



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2009년09월07일
(11) 등록번호 10-0915847
(24) 등록일자 2009년08월31일

- (51) Int. Cl.
G11B 20/10 (2006.01)
- (21) 출원번호 10-2002-7016141
- (22) 출원일자 2002년03월19일
심사청구일자 2007년03월19일
- (85) 번역문제출일자 2002년11월28일
- (65) 공개번호 10-2003-0007736
- (43) 공개일자 2003년01월23일
- (86) 국제출원번호 PCT/IB2002/000896
- (87) 국제공개번호 WO 2002/80524
국제공개일자 2002년10월10일
- (30) 우선권주장
09/822,447 2001년03월30일 미국(US)
- (56) 선행기술조사문헌
W00058967 A1*
W09849685 A
W09962008 A
*는 심사관에 의하여 인용된 문헌

- (73) 특허권자
코닌클리케 필립스 일렉트로닉스 엔.브이.
네델란드왕국, 아인드호펜, 그로네보르스베그 1
- (72) 발명자
디미트로바, 네벤카
네델란드, 아아아인드호펜5656, 프로프. 홀스트란6
아그니호트리, 랄리사
네델란드, 아아아인드호펜5656, 프로프. 홀스트란6
(뒷면에 계속)
- (74) 대리인
장훈

전체 청구항 수 : 총 23 항

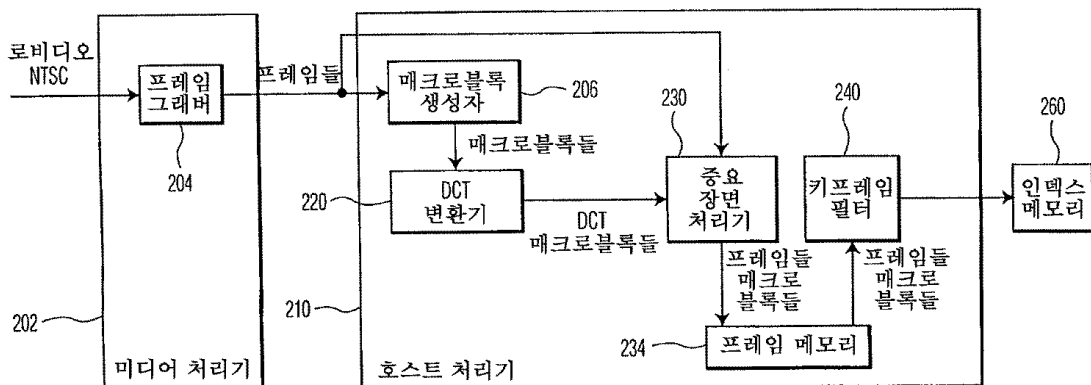
심사관 : 안지현

(54) 스트리밍 비디오 북마크들

(57) 요약

본 발명은 저장된 비디오 콘텐츠 중의 관심있는 부분을 북마크하는 방법, 장치 및 시스템들에 관한 것이다. 시청자가 비디오를 시청하면서 관심있는 부분을 발견할 때, 시청자들은 비디오의 특정 세그먼트를 북마크하고 이후에 상대적으로 간단하게 그 세그먼트로 되돌아갈 수 있다. 이는 버튼을 누르거나, 마우스로 클릭하거나, 다르게는 신호를 장치로 전송하여 비디오의 관심있는 특정 위치를 표시하여, 달성된다. 또한 프레임 식별자들을 이용하여 인덱스로부터 원하는 비디오를 선택하고, 이후 다수의 비디오들을 포함하는 매체로부터 비디오를 검색할 수 있다.

대표도



(72) 발명자

맥지, 토마스, 에프.

네델란드, 아아아인드호펜5656, 프로프. 홀스트란6

자신스키, 라두, 에스.

네델란드, 아아아인드호펜5656, 프로프. 홀스트란6

(81) 지정국

국내특허 : 중국, 일본, 대한민국

EP 유럽특허 : 오스트리아, 벨기에, 스위스, 독일, 덴마크, 스페인, 프랑스, 영국, 그리스, 아일랜드, 이탈리아, 룩셈부르크, 모나코, 네덜란드, 포르투갈, 터키, 스웨덴, 사이프러스, 핀란드

특허청구의 범위

청구항 1

비디오의 세그먼트들을 형성하는 방법에 있어서:

일련의 비디오 프레임들을 포함하는 비디오 콘텐츠를 입력받아 상기 비디오를 디스플레이하는 단계;

상기 비디오의 디스플레이 동안 북마크 신호를 송신하여, 상기 북마크 신호가 송신될 때 디스플레이되고 있는 프레임을 선택하는 단계;

상기 선택된 프레임의 특성적 특징을 식별하고(503) 상기 프레임에 대응하는 표시점(marking point)을 저장 매체 상에 기록하는(503) 단계; 및

상기 선택된 프레임을 포함하고 세그먼트 시작점 및 세그먼트 종료점을 갖는 상기 비디오의 세그먼트를 식별하여(504) 상기 세그먼트의 적어도 일부분을 저장하는(505) 단계를 포함하고,

상기 세그먼트 시작점으로부터 상기 비디오의 미리 규정된 길이가 저장되거나(505) 상기 세그먼트 시작점은 상기 선택된 프레임 전에 미리 규정되는 상기 비디오의 길이로 결정되는(504), 비디오 세그먼트 형성 방법.

청구항 2

제 1 항에 있어서, 상기 프레임의 특성적 특징은 상기 프레임의 영상, 음성 또는 텍스트 성질들을 이용하여 생성되는 프레임 시그니처(frame signature)로부터 얻어지는, 비디오 세그먼트 형성 방법.

청구항 3

제 1 항에 있어서, 상기 프레임의 특성적 특징은 자신의 프레임 번호인, 비디오 세그먼트 형성 방법.

청구항 4

제 1 항에 있어서, 상기 프레임의 특성적 특징은 상기 비디오에서의 자신의 타임 스탬프(time stamp)인, 비디오 세그먼트 형성 방법.

청구항 5

제 1 항에 있어서, 상기 세그먼트의 시작점과 종료점은 상기 선택된 프레임의 전과 후의 상기 비디오의 음성, 영상 또는 트랜스크립트(transcript) 부분들에서 정보의 의미 있는 변화들을 검출하는 것에 기초하는, 비디오 세그먼트 형성 방법.

청구항 6

제 5 항에 있어서, 변화들을 검출하기 위해 해석되는 상기 정보는 상기 비디오의 DCT 계수들, 색상, 에지, 형태, 무음, 스피치, 음악, 자막 및 음성 콘텐츠와, 이들의 조합들로 구성되는 그룹으로부터 선택되는, 비디오 세그먼트 형성 방법.

청구항 7

제 1 항에 있어서, 상기 세그먼트는 상기 비디오와 동일한 매체 상에 저장되는(505), 비디오 세그먼트 형성 방법.

청구항 8

삭제

청구항 9

삭제

청구항 10

제 1 항에 있어서, 상기 저장된 세그먼트는 EPG 데이터, 상기 세그먼트의 프레임, 또는 상기 세그먼트로부터의 트랜스크립트 정보 및 이들의 조합들 중 적어도 하나를 포함하는, 비디오 세그먼트 형성 방법.

청구항 11

추후 검색을 위해 비디오의 세그먼트들을 식별하기 위한 시스템에 있어서,

일련의 비디오 프레임들을 포함하는 비디오 콘텐츠를 디스플레이할 수 있는 비디오 플레이어와 관련하여 작동하도록 구성된 북마커(bookmarker)를 포함하며,

상기 북마커는 북마크 신호를 상기 북마커로 송신하도록 구성된 신호 전송기로부터의 신호에 응답하여, 상기 북마커는 상기 북마크 신호가 수신되었을 때 디스플레이되고 있는 상기 비디오의 선택된 프레임을 식별하도록 (503) 구성되고,

또한, 상기 북마커는 상기 비디오가 이미 세그먼트 시작점 및 세그먼트 종료점을 갖는 세그먼트들로 나누어졌을 때 상기 선택된 프레임의 특성적 특징을 식별하여 상기 특성적 특징에 대응하는 표시점을 저장 매체 상에 기록하도록 구성되고, 상기 선택된 프레임을 포함하는 상기 세그먼트의 적어도 일부분을 저장 매체 상에 저장하며 (505), 또는 상기 북마커는 상기 비디오를 상기 세그먼트들로 나눌 수 있는(504) 분할 특징을 가지고 상기 북마커는 상기 선택된 프레임을 포함하는 상기 세그먼트의 적어도 일부분을 상기 저장 매체에 저장하도록(505) 구성되고,

상기 북마커는 상기 세그먼트의 적어도 일부분을 상기 저장 매체 상에 저장하도록(505) 구성되거나, 상기 북마커는 상기 세그먼트 시작점으로부터 상기 비디오의 미리 규정된 길이를 저장하도록(505) 구성되는, 비디오 세그먼트 식별 시스템.

청구항 12

제 11 항에 있어서, 상기 선택된 프레임의 특성적 특징은 상기 프레임의 영상, 음성 또는 텍스트 성질들을 이용하여 생성되는 프레임 시그니처로부터 얻어지는, 비디오 세그먼트 식별 시스템.

청구항 13

제 11 항에 있어서, 상기 프레임의 특성적 특징은 자신의 프레임 번호이고, 상기 북마커는 상기 프레임 번호를 결정하도록 구성되는, 비디오 세그먼트 식별 시스템.

청구항 14

제 11 항에 있어서, 상기 프레임의 특성적 특징은 상기 비디오에서의 자신의 타임 스탬프(time stamp)이고, 상기 북마커는 상기 비디오에서 상기 타임 스탬프를 결정하도록 구성되는, 비디오 세그먼트 식별 시스템.

청구항 15

제 11 항에 있어서, 상기 북마커는 상기 세그먼트의 시작점 및 종료점을 결정하기 위해서 상기 비디오의 음성, 영상 또는 트랜스크립트 부분들에서 정보의 의미 있는 변화들을 검출하도록 구성되는 해석 엔진(analysis engine)을 포함하는, 비디오 세그먼트 식별 시스템.

청구항 16

제 15 항에 있어서, 상기 해석 엔진은 상기 비디오의 DCT 계수들, 색상, 에지, 형태, 무음, 스피치, 음악, 자막 및 음성 콘텐츠와, 이들의 조합들로 구성되는 그룹으로부터 선택되는 정보를 해석하는, 비디오 세그먼트 식별 시스템.

청구항 17

삭제

청구항 18

삭제

청구항 19

원하는 비디오를 식별하는 방법에 있어서:

상기 원하는 비디오의 프레임을 선택하고 상기 프레임과 관련된 식별 특징들을 기록하는 단계(503);

다중 비디오들을 포함하는 저장 매체 상에 상기 원하는 비디오를 저장하는 단계;

상기 식별 특징들을 검색함으로써(503) 상기 원하는 비디오를 위해 상기 저장 매체를 검색하는 단계; 및

상기 식별 특징들을 갖는 상기 프레임을 포함하는 상기 비디오를 검색하는 단계(503)를 포함하고,

상기 매체상에 저장된 상기 비디오들의 세그먼트들은 비디오 인덱스로 컴파일되고 상기 원하는 비디오의 세그먼트는 상기 선택된 프레임을 포함하거나, 상기 비디오들은 제 1 위치에 저장되어 상기 인덱스는 인터넷을 통해서 원격지에 있는 이용자에게 전송되는, 비디오 식별 방법.

청구항 20

제 19 항에 있어서, 상기 프레임의 상기 식별 특징들은 상기 비디오의 영상, 음성 또는 텍스트 성질들을 이용하여 생성된 프레임 시그니처로부터 얻어지는, 비디오 식별 방법.

