(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2020년01월08일
(11) 등록번호 10-2064023
(24) 등록일자 2020년01월02일

(51) 국제특허분류(Int. Cl.)
H04N 19/107 (2014.01) *H04N 19/89* (2014.01)
(21) 출원번호 10-2014-7005589
(22) 출원일자(국제) 2012년09월02일
심사청구일자 2017년08월04일
(85) 번역문제출일자 2014년02월28일
(65) 공개번호 10-2014-0054164
(43) 공개일자 2014년05월08일
(86) 국제출원번호 PCT/US2012/053583
(87) 국제공개번호 WO 2013/033679
국제공개일자 2013년03월07일
(30) 우선권주장
1115201.4 2011년09월02일 영국(GB)
13/274,881 2011년10월17일 미국(US)
(56) 선행기술조사문헌
US20080247469 A1
Sarat Chandra Vadapalli et al: "Efficient Alternative to Intra Refresh using Reliable Reference Frames", Multimedia and Expo, 2007 IEEE International conference, 1 July 2007.
Yuan Zhang et al., Joint Source/Channel Rate - Prone Networks, IEEE Transactions on Multimedia, IEEE, 2007, Vol. 9, No.3, pp.445-454.

(73) 특허권자
마이크로소프트 코포레이션
미국 워싱턴주 (우편번호 : 98052) 레드몬드 원 마이크로소프트 웨이
(72) 발명자
자오 데이비드
아일랜드 2 더블린 서 존 로저손스 퀘이 70 인터내셔널 페이턴츠 스카이프
닐슨 마티아스
아일랜드 2 더블린 서 존 로저손스 퀘이 70 인터내셔널 페이턴츠 스카이프
(뒷면에 계속)
(74) 대리인
제일특허법인(유)

전체 청구항 수 : 총 28 항 심사관 : 김영태

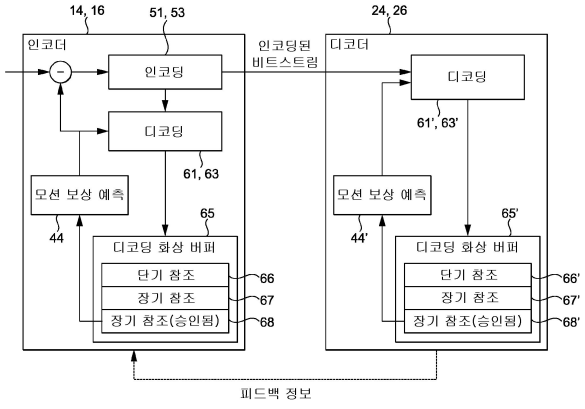
(54) 발명의 명칭 **무-에러 기준 프레임을 이용한 비디오 리프레시 기법**

(57) 요 약

복수의 프레임의 프레임 각각 내 인코딩될 복수의 타깃 이미지 부분의 타깃 이미지 부분 각각에 대해, 타깃 이미지 부분에 대한 왜곡의 추정 및 상기 타깃 이미지 부분을 인코딩하기 위해 요구되는 비트율의 양을 포함하는 함수를 최적화함으로써 인코딩 모드의 세트 중에서 선호되는 인코딩 모드를 선택하여, 율-왜곡 최적화 프로세스를

(뒷면에 계속)

대 표 도



수행하는 방법이 제공되며, 상기 왜곡의 추정은 소스 코딩 왜곡과 채널을 통할 때 일어날 수 있는 손실로 인해 겪게 될 왜곡의 추정을 기초로 하며, 선택된 인코딩 모드를 이용해 타깃 이미지 부분을 인코딩된 비디오 스트림으로 인코딩하고, 상기 인코딩된 비디오 스트림을 채널을 통해 송신한다. 상기 복수의 프레임 중 현재 프레임에 대한 상기 율-왜곡 최적화 프로세스는 상기 복수의 프레임 중 이전 프레임을 기초로 수신 단말기로부터 수신된 피드백에 따라 수행된다.

(72) 발명자

바핀 레나트

아일랜드 2 더블린 서 존 로저손스 퀘이 70 인터내셔널 페이턴츠 스카이프

제프레모브 안드레이

아일랜드 2 더블린 서 존 로저손스 퀘이 70 인터내셔널 페이턴츠 스카이프

앤더슨 소렌 뱅

아일랜드 2 더블린 서 존 로저손스 퀘이 70 인터내셔널 페이턴츠 스카이프

칼슨 폰투스

아일랜드 2 더블린 서 존 로저손스 퀘이 70 인터내셔널 페이턴츠 스카이프

명 세 서

청구범위

청구항 1

손실 채널(lossy channel)을 통한 수신 단말기의 디코더로의 송신을 위해 송신 단말기의 인코더에서 비디오 스트림을 인코딩하는 방법으로서,

상기 방법은,

복수의 프레임의 프레임 각각 내 인코딩될 복수의 타깃 이미지 부분(target image portion)의 타깃 이미지 부분 각각에 대해, 상기 타깃 이미지 부분에 대한 왜곡의 추정 및 상기 타깃 이미지 부분을 인코딩하기 위해 요구되는 비트율의 양(measure)을 포함하는 함수를 최적화함으로써, 인코딩 모드의 세트 중 선호되는 인코딩 모드를 선택하여, 율-왜곡 최적화 프로세스(rate-distortion optimization process)를 수행하는 단계 - 상기 왜곡의 추정은 소스 코딩 왜곡 및 상기 채널을 통할 때 일어날 수 있는 손실로 인해 겪게 될 왜곡의 추정을 기초로 하고, 상기 프레임 중 현재 프레임에 대한 상기 율-왜곡 최적화 프로세스는 상기 프레임 중 이전 프레임을 기초로 한 상기 수신 단말기로부터 수신된 피드백에 따라 수행됨 - 와,

상기 선택한 인코딩 모드를 이용해서 상기 타깃 이미지 부분을 인코딩된 비디오 스트림으로 인코딩하는 단계와,

상기 인코딩된 비디오 스트림을 상기 채널을 통해 송신하는 단계와,

상기 인코더에서 상기 디코더의 인코더-측 인스턴스를 실행하고, 디코딩된 화상 버퍼에 저장된 각각의 프레임과 연관된 에러 전파 왜곡 맵을 저장하는 상기 디코딩된 화상 버퍼를 상기 인코더에서 유지하는 단계와,

상기 피드백에 기초해서, 상기 에러 전파 왜곡 맵을 각각의 후속하는 인코딩 모드 선택에 사용하기 위해서 업데이트하는 단계 - 상기 업데이트는 율-왜곡 성능을 개선함 -

를 포함하고,

상기 인코딩 모드의 세트는, 상기 디코딩된 화상 버퍼에 저장된 확인 응답되지 않은(non-acknowledged) 프레임이나 혹은 프레임의 일부 내의 대응하는 장기 기준(long-term reference)에 대해서, 상기 타깃 이미지 부분을 인코딩하는 확인 응답되지 않은 장기 기준 인터 예측 모드를 포함하고,

상기 피드백에 따라 상기 율-왜곡 최적화 프로세스가 수행되는 것은, 추정 손실 확률 및 마지막으로 확인 응답된 이전 프레임 또는 이전 프레임의 확인 응답된 부분 이후의 시간에 따라서, 가중된 왜곡 추정을 구하는 것을 포함하는

비디오 스트림 인코딩 방법.

청구항 2

제 1 항에 있어서,

상기 피드백은 상기 이전 프레임의 적어도 일부분이 수신되었다는 확인 응답(acknowledgement) 및 상기 이전 프레임의 적어도 일부분이 수신되지 않았다는 리포트 중 하나를 포함하고,

상기 리포트 및 상기 확인 응답은 상기 수신 단말기의 상기 디코더에서 판정되는

비디오 스트림 인코딩 방법.

청구항 3

제 2 항에 있어서,

상기 인코딩 모드의 세트는, 상기 확인 응답된 이전 프레임 또는 상기 이전 프레임의 확인 응답된 부분 내의 대

응하는 기준 부분에 대해서 상기 타깃 이미지 부분을 인코딩하는 확인 응답된 기준 인터 예측 모드를 더 포함하고,

상기 확인 응답된 이전 프레임은 상기 이전 프레임에 대응하는 확인 응답에 의해서 확인 응답되며,

상기 이전 프레임의 상기 확인 응답된 부분은 상기 이전 프레임의 상기 적어도 일부분에 대응하는 확인 응답에 의해서 확인 응답되는

비디오 스트림 인코딩 방법.

청구항 4

제 3 항에 있어서,

상기 확인 응답된 기준 인터 예측 모드는, 상기 기준 부분이 수신된 것으로 확인 응답되고, 이에 대해서 상기 기준 부분이 인코딩된 모든 것도 역시 수신된 것으로 확인 응답되며, 이로써 상기 기준 부분이 어떠한 에러 전파도 발생시키지 않는다는 것이 알려진 것을 조건으로, 상기 인코딩 모드 선택에서 사용되도록 이용 가능한

비디오 스트림 인코딩 방법.

청구항 5

제 3 항에 있어서,

상기 피드백에 따라 상기 율-왜곡 최적화 프로세스가 수행되는 것은, 상기 확인 응답을 포함하는 상기 피드백이 수신된 것을 조건으로, 손실로 인한 상기 왜곡의 추정은 없는 것으로 설정하는 것을 포함하는

비디오 스트림 인코딩 방법.

청구항 6

제 3 항에 있어서,

상기 디코딩된 화상 버퍼는, 단기 기준과 장기 기준을, 상기 디코더의 상기 인코더-측 인스턴스에 의해 재구성된 기준 이미지 데이터의 형태로 저장하고,

상기 단기 기준은 연속적인 프레임(successive frame)에 의해 자동으로 덮어 써지고, 상기 장기 기준은 명시적 제거 명령어(explicit removal command)의 조건 하에서 제거되며,

상기 확인 응답된 기준 인터 예측 모드는, 상기 수신 단말기에서 수신되었다고 확인 응답되어 있는 상기 디코딩된 화상 버퍼 내 상기 장기 기준 중 대응하는 것에 대해서 상기 타깃 이미지 부분을 인코딩하는

비디오 스트림 인코딩 방법.

청구항 7

제 1 항에 있어서,

상기 인코딩 모드의 세트는, 적어도 인트라 프레임 모드, 적어도 하나의 확인 응답되지 않은 인터 프레임 인코딩 모드, 및 확인 응답된 인터 예측 모드를 포함하는

비디오 스트림 인코딩 방법.

청구항 8

제 1 항에 있어서,

상기 인코딩 모드의 세트는, 확인 응답되지 않은 단기 기준 인터 예측 모드를 더 포함하는

비디오 스트림 인코딩 방법.

청구항 9

제 3 항에 있어서,

상기 확인 응답된 기준 인터 예측 모드는, 상기 디코딩된 화상 버퍼 내의 상기 수신 단말기에서 수신되었다고 확인된 대응하는 단기 참조와 관련하여 상기 타깃 이미지 부분을 인코딩하는

비디오 스트림 인코딩 방법.

청구항 10

제 1 항에 있어서,

상기 인코딩 모드의 세트는, 인터 인코딩된 참조 부분으로부터 타깃 이미지 부분의 인트라 인코딩을 가능하게 하는 인트라 인코딩 모드(intra encoding mode)를 포함하는

비디오 스트림 인코딩 방법.

청구항 11

제 2 항에 있어서,

상기 피드백에 따라 상기 율-왜곡 최적화 프로세스가 수행되는 것은, 상기 확인 응답 및 상기 리포트 중 적어도 하나에 따라서 상기 이전 프레임 또는 상기 이전 프레임의 일부분에 대한 상기 왜곡의 추정을 조정하는 것, 및 상기 조정된 왜곡의 추정을 상기 현재 프레임과 관련해서 사용하도록 순방향으로(forward) 전파시키는 것을 포함하는

비디오 스트림 인코딩 방법.

청구항 12

제 1 항에 있어서,

상기 일어날 수 있는 손실로 인해 겪게 될 상기 왜곡의 추정은, 상기 타깃 이미지 부분의 예측에 의존할 상기 타깃 이미지 부분의 히스토리 내 기준 부분이 도착하지 않음으로 인해서, 상기 타깃 이미지 부분이 상기 채널을 통해서 도착하는 경우에 겪게 될 상기 왜곡의 추정을 나타내는 제 1 기여분(contribution)과, 은폐(concealment)로 인해 겪게 될 상기 왜곡의 추정을 나타내는 제 2 기여분에 기초하는

비디오 스트림 인코딩 방법.

청구항 13

제 12 항에 있어서,

상기 제 2 기여분은, 상기 타깃 이미지 부분이 채널을 통해서 손실되는 경우 상기 타깃 이미지 부분의 손실을 은폐하기 위해 사용될 이미지 부분과 관련된 상기 타깃 이미지 부분의 은폐 왜곡의 양을 나타내는 기여분 및 상기 타깃 이미지 부분의 은폐에 의존할 상기 타깃 이미지 부분의 히스토리 내 이미지 부분의 손실로 인해 겪게 될 왜곡의 추정을 나타내는 기여분을 포함하는

비디오 스트림 인코딩 방법.

청구항 14

제 12 항에 있어서,

상기 피드백에 따라 상기 율-왜곡 최적화 프로세스가 수행되는 것은, 상기 확인 응답을 포함하는 상기 피드백이 수신된 것을 조건으로 상기 이전 프레임에 대해 상기 제 2 기여분을 없는 것으로 설정하는 것, 및 수신되지 않았다는 리포트를 포함하는 피드백이 수신된 것을 조건으로 상기 이전 프레임에 대해 상기 제 1 기여분을 없는 것으로 설정하는 것 중 적어도 하나를 포함하는

비디오 스트림 인코딩 방법.

청구항 15

손실 채널을 통한 수신 단말기의 디코더로의 송신을 위해 비디오 스트림을 인코딩하는 송신 단말기로서,

복수의 프레임의 프레임 각각 내 인코딩될 복수의 타깃 이미지 부분의 타깃 이미지 부분 각각에 대해, 상기 타깃 이미지 부분에 대한 왜곡의 추정 및 상기 타깃 이미지 부분을 인코딩하기 위해 요구되는 비트율의 양을 포함하는 함수를 최적화함으로써, 인코딩 모드의 세트 중 선호되는 인코딩 모드를 선택하는 것을 포함하는, 율-왜곡 최적화 프로세스를 수행하도록 구성된 인코더와,

상기 인코딩된 비디오 스트림을 상기 채널을 통해 송신하도록 배열된 송신기

를 포함하고,

상기 왜곡의 추정은 소스 코딩 왜곡 및 상기 채널을 통할 때 일어날 수 있는 손실로 인해 겪게 될 왜곡의 추정을 기초로 하고,

상기 인코더는 또한 상기 인코더에서 상기 디코더의 인코더-측 인스턴스를 실행하고, 디코딩된 화상 버퍼에 저장된 각각의 프레임과 연관된 에러 전파 왜곡 맵을 저장하는 상기 디코딩된 화상 버퍼를 상기 인코더에서 유지하도록 더 구성되고,

상기 인코딩 모드의 세트는, 상기 디코딩된 화상 버퍼에 저장된 확인 응답되지 않은 프레임이나 혹은 프레임의 일부 내의 대응하는 장기 기준에 대해서, 상기 타깃 이미지 부분을 인코딩하는 확인 응답되지 않은 장기 기준 인터 예측 모드를 포함하고,

상기 인코더는 상기 선택한 인코딩 모드를 이용해서 상기 타깃 이미지 부분을 인코딩된 비디오 스트림으로 인코딩하도록 배치되며,

상기 프레임 중 현재 프레임에 대한 상기 율-왜곡 최적화 프로세스는 상기 프레임 중 이전 프레임을 기초로 한 상기 수신 단말기로부터 수신된 피드백에 따라 수행되고,

상기 인코더는 또한 상기 피드백에 기초해서, 상기 에러 전파 왜곡 맵을 각각의 후속하는 인코딩 모드 선택에 사용하기 위해서 업데이트해서 율-왜곡 성능을 개선하도록 더 구성되며,

상기 피드백에 따라 상기 율-왜곡 최적화 프로세스가 수행되는 것은, 추정 손실 확률 및 마지막으로 확인 응답된 이전 프레임 또는 이전 프레임의 확인 응답된 부분 이후의 시간에 따라서, 가중된 왜곡 추정을 구하는 것을 포함하는

송신 단말기.

