(주) 이디 리서치

한국과학기술원

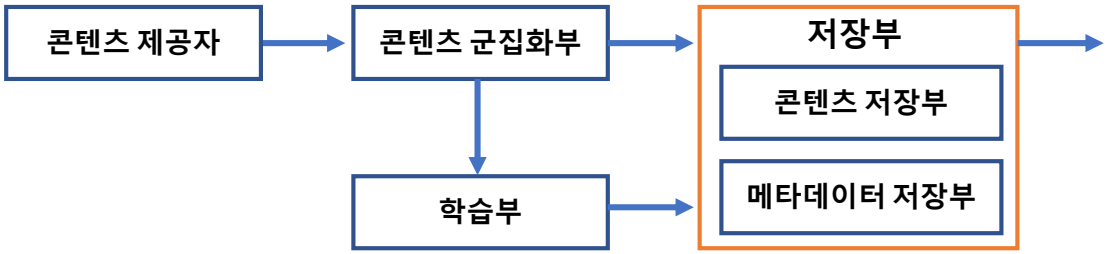
네트워크 및 단말기 리소스를 고려한 딥러닝 기반 콘텐츠 비디오 전송기술

네트워크 및 단말기 리소스를 고려한 딥러닝 기반 콘텐츠 비디오 전송기술

기술개요

- 개별 콘텐츠의 특성에 적합하도록 **사용자 단말에 콘텐츠를 전송**할 수 있는 기술
- 적응형 비디오 스트리밍에 초해상도 심층 신경망을 적용하여, 네트워크가 혼잡해질 때 클라이언트 **비디오 화질이 급격히 저하되는 한계점을 해결**하는 기술
- 콘텐츠 전송 서버는 콘텐츠 군집화부, 학습부 및 저장부를 포함하며, 저장부는 콘텐츠 저장부 및 메타데이터 저장부로 구성되며, 각 구성요소는 하드웨어 칩 또는 소프트웨어로 구현되고 마이크로프로세서가 각 구성요소에 대응하는 소프트웨어의 기능을 실행함
- 콘텐츠 군집화부는 콘텐츠 제공자로부터 제공되는 다수의 콘텐츠를 유사도에 따라 군집화하고, 콘텐츠 제공 서버는 다수의 콘텐츠를 군집화 하기위해 콘텐츠 제공자로부터 제공되는 콘텐츠에 대한 메타데이터 및 이미지의 유사도를 판단하는 기계학습 기반의 모델을 이용하며, 콘텐츠 군집화부는 이미지 분류를 학습한 신경망 모델을 통해 콘텐츠의 유사도를 판단하여 제공되는 다수의 콘텐츠를 군집화할 수 있음

콘텐츠 전송 서버의 구성도



네트워크 및 단말기 리소스를 고려한 딥러닝 기반 콘텐츠 비디오 전송기술

경쟁기술과 비교

기존 비디오 전송 기술 문제점

- 기존의 비디오 전송 시스템은 클라이언트가 네트워크 대역폭을 측정한 후, 이에 적합한 화질의 비디오를 다운로드하는 적응형 스트리밍(Adaptive streaming)을 사용하며, 네트워크가 혼잡할 때 사용자가 보는 비디오 화질이 급격히 저하되는 근본적인 한계점이 존재함
- 기존의 비디오 압축 알고리즘은 적응형 스트리밍에서 매끄러운 화질 전환을 위해 영상 데이터를 2-10초로 설정하는게 일반적이나 비디오에는 중복이 훨씬 더 긴 시간에 걸쳐 존재하여 기존의 압축 방법은 이를 활용하지 못함

딥러닝 기반 콘텐츠 비디오 전송 기술

- 이미지 해상도를 향상시키는 기술로 저해상도 미디어로부터 고해상도의 이미지를 복구하는 기술이며, 콘텐츠 인지 기반 슈퍼 레졸루션을 이용한 콘텐츠 복원 모델은 적응형 스트리밍의 대안이 될 수 있음
- 사용자 단말의 연산 능력을 이용하여 컨텐츠 복원 성능을 구현함으로써, 네트워크 환경에 제한되지 않고 사용자가 원하는 품질의 콘텐츠를 제공함
- 유사한 콘텐츠들을 군집화 하여 컨텐츠들의 유사성을 이용해 콘텐츠 복원 모델을 학습시킴으로써, 사용자에게 높은 품질의 콘텐츠를 제공하면서도 전송에 필요한 네트워크 대역폭은 감소시킬 수 있음

네트워크 및 단말기 리소스를 고려한 딥러닝 기반 콘텐츠 비디오 전송기술

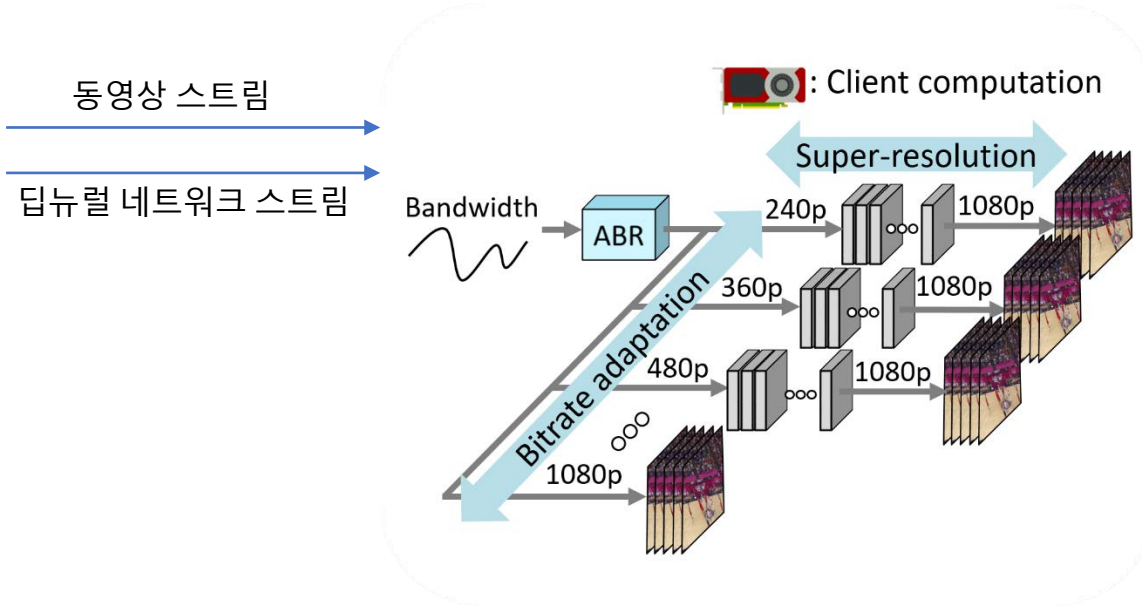
기술특징

- 적응형 스트리밍에 초해상화 심층 신경망을 적용하고, 클라이언트는 네트워크 환경이 좋지 않을 때, 본인의 컴퓨팅 자원을 **심층 신경망을 작동시키는데 활용하여 비디오 화질을 개선**함

- 저해상도의 이미지를 고해상도의 이미지로 복원함

* ABR(available bit rate) : 네트워크에서 혼잡 수준에 따라 대역폭을 조절

음성 파장으로부터의 주파수별 분리과정 도식도



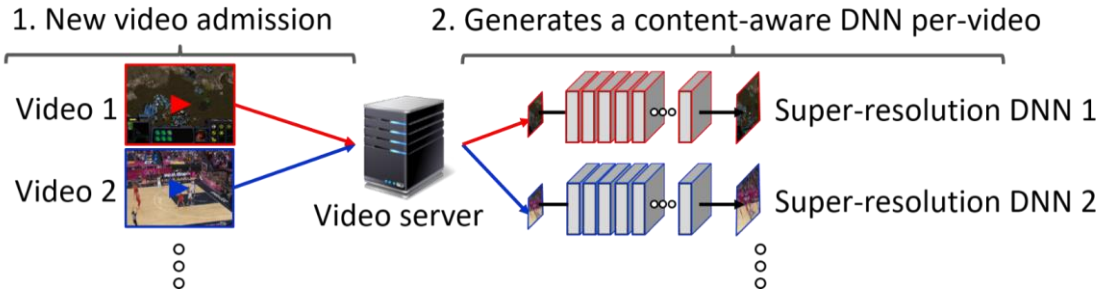
네트워크 및 단말기 리소스를 고려한 딥러닝 기반 콘텐츠 비디오 전송기술

기술특징

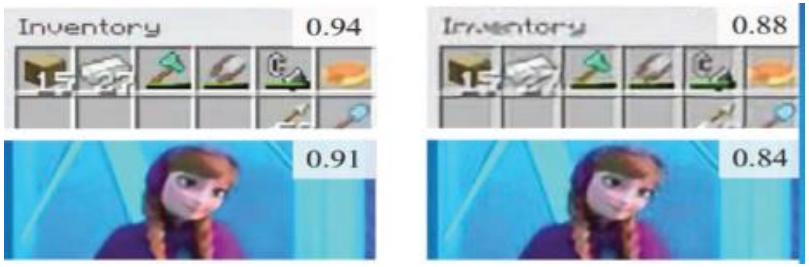
- 비디오 영상 마다 콘텐츠 인지 **초해상화 심층 신경망(super-resolution DNN)**을 통해 **좋은 비디오 화질을 제공**함
- 기존의 딥러닝 시스템에서는 학습 데이터와 테스트 데이터의 불일치로 인해 성능을 보장할 수 없었으나, **콘텐츠 인지 심층 신경망은 학습 데이터와 테스트 데이터가 같아 높은 성능을 안정적으로 제공**함

* 심층신경망(DNN : Deep Neural Network)
- 입력층과 출력층 사이에 여러 개의 은닉층들로 이뤄진 인공신경망
- 분류 및 수치 예측 가능

콘텐츠 인지 초해상화 심층 신경망 제공



콘텐츠 인지 신경망(좌)과 일반적인 신경망(우) (화질표기 : SSIM)



네트워크 및 단말기 리소스를 고려한 딥러닝 기반 콘텐츠 비디오 전송기술

기술특징

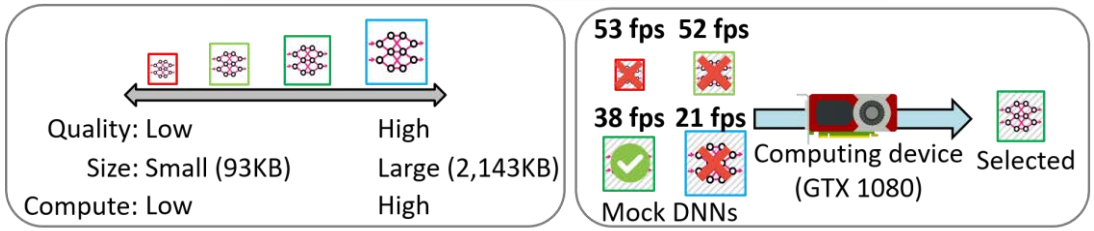
- 기존의 비디오 전송 방식과 다르게 콘텐츠 제공자가 **비디오 뿐만 아니라 이에 적합한 콘텐츠 인지 심층 신경망을 제공**함
- 클라이언트는 비디오와 심층 신경망을 동시에 다운 받기위해서, **강화 학습을 이용**하여 **비디오 화질에 끼치는 영향을 최소화** 하면서 심층 신경망을 빠르게 다운 받는 알고리즘을 개발함
- 비디오와 딥뉴럴 네트워크를 동시에 전송, 강화 학습을 기반으로 네트워크 오버헤드를 최소화함

네트워크 및 단말기 리소스를 고려한 딥러닝 기반 콘텐츠 비디오 전송기술

기술특징

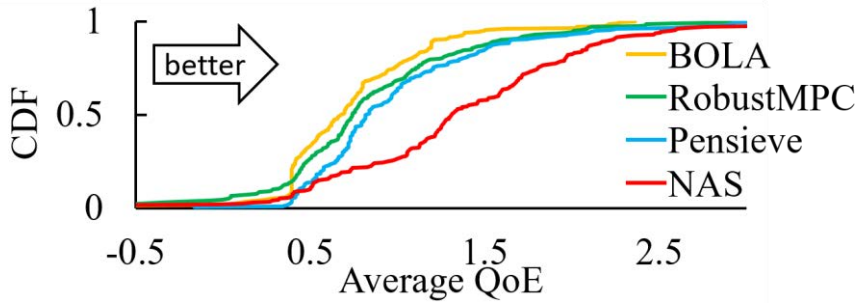
- 클라이언트 컴퓨팅 자원을 사용하나, 클라이언트는 서로 다른 컴퓨팅 자원을 가지고 있음
- 콘텐츠 제공자는 여러 크기의 심층 신경망을 제공하고, 클라이언트는 이 중에 적합한 것을 선택하여 다운로드 하며, **저가부터 고가의 Nvidia GPU에서 작동이 가능**함

