



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2020년11월06일
(11) 등록번호 10-2175020
(24) 등록일자 2020년10월30일

- (51) 국제특허분류(Int. Cl.)
H03M 7/30 (2006.01) G06F 7/24 (2006.01)
G06K 9/38 (2006.01) G06T 1/20 (2018.01)
G11B 20/14 (2006.01) H03M 7/14 (2006.01)
H03M 7/40 (2006.01)
- (52) CPC특허분류
H03M 7/30 (2019.01)
G06F 7/24 (2013.01)
- (21) 출원번호 10-2016-7022055
- (22) 출원일자(국제) 2015년02월20일
심사청구일자 2019년12월26일
- (85) 번역문제출일자 2016년08월11일
- (65) 공개번호 10-2016-0123302
- (43) 공개일자 2016년10월25일
- (86) 국제출원번호 PCT/EP2015/025007
- (87) 국제공개번호 WO 2015/124323
국제공개일자 2015년08월27일
- (30) 우선권주장
1403038.1 2014년02월20일 영국(GB)
- (56) 선행기술조사문헌
JP2005051406 A

- (73) 특허권자
구루로직 마이크로시스템스 오이
핀란드 투르쿠 20100 린난카투 34
- (72) 발명자
월너 카이
핀란드 37550 램팔라 로힐랜티에 3 아이 15
칼레보 오시
핀란드 37800 아카아 케툰한타 1
- (74) 대리인
김태홍, 김진희

전체 청구항 수 : 총 25 항

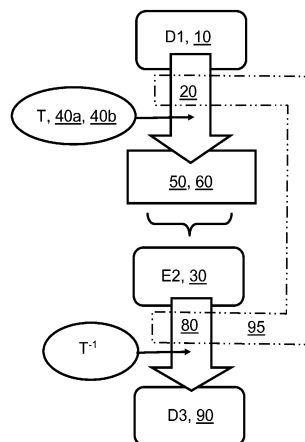
심사관 : 조춘근

(54) 발명의 명칭 데이터의 소스-인코딩 및 디코딩 디바이스 및 방법

(57) 요약

데이터(D1, 10)를 인코딩하여 대응하는 인코딩된 데이터(E2, 30)를 생성하는 인코더(20)가 제공되며, 인코딩된 데이터(E2, 30)는 그 인코딩된 데이터(E2, 30)에서 표현될 하나 이상의 심볼의 빈도, 확률 또는 레인지를 나타내는 정보를 포함한다. 인코더(20)는 하나 이상의 심볼 중의 적어도 하나의 심볼에 대한 빈도, 확률 또는 레인지를 (뒷면에 계속)

대표도 - 도1



나타내는 정보가 상기 인코딩된 데이터(E2, 30)에 포함되어 있는지의 여부를 나타내는 추가 정보를 상기 인코딩된 데이터(E2, 30)에 포함시키도록 동작 가능하다. 인코더(20)는 상기 하나 이상의 심볼 중의 적어도 하나의 심볼에 대한 빈도, 확률 또는 레인지를 나타내는 정보가 상기 인코딩된 데이터(E2, 30)에 포함되어 있는지의 여부에 관한, 단일 가용성 비트 정보로서 표현된 추가 정보를 인코딩된 데이터(E2, 30)에 포함시키도록 동작 가능하다. 유리하게는, 인코더(20)는 인코딩된 데이터(E2, 30)에서의 빈도, 확률 또는 레인지 정보를 나타내는 정보의 내포를 단일 가용성 비트 값 "1"을 이용하여 나타내고, 상기 인코딩된 데이터(E2, 30)에서의 빈도, 확률 또는 레인지 정보를 나타내는 정보의 내포하지 않음을 단일 가용성 비트 값 "0"을 이용하여 나타내도록 동작 가능하다.

(52) CPC특허분류

- G06K 9/38* (2013.01)
 - G06T 1/20* (2013.01)
 - G11B 20/1419* (2013.01)
 - H03M 7/14* (2013.01)
 - H03M 7/3071* (2013.01)
 - H03M 7/4006* (2013.01)
 - H03M 7/6017* (2013.01)
-

명세서

청구범위

청구항 1

레인지 코딩을 채택해서 상이한 심볼들을 포함하는 데이터를 인코딩하여 대응하는 인코딩된 데이터를 생성하는 인코더에 있어서,

상기 인코딩된 데이터는 그 인코딩된 데이터에 나타낼 인코딩된 상기 상이한 심볼들의 빈도, 확률 또는 레인지 값을 나타내는 레인지 코딩을 위한 정보를 포함하고,

상기 인코더는, 상기 상이한 심볼들에 대한 빈도, 확률 또는 레인지 값의 상기 정보가 상기 인코딩된 데이터에 포함되어 있는지의 여부를 나타내는 추가 정보를 상기 인코딩된 데이터에 포함시키도록 동작 가능하며,

상기 추가 정보는 상기 빈도, 확률 또는 레인지 값이 관련되는 개별 심볼을 나타내는 것인 인코더.

청구항 2

제1항에 있어서, 상기 인코더는, 상기 심볼들에 대한 빈도, 확률 또는 레인지 값의 상기 정보가 상기 인코딩된 데이터에 포함되어 있는지의 여부를 나타내는, 단일 가용성 비트(single availability bit) 정보로서 표현된 상기 추가 정보를 상기 인코딩된 데이터에 포함시키도록 동작 가능한 것인 인코더.

청구항 3

제2항에 있어서, 상기 인코더는, 상기 인코딩된 데이터에서의 상기 심볼들에 대한 빈도, 확률 또는 레인지 값의 상기 정보의 내포(inclusion)를 단일 가용성 비트 값 "1"을 이용하여 나타내고, 상기 인코딩된 데이터에서의 상기 심볼들에 대한 빈도, 확률 또는 레인지 값의 상기 정보의 내포하지 않음(non-inclusion)을 단일 가용성 비트 값 "0"을 이용하여 나타내도록 동작 가능한 것인 인코더.

청구항 4

제1항에 있어서, 상기 인코딩된 데이터에 나타낼 상기 심볼들의 빈도, 확률 또는 레인지 값의 상기 정보는 상기 인코더에 의해, 인코딩될 데이터의 하나 이상의 특성의 함수로서 동적으로 변하는 것인 인코더.

청구항 5

제4항에 있어서, 상기 하나 이상의 특성은, 상기 데이터의 유형(type), 상기 데이터의 성분(content), 상기 데이터의 구조, 상기 데이터의 사이즈, 상기 데이터에 존재하는 값의 레인지, 상기 데이터에 존재하는 엘리먼트의 사이즈 또는 상기 데이터와 연관된 메타데이터 중 적어도 하나를 포함하는 것인 인코더.

청구항 6

제4항에 있어서, 상기 인코더는, 인코딩될 데이터를 복수의 부분으로 세분하고, 각 부분마다 상기 심볼들에 대한 빈도, 확률 또는 레인지 값의 상기 정보가 상기 인코딩된 데이터에 포함되어 있는지의 여부를 나타내는 대응하는 추가 정보를 개별적으로 생성하도록 동작 가능한 것인 인코더.

청구항 7

제1항에 있어서, 상기 인코더는, 데이터를 분석하여 그 데이터에서 발생하는 최소(min) 및 최대(max) 값을 결정하도록 동작 가능하고, 상기 최소(min) 및 최대(max) 값을 기술하는 정보는 상기 인코더에 의해 상기 인코딩된 데이터에 포함되며, 상기 최소(min) 및 최대(max) 값은 레인지 코딩을 채택할 때에 가능하게 이용할 수 있는 심볼 값의 레인지를 한정하는 것인 인코더.

청구항 8

제7항에 있어서, 상기 인코더는 상기 심볼들에 대한 빈도, 확률 또는 레인지 값의 상기 정보가 상기 인코딩된 데이터에 포함되어 있는지의 여부를 나타내는 가용성 신호를 이용하지 않고서 상기 인코딩된 데이터에서 데이터

에 존재한 상기 최소(min) 및 최대(max) 값에 의해 한정되는 가능하게 이용할 수 있는 심볼 값의 레인지에서의 최초 심볼 및 최후 심볼 중 적어도 하나를 표현하도록 동작 가능한 것인 인코더.

청구항 9

제1항에 있어서, 상기 인코더는, 레인지 인코딩을 데이터에 적용하여 인코딩된 데이터를 생성하기 이전에, 인코딩될 데이터의 엔트로피를 변경하도록 동작 가능한 것인 인코더.

청구항 10

제1항에 있어서, 상기 인코더에 의해 인코딩될 데이터는, 캡처된 오디오 신호, 유전학 시퀀싱(genetic sequencing) 데이터, 캡처된 비디오 신호, 캡처된 이미지, 텍스트 데이터, 지진계 데이터(seismographic data), 센서 신호, 아날로그-디지털(ADC) 변환 데이터, 생체 신호 데이터, 캘린더 데이터, 경제 데이터, 수학 데이터 중 적어도 하나에 대응하는 데이터인 것인 인코더.

청구항 11

인코더에서 레인지 코딩을 채택해서 상이한 심볼들을 포함하는 데이터를 인코딩하여 대응하는 인코딩된 데이터를 생성하는 데이터 인코딩 방법에 있어서,

상기 인코딩된 데이터는 그 인코딩된 데이터에 나타낼 인코딩된 상기 상이한 심볼들의 빈도, 확률 또는 레인지 값을 나타내는 레인지 코딩을 위한 정보를 포함하고, 상기 방법은,

상기 인코더를 이용해서, 상기 상이한 심볼들에 대한 빈도, 확률 또는 레인지 값의 상기 정보가 상기 인코딩된 데이터에 포함되어 있는지의 여부를 나타내는 추가 정보를 상기 인코딩된 데이터에 포함시키는 단계를 포함하고,

상기 추가 정보는 상기 빈도, 확률 또는 레인지 값이 관련되는 개별 심볼을 나타내는 것인 데이터 인코딩 방법.

청구항 12

제11항에 있어서, 상기 방법은,

상기 인코더를 이용해서, 상기 심볼들에 대한 빈도, 확률 또는 레인지 값의 상기 정보가 상기 인코딩된 데이터에 포함되어 있는지의 여부를 나타내는, 단일 가용성 비트 정보로서 표현된 상기 추가 정보를 상기 인코딩된 데이터에 포함시키는 단계를 더 포함하는 것인 데이터 인코딩 방법.

청구항 13

제12항에 있어서, 상기 방법은,

상기 인코더를 이용해서, 상기 인코딩된 데이터에서의 상기 심볼들에 대한 빈도, 확률 또는 레인지 값의 상기 정보의 내포를 단일 가용성 비트 값 "1"을 이용하여, 그리고 상기 인코딩된 데이터에서의 상기 심볼들에 대한 빈도, 확률 또는 레인지 값의 상기 정보의 내포하지 않음을 단일 가용성 비트 값 "0"을 이용하여 나타내는 단계를 더 포함하는 것인 데이터 인코딩 방법.

청구항 14

제11항에 있어서, 상기 방법은,

상기 인코더를 이용해서, 상기 인코딩된 데이터에 나타낼 상기 심볼들의 빈도, 확률 또는 레인지 값의 상기 정보를, 인코딩될 데이터의 하나 이상의 특성의 함수로서 동적으로 변화시키는 단계를 포함하는 것인 데이터 인코딩 방법.

