(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2011-0089935
(43) 공개일자 2011년08월10일

(51) Int. Cl.
H04N 5/93 (2006.01) *H04N 7/08* (2006.01)
(21) 출원번호 10-2010-0009451
(22) 출원일자 2010년02월02일
심사청구일자 2010년02월02일

(71) 출원인
인하대학교 산학협력단
인천 남구 용현동 253 인하대학교
(72) 발명자
정동석
서울특별시 강남구 도곡동 467번지 삼성타워팰리스 D-507
조주희
경기도 시흥시 대야동 극동아파트 104동 404호
진주경
경기도 부천시 원미구 중1동 미리내마을 920동 1604호
(74) 대리인
이원희

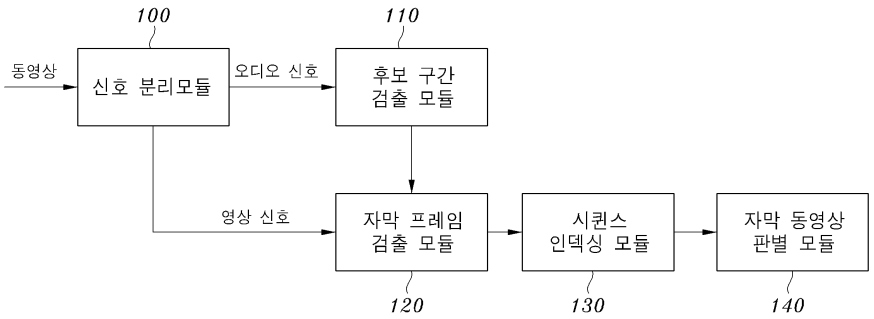
전체 청구항 수 : 총 12 항

(54) 자막이 입혀진 동영상을 검출하기 위한 장치 및 방법

(57) 요 약

본 발명은 동영상을 오디오 신호와 영상 신호로 분리하는 신호 분리 모듈, 상기 분리된 오디오 신호를 오디오 프레임으로 분할하고, 각 오디오 프레임의 음성 에너지를 구하여 자막 검출을 위한 후보 구간을 결정하는 후보 구간 검출 모듈, 상기 결정된 후보 구간에 해당하는 영상 프레임을 디코딩하여 에지 성분을 구하고, 라인 스캐닝을 통해 각 라인별 에지 성분 총량을 구하여 자막이 입혀진 영상 프레임을 판별하는 자막 프레임 검출 모듈, 상기 자막 프레임 검출 모듈에서의 판별 결과에 따라 영상 프레임별 자막 유무를 기록하여 자막 시퀀스 바이너리 맵을 생성하는 시퀀스 인덱싱 모듈로 구성되어, 대용량 데이터베이스로부터 자막이 입혀진 동영상을 자동적으로 선별하므로, 콘텐츠 공급자의 편의성과 사용자의 만족도를 높일 수 있다.

대 표 도 - 도1



특허청구의 범위

청구항 1

동영상을 오디오 신호와 영상 신호로 분리하는 신호 분리 모듈;

상기 분리된 오디오 신호를 오디오 프레임으로 분할하고, 각 오디오 프레임의 음성 에너지를 구하여 자막 검출을 위한 후보 구간을 결정하는 후보 구간 검출 모듈;

상기 결정된 후보 구간에 해당하는 영상 프레임을 디코딩하여 에지 성분을 구하고, 라인 스캐닝을 수행하여 자막이 입혀진 영상 프레임을 판별하는 자막 프레임 검출 모듈;및

상기 자막 프레임 검출 모듈에서의 판별 결과에 따라 영상 프레임별 자막 유무를 기록하여 자막 시퀀스 바이너리 맵을 생성하는 시퀀스 인덱싱 모듈;

을 포함하는 자막이 입혀진 동영상을 검출하기 위한 장치.

청구항 2

제1항에 있어서,

상기 생성된 자막 시퀀스 바이너리 맵에서 자막 존재 표시의 밀도와 분포를 이용하여 자막이 입혀진 동영상을 판별하는 자막 동영상 판별 모듈;을 더 포함하는 자막이 입혀진 동영상을 검출하기 위한 장치.

청구항 3

제1항에 있어서,

상기 후보 구간 검출 모듈은,

상기 오디오 신호를 일정 간격으로 분할하여 복수의 오디오 프레임을 생성하는 오디오 신호 분할부;

상기 생성된 각 오디오 프레임의 음성 에너지를 구하고, 일정한 윈도우 사이즈내에 있는 오디오 프레임의 평균 에너지를 구하는 음성 에너지 계산부;

상기 구해진 평균 에너지를 평균과 분산을 이용하여 정규화하는 정규화부;및

상기 정규화된 평균 에너지가 미리 정해진 제1 임계치 이상인 오디오 프레임의 개수를 확인하여, 그 개수가 일정 개수 이상인 경우 해당 구간을 후보 구간으로 결정하는 후보 구간 결정부;를 포함하는 자막이 입혀진 동영상을 검출하기 위한 장치.

청구항 4

제3항에 있어서,

상기 음성 에너지 계산부는 $E_m = \frac{1}{N_a} log \sum_{n=1}^{N_a} |x(n)|$ 를 이용하여 음성 에너지(Em)를 구하되, N_a는 윈도우 사이즈, x(n)은 n번째 오디오 샘플인 것을 특징으로 하는 자막이 입혀진 동영상을 검출하기 위한 장치.

청구항 5

제3항에 있어서,

상기 윈도우 사이즈는 오디오 프레임의 샘플 수보다 크게 두어 연속하는 두 개의 오디오 프레임이 중첩되도록

구성하는 것을 특징으로 하는 자막이 입혀진 동영상을 검출하기 위한 장치.

청구항 6

제3항에 있어서,

상기 정규화부는 상기 음성 에너지, 초당 오디오 프레임의 개수 및 오디오 신호의 초단위 길이를 이용하여 평균과 분산을 구하고, 상기 구해진 평균과 분산을 이용하여 정규화하는 것을 특징으로 하는 자막이 입혀진 동영상을 검출하기 위한 장치.

