



특집논문 (Special Paper)

방송공학회논문지 제29권 제6호, 2024년 11월 (JBE Vol.29, No.6, November 2024)

<https://doi.org/10.5909/JBE.2024.29.6.809>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

드론 동영상을 통한 광학 흐름 기반 프레임 보간 모델 성능 분석

오명훈^{a)}, 이성배^{a)}, 김규현^{a)†}

Optical flow based Frame Interpolation Model Analysis for Drone Captured Video Sequence

Myeounghun Oh^{a)}, Seongbae Rhee^{a)}, and Kyuheon Kim^{a)†}

요 약

광학 흐름(Optical Flow) 기반 동영상 프레임 보간법은 실제 존재하는 프레임을 입력으로 받아, 추정된 광학 흐름을 통해 프레임 사이 임의의 시간에 해당하는 프레임을 생성하는 기술이다. 이러한 동영상 프레임 보간법은 주로 Vimeo90k, DAVIS 등과 같이 촬영된 카메라의 고도 정보가 크게 고려되지 않는 평면적 움직임만 존재하는 영상에 대해서 실험이 진행되었으며, 이때의 광학 흐름은 단방향적 움직임을 보인다. 그러나 드론 영상 등과 같이 촬영하는 카메라의 고도에 따라 피사체와의 거리가 변동되는 동영상은 광학 흐름이 단방향적인 움직임뿐만 아니라 수렴(Convergence)과 발산(Divergence)할 것으로 예상되는데, 이러한 광학 흐름에서 프레임 보간법의 성능 분석 실험은 부족하다. 따라서 본 논문은 드론 영상 등 카메라의 고도 정보에 따라 수렴과 발산 형태의 광학 흐름을 보이는 동영상을 통해 광학 흐름 기반 프레임 보간법의 성능 분석을 진행하고자 한다.

Abstract

Optical flow-based video frame interpolation is a technology that receives an actual frame as input and generates a frame corresponding to an arbitrary time between the frames through the estimated optical flow. This video frame interpolation method was mainly conducted on images with only flat movements, such as Vimeo90k and DAVIS, where altitude information of photographed cameras was not largely considered, and the optical flow at this time shows a unidirectional movement. However, in videos in which the distance to the subject changes depending on the altitude of the camera being photographed, such as drone images, the optical flow is expected to converge and diverge as well as unidirectional movement, and experiments to analyze the performance of frame interpolation in this optical flow are insufficient. Therefore, this paper aims to analyze the performance of optical flow-based frame interpolation method through videos showing optical flow in the form of convergence and divergence according to the altitude information of the camera such as drone images.

Keyword : Optical Flow, Frame Interpolation, Unidirectional movement, Drone images, Camera altitude

I. 서론

동영상 프레임 보간법은 실제 존재하는 프레임을 입력으로 하여, 움직임을 추정한 뒤 이를 기반으로 프레임을 보간하는 기술이다^[1]. 주로 슬로우 모션 생성, 프레임률 향상, 비디오 복원 등에 프레임 보간 기술이 사용되었으며, 크게 두 가지 절차인 움직임 정보 추정 단계와 픽셀 합성 단계로 진행되었다. 보통 프레임 사이 움직임 정보를 광학 흐름에 기반하여 추정하며 이후 앞서 추정된 광학 흐름을 이용해 픽셀을 합성하는 방식으로 프레임을 보간하였다.

그러나 움직임 정보가 광학 흐름에 기반하였기 때문에 광학 흐름 추정에 오류가 발생할 수 있는 급격한 조도 변화, 큰 움직임, 폐색 등의 상황이 발생할 경우 아티팩트(Artifact)로 인해 광학 흐름 기반 프레임 보간법의 성능 또한 떨어지는 문제가 발생하였다. 때문에, 앞선 문제 상황에서 광학 흐름 기반 프레임 보간법이 아닌 다른 방식으로 프레임 보간을 하려는 시도가 있었으며, 그중 하나는 커널 기반 기술을 통한 프레임 보간이다. 최초의 커널 기반 보간법인 AdaConv(Adaptive Convolution)^[2]은 보간된 프레임의 픽셀 합성을 두 입력 프레임 간의 국소적 컨볼루션(Convolution)으로 간주하여, 보간된 프레임의 특정 픽셀 위치에 대해 입력 영상의 주변 화소들을 컨볼루션으로 반영하여 커널 크기만큼 움직임에 대하여 효과적으로 처리하였다^[3]. 때문에, 급격한 조도 변화와 같은 광학 흐름 정보가 부정확한 상황일 경우 프레임 보간을 효과적으로 할 수 있

었다. 그러나 커널 크기를 넘어서는 움직임에 대해서는 추정이 힘들었으며 영상 크기에 따라 GPU 메모리 사용량이 기하급수적으로 늘어나는 문제가 존재하며 이후 등장한 모델들 또한 커널 크기로 인한 움직임 정보 추정에 제약을 받고 있기 때문에 광학 흐름 기반 프레임 보간법은 여전히 주요한 프레임 보간 방법으로 사용되고 있다.

광학 흐름 기반 프레임 보간법은 특정 픽셀의 다음 프레임의 움직임을 추정하는 광학 흐름에 기반하여 보간된 프레임의 픽셀을 합성한다. 광학 흐름 자체가 프레임별 동일 위치의 픽셀 변위는 작다는 것을 전제로 하고 있기 때문에^[4] 조도 변화, 큰 움직임, 폐색 등의 상황에선 광학 흐름 추정 자체가 부정확한 특성이 있으며, 이러한 이유로 광학 흐름 기반 프레임 보간법은 프레임의 해상도를 다양하게 조정해 가며 점진적으로 광학 흐름을 근사화하는 방법을 사용하고 있다.