청구항 21

삭제

청구항 22

삭제

청구항 23

전자적으로 저장된 콘텐츠를 처리하는 방법에 있어서,

비디오들, 음성, 이미지들, 텍스트 및 이들의 조합으로 구성되는 그룹으로부터 선택되는 콘텐츠 항목들을 전자적으로 저장하고 디스플레이하는 단계,

상기 콘텐츠를 디스플레이하면서 상기 콘텐츠 항목을 식별하기 위한 충분한 식별 정보를 갖는 상기 콘텐츠 항목의 선택된 세그먼트를 저장(505)함으로써 북마크를 생성(503)하는 단계,

이후, 상기 항목을 식별하는 상기 세그먼트만을 저장 장치상에 보유하는 단계, 및

이후, 나중에, 상기 항목을 식별하는 상기 정보를 비교하여 동일한 항목을 식별하는 단계를 포함하는, 콘텐츠 처리 방법.

청구항 24

제 23 항에 있어서, 상기 북마크들은 원격 장치에 백업되는, 콘텐츠 처리 방법.

청구항 25

제 23 항에 있어서, 상기 북마크들은 EPG 데이터, 트랜스크립트 정보, 자막 정보 또는 비디오 특징 정보 중 적어도 하나에 기초하여 카테고리화되는, 콘텐츠 처리 방법.

청구항 26

제 23 항에 있어서, 상기 북마크들은 웹 페이지들, 무선 접속을 통한 이동 통신 장치들, PDA들 및 컴퓨터화된 시계들 중 적어도 하나에 의해서 액세스되는, 콘텐츠 처리 방법.

청구항 27

제 23 항에 있어서, 상기 북마크들은 다중 레벨들로 저장되고, 서로 다른 개개들은 서로 다른 레벨들로 액세스되는, 콘텐츠 처리 방법.

청구항 28

제 23 항에 있어서, 제 1 저장 장치를 이용하여 상기 세그먼트를 저장 매체에 저장함으로써 상기 북마크가 생성되어(503) 다른 장치와 연관된 저장 매체로 전달되는, 콘텐츠 처리 방법.

청구항 29

제 23 항에 있어서, 상기 콘텐츠 항목은 상기 북마크가 생성된 후에(503) 그것이 저장되었던 매체로부터 삭제되고, 반면 상기 북마크만을 저장하는, 콘텐츠 처리 방법.

청구항 30

삭제

명세서

기술분야

<1> 본 발명은 저장된 비디오 콘텐츠에 액세스하는 방법에 관한 것이며, 보다 상세하게는 나중에 편리하게 검색할 수 있도록 비디오 신호 중에 포함된 의미있는 세그먼트들(meaningful segments)을 식별하기 위해 비디오 콘텐츠를 북마크하는 방법과 장치에 관한 것이다.

배경기술

<2> 이용자들은 바로 시청하거나 나중에 시청하기 위해서 종종 VHS 형식, DVD, 디스크들, 파일들 또는 다른 형태로 저장된 다수의 비디오들을 갖고 있다. 종종 비디오들은 길이가 상당하며, 다양한 콘텐츠를 포함하고 있다. 예를 들어 시청자는 다양한 텔레비전 프로그램들 또는 개인적인 활동들을 기록하는 것을 포함하여 몇 시간 분량의 콘텐츠를 단일한 비디오 카세트, 하드 디스크 또는 다른 저장 매체 상에 기록해 둘 수 있다. 종종 시청자들이 비디오 중에서 특히 중요한 부분들(significant portions)로 되돌아가는 것은 어려운 일이다. 프레임 카운터 숫자들을 기록해 두거나 시간 정보를 기록해 두는 것은, 특히 비디오를 시청하는 중에는 종종 불편하다.

<3> 이용자들은 종종 좌절감을 안겨주는 무작정적인 방법들을 이용하여 특별히 관심있는 세그먼트들로 되돌아가려 한다. 예를 들어, 시청자가 다수의 코미디언들 또는 피겨 스케이트 선수들의 연기들을 포함하는 비디오를 기록하거나 얻을 수 있지만 이들 각각들 중에서 상대적으로 소수의 연기들에만 관심을 가질 수도 있다. 또한, 시청자는 이들 방송을 기록하면서 슈퍼볼(Superbowl)이나 월드 시리즈(World Series)를 시청할 수도 있으며, 게임 중에서 대역섯 군데의 기억나는 플레이 내용들로 되돌아가고 싶어할 수도 있다.

<4> 특별히 관심있는 세그먼트들로 위치시키는 기존의 방법들은 이용하기가 불편하며, 따라서 비디오의 의미있는 세그먼트를 북마크하기 위한 개선된 장치 및 방법을 제공할 필요가 있다.

발명의 상세한 설명

<5> 대략적으로 말하자면, 본 발명에 따르면 저장된 비디오 콘텐츠 중의 관심있는 영역을 북마크하는 방법, 장치 및 시스템들이 제공된다. 시청자가 비디오를 시청하고 관심있는 영역을 발견하면, 비디오의 이 특정 세그먼트를 북마크하고 나중에 비교적 간단하게 북마크한 세그먼트로 되돌아갈 수 있게 된다. 이는 버튼을 누르거나 마우스로 클릭하거나 다르게는 신호를 장치로 송신하여 관심있는 비디오의 특정 위치를 표시하여 달성된다. 따라서 전체 세그먼트의 경계들은 비디오 신호 중의 영상, 음성 및 트랜스크립트 부분들을 해석하여 다양한 슈퍼히스토그램들(superhistograms), 프레임 시그니처들(frame signatures), 장면 변환 검출 방법들, 자막 기능 정보, 음성 정보 등을 이용하여 자동으로 식별할 수 있게 된다. 색상, 에지 및 형태에서의 변화들을 분석하기 위해서 영상 정보를 해석하여 페이스 변화들(face changes), 키프레임들, 비디오텍스트들 등에 의한 각각의 변화들을 결정할 수 있다. 다양한 음성 특징들, 예를 들어 무음, 잡음, 스피치, 음악 및 이들의 조합들을 해석하여 세그먼트의 시작과 종료를 결정한다. 자막 기능 정보 또한 단어들, 카테고리들 등에 관해서 해석될 수 있다. 이 정보를 처리하여 비디오의 의미있는 세그먼트의 경계들을 결정하게 되면, 북마크는 단순히 비디오의 특정 지점에 대응하

는 것이 아니라 자동으로 생성된 콘텐츠 세그먼트 전체와 대응하게 된다.

- <6> 따라서, 본 발명에 따른 북마크 방법들, 시스템들 및 장치들은 간단하게 이용자가 비디오의 관심있는 세그먼트로 되돌아가도록 하는 것뿐만 아니라, 이용자가 세그먼트의 시작 부분으로 갈 수 있고, 선택적으로 관심있는 특정 세그먼트만을 시청하거나, 스크롤을 하거나, 관심있는 연속적인 세그먼트만을 시청할 수도 있다.
- <7> 예를 들어, 한편으로는 특정 화자(speaker)가 시트콤의 비디오에서 말하는 동안 북마크 신호가 송신된다면, 북마크 신호가 전달되었을 때의 당시 화자를 식별하는 것은, 상기 화자가 언제 스피치를 시작하고 중지했는지를 결정하는 것에 의해, 세그먼트 경계들을 식별할 수 있다. 이 정보는, 예를 들어 영화의 세그먼트를 식별하는 것과 같은 특정 종류에 대해서는 유용할 수 있지만 다른 것들에 대해서는 그렇지 않다. 색상 팔레트 신호의 변화와 같은 히스토그램 정보 또한 세그먼트 변화들을 식별하는데 도움이 된다. 자막 및 자연어 처리(natural language processing) 기술들은 그 다음 주제로부터 하나의 주제를 나타내는데 추가적인 정보를 제공하며, 또한 주제들, 대화들 및 기타 등에 기초하여 경계들을 식별하는데도 도움이 된다. 전술한 세그먼트 식별 기술들로부터의 증거를 선택하거나 조합하는 것에 의해서, 세그먼트의 경계들을 결정하고 확립할 수 있게 된다. 또한 세그먼트들을 게다가 식별하기 위해서 전체적으로 프로그램의 구조를 해석하는 것과 이상의 것들을 결합시킬 수도 있다.
- <8> 본 발명의 한 실시예에 있어서, 북마크 신호는 프레임을 식별하며, 세그먼트는 시간, 예를 들어 30초 또는 1분과 같은 시간 또는 예를 들어 선택된 프레임의 전과 후와 같은 선택된 프레임 카운트와 같은 비디오 길이에 기초한다. 대안적으로는, 세그먼트는 미리 규정된 길이, 예를 들어 세그먼트의 시작으로부터 30 초 또는 1 분의 길이로 설정될 수도 있다. 따라서, 북마크 신호가 길이가 긴 세그먼트의 마지막으로 향해서 송신되는 경우는, 세그먼트의 제 1 부분 및 가능하게는 북마크 신호를 이용한 바로 그 부분만이 저장될 수도 있다. 각각의 세그먼트에는 EPG 데이터, 프레임 또는 트랜스크립트 정보 또는 이들의 조합들이 포함될 수도 있다. 세그먼트의 인덱스들은, 예를 들어 인터넷 또는 월드 와이드 웹을 통해서 원격 위치로부터 검사될 수 있고, 이와 같은 인덱스를 검색하는 것에 의해서 비디오들을 선택할 수도 있다.
- <9> 본 발명의 한 실시예에 있어서, 새로운 장면들은 비디오가 시청된다는 기초 하에 검출된다. 북마크 신호가 활성화되는 경우, 시스템은 이후 장면의 마지막을 검색하며, 북마크된 장면을 기록/인덱스하거나 해당 장면을 독립적으로 저장한다.
- <10> 본 발명의 한 실시예에 있어서, 비디오를 시청하고 있는 이용자가 북마크 특징을 활성화하는 경우, 개별적인 프레임의 독특한 특성이 기록된다. 이후, 이용자가 저장 매체 내에 다량의 비디오 콘텐츠를 갖고 있고 북마크된 장면 또는 세그먼트로 되돌아가고 싶어하지만, 영화, 텔레비전 프로그램 또는 스포츠 경기의 식별부를 기억하지 못하는 경우, 독특하거나 상대적으로 독특한 식별자와 같은 프레임의 특성들을 검색하여 장면(또는 전체 작품)을 검색할 수 있다. 따라서 시청자는 원하는 장면을 찾을 때까지 일련의 비디오 북마크들을 스크롤할 수 있게 되며, 해당 장면 또는 작품의 시작 부분으로 직접 갈 수 있다. 이용자들은 비디오뿐만 아니라 음악, 음성 및 다른 저장된 콘텐츠에 대해서도 즐겨 찾는 북마크된 세그먼트의 개인 목록을 보유할 수 있으며, 프레임 식별자 또는 세그먼트 식별자를 콘텐츠 소스로 송신하는 것에 의해서 다양한 인터넷 또는 웹 액세스형 콘텐츠 소스들로부터 콘텐츠에 액세스할 수 있다.
- <11> 본 발명에 따른 북마크들은 원격 장치, 예를 들어 PDA, 또는 다른 컴퓨터화된 저장 장치에 백업될 수도 있다. 이와 같은 장치는 북마크들을 EPG 데이터, 프레임 정보, 트랜스크립트 정보를 해석하는 것에 의해서 또는 키워드 검색을 수행하는 것에 의해서, 또는 다른 비디오 특징들로 카테고리화할 수 있다. 사실 본 발명에 따른 시스템들 및 방법들은 음성 북들, 음악, 라디오 방송 프로그램들, 텍스트 문서들, 멀티미디어 프리젠테이션들, 사진들 또는 다른 이미지들, 등으로부터의 세그먼트와 같은 다양한 종류의 전자적인 콘텐츠를 북마크하고 카테고리화하는 데에도 이용할 수 있다. 또한 서로 다른 레벨들로 북마크들을 저장하여 특정 사생활 및/또는 부모들의 시청 제한 관련 문제점들을 해결할 수도 있다. 본 발명의 특정 실시예들에 있어서, 웹 페이지들, 이동 통신 장치들, PDA들, 시계들 및 다른 전자 장치들을 통해서 북마크들에 액세스할 수 있다.
- <12> 따라서 각각에는 EPG 데이터, 텍스트 관련 데이터 또는 다른 어떤 정보뿐만 아니라 비디오의 다양한 측면에 관한 북마크들을 저장할 수 있다. 이 텍스트 관련 정보는 트랜스크립트, 시놉시스 또는 배우와 관련된 EPG 데이터, 키프레임 등의 일부 또는 전체일 수도 있다. 이 정보는 세그먼트 및 북마크를 특정화하는데에도 이용될 수도 있다.
- <13> 따라서 본 발명의 목적은 기존의 방법들, 시스템들 및 장치들에서의 단점들을 극복하는 비디오 및 다른 콘텐츠

를 북마크하고 검사하기 위한 개선된 방법, 시스템 및 장치를 제공하는 것이다.

