



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2010-0123204
(43) 공개일자 2010년11월24일

(51) Int. Cl.

H04N 5/93 (2006.01) G11B 27/10 (2006.01)

H04N 5/445 (2006.01)

(21) 출원번호 10-2009-0042284

(22) 출원일자 2009년05월14일

심사청구일자 2009년05월14일

(71) 출원인

인하대학교 산학협력단

인천 남구 용현동 253 인하대학교

(72) 발명자

박승보

인천광역시 남구 용현3동 인하대학교 정보공학과
하이테크관 1207호

김유원

인천광역시 남구 용현3동 인하대학교 정보공학과
하이테크관 1207호

(뒷면에 계속)

(74) 대리인

이선재, 김충석

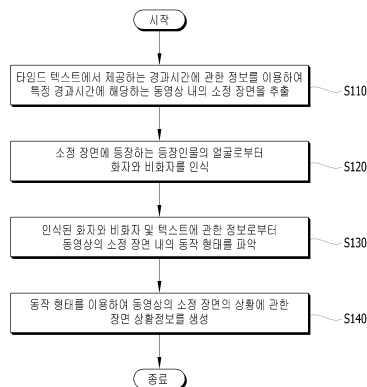
전체 청구항 수 : 총 12 항

(54) 동영상 상황정보 어노테이션 방법 및 장치

(57) 요약

동영상 상황정보를 어노테이션하기 위하여, 경과시간 및 텍스트를 포함하는 타임드 텍스트에서 제공하는 경과시간에 관한 시간정보를 이용하여 특정 경과시간에 해당하는 동영상 내의 소정 장면을 추출하고, 소정 장면 내에 등장하는 적어도 하나의 등장인물의 얼굴로부터 화자와 비화자를 인식한 후, 인식된 화자와 비화자 및 타임드 텍스트의 텍스트에 관한 정보로부터 동영상의 소정 장면 내의 동작 형태를 파악하고, 동작 형태를 이용하여 동영상의 소정 장면의 상황에 관한 장면 상황정보를 생성한다. 이에 따라, 소정 장면에 대한 상황을 어노테이션할 수 있고, 이를 이용하여 동영상 내에서 원하는 위치를 용이하게 검색하고 동영상의 내용을 효과적으로 축약할 수 있다.

대표도 - 도1



(72) 발명자

차상진

인천광역시 남구 용현3동 인하대학교 정보공학과
하이테크관 1207호

조근식

인천광역시 남구 용현3동 인하대학교 컴퓨터정보공
학과

특허청구의 범위

청구항 1

경과시간 및 텍스트를 포함하는 타임드 텍스트(timed text)에서 제공하는 상기 경과시간에 관한 시간정보를 이용하여 특정 경과시간에 해당하는 동영상 내의 소정 장면을 추출하는 단계;

상기 소정 장면 내에 등장하는 적어도 하나의 등장인물의 얼굴로부터 화자와 비화자를 인식하는 단계;

상기 인식된 화자와 비화자 및 상기 타임드 텍스트의 텍스트에 관한 정보로부터 상기 동영상의 소정 장면 내의 동작 형태를 파악하는 단계; 및

상기 동작 형태를 이용하여 상기 동영상의 소정 장면의 상황에 관한 장면 상황정보를 생성하는 단계를 포함하는 동영상 상황정보 어노테이션(annotation) 방법.

청구항 2

제1항에 있어서, 상기 화자와 비화자를 인식하는 단계는,

상기 등장인물의 얼굴로부터 좌안 및 우안의 위치를 검출하고, 상기 좌안 및 상기 우안 사이의 제1 거리를 산출하는 단계;

상기 좌안의 위치로부터 아래로 상기 제1 거리만큼 이격된 위치를 상기 등장인물의 입을 나타내는 박스의 좌상단 위치로 설정하고, 상기 우안의 위치로부터 아래로 상기 제1 거리에 소정 상수값인 제2 거리를 더한 거리만큼 이격된 위치를 상기 등장인물의 입을 나타내는 박스의 우하단 위치로 설정하는 단계; 및

상기 좌상단 위치 및 상기 우하단 위치가 정의하는 상기 박스의 내부 이미지의 히스토그램 변화를 이용하여 상기 변화가 존재하는 얼굴을 상기 화자로 결정하고 상기 변화가 존재하지 않는 얼굴을 상기 비화자로 결정하는 단계를 포함하는 것을 특징으로 하는 동영상 상황정보 어노테이션 방법.

청구항 3

제1항에 있어서, 상기 동영상의 소정 장면 내의 상기 동작 형태는,

상기 화자의 수, 상기 비화자의 수 및 상기 텍스트의 존재 여부를 이용하여 기 설정된 규칙을 기초로 파악되는 것을 특징으로 하는 동영상 상황정보 어노테이션 방법.

청구항 4

제1항에 있어서, 상기 장면 상황정보는,

상기 경과시간, 상기 화자의 이름, 상기 비화자의 이름, 상기 동작 형태 및 상기 텍스트의 내용 중 적어도 하나를 포함하는 XML 파일로 생성되는 것을 특징으로 하는 동영상 상황정보 어노테이션 방법.

청구항 5

동영상 내의 소정 장면에 등장하는 적어도 하나의 등장인물의 얼굴로부터 화자와 비화자를 인식하고, 인식된 상기 화자와 비화자에 관한 제1 상황정보를 제공하는 영상 처리부;

경과시간 및 텍스트를 포함하는 타임드 텍스트에 관한 제2 상황정보를 제공하는 텍스트 처리부; 및

상기 영상 처리부로부터 제공된 상기 제1 상황정보 및 상기 텍스트 처리부로부터 제공된 상기 제2 상황정보로부터 상기 동영상의 소정 장면의 상황에 관한 장면 상황정보를 생성하는 장면상황 생성부를 포함하는 동영상 상황정보 어노테이션 장치.

