특집논문 (Special Paper)
방송공학회논문지 제27권 제3호, 2022년 5월 (JBE Vol.27, No.3, May 2022)
https://doi.org/10.5909/JBE.2022.27.3.295
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

MPEG CDVA 전역 특징 서술자 압축 방법

김 준 수[a)‡], 조 원[b)], 임 근 택[c)], 윤 정 일[a)], 곽 상 운[a)], 정 순 흥[a)], 정 원 식[a)], 추 현 곤[a)], 서 정 일[a)], 최 유 경[c)]

Compression Method for MPEG CDVA Global Feature Descriptors

Joonsoo Kim[a)‡], Won Jo[b)], Guentaek Lim[c)], Joungil Yun[a)], Sangwoon Kwak[a)], Soon-heung Jung[a)], Won-Sik Cheong[a)], Hyon-Gon Choo[a)], Jeongil Seo[a)], and Yukyung Choi[c)]

요 약

본 논문은 동영상의 시각적 특징을 추출하는 MPEG CDVA 표준 기술에서 개별 프레임의 전역적인 특징을 표현하는 scalable Fisher vector (SCFV)의 새로운 압축 방법을 제안한다. CDVA 표준은 전역 특징 서술자에 대한 시간적 중복성 제거 기법을 도입하였으며, 구체적으로 부호화 단위 세그먼트 내의 SCFV 들이 서로 유사할 가능성이 높다는 점을 활용하여 SCFV에 대한 차분을 부호화하는 방식을 사용하고 있다. 그러나 SCFV의 구조적 특징에 의해 SCFV의 차분을 부호화 한 결과물이 원본 데이터보다도 용량이 큰 경우가 발생하게 된다. 이와 같은 현상을 방지하기 위해 비대칭적 SCFV의 차분 계산 방법과 변경된 SCFV 차분을 활용하여 원본 SCFV를 복원하는 새로운 방법을 제안하였다. FIVR 데이터셋을 활용한 실험결과는 전역 특징 서술자의 압축 효율이 기존 CDVA Experimental Model에 대비하여 유의미하게 증가함을 보여준다.

Abstract

In this paper, we propose a novel compression method for scalable Fisher vectors (SCFV) which is used as a global visual feature description of individual video frames in MPEG CDVA standard. CDVA standard has adopted a temporal descriptor redundancy removal technique that takes advantage of the correlation between global feature descriptors for adjacent keyframes. However, due to the variable length property of SCFV, the temporal redundancy removal scheme often results in inferior compression efficiency. It is even worse than the case when the SCFVs are not compressed at all. To enhance the compression efficiency, we propose an asymmetric SCFV difference computation method and a SCFV reconstruction method. Experiments on the FIVR dataset show that the proposed method significantly improves the compression efficiency compared to the original CDVA Experimental Model implementation.

Keyword : Visual search, descriptor, Fisher vector, compression

Ⅰ. 서 론

시각 정보 기반 영상 검색은 내용 기반 검색 (Content-based retrieval)의 한 종류로, 영상의 시각적 정보에 대한 분석을 통해 유사한 시각적 특징을 가지는 영상을 찾는 임무를 말한다. 이와 같은 임무의 수행을 위해서는 영상들의 유사 여부를 판정하기 위한 지표들이 필요하며, 이 때문에 영상의 시각적 특성을 추출하는 방법이 성능에 결정적인 영향을 주게 된다. 전통적인 시각 특징으로는 BoW, VLAD, 피셔 벡터 (Fisher vector) 등이 흔히 고려되었으며, 2012년 이후로는 딥 러닝 기반의 영상 검색 기법들이 좋은 성능을 보여주고 있다[1-5]. 시각 정보 기반 영상 검색 기술의 발달에 힘입어 MPEG은 모바일 환경에서의 이미지 기반 검색을 위한 영상 특징 서술자 추출 표준인 Compact Descriptors for Visual Search (CDVS)를 개발하여 2015년에 발표하였고, 이어서 비디오 입력에 대해 영상 특징 서술자를 추출하는 Compact Descriptors for Video Analysis (CDVA)를 2019년에 발표하였다[6,7]. CDVA는 비디오의 시각적 특성을 개별 프레임들의 시각적 특성 모음으로 표현하며, 각 프레임의 시각적 특성은 앞서 개발된 CDVS의 handcraft 영상 특징 서술자와 딥러닝 기반 영상 특징 서술자의 조합으로 표현된다. 시각적 특징 정보는 많은 경우에 연속한 다수의 프레임에 걸쳐 유지되는 경향이 있으며, 이와 같은 정보의 중복을 최소화하기 위해 시간적 중복성 제거 기법이 적용될 필요가 있다[8]. CDVA에 적용된 시간적 중복성 제거 기법은 거의 동일한 내용을 담고 있는 프레임들을 분석 대상에서 배제하는 기법과 약간의 유사성이 있는 프레임들 간의 영상 특징 중복성을 제거하는 기법으로 나누어 볼 수 있는데, 가변 길이 영상 특징 서술자에 대해 프레임 간 중복 제거를 수행하는 과정에서 서술자 압축 효율이 크게 감소하는 현상이 발생하고 있다.

본 논문에서는 가변 길이 영상 특징 서술자에 대한 프레임 간 중복성 제거를 통해 서술자 압축 효율을 개선하는 방법을 제안한다. 먼저 CDVA 기술의 구성 요소에 대한 분석을 통해 프레임 간 중복성 제거 효율이 낮은 경우를 분석하도록 한다. 이어서 가변 길이 영상 특징 서술자의 특성을 고려한 중복성 제거 방안을 설명하고, 제안된 방법의 적용에 따른 서술자 압축 효율 개선을 실험적으로 검증한다.

Ⅱ. CDVA 기반 영상 특징 서술자 추출 및 압축

1. CDVA 기반 영상 검색 기술 개요

본 절에서는 CDVA 기반 영상 검색 프레임워크와 CDVA 서술자를 추출하는 전체적인 절차를 설명하도록 한다. 서비스가 이루어지는 상황은 그림 1에서 표현된 것과 같이 유저가 CDVA 표준을 따르는 영상 특징 추출 인코더를 가지고 있고, 검색 서비스 제공자는 CDVA 디코더 및 영상 데이터베이스를 가지고 있는 상황으로 가정한다. 유저는 쿼리 영상과 유사한 영상을 데이터베이스에서 검색하여 유사 영상 목록을 반환하는 인출 (retrieval) 동작 혹은 쿼리 영상과 특정 레퍼런스 영상의 유사도를 평가하여 유사 영상 판정 결과를 반환하는 매칭 (matching) 동작을 요청할 수 있다[7]. 어떤 동작을 수행할 것인지와 무관하게 유저 측의 CDVA 인코더는 쿼리 영상으로부터 시각적 특징들을 추출한 뒤, 추출된 특징 정보를 CDVA 서술자로 구성하고 이를 비트스트림화 하여 서버 측으로 송신한다. 서버 측에서는 수신된 비트스트림을 CDVA 서술자로 복호화 한 뒤에 동작 모드에 따라 매칭 혹은 인출 동작을 수행한다. 매칭 동작을 수행하는 경우에는 쿼리 영상의 CDVA 서술자와 레퍼런스 영상의 CDVA 서술자의 유사도 점수에 기

a) 한국전자통신연구원(Electronics and Telecommunications Research Institute)
b) 세종대학교 인공지능학과(Department of Artificial Intelligence, Sejong University)
c) 세종대학교 지능기전공학부(Department of Intelligent Mechatronics Engineering, Sejong University)
‡ Corresponding Author : 김준수(Joonsoo Kim)
E-mail: joonsookim@etri.re.kr
Tel: +82-42-860-5778
ORCID:https://orcid.org/0000-0002-6470-0773
※본 논문은 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2020-0-00011, 기계를 위한 영상부호화 기술, No. 2021-0-02067, 다목적 비디오 검색을 위한 차세대 인공신경망 기술 개발).
※ This work was supported by Institute for Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) under Grant 2020-0-00011 (Video Coding for Machine) and Grant 2021-0-02067 (Next Generation AI for Multi-purpose Video Search).
· Manuscript April 12, 2022; Revised May 13, 2022; Accepted May 13, 2022.

