(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2014-0019335
(43) 공개일자 2014년02월14일

(51) 국제특허분류(Int. Cl.)
H04N 19/136 (2014.01) *H04N 19/142* (2014.01)
(21) 출원번호 10-2013-7022634
(22) 출원일자(국제) 2012년01월26일
심사청구일자 없음
(85) 번역문제출일자 2013년08월27일
(86) 국제출원번호 PCT/US2012/022720
(87) 국제공개번호 WO 2012/103332
국제공개일자 2012년08월02일
(30) 우선권주장
61/437,193 2011년01월28일 미국(US)
61/437,211 2011년01월28일 미국(US)

(71) 출원인
아이 이오, 엘엘씨
미국 캘리포니아 94301, 팰로앨토 스위트 넘버3, 유니버시티 애비뉴 165
(72) 발명자
게레로, 로돌포 바르가스
미국, 캘리포니아 94301, 팰러 앨토, 스위트 넘버3, 유니버시티 애비뉴 165
(74) 대리인
특허법인씨엔에스

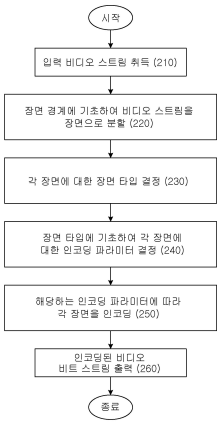
전체 청구항 수 : 총 34 항

(54) 발명의 명칭 **장면 타입에 기초한 비디오 스트림 인코딩**

(57) 요 약

비디오 스트림 또는 이미지를 위한 인코더가 본 명세서에 설명된다. 인코더는 입력 비디오 스트림을 취득하여 입력 비디오 스트림의 인스턴스를, 적어도 근사적으로, 복구하기 위하여 디코더에서 디코딩될 수 있는 인코딩된 비디오 스트림을 출력한다. 인코더는 먼저 장면 경계(scene boundary)를 식별하고 파라미터 세트를 이용하여 장면 경계 사이의 프레임을 인코딩함으로써 비디오 스트림을 인코딩한다. 적어도 2개의 상이한 장면 시컨스에 대하여, 상이한 파라미터 세트가 사용되어, 적응적이고 장면에 기초한 인코딩을 제공한다.

대 표 도 - 도2



특허청구의 범위

청구항 1

임의의 주어진 장면 타입을 인코딩하기 위하여 비디오 인코더에 의해 사용되는 복수의 인코더 파라미터의 하나 이상의 미리 정의된 세트를 각각 갖는 장면 타입을 이용하여 비디오 스트림을 인코딩하는 방법에 있어서,

입력 비디오 스트림을 취득하는 단계;

장면 경계에 기초하여 상기 입력 비디오 스트림을 복수의 장면으로 분할하는 단계로서, 각각의 장면은 시간적으로 인접한 복수의 이미지 프레임을 포함하고, 주어진 장면 경계는 상기 입력 비디오 스트림에서 2개의 시간적으로 인접한 이미지 프레임의 관련성에 따라 결정되는, 상기 분할하는 단계;

상기 복수의 장면의 각각에 대하여 장면 타입을 결정하는 단계; 및

상기 장면 타입에 따라 상기 복수의 장면의 각각을 인코딩하는 단계

를 포함하는,

비디오 스트림 인코딩 방법.

청구항 2

제1항에 있어서,

각각의 장면 타입은 하나 이상의 기준에 기초하여 결정되고,

상기 하나 이상의 기준은,

상기 입력 비디오 스트림의 타임라인에서의 주어진 장면의 위치;

상기 주어진 장면의 길이;

상기 주어진 장면에서의 모션 예측;

상기 주어진 장면에서의 이전 장면으로부터의 유효 차이;

상기 주어진 장면의 스펙트럼 데이터 크기;

상기 주어진 장면에서의 광학 문자 인식; 또는

상기 주어진 장면의 스크린 플레이 구조 정보

를 포함하는,

비디오 스트림 인코딩 방법.

청구항 3

제1항에 있어서,

상기 장면 타입을 결정하는 단계는, 얼굴 인식을 활용하는 단계를 더 포함하는,

비디오 스트림 인코딩 방법.

청구항 4

제2항에 있어서,

상기 스크린 플레이 구조 정보는 상대적인 관심 파라미터를 포함하고,

상기 상대적인 관심 파라미터는 상기 주어진 장면을 포함하는 상기 입력 비디오 스트림의 세그먼트에 대하여 예상되는 시청자 관심의 상대적인 양의 미리 결정되는 예측을 근사화하는,

비디오 스트림 인코딩 방법.

청구항 5

제2항에 있어서,

상기 스크린 플레이 구조 정보는,

시간 범위 정의;

상기 주어진 장면으로부터의 문자 정보;

상기 주어진 장면과 관련된 오디오 콘텐츠;

상기 주어진 장면과 관련된 폐쇄 자막 방송 정보; 또는

상기 주어진 장면과 관련된 메타 데이터

중 하나 이상을 포함하는,

비디오 스트림 인코딩 방법.

청구항 6

제1항에 있어서,

주어진 장면 타입은,

빠른 모션;

정지;

토킹 헤드(talking head);

문자;

대부분의 블랙 이미지;

짧은 장면;

스크롤 크레디트(scroll credit);

타이틀 장면;

기타; 또는

디폴트

중 하나 이상을 포함하는,

비디오 스트림 인코딩 방법.

청구항 7

제1항에 있어서,

제1 이미지 프레임이 상기 입력 비디오 스트림의 타임라인에서 제2 이미지 프레임에 대하여 하나 이상의 인접한 위치를 가질 때, 상기 제1 이미지 프레임이 상기 제2 이미지 프레임에 시간적으로 인접하다는 것을 결정하는 단

계를 더 포함하는,

비디오 스트림 인코딩 방법.

청구항 8

제1항에 있어서,

상기 입력 비디오 스트림에서 2개의 시간적으로 인접한 이미지 프레임의 관련성을 결정하는 것은,

각 이미지 프레임의 하나 이상의 고주파수 요소를 스케일링하는 단계;

각 이미지 프레임의 상기 하나 이상의 고주파수 요소를 제거하는 단계;

시간적으로 인접한 이미지 프레임 사이의 차이를 결정하기 위하여 상기 이미지 프레임을 분석하는 단계로서, 점수가 상기 차이에 기초하여 계산되는, 상기 분석하는 단계; 및

상기 점수가 미리 설정된 한계를 초과할 때 상기 이미지 프레임 사이의 비관련성 정도를 식별하는 단계

를 포함하고,

상기 미리 설정된 한계 점수는 장면 변경이 발생하는 임계값인,

비디오 스트림 인코딩 방법.

청구항 9

제8항에 있어서,

상기 차이는 회귀 필터 또는 적응 필터 중 하나에 의해 추적되는,

비디오 스트림 인코딩 방법.

청구항 10

제1항에 있어서,

상기 미리 결정된 인코더 파라미터는,

모션 예측 범위 검색;

디블록킹 양 인자;

양자화기; 또는

기준 프레임 수

중 하나 이상을 포함하는,

비디오 스트림 인코딩 방법.

청구항 11

임의의 주어진 장면 타입을 인코딩하기 위하여 비디오 인코더에 의해 사용되는 복수의 인코더 파라미터의 하나 이상의 미리 정의된 세트를 각각 갖는 장면 타입을 이용하여 비디오 스트림을 인코딩하는 방법에 있어서,

입력 비디오 스트림을 취득하는 단계;

장면 전환이 발생하는 상기 입력 비디오 스트림 내의 위치를 표시하는 장면 경계 정보를 취득하는 단계로서, 장면 전환은 상기 입력 비디오 스트림에서 2개의 시간적으로 인접한 이미지 프레임의 관련성에 따라 결정되는, 상

기 장면 경계 정보를 취득하는 단계;

상기 장면 경계 정보에 기초하여 상기 입력 비디오 스트림을 복수의 장면으로 분할하는 단계로서, 각각의 장면은 시간적으로 인접한 복수의 이미지 프레임을 포함하는, 상기 분할하는 단계;

상기 복수의 장면의 각각에 대하여 장면 타입을 결정하는 단계; 및

상기 장면 타입에 따라 상기 복수의 장면의 각각을 인코딩하는 단계

를 포함하는,

비디오 스트림 인코딩 방법.

청구항 12

제11항에 있어서,

각각의 장면 타입은 하나 이상의 기준에 기초하여 결정되고,

상기 하나 이상의 기준은,

상기 입력 비디오 스트림의 타임라인에서의 주어진 장면의 위치;

상기 주어진 장면의 길이;

상기 주어진 장면에서의 모션 예측;

상기 주어진 장면에서의 이전 장면으로부터의 유효 차이;

상기 주어진 장면의 스펙트럼 데이터 크기;

상기 주어진 장면에서의 광학 문자 인식; 또는

상기 주어진 장면의 스크린 플레이 구조 정보

를 포함하는,

비디오 스트림 인코딩 방법.

청구항 13

제12항에 있어서,

상기 스크린 플레이 구조 정보는 상대적인 관심 파라미터를 포함하고,

상기 상대적인 관심 파라미터는 상기 주어진 장면을 포함하는 상기 입력 비디오 스트림의 세그먼트에 대하여 예상되는 시청자 관심의 상대적인 양의 미리 결정되는 예측을 근사화하는,

비디오 스트림 인코딩 방법.

청구항 14

제12항에 있어서,

상기 스크린 플레이 구조 정보는,

시간 범위 정의;

상기 주어진 장면으로부터의 문자 정보;

상기 주어진 장면과 관련된 오디오 콘텐츠;

상기 주어진 장면과 관련된 폐쇄 자막 방송 정보; 또는

상기 주어진 장면과 관련된 메타 데이터

중 하나 이상을 포함하는,

비디오 스트림 인코딩 방법.

청구항 15

제12항에 있어서,

상기 장면 타입을 결정하는 단계는, 얼굴 인식을 활용하는 단계를 더 포함하는,

비디오 스트림 인코딩 방법.

청구항 16

제11항에 있어서,

주어진 장면 타입은,

빠른 모션;

정지;

토킹 헤드;

문자;

스크롤 크레디트;

타이틀 장면;

대부분의 블랙 이미지; 또는

짧은 장면

중 하나 이상을 포함하는,

비디오 스트림 인코딩 방법.

청구항 17

제11항에 있어서,

제1 이미지 프레임이 상기 입력 비디오 스트림의 타임라인에서 제2 이미지 프레임에 대하여 하나 이상의 인접한 위치를 가질 때, 상기 제1 이미지 프레임이 상기 제2 이미지 프레임에 시간적으로 인접하는,

비디오 스트림 인코딩 방법.

청구항 18

제11항에 있어서,

상기 미리 결정된 인코더 파라미터는,

모션 예측 범위 검색;

디블로킹 양 인자;

양자화기; 또는

기준 프레임 수

중 하나 이상을 포함하는,

비디오 스트림 인코딩 방법.

청구항 19

임의의 주어진 장면 타입을 인코딩하기 위하여 비디오 인코더에 의해 사용되는 복수의 인코더 파라미터의 하나 이상의 미리 정의된 세트를 각각 갖는 장면 타입을 이용하여 비디오 스트림을 인코딩하는 비디오 인코딩 장치에 있어서,

입력 비디오 스트림을 취득하는 입력 모듈;

장면 경계에 기초하여 상기 입력 비디오 스트림을 복수의 장면으로 분할하는 비디오 처리 모듈로서, 각각의 장면은 시간적으로 인접한 복수의 이미지 프레임을 포함하고, 상기 입력 비디오 스트림에서 2개의 시간적으로 인접한 이미지 프레임의 관련성에 따라 주어진 장면 경계를 결정하는 상기 비디오 처리 모듈; 및

상기 장면 타입에 따라 상기 복수의 장면의 각각을 인코딩하는 비디오 인코딩 모듈

을 포함하고,

상기 비디오 처리 모듈은, 상기 복수의 장면의 각각에 대하여 장면 타입을 결정하는,

비디오 인코딩 장치.

청구항 20

제19항에 있어서,

상기 비디오 처리 모듈은 하나 이상의 기준에 기초하여 각각의 장면 타입을 결정하고,

상기 하나 이상의 기준은,

상기 입력 비디오 스트림의 타임라인에서의 주어진 장면의 위치;

상기 주어진 장면의 길이;

상기 주어진 장면에서의 모션 예측;

상기 주어진 장면에서의 이전 장면으로부터의 유효 차이;

상기 주어진 장면의 스펙트럼 데이터 크기;

상기 주어진 장면에서의 광학 문자 인식; 또는

상기 주어진 장면의 스크린 플레이 구조 정보

를 포함하는,

비디오 인코딩 장치.

청구항 21

제20항에 있어서,

상기 비디오 인코딩 장치에 의해 활용되는 상기 스크린 플레이 구조 정보는 상대적인 관심 파라미터를 포함하고,

상기 상대적인 관심 파라미터는 상기 주어진 장면을 포함하는 상기 입력 비디오 스트림의 세그먼트에 대하여 예상되는 시청자 관심의 상대적인 양의 미리 결정되는 예측을 근사화하는,

비디오 인코딩 장치.

