



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2010-0000256
(43) 공개일자 2010년01월06일

(51) Int. Cl.

H04N 7/08 (2006.01) H04N 5/445 (2006.01)

(21) 출원번호 10-2008-0059686

(22) 출원일자 2008년06월24일

심사청구일자 2008년06월24일

(71) 출원인

인하대학교 산학협력단

인천 남구 용현동 253 인하대학교

(72) 발명자

박승보

인천 남구 용현3동 인하대학교 정보공학과 하이테크관1207호

조근식

인천 연수구 동춘2동 송도신도시 풍림아파트 601-1302호

오경진

인천 남구 용현3동 인하대학교 정보공학과 하이테크관1207호

(74) 대리인

황광연, 강동호, 오정환

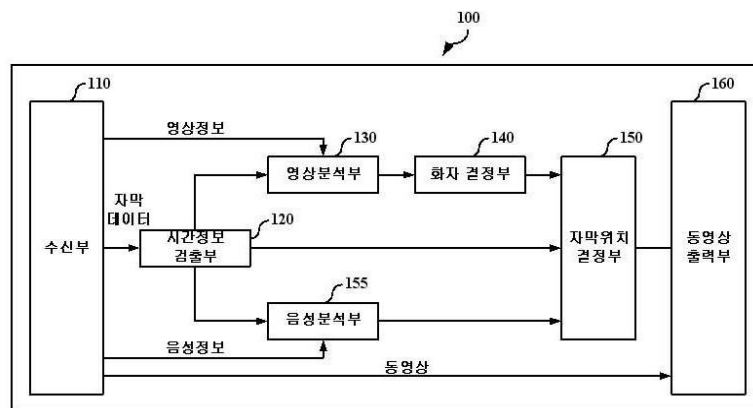
전체 청구항 수 : 총 5 항

(54) 동영상의 자막 표시 방법

(57) 요약

동영상을 구성하는 음성정보와 영상정보를 이용하여 결정된 화자의 위치에 따라 동영상의 자막이 출력되는 위치를 적응적으로 결정할 수 있는 본 발명의 일 측면에 따른 동영상의 자막 표시 방법은 동영상에 표시될 자막으로부터 상기 자막이 표시될 시간정보를 획득하는 단계; 상기 동영상을 구성하는 영상 프레임들 중 상기 시간정보에 상응하는 영상 프레임들로부터 얼굴을 검출하는 단계; 상기 시간정보에 상응하는 각 영상 프레임들로부터 검출된 얼굴의 특징점의 변경여부를 이용하여 상기 시간정보에 상응하는 영상 프레임들에 해당하는 화자를 결정하는 단계; 상기 결정된 화자의 얼굴을 이용하여 상기 시간정보에 상응하는 각 영상 프레임들 내에서 상기 화자의 얼굴영역 범위를 결정하는 단계; 및 상기 자막이 표시될 영역을 상기 화자의 얼굴영역 범위에 따라 적응적으로 결정하여 상기 결정된 영역에 상기 자막을 표시하는 단계를 포함한다.

대표도



특허청구의 범위

청구항 1

동영상에 표시될 자막으로부터 상기 자막이 표시될 시간정보를 획득하는 단계;

상기 동영상을 구성하는 영상 프레임들 중 상기 시간정보에 상응하는 영상 프레임들로부터 얼굴을 검출하는 단계;

상기 시간정보에 상응하는 각 영상 프레임들로부터 검출된 얼굴의 특징점의 변경여부를 이용하여 상기 시간정보에 상응하는 영상 프레임들에 해당하는 화자를 결정하는 단계;

상기 결정된 화자의 얼굴을 이용하여 상기 시간정보에 상응하는 각 영상 프레임들 내에서 상기 화자의 얼굴영역 범위를 결정하는 단계; 및

상기 자막이 표시될 영역을 상기 화자의 얼굴영역 범위에 따라 적응적으로 결정하여 상기 결정된 영역에 상기 자막을 표시하는 단계를 포함하는 것을 특징으로 하는 동영상의 자막 표시 방법.

청구항 2

제1항에 있어서, 상기 자막 표시 단계는,

상기 화자의 얼굴영역 범위로부터 소정 거리 이내의 영역들 중 상기 자막이 표시될 후보 영역을 설정하는 단계; 및

상기 후보 영역들 중 색편차가 가장 적고 채도가 가장 높은 후보 영역을 상기 자막이 표시될 영역으로 결정하는 단계를 포함하는 것을 특징으로 하는 동영상의 자막 표시 방법.

청구항 3

제 1 항에 있어서, 상기 자막 표시 단계에서,

상기 자막은 상기 자막이 표시될 영역의 색상과 보색관계에 있는 색상으로 표시하는 것을 특징으로 하는 동영상의 자막표시 방법.

