

딥러닝 기반 비디오 프레임 보간 기술 연구 동향

□ 허진강, 윤기환, 김성제, 정진우 / 한국전자기술연구원

요약

비디오 프레임 보간 기술은 연속되어 있는 두 개의 프레임 사이의 중간 프레임을 생성하는 기술로 비디오의 프레임율을 늘리거나 슬로우 모션 영상을 생성 시 사용된다. 최근 딥러닝 기술의 발전에 따라 다양한 알고리즘의 비디오 프레임 보간 기술이 연구되고 있다. 본 고에서는 이러한 알고리즘들을 커널 기반 방식과 플로우 기반 방식으로 분류하고, 각 범주에 속하는 대표적인 알고리즘들의 특징 및 한계점에 대해 살펴본다.

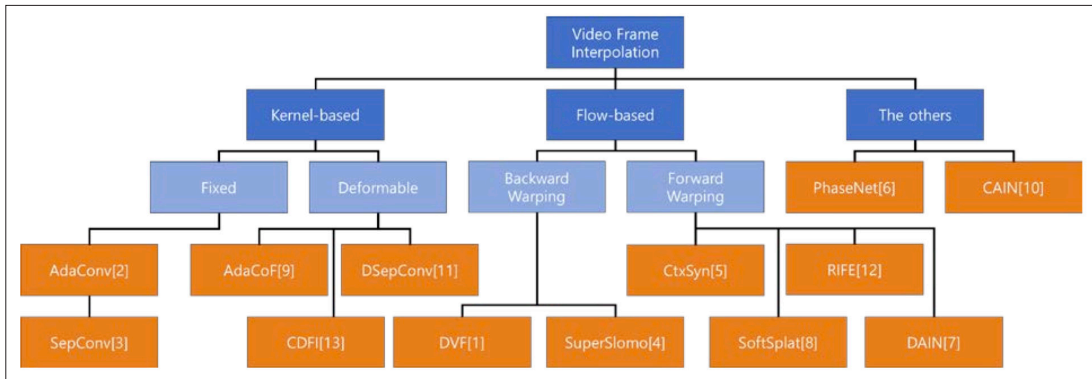
1. 서론

비디오에서 초당 재생되는 프레임의 횟수인 프레임율(Frame Rate)은 디지털 비디오의 필수 구성 요소 중 하나이다. 더 높은 해상도로 인해 동영상 품질이 향상된다는 인식과 마찬가지로 프레임율을 높이는 것도

동영상 품질에 중대한 영향을 미치게 된다. 이로 인해 MPEG(Motion Picture Expert Group)과 같은 국제 표준에서 영상의 비디오 포맷을 정의할 시 해상도 뿐만 아니라 프레임율도 함께 지정하고 있다[14].

비디오 프레임 보간(Video Frame Interpolation, 이하 VFI) 기술은 연속되어 있는 두 개의 프레임 사이의 그 중간 프레임을 생성하는 기술로 비디오의 프레임율을 늘리거나 슬로우 모션(Slow Motion) 영상을 생성 시 사용된다. VFI는 낮은 프레임율을 가지는 영상이 시각적으로 움직임의 불연속성(Judder)을 보일 때 프레임율을 증가시켜 영상을 부드럽게 재생할 수 있어 텔레비전이나 셋톱박스 등에서는 일반적으로 적용되고 있다. 또한 VFI 기술을 적용하여 프레임율 변환이나 슬로우 모션을 생성하는 다양한 소프트웨어가 시판되고 있다[16, 17].

※ 이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2021-0-00087, SD/HD급 저화질 미디어의 고품질 변환 기술 개발)

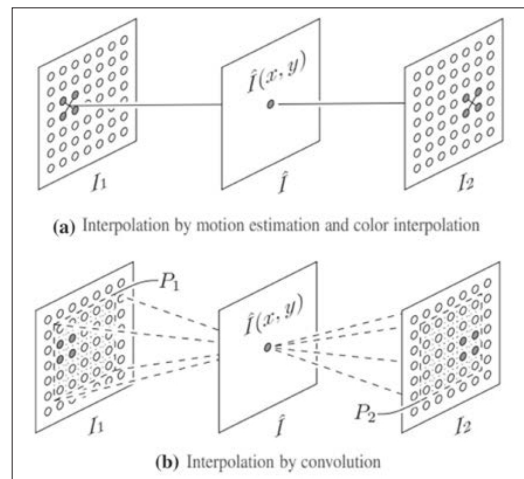


<그림 1> 딥러닝 기반 다양한 VFI 기술

현재 딥러닝 기술은 영상, 음성과 같은 미디어 분야에서 전통적인 핸드 크래프트 방법에 비해 괄목할 만한 발전을 이루었으며 실산업에서 활발히 활용되고 있다. VFI도 마찬가지로 2015년 첫 딥러닝 기술이 적용된 논문이 발표된 이래로 전통적인 방법에 비해 큰 성능 향상을 달성하였다[14]. 이로 인해 최근 연구는 대부분 딥러닝에 의한 VFI에 집중되어 있고 본 고에서는 딥러닝 기반 VFI에 대하여 서술한다.

