



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2017년02월23일
(11) 등록번호 10-1709085
(24) 등록일자 2017년02월16일

(51) 국제특허분류(Int. Cl.)
G06F 17/30 (2006.01) G06N 3/02 (2006.01)
G06N 3/08 (2006.01)
(52) CPC특허분류
G06F 17/30846 (2013.01)
G06F 17/30705 (2013.01)
(21) 출원번호 10-2015-0180362
(22) 출원일자 2015년12월16일
심사청구일자 2015년12월16일
(56) 선행기술조사문헌
KR1020020075956 A*
JP2002525735 A*
KR1020130108427 A
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
서강대학교산학협력단
서울특별시 마포구 백범로 35 (신수동, 서강대학교)
(72) 발명자
남종호
서울특별시 마포구 마포대로 195, 307동 1501호
(마포래미안 푸르지오)
최기석
경기도 고양시 일산서구 강선로 71 709동 601호
(주엽동, 강선마을7단지아파트)
홍호탁
경기도 용인시 수지구 정평로 41, 605동 1105호
(풍덕천동, 우성아파트)
(74) 대리인
이지연

전체 청구항 수 : 총 4 항

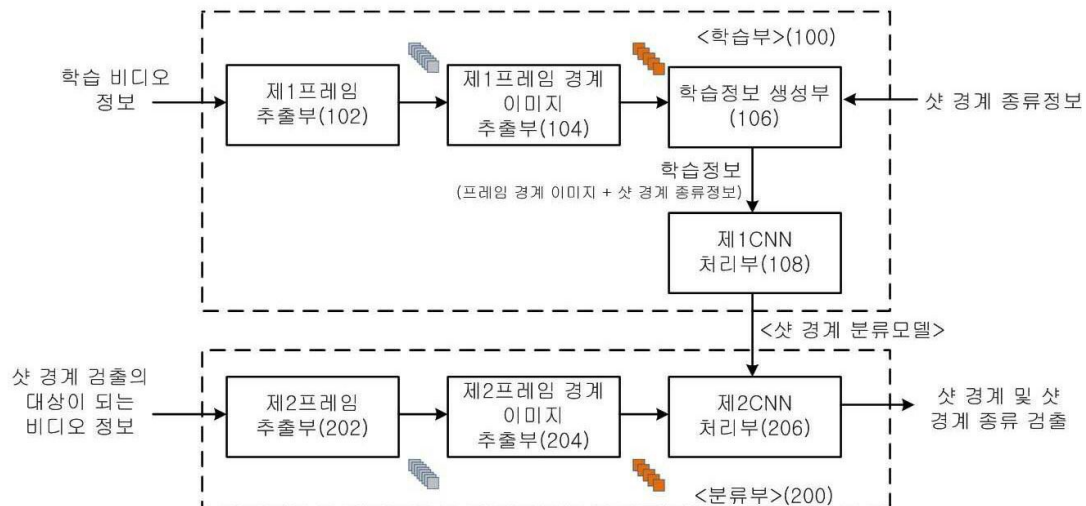
심사관 : 박승철

(54) 발명의 명칭 컨볼루션 신경망을 이용한 샷 경계 검출 방법 및 장치

(57) 요약

본 발명에 따르는 컨볼루션 신경망을 이용한 샷 경계 검출 장치는, 샷 경계 검출의 대상이 되는 비디오 정보를 입력받아 제1프레임들을 추출하는 제1프레임 추출부; 상기 제1프레임 추출부가 출력하는 제1프레임들 중 인접한 다수개의 제1프레임 각각에 대해 선택된 일부영역을 결합하여 제1프레임 경계 이미지를 생성하는 제1프레임 경계 이미지 추출부; 및 상기 제1프레임 경계 이미지를 제공받아, 상기 제1프레임 경계 이미지를 샷 경계 분류모델에 따라 분류하여 상기 제1프레임 경계 이미지가 샷 경계로 분류되는지 여부를 출력하는 제1컨볼루션 신경망 처리부;를 포함하며, 상기 샷 경계 분류모델은 샷 경계 종류별로 분류된 제2프레임 경계 이미지 정보를 포함하는 학습정보로부터 생성한 것임을 특징으로 한다.

대표도



(52) CPC특허분류

G06F 17/30858 (2013.01)

G06N 3/02 (2013.01)

G06N 3/08 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 R0126-15-1112

부처명 미래창조과학부

연구관리전문기관 정보통신기술진흥센터

연구사업명 정보통신·방송 기술개발 및 표준화 사업

연구과제명 퍼스널 미디어가 연결·공유·결합하여 재구성 가능케 하는 복합 모달리티 기반 미디어 응

용 프레임워크 개발

기 여 율 1/1

주관기관 한국전자통신연구원

연구기간 2015.03.01 ~ 2018.02.28

명세서

청구범위

청구항 1

컨볼루션 신경망을 이용한 샷 경계 검출 장치에 있어서,

샷 경계 검출의 대상이 되는 비디오 정보를 입력받아 제1프레임들을 추출하는 제1프레임 추출부;

상기 제1프레임 추출부가 출력하는 제1프레임들 중 인접한 다수개의 제1프레임 각각에 대해 선택된 일부영역을 결합하여 제1프레임 경계 이미지를 생성하는 제1프레임 경계 이미지 추출부;