서로 다른 컴퓨팅 자원을 가진 클라이언트의 적응 과정



- 최신의 비디오 전송 방식보다 **같은 양의 대역폭을 사용**하면서 사용자의 **체감 품질(Quality of Experience, QoE)**을 DASH표준인 BOLA 대비 92% 이상 **향상시킴**(평균 1.3Mbps인 실 네트워크 환경)
- 같은 QoE를 제공하면서 콘텐츠 제공자가 사용하는 **대역폭을 17.13%만큼 감소**시킴

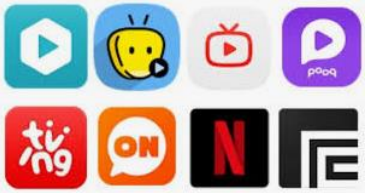
서로 다른 컴퓨팅 자원을 가진 클라이언트의 적응 과정



네트워크 및 단말기 리소스를 고려한 딥러닝 기반 콘텐츠 비디오 전송기술

응용분야

- **OTT(Over the Top) 서비스**(CJ 헬로비전, 방송사, 통신사, SK플래닛, 다음, 네이트, 구글, 넷플릭스, 아마존)
- **소셜미디어**(트위터, 페이스북, 인스타그램, 유튜브)
- **화상회의 시스템**(폴리콤, 로지텍, 새하컴즈, 건우씨엔에스)
- **보안 시스템**(ADT캡스, S1, 시큐아이넷, 라온시큐리티)
- **게임산업**



OTT 서비스

소셜미디어

화상회의시스템

보안시스템

관련특허

No.	특허번호	발명의 명칭
1	10-2018-0003377	신경망을 이용한 컨텐츠 인지 기반 컨텐츠 전송 서버 장치 및 방법
2	US 15924637	Server Apparatus and Method for Content Delivery based on Content-aware Neural Network
3	10-2018-0116404	컨텐츠 인지 신경망을 이용하여 실시간으로 적응형 비디오를 전송하는 방법 및 장치
4	PCT KR2018/011602	Method for Computational Adaptive Video Delivery Utilizing Real-time Content-aware Neural Network

(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2019년12월02일
(11) 등록번호 10-2050780
(24) 등록일자 2019년11월26일

(51) 국제특허분류(Int. Cl.)
G06K 9/62 (2006.01) *G06K 9/46* (2006.01)
G06N 3/08 (2006.01) *H04N 21/231* (2011.01)
(52) CPC특허분류
G06K 9/627 (2013.01)
G06K 9/46 (2013.01)
(21) 출원번호 10-2018-0003377
(22) 출원일자 2018년01월10일
심사청구일자 2018년01월10일
(65) 공개번호 10-2019-0093746
(43) 공개일자 2019년08월12일
(56) 선행기술조사문헌
KR101289758 B1*
KR1020110049570 A*
KR1020170096298 A*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
한국과학기술원
대전광역시 유성구 대학로 291(구성동)
(72) 발명자
한동수
대전광역시 유성구 대학로 291
여현호
대전광역시 유성구 대학로 291
도성현
대전광역시 유성구 대학로 291
(74) 대리인
이철희

전체 청구항 수 : 총 12 항 심사관 : 강현일

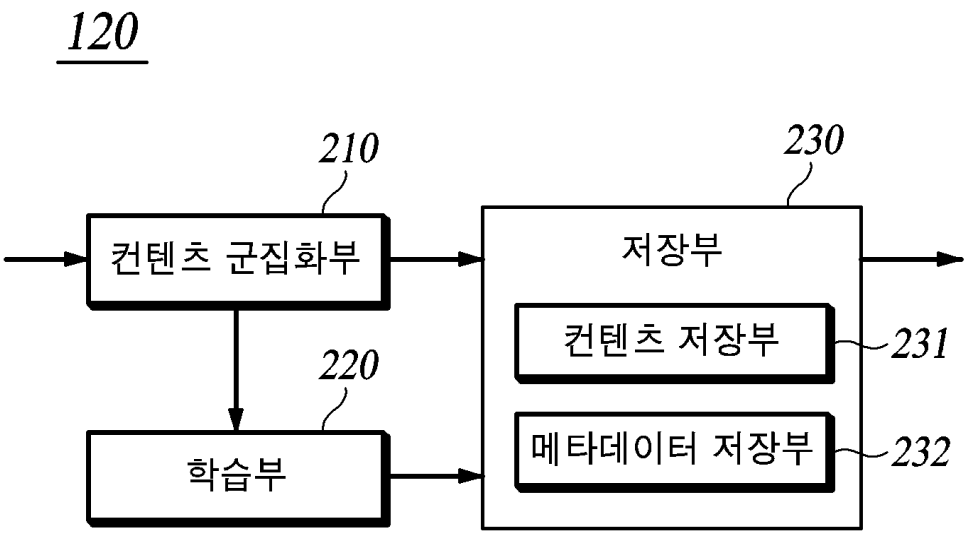
(54) 발명의 명칭 **신경망을 이용한 컨텐츠 인지 기반 컨텐츠 전송 서버 장치 및 방법**

(57) 요 약

신경망을 이용한 컨텐츠 인지 기반 컨텐츠 전송 서버 장치 및 방법을 개시한다.

컨텐츠 제공자로부터 제공되는 다수의 컨텐츠를 유사도에 기반하여 군집화하는 컨텐츠 군집화부; 군집화 결과에 따라 각 군집에 포함된 컨텐츠를 이용하여 군집별 컨텐츠 복원 모델을 학습시키는 학습부; 상기 다수의 컨텐츠 및 상기 군집별 컨텐츠 복원 모델을 저장하는 저장부; 및 사용자로부터 요청된 컨텐츠 및 상기 요청된 컨텐츠가 포함된 군집에 대응하는 컨텐츠 복원 모델을 사용자 단말에 전송하는 전송부를 포함하는 컨텐츠 전송 서버 장치를 제공한다.

대 표 도 - 도2



(52) CPC특허분류
G06N 3/08 (2013.01)
H04N 21/23109 (2013.01)

명 세 서

청구범위

청구항 1

컨텐츠 제공자로부터 제공되는 다수의 컨텐츠를 상기 다수의 컨텐츠의 메타데이터를 이용하거나 또는 이미지 분류를 학습한 기계학습 모델을 이용함으로써 유사도에 기반하여 군집화하는 컨텐츠 군집화부;

군집화 결과에 따라 각 군집에 포함된 컨텐츠를 이용하여 군집별 컨텐츠 복원 모델을 학습시키고, 상기 군집별 컨텐츠 복원 모델과 상기 각 군집 내 컨텐츠 간의 연관 정보를 매니페스트 파일에 기록하는 학습부;

상기 다수의 컨텐츠 및 상기 군집별 컨텐츠 복원 모델을 저장하는 컨텐츠 저장부; 및

사용자 단말로부터 요청된 컨텐츠 및 상기 요청된 컨텐츠가 포함된 군집에 대응하는 컨텐츠 복원 모델을 상기 매니페스트 파일을 통해 상기 사용자 단말에 전송하는 전송부

를 포함하는 컨텐츠 전송 서버 장치.

청구항 2

제 1항에 있어서,

상기 컨텐츠 군집화부는,

인공신경망 기반 이미지 분류 모델을 통해 상기 유사도를 판단하여 군집화하는 것을 특징으로 하는, 컨텐츠 전송 서버 장치.

청구항 3

제 2항에 있어서,

상기 컨텐츠 군집화부는,

상기 컨텐츠 제공자로부터 새로운 컨텐츠를 제공받으면, 상기 이미지 분류 모델을 통해 상기 새로운 컨텐츠와 기존의 군집들 간의 유사도를 판단하고, 상기 새로운 컨텐츠와 중복성이 가장 높은 군집에 상기 새로운 컨텐츠를 군집화하고, 상기 유사도가 일정 수준 이하인 경우 새로운 컨텐츠를 새로운 군집에 군집화하는 것을 특징으로 하는, 컨텐츠 전송 서버 장치.

청구항 4

제 1항에 있어서,

상기 학습부는,

상기 다수의 컨텐츠를 압축시켜 대체 컨텐츠를 생성하고, 상기 대체 컨텐츠로부터 컨텐츠 원본을 출력하도록 상기 컨텐츠 복원 모델을 학습시키는 것을 특징으로 하는, 컨텐츠 전송 서버 장치.

청구항 5

제 4항에 있어서,

상기 학습부는,

상기 군집에 포함된 컨텐츠의 저해상도 영상을 생성하고, 상기 컨텐츠 복원 모델에 상기 저해상도 영상으로부터 컨텐츠 원본 영상을 출력하도록 학습시키는 것을 특징으로 하는, 컨텐츠 전송 서버 장치.

청구항 6

제 4항에 있어서,

상기 학습부는,

상기 군집에 포함된 컨텐츠의 휘도 영상 또는 윤곽선 영상을 추출하여, 상기 컨텐츠 복원 모델에 상기 휘도 영상 또는 윤곽선 영상으로부터 컨텐츠 원본을 출력하도록 학습시키는 것을 특징으로 하는, 컨텐츠 전송 서버 장치.

청구항 7

제 4항에 있어서,

상기 학습부는,

상기 컨텐츠 복원 모델에 상기 군집에 포함된 컨텐츠에 대해 프레임 보간을 학습시키는 것을 특징으로 하는, 컨텐츠 전송 서버 장치.

청구항 8

제 4항에 있어서,

상기 전송부는,

상기 사용자 단말의 네트워트 연결 상태에 따라 상기 요청된 컨텐츠를 상기 대체 컨텐츠로 대체하여 전송하는 것을 특징으로 하는, 컨텐츠 전송 서버 장치.

청구항 9

컨텐츠 제공 시스템의 컨텐츠 전송 서버에서 사용자 단말에 컨텐츠를 전송하기 위한 방법에 있어서,

컨텐츠 제공자로부터 제공되는 다수의 컨텐츠를 상기 다수의 컨텐츠의 메타데이터를 이용하거나 또는 이미지 분류를 학습한 기계학습 모델을 이용함으로써 유사도에 기반하여 군집화하는 과정,

군집화 결과에 따라 각 군집에 포함된 컨텐츠를 이용하여 군집별 컨텐츠 복원 모델을 학습시키는 과정,

상기 군집별 컨텐츠 복원 모델과 각 군집 내 컨텐츠 간의 연관 정보를 매니페스트 파일에 기록하는 과정, 및

사용자로부터 요청된 컨텐츠 및 상기 요청된 컨텐츠가 포함된 군집에 대응하는 컨텐츠 복원 모델을 사용자 단말에 전송하는 과정

를 포함하는 컨텐츠 전송 방법.

청구항 10

제 9항에 있어서,

상기 군집화하는 과정은,

인공신경망 기반 이미지 분류 모델을 통해 상기 유사도를 판단하여 군집화하는 것을 특징으로 하는, 컨텐츠 전송 방법.

청구항 11

제 9항에 있어서,

상기 학습시키는 과정은,

상기 다수의 컨텐츠를 압축시켜 대체 컨텐츠를 생성하고, 상기 대체 컨텐츠로부터 컨텐츠 원본을 출력하도록 상기 컨텐츠 복원 모델을 학습시키는 것을 특징으로 하는, 컨텐츠 전송 방법.

청구항 12

제 11항에 있어서,

상기 전송하는 과정은,

상기 사용자 단말의 네트워트 연결 상태에 따라 상기 요청된 컨텐츠를 상기 대체 컨텐츠로 대체하여 전송하는 것을 특징으로 하는, 컨텐츠 전송 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 신경망을 이용한 컨텐츠 인지 기반 컨텐츠 전송 서버 장치 및 방법에 관한 것이다.

배 경 기 술

[0002] 이 부분에 기술된 내용은 단순히 본 실시예에 대한 배경 정보를 제공할 뿐 종래기술을 구성하는 것은 아니다.

[0003] 인터넷을 통한 비디오 전송이 급격하게 증가하고 있으며, 증강 현실이나 가상 현실을 제공하는 스트리밍 서비스가 등장함에 따라 인터넷 비디오가 전체 인터넷 트래픽에서 차지하는 비중이 높아지고 있다.