청구항 6

제5항에 있어서, 상기 영상 처리부는,

상기 동영상 내의 소정 장면으로부터 상기 등장인물의 얼굴을 추출하고, 추출된 상기 얼굴이 누구인지 인식하는 얼굴 인식부; 및

상기 얼굴 인식부에서 인식된 얼굴 중에서 상기 화자와 비화자가 누구인지를 결정하는 화자 결정부를 포함하는 것을 특징으로 하는 동영상 상황정보 어노테이션 장치.

청구항 7

제5항에 있어서,

상기 동영상 내에 등장하는 등장인물의 얼굴을 저장한 얼굴 데이터베이스를 더 포함하고,

상기 얼굴 인식부는, 상기 얼굴 데이터베이스를 이용하여 추출된 상기 얼굴이 누구인지를 인식하는 것을 특징으로 하는 동영상 상황정보 어노테이션 장치.

청구항 8

제5항에 있어서, 상기 텍스트 처리부는,

상기 동영상의 소정 장면에 대하여 상기 경과시간에 관한 시간정보를 포함하는 시간 정보부; 및

상기 동영상의 상기 경과시간에 해당하는 텍스트의 내용에 관한 텍스트정보를 포함하는 텍스트 정보부를 포함하는 것을 특징으로 하는 동영상 상황정보 어노테이션 장치.

청구항 9

제8항에 있어서,

상기 시간 정보부는 상기 시간정보를 상기 영상 처리부에 전달하고,

상기 영상 처리부는 상기 시간 정보부로부터 전달된 시간정보를 이용하여 상기 소정 장면을 정하는 것을 특징으로 하는 동영상 상황정보 어노테이션 장치.

청구항 10

제5항에 있어서,

상기 장면상황 생성부는,

상기 영상 처리부 및 상기 텍스트 처리부로부터 각각 제공된 상기 제1 상황정보 및 상기 제2 상황정보를 이용하여 상기 장면 상황정보를 추출하는 상황정보 추출부; 및

추출된 상기 장면 상황정보를 XML 파일로 변환하여 저장하는 XML 생성부를 포함하는 것을 특징으로 하는 동영상 상황정보 어노테이션 장치.

청구항 11

제5항에 있어서,

상기 화자의 수, 상기 비화자의 수 및 상기 텍스트의 존재 여부를 이용하여 동작 형태를 설정한 규칙들을 저장한 규칙 데이터베이스를 더 포함하고,

상기 장면 상황정보는 상기 규칙 데이터베이스에 저장된 규칙들을 이용하여 설정된 동작 형태를 포함하는 것을 특징으로 하는 동영상 상황정보 어노테이션 장치.

청구항 12

제11항에 있어서, 상기 XML 파일은,

상기 경과시간, 상기 화자의 이름, 상기 비화자의 이름, 상기 동작 형태 및 상기 텍스트의 내용 중 적어도 하나를 포함하는 것을 특징으로 하는 동영상 상황정보 어노테이션 장치.

명세서

발명의 상세한 설명

기술분야

[0001] 본 발명은 동영상 상황정보 어노테이션 방법 및 장치에 관한 것으로, 더욱 상세하게는 얼굴 인식을 이용한 동영상 상황정보 어노테이션 방법 및 장치에 관한 것이다.

배경기술

[0002] 최근, 인터넷과 디지털 멀티미디어의 발달로 인터넷 등의 다양한 수단을 통하여 각종 서비스업자가 제작한 동영상 및 개인이 만든 UCC 동영상 등 다양한 매체로부터 생성된 동영상이 기하급수적으로 증가되어 왔다.

[0003] 이에 따라, 상기와 같은 수많은 동영상 속에서 원하는 동영상을 찾으려는 필요성과 동영상의 축약에 대한 필요성이 지속적으로 증대되어 왔다. 이러한 동영상의 검색과 축약을 위하여, 동영상 정보에서 의미를 추출하여 영상 정보를 어노테이션(annotation)해 두는 작업이 절실히 요청된다.

[0004] 일반적으로, 상기 동영상의 어노테이션 방식은 텍스트 기반의 방식과 시각정보 방식으로 분류된다. 상기 텍스트 기반의 방식은 이미지의 주변의 텍스트 정보로부터 이미지의 내용을 추출하는 방식이고, 상기 시각정보 방식은 이미지 자체에서 시각정보를 추출하여 어노테이션하는 방식이다.

[0005] 상기 텍스트 기반의 방식은, 동영상 주변에서 동영상에 대하여 설명한 정보나 동영상과 관련된 정보가 존재하지 않는 경우도 있어 일정한 한계점을 가진다. 또한, 종래의 시각정보 방식은 이미지 자체에서 시각정보를 추출하여 어노테이션하므로 주변에 관련 정보가 없더라도 동영상 자체에서 직접 정보를 추출하는 점에서 장점이 있으나, 아직 정보 추출 방식에 대한 연구가 미진한 상태이다.

발명의 내용

해결 하고자 하는 과제

[0006] 따라서, 본 발명이 해결하고자 하는 과제는 동영상 자체 내에서 의미를 추출할 수 있는 동영상 상황정보 어노테이션 방법을 제공하는 것이다.

[0007] 본 발명이 해결하고자 하는 다른 과제는 동영상 자체 내에서 의미를 추출할 수 있는 동영상 상황정보 어노테이션 장치를 제공하는 것이다.

