

## 컬러 히스토그램과 CNN 모델을 이용한 객체 추적

# Object Tracking using Color Histogram and CNN Model

박성준 · 백중환\*

한국항공대학교 항공전자정보공학부

Sung-Jun Park · Joong-Hwan Baek\*

School of Electronics and Information Engineering, Korea Aerospace University, Goyang, 10540, Korea

### [요 약]

본 논문에서는 컬러 히스토그램과 CNN 모델을 이용한 객체 추적 기법 알고리즘을 제안한다. CNN (convolutional neural network) 모델 기반 객체 추적 알고리즘인 GOTURN (generic object tracking using regression network)의 정확도를 높이기 위해 컬러 히스토그램 기반 mean-shift 추적 알고리즘을 합성하였다. 두 알고리즘을 SVM (support vector machine)을 통해 분류하여 추적 정확도가 더 높은 알고리즘을 선택하도록 설계하였다. Mean-shift 추적 알고리즘은 객체 추적에 실패할 때 경계 박스가 큰 범위로 움직이는 경향이 있어 경계 박스의 이동거리에 제한을 두어 정확도를 향상 시켰다. 또한 영상 평균 밝기, 히스토그램 유사도를 고려하여 두 알고리즘의 추적 시작 위치를 초기화하여 성능을 높였다. 결과적으로 기존 GOTURN 알고리즘보다 본 논문에서 제안한 알고리즘이 전체적으로 정확도가 1.6% 향상되었다.

### [Abstract]

In this paper, we propose an object tracking algorithm based on color histogram and convolutional neural network model. In order to increase the tracking accuracy, we synthesize generic object tracking using regression network algorithm which is one of the convolutional neural network model-based tracking algorithms and a mean-shift tracking algorithm which is a color histogram-based algorithm. Both algorithms are classified through support vector machine and designed to select an algorithm with higher tracking accuracy. The mean-shift tracking algorithm tends to move the bounding box to a large range when the object tracking fails, thus we improve the accuracy by limiting the movement distance of the bounding box. Also, we improve the performance by initializing the tracking start positions of the two algorithms based on the average brightness and the histogram similarity. As a result, the overall accuracy of the proposed algorithm is 1.6% better than the existing generic object tracking using regression network algorithm.

**Key word** : CNN, GOTURN, Mean-shift, SVM, Color histogram.

<https://doi.org/10.12673/jant.2019.23.1.77>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 31 January 2019; Revised 7 February 2019

Accepted (Publication) 27 February 2019 (28 February 2019)

\*Corresponding Author; Joong-Hwan Baek

Tel: +82-2-300-0125

E-mail: jhbaek@kau.ac.kr

## I. 서론

4차 산업 혁명의 핵심 기술인 인공지능 기술의 발달로 인하여 객체 추적에 대한 연구 및 응용이 감시 시스템, 음향, 스포츠 등 다양한 분야에서 진행되고 있다. 또한 사람이 관찰하지 못하는 부분이나 도달할 수 없는 곳에서의 촬영을 위해 개인 소유의 무인항공기 많아지고 무인항공기에서의 360도 영상 촬영, 하나의 객체를 추적하며 비행하는 등의 연구가 활발히 진행 중이다. 최근 고화질 영상, 360도 영상 콘텐츠 등에서 픽셀 수 증가로 알고리즘이 처리해야 할 데이터의 수가 더욱 많아지고 있다. 따라서 무인항공기를 포함한 다양한 분야에서 기기의 오작동을 줄이고 응용이 되기 위해서는 좀 더 정확하고 구현 속도가 빠른 알고리즘이 필요하다. 이에 본 논문에서는 컬러 히스토그램 정보와 CNN 모델을 이용한 객체 추적 기법을 제안한다.

본 논문에서는 실시간 이상의 속도로 객체를 추적하기 위해 CNN 모델 기반 객체 추적기인 GOTURN 알고리즘을 사용하여 객체를 추적한다[1]. 그리고 Mean-shift 추적 알고리즘을 GOTURN과 융합하여 색 정보를 주어 객체 추적 실패 확률을 낮추었다. 두 추적 알고리즘의 실행 시기는 SVM을 활용하여 분류하였고 그에 대한 트레이닝 dataset은 VOT(visual object tracking) challenge에서 사용한 VOT2014, VOT2015를 사용하였다[2]-[4]. 각 추적기가 예측한 위치의 히스토그램을 비교하여 더욱 유사한 쪽으로 추적기의 시작 위치를 초기화 시켜 알고리즘의 전체적인 정확도를 높였다.