청구항 4

제 1 항에 있어서,

상기 자막 표시 단계 이전에, 상기 동영상을 구성하는 음성정보를 분석하여 음성이 발생하는 공간적 위치를 결정하는 단계를 더 포함하고,

상기 자막 표시 단계에서, 상기 음성의 공간적 위치를 함께 이용하여 상기 자막이 표시될 영역을 결정하는 것을 특징으로 하는 동영상의 자막 표시 방법.

청구항 5

제 4 항에 있어서,

상기 음성의 공간적 위치는 각 채널별로 출력되는 음성신호의 차이 및 음성주파수 대역 중 적어도 하나를 이용하여 결정되는 것을 특징으로 하는 동영상의 자막 표시 방법.

명 세 서

발명의 상세한 설명

기술 분야

<1> 본 발명은 영상 표시에 관한 것으로, 보다 구체적으로, 동영상의 자막을 표시하는 방법에 관한 것이다.

배경 기술

- <2> 최근 과학기술의 발전 및 경제수준의 향상으로 인해 초고속 인터넷과 같은 통신망의 보급과 초고속 통신망 이용자의 급격한 증가가 이루어졌고, 초고속 통신망 이용자의 급격한 증가는 통신망을 통한 다양한 영상 콘텐츠의 보급화를 가능하게 하였다.
- <3> 상술한 영상 콘텐츠의 종류에는 교육, 의료, 또는 과학 등과 같은 전문 분야의 동영상과 방송, 또는 뮤직 비디오 등과 같은 오락/문화 분야의 동영상이 포함된다.
- <4> 이와 같은 동영상 중 외국어로 제작되어 있는 동영상의 경우 시청자의 이해를 돕기 위한 자막이 동영상 내에 표시되는데, 시청자들은 이러한 자막을 통해 동영상에 등장하는 화자들의 대화 내용을 원활하게 이해할 수 있게 된다.
- <5> 한편, 최근에는 외국어로 제작된 동영상 뿐만 아니라 국어로 제작된 동영상의 경우라도 동영상의 재미를 더하기 위해 동영상 내에 자막을 표시하는 경우도 있다.
- <6> 그러나, 기존에는 이러한 동영상의 자막을 표시함에 있어서, 동영상 내에 등장하는 화자의 위치에 관계 없이 화면의 일편, 예컨대, 화면의 하단부에 자막을 고정적으로 표시하였기 때문에, 시청자가 동영상에 등장하는 화자의 대화 내용을 이해하기 위해서는 동영상과 동영상의 하단부에 표시되는 자막을 번갈아 가며 시청할 수 밖에 없어서 동영상에 대한 시각적 몰입도가 저하될 뿐만 아니라, 이로 인해 동영상에 대한 이해도가 떨어질 수 밖에 없다는 문제점이 있다.

발명의 내용

해결 하고자하는 과제

- <7> 본 발명은 상술한 문제점을 해결하기 위한 것으로서, 동영상을 구성하는 음성정보와 영상정보를 이용하여 결정된 화자의 위치에 따라 동영상의 자막이 출력되는 위치를 적응적으로 결정할 수 있는 동영상의 자막 표시 방법을 제공하는 것을 기술적 과제로 한다.

과제 해결수단

- <8> 상술한 기술적 과제를 달성하기 위한 본 발명의 일 측면에 따른 동영상의 자막 표시 방법은 동영상에 표시될 자막으로부터 상기 자막이 표시될 시간정보를 획득하는 단계; 상기 동영상을 구성하는 영상 프레임들 중 상기 시간정보에 상응하는 영상 프레임들로부터 얼굴을 검출하는 단계; 상기 시간정보에 상응하는 각 영상 프레임들로부터 검출된 얼굴의 특징점의 변경여부를 이용하여 상기 시간정보에 상응하는 영상 프레임들에 해당하는 화자를 결정하는 단계; 상기 결정된 화자의 얼굴을 이용하여 상기 시간정보에 상응하는 각 영상 프레임들 내에서 상기 화자의 얼굴영역 범위를 결정하는 단계; 및 상기 자막이 표시될 영역을 상기 화자의 얼굴영역 범위에 따라 적응적으로 결정하여 상기 결정된 영역에 상기 자막을 표시하는 단계를 포함한다.
- <9> 일 실시예에 있어서, 상기 자막 표시 단계는, 상기 화자의 얼굴영역 범위로부터 소정 거리 이내의 영역들 중 상기 자막이 표시될 후보 영역을 설정하는 단계; 및 상기 후보 영역들 중 색편차가 가장 적고 채도가 가장 높은 후보 영역을 상기 자막이 표시될 영역으로 결정하는 단계를 포함하는 것을 특징으로 한다.
- <10> 이때, 상기 자막 표시 단계에서, 상기 자막은 상기 자막이 표시될 영역의 색상과 보색관계에 있는 색상으로 표시하는 것을 특징으로 한다.
- <11> 한편, 상술한 동영상의 자막 표시 방법은 상기 자막 표시 단계 이전에, 상기 동영상을 구성하는 음성정보를 분석하여 음성이 발생하는 공간적 위치를 결정하는 단계를 더 포함하고, 상기 자막 표시 단계에서, 상기 음성의 공간적 위치를 함께 이용하여 상기 자막이 표시될 영역을 결정하는 것을 특징으로 한다.
- <12> 이때, 상기 음성의 공간적 위치는 각 채널별로 출력되는 음성신호의 차이 및 음성주파수 대역 중 적어도 하나를 이용하여 결정되는 것을 특징으로 한다.