실시예

- <19> 본 발명을 더욱 완전하게 설명하기 위해서, 첨부한 도면들과 관련하여 후술하는 발명의 상세한 설명을 참조하도록 한다.
- <20> 종종 시청자는 그들이 시청하고 있는 비디오의 세그먼트를 추후 검색을 위해 북마크하고 싶어한다. 비디오를 북마크하면 관심있는 특정 세그먼트로 훨씬 더 용이하게 되돌아 갈 수 있게 된다. 이용자가 생방송을 시청하거나 테이프, 디스크, DVD, VHS 테이프 또는 다른 것에 저장된 비디오를 시청하는 경우, 이들은 표시점을 입력하기 위해서 버튼을 누르거나 또는 다른 방식으로 비디오에 전자적으로 결합된 장치로 신호를 송신할 수 있다. 이 표시점(또는 프레임 시그니처)은 (제어 영역과 같은) 테이프의 자유 영역 또는 비디오가 기록되는 매체상에 기록될 수 있으며, 또는 테이프의 특정 지점에 대한 시간 또는 프레임 카운트는 독립적인 저장 매체 상에 기록될 수도 있다.
- <21> 도 5는 프로세스 흐름을 도시한 것이다. 입력되는 비디오는 단계(501)에서 프레임들로 나뉜다(포맷). 각각의 프레임들의 다음에 단계(502)에서 시그니처를 개발하고 저장한다. 이용자가 북마크하기 위해서 프레임을 선택한 경우라면, 단계(503)에서 프레임을 식별하고 비디오 정보 및 프레임 위치를 갖는 시그니처가 북마크로서 저장된다. 이후 단계(504)에서 북마크 주위의 경계들이 식별되고, 이들 정보가 저장된다. 단계(505)에서 세그먼트 경계들과 같은 세그먼트 식별 또는 비디오는 이용자에 따라서 저장될 수 있다.
- <22> 본 발명의 한 실시예에 있어서, 이용자는 PDA, 서버 또는 다른 저장 장치에 북마크들을 저장할 수도 있다. 이는 록업 테이플과 같은 역할을 한다. 이용자는 또한 북마크 또는 프레임 정보, 예를 들어 외부 서버에 저장된 비디오의 프레임 정보,를 비교함으로써, 특정 비디오를 시청했는지 또는 얻었는지를 검증할 수도 있다. 시청자는 비디오를 다운로드할 수도 있으며, 시청하고 나서 비디오를 삭제하고 북마크(들)만 보유하다가, 이후 추가로 시청하고 싶어질 때 외부 소스로부터 비디오를 검색할 수도 있다. 따라서 저장 소스들이 극대화될 수 있고, 집중된 콘텐츠 저장 소스들의 효율성을 활용할 수 있다.
- <23> 본 발명의 한 실시예에 있어서, 시청자가 비디오를 클릭하면, 그 시점에 디스플레이되는 프레임이 해석을 위해 추출된다. 시그니처(signature), 히스토그램(histogram), 자막 또는 다른 어떤 로우-레벨의 특징 또는 특징들의 조합이 상기 프레임을 대표할 수 있다. 이하에서 예들이 제공될 것이다.
- <24> 본 발명에 따른 시스템들이 북마크 신호가 활성화된 정확한 지점으로 되돌아가도록 설정될 수도 있지만, 개선된 시스템들 또는 애플리케이션들에 있어서, 비디오의 의미있는 세그먼트를 북마크할 수도 있으며, 이용자들은 세그먼트를 시청한 이후 관심있다고 발견하기까지는 세그먼트를 북마크하지 않겠다고 결정할 수도 있기 때문에, 정확한 지점으로 되돌아가거나 세그먼트의 중간이나 세그먼트의 마지막 부분보다는 의미있는 세그먼트의 시작 부분으로 되돌아 갈 수 있는 옵션을 가질 수도 있다.
- <25> 북마크와 대응하는 세그먼트를 식별하는 것은 다양한 방식으로 달성할 수 있다. 예를 들어, 본 발명의 선호되는 실시예에 있어서, 전체 비디오 또는 그 비디오의 상당한 부분들은 본 발명에 따라서 해석되어 세그먼트들로 나뉘어질 수 있다. 이후, 북마크 신호가 활성화되었을 때, 신호가 활성화되었을 때에 발생하는 세그먼트(또는 이전 세그먼트 또는 둘 다)를 북마크할 수 있게 된다. 본 발명의 다른 실시예에 있어서, 북마크 신호가 활성화되기 전까지는 세그먼트의 경계들을 결정하기 위한 해석이 수행되지 않는다. 이 정보(비디오 시그니처, 테이프의 시작 및 종료 시간, 프레임 카운트 등등)는 위에서 식별된 동일한 위치에 저장될 수 있다.
- <26> 본 발명의 또 다른 실시예에 있어서, 비디오들, 음성, 이미지들, 텍스트 및 이들의 조합들 등등과 같은 콘텐츠의 항목들을 식별하는 방법은, 콘텐츠 항목을 식별하는데 충분한 식별 정보를 갖는 콘텐츠 항목 중의 선택된 세그먼트를 포함하는 북마크를 생성하는 것에 의해서, 또한 예를 들어 서비스 제공자의 저장 매체와 같은 저장 매체 상에서 항목을 식별하는 세그먼트를 보유하는 것에 의해서, 수행될 수 있다. 이용자들은 이후 그들의 선택에 따라서 원격지에서 북마크들을 다운로드할 수 있게 된다. 따라서 이용자들은 북마크가 생성된 콘텐츠의 원래 항목을 식별하는데 북마크들을 이용할 수 있게 된다. 개인 프로파일들에 따라 이들 북마크들의 다운로드들이 이루어질 수 있다.
- <27> DCT 프레임 시그니처들(DCT Frame Signatures)

- <28> 시청자가 프레임을 선택할 때, DCT(이산 코사인 변환, Discrete Cosine Transform) 계수들을 합성하여 한 가지 종류의 프레임 시그니처를 유도할 수 있다. 프레임 시그니처 표시는 프레임 내의 유사하게 가중치가 부가된 DCT 블록들을 각각 그룹화하기 위해서 유도된다, 즉 프레임 시그니처는 프레임 내의 영역 시그니처들로부터 유도된다. 각각의 영역 시그니처는 이하의 단락에서 설명하는 바와 같이 블록 시그니처들로부터 유도된다. 정성적으로는, 프레임 시그니처들은 식별 가능한 오브젝트들을 나타내는 비디오 프레임들 내의 두드러진 영역들에 대한 정보를 포함하고 있다. 이 프레임 시그니처는 이후 비디오의 이 부분을 검색하는데 이용된다.
- <29> 도 3을 참조하면, 블록, 영역 및 프레임 시그니처들을 추출하는 것은 다음과 같이 실행된다. DC 및 AC 계수들의 최대값에 기초하여, 비디오 프레임(302) 내의 각각의 블록(301)에 대해서 시그니처를 유도한다. 이후 영역 시그니처들을 유도하기 위해, 유사한 시그니처들을 갖는 블록들(301)을 비교하여, 블록들(301) 그룹들의 크기 및 위치를 결정한다.
- <30> 블록 시그니처(310)는 8 비트 길이이며, 이 중에서 세 개의 비트들(320)이 DC 시그니처에 이용되며, 다섯 개의 비트들(330)은 AC 값들에 이용된다. 시그니처(310)의 DC 부분(320)은, DC 값이 특정 값들의 범위 내에(예를 들어 -2400 내지 2400)해당하는 범위를 결정하여, 유도된다. 이 범위는 미리 선택된 간격들 숫자로 나누어질 수도 있다. 본 경우 있어서, 여덟 개의 간격들이 이용되었다(여덟 개의 값들은 세 개의 비트들로 표시되었다). 애플리케이션의 종류에 따라서, 전체 시그니처의 크기는 더 큰 숫자의 간격들을 수용하도록 변경될 수도 있으며, 따라서 더 미세한 입도(granularity) 표시를 수용할 수도 있다. 각각의 간격에는 DC 값들의 범위로부터 시그니처의 DC 부분(320)까지 미리 규정된 맵핑이 할당된다. 다섯 개의 비트들(330)이 이용되어, AC 값들의 내용을 나타낸다. 각각의 AC 값은 문턱값(threshold), 예를 들어 200,과 비교하여 이 값이 문턱값보다 더 크다면, AC 시그니처 내의 대응하는 비트를 1로 설정한다. 한 가지 실시예를 도 3에 나타내었으며, 여기에서 값 370만이 200인 문턱값보다 크다.
- <31> 도 3에서 도시한 바와 같이, 다섯 개의 비트들을 이용하여 AC 값들의 콘텐츠를 나타내고 있다. 각각의 AC 값은 문턱값과 비교하여, 이 값이 문턱값보다 크다면, AC 시그니처 내의 대응하는 비트를 1로 설정한다.
- <32> 각각의 프레임에 대한 블록 시그니처들을 유도한 이후, 유사하게 평가된 블록 시그니처들의 영역들이 결정된다. 영역들은 유사한 시그니처들을 공유하는 두 개 또는 그 이상의 블록들로 구성된다. 이 프로세스에 있어서, 영역 성장법(region growing method)이 이미지 내의 영역들을 분리시키는데에 이용될 수도 있다. 통상적으로 영역 성장법은 영역들을 검출하기 위해서 픽셀 색상(pixel color) 및 이웃 개념(neighborhood concepts)을 이용한다. 본 발명의 한 실시예에 있어서, 영역들을 성장시키기 위한 기반으로서는 블록 시그니처를 이용하였다. 따라서 각각의 영역에는 영역 시그니처가 할당되는데, 예를 들면 *regionSignature(mblockSignature, regionSize, Rx, Ry)*와 같으며, 여기에서 *Rx*와 *Ry*는 영역의 중심 좌표이다. 각각의 영역은 대략적으로 이미지 내의 오브젝트에 대응한다.
- <33> 선택된 프레임은 DCT 블록들 중에서 가장 돌출된 그룹들(영역들)로 나타낼 수 있다. *n* 워드 길이의 시그니처는 *n*이 중요한 영역의 숫자(애플리케이션에 의해서 한정됨)를 결정하고 워드는 미리 규정된 바이트 수(number of bytes)로 구성되는 프레임용으로 유도된다. 각각의 프레임은 다수의 돌출된 영역들로 표시된다. 본 발명의 한 실시예에 있어서, 이미지 내의 영역들의 숫자는 제한되고, 가장 큰 영역만이 보유된다. 하나의 프레임이 다수의 영역들로 표시되기 때문에, 프레임들 간의 유사성은, 이들의 블록 시그니처, 크기 및 위치에 기초하여 유사한 영역들의 숫자를 선택하여 제어할 수 있다. 영역들은 영역 크기에 따라서 분류될(sorted) 수 있고, 이후 최상위 *n* 개의 영역 시그니처들이 프레임의 대표값, *frame(regionSignature1, ..., regionSignature_n)*으로서 선택될 수 있다. 이 키프레임들의 대표값은 이미지들의 영상적인 외관에 기초하고, 이미지들에 관한 어떠한 의미론을 기술하고자 시도하지는 않았다는 것이 주의된다.
- <34> 프레임 검색
- <35> 비디오 내에서 위치를 찾아내기 위해서, 프레임 비교 프로세스는 프레임들의 목록에서 북마크된 프레임(F')과 전체 프레임들(F')을 비교한다. 이들 각각의 영역 시그니처들은 그들의 크기에 따라서 비교된다.

$$frame_difference = \sum_{i=1}^n |region_size'_i - region_size_i|$$

- <36>
- <37> 동일한 중심을 구비한 프레임 시그니처 내의 영역들에 대해서 프레임 차이를 계산할 수 있다. 이 경우, 오브젝트들의 위치뿐만 아니라 시그니처값도 고려한다. 한편, 위치가 부적절하고, 영역 크기들만 비교하고 영역들의

위치는 무시하고자 하는 경우들이 있다.

