Journal of The Institute of Electronics and Information Engineers Vol.58, NO.2, February 2021 https://doi.org/10.5573/ieie.2021.58.2.33
ISSN 2287-5026(Print) / ISSN 2288-159X(Online)

논문 2021-58-2-4

유사도 추정 기반 플렌옵틱 영상 내 단일 객체 추적 기술

(Single Object Tracking in Plenoptic Sequences via Similarity Estimation)

오 희 석*

(Heeseok Oh©)

요 약

단일 객체 추적은 컴퓨터비전의 오래된 기술 분야로써 감시, 국방 및 자율주행을 비롯한 다양한 응용 기술에 활용된다. 최근의 2차원 영상 내 객체 추적 기술들은 Siamese 구조의 심층신경망을 통해 추출된 타겟 객체와 탐색 영역의 특징 간 유사도를 추정함으로써 이루어진다. 이를 통해 추적의 신뢰성과 실시간성 측면에서 전통적인 필터 기반의 객체 추적 기술 대비 비약적인 성능 향상을 이루었으나 여전히 occlusion 발생 시 객체 추적의 빈번한 실패는 명확한 해결방안을 찾기 어려운 고난이도의 기술적 문제점으로 지적되어왔다. 이에 본 논문에서는 플렌옵틱 영상 기반으로 그 특성을 적극 활용하여 occlusion 발생에도 강건한 성능을 보장하는 객체 추적 알고리즘을 제안한다. 복수의 카메라를 통해 렌더링 된 플렌옵틱 영상은 포컬스택으로 표현되며, 서로 다른 초점 영역을 나타내는 다수의 포컬플레인으로 구성된다. 일반적인 2차원 영상과는 달리 플렌옵틱 영상의 포컬스택은 occlusion 발생 시에도 소수의 특정 초점 영역에서 타겟 객체의 추적을 위한 외형 정보를 포함하며 추적의 성공 가능성이 존재한다. 따라서 본 논문에서는 해당 정보의 활용을 위해 Siamese 신경망을 통해 추출된 타겟 객체와 포컬플레인 영상의 고차원 특징 간 유사도를 추정함으로써 객체를 추적하는 플렌옵틱 객체 추적 모델을 구현하였다. 또한, 다수의 포컬플레인 입력으로 인한 오류를 최소화하고자 프레임별로 탐색 영역을 제한하는 알고리즘을 제안한다. 실제 플렌옵틱 영상에 제안하는 알고리즘 적용 시, occlusion 발생의 경우에도 기존 2차원 객체 추적 기술 대비 향상된 성능의 객체 추적이 가능함을 실험적으로 확인하였다.

Abstract

Single object tracking is one of the conventional fields in computer vision, and which is being employed by various applications including surveillance, defense, and autonomous driving. Recent 2D object tracking techniques adopt a similarity estimation between the extracted features from a target object and search regions by feeding them into a Siamese network. Such deep learning based object tracking methods have led the improved performances regarding both robustness and real-time capability, however, tracking the partially or even fully occluded object is challenging, and which is still remained as an insurmountable technical huddle in the related field. In order to resolve this problem, we introduce the novel plenoptic object tracking method guaranteeing the reliable performance when occlusion occurs. A focal stack can be rendered by plenoptic imaging with the calibrated multiple cameras, and which consists of several focal planes representing different focus regions. Differ to general 2D sequences, some of focal planes in a focal stack provide weak appearance information to track the target object stemmed from disparities of the separately located cameras. Thus, we utilize such characteristics to track an object in plenoptic sequences by estimating similarity between the features of target object and focal planes in hyperspace which are captured by a weight-sharing structured network. Additionally, towards minimizing the plenoptic object tracking error mainly caused by an exhaustive search over all focal planes, the adaptive search region restriction algorithm is also proposed. Through applying the proposed plenoptic object tracking scheme, the results show that promising performance can be achieved when even a target object is invisible.

Keywords : Object tracking, plenoptic sequence, similarity estimation, search region restriction

*정회원, 한성대학교 IT융합공학부 (Department of IT Convergence Engineering, Hansung University)
©Corresponding Author(E-mail: ohhs@hansung.ac.kr)
※ 본 연구 논문은 과학기술정보통신부의 출연금으로 수행하고 있는 한국전자통신연구원 “중대형 공간용 초고해상도 비정형 플렌옵틱 동영상 저작 재생 플랫폼 기술 개발 (2020-0-00457)” 위탁연구과제의 연구결과입니다.
Received ; December 16, 2020 Revised ; January 7, 2021 Accepted ; January 14, 2021

Ⅰ. 서 론

플렌옵틱(plenoptic) 영상 기술은 복수 카메라의 어레이로 구성된 촬영기기를 통해 실 공간 내에서 임의의 방향으로 진행하는 빛의 공간 및 각도 정보를 기초로, 광선의 방향까지 활용해 영상을 재구성함으로써 다시점 영상 생성, 재초점 설정 및 3차원 깊이정보 추출 등이 달성 가능한 차세대 영상 처리 기술이다[1]. 어레이 기반의 캘리브레이션이 완료된 복수 카메라로 획득한 다시점 2차원 영상들은 광학적 특성을 십분 활용하여 서로 다른 초점 영역을 갖는 다수의 포컬플레인(focal plane) 영상들로 층층이 구성된 포컬스택(focal stack)으로 렌더링 될 수 있다. 이러한 포컬스택은 그 특성 상 방대한 양의 초점 및 깊이 정보를 포함함에 따라 일반적인 2차원 영상 대비 다양한 차세대 콘텐츠 응용 분야에 활용이 가능하다. 기존 플렌옵틱 영상 관련 기술은 하드웨어 측면에서 획득 기기 및 재생 단말 기기의 성능 향상을 위한 연구가 주를 이루었으나, 최근에는 플렌옵틱 영상 기반의 콘텐츠 제작 및 스트리밍을 위한 압축 기술 및 가시화에 관한 연구로 활용 분야의 범위가 대폭 확대되었다. 특히 컴퓨터비전 분야에서도 플렌옵틱 영상을 활용한 다양한 응용 사례가 선보이고 있으며, 객체 검출 및 분류에서의 우수한 성과가 보고되었다[2, 3].

반면, 전 세계적으로 활발하게 연구 중인 단일 객체 추적의 경우, 자율주행, 국방 및 감시 분야에 필수적인 요소 기술임에도 불구하고 현존하는 대부분의 알고리즘이 단일 2차원 혹은 다시점 2차원 영상 기반의 기술로 국한되어 있다. 따라서 비정형 객체, 급작스러운 움직임, 조명과 배경의 변화, 그리고 occluder에 의해 타겟 객체가 가려지는 occlusion을 포함하는 영상과 관련하여 여전히 고신뢰 추적이 불가한 상황이며, 이의 해결방안이 쉽게 도출되기 힘든 고난이도의 기술적 한계가 존재한다[4].