효과

- <13> 본 발명에 따르면 동영상을 구성하는 영상정보 및 음성정보를 이용하여 결정된 화자의 위치에 따라 동영상의 자막을 화자 주변에 표시함으로써 동영상에 대한 시청자의 시각적 몰입도를 높일 수 있다는 효과가 있다.

<14> 또한, 본 발명에 따르면 청각장애가 있는 시청자도 동영상 내에서 2인 이상의 화자가 대화하는 경우에 동영상의 화자의 대화 내용을 쉽게 이해할 수 있도록 하는 효과가 있다.

발명의 실시를 위한 구체적인 내용

<15> 이하 첨부된 도면을 참조하여 본 발명의 실시예에 대해 상세히 설명한다.

<16> 도 1은 본 발명의 일 실시예에 따른 동영상 재생 장치를 나타내는 블록도이다. 이러한 동영상 재생 장치(100)는 디스플레이(미도시)를 통해 동영상 및 동영상의 자막을 표시하는 것으로서, 도시된 바와 같이 수신부(110), 시간정보 검출부(120), 영상분석부(130), 화자 결정부(140), 자막위치 결정부(150), 및 동영상 출력부(160)를 포함한다.

<17> 수신부(110)는 저장매체(미도시)로부터 동영상 및 동영상의 자막 데이터를 수신하고, 수신된 동영상을 영상정보 및 음성정보로 분리하여 영상정보를 후술할 영상분석부(130)로 제공하고, 수신된 동영상의 자막 데이터를 후술할 시간정보 검출부(120)에 제공한다.

<18> 상술한 저장매체는 동영상 재생 장치(100)와 별개로 구성될 수도 있지만 동영상 재생 장치(100)내에 포함될 수도 있을 것이다. 또한, 상술한 실시예에 있어서 수신부(110)는 저장매체로부터 동영상 및 동영상의 자막 데이터를 수신하는 것으로 기재하였지만, 변형된 실시예에 있어서는 동영상 및 동영상의 자막 데이터를 인터넷을 통해 스트리밍 형식으로 다운로드 받을 수도 있을 것이다.

<19> 시간정보 검출부(120)는 수신부(110)로부터 제공되는 자막 데이터로부터 각 자막들이 표시될 시간정보 또는 영상 프레임 정보를 검출하여 영상분석부(130)에 제공한다. 여기서, 자막은 smi, idx, sub, srt, psb, ssa, ass, 및 usf 등과 같은 파일형식일 수 있으며, smi 파일 형식으로 제작된 자막의 일예가 아래에 기재되어 있다.

<20> <SYNC start=46085><P Class=KRCC>

<21> 멀더! 이것 좀 봐요.

<22> <SYNC start=48341><P Class=KRCC>

<23> 왜그래요? 스킬리.

<24> 위의 예에서 "멀더! 이것 좀 봐요."라는 자막은 46.085초부터 표시됨을 알 수 있고, "왜그래요? 스킬리."라는 자막은 48.341초부터 표시됨을 알 수 있다. 이러한 자막 데이터로부터 시간정보 검출부(120)는 "멀더! 이것 좀 봐요."라는 자막이 표시될 시간정보인 46.085초~48.341초라는 정보를 획득하게 되는 것이다.

<25> 한편, 상술한 실시예에 있어서는 자막 데이터로부터 시간정보를 추출하는 것으로 기재하였지만 해당 자막이 표시될 프레임 번호가 자막 데이터에 포함되어 있는 경우에는 프레임 번호를 직접 추출하여 영상분석부(130)로 제공할 수도 있을 것이다.

<26> 영상분석부(130)는 시간정보 검출부(120)로부터 제공된 시간정보에 상응하는 각 영상프레임으로부터 얼굴 및 얼굴 위치를 검출하여 화자 결정부(140)에 제공한다. 영상분석부(130)는 다양한 방법을 이용하여 얼굴을 검출할 수 있는데, 일 실시예에 있어서 얼굴 검출부는 스킨컬러 추출방법을 이용하여 얼굴을 검출할 수 있다.