청구항 15

제14항에 있어서, 상기 방법은,

상기 인코더를 이용해서, 인코딩될 데이터를 복수의 부분으로 세분하고, 각 부분마다 상기 심볼들에 대한 빈도, 확률 또는 레인지 값의 상기 정보가 상기 인코딩된 데이터에 포함되어 있는지의 여부를 나타내는 대응하는 추가 정보를 개별적으로 생성하는 단계를 포함하는 것인 데이터 인코딩 방법.

청구항 16

제11항에 있어서, 상기 방법에 의해 인코딩될 데이터는, 유전학 시퀀스 정보, 캡처된 오디오 신호, 캡처된 비디오 신호, 캡처된 이미지, 텍스트 데이터, 지진계 데이터, 센서 신호, 아날로그-디지털(ADC) 변환 데이터, 생체 신호 데이터, 캘린더 데이터, 경제 데이터, 수학 데이터 중 적어도 하나에 대응하는 데이터인 것인 데이터 인코딩 방법.

청구항 17

레인지 코딩을 채택하여 상이한 심볼들을 포함하는 데이터로부터 인코더에 의해 생성되는 인코딩된 데이터를 디코딩하는 디코더에 있어서,

상기 인코딩된 데이터는 그 인코딩된 데이터에 나타낼 인코딩된 상기 상이한 심볼들의 빈도, 확률 또는 레인지 값을 나타내는 레인지 코딩을 위한 정보를 포함하고,

상기 인코더는, 상기 상이한 심볼들에 대한 빈도, 확률 또는 레인지 값의 상기 정보가 상기 인코딩된 데이터에 포함되어 있는지의 여부를 나타내는 추가 정보를 상기 인코딩된 데이터에 포함시키도록 동작 가능하며,

상기 추가 정보는 상기 빈도, 확률 또는 레인지 값이 관련되는 개별 심볼을 나타내는 것인 디코더.

청구항 18

제17항에 있어서, 상기 디코더는 데이터 프로세싱 하드웨어를 포함하고,

상기 데이터 프로세싱 하드웨어는,

(a) 상기 심볼들에 대한 빈도, 확률 또는 레인지 값의 상기 정보가 상기 인코딩된 데이터에 포함되어 있는지의 여부를 나타내는 가용성 신호를 수신하고,

(b) 상기 인코딩된 데이터 내에 송신된 빈도, 확률 또는 레인지 값을 수신하며,

(c) (a)와 (b)로부터 완전한 빈도 테이블, 확률 테이블 또는 레인지 테이블을 구성하고,

(d) 상기 완전한 빈도 테이블, 확률 테이블 또는 레인지 테이블을 이용해서, 상기 인코딩된 데이터를 레인지 디코딩하여 디코딩된 출력 데이터를 생성하도록

동작 가능한 것인 디코더.

청구항 19

제17항에 있어서, 상기 디코더는 상기 디코딩된 출력 데이터에 트랜스코딩(transcoding)을 실행하여 대응하는 트랜스코딩된 데이터를 생성하도록 동작 가능한 것인 디코더.

청구항 20

디코더에서 인코딩된 데이터를 디코딩하는 방법에 있어서,

레인지 코딩을 채택하여 상이한 심볼들을 포함하는 데이터로부터 인코딩된 데이터를 생성한 인코더로부터의 인코딩된 데이터를 디코딩하는 단계를 포함하고,

상기 인코딩된 데이터는 그 인코딩된 데이터에 나타낼 인코딩된 상기 상이한 심볼들의 빈도, 확률 또는 레인지 값을 나타내는 레인지 코딩을 위한 정보를 포함하고,

상기 인코딩된 데이터는 상기 상이한 심볼들에 대한 빈도, 확률 또는 레인지 값의 상기 정보가 상기 인코딩된 데이터에 포함되어 있는지의 여부를 나타내는 추가 정보를 포함하며,

상기 추가 정보는 상기 빈도, 확률 또는 레인지 값이 관련되는 개별 심볼을 나타내는 것인 데이터 디코딩 방법.

청구항 21

제20항에 있어서, 상기 방법은,

(a) 상기 심볼들에 대한 빈도, 확률 또는 레인지 값의 상기 정보가 상기 인코딩된 데이터에 포함되어 있는지의

여부를 나타내는 가용성 신호를 수신하는 단계와,

(b) 인코딩된 데이터 내에 송신된 빈도, 확률 또는 레인지 값을 수신하는 단계와,

(c) 단계 (a)와 (b)로부터 완전한 빈도 테이블, 확률 테이블 또는 레인지 테이블을 구성하는 단계와,

(d) 상기 완전한 빈도 테이블, 확률 테이블 또는 레인지 테이블을 이용해서, 상기 인코딩된 데이터를 레인지 디코딩하여 디코딩된 출력 데이터를 생성하는 단계

를 포함하는 것인 데이터 디코딩 방법.

청구항 22

제20항에 있어서, 상기 방법은, 상기 디코딩된 출력 데이터에 트랜스코딩을 실행하여 대응하는 트랜스코딩된 데이터를 생성하는 단계를 포함하는 것인 데이터 디코딩 방법.

청구항 23

코덱에 있어서,

레인지 코딩을 채택해서 상이한 심볼들을 포함하는 데이터를 인코딩하여 대응하는 인코딩된 데이터를 생성하는 인코더로서, 상기 인코딩된 데이터는 그 인코딩된 데이터에 나타낼 인코딩된 상기 상이한 심볼들의 빈도, 확률 또는 레인지 값을 나타내는 레인지 코딩을 위한 정보를 포함하고, 상기 인코더는, 상기 상이한 심볼들에 대한 빈도, 확률 또는 레인지 값의 상기 정보가 상기 인코딩된 데이터에 포함되어 있는지의 여부를 나타내는 추가 정보를 상기 인코딩된 데이터에 포함시키도록 동작 가능하며, 상기 추가 정보는 상기 빈도, 확률 또는 레인지 값이 관련되는 개별 심볼을 나타내는 것인 상기 인코더와,

상기 인코더에 의해 생성되는 인코딩된 데이터를 디코딩하는 디코더

를 포함하는 코덱.

청구항 24

컴퓨터 판독 가능한 명령어가 저장되어 있는 비일시적 컴퓨터 판독 가능한 저장 매체에 저장된 컴퓨터 프로그램에 있어서,

상기 컴퓨터 판독 가능한 명령어는 방법을 구현하도록 인코더의 컴퓨팅 하드웨어에 의해 실행 가능하며, 상기 방법은,

레인지 코딩을 채택하여 상이한 심볼들을 포함하는 데이터를 인코딩하는 단계로서, 상기 인코딩된 데이터는 그 인코딩된 데이터에 나타낼 인코딩된 상기 상이한 심볼들의 빈도, 확률 또는 레인지 값을 나타내는 레인지 코딩을 위한 정보를 포함하는, 상기 데이터 인코딩 단계와,

상기 인코더를 이용해서, 상기 상이한 심볼들에 대한 빈도, 확률 또는 레인지 값의 상기 정보가 상기 인코딩된 데이터에 포함되어 있는지의 여부를 나타내는 추가 정보를 상기 인코딩된 데이터에 포함시키는 단계를 포함하고,

상기 추가 정보는 상기 빈도, 확률 또는 레인지 값이 관련되는 개별 심볼을 나타내는 것인 비일시적 컴퓨터 판독 가능한 저장 매체에 저장된 컴퓨터 프로그램.

청구항 25

컴퓨터 판독 가능한 명령어가 저장되어 있는 비일시적 컴퓨터 판독 가능한 저장 매체에 저장된 컴퓨터 프로그램에 있어서,

상기 컴퓨터 판독 가능한 명령어는 방법을 구현하도록 디코더의 컴퓨팅 하드웨어에 의해 실행 가능하며, 상기 방법은,

레인지 코딩을 채택하여 상이한 심볼들을 포함하는 데이터로부터 인코딩 데이터를 생성한 인코더로부터의 인코딩된 데이터를 디코딩하는 단계를 포함하고,

상기 인코딩된 데이터는 그 인코딩된 데이터에 나타낼 인코딩된 상기 상이한 심볼들의 빈도, 확률 또는 레인지

값을 나타내는 레인지 코딩을 위한 정보를 포함하고,

상기 인코딩된 데이터는 상기 상이한 심볼들에 대한 빈도, 확률 또는 레인지 값의 상기 정보가 상기 인코딩된 데이터에 포함되어 있는지의 여부를 나타내는 추가 정보를 포함하며,

상기 추가 정보는 상기 빈도, 확률 또는 레인지 값이 관련되는 개별 심볼을 나타내는 것인 비밀시적 컴퓨터 판독 가능한 저장 매체에 저장된 컴퓨터 프로그램.

발명의 설명

기술 분야

[0001] 본 개시내용은 데이터(D1)를 인코딩하여 대응하는 인코딩된 데이터(E2)를 생성하는 방법에 관한 것이다. 또, 본 개시내용은 전송한 인코딩된 데이터(E2)를 디코딩하여 대응하는 디코딩된 데이터(D3) 및/또는 트랜스코딩된 버전의 디코딩된 데이터(D3)를 생성하는 방법에 관한 것이기도 하다. 또한, 본 개시내용은 전송한 방법들을 구현하도록 동작 가능한 인코더 및 디코더에 관한 것이며, 이 인코더와 디코더를 조합하여 구성한 것이 코덱이다. 부가적으로, 본 개시내용은 컴퓨터 판독 가능한 명령어가 저장되어 있는 비밀시적 컴퓨터 판독 가능한 저장 매체를 포함하는 컴퓨터 프로그램 제품에 관한 것이며, 이 컴퓨터 판독 가능한 명령어는 전송한 방법을 실행하도록 컴퓨팅 하드웨어에 의해 실행될 수 있다.

배경 기술

[0002] 개관에 있어서, 도 1에 도시하는 바와 같이, 인코더(20)에서 입력 데이터(D1; 10)를 인코딩하여 대응하는 인코딩된 출력 데이터(E2; 30)를 생성하는 알려진 인코딩 방법은 한가지 이상의 변환(40a, 40b)을 입력 데이터(D1; 10)에 적용하여, 대응하는 인코딩된 변환 데이터(50)를 생성하는 것을 수반하는데, 인코딩된 변환 데이터(50)는 채택된 하나 이상의 변환을 정의하는 하나 이상의 코딩 테이블을 정의하는 코딩 테이블 데이터(60)와 연관되어 있다. 인코딩된 변환 데이터(50) 및 코딩 테이블 데이터(60), 즉 조합하여 인코딩된 출력 데이터(E2; 30)는 종종 데이터 캐리어 및/또는 통신 네트워크를 통해, 인코딩된 출력 데이터(E2; 30)를 디코딩하여 대응하는 디코딩된 데이터(D3; 90)를 생성하도록 동작 가능한 하나 이상의 디코더(80)에 전달된다. 선택사항으로서, 디코딩된 데이터(D3)는, 예컨대 하나 이상의 디코더(80)에서 트랜스코딩이 일어날 경우에는 재작성된 버전의 데이터(D1)이지만, 이것에 한정되지는 않는다. 종종, 인코딩된 출력 데이터(E2; 30)는 입력 데이터(D1)에 비해 압축되는 것이 바람직하다. 또, 인코딩된 출력 데이터(E2; 30)도 실질적으로 손실없이 압축되어, 디코딩된 데이터(D3; 90)가 입력 데이터(D1; 10)와 비교하여 가능한 한 정확한 것이 바람직하다. 입력 데이터(D1; 10)에 대해 인코딩된 출력 데이터(E2; 30)에서 달성 가능한 데이터 압축은, 코딩 테이블 데이터(60)가 인코딩된 변환 데이터(50)에 비해 사이즈에 있어서 유효하다면, 즉 코딩 테이블 데이터(60)가 유효 데이터 오버헤드에 대응하면, 잠재적으로 비효율적이다. 소정의 알려진 인코더(20)의 구현예에 있어서, 코딩 테이블 데이터(60)는 입력 데이터(D1; 10)에서의 원래의 심볼과 같은, 인코딩된 변환 데이터(50)에 존재하는 코딩된 "심볼"의 빈도, 확률 또는 레인지를 나타내는 정보를 포함한다. 코딩 테이블 데이터(60)가 그것에 한정되지는 않는다. 선택사항으로서, 코딩 테이블 데이터(60)는 사용된 변환 및 그 파라미터에 관한 정보도 포함한다.