청구항 16

제 15 항에 있어서,

상기 인코더는, 상기 피드백이 상기 이전 프레임의 적어도 일부분이 수신되었다는 확인 응답 및 상기 이전 프레

임의 적어도 일부분이 수신되지 않았다는 리포트 중 하나를 포함하도록, 더 구성되고,

상기 리포트 및 상기 확인 응답은 상기 수신 단말기의 상기 디코더에서 판정되는

송신 단말기.

청구항 17

제 16 항에 있어서,

상기 인코더는 또한, 상기 인코딩 모드의 세트가 상기 확인 응답된 이전 프레임 또는 상기 이전 프레임의 확인 응답된 부분 내의 대응하는 기준 부분에 대해서 상기 타깃 이미지 부분을 인코딩하는 확인 응답된 기준 인터 예측 모드를 포함하도록, 더 구성되고,

상기 확인 응답된 이전 프레임은 상기 이전 프레임에 대응하는 확인 응답에 의해서 확인 응답되며,

상기 이전 프레임의 상기 확인 응답된 부분은 상기 이전 프레임의 상기 적어도 일부분에 대응하는 확인 응답에 의해서 확인 응답되는

송신 단말기.

청구항 18

제 17 항에 있어서,

상기 인코더는 또한, 상기 확인 응답된 기준 인터 예측 모드가, 상기 기준 부분이 수신된 것으로 확인 응답되고, 이에 대해서 상기 기준 부분이 인코딩된 모든 것도 역시 수신된 것으로 확인 응답되며, 이로써 상기 기준 부분이 어떠한 에러 전파도 발생시키지 않는다는 것이 알려진 것을 조건으로, 상기 인코딩 모드 선택에서 사용되도록 이용 가능하도록, 더 구성되는

송신 단말기.

청구항 19

제 17 항에 있어서,

상기 인코더는 또한, 상기 피드백에 따라 상기 율-왜곡 최적화 프로세스가 수행되는 것이, 상기 확인 응답을 포함하는 상기 피드백이 수신된 것을 조건으로, 손실로 인한 상기 왜곡의 추정은 없는 것으로 설정하는 것을 포함하도록, 더 구성되는

송신 단말기.

청구항 20

제 16 항에 있어서,

상기 인코더는 또한,

상기 피드백에 따라 상기 율-왜곡 최적화 프로세스가 수행되는 것이, 상기 피드백에 따라서 상기 이전 프레임 또는 상기 이전 프레임의 일부분에 대한 상기 왜곡의 추정을 조정하는 것, 및 상기 조정된 왜곡의 추정을 상기 현재 프레임과 관련해서 사용하도록 순방향으로 전파시키는 것을 포함하도록, 또한

상기 피드백에 따라 상기 율-왜곡 최적화 프로세스가 수행되는 것이, 상기 확인 응답 및 상기 리포트 중 적어도 하나에 따라서 상기 이전 프레임 또는 상기 이전 프레임의 일부분에 대한 상기 왜곡의 추정을 조정하는 것, 및 상기 조정된 왜곡의 추정을 상기 현재 프레임과 관련해서 사용하도록 순방향으로 전파시키는 것을 포함하도록

더 구성되는

송신 단말기.

청구항 21

손실 채널을 통한 수신 단말기의 디코더로의 송신을 위해 송신 단말기의 인코더에서 비디오 스트림을 인코딩하기 위한 명령어가 저장된 컴퓨터 판독가능 저장 매체로서,

상기 명령어는, 상기 송신 단말기 상에서 실행될 때 동작을 수행하도록 구성된 저장된 코드를 포함하며,

상기 동작은,

복수의 프레임의 프레임 각각 내 인코딩될 복수의 타깃 이미지 부분의 타깃 이미지 부분 각각에 대해, 상기 타깃 이미지 부분에 대한 왜곡의 추정 및 상기 타깃 이미지 부분을 인코딩하기 위해 요구되는 비트율의 양을 포함하는 함수를 최적화함으로써, 인코딩 모드의 세트 중 선호되는 인코딩 모드를 선택하여, 율-왜곡 최적화 프로세스를 수행하는 동작 - 상기 왜곡의 추정은 소스 코딩 왜곡 및 상기 채널을 통할 때 일어날 수 있는 손실로 인해 겪게 될 왜곡의 추정을 기초로 하고, 상기 프레임 중 현재 프레임에 대한 상기 율-왜곡 최적화 프로세스는 상기 프레임 중 이전 프레임을 기초로 한 상기 수신 단말기로부터 수신된 피드백에 따라 수행됨 - 과,

상기 선택한 인코딩 모드를 이용해서 상기 타깃 이미지 부분을 인코딩된 비디오 스트림으로 인코딩하는 동작과,

상기 인코딩된 비디오 스트림을 상기 채널을 통해 송신하는 동작과,

상기 인코더에서 상기 디코더의 인코더-측 인스턴스를 실행하고, 디코딩된 화상 버퍼에 저장된 각각의 프레임과 연관된 에러 전파 왜곡 맵을 저장하는 상기 디코딩된 화상 버퍼를 상기 인코더에서 유지하는 동작과,

상기 피드백에 기초해서, 상기 에러 전파 왜곡 맵을 각각의 후속하는 인코딩 모드 선택에 사용하기 위해서 업데이트하는 동작 - 상기 업데이트는 율-왜곡 성능을 개선함 -

을 포함하고,

상기 인코딩 모드의 세트는, 상기 디코딩된 화상 버퍼에 저장된 확인 응답되지 않은 프레임이나 혹은 프레임의 일부 내의 대응하는 장기 기준에 대해서, 상기 타깃 이미지 부분을 인코딩하는 확인 응답되지 않은 장기 기준 인터 예측 모드를 포함하고,

상기 피드백에 따라 상기 율-왜곡 최적화 프로세스가 수행되는 것은, 추정 손실 확률 및 마지막으로 확인 응답된 이전 프레임 또는 이전 프레임의 확인 응답된 부분 이후의 시간에 따라서, 가중된 왜곡 추정을 구하는 것을 포함하는

컴퓨터 판독가능 저장 매체.

청구항 22

제 21 항에 있어서,

상기 피드백은 상기 이전 프레임의 적어도 일부분이 수신되었다는 확인 응답 및 상기 이전 프레임의 적어도 일부분이 수신되지 않았다는 리포트 중 하나를 포함하고,

상기 리포트 및 상기 확인 응답은 상기 수신 단말기의 상기 디코더에서 판정되는

컴퓨터 판독가능 저장 매체.

청구항 23

제 22 항에 있어서,

상기 코드는 또한, 상기 인코딩 모드의 세트가 상기 확인 응답된 이전 프레임 또는 상기 이전 프레임의 확인 응

답된 부분 내의 대응하는 기준 부분에 대해서 상기 타깃 이미지 부분을 인코딩하는 확인 응답된 기준 인터 예측 모드를 포함하도록, 더 구성되고,

상기 확인 응답된 이전 프레임은 상기 이전 프레임에 대응하는 확인 응답에 의해서 확인 응답되며,

상기 이전 프레임의 상기 확인 응답된 부분은 상기 이전 프레임의 상기 적어도 일부분에 대응하는 확인 응답에 의해서 확인 응답되는

컴퓨터 판독가능 저장 매체.

청구항 24

제 23 항에 있어서,

상기 코드는 또한, 상기 확인 응답된 기준 인터 예측 모드가, 상기 기준 부분이 수신된 것으로 확인 응답되고, 이에 대해서 상기 기준 부분이 인코딩된 모든 것도 역시 수신된 것으로 확인 응답되며, 이로써 상기 기준 부분이 어떠한 에러 전파도 발생시키지 않는다는 것이 알려진 것을 조건으로, 상기 인코딩 모드 선택에서 사용되도록 이용 가능하도록, 더 구성되는

컴퓨터 판독가능 저장 매체.

청구항 25

제 22 항에 있어서,

상기 코드는 또한, 상기 피드백에 따라 상기 율-왜곡 최적화 프로세스가 수행되는 것이, 상기 확인 응답을 포함하는 상기 피드백이 수신된 것을 조건으로, 손실로 인한 상기 왜곡의 추정은 없는 것으로 설정하는 것을 포함하도록, 더 구성되는

컴퓨터 판독가능 저장 매체.

청구항 26

제 22 항에 있어서,

상기 인코더는 또한, 상기 피드백에 따라 상기 율-왜곡 최적화 프로세스가 수행되는 것이, 상기 확인 응답 및 상기 리포트 중 적어도 하나에 따라서 상기 이전 프레임 또는 상기 이전 프레임의 일부분에 대한 상기 왜곡의 추정을 조정하는 것, 및 상기 조정된 왜곡의 추정을 상기 현재 프레임과 관련해서 사용하도록 순방향으로 전파시키는 것을 포함하도록, 더 구성되는

컴퓨터 판독가능 저장 매체.

청구항 27

제 1 항에 있어서,

상기 피드백에 따르는 것은, 상기 인코딩 모드의 세트 중 적어도 하나의 모드에 대해서 상기 피드백으로부터 결정된 시간 계산에 기초해서 확률을 업데이트하는 것 및 상기 업데이트된 확률로, 상기 채널을 통할 때 일어날 수 있는 손실로 인해 겪게 될 왜곡의 추정의 적어도 일부를 가중하는 것을 포함하는

비디오 스트림 인코딩 방법.

청구항 28

제 1 항에 있어서,

상기 피드백에 따라 상기 율-왜곡 최적화 프로세스가 수행되는 것은, 상기 피드백에 따라서 상기 이전 프레임 또는 상기 이전 프레임의 일부분에 대한 상기 왜곡의 추정을 조정하는 것, 및 상기 조정된 왜곡의 추정을 상기 현재 프레임과 관련해서 사용하도록 순방향으로 전파시키는 것을 포함하는

비디오 스트림 인코딩 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 비디오 신호의 부분들을 인코딩하기 위한 인코딩 모드를 선택할 때 비트율(bitrate)과 왜곡(distortion) 간의 트레이드 오프(trade-off)의 균형을 찾는 것과 관련된다. 본 발명은 특히, 실시간으로 비디오 스트림을 인코딩할 때, 즉, 인코더가 스트림을 카메라 등으로부터 수신하자마자 동적으로 인코딩해야 하는 실시간 비디오 스트림(live video stream), 가령, 비디오 콜(video call)의 경우에, 적용 가능할 수 있다.

배 경 기 술

[0002] 인코딩될 비디오 데이터의 스트림이 도 1a에 개략적으로 도시되어 있다. 상기 스트림은 각각 서로 다른 시점에서의 비디오 이미지를 나타내는 복수의 프레임(F)을 포함한다. 해당 분야의 통상의 기술자에게 익숙한 바와 같이, 인코딩의 목적으로, 각각의 프레임(F)은 부분(portion)들로 분할되고 각각의 부분은 또한 더 작은 서브-부분(sub-portion)으로 세분될 수 있으며, 각각의 부분 또는 서브-부분은 복수의 픽셀을 포함한다. 예를 들어, 하나의 용어에 따르면, 인코딩될 비디오 스트림의 각각의 프레임은 매크로블록(macroblock)(MB)으로 분할되고 각각의 매크로블록은 블록 또는 서브블록(b)으로 세분되며, 각각의 블록 또는 서브블록은 복수의 픽셀을 포함한다. 각각의 프레임은 또한 독립적으로 디코딩 가능한 슬라이스(slice)(S)로 분할될 수도 있고, 각각의 슬라이스는 하나 이상의 매크로블록(MB)을 포함한다. 도 1a에 도시된 분할은 단지 예시 목적으로 개략화된 것이며 임의의 실제 인코딩 스킴(encoding scheme)에 대응함을 의미하지 않고, 예컨대, 각각의 프레임은 더 많은 개수의 매크로블록을 포함할 가능성이 높음에 주의해야 한다.

[0003] 비디오 코딩이 사용될 수 있는 예시적 통신 시스템이 도 2의 블록도에서 개략적으로 도시된다. 통신 시스템은 송신 제 1 단말기(12)와 수신 제 2 단말기(22)를 포함한다. 예들 들어, 각각의 단말기(12, 22)는 모바일 폰 또는 스마트 폰, 태블릿, 랩톱 컴퓨터, 데스크톱 컴퓨터, 또는 그 밖의 다른 가정용 기기, 가령, 텔레비전 세트, 셋-톱 박스, 스테레오 시스템 등을 포함할 수 있다. 제 1 단말기 및 제 2 단말기(12, 22)는 각각 통신 네트워크(32)에 동작 가능하게 연결되며, 이에 따라 송신 제 1 단말기(12)는 신호를 송신하도록 구성되고, 상기 신호는 수신 제 2 단말기(22)에 의해 수신될 것이다. 물론 송신 단말기(12)가 수신 단말기(22)로부터 신호를 수신하는 것 및 그 반대의 경우도 가능할 수 있지만, 설명 목적으로, 본원에서 송신은 제 1 단말기(12) 관점에서 그리고 수신은 제 2 단말기(22) 관점에서 기술된다. 통신 네트워크(32)는 예를 들어, 패킷-기반 네트워크, 가령, 광역 인터넷 및/또는 로컬 영역 네트워크 및/또는 모바일 셀룰러 네트워크를 포함할 수 있다.

[0004] 제 1 단말기(12)는 저장 매체(14), 가령, 플래시 메모리 또는 그 밖의 다른 전자 메모리, 자기 저장 장치, 및/또는 광학 저장 장치를 포함한다. 제 1 단말기(12)는 하나 이상의 코어를 갖는 CPU 형태로 된 프로세싱 장치(16); 송수신기(transceiver), 가령, 적어도 송신기(18)를 갖는 유선 또는 무선 모뎀; 및 단말기(12)의 나머지와 동일한 케이스 내에 하우징되거나 하우징되지 않을 수 있는 비디오 카메라(15)를 더 포함한다. 저장 매체(14), 비디오 카메라(15) 및 송신기(18) 각각은 프로세싱 장치(16)에 동작 가능하게 연결되고, 송신기(18)는 유선 또는 무선 링크를 통해 네트워크(32)에 동작 가능하게 연결된다. 마찬가지로, 제 2 단말기(22)는 저장 매체(24), 가령, 전자, 자기, 및/또는 광학 저장 장치; 및 하나 이상의 코어를 포함하는 CPU의 형태로 된 프로세싱 장치(26)를 포함한다. 제 2 단말기는 송수신기, 가령, 적어도 수신기(28)를 갖는 유선 또는 무선 모뎀; 단말기(22)의 나머지와 동일한 케이스 내에 하우징될 수도 있고 하우징되지 않을 수도 있는 스크린(25)을 포함한다. 제 2 단말기의 저장 매체(24), 스크린(25) 및 수신기(28)는 각자의 프로세싱 장치(26)에 동작 가능하게 연결되고, 수신기(28)는 유선 또는 무선 링크를 통해 네트워크(32)에 동작 가능하게 연결된다.

[0005] 제 1 단말기(12) 상의 저장 매체(14)는 상기 프로세싱 장치(16) 상에서 실행되도록 구성된 비디오 인코더(video encoder)를 적어도 저장한다. 실행될 때, 상기 인코더는 비디오 카메라(15)로부터 "원시(raw)" (인코딩되지 않

은) 입력 비디오 스트림을 수신하고, 상기 비디오 스트림을 더 낮은 비트율 스트림으로 압축하도록 인코딩하며, 송신기(18) 및 통신 네트워크(32)를 통해 제 2 단말기(22)의 수신기(28)로 상기 인코딩된 비디오 스트림을 출력한다. 제 2 단말기(22) 상의 저장 매체는 자신의 고유 프로세싱 장치(26) 상에서 실행되도록 구성된 비디오 디코더를 적어도 저장한다. 실행될 때 상기 디코더는 수신기(28)로부터 인코딩된 비디오 스트림을 수신하고 이를 상기 스크린(25)으로 출력되도록 디코딩한다. 인코더 및/또는 디코더를 지칭하기 위해 사용될 수 있는 일반적인 용어가 코덱(codec)이다.

[0006] 비디오 코덱의 목적은 가능한 가장 높은 품질을 유지하면서 비디오 신호를 전송하는 데 필요한 비트율을 감소시키는 것이다. 이 목적은 통계적 중복(statistical redundancy) 및 (인간의 시각계의 감각과 관련된) 지각적 무관성(perceptual irrelevancy)을 활용함으로써, 달성된다.

[0007] 오늘날의 대부분의 비디오 코덱은 타 픽셀 블록들로부터의 픽셀 블록의 예측, 예측 잔차(prediction residual)의 변환, 변환 계수의 양자화, 및 양자화 인덱스(quantization indices)의 엔트로피 코딩을 포함하는 아키텍처를 기반으로 한다. 이들 단계들은 중복과 무관성을 감소시키는 데 기여한다.