청구항 22

제20항에 있어서,

상기 비디오 인코딩 장치에 의해 활용되는 상기 스크린 플레이 구조 정보는,

시간 범위 정의;

상기 주어진 장면으로부터의 문자 정보;

상기 주어진 장면과 관련된 오디오 콘텐츠;

상기 주어진 장면과 관련된 폐쇄 자막 방송 정보; 또는

상기 주어진 장면과 관련된 메타 데이터

중 하나 이상을 포함하는,

비디오 인코딩 장치.

청구항 23

제20항에 있어서,

상기 비디오 처리 모듈은 장면 타입을 결정하기 위하여 얼굴 인식을 활용하는,

비디오 인코딩 장치.

청구항 24

제19항에 있어서,

주어진 장면 타입은,

빠른 모션;

정지;

토킹 헤드(talking head);

문자;

대부분의 블랙 이미지;

짧은 장면;

스크롤 크레디트(scroll credit);

타이틀 장면;

기타; 또는

디폴트

중 하나 이상을 포함하는,

비디오 인코딩 장치.

청구항 25

제19항에 있어서,

상기 비디오 처리 모듈은, 제1 이미지 프레임이 상기 입력 비디오 스트림의 타입라인에서 제2 이미지 프레임에 대하여 하나 이상의 인접한 위치를 가질 때, 상기 제1 이미지 프레임이 상기 제2 이미지 프레임에 시간적으로 인접하다는 것을 더 결정하는,

비디오 인코딩 장치.

청구항 26

제19항에 있어서,

상기 입력 비디오 스트림에서 2개의 시간적으로 인접한 이미지 프레임의 관련성에 대한 상기 비디오 처리 모듈의 결정은,

각 이미지 프레임의 하나 이상의 고주파수 요소를 스케일링하는 것;

각 이미지 프레임의 상기 하나 이상의 고주파수 요소를 제거하는 것;

시간적으로 인접한 이미지 프레임 사이의 차이를 결정하기 위하여 상기 이미지 프레임을 분석하는 것으로서, 점수가 상기 차이에 기초하여 계산되는 것; 및

상기 점수가 미리 설정된 한계를 초과할 때 상기 이미지 프레임 사이의 비관련성 정도를 식별하는 것

을 포함하고,

상기 미리 설정된 한계 점수는 장면 변경이 발생하는 임계값인,

비디오 인코딩 장치.

청구항 27

제26항에 있어서,

상기 비디오 처리 모듈은 상기 차이를 추적하도록 회귀 필터 또는 적응 필터 중 하나를 활용하는,

비디오 인코딩 장치.

청구항 28

제19항에 있어서,

상기 비디오 인코딩 모듈에 의해 활용되는 상기 미리 결정된 인코더 파라미터는,

모션 예측 범위 검색;

양자화기; 또는

기준 프레임 수

중 하나 이상을 포함하는,

비디오 인코딩 장치.

청구항 29

임의의 주어진 장면 타입을 인코딩하기 위하여 비디오 인코더에 의해 사용되는 복수의 인코더 파라미터의 하나 이상의 미리 정의된 세트를 각각 갖는 장면 타입을 이용하여 비디오 스트림을 인코딩하는 비디오 인코딩 장치에 있어서,

입력 비디오 스트림을 취득하는 취득 수단;

장면 경계에 기초하여 상기 입력 비디오 스트림을 복수의 장면으로 분할하는 분할 수단으로서, 각각의 장면은 시간적으로 인접한 복수의 이미지 프레임을 포함하고, 상기 입력 비디오 스트림에서 2개의 시간적으로 인접한 이미지 프레임의 관련성에 따라 주어진 장면 경계를 결정하는, 상기 분할 수단;

상기 복수의 장면의 각각에 대하여 장면 타입을 결정하는 결정 수단으로서, 각 장면 타입은 상기 주어진 장면 타입을 인코딩하기 위하여 비디오 인코더에 의해 사용되는 복수의 미리 결정된 인코더 중 하나 이상과 관련되는, 상기 결정 수단; 및

상기 복수의 장면의 각각과 관련된 상기 장면 타입에 따라 결정된 상기 주어진 장면의 이전에 결정된 인코더 파라미터에 기초하여 상기 복수의 장면의 각각을 인코딩하는 인코딩 수단

을 포함하는,

비디오 인코딩 장치.

청구항 30

임의의 주어진 장면 타입을 인코딩하기 위하여 비디오 인코더에 의해 사용되는 복수의 인코더 파라미터의 하나 이상의 미리 정의된 세트를 각각 갖는 장면 타입을 이용하여 비디오 스트림을 인코딩하는 방법에 있어서,

입력 비디오 스트림을 취득하는 단계;

장면 경계에 기초하여 상기 입력 비디오 스트림을 복수의 장면으로 분할하는 단계로서, 각각의 장면은 시간적으로 인접한 복수의 이미지 프레임을 포함하고, 주어진 장면 경계가 상기 입력 비디오 스트림의 스크린 플레이 구조 정보에 따라 결정되는, 상기 분할하는 단계;

상기 복수의 장면의 각각에 대하여 장면 타입을 결정하는 단계; 및

상기 장면 타입에 따라 상기 복수의 장면의 각각을 인코딩하는 단계

를 포함하는,

비디오 스트림 인코딩 방법.

청구항 31

제30항에 있어서,

제1 이미지 프레임이 상기 입력 비디오 스트림의 타임라인에서 제2 이미지 프레임에 대하여 하나 이상의 인접한 위치를 가질 때, 상기 제1 이미지 프레임이 상기 제2 이미지 프레임에 시간적으로 인접하다는 것을 결정하는 단계를 더 포함하는,

비디오 스트림 인코딩 방법.

청구항 32

제30항에 있어서,

상기 스크린 플레이 구조 정보는 상대적인 관심 파라미터를 포함하고,

상기 상대적인 관심 파라미터는 상기 입력 비디오 스트림의 복수의 비디오 세크먼트의 각각에 대하여 예상되는 시청자 관심의 상대적인 양의 미리 결정된 예측을 근사화하고,

상기 복수의 비디오 세그먼트의 각각은 복수의 장면을 포함할 수 있는,

비디오 스트림 인코딩 방법.

청구항 33

제30항에 있어서,

상기 스크린 플레이 구조 정보는,

시간 범위 정의;

상기 주어진 장면으로부터의 문자 정보;

상기 주어진 장면과 관련된 오디오 콘텐츠;

상기 주어진 장면과 관련된 폐쇄 자막 방송 정보; 또는

상기 주어진 장면과 관련된 메타 데이터

중 하나 이상을 더 포함하는,

비디오 스트림 인코딩 방법.

청구항 34

제30항에 있어서,

주어진 장면 타입은,

액션 장면;

느린 움직임 장면;

타이틀 장면;

오프닝 장면;

크레디트 장면;

헤드샷(head-shot) 장면; 또는

대화 장면

중 하나 이상을 포함하는,

비디오 스트림 인코딩 방법.

명 세 서

기 술 분 야

[0001] [관련 출원에 대한 교차 참조]

[0002] 본 출원은 전문이 본 명세서에 참조로서 명시적으로 편입되는 2011년 1월 28일 출원된 미국 가특허 출원 No. 61/437,193 및 2011년 1월 28일 출원된 미국 가특허 출원 No. 61/437,211에 대한 우선권을 주장한다.

[0003] [기술분야]

[0004] 본 발명은 비디오 인코딩 기술에 관한 것으로, 더욱 상세하게는, 비디오 인코딩을 위한 비디오 인코딩 파라미터의 자동 선택에 관한 것이다.

배 경 기 술

[0005] 비디오 스트리밍이 일상적인 사용자들 사이에서 인기 및 사용이 계속 상승하고 있지만, 극복되어야 할 여러 가지 내재하는 한계가 있다. 예를 들어, 사용자들은 종종 비디오 스트림을 획득하기 위하여 제한적인 대역폭만을 갖는 인터넷을 통해 비디오를 시청하기 원한다. 예를 들어, 사용자들은 이동 전화 연결 또는 가정용 무선 연결을 통해 비디오 스트림을 획득하기 원할 수 있다. 일부 시나리오에서, 사용자들은 콘텐츠를 스풀링함으로써 충분한 대역폭의 부족을 보상한다(즉, 최종적인 시청을 위하여 로컬 스토리지에 콘텐츠를 다운로드한다). 이 방법은 여러 가지 단점으로 가득 차 있다. 먼저, 사용자는 실제의 "런타임(run-time)" 경험을 가질 수 없다 - 즉, 사용자는 프로그램을 시청할지를 결정할 때 이를 볼 수 없다. 대신에, 사용자는 프로그램을 시청하기 전에 콘텐츠가 스풀링되기 위하여 상당한 지연을 겪어야만 한다. 다른 단점은 스토리지의 유용성이다 - 제공자(provider) 또는 사용자는, 단기간이라 하더라도, 스풀링된 콘텐츠가 저장될 수 있다는 것을 보장하도록 스토리지 리소스에 대한 책임을 져야 하여, 고가의 스토리지 리소스의 불필요한 활용을 야기한다.

[0006] 비디오 스트림(통상적으로 이미지 부분과 오디오 부분을 포함한다)은, 특히 고해상도(예를 들어, HD 비디오)에서, 상당한 대역폭을 필요로 할 수 있다. 오디오는 통상적으로 훨씬 더 적은 대역폭을 필요로 하지만, 종종 고려될 필요가 여전히 있다. 하나의 스트리밍 비디오 접근 방식은 비디오 스트림을 아주 많이 압축하여 사용자가 런타임으로 또는 실질적으로 즉각적으로(즉, 실질적인 스풀링 지연을 겪지 않으면서) 콘텐츠를 시청할 수 있게 하도록 빠른 비디오 전달을 가능하게 하는 것이다. 통상적으로, 비가역 압축(lossy compression)(즉, 완전히 가역적이지 않은 압축)은 무손실 압축보다 더 많은 압축을 제공하지만, 너무 많은 비가역 압축은 바람직하지 않은 사용자 경험을 제공한다.

[0007] 디지털 비디오 신호를 전송하는데 요구되는 대역폭을 감소시키기 위하여, (비디오 데이터 압축의 목적으로) 디지털 비디오 신호의 데이터 레이트(data rate)가 실질적으로 감소될 수 있는 효율적인 디지털 비디오 인코딩을 사용하는 것이 잘 알려져 있다. 정보 처리 상호 운용(interoperability)을 보장하기 위하여, 비디오 인코딩 표준은 많은 전문적인 애플리케이션 및 소비자 애플리케이션에서 디지털 비디오의 채용을 용이하게 하는데 주요한 역할을 하여 왔다. 대부분의 영향력 있는 표준은 일반적으로 ITU-T(International Telecommunications Union) 또는 ISO/IEC(International Organization for Standardization/International Electrotechnical Committee)의 MPEG(Motion Pictures Experts Group) 15 위원회에 의해 개발되어 왔다. 권장 사항으로서 알려진 ITU-T 표준은 통상적으로 실시간 통신(예를 들어, 비디오 컨퍼런싱)에 목표를 두고 있는 반면, 대부분의 MPEG 표준은 저장(예를 들어, DVD(Digital Versatile Disc)) 및 방송(예를 들어, OVB(Digital Video Broadcast) 표준)에 대하여 최적화된다.

[0008] 현재, 대다수의 표준화된 비디오 인코딩 알고리즘은 하이브리드 비디오 인코딩에 기초한다. 하이브리드 비디오 인코딩 방법은 통상적으로 원하는 압축 이득을 획득하도록 여러 가지 상이한 무손실 압축 스킴 및 비가역 압축 스킴을 결합한다. 또한, 하이브리드 비디오 인코딩은 ITV-T 표준(H.261, H.263과 같은 H.26x 표준)과 ISO/IEC 표준(MPEG-1, MPEG-2 및 MPEG-4와 같은 MPEG-X 표준)의 근거이다. 최근의 진보된 대부분의 비디오 인코딩 표준은 현재 ITV-T 그룹 및 ISO/IEC MPEG 그룹의 합동 팀인 JVT(joint video team)에 의한 표준화 노력의 결과인 H.264/MPEG-4 AVC(advanced video coding)으로 표시되는 표준이다.

[0009] H.264 표준은 MPEG-2와 같은 확립된 표준으로부터 알려진 블록 기반 모션 보상 하이브리드 변환 코딩과 동일한 원리를 채용한다. 따라서, H.264 구문(syntax)은 화면(picture)-블록 헤더, 부분(slice)-블록 헤더 및 매크로-블록 헤더와 같은 보통의 헤더 계층과 모션-벡터, 블록-변환 계수, 양자화기 스케일(quantizer scale) 등과 같은 데이터로서 조직화된다. 그러나, H.264 표준은 비디오 데이터의 콘텐츠를 나타내는 VCL(Video Coding Layer)과 데이터의 포맷을 지정하고 헤더 정보를 제공하는 NAL(Network Adaptation Layer)을 분리한다.