반하여 매칭 여부를 판정하며, 인출 동작을 수행하는 경우에는 쿼리 영상과 각 데이터베이스 영상들 사이의 유사도를 평가하여 유사도 순으로 데이터베이스 영상 목록을 작성하여 유저에게 최종 결과를 제공한다. CDVA 표준은 인코더 만을 정의하고 있기 때문에 서버 측에서 사용되는 구체적인 매칭, 인출 알고리즘은 구현에 따라 상이할 수 있는데, 예를 들어 인출을 위해 단순히 전수 조사를 통한 검색을 수행할 수도 있고 검색 속도 개선을 위해 최근접 이웃 탐색 알고리즘을 적용할 수도 있다.

이제 CDVA 서술자의 구성에 대해 설명하도록 한다. CDVA 서술자를 구성하는 영상의 시각적 특징 정보는 개별 프레임에 대한 특징 정보의 모음으로 구성되며, 개별 프레임에 대한 특징 정보는 구동 옵션에 따라 전역 특징 서술자 (global feature descriptor), 국소 특징 서술자 (local feature descriptor), 그리고 딥 특징 서술자 (deep feature descriptor)의 조합으로 구성된다. 전역 및 국소 특징 서술자는 CDVA에 앞서 개발된 Compact Descriptors for Visual Search (CDVS) 표준 기술의 방식을 그대로 따라 산출된다[9,10]. 구체적으로, 국소 특징 서술자는 scale-invariant feature transform (SIFT) 특징과 유사한 멀티 스케일 국소 특징 추출기와 histogram of oriented Gaussian (HOG)의 조합을 활용하여 산출되며, 국소 특징 서술자의 분포 특성을 피셔 벡터로 요약하여 전역 특징 서술자를 얻는다. 딥 특징 서술자는 인공신경망을 활용하여 산출된 특징맵에 Nested Invariance Pooling (NIP)을 적용하여 산출된 고정 길이 벡터로, 반복적 풀링 연산을 통해 공간적 변환에 강인한 특성을 가진다[11].

프레임 단위 서술자는 많은 경우에 높은 시간적 중복도를 가지기 때문에, CDVA 개발 과정에서 다양한 시간적 중복성 제거 방안이 제안되었다[12]. 먼저 거의 유사한 내용을 담고 있는 프레임들에 대해 중복으로 서술자를 산출하지 않기 위해 프레임 샘플링을 수행하는 방식들이 제안되었다. 구체적으로, Bailer와 Huang 등은 각각이 제안한 영상 특징 서술자 산출 파이프라인에 시간적 서브샘플링을 적용하여 비트율과 검색 성능을 조율할 수 있음을 보고하였고, Balestri 등은 단순한 시간적 서브샘플링에 더하여 색상 히스토그램 기반 프레임 선별을 추가적으로 수행하는 방안을 제안하였다[13-15]. 이와 같은 기술들은 모두 표준에 반영되어 있지만 프레임의 구체적인 선별 방식이 CDVA 서술자의 상호 운용성에 영향을 주지 않기 때문에 참고 (informative) 사항으로서 포함되어 있다. 선별된 프레임들에 대해 프레임 단위 서술자를 산출한 이후에 프레임 간 중복성 제거를 수행하는 기술들도 다수 제안되었는데, 예를 들어 Huang 등은 연속한 프레임의 전역 특징 서술자가 일정 수준 이상의 유사도를 가지는 경우에 현재 프레임의 서술자를 폐기하고 이전 프레임의 서술자를 재활용하는 전

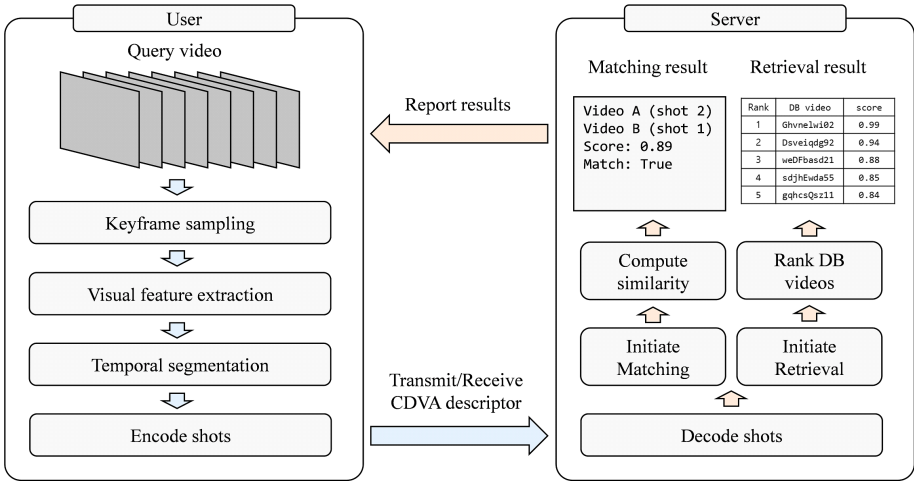


그림 1. CDVA에 기반하여 유사 동영상 판정 및 검색 기술 개념도
Fig 1. Schematic diagram of video matching and retrieval based on CDVA

파 부호화 (propagative coding) 방식을 제안하였다[16]. Bailer 등과 Balestri 등은 각각 비디오를 전역 특징 유사도에 근거하여 분할한 샷 (shot) 개념을 활용한 방식들을 제안하였는데, 전자의 경우는 샷 내에서 기준 프레임 대비 일정 수준 이상으로 유사한 전역 특징 서술자를 폐기하는 방식이고 후자의 경우는 샷 내부의 전역 특징이 충분히 유사해질 때까지 세밀하게 샷 분할을 수행한 뒤에 샷 내부에서 하나의 대표값을 선택하는 방식이다[17,18]. 여기에 더하여 서술자를 폐기하는 대신 잔차 정보를 전송하는 방식으로 무손실 압축을 수행하는 기술 또한 제안되었으며, 구체적으로는 Huang 등과 Bailer 등이 각각 국소 특징 서술자와 전역 특징 서술자에 대해 예측 부호화 (predictive coding) 기반 무손실 압축 방식을 제안하였다. CDVA 표준에는 샷 단위의 서술자 부호화를 수행하되 전역 특징 서술자와 국소 특징 서술자에 대해 각각 무손실 예측 부호화와 유사도 기반 필터링을 수행하는 방안이 채택되었으며, 영상 특징 서술자 부호화를 수행하는 구체적인 방법이 표준에 규정 (normative)되어 있다.