<27> 스킨컬러 추출 방법을 이용하는 경우, 영상분석부(130)는 영상 프레임 내에서 피부색과 임계치 이내의 색편차 값을 갖는 영역을 흰색으로 표시하고, 나머지 영역을 검은색으로 표시한 후 흰색으로 표시된 부분에서 눈을 찾아내게 된다. 이때, 눈은 얼굴에서 기하학적으로 움푹 패인 형상이며, 피부색과는 확연히 다른 색이므로 흰색 영역 내에서 쉽게 찾아낼 수 있게 된다. 이러한 방법에 의하는 경우 가운데 검은 점이 2개 찍힌 흰색 영역이 얼굴로 결정되게 된다.

<28> 한편, 스킨컬러 추출방식을 이용하는 경우 발생될 수 있는 단점(예컨대, 옆 모습이라 던지 안경을 쓴 얼굴을 검출하지 못할 수 있다는 점)을 보완하기 위해 움직임에 기반한 예지 차영상을 이용하여 얼굴의 윤곽을 찾아내는 방식을 이용하거나, 눈의 깜빡임을 이용하여 얼굴을 검출하는 방식을 함께 이용할 수도 있다.

<29> 이러한 방법 이외에도 얼굴의 지역적 특성(local features)에 근거한 방법 또는 얼굴 전체 형상(template based)에 근거한 방법 등과 같은 다양한 방법을 이용하여 각 영상 프레임으로부터 얼굴을 검출할 수도 있을 것이다.

<30> 화자결정부(140)는 영상분석부(130)에 의해 검출된 각 영상 프레임 내의 얼굴을 이용하여 자막에 해당하는 각

영상 프레임 내에서 화자를 결정한다. 동영상은 대개 1초에 24~30개의 영상프레임으로 구성되고, 각 영상 프레임에는 화자 이외의 다른 캐릭터들도 등장할 수 있기 때문에, 영상분석부(130)에 의해 검출된 얼굴이 모두 화자의 얼굴이라고 할 수는 없다. 따라서, 화자결정부(140)는 영상분석부(130)에 의해 검출된 얼굴 중 화자의 얼굴을 검출함으로써 현재 프레임 내에서 말을 하고 있는 화자를 결정하는 것이다.