[0004] 인터넷 비디오 전송 기술은 컨텐츠 전송 네트워크(content delivery networks, CDNs)에서부터 HTTP 적응적 스트리밍(adaptive streaming) 및 QoE(Quality of Experience)를 위한 데이터를 이용한 최적화에 이르기까지, 한정된 네트워크 자원 내에서 사용자에게 최상의 화질을 제공할 수 있도록 하기 위해 다양한 기술이 제안된 바 있다.

[0005] 그러나, 현재의 비디오 전송 기술의 경우, 비디오를 단지 비트스트림으로만 취급함으로써 컨텐츠의 종류와 상관없이 동일한 기술을 적용하고 있으며, 비디오 인코딩도 단순히 짧은 시간 스케일(프레임 내)에서 발생하는 공간 및 시간 중복성을 이용한 신호 처리 기술(이산 코사인 변환 및 프레임 간 예측)을 사용하는 것에 그치고 있다.

발명의 내용

해결하려는 과제

[0006] 본 발명은, 개별 컨텐츠의 특성에 적합하도록 사용자 단말에 컨텐츠를 전송할 수 있는 컨텐츠 인지 기반 컨텐츠 전송 서버 장치 및 그 전송 방법을 제공하는 데 주된 목적이 있다.

과제의 해결 수단

[0007] 본 발명의 일 실시예에 의하면, 컨텐츠 제공자로부터 제공되는 다수의 컨텐츠를 유사도에 기반하여 군집화하는 컨텐츠 군집화부; 군집화 결과에 따라 각 군집에 포함된 컨텐츠를 이용하여 군집별 컨텐츠 복원 모델을 학습시키는 학습부; 상기 다수의 컨텐츠 및 상기 군집별 컨텐츠 복원 모델을 저장하는 저장부; 및 사용자로부터 요청된 컨텐츠 및 상기 요청된 컨텐츠가 포함된 군집에 대응하는 컨텐츠 복원 모델을 사용자 단말에 전송하는 전송부를 포함하는 컨텐츠 전송 서버 장치를 제공한다.

[0008] 상기 장치의 실시예들은 다음의 특징들을 하나 이상 더 포함할 수 있다.

[0009] 상기 컨텐츠 군집화부는, 인공신경망 기반 이미지 분류 모델을 통해 상기 유사도를 판단하여 군집화할 수 있다.

[0010] 상기 학습부는, 상기 다수의 컨텐츠를 압축시켜 대체 컨텐츠를 생성하고, 상기 대체 컨텐츠로부터 컨텐츠 원본을 출력하도록 상기 컨텐츠 복원 모델을 학습시킬수 있다.

[0011] 상기 전송부는, 상기 사용자 단말의 네트워트 연결 상태에 따라 상기 요청된 컨텐츠를 상기 대체 컨텐츠로 대체하여 전송할 수 있다.

[0012] 본 발명의 일 실시예에 의하면, 컨텐츠 제공 시스템의 컨텐츠 전송 서버에서 사용자 단말에 컨텐츠를 전송하기 위한 방법에 있어서, 컨텐츠 제공자로부터 제공되는 다수의 컨텐츠를 유사도에 기반하여 군집화하는 과정, 군집화 결과에 따라 각 군집에 포함된 컨텐츠를 이용하여 군집별 컨텐츠 복원 모델을 학습시키는 과정, 사용자로부터 요청된 컨텐츠 및 상기 요청된 컨텐츠가 포함된 군집에 대응하는 컨텐츠 복원 모델을 사용자 단말에 전송하는 과정을 포함하는 컨텐츠 전송 방법을 제공한다.

발명의 효과

[0013] 이상에서 설명한 바와 같이 본 실시예에 의하면, 유사한 컨텐츠들을 군집화하여 컨텐츠들의 유사성을 이용해 컨텐츠 복원 모델을 학습시킴으로써, 사용자에게 높은 품질의 컨텐츠를 제공하면서도 전송에 필요한 네트워크 대역폭은 감소시킬 수 있는 효과가 있다.

[0014] 또한, 본 실시예에 의하면, 사용자 단말에 이러한 컨텐츠 복원 모델을 제공하고 사용자 단말의 연산 능력을 이

용하여 컨텐츠 복원 성능을 구현함으로써, 네트워크 환경에 제한되지 않고 사용자가 원하는 품질의 컨텐츠를 제공할 수 있는 효과가 있다.

도면의 간단한 설명

[0015] 도 1은 본 발명의 일 실시예에 따른 컨텐츠 제공 서비스 시스템을 개략적으로 나타낸 도면이다.

도 2는 본 발명의 일 실시예에 따른 컨텐츠 전송 서버의 구성을 개략적으로 도시한 도면이다.

도 3은 본 발명의 일 실시예에 따른 컨텐츠 전송 서버에서의 컨텐츠 복원 모델 학습 방법을 설명하기 위한 도면이다.

도 4는 본 발명의 일 실시예에 따른 군집별 컨텐츠 복원 모델을 이용한 컨텐츠 전송 방법을 설명하기 위한 도면이다.

도 5는 본 발명의 일 실시예에 따른 컨텐츠 전송 서버의 컨텐츠 전송 방법을 도시한 흐름도이다.

도 6은 본 발명의 일 실시예에 따른 방법 및 종래 기술에 의한 비디오 화질 개선 결과를 비교하여 도시한 도면이다.

도 7 및 도 8은 본 발명의 일 실시예에 따른 방법 및 종래 기술에 의한 비디오 전송 성능을 비교하여 도시한 도면이다.

도 9는 종래 기술에 의한 비디오 인코딩 및 디코딩 결과를 비교하여 도시한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0016] 이하, 본 발명의 일부 실시예들을 예시적인 도면을 통해 상세하게 설명한다. 각 도면의 구성요소들에 참조부호를 부가함에 있어서, 동일한 구성요소들에 대해서는 비록 다른 도면상에 표시되더라도 가능한 한 동일한 부호를 가지도록 하고 있음에 유의해야 한다. 또한, 본 발명을 설명함에 있어, 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 발명의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략한다.

[0017] 또한, 본 발명의 구성 요소를 설명하는 데 있어서, 제 1, 제 2, A, B, (a), (b) 등의 용어를 사용할 수 있다. 이러한 용어는 그 구성 요소를 다른 구성 요소와 구별하기 위한 것일 뿐, 그 용어에 의해 해당 구성 요소의 본질이나 차례 또는 순서 등이 한정되지 않는다. 명세서 전체에서, 어떤 부분이 어떤 구성요소를 '포함', '구비'한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 또한, 명세서에 기재된 '~부', '모듈' 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.

[0018] 아래에서는 첨부한 도면을 참고하여 본 발명의 실시예에 대하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략한다.

[0019] 본 명세서에서 컨텐츠는 방송 컨텐츠, 오디오 또는 비디오 컨텐츠 등 다양한 멀티미디어 컨텐츠를 포함한다. 컨텐츠는 소정의 미디어 컨텐츠이면 그 내용이나 형식에는 제한이 없다. 따라서, 컨텐츠에는 컨텐츠 제작 사업자가 제작한 컨텐츠 뿐만 아니라 개인이 만들어 배포하는 UCC(User Creative Contents)가 포함된다.

[0020] 본 명세서에서 메타데이터는 데이터에 관한 데이터(data about data)로서 컨텐츠에 대한 부가 정보를 의미한다. 메타데이터는 매니페스트 파일(manifest file)에 기록되어 사용자 단말에 전송되며, 사용자 단말은 메타데이터를 이용하여 원하는 컨텐츠를 요청하는 등 컨텐츠 제공 서비스를 받을 수 있다.

[0021] 도 1은 본 발명의 일 실시예에 따른 컨텐츠 제공 서비스 시스템을 개략적으로 나타낸 도면이다.

[0022] 도 1을 참조하면, 본 발명의 일 실시예에 따른 컨텐츠 제공 서비스 시스템은 컨텐츠 제공자(110), 컨텐츠 제공 서버(120) 및 사용자 단말(130)을 포함하여 구성된다.

[0023] 컨텐츠 제공자(110)는 컨텐츠 제공 서비스를 위한 컨텐츠를 생성하여 제공한다. 컨텐츠 제공자(110)는 서비스 목적에 따라 다양한 컨텐츠를 생성할 수 있다.

[0024] 예컨대, VOD(Video On Demand) 서비스를 제공하는 컨텐츠 제공자(110)는 비디오/오디오 형태의 컨텐츠를 생성하

여 제공할 수 있고, 개인 방송을 제공하는 컨텐츠 제공자(110)는 라이브 스트림 형태의 컨텐츠를 생성하여 제공할 수 있다. 또한, 다시점 영상 공유 서비스를 제공하는 컨텐츠 제공자(110)는 영상의 깊이 정보를 포함하는 컨텐츠를 생성하여 제공할 수 있다. AR(Augmented Reality) 서비스를 제공하는 컨텐츠 제공자(110)는 360 카메라를 통해 스티칭된 컨텐츠를 생성하여 제공할 수 있다.

[0025] 컨텐츠 제공 서버(120)는 컨텐츠 제공자(110)에 의해 제공된 컨텐츠를 서비스 및 컨텐츠의 종류에 적합한 전송 프로토콜을 통해 사용자 단말(130)에 전송한다. 컨텐츠 제공 서버(120)는 다양한 컨텐츠와 서비스를 사용자 단말에 따라 적합한 제공 방식을 선정하고, 선정한 방식을 통해 컨텐츠 제공자(110)로부터 제공받은 컨텐츠를 사용자 단말(130)에 전송한다. 컨텐츠 제공 서버(120)는 컨텐츠 제공자(110)로부터 생성되어 제공되는 새로운 컨텐츠를 지속적으로 제공받아 사용자 단말(130)에 새로운 컨텐츠를 전송할 수 있다.

[0026] 예컨대, VOD 서비스를 제공하기 위해 RTSP(Real Time Streaming Protocol)을 이용하여 컨텐츠 제공자(110)로부터 제공받은 VOD 컨텐츠를 전송할 수 있고, 개인 방송 컨텐츠를 제공하는 컨텐츠 제공자(110)로부터 제공받은 개인 방송 컨텐츠를 전송하기 위해 HLS(HTTP Live Streaming) 프로토콜 또는 MPEG-TS(Moving Picture Experts Group - Transport Stream) 프로토콜을 통해 전송할 수 있다.

[0027] 또한, 컨텐츠 제공 서버(120)는 컨텐츠 제공자(110)로부터 제공받은 다수의 컨텐츠를 유사도에 기반하여 군집화하고, 각 군집에 대해 서로 다른 신경망 기반의 학습 모델을 통해 각 군집별로 군집 내 컨텐츠들을 이용한 컨텐츠 복원 모델을 생성한다. 컨텐츠 복원 모델은, 저화질 또는 압축된 컨텐츠로부터 고화질 또는 고품질의 컨텐츠를 출력하도록 학습된 신경망 기반의 모델로, 저해상도의 영상으로부터 고해상도의 영상을 출력하도록 학습하거나, 흑백 또는 윤곽선으로 이루어진 영상으로부터 원본 영상을 복원하여 출력하도록 학습하거나, 영상의 프레임 간 보간을 학습하여 압축된 영상으로부터 보간된 프레임이 포함된 고품질의 영상을 출력할 수 있다.

[0028] 컨텐츠 제공 서버(120)는 다수의 컨텐츠를 군집화하기 위해 컨텐츠 제공자(110)로부터 제공되는 컨텐츠에 대한 메타 데이터를 이용하거나, 이미지의 유사도를 판단하는 기계학습 기반의 모델을 이용할 수 있다. 구체적으로, 이미지 분류(classification)를 학습한 신경망 모델을 통해 컨텐츠의 유사도를 판단하여 제공되는 다수의 컨텐츠를 군집화할 수 있다.

[0029] 컨텐츠 제공 서버(120)는 각 군집에 포함되어 있는 유사도가 높은 컨텐츠들을 이용하여, 저화질 또는 압축된 형태의 컨텐츠로부터 고화질 또는 고품질의 컨텐츠를 출력하도록 컨텐츠 복원 모델을 학습시킬 수 있다. 군집 내에 포함되어 있는 컨텐츠들은 유사도가 높은 컨텐츠로 서로 공유하고 있는 중복된 정보가 많기 때문에, 동일한 컨텐츠 복원 모델을 이용할 수 있다.

[0030] 사용자 단말(130)은 컨텐츠 제공자(110)에 의해 제공되는 컨텐츠와 컨텐츠 복원 모델을 컨텐츠 전송 서버(120)를 통해 수신한다. 사용자 단말(130)은 스마트폰, 태블릿 PC, 노트북 또는 데스크탑 등일 수 있으며, 사용자 단말(130)의 종류에 따라 적합한 컨텐츠 제공 모델이 제공될 수 있다.

