Journal of The Korea Society of Computer and Information
Vol. 24 No. 8, pp. 19-28, August 2019
www.ksci.re.kr
https://doi.org/10.9708/jksci.2019.24.08.019

Online Video Synopsis via Multiple Object Detection

JaeWon Lee*, DoHyeon Kim*, Yoon Kim*

Abstract

In this paper, an online video summarization algorithm based on multiple object detection is proposed. As crime has been on the rise due to the recent rapid urbanization, the people's appetite for safety has been growing and the installation of surveillance cameras such as a closed-circuit television(CCTV) has been increasing in many cities. However, it takes a lot of time and labor to retrieve and analyze a huge amount of video data from numerous CCTVs. As a result, there is an increasing demand for intelligent video recognition systems that can automatically detect and summarize various events occurring on CCTVs. Video summarization is a method of generating synopsis video of a long time original video so that users can watch it in a short time. The proposed video summarization method can be divided into two stages. The object extraction step detects a specific object in the video and extracts a specific object desired by the user. The video summary step creates a final synopsis video based on the objects extracted in the previous object extraction step. While the existed methods do not consider the interaction between objects from the original video when generating the synopsis video, in the proposed method, new object clustering algorithm can effectively maintain interaction between objects in original video in synopsis video. This paper also proposed an online optimization method that can efficiently summarize the large number of objects appearing in long-time videos. Finally, Experimental results show that the performance of the proposed method is superior to that of the existing video synopsis algorithm.

▸Keyword: Video Synopsis, Video Summarization, Object-based Video Recognition, Object Detection, CCTV(closed-circuit television)

I. Introduction

최근 도시 전역에는 CCTV(closed-circuit television)와 같은 감시카메라의 설치가 급증하고 있다. 감시카메라는 물리 보안에서 핵심이 되는 장비로써 우리 사회의 안전을 책임지고 있으며 일반적으로 고정된 위치에서 24시간 동안 녹화된다. 도시 전역에 많은 수의 감시카메라가 설치되고 있지만 이렇게 녹화된 긴 시간의 비디오 데이터에서 특정 용도에 따라 사람이 직접 이벤트를 검색하고 분석하는 작업은 많은 시간과 노동력이 필요하다는 문제가 있다. 이에, 감시카메라에서 녹화된 비디오에서 특정 이벤트를 효율적이고 효과적으로 검색하고 조회할 수 있는 비디오 요약(video condensation) 방법에 관한 요구와 관심이 점차 증가하고 있다. 비디오 요약이란 원본 비디오로부터 사용자가 원하는 의미 있는 정보를 검색하고 추출하는 방식의 일종으로, 대표적인 비디오 요약 방법으로는 비디오 빨리 감기(video fast-forward), 비디오 스키밍(video skimming), 비디오 초록(video abstraction), 비디오 개괄(video summarization), 비디오 몽타주(video montage) 그리고 비디오 시놉시스(video synopsis)와 같은 방법들이 있다. 비디오

• First Author: JaeWon Lee, Corresponding Author: Yoon Kim
*JaeWon Lee (insurgent92@kangwon.ac.kr), Dept. of Computer Engineering, Kangwon National University
*DoHyeon Kim (abc3698@kangwon.ac.kr), Dept. of Computer Engineering, Kangwon National University
*Yoon Kim (yooni@kangwon.ac.kr), Dept. of Computer Engineering, Kangwon National University
• Received: 2019. 07. 16, Revised: 2019. 08. 08, Accepted: 2019. 08. 12.
• This study has been worked with the support of a research grant of Kangwon National University in 2017.

요약은 일반적으로 프레임 기반의 비디오 요약 방법과 객체 기반의 비디오 요약 방법으로 분류할 수 있다.

프레임 기반의 비디오 요약 방법은 비디오 요약을 위한 기본 처리 단위가 이미지 프레임이다. 프레임 기반의 비디오 요약 방법에는 비디오 빨리 감기, 비디오 스키밍, 비디오 초록 그리고 비디오 개괄과 같은 방법이 있다. 프레임 기반의 비디오 요약 방법은 사전에 사용자가 정의한 기준을 바탕으로 긴 시간의 비디오에서 적절한 키 프레임을 선택하고 불필요한 프레임을 제거하는 방식으로 동작한다. 하지만 프레임 기반의 비디오 요약 방법은 이미지 프레임 단위로 비디오를 요약하기 때문에 긴 비디오를 매우 집약적으로 압축하여 아주 짧은 요약 비디오를 생성하는 데 한계가 있으며 사용자가 원하는 특정 이벤트를 정확히 인식할 수 없다는 문제점이 있다.

반면 객체 기반의 비디오 요약 방법은 프레임 기반의 비디오 요약 방법보다 비교적 최근에 많은 연구가 진행 중인 분야로 비디오 요약을 위한 기본 처리 단위가 원본 비디오에서 추출한 객체이다. 객체 기반의 비디오 요약 방법은 긴 시간의 비디오 프레임에서 사용자가 원하는 이벤트를 직접 인식 및 추출하고 이를 기반으로 짧은 요약 비디오을 생성하는 기법이다. 객체 기반의 비디오 요약 방법은 프레임 기반의 비디오 요약 방법과 비교하여 훨씬 더 집약적이고 짧은 요약 비디오를 생성할 수 있으며, 다양한 컴퓨터 비전 알고리즘을 이용하여 여러 이벤트를 인식할 수 있다는 장점이 있다. 객체 기반의 비디오 요약 방법에는 비디오 몽타주, 비디오 시놉시스와 같은 방법이 있다. 비디오 몽타주는 시·공간적으로 객체들을 압축하여 요약 비디오를 생성하는 방법이다[1]. 하지만 비디오 몽타주에서 사용하는 시·공간적 압축 기법은 요약 비디오 생성 시 원본 비디오에서 각 객체가 가지는 고유한 공간 정보와 움직임 정보를 고려하지 않기 때문에 요약 비디오에서는 객체의 고유한 정보들을 잃는 문제가 있다.