[0003] 일부 알려진 구현예는 코딩 테이블 데이터(60)에 포함된 확률을 변경하기 위해 컨텍스트 정보를 이용한다. 컨텍스트 정보의 이용은 확률을 변경하기에 매우 효율적인 방법이지만, 소정의 데이터 유형에만, 예컨대 이미지에서의 픽셀의 값, 또는 이미지 정보가 구현된 후의 값의 코딩 시에만 효율적이다.

[0004] 레인지 코덱(range codec)의 기본 구현이 예컨대 Wikipedia Range coding site [Range encoding - Wikipedia, the free encyclopedia (accessed December 13, 2013), URL: http://en.wikipedia.org/wiki/Range_encoding]에 기재되어 있다. 이 기본 구현은 적어도 하나의 <EOM> (end-of-message) 심볼을 이용한다. 산술 코딩(Arithmetic coding), URL: http://en.wikipedia.org/wiki/Arithmetic_coding이 레인지 코딩과 매우 유사하지만, 산술 코딩은 하나의 부동소수점 수 [0, 1]와 <EOD> (end-of-data) 심볼을 이용한다. 또한, 레인지 코딩은 비디오 코딩 및 이미지 코딩 등의 데이터 압축을 필요로 하는 적용분야에서 폭넓게 이용된다. 이 방법은 다른 가능한 심볼의 확률이 알려진 경우에 대량의 데이터 세트를 코딩할 때에 매우 유용하다. 그러나, 코딩될 데이터의 특성상 하나 이상의 디코더(80)에서 심볼의 확률이 미리 알려지지 않게 되어 있다면, 인코딩된 데이터(E2; 30)와 함께 심볼의 확률을 송신해야 한다. 그렇지 않으면, 레인지 코딩의 압축 효율이 크게 손상된다.

[0005] 빈도, 확률 또는 레인지 값의 전달은 추가 데이터의 내포를 필요로 하는데, 즉, "데이터를 소모"하기 때문에, 인코더(20)에서 달성 가능한 전체 압축 효율에 영향을 미친다. 또한, 더욱 정확한 심볼 빈도, 확률 또는 레인지

값이 송신될 때에, 레인지 코딩 압축 효율의 향상을 달성할 수 있다. 그러나, 매우 정확한 빈도, 확률 또는 레인지 값의 송신에는 더 큰 데이터 사이즈를 필요로 한다. 다시 말해, 심볼에 대한 정확한 빈도, 확률 또는 레인지 값을 송신하는 것과 인코딩된 데이터(E2)에서의 최상의 가능한 압축 효율을 달성하는 것 사이에는 트레이드 오프가 존재한다.

[0006] 동시대적 상황에 있어서, 세계적으로 전달 및 저장되는 데이터 양이 급속하게 증가하고 있다. 전송 대역폭과 데이터 스토리지가 제한적인 파라미터이기 때문에, 효율적인 데이터 압축 방법의 필요는 매우 바람직하다. 이에, 예컨대 압축될 데이터의 사이즈가 매우 크지 않은 경우에, 레인지 코딩을 이용함으로써 달성 가능한 강화된 압축 효율을 제공하는 데이터 압축 방법의 개선에 대한 요구가 있다. 따라서, 레인지 코딩 및 유사 방법을 이용하여 이전에 실행되지 못한 그리고/또는 효과를 보지 못한 경우에 레인지 코딩이 이용되게 할 수 있는 방법에 대한 요구가 있다.

[0007] 선행 공개된 유럽 특허문헌 EP 0166607 A2 (King Reginald Alfred)에는, (오디오) 데이터의 샘플, 구체적으로 시간 인코딩된 심볼(time encoded symbol, TES)의 전송의 압축에 있어서 선형 표현(고정 길이 코드) 대신에 비선형 방식(가변 길이 코드)으로 비트를 제공함으로써 잠재적으로 개선되는 방법이 기재되어 있다. 이 방법은 프리커서 코드를 사용하는데, 이 코드는 인코더로부터 디코더로 엘리먼트의 발생의 확률 서열을 전달하는데 이용된다. 즉, 이 방법은 엘리먼트가 도착하는 동안에 실제로 전송하는 것은 확률이 아니라 그 서열이다.

[0008] 선행 특허 문헌 US 6650996 B1 (Garmin Ltd.)에는, 데이터 구조를 채택하는 시스템, 디바이스 및 방법이 기재되어 있으며, 여기에서는 정규 허프만 인코딩 데이터를 디코딩하는 디코딩 구조를 정의하는 제1 데이터 필드와, 심볼 테이블을 정의하는 제2 데이터 필드가 이용된다. 또한, 심볼 테이블은 저빈도 인덱스와 연관된 심볼을 제공하도록 적용된다.

발명의 내용

[0009] 본 개시내용은 데이터(D1)를 인코딩하여 대응하는 인코딩된 데이터(E2)를 생성하는 개선된 방법을 제공하며, 이 방법에 의해 예컨대 레인지 코딩의 상황에서 인코딩된 데이터(E2)의 데이터 압축이 향상된다.

[0010] 본 개시내용은 또한 전술한 개선된 방법을 채택하는 개선된 인코더를 제공하는 것을 목적으로 한다.

[0011] 제1 양태에 따르면, 데이터(D1)를 인코딩하여 대응하는 인코딩된 데이터(E2)를 생성하는 인코더가 제공되며, 인코딩된 데이터(E2)는 이 인코딩된 데이터(E2)에서 표현될 하나 이상의 심볼의 빈도, 확률 또는 레인지를 나타내는 정보를 포함하고, 상기 인코더는 하나 이상의 심볼 중의 적어도 하나의 심볼에 대한 빈도, 확률 또는 레인지를 나타내는 정보가 상기 인코딩된 데이터(E2)에 포함되어 있는지의 여부를 나타내는 추가 정보를 상기 인코딩된 데이터에 포함시키도록 동작 가능한 것을 특징으로 한다.

[0012] 본 발명은, 인코딩될 데이터(D1)에서의 심볼의 발생의 함수로서, 인코딩된 데이터(E2)에서의 빈도, 확률 또는 레인지 정보의 선택적 내포가 코딩 확률 - 예를 들어 소량의 데이터(D1)를 인코딩할 때에 매우 효과적인 방법으로 예컨대 달성 가능한 압축률 - 을 향상시킬 수 있다는 점에서 유리하다.

[0013] 전술한 표현 "...의 여부를 나타내는 정보"에 대해서는 이하에서 상세하게 설명한다.

[0014] 제1 양태에 따른 인코더는, 개선된 방법을 이용하여 생성되는 인코딩된 데이터와 연관시켜 레인지 코딩 빈도, 확률 또는 레인지 값의 효율적인 전달이 달성되게 할 수 있는 개선된 방법을 채택한다는 점에서 유리하다.

[0015] 그러나, 전술한 개선된 방법은 코딩될 심볼 수가 가능한 심볼 수에 비해 매우 많지 않은 경우에 레인지 코딩을 이용할 수 있다. 알려진 레인지 코딩 기법에서는, 빈도, 확률 또는 레인지 값을 전달하는데 요구되는 비트 수 때문에 그러한 데이터를 인코딩하는 것이 적절하지 못한 그리고/또는 실용적이지 못했는데, 즉 알려진 레인지 코딩 기법을 이용하여 빈도, 확률 또는 레인지 정보를 인코더로부터 대응하는 디코더로 전달하게 되면 연관된 비실용적인 데이터 오버헤드가 발생하게 된다.

[0016] 또한, 제1 양태에 따른 개선된 방법은 하나 이상의 <EOM> 심볼의 유무에 관계없이 이용될 수 있으며, "EOM"은 "end-of-message"를 나타내는데 이용된다. 추후에 설명하는 실시형태의 예들은 모두 하나 이상의 <EOM> 심볼이 없는 것으로 보여지지만, 하나 이상의 <EOM> 심볼은, 예컨대 그것에 대해 확률 값 1을 사용하고 기타 확률 값으로부터 1을 차감함으로써 실시형태의 모든 예에 경우에 따라 추가될 수 있다. 이러한 솔루션에는 확률 값이 이미 알려져 있기 때문에, 새로운 확률 값의 전달을 필요로 하지 않는다. <EOM> 심볼의 확률은 또한 경우에 따라 데이터 심볼의 양에 기초하여 산출될 수도 있다. 그 확률이 산출될 때에, <EOM> 심볼이 확률 테이블에

추가되고, 그것에는 빈도 값 1이 주어질 것이다. 확률의 합은 또한 1씩 증가하며 이것은 모든 심볼 확률에 영향을 끼친다. 이것은 또한 추가 확률 값이 <EOM> 심볼에 대해 전달되는 것도 필요로 한다.

[0017] 제1 양태에 따른 개선된 방법은 임의 유형의 데이터를 인코딩하는데 이용될 수 있고, 이에, 잠재적으로 이용될 분야가 넓으며, 인코딩될 상이한 "임의 유형의 데이터"는 경우에 따라, 유전학 시퀀싱(genetic sequencing) 데이터, DNA 시퀀싱 데이터, RNA 시퀀싱 데이터, 캡처된 오디오 신호, 캡처된 비디오 신호, 캡처된 이미지, 텍스트 데이터, 지진계 데이터(seismographic data), 센서 데이터, 아날로그-디지털(ADC) 변환 데이터, 생체 신호 데이터, 캘린더 데이터, 경제 데이터, 수학 데이터, 이진 데이터 중 하나 이상을 포함하지만, 이들에 한정되지는 않는다. 개선된 방법에 있어서, 확률을 변경하는 다른 방법은, 데이터를 코딩하여 대응하는 인코딩된 데이터를 생성하면서 적응적 확률 및 레인지를 이용하는 것이다. 적응적 확률은 데이터 스트림 내 상이한 부분에서, 특히 심볼의 확률이 가변적으로, 그러나 느리게 변한다면 대량의 데이터 스트림을 인코딩하는데 효과적이다. 그러나, 적응적 확률은 소량의 데이터 스트림에는 덜 적합한데, 그 확률을 데이터 스트림 내의 심볼의 정확한 확률과 매칭하도록 적응시키는 데에 추가 컴퓨팅 리소스 및 시간이 필요하기 때문이다.