[0031] 컨텐츠 복원 모델은 사용자 단말(130)이 수신하는 매니페스트 파일(manifest file)의 메타 데이터에 포함될 수 있다. 사용자 단말(130)은 매니 페스트 파일을 통해 컨텐츠에 접근 또는 컨텐츠를 요청하는 것 이외에, 해당 컨텐츠에 적합한 컨텐츠 복원 모델을 수신할 수 있다.

[0032] 사용자 단말(130)은 수신한 컨텐츠 및 컨텐츠 복원 모델을 이용하여, 원하는 품질의 컨텐츠를 생성할 수 있다. 즉, 제공받은 컨텐츠 복원 모델을 이용하여 사용자 단말(130) 내에서의 연산을 통해 고품질의 컨텐츠를 생성할 수 있기 때문에 사용자는 네트워크 환경이 좋지 않더라도 고품질의 컨텐츠를 제공받을 수 있게 된다.

[0033] 사용자 단말(130)은 각 컨텐츠의 서비스 타입이 지원하는 전송 프로토콜을 통해 각 컨텐츠 및 컨텐츠 복원 모델을 수신한다. 예컨대, VOD 서비스를 제공받기 위하여 RTSP 및 HLS 프로토콜을 통해 컨텐츠를 수신하거나, 개인 방송 서비스를 제공받기 위해 라이브 스트림 형태의 컨텐츠를 HLS 프로토콜 또는 MPEG-TS 프로토콜을 통해 수신할 수 있다.

[0034] 도 2는 본 발명의 일 실시예에 따른 컨텐츠 전송 서버의 구성을 개략적으로 도시한 도면이다.

[0035] 도 2를 참조하면, 본 실시예의 컨텐츠 전송 서버(120)는 컨텐츠 군집화부(210), 학습부(220) 및 저장부(230)를 포함하며, 저장부(230)는 컨텐츠 저장부(231) 및 메타데이터 저장부(232)를 포함하여 구성된다. 도 2에 도시한 각 구성요소는 하드웨어 칩으로 구현될 수 있으며, 또는 소프트웨어로 구현되고 마이크로프로세서가 각 구성요소에 대응하는 소프트웨어의 기능을 실행하도록 구현될 수도 있다.

[0036] 컨텐츠 군집화부(210)는 컨텐츠 제공자로부터 제공되는 다수의 컨텐츠를 유사도에 기반하여 군집화한다. 컨텐츠

제공 서버(120)는 다수의 컨텐츠를 군집화하기 위해 컨텐츠 제공자(110)로부터 제공되는 컨텐츠에 대한 메타 데이터를 이용하거나, 이미지의 유사도를 판단하는 기계학습 기반의 모델을 이용할 수 있다. 구체적으로, 컨텐츠 군집화부(210)는 이미지 분류(classification)를 학습한 신경망 모델을 통해 컨텐츠의 유사도를 판단하여 제공되는 다수의 컨텐츠를 군집화할 수 있다.

[0037] 예컨대, 컨텐츠 제공자(110)로부터 다양한 스포츠 관련 영상을 제공받은 경우에, 컨텐츠 제공자(110)가 각 컨텐츠에 대하여 생성한 메타 데이터를 분석하여 농구 경기, 축구 경기, 야구 경기 등으로 유사한 컨텐츠들을 군집화할 수 있다. 또는 기계학습을 통해 이미지 분류를 학습한 이미지 분류 모델을 이용하여 제공받은 컨텐츠의 프레임을 분석하여 유사도에 따라 농구 경기, 축구 경기, 야구 경기 등으로 분류하여 군집화할 수도 있다.

[0038] 컨텐츠 군집화부(210)는 컨텐츠 제공자(110)로부터 새로운 컨텐츠를 제공받으면, 이미지 분류 모델을 통해 새로운 컨텐츠와 기존의 군집들 간의 유사도를 판단한다. 컨텐츠 군집화부(210)는 새로운 컨텐츠가 기존에 존재하는 군집과 중복성이 높다면 중복성이 가장 높은 군집으로 새로운 컨텐츠를 분류하고, 중복성이 일정 수준 이하인 경우 유사도가 높은 군집이 없으므로 새로운 컨텐츠를 새로운 군집으로 분류할 수 있다.

[0039] 학습부(220)는 군집화 결과에 따라 각 군집에 포함된 컨텐츠를 이용하여 군집별 컨텐츠 복원 모델을 학습시킨다. 컨텐츠 복원 모델은, 저화질 또는 압축된 컨텐츠로부터 고화질 또는 고품질의 컨텐츠를 출력하도록 학습된 신경망 기반의 모델로, 저해상도의 영상으로부터 고해상도의 영상을 출력하도록 학습하거나, 흑백 또는 윤곽선으로 이루어진 영상으로부터 원본 영상을 복원하여 출력하도록 학습하거나, 압축된 영상으로부터 보간된 프레임이 포함된 고품질의 영상을 출력하도록 영상의 프레임간 보간을 학습할 수 있다.

[0040] 학습부(220)는 각 군집에 포함되어 있는 유사도가 높은 컨텐츠들을 이용하여, 저화질 또는 압축된 형태의 컨텐츠로부터 고화질 또는 고품질의 컨텐츠를 출력하도록 컨텐츠 복원 모델을 학습시킬 수 있다. 군집 내에 포함되어 있는 컨텐츠들은 유사도가 높은 컨텐츠로 서로 공유하고 있는 중복된 정보가 많기 때문에, 동일한 컨텐츠 복원 모델을 이용할 수 있다.

[0041] 예컨대, 축구 경기의 경우, 축구장과 같은 배경이나 플레이어 등은 영상 전체를 통하여 반복하여 나타나고, 여러 축구 경기 영상에서 동일한 배경 및 플레이어가 나타날 수 있다. 또한, 동일한 축구 경기장이나 동일한 플레이어가 아니라 하더라도 경기장의 잔디 색이나 축구장 전경에 나타나는 관중 모습 등과 같이 축구 경기 영상이 공유하는 중복된 정보가 많다. 따라서, 이렇게 유사한 컨텐츠를 군집화하여 컨텐츠 복원 모델을 학습시키는 경우에, 다양한 경기 영상에 적용되는 뛰어난 컨텐츠 복원 성능을 보여줄 수 있으며, 군집에 포함된 모든 경기 영상이 해당 컨텐츠 복원 모델을 공유할 수 있다.

[0042] 학습부(220)는 CNN(convolutional neural network)과 같이 이미지 처리에 적합한 신경망을 이용하여 각 군집 내 컨텐츠들에 대한 컨텐츠 복원 모델을 생성한다. 학습부(220)는 군집별 컨텐츠 복원 모델과 군집 내 컨텐츠를 연관시키고 해당 연관 정보를 메타데이터로서 매니페스트 파일에 기록한다. 학습부(220)는 컨텐츠 복원 모델의 학습 내용에 따라 컨텐츠의 저화질 또는 압축된 형태의 컨텐츠(이하 '대체 컨텐츠'라 함)를 생성하고, 대체 컨텐츠를 매니페스트 파일에 기록한다. 즉, 매니페스트 파일에는 컨텐츠 복원 모델과 대체 컨텐츠가 포함되어 있을 수 있다,

[0043] 컨텐츠를 군집화하지 않고, 모든 컨텐츠에 적용되는 컨텐츠 복원 모델을 학습시키는 경우에, 연산을 위한 비용이 증가하며 모든 컨텐츠에 대해 고른 컨텐츠 복원 성능을 보여줄 수 없다. 따라서, 본 실시예에서는 컨텐츠 제공자로부터 제공되는 다수의 컨텐츠들을 유사도가 높은 컨텐츠들끼리 군집화하고, 각 군집에 대해 개별적으로 군집 내에 포함된 컨텐츠들을 이용하여 학습시킨 컨텐츠 복원 모델을 생성함으로써, 연산 비용을 감소시킬 수 있을 뿐 아니라 뛰어난 컨텐츠 복원 성능을 보여줄 수 있다.

[0044] 저장부(230)는 다수의 컨텐츠 및 군집별 컨텐츠 복원 모델을 저장한다. 컨텐츠 저장부(231)는 컨텐츠 제공자(110)로부터 수신한 컨텐츠를 저장한다. 저장된 컨텐츠는 컨텐츠 원본 파일일 수도 있고, 경우에 따라 저화질 또는 압축된 형태의 컨텐츠(대체 컨텐츠)일 수 있다. 메타데이터 저장부(232)는 컨텐츠에 관련된 메타데이터를 저장한다.

[0045] 도면에 도시되지는 않았지만, 컨텐츠 전송 서버(120)는 사용자로부터 요청된 컨텐츠 및 요청된 컨텐츠가 포함된 군집에 대응하는 컨텐츠 복원 모델을 사용자 단말에 전송하는 전송부를 더 포함할 수 있다. 전송부(미도시)에서는, 사용자 단말(130)과의 네트워크 연결 상태를 고려하여 요청된 컨텐츠의 원본 또는 대체 컨텐츠를 컨텐츠 복원 모델과 함께 전송할 수 있다.

[0046] 도 3은 본 발명의 일 실시예에 따른 컨텐츠 전송 서버에서의 컨텐츠 복원 모델 학습 방법을 설명하기 위한 도면

이다.

[0047] 도 4는 본 발명의 일 실시예에 따른 군집별 컨텐츠 복원 모델을 이용한 컨텐츠 전송 방법을 설명하기 위한 도면이다.

[0048] 컨텐츠 전송 서버(120)에서는 컨텐츠 제공자로부터 제공되는 다수의 컨텐츠를 유사도에 기반하여 군집화한다. 다수의 컨텐츠를 군집화하기 위해 컨텐츠 제공자로부터 컨텐츠와 함께 제공되는 컨텐츠에 대한 메타 데이터를 이용하거나, 이미지의 유사도를 판단하는 기계학습 기반의 모델을 이용할 수 있다.

[0049] 컨텐츠 복원 모델은 각 군집에 대해 개별적으로 생성된다. 군집은 서로 유사한 컨텐츠를 포함하고 있기 때문에, 동일한 컨텐츠 복원 모델을 공유할 수 있다. 컨텐츠 복원 모델은, 저화질 또는 압축된 컨텐츠로부터 고화질 또는 고품질의 컨텐츠를 출력하도록 학습된 신경망 기반의 모델로, 저해상도의 영상으로부터 고해상도의 영상을 출력하도록 학습하거나, 흑백 또는 윤곽선으로 이루어진 영상으로부터 원본 영상을 복원하여 출력하도록 학습하거나, 압축된 영상으로부터 보간된 프레임이 포함된 고품질의 영상을 출력하도록 영상의 프레임간 보간을 학습할 수 있다.

[0050] 예컨대, 군집 A는 축구 경기, 군집 B는 농구 경기, 군집 C는 야구 경기인 경우에, 학습 모델 A는 군집 A에 포함된 컨텐츠인 축구 경기를 학습한 컨텐츠 복원 모델이고, 학습 모델 B는 농구 경기를 학습한 컨텐츠 복원 모델이고, 학습 모델 C는 야구 경기를 학습한 컨텐츠 복원 모델이다.

[0051] 도 4를 참조하면, 컨텐츠 전송 서버(120)는 사용자 단말에 컨텐츠와 함께 컨텐츠 복원 모델을 전송한다. 사용자 단말1(131) 및 사용자 단말3(133)과의 네트워크 연결이 좋지 않은 상황이고, 사용자 단말2(132)와의 네트워크 연결은 양호한 경우를 가정한다. 사용자 단말1(131) 및 사용자 단말3(133)과의 네트워크 연결 상황이 좋지 않은 경우에, 대체 컨텐츠와 함께 컨텐츠 복원 모델을 전송하여 사용자 단말1(131) 및 사용자 단말3(133)에서 자체적으로 연산하여 고품질의 컨텐츠를 생성할 수 있다. 또한, 사용자 단말2(132)와 같이 네트워크 연결 상황이 양호한 경우에는 컨텐츠 전송 서버(120)에서 고품질의 컨텐츠를 직접 전송하는 것도 가능하다. 이 경우에도, 대체 컨텐츠 및 컨텐츠 복원 모델을 전송할 수도 있으며, 고품질의 컨텐츠와 함께 컨텐츠 복원 모델을 전송하여 사용자가 원하는 품질의 컨텐츠를 제공받을 수 있도록 할 수 있다.

[0052] 도 5는 본 발명의 일 실시예에 따른 컨텐츠 전송 서버의 컨텐츠 전송 방법을 도시한 흐름도이다.