본 논문에서는 특히 기존 2차원 영상 기반 객체 추적 기술에서 극복이 요원했던 occlusion 발생 시의 추적 실패에 따른 한계 극복을 위해 플렌옵틱 영상 시퀀스를 활용하여 고성능의 객체 추적을 보장하는 기술을 제안한다. 그림 1에서는 플렌옵틱 영상 시퀀스를 활용한 객체 추적의 장점을 가시적으로 보여준다. 기존 2차원 영상 기반 객체 추적 기술의 경우, 타겟 객체가 occluder에 의해 가려질 시에 객체의 외형을 비롯한 어떠한 특징 정보도 존재하지 않으므로 기존의 알고리즘을 통해서는 객체 추적이 불가능 하다. 반면 플렌옵틱 시퀀스의 경우에는 서로 다른 초점 정보를 갖는 다수의 포컬

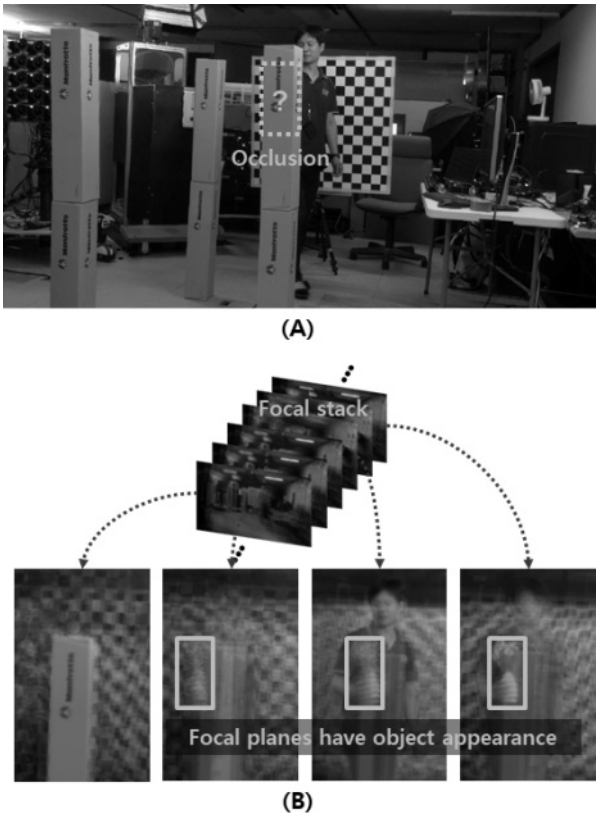

그림 1. Occlusion 발생 시 2차원 영상 객체 추적과 플렌옵틱 영상 객체 추적 비교 (A) 2차원 영상 내 타겟 객체 정보 부재 (B) 플렌옵틱 영상 내 약한 외형 정보 존재

Fig. 1. Comparison of 2D and plenoptic object tracking when occlusion occurs. (A) Target object is invisible in 2D sequence. (B) Weak appearance of target object exists in plenoptic sequence.

플레인 영상으로 이루어진 포컬스택 전체를 프레임 단위로 객체 추적에 활용 가능하다. 즉, 복수의 카메라 간 시차를 반영하여 렌더링 된 포컬스택 내 특정 포컬플레인에서는 occluder에 의해 가려진 타겟 객체의 추적을 위한 외형적 단서가 미약하게나마 존재하므로 최적의 초점 영역을 검출하여 이를 기준으로 적절한 탐색이 수행된다 가정할 때, occlusion 발생에도 불구하고 성공적인 객체 추적 달성의 가능성이 존재한다.

본 논문에서는 이와 같은 플렌옵틱 영상의 특성을 적극적으로 활용하여 포컬스택 내 다수의 포컬플레인 영상에 대해 심층신경망을 통해 추출된 특징(feature) 간 유사도(similarity)를 계산함으로써 occlusion 발생에도 불구하고 강인한 성능을 보이는 플렌옵틱 객체 추적 알고리즘을 제안한다. 더불어 추가적으로 본 플렌옵틱 객체 추적의 성능 향상을 위해 프레임 별 탐색 영역을 제한함으로써 해당 탐색 영역 내 선택적인 유사도 계산을 통한 신뢰도 높은 추적 결과 도출 방법을 고안하였다.

Ⅱ. 선행 연구 및 한계

1. 2차원 영상 내 단일 객체 추적

고전적인 실시간 객체 추적 기술은 주로 영상 내 추적을 위한 객체의 특징을 별도로 추출한 후, 상관 필터(correlation filter)와 같은 특정 커널을 활용하여 응답 에너지가 최대가 되는 영역을 탐색하는 방식으로 동작하였다[5]. Optical flow를 비롯한 컬러 히스토그램, 실루엣, SIFT(scale invariant feature transform), 혹은 HOG(histogram of gradient)와 같은 기법을 주로 활용하던 기존의 객체 추적은 심층신경망의 등장과 함께 자동화된 특징의 학습 형태로 발전하였다[6]. 하지만 객체 추적은 여타 컴퓨터비전 분야와는 달리 긴 영상 시퀀스에 대한 일반화된 표현 학습의 난해함으로 인하여 상대적으로 심층신경망이 적극적으로 활용되지 않는 경향이 두드러졌다. Nam 등은 영상 시퀀스에 대한 일반적 표현을 우선적으로 학습하고, 이후 종단의 전연결층은 온라인 학습을 통해 특정 영상에 과적합시키는 형태의 MDNet[7]을 제안함으로써 심층신경망 기반 객체 추적의 우수한 성능을 증명하였으나, 온라인 학습의 특성상 실시간성을 보장하기 어렵다는 단점이 존재하였다. 이에 Held 등은 비교적 단순한 feed-forward 신경망을 구성하여 외형과 움직임에 강인한 일반적인 표현을 학습함으로써 바운딩박스(bounding box) 좌표 예측이 가능한 GOTURN[8]을 제안하였다. GOTURN은 타겟 객체와 탐색 영역을 각각 가중치를 공유하는 신경망에 입력하여 특징들을 추출하고 이들을 활용하여 전연결 신경망이 직접적으로 좌표를 추론하는 방식으로, 실시간 동작과 동시에 높은 성능을 보장함으로써 이후 유사도 기반의 객체 추적 기술 발전에 영감을 주었다.

2. 유사도 학습 기반 객체 추적

단순 클래스의 분류를 위함이 아닌 지표 학습(metric learning)을 목적으로 하여 Siamese 신경망은 기존부터 널리 이용되어왔다. Siamese 신경망은 가중치를 공유하는 두 개의 신경망에서 추출된 특징들의 특징 공간 내 거리 계산을 통해 두 입력이 같은 클래스인지 여부를 판별함으로써 one-shot 학습 및 얼굴 검증과 같은 분야에서 우수한 성능을 보였다[9, 10].

Siamese 신경망을 이용해 타겟 객체와 탐색 영역 간 거리 계산을 통한 객체 추적으로의 적용이 꾸준히 시도되었고, Bertinetto 등은 Siamese 신경망 구조를 활용하여 유사도를 계산하는 객체 추적 기술인 SiamFC를 발표하였다[11]. 그림 2는 SiamFC의 구조를 나타내며 z는 타겟 객체 영상, x는 탐색 영역 영상을 의미한다. 콘볼루션만으로 이루어진 신경망 φ을 통해 z와 x가 각각 특징맵으로 임베딩되며 두 특징맵 $\varphi(z)$과 $\varphi(x)$ 간 2차원 교차상관(cross-correlation) 연산을 통해 최종적으로 유사도를 계산한 후, 가장 높은 유사도 값을 가지는 위치를 추적 객체로 결정한다.

