



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2013년11월08일
(11) 등록번호 10-1326443
(24) 등록일자 2013년10월31일

(51) 국제특허분류(Int. Cl.)
H04N 7/26 (2006.01) H04N 7/50 (2006.01)
(21) 출원번호 10-2012-7021981(분할)
(22) 출원일자(국제) 2006년09월27일
심사청구일자 2012년08월22일
(85) 번역문제출일자 2012년08월22일
(65) 공개번호 10-2012-0099806
(43) 공개일자 2012년09월11일
(62) 원출원 특허 10-2010-7019260
원출원일자(국제) 2006년09월27일
심사청구일자 2011년09월27일
(86) 국제출원번호 PCT/US2006/037948
(87) 국제공개번호 WO 2007/038701
국제공개일자 2007년04월05일
(30) 우선권주장
60/721,416 2005년09월27일 미국(US)
(뒷면에 계속)
(56) 선행기술조사문헌
KR1020050089721 A
MIHAELA VAN DER SCHAAR ET AL: "A Hybrid
Temporal-SNR Fine-Granular Scalability for
Internet Video", IEEE TRANSACTIONS ON
CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY,
vol. 11, no. 3 ,2001.03.
KWON ET AL : "Overvi ew o f H. 264/MPEG-4
part 10", JOURNAL OF VISUAL COMMUNICATION AND
IMAGE REPRESENTATION, vol. 17, no. 2 ,
2006.04., p186-216
RUSERT T ET AL: "Enhanced inter frame wavelet
video coding considering the interrelation of
spatio-temporal transform and motion
compensation" SIGNAL PROCESSING IMAGE
COMMUNICATION, ELSEVIER SCIENCE P

(73) 특허권자
퀄컴 인코포레이티드
미국 92121-1714 캘리포니아주 샌 디에고 모어하
우스 드라이브 5775
(72) 발명자
라빈드란 비자알라크시미 알
미국 92130 캘리포니아주 샌디에고 칼레 마르 데
발레나스 4272
워커 고든 켄트
미국 92064 캘리포니아주 포웨이 헌팅턴 게이트
드라이브 14484
(뒷면에 계속)
(74) 대리인
특허법인코리어나

전체 청구항 수 : 총 18 항

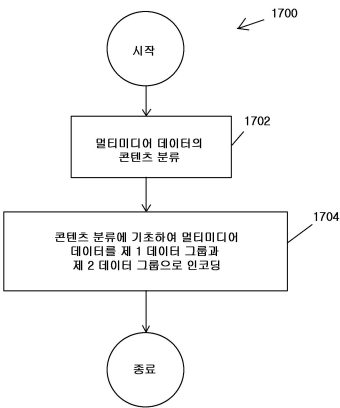
심사관 : 조우연

(54) 발명의 명칭 콘텐츠 정보에 기초한 스케일러빌리티 기술

(57) 요약

콘텐츠 정보를 사용하여 멀티미디어 데이터를 인코딩하는 장치 및 방법이 설명된다. 멀티미디어 데이터를 처리하는 방법은 멀티미디어 데이터의 콘텐츠를 분류하는 단계, 및 콘텐츠 분류에 기초하여 멀티미디어 데이터를 제 1 데이터 그룹과 제 2 데이터 그룹으로 인코딩하는 단계를 포함하고, 제 1 데이터 그룹은 일 계수를 포함하고, 제 2 데이터 그룹은 제 1 데이터 그룹 계수와 연관된 제 1 차분 세분화를 포함한다. 콘텐츠 정보를 사용하여 멀티미디어 데이터를 인코딩하는 장치는 멀티미디어 데이터의 콘텐츠를 분류하고, 콘텐츠 분류 데이터를 제공하도록 구성된 콘텐츠 분류 모듈, 및 콘텐츠 분류에 기초하여 멀티미디어 데이터를 제 1 데이터 그룹과 제 2 데이터 그룹으로 인코딩하도록 구성된 인코더를 포함하고, 제 1 데이터 그룹은 일 계수를 포함하고, 제 2 데이터 그룹은 제 1 데이터 그룹 계수와 연관된 제 1 차분 세분화를 포함한다.

대표도 - 도17a



(72) 발명자	(30) 우선권주장
티안 타오	60/727,640 2005년10월17일 미국(US)
미국 92126 캘리포니아주 샌디에고 뉴 살렘 스트리트 8680 넘버 145	60/727,643 2005년10월17일 미국(US)
바미디파티 파니쿠마르	60/727,644 2005년10월17일 미국(US)
미국 92121 캘리포니아주 샌디에고 제네시 애비뉴 9645 넘버 아이1	60/730,145 2005년10월24일 미국(US)
시 팡	60/789,048 2006년04월03일 미국(US)
미국 92130 캘리포니아주 샌디에고 칼레 마 데 아모니아 4460	60/789,377 2006년04월04일 미국(US)
천페이송	
미국 92126 캘리포니아주 샌디에고 카미니토 알바레즈 10859	
수브라마니아 시타라만 가나파시	
미국 92126 캘리포니아주 샌디에고 힐러리 드라이브 9364 넘버 2204	
오구즈 세이풀라 할리트	
미국 92122 캘리포니아주 샌디에고 피오레 테라스 5240 넘버 417	

특허청구의 범위

청구항 1

멀티미디어 데이터를 인코딩하는 방법으로서,

멀티미디어 데이터의 콘텐츠를 분류하는 단계로서, 상기 콘텐츠는 상기 멀티미디어 데이터의 텍스처의 레벨, 상기 멀티미디어 데이터의 시각적으로 인식된 품질의 레벨, 상기 멀티미디어 데이터의 소리-대-잡음 비 (SNR) 중 적어도 하나를 포함하는 타깃 품질 레벨에 기초하여 분류되는, 상기 분류하는 단계; 및

상기 콘텐츠 분류에 기초하여 상기 멀티미디어 데이터를 제 1 데이터 그룹 및 제 2 데이터 그룹으로 인코딩하는 단계로서, 상기 제 1 데이터 그룹은 일 계수를 포함하고, 상기 제 2 데이터 그룹은 상기 제 1 데이터 그룹 계수와 관련된 제 1 차분 세분화 (differential refinement) 를 포함하며, 상기 제 1 데이터 그룹은 베이스 계층이고, 상기 제 2 데이터 그룹은 인핸스먼트 계층인, 상기 인코딩하는 단계를 포함하는, 멀티미디어 데이터 인코딩 방법.

청구항 2

제 1 항에 있어서,

통신 모듈을 통하여 상기 타깃 품질을 수신하는 단계를 더 포함하고,

상기 타깃 품질은 적어도 부분적으로 사용자에게 의해 선택되는, 멀티미디어 데이터 인코딩 방법.

청구항 3

제 1 항에 있어서,

상기 인코딩하는 단계는, 상기 멀티미디어 데이터의 상기 콘텐츠 분류에 기초하여 비트 레이트를 결정하고, 상기 비트 레이트에 기초하여 상기 멀티미디어 데이터를 인코딩하는 단계를 포함하는, 멀티미디어 데이터 인코딩 방법.

청구항 4

제 1 항에 있어서,

상기 콘텐츠를 분류하는 단계는, 상기 멀티미디어 데이터의 복잡도를 결정하는 단계를 포함하고,

선택된 멀티미디어 데이터는 상기 멀티미디어 데이터의 상기 복잡도에 기초하여 인코딩되는, 멀티미디어 데이터 인코딩 방법.

청구항 5

제 1 항에 있어서,

상기 인코딩하는 단계는, 상기 제 1 데이터 그룹만을 디코딩하거나 또는 상기 제 1 데이터 그룹 및 상기 제 2 데이터 그룹을 디코딩하여 단일 결합된 데이터 그룹이 되도록 상기 멀티미디어 데이터를 인코딩하는 단계를 포함하는, 멀티미디어 데이터 인코딩 방법.

청구항 6

제 1 항에 있어서,

상기 제 1 차분 세분화는 선택된 비디오 프레임과 상기 제 1 데이터 그룹을 디코딩한 것으로부터 얻어진 프레임 데이터의 차이를 나타내는, 멀티미디어 데이터 인코딩 방법.

청구항 7

멀티미디어 데이터를 인코딩하는 방법으로서,

멀티미디어 데이터의 콘텐츠를 분류하는 단계로서, 상기 콘텐츠는 타깃 품질 레벨에 기초하여 분류되는, 상기

분류하는 단계;

상기 콘텐츠 분류에 기초하여 상기 멀티미디어 데이터를 제 1 데이터 그룹 및 상기 제 2 데이터 그룹으로 인코딩하는 단계로서, 상기 제 1 데이터 그룹은 일 계수를 포함하고, 상기 제 2 데이터 그룹은 상기 제 1 데이터 그룹 계수와 관련된 제 1 차분 세분화 (differential refinement) 를 포함하며, 상기 제 1 데이터 그룹은 베이스 계층이고, 상기 제 2 데이터 그룹은 인헨스먼트 계층인, 상기 인코딩하는 단계;

원래 베이스 계층 나머지 에러 계수 또는 원래 인헨스먼트 계층 나머지 에러 계수 중 하나로부터 상기 계수를 선택하는 단계; 및

상기 계수 및 상기 원래 인헨스먼트 계층 나머지 에러 계수에 기초하여 상기 제 1 차분 세분화를 계산하는 단계를 포함하는, 멀티미디어 데이터 인코딩 방법.

청구항 8

제 1 항에 있어서,

상기 인코딩하는 단계는 상기 제 1 데이터 그룹 내의 매크로블록 헤더 정보 및 모션 벡터 정보를 인코딩하는 단계를 더 포함하는, 멀티미디어 데이터 인코딩 방법.

청구항 9

제 1 항에 있어서,

상기 인코딩하는 단계는 상기 제 1 데이터 그룹을 제 1 스텝 크기로 양자화하는 단계, 및 상기 제 2 데이터 그룹을 제 2 스텝 크기로 양자화하는 단계를 더 포함하고, 상기 제 1 스텝 크기 및 상기 제 2 스텝 크기는 스케일 인자에 의해 상관되는, 멀티미디어 데이터 인코딩 방법.

청구항 10

멀티미디어 데이터를 인코딩하는 장치로서,

멀티미디어 데이터의 콘텐츠를 분류하기 위한 수단으로서, 상기 콘텐츠는 상기 멀티미디어 데이터의 텍스처의 레벨, 상기 멀티미디어 데이터의 시각적으로 인식된 품질의 레벨, 상기 멀티미디어 데이터의 소리-대-잡음 비 (SNR) 중 적어도 하나를 포함하는 타겟 품질 레벨에 기초하여 분류되는, 상기 분류하기 위한 수단; 및

상기 콘텐츠 분류에 기초하여 상기 멀티미디어 데이터를 제 1 데이터 그룹 및 제 2 데이터 그룹으로 인코딩하기 위한 수단으로서, 상기 제 1 데이터 그룹은 일 계수를 포함하고, 상기 제 2 데이터 그룹은 상기 제 1 데이터 그룹 계수와 관련된 제 1 차분 세분화 (differential refinement) 를 포함하며, 상기 제 1 데이터 그룹은 베이스 계층이고, 상기 제 2 데이터 그룹은 인헨스먼트 계층인, 상기 인코딩하기 위한 수단을 포함하는, 멀티미디어 데이터 인코딩 장치.

청구항 11

제 10 항에 있어서,

상기 타겟 품질을 수신하기 위한 수단을 더 포함하고,

상기 타겟 품질은 적어도 부분적으로 사용자에게 의해 선택되는, 멀티미디어 데이터 인코딩 장치.

청구항 12

멀티미디어 데이터를 인코딩하도록 구성된 장치로서,

멀티미디어 데이터의 콘텐츠를 분류하고 콘텐츠 분류 데이터를 제공하도록 구성된 콘텐츠 분류 모듈로서, 상기 콘텐츠는 상기 멀티미디어 데이터의 텍스처의 레벨, 상기 멀티미디어 데이터의 시각적으로 인식된 품질의 레벨, 상기 멀티미디어 데이터의 소리-대-잡음 비 (SNR) 중 적어도 하나를 포함하는 타겟 품질 레벨에 기초하여 분류되는, 상기 콘텐츠 분류 모듈; 및

상기 콘텐츠 분류에 기초하여 상기 멀티미디어 데이터를 제 1 데이터 그룹 및 제 2 데이터 그룹으로 인코딩하도록 구성된 인코더로서, 상기 제 1 데이터 그룹은 일 계수를 포함하고, 상기 제 2 데이터 그룹은 상기 제 1 데이터 그룹 계수와 관련된 제 1 차분 세분화 (differential refinement) 를 포함하며, 상기 제 1 데이터 그룹은 베

이스 계층이고, 상기 제 2 데이터 그룹은 인헨스먼트 계층인, 상기 인코더를 포함하는, 멀티미디어 데이터를 인코딩하도록 구성된 장치.

청구항 13

제 12 항에 있어서,

통신 모듈을 통하여 상기 타깃 품질을 수신하는 것을 더 포함하고,

상기 타깃 품질은 적어도 부분적으로 사용자에게 의해 선택되는, 멀티미디어 데이터를 인코딩하도록 구성된 장치.

청구항 14

그 내부에 저장된 명령들을 포함하는 머신-판독가능 매체로서, 실행시 머신으로 하여금:

멀티미디어 데이터의 콘텐츠를 분류하는 것으로서, 상기 콘텐츠는 상기 멀티미디어 데이터의 텍스처의 레벨, 상기 멀티미디어 데이터의 시각적으로 인식된 품질의 레벨, 상기 멀티미디어 데이터의 소리-대-잡음 비 (SNR) 중 적어도 하나를 포함하는 타깃 품질 레벨에 기초하여 분류되는, 상기 분류하는 것; 및

상기 콘텐츠 분류에 기초하여 상기 멀티미디어 데이터를 제 1 데이터 그룹 및 제 2 데이터 그룹으로 인코딩하는 것으로서, 상기 제 1 데이터 그룹은 일 계수를 포함하고, 상기 제 2 데이터 그룹은 상기 제 1 데이터 그룹 계수와 관련된 제 1 차분 세분화 (differential refinement) 를 포함하는, 상기 인코딩하는 것을 하도록 하는, 머신-판독가능 매체.

청구항 15

제 14 항에 있어서,

통신 모듈을 통하여 상기 타깃 품질을 수신하는 것을 더 하도록 하고,

상기 타깃 품질은 적어도 부분적으로 사용자에게 의해 선택되는, 머신-판독가능 매체.

청구항 16

멀티미디어 데이터의 콘텐츠를 분류하는 것으로서, 상기 콘텐츠는 상기 멀티미디어 데이터의 텍스처의 레벨, 상기 멀티미디어 데이터의 시각적으로 인식된 품질의 레벨, 상기 멀티미디어 데이터의 소리-대-잡음 비 (SNR) 중 적어도 하나를 포함하는 타깃 품질 레벨에 기초하여 분류되는, 상기 분류하는 것; 및

상기 콘텐츠 분류에 기초하여 상기 멀티미디어 데이터를 제 1 데이터 그룹 및 제 2 데이터 그룹으로 인코딩하는 것으로서, 상기 제 1 데이터 그룹은 일 계수를 포함하고, 상기 제 2 데이터 그룹은 상기 제 1 데이터 그룹 계수와 관련된 제 1 차분 세분화 (differential refinement) 를 포함하는, 상기 인코딩하는 것을 하도록 구성된, 프로세서.

청구항 17

제 16 항에 있어서,

통신 모듈을 통하여 상기 타깃 품질을 수신하는 것을 더 하도록 구성되고,

상기 타깃 품질은 적어도 부분적으로 사용자에게 의해 선택되는, 프로세서.

청구항 18

제 16 항에 있어서,

상기 프로세서와 연관된 메모리를 더 포함하는, 프로세서.

명세서

기술 분야

[0001] 본 특허 출원은 실시간 스트리밍용 비디오 데이터의 비디오 트랜스코딩을 위한 장치 및 방법에 관한 것으로, 더욱 상세하게는, 이동 방송 애플리케이션에서 실시간 스트리밍용 비디오 데이터를 트랜스코딩하는 것에 관한 것이다.

배경 기술

[0002] 대역폭 자원과 이용 가능한 대역폭의 변화성이 제한되기 때문에, 무선 비디오 스트리밍 및 비디오 전화통신과 같은 다수의 멀티미디어 애플리케이션에서는 효율적인 비디오 압축이 유용하다. MPEG-4 (ISO/IEC), H.264 (ITU) 또는 유사한 비디오 코딩과 같은 일정 비디오 코딩 표준은 무선 방송과 같은 애플리케이션에 아주 적합한 고효율 코딩을 제공한다. 일반적으로, 예를 들어, 디지털 TV 프리젠테이션과 같은 몇몇 멀티미디어 데이터는 MPEG-2 와 같은 다른 표준에 따라 코딩된다. 따라서, 트랜스코더를 사용하여, 한 표준 (예를 들어, MPEG-2) 에 따라 코딩된 멀티미디어 데이터를 무선 방송 전에 다른 표준 (예를 들어, H.264) 으로 트랜스코딩하거나 변환하게 된다.

[0003] 개선 비율 최적화된 코덱 (improvements rate optimized codec) 은 에러 내성, 에러 복구 및 스케일러빌리티에 있어서 이점을 제공할 수 있다. 또한, 멀티미디어 데이터 자체로부터 결정된 정보를 사용하면, 에러 내성, 에러 복구 및 스케일러빌리티를 비롯한 인코딩에 대한 추가적인 개선을 제공할 수 있다. 따라서, 멀티미디어 데이터의 고효율 처리 및 압축을 제공하고, 멀티미디어 데이터 자체로부터 결정된 정보를 사용하고, 스케일러블 (scalable) 하고, 에러를 복원하며, 멀티미디어 정보를 스트리밍하는 이동 방송을 비롯한 다수의 멀티미디어 데이터 애플리케이션에서 사용되는 트랜스코더에 대한 요구가 존재한다.

발명의 내용

해결하려는 과제

[0004] 대역폭 자원과 이용 가능한 대역폭의 변화성이 제한되기 때문에, 무선 비디오 스트리밍 및 비디오 전화통신과 같은 다수의 멀티미디어 애플리케이션에서는 효율적인 비디오 압축이 유용하다. MPEG-4 (ISO/IEC), H.264 (ITU) 또는 유사한 비디오 코딩과 같은 일정 비디오 코딩 표준은 무선 방송과 같은 애플리케이션에 아주 적합한 고효율 코딩을 제공한다. 일반적으로, 예를 들어, 디지털 TV 프리젠테이션과 같은 몇몇 멀티미디어 데이터는 MPEG-2 와 같은 다른 표준에 따라 코딩된다. 따라서, 트랜스코더를 사용하여, 한 표준 (예를 들어, MPEG-2) 에 따라 코딩된 멀티미디어 데이터를 무선 방송 전에 다른 표준 (예를 들어, H.264) 으로 트랜스코딩하거나 변환하게 된다.

[0005] 개선 비율 최적화된 코덱 (improvements rate optimized codec) 은 에러 내성, 에러 복구 및 스케일러빌리티에 있어서 이점을 제공할 수 있다. 또한, 멀티미디어 데이터 자체로부터 결정된 정보를 사용하면, 에러 내성, 에러 복구 및 스케일러빌리티를 비롯한 인코딩에 대한 추가적인 개선을 제공할 수 있다. 따라서, 멀티미디어 데이터의 고효율 처리 및 압축을 제공하고, 멀티미디어 데이터 자체로부터 결정된 정보를 사용하고, 스케일러블 (scalable) 하고, 에러를 복원하며, 멀티미디어 정보를 스트리밍하는 이동 방송을 비롯한 다수의 멀티미디어 데이터 애플리케이션에서 사용되는 트랜스코더에 대한 요구가 존재한다.

과제의 해결 수단

[0006] 설명되고 도시된 각각의 독창적인 콘텐츠 기반 트랜스코딩 장치 및 방법은 몇몇 양태를 갖고, 그 단 하나의 양태도 자신의 바람직한 속성에 대해 단독으로 책임지지 않는다. 이하, 본 개시내용의 범위를 제한함 없이, 더 현저한 특징을 간략히 설명할 것이다. 이러한 설명을 고려한 후에, 특히, "상세한 설명" 으로 지칭되는 섹션을 읽은 후에, 이러한 콘텐츠 구동 트랜스코딩의 특징이 멀티미디어 데이터 처리 장치 및 방법에 대한 개선을 어떻게 제공하는지를 이해할 것이다.

[0007] 본원에서 설명되는 독창적인 양태는 인코더 (예를 들어, 트랜스코더에서 사용되는 인코더) 의 여러 모듈 또는 컴포넌트에서, 또한 멀티미디어 데이터를 인코딩하는 여러 방법에 대해 콘텐츠 정보를 사용하는 것에 관한 것이다. 트랜스코더는 콘텐츠 정보를 사용하여 멀티미디어 데이터를 트랜스코딩하는 것을 조정할 수 있다. 콘텐츠 정보는, 예를 들어, 비디오와 함께 수신되는 메타데이터와 같이, 다른 소스로부터 수신될 수 있다. 트랜스코더는 여러 상이한 처리 동작을 통해 콘텐츠 정보를 생성하도록 구성될 수 있다. 몇몇 양태에서, 트랜스코더는 멀티미디어 데이터의 콘텐츠 분류를 생성하고, 그 다음에, 이 콘텐츠 분류는 하나 이상의 인코딩 프로세스에서 사용된다. 몇몇 양태에서, 콘텐츠 구동 트랜스코더는 멀티미디어 데이터의 공간 및 시간 콘텐츠

정보를 결정할 수 있고, 채널에 걸쳐 콘텐츠 인식 균일 (content-aware uniform) 품질을 위한 콘텐츠 정보와, 콘텐츠 분류 기반 압축/비트 할당을 사용할 수 있다.

[0008] 몇몇 양태에서는, 멀티미디어 데이터의 콘텐츠 정보 (예를 들어, 메타데이터, 콘텐츠 매트릭 및/또는 콘텐츠 분류) 가 얻어지거나 계산되고, 그 다음에, 인코딩을 위한 멀티미디어 데이터의 처리에서 사용하기 위해 트랜스코더의 컴포넌트에 제공된다. 예를 들어, 전처리기는 장면 전환 검출, IVTC (inverse telecine) 수행, 디인터레이싱, 모션 보상 및 잡음 억제 (예를 들어, 2D 웨이브렛 변환), 및 공간-시간 잡음 저감 (예를 들어, 아티팩트 제거, 디링잉, 디블록킹 및/또는 디노이징) 에 대한 일정 콘텐츠 정보를 사용할 수 있다. 또한, 몇몇 양태에서, 전처리기는, SD (standard definition) 에서 QVGA (Quarter Video Graphics Array) 로 다운샘플링 시, 적절한 "안전 (safe)" 및 "액션 핸들링 (action handling)" 영역을 결정하는 것과 같이, 공간 해상도 다운샘플링에 대한 콘텐츠 정보를 사용할 수 있다.

[0009] 몇몇 양태에서는, 인코더가 콘텐츠 정보를 계산하도록 구성되는 콘텐츠 분류 모듈을 포함한다. 그 인코더는 콘텐츠 분류를 사용하여, 각 MB 에 대한 QP (quantization parameter) 를 결정하는데 있어서 비트 레이트 제어 (예를 들어, 비트 할당) 를 수행하고, 컬러 ME (motion estimation) 의 수행, MV (motion vector) 예측의 수행과, 베이스 계층과 인헨스먼트 계층을 제공하는데 있어서의 스케일러빌리티를 수행하며, 또한, 콘텐츠 분류를 사용하여, 예를 들어, 적응 인트라 리프레시나 경계 정렬 프로세스를 비롯한 예측 계층 구성과 에러 내성 방식에 영향을 줌으로써, 그리고, 인헨스먼트 계층에서 중복 I-프레임 데이터를 제공함으로써 에러 내성을 수행한다. 몇몇 양태에서, 트랜스코더는 채널에 걸쳐 최적 멀티미디어 데이터 품질을 유지하기 위해 데이터 멀티플렉서와 함께 콘텐츠 분류를 사용한다. 몇몇 양태에서, 인코더는 콘텐츠 분류 정보를 사용하여, I-프레임이 인코딩된 데이터에 주기적으로 나타나도록 함으로써, 고속 채널 스위칭을 허용할 수 있다. 또한, 이러한 구현은, 랜덤 액세스 스위칭과 에러 내성 (예를 들어, 콘텐츠 분류에 기초함) 이 예측 계층 구성을 통해 효과적으로 결합하여 에러에 대한 견고성을 증가시키면서 코딩 효율을 개선할 수 있도록, 에러 내성을 위해 인코딩된 데이터에서 필요할 수도 있는 I-블록을 사용할 수도 있다.

[0010] 일 양태에서, 멀티미디어 데이터를 처리하는 방법은 멀티미디어 데이터의 콘텐츠를 분류하는 단계 및 콘텐츠 분류에 기초하여 멀티미디어 데이터를 제 1 데이터 그룹과 제 2 데이터 그룹으로 인코딩하는 단계를 포함하고, 제 1 데이터 그룹은 일 계수를 포함하고, 제 2 데이터 그룹은 제 1 데이터 그룹 계수와 연관된 제 1 차분 세분화를 포함한다. 인코딩 단계는 멀티미디어 데이터의 콘텐츠 분류에 기초하여 비트 레이트를 결정하는 단계, 및 비트 레이트에 기초하여 멀티미디어 데이터를 인코딩하는 단계를 포함할 수 있다. 콘텐츠 분류 단계는 멀티미디어 데이터의 복잡도를 결정하는 단계를 포함할 수 있고, 선택된 멀티미디어 데이터는 멀티미디어 데이터의 복잡도에 기초하여 인코딩된다. 멀티미디어 데이터 복잡도는 시간 복잡도 또는 공간 복잡도를 포함할 수 있고, 또는 시간 복잡도 및 공간 복잡도를 포함할 수 있다. 인코딩 단계는 단지 제 1 데이터 그룹만을 디코딩하는 것을 허용하거나 제 1 데이터 그룹과 제 2 데이터 그룹을 단일 결합된 데이터 그룹으로 디코딩하는 것을 허용하도록 멀티미디어 데이터를 인코딩하는 단계를 포함할 수 있다. 제 1 차분 세분화는 제 1 데이터 그룹의 디코딩 결과로서 생성된 프레임 데이터와 선택된 비디오 프레임 간의 차이를 표시할 수 있다. 제 1 데이터 그룹은 베이스 계층일 수 있고, 제 2 데이터 그룹은 인헨스먼트 계층일 수 있다. 또한, 본 발명의 방법은 원래 베이스 계층 나머지 에러 계수 또는 원래 인헨스먼트 계층 나머지 에러 계수 중 하나로부터 계수를 선택하는 단계, 및 선택된 계수 및 원래 인헨스먼트 계층 나머지 에러 계수에 기초하여 제 1 차분 세분화를 계산하는 단계를 포함할 수 있다. 또한, 인코딩 단계는 제 1 데이터 그룹 내의 매크로블록 헤더 정보와 모션 벡터 정보를 인코딩하는 단계를 더 포함할 수 있다. 인코딩 단계는 제 1 데이터 그룹을 제 1 스텝 크기로 양자화하는 단계, 및 제 2 데이터 그룹을 제 2 스텝 크기로 양자화하는 단계를 더 포함할 수 있고, 제 1 스텝 크기와 제 2 스텝 크기는 스케일 인자에 의해 상관된다. 또한, 인코딩 단계는 제 1 데이터 그룹을 인코딩하는데 사용되는 제 1 양자화 스텝 크기를 갖는 제 1 양자화 파라미터를 결정하는 단계, 및 제 2 데이터 그룹을 인코딩하는데 사용되는 제 2 양자화 스텝 크기를 갖는 제 2 양자화 파라미터를 결정하는 단계를 더 포함할 수 있고, 제 1 및 제 2 양자화 파라미터는 선택된 프레임 데이터의 콘텐츠 정보에 기초하여 결정되고, 상기 제 1 양자화 스텝 크기는 상기 제 2 양자화 스텝 크기보다 덜 세분화된다. 다른 양태에서, 인코딩 단계는 I-프레임, P-프레임 또는 그 임의의 조합을 사용하여 제 1 데이터 그룹을 인코딩하는 단계, 및 I-프레임, P-프레임, B-프레임 또는 그 임의의 조합을 사용하여 제 2 데이터 그룹을 인코딩하는 단계를 포함한다.

[0011] 다른 양태에서, 멀티미디어 데이터를 인코딩하는 장치는 멀티미디어 데이터의 콘텐츠를 분류하는 수단, 및 콘텐츠 분류에 기초하여 멀티미디어 데이터를 제 1 데이터 그룹과 제 2 데이터 그룹으로 인코딩하는 수단을 포함하고, 제 1 데이터 그룹은 일 계수를 포함하고, 제 2 데이터 그룹은 제 1 데이터 그룹 계수와 연관된 제 1 차분

세분화를 포함한다. 인코딩 수단은 멀티미디어 데이터의 콘텐츠 분류에 기초하여 비트 레이트를 결정하는 수단, 및 비트 레이트에 기초하여 멀티미디어 데이터를 인코딩하는 수단을 포함할 수 있다. 콘텐츠 분류 수단은 멀티미디어 데이터의 복잡도를 결정하는 수단을 포함할 수 있고, 선택된 멀티미디어 데이터는 멀티미디어 데이터의 복잡도에 기초하여 인코딩되고, 복잡도는 시간 복잡도 또는 공간 복잡도를 포함하거나, 시간 복잡도 및 공간 복잡도를 포함한다. 인코딩 수단은 단지 제 1 데이터 그룹만을 디코딩하는 것을 허용하거나 제 1 데이터 그룹과 제 2 데이터 그룹을 단일 결합된 데이터 그룹으로 디코딩하는 것을 허용하는 수단을 포함할 수 있다.

[0012] 다른 양태에서, 본 발명의 장치는 멀티미디어 데이터의 콘텐츠를 분류하고 콘텐츠 분류 데이터를 제공하도록 구성된 콘텐츠 분류 모듈, 및 콘텐츠 분류에 기초하여 멀티미디어 데이터를 제 1 데이터 그룹과 제 2 데이터 그룹으로 인코딩하도록 구성된 인코더를 포함하고, 제 1 데이터 그룹은 일 계수를 포함하고, 제 2 데이터 그룹은 제 1 데이터 그룹 계수와 연관된 제 1 차분 세분화를 포함한다. 인코더는 콘텐츠 분류에 기초하여 비트 할당을 결정하도록 구성된 비트 레이트 컴포넌트를 포함할 수 있고, 인코딩 컴포넌트는 또한 비트 할당을 사용하여 선택된 멀티미디어 데이터를 인코딩하도록 구성된다.

[0013] 다른 양태에서, 머신 판독가능 매체는, 실행 시, 머신으로 하여금, 멀티미디어 데이터의 콘텐츠를 분류하게 하고, 콘텐츠 분류에 기초하여 멀티미디어 데이터를 제 1 데이터 그룹과 제 2 데이터 그룹으로 인코딩하게 하는 명령들을 포함하고, 제 1 데이터 그룹은 일 계수를 포함하고, 제 2 데이터 그룹은 제 1 데이터 그룹 계수와 연관된 제 1 차분 세분화를 포함한다.

[0014] 다른 양태에서, 프로세서는 멀티미디어 데이터의 콘텐츠를 분류하고, 콘텐츠 분류에 기초하여 멀티미디어 데이터를 제 1 데이터 그룹과 제 2 데이터 그룹으로 인코딩하도록 구성되고, 제 1 데이터 그룹은 일 계수를 포함하고, 제 2 데이터 그룹은 제 1 데이터 그룹 계수와 연관된 제 1 차분 세분화를 포함한다.

도면의 간단한 설명

[0015] 도 1a 는 상이한 비디오 포맷 간의 트랜스코딩을 위한 트랜스코더를 포함한 미디어 방송 시스템의 블록도이다.

도 1b 는 멀티미디어 데이터를 인코딩하고, 인코딩된 제 1 데이터 그룹과 인코딩된 제 2 데이터 그룹을 제공하도록 구성된 인코더의 블록도이다.

도 1c 는 멀티미디어 데이터를 인코딩하도록 구성된 프로세서의 블록도이다.

도 2 는 도 1 의 시스템의 트랜스코더의 일 예의 블록도이다.

도 3 은 도 2 의 트랜스코더에서 사용되는 파서 (parser) 의 동작을 도시한 흐름도이다.

도 4 는 도 2 의 트랜스코더에서 사용되는 디코더의 동작을 도시한 흐름도이다.

도 5 는 도 2 의 트랜스코더에 의해 수행되는 동작 순서를 도시한 시스템 타이밍도이다.

도 6 은 도 2 의 트랜스코더에서 사용될 수도 있는 전처리기의 동작 및 기능 순서를 도시한 흐름도이다.

도 7 은 도 2 의 트랜스코더에서 사용될 수도 있는 예시적인 2-패스 인코더의 블록도이다.

도 8 은 콘텐츠 분류와 텍스처 및 모션 값을 연관시키는 방법의 일 양태를 도시한 분류 차트의 일 예를 도시한다.

도 9 는 도 7 의 인코더에서 사용되는 것과 같은 콘텐츠 분류에 대한 예시적인 동작을 도시한 흐름도이다.

도 10 은 도 7 의 인코더에서 사용되는 것과 같은 레이트 제어의 동작을 도시한 흐름도이다.

도 11 은 도 7 의 인코더에서 사용되는 것과 같은 예시적인 모션 추정기의 동작을 도시한 흐름도이다.

도 12 는 도 7 의 인코더에서 사용되는 것과 같은 예시적인 모드 결정 인코더 기능의 동작을 도시한 흐름도이다.

도 13 은 도 7 의 인코더에서 사용되는 스케일러빌리티를 달성하는 예시적인 동작을 도시한 흐름도이다.

도 14 는 예를 들어, 도 7 의 인코더에서 일어나는 것과 같은 레이트 왜곡 데이터 흐름을 일으키는 예시적인 동작을 도시한 흐름도이다.

- 도 15 는 인코딩 복잡도, 할당된 비트 및 인간 시각 품질 간의 관계를 도시한 그래프이다.
- 도 16 은 비선형 장면 검출 공식을 도시한 그래프이다.
- 도 17a 는 획득되었거나, 수신되었거나, 또는 그렇지 않은 경우에는 액세스 가능한 멀티미디어 데이터의 처리를 도시한 흐름도이다.
- 도 17b 는 멀티미디어 인코딩 시스템의 블록도이다.
- 도 18 은 모션 추정/보상을 사용한 디인터레이싱 프로세스를 도시한 도면이다.
- 도 19 는 멀티미디어 통신 시스템의 블록도이다.
- 도 20 은 인헨스먼트 계층과 베이스 계층에서 비디오 비트스트림의 구성을 도시한 도면이다.
- 도 21 은 비디오 프레임 경계에 슬라이스를 정렬시킨 것을 도시한 도면이다.
- 도 22 는 예측 계층 구성을 도시한 블록도이다.
- 도 23 은 콘텐츠 정보에 기초하여 멀티미디어 데이터를 인코딩하는 방법을 도시한 프로세스 흐름도이다.
- 도 24 는 콘텐츠 정보 레벨에 기초하여 데이터 경계를 정렬시키도록 멀티미디어 데이터를 인코딩하는 방법을 도시한 프로세스 흐름도이다.
- 도 25 는 데이터 프레임의 안전한 액션 영역과 안전한 타이틀 영역을 도시한 그래프이다.
- 도 26 은 데이터 프레임의 안전한 액션 영역을 도시한 그래프이다.
- 도 27 은 멀티미디어 콘텐츠 정보에 기초한 적응 인트라 리프레시를 사용하여 멀티미디어 데이터를 인코딩하는 프로세스를 도시한 프로세스 흐름도이다.
- 도 28 은 멀티미디어 콘텐츠 정보에 기초한 중복 I 프레임을 사용하여 멀티미디어 데이터를 인코딩하는 프로세스를 도시한 프로세스 흐름도이다.
- 도 29 는 현재 프레임과 이전 프레임 MV_p 간, 또한 현재 프레임과 다음 프레임 MV_n 간 모션 보상 벡터를 도시한다.
- 도 30 은 샷 검출을 도시한 프로세스 흐름도이다.
- 도 31 은 베이스 및 인헨스먼트 계층을 인코딩하는 것을 도시한 프로세스 흐름도이다.
- 도 32 는 매크로블록을 인코딩하는 것을 개략적으로 도시한다.
- 도 33 은 베이스 계층과 인헨스먼트 계층을 인코딩하는 모듈을 개략적으로 도시한다.
- 도 34 는 베이스 계층과 인헨스먼트 계층 계수 선택 프로세스의 일 예를 도시한다.
- 도 35 는 베이스 계층과 인헨스먼트 계층 계수 선택 프로세스의 다른 예를 도시한다.
- 도 36 은 베이스 계층과 인헨스먼트 계층 계수 선택 프로세스의 다른 예를 도시한다.
- 도 37 은 콘텐츠 정보에 기초하여 멀티미디어 데이터를 인코딩하는 것을 도시한 프로세스 흐름도이다.
- 도 38 은 역 텔레시네 프로세스에서 가능한 시스템 결정을 도시한 도면이다.
- 도 39 는 더블록킹 프로세스에 의해 필터링되는 매크로블록에서의 경계를 도시한다.
- 도 40 은 공간-시간 디인터레이싱 프로세스를 도시한 도면이다.
- 도 41 은 1-D 다상 (poly-phase) 리샘플링의 일 예를 도시한다.
- 도 42 는 비디오 스트리밍에서 적응 GOP 구조의 일 예를 도시한 흐름도이다.
- 필요에 따라, 도면 중 몇몇 그림에서 동일 참조 부호는 동일 구성요소를 지칭한다는 것에 주목하자.