이를 해결하기 위하여 원본 비디오에서 나타나는 객체의 고유한 정보를 최대한 유지하려는 방법인 비디오 시놉시스가 제안되었다[2, 3, 4]. 비디오 시놉시스는 원본 비디오에서 추출한 객체를 시간 도메인 상에서 재배열하는 방법으로 비디오를 요약하는 방법이다. 비디오 시놉시스는 다른 비디오 요약 방법들과 비교하여 훨씬 더 짧은 요약 비디오를 생성하면서도 원본 비디오의 정보를 최대한 유지할 수 있다. 비디오 시놉시스 방법으로 생성한 요약 비디오를 이용하여 사용자는 아주 긴 시간의 비디오를 단 몇 분 안에 살펴볼 수 있다. 비디오 시놉시스의 기본 연산 단위는 객체 튜브이다. 비디오 시놉시스는 이렇게 객체 튜브를 시·공간적으로 재배열하는 방식으로 원본 비디오 대비 훨씬 더 짧은 요약 비디오를 생성한다. 비디오 시놉시스는 다른 방법들과 비교하여 요약률(condensation ratio)이 높으며 객체 튜브를 기반으로 요약을 하므로 원본 비디오에서 나타나는 객체의 움직임과 같은 객체 고유의 특성을 최대한 보존할 수 있다. 이러한 특징 때문에 비디오 시놉시스는 특히 감시카메라와 같이 고정된 카메라에서 24시간 동안 녹화되는 긴 시간의 비디오를 짧게 요약하는 데 적합하다. 비디오 시놉시스 알고리즘에서 가장 핵심은 원본 비디오에서 나타나는 객체들의 움직임을 요약 비디오에서도 가능한 잘 보존해야 한다는 점이다. 본 논문에서는 다중 객체검출 기반의 온라인 비디오 요약 알고리즘을 제안한다. 기존의 비디오 시놉시스 방법들은 원본 비디오에서 등장하는 단일 객체의 움직임은 보존할 수 있었지만, 객체 간의 상호작용은 올바르게 표현할 수 없었다. 제안하는 방법에서는 객체 간의 상호작용을 비용함수에 명시하고 이를 최적화하여 객체 간의 상호작용을 보존할 수 있는 새로운 비디오 시놉시스 방법과 이를 위한 새로운 비디오 시놉시스 프레임워크를 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 비디오 시놉시스와 관련된 연구를 기술하고 3장에서는 제안하는 알고리즘의 구체적인 내용을 기술한다. 그리고 4장에서는 실험을 통하여 제안하는 알고리즘의 우수성을 입증한다. 그리고 최종적으로 5장에서 결론을 맺는다.

II. Related Works

이번 장에서는 비디오 시놉시스와 관련된 연구에 관해 설명한다. 비디오 요약 방법을 프레임 기반의 비디오 요약 방법과 객체 기반의 비디오 요약 방법으로 분류하여 설명하고 기존의 오프라인과 온라인 비디오 시놉시스 프레임워크에 관해 설명한다.

2.1 Frame-based Video Condensation

프레임 기반의 비디오 요약 방법에는 키 프레임 선택과 비디오 스키밍이 있다[5, 6, 7]. 키 프레임 선택 방법은 사전에 정의한 기준에 따라서 원본 비디오로부터 특정 프레임을 선택하는 기법이다. 키 프레임을 가장 쉽게 추출하는 방법은 원본 비디오에서 시간상 일정한 비율로 프레임을 뽑는 것이다. 하지만 이 방법은 사용자가 원하는 기준을 반영할 수 없다. Bennett은 원본 비디오에서 중요한 정보를 추출하기 위해 비 균일적으로 프레임을 뽑은 방법을 제안하였다[5]. 그리고 이와 유사하게 Petrovic은 적응적 비디오 빨리 감기(adaptive video fast-forward)를 제안하였는데 이 방법은 원본 비디오로부터 객체의 움직임과 같이 사용자가 사전에 정의한 기준으로 의미 있는 정보가 포함된 프레임을 추출한다[6, 7]. 그리고 정확한 키 프레임 추출을 위한 다양한 기준이 연구되었는데 이 방법들은 공통으로 원본 비디오에서 움직임이 없는 프레임과 같은 사용자의 관심이 비교적 적은 프레임을 제외하는 방법이다[8, 9, 10, 11, 12]. 그리고 키 프레임 선택 시 이미지뿐만 아니라 소리 신호와 같은 다른 도메인 정보 또한 도움이 될 수 있는데 [13], [14], [15]과 같은 방법들은 이러한 추가적인 정보를 이용하는 방법을 제안하였다. 키 프레임 선택 방법은 정적인 요약이라고 할 수 있다. 반면 비디오 스키밍은 조금 더 동적인 비디오 요약 방법이다. 비디오 스키밍은 키 프레임 선택 방법처럼 고정된 프레임만을 뽑아내는 것이 아니라 사전에 정의된 기준을 바탕으로 연속된 비디오를 추출하고 이어 붙이는 기술이다[16, 17]. 일반적으로 프레임 기반의 비디오 요약 방법은 효율적이며 많은 경우에 추출한

키 프레임으로 원본 비디오의 정보를 올바르게 표현할 수 있지만, 프레임을 버리면서 정보 손실이 발생할 수 있고 요약 비디오에서 발생하는 프레임 간의 끊김 현상으로 요약된 비디오가 다소 부자연스러울 수 있다.

2.2 Object-based Video Condensation

객체 기반의 비디오 요약 방법은 원본 비디오를 짧게 요약하는 방법의 일종으로 원본 비디오에 나타나는 객체들을 추출하여 시·공간 도메인에서 사상하는 방식이다. 기존에 객체 기반의 비디오 요약 방법에 관한 다양한 연구가 있었다. [1], [18], [19], [20]은 객체 기반의 비디오 요약 알고리즘을 제안한 초기 연구로써 이후의 많은 객체 기반의 비디오 요약 연구들에 영향을 미쳤다. Kang은 시·공간 도메인에서의 사상을 이용해 원본 비디오를 짧게 요약하는 방법을 처음 제안하였다[1]. 이 방법은 최적의 사상 함수를 계산하기 위해서 최초 적합(first-fit)과 그래프 컷(graph-cut) 알고리즘을 통한 최적화 기법을 이용하였고 이를 통해 시각 정보를 최대화한 요약 비디오를 생성할 수 있었다. 하지만 이 방법으로 생성된 요약 비디오은 객체 간의 경계에서 이음매(seams) 현상이 발생하여 객체 간 경계가 시각적으로 부드럽지 않고 공간적 왜곡이 발생하는 문제점이 있다. 이를 해결하기 위하여 Rav Acha는 Kang이 제안한 방법과 유사한 방법이지만 객체를 오직 시간 축으로만 사상시키는 방법을 제안하였다[18]. 그리고 Pritch는 [20]에서 제안하는 방법을 확장해서 감시카메라와 같은 고정된 비디오가 대상인 시놉시스 비디오 생성 방법을 제안하였다[19]. [18]에서는 online phase와 response phase라는 개념을 정립하여 현재 대부분의 비디오 시놉시스 연구의 바탕이 되는 비디오 시놉시스 프레임워크를 구축하였다. Online phase는 객체 튜브를 추출하고 배경 이미지을 생성하는 과정이며 response phase는 online phase에서 추출한 객체 튜브를 시간 축으로 사상하여 객체들을 재배열하는 과정이다. 또한, 24시간 동안 촬영되는 연속된 비디오를 처리하는 방법과 조명 변화에 강인한 배경생성 알고리즘을 제안하였다. Xu도 이와 유사한 시놉시스 비디오 생성 알고리즘을 제안하였다[2]. 이 방법은 객체를 시·공간상에 존재하는 집합으로 문제를 정의하고 시각 정보가 최대화되는 객체들의 재배열을 계산하여 짧은 요약 비디오를 생성한다. 그리고 새로운 객체 튜브 재배열 알고리즘과 기존의 방법의 속도를 개선하는 방법이 등장하였다. Nie는 공간 도메인을 최대한 활용할 수 있는 비디오 시놉시스 방법을 제안하였다[3]. 이 방법은 시간 도메인뿐만 아니라 시·공간 도메인에서도 객체를 사상하는 방법을 제안하였다. 이를 통해 시놉시스 비디오의 길이를 줄이고 공간 도메인에서의 사상을 통해 객체 간의 겹침을 감소시킬 수 있었다. 그리고 다중계층 패치 재배치(multi-level patch relocation, MPR)를 통하여 공간적으로 확장된 배경 이미지을 생성하였고, 이렇게 확장된 배경 이미지에 객체들을 공간적으로 사상하였다. 또한, 비디오 시놉시스 알고리즘의 시간을 단축하기 위한 다양한 연구들도 있었다. [5], [21]에서는 객체 튜브를 재배열하는 과정을 최대 사후 확률(maximum a posteriori probability, MAP)로 수식화하고 계산량이 많은 별도의 최적화 과정 없이 시놉시스 테이블을 온라인으로 갱신하는 실시간 비디오 시놉시스 알고리즘을 제안하였다. 하지만 [4], [21]에서 제안한 온라인 알고리즘은 객체가 등장하고 사라지는 전체 프레임을 고려하지 않고 객체의 최초 등장 위치만 고려하기 때문에 시놉시스 비디오에서 발생하는 객체 간의 겹침을 정확하게 예측할 수 없다는 문제점이 있다. Zhu는 다른 온라인 비디오 시놉시스 방법을 제안하였다[22, 23]. 이 방법은 단계별 최적화 문제를 이용하여 객체 튜브를 재배열하고 그래픽 처리장치(graphics processing unit, GPU)와 다중 코어 프로세서(multi-core processor)를 이용한 병렬처리 기법으로 계산을 가속하는 방법을 제안하였다. 이를 통해 기존의 방법보다 훨씬 효율적인 메모리 사용 및 향상된 계산 속도로 실시간 처리가 가능한 시놉시스 생성 방법을 제안하였다. 그 외에도 비디오 시놉시스 알고리즘의 성능을 향상하기 위한 다양한 방법이 있었는데, 객체 튜브를 올바르게 만드는 방법, 원본 비디오에서 등장하는 객체의 움직임 정보를 최대한 유지하려는 방법, 이상 행동 감지를 위한 비디오 시놉시스 비디오와 같은 방법 또한 제안되었다[24, 25, 26, 27, 28].

