

소리 파형 기반 동영상 간 시간축 동기화 기법

*김신 **윤경로

건국대학교 컴퓨터공학과

*new.xin22@gmail.com

Timeline Synchronization of video clips based on sound waveform

*Kim, Shin **Yoon, Kyoungro

Department of Computer Science Engineering, Konkuk University

요약

파노라마 영상은 최근 몇 년 사이 스마트 폰을 사용하는 일반 사용자에게 손쉽게 제작할 수 있게 되었다. 하지만 파노라마 동영상은 기술상의 어려움이 있다. 파노라마 동영상을 제작하기 위해서는 두 대 이상의 스마트 폰으로 동시에 촬영을 하여야 하는데, 하드웨어의 차이로 인해 시간축을 동기화하는데 어려움이 발생한다. 이러한 어려움을 극복하기 위해 영상 정보를 이용한 시간축 동기화도 시도되었다.

하지만 영상 정보를 이용한 시간축 동기화에도 어려움이 있다. 영상 정보를 이용하기 때문에 계산복잡도가 높고 영상 내 움직임으로 인해 시간축 동기화 정확도도 떨어진다. 본 논문에서는 영상 정보를 이용하는 대신 소리 파형을 이용한 시간축 동기화 기법을 제안한다. 소리 파형 이미지를 이용하여 시간축을 동기화하기 때문에 정확도가 향상되고 시간축 동기화 프로세싱 시간이 단축되는 것을 확인할 수 있다.

1. 서론

최근 들어서 파노라마 영상은 일반 사용자들도 스마트 폰으로 이용하여 쉽게 제작할 수 있게 되었다. 하지만 아직까지 스마트 폰을 이용한 파노라마 동영상 제작에는 어려움이 있다. 동시에 촬영할 때 여러 스마트 폰으로 다각도에서 촬영하여 파노라마 동영상을 제작하여야 하는데, 스마트 폰 내 하드웨어 차이로 인해 동시 시간을 파악하기가 어렵기 때문이다.

하드웨어 간의 차이로 인해 여러 동영상 간의 동일 시간을 구하기 힘든 점을 극복하기 위해 영상 정보를 이용하여 여러 동영상간의 시간축 동기화하는 기법[1, 2, 3, 4]이 연구되었다. 하지만 영상 정보를 이용한 방식은 계산량이 많고 움직임이 있는 경우에는 정확도가 떨어지는 문제점이 있다.

동영상은 영상과 오디오로 구성되어 있다. 즉, 오디오도 동영상간의 시간축 동기화에 사용할 수 있는 수단임을 의미한다. 흔히 알고 있는 오디오 특징으로는 오디오 핑거프린팅, 스펙트로그램, 소리 파형 등이 있다. 이 중 쉽게 이해하고 다룰 수 있는 것은 소리 파형이다.

본 논문에서는 동영상 간 시간축 동기화를 위해 소리 파형을 이용한 시간축 동기화 알고리즘을 제안한다. 2장에서는 영상 정보 기반 시간축 동기화 방법 및 단점에 대해 설명하고 3장에서는 본 논문에서 제안하는 소리 파형 기반 시간축 동기화 기법에 대해 서술한다. 4장에서 실험 결과에 대해 설명하고 5장에서 본 논문의 결론을 서술한다.

2. 영상 정보 기반 시간축 동기화

영상 정보를 이용한 시간축 동기화 기법[1, 2, 3, 4]이 수년간 연구되었다. 이 중 SURF 알고리즘[5]을 이용하여 영상 특징을 추출하여 매칭하는 방식의 시간축 동기화 알고리즘[1, 2]도 연구 되었는데, 동영상 내 모든 프레임에서 추출한 특징들을 비교하여 유사도가 가장 높을 때가 동일한 시간축임을 유추하는 방식이다.

하지만 위와 같이 특징을 추출하는 방식같이 영상 정보를 이용할 때의 문제점은 크게 두 가지가 있다. 첫 번째로 동영상 내에 움직임이 있기 때문에 시간축 동기화 정확도가 떨어진다. 동영상 내의 큰 움직임, 작은 움직임 모두 시간축 동기화 정확도에 영향을 끼친다. 또한 움직임이 없는 경우에는 시간축 동기화가 불가능하다는 문제점도 있다. 두 번째로, 모든 프레임에서 키포인트를 찾고 특징을 추출하기 때문에 동영상의 길이가 길수록 오래 걸린다.

3. 소리 파형 기반 시간축 동기화 기법

본 논문에서는 동영상 간의 시간축 동기화를 위해 소리 파형을 이용하는 방식을 제안한다. 소리 파형을 이용한 시간축 동기화를 위해 먼저 ffmpeg을 통해 여러 동영상에서 추출한 파형을 이미지로 만들고 파형 이미지간 비교를 위해 파형 영상을 이진 영상으로 만들었다.

