

Online Multi-Object Tracking by Learning Discriminative Appearance with Fourier Transform and Partial Least Square Analysis

Seong-Ho Lee*, Seung-Hwan Bae*

*M. S. candidate, Dept. of Computer Science and Engineering, Incheon National University, Incheon, Korea

*Assistant Professor, Dept. of Computer Science and Engineering, Incheon National University, Incheon, Korea

[Abstract]

In this study, we solve an online multi-object problem which finds object states (i.e. locations and sizes) while conserving their identifications in online-provided images and detections. We handle this problem based on a tracking-by-detection approach by linking (*or* associating) detections between frames. For more accurate online association, we propose novel online appearance learning with discrete fourier transform and partial least square analysis (PLS). We first transform each object image into a Fourier image in order to extract meaningful features on a frequency domain. We then learn PLS subspaces which can discriminate frequency features of different objects. In addition, we incorporate the proposed appearance learning into the recent confidence-based association method, and extensively compare our methods with the state-of-the-art methods on MOT benchmark challenge datasets.

▶ **Key words:** Vision-based tracking, multi-object tracking, appearance learning, image fourier transform, data association, surveillance system, recognition.

[요 약]

본 연구는 온라인 다중 객체 추적 환경에서 모든 객체의 상태(예. 위치 및 크기) 및 identifications (IDs)를 추적하는 문제를 다룬다. 프레임들 간 검출 결과들을 연관하여 객체들의 궤도를 점진적으로 완성하는 tracking-by-detection 접근법을 기반으로 온라인 다중 객체 추적 문제를 해결하고자 한다. 정확한 온라인 연관을 수행하기 위해 이산 푸리에 변환과 부분 최소 제곱법 (partial least square, PLS) 분석을 기반으로 하는 새로운 온라인 외형 학습 방법을 제안한다. 즉, 먼저 주파수 도메인에서 추적에 용이한 객체 특징량을 추출하기 위해 추적 객체에 대한 이미지를 푸리에 이미지로 변환한다. 나아가 객체간의 주파수 특징을 보다 잘 구별할 수 있도록 PLS기반 부분 공간을 학습한다. 제안된 외형 학습을 최신 신뢰도 기반 연관 기법과 결합하였고, 다중 객체 추적평가 분야에서 국제적으로 공인된 MOT 벤치마크 챌린지 데이터 셋에서 최신 다중 객체 추적 알고리즘과 비교평가를 수행하였다.

▶ **주제어:** 비전 기반 추적, 다중 객체 추적, 외형 학습, 이미지 푸리에 변환, 데이터 연관, 감시 시스템, 인식

-
- First Author: Seong-Ho Lee, Corresponding Author: Seung-Hwan Bae
 - *Seong-Ho Lee (shLee_CV@inu.ac.kr), Dept. of Computer Science and Engineering, Incheon National University
 - *Seung-Hwan Bae (shbae@inu.ac.kr), Dept. of Computer Science and Engineering, Incheon National University
 - Received: 2020. 01. 14, Revised: 2020. 02. 03, Accepted: 2020. 02. 03.

I. Introduction

다중 객체 추적(Multi-Object Tracking, MOT)은 다수의 객체들을 추적하기 위해 검출 결과 간 연관을 수행하는 과정으로 최근 몇 년간 컴퓨터 비전 분야에서 활발하게 연구되고 있는 분야 중 하나이다. 특히 다중 객체 추적기술은 감지 시스템[1], 자율 주행 시스템[2]과 같은 실생활 분야에 기반 기술로 광범위하게 사용되고 있다.

다중 객체 추적 방식은 객체 탐지 결과와의 결합 유무에 따라 크게 detection-free-tracking 방식[3]과 tracking-by-detection 방식[4]으로 구분된다. 최근 Mask-RCNN [5], YOLACT [6] 등 고성능 검출기들이 개발됨에 따라 tracking-by-detection 방식으로 대부분의 연구가 진행되고 있다. 위 방식은 검출된 객체 정보를 기반으로, 각 프레임 간의 동일 객체에 대한 검출 반응을 연관하여 각 객체에 대한 전체 궤도를 완성한다. 따라서 보다 강건한 다중 객체 추적을 위해 다양한 데이터 연관 알고리즘이 개발되었다.

일반적으로 데이터 연관을 위해서는 추적 객체 또는 검출 반응간의 affinity를 평가할 수 있도록 객체 affinity 모델이 필요하다. 다중 객체 추적에서는 최근 appearance (외형), motion (모션), shape (형태) 모델 등을 사용하여 affinity 모델을 설계했다. 특히, appearance 모델은 이미지 기반 객체 추적에서 객체간의 구분을 위해 가장 중요하다. 전통적으로 RGB, HSV, YCbCr 같은 color 히스토그램을 추출하는 방식이 사용되었다. 하지만 이러한 방식은 객체들의 외형이 유사한 색상 분포도를 가질 경우 추적 성능이 저하된다.

이러한 문제를 해결하기 위해, 본 논문에서는 이산 푸리에 변환과 부분 최소 제곱법(partial least square, PLS) 분석을 기반으로 하는 새로운 온라인 외형 학습 방법을 제안한다. 구체적으로, 각 객체에 대한 이미지를 푸리에 이미지로 변환하여, 동일한 주파수 도메인에서 각 객체에 대한 이미지 주파수 패턴(또는 이미지)을 추출한다. 추적 객체들에 대한 주파수 이미지 및 라벨(또는 identification)을 가지고 PLS 부분 공간을 학습하여 객체들 간의 외형 분별력을 극대화 한다.

제한된 외형 학습을 최근 개발된 신뢰도 기반 연관 알고리즘[4]과 결합하여 온라인 다중 객체 추적 시스템을 구현하였다. 국제적으로 공인된 MOT 벤치마크 챌린지 서버에서 다중 객체 추적 성능을 평가하고, 최근 개발된 다중 객체 추적기들과 비교 평가 하였다.

본 논문은 다음과 같이 구성된다. 2장에서는 제안하는 다중 객체 추적 및 외형 학습 기법과 관련된 연구에 대해 기술하고, 3장에서는 본 연구의 핵심인 신뢰도 기반의 데

이터 연관과 분별력 있는 외형 학습 알고리즘을 논한다. 4장에서는 MOT 벤치마크 챌린지 데이터 셋에서, 다양한 외형 알고리즘 적용하여 ablation study를 진행하였고, 최근 개발된 다중 객체 추적기와 비교 평가도 수행하였다. 5장에서는 본 논문의 결론을 도출한다.