2.3 Offline Video Synopsis Framework

객체 기반의 비디오 시놉시스는 원본 비디오에서 객체 튜브를 추출하고 튜브의 순서를 재배열하는 방식으로 비디오를 짧게 요약하는 방법이다. 객체 튜브의 순서를 재배열하는 방식으로 원본 비디오에서 서로 다른 시간에 등장한 객체를 동시에 나타낼 수 있다. 비디오 시놉시스는 오프라인 기반 프레임워크가 먼저 연구되었다[18]. 오프라인 비디오 시놉시스 방법은 여러 가지 제약조건을 이용하는데 객체 간 시간 순서를 유지하거나 객체 간의 겹침을 줄이고 전경 객체와 배경 이미지와의 일관성(consistency)을 유지하기 위한 목적으로 제약조건을 이용할 수 있다. 따라서 오프라인 비디오 시놉시스를 제약 최적화(constraint optimization)를 해결하는 문제로 접근할 수 있다. 일반적으로 오프라인 비디오 시놉시스는 식 (1)과 같은 에너지 함수를 최소화함으로써 각 객체 튜브가 시놉시스 비디오에서 최적의 시점에 등장할 수 있도록 결정한다.

$$E(l) = \sum_{i \in Q} E_u(l_i) + \sum_{i,j \in Q} E_p(l_i, l_j) \qquad (1)$$

식 (1)에서 Q는 원본 비디오에서 추출한 모든 튜브의 집합, l_i은 i번째 객체 튜브의 시간 레이블이며 오프라인 방법의 경우에는 $1 \le l_i \le M$의 범위를 지닌다. 여기에서 M은 시놉시스 비디오의 전체 프레임 수이다. 일반적으로 오프라인 기반의 비디오 시놉시스 방법들은 시놉시스 비디오의 전체 프레임 길이를 사용자가 사전에 정의한다. 그리고 E_u와 E_p는 각각 단항(unary), 쌍(pairwise) 에너지 함수이며 여러 가지 제약조건을 포함하기 위하여 사용한다. 하지만 식 (1)의 에너지 함수를 최소화하는 과정은 너무 큰 탐색 공간으로 계산 량이 많으며 최적화 과정에서 원본 비디오에 등장하는 모든 객체 튜브를 메모리에 저장해야 하므로 많은 크기의 저장 공간이 필요하다.

2.4 Online Video Synopsis Framework

오프라인 비디오 시놉시스와는 달리 온라인 기반의 비디오 시놉시스 프레임워크는 각 객체 튜브별로 최적화 과정을 수행한다[22]. 이는 단계별 최적화 문제로 볼 수 있다. 객체 튜브 i가 있을 때 일반적인 온라인 비디오 시놉시스의 에너지 함수는 식 (2)와 같다.

$$E(l_i) = E_u(l_i) + \sum_{j \in Q'} E_p(l_i|l_j) \qquad (2)$$

집합 Q'는 원본 비디오에서 추출한 전체 객체 튜브 집합 Q의 부분집합으로($Q' \subset Q$), 비디오 요약 시점에 따라 동적으로 변한다. 집합 Q'는 전체 객체 집합인 Q와 비교해서 크기가 훨씬 작다. 오프라인 방법과 비교하여 l_i의 범위인 $1 \le l_i \le n$가 훨씬 더 작아질 수 있으므로($n \ll M$) 빠른 연산이 가능하다. 일반적으로 온라인 비디오 시놉시스 방법에서는 식 (2)를 통해서 최적의 시놉시스 비디오를 생성한다. 이 방법은 식 (1)의 최적값에 근사해를 구하는 방법이다. 식 (1)과 비교해서 식 (2)의 탐색 공간은 훨씬 더 작으며 이로 인해 온라인 비디오 시놉시스는 훨씬 더 작은 메모리 사용량으로 실시간으로 객체 튜브를 처리한다.

III. Proposed Scheme

이번 장에서는 제안하는 객체 기반의 온라인 시놉시스를 구체적으로 설명한다.

3.1. Online Video Synopsis Framework

본 논문에서는 효율적인 비디오 시놉시스 알고리즘을 제안한다. 비디오 시놉시스의 목적은 원본 비디오에서 중요한 정보를 추출하여 원본 비디오보다 짧은 요약 비디오를 생성하는 것이다. 시놉시스 비디오를 생성하는 과정은 Fig. 1과 같다. Fig. 1의 상단 이미지는 비교적 시간이 긴 원본 비디오이다. 원본 비디오에서는 보라색 상의를 입은 보행자와 파란색 상의를 입고 자전거를 타고 있는 사람이 일정 시간 간격을 두고 서로 다른 시간에 등장한다. Fig. 1의 하단 이미지는 비디오 시놉시스로 요약된 시놉시스 비디오이다. 요약된 시놉시스 비디오에서는 원본 비디오에서 서로 다른 시간에 등장하던 두 객체가 같은 시간에 등장한다. 이러한 방식으로 비디오 시놉시스 알고리즘은 정보가 적다고 판단되는 불필요한 프레임을 제거하고 각 객체가 나타나는 시간을 재배열한다. 시놉시스 비디오는 원본 비디오와 비교해서 훨씬 더 객체가 집약적으로 등장하며 전체 비디오 시간이 짧다. 제안하는 비디오 시놉시스 프레임워크는 Fig. 2와 같다.