청구항 7

제1항에 있어서,

상기 자막 프레임 검출 모듈은,

상기 후보 구간 검출 모듈에서 결정된 후보 구간에 해당하는 영상 프레임을 디코딩하는 디코딩부;

상기 디코딩된 영상 프레임을 흑백으로 변환하며 그 사이즈를 정규화하는 영상 프레임 정규화부;

상기 정규화된 영상 프레임에 대해 현재 픽셀과 인접 픽셀과의 차를 이용하여 에지 성분을 구하는 에지 성분 결정부;

상기 영상 프레임에 대해 가로축 방향으로 라인 스캐닝을 수행하여 각 라인별 에지 성분 총량을 구하는 라인 스캐닝부;및

상기 영상 프레임을 세로축 방향으로 탐색하여 에지 성분 총량이 제2 임계치를 초과하는 라인의 개수가 일정 개수 이상이면, 자막이 입혀진 영상 프레임으로 판별하는 자막 프레임 판별부;를 포함하는 자막이 입혀진 동영상을 검출하기 위한 장치.

청구항 8

제7항에 있어서,

상기 라인 스캐닝부는 각 라인에 있는 에지 성분에 현재 픽셀과 인접 픽셀간 거리에 따른 가중치를 적용하여 에지 성분 총량을 구하는 것을 특징으로 하는 자막이 입혀진 동영상을 검출하기 위한 장치.

청구항 9

(a)동영상을 오디오 신호와 영상 신호로 분리하는 단계;

(b)상기 분리된 오디오 신호를 오디오 프레임으로 분할하고, 각 오디오 프레임의 음성 에너지를 구하여 자막 검출을 위한 후보 구간을 결정하는 단계;및

(c)상기 결정된 후보 구간에 해당하는 영상 프레임을 디코딩하여 에지 성분을 구하고, 라인 스캐닝을 통해 자막이 입혀진 영상 프레임을 판별하는 단계;

(d)상기 판별 결과에 따라 영상 프레임별 자막 유무를 기록하여 자막 시퀀스 바이너리 맵을 생성하는 단계;

를 포함하는 자막이 입혀진 동영상을 검출하기 위한 방법.

청구항 10

제9항에 있어서,

상기 생성된 자막 시퀀스 바이너리 맵에서 자막 존재 표시의 밀도와 분포를 이용하여 자막이 입혀진 동영상을

판별하는 단계;를 더 포함하는 자막이 입혀진 동영상을 검출하기 위한 방법.

청구항 11

제9항에 있어서,

상기 (b)단계는, 상기 오디오 신호를 일정 간격으로 분할하여 복수의 오디오 프레임을 생성하는 단계;

상기 생성된 각 오디오 프레임의 음성 에너지를 구하고, 일정한 윈도우 사이즈 내에 있는 오디오 프레임의 평균 에너지를 구하는 단계;

상기 구해진 평균 에너지를 평균과 분산을 이용하여 정규화하는 단계;및

상기 정규화된 평균 에너지가 제1 임계치 이상인 오디오 프레임의 개수를 확인하여, 그 개수가 일정 개수 이상인 경우 해당 구간을 후보 구간으로 결정하는 단계;를 포함하는 자막이 입혀진 동영상을 검출하기 위한 방법.

청구항 12

제9항에 있어서,

상기 (c)단계는, 상기 결정된 후보 구간에 해당하는 영상 프레임을 디코딩하는 단계;

상기 디코딩된 영상 프레임을 흑백으로 변환하며 그 사이즈를 정규화하는 단계;

상기 정규화된 영상 프레임에 대해 현재 픽셀과 인접 픽셀과의 차를 이용하여 에지 성분을 구하는 단계;

상기 영상 프레임에 대해 가로축 방향으로 라인 스캐닝을 수행하여 각 라인별 에지 성분 총량을 구하는 단계;및

상기 영상 프레임을 세로축 방향으로 탐색하여 에지 성분 총량이 제2 임계치를 초과하는 라인의 개수가 일정 개수 이상이면, 자막이 입혀진 영상 프레임으로 판별하는 단계;를 포함하는 자막이 입혀진 동영상을 검출하기 위한 방법.

명 세 서

기 술 분 야

[0001] 본 발명은 자막이 입혀진 동영상을 검출하기 위한 장치 및 방법에 관한 것으로, 더욱 상세하게는 동영상을 오디오 신호와 영상 신호로 분리한 후, 상기 오디오 신호의 음성 에너지를 구하여 자막 검출을 위한 후보 구간을 결정하고, 상기 결정된 후보 구간에 해당하는 영상 프레임을 디코딩하여 에지 성분을 구하고, 라인 스캐닝을 통해 자막이 입혀진 영상 프레임을 판별하고, 그 판별 결과에 따라 자막 시퀀스 바이너리 맵을 생성하여 자막이 입혀진 동영상을 판별하는 자막이 입혀진 동영상을 검출하기 위한 장치 및 방법에 관한 것이다.

배 경 기 술

[0002] 오늘날 전세계적으로 방대한 양의 동영상이 만들어지며 여러 사람들에게 공유되고 있다. 이에 발맞추어 동영상 콘텐츠 제공 서비스의 질 또한 하루가 다르게 발전하고 있으며, 개인 맞춤형 서비스에 대한 사용자의 욕구도 점차 커지고 있다.

[0003] 또한, 멀티미디어 콘텐츠를 찾는 사람들은 자신이 원하는 콘텐츠를 좀더 쉽고 빠르게 찾고 싶어한다. 이러한 사용자 욕구를 반영하기 위한 동영상 검색 서비스는 다양하게 제공되고 있다.

[0004] 동영상 콘텐츠를 공유하기 위한 공간적 제약성이 적어짐에 따라 웹 서비스를 통해 다른 나라에서 제작된 동영상들을 빠르게 접할 수 있게 되었다.

[0005] 사용자는 외화 시리즈 물 제목과 같은 키워드를 입력하면, 서비스 제공자로부터 관련 동영상들의 리스트와 동영상으로의 링크를 제공받게 된다. 이와 동시에 사용자는 보고자 선택한 에피소드의 자막을 같이 검색하거나 동영

상 자체에 관련 자막이 입혀진 에피소드를 검색 결과 리스트로부터 찾아낼 수 있다.

[0006] 그러나 자막이 입혀진 동영상인지 아닌지를 판별하기 위해서는 해당 동영상을 일일이 재생해야 하는 불편함이 있었다.