이러한 광학 흐름 기반 프레임 보간 모델은 주로 Vimeo-90k, DAVIS 등의 데이터셋을 대상으로 평가가 진행되었는데^[5], 이와 같은 동영상은 보통 카메라의 Z축은 고정된 상태에서 X, Y축의 움직임만이 변화하는 평면적인 양상의 동영상이며, 이때 광학 흐름은 단편향적으로 나타난다.

그러나 드론을 이용한 영상은 고도가 자유롭게 조정 가능하다는 점에서 앞선 동영상과 다른 특성을 갖는다^[6]. 일례로, 전진비행과 더불어 저고도에서 고고도로 상승하며 촬영된 영상은 배경이 움직이는 X, Y축 움직임과 더불어 Z축 움직임으로 인해 피사체가 작아지게 된다. 이처럼 고도 변화로 인해 광학 흐름 또한 피사체의 X, Y만 변화하는 단편향적 움직임만이 아닌 피사체의 축소 및 확장하게 되어 광학 흐름은 수렴 또는 발산 형태로 나타날 것으로 예상되는데, 이러한 광학 흐름에서 프레임 보간법의 성능 분석은 부족한 실정이다. 따라서 본 논문에서는 드론 영상과 같이 고도 변화로 인해 다양한 양상의 광학 흐름이 나타나는 영상에 대하여 광학 흐름 기반 프레임 보간법의 성능을 분석하고자 한다.

II. 배경 기술

1. 광학 흐름 기반 프레임 보간법

광학 흐름 기반 프레임 보간법에선 중간 프레임을 생성

a) 경희대학교 전자공학과(Department of Electronic Engineering, Kyung Hee University)

* Corresponding Author : 김규현(Kyuheon Kim)

E-mail: kyuheonkim@khu.ac.kr

Tel: +82-31-201-3810

ORCID: <https://orcid.org/0000-0003-1553-936X>

※ 이 논문의 연구 결과 중 일부는 한국방송·미디어공학회 2024년 하계 학술대회에서 발표한 바 있음.

※ This work was supported by Korea Research Institute for defense Technology planning and advancement(KRIT) grant funded by the Korea government (DAPA(Defense Acquisition Program Administration)) (21-106-A00-007, Space-Layer Intelligent Communication Network Laboratory, 2022) and supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2024-2021-0-02046) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation).

· Manuscript September 4, 2024; Revised October 4, 2024; Accepted October 4, 2024.

하기 위해서 프레임 N과 프레임 N+1 사이의 특정 시간 t에 대한 광학 흐름을 추정해야 하며, 이를 추정하는 방식에 따라 그림 1과 같이 Forward warping, Backward warping으로 구분된다^[7]. Forward warping은 원본 프레임에서 보간될 프레임으로 광학 흐름을 추정하고 원본 프레임의 픽셀 하나가 보간될 프레임으로 매핑(mapping)되는 방식으로 이루어지며, Backward warping은 보간될 프레임에서 원본 프레임으로 광학 흐름을 추정하고 보간될 프레임의 픽셀 하나를 원본 프레임에서 매핑하는 방식으로 이루어진다. 때문에 Forward warping에선 원본 프레임의 픽셀이 보간될 프레임의 픽셀에 겹치게 매핑이 되거나 혹은 보간될 프레임에 원본 프레임의 픽셀이 매핑되지 않아 그림 1의 (a)와 같은 폐색(Occlusion) 등이 발생하는 경우가 있기에 그림 1의 (b)와 같은 Backward warping 방식이 보편적으로 사용되고 있다.

Forward warping은 프레임 N에서 프레임 N+1 사이의 특정 시간 t에 대한 광학 흐름을 추정하며 이미지를 보간한다. 일례로, 프레임 0과 프레임 1 사이 광학 흐름을 $F_{0 \rightarrow 1}$ 이라 할 경우, 보간된 프레임 t는 시간축 $t \in (0, 1)$ 에 속하게 된다. 프레임 0에서 보간된 프레임 사이 광학 흐름은 식 (1)과 같이 표기되며, 역방향 광학 흐름인 프레임 1에서 보간된 프레임 사이 광학 흐름은 식 (2)와 같이 표기되게 된다.

$$F_{0 \rightarrow t} = tF_{0 \rightarrow 1} \quad (1)$$

$$F_{1 \rightarrow t} = (1 - t)F_{1 \rightarrow 0} \quad (2)$$

이후 앞서 구한 광학 흐름을 이용하여 식 (3), (4)와 같

이 t 시간 프레임 I_t 픽셀 매핑을 통해 보간하게 된다. 픽셀 매핑 방식에는 합산 매핑, 평균 매핑, 선형 매핑, 소프트맥스 매핑이 있는데^[8], $\vec{\sigma}$ 는 소프트맥스 매핑을 표기한 기호이다. 소프트맥스 매핑을 제외한 다른 매핑 방식들은 픽셀 겹침(overlapping) 문제로 인해 현재 Forward warping의 픽셀 매핑 방식에선 소프트맥스 방식이 가장 좋은 성능을 달성하고 있다. Z값은 픽셀 각각의 깊이와 유사한 개념으로^[9], 네트워크를 통해 얻어지는 가중치이며 원본 픽셀에 가중치들이 지수함수 꼴로 곱해져 프레임을 보간하게 된다.