[0018] 제1 양태에 따른 개선된 방법은 인코더에서 이용 가능하고 개선된 방법의 대응하는 역방법은 대응하는 인코더에서 이용 가능하며, 전술한 바와 같이 인코더와 디코더를 조합한 것이 코덱을 구성한다. 선택사항으로서, 디코더에서의 디코딩된 데이터를 재포맷하기 위해, 예컨대 호환성 데이터(compatible data)를 디코더의 데이터 렌더링 장치에 또는 디코더의 데이터 저장 장치에 제공하기 위해 디코더에서 트랜스코딩(transcoding)이 수행된다. 개선된 방법은 데이터 스트림 심볼 중에 존재할 수도 있는 상이한 심볼에 대해 레인지 값, 확률 값 또는 빈도 값이 전달되어야 하는지의 여부를 기술하기 위해 0 비트 및 1 비트 정보를 이용한다. 이 개선된 방법은 경우에 따라 하나 이상의 확률 임계 값을 이용하여, 어떤 심볼이 그 자신의 전달 확률 값, 즉 1 비트를 필요로 하는지 그리고 어떤 심볼이 예컨대 0 비트를 이용하여 이스케이프(escape)될 수 있는지를 결정한다. 선택사항으로서, 모든 이스케이프 데이터 심볼은 인코더에서 그리고 디코더에서 확률 값 1을 이용할 수 있다. 대안적으로, 그 자신의 빈도 또는 확률 값, 즉 레인지를 갖지 않는 그 모든 데이터 심볼은 레인지 코딩 방식으로, 경우에 따라 이스케이프 심볼, 즉 레인지에 의해 전달된다. 부가적으로, 그 이스케이프 심볼에 대한 심볼 값 전부는 경우에 따라 압축 없이 예를 들어 원래의 값으로서 전달된다.

[0019] 다음과 같은 개념 (i) 내지 (iii)에 대해 이하에서 상세하게 설명할 것이다.

[0020] (i) "코딩된 심볼이 알려져 있음" 또는

[0021] (ii) 인코더에서, "하나 이상의 심볼 중의 적어도 하나의 심볼에 대한 빈도, 확률 또는 레인지를 나타내는 정보가 인코딩된 데이터에 포함되어 있는지의 여부" 또는

[0022] (iii) 디코더에서, "인코딩된 데이터 내의 인코딩된 어떤 심볼이 빈도, 확률 또는 레인지 정보를 이용하여 디코딩되는지"

[0023] 본 개시내용의 실시형태에 있어서, 심볼 그 자체가 항상 알려져 있어야 하는 것은 중요하지 않고, 확률이 관계하는 것이 어떤 심볼인가가 중요하다. 이에, 본 개시내용의 실시형태에 있어서, 예를 들어 데이터의 큐에 대해, 확률이 관계하는 심볼에 대해 알려져 있지만, 어떤 실제 심볼이 실행중인지의 상황에 대해서 반드시 알고 있을 필요는 없다. 이에, 예를 들어 디코더에 전달되며 적어도 하나의 심볼에 대해 이용 가능한 확률 정보가 있어야 한다. 그러나, 어떤 실제 심볼이 실행중인지에 대해서는 아는 것은 중요하지 않다. 또한 선택사항으로서, 심볼의 값을 실제로 알고 있을 가능성도 있으며, 그러한 경우에 그 정보도 물론 이용 가능하다.

[0024] 다음으로, 추가 예에 대해 설명할 것이다. 빈도, 확률 또는 레인지 테이블이 전달될 때에, 전달에 있어서 중요한 파라미터는 테이블의 최대 인덱스와, 전달된 확률로 코딩되는 심볼에 대한 정보이다. 테이블의 최대 인덱스는 상이한 심볼의 수를 나타내거나, 가능한 상이한 심볼의 수는 전달된 테이블에서 이용 가능하며, 즉 입력 데이터(D1) 내에 존재한다. 예는 다음과 같다.

[0025] 4, 3, 0, 1, 0, 4, 3, 4

[0026] 실제의 최대 인덱스는 4이고 (그리고 최소 인덱스는 0이다), 이것은 데이터 내 존재하는 상이한 심볼(0, 1, 2, 3, 4)이 5개(max - min + 1 = 4 - 0 + 1)임을 의미한다. 데이터(D1) 내에 사실상 상이한 심볼이 4개(0, 1, 3, 4)만 존재할 경우, 테이블은 가능한 상이한 심볼수 4 대신에, 최대 인덱스로서 3, 즉 이용 가능한 상이한 심볼수를 이용해서 또한 전달될 수 있다. 값 '3'을 테이블의 최대 인덱스 값에 이용할 경우, 선택사항으로서, 각각의 테이블 인덱스에 무슨 심볼이 이용되는지에 관한 정보를 전달하기 위해 일부 다른 메커니즘이 채택된다.

- [0027] 이러한 방법에 대해서는 본 개시내용의 다른 곳에 상세하게 설명되어 있다. 심볼이 순서를 따른다면, 한가지 가능성은 실제 최대 인덱스(4)와 가용성 비트(예컨대, 이 경우에는 11011)를 전달하는 것이다. 이러한 종류의 전달은 "인덱스"라고 불리는 테이블 인덱스 0=심볼 0이 되도록 맵핑하여, 심볼 쌍은 (0, 0)이 될 것이다. 마찬가지로, 인덱스 및 심볼 쌍의 나머지는 (1, 1), (2, 3) 및 (3, 4)이다. 또한, 상이한 테이블 인덱스에 대해 사용된 인덱스를 정의하는 데에 직접 인덱스 및 심볼 쌍을 이용한 다음 테이블의 최대 인덱스를 3으로서 또는 상이한 인덱스 및 심볼 쌍의 카운트를 4로서 전달하는 것도 가능하다. 이러한 방법을 채택함으로써, 코드 테이블 인덱스는 통상 순서를 따르며, 심볼만이 전달되어야 한다. 이러한 방법은, 전달된 테이블 내의 심볼이 그 빈도에 기초하여 정렬될 때에 채택되는 것이 매우 유리하다.
- [0028] 예를 들어, 이 경우에 인덱스 및 심볼 쌍은 (0, 4), (1, 0), (2, 3), 및 (3, 1)일 수 있고, 값 4, 0, 3, 1만 순서를 따라 전달된다. 이 상황에서,
- [0029] (a) 그 사용되는 인덱스 및 심볼 쌍은 미리 정의되어 있다.
- [0030] (b) 때때로, 사용된 심볼의 인덱스 및 인덱스 쌍 테이블이 전달되고, 다른 때에는 인덱스 및 심볼 쌍이 인코딩된 데이터와 함께 전달된다.
- [0031] (c) 또 다른 때에는, 디코더가 알려진 위치로부터 이용된 인덱스 및 심볼 쌍을 검색할 수 있다.
- [0032] (d) 또 다른 상황에서는, 디코더가 인코딩된 데이터와 함께 전달된 위치로부터 이용된 인덱스 및 심볼 쌍을 검색할 수 있다.
- [0033] 제1 양태에 따른 전술한 개선된 방법을 채택하는 인코더에서 인코딩될 데이터의 부분의 빈도는 종종 서로 상이하며, 종종 그 상대적 데이터 엔트로피도 서로 상이하며, 이러한 이유에서, 인코딩 데이터(E2)를 생성하기 위해 인코딩할 때에 데이터를 다수의 부분으로 분할, 즉 세분하는 것이 유리하다. 유리하게는, 상이한 코드 테이블이 상이한 부분에 이용된다. 개선된 방법은, 예를 들어, 하나의 대형 데이터 청크가 더 작은 부분으로, 즉 더 작은 데이터 청크로 더욱 효율적으로 분할되는 것을 가능하게 하는데, 그렇게 되면 코드 테이블 또는 빈도 테이블의 전달이 최적화되기 때문이다. 이러한 인코딩될 데이터의 분할은 분할 시에, 인코딩될 데이터(E1)의 엔트로피에 대해 상당한 효과를 줄 수 있으며, 그래서 대응하는 인코딩되지 않은 데이터(D1)에 비해 인코딩된 데이터(E2)의 향상된 압축 정도를 제공하는 것이 가능하다.
- [0034] 선택사항으로서, 전술한 제1 양태에 따르면, 인코더는 하나 이상의 심볼 중의 적어도 하나의 심볼에 대한 빈도, 확률 또는 레인지를 나타내는 정보가 인코딩된 데이터(E2)에 포함되어 있는지의 여부에 관한, 단일 가용성 비트 정보로서 표현될 추가 정보를 인코딩된 데이터(E2)에 포함시키도록 동작 가능하다. 또 선택사항으로서, 인코더는 인코딩된 데이터(E2)에서의 빈도, 확률 또는 레인지 정보를 나타내는 정보의 내포(inclusion)를 단일 가용성 비트 값 "1"을 이용하여 나타내고, 인코딩된 데이터(E2)에서의 빈도, 확률 또는 레인지 정보를 나타내는 정보의 내포하지 않음(non-inclusion)을 단일 가용성 비트 값 "0"을 이용하여 나타내도록 동작 가능하다. 이러한 가용성 비트 값은 반대로도 이용될 수 있다. 또한, 이러한 가용성 정보는 이용 가능한 인덱스의 수 그리고 존재하는 정보의 인덱스로서 전달될 수 있다. 이들 인덱스는 경우에 따라 그 인덱스 그대로 또는 델타 코딩을 이용하여 선택적으로 전달된다. 선택사항으로서, 상이한 코딩 방법이 가용성 정보를 전달하는데 필요한 데이터 양을 저감하는 데에 이용된다. 선택된 코딩 방법은 경우에 따라, 이용 가능한 인덱스의 수를 포함하거나, 상기 코딩 방법은 인코딩 테이블이 전달될 때에 심볼 전체에 대해 가용성 비트 값이 전달되는 것을 필요로 한다.
- [0035] 선택사항으로서, 인코더에서, 인코딩된 데이터(E2)에서 표현될 하나 이상의 심볼의 빈도, 확률 또는 레인지를 나타내는 정보는 인코딩될 데이터의 본성, 즉 특성의 함수에 따라 인코더에 의해 동적으로 변하며, 이러한 특성은 경우에 따라, 데이터(D1)의 구조, 데이터(D1)의 사이즈, 데이터(D1)에 존재하는 값의 레인지, 데이터(D1)에 존재하는 엘리먼트의 사이즈, 데이터(D1)와 연관된 메타데이터 중 하나 이상과 관련된다. 또한, 이러한 특성은 예컨대 데이터(D1)의 유형(type), 데이터(D1)의 성분(content), 및/또는 데이터(D1)의 통계를 나타낸다. 또 선택사항으로서, 인코더는 인코딩될 데이터(D1)를 복수의 부분으로 세분하고, 각 부분마다 하나 이상의 심볼 중의 적어도 하나의 심볼에 대한 빈도, 확률 또는 레인지를 나타내는 정보가 인코딩된 데이터(E2)에 포함되어 있는지의 여부를 나타내는 대응하는 추가 정보를 개별적으로 생성하도록 동작 가능하다.
- [0036] 선택사항으로서, 인코더는, 유전학 시퀀싱 정보, DNA 시퀀싱 정보, RNA 시퀀싱 정보, 캡처된 오디오 신호, 캡처된 비디오 신호, 캡처된 이미지, 텍스트 데이터, 지진계 데이터, 센서 데이터, 아날로그-디지털(ADC) 변환 데이터, 생체 신호 데이터, 캘린더 데이터, 경제 데이터, 수학 데이터, 이진 데이터 중 적어도 하나에 대응하나 이들에 한정되지 않는 데이터를 인코딩하도록 동작 가능하다.