[0007] 따라서, 매일 새로운 동영상이 추가/삭제되는 대용량의 동영상 데이터베이스에서 자막이 입혀진 동영상을 자동으로 판별할 수 있는 기능이 요구되고 있는 실정이다.

발명의 내용

해결하려는 과제

[0008] 본 발명의 목적은 자막이 입혀진 동영상을 대용량 데이터베이스 내에서 구분해 낼 수 있는 자막이 입혀진 동영상을 검출하기 위한 장치 및 방법을 제공하는데 있다.

[0009] 본 발명의 다른 목적은 1차원 키워드 중심의 동영상 검색 방법을 통해 사용자 맞춤형 동영상 검색 서비스를 지원할 수 있도록 대용량 데이터베이스로부터 자막이 입혀진 동영상을 자동적으로 선별하는 자막이 입혀진 동영상을 검출하기 위한 장치 및 방법을 제공하는데 있다.

[0010] 또한, 사용자가 동영상을 검색하고자 키워드를 입력함과 동시에 동영상 리스트에 자막의 존재 여부가 표시되는 자막이 입혀진 동영상을 검출하기 위한 장치 및 방법을 제공하는데 있다.

과제의 해결 수단

[0011] 상기 목적들을 달성하기 위하여 본 발명에 따르면, 동영상을 오디오 신호와 영상 신호로 분리하는 신호 분리 모듈, 상기 분리된 오디오 신호를 오디오 프레임으로 분할하고, 각 오디오 프레임의 음성 에너지를 구하여 자막 검출을 위한 후보 구간을 결정하는 후보 구간 검출 모듈, 상기 결정된 후보 구간에 해당하는 영상 프레임을 디코딩하여 에지 성분을 구하고, 라인 스캐닝을 통해 각 라인별 에지 성분 총량을 구하여 자막이 입혀진 영상 프레임을 판별하는 자막 프레임 검출 모듈, 상기 자막 프레임 검출 모듈에서의 판별 결과에 따라 영상 프레임별 자막 유무를 기록하여 자막 시퀀스 바이너리 맵을 생성하는 시퀀스 인덱싱 모듈을 포함하는 자막이 입혀진 동영상을 검출하기 위한 장치가 제공된다.

[0012] 또한, 상기 생성된 자막 시퀀스 바이너리 맵에서 자막 존재 표시의 밀도와 분포를 이용하여 자막이 입혀진 동영상을 판별하는 자막 동영상 판별 모듈을 더 포함하여 구성할 수 있다.

[0013] 상기 후보 구간 검출 모듈은, 상기 오디오 신호를 일정 간격으로 분할하여 복수의 오디오 프레임을 생성하는 오디오 신호 분할부, 상기 생성된 각 오디오 프레임의 음성 에너지를 구하고, 일정한 윈도우 사이즈 내에 있는 오디오 프레임의 평균 에너지를 구하는 음성 에너지 계산부, 상기 구해진 평균 에너지를 평균과 분산을 이용하여 정규화하는 정규화부, 상기 정규화된 평균 에너지가 미리 정해진 제1 임계치 이상인 오디오 프레임의 개수를 확인하여, 그 개수가 일정 개수 이상인 경우 해당 구간을 후보 구간으로 결정하는 후보 구간 결정부를 포함한다.

[0014] 상기 음성 에너지 계산부는 $$E_m = \frac{1}{N_a} log \sum_{n=1}^{N_a} |x(n)|$$ 를 이용하여 음성 에너지(Em)를 구하되, N_a는 윈도우 사이즈, x(n)은 n번째 오디오 샘플을 말한다.

[0015] 상기 윈도우 사이즈는 오디오 프레임의 샘플 수보다 크게 두어 연속하는 두 개의 오디오 프레임이 중첩되도록 구성한다.

[0016] 상기 정규화부는 상기 음성 에너지, 초당 오디오 프레임의 개수, 오디오 신호의 초단위 길이를 이용하여 평균과 분산을 구하고, 상기 구해진 평균과 분산을 이용하여 정규화한다.

[0017] 상기 자막 프레임 검출 모듈은 상기 후보 구간 검출 모듈에서 결정된 후보 구간에 해당하는 영상 프레임을 디코딩하는 디코딩부, 상기 디코딩된 영상 프레임을 흑백으로 변환하며 그 사이즈를 정규화하는 영상 프레임 정규화부, 상기 정규화된 영상 프레임에 대해 현재 픽셀과 인접 픽셀과의 차를 이용하여 에지 성분을 구하는 에지 성

분 결정부, 상기 영상 프레임에 대해 가로축 방향으로 라인 스캐닝을 수행하여 각 라인별 에지 성분 총량을 구하는 라인 스캐닝부, 상기 영상 프레임을 세로축 방향으로 탐색하여 에지 성분 총량이 제2 임계치를 초과하는 라인의 개수가 일정 개수 이상이면, 자막이 입혀진 영상 프레임으로 판별하는 자막 프레임 판별부를 포함한다.

[0018] 상기 라인 스캐닝부는 각 라인에 있는 에지 성분에 현재 픽셀과 인접 픽셀간의 거리에 따른 가중치를 적용하여 에지 성분 총량을 구한다.

[0019] 또한, 본 발명에 따르면, (a)동영상을 오디오 신호와 영상 신호로 분리하는 단계, (b)상기 분리된 오디오 신호를 오디오 프레임으로 분할하고, 각 오디오 프레임의 음성 에너지를 구하여 자막 검출을 위한 후보 구간을 결정하는 단계, (c)상기 결정된 후보 구간에 해당하는 영상 프레임을 디코딩하여 에지 성분을 구하고, 라인 스캐닝을 통해 자막이 입혀진 영상 프레임을 판별하는 단계, 상기 판별 결과에 따라 영상 프레임별 자막 유무를 기록하여 자막 시퀀스 바이너리 맵을 생성하는 단계를 포함하는 자막이 입혀진 동영상을 검출하기 위한 방법이 제공된다.

[0020] 상기 방법은 상기 생성된 자막 시퀀스 바이너리 맵에서 자막 존재 표시의 밀도와 분포를 이용하여 자막이 입혀진 동영상을 판별하는 단계를 더 포함할 수도 있다.