상기 제1프레임 경계 이미지를 제공받아, 상기 제1프레임 경계 이미지를 샷 경계 분류모델에 따라 분류하여 상기 제1프레임 경계 이미지가 샷 경계로 분류되는지 여부를 출력하는 제1컨볼루션 신경망 처리부;

학습 비디오 정보를 입력받아 제2프레임들을 추출하는 제2프레임 추출부;

상기 제2프레임들 중 인접한 다수개의 프레임에 대해 선택된 일부영역을 결합하여 제2프레임 경계 이미지를 생성하는 제2프레임 경계 이미지 추출부;

상기 제2프레임 경계 이미지와 외부로부터 제공받은 샷 경계 종류정보를 결합하여 학습정보를 생성하여 출력하는 학습정보 생성부; 및

상기 학습정보를 제공받아, 그 학습정보를 토대로 샷 경계 종류별 제2프레임 경계 이미지를 학습하여 상기 샷 경계 분류모델을 생성하는 제2컨볼루션 신경망 처리부;를 포함하며,

상기 샷 경계 분류모델은 샷 경계 종류별로 분류된 제2프레임 경계 이미지 정보를 포함하는 학습정보로부터 생성한 것임을 특징으로 하는 컨볼루션 신경망을 이용한 샷 경계 검출 장치.

청구항 2

삭제

청구항 3

제1항에 있어서,

상기 제1 또는 제2프레임 경계 이미지는,

인접한 다수개의 프레임 각각을 다수개로 분할하고,

다수개의 프레임 각각에서 서로 다른 위치에 위치하는 분할영역들을 선택적으로 취하여 결합하여 하나의 이미지를 생성한 것임을 특징으로 하는 컨볼루션 신경망을 이용한 샷 경계 검출 장치.

청구항 4

컨볼루션 신경망을 이용한 샷 경계 검출 방법에 있어서,

샷 경계 검출의 대상이 되는 비디오 정보를 입력받아 제1프레임들을 추출하는 단계;

상기 제1프레임들 중 인접한 다수개의 제1프레임 각각에 대해 선택된 일부영역을 결합하여 제1프레임 경계 이미지를 생성하는 단계;

상기 제1프레임 경계 이미지를 제공받아, 상기 제1프레임 경계 이미지를 샷 경계 분류모델에 따라 분류하여 상기 제1프레임 경계 이미지가 샷 경계로 분류되는지 여부를 출력하는 단계;

학습 비디오 정보를 입력받아 제2프레임들을 추출하는 단계;

상기 제2프레임들 중 인접한 다수개의 제2프레임에 대해 선택된 일부영역을 결합하여 제2프레임 경계 이미지를 생성하는 단계;

상기 제2프레임 경계 이미지와 외부로부터 제공받은 샷 경계 종류정보를 결합하여 학습정보를 생성하여 출력하

는 단계; 및

상기 학습정보를 제공받아, 그 학습정보를 토대로 샷 경계 종류별 제2프레임 경계 이미지를 학습하여 상기 샷 경계 분류모델을 생성하는 단계;를 포함하며,

상기 샷 경계 분류모델은 샷 경계 종류별로 분류된 제2프레임 경계 이미지 정보를 포함하는 학습정보로부터 생성한 것임을 특징으로 하는 컨볼루션 신경망을 이용한 샷 경계 검출 방법.

청구항 5

삭제

청구항 6

제4항에 있어서,

상기 제1 또는 제2프레임 경계 이미지는,

인접한 다수개의 프레임 각각을 다수개로 분할하고,

다수개의 프레임 각각에서 서로 다른 위치에 위치하는 분할영역들을 선택적으로 취하여 결합하여 하나의 이미지를 생성한 것임을 특징으로 하는 컨볼루션 신경망을 이용한 샷 경계 검출 방법.

발명의 설명

기술 분야

- [0001] 본 발명은 영상처리기술에 관한 것으로, 더욱 상세하게는 학습 비디오 정보로부터 추출한 프레임들 중 인접한 다수개의 프레임을 결합하여 프레임 경계 이미지를 생성하고, 이 프레임 경계 이미지를 샷 경계 종류에 따라 분류하여 컨볼루션 신경망에 입력하여 샷 경계 분류 모델을 생성하고, 그 샷 경계 분류 모델을 적용한 컨볼루션 신경망을 통해 샷 분류 대상 비디오 정보로부터 생성한 프레임 경계 이미지가 어느 샷 경계 종류로 분류되는지를 토대로 샷 경계를 검출하는 컨볼루션 신경망을 이용한 샷 경계 검출 방법 및 장치에 관한 것이다.