그림 2는 CDVA Experimental Model (CXM)을 기준으로 CDVA 서술자 산출 과정을 나타낸 순서도이다. 산출 과정은 크게 키프레임 샘플링, 프레임 단위 특징 추출, 시간적 공간 분할, 그리고 샷 단위 부호화 순서로 진행된다. 이 중 키프레임 샘플링을 포함한 일부 과정들은 인코더 제작자의 재량에 따라 다르게 구현될 수 있으나, 편의상 여기에서는 CXM 구현 내용을 기준으로 설명하도록 한다. 먼저 키프레임 샘플링은 단순한 균등 서브샘플링과 색상 히스토그램 변화에

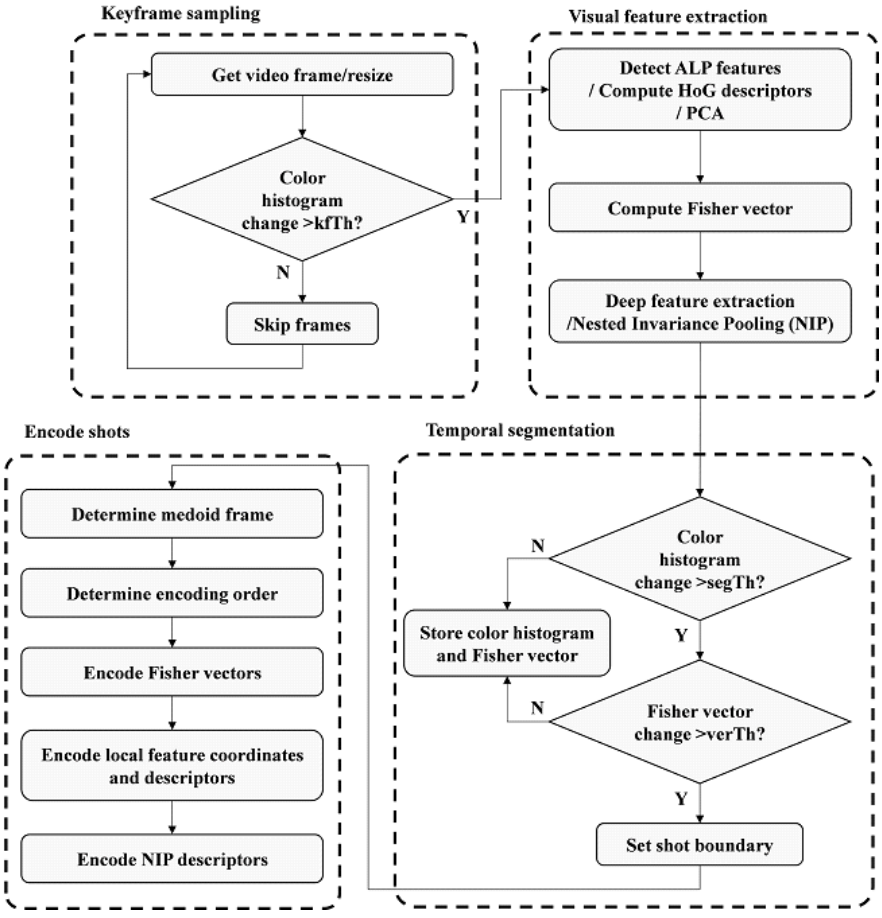


그림 2. CDVA Experimental Model에 구현된 영상 특징 서술자 추출 절차
Fig. 2. CDVA descriptor extraction procedure that is implemented in CDVA Experimental Model

기반한 추가적인 필터링으로 구현될 수 있다. 구체적으로, 동영상 프레임을 정해진 갯수씩 넘겨가며 키프레임을 샘플링하되, 새로운 프레임의 색상 히스토그램이 직전 키프레임의 색상 히스토그램과 일정 수준이상 유사한 경우 키프레임으로 선택하지 않는다. 이와 같은 방법으로 선택된 키프레임들에 대해 국소 특징점 검출을 수행하게 되며, 각 특징점에 대해 32차원 서술자들이 산출된다. 국소 특징점 서술자들의 분포 특성은 혼합 가우시안 모델을 가정한 피셔 벡터로 요약되며, 이 때 혼합 가우시안 모델의 512개 중심점 (centroid)에 대한 평균 및 분산으로는 미리 학습된 파라미터를 활용한다. 딥 특징 서술자는 이와는 완전히 별도의 과정을 통해 산출되는데, 먼저 이미지를 0도, 90도, 180도, 270도 회전한 뒤 인공신경망에 입력하여 4개의 특징맵을 얻는다. 이 특징맵들의 여러 윈도우 크기와 위치에 대해 각각 풀링을 수행한 뒤에 풀링된 벡터를 재차 풀링하는 과정을 반복하게 되며, 결과적으로 512차원 고정 길이 벡터가 산출되게 된다. 3종의 서술자들은 미리 정의된 임계값에 기반하여 양자화 되는데, 전역과 딥 특징 서술자의 경우는 이진 논리값으로, 국소 특징 서술자는 삼진 논리값으로 표현되게 된다.

이어지는 시간적 구간 분할은 샷 내에 일정 수준 이상 특징이 유사한 키프레임만이 포함되도록 하는 것이 목적으로, 샷의 첫번째 키프레임과의 색상 히스토그램 차이 및 전역 특징 서술자 유사도를 고려하여 일정 임계값 이상의 차이가 발생하는 경우 샷의 경계에 도달하였음을 판정한다. 동일한 샷에 속하는 키프레임은 일정 수준 이상의 유사도를 가지기 때문에 프레임 수준 서술자들 간의 중복성이 존재하며, 이러한 점을 고려하여 샷 단위 부호화 단계에서는 예측 부호화를 통한 서술자 압축을 수행하게 된다. 샷 단위 부호화 단계에 단계에 진입하게 되면 먼저 전역 특징 서술자를 기준으로 중앙자 (medoid)를 찾아 대표 프레임으로 지정하게 되며, 나머지 프레임들에 대해서는 중앙자와의 차분에 해당하는 정보만 남겨 이진 산술 부호화를 수행한다. 이와 같은 부호화 기법은 샷 내부 키프레임 서술자의 정보 중복성을 제거하여 압축 이득으로 이어지는 것이 일반적이다. 그러나 CXM을 활용하여 CDVA 서술자를 추출해보면 출력 비트스트림에서 전역 특징 서술자의 크기가 원본 데이터보다도 오히려 더 큰 경우가 흔하게 발생한다. 이러한 현상은 전역 특징 서술자가 고정된 크기의 피셔 벡터가 아닌 가변 길이의 scalable Fisher vector (SCFV)라는 점에 기인하는데, 자세한 내용은 다음 절에서 설명하도록 한다.