II. Related works

최근 tracking-by-detection 방식의 다중 객체 추적 방법은 연관 방식에 따라 크게 배치 추적 방식과 온라인 추적 방식으로 구분된다. 배치 추적 방식[7]은 전체 궤도를 형성하기 위해 전체 프레임에서 검출된 모든 객체 정보에 대해 반복적인 검출 반응 연관을 수행한다. 전체 프레임에 대한 객체 검출 정보를 사용하여 추적 궤도를 형성하기 때문에 온라인 추적 방식보다 좋은 추적 성능을 보인다. 그러나, 사전에 전체 프레임에 대한 객체 검출 정보를 가지고 있어야하기 때문에 실시간 어플리케이션에 적용하기는 어렵다. 이와 반대로, 온라인 추적 방식[4, 8]은 미래 프레임에 대한 정보 없이 과거와 현재 프레임의 객체 검출 정보만을 사용하여 추적 궤도 형성을 위한 검출 반응 연관을 수행하는 방식이다. 따라서, 온라인 추적 방식이 배치 추적 방식에 비해 실시간 어플리케이션에 더욱 적합하다. 온라인 추적 방식은 미래 프레임 정보의 부재로 인하여, 긴 기간의 폐색(occlusion)이나 객체 외형 변화에 대해 취약하며, 추적 궤도 형성을 위한 검출 반응 연관이 어려워지는 문제가 발생한다. 이는 트랙 분리(track fragment) 및 IDs 전환(identity switches) 현상을 발생시켜 추적 성능을 저하시킨다. 이러한 문제를 해결하기 위해 추적 객체 또는 검출 반응간의 구별력 있는 객체 affinity 모델을 생성하는 것은 매우 중요하다.

예측되지 않는 카메라 및 객체의 움직임에 대해서 affinity를 평가하기 위해 추적 객체의 속도 및 가속도를 판단하는 칼만 필터[9] 또는 베이저안 필터[10]를 사용하는 motion 유사도를 평가하는 모델이 제시되었다. 또한, 객체 간 구분을 위해 여러 가지 appearance 모델을 추출하는 방식들이 제시되었다. color 히스토그램[11], 기울기 히스토그램(HoG)[12] 등을 추출하여 외형 유사도를 비교하는 방식과 외형 특징량 추출에 대한 계산 복잡도를 감소시키고 외형 구별 성능을 향상시키기 위해 부분 공간을 학습 하는 방식(subspace learning)[13, 14]을 포함하는 얇은 모델(shallow model) 방식이 제시되었다. 또한, 최근 GPU의 발달과 함께 CNN (convolutional neural networks)[4, 8,

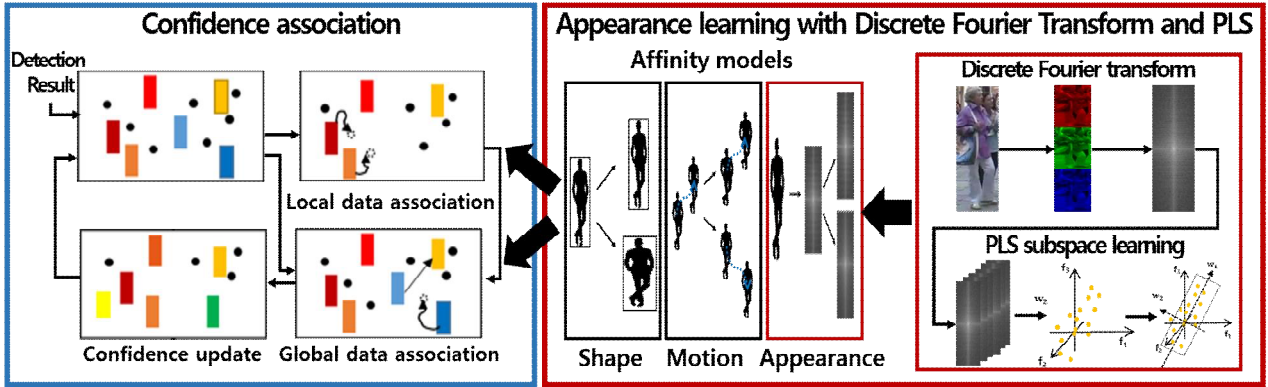


Fig. 1. The overall framework consists of confidence-based data association and appearance learning with Fourier transform and partial least square analysis.

15]을 이용하여 구별력 있는 외형 특징량을 추출하는 깊은 모델(deep model) 기반 방식이 제시되었다.

또한 최근에 푸리에 변환을 이용하여 객체 추적을 진행하는 방식이 제시되었다. [16, 17]은 푸리에 변환을 이용하여 이미지 특징량을 주파수 성분으로 변환하여 외형 특징량 연산에 있어 상관(correlation) 및 컨볼루션(convolution) 연산을 높은 계산 효율성으로 수행하는 단일 객체 추적 방식이 제시되었다.

본 논문은 구별력 높은 appearance 모델을 추출하기 위해 푸리에 변환을 통해 이미지를 푸리에 이미지로 변환하여 주파수 패턴 및 특징을 추출하고, 각 추적 객체에 대한 주파수 특징의 구별력을 향상시키기 위해 PLS 부분 공간을 학습하는 온라인 외형 학습 방법을 제안한다.

III. Methodology

본 연구는 최근 다중 객체 추적에서 적용되고 있는 신뢰도 기반 연관 방식[4]과 본 논문에서 제안한 외형 학습 방법을 결합하였다. 결합된 다중 객체 추적 시스템 구조는 Fig. 1에서 보여준다. Fig. 1의 신뢰도 기반 연관 방식에서 각 박스들은 트랙렛(tracklet)을 의미한다. 신뢰도가 높은 트랙렛은 붉은색 계열로 표시되며 신뢰도가 낮은 트랙렛은 파란색 계열로 표시된다. 또한 검은 점들은 객체 검출 결과를 나타낸다.

3.1 Confidence Association

신뢰도 기반의 연관 방식을 이용한 다중 객체 추적은 트랙렛의 신뢰도에 따라 적응적인 객체 연관을 수행한다. 신뢰도가 높은 트랙렛은 검출 결과와의 지역 연관을 수행한다. 이와 반대로, 신뢰도가 낮은 트랙렛은 다른 트랙렛 또는 검출 결과와의 전역 연관을 수행한다.