발명을 실시하기 위한 구체적인 내용

35 U.S.C. § 119 에 따른 우선권 주장

[0016]

- [0017] 본 특허 출원은 (a) 2005 년 9 월 27 일 출원된 발명의 명칭이 "A VIDEO TRANSCODER FOR REAL-TIME STREAMING AND MOBILE BROADCAST APPLICATIONS" 인 미국 가출원 제 60/721,416 호, (b) 2006 년 4 월 4 일 출원된 발명의 명칭이 "A VIDEO TRANSCODER FOR REAL-TIME STREAMING AND MOBILE BROADCAST APPLICATIONS" 인 미국 가출원 제 60/789,377 호, (c) 2005 년 10 월 17 일 출원된 발명의 명칭이 "METHOD AND APPARATUS FOR SPATIO-TEMPORAL DEINTERLACING AIDED BY MOTION COMPENSATION FOR FIELD-BASED VIDEO" 인 미국 가출원 제 60/727,643 호, (d) 2005 년 10 월 17 일 출원된 발명의 명칭이 "METHOD AND APPARATUS FOR SHOT DETECTION IN VIDEO STREAMING" 인 미국 가출원 제 60/727,644 호, (e) 2005 년 10 월 17 일 출원된 발명의 명칭이 "A METHOD AND APPARATUS FOR USING AN ADAPTIVE GOP STRUCTURE IN VIDEO STREAMING" 인 미국 가출원 제 60/727,640 호, (f) 2005 년 10 월 24 일 출원된 발명의 명칭이 INVERSE TELECINE ALGORITHM BASED ON STATE MACHINE" 인 미국 가출원 제 60/730,145 호, 및 (g) 2006 년 4 월 3 일 출원된 발명의 명칭이 "SPATIO-TEMPORAL DEINTERLACING AIDED BY MOTION COMPRESSION FOR FIELD-BASED MULTIMEDIA DATA" 인 미국 가출원 제 60/789,048 호에 대한 우선권을 주장한다. 이들 7 개의 가출원 모두는 본원의 양수인에게 양도되며, 본원에서 명확하게 참조로서 병합하고 있다.
- [0018] 동시 계속 중인 특허 출원에 대한 참조
- [0019] 본 특허 출원은, 2006 년 3 월 10 일 출원된 발명의 명칭이 "CONTENT CLASSIFICATION FOR MULTIMEDIA PROCESSING" 인 미국 특허 출원 제 11/373,577 호에 관한 것으로, 이 특허 출원은 본원의 양수인에게 양도되며, 본원에서 명확하게 참조로서 병합하고 있다.
- [0020] 다음 상세한 설명은 본 개시내용에서 논의되는 일정 양태에 관한 것이다. 그러나, 본 발명은 다수의 상이한 방법으로 구현될 수 있다. 본 명세서에서, "하나의 양태" 또는 "일 양태" 에 대한 참조는 그 양태와 관련하여 설명되는 특정 특징, 구조 또는 특성이 적어도 일 양태에 포함되는 것을 의미한다. 본 명세서의 여러 위치에서, "일 양태에서", "일 양태에 따르면", 또는 "몇몇 양태에서" 와 같은 구의 사용은 반드시 모두 동일한 양태를 지칭하는 것은 아니고, 기타 양태를 제외한 서로 별개의 또는 다른 양태를 지칭하는 것도 아니다. 또한, 몇몇 양태에 의해 구현될 수도 있지만 다른 양태에 의해서는 구현될 수 없는 다양한 특징이 설명되어 있다. 이와 유사하게, 몇몇 양태에 대한 요건이 될 수도 있지만 다른 양태에 대한 요건은 될 수 없는 다양한 요건이 설명된다.
- [0021] 다음 설명은 본 발명의 예의 완전한 이해를 제공하는 상세를 포함한다. 그러나, 당업자라면, 일 예 또는 양태에 따른 프로세스 또는 디바이스의 모든 상세가 본원에 설명되거나 도시되어 있지 않더라도, 그 예를 실시할 수도 있다는 것을 이해할 수 있다. 예를 들어, 전기 컴포넌트는 본 발명의 예를 너무 상세히 설명하여 불명확하게 하지 않도록 그 전기 컴포넌트의 모든 전기 접속 또는 모든 전기 소자를 도시하지 않은 블록도로 도시될 수도 있다. 다른 경우에는, 그러한 컴포넌트, 기타 구조 및 기술을 도시하여, 본 발명의 예를 더 상세히 설명할 수도 있다.
- [0022] 본 개시내용은 인코딩되는 멀티미디어 데이터의 콘텐츠 정보를 사용하여 인코딩 및 트랜스코딩 장치 및 방법을 제어하는 것에 관한 것이다. (멀티미디어 데이터의) "콘텐츠 정보" 또는 "콘텐츠" 는 멀티미디어 데이터의 콘텐츠 관련 정보를 의미하는 광범위한 용어이고, 예를 들어, 메타데이터, 멀티미디어 데이터로부터 계산된 메트릭 및 예를 들어, 콘텐츠 분류와 같은 하나 이상의 메트릭과 연관된 콘텐츠 관련 정보를 포함할 수 있다. 콘텐츠 정보는 특정 애플리케이션에 따라 인코더에 제공되거나 인코더에 의해 결정될 수 있다. 콘텐츠 정보는 장면 전환 검출, 시간 처리 (temporal processing), 공간-시간 잡음 저감, 다운샘플링, 양자화를 위한 비트 레이트 결정, 스케일러빌리티, 에러 내성, 방송 채널에 걸쳐 최적 멀티미디어 유지 및 고속 채널 스위칭을 포함한, 다양한 양태의 멀티미디어 데이터 인코딩에 사용될 수 있다. 이들 양태 중 하나 이상을 사용하면, 트랜스코더는 멀티미디어 데이터의 처리를 조정하고, 콘텐츠 관련 인코딩된 멀티미디어 데이터를 생성할 수 있다. 또한, 트랜스코딩 양태를 설명하는 본원의 설명 및 도면은 인코딩 양태 및 디코딩 양태에 적용 가능할 수 있다.
- [0023] 트랜스코더 장치 및 방법은 한 포맷에서 다른 포맷으로 트랜스코딩하는 것에 관한 것으로, 예시적인 몇몇 양태에 따르면, 무선 채널을 통해 이동 디바이스로 송신하기 위해 MPEG-2 비디오를 인핸스드, 스케일러블 H.264 포맷으로 트랜스코딩하는 것과 관련하여 본원에 구체적으로 설명되어 있다. 그러나, MPEG-2 비디오를 H.264 포맷으로 트랜스코딩하는 것에 대한 설명은 본 발명의 범위를 한정하기 위한 것이 아니라, 본 발명의 몇몇 양태를 예시하기 위한 것일 뿐이다. 개시된 장치 및 방법은 랜덤 액세스 및 레이아웃 기능을 사용하여 에러 내성 인코딩을 지원하는 고효율 아키텍처를 제공하고, 또한, MPEG-2 및 H.264 이외의 비디오 포맷을 트랜스코딩

및/또는 인코딩하는 것에 적용 가능할 수 있다.

- [0024] 본원에서 사용되는 "멀티미디어 데이터" 또는 단순히 "멀티미디어" 는 비디오 데이터 (오디오 데이터를 포함할 수 있음), 오디오 데이터, 또는 비디오 데이터와 오디오 데이터 모두를 포함하는 광범위한 용어이다. 본원에서 광범위한 용어로서 사용되는 "비디오 데이터" 또는 "비디오" 는 프레임 기반 또는 필드 기반 데이터를 지칭하는 광범위한 용어로서, 이 데이터는 하나 이상의 이미지 또는 이미지 관련 시퀀스, 텍스트 포함 (containing text), 이미지 정보 및/또는 오디오 데이터를 포함하고, 달리 지정되지 않는 한, 멀티미디어 데이터 (예를 들어, 그 용어는 서로 교환 가능하게 사용될 수 있음) 를 지칭하는데 사용될 수 있다.
- [0025] 이하, 멀티미디어 데이터를 인코딩하기 위해 콘텐츠 정보를 사용할 수 있는 여러 트랜스코더 컴포넌트의 예와 프로세스의 예를 설명한다.
- [0026] 도 1a 는 멀티미디어 데이터 방송 시스템 (100) 의 몇몇 양태의 데이터 흐름을 도시한 블록도이다. 시스템 (100) 에서, 멀티미디어 데이터 제공자 (106) 는 인코딩된 멀티미디어 데이터 (104) 를 트랜스코더 (200) 에 통신한다. 인코딩된 멀티미디어 데이터 (104) 는 트랜스코더 (200) 에 의해 수신되고, 이 트랜스코더 (200) 는 블록 (110) 에서 멀티미디어 데이터 (104) 를 미가공 멀티미디어 데이터로 처리한다. 블록 (110) 에서의 처리는 인코딩된 멀티미디어 데이터 (104) 를 디코딩 및 파싱하고, 또한 멀티미디어 데이터를 처리하여 다른 포맷으로 인코딩하기 위해 준비하는 것이다. 디코딩된 멀티미디어 데이터는 블록 (112) 으로 제공되어, 멀티미디어 데이터를 소정의 멀티미디어 포맷 또는 표준으로 인코딩한다. 일단 멀티미디어 데이터가 인코딩되면, 블록 (114) 에서, 예를 들어, 무선 방송 시스템 (예를 들어, 셀룰러 전화 방송 네트워크) 을 통해, 또는 다른 통신 네트워크를 통해 송신을 위해 준비된다. 몇몇 양태에서, 수신된 멀티미디어 데이터 (104) 는 MPEG-2 표준에 따라 인코딩되었다. 트랜스코딩된 멀티미디어 데이터 (104) 가 인코딩된 후에, 트랜스코더 (200) 는 멀티미디어 데이터를 H.264 표준으로 인코딩한다.
- [0027] 도 1b 는 도 1a 의 블록 (110 및 112) 에서의 처리를 수행하도록 구성될 수 있는 트랜스코더 (130) 의 블록도이다. 트랜스코더 (130) 는 멀티미디어 데이터를 수신하고, 멀티미디어 데이터를 패킷화된 기본 스트림 (예를 들어, 서브타이틀, 오디오, 메타데이터, "미가공" 비디오, CC 데이터 및 프리젠테이션 타임스탬프) 으로 디코딩 및 파싱하고, 원하는 포맷으로 인코딩하며, 추후 처리 또는 송신을 위해 인코딩된 데이터를 제공하도록 구성될 수 있다. 트랜스코더 (130) 는 인코딩된 데이터를 2 개 이상의 데이터 그룹, 예를 들어, 인코딩된 제 1 데이터 그룹과 인코딩된 제 2 데이터 그룹 (이는 계층적 인코딩으로 지칭됨) 으로 제공하도록 구성될 수 있다. 양태들의 몇몇 예에서, 계층적 인코딩 방식에 따라 여러 데이터 그룹 (또는 계층) 은 상이한 품질 레벨로 인코딩될 수 있고, 제 1 데이터 그룹에서 인코딩된 데이터가 제 2 데이터 그룹에서 인코딩된 데이터보다 낮은 품질 (예를 들어, 표시 시, 낮은 시각 품질 레벨) 을 갖도록 포맷될 수 있다.
- [0028] 도 1c 는 멀티미디어 데이터를 트랜스코딩하도록 구성될 수 있으며, 도 1a 의 블록 (110 및 112) 으로 도시된 처리의 일부 또는 모두를 수행하도록 구성될 수 있는 프로세서 (140) 의 블록도이다. 프로세서 (140) 는 디코딩, 파싱, 전처리 (preprocessing) 및 인코딩을 비롯한, 본원에 설명된 트랜스코딩 프로세스 중 하나 이상을 수행하는 모듈 (124a...n) 을 포함할 수 있고, 처리를 위해 콘텐츠 정보를 사용할 수 있다. 또한, 프로세서 (140) 는 내부 메모리 (122) 를 포함하고, 다른 디바이스를 통해 직접 또는 간접으로 외부 메모리 (120) 와 통신하도록 구성될 수 있다. 또한, 프로세서 (140) 는, 멀티미디어 데이터를 수신하며, 제 1 데이터 그룹에서 인코딩된 데이터 및 제 2 데이터 그룹에서 인코딩된 데이터와 같은 인코딩된 데이터를 제공하는 것을 비롯한, 프로세서 (140) 외부의 하나 이상의 디바이스와 통신하도록 구성된 통신 모듈 (126) 을 포함한다. 양태들의 몇몇 예에서, 계층적 인코딩 방식에 따라 여러 데이터 그룹 (또는 계층) 은 상이한 품질 레벨로 인코딩될 수 있고, 제 1 데이터 그룹에서 인코딩된 데이터가 제 2 데이터 그룹에서 인코딩된 데이터보다 낮은 품질 (예를 들어, 표시 시, 낮은 시각 품질 레벨을 제공함) 을 갖도록 포맷될 수 있다.
- [0029] 트랜스코더 (130) 또는 (트랜스코딩을 위해 구성된) 전처리기 (140), 그 컴포넌트 및 거기에 포함된 프로세스는 하드웨어, 소프트웨어, 펌웨어, 미들웨어, 마이크로코드 또는 그 임의의 조합으로 구현될 수 있다. 예를 들어, 파서, 디코더, 전처리기 또는 인코더는 자립형 컴포넌트일 수도 있고, 하드웨어, 펌웨어, 미들웨어로서 다른 디바이스의 컴포넌트에 통합될 수도 있고, 또는 프로세서 상에서 실행되는 마이크로코드 또는 소프트웨어에 구현될 수도 있고, 또는 그 조합일 수도 있다. 소프트웨어, 펌웨어, 미들웨어 또는 마이크로코드에 구현 시, 모션 보상, 샷 분류 및 인코딩 프로세스를 수행하는 프로그램 코드 또는 코드 세그먼트는 저장 매체와 같은 머신 판독가능 매체에 저장될 수도 있다. 코드 세그먼트는 절차, 함수, 서브프로그램, 프로그램, 루틴, 서브루틴, 모듈, 소프트웨어 패키지, 클래스, 또는 명령, 데이터 구조 또는 프로그램 명령문의 임의의 조합을 표

현할 수도 있다. 코드 세그먼트는 정보, 데이터, 인수 (argument), 파라미터 또는 메모리 콘텐츠를 송수신함으로써 다른 코드 세그먼트 또는 하드웨어 회로에 연결될 수도 있다.

[0030] 트랜스코더 아키텍처의 예시적인 예

[0031] 도 2 는 도 1 의 멀티미디어 방송 시스템 (100) 에 도시된 트랜스코더 (200) 에 사용될 수도 있는 트랜스코더의 일 예의 블록도를 도시한다. 트랜스코더 (200) 는 아래에 상세히 설명되는 파서/디코더 (202), 전처리기 (226), 인코더 (228) 및 동기화 계층 (240) 을 포함한다. 본원에 설명되는 바와 같이, 트랜스코더 (200) 는 트랜스코딩 프로세스의 하나 이상의 양태에 대한 멀티미디어 데이터 (104) 의 콘텐츠 정보를 사용하도록 구성된다. 콘텐츠 정보는, 멀티미디어 메타데이터를 통해 트랜스코더 (200) 외부에 있는 소스로부터 얻어질 수 있고, 또는 트랜스코더에 의해, 예를 들어, 전처리기 (226) 에 의해 또는 인코더 (228) 에 의해 계산될 수 있다.

도 2 에 도시된 컴포넌트는 하나 이상의 트랜스코딩 프로세스에 대한 콘텐츠 정보를 사용하는 트랜스코더에 포함될 수 있는 컴포넌트를 예시하는 것이다. 특정 구현에서는, 트랜스코더 (200) 의 컴포넌트 중 하나 이상을 제외할 수도 있고, 또는 추가 컴포넌트를 포함시킬 수도 있다. 또한, 프로세스 또는 디바이스의 모든 상세가 본원에 설명되지는 않더라도, 당업자라면 본 발명을 실시할 수 있도록, 트랜스코더 및 트랜스코딩 프로세스의 부분이 설명된다.

[0032] 도 5 는 트랜스코더 (200) 의 여러 컴포넌트 및/또는 프로세스 동작의 시간 관계를 그림으로 도시한 타이밍도를 도시한다. 도 5 에 도시된 바와 같이, MPEG-2 비디오와 같은 인코딩된 스트리밍 비디오 (104; 인코딩된 멀티미디어 데이터) 는 파서 (205; 도 2) 에 의해 임의의 시간 0 에서 처음 수신된다. 그 다음에, 예를 들어, 디코더 (241) 와 함께 파서 (205) 에 의해, 비디오 스트림이 파싱되고 (501), 디멀티플렉싱되고 (502) 되며 디코딩된다 (503). 도시된 바와 같이, 이들 프로세스는 전처리기 (226; 도 2) 에 처리 데이터의 스트림 출력을 제공하도록, 약간의 타이밍 오프셋으로 동시에 일어날 수 있다. 시점 T_1 (504) 에서, 일단 전처리기 (226) 가 처리 결과의 출력을 시작하기에 충분한 데이터를 디코더 (214) 로부터 수신하면, 나머지 처리 단계는 사실상 순차적으로 이루어지는데, 제 1 패스 인코딩 (505), 제 2 패스 인코딩 (506) 및 재인코딩 (507) 은, 전처리 후에 시점 T_f (508) 에서 재인코딩이 완료될 때까지 순서대로 일어난다.

[0033] 본원에 설명된 트랜스코더 (200) 는 다양한 멀티미디어 데이터를 트랜스코딩하도록 구성될 수 있고, 트랜스코딩되는 멀티미디어 데이터의 타입에 관계없이 다수의 프로세스가 적용된다. 본원에 제공된 몇몇 예가 특히 MPEG-2 데이터를 H.264 데이터로 트랜스코딩하는 것에 관한 것이지만, 이들 예는 본 개시내용을 그러한 데이터에 한정하기 위한 것은 아니다. 아래에 설명되는 인코딩 양태는 임의의 적당한 멀티미디어 데이터 표준을 다른 적당한 멀티미디어 데이터 표준으로 트랜스코딩하는 것에 적용될 수 있다.

[0034] 파서/디코더

[0035] 도 2 를 다시 참조하면, 파서/디코더 (202) 는 멀티미디어 데이터 (104) 를 수신한다. 파서/디코더 (202) 는 멀티미디어 데이터 (104) 를 수신하고 그 데이터를 비디오 기본 스트림 (206; video elementary stream; video ES), 오디오 ES (208), PTS (210; presentation time stamp) 및 서브타이틀 (212) 과 같은 다른 데이터로 파싱하는 전송 스트림 파서 ("parser") (205) 를 포함한다. ES 는 단일 비디오 또는 오디오 인코더로부터 한 가지 타입의 데이터 (비디오 또는 오디오) 를 전달한다. 예를 들어, 비디오 ES 는 시퀀스 헤더 및 그 시퀀스의 모든 서브파트를 포함한 데이터 시퀀스에 대한 비디오 데이터를 포함한다. 패킷화된 기본 스트림, 또는 PES 는 패킷으로 만들어진 단일 ES 로 이루어지고, 통상, 그 각각은 추가된 패킷 헤더로 시작한다. PES 스트림은 하나의 소스로부터, 예를 들어, 하나의 비디오 또는 오디오 인코더로부터 한 가지 타입의 데이터만을 포함한다. PES 패킷은 전송 패킷의 고정 패킷 길이에 대응하지 않는 가변 길이를 갖고, 전송 패킷보다 훨씬 더 길어질 수도 있다. 전송 패킷이 PES 스트림으로 형성될 때, PES 헤더는 전송 패킷 헤더 바로 다음에, 전송 패킷 페이로드의 시작에 위치할 수 있다. 나머지 PES 패킷 콘텐츠는, PES 패킷 모두를 사용할 때까지, 연속하는 전송 패킷의 페이로드를 채운다. 최종 전송 패킷은 예를 들어, 바이트 (예를 들어, 바이트 = 0xFF (모두 1)) 를 채움으로써, 고정 길이까지 채워질 수 있다.

[0036] *파서 (205) 는 본원에 도시된 파서/디코더 (202) 의 부분인 디코더 (214) 에 비디오 ES (206) 를 통신한다. 다른 구성에서는, 파서 (205) 와 디코더 (214) 가 별개의 컴포넌트이다. PTS (210) 는 트랜스코더 PTS 생성기 (215) 에 송신되고, 이 트랜스코더 PTS 생성기 (215) 는 트랜스코더 (200) 로부터 방송 시스템에 송신되는 데이터를 배열하는데 사용되는 트랜스코더 (200) 에 고유한 별개의 프리젠테이션 타임스탬프를 생성할 수 있다.

트랜스코더 PTS 생성기 (215) 는 트랜스코더 (200) 의 동기화 계층 (240) 에 데이터를 제공하여 데이터 방송의 동기화를 조정하도록 구성될 수 있다.

[0037] 도 3 은 상술한 여러 패킷화된 기본 스트림의 파싱 아웃 시, 파서 (205) 가 따를 수도 있는 프로세스 (300) 의 일 예의 흐름도를 도시한다. 멀티미디어 데이터 (104) 가 콘텐츠 제공자 (106) 로부터 수신될 때 (도 1), 블록 (302) 에서 프로세스 (300) 는 시작한다. 파서 (205) 의 초기화가 수행되는 블록 (304) 으로 프로세스 (300) 가 진행한다. 초기화는 독립적으로 생성된 취득 커맨드 (306) 에 의해 트리거될 수도 있다. 예를 들어, 파서 (205) 와 독립적이며 외부에서 수신된 TV-스케줄 및 채널 라인업 정보에 기초하는 프로세스는 취득 커맨드 (306) 을 생성할 수도 있다. 또한, 실시간 전송 스트림 (TS) 버퍼 디스크립터 (308) 를 입력하여, 초기화와 메인 처리 모두를 도울 수도 있다.

[0038] 블록 (304) 에 도시된 바와 같이, 초기화는 커맨드 구문 검증을 취득하는 것, 제 1 패스 PSI/PSIP/SI (program specific information/program and system information protocol/system information) 처리를 수행하는 것, 특히 취득 커맨드 또는 PSI/PSIP/SI 일관성 검증에 관련된 처리를 수행하는 것, 각 PES 에 대한 PES 버퍼를 할당하는 것 및 (예를 들어, 원하는 취득 시작 인스턴트와의 정렬을 위한) 타이밍을 설정하는 것을 포함할 수 있다. PES 버퍼는 파싱된 ES 데이터를 유지하고, 각각의 파싱된 ES 데이터를 대응하는 오디오 디코더 (216), 테스트 인코더 (220), 디코더 (214) 또는 트랜스코더 PTS 생성기 (215) 에 통신한다.

[0039] 초기화 후에, 수신된 멀티미디어 데이터 (104) 의 메인 처리를 위한 블록 (310) 으로 프로세스 (300) 가 진행한다. 블록 (310) 에서의 처리는, 들어오는 멀티미디어 데이터가 적절한 PES 버퍼 내에 전달되도록, 타깃 패킷 식별자 (packet identifier; PID) 필터링, 연속적인 PSI/PSIP/SI 감시 및 처리, 및 (예를 들어, 원하는 취득 지속기간을 달성하기 위한) 타이밍 프로세스를 포함할 수 있다. 블록 (310) 에서 멀티미디어 데이터 처리의 결과로서, PES 버퍼의 프로그램 디스크립터 및 인디케이션인 "판독 (read)" 이 생성되어, 아래에서 설명하는 바와 같이, 디코더 (214; 도 2) 와 인터페이싱할 것이다.

[0040] 블록 (310) 후에, 프로세스 (300) 가 블록 (314) 으로 진행하여, 타이머 인터럽트의 생성 및 PES 버퍼 소비에 따른 PES 버퍼의 비움을 비롯한, 파싱 동작의 종료가 일어난다. 오디오, 비디오 및 서브타이틀 스트림과 같은 그 디스크립터에서 인용된 프로그램의 모든 관련 기본 스트림에 대해 PES 버퍼가 존재할 것이라는 것에 주목하자.

[0041] 다시 도 2 를 참조하면, 파서 (205) 는 트랜스코더 구현에 대응하는 오디오 디코더 (216) 에 오디오 ES (208) 를 송신하고, 인코딩된 텍스트 (216) 를 동기화 계층 (240) 에 제공하며, 오디오 정보를 디코딩한다. 서브타이틀 정보 (212) 는 텍스트 인코더 (220) 에 전달된다. 또한, 디코더 (214) 로부터의 CC (closed captioning) 데이터 (218) 는 텍스트 인코더 (220) 에 제공되고, 이 텍스트 인코더 (220) 는 서브타이틀 정보 (212) 와 CC 데이터 (218) 를 트랜스코더 (200) 에 의해 영향을 받은 포맷으로 인코딩한다.

[0042] 또한, 파서/디코더 (202) 는 디코더 (214) 를 포함하고, 이 디코더 (214) 는 비디오 ES (206) 를 수신한다. 디코더 (214) 는 비디오 데이터와 연관된 메타데이터를 생성할 수 있고, 인코딩된 비디오 패킷화 기본 스트림을 미가공 비디오 (224) 로 (예를 들어, 표준 화질 포맷으로) 디코딩하며, 비디오 ES 스트림 내의 비디오 CC 데이터를 처리한다.

[0043] 도 4 는 디코더 (214) 에 의해 수행될 수 있는 디코딩 프로세스 (400) 의 일 예를 도시한 흐름도를 도시한다. 블록 (402) 에서, 프로세스 (400) 는 비디오 기본 스트림 데이터 (206) 의 입력으로 시작한다. 프로세스 (400) 는 디코더가 초기화되는 블록 (404) 으로 진행한다. 초기화는 비디오 시퀀스 헤더 (VSH) 의 검출, 제 1 패스 VSH, 비디오 시퀀스 (VS) 및 VS 디스플레이 확장 처리 (비디오 포맷, 색 프라임리 및 행렬 계수를 포함) 의 수행 및 데이터 버퍼를 할당하여 디코딩된 영상, 관련 메타데이터 및 CC 데이터를 각각 버퍼링하는 것을 비롯한, 다수의 작업을 포함할 수도 있다. 또한, 파서 (205) 에 의해 제공되는 비디오 PES 버퍼 '판독' 정보 (406) 가 입력된다 (예를 들어, 이는 도 3 의 블록 (310) 에서 프로세서 (300) 에 의해 생성될 수 있음).

[0044] 블록 (404) 에서 초기화 후에, 프로세스 (400) 는 비디오 ES 의 메인 처리가 디코더 (214) 에 의해 수행되는 블록 (408) 으로 진행한다. 메인 처리는 새로운 데이터 가용도에 대한 비디오 PES 버퍼 '판독' 정보 또는 "인터페이스" 의 폴링 (polling), 비디오 ES 의 디코딩, 영상 경계에서 화소 데이터의 재구성 및 저장, 비디오와 a/v 의 동기화, 영상 경계에서 메타데이터의 생성 및 저장, 및 영상 경계에서 CC 데이터의 저장을 포함한다. 메인 처리 (408) 의 결과 블록 (410) 은 시퀀스 디스크립터, 디코딩된 영상 버퍼 디스크립터, 메타데이터 버퍼 디스크립터 및 CC 데이터 버퍼 디스크립터의 생성을 포함한다.

- [0045] 메인 처리 (408) 후에, 프로세스 (400) 는 종료 프로세스를 수행하는 블록 (412) 으로 진행한다. 종료 프로세스는 소정의 임계값보다 큰 특정 지속기간 동안 새로운 데이터가 발생하지 않음, 시퀀스 종료 코드의 검출 및/또는 명시적 종료 신호의 검출을 비롯한, 종료 조건을 결정하는 것을 포함할 수 있다. 종료 프로세스는, 아래에 설명되는 전처리기에 의해, CC 데이터 버퍼의 소비에 따라 디코딩된 영상, 관련 메타데이터 및 CC 데이터 버퍼의 비움을 더 포함할 수 있다. 프로세스 (400) 는 블록 (414) 에서 종료하고, 그 블록 (414) 에서, 비디오 ES 가 입력으로서 수신되기를 대기하는 상태에 들어갈 수 있다.
- [0046] 전처리
- [0047] 도 2, 또한 더 상세하게는 도 6 은 하나 이상의 전처리 동작에 대한 콘텐츠 정보를 사용할 수 있는 전처리 (226) 의 예시적인 양태를 도시한다. 전처리 (226) 는 파서/디코더 (202) 로부터 메타데이터 (222) 와 디코딩된 "미가공" 비디오 데이터 (224) 를 수신한다. 전처리 (226) 는 비디오 데이터 (224) 와 메타데이터 (222) 에 대한 일정 타입의 처리를 수행하고, 처리된 멀티미디어 (예를 들어, 베이스 계층 기준 프레임, 인헨스먼트 계층 기준 프레임, 대역폭 정보, 콘텐츠 정보) 와 비디오를 인코더 (228) 에 제공하도록 구성된다. 그러한 멀티미디어 데이터의 전처리는 데이터의 시각적인 명료도, 안티-에일리어싱 (anti-aliasing), 및 압축 효율을 개선할 수 있다. 통상, 전처리 (226) 는 파서 디코더 (202) 내의 디코더 (214) 에 의해 제공된 비디오 시퀀스를 수신하고, 이 비디오 시퀀스를 인코더 (228) 에 의한 추후 처리 (예를 들어, 인코딩) 를 위한 순차 비디오 시퀀스로 변환한다. 몇몇 양태에서, 전처리 (226) 는 역 텔레시네, 디인터레이싱, 필터링 (예를 들어, 아티팩트 제거, 디링잉, 디블록킹 및 디노이징), 리사이징 (예를 들어, 표준 화질에서 QVGA (Quarter Video Graphics Array) 로의 공간 해상도 다운샘플링) 및 GOP 구조 생성 (예를 들어, 계산 복잡도 맵 생성, 장면 전환 검출 및 페이드/플래시 검출) 을 비롯한, 다수의 동작을 위해 구성될 수 있다.
- [0048] 전처리 (226) 는 디코더로부터의 메타데이터를 사용하여 전처리 동작 중 하나 이상에 영향을 줄 수 있다. 메타데이터는 멀티미디어 데이터의 콘텐츠 ("콘텐츠 정보") 를 설명하거나 분류하는 것과 관련된 정보를 포함할 수 있고, 특히, 메타데이터는 콘텐츠 분류를 포함할 수 있다. 몇몇 양태에서, 메타데이터는 인코딩 동작에 필요한 콘텐츠 정보를 포함하지 않는다. 이와 같은 경우에, 전처리 (226) 는 콘텐츠 정보를 결정하고 전처리 동작에 대한 콘텐츠 정보를 사용하고/하거나 예를 들어, 디코더 (228) 와 같은 트랜스코더 (200) 의 다른 컴포넌트에 콘텐츠 정보를 제공하도록 구성될 수 있다. 몇몇 양태에서, 전처리 (226) 는 그와 같은 콘텐츠 정보를 사용하여, GOP 분할에 영향을 주고, 적당한 타입의 필터링을 결정하고/하거나 인코더에 통신되는 인코딩 파라미터를 결정할 수 있다.
- [0049] 도 6 은 전처리 (226) 에 포함될 수 있는 여러 프로세스 블록의 예시적인 예를 도시하고, 전처리 (226) 에 의해 수행될 수 있는 처리를 도시한다. 본 예에서, 전처리 (226) 는 메타데이터와 비디오 (222, 224) 를 수신하고, (처리된) 메타데이터와 비디오를 포함한 출력 데이터 (614) 를 인코더 (228) 에 제공한다. 통상, 3 가지 타입의 비디오가 수신될 수도 있다. 첫째로, 수신된 비디오는 디인터레이싱을 필요로 하지 않는 순차 비디오일 수 있다. 둘째로, 비디오 데이터는 24 fps 영화 시퀀스로부터 변환된 인터레이스된 비디오, 텔레시네된 비디오일 수 있고, 이 경우에는, 역 텔레시네 동작이 필요할 수도 있다. 셋째로, 그 비디오는 텔레시네되지 않은 인터레이스된 비디오일 수 있다. 전처리 (226) 는 아래에서 설명되는 이들 타입의 비디오를 처리할 수 있다.
- [0050] 블록 (601) 에서, 전처리 (226) 는 수신된 비디오 데이터 (222, 224) 가 순차 비디오인지를 결정한다. 몇몇 경우에, 이러한 결정은, 메타데이터가 그러한 정보를 포함하는 경우에 메타데이터로부터, 또는 비디오 데이터 자체를 처리하여 결정될 수 있다. 예를 들어, 아래에 설명되는 역 텔레시네 프로세스는, 수신된 비디오 (222) 가 순차 비디오인지를 결정할 수 있다. 순차 비디오인 경우에, 프로세스는 백색 가우스 잡음과 같은 잡음을 줄이기 위해 비디오에 대해 필터링 (예를 들어, 디노이저 (denoiser)) 동작이 수행되는 블록 (607) 으로 진행한다. 비디오 데이터 (222, 224) 가 순차 비디오가 아닌 경우에, 블록 (601) 에서, 프로세스는 블록 (604) 의 위상 검출기 (604) 로 진행한다.
- [0051] 위상 검출기 (604) 는 텔레시네에서 시작한 비디오와 표준 방송 포맷에서 시작한 비디오를 구별한다. 비디오가 텔레시네되었다는 결정이 이루어지면, 위상 검출기 (604) 에 존재하는 "예" 결정 경로), 텔레시네된 비디오는 역 텔레시네 (606) 에서 그 원래 포맷으로 리턴된다. 중복 프레임이 식별 및 제거되고, 동일 비디오 프레임으로부터 얻어진 필드는 완전한 이미지로 재구성된다. 재구성된 영화 이미지의 시퀀스가 1 초의 1/24 의 일정한 간격으로 사진 기록되기 때문에, GOP 분할기 (612) 또는 디코더 (228) 에서 수행되는 모션 추정 프로세스는 불규칙한 시간축을 갖는 텔레시네된 데이터보다는 역 텔레시네된 이미지를 사용하여 더 정확해진다.

[0052] 일 양태에 따르면, 위상 검출기 (604) 는 비디오 프레임의 수신 후에 일정 결정을 행한다. 이들 결정은 (i) 텔레시네 출력으로부터의 현재 비디오와 3:2 폴 다운 위상이 도 38 에 도시된 5 개의 위상 P0, P1, P2, P3 및 P4 중 하나인지 여부와, (ii) 그 비디오가 통상의 NTSC 로서 생성되었다는 것을 포함한다. 그 결정은 위상 P5 로서 지칭된다. 이들 결정은 도 2 에 도시된 위상 검출기 (604) 의 출력으로서 나타난다. "예" 로 표시된 위상 검출기 (604) 로부터의 경로는 역 텔레시네 (606) 를 작동시키는데, 이는, 정확한 폴 다운 위상이 제공되어, 동일 사진 이미지로부터 형성된 필드를 분류하여 그들을 결합시킬 수 있다는 것을 나타낸다. 이와 유사하게, "아니오" 로 표시된 위상 검출기 (604) 로부터의 경로는 디인터레이서 (605) 를 작동시켜, 외관 NTSC 프레임을 최적 처리를 위한 필드로 분리한다. 위상 검출기 (604) 는 상이한 타입의 비디오를 어느 때라도 수신할 수 있기 때문에 연속하여 비디오 프레임을 분석할 수 있다. 일 예로서, NTSC 표준에 따르는 비디오가 광고로서 비디오 내에 삽입될 수도 있다. 역 텔레시네 후에, 결과로서 생성된 순차 비디오는 백색 가우스 잡음을 줄이는데 사용될 수 있는 디노이저 (필터) (607) 로 송신된다.

[0053] 통상의 NTSC 비디오가 인식되는 경우에 (위상 검출기 (601) 로부터 "아니오" 경로인 경우에), 통상의 NTSC 비디오는 압축을 위해 디인터레이서 (605) 에 송신된다. 디인터레이서 (605) 가 인터레이싱된 필드를 순차 비디오로 변환한 후에, 디노이징 동작을 순차 비디오에 대해 수행할 수 있다. 디인터레이싱 처리의 일 예시적인 예는 아래에 설명되어 있다.

[0054] 텔레비전과 같은 통상의 아날로그 비디오 디바이스는 인터레이싱 방식으로 비디오를 렌더링하는데, 즉, 그러한 디바이스는 홀수 주사선 (홀수 필드) 와, 짝수 주사선 (짝수 필드) 를 송신한다. 신호 샘플링 관점에서, 이는 다음 수학식에 의해 설명된 패턴에 있어서 공간-시간 서브샘플링과 등가이다.

수학식 1

$$F(x, y, n) = \begin{cases} \Theta(x, y, n), & \text{if } y \bmod 2 = 0 \text{ 짝수 필드의 경우,} \\ \Theta(x, y, n), & \text{if } y \bmod 2 = 1 \text{ 홀수 필드의 경우,} \\ \text{소거,} & \text{otherwise,} \end{cases}$$

[0056] 여기서, Θ 는 원래 프레임 영상을 나타내고, F 는 인터레이싱된 필드를 나타내며, (x, y, n) 은 각각 화소의 수평, 수직 및 시간 위치를 나타낸다.

[0057] 일반성을 잃지 않으면서, 본 개시내용 전체에서 $n = 0$ 이 짝수 필드인 것으로 가정하여, 상기 수학식 1 을 다음 식과 같이 단순화할 수 있다.

수학식 2

$$F(x, y, n) = \begin{cases} \Theta(x, y, n), & \text{if } y \bmod 2 = n \bmod 2, \\ \text{소거,} & \text{otherwise,} \end{cases}$$

[0059] 수평 차원에서는 데시메이션 (decimation) 이 수행되지 않기 때문에, 서브샘플링 패턴은 다음 $n \sim y$ 좌표에서 도시될 수 있다.

[0060] 디인터레이서의 목적은 인터레이싱된 비디오 (필드 시퀀스) 를 인터레이싱되지 않은 순차 프레임 (프레임 시퀀스) 으로 변환하기 위한 것이다. 즉, 짝수 및 홀수 필드를 보간하여, 완전한 프레임 영상을 "복원" 하거나 생성하기 위한 것이다. 이는 수학식 3 에 표현될 수 있다.

수학식 3

$$F_o(x, y, n) = \begin{cases} F(x, y, n), & y \bmod 2 = n \bmod 2, \\ F_i(x, y, n), & \text{otherwise}, \end{cases}$$

[0061]

[0062] 여기서, F_i 는 손실된 화소에 대한 디인터레이싱 결과를 나타낸다.

[0063] 도 40 은 Wmed 필터링과 모션 추정을 사용하여 인터레이싱된 멀티미디어 데이터로부터 순차 프레임을 생성하는 디인터레이서 (605) 의 일 양태의 몇몇 양태를 도시한 블록도이다. 도 40 의 상부는 현재 필드, 2 개의 이전 필드 (PP 필드 및 P 필드), 및 2 개의 후속 필드 (다음 필드 및 다음 다음 필드) 로부터의 정보 사용하여 생성될 수 있는 모션 세기 맵 (4002) 을 도시한다. 모션 세기 맵 (4002) 은 현재 프레임을 2 개 이상의 상이한 모션 레벨로 분류 또는 분할하고, 아래에서 더 상세히 설명되는 공간-시간 필터링에 의해 생성될 수 있다. 몇몇 양태에서, 모션 세기 맵 (4002) 은 아래 수학식 4 내지 수학식 8 을 참조하여 설명되는 것과 같이 정적 영역, 슬로우 모션 영역 및 패스트 모션 영역을 식별하도록 생성된다. 예를 들어, Wmed 필터 (4004) 와 같은 공간-시간 필터는 모션 세기 맵에 기초한 기준을 사용하여 인터레이싱된 멀티미디어 데이터를 필터링하고, 공간-시간 임시 디인터레이싱된 프레임을 생성한다. 몇몇 양태에서, Wmed 필터링 프로세스는, $[-1, 1]$ 의 수평 근방, $[-3, 3]$ 의 수직 근방 및 도 40 에 도시된 5 개의 필드 (PP 필드, P 필드, 현재 필드, 다음 필드, 다음 다음 필드) 에 의해 표현되는 5 개의 인접 필드의 시간 근방을 수반하고, Z^{-1} 은 일 필드의 지연을 나타낸다. 현재 필드와 관련하여, 다음 필드와 P 필드는 넌패리티 (non-parity) 필드이고, PP 필드와 다음 다음 필드는 패리티 필드이다. 공간-시간 필터링에 사용되는 "근방" 은 필터링 동작 동안에 실제로 사용되는 필드와 화소의 공간 및 시간 위치를 지칭하고, 예를 들어, 도 6 및 도 7 에서 도시된 것과 같은 "애퍼처 (aperture)" 로서 도시될 수 있다.

[0064] 또한, 디인터레이서 (605) 는 Wmed 필터 (4004) 에 의해 생성되는 공간-시간 임시 디인터레이싱된 프레임을 필터링하도록 구성된 디노이저 (4006; 디노이징 필터) 를 포함할 수 있다. 공간-시간 임시 디인터레이싱된 프레임을 디노이징함으로써, 특히 소스 인터레이싱된 멀티미디어 데이터가 백색 잡음에 의해 오염되는 경우에 후속 모션 검색 프로세스를 더 정확하게 한다. 또한, Wmed 영상에서 짝수 행과 홀수 행 간의 에일리어스를 적어도 부분적으로 제거할 수 있다. 디노이저 (4006) 는 웨이블릿 축소 및 웨이블릿 위너 필터 기반 디노이저를 비롯한 다양한 필터로서 구현될 수 있다. 디노이저 필터는, 모션 보상 정보를 사용하여 더 처리하기 전에, 후보 Wmed 프레임에서 잡음을 제거하는데 사용될 수 있고, Wmed 프레임에 존재하는 잡음을 제거할 수 있으며, 신호의 주파수 콘텐츠에 관계없이 존재하는 신호를 보유할 수 있다. 웨이블릿 필터를 포함한 다양한 타입의 디노이징 필터를 사용할 수 있다. 웨이블릿은 공간과 스케일링 영역 모두에서 주어진 신호를 국부화하는데 사용되는 함수 클래스이다. 웨이블릿의 기초가 되는 기본 사상은, 웨이블릿 표현에서의 작은 변화가 그에 대응하여 원래 신호에서 작은 변화를 생성하도록, 상이한 스케일 또는 해상도로 신호를 분석한다는 것이다.

[0065] 또한, 웨이블릿 축소 또는 웨이블릿 위너 필터는 디노이저로서 적용될 수 있다. 웨이블릿 축소는 잡음 신호를 웨이블릿 변환하고, 그 다음에, 큰 계수에 영향을 주지 않으면서 작은 웨이블릿 계수를 0 (또는 작은 값) 으로 축소하는 것으로 이루어진다. 끝으로, 역 변환을 수행하여 추정된 신호를 취득한다.

[0066] 디노이징 필터링은 잡음 환경에서 모션 보상의 정확도를 향상시킨다. 웨이블릿 축소 디노이징은 웨이블릿 변환 영역의 축소를 수반할 수 있고, 통상 3 개의 단계를 포함한다: 선형 순방향 웨이블릿 변환, 비선형 축소 디노이징 및 선형 역 웨이블릿 변환. 위너 필터는 추가적인 잡음 및 블러링에 의해 저하되는 이미지를 개선하는데 사용될 수 있는 MSE-최적 필터이다. 일반적으로, 그러한 필터는, 당해 기술분야에서 공지되어 있으며, 예를 들어, 상기 참조한 "Ideal spatial adaptation by wavelet shrinkage" 에서 설명되어 있고, "S. P. Ghael, A. M. Sayeed 및 R. G. Baraniuk 에 의한 "Improvement Wavelet denoising via empirical Wiener filtering" (*Proceedings of SPIE*, vol. 3169, pp. 389-399, San Diego, 1997 년 7 월) 에 설명되어 있으며, 그 전체를 본원에서 명확히 참조로서 병합하고 있다.