[0021] 상기 (b)단계는, 상기 오디오 신호를 일정 간격으로 분할하여 복수의 오디오 프레임을 생성하는 단계, 상기 생성된 각 오디오 프레임의 음성 에너지를 구하고, 일정한 윈도우 사이즈 내에 있는 오디오 프레임의 평균 에너지를 구하는 단계, 상기 구해진 평균 에너지를 평균과 분산을 이용하여 정규화하는 단계, 상기 정규화된 평균 에너지가 제1 임계치 이상인 오디오 프레임의 개수를 확인하여, 그 개수가 일정 개수 이상인 경우 해당 구간을 후보 구간으로 결정하는 단계를 포함한다.

[0022] 상기 (c)단계는, 상기 결정된 후보 구간에 해당하는 영상 프레임을 디코딩하는 단계, 상기 디코딩된 영상 프레임을 흑백으로 변환하며 그 사이즈를 정규화하는 단계, 상기 정규화된 영상 프레임에 대해 현재 픽셀과 인접 픽셀과의 차를 이용하여 에지 성분을 구하는 단계, 상기 영상 프레임에 대해 가로축 방향으로 라인 스캐닝을 수행하여 각 라인별 에지 성분 총량을 구하는 단계, 상기 영상 프레임을 세로축 방향으로 탐색하여 에지 성분 총량이 제2 임계치를 초과하는 라인의 개수가 일정 개수 이상이면, 자막이 입혀진 영상 프레임으로 판별하는 단계를 포함한다.

발명의 효과

[0023] 상술한 바와 같이 본 발명에 따르면, 대용량 데이터베이스로부터 자막이 입혀진 동영상을 분류해 냄으로써 콘텐츠를 보유하고 있는 서비스 공급자가 좀더 편하게 사용자 맞춤형 서비스를 제공해 줄 수 있다.

[0024] 또한, 1차원 키워드 중심의 동영상 검색 방법을 통해 사용자 맞춤형 동영상 검색 서비스를 지원할 수 있도록 대용량 데이터베이스로부터 자막이 입혀진 동영상을 자동적으로 선별하므로, 콘텐츠 공급자의 편의성과 사용자의 만족도를 높일 수 있다.

[0025] 또한, 사용자가 동영상을 검색하고자 키워드를 입력함과 동시에 동영상 리스트에 자막의 존재 여부가 표시되므로, 사용자는 좀 더 편하게 보고 싶은 동영상을 찾아볼 수 있다.

도면의 간단한 설명

[0026] 도 1은 본 발명에 따른 자막이 입혀진 동영상을 검출하기 위한 장치의 구성을 개략적으로 나타낸 블럭도.

도 2는 도 1에 도시된 후보 구간 검출 모듈의 구성을 구체적으로 나타낸 블럭도.

도 3은 도 1에 도시된 자막 프레임 검출 모듈의 구성을 구체적으로 나타낸 블럭도.

도 4는 본 발명에 따른 자막이 입혀진 동영상을 검출하기 위한 방법을 나타낸 흐름도.

도 5는 본 발명에 따른 자막 동영상 검출 장치가 자막 검출을 위한 후보 구간을 검출하는 방법을 나타낸 흐름도.

도 6은 본 발명에 따른 자막 동영상 검출 장치가 자막이 입혀진 영상 프레임을 검출하는 방법을 나타낸 흐름도.

도 7은 본 발명에 따른 라인 스캐닝 방법을 설명하기 위한 예시도.

도 8은 본 발명에 따른 시퀀스 인덱싱 방법을 설명하기 위한 예시도.

발명을 실시하기 위한 구체적인 내용

[0027] 본 발명의 전술한 목적과 기술적 구성 및 그에 따른 작용 효과에 관한 자세한 사항은 본 발명의 명세서에 첨부된 도면에 의거한 이하 상세한 설명에 의해 보다 명확하게 이해될 것이다.

[0028] 도 1은 본 발명에 따른 자막이 입혀진 동영상을 검출하기 위한 장치의 구성을 개략적으로 나타낸 블럭도이다.

[0029] 도 1을 참조하면, 자막이 입혀진 동영상을 검출하기 위한 장치는 동영상을 오디오 신호와 영상 신호로 분리하는 신호 분리 모듈(100), 상기 오디오 신호에서 음성이 존재하는 구간을 자막 검출을 위한 후보 구간으로 결정하는 후보 구간 검출 모듈(110), 상기 결정된 후보 구간에 해당하는 영상 프레임을 디코딩하여 자막이 입혀진 영상 프레임을 검출하는 자막 프레임 검출 모듈(120), 상기 판별 결과에 따라 영상 프레임별 자막 유무를 기록하여 자막 시퀀스 바이너리 맵을 생성하는 시퀀스 인덱싱 모듈(130)을 포함한다.

[0030] 상기 후보 구간 검출 모듈(110)은 상기 오디오 신호를 일정 간격으로 분할하여 복수의 오디오 프레임을 생성하고, 각 오디오 프레임의 음성 에너지를 구하여 정규화한다. 그런 다음 상기 후보 구간 검출 모듈(110)은 상기 정규화된 평균 음성 에너지가 임계치 이상인 오디오 프레임의 개수를 확인하고, 그 개수가 일정 개수 이상인 경우 해당 구간을 후보 구간으로 결정한다. 상기 후보 구간 검출 모듈(110)에 대한 상세한 설명은 도 2를 참조하기로 한다.

[0031] 상기 자막 프레임 검출 모듈(120)은 상기 후보 구간 검출 모듈(110)에서 결정된 후보 구간에 해당하는 영상 프레임을 디코딩하여 에지 성분을 구하고, 라인 스캐닝(line scanning)을 통해 각 라인별 에지 성분 총량을 구하여 자막이 입혀진 영상 프레임을 판별한다. 상기 자막 프레임 검출 모듈(120)에 대한 상세한 설명은 도 3을 참조하기로 한다.

[0032] 상기 시퀀스 인덱싱 모듈(130)은 상기 자막 프레임 검출 모듈의 판정 결과를 입력받아 영상 프레임의 자막 존재 여부를 순차적으로 기록한 자막 시퀀스 바이너리 맵(SC)을 생성한다. 즉, 상기 시퀀스 인덱싱 모듈(130)은 자막이 존재하는 영상 프레임인 경우 1(참)로 표시하고, 자막이 존재하지 않은 경우 0(거짓)으로 표시하여, 도 8과 같이 자막 존재 여부에 따라 1과 0이 순차적으로 기록된 자막 시퀀스 바이너리 맵을 생성한다.