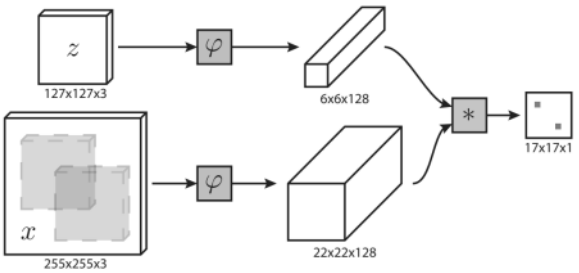

그림 2. SiamFC에서의 Siamese 신경망 구조 및 유사도 계산[11]

Fig. 2. Visual object tracking based on similarity estimation by employing Siamese architecture[11].

SiamFC이 보여준 성공적인 실시간 고신뢰 추적 성능에 기인해 Siamese 구조를 활용한 유사도 기반의 객체 추적 기술이 잇따라 발표되었다. Valmadre 등은 기존 SiamFC의 구조에 온라인 학습을 통한 상관필터의 튜닝 과정이 추가된 경량의 추적 모델 CFNet을 발표하였다[12]. He 등은 타겟 객체의 급격한 외형 변화에 불변성을 지닐 수 있도록 유사도 뿐 아닌 고차원 특징의 의미론적 비교까지 수행하는 객체 추적 모델인 SA-Siam을 개발하였다[13]. Li 등은 객체 검출 모델인 Faster R-CNN[14]에서 활용된 RPN(region proposal network)을 추가하여 객체와 배경의 분류 및 바운딩박스 예측을 동시에 수행하는 SiamRPN[15]을 제안하였고, 이듬해 특징 채널 별 교차상관 연산의 도입과 더불어 객체 위치 정보의 상실 없이 보다 깊은 구조의 백본 신경망을 이용할 수 있도록 모델을 수정하여 객체 추적의 성능을 높인 SiamRPN++[16]를 발표하였다. Wang 등은 multi-task 학습을 통해 객체 추적의 성능을 향상시킴과 동시에 객체 분리까지 수행하는 SiamMask[17]를 제안하였으며, Guo 등은 RPN기반 구조 내 앵커(anchor)의 스케일 및 종횡비 결정과 튜닝 과정이 사용자의 경험에 지나치게 의존적임을 지적하고 이를 해결하고자 픽셀 단위로 분류 및 회귀를 수행하는 SiamCAR[18]를 제안함으로써 객체 추적 분야에서의 우수한 성능을 달성하였다.

Siamese 신경망 기반의 2차원 영상에 대한 객체 추적 기술은 현재까지 전 세계적으로 활발히 연구되고 있

는 주제이나 전통적으로 지적되었던 occlusion을 비롯한 비정형 객체 추적, 급작스러운 조명 및 배경 변화를 포함하는 영상에 대해 여전히 기술적 해결을 요하는 문제점들이 존재한다.

3. 기존 플렌옵틱 영상 내 객체 추적

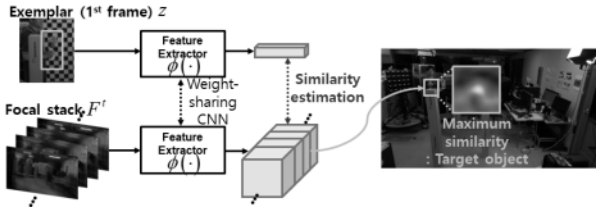

그림 3. 플렌옵틱 영상 시퀀스에서의 유사도 계산을 통한 객체 추적 개념도

Fig. 3. Framework of the proposed plenoptic object tracking based on similarity estimation.

현존하는 대부분의 객체 추적 알고리즘은 2차원 영상 혹은 다중 카메라 기반의 다시점 2차원 영상을 위한 기술이며, 서로 다른 초점영역의 포컬플레인이 입력되는 플렌옵틱 시퀀스의 특성을 면밀히 활용한 객체 추적 기술 개발은 거의 진행되지 않은 상태이다.

Kim 등은 입력되는 포컬스택에 대하여 가장 선명한 포컬플레인이 추적하고자 하는 객체를 포함한다는 가정하에 Sum-modified-Laplacian 연산을 수행하여 선명도를 계산하였다[19]. 가장 높은 선명도를 보이는 포컬플레인을 매 프레임별로 선택하여 하나의 2차원 시퀀스로 재구성한 후 AdaBoost 알고리즘을 통해 객체를 추적하는 방식을 제안하였다. Bae 등은 포컬스택 내 추적 대상이 되는 객체에 초점이 명확한 포컬플레인은 소수 존재하므로, 해당 포컬플레인 후보군의 이미지 선명도 향상 필터 연산을 통해 보다 정확한 객체 추적이 진행될 수 있도록 알고리즘을 개선하였다[20]. 또한, 객체의 3차원 이동에 따른 크기 변화 반영을 위하여 일정 비율로 바운딩박스의 스케일을 조절하는 기술을 제안하였다. 이후 Bae 등은 단순히 선명도 기준이 아닌, ImageNet으로 기 학습된 VGG16 모델을 통해 타겟 객체 영상과 탐색 영역으로부터 추출된 특징벡터 간 코사인 유사도를 비교함으로써 콘텐츠의 특성까지 고려해 포컬플레인을 선택하는 알고리즘을 개발하였다[21].

이렇듯 기존 플렌옵틱 시퀀스 내 객체 추적을 위한 접근 방식은 선명도와 같은 특정 수치를 토대로 한 최적의 포컬플레인 선택에 집중하였다. 이는 곧 프레임별 선택된 포컬플레인을 취합한 새로운 2차원 영상 시퀀스 구성 후 2차원 객체 추적 알고리즘을 통해 영역을 탐색하는 방법과 동일하며, 배경이 선명하거나 occluder가 선명할 경우 추적하고자 하는 타겟 객체와 전혀 무관한 초점 영역을 갖는 포컬플레인이 선택될 가능성이 높다는 문제점이 존재한다. 또한 end-to-end 프로세스가 아닌 포컬플레인의 선택과 객체 추적이 별도로 구성됨으로 인해 최적의 포컬플레인 선택 시에는 추적을 위하여 추출한 객체의 특징이 효과적으로 반영되지 못한다는 단점을 갖는다.

Ⅲ. 제안 기술

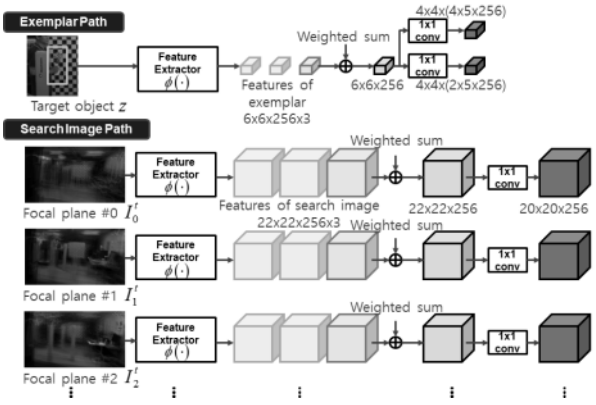

그림 4. 플렌옵틱 객체 추적 베이스라인에서의 특징 추출

Fig. 4. Feature extraction for plenoptic object tracking in the baseline model.