$$I_t \approx \vec{\sigma}(I_0, F_{0 \rightarrow t}) = \frac{\Sigma(\exp(Z) \cdot I_0, F_{0 \rightarrow t})}{\Sigma(\exp(Z), F_{0 \rightarrow t})} \quad (3)$$

$$I_t \approx \vec{\sigma}(I_1, F_{1 \rightarrow t}) = \frac{\Sigma(\exp(Z) \cdot I_1, F_{1 \rightarrow t})}{\Sigma(\exp(Z), F_{1 \rightarrow t})} \quad (4)$$

Backward warping 또한 프레임 0과 프레임 1 사이 특정 시간 t에 대한 광학 흐름을 추정해야 하며, 이때 t에 대한 광학 흐름은 식 (5), (6)에서 나타나는 바와 같이 프레임 0에서 프레임 1로의 광학 흐름($F_{0 \rightarrow 1}$)과 프레임 1에서 프레임 0으로의 광학 흐름($F_{1 \rightarrow 0}$)을 통해 추정된다^[10].

$$F_{t \rightarrow 1} = (1 - t)F_{0 \rightarrow 1} \quad (5)$$

$$F_{t \rightarrow 1} = -(1 - t)F_{1 \rightarrow 0} \quad (6)$$

보간된 프레임에서 프레임 0으로의 광학 흐름은 식 (7)과

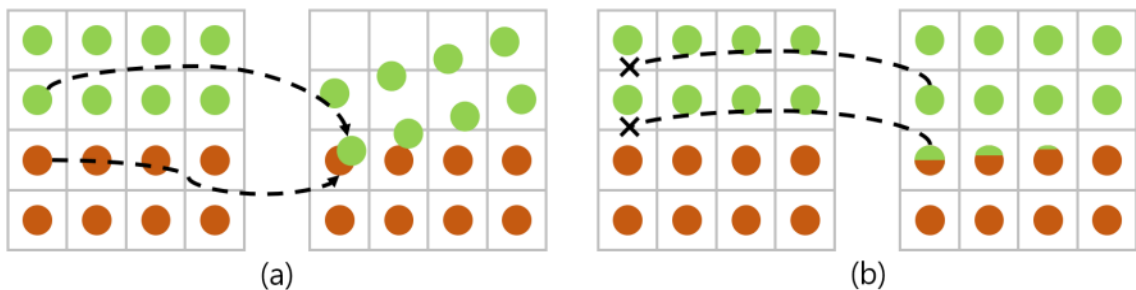


그림 1. Forward Warping 및 Backward Warping (a) Forward Warping and occlusion (b) Backward Warping
Fig. 1. Forward Warping and Backward Warping (a) Forward Warping and occlusion (b) Backward Warping

같이 표기되며, 양방향 광학 흐름을 사용하여 근사할 경우, 식 (8)과 같이 표기된다. 마찬가지로 보간된 프레임에서 프레임 1로의 광학 흐름을 양방향 광학 흐름을 사용해 근사할 경우, 식 (9)와 같이 표기된다. 이를 통해 식 (10)과 같이 프레임을 합성하며, 해당 식에서 사용되는 $g(\cdot, \cdot)$ 는 역방향 와핑 함수^[11], α_0 는 사용자 파라미터이다. 식 (10)과 식 (3), (4)를 통해 알 수 있듯이, Forward warping과 Backward warping은 원본 프레임에서 보간된 프레임으로 광학 흐름을 추정하고 이후 합성하느냐, 보간된 프레임에서 원본 프레임으로 광학 흐름을 추정하고 합성하느냐의 차이가 존재한다.

$$F_{t \rightarrow 0}(X) = tF_{1 \rightarrow 0}(X) \quad (7)$$

$$F_{t \rightarrow 0} = -(1-t)tF_{0 \rightarrow 1} + t^2F_{1 \rightarrow 0} \quad (8)$$

$$F_{t \rightarrow 1} = (1-t)^2F_{0 \rightarrow 1} - t(1-t)F_{1 \rightarrow 0} \quad (9)$$

$$I_t = \alpha_0 \odot g(I_0, F_{t \rightarrow 0}) + (1 - \alpha_0) \odot g(I_1, F_{t \rightarrow 1}) \quad (10)$$

2. 드론 동영상의 광학 흐름 특성

광학 흐름은 두 개 영상 간의 픽셀 차이를 통해 측정하는 것으로 그림 2에서 나타난 바와 같이 (a)의 #번째와 #번째 프레임 간의 광학 흐름은 xx 네트워크를 통해 측정하면, 이는 그림 2 (b)에 나타난 것과 같이 이동성이 높은 부분은 상대적으로 짙은 색깔로 나타내어 표시될 수 있다. 본 논문에서 사용하는 DAVIS, Vimeo90k, KITTI와 같은 동영상에서는 프레임 0과 1 사이의 움직임이 주로 X, Y축으로 나타나기에 단편향적으로 나타난다.

