMN Unusual Crowd Activity were used as training datasets to detect abnormal behavior. For the proposed method, the frame-level AUC 0.9450 and EER 0.1317 were shown in the UCSD Ped2 dataset, which shows performance improvement compared to the models in the previous studies.

▸ **Key words**: Abnormal object detection, Generative adversarial network, Dense optical flow, Image-to-image translation, Image preprocessing

[요 약]

이상 객체란 일반적이고 평범한 행동을 취하는 객체가 아닌 비정상적이고 흔하지 않은 행동을 하여 관찰이나 감시·감독을 필요로 하는 사람, 물체, 기계 장치 등을 뜻한다. 이를 사람의 지속적인 개입 없이 인공지능 알고리즘을 통해 탐지하기 위해서 광학 흐름 기법을 활용한 시간적 특징의 특이도를 관찰하는 방법이 많이 활용되고 있으며, 이 기법은 정해진 표현 범위가 없는 수많은 이상 행동을 식별하기에 적합하다. 본 연구에서는 생성적 적대 신경망(Generative Adversarial Network, GAN)으로 입력 영상 프레임을 광학 흐름 영상으로 변환하는 알고리즘을 학습시켜 비정상적인 상황을 식별한다. 특히 생성적 적대 신경망 모델이 입력 영상에 대한 중요한 특징 정보를 학습하고, 그 외 불필요한 이상치를 제외시키기 위한 전처리 과정과 학습 후 테스트 데이터셋에서 식별 정확도를 높이기 위한 후처리 과정을 고도화하여 전체적인 모델의 이상 행동 식별 성능을 향상시키는 기법을 제안한다. 이상 행동을 탐지하기 위한 학습 데이터셋으로 UCSD Pedestrian, UMN Unusual Crowd Activity를 활용하였으며, UCSD Ped2 데이터셋에서 프레임 레벨 AUC 0.9450, EER 0.1317의 수치를 보이며 이전 연구에서 도출된 성능 지표 대비 성능 향상이 확인되었다.

▸ **주제어**: 이상객체탐지, 생성적 적대 신경망, 밀집 광학 흐름, 이미지 대 이미지 변환, 이미지 전처리

• First Author: Hyunseok Lim, Corresponding Author: Jeonghwan Gwak
*Hyunseok Lim (hyunseoki@ut.ac.kr), Dept. of Software, Korea National University of Transportation
*Jeonghwan Gwak (jgwak@ut.ac.kr), Dept. of Software, Korea National University of Transportation
• Received: 2021. 03. 23, Revised: 2021. 04. 16, Accepted: 2021. 04. 16.

http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945

I. Introduction

이상 상황에 대한 감지·관찰은 정해진 동작이나 구조화된 로직이 내장된 기계 및 컴퓨터 시스템에 비해 행동이 자유로운 사람을 대상으로 요구되거나 취하는 경우가 많다. 특히 이상 상황에서 비정상 행동을 취하는 이상 객체는 그 움직임의 크기가 정상적인 객체보다 크거나 행동 변화의 속도가 빠른 경우가 일반적이다. 이러한 시각적 변화는 감지·관찰을 위한 카메라 영상에서 프레임 내 화소 값의 변화와 이동성으로 표현될 수 있으며, 광학 흐름(Optical flow) 알고리즘은 이러한 화소의 이동 방향과 속도를 파악할 수 있어 정량적 이상 객체 탐지의 핵심 기술로 활용될 수 있다. 광학 흐름 알고리즘을 기반한 여러 이상 객체 탐지 연구[1-3]에서는 광학 흐름 특징 분포의 범위를 분류하여 프레임 상의 이상 유무를 식별하거나, 움직임이 빠른 객체에 대한 지역화(Localization)를 수행하는 성과를 냈다. 하지만 정상적인 상황에서 객체가 나타내는 행동 패턴이나 움직임에 비해 이상 상황에서 나타나는 행동 패턴은 그 표현 범위가 넓고, 각 개별 상황에 대한 레이블링 처리가 어려운 점에 기인하여 비지도 학습에 기반한 이상 상황 식별 모델링을 구현하는 것이 일반적이다[4-6]. 정상 상황에 대한 광학 흐름 특징들만 학습한 모델은 이상 객체가 출현하는 영상에서 나타나는 광학 흐름 특징들을 적절하게 복원하지 못한다는 특성을 활용하여 모든 이상 상황에 대한 식별이 가능하다.

본 연구에서는 정상 상황에 대한 광학 흐름 특징들을 기반으로 비지도 학습 모델을 구현하고 데이터의 중요한 특징 정보를 보존하거나 이상치를 제거하는 전처리 및 후처리 기법을 통해 최종 식별 성능을 향상시키는 과정에 대한 연구를 진행한다.

II. Related works

1. Generative adversarial network

특정 문장을 영어에서 한국어로 번역하거나, 카메라의 아날로그 신호가 RGB 이미지 데이터로 변환되는 등 변환 작업(Translation task)은 이미지 처리, 컴퓨터 그래픽이나 컴퓨터 비전 및 자연어 처리(NLP) 분야에서 많이 활용되고 있다. 하지만 이러한 변환 과정을 수행하는 알고리즘은 각 작업마다 특수적으로 구현되어 문제를 해결하므로 영상의 형태나 표현 색상 범위가 달라질 때 마다 다시 특징 분석을 하고 학습 파라미터를 설정해야 한다. 또한 데이터셋을 구축하기 위한 레이블링 과정이 요구되고 이는 실무에서 해당 모델을 활용하기 어렵게 만들고 범용적이지 못하다.