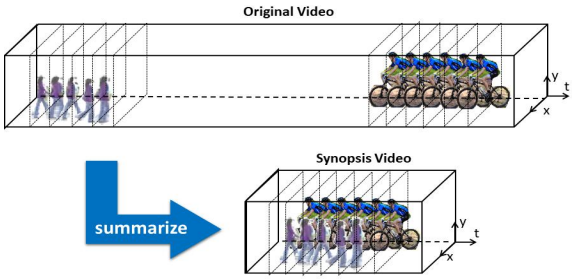


Fig. 1. Video Synopsis

비디오 시놉시스는 크게 객체 튜브 추출단계(object tube extraction), 객체 튜브 재배열단계(object tube rearrangement), 객체 합성단계(object stitching)로 구성된다.

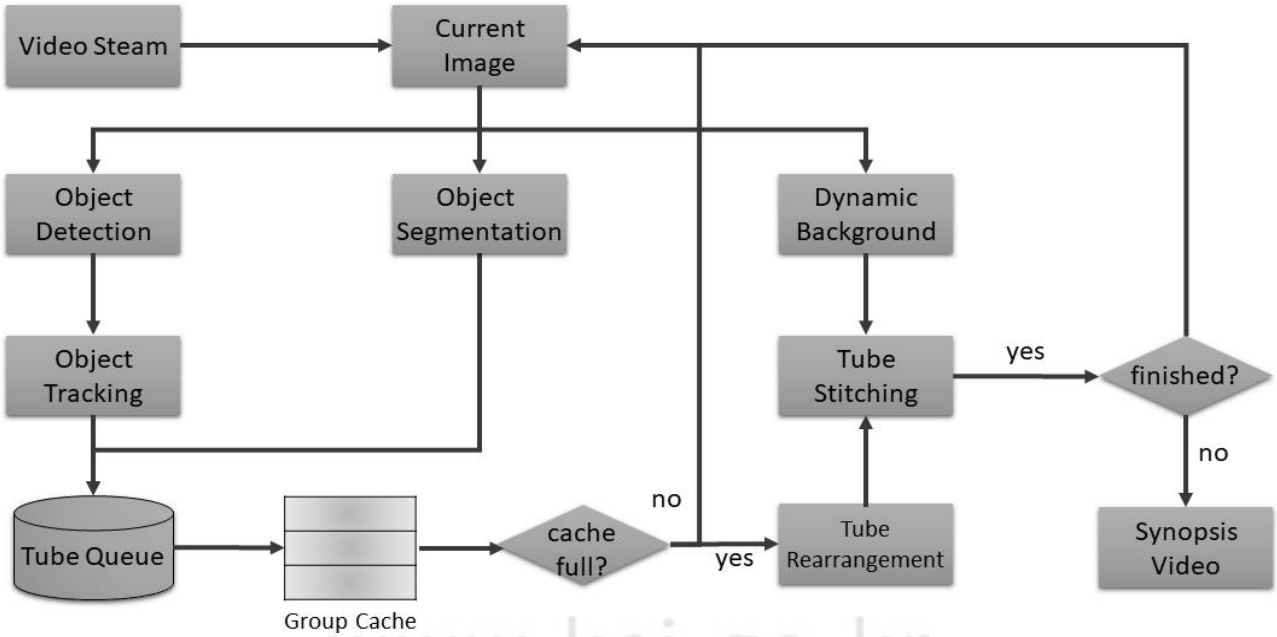


Fig. 2. Proposed Video Synopsis Framework

객체 튜브 추출단계는 원본 비디오 스트림의 각 프레임에서 객체를 추출하고 추출된 객체를 바탕으로 비디오 시놉시스의 기본 처리 단위인 객체 튜브를 생성한다. 객체 튜브 추출단계는 세부적으로 객체 검출(object detection)단계, 객체 추적(object tracking)단계 그리고 객체 분할(object segmentation)단계로 이루어진다. 객체 검출단계에서는 현재 이미지 프레임 내에 존재하는 관심 있는 객체(사람, 자동차 등)의 위치 정보를 얻는다. 객체 추적단계에서는 검출된 객체에 고유 아이디를 부여하고 시간상으로 변하는 객체의 위치 정보를 지속해서 파악하고 기록한다. 마지막으로 객체 분할단계에서는 이미지 내에 객체가 포함되는 화소 영역을 찾는다. 객체 분할단계에서는 객체의 이진 마스크(binary mask)를 생성한다. 생성된 이진 마스크는 객체 합성단계에서 전경(foreground) 객체와 배경(background) 이미지를 자연스럽게 합성하는 객체 합성단계에서 사용한다. 객체 튜브 추출단계와 병렬로 이루어지는 동적 배경생성(dynamic background construction)단계는 객체 합성단계에서 필요한 배경 이미지를 생성한다. 이 과정에서 배경 이미지는 일정 프레임의 가중평균(weighted average)을 이용하여 매 프레임 생성한다. 객체 튜브 추출단계로 객체 튜브가 만들어지면 각 객체 튜브는 객체 튜브 큐(object tube queue)에 차례로 저장한다.

튜브 재배열 단계에서는 비용함수를 정의하고 최적화 알고리즘을 통해서 시놉시스 비디오에서 각 객체 튜브의 최적 등장 시점을 계산한다. 대부분의 온라인 비디오 시놉시스 방법들은 튜브 재배열 과정을 단일 객체 튜브 단위로 진행하지만 제안하는 방법에서는 특정 튜브들을 그룹 단위로 군집화하여 최적화를 진행하며 이를 통해 비디오 시놉시스 알고리즘의 전체 연산량을 줄일 수 있고 시놉시스 비디오에서 원본 비디오에서 발생하는 객체의 상호작용을 보존할 수 있다. 최종적으로 객체 합성단계는 재배열된 객체와 배경을 합성하여 최종적인 시놉시스 비디오를 생성한다.

3.2. Object Tube Extraction

비디오 시놉시스 알고리즘의 기본 처리 단위는 객체 튜브이다. 객체 튜브는 시·공간상의 객체의 움직임을 표현하는 3차원 볼륨으로 Fig. 3는 이를 보여준다. 객체 튜브는 식 (3)와 같다.