그러나 드론 영상의 경우에는 지상 촬영과 달리 공간에 제약받지 않으며 고도 변화가 자유롭다는 특징으로 인해 단편적인 움직임뿐만 아니라 회전(Rotation), 수렴(Convergence), 발산(Divergence) 등 다양한 양상의 광학 흐름이 포함될 수 있다. 일례로, 드론이 고도를 상승하며 촬영하는 경우, 동영상 프레임 내의 피사체의 크기는 점차 축소되기에 그림 3에서 나타나는 바와 같이 영상 중심 방향으로 수렴하는 광학 흐름이 나타날 수 있다. 이같이 다양한 방향성



그림 2. 일반적인 평면 동영상에서 광학 흐름 (a) 17th 프레임 (b) 18th 프레임 (c) 단편향적 형태의 광학 흐름
Fig. 2. Optical flow in a typical planar video (a) 17th frame (b) 18th frame (c) Optical flow in undirectional form



그림 3. 드론의 고도 상승 시 광학 흐름 (a) 13th 프레임 (b) 14th 프레임 (c) 수렴하는 형태의 광학 흐름
Fig. 3. Optical flow when the drone rise in altitude (a) 13th frame (b) 14th frame (c) Converging optical flow

이 포함되는 광학 흐름에서 동영상 프레임 보간의 성능은 저하될 수 있기에, 본 논문은 해당 광학 흐름에서 광학 흐름 기반 알고리즘별 성능을 확인 및 검증하고자 한다.

III. 본 론

본 논문에서는 다양한 방향성이 포함되는 드론 동영상에 선 광학 흐름 추정 순서에 따라 기존 단편향적 광학 흐름을 보이는 데이터셋과 결과가 다를 수 있다는 가정하에, 광학 흐름 추정 순서가 다른 프레임 보간 방식인 Forward warping, Backward warping 두 가지로 구분됨에 따라 모델을 와핑(Warping) 방식에 따라 나누며 이후 광학 흐름만 사용한 모델, 광학 흐름 및 픽셀의 상관관계(Correlation)나 코스트 볼륨(Cost Volume) 등의 추가적인 정보를 활용한 모델과 같이 세부적인 방식으로 나눈다.

Forward Warping에 사용한 모델은 Softmax Splatting^[8], M2M(Many-to-Many Splatting)^[12]를 사용하였으며 Backward Warping에 사용한 모델은 Super Slomo^[10], RIFE (Real-Time Intermediate Flow Estimation)^[5]를 사용하였다. Backward Warping과 추가적인 다른 정보를 사용한 모델은 BMBC(Bilateral Motion Estimation with Bilateral Cost Volume)^[13], AMT(All-Pairs Multi-Field Transforms)^[14]를 사용하였다.

Softmax Splatting 모델은 Forward Warping에서 발생하는 픽셀 겹침 문제를 방지하기 위해 소프트맥스 함수를 사용하여 원본 픽셀의 가중치를 계산하고 이를 기반으로 픽셀을 합성하게 된다. 내부는 피쳐(feature) 피라미드 구조를 사용하여 원본을 대상으로 각각 다른 스케일을 가진 피쳐들이 산출되게 된다. M2M 모델은 기존의 Forward warping 방식은 보간될 프레임에 픽셀을 직접 매핑하는 방식으로 인해 불필요한 구멍(hole)이 생길 수 있는 문제를 방지하기 위해 원본 픽셀을 보간될 프레임의 여러 픽셀에 매핑시키는 다대다 매핑 방식을 사용한다. 내부는 여러 움직임 정제(Refine) 네트워크를 통해 다수의 광학 흐름을 산출하며 이를 통해 다대다 매핑을 수행하게 된다. Super Slomo 모델은 기존의 방법들은 주로 단일 프레임 보간에 초점을 맞추었기 때문에, 다중 프레임 보간을 하기 위해 양방향 광학 흐름

을 사용해 중간 광학 흐름을 근사화하고 이를 통해 와핑하는 방식으로 보간될 프레임을 합성하였다. 광학 흐름 계산과 광학 흐름 보간을 위해 U-Net 구조를 채택하였다. RIFE 모델은 중간 광학 흐름을 근사할 경우 물체 이동 문제로 인해 광학 흐름 근사가 쉽지 않은 문제를 해결하기 위해 IFNET이라는 네트워크를 통해 중간 흐름을 직접적으로 근사하는 방법을 제시하였다. 연속된 IFNET에 다양한 스케일로 변형된 원본 영상이 들어가 중간 광학 흐름이 추정되며 이를 통해 보간 프레임을 합성한다. BMBC 모델은 기존 광학 흐름만 사용한 모델과 달리 두 영상 간의 상관관계를 이용한 코스트 볼륨(Cost Volume)과 문맥(Context) 정보를 이용한 문맥 맵을 사용한다. 광학 흐름 추정이 부정확하다는 가정에 기반하여, 문맥 맵과 코스트 볼륨을 통해 더 정확한 광학 흐름을 추정하게 되며 문맥 맵은 보조적인 수단으로 사용한다. AMT 모델은 대부분의 모델들이 광학 흐름을 근사하기 위해 콘볼루션을 사용한 U-Net 구조를 채택하여 부정확한 광학 흐름이 근사 된다고 가정하여 상관(Correlation) 인코더, 문맥 인코더를 결합해 중간 광학 흐름을 근사하고 이를 통해 프레임을 보간하였다.