[0067] 몇몇 양태에서, 디노이징 필터는 이원직교 3 차 B-스플라인 웨이블릿 필터의 일 양태에 기초한다. 그러한 일 필터는, 다음 순방향 및 역방향 변환에 의해 정의될 수 있다:

수학식 4

$$h(z) = \frac{3}{4} + \frac{1}{2}(z + z^{-1}) + \frac{1}{8}(z + z^{-2}) \quad (\text{순방향 변환})$$

[0068]

수학식 5

$$g(z) = \frac{5}{4}z^{-1} - \frac{5}{32}(1 + z^{-2}) - \frac{3}{8}(z + z^{-3}) - \frac{3}{32}(z^2 + z^{-4}) \quad (\text{역방향 변환})$$

[0069]

[0070]

디노이징 필터를 적용하면, 잡음 환경에서 모션 보상의 정확도를 증가시킬 수 있다. 그러한 필터의 구현은 그 전체를 본원에서 명확히 참조로서 병합하고 있는 "Ideal spatial adaptation by wavelet shrinkage" (D.L. Donoho 및 I.M. Johnstone, *Biometrika*, vol. 8, pp. 425-455, 1994 년) 에 더 설명되어 있다.

[0071]

도 40 의 하부는 인터레이싱된 멀티미디어 데이터의 모션 정보 (예를 들어, 모션 벡터 후보, 모션 추정, 모션 보상) 를 결정하는 일 양태를 도시한다. 특히, 도 40 은, 선택된 프레임의 모션 보상된 임시 순차 프레임들 생성한 후에, Wmed 임시 프레임과 결합하여 결과로서 생성된 "최종" 순차 프레임 (디인터레이싱된 현재 프레임 (4014) 으로서 도시됨) 을 형성하는데 사용되는 모션 추정 및 모션 보상 방식을 도시한다. 몇몇 양태에서, 인터레이싱된 멀티미디어 데이터의 모션 벡터 (MV) 후보 (또는 추정) 는 외부 모션 추정기로부터 디인터레이서 에 제공되고, 양방향 모션 추정기 및 보상기 (4018; ME/MC) 에 대한 시작점을 제공하는데 사용된다. 몇몇 양태에서, MV 후보 선택기 (4022) 는 예를 들어, 디인터레이싱된 이전 프레임 (4020) 내의 블록과 같은 이전 처리된 블록의 MV 와 같은 처리 중인 블록의 MV 후보를 위한 근방 블록에 대해 미리 결정된 MV 를 사용한다. 모션 보상은 이전 디인터레이싱된 프레임 (70) 과 다음 (예를 들어, 장래) Wmed 프레임 (4008) 에 기초하여, 양 방향으로 행해질 수 있다. 현재 Wmed 프레임 (4010) 과 모션 보상된 (MC) 현재 프레임 (4016) 은 결합기 (4012) 에 의해 병합 또는 결합된다. 결과로서 디인터레이싱된 현재 프레임 (4014), 즉, 순차 프레임은 다시 ME/MC (4018) 에 제공되어 디인터레이싱된 이전 프레임 (4020) 으로서 사용되고, 또한 후속 처리를 위해 디인터레이서 (605) 외부로 통신된다.

[0072]

Wmed + MC 디인터레이싱 방식을 사용하여, 인트라 필드 보간으로부터 인터 필드 보간을 포함한 디인터레이싱 예측 방식을 분리하는 것이 가능하다. 즉, 공간-시간 Wmed 필터링은 주로 인트라 필드 보간 목적으로 사용될 수 있지만, 인터 필드 보간은 모션 보상 동안에 수행될 수 있다. 이는, Wmed 결과의 피크 신호 대 잡음비를 줄이지만, 정확하지 않은 인터 필드 예측 모드 결정으로부터의 불량 화소가 Wmed 필터링 프로세스에서 제거될 것이므로, 모션 보상 후의 시각 품질은 더 만족스럽게 적용된다.

[0073]

적절한 역 텔레시네 또는 디인터레이싱 처리 후에, 블록 (608) 에서, 순차 비디오는 에일리어스 억제 및 리샘플링 (예를 들어, 리사이징) 을 위해 처리된다. 몇몇 리샘플링 양태에서, 다상 리샘플러는 영상 크기 리사이징을 위해 구현된다. 다운샘플링의 일 예에서, 원래 영상과 리사이징된 영상 간의 비율은 p/q 일 수 있고, 여기서, p 와 q 는 비교적 프라임 정수이다. 총 위상 수는 p 이다. 몇몇 양태에서 다상 필터의 차단 주파수는 약 0.5 의 리사이징 계수에 대해 0.6 이다. 차단 주파수는 리사이징된 시퀀스의 고주파수 응답을 향상시키기 위해 리사이징 비율과 정확히 일치하지 않아도 된다. 이로 인해, 반드시 약간의 에일리어싱을 허용하게 된다. 그러나, 인간의 눈이 흐릿하며 에일리어싱 없는 영상보다는 뚜렷하지만 덜 에일리어싱된 영상을 선호한다는 것이 널리 공지되어 있다.

[0074]

도 41 은 다상 리샘플링의 일 예를 도시하는 것으로서, 리사이징 비율이 $3/4$ 인 경우의 위상을 도시한다. 또한, 도 41 에 도시된 차단 주파수는 $3/4$ 이다. 도 41 에서 원래 화소는 수직축에 대해 도시되어 있다. 또한, sinc 함수는 그 수직축을 중심으로 도시되어 필터 파형을 나타낸다. 차단 주파수가 리샘플링 비율과 정확히 동일하게 선택되었기 때문에, sinc 함수의 영점은 리사이징 후의 화소 위치와 겹치고, 도 41 에는 교차하는 것으로 도시되어 있다. 리사이징 후의 화소 값을 찾기 위해, 그 기여는 다음 식에서 표현된 원래 화소로부터 합계될 수 있다.

수학식 6

$$v(x) = \sum_{i=-\infty}^{\infty} u(i) \times \sin c(\pi f_c(i-x))$$

[0075]

[0076]

[0077]

[0078]

[0079]

[0080]

[0081]

[0082]

여기서, f_c 는 차단 주파수이다. 상기 1-D 다상 필터는 수평 차원과 수직 차원 모두에 적용될 수 있다.

리샘플링 (리사이징) 의 다른 양태는 오버스캔을 설명한다. NTSC 텔레비전 신호에 있어서, 이미지는 486 개의 주사선을 갖고, 디지털 비디오에서 각 주사선 상에 720 개의 화소를 갖는다. 그러나, 사이즈와 화면 포맷 간의 불일치 때문에 텔레비전에서 전체 이미지 모두를 볼 수 있는 것은 아니다. 볼 수 없는 이미지 부분은 오버스캔으로 지칭된다.

방송사업자가 가능한 한 많은 텔레비전이 볼 수 있는 영역에 유용한 정보를 두는 것을 돕기 위해, SMPTE (Society of Motion Picture & Television Engineers) 는 안전한 액션 영역과 안전한 타이틀 영역으로 지칭되는 액션 프레임의 특정 크기를 정의하였다. *Specifications for Safe Action and Safe Title Areas Test Pattern for Television Systems* 에 대한 SMPTE 권고 실무 RP 27.3-1989 년 참조. 안전 액션 영역은 "모든 중요한 액션이 일어나야 하는" 영역으로서 SMPTE 에 의해 정의된다. 안전 타이틀 영역은 "가정용 텔레비전 수상기 대부분에서 가시성을 확보하도록 모든 유용한 정보가 제한될 수 있는" 영역으로서 정의된다.

예를 들어, 도 25 를 참조하면, 안전 액션 영역 (2510) 은 화면의 중심 90% 을 차지하고, 모든 주위에서 5 % 경계를 준다. 안전 타이틀 영역 (2505) 은 화면의 중심 80% 를 차지하고, 10% 경계를 준다. 이하, 도 26 을 참조하면, 안전 타이틀 영역이 너무 작기 때문에, 이미지에 더 많은 콘텐츠를 추가하기 위해, 몇몇 지국은 백색 장방형 윈도우 (2615) 내에 있는 안전 액션 영역 안으로 텍스트를 밀어 넣을 것이다.

통상, 흑색 경계는 오버스캔에서 보일 수도 있다. 예를 들어, 도 26 에서, 흑색 경계는 이미지의 상부 측 (2620) 과 하부 측 (2625) 에 나타난다. 이들 흑색 경계는 오버스캔에서 제거될 수 있는데, 그 이유는, H.264 비디오가 모션 추정에서 경계 확장을 사용하기 때문이다. 확장된 흑색 경계는 나머지를 증가시킨다. 대략, 그 경계가 2% 만큼 줄어들 수 있고, 그 다음에, 리사이징을 행할 수 있다. 그에 따라, 리사이징을 위한 필터를 생성할 수 있다. 절단 (truncation) 을 행하여 다상 다운샘플링 전에 오버스캔을 제거할 수 있다.

다시 도 6 을 참조하면, 그 다음에, 순차 비디오는 디블록킹 및 디링잉 동작을 수행하는 블록 (610) 으로 진행된다. 2 가지 타입의 아티팩트인 "블록킹" 과 "링잉" 은 비디오 압축 애플리케이션에서 보통 일어난다. 블록킹 아티팩트는, 압축 알고리즘이 각 프레임을 블록 (예를 들어, 8×8 개의 블록) 으로 분할하기 때문에, 일어난다. 각 블록은 얼마간의 작은 에러를 갖도록 재구성되고, 블록 에지에서의 에러는 인접 블록 에지에서의 에러와 종종 대조를 이루어 블록 경계를 볼 수 있게 한다. 이와는 달리, 링잉 아티팩트는, 이미지 특징의 에지 주위의 왜곡으로서 나타난다. 링잉 아티팩트는, 인코더가 고주파수 DCT 계수를 양자화하는데 있어서 너무 많은 정보를 버리기 때문에, 일어난다. 몇몇 예시적인 예에서, 디블록킹과 디링잉 모두는 저역 통과 FIR (finite impulse response) 필터를 사용하여 이들 보이는 아티팩트를 숨길 수 있다.

디블록킹 처리의 일 예에서, 디블록킹 필터는, 프레임 경계에 있는 에지와 디블록킹 필터 프로세스가 디세이블되는 임의의 에지를 제외하고는, 모든 4×4 개의 블록 에지에 적용될 수 있다. 이러한 필터링 프로세스는 매크로블록 주소가 증가하는 순서로 처리되는 프레임 내의 모든 매크로블록과 함께 프레임 구성 프로세스의 완료 후에 매크로블록 기반으로 수행될 것이다. 각각의 매크로블록마다, 우선, 수직 에지가 좌측에서 우측으로 필터링된 다음에, 수평 에지가 상부에서 하부로 필터링된다. 도 39 에 도시된 바와 같이, 루마 디블록킹 필터 프로세스는 4개의 16-샘플 에지에 대해 수행되고, 각각의 크로마 컴포넌트에 대한 디블록킹 필터 프로세스는 수평 방향과 수직 방향에 대해 2 개의 8-샘플 에지에 대해 수행된다. 이전 매크로블록에 대한 디블록킹 프로세스 동작에 의해 이미 수정되었을 수도 있는 현재 매크로블록의 위와 좌측에 있는 샘플 값은 현재 매크로블록에 대한 디블록킹 필터 프로세스에 대한 입력으로서 사용될 것이고, 또한 현재 매크로블록의 필터링 동안에 수정될 수도 있다. 수직 에지의 필터링 동안에 수정된 샘플 값은 동일 매크로블록에 대한 수평 에지의 필터링을 위한 입력으로서 사용될 수 있다. 디블록킹 프로세스는 루마 및 크로마 컴포넌트에 대해 개별적으로 호출될 수 있다.

- [0083] *디링잉 프로세스의 일 예에서, 2-D 필터는 에지 근방 영역을 평활화하는데 적응 가능하게 적용될 수 있다. 에지 화소는 블러링을 피하기 위해 필터링을 거의 하지 않거나 전혀 하지 않는다.
- [0084] GOP 분할기
- [0085] 디블록킹과 디링잉 후에, 순차 비디오는 GOP 분할기 (612) 에 의해 처리된다. GOP 포지셔닝은 샷 변화 검출, 복잡도 맵 생성 (예를 들어, 시간, 공간 대역폭 맵) 및 적응 GOP 분할을 포함할 수 있다. 이들은 아래에 각각 설명되어 있다.
- [0086] A. 장면 전환 검출
- [0087] 샷 검출은, GOP (group of picture) 내의 프레임이 장면 전환이 일어난 데이터를 나타내는 때를 결정하는 것에 관한 것이다. 일반적으로, GOP 내에서, 프레임은 임의의 2 개 또는 3 개의 (또는 그 이상) 인접 프레임에서 상당한 변화를 갖지 않거나 느린 변화 또는 빠른 변화일 수도 있다. 물론, 이들 장면 전환 분류는 또한 필요한 경우 특정 애플리케이션에 따라 더 큰 변화 레벨로 분류될 수 있다.
- [0088] 샷 또는 장면 전환 검출은 효율적인 비디오 인코딩에 있어서 중요하다. 통상, GOP 가 상당히 변하지 않고 있는 경우에는, 다수의 예측 프레임이 비디오를 충분히 인코딩한 다음에, GOP 의 시작에서 I-프레임이 일어나므로, 후속 디코딩 및 비디오 디스플레이는 시각적으로 허용 가능하다. 그러나, 장면이 갑작스럽게 또는 느리게 변하고 있을 때에는, 그 다음에 디코딩된 시각적으로 허용 가능한 결과를 생성하기 위해 추가적인 I-프레임과 덜 예측적인 인코딩 (P-프레임과 B-프레임) 을 필요로 할 수도 있다.
- [0089] 이하, 기존 인코딩 시스템의 성능을 개선하는 샷 검출과 인코딩 시스템 및 방법을 설명한다. 그러한 양태는 전처리 (226) 의 GOP 분할기 (612) 에 구현될 수 있고, 또는 전처리 유무에 관계없이 동작할 수도 있는 인코더 디바이스에 포함될 수 있다 (도 7). 그러한 양태는, 갑작스런 장면 전환이 일어났는지, 장면이 느리게 변하고 있는지, 또는 비디오 인코딩을 특히 복잡하게 할 수 있는 장면 내의 카메라 플래시라이트가 존재하는지를 결정하기 위해, 비디오 데이터의 인접 프레임 간의 통계적 비교를 포함한 통계치 (또는 메트릭) 를 사용한다. 그 통계치는 전처리로부터 얻어질 수 있고, 그 다음에, 인코딩 디바이스로 송신되거나, (예를 들어, 모션 보상을 수행하도록 구성된 프로세서에 의해) 인코딩 디바이스에서 생성될 수 있다. 결과로서 생성된 통계치는 장면 전환 검출 결정을 돕는다. 트랜스코딩을 행하는 시스템에서는, 종종 적당한 전처리 또는 구성 가능한 프로세서가 존재한다. 전처리가 모션 보상 활용 디인터레이싱을 수행하는 경우에, 모션 보상 통계치를 사용할 수 있고, 즉시 사용 가능하다. 그러한 시스템에서, 샷 검출 알고리즘은 시스템 복잡도를 약간 증가시킬 수도 있다.
- [0090] 본원에 설명된 샷 검출기의 예시적인 예는 이전 프레임, 현재 프레임 및 다음 프레임으로부터의 통계치를 사용하기만 하면 되므로, 매우 낮은 지연시간을 갖는다. 샷 검출기는 갑작스러운 장면 전환, 크로스페이딩 (cross-fading) 과 기타 느린 장면 전환, 및 카메라 플래시라이트를 비롯한 몇몇 상이한 타입의 샷 이벤트를 구별한다. 인코더에서 상이한 전략으로 상이한 타입의 샷 이벤트를 결정함으로써, 인코딩 효율과 시각 품질이 향상된다.
- [0091] 장면 전환 검출은 고정 간격으로 I-프레임을 삽입함으로써 지능적으로 비트를 보전하기 위한 임의의 비디오 코딩 시스템에 사용될 수 있다. 몇몇 양태에서, (예를 들어, 메타데이터에 통합되거나 전처리 (226) 에 의해 계산되는) 전처리에 의해 얻어진 콘텐츠 정보는 장면 전환 검출에 사용될 수 있다. 예를 들어, 콘텐츠 정보에 따라, 아래에 설명되는 임계값 및 기타 기준은 상이한 타입의 비디오 콘텐츠에 대해 동적으로 조절될 수도 있다.
- [0092] 통상, 비디오 인코딩은 구조화된 GOP (group of picture) 에 대해 작용한다. 통상, GOP 는 인트라 코딩된 프레임 (I-프레임) 에서 시작하고, 일련의 P (예측) 또는 B (양방향) 프레임이 뒤따른다. 통상, I-프레임은 프레임을 디스플레이하는데 필요한 모든 데이터를 저장할 수 있고, B-프레임은 이전 및 다음 프레임의 데이터에 의존하며 (예를 들어, 단지 이전 프레임으로부터 변경된 데이터를 포함하거나, 다음 프레임의 데이터와 상이함), P-프레임은 이전 프레임으로부터 변경된 데이터를 포함한다. 통상의 사용 시, I-프레임은 인코딩된 비디오에서 P-프레임과 B-프레임이 산재된다. 크기 (예를 들어, 프레임을 인코딩하는데 사용되는 비트 수) 면에서, 통상, I-프레임은 P-프레임보다 훨씬 더 크고, 또한, P-프레임은 B-프레임보다 크다. 효율적인 인코딩, 송신 및 디코딩 처리를 위해, GOP 의 길이는 큰 I-프레임으로부터의 유효 손실을 줄이기에 충분히 길어야

하고, 인코더와 디코더 간의 부정합, 또는 채널 손상을 해결하기에 충분히 짧아야 한다. 또한, P 프레임 내의 매크로블록 (MB) 은 동일한 이유로 인트라 코딩될 수 있다.

[0093] 종종 불필요한 I-프레임을 고정 간격으로 삽입하는 대신에, 비디오 인코더에 대해 장면 전환 검출을 사용하여, GOP 길이에 기초하여 적절한 GOP 길이를 결정하고, I-프레임을 삽입할 수 있다. 통상, 실제 스트리밍 비디오 스트림에서, 통신 채널은 비트 에러나 패킷 손실에 의해 손상된다. I 프레임 또는 I MB 를 어느 곳에 배치해야 하는지는 디코딩된 시각 품질과 시청 경험에 상당히 영향을 줄 수도 있다. 하나의 인코딩 방식은 나란히 배열된 이전 영상 또는 영상 부분으로부터 상당한 변화를 가질 수도 있는 영상 또는 영상 부분에 대한 인트라 코딩된 프레임을 사용하는 것이다. 통상, 이들 영역은 모션 추정을 사용하여 효과적·효율적으로 예측될 수 없고, 그러한 영역이 인트라 프레임 코딩 기술 (예를 들어, B-프레임과 P-프레임을 사용하여 인코딩하는 기술) 에서 면제된다면, 인코딩은 더 효율적으로 행해질 수 있다. 채널 손상의 문맥에서, 이들 영역은 에러 전파의 영향을 받을 가능성이 있는데, 이러한 에러 전파는 인트라 프레임 인코딩에 의해 감소 또는 제거 (또는 거의 제거됨) 될 수 있다.

[0094] GOP 비디오의 일부는 2 개 이상의 카테고리로 분류될 수 있고, 여기서, 각 영역은 특정 구현에 의존할 수도 있는 상이한 인트라 프레임 인코딩 기준을 가질 수 있다. 일 예로서, 비디오는 3 개의 카테고리로 분류될 수 있다: 갑작스런 장면 전환, 크로스페이딩과 기타 느린 장면 전환, 및 카메라 플래시라이트.

[0095] 갑작스런 장면 전환은 통상, 카메라 동작에 의해 발생하는 이전 프레임과 상당히 상이한 프레임을 포함한다. 이들 프레임의 콘텐츠가 이전 프레임의 콘텐츠와 상이하기 때문에, 갑작스런 장면 전환은 I 프레임으로서 인코딩되어야 한다.

[0096] 크로스페이딩과 기타 느린 장면 전환은, 통상 카메라 샷의 컴퓨터 처리에 의해 발생하는 느린 장면 전환을 포함한다. 2 개의 상이한 장면의 점진적인 블렌딩은 사람 눈에 더 즐겁게 보일 수도 있지만, 비디오 코딩하는데 있어서 어려움을 제공한다. 모션 보상은 이들 프레임의 비트레이트를 효과적으로 줄일 수 있고, 이들 프레임에 대해 더 많은 인트라 MB 가 갱신될 수 있다.

[0097] 카메라 플래시라이트 또는 카메라 플래시 이벤트는, 프레임 콘텐츠가 카메라 플래시를 포함하는 경우에 발생한다. 그러한 플래시는 지속 기간이 비교적 짧고 (예를 들어, 일 프레임), 매우 밝아서, 플래시를 표현하는 프레임 내의 화소는 인접 프레임 상의 대응하는 영역에 비해 상당히 높은 휘도를 나타낸다. 카메라 플래시라이트는 영상의 휘도를 갑자기 또한 신속히 시프트한다. 통상, 카메라 플래시라이트의 지속 기간은 보통 44 ms 로 규정되어 있는 HVS (human vision system) 의 시간 마스킹 지속 기간보다 짧다. 사람 눈은 이들 짧은 밝기 버스트의 품질에 민감하지 않으므로, 덜 세분화되어 인코딩될 수 있다. 플래시라이트 프레임이 모션 보상을 사용하여 효과적으로 처리될 수 없고 추후 프레임에 대한 좋지 않은 예측 후보이기 때문에, 이들 프레임의 덜 세분화된 인코딩은 추후 프레임의 인코딩 효율을 줄이지 않는다. 플래시라이트로서 분류된 장면은 "인위적인" 고휘도 때문에 다른 프레임을 예측하는데 사용되지 않아야 하고, 같은 이유로 다른 프레임은 이들 프레임을 예측하는데 효과적으로 사용될 수 없다. 일단 식별되면, 이들 프레임이 비교적 높은 처리량을 필요로 할 수도 있기 때문에, 이들 프레임을 제거할 수 있다. 하나의 옵션은 카메라 플래시라이트 프레임을 제거하고, 그 대신에 DC 계수를 인코딩하는 것이다: 그러한 솔루션은 간단하고, 계산상 빠르고 많은 비트를 절약한다.

[0098] 상기 내용 중 임의의 내용이 프레임이 검출되는 것으로 분류하면, 샷 이벤트가 선언된다. 샷 검출은 인코딩 품질을 개선하는데 유용할 뿐만 아니라, 비디오 콘텐츠 검색 및 인덱싱을 식별하는 것을 도울 수도 있다. 이하, 장면 검출 프로세스의 일 예시적인 양태가 설명된다. 본 예에 따르면, 우선, 샷 검출 프로세스는 샷 검출을 위해 처리중인 선택된 프레임에 대한 정보, 또는 메트릭을 계산한다. 메트릭은 비디오의 양방향 모션 추정 및 보상으로부터의 정보, 및 기타 휘도 기반 메트릭을 포함할 수 있다.

[0099] 양방향 모션 추정/보상을 수행하기 위해, 비디오 시퀀스는, 프레임들 중 2 개의 프레임, 즉, 프레임들 중 가장 인접한 이웃 프레임인, 과거의 한 프레임 및 미래의 한 프레임의 블록과 현재 프레임의 모든 8×8 블록을 정합시키는 양방향 모션 보상기를 사용하여 전처리될 수 있다. 모션 보상기는 모든 블록마다 모션 벡터와 차이 메트릭을 생성한다. 도 29 는 과거 프레임 (P) 및 미래 (또는 다음) 프레임 (N) 에 현재 프레임 (C) 의 화소를 정합시키는 일 예를 도시한 도면으로서, 정합된 화소에 대한 모션 벡터 (과거 모션 벡터 (MV_P) 및 미래 모션 벡터 (MV_N)) 를 도시한다. 통상, 양방향 모션 벡터 생성 및 관련 인코딩의 일반적인 설명은 도 32 를 참조하여 아래에 설명된다.

[0100] 양방향 모션 정보 (예를 들어, 대응하는 인접 프레임에서 (가장 정합되는) MB 를 식별하는 모션 정보) 를 결정 한 후에, 다음 프레임 및 이전 프레임과 현재 프레임을 다양하게 비교함으로써, (예를 들어, GOP 분할기 (612) 내의 모션 보상기 또는 다른 적당한 컴포넌트에 의해) 추가 메트릭을 생성할 수 있다. 모션 보상기는 모든 블록마다 차이 메트릭을 생성할 수 있다. 차이 메트릭은 SSD (sum of square difference) 또는 SAD (sum of absolute difference) 일 수 있다. 일반성을 잃지 않으면서, 본원에서 SAD 는 일 예로서 사용된다.

[0101] 모든 프레임에 대해, "콘트라스트비" 로도 지칭되는 SAD 비는 다음 수학적식과 같이 계산된다.

[0102] [수학적식 6a]

$$\gamma = \frac{\varepsilon + SAD_P}{\varepsilon + SAD_N}$$

[0103]

[0104] 여기서, SAD_P 및 SAD_N 은 각각 순방향 및 역방향 차이 메트릭의 절대 차이의 합이다. 분모는 "divide-by-zero" 에러를 방지하기 위해 작은 양의 수 ε 을 포함한다는 것에 주목해야 한다. 또한, 분자는 분모에 있는 "1" 의 영향과 균형을 맞추기 위해 ε 을 포함한다. 예를 들어, 이전 프레임, 현재 프레임 및 다음 프레임이 동일하면, 모션 검색은 $SAD_P = SAD_N = 0$ 이라는 결과를 나타낼 것이다. 이러한 경우에, 상기 계산은 "0" 또는 무한대 대신에 $\gamma = 1$ 이라는 결과를 생성한다.

[0105] 휘도 히스토그램은 모든 프레임에 대해 계산될 수 있다. 통상, 멀티미디어 이미지는 8 비트의 휘도 깊이 (예를 들어, "빈 (bin)" 의 개수) 를 갖는다. 몇몇 양태에 따른 휘도 히스토그램을 계산하는데 사용되는 휘도 깊이는 그 히스토그램을 얻기 위해 16 으로 설정될 수 있다. 다른 양태에서, 휘도 깊이는 처리되는 데이터 타입, 가용 계산 전력 또는 기타 소정의 기준에 따라 수도 있는 적절한 개수로 설정될 수 있다. 몇몇 양태에서, 휘도 깊이는 데이터 콘텐츠와 같은 계산되거나 수신된 메트릭에 기초하여 동적으로 설정될 수 있다.

[0106] 아래 수학적식은 휘도 히스토그램 차이 (람다) 를 계산하는 일 예를 도시한다:

수학적식 7

$$\lambda = \frac{\sum_{i=1}^{16} |N_{Pi} - N_{Ci}|}{N}$$

[0107]

[0108] 여기서, N_{Pi} 는 이전 프레임에 대한 i 번째 빈 내의 블록 개수이고, N_{Ci} 는 현재 프레임에 대한 i 번째 빈 내의 블록 개수이며, N 은 일 프레임 내의 총 블록 개수이다. 이전 및 현재 프레임의 휘도 히스토그램 차이가 완전히 다르면 (또는 공통 원소를 갖지 않으면), $\lambda = 2$ 이다.

[0109] 이러한 정보를 사용하여, 프레임 차이 메트릭 (D) 은 다음 식과 같이 계산된다:

수학적식 8

$$D = \frac{\gamma_C}{\gamma_P} + A\lambda(2\lambda + 1)$$

[0110]

[0111] 여기서, A 는 애플리케이션에 따라 선택된 상수이고, $\gamma_C = \frac{\varepsilon + SAD_P}{\varepsilon + SAD_N}$, 및 $\gamma_P = \frac{\varepsilon + SAD_{PP}}{\varepsilon + SAD_C}$ 이다.

[0112] 프레임 차이 메트릭이 수학적식 9 에 도시된 기준을 만족하는 경우에, 선택된 (현재) 프레임은 갑작스런 장면 전환 프레임으로서 분류된다.

수학식 9

$$D = \frac{\gamma_C}{\gamma_P} + A\lambda(2\lambda + 1) \geq T_1$$

[0113]

[0114] 여기서, A 는 애플리케이션에 따라 선택된 상수이고, T_1 은 임계값이다.

[0115] 시뮬레이션이 나타내는 일 예에서, $A = 1$ 및 $T_1 = 5$ 로 설정하면, 양호한 검출 성능이 달성된다. 현재 프레임이 갑작스런 장면 전환 프레임이면, γ_C 는 커야하고, γ_P 는 작아야 한다. 메트릭이 문맥의 액티비티 레

벨로 정규화되도록, 단지 γ_C 만을 사용하는 대신에 비율 $\frac{\gamma_C}{\gamma_P}$ 를 사용할 수 있다.

[0116] 상기 기준은 휘도 히스토그램 차이 람다 (λ) 를 비선형으로 사용한다는 것에 주목하자. 도 16 은 $\lambda * (2\lambda + 1)$ 이 볼록 함수인 것을 도시한다. λ 가 작으면 (예를 들어, 0 에 가까우면), 프리엠퍼시스를 거의 갖지 않는다. λ 가 커지면, 볼록 함수에 의해 더 많은 엠퍼시스가 수행된다. 이러한 프리엠퍼시스를 사용하여, 1.4 보다 큰 임의의 λ 에 대해, 임계값 T_1 이 5 로 설정되면, 갑작스런 장면 전환이 검출된다.

[0117] 일정 수의 연속 프레임에 대해, 장면 세기 메트릭 D 가 수학식 10 에 표현된 기준을 만족하면, 현재 프레임이 크로스페이딩 또는 느린 장면 전환인 것으로 결정된다.

수학식 10

$$T_2 \leq D < T_1$$

[0118]

[0119] 여기서, T_1 은 위에서 사용된 동일 임계값이고, T_2 는 다른 임계값이다.

[0120] 통상, 플래시라이트 이벤트로 인해, 휘도 히스토그램이 더 밝은 쪽으로 이동하게 된다. 이러한 예시적인 양태의 카메라에서는, 휘도 히스토그램 통계치를 사용하여, 현재 프레임이 카메라 플래시라이트를 포함하는지를 결정한다. 수학식 11 및 수학식 12 에 나타난 바와 같이, 샷 검출 프로세스는, 현재 프레임 마이너스의 휘도가 이전 프레임의 휘도보다 일정 임계값 T_3 만큼 큰지를 결정할 수 있고, 현재 프레임의 휘도가 다음 프레임의 휘도보다 임계값 T_3 만큼 큰지를 결정할 수 있다.

수학식 11

$$\bar{Y}_C - \bar{Y}_P \geq T_3$$

[0121]

수학식 12

$$\bar{Y}_C - \bar{Y}_N \geq T_3$$

[0122]

[0123] 상기 기준을 만족하지 않으면, 현재 프레임은 카메라 플래시라이트를 포함한 것으로서 분류되지 않는다. 상기 기준을 만족하면, 아래 수학식에 표현된 바와 같이, 샷 검출 프로세스는 역방향 차이 메트릭 SAD_b 와 순방향 차이 메트릭 SAD_f 이 일정 임계값 T_4 보다 큰지를 결정한다.

수학식 13

$$SAD_P \geq T_4$$

[0124]

수학식 14

$$SAD_N \geq T_4$$

[0125]

[0126] 여기서, \bar{Y}_C 는 현재 프레임의 평균 휘도이고, \bar{Y}_P 는 이전 프레임의 평균 휘도이고, \bar{Y}_N 은 다음 프레임의 평균 휘도이며, SAD_P 및 SAD_N 은 현재 프레임과 연관된 순방향 및 역방향 차이 메트릭이다.

[0127] 샷 검출 프로세스는, 우선 현재 프레임의 휘도가 이전 프레임의 휘도와 다음 프레임의 휘도보다 큰지를 결정함으로써 카메라 플래시 이벤트를 결정한다. 현재 프레임의 휘도가 이전 프레임의 휘도와 다음 프레임의 휘도보다 크지 않으면, 그 프레임은 카메라 플래시 이벤트가 아니고, 현재 프레임의 휘도가 이전 프레임의 휘도와 다음 프레임의 휘도보다 크면, 카메라 플래시 이벤트일 수도 있다. 그 다음에, 샷 검출 프로세스는 역방향 차이 메트릭이 임계값 T_3 보다 큰지 여부와, 순방향 차이 메트릭이 임계값 T_4 보다 큰지 여부를 평가할 수 있고, 이들 양 조건을 만족하면, 샷 검출 프로세스는 카메라 플래시라이트를 갖는 것으로서 현재 프레임을 분류한다. 상기 기준을 만족하지 않으면, 그 프레임은 어떤 타입의 샷 이벤트로 분류되지 않고, 또는 그 프레임에 대해 수행되는 인코딩을 식별하는 디폴트 분류 (예를 들어, 드롭 프레임, I-프레임으로 인코딩됨) 가 주어질 수 있다.

[0128] 상기와 같이, T_1 , T_2 , T_3 및 T_4 에 대한 몇몇 예시적인 값이 존재한다. 통상, 이들 임계값은 샷 검출의 특정 구현의 테스트를 통해 선택된다. 몇몇 양태에서, 임계값 T_1 , T_2 , T_3 및 T_4 중 하나 이상은 미리 정해지고, 그러한 값은 인코딩 디바이스에서 샷 분류기 내에 통합된다. 몇몇 양태에서, 임계값 T_1 , T_2 , T_3 및 T_4 중 하나 이상은 샷 분류기에 공급된 정보 (예를 들어, 메타데이터) 의 사용에 기초하여 또는 샷 분류기 자체에 의해 계산된 정보에 기초하여 처리 동안에 (예를 들어, 동적으로) 설정될 수 있다.

[0129] 통상, 샷 검출 정보를 사용한 비디오 인코딩은 인코더에서 수행되지만, 본원에서는 샷 검출 개시내용의 완전성을 위해 설명된다. 도 30 을 참조하면, 인코딩 프로세스 (301) 는 샷 검출 정보를 사용하여, 프레임 시퀀스에서 검출된 샷에 기초하여 비디오를 인코딩할 수 있다. 프로세스 (301) 는 블록 (303) 으로 진행하고, 현재 프레임이 갑작스런 장면 전환으로서 분류되는지를 알기 위해 검사한다. 현재 프레임이 갑작스런 장면 전환으로서 분류되면, 블록 (305) 에서, 현재 프레임을 I-프레임으로서 분류할 수 있고, GOP 경계를 결정할 수 있다. 현재 프레임이 갑작스런 장면 전환으로서 분류되지 않으면, 프로세스 (301) 는 블록 (307) 으로 진행하고, 즉, 블록 (309) 에서 현재 프레임이 느리게 변하는 장면의 일부로서 분류되면, 현재 프레임과 느리게 변하는 장면 내의 다른 프레임을 예측 프레임 (예를 들어, P-프레임 또는 B-프레임) 으로서 인코딩할 수 있다. 그 다음에, 프로세스 (301) 는 블록 (311) 으로 진행하여, 현재 프레임이 카메라 플래시를 포함한 플래시라이트 장면으로서 분류되는지를 검사한다. 현재 프레임이 플래시라이트 장면으로서 분류되면, 블록 (313) 에서, 현재 프레임은 예를 들어, 그 프레임에 대한 DC 계수 제거 또는 인코딩과 같은 특별한 처리에 대해 식별될 수 있고, 현재 프레임이 플래시라이트 장면으로서 분류되지 않으면, 현재 프레임의 어떤 분류도 이루어지지 않고, 현재 프레임을 다른 기준에 따라 인코딩할 수 있는데, I-프레임으로 인코딩하거나 드롭할 수 있다.

[0130] 상술한 양태에서, 압축될 프레임과 그 인접한 2 개의 프레임 간의 차이 양은 프레임 차이 메트릭 D 에 의해 표시된다. 상당한 양의 단방향 휘도 변화가 검출되면, 이는 프레임에서 크로스페이딩 효과를 나타낸다. 크로스페이딩이 더 현저해지면, B-프레임을 사용함으로써 더 많은 이득을 달성할 수 있다. 몇몇 양태에서, 수정된 프레임 차이 메트릭은 아래 수학식에 표현된 것과 같이 사용된다.

수학식 15

$$D_1 = \begin{cases} \left(1 - \alpha + 2\alpha \frac{|d_P - d_N|}{d_P + d_N}\right) \times D, & \text{if } Y_P - \Delta \geq Y_C \geq Y_N + \Delta \text{ or } Y_P + \Delta \leq Y_C \leq Y_N - \Delta, \\ D, & \text{otherwise,} \end{cases}$$

여기서, $d_P = |Y_C - Y_P|$ 및 $d_N = |Y_C - Y_N|$ 은 각각 현재 프레임과 이전 프레임 간의 루마 차이 및 현재 프레임과 다음 프레임 간의 루마 차이이고, Δ 는 구현에 따라 정규 실험에서 결정될 수 있는 상수를 나타내며, α 는 0 과 1 사이의 값을 갖는 가중 변수이다.

B. 대역폭 맵 생성

또한, 전처리 (226; 도 6) 는 멀티미디어 데이터를 인코딩하는데 사용될 수 있는 대역폭 맵을 생성하도록 구성될 수 있다. 몇몇 양태에서는, 그 대신에, 인코더 (228; 도 7) 내의 콘텐츠 분류 모듈 (712) 이 대역폭 맵을 생성한다.

인간 시각 품질 V 는 인코딩 복잡도 C 와 할당된 비트 B (또한, 대역폭으로도 지칭됨) 모두의 함수일 수 있다. 도 15 는 이러한 관계를 도시한 그래프이다. 인코딩 복잡도 메트릭 C 는 인간 시각 관점으로 공간 및 시간 주파수를 고려한다. 사람 눈에 더 민감한 왜곡의 경우에는, 이에 대응하여 복잡도 값이 높아진다. 통상, V 는 C 에 따라 단조 감소하고 있고, B 에 따라 단조 증가하는 것으로 가정할 수 있다.

일정한 시각 품질을 달성하기 위해, 인코딩될 i 번째 물체 (프레임 또는 MB) 에는 바로 아래 2 개의 수학식에 표현된 기준을 만족하는 대역폭 B_i 가 할당된다.

수학식 16

$$B_i = B(C_i, V)$$

수학식 17

$$B = \sum_i B_i$$

바로 위에 있는 2 개의 수학식에서, C_i 는 i 번째 물체의 인코딩 복잡도이고, B 는 전체 가용 대역폭이며, V 는 일 물체에 대해 달성된 시각 품질이다. 인간 시각 품질은 일 수학식으로 공식화하는 것이 어렵다. 따라서, 상기 수학식 세트는 정확히 정해지지 않는다. 그러나, 3-D 모델이 모든 변수에서 연속적인 것으로 가정

하면, 대역폭 비율 (B_i/B) 은 (C, V) 쌍 근방 내에서 변하지 않는 것으로서 취급될 수 있다. 대역폭 비율 β_i 는 아래 표현된 수학식에서 정해진다.

수학식 18

$$\beta_i = B_i/B$$

[0141] 그 다음에, 비트 할당은 다음 수학식에서 표현된 것과 같이 정해질 수 있다:

수학식 19

$$\beta_i = \beta(C_i)$$

$$1 = \sum_i \beta_i \quad \text{for } (C_i, V) \in \delta(C_0, V_0)$$

[0142]

[0143] 여기서, δ 는 "근방" 을 나타낸다.

[0144] 인코딩 복잡도는 인간 시각 감도 (공간과 시간 모두) 에 의해 영향을 받는다. 지로드의 인간 시각 모델은 공간 복잡도를 규정하는데 사용될 수 있는 모델의 일 예이다. 이러한 모델은 로컬 공간 주파수와 주위 조명을 고려한다. 그 결과로서 생성된 메트릭은 D_{csat} 로 지칭된다. 프로세스 내의 전처리 시점에서는, 영상 이 인트라 코딩되는지 인터 코딩되는지 여부가 알려지지 않고, 양쪽에 대한 대역폭 비율이 생성된다. 비트 는 상이한 비디오 물체의 β_{INTRA} 간의 비율에 따라 할당된다. 인트라 코딩된 영상의 경우에, 대역폭 비율은 다음 수학식으로 표현된다.