본 논문에서는 그림 3에서와 같이 객체 추적 알고리즘에 포함된 Siamese 신경망 기반의 유사도 계산을 최적의 포컬플레인 선택부에 도입하여 포컬플레인 선택과 객체 추적이 별도로 동작하지 않도록 설계하였다. z는 타겟 객체 영상(exemplar)을 의미하며 F^t는 t번째 프레임에서의 포컬스택이다. F^t는 K개의 서로 다른 초점영역을 갖는 k번째 포컬플레인 영상 I_k^t으로 구성되어 있으며 ($F^t = \{I_k^t \mid k = 1,\ldots,K\}$), 가중치를 공유하는 CNN $\phi(\bullet)$를 통해 z를 비롯한 F^t 내 모든 I_k^t에 대하여 특징을 추출한다. 해당 특징 간 상관연산을 통해 유사도를 계산하며, 가장 높은 유사도를 갖는 영역을 객체로 판단한다. Ⅱ.2에서 서술한 Siamese 신경망 활용 특징 기반 유사도 계산은 이렇듯 포컬스택 내 찾고자 하는 객체와 가장 유사한 특징 영역을 정량적으로 탐색 가능하게 함에 따라 다수의 초점영역을 포함한 플렌옵틱 영상 시퀀스에 쉽게 이식 가능하다는 장점이 존재한다. 또한 이는 객체 추적에 필요한 특징을 활용하

여 다수의 초점영역에 대해 탐색을 수행함과 동시에, 기존에 연구되어왔던 최적의 포컬플레인 선택 기법들과 상호보완적으로 동작 가능함으로써 추적의 오류를 최소화 할 수 있음을 의미한다.

1. 제안 플렌옵틱 객체 추적 베이스라인 (Baseline)

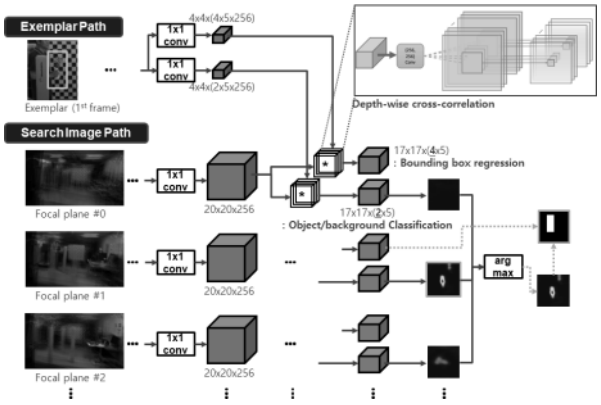

그림 5. 플렌옵틱 객체 추적에서의 유사도 계산을 통한 바운딩박스 좌표 예측 및 객체 분류

Fig. 5. Bounding-box regression and object classification through similarity estimation for plenoptic object tracking.

본 연구에서 제안하는 플렌옵틱 객체 추적 베이스라인 모델은 2차원 객체 추적 데이터셋에 의해 기 학습된 SiamRPN++[16] 기반으로 구성되었다. 그림 4에서와 같이 첫 프레임에서 사용자가 추적하고자 하는 객체를 2차원 영상 위에서 지정한 뒤, 이후 프레임에 대해 포컬스택 내 모든 포컬플레인을 대상으로 특징을 추출한다. 특징 추출을 위한 백본 신경망은 ResNet50 모델을 이용하였으며 종단의 전연결을 위한 벡터형태의 특징이 아닌, 백본 내 세 층의 특징을 일차적으로 추출하고 이들의 가중합 결과를 유사도 계산을 위한 최종 특징으로 활용한다. 이때 각 특징의 가중치는 1로 통일하였다[16]. 타겟 객체 영상 z에서는 4x4x(4x5x256) 크기의 특징과 4x4x(2x5x256) 크기를 갖는 특징이 추출된다. k번째 포컬플레인 I_k^t에 대해 20x20x256 크기의 특징을 추출하며, 따라서 전체 포컬스택 F^t에 대해 K개의 특징이 추출된다.

그림 5에서는 추출된 특징 간 유사도 계산에 기반한 객체 추적을 나타내었다. z와 I_k^t에서 추출된 특징들은 1x1 콘볼루션 연산을 통해 정제되며, 백본에 연이어 연결된 RPN으로 입력된다. 최종 특징 간 채널별 교차 상관(depth-wise cross-correlation) 연산을 수행함으로써 앵커 별 5개의 바운딩박스 표현을 위한 각 4개의 좌표 (x,y,w,h) 정보를 포함하는 17x17x(4x5) 크기의 맵이 예측되며, 앵커 별 5개의 바운딩박스에 대해 객체와 배경의 분류를 수행하는 17x17x(2x5) 크기의 유사도 점수 맵 R^t이 계산된다. 최종적으로는 해당 유사도 점수 맵 R^t에서 가장 높은 값을 갖는 바운딩박스 번호에 해당하는 좌표값을 객체의 위치로 지정함으로써 추적이 이루어지도록 구현하였다.

2. 포컬플레인 탐색 영역 제한

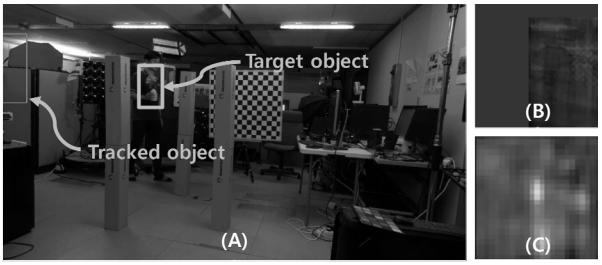

그림 6. (A) 베이스라인 모델에서의 플렌옵틱 객체 추적 실패 (B) 선택된 포컬플레인 (C) 최대 유사도 점수 맵.

Fig. 6. (A) Failure case of plenoptic object tracking when using the baseline model. (B) Selected focal plane. (C) Maximum similarity score map.

베이스라인 모델을 통해 타겟 객체 영상을 기반으로 포컬스택 내 모든 포컬플레인에 대해 유사도를 계산하여 최대 유사도 영역을 추적 객체로 판단하였을 시, 다수 영상에서 포컬스택으로의 재구성 렌더링으로 인해 객체 잔상이 여러 포컬플레인에 걸쳐 존재하므로, 실제 객체와는 무관한 위치로 추적되는 문제점이 발생한다. 또한 방대한 영상 영역에 대해 유사도를 계산함으로 인해 오추적 확률이 증가하며, 계산량 측면에서도 단점을 갖는다. 그림 6에서는 베이스라인 모델 이용 시의 플렌옵틱 객체 추적 실패 영상과 해당 프레임에서 선택된 포컬플레인 및 유사도 점수 맵을 가시화하였다. 타겟 객체와는 상이한 포컬플레인에서 최대 유사도 점수가 계산됨으로 인해 추적이 실패하며, 이는 Siamese 신경망 기반의 포컬스택 내 모든 초점영역에 대한 유사도 계산만을 활용할 시의 한계점을 나타내며, 객체 추적을 위한 포컬플레인 탐색 영역의 제한이 가능한 별도의 알고리즘이 요구됨을 의미한다.