본 논문은 앞선 모델들의 단편향적 광학 흐름에 대한 성능 분석을 위해 480p, 1080p의 해상도를 가지며 비디오 분할(Segmentation)을 위해 벤치마킹된 DAVIS^[15] 데이터셋을 사용하며, 또한 다양한 방향성이 포함되는 두 가지 드론 데이터셋을 통한 실험으로 모델들의 성능을 확인 및 검증한다. 첫 번째 데이터셋은 1000개의 60fps로 촬영된 4K 해상도로 이루어진 Inter4K^[16] 데이터셋의 (15, 21, 75, 177, 215, 224, 242, 405, 555, 712) 데이터를 이용한 광학 흐름의 수렴, 발산, 회전 형태가 두드러지지 않는 일반적인 드론 데이터이며, 두 번째 데이터셋은 드론 동영상 중 광학 흐름의 수렴, 회전, 발산 형태가 두드러지는 커스텀 데이터셋이다. 실험은 각각의 데이터셋을 대상으로 N 번째 프레임과 N+2 번째 프레임을 입력으로 하여 N+1 번째 프레임을 보간하여 이를 원본과 PSNR, SSIM 평가지표를 통해 비교한다. 이후 원본과 보간되는 프레임 사이 간격인 Pivot Interval을 늘려 N 번째 프레임과 N+4 번째 프레임을 입력으로 하여 N+2 번째 프레임을 보간하여 이를 원본과 비교한다. 마찬가지로, N 번째 프레임과 N+6 번째 프레임을 입력으로 하여 N+3 번째 프레임을 보간한다.

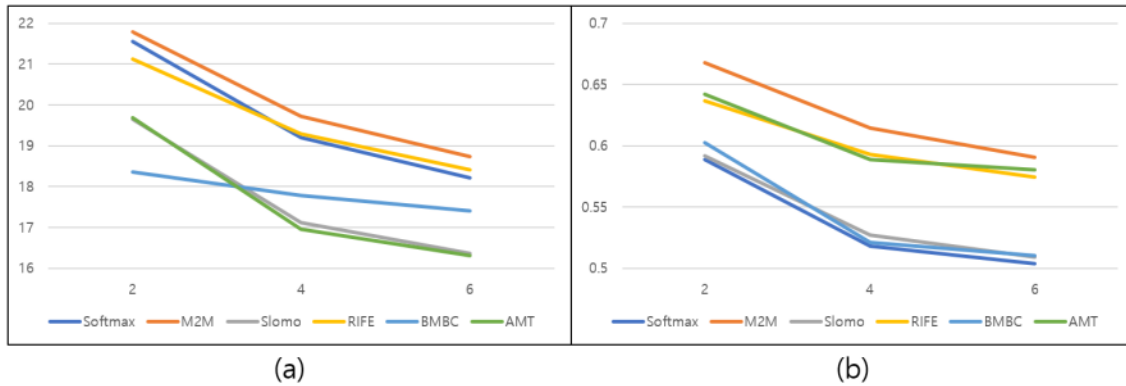


그림 4. Pivot Interval에 따른 DAVIS 데이터셋의 성능 평가 결과 (a) PSNR 평가지표 (b) SSIM 평가지표

Fig. 4. Performance evaluation results of DAVIS dataset according to pivot interval (a) PSNR indicators (b) SSIM indicators

표 1. DAVIS 데이터셋에 대한 프레임 보간 성능 비교

Table 1. Comparison of VFI performance for DAVIS dataset

Pivot Interval(2)	DAVIS	
	PSNR	SSIM
Softmax	21.55	0.5889
M2M	21.78	0.6679
Slomo	19.64	0.5920
RIFE	21.12	0.6369
BMBC	18.36	0.6026
AMT	19.68	0.6423

표 1은 DAVIS 데이터셋을 대상으로 앞선 모델들의 N 번째 프레임과 N+2 번째 프레임을 원본으로 하여 N+1 번째 프레임을 보간한 실험 결과이며 Forward Warping에 속하는 M2M 방식이 가장 우수한 성적을 기록하였다. 이후, Softmax, RIFE, AMT, Slomo, BMBC 순으로 성능이 우수한 것으로 나타났다. 이러한 현상은 Pivot Interval을 2, 4, 6으로 증가시킨 그림 4에서도 동일하게 관찰되었으며, 원본 프레임과 보간되는 프레임 사이 간격을 늘려서 실험을 진행하더라도 M2M 모델이 가장 우수한 성능을 기록하였다.

그러나 DAVIS 데이터셋이 아닌 첫 번째 드론 데이터셋을 사용한 경우 다른 결과가 나타났다. 표 2는 Inter4K 데이터셋의 드론 동영상에 대한 모델들의 실험 결과이며, 전반적으로 DAVIS 데이터셋에 비해 PSNR, SSIM 평가지표에서 개선된 성능을 보였으며, 이는 해당 데이터셋이 DAVIS 데이터셋에 비해 광학 흐름 추정이 용이하였기에 나온 결

과로 보인다. PSNR 평가지표에선 Softmax Splatting 모델이 가장 우수한 측정 결과를 기록하였으며, SSIM 평가지표에선 AMT 모델이 가장 우수한 측정 결과를 기록하였다. 그러나 앞선 결과는 Pivot Interval을 증가시켰을 경우, 상기의 결과와 다르게 그림 5와 같이 PSNR 평가지표에선 원본 프레임과 보간된 프레임의 간격에 상관없이 Softmax Splatting이 가장 우수한 성능을 기록하였으며, SSIM 평가지표에선 AMT 모델이 가장 우수한 성능을 기록하다 보간 간격이 6으로 증가할 때 M2M 모델이 가장 우수한 성능을 기록하였다. Pivot Interval이 증가하며 광학 흐름 추정이 어려워졌고, DAVIS의 실험 결과에서 AMT 모델이 낮은 성능을 기록한 것을 고려할 때, AMT 모델은 광학 흐름 추정에 프레임 보간 성능이 크게 영향받는 것으로 보인다. 또한 앞선 DAVIS 데이터셋에선 Pivot Interval 증가에 따라 전체 모델들의 성능이 하락한 반면, 드론 데이터셋에 대해선 전체 모델들이 소폭 상승하는 모습을 보였다.