- <31> 화자결정부(140)가 영상분석부(130)에 의해 검출된 각 영상 프레임 내의 얼굴을 이용하여 화자를 결정하는 방법을 도 2를 이용하여 보다 구체적으로 설명한다.
- <32> 화자결정부(140)는 영상분석부(130)에 의해 검출된 각 영상 프레임 별 얼굴들 가운데 특징점이 변경되는 얼굴을 화자의 얼굴로 결정함으로써 해당 영상 프레임에 대한 화자를 결정한다. 일 실시예에 있어서, 특징점은 얼굴 중 눈 아래의 부위, 예컨대, 입(210)이나 턱 주변의 영역(212)일 수 있다. 이는 일반적으로 사람이 말을 할 때, 입 또는 턱 주변의 영역(212)이 움직이기 때문이다.
- <33> 예컨대, 도 2에 도시된 바와 같이, 해당 자막이 표시될 시간정보에 상응하는 영상 프레임이 n 프레임 및 n+1 프레임인 경우 n+1 프레임에서 얼굴(200)의 특징점(210, 212)이 변경되었기 때문에, 도 2의 얼굴(200)은 화자의 얼굴로 결정하게 되는 것이다.
- <34> 예컨대, n 프레임에서 검출된 얼굴의 폭과 높이가 200 X 200의 크기인 경우, 검출된 얼굴을 폭 30개와 높이 30개로 균등하게 분할한 후, n+1 프레임에서 검출된 얼굴도 동일하게 균등 분할하여 양 프레임을 서로 겹쳐주었을 때, n+1 프레임 내에서 눈 밑의 부분의 영역들의 색상이 n 프레임 중 동일한 영역의 색상과 다른 얼굴을 화자로 결정하게 되는 것이다.
- <35> 한편, 화자결정부(140)는 화자가 결정되는 경우, 화자로 결정된 얼굴의 위치를 후술할 자막위치 결정부(150)로 제공한다.
- <36> 자막위치 결정부(150)는 화자결정부(140)에 의해 결정된 화자의 얼굴위치를 이용하여 해당 자막이 표시될 위치를 결정한다. 구체적으로, 자막위치 결정부(150)는 먼저, 화자의 얼굴위치를 이용하여 자막이 표시될 후보 영역을 결정한다. 일 실시예에 있어서 자막이 표시될 후보 영역은 도 3에 도시된 바와 같이, 자막이 표시될 각 영상 프레임들로부터 검출된 화자 얼굴영역 범위(310)으로부터 소정 거리 이내의 영역, 예컨대, 화자의 얼굴영역 범위(310)으로부터 1줄 높이 이상 벗어나지 않거나, 수평축이 화자의 얼굴영역 범위(310)으로부터 2글씨 크기만큼을 벗어나지 않는 영역을 자막이 표시될 후보 영역으로 결정할 수 있다. 이때, 화자의 얼굴영역 범위(310)는 자막이 표시될 각 영상 프레임들로부터 추출된 화자의 얼굴영역이 각 영상 프레임에 따라 변경되는 영역을 의미한다. 도 3에서는 자막이 표시될 영상 프레임이 3개인 경우, 3개의 영상 프레임 각각에 포함된 화자의 얼굴이 변경되는 영역이 화자의 얼굴영역 범위(310)로 결정되었다.
- <37> 이때, 이러한 조건을 만족하는 후보영역(320)이 없을 경우에는 도 4에 도시된 바와 같이, 화자의 얼굴영역 범위(310)의 좌측 또는 우측의 영역 중 화자의 얼굴영역 범위(310)를 침범하지 않으면서 자막의 내용을 최소 1회 이상으로 줄 바꿈 하여 표시할 수 있는 지역을 후보영역으로 설정할 수 있다.
- <38> 한편, 이러한 후보 영역들 중 자막을 표시하였을 때 자막이 화면을 벗어나게 되는 영역이나 화자를 침범하게 되는 영역은 후보 영역에서 제외할 수 있다.
- <39> 이후, 자막위치 결정부(150)는 후보영역들 중 자막표시영역을 결정하는데, 일 실시예에 있어서, 자막위치 결정부(150)는 후보영역들 중 색편차가 가장 적고, 채도가 가장 높은 영역을 자막표시영역으로 결정할 수 있다. 만약, 결정된 자막표시영역의 채도값이 임계치 이하일 경우, 자막의 색상은 결정된 자막표시영역의 평균색상과 보색인 색상으로 결정할 수 있다.
- <40> 상술한 실시예에 있어서는 자막위치 결정부(150)가 화자 결정부(140)에 의해 결정된 화자의 얼굴영역 범위를 이용하여 자막위치를 결정하는 것으로 기재하였지만, 변형된 실시예에 있어서는 자막위치를 결정함에 있어서 동영상 구성하는 음성정보를 함께 이용하여 또는 음성정보 만을 이용하여 자막위치를 결정할 수도 있을 것이다. 이를 위해, 동영상 재생 장치(100)는 음성분석부(155)를 더 포함할 수 있다.
- <41> 음성분석부(155)는 상술한 수신부(110)로부터 동영상을 구성하는 음성정보를 수신하고, 시간정보 검출부(120)로부터 해당 자막이 표시될 시간정보를 수신하여 해당 시간정보에 상응하는 영상 프레임의 음성정보로부터 음성이 발생하는 공간적 위치를 결정하고, 결정된 공간적 위치를 상술한 화자결정부(140)로 제공한다.
- <42> 구체적으로, 음성분석부(155)는 먼저, 수신부(110)로부터 제공되는 음성정보의 규격을 분석한다. 일 실시예에 있어서, 음성정보의 규격은 각각의 채널별 음성신호가 다르게 출력되는 스테레오, 돌비 디지털, 돌비 써라운드,

돌비 프로로직, 디지털 써라운드, 또는 DTS등이 있을 수 있으며, 음성분석부(155)는 해당 동영상의 음성규격이 어떤 것에 해당하는지 여부를 분석하는 것이다.