[0008] 다음의 문헌들이 참조된다:

[0009] [1] ITU-T, Recommendation H.264, "Advanced video coding for generic audiovisual services", 2007;

[0010] [2] Zhang외, "Error resilience video coding in H.264 encoder with potential distortion tracking", In Proc. IEEE International Conference on Image Processing, pp.163-166, 2004;

[0011] 일반적으로 예측은 현재 프레임이 아닌 다른 비디오 프레임 내 픽셀들로부터 수행되고(인터 예측(inter prediction)), 동일 프레임 내 픽셀들로부터 수행된다(인트라 예측(intra prediction)). 즉, 인트라 프레임 인코딩(intra frame encoding)을 이용해 인코딩되는 경우, 블록, 서브블록, 또는 그 밖의 다른 프레임의 부분(타깃 블록 또는 부분)이 동일한 프레임의 또 다른 블록, 서브블록, 또는 이미지 부분(기준 블록 또는 부분)에 관련해(relative to) 인코딩되고, 인터 프레임 인코딩을 이용해 인코딩되는 경우, 타깃 블록 또는 부분은 다른 프레임 내 기준 블록 또는 부분에 관련해 인코딩된다. 일반적으로 이 프로세스는 예측 또는 예측 코딩이라고 지칭된다. 따라서 인터 또는 인트라 예측 모듈은 예측, 가령, 인트라 프레임 인코딩의 경우 이웃 블록 또는 서브블록 및/또는 인터 프레임 인코딩의 경우 모션 벡터(motion vector)의 표시의 형태로 된 예측을 생성할 것이다. 일반적으로, 인코더는 예측 블록과 실제 블록 간(또는 예측 서브블록과 실제 서브블록 간, 등) "잔재(left over)" 차이를 나타내는 잔차 신호(residual signal)를 생성한다. 그 후, 잔차, 모션 벡터 및 인트라 예측과 연관된 임의의 요구되는 데이터가 일반적으로 추가 코딩 스테이지, 가령, 양자화기(quantizer) 및 엔트로피 인코더(entropy encoder)를 통해 인코딩된 비디오 스트림으로 출력된다. 따라서 비디오의 대부분의 블록은 블록들 간 차이 면에서 인코딩될 수 있고, 이는 절대 픽셀 값을 인코딩하는 것보다 더 적은 비트를 필요로 하며, 따라서 비트율을 절약한다. 일반적으로 인트라 예측 인코딩은 인터 예측보다 더 많은 비트를 필요로 하지만, 여전히 절대 값 인코딩보다는 절약을 나타낸다. 비디오에 대한 적합한 인터 및 인트라 인코딩 기법의 세부사항은 해당 분야의 통상의 기술자에게 친숙할 것이다.

[0012] 오늘날의 코덱은 하나의 프레임 내 서로 다른 부분에 대한 서로 다른 예측 인코딩 모드의 사용을 가능하게 한다. 서로 다른 코딩 옵션을 가질 수 있음이 비디오 코덱의 율-왜곡(rate-distortion) 효율을 증가시킨다. 모든 프레임 영역에 대해 최적 코딩 표현이 발견되어야 한다. 일반적으로, 이러한 영역은 가령 16×16 픽셀의 매크로블록(macroblock)이다. 즉, 매크로블록 각각에 대해 개별적으로 인트라 예측 또는 인터 예측 모드가 선택되어, 동일 프레임 내 서로 다른 매크로블록이 서로 다른 모드로 인코딩될 수 있는 것이 가능하다. 또한 일부 코덱에서 매크로블록의 서로 다른 레벨의 분할을 기초로 서로 다른 모드를 이용하는 것, 가령, 매크로블록 내 4×4 서브블록 각각에 대해 개별 예측이 수행되는 고 복잡도 모드 또는 8×8 또는 8×16 블록만 기초로 또는 심지어 전체 매크로블록을 기초로 예측이 수행되는 저 복잡도 모드를 선택하는 것이 가능하다. 또한 이용 가능한 모드는 예측을 수행하기 위한 서로 다른 옵션을 포함할 수 있다. 도 1b에 개략적으로 도시된 예시의 경우, 하나의 인트라 모드에서, 4×4 서브블록(b)의 픽셀이, 바로 위 서브블록으로부터의 이웃하는 픽셀들로부터 하향 외삽(extrapolate)함으로써, 또는 바로 좌측 서브블록으로부터의 이웃하는 픽셀로부터 측향 외삽함으로써, 결정될 수 있다. 일부 코덱에서 "스킵 모드(skip mode)"라고 지칭되는 또 다른 특수 예측 모드가 제공될 수 있으며, 이는 인터 모드의 대안적 유형이라고 간주될 수 있다. 스킵 모드(PSkip)에서, 타깃의 모션 벡터가 상단 및 좌측까지의 모션 벡터를 기초로 추론되고 잔차 계수의 어떠한 인코딩도 존재하지 않는다. 모션 벡터가 추론되는 방식은 모션 벡터 예측과 일관되며, 따라서 모션 벡터 차이가 0이고 따라서 매크로블록이 스킵 블록임을 시그널링하는 것만 요구된다.

[0013] 도 3은 인코더, 가령, 송신 단말기(12)에서 구현될 수 있는 인코더를 개략적으로 도시하는 하이-레벨 블록도이다. 상기 인코더는, 이산 코사인 변환(DCT: discrete cosine transform) 모듈(51), 양자화기(53), 역방향 변환 모듈(inverse transform module)(61), 역방향 양자화기(63), 인트라 예측 모듈(41), 인터 예측 모듈(43), 및 감산 스테이지(subtraction stage)(-)를 포함한다. 인코더는 또한 스위치(47) 및 모드 선택 모듈(49)을 포함한다. 각각의 모듈은 송신 단말기의 저장 매체(14) 상에 저장되고 코드의 일부분으로서 구현되고 자신의 프로세싱 장치(16) 상에서 실행되도록 구성되는 것이 바람직하지만, 전용 하드웨어 회로에서 이들 중 일부 또는 전부가 전체적으로 또는 부분적으로 구현될 가능성이 배제되지 않는다.

[0014] 스위치(47) 및 모드 선택 모듈(49) 각각이 복수의 매크로블록(MB)을 포함하는 입력 비디오 스트림의 인스턴스를 수신하도록 구성된다. 상기 모드 선택 모듈(49)은 각각의 매크로블록에 대해 코딩 모드 "o"를 선택하도록 구성되고 멀티플렉서(47)에 동작 가능하게 연결되어 역방향 양자화기(63)의 출력을, 선택된 모드에 따라 적절하게, 인트라 예측 모듈(41)의 입력 또는 인터 예측 모듈(43)의 입력으로 전달하도록 제어할 수 있다. 또한 모드 선택 모듈(49)은 관련 예측 모듈(41, 43)에게 선택된 모드 "o"를 나타내고(가령, 4×4 파티션 모드, 8×8 모드, 스킵 모드 등을 나타내고), 다음 프레임에 대한 모드를 선택할 때 사용되기 위해 예측 모듈(41, 43)로부터 피드백되는 정보를 수신하도록 구성될 수 있다. 그 후 인트라 예측 모듈(41) 또는 인터 예측 모듈(43)의 출력이 감산 스테이지(-)의 입력으로 연결되며, 상기 감산 스테이지는 자신의 다른 입력에서 인코딩되지 않은 입력 비디오 스트림을 수신하고 이들의 인코딩되지 않은 대응물(unencoded counterpart)로부터 예측 블록을 빼서, 잔차 신호를 생성하도록 구성된다. 그 후 잔차 블록이 변환(DCT) 모듈(51)을 통해 전달되어 이들의 잔차 값이 주파수 영역으로 변환되고, 그 후 양자화기(53)로 전달되어 변환된 값이 이산 양자화 인덱스로 변환된다. 변환되고 양자화된 신호가 역방향 양자화기(63) 및 역방향 변환 모듈(61)을 통해 피드백되어 선택된 예측 모듈(41, 43)에 의해 사용되기 위한 (디코더에서 보일) 블록 또는 서브블록의 예측 버전을 생성할 수 있다. 예측 모듈(41, 43)에서 사용되는 예측의 표시, 인터 예측 모듈(43)에 의해 생성된 모션 벡터 및 변환 및 양자화 모듈(51, 53)에 의해 생성되는 잔차의 변환되고 양자화된 인덱스는 모두, 일반적으로, 추가적인 무손실 인코딩 스테이지, 가령, 해당 분야에 공지되어 있는 무손실 인코딩 기법을 이용해 예측 값과 양자화되고 변환된 인덱스가 추가로 압축될 수 있는 엔트로피 인코더(도시되지 않음)를 통해, 인코딩된 비디오 스트림에 포함되도록 출력된다.

[0015] 상기의 내용에 따르면, 따라서 코딩 표현은 블록 파티션 정보, 예측 모드, 모션 벡터, 양자화 정확도(quantization accuracy) 등을 포함할 수 있다. 최적 코딩 옵션은 비디오 콘텐츠, 비트율, 사전 코딩 판정 등에 따라 달라진다. 일반적으로 변환 계수의 양자화의 정확도는 비트율 제한(bit rate constraint)을 충족시키도록 선택된다. 덧붙여, 왜곡이 최소화되어야 한다.

[0016] 예를 들어, H.264 비디오 코더가 예측 모드를 선택하는 데 우수한 유연성을 제공한다[1]. 루마 성분(luma component)의 인터 예측을 위해, 16×16 픽셀의 매크로블록은 16×16 픽셀의 블록 하나, 또는 16×8 픽셀의 블록 2개, 또는 8×16 픽셀의 블록 2개, 또는 8×8 픽셀의 블록 4개로서 표현될 수 있다. 덧붙여, 8×8 블록은 8×8 픽셀의 블록 하나, 또는 8×4 픽셀의 서브블록 2개, 또는 4×8 픽셀의 서브블록 2개, 또는 4×4 픽셀의 서브블록 4개로 표현될 수 있다. 매크로블록의 허용된 파티션 각각에 대해 인터 예측이 시도된다. 한 블록의 인터 예측은 기준 프레임(들) 및 모션 벡터(들)(각각의 기준 프레임 내에서 기준 블록으로부터의 공간 이격거리(spatial shift))을 인덱싱함으로써 표현되며, 이는 통상적으로 서브-픽셀 정밀도에 의해 추정된다. 루마 성분의 인트라 예측을 위해, 16×16 블록의 경우 네 가지 가능한 모드가 존재하고 4×4 서브블록의 경우 아홉 가지 가능한 모드가 존재한다. 덧붙여, 크로마 성분(chroma component)에 대해 네 가지 가능한 모드가 존재한다. 인터 예측 모드와 인트라 예측 모드의 성능을 비교함으로써 최상의 예측 모드가 선택된다.

[0017] 비디오 코덱, 가령, H.264 AVC [1]의 율-왜곡 성능은 매크로블록 모드 선택 o의 성능에 크게 좌우된다. 즉, 가령, 인트라 모드 또는 인터 모드를 이용할 때 율-왜곡 트레이드 오프의 측면에서 매크로블록이 최상으로 인코딩되는지 여부를 결정하는 절차이다. 강건성(robustness) 관점에서, (제한된 인트라 예측, 즉, 인터 예측된 매크로블록으로부터의 인트라 예측의 사용이 금지된다고 가정할 때) 인트라 코딩된 매크로블록이 시간 에러 전파(temporal error propagation)를 중단시키기 때문에 유익하다. 그러나 일반적으로 인트라 코딩된 매크로블록이 인터 코딩된 매크로블록에 비교할 때 율(rate)의 측면에서 고비용이며, 따라서 특정 비트 예산(bit budget) 및 채널 상태가 주어질 때 디코더에서의 왜곡(가령, 평균 왜곡)이 최소화되도록 인트라 코딩된 매크로블록을 체계적으로 도입하는 것이 중요하다. [2](Zhang외)는 디코더에서의 기대 평균 SSD(sum of squared difference)의 최소화를 기초로 하여 인트라 코딩된 매크로블록을 도입하기 위해 이러한 체계적 프레임워크를 제안한다. 잠재 왜곡을 추적함으로써, Zhang외는 인코더 율-왜곡 루프 내에서 인터 매크로블록에 대한 비용을 계산할 때, 소스 코딩 왜곡에 추가된 (디코더에서의) 기대 에러-전파 왜곡과 관련된 바이어스 항(bias term)을 계산할 수 있다.

[0018] 상기 율-왜곡 성능 최적화 문제는 비트율 제한 R 하의 왜곡을 최소화하는 측면에서 공식화될 수 있다. 종종 라그랑주(Lagrangian) 최적화 프레임워크가 문제를 해결하기 위해 사용되며, 이에 따라서 최적화 기준(optimization criterion)이 다음과 같이 공식화될 수 있다:

[0019] $$J = D(m, o) + \lambda R(m, o) \quad (1)$$

[0020] 여기서, J는 라그랑주 함수를 나타내고, D는 왜곡의 양(모드 o와 매크로블록 m 또는 매크로블록 서브-파티션의 함수)를 나타내며, R은 비트율이고, λ는 왜곡과 율 간의 트레이드 오프를 정의하는 파라미터이다. 흔히 사용되는 왜곡 양은 원본 픽셀과 재구성된 픽셀 간 SSD(sum of squared difference) 또는 원본 픽셀과 예측 픽셀 간 SAD(sum of absolute difference)이다.

[0021] 이 적용예에서 라그랑주 최적화 문제를 해결하는 것이 라그랑주 함수 J를 최소화하는 인코딩 모드 o를 찾는 것을 의미하며, 여기서, 라그랑주 함수 J는 왜곡을 나타내는 항, 비트율을 나타내는 항, 및 둘 사이의 트레이드 오프(tradeoff)를 나타내는 팩터(factor)("라그랑주 승수(Lagrange multiplier)") 을 적어도 포함한다. 인코딩 모드 o가 더 철저하거나 더 우수한 품질의 인코딩 모드를 위해 변함에 따라, 왜곡 항 D는 감소될 것이다. 그러나 동시에 율 항 R이 증가할 것이고, λ에 종속적인 특정 포인트에서 R의 증가가 D의 감소를 능가할 것이다. 따라서 수식 J가 일부 최소 값을 가질 것이며, 이가 발생하는 인코딩 모드 o가 최적 인코딩 모드라고 여겨진다.

[0022] 이러한 관점에서, 비트율 R, 또는 심지어 항 λR은 이 항이 최적 인코딩 모드를 계속 증가하는 품질로부터 후퇴하게 한다는 점에서 최적화를 제한한다. 이 최적 균형이 발견되는 모드가 λ에 종속적일 것이며, 따라서 λ는 비트율과 왜곡 간의 트레이드 오프를 나타낸다고 여겨질 수 있다.

[0023] 코딩 판정을 선택하는 프로세스에서 라그랑주 최적화가 흔히 사용되고, 매 프레임 영역(가령, 16×16 픽셀의 매크로블록)에 적용된다. 일반적으로, 왜곡은 모든 프로세싱 스테이지를 설명하도록 평가될 수 있다. 이들은 예측, 변환, 및 양자화를 포함한다. 덧붙여, 재구성된 픽셀을 계산하기 위해, 역방향 양자화, 역방향 변환, 및 역방향 예측의 단계가 수행되어야 한다. 종종 SSD가, SAD에 비교할 때 더 고품질을 도출하기 때문에 왜곡 기준으로 선호된다. 흔히, 또한 율은 모든 필요한 파라미터, 가령, 예측 및 양자화된 변환 계수를 기술하는 파라미터의 코딩을 설명한다[4].

[0024] [2]에서, Zhang외, 즉 저자들은 소스 코딩뿐 아니라 채널 에러에 의한 디코더에서의 잠재적 왜곡, 즉, 신호가 채널을 통해 송신될 때 데이터의 손실로 인해 겪게 될 가능성이 있는 왜곡을 추정한다. 그 후 상기 추정된 잠재적 왜곡은 (채널 에러의 확률이 존재하는 경우) 모드 선택을 인트라 코딩 쪽으로 편향시키기 위해 간접적으로 사용된다.