표 2. 첫 번째 드론 데이터셋에 대한 프레임 보간 성능 비교

Table 2. Comparison of VFI performance for No.1 Drone dataset

Pivot Interval(2)	Drone	
	PSNR	SSIM
Softmax	34.92	0.9414
M2M	34.37	0.9475
Slomo	32.06	0.9444
RIFE	33.90	0.9433
BMBC	29.10	0.9131
AMT	34.39	0.9486

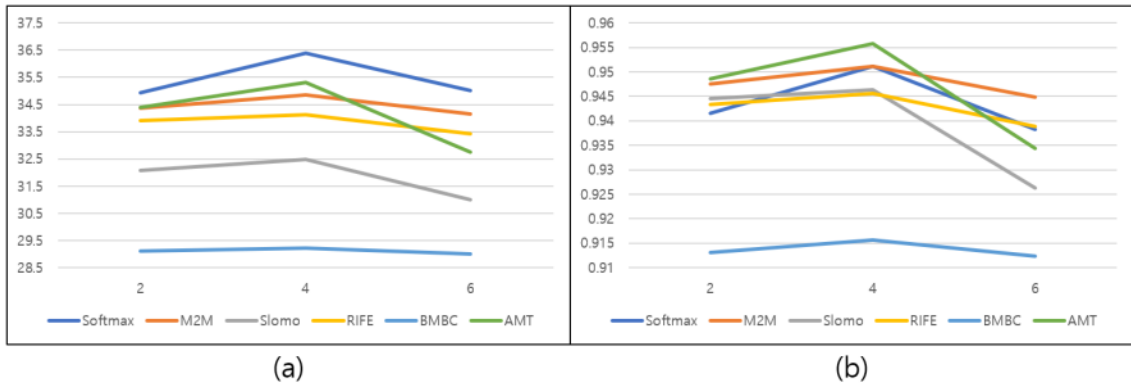


그림 5. Pivot Interval에 따른 드론 데이터셋 1번의 성능 평가 결과 (a) PSNR 평가지표 (b) SSIM 평가지표

Fig. 5. Performance evaluation results of Drone dataset No.1 according to pivot interval (a) PSNR indicator (b) SSIM indicator

표 3. 두 번째 드론 데이터셋에 대한 프레임 보간 성능 비교

Table 3. Comparison of VFI performance for No.2 Drone dataset

Pivot Interval(2)	Drone	
	PSNR	SSIM
Softmax	34.66	0.9374
M2M	34.14	0.9454
Slomo	31.19	0.9321
RIFE	34.02	0.9453
BMBC	28.74	0.9174
AMT	35.25	0.9502

표 3은 수렴, 발산, 회전 양상의 광학 흐름이 두드러지는 두 번째 드론 데이터셋을 대상으로 한 모델들의 실험 결과이며 이전과 다른 결과가 나타났다. AMT 모델이 PSNR,

SSIM 평가지표에서 가장 우수한 성능을 기록하였다. 그러나 Pivot Interval을 증가시켜 실험을 진행하였을 때, 보간 프레임과 원본 프레임 사이 간격이 증가할수록 성능이 하락하여 그림 6과 같이 보간 간격이 6에 해당될 때, PSNR 평가지표에선 Softmax Splatting, SSIM 평가지표에선 M2M 모델이 가장 우수한 성능을 기록하였다. 아울러, 이전 드론 데이터에서 관측되었던 Pivot Interval 증가에 따른 보간 성능 향상은 관측되지 않았으며, 보간 간격이 증가함에 따라 모든 모델의 성능이 지속적으로 하락하는 모습을 보여주었다. 상기에서 설명한 바와 같이 두 번째 드론 영상의 경우, 수렴, 발산 및 회전의 특징을 갖고 있기에 이에 대한 개별적인 평가가 필요하여, 표 4와 5에서 나타나는 바와 같이 데이터를 나누고 실험을 진행하였다.

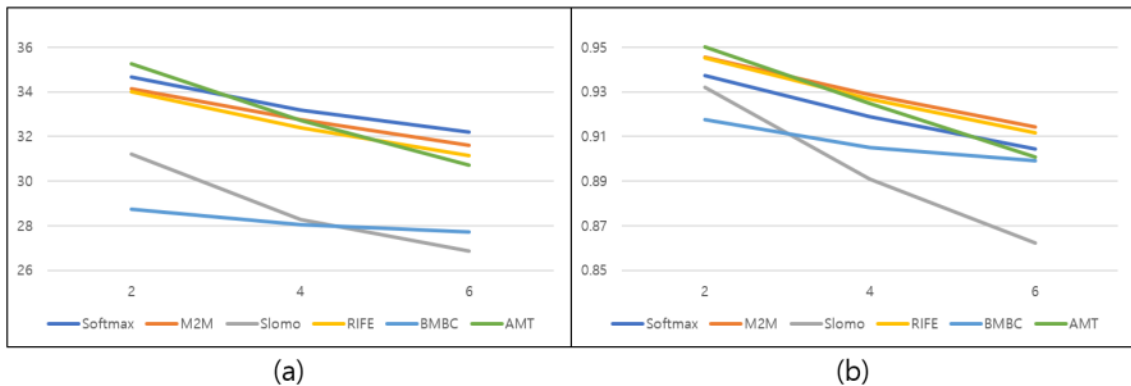


그림 6. Pivot Interval에 따른 드론 데이터셋 2번의 성능 평가 결과 (a) PSNR 평가지표 (b) SSIM 평가지표

Fig. 6. Performance evaluation results of Drone dataset No.2 according to pivot interval (a) PSNR indicator (b) SSIM indicator