[0053] 컨텐츠 전송 서버는 컨텐츠 제공자로부터 컨텐츠를 수신한다(S510). 컨텐츠는 방송 컨텐츠, 오디오 또는 비디오 컨텐츠 등 다양한 멀티미디어 컨텐츠를 포함한다.

[0054] 다음으로, 컨텐츠 전송 서버는 수신된 컨텐츠를 군집화한다(S520). 컨텐츠 제공자로부터 제공받은 다수의 컨텐츠를 유사도에 기반하여 군집화한다. 다수의 컨텐츠를 군집화하기 위해 컨텐츠 제공자로부터 제공되는 컨텐츠에 대한 메타 데이터를 이용하거나, 이미지의 유사도를 판단하는 기계학습 기반의 모델을 이용할 수 있다.

[0055] 컨텐츠 제공 서버는 군집별 학습 모델을 학습시킨다(S530). 각 군집에 포함되어 있는 유사도가 높은 컨텐츠들을 이용하여, 저화질 또는 압축된 형태의 컨텐츠로부터 고화질 또는 고품질의 컨텐츠를 출력하도록 컨텐츠 복원 모델을 학습시킬 수 있다. 군집 내에 포함되어 있는 컨텐츠들은 유사도가 높은 컨텐츠로 서로 공유하고 있는 중복된 정보가 많기 때문에, 동일한 컨텐츠 복원 모델을 이용할 수 있다.

[0056] 사용자로부터 컨텐츠 요청이 있는 경우 컨텐츠-학습 모델 쌍을 사용자 단말에 전송한다(S540). 이 때, 전송되는 컨텐츠는 네트워크 연결 상황에 따라 대체 컨텐츠 이거나 컨텐츠 원본일 수 있다. 컨텐츠 복원 모델은 사용자 단말이 수신하는 매니페스트 파일(manifest file)의 메타 데이터에 포함될 수 있다. 사용자 단말은 매니 페스트 파일을 통해 컨텐츠에 접근 또는 컨텐츠를 요청하는 것 이외에, 해당 컨텐츠에 적합한 컨텐츠 복원 모델을 요청하여 수신할 수 있다.

[0057] 이하, 본 발명의 일 실시예에 따른 컨텐츠 복원 모델의 구현 및 이를 이용한 비디오 전송 방법을 구체적으로 설명한다.

[0058] 1.고해상도 복원

[0059] 슈퍼 해상도(Super-resolution imaging, SR)는 이미지 해상도를 향상시키는 기술로 저해상도 미디어로부터 고해상도의 이미지를 복구하는 기술이다. 이하에서는, 컨텐츠 인지 기반 슈퍼 레졸루션을 이용한 컨텐츠 복원 모델을 통한 비디오 전송 방법에 대해 설명한다. 이러한 컨텐츠 복원 모델은 적응형 스트리밍의 대안이 될 수 있으며, 안정적이고 향상된 품질을 제공할 수 있게 한다.

[0060] 컨텐츠 복원 모델로서 이미지 슈퍼 해상도 복원을 위해 심층 컨볼루션 신경망을 이용한다. 컨텐츠 인지 기반의 모델을 생성하기 위하여, 시리즈로 구성된 컨텐츠의 각 에피소드를 군집으로 하여 학습 데이터로 이용하였다. 구체적으로, 유투브에서 제공되는 2012 런던 올림픽의 농구 경기, 유투브에서 제공되는 2012 런던 올림픽의 100m 및 200m 경주 남자 결승전, 컴퓨터 게임(스타크래프트)의 플레이 영상 및 유투브의 공식 코난 오브라이언 쇼 채널로부터 제공되는 코난의 모놀로그 에피소드로 이루어진 총 4개의 데이터셋을 이용하였다. 농구 경기의 경우, 전반전을 학습 비디오로 사용하였고, 후반전을 테스트 비디오로 사용하였다. 나머지 데이터 셋에 대해서는, 학습을 위한 비디오와 테스트를 위한 비디오를 나누어 사용하였다.

[0061] 비교를 위한 신경망 모델로는 유사도에 구분없이 슈퍼 해상도 복원을 위한 벤치마크 데이터 셋을 학습한 신경망 모델(content-agnostic DNN)을 사용하였고, 보간법(nearest-neighbor interpolation)을 이용하여 해상도를 복원하는 방법을 최소 성능 기준으로 사용하였다.

[0062] 도 6은 본 발명의 일 실시예에 따른 방법 및 종래 기술에 의한 비디오 화질 개선 결과를 비교하여 도시한 도면이다.

[0063] 도 6의 (a)는 원본 비디오를 도시한 것이고, (b)는 본 실시예에 의한 컨텐츠 복원 모델을 이용한 경우, (c)는 종래 VDSR 모델을 이용한 경우, (d)는 보간법을 이용한 경우 획득한 슈퍼 레졸루션 비디오를 도시한 것이다. 도 6에 도시된 바와 같이, 본 실시예에 의하면, 컨텐츠를 유사도에 따라 분류하여 해상도 복원을 학습함으로써, 향상된 화질 개선 성능을 제공할 수 있다. 특히, 도 6에서 게임 화면을 복원한 경우를 참조하면, 본 실시예에 의하면 또렷한 텍스트 복원 성능을 보여줌을 알 수 있다.

[0064] 도 7 및 도 8은 본 발명의 일 실시예에 따른 방법 및 종래 기술에 의한 비디오 전송 성능을 비교하여 도시한 도면이다.

[0065] 도 7은 비디오 전송에서의 비트레이트(bitrate)와 품질의 관계를 도시한 것이다. 구체적으로, 도 7의 (b)에서 확인할 수 있듯이, 컨텐츠 인지 기반 컨텐츠 복원 모델에 의한 1.1Mbps 비디오는 종래의 보간법을 이용한 경우의 2.2Mbps 비디오보다 더 품질이 뛰어난 것을 확인할 수 있고, 결과적으로 본 실시예를 이용하는 경우에 네트워크 대역폭 사용을 50% 이상 줄일 수 있다. 즉, 본 실시예에 의하면, 적은 대역폭을 사용하여 높은 수준의 화질을 보여주는 비디오를 전송할 수 있다.

[0066] 도 8은 동일한 품질의 비디오를 전송하는 경우의 데이터 사용량을 도시한 것이다. 본 실시예에 의한 신경망 기반 컨텐츠 복원 모델은 7.8MB의 사이즈를 갖는다. 컨텐츠 복원 모델의 사이즈는 신경망의 파라미터 설정에 따라 변경될 수 있다. 신경망의 파라미터 갯수가 증가하면 복원 성능이 좋아지지만, 컨텐츠 복원 모델의 크기가 커지는 단점이 있다. 신경망 성능이 도 8의 (a)를 참조하면 2분 이내에, (b)를 참조하면 20초 이내에, 본 실시예에 의한 컨텐츠 복원 모델을 전송하는 데이터 전송 비용이 보상됨을 알 수 있다.

[0067] 도시된 성능 비교 결과를 참조하면, 본 실시예에 의하면 유사도가 높은 컨텐츠를 지속적으로 시청하는 사용자 단말의 경우에는 최초 컨텐츠 복원 모델을 전송한 이후에는 해당 모델을 이용하여 지속적으로 대체 컨텐츠를 복원할 수 있기 때문에 더 효율적일 수 있다. 또한, 파라미터를 로드하는데 소요되는 시간이 짧기 때문에 학습 모델을 초기화하는데에도 긴 시간이 필요하지 않으며 고해상도의 이미지를 복원하는 데에도 짧은 시간이 소요된다.

[0068] 2. 원본 영상 복원

[0069] GANs(Generative Adversarial Networks)는 이미지의 간단한 설명이 주어지면 실제의 이미지와 구분할 수 없는 이미지를 합성하는 신경망이다. 이러한 GANs을 이용하여, 중복성이 적은 비디오에 대해서도 높은 품질의 비디오를 생성할 수 있다. 본 실시예에서 대체 컨텐츠로는 컨텐츠를 YCbCr 색상 공간에서 채도를 제거하고 원본 비디오의 휘도(Y)만을 포함하여 데이터를 표현하는 LUM과 에지(edge) 검출 알고리즘을 이용하여 각 프레임의 경계선을 추출하고 1 비트 양자화를 통해 흑백 이미지를 생성하는 EDGE를 예를 들어 설명한다.

[0070] 도 9는 종래 기술에 의한 비디오 인코딩 및 디코딩 결과를 비교하여 도시한 도면이다.

[0071] 도 9의 (c) 및 (d)는 각 농구 경기, 컴퓨터 게임(스타크래프트) 영상에 적용된 LUM 및 EDGE의 예제 프레임을 도시한 것이다. 도면을 참조하면, 원본 영상에 비하여 훨씬 적은 양의 정보가 포함되어 있음을 알 수 있다. 본 실시예에서, 데이터 셋에 포함된 비디오를 학습 데이터로 이용하여 GAN 네트워크를 학습시키고 LUM 및 EDGE의 이미지를 생성하도록 한다. 예컨대, LUM의 경우에, 네트워크는 휘도 값으로부터 원래의 이미지(채도 포함)를 합성한다. LUM 및 EDGE를 이용한 이미지를 비슷한 품질의 JPEG 이미지와 비교한다.

[0072] 비교 결과, LUM(20.33KB)는 유사한 품질의 이미지를 전송하기 위해 JPEG(22.84KB)에 비하여 11% 감소된 데이터 사용을 보여준다. 도 9의 (e)는 이러한 LUM을 이용하여 이미지를 복원한 결과를 보여준다. 복원된 색이 원본과 거의 동일함을 알 수 있다. 즉, 채도의 경우 이러한 신경망 기반의 학습 모델을 통하여 중복된 정보가 잘 학습되는 요소임을 알 수 있으며, 본 발명의 일 실시예에 따라 비디오 전송 과정에서 이용되는 경우 뛰어난 성능을 보여줄 수 있다.

[0073] EDGE(3.65KB)는 JPEG(9.29KB)와 유사한 품질의 이미지를 전송하기 위해 훨씬 더 적은 데이터를 사용한다. 도 9의 (f)는 이러한 EDGE를 이용하여 디코딩된 이미지를 보여준다. 객체의 외곽선에 약간의 왜곡이 있다는 점을 제외하면, 생성된 이미지의 색상은 원본의 색상과 거의 일치한다. 이는, 비디오에 장기적인 중복성이 있는 경우에 윤곽선으로 구성된 흑백 이미지가 원본 이미지를 복원하기에 충분한 정보를 포함하고 있음을 나타낸다.

[0074] 3. 프레임 보간

[0075] 심층 신경망(DNN)을 이용한 프레임 보간 학습은, 신호 처리 기반의 프레임 보간에 비해 더 나은 성능을 보여주고 있다. 따라서, 본 실시예에 의한 컨텐츠 복원 모델이 컨텐츠들에 대한 프레임 보간을 학습하는 경우에, 대체 컨텐츠로서 프레임이 압축된 컨텐츠를 생성하고, 프레임 보간을 학습한 컨텐츠 복원 모델을 함께 전송하는 경우에 기존의 신호 처리 기반의 프레임 보간에 비해 아티팩트가 적게 나타나고 영상간 연결이 더 자연스러운 컨텐츠를 제공할 수 있다.

[0076] 이상의 설명은 본 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 다양한 수정 및 변형이 가능할 것이다. 따라서, 본 실시예들은 본 실시예의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예에 의하여 본 실시예의 기술 사상의 범위가 한정되는 것은 아니다. 본 실시예의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 실시예의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

[0077] 전술한 바와 같이, 도 5에 기재된 방법은 프로그램으로 구현되고 컴퓨터로 읽을 수 있는 기록매체에 기록될 수 있다. 본 발명의 일 실시예에 따른 컨텐츠 전송 방법을 구현하기 위한 프로그램이 기록되고 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 이러한 컴퓨터가 읽을 수 있는 기록매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피디스크, 광 데이터 저장장치 등을 포함한다. 또한 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수도 있다. 또한, 본 실시예를 구현하기 위한 기능적인(Functional) 프로그램, 코드 및 코드 세그먼트들은 본 실시예가 속하는 기술분야의 프로그래머들에 의해 용이하게 추론될 수 있을 것이다.

