

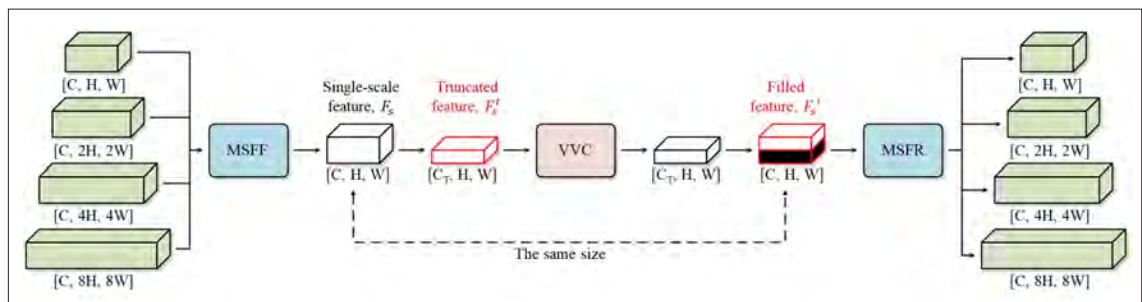
VCM을 위한 QP 적응적 특징 채널 절삭을 이용한 효율적인 다중스케일 특징 압축 기법

윤용욱 / 한국항공대학교 Media Communication Lab

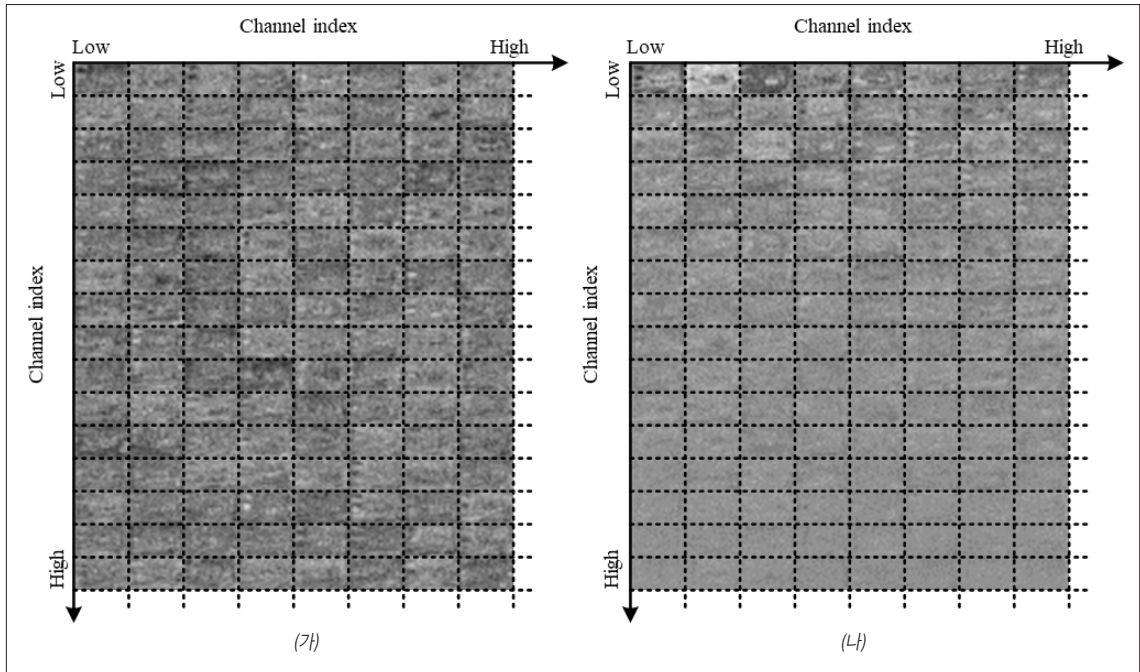
인공지능 기술을 활용한 객체검출(object detection), 객체추적(object tracking) 등의 머신비전 응용이 확산되고 있다. 따라서 이러한 머신비전 응용에서 방대한 양의 비디오 데이터를 처리하고 전송할 수 있는 표준 기술이 요구되고 있다. 사람의 시각 시스템(Human Visual System, HVS) 특성을 고려하여 설계된 기존 비디오 압축 기술은 기계의 머신비전 임무(task) 수행에는 비효율적일 수 있다. 즉, 지능형 분석에 필요한 중요한 정보가 손실되거나 분석에 불필요한 정보가 전송될 수 있다. 이에 따라 MPEG에서는 기계가 사용하는 비디오를 보다 효과적으로 부호화할