수학식 20

$$\beta_{INTRA} = \beta_{0INTRA} \log_{10}(1 + \alpha_{INTRA} Y^2 D_{csat})$$

[0145]

[0146] 상기 수학식에서, Y 는 매크로블록의 평균 휘도 성분이고, α_{INTRA} 는 휘도 제곱과 그 다음에 오는 D_{csat} 항에 대한 가중 계수이며, β_{0INTRA} 는 $1 = \sum_i \beta_i$ 를 보장하는 정규화 계수이다. 예를 들어,

$\alpha_{INTRA} = 4$ 에 대한 값은 양호한 시각 품질을 달성한다. 콘텐츠 정보 (예를 들어, 콘텐츠 분류) 를 사용하여, 특정 비디오 콘텐츠에 대한 양호한 시각 품질 레벨에 대응하는 값으로 α_{INTRA} 를 설정할 수 있다. 일 예에서, 비디오 콘텐츠가 "토크 헤드 (talking head)" 뉴스 방송을 포함하면, 정보 이미지 또는 비디오의 표시 가능 부분이 오디오 부분보다 덜 중요한 것으로 간주될 수도 있기 때문에, 시각 품질 레벨을 더 낮게 설정할 수도 있고, 더 적은 비트를 할당하여 데이터를 인코딩할 수 있다. 다른 예에서, 비디오 콘텐츠가 스포츠 이벤트를 포함하면, 표시된 이미지가 시청자에게 더 중요할 수도 있기 때문에, 콘텐츠 정보를 사용하여, 높은 시각 품질 레벨에 대응하는 값으로 α_{INTRA} 를 설정할 수도 있으므로, 더 많은 비트를 할당하여 데이터를 인코딩할 수 있다.

[0147] 이러한 관계를 이해하기 위해서는, 인코딩 복잡도에 따라 대역폭을 대수적으로 할당한다는 것에 주목해야 한다.

휘도 제곱 항 Y^2 은, 크기가 더 큰 계수가 인코딩을 위해 더 많은 비트를 사용한다는 사실을 나타낸다. 로그가 음의 값을 얻는 것을 방지하기 위해, 괄호 내의 항에 1 을 더한다. 또한, 다른 밑을 갖는 로그를 사용할 수도 있다.

[0148] 시간 복잡도는 프레임 차이 메트릭의 측정에 의해 결정되고, 이는 절대 차이의 합 (SAD) 으로서 프레임 차이 메트릭과 함께 모션 양 (예를 들어, 모션 벡터) 을 고려하여 2 개의 연속 프레임 간의 차이를 측정한다.

[0149] 인터 코딩된 영상에 대한 비트 할당은 시간 복잡도뿐만 아니라 공간 복잡도도 고려할 수 있다. 이는 아래 수학식으로 표현된다:

수학식 21

$$\beta_{INTER} = \beta_{0INTER} \log_{10} \left(1 + \alpha_{INTER} \cdot SSD \cdot D_{csat} \exp \left(-\gamma \|MV_P + MV_N\|^2 \right) \right)$$

[0150]

[0151]

상기 수학식에서, MV_P 와 MV_N 은 현재 MB 에 대한 순방향 및 역방향 모션 벡터이다 (도 29 참조). 인트라

코딩된 대역폭 공식 내의 Y^2 은 제곱 차이의 합 (SSD) 에 의해 대체된다. 상기 수학식에서

$\|MV_P + MV_N\|^2$ 의 역할을 이해하기 위해서는, 인간 시각 시스템의 다음 특성에 주목해야 한다: 평탄하고

예측 가능한 모션 (작은 $\|MV_P + MV_N\|^2$) 을 경험한 영역은 주의를 끌고, 통상, 사람 눈이 쫓아갈 수 있으며, 정지 영역보다 임의의 더 많은 왜곡을 허용할 수 있다. 그러나, 빠르거나 예측할 수 없는 모션 (큰

$\|MV_P + MV_N\|^2$) 을 경험한 영역은 사람 눈이 쫓아갈 수 없고, 상당한 양자화를 허용할 수 있다. 실험

은 $\alpha_{INTER} = 1$, $\gamma = 0.001$ 이 양호한 시각 품질을 달성하는 것을 나타낸다.

[0152]

C. 적응 GOP 분할

[0153]

또한, 전처리 (226) 에 의해 수행될 수도 있는 처리의 다른 예시적인 예에서, 도 6 의 GOP 분할기 (612) 는 함께 코딩된 영상 그룹의 구성을 적응 변경할 수도 있고, MPEG2 를 사용한 일 예를 참조하여 설명된다. 몇몇 오래된 비디오 압축 표준 (예를 들어, MPEG2) 은 GOP 가 일정한 구조를 갖는 것을 필요로 하지 않지만, 일 구조가 부과될 수 있다. MPEG2 시퀀스는 항상 I 프레임으로 시작하는데, 즉, 이전 영상을 참조함 없이 인코딩된 프레임으로 시작한다. 통상, MPEG2 GOP 포맷은 I 프레임 다음에 오는 P 또는 예측 영상의 GOP 내의 간격을 고정함으로써 인코더에서 사전에 조정된다. P 프레임은 이전 I 또는 P 영상에서 부분적으로 예측된 영상이다. 시작 I 프레임과 그 다음에 오는 P 프레임 간의 프레임은 B-프레임으로서 인코딩된다. "B" 프레임 (B 는 양방향을 나타냄) 은 이전 및 다음 I 또는 P 영상을 개별적으로 또는 동시에 기준으로 사용할 수 있다. 평균적으로 I 프레임을 인코딩하는데 필요한 비트 수는 P 프레임을 인코딩하는데 필요한 비트 수를 초과한다; 이와 마찬가지로, 평균적으로 P 프레임을 인코딩하는데 필요한 비트 수는 B-프레임을 인코딩하는데 필요한 비트 수를 초과한다. 스킵된 프레임은, 만약 사용된다하더라도, 그 표현에 어떤 비트도 필요로 하지 않을 것이다.

[0154]

P 및 B-프레임 사용에 기초가 되는 개념인, 최근의 압축 알고리즘에서 비디오를 표현하는데 필요한 데이터 레이트를 줄이도록 프레임을 스킵하는 것은 시간 중복을 제거하는 것이다. 시간 중복이 높으면, 즉, 영상 간에 변화가 거의 없으면, 먼저 디코딩된 I 또는 P 영상이 다른 P 또는 B 영상을 디코딩하기 위한 기준으로서 나중에 사용되기 때문에, P, B 또는 스킵된 영상의 사용은 비디오 스트림을 효율적으로 표현한다.

[0155]

적응 GOP 분할은 이러한 개념을 적응적으로 사용하는 것에 기초한다. 프레임 간의 차이는 양자화되고, 양자화된 차이에 대해 적당한 테스트를 수행한 후에 I, P, B 또는 스킵된 프레임에 의해 영상을 표현하려는 결정이 자동으로 이루어진다. 적응 구조는 고정된 GOP 구조에서는 이용 가능하지 않은 이점을 갖는다. 고정된 구조는 콘텐츠에서 변화가 거의 발생하지 않은 확률을 무시할 것이고, 적응 절차로 인해, 훨씬 많은 B-프레임이 각각의 I 와 P 사이에, 또는 2 개의 P 프레임 사이에 삽입되는 것을 허용함으로써, 프레임 시퀀스를 충분히 표현하는데 필요한 비트 수를 줄일 것이다. 이와 반대로, 비디오 콘텐츠의 변화가 아주 큰 경우에는, 예측 프레임과 기준 프레임 간의 차이가 너무 크기 때문에, P 프레임의 효율이 많이 감소할 것이다. 이들 조건 하에서는, 물체 정합 (matching objects) 이 모션 검색 영역 밖으로 나갈 수도 있고, 또는 물체 정합 간의 유사성이 카메라 앵글의 변화로 인한 왜곡 때문에 감소한다. 이때, P 프레임 또는 I 프레임 및 그 인접한 P 프레임은 서로 더 가까워지도록 선택되어야 하고, 더 적은 B-프레임을 삽입해야 한다. 고정된 GOP 는 그 조정을 행할 수 없다.

[0156]

본원에 개시된 시스템에서, 이들 조건은 자동으로 감지된다. GOP 구조는 유연하고, 콘텐츠에서의 이들 변화에 적응하도록 이루어진다. 본 발명의 시스템은 프레임 차이 메트릭을 평가하는데, 이는, 동일한 부가적인

거리 특성을 사용하여 프레임 간의 거리 측정으로서 간주될 수 있다. 개념상, 프레임 간 거리가 d_{12} 및 d_{23} 인 프레임 F_1 , F_2 및 F_3 이 주어지면, F_1 과 F_3 간의 거리는 적어도 $d_{12} + d_{23}$ 으로서 간주된다. 프레임 할당은 이러한 거리 유사 메트릭에 기초하여 이루어진다.

- [0157] GOP 분할기는 영상 타입을 수신되는 그대로 프레임에 할당함으로써 동작한다. 영상 타입은 각 블록을 코딩 하는데 필요할 수도 있는 예측 방법을 나타낸다.
- [0158] I-영상은 다른 영상을 참조함 없이 코딩된다. I-영상이 그 자체로 동작하기 때문에, 이는 디코딩을 시작할 수 있는 데이터 스트림 내의 액세스 포인트를 제공한다. 그 전 프레임까지의 "거리" 가 장면 전환 임계값을 초과하면, 프레임에 I 인코딩 타입이 할당된다.
- [0159] P-영상은 모션 보상된 예측에 대해 이전 I 또는 P 영상을 사용할 수 있다. P-영상은 인코딩 기준으로 예측되는 블록에서 변위될 수도 있는 이전 필드 또는 프레임 내의 블록을 사용한다. 기준 블록을 고려 중인 블록에서 제거한 후에, 통상, 공간 중복의 제거를 위한 이산 코사인 변환을 사용하여, 나머지 블록을 인코딩한다. 일 프레임과 P 프레임에 할당된 마지막 프레임 간의 "거리" 가 통상 제 1 임계값보다 낮은 제 2 임계값을 초과하면, 그 프레임에 P 인코딩 타입이 할당된다.
- [0160] 상술한 바와 같이, B-프레임 영상은 모션 보상을 위해 이전 및 다음 P- 또는 I-영상을 사용할 수 있다. B 영상 내의 블록은 순방향, 역방향 또는 양방향 예측될 수 있다; 또는 그 블록은 다른 프레임을 참조함 없이 인트라 코딩될 수 있다. H.264 에서, 기준 블록은 32 개 프레임으로부터의 32 개 블록의 선형 조합일 수 있다. 그 프레임을 I 또는 P 타입이 되도록 할당할 수 없는 경우에는, 그 프레임과 바로 이전 프레임 간의 "거리" 가 통상 제 2 임계값보다 낮은 제 3 임계값보다 크면, 그 프레임을 B 타입이 되도록 할당한다.
- [0161] 프레임을 인코딩되는 B-프레임이 되도록 할당할 수 없는 경우에는, "스킵 프레임" 상태를 할당한다. 그 프레임이 사실상 이전 프레임의 사본이기 때문에, 이러한 프레임을 스킵할 수 있다.
- [0162] 표시 순서로 인접 프레임 간의 거리를 양자화하는 메트릭의 평가는 이러한 처리에서 가장 먼저 일어나는 부분이다. 이러한 메트릭은 상기 참조되는 거리이다; 이를 사용하여, 모든 프레임은 그 적절한 타입에 대해 평가된다. 따라서, I 및 인접한 P, 또는 2 개의 연속적인 P 프레임 간의 간격이 변할 수 있다. 메트릭 계산은 블록 기반 모션 보상을 사용하여 비디오 프레임을 처리함으로써 시작하고, 일 블록은 통상 16×16 화소로 이루어진 비디오 압축의 기본 단위이지만, 8×8 , 4×4 및 8×16 과 같은 다른 블록 크기도 가능하다. 2 개의 디인터레이싱된 필드로 이루어진 프레임의 경우에, 모션 보상은 필드 단위로 행해질 수 있고, 기준 블록에 대한 검색은 프레임보다는 필드에서 일어나게 된다. 현재 프레임의 첫 번째 필드 내의 블록의 경우에, 순방향 기준 블록은 그 다음에 오는 프레임의 필드에서 발견된다; 이와 마찬가지로, 역방향 기준 블록은 현재 필드 바로 앞에 오는 프레임의 필드에서 발견된다. 현재 블록은 보상된 필드 내에 모아진다. 이러한 프로세스는 프레임의 두 번째 필드에서 계속된다. 2 개의 보상된 필드를 조합하여 순방향 및 역방향 보상된 프레임을 형성한다.
- [0163] 역 텔레시네 (606) 에서 생성된 프레임의 경우에는, 단지 재구성된 필름 프레임만 있기 때문에, 기준 블록에 대한 검색은 프레임 단위로만 행해진다. 2 개의 기준 블록 및 2 개의 차이인, 순방향 및 역방향이 발견되어, 또한 순방향 역방향 보상된 프레임으로 된다. 요약하면, 모션 보상은 모든 블록에 대해 모션 벡터와 차이 메트릭을 생성한다; 그러나, 일 블록은 디인터레이서 (605) 의 출력인 경우에 NTSC 필드의 일부이고, 역 텔레시네의 출력이 처리되는 경우에 필름 프레임의 일부이다. 그 메트릭의 차이는, 순방향 또는 역방향 차이가 평가되고 있는지 여부에 따라, 이전 필드 또는 프레임에서 또는 그 바로 다음에 오는 필드 또는 프레임에서, 고려 중인 필드 또는 프레임과 그 최적 정합 블록 간에 평가된다는 것에 주목하자. 단지 휘도 값만이 이러한 계산 내에 입력된다.
- [0164] 이와 같이, 모션 보상 단계는 2 개의 차이 세트를 생성한다. 이들 차이 세트는, 적절한 때에 바로 앞에 있는 프레임과 바로 뒤에 있는 프레임에서 선택된 기준 블록 내의 휘도 값과 휘도의 현재 값 블록 간에 존재한다. 각각의 순방향 및 각각의 역방향 차이의 절대값은 각 화소에 대해 결정되고, 그 각각은 전체 프레임에 걸쳐 개별적으로 더해진다. 일 프레임을 포함하는 디인터레이싱된 NTSC 필드가 처리될 때, 양쪽 필드는 2 개의 덧셈에 포함된다. 이러한 방법으로, SAD_p 및 SAD_n , 순방향 및 역방향 차이의 절대값 합이 구해진다.
- [0165] 모든 프레임에 대해, SAD 비율은 다음 관계식을 사용하여 계산된다.

수학식 22

$$\gamma = \frac{\varepsilon + SAD_P}{\varepsilon + SAD_N}$$

[0166]

[0167] 여기서, SAD_P 및 SAD_N 은 각각 순방향 및 역방향 차이의 절대값 합이다. 작은 양의 실수 ε 를 분자에 더하여 "divide-by-zero" 에러를 방지한다. 유사한 ε 항을 분모에 더하여, 또한 SAD_P 또는 SAD_N 이 0 에 가까워질 때, γ 의 감도를 줄인다.

[0168] 다른 양태에서, 그 차이는 SSD (sum of squared difference) 일 수 있고, SAD (sum of absolute difference), 또는 SATD 일 수 있는데, 여기서, 화소 값 블록은, 블록 요소 내의 차이를 취하기 전에 2 차원 이산 코사인 변환을 적용함으로써 변환된다. 그 합은 액티브 비디오의 영역에 대해 평가되지만, 다른 양태에서는 더 작은 영역을 사용할 수도 있다.

[0169] 또한, 수신된 것과 같은 (모션 보상되지 않은) 모든 프레임의 휘도 히스토그램이 계산된다. 히스토그램은, 이용 가능한 경우에, DC 계수, 즉, 2 개의 2 차원 이산 코사인 변환을 휘도 값 블록에 적용한 결과인 16×16 계수 어레이 내의 (0, 0) 계수에 대해 작용한다. 이와 마찬가지로, 16×16 블록 내의 256 개 휘도 값의 평균 값을 히스토그램에서 사용할 수도 있다. 휘도 깊이가 8 비트인 이미지의 경우에, bin의 수는 16 으로 설정된다. 다음 메트릭은 히스토그램 차이를 평가한다.

수학식 23

$$\lambda = \frac{1}{N} \sum_{i=1}^{16} |N_{Pi} - N_{Ci}|$$

[0170]

[0171] 상기 수학식에서, N_{Pi} 는 i 번째 bin에서 이전 프레임으로부터의 블록 수이고, N_{Ci} 는 i 번째 bin에 속하는 현재 프레임으로부터의 블록 수이며, N 은 프레임 내의 전체 블록 수이다.

[0172] 이들 중간 결과를 모아서, 다음과 같은 현재 프레임 차이 메트릭을 형성한다.

수학식 24

$$D = \frac{\gamma_C}{\gamma_P} + \lambda(2\lambda + 1)$$

[0173]

[0174] 여기서, γ_C 는 현재 프레임에 기초한 SAD 비율이고, γ_P 는 이전 프레임에 기초한 SAD 비율이다. 장면이 평활한 모션을 갖고, 그 루마 히스토그램이 거의 변하지 않으면, $D \approx 1$ 이다. 현재 프레임이 갑작스런 장

면 전환을 표시하면, γ_C 는 커질 것이고, γ_P 는 작아져야 한다. γ_C 만을 사용하는 대신에 $\frac{\gamma_C}{\gamma_P}$ 비율을 사용하여, 문맥의 액티비티 레벨로 메트릭을 정규화한다.

[0175] 도 42 는 프레임에 압축 타입을 할당하는 프로세스를 도시한다. 수학식 19 에서 규정된 현재 프레임 차이인 D 는 프레임 할당과 관련하여 이루어진 결정에 대한 기준이다. 결정 블록 (4202) 이 나타내는 바와 같이, 고려 중인 프레임이 시퀀스 내의 첫 번째 프레임인 경우에, "예" 로 표시된 결정 경로가 블록 (4206) 다음에 옴으로써, 그 프레임이 I 프레임인 것으로 선언한다. 블록 (4208) 에서 누적된 프레임 차이가 0 으로 설정되고, (블록 (4210) 에서) 프로세스는 시작 블록으로 리턴한다. 고려 중인 프레임이 시퀀스 내의 첫 번째 프

레이미가 아니면, 그 결정이 이루어진 블록 (4202) 다음에 "아니오" 로 표시된 결정 경로가 오고, 테스트 블록 (4204) 에서, 현재 프레임 차이가 장면 전환 임계값에 대해 테스트된다. 현재 프레임 차이가 그 임계값보다 큰 경우에는, "예" 로 표시된 결정 경로가 블록 (4206) 으로 진행하여, 또 I-프레임을 할당하게 된다.

[0176] 현재 프레임 차이가 장면 전환 임계값보다 작은 경우에, "아니오" 경로는 블록 (4212) 으로 진행하여, 현재 프레임 차이에 누적된 프레임 차이가 더해진다. 결정 블록 (4214) 에서 흐름도를 통해 계속하여, 누적된 프레임 차이는 임계값 t 와 비교되는데, 이는 일반적으로 장면 전환 임계값보다 작다. 누적된 프레임 차이가 t 보다 큰 경우에, 제어는 블록 (4216) 으로 이동하고, 그 프레임에는 P 프레임이 할당된다; 그 다음에, 단계 (4218) 에서, 누적된 프레임 차이는 0 으로 리셋된다. 누적된 프레임 차이가 t 보다 작은 경우에, 제어는 블록 (4214) 에서 블록 (4220) 으로 이동한다. 여기에서, 현재 프레임 차이가 τ 와 비교되는데, 이는 t 보다 작다. 현재 프레임 차이가 τ 보다 작은 경우에, 블록 (4222) 에서 그 프레임은 스킵될 것으로 할당되고, 그 다음에, 프로세스가 리턴한다; 현재 프레임 차이가 τ 보다 큰 경우에, 블록 (4226) 에서 그 프레임은 B-프레임이 되도록 할당된다.

[0177] 인코더

[0178] 다시 도 2 를 참조하면, 트랜스코더 (200) 는 전처리기 (226) 로부터 처리된 메타데이터와 미가공 비디오를 수신하는 인코더 (228) 를 포함한다. 메타데이터는 소스 비디오 (104) 에서 원래 수신된 임의의 정보와 전처리기 (226) 에 의해 계산된 임의의 정보를 포함할 수 있다. 인코더 (228) 는 제 1 패스 인코더 (230), 제 2 패스 인코더 (232) 및 재인코더 (234) 를 포함한다. 또한, 인코더 (228) 는 전처리기 (226) 뿐만 아니라 제 1 패스 인코더 (230), 재인코더 (234) 에도 제 2 패스 인코더 (232) 로부터의 정보 (예를 들어, 메타데이터, 예러 내성 정보, 콘텐츠 정보, 인코딩된 비트레이트 정보, 베이스 계층과 인헨스먼트 계층 균형 정보 및 양자화 정보) 를 제공할 수 있는 트랜스코더 컨트롤 (231) 로부터 입력을 수신한다. 인코더 (228) 는 예를 들어, 콘텐츠 분류 모듈 (712; 도 7) 에 의해 인코더 (228) 자체에 의해 생성되는 콘텐츠 정보 및/또는 전처리기 (226) 로부터 수신된 콘텐츠 정보를 사용하여 수신된 비디오를 인코딩한다.

[0179] 도 7 은 도 2 에 도시된 인코더 (228) 에 사용될 수도 있는 예시적인 2-패스 인코더에 포함될 수 있는 기능 모듈의 블록도를 도시한다. 도 7 에는 기능 블록의 여러 양태가 도시되어 있지만, 도 7 및 본원의 설명은 인코더에 통합될 수 있는 모든 기능을 반드시 다루는 것은 아니다. 따라서, 이하, 그 기능 모듈의 일정 양태는 아래의 베이스 및 인헨스먼트 계층 인코딩의 설명 다음에 설명된다.

[0180] 베이스 계층 및 인헨스먼트 계층 인코딩

[0181] 인코더 (228) 는 SNR 스케일러를 인코더일 수 있고, 이는 전처리기 (226) 로부터의 미가공 비디오와 메타데이터를 인코딩된 제 1 데이터 그룹 (또한, 베이스 계층으로 지칭됨) 와 인코딩된 하나 이상의 추가 데이터 그룹 (또한, 인헨스먼트 계층으로 지칭됨) 으로 인코딩할 수 있다. 인코딩 알고리즘은 베이스 계층 및 인헨스먼트 계층 계층을 생성하는데, 이들 계층은, 디코딩 시, 양쪽 계층을 디코딩에 사용할 수 있을 때 디코더에서 결합될 수도 있다. 양쪽 계층을 이용할 수 없는 경우에, 베이스 계층의 인코딩은 베이스 계층이 단일 계층으로서 디코딩되는 것을 허용한다.

[0182] 도 31 을 참조하여 이와 같은 다층 인코딩 프로세스의 일 양태를 설명한다. 블록 (321) 에서, I 프레임은 완전히 인트라 코딩된 매크로블록 (인트라 코딩된 MB) 으로 인코딩된다. H.264 에서, I 프레임 내의 인트라 코딩된 MB 는 완전히 이용된 공간 예측으로 인코딩되어, 상당한 양의 코딩 이득을 제공한다. 2 개의 서브모드가 존재한다: 인트라 4×4 및 인트라 16×16 . 베이스 계층이 공간 예측에 의해 제공된 코딩 이득을 이용하려면, 인헨스먼트 계층을 인코딩 및 디코딩 하기 전에 베이스 계층을 인코딩 및 디코딩해야 한다. I 프레임의 2 패스 인코딩 및 디코딩이 사용된다. 베이스 계층에서, 베이스 계층 양자화 파라미터 QP_b 는 덜 세분화된 양자화 스텝 크기를 변환 계수에 제공한다. 원래 프레임과 재구성된 베이스 계층 프레임 간의 화소 단위 차이는 인헨스먼트 계층에서 인코딩될 것이다. 인헨스먼트 계층은 양자화 파라미터 QP_e 를 사용하여, 더 미세한 양자화 스텝 크기를 제공한다. 도 2 의 인코더 (228) 와 같은 인코딩 수단은 블록 (321) 에서 인코딩을 수행할 수 있다.

[0183] 블록 (323) 에서, 인코더는 처리되는 GOP 에서 P 및/또는 B-프레임에 대한 베이스 계층 데이터와 인헨스먼트 계층 데이터를 인코딩한다. 인코더 (228) 와 같은 인코딩 수단은 블록 (323) 에서 인코딩을 수행할 수 있다. 블록 (325) 에서, 인코딩 프로세스는 인코딩할 P 또는 B-프레임이 더 있는지를 검사한다. SNR 스케일러

블 인코더 (228) 와 같은 인코딩 수단은 행동 (325) 을 수행할 수 있다. P 또는 B-프레임이 더 남아 있는 경우에는, GOP 내의 모든 프레임을 인코딩하는 것을 종료할 때까지, 단계 (323) 를 반복한다. P 및 B-프레임은 인터 코딩된 매크로블록 (인터 코딩된 MB) 으로 이루어져 있지만, 아래에서 설명되는 바와 같이, P 및 B-프레임에 인트라 코딩된 MB 가 존재할 수 있다.

[0184] 디코더가 베이스 계층과 인헨스먼트 계층 데이터를 구별하기 위해, 인코더 (228) 는 블록 (327) 에서 오버헤드 정보를 인코딩한다. 오버헤드 정보의 타입은, 예를 들어, 계층 수를 식별하는 데이터, 베이스 계층으로서 계층을 식별하는 데이터, 인헨스먼트 계층으로서 계층을 식별하는 데이터, 계층 간 상호 관계 (예를 들어, 계층 2 는 베이스 계층 1 에 대한 인헨스먼트 계층이고, 또는 계층 3 은 인헨스먼트 계층 2 에 대한 인헨스먼트 계층임) 를 식별하는 데이터, 또는 인헨스먼트 계층 스트링 내의 최종 인헨스먼트 계층으로서 계층을 식별하는 데이터를 포함할 수 있다. 오버헤드 정보는 별개의 데이터 메시지에 관련되거나 포함되는 베이스 및/또는 인헨스먼트 계층 데이터와 접속되는 헤더에 포함될 수 있다. 도 2 의 인코더 (228) 와 같은 인코딩 수단은 블록 (327) 에서 프로세스를 수행할 수 있다.

[0185] 단일 계층 디코딩을 갖기 위해, 역 양자화 전에 2 개 계층의 계수를 결합한다. 따라서, 2 개 계층의 계수는 인터랙티브 방식으로 생성되어야 한다; 그렇지 않은 경우에는, 상당한 양의 오버헤드가 발생할 수 있다. 오버헤드가 증가하는 하나의 이유는 베이스 계층 인코딩과 인헨스먼트 계층 인코딩이 상이한 시간 기준을 사용할 수 있기 때문이다. 양쪽 계층을 이용할 수 있을 때 역양자화 전에 디코더에서 결합될 수 있는 베이스 계층과 인헨스먼트 계층 계수를 생성하기 위한 알고리즘이 요구된다. 이와 동시에, 인헨스먼트 계층을 이용할 수 없거나 디코더가 예를 들어, 전원 절약과 같은 이유 때문에 인헨스먼트 계층을 디코딩하지 않기로 결정하는 경우에, 그 알고리즘은 허용 가능한 베이스 계층 비디오를 제공해야 한다. 이하, 그러한 프로세스의 일 예시적인 예의 상세는 바로 다음에 오는 표준 예측 코딩의 간단한 설명 문맥에서 더 설명된다.

[0186] P-프레임 (또는 임의의 인터 코딩된 섹션) 은 현재 영상 내의 영역과 기준 영상 내의 최적 정합 예측 영역 간의 시간 중복을 이용할 수 있다. 기준 프레임 내의 최적 정합 예측 영역의 위치는 모션 벡터에서 인코딩될 수 있다. 현재 영역과 최적 정합 기준 예측 영역 간의 차이는 나머지 에러 (또는 예측 에러) 로서 공지되어 있다.

[0187] 도 32 는 예를 들어, MPEG-4 에 따른 P-프레임 구성 프로세스의 일 예를 도시한다. 프로세스 (331) 는 도 31 의 블록 (323) 에서 일어날 수 있는 예시적인 프로세스를 더 상세히 도시한다. 프로세스 (331) 는 5×5 매크로블록으로 이루어진 현재 영상 (333) 을 포함하는데, 본 예에 따른 매크로블록의 수는 임의적이다. 매크로블록은 16×16 화소로 이루어진다. 화소는 8 비트 휘도 값 (Y) 과 2 개의 8 비트 색차 값 (Cr 및 Cb) 에 의해 규정될 수 있다. MPEG 에 따르면, Y, Cr 및 Cb 성분은 4:2:0 포맷으로 저장될 수 있는데, Cr 및 Cb 성분은 X 및 Y 방향으로 2 만큼 다운샘플링된다. 따라서, 각각의 매크로블록은 256 개의 Y 성분, 64 개의 Cr 성분 및 64 개의 Cb 성분으로 이루어질 것이다. 현재 영상 (333) 의 매크로블록 (335) 은 현재 프레임 (333) 과 상이한 시점에서 기준 영상 (337) 으로부터 예측된다. 기준 영상 (337) 에서 검색을 행하여, Y, Cr 및 Cb 값의 관점에서, 인코딩되는 현재 매크로블록 (335) 과 가장 가까운 최적 정합 매크로블록 (339) 의 위치를 찾는다. 기준 영상 (337) 에서 최적 정합 매크로블록 (339) 의 위치는 모션 벡터 (341) 에 인코딩된다. 기준 영상 (337) 은, 현재 영상 (323) 의 구성 전에 디코더가 재구성하게 될 I-프레임 또는 P 프레임일 수 있다. 현재 매크로블록 (335) 에서 최적 정합 매크로블록 (339) 을 빼면 (Y, Cr 및 Cb 성분 각각의 차이가 계산됨), 나머지 에러 (343) 가 된다. 나머지 에러 (343) 는 2D 이산 코사인 변환 (345; DCT) 을 사용하여 인코딩된 후에, 양자화된다 (347). 양자화 (347) 를 수행하여, 예를 들어, 고주파수 계수에 더 적은 비트를 할당하면서 저주파수 계수에 더 많은 비트를 할당함으로써 공간 압축을 제공할 수 있다. 정보를 식별하는 기준 영상 (333) 및 모션 벡터 (341) 와 함께, 나머지 에러 (343) 의 양자화된 계수는 현재 매크로블록 (335) 을 표현한 인코딩된 정보이다. 인코딩된 정보는 추후 사용을 위해 메모리에 저장되거나, 예를 들어, 에러 정정 또는 이미지 인헨스먼트를 위해 조작되거나, 또는 네트워크 (349) 를 통해 송신될 수 있다.

[0188] 인코딩된 모션 벡터 (341) 와 함께, 나머지 에러 (343) 의 인코딩된 양자화된 계수를 사용하여, 후속하는 모션 추정 및 보상을 위한 기준 프레임의 일부로서 사용되는 인코더에서 현재 매크로블록 (335) 을 재구성할 수 있다. 인코더는 이러한 P 프레임 재구성을 위해 디코더의 절차를 에뮬레이션할 수 있다. 디코더의 에뮬레이션으로 인해, 인코더와 디코더 모두가 동일 기준 영상과 함께 작업하게 된다. 이때, 추후 인터 코딩을 위해 인코더에서 행해지는지, 또는 디코더에서 행해지는지 여부에 관계없이, 재구성 프로세스가 제공된다. P 프레임의 재구성은 기준 프레임 (또는 참조되고 있는 영상이나 프레임의 일부) 이 재구성된 후에 시작할 수 있다. 인코딩되어 양자화된 계수를 역양자화 (351) 한 후에, 2D IDCT (Inverse DCT) 를 수행 (353) 하면,

나머지 에러 (355) 가 디코딩되거나 재구성된다. 인코딩된 모션 벡터 (341) 를 디코딩하고, 이를 사용하여, 미리 재구성된 기준 영상 (337) 에서 미리 재구성된 최적 정합 매크로블록 (357) 의 위치를 찾는다. 그 다음에, 재구성된 나머지 에러 (355) 를 재구성된 최적 정합 매크로블록 (357) 에 추가하여, 재구성된 매크로블록 (359) 을 형성한다. 재구성된 매크로블록 (359) 은 메모리에 저장되거나, 다른 재구성된 매크로블록과 함께 하나의 영상으로 또는 독립적으로 표시되거나, 또는 이미지 인헨스먼트를 위해 더 처리될 수 있다.

[0189] B-프레임 (또는 양방향 예측을 사용하여 코딩된 임의의 섹션) 은 현재 영상 내의 영역, 이전 영상 내의 최적 정합 예측 영역 및 후속 영상 내의 최적 정합 예측 영역 간의 시간 중복을 이용할 수 있다. 후속 최적 정합 예측 영역과 이전 최적 정합 예측 영역을 결합하여, 결합된 양방향 예측 영역을 형성한다. 현재 영상 영역과 최적 정합 결합된 양방향 예측 영역 간의 차이는 나머지 에러 (또는 예측 에러) 이다. 후속 기준 영상 내의 최적 정합 예측 영역과 이전 기준 영상 내의 최적 정합 예측 영역의 위치는 2 개의 모션 벡터에 인코딩될 수 있다.

[0190] 도 33 은 인코더 (228) 에 의해 수행될 수 있는 베이스 계층 및 인헨스먼트 계층 계수의 인코딩을 위한 인코더 프로세스의 일 예를 도시한다. 베이스 및 인헨스먼트 계층을 인코딩하여 SNR 스케일러블 비트스트림을 제공한다. 도 33 은 도 31 의 단계 (323) 에서 행해지는 것과 같은 인터 MB 나머지 에러 계수를 인코딩하는 일 예를 도시한다. 그러나, 유사한 방법을 사용하여, 인트라 MB 계수를 인코딩할 수도 있다. 도 2 의 인코더 컴포넌트 (228) 와 같은 인코딩 수단은 도 33 에 도시된 프로세스와 도 32 의 단계 (323) 를 수행할 수 있다. (인코딩될) 원래 비디오 데이터 (406) (본 예에서, 비디오 데이터는 루마 및 크로마 정보를 포함함) 는 베이스 계층 최적 정합 매크로블록 루프 (302) 와 인헨스먼트 계층 최적 정합 매크로블록 루프 (365) 에 입력된다. 양 루프 (363 및 365) 의 목적은 각각 덧셈기 (367 및 369) 에서 계산되는 나머지 에러를 최소화하는 것이다. 루프 (363 및 365) 는 도시된 바와 같이 동시에 수행될 수 있고, 또는 순차적으로 수행될 수 있다. 루프 (363 및 365) 는, 최적 정합 매크로블록과 원래 데이터 (361) 간의 나머지 에러를 최소화하는 최적 정합 매크로블록을 식별하도록, 기준 프레임을 포함한 버퍼 (371 및 373) 검색 논리를 각각 포함한다 (버퍼 (371 및 373) 는 동일 버퍼일 수 있음). 통상 베이스 계층 루프 (363) 가 인헨스먼트 계층 루프 (365) 보다 덜 세분화된 양자화 스텝 크기 (QP 값이 더 높음) 를 사용할 것이므로, 루프 (363 및 365) 의 나머지 에러는 상이할 것이다. 변환 블록 (375 및 377) 은 각 루프의 나머지 에러를 변환한다.

[0191] 그 다음에, 변환된 계수는 선택기 (379) 에서 베이스 계층과 인헨스먼트 계층 계수 내에 파싱된다. 아래에 설명하는 바와 같이, 선택기 (379) 의 파싱은 몇몇 형태를 취할 수 있다. 파싱 기술의 하나의 공통 특징은

인헨스먼트 계층 계수 C'_{enh} 가 베이스 계층 계수 C'_{base} 에 대해 차분 세분화 (differential refinement) 되도록 계산된다는 것이다. 베이스 계층에 대해 세분화되도록 인헨스먼트 계층을 계산함으로써, 디코더가 단독으로 베이스 계층 계수를 디코딩하여 이미지의 적당한 표현을 가질 수 있고, 또는 베이스 및 인헨스먼트 계층 계수를 결합하여 이미지의 표현을 세분화할 수 있다. 그 다음에, 선택기 (379) 에 의해 선택된 계수를 양자화기 (381 및 383) 에 의해 양자화한다. 양자화된 계수 \tilde{C}'_{base} 및 \tilde{C}'_{enh} (각각 양자화기 (381 및 383) 를 사용하여 계산됨) 는 메모리에 저장될 수 있고, 또는 네트워크를 통해 디코더로 송신될 수 있다.

[0192] 디코더에서 매크로블록의 재구성을 정합시키기 위해, 역양자화기 (385) 는 베이스 계층 나머지 에러 계수를 역양자화한다. 역양자화된 나머지 에러 계수는 버퍼 (371) 에서 찾은 최적 정합 매크로블록으로 역변환 (387) 및 가산 (389) 되어, 재구성된 매크로블록이 디코더에서 재구성될 매크로블록과 정합하게 된다. 양자화기 (383), 역양자화기 (391), 역변환기 (393), 덧셈기 (397) 및 버퍼 (373) 는 베이스 계층 루프 (363) 에서 행해진 것과 유사한 계산을 인헨스먼트 루프 (365) 에서 수행한다. 또한, 덧셈기 (393) 를 사용하여, 인헨스먼트 계층의 재구성에 사용되는 역양자화된 인헨스먼트 계층 및 베이스 계층 계수를 결합한다. 통상, 인헨스먼트 계층 양자화기 및 역양자화기는 베이스 계층보다 미세한 양자화기 스텝 크기 (QP 가 낮음) 를 이용할 것이다.

[0193] 도 34, 도 35 및 도 36 은 도 33 의 선택기 (379) 에서 사용될 수 있는 베이스 계층 및 인헨스먼트 계층 계수 선택 프로세스의 예를 도시한다. 도 2 의 인코더 (228) 와 같은 선택 수단은 도 34, 도 35 및 도 36 에 도시된 프로세스를 수행할 수 있다. 도 34 를 예로 들면, 변환된 계수는 다음 수학식에 표현된 바와 같이 베이스 및 인헨스먼트 계층 계수 내에 파싱된다.

수학식 25

$$C'_{base} = \begin{cases} 0, & \text{if } C_{base} \text{ and } C_{enh} \text{ 가 반대 부호} \\ \min(C_{base}, C_{enh}), & \text{otherwise} \end{cases}$$

[0194]

수학식 26

$$C'_{enh} = C_{enh} - Q_b^{-1}(Q_b(C'_{base}))$$

[0195]

[0196] 여기서, "min" 함수는 2 개의 인수 (argument) 의 수학적인 최소 또는 최소 크기일 수 있다. 수학식 25 는 블록 (401) 으로서 도시되어 있고, 수학식 26 은 도 34 의 덧셈기 (510) 로서 도시되어 있다. 수학식 26 에
서, Q_b 는 베이스 계층 양자화기 (381) 를 나타내고, Q_b^{-1} 는 베이스 계층의 역양자화기 (385) 를 나타낸다. 수학식 26 은 인헨스먼트 계층 계수를 수학식 25 를 사용하여 계산된 베이스 계층 계수의 차분 세분화로 변환한다.

[0197] 도 35 는 베이스 계층 및 인헨스먼트 계층 계수 선택기 (379) 의 다른 예를 도시한다. 본 예에서, 블록 (405) 에 포함된 수학식 (.) 은 다음 수학식을 표현한다.

수학식 27

$$C'_{base} = \begin{cases} C_{base}, & \text{if } |Q_b^{-1}Q_b(C_{base}) - C_{enh}| < |C_{enh}| \\ 0, & \text{otherwise} \end{cases}$$

[0198]

[0199] 덧셈기 (407) 는 다음 2 개의 수학식에 도시된 것과 같은 인헨스먼트 계층 계수를 계산한다:

수학식 28

$$C'_{enh} = C_{enh} - Q_b^{-1}(Q_b(C'_{base}))$$

[0200]

[0201] 여기서, C'_{base} 는 수학식 27 에 의해 주어진다.

[0202] 도 36 은 베이스 계층 및 인헨스먼트 계층 선택기 (379) 의 다른 예를 도시한다. 본 예에서, 베이스 계층 계수는 변하지 않고, 인헨스먼트 계층 계수는 양자화/역양자화된 베이스 계층 계수 및 원래 인헨스먼트 계층 계수 간의 차이와 같다.

[0203] 베이스 및 인헨스먼트 계층 나머지 예러 계수 외에도, 디코더는 어떻게 MB 가 인코딩되는지를 식별하는 정보를 필요로 한다. 도 2 의 인코더 컴포넌트 (228) 와 같은 인코딩 수단은 오버헤드 정보를 인코딩할 수 있고, 이 오버헤드 정보는 예를 들어, MB 맵과 같은 인트라 코딩 및 인터 코딩된 부분의 맵을 포함할 수 있으며, 이 MB 맵에서, 매크로블록 (또는 서브 매크로블록) 은 인트라 코딩되거나 인터 코딩된 것으로서 식별된다 (또한, 예를 들어, 순방향, 역방향 또는 양방향을 비롯한 어느 타입의 인터 코딩인지가 식별되고, 인터 코딩된 부분이 어느 프레임(들)을 참조하는지가 식별됨). 예시적인 양태에서, MB 맵과 베이스 계층 계수는 베이스 계층에서 인코딩되고, 인헨스먼트 계층 계수는 인헨스먼트 계층에서 인코딩된다.