3.1.1 Track Confidence

프레임 t 에서의 객체 검출 결과를 $\mathbf{z}_t^i = [z_x, z_y, z_w, z_h]$ 라 나타낸다. 여기서, z_x, z_y, z_w, z_h 는 각각 객체 검출 결과의 x, y 좌표와 너비, 높이를 의미한다. i 번째 트랙렛 K^i 은 다중 객체 추적에서 객체 연관을 통해 연관된 객체 검출 정보 집합으로, 이를 $K^i = \{\mathbf{z}_t^i | v^i(\tau), 1 \leq t_s \leq \tau \leq t_e \leq t\}$ 와 같이 정의할 수 있다. 여기서, $v^i(\tau)$ 는 i 번째 트랙렛의 존재 유무를 나타내며 트랙렛이 존재할 때 1, 그렇지 않으면 0을 출력한다. 또한, t_s^i 와 t_e^i 는 각각 트랙렛의 시작 및 끝 프레임을 나타낸다.

i 번째 트랙렛 K^i 의 신뢰도 $conf(K^i)$ 는 실제 객체 추적 궤도와 비교하여 트랙렛의 추적 궤도가 얼마나 지속적으로 유사하게 추적되고 있는지를 판단하는 값이라고 정의할 수 있다. 따라서, 신뢰도가 높은 트랙렛은 오랜 시간 동안 다른 장애물에 의해 가려지지 않고 추적 궤도를 따라 추적되고 있으며, 트랙렛과 연관된 검출 결과 간의 높은 평균 affinity 점수를 가진다. 이와 반대로 신뢰도가 낮은 트랙렛은 장애물에 의해 빈번히 가려지고 짧은 시간 동안만 추적 궤도를 따라 추적되며, 트랙렛과 연관된 검출 결과 간의 낮은 평균 affinity 점수를 가진다. 이를 고려하여 트랙렛의 신뢰도를 수식으로 정의할 수 있다.

$$conf(K^i) = \left(\frac{1}{L} \sum_{\tau \in [t_s^i, t_e^i], v^i(\tau)=1} \Lambda(K^i, \mathbf{z}_\tau^i) \times (1 - \exp^{-\gamma \cdot \sqrt{L-o}}) \right) \quad (1)$$

i 번째 트랙렛의 신뢰도에 대한 추적 연관 프레임의 길이는 L 이며, o 는 장애물에 의한 폐색에 따른 추적 손실 프레임의 수이며, $o = t_e^i - t_s^i + 1 - L$ 로 정의된다. $\Lambda(K^i, \mathbf{z}_\tau^i)$ 는 트랙렛과 연관된 검출 결과 간의 affinity 점수를 나타내고, γ 는 검출기 성능에 따라 변경되는 신뢰도 조절 파라미터(본 논문에서 1.2로 설정)이다. $\Lambda(K^i, \mathbf{z}_\tau^i)$ 의 평균 정확도 점수 계산식은 수식 (7)에서 보여준다.

다중 객체 추적에서 신뢰도 기반 연관 기법은 해당 트랙렛의 신뢰도를 기반으로 두 가지의 연관 문제를 다룬다.

(p1) 신뢰도가 높은 트랙렛 $K^{i(hi)}$ 이 온라인으로 제공되는 검출 결과와의 연관 (지역 연관)

(p2) 신뢰도가 낮은 트랙렛 $K^{i(lo)}$ 이 다른 트랙렛이나 지역 연관에서 연관되지 않은 검출 결과와의 연관 (전역 연관)

지역 연관(p1)에서 h 개의 신뢰도가 높은 트랙렛 $K^{i(hi)}$ 와 d 개의 τ 프레임에서 온라인으로 제공되는 검출 결과로 구성된 검출결과 셋 $Z_\tau = \{\mathbf{z}_\tau^j\}_{j=1}^{d+1}$ 으로 연관을 수행한다. 지역 연관 점수 행렬 S^{local} 을 정의하면 아래와 같다.

$$S^{local} = [s_{ij}]_{h \times d}, \quad s_{ij} = -\Lambda(K^{i(hi)}, \mathbf{z}_\tau^j), \quad \mathbf{z}_\tau^j \in Z_\tau \quad (2)$$

전역 연관 (p2)에서 l 개의 신뢰도가 낮은 트랙렛 $K^{i(lo)}$ 은 빈번한 폐색으로 인해 트랙 분리가 될 가능성이 크기 때문에, 신뢰도가 높은 트랙렛 $K^{i(hi)}$ 에 비해 여러 가지 상황을 고려하여 연관을 수행한다.

상황 A: l 개의 신뢰도가 낮은 트랙렛 $K^{i(lo)}$ 와 h 개의 신뢰도가 높은 트랙렛 $K^{i(hi)}$ 이 연관되는 경우

$$A = [a_{ij}]_{l \times h}, \quad a_{ij} = -\Lambda(K^{i(lo)}, K^{j(hi)}) \quad (3)$$

상황 B: l 개의 신뢰도가 낮은 트랙렛 $K^{i(lo)}$ 가 추적 종료되는 경우

$$B = \text{diag}[b_1, \dots, b_l], \quad b_i = -(1 - \text{conf}(K^{i(lo)})) \quad (4)$$

상황 C: l 개의 신뢰도가 낮은 트랙렛 $K^{i(lo)}$ 가 η 개의 지역 연관에서 연관되지 않은 검출 결과 y_τ^j 와 연관되는 경우

$$C = [c_{ij}]_{l \times \eta}, \quad c_{ij} = -\Lambda(K^{i(lo)}, y_\tau^j) \quad (5)$$

위 상황 A, B, C에 대해서 고려하여 전역 연관 점수 행렬 S^{global} 을 정의하면 아래와 같다.

$$S^{global} = [s_{ij}]_{(l+\eta) \times (h+l)} = \begin{bmatrix} A_{l \times h} & B_{l \times l} \\ -\theta_{\eta \times h} & C_{\eta \times l} \end{bmatrix} \quad (6)$$

θ 는 연관 임계값(본 논문에서는 0.5로 설정)으로 높은 연관 스코어를 가진 연관 쌍에 대해서만 연관이 수행되도록 한다. 지역 연관 점수 행렬 S^{local} 과 전역 연관 점수 행렬 S^{global} 이 계산된 후 최적의 매칭 쌍을 찾는 헝가리안 알고리즘[18]을 이용하여 객체 간 최적의 매칭 쌍을 찾아 객체 연관을 수행한다. 그 후 수식 (1)을 사용하여, 트랙렛의 신뢰도를 업데이트하며 다중 객체 추적을 진행한다.