배경 기술

- [0002] 인터넷과 디지털 영상 기술의 발달로 디지털 동영상 저작물이 점점 많아지고 있다. 이렇게 많은 데이터를 구조적으로 기록하고 이후 임의의 사용자의 검색을 가능하기 위해선 단순 텍스트 검색보단 이미지의 특성과 영상의 정보를 이용한 검색이 더 효율적이다. 이러한 비디오 검색을 위해 비디오 파싱(parsing) 기능이 요구되었다. 상기 비디오 파싱은 비디오 분할과 비디오 색인으로 이루어져 있다. 상기 비디오 분할단계에서 비디오 스트림은 비디오 기본 단위인 샷(shot)과 장면(scene)으로 분할된다. 상기 샷이란 영상 제작자의 편집을 거치지 않은 연속적인 카메라 이동으로 얻어진 가장 작은 단위의 비디오 데이터이다. 상기 장면(scene)은 같은 의미를 가지는 연속된 여러 개의 샷이 모여서 이루어진 의미적인 데이터이다. 상기 샷은 컷(cut), 디졸브(dissolve), 페이드(fade), 와이프(wipe) 등과 같이 여러 가지 변화로 연결되어 있다. 이렇게 변화되는 샷의 경계를 찾는 것은 비디오 내용 기반 검색을 위하여 전체 비디오 데이터를 효과적으로 구성하는데 가장 기본이 되는 핵심 기술이다.
- [0003] 현재까지 샷 경계 검출을 위하여 많은 연구가 진행되어 왔다. 많은 연구 결과들에서 급격한 장면 변환과 점진적 장면 변환을 위한 탐지모델을 각각 따로 정의하고 있으며, 이런 경우 특징 선택 및 변수 설정을 모델별로 수행하였다.
- [0004] Zhang et al.[H. J. Zhang, A. Kankanhalli, and S. W. Smoliar, "Automatic partitioning of full-motion video," ACM Multimedia Systems, 1: pp.10-28, 1993.]은 연속된 프레임들의 히스토그램 차를 이용하는 트윈 컴패리슨(twin-comparison) 방법을 이용하여 갑작스런 장면 변화와 점진적 장면 변화를 검출하였다. 이 방법은 비교적 좋은 결과를 나타내지만, 응용 대상에 따라 적절한 임계값을 설정할 수 있는지 여부에 따라 성능이 좌우되는 문제가 있었다.
- [0005] Alattar[A. M. Alattar, "Detecting and compressing dissolve regions in video sequences with DVI multimedia image compression algorithm," ISCAS, 13-16, 1993.] 그리고 Meng et al.[J. Meng, Y. Juan, S. F. Chang, "Scene change detection in a MPEG compressed video sequence," IS&T/SPIE Symposium, Proceedings, vol. 2419, Feb. 1995.]은 점진적 장면 변화의 하나인 디졸브 검출을 위해 동영상의 각 프레임의

분산으로부터 만들어지는 분산 곡선의 특징을 이용하였다. 상기 분산 곡선은 디졸브 구간에서 아래로 볼록한 포물선 모양을 나타나게 되는데, Alattar는 디졸브 구간의 시작과 끝에서 분산 곡선의 2차 미분값이 음의 최소값으로 나타남을 증명하고 이를 음의 최소값이 일정한 임계값을 넘는 구간에 대해 그 구간의 평균이 또 다른 임계값보다 크고 그 구간의 길이가 일정한 길이보다 큰 구간을 디졸브 구간으로 정의하였다. 또한 Meng은 분산 곡선의 1차 미분에서 인접한 최대값과 최소값의 차이를 기준으로 디졸브 구간을 찾았다. 그러나 이들의 방법은 1, 2차 미분에서 최대값과 최소값 또는 다른 조건을 만족할 수 있도록 임계값을 설정해야 하는데, 대상 동영상에 따라 사용자의 경험에 의존하여 적절한 임계값을 설정할 수 있는가에 따라 성능이 좌우되는 문제가 있었다. 즉 사용자에게 의한 임계값 설정에 따라 샷 경계 검출의 성능이 달라지고, 샷 경계 종류에 따라 구성된 탐지 모델들을 각각 고려해서 임계값을 적절하게 설정해주어야 한다는 문제점이 있었다.

[0006] 이와같이 샷 경계를 검출하여 샷들을 분할하는 작업은 효과적인 비디오 검색 시스템을 구축하기 위한 기본적인 작업으로, 효과적인 비디오 분석을 위해서는 신뢰할만한 샷 경계 검출 기술이 요구되었다.

[0007] 이러한 요구에 부응하여 많은 종류의 샷 경계 검출 기법들이 제안되었으며, 종래의 기법들은 샷 경계 종류에 따라 각각의 탐지 모델을 만들어 포괄적인 샷 경계 검출이 어려웠고, 사용자가 설정한 임계값을 이용하여 샷 경계를 검출하므로 사용자의 경험과 소양에 따라 성능이 좌우되는 문제가 있었다.

선행기술문헌

특허문헌

[0008] (특허문헌 0001) 한국특허공개 제1020050041761호
(특허문헌 0002) 한국특허공개 제1020050089498호

발명의 내용

해결하려는 과제

[0009] 본 발명은 학습 비디오 정보로부터 추출한 프레임들 중 인접한 다수개의 프레임을 결합하여 프레임 경계 이미지를 생성하고, 이 프레임 경계 이미지를 샷 경계 종류에 따라 분류하여 컨볼루션 신경망에 입력하여 샷 경계 분류 모델을 생성하고, 그 샷 경계 분류 모델을 적용한 컨볼루션 신경망을 통해 샷 분류 대상 비디오 정보로부터 생성한 프레임 경계 이미지가 어느 샷 경계 종류로 분류되는지를 토대로 샷 경계를 검출하는 컨볼루션 신경망을 이용한 샷 경계 검출 방법 및 장치를 제공하는 것을 그 목적으로 한다.