[0010] 또한, H.264는 인코딩 파라미터에 대한 훨씬 증가된 선택을 허용한다. 예를 들어, 이는 16x16 매크로 블록의 더욱 정교한 분할 및 조작을 허용하여, 이에 의해 예를 들어, 모션 압축 과정이 4x4 크기와 같이 작은 매크로

블록의 세그먼트화에 수행될 수 있다. 또한, 샘플 블록의 모션 보상 예측을 위한 선택 과정이, 단지 인접한 화면 대신에, 저장된 이전에 디코딩된 다수의 화면을 포함할 수 있다. 단일 프레임 내에서의 인트라 코딩(intra coding)으로도, 동일한 프레임으로부터의 이전에 디코딩된 샘플을 이용하여 블록 예측을 형성하는 것이 가능하다. 또한, 모션 보상에 이어지는 결과에 따른 예측 오차는, 전통적인 8x8 크기 대신에, 4x4 블록 크기에 기초하여 변환되고 양자화될 수 있다. 또한, 블록 아티팩트를 감소시키는 인루프(in-loop) 디블로킹(deblocking) 필터가 사용될 수 있다.

[0011] H.264 표준은, 가능한 코딩 결정(coding decision) 및 파라미터의 개수를 확장하면서 비디오 데이터의 동일한 전체 구성을 이용한다는 점에서, H.262/MPEG-2 비디오 인코딩 구문의 확대집합(superset)으로 고려될 수 있다. 다양한 코딩 결정을 갖는 결과는 비트 레이트와 화질 사이의 양호한 트레이드 오프가 획득될 수 있다는 것이다. 그러나, H.264 표준이 블록 기반의 코딩의 전형적인 아티팩트를 상당히 감소시킬 수 있다는 것이 일반적으로 인정되지만, 다른 아티팩트를 두드러지게 할 수도 있다. H.264가 다양한 코딩 파라미터에 대한 증가된 개수의 가능한 값을 허용하고 이에 따라 인코딩 과정을 개선하기 위한 잠재성을 증가시킨다는 사실은, 또한 비디오 인코딩 파라미터의 선택에 대한 민감성을 증가시키는 결과를 초래한다.

[0012] 다른 표준과 유사하게, H.264 표준은 비디오 인코딩 파라미터를 선택하기 위한 규범적인 절차를 특정하지 않지만, 기준 구현(reference implementation)을 통해 코딩 효율, 비디오 품질 및 구현의 실현 가능성 사이의 적합한 트레이드 오프를 획득하는 것과 같이 비디오 인코딩 파라미터를 선택하는데 사용될 수 있는 다수의 기준을 기술한다. 그러나, 기술된 기준은 모든 종류의 콘텐츠 및 애플리케이션에 적합한 코딩 파라미터의 최적의 또는 적합한 선택을 항상 제공할 수 없다. 예를 들어, 기준은 비디오 신호의 특성에 최적이거나 바람직한 비디오 인코딩 파라미터의 선택을 제공하지 않거나 또는 기준은 현재의 애플리케이션에 적합하지 않은 인코딩된 신호의 특성을 획득하는데 기초할 수 있다.

[0013] 따라서, 비디오 인코딩을 위한 개선된 시스템이 유익할 것이다.

[0014] 관련 기술에 대한 전술한 예와 그와 관련된 한정은 예시적이고 비독점적인 것으로 의도된다. 관련 기술의 다른 한정은 본 명세서를 읽고 도면을 연구함에 따라 명백하게 될 것이다.

발명의 내용

[0015] 본 교시 내용은 비디오 스트림을 인코딩하기 위한 다양한 방법, 시스템 및 인코더를 고려한다. 비디오 스트림을 인코딩하기 위한 인코더가 본 명세서에 소개된다. 인코더는 입력 비디오 스트림을 취득하여 입력 비디오 스트림의 인스턴스를, 적어도 근사적으로, 복구하기 위하여 디코더에서 디코딩될 수 있는 인코딩된 비디오 스트림을 출력한다. 본 발명의 실시예들에서, 인코더는 먼저 장면 경계(scene boundary)를 식별하고 파라미터 세트를 이용하여 장면 경계 사이의 프레임을 인코딩함으로써 비디오 스트림 또는 저장된 시컨스(즉, 하나 이상의 이미지 프레임을 포함하는 "장면 시컨스")를 인코딩한다.

[0016] 본 발명의 일 실시예에서, 장면 변경은, 장면 변경의 각 측에서의 2개의 프레임 사이의 차이가 동일한 시야각(viewing angle)으로부터 다른 프레임 샷(frame shot)만큼 압축가능하지 않도록 카메라가 갑자기 하나의 시야각으로부터 다른 시야각으로 변경하는 비디오 스트림에서 식별될 수 있다. 적어도 2개의 상이한 장면 시컨스에 대하여, 상이한 파라미터 세트가 사용되어, 이는 적응적이고 장면 기반의 인코딩을 제공한다.

[0017] 본 발명의 내용은 아래의 발명을 실시하기 위한 구체적인 내용에서 더 설명되는 개념 중에서 선택된 것을 간단한 형태로 소개하기 위하여 제공된다. 본 발명의 내용은 청구된 대상의 주요 특징 또는 필수적인 특징을 식별하려고 의도되지 않으며, 청구된 대상의 범위를 제한하는데 사용되려고 의도되지 않는다.

도면의 간단한 설명

[0018] 본 발명의 하나 이상의 실시예들은 예로서 예시되며 유사한 도면 부호가 유사한 구성 요소를 나타내는 첨부된 도면에 의해 한정되지 않는다.

도 1은 인코더의 일례를 도시한다.

도 2는 입력 비디오 스트림을 인코딩하는 방법의 단계들을 예시한다.

도 3은 입력 비디오 스트림을 복수의 장면 시컨스로 분할하는 방법의 단계들을 예시한다.

도 4는 장면 시컨스의 장면 타입을 결정하는 방법의 단계들을 예시한다.

도 5a 내지 5f는 프레임에서의 모션 예측을 설명하는 블록도이다.

도 6은 본 명세서에 설명된 소정의 기술을 구현하는 인코더를 구현하는데 사용될 수 있는 처리 시스템의 블록도이다.

발명을 실시하기 위한 구체적인 내용

[0019] 본 발명의 다양한 양태가 설명될 것이다. 다음의 설명은 이러한 예들의 설명을 완전히 이해하고 가능하게 하기 위한 구체적인 상세를 제공한다. 그러나, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 이러한 상세의 많은 부분 없이 본 발명이 실시될 수 있다는 것을 이해할 것이다. 또한, 널리 알려진 일부 구조 또는 기능은 관련 설명을 불필요하게 흐리게 하는 것을 방지하기 위하여 상세히 도시되거나 설명되지 않을 수 있다. 도면이 기능적으로 분리된 컴포넌트를 도시하더라도, 이러한 도시는 단지 예시적인 목적을 위한 것이다. 이 도면에서 묘사된 컴포넌트들이 임의로 결합되거나 개별 컴포넌트로 분할될 수 있다는 것은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 명백할 것이다.

[0020] 아래에 제공된 설명에서 사용된 용어는, 본 발명의 소정의 특정 예에 대한 상세한 설명과 함께 사용되고 있더라도, 최광의의 타당한 방식으로 해석되도록 의도된다. 어떤 용어는 아래에서 강조될 수 있다; 그러나, 임의의 제한된 방식으로 해석되도록 의도되는 임의의 용어는 본 발명을 실시하기 위한 구체적인 내용 항목에서 명시적이고 구체적으로 그와 같이 정의될 것이다.

[0021] 본 명세서에서 "일 실시예", "하나의 실시예" 등에 대한 참조는, 설명되는 특정 특징, 구조 또는 특성이 본 발명의 적어도 하나의 실시예에 포함된다는 것을 의미한다. 본 명세서에서 이러한 문구의 사용은 반드시 모두 동일한 실시예를 말하는 것은 아니다.

[0022] 본 발명의 일 실시예에서, 인코더는 입력 비디오 스트림을 취득하여 입력 비디오 스트림의 인스턴스를, 적어도 근사적으로, 복구하기 위하여 디코더에서 디코딩될 수 있는 인코딩된 비디오 스트림을 출력하도록 제공된다. 인코더는, 입력 비디오 스트림을 취득하는 입력 모듈; 비디오 스트림을 장면 경계에 기초하여 복수의 장면으로 분할하기 위한 비디오 처리 모듈로서, 입력 비디오 스트림에서 2개의 시간적으로 인접한 이미지 프레임의 관련성에 따라 주어진 장면 경계를 결정하는 비디오 처리 모듈; 복수의 장면의 각각에 대한 장면 타입을 더 결정하는 상기 비디오 처리 모듈 - 각각의 장면 타입은 주어진 장면 타입을 인코딩하기 위하여 비디오 인코더에 의해 사용되는 미리 결정된 복수의 인코더 파라미터 중 하나 이상과 관련됨 -; 및 복수의 장면의 각각과 관련된 장면 타입에 따라 복수의 장면의 각각을 인코딩하는 비디오 인코딩 모듈을 포함한다.

[0023] 이러한 방식으로, 인코더는 인코딩되는 입력 비디오 스트림에서 각각의 장면에 가장 적합한 품질로 입력 비디오 스트림을 인코딩할 수 있다.

[0024] 도 1은 본 발명의 일 실시예에 따른 인코더(100)의 일례를 도시한다. 인코더(100)는 입력 비디오 스트림(110)을 취득하여 입력 비디오 스트림(110)의 인스턴스를, 적어도 근사적으로, 복구하기 위해 디코더에서 디코딩될 수 있는 인코딩된 비디오 스트림(120)을 출력한다. 인코더(100)는 입력 모듈(102), 비디오 처리 모듈(104) 및 비디오 인코딩 모듈(106)을 포함한다. 인코더(100)는 하드웨어, 소프트웨어 또는 임의의 적합한 조합으로 구현될 수 있다. 인코더(100)는 파라미터 입력 모듈, 파라미터를 저장하기 위한 메모리 등과 같은 다른 컴포넌트를 포함할 수 있다. 인코더(100)는 본 명세서에 구체적 설명되지 않은 다른 비디오 처리 기능들을 수행할 수 있다.

[0025] 입력 모듈(102)은 입력 비디오 스트림(110)을 취득한다. 입력 비디오 스트림(110)은 임의의 적합한 형태를 취할 수 있으며, 메모리와 같은 다양한 적합한 소스 중 어느 것으로부터 또는 심지어 생방송 공급으로부터 유래할 수 있다.

[0026] 비디오 처리 모듈(104)은 입력 비디오 스트림(110)을 분석하여, 비디오 스트림(110)을 복수의 장면으로, 이러한 복수의 장면의 각각에 대한 해당하는 비디오 인코딩 파라미터와 함께 분할한다. 일 실시예에서, 비디오 처리 모듈(104)은 장면 경계에 기초하여 비디오 스트림을 복수의 장면으로 분할하고, 장면 경계는 입력 비디오 스트림에서 2개의 시간적으로 인접한 이미지 프레임의 관련성에 따라 결정된다. 비디오 처리 모듈(104)은 복수의 장면의 각각에 대한 장면 타입을 더 결정한다. 마지막으로, 비디오 처리 모듈(104)은 각 장면 타입을 미리 결정된 복수의 인코더 파라미터 중 하나 이상과 관련시켜 각 장면을 인코딩하기 위하여 비디오 인코더(106)에 의해 사용되는 비디오 인코딩 파리미터를 결정한다. 파라미터는 각 장면 타입에 대하여 미리 정의되거나, 비디오 스트림 처리 동안 계산되고 그리고/또는 조정될 수 있다. 비디오 인코딩 모듈(106)은 비디오 처리 모듈(104)로부터 복수의 장면 및 해당하는 비디오 인코딩 파라미터를 공급받아 해당하는 인코딩 파라미터에 따라 복수의 장면의 각각을 인코딩하여 인코딩된 비디오 스트림(120)을 출력한다.

[0027] 도 2는 입력 비디오 스트림을 인코딩하는 방법(200)의 단계들을 도시한다. 본 방법(200)은 입력 비디오 스트림의 인스턴스를, 적어도 근사적으로, 복구하기 위해 디코더에서 디코딩될 수 있는 인코딩된 비디오 비트 스트림으로 입력 비디오 스트림을 인코딩한다. 단계 210에서, 본 방법은 인코딩될 입력 비디오 스트림을 취득한다. 단계 220에서, 비디오 스트림은 장면 경계에 기초하여 복수의 장면으로 분할된다. 여기에서, 도 3을 참조하여 아래에서 더욱 상세히 논의되는 바와 같이, 본 방법은 입력 비디오 스트림에서 2개의 시간적으로 인접한 이미지 프레임의 관련성에 따라 장면 경계를 결정한다. 그러나, 다양한 다른 적합한 메커니즘 중 어느 것도 장면 타입들을 구별하는데 활용될 수 있다. 그 다음, 단계 230에서, 과정은 복수의 장면의 각각에 대한 장면 타입을 결정한다. 단계 240에서, 과정은, 역시 아래에서 더욱 상세히 논의되는 바와 같이, 각각의 장면 타입을 적합한 미리 결정된 인코더 파라미터로 매핑함으로써 각 장면을 인코딩하기 위하여 비디오 인코딩 파라미터를 결정한다. 단계 250에서, 과정은 각 장면의 해당하는 비디오 인코딩 파라미터(예를 들어, 단계 240에서 결정되는 바와 같은)에 따라 장면을 인코딩한다. 단계 260에서, 과정은 인코딩된 비디오 비트 스트림을 출력한다.