부호의 설명

[0078] 100: 컨텐츠 제공 시스템

110: 컨텐츠 제공자

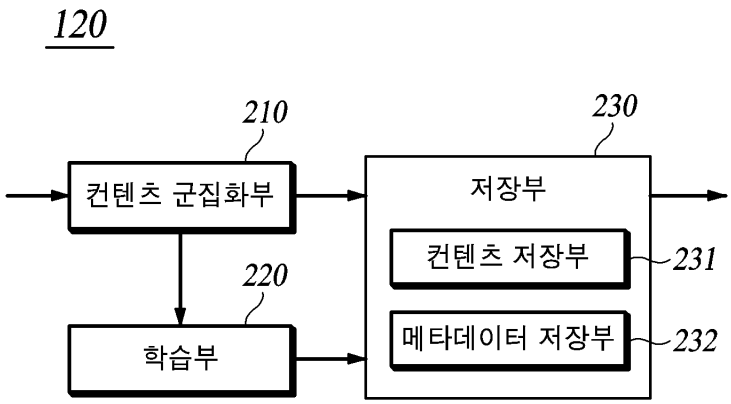
120: 컨텐츠 전송 서버

130: 사용자 단말

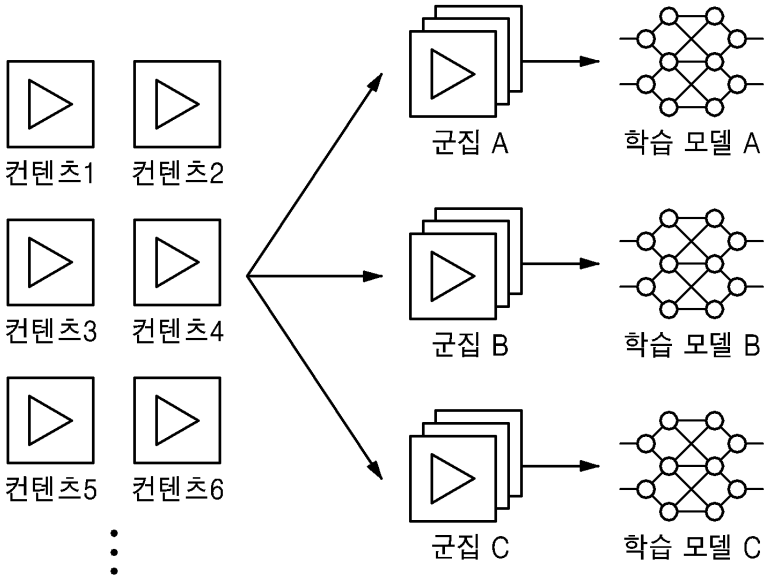
도면

도면1

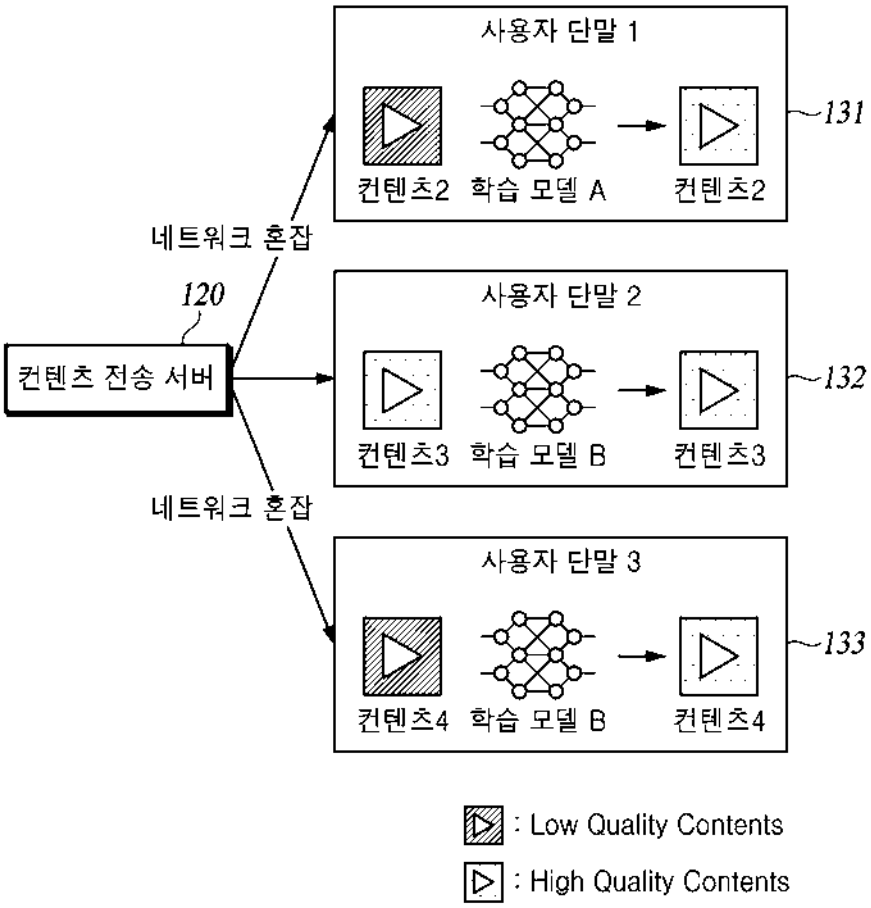
도면2


120
210
230
컨텐츠 군집화부
저장부
컨텐츠 저장부
231
220
학습부
메타데이터 저장부
232

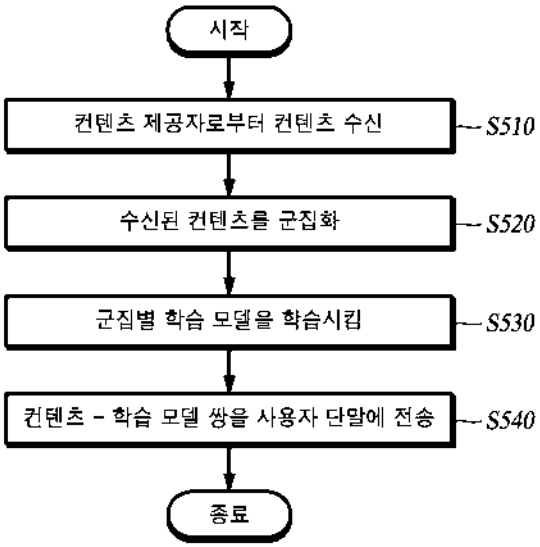
도면3


컨텐츠1
컨텐츠2
컨텐츠3
컨텐츠4
컨텐츠5
컨텐츠6
군집 A
군집 B
군집 C
학습 모델 A
학습 모델 B
학습 모델 C

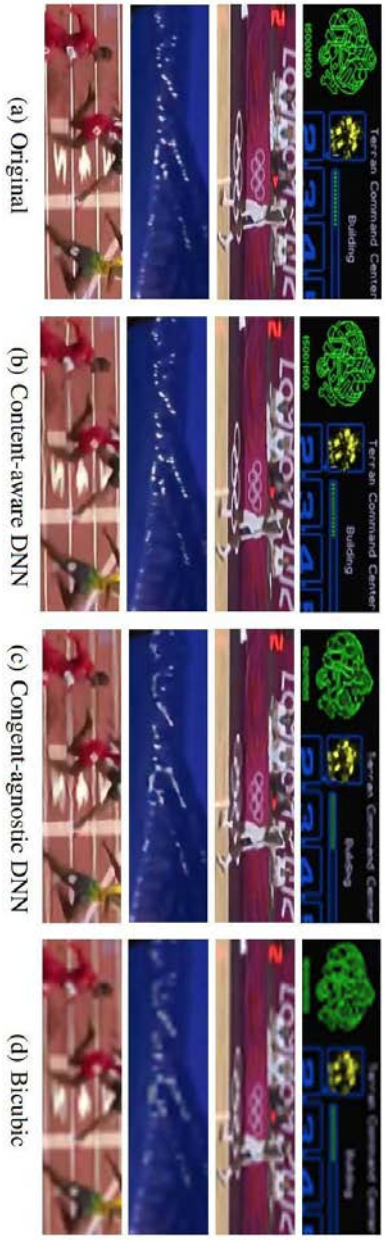
도면4


사용자 단말 1
131
컨텐츠2
학습 모델 A
컨텐츠2
네트워크 혼잡
120
컨텐츠 전송 서버
사용자 단말 2
132
컨텐츠3
학습 모델 B
컨텐츠3
네트워크 혼잡
사용자 단말 3
133
컨텐츠4
학습 모델 B
컨텐츠4
: Low Quality Contents
: High Quality Contents

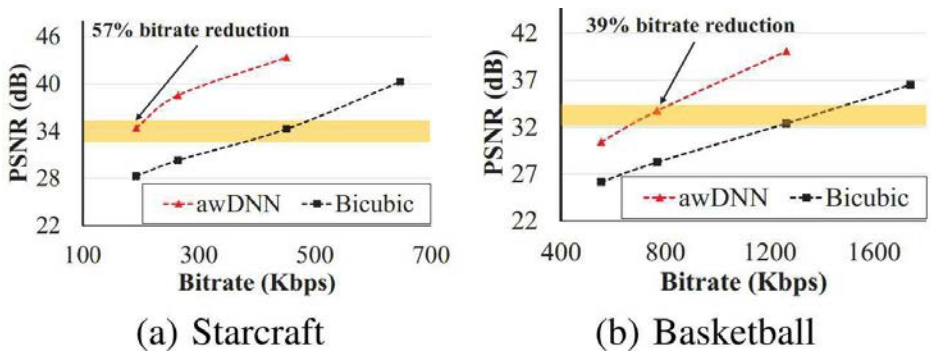
도면5


시작
컨텐츠 제공자로부터 컨텐츠 수신
S510
수신된 컨텐츠를 군집화
S520
군집별 학습 모델을 학습시킴
S530
컨텐츠 - 학습 모델 쌍을 사용자 단말에 전송
S540
종료

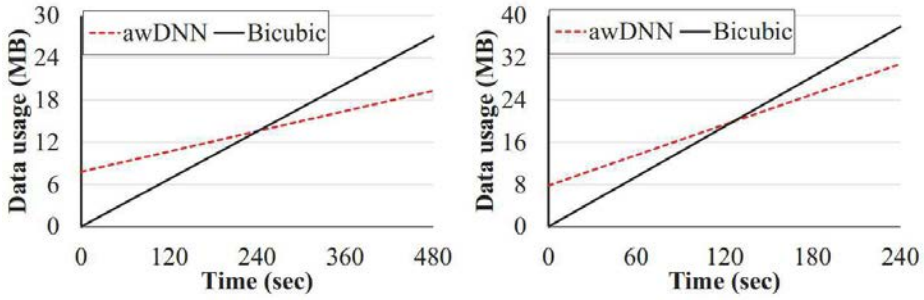
도면6


(a) Original
(b) Content-aware DNN
(c) Congent-agnostic DNN
(d) Bicubic

도면7

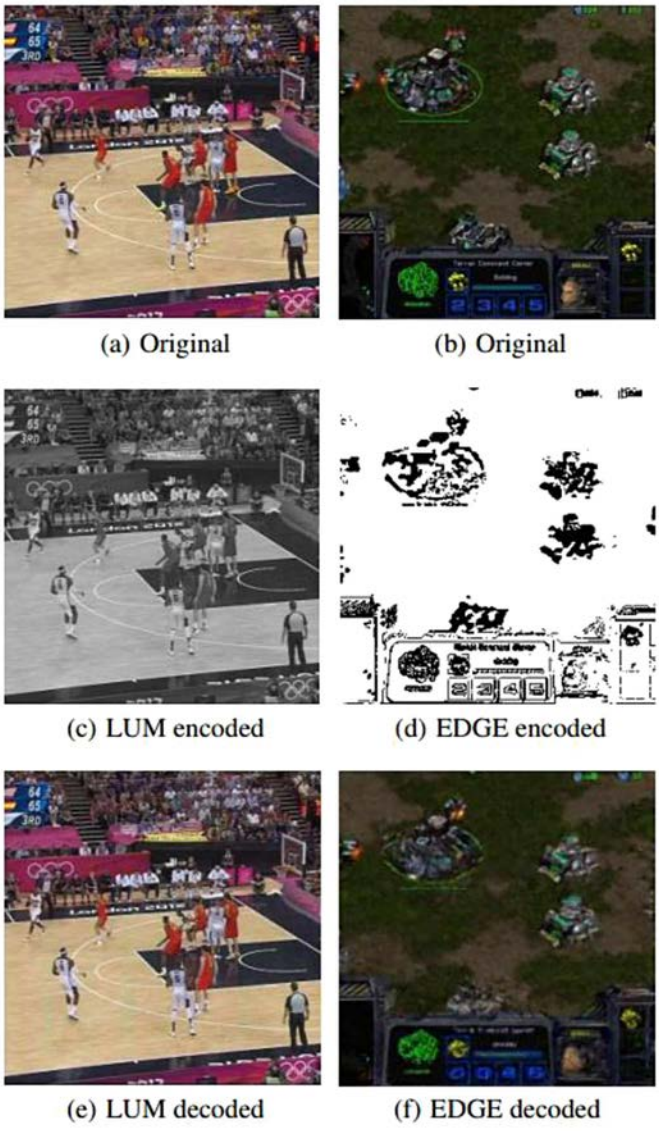

57% bitrate reduction
PSNR (dB)
46
40
34
28
22
awDNN
Bicubic
100
300
500
700
Bitrate (Kbps)
(a) Starcraft
39% bitrate reduction
PSNR (dB)
42
37
32
27
22
awDNN
Bicubic
400
800
1200
1600
Bitrate (Kbps)
(b) Basketball