[0033] 상기 자막이 입혀진 동영상을 검출하기 위한 장치는 상기 시퀀스 인덱싱 모듈(130)에서 생성된 자막 시퀀스 바이너리 맵을 통해 최종적으로 자막이 입혀진 동영상을 판별하는 자막 동영상 판별 모듈(140)을 더 포함하여 구성할 수 있다.

[0034] 상기 자막 동영상 판별 모듈(140)은 상기 자막 시퀀스 바이너리 맵에서 자막 존재 표시가 임계치 이상인 경우 해당 동영상을 자막이 입혀진 동영상으로 판별한다. 즉, 상기 자막 동영상 판별 모듈(140)은 참(1)과 거짓(0)로 표시된 자막 시퀀스 바이너리 맵에서 참(1)의 분포와 밀도를 비교하여 최종적으로 동영상 내 자막의 유무를 판별하게 된다.

[0035] 예를 들어, 상기 자막 시퀀스 바이너리 맵을 분석한 결과 참이 거짓보다 일정 비율 더 많고 참의 분포가 몰려 있으면, 해당 동영상을 자막이 입혀진 동영상으로 판별할 수 있다.

[0036] 도 2는 도 1에 도시된 후보 구간 검출 모듈의 구성을 구체적으로 나타낸 블럭도이다.

[0037] 도 2를 참조하면, 후보 구간 검출 모듈(110)은 오디오 신호 분할부(112), 음성 에너지 계산부(114), 정규화부(116), 후보 구간 결정부(118)를 포함한다.

[0038] 상기 오디오 신호 분할부(112)는 상기 오디오 신호를 일정 간격으로 분할하여 복수의 오디오 프레임을 생성한다. 즉, 상기 오디오 신호 분할부(112)는 상기 오디오 신호를 초당 M frame per second(fps)로 분할하여 오디오 프레임을 생성한다. 예를 들어 동영상의 오디오 신호가 16KHz로 샘플링된 경우 100fps라 하면 160개의 샘플이 하나의 오디오 프레임이 된다.

[0039] 상기 음성 에너지 계산부(114)는 상기 오디오 신호 분할부(112)에서 생성된 각 오디오 프레임의 음성 에너지를

구하고, 일정한 윈도우 사이즈 내에 있는 오디오 프레임에 대한 평균 에너지를 구한다. 이때, 윈도우 사이즈는 오디오 프레임의 샘플 수보다 크게 두어 연속하는 두 개의 프레임이 중첩이 되도록 구성한다.

[0040] 예를 들어, 16KHz 오디오를 100fps로 분할한 경우 윈도우 사이즈를 256으로 두는 것과 같다. 윈도우 내 평균 에너지(Em)를 구하는 수식은 수학식 1과 같다.

수학식 1

[0041] $$E_m = \frac{1}{N_a} log \sum_{n=1}^{N_a} |x(n)|$$

[0042] 여기서, 상기 Na는 윈도우 사이즈, x(n)은 n번째 오디오 샘플을 나타낸다.

[0043] 상기 정규화부(116)는 상기 음성 에너지 계산부(114)에서 구해진 평균 에너지, 초당 오디오 프레임의 개수, 오디오 신호의 초단위 길이를 이용하여 평균과 분산을 구하고, 상기 구해진 평균과 분산을 이용하여 상기 평균 에너지를 정규화한다. 즉, 상기 정규화부(116)는 상기 구해진 각 프레임의 평균 에너지들을 T_a초 만큼의 구간에 대해 정규화해주는 작업을 수행한다.

[0044] 상기 후보 구간 결정부(118)는 상기 정규화된 평균 음성 에너지가 임계치 이상인 오디오 프레임의 개수를 확인하고, 그 개수가 일정 개수 이상인 경우 해당 구간을 후보 구간으로 결정한다. 즉, 상기 후보 구간 결정부(118)는 상기 정규화된 평균 오디오 프레임 에너지를 입력으로 영상 프레임 디코딩을 위한 후보 구간을 결정하게 된다. 이는 음성신호가 활성화된 구간의 영상 프레임을 디코딩하여 자막을 검출함으로써 자막이 존재할 가능성이 높은 구간에 대해서만 자막 판별을 수행하기 위함이다.

[0045] 도 3은 도 1에 도시된 자막 프레임 검출 모듈의 구성을 구체적으로 나타낸 블럭도이다.

[0046] 도 3을 참조하면, 상기 자막 프레임 검출 모듈(120)은 디코딩부(121), 영상 프레임 정규화부(122), 에지 성분 결정부(123), 라인 스캐닝부(124), 자막 프레임 판별부(125)를 포함한다.

[0047] 상기 디코딩부(121)는 후보 구간 검출 모듈에서 검출된 후보 구간에 해당하는 영상 프레임을 디코딩한다. 즉, 상기 디코딩부(121)는 상기 후보 구간 검출 모듈에서 결정된 후보 구간에 대해 자막을 검출하기 위해 영상 프레임 시퀀스를 D_s초 단위로 샘플링하여 디코딩을 수행한다.

[0048] 예를들어 30fps(frame per second) 영상의 경우 D_s를 1초로 한다면, 영상 프레임 시퀀스 30장당 1장씩 디코딩을 수행하는 것과 같다.

[0049] 상기 영상 프레임 정규화부(122)는 상기 디코딩부(121)에서 디코딩된 영상 프레임을 흑백으로 변환하며 그 사이즈를 정규화한다. 즉, 상기 영상 프레임 정규화부(122)는 컬러 영상 프레임을 흑백 영상 프레임으로 변환함과 동시에 가로 길이 W_t, 세로 길이 H_t를 갖는 영상으로 크기를 정규화한다.

[0050] 상기 에지 성분 결정부(123)는 상기 정규화된 영상 프레임에 대해 현재 픽셀과 인접 픽셀과의 차를 구하고, 상기 구해진 차가 일정값 이상인 픽셀을 에지 성분이라고 결정한다.

[0051] 상기 라인 스캐닝부(124)는 상기 영상 프레임에 대해 가로축 방향으로 라인 스캐닝을 수행하여 각 라인별 에지 성분 총량을 구한다. 즉, 상기 라인 스캐닝부(124)는 각 라인에 있는 에지 성분에 현재 픽셀과 인접 픽셀간의 거리에 따른 가중치를 적용하여 에지 성분 총량을 구한다. 상기 에지 성분 총량(S_j)을 구하는 식은 다음과 같다.