전통적인 VFI는 움직임 또는 플로우(Optical Flow) 추정과 합성 단계로 구성되어 있다. 마찬가지로 딥러닝 기반 VFI도 대부분의 알고리즘은 이와 같은 형태로 구성되어 있으며 전통적인 방법이 아닌 딥러닝 네트워크를 사용하는 것에 그 차별점이 있다. 프레임 간 배경 및 객체의 움직임을 얼마나 정확히 추정해 내는 것이 성능을 가른다고 볼 수 있다. 움직임 추정을 위해 딥러닝 기반 VFI는 크게 커널(Kernel) 기반 알고리즘과 플로우 기반 알고리즘으로 구분할 수 있다. 커널 기반 알고리즘은 2017년 Niklaus에 의해 처음으로 발표되었으며, 그 후 Separable이나 Deformable 등 많은 개선된 알고리즘들이 발표되고 있다[2, 3, 9, 11]. 커널 기반 알고리즘은 보간 프레임에 대한 픽셀 합성을 두 개의 입력 프레임에 대한 지역(Local) 컨볼루션으로 보고 픽셀마

다 적응적으로 컨볼루션 커널을 생성하도록 한다(<그림 2-(b)>). 이런 컨볼루션 커널은 입력 프레임 사이의 움직임과 프레임 합성을 위한 계수(coefficients)를 모두 포함한다. 한편 딥러닝 기반 플로우 산출 알고리즘은 Smagt에 의해 FlowNet이란 이름으로 2015년 ICCV에서 발표되었으며[15], 2017년에는 플로우 기반 VFI가 Liu에 의해 CVPR에서 발표되었다[1]. 커널과 플로우를 동시에 사용하여 성능을 높이려는 시도도 존재하는 등 최근에도 관련 연구가 활발하게 진행 중이다[7]. 이외에



<그림 2> 플로우 기반 보간 기술(a)과 커널 기반 보간 기술(b)

도 커널 또는 플로우를 사용하지 않는 PhaseNet이나 CAIN 과 같은 알고리즘들도 발표되었다[6, 10].

본 고에서는 커널 기반과 플로우 기반 주요 알고리즘들에 대하여 살펴본다.

II. 딥러닝 기반 비디오 프레임 보간 기술

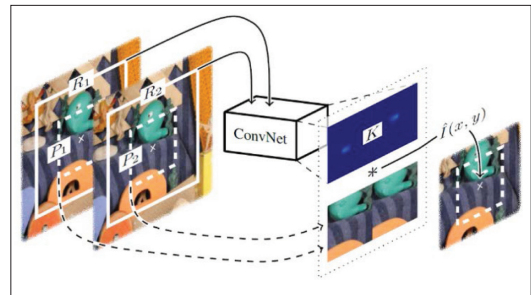
1. 커널 기반 비디오 프레임 보간 기술

전통적인 VFI는 플로우 기반 알고리즘으로 객체의 움직임을 추정하고 추정된 위치의 주변 화소 값을 가중 평균하여 중간 프레임의 해당 화소를 생성한다 (<그림 2-(a)>). 하지만 전통적인 플로우 기반 VFI는 큰 움직임, 폐색(Occlusion), 밝기 변화와 같은 영역에서 플로우 추정의 부정확성으로 인해 이와 같은 영역을 처리하는데 어려움이 있다. 부정확한 플로우 추정의 문제를 피하기 위해 Niklaus는 최초의 커널 기반 알고리즘인 Adaptive Convolution(AdaConv)을 2017년 CVPR에서 발표하였다[2]. AdaConv는 두 입력 프레임의 모든 화소(pixel)에 대하여 화소마다 적응적으로 컨볼루션 할 수 있는 지역(local) 컨볼루션 커널을 추정하고, 추정된 커널과 두 입력 프레임의 패치에 대하여 컨볼루션 함으로써 중간 프레임을 보간한다(<그림 2-(b)>). AdaConv는 화소마다 적응적인 커널 추정을 위해 <그림 3>의 ConvNet인 Convolution Neural Network(CNN)를 사용하였다. 논문에서 ConvNet의 입력 R1, R2는 79x79 크기의 패치이고 출력은 2D 커널로 41x41 크기를 갖는다. ConvNet에 의해 추정된 커널 K_1, K_2 를 이용하여 <수식 1>과 같은 컨볼루션 연산을 통해 (x, y) 위치에 대한 중간 프레임을 생성한다. <수식 1>의 *은

컨볼루션을 의미한다. AdaConv는 커널 크기만큼의 주변 화소들을 반영하기 때문에 패치 내의 움직임에 대하여 효과적으로 처리할 수 있다. AdaConv는 최초로 딥러닝을 사용한 커널 기반 VFI라는 것에 그 의의가 있으며, 이후 다양한 알고리즘들이 AdaConv의 구조를 차용하여 개발되고 있다. 그러나 AdaConv는 GPU 메모리 한계로 인하여 커널 크기에 제한이 있어 커널 크기를 벗어나는 움직임을 다루기 어렵다는 한계가 있다. 실제로 41x41의 2D 커널일 경우 1080p 영상을 기준으로 약 26GBytes GPU 메모리가 필요한데 이는 대부분의 GPU에서 동작이 불가능하다.