- <43> 이후, 음성분석부(155)는 각 채널별 음성신호의 차이 또는 음성주파수 대역을 이용하여 음성이 발생하는 공간적 위치를 결정한다. 일반적으로 동영상의 음성정보는 해당 동영상을 구성하는 영상 프레임 내에서 발생하는 음원의 위치와 일치시켜주기 위해 좌우 2 채널이나 그 이상의 채널로 저장되므로, 이러한 각 채널의 차이를 이용한다면 음성의 공간적 위치를 유추할 수 있게 되는 것이다. 보다 구체적으로 음성신호의 차이를 이용하는 경우, 분석된 음성 규격이 스테레오 또는 돌비디지털이라고 가정하면 이러한 음성규격은 음성을 별도의 마이크를 이용하여 서로 다른 방향에서 녹음하는 것이므로, 채널 별로 음성의 크기가 상이할 수 밖에 없고, 채널 별로 음성이 마이크까지 도달하는데 걸리는 시간차이(위상차이)가 발생할 수 밖에 없으므로, 음성분석부(155)는 이러한 차이를 이용하여 음성발생위치의 좌우를 결정하게 된다. 일 실시예에 있어서 음성분석부(155)는 저음대역에서는 음성신호의 시간차이를 이용하여 음성의 좌우위치를 결정하고, 고음대역에서는 음성의 크기차이로 좌우위치를 결정하게 된다.
- <44> 한편, 음성 주파수 대역을 이용하는 경우, 음성분석부(155)는 음성의 스펙트럼 특성을 이용하여 음성 발생위치의 상하를 결정하게 된다.
- <45> 이러한 방법을 통해 결정된 음성의 공간적 위치가 자막위치 결정부(150)로 제공되면, 자막위치 결정부(150)는 결정된 음성의 공간적 위치를 함께 고려하여 자막표시영역을 결정하게 된다.
- <46> 한편, 자막위치 결정부(150)는 상술한 방법들을 통해서 자막이 표시될 영역을 결정할 수 없는 경우, 예컨대, 시간정보에 상응하는 영상 프레임에 화자가 등장하지 않아 화자의 얼굴 및 얼굴 위치를 검출할 수 없는 경우, 영상 프레임에 화자가 등장하지만 영상 프레임이 어두워 화자를 인식할 수 없는 경우, 또는 음성정보가 음성이 발생하는 공간적 위치를 결정하기에 적당하지 않은 음성규격인 경우에는 기본적으로 설정된 위치(화면의 우측 또는 하단과 같은 일편)를 자막이 표시될 위치로 결정할 수 있다.
- <47> 이러한 다양한 상황을 고려하여 자막위치 결정부(150)에 의해 결정된 자막위치를 표로 정리하면 도 5와 같다.
- <48> 다시 도 1을 참조하면, 동영상 출력부(160)는 동영상 및 동영상의 자막을 디스플레이(미도시)를 통해 출력하는 것으로서, 이때, 동영상의 자막은 상술한 자막위치 결정부(150)에 의해 결정된 자막표시영역에 출력된다. 즉, 동영상 출력부(160)는 도 6에 도시된 바와 같이, 영상 프레임 내에서 화자의 주위에 자막이 출력되도록 하는 것이다.
- <49> 이하에서는, 도 7을 참조하여 동영상의 자막 표시 방법을 설명하기로 한다.
- <50> 도 7은 본 발명의 일 실시예에 따른 동영상의 자막 표시 방법을 보여주는 플로우차트이다. 먼저, 동영상 및 동영상의 자막 데이터가 수신되면(S700), 동영상의 자막 데이터로부터 각 자막이 표시될 시간정보를 획득한다(710단계). 일 실시예에 있어서, 시간정보는 해당 자막이 표시될 시간에 대한 정보이거나, 해당 자막이 표시될 프레임 번호일 수 있다.
- <51> 다음으로, 동영상을 구성하는 영상 프레임들 중 해당 자막이 표시될 시간정보에 상응하는 영상 프레임들로부터 얼굴 및 얼굴 위치를 검출한다(7620). 일 실시예에 있어서, 얼굴검출은 스किन컬러 방법을 이용하여 수행될 수 있으며, 스किन컬러 방법과 함께 움직임에 기반한 에지 차영상을 이용하여 얼굴의 윤곽을 찾아내는 방식 또는 눈의 깜빡임을 이용하여 얼굴을 검출하는 방법을 이용할 수도 있다.
- <52> 이후, 검출된 얼굴의 특징점의 변경여부를 판단하여 시간정보에 상응하는 각 영상 프레임들에 해당하는 화자를 결정한다(S730). 여기서 화자를 결정하는 이유는 각 영상 프레임 내에는 화자 이외의 다른 캐릭터들도 등장할 수 있기 때문에, S620에서 검출된 모든 얼굴이 화자의 얼굴이라고 할 수는 없기 때문에, 검출된 얼굴 중 화자의 얼굴을 검출함으로써 현재 프레임 내에서 말을 하고 있는 화자를 결정하기 위한 것이다.
- <53> 화자를 결정함에 있어서, 검출된 얼굴 내에서 특징점이 변경되는 얼굴을 화자의 얼굴로 결정할 수 있고, 이때, 특징점은 얼굴 중 눈 아래의 부위, 예컨대, 입이나 턱 주변의 영역일 수 있다. 보다 구체적으로 설명하면, 해당 자막이 표시될 시간정보에 상응하는 영상 프레임이 n 프레임 및 n+1프레임인 경우, n 프레임 및 n+1 프레임에서 검출된 얼굴들의 차이를 산출하고, 산출된 차이를 통해 얼굴의 특징점에 해당하는 부분이 변경된 것으로 판단되는 얼굴을 화자의 얼굴로 결정하는 것이다.
- <54> 다음으로, S730에서 결정된 화자의 얼굴을 이용하여 자막이 표시될 시간정보에 상응하는 각 영상 프레임들 내에서 화자의 얼굴영역 범위를 결정한다(S740). 일 실시예에 있어서, 각 영상 프레임들에서 검출된 얼굴을 서로

겹쳤을 때 발생하는 전체 얼굴 영역을 화자의 얼굴영역 범위로 결정한다.