수학식 2

[0052] $$s_j = \sum_{1}^{W_c} w(d)\delta(i,j)$$

[0053] 여기서, 상기 w(d)는 현재 픽셀과 인접 픽셀간의 거리에 따른 가중치를 말하고, $\delta(i,j)$는 픽셀을 나타낸 것으로 에지 성분인 경우 1, 에지 성분이 아닌 경우 0의 값을 가진다. 상기 가중치는 거리가 가까울수록 큰 값을 부여하는 형태일 수 있다.

[0054] 상기 라인 스캐닝부(124)는 상기 영상 프레임의 에지로부터 영상 프레임 상의 자막 영역을 검출하기 위해 영상 프레임 내에서 높이 j를 갖는 가로축의 한 라인에서의 에지 성분 총량(s_j)을 계산한다. 이는 가로축 방향의 라인 탐색을 수행함으로써 구하게 된다.

[0055] 상기 자막 프레임 판별부(125)는 상기 영상 프레임을 세로축 방향으로 탐색하여, 에지 성분 총량이 임계치를 초과하는 라인의 개수가 일정 개수 이상이면, 자막이 입혀진 영상 프레임으로 판별한다.

[0056] 도 4는 본 발명에 따른 자막이 입혀진 동영상을 검출하기 위한 방법을 나타낸 흐름도이다.

[0057] 도 4를 참조하면, 자막이 입혀진 동영상을 검출하기 위한 장치(이하에서는, 자막 동영상 검출 장치라고 칭하기로 함)는 동영상을 오디오 신호와 영상 신호로 분리한다(S400).

[0058] 그런 다음 상기 자막 동영상 검출 장치는 상기 분리된 오디오 신호를 오디오 프레임으로 분할하고, 각 오디오 프레임의 음성 에너지를 구하여 자막 검출을 위한 후보 구간을 결정한다(S402). 즉, 상기 자막 동영상 검출 장치는 상기 분리된 오디오 신호로부터 음성 존재 여부를 판단하여 음성이 존재하는 구간을 자막 검출을 위한 후보 구간으로 결정한다. 상기 자막 검출을 위한 후보 구간을 결정하는 방법에 대한 상세한 설명은 도 5를 참조하기로 한다.

[0059] 상기 S402의 수행 후, 상기 자막 동영상 검출 장치는 상기 결정된 후보 구간에 해당하는 영상 프레임을 디코딩하여 에지 성분을 구하고, 라인 스캐닝을 통해 자막이 입혀진 영상 프레임을 판별한다(S404).

[0060] 상기 자막이 입혀진 영상 프레임을 판별하는 방법에 대한 상세한 설명은 도 6을 참조하기로 한다.

[0061] 상기 S404의 수행 후, 상기 자막 동영상 검출 장치는 상기 판별 결과에 따라 영상 프레임별 자막 유무를 순차적으로 기록하여 자막 시퀀스 바이너리 맵(SC)을 생성한다(S406).

[0062] 그런 다음 상기 자막 동영상 검출 장치는 해당 동영상에 자막이 입혀져 있는지의 여부를 판단하기 위하여 상기 자막 시퀀스 바이너리 맵에서 자막 존재 표시가 임계치 이상인지의 여부를 판단한다(S408). 즉, 상기 자막 동영상 검출 장치는 참의 분포가 거짓의 분포보다 일정 비율 이상인지의 여부를 판단한다.

[0063] 상기 S408의 판단결과 임계치 이상이면, 상기 자막 동영상 검출 장치는 해당 동영상에는 자막이 입혀져 있다고 판단하고(S410), 임계치 이상이 아니면 상기 동영상은 자막이 입혀져 있지 않다고 판단한다(S412).

[0064] 상기와 같은 과정을 통해 대용량 동영상 데이터베이스로부터 자막이 입혀진 동영상을 자동으로 분류해낼 수 있다.

[0065] 상기 방법을 통해 사용자가 동영상을 검색하고자 키워드를 입력하면, 상기 자막 동영상 검출 장치는 구비된 동영상 데이터베이스에서 상기 키워드에 상응하는 동영상 리스트를 출력하는데, 상기 동영상 리스트에는 자막이 입혀진 동영상인지의 여부가 표시되어 있다.

[0066] 따라서, 검색한 동영상이 자막이 입혀진 동영상인지 아닌지를 판별하기 위해서 해당 동영상을 일일이 재생해야 하는 불편함을 해소하였다.

[0067] 도 5는 본 발명에 따른 자막 동영상 검출 장치가 자막 검출을 위한 후보 구간을 결정하는 방법을 나타낸 흐름도

이다.

[0068] 도 5를 참조하면, 자막 동영상 검출 장치는 입력된 오디오 신호를 일정 간격으로 분할하여 복수의 오디오 프레임을 생성하고(S500), 상기 생성된 각 오디오 프레임의 음성 에너지를 계산한다(S502). 이때, 상기 자막 동영상 검출 장치는 전체 오디오 신호에 대해 일정 크기로 윈도우 사이즈를 정하고, 그 윈도우 사이즈 내에 있는 오디오 프레임에 대한 평균 에너지를 구한다.

[0069] 그런 다음 상기 자막 동영상 검출 장치는 상기 구해진 평균 에너지를 평균과 분산을 이용하여 정규화하고(S504), 상기 정규화된 평균 에너지가 임계치 이상인 오디오 프레임의 개수를 확인하고, 그 개수가 일정 개수 이상인지의 여부를 판단한다(S506).

[0070] 상기 S506의 판단결과 일정 개수 이상이면, 상기 자막 동영상 검출 장치는 해당 구간을 자막 검출을 위한 후보 구간으로 결정한다(S508).

[0071] 도 6은 본 발명에 따른 자막 동영상 검출 장치가 자막이 입혀진 영상 프레임을 검출하는 방법을 나타낸 흐름도, 도 7은 본 발명에 따른 라인 스캐닝 방법을 설명하기 위한 예시도이다.

[0072] 도 6을 참조하면, 자막 동영상 검출 장치는 자막 검출을 위한 후보 구간에 해당하는 영상 프레임을 디코딩하고(S600), 상기 디코딩된 영상 프레임을 흑백으로 변환하며 그 사이즈를 정규화한다(S602).