과제 해결수단

[0008] 본 발명의 예시적인 일 실시예에 따른 동영상 상황정보 어노테이션(annotation) 방법은, 경과시간 및 텍스트를 포함하는 타임드 텍스트(timed text)에서 제공하는 상기 경과시간에 관한 시간정보를 이용하여 특정 경과시간에 해당하는 동영상 내의 소정 장면을 추출하는 단계, 상기 소정 장면 내에 등장하는 적어도 하나의 등장인물의 얼굴로부터 화자와 비화자를 인식하는 단계, 상기 인식된 화자와 비화자 및 상기 타임드 텍스트의 텍스트에 관한 정보로부터 상기 동영상의 소정 장면 내의 동작 형태를 파악하는 단계 및 상기 동작 형태를 이용하여 상기 동영상의 소정 장면의 상황에 관한 장면 상황정보를 생성하는 단계를 포함한다.

[0009] 일 실시예로, 상기 화자와 비화자를 인식하는 단계는, 상기 등장인물의 얼굴로부터 좌안 및 우안의 위치를 검출하고, 상기 좌안 및 상기 우안 사이의 제1 거리를 산출하는 단계, 상기 좌안의 위치로부터 아래로 상기 제1 거리만큼 이격된 위치를 상기 등장인물의 입을 나타내는 박스의 좌상단 위치로 설정하고, 상기 우안의 위치로부터 아래로 상기 제1 거리에 소정 상수값인 제2 거리를 더한 거리만큼 이격된 위치를 상기 등장인물의 입을 나타내는 박스의 우하단 위치로 설정하는 단계 및 상기 좌상단 위치 및 상기 우하단 위치가 정의하는 상기 박스의 내부 이미지의 히스토그램 변화를 이용하여 상기 변화가 존재하는 얼굴을 상기 화자로 결정하고 상기 변화가 존재하지 않는 얼굴을 상기 비화자로 결정하는 단계를 포함할 수 있다.

[0010] 상기 동영상의 소정 장면 내의 상기 동작 형태는, 상기 화자의 수, 상기 비화자의 수 및 상기 텍스트의 존재 여부를 이용하여 기 설정된 규칙을 기초로 파악될 수 있다.

[0011] 상기 장면 상황정보는, 상기 경과시간, 상기 화자의 이름, 상기 비화자의 이름, 상기 동작 형태 및 상기 텍스트의 내용 중 적어도 하나를 포함하는 XML 파일로 생성될 수 있다.

[0012] 본 발명의 예시적인 일 실시예에 따른 동영상 상황정보 어노테이션 장치는 영상 처리부, 텍스트 처리부 및 장면 상황 생성부를 포함한다. 상기 영상 처리부는 동영상 내의 소정 장면에 등장하는 적어도 하나의 등장인물의 일

굴로부터 화자와 비화자를 인식하고, 인식된 상기 화자와 비화자에 관한 제1 상황정보를 제공한다. 상기 텍스트 처리부는 경과시간 및 텍스트를 포함하는 타임드 텍스트에 관한 제2 상황정보를 제공한다. 상기 장면상황 생성부는 상기 영상 처리부로부터 제공된 상기 제1 상황정보 및 상기 텍스트 처리부로부터 제공된 상기 제2 상황정보로부터 상기 동영상의 소정 장면의 상황에 관한 장면 상황정보를 생성한다.

[0013] 일 실시예로, 상기 영상 처리부는 얼굴 인식부 및 화자 결정부를 포함할 수 있다. 상기 얼굴 인식부는 상기 동영상 내의 소정 장면으로부터 상기 등장인물의 얼굴을 추출하고, 추출된 상기 얼굴이 누구인지 인식한다. 상기 화자 결정부는 상기 얼굴 인식부에서 인식된 얼굴 중에서 상기 화자와 비화자가 누구인지를 결정한다.

[0014] 상기 동영상 상황정보 어노테이션 장치는 얼굴 데이터베이스를 더 포함할 수 있다. 상기 얼굴 데이터베이스는 상기 동영상 내에 등장하는 등장인물의 얼굴을 저장한다. 이 경우, 상기 얼굴 인식부는 상기 얼굴 데이터베이스를 이용하여 추출된 상기 얼굴이 누구인지를 인식할 수 있다.

[0015] 일 실시예로, 상기 텍스트 처리부는 시간 정보부 및 텍스트 정보부를 포함할 수 있다. 상기 시간 정보부는 상기 동영상의 소정 장면에 대하여 상기 경과시간에 관한 시간정보를 포함한다. 상기 텍스트 정보부는 상기 동영상의 상기 경과시간에 해당하는 텍스트의 내용에 관한 텍스트정보를 포함한다. 상기 시간 정보부는 상기 시간정보를 상기 영상 처리부에 전달하고, 상기 영상 처리부는 상기 시간 정보부로부터 전달된 시간정보를 이용하여 상기 소정 장면을 정할 수 있다.

[0016] 일 실시예로, 상기 장면상황 생성부는 상황정보 추출부 및 XML 생성부를 포함할 수 있다. 상기 상황정보 추출부는 상기 영상 처리부 및 상기 텍스트 처리부로부터 각각 제공된 상기 제1 상황정보 및 상기 제2 상황정보를 이용하여 상기 장면 상황정보를 추출한다. 상기 XML 생성부는 추출된 상기 장면 상황정보를 XML 파일로 변환하여 저장한다.

[0017] 상기 동영상 상황정보 어노테이션 장치는 상기 화자의 수, 상기 비화자의 수 및 상기 텍스트의 존재 여부를 이용하여 동작 형태를 설정한 규칙들을 저장한 규칙 데이터베이스를 더 포함할 수 있으며, 상기 장면 상황정보는 상기 규칙 데이터베이스에 저장된 규칙들을 이용하여 설정된 동작 형태를 포함할 수 있다.