- [0037] 제2 양태에 따르면, 인코더에서 데이터(D1)를 인코딩하여 대응하는 인코딩된 데이터(E2)를 생성하는 방법이 제공되며, 인코딩된 데이터(E2)는 그 인코딩된 데이터(E2)에서 표현될 하나 이상의 심볼의 빈도, 확률 또는 레인지를 나타내는 정보를 포함하고, 상기 방법은, (a) 인코더를 이용해서, 하나 이상의 심볼 중의 적어도 하나의 심볼에 대한 빈도, 확률 또는 레인지를 나타내는 정보가 인코딩된 데이터(E2)에 포함되어 있는지의 여부를 나타내는 추가 정보를 상기 인코딩된 데이터(E2)에 포함시키는 단계를 포함한다.
- [0038] 선택사항으로서, 상기 방법은, (b) 인코더를 이용해서, 하나 이상의 심볼 중의 적어도 하나의 심볼에 대한 빈도, 확률 또는 레인지를 나타내는 정보가 인코딩된 데이터(E2)에 포함되어 있는지의 여부에 관한, 단일 가용성 비트 정보로서 표현된 추가 정보를 인코딩된 데이터(E2)에 포함시키는 단계를 더 포함한다.
- [0039] 선택사항으로서, 상기 방법은, (c) 인코더를 이용해서, 인코딩된 데이터(E2)에서의 빈도, 확률 또는 레인지 정보를 나타내는 정보의 내포(inclusion)를 단일 가용성 비트 값 "1"을 이용하여 나타내고, 인코딩된 데이터(E2)에서의 빈도, 확률 또는 레인지 정보를 나타내는 정보의 내포하지 않음(non-inclusion)을 단일 가용성 비트 값 "0"을 이용하여 나타내는 단계를 더 포함한다. 선택사항으로서, 값 "0" 및 "1"은 반대로 이용되어 상기 내포 또는 내포하지 않음을 나타낼 수도 있다.
- [0040] 선택사항으로서, 상기 방법은, 인코더를 이용해서, 인코딩된 데이터(E2)에서 표현될 하나 이상의 심볼의 빈도, 확률 또는 레인지를 나타내는 정보를, 인코딩된 데이터(D1)의 본성, 즉 하나 이상의 특성의 함수에 따라 동적으로 변화시키는 단계를 포함한다. 기술적인 용어로서 "본성(nature)" 및 "특성(characteristics)"의 정의에 관해서는 전술한 바를 참조하면 된다.
- [0041] 선택사항으로서, 상기 방법은, 인코더를 이용해서, 인코딩된 데이터(D1)를 복수의 부분으로 세분하고, 각 부분마다 하나 이상의 심볼 중의 적어도 하나의 심볼에 대한 빈도, 확률 또는 레인지를 나타내는 정보가 인코딩된 데이터(E2)에 포함되어 있는지의 여부를 나타내는 대응하는 추가 정보를 개별적으로 생성하는 단계를 포함한다.
- [0042] 선택사항으로서, 상기 방법은, 유전학 시퀀싱 데이터, 캡처된 오디오 신호, 캡처된 비디오 신호, 캡처된 이미지, 텍스트 데이터, 센서 데이터 중 적어도 하나에 대응하나 이들에 한정되지 않는 데이터(D1)를 인코딩하는 단계를 포함한다. 종종, 생물학적 샘플로부터 전기영동 장치에 의해 생성되는 유전학 시퀀싱 데이터는 상당히 하며, 종종 1 테라바이트 데이터의 오더(order)에 근접하여 핵염기 및/또는 아미노산을 정의한다. 이러한 데이터의 통신과 함께 이러한 데이터의 취급은 중요한 동시대의 기술적 과제를 나타낸다.
- [0043] 제3 양태에 따르면, 제1 양태에 따른 인코더에 의해 생성되는 인코딩된 데이터(E2)를 디코딩하는 디코더가 제공된다.
- [0044] 제4 양태에 따르면, 디코더에서 인코딩된 데이터(E2)를 디코딩하는 방법이 제공되며, 이 방법은 제2 양태에 따른 방법의 단계들의 역(inverse) 단계들을 수행하는 것을 포함한다.
- [0045] 선택사항으로서, 상기 방법은,
- [0046] (a) 가용성 신호를 수신하는 단계와,
- [0047] (b) 인코딩된 데이터(E2) 내에 송신된 빈도, 확률 또는 레인지 테이블을 수신하는 단계와,
- [0048] (c) 단계 (a)와 (b)로부터 완전한 빈도 테이블, 확률 테이블 또는 레인지 테이블을 구성하는 단계와,
- [0049] (d) 완전한 빈도 테이블, 확률 테이블 또는 레인지 테이블을 이용해서, 상기 인코딩된 데이터(E2)를 레인지 디코딩하여 디코딩된 출력 데이터(D3)를 생성하는 단계를 포함한다.
- [0050] 선택사항으로서, 상기 방법에 있어서, 가용성 신호는 적어도 하나의 가용성 표시 비트를 포함한다.
- [0051] 선택사항으로서, 상기 방법에 있어서, 가용성 신호는 적어도 하나의 가용성 표시 코딩 방법을 포함한다.
- [0052] 선택사항으로서, 상기 방법에 있어서, 가용성 신호는 적어도 하나의 가용성 표시 인덱스 및/또는 인덱스 수를 포함한다.
- [0053] 제5 양태에 따르면, 제1 양태에 따른 적어도 하나의 인코더와, 제3 양태에 따른 적어도 하나의 디코더를 포함하는 코덱이 제공된다.
- [0054] 제6 양태에 따르면, 컴퓨터 관독 가능한 명령어가 저장되어 있는 비일시적 컴퓨터 관독 가능한 매체를 포함하는 컴퓨터 프로그램 제품이 제공되며, 이 컴퓨터 관독 가능한 명령어는 제2 양태에 따른 방법을 구현하도록 제1 양

태의 인코더의 컴퓨팅 하드웨어에 의해 실행 가능하다.

[0055] 제7 양태에 따르면, 컴퓨터 판독 가능한 명령어가 저장되어 있는 비일시적 컴퓨터 판독 가능한 저장 매체를 포함하는 컴퓨터 프로그램 제품이 제공되며, 이 컴퓨터 판독 가능한 명령어는 제4 양태에 따른 방법을 구현하도록 제3 양태의 디코더의 컴퓨팅 하드웨어에 의해 실행 가능하다.

[0056] 본 발명의 특징들은 첨부하는 청구범위에 의해 정해지는 발명의 범주로부터 이탈하지 않고서 다양한 조합으로 조합될 수도 있음은 물론이다.

도면의 간단한 설명

[0057] 이하, 본 개시내용의 실시형태들에 대해 다음의 도면을 참조하여 예시적으로만 설명한다.

도 1은 인코더 및 대응하는 디코더의 예시로서, 인코더는 입력 데이터(D1)를 인코딩하여 대응하는 인코딩된 출력 데이터(E2)를 생성하도록 동작 가능하고, 디코더는 인코딩된 데이터(E2)를 디코딩하여 대응하는 디코딩된 데이터(D3)를 생성하도록 동작 가능하다.

도 2는 본 개시내용에 따른 데이터 인코딩 방법의 흐름도이다.

도 3은 본 개시내용에 따른 데이터 디코딩 방법의 흐름도이다.

첨부하는 도면에 있어서, 밑줄 있는 도면부호는 그 도면부호가 위치하는 아이템 또는 그 도면부호에 인접한 아이템을 나타내는데 이용된다. 도면부호에 밑줄이 없고 연관 화살표를 수반하는 경우, 그 밑줄 없는 도면부호는 화살표가 가리키는 일반적인 아이템을 식별하는데 이용된다.

발명을 실시하기 위한 구체적인 내용

[0058] 개관에 있어서, 데이터(D1)를 인코딩하는 레인지 코딩 채택 시에, 코딩되는 데이터(D1) 내의 심볼의 확률이 알려져야 하며, 그렇지 않으면 그러한 레인지 코딩의 달성 가능한 압축 효율이 그에 따라 심각하게 손상된다. 각각의 심볼의 확률은 코딩 동안의 "레인지(range)"를 정의하는데, 확률이 높을수록 레인지가 넓다. 심볼이 코딩될 때에, 대응하는 레인지는 코딩된 심볼의 확률에 따라 업데이트된다. 레인지가 상당히 작아지면, 레인지 코더는 예를 들어 바이트를 방출하고 레인지를 증가시킨다. 이에, 공통 심볼이 많은 경우에는, 그 레인지가 천천히 축소되고, 바이트가 방출되기 전에 더 많은 심볼이 코딩될 수 있다.

[0059] 인코더에서 레인지 코딩을 구현할 때에, 심볼에 대한 빈도 테이블 인덱스, 확률 테이블 인덱스 또는 레인지 테이블 인덱스, 또는 빈도, 확률 또는 레인지 값이 일부 소정의 정밀도를 이용하여 인코더에 의해 생성되는 인코딩된 데이터(E2)로 송신되며, 대응하는 디코더에서 이미 알려지지 않다면, 즉 선택적이라면, 경우에 따라, 인코더는 인코딩된 데이터(D1)의 서로 상이한 부분에 대해 빈도 테이블 인덱스, 확률 테이블 인덱스, 레인지 테이블 인덱스, 빈도 값, 확률 값 및 레인지 값을 이용하여 동적으로 전환하도록 동작 가능하다. 예를 들어, 레인지 코덱 내에서 8 비트 데이터가 코딩되고 있는 경우에는, 잠재적으로 2^8 심볼, 즉 256 심볼이 발생한다. 8 비트, 즉 1 바이트가 각각의 빈도, 확률 또는 레인지 값을 송신하는데 이용되었다면, 빈도, 확률 또는 레인지 값 전부를 송신하는 데에는 256 바이트가 필요하게 된다. 일부 경우에 있어서, 레인지 인코더로부터 취득되는 압축 데이터 양보다 256 바이트가 잠재적으로 더 크기 때문에, 빈도, 확률 또는 레인지 값을 전달하는 데에 더 적은 바이트를 이용해야 할 분명한 필요가 있고, 다시 말해 인코더로부터 빈도, 확률 또는 레인지 값을 디코더에 전달하는 것은 인코딩된 데이터의 실제 양보다 잠재적으로 더 많은 데이터 오버헤드를 나타내며, 이것은 더 낮은 압축률 성능에 대응한다.

[0060] 본 개시내용에 따른 데이터(D1) 인코딩 방법은, 심볼 빈도, 확률 또는 레인지 값 외에도, 가용성 신호, 예를 들어 매 심볼당 1 가용성 비트도 송신된다는 점에서, 알려진 방법과 구별된다. 이 가용성 비트의 값은, 인코딩된 데이터(E2)에서 표현되며 충분히 높은 확률을 갖는 모든 심볼에 대해 "일", 즉 "1", 그리고 인코딩된 데이터(E2)에서 표현되지 않는 모든 심볼에 대해서는 "제로", 즉 "0", 또 인코딩된 데이터(E2)에서 표현되지만 상대적으로 낮은 확률을 가진 모든 심볼에 대해서도 "제로"인 것이 유리하다.