[0073] 그런 다음 상기 자막 동영상 검출 장치는 상기 정규화된 영상 프레임의 각 픽셀에 대해 인접 픽셀과의 차를 이용하여 에지 성분을 구하고(S604), 상기 영상 프레임에 대해 가로축 방향으로 라인 스캐닝을 수행하여 각 라인별 에지 성분 총량을 구한다(S606). 즉, 상기 자막 동영상 검출 장치가 라인 스캐닝하는 방법에 대하여 도 7을 참조하기로 한다.

[0074] 상기 자막 동영상 검출 장치는 최종적으로 자막 영역을 검출하기 위해 영상 프레임의 세로축 방향으로 모든 높이 값 j에 대해 s_j를 탐색하는 과정을 수행한다. 여기서, 영상 프레임은 H_t 개의 라인을 갖고 있으므로, 상기 자막 동영상 검출 장치는 각 라인별로 스캐닝하여 H_t 개의 에지 성분 총량(s_j)를 구한다.

[0075] 상기 S606의 수행 후, 상기 자막 동영상 검출 장치는 상기 영상 프레임에 대해 세로축 방향으로 각 라인별 에지 성분 총량을 탐색하여(S608), 에지 성분 총량이 임계치를 초과하는 라인의 개수를 구한다(S610).

[0076] 그런 다음 상기 자막 동영상 검출 장치는 상기 구해진 개수가 일정 개수 이상인지를 판단하여(S612), 일정 개수 이상이면 해당 영상 프레임을 자막이 입혀진 영상 프레임으로 판별한다(S614).

[0077] 만약, 상기 S612의 판단결과 일정 개수 이상이 아니면, 해당 영상 프레임을 자막이 없는 영상 프레임으로 판별한다(S616).

[0078] 이와 같이, 본 발명이 속하는 기술분야의 당업자는 본 발명이 그 기술적 사상이나 필수적 특징을 변경하지 않고서 다른 구체적인 형태로 실시될 수 있다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적인 것이 아닌 것으로서 이해해야만 한다. 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 등가개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

부호의 설명

[0079]

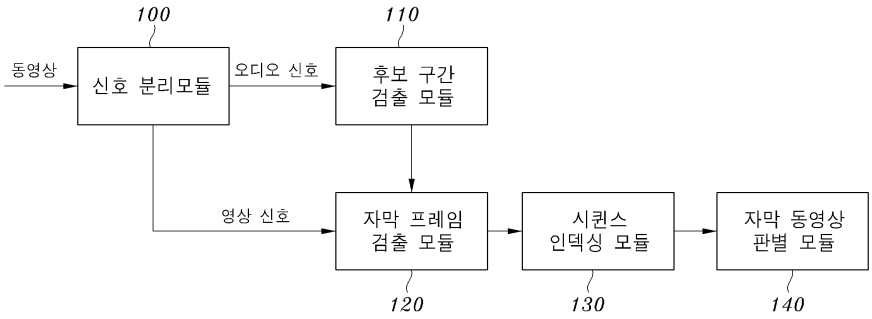
100 : 신호 분리 모듈	110 : 후보 구간 검출 모듈
112 : 오디오 신호 분할부	114 : 음성 에너지 계산부
116 : 정규화부	118 : 후보 구간 결정부
120 : 자막 프레임 검출 모듈	121 : 디코딩부
122 : 영상 프레임 정규화부	123 : 에지 성분 계산부