도면8



(a) Starcraft (b) Basketball

도면9



(a) Original (b) Original

(c) LUM encoded (d) EDGE encoded

(e) LUM decoded (f) EDGE decoded

(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2020년07월02일
(11) 등록번호 10-2129115
(24) 등록일자 2020년06월25일

(51) 국제특허분류(Int. Cl.)
H04N 21/466 (2011.01) *H04N 21/2343* (2011.01)
H04N 21/2385 (2011.01) *H04N 21/274* (2011.01)
(52) CPC특허분류
H04N 21/4666 (2013.01)
H04N 21/2343 (2013.01)
(21) 출원번호 10-2018-0116404
(22) 출원일자 2018년09월28일
심사청구일자 2018년09월28일
(65) 공개번호 10-2020-0037015
(43) 공개일자 2020년04월08일
(56) 선행기술조사문헌
KR1020120012089 A*
KR1020170101287 A
JP2018523182 A
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
한국과학기술원
대전광역시 유성구 대학로 291(구성동)
(72) 발명자
한동수
대전광역시 유성구 대학로 291
여현호
대전광역시 유성구 대학로 291
(뒷면에 계속)
(74) 대리인
이철희

전체 청구항 수 : 총 23 항 심사관 : 홍기완

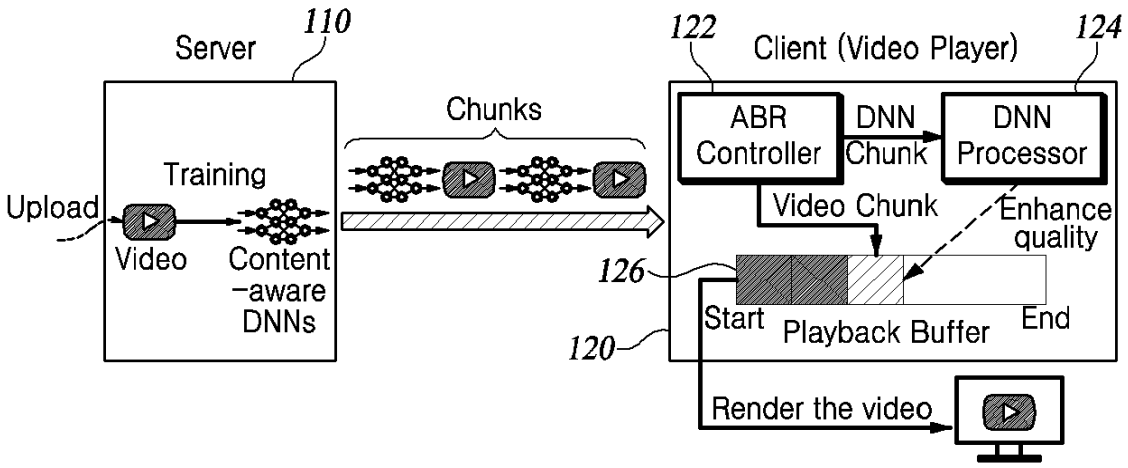
(54) 발명의 명칭 **컨텐츠 인지 신경망을 이용하여 실시간으로 적응형 비디오를 전송하는 방법 및 장치**

(57) 요 약

컨텐츠 인지 신경망을 이용하여 실시간으로 적응형 비디오를 전송하는 방법 및 장치를 개시한다.

본 실시예의 일 측면에 의하면, 컨텐츠 인지 신경망을 이용하여 실시간으로 적응형 비디오를 전송하는 클라이언트를 지원하는 서버의 방법은, 비디오를 다운로드 받는 과정, 적어도 하나의 해상도에 따라 상기 다운로드 받은 비디오를 인코딩하는 과정, 상기 인코딩된 비디오를 일정 크기로 분할하는 과정, 상기 인코딩된 비디오를 컨텐츠 인식 DNN(Deep Neural Network)을 이용해 학습하는 과정, 상기 학습된 컨텐츠 인식 DNN에 관한 정보, 상기 인코딩된 비디오에 관한 정보를 포함하는 설정 파일을 생성하는 과정, 및 상기 클라이언트의 요청에 따라 상기 설정 파일을 전송하는 과정을 포함한다.

대 표 도 - 도1



(52) CPC특허분류
H04N 21/2385 (2013.01)
H04N 21/274 (2013.01)
(72) 발명자
정영목
대전광역시 유성구 대학로 291
김재홍
대전광역시 유성구 대학로 291
신진우
대전광역시 유성구 대학로 291 (구성동,한국과학기술원)

공지예외적용 : 있음

명 세 서

청구범위

청구항 1

컨텐츠 인지 신경망을 이용하여 실시간으로 적응형 비디오를 전송하는 서버의 방법에 있어서,

비디오를 다운로드 받는 과정,

적어도 하나의 해상도에 따라 상기 다운로드 받은 비디오를 인코딩하는 과정,

상기 인코딩된 비디오를 일정 크기로 분할하는 과정,

상기 인코딩된 비디오를 컨텐츠 인식 DNN(Deep Neural Network)을 이용해 학습하는 과정,

상기 학습된 컨텐츠 인식 DNN에 관한 정보, 상기 인코딩된 비디오에 관한 정보를 포함하는 설정 파일을 생성하는 과정, 및

클라이언트의 요청에 따라 상기 설정 파일을 전송하는 과정을 포함하는 방법.

청구항 2

제1항에 있어서,

상기 인코딩된 비디오에 관한 정보는,

상기 인코딩된 비디오의 저장 위치, 해상도, 및 비트 전송율 중 적어도 하나를 포함하는 것을 특징으로 하는 방법.

청구항 3

제1항에 있어서,

상기 학습된 컨텐츠 인식 DNN에 관한 정보는,

상기 학습된 컨텐츠 인식 DNN의 레이어 수, 채널 수, 저장 위치, 크기, 및 향상된 품질의 정도 중 적어도 하나를 포함하는 것을 특징으로 하는 방법.

청구항 4

제1항에 있어서,

상기 컨텐츠 인식 DNN은 반드시 수행되는 필수 구성 요소와 선택적으로 수행될 수 있는 선택적 구성 요소로 구성됨을 특징으로 하는 방법.

청구항 5

컨텐츠 인지 신경망을 이용하여 실시간으로 적응형 비디오를 전송하는 서버 장치에 있어서,

비디오를 다운로드 받고, 클라이언트의 요청에 따라 설정 파일을 전송하는 송수신부와,

적어도 하나의 해상도에 따라 상기 다운로드 받은 비디오를 인코딩하고, 상기 인코딩된 비디오를 일정 크기로 분할하고, 상기 인코딩된 비디오를 컨텐츠 인식 DNN(Deep Neural Network)을 이용해 학습하여, 상기 학습된 컨텐츠 인식 DNN에 관한 정보, 상기 인코딩된 비디오에 관한 정보를 포함하는 상기 설정 파일로 생성하는 제어부를 포함하는 서버 장치.

청구항 6

제5항에 있어서,

상기 인코딩된 비디오에 관한 정보는,

상기 인코딩된 비디오의 저장 위치, 해상도, 및 비트 전송율 중 적어도 하나를 포함하는 것을 특징으로 하는 서버 장치.

청구항 7

제5항에 있어서,

상기 학습된 컨텐츠 인식 DNN에 관한 정보는,

상기 학습된 컨텐츠 인식 DNN의 레이어 수, 채널 수, 저장 위치, 크기, 및 향상된 품질의 정도 중 적어도 하나를 포함하는 것을 특징으로 하는 서버 장치.

청구항 8

제5항에 있어서,

상기 컨텐츠 인식 DNN은 반드시 수행되는 필수 구성 요소와 선택적으로 수행될 수 있는 선택적 구성 요소로 구성됨을 특징으로 하는 서버 장치.

청구항 9

컨텐츠 인지 신경망을 이용하여 실시간으로 적응형 비디오를 다운로드 받는 장치의 방법에 있어서,

서버 장치로부터 다운로드 받을 비디오에 대한 설정 파일을 다운로드 받는 과정,

상기 설정 파일에 저장된 정보를 이용하여 컨텐츠 인식 DNN의 수행에 필요한 추론 시간을 측정하는 과정,

상기 장치의 성능과 상기 측정된 추론 시간을 이용하여 다운로드 받을 대상을 결정하는 과정,

상기 결정된 대상에 대하여 상기 서버 장치로부터 다운로드 받는 과정,

상기 다운로드 받은 대상이 비디오이면 버퍼에 저장하고, 상기 다운로드 받은 대상이 컨텐츠 인식 DNN의 일부이면 상기 컨텐츠 인식 DNN에 추가하는 과정, 및

상기 컨텐츠 인식 DNN을 이용해 상기 버퍼에 저장된 비디오의 품질을 개선하는 과정을 포함하는 방법.

청구항 10

제9항에 있어서,

상기 품질이 개선된 비디오를 실시간으로 재생하거나 또는 전송하는 과정을 포함하는 방법.

청구항 11

제9항에 있어서,

상기 추론 시간을 측정하는 과정은,

상기 설정 파일에 저장된 컨텐츠 인식 DNN에 관한 정보를 이용해 임의로 컨텐츠 인식 DNN을 구성해 상기 컨텐츠 인식 DNN의 수행에 필요한 추론 시간을 측정하는 과정임을 특징으로 하는 방법.

청구항 12

제9항에 있어서,

상기 설정 파일은

상기 다운로드 받을 비디오의 저장 위치, 해상도, 및 비트 전송율 중 적어도 하나를 포함하는 것을 특징으로 하는 방법.

청구항 13

제9항에 있어서,

상기 설정 파일은

상기 컨텐츠 인식 DNN의 레이어 수, 채널 수, 저장 위치, 크기, 및 향상된 품질의 정도 중 적어도 하나를 포함하는 것을 특징으로 하는 방법.

청구항 14

제9항에 있어서,

상기 컨텐츠 인식 DNN은 반드시 수행되는 필수 구성 요소와 선택적으로 수행될 수 있는 선택적 구성 요소로 구성됨을 특징으로 하는 방법.

청구항 15

제9항에 있어서,

상기 다운로드 받을 대상을 결정하는 과정은,

보강학습을 프레임 워크를 사용하고 A3C(asynchronous advantage actor-critic)을 딥러닝 알고리즘으로 선택하여 상기 다운로드 받을 대상을 결정하는 과정임을 특징으로 하는 방법.

청구항 16

컨텐츠 인지 신경망을 이용하여 실시간으로 적응형 비디오를 다운로드 받는 장치에 있어서,

서버 장치로부터 다운로드 받을 비디오에 대한 설정 파일을 다운로드 받고 결정된 대상에 대하여 상기 서버로부터 다운로드 받는 송수신부와

상기 설정 파일에 저장된 정보를 이용하여 컨텐츠 인식 DNN의 수행에 필요한 추론 시간을 측정하여, 클라이언트의 성능과 상기 측정된 추론 시간을 이용하여 다운로드 받을 상기 대상을 결정하고, 상기 다운로드 받은 대상이 비디오이면 버퍼에 저장하고, 상기 다운로드 받은 대상이 컨텐츠 인식 DNN의 일부이면 상기 컨텐츠 인식 DNN에 추가하고, 상기 컨텐츠 인식 DNN을 이용해 상기 버퍼에 저장된 비디오의 품질을 개선하는 제어부를 포함하는 장치.

청구항 17

제16항에 있어서,

상기 제어부는,

상기 품질이 개선된 비디오를 실시간으로 재생함을 특징으로 하는 장치.

청구항 18

제16항에 있어서,

상기 송수신부는,

상기 품질이 개선된 비디오를 실시간으로 전송함을 특징으로 하는 장치.

청구항 19

제16항에 있어서,

상기 제어부는,

상기 설정 파일에 저장된 컨텐츠 인식 DNN에 관한 정보를 이용해 임의로 컨텐츠 인식 DNN을 구성해 상기 컨텐츠 인식 DNN의 수행에 필요한 추론 시간을 측정함을 특징으로 하는 장치.