[0025] Zhang의 "종단간(end-to-end)" 왜곡 수식은 SSD(sum of squared difference) 왜곡 양을 기초로 하며, 손실된 매크로블록에 대한 베르누이(Bernoulli) 분포를 가정한다. 최적 매크로블록 모드 o_{opt}는 다음과 같이 주어진다:

[0026] $$o_{opt} = \arg\min_{o}(D_s(m,o) + D_{ep_ref}(m,o) + \lambda R(m,o)) \quad (2)$$

[0027] 여기서, $D_s(m,o)$는 매크로블록 m 및 매크로블록 모드 o에 대한 원본 픽셀과 재구성된 픽셀 간의 SSD 분포를 지칭하고, R은 총 율(total rate)이고, λ는 왜곡과 율 항과 관련된 라그랑주 승수이다. $D_{ep_ref}(m,o)$는 에러 전파로 인한 디코더 내 기준 매크로블록 내에서의 기대 왜곡을 나타낸다. 따라서 $D_{ep_ref}(m,o)$는 에러 전파 왜곡이 너무 커질 경우 최적화를 인트라 코딩 쪽으로 편향시키는 바이어스 항을 제공한다. 인트라 코딩된 매크로블록 모드에 대해 $D_{ep_ref}(m,o)$는 0(zero)이다. 수식 $D_s(m,o) + D_{ep_ref}(m,o) + \lambda R(m, o)$ 은 라그랑주 함수 J의 하나의 사례라고 여겨질 수 있다. $Argmin_o$ 는 수식 J의 값이 최소일 때의 논항(argument) o의 값을 출력한다.

[0028] [2]에서 항 $D_{ep_ref}(m,o)$은 객체의 모션을 따르며 현 모션 벡터를 이용해 총 왜곡 맵(total distortion map)으로부터 계산된다. 총 기대 에러 전파 왜곡 맵 D_{ep} 는 에러 은폐(error concealment)의 수행에 의해 구동되고 각각의 매크로블록 선택 후 다음과 같이 업데이트된다:

[0029] $$D_{ep}(m(k), n+1) = (1-p)D_{ep_ref}(m(k), n, o_{opt}) + p(D_{ec\text{-}rec}(m(k), n, o_{opt}) + D_{ec\text{-}ep}(m(k), n))$$

[0030] (3)

[0031] 여기서, n은 프레임 번호이고, m(k)는 매크로블록 m의 k번째 서브-파티션(즉, 블록 또는 서브블록)을 지칭하며, p는 패킷 손실의 확률이고, D_{ec-rec}는 인코더에서 재구성된 픽셀과 에러 은폐된 픽셀 간의 SSD를 나타내며, D_{ec-ep}는 인코더와 디코더에서의 에러 은폐된 픽셀의 기대 SSD이다.

[0032] [2]에서 D_{ep}는 프레임의 매크로블록 각각에 걸친 4×4 그리드 상에 저장되는데, 즉, 매크로블록 당 D_{ep}의 값이 16개이며, 따라서 각각의 매크로블록의 4×4 픽셀 서브블록 당 D_{ep} 값이 1개이다. 도 1c에서 도시된 바와 같이, 그 후 $D_{ep_ref}(m(k),o)$의 계산, 즉, 시점 n에서의 프레임의 매크로블록 m 내의 서브블록 k에 대한 기대 에러-전파 기준 왜곡의 계산이, 시점 n-1의 선행 프레임으로부터의 4개의 서브블록에서의 D_{ep}의 값들 가중합(weighted sum)으로서 수행된다. 가중치는 관심 블록 m에 대한 모션 벡터로부터 결정된다. 즉, 다음과 같다:

[0033] $$D_{ep_ref}(m(k),n)=\sum_{i=1}^{4} w_i D_{ep}\left(q_i(k_i),n-1\right) \quad (4)$$

[0034] 여기서, 가중치 w_i는 겹치는 면적에 비례하고, $q_i(k_i)$는 선행하는 프레임 n-1에서의 매크로블록 q_i의 서브블록 k_i를 나타낸다.

[0035] 도 1c는 예시적 서브블록 b1 ... b4를 참조하여(이 예시에서 k는 b1에 대응하고 i는 b1 ... b4를 카운트함), 모션 벡터 및 기대 에러-전파 왜곡 맵으로부터의 기대 에러-전파 기준 왜곡의 계산을 도시한다.

발명의 내용

해결하려는 과제

과제의 해결 수단

[0036] Zhang외 [2]의 프로세스는 채널을 통한 손실의 가능도에 대해 순수하게 인코더에서 만들어진 선험적 확률 가정만을 기초로 한다.

[0037] 그러나, 일부 기존 통신 시스템은 수신기에서 송신기로 특정 정보를 되돌려 리포팅하기 위한 목적 및/또는 제어 목적으로 피드백 메커니즘을 제공한다. 예를 들어, 인코더는 디코더에서 올바르게 도착한 프레임이 무엇인지에 대한 정보 및/또는 송신 시 손실된 프레임이 무엇인지에 대한 정보를 되돌려 수신할 수 있고, 이에 응답하여 에러 전파를 중단시키기 위해 인트라 프레임을 생성할 수 있다. 그러나 기존 메커니즘은 단지 전체 인트라 프레임의 생성만 트리거한다는 점에서 더무 단순하고, 또한 확인 응답(acknowledge)될 예정이거나 손실로 리포트된 그 밖의 다른 프레임 또는 프레임의 일부분의 손실로 인해 겪게 될 가능성이 높은 왜곡의 임의의 확률 추정을 포함하지 않는다.

[0038] Zhang외 [2]에 의한 알고리즘은 임의의 피드백의 가능한 사용 또는 가용성을 고려하지 않고 에러 채널(erroneous channel)을 통해 비디오를 송신하는 시나리오만 고려하며, 따라서 Zhang의 프로세스는 채널의 어떠한 실제 경험적 지식(a posteriori knowledge)도 기초로 하지 않는다.

[0039] 그 밖의 다른 관점에서 보자면, 피드백의 종래의 사용은 전체 인트라 프레임의 생성만 트리거하며, 하나의 프레임 내 개별 부분 레벨(가령, 매크로블록(macroblock) 단위)에서는 어떠한 모드 선택도 없다. 또한, 종래의 피드백의 사용은 채널을 통할 때 일어날 수 있는 손실로 인해 겪게 될 왜곡의 추정을 포함하지 않는다.

[0040] 다른 한편으로, 본 발명은 디코더에서 인코더로 피드백되는 정보, 가령, 패킷 및/또는 프레임 도착 상태를 사용하여, 인코더에서 손실-적응적 율-왜곡 최적화 프로세스를 추가로 적응시켜, Zhang외에 의한 방법에 비교할 때 전체적인 율-왜곡 성능을 개선하는 것을 목적으로 가진다.

[0041] 본 발명의 첫 번째 실시예는 단기 기준(short-term reference) 및 장기 기준(long-term reference)의 시스템을 사용할 수 있다. 예를 들어, H.264 AVC 표준은 특정 기준 프레임을 이른바 "장기" 기준으로 만드는 기능을 지원한다. 이들 장기 기준은 명시적으로 제거될 때까지 디코딩 화상 버퍼(decoded picture buffer)에 유지된다. 이

는, 디코딩 화상 버퍼에서 새로운 단기 기준이 가장 오래된 단기 기준 프레임을 덮어 쓰는 "단기" 기준 프레임과 반대이다.

[0042] 본 발명의 첫 번째 실시예에 따라, 피드백 메커니즘은 인코더가 (손실된 프레임이 무엇인지에 대한 정보에 추가로) 디코더에서 이용 가능한 가장 마지막에 확인 응답된 장기 기준이 무엇인지를 알게 하도록 사용될 수 있다. 이하에서, 확인 응답된 기준은 자체만 확인 응답된 기준이 아닌 확인 응답된 무-에러(error-free) 기준(어떠한 에러-전파 왜곡도 없는 기준)을 의미한다. 즉, 기준은, 기준의 히스토리의 엄격한 확인응답 없이 현재 기준만 수신된 것으로 확인되는 것과 다른, 기준이 수신된 것으로 확인되고, 상기 기준의 히스토리 내 관련 모든 것도 역시 수신된 것으로 확인되어, 어떠한 에러 전파도 없다고 알려질 수 있는 엄격한 정의에 따라 확인되어야 하는 것이 바람직하다. 자체만 확인 응답된 프레임 내의, 무-에러 블록과 관련해 인코딩된 부분이 역시 무-에러이다(어떠한 전파 에러도 없음).

[0043] 확인 응답된 장기 기준 프레임을 기초로 하는 인터 예측은 인트라 코딩과 유사하게 디코더에서의 에러 전파를 중단시키도록 사용될 수 있다. 장기 기준으로부터의 인터 예측을 이용하는 것의 이점은 일반적으로 인터 예측은 특정 왜곡 레벨에 대해 더 낮은 비트율을 도출한다는 것이다.

[0044] 인터 예측에 대해 확인 응답된 장기 기준을 이용함으로써, 본 발명의 이 첫 번째 실시예는 인트라 코딩과 유사하게 에러 전파를 중단시킬 수 있지만, 일반적으로 더 낮은 비트율과 연관된 추가적인 매크로블록 코딩 모드가, 예를 들어, Zhang [2]의 프레임워크에서 사용될 수 있게 한다.

[0045] Zhang외 [2]에 의한 알고리즘은 단 두 개의 서로 다른 유형의 코딩 모드, 즉, 인트라 코딩과 인커 코딩을 고려한다. 이 경우, 수식(3)에서의 에러-전파 기준 왜곡 $D_{ep\text{-}ref}(m,o)$이 인트라 코딩된 매크로블록 모드에 대해서만 0(zero)이다. 그러나 본 발명의 첫 번째 실시예는 확인 응답된 장기 기준으로부터의 인터 코딩을 포함하도록 이용 가능한 코딩 모드의 세트를 증강시킨다. 그 후, $D_{ep\text{-}ref}(m,o)$가 인트라 코딩에 대해서뿐 아니라 확인 응답된 기준으로부터의 인터 코딩에 대해서도 0으로 설정된다. 이러한 코딩 모드의 이점은 인트라 코딩과 유사한 방식이지만 일반적으로 더 낮은 비트율에서 에러 전파를 중단시킬 수 있다는 것이다.

[0046] 본 발명의 첫 번째 실시예의 변형은, 어떠한 피드백도 아직 수신되지 않은 특정 기준(가령, 장기 기준)에 대해, 그럼에도 불구하고, 이 기준이 디코더로부터의 중간 리포트를 기초로 무-에러 확인 응답된 기준이 될 특정 확률을 가진다는 아이디어를 이용한다. 예를 들어, 확인 응답되지 않은 장기 기준을 기초로 하는 인터 예측의 이용을 확인 응답되지 않은 단기 기준을 기초로 하는 인터 예측의 이용으로부터 구별하는 또 다른 이용 가능한 인코딩 모드가 도입될 수 있다. 확인 응답되지 않은 장기 기준에 대해, 에러 전파 왜곡의 추정 $D_{ep\text{-}ref}(m,o)$는 손실의 확률의 선험적 추정에 종속적으로(즉, p 자체는 피드백을 기초로 하지 않음) 그리고 확인 응답되지 않은 기준의 히스토리 내 마지막(가장 최근) 확인 응답된 장기 기준 이후로부터의 시간을 기초로 감소된다.

[0047] 첫 번째 실시예의 또 다른 변형에서, 디코딩 화상 버퍼 내 단기 기준의 개수에 비교할 때 패킷의 왕복 시간(RTT)이 충분히 낮을 때(왕복 시간은 패킷이 송신기에서 수신기로 이동하고 다시 돌아오는 데 걸리는 시간이다), 동일한 개념이 단기 기준에 적용될 수 있다. 즉, 충분히 작은 RTT에 대해, 대안적 또는 추가적으로, 단기 기준을 확인 응답된 것으로 인식하고, 그 후 상기 단기 기준을 앞서 언급된 확인 응답된 장기 기준과 유사한 방식으로 사용될 수 있는 것이 가능하다. 다시 말하면, 요구되는 수식(3)의 알고리즘 변경은 $D_{ep\text{-}ref}(m,o)$가 인트라 코딩뿐 아니라 확인 응답된 기준으로부터의 인터 코딩에 대해서도 0으로 설정되는 것이다.

[0048] 또한 확인응답은 전체 프레임 단위로 이뤄질 필요는 없다. 대신, 프레임의 일부분, 가령, 슬라이스에 대해 확인응답을 수신하고, 확인응답 또는 확인응답의 부재(또는 수신되지 않음에 대한 명시적 리포트) 에 따라 서로 다른 일부분들을 서로 다르게 취급하도록 구성될 수 있다.

[0049] 본 발명의 두 번째 실시예에서, 디코더에서 인코더로 피드백되는 정보, 가령, 패킷 및/또는 프레임 도착 상태가 인코더에서 가능한 왜곡 맵을 조정하고, 이에 따라 Zhang외의 방법에 비교할 때 전체적은 율-왜곡 성능을 개선하도록 사용된다.

[0050] 두 번째 실시예에 따라, 가능한 에러 전파 왜곡 맵은 인코더의 디코딩 화상 버퍼에 프레임 또는 슬라이스 각각과 연관되어 (에러 은폐 재구성 왜곡 맵, 에러 은폐 에러 전파 맵, 대응하는 모드 결정 및 모드 벡터 정보와 함께) 저장된다. 그 후 이 두 번째 실시예는 디코더로부터의 피드백 정보를 이용하여 가능한 왜곡 맵을 업데이트할 수 있다. 상기 피드백 정보는 정제된 가능한 왜곡 추적을 촉진시켜 더 우수한 율-왜곡 성능을 도출할 수 있다.

[0051] 인코더가 특정 프레임이 디코더에서 도착했음을 시그널링하는 피드백 정보를 수신한 경우, 에러 은폐 기여분(error concealment contribution)이 수식(3)의 에러 전파 왜곡 맵에서 삭제될 수 있다. 역으로, 특정 프레임 또는 슬라이스가 디코더에서 손실되었다고 시그널링하는 피드백 정보가 수신된 경우, 에러 은폐 왜곡으로부터의 기여분, 즉, (p로 정규화된) 수식(3)의 우변의 두 번째 항과 세 번째 항만 포함하도록 연관된 에러 전파 왜곡 맵이 재계산된다.

[0052] 그 후, 왕복 시간(RTT)이 디코딩 화상 버퍼 내 기준 화상의 개수와 비교해서 작을 때, 수식(3)을 재귀적으로 이용하여 시점 n-RTT에서 조정된 잠재적 에러-전파 맵을 시점 n-1에서의 에러 전파 왜곡 맵으로 전파시키는 것이 가능하다. 그 후 시점 n-1에서의 업데이트된 에러 전파 왜곡 맵이 모드 선택 프로세스 (2)에서 사용되는 시점 n에서의 D_{ep_ref}의 계산의 기초가 될 것이다. 이는 잠재적 왜곡 맵의 더 정확한 추적을 도출하고, 따라서 시스템의 전체적인 율-왜곡 성능을 개선한다.

[0053] 지금까지 특정 예시적 실시예를 간략히 기재했지만, 더 일반적으로 본 발명은 이하의 내용에서 제공되느 시스템, 컴퓨터 프로그램 프로덕트, 및 장치를 제공한다.

[0054] 본 발명의 하나의 양태에 따르면, 손실 채널(lossy channel)을 통한 수신 단말기의 디코더로의 송신을 위해 송신 단말기의 인코더에서 비디오 스트림을 인코딩하는 방법이 제공되며, 상기 방법은 복수의 프레임의 프레임 각각 내 인코딩될 복수의 타깃 이미지 부분의 타깃 이미지 부분 각각에 대해, 타깃 이미지 부분에 대한 왜곡의 추정 및 상기 타깃 이미지 부분을 인코딩하기 위해 요구되는 비트율의 양을 포함하는 함수를 최적화함으로써 인코딩 모드의 세트 중에서 선호되는 인코딩 모드를 선택하여 율-왜곡 최적화 프로세스를 수행하는 단계 - 상기 왜곡의 추정은 소스 코딩 왜곡 및 채널을 통할 때 일어날 수 있는 손실로 인해 겪게 될 왜곡의 추정을 기초로 함 - 선택된 인코딩 모드를 이용해 상기 타깃 이미지 부분을 인코딩된 비디오 스트림으로 인코딩하는 단계, 및 채널을 통해 인코딩된 비디오 스트림을 송신하는 단계 - 상기 프레임 중 현재 프레임에 대한 상기 율-왜곡 최적화 프로세스는 상기 프레임 중 이전 프레임을 기초로 수신 단말기로부터 수신된 피드백에 따라 수행됨 - 를 포함한다.

[0055] 상기 피드백은 상기 이전 프레임의 적어도 일부분이 수신되었다는 확인응답, 및 상기 이전 프레임의 적어도 일부분이 수신되지 않았다는 리포트 중 하나를 포함할 수 있다.