[0028] 전술한 과정은 다음 부분에서 더욱 상세하게 상술된다. 입력 비디오 스트림은 통상적으로 여러 이미지 프레임을 포함한다. 각 이미지 프레임은 통상적으로 입력 비디오 스트림에서 별개의 "시간 위치"에 기초하여 식별될 수 있다. 실시예에서, 입력 비디오 스트림은 나누어서 또는 별개의 세그먼트로 인코더에 사용될 수 있는 스트림일 수 있다. 이러한 경우에, 인코더는 심지어 전체 입력 비디오 스트림을 취득하기 전에 롤링(rolling) 기반으로 인코딩된 비디오 비트 스트림을 스트림으로서 (예를 들어, HDTV와 같은 최종 소비자 장치로) 출력한다.

[0029] 실시예에서, 입력 비디오 스트림과 인코딩된 비디오 비트 스트림은 스트림 시컨스로서 저장된다. 여기에서, 인코딩은 미리 수행될 수 있고, 그 다음, 인코딩된 비디오 스트림은 나중에 소비자 장치로 스트리밍될 수 있다. 여기에서, 인코딩은 소비자 장치로 스트리밍되기 전에 전체 비디오 스트림에 완전히 수행된다. 또한, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 고려될 수 있는 바와 같이, 비디오 스트림의 사전, 사후 또는 "인라인(inline)" 인코딩 혹은 그 조합의 다른 예가 본 명세서에서 소개된 기술과 함께 고려될 수 있다

는 것이 이해된다.

[0030] 실시예에서, 입력 비디오 스트림에서의 장면 경계는 먼저 각 이미지 프레임에 존재하는 임의의 고주파수 요소를 스케일링하여 제거함으로써 결정된다. 다음으로, 입력 비디오 스트림의 타임라인에서 시간적으로 서로 인접한 2개의 이미지 프레임 사이의 차이가 결정된다. 일부 경우에, 예를 들어, 2개의 이미지 사이의 차이는 회귀(recursive) 필터 또는 적응(adpative) 필터를 이용하여 구별될 수 있다. 계산된 차이가 장면 변경을 알리는 설정된 임계값을 초과하면, 2개의 이미지 프레임은 2개의 상이한 장면 시컨스의 일부로 결정되고, 따라서 장면 경계가 2개의 이미지 프레임 사이에 구축된다. 시간적으로 인접한 이미지 프레임 사이의 장면 경계 결정 과정을 반복함으로써, 입력 비디오 스트림은, 예를 들어, 정렬된 장면 시컨스 세트로 분할될 수 있다.

[0031] 실시예에서, 도 2를 참조하여 예시된 바와 같이, 장면 타입이 인코딩 과정과 함께 복수의 장면 시컨스의 각각에 대하여 결정될 수 있다. 일부 경우에, 장면 시컨스 타입은 다음의 장면 시컨스 파라미터 중 하나 이상을 활용하여 결정될 수 있다: (i) 입력 스트림의 타임라인에서의 장면 시컨스의 위치; (ii) 장면 시컨스의 길이; (iii) 이의 위치 벡터 예측; (iv) 이전 프레임으로부터의 장면 시컨스의 유효 차이; (v) 장면 시컨스의 스펙트럼 데이터 크기; (vi) 광학 문자 인식을 이용하여 결정된 장면의 문자 콘텐츠; (vii) 스크린 플레이(screenplay) 구조 정보에 기초한 장면의 스크린 플레이 속성 등. 또한, 일부 경우에, 얼굴 인식이 장면 시컨스가 개인의 얼굴을 포함하는지 여부를 결정하도록 장면 타입의 결정에 사용될 수 있다.

[0032] 주어진 장면 타입은, 예를 들어, "빠른 모션", "정지(static)", "토킹 헤드(talking head)", "문자", "스크롤 크레디트(scroll credit)", "대부분의 블랙 이미지", "5 프레임 이하의 짧은 장면" 등을 포함할 수 있다. 일부 경우에, 장면 시컨스는 특정 장면 타입에 할당될 수 없다. 다른 경우에, 할당된 장면 시컨스는, "기타(miscellaneous)", "알 수 없음", "디폴트" 등의 장면 타입을 포함할 수 있다.

[0033] 실시예에서, 장면 타입이 할당되면, 장면 시컨스가 인코딩된다. 일부 경우에, 이러한 인코딩은 소프트웨어 또는 하드웨어 명령어 세트에 따라 파라미터화된 인코딩 과정을 실행함으로써 수행될 수 있다. 여기에서, 일부 경우에, 매우 최적화된 파라미터 세트가 장면 타입에 따라 인코딩의 상세를 제어하는데 활용될 수 있다. 복수의 파라미터가 장면 타입 데이터베이스 또는 다른 데이터 구조나 기계 학습 시스템에 저장될 수 있다. 예시된 예에서, 메모리에 저장되고 인코더에 의해 액세스 가능한 데이터베이스는 표 1에 예시된 구조를 가질 수 있다. 인코딩을 위해 사용되지만, 장면 타입 데이터베이스에 구체적으로 설정되지 않은 파라미터는 인코딩 과정의 초기에 결정된 디폴트 파라미터값을 활용할 수 있다. 일부 경우에, 디폴트 파라미터값은 입력 비디오 스트림을 인코딩하는데 활용되는 인코딩 표준에 의해 추천된 값에 기초하여 결정될 수 있다.

표 1

[0034]

장면 타입	모션 예측 범위 검색	양자화기	레퍼런스 프레임 수	파라미터 4 내지 N...
빠른 모션	높은 모션 예측 범위 검색	높은 디블로킹을 갖는 낮은 양자화기	낮은 레퍼런스 프레임 수	
정지	낮은 모션 예측 범위 검색	높은 양자화기	높은 레퍼런스 프레임 수	
토킹 헤드	낮은 모션 예측 범위 검색	얼굴 영역을 강조한 중간 범위 양자화기	낮은 레퍼런스 프레임 수	
문자	낮은 모션 예측 범위 검색	낮은 양자화기, 낮은 디블로킹	중간 레퍼런스 프레임 수	
대부분 블랙	낮은 모션 예측 범위 검색	매우 낮은 양자화기, 디블로킹 없음	레퍼런스 프레임 수 없음	
짧은 장면	높은 모션 예측 범위 검색	낮은 양자화기	낮은 레퍼런스 프레임 수	

디폴트	중간 모션 예측 범위 검색	중간 양자화기	중간 레퍼런스 프레임 수	
...				

[0035] 도 3은 입력 비디오 스트림에서 장면 경계를 결정하는 방법 또는 과정(300)의 단계들을 도시한다. 단계 310에서, 과정은 장면 경계가 결정될 필요가 있는 현재 프레임 (i)과 이전 프레임 (i-1)로부터 고주파수 요소를 스케일링한다. 적어도 일부 실시예에서, 단계 320에서, 과정은 현재 프레임 (i)과 이전 프레임 (i-1)로부터 고주파수 요소를 제거한다. 일 실시예에서, 변환 코더(transform coder)는 이미지 프레임 내의 픽셀 데이터를 주파수 계수로 변환한다. 주파수 도메인에서, 낮은 주파수 데이터는 높은 주파수 데이터보다 더 큰 인간의 인지적 중요도(human perceptual importance)를 갖는다. 단계 310 및 320은 분석이 프레임의 인지적으로 중요한 낮은 주파수 요소에 기초할 수 있게 한다.

[0036] 단계 330에서, 현재 프레임 (i)의 루마(luma)가 계산된다. 루미네슨스(luminescence)로도 알려진 루마 값은 이미지의 휘도(brightness)를 나타낸다(이미지의 "블랙 및 화이트" 또는 무색(achromatic) 부분).

[0037] 단계 340에서, 현재 프레임 (i) 및 이전 프레임 (i-1)에 기초하여 투영(projection)의 루마 값이 계산된다. 투영은 이전 프레임 (i-1)에 기초한 부분 공간(subspace)으로의 현재 프레임 (i)의 투영이다. 부분 공간은 이전 프레임 (i-1)의 특이값 분해(singular value decomposition)에 의해 획득된다.

[0038] 단계 350에서, 단계 330 및 340에서 획득된 루미 값 사이의 차이에 기초한 오차값(residual value)이 계산된다. 단계 360에서, 과정은, 예를 들어, 회귀 필터 또는 적응 필터를 이용하여 임의의 오차값을 필터링하고 0 내지 1의 점수 범위로 나머지 값을 매핑한다. 적응 필터는 오차값에서 임의의 아티팩트를 회귀적으로 필터링하는데 도움을 준다. 단계 370에서, 정규화된 점수가 제1 임계값보다 더 클 때, 과정은 장면 변경을 알리고 입력 비디오 스트림에서 장면 경계를 마킹한다. 일 실시예에서, 이러한 제1 임계값의 예시적인 값은 0.65이다. 단계 380에서, 일부 경우에, 단계 310 내지 370이 입력 비디오 스트림 내의 각 프레임에 대하여 반복되어 입력 비디오 스트림을 정렬된 장면 시컨스로 분할한다.

[0039] 도 4는 주어진 장면 시컨스에 대하여 장면 타입을 결정하는 과정(400)의 단계들을 도시한다. 단계 410에서, 과정은 입력 비디오 스트림의 타임라인에서 장면의 위치를 결정한다. 장면의 위치에 기초하여, 점수가 예를 들어 1 내지 5의 범위에 할당된다. 예시적인 예에서, 1의 점수는 장면이 입력 비디오 스트림의 처음에 있다는 것을 나타낼 수 있고, 5의 점수는 장면이 입력 비디오 스트림의 마지막에 있다는 것을 나타낼 수 있다.

[0040] 단계 420에서, 과정은 장면 시컨스의 재생 시간(play-time) 길이를 결정하고 상응하는 점수를 (예를 들어, 1 내지 5의 범위에) 할당한다. 예시된 예에서, 1의 점수는 10초 미만의 장면 길이를 의미할 수 있고, 5의 점수는 50초보다 더 긴 장면 길이를 의미할 수 있다.

[0041] 단계 430에서, 과정은 장면 시컨스에서 모션 예측을 수행하고 상응하는 점수를 (예를 들어, 1 내지 5의 범위에) 할당한다. 예를 들어, 1의 점수는 모션 벡터가 거의 없거나 전혀 없는 장면을 의미할 수 있고, 5의 점수는 장면에 걸쳐 큰 모션 벡터를 갖는 장면을 의미할 수 있다. 모션 예측(Motion Estimation(ME))은 일반적으로 압축 동안에 비디오 시컨스에서의 시간적 중복성(temporal redundancty)을 탐색하는데 사용되는 기술이다. 시간적 중복성은 이웃하는 프레임들이 유사한 픽셀 영역을 매우 자주 공유한다는 사실로부터 발생한다. 따라서, 모션 예측의 목적은 이웃하는 프레임에 걸쳐 이러한 유사한 영역(매크로 블록)의 이동을 예측하기 위한 것으로, 따라서 이들이 구별하여 인코딩될 수 있게 한다. 블록 기반의 ME에서, 유사한 영역의 변위는 블록 매칭 알고리즘(Block-Matching Algorithm)에 의해 계산되는 모션 벡터에 의해 표현된다.

[0042] 단계 440에서, 과정은 주어진 장면 시컨스와 이전의 장면 시컨스 사이의 유효 차이(effective difference)를 계산한다. 유효 차이에 기초하여, 과정은 상응하는 점수를 (예를 들어, 1 내지 5의 범위에) 할당한다. 예시된 예에서, 1의 점수는 장면들 사이에서 차이가 거의 없다는 것을 나타낼 수 있고, 5의 점수는 xyz 임계값보다 더 큰 차이를 나타낼 수 있다. 예시적인 경우에, 유효 차이는 단계 310 내지 370을 참조하여 전술한 것과 동일한 점수 부여 원리를 이용하여 계산될 수 있다.