여러 동영상 중 가장 짧은 동영상의 파형 영상의 절반(그림 1에서

의 파란색 박스 부분 이미지)을 기준 파형 영상으로 삼아 다른 동영상의 파형 영상(그림 1의 아래 이미지)과의 비교를 통해 유사도를 측정하였다. 파형 영상간 비교 연산으로 XOR 연산을 이용하였으며, XOR 연산 후 흰 색의 픽셀 수가 가장 적을 때를 동일한 시간대임을 가정하였다.

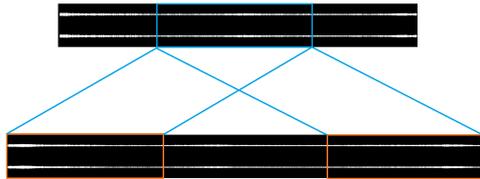


그림 1. 소리 파형을 이용한 시간축 동기화 알고리즘

4. 실험 결과

본 논문에서는 실험을 위해 스마트 폰으로 취득한 초당 60 프레임 동영상 세트 3개를 이용하였다. 영상 정보를 이용한 시간축 동기화 성능[2]과 소리 파형 기반 시간축 동기화 성능을 비교하기 위해 동일한 테스트 세트를 이용하여 시간축 동기화 정확도를 측정하였다. 본 논문에서 제안한 소리 파형을 이용한 시간축 동기화 기법이 영상 정보를 이용한 시간축 동기화 기법보다 정확도가 향상된 것을 표1을 통해 확인할 수 있다.

표 1. 동기화 이후 동영상 간의 시간축 프레임 차이 (프레임)

	영상 정보 기반 알고리즘[2]	소리 파형 기반 알고리즘
테스트 세트 1	7	3
테스트 세트 2	15	5
테스트 세트 3	50	3

표2를 통해 시간축 동기화 프로세싱에 걸리는 시간이 영상 정보를 이용했을 때에 비해 압도적으로 줄어든 것을 확인할 수 있다. 영상 정보를 이용하는 경우 모든 프레임에서 특징을 추출해야 하는 반면, 소리 파형을 이용하는 경우 ffmpeg을 이용하여 소리 파형을 영상으로 만든 후 이진 영상을 가지고 연산하기 때문에 계산복잡도가 줄었기 때문이다. 평균적으로 0.07%로 계산량이 줄어든 것을 표2를 통해 알 수 있다.

표 2. 시간축 동기화 프로세싱 시간 (초)

	영상 정보 기반 알고리즘[2]	소리 파형 기반 알고리즘	향상도 (%)
테스트 세트 1	844	1	0.11
테스트 세트 2	1334	1	0.07
테스트 세트 3	2138	1	0.05

그림2를 통해 영상 정보 기반 시간축 동기화[4] 알고리즘을 이용한 결과물과 본 논문에서 제안한 소리 파형 기반 시간축 동기화 알고리즘을 이용한 결과물을 확인할 수 있으며 영상 정보 기반보다 소리 파형 기반의 결과물이 좀 더 정확한 것을 알 수 있다.



(a)



(b)

그림 2. 소리 파형 기반 시간축 동기화 결과(a)와 영상 정보 기반 시간축 동기화 결과(b) 비교 예시

5. 결론

본 논문에서는 동영상간의 시간축 동기화를 위해 소리 파형을 이용한 기법을 제시하였다. 영상 정보를 이용한 시간축 동기화 기법에 비해 정확도가 향상되었으며 계산량도 줄어든 것을 확인할 수 있었다.

하지만 본 논문에서 제시한 기법은 소리가 없는 상황에서는 시간축 동기화를 할 수 없는 단점을 가지고 있으며, 소리가 없는 상황에서도 동영상 간의 시간축을 구할 수 있는 기법을 연구해야 할 것이다.

감사의 글

이 논문은 2017년 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임(B0126-16-1013 퍼즐형 Ultra-wide viewing 공간미디어 생성 및 소비 기술 개발)

참 고 문 헌

[1] 고명준, 윤경로, 이미지 매칭 기반 동영상간 시간축 동기화 기법. 한국방송미디어공학회 학술발표대회 논문집, 144-145. 2016년 11월.

[2] 김신, 윤경로, 동영상간 시간축 동기화 최적화 기법, 한국방송미디어공학회 학술발표대회 논문집, 109-110, 2017년 6월.

[3] Shrestha, Prarthana, et al. "Synchronization of multiple video recordings based on still cameraflashes." Proceedings of the 14th ACM international conference onMultimedia. ACM, 2006..

[4] Lei, Cheng, and Yee-Hong Yang. "Tri-focal tensor-based multiple video synchronization with subframeoptimization." IEEE Transactions on Image Processing 15.9 , pp. 2473-2480,2006.

[5] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in Proceedings of the European Conference on Computer Vision, pp. 404 -417, 2006