[0204] P-프레임 및 B-프레임은 인터 MB 뿐만 아니라 인트라 코딩된 MB 도 포함할 수 있다. 하이브리드 비디오 인

코더가 RD (rate distortion) 최적화를 사용하여 P 또는 B-프레임 내의 일정 매크로블록을 인트라 코딩된 MB 로서 인코딩하기로 결정하는 것이 일반적이다. 인트라 코딩된 MB 가 인헨스먼트 계층 인터 MB 에 의존하지 않는 단일 계층 디코딩을 갖기 위해, 임의의 인접한 인터 MB 가 베이스 계층 인트라 코딩된 MB 의 공간 예측에 사용되지 않는다. 베이스 계층 P 또는 B-프레임에서 인트라 코딩된 MB 의 경우에, 계산 복잡도가 인헨스먼트 계층 디코딩 동안에 불변 상태를 유지하기 위해서는, 인헨스먼트 계층에서의 세분화를 스킵할 수 있다.

[0205] P 또는 B-프레임 내의 인트라 코딩된 MB 는 인터 MB 보다 많은 비트를 필요로 할 수도 있다. 이러한 이유로, P 또는 B-프레임 내의 인트라 코딩된 MB 는 더 높은 QP 로 베이스 계층 품질에서만 인코딩될 수 있다.

이로 인해, 비디오 품질이 약간 저하되지만, 상술한 바와 같이 베이스 및 인헨스먼트 계층에서 인터 MB 계층을 사용하여 나중 프레임에서 세분화되면, 이러한 저하를 인식할 수 없을 것이다. 2 가지 이유 때문에 이러한 저하를 인식할 수 없게 된다. 첫 번째 이유는 HVS (human visual system) 의 특징 때문이고, 다른 이유는 인터 MB 가 인트라 MB 를 세분화한다는 사실 때문이다. 제 1 프레임과 제 2 프레임 간에 위치가 변하는 물체의 경우에, 제 1 프레임 내의 몇몇 화소는 제 2 프레임에서 볼 수 없게 되고 (숨겨질 정보), 제 2 프레임 내의 몇몇 화소는 제 1 프레임 동안에 볼 수 있게 된다 (알려진 정보). 사람 눈은 알려진 정보와 숨겨질 시각 정보에 민감하지 않다. 알려진 정보의 경우에는, 낮은 품질로 인코딩되더라도, 사람 눈이 그 차이를 구별하지 못할 수도 있다. 동일 정보가 다음 P 프레임에 남아 있으면, 인헨스먼트 계층이 낮은 QP 를 갖기 때문에, 인헨스먼트 계층에서 다음 P 프레임이 그 동일 정보를 세분화할 수 있는 가능성이 클 것이다.

[0206] P 또는 B-프레임에서 인트라 코딩된 MB 를 도입하는 다른 일반적인 기술은 인트라 리프레시 (Intra Refresh) 로 공지되어 있다. 이 경우에, 표준 R-D 최적화는 MB 가 인터 코딩된 MB 일 것을 요구하더라도, 몇몇 MB 는 인트라 코딩된 MB 로서 코딩된다. 베이스 계층에 포함된 이들 인트라 코딩된 MB 는 QP_b 또는 QP_e 를 사용하여 인코딩될 수 있다. QP_e 가 베이스 계층에 사용되면, 인헨스먼트 계층에서는 어떤 세분화도 필요하지 않다.

QP_b 가 베이스 계층에 사용되면, 세분화를 필요로 할 수도 있지만, 그렇지 않은 경우에는, 인헨스먼트 계층에서, 품질 저하를 인식할 수 있을 것이다. 코딩 효율 면에서 인터 코딩이 인트라 코딩보다 효율적이므로, 인헨스먼트 계층에서의 이들 세분화는 인터 코딩될 것이다. 이와 같이, 베이스 계층 계층수는 인헨스먼트 계층에 사용되지 않을 것이다. 따라서, 새로운 연산을 도입함 없이 인헨스먼트 계층에서 품질이 개선된다.

[0207] B-프레임은 B-프레임이 제공하는 높은 압축 품질 때문에 인헨스먼트 계층에서 일반적으로 사용된다. 그러나, B-프레임은 P 프레임의 인트라 코딩된 MB 를 참조할 필요가 있을 수도 있다. B-프레임의 화소가 인헨스먼트 계층 품질로 인코딩되지 않는다면, 상술한 바와 같이, P 프레임 인트라 코딩된 MB 의 낮은 품질 때문에 너무 많은 비트를 필요로 할 수도 있다. 상술한 바와 같이, HVS 의 품질을 이용함으로써, B-프레임 MB 는 P 프레임의 낮은 품질로 인트라 코딩된 MB 를 참조하여 낮은 품질로 인코딩될 수 있다.

[0208] P 또는 B-프레임에서 인트라 코딩된 MB 의 한 가지 극단적인 경우는, P 또는 B-프레임 내의 모든 MB 가 인코딩되는 비디오의 장면 전환의 존재로 인해 인트라 모드에서 인코딩되는 경우이다. 이 경우에는, 전체 프레임을 베이스 계층 품질로 코딩할 수 있고, 인헨스먼트 계층에서는 어떤 세분화도 이루어지지 않는다. B-프레임에서 장면 전환이 일어나는 경우에, B-프레임이 인헨스먼트 계층에서 인코딩될 뿐이라고 가정하면, B-프레임은 베이스 계층 품질로 인코딩되거나 단순히 드롭될 수 있다. 장면 전환이 P 프레임에서 일어나면, 어떤 변화도 필요로 하지 않지만, P 프레임을 드롭하거나 베이스 계층 품질에서 인코딩할 수 있다. 스케일러블 계층 인코딩은 발명의 명칭이 "SCALABLE VIDEO CODING WITH TWO LAYER ENCODING AND SINGLE LAYER DECODING" 인 동시에 계속 중인 미국 특허 출원 [대리인 docket/ref. no. 050078] 에 더 설명되어 있고, 이는 본원의 양수인이 소유하고 있으며 그 전체를 본원에서 참조로서 병합하고 있다.

[0209] 인코더 제 1 패스 부분

[0210] 도 7 은 도 2 의 인코더 (228) 의 예시적인 예를 도시한다. 도시된 블록은 인코더 (228) 에 포함될 수 있는 여러 인코더 처리를 도시한다. 본 예에서, 인코더 (228) 는 경계선 (704) 위의 제 1 패스 부분 (706) 및 경계선 (704) 아래의 제 2 패스 부분 (706; 도 2 의 제 2 패스 인코더와 재인코더 (234) 의 기능을 포함함) 을 포함한다.

[0211] 인코더 (228) 는 전처리기 (226) 로부터 메타데이터 및 미가공 비디오를 수신한다. 메타데이터는, 비디오

콘텐츠 정보에 관련된 메타데이터를 비롯한, 전처리기 (226) 에 의해 수신 또는 계산되는 임의의 메타데이터를 포함할 수 있다. 인코더 (228) 의 제 1 패스 부분 (702) 은 제 1 패스 인코딩 (702) 에 포함될 수 있는 예시적인 프로세스를 도시하는데, 이는 그 기능과 관련하여 아래에 설명되어 있다. 당업자라면, 그러한 기능이 여러 형태 (예를 들어, 하드웨어, 소프트웨어, 펌웨어 또는 그 조합) 로 수록될 수 있다는 것을 알 수 있다.

[0212] 도 7 은 AIR (adaptive intra refresh) 모듈을 도시한다. AIR 모듈 (710) 은 메타데이터에 기초하여 I-프레임을 인스턴스화하는 I-프레임 인스턴스화 모듈 (708) 에 대한 입력을 제공한다. 또한, 제 1 패스 부분 (702) 은 메타데이터 및 비디오를 수신하고 비디오 관련 콘텐츠 정보를 결정하도록 구성된 콘텐츠 분류 모듈 (712) 을 포함할 수 있다. 콘텐츠 정보는 레이트 제어 비트 할당 모듈 (714) 에 제공될 수 있고, 또한 이 레이트 제어 비트 할당 모듈 (714) 은 메타데이터와 비디오를 수신한다. 레이트 제어 비트 할당 모듈 (714) 은 레이트 비트 제어 정보를 결정하고, 이 정보를 모드 결정 모듈 (715) 에 제공한다. 콘텐츠 정보와 비디오 오는 인트라 모델 (왜곡) 모듈 (716) 에 제공될 수 있고, 이 인트라 모델 (왜곡) 모듈 (716) 은 인트라 코딩 왜곡 정보를 모드 결정 모듈 (715) 및 베이스 및 인헨스먼트 계층에 대한 스케일러빌리티 레이트 왜곡 모듈 (718) 에 제공한다. 비디오 및 메타데이터는 모션 추정 (왜곡) 모듈 (720) 에 제공되어, 인터 코딩 왜곡 정보를 베이스 및 인헨스먼트 계층에 대한 스케일러빌리티 레이트 왜곡 모듈 (718) 에 제공한다. 베이스 및 인헨스먼트 계층에 대한 스케일러빌리티 레이트 왜곡 모듈 (718) 은 모션 추정 모듈 (720) 및 인트라 모델 왜곡 모듈 (716) 로부터의 왜곡 추정을 사용하여 스케일러빌리티 레이트 왜곡 정보를 결정하고, 이 정보를 모드 결정 모듈 (715) 에 제공한다. 또한, 모드 결정 모듈 (715) 은 슬라이스/MB 순서화 모듈 (722) 로부터 입력을 수신한다. 슬라이스/MB 순서화 모듈 (722) 은 (제 2 패스 부분 (706) 에 도시된) 여러 내성 모듈 (740) 로부터 입력을 수신하고, 여러 내성을 위한 액세스 유닛 경계를 갖는 비디오의 독립적으로 인코딩 가능한 부분 (슬라이스) 을 정렬하는 것에 대한 정보를 모드 결정 모듈 (715) 에 제공한다. 모드 결정 모듈 (715) 은 그 입력에 기초하여 모드 정보를 인코딩하는 것을 결정하고, "최적" 코딩 모드를 제 2 패스 부분 (706) 에 제공한다. 이하, 그러한 제 1 패스 부분 (702) 인코딩의 몇몇 예에 대한 예시적인 설명을 제공한다.

[0213] 상술한 바와 같이, 콘텐츠 분류 모듈 (712) 은 전처리기 (226) 에 의해 공급된 메타데이터와 미가공 비디오를 수신한다. 몇몇 예에서, 전처리기 (226) 는 멀티미디어 데이터로부터 콘텐츠 정보를 계산하고, 콘텐츠 정보를 콘텐츠 분류 모듈 (712) 에 (메타데이터 형식으로) 제공하는데, 이 콘텐츠 분류 모듈 (712) 은 콘텐츠 정보를 사용하여, 멀티미디어 데이터에 대한 콘텐츠 분류를 결정할 수 있다. 몇몇 다른 양태에서, 콘텐츠 분류 모듈 (712) 은 멀티미디어 데이터로부터 여러 콘텐츠 정보를 결정하도록 구성되고, 또한 콘텐츠 분류를 결정하도록 구성될 수 있다.

[0214] 콘텐츠 분류 모듈 (712) 은 여러 타입의 콘텐츠를 갖는 비디오에 대해 상이한 콘텐츠 분류를 결정하도록 구성될 수 있다. 상이한 콘텐츠 분류로 인해, 예를 들어, 양자화 파라미터, 모션 추정, 스케일러빌리티, 여러 내성을 결정하고, 채널 전체에서 최적 멀티미디어 데이터 품질을 유지하며, 고속 채널 스위칭 방식 (예를 들어, I-프레임으로 하여금 고속 채널 스위칭을 주기적으로 허용하도록 강제하는 방식) 을 위한 비트 레이트 (예를 들어, 비트 할당) 을 결정하는 것과 같이, 멀티미디어 데이터를 인코딩하는 양태에 따라, 상이한 파라미터가 사용될 수 있다. 일 예에 따르면, 인코더 (228) 는 콘텐츠 분류에 기초하여 R-D (rate-distortion) 최적화와 비트 레이트 할당을 결정하도록 구성된다. 콘텐츠 분류를 결정함으로써, 멀티미디어 데이터를 콘텐츠 분류에 기초하여 원하는 비트 레이트에 대응하는 소정의 품질 레벨로 압축할 수 있게 된다. 또한, 멀티미디어 데이터의 콘텐츠를 분류함으로써 (예를 들어, 인간 시각 시스템에 기초하여 콘텐츠 분류를 결정함으로써), 수신 장치의 디스플레이 상에서 전달되는 멀티미디어 데이터의 결과로서 생성된 인식 가능한 품질은 비디오 콘텐츠에 의존하게 된다.

[0215] 콘텐츠 분류 모듈 (712) 이 콘텐츠를 분류하기 위해 경험하는 절차의 일 예로서, 도 9 는 콘텐츠 분류 모듈 (712) 이 동작할 수도 있는 예시적인 프로세스를 도시한 프로세스 (900) 를 도시한다. 도시된 바와 같이, 프로세스 (900) 는, 콘텐츠 분류 모듈 (712) 이 미가공 멀티미디어 데이터와 메타데이터를 수신하는 입력 블록 (902) 에서 시작한다. 그 다음에, 프로세스 (900) 는, 콘텐츠 분류 모듈 (712) 이 멀티미디어 데이터의 공간 정보와 시간 정보를 결정하는 블록 (904) 으로 진행한다. 몇몇 양태에서, 공간 및 시간 정보는 공간 및 시간 마스킹 (예를 들어, 필터링) 에 의해 결정된다. 공간 및 시간 정보는 장면 전환 데이터 및 MV (motion vector) 평활화를 포함한 메타데이터에 기초하여 결정될 수 있다. 그 다음에, 프로세스 (900) 는 공간 복잡도, 시간 복잡도 및 감도 추정을 수행하는 블록 (912) 으로 진행한다. 그 다음에, 프로세스 (900) 는, 블록 (904 및 912) 에서 결정된 공간, 시간 및 감도 데이터의 결과에 기초하여 멀티미디어 데이터의 콘텐츠를 분류하는 블록 (916) 으로 진행한다. 또한, 블록 (916) 에서, 특정 R-D (rate-distortion) 곡선을 선택할 수 있

고/있거나 R-D 곡선을 갱신할 수 있다. 그 다음에, 프로세스 (900) 는 출력 블록 (918) 으로 진행하고, 이 블록 (918) 에서, 출력은 공간 및 시간 액티비티 (예를 들어, 콘텐츠 분류) 를 나타내는 복잡도-왜곡 맵 또는 값, 및/또는 선택된 R-D 곡선을 포함할 수 있다. 다시 도 7 을 참조하면, 상술한 바와 같이, 콘텐츠 분류 모듈 (712) 은 레이트 제어 비트 할당 모듈 (714), 인트라 모델 (왜곡) 모듈 (716), 또한 I-프레임 인스턴스화 모듈 (708) 에 출력을 제공한다.

[0216] 콘텐츠 정보

[0217] 아래에 더 설명되는 바와 같이, 콘텐츠 분류 모듈 (712) 은 공간 복잡도, 시간 복잡도, 콘트라스트비 값, 표준 편차 및 프레임 차이 메트릭을 비롯한 다양한 콘텐츠 관련 메트릭을 포함한 다양한 콘텐츠 정보를 멀티미디어 데이터로부터 계산하도록 구성될 수 있다.

[0218] 콘텐츠 분류 모듈 (712) 은 멀티미디어 데이터의 공간 복잡도 및 시간 복잡도를 결정하고, 또한 공간 복잡도에 대한 텍스처 값과 시간 복잡도에 대한 모션 값을 연관시키도록 구성될 수 있다. 콘텐츠 분류 모듈 (712) 은 전처리기 (226) 로부터 인코딩되는 멀티미디어 데이터의 콘텐츠 관련 전처리된 콘텐츠 정보를 수신하고, 또는 다른 방법으로는, 전처리기 (226) 는 콘텐츠 정보를 계산하도록 구성될 수 있다. 상술한 바와 같이, 콘텐츠 정보는, 예를 들어, 하나 이상의 D_{csat} 값, 콘트라스트비 값, MV (motion vector) 및 SAD (sum of absolute difference) 를 포함할 수 있다.

[0219] 일반적으로, 멀티미디어 데이터는 하나 이상의 이미지 시퀀스, 또는 프레임을 포함한다. 각 프레임은 처리를 위해 화소 블록으로 분해될 수 있다. 공간 복잡도는 일반적으로 프레임 내의 공간 상세 레벨의 측정을 설명하는 광의의 용어이다. 대체로 단조롭거나 변하지 않거나 약하게 변하는 휘도 및 색차 영역을 갖는 장면은 낮은 공간 복잡도를 가질 것이다. 공간 복잡도는 비디오 데이터의 텍스처와 연관된다. 본 양태에서, 공간 복잡도는, 로컬 공간 주파수 및 주위 조명의 함수로서 각 블록에 대해 계산되는 인간 시각 감도 메트릭 (D_{csat} 로 지칭됨) 에 기초한다. 당업자라면, 시각 이미지의 공간 주파수 패턴과 조명 및 콘트라스트 특성을 사용하는 기술은 인간 시각 시스템을 이용한다는 알 수 있다. 다수의 감도 메트릭이 인간 시각 시스템의 원근적 제한을 이용하는 것으로서 공지되어 있고, 본원에 설명된 방법과 함께 사용될 수 있다.

[0220] 시간 복잡도는 프레임 시퀀스 내의 프레임 간에 참조되는 것과 같은 멀티미디어 데이터의 모션 레벨 측정을 설명하는데 일반적으로 사용되는 광의의 용어이다. 모션이 거의 없거나 전혀 없는 장면 (예를 들어, 비디오 데이터의 프레임 시퀀스) 은 낮은 시간 복잡도를 갖는다. 시간 복잡도는 각 매크로블록에 대해 계산될 수 있고, D_{csat} 값, 모션 벡터 및 한 프레임과 다른 프레임 (예를 들어, 기준 프레임) 간의 절대 화소 차이의 합에 기초할 수 있다.

[0221] 프레임 차이 메트릭은 예측기와 현재 매크로블록 간의 SAD (sum of absolute difference) 로서 표현된 나머지 에너지와 함께 모션 양 (예를 들어, 모션 벡터 또는 MV) 을 고려하여 2 개의 연속 프레임 간의 차이 측정을 제공한다. 또한, 프레임은 양방향 또는 단방향 예측 효율의 측정을 제공한다.

[0222] 모션 보상된 디인터레이싱을 수행할 수도 있는 전처리기로부터 수신된 모션 정보에 기초하는 프레임 차이 메트릭의 일 예는 다음과 같다. 디인터레이서가 양방향 모션 추정을 수행하므로, 양방향 모션 벡터와 SAD 정보를 이용할 수 있게 된다. 각 매크로블록에 대한 SAD_MV 에 의해 표현된 프레임 차이는 다음과 같이 구할 수 있다.

수학식 29

[0223]
$$SAD_MV = \log_{10} [SAD * \exp (-\min(1, MV))]$$

[0224] 여기서, $MV = \text{Square_root}(MV_x^2 + MV_y^2)$, $SAD = \min(SAD_N, SAD_P)$, 여기서, SAD_N 은 역방향 기준 프레임으로부터 계산된 SAD 이고, SAD_P 는 순방향 기준 프레임으로부터 계산된 SAD 이다.

[0225] 프레임 차이를 추정하는 다른 접근법은 수학식 6 내지 수학식 8 을 참조하여 위에 설명되어 있다. SAD 비율

(또는 콘트라스트비) γ 는 수학식 6 에서 상술한 것과 같이 계산될 수 있다. 또한, 모든 프레임의 휘도 히스토그램이 결정될 수 있고, 히스토그램 차이 λ 는 수학식 7 로부터 계산된다. 프레임 차이 D 는 수학식 8 에 도시된 것과 같이 계산될 수 있다.

[0226] 일 예시적인 예에서, 콘트라스트비와 프레임 차이 메트릭은 다음 방식으로 이용되어 비디오 콘텐츠 분류를 얻음으로써, 주어진 비디오 시퀀스의 특징을 신뢰성 있게 예측할 수 있다. 본원에는 인코더 (228) 에서 일어나는 것으로서 설명되어 있지만, 전처리기 (226) 는 콘텐츠 분류 (또는 다른 콘텐츠 정보) 를 결정하고, 메타데이터를 통해 콘텐츠 분류를 인코더 (228) 에 전달하도록 구성될 수도 있다. 아래 예에서 설명된 프로세스는 R-D 곡선 기반 분석으로부터 얻어진 분류와 유사하게, 콘텐츠를 8 개의 가능한 클래스로 분류한다. 분류 프로세스는 각 슈퍼프레임에 대해서 장면 복잡도와 그 슈퍼프레임 내의 장면 전환 발생 횟수에 따라 0 과 1 사이의 범위에 있는 값을 출력한다. 전처리기 내의 콘텐츠 분류 모듈은 각 슈퍼프레임에 대해 다음 (1) 내지 (5) 의 단계를 실행하여, 프레임 콘트라스트와 프레임 차이 값으로부터 콘텐츠 분류 메트릭을 얻을 수 있다.

- [0227] 1. 매크로블록 콘트라스트 값으로부터 평균 프레임 콘트라스트와 프레임 콘트라스트 편차 계산
- [0228] 2. 시뮬레이션으로부터 얻어진 값 (이 값은 각각 40 과 5 임) 을 사용하여 프레임 콘트라스트와 프레임 차이 값 정규화
- [0229] 3. 예를 들어, 다음과 같은 일반화된 수학식을 사용하여 콘텐츠 분류 메트릭 계산:

수학식 30

$$\begin{aligned} \text{CCMetric} &= \text{CCW1} * \text{I_Frame_Contrast_Mean} + \text{CCW2} \\ & * \text{Frame_Difference_Mean} - \\ & \text{CCW3} * \text{I_Contrast_Deviation}^2 * \exp(\text{CCW4} * \text{Frame_Difference_Deviation}^2) \end{aligned}$$

- [0230]
- [0231] 여기서, CCW1, CCW2, CCW3 및 CCW4 는 가중 계수이다. 본 예에서, CCW1 에 대해 0.2 의 값이 선택되고, CCW2 에 대해 0.9 의 값이 선택되고, CCW3 에 대해 0.1 의 값이 선택되며, CCW4 의 값에 대해 -0.00009 의 값이 선택된다.
- [0232] 4. 슈퍼프레임 내의 장면 전환 횟수 결정. 일반적으로, 슈퍼프레임은 특정 기간에 표시될 수 있는 영상 또는 프레임 그룹을 지칭한다. 통상, 그 기간은 일 초이다. 몇몇 양태에서, 슈퍼프레임은 30 개의 프레임 (30/fps 비디오의 경우) 을 포함한다. 다른 양태에서, 슈퍼프레임은 24 개의 프레임을 포함한다 (24/fps 비디오). 장면 전환 횟수에 따라, 다음 경우 중 하나의 경우가 실행된다.
- [0233] (a) 장면 전환 없음 : 슈퍼프레임에서 어떤 장면 전환도 없는 경우에, 메트릭은 다음 수학식에 표현된 것과 같은 프레임 차이 값에만 전적으로 의존한다.

수학식 31

$$\begin{aligned} \text{CCMetric} &= (\text{CCW2} + (\text{CCW1}/2)) * \text{Frame_Difference_Mean} - (\text{CCW3} - \\ & (\text{CCW1}/2)) * 1 * \\ & \exp(-\text{CCW4} * \text{Frame_Difference_Deviation}^2) \end{aligned}$$

- [0234]
- [0235] (b) 단일 장면 전환 : 슈퍼프레임에서 단일 장면 전환이 관찰되는 경우에, 아래에 표현된 것과 같이, 디폴트 수학식을 사용하여 메트릭을 계산할 것이다:

수학식 32

$$\begin{aligned} \text{CCMetric} &= \text{CCW1} * \text{I_Frame_Contrast_Mean} + \text{CCW2} \\ &* \text{Frame_Difference_Mean} - \\ &\text{CCW3} * \text{I_Contrast_Deviation}^2 * \exp(\text{CCW4} * \text{Frame_Difference_Deviation}^2) \end{aligned}$$

[0236]

[0237]

(c) 2 개의 장면 전환 : 주어진 슈퍼프레임에서 최대 2 개의 장면 전환이 존재하는 것으로 관찰되는 경우에는, 다음 수학식에 표현된 바와 같이, 첫 번째 슈퍼프레임이 나중 슈퍼프레임보다 어쨌든 빨리 리프레시될 것이므로, 마지막 슈퍼프레임에는 첫 번째 슈퍼프레임보다 큰 가중치가 주어진다 :

수학식 33

$$\begin{aligned} \text{CCMetric} &= 0.1 * \text{I_Frame_Contrast_Mean1} + \\ &\text{CCW1} * \text{I_Frame_Contrast_Mean2} + \\ &(\text{CCW2} - 0.1) * \text{Frame_Difference_Mean} - \text{CCW3} * \text{I_Contrast_Deviation1}^2 \\ &* \text{I_Contrast_Deviation2}^2 * \exp(\text{CCW4} * \text{Frame_Difference_Deviation}^2) \end{aligned}$$

[0238]

[0239]

(d) 3 개 이상의 장면 전환: 주어진 슈퍼프레임이 3 개보다 많은 (예를 들어, N 개의) I 프레임을 갖는 것으로 관찰되는 경우에는, 다음 수학식에 표현된 바와 같이, 마지막 I 프레임에는 더 큰 가중치가 주어지고, 모든 다른 I 프레임에는 0.05 의 가중치가 주어진다.

수학식 34

$$\begin{aligned} \text{CCMetric} &= 0.05 * \text{I_Frame_Contrast_Mean}_{(1...N-1)} + \text{CCW1} * \\ &\text{I_Frame_Contrast_Mean}_{(N)} + (\text{CCW2} - (0.05 * (N-1))) * \text{Frame_Difference_Mean} \\ &- \\ &\text{CCW3} * \text{I_Contrast_Deviation}_{(N)}^2 * \text{I_Contrast_Deviation}_{(1...N-1)}^2 \\ &* \exp(\text{CCW4} * \text{Frame_Difference_Deviation}^2) \end{aligned}$$

[0240]

[0241]

5. 프레임 차이 평균이 0.05 보다 작은 낮은 모션 변경의 경우에는 메트릭에 대해 보정을 사용할 수도 있다. 0.33 의 오프셋 (CCOFFSET) 은 CCMetric 에 더해질 것이다.

[0242]

콘텐츠 분류 모듈 (712) 은 D_{csat} 값, 모션 벡터 및/또는 절대 차이의 합을 사용하여, 매크로블록에 대한 공간 복잡도를 나타내는 값 (또는 지정된 비디오 데이터량) 을 결정한다. 시간 복잡도는 프레임 차이 메트릭의 측정에 의해 결정된다 (2 개의 연속 프레임 간의 차이는 모션 벡터와 함께 모션량, 및 프레임 간의 절대 차이의 합을 고려함).

[0243]

몇몇 양태에서, 콘텐츠 분류 모듈 (712) 은 대역폭 맵을 생성하도록 구성될 수 있다. 예를 들어, 전처리기 (226) 가 대역폭 맵을 생성하지 않으면, 대역폭 맵 생성은 콘텐츠 분류 모듈 (712) 에 의해 수행될 수 있다.

[0244]

텍스처 및 모션 값 결정

[0245]

멀티미디어 데이터 내의 각 매크로블록에 대해, 콘텐츠 분류 모듈 (712) 은 공간 복잡도를 텍스처 값과 연관시키고, 시간 복잡도를 모션 값과 연관시킨다. 텍스처 값은 멀티미디어 데이터의 휘도 값에 관한 것으로서, 낮은 텍스처 값은 데이터의 인접 화소의 휘도 값의 변화가 작은 것을 나타내고, 높은 텍스처 값은 데이터의 인접 화소의 변화가 큰 것을 나타낸다. 일단 텍스처와 모션 값이 계산되면, 콘텐츠 분류 모듈 (712) 은 모션과 텍스처 정보 모두를 고려함으로써 콘텐츠 분류를 결정한다. 콘텐츠 분류 모듈 (712) 은 분류된 비디오 데이터의 텍스처를 상대 텍스처 값, 예를 들어, "낮은" 텍스처, "중간" 텍스처, 또는 "높은" 텍스처와 연관시키고, 이는 일반적으로 매크로블록의 휘도 값의 복잡도를 나타낸다. 또한, 콘텐츠 분류 모듈 (712) 은 분류된 비디오 데이터에 대해 계산된 모션 값을 상대 모션 값, 예를 들어, "낮은" 모션, "중간" 모션, 또는 "높은" 모

선과 연관시키고, 이는 일반적으로 매크로블록의 모션량을 나타낸다. 다른 양태에서는, 더 많거나 적은 모션 및 텍스처 카테고리를 사용할 수 있다. 그러면, 콘텐츠 분류 메트릭은 연관된 텍스처 및 모션 값을 고려함으로써 결정된다.

[0246] 도 8은 어떻게 텍스처 및 모션 값이 콘텐츠 분류와 연관되는지를 도시한 분류 차트의 일 예를 도시한다. 당업자라면, 예를 들어, 룩업 테이블 또는 데이터베이스에서 그러한 분류 차트를 구현하는 다수의 방법을 알 수 있다. 분류 차트는 비디오 데이터 콘텐츠의 소정의 평가에 기초하여 생성된다. 비디오 데이터 분류를 결정하기 위해, (x 축에서) "낮은", "중간", 또는 "높은" 텍스처 값은 (y 축에서) "낮은", "중간", 또는 "높은" 모션 값과 상호 참조된다. 교차 블록에 표시된 콘텐츠 분류는 비디오 데이터에 할당된다. 예를 들어, "높은" 텍스처 값과 "중간" 모션 값은 "7"이라는 분류를 생성한다. 도 8은 본 예에서 8개의 상이한 콘텐츠 분류와 연관되는 상대 텍스처 및 모션 값의 여러 조합을 도시한다. 몇몇 다른 양태에서는, 더 많거나 적은 분류를 사용할 수 있다. 콘텐츠 분류의 예시적인 양태의 추가 설명은, 2006년 3월 10일 출원된 발명의 명칭이 "Content Classification for Multimedia Processing"인 동시 계속 중인 미국 특허 출원 번호 제 11/373,577 호에 개시되어 있고, 상기 출원은 본원의 양수인에게 양도되고 본원에서 명확히 참조로서 병합하고 있다.

[0247] 레이트 제어 비트 할당

[0248] 본원에서 상술한 바와 같이, 인코딩 알고리즘에서 멀티미디어 데이터 콘텐츠를 사용하여, 인식 가능한 비디오 품질을 일정하게 유지하면서 비트 관리를 효과적으로 개선할 수 있다. 예를 들어, 장면 전환 검출, 인코딩 비트 레이트 할당 제어 및 FRUC (frame rate up conversion)를 위한 알고리즘에서 분류 메트릭을 사용할 수 있다. 코덱 (COmpressor/DECompressor; CODEC) 시스템 및 디지털 신호 처리 알고리즘은 일반적으로 비디오 데이터 통신에서 사용되고, 대역폭을 절약하도록 구성될 수 있지만, 품질과 대역폭 절약 간에 절충이 존재한다. 최적 코덱은 비디오 품질을 최소로 저하시키면서 최대의 대역폭 절약을 제공한다.

[0249] 예시적인 일 예에서, 레이트 제어 비트 할당 모듈 (714)은 콘텐츠 분류를 사용하여, 비트 레이트 (예를 들어, 멀티미디어 데이터를 인코딩하는데 할당된 비트 수)를 결정하고, 인코더 (228)의 다른 프로세스 및 컴포넌트에 의한 사용을 위해 메모리 내에 비트 레이트를 저장한다. 비디오 데이터의 분류로부터 결정된 비트 레이트는 일관된 품질 레벨로 멀티미디어 데이터를 제공하면서 대역폭을 절약하는 것을 도울 수 있다. 일 양태에서는, 8개의 상이한 콘텐츠 분류 각각과 상이한 비트 레이트를 연관시킬 수 있고, 그 다음에, 그 비트 레이트를 사용하여 멀티미디어 데이터를 인코딩한다. 결과로서 생성된 효과는, 멀티미디어 데이터의 상이한 콘텐츠 분류에 인코딩을 위한 상이한 수의 비트가 할당되지만, 디스플레이 상에 표시될 때 인식된 품질이 유사하거나 일관된다는 것이다.

[0250] 일반적으로, 더 높은 콘텐츠 분류를 갖는 멀티미디어 데이터는 더 높은 레벨의 모션 및/또는 텍스처를 나타내고, 인코딩 시 더 많은 비트가 할당된다. 더 낮은 분류 (적은 텍스처와 모션을 나타냄)를 갖는 멀티미디어 데이터는 더 적은 비트가 할당된다. 특정 콘텐츠 분류를 갖는 멀티미디어 데이터의 경우, 비트 레이트는 멀티미디어 데이터를 보기 위한 선택된 타겟의 인식된 품질 레벨에 기초하여 결정될 수 있다. 멀티미디어 데이터 품질의 결정은 인간 뷰잉 (human viewing)과 멀티미디어 데이터의 그레이딩 (grading)에 의해 결정될 수 있다. 몇몇 다른 양태에서, 멀티미디어 데이터 품질의 추정은 예를 들어, 신호 대 잡음비 알고리즘을 사용하는 자동 테스트 시스템에 의해 행해질 수 있다. 일 양태에서, (예를 들어, 5개의) 표준 품질 레벨 세트 및 각각의 특정 품질 레벨을 달성하는데 필요한 대응 비트 레이트는 각각의 콘텐츠 분류의 멀티미디어 데이터에 대해 미리 결정된다. 품질 레벨 세트를 결정하기 위해, 특정 콘텐츠 분류의 멀티미디어 데이터는, 일정 비트 레이트를 사용한 인코딩 시, 멀티미디어 데이터의 시각적으로 인식된 멀티미디어 데이터 품질의 수치 표시를 제공하는 MOS (Mean Opinion Score)를 생성함으로써 평가될 수 있다. MOS는 1 내지 5의 범위에 있는 단일 숫자로서 표현될 수 있고, 여기서, 1은 가장 낮게 인식된 품질이고, 5는 가장 높게 인식된 품질이다. 다른 양태에서, MOS는 5개보다 많은 품질 레벨을 갖거나 5개보다 적은 품질 레벨을 가질 수 있고, 각 품질 레벨의 상이한 설명을 사용할 수 있다.

[0251] 멀티미디어 데이터 품질의 결정은 인간 뷰잉과 멀티미디어 데이터의 그레이딩에 의해 결정될 수 있다. 몇몇 다른 양태에서, 멀티미디어 데이터 품질의 추정은 예를 들어, 신호 대 잡음비 알고리즘을 사용하는 자동 테스트 시스템에 의해 행해질 수 있다. 일 양태에서, (예를 들어, 5개의) 표준 품질 레벨 세트 및 각각의 특정 품질 레벨을 달성하는데 필요한 대응 비트 레이트는 각 콘텐츠 분류의 멀티미디어 데이터에 대해 미리 결정된다.

[0252] 일정 콘텐츠 분류의 멀티미디어 데이터에 대한 비트 레이트와 시각적으로 인식된 품질 레벨 간의 관계를 이해하

는 것은 타깃 (예를 들어, 원하는) 품질 레벨을 선택함으로써 결정될 수 있다. 비트 레이트를 결정하는데 사용되는 타깃 품질 레벨은 미리 선택되거나, 사용자에게 의해 선택되거나, 자동 프로세스 또는 사용자에게 다른 프로세스로부터의 입력을 필요로 하는 반자동 프로세스를 통해 선택되거나, 또는 소정의 기준에 기초하여 인코딩 디바이스나 시스템에 의해 동적으로 선택될 수 있다. 타깃 품질 레벨은 예를 들어, 인코딩 애플리케이션 타입, 또는 멀티미디어 데이터를 수신하는 클라이언트 디바이스 타입에 기초하여 선택될 수 있다.

[0253] 도 7 에 도시된 예에서, 레이트 제어 비트 할당 모듈 (714) 은 콘텐츠 분류 모듈 (712)로부터 데이터를 수신하고, 또한 전처리기 (226)로부터 메타데이터를 직접 수신한다. 레이트 제어 비트 할당 모듈 (714)은 인코더 (228)의 제 1 패스 부분에 상주하고, 레이트 제어 미세 조정 모듈 (738)은 제 2 패스 부분 (706)에 상주한다. 이러한 2-패스 레이트 제어 양태는, 제 1 패스 (레이트 제어 비트 할당 모듈 (714))가 하나의 슈퍼프레임 룩어헤드 (look-ahead)를 사용하여 문맥 적응 비트 할당을 수행 (예를 들어, 256 kbps의 장기간 평균 비트 레이트 타깃팅)하며 피크 레이트를 제한하고, 제 2 패스 (레이트 제어 미세 조정 모듈 (738))가 2-계층 스케일러빌리티에 대해 제 1 패스 결과를 세분화하며 레이트 적응을 수행하도록 구성된다. 레이트 제어는 4개의 레벨에 대해 작용한다: (1) GOP 레벨 - I, P, B 및 F 프레임의 비트 분포를 GOP 내에서 균일하지 않게 제어함; (2) 슈퍼프레임 레벨 - 최대 슈퍼프레임 크기에 대한 하드 리미트 (hard limit)를 제어함; (3) 프레임 레벨 - 콘텐츠 정보 (예를 들어, 콘텐츠 분류)에 기초하는, 멀티미디어 데이터 프레임의 공간 및 시간 복잡도에 따라 비트 요건을 제어함; 및 (4) 매크로블록 레벨 - 콘텐츠 정보 (예를 들어, 콘텐츠 분류)에 기초하는, 공간 및 시간 복잡도 맵에 기초하여 매크로블록의 비트 할당을 제어함.

[0254] 도 10에는 레이트 제어 모듈 (714)의 동작의 예시적인 흐름도가 도시되어 있다. 도 10에 도시된 바와 같이, 프로세스 (1000)는 입력 블록 (1002)에서 시작한다. 레이트 제어 모듈 (714)은 여러 입력을 수신하지만, 그 모든 입력이 반드시 도 7에 도시되어 있는 것은 아니다. 예를 들어, 입력 정보는 전처리기 (226)로부터의 메타데이터, 타깃 비트레이트, 인코더 버퍼 크기 (또는, 등가물로서, 레이트 제어에 대한 최대 지연 시간), 초기 레이트 제어 지연 및 프레임 레이트 정보를 포함할 수 있다. 추가 입력 정보는, 예를 들어, (장면 전환 정보를 비롯한) GOP의 최대 슈퍼프레임 크기, 길이 및 P/B-프레임 분포, 필요한 베이스 및 인헨스먼트 계층의 배열, 장래 30 프레임 동안 GOP 내의 영상에 대한 복잡도-왜곡 메트릭을 비롯한, GOP (group of pictures) 레벨의 입력을 포함할 수 있다. 다른 입력 정보는, (콘텐츠 분류 모듈 (712)로부터 수신된) 현재 영상에 대한 복잡도-왜곡 맵, 양자화 파라미터 (QP), 및 (슬라이딩 윈도우에 맞는) 과거 30 프레임의 비트 브레이크다운을 비롯한, 영상 레벨의 입력을 포함한다. 끝으로, 매크로블록 (MB) 레벨의 입력 정보는, 예를 들어, 기준 영상에서 수집된 매크로블록 (MB)의 MAD (mean absolute difference), 및 (스킵 여부에 관계없이) 양자화 후의 매크로블록의 CBP (coded block pattern)를 포함한다.

[0255] 블록 (1002)에서의 입력 후에, 프로세스 (1000)는 비트스트림을 인코딩하기 위한 초기화 블록 (1004)으로 진행한다. 이와 동시에, 버퍼 초기화 (1006)가 수행된다. 다음으로, 블록 (1008)에 도시된 바와 같이 GOP가 초기화되고, GOP 비트 할당 (1010)이 초기화의 일부로서 수신된다. GOP 초기화 후에, 흐름은 블록 (1012)으로 진행하여, 슬라이스가 초기화된다. 이러한 초기화는 블록 (1014)에 도시된 바와 같이 헤더 비트의 갱신을 포함한다. 블록 (1004, 1008 및 1012)의 초기화가 수행된 후에, 블록 (1016)에 의해 도시된 바와 같이, 기본 유닛 또는 매크로블록 (MB)에 대한 레이트 제어 (RC)가 실행된다. 블록 (1016)에서 매크로블록의 레이트 제어 결정의 일부로서, 인코더 (228)에서 인터페이스를 통해 입력이 수신된다. 이들 입력은 매크로블록 (MB) 비트 할당 (1018), 2차 모델 파라미터의 갱신 (1020) 및 중앙값으로부터의 중앙값 절대 편차 ("MAD", 견고한 분산 추정) 파라미터 (1022)의 갱신을 포함할 수 있다. 그 다음에, 프로세스 (1000)는 하나의 영상을 인코딩한 후의 동작을 실행하는 블록 (1024)으로 진행한다. 블록 (1026)에 의해 도시된 바와 같이, 이러한 절차는 버퍼 파라미터의 갱신을 수신하는 것을 포함한다. 그 다음에, 프로세스 (1000)는, 출력 블록 (1028)으로 진행하고, 도 7에 도시된 바와 같이, 레이트 제어 모듈 (714)은 모드 결정 모듈 (715)에 의해 사용되는 각 매크로블록 (MB)에 대한 양자화 파라미터 (QP)를 출력한다.