[0043] 단계 450에서, 과정은 장면 시컨스의 스펙트럼 데이터 크기를 결정한다. 스펙트럼 데이터 크기에 기초하여, 점수가 예를 들어 1 내지 5의 범위에 할당된다. 예시적인 에에서, 1의 점수는 낮은 스펙트럼 데이터를 갖는 장면을 나타낼 수 있고, 5의 점수는 높은 스펙트럼 데이터를 갖는 장면을 나타낼 수 있다. 일 실시예에서, 변환 코딩 기술은 비디오 데이터를 이미지 프레임의 주파수 도메인 범위가 스펙트럼 데이터 크기를 나타내는 주파수(또는 스펙트럼) 도메인으로 변환한다. 변환 코더는 이미지 프레임 내의 픽셀 데이터를 주파수 계수로 변환한다. 주파수 도메인에서, 저주파수 데이터는 고주파수 데이터보다 더 큰 인간의 인지적 중요도를 가진다.

[0044] 단계 460에서, 과정은 선택적으로(또는 일부 경우에 강제적으로), 예를 들어, 얼굴 인식 소프트웨어를 이용하여, 장면 시컨스에서 얼굴 구조에 대한 검색을 수행한다. 검색 결과에 따라, 점수가, 예를 들어, 1 내지 5의 범위에 할당될 수 있다. 여기에서, 예시적인 예에서, 1의 점수는 인식된 얼굴 구조가 없다는 것을 나타낼 수 있고, 5의 점수는 장면이 많은 수의 얼굴 구조를 갖는다는 것을 나타낼 수 있다.

[0045] 단계 470에서, 과정은 장면 시컨스 내에서 광학 문자 인식(optical character recognition(OCR))을 수행하여 장면 시컨스에서 임의의 문자 정보를 식별한다. OCR은 이미지 파일에서 그림 콘텐츠와 문자 콘텐츠를 구별하는 데 도움을 준다. OCR은 패턴 인식, 인공 지능 및 컴퓨터 비전을 활용하여 구별을 수행한다. OCR 분석에 기초하여, 과정은 상응하는 점수를 (예를 들어, 1 내지 5의 범위에) 할당한다. 예시적인 예에서, 1의 점수는 장면 시컨스 내에서 어떠한 문자 콘텐츠도 없다는 것을 나타낼 수 있고, 5의 점수는 문자 콘텐츠가 장면 시컨스의 콘텐츠의 적어도 30퍼센트, 즉 영화의 크레디트를 구성하는 것을 나타낼 수 있다.

[0046] 단계 480에서, 과정은 장면과 관련된 스크린 플레이 구조 정보를 결정한다. 적어도 일부 실시예에서, 스크린 플레이 구조 정보는 상대적인 관심 파라미터이다. 상대적인 관심 파라미터는 주어진 장면 시컨스에 대하여 예측되는 시청자의 관심의 상대적인 양이다. 일부 경우에, 상대적인 관심 파라미터는 주어진 장면이 일부가 되는 주어진 비디오 세그먼트에 대하여 예측되는 시청자 관심의 상태적인 양을 근사화한다. 분석에 기초하여, 과정은 상응하는 점수를 (예를 들어, 1 내지 5의 범위에) 할당한다. 예시적인 예에서, 1의 점수는 장면 시컨스의 콘텐츠에서의 낮은 시청자 관심을 나타낼 수 있고, 5의 점수는 장면 시컨스의 콘텐츠에서의 높은 시청자 관심을 나타낼 수 있다.

[0047] 단계 490에서, 과정은 단계 410 내지 480으로부터의 점수에 기초하여 장면 시컨스에 대한 장면 타입을 결정한다. 일 실시예에서, 단계 410 내지 480으로부터의 점수를 이용한 장면 타입의 결정은 폭포수 프로세스(waterfall process)에 기초할 수 있다. 폭포수 프로세스는 시작부터 종료까지의 선형의 순차적인 결정 프로세스이며, 이 프로세스는 일반적으로 최종 결정까지의 자신의 경로를 따라 도달한 임의의 중간 결론을 다시 논의하지 않는다.

[0048] 예시적인 예에서, 영화의 종료시에 크레디트를 포함하는 장면은 일반적으로 상향 또는 하향 방향으로 이동하는 문자를 가질 것이다. 이러한 장면 시컨스는 일반적으로 2 이하의 점수가 부여되고 문자의 방향에 따라 위 또는 아래로 향하는 작지만 일정한 이동 백터를 가진다. 또한, 장면 시컨스는 일반적으로 예를 들어 장면 콘텐츠의 30% 이상을 구성하는 영화 크레디트 형태의 문자를 포함한다. 광학 문자 인식 과정은 일반적으로 장면에 4 이상의 점수를 부여한다. 영화 크레디트가 일반적으로 모든 영화의 일부이고 마지막 영화 타임라인의 내용 부분을 구성한다면, 폭포수 프로세스는 다른 장면 타입에 대한 체크를 수행하기 전에 먼저 장면이 "스크롤 크레디트" 타입인지 여부를 알도록 체크할 수 있다. 본 예에서, 2의 점수는 관련된 장면 타입이 "스크롤 크레

디트" 타입인 것을 강력하게 제안하고, 따라서 장면 타입 결정은 장면이 이와 같이 태그되기만 하면 그 장면에 대하여 종료될 수 있다. 장면 타입이 "스크롤 크레디트" 타입이 아닌 것으로 결정되면, 폭포수 프로세스는 이것이 "스크롤 크레디트" 타입 이외의 장면 타입 중 하나인지 알기 위하여 장면 시컨스를 체크한다. 또한, 주어진 장면이 특정 타입이 아니라는 판단을 폭포수 프로세스가 하면, 과정은 일반적으로 그 특정 장면 타입에 대하여 장면을 재평가하지 않는다.

[0049] 다른 예시적인 예에서, 삼나무를 캡쳐하는 장면은 일반적으로 나무의 녹색 나뭇잎과 그 주변을 포함한다. 나뭇잎은 일반적으로 장면 콘텐츠의 주요 부분을 구성한다. 나무 자체가 정지된 상태를 유지하고 가지 및 잎이 최소의 모션을 가지기 때문에, 이러한 장면은 모션 벡터가 거의 없거나 무작위적인 모션 벡터를 가질 것이다. 모션 예측 점수는 0에 가까울 것이다. 또한, 장면에서의 문자의 임의의 인식은 일반적으로 장면의 간단한 설명일 것이어서, 낮은 문자 콘텐츠 점수를 제공할 것이다. 그러나, 장면의 나뭇잎에서의 녹색이 스펙트럼 데이터의 고주파수 도메인에서 캡쳐될 것이기 때문에, 스펙트럼 분석은 높은 점수를 제공할 것이다. 이전에 논의된 바와 같이, 저주파수 스펙트럼 데이터는 고주파수 데이터보다 더 큰 인간의 인지적 중요도를 가지며, 더 낮은 품질에서 높은 스펙트럼 데이터 점수를 갖는 이미지의 인코딩을 가능하게 한다. 점수에 기초하여, 폭포수 프로세스는 장면 시컨스가 "정지 장면" 타입인 것으로 결정할 것이고, 이는 저품질 잔여 인코딩 및 낮은 디블로킹 필터링이 이어지는 제1 프레임에서의 높은 인코딩 품질을 필요로 할 것이다.

[0050] 전술한 설명은 장면 경계 및 장면 타입을 결정하는 과정을 예시하였다. 장면 타입을 결정하는 목적으로, 적어도 일부 경우에, 모션 벡터에 의해 표현되는 장면 시컨스에서의 모션의 크기를 결정하기 위하여 장면에서 모션 예측을 결정하고 분석하는 것이 유용하다. 도 5a 내지 5f는, 여기에서 더욱 상세히 설명되는 바와 같이, 장면 시컨스에서 모션 예측의 예시적인 과정을 도시한다. 모션 예측(ME)은 일반적으로 압축 동안에 비디오 시컨스에서 시간적 중복성을 탐색하는데 사용되는 기술이다. 시간적 중복성은 이웃하는 프레임들이 유사한 픽셀 영역을 매우 자주 공유한다는 사실로부터 발생한다. 따라서, 모션 예측의 목적은 이웃하는 프레임에 걸쳐 이러한 유사한 영역(매크로 블록)의 이동을 예측하기 위한 것으로, 따라서 이들이 구별하여 인코딩될 수 있게 한다. 블록 기반의 ME에서, 유사한 영역의 변위는 블록 매칭 알고리즘에 의해 계산되는 모션 벡터에 의해 표현된다.

[0051] 일 실시예에서, 블록 매칭 알고리즘(BMA)은 이미지 프레임에서 유사한 블록을 검색하여 모션 벡터를 생성한다. BMA는 유사한 블록이 검색되는 동안 검색 윈도우의 특정 포인트에서만 보는 빠른 검색 방식을 사용한다. 다중 해상도 모션 예측으로 알려진 다른 방식에서, ME는 계층적으로 수행되어, 특정 프레임 영역에 대한 모션 벡터를 계산하고 이를 각 레벨에서 다듬는다. 다른 전략은, ME 스테이지를 동시에 실행하기 위하여, BMA에서 유사점을 찾는 것을 살펴본다.

[0052] 도 5는 모션 예측에 대한 예시적인 접근 방식을 예시한다. 여기에서, 실시예에서, 이후의 프레임(예를 들어, 상이한 위치에서의 프레임)에서 발견되는 하나의 프레임의 매크로 블록의 이미지는 모션 벡터를 이용하여 통신된다. 도 5a 및 5b는 기준 프레임 및 희망 프레임을 각각 나타낸다. 프레임들은, 예를 들어 4x4 내지 16x16 범위의 크기로, 매크로 블록으로 분할된다. 실시예에서, 기준 프레임 내의 각 매크로 블록은 임의의 매크로 블록들 사이의 매칭을 검출하기 위하여 희망 프레임 내의 매크로 블록에 비교된다. 도 5c 및 5d는 서로에 대하여 비교되는, 각각 매크로 블록으로 나누어지는 기준 프레임 및 희망 프레임을 도시한다. 도 5e는, 매크로 블록이 해당하는 프레임에서 동일한 그리드 위치에 있지 않더라도, 희망 프레임 내의 매크로 블록과 매칭하는 기준 프레임으로부터의 매크로 블록을 나타낸다. 도 5f는 기준 프레임 내의 매크로 블록의 위치에 대하여 희망 프레임 내의 매크로 블록의 위치를 통신하기 위하여 인코더에 의해 생성되는 모션 벡터를 나타낸다. 따라서, 모션 예측은 장면 시컨스에서 모션 벡터를 결정하는데 도움을 주어, 장면 타입의 결정이 장면 시컨스에서 모션 벡터의 크기에 의해 영향을 받을 수 있게 한다.

[0053] 전술한 설명은 모션 예측에 기초하여 장면 타입을 결정하는 과정을 예시하였다. 장면 타입을 결정하는 목적으로, 모션 예측에 더하여, 적어도 일부 경우에, 장면과 관련된 스크린 플레이 구조 정보를 결정하고 분석하는 것이 유용하다. 스크린 플레이 구조 정보는 적합한 장면 타입을 결정하기 위하여 영화 스토리 라인의 일반적인

구성을 활용하여, 주어진 장면의 적절한 인코딩을 허용한다.

[0054] 영화는 일반적으로 스크린 플레이에 기초한다. 스크린 플레이는 시청자의 관심을 끌도록 구성된다. "자극(bite) 및 전환(switch)"라 불리는 영화 스크린 플레이의 제1 부분은 일반적으로 대부분의 사람들이 전체 영화를 시청할지 여부를 결정할 때이다. 따라서, 여기에서의 이미지 품질은 시청자의 시청 경험을 손상시키지 않도록 매우 높은 것으로 예상된다. "등장인물 전개(character development)"라 불리는 영화 스크린 플레이의 다음 부분은 일반적으로 낮은 시청자 관심을 얻으며, 따라서 이전 세그먼트보다 더 낮은 이미지 품질을 가질 수 있다. 영화의 이후 세그먼트는 일반적으로 영화의 줄거리를 구성하며, 여기에서 시청자 관심은 이전 세그먼트에 비하여 더 높다. 이미지 품질은 이전 품질보다 더 높아야 한다. 영화의 다음 세그먼트는 영화의 가장 중요한 부분이고 이미지의 품질이 높을 필요가 있는 "클라이막스"이다. 최종 세그먼트는 매우 낮은 시청자 관심을 얻는 영화의 "크레디트"이다. 세그먼트는 시청자의 시청 경험에 영향을 미치지 않으면서 더 낮은 품질의 이미지를 활용할 수 있다.

[0055] 일 실시예에서, 장면 타입을 결정하는데 사용되는 스크린 플레이 구조 정보는 영화의 타임라인에 기초할 수 있다. 예를 들어, 주어진 장면 시컨스가 영화의 시작의 일부일 때, 장면 시컨스는 높은 시청자 관심을 얻는 "자극 및 전환" 장면으로서 분류될 수 있다. 장면 시컨스는 범위에서 높은 시청자 관심을 나타내는 5의 점수가 부여될 수 있다. 다른 예로서, 주어진 장면 시컨스가 영화에서 30분 지나 있다면, 영화 세그먼트는 등장인물 전개와 관련된다고 가정할 수 있다. 등장인물 전개 세그먼트는 낮은 시청자 관심을 얻는다. 따라서, 등장인물 전개의 일부인 임의의 장면 시컨스는 범위에서 2 이하의 점수가 부여될 수 있다. 따라서, 타임라인 정보는 장면 타입을 결정하는데 도움을 준다.