본 논문의 2장에서는 객체 추적 알고리즘인 GOTURN과 Mean-shift 추적 알고리즘에 대해 설명한다. 3장에서는 두 알고리즘을 분류하는 SVM에 대해 설명하고 4장에서 설계한 알고리즘의 구조도를 보인다. 5장에서 실험 결과를 보이며, 6장에서 결론을 맺는다.

## II. Tracker

본 연구에서는 다양한 분야에서 응용이 될 수 있는 알고리즘을 설계하기 위해 어떠한 객체든 추적을 할 수 있어야 하고 실시간 성능 이상의 성능을 보이는 객체 추적 알고리즘을 선택하였다. 최근 VOT Challenge 과제에서 최고의 성능을 보여준 추적기인 MDnet, C-COT 등을 사용하려 했으나 속도가 실시간 성능 이하였고 최근 CNN기반 추적기인 GOTURN은 실시간 성능 이상의 성능을 보였다[5]. 성능 또한 MDnet과 C-COT와 비교하였을 때 큰 차이를 보이지 않아 본 연구의 기준 추적기로 선정하였다. 그러나 GOTURN 알고리즘은 실험을 통해 확인한 결과 객체 추적을 실패하는 경우가 있어, 이를 보완하기 위해 컬러 히스토그램을 기반으로 추적하고 일반적인 모든 객체를 추적할 수 있는 Mean-shift 추적 알고리즘을 GOTURN에 융합하였다. 또한 본 연구에서 사용한 객체 추적 알고리즘에 대해 원래와 같은 기본 매개 변수를 사용하여 구현하였다.

### 2-1 GOTURN

GOTURN은 ALOV300++와 ImageNet 데이터 셋에서 얻은 학습 데이터에 fine-tuning과 online learning 없이 완전히 의존한다. 그림 1과 같이 학습 단계에서 연속된 프레임 쌍을 무작위로 선택하고 이전 프레임에서 추적할 객체 영역을 지정하여 자르고 현재 프레임에서는 동일 위치의 영역을 잘라 convolution layer에서 객체와 객체 주변에 대한 정보를 학습하고 fully-connected layer에서 두 convolution layer weight 값을 공유하여 현재 프레임에서 경계 상자의 위치를 예측하도록 학습한다. 여기서 그림 2와 같이 동영상보다 정적인 이미지의 객체를 마치 움직인 듯이 잘라 학습을 더 많이 시켰기 때문에 학습이 더 많이 된 객체를 추적하려는 경향이 크다. 하지만 이전 프레임에서 무엇을 추적할지 설정을 하기 때문에 일반적인 객체들을 모두 추적할 수 있다.

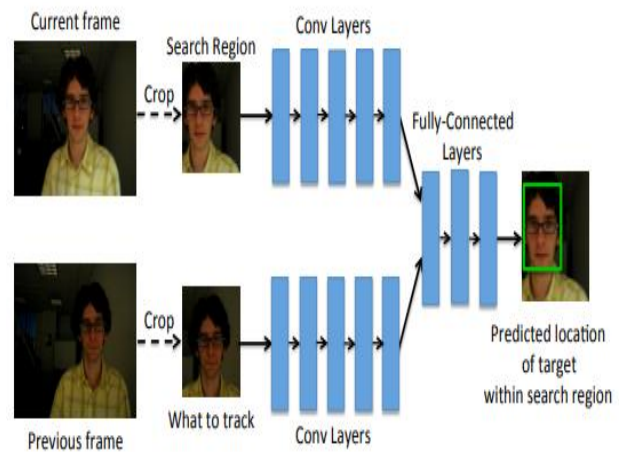


그림 1. GOTURN 알고리즘 네트워크 구조도  
Fig. 1. GOTURN algorithm network architecture.



그림 2. GOTURN의 정적인 이미지 학습 예시  
Fig. 2. Example of GOTURN's static image learning.

따라서 **offline** 학습 데이터만을 사용하여 속도가 빠르고 어떠한 객체를 추적할 수 있다. 하지만 모션정보가 없어 추적 객체의 일부 또는 전체가 다른 객체에 가려지게 되면 성능이 현저히 떨어진다. 또한 추적에 한 번이라도 실패하게 되면 실패한 추적의 위치를 토대로 계속해서 추적해 나가기 때문에 초기 추적 대상을 다시 찾지 못하게 된다. 따라서 본 연구에서는 초기 설정한 객체를 GOTURN이 추적에 실패했을 때에도 초기 설정 객체의 색 정보를 가지고 있으면 다시 추적할 것이라고 판단하였고, 앞서 언급한 단점들을 보완하기 위해 본 설계에서는 색 정보기반 추적기인 Mean-shift 알고리즘을 사용하였다.