124 : 라인 스캐닝부　　125 : 자막 프레임 판별부

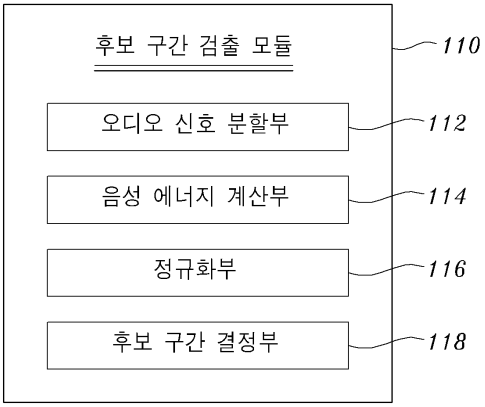
130 : 시퀀스 인덱싱 모듈　　140 : 자막 동영상 판별 모듈

도면

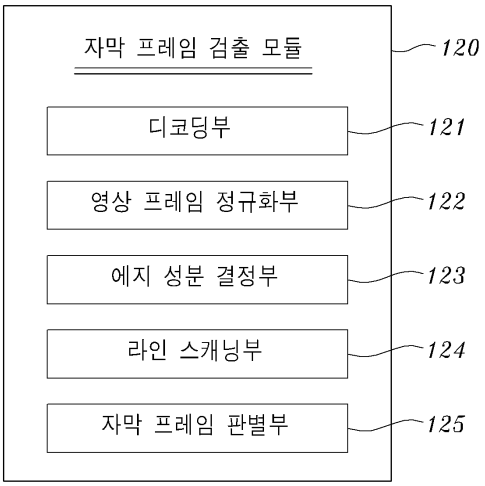
도면1



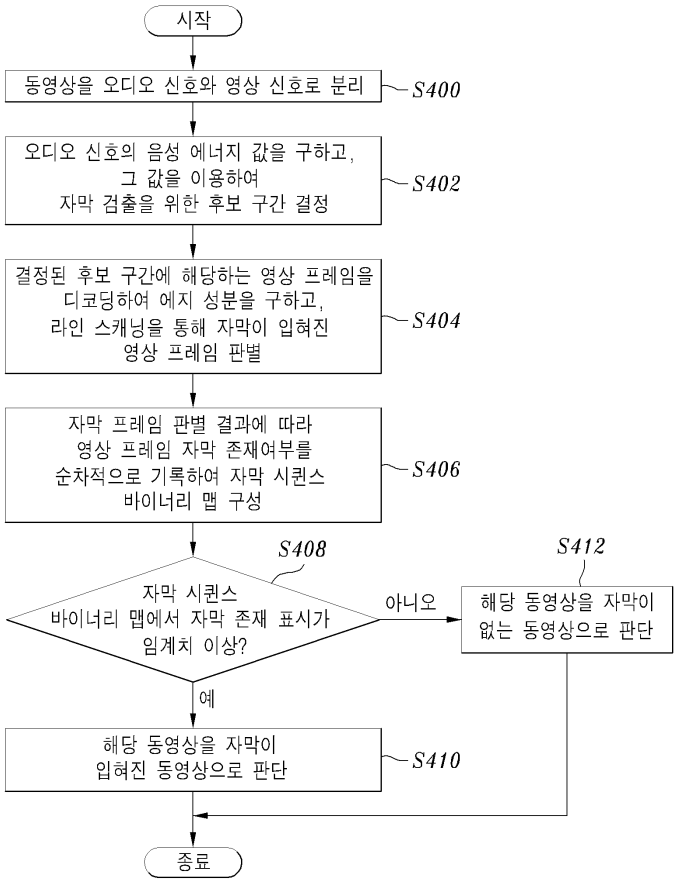
도면2



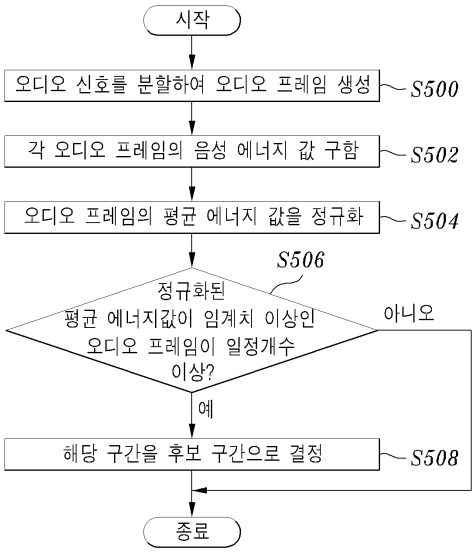
도면3



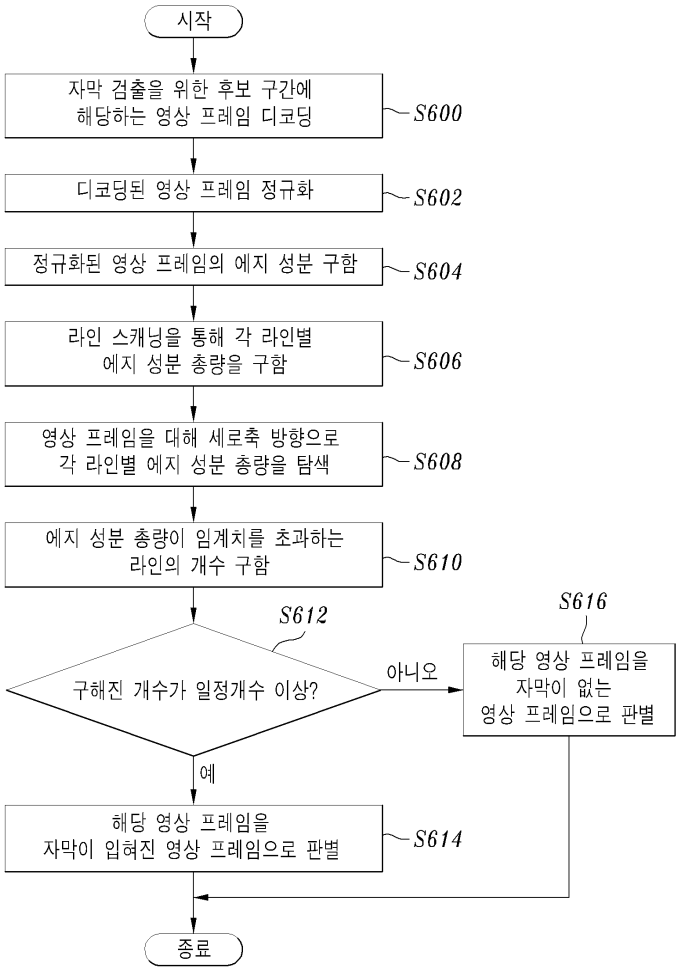
도면4


시작
동영상을 오디오 신호와 영상 신호로 분리
S400
오디오 신호의 음성 에너지 값을 구하고,
그 값을 이용하여
자막 검출을 위한 후보 구간 결정
S402
결정된 후보 구간에 해당하는 영상 프레임을
디코딩하여 에지 성분을 구하고,
라인 스캐닝을 통해 자막이 입혀진
영상 프레임 판별
S404
자막 프레임 판별 결과에 따라
영상 프레임 자막 존재여부를
순차적으로 기록하여 자막 시퀀스
바이너리 맵 구성
S406
S408
자막 시퀀스
바이너리 맵에서 자막 존재 표시가
임계치 이상?
아니오
S412
해당 동영상을 자막이
없는 동영상으로 판단
예
해당 동영상을 자막이
입혀진 동영상으로 판단
S410
종료

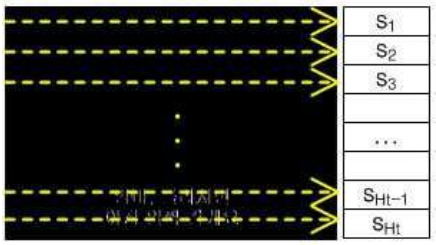
도면5


시작
오디오 신호를 분할하여 오디오 프레임 생성
S500
각 오디오 프레임의 음성 에너지 값 구함
S502
오디오 프레임의 평균 에너지 값을 정규화
S504
S506
정규화된
평균 에너지값이 임계치 이상인
오디오 프레임이 일정개수
이상?
아니오
예
해당 구간을 후보 구간으로 결정
S508
종료

도면6


시작
자막 검출을 위한 후보 구간에 해당하는 영상 프레임 디코딩
S600
디코딩된 영상 프레임 정규화
S602
정규화된 영상 프레임의 에지 성분 구함
S604
라인 스캐닝을 통해 각 라인별 에지 성분 총량을 구함
S606
영상 프레임을 대해 세로축 방향으로 각 라인별 에지 성분 총량을 탐색
S608
에지 성분 총량이 임계치를 초과하는 라인의 개수 구함
S610
S612
구해진 개수가 일정개수 이상?
아니오
S616
해당 영상 프레임을 자막이 없는 영상 프레임으로 판별
예
해당 영상 프레임을 자막이 입혀진 영상 프레임으로 판별
S614
종료

도면7


S1
S2
S3
...
SHt-1
SHt

도면8