3.1.2 Affinity Models

수식 (1) 과 수식 (2), 수식 (6)에서 보이듯이 신뢰도 기반의 다중 객체 추적을 위해 신뢰도 계산과 지역 및 전역 연관을 수행하기 위해서는 비교 객체들 간의 affinity 점수 계산이 필요하다. 해당 수식은 아래와 같다.

$$\Lambda(p, q) = \Lambda^S(p, q) \cdot \Lambda^M(p, q) \cdot \Lambda^F(p, q) \quad (7)$$

p, q 는 검출 결과 또는 트랙렛을 나타내며 이들 간의 비교를 수행하기 위해 객체의 shape 모델 Λ^S 과 motion 모델 Λ^M , appearance 모델 Λ^F 을 각각 고려한다.

Shape 모델 Λ^S : 검출 결과 및 트랙렛의 높이 대비 너비 비율과 같은 객체의 모양을 고려하여, shape affinity 점수를 계산한다. 수식은 아래와 같다.

$$\Lambda^S(p, q) = \exp\left(-\left\{\frac{\hat{z}_h^p - \hat{z}_h^q}{\hat{z}_h^p + \hat{z}_h^q} + \frac{\hat{z}_w^p - \hat{z}_w^q}{\hat{z}_w^p + \hat{z}_w^q}\right\}\right) \quad (8)$$

여기서, \hat{z} 는 칼만 필터에 의해 갱신된 위치 정보를 나타낸다.

Motion 모델 Λ^M : 검출 결과 및 트랙렛의 움직임을 고려하여, 칼만 필터에 의한 위치 예측을 기반으로 전방향 예측과 후방향 예측 위치에 대한 가우시안 분포로 motion affinity 점수를 계산한다.

$$\Lambda^M(p, q) = N(\hat{\mathbf{z}}_{tail}^p + \mathbf{v}_F^p \Phi; \hat{\mathbf{z}}_{head}^q, \mathbf{O}^F) \times N(\hat{\mathbf{z}}_{head}^q + \mathbf{v}_B^q \Phi; \hat{\mathbf{z}}_{tail}^p, \mathbf{O}^B) \quad (9)$$

여기서, $\hat{\mathbf{z}}_{tail}^p$ 은 검출 결과 또는 트랙렛 p 의 최근 갱신된 위치정보를 나타내며, $\hat{\mathbf{z}}_{head}^q$ 는 검출 결과 또는 트랙렛 q 의 처음으로 갱신된 위치 정보이다. 위 두 프레임 격차는 Φ 이다. \mathbf{v}_F^p 는 객체 p 의 계산된 전방향 속도이고, \mathbf{v}_B^q 는 객체 q 의 계산된 후방향 속도를 나타낸다.

Appearance 모델 Λ^F : 검출 결과 및 트랙렛의 외형을 고려하여 appearance affinity 점수를 계산한다.

$$\Lambda^F(p, q) = \max\left(\sum \sqrt{\mathbf{f}_{proj}^p \mathbf{f}_{proj}^q}, 0\right) \quad (10)$$

본 연구는 검출 결과 또는 트랙렛 p, q 에 대한 이미지를 이산 푸리에 변환을 적용하여 푸리에 이미지 $\mathbf{f}_{freq}^p, \mathbf{f}_{freq}^q$ 로 변환한 후 주파수 패턴에 대한 특징을 추출한다. 또한, 구별력 높은 외형 구별력을 위해 주파수 특징의 PLS 부분 공간 학습을 사용하여 가중치 행렬 W_{pls} 을 생성한다. 이를 수식(18)과 같이 해당 가중치 행렬에 푸리에 이미지 $\mathbf{f}_{freq}^p, \mathbf{f}_{freq}^q$ 를 투영하여, 갱신된 주파수 특징량 $\mathbf{f}_{proj}^p, \mathbf{f}_{proj}^q$ 을 생성한다. 해당 appearance 모델 추출 및 학습에 대해서는 3.2에서 자세히 설명하며, 본 연구에서 제안한 appearance 모델을 이용한 실험결과는 4에서 보여준다. 마지막으로, Bhattacharyya 거리 함수를 이용하여 appearance affinity 점수를 계산한다.

3.2 Discriminative appearance learning

본 연구는 구별력 높은 appearance model을 추출하기 위해 이산 푸리에 변환을 이용하여 객체의 이미지를 푸리

에 이미지로 변환하여 객체 이미지에서의 주파수 특징을 추출한다. 또한, 부분 최소 제곱법 기반 외형 구별력 학습을 수행하여 외형 구별 성능을 향상시킨다.

3.2.1 Discrete Fourier Transform

트랙렛 간의 외형 구별을 위해, 해당 트랙렛의 이미지 패치에 대해서 64×32 의 크기로 리사이징을 수행한다. 해당 이미지의 RGB채널에 대한 정보를 잃지 않기 위해 각 채널을 나누고 연결하는 과정을 수행한다. 해당 과정이 수행된 이미지 패치 \mathbf{f}_{img} 는 192×32 의 크기를 가진다. 이미지 패치 \mathbf{f}_{img} 를 푸리에 변환하기 위해서 x 축과 y 축을 둘다 고려해야하기 때문에, 2차원 신호에 대한 이산 푸리에 변환을 수행한다. 해당 변환 식은 아래와 같다.

$$f(x,y) = \sum_{u=0}^{W-1} \sum_{v=0}^{H-1} F(u,v) e^{j2\pi(ux/W+vy/H)} \quad (11)$$

$$F(u,v) = \frac{1}{WH} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} f(x,y) e^{-j2\pi(ux/W+vy/H)} \quad (12)$$

수식 (12)에서 보듯이, x 축 방향과 y 축 방향에 대한 이미지 $f(x,y)$ 를 주파수의 주기함수 성분 계수 $F(u,v)$ 로 변환한다. u, v 는 각각 x 축 방향으로의 주파수 성분 계수와 y 축 방향으로의 주파수 성분 계수로 나타난다. 해당 이산 푸리에 변환으로 얻어지는 주기함수 성분 계수 $F(u,v)$ 는 복소수형태로 아래와 같다.

$$F(u,v) = R(u,v) + jI(u,v) \quad (13)$$

$R(u,v)$ 는 실수영역을 나타내며, $I(u,v)$ 는 허수영역을 나타낸다. 해당 복소수형태로 구성된 주기함수 성분 계수 $F(u,v)$ 에 대해 크기 $|F(u,v)|$ 를 계산하여 푸리에 스펙트럼을 계산하고, 해당 스펙트럼을 벡터화하여 푸리에 이미지 특징량 \mathbf{f}_{freq} 으로 적용한다. 해당 푸리에 이미지 특징량의 크기는 6144×1 이다. 또한, 본 연구에서는 속도 개선을 위해 분할 정복 기법을 사용해 푸리에 변환 연산 복잡도를 줄이는 고속 푸리에 변환을 사용하여 연산 속도를 향상하였다.