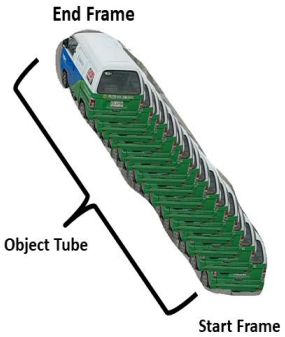


Fig. 3. Object Tube Example

$$Tube_i = \{O_i^s, O_i^{s+1}, ..., O_i^e\} \qquad (3)$$

O_i^s는 t번째 프레임에 등장하는 i번째 객체 인스턴스이다. 각 인스턴스는 전경 마스크와 위치 정보를 가지고 있다. $I(x,y,t)$는 $(1 \le x \le W, 1 \le y \le H, 1 \le t \le T)$를 만족하는 t번째 프레임의 위치(x,y)에서의 화소 값이다. 객체 튜브를 생성하기 위해서는 객체검출 및 추적이 필요하다. 최근 기계학습을 이용한 다양한 객체 검출 알고리즘이 연구되고 있다[29, 30, 31, 32, 33]. 본 논문에서는 효율적인 실시간 객체 검출을 위해 YOLOv2 객체 검출기를 사용하였다[29]. 그리고 객체 간 유사도를 기반으로 하는 객체 추적 알고리즘을 사용하였다[34]. 객체 추적을 위한 객체 간의 유사도는 식 (4)과 같다.

$$D(O_i^{t_1}, O_i^{t_2}) = d_u(O_i^{t_1}, O_i^{t_2}) \cdot d_c(H_i^{t_1}, H_i^{t_2}) \qquad (4)$$

식 (4)에서 $d_u(O_i^{t_1}, O_i^{t_2})$는 두 객체 튜브 $O_i^{t_1}$와 $O_i^{t_2}$간의 유클리디안 거리(euclidean distance)이고 $d_c(O_i^{t_1}, O_i^{t_2})$는 카이-제곱 거리(Chi-square distance)이다. 객체 튜브를 만들기 위해서 간단한 탐욕 알고리즘(greedy algorithm)을 통해 이전 프레임의 객체와 현재 프레임의 객체 간의 유사도가 큰 객체를 서로 연결한다. 하지만 이와 같은 객체 추적 알고리즘은 객체 검출단계에서 얻은 객체의 위치 정보를 이용하므로 객체 검출기의 성능의 큰 영향을 받는다. 따라서 객체 검출기가 객체를 잘못 검출하지 경우 객체 튜브의 생성에 영향을 미칠 수 있다. 제안하는 방법에서 이용한 YOLOv2 객체 검출기가 특정 프레임에서 객체를 검출하지 못하는 문제가 발생할 수 있지만, 객체 튜브의 시공간적 특성을 이용하여 이 문제를 보완할 수 있다.

3.3. Object Tube Rearrangement

객체 튜브를 재배열하는 과정은 비디오 요약을 위해 각 객체를 시놉시스 비디오 상에 사상시키는 과정으로 비디오 시놉시스 알고리즘의 핵심이다. 객체 튜브 재배열은 식 (5)와 같은 에너지 함수를 최소화함으로써 수행한다.

$$E(l) = E_l(l_i) + \sum_{j \in Q'} (\lambda_1 E_o(l_i|l_j) + \lambda_2 E_t(l_i|l_j)) \qquad (5)$$

식 (5)에는 세 가지 에너지 함수인 E_l, E_o, E_t가 있다. E_l는 시놉시스 비디오의 길이에 제약을 주는 에너지 함수로 새로운 객체 튜브가 시놉시스 비디오에 추가될 때 증가하는 시놉시스 비디오의 길이가 길어지면 E_l가 증가한다. 즉, E_l은 시놉시스 비디오의 길이가 짧을수록 작아진다. E_o는 시놉시스 비디오 상의 객체 간의 겹침 정도에 제약을 주는 에너지 함수이다. 시놉시스 비디오 상에 객체 간의 겹침이 많아지면 E_o의 값이 증가한다. 식 (5)에서 λ_1은 E_o의 가중치를 조절하는 매개변수이며 λ_1의 값을 조절하여 객체 간의 겹침 정도를 조절할 수 있으며 Fig. 4은 λ_1에 따른 겹침 정도를 보여준다. 그리고 마지막으로 E_t는 시놉시스 비디오에 등장하는 객체가 원본 비디오에서의

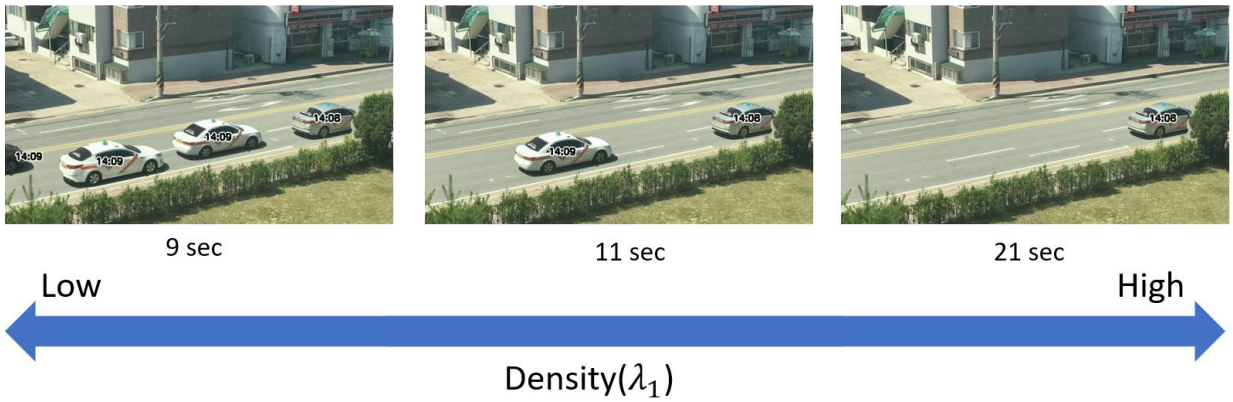


Fig. 4. Change in Density of Video Synopsis Results

상호작용을 보존하도록 제약을 주는 에너지 함수이다. E_t는 시놉시스 비디오에 등장하는 객체 중 원본 비디오에서의 상호작용을 유지하지 못하는 객체의 수에 따라서 증가한다. λ_2는 E_t의 가중치를 조절하는 매개변수이며 $\lambda_2 = 0$인 경우 시놉시스 비디오에서 객체 간 상호작용을 유지하지 않으며 반대로 $\lambda_2 = 1$인 경우 시놉시스 비디오에서 객체 간 상호작용을 유지하며 그 결과는 Fig. 5와 같다.