따라서 본 논문에서는 프레임 별 포컬플레인 탐색 영역을 제한하는 알고리즘을 제안한다. 그림 7에서는 프레임 별 포컬플레인 탐색 영역 제한 방법을 도식화하였다. $t-1$번째 프레임의 포컬스택 F_r^{t-1}에서 유사도

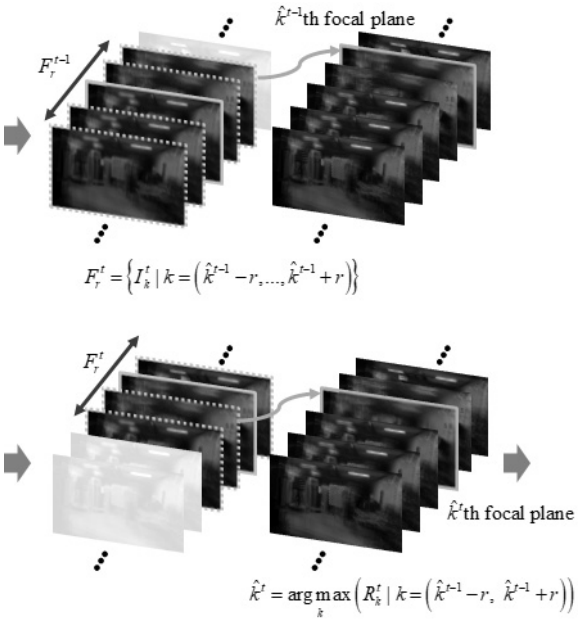

그림 7. 프레임 별 포컬플레인 탐색 영역 제한 알고리즘
Fig. 7. Search region restriction algorithm at each frame.

점수 계산을 통해 추적 객체를 결정하며, 최대 유사도 점수를 갖는 포컬플레인 $\hat{k}^{t-1}$이 선택된다. t번째 프레임에서는 $\hat{k}^{t-1}$ 포컬플레인을 중심으로 탐색 영역이 r개의 포컬플레인 영상으로 제한된 새로운 포컬스택 $F_r^t = \left\{ I_k^t | k = \left(\hat{k}^{t-1} - r, \ldots, \hat{k}^{t-1} + r \right) \right\}$이 구성되어, 해당 F_r^t에서의 유사도 R_r^t 계산을 통해 $\hat{k}^t = \arg max_k \left\{ R_k^t | k = \left(\hat{k}^{t-1} - r, \ldots, \hat{k}^{t-1} + r \right) \right\}$번 째 포컬플레인에서의 객체를 추적하는 방식을 도입하였다.

본 방법은 곧 $t-1$프레임에서의 타겟 객체 영상과 가장 높은 유사도를 갖는 포컬플레인을 중심으로 이후 t프레임에서의 포컬스택 내 후보 그룹을 재구성하여 특징을 추출함으로써 보다 정확한 객체 추적을 도모하고자 함이며, r의 값이 매우 클 경우에는 베이스라인 모델과 동일하게 동작하고 r의 값이 매우 작을 경우에는 2차원 영상에서의 객체추적과 유사하게 동작한다. 본 논문에서의 탐색 영역은 실험적으로 $r = 3$으로 결정하였고, 따라서 매 프레임 별 총 7개의 포컬플레인 영상을 이용해 객체를 추적한다.

3. $t = 1$에서의 포컬플레인 탐색 영역 선정

Ⅲ.2절에서의 포컬플레인 탐색 영역 제한은 직전 프레임에서의 유사도를 기반으로 이루어진다. 이 경우 문제가 되는 것은 $t = 1$에서는 어떠한 방식으로 포컬스택 내 탐색 영역을 제한할 것인가에 관한 것으로, $t = 1$에서의 포컬스택 내 모든 포컬플레인에 관해 유사도를 계산할 경우 베이스라인 모델이 가지는 단점을 해결하지 못하며, 또한 직전 프레임이 존재하지 않기에 탐색 영역 제한 알고리즘이 동작할 수 없어 적절한 탐색 영역을 다른 접근으로 선정할 필요가 있다.

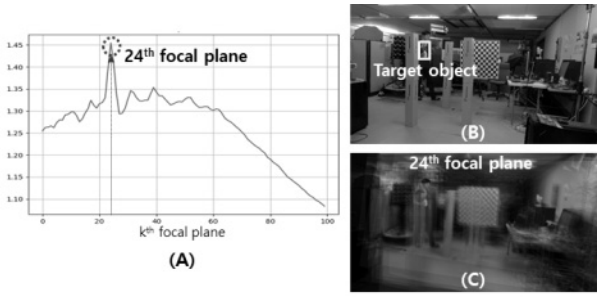

그림 8. (A) 포컬플레인 별 선명도 계산 결과 (B) 타겟 객체 영상 (C) $t = 1$에서 가장 높은 선명도 값을 갖는 포컬플레인
Fig. 8. (A) Calculated sharpness for each focal plane. (B) Target object image. (C) Focal plane having the maximum sharpness at $t = 1$ frame.

본 논문에서는 기존 플렌옵틱 영상 시퀀스 내 선명도 기반의 최적의 포컬플레인 선택을 위한 기술을 $t = 1$에 대하여 적용하며, 간단하게 픽셀 간 그래디언트 크기 $G_x(I_k^1)$와 $G_y(I_k^1)$를 활용하여 k번째 포컬플레인 영상 I_k^1의 선명도 $S(I_k^1)$를 정량적으로 계산하였다.

$$\begin{aligned} S(I_k^1) &= \left| G_x(I_k^1) \right| + \left| G_y(I_k^1) \right|, \\ G_x(I_k^1) &= I_k^1(x,y) - I_k^1(x+1, y+1), \\ G_y(I_k^1) &= I_k^1(x+1,y) - I_k^1(x, y+1). \end{aligned} \quad (1)$$

가장 높은 $S(I_k^1)$ 값을 갖는 $\hat{k}^1$번째 포컬플레인을 기준으로 r개의 포컬플레인으로 제한된 새로운 포컬스택 $F_r^1 = \left\{ I_k^1 | k = \left(\hat{k}^1 - r, \ldots, \hat{k}^1 + r \right) \right\}$을 구성하며, 이후 프레임부터 프레임 별 포컬플레인 탐색 영역 제한 알고리즘이 동작함으로써 객체를 추적한다. 그림 8에서는 $S(I_k^1)$ 계산을 통해 $t = 1$프레임에서의 포컬플레인 선택 결과 예시를 나타내었다. 해당 플렌옵틱 영상 시퀀스에서 가장 높은 선명도 값을 갖는 포컬플레인의 경우, 추적하고자 하는 타겟 객체를 기준으로 명확한 초점 영역이 형성되어 있음을 확인할 수 있다. 눈여겨 볼 점은 이러한 선명도 기반의 포컬플레인 선택은 사용자가 지정한 타겟 객체 직후 프레임인 $t = 1$에서만 적용 가능한 기술로써, 사용자 역시 어떠한 occlusion도 존재하지 않는 영상에서만 육안으로 타겟 객체 지정이 가능하기 때문이다. 이후 프레임부터 $S(I_k^1)$값을 객체 추적

에 이용하였을 시에는 occluder가 선명한 포컬플레인이 선택되어 추적의 오류가 누적된다.