[0056] 일 실시예에서, 장면 타입을 결정하는데 사용되는 스크린 플레이 구조 정보는 상대적인 관심 파라미터일 수 있으며, 상대적인 관심 파라미터는 입력 비디오 스트림의 주어진 세그먼트에 대하여 예상되는 예측 사용자 관심을 근사화한다. 상대적인 관심 파라미터는 시청자에 의해 미리 결정될 수 있거나, 영화 감독으로부터의 출력에 기초할 수 있다. 정보는 입력 스트림의 메타 데이터의 일부로서 입력 비디오 스트림에 포함될 수 있다. 메타 데이터를 파싱(parsing)함으로써, 상대적인 관심 파라미터가 결정될 수 있다. 미리 결정된 상대적인 관심 파라미터는 입력 비디오 스트림에서의 각각의 주어진 장면 시컨스에 대하여 또는 복수의 장면 시컨스를 포함하는 입력 비디오 스트림의 주어진 세그먼트에 대하여 정의될 수 있다. 상대적인 관심 파라미터가 높은 시청자 관심을 나타낼 때, 점수는 4 이상으로 설정될 수 있다. 상대적인 관심 파라미터가 낮은 시청자 관심을 나타낼 때, 점수는 2 이하로 설정될 수 있다. 따라서, 상대적인 관심 파라미터는 장면 타입을 결정하는데 활용될 수 있다.

[0057] 일 실시예에서, 장면 타입을 결정하는데 사용되는 스크린 플레이 구조 정보는장면 시컨스 내의 문자 콘텐츠에 기초할 수 있거나 또는 장면 시컨스와 관련된 폐쇄 자막 방송(closed captioning)에 기초할 수 있다. 양 경우에, 문자 정보는 영화의 스크린 플레이 시컨스를 결정하는데 사용된다. 스크린 플레이 시컨스는 낮은 관심도의 장면에 대하여 1의 점수를 그리고 높은 관심도의 장면에 대하여 5의 점수를 부여하여 주어진 장면에 대한 시청자 관심을 결정하는데 활용될 수 있다. 따라서, 문자 콘텐츠 정보는 장면 타입을 결정하는데 활용될 수 있다.

[0058] 다른 실시예에서, 장면 타입을 결정하는데 사용되는 스크린 플레이 구조 정보는 장면 시컨스와 관련된 오디오 콘텐츠에 기초할 수 있다. 오디오 콘텐츠는, 예를 들어, 오디오 콘텐츠의 라우드니스(loudness)(진폭), 음성, 침묵, 언어 인식, 언어 차이, 음악 작품, 음향 효과, 서라운드 사운드 등일 수 있다. 예시적인 예에서, 오디오 콘텐츠의 라우드니스는 장면 시컨스가 그 일부가 되는 스크린 플레이 세그먼트를 결정하는데 사용될 수 있다. 영화에서 액션 세그먼트는 일반적으로 그와 관련된 소리가 큰 오디오 콘텐츠를 가진다. 소리가 큰 오디오 콘텐츠는 시청자의 최대 관심을 얻기 위하여 필요하다. 또한, 액션 장면은 통상적으로 소리가 큰 오디오 콘텐츠를 생성하는 폭발과 같은 특수 효과를 포함한다. 한편, 등장인물 전개와 관련된 영화 세그먼트는 일반적으로 통상적인 범위의 인간의 가청 진폭의 대화를 포함하며 폭발과 같은 특수 효과를 거의 포함하지 않는다. 통상적으로 시청자 관심은 영화의 등장인물 전개 단계에서 낮다. 따라서, 오디오 콘텐츠의 라우드니스는 낮은 진폭의 장면에 대하여 1의 점수를 그리고 높은 진폭의 장면에 대하여 5의 점수를 부여하여 주어진 장면에 대한 시청자 관심

을 결정하는데 활용될 수 있다. 따라서, 오디오 콘텐츠의 진폭(라우드니스)은 시청자의 관심에 기초하여 장면 타입을 결정하는데 활용될 수 있다.

[0059] 다른 예시적인 예에서, 장면 시컨스와 관련된 음향 효과는 장면 시컨스가 그 일부가 되는 스크린 플레이 세그먼트를 결정하는데 사용될 수 있다. 오디오 콘텐츠 내의 증가하는 템포와 같은 특수한 음향 효과는 높은 시청자 관심을 얻는 영화에서 흥미있는 예상 밖의 전개로의 준비, 흥분되는 액션 시컨스 등를 나타내는데 사용될 수 있다. 한편, 거의 없는 음향 효과는 대화와 관련된 영화 세그먼트와 관련된다. 대화가 통상적으로 음향 효과로 더 강조될 수 있는 극적인 감정 이동이 부족하기 때문에, 세그먼트는 일반적으로 음향 효과가 부족하다. 따라서, 오디오 콘텐츠의 특수 효과는 낮은 음향 효과의 장면에 대하여 1의 점수를 그리고 풍부한 음향 효과의 장면에 대하여 5의 점수를 부여하여 주어진 장면에 대한 시청자 관심을 결정하는데 활용될 수 있다. 따라서, 오디오 콘텐츠의 음향 효과는 시청자의 관심에 기초하여 장면 타입을 결정하는데 활용될 수 있다.

[0060] 도 6은 인코더와 같이 임의의 전술한 기술을 구현하는데 사용될 수 있는 처리 시스템에 대한 블록도이다. 소정의 실시예에서, 도 6에 도시된 컴포넌트의 적어도 일부는 2개 이상의 물리적으로 분리되지만 연결된 컴퓨팅 플랫폼 또는 박스 사이의 분산될 수 있다는 것에 유의하여야 한다. 처리는 통상적인 서버-클래스 컴퓨터, PC, 이동 통신 장치(예를 들어, 스마트폰) 또는 임의의 다른 공지되거나 통상적인 처리/통신 장치를 나타낼 수 있다.

[0061] 도 6에 도시된 처리 시스템(601)은 하나 이상의 프로세서(610), 즉 중앙 처리 장치(CPU)와, 메모리(620)와, 이더넷 어댑터 및/또는 무선 통신 서브 시스템(예를 들어, 휴대 전화, 와이파이, 블루투스 등)과 같은 적어도 하나의 통신 장치(640)와, 하나 이상의 I/O 장치(670, 680)를 포함하며, 그 모두는 상호 연결부(690)를 통해 서로 연결된다.

[0062] 프로세서(들)(610)는 컴퓨터 시스템(601)의 동작을 제어하고, 하나 이상의 프로그래머블 범용 또는 특수 목적 마이크로프로세서, 마이크로 컨트롤러, ASIC(application specific integrated circuit), 프로그래머블 논리 장치(programmable logic device(PLD)) 또는 이러한 장치들의 조합이거나 이를 포함할 수 있다. 상호 연결부(690)는 하나 이상의 버스, 직접 연결부 및/또는 다른 종류의 물리적 연결부를 포함할 수 있으며, 본 발명이 속하는 기술 분야에서 널리 알려진 것과 같은 다양한 브릿지, 컨트롤러 및/또는 어댑터를 포함할 수 있다. 상호 연결부(690)는 하나 이상의 어댑터를 통해 PCI(Peripheral Component Interconnect) 버스, 하이퍼 트랜스포트(HyperTransport) 또는 ISA(industry standard architecture) 버스, SCSI(small computer system interface) 버스, USB(universal serial bus) 또는 IEEE(Institute of Electrical and Electronics Engineers) 표준 1394 버스(가끔 "파이어와이어(Firewire)"라 함)의 형태와 같은 하나 이상의 확장 버스에 연결될 수 있는 "시스템 버스"를 더 포함할 수 있다.

[0063] 메모리(620)는, ROM(read-only memory), RAM(random access memory), 플래시 메모리, 디스크 드라이브 등과 같은 하나 이상의 종류의 하나 이상의 메모리 장치이거나 이를 포함할 수 있다. 네트워크 어댑터(640)는 처리 시스템(601)이 통신 링크를 통해 원격 처리 시스템과 데이터를 통신할 수 있게 하기에 적합한 장치이고, 예를 들어, 종래의 전화 모뎀, 무선 모뎀, DSL(Digital Subscriber Line) 모뎀, 케이블 모뎀, 무선 트랜스시버, 위성 트랜스시버, 이더넷 어댑터 등일 수 있다. I/O 장치(670, 680)는, 예를 들어, 마우스, 트랙볼, 터치패드 등과 같은 지시 장치; 키보드; 음성 인식 인터페이스를 갖는 마이크; 오디오 스피커; 디스플레이 장치 등을 포함할 수 있다. 그러나, 적어도 일부 환경에서는 서버가 그러하듯이, 이러한 I/O 장치는 서버로서 독점적으로 동작되고 직접적인 사용자 인터페이스를 제공하지 않는 시스템에서는 불필요할 수 있다. 예시된 부품 세트에서의 다른 변동은 본 발명과 일치하는 방식으로 구현될 수 있다.

[0064] 전술한 동작을 수행하도록 프로세서(들)(610)를 프로그래밍하기 위한 소프트웨어 및/또는 펌웨어(630)는 메모리(620) 내에 저장될 수 있다. 소정의 실시예에서, 이러한 소프트웨어 또는 펌웨어는 초기에 컴퓨터 시스템(601)을 통해(예를 들어, 네트워크 어댑터(640)를 통해) 원격 시스템으로 이를 다운로드함으로써 컴퓨터 시스템

(601)에 제공할 수 있다.

[0065] 위에서 소개된 기술은, 예를 들어, 소프트웨어 및/또는 펌웨어로 프로그래밍된 프로그래머블 회로(예를 들어, 하나 이상의 마이크로프로세서)에 의해, 또는 특수 목적의 하드웨어 내장형(hardwired) 회로에서 전적으로, 혹은 이러한 형태의 조합으로 구현될 수 있다. 특수 목적의 하드웨어 내장형 회로는 ASIC(application-specific integrated circuit), PLD(programmable logic device), FPGA(field-programmable gate array) 등의 형태일 수 있다.

[0066] 여기에서 소개된 기술을 구현하는데 사용하기 위한 소프트웨어 또는 펌웨어는 기계 판독 가능한 저장 매체에 저장될 수 있고, 하나 이상의 범용 또는 특수 목적 프로그래머블 마이크로프로세서에 의해 실행될 수 있다. "기계 판독 가능한 저장 매체"는, 용어가 본 명세서에서 사용되는 바와 같이, 기계(기계는, 예를 들어, 컴퓨터, 네트워크 장치, 휴대 전화, PDA(personal digital assistant), 제작 도구, 하나 이상의 프로세서를 갖는 임의의 장치 등일 수 있다)에 의해 액세스 가능한 형태로 정보를 저장할 수 있는 임의의 메커니즘을 포함한다. 예를 들어, 기계가 액세스 가능한 저장 매체는, 재기록 가능/재기록 불가능 매체(예를 들어, ROM(read-only memory); RAM(random access memory); 자기 디스크 저장 매체; 광학 저장 매체; 플래시 메모리 장치 등) 등을 포함한다.

[0067] "논리부(logic)"라는 용어는, 본 명세서에서 사용되는 바와 같이, 예를 들어, 특정 소프트웨어 및/또는 펌웨어, 특수 목적의 하드웨어 내장형 회로 또는 그 조합으로 프로그래밍된 프로그래머블 회로를 포함할 수 있다.

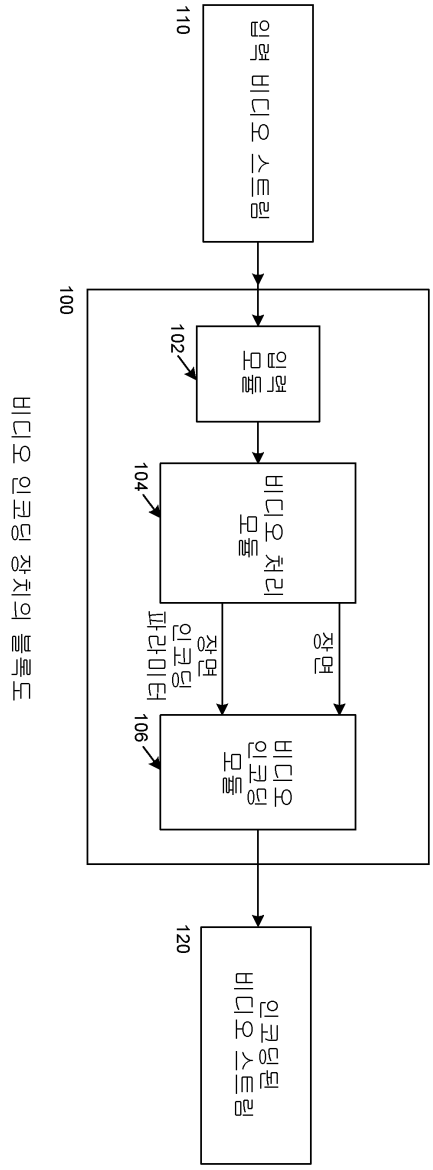
[0068] 청구된 대상의 다양한 실시예에 대한 전술한 설명이 예시 및 설명의 목적을 위해 제공되었다. 청구 대상을 개시된 특정한 형태에 한정시키거나 소진적인 것으로 의도되지 않는다. 많은 수정 및 변형이 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 자명할 것이다. 본 발명의 원리와 그의 실용적인 애플리케이션을 최선으로 설명하기 위하여 실시예들이 선택되고 설명되었으며, 이에 의해 관련 분야에서 통상의 지식을 갖는 자가 청구 대상과 다양한 실시예를 이해할 수 있게 하며, 특정 용도에 적합한 다양한 변형이 고려될 수 있게 한다.