**2-2 Mean-shift 추적기**

Mean-shift란 데이터 분포의 중심을 찾는 방법으로 현재 자신의 주변에서 데이터가 가장 밀집된 방향으로 이동하는 방식의 알고리즘이다. Mean-shift가 영상 내 객체 추적에서 사용될 경우 지정된 객체의 색상 기반 히스토그램과 현재 입력 영상의 히스토그램을 비교하여 가장 유사한 히스토그램을 갖는 영역을 찾는 방식으로 사용된다. 간단한 예를 그림 3에 보인다. 그림 3의 object model의 히스토그램은 식(1)로 나타낼 수 있다 [6]-[7]. 여기서  $Hist_{model}$ 은 객체의 히스토그램 모델이고  $p_v$ 는 모델 히스토그램 각 픽셀의 확률이다.

$$Hist_{model} = p_v (v = 1 \dots n) \tag{1}$$

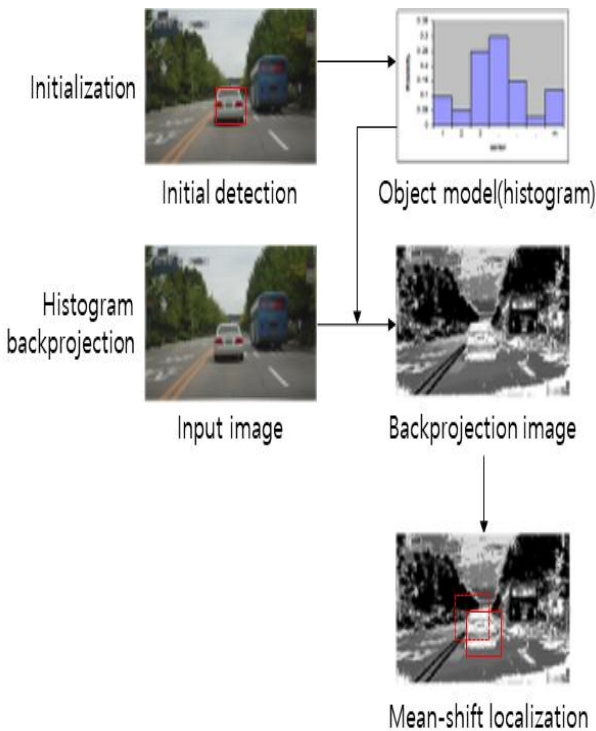


그림 3. Mean-shift 추적 알고리즘 구조도  
Fig. 3. Algorithm structure of Mean-shift Tracker.

그리고 입력 영상 Input의 픽셀 값  $x$ 를 식(2)를 사용하여 0~1 사이의 확률 값  $w(x)$ 로 변경시킨다.  $Hist_{bkg}$ 는 배경의 히스토그램이다.

$$w(x) = \sqrt{\frac{Hist_{model}(Input(x))}{Hist_{bkg}(Input(x))}} \tag{2}$$

식 (2)의 확률 값  $w(x)$ 와 Epanechnikov 커널 함수  $K$ 를 이용해 mean-shift 이동거리  $\Delta x$ 를 구할 수 있다.

$$\Delta x = \frac{\sum w(x)(x - \bar{x}_{old})K(x - \bar{x}_{old})}{\sum w(x)K(x - \bar{x}_{old})} \tag{3}$$

본 논문에서는 색상 기반 객체 추적과 탐색 반경의 크기가 유연한 Tomas Vojir의 robust scale-adaptive Mean-shift tracker 알고리즘을 사용한다[8]. 이 알고리즘은 Mean-shift 추적기의 중심점을 찾는 방식과 같다. 하지만 중심점을 찾은 후 윈도우 사이즈의 스케일을 변화시키며 지정한 객체의 히스토그램 유사도가 가장 높은 윈도우 크기를 찾는 점이 다르며 기존의 Mean-shift 추적기보다 높은 성능을 보인다.

**III. 분류기**

GOTURN과 Mean-shift 추적 알고리즘 중 하나를 선택하기 위해 본 연구에서는 SVM을 활용하여 분류하였다.