Ⅳ. 실험 결과

1. 객체 추적 검증용 플렌옵틱 영상

제안하는 플렌옵틱 영상 객체 추적 알고리즘의 성능 평가를 위하여 두 가지 플렌옵틱 영상을 촬영해 활용하였다. V3 시퀀스는 25개의 카메라가 동일 간격의 5x5 어레이 형태로 구성되어 촬영한 정형 플렌옵틱 영상이며 이를 통해 프레임 별 100장의 포컬플레인 영상을 포함하는 포컬스택으로 렌더링되었다. NV4 시퀀스는 10개의 카메라가 정해진 규격 없이 임의의 위치에서 시차를 가지고 촬영한 비정형 플렌옵틱 영상으로, 프레임 별 101장의 포컬플레인 영상을 포함하는 포컬스택으로 구성되어있다. 두 시퀀스 모두 객체의 종횡 및 깊이 이동과 동시에 occlusion 발생을 포함하고 있으며, 테스트 시퀀스의 구체적인 사양은 표 1에 나타내었다.

표 1. 실험에 이용된 플렌옵틱 시퀀스
Table1. Tested plenoptic sequences for object tracking.

시퀀스 이름	*V3*	*NV4*
프레임 수	150	280
카메라 수	25	10
포컬플레인 수	100	101
형태	정형 플렌옵틱	비정형 플렌옵틱
해상도	FHD (1920x1080)	

본 실험에서 객체 추적을 위해 이용한 플렌옵틱 영상 시퀀스는 공개된 데이터셋이 아님에 따라 객체 추적의 정확도 평가를 위한 ground-truth가 존재하지 않으므로 추적 객체에 대한 ground-truth 제작이 요구된다. 이에 프레임 별 추적 대상 객체의 (x,y,w,h)의 어노테이션을 수작업으로 진행하였으며, 해당 바운딩박스 좌표 정보를 별도로 저장하여 성능 검증에 활용하였다.

2. 플렌옵틱 객체 추적 성능

객체 추적의 성능은 추적하고자 하는 타겟 객체를 중심으로 그려지는 바운딩박스의 좌표를 기준으로 계산되며, 그림 9과 같이 크게 두 가지의 지표가 널리 활용된다[19]. 첫 번째 지표는 추적 거리(distance)로써, 두 바운

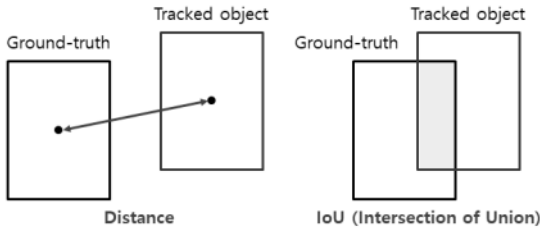

그림 9. 객체 추적 성능 평가 지표: 거리 (좌), IoU (우)
Fig. 9. Metrics to measure the tracking performance: distance (left) and IoU (right).

딩박스의 중심 좌표를 기준으로 측정한다. 즉, 추적하고자 하는 객체의 ground-truth 중심좌표 (x_{GT},y_{GT})와 플렌옵틱 객체 추적의 결과로 도출되는 바운딩박스의 중심좌표 $(\hat{x},\hat{y})$ 간 유클리드 거리 계산을 통해 수행된다.

$$distance=\sqrt{(x_{GT}-\hat{x})^2+(y_{GT}-\hat{y})^2}. \quad (2)$$

즉, 객체 추적의 오차를 픽셀 단위의 거리로 표현하며, 일반적으로 전체 영상의 크기 대비 특정 거리 기준 이내에 오차가 존재하는지를 계산하여 성공 여부를 판단한다[22].

두 번째 지표는 IoU(Intersection over Union)로써 중심좌표 (x,y) 뿐만 아닌 바운딩박스의 넓이와 높이 (w,h)를 함께 고려하는 지표로, ground-truth의 바운딩박스 B_{GT}와 객체 추적 결과 바운딩박스 $\hat{B}$에 대해 다음과 같이 계산된다.

$$IoU=\frac{|\hat{B}\cap B_{GT}|}{|\hat{B}\cup B_{GT}|}\times 100. \quad (3)$$

$|\bullet|$는 픽셀의 수를 계산하는 연산자이며, 일반적으로 VOT(Visual Object Tracking) challenge와 같은 대회에서는 IoU 50%를 기준으로 객체 추적의 성공과 실패 여부를 판단한다[22].

그림 10에서는 정형 플렌옵틱 시퀀스 V3에 대한 2차원 객체 추적(SiamRPN++[16])과 제안하는 플렌옵틱 객체 추적 알고리즘의 결과를 정성적으로 비교하였다. 밝은 바운딩박스는 ground-truth를 나타내며, 어두운 바운딩박스는 객체 추적 결과를 나타낸다. 2차원 객체 추적의 경우 occlusion 발생 시, 객체의 추적에 실패함과 동시에 이후 유사한 형태를 띠는 다른 객체로 추적이 옮겨감을 확인할 수 있다. 반면, 제안하는 객체 추적 기술을 통해 occlusion 발생 및 해당 시점 이후에도 정확하게 객체를 추적함을 확인할 수 있다. 그림 11에서는 V3에 대한 추적 성능을 거리와 IoU를 이용해 정량적으

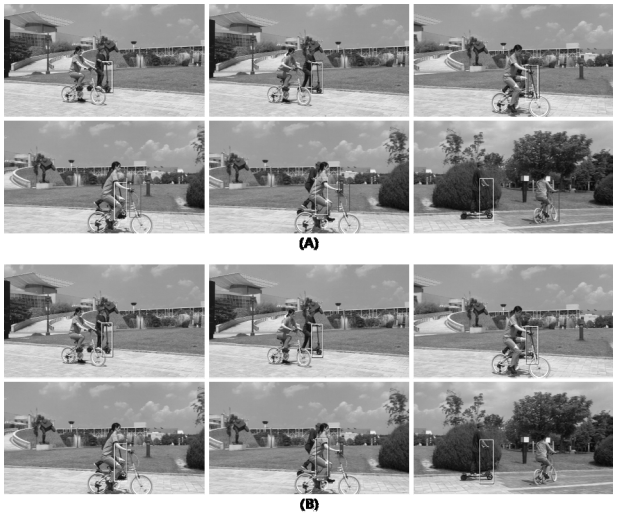

그림 10. 플렌옵틱 영상 V3 객체 추적 결과: 순서대로 1번, 50번, 90번, 100번, 105번, 150번 프레임 (A) 2차원 객체 추적 결과 (SiamRPN++[16]) (B) 제안하는 플렌옵틱 객체 추적 결과 (확대하여 세부 그림 확인)

Fig. 10. Results of object tracking on plenoptic sequence V3: 1st, 50th, 90th, 100th, 105th, and 150th frame in order. (A) 2D object tracking (SiamRPN++[16]). (B) Proposed object tracking (Please zoom-in to see details).