[0018] 상기 XML 파일은, 예를 들면, 상기 경과시간, 상기 화자의 이름, 상기 비화자의 이름, 상기 동작 형태 및 상기 텍스트의 내용 중 적어도 하나를 포함할 수 있다.

효 과

[0019] 본 발명에 따르면, 동영상의 특정 장면에서 등장인물의 얼굴을 인식한 후 화자와 비화자를 결정하고, 이와 함께 텍스트 정보를 이용하여 소정 장면의 상황에 관한 장면 상황정보를 생성함으로써, 상기 소정 장면에 대한 상황을 이미지 자체로부터 추출하여 어노테이션할 수 있고, 이를 이용하여 동영상 내에서 원하는 위치를 용이하게 검색하고 동영상의 내용을 효과적으로 축약할 수 있다.

[0020] 따라서, 동영상 검색 서비스용으로 활용될 수 있고, 영화나 UCC 동영상 서비스 제공 사이트 등에서 효과적으로 응용 서비스가 가능하며, 청각 장애우 및 외국어 교육용 소프트웨어 등으로 널리 활용 가능하다.

[0021] 또한, XML 파일 형식으로 장면 상황정보를 생성하는 경우 이식성이 뛰어나므로, 다양한 시스템에서 보다 용이하고 편리하게 이용될 수 있다.

발명의 실시를 위한 구체적인 내용

[0022] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 형태를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 본문에 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 개시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.

[0023] 제1, 제2 등의 용어는 다양한 구성 요소들을 설명하는데 사용될 수 있지만, 상기 구성 요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성 요소를 다른 구성 요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성 요소는 제2 구성 요소로 명명될 수 있고, 유사하게 제2 구성 요소도 제1 구성 요소로 명명될 수 있다.

[0024] 본 출원에서 사용한 용어는 단지 특정한 실시예들을 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "가지다" 등의 용어는 명세서에 기재된 특징, 숫자, 단계, 동작, 구성 요소, 부분품 또는 이들

을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성 요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

- [0025] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 갖는다.
- [0026] 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 갖는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.
- [0027] 이하, 첨부한 도면들을 참조하여, 본 발명의 바람직한 실시예들을 보다 상세하게 설명한다.
- [0028] 도 1은 본 발명의 일 실시예에 의한 동영상 상황정보 어노테이션 방법을 나타낸 흐름도이다.
- [0029] 도 1을 참조하면, 본 발명의 일 실시예에 따라 동영상 상황정보를 어노테이션(annotation)을 하기 위하여, 먼저 타임드 텍스트(timed text)에서 제공하는 경과시간에 관한 정보를 이용하여 특정 경과시간에 해당하는 동영상 내의 소정 장면을 추출한다(S110).
- [0030] 타임드 텍스트란 오디오, 비디오와 같은 매체와 동기화된 텍스트를 의미하는 것으로, 상기 타임드 텍스트는 영상 처음을 기준으로 경과된 시간에 해당하는 경과시간 및 상기 경과시간에 동기화된 텍스트를 포함한다. 상기 경과시간은 실제 경과된 시간일 수도 있고, 경과된 프레임 수일 수도 있다. 예를 들면, 상기 경과시간은 밀리세컨드(millisecond)단위로 표시될 수 있다.
- [0031] 일 실시예로, 상기 타임드 텍스트는 동영상과 함께 저장된 자막 정보로 DVD의 자막 정보, 컴퓨터의 동영상 재생기의 SMI 형식(예를 들면, Windows Media Player)이나 RealText(예를 들면, RealPlayer), W3C에서 제안하고 있는 DFXP 등을 포함할 수 있다.
- [0032] 이어서, 상기 소정 장면 내에 등장하는 적어도 하나의 등장인물의 얼굴로부터 화자와 화자가 아닌 자(이하, "비화자"라 함)를 인식한다(S120).
- [0033] 상기 화자와 비화자는 등장인물의 얼굴의 변화를 통하여 인식될 수 있다. 일 실시예로, 상기 화자와 비화자는 입의 움직임을 이용하여 판단할 수 있다. 비화자는 입과 입 주변의 움직임이 거의 없어 시간에 따른 프레임에 대하여 변화가 거의 없다. 반면, 화자는 입과 입 주변이 움직임이므로 시간에 따른 프레임에 대하여 변화가 비교적 크다. 따라서, 이러한 입과 입 주변의 움직임을, 예를 들어 각 프레임별 히스토그램 그래프를 이용하여 판단할 수 있다.
- [0034] 도 2는 도 1의 화자와 비화자를 인식하는 단계의 구체적인 예를 나타낸 흐름도이고, 도 3은 도 2에 나타난 화자와 비화자를 인식하는 단계를 설명하기 위한 개념도이다.
- [0035] 도 2 및 도 3을 참조하면, 상기 화자와 비화자를 인식하기 위하여, 먼저 동영상의 장면(100) 내에 등장하는 등장인물(110)의 얼굴로부터 좌안의 위치((x1,y1)) 및 우안의 위치((x2,y2))를 검출하고, 상기 좌안 및 상기 우안 사이의 제1 거리(d1)를 산출한다(S122). 사람의 눈동자는 색의 특징으로 인하여 입보다 검출이 훨씬 용이하므로, 좌안과 우안의 위치를 먼저 검출한 후, 이로부터 상대적인 입의 위치를 검출할 수 있다. 일반적인 사람의 경우, 상기 제1 거리(d1)를 기준으로 입의 위치를 파악할 수 있다.
- [0036] 이어서, 상기 좌안의 위치로부터 아래로 상기 제1 거리(d1)만큼 이격된 위치를 상기 등장인물의 입을 나타내는 박스의 좌상단 위치((x1,y1+d1))로 설정하고(S124), 상기 우안의 위치로부터 아래로 상기 제1 거리에 소정의 제2 거리(d2)를 더한 거리만큼 이격된 위치를 상기 등장인물의 입을 나타내는 박스의 우하단 위치((x2,y2+d1+d2))로 설정한다(S126). 상기 제2 거리(d2)는 상기 제1 거리(d1)보다 작은 값 중에서 선정하여 사용될 수 있으며, 일 예로 제1 거리의 1/3 내지 2/3의 범위에서 설정될 수 있다. 이로써, 등장인물의 입을 포함하는 박스를 설정할 수 있다.
- [0037] 이어서, 상기 좌상단 위치 및 상기 우하단 위치가 정의하는 상기 박스의 내부 이미지의 히스토그램 변화를 이용하여 상기 변화가 존재하는 얼굴을 상기 화자로 결정하고 상기 변화가 존재하지 않는 얼굴을 상기 비화자로 결정한다(S128). 즉, 상기 박스 내부 이미지의 히스토그램 변화가 있는 경우 입이 움직이고 있는 것으로 판단할 수 있으며, 이러한 경우를 화자의 입으로 판단할 수 있다.