하지만 생성적 적대 신경망(Generative Adversarial Network, GAN) [7]은 학습 과정에서 제공되는 레이블 정보 없이 생성자(Generator)와 판별자(Discriminator)가 경쟁적으로 각 네트워크의 가중치를 최적화시키는 과정에서 기존 데이터에 존재하지 않는 새로운 유사 데이터를 생성할 수 있다는 점에서 기존 지도 학습 모델링과 큰 차이를 보인다. 이러한 생성적 적대 신경 모델은 인간의 직접적인 파라미터 튜닝의 수고를 덜어주고 여러 데이터셋에 적응가능한 공통 프레임워크로 활용할 수 있다.

2. Optical flow

최근 이상 행동을 탐지하기 위한 딥러닝 시스템은 학습 데이터셋의 평범하고 일반적인 샘플만을 입력 데이터로 주입하여 모델이 정상적인 움직임에 대한 행동 패턴만을 학습하고 기억할 수 있게 하여, 이상 행동이 나타나는 테스트 데이터셋에서는 모델의 패턴 인식 능력이 떨어지는 특성을 활용해 복원 오차(Reconstruction error)가 높은 영역에 대한 지역화를 수행하여 화소 단위에서의 이상 행동 탐지를 수행하고 있다[8]. 이미지 한 장만으로는 그 행동에 대한 시각적 판단을 내릴 만한 정보가 부족하므로 일반적으로 동영상 프레임으로 구성된 영상 시퀀스를 기반한 행동 패턴 분석을 수행한다. 영상 프레임에서 행동 패턴에 대한 정량적 단위를 측정하기 위해서는 이전 프레임과 현재 프레임 사이의 화소 값 차이를 근거해 계산한다. 컴퓨터 비전 분야에서 광학 흐름 기법은 영상 프레임 사이의 시간적 연속성과 화소와 화소가 이웃하는 점들 사이의 공간적인 연속성을 계산하여 객체(화소)에 대한 이동 속력, 방향성을 수치화하여 표현할 수 있다.

3. Image-to-Image translation network

정상 행동만 포함된 학습 데이터셋은 일반적으로 고정된 카메라로부터 원거리에서 객체들이 움직이는 공간이나 환경을 장시간 촬영한 동영상으로 구성되어 있다. 이러한 영상 프레임은 배경과 움직이는 객체를 모두 포함하고 있는 공간적 특징으로 표현되며, 해당 프레임과 이전 프레임과의 차이를 계산한 광학 흐름 프레임은 해당 프레임에 대한 시간적 특징으로 표현된다. 그러므로 생성적 적대 신경망의 인코더 단으로 들어오는 공간적 특징 벡터는 정상 행동만을 포함한 영상 프레임이 되며, 디코더 영역에서는 해당 프레임에 대한 시간적 특징 벡터인 광학 흐름으로 구성할 수 있다.

D. Pathak의 연구[9]에서는 생성적 적대 모델이 특징을 학습하고 손실을 최소화시키는 과정에서 어떠한 손실 함수를 사용했을 때 생성 성능이 향상되는 지에 대한 시각을 제공하였다. Isola, Phillip의 연구[10]에서는 생성적 적대 신경망 모델이 영상에서 다른 영상으로 맵핑하는 과정에서 여러 손실 함수에 대한 각 출력 이미지의 비교 실험과 특징 추출을 통해 데이터셋의 특징 분포에 상관없이 판별자의 판별 능력을 뛰어넘는 수준의 생성 성능을 끌어내어, 다양한 영상 변환 작업에서 높은 복원율을 달성하고 광범위한 응용 가능성을 시사하였다. 해당 연구를 이상행동탐지 문제에 적용하여 유의미한 결과를 낸 Ravanbakhsh의 연구[11]에서는 전체 입력 이미지에 대한 광학 흐름 출력 결과를 내도록 생성적 적대 신경망을 학습시켰다. 하지만 네트워크가 관심 객체에 대한 특징만을 학습하도록 유도하기 위한 장치가 없고 원본 데이터셋을 Patch 단위로 나누지 않고 그대로 입력 데이터로 주입하였는데, 이러한 이유 때문에 결과 프레임에서의 배경이 제대로 표현되지 않아 회색조를 띄며 객체가 이동하는 방향성 특징을 제대로 학습하지 못해 결과 프레임의 선명도가 다소 떨어지는 것이 특징이다. 이는 복원 오차를 계산하는 과정에서 이상치로서 작용하게 되고, 결과적으로 탐지 성능을 떨어뜨리는 요인이 된다. 본 연구에서는 해당 문제점에 대한 적절한 전처리 과정을 적용하여 생성적 적대 신경망이 관심있는 특징에만 집중하여 학습할 수 있도록 유도한다.

III. The Proposed Scheme

1. Overview

본 연구의 이상 행동 탐지 시스템의 주요 모듈은 움직이는 객체 추출, 시공간 변환, Adversarial 손실 계산, 후처리로 구성되어 있다.