SVM이란 기계 학습의 분야 중 하나로 지도 학습을 이용한 이진 분류기이다. SVM은 주어진 데이터 집합을 두 범주로 나눌 때 분류하려는 데이터를 선형으로 분리하는 최적의 선형 경계를 찾는 알고리즘이다. SVM의 목표는 서로 다른 범주의 데이터들을 가장 큰 마진(margin)으로 분리해내는 모델을 찾는 것이다.

지도 학습을 이용하여 모델을 만들며 이 모델을 기준으로 새로운 데이터가 어느 카테고리에 속할지 판단할 수 있다.

그림 4의 경우에서 주어진 데이터들에 대해 H1은 분류에 실패

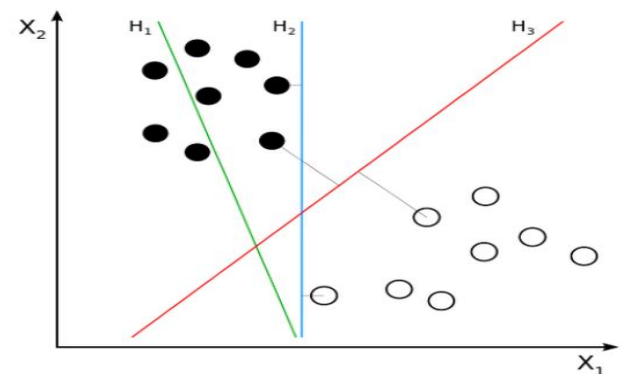


그림 4. SVM 기존 데이터 분류 예  
Fig. 4. SVM example of existing data classification.

패하였고 H2와 H3는 분류에 성공하였다. 하지만 H2의 경우 마진이 작기 때문에 그림 5와 같이 새로운 데이터가 들어왔을 때 분류에 실패할 가능성이 크다. 반면 H3는 주어진 데이터에서 최대 마진을 갖는 모델을 만들어놓았으므로 새로운 데이터가 들어왔을 때 더 안정적으로 분류할 수 있다[9].

본 설계에서 SVM의 학습 데이터 형태로 4개의 히스토그램 유사도와 2개의 라벨로 구성하였다. 4개의 히스토그램 유사도는 각각 첫 번째 프레임에서의 객체와 현재 프레임에서 GOTURN과 Mean-shift 추적기가 찾은 객체와의 히스토그램 유사도, 3번째 전 프레임에서의 객체와 현재 프레임에서 GOTURN과 Mean-shift 추적기가 찾은 객체와의 히스토그램 유사도로 구성된다. 3번째 전 프레임을 사용한 이유는 GOTURN이 객체 추적에 실패하기까지 총 3프레임이 걸리는 것을 실험적으로 확인하였기 때문이다.

SVM의 학습 데이터 라벨은 GOTURN과 Mean-shift 추적기의 결과 좌표( $A_b$ )와 ground truth( $A_G$ )를 각각 비교하여 IOU(intersection of union)가 높은 쪽을 라벨로 선택하였다. 둘 다 IOU가 0인 경우 두 추적기중 ground truth와의 거리가 짧은 것을 라벨로 선택하였다. IOU는 두 경계 상자  $A_G$ 와  $A_b$ 의 교집합 영역에서 합집합 영역으로 나눈 것으로 객체 검출의 정확성을 판단할 때 사용하는 지표이다. 그림 7과 식 (4)에 IOU의 개념을 보인다[10].

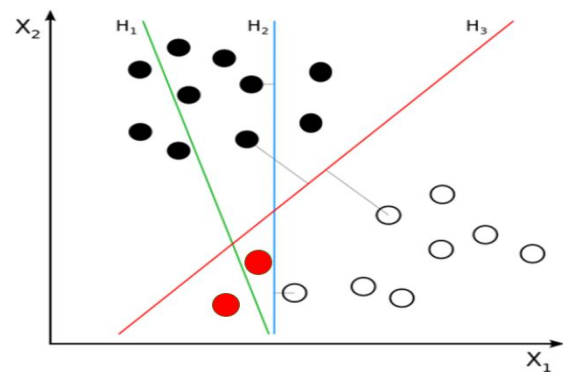


그림 5. SVM 새로운 데이터 추가 분류 예  
Fig. 5. SVM classification example for new additional data.