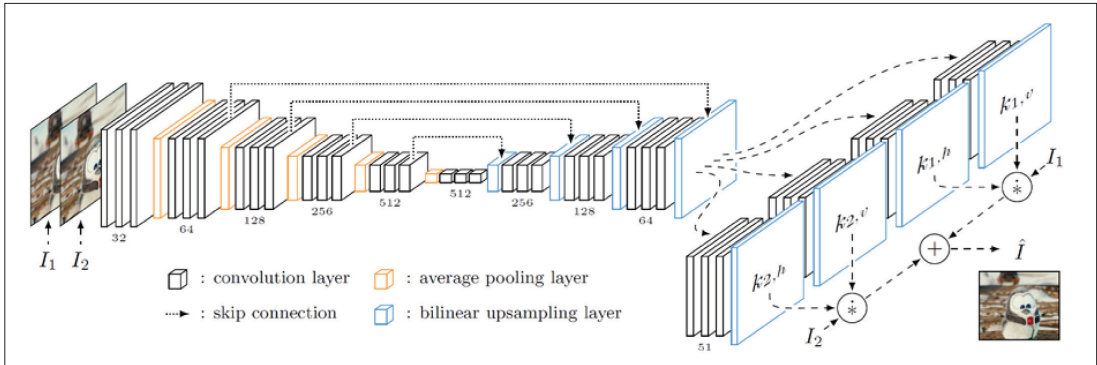
$$\hat{I}(x, y) = K_1(x, y) * P_1(x, y) + K_2(x, y) * P_2(x, y)$$

<수식 1> AdaConv[2]의 중간 프레임 보간 수식



<그림 3> AdaConv[2]의 전체 구조

이런 문제점을 해결하기 위해 Niklaus는 2017년 ICCV 에서 2D 커널을 2개의 수직, 수평의 1D 커널로 대체하는 Separable Convolution(SepConv)를 발표하였다[3]. SepConv는 이로 인해 N^2 의 커널 크기를 $2N$ 으로 감소시켰으며, 1080p 영상 기준으로 커널에 필요한 용량은 1.27GBytes로 AdaConv 대비 필요 메모리는 약 1/20에 불과하다. 필요 메모리 감소로 SepConv는 1D 커널 크기를 51로 증가시켜 AdaConv보다 더

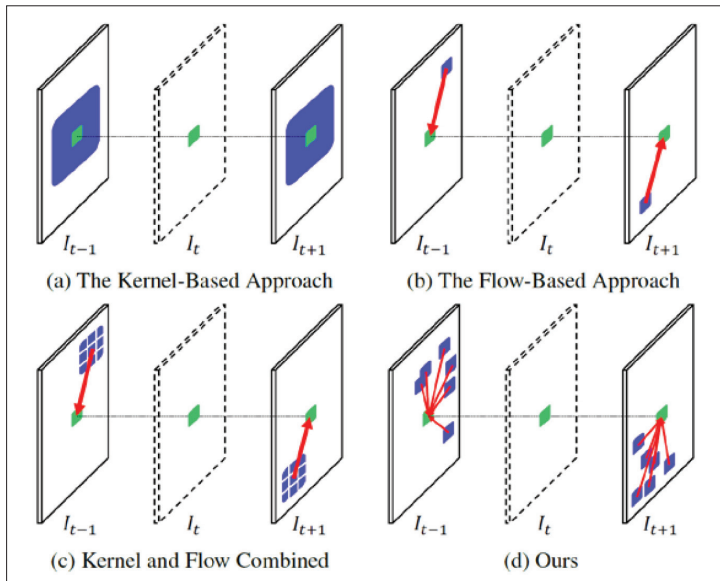


<그림 4> SepConv[3]의 전체 네트워크 구조

큰 움직임을 추정할 수 있게 하였다. 또한 SepConv는 <그림 4>와 같이 1D 커널 추정을 위해 직접 연결(Skip Connection) 구조를 갖는 U-Net[20] 형식의 네트워크를 사용하였으며, 이 구조는 이후 다양한 커널 기반 VFI에 사용되었다[9, 11, 13]. SepConv는 비록 분리된 1D 컨볼루션을 사용했음에도 불구하고 우수한 네트워크 구조와 AdaConv보다 큰 커널 크기로 AdaConv에 비

해 우수한 성능을 제공하였다. 하지만 여전히 커널 크기를 넘어서는 움직임을 갖는 객체에 대해서는 보간 성능이 저하되는 한계를 가지고 있다.

커널 크기의 제한으로 인한 문제를 해결하기 위해 Lee는 커널의 형태가 영상의 특징에 따라 변형되는 Adaptive Collaboration of Flow(AdaCoF) 알고리즘을 2020년 CVRP에서 발표하였다[9]. <그림 5>는 기