데이터셋 전처리 과정에서는 시공간 변환 네트워크의 입력으로 주입될 영상 프레임을 최적화하는 과정이 진행되고, 변환 모델은 학습 시 실제 프레임 영상으로부터 Fake 광학 흐름 프레임을 생성하는 과정을 학습한다. 그 뒤 테스트 데이터셋으로부터 추출된 Fake 광학 흐름 프레임과 광학 흐름 알고리즘으로 계산된 Real 광학 흐름 프레임의 화소 차이를 계산하고 후처리를 통해 차이가 많이 발생되는 영역에 대한 강조 후 원본 영상 프레임과 합성하여 이상 행동 객체에 대한 지역화 처리를 마친다. 전체적인 시스템 구성도는 Fig 1.으로 요약된다.

이미지 대 이미지 학습 모델은 입력 영상을 출력 영상으로 변환하기 위해서 필요한 특징들을 학습하게 되는데 이는 이상 행동을 탐지하기 위해 정상적인 행동 패턴만을 학습시키기 위해서 요구되는 손실함수가 적용되기에 적절한 모델링이다. 카메라로부터 촬영된 정상적인 행동 패턴만 포함된 영상 프레임을 광학 흐름 출력 영상으로 변환하는 것은 생성자로 하여금 정상적인 객체 움직임에 대한 광학 흐름 특징들만을 학습하게 할 수 있으며, 이상 객체로부터 발생되는 비정상 행동 패턴에 대한 광학 흐름 출력은 정상 광학 흐름과 거의 유사한 색상 스케일을 가지게 된다. 모델

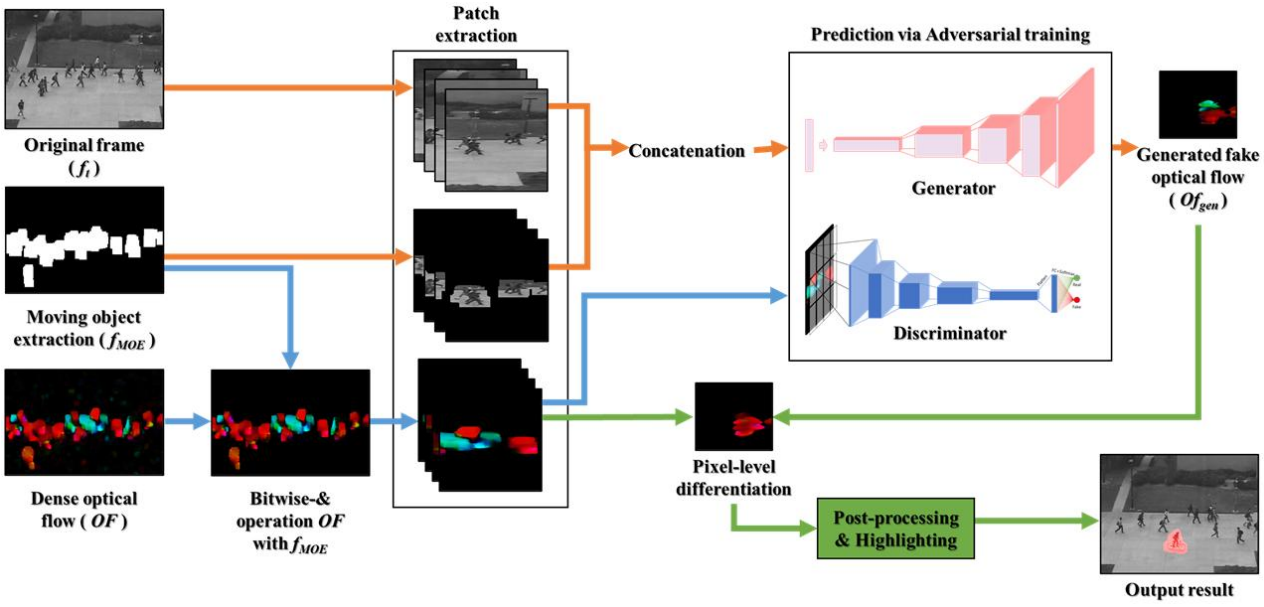

Fig. 1. Overall architecture of the GAN-based abnormal object detection system

은 학습 시에 정상적인 행동만을 포함한 데이터셋만 받아들였고 이러한 영상에는 사람이 일반적인 속도로 걷는 단조로운 움직임 속력과 방향성만을 포함한다. 하지만 테스트 데이터셋에서는 정상적인 패턴 영상 외에 빠른 속도로 달리는 장면, 커다란 물건을 들고 걷거나 자전거 또는 스케이트 보드를 타는 등의 행동 패턴을 보이는 프레임이 포함되며, 이러한 화소 분포와 변화도는 생성자가 최소화된 손실을 내면서 정상적으로 결과 영상을 생성하기 어려워진다. 결국 해당 이상 객체가 위치해 있는 영역에서 계산된 광학 흐름 정보와의 값 차이가 벌어져 수치값을 비교했을 때 그 차이가 커지게 된다. 입력 영상에 대한 광학 흐름 계산 결과를 RGB 영상으로 변환하여 출력하게 되면 해당 영역의 차이가 크게 나타나는 것을 시각적으로 확인할 수 있으며, 실제로 계산된 Real 광학 흐름 영상과 모델의 생성자가 계산한 Fake 광학 흐름 영상의 화소 차이를 계산하게 되면 이상 행동 객체가 위치해 있는 영역을 지역화할 수 있다.