[0256] 모션 추정

[0257] 모션 추정 모듈 (720)은 전처리기 (226)로부터 메타데이터와 미가공 비디오의 입력을 수신하고, 블록 크기, 모션 벡터, 왜곡 메트릭 및 기준 프레임 식별자를 포함할 수 있는 출력을 모드 결정 모듈 (715)에 제공한다.

도 11은 모션 추정 모듈 (720)의 예시적인 동작을 도시한다. 도시된 바와 같이, 프로세스 (1100)는 입력 (1102)으로 시작한다. 프레임 레벨에서, 모션 추정 모듈 (720)은 기준 프레임 ID 및 모션 벡터의 입력을 수신한다. 매크로블록 레벨에서, 입력 (1102)은 입력 화소 및 기준 프레임 화소를 포함한다. 프로세스 (1100)는, 컬러 모션 추정 (ME) 및 모션 벡터 예측이 수행되는 단계 (1104)로 진행한다. 이러한

프로세스를 실행하기 위해, MPEG-2 모션 벡터, 루마 모션 벡터 (MV) (1106), 모션 벡터 평활화 (1108) 및 비인과적 모션 벡터 (1110) 를 비롯한 여러 입력이 수신된다. 다음으로, 프로세스 (1100) 는, 16 진 또는 다이아몬드 검색 방법과 같은, 모션 벡터 검색 알고리즘 또는 방법론이 수행되는 블록 (1112) 으로 진행한다. 블록 (1112) 에서의 처리에 대한 입력은 블록 (1114) 에 의해 도시된 바와 같이, SAD (sum of absolute difference), SSD (sum of squared difference) 및/또는 다른 메트릭을 포함할 수도 있다. 일단 모션 벡터 검색이 수행되면, 프로세스 (1100) 는 종료 블록 (1116) 으로 진행하여, 종료 처리가 수행된다. 그 다음에, 프로세스 (100) 는 출력 블록 (1118) 에서 종료하고, 이 출력 블록 (1118) 은 블록 크기, 모션 벡터 (MV), 왜곡 메트릭 및 기준 프레임 식별자의 출력을 생성한다.

[0258] 베이스 및 인헨스먼트 계층에 대한 스케일러빌리티 R-D

[0259] 도 13 은 스케일러빌리티 R-D 모듈 (718) 에 의해 수행될 수 있는 스케일러빌리티 프로세스 (1300) 의 예시적인 흐름도를 도시한다. 프로세스 (1300) 는 시작 블록 (1302) 에서 시작하고, 블록 (1304) 으로 진행하여, 스케일러빌리티 R-D 모듈 (718) 이 모션 추정 모듈 (720) 로부터 입력을 수신하고 모션 추정을 수행한다. 모션 추정은 블록 (1306) 에 의해 표시된 바와 같이, 베이스 계층 기준 프레임, 인헨스먼트 계층 기준 프레임 및 코딩될 원래 프레임의 입력에 의존한다. 그러한 정보는 GOP 분할기 (612) 에 의해 계산될 수 있고, 예를 들어, 메타데이터를 통해 스케일러빌리티 R-D 모듈 (718) 에 전달될 수 있다. 프로세스 (1300) 는 블록 (1308) 으로 진행하여, 베이스 계층 데이터와 인헨스먼트 계층 데이터의 스케일러빌리티 정보를 결정한다. 그 다음에, 블록 (1310) 에 도시된 바와 같이, 베이스 계층 인코딩이 수행되고, 그 다음에, 블록 (1312) 에서 인헨스먼트 계층 인코딩이 수행된다. 블록 (1314) 에 도시된 바와 같이, 인헨스먼트 계층의 인코딩은 층간 예측에 대한 베이스 계층 코딩 결과를 입력으로서 사용하므로, 베이스 계층 인코딩 후에 임의로 수행된다. 이는, 발명의 명칭이 "SCALABLE VIDEO CODING WITH TWO LAYER ENCODING AND SINGLE LAYER DECODING" 인 동시 계속 중인 미국 특허 출원 [Attorney docket/ref. no. 050078] 에 더 설명되어 있다. 인코딩이 완료된 후에, 프로세스 (1300) 는 블록 (1316) 에서 종료한다.

[0260] 슬라이스/매크로블록 순서화

[0261] 또한, 제 1 패스 부분 (702) 은 슬라이스/매크로블록 순서화 모듈 (722) 을 포함하고, 이 슬라이스/매크로블록 순서화 모듈 (722) 은 제 2 패스 부분 내의 여러 내성 모듈 (740) 로부터 입력을 수신하고, 슬라이스 정렬 정보를 모드 결정 모듈 (715) 에 제공한다. 슬라이스는 독립적으로 디코딩 가능하게 (엔트로피 디코딩) 코딩된 비디오 데이터의 청크이다. 액세스 유닛 (AU) 은 정확히 하나의 프라이머리 코딩된 영상을 항상 포함하는 NAL 유닛 세트를 각각 포함한 코딩된 비디오 프레임이다. 프라이머리 코딩된 영상 외에도, 액세스 유닛은 하나 이상의 중복 코딩된 영상 또는 코딩된 영상의 슬라이스 또는 슬라이스 데이터 분할을 포함하지 않는 다른 NAL 유닛을 포함할 수도 있다. 액세스 유닛의 디코딩은 디코딩된 영상을 항상 생성한다.

[0262] 프레임은 최고 시간 다이버시티를 제공하는 물리 계층 패킷 (TDM 캡슐로 지칭됨) 의 시분할 다중화 블록일 수 있다. 슈퍼프레임은 하나의 시간 단위 (예를 들어, 1 초) 에 대응하고, 4 개의 프레임을 포함한다. 슬라이스 및 AU 경계를 시간 영역 내의 프레임 경계와 정렬시킴으로써, 손상된 데이터를 가장 효율적으로 분리 및 한정하게 된다. 딥 페이드 동안에, TDM 캡슐 내의 연속 데이터 대부분은 에러에 의해 영향을 받는다. 시간 다이버시티 때문에, 나머지 TDM 캡슐은 손상되지 않을 확률이 높다. 손상되지 않은 데이터를 이용하여, 손실된 데이터를 복원하고, 영향을 받은 TDM 캡슐로부터 손실된 데이터를 은폐할 수 있다. 유사한 논리는, 데이터 심벌이 변조하는 주파수 서브캐리어 내의 분리를 통해 주파수 다이버시티를 달성하는 주파수 영역 다중화 (FDM) 에 적용된다. 또한, 유사 논리는 (송신기 및 수신기 안테나에서의 분리를 통하는) 공간 다이버시티와, 무선 네트워크에서 종종 적용되는 다른 형태의 다이버시티에 적용된다.

[0263] 슬라이스 및 AU 를 프레임에 정렬시키기 위해서는, 외부 코드 (FEC) 코드 블록 생성과 MAC 계층 캡슐화도 또한 정렬되어야 한다. 도 20 은 코딩된 비디오데이터 또는 비디오 스트림을 슬라이스와 AU 로 구성한 것을 도시한다. 코딩된 비디오는 예를 들어, 계층적 비디오 코딩이 적용되는 베이스 계층 비트스트림과 인헨스먼트 계층 비트스트림과 같은 하나 이상의 비트스트림으로 구성될 수도 있다.

[0264] 도 20 에 도시된 바와 같이, 비디오 비트스트림은 프레임 1 (2005), 프레임 3 (2010) 및 프레임 M (2015) 단위로 AU 를 포함한다. 도시된 바와 같이, AU 는 슬라이스 1 (2020), 슬라이스 2 (2025) 및 슬라이스 N (2030) 단위로 데이터 슬라이스를 포함한다. 각각의 슬라이스 시작은 시작 코드에 의해 식별되고, 네트워크 적용에 제공된다. 일반적으로, I-프레임 또는 인트라 코딩된 AU 는 크고, 그 다음에 P-프레임 또는 순방향 예측 프레임이 오고, 그 다음에 B-프레임이 온다. AU 를 다수의 슬라이스로 코딩함으로써, 슬라이스 전체에

서 공간 예측이 제한되며 슬라이스 헤더 또한 오버헤드에 기여하기 때문에, 코딩된 비트레이트의 관점에서 상당한 오버헤드 비용을 발생시킨다. 슬라이스 경계가 재동기화 지점이기 때문에, PLP 가 손상된 이후에, 연속적인 물리 계층 패킷을 슬라이스에 제한함으로써, 에러를 제어하고, PLP 내의 슬라이스에 에러가 제한되지만, PLP 가 다수의 슬라이스 또는 다수의 슬라이스의 부분을 포함하면, 에러가 PLP 내의 모든 슬라이스 또는 슬라이스 부분에 영향을 미칠 것이다.

[0265] 통상 I-프레임이 예를 들어, 수십 kbits 정도로 크기 때문에, 다수의 슬라이스로 인한 오버헤드는 전체 I-프레임 크기 또는 전체 비트레이트의 큰 비율이 아니다. 또한, 인트라 코딩된 AU 내에 더 많은 슬라이스를 가짐으로써, 더 양호하고 더 빈번한 재동기화를 가능하게 하고, 더 효율적인 공간 에러 은폐를 가능하게 한다. 또한, P 및 B-프레임이 I-프레임에서 예측되기 때문에, I-프레임은 비디오 비트스트림에서 가장 중요한 정보를 전달할 수 있다. 또한, I-프레임은 채널 획득을 위한 랜덤 액세스 포인트로서 기능을 한다.

[0266] 이하, 도 21 을 참조하면, I-프레임을 프레임 경계와 정렬시키고, 또한 I AU 를 갖는 슬라이스를 프레임 경계와 신중히 정렬시킴으로써, 가장 효율적인 에러 제어, 에러 보호 (프레임 1 (2105) 에 속한 하나의 슬라이스가 손실되더라도, 프레임 2 (2110) 가 프레임 1 (2105) 로부터 상당한 시간이 분리되므로 프레임 2 (2110) 에 속하는 슬라이스가 손상되지 않을 확률이 높기 때문임) 을 가능하게 하고, 재동기화 및 에러 은폐를 통해 에러 내성을 수행할 수 있다.

[0267] 통상 P-프레임이 몇 kbits 정도의 크기를 갖기 때문에, P-프레임의 슬라이스 및 정수 개의 P-프레임을 프레임 경계에 정렬시킴으로써, (I-프레임과 유사한 이유로) 치명적인 효율 손실 없이 에러 내성을 가능하게 한다. 시간 에러 은폐는 그러한 양태에서 사용될 수 있다. 다른 방법으로는, 연속적인 P-프레임을 분산시켜 상이한 프레임에서 도착하게 함으로써, P-프레임 간에 시간 다이버시티를 추가하는데, 이는 시간 은폐가 미리 채구성된 I 또는 P 프레임으로부터의 모션 벡터 및 데이터에 기초하기 때문이다. B-프레임은 매우 작은 (수백 비트) 프레임부터 매우 큰 (몇천 비트) 프레임까지 존재할 수 있다. 따라서, 치명적인 효율 손실 없이 에러 내성을 달성하기 위해서는, 정수 개의 B-프레임을 프레임 경계에 정렬시키는 것이 바람직하다.

[0268] 모드 결정 모듈

[0269] 도 12 는 모드 결정 모듈 (715) 의 동작의 몇몇 예를 도시한다. 도시된 예와 같이, 프로세스 (1200) 는 입력 블록 (1202) 에서 시작한다. 일 예시적인 예에서, 모드 결정 모듈 (715) 에 대한 여러 정보 입력은 슬라이스 타입, Intra 4x4cost, Intra 16x16cost, IntraUV 8x8cost, IntraY 16x16 Mode, IntraUV Mode, MVD (motion vector data), QP (quantization parameters), SpPredMB4x4Y, SpPredMB16x16Y, SpPredMB8x8u, SpPredMB8x8v, 레이트 왜곡 플래그, 미가공 YMB 화소, 미가공 UMB 화소 및 미가공 VMB 화소를 포함한다. 그 다음에, 프로세스 (1200) 는 인코딩 초기화 블록 (1204) 로 진행하여, 블록 (1206) 에 의해 표시된 것과 같이 인코더 초기화를 명령하는 입력 신호 또는 인터페이스에 의해 개시될 수 있다. 그 초기화는 허용된 모드 설정 (skip, direct 포함), 모드 가중치 설정 (필요에 따라, 디폴트는 모든 노드에 대해 동일한 가중치일 것임), 및 버퍼 설정을 포함할 수 있다. 초기화 후에, 프로세스 (1200) 는 블록 (1028) 으로 진행하여, 각각의 허용된 모드에 대한 매크로블록 (MB) 모드 비용 계산, 가중 인자로 각각의 MB 모드 비용 가중화 및 최소 MB 모드 비용 모드 선택을 비롯한, 모드 결정을 위한 메인 처리가 수행된다. 블록 (1210 및 1212) 에 도시된 바와 같이, 이들 동작과 관련된 입력은 모션 추정 (예를 들어, MVD 및 예측), 및 공간 예측 (예를 들어, 모든 인트라 비용 및 예측) 을 포함한다. 블록 (1214) 에서 엔트로피 인코딩이 모드 결정 모듈 (715) 과 인터페이스하여, 그 중에서 특히 압축 레이트를 향상시킨다. 프로세스 (1200) 는 블록 (1216) 으로 진행하여, 버퍼를 갱신함으로써, 인코더 제 2 패스 부분 (706) 으로 정보를 전달한다. 끝으로, 프로세스 (1200) 는 블록 (1218) 으로 진행하여, 인코더 제 2 패스 부분 (706) 으로 "최적" 인코딩 모드를 전달할 수 있다.

[0270] 인코더 제 2 패스 부분

[0271] 다시 도 7 을 참조하면, 인코더 (228) 의 제 2 패스 부분 (706) 은 인코딩의 제 2 패스를 수행하기 위한 제 2 패스 인코더 모듈 (232) 을 포함한다. 제 2 패스 인코더 (232) 는 모드 결정 모듈 (715) 로부터의 출력을 수신한다. 제 2 패스 인코더 (232) 는 MC/변환 양자화 모듈 (726) 및 ZZ (Zig Zag)/엔트로피 인코더 (728) 를 포함한다. 제 2 패스 인코더 (232) 의 결과는 스케일러빌리티 모듈 (730) 과 비트스트림 패킹 모듈 (731) 에 출력되고, 또한, (도 2 에 도시된) 동기화 계층 (240) 을 통해 트랜스코더 (200) 가 송신하기 위해 인코딩된 베이스 및 인헨스먼트 계층에 출력된다. 도 2 에 도시된 바와 같이, 제 2 패스 인코더 (232) 와 제 인코더 (234) 로부터의 베이스 및 인헨스먼트 계층은 동기화 계층 (240) 에 의해 모아져, 베이스 및 인헨스먼트 계층, 데이터 PES (244; 예를 들어, CC 및 다른 텍스트 데이터) 및 오디오 PES (246) 를 비롯한 패킷화된 PES

(242) 로 된다는 것에 주목하자. 오디오 인코더 (236) 는 디코딩된 오디오 정보 (218) 를 수신하고, 또한, 그 정보를 인코딩하여 인코딩된 정보 (238) 를 동기화 계층 (240) 에 출력한다는 것에 주목하자.

[0272] 재인코더

[0273] 다시 도 7 을 참조하면, 인코더 제 2 패스 부분 (706) 은 또한 도 2 의 재인코더 (234) 에 대응하는 재인코더 (234) 를 포함한다. 또한, 재인코더 (234) 는 제 1 패스 부분 (702) 의 출력을 수신하고, MC/변환 양자화 (726) 및 ZZ/엔트로피 코딩 (728) 부분을 포함한다. 또한, 스케일러빌리티 모듈 (730) 은 재인코더 (234) 에 출력한다. 재인코더 (234) 는 결과로서 생성된 베이스 및 인헨스먼트 계층을 재인코딩으로부터 비트스트림 패킹 모듈 (731) 로 출력하여, 동기화 장치 (예를 들어, 도 2 에 도시된 동기화 계층 (240)) 로 송신한다. 또한, 도 7 의 인코더 (228) 예는, 비트스트림 패킹 피드백을 제 2 패스 인코더 (232) 내의 MC/변환 양자화 모듈 (234) 과 재인코더 (234) 내의 ZZ/엔트로피 모듈 (736) 모두에 제공하여, (예를 들어, 압축 효율을 증가시키는 것과 같이) 제 2 패스 인코딩을 조정하는 것을 돕는 레이트 제어 미세 조정 모듈 (738) 을 포함한다.

[0274] 에러 내성 모듈

[0275] 또한, 도 7 에 도시된 인코더 (228) 는 제 2 패스 부분 (706) 에 에러 내성 모듈 (740) 을 포함한다. 에러 내성 모듈 (740) 은 비트스트림 패킹 모듈 (730) 및 슬라이스/MB 순서화 모듈 (722) 과 통신한다. 에러 내성 모듈 (740) 은 전처리 (228) 로부터 메타데이터를 수신하고, 예를 들어, 슬라이스 및 액세스 유닛을 프레임 경계에 정렬시키는 방식, 예측 계층 구성 및 적응 인트라 리프레시와 같은 에러 내성 방식을 선택한다. 에러 내성 방식의 선택은 메타데이터에서 수신된 정보에 기초하여 행해질 수 있고, 또는 비트스트림 패킹 모듈 (731) 및 슬라이스/MB 순서화 모듈 (722) 로부터 에러 내성 모듈에 전달된 정보로부터 행해질 수 있다. 에러 내성 모듈 (740) 은 제 1 패스 부분 (702) 내의 슬라이스/매크로블록 (MB) 순서화 모듈에 정보를 제공하여, 선택된 에러 내성 프로세스를 구현한다. 에러가 발생하기 쉬운 환경을 통한 비디오 송신은, 뷰잉 사용자에게 더 명확하며 에러로 덜 채워진 데이터를 표시할 수 있는 에러 내성 전략 및 알고리즘을 사용할 수 있다. 아래의 에러 내성 설명은 현재 또는 장래 애플리케이션, 전송 및 물리 계층 또는 다른 기술의 임의의 하나 또는 그 조합에 적용될 수 있다. 효과적인 에러 견고성 알고리즘은 낮은 대기 시간 및 높은 처리량과 같은 통신 시스템의 원하는 특성과 관련하여 OSI 계층 간의 에러 감응 특성과 에러 보호 능력의 이해를 완전하게 한다. 에러 내성 처리는 예를 들어, 멀티미디어 데이터의 콘텐츠 분류와 같은 멀티미디어 데이터의 콘텐츠 정보에 기초할 수 있다. 주요 이점 중 하나는 페이딩 및 다경로 채널 에러로부터의 복구성이다. 아래에 설명되는 에러 내성 접근법은 특히, 인코더 (228) 내에 (예를 들어, 더욱 상세하게는, 에러 내성 모듈 (740) 및 슬라이스/MB 순서화 모듈 (722) 내에) 통합될 수 있으며, 에러가 발생하기 쉬운 환경에서의 데이터 통신으로 일반적으로 확장될 수 있는 프로세스에 관한 것이다.

[0276] 에러 내성

[0277] 예측 기반 하이브리드 압축 시스템의 경우에, 인트라 코딩된 프레임은 임의의 시간 예측 없이 독립하여 코딩된다. 인트라 코딩된 프레임은 과거 프레임 (P-프레임) 과 장래 프레임 (B-프레임) 으로부터 임시로 예측될 수 있다. 최적 예측기는 (하나 이상의) 기준 프레임에서 검색 프로세스를 통해 식별될 수 있고, SAD 와 같은 왜곡 측정을 사용하여 최적 정합을 식별한다. 현재 프레임의 예측 코딩된 영역은 예를 들어, 세그먼테이션을 통해 물체로서 식별되는 화소 그룹 또는 가변 크기 및 형상 (16x16, 32x32, 8x4 등) 을 갖는 블록일 수 있다.

[0278] 통상, 시간 예측은 다수의 프레임 (예를 들어, 10 개 내지 수십 개 프레임) 에 대해 확장되고, 프레임이 I-프레임으로서 코딩될 때 종료하며, GOP 는 통상 I-프레임 주파수에 의해 규정된다. 최대 코딩 효율의 경우에, GOP 는 장면이고, 예를 들어, GOP 경계는 장면 경계와 정렬되고, 장면 전환 프레임은 I-프레임으로서 코딩된다. 낮은 모션 시퀀스는 비교적 정적인 배경을 포함하고, 모션은 일반적으로 전경 물체에 제한된다. 그러한 낮은 모션 시퀀스의 콘텐츠 예는 뉴스 및 기상 예보 프로그램을 포함하고, 가장 많이 시청되는 콘텐츠의 30% 이상이 이러한 성질을 갖는다. 낮은 모션 시퀀스에서, 대부분의 영역은 인트라 코딩되고, 예측된 프레임은 중간 예측된 프레임을 통해 다시 I-프레임을 참조한다.

[0279] 도 22 를 참조하면, I-프레임 내의 인트라 코딩된 블록 (2205) 은 코딩된 프레임 (또는 AU) P1 내의 인트라 코딩된 블록 (2210) 에 대한 예측기이다. 본 예에서, 이들 블록 영역은 배경의 정지 부분이다. 연속적인 시간 예측을 통해, 에러에 대한 인트라 코딩된 블록 (2205) 의 감도가 상승하는데, 그 이유는 인트라 코딩된 블록 (2205) 이 "양호한" 예측기이기 때문이다 (또한, 이는 그 "중요성" 이 더 높다는 것을 의미함). 또한, 인트라

라 코딩된 블록 (2205) 은, 예측 체인으로 지칭되는 이러한 시간 예측 체인에 의해, (도 22 의 예에서 장면 지속 기간 동안에) 디스플레이에서 더 길게 계속된다.

[0280] 예측 계층 구성은 부모가 위 (인트라 코딩된 블록 (2205)) 에 있으며 자식이 아래에 있는 이러한 중요성 레벨 또는 영속성 측정에 기초하여 생성된 블록 트리로서 규정된다. P1 에서 인트라 코딩된 블록 (2215) 은 계층 구성의 제 2 레벨에 있다는 것에 주목하자. 리프는 예측 체인을 종료하는 블록이다.

[0281] 예측 계층 구성은 콘텐츠 타입 (예를 들어, 단지 뉴스뿐만 아니라 음악과 스포츠) 에 관계없이 비디오 시퀀스에 대해 생성될 수 있고, 예측 기반 비디오 (및 데이터) 압축에 일반적으로 적용 가능하다 (이는 본원에 설명된 모든 양태에 적용 가능함). 일단 예측 계층 구성이 확립되면, 아래에 설명되는 적응 인트라 리프레시와 같은 예러 내성 알고리즘은 더 효과적으로 적용될 수 있다. 중요성 측정은 예를 들어, 은폐 동작을 통해 예러로부터 주어진 블록의 복구성에 기초할 수 있고, 적응 인트라 리프레시를 적용하여 예러에 대한 코딩된 비트스트림의 복원성을 향상시킬 수 있다. 중요성 추정, 블록이 예측기 (또한, 영속성 메트릭으로 지칭됨) 로서 사용되는 횟수에 기초할 수 있다. 또한, 영속성 메트릭을 사용하여 예측 예러 전파를 저지함으로써 코딩 효율을 개선한다. 또한, 영속성 메트릭은 중요성이 더 높아진 블록에 대한 비트 할당을 증가시킨다.

[0282] 적응 인트라 리프레시

[0283] 적응 인트라 리프레시는 멀티미디어 데이터의 콘텐츠 정보에 기초할 수 있는 예러 내성 기술이다. 인트라 리프레시 프로세스에서는, 몇몇 MB 를 인트라 코딩하지만, 표준 R-D 최적화는 몇몇 MB 가 인트라 코딩된 MB 이어야 한다는 것을 요구할 것이다. AIR 는 모션 가중된 인트라 리프레시를 사용하여 P 또는 B-프레임에서 인트라 코딩된 MB 를 도입한다. 베이스 계층에 포함된 이들 인트라 코딩된 MB 는 QP_b 또는 QP_e 를 사용하여 인코딩될 수 있다. QP_e 가 베이스 계층에 사용되면, 인헨스먼트 계층에서는 어떤 세분화도 행해지지

않아야 한다. QP_b 가 베이스 계층에 사용되면, 세분화가 적절할 수도 있지만, 그렇지 않은 경우에는, 인헨스먼트 계층에서, 품질 저하를 인식할 수 있을 것이다. 코딩 효율 면에서 인트라 코딩이 인트라 코딩보다 효율적이기 때문에, 인헨스먼트 계층에서의 이들 세분화는 인트라 코딩될 것이다. 이러한 방법으로, 베이스 계층 계수는 인헨스먼트 계층에 사용되지 않을 것이고, 그 품질은 새로운 연산을 도입함 없이 인헨스먼트 계층에서 개선된다.

[0284] 몇몇 양태에서, 적응 인트라 리프레시는 모션 가중된 방식 대신에, 또는 모션 가중된 방식에 더하여 멀티미디어 데이터의 콘텐츠 정보 (예를 들어, 콘텐츠 분류) 에 기초할 수 있다. 예를 들어, 콘텐츠 분류가 비교적 높은 경우에는 (예를 들어, 장면이 높은 공간 및 시간 복잡도를 갖는 경우에는), 적응 인트라 리프레시가 P 또는 B-프레임 내에 비교적 많은 인트라 코딩된 MB 를 도입할 수 있다. 다른 방법으로는, 콘텐츠 분류가 비교적 낮은 경우에는 (낮은 공간 및/또는 시간 복잡도를 갖는 덜 동적인 장면을 나타내는 경우에는), 적응 인트라 리프레시가 P 및 B-프레임에 인트라 코딩된 더 적은 MB 를 도입할 수 있다. 예러 내성을 개선하기 위한 그러한 메트릭 및 방법은 무선 멀티미디어 통신 환경에서 적용될 수 있을 뿐만 아니라, 통상적으로 (예를 들어, 그래픽 렌더링에서) 데이터 압축 및 멀티미디어 처리에서도 적용될 수 있다.

[0285] 채널 변경 프레임

[0286] 본원에서 규정되는 것과 같은 채널 변경 프레임 (CSF) 은 고속 채널 획득 및 이에 따른 방송 멀티플렉스에서 스트림 간의 고속 채널 변경을 위해 방송 스트림 내의 적당한 위치에 삽입된 랜덤 액세스 프레임을 설명하는 광의의 용어이다. 채널 변경 프레임은 또한 예러 견고성을 증가시키는데, 그 이유는, 프라이머리 프레임이 잘못 송신되는 경우에 사용될 수 있는 중복 데이터를 채널 변경 프레임이 제공하기 때문이다. 통상, I-프레임 또는 H.264 내의 순차 디코더 리프레시 프레임과 같은 순차 I-프레임은 랜덤 액세스 포인트로서 기능을 한다. 그러나, 빈번한 I-프레임으로 인해 (또는, 짧은 GOP (이는 장면 지속 기간보다 짧음) 로 인해), 압축 효율이 상당히 떨어진다. 예러 내성을 위해 인트라 코딩 블록이 필요할 수도 있기 때문에, 예측 계층 구성을 통해 랜덤 액세스와 예러 내성을 효과적으로 결합하여, 예러에 대한 견고성을 증가시키면서 코딩 효율을 향상시킬 수 있다.

[0287] 랜덤 액세스 스위칭 및 예러 견고성의 개선은 동시에 달성될 수 있고, 콘텐츠 분류와 같은 콘텐츠 정보에 기초할 수 있다. 낮은 모션 시퀀스의 경우에는, 예측 체인이 길고, 슈퍼프레임 또는 장면을 재구성하는데 필요한 정보의 상당 부분이 장면 시작에서 일어난 I-프레임에 포함된다. 채널 예러는 버스티 (bursty) 하는 경

향이 있고, 페이드가 일어나고 FEC 및 채널 코딩이 실패하면, 상당한 나머지 에러가 발생하여, 은폐가 실패한다. 이는 특히 낮은 모션 (및 이에 따른 낮은 비트 레이트) 시퀀스의 경우에 심하게 되는데, 그 이유는, 코딩된 데이터 양이 비디오 비트스트림 내에서 양호한 시간 다이버시티를 제공하기에 충분하지 않기 때문이다. 또한 이들 낮은 모션 시퀀스가 재구성에 유용한 모든 비트를 렌더링하는 고도로 압축 가능한 시퀀스이기 때문이다. 높은 모션 시퀀스는 콘텐츠 성질 때문에 에러에 더 견고하다 - 모든 프레임에서 더 많은 새로운 정보는 독립적으로 디코딩 가능하며 원래 에러에 대한 복원성 더 뛰어난 코딩된 인트라 블록의 수를 증가시킨다.

예측 계층 구성에 기초한 적응 인트라 리프레시는 높은 모션 시퀀스에 대해 높은 성능을 달성하지만, 낮은 모션 시퀀스의 경우에는 성능 개선이 현저하지 않다. 따라서, 대부분의 I-프레임을 포함한 채널 변경 프레임은 낮은 모션 시퀀스에 대한 양호한 다이버시티 소스가 된다. 슈퍼프레임에서 에러가 발생하면, 예측으로 인해 손실된 정보를 복구하며 에러 내성이 달성되는 CSF로부터 연속적인 프레임에서의 디코딩이 시작한다.

[0288] 비교적 높은 콘텐츠 분류 (예를 들어, 6 내지 8) 를 갖는 시퀀스와 같은 높은 모션 시퀀스의 경우에, CSF 는 SF 에서 지속되는 블록으로 이루어질 수 있다 - 이들 블록은 양호한 예측기이다. CSF 의 모든 다른 영역은 코딩될 필요가 없는데, 그 이유는, 이들 영역이 짧은 예측 체인을 갖는 블록이기 때문이다 (이는, 이들 블록이 인트라 블록으로 종료된다는 것을 의미함). 따라서, CSF 는 에러 발생 시 예측으로 인해 손실된 정보로부터 계속 복구하는 기능을 한다. 낮은 모션 시퀀스에 대한 CSF 는 I-프레임의 크기와 동등하고, CSF 는 더 큰 양자화를 통해 더 낮은 비트 레이트에서 코딩될 수 있지만, 높은 모션 시퀀스에 대한 CSF 는 대응하는 I-프레임보다 훨씬 작다.

[0289] 예측 계층 구성에 기초한 에러 내성은 스케일러빌리티와 잘 작용할 수 있고, 매우 효율적인 계층적 코딩을 달성할 수 있다. 물리 계층 기술에서 계층형 변조를 지원하는 스케일러빌리티는 비디오 비트스트림을 특정 대역폭 비율로 데이터 분할하는 것을 필요로 할 수도 있다. 이들 비율이 항상 (예를 들어, 최소 오버헤드를 갖는) 최적 스케일러빌리티에 대한 이상적인 비율인 것은 아니다. 몇몇 양태에서는, 1:1 대역폭 비율을 갖는 2-계층 스케일러빌리티가 사용된다. 비디오 비트스트림을 같은 크기의 2-계층으로 분할하는 것이, 낮은 모션 시퀀스의 경우에는 효율적이지 않을 수도 있다. 낮은 모션 시퀀스의 경우에, 모든 헤더 및 메타데이터 정보를 포함한 베이스 계층은 인헨스먼트 계층보다 크다. 그러나, 낮은 모션 시퀀스에 대한 CSF 가 더 크기 때문에, 그 CSF 는 인헨스먼트 계층 내의 나머지 대역폭에서 잘 맞게 된다.

[0290] 높은 모션 시퀀스는 1:1 데이터 분할을 최소 오버헤드로 달성할 수 있도록 충분한 나머지 정보 (residual information) 를 갖는다. 또한, 그러한 시퀀스에 대한 채널 변경 프레임은 높은 모션 시퀀스의 경우에 훨씬 작아진다. 따라서, 예측 계층 구성에 기초한 에러 내성은 높은 모션 시퀀스의 경우에도 스케일러빌리티와 잘 작용할 수 있다. 중간 정도의 모션 클립에 대해 상술한 개념을 확장하는 것은, 이들 알고리즘의 설명에 기초하여 가능하고, 통상, 제안된 개념은 비디오 코딩에 적용된다.

[0291] 멀티플렉서

[0292] 몇몇 인코더 양태에서, 멀티플렉서는 인코더에 의해 생성된 다수의 멀티미디어 스트림을 인코딩하는데 사용될 수 있고, 방송용으로 인코딩된 비트를 준비하는데 사용될 수 있다. 예를 들어, 도 2 에 도시된 인코더 (228) 의 예시적인 양태에서, 동기화 계층 (240) 은 멀티플렉서를 포함한다. 멀티플렉서는 비트 레이트 할당 제어를 제공하도록 구현될 수도 있다. 추정된 복잡도가 멀티플렉서에 제공될 수 있고, 그 다음에, 이 멀티플렉서는 이들 비디오 채널에 대해 예상되는 인코딩 복잡도에 따라 멀티플렉싱된 비디오 채널 그룹에 대해 가용 대역폭을 할당할 수 있고, 또한, 멀티플렉싱된 비디오 스트림의 그룹에 대한 대역폭이 비교적 일정하더라도 특정 채널의 품질이 비교적 일정한 상태를 유지할 수도 있도록 해준다. 이로 인해, 채널 그룹 내의 채널이 비교적 일정한 비트 레이트와 가변 시각 품질을 갖기 보다는, 가변 비트 레이트와 비교적 일정한 시각 품질을 갖게 된다.

[0293] 도 18 은 다수의 멀티미디어 스트림 또는 채널 (1802) 을 인코딩하는 시스템을 도시한 블록도이다. 멀티미디어 스트림 (1802) 은 각각의 인코더 (1804) 에 의해 인코딩되고, 이 인코더 (1804) 는 멀티플렉서 (1806; MUX) 와 통신하고 있고, 또한, 이 멀티플렉서 (1806) 는 송신 매체 (1808) 와 통신하고 있다. 예를 들어, 멀티미디어 스트림 (1802) 은 뉴스 채널, 스포츠 채널, 영화 채널 등과 같은 여러 콘텐츠 채널에 대응할 수 있다. 인코더 (1804) 는 멀티미디어 스트림 (1802) 을 시스템에 지정된 인코딩 포맷으로 인코딩한다. 비디오 스트림의 인코딩 환경에서 설명되지만, 개시된 기술의 원리 및 이점은 예를 들어, 오디오 스트림을 비롯한 멀티미디어 스트림에 일반적으로 적용 가능하다. 인코딩된 멀티미디어 스트림은 멀티플렉서 (1806) 에 제공되고, 이 멀티플렉서 (1806) 는 여러 인코딩된 멀티미디어 스트림을 결합하고, 결합된 스트림을 송신을 위한 송

신 매체 (1808) 에 송신한다.

- [0294] 송신 매체 (1808) 는 DirecTV® 와 같은 디지털 위성 통신, 디지털 케이블, 유무선 인터넷 통신, 광 네트워크, 셀 전화 네트워크 등과 같은 여러 매체에 대응할 수 있지만, 이에 한정되는 것은 아니다. 송신 매체 (1808) 는 예를 들어, 무선 주파수 (RF) 에 대한 변조를 포함할 수 있다. 통상, 스펙트럼 제약으로 인해, 송신 매체는 제한된 대역폭을 갖고, 멀티플렉서 (1806) 로부터 송신 매체로의 데이터는 비교적 일정한 비트 레이트 (constant bit rate; CBR) 로 유지된다.
- [0295] 통상의 시스템에서, 멀티플렉서 (1806) 의 출력에서 일정한 비트 레이트 (CBR) 를 사용하기 위해서는, 멀티플렉서 (1806) 에 입력되는 인코딩된 멀티미디어 또는 비디오 스트림이 또한 CBR 일 것을 요구할 수도 있다. 배경기술에서 설명한 바와 같이, 비디오 콘텐츠 인코딩 시, CBR 을 사용하면, 가변 시각 품질로 되므로, 일반적으로 바람직하지 않다.
- [0296] 도시된 시스템에서, 인코더 (1804) 중 2 개 이상의 인코더는 입력 데이터의 예상 인코딩 복잡도를 전달한다. 이에 응답하여, 인코더 (1804) 중 하나 이상의 인코더는 멀티플렉서 (1806) 로부터 적응 비트 레이트 제어를 수신할 수도 있다. 이로 인해, 인코더 (1804) 는, 준-가변 (quasi-variable) 비트 레이트 방식으로 그 비디오 프레임에 대해 더 높은 비트 레이트 또는 더 높은 대역폭 (더 많은 프레임당 비트) 을 수신하도록 비교적 복잡한 비디오를 인코딩할 수 있을 것이다. 이로 인해, 멀티미디어 스트림 (1802) 은 일정한 시각 품질로 인코딩될 수 있다. 비교적 복잡한 비디오를 인코딩하는 특정 인코더 (1804) 에 의해 사용되는 여분 대역폭은, 인코더가 일정한 비트 레이트로 동작하도록 구현되었다면 다른 비디오 스트림 (1804) 을 인코딩하는데 사용되었을 비트로부터 발생한다. 이로 인해, 멀티플렉서 (1806) 의 출력이 일정한 비트 레이트 (CBR) 로 유지된다.
- [0297] 개개의 멀티미디어 스트림 (1802) 이 비교적 "버스티" 하지만 (즉, 사용되는 대역폭에서 변하지만), 다수의 비디오 스트림의 누적 함은 덜 버스티할 수 있다. 덜 복잡한 비디오를 인코딩하는 채널로부터의 비트 레이트는, 예를 들어, 멀티플렉서 (1806) 에 의해, 비교적 복잡한 비디오를 인코딩하는 채널에 재할당될 수 있고, 이로 인해, 결합된 비디오 스트림의 시각 품질을 전체로서 향상시킬 수 있다.
- [0298] 인코더 (1804) 는 함께 인코딩 및 멀티플렉싱되는 비디오 프레임 세트의 복잡도 인디케이션을 멀티플렉서 (1806) 에 제공한다. 멀티플렉서 (1806) 의 출력은 송신 매체 (1808) 에 지정된 비트 레이트보다 높지 않은 출력을 제공해야 한다. 복잡도 인디케이션은 상술한 콘텐츠 분류에 기초하여 선택된 품질 레벨을 제공할 수 있다. 멀티플렉서 (1806) 는 복잡도 인디케이션을 분석하고, 할당된 개수의 비트 또는 대역폭을 여러 인코더 (1804) 에 제공하고, 인코더 (1804) 는 이러한 정보를 사용하여 그 비디오 프레임 세트를 인코딩한다. 이로 인해, 비디오 프레임 세트가 개별적으로 가변 비트 레이트일 수 있고, 또한 일정한 비트 레이트를 그룹으로서 달성할 수 있다.
- [0299] 또한, 통상 임의의 일반적인 압축기에서 품질 기반 멀티미디어 압축을 가능하게 하는데 있어서 콘텐츠 분류가 사용될 수 있다. 본원에서 설명되는 콘텐츠 분류와 그 방법 및 장치는 임의의 멀티미디어 데이터의 품질 기반 및/또는 콘텐츠 기반 멀티미디어 처리에서 사용될 수도 있다. 일 예로서, 본원에서 설명되는 콘텐츠 분류와 그 방법 및 장치가 통상 임의의 일반적인 압축기에서 멀티미디어를 압축하는데 사용된다. 다른 예로서, 보간, 리샘플링, 인헨스먼트, 복원 및 프리젠테이션 동작과 같은 임의의 압축해제기 또는 디코더 또는 후처리기에서 압축해제 또는 디코딩하는 것이 있다.
- [0300] 이하, 도 19 를 참조하면, 통상적인 비디오 통신 시스템은 통신 네트워크에 의해 접속된 비디오 인코더 및 비디오 디코더로 이루어진 비디오 압축 시스템을 포함한다. 무선 네트워크는 에러가 발생하기 쉬운 네트워크의 한 클래스이고, 여기서, 통신 채널은 경로 손실 외에도 이동 시나리오에서 로그 정규 페이딩 또는 섀도잉 (shadowing) 및 다경로 페이딩을 나타낸다. 채널 에러를 해결하고, 애플리케이션 계층 데이터에 대한 신뢰성 있는 통신을 제공하기 위해, RF 변조기는 컨벌루션 코딩 또는 터보 코딩과 같은 채널 코딩 및 인터리버를 비롯한 순방향 에러 정정을 포함한다.
- [0301] 비디오 압축은 소스 비디오의 중복을 감소시키고, 코딩된 비디오 데이터의 각 비트에서 전달되는 정보 양을 증가시킨다. 이로 인해, 심지어 코딩된 비디오의 더 작은 부분이 손실된 경우에도 품질에 대한 영향을 증가시킨다. 비디오 압축 시스템에 고유한 공간 및 시간 예측은 손실을 악화시키고, 에러를 전파시켜 재구성된 비디오에서 아티팩트를 볼 수 있게 된다. 비디오 인코더에서의 에러 내성 알고리즘 및 비디오 디코더에서의 에러 복구 알고리즘은 비디오 압축 시스템의 에러 견고성을 향상시킨다.
- [0302] 통상, 비디오 압축 시스템은 그 기초가 되는 네트워크를 알지 못한다. 그러나, 에러가 발생하기 쉬운 네트

위크의 경우에, 링크/물리 계층에서 FEC 및 채널 코딩을 사용하여 애플리케이션 계층 내의 에러 보호 알고리즘을 통합 또는 정렬시키는 것이 매우 바람직하고, 시스템의 에러 성능을 향상시키는데 있어서 최고 효율을 제공한다.