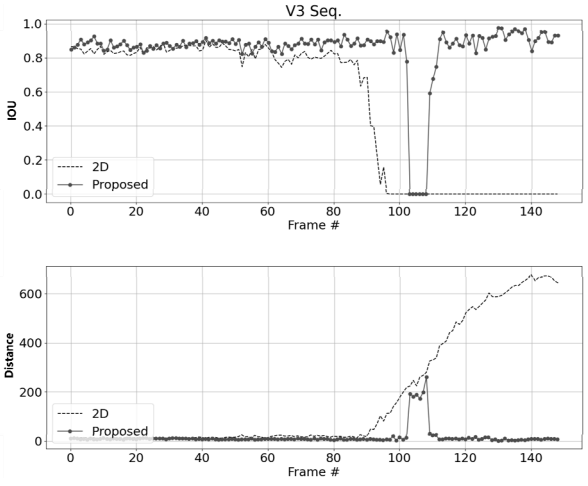

그림 11. V3 객체 추적 결과: IoU (위), 거리 (아래)

Fig. 11. Results of object tracking on plenoptic sequence V3: IoU (upper), tracking distance (lower)

로 표현하였다. 실선으로 나타난 제안하는 플렌옵틱 객체 추적 기술이 기존의 2차원 영상 객체 추적 기술 대비 높은 성능을 보임을 확인 가능하다.

그림 12에서는 비정형 플렌옵틱 시퀀스 NV4에 대한 2차원 객체 추적과 제안하는 플렌옵틱 객체 추적 알고리즘의 결과를 가시적으로 표현하였다. 밝은 바운딩박스는 ground-truth를 나타내며, 어두운 바운딩박스는 객체 추적 결과를 나타낸다. 2차원 객체 추적 대비 제안하는 객체 추적 기술은 여러 번의 occlusion발생에도

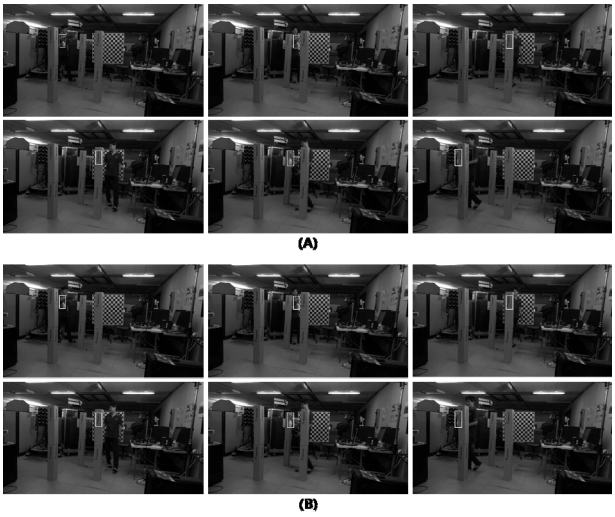

그림 12. 플렌옵틱 영상 NV4 객체 추적 결과: 순서대로 1번, 60번, 90번, 128번, 180번, 240번 프레임 (A) 2차원 객체 추적 결과 (SiamRPN++[16]. (B) 제안하는 플렌옵틱 객체 추적 결과 (확대하여 세부 그림 확인)

Fig. 12. Results of object tracking on plenoptic sequence NV4: 1st, 60th, 90th, 128th, 180th, and 240th frame in order. (A) 2D object tracking (SiamRPN++[16]). (B) Proposed object tracking (Please zoom-in to see details).

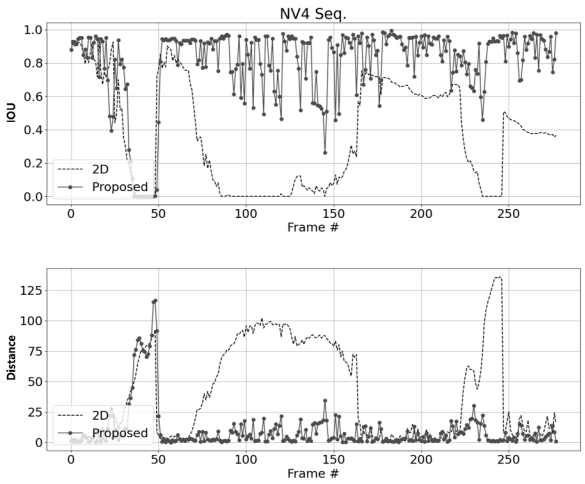

그림 13. NV4 객체 추적 결과: IoU (위), 거리 (아래)

Fig. 13. Results of object tracking on plenoptic sequence NV4: IoU (upper), tracking distance (lower).

강건하게 객체를 추적함을 확인할 수 있다. 그림 13에서는 NV4에 대한 추적 성능을 거리와 IoU를 이용해 정량적으로 표현하였으며, 실선으로 나타난 제안 기술이 기존의 2차원 영상 객체 추적 기술 대비 전체적으로 향상된 성능을 달성하였다.

본 논문에서 제안한 플렌옵틱 객체 추적 알고리즘의

표 2. 객체 추적 결과의 ablation study
Table2. Ablation study of the proposed algorithm.

시퀀스 이름	*V3*		*NV4*	
성능 지표	거리 (픽셀)	IoU (%)	거리 (픽셀)	IoU (%)
2D 객체 추적 (SiamRPN++[16])	81.80	30.98	71.85	20.13
플렌옵틱 객체 추적 베이스라인	5.54	76.28	132.00	7.45
베이스라인+ 포컬스택 내 탐색 영역 제한	4.12	84.43	3.56	77.46
베이스라인+ 포컬스택 내 탐색 영역 제한 + 선명도 계산	**3.08**	**91.66**	**3.37**	**83.04**

요소별 성능 확인을 위해 ablation study를 수행하였고, 모든 프레임에 대한 성능 지표의 평균을 계산한 결과는 표 2에 나타내었다. 2차원 객체 추적 기술보다 제안하는 유사도 기반 플렌옵틱 객체 추적 기술이 거리와 IoU 측면에서 탁월한 성능을 보임을 확인할 수 있으나, NV4의 경우에는 제안 기술의 베이스라인 모델이 2차원 객체 추적 기술을 이용할 때보다 오히려 낮은 성능을 나타낸다. 이는 상기 Ⅲ.2절에서 서술한 바와 같이 최대 유사도만을 활용하여 객체 추적을 수행할 시, 객체 잔상이 여러 포컬플레인에 걸쳐 존재함에 실제 객체와는 무관한 위치로 추적되는 문제점에 기인하며, 제안하는 포컬플레인 탐색 영역 제한 및 $t=1$에서의 선명도 계산에 따른 초기 탐색 영역 선정에 의해 월등한 성능 향상이 가능함을 의미한다.