[0061] 본 개시내용에 따른 데이터(D1) 인코딩 방법에 있어서, 심볼이 충분히 낮은 확률을 갖는지의 여부는, 예를 들어 심볼의 확률을 임계치와 비교함으로써, 경우에 따라, 1 임계치보다 높은지 확인된다. 심볼을 인코딩할 때에, 인코딩된 데이터(E2)에 존재하지 않거나 확률이 임계치보다 낮은 심볼에 대한 확률은 '제로'(이스케이프 코드가 이용되고 그 심볼에 대해 유지되는 레인지가 없음)이거나 '일'(즉, 최소의 가능한 레인지를 갖는 낮은 확률 심

볼이 전달)이고, 즉 '0' 또는 '1'은 인코딩된 데이터(E2; 30)를 생성하는 인코더(20)에 제공되는 입력 데이터(D1; 10)에 이용되는 전체 레인지의 사이즈와 무관하며, 이러한 것은 인코더(20)와 하나 이상의 디코더(80)가 개선된 방법(T)과 그 역방법(T⁻¹)을 각각 채택할 때에 관계된다. 빈도, 확률 또는 레인지 및 가용성 신호(예를 들어, 이 경우에는 가용성 비트)를 송신하는데 필요한 인코딩된 데이터의 양 A(비트)는, 충분히 높은 확률을 가진 인코딩된 데이터 내에 존재하는, 또는 이용 가능한 심볼의 수 N_S를, 각각의 빈도, 확률 또는 레인지 값에 이용되는 비트의 수 N_B와 곱해서 모든 심볼 0_S에 대한 1 가용성 비트를 더한 것이며, 다시 말해, 다음과 같다.

[0062] $A = N_S * N_B + 0_S$ (식 1)

[0063] 본 개시내용에 다른 개선된 방법의 다음 예는 입력 데이터(D1; 10)로서 데이터 값의 세트를 인코딩하여 대응하는 인코딩된 데이터(E2; 30)를 생성하는 인코더(20)에 적용된다. 이 예에서는, 코딩된 데이터 스트림(D1)이, 20 심볼(=값), 즉 0 내지 19를 포함할 수 있는 150 데이터 값을 포함하지만, 그 20개의 가능한 데이터 값 중 8개만이, 즉 min=2 및 max=19가 현재 데이터 스트림에서 정확하게 표현된다. 이들 min 및 max 값(min = "최소", max = "최대")은, 예컨대 Gurulogic Microsystems Oy에 의해 2013년 3월 1일에 출원된 특허 출원 GB1303661.1에 개시된 ODelta 코딩 방법에서 설명한 바와 같이 경우에 따라 전달되며, "ODelta" 인코딩은 데이터의 특히 효율적인 인코딩을 제공하기 위한 랩어라운드 연산(wraparound computation) 및 오프셋을 가진 델타 인코딩 방법에 관한 것이다. 이것은 가용성 신호는 전달하는데 필요한 일부 비트의 감소를 가능하게 한다. 또한, 이것은 레인지 코딩이 채택되기 전에, 예컨대 ODelta 코딩 등의, 엔트로피를 변경하는 일부 방법이, 이용될 경우에 가능하게 이용할 수 있는 심볼 값의 데이터 레인지도 삭감한다. 엔트로피를 변경하는 방법 및 레인지 코딩 모두가 동일한 정보를 이용하고 그 정보가 인코더로부터 대응하는 디코더에 한번만 전달될 경우에 min 및 max 값의 이용과 연관된 최상의 효과가 달성된다. 이하의 표 1은 예컨대 심볼의 빈도 및 확률을 나타낸다. 확률은 빈도를 데이터 값의 총 수로 나누어서 산출된다.

[0064] [표 1] 본 개시내용에 따른 레인지 인코딩 예

값 (데이터 내 심볼)	빈도	확률	7비트로 된 확률	7비트로 된 확률 (이스케이프)	7비트로 된 확률 (변경)	6비트로 된 확률 (변경)
2	7	0.047	6	6	6	2
4	2	0.013	2	(0)	(1)	(1)
7	81	0.54	69	69	63	28
9	1	0.007	1	(0)	(1)	(1)
12	39	0.26	33	33	30	13
13	9	0.06	8	8	7	3
14	3	0.02	2	2	2	(1)
19	8	0.053	7	7	6	3
이스케이프	0	0.0	(0)	3	x 12	x 12
합	150	1.000	128	128	114+14=128	49+15=64

[0065]

[0066] 표 1에서는, 발생 확률이 매우 낮은 두어개 심볼이 있다. 확률을 전달하는데 7 비트가 이용되는 경우에, 각각의 확률을 기술하는데 동일한 양의 비트가 사용되었다면, 확률 전부를 표현하는 데에는 20*7 = 140 비트가 필요하게 된다. 예를 들어 심볼당 1 가용성 비트를 송신함으로써, 스트림, 즉 인코딩된 데이터(E2; 30) 내의 심볼의 존재가 표현되면, 확률 전부를 표현하는 데에 여전히 8*7 + 20*1 = 76 비트가 필요하게 된다.

[0067]

본 개시내용에 따른 개선된 방법을 구현하는 이 최초 실시형태는, 전달된 확률의 수를 20에서 8로 감소시키기 위해 별도의 가용성 비트를 이용하기 때문에, 알려진 솔루션과 상이하다. 후술하는 본 개시내용의 다른 실시형태에 대해, 경우에 따라, 주어진 확률이 디코더(80)에 송신되는지의 여부를 결정하는 임계치, 예컨대 그 임계치가 경우에 따라 7 비트 확률로 0.018에 설정되는 것을 보장하는 것이 유리하다. 임계치의 변경은, 확률을 표현하여 하나 이상의 디코더(80)에 전달하는데 필요한, 인코딩된 데이터(E2; 30) 내의 비트 수를 변화시킨다. 표 1에서 기술하는 데이터에 대한 임계치로서 0.10이 선택되면, 2개의 심볼만 남게 될 것이고, 인코딩된 데이터(E2; 30)에서 확률을 표현하는 데에 2*x + 20*1 비트가 필요하게 되는데 여기서 x는 확률을 표현하는데 이용되는 비트 카운트를 나타낸다.

[0068]

본 개시내용의 개선된 방법을 구현할 때에 이용하기 위한 최적의 임계치를 구하는 것은, 인코더(20)로부터 하나 이상의 디코더(80)에 송신될 인코딩된 데이터(E2; 30)의 양에 그리고 그 인코딩된 데이터(E2; 30)에서 표현되는 심볼의 빈도에 종속된다. 하나 이상의 전술한 임계치를 결정하는 방식에 대해 이제 상세하게 설명한다. 예를 들어,

[0069] T = 임계치

[0070] B = 확률을 표현하는데 이용되는 비트 카운트

[0071] 라고 하면, 예컨대, 앞의 표 1에서는, 6 또는 7 비트 중 하나로 표현 가능하다. 바람직하게는, 임계치는 주어진 권고(Recommended) 한도 내에 있고, 식 2(Eq.2)에 의해 결정되는 바와 같이, 임계 하한(T_{Low})보다 작지 않다.

$$T_{Low} = \frac{1}{2^B}$$

$$\frac{2}{2^B} \leq T_{Recommended} \leq \frac{8}{2^B} \quad \text{Eq. 2}$$

[0072]

[0073] 표 1의 하위 4개의 열에는 하나 이상의 디코더(80)에 송신되는 실제 확률이 표시된다. "7 비트로 된 확률" 열에는, 가용성 비트의 유무와 관계 없이 8 또는 20 데이터 값 전부에 대해 7 비트가 이용되는 경우에 송신된 확률이 표시되며, 이 경우 확률 신호는 스트림 내의 심볼의 존재를 표현하는 신호를 나타낸다. "7 비트로 된 확률 (이스케이프)"은 본 개시내용에 따른 개선된 방법의 제1 실시형태를 이용하여 송신된 확률에 대응하며, "7(또는 6) 비트로 된 확률 (변형)"은 본 개시내용에 따른 개선된 방법의 제2 실시형태를 이용하여 송신된 확률에 대응한다.

[0074] 표 1의 하위 3개의 열에 있는 괄호 안에 표시된 확률은 제로 가용성 비트를 이용해서 표현되기 때문에, 하나 이상의 디코더(80)에 사실상 송신되지 않는다. 이 예에서는, 하나 이상의 디코더(80)에 송신된 확률이 알려진 인코딩 방법으로 송신된 확률에 가깝기 때문에, 본 개시내용에 따른 개선된 방법을 이용하는 것은 실제 레인지 코딩의 효율에 유효하게 영향을 미치지 않는지만, 빈도, 확률 또는 레인지 값에 대해 송신되기에 필요한 비트 수를 삭감함으로써 전체 압축 효율을 상당히 향상시킬 수 있다. 또한, 이스케이프된 경우에, 또는 알려진 방법으로 송신되는 확률의 합은 128인 반면, 제2 실시형태에 따른 개선된 방법은 표에 나타내는 바와 같이 송신되는 확률의 합, 즉 114 또는 49 중 하나를 포함하는 것을 이해해야 될 것이다. 이것은 제로 가용성 비트로 표현된 각각의 데이터 값에 대해 확률 값 '1'을 유지하기 때문에 발생하고, 이들 데이터 값이 14 또는 15개가 있으며, 스트림 내에는 12 데이터 값은 존재하지 않고, 임계치보다 낮은 확률을 갖는 값은 2 또는 3개이다.

[0075] 이 예에서, 마찬가지로 7 비트와 임계치를 이용하여, 심볼 4와 9는 데이터 스트림 내에 존재하지 않는 12개의 심볼 전부와 함께, 제로 가용성 비트로 표현된다. 원래는, 데이터 스트림 내에 존재하지 않는 심볼에 대해서는, 이스케이프(escape)라고 불리는 레인지가 유지되지 않았다. 이제, 그 자신의 확률을 갖지 않는 심볼 4와 9도 전달하기 위한 메커니즘이 필요하게 될 것이다. 이것은 후술하는 2개의 상이한 실시형태를 이용해서 효과적으로 행해질 수 있다.

[0076] 앞의 표 1에서 "(이스케이프)"라고 표시된 제1 실시형태는 별도의 이스케이프 심볼 및 그것의 연관된 레인지가 형성되는 경우이다. 심볼이 그 자신의 레인지를 갖지 않을 경우, 그 심볼은 이스케이프(레인지)로 전달된다. 부가적으로, 심볼 값도 그 자신의 데이터 스트림 내에 전달된다. 심볼 값은 본 예에서 5 비트로 전달될 수 있는데, 심볼 값 레인지 0 내지 19이기 때문이다. 선택사항으로서, 가능한 모든 심볼로 나뉜 이용 가능한 심볼의 비가 높을 경우, 즉 0.25보다 높을 경우, 심볼 값은 그 심볼 값을 그것의 대응하는 가용성 0 비트 포지션과 함께 전달함으로써, 약간 더 작은 양의 비트로 전달되는 것이 유리하다. 예를 들어, 본 예의 경우, 비트 스트림, 즉 "00100001000011100001"에 있어서, 심볼 값 4는 가용성 0 비트 3에 대응하고 심볼 값 9는 가용성 0 비트 7에 대응한다. 이에, 스트림 내에서 이용 가능한 가용성 0 비트가 14개 뿐이고 4 비트만이 이들을 제공하는데, 즉 인코딩 데이터(E2)에서 이들을 표현하는데 필요하다. 별개의 심볼 값 스트림의 전달이 경우에 따라 회피되고, 즉 레인지 코딩의 확률은 정상 확률로부터 전체 심볼의 동일한 확률로, 또는 동일 실제 데이터 값을 전달하기 위한 이스케이프(레인지) 심볼 전달 후에 가능한 것인 전체 가용성 0 비트 심볼의 동일한 확률로서 샌드위치된 다음, 실제 심볼 값이 전달될 때에, 확률은 다시 정상 확률로 되돌아간다.