[0056] 본 발명의 첫 번째 실시예에서, 인코딩 모드의 세트는 타깃 이미지 부분을 확인 응답된 이전 프레임 또는 상기 이전 프레임의 확인 응답된 부분 내 대응하는 기준 부분과 관련하여 인코딩하는 확인 응답된 기준 인터 예측 모드를 포함한다.

[0057] 상기 확인 응답된 기준 인터 예측 모드는, 기준 부분이 수신된 것으로 확인 응답되고, 이에 대해서 기준 부분이 인코딩된 모든 것도 역시 수신된 것으로 확인 응답되며, 이로써 기준 부분이 어떠한 에러 전파도 발생시키지 않는다는 것이 알려진 것을 조건으로, 인코딩 모드 선택에서 사용되도록 이용 가능할 수 있다.

[0058] 상기 피드백에 따르는 손실-적응적 율-왜곡 최적화 프로세스의 수행은, 상기 확인 응답을 포함하는 피드백이 수신된 경우, 손실로 인한 왜곡의 추정을 없는 것으로 설정하는 것을 포함할 수 있다.

[0059] 상기 방법은, 인코더에서 디코더의 인코드-측 인스턴스를 실행하는 단계, 및 단기 및 장기 기준을 디코더의 상기 인코더-측 인스턴스에 의해 재구성된 기준 이미지 데이터의 형태로 저장하는 디코딩 화상 버퍼를 인코더에 유지하는 단계를 포함하며, 상기 단기 기준은 연속적인 프레임에 의해 자동으로 덮어 써질 수 있는 반면에, 장기 기준은 명시적인 제거 명령어가 있는 경우 제거될 수 있고, 확인 응답된 기준 인터 예측 모드는 상기 디코딩 화상 버퍼 내의 장기 기준들 중, 수신 단말기에서 수신된 것으로 확인 응답된 대응하는 장기 기준과 관련하여 상기 타깃 이미지 부분을 인코딩할 수 있다.

[0060] 인코딩 모드의 세트는 인트라 프레임 모드, 적어도 하나의 확인 응답되지 않은 인터 프레임 인코딩 모드, 및 상기 확인 응답된 인터 예측 모드를 적어도 포함할 수 있다.

[0061] 상기 인코딩 모드의 세트는, 디코딩 화상 버퍼에 저장된 확인 응답되지 않은 프레임 또는 프레임의 부분 내 대응하는 장기 기준에 관련하여 상기 타깃 이미지 부분을 인코딩하는 확인 응답되지 않은 장기 기준 인터 예측 모드를 포함하고, 상기 피드백에 따르는 상기 손실-적응적 율-왜곡 최적화 프로세스의 수행은, 추정 손실 확률 및 마지막 확인 응답된 이전 프레임 또는 이전 프레임의 확인 응답된 부분 이후의 시간에 따라서, 손실로 인한 왜곡의 추정을 구하는 것을 포함한다.

[0062] 인코딩 모드의 세트는 확인 응답되지 않은 단기 기준 인터 예측 모드를 더 포함할 수 있다.

[0063] 상기 방법은 인코더에서 디코더의 인코더-측 인스턴스를 실행하는 단계, 및 단기 기준 및 장기 기준을 상기 디코더의 상기 인코더-측 인스턴스에 의해 재구성된 기준 이미지 데이터의 형태로 저장하는 디코딩 화상 버퍼를 인코더에 유지하는 단계를 포함하고, 상기 단기 기준은 연속적인 프레임에 의해 자동으로 덮어 써질 수 있는 반면에, 장기 기준은 명시적 제거 명령가 있을 경우 제거될 수 있고, 상기 확인 응답된 기준 인터 예측 모드는 상기 디코딩 화상 버퍼 내 상기 수신 단말기에서 수신된 것으로 확인 응답된 대응하는 단기 기준에 관련하여 상기 타깃 이미지 부분을 인코딩할 수 있다.

[0064] 상기 인코딩 모드의 세트는 인터 인코딩된 기준 부분으로부터 타깃 이미지 부분의 인트라 인코딩을 가능하게 하는 비제한 인트라 인코딩 모드를 포함할 수 있다.

[0065] 본 발명의 두 번째 실시예에서, 상기 피드백에 따르는 손실-적응적 율-왜곡 최적화 프로세스의 수행은, 상기 피드백에 따라 이전 프레임 또는 이전 프레임의 일부분에 대해 왜곡의 추정을 조정하는 것, 및 조정된 왜곡의 추정을 현재 프레임과 관련해 사용되도록 순방향으로(forward) 전파시키는 것을 포함할 수 있다.

[0066] 일어날 수 있는 손실로 인해 겪게 될 왜곡의 추정은, 타깃 부분이 채널을 통해 도착하는 경우에, 타깃 부분의 히스토리 내, 상기 타깃 부분의 예측에 의존하는 기준 부분의 도착하지 않음으로 인해 겪게 될 왜곡의 추정을 나타내는 제 1 기여분과, 은폐로 인해 겪게 될 왜곡의 추정을 나타내는 제 2 기여분을 기초로 할 수 있다.

[0067] 상기 제 2 기여분은 타깃 부분이 채널을 통해 손실된 경우 상기 타깃 부분의 손실을 은폐하기 위해 사용될 이미지 부분에 관련된 타깃 부분의 은폐 왜곡의 양를 나타내는 기여분과, 타깃 부분의 은폐가 따를 타깃 부분의 히스토리 내 이미지 부분의 손실로 인해 겪게 될 왜곡의 추정을 나타내는 기여분을 포함할 수 있다.

[0068] 상기 피드백에 따르는 손실-적응적 율-왜곡 최적화 프로세스의 수행은, 상기 확인응답을 포함하는 상기 피드백이 수신된 경우 이전 프레임에 대해 상기 제 2 기여분을 없는 것으로 설정하는 것, 및 수신되지 않음의 리포트를 포함하는 피드백이 수신된 경우 이전 프레임에 대해 상기 제 1 기여분을 없는 것으로 설정하는 것을 포함할 수 있다.

[0069] 본 발명의 또 다른 양태에 따르면, 손실 채널을 통한 수신 단말기의 디코더로의 송신을 위해, 비디오 스트림을 인코딩하기 위한 송신 단말기가 제공되며, 상기 송신 단말기는, 복수의 프레임의 프레임 각각 내 인코딩될 복수의 타깃 이미지 부분의 타깃 이미지 부분 각각에 대해, 타깃 이미지 부분에 대한 왜곡의 추정 및 상기 타깃 이미지 부분을 인코딩하기 위해 요구되는 비트율의 양을 포함하는 함수를 최적화함으로써, 인코딩 모드의 세트 중에서 선호되는 인코딩 모드를 선택하여, 율-왜곡 최적화 프로세스를 수행하도록 구성된 인코더 - 왜곡의 추정은 소스 코딩 왜곡 및 상기 채널을 통할 때 일어날 수 있는 손실로 인해 겪게 될 왜곡의 추정을 기초로 하며, 상기 인코더는 선택된 인코딩 모드를 이용해 상기 타깃 이미지 부분을 인코딩된 비디오 스트림으로 인코딩하도록 구성됨 - , 및 상기 채널을 통해 상기 인코딩된 비디오 스트림을 송신하도록 구성된 송신기를 포함하며, 상기 인코더는 상기 프레임 중 현재 프레임에 대한 상기 율-왜곡 최적화 프로세스가 상기 프레임 중 이전 프레임을 기초로 한 상기 수신 단말기로부터 수신된 피드백에 따라 수행되도록 구성된다.

[0070] 실시예에서, 상기 인코더는 앞서 언급된 방법 특징들 중 임의의 것에 따르는 동작을 수행하도록 더 구성될 수 있다.

[0071] 본 발명의 또 다른 양태에 따르면, 손실 채널을 통한 수신 단말기의 디코더로의 송신을 위해, 송신 단말기에서 비디오 스트림을 인코딩하기 위한 컴퓨터 프로그램 프로덕트가 제공되며, 상기 컴퓨터 프로그램 프로덕트는 컴퓨터 판독형 매체 상에 포함되고 송신 단말기에서 실행될 때 다음의 동작을 수행하도록 구성된 코드를 포함한다: 복수의 프레임의 프레임 각각 내 인코딩될 복수의 타깃 이미지 부분의 타깃 이미지 부분 각각에 대해, 타깃 이미지 부분에 대한 왜곡의 추정 및 상기 타깃 이미지 부분을 인코딩하기 위해 요구되는 비트율의 양을 포함하는 함수를 최적화함으로써 인코딩 모드의 세트 중에서 선호되는 인코딩 모드를 선택하여 율-왜곡 최적화 프로세스를 수행하는 단계 - 상기 왜곡의 추정은 소스 코딩 왜곡 및 채널을 통할 때 일어날 수 있는 손실로 인해 겪게 될 왜곡의 추정을 기초로 함 - 선택된 인코딩 모드를 이용해 상기 타깃 이미지 부분을 인코딩된 비디오 스트림으로 인코딩하는 단계, 및 채널을 통해 인코딩된 비디오 스트림을 송신하는 단계 - 상기 프레임 중 현재 프레임에 대한 상기 율-왜곡 최적화 프로세스는 상기 프레임 중 이전 프레임을 기초로 수신 단말기로부터 수신된 피드백에 따라 수행됨 - .

[0072] 실시예에서, 코드는 실행될 때 앞서 언급된 방법의 특징들 중 임의의 것에 따르는 동작을 수행하도록 더 구성될

수 있다.

도면의 간단한 설명

[0073] 본 발명 및 본 발명이 실행되는 방식을 더 잘 이해하기 위해, 다음의 첨부된 도면이 예로서 참조된다.

도 1a는 비디오 스트림의 개략적 표현이다.

도 1b는 일부 인트라 예측 코딩 모드의 개략적 표현이다.

도 1c는 에러 전파 왜곡의 계산의 개략적 표현이다.

도 2는 통신 시스템의 개략적 블록도이다.

도 3은 인코더의 개략적 블록도이다.

도 4는 디코더에서 인코더로의 피드백을 이용하는 시스템의 개략적 블록도이다.

발명을 실시하기 위한 구체적인 내용

[0074] 이하에서, 손실-적응적 율-왜곡 최적화 프로세스를 추가로 적응시키고, 이로써, 전체 율-왜곡 성능을 개선하기 위해 디코더에서 인코더로 피드백되는 정보, 가령, 패킷 및/또는 프레임 도착 상태를 이용하는 인코딩 시스템 및 방법이 기재된다. 상기 인코더는 상기 인코더는 도 1에 도시된 유형의 비디오 스트림을 인코딩하기 위해 사용될 수 있고, 통신 시스템, 가령, 도 2의 통신 시스템에서 구현될 수 있다.

[0075] 언급된 바와 같이, 모드 선택은 라그랑주(Lagrangian)형 함수를 최적화(가령, 최소화)하는 것을 포함할 수 있다:

[0076] $$J = D(m, o) + \lambda R(m, o) \quad (1)$$

[0077] 여기서, J는 라그랑주 함수를 나타내고, D는 왜곡의 양(모드 o와 매크로블록 m 또는 매크로블록 서브-파티션의 함수)을 나타내며, R은 비트율이고, λ는 왜곡과 율 간의 트레이드 오프를 정의하는 파라미터이다.

[0078] 종래의 경우, 왜곡 항 D는 소스 코딩 왜곡만 고려한다, 즉, 인코더 내 결함으로 인한 왜곡, 가령, 양자화에 의해 야기되는 왜곡만 고려한다. 채널을 통과할 때의 데이터 손실로 인해 야기될 수 있는 왜곡, 가령, 패킷-기반 네트워크(32)를 통한 송신 시의 패킷 손실은 고려하지 않는다.

[0079] 다른 한편, 손실 적응적 기법(loss adaptive technique), 가령, 본 발명과 Zhang [2]에서 기재된 기법은 소스 인코딩과 채널을 통한 데이터의 손실로 인한 왜곡 모두를 고려하는 "종단간(end-to-end)" 왜곡을 정의하려 시도한다. 특정(타깃) 블록, 매크로블록 또는 서브블록에 대한 종단간 왜곡은 다음과 같이 기술될 수 있다:

[0080] $$D = (1-p)D_{arrival} + pD_{loss} \quad (5)$$

[0081] 여기서, $D_{arrival}$은 타깃 블록이 디코더에 도달할 경우 겪게 될 왜곡의 추정치이고, D_{loss}는 채널을 통한 패킷의 손실로 인해, 가령, 패킷 기반 네트워크(32)를 통한 블록을 포함하는 패킷의 손실로 인해 타깃 블록이 디코더에 도달하지 않을 경우 겪게 될 왜곡의 추정치이다. 파라미터 p는 관심 블록 또는 이미지 부분의 손실을 초래하는 채널을 통한 손실 이벤트가 발생할 확률의 추정치, 가령, 패킷 손실의 확률의 추정치이다. 편의상, 본원에서 관련 레벨의 프레임 파티션(가령, 특정 표준, 가령, H.264의 블록 또는 서브블록)을 일반적으로 지칭하기 위해 용어 "블록"이 사용될 수 있다.

[0082] $D_{arrival}$은 소스 코딩 왜곡뿐 아니라 블록의 과거의 왜곡, 즉, 타깃 블록을 예측할 하나 이상의 기준 블록의 왜곡으로 인해 야기될 왜곡도 나타낸다. 따라서 $D_{arrival}$는 소스 코딩 왜곡 항 D_s와 예측되는 타깃 블록의 히스토리에서의 왜곡(즉, 타깃 블록으로 전달될 타깃 블록의 기준 블록에서의 왜곡)을 나타내는 에러 전파 왜곡 항 D_{ep_ref}를 모두 포함한다:

[0083] $$D_{arrival} = D_s + D_{ep_ref} \quad (6)$$

[0084] D_{loss}는 은폐(concealment)로 인한 손실을 포함한다. 타깃 블록이 수신되지 않은 경우, 디코더는 이전에 디코딩된 블록을 동결(freeze)하는 것, 또는 하나 이상의 성공적으로 디코딩된 블록으로부터(현재 프레임 및/또는 이전 프레임으로부터) 내삽(interpolate) 또는 외삽(extrapolate)하는 것을 포함할 수 있는 은폐 알고리즘을 적용할 것이다. 따라서 D_{loss}는 이러한 은폐 프로세스로 인한 왜곡으로 식별될 수 있다:

[0085] $$D_{loss} = D_{ec} \quad (7)$$

[0086] 따라서 수식(5)을 검사하면, 항 D_s가 손실이 전혀 없을 경우 겪게 될 왜곡의 추정치를 나타내고, 항 D_{ec}는 타깃 블록이 손실될 경우 겪게 될 왜곡의 추정치를 나타내며, 항 D_{ep_ref}는 타깃 블록이 성공적으로 수신되지만 이의 히스토리 내 무언가가 손실된 경우(타깃 블록의 기준 블록이 손실되거나 기준 블록의 기준 블록이 손실된 등의 경우) 겪게 될 왜곡의 추정치를 나타낸다.

[0087] D_s와 D_{ep_ref}는 인코딩 모드 선택 o의 함수이다. D_{ec}는 모드 선택 o의 함수가 아니며, 따라서 라그랑주 표현식에서 빠진다(손실된 블록 - 여전히 손실된 상태임 - 이 어떻게 인코딩되었는지는 중요하지 않다). 그러므로 최적화는 다음과 같이 써질 수 있다:

[0088] $$o_{opt} = \arg\min_{o}(D_s(m,o) + D_{ep_ref}(m,o) + \lambda R(m,o)), \quad (2)$$

[0089] D_s는 인코더에서 알려질 수 있는 정보, 가령, 원시 입력 샘플 값 s와 재구성된 샘플 값 $\hat{s}$ 간의 차이를 기초로 하기 때문에 확정적이다. 인코더는 인코더 측에서 디코더의 병렬 인스턴스(또는 이에 근사한 것(approximation))를 실행한다 - 인터 예측 모듈(43)을 상세히 나타내는 도 3의 삽입도를 참조할 수 있다. 인터 예측 모듈(43)은, 예측된 샘플 $\hat{s}_{pred}$과 재구성된 잔차 $\hat{r}$를 조합함으로써 재구성된 샘플 $\hat{s}$를 결정(즉, 각각의 샘플 인덱스 i에 대해 $\hat{s}_i = \hat{r}_i + \hat{s}_{pred}$)하도록 배열된 모션 보상 예측(MCP)(motion compensation prediction) 블록(44) 및 가산 스테이지(+)를 포함한다.