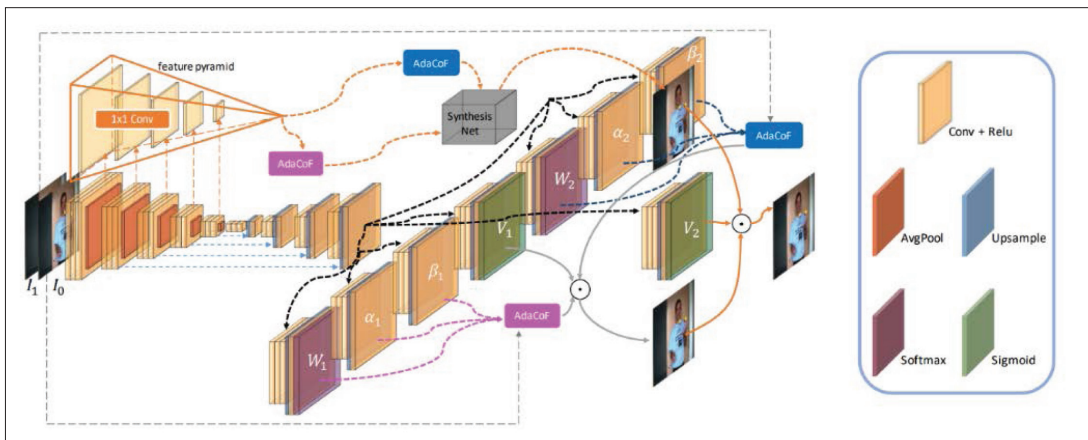


<그림 5> AdaCoF[9] (d)와 기존 방법들과의 비교

존 VFI들과 AdaCoF 알고리즘의 차이점을 보여 준다. SepConv와 같은 기존의 커널 기반 VFI는 입력 영상의 참조 영역이 크다는 장점이 있지만 항상 고정된 영역만 참조하여 움직임이 그 영역이 벗어날 시 보간 성능이 저하된다는 문제점이 있다(〈그림 5-(a)〉)[2, 3]. 플로우 기반 VFI의 경우에는 참조 영역이 제한되지 않지만 복잡한 움직임 등에서 플로우의 정확도가 낮을 경우 성능이 크게 저하된다는 문제가 있다(〈그림 5-(b)〉)[4]. 커널과 플로우를 조합한 알고리즘의 경우 플로우 주변으로 작은 커널을 사용하여 플로우 알고리즘에 비해 일부 주변 화소를 더 참조하지만 커널의 영역은 고정되어 있기 때문에 더 개선할 여지는 남아있다(〈그림 5-(c)〉)[7, 21]. 기존 알고리즘들의 문제를 해결하기 위해 AdaCoF는 각 커널 위치에 (x,y) 오프셋을 추가하여 커널의 위치에 대한 제한을 제거하였다(〈그림 5-(d)〉). 또한 학습 단계에서 커널 확장 변수(Dilation)를 추가해 더 넓은 범위의 오프셋을 가지게 하였다. 이 알고리즘은 디포머블(Deformable) 컨볼루션과 유사한 개념을 가지고 있으나 디포머블 컨볼루션이 모든 화소에서 동일한 가중치를 가지는데 반하여 화소마다 커널의 가중치가 계

산된다는 점이 차이가 있다. 이런 성질 때문에 본 고에서는 AdaCoF를 커널 기반 VFI로 분류하였다. 2020년 AAAI에 유사한 개념을 가진 Deformable Separable Convolution(DSepConv)이 Cheng에 의해 제안되었는데 AdaCoF와의 주요 차이점은 Separable 컨볼루션을 사용하였다는 데에 있다[11].

Ding은 2021년 CVPR에서 기존 VFI는 대부분 거대한 네트워크를 사용하기 때문에 과파라미터화(over-parameterized) 되었고 이로 인해 모바일 장비 같은 소형 기기에서의 사용이 제한되는 점을 지적했다. 그들은 네트워크 크기를 줄이기 위해 L1 정규화 기반의 프루닝(pruning) 기법을 AdaCoF에 적용하였다. 프루닝 적용 결과 AdaCoF 대비 파라미터의 수는 약 1/9로 줄었지만, 성능 저하는 평균 약 0.3dB로 크지 않음을 보였다. 또한 압축된 네트워크의 성능을 개선하기 위해 커널 추출 네트워크의 인코더에서 계층별 문맥 정보를 피라미드 구조로 추출하고 추출된 문맥 정보를 별도의 합성(synthesis) 네트워크를 통해 문맥 기반 중간 프레임 생성하도록 하였다. 문맥 기반 중간 프레임은 압축된 네트워크를 통해 보간된 중간 프레임과 재 합성되어 최종



〈그림 6〉 CDFI[13] 전체 네트워크 구조

중간 프레임을 보간하도록 한다. 이 논문은 최초로 기존 알고리즘들이 과파라미터화 되었고 개선할 여지가 많다는 것을 보였다는데 그 의미가 있다.