[0303] 도 14 는 프레임을 인코딩하는 인코더 (228) 에서 발생할 수 있는 레이트-왜곡 데이터 흐름의 일 예를 도시한다. 프로세스 (1400) 는 시작 블록 (1402) 에서 시작하고, 결정 블록 (1404) 으로 진행하여, (예를 들어, 메타데이터를 통해) 전처리기 (226) 로부터 장면 전환 검출기 입력 (1410) 을 수신하고, 에러 내성 입력 (1406) 을 획득한다. 그 정보가 선택된 프레임이 I 프레임인 것을 나타내면, 프로세스는 그 프레임을 인트라 코딩한다. 그 정보가 선택된 프레임이 P 또는 B 프레임인 것을 나타내면, 프로세스는 인트라 코딩 및 모션 추정 (인터) 코딩을 사용하여 그 프레임을 인코딩한다.

[0304] 블록 (1404) 의 조건에 대해 긍정 조건이 일어난 후에, 프로세스 (1400) 는 준비 블록 (1414) 으로 진행하여, R-D 곡선에 기초하여 레이트 R 을 $R = R_{qual}$ 인 값, 즉, 원하는 타겟 품질로 설정한다. 이러한 설정은 R-D 곡선을 포함한 데이터 블록 (1416) 으로부터 수신된다. 그 다음에, 프로세스 (1400) 는 블록 (1418) 으로 진행하여, 블록 (1420) 의 콘텐츠 분류 프로세스로부터의 이미지/비디오 액티비티 정보 (예를 들어, 콘텐츠 분류) 에 기초하여 레이트 제어 비트 할당 $\{Q_{pi}\}$ 을 수행한다.

[0305] 또한, 레이트 제어 비트 할당 블록 (1418) 은 블록 (1422) 에서의 모션 추정을 위해 사용된다. 또한, 모션 추정 (1422) 은 전처리기 (1412) 로부터 메타데이터의 입력, 블록 (1424) 으로부터 모션 벡터 평활화 (MPEG-2 + 이력) 및 블록 (1426) 으로부터 다수의 기준 프레임 (인과적 + 비인과적 매크로블록 MB) 을 수신할 수 있다. 그 다음에, 프로세스 (1400) 는 블록 (1428) 으로 진행하여, 인트라 코딩된 모드에 대한 레이트 계산을 레이트 제어 비트 할당 (Q_{pi}) 에 대해 결정한다. 그 다음에, 프로세스 (1400) 는 블록 (1430) 으로 진행하여, 모드 및 양자화 파라미터를 결정한다. 블록 (1430) 의 모드 결정은 모션 추정 블록 (1422) 입력, 에러 내성 (1406) 입력 및 스케일러빌리티 R-D 에 기초하여 이루어지고, 이는 블록 (1432) 에서 결정된다. 일단 모드가 결정되면, 흐름은 블록 (1432) 으로 진행한다. 블록 (1430) 으로부터 블록 (1432) 으로의 흐름은, 데이터가 인코더의 제 1 패스로부터 제 2 패스 부분으로 전달될 때에 일어난다는 것에 주목하자.

[0306] 블록 (1432) 에서, 인코더 (228) 의 제 2 패스에 의해 변환 및 양자화가 수행된다. 변환/양자화 프로세스는 블록 (1444) 으로 표시된 바와 같이 조정 또는 미세 조정된다. 이러한 변환/양자화 프로세스는 레이트 제어 미세 조정 모듈 (도 7) 에 의해 영향을 받을 수도 있다. 그 다음에, 프로세스 (1400) 는 지그재그 소팅 및 엔트로피 코딩을 위한 블록 (1434) 으로 진행하여 인코딩된 베이스 계층을 생성한다. 지그재그 소팅은 양자화된 데이터를 인코딩에 효과적인 포맷으로 준비한다. 엔트로피 코딩은 일련의 비트 코드를 사용하여 일련의 가능한 심벌을 표현하는 압축 기술이다. 또한, 도 31 내지 도 36 을 참조하여 상술한 바와 같이, 변환/양자화 블록 (1432) 의 인헨스먼트 계층 결과는 덧셈기 (1436) 에 송신되고, 이 덧셈기 (1436) 는 베이스 계층을 빼고, 그 결과를 인헨스먼트 계층을 위한 ZZ/엔트로피 코더 (1438) 에 송신한다. 또한, 물론, 레이트 제어에 의해 사용되는 비트 레이트의 장단기 이력을 결정하는 동작 및 트루 레이트의 콘텐츠 분류 (1420) 를 갱신하기 위해 인헨스먼트 계층이 피드백된다 (라인 (1440) 의 트루 레이트 갱신 참조).

[0307] 도 17a 는 획득되었거나, 수신되었거나, 또는 그렇지 않은 경우에는 액세스 가능한 멀티미디어 데이터의 처리를 도시한 흐름도이다. 프로세스 (1700) 가 시작하고, 블록 (1702) 에서, 멀티미디어 데이터의 콘텐츠를 분류한다. 일 예시적인 양태에서, 콘텐츠 분류는 예를 들어, 도 7 의 콘텐츠 분류 모듈 (712) 과 같은 분류 수단에 의해 수행될 수 있다. 프로세스 (1700) 는 블록 (1704) 으로 진행하여, 콘텐츠 분류에 기초하여 멀티미디어 데이터를 제 1 데이터 그룹과 제 2 데이터 그룹으로 인코딩한다. 이러한 인코딩은, 제 1 데이터 그룹이 일 계층을 포함하고, 제 2 데이터 그룹이 제 1 데이터 그룹 계층과 연관된 제 1 차분 세분화를 포함하도록, 수행된다. 이는 예를 들어, 도 7 의 인코더 (228) 와 같은, 본원에 설명된 인코딩 수단에 의해 수행될 수 있다.

[0308] 도 17b 는 도 17a 에 도시된 프로세스를 수행할 수 있는 멀티미디어 인코딩 시스템 (1710) 의 블록도이다. 몇몇 양태에서, 멀티미디어 인코딩 시스템 (1710) 은 트랜스코더 (200) 와 같은 트랜스코더일 수 있다. 다른 양태에서, 인코딩 시스템 (1710) 은 트랜스코더의 일부를 포함할 수 있다. 멀티미디어 인코딩 시스템 (1710) 은 멀티미디어 데이터의 콘텐츠를 분류하는 수단, 즉, 멀티미디어 데이터의 콘텐츠를 분류하는 모듈 (712) 을 포함한다. 콘텐츠를 분류하는 수단은, 예를 들어, 전처리기 (예를 들어, 전처리기 (226)) 또는 인코더 (예를 들어, 인코딩 (228)) 내의 분류 모듈일 수 있다. 또한, 인코딩 시스템 (1710) 은 멀티미디어 테

이터를 인코딩하는 수단, 즉, 멀티미디어 데이터를 인코딩하는 모듈 (1714) 을 포함하고, 이 모듈 (1714) 은 콘텐츠 분류에 기초하여 멀티미디어 데이터를 제 1 데이터 그룹과 제 2 데이터 그룹으로 인코딩하도록 구성될 수 있고, 여기서, 이러한 인코딩은, 제 1 데이터 그룹이 일 계수를 포함하고, 제 2 데이터 그룹이 제 1 데이터 그룹 계수와 연관된 제 1 차분 세분화를 포함하도록, 수행된다. 또한, 상술한 바와 같은 다른 트랜스코더 컴포넌트는 인코딩 시스템 (1710) 에 포함될 수도 있다.

[0309] 도 23, 도 24, 도 27 및 도 28 은, 멀티미디어 데이터를 인코딩하는 방법을 예시하며 본원에 설명된 양태를 구현하는 프로세스 흐름도이다. 도 23 은 콘텐츠 정보에 기초하여 멀티미디어 데이터를 인코딩하는 프로세스 (2300) 를 도시한 프로세스 흐름도이다. 블록 (2305) 에서, 프로세스 (2300) 는 인코딩된 멀티미디어 데이터를 수신하고, 블록 (2310) 에서, 프로세스 (2300) 는 멀티미디어 데이터를 디코딩한다. 블록 (2315) 에서, 프로세스 (2300) 는 디코딩된 멀티미디어 데이터와 연관된 콘텐츠 정보를 결정한다. 블록 (2320) 에서, 프로세스 (2300) 는 콘텐츠 정보에 기초하여 멀티미디어 데이터를 인코딩한다.

[0310] 도 24 는 콘텐츠 정보 레벨에 기초하여 데이터 경계를 정렬시키도록 멀티미디어 데이터를 인코딩하는 프로세스 (2400) 를 도시한 프로세스 흐름도이다. 블록 (2405) 에서, 프로세스 (2400) 는 멀티미디어 데이터와 연관된 콘텐츠 정보를 얻는데, 이는 예를 들어, 도 7 에 도시된 콘텐츠 분류 모듈 (712) 또는 전처리기 (226) 에 의해 행해질 수 있다. 블록 (2410) 에서, 프로세스 (2400) 는 콘텐츠 정보에 기초하여 데이터 경계를 정렬시키도록 멀티미디어 데이터를 인코딩한다. 예를 들어, 슬라이스 경계와 액세스 유닛 경계는 인코딩된 멀티미디어 데이터의 콘텐츠 분류에 기초하여 프레임 경계와 정렬된다. 그 다음에, 후속 처리 및/또는 이동 디바이스로의 송신을 위해 인코딩된 데이터를 이용할 수 있고, 프로세스 (2400) 는 종료한다.

[0311] 도 27 은 콘텐츠 정보에 기초한 적응 인트라 리프레시 방식을 사용하여 데이터를 인코딩하는 프로세스 (2700) 를 도시한 프로세스 흐름도이다. 프로세스 (2700) 시작 시, 멀티미디어 데이터는 획득되어 있다. 블록 (2705) 에서, 프로세스 (2700) 는 멀티미디어 데이터의 콘텐츠 정보를 얻는다. 상술한 바와 같이, 콘텐츠 정보의 획득은, 예를 들어, 전처리기 (226) 또는 콘텐츠 분류 모듈 (712) 에 의해 수행될 수 있다. 프로세스 (2700) 는 블록 (2710) 으로 진행하고, 이 블록 (2710) 에서, 적응 인트라 리프레시 에러 내성 방식을 사용하여 멀티미디어 데이터를 인코딩하는데, 적응 인트라 리프레시 에러 내성 방식은 콘텐츠 정보에 기초하고 있다. 블록 (2710) 의 기능은 인코더 (228) 에 의해 수행될 수 있다. 인코딩된 데이터는 후속 처리 및 송신을 위해 이용 가능하게 되고, 그 다음에, 프로세스 (2700) 는 종료한다.

[0312] 도 28 은 멀티미디어 콘텐츠 정보에 기초한 중복 I 프레임을 사용하여 멀티미디어 데이터를 인코딩하는 프로세스를 도시한 프로세스 흐름도이다. 프로세스 (2800) 시작 시, 처리를 위해 멀티미디어 데이터를 이용할 수 있다. 블록 (2805) 에서, 프로세스 (2800) 는 멀티미디어 데이터의 콘텐츠 정보를 얻는다. 상술한 바와 같이, 이는 예를 들어, 전처리기 (226) 및/또는 인코더 (228) 에 의해 행해질 수 있다. 블록 (2810) 에서, 프로세스 (2800) 는 콘텐츠 정보에 기초하여 하나 이상의 추가 I-프레임을 인코딩된 데이터 내에 삽입하도록 멀티미디어 데이터를 인코딩한다. 이는, 에러 내성 방식과 관련하여 상술한 바와 같이 인코더 (228) 에 의해 행해져서, 사용된 에러 내성 방식에 따라 베이스 계층 또는 인핸스먼트 계층 내에 I-프레임을 삽입할 수 있다. 블록 (2810) 다음에, 후속 처리 및/또는 이동 디바이스로의 송신을 위해 인코딩된 데이터를 이용할 수 있다.

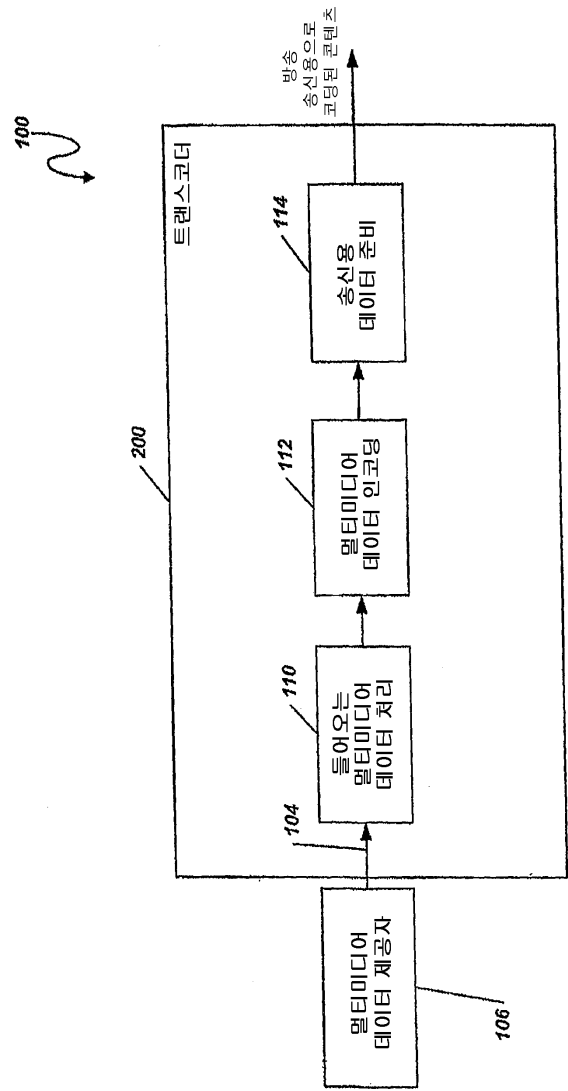
[0313] 본원에 설명된 방법은 당업자에게 공지된 다양한 통신 하드웨어, 프로세서 및 시스템 상에 구현될 수도 있다는 것에 주목해야 한다. 예를 들어, 클라이언트가 본원에 설명된 것과 같이 동작하기 위한 일반 요건은, 클라이언트가 콘텐츠 및 정보를 표시하는 디스플레이, 클라이언트의 동작을 제어하는 프로세서 및 클라이언트의 동작 관련 데이터 및 프로그램을 저장하는 메모리를 갖는 것이다. 일 양태에서, 클라이언트는 셀룰러 전화기이다. 다른 양태에서, 클라이언트는 통신 기능을 갖는 핸드헬드 컴퓨터이다. 또 다른 양태에서, 클라이언트는 통신 기능을 갖는 개인용 컴퓨터이다. 또한, GPS 수신기와 같은 하드웨어를 클라이언트에 통합하여, 여러 양태를 구현할 수도 있다. 본원에 개시된 양태와 관련하여 설명된 여러 예시적인 논리, 논리 블록, 모듈 및 회로는 범용 프로세서, DSP (digital signal processor), ASIC (application specific integrated circuit), FPGA (field programmable gate array) 또는 다른 프로그램 가능 논리 디바이스, 이산 게이트 또는 트랜지스터 논리, 이산 하드웨어 컴포넌트, 또는 본원에 설명된 기능을 수행하도록 설계된 그 임의의 조합을 사용하여 구현 또는 수행될 수도 있다. 범용 프로세서는 마이크로프로세서일 수도 있지만, 다른 방법으로는, 범용 프로세서는 임의의 통상적인 프로세서, 컨트롤러, 마이크로컨트롤러, 또는 상태 머신일 수도 있다. 또한, 프로세서는 예를 들어, DSP 와 마이크로프로세서의 조합, 복수의 마이크로프로세서, DSP 코어와 함께 하나 이상의 마이크로프로세서, 또는 임의의 다른 그러한 구성과 같은 컴퓨팅 디바이스의 조합으로서 구현될 수도 있

다.

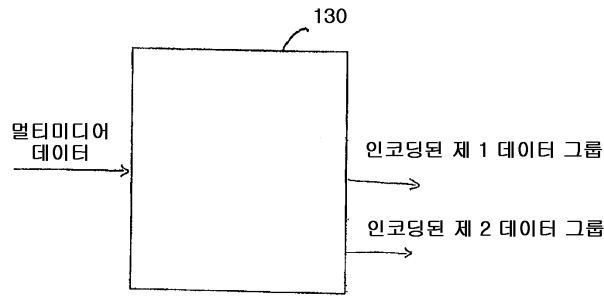
- [0314] 본원에 개시된 양태와 관련하여 설명된 여러 예시적인 논리, 논리 블록, 모듈 및 회로는 범용 프로세서, DSP (digital signal processor), ASIC (application specific integrated circuit), FPGA (field programmable gate array) 또는 다른 프로그램 가능 논리 디바이스, 이산 게이트 또는 트랜지스터 논리, 이산 하드웨어 컴포넌트, 또는 본원에 설명된 기능을 수행하도록 설계된 그 임의의 조합을 사용하여 구현 또는 수행될 수도 있다. 범용 프로세서는 마이크로프로세서일 수도 있지만, 다른 방법으로는, 범용 프로세서는 임의의 통상적인 프로세서, 컨트롤러, 마이크로컨트롤러, 또는 상태 머신일 수도 있다. 또한, 프로세서는 예를 들어, DSP 와 마이크로프로세서의 조합, 복수의 마이크로프로세서, DSP 코어와 함께 하나 이상의 마이크로프로세서, 또는 임의의 다른 그러한 구성과 같은 컴퓨팅 디바이스의 조합으로서 구현될 수도 있다.
- [0315] 본원에 개시된 방법 및 장치는 한 포맷으로 인코딩된 비디오 데이터를 다른 포맷으로 인코딩된 비디오 데이터로 트랜스코딩하는 것을 제공하는데, 여기서, 인코딩은 비디오 데이터의 콘텐츠에 기초하고 있고, 인코딩은 예러에 대한 복원성이 있다. 본원에 개시된 예와 관련하여 설명된 방법 또는 알고리즘은 하드웨어에 직접, 프로세서에 의해 실행되는 소프트웨어 모듈, 펌웨어 또는 이들 중 2 개 이상의 조합으로 수록될 수도 있다. 소프트웨어 모듈은 RAM 메모리, 플래시 메모리, ROM 메모리, EPROM 메모리, EEPROM 메모리, 레지스터, 하드 디스크, 착탈식 디스크, CD-ROM 또는 당해 기술분야에서 공지된 임의의 다른 형태의 저장 매체에 상주할 수도 있다. 예시적인 저장 매체는 프로세서에 연결되어, 프로세서가 저장 매체로부터 정보를 판독하고, 저장 매체에 정보를 기입할 수 있도록 한다. 다른 방법으로는, 저장 매체는 프로세서와 일체로 형성될 수도 있다. 프로세서 및 저장 매체는 ASIC 에 상주할 수도 있다. ASIC 은 사용자 단말기에 상주할 수도 있다. 다른 방법으로는, 프로세서와 저장 매체는 사용자 단말기에서 이산 컴포넌트로서 상주할 수도 있다.
- [0316] 상술한 예는 단지 예시적인 것으로서, 당업자라면 본원에 설명된 본 발명의 개념에서 이탈함이 없이 상술한 예를 다양하게 사용하고, 상술한 예로부터 다양하게 진보를 달성할 수도 있다. 당업자라면 이들 예의 여러 변형을 쉽게 알 수도 있고, 본원에 규정된 일반적인 원리는 본원에 설명된 신규한 양태의 사상 및 범위로부터 이탈함이 없이 예를 들어, 인스턴스 메시징 서비스 또는 임의의 일반적인 무선 데이터 통신 애플리케이션에서, 다른 예에 적용될 수도 있다. 따라서, 본 개시내용의 범위는 본원에 도시된 예에 한정되는 것이 아니라, 본원에 개시된 신규한 특징 및 원리와 일치하는 가장 넓은 범위가 주어진다. 본원에서 "예시적" 이란 단어는 오로지 "예, 실례, 예증으로서 기능을 하는 것" 을 의미하기 위해 사용될 뿐이다. "예시적" 인 것으로서 본원에 설명된 임의의 예는 다른 예보다 반드시 바람직하거나 유익한 것으로서 해석되지 않아야 한다. 따라서, 본원에 개시된 신규한 양태는 다음 청구항의 범위에 의해서만 규정되어야 한다.