[0090] 인터 인코딩의 경우, 인코더에서 예측된 샘플 $\hat{s}_{pred}$은 기준 블록의 샘플 $\hat{s}_{ref}$과 동일할 수 있다(기준 프레임에서의 기준 블록은 타깃 프레임에 대해 모션 벡터만큼만 오프셋(offset)되어 있음 - 도 1c를 참조하며, 다시 짧게 언급될 것이다).

[0091] 따라서 인코더는 인코더 및 디코더 단에서 나타날 때의 실제 샘플 s과 재구성된 샘플 $\hat{s}$ 간의 차이를 결정할 수 있다(이는 지금까지 디코더에서 겪는 추가 왜곡을 야기할 손실의 확률을 무시한다). 관심 타깃 블록의 모든 샘플 인덱스 i에 대한 샘플의 차이가, 예를 들어 SSD(sum square difference) 에러로서 계산될 수 있다:

[0092] $$D_s = \sum_i [(s_i - \hat{s}_i)^2] \quad (8)$$

[0093] 그러나, 인코딩된 데이터가 (가령, 패킷-기반 네트워크(32)를 통해) 송신될 채널과 관련된 일부 추정을 기초로 할 D_{ep_ref}는 아직 추정되지 않았다.

[0094] 이를 이루기 위해, 인코더 내 모드 선택 모듈(49)은 가장 최근 인코딩된 프레임 내 각각의 매크로블록 또는 매크로블록의 파티션의 왜곡을 설명하는 에러 전파 왜곡 맵 D_{ep}를 유지하도록 구성될 수 있다. 또한 모드 선택 모듈(49)은 타깃 블록이 예측될 기준 블록을 포함하는 패킷이 채널을 통해 손실될 확률 p를 결정(그리고 따라서 묵시적으로 또는 명시적으로, 패킷이 도착할 확률 $1-p$을 결정)하도록 배열된다. 확률 p는 통계 모델링을 기초로 하는 설계 단계에서 지정될 수 있고, 이때 모드 선택 모듈(49)은 메모리(14)로부터 값을 불러옴으로써, p를 결정한다. 또는 모드 선택 모듈(49)이 수신기(22)로부터의 피드백을 기초로 p를 결정한다.

[0095] 에러 전파 맵은 다음과 같이 표현될 수 있다:

[0096] $$D_{ep} = (1-p)D_{ep_arrival} + pD_{loss} \quad (9)$$

[0097] 에러 전파 맵 D_{ep}는 가장 최근 인코딩된 프레임 내 매크로블록 m에 대한, 또는 더 바람직하게는, 각각의 서브 파티션(블록 또는 서브-블록) m(k)에 대한 왜곡 추정치를 포함한다. 따라서 이는 더 명확히 다음과 같이 써질 수 있다:

[0098] $$D_{ep}(m(k)) = (1-p)D_{ep_arrival}(m(k)) + pD_{loss}(m(k)) \quad (10)$$

[0099] 여기서, m(k)는 매크로블록 m의 k번째 서브-파티션(가령, 서브-블록)을 나타내고, p는 패킷 손실의 확률을 나타낸다.

[0100] 앞서 언급된 바와 같이 D_{loss}는 D_{ec}와 동일하다. $D_{ep_arrival}$은 채널을 통한 차이, 즉, 인코더에서 재구성된 샘플과 디코더에서 재구성된 샘플 간의 차이를 나타낸다. 예를 들어, 이는 SSD(sum of squared difference)의 항으로 정량화될 수 있다:

[0101] $$D_{ep_arrival} = \sum_i (\hat{s}_i - \tilde{s}_i)^2 \quad (11)$$

[0102] 여기서 $\tilde{s}_i$는 소스 코딩 왜곡과 채널로 인한 왜곡을 모두 고려한, 디코더에서 수신된 (인덱스 i의) 샘플이다. 즉, s_i는 원시의 인코딩되지 않은 입력 샘플이고, $\hat{s}_i$는 소스 코딩 왜곡(가령, 양자화로 인한 왜곡)을 고려한, 인코더에서 재구성된 샘플이며, $\tilde{s}_i$는 채널 손실 영향을 포함해 총 종단간 왜곡을 고려한 샘플이며, $s_i \rightarrow \hat{s}_i \rightarrow \tilde{s}_i$이다.

[0103] $D_{ep_arrival}$은 다음과 같이 확장될 수 있다:

[0104] $$D_{ep_arrival} = \sum_i ((\hat{s}_{ref} + \hat{r}_i) - (\tilde{s}_{ref} + \hat{r}_i))^2 \quad (12)$$

[0105] 여기서, $\hat{r}_i$는 재구성된 잔차의 샘플이다. 따라서:

[0106] $$D_{ep_arrival} = \sum_i (\hat{s}_{ref} - \tilde{s}_{ref})^2 = D_{ep_ref} \quad (13)$$

[0107] 따라서 수식 (9)에서 치환하면, 에러 전파 맵은 다음과 같이 다시 써질 수 있다:

[0108] $$D_{ep} = (1-p)D_{ep_ref} + pD_{ec} \quad (14)$$

[0109] 또는,

[0110] $$D_{ep}(m(k)) = (1-p)D_{ep_ref}(m(k)) + pD_{ec}(m(k)) \quad (15)$$

[0111] 모드 최적화 문제를 고려하면, 또한 다음과 같이도 써질 수 있다:

[0112] $$D_{ep}(m(k), n+1) = (1-p)D_{ep_ref}(m(k), n, o_{opt}) + pD_{ec}(m(k), n, o_{opt}) \quad (16)$$

[0113] 여기서 n은 프레임 번호이며, 즉, $D_{ep}(n+1)$은, 이전 시점 n에서의 프레임에 대한 기존 결정 o_{opt}와 왜곡 $D_{ep}(n)$이 주어질 때, 시점 n+1에서의 프레임에 대해 모드를 선택하기 위해 사용될 에러 전파 맵이다.

[0114] Zhang [2]에서처럼, D_{ec} 항도 역시 다음과 같이 확장될 수 있다:

[0115] $$D_{ep}(m(k), n+1) = (1-p)D_{ep_ref}(m(k), n, o_{opt}) + p(D_{ec\text{-}rec}(m(k), n, o_{opt}) + D_{ec\text{-}ep}(m(k), n))$$

(3)

[0116] 여기서, D_{ec-rec}는 인코더에서 재구성된 픽셀과 에러 은폐된 픽셀 간 SSD를 나타내고, D_{ec-ep}는 인코더와 디코더에서의 에러 은폐된 픽셀들 간 기대 SSD이다.

[0117] 수식(3)을 조사해보면, 앞서 설명한 바와 같이, 항 D_{ep_ref}는 타깃 블록이 성공적으로 수신됐지만 이의 히스토리 내 무언가가 손실된 경우(타깃 블록의 기준 블록이 손실, 또는 기준 블록의 기준 블록이 손실된 경우 등) 겪게 될 왜곡을 나타낸다. 덧붙여, D_{ec-rec}는 은폐 알고리즘 자체의 속성으로 인한 왜곡의 추정치(예측에 대한 고유 소스 코딩 왜곡 D_s와 다소 유사함)를 나타낸다. D_{ec-ep}는 타깃 블록이 손실되고(따라서 디코더에서 은폐될 필요가 있고) 은폐된 타깃 블록의 히스토리 내 무언가도 손실된 경우(은폐가 이뤄진 블록이 손실된 경우, 또는 블록의 예측 또는 은폐가 이뤄지는 블록이 손실된 경우 등), 겪게 될 왜곡의 추정치를 나타낸다.

[0118] 따라서 왜곡 맵 D_{ep}는 D_{ec-rec}로부터 그리고 부분적으로 D_{ec-ep}로부터 도출된, 새로운 손실로 인한 기여분(contribution); 및 D_{ep_ref}로부터 그리고 부분적으로 D_{ec-ep}로부터 도출된 과거 손실로 인한 기여분을 포함한다.

[0119] 시퀀스의 첫 번째 프레임에 대해, 프레임은 인트라 코딩에 의해 코딩될 것이며, 이 경우, $D_{ep_ref}=0$이고, 따라서 $D_{ep}=pD_{ec}$이다.

[0120] 에러 은폐 왜곡 D_{ec}는 모드 선택 모듈(49)에 의해 계산된다. 항 D_{ec-rec}는 은폐 알고리즘의 지식을 기초로 하며, 사용되는 특정 에러 은폐 알고리즘에 따라 달라질 수 있다. D_{ec-ep}는 기존(가장 최근) 왜곡 맵을 기초로, D_{ep-ref}와 유사한 방식으로, 가령, 기본 은폐 알고리즘의 경우 공동 위치 블록(co-located block)의 왜곡을 복사함으로써, 또는 모션을 외삽하도록 시도하는 더 복잡한 은폐가 사용되는 경우, 복수의 이전 인코딩된 블록(b1-b4)으로부터의 왜곡의 가중합을 계산함으로써(유추에 의해 이하의 도 1c와 관련된 설명을 참조), 계산된다. D_{ec}를 계산하는 그 밖의 다른 방식이 사용될 수 있는데 - D_{ec}는 인코더에서의 재구성된 샘플과 디코더에 의해 볼 때의 에러 은폐된 샘플(즉, 복사되고, 손실된 프레임 또는 영역을 은폐하기 위해 이전 수신된 프레임 또는 동일 프레임의 수신된 영역으로부터 내삽 또는 외삽된 샘플) 간 차이의 임의의 추정일 수 있다.

[0121] 모드 선택 모듈(49)은, 이제 기존 에러 맵의 지식으로부터의 D_{ep_ref}의 계산을 포함해 각각의 모드 선택 결정에 따라 업데이트함으로써, 각각의 뒤 따르는 인터 예측된 프레임에 대한 에러 전파 맵을 유지한다. 인터 예측(모션 추정)의 경우, Zhang [2]에 따라, 이는 관심 프레임에 대한 모션 벡터를 이용해 이뤄진다.

[0122] 이의 예가 도 1c에 도시되어 있다. 4개의 예시적 블록(b1, b2, b3 및 b4)가 (시점 n-1에서의) 기준 프레임 F_n에서 나타나고, 기준 프레임은 이미 인코딩된 것이다. (다음 시점 n에서의) 타깃 프레임 F_n의 블록은 기준 프레임 F_{n-1}로부터 예측될 것이다. 예를 들어, 타깃 프레임 F_n 내 타깃 블록 b1를 고려할 수 있다. 이러한 목적으로 모션 예측 모듈(44)은 타깃 프레임 F_n 내 타깃 블록과 기준 프레임 F_{n-1} 내 기준 블록(점선으로 나타남) 간 차이를 정의하는 모션 벡터를, 기준 블록이 기준 프레임 F_{n-1} 내 오프셋 위치에서 타깃 프레임 F_n 내 타깃 블록 b1'의 위치로 번역(translate)될 때 타깃 블록 b1의 최상의 추정치를 제공하도록 결정한다. 따라서 점선의 기준 블록이 반드시 기준 프레임 F_{n-1} 내 인덱싱 가능한 블록일 필요는 없다, 즉, 기준 프레임의 지정 서브분할일 필요는 없고, 어떠한 임의의 크기만큼도 오프셋될 수 있다(그리고 실제로 픽셀의 분수만큼 오프셋될 수 있다). 따라서 기준 블록은 4개의 실제 인덱싱 가능한 블록 b1 , b2, b3 및 b4으로부터의 기여분로 구성된다.

[0123] 따라서 에러 전파 맵 $D_{ep}(n+1)$의 업데이트에서 사용되기 위한 D_{ep_ref}를 결정하기 위해 모드 선택 모듈(49)에 의해 수행되는 기존 계산은 기존 맵 $D_{ep}(n)$에서 블록 또는 서브블록 b1 내지 b4에 대해 기록된 왜곡의 가중합을 계산하는 것을 포함한다:

$$D_{ep-ref} = \sum_{i=1}^{4} w_i D_{ep}(i)$$

[0124] (17)

[0125] 또는 더 명확하게,

[0126] $$D_{ep_ref}(m(k),n)=\sum_{i=1}^{4} w_i D_{ep}(b_i,n-1) \quad (4b)$$

[0127] 여기서 w_i는 블록 또는 서브 블록 b_i로부터의 기여분을 나타내는 가중치이고, $D_{ep}(i)$는 블록 또는 서브 블록 b_i에 대한 에러 전파 맵 항목이다.

[0128] 지금까지 초기 에러 전파 맵 D_{ep}를 결정하고, 상기 에러 전파 맵을 이용해 그 다음 코딩을 위한 최적 코딩 모드 결정 o_{opt}을 선택하며, 코딩 결정을 이용해 맵 D_{ep}를 업데이트하고, 그 후 또 다음 번 코딩 결정에 업데이트된 맵을 이용하는 등의 기존의 프로세스가 기재됐으며, 여기서, 에러 전파 맵은 채널을 통한 손실의 추정된 영향을 포함하는 종단간 왜곡을 나타내는데, Zhang [2]을 다시 참조할 수 있다. 이는 본원에서 손실 적응적 율-왜곡 최적화(LARDO)(loss-adaptive rate-distortion optimization)라고 지칭될 수 있다.

[0129] 그러나, Zhang외 [2]의 프로세스는 채널을 통한 손실의 가능성에 대해 순수하게 인코더에서 만들어진 선험적 확률 가정만 기초로 한다.

[0130] 본 발명은 디코더에서 인코더로 피드백되는 정보, 가령, 패킷 및/또는 프레임 도착 상태를 사용함으로써, Zhang의 개선을 제공하여, 인코더에서 손실-적응적 율-왜곡 최적화 프로세스를 추가로 적응시키고, 따라서 전체 율-왜곡 성능을 개선할 수 있다.

[0131] 도 4는 본 발명을 구현하도록 사용될 수 있는 인코더 및 디코더의 시스템을 도시하는 개략적 블록도이다. 인코더는 저장장치(14) 및 프로세싱 장치(16)에서 현시(manifest)되고, 디코더는 수신 단말기(22)의 저장 매체(24) 및 프로세싱 장치(26)에서 현시되는 것이 바람직하다. 송신 단말기(12) 상의 인코더는 인코딩 모듈 및 디코더에서 수행되는 디코딩을 미러링하거나 근사하는 디코딩 모듈의 인코더-측 인스턴스를 포함한다. 상기 인코딩 모듈은 순방향 변환 모듈(forward transform module)(51) 및 양자화기(53)를 포함할 뿐 아니라, 하나 이상의 그 밖의 다른 스테이지, 가령, 엔트로피 인코더(도시되지 않음)까지 포함할 수도 있다. 인코더-측 디코딩 모듈은 역방향 양자화기(63) 및 역방향 변환 모듈(61)을 포함할 뿐 아니라 그 밖의 다른 스테이지, 가령, 엔트로피 디코더까지 포함할 수도 있다. 또한 인코더는 모션 보상 예측 (MCP: motion compensation prediction) 모듈(44), 및 감산 스테이지(-)를 포함한다. 이들 인코더 요소들 간 연결을 설명하기 위해 도 3이 다시 참조된다.

[0132] 덧붙여, 도 3에 도시되어 있지 않지만, 인코더는 인코더-측 디코딩 모듈(61, 63)과 모션 보상 예측 모듈(44) 간의 경로에 연결된 디코딩 화상 버퍼(decoded picture buffer)(65)도 포함한다. 상기 디코딩 화상 버퍼(65)는 복수의 구성 버퍼 영역을 포함하며, 상기 구성 버퍼 영역 각각은 단기 기준(short-term reference) 또는 장기 기준(long-term reference)을 보유하는 것으로 마킹될 수 있다. H.264에서, 단 하나의 실제 버퍼만 존재하며, 여기서 장기 기준을 나타내도록 사용된다(그러나 그 밖의 다른 구현예에서 개별적이고 특화된 단기 버퍼 및 장기 버퍼의 가능성이 배제되는 것은 아니다). 도 4에서 디코딩 화상 버퍼(65)가 하나 이상의 단기 기준(66), 하나 이상의 미확인(unacknowledge)된 장기 기준(67), 및 하나 이상의 확인(acknowledge)된 장기 기준(68)을 보유하는 것으로 도시된다.

[0133] 각각의 구성 버퍼 영역이 하나 이상의 이전에 인코딩된 프레임 또는 슬라이스의 재구성된 버전(즉, 인코더에서 보일 것 같은 프레임 또는 슬라이스를 나타내기 위해 디코딩 모듈(61, 63)의 인코더-측 인스턴스에 의해 인코딩되고 그 후 다시 디코딩된 버전)을 저장하도록 동작 가능하다. 이전에 인코딩된 프레임 또는 슬라이스의 이들 재구성된 버전은 현재 프레임 또는 슬라이스의 인터 예측 코딩에서 기준으로서 사용되도록, 즉, 인코딩될 타깃 블록이 버퍼에서 기준 블록에 대해 코딩될 수 있도록, 제공된다.