2. Moving object extraction

이미지 변환 네트워크에서 생성자는 입력 영상을 기반으로 출력 영상을 생성하게 되며, 영상이 달라질 때 마다 해당 영상에 대한 적절한 특징들을 잡아내어 결과를 도출한다. 모델 학습 시에는 입력 영상과 매핑이 되는 결과 영상을 짝지어 생성자와 판별자 신경망의 입력으로 주입되며, 이러한 매핑이 적절하게 이뤄지고 출력 영상이 적절하게 표현되기 위해서는 그에 맞는 충분한 가이드라인 정보가 입력 영상에 포함되어야 한다. 이상 행동 감지를 위한 학습 데이터셋에서는 일반적인 평범한 패턴만을 포함한 프레임 시퀀스로 구성되어 있으며, 모델에게 관심을 집중시켜야 할 대상은 해당 프레임에서 움직임이 발생하는 객체가 위치해 있는 영역이다. 대부분의 이상 행동은 움직이는 객체에 의해서 유발되므로 객체를 포함하는 배경 정보는 그렇게 중요치 않다. 또한 배경은 나뭇잎이 바람에 날리거나, 촬영 카메라의 떨림에 의해 일반적인 상황에서도 미세한 움직임이 있을 수 있으며 이러한 움직임은 모델의 학습에서 중요한 부분이 아니다. 그러므로 움직임이 크게 나타나는 객체를 원본 영상에서 구분하고 이 영역에 대한 윤곽을 구분지어 가이드라인을 부여할 수 있다.

동영상은 연속적인 프레임의 화소 변화를 시간 순으로 나열한 데이터이다. 영상 프레임에서 움직이는 객체를 추출하기 위해서는 이전 프레임 f_{t-1}과 현재 프레임 f_t의 화소 차이를 회색조 범위인 0부터 255 사이의 화소값으로 계산하여 표현할 수 있으며, 이를 f_{MOE}이라고 할 때 수식 (1)로서 계산된다.

$$f_{MOE} = | f_{t-1} - f_t | \quad (1)$$

차이가 크게 나타나는 화소는 동영상 데이터 관점에서 변화가 크고 빠르게 움직이는 영역으로 판단될 수 있으며 행동 데이터셋에서는 관심 객체의 영역에서 그 변화가 크게 나타나게 된다. 동시에 배경에서의 작은 화소 값 변화 또한 포함되어 계산되므로 Fig 2.의 두 번째 그림처럼 화소값으로서 그 강도를 표현할 수 있다. 이러한 배경 노이즈 차이를 제외시키고 객체의 움직임을 강조하기 위해서는 임계치 및 이진화 처리와 모폴로지 연산을 적용하는 과정이 요구된다. Fig 2.의 세 번째 그림은 그 결과를 보여주고 있다.

영상 프레임의 움직이는 객체에 대한 추출 과정은 시공간 변환 네트워크가 광학 흐름 프레임을 생성하는 과정에서 관심 있는 화소 영역에 대한 윤곽선 정보를 제공하여 이상치의 생성을 억제하는 역할을 한다. 즉, 원본 프레임과 객체 추출 프레임을 Patch 단위로 분할하고 연결한 2개의 채널을 가진 데이터가 시공간 변환 네트워크의 입력으로 주입된다.

3. Optical flow calculation

광학 흐름은 연속적인 영상 프레임에서 나타나는 화소의 이동 방향과 속력을 정량적으로 표현하는 객체 추적 기법 중 하나이다. 프레임 내 화소 움직임의 크기와 방향을 2채널로 표현하며, Fig 3.와 같이 움직임의 방향성은 색상 Hue 값으로, 움직임의 크기는 명도 Value로 스케일을 정규화하여 시각화할 수 있다.

Fig. 2. Extraction process of moving objects

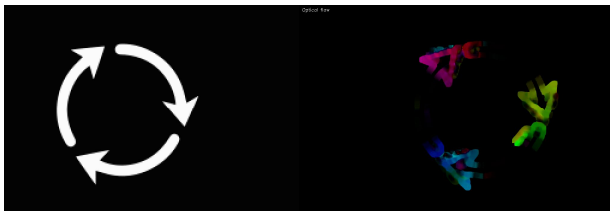

Fig. 3. Visualization of dense optical flow direction and intensity for moving image frames

특히 밀집 광학 흐름는 희소 광학 흐름 기법에 비해 시간 복잡도가 커지는 대신 모든 화소 포인트들에 대한 광학 흐름을 계산하여 특성 포인트들의 정확도와 해상도를 늘려주는 이점이 있다. 본 연구에서는 밀집 광학 흐름 알고리즘을 활용하여 실제 영상 프레임 사이의 객체 이동성을 추적한다.

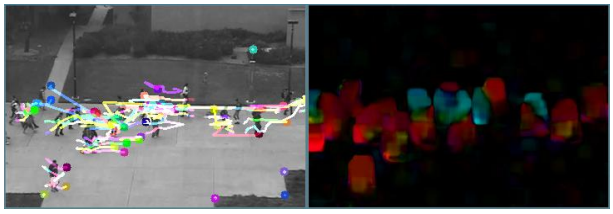

Fig. 4. Difference in the calculation method between sparse optical flow(left) and dense optical flow(right)