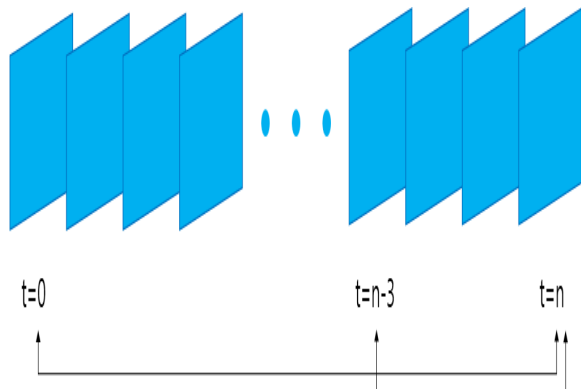


그림 6. SVM 학습 데이터 형태  
Fig. 6. SVM learning data type.

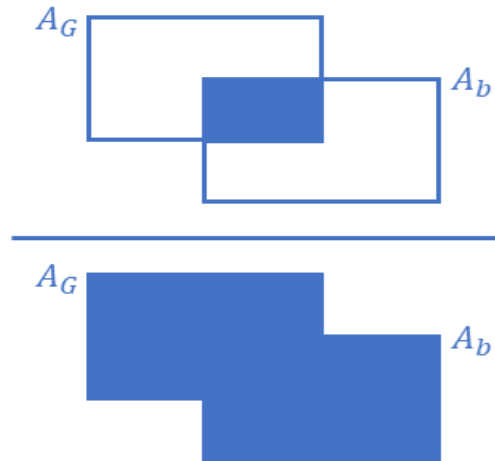


그림 7. IOU  
Fig. 7. IOU.

$$IoU = \frac{A_G \cap A_b}{A_G \cup A_b} \quad (4)$$

구성된 학습 데이터와 라벨로 SVM을 학습시켜 이진 분류 모델을 만들었다.

#### IV. 알고리즘 구조도

그림 8은 본 논문에서 제안한 알고리즘의 구조도이다. 히스토그램 유사도를 사용한 데이터와 IOU를 사용한 라벨을 SVM에 학습시켜 GOTURN을 사용할지 Mean-shift 추적기를 사용할지 선택을 한다. 그리고 선택된 추적기가 예측한 경계 박스의 히스토그램과 세 프레임 전의 히스토그램을 비교하여 유사하면 선택되지 않은 추적기를 초기화 시켜 정확도를 높이고 유사하지 않다면 다시 SVM이 더 정확한 추적기를 선택하게 한다. 또한 Mean-shift 추적기는 초기 히스토그램 모델만을 사용하여 객체를 추적하기 때문에 추적 객체가 아닌 다른 곳으로 경계 박스가 튀는 현상이 생긴다. 또한 객체의 밝기에 민감해 어두운 곳과 밝은 곳에서 객체의 히스토그램 모델을 저장하게 되면 영상의 밝기가 변화하게 되어 추적이 되지 않는다. 따라서 추적하는 객체의 픽셀 평균을 구해 밝기를 1~10 단계로 선형적으로 나누어 전체 밝기 단계가 3 이하(영상 전체 평균 0 ~ 77)이거나 8 이상(영상 전체 평균 205 ~ 255)일 때는 GOTURN이 실행되게 하여 추적 성능을 높였다. 그림 8에서 경계 박스의 이동 거리를 Mean-shift distance로 명명하였고 임계 값을 영상의 대각선 길이의 100분의 1로 설정하였다. 따라서 직전 프레임과 현재 프레임에서 임계 값 이상으로 경계 박스가 움직이게 되면 추적에 실패할 확률이 크기 때문에 GOTURN을 실행시켰다. 또한 객체 영역 밝기 단계(brightness level)가 8보다 크거나 3보다 작으면 GOTURN을 실행시켜 주어 알고리즘의 정확도를 향상시켰다.