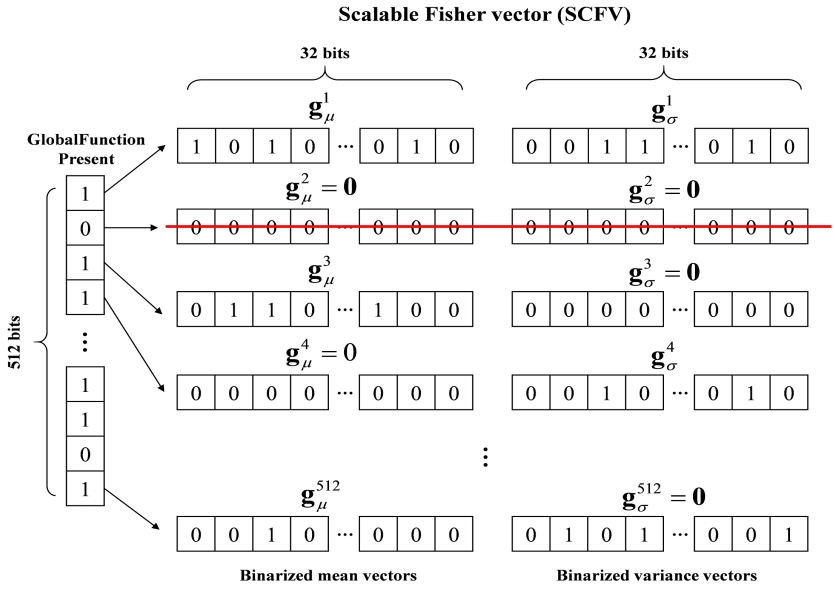


그림 3. Scalable Fisher vector를 구성하는 GlobalFunctionPresent 벡터와 피셔 서브벡터의 구조
Fig. 3. Structure of Scalable Fisher vector including the GlobalFunctionPresent vector and the Fisher subvectors

2. 전역 특징 서술자의 압축 방법

본 절에서는 전역 특징 서술자의 구체적인 형상 및 압축 방법, 그리고 기존 압축 알고리즘의 한계에 대해 설명한다. 앞서 설명한 바와 같이 개별 키프레임에 대한 전역 특징 서술자는 국소 특징 서술자를 결합한 피셔 벡터인데, 피셔 벡터는 512개의 중심점에 대응되는 32차원 평균 서브벡터 g_μ와 32차원 분산 서브벡터 g_σ들로 구성되어 총 512 x 64 비트의 크기를 가진다. 그러나 많은 경우에 피셔 벡터를 이진화 한 이후에는 상당 수의 서브벡터가 영벡터가 되게 된다. 이 점을 활용하면 굳이 32차원 영벡터들을 여러 번 보내는 대신 영벡터가 아닌 서브벡터들과 각 중심점에 해당하는 서브벡터가 영벡터인지 알려주는 512 비트를 추가로 보내어 전체적인 전송 대상 데이터 용량을 크게 줄일 수 있으며, 이러한 개념은 CDVS 표준에 이미 반영되어 있다[10]. CDVS 표준은 그림 3과 같이 512 비트 길이의 GlobalFunctionPresent 벡터와 영벡터가 아닌 서브벡터들로 전역 특징 서술자를 구성하고 있으며, 이와 같은 데이터 구성을 scalable Fisher vector라 부르고 있다.

앞서 설명한 바와 같이 각 샷의 전역 특징 서술자를 부호화할 때, 중앙자의 SCFV를 예측 서술자로 활용하는 부호화를 수행하게 되는데, 원래 의도대로라면 샷 내부 키프레임들의 SCFV들이 서로 유사하여 중앙자 SCFV와 부호화 대상 SCFV의 차분이 부호화 대상 SCFV보다 짧은 크기를 가져야 할 것이다. 그러나 실제로는 기대한 것과 반대로 SCFV의 차분이 원본 SCFV보다 큰 크기를 가지게 되는 경우가 빈번하게 발생한다. 그림 4는 SCFV의 차분을 계산하는 과정에서 오히려 부호화 대상 데이터의 크기가 증가하는 경우를 표현하고 있다. 두 개의 SCFV A와 B가 있다고 하고, 각각의 서브벡터를 $[\mathbf{a}_1\ \mathbf{a}_2\ \dots\ \mathbf{a}_{512}]$와 $[\mathbf{b}_1\ \mathbf{b}_2\ \dots\ \mathbf{b}_{512}]$로 표기하도록 하자. 배타적 논리합 (XOR)연산을 수행할 두 서브벡터의 조합은 5가지 경우가 가능한데, 두 벡터 모두 영벡터인 경우, A측 서브벡터만 영벡터인 경우, B측 서브벡터만 영벡터인 경우, 양측 서브벡터가 모두 영벡터가 아니고 두 벡터가 서로 다른 경우, 그리고 양측 서브벡터가 모두 영벡터가 아닌 동시에 서로 같은 벡터인 경우이다. 첫번째 경우와 다섯번째 경우에는 XOR 연산 결과 영벡터를 얻게 되고, 나머지 세 가지 경우에는 영벡터가 아닌 벡터를 얻게 된다. 이제 SCFV B 대신 SCFV A와 B의 차분을 전송하고자 하는 경우를 가정하고, SCFV 차분과 SCFV B의 상대적인 길이를 비교해본다. SCFV 차분의 크기가 SCFV B보다 증가하는 경우는 B측의 서브벡터만 영벡터인 경우인데, 이 경우 사실상 A측의 서브벡터를 중복으로 2회 보내는 효과가 발생하게 된다. 이러한 경우는 실제로 매우 빈번하게 발생하는데, 다른 서브벡터 조합에서 발생한 예측 부호화 이득이 중복 전송에 의한 손해를 상쇄하지 못하게 되면 오히려 예측 부호화를 통해 얻은 비트스트림의 크기가 SCFV 원본보다 더 커지게 된다.

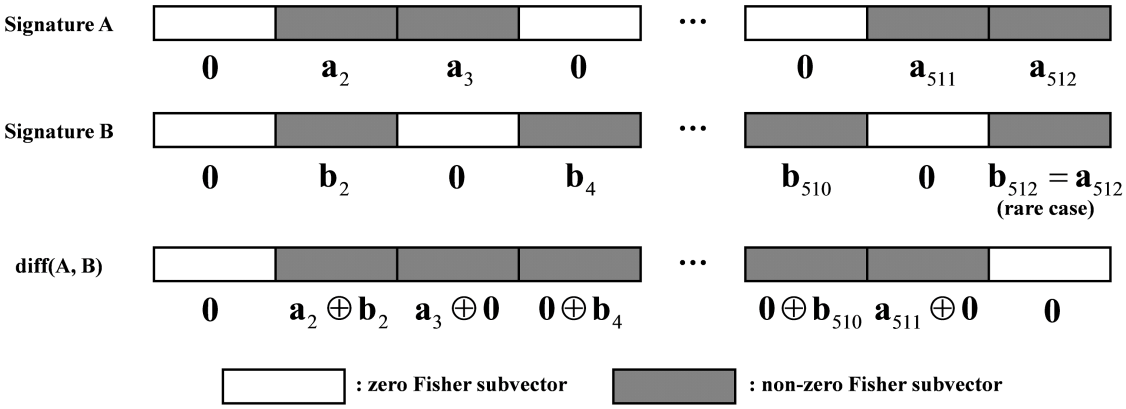


그림 4. SCFV A, B의 차분을 산출하는 과정에서 차분 벡터의 크기가 원본 SCFV보다 증가하는 사례
Fig. 4. Visualization of the case where the difference of two SCFV signatures A, B becomes larger than the original signature