[0038] 상기와 같은 화자와 비화자를 인식하는 방법은 등장인물의 옆 모습이 나타난 경우에는 적절히 변형되어 활용될 수 있다. 즉, 좌안과 우안 사이의 거리를 이용하는 대신, 하나의 눈의 크기를 검출하고 이를 이용하여 입의 위치를 파악하여 입의 움직임을 관찰할 수 있다.

[0039] 한편, 상기와 같은 화자와 비화자를 인식하는 과정 중, 혹은 인식하는 과정 이전에 상기 특정 장면에서 나타나는 인물들에 대하여 등장인물 중 누구인지 소정의 등장인물의 얼굴 데이터베이스를 통하여 비교함으로써 얼굴 인식을 선행할 수 있다.

[0040] 다시 도 1을 참조하면, 다음으로, 상기 인식된 화자와 비화자 및 상기 타임드 텍스트의 텍스트에 관한 정보로부터 상기 동영상의 소정 장면 내의 동작 형태를 파악한다(S130).

[0041] 상기 동작 형태는, 예를 들면, 상기 화자의 수, 상기 비화자의 수 및 상기 텍스트의 존재 여부를 이용하여 기 설정된 규칙을 기초로 파악될 수 있다. 구체적인 일 예로 다음과 같은 규칙을 기초로 상기 동작 형태를 파악할 수 있다. 표 1은 등장인물 간의 동작 형태 판단 규칙의 일 예를 나타낸다.

표 1

번호	등장인물수	화자수	비화자수	텍스트유무	동작 형태
1	2	1	1	있음	(화자) say to (청자)
2	1	0	1	있음	(청자) listen
3	1	1	0	있음	(화자) speak alone
4	3이상	1	2이상	있음	(화자) talk with (청자들)
5	2이상	0	2이상	있음	(청자들) listen
6	0	0	0	있음	nothing(or open caption)
7	-	-	-	없음	nothing

[0043] 표 1을 참조하면, 화자와 비화자의 수와 텍스트의 유무로부터 동작 형태를 판단할 수 있다.

[0044] 화자가 인식되고 비화자도 인식된 상태에서 텍스트가 존재하는 경우, 번호 1과 같이 화자와 비화자가 각 1인으로 인식되면 화자와 비화자(청자) 사이에 말하고 있는 상태 'say to'로 판단할 수 있고, 번호 4와 같이 화자가 1인이고 비화자가 2인 이상으로 인식되면 화자가 비화자(청자)들을 대상으로 연설하거나 토론을 하는 상태 'talk with'로 판단할 수 있다.

[0045] 화자가 인식되지 않고 비화자만 인식된 상태에서 텍스트가 존재하는 경우, 영상에 나타나지 않는 화자가 존재하는 것이므로 영상에 보이는 비화자가 청자가 되며 청자가 영상에 보이지 않는 화자의 말을 듣고 있는 상태 'listen'으로 판단할 수 있다. 이때, 번호 2와 같이 비화자가 1인으로 인식되면 비화자(청자)가 전화를 받고 있거나, 범인이 심문을 당하는 장면 등에 해당하며, 번호 5와 같이 비화자가 2인 이상으로 인식되면 비화자(청자)가 강연을 듣고 있거나 수업에 받고 있는 장면 등에 해당한다.

[0046] 번호 3과 같이 비화자가 인식되지 않고 화자만 인식된 상태에서 텍스트가 존재하는 경우, 장면 속에서 화자가 혼자 말하는 상태 'speak alone'으로 판단할 수 있다. 이 경우, 화자가 전화 통화중이거나, 연설자 등에 해당할 수 있다.

[0047] 번호 6과 같이 화자나 비화자 모두 인식되지 않고 텍스트만 존재하는 경우, 장면에 등장인물과 관계된 어떠한 상황도 존재하지 않는 상태 'nothing'으로 판단할 수 있다. 이 경우, 텍스트는 오픈 캡션(open caption)과 같이 장면에 대한 단순한 설명이나 내레이션(narration), 또는 장면 중 나오는 문자에 대한 번역 등에 해당할 수 있다.

[0048] 번호 7은 화자나 비화자의 인식과 관계없이 텍스트가 없는 경우로, 장면에 등장인물 사이의 대화가 이루어지고 있지 않은 상태 'nothing'으로 판단할 수 있다. 화자나 비화자 모두 인식되지 않은 경우 풍경 등을 나타내는 장면일 수 있으며, 화자나 비화자가 인식된 경우(화자가 인식된 경우는 텍스트가 없으므로 잘못 인식이 되었거나 입 모양만 내는 경우일 것이다) 등장인물이 대기 상태에 있거나, 서로 마주보는 상태 등에 해당할 수 있다.