[0077] 제2 실시형태는 가용성 0 비트 심볼에 확률 '1'이 할당되고, 그 레인지의 나머지는 확률의 전체 레인지로부터 가용성 1 비트 심볼에 대해 유지된다. "(변형)"으로 표 1에 표시되는 그러한 솔루션도 설명한다. 이제, 앞의 7 비트 예에서는 12 + 2 이스케이프 심볼이 존재하며, 이들에는 128개의 이용 가능한 값 중에서 14개 값이 할당된다. 나머지 114개 값은 가용성 1 비트가 할당되는 6 심볼에 이용된다. 표 1의 최후 열에 나타난 레인지 확률에 대해 6 비트를 이용해서 유사한 예를 제시하는데, 본 예에서는 임계 값이 0.4로 선택된다. 전술한 바와 같이, 이 때 비트 스트림 내의 가용성 0 비트 심볼은 이스케이프 심볼이 되는 것으로서 이될 수 있으며, 이들 심볼 전체에는 표 1에서 괄호 안에 표시하는 바와 같이 그 자신이 확률 값이 할당될 것이다.

- [0078] 확률 값을 전달하는데 필요한 비트의 양에 대해 이제 설명한다. 조합된 (이스케이프) 솔루션은 $7*7 + 20*1 = 69$ 비트를 이용한다. 본 예에서는, 본 개시내용에 따른 제1 실시형태와 비교해서 확률 테이블 전달 시에 $7(76 - 69)$ 비트만 절감하지만 추가 이스케이프 심볼 값 전달에는 $3*4 = 12$ 비트를 이용하기 때문에 양호한 솔루션이 아니다. 이에, 원래의 심볼과 비교해 하나 이상의 이스케이프 심볼의 더 양호한 확률은 데이터 사이즈의 증가를 보상하기에 충분하지 않다. 이 때, 다른 심볼들의 확률은 동일하고 그런데다 이들 심볼을 전달함으로써 이익이 달성되지도 않는다.
- [0079] 이 개선된 방법은 다른 식으로도, 예컨대 대량의 심볼이 이스케이프되지만 이들의 확률이 매우 높지 않은 경우에도 유용하다. 때때로, 더 적은 양의 비트를 원래의 심볼보다는 이스케이프 심볼에 이용하는 것도 실현될 수 있는데, 최소 확률은 조합된 이스케이프 코드에 의해 어느 경우에도 전달되기 때문이다. 변경된 확률을 갖는 제 2 실시형태는 $6*7 + 20*1 = 62$ 비트(7 비트를 이용하는 경우) 또는 $5*6 + 20*1 = 50$ 비트(6 비트를 이용하는 경우)를 이용한다. 확률의 전달은, 경우에 따라, 인코딩된 데이터(E2; 30) 내에 표현되는 가장 빈도가 높은 심볼의 확률을 표시하는 정보를 알거나 전달함으로써, 훨씬 더 좋게 최적화된다.
- [0080] 이제, 가장 높은 확률 값이 64 미만이고, 32 미만인 경우에, 확률은 경우에 따라, 본 예에서는 6 비트(그리고 5 비트)를 이용해 선택적으로 전달되어, 전체 확률 테이블 전달에는 $6*6 + 20*1 = 56$ 비트(7 비트를 이용하는 예)와 $5*5 + 20*1 = 45$ 비트(6 비트를 이용하는 예)만을 필요로 한다. 또한, 이 $20*1$ 가용성 비트 스트림의 전달은 경우에 따라 압축되며, 가용성 비트의 압축 외에도, 빈도 테이블 또는 확률 값을 압축함으로써 본 개시내용의 실시형태를 구현하는 것도 실현 가능하며, 이러한 압축은 하나 이상의 압축 방법을 이용하여 선택적으로 구현된다. 이것은 특히, 가능하게 이용할 수 있는 심볼이 다수이고 상이한 가용성 0 비트 및 가용성 1 비트 확률이 존재할 때에, 즉 다수의 0 또는 다수의 1이 있을 때의 경우, 또는 이들 심볼이 인코딩된 데이터(E2; 30) 내의 비트 스트림의 상이한 부분에 배치되는 경우이다. 선택사항으로서, 예컨대 min 및/또는 max 값을 전달함으로써, 이용 가능한 심볼의 레인지를 별도로 제한하는 경우에, 최초 및/또는 최후 심볼이 스트림 내에서 이용 가능하다고 알려져 있고, 이들 심볼은 가용성 신호, 예를 들어 그것에 대한 가용성 비트를 이용하여 통신될 필요가 없거나 전달될 필요가 없다. 가용성 신호, 예컨대 min+1 내지 max-1의 레인지에 대한 가용성 비트를 전달할 필요만 있다. 이 표 1의 예에서는 심볼 존재가 표현된 경우, 즉 제1 개선된 방법과 비교해서 31 (76 - 45) 비트가 절감된다. 이 절감은 알려진 방법과 비교하면 95 (140 - 45) 비트이다. 이러한 절감은 심볼의 덜 정확한 확률로 손실되지 않으며 그래서 이들 개선된 방법은 본 예에서 데이터를 전달하기에 확실히 바람직한 방법이다.
- [0081] 선택사항으로서, 확률 정보는 인코딩된 데이터(E2; 30) 내에서 훨씬 덜 정확한 확률 값으로 또는 높은 임계 값을 이용하여 전달되며, 그래서 인코딩된 데이터(E2; 30)를 통한 확률 테이블 전달에 있어서 훨씬 더 많은 데이터 사이즈 절감을 달성할 수 있다. 확률 테이블 전달에 있어서의 절감은 엔트로피 코딩에 있어서 통상 일부 손실이 있게 한다. 이러한 손실은 본 개시내용의 실시형태의 구현 시에 최상의 가능한 테이블 전달 방법을 선택하는 경우에 고려되어야 한다.
- [0082] 본 개시내용의 개선된 방법은 일부 경우에 있어서 원래 데이터에도 선택적으로 이용되는데, 데이터 사이즈를 절감하기 위해, 예컨대 데이터베이스 저장시에 캐리어 및 동류에 대하여 이용된다. 앞에서 제시한 예에서는, 원래의 데이터가 단지 150 심볼 그리고 8개의 상이한 값을 포함하고, 그래서 원래의 $150*5 = 750$ 비트는 원래의 데이터에 대해 가용성 비트 정보를 직접 이용함으로써 $150*3 + 20*1 = 470$ 비트로도 전달될 수 있는 것을 이해하게 될 것이다. 엔트로피 코딩 및 테이블 전달에 의한 레인지 코딩은 원래 데이터보다 더 작은 데이터 사이즈를 달성하는 것이 가능하다. 데이터의 엔트로피를 변형하는 상이한 방법들도 레인지 코딩 이전에 이용할 수 있음을 이해하게 될 것이다. 마찬가지로, 이들의 역동작(inverse operation)이 레인지 코딩 후에 행해져야 한다. 엔트로피를 변형하는 이러한 방법은, 예컨대 델타 코딩, ODelta 코딩, 즉 Gurulogic Microsystems Oy에 의해 2013년 3월 1일에 출원한 특허 출원 GB1303661.1에 개시되어 있는 방법(이 특허문헌은 참조로 본 명세서에 포함되며, ODelta 인코딩은 델타 코딩, 오프셋 연산 및 램퍼라운드 연산의 조합을 채택함), 엔트로피 변형기(EM, entropy modifier), Gurulogic Microsystems Oy에 의해 2013년 3월 1일에 출원한 특허 출원 GB1303658.7에 개시되어 있는 방법(이 특허문헌은 참조로 본 명세서에 포함됨), 런 렱스 인코딩(RLE, run length encoding), 및 분할된 런 렱스 인코딩(SRLE, split run length encoding), Gurulogic Microsystems Oy에 의해 2013년 3월 1일에 출원한 특허 출원 GB1303660.3에 개시되어 있는 방법(이 특허문헌은 참조로 본 명세서에 포함됨)을 포함한다.
- [0083] 전술한 바와 같이, 본 개시내용에 따른 개선된 방법은 레인지 코딩을 구현할 때에 바람직하게 경우에 따라 하나 이상의 <EOM> 심볼을 이용하지 않는데, 인코딩될 데이터 양(= 디코딩될 데이터의 양) 또는 인코딩된 데이터의 양(디코딩되어야 하는 양)이 그 대신에 전달될 수 있기 때문이다. 레인지 인코딩 후의 인코딩된 데이터(E2; 3

0)에 존재하는 인코딩된 데이터의 양, 또는 인코딩될 데이터의 양은 레인지 디코딩으로 데이터(D3; 90)를 디코딩할 때에 이용될 수 있다. 또한, 본 개시내용에 따르면, 인코딩된 데이터(E2; 30)를 생성하기 위한 레인지 코딩 후의 데이터 다음에 추가 <EOM> 심볼을 전달할 필요가 없다.

[0084] 선택사항으로서, 확률 값을 예컨대 인코딩된 데이터(E2; 30)를 통해(그러나 이것에 제한되지 않음) 전달할 때에, 확률 값들의 합이 알려져 있고, 예컨대 미리 설정되어 있는 경우에, 또는 하나 이상의 디코더(80)에 이미 전달된 경우에 최후 확률 값을 전달할 필요가 없다. 데이터에 존재하는 심볼, 이스케이프, <EOM> 또는 실제 심볼에 대한 최후 확률은 경우에 따라, 이전에 전달된 값 전부의 확률 합으로부터 감산함으로써 산출된다. 일부 경우에는 그 합이 알려져 있어야 하므로, 확률 값의 디코딩은 하나 이상의 디코더(80)에서 적절하게 실행될 수 있다. 이에 일부 비트는 경우에 따라, 최후 확률 값을 전달하지 않음으로써 미리 데이터 사이즈로부터 절감될 수 있다.

[0085] 본 개시내용의 예들이 레인지 코딩에 대해 제시되지만, 이들은 인코더(20)에서 구현되는 산술 코딩 및 이진 산술 코딩에도 필요한 부분만 수정하여 채택되는 것이 가능하다. 이러한 산술 코딩 및 이진 산술 코딩이 인코더(20)에서 구현될 때에, 이스케이프 코드는 실제 코딩 프로세스에는 사용되지 않지만, 대신에 유사한 추후 아이템에 대한 확률 레인지를 송신하는 방법을 결정하기 위한 것이다. 그러나, 이러한 방법은 본 개시내용의 전술한 개선된 방법보다는 덜 효율적인데, 개선된 방법은 가용성 신호와, 소정의 심볼이 존재하는 것과 그 소정의 심볼의 연관된 확률의 표시를 전달하기 위한 확률 값을 이용하며, 이에 대해서는 공지된 *"Compression ratio is generally lower, than in static <http://compressions.sourceforge.net/Arithmetic.html#static> coding. Although for short input sequences benefit of not sending statistics data can be more significant"*를 참조하면 된다. 이에, 본 개시내용의 개선된 방법은 데이터 인코딩 기술에 있어서 상당한 기술적 진보를 나타내고 있다.