- <38> 프레임 차이가 영이면, 정합 프레임(matching frame)으로부터 위치 정보를 이용하여 비디오의 그 섹션을 검색할 수 있다.
- <39> 기타 프레임 시그니처 종류들
- <40> 이전 및/또는 이후의 프레임 간의 최대 절대값 차이(MAD, Maximum Absolute Difference)와 같은 프레임들로부터의 특징들의 조합을 이용하여 시그니처들을 생성할 수 있다. 프레임의 강도(intensity), 프레임에 이용된 비트레이트(bitrate), 프레임이 비월 주사인지(interlaced) 순차 주사인지(progressed), 프레임이 16:9 포맷 또는 4:3 포맷으로부터인지 등등. 프레임을 식별하기 위해서 상기 종류의 정보는 임의의 조합을 이용하여도 무방하며, 전술한 바와 유사하게 개발된 검색 프로세스를 이용한다.
- <41> 색상 히스토그램들
- <42> 전술한 시그니처들을 이용하는 것 대신에, 프레임에 대한 색상 히스토그램을 계산하고 이를 검색에 이용할 수도 있다. 색상 히스토그램은 어떠한 숫자의 빈들(bins)로 구성되어도 무방하다.
- <43> 자막
- <44> 자막 데이터 역시 세그먼트를 표시하는 키워드를 추출하여 세그먼트를 북마크하는데 이용할 수 있다.
- <45> 조합들
- <46> 또한 전술한 것의 임의의 조합 또한 프레임 또는 섹션을 북마크하는데 이용할 수 있다.
- <47> 세그먼트들을 한정하기
- <48> 세그먼트들을, 시청자가 비디오의 시작 및 종료점을 클릭해서, 시청자가 수동으로 북마크할 수도 있다. 대안적으로 슈퍼히스토그램(superhistogram)과 같은 기술을 이용하여 북마크 작업을 자동화할 수도 있다. 프레임의 경계들을 결정하기 위한 자동화 기술에 대해서는 이하에서 설명하기로 한다. 예를 들어, 어떤 장면은 종종 특정 색상 팔레트(color palette)를 보유한다. 장면의 변화는 대개 색상 팔레트의 브레이크(break)를 수반한다. 히스토그램들을 추출하기 위해서 비디오가 동작 중일 때, 자동 비디오 해석을 수행할 수 있다. 시청자가 비디오를 클릭하면, 그 프레임에 대한 색상 히스토그램을 이미 캡처한 프레임들과 비교하여 프레임의 시작을 식별하고 이후 동일한 비교를 행하여 장면의 마지막을 검색한다. 이제 이 정보를 이용하면 시청자에게 관심있는 세그먼트만을 저장할 수 있게 된다. 이 정보는 또한 전체 비디오의 보다 의미있는 검색용으로 이용될 수도 있다. 예를 들어, 시청자가 클릭한 위치로 직접 가는 것 대신, 실제로는 이 프레임을 포함하는 장면의 시작 부분으로 갈 수도 있다.
- <49> 예
- <50> 시청자는 오즈의 마법사(the Wizard of Oz) 영화를 시청하고 있다. 현재 장면에는 도로시, 양철 나뭇꾼, 겁쟁이 사자와 허수아비가 양귀비밭을 지나서 에메랄드시로 들어가는 프레임들이 포함되어 있다. 시청자가 비디오, 예를 들어 다른 색상의 말이 지나갈 때를 클릭한다. 본 발명의 한 실시예에 있어서, 프레임/장면 해석은 연속적이다. 시스템은 선택된 프레임을 추출하고, 예를 들어 DCT 프레임 시그니처와 함께 색상 히스토그램의 둘 다를 동시에 생성한다. 해석 프로그램은 동일한 색상 팔레트에 속하지 않는 프레임을 찾을 때까지 이미 저장된 프레임들을 검색한다. 이는 장면의 시작을 의미한다. 프로그램은 색상 팔레트에서의 다른 중요한 변화에 의해서 장면의 마지막에 위치할 때까지 비디오를 계속 해석한다. 이용자가 비디오 전체를 기록하기로 했다면, 시작 및 종료 점들이 표시된다. 본 발명의 다른 실시예에 있어서, 단지 특정 세그먼트만이 저장된다. 한편으로 프로그램은 개별적인 프레임들에 대한 DCT 프레임 정보를 해석하고 저장한다. 조금 뒤에, 시청자가 북마크된 프레임을 시청하고 비디오의 그 부분을 검색하기로 결정하는 경우, DCT 프레임 정보가 일치하는 부분이 발견될 때까지 저장된 정보와 비교된다. 이후 비디오의 상기 부분을 검색하는데에 상기 프레임 주위로 표시된 지점들을 이용한다.
- <51> 또한, 비디오의 분할은 참조에 의해서 콘텐츠가 본원에 합체되는 미국 특허 제 6,137,544 호 및 제 6,125,229 호에서 개시된 것과 같은 해석 기술들을 이용하여 수행할 수 있다.
- <52> 또한, 비디오 신호의 분할은 베이저언 엔진 또는 BE(Bayesian Engine)로 지칭되는 층화 확률 시스템(layered probabilistic system)을 이용하여 수행할 수도 있다. 이와 같은 시스템은 제이. 펄(J. Pearl), "Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference", 모간 카우프만 출판

사(Morgan Kaufmann Publishers, Inc), 산 마테오(San Mateo), 캘리포니아 (1988)에 기술되어 있다. 이와 같은 시스템은 도 9를 참조하면 이해할 수 있다.

<53> 도 4는 로우-레벨(410), 미들-레벨(420) 및 하이-레벨(430)의 세 개의 층으로 된 세 개 층 확률 프레임워크(probabilistic framework)를 도시한 것이다. 로우-레벨층(410)은 비디오 신호(401)용의 신호 처리 파라미터들을 설명한다. 이들에는 색상, 에지 및 형태와 같은 영상 특징들, 평균 에너지, 대역폭, 피치(pitch), 멜 주파수 켈스트럴 계수들(mel-frequency cepstral coefficients), 선형 예측 부호화 계수들(linear prediction coding coefficients) 및 제로 크로싱들(zero-crossings)과 같은 음성 파라미터들; 및 자막의 ASCII 문자들로부터 추출할 수 있는 트랜스크립트가 포함된다. 자막 정보를 이용할 수 없는 경우에는 음성 인식법들(voice recognition methods)을 이용하여 음성을 트랜스크립트 문자들로 변환할 수 있다.

<54> 화살표는 미들-레벨(420) 특징들을 생성한 로우-레벨(410) 특징들의 조합들을 나타낸다. 미들-레벨(420) 특징들은 전체 프레임들 또는 프레임들의 집합들과 관련되어 있는 반면에, 로우-레벨(410) 특징들은 픽셀들 또는 단기간 간격들(short time intervals)과 연관되어 있다. 키프레임들(keyframes)(한 화면(short)의 제 1 프레임), 페이스들(faces) 및 비디오텍스트는 미들-레벨 영상 특징들이다. 무음, 잡음, 스피치, 음악 및 이들의 조합들은 미들-레벨(420) 특징들이다. 키워드 및 자막/트랜스크립트 카테고리 또한 미들-레벨(420)의 일부이다.

<55> 하이-레벨 특징들은 서로 다른 양상(modality)들에 걸쳐서 미들-레벨(420) 특징들의 통합을 통해서 얻어지는 미분적인 비디오 콘텐츠를 설명할 수 있다.

<56> 이 액세스법은 매우 적절한 것으로, 그 이유는 확률론적인 프레임워크들이 불확실한 정보를 취급하도록 설계되어 있고, 이들이 정보의 통합을 표시하는데 적절하기 때문이다. BE의 확률론적인 통합은 인트라(intra) 또는 인터(inter) 양상들(modality) 중 하나를 채택한다. 인트라 양상 통합(intra-modality integration)은 단일 도메인 내의 특징들의 통합에 관한 것이다. 예를 들면, 비디오텍스트에 대한 색상, 에지 및 형태 정보의 통합은 인트라 양상 통합을 나타내는 것인데, 그 이유는 이 모든 것들이 영상 도메인을 대신하기 때문이다. 영상 카테고리의 페이스 및 비디오텍스트와 미들-레벨 음성 카테고리의 통합은 인터 양상의 예를 제공한다.

<57> 베이저안 네트워크들은 지향 비순환 그래프(DAG, directed acyclical graph)이며 여기에서 노드는 (확률론적인) 변수들에 대응한다. 아크(arcs)는 링크된 변수들 간의 직접적인 인과 관계를 나타낸다. 이들 링크들의 강도는 조건부 확률 분포들(cpd : conditional probability distributions)로 주어진다. 보다 형식적으로는, N 개의 변수들을 가지는 집합 $A(\underline{1}, \dots, \underline{N})$ 를 DAG라 정의하자. 각각의 변수에 대해서, A 의 변수들로 구성된 하위 집합, $9_{\underline{i}}$ 가 존재하며, \underline{i} 의 부모 집합, 즉 \underline{i} 의 조상은 DAG 내에서 $P(\underline{i}|9_{\underline{i}}) = P(\underline{i}|\underline{1}, \dots, \underline{i-1})$ 이고, 여기에서 $P(\cdot|\cdot)$ 가 cpd이므로 엄밀히 말하자면 양이다. 이제, 결합 확률 밀도 함수(pdf : probability density function)를 $P(\underline{1}, \dots, \underline{N})$ 로 놓고 연쇄 법칙을 이용하면,

$$P(\underline{1}, \dots, \underline{N}) = P(\underline{N}|\underline{N-1}, \dots, \underline{1}) \times \dots \times P(\underline{2}|\underline{1})P(\underline{1})$$

<58> 가 된다. 이 방정식에 따르면 부모 집합인 $9_{\underline{i}}$ 는 \underline{i} 및 $\{\underline{1}, \dots, \underline{N}\} \setminus \underline{N}_{\underline{i}}$ 가 조건부적으로 주어진 $9_{\underline{i}}$ 와는 독립적으로 되는 성질을 가지게 된다.