[0049] 다시 도 1을 참조하면, 이어서 상기 동작 형태를 이용하여 상기 동영상의 소정 장면의 상황에 관한 장면 상황정보를 생성한다(S140).

[0050] 상기 장면 상황정보는, 예를 들면, 상기 경과시간, 상기 화자의 이름, 상기 비화자(청자)의 이름, 상기 동작 형태 및 상기 텍스트(자막)의 내용 중 적어도 하나를 포함할 수 있다. 상기 장면 상황정보는 어노테이션을 위한

여러 가지 형태로 제공될 수 있으며, 일 예로 XML 파일로 생성될 수 있다.

- [0051] 도 4는 도 1의 장면 상황정보를 생성한 결과인 XML 파일의 일 예를 나타낸다.
- [0052] 도 4를 참조하면, XML 파일(200)은 경과시간(210), 화자의 이름(220), 비화자의 이름(230), 동작 형태(240), 텍스트의 내용(250)을 포함한다.
- [0053] 일 예로, 도 4에 도시된 XML 파일(200)의 장면 상황정보를 이용하면, "1000"이라는 시간이 경과된 때 화자인 "갑돌이"가 청자인 "갑순이"에게 "Where are you going?"이라는 "말을 하고 있다"는 사실을 알 수 있다.
- [0054] 상기 XML 파일(200)은 상기 경과시간(210)에서의 장면 상황정보를 식별하기 위한 아이디(260)를 더 포함할 수 있다. 도 4에서는, 상기 아이디(260)가 상기 경과시간(210)과 같은 값으로 설정되어 있다.
- [0055] 이와 같이 생성된 장면 상황정보를 이용하여 상기 소정 장면에 대한 상황을 어노테이션할 수 있고, 이를 이용하여 동영상 내에서 원하는 위치를 용이하게 검색하고 동영상의 내용을 효과적으로 축약할 수 있다.
- [0056] 도 5는 본 발명의 일 실시예에 의한 동영상 상황정보 어노테이션 장치를 도시한 블록도이다.
- [0057] 도 5를 참조하면, 본 발명의 일 실시예에 의한 동영상 상황정보 어노테이션 장치(300)는 영상 처리부(310), 텍스트 처리부(320) 및 장면상황 생성부(330)를 포함한다.
- [0058] 상기 영상 처리부(310)는 동영상 내의 소정 장면에 등장하는 적어도 하나의 등장인물의 얼굴로부터 화자와 비화자를 인식하고, 인식된 상기 화자와 비화자에 관한 제1 상황정보를 제공한다.
- [0059] 일 실시예로, 상기 영상 처리부(310)는 얼굴 인식부(312) 및 화자 결정부(314)를 포함할 수 있다.
- [0060] 상기 얼굴 인식부(312)는 상기 동영상 내의 소정 장면으로부터 상기 등장인물의 얼굴을 추출하고, 추출된 상기 얼굴이 누구인지 인식한다. 이 경우, 상기 동영상 상황정보 어노테이션 장치(300)는 얼굴 데이터베이스(340)를 더 포함할 수 있다. 상기 얼굴 데이터베이스(340)에는 상기 동영상 내에 등장하는 등장인물의 얼굴이 저장된다. 상기 얼굴 인식부(312)는 등장인물의 얼굴이 저장되어 있는 상기 얼굴 데이터베이스(340)를 이용하여, 추출된 상기 얼굴이 상기 등장인물 중 누구인지를 인식할 수 있다.
- [0061] 상기 화자 결정부(314)는 상기 얼굴 인식부(312)에서 인식된 얼굴 중에서 상기 화자와 비화자가 누구인지를 결정한다. 상기 화자 결정부(314)는 다양한 방법을 이용하여 상기 화자와 비화자를 결정할 수 있으며, 일 실시예로 도 1 내지 도 4에서 설명된 방법을 이용하여 상기 화자와 비화자를 결정할 수 있다. 상기 제1 상황정보는 결정된 상기 화자와 비화자에 관한 정보를 포함할 수 있으며, 상기 화자와 비화자에 관한 정보는 후술되는 장면 상황 생성부(330)로 제공될 수 있다.
- [0062] 상기 텍스트 처리부(320)는 경과시간 및 텍스트를 포함하는 타임드 텍스트에 관한 제2 상황정보를 제공한다. 상기 타임드 텍스트에 관한 설명은 도 1에서 설명된 바와 실질적으로 동일하므로 중복되는 상세한 설명은 생략한다.
- [0063] 일 실시예로, 상기 텍스트 처리부(320)는 시간 정보부(322) 및 텍스트 정보부(324)를 포함할 수 있다.
- [0064] 상기 시간 정보부(322)는 상기 동영상의 소정 장면에 대하여 상기 경과시간에 관한 시간정보를 포함한다. 상기 경과시간에 관한 시간정보는 도 1에서 설명한 바와 같이 실제 경과된 시간 또는 경과된 프레임 수에 관한 정보일 수 있다. 상기 시간 정보부(322)는 상기 시간정보를 상기 영상 처리부(310)에 전달하고, 상기 영상 처리부(310)는 상기 시간 정보부(322)로부터 전달된 시간정보를 이용하여 상기 소정 장면을 정할 수 있다. 즉, 상기 영상 처리부(310)는 상기 시간 정보부(322)로부터 전달된 시간정보를 이용하여 특정 경과시간에 해당하는 동영상 내의 소정 장면을 추출할 수 있다.
- [0065] 상기 텍스트 정보부(324)는 상기 동영상의 상기 경과시간에 해당하는 텍스트의 내용에 관한 텍스트정보를 포함한다. 예를 들면, 상기 텍스트정보는 상기 경과시간에 나타나게 되는 텍스트의 내용이다. 상기 제2 상황정보는 상기 텍스트정보를 포함할 수 있으며, 상기 텍스트정보는 후술되는 장면상황 생성부(330)에 제공될 수 있다.
- [0066] 상기 장면상황 생성부(330)는 상기 영상 처리부(310)로부터 제공된 상기 제1 상황정보 및 상기 텍스트 처리부(320)로부터 제공된 상기 제2 상황정보로부터 상기 동영상의 소정 장면의 상황에 관한 장면 상황정보를 생성한다.
- [0067] 일 실시예로, 상기 장면상황 생성부(330)는 상황정보 추출부(332) 및 XML 생성부(334)를 포함할 수 있다.