- <55> 이후, S740에서 결정된 화자의 얼굴영역 범위로부터 소정 거리 이내에 예컨대, 화자의 얼굴영역 범위로부터 1줄 높이 이상 벗어나지 않거나, 수평축이 화자의 얼굴영역 범위로부터 2글씨 크기만큼을 벗어나지 않는 영역을 자막이 표시될 후보 영역을 설정한다(S750). 이때, 이러한 조건을 만족하는 영역이 존재하지 않는 경우에는 화면의 일편(예컨대, 화면의 우측 또는 하단)을 후보 영역으로 설정할 수 있다.
- <56> 다음으로, 자막이 표시될 후보 영역 중 자막표시영역을 선택한다(S760). 일 실시예에 있어서, 자막표시영역은 후보 영역들 중 색편차가 가장 적고 채도가 가장 높은 영역으로 결정할 수 있다.
- <57> 한편, 상술한 실시예에 있어서는 각 영상 프레임 내에서 검출된 화자의 얼굴을 이용하여 자막의 위치를 결정하는 것으로 기재하였지만, 변형된 실시예에 있어서는 해당 시간정보에 상응하는 영상 프레임을 구성하는 음성정보를 분석함으로써 음성의 공간적 위치를 결정하고(S770), 결정된 음성의 공간적 위치를 함께 이용하여 자막표시영역을 결정할 수도 있을 것이다. 이때, 음성의 공간적 위치는 음성신호의 규격을 분석한 후, 분석된 음성 규격에 따라 각 채널 별 음성신호의 차이 및 음성신호의 주파수 스펙트럼 특성을 이용하여 결정할 수 있다.
- <58> 이후, 마지막으로, 해당 동영상 및 동영상의 자막을 출력하되, S770단계에서 결정된 자막표시영역에 해당 자막을 출력한다(S780). 이때, 자막은 자막표시영역의 색상과 보색관계에 있는 색상으로 출력되도록 할 수 있다.
- <59> 상술한 동영상의 자막 표시 방법은 다양한 컴퓨터 수단을 이용하여 수행될 수 있는 프로그램 형태로도 구현될 수 있는데, 이때 동영상의 자막 표시 방법을 수행하기 위한 프로그램은 하드 디스크, CD-ROM, DVD, 롬(ROM), 램(RAM), 또는 플래시 메모리와 같은 컴퓨터로 판독할 수 있는 기록 매체에 저장된다.
- <60> 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.
- <61> 그러므로, 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적인 것이 아닌 것으로 이해해야만 한다. 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 등가 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

도면의 간단한 설명

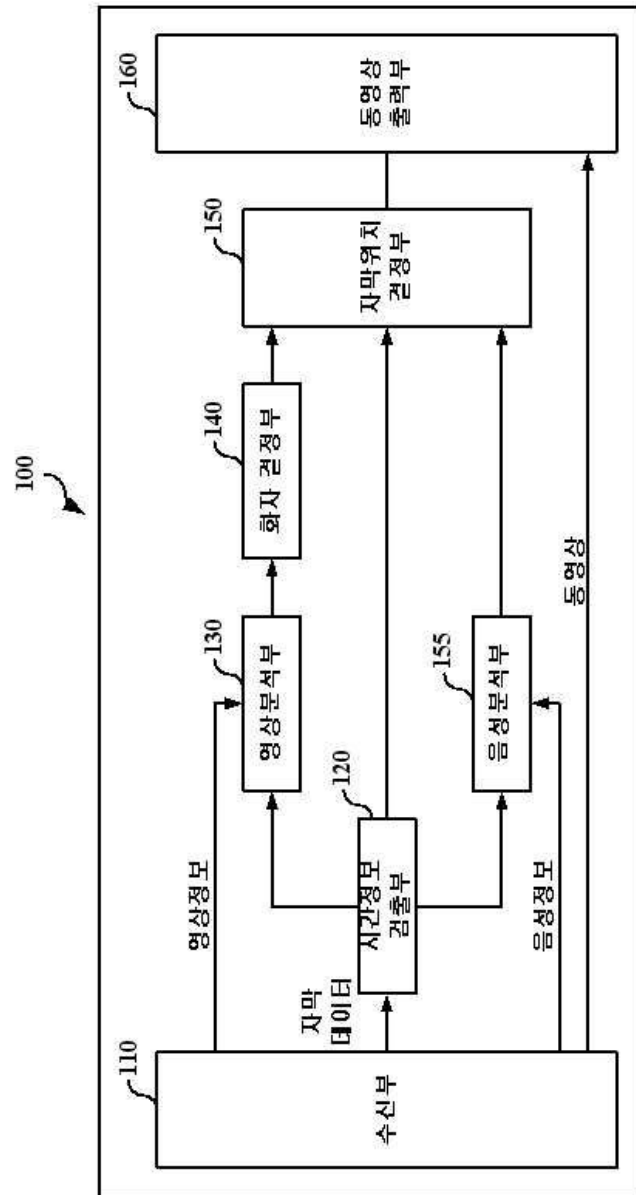
- <62> 도 1은 본 발명의 일 실시예에 따른 동영상 재생 장치의 개략적이 블록도.
- <63> 도 2는 검출된 얼굴의 특징점 변경여부를 이용하여 화자를 결정하는 방법을 보여주는 도면.
- <64> 도 3 및 도 4는 자막표시 후보영역을 보여주는 도면.
- <65> 도 5는 음성정보 및 영상정보에 따른 자막위치를 보여주는 도면.
- <66> 도 6은 본 발명의 일 실시예에 따라 자막을 표시한 동영상을 보여주는 도면.
- <67> 도 7은 본 발명의 일 실시예에 따른 동영상의 자막 표시 방법을 보여주는 플로우차트.