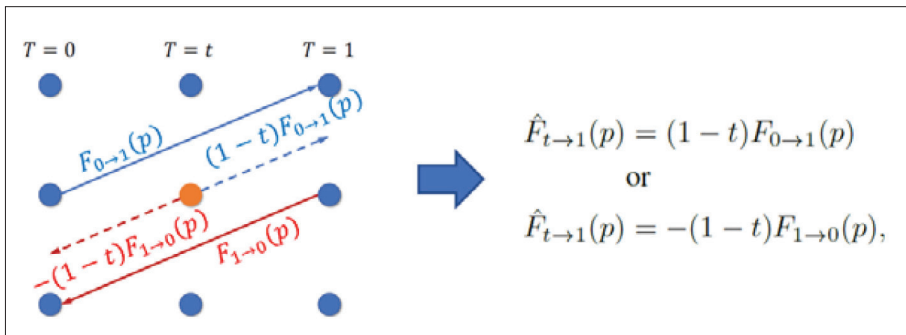
2. 플로우 기반 비디오 프레임 보간 기술

플로우 기반 VFI는 두 개의 연속적인 입력 프레임(I_0, I_1)과 중간 프레임($I_t, 0 < t < 1$) 사이의 플로우($\hat{F}_{t \rightarrow 0}, \hat{F}_{t \rightarrow 1}$)를 추정하고, 추정된 플로우를 이용해서 중간 프레임을 보간하는 기술이다. 플로우 기반 VFI는 넓은 범위의 수용 영역(Receptive Field)을 갖는 플로우 추정 네트워크 설계가 가능하다. 따라서 커널 기반 VFI와 달리 커널 크기에 대한 제한이 없기 때문에 더 역동적인 움직임에 대한 추정이 가능하다.

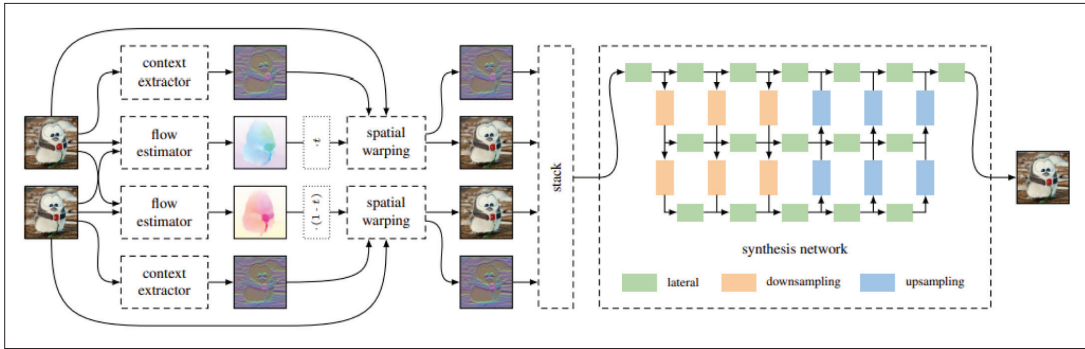
2018년 CVPR에서 Jiang은 두 개의 연속적인 입력 프레임 사이에 다중 프레임을 보간하는 Super Slow Motion(SuperSlomo)을 발표하였다[4]. 커널 기반 VFI는 여러 프레임을 생성 할 경우 반복적인 방법으로 보간해야 한다. 예를 들면 $I_{0.5}$ 프레임을 보간 후에 I_0 프레임과 $I_{0.5}$ 프레임을 이용해서 $I_{0.25}$ 프레임을 보간한다. 이 방법은 최소 두 가지 문제점을 내제하고 있는데, 먼저 반복적인 방법을 통해 다중 프레임을 보간하기 때

문에 병렬처리가 불가능하다. 따라서 속도 저하 문제를 발생시킨다. 두 번째로 $I_{0.5}$ 프레임에서 발생하는 오차가 $I_{0.25}$ 에 누적되므로 에러에 취약하다. 이러한 문제를 해결하기 위해 SuperSlomo는 두 입력 프레임 I_0, I_1 의 양방향 플로우($F_{t \rightarrow 0}, F_{t \rightarrow 1}$)로부터 임의 시간 $I_{t, 0 < t < 1}$ 프레임의 플로우($\hat{F}_{t \rightarrow 0}, \hat{F}_{t \rightarrow 1}$)를 직접적으로 근사화 하는 방법을 제시함으로써 다중 프레임 VFI를 병렬처리 가능하도록 제안했다.

일반적으로 플로우 기반 VFI는 추정된 플로우를 통해 두 입력 프레임을 와핑하고 이후 중간 프레임을 합성한다. 이때 와핑 방법으로는 전방향(forward) 와핑과 역방향(backward) 와핑이 존재하는데 앞서 설명한 SuperSlomo는 역방향 와핑에 해당한다. 역방향 와핑 방법의 경우 중간 프레임으로부터 두 입력 프레임 방향의 플로우를 추정하기 때문에 중간 프레임의 모든 화소가 플로우를 가진 상태로 합성된다. 따라서 합성 과정에서 구멍(Hole)이 발생하지 않는 장점이 있다. 반면 전방향 와핑 방법은 두 입력 프레임으로부터 중간 프레임 방향으로 플로우를 추정하기 때문에 합성 후 구멍이 생길 수 있는 단점이 있다. 2018년 CVPR에서 Niklaus는 전방향 와핑의 문제점을 해결하기 위해 문맥(Context) 정보와 합성 네트워크를 사용하는 Context-Aware



<그림 7> SuperSlomo의 임의 시간 t 프레임의 플로우 근사화 방법

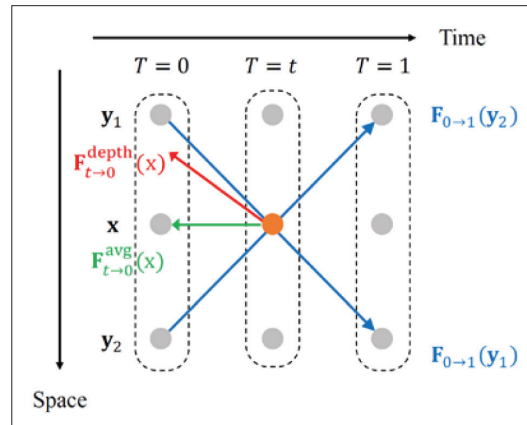


<그림 8> CtxSyn의 전체 구조

Synthesis(CtxSyn)[5]를 제안했다.