[0134] 디코딩 화상 버퍼(65)는 인코딩되는 연속적인 프레임 또는 슬라이스 각각을 이용해 단기 기준(66)이 자동으로 업데이트되도록 구성된다. 즉, 프레임 또는 슬라이스 각각이 인코딩될 때, 해당 새로운 프레임 또는 슬라이스의 디코딩된 버전이 이전에 단기 버퍼에 보유됐던 또 다른 최근 기준 프레임 또는 슬라이스를 자동으로 덮어 쓴다. 바람직한 실시예에서, 디코딩 화상 버퍼(65)는 복수의 단기 기준(66)을 보유할 수 있고 가장 오래된 단기 기준은 항상 버퍼에서 덮어 써진 기준이다. 이를 이루기 위해 어떠한 추가적인 조건도 필요하지 않다.

[0135] 언급된 바와 같이, 또한 H.264 AVC 표준에 의해 특정 기준 프레임 또는 슬라이스가 장기 기준(67, 68)으로서 마킹될 수 있다. 이들 장기 기준은 명시적으로 제거될 때까지 디코딩 화상 버퍼 내에 유지된다. 즉, 이들은 잇따라 인코딩된 프레임 또는 슬라이스에 의해 자동으로 덮어 써지지 않으며, 대신, 인코더의 또 다른 동작 또는 요소가 트리거하는 추가 조건, 가령, 메모리 관리 명령어 같은 제어 명령어가 있을 때만 덮어 써지거나 그 밖의

다른 방식으로 제거된다. 제어 명령어는 취해질 동작이 무엇인지를 결정하는 인코더의 제어기(도시되지 않음)에 의해서 발행(issue)될 수 있다. 장기 버퍼를 비우기 위한 명령어가, 인코딩되는 비트스트림의 헤더, 가령, 슬라이스 헤더 내에서, 디코더로 전송될 수 있다. 또한 유사한 기능이 그 밖의 다른 표준에 포함될 수 있다.

[0136] 수신 단말기(24) 상의 디코더가 모션 보상 예측 모듈(44)의 디코더-측 인스턴스(44', 61', 63', 65', 66', 67' 및 68'), 디코딩 모듈(61, 63), 및 대응하는 단기 및 장기 기준(66, 67 및 68)을 저장하도록 구성된 디코딩 화상 버퍼(65)를 포함한다.

[0137] 수신 단말기(24) 상의 디코더는 피드백 채널을 통해 송신 단말기(12) 상의 인코더와 통신하도록 구성된다. 상기 피드백은 비디오 스트림이 수신 단말기(22)로 송신될 때 이용하는 것과 동일한 네트워크(32), 가령, 동일한 패킷-기반 네트워크, 가령, 인터넷을 통하는 것이 바람직하지만, 대안적 피드백 메커니즘의 가능성이 배제되는 것은 아니다.

[0138] 예를 들어, 장기 기준은 이하와 같이 제어기에 의해 관리될 수 있다. 말하자면, 디코딩 화상 버퍼에 두 개의 장기 기준(가령, It pos 0 및 It pos 1)을 유지하는 것이 결정된다. 인코딩되는 첫 번째 프레임(시점 t0)이 It pos 0에 배치될 수 있다. 상기 첫 번째 프레임이 디코더에 도착할 것(디코더로부터의 피드백이 하나의 RTT 후 도착할 것)임 가정될 수 있고, 따라서 It pos 0은 초기에 무-에러 확인(error-free acknowledged)된 것으로 마킹된다. 장기 기준으로서 마킹되는 다음 프레임은 시점 t0+RTT에서의 프레임이며 It pos 1에 놓인다. 인코더가 디코더로부터 It pos 1에서의 기준이 도착했음(그리고 어떠한 에러 전파도 포함하지 않음)을 가리키는 피드백을 받은 경우, It pos 1은 무-에러 확인 응답된 것으로 마킹되고, 다음 장기 기준 프레임(시점 t0 + 2*RTT)이 It pos 0에 놓인다. 따라서 두 개의 위치가 무-에러 확인 응답된 하나의 위치와 일시적으로 프로빙(probing)을 위해 사용되는 하나의 위치가 항상 존재하는 핑-퐁 버퍼(ping-poing buffer)를 형성한다. 이러한 방식으로 디코딩 화상 버퍼 내에, 손실의 경우 복원 프레임(recovery frame)을 생성하도록 사용될 수 있는 꽤 최근 확인 응답된 무-에러 기준이 항상 존재할 것이다. 원칙적으로, It-ref 프레임이 현재 시점 위치에 가까울수록 인터 코딩이 효율적이며, 따라서 복원 프레임이 (비트로) 더 작아질 것이다. 그러나 이는 인코더의 제어기가 장기 기준을 관리하도록 사용하도록 구성될 수 있는 하나의 예시적 전략에 불과하며, 화상 버퍼에서 장기 기준 및 단기 기준의 시스템을 관리하는 그 밖의 다른 방식이 가능하다(가령, 훨씬 더 우수한 방식은 더 많은 It-ref를 제공하는 것이다).

[0139] 도 4의 예시적 실시예를 참조하여, 본 발명은 디코더에서 디코딩 화상 버퍼(65')에 대한 정보를 포함하기 위한 피드백을 고려한다. 이러한 피드백이 주어지면, 인코더는 가령, 디코더 내 어느 프레임 또는 슬라이스가 어떠한 에러-전파 왜곡도 포함하지 않고 디코딩되었는지를 알 수 있다. 도 4에서, 디코딩 화상 버퍼(65) 내 항목(68)이 이러한 확인 응답된 무-에러 프레임을 가리킨다. 디코딩 화상 버퍼(65) 내 항목(67)은 확인 응답되지 않은 프레임을 가리킨다. 인코더가 (어느 프레임이 손실되었는지에 대한 정보에 추가로) 디코더에서 이용 가능한 가장 최근 확인 응답된 장기 기준이 무엇인지를 알게 하기 위한 피드백 메커니즘이 사용될 수 있다. 이하에서, 확인 응답된 기준은 단순히 그 자체만 확인 응답된 기준이 아니라 확인 응답된 무-에러 기준(즉, 어떠한 에러-전파 왜곡도 없는 참조)을 의미하는 것이 바람직하다. 즉, 기준은, 기준의 히스토리의 엄격한 확인 없이 단지 현재 기준만 수신된 것으로 확인되는 것과는 다른, 기준이 수신된 것으로 확인되고 상기 기준의 히스토리 내 관련된 모든 것도 역시 수신된 것으로 확인되는 엄격한 정의에 따라 확인되어서, 어떠한 전파도 존재하지 않는 것으로 알려질 수 있는 것이 바람직하다. 자체만 확인 응답된 프레임 내의, 무-에러 블록과 관련해 인코딩된 부분이 역시 무-에러다(어떠한 전파 에러도 없음).

[0140] 본 발명의 첫 번째 실시예에 따라, 확인 응답된 장기 기준 프레임(또는 슬라이스)을 기초로 하는 인터 예측이, 인트라 코딩과 유사한 방식으로, 디코더에서의 에러 전파를 중단시키기 위해 사용될 수 있다. 장기 기준으로부터 인터 예측을 이용하는 이점은 일반적으로 인터 예측은 특정 왜곡 수준에 대해 더 낮은 비트율을 도출한다는 것이다.

[0141] 인터 예측에 대해 확인 응답된 장기 기준을 이용함으로써, 본 발명의 첫 번째 실시예는 가령, Zhang [2]의 프레임워크 내에서 추가 매크로블록 코딩 모드가 사용될 수 있게 하며, 이는 인트라 코딩과 유사하게, 그러나 일반적으로 더 낮은 연관 비트율로 에러 전파를 중단시킬 수 있다.

[0142] Zhang외 [2]에 의한 알고리즘은 단 두 개의 서로 다른 유형의 코딩 모드, 즉, 인트라 코딩과 인터 코딩을 고려한다. 이 경우, 에러-전파 기준 왜곡 $D_{ep\text{-}ref}(m,o)$는 수식(3)에서 인트라 코딩된 매크로블록 모드에 대해서만 0(zero)이다. 그러나 본 발명의 첫 번째 실시예는 확인 응답된 장기 기준으로부터의 인터 코딩을 포함하도록 이용 가능한 코딩 모드의 세트를 증강시킨다. 이러한 코딩 모드의 이점은 인트라 코딩과 유사한 방식이지만, 일반

적으로 더 낮은 비트율에서 에러 전파를 중단시킬 수 있다는 것이다.

[0143] 앞서 제시된 알고리즘의 요구되는 변경은 코딩 모드가 확인 응답된 기준 프레임으로부터의 인터 인코딩일 때에도 수식(2) 및 (3)에서 에러-전파 기준 왜곡이 0으로 설정되는 것이다, 즉:

[0144]
$$D_{ep-ref}(m,o)=\begin{cases}\sum_k w_k D_{ep}(m'(k)), & \forall o \in \text{"nonackn. INTER"} \\ 0 \quad INTRA \quad and \quad \text{"ackn. INTER"}\end{cases}$$

[0145] 첫 번째 실시예의 특정 변형에서, 상기의 예시가 다음과 같은 방식으로 수정될 수 있다. 인코더가, 가령, 왕복 시간(RTT)에 따라 달라지는 규칙적인 간격에서, 기준 프레임을 장기 기준이라고 마킹하는 시나리오를 고려할 수 있다. 따라서 (도 4에 도시된 것과 같은) 디코딩 화상 버퍼(65)가 어떤 시점 인스턴스에서 단기 기준(66), "확인 응답되지 않은(non-acknowledged)" 장기 기준(67), 및 확인 응답된 장기 기준(68)을 포함할 것이다. 엄격한 정의에 따라, "확인 응답되지 않은" 장기 기준(67)이 디코더에서 확인되기 위해, 확인 응답되지 않은 장기 기준 전에 어떠한 손실도 없을 것, 즉, 확인 응답된 장기 기준이 어떠한 에러-전파 왜곡 없이 디코딩된 프레임인 것이 요구될 것이다. 디코더에 의해 어느 프레임이 수신되었는지에 대한 정보를 운반하는 피드백이 인코더 측에서 수신된다. 인코더가 어떤 프레임이 확인 응답되지 않은 장기 기준이라고 마킹되었는지를 정확히 알고 있고 패킷 또는 프레임 손실 확률의 추정 p을 가진다고 가정하면, 확인 응답되지 않은 장기 기준이 확인 응답된 장기 기준으로 변환될 확률에 대한 모델을 설정(setup)하는 것이 가능하다. 프레임 손실 확률이 p이고 두 개의 장기 기준 간에 알려진 또는 지정된 간격(interval) L이 있다면, 확인 응답되지 않은 장기 기준이 확인될 선험적 확률(a priori probability)은 $(1-p)^L$이다. 점진적으로, 양의 피드백 정보(positive feedback information)가 수신될 때, 가령, 디코더가 지금가지 L개의 프레임 밖에서 l을 수신했을 때, 장기 기준이 확인될 확률은 $(1-p)^L$에서 $(1-p)^{(L-l)}$로 변경된다. 이제, 기술된 확률 모델이 앞서 언급된 $D_{ep-ref}(m,o)$에 대한 수식을 일반화하도록 사용될 수 있다:

[0146]
$$D_{ep-ref}(m,o)=\begin{cases}\sum_k w_k D_{ep}(m'(k)), & \forall o \in INTER \quad \text{"ShortTerm"} \\ (1-p)^{L-l}\sum_k w_k D_{ep-LTref}(m'(k)), & \forall o \in INTER \quad \text{"non-ackn LongTerm"} \\ 0 \quad INTRA \quad and \quad \text{"ackn. LongTerm"}\end{cases}$$

[0147] 여기서, $D_{ep-LTref}(m(k))$는 확인 응답되지 않은 장기 기준에 대한 기대 에러-전파(단순히, 해당 프레임의 D_{ep}의 복사(copy))를 나타낸다.

[0148] 따라서, 첫 번째 실시예의 이 변형은 확인 응답되지 않은 장기 기준을 기초로 한 인터 예측의 사용을 확인 응답되지 않은 단기 기준을 기초로 한 인터 예측의 사용으로부터 구별하는 또 다른 인코딩을 도입한다. 이는, 어떠한 피드백도 아직 수신되지 않은 특정 장기 기준에 대해, 그럼에도 불구하고, 디코러로부터의 중간 리포트를 기초로 무-에러 확인 응답된 기준이 될 특정 확률을 가진다. 따라서 확인 응답되지 않은 장기 기준에 대해, 에러 전파 왜곡의 추정 $D_{ep-ref}(m,o)$는 손실 확률의 선험적 추정(즉, p 자체는 피드백을 기초로 하지 않음)에 따라 그리고 확인 응답되지 않은 기준의 히스토리에서의 마지막(가장 최근) 확인 응답된 장기 기준으로부터의 시간(또는 이에 상응하는 프레임의 번호)을 기초로 하여 가중될 수 있다. 가중치는 왜곡의 추정을 감쇠(attenuate)시켜, 추정을 감소 또는 약화시킬 수 있다.

[0149] 상기의 논리는 단지 예시적 실시예에 불과하며, 앞서 언급된 형태는 모든 손실이 장기 기준이라고 마킹된 프레임에 영향을 미친다고 가정한다는 점에서 다소 보수적임을 알아야 한다. 장기 기준 프레임에 영향을 미치는 프레임 또는 슬라이스의 손실만 고려하도록 상기의 논리는 정제될 수 있다.

[0150] 또 다른 일반화에서, 장기 기준이 디코더에 의해 수신되었지만 상기 장기 기준에 앞서 손실이 있었던 경우를 고려할 수 있다. 따라서 장기 기준은 상기의 엄격한 정의에 따라 "확인"되지 않았다. 그러나 프레임의 모든 매크로블록에 대해 모드 결정이 저장되었다면, 대응하는 모드가 인트라로 설정된 부분 또는 확인 응답된 장기 기준이 0 에러-전파 기준 왜곡과 연관되고, 따라서 이들 영역으로부터 참조되는 인터 예측이 인트라 코딩에 비교할

때 더 낮을 가능성이 있는 율-왜곡 비용에서 에러 전파를 중단시킨다.

[0151] 상기의 예시적 실시예가 모두 장기 기준 및 피드백 리포팅의 개념을 기초로 하더라도, 디코딩 화상 버퍼 내 단기 기준의 개수에 비교할 때 왕복 시간(RTT)이 충분히 작을 때 동일한 개념이 단기 기준에도 적용될 수 있다. 즉, 충분히 작은 RTT에 대해, 대안적 또는 추가적으로, 단기 기준을 확인 응답된 것으로 인식할 수 있으며, 그렇다면, 이는 앞서 언급된 확인 응답된 장기 기준과 유사한 방식으로 사용될 수 있다. 이전 섹션의 수식(3)의 요구되는 알고리즘 변경은, 인트라 코딩뿐 아니라 확인 응답된 기준으로부터의 인터 코딩에 대해서도 $D_{ep_ref}(m,o)$가 0으로 설정되는 것이다.

[0152] 앞서 언급된 첫 번째 실시예에 대한 일반화가 강건성과 소스 코딩 간의 트레이드 오프를 이루는 데 유연성을 개선한다.

[0153] 추가로, LARDO는 디폴트로 제한된 인트라 예측, 즉, 인터 예측된 매크로블록으로부터의 인트라 예측이 금지됨을 가정한다. 그러나 본 발명자는 제한된 인트라 예측이 실제로 (평이하게 변하는 화상 영역(smooth gradient picture area) 상에서 특히) 심각한 코딩 왜곡을 초래할 수 있음을 관측했다. 따라서 본 발명의 특히 바람직한 변형에서, LARDO는 제한된 인트라 예측 없이 실행되어야 한다. (인터 예측된 매크로블록으로부터 예측될 때) 인트라 코딩 모드는 에러-전파 기준 왜곡과 연관되며, 따라서 에러-전파 기준 왜곡과 연관되지 않는 유일한 모드는 확인 응답된 무-에러 기준 화상으로부터의 인터 예측임을 알 수 있다.

[0154] 본 발명의 두 번째 실시예에서, 디코더에서 인코더로 피드백되는 정보, 가령, 패킷 및/또는 프레임 도착 상태가 사용되어 인코더 내 가능한 왜곡 맵을 조정하고, 따라서 Zhang외의 방법에 비교할 때 전체 율-왜곡 성능을 개선할 수 있다.