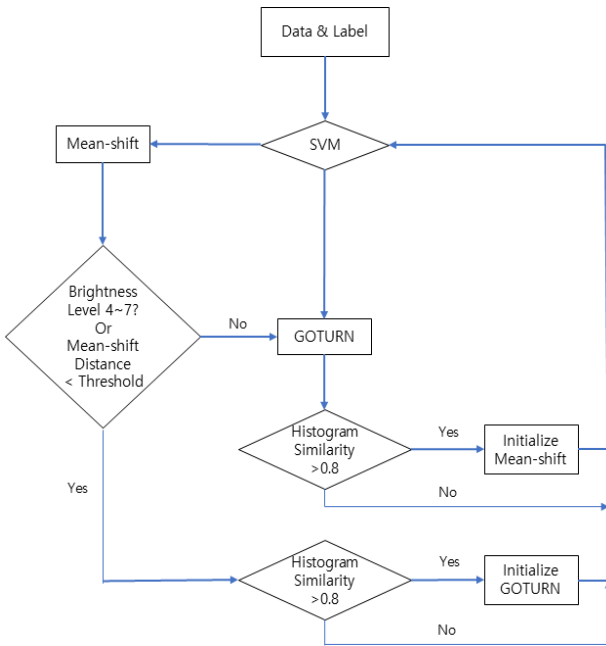


그림 8. 알고리즘 구조도  
Fig. 8. Algorithm structure.

### V. 실험

알고리즘의 성능 평가를 위해 ground-truth 경계 상자와 본 논문에서 제안한 알고리즘이 예측한 경계 상자의 IOU 값을 사용하였다. 매 프레임 IOU값에 사용된 비디오 시퀀스 전체 프레임 수를 나누어 영상 별 평균 IOU를 구하였다. 테스트 데이터 셋으로는 객체 추적 알고리즘 대회 visual object tracking challenge에서 사용한 총 25개의 영상 시퀀스가 있는 VOT2014를 사용하였다. IOU가 50%가 넘을 때 추적에 성공하였고 50% 미만일 때 추적에 실패했다고 판단을 하였다. 전체 영상 시퀀스의 IOU 전체 평균은 프레임 별 IOU 값의 합에 전체 프레임 수를 나눈 값이다. GOTURN은 48.88%, 본 논문에서 제안한 알고리즘은 50.43%로 약 1.6%의 성능 향상을 보였다. 최종적으로 전체 프레임 수는 10214 프레임, GOTURN이 추적에 성공한 전체 프레임 수는 6151 프레임, 본 논문에서 제안한 알고리즘은 6178 프레임이었다.

그림 9에 객체간의 겹침 현상에 있어서 기존의 GOTURN 보다 성능이 크게 향상된 사례를 보인다. 대표적으로 4, 15, 25번 시퀀스가 대표적인 사례이다. 그림 10에는 본 논문에서 제안한 알고리즘이 히스토그램 모델의 영향으로 인해 경계 상자가 기존 GOTURN 보다 다소 정적이어서 성능이 하락된 사례를 보인다. 대표적으로 12, 17, 19번 시퀀스가 대표적인 사례이다.

25개의 영상 시퀀스 별 성능 비교를 표 1과 그림 11에 나타내었다. 평균 IOU가 50%가 넘는 영상 시퀀스에서 작은 차이로 성능이 하락된 경우가 있었지만 평균 IOU가 50%보다 낮고 기존 GOTURN에서는 추적이 잘 되지 않았던 영상 시퀀스에서 추적 정확도가 크게 향상됨을 확인하였다.

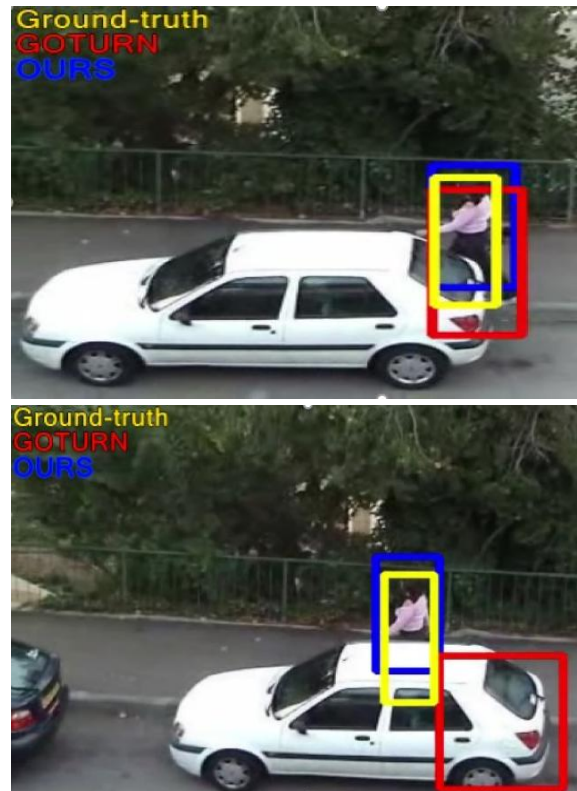
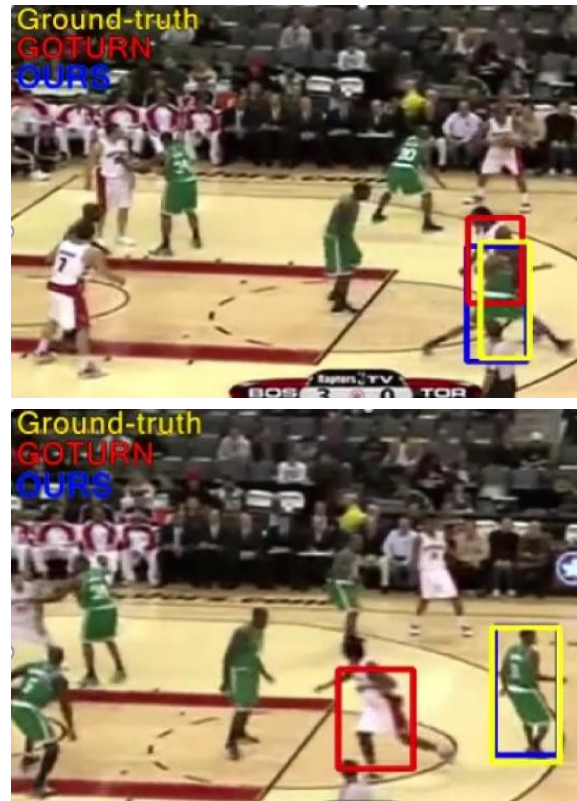