<그림 8>은 CtxSyn의 전체 네트워크 구조이다. 문맥 정보를 추출하기 위해 ResNet-18의 Conv1 계층을 사용했으며, 추출된 문맥 정보는 두 입력 프레임과 함께 전방향 와핑된다. 전방향 와핑은 임의의 시간 t 프레임으로 근사화된 플로우를 사용한다. 이후 와핑 된 정보들은 GridNet으로 구성된 합성 네트워크의 입력으로 사용되며, 이 때 전방향 와핑 과정에서 발생할 수 있는 구멍들은 합성 네트워크에서 보상된다. CtxSyn는 기존 논문들과 달리 최초로 문맥 정보와 합성 네트워크를 사용하였고, 이후 다양한 알고리즘에서 문맥 정보와 합성 네트워크를 사용했다[13, 8].

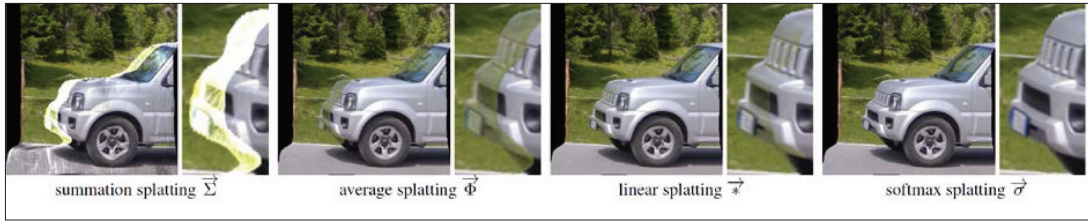
다음으로 소개할 알고리즘은 깊이 정보를 활용해서 플로우 근사화 성능을 향상 시킨 2019년 CVPR에 제안한 Bao의 Depth-Aware Video Frame Interpolation(DAIN)[7]이다. DAIN은 대부분의 알고리즘들이 명확하지 않은 방법으로 플로우를 근사화 하는 것에 문제를 제기한다. 기존의 알고리즘들은 대부분 중간 프레임의 플로우를 근사화 할 때 보간 화소를 지나는 다중 플로우들을 단순 평균을 취함으로써 플로우를 추정했다. 하지만 이런 방법은 객체의 깊이 정보를 고려하지 않기 때문에 객체 간 겹침이 발생하는 영역에서 부정



<그림 9> DAIN의 깊이 정보를 이용한 플로우 근사화 방법

확한 플로우를 생성한다. 따라서 이 문제를 해결하고자 DAIN은 깊이 정보를 활용하여 보간 화소의 플로우 정확도를 향상시킨다. <그림 9>는 DAIN에서 제안한 플로우 근사화 방법이다. 보간 화소인 주황색 화소는 두 개의 플로우가 지나는 것을 볼 수 있다. 이 때 두 개의 플로우를 단순 평균 취하여 추정할 경우 초록색 플로우로 근사화 되지만, 깊이 정보를 반영하여 플로우(y_1)이 더 높은 가중치를 가지고 있다고 했을 때 빨간색 플로우로 근사화 되는 것을 볼 수 있다.