3.2.2 PLS Subspace Learning

본 연구는 대표적인 PLS기반의 부분 공간 학습 알고리즘인 NIPALS [19] 방식을 적용하여 다음과 같이 PLS 가중치 행렬 W_{pls} 을 생성한다.

우선, 각 트랙렛에 대해 초기 Δ 프레임(본 논문에서는 30로 설정) 동안의 n 개의 특징 차원(본 논문에서는 6144로 설정)을 가진 푸리에 이미지 특징량 \mathbf{f}_{freq} 을 m 개 포함하는 크기 $m \times n$ 의 푸리에 이미지 특징량 부분 행렬 U 를 생성한다.

$$U = \{\mathbf{f}_{freq}^1, \mathbf{f}_{freq}^2, \dots, \mathbf{f}_{freq}^m\} \quad (14)$$

푸리에 이미지 특징량 부분 행렬 U 에 대해 m 개의 특징 차원을 가진 레이블 스코어 벡터 \mathbf{u} 를 생성하여 수식 (15)을 적용하여 PLS 가중치 벡터 \mathbf{w} 를 생성한다.

$$\begin{aligned} \mathbf{w} &= \frac{U^T \mathbf{u}}{\mathbf{u}^T \mathbf{u}}, \\ \mathbf{w} &= \frac{\mathbf{w}}{\|\mathbf{w}\|} \end{aligned} \quad (15)$$

또한, m 개의 특징 차원을 가지고 있는 특징 스코어 벡터 \mathbf{t} 를 수식 (16)를 적용하여 계산한다.

$$\mathbf{t} = U\mathbf{w} \quad (16)$$

특징 스코어 벡터 \mathbf{t} 와 푸리에 이미지 특징량 부분 행렬 U 를 사용하여 레이블 로딩 값 q 를 계산하고 해당 레이블 로딩 값 q 를 사용하여 레이블 스코어 벡터 \mathbf{u} 를 갱신한다.

$$\begin{aligned} q &= \frac{U^T \mathbf{t}}{\mathbf{t}^T \mathbf{t}} \\ \bar{\mathbf{u}} &= \frac{Uq}{\sqrt{q^T q}} \end{aligned} \quad (17)$$

갱신된 레이블 스코어 벡터 $\bar{\mathbf{u}}$ 와 기존 레이블 스코어 \mathbf{u} 를 비교하여 그 차이가 θ_{pls} (본 논문에서는 1×10^{-5} 로 설정)미만이 될 때까지 수식 (14)-(17)과정을 반복 수행을 한다. 반복 수행 이후 계산된 가중치 벡터 \mathbf{w} 를 저장한다.

위 과정을 τ 회(본 논문에서 60으로 설정) 반복 수행하며 나온 가중치 벡터를 추가하여 PLS 부분 공간 가중치 행렬 $W_{PLS} = \{\mathbf{w}^1, \mathbf{w}^2, \dots, \mathbf{w}^\tau\}$ 을 생성한다.

i 번째 트랙렛 K^i 에 대한 학습된 PLS 부분 공간 가중치 행렬 W_{PLS} 에 i 번째 트랙렛의 푸리에 이미지 특징량 \mathbf{f}_{freq}^i 을 투영하여, 갱신된 푸리에 이미지 특징량 \mathbf{f}_{proj}^i 을 생성한다. 해당 수식은 아래와 같다.

$$\mathbf{f}_{proj}^i = W_{PLS} \mathbf{f}_{freq}^i \quad (18)$$

갱신된 푸리에 이미지 특징량 \mathbf{f}_{proj}^i 는 수식 (10)에서 보듯이 객체들 간의 affinity 점수 계산에 사용된다.

IV. Experiments

4.1 Evaluation Setup

4.1.1 Dataset

다중 객체 추적에서 본 논문에서 제안된 외형 학습 방식의 성능을 평가하기 위해 보행자 추적에 대해 국제적으로 공인된 MOT16 벤치마크 데이터 셋[20]을 사용하였다. MOT16 데이터 셋은 7개의 학습 영상과 7개의 테스트 영상으로 구성되어 있으며 각각의 영상은 다른 각도, 다른

Table 1. Performance comparison with different appearance models on the 2016 MOT challenge benchmark training dataset sequences (MOT16-09, MOT16-10, MOT16-11, and MOT16-13).

Method	Sequence	MOTA ↑	MOTP ↑	FAF ↓	MT ↑	ML ↓	FP ↓	FN ↓	IDS ↓	FG ↓	Hz ↑
(a1) with RGB hist.	MOT16-09	31.18%	73.81%	1.4	8.00%	76.00%	725	2846	47	83	9.70
	MOT16-10	16.72%	73.30%	1.1	1.85%	29.63%	739	9369	151	259	9.78
	MOT16-11	44.74%	78.34%	0.8	11.60%	37.68%	752	4288	30	40	11.96
	MOT16-13	14.63%	73.04%	1.2	4.67%	31.78%	929	8821	25	109	16.26
	Avg.	29.80%	76.87%	1.7	4.31%	56.86%	2802	26088	302	488	11.39
(a2) with discrete fourier transform	MOT16-09	39.13%	74.23%	1.4	20.00%	12.00%	751	2402	47	75	9.93
	MOT16-10	24.54%	74.24%	1.3	1.85%	50.00%	860	8337	98	199	10.78
	MOT16-11	44.96%	78.31%	0.8	11.59%	50.72%	738	4281	30	41	12.73
	MOT16-13	14.58%	72.94%	1.3	4.67%	65.42%	968	8789	24	107	14.85
	Avg.	28.47%	75.38%	1.2	7.45%	52.94%	3317	23809	119	422	12.05
(a3) with discrete fourier transform and PLS	MOT16-09	38.67%	73.64%	2.2	20.00%	8.00%	1162	2016	46	88	3.95
	MOT16-10	31.78%	73.93%	2.0	12.96%	44.44%	1289	7051	63	179	4.19
	MOT16-11	44.64%	78.35%	0.8	10.14%	50.72%	744	4305	29	40	12.88
	MOT16-13	14.42%	72.94%	1.3	4.67%	65.42%	986	8789	24	107	16.80
	Avg.	30.62%	75.05%	1.5	9.41%	51.37%	4181	22161	162	414	7.01

움직임을 가진 카메라로 촬영되었다. 모든 영상에 대해 DPM v5[21] 검출기 기반의 객체 검출 결과를 제공한다. 다중 객체 추적 성능의 공평한 비교를 위해 본 논문은 MOT16 벤치마크 데이터 셋에서 제공하는 검출 결과 및 GT를 이용하여 성능을 평가한다.