Ⅲ. 제안된 전역 특징 서술자 압축 방법

1. 비대칭적 SCFV 차분 계산 및 서술자 재건 방법

앞서 살펴본 전역 특징 서술자 예측 부호화의 한계를 극복하기 위해, 새로운 SCFV 차분 계산 방법 및 원본 SCFV 복원 방법을 제안하도록 한다. 단순한 XOR 연산을 통해 SCFV의 차분을 계산할 때, 가장 큰 문제가 발생하는 부분은 부호화를 수행하려는 프레임의 SCFV측 서브벡터가 영벡터이고 중앙자 측에 대응되는 서브벡터는 영벡터가 아닌 경우이다. 제안하는 방법의 핵심은 이 경우에 한하여 서브벡터의 잔차값 (residual)을 영벡터로 정의하는 것인데, 구체적으로 다음과 같은 비가환 (noncommutative) 연산으로 표현이 가능하다:

$$\begin{aligned}&\text{asymDiff}(A,B)=\text{Concat}\left[\text{subDiff}(\mathbf{a}_i,\mathbf{b}_i)\right]_{i=1,2,\ldots,512}\\&\text{subDiff}(\mathbf{a}_i,\mathbf{b}_i)=\begin{cases}\mathbf{0} & \mathbf{b}_i=\mathbf{0}\\ \mathbf{a}_i\oplus\mathbf{b}_i & \text{otherwise}\end{cases}\end{aligned}\quad(1)$$

여기에서 $A = [\mathbf{a}_1\ \mathbf{a}_2\ \ldots\ \mathbf{a}_{512}]$와 $B = [\mathbf{b}_1\ \mathbf{b}_2\ \ldots\ \mathbf{b}_{512}]$는 SCFV 및 그 서브벡터들을 의미하며, ⊕는 XOR 연산을 의미한다. 위와 같이 잔차값을 정의하는 경우, asymDiff(A, B)의 서브벡터가 영벡터가 되는 경우는 양 측의 서브벡터가 동일한 벡터인 경우와 B측의 서브벡터가 영벡터인 경우의 두 가지가 존재하게 된다. 이로 인해 중앙자와 잔차값으로부터 SCFV를 재건하고자 할 때 모호성이 발생하게 되는데, 이를 해결하기 위해 GlobalFunctionPresent 벡터를 활용하도록 한다.

GlobalFunctionPresent 벡터는 앞선 절에서 설명한 바와 같이 SCFV의 각 서브벡터가 영벡터인지를 표시하는 비트들의 모음인데, 이에 따라 기존 CDVA 표준에서 GlobalFunctionPresent의 i-번째 비트가 1로 설정되어 있으면 반드시 i-번째 서브벡터는 영벡터가 아니다. 제안되는 방법에서는 위에서 언급된 모호성을 해결하기 위해 특정한 조건 하에 한하여 서브벡터가 영벡터임에도 불구하고 해당 서브벡터에 대응되는 GlobalFunctionPresent 비트를 1로 설정하도록 한다. 구체적으로, 그림 5의 좌측에 표현된 것과 같이 asymDiff(medoid, G)를 산출함에 있어 중앙자 측과 G측의 i-번째 서브벡터가 영벡터가 아닌 동시에 서로 같은 경우에 한하여 XOR 연산 결과가 영벡터임에도 불구하고 GlobalFunctionPresent의 i-번째 비트를 1로 설정한다.

여기에서 mi와 gi는 각각 중앙자와 G의 i-번째 서브벡터를 의미한다. 이와 같은 방식으로 잔차값을 산출하는 경우,

$$\text{GlobalFunctionPresent}[i]=\begin{cases}1 & \mathbf{m}_i\neq\mathbf{0}\text{ and }\mathbf{g}_i\neq\mathbf{0}\\ \text{bool}(\text{subDiff}(\mathbf{m}_i,\mathbf{g}_i)=0) & \text{otherwise}\end{cases}\quad(2)$$

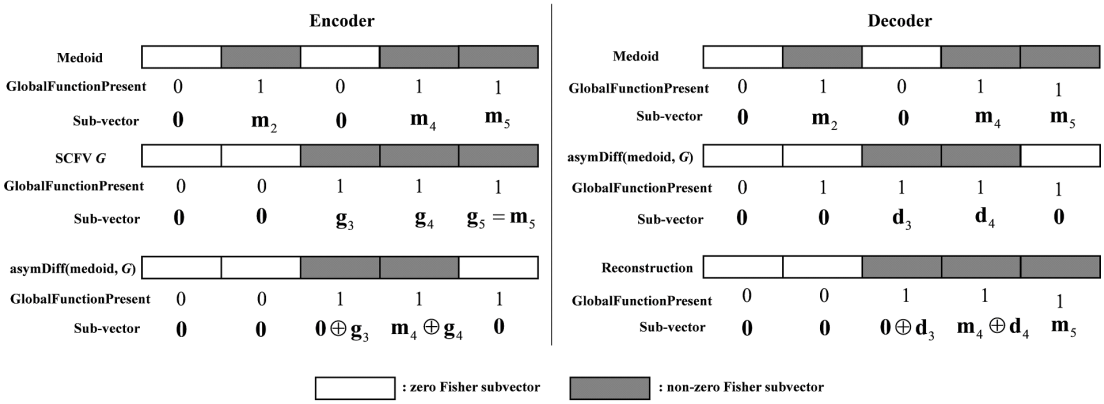


그림 5. 비대칭적 차분 계산에 기반한 SCFV 잔차값 산출 및 원본 SCFV 복원 방법
Fig. 5. Illustration of the SCFV residual computation and reconstruction method based on the proposed asymmetric difference

asymDiff(medoid, G)는 언제나 G와 동일한 GlobalFunctionPresent 벡터를 가지게 된다. 결과적으로 차분 계산 과정에서 서브벡터의 갯수가 오히려 증가하는 현상이 발생하지 않으며, asymDiff(medoid, G)이 G에 비해 희소벡터 (sparse vector)에 가까운 성질을 가질 확률이 높기 때문에 산술부호화를 통한 안정적인 압축 이득을 얻을 수 있게 된다. 단, CDVA 표준이 상호운용성 보장을 위해 단순 XOR 연산 기반 SCFV 압축을 수행하도록 강제하고 있기 때문에, 비대칭 차분 계산을 도입하여 제작된 인코더가 CDVA 표준을 따르지 못하게 되는 한계가 있다.

이제 그림 5의 우측과 같이 디코더가 중앙자와 비대칭 차분 asymDiff(medoid, G)로부터 G를 재건할 수 있게 된다. 비대칭 차분의 i-번째 서브벡터가 영벡터가 아닌 경우는 기존과 동일하게 XOR 연산을 통해 G의 서브벡터를 재건할 수 있고, 반대로 영벡터인 경우는 두 가지 경우에 대해 아래와 같이 재건을 수행한다:

$$g_i = \begin{cases} m_i & \text{GlobalFunctionPresent}[i] = 1 \\ 0 & \text{GlobalFunctionPresent}[i] = 0 \end{cases} \quad (3)$$

여기에서 GlobalFunctionPresent는 비대칭 차분에 대한 것이다. 이와 같은 복원 절차가 G를 정확하게 복원함은 그림 5에 표현된 5가지 경우로부터 쉽게 확인이 가능하다.

서술자 재건 과정에서의 모호성을 해결하는 방법은 사실 한 가지 이상 더 존재하는데, 예를 들어 mi가 영벡터가 아니고 g_i가 영벡터인 경우에 대해 GlobalFunctionPresent의 i-번째 비트를 1로 설정할 수 있을 것이다. 그러나 일반적으로 m_i와 g_i가 모두 영벡터가 아닌 동시에 동일한 벡터일 확률이 G측 서브벡터만 영벡터일 확률보다 더 희박할 것이므로 수식 (3)과 같이 GlobalFunctionPresent를 설정할 때 1로 설정된 비트의 갯수가 더 적을 것으로 예상할 수 있다.