한편 2020년 CVPR에서 Niklaus는 전방향 와핑 시

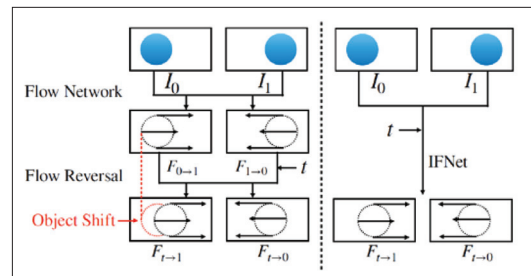


<그림 10> Niklaus가 제시한 4가지 Splatting 방법의 결과물

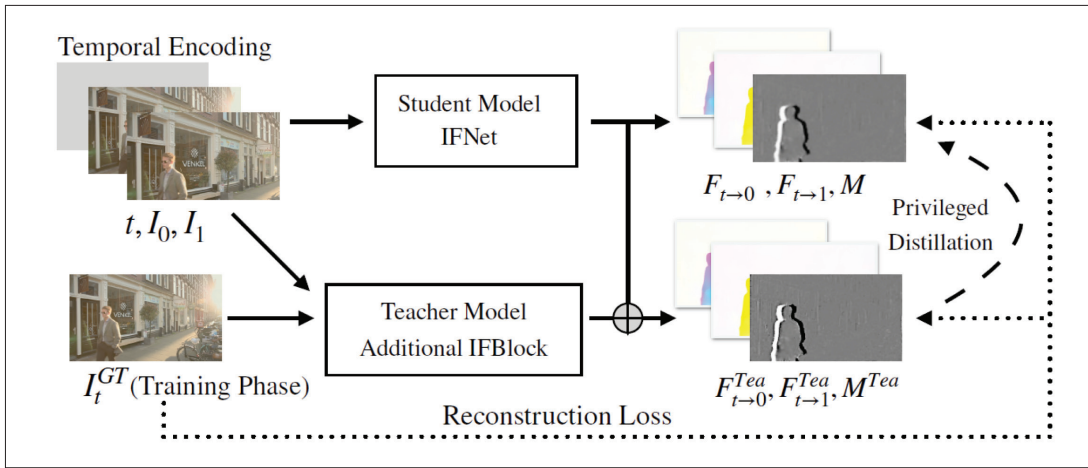
입력 프레임의 다중 화소를 보간 프레임의 단일 화소에 매핑할 때, 어떤 화소를 선택할지 결정해야하는 모호성 문제를 수학적으로 정의하고 이것을 해결하기 위해 Softmax Splatting(SoftSplat) 방법을 제안했다. Splatting이란 다중 화소를 단일 화소에 매핑하는 방법으로 Niklaus는 Summation Splatting, Average Splatting, Linear Splatting, Softmax Splatting 총 4가지를 제시하였다. 먼저 Summation Splatting 방법이란 다중 화소를 합하여 단일 화소에 매핑하는 방법이다. 이것은 최대 화소 값을 초과하는 문제로 보간 화소가 포화되는 문제가 있다. 두 번째로 Average Splatting 방법은 다중 화소를 평균 취함으로써 단일 화소에 매핑하는 방법으로 화소값이 포화되는 문제는 없지만 배경과 객체가 결합되는 문제가 발생한다. 세 번째로 Linear Splatting 방법은 화소 간 중요도 정보(Z)를 이용해서 다중 화소에 가중치를 취하여 단일 화소에 매핑하는 방법이다. 이 방법은 비교적 배경과 객체가 정확히 분리되지만, 중요도 정보를 선형적으로 반영하기 때문에 왜곡하는 과정에서 변형(Translation)되는 문제가 있다. 좀 더 자세히 설명하면, 중요도 정보를 깊이 정보의 역수라고 가정할 때, 가까이 있는 객체의 깊이 정보가 1이면 $Z=1/1$ 이 되고 배경의 깊이 정보가 10이면 $Z=1/10$ 이 되어 분명한 차이를 볼 수 있다. 반면 객체의 깊이 정보가 101이면 $Z=1/101$ 이 되고 배경의 깊이 정보가 110이면 $Z=1/110$ 이 되어서 객체와 배경의 중요도 값이 큰 차이

가 없는 것을 볼 수 있다. 이런 문제를 해결하기 위해 마지막으로 중요도 정보에 지수함수를 취하는 Softmax Splatting 방법을 제안했다. Softmax Splatting 방법은 중요도 정보에 지수함수를 취하여 중요도 정보의 변형을 방지하여 다른 Splatting 방법에 비해 높은 성능을 제공할 수 있었다(<그림 10>).

마지막으로 2020년 ArXiv에 Huang는 실시간으로 중간 프레임을 보간할 수 있는 Real-Time Interpolation Flow Estimation(RIFE)[12]를 제안했다. Huang은 기존 논문들이 서드 파티의 최신 플로우 추정기를 사용하는 것은 VFI 기술 응용 분야에 최적화 되지 않았다고 주장한다. <그림 11>을 보면 기존 논문들의 경우 두 입력 프레임 간 양방향 플로우를 추정하고 중간 프레임의 플로우를 근사화 하기 위해 반전(Flow Reversal)하고 이동(Object Shift) 시킨다. 이럴 경우 양방향에서 오는 플로우가 중간 프레임에서 정확히 일치하지 않을 수 있다.



<그림 11> 기존 논문(왼쪽)들과 RIFE(오른쪽)의 플로우 추정 방법 비교



<그림 12> RIFE[12]의 Privileged Distillation 학습 방법

반면 RIFE는 직접 설계한 IFNet 네트워크를 통해 두 입력 프레임으로부터 직접 중간 프레임의 플로우를 추정하여 양방향 플로우를 근사화해서 생기는 불일치 현상을 근본적으로 제거하였다. 또한 이 방법은 플로우를 한 번만 산출하고 경량화로 인해 기존 플로우 추정기(PWC-Net[18], RAFT[19]) 대비 약 6~15배 정도 속도를 향상시킬 수 있다. 이로 인해 640x480 크기의 영상에 대하여 실시간 보간을 달성하였다.

또한 RIFE는 속도 향상으로 발생할 수 있는 성능 저하 문제를 Privileged Distillation 학습 모델을 통해 해결했다. Privileged Distillation 학습 모델은 학습 단계에서만 사용되는 방법으로 Teacher 모델과 Student 모델로 분류해서 학습하는 방법이다. RIFE는 Teacher 모델로서 IFNet 네트워크에 하나의 계층을 추가하여 중간 프레임의 원본 데이터에 접근할 수 있도록 설계했다. 학습 과정에서 Student 모델인 IFNet 네트워크는 Distillation Loss를 통해 원본 데이터에 접근 가능한 Teacher 모델의 지도를 받아 더 높은 성능을 나타낼 수 있다. <그림 12>는 RIFE의 Privileged Distillation 학습 방법을 표현한 그림이다.