[0010]

과제의 해결 수단

[0011] 상기한 목적을 달성하기 위한 본 발명에 따르는 컨볼루션 신경망을 이용한 샷 경계 검출 장치는, 샷 경계 검출의 대상이 되는 비디오 정보를 입력받아 제1프레임들을 추출하는 제1프레임 추출부; 상기 제1프레임 추출부가 출력하는 제1프레임들 중 인접한 다수개의 제1프레임 각각에 대해 선택된 일부영역을 결합하여 제1프레임 경계 이미지를 생성하는 제1프레임 경계 이미지 추출부; 및 상기 제1프레임 경계 이미지를 제공받아, 상기 제1프레임 경계 이미지를 샷 경계 분류모델에 따라 분류하여 상기 제1프레임 경계 이미지가 샷 경계로 분류되는지 여부를 출력하는 제1컨볼루션 신경망 처리부;를 포함하며, 상기 샷 경계 분류모델은 샷 경계 종류별로 분류된 제2프레임 경계 이미지 정보를 포함하는 학습정보로부터 생성한 것임을 특징으로 한다.

발명의 효과

[0012] 상기한 본 발명은 학습 비디오 정보로부터 추출한 프레임들 중 인접한 다수개의 프레임을 결합하여 프레임 경계 이미지를 생성하고, 이 프레임 경계 이미지를 샷 경계 종류에 따라 분류하여 컨볼루션 신경망에 입력하여 샷 경계 분류 모델을 생성하고, 그 샷 경계 분류 모델을 적용한 컨볼루션 신경망을 통해 샷 분류 대상 비디오 정보로부터 생성한 프레임 경계 이미지가 어느 샷 경계 종류로 분류되는지를 토대로 샷 경계를 검출함으로써, 샷 경계 검출의 성능을 일정하게 유지할 수 있는 효과를 야기한다.

도면의 간단한 설명

- [0013] 도 1은 본 발명의 바람직한 실시예에 따른 프레임 경계 이미지 생성과정을 예시한 도면.
 도 2는 본 발명의 바람직한 실시예에 따른 프레임 경계 이미지들을 샷 경계 종류에 따라 구분하여 예시한 도면.
 도 3은 본 발명의 바람직한 실시예에 따른 샷 경계 종류에 따른 프레임 경계 이미지들의 특성을 정리한 표를 도시한 도면.
 도 4는 본 발명의 바람직한 실시예에 따른 컨볼루션 신경망을 이용한 샷 경계 검출 장치의 구성도.
 도 5는 본 발명의 성능실험을 위한 훈련 비디오 정보의 구성을 정리한 표를 도시한 도면.
 도 6은 본 발명의 성능실험 결과를 정리한 표를 도시한 도면.