[0155] 상기 두 번째 실시예에 따라, 각각의 프레임 또는 한 프레임의 슬라이스에 대한 에러 전파 왜곡 맵 D_{ep}이 인코더의 디코딩 화상 버퍼(65) 내 해당 프레임 또는 슬라이스와 연관되어 저장된다. 또한 각각의 프레임 또는 슬라이스에 대해, 인코더에서의 디코딩 화상 버퍼(65)는 대응하는 에러 은폐 재구성 왜곡 맵 D_{ec-rec}, 에러 은폐 에러 전파 맵 D_{ec-ep}, 대응하는 모드 결정 o, 및 모션 벡터 정보를 저장한다. 그 후, 디코더로부터의 피드백 정보가 인코더에 의해 사용되어 왜곡 맵을 업데이트할 수 있다. 다시 수식(3)을 참조할 수 있다. 피드백 정보에 의해 추정 왜곡 추적이 정제될 수 있어서, 더 우수한 율-왜곡 성능을 산출할 수 있다.

[0156] 이는 다음과 같이 이뤄지는 것이 바람직하다. 인코더가 특정 프레임 또는 슬라이스가 디코더에 성공적으로 도착했음을 시그널링하는 피드백 정보를 수신한 경우, 에러 은폐 기여분 D_{ec-rec} 및 D_{ec-ep}가 수식(3)의 에러 전파 왜곡 맵 D_{ep}에서 제거될 수 있다. 역으로, 특정 프레임 또는 슬라이스가 디코더에서 손실되었다고 시그널링하는 피드백 정보가 수신된 경우, 에러 은폐 왜곡으로부터의 기여분, 즉, 수식(3)의 우변에서 두 번째 항 및 세 번째 항, 즉 (선험적 손실 확률 추정치 p에 의해 정규화된) D_{ec-rec} 및 D_{ec-ep}만 포함하도록, 연관된 에러 전파 왜곡 맵 D_{ep}이 재계산된다.

[0157] 그 후, 디코딩 화상 버퍼 내 기준 화상의 수에 비교할 때 왕복시간(RTT: round trip time)이 작은 경우, 수식(3)을 재귀적으로 적용함으로써, 시점 n-RTT에서 조정된 잠재적 에러-전파 맵 D_{ep}를 시점 n-1에서의 에러 전파 왜곡 맵으로 전파하는 것이 가능하다. 그 후 시점 n-1에서의 업데이트된 에러 전파 왜곡 맵 D_{ep}이 모드 선택 프로세스 (2)에서 사용되는 시점 n에서의 D_{ep_ref}의 계산의 기초가 될 것이다. 이는 잠재적 왜곡 맵의 더 정확한 추적을 도출하고, 따라서 시스템의 전체 율-왜곡 성능을 개선한다.

[0158] 상기의 실시예는 단지 예시로서 기재된 것임을 알 것이다.

[0159] 일반적으로, 앞의 내용은 슬라이스, 매크로블록, 및 블록 또는 서브블록과 관련하여 기재되었지만, 반드시 이들로의 한정을 의도한 것이 아니며, 본원에 기재된 아이디어는 프레임을 분할 또는 세분하는 임의의 구체적 방식에 한정되지 않는다. 덧붙여, 왜곡 맵은 전체 프레임 또는 프레임 내 영역을 커버할 수 있고 코딩 결정 프로세스는 전체 프레임에 대해 적용되거나 한 프레임 내 한 영역에 대해서만 적용될 수 있다. 또한, 예측 블록 입도(granularity)가 왜곡 맵 입도와 동일해야 하는 것은 아니며 심지어 관련되어야 하는 것도 아니다(그러나 가능성이 배제되는 것은 아니다).

[0160] 종종 SSD(sum of squared differences)가 SAD(sum of absolute differences)에 비교할 때 더 높은 품질을 도출하기 때문에 차이의 양으로서 선호되지만, SAD의 가능성 또는 그 밖의 다른 가능성이 배제되는 것은 아니며, 일

반적으로 본 발명은 왜곡을 정량화하기 위한 토대로서 샘플들 간 차이의 어떠한 양으로도 이용해 구현될 수 있다.

[0161] 일반적으로, 율의 양이 모든 요구되는 파라미터, 가령, 예측 및 양자화된 변환 계수를 기술하는 파라미터의 코딩을 설명한다. 본원에서 이러한 유형의 최적화는 완전 율-왜곡 최적화(RDO)라고 지칭될 수 있다. 그러나 복잡도가 더 낮은 실시예에서, 전부가 아닌 일부 프로세싱 스테이지의 영향만 고려함으로써, 가령, 예측의 영향만 고려함으로써, 왜곡 및/또는 율 항이 근사화될 수 있다.

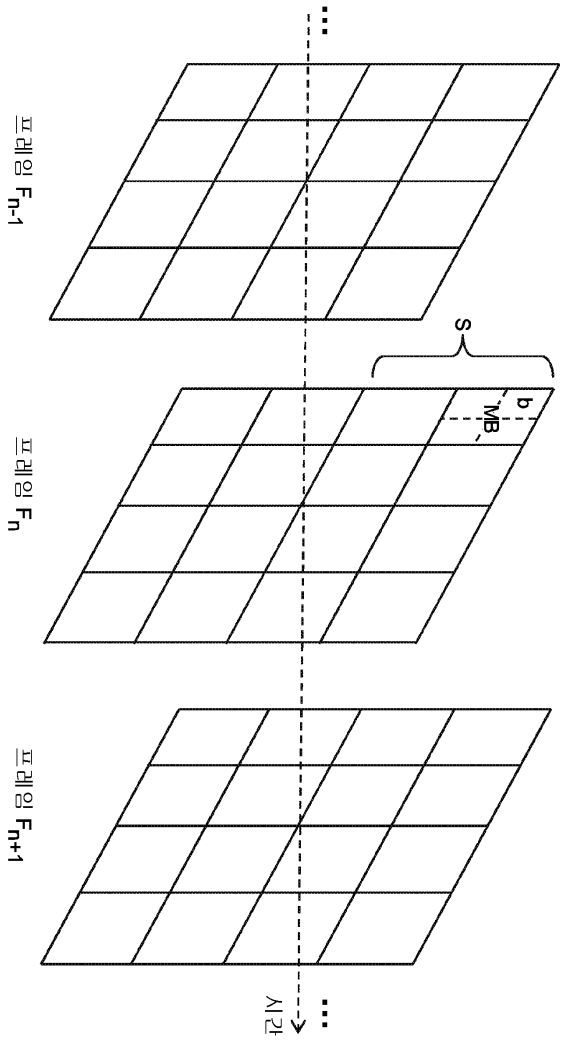
[0162] 덧붙여, 본 발명의 특정 실시예에 따라, 본 발명이 2개의 프레임 n-1과 n, 또는 n과 n+1, 또는 이러한 식으로 기술된 경우, 이들이 반드시 2개의 인접한 프레임을 지칭하는 것은 아니다(그러나 기존 코덱에서는 이러한 경우에 해당할 수 있다). 일부 실시예에서, 더 이른 프레임과 관련해 인터 예측이 수행될 수 있고, 임의의 이전 인코딩된 프레임 또는 이미지 부분 및 이로부터 예측될 이후 프레임 또는 부분을 지칭하기 위해 일반적으로 n-1과 n, 또는 n과 n+1이 본 발명과 관련해 사용될 수 있다.

[0163] 이 출원에서 손실로 인한 기여분이 언급되는 경우, 또는 "만약" 데이터가 채널 등을 통해 손실된다면 발생할 어떠한 것을 언급하는 경우, 이는 디코더가 겪게 될 수 있는 것에 대해 인코더에 의해 만들어지는 확률적 가정(가령, p)과 관련될 뿐이며 인코더는 당연히 발생할 일을 알지 못한다. 확률적 가정은 통계적 네트워크 모델링을 기초로 설계 단계에서 지정되거나, 및/또는 디코더로부터의 피드백을 기초로 동적으로 결정될 수도 있다.

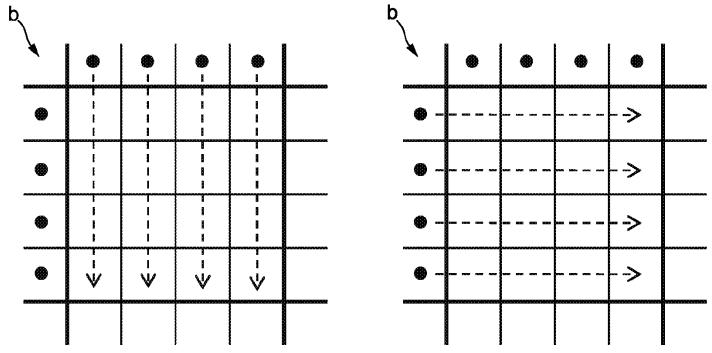
[0164] 본원의 내용을 제공 받은 해당 분야의 통상의 기술자에게 그 밖의 다른 변형예가 자명할 수 있다. 본 발명의 범위는 기재된 실시예에 의해 제한되지 않고 이하의 특허청구범위에 의해서만 제한된다.

도면

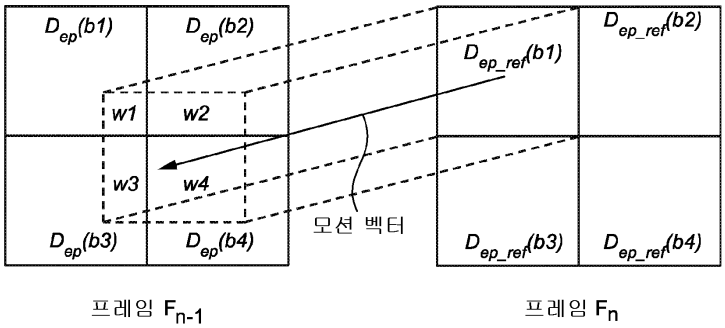
도면1a



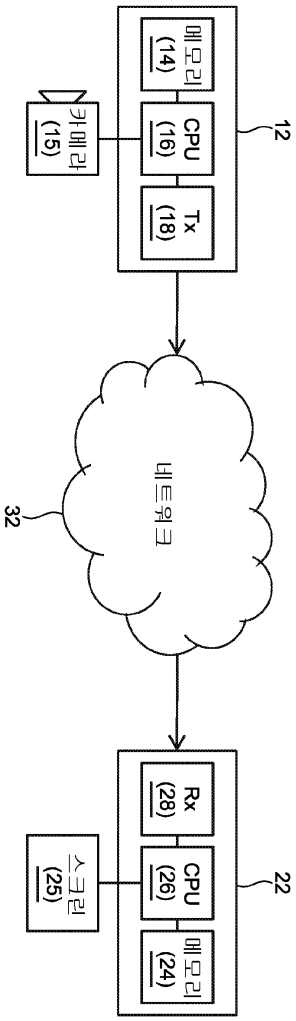
도면1b



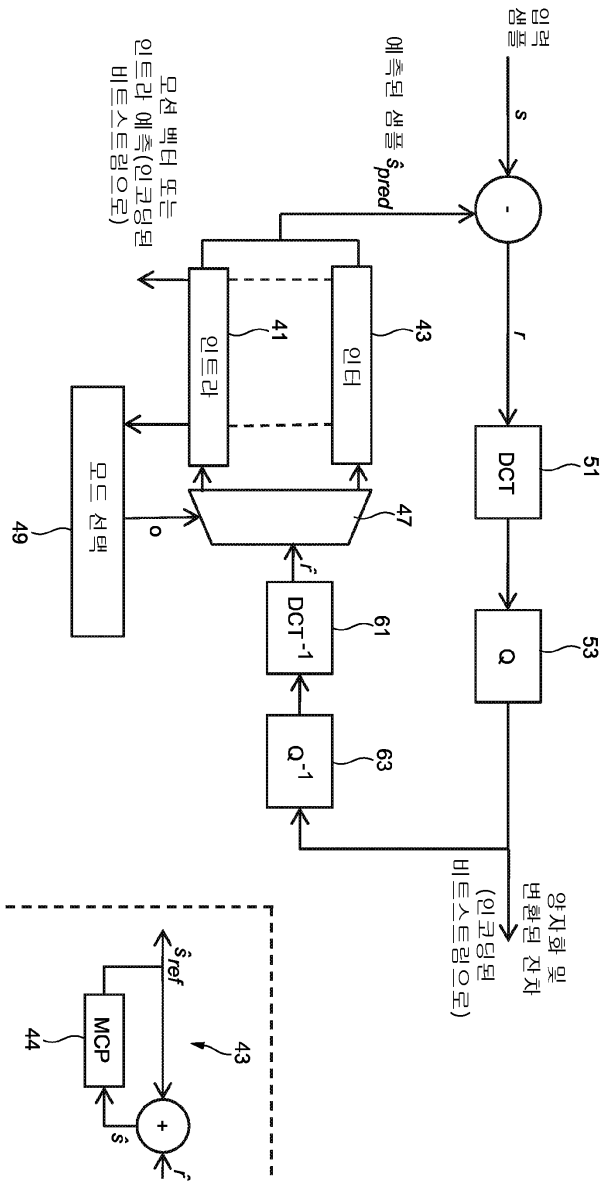
도면1c


$D_{ep}(b1)$
$D_{ep}(b2)$
w1
w2
w3
w4
$D_{ep}(b3)$
$D_{ep}(b4)$
모션 벡터
$D_{ep_ref}(b1)$
$D_{ep_ref}(b2)$
$D_{ep_ref}(b3)$
$D_{ep_ref}(b4)$
프레임 F_{n-1}
프레임 F_n

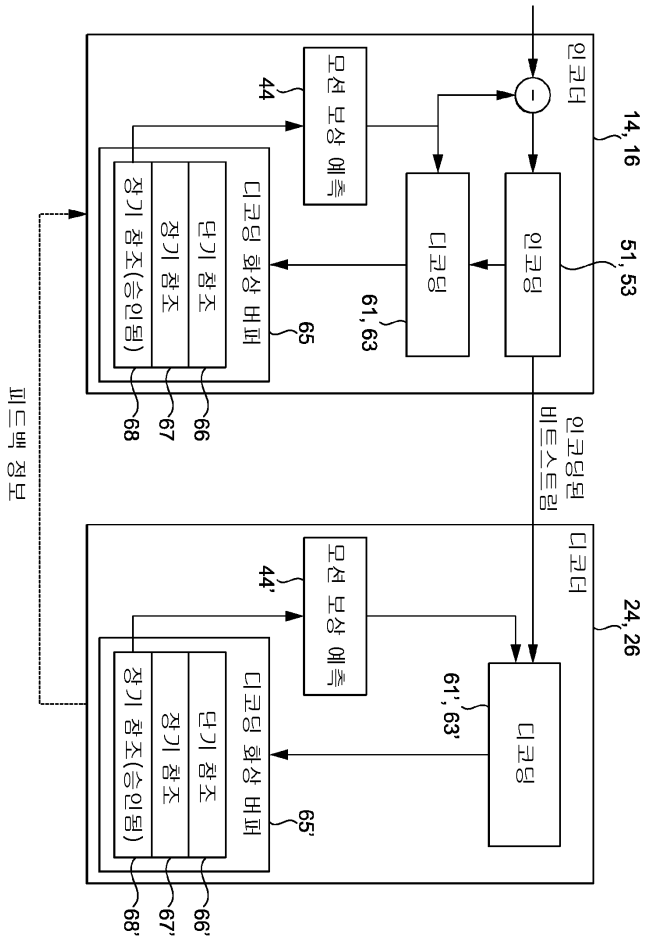
도면2


메모리 (14)
CPU (16)
Tx (18)
12
카메라 (15)
네트워크
32
Rx (28)
CPU (26)
메모리 (24)
22
스크린 (25)

도면3


입력 샘플
s
예측된 샘플 $\hat{s}_{pred}$
모션 벡터 또는 인트라 예측(인코딩된 비트스트림으로)
-
r
41
인트라
43
인터
51
DCT
53
Q
양자화 및 변환된 잔자 (인코딩된 비트스트림으로)
47
o
49
모드 선택
r'
61
DCT^{-1}
63
Q^{-1}
$\hat{s}_{ref}$
MCP
44
43
$\hat{s}$
+
r'

도면4


인코더
14, 16
모션 보상 예측
44
디코딩 화상 버퍼
단기 참조
장기 참조
장기 참조(승인됨)
65
66
67
68
디코딩
61, 63
인코딩
51, 53
인코딩된 비트스트림
디코더
24, 26
모션 보상 예측
44'
디코딩 화상 버퍼
단기 참조
장기 참조
장기 참조(승인됨)
65'
66'
67'
68'
디코딩
61', 63'
피드백 정보