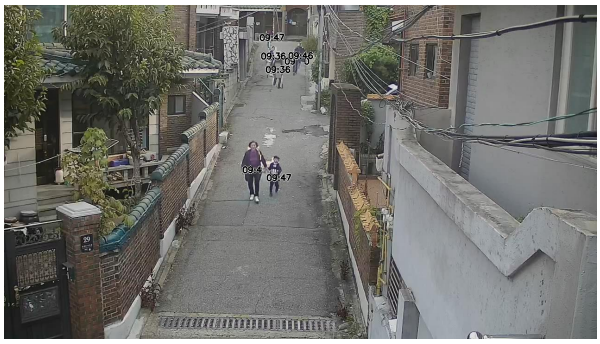


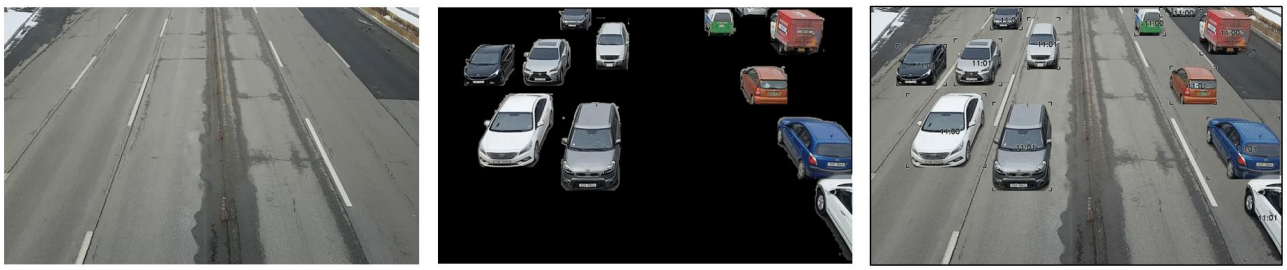
Fig. 5. Interaction Between objects

3.4. Stitching Background

객체 튜브를 재배열하는 단계를 수행하고 나면, 최종적인 시놉시스 비디오를 생성하기 위해서 객체와 배경 이미지를 합성(stitching)해야 한다. Fig. 6은 이 과정을 보이고 있다. 이때 객체 분할단계에서 얻은 객체의 이진 마스크를 사용하는데, 실제로는 객체 분할 시 이진 마스크가 객체가 속한 화소 영역을 완벽하게 표현하는 것은 어렵다. 이렇게 객체 화소를 부정확하게 표현한 이진 마스크를 이용하여 객체 튜브를 배경 이미지에 합성할 경우 결과 이미지가 부자연스러울 수 있다. 본 논문에서는 Poisson Image Editing 기법을 사용하여 이 문제를 해결하였다[35]. Poisson Image Editing은 그레디언트(gradient)의 변화를 이용하여 두 이미지를 자연스럽게 합성하는 방법이다. Poisson Image Editing을 통해서 객체와 배경 이미지를 자연스럽게 합성할 수 있으며 이 방법은 다른 객체 기반의 비디오 시놉시스 방법에서도 많이 사용되었다[3, 18, 19]. 특히 [18]에서는 Poisson Image Editing의 변형된 방법을 사용하는데 이는 기존의 방식에서 원본 이미지의 형태를 잘 보존할 수 있는 새로운 정규화 항을 추가하였으며 식은 식 (6)과 같다.

$$\min_s \sum_{\Omega} [(\triangle s - \triangle f)^2 + \lambda (s-f)^2], \quad s.t. \quad s_{\partial\Omega} = b_{\partial\Omega} \tag{6}$$

식 (6)에서 Ω는 경계 $\partial\Omega$를 가지는 이미지 도메인이다. f, b는 각각 전경 객체(튜브)와 배경 픽셀이다. 그리고 λ는 정규화 항의 가중치이다. [36]에서는 그레디언트 도메인에서의 이미지 합성이 아주 효율적이라는 연구결과를 제시하였으며 본 연구에서도 [18]에서 제시한 변형된 식을 이용하여 이미지 합성 방법을 사용하였다. 또한, 최근 다양한 이미지 합성 방법이 활발히 연구되고 있으며, 특히 딥러닝을 이용한 이미지 합성 방법을 활용한다면, 복잡한 장면에서도 더 자연스럽게 합성된 시놉시스 비디오를 생성할 수 있는 좋은 대안이 될 수 있다[37, 38].



(a) background Image (b) foreground(objects) image (c) synopsis image

Fig. 6. Object Stitching

IV. Experimental Results

4.1. Experiment Method

본 논문에서 제안하는 다중 객체검출 기반의 비디오 요약 방법의 성능을 평가하기 위해서 실험을 진행하였다. 실험에는 총 4개의 비디오를 이용하였고 해당 비디오 정보는 Table 1와 같다. Cross Road, Street, Sidewalk, Hall은 [17]에서 제공한 공개 데이터 세트로 세 가지는 실외 장면(Cross Road, Street, Sidewalk)이고 한 가지는 실내 장면이다(Hall). Cross Road는 다수의 자동차가 다니는 교차로에서 촬영한 장면이다. Street는 다수의 보행자와 주차된 자동차가 있는 거리에서 촬영한 장면이다. Sidewalk는 다수의 보행자와 오토바이가 등장하는 인도 주변에서 촬영한 장면이다. 마지막으로 Hall은 유일하게 실내에서 촬영한 장면으로 건물 홀을 출입하는 모습을 촬영한 장면이다. 네 가지 데이터 세트 모두 320 x 240의 해상도의 비디오이며 본 논문에서는 정량적인 평가를 위하여 해당 데이터 세트를 이용하였다. 실험 환경은 Table 2과 같다.

Table 1. Dataset Overview

Video	Duration	Resolution	FPS
Cross Road	01:04:59	320 x 240	18
Street	01:13:33	320 x 240	18
Hall	01:01:49	320 x 240	18
Sidewalk	00:58:18	320 x 240	30

Table 2. Experiments Environment

Operating System	Microsoft Windows 10 for Education
Hardware	CPU : Intel i7-6700K 4.00 GHz 8 cores GPU : NVIDIA GeForce GTX 1080 Ti (11GB) RAM : 32GB
Compiler	Microsoft Visual Studio 2015

4.2. Performance Metrics

정량적인 평가를 위해 세 가지 성능 측정 지표를 이용하였다. 우선, 첫 번째 지표는 프레임 감소율(frame reduction rate, FR)이다. 프레임 감소율은 원본 비디오가 시간상으로 얼마나 압축되었는지를 나타내는 지표로서 원본 비디오의 프레임 수 대비 시놉시스 비디오의 프레임 수가 얼마나 감소했는지를 계산한다. 프레임 감소율은 식 (7)과 같다. 프레임 감소율이 낮을수록 원본 비디오과 비교하여 시놉시스 비디오의 프레임 수가 더 많이 감소했음을 의미한다.

$$FR = \frac{\#frames\ of\ synopsis\ video}{\#frames\ of\ original\ video} \quad (7)$$

두 번째 지표는 평균 프레임 응집률(averag e frame compact rate, CR)이다. 평균 프레임 응집률은 원본 비디오가 비디오 요약으로 공간적으로 얼마나 압축되었는지를 나타내는 지표로서 한 프레임 내에 객체가 얼마나 등장하는지를 계산한다. 프레임 응집률은 식 (8)와 같이 나타낼 수 있다.

$$CR = \frac{1}{W \bullet H \bullet \#V_S} \sum_{s=1}^{\#V_s} \sum_{x=1}^{W} \sum_{y=1}^{H} \{1 | \text{if } v(x,y,s) \in \text{foreground in } V_s\} \quad (8)$$

식 (8)에서 $v(x,y,s)$는 시놉시스 비디오 V_S의 s번째 프레임의 화소(x,y)를 나타낸다. W와 H는 각각 프레임의 너비(width)와 높이(height)를 의미하며 $\#V_S$는 시놉시스 비디오 V_S의 총 프레임 수를 나타낸다. 높은 평균 프레임 응집률은 객체의 응집도가 높음을 의미한다.