발명을 실시하기 위한 구체적인 내용

- [0014] 본 발명에 따르는 컨볼루션 신경망(Convolutional Neural Networks, 이하, CNN이라 칭함)을 이용한 샷 경계 검출은, 학습 비디오 정보로부터 추출한 프레임들 중 인접한 다수개의 프레임을 결합하여 프레임 경계 이미지를 생성하고, 이 프레임 경계 이미지를 샷 경계 종류에 따라 분류하여 CNN에 입력하여 샷 경계 분류 모델을 생성하고, 그 샷 경계 분류 모델을 적용한 CNN을 통해 샷 분류 대상 비디오 정보로부터 생성한 프레임 경계 이미지가 어느 샷 경계 종류로 분류되는지를 토대로 샷 경계를 검출함으로써, 샷 경계 검출의 성능을 일정하게 유지할 수 있다.
- [0015] 특히 상기 CNN으로는 GoogLeNet을 채용할 수 있으며, 이는 ILSVRC 2014에서 우승한 컨볼루션 신경망으로 클래스 분류와 탐색에 뛰어난 성능을 보인다. 이에 본 발명에서는 샷 경계별로 인접한 다수개의 프레임을 결합하여 [프레임 경계 이미지]를 생성하고, 그 프레임 경계 이미지를 상기 GoogLeNet의 입력으로 주어 샷 경계별 프레임 경계 이미지의 특성을 학습하여 샷 경계의 종류를 분류해낼 수 있는 샷 경계 분류 모델을 생성하고, 그 샷 경계 분류 모델을 통해 샷 분류 대상 비디오 정보로부터 생성한 프레임 경계 이미지가 어느 샷 경계의 특성에 부합되는지를 토대로 샷 경계의 종류로 분류되게 함으로써, 샷 경계 여부를 검출한다. 이는 상기 프레임 경계 이미지에는 샷 경계 각각에 따른 특성을 포함하고 있을 것으로 기대하여, 이 특성에 따라 CNN이 분류 작업을 수행하게 하는 것이다. 이와 같이 본 발명은 비디오 정보의 샷 경계를 검출하는 것을 CNN을 이용한 분류 문제로 처리한다.
- [0016] 상기 [프레임 경계 이미지]는 N개 크기의 윈도우에 포함된 인접한 N개의 프레임에 대해 각기 다르게 선택된 일부영역을 결합하여 N개의 프레임을 한 장의 이미지로 나타낸 것으로, 샷 경계의 포함여부와 샷 경계의 종류별로 상이한 특성을 가진다.
- [0017] 이러한 [프레임 경계 이미지]의 생성과정을 도 1을 참조하여 설명한다. 상기 도 1의 (a)에 예시한 프레임 경계 이미지는 인접한 두 프레임, 즉 A 및 B 프레임 각각을 반씩 균등분할하고, A 프레임의 좌측 절반 부분인 A' 영역과 B 프레임의 우측 절반 부분인 B' 영역을 결합하여 형성한 것이다. 그리고 도 1의 (b)에 예시한 프레임 경계 이미지는 인접한 네 프레임인 A 내지 D 프레임 각각을 4개로 균등분할하고, A 프레임의 1/4 부분인 A⁰ 영역과 B 프레임의 2/4 부분인 B¹ 영역과 C 프레임의 3/4 부분인 C² 영역과 D 프레임의 4/4 부분인 D³ 영역을 결합하여 형성한 것이다. 상기한 예에서는 네 프레임의 부분들을 순서대로 조합하는 것만을 예시하였으나, 인접된 프레임의 수 및 조합 순서는 가변될 수 있으며, 이는 본 발명에 의해 당업자에게 자명하다.
- [0018] 상기한 바와 같이 형성된 [프레임 경계 이미지]는 샷 경계에 따라 상이한 특성을 가지며, 이를 예시한 것이 도 2이다. 여기서, 상기 도 2는 편이상 인접한 두 프레임을 결합하여 프레임 경계 이미지를 생성한 예를 도시한 것이다. 상기 도 2의 (a)는 샷 경계가 없는 경우(non)의 인접한 두 프레임 사이의 프레임 경계 이미지를 나타내고, (b)는 샷 경계 중 컷(cut)에서 만들어진 프레임 경계 이미지이고, (c)는 샷 경계 중 디졸브(dissolve)에서 만들어진 프레임 경계 이미지이고, (d)는 샷 경계 중 페이드(fade)에서 만들어진 프레임 경계 이미지이다.
- [0019] 이러한 프레임 경계 이미지들을 참조하면, 샷 경계에서 만들어진 프레임 경계 이미지 (b),(c),(d)는 샷 경계가 없는 일반 프레임으로부터 만들어진 프레임 경계 이미지 (a)와는 구별되는 특성이 있다. 즉, (b)는 장면이 급진적으로 변화하여 중앙선을 기준으로 서로 다른 장면이 합쳐짐에 따라 명확한 경계선이 형성되고, (c)는 디졸브의 특징이 반영되어 서로 다른 장면이 겹쳐서 보이게 나타나고, (d)는 페이드의 특징이 반영되어 중앙선을 기준

으로 명확한 밝기 차이가 나타난다.