[0069] 여기에서 제공된 본 발명에 대한 교시 내용은 반드시 전술한 시스템일 필요가 없는 다른 시스템에 적용될 수 있다. 전술한 다양한 실시예의 요소 및 동작은 다른 실시예를 제공하기 위하여 조합될 수 있다.

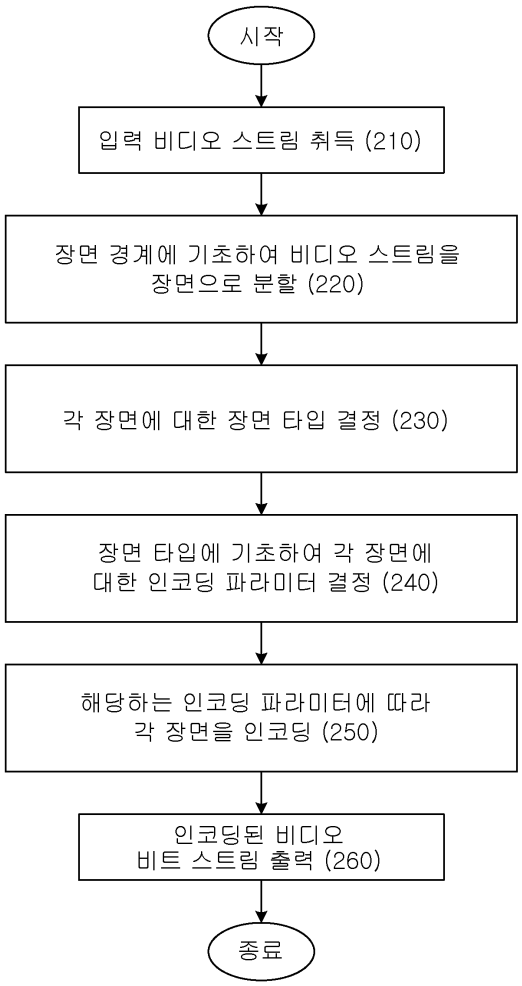
[0070] 전술한 설명은 본 발명의 소정의 실시예를 설명하고 고려되는 최선의 형태를 설명하지만, 전술한 것이 본문에서 얼마나 상세한 지에 관계없이, 본 발명은 많은 방법으로 실시될 수 있다. 실시예에 대한 상세는, 본 명세서에 개시된 발명에 의해 여전히 포함되면서, 그 상세한 구현에서 상당히 변경될 수 있다. 전술한 바와 같이, 본 발명의 소정의 특징 또는 양태를 설명할 때 사용되는 특정 용어는, 그 용어가 관련되는 발명의 임의의 특수한 특성, 특징 또는 양태에 제한되는 것으로 용어가 재정의되는 것을 의미하도록 취급되어서는 안 된다. 일반적으로, 이어지는 특허청구범위에 사용되는 용어는, 전술한 발명을 실시하기 위한 구체적인 내용 부분이 그러한 용어를 명시적으로 정의하지 않는 한, 본 발명을 명세서에 개시된 특정 실시예로 한정하도록 고려되어서는 안 된다. 따라서, 본 발명의 실제 범위는 개시된 실시예 뿐만 아니라 특허청구범위 하에서 본 발명을 실시하거나 구현하는 모든 균등한 방법을 포함한다.

도면

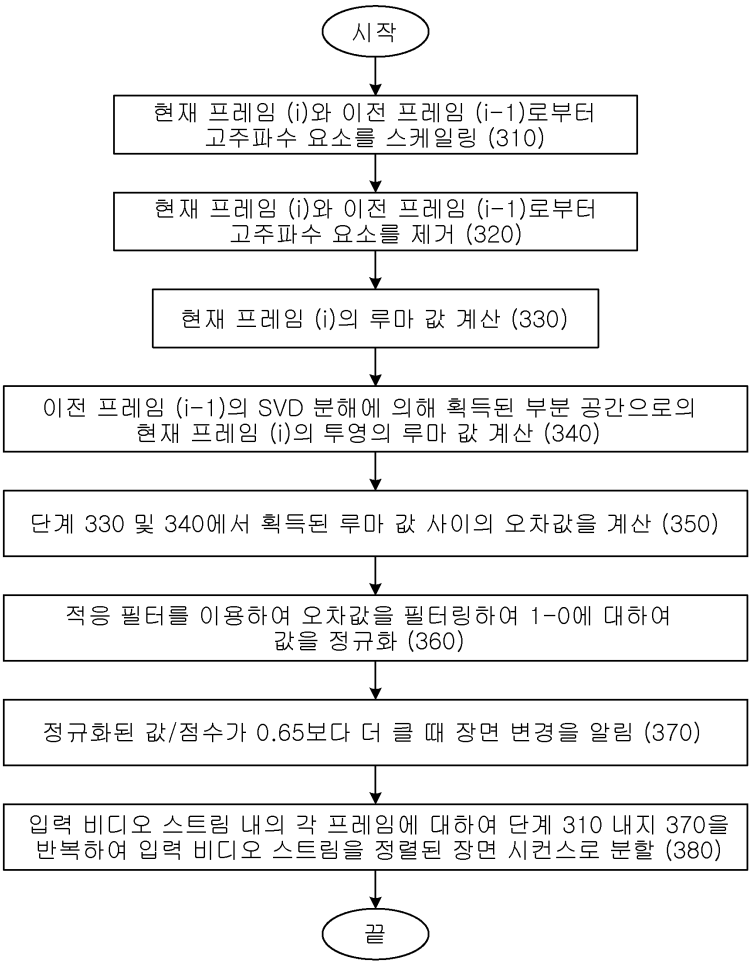
도면1


비디오 인코딩 장치의 블록도
입력 비디오 스트림
110
100
입력 모듈
102
비디오 처리 모듈
104
장면 인코딩 파라미터
장면
비디오 인코딩 모듈
106
인코딩된 비디오 스트림
120

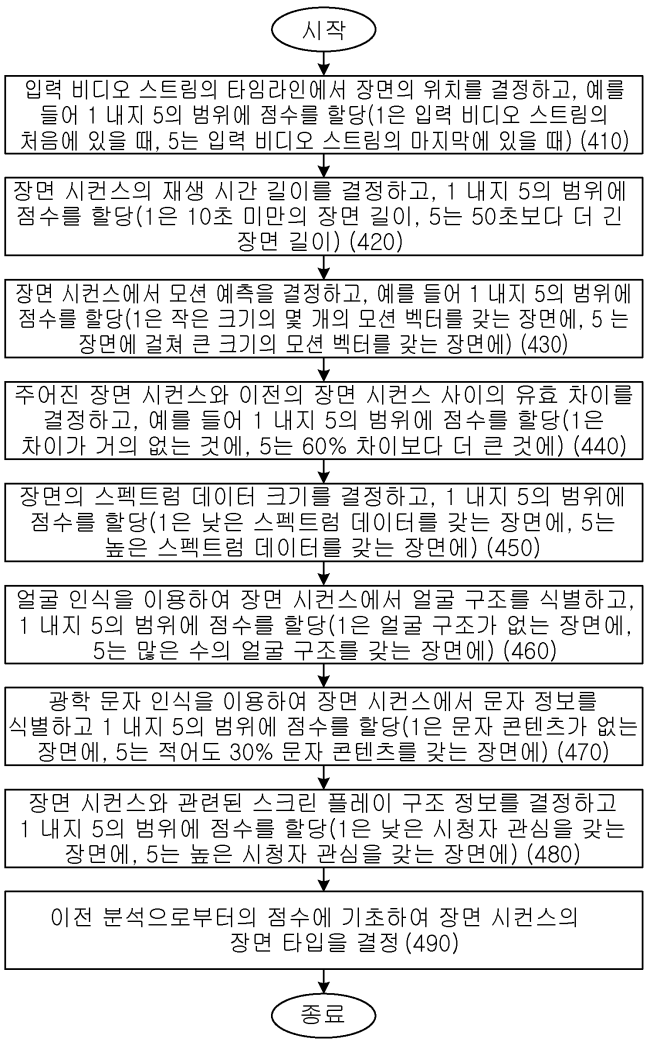
도면2


시작
입력 비디오 스트림 취득 (210)
장면 경계에 기초하여 비디오 스트림을 장면으로 분할 (220)
각 장면에 대한 장면 타입 결정 (230)
장면 타입에 기초하여 각 장면에 대한 인코딩 파라미터 결정 (240)
해당하는 인코딩 파라미터에 따라 각 장면을 인코딩 (250)
인코딩된 비디오 비트 스트림 출력 (260)
종료

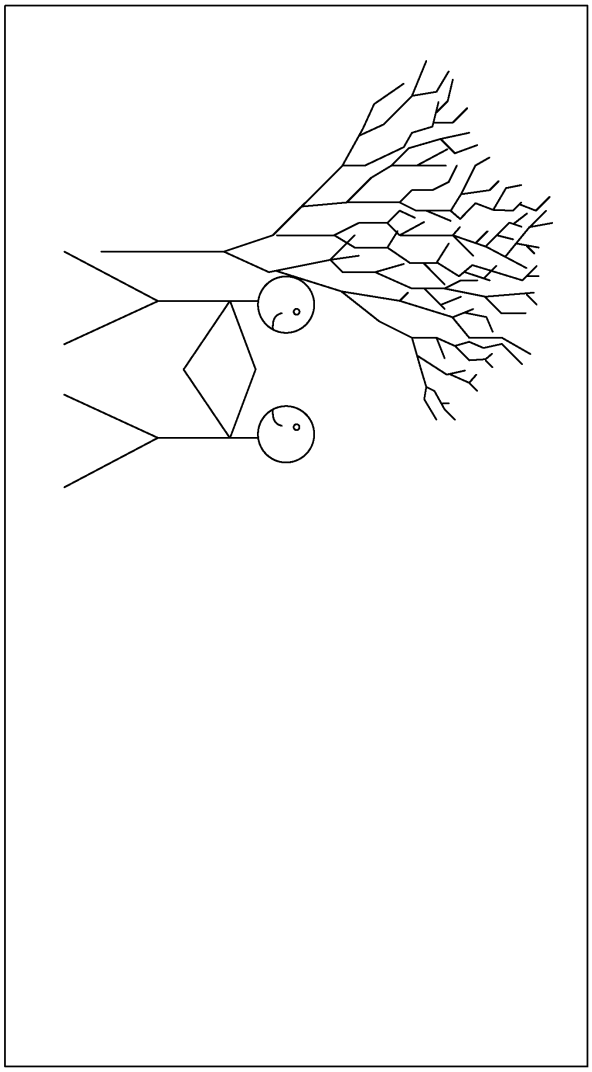
도면3


시작
현재 프레임 (i)와 이전 프레임 (i-1)로부터 고주파수 요소를 스케일링 (310)
현재 프레임 (i)와 이전 프레임 (i-1)로부터 고주파수 요소를 제거 (320)
현재 프레임 (i)의 루마 값 계산 (330)
이전 프레임 (i-1)의 SVD 분해에 의해 획득된 부분 공간으로의 현재 프레임 (i)의 투영의 루마 값 계산 (340)
단계 330 및 340에서 획득된 루마 값 사이의 오차값을 계산 (350)
적응 필터를 이용하여 오차값을 필터링하여 1-0에 대하여 값을 정규화 (360)
정규화된 값/점수가 0.65보다 더 클 때 장면 변경을 알림 (370)
입력 비디오 스트림 내의 각 프레임에 대하여 단계 310 내지 370을 반복하여 입력 비디오 스트림을 정렬된 장면 시컨스로 분할 (380)
끝