도면

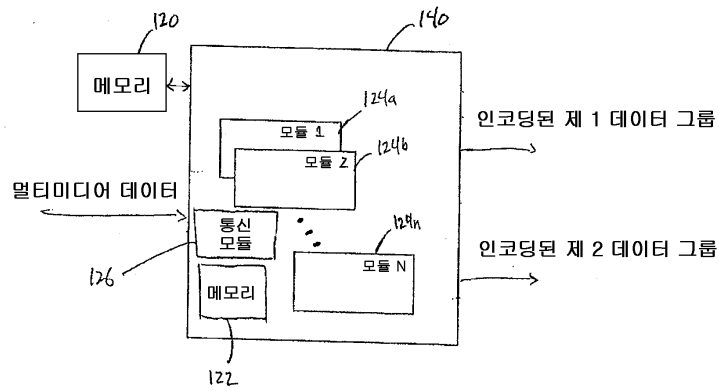
도면1



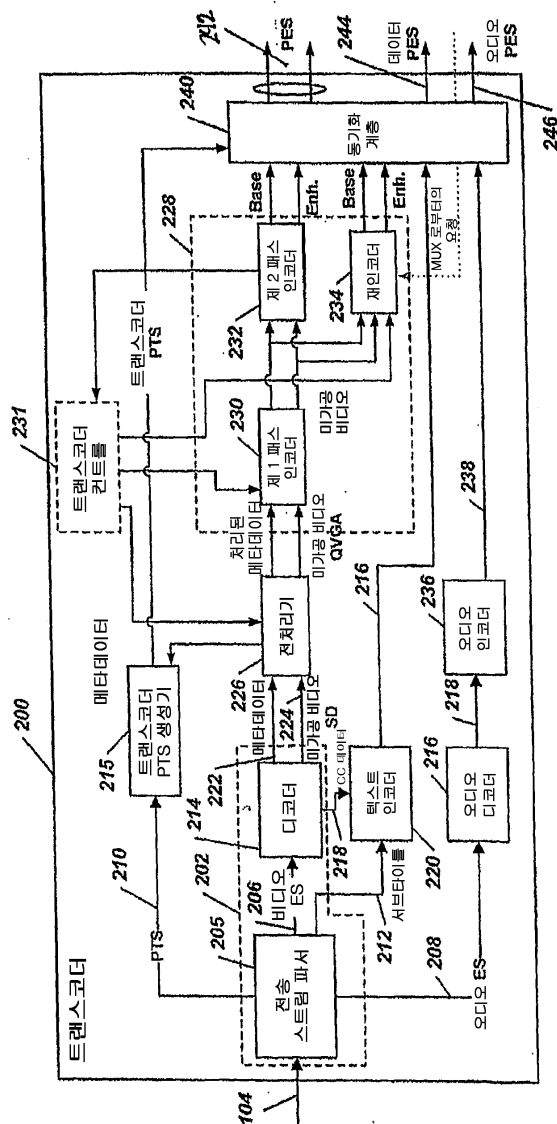
도면1b



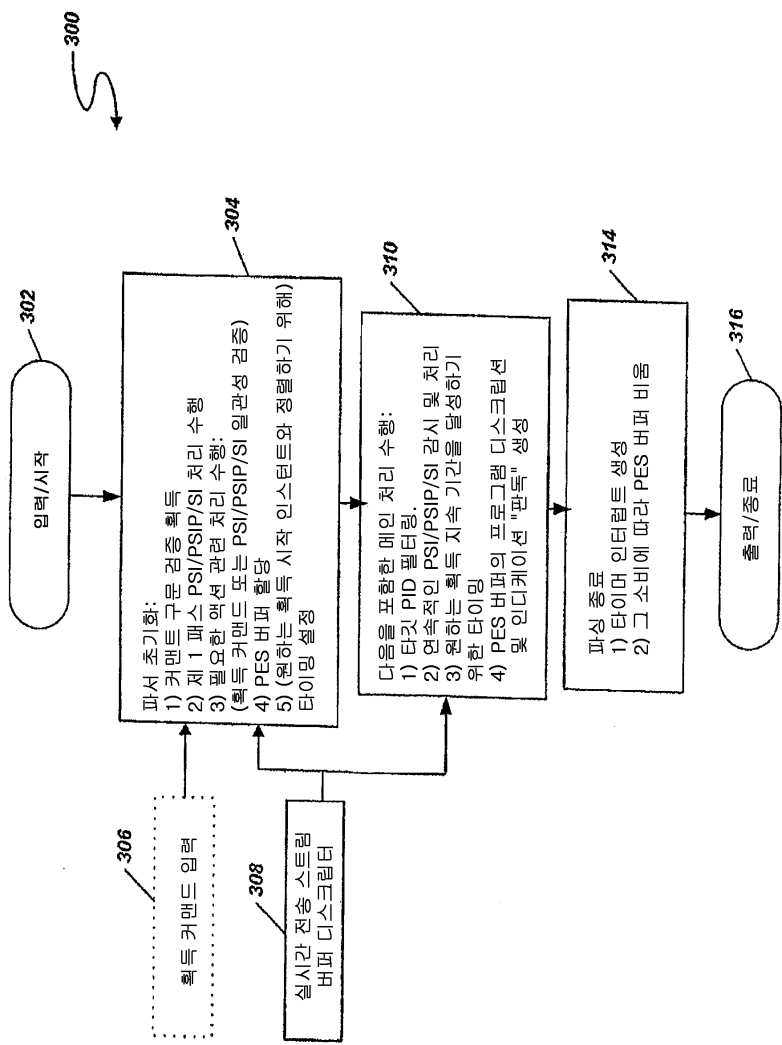
도면1c



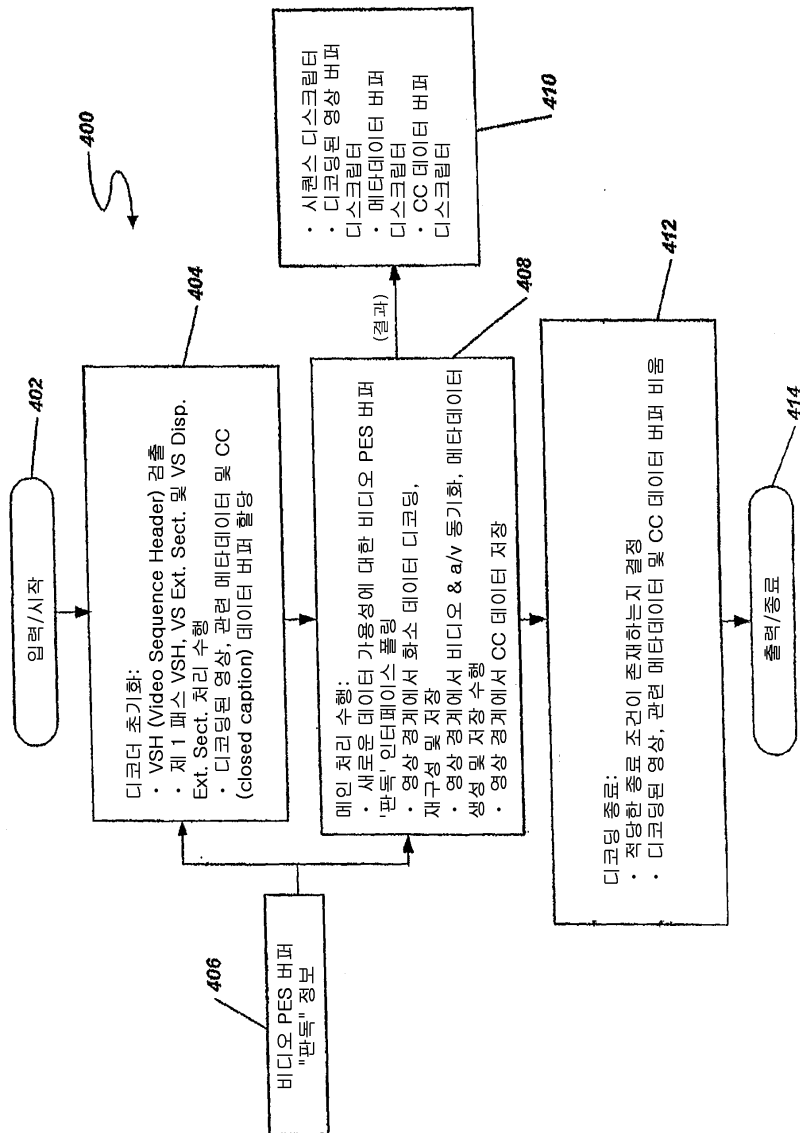
도면2



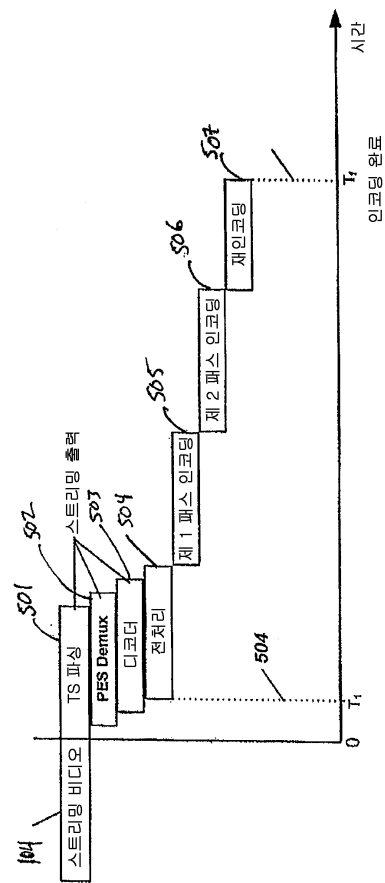
도면3



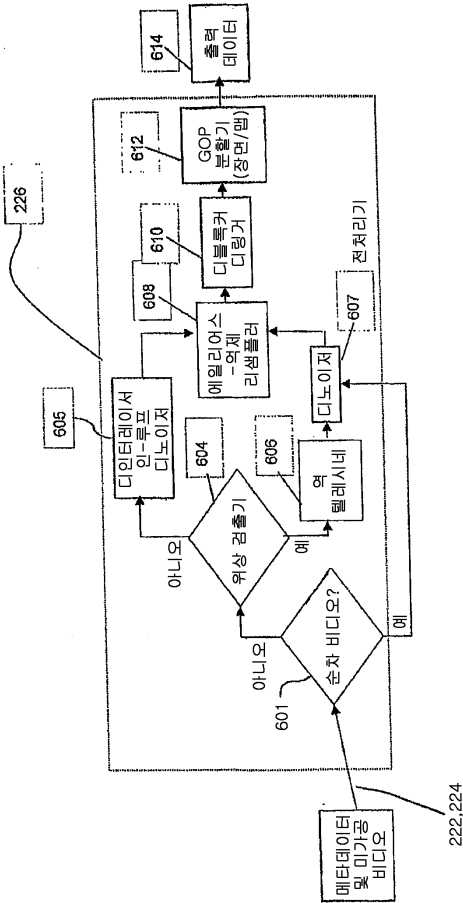
도면4



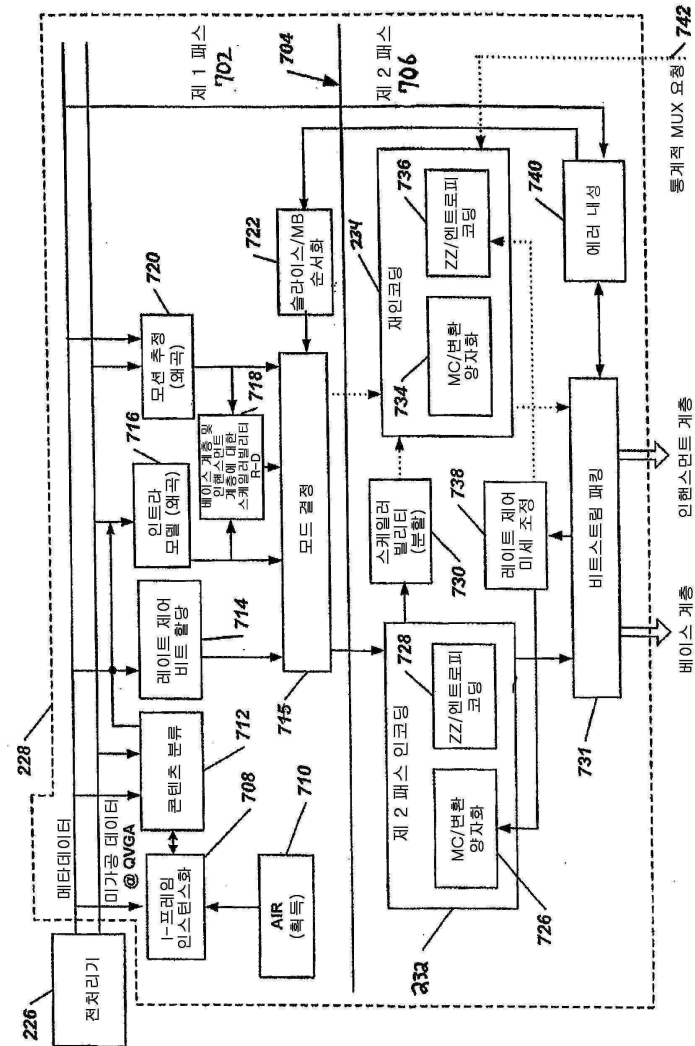
도면5



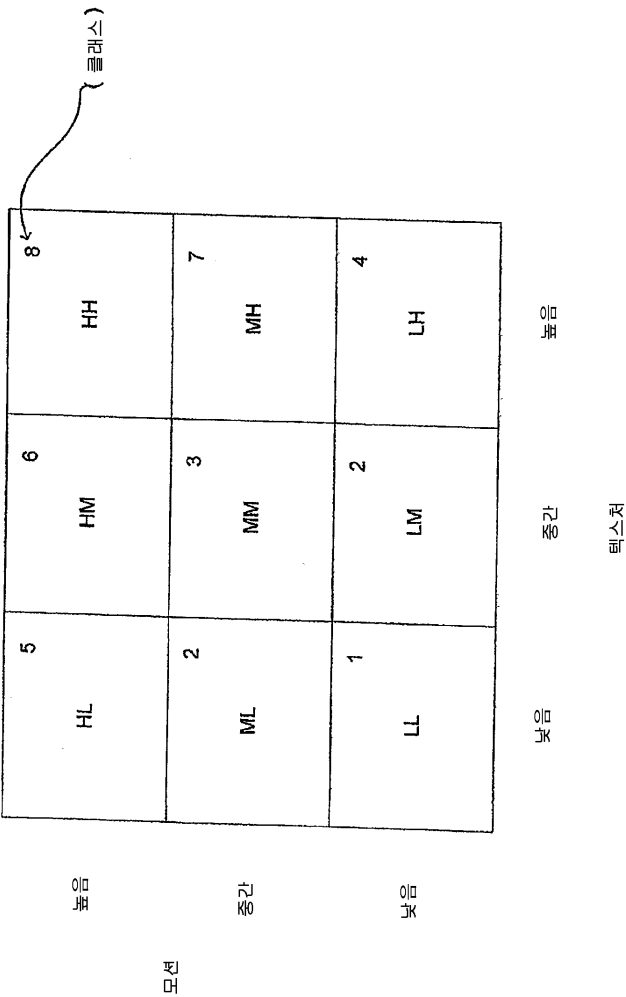
도면6



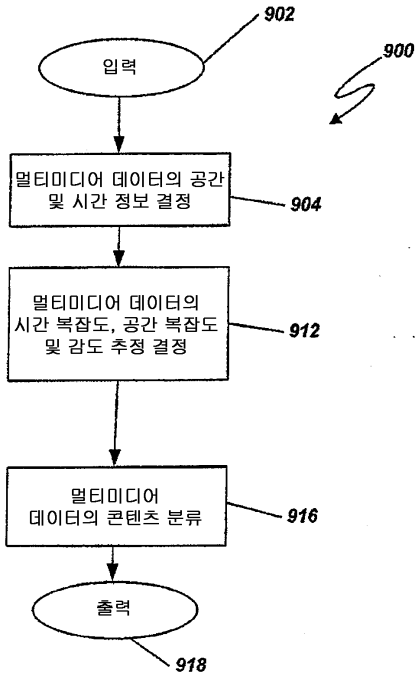
도면7



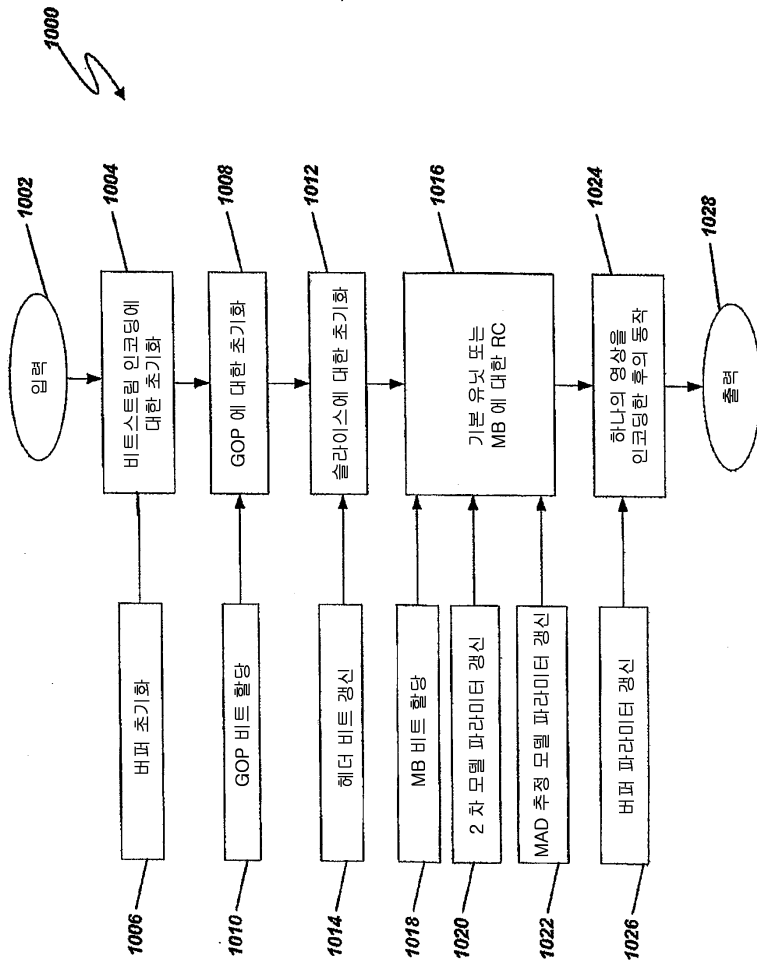
도면8



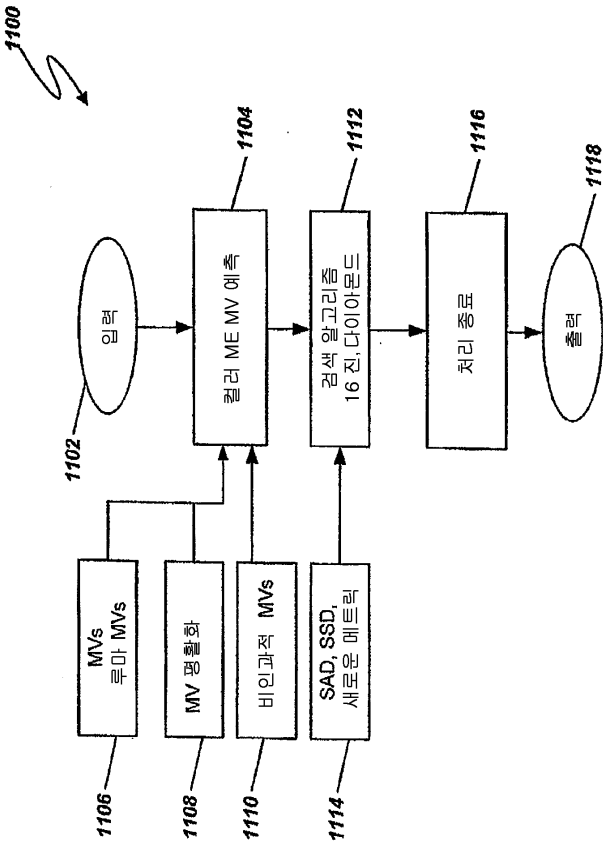
도면9



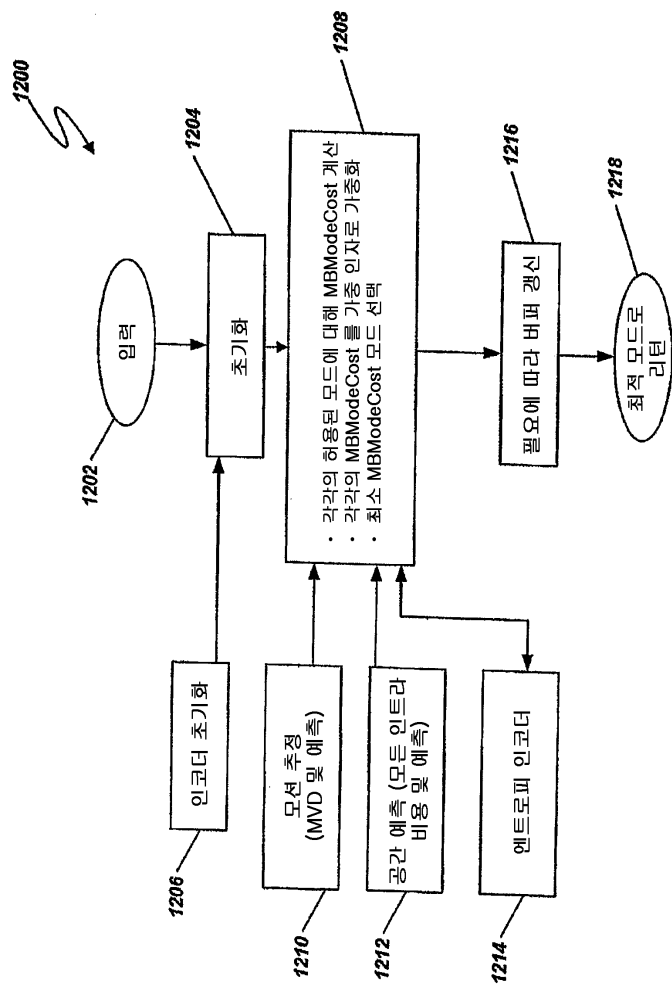
도면10



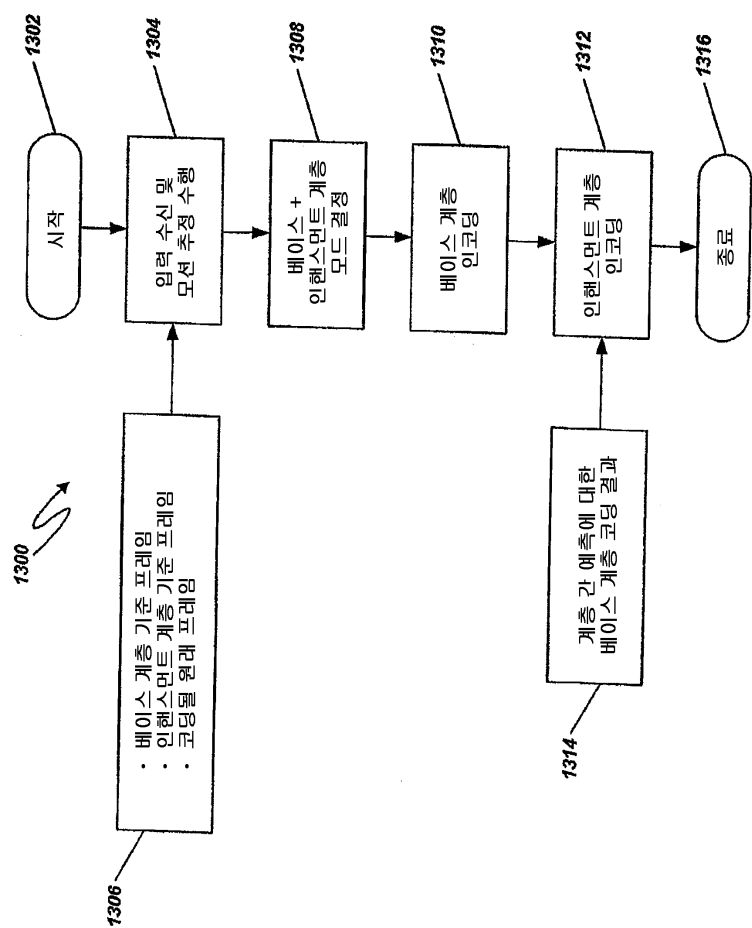
도면11



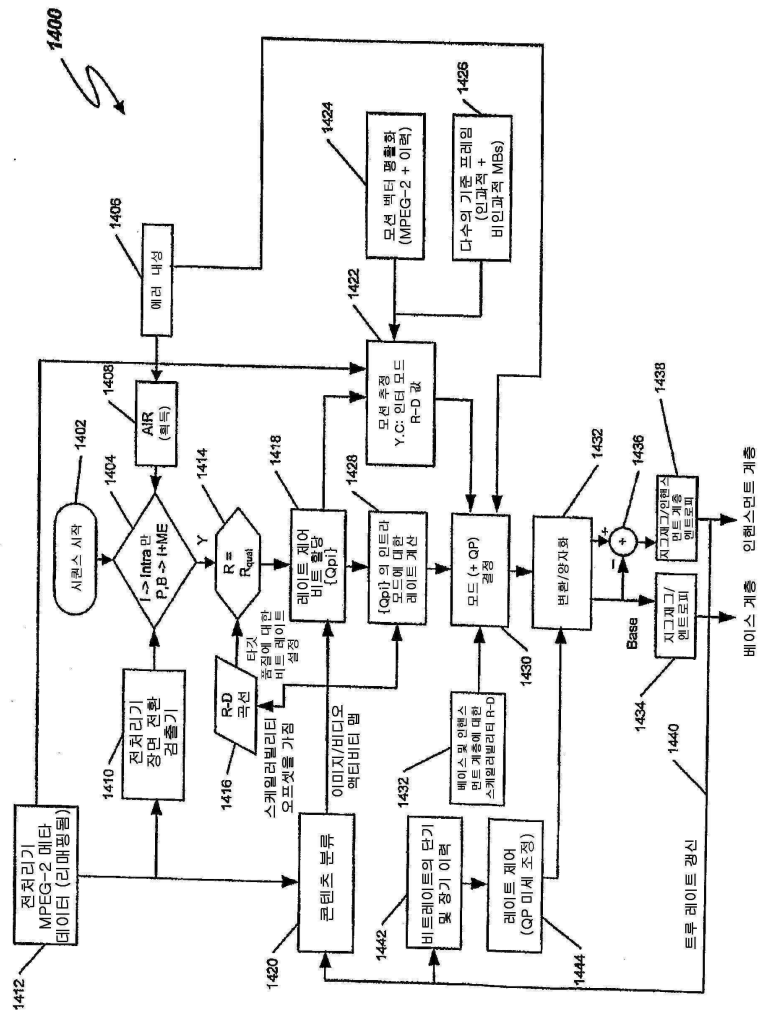
도면12



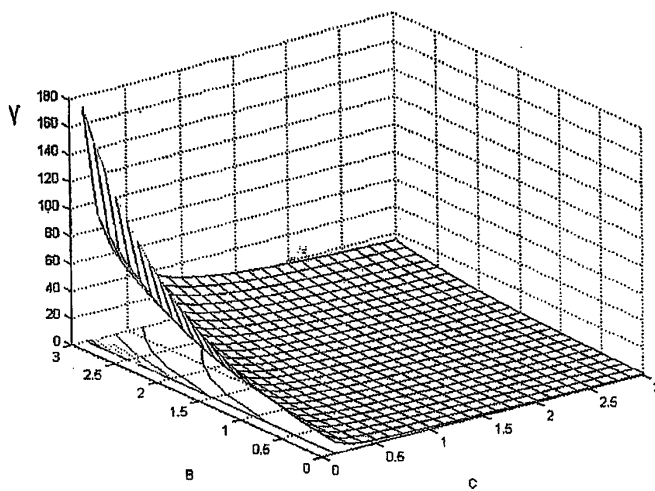
도면13



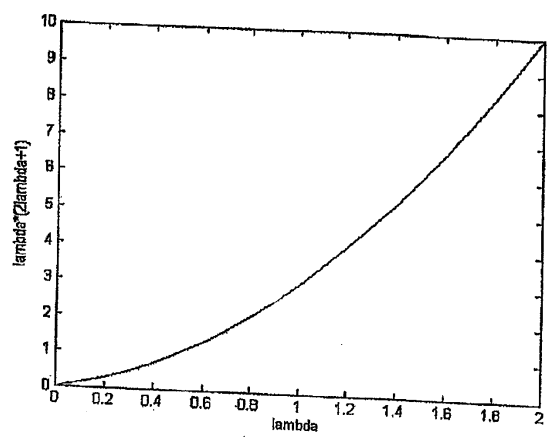
도면14



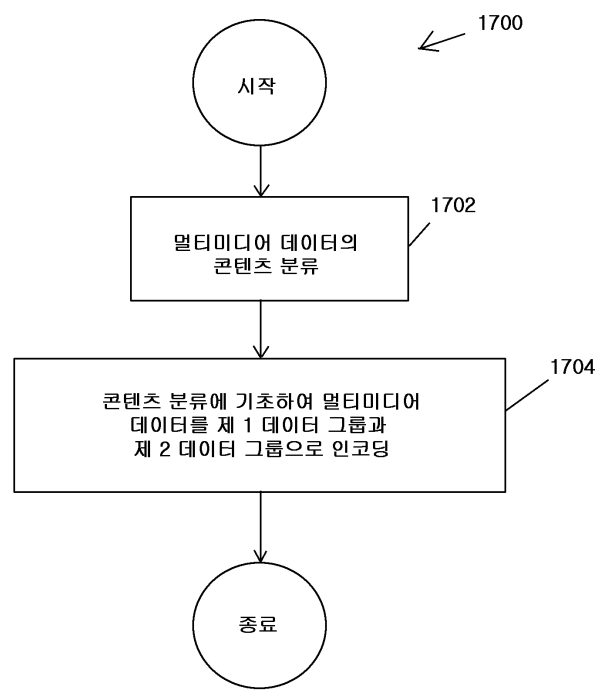
도면15



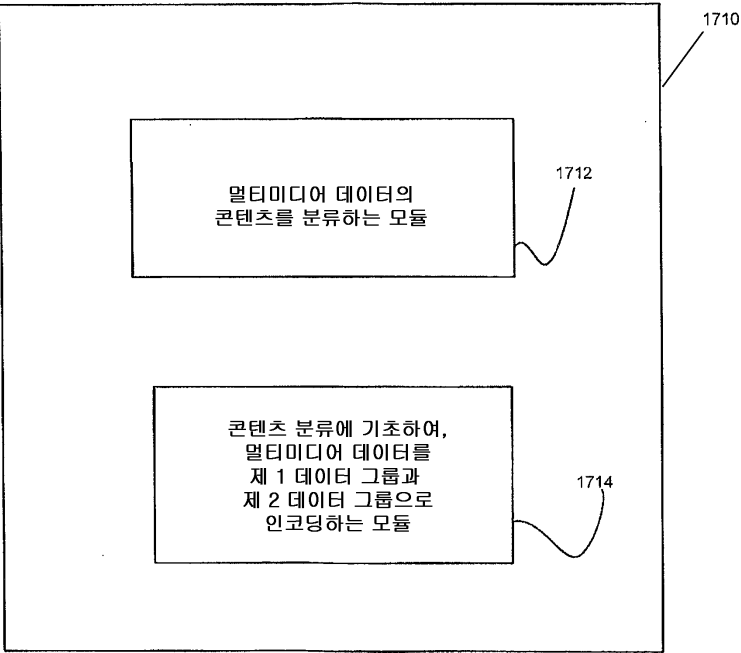
도면16



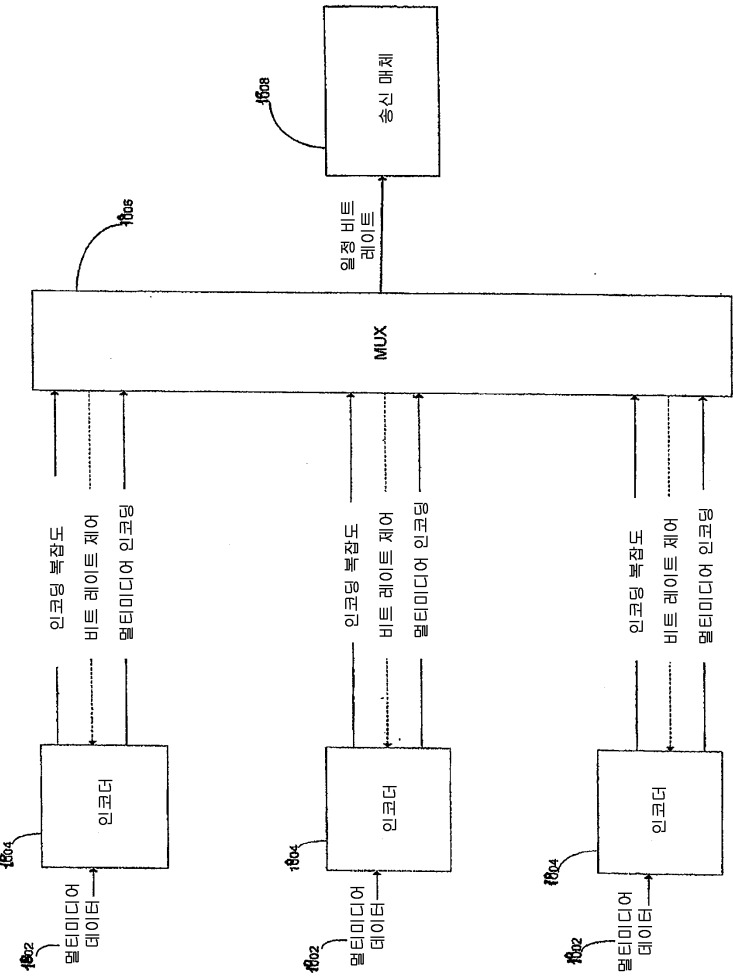
도면17a



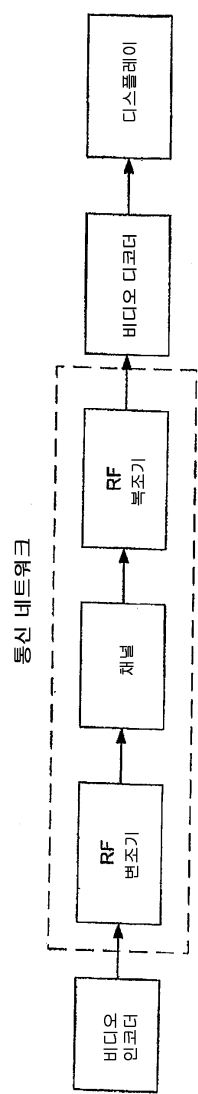
도면17b



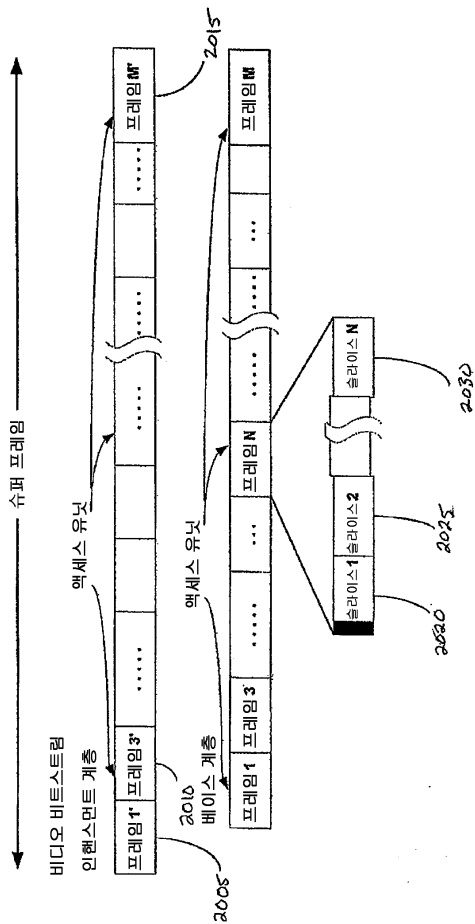
도면18



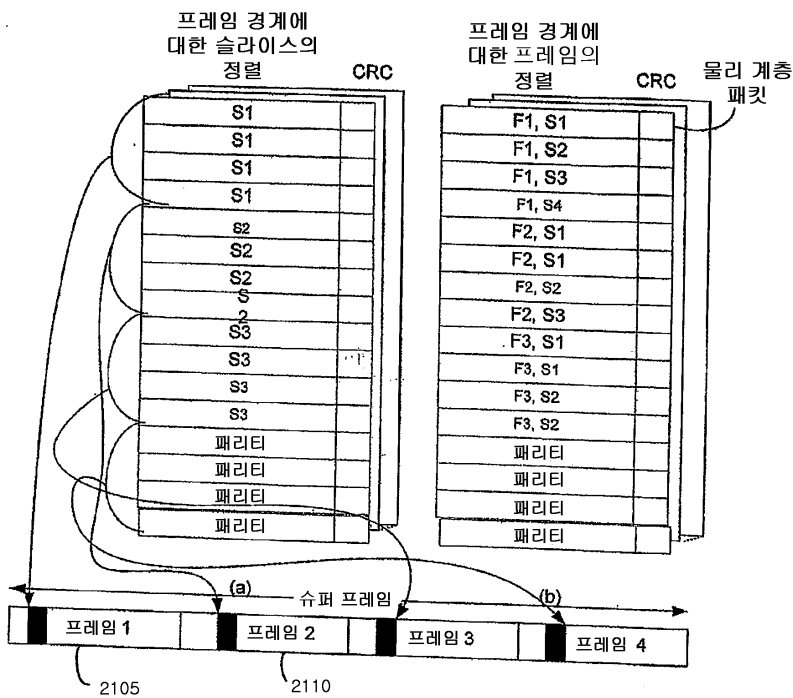
도면19



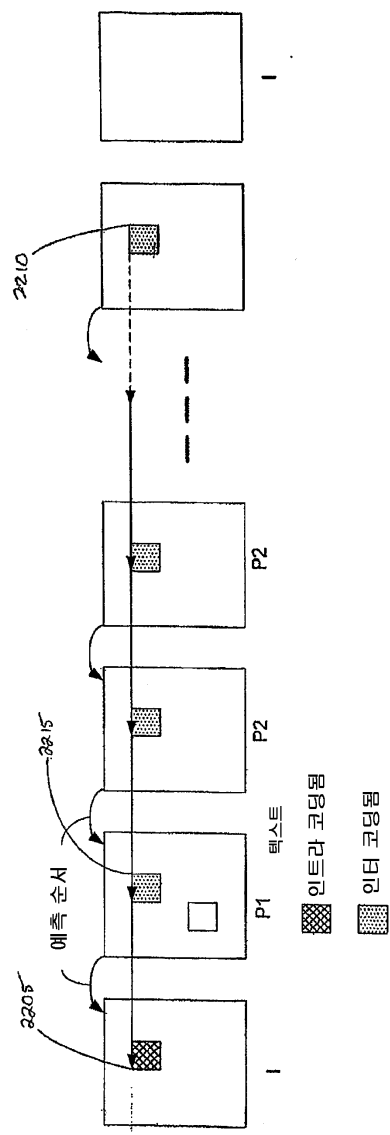
도면20



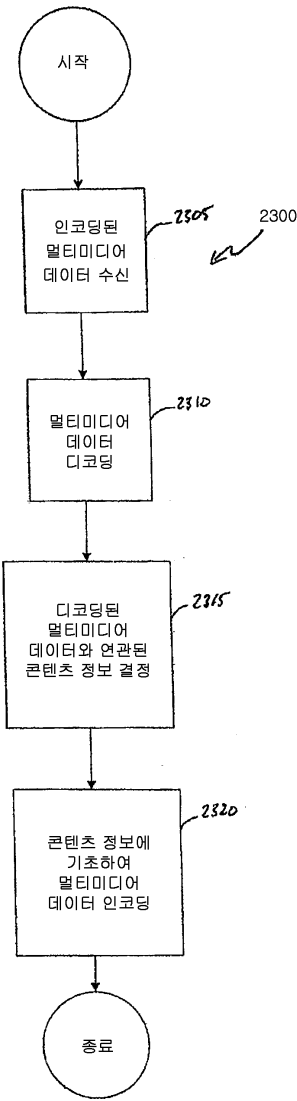
도면21



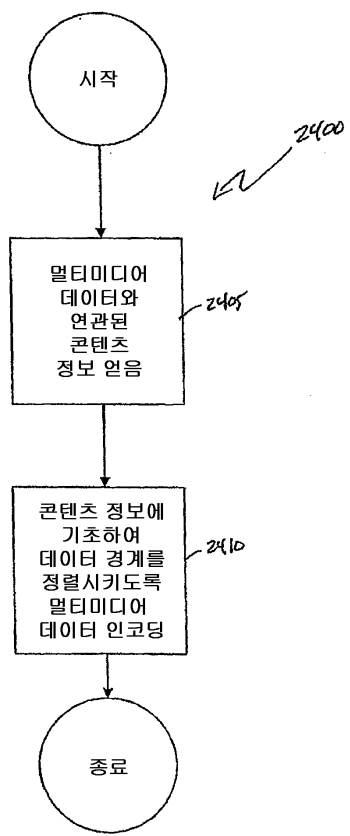
도면22



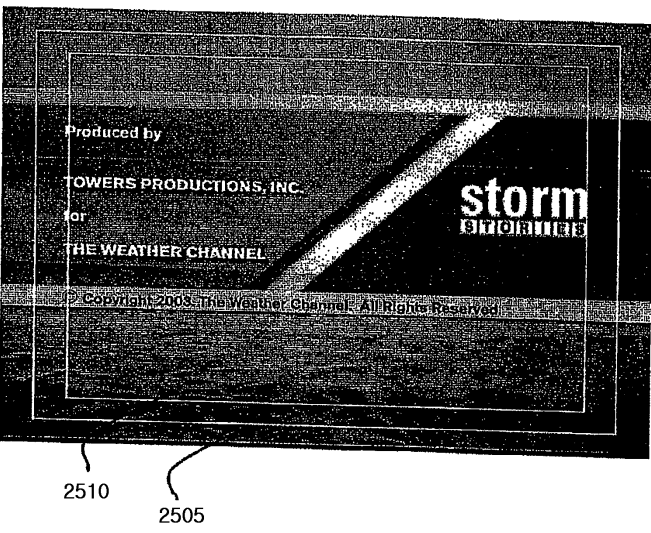
도면23



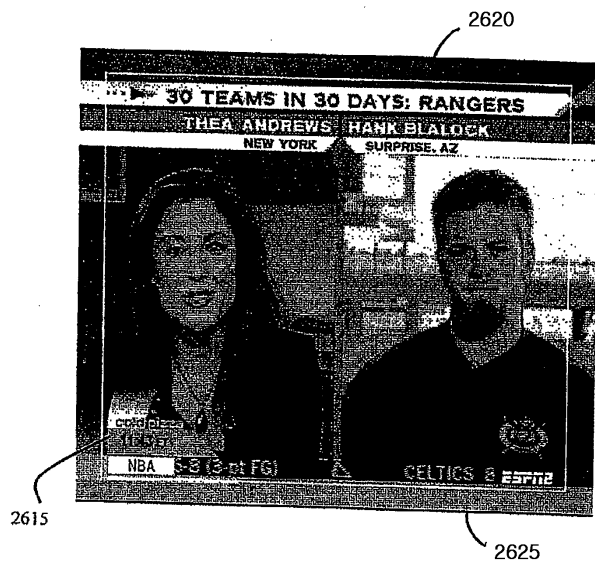
도면24



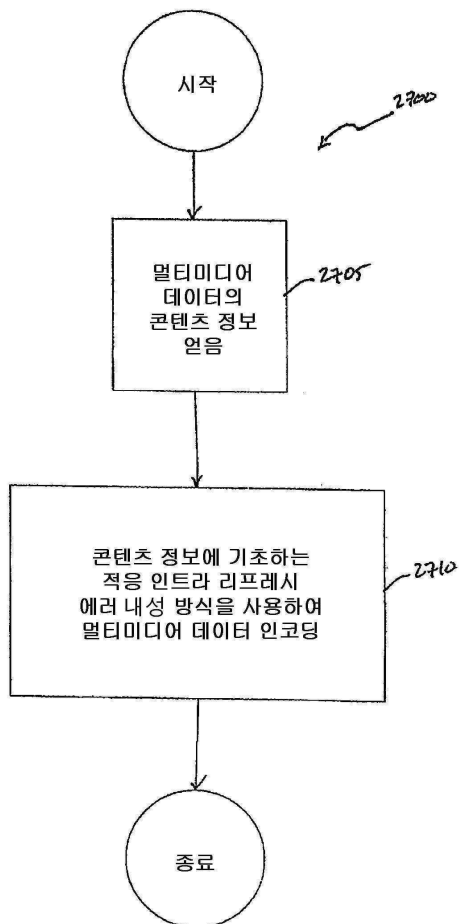
도면25



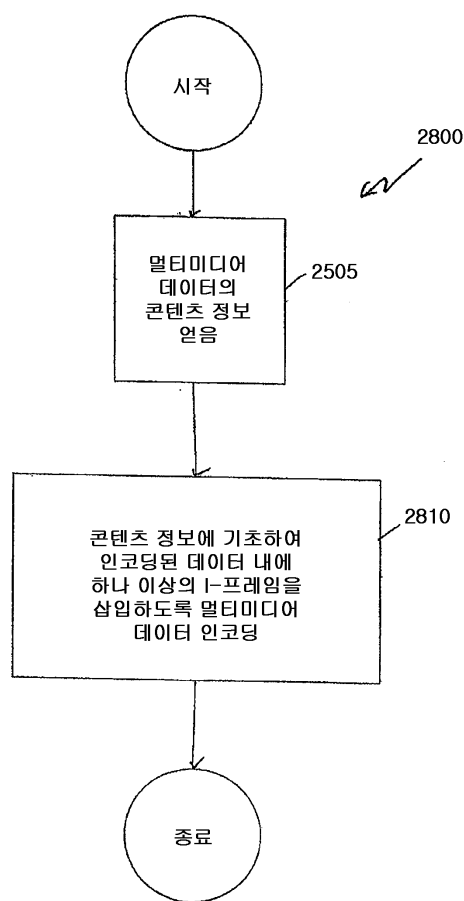
도면26



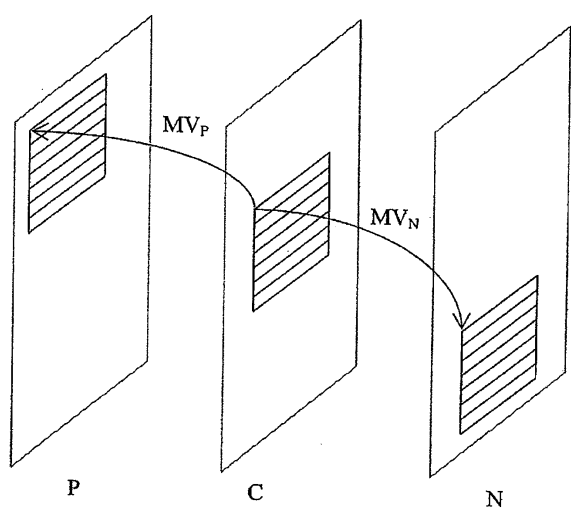
도면27



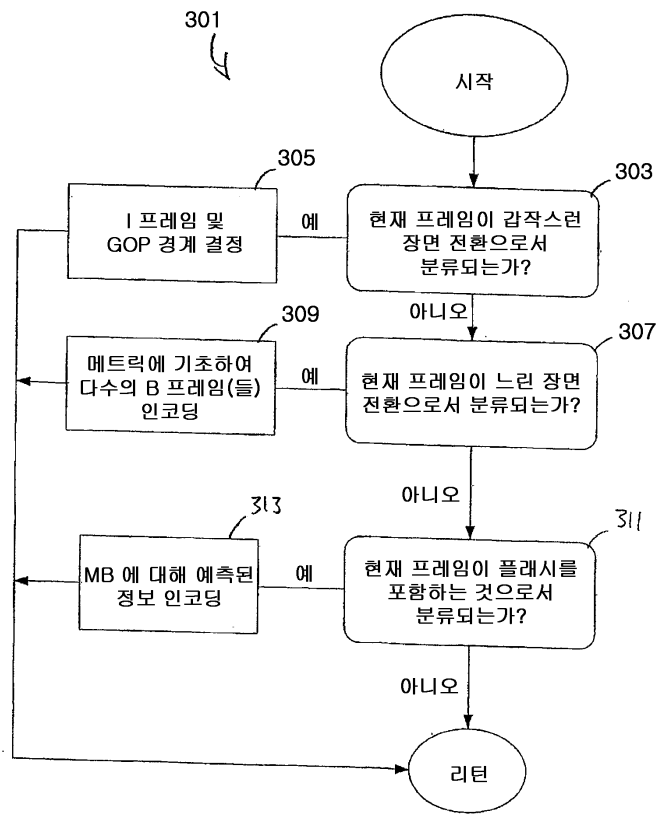
도면28



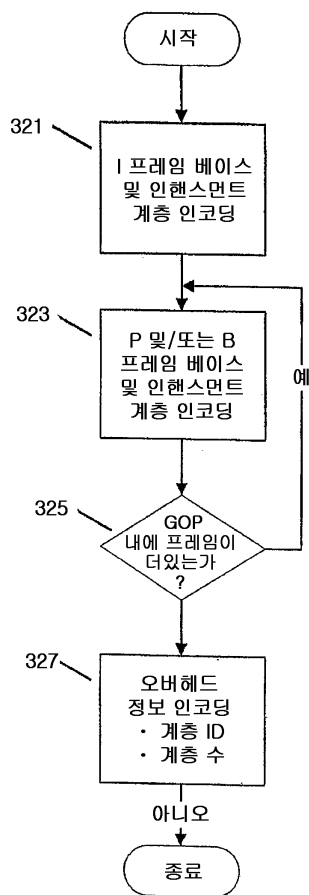
도면29



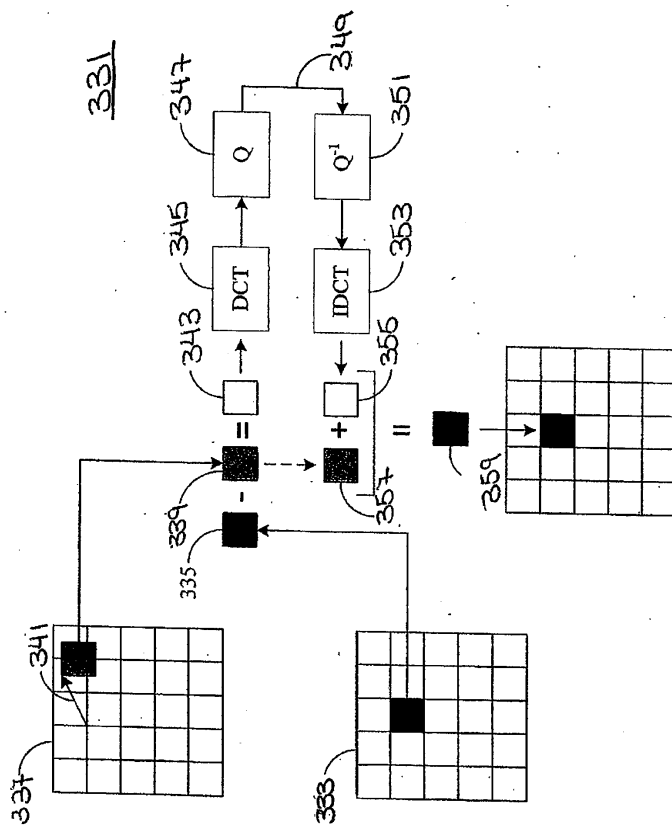
도면30



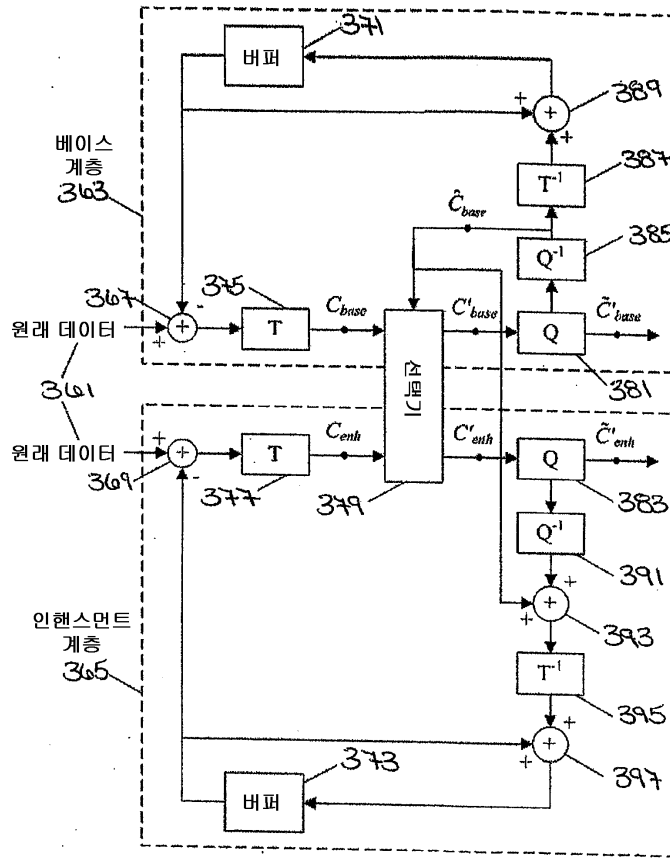
도면31



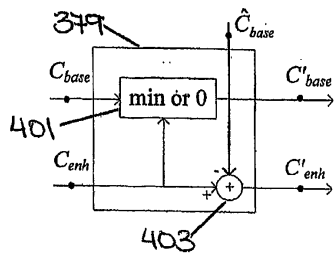
도면32



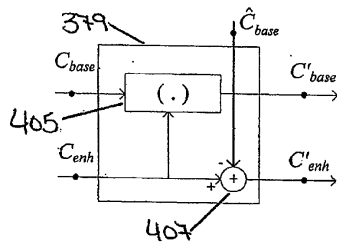
도면33



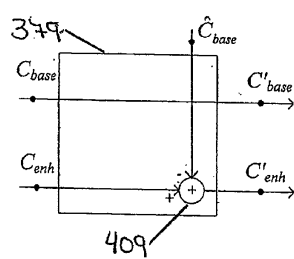
도면34



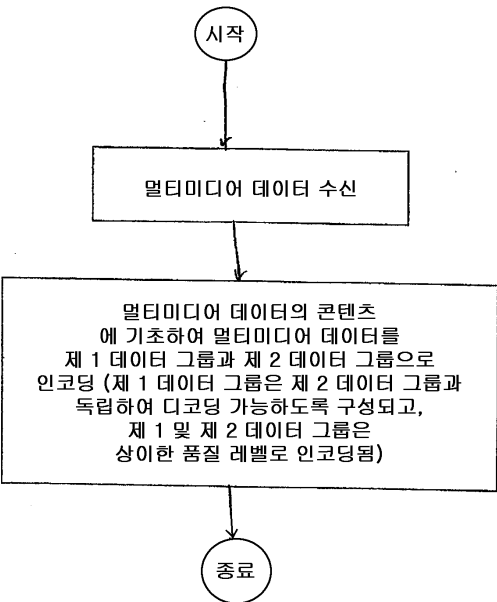
도면35



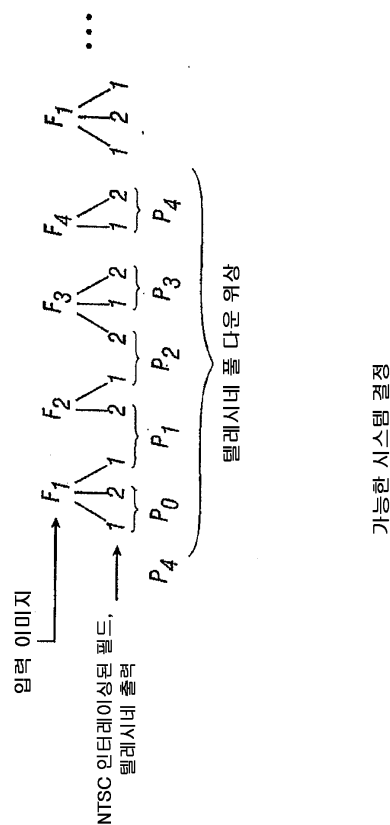
도면36



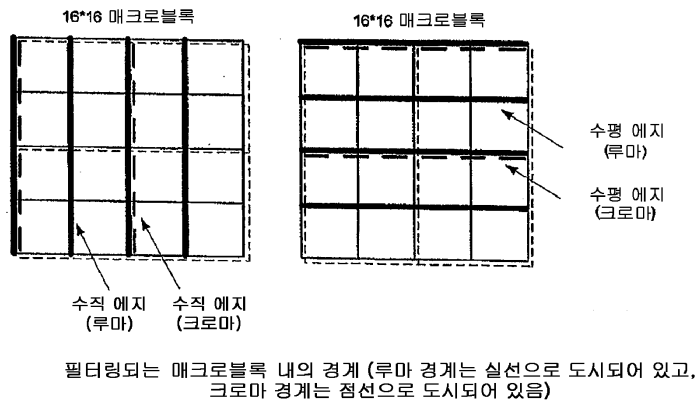
도면37



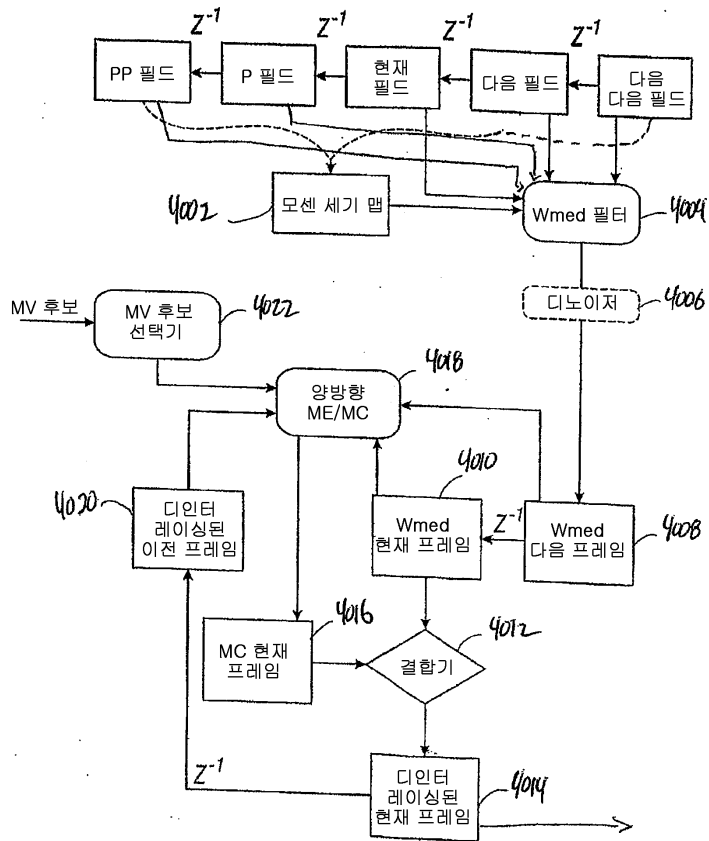
도면38



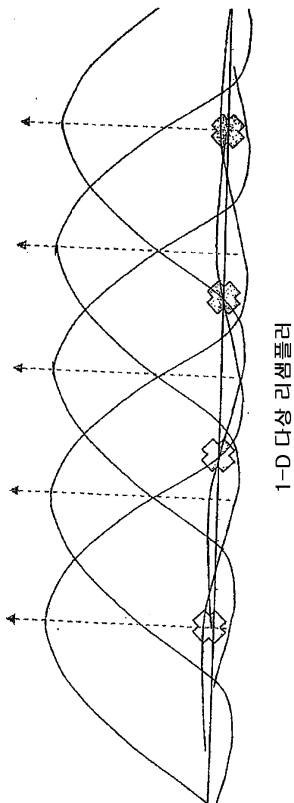
도면39



도면40



도면41



도면42