III. 맺음말

본 기고에서 우리는 딥러닝 기반 VFI 기술들에 대해 커널 기반 VFI, 플로우 기반 VFI로 분류하고 주요 기술들에 대한 특징들을 살펴보았다. 전통적인 VFI 기술은 복잡한 움직임, 폐색 영역 등의 난제로 인해 만족할 만한 품질을 보장하기 어려워 적용 범위가 제한적이었다. 그러나 딥러닝의 등장으로 VFI 성능은 비약적으로 발전하였고 이를 이용한 소프트웨어가 등장기도 하는 등 활용 영역이 넓어지고 있다. 최근에는 매우 높은 화질을 요구하는 방송 콘텐츠에서도 딥러닝 기반 VFI를 이용하여 프레임율을 증가시키려는 시도도 존재한다. 그러나 고해상도 영상과 복잡한 움직임에서는 여전히 만족할 만한 품질을 얻지 못하고 있으며 개선할 여지는 여전히 많이 남아있다[12]. 또한 딥러닝 기반 알고리즘은 높은 연산량을 가지기 때문에 실사용을 위하여는 모델 압축이나 경량화에 대한 추가적인 연구가 필요하다. 따라서 향후 성능 개선과 경량 알고리즘의 발전으로 딥러닝 기반 VFI 기술이 동영상 플레이어, 슬로우 모션 생성, 라이브 스트리밍, 모바일 등 다양한 분야에서 광범위하게 사용되길 기대한다.

참고 문헌

- [1] Liu, Ziwei, et al. "Video frame synthesis using deep voxel flow," Proceedings of the IEEE International Conference on Computer Vision, 2017.
- [2] Niklaus, Simon, Long Mai, and Feng Liu. "Video frame interpolation via adaptive convolution," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [3] Niklaus, Simon, Long Mai, and Feng Liu. "Video frame interpolation via adaptive separable convolution," Proceedings of the IEEE International Conference on Computer Vision, 2017.
- [4] Niklaus, Simon, and Feng Liu. "Context-aware synthesis for video frame interpolation," Proceedings of the IEEE conference on computer vision and pattern recognition, 2018.
- [5] Jiang, Huaizu, et al. "Super slomo: High quality estimation of multiple intermediate frames for video interpolation," Proceedings of the IEEE conference on computer vision and pattern recognition, 2018.
- [6] Meyer, Simone, et al. "Phasenet for video frame interpolation," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.
- [7] Bao, Wenbo, et al. "Depth-aware video frame interpolation," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.
- [8] Niklaus, Simon, and Feng Liu. "Softmax splatting for video frame interpolation," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.
- [9] Lee, Hyeongmin, et al. "Adacof: Adaptive collaboration of flows for video frame interpolation," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.
- [10] Choi, Myungsub, et al. "Channel attention is all you need for video frame interpolation," Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, No. 07, 2020.
- [11] Cheng, Xianhang, and Zhenzhong Chen. "Video frame interpolation via deformable separable convolution," Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, No. 07, 2020.
- [12] Huang, Zhewei, et al. "Rife: Real-time intermediate flow estimation for video frame interpolation." arXiv preprint arXiv:2011.06294 (2020).
- [13] Ding, Tianyu, et al. "Cdfi: Compression-driven network design for frame interpolation," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.
- [14] BT,2020-2 (10/2015) : Parameter values for ultra-high definition television systems for production and international programme exchange
- [15] Dosovitskiy, Alexey, et al. "Flownet: Learning optical flow with convolutional networks," Proceedings of the IEEE international conference on computer vision, 2015.
- [16] <https://www.svp-team.com/>
- [17] <https://topten.ai/convert-videos-to-60-fps-with-ai-interpolation/>
- [18] Sun, Deqing, et al. "Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume," Proceedings of the IEEE conference on computer vision and pattern recognition, 2018.
- [19] Teed, Zachary, and Jia Deng. "Raft: Recurrent all-pairs field transforms for optical flow," European conference on computer vision, Springer, Cham, 2020.
- [20] O. Ronneberger, P. Fischer, and T. Brox, U-net: Convolutional networks for biomedical image segmentation, In MIC-CAI, 2015.
- [21] Bao, Wenbo, et al. "Memc-net: Motion estimation and motion compensation driven neural network for video interpolation and enhancement," IEEE transactions on pattern analysis and machine intelligence 43,3 (2019): 933-948.

필자 소개



허진강

- 2015년 : 강릉원주대학교 학사
- 2018년 : 연세대학교 석사
- 2017년 ~ 현재 : 한국전자기술연구원 연구원
- 주관심분야 : 인공지능 기반 영상 처리



윤기환

- 2020년 : 단국대학교 전자전기공학부 학사
- 2020년 ~ 현재 : 서울시립대학교 전자전기컴퓨터공학부 석박사과정



김성제

- 2011년 : 연세대학교 전기전자공학과 박사
- 2011년 ~ 2015년 : 삼성전자 S.LSI 책임연구원
- 2015년 ~ 현재 : 한국전자기술연구원 책임연구원
- 주관심분야 : 미디어 신호처리, 인공지능



정진우

- 2011년 : 연세대학교 전기전자공학과 박사
- 2011년 ~ 2015년 : 삼성전자 VD사업부 책임연구원
- 2016년 ~ 현재 : 한국전자기술연구원 책임연구원
- 주관심분야 : 미디어 신호처리, 인공지능