세 번째 지표는 시간순서 이상률(chronological disorder ratio, CDR)이다. 이 지표는 시놉시스 비디오에서 등장하는 객체 간의 시간적 어긋남을 측정하는 지표로 원본 비디오 대비 시간순서 이상률은 시놉시스 비디오에 등장하는 객체 간의 시간순서의 어긋남을 측정한다. 시간순서 이상률은 식 (9)과 같이 나타낼 수 있다.

$$CDR = \frac{1}{\#F} \sum_{P_m \in F} \sum_{P_{m'} \in F} \{1 | D^O(P_m, P_{m'}) D^S(P_m, P_{m'}) < 0\} \quad (9)$$

F는 시놉시스 비디오에 등장하는 모든 객체를 포함하는 집합이며, $\#F$는 집합 F에 속한 객체들의 수이다. 그리고 $D^O(P_m, P_{m'})$, $D^S(P_m, P_{m'})$는 각각 원본, 시놉시스 비디오에서 두 객체가 처음 등장하는 프레임 인덱스의 차이 값이다. 시놉시스 비디오에서 시간상으로 어긋난 객체가 많을수록 시간순서 이상률이 증가한다.

Table 3. Experimental Results

Dataset	Method	FR	CR	CDR
Cross Road	[18[	-	0.171	155.395
	[4]	0.181	0.137	0.127
	[21]	0.112	0.175	0.059
	Proposed	**0.122**	**0.182**	**0.044**
Street	[18]	-	0.049	144.585
	[4]	0.237	0.058	0.668
	[21]	0.269	0.064	0.619
	Proposed	**0.250**	**0.072**	**0.720**
Hall	[18]	-	0.039	31.278
	[4]	0.215	0.044	0.251
	[21]	0.169	0.040	0.134
	Proposed	**0.159**	**0.035**	**0.092**
Sidewalk	[18]	-	0.067	35.297
	[4]	0.174	0.069	0.205
	[21]	0.165	0.069	1.596
	Proposed	**0.145**	**0.082**	**0.125**

4.3. Results

제안하는 방법의 성능을 [4], [18], [21]의 방법과 비교하였다. [18]에서 제안하는 방법의 MRF(Markov random field) 알고리즘은 계산 복잡도가 너무 크다. [18]의 방법은 원본 비디오 내의 객체의 수가 늘어날수록 계산 복잡도가 커지므로 비디오 내의 전경 객체의 수를 줄여서 실험을 진행하였다. 수천 개의 전경을 가지고 있는 긴 시간의 비디오의 경우 여러 개의 짧은 비디오로 나누어서 실험을 진행하였다. 이로 인해 각 비디오는 균일한 전경 객체 수를 가지고 있다. 이 경우에는 각각의 비디오가 같은 계산 복잡도를 가진다. Table 3는 제안하는 방법과 [4], [18], [21]에서 제안하는 방법의 시놉시스 비디오 FR, CR, CDR을 비교한 결과이다. 실험을 통해서 [4], [21]에 비교하여 제안하는 방법의 우수함을 확인할 수 있었다. [18]의 경우에는 시놉시스 비디오의 길이가 사용자가 사전에 정하는 매개변수이기 때문에 FR를 측정하는 실험에서 제외하였다. 또한, 실험을 통해서 [4], [18], [21]에 비교하여 제안하는 방법의 CR과 CDR이 우수함을 확인할 수 있었다. 특히 CDR의 결과를 통해 본 논문에서 제안한 그룹 방식의 객체 튜브 재배열 알고리즘을 사용하면 원본 비디오에서 나타나는 객체의 상호작용 및 시간 순서를 효과적으로 유지할 수 있음을 알 수 있었다.

IV. Conclusions

본 연구에서는 다중 객체 검출 기반의 효율적인 온라인 비디오 시놉시스 알고리즘을 제안하였다. 비디오 시놉시스는 객체 기반의 요약 알고리즘으로 긴 시간의 비디오를 짧게 요약하는 기법이다. 시놉시스 비디오에서 원본 비디오에 등장하는 객체의 특성을 유지하기 위해서는 원본 비디오에서의 객체의 움직임과 객체 간 상호작용을 유지해야 한다. 기존에 제안된 비디오 시놉시스 방법들은 객체의 움직임을 보존하지만, 대부분은 객체 간의 상호작용을 고려하지 않는다. 본 연구에서 제안하는 방법은 원본 비디오에서 같은 시간에 나타나는 객체 중에 상호작용이 발생했다고 판단하는 객체들을 군집화하여 시놉시스 비디오에서도 같은 시간대에 등장하도록 알고리즘을 설계하였다. 또한, 시놉시스 비디오는 사용자의 이용 목적에 따라서 그 결과가 달라질 수 있어야 한다. 이를 위해서 제안하는 방법에서는 새로운 비용함수를 제안하였고 새롭게 정의한 비용함수의 매개변수의 값에 따라 시놉시스 비디오의 전체 길이, 객체 간의 겹침 정도를 동적으로 조절할 수 있도록 설계하였다. 그리고 제안하는 방법의 성능을 평가하기 위하여 실험을 진행하였고 제안하는 방법의 우수성을 입증하였다. 제안하는 방법은 비용함수의 일부 매개변수를 사용자 수준에서 직접 조절할 수 있다는 장점이 있지만, 매개변수의 적절한 값을 직접 찾아야 하는 문제가 있다. 따라서 향후 사용자가 적절한 매개변수를 선택할 수 있도록 요약 비디오 길이를 사전에 예측하는 방법을 연구할 필요성이 있을 것으로 생각한다.

REFERENCES

[1] H. W. Kang, Y. Matsushita, X. Tang, and X. Q. Chen, "Space-time video montage," 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 1331-1338, New York, USA, June 2006.

[2] M. Xu, S. Z. Li, B. Li, X. T. Yuan, and S. M. Xiang, "A set theoretical method for video synopsis," Proceedings of the 1st ACM international conference on Multimedia information retrieval, pp. 366-370, Vancouver, Canada, October 2008.

[3] Y. Nie, C. Xiao, H. Sun, and P. Li, "Compact video synopsis via global spatiotemporal optimization," IEEE transactions on visualization and computer graphics, Vol. 19, No. 10, pp. 1664-1676, October 2013.

[4] C. R. Huang, H. C. Chen, and P. C. Chung, "Online surveillance video synopsis," 2012 IEEE International Symposium on Circuits and Systems, pp. 1843-1846, Seoul, South Korea, May 2012.

[5] E. Bennett, P. Eric, and L. McMillan, "Computational time-lapse video," ACM Transactions on Graphics, Vol. 26, No. 3, pp. 102, July 2007.

[6] N. Petrovic, N. Jojic, and T. S. Huang, "Adaptive video fast forward," Multimedia Tools and Applications, Vol. 26, No. 3, pp. 327-344, August 2005

[7] J. Nam, and A. H. Tewfik, "Video abstract of video," 1999 IEEE Third Workshop on Multimedia Signal Processing (Cat. No. 99TH8451), pp. 117-122, September 1999.

[8] X. Zhu, X. Wu, J. Fan, A. K. Elmagarmid, and W. G. Aref, "Exploring video content structure for hierarchical summarization," Multimedia Systems, Vol. 10, No. 2, pp. 98-115, August 2004.

[9] T. Liu, X. Zhang, J. Feng, and K. T. Lo, "Shot reconstruction degree: a novel criterion for key frame selection," Pattern recognition letters, Vol. 25, No. 12, pp. 1451-1457, September 2004.