2. 실험 결과

본 절에서는 제안된 SCFV 압축 방법의 검증을 위한 실험 환경 및 서술자 압축 실험 결과를 설명한다. 실험에 사용된 CDVA 표준 소프트웨어는 CXM 6.5 버전이며, 제안된 알고리즘의 구현을 위해 SCFV 비대칭 차분 계산을 추가적으로 구현하여 인코더 측에 반영하였다. 디코더 측에는 그림 5에 표현된 SCFV 재건 방법을 구현하였으며, 이를 통해 제안된 SCFV 압축 방법의 적용 여부와 무관하게 동일한 비디오 검색 결과를 얻음을 확인하였다. 압축 성능 평가를 위한 테스트 영상은 비디오 검색 데이터셋인 Fine-Grained Video Retrieval (FIVR) 데이터셋을 활용하였으며, 구체적으로는 FIVR 5K에 속한 50개의 쿼리 영상을 사용하였다[19,20]. 그림 6에서 볼 수 있는 것과 같이 쿼리 영상들은 화재, 항공기 추락 등을 포함한 다양한 사건에 관한 것들이며, 총 분량은 61,025 프레임이다. 일부 영상은 담화문 발표나 예배 영상과 같이 정적인 영상이며, 다수 인원이 대피하거나 급작스러운 폭발이 일어나는 동적인 영상도 포함되어 있다. 실험은 CXM의 3종의 비트레이트 옵션 (‘16K’, ‘64K’, ‘256K’)에 대해 수행되었으며, 비트레이트 옵션에 따라 시

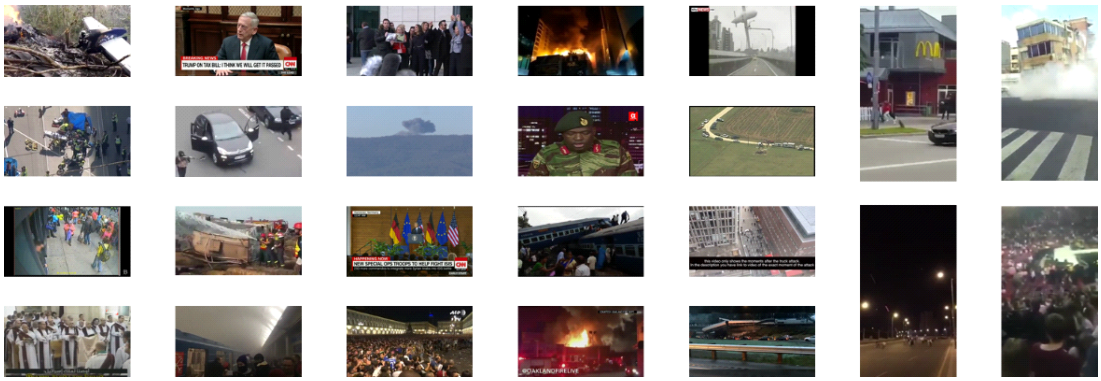


그림 6. FIVR 데이터셋에 포함된 쿼리 영상의 예시
Fig. 6. Examples of the query videos that are included in FIVR dataset

간적 서브샘플링 레이트와 색상 히스토그램 변화 임계값이 변화하므로 키프레임의 갯수가 변화하게 된다. CXM은 SCFV 유사도 임계값을 기반으로 손실 압축을 수행할 수 있도록 구현되어 있는데, 여기에서는 임계값을 매우 큰 값으로 설정하여 무손실 압축을 수행하도록 설정하였다.

표 1은 세 가지 비트레이트 조건 하에서 프레임 간 중복이 제거되지 않은 SCFV, 기존 CXM으로 압축된 전역 특징 서술자, 그리고 제안된 알고리즘을 적용하여 압축된 전역 특징 서술자를 비교한 것이다. 기입된 서술자 크기는 키프레임 당 평균 크기를 의미하며, 단일 프레임에 대한 SCFV 원본 크기는 최대 4160 bytes이다. 프레임 간 서술자 중복 제거를 수행하지 않은 'All-Intra' 조건 하에서 평균적인 SCFV 크기는 약 1560 bytes였으며, 이로부터 FIVR 5K 쿼리 영상에 대해서는 평균적으로 512개 중심점 중 187개 정도가 영벡터가 아닌 서브벡터에 대응된다는 것을 알 수 있다. 상당 수의 중심점에 대응되는 서브벡터가 영벡터이므로, SCFV 차분 계산 과정에서 데이터의 양이 증가하는 현상이 자주 발생할 것으로 예상할 수 있는데, 실제로 중앙자와의 차분을 계산하는 단계에서 두 SCFV의 Global-FunctionPresent 벡터를 비교하여 Intersection over Union (IoU)을 산출하고 그 빈도를 표시하면 그림 7과 같은 결과를 얻는다. 이 때, IoU는 영벡터가 아닌 서브벡터에 대응되는 중심점들의 교집합 목록 길이를 합집합 목록 길이로 나눈 것으로 정의한다. 대부분의 경우 IoU는 0.4에서 0.6 사이에 분포하며, 이는 XOR 연산 결과로 얻는 영벡터가 아닌 서브벡터 중 절반 정도는 한 쪽 서브벡터가 영벡터인 상태에서 계산된 것이라는 것을 의미한다. 중앙자가 특별히 더 희소 벡터에 가까운 특성을 가지지 않는다고 가정하면 약

표 1. SCFV 차분 계산 방법에 따른 전역 특징 서술자의 크기 비교
Table 1. Comparison of the size of the global feature descriptors depending on the type of SCFV difference

Bitrate	# of keyframes	All-Intra	Naive exclusive OR (conventional)		Asymmetric difference (proposed)	
		Size per frame (bytes)	Size per frame (bytes)	Relative size (%)	Size per frame (bytes)	Relative size (%)
16K	1775	1558.86	1959.67	125.7	1535.79	98.5
64K	2353	1560.13	1968.49	126.2	1534.83	98.4
256K	3333	1567.46	1970.30	125.7	1539.43	98.2

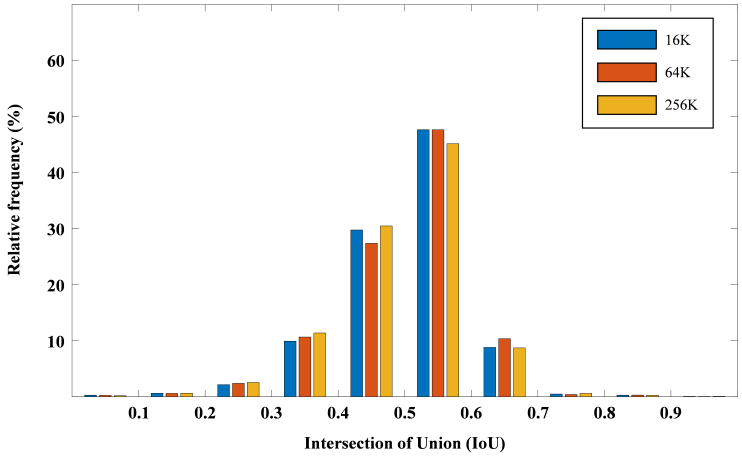


그림 7. 50개 테스트 영상에 대해 수행된 모든 SCFV 차분 계산 과정에서 영벡터가 아닌 서브벡터에 대응되는 중심점들의 겹침 정도를 표현한 히스토그램
Fig. 7. A histogram visualizing the degree of overlap for centroids that corresponds to the non-zero subvectors when computing SCFV differences for the 50 test videos

4분의 1정도 빈도로 중앙자 측의 서브벡터가 중복 전송되는 효과가 발생하게 된다.