<도면의 주요부분에 대한 부호의 설명>

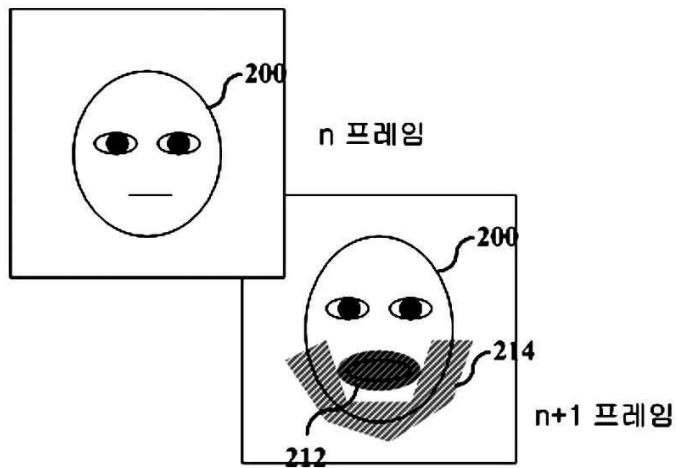
- <69> 100: 동영상 재생 장치 110: 수신부
- <70> 120: 시간정보 검출부 130: 영상분석부
- <71> 140: 화자결정부 150: 자막위치 결정부
- <72> 155: 음성분석부 160: 동영상 출력부

도면

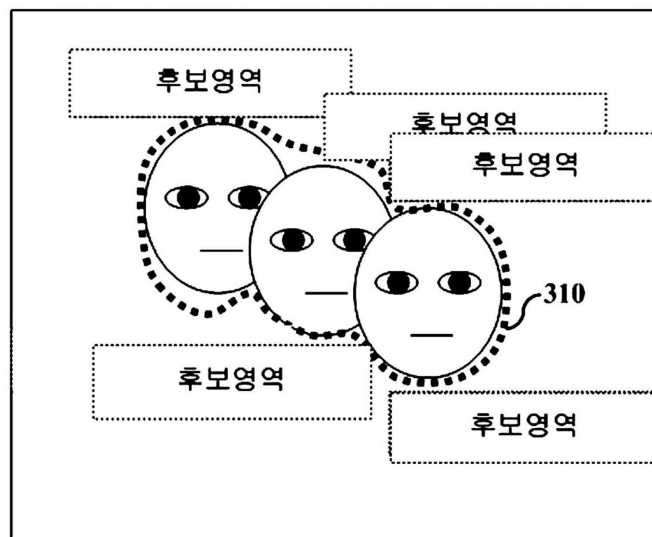
도면1



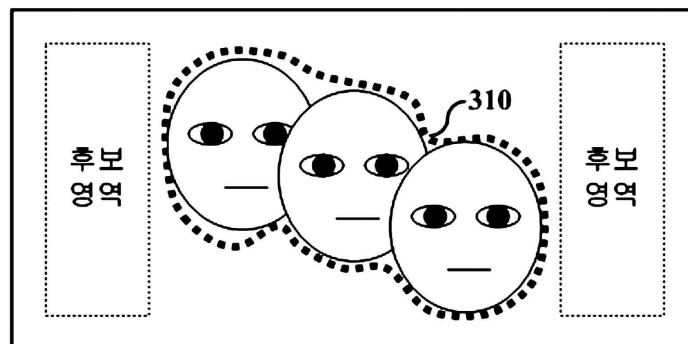
도면2



도면3



도면4



도면5

음성정보	영상정보	자막위치
○	○	화자의 얼굴영역 범위 및 음성의 공간적 위치고려하여 자막위치 결정
○	x	음성의 공간적 위치만을 고려하여 자막위치 결정
x	○	화자의 얼굴영역범위를 고려하여 자막위치 결정
x	x	기본으로 설정된 영역을 자막위치로 결정

도면6



도면7