4.1.2 Evaluation metrics

Table.1에서 보이듯이, 모든 실험에서 평가를 위해 MOT16 벤치마크에서 사용되는 측정항목들을 사용하였다. 다중 객체 추적 정확도 (MOTA ↑)와 다중 객체 추적 정밀도 (MOTP ↑), 추적된 추적 궤도 비율 (MT ↑), 손실된 추적 궤도 비율 (ML ↓), 트랙 분리 수 (FG ↓), 프레임당 오탐률 (FAF ↓), 거짓긍정 수 (FP ↓), 거짓부정 수 (FN ↓), IDs 전환 수 (IDS ↓), 초당 프레임 속도 (Hz ↑)를 나타낸다. 여기서, ↑는 높을수록 좋은 성능임을 나타내며, ↓는 낮을수록 좋은 성능임을 나타낸다.

4.1.3 System parameters

본 실험에서 사용된 모든 파라미터는 실험을 통해 조정했으며, 모든 실험에서 동일한 파라미터를 적용했다. Fig. 1에서 보이듯이, 제안한 다중 객체 추적 프레임 워크는 신뢰도 기반 연관[4] 과 푸리에 변환과 PLS를 이용한 외형 학습으로 구성되어 있다. 이미지 패치 리사이징 크기 파라미터(본 논문에서는 64×32 로 설정)를 제외한 신뢰도 기반 연관과 관련된 대부분의 파라미터는 [4]와 일치한다. 외형 모델 및 학습과 관련된 파라미터는 n , Δ , τ 가 있고 각각 6144, 30, 60로 설정된다. 제안한 외형 학습 방식의 성능을 확인하기 위해 사용된 컴퓨터 환경은 Intel I5 CPU @3.40GHz, 32.00GB Memory, Windows 10 64bit 조건에서 수행하였고, 사용된 SW들은 Matlab R2019a를 사용하였다.

4.2 Ablation study

신뢰도 기반 다중 객체 추적 시스템에서 본 논문에서 제안한 외형 학습 방식을 평가하기 위해, 다양한 외형 알고리즘을 적용한 신뢰도 기반 다중 객체 추적 시스템을 구현하였다.

- (a1) RGB히스토그램 기반 appearance 모델 적용
- (a2) 푸리에 변환을 적용한 appearance 모델 적용
- (a3) 푸리에 변환과 PLS 외형 학습 방식을 적용한 appearance 모델 적용

다중 객체 추적 시스템 (a1)-(a3)는 외형 알고리즘과 관련된 파라미터를 제외한 대부분의 파라미터는 동일하다. (a1)은 이미지 패치에 대해 각 채널에 따라 48개의 bin을 생성하여, 144차원의 RGB히스토그램을 appearance 모델로 적용한다. (a2)는 이미지 패치 각 채널에 대해 이산 푸리에 변환을 적용하여, 푸리에 이미지 특징량 f_{freq}^i 을 생성하여 6144차원의 appearance 모델로 적용한다. (a3)은 본 논문에서 제안한 방식으로, 수식 (18)에서 보이듯이, PLS 가중치 행렬 W_{PLS} 을 생성하여 푸리에 이미지 특징량 f_{freq}^i 에 투영하여 갱신된 푸리에 이미지 특징량 f_{proj}^i 을 생성하여 60차원의 appearance 모델을 적용한다.

Table. 1은 (a1)-(a3)에 대해 MOT16 학습 데이터 셋 (MOT16-09, MOT16-10, MOT16-11, MOT16-13)에서의 다중 객체 추적 성능을 보여준다. 본 논문에서 제안한 (a3) 방식의 평균 MOTA점수가 (a1)와 (a2) 방식보다 높음을 알 수 있다. 특히, (a1)와 (a2) 대비 (a3)는 MOTA와 MT 성능이 향상되었다. 이는, 푸리에 변환을 이용하여 이미지의 주파수 패턴을 비교하는 푸리에 이미지 기반 appearance 모델이 RGB히스토그램 기반의 외형 모델보다 높은 외형 구별력을 가지면서 객체 추적이 진행됨을 의미한다. (a2)와 (a3)를 비교하였을 때, MOTA와 MT 성능이 향상되고

Table 2. Performance comparison with other MOT systems on the 2016 MOT Challenge benchmark. The results are sorted according to settings and MOTA scores.

Method	Setting	Learning	MOTA \uparrow	MOTP \uparrow	FAF \downarrow	MT \uparrow	ML \downarrow	FP \downarrow	FN \downarrow	IDS \downarrow	FG \downarrow	Hz \uparrow
proposed (FTPLS)	Online	Shallow	38.2%	74.8%	3.2	9.6%	44.0%	18915	93051	689	2006	5.1
Tracker 16 [8]	Online	Deep	54.4%	78.2%	0.6	19.0%	36.9%	3280	79149	682	1480	1.5
CDA_DDA Lv2 [4]	Online	Deep	43.9%	74.7%	1.1	10.7%	44.4%	6450	95175	676	1795	0.5
GMPHD_DAL [22]	Online	Deep	35.1%	76.6%	0.4	7.0	51.4%	2350	111886	4047	5338	3.5
AM_ADM [14]	Online	Shallow	40.1%	75.4%	1.4	7.1%	46.2%	8503	99891	789	1736	5.8
EAMTT_pub [23]	Online	Shallow	38.8%	75.1%	1.4	7.9	49.1%	8114	102452	965	1657	11.8
OVBT [24]	Online	Shallow	38.4%	75.4%	1.9	7.5%	47.3%	11517	99463	1321	2140	0.3
JCmin_MOT [25]	Online	Shallow	36.7%	75.9%	0.5	7.5%	54.4%	2936	111890	667	831	14.8
GMPHD_HDA [26]	Online	Shallow	30.5%	75.4%	0.9	4.6%	59.7%	5169	120970	539	731	13.6
QuadMOT 16 [27]	Batch	Deep	44.7%	76.4%	1.1	14.6%	44.9%	6388	94775	745	1096	1.8
GMMCP [28]	Batch	Shallow	38.1%	75.8%	1.1	8.6%	50.9%	6607	105315	937	1669	0.5
DP_NMS [29]	Batch	Shallow	26.2%	76.3%	0.6	4.1%	67.5%	3689	130557	365	638	5.9

FG 수는 감소한다. 이는, 푸리에 이미징 특징량에 PLS 부분 공간 학습을 적용하여, 갱신된 푸리에 이미징 기반 appearance 모델이 구별력 높은 외형 구별력을 가지므로 폐색 및 외형 변화에 강건해져 트랙 분리 현상을 방지해준다는 것을 알 수 있다. Fig. 2는 본 논문에서 제안한 외형 학습 방식을 제안한 다중 객체 추적과 기존 RGB 히스토그램 기반 appearance 모델 다중 객체 추적의 출력 결과를 보여준다. 본 연구에서 제안된 방식의 트랙렛의 신뢰도가 높게 추적됨을 알 수 있다. 이는, 우리의 외형 학습 방식이 트랙렛의 IDs를 유지하고 트랙 분리 현상을 방지하며 추적이 진행되는 효과가 있음을 보여준다.