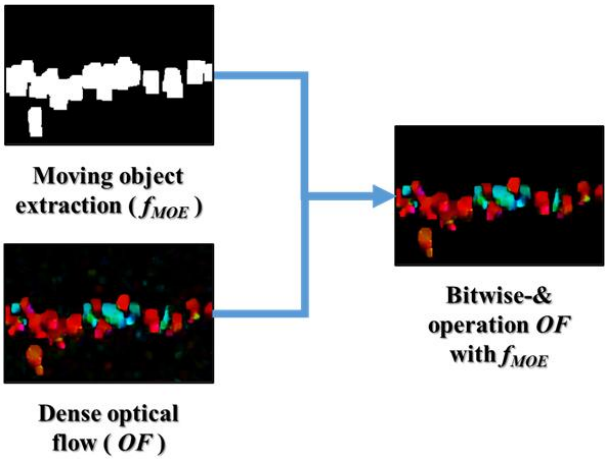

Fig. 5. Process of extracting optical flow features only for objects of interest

그러나 학습 데이터로 쓰이는 현실의 영상 프레임에는 실제 관심을 가져야 하는 객체의 움직임 뿐만 아니라 배경에 존재하는 나뭇잎의 미세한 흔들림, 카메라의 떨림에 의한 화소의 이동성 또한 포함되어 계산이 된다. 이러한 이상치들은 생성적 적대 신경망 모델이 주요 객체에 대한 공간적 특징을 광학 흐름의 시간적 특징으로 변환하는 과정에서 성능을 떨어뜨리는 원인으로 작용될 수 있다. 그러므로 원본 프레임으로부터 이동하는 객체의 영역을 추출한 프레임 f_{MOE}와의 Bitwise-& 연산을 적용하여 관심 객체에 대한 광학 흐름 특징만을 남겨두는 추가적인 처리가 Fig. 5과 같이 필요하다.

4. Patch extraction

대부분의 영상 프레임은 1:1 비율의 화면비가 아니며, 딥러닝 네트워크의 입력으로 주입되는 영상 크기보다 크게 촬영되는 것이 일반적이다. 이러한 영상의 크기를 그대로 재조정하여 네트워크로 주입하는 것은 모델이 학습을 하는 과정에서 적절한 공간적 특징을 뽑지 못하며, 원본 영상의 해상도와 모델의 입력 해상도의 차이가 클수록 비효율적이다. 이러한 공간적 특징을 좀 더 효율적으로 추출하고 모델이 지역적 화소의 패턴을 학습하기 위해서 Patch 기반 모델 학습 방식을 채택한다. Patch는 입력 영상의 지역적 특징을 비교적 잘 보존하며, 움직이는 관심 객체의 외형 정보를 잘 나타낸다. Patch 추출 시 커널의 크기는 $\frac{W}{\lambda} \times H \times C$로 정의될 수 있으며, W는 프레임의 가로 길이, H는 세로 길이, C는 채널 수로 정의된다. C는 회색톤의 영상일 때 1로 지정되며, RGB 색상 이미지의 경우 3으로 설정된다. λ는 영상 프레임의 가로 길이에서 Patch로 추출할 비율을 지정하는 Scale value로서 작용한다. 보통 영상의 프레임은 가로의 길이가 세로의 길이에 비해 긴편이 일반적이므로 네트워크의 입력 크기로 줄이는 과정에서 가로의 길이를 등분하는 것이 좀 더 많은 정보를 보존할 수 있다. 이 값의 변화량에 따라 그 정도를 조절할 수 있으며, 본 연구에서는 4로 설정하고 Stride S는 $\frac{W}{\lambda}$로 지정하여 한 영상 프레임에서 4개의 Patch가 생성되도록 한다. 각 f, f_{MOE}, OF 프레임에 대해서 일괄 Patch 추출을 진행하여 f_{patch}, $f_{MOE}^{bit\ patch}$는 생성자 네트워크의 입력으로, OF_{patch}^{bit}는 판별자 네트워크의 입력으로 주입한다.

5. Spatio-temporal translation network

GAN 네트워크는 공간적 특징 정보를 입력으로 받는 생성자가 그에 상응하는 시간적 특징 정보인 광학 흐름 영상을 생성해 낼 수 있도록 관심 객체의 특징 정보를 학습해야 한다. 그 과정으로 네트워크 입력에는 원본 영상 프레임의 정보 f와 관심 객체 영역을 강조한 f_{MOE} 프레임 정보가 채널 2개로 결합되어 주입되며, 판별자는 광학 흐름 알고리즘으로 계산된 Real 광학 흐름 프레임과 생성자가 생성한 Fake 광학 흐름 프레임을 비교하면서 생성자의 생성 능력을 향상시킨다.

이러한 영상 변환 네트워크는 U-net 아키텍처를 기반으로 한 이미지 대 이미지 변환 네트워크[10] 통해 구현된다. 생성자 G는 학습 및 예측 시 모두 작동되는데, 기본 GAN[7] 모델에서는 입력 데이터와 랜덤 노이즈 z를 받는 반면, 이미지 대 이미지 변환 네트워크에서는 공간적 특징 정보가 포함된 입력 데이터를 기반으로 시간적 특징 정보를 생성해야 하는 상황에서 랜덤 노이즈 z가 효율적이지 못하다. 그러므로 추가적인 가우시안 노이즈 z 대신 Drop-out을 포함한 형태의 네트워크 아키텍처로 구성된다.