- [0068] 상기 상황정보 추출부(332)는 상기 영상 처리부(310) 및 상기 텍스트 처리부(320)로부터 각각 제공된 상기 제1 상황정보 및 상기 제2 상황정보를 이용하여 상기 장면 상황정보를 추출한다. 상기 장면 상황정보는 상기 동영상의 소정 장면 내의 등장인물의 동작 형태를 포함할 수 있다. 상기 동작 형태는 기 설정된 규칙을 기초로 파악될 수 있다.
- [0069] 예를 들면, 상기 동영상 상황정보 어노테이션 장치(300)는 규칙 데이터베이스(350)를 더 포함할 수 있다. 상기 규칙 데이터베이스(350)에는 상기 화자의 수, 상기 비화자의 수 및 상기 텍스트의 존재 여부를 이용하여 동작 형태를 설정한 규칙들이 저장된다. 상기 장면 상황정보는 상기 규칙 데이터베이스(350)에 저장된 규칙들을 이용하여 설정된 등장인물의 동작 형태를 포함할 수 있다. 상기 규칙들은 예를 들어 표 1에서 설명된 바와 같이 설정될 수 있다.
- [0070] 상기 XML 생성부(334)는 추출된 상기 장면 상황정보를 XML 파일로 변환하여 저장한다. 상기 XML 파일은, 예를 들면, 상기 경과시간, 상기 화자의 이름, 상기 비화자의 이름, 상기 동작 형태 및 상기 텍스트의 내용 중 적어도 하나를 포함할 수 있다. 상기 XML 파일은 일 예로 도 4에 도시된 바와 같이, 경과시간(210), 화자의 이름(220), 비화자의 이름(230), 동작 형태(240), 텍스트의 내용(250)을 포함할 수 있다.
- [0071] 이와 같이 생성된 장면 상황정보를 이용하여 상기 소정 장면에 대한 상황을 어노테이션할 수 있고, 이를 이용하여 동영상 내에서 원하는 위치를 용이하게 검색하고 동영상의 내용을 효과적으로 축약할 수 있다.
- [0072] 앞서 설명한 본 발명의 상세한 설명에서는 본 발명의 바람직한 실시예들을 참조하여 설명하였지만, 해당 기술분야의 숙련된 당업자 또는 해당 기술분야에 통상의 지식을 갖는 자라면 후술될 특허청구범위에 기재된 본 발명의 사상 및 기술 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 및 변경시킬 수 있을 것이다. 따라서, 전술한 설명 및 아래의 도면은 본 발명의 기술사상을 한정하는 것이 아닌 본 발명을 예시하는 것으로 해석되어야 한다.

도면의 간단한 설명

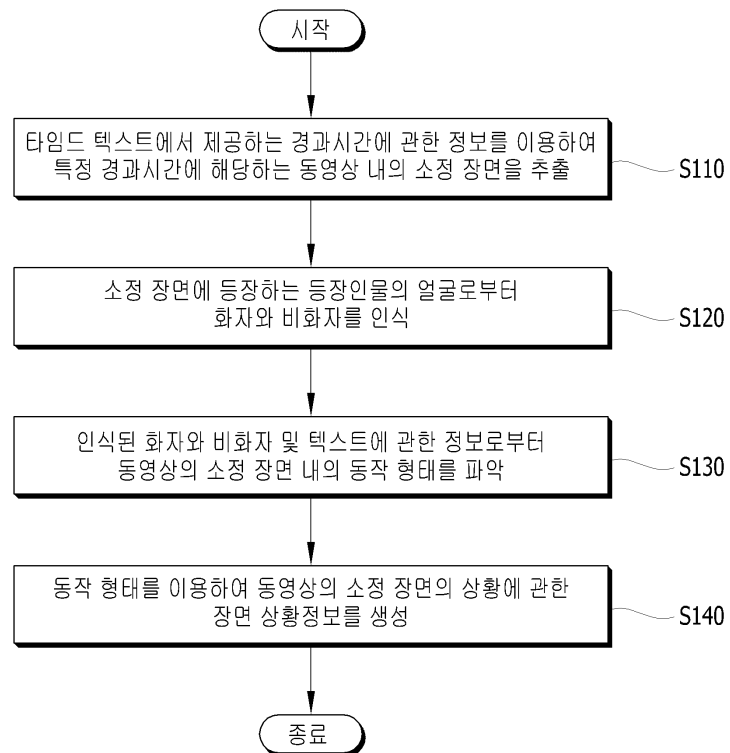
- [0073] 도 1은 본 발명의 일 실시예에 의한 동영상 상황정보 어노테이션 방법을 나타낸 흐름도이다.
- [0074] 도 2는 도 1의 화자와 비화자를 인식하는 단계의 구체적인 예를 나타낸 흐름도이다.
- [0075] 도 3은 도 2에 나타난 화자와 비화자를 인식하는 단계를 설명하기 위한 개념도이다.
- [0076] 도 4는 도 1의 장면 상황정보를 생성한 결과인 XML 파일의 일 예를 나타낸다.
- [0077] 도 5는 본 발명의 일 실시예에 의한 동영상 상황정보 어노테이션 장치를 도시한 블록도이다.