<59> 도 4에 있어서, BE이 흐름도는 세 개의 층들로 형성되어 있는 DAG의 구조를 갖고있다. 각각의 층에 있어서, 각각의 요소는 DAG 내의 노드에 대응한다. 지향 아크들은 직전 층의 하나 또는 그 이상의 노드들과 주어진 층 내의 하나의 노드와 결합한다. 아크들의 두 개의 집합들은 세 개의 층들의 요소들과 결합한다. 주어진 층에 대해서 및 주어진 요소에 대해서 전술한 바와 같이 결합 pdf(확률 밀도 함수, probability density function)를 계산한다. 보다 상세하게는, 제 1 번째 층과 관련된 임의의 요소(노드) $i^{(1)}$ 에 대해서, 결합 pdf는 다음과 같다.

수학식 1

$$P^{(l)}(c_{i(0)}^{(l)}, g^{(l-1)}, \dots, g^{(2)}) = P(c_{i(0)}^{(l)} | g^{(l)})$$

$$\times \{P(c_{i(1)}^{(l-1)} | g^{(l-1)}) \dots P(c_{N(l-1)}^{(l-1)} | g^{(l-1)})\} \dots$$

$$\times \{P^2_{i_1} | g^2_{i_1}) \dots P^2_{i_2} | g^2_{i_2})\},$$

<60>

<61> 여기에서 각각의 요소 $c_{i(1)}^{(1)}$ 에 대해서 부모 집합 $g_i^{(1)}$ 이 존재하며, 주어진 레벨 l , 즉, 대한 부모 집합들의 합집합은 $g^{(1)(df)};_{i=1}^{N(1)} g_i^{(1)}$ 이 된다. 각각의 레벨에 대해서 서로 다른 부모 집합들 간에 중첩이 있을 수도 있다.

<62> BE에 의해서 수행된 토픽 분할(및 분류)은 도 4의 제 3 층(하이-레벨)에 도시되어 있다. 멀티미디어 콘텐츠의 복잡한 특성은 다중 도메인들에 걸친 통합을 필요로 한다. 음성, 영상 및 트랜스크립트 도메인들로부터 포괄적인 데이터 집합을 이용하는 것이 유리하다.

<63> 도 4의 BE 구조에 있어서, 각각의 삼 층들에 대해서, 각각의 노드 및 화살표는 cpd와 관련되어 있다. 로우-레벨 층에 있어서 cpd의 것들은 전술한 바와 같이 AE에 의해서 할당된다. 미들-레벨의 층에 대해서, 날씨, 국제, 범죄, 스포츠, 영화, 패션, 기술 주식, 음악, 자동차, 전쟁, 경제, 에너지, 주식, 폭동, 재정, 국내(사건들), 바이오 테크, 재난, 예술 및 정치와 같은 20 개의 (예를 들면) 자막 카테고리들이 생성된다. 키워드들 및 카테고리들이 연관 테이블로 형성된 각각의 카테고리에 대해서 지식 트리(tree)를 이용하는 것이 유리하다. 통계적인 처리를 한 후, 시스템은 카테고리 투표 히스토그램들(category vote histograms)을 이용하여 카테고리화를 수행한다. 자막들 파일 내의 한 단어가 지식 기반 키워드와 일치하는 경우라면, 이후 대응하는 카테고리는 한 표를 얻게 된다. 각각의 카테고리에 대한 확률은 키워드당 전체 투표수들 및 자막들의 문장에 대한 전체 투표수들 사이의 비율에 의해서 주어진다.

<64> 본 발명에 따른 시스템은 비상업적인 부분들을, 예를 들어 (BE에 의해서 수행되는) 재정 뉴스와 토크 쇼들의 두 가지 하이-레벨 카테고리들에 기초하여 세그먼트들로 분류하는 것과 같이, TV 프로그램을 상업적인 부분과 비상업적인 부분들로 분할하는 분할을 수행할 수 있다.

<65> 초기 분할은 자막 데이터를 이용하여 비디오를 프로그램과 상업적인 세그먼트들로 나눌 수 있다. 다음에 단일, 이중, 및 삼중 화살표들에 대해서 프로그램 세그먼트들의 자막들을 해석한다. 이중 화살표는 화자의 변경을 나타낸다. 시스템은 연속적인 이중 화살표들 사이의 텍스트를 시작 시간과 종료 시간으로 표시하여, 이를 원자화된 자막들 유닛(atomic closed captions unit)으로 이용한다. 본 발명에 따른 시스템들은 이들 유닛들을 분할 구축 블록들(segmenting building blocks)로 이용할 수 있다. (예를 들어 재정 뉴스이든 토크 쇼이든지 간에) 임의 세그먼트의 하이-레벨 인덱스를 결정하기 위해서 스카우트(Scout)가 두 개의 결합 확률(joint probabilities)을 계산한다. 이들은 다음과 같이 정의된다.

수학식 2

$$p\text{-FIN-TOPIC} = p\text{-VTEXT} * p\text{-KWORDS} * p\text{-FACE} *$$

$$p\text{-AUDIO-FIN} * p\text{-CC-FIN} * p\text{-FACETEXT-FIN}$$

<66>

수학식 3

$$p\text{-TALK-TOPIC} = p\text{-VTEXT} * p\text{-KEYWORDS} * p\text{-FACE} * p\text{-AUDIO-TALK} * p\text{-CC-TALK} * p\text{-FACETEXT-TALK}$$

<67>

<68>

재정 뉴스에 대한 p-AUDIO-FIN 및 토크 쇼들에 대한 p-AUDIO-TALK의 음성 확률들은 서로 다른 개별적인 음성 카테고리 확률들의 조합에 의해서 생성된다. 재정 뉴스에 대한 p-CC-FIN 및 토크 쇼들에 대한 p-CC-TALK의 자막 확률들은 20 개의 확률들 목록 중에서 최대의 확률이 되도록 선택된다. p-FACETEXT-FIN 및 p-FACETEXT-TALK의 페이스 및 비디오텍스트 확률들은, 각각의 개별적인 자막 유닛에 대해서, 페이스 및 텍스트 발생의 확률을 결정하는 p-FACE와 p-TEXT의 페이스 및 비디오텍스트 확률들과 비교하여 얻는다. 하나의 발견적인 이용법은 토크 쇼들이 페이스들에 의해서 좌우되는 것에 반해서 재정 뉴스는 페이스들 및 텍스트의 양자를 이용한다는 사실에 기초한다. 하이-레벨 분할은 p-FIN-TOPIC 및 p-TALK-TOPIC의 새로운 쌍의 확률들을 계산하는 것에 의해서 각각의 자막에 행해진다. 최고값은 세그먼트의 분류가 재정 뉴스인지 아니면 토크쇼인지를 나타낸다.

<69>

세그먼트의 경계들을 식별하는데 이용할 수 있는 다른 유용한 정보를 추출하기 위해서 비디오 신호를 전처리할 수도 있다. 예를 들어, 신호를 음성, 영상 및 트랜스크립트 도메인들로 나누어 비디오 세그먼트들로부터 의미론적으로 인덱스된 정보를 생성할 수도 있다. 도메인들의 콘텐츠를 해석하며, 세그먼트의 경계들을 보다 정확하게 식별하기 위해서 해석이 결합된다.

<70>

비디오의 전처리는, 해석과 비디오 전처리를 결합시킨 해석 엔진(AE, Analysis Engine)을 이용하여 달성될 수 있다. AE는, 예를 들어 MPEG-2 입력을 받아들일 수 있으며, 자막(cc, closed caption) 정보를 추출할 수 있다. 또한 AE는 이하에서 설명하는 바와 같이 부가적인 특징을 추출하기 위해서 음성 및 영상 해석을 수행할 수 있다. 이 정보는 이후 세그먼트 경계들의 식별에 결합된다.

<71>

선호되는 AE 중의 하나는 필립스(Philips) 상표의 트리미디어(TriMedia™) 트리오덱 카드(Triodec card)가 있다. 이 카드는 TM 1000 처리기와 8 Mb의 메모리를 갖추고 있다. WinNT 운영 체제를 구비한 600 MHz PIII 컴퓨터에서 호스트되는 모드를 이용하여 동작시킬 수도 있다. 해석이 수행되는 트리미디어로 "C" 코드를 다운로드할 수도 있다.

<72>

AE는 먼저 그림들의 그룹(GOP, group of picture) 중의 연속적인 I-프레임들(중간 프레임들) 간의 현저한 차이점을 검출했을 때 새로운 키프레임을 추출하여 화면 검출을 수행한다. AE는 히스토그램과 매크로블록(macroblock)의 프레임을 "차이나게 하는" 프레임의 구현을 바탕으로 두 개의 DCT를 채택할 수도 있다. 유니컬러 키프레임들 또는 이미 추출된 키프레임들과 유사한 것으로 보이는 프레임들은 1 바이트 프레임 시그니처를 이용하여 필터링 될 수 있다. 이 키프레임 추출은 비압축 이미지와 키프레임 번호 및 장면의 확률이 부여된 목록을 생성한다. AE는 연속적인 I-프레임들 간의 차이점들을 이용하여 문턱값 이상의 상대적인 양에 대한 확률에 기초한다. 시스템은 이후 키프레임들을 비디오텍스트 검출용으로 전달한다. AE는 비디오텍스트의 이용 및 비압축 이미지에 대한 예지 기반의 방법을 검색하며, 텍스트의 존재 및 비존재에 대해서 각각의 프레임에 태그를 달 수 있다. 이들 키프레임들은 페이스들의 존재에 관해서 해석될 수 있다.

<73>

추출 프로세스는 완전한 타임 스탬프된 프로그램 트랜스크립트(a complete time-stamped program trascript)를 생성할 수 있다. 관련된 비디오와 트랜스크립트 데이터를 정렬하기 위해서 타임 스탬프들을 이용할 수 있다.

<74>

다중의, 예를 들어 20 개의 로우-레벨 음성 파라미터들을 비디오 신호의 음성 부분에서 추출하여 분할 프로세스에서의 해석에 이용할 수도 있다. 이는, 예를 들어 PC 상에서 .wav 파일들을 이용하여 달성된다. 비디오 전처리의 출력물은 이후 분할 프로세스에서 이용이 가능하다.

<75>

비디오 신호의 분할에 대한 추가적인 방법들 및 시스템들은 이하에 기술되어 있으며, 참조에 의해서 전체 개시물들이 본원에 합체되는 1999년 1월 캘리포니아 산 호세에서의 Proc. of SPIE Conf. on Storage and Retrieval for Image and Video Databases의 pp. 243-251에서 T. McGee 및 N. Dimitrova의 "Parsing TV Programs For Identification and Removal of Non-Story Segments"와, 1999년 9월 보스턴에서의 SPIE Conference on Multimedia Storage and Archiving Systems IV의 pp. 2-10에서 N. Dimitrova, H. Elenbaas 및 T. McGee의 "PNRS-Personal News Reretrieval System"과, 1995년 가을 AAAI Symposium on Computational Models for Integrating Language and Vision에서 A. Hauptmann 및 M. Smith의 "Text, Speech, and Vision For Video

Segmentation: The Informedia Project"에 기재되어 있다.