생성자 G와 판별자 D를 학습하기 위해서 목적함수는 L1 Loss L_{L1}와 GAN Loss L_{GAN}로 구성되며, 각 함수는 수식 (2)와 수식 (3)으로 정의된다.

$$L_{L1}(G) = E_{x,y}[\,||\,y - G(x)\,||\,] \tag{2}$$

$$\begin{aligned} &L_{GAN}(G, D) \\ &= E_y[\log D(y)] + E_x[\log(1 - D(G(X)))] \end{aligned} \tag{3}$$

L1 Loss L_{L1}은 생성자가 원본 프레임의 공간적 특징 정보와 광학 흐름의 시간적 특징정보 사이에서 공통적으로 표현되는 관심 객체의 외형 정보를 학습하는 목적으로 활용되며, GAN Loss L_{GAN}은 생성자 G가 판별자 D를 속일 만큼의 정확성을 가진 광학 흐름 생성 능력을 경쟁적으로 학습하기 위해서 요구되는 손실 함수이다.

6. Adversarial loss calculation

학습 단계가 끝난 후 모델은 정상 상황에서 관심 객체의 움직임에 대한 일반적인 행동 패턴과 특징을 시간적 특징인 광학 흐름 데이터로 표현할 수 있는 능력을 지닌 상태이다. 예측 단계에서는 테스트 데이터셋을 생성자 G의 입력으로 주입하고 출력으로 나오는 Fake 광학 흐름 데이터를 얻는다. 생성자 G가 만들어낸 광학 흐름에는 정상 행동 패턴의 특징만을 활용해 테스트 데이터셋에 존재하는 모든 객체의 활동들을 표현하게 되므로 이상 행동이 발생하는 프레임에 대해서도 생성자 G는 정상 객체와 유사한 광학 흐름 방향성과 강도를 화소값으로 표현할 것이다. 그러나 광학 흐름 알고리즘을 통해 테스트 데이터셋을 계산한 Real 광학 흐름 프레임은 생성자 G가 임의로 만들어 내는 수치값이 아닌 화소 값의 이동 방향과 강도를 수치적으로 계산한 실제 데이터가 되므로 두 Fake, Real 광학 흐름 사이에 오차가 발생하게 된다. 특히 이상 행동이 발생하는 객체의 영역에서는 그 차이가 더 커지게 되므로 이 두 프레임 사이의 광학 흐름 색상 값 차이 $\triangle OF$를 수식 (4)로 계산하게 되면 이상 행동을 유발하는 영역에 대한 지역화가 가능해진다.

$$\triangle OF = OF_{real} - OF_{fake} \tag{4}$$

이 차이의 정도에 따른 임계치 설정을 통해 이상 행동이 발생하는 영역에 대한 모델의 민감도를 설정할 수 있다.

일반적인 이미지 포맷으로 저장되는 영상 데이터는 RGB 또는 BGR 방식으로 표현되는데, 이러한 3채널 색상 스케일에 대한 두 광학 흐름 프레임의 차이를 계산하는 것은 광학 흐름의 방향 정보가 색상 정보로써 표현되므로 방향성 정보의 차이까지 나타나게 된다. 하지만 이상 행동을 탐지하는 과정에서 객체가 움직이는 방향 정보는 많은 군중에 밀집된 공간에서 그리 중요한 정보가 아닐 수 있다. 그러므로 RGB 색상 스케일을 색조 (Hue), 채도 (Saturation), 명도 (Value)로 구성된 HSV 스케일로 변환하여 움직임의 강도에 대한 차이를 Value 값으로 계산하는 것이 더 유의미하다.

7. Postprocessing for accurate localization

$\triangle OF$는 유의미적 지역화가 아닌 단순한 화소값의 차이이므로 False positive가 많이 발생하게 되어, 후처리 과정을 통해 이러한 이상치를 줄여야 한다. 관심 객체는 일반적으로 사람의 눈에 보일 만큼의 충분한 면적을 가지고 있기에 일정 면적 이하의 작은 영역은 탐지에서 제외시키고, 각 영역당 가장 큰 화소값이 임계치 이하인 영역은 삭제하는 과정이 필요하다.

IV. Experiment Results

본 연구에 대한 실험은 이상 행동 데이터셋인 UCSD Pedestrian[12], UMN Unusual Crowd Activity[13]를 활용하여 진행하였으며, 프레임 차원에서 이상 행동의 발생 유무를 판별하는 성능 지표를 측정하고 유사한 기존 연구에서 보여준 성능 지표와 비교하였다.

1. Dataset

UCSD Pedestrian 데이터셋은 ped1과 ped2 두가지 영상 종류로 나눠지며 각 영상은 여러개의 시퀀스 별 프레임 파일로 제공된다. UMN Unusual Crowd Activity는 여러 시퀀스가 단일 영상 파일로 합쳐져 제공되며 상단에 캡션으로 이상 행동 발생 유무를 표현하고 있다.