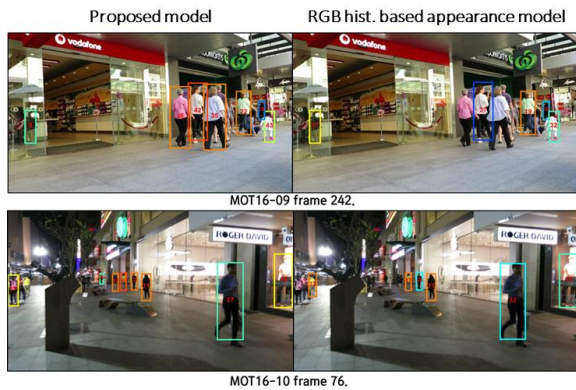


Fig. 2. Tracking results by using our proposed model (left) and RGB histogram based model (right) on the MOT16-09 and MOT16-10 sequences.

4.3 Comparison with SOTA on MOT16 benchmark challenge

본 실험에서는 제안된 외형 학습과 신뢰도 기반 연관 방법을 통합한 다중 객체 추적 시스템과 다른 최신 다중 객체 추적기와의 비교를 수행하기 위해 MOT16 벤치마크 챌린지[20]에 추적 성능을 평가하였다. Table. 2는 본 연구에서 제안된 다중 객체 추적 시스템과 다른 최신 다중 객체 추적기와의 성능 비교표를 보여준다. 해당 성능 비교표는 연관 방식에 따라 온라인 추적 또는 배치 추적으로 추적기를 분류하였고 추가적으로 외형 모델에서 사용하는 appearance model에 따라 얇은 모델과 깊은 모델을 분류하였다. Table. 2에서 보이듯이, 제안된 다중 객체 추적 시스템은 얇은 모델을 사용하는 다른 최신 다중 객체 추적기와 비슷한 추적 성능[14, 23, 24, 28]을 보여주었다. 또한 제안된 다중 객체 추적기는 깊은 모델을 사용하는 다중 객체 추적기[4, 8, 22, 27]보다 빠른 속도를 보였다. 이는 본 논문에서 제안한 외형 학습 방식이 다중 객체 추적에 효과가 있음을 보여주고 있다.

V. Conclusions

본 논문에서는 다중 객체 추적 성능 향상을 위해 이미지를 이산 푸리에 변환을 이용하여 푸리에 이미지로 변환하

여 주파수 패턴을 추출하고 부분 최소 제곱법 분석 방법을 사용하여 객체 외형 구별 성능을 향상하는 방법을 제안하였다. 또한 본 논문에서 제안된 외형 학습 방식을 신뢰도 기반 연관 방법에 적용하여 다중 객체 추적 시스템을 제안하였다. 제안된 외형 학습 방식의 성능을 보여주기 위해, 다양한 외형 알고리즘을 적용한 다중 객체 추적기와 추적 성능을 비교하여 제안된 방식의 외형 구별 성능을 검증하였다. 또한, 국제적으로 공인된 MOT 벤치마크 챌린지 서버에서 성능을 평가하여, 다른 최신 다중 객체 추적기와 비교 평가를 통해 본 논문에서 구현한 다중 객체 추적기의

효용성을 보여주었다.

향후 연구로서, 제안된 appearance 모델의 외형 구별 성능을 향상하기 위해 푸리에 이미지의 노이즈 제거 기법을 적용할 예정이며, 주파수 성분의 특징을 이용하여 다중 객체 추적에서 CNN연산의 계산 효율성을 향상시키는 연구를 진행할 예정이다.



Fig. 3. (a)–(n) Tracking results using the proposed model on the 2016 MOT Challenge dataset.

ACKNOWLEDGEMENT

This work was supported by Incheon National University (International Cooperative) Research Grant in 2019