3. 플렌옵틱 객체 추적 실패 프레임 고찰

그림 11과 13에서 확인할 수 있듯, 제안하는 유사도 기반 플렌옵틱 객체 추적 알고리즘에서도 occlusion 발생 시 간헐적으로 객체를 놓치는 현상이 발생한다. 그림 14에서는 NV4 시퀀스에서 제안하는 플렌옵틱 객체 추적 알고리즘이 실패하는 순간의 프레임을 가시화하였다. 해당 시퀀스에서의 40번째 프레임에서는 객체 추적 알고리즘이 오동작하며 이때의 최대 유사도를 갖는 제한된 포컬스택 내 포컬플레인 번호는 27번이다. 해당 포컬플레인 영상을 살펴보면, 추적하고자 하는 객체가 occluder에 의해 가려져 총 101장의 포컬플레인 중 어떠한 초점영역에서도 육안으로의 객체 형태 확인이 불가하였다. 따라서 occlusion 발생 시, 포컬스택 내 객체 추적이 가능한 정보를 갖는 포컬플레인이 전혀 존재하지 않음으로 인해 2차원 영상에서의 객체 추적과 동일한 케이스가 되며, 이는 플렌옵틱 영상 특성 상 카메라로부터의 촬영 거리가 매우 먼 영역에 대해 복수 카메라 간 시차가 거의 존재하지 않음에 기인한다. 이처럼 플렌옵틱 영상 획득 시, 카메라 간 시차 없이 렌더링되어 객체 정보가 전혀 없는 포컬스택에서는 제안하는 알고리즘 역시 한계를 가지며, 이러한 문제점의 해결을 위해 보다 진보한 형태의 추적 알고리즘 개발이 필요할 것이다.

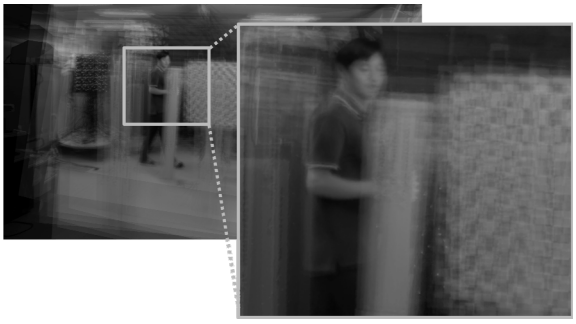

그림 14. 객체 추적 실패 프레임(NV4 시퀀스 40번)과 해당 프레임에서 선택된 포컬플레인(27번)
Fig. 14. Failed frame to track the object (40th frame of NV4 sequence) and the selected focal plane at corresponding frame by the proposed algorithm (27th focal plane).

Ⅴ. 결 론

본 논문에서는 플렌옵틱 시퀀스의 특성을 활용하여 기존 2차원 객체 추적 기술들이 해결하지 못했던 occlusion 발생 시의 타겟 객체 추적 성공을 도모하였다. 이를 위해 Siamese 심층신경망 기반의 추적 모델을 활용하여 다수의 포컬플레인에 대한 타겟 객체와의 유사도 계산을 수행하였으며, 모든 초점 영역에 대한 유사도 계산에 의해 야기되는 오류를 최소화하고자 프레임별로 포컬스택 내 탐색 영역을 제한하는 알고리즘을 제안하였다. 실제 정형 및 비정형 플렌옵틱 영상 내 객체 추적에 제안 기술을 적용한 결과, occlusion 발생 시에도 기존 기술 대비 월등한 객체 추적 성능 달성이 가능함을 정량적으로 확인하였다. 일반 영상과는 차별화된 플렌옵틱의 영상 활용으로 기존 고난이도 문제의 해결 방법을 제시함으로써, 향후 전개될 플렌옵틱 영상 기반 차세대 콘텐츠 분야에서의 전반적인 기술적 향상과 관련 응용 분야의 확대를 기대할 수 있다.

REFERENCES

[1] W. H. Son, H. W. Jang, S. J. Bae, S. J. Park, J. W. Kim, and D. H. Kim, "Plenoptic image processing technology trends," *Electronics and Telecommunications Trends*, vol. 31, no. 4, pp. 1-12, 2016.

[2] S. Wanner, C. Straehle, and B. Goldluecke, "Globally consistent multi-label assignment on the ray space of 4D light fields," *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2013.

[3] A. Shimada, H. Nagahara, and R. Taniguchi, "Change detection on light field for active video surveillance," *IEEE Conf. Advanced Video and Signal Based Surveillance (AVSS)*, 2015.

[4] J. Kwon, and K. Lee, "Visual tracking decomposition," *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2010.

[5] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernalized correlation filters," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 37, no. 3, pp. 583-596, March 2015.

[6] M. Danelljan, G. Hager, F. Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," *Int'l Conf. Computer Vision (ICCV)*, 2015.

[7] H. Nam, and B. Han, "Learning multi-domain convolutional neural networks for visual tracking," *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016.

[8] D. Held, S. Thrun, and S. Savarese, "Learning to track at 100 FPS with deep regression networks," *European Conf. Computer Vision (ECCV)*, 2016.

[9] K. Gregory, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," *Int'l Conf. Machine Learning (ICML)*, 2015.

[10] F. Schroff, D. Kalenichenko, J. Philbin, "FaceNet: a unified embedding for face recognition and clustering," *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2015.

[11] L. Bertinetto, J. Valmadre, J. F. Henrique, A. Vedaldi, and P. H. S. Torr, "Fully-convolutional Siamese networks for object tracking," European Conf. Computer Vision (ECCV), 2016.

[12] J. Valmadre, L. Bertinetto, J. F. Henriques, A. Vedaldi, and P. H. S. Torr, "End-to-end representation learning for correlation filter based tracking," *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017.

[13] A. He, C. Luo, X. Tian, and W. Zeng, "A twofold Siamese network for real-time object tracking," *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2018.

[14] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *Advanced in Neural Information Processing Systems (NIPS)*, 2015.

[15] B. Li, J. Yan, Z. Zhu, and X. Hu, "High performance visual tracking with Siamese region proposal network," *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2018.

[16] B. Li, W. Wu, Q. Wang, F. Zhang, J. Xing, and J. Yan, "SiamRPN++: evolution of Siamese visual tracking with very deep network," *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2019.

[17] Q. Wang, L. Zhang, L. Bertinetto, W. Hu, P. H. S. Torr "Fast online object tracking and segmentation: a unifying approach," *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2019.

[18] D. Guo, J. Wang, Y. Cui, Z. Wang, and S. Chen "SiamCAR: Siamese fully convolutional classification and regression for visual tracking," *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2020.

[19] J. W. Kim, S. Bae, S. Park, and D. H. Kim, "Object tracking using plenoptic image sequences," *SPIE Three-Dimensional Imaging, Visualization*, and Display, 2017.

[20] D. H. Bae, J. W. Kim, H. C. Noh, D. H. Kim, and J. Heo, "Plenoptic imaging techniques for improving accuracy and robustness of object tracking," *SPIE Three-Dimensional Imaging, Visualization, and Display*, 2018.

[21] D. H. Bae, J. W. Kim, and J. Heo, "Content-aware focal plane selection and proposals for object tracking on plenoptic image sequences," *Sensors,* vol. 19, no. 48, pp. 1-20, 2019.

[22] M. Kristan, J. Matas, A. Leonardis, T. Vojir, R. Pflugfelder, G. Fernandez, G. Nebehay, F. Porikli, and L. Vcehovin, "A novel performance evaluation methodology for single-target trackers," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 38, no. 11, pp. 2137-2144, Nov. 2016.

저 자 소 개

오 희 석(정회원)

2017년 연세대학교 전기전자공학과 박사 졸업

2017년 삼성전자 DMC 연구소 책임연구원

2017년～2020년 한국전자통신연구원 선임연구원

2020년～현재 한성대학교 IT융합공학부 조교수

<주관심분야 : 영상처리, 컴퓨터비전, 혼합현실, 심층생성모델 등>