표 4. 두 번째 드론 데이터셋에 대한 PSNR 성능지표

Table 4. Comparison of PSNR performance for No.2 Drone dataset

Pivot Interval(2)	Convergence	Divergence	Rotation
Softmax	40.71	38.86	24.43
M2M	39.79	37.90	24.74
Slomo	34.07	34.99	24.53
RIFE	39.38	37.03	25.66
BMBC	29.94	31.75	24.55
AMT	41.7	38.38	25.67
Average	37.59	36.48	24.93

표 5. 두 번째 드론 데이터셋에 대한 SSIM 성능지표

Table 5. Comparison of SSIM performance for No.2 Drone dataset

Pivot Interval(2)	Convergence	Divergence	Rotation
Softmax	0.9855	0.9800	0.8467
M2M	0.9859	0.9807	0.8696
Slomo	0.9767	0.9757	0.8440
RIFE	0.9831	0.9765	0.8762
BMBC	0.9524	0.9573	0.8426
AMT	0.9905	0.9814	0.8789
Average	0.9790	0.9752	0.8596

상기에서 설명한 바와 같이, 드론 영상은 일반 영상 대비 수렴, 발산 및 회전이 많이 나타나기에 이를 기준으로 표 4와 5에서는 두 번째 드론 데이터셋을 수렴, 발산, 회전에 해당하는 영상으로 나누고 실험을 진행하였다. 수렴하는 영상은 PSNR 평균이 37.59로 가장 높았으며, 발산하는 영상은 36.48, 회전하는 영상은 24.93의 가장 낮은 평균을 기록하였다. 마찬가지로, SSIM 평가지표에서도 수렴하는 영상이 0.9790의 가장 높은 평균을 기록하였으며, 발산 영상은 0.9752, 회전 영상은 0.8596의 가장 낮은 평균을 기록하였다. PSNR 표준편차는 수렴 영상에서 4.595, 발산 영상은 2.690, 회전 영상에서 0.5781을 기록하였으며 SSIM 표준편차는 수렴 영상에서 0.0137, 발산 영상에서 0.0091, 회전 영상에서 0.0170을 기록하였다. 회전 영상의 경우 전체적으로 보간 성능이 나빠져 표준편차도 0.5781로 적은 모습을 보이고 있는데, 이는 전체적으로 보간 성능이 나빠져 PSNR 표준편차가 2.204를 기록한 DAVIS 데이터셋과는 대조적이다. 단순한 광학 흐름 추정의 어려움 때문에 보간 성능이 나쁜 것이라면, DAVIS와 같이 모델 별로 성능이

크게 차이가 나야 하지만, 그렇지 않다는 것은 회전 영상의 특징이 프레임 보간에 많은 어려움을 야기하는 것으로 파악된다.

IV. 결 론

본 논문에서는 현재 Vimeo90K, UCF101, DAVIS 등과 같이 단편향적 광학 흐름 양상을 보이는 데이터셋을 대상으로 광학 흐름 프레임 보간법이 평가되었다는 점에 착안하여, 드론 영상과 같은 카메라의 고도 정보에 따라 피사체와 거리가 변동되며 수렴, 발산, 회전 등 다양한 광학 흐름 양상을 보이는 영상을 대상으로 성능 분석을 진행하였다. Warping 기법에 따라 모델을 우선 분류하며, 이후 세부적으로 모델을 분류해 프레임 보간 간격을 변경하며 실험을 진행하였다. DAVIS 데이터셋에 대해선 Forward Warping 기법이 비교적 우수한 성능을 보였으며, 첫 번째 드론 영상에 대해선 Backward Warping 기법과 추가적인 정보를 사용한 AMT 모델이 우수한 성능을 보였으나, Pivot Interval에 따라 결과는 상이하게 나타났다. 수렴, 발산, 회전의 광학 흐름 양상이 명확하게 나타나는 두 번째 드론 영상에 대해서 실험을 진행하였으며 회전 영상은 가장 낮은 PSNR, SSIM 평균을 기록하며 가장 작은 PSNR 표준편차를 기록하였다. DAVIS 데이터셋도 마찬가지로 가장 낮은 PSNR 평균을 기록하였지만, 모델 성능의 표준편차는 크게 기록되었으며, 때문에 회전 영상은 단순히 광학 흐름 추정의 어려움 때문 보다는 회전이 갖는 특징으로 인해 프레임 보간에 어려움을 갖는 것으로 파악되기에 회전으로 인한 추가적인 요인을 밝혀내기 위해 연구가 필요할 것으로 예상된다.