- <76> 이하의 기술들 또한 세그먼트 경계들을 식별하기에 유용한 정보를 얻는데 이용될 수 있다.
- <77> 장면 검출(cut detection) : 여기에서는 두 개의 연속적인 비디오 프레임들을 비교하여 돌연한 장면 변화들(하드 커트들(hard cuts))인지 소프트 변환(화면의 디졸브(dissolve), 페이드 인(fade-in) 및 페이드 아웃(fade-out))인지를 식별한다. 장면 검출의 설명은 참조에 의해서 전체 개시물이 본원에 합체되는 1997년 Proc. ACM Conf. on Knowledge and Information Management의 pp. 113-120에서 제목이 "Video Keyframe Extraction and Filtering: A Keyframe is Not a Keyframe to Everyone"인 N. Dimitrova, T. McGee, H. Elenbaas의 간행물에 제공되어 있다.
- <78> 페이스 검출(face detection) : 여기에서는 스킨 톤(skin tone)을 포함하고 달걀 모양의 형상에 대응하는 비디오 프레임들 중의 영역들이 식별된다. 페이스 검출의 설명은 참조에 의해서 전체 개시물이 본원에 합체되는 1999년 11월, Pattern Recognition Letters의 Vol. 20, No. 11에서 제목이 "Face Detection for Image Annotation"인 Gang Wei 및 Ishwar K. Sethi에 의한 간행물에 제공되어 있다.
- <79> 텍스트 검출(text detection) : 여기에서는 오버레이된 또는 중첩된 텍스트와 같이 비디오 프레임 내에 나타나는 텍스트가 식별된다. 텍스트 검출에 대한 설명은 참조에 의해서 전체 개시물이 본원에 합체되는 IEEE Conference on Computer Vision and Pattern Recognition 1999와 연계하여 개최된 1999년 6월 콜로라도의 포트 콜린스(Fort Collins)에서의 Proceedigns of IEEE Workshop on CBAIVL에서 L. Agnihotri 및 N. Dimitrova의 제목이 "Text Detectoin in Video Segments"에서 제공되어 있다. 본 발명의 선호되는 실시예에 있어서, 일단 텍스트가 검출되면, 본 기술 분야에서 공지된 광학 문자 인식 기술(OCR : optical character recognition)을 채택하였다.
- <80> 삭제
- <81> 움직임 추정/분할/검출(motion estimation/segmentation/detection) : 여기에서는 움직이는 물체들이 비디오 시퀀스들 내에 있는지가 결정되고, 이동 중인 물체의 궤도가 해석된다. 비디오 시퀀스들 내에서 물체들의 움직임을 결정하기 위해서는, 광학 유동 추정법(optical flow estimation), 움직임 보상(motion compensation) 및 움직임 분할(motion segmentation)과 같은 공지된 오퍼레이션들 채택하는 것이 유리하다. 움직임 추정/분할/검출에 대한 설명은 참조에 의해서 전체 개시물이 본원에 합체되는 1993년 봄, International Journal of Computer Vision, Vol. 10, No. 2, pp. 157-182에서 Patrick Bouthemy 및 Francois Edouard의 제목이 "Motion Segmentation and Qualitative Dinamic Scene Analysiss from an Image Sequence"인 간행물에 제공되어 있다.
- <82> 카메라 움직임(camera motion) : 여기에서는 다섯(5) 개 집합의 글로벌 카메라 파라미터들이 채택되며, 유리하게는 두(2) 개의 병진 및 세(3) 개의 회전 파라미터가 채택된다. 3-D 카메라 움직임은 이후 순수 추적(수평 방향의 움직임), 부밍(수직 방향의 움직임, booming), 돌링(깊이 방향으로의 움직임, dollying), 페닝(수직 글로벌 축에 대한 회전, panning), 틸팅(수평 축에 대한 회전, tilting) 및 롤링(z 축에 대한 회전, rolling) 또는 이들 움직임들의 조합들로 분류된다. 이 정보는 예를 들어 "정지(static)", "줌(zoom)" 및/또는 "스팬(span)"으로 비디오 화면을 분류하는데 이용될 수 있으며, 게다가 화면을 생성하고자 하는 감독의 의도를 결정하는데 이용될 수도 있다. 카메라 움직임 검출에 대한 설명은 참조에 의해서 전체 개시물이 본원에 합체되는 1994년, IEEE Transaction on PAMI, Vol. 6, number 1, pp. 13-27에서의 R. Y. Tsai 및 T. S. Huang의 제목이 "Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces"인 간행물에 제공되어 있다.
- <83> 음성 분할에는 다음과 같은 종류들의 텔레비전 프로그램들의 해석이 포함된다. 스피치의 텍스트로의 전환, 음성 효과들 및 이벤트 검출, 화자 식별, 프로그램 식별, 음악 분류 및 화자 식별에 기초한 대화 검출. 음성 분할에는 음성 신호를 스피치 및 비-스피치 부분들로 나누는 것이 포함될 수도 있다. 음성 분할의 제 1 단계는 대역폭, 에너지 및 피치 등과 같은 로우-레벨의 음성 특징들을 이용한 세그먼트 분류를 포함할 수 있다. 그 후에 동시에 발생하는 음성 성분들(예를 들어 음악 및 스피치)을 분리하여 각각 독립적으로 해석할 수 있도록 채널 분리를 채택할 수도 있다.
- <84> 이후, 텔레비전 프로그램의 음성 부분은 스피치의 텍스트로의 변환, 음성 효과들 및 이벤트 검출 및 화자 식별과 같은 서로 다른 방식으로 처리한다. 음성 분할은 본 기술 분야에서 공지되어 있으며, 참조에 의해서 전체 개시물이 본원에 합체되는 1996년 가을, IEEE Multimedia, pp. 27-36에서의 E. Wold 및 T. Blum의 제목이

"Content-Based Classification, Search, and Retrieval of Audio"인 간행물에서 대략 설명되어 있다.

- <85> 텔레비전 신호의 음성 부분 중의 스피치 세그먼트들을 배경 잡음 또는 음악으로부터 식별하거나 분리하기만 하면, 스피치의 텍스트로의 변환(본 기술 분야에서 공지되어 있으며, 예를 들어 참조에 의해서 전체 개시물이 본원에 합체되는 1998년 2월 8일-11일, 버지니아, DARPA Broadcast News Transcription and Understanding Workshop에서 P. Beyerlein, X. Aubert, R. Haeb-umbach, D. Klakow, M. Ulrich, A. Wendmuth 및 P. Wilcox의 제목이 "Automatic Transcription of English Broadcast News"인 간행물 참조)이 채택될 수 있다. 자막이 텔레비전 프로그램의 음성 부분의 트랜스크립트를 제공할 수 없는 경우, 스피치의 텍스트로의 변환이 중요하다.
- <86> 음성 효과들을 이용하여 이벤트들을 검출(본 기술 분야에서 공지되어 있으며, 예를 들어 참조에 의해서 전체 개시물이 본원에 합체되는 1997년 캘리포니아 멘로 파크(Menlo Park)의 AAAI Press의 Intelligent Multimedia Information Retrieval, pp. 113-135에서 T. Blum, D. Keislar, J. Wheaton 및 E. Wold의 제목이 "Audio Databases with Content-Based Retrieval"인 간행물 참조)할 수도 있다. 이벤트들은 특정 이벤트들과 연관된 사운드들을 식별하는 것에 의해서 검출될 수도 있다. 예를 들어, 아나운서가 스포츠 경기에서 "골(goal)"이라고 외치는 것은 세그먼트를 식별하는데 있어서 중요한 부분으로 검출될 수 있다.
- <87> 화자 식별(본 기술 분야에서 공지되어 있으며, 예를 들어 참조에 의해서 전체 개시물이 본원에 합체되는 1997년 2월, 캘리포니아, 산 호세에서의 IS&T SPIE Proceedings: Storage and Retrieval for Image and Video Databases V, pp. 218-225에서 Nilesh V. Patel 및 Ishwar K. Sethi의 제목이 "Video Classification Using Speaker Identification"인 간행물 참조)에는 음성 신호 내에 존재하는 스피치의 성문을 해석하여 스피치 중인 사람의 신원을 결정하는 것을 포함할 수도 있다. 예를 들어 언제 화자가 변경되는지를 결정하고 세그먼트 경계를 식별하는데 도움을 주도록, 화자 식별을 이용할 수도 있다.
- <88> 프로그램 식별에는 텔레비전 프로그램을 식별하기 위해서 음성/데이터/영상 신호 중의 음성 부분을 해석하는 것이 포함된다. 이는, 세그먼트 경계들을 표시함에 있어서, 프로그램의 종류에 따라서 이들 경계들을 다르게 설정할 수 있도록 결정하기 위한 표준이기 때문에, 특히 유용하다.
- <89> 음악 분류에는 음성 신호 중의 비-스피치 부분을 해석하여 어떤 종류(클래식, 락, 재즈 등)의 음악이 존재하는지를 결정하는 것이 포함된다. 이는, 예를 들어 음성 신호 중의 비-스피치 부분의 주파수, 피치, 음색, 사운드 및 멜로디를 해석하고 해석의 결과들을 알려진 특징들을 갖는 특정 종류들의 음악과 비교하여 달성된다. 음악 분류는 공지된 기술이며, 1999년 10월 17일 ~ 20일, 뉴욕, 뉴 팔츠(New Paltz)에서의 1999 IEEE Workshop on Application of Signal Processing to Audio and Acoustics에서 Eric D. Scheirer의 제목이 "Towards Music Understanding Without Separation: Segmenting Music with Correlogram Comodulation"인 간행물에서 대략적으로 제공되어 있다.
- <90> 음성 및 비디오 신호들의 분할 이후, 분할된 음성 및 비디오 신호들의 다양한 부분들을 결합(통합)시킬 수 있다.
- <91> 이하의 비제한적인 실시예를 이용하여 세그먼트들의 인덱스를 생성할 수 있다. 이후 이용자가 기존의 테이프(또는 파일, DVD, 디스크들, 등, 일반적으로 테이프들로 칭함)를 시청할 때 또는 신규 테이프에 기록하는 중에 이들 세그먼트들 중의 하나 내의 북마크 신호를 활성화하는 경우, 상기 세그먼트가 선택된다. 세그먼트들은 미리 설정되어 있거나 비디오를 시청하면서 설정될 수도 있다. 장면의 시작을 식별할 수 있으며, 북마크 활성화 신호가 활성화된 경우는 장면의 마지막을 식별할 수 있으며, 상기 장면이(세그먼트가) 인덱스된다. 예를 들어 인덱스를 저장하는데 테이프의 시작 부분을 이용할 수 있다. 대안적으로는 세그먼트들의 식별은 관련된 파일에 저장될 수도 있다. 본 실시예에 대해서는, 30 초의 "빈(blank)" 또는 재기록 가능한 테이프가 바람직하다. 파일에 대해서는, 영상 인덱스용으로 선택되는 영역은 파일 내의 아무 위치나 가능하며, 시스템에 의해서 자동으로 예약되거나 이용자에 의해서 수동적으로 선택될 수도 있다. 인덱스에는 영상 이미지들, 음성, 텍스트 또는 이들의 임의의 조합이 포함될 수도 있다. 본 실시예용으로 영상 이미지들 및 텍스트를 제공하였다.
- <92> 아카이브 프로세스(archival process) 중에 비디오 콘텐츠가 해석되며, 비디오 해석 프로세스 중에 영상 인덱스가 생성된다. 비디오 해석 프로세스에서는, 자동으로 중요 장면이 검출되고 키프레임이 선택된다. 중요 장면 검출은 장면 변화, 즉 "장면들(cuts)"(비디오 장면 검출 또는 분할 검출)을 식별하고 정지 장면을 식별(정지 장면 검출)하는 프로세스일 수도 있다. 각각의 장면에 대해서, 키프레임으로 지칭되는 특정 대표 프레임이 추출된다. 비디오 테이프와 같은 소스 비디오(source video)의 각각의 키프레임에 키프레임 필터링 및 선택 프로세스를 적용하여 선택적으로 선택된 키프레임들로부터 영상 인덱스를 생성한다. 분명히 소스 테이프를 참조하고 있지만,

소스 비디오는 파일, 디스크, DVD, 기타 저장 수단들 또는 전송원으로부터(예를 들어 홈 비디오를 기록하는 중에) 직접 가져올 수도 있다.