- [0020] 이러한 프레임 경계 이미지의 특성을 정리한 표를 도시한 것이 도 3이다. 상기 도 3을 참조하면, 넌(non)의 프레임 경계 이미지는 샷 경계가 없는 여러 프레임으로부터 획득된 것으로 밝기값의 차이가 없고 겹침이 없고 경계선이 없는 특성을 가지고, 페이드(fade)의 프레임 경계 이미지는 어두워지면서 샷이 변경되는 두 프레임으로부터 획득된 것으로 명확한 밝기 차이가 있는 특성을 가지고, 디졸브(dissolve)의 프레임 경계 이미지는 여러 프레임의 겹쳐지면서 샷이 변경되는 여러 프레임으로부터 획득된 것으로 명확한 겹침이 있는 특성을 가지고, 컷(cut)의 프레임 경계 이미지는 프레임과 프레임이 급격하게 변경되므로 중앙부분에 경계선이 있는 특성을 가진다.
- [0021] 이에 본 발명은 상기한 바와 같이 프레임 경계 이미지에는 샷 경계의 종류에 따라 각기 상이한 특성을 포함하고 있으므로 CNN을 통해 상기 특성을 바탕으로 프레임 경계 이미지의 샷 경계의 종류를 분류하도록 함으로써 샷 경계 여부를 판별한다.
- [0022] <컨볼루션 신경망을 이용한 샷 경계 검출 장치의 구성>
- [0023] 이러한 본 발명의 바람직한 실시예에 따른 컨볼루션 신경망을 이용한 샷 경계 검출 장치의 구성을 도 4를 참조하여 설명한다.
- [0024] 상기 컨볼루션 신경망을 이용한 샷 경계 검출 장치는 크게 학습부(100)와 분류부(200)로 구성된다. 상기 학습부(100)는 다수개의 프레임을 결합하여 프레임 경계 이미지를 생성하고, 이 프레임 경계 이미지와 샷 경계 종류정보를 결합하여 학습정보를 생성하고, 그 학습정보를 컨볼루션 신경망에 통해 학습하여 샷 경계 분류모델을 생성하며, 이는 제1프레임 추출부(102)와 제1프레임 경계 이미지 추출부(104)와 학습정보 생성부(106)와 제1CNN 처리부(108)로 구성된다. 그리고 상기 분류부(200)는 샷 경계 검출의 대상이 되는 비디오 정보를 입력받아 프레임 경계 이미지를 생성하고, 이를 상기 샷 경계 분류모델을 토대로 샷 경계 종류별로 분류함으로써 샷 경계 여부를 판별하며, 이는 제2프레임 추출부(202)와 제2프레임 경계 이미지 추출부(204)와 제2CNN 처리부(206)로 구성된다. 특히 상기 프레임 경계 이미지는 인접된 다수개의 프레임 각각에 대해 일부를 각기 다르게 선택하여 결합하여 생성한 것으로, 이는 샷 경계 종류에 따라 상이한 특성을 가지며, 이 특성을 CNN이 학습하여 샷 경계의 종류를 분류할 수 있게 된다.
- [0025]
- [0026] <학습부>
- [0027] 상기 학습부(100)의 구성 및 동작을 좀더 상세히 설명한다.
- [0028] 상기 학습부(100)는 지도 학습(Supervised-learning) 방법을 따르고 있기 때문에 학습에 필요한 샷 경계 정보는 사전에 사람으로부터 제공받는다.
- [0029] 상기 학습부(100)의 제1프레임 추출부(102)는 학습 비디오 정보를 입력받아 프레임들을 추출하여 제1프레임 경계 이미지 추출부(104)로 제공한다. 상기 제1프레임 경계 이미지 추출부(104)는 도 1에 도시한 바와 같이 제1프레임 추출부(102)가 순차적으로 입력되는 프레임들 중 인접된 다수개의 프레임 각각에 대해 일부를 선택적으로 취한 후에 하나의 이미지로 결합하여 프레임 경계 이미지를 생성한다.
- [0030] 상기 제1프레임 경계 이미지 추출부(104)는 상기 프레임 경계 이미지를 학습정보 생성부(106)로 제공한다.
- [0031] 상기 학습정보 생성부(106)는 제1프레임 경계 이미지 추출부(104)가 제공하는 프레임 경계 이미지에 대한 샷 경계 종류 정보를 외부로부터 입력받아, 샷 경계 분류 라벨정보가 결합된 샷 경계 이미지 형태의 학습정보를 생성하여 제1CNN 처리부(108)에 제공한다.
- [0032] 상기 제1CNN 처리부(108)는 상기 학습정보를 제공받아 컨볼루션 신경망으로 학습하여, 프레임 경계 이미지에 대해 샷 경계 종류별로 분류하는 샷 경계 분류 모델을 생성하여 상기 분류부(200)에 제공한다.
- [0033] 상기 분류부(200)의 제2프레임 추출부(202)는 샷 경계 검출의 대상이 되는 비디오 정보를 입력받아 프레임들을 추출하고, 이를 제2프레임 경계 이미지 추출부(204)로 제공한다.
- [0034] 상기 제2프레임 경계 이미지 추출부(204)는 상기 제2프레임 추출부(202)가 제공하는 프레임들을 제공받아, 인접된 다수개의 프레임 각각에 대해 일부를 선택적으로 결합하여 프레임 경계 이미지를 생성한다. 여기서, 이러한 프레임 경계 이미지를 생성하기 위한 프레임 분할 및 선택은 제1프레임 경계 이미지 추출부(104)와 동일한 방식

으로 결정된다.