참 고 문 헌 (References)

- [1] Dong, J., Ota, K., & Dong, M. (2023). Video frame interpolation: A comprehensive survey. *ACM Transactions on Multimedia Computing, Communications and Applications*, 19(2s), 1-31. doi: <https://doi.org/10.1145/3556544>
- [2] Niklaus, S., Mai, L., & Liu, F. (2017). Video frame interpolation via adaptive convolution. In *Proceedings of the IEEE conference on*

- computer vision and pattern recognition (pp. 670-679).
doi: <https://doi.org/10.1109/cvpr.2017.244>
- [3] Heo, J. G., Yun, G. H., Kim, S. J., & Jeong, J. U. (2022). Trend of research on video frame interpolation technology based on deep learning. *Broadcasting and Media Magazine*, 27(2), 51-61.
- [4] Zhai, M., Xiang, X., Lv, N., & Kong, X. (2021). Optical flow and scene flow estimation: A survey. *Pattern Recognition*, 114, 107861.
doi: <https://doi.org/10.1016/j.patcog.2021.107861>
- [5] Huang, Z., Zhang, T., Heng, W., Shi, B., & Zhou, S. (2022, October). Real-time intermediate flow estimation for video frame interpolation. In *European Conference on Computer Vision* (pp. 624-642). Cham: Springer Nature Switzerland.
doi: https://doi.org/10.1007/978-3-031-19781-9_36
- [6] Gang, G. H. (2017). Aerial photography techniques using drones. *Broadcasting and Media Magazine*, 22(2), 18-30.
- [7] Shimizu, J., Sun, H., & Katto, J. (2022, February). Forward and backward warping for optical flow-based frame interpolation. In *2022 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)* (pp. 082-086). IEEE.
doi: <https://doi.org/10.1109/icaaic54071.2022.9722682>
- [8] Niklaus, S., & Liu, F. (2020). Softmax splatting for video frame interpolation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5437-5446).
doi: <https://doi.org/10.1109/cvpr42600.2020.00548>
- [9] Bao, W., Lai, W. S., Ma, C., Zhang, X., Gao, Z., & Yang, M. H. (2019). Depth-aware video frame interpolation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3703-3712).
doi: <https://doi.org/10.1109/cvpr.2019.00382>
- [10] Jiang, H., Sun, D., Jampani, V., Yang, M. H., Learned-Miller, E., & Kautz, J. (2018). Super slo-mo: High quality estimation of multiple intermediate frames for video interpolation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 9000-9008).
doi: <https://doi.org/10.1109/cvpr.2018.00938>
- [11] Zhou, T., Tulsiani, S., Sun, W., Malik, J., & Efros, A. A. (2016). View synthesis by appearance flow. In *Computer Vision - ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11 - 14, 2016, Proceedings, Part IV 14* (pp. 286-301). Springer International Publishing.
doi: https://doi.org/10.1007/978-3-319-46493-0_18
- [12] Hu, P., Niklaus, S., Sclaroff, S., & Saenko, K. (2022). Many-to-many splatting for efficient video frame interpolation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3553-3562).
doi: <https://doi.org/10.1109/cvpr52688.2022.00354>
- [13] Park, J., Ko, K., Lee, C., & Kim, C. S. (2020). Bmbc: Bilateral motion estimation with bilateral cost volume for video interpolation. In *Computer Vision - ECCV 2020: 16th European Conference, Glasgow, UK, August 23 - 28, 2020, Proceedings, Part XIV 16* (pp. 109-125). Springer International Publishing.
doi: https://doi.org/10.1007/978-3-030-58568-6_7
- [14] Li, Z., Zhu, Z. L., Han, L. H., Hou, Q., Guo, C. L., & Cheng, M. M. (2023). Amt: All-pairs multi-field transforms for efficient frame interpolation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9801-9810).
doi: <https://doi.org/10.1109/cvpr52729.2023.00945>
- [15] Pont-Tuset, J., Perazzi, F., Caelles, S., Arbeláez, P., Sorkine-Hornung, A., & Van Gool, L. (2017). The 2017 davis challenge on video object segmentation. *arXiv preprint arXiv:1704.00675*.
- [16] Stergiou, A., & Poppe, R. (2022). Adapool: Exponential adaptive pooling for information-retaining downsampling. *IEEE Transactions on Image Processing*, 32, 251-266.
doi: <https://doi.org/10.1109/tip.2022.3227503>

저 자 소 개



오 명 훈

- 2023년 8월 : 경희대학교 전자공학과 공학사
- 2023년 9월 ~ 현재 : 경희대학교 전자정보융합공학과 공학석사
- ORCID : <https://orcid.org/0009-0001-9935-5922>
- 주관심분야 : 딥러닝, 영상처리, 멀티미디어 시스템

저 자 소 개



이 성 배

- 2019년 2월 : 경희대학교 전자공학과 공학사
- 2021년 2월 : 경희대학교 전자정보융합공학과 공학석사
- 2021년 3월 ~ 현재 : 경희대학교 전자정보융합공학과 박사과정
- ORCID : <https://orcid.org/0000-0002-7392-6870>
- 주관심분야 : 딥러닝, 영상처리, 멀티미디어 시스템



김 규 현

- 1989년 2월 : 한양대학교 전자공학과 공학사
- 1992년 9월 : 영국 University of Newcastle upon Tyne 전기전자공학과 공학석사
- 1996년 7월 : 영국 University of Newcastle upon Tyne 전기전자공학과 공학박사
- 1996년 ~ 1997년 : 영국 University of Sheffield, Research Fellow
- 1997년 ~ 2006년 : 한국전자통신연구원 대화형미디어연구팀장
- 2006년 ~ 현재 : 경희대학교 전자정보대학 교수
- ORCID : <http://orcid.org/0000-0003-1553-936X>
- 주관심분야 : 디지털 방송, 영상처리, 멀티미디어 통신, 디지털 대화형 방송