표 1의 'Naive exclusive OR' 조건은 제안된 알고리즘이 적용되지 않은 CXM 구동 결과를 의미하며, 이 경우 프레임 당 SCFV의 평균 크기는 약 1960~1970 bytes였다. 이는 원본 SCFV 대비 약 25% 오히려 증가한 것으로, 이러한 증가는 앞에서 살펴본 중복 전송 효과로 대부분 설명이 가능하다. 반면, 본 논문에서 제안된 비대칭적 SCFV 차분 계산 방법을 적용하게 되면 모든 비트레이트 모드에 대해 일정 수준의 압축 이득을 볼 수 있게 된다. 언제나 압축 이득이 발생하는 가장 큰 이유는 이는 앞서 설명한 바와 같이 차분 벡터가 재건 대상 SCFV보다 더 커지는 일이 발생하지 않기 때문으로 해석할 수 있다. 단, 제안된 방법은 영벡터가 아닌 두 서브벡터가 완전히 일치하는 경우에 기존 방식보다 압축 효율에서 손해를 보게 되는데, 이와 같은 경우가 빈번하게 발생하는 특수한 데이터셋에 대해서는 기존의 전역 특징 서술자가 더 높은 효율을 보일 가능성이 있다. 그럼에도 불구하고 실험 결과에서도 확인할 수 있는 것과 같이 무손실 서술자 부호화 기법은 손실 압축을 수행하는 과정에서 산출된 SCFV 및 SCFV 유사도 평가 결과를 재활용하여 보조적 압축 이득을 획득하는 수준에 머무르고 있으며, 따라서 기존의 SCFV 무손실 압축 기법을 활용하여 큰 압축 이득을 볼 수 있는 몇 가지 특수한 사례를 제외하면 기존 기법에서 빈번하게 발생하는 압축 효율 손실을 원천적으로 차단하는 것이 범용성 측면에서 더 합리적일 수 있다.

Ⅳ. 결 론

본 논문에서는 CDVA 서술자를 구성하는 전역 특징 서술자의 압축 방법을 제안하였다. SCFV의 구조적 특징에 의해 프레임 간 서술자 중복성 제거 과정에서 오히려 부호화 대상 데이터의 양이 증가하는 원인을 분석하였고, 이러한 현상이 일어나는 것을 방지하는 방법으로 비대칭적 SCFV 차분 연산의 도입을 제안하였다. 제안된 비대칭적 SCFV 차분 계산을 기존 CXM에 구현하여 FIVR 데이터셋에 포함된 50개 쿼리 영상에 대해 전역 특징 서술자 압축 성능을 확인하였고, 부호화 이후 원본 데이터보다 오히려 크기가 증가하는 기존 압축 방식과 달리 3종의 비트레이트 조건에 대해 모두 일정 수준의 압축 효율을 얻을 수 있음을 확인하였다.

참 고 문 헌 (References)

[1] D. Nistér, and H. Stewénius, "Scalable Recognition with a Vocabulary Tree," Proceeding of Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY, USA, pp. 2161-2168, 2006.
doi: https://doi.org/10.1109/CVPR.2006.264

[2] H. Jégou, M. Douze, C. Schmid, and P. Pérez, "Aggregating Local Descriptors into a Compact Image Representation," Proceeding of Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, pp. 3304-3311, 2010.
doi: https://doi.org/10.1109/CVPR.2010.5540039

[3] F. Perronnin, Y. Liu, J. Sánchez, and H. Poirier, "Large-Scale Image Retrieval with Compressed Fisher Vectors," Proceeding of Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 3384-3391, 2010.
doi: https://doi.org/10.1109/CVPR.2010.5540009

[4] A. Babenko, A. Slesarev, A. Chigorin, and V. Lempitsky, "Neural Codes for Image Retrieval," Proceeding of European Conference on Computer Vision, Zurich, Switzerland, pp. 584-599, 2014.

[5] A. Gordo, J. Almazán, J. Revaud, and D. Larlus, "Deep Image Retrieval: Learning Global Representations for Image Search," Proceeding of European Conference on Computer Vision, Amsterdam, Netherlands, pp. 241-257, 2016.

[6] L.-Y. Duan, V. Chandrasekhar, J. Chen, J. Lin, Z. Wang, T. Huang, B. Girod, and W. Gao, "Overview of the MPEG-CDVS Standard," IEEE Transactions on Image Processing, Vol. 25, No. 1, pp. 179-194, Nov. 2015.
doi: https://doi.org/10.1109/TIP.2015.2500034

[7] L.-Y. Duan, Y. Lou, Y. Bai, T. Huang, W. Gao, V. Chandrasekhar, J. Lin, S. Wang, and A. C. Kot, "Compact Descriptors for Video Analysis: The Emerging MPEG Standard," IEEE MultiMedia, Vol. 26, No. 2, pp. 44-54, Oct. 2018.
doi: https://doi.org/10.1109/MMUL.2018.2873844

[8] Z. Huang, L.-Y. Duan, J. Lin, S. Wang, S. Ma, and T. Huang, "An Efficient Coding Framework for Compact Descriptors Extracted from Video Sequence," Proceeding of IEEE International Conference on Image Processing, Quebec City, QC, Canada, pp. 3822-3826, 2015.
doi: https://doi.org/10.1109/ICIP.2015.7351520

[9] Y Uchida, and S. Sakazawa, "Image Retrieval with Fisher Vectors of Binary Features," Proceeding of Asian Conference on Pattern Recognition, Naha, Japan, pp. 23-28, 2013.
doi: https://doi.org/10.1109/ACPR.2013.6

[10] Y. Wu, F. Gao, Y. Huang, J. Lin, V. Chandrasekhar, J. Yuan, and L.-Y. Duan, "Codebook-Free Compact Descriptors for Scalable Visual

Search," IEEE Transactions on Multimedia, Vol. 21. No. 2, pp. 388-401, July, 2018.
doi: https://doi.org/10.1109/TMM.2018.2856628

[11] J. Lin, L.-Y. Duan, S. Wang, Y. Bai, Y. Lou, V. Chandrasekhar, T. Huang, A. Kot, and W. Gao, "HNIP: Compact Deep Invariant Representations for Video Matching, Localization, and Retrieval," IEEE Transactions on Multimedia, Vol. 19, No. 9, pp. 1968-1983, June, 2017.
doi: https://doi.org/10.1109/TMM.2017.2713410

[12] M. Bober, W. Bailer, and J. Chen, "Results of the Call for Proposals on CDVA," ISO/IEC JTC1/SC29/WG11, N15938, San Diego, USA, Feb. 2016.

[13] W. Bailer, "JRS response to CDVA Core Experiment 1," ISO/IEC JTC1/SC29/WG11, m38519, Geneva, CH, May. 2016.