그림 9. 추적 성능 향상 예  
Fig. 9. Case of tracking performance improvement.

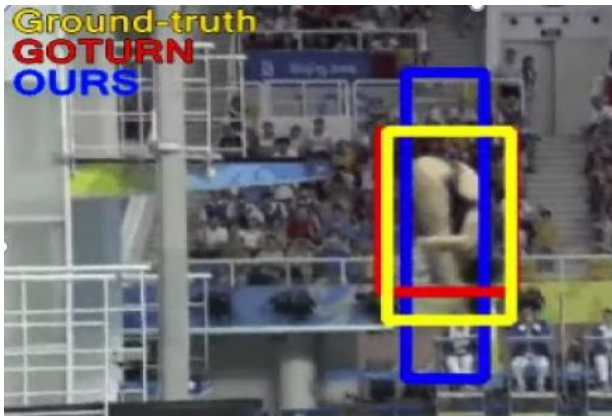


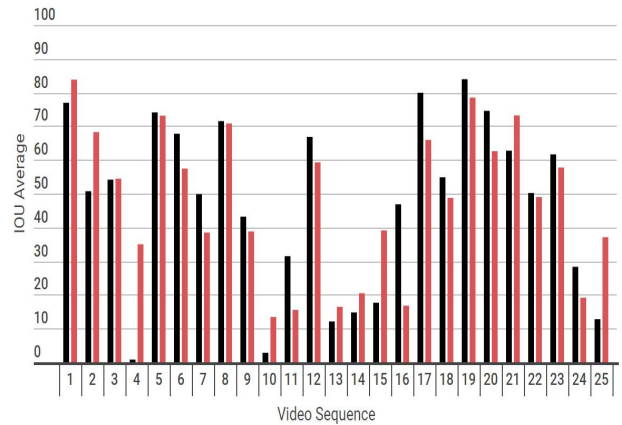
그림 10. 추적 성능 하락 예

Fig. 10. Case of tracking performance decline.

표 1. 영상 시퀀스 별 성능 비교

Table 1. Performance comparison of each video sequence.

Video Sequence	Frame Number	IOU Average(%)		Tracking Success Frame	
		GOTURN	Ours	GOTURN	Ours
1	603	76.93	83.78	596	600
2	725	50.70	68.23	474	631
3	271	54.13	54.40	159	164
4	350	0.78	34.97	2	141
5	252	74.06	73.09	252	242
6	770	67.74	57.42	749	627
7	219	49.83	38.43	81	43
8	1210	71.48	70.77	1105	1079
9	292	43.16	38.77	150	75
10	436	2.80	13.42	5	29
11	310	31.44	15.52	72	51
12	207	66.80	59.27	184	148
13	244	12.09	16.44	16	42
14	267	14.76	20.46	28	52
15	307	17.64	39.09	69	169
16	164	46.82	16.77	98	27
17	371	79.89	65.92	371	348
18	400	54.83	48.73	239	226
19	201	83.91	78.45	201	167
20	172	74.53	62.58	172	163
21	282	62.72	73.18	243	257
22	264	50.17	48.97	172	161
23	569	61.62	57.72	517	406
24	731	28.29	19.12	86	26
25	597	12.76	37.08	110	304



● GOTURN ● Ours

그림 11. 영상 시퀀스 별 IOU 평균 비교

Fig. 11. IOU average comparison of each video sequence.