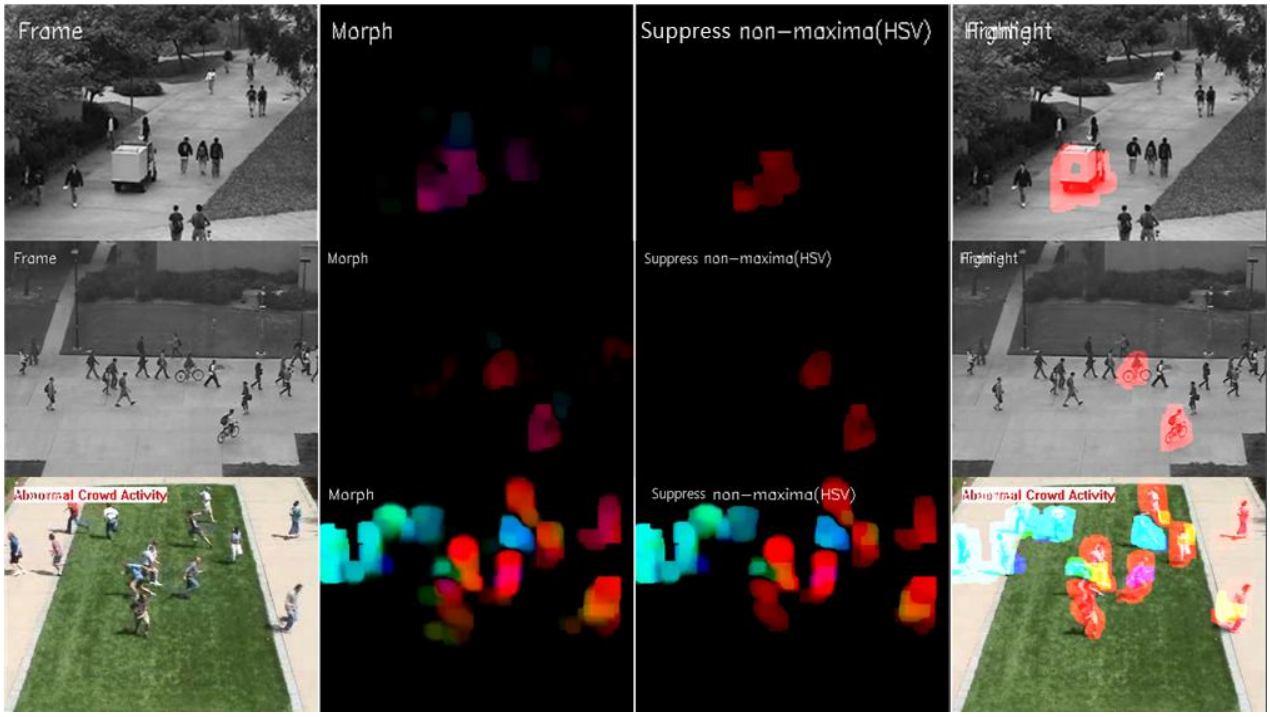

Fig. 6. Abnormal behavior detection process(Input frame, Morph result, Non-maxima suppression, Highlight) in each dataset

Fig. 7. UCSD Ped1(left), UCSD Ped2(center), UMN(right) dataset

2. Evaluation

프레임 차원에서의 이상 행동 유무를 판단하기 위해서 테스트 데이터셋의 Ground truth 프레임과 모델의 예측 프레임을 비교한다. Ground truth가 존재하는 프레임에 대해서 모델이 최소 하나 이상의 Positive(양성)를 식별 했을 때 True positive로 판단하며, Ground truth가 존재하지 않는 프레임에 대해서 모델이 하나 이상의 Positive 예측을 할 때 False positive로 판단한다. 화소 차이의 정도에 따라 임계치를 설정하여 그 이상의 화소값을 가진 영역에 대해서는 이상 행동으로, 그 이하는 정상 행동으로 판단할 수 있으며, 이러한 Positive와 Negative 수치에 대한 비율을 표현하기 위해서 TPR(True Positive Rate)와 FPR(False Positive Rate)을 활용한다.

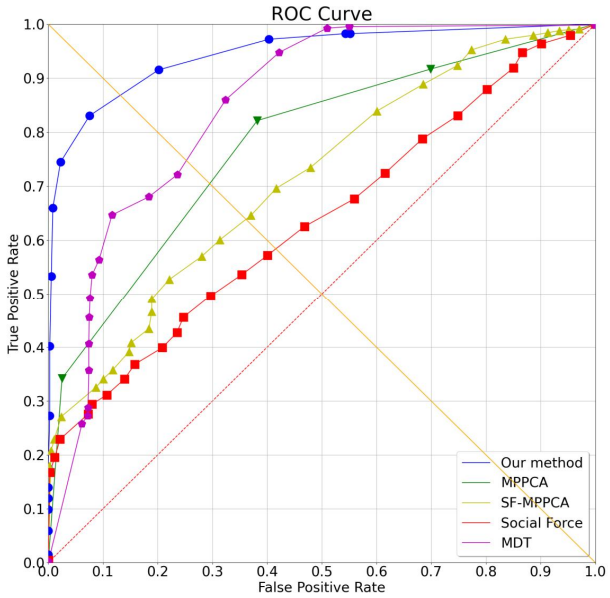

Fig. 8. Performance comparison with previous studies

$$TPR = \frac{TP}{TP + FN} \quad , \quad FPR = \frac{FP}{FP + TN} \qquad (5)$$

TPR은 모든 Ground truth 중에서 얼마나 많은 객체가 True positive로 탐지되는지의 비율로서 False negative를 피하는 것이 중요할 때 활용된다. 이상 행동을 놓치는 것을 최소화하고 모든 이상 행동을 탐지하는 것이 비즈니스 목표로서 더 적합하므로 TPR 대비 FPR을 ROC 곡선으로 시각화하여 최적의 임계치 값을 선정할 수

있다. Fig 6.은 테스트 데이터셋을 대상으로 수행된 이상 객체 탐지 결과 프레임을 보여주고 있다. Fig 8.은 UCSD Ped2 데이터셋에 대한 학습된 모델 및 기존 연구의 AUC 수치를 보여주고 있으며 본 연구의 AUC 수치가 더 높은 성능을 내고 있음을 보여주고 있다.