[14] Z.Huang, L. Wei, S. Wang, L.-Y. Duan, J. Chen, A. Kot, S. Ma, T. Huang, and W. Gao, "PKU's Response to CDVA CE1," ISO/IEC JTC1/SC29/WG11, m38625, Geneva, CH, May. 2016.

[15] M. Balestri, G. Francini, S. Lepsoy, M. Bober, S. Husain, and S. Paschalakis "BRIDGET Response to the MPEG CfP for Compact Descriptors for Video Analysis (CDVA) - Search and Retrieval," ISO/IEC JTC1/SC29/WG11, m37880, San Diego, USA, Feb. 2016.

[16] Z. Huang, L.-Y. Duan, J. Chen, L. Wei, T. Huang, and W. Gao, "PKU's Response to MPEG CfP for Compact Descriptor for Visual Analysis," ISO/IEC JTC1/SC29/WG11, m37636, San Diego, USA, Feb. 2016.

[17] W. Bailer and S. Wechtitsch, "JRS Response to Call for Proposals for Technologies Compact Descriptors for Video Analysis(CDVA) - Search and Retrieval," ISO/IEC JTC1/SC29/WG11, m37794, San Diego, USA, Feb. 2016.

[18] M. Balestri, G. Francini, S. Lepsoy, M. Bober, and S. Husain, "BRIDGET Report on CDVA Core Experiment 1 (CE1)," ISO/IEC JTC1/SC29/WG11, m38664, Geneva, CH, May. 2016.

[19] G. Kordopatis-Zilos, S. Papadopoulos, I. Patras, and I. Kompatsiaris, "FIVR: Fine-Grained Incident Video Retrieval," IEEE Transactions on Multimedia, Vol. 21, No. 10, pp. 2638-2652, Oct., 2019.
doi: https://doi.org/ 10.1109/TMM.2019.2905741

[20] G. Kordopatis-Zilos, S. Papadopoulos, I. Patras, and I. Kompatsiaris, "ViSiL: Fine-Grained Spatio-Temporal Video Similarity Learning," Proceeding of the IEEE/CVF International Conference on Computer Vision, Seoul, South Korea, pp. 6351-6360, 2019.
doi: https://doi.org/ 10.1109/ICCV.2019.00645

저 자 소 개

김 준 수

- 2012년 2월 : 서울대학교 전기컴퓨터공학부 학사
- 2017년 2월 : 서울대학교 전기정보공학부 박사
- 2017년 2월 ~ 현재 : ETRI 실감미디어연구실 선임연구원
- ORCID : https://orcid.org/0000-0002-6470-0773
- 주관심분야 : Light field, VR/AR, 컴퓨터 비전, 기계를 위한 영상 부호화

조 원

- 2021년 2월 : 세종대학교 지능기전공학부 학사
- 2021년 3월 ~ 현재 : 세종대학교 인공지능학과 석사과정
- ORCID : https://orcid.org/0000-0001-6642-5667
- 주관심분야 : 컴퓨터 비전, 기계학습, 비디오 이해, 능동학습, 자기지도학습

저 자 소 개

임 근 택

- 2023년 2월: 세종대학교 지능기전공학부 학사
- 2023년 3월 ~ 현재 : 세종대학교 지능기전공학부 석사과정
- ORCID : https://orcid.org/0000-0003-2204-6109
- 주관심분야 : 기계학습, 컴퓨터 비전, 비디오 이해

윤 정 일

- 1996년 2월 : 전북대학교 제어계측공학 학사
- 1998년 2월 : GIST 메카트로닉스공학부 석사
- 2005년 2월 : GIST 메카트로닉스공학부 박사
- 2005년 2월 ~ 현재 : ETRI 실감미디어연구실 책임연구원
- ORCID : https://orcid.org/0000-0002-9848-3041
- 주관심분야 : 이머시브 미디어, 컴퓨터 비전, 기계를 위한 영상 부호화

곽 상 운

- 2012년 2월 : 연세대학교 전기전자공학부 학사
- 2014년 2월 : KAIST 전기전자공학부 석사
- 2014년 2월 ~ 현재 : ETRI 실감미디어연구실 선임연구원
- ORCID : https://orcid.org/0000-0002-9295-8966
- 주관심분야 : 이머시브 미디어, 컴퓨터 비전, 기계를 위한 영상 부호화

정 순 흥

- 2001년 2월 : 부산대학교 전자공학과 학사
- 2003년 2월 : KAIST 전기전자공학부 석사
- 2016년 2월 : KAIST 전기전자공학부 박사
- 2005년 4월 ~ 현재 : ETRI 실감미디어연구실 책임연구원
- ORCID : https://orcid.org/0000-0003-2041-5222
- 주관심분야 : 이머시브 미디어, 컴퓨터 비전, 영상부호화, 실감 방송

정 원 식

- 1992년 2월 : 경북대학교 전자공학과 (공학사)
- 1994년 2월 : 경북대학교 대학원 전자공학과 (공학석사)
- 2000년 2월 : 경북대학교 대학원 전자공학과 (공학박사)
- 2000년 5월 ~ 현재 : 한국전자통신연구원 책임연구원
- ORCID : https://orcid.org/0000-0001-5430-2969
- 주관심분야 : 3DTV 방송 시스템, 라이트필드 이미징, 영상부호화, 딥러닝기반 신호처리, 멀티미디어 표준화

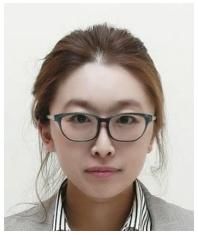
저 자 소 개

추 현 곤

- 1998년 2월 : 한양대학교 전자공학과 (공학사)
- 2000년 2월 : 한양대학교 전자공학과 (공학석사)
- 2005년 2월 : 한양대학교 전자통신전파공학과 (공학박사)
- 2005년 2월 ~ 현재 : 한국전자통신연구원 선임연구원
- 2015년 1월 ~ 2017년 1월 : 한국전자통신연구원 디지털홀로그래피연구실장
- 2017년 9월 ~ 2018년 8월 : Warsaw University of Technology, Poland 방문연구원
- ORCID : https://orcid.org/0000-0002-0742-5429
- 주관심분야 : Computer vision, 3D imaging and holography, 3D depth imaging, 3D broadcasting system

서 정 일

- 1994년 2월 : 경북대학교 전자공학과 (공학사)
- 1996년 2월 : 경북대학교 대학원 전자공학과 (공학석사)
- 2005년 8월 : 경북대학교 대학원 전자공학과 (공학박사)
- 1998년 2월 ~ 2000년 10월 : LG반도체 주임연구원
- 2010년 8월 ~ 2011년 7월 : 영국 Southampton University, ISVR 방문연구원
- 2000년 11월 ~ 현재 : 한국전자통신연구원 실감미디어연구실 실장
- ORCID : http://orcid.org/0000-0001-5131-0939
- 주관심분야 : 오디오 신호처리, 실감음향, 디지털방송, 멀티미디어 표준화



최 유 경

- 2006년 2월 : 숭실대학교 전자공학부 학사
- 2008년 2월 : 연세대학교 전기전자공학부 석사
- 2018년 2월 : KAIST 전기전자공학부/로봇공학학제전공 박사
- 2018년 2월 ~ 현재: 세종대학교 지능기전공학부 조교수
- ORCID : https://orcid.org/0000-0002-9970-0132
- 주관심분야 : 컴퓨터 비전, 로보틱스