## VI. 결 론

본 논문에서는 일반적인 객체를 모두 추적할 수 있고 속도가 빠르고 성능이 우수한 CNN 모델 기반 객체 추적기인 GOTURN을 사용하여 다양한 분야에 적용할 수 있는 범용의 객체 추적기를 설계하였다. 그러나 GOTURN은 객체의 위치 정보만을 학습하여 색으로 쉽게 구분할 수 있는 환경에서도 객체 추적을 실패하는 경우가 있었다. 이러한 문제점을 해결하기 위해 컬러 히스토그램 정보로 객체를 추적하는 Mean-shift 추적기와 결합을 제안하였다. 또한 매 프레임 SVM 이진 분류기를 사용하여 두 추적기 중 하나의 추적기를 선택해 기존의 GOTURN과 비교하여 약 1.6%의 성능 향상을 보였다.

## Acknowledgments

본 연구는 경기도의 지역협력연구센터 사업의 일환으로 수행하였음.[GRRC항공2018-B04, 공간/미디어 융합형 인터랙티브 VR 플레이어 및 서비스 개발]

## References

[1] D. Held, S. Thrun, and S. Savarese, "Learning to track at 100 fps with deep regression networks," in *European Conference on Computer Vision*, Netherlands: NL, pp. 749-765, 2016.

[2] C. C. Chang, and C. J. Lin. "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology(TIST)*, Vol. 2, No. 27, pp. 170-197, 2011.

[3] M. Kristan, R. Pflugfelder, A. Leonardis, J. Matas, L.

- Čehovin, G. Nebehay, and A. Dimitriev, "The visual object tracking vot2014 challenge results," [Online] 2014. Available: [http://votchallenge.net/vot2014/download/vot\\_2014\\_paper.pdf](http://votchallenge.net/vot2014/download/vot_2014_paper.pdf).
- [4] M. Kristan, J. Matas, A. Leonardis, M. Felsberg, L. Čehovin, G. Fernandez, et al. "The visual object tracking vot2015 challenge results," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, Chile: CHI, pp. 1-23, 2015.
- [5] U. Kart, J. K. Kämäräinen, L. Fan, and M. Gabbouj, "Evaluation of Visual Object Trackers on Equirectangular Panorama," in *International Conference on Computer Vision Theory and Applications (VISAPP)*, Portugal: POR, pp. 25-32, 2018.
- [6] darkprogrammer. Mean-shift Tracker Image [Internet]. Available: <http://darkpgmr.tistory.com/64>
- [7] D. Comaniciu, V. Ramesh and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference*, Hilton Head Island, South Carolina: SC, pp. 142-149, 2000.
- [8] T. Vojir, N. Jana and M. Jiri, "Robust scale-adaptive mean-shift for tracking," *Pattern Recognition Letters*, Vol. 49, pp. 250-258, 2014.
- [9] bskyvision. SVM data classification Image [Internet]. Available: <https://bskyvision.com/163>
- [10] L. Čehovin, M. Kristan, A. Leonardis, "Is my new tracker really better than yours?" in *2014 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Steamboat Springs, Colorado: CO, pp. 540-547, 2014.



**박 성 준 (Sung-Jun Park)**

2019년 2월 : 한국항공대학교 전자 및 항공전자공학(공학사)  
 2019년 ~ 현재 : 한국항공대학교 전자 및 항공전자공학 석사과정  
 ※관심분야 : 컴퓨터비전, 영상처리



**백 중 환 (Joong-Hwan Baek)**

1981년 2월 : 한국항공대학교 항공통신공학 (공학사)  
 1987년 7월 : 오클라호마주립대학원 전기 및 컴퓨터공학 (공학석사)  
 1991년 7월 : 오클라호마주립대학원 전기 및 컴퓨터공학 (공학박사)  
 1992년 ~ 현재 : 한국항공대학교 항공전자정보공학부 교수  
 ※관심분야 : 영상처리, 패턴인식, 멀티미디어, 가상현실