- [0035] 상기 제2프레임 경계 이미지 추출부(204)는 상기 인접된 다수개의 프레임 각각에 대해 일부를 선택적으로 결합한 프레임 경계 이미지를 제2CNN 처리부(206)로 제공한다.
- [0036] 상기 제2CNN 처리부(206)는 상기 학습부(100)가 제공한 샷 경계 분류모델에 따라 상기 프레임 경계 이미지를 샷 경계의 종류로 분류하여, 상기 프레임 경계 이미지가 미리 정해진 샷 경계의 종류, 즉 컷, 디졸브, 페이드의 샷 경계에 해당되는 프레임 경계 이미지로 분류되는지 여부를 토대로 샷 경계 여부를 판별하고 그 결과를 출력한다.
- [0037] 이러한 본 발명에 따르는 컨볼루션 신경망을 이용한 샷 경계 검출 장치가 수행하는 기능은 소프트웨어의 형태로 구성될 수도 있으며, 이는 본 발명에 의해 당업자에게 자명하다. 또한 소프트웨어의 형태로 구성된 컨볼루션 신경망을 이용한 샷 경계 검출 프로그램은 소정의 프로그램 데이터가 저장된 기억장치(ROM, RAM, 하드디스크 등)와, 상기 프로그램 데이터를 실행하는 CPU에 의해 실현된다. 이 경우, 각 프로그램 데이터는 CD-ROM과 USB 메모리 등의 기록매체에 기록되어 제공될 수 있다.
- [0038] <성능 평가>
- [0039] 이러한 본 발명에 따르는 컨볼루션 신경망을 이용한 샷 경계 검출은, 학습과 분류로 이루어진다. 상기 학습 단계에서는 비디오 정보에서 추출한 프레임들 중 인접한 다수개의 프레임의 결합을 통해 프레임 경계 이미지를 얻는다. 이 프레임 경계 이미지 중 샷 경계를 나타내는 이미지들을 골라 컨볼루션 신경망에 입력하여 샷 경계를 학습시킨다. 그리고 상기 분류 단계에서는 학습을 통해 구성된 모델을 가지고 새로운 비디오 정보의 프레임 경계 이미지를 분류하여 샷 경계를 검출한다. 이러한 본 발명에 따르는 샷 경계 검출의 성능을 평가하기 위해 정확률(precision)과 재현율(recall)을 평가 지표로 사용한다.
- [0040] 본 발명의 성능 실험에는 KBS <생로병사의 비밀> 13개의 동영상을 사용하였다. 먼저 학습에 10개의 동영상상을 사용하고, 분류에 3개의 동영상상을 사용하였다. 상기 <생로병사의 비밀>은 컷, 디졸브, 페이드를 사용하여 샷 전환을 한다. 비디오 길이는 평균 48분이고 해상도는 480x360이다. 본 발명에서는 이미지의 중심을 기준으로 256x256으로 이미지를 잘라서 사용하였다.
- [0041] 먼저 학습단계에서 10개의 동영상에서 총 22818개의 샷 경계를 나타내는 프레임 경계 이미지를 추출했다. 여기서, 25%를 유효(validation)로 하여 훈련(train) 이미지는 17115개, 유효(validation) 이미지는 5703개로 구성했다. 즉, 학습에 사용된 경계 이미지 개수를 정리한 도 5의 표를 참조하면, 훈련시에는 넌(non)이 4430개, 컷(cut)이 4237개, 디졸브(dissolve)가 4150개, 페이드(fade)가 4298개로 총 17115개가 사용되고, 그 중 유효 이미지는 넌(non)이 1476개, 컷(cut)이 1412개, 디졸브(dissolve)가 1383개, 페이드(fade)가 1432개로 총 5703개가 사용된다.
- [0042] 이러한 학습단계를 거쳐 구성된 모델을 이용하여 분류단계에서는 새로운 프레임 경계 이미지를 모델의 입력으로 하여 이미지를 분류함으로써 샷 경계를 검출한다.
- [0043] 도 6의 표 2는 본 발명에 따르는 샷 경계를 검출한 결과를 나타낸 것이다. 상기 표 2에서 C는 컷 경계 이미지이고, D는 디졸브 경계 이미지, F는 페이드 경계 이미지이다. N_r 은 실제 경계 이미지 개수, N_d 은 검색된 경계 이미지 개수, N_s 은 잘 인식된 경계 이미지 개수, N_f 은 오 인식된 경계 이미지 개수, N_m 은 놓친 경계 이미지 개수를 나타낸다. 정확률(P.)은 N_s/N_d 로 나타내고 재현율(R.)은 N_s/N_r 로 나타낸다.
- [0044] 상기한 성능실험의 결과는 페이드 검출에서 페이드를 놓치는 경우가 많아 재현율이 낮게 나왔다. 그러나 정확률은 100%로 좋은 결과를 보여준다. 그리고 컷과 디졸브도 정확률이 100%, 97%, 재현율이 97%, 96%로 좋은 결과를 내고 있다.
- [0045] 이러한 본 발명의 컨볼루션 신경망을 이용한 샷 경계 검출은 샷 경계를 사용자의 설정없이 한 모델로 포괄적으로 검출할 수 있다. 특히 샷 경계에 대한 프레임 경계 이미지로 기계를 학습시켜 모델을 구성하고, 비디오 정보에서 추출한 프레임들에서 프레임 경계 이미지를 만들어 학습단계에서 구성한 모델을 통해 경계 이미지를 분류함으로써 샷 경계를 검출한다.
- [0046] 이와 같이 기계가 학습을 통해 스스로 모델을 구성하고 임계값을 설정하여 기존에 사용자가 모델을 구성하고,

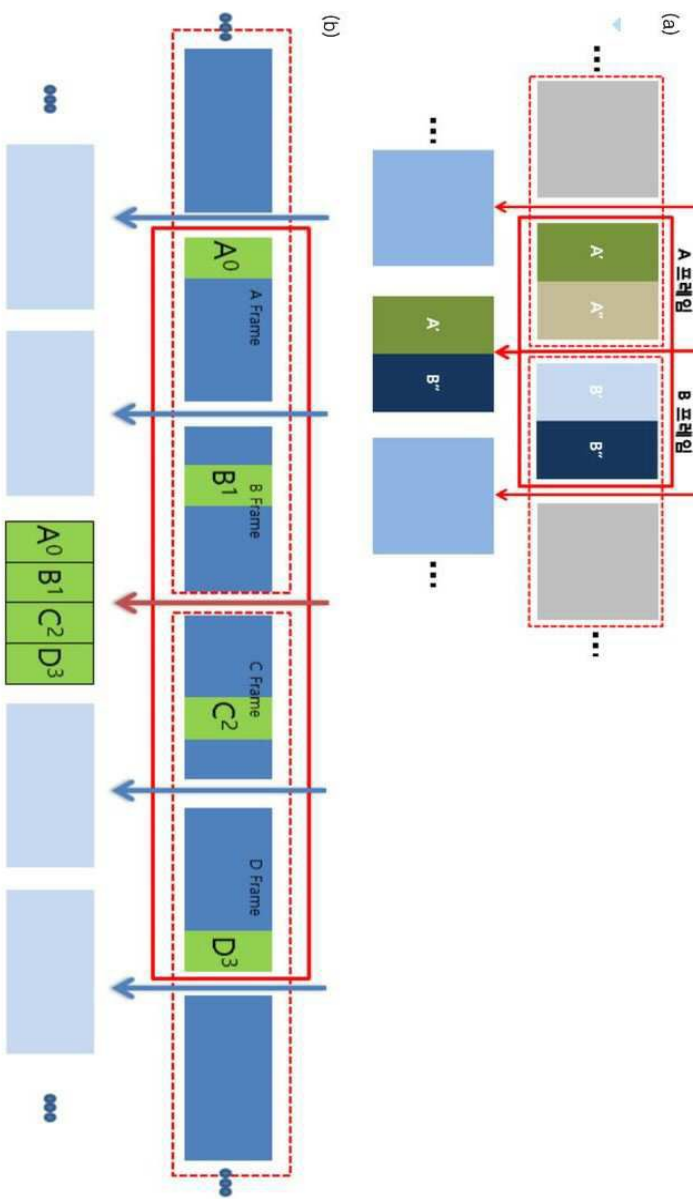
임계값을 설정했던 방법과 달리 샷 경계 검출의 성능을 일정하게 유지시킬 수 있다.