- <93> 비디오 테이프의 인덱스 작업에 있어서, 인덱스는 일반적으로 소스 테이프에 저장된다. MPEG1, MPEG2, MPEG4, Motion JPEG 파일, 또는 비디오 CD, DVD, 또는 다른 저장 장치로부터 또는 방송 스트림으로부터 다른 비디오 파일의 비디오 인덱스 작업에 있어서, 인덱스는 하드 디스크, 또는 다른 저장 매체 상에 저장될 수도 있다.
- <94> MPEG 파일과 같은 이전에 저장된 영상 정보를 이용하여 다른 아카이브 장치들에 대해서 유사한 프로세스를 계속할 수도 있지만, 음성 및/또는 텍스트를 포함하고 있을 수도 있는 이전에 기록된 소스 비디오를 이용하여 소스 테이프에 대한 비디오 기록 프로세스가 도 1에 도시되어 있다. 이러한 프로세스에 있어서, 영상 인덱스는 소스 비디오에 기초하여 생성된다. 이용자가 기록하고자 하는 소스 테이프에 대한 제 2 프로세스는 기록과 동시에 영상 인덱스를 생성한다.
- <95> 도 1은 비디오 테이프에 대한 (이전에 기록된 소스 테이프에 대한) 제 1 프로세스의 일례를 도시한 것이다. 단계(101)에서, 필요하다면, VCR과 같은 재생/기록 장치를 이용하여 소스 비디오를 되감는다. 단계(102)에서, 소스 비디오가 재생된다. 소스 비디오로부터의 신호들은 텔레비전, VCR 또는 다른 처리 장치에서 수신된다. 단계(103)에서, 처리 장치 내의 미디어 처리기 또는 외부 처리기는 비디오 신호들을 수신하고, 이 비디오 신호들을 픽셀 데이터를 나타내는 프레임들로 포맷한다(프레임 그래빙 : frame grabbing).
- <96> 단계(104)에서, 호스트 처리기는 각각의 프레임을 블록들로 분리하고, 이 블록들 및 이들과 관련된 데이터를 변환하여 DCT(discrete cosine transform) 계수들을 생성하고; 중요 장면 검출 및 키프레임 선택을 수행하고; 및 데이터 구조로서 키프레임들을 메모리, 디스크 또는 다른 저장 매체 내에 생성하고 저장한다. 단계(105)에서, 시작 위치로 소스 테이프를 다시 감고, 단계(106)에서, 소스 테이프는 정보를 기록할 수 있게 설정된다. 단계(107)에서, 메모리로부터 소스 테이프를 시그니처를 전송하고, 영상 인덱스를 생성한다. 이후 영상 인덱스를 시청할 수 있도록 테이프를 다시 감을 수도 있다.
- <97> 기록 중 이용자가 테이프에 영상 인덱스를 생성하고자 원하는 경우, 위의 프로세스가 약간 변경된다. 단계들(101 및 102) 대신에, 도 1의 단계(112)에서 도시한 바와 같이, 비디오(필름 등)가 기록됨에 따라서 단계(103)의 프레임 그래빙 프로세스가 진행된다. 게다가, 테이프, 또는 파일이 한번에 완전하게 기록되지 않은 경우 부분적으로 생성된 비디오 인덱스를 테이프, 파일 등에 저장하거나, 나중에 추가하기 위해서 테이프 메모리 내에 저장할 수도 있다.
- <98> 단계들(103 및 104)은 도 2a 및 도 2b에 보다 더 상세하게 도시되어 있다. 비디오는 아날로그(연속 데이터) 또는 디지털(이산 데이터) 형태 중의 하나로 표시될 수 있다. 본 실시예에는 디지털 영역에서 동작하며, 따라서 디지털 형태를 이용하여 처리하며, 아날로그를 디지털로 변환하는 변환기를 포함할 수도 있다. 따라서 소스 비디오 또는 비디오 신호는 충분히 높은 속도로 디스플레이되는 일련의 개별적인 이미지들 또는 비디오 프레임들(본 실시예에서는 초당 30 프레임들)이므로 디스플레이된 연속적인 이미지들은 연속적인 그림 스트림으로 나타나게 된다. 이들 비디오 프레임들은 비압축 (NTSC 또는 로 비디오(raw video))이거나 MPEG, MPEG2, MPEG4, Motion JPEG 등과 같은 포맷에서 압축 데이터일 수도 있다.
- <99> 비압축된 비디오 내의 정보는 Intel.RTM. Smart Video Recorder III에서 표시되는 것과 같은 프레임 그래빙 기술을 이용하여 미디어 처리기(202)에서 먼저 프레임들로 분할된다. 다른 프레임 크기들을 이용할 수 있음에도 불구하고, 본 실시예에서의 프레임은 하나의 텔레비전, 비디오, 또는 다른 영상 이미지를 나타내며, 352 X 240 픽셀들을 포함한다.
- <100> 본 실시예에 있어서 프레임들은 호스트 처리기(210)에서 8 X 8 픽셀들의 블록들로 나누어진다(도 2a). 이들 블록들 및 널리 보급된 방송 표준인 CCIR-601을 이용하여, 매크로블록 생성자(206)(도 2a)는 루미넌스 블록들(luminance blocks)을 생성할 수 있으며, 색상 정보를 평균화하여 크로미넌스 블록들(chrominance blocks)을 생성한다. 루미넌스 및 크로미넌스 블록들은 매크로블록을 형성한다. 본 실시예에 있어서는 4:2:0을 이용하였지만 당업자는 다른 포맷들, 예를 들어 4:1:1 및 4:2:2 등을 이용할 수도 있다. 4:2:0에서, 매크로블록은 여섯 개의 블록들, 네 개의 루미넌스(Y1, Y2, Y3 및 Y4)와, 매크로블록 내에서 각각의 블록이 8 X 8 픽셀들인 두 개의 크로미넌스(Cr 및 Cb)를 구비한다.
- <101> 또한 비디오 신호는 Motion JPEG(Joint Photographic Experts Group) 및 MPEG(Motion Pictures Experts Group)와 같은 압축 표준을 이용하여 압축 이미지를 나타낼 수도 있다. 신호가 MPEG가 아니거나 다른 압축된 신호인 경우, 도 2b에 도시한 바와 같이 MPEG 신호는 프레임 파서(frame parser)(205)에 의해서 프레임 또는 비트

스트림 파싱 기술을 이용하여 프레임들로 나누어진다. 이후 프레임들은 미디어 처리기(203) 내의 엔트로피 디코더(entropy decoder)(214)로 및 테이블 지정자(table specifier)(216)로 전송된다. 엔트로피 디코더(214)는, 예를 들어 허프만 디코딩(Huffman decoding) 또는 다른 디코딩 기술을 이용하여, 테이블 지정자(216)로부터의 데이터를 이용하여 MPEG 신호를 복호화한다.

<102> 복호화된 신호는 이후 테이블 지시자(216)로부터의 데이터를 이용하여 복호화된 신호를 역양자화하는(dequantize) 디퀀타이저(dequantizer)(218)로 공급된다. 도면에서는 미디어 처리기(203)에서 수행되는 것으로 도시했지만, 이들 단계들(단계들(214-218))은 미디어 처리기(203), 호스트 처리기(211) 또는 심지어 이용된 장치들에 의존하는 다른 외부 장치에서도 수행할 수 있다.

<103> 본 발명의 방법을 이용하여 DCT(Discrete Cosine Transform) 계수를 비교할 수 있다. 먼저, 각각의 수신된 프레임을 호스트 처리기(210)에서 개별적으로 처리하여 매크로블록들을 생성한다. 호스트 처리기(210)는 공간 정보를 포함하는 매크로블록을 처리하고, 이산 코사인 변환기(220)를 이용하여 DCT 계수들을 추출하고, 예를 들면 DCT 계수들로 형성된 여섯 개의 8 X 8 블록들을 생성한다. 또한 호스트 처리기(210)는 중요 장면 처리기(a significant scene processor)(230)를 이용하여 각각의 매크로블록들을 처리한다. 본 발명에서는 키프레임 필터링 방법을 이용하여 반복적인 프레임들과 선택된 다른 종류의 프레임들을 필터링하여 프레임 메모리(234) 내에 저장된 키프레임들의 숫자를 감소시킬 수 있다. 키프레임 필터링은 충분한 장면 검출후에 호스트 처리기(210)의 키프레임 필터기(240)에 의해 실행될 수 있다. 이후 정보는 인덱스 메모리(260)로 출력된다.

<104> 대안적으로는, 시스템이 서로 다른 처리 단계들에 액세스를 허용하는 인코딩을 할 수 있는 경우(예를 들어, 미디어 처리기 내에서)에는, 호스트 처리기로 DCT 계수들을 직접 전달할 수도 있다. 이들 모든 액세스법들에 있어서, 처리는 최대 실시간으로 수행된다.

<105> 도 1의 단계(104)에서, 예를 들어, **Intel® Pentium®** 칩 또는 다른 멀티 처리기, **Philips® Trimedia™** 칩 또는 다른 멀티미디어 처리기; 컴퓨터; 개선된 VCR, 기록/재생 장치, 또는 텔레비전; 또는 다른 처리기일 수도 있는 호스트 처리기(210)는 중요 장면의 검출, 키프레임의 선택 및 예를 들어 하드 디스크, 파일, 테이프, DVD 또는 다른 저장 매체와 같은 인덱스 메모리 내에 데이터 구조를 형성하고 저장한다.

<106> 비디오 검색

<107> 일단 비디오 테이프 또는 파일이 영상 인덱스를 구비하게 되면, 이용자가 영상 인덱스로 액세스를 원할 수도 있다. 비디오 검색 프로세스는 이용자에게 이용 가능한 형태로 영상 인덱스를 디스플레이해 준다. 이용자는 영상 인덱스를 둘러보고 향해할 수 있으며, 소스 테이프 또는 MPEG 파일 상의 선택된 지점으로 빨리 감기가 된다.

<108> 소스 비디오는, 예를 들어, VCR 또는 재생 장치에 의해서, 필요하다면, 한 실시예에서 테이프의 시작 부분의 영상 인덱스 위치로 되감긴다. 소스 비디오가 MPEG 파일 또는 디스크 상에 있는 경우라면, 포인터가 저장 위치의 시작 부분을 지시할 수 있으며, 되감을 필요가 없게 된다. 유사하게, 다른 저장 수단들도 적절히 영상 인덱스의 시작 부분을 설정하도록 할 수 있다.

<109> 영상 인덱스는 VCR 헤드, 컴퓨터, 또는 다른 하드웨어 장치에 의해서 소스 비디오로부터 관독되어, 심지어 PDA, 팜톱 컴퓨터, 휴대폰 또는 다른 무선 저장 장치와 같은 어떠한 종류의 저장 장치라도 무방한 인덱스 메모리에 저장된다. VCR 내의, 예를 들어, 처리기는 소스 비디오의 인덱스 메모리로부터 키프레임들을 검색할 수 있다. 검색된 키프레임들은, 다른 프레임 크기들을 자동으로 또는 이용자에 의해서 수동적으로 용이하게 선택할 수 있음에도 불구하고, 크기를 예를 들어, 120 X 80 픽셀들로 축소하도록, 처리된다.

처리된 프레임들은, 처리된 키프레임들을 디스플레이 메모리에 기록하고, 이들을 컴퓨터 디스플레이, 텔레비전 스크린 등과 같은 이용자 인터페이스에 디스플레이하는, 호스트 처리기로 전송된다.

