

하이라이트 비디오 생성을 위한 데이터셋 구축을 위한 비디오 탐색 알고리즘

송기연¹, 이재환²

¹한림대학교 데이터사이언스학부 학부생

²한림대학교 경영학과 교수

rldus1203@naver.com, jaehwan@hallym.ac.kr

Video Retrieval Algorithm for Building a Dataset for Highlight Video Generation

Gi-Yeon Song¹, Jaehwan Lee²

¹Division of Data Science, Hallym University

²Dept. of Business Administration, Hallym University

Abstract

본 연구에서는 특정 비디오에서 추출된 비디오 클립이 어떤 비디오에서 추출된 것인지 탐색하는 알고리즘을 제안한다. 국내 이스포츠 리그 중 하나인 LCK의 경기 영상과 하이라이트 영상을 수집하여 알고리즘의 성능을 테스트하였다. 본 연구에서 제안한 알고리즘은 하이라이트 비디오 추출 모델 개발에 필요한 비디오-하이라이트 클립 데이터셋을 구축하는 데 도움이 될 것이라 기대한다.

인지 추출하여 하이라이트 영상 제작 인공지능 학습 데이터셋을 구축하는 데 사용될 수 있다.

1. Introduction

비디오를 효과적으로 요약하는 것은 영상 콘텐츠를 다양한 용도로 활용하는 데 중요한 요소이다. 특히 스포츠 경기나 게임과 같은 실시간 이벤트의 하이라이트를 자동으로 추출하는 기술은 많은 관심을 받고 있다[1]. [2]의 연구에서는 시청자 채팅 반응을 활용하여 멀티모달 및 캐릭터 수준의 CNN-RNN 모델 아키텍처를 사용한 비디오 하이라이트를 예측하는 방법을 제안하였다. 이러한 하이라이트 추출 알고리즘은 생성형 인공지능 기술의 한 영역으로 HitPaw Edimakor¹⁾, Opus Clip²⁾과 같은 플랫폼이 그 예이다.

그러나 하이라이트 추출 알고리즘을 효과적으로 학습시키기 위해서는 풀 영상과 하이라이트 영상 사이의 정확한 대응 관계를 알고 있어야 한다. 하지만 이런 데이터셋을 구축하는 것은 매우 어려운 문제이다.

따라서 본 연구에서는 비디오-하이라이트 클립 데이터셋을 구축하기 위한 비디오 탐색 알고리즘을 제안한다. 제안된 알고리즘은 하이라이트 클립이 어떤 풀 영상에 속하는지 찾아내고, 풀 영상의 어느 위치

2. Video Matching

두 비디오에서 동일한 프레임을 찾는 것은 매우 어렵고 낮은 매칭 가능성을 가지고 있다[3]. 따라서 이미 하이라이트 클립이 존재한다고 하여도, 하이라이트 클립이 어떤 영상의 어느 부분에서 추출됐는지 탐색하는 것은 도전적인 작업이다.

비디오 매칭 문제는 문자열에서 부분 문자열을 찾는 문자열 매칭(string-matching) 문제와 유사하게 보인다. 대표적 문자열 매칭 알고리즘에는 보이어-무어 알고리즘이 있다[4]. 보이어-무어 알고리즘은 문자열 검색을 위한 효율적인 알고리즘으로, 특정 패턴을 빠르게 찾는 데 사용된다. 이 알고리즘은 검색 대상 문자열을 오른쪽에서 왼쪽으로 스캔하면서 패턴의 일치 여부를 확인하고, 검색 시 패턴의 마지막 문자부터 시작하여 패턴 내의 문자가 검색 대상 문자열에서 발견되는 위치를 찾아낸다. 패턴 내의 문자가 검색 대상 문자열에 없는 경우에는 한 번에 여러 문자를 건너뛰어 검색 속도를 높인다[5].

문자열 매칭 문제를 해결하기 위한 다양한 알고리즘이 제시되어있다. 영어 알파벳은 대소문자를 포함하여 52자인데, 문자열 매칭은 특수문자나 다양한 언어를 고려하여도 한 글자가 가질 수 있는 경우의

1) <https://www.hitpaw.net/>

2) <https://www.opus.pro/>

수가 제한적이라고 볼 수 있다. 반면 비디오의 경우 HD영상은 가로 720개, 세로 1,280개의 픽셀로 구성되며 각 픽셀은 RGB라는 3차원 값을 가진다. HD영상의 1개 프레임이 가질 수 있는 경우의 수는 2,764,800개로 문자열에 비해 매우 많으며, 따라서 문자열 매칭 알고리즘을 비디오 매칭 알고리즘에 적용하긴 어렵다.