REFERENCES

- [1] D. Xie, W. Hu, T. Tan, and J. Peng, "A multi-object tracking system for surveillance video analysis," Proceedings of the 17th International Conference on Pattern Recognition, Vol. 4., pp. 767-770, Aug. 2004, DOI: 10.1109/ICPR.2004.1333885
- [2] Y. Ye, L. Fu and B. Li, "Object detection and tracking using multi-layer laser for autonomous urban driving," IEEE International Conference on Intelligent Transportation Systems, pp. 259-264. Nov. 2016, DOI: 10.1109/ITSC.2016.7795564
- [3] K. Fragkiadaki and J. Shi, "Detection free tracking: Exploiting motion and topology for segmenting and tracking under entanglement," IEEE Conference on Computer Vision and Pattern Recognition, pp. 2073-2080, June 2011, DOI: 10.1109/CVPR.2011.5995366
- [4] S. Bae and K. Yoon, "Confidence-Based Data Association and Discriminative Deep Appearance Learning for Robust Online Multi-Object Tracking," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol 40, no. 3, pp. 595-610, Nov. 2018, DOI: 10.1109/TPAMI.2017.2691769
- [5] K. He, G. Gkioxari, P. Dollar, and R. Girshick "Mask R-CNN", Proceedings of the IEEE international conference on computer vision, pp. 2961-2969, Oct., 2017. DOI: 10.1109/iccv.2017.322
- [6] D. Bolya, C. Zhou, F. Xiao, and J. Lee "YOLACT: Real-time Instance Segmentation", Proceedings of the IEEE international conference on computer vision, pp. 9157-9166, Oct., 2019. arXiv:1904.02689
- [7] C. Kim, F. Li, A. Ciptadi, and J. M. Rehg "Multiple hypothesis tracking revisited", Proceedings of the IEEE international conference on computer vision, pp. 4696-4704, Dec. 2015, DOI: 10.1109/iccv.2015.533
- [8] P. Bergmann, T. Meinhardt, and L. Leal-Taixe, "Tracking without bells and whistles", Proceedings of the IEEE international conference on computer vision, pp. 941-951, Oct. 2019. arXiv:1903.05625
- [9] X. Li, K. Wang, W. Wang and Y. Li, "A multiple object tracking method using Kalman filter," The 2010 IEEE International Conference on Information and Automation, Harbin, pp. 1862-1866., June 2010, DOI: 10.1109/ICINFA.2010.5512258
- [10] J. Yoon, M. Yang, J. Lim, and K. Yoon, "Bayesian Multi-Object Tracking Using Motion Context from Multiple Objects", The 2015 IEEE Winter Conference on Applications of Computer Vision, pp. 33-40, Jan. 2015, DOI: 10.1109/WACV.2015.12
- [11] B. Benfold, and I. Reid, "Stable multi-target tracking in real-time surveillance video", IEEE Conference on Computer Vision and Pattern Recognition, pp. 3457-3464, June 2011, DOI: 10.1109/CVPR.2011.5995667
- [12] J. Ge, Y. Luo, and G. Tei, "Real-time pedestrian detection and tracking at nighttime for driver-assistance systems" IEEE Trans. on Intelligent Transportation Systems, vol. 10, no. 2, pp. 283-298, June 2009, DOI: 10.1109/TITS.2009.2018961
- [13] W. Hu, X. Li, W. Luo, X. Zhang, S. Maybank and Z. Zhang, "Single and Multiple Object Tracking Using Log-Euclidean Riemannian Subspace and Block-Division Appearance Model," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 34, no. 12, pp. 2420-2440, Dec. 2012. DOI: 10.1109/TPAMI.2012.42
- [14] S. Lee, M. Kim and S. Bae, "Learning Discriminative Appearance Models for Online Multi-Object Tracking With Appearance Discriminability Measures," IEEE Access, vol. 6, pp. 67316-67328, Nov. 2018. DOI: 10.1109/ACCESS.2018.2879535
- [15] J. Son, M. Baek, M. Cho and B. Han, "Multi-object Tracking with Quadruplet Convolutional Neural Networks," IEEE Conference on Computer Vision and Pattern Recognition, pp. 3786-3795, July 2017. DOI: 10.1109/CVPR.2017.403
- [16] J. F. Henriques, R. Caseiro, P. Martins and J. Batista, "High-Speed Tracking with Kernelized Correlation Filters," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 3, pp. 583-596, Aug. 2014. DOI: 10.1109/TPAMI.2014.2345390
- [17] Z. He, Z. Zhang and C. Jung, "Fast Fourier Transform Networks for Object Tracking Based on Correlation Filter," IEEE Access, vol. 6, pp. 6594-6601, Jan. 2018. DOI: 10.1109/ACCESS.2018.2790942
- [18] F. Bourgeois and J. C. Lassalle, "An extension of the munkres algorithm for the assignment problem to rectangular matrices," Communications of the ACM, vol. 14, no. 12, pp. 802-804, Dec. 1971. DOI: 10.1145/362919.362945
- [19] H. Wold "Path models with latent variables: The NIPALS approach." Quantitative sociology. Academic Press, pp. 307-357, 1975. DOI: 10.1016/b978-0-12-103950-9.50017-4
- [20] A. Milan, L. Leal-Taixé, I. D. Reid, S. Roth, and K. Schindler, "MOT16: A benchmark for multi-object tracking," CoRR, Mar. 2016. arXiv:1603.00831
- [21] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 32, no. 9, pp. 1627-1645, Sep. 2010. DOI: 10.1109/TPAMI.2009.167
- [22] N. L. Baisa, "Online multi-object visual tracking using a

- GM-PHD filter with deep appearance learning,” International Conference on Information Fusion, July 2019, arXiv:1907.13347
- [23] R. Sanchez-Matilla, F. Poiesi, and A. Cavallaro, “Online multi-target tracking with strong and weak detections,” European Conference on Computer Vision, pp. 84-99, Nov. 2016. DOI: 10.1007/978-3-319-48881-3_7
- [24] Y. Ban, S. Ba, X. Alameda-Pineda, and R. Horaud, “Tracking Multiple Persons Based on a Variational Bayesian Model,” European Conference on Computer Vision, ser. Lecture Notes in Computer Science, vol. 9914, pp. 52-67, Oct. 2016. DOI: 10.1007/978-3-319-48881-3_5
- [25] A. Boragule and M. Jeon, “Joint cost minimization for multi-object tracking,” IEEE Conference on Advanced Video and Signal Based Surveillance, pp. 1-6, Aug. 2017. DOI: 10.1109/avss.2017.8078481
- [26] Y. Song and M. Jeon, “Online multiple object tracking with the hierarchically adopted gm-phd filter using motion and appearance,” IEEE Conference on Consumer Electronics-Asia, pp. 1-4. Oct. 2016. DOI: 10.1109/icce-asia.2016.7804800
- [27] J. Son, M. Baek, M. Cho, and B. Han, “Multi-object tracking with quadruplet convolutional neural networks,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3786-3795, July 2017. DOI: 10.1109/cvpr.2017.403
- [28] A. Dehghan, S. M. Assari, and M. Shah, “GMMCP tracker: Globally optimal generalized maximum multi clique problem for multiple object tracking,” IEEE Conference on Computer Vision and Pattern Recognition, pp. 4091-4099, June 2015. DOI: 10.1109/cvpr.2015.7299036
- [29] H. Pirsivash, D. Ramanan, and C. C. Fowlkes, “Globally-optimal greedy algorithms for tracking a variable number of objects,” IEEE Conference on Computer Vision and Pattern Recognition, pp. 1201-1208, June 2011. DOI: 10.1109/cvpr.2011.5995604

Authors



Seong-Ho Lee received the B.S. degrees in computer science and engineering from Incheon National University, South Korea, in 2019. He is currently pursuing the M.S degree with the Department of Computer

Science and Engineering at Incheon National University. His current research interests are multi-object tracking, object detection, super resolution, and deep learning.



Seung-Hwan Bae received the BS degree in information and communication engineering from Chungbuk National University, in 2009 and the MS and PhD degrees in information and communications from the Gwangju

Institute of Science and Technology (GIST), in 2010 and 2015, respectively. He was a senior researcher at Electronics and Telecommunications Research Institute (ETRI) in Korea from 2015 to 2017. He is currently an assistant professor in the Department of Computer Science and Engineering at Incheon National University, South Korea. His research interests include multi-object tracking, object detection, deep learning, dimensionality reduction, medical image analysis, and generative adversarial networks.