<110> 삭제

<111> 일단 비디오 인덱스가 관독되면 소스 비디오를 정지시킬 수 있다. 비디오 인덱스 시스템 또는 소프트웨어는 키프레임들을, 예를 들어 컴퓨터 모니터 또는 텔레비전 스크린과 같은 디스플레이 상에 디스플레이 되도록 한다. 영상 인덱스는 이용자가 원하는 경우 출력될 수 있다. 또한 이용자는 영상 인덱스로부터 특정 키프레임을 선택할 수도 있다. 이용자가 상기 특정 키프레임에서의 소스 비디오를 시청하고자 원한다면, 소스 매체는 키프레임이 추출된 대응하는 지점으로 자동으로 되감기게 되어, 소스를 재생할 수 있게 된다. 대안적으로는, 카운터를

표시하여 이용자로 하여금 소스 비디오를 빨리 되감도록 하거나, 영상 인덱스부터 선택된 키프레임까지 소스 비디오 재생이 가능하게 할 수도 있다.

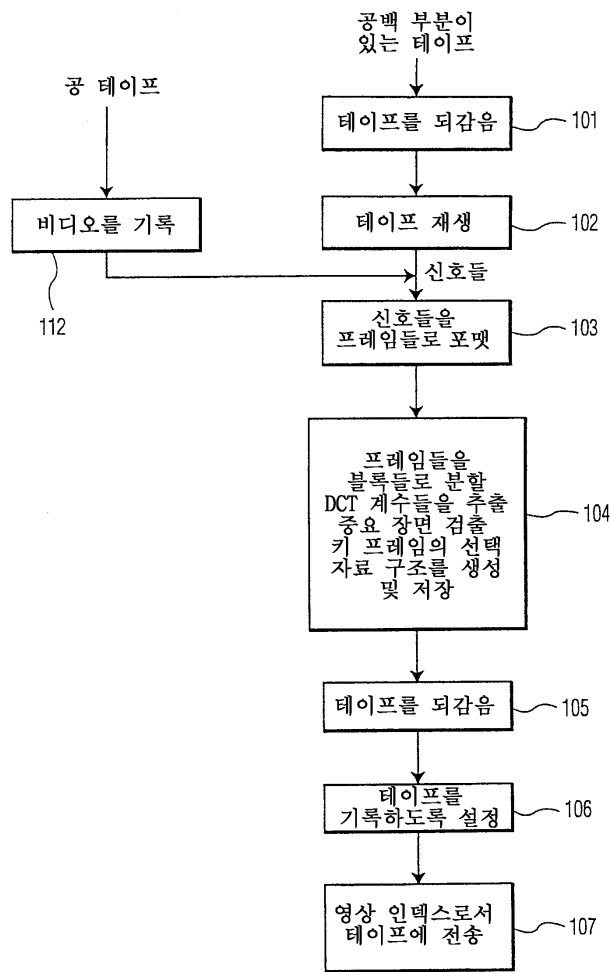
- <112> 본 발명의 부가적인 특징으로는, 이용자가 임의의 지점에서 비디오 테이프의 재생을 정지할 수 있고, 이 비디오 테이프에 대한 영상 인덱스에 액세스할 수 있다는 것이다. 이는 세션 중에 비디오 테이프가 처음으로 이용되었을 때 영상 인덱스를 저장하기 위한 메모리 또는 버퍼를 필요로 할 수도 있다.
- <113> 본 발명은 DCT 계수들을 이용하여 설명하였으나, 파장 계수들(wavelength coefficients)과 같은 대표값을 이용하거나 또는 하위 영역에 대한 대표값을 부여하는 이미지의 하위 영역에서 동작하는 함수를 대신 이용할 수도 있다. 이는 중요 장면 검출뿐만 아니라 키프레임 필터링에서도 이용할 수 있다.
- <114> 본 발명을 선호되는 실시예들과 관련시켜 설명하였지만, 전술한 대략적인 원리들 내에서의 변경들이 당업자에게는 명백한 것으로 이해될 것이고, 따라서 본 발명은 선호되는 실시예들에 제한되지 않고 그러한 변경들을 포함하는 것으로 해석되어야 한다.
- <115> 본 발명은 이용자가 비디오를 시청함에 따라, 북마크되는 세그먼트를 제공하기 위한 단서들을 비디오 스트림들에서 찾기 위해 비디오 분할(video segmentation), 비디오 파싱(video parsing), 스피치 인식(speech recognition), 문자 인식(character recognition) 및 오브젝트 스폿팅(object spotting)과 같은 기존의 기술들을 이용하고 개조할 수도 있다. 대안적으로는, 비디오는 이전에 북마크될 수 있으며, 이용자는 원하는 세그먼트에서 원하는 세그먼트로 건너 뛸 수 있다.
- <116> 첨부한 도면들을 참조하여 본 발명의 선호되는 실시예들을 설명하였지만, 본 발명이 실시예들에만 제한된 것이 아니고, 첨부된 청구 범위에 따른 본 발명의 사상 또는 정신으로부터 벗어나지 않고도 당업자에 의한 다양한 변화 및 변경들이 가능하다.

도면의 간단한 설명

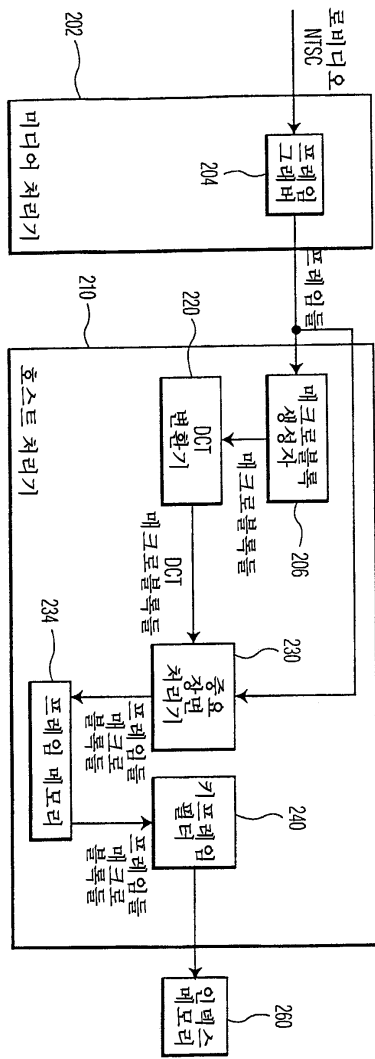
- <14> 도 1은 본 발명의 실시예들에 따라 비디오 콘텐츠를 분할하기 위한 비디오 해석 프로세스를 설명하는 도면.
- <15> 도 2a 및 도 2b는 본 발명의 실시예들에 따라 세그먼트들의 영상 인덱스를 생성하는데 이용된 장치들의 블록도들.
- <16> 도 3은 본 발명의 실시예들에 따라 비디오 이미지로부터 프레임 정보의 선택을 도시한 개략도.
- <17> 도 4는 본 발명의 실시예들에 따라 분할 해석의 세 가지 레벨들을 도시한 차트.
- <18> 도 5는 입력되는 비디오에 대한 프로세스 흐름을 도시한 도면.

도면

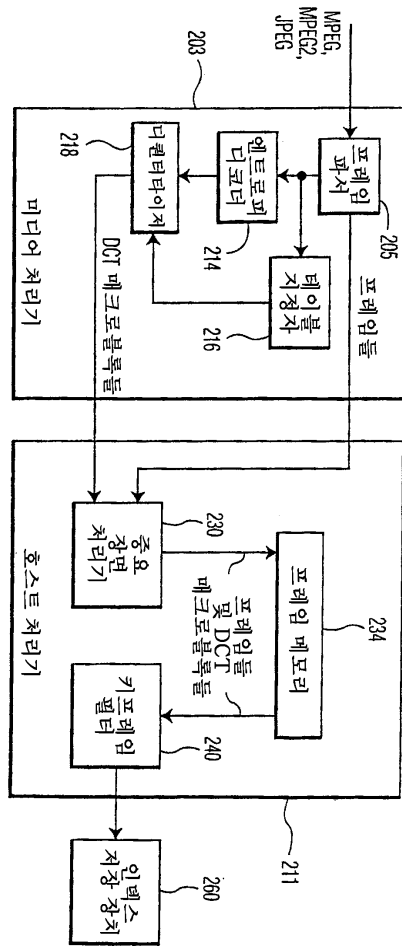
도면1



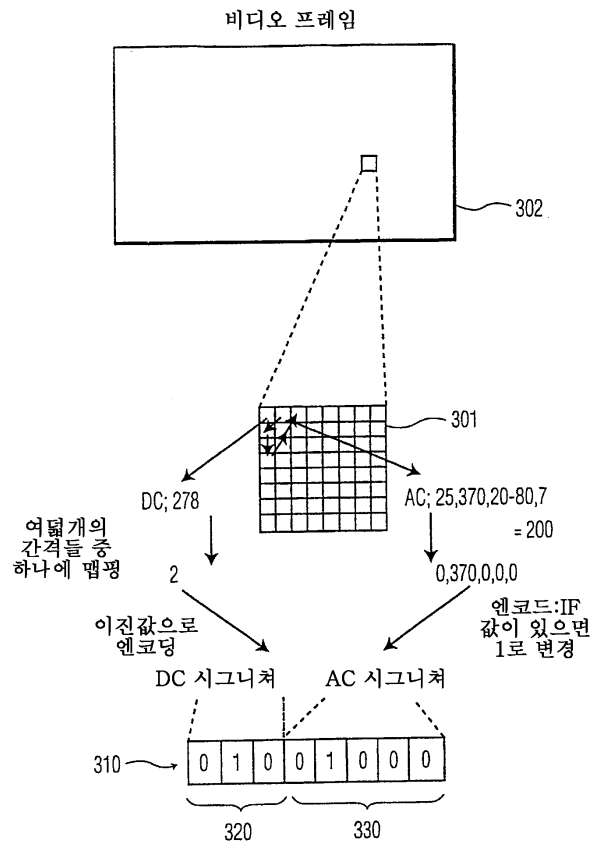
도면2a



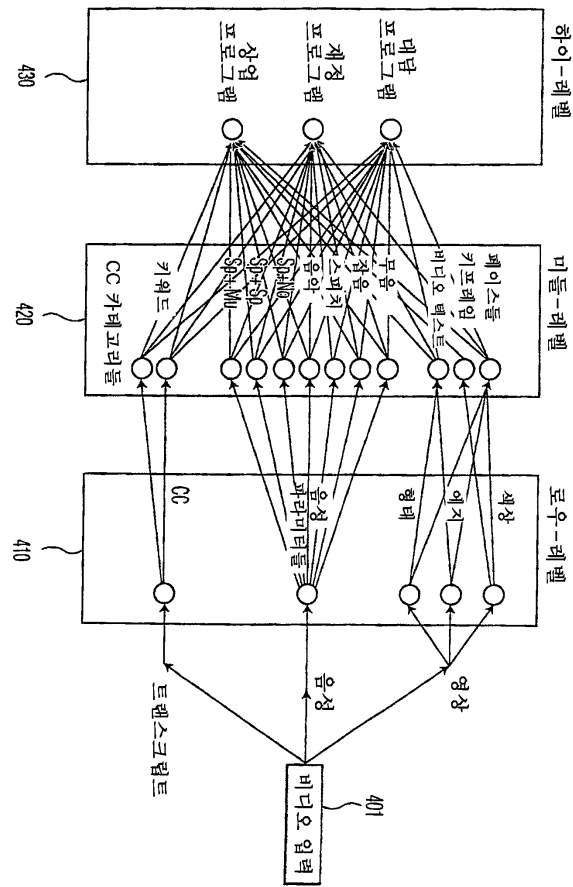
도면2b



도면3



도면4



도면5