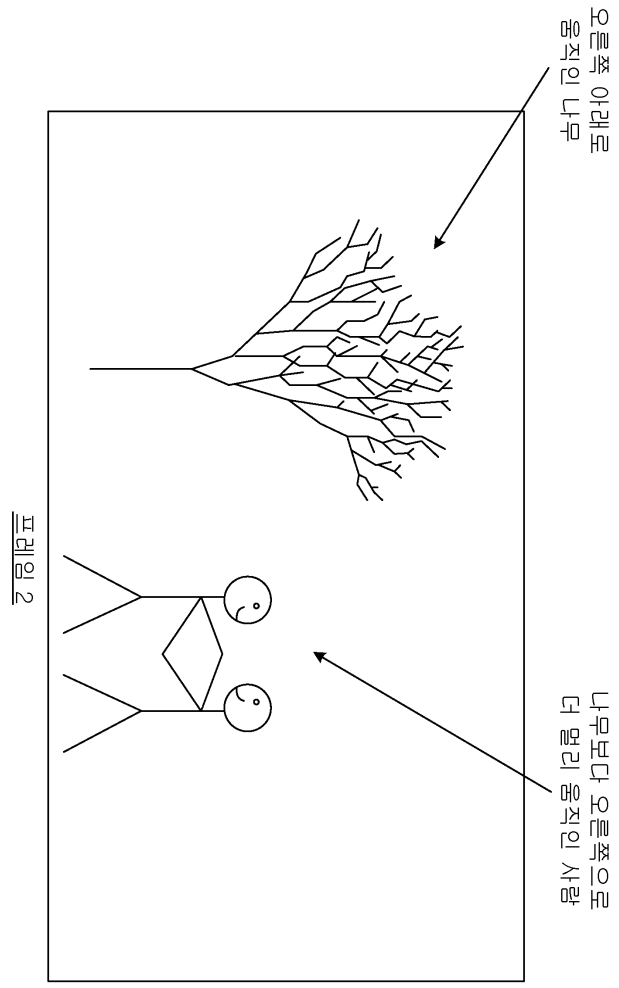
도면4


시작
입력 비디오 스트림의 타임라인에서 장면의 위치를 결정하고, 예를 들어 1 내지 5의 범위에 점수를 할당(1은 입력 비디오 스트림의 처음에 있을 때, 5는 입력 비디오 스트림의 마지막에 있을 때) (410)
장면 시컨스의 재생 시간 길이를 결정하고, 1 내지 5의 범위에 점수를 할당(1은 10초 미만의 장면 길이, 5는 50초보다 더 긴 장면 길이) (420)
장면 시컨스에서 모션 예측을 결정하고, 예를 들어 1 내지 5의 범위에 점수를 할당(1은 작은 크기의 몇 개의 모션 벡터를 갖는 장면에, 5 는 장면에 걸쳐 큰 크기의 모션 벡터를 갖는 장면에) (430)
주어진 장면 시컨스와 이전의 장면 시컨스 사이의 유효 차이를 결정하고, 예를 들어 1 내지 5의 범위에 점수를 할당(1은 차이가 거의 없는 것에, 5는 60% 차이보다 더 큰 것에) (440)
장면의 스펙트럼 데이터 크기를 결정하고, 1 내지 5의 범위에 점수를 할당(1은 낮은 스펙트럼 데이터를 갖는 장면에, 5는 높은 스펙트럼 데이터를 갖는 장면에) (450)
얼굴 인식을 이용하여 장면 시컨스에서 얼굴 구조를 식별하고, 1 내지 5의 범위에 점수를 할당(1은 얼굴 구조가 없는 장면에, 5는 많은 수의 얼굴 구조를 갖는 장면에) (460)
광학 문자 인식을 이용하여 장면 시컨스에서 문자 정보를 식별하고 1 내지 5의 범위에 점수를 할당(1은 문자 콘텐츠가 없는 장면에, 5는 적어도 30% 문자 콘텐츠를 갖는 장면에) (470)
장면 시컨스와 관련된 스크린 플레이 구조 정보를 결정하고 1 내지 5의 범위에 점수를 할당(1은 낮은 시청자 관심을 갖는 장면에, 5는 높은 시청자 관심을 갖는 장면에) (480)
이전 분석으로부터의 점수에 기초하여 장면 시컨스의 장면 타입을 결정 (490)
종료

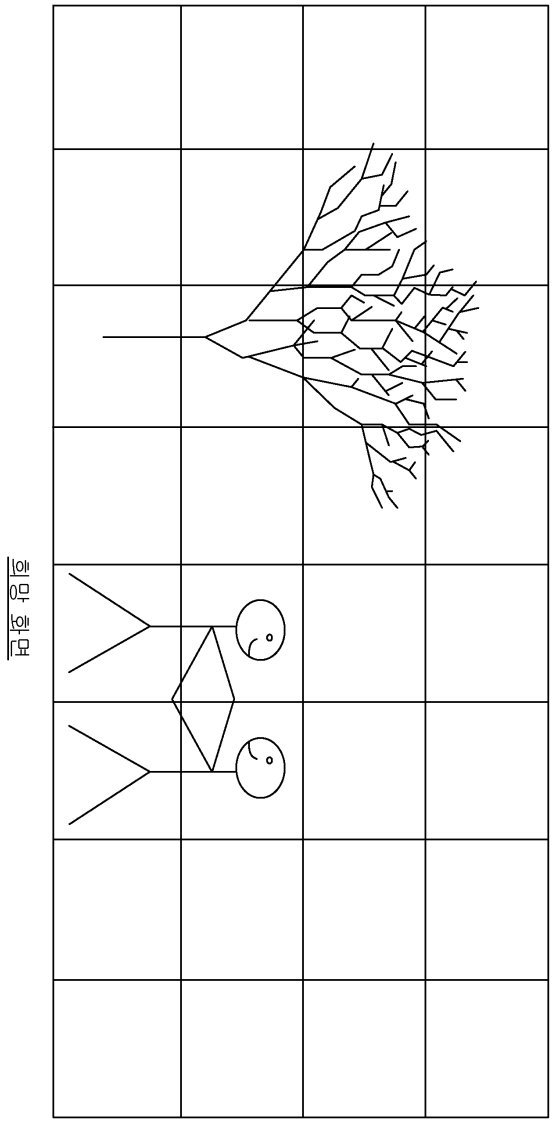
도면5a



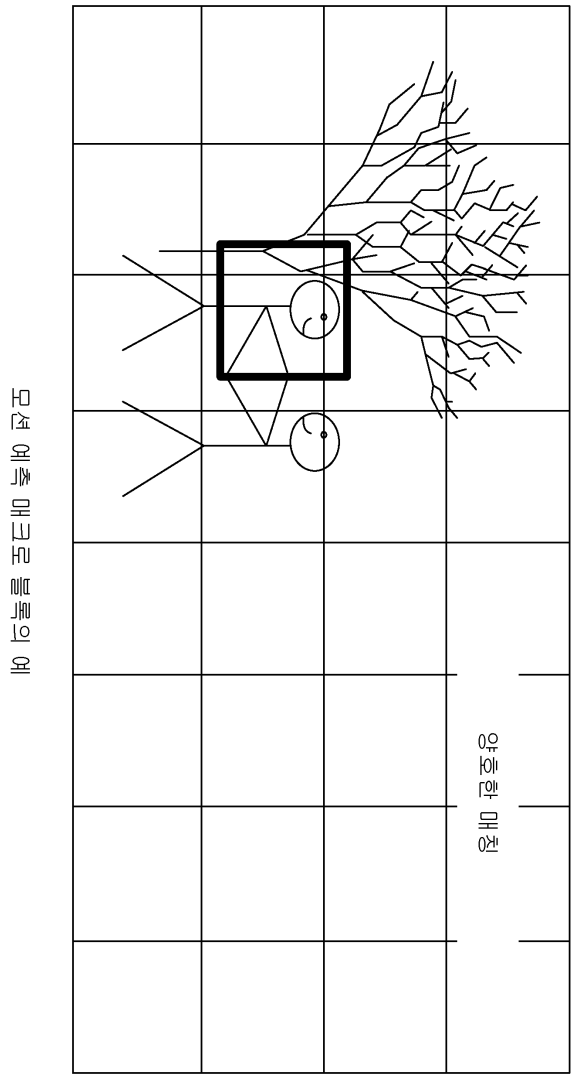
도면5b



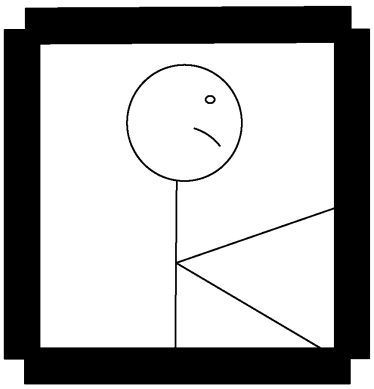
도면5c



도면5d

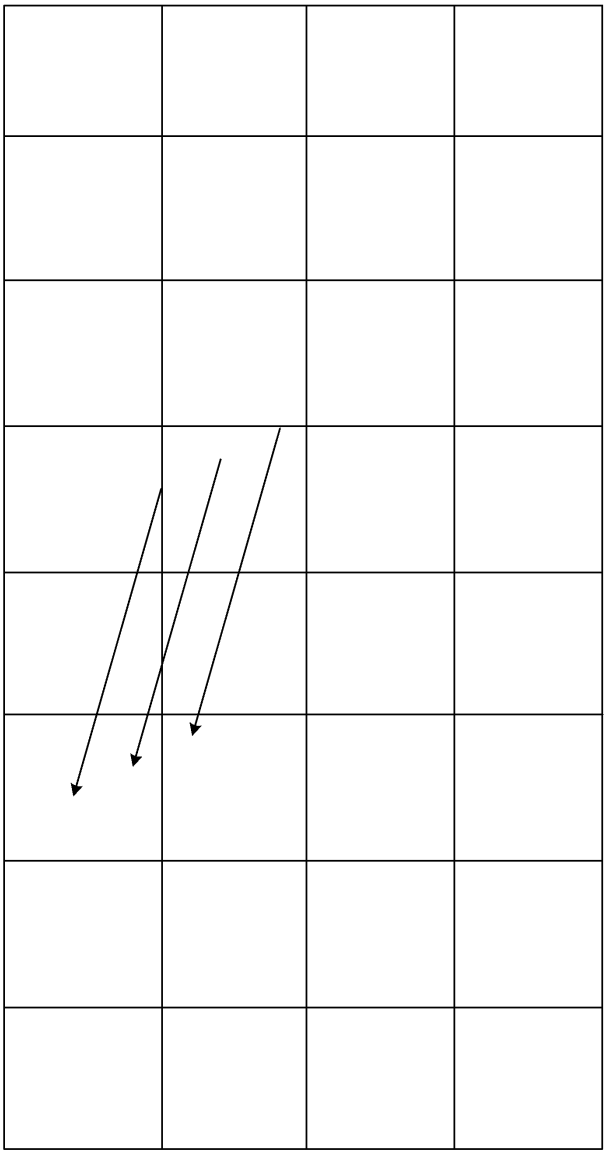


도면5e

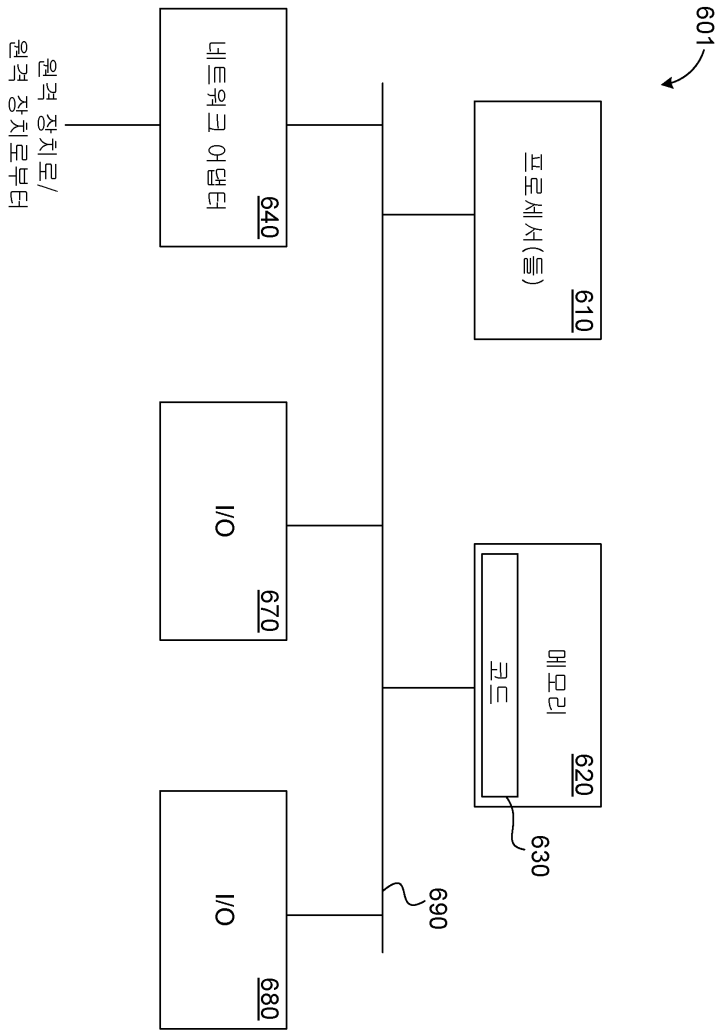


코딩될 매크로 블록

도면5f



도면6


601
프로세서(들)
610
메모리
620
코드
630
네트워크 어댑터
640
I/O
670
I/O
680
690
원격 장치로/
원격 장치로부터