[0086] 본 개시내용에 따른 개선된 인코딩 방법은 예컨대 개인용 컴퓨터, 휴대용 무선 가능 컴퓨터, 패블릿 컴퓨터 (phablet computer), 태블릿 컴퓨터, 스마트폰, 모바일폰("셀폰"), 시청각(audio-visual) 장치, 디지털 카메라, 비디오 카메라, 과학적 기구, 유전자 시퀀싱 장치, 텔레비전, 감시 장치에 있어서(그러나 이들에 한정되지는 않음) 데이터 취급 및 데이터 생성 디바이스에도 이용될 수 있다. 개선된 방법은 경우에 따라서는 디지털 하드웨어, 예컨대 주문형 반도체(ASIC)로 구현된다. 대안으로 또는 부가적으로, 개선된 방법은 비일시적(비일시) 머신 판독 가능한 저장 매체 상에 기록되는 하나 이상의 소프트웨어 제품을 이용하여 구현될 수 있으며, 이 소프트웨어 제품은 컴퓨팅 하드웨어 상에서 실행 가능하다. 선택사항으로서, 하나 이상의 소프트웨어 제품은 "소프트웨어 앱"으로서 통신 네트워크를 통해, 예컨대 무선 가능 휴대용 통신 디바이스, 예를 들어 모바일폰에 다운로드 가능하다. 인코더(20) 및 대응하는 디코더(80)는, 예컨대 휴대용 신호 캡처 장치, 예를 들어 디지털 카메라, 인터넷 가능 멀티미디어 장치, 게임 플레이 장치, 감시 카메라, 환경 모니터링 장치 등(그러나 이들에 한정되지는 않음)에 이용되는 코덱(95)으로서 구현될 수 있다.

[0087] 이제 도 2를 참조하면, 인코더(20)에서 데이터(D1; 10)를 인코딩하여 인코딩된 데이터(E2; 30)를 생성하는 방법의 단계들을 보여주고 있다. 이들 단계는 경우에 따라, 도시하는 바와 같은 순서로 수행되지만, 단계들 중 소정의 것은 선택사항이며, 이들의 실행 순서는 달라질 수도 있다.

[0088] [표 2] 데이터 인코딩 방법의 단계들

단계	동작	세부사항
100	데이터(D1; 10)를 수신하고 그것의 데이터 특성을 분석하여 데이터 내 심볼의 발생 확률을 기술하는 정보를 결정	비선택사항
110	데이터(D1; 10)를 복수의 부분으로 세분하는 것이 유리한지의 여부를 결정, 각각의 부분은 데이터 내의 심볼의 발생 확률과 연관되어 별개로 인코딩됨	선택사항
120	가용성 신호로 데이터 내 심볼의 발생 확률을 기술하는 정보에 의해 조정된 심볼이 인코딩된 데이터(E2; 30) 내에 표현되는지의 여부를 나타내는 데이터를 생성, 가용성 신호는 예를 들어 단일 가용성 비트 정보로서, 예컨대 "1" 또는 "0" 또는 코딩 방법, 또는 인덱스의 수 및 그 값으로서 표현됨	비선택사항
130	경우에 따라 하나 이상의 임계 한도를 적용하여 데이터(D1; 10)에 존재하는 하나 이상의 심볼에 관한 확률 정보의 송신 여부를 결정	선택사항
140	단계 100으로부터의 데이터 내 심볼의 발생 확률을 기술하는 정보 및 단계 120의 표시 데이터와 함께, 데이터(D1; 10)의 심볼을 인코딩하여 인코딩된 데이터(E2; 30)를 생성	비선택사항

[0089]

[0090]

하나 이상의 디코더(80)에서는, 실질적으로 표 2의 역순서가 행해지는데, 즉 단계 100에서 도출된 인코딩된 데이터(E2; 30)로부터 빈도, 확률 또는 레인지 정보가 추출되며, 단계 120으로부터의 데이터는, 인코딩된 데이터(E2; 30) 내의 인코딩된 어떤 심볼이 빈도, 확률 또는 레인지 정보를 이용해서 디코딩되는지, 그리고 어떤 다른 심볼이 그러한 방식으로 디코딩되지 않는지를 결정하는데 채택되며, 그러한 연산의 결과들은 디코딩된 데이터(E3; 90)를 제공하도록 함께 어셈블된다. 선택사항으로서, 하나 이상의 디코더(80)는 데이터(D1)와 데이터(D3)가 서로 상이해지게 트랜스코딩을 수행하도록 구성 가능하다. 이러한 트랜스코딩은, 예를 들어 하나 이상의 디코더(80)가 서로 상이한 경우에 유리하며, 데이터(D3)는 상이한 프레젠테이션 장치에 맞게(예컨대, 픽셀 디스플레이 해상도 및 디스플레이 중횡비) 커스터마이징되도록 트랜스코딩을 필요로 하는 서로 상이한 프레젠테이션 장치 상에 렌더링된다.

[0091]

이제 도 3을 참조하면, 인코딩된 데이터(E2; 30)를 디코딩하여 디코딩된 데이터(D3; 90)를 생성하는 방법의 단계 200 내지 단계 240을 나타내며, 선택사항으로서, 디코딩된 데이터(D3; 90)는 데이터(D1; 10)의 제작성 버전이나, 이것에 한정되지는 않는다. 동작에 있어서, 단계 200에서, 디코더(80)는 가용성 신호, 즉 가용성 비트, 또는 디코딩 방법, 또는 인덱스의 수 및 인덱스 그 자체를 수신한다. 또한, 단계 210에서, 디코더(80)가 (데이터(E2; 30)에 포함되는) 데이터 수에 있어서 송신된 빈도, 확률 또는 레인지 테이블(60)을 수신하여, 디코더(80)는 완전한 빈도 테이블, 확률 테이블 또는 레인지 테이블을 구성한다. "0" 비트 슬롯은 "0" 또는 "1"의 값에 설정되는 것이 유리하고, 그렇게 된다면, "1" 비트 슬롯은 수신된 빈도, 확률 또는 레인지의 값에 설정될 것이다. 빈도 테이블이 수신되면, 그 테이블은 확률 테이블로 변환되어 레인지 코딩에서 이용될 수 있다. 또한 인코더(20)는 레인지 코딩을 이용하여 인코딩을 효율적으로 실행하기 전에 항상 빈도 테이블을 확률 테이블로 변환하는 것을 필요로 하는 것을 이해하게 될 것이다. 단계 220 내지 단계 240에서, 디코더(80)가, 결합된 값의 합이 2의 급수인, 확률 테이블을 구성한 후에, (데이터(E2; 30)에 포함되어 있는) 수신된 변환 데이터(50)를 레인지 디코딩하여 디코딩된 출력 데이터(D3; 90)를 생성하는 것이 가능하다. 단계 200 내지 단계 240은 실질적으로 도 2의 단계 100 내지 단계 140의 역에 해당한다.

[0092]

인코더(20) 및 디코더(80)는 예컨대 무선 통신 디바이스, 오디오 장치, 시청각 장치, 컴퓨팅 장치, 화상 회의 장치, 감시 장치, 과학적 측정 장치, 유전학 시퀀싱 장치, 유전학 시퀀스 분석 장치, 통신 시스템 하드웨어 및/또는 소프트웨어, 통신 네트워크 하드웨어 및/또는 소프트웨어, 그러나 이들에 한정되지 않는 다양한 데이터 취급 장치에 채택될 수 있다.

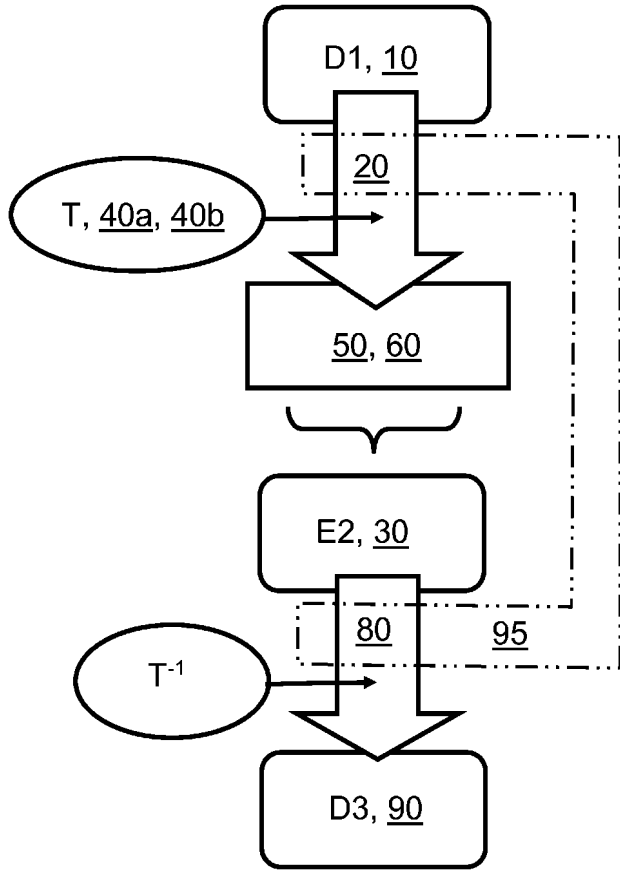
[0093]

앞에서 전술한 본 발명의 실시형태의 변형이, 첨부하는 청구범위에 의해 정해지는 발명의 범주로부터 이탈하지 않고서 가능할 수 있다. 본 발명을 기술하고 청구하는데 사용되는 "포함하는", "통합하는", "이루어지는", "구비하는", "이다"는 비배타적으로 해석되도록, 즉 존재하는 것으로 명시적으로 설명하지 않는 아이템, 구성요소 또는 엘리먼트들에도 허용되게 의도된다. 또한, 단수로 칭해지더라도 복수에 관련되는 것으로 해석되어야 한다. 첨부하는 청구범위 내에서 괄호 안의 숫자는 청구범위의 이해를 돕기 위한 것으로 이들 청구범위에 의해 정해지

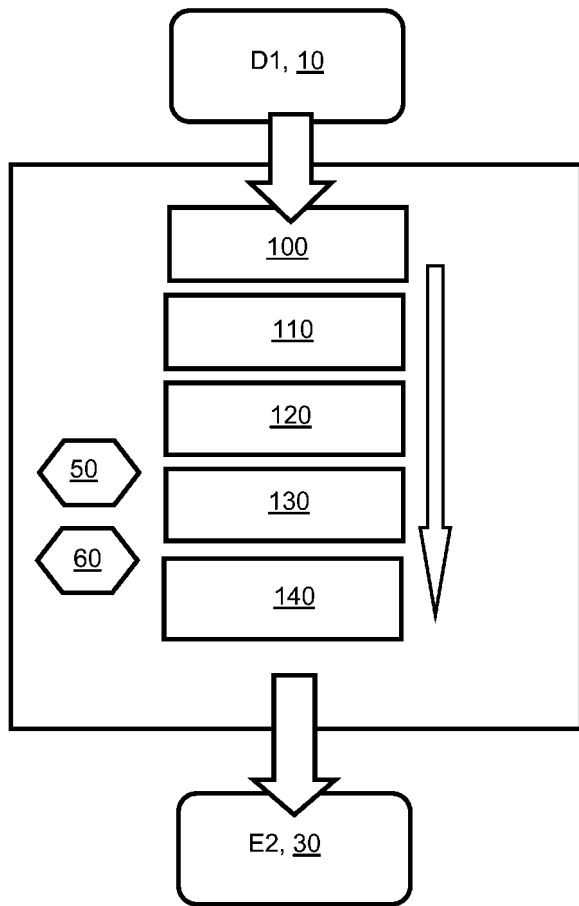
는 발명의 대상을 어떤 식으로도 제한하는 것으로 해석되어서는 안 된다.

도면

도면1



도면2



도면3