V. Conclusions

본 연구에서는 Generative 모델의 패턴 학습에 기반하여 정상 행동의 공간적 움직임 특징들을 학습하고, 이상 행동이 나타나는 공간에서 광학 흐름 연산과의 차이를 활용해 이상 객체에 대한 지역화를 수행하였다. 특히 생성자가 생성한 Fake 광학 흐름과 실제 연산을 통해 도출된 Real 광학 흐름과의 차이를 계산하는 과정에서 유의미한 탐지를 위해 추가적인 후처리 과정을 거치면서 정확도를 향상시켰다.

추후 각 이상행동에 대한 캡션 처리와 느린 속도로 움직이거나 이동하지 않는 객체에 대한 추적과 행동 패턴 탐지를 강화하고 카메라의 촬영 구도로 인해 발생되는 원근감에 의한 속도 차이를 감응하는 연구를 진행하여 현실 환경에 강인한 이상 객체 탐지 시스템을 구현할 계획이다.

ACKNOWLEDGEMENT

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (Grant No. NRF-2020R1I1A3074141), the Brain Research Program through the NRF funded by the Ministry of Science, ICT and Future Planning (Grant No. NRF-2016M3C7A1905477), and "Regional Innovation Strategy (RIS)" through the NRF funded by the Ministry of Education.

REFERENCES

[1] A. Das K.M. and O. V. R. Murthy, "Optical Flow Based Anomaly Detection in Traffic Scenes," 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), pp. 1-7, 2017, DOI: 10.1109/ICCIC.2017.8524181

[2] Y. Gao et al., "Anomaly Detection for Videos of Crowded Scenes based on Optical Flow Information," 2018 3rd International Conference on Advanced Robotics and Mechatronics (ICARM), pp. 869-879, 2018, DOI: 10.1109/ICARM.2018.8610722

[3] E. Duman and O. A. Erdem, "Anomaly Detection in Videos Using Optical Flow and Convolutional Autoencoder," in IEEE Access, vol. 7, pp. 183914-183923, 2019, DOI: 10.1109/ACCESS.2019.2960654

[4] M. Ravanbakhsh, M. Nabi, E. Sangineto, L. Marcenaro, C. Regazzoni and N. Sebe, "Abnormal event detection in videos using generative adversarial nets," 2017 IEEE International Conference on Image Processing (ICIP), pp. 1577-1581, 2017, DOI: 10.1109/ICIP.2017.8296547

[5] M. Ravanbakhsh, E. Sangineto, M. Nabi and N. Sebe, "Training Adversarial Discriminators for Cross-Channel Abnormal Event Detection in Crowds," 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1896-1904, 2019, DOI: 10.1109/WACV.2019.00206

[6] W. Liu, W. Luo, D. Lian and S. Gao, "Future Frame Prediction for Anomaly Detection - A New Baseline," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6536-6545, 2018, DOI: 10.1109/CVPR.2018.00684

[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," Proc. Adv. Neural Inf. Process. Syst., pp. 26722680, 2014

[8] M. Ravanbakhsh, M. Nabi, E. Sangineto, L. Marcenaro, C. Regazzoni, and N. Sebe, "Abnormal event detection in videos using generative adversarial nets," IEEE Int. Conf. Image Process. (ICIP), pp. 15771581, Sep. 2017

[9] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell and A. A. Efros, "Context Encoders: Feature Learning by Inpainting," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2536-2544, 2016, DOI: 10.1109/CVPR.2016.278

[10] P. Isola, J. Zhu, T. Zhou and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5967-5976, 2017, DOI: 10.1109/CVPR.2017.632

[11] Ravanbakhsh, Mahdyar, Moin Nabi, E. Sangineto, L. Marcenaro, C. Regazzoni and N. Sebe. "Abnormal event detection in videos using generative adversarial nets," 2017 IEEE International Conference on Image Processing, pp. 1577-1581, 2017, DOI: 10.1109/ICIP.2017.8296547

[12] V. Mahadevan, W. Li, V. Bhalodia and N. Vasconcelos, "Anomaly Detection in Crowded Scenes," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010, (Accessed Jan., 2, 2021).

[13] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," IEEE Conf. Comput. Vis. Pattern Recognit., pp. 935942, Jun. 2009, (Accessed Jan., 2, 2021).

Authors

Hyunseok Lim received the B.S degree in Software engineering from Korea National University of Transportation (KNUT) in 2019. From 2020, he has been working as a Master's student and researcher Artificial Machine Intelligence Lab in KNUT. His research interest is deep learning, computer vision, and abnormal object detection using machine learning and adversarial training.

Jeonghwan Gwak received the Ph.D. degree in machine learning and artificial intelligence from Gwangju Institute of Science and Technology, Gwangju, Korea in 2014. From 2002 to 2007, he had worked for several companies and research institutes as a researcher and a chief technician. From 2014 to 2016, he worked as a postdoctoral researcher in GIST, and from 2016 to 2017 as a research professor. From 2017 to 2019, he was a research professor in Biomedical Research Institute & Department of Radiology at Seoul National University Hospital, Seoul, Korea. From 2019, he joined Korea National University of Transportation as an assistant professor, and he is the director of the Applied Machine Intelligence laboratory. His current research interests include deep learning, computer vision, signal and image processing, AIoT, evolutionary algorithms and optimization, fuzzy sets and systems, and relevant applications of biomedical and visual surveillance systems.