수 있는 새로운 비디오 압축 표준인 VCM(Video Coding for Machines) 표준 개발을 활발히 진행하고 있다.

본 연구는 머신비전 네트워크의 분할 포인트(split point)로부터 추출된 다중스케일 특징을 효율적으로 압축하는 eMSFC(efficient Multi-Scale Feature Compression) 기법을 제안한다. 제안하는 eMSFC는 SLS(selective learning strategy)로 학습된 MSFC 네트워크로부터 중요도 순서로 나열된 단일스케일 특징을 생성하고, VVC(Versatile Video Coding)로 2D 프레임 패킹된 단일스케일 특징을 효율적으로 압축한다. 제안하



<그림 1> 제안하는 eMSFC 기법의 프레임워크



<그림 2> 2D 프레임 패킹된 특징맵의 예

는 eMSFC의 파이프라인은 <그림 1>과 같다. 제안하는 eMSFC는 머신비전 네트워크의 임의의 분할 포인트로부터 추출된 다중스케일 특징을 입력으로 한다. 다중스케일 특징은 MSFF(Multi-Scale Feature Fusion) 모듈을 통해 단일스케일 특징으로 변환된다. 이때, SLS로 인해 중요도 순서로 나열된 단일스케일 특징이 생성되기 때문에, 양자화 파라미터(QP)에 따라 중요치 않은 특징을 채널 단위로 절삭이 가능하다. 절삭된 특징은 2D 프레임으로 패킹되어 VVC로 압축되고, 디코더는 단일스케일 특징을 다중스케일 특징으로 복원하여 머신비전 네트워크로 다시 입력되어 머신비전 임무를 수행한다.

<그림 2>는 128개의 채널을 갖는 단일스케일 특징을 2D 프레임 패킹했을 때의 특징맵을 보여준다. <그림 2-(가)>는 SLS를 적용하지 않았을 때 생성되는 단일스케일 특징이며, <그림 2-(나)>는 SLS를 적용했을 때 생성되는 단일

스케일 특징이다. <그림 2>와 같이, SLS로 인해 낮은 인덱스의 채널은 복잡한 특징을 보여주지만, 높은 채널 인덱스일수록 단순한 특징을 보여준다. 이로 인해, 특징맵의 공간적 상관성이 높아져 VVC로 효율적인 압축이 가능하다.

<표 1>은 객체분할, 검출, 추적 임무에 대한 제안 방법

<표 1> 제안하는 eMSFC의 성능 평가 결과

Task	Dataset	BD-rate
Instance segmentation	OpenImagesV6	-98.72%
Object detection	OpenImagesV6	-98.34%
Object tracking	TVD1	-97.76%
	TVD2	-97.75%
	TVD3	-98.71%
	TVD-overall	-98.09%

졸업논문 소개

의 성능 평가 결과를 보여준다. 제안 방법은 모든 임무에 대해 우수한 성능을 보여준다. 제안 방법은 MPEG-VCM의 CfE(Call for Evidence) 응답 기술로 제출되었으며, 응

답 기술 중 가장 높은 성능을 보여주면서, MSFC를 활용한 특징 압축 기술의 우수함을 보였다.



윤 용 옥

- 2017년 2월 : 한국항공대학교 항공전자정보공학과 학사
- 2019년 2월 : 한국항공대학교 항공전자정보공학과 석사
- 2023년 8월 : 한국항공대학교 항공전자정보공학과 박사
- 2023년 8월 ~ 현재 : Tencent America
- 주관심분야 : 비디오 코딩, 딥러닝