3. Algorithm

본 연구에서 제안하는 알고리즘은 다음과 같다. 풀 영상을 F, 하이라이트 영상을 H라고 하자. F와 H는 같은 해상도를 갖고 있다. 영상 F는 n개의 프레임으로 구성되어있고 영상 H는 m개의 프레임으로 구성되어있다.

$$F = \{f_1, f_2, \dots, f_n\}$$

$$H = \{h_1, h_2, \dots, h_m\}$$

```

begin
  Define sampling area randomly
  for i=1...n do
    Compare sampling area of fi and h1;
    if same
      Compare entire frame of fi+m-1 and hm;
      if same
        return i;
    end
  end
  return -1;
end
    
```

Figure 1. Proposed algorithm

4. Research Methods

제안한 알고리즘의 성능을 테스트하기 위해 국내 이스포츠 종목 중 하나인 리그 오브 레전드 게임의 프로 리그 LCK(Legendary Champions Korea) 영상을 수집하였다. LCK의 유튜브 채널에는 전체 경기 영상과 하이라이트 영상이 모두 업로드되어있으며, 하이라이트 영상은 Figure 2와 같이 전체 영상의 일부 클립을 연결한 형태로 구성되어있다.

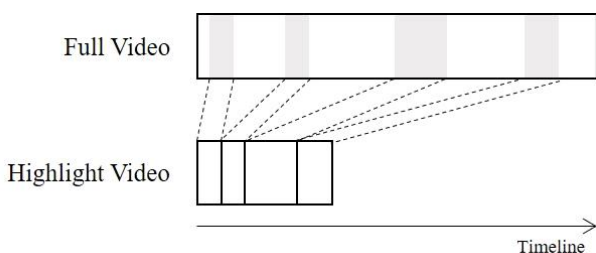


Figure 2. Full video and highlight video

실험에 사용할 데이터 셋은 LCK 2024 Spring 시즌 기준 55개의 풀 영상(실시간 스트리밍), 98개의 하이라이트 영상이다. 경기 영상은 1주차부터 9주차까지의 정규시즌, 1라운드부터 4라운드까지의 플레이오프, 최종 결승전 경기를 포함하고 있다. 데이터 수집 후 탐색 시간 등의 실험을 수행할 예정이다.

5. Conclusion

본 연구에서는 비디오 매칭 알고리즘을 제안하였고, 제안한 알고리즘의 성능을 LCK 영상을 활용하여 테스트하였다. 제안된 알고리즘은 프레임을 분할하여 평균을 계산하는 기존 알고리즘에 비해 뛰어난 성능을 보일 것이라 예상된다.

제안한 알고리즘을 활용해 하이라이트 영상 정답 데이터셋을 구축할 수 있으며, 구축된 데이터셋은 하이라이트 영상 제작 인공지능의 개발과 이를 통한 시청 경험을 개선하는 데에 활용될 수 있을 것으로 기대한다.

References

- [1] 최승혁, 박상준, 한재현, & 이현아. (2019). 게임 시청자 실시간 한글 채팅의 언어 자질에 기반한 하이라이트 탐지. 디지털콘텐츠학회논문지, 20(2), 335-342.
- [2] Fu, C. Y., Lee, J., Bansal, M., & Berg, A. C. (2017). Video highlight prediction using audience chat reactions. EMNLP.
- [3] Sand, P., & Teller, S. (2004). Video matching. Acm transactions on graphics (tog), 23(3), 592-599.
- [4] Choudhary, R., Rasool, A., & Khare, N. (2012). Variation of Boyer-Moore String Matching Algorithm: A Comparative Analysis. International Journal of Computer Science and Information Security, 10(2), 95-101.
- [5] Chakraborty, U. K., & Sharma, P. (2016). String Searching Algorithm: A Predictive Approach. International Conference on Communication Computing & Visualization.