[0078] <주요 도면번호에 대한 간단한 설명>

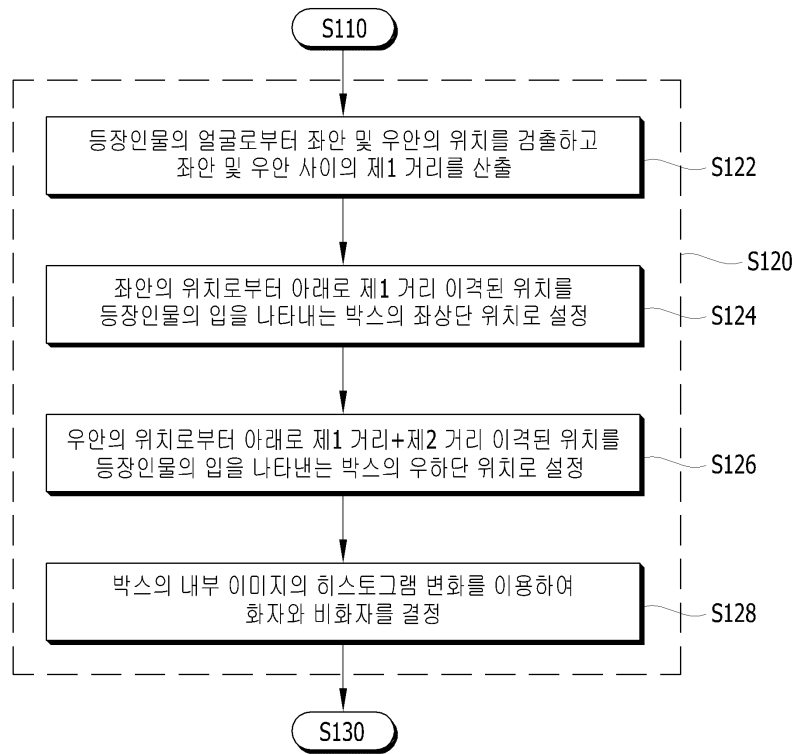
- | | |
|--------------------------------|-----------------|
| [0079] 100 : 동영상의 장면 | 200 : XML 파일 |
| [0080] 300 : 동영상 상황정보 어노테이션 장치 | |
| [0081] 310 : 영상 처리부 | 312 : 얼굴 인식부 |
| [0082] 314 : 화자 결정부 | 320 : 텍스트 처리부 |
| [0083] 322 : 시간 정보부 | 324 : 텍스트 정보부 |
| [0084] 330 : 장면상황 생성부 | 332 : 상황정보 추출부 |
| [0085] 334 : XML 생성부 | 340 : 얼굴 데이터베이스 |
| [0086] 350 : 규칙 데이터베이스 | |

도면

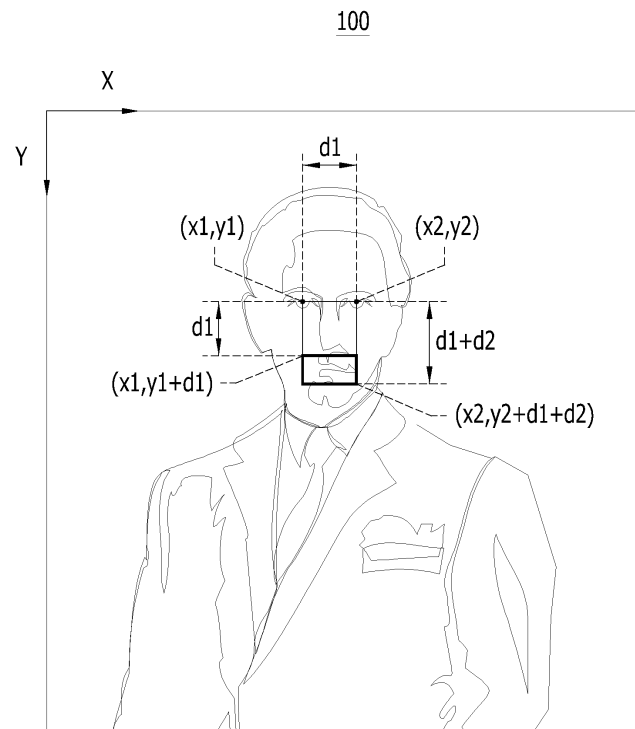
도면1



도면2

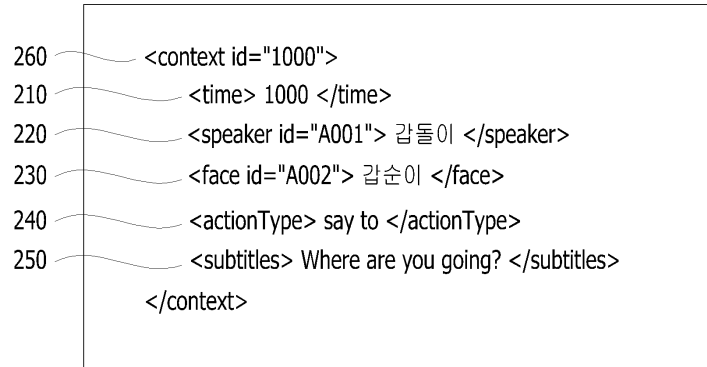


도면3



도면4

200



도면5