부호의 설명

[0047]

- 100 : 학습부
- 102 : 제1프레임 추출부
- 104 : 제1프레임 경계 추출부
- 106 : 학습정보 생성부
- 108 : 제1CNN 처리부
- 200 : 분류부
- 202 : 제1프레임 추출부
- 204 : 제2프레임 경계 이미지 추출부
- 206 : 제2CNN 처리부

도면
도면1



도면2



(a) 일반 프레임



(b) 컷



(c) 디졸브

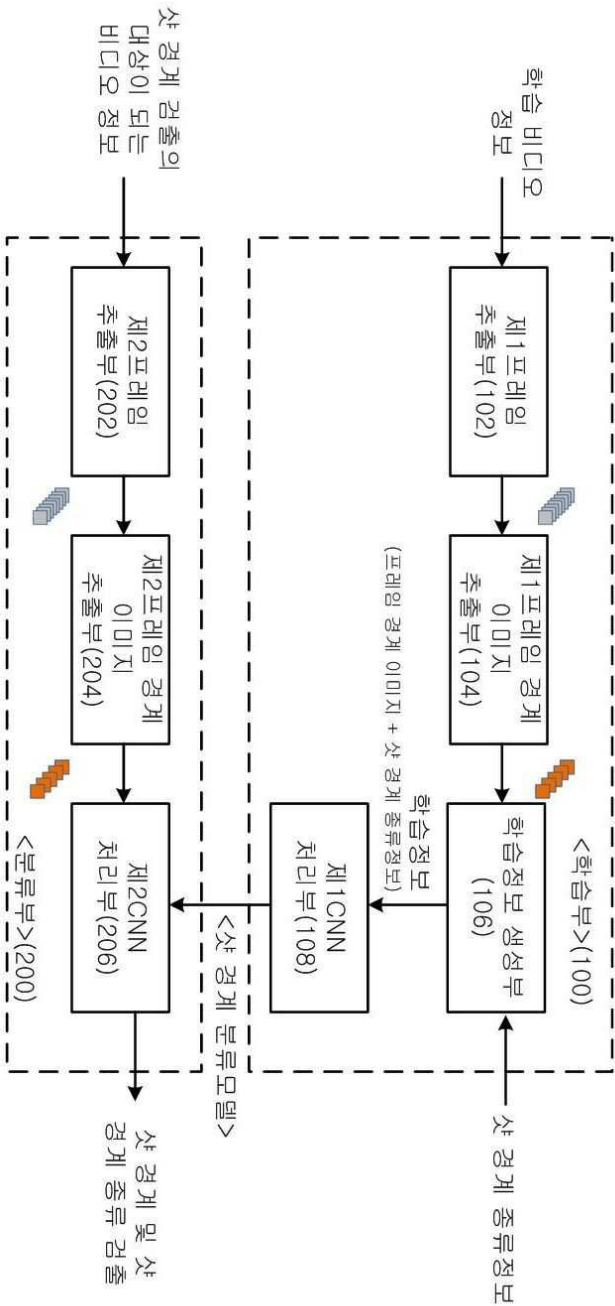


(d) 페이드

도면3

샷 경계의 종류	특성	예시
non	없음	
cut	명확한 경계선	
fade	명확한 밝기 차이	
dissolve	명확한 겹침	

도면4



도면5

	train	validation
non	4430	1476
cut	4237	1412
dissolve	4150	1383
fade	4298	1432
합계	17115	5703

도면6

계통	N _r			N _f			N _s			N _j			N _m			P.(%)			R.(%)		
	C	D	F	C	D	F	C	D	F	C	D	F	C	D	F	C	D	F	C	D	F
27384	195	279	217	194	286	146	194	274	146	0	12	0	1	5	59	100	96	100	100	98	67
27484	120	388	51	115	352	34	115	346	34	0	6	0	5	22	11	100	98	100	96	94	67
27584	125	274	286	117	289	150	117	261	150	0	8	0	8	13	49	100	97	100	94	95	73
평균																100	97	100	97	96	69