[10] B. T. Truong, and S. Venkatesh, "Video abstraction: A systematic review and classification," ACM transactions on multimedia computing, communications, and applications, Vol. 3, No. 1, February 2007.

[11] C. Gianluigi, and S. Raimondo, "An innovative algorithm for key frame extraction in video summarization," Journal of Real-Time Image Processing, Vol. 1, No. 1, pp. 69-88, March 2006.

[12] H. Liu, W. Meng, and Z. Liu, "Key frame extraction of online video based on optimized frame difference," 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery, pp. 1238-1242, Sichuan, China, May 2012.

[13] C. M. Taskiran, Z. Pizlo, A. Amir, D. Ponceleon, and E. J. Delp, "Automated video program summarization using speech transcripts," IEEE Transactions on Multimedia, Vol. 8, No. 4, pp. 775-791, August 2006.
[14] Y. F. Ma, X. S. Hua, L. Lu, and H. J. Zhang, "A generic framework of user attention model and its application in video summarization," IEEE Transaction on multimedia, Vol. 7, No. 5, pp. 907-919, October 2005.
[15] X. Zhu, C. C. Loy, and S. Gong, "Video synopsis by heterogeneous multi-source correlation," Proceedings of the IEEE International Conference on Computer Vision, pp. 81-88, Sydney, Australia, December 2013.
[16] S. Benini, P. Migliorati, and R. Leonardi, "Hidden Markov models for video skim generation," Eighth International Workshop on Image Analysis for Multimedia Interactive Services, pp. 6-6, Santorini, Greece, June 2007.
[17] S. Benini, P. Migliorati, and R. Leonardi, "A statistical framework for video skimming based on logical story units and motion activity," 2007 International Workshop on Content-Based Multimedia Indexing, pp. 152-156, Bordeaux, France, June 2007.
[18] Y. Pritch, A. Rav-Acha, and S. Peleg, "Nonchronological video synopsis and indexing," IEEE Trans. on pattern analysis and machine intelligence, Vol. 30, No. 11, pp. 1971-1984, November 2008.
[19] Y. Pritch, A. Rav-Acha, A. Gutman, and S. Peleg, "Webcam synopsis: Peeking around the world," 2007 IEEE 11th International Conference on Computer Vision, pp. 1-8, Rio de Janeiro, Brazil, December 2007.
[20] A. Rav-Acha, Y. Pritch, and S. Peleg, "Making a long video short: Dynamic video synopsis," 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 435-441, New York, USA, July 2006.
[21] C. R. Huang, P. C. Chung, D. K. Yang, H. C. Chen, and G. J. Huang, "Maximuma PosterioriProbability Estimation for Online Surveillance Video Synopsis," IEEE Transactions on circuits and systems for video technology, Vol. 24, No. 8, pp. 1417-1429, August 2014.
[22] S. Feng, Z. Lei, D. Yi, and S. Z. Li, "Online content-aware video condensation," 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2082-2087, Providence, USA, June 2012.
[23] J. Zhu, S. Feng, D. Yi, S. Liao, Z. Lei, and S. Z. Li, "High-performance video condensation system," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 25. No. 7, pp. 1113-1124, July 2014.
[24] L. Sun, J. Xing, H. Ai, and S. Lao, "A tracking based fast online complete video synopsis approach," Proceedings of the 21st international conference on pattern recognition, pp. 1956-1959, Tsukuba, Japan, November 2012.
[25] M. Lu, Y. Wang, and G. Pan, "Generating fluent tubes in video synopsis," 2013 IEEE international conference on acoustics, speech and signal processing, pp. 2292-2296, Vancouver, Canada, May 2013.
[26] W. Fu, L. Gui, H. Lu, and S. Ma, "Online video synopsis of structured motion," Neurocomputing, Vol. 135, No. 5, pp. 155-162, July 2014.
[27] R. Zhong, R. Hu, Z. Wang, and S. Wang, "Fast synopsis for moving objects using compressed video," IEEE signal processing letters, Vol. 21, No. 7, pp. 834-838, July 2014.
[28] W. Lin, Y. Zhang, J. Lu, B. Zhou, J. Wang, and Y. Zhou, "Summarizing surveillance videos with local-patch-learning-based abnormality detection, blob sequence optimization, and type-based synopsis," Neurocomputing, Vol. 155, No. 1, pp. 84-98, May 2015.
[29] J. Redmon, and A. Farhadi, "YOLO9000: better, faster, stronger," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7263-7271, Honolulu, USA, July 2017.
[30] T. Y. Lin, P. Goyal, R. Girshick, and K. He, "Focal loss for dense object detection," Proceedings of the IEEE international conference on computer vision, pp. 2980-2988, Venice, Italy, October 2017.
[31] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask r-cnn," Proceedings of the IEEE international conference on computer vision, pp. 2017, pp. 2961-2969, Venice, Italy, October 2017.
[32] S. Zhang, L. Wen, X. Bian, Z. Lei, and S. Z. Li, "Single-shot refinement neural network for object detection," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4203-4212, Salt Lake City, USA, June 2018.
[33] Q. ZHAO, T. Sheng, Y. Wang, Z. Tang, Y. Cheng, L. Cai, and H. Ling, "M2det: A single-shot object detector based on multi-level feature pyramid network," Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, No. 1, pp. 9259-9266, July 2019.
[34] J. Jin, F. Liu, Z. Gan, and Z. Cui, "Online video synopsis method through simple tube projection strategy," 2016 8th International Conference on Wireless Communications & Signal Processing, pp. 1-5, Yangzhou, China, October 2016.
[35] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," ACM Transactions on graphics, Vol. 22, No. 3, pp. 313-318, July 2003.
[36] A. Agarwala, "Efficient gradient-domain compositing using quadtrees," ACM Transactions on Graphics, Vol.

26, No. 3, July 2007.
[37] Z. Fangneng, J. Huang, and S. Lu, “Adaptive Composition GAN towards Realistic Image Synthesis,” arXiv preprint, arXiv:1905.04693, May 2019.
[38] H. Wu, S. Zheng, J. Zhang, and K. Huang, “Gp-gan: Towards realistic high-resolution image blending,” arXiv preprint, arXiv:1703.07195, August 2019.

Authors

JaeWon Lee received the B.S. degree in Department of Computer and Communications Engineering at Kangwon National University, in 2017 and has been in M.S. degree program since 2017. His research interests are in the areas of machine learning, and computer vision.

DoHyun Kim received a B.S. degree and an M.S. degree in 2016. He is now a Ph.D. candidate in Department of Computer and Communications Engineering at Kangwon National University. His research interests are in the areas of machine learning, and computer vision

Yoon Kim received a B.S. degree in 1993, an M.S. degree in 1995, and a Ph.D. degree in 2003, in electronic engineering with the Department of Electronic Engineering from Korea University In 2004, he joined the Department of Computer and Communications Engineering, Kangwon National University, where he is currently a professor. From 1995 to 1999, he was with the LG-Philips LCD Co., where he was involved in research and development on digital image equipment. His research interests are in the areas of machine learning, multimedia communications, and computer vision.