청구항 20

제16항에 있어서,

상기 설정 파일은

상기 다운로드 받을 비디오의 저장 위치, 해상도, 및 비트 전송율 중 적어도 하나를 포함하는 것을 특징으로 하

는 장치.

청구항 21

제16항에 있어서,

상기 설정 파일은

상기 컨텐츠 인식 DNN의 레이어 수, 채널 수, 저장 위치, 크기, 및 향상된 품질의 정도 중 적어도 하나를 포함하는 것을 특징으로 하는 장치.

청구항 22

제16항에 있어서,

상기 컨텐츠 인식 DNN은 반드시 수행되는 필수 구성 요소와 선택적으로 수행될 수 있는 선택적 구성 요소로 구성됨을 특징으로 하는 장치.

청구항 23

제16항에 있어서,

상기 제어부는,

보강학습을 프레임 워크를 사용하고 A3C(asynchronous advantage actor-critic)을 딥러닝 알고리즘으로 선택하여 상기 다운로드 받을 대상을 결정함을 특징으로 하는 장치.

발명의 설명

기 술 분 야

[0001] 본 발명은 컨텐츠 인지 신경망을 이용하여 실시간으로 적응형 비디오를 전송하는 방법 및 장치에 관한 것이다.

배 경 기 술

[0002] 이 부분에 기술된 내용은 단순히 본 발명에 대한 배경 정보를 제공할 뿐 종래기술을 구성하는 것은 아니다.

[0003] 비디오 스트리밍 서비스는 지난 수십년간 빠르게 성장했다. 이러한 비디오 스트리밍 서비스의 품질은 전송 대역폭에 따라 결정되기 때문에 네트워크의 상태가 좋지 않으면 사용자의 QoE(Quality of Experience)도 저하된다. 이를 해결하기 위한 방법으로 서버 측에서는 분산 컴퓨팅 기술이 이용되고 있으며, 클라이언트(즉, 사용자) 측에서는 ABR(Adaptive Bit-Rate) 스트리밍이 전송 대역폭이 변경되거나 시간과 공간에서 변형되는 문제를 해결하는데 이용되고 있다. 그렇지만 이러한 기술들이 전송 대역폭으로부터 완전히 자유로울 수 있는 것은 아니다.

[0004] 그 외 비디오 스트리밍 서비스의 품질을 개선시키기 위해 더 좋은 코덱 선택, 적응형 비트 전송율 알고리즘의 최적화, 더 좋은 서버와 CDNs(Content Distribution Networks) 선택, 중앙 제어 평면을 통한 클라이언트와 서버간의 코디네이션 등이 이용되고 있다.

발명의 내용

해결하려는 과제

[0005] 본 실시예는, 고품질의 비디오 스트리밍 서비스를 사용자에게 제공하는데 주된 목적이 있다. 본 실시예에 의하면 동일 품질의 비디오 스트리밍 서비스를 종래 기술보다 좁은 전송 대역폭으로 사용자에게 제공할 수 있다.

과제의 해결 수단

[0006] 본 실시예의 일 측면에 의하면, 컨텐츠 인지 신경망을 이용하여 실시간으로 적응형 비디오를 전송하는 클라이언트를 지원하는 서버의 방법은, 비디오를 다운로드 받는 과정, 적어도 하나의 해상도에 따라 상기 다운로드 받은 비디오를 인코딩하는 과정, 상기 인코딩된 비디오를 일정 크기로 분할하는 과정, 상기 인코딩된 비디오를 컨텐츠 인식 DNN(Deep Neural Network)을 이용해 학습하는 과정, 상기 학습된 컨텐츠 인식 DNN에 관한 정보, 상기 인코딩된 비디오에 관한 정보를 포함하는 설정 파일을 생성하는 과정, 및 상기 클라이언트의 요청에 따라 상기

설정 파일을 전송하는 과정을 포함한다.

[0007] 상기 시스템의 실시예들은 컨텐츠 인지 신경망을 이용하여 실시간으로 적응형 비디오를 전송하는 클라이언트를 지원하는 서버 장치는, 비디오를 다운로드 받고, 상기 클라이언트의 요청에 따라 설정 파일을 전송하는 송수신부와, 적어도 하나의 해상도에 따라 상기 다운로드 받은 비디오를 인코딩하고, 상기 인코딩된 비디오를 일정 크기로 분할하고, 상기 인코딩된 비디오를 컨텐츠 인식 DNN(Deep Neural Network)을 이용해 학습하여, 상기 학습된 컨텐츠 인식 DNN에 관한 정보, 상기 인코딩된 비디오에 관한 정보를 포함하는 상기 설정 파일로 생성하는 제어부를 포함할 수 있다.

[0008] 본 실시예의 다른 측면에 의하면, 컨텐츠 인지 신경망을 이용하여 실시간으로 적응형 비디오를 전송하는 클라이언트의 방법은, 서버 장치로부터 다운로드 받을 비디오에 대한 설정 파일을 다운로드 받는 과정, 상기 설정 파일에 저장된 정보를 이용하여 컨텐츠 인식 DNN의 수행에 필요한 추론 시간을 측정하는 과정, 상기 클라이언트의 성능과 상기 측정된 추론 시간을 이용하여 다운로드 받을 대상을 결정하는 과정, 상기 결정된 대상에 대하여 상기 서버 장치로부터 다운로드 받는 과정, 상기 다운로드 받은 대상이 비디오이면 버퍼에 저장하고, 상기 다운로드 받은 대상이 컨텐츠 인식 DNN의 일부이면 상기 컨텐츠 인식 DNN에 추가하는 과정, 상기 컨텐츠 인식 DNN을 이용해 상기 버퍼에 저장된 비디오의 품질을 개선하는 과정, 및 상기 품질이 개선된 비디오를 실시간으로 재생하는 과정을 포함한다.

[0009] 본 실시예의 다른 측면에 의하면, 컨텐츠 인지 신경망을 이용하여 실시간으로 적응형 비디오를 전송하는 클라이언트 장치는, 서버 장치로부터 다운로드 받을 비디오에 대한 설정 파일을 다운로드 받고 결정된 대상에 대하여 상기 서버로부터 다운로드 받는 송수신부와 상기 설정 파일에 저장된 정보를 이용하여 컨텐츠 인식 DNN의 수행에 필요한 추론 시간을 측정하여, 상기 클라이언트의 성능과 상기 측정된 추론 시간을 이용하여 다운로드 받을 상기 대상을 결정하고, 상기 다운로드 받은 대상이 비디오이면 버퍼에 저장하고, 상기 다운로드 받은 대상이 컨텐츠 인식 DNN의 일부이면 상기 컨텐츠 인식 DNN에 추가하고, 상기 컨텐츠 인식 DNN을 이용해 상기 버퍼에 저장된 비디오의 품질을 개선하여, 상기 품질이 개선된 비디오를 실시간으로 재생하는 제어부를 포함한다.

발명의 효과

[0010] 이상에서 설명한 바와 같이 본 실시예에 의하면, 고품질의 비디오 스트리밍 서비스를 사용자에게 제공할 수 있다. 또한, 컨텐츠 인식 DNN(Deep Neural Network)을 이용함으로써, 고품질의 비디오 스트리밍 서비스를 실시간으로 사용자에게 제공할 수 있으며, 클라이언트는 자신의 하드웨어 자원(또는 성능)을 실시간으로 고려하여 비디오 스트리밍 서비스를 최적화할 수 있다. 본 실시예에 의하면, 컨텐츠 제공에 소요되는 전송 대역폭을 감소시켜 컨텐츠 제공자 혹은 CDNs 운영자는 같은 품질의 비디오를 더 적은 비용으로 전달할 수 있다.

도면의 간단한 설명

[0011] 도 1은 본 개시의 일 실시예에 따른 시스템의 개요를 나타낸 도면,

도 2는 적응형 비트 전송율을 지원하는 DNN을 나타낸 도면,

도 3은 본 개시의 일 실시예에 따라 적응형 비트 전송율을 지원하는 DNN을 나타낸 도면,

도 4는 본 개시의 일 실시예에 따른 컨텐츠 인식 DNN을 간략히 나타낸 도면,

도 5는 본 개시의 일 실시예에 따른 확장가능한 컨텐츠 인식 DNN을 나타낸 도면,

도 6은 본 개시의 일 실시예에 따른 서버의 순서도를 나타낸 도면,

도 7은 본 개시의 일 실시예에 따른 클라이언트의 순서도를 나타낸 도면,

도 8은 9가지 비디오 에피소드에 대해 본 개시의 일 실시예와 종래 기술의 평균 QoE를 비교해 도시한 도면,

도 9는 본 개시의 일 실시예와 종래 기술의 누적된 QoE를 비교해 도시한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0012] 이하, 본 발명의 일부 실시예들을 예시적인 도면을 통해 상세하게 설명한다. 각 도면의 구성요소들에 참조부호를 부가함에 있어서, 동일한 구성요소들에 대해서는 비록 다른 도면상에 표시되더라도 가능한 한 동일한 부호를 가지도록 하고 있음에 유의해야 한다. 또한, 본 발명을 설명함에 있어, 관련된 공지 구성 또는 기능에 대한 구

체적인 설명이 본 발명의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략한다.

[0013] 또한, 본 발명의 구성 요소를 설명하는 데 있어서, 제 1, 제 2, A, B, (a), (b) 등의 용어를 사용할 수 있다. 이러한 용어는 그 구성 요소를 다른 구성 요소와 구별하기 위한 것일 뿐, 그 용어에 의해 해당 구성 요소의 본질이나 차례 또는 순서 등이 한정되지 않는다. 명세서 전체에서, 어떤 부분이 어떤 구성요소를 '포함', '구비'한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 또한, 명세서에 기재된 '…부', '모듈' 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.

[0014] 도 1은 본 개시의 일 실시예에 따른 시스템의 개요를 나타낸 도면이다.

[0015] 도 1을 참조하면, 본 개시의 일 실시예에 따른 시스템인 NAS(Neural Adaptive Streaming, 이하 'NAS')는 적어도 하나의 비디오를 저장하고 있는 서버(110)와 서버(110)로부터 비디오를 다운로드하면서 실시간으로 재생할 수 있는 클라이언트(120)로 구성될 수 있다.

[0016] NAS는 DASH(Dynamic Adaptive Streaming over HTTP)에서 표준화한 HTTP 적응형 스트리밍 뿐만 아니라 다른 스트리밍에도 구현될 수 있다.

[0017] 서버(110)는 비디오가 전송되면 다양한 비트 전송율(bitrate)로 인코딩한 후 여러 개의 청크(chunk)로 나누어 저장한다. 또한, 서버(110)는 품질 향상을 위해 클라이언트(120)에서 이용할 컨텐츠 인식 DNN(Deep Neural Network, 이하 'DNN')을 학습시킨다. 이후 서버(110)는 학습된 컨텐츠 인식 DNN에 관한 정보, 비트 전송율, 해상도, 인코딩된 비디오를 다운로드 받을 수 있는 URL 등을 설정 파일(예컨대, manifest 파일 등)에 저장할 수 있다.

[0018] 서버(110)는 상기 비디오를 전송받고, 상기 설정 파일을 전송하는 송수신부(미도시)와 상기 다운로드 받은 비디오를 인코딩하여 일정 크기로 분할하고, 컨텐츠 인식 DNN을 이용해 학습하는 제어부(미도시)로 구성될 수 있다.

[0019] 클라이언트(120)는 비디오, 컨텐츠 인식 DNN, 설정 파일 등을 서버(110)로부터 다운로드 받을 수 있다. 클라이언트(120)는 먼저 상기 설정 파일을 다운로드 받아 클라이언트(120)에서 사용할 컨텐츠 인식 DNN과 비트 전송율을 결정하고, 결정된 컨텐츠 인식 DNN과 비트 전송율의 비디오를 다운로드 받을 수 있다. 클라이언트(120)는 ABR controller(122), DNN 프로세서(124)와 재생할 비디오를 저장할 버퍼(126)를 포함할 수 있다. ABR controller(122)는 현재 클라이언트(120)의 성능을 고려하여 재생할 비디오의 비트 전송율 및 다운로드할 대상으로 컨텐츠 인식 DNN 또는 비디오를 선택할 수 있다. DNN 프로세서(124)는 경량화된 메커니즘을 이용하여 클라이언트(120)의 리소스에 맞는 최적의 가용 메커니즘을 선택할 수 있다. 클라이언트(120)는 서버(110)로부터 컨텐츠 인식 DNN을 다운로드 받으면 DNN 프로세서(124)로 전달하고, 비디오를 다운로드 받으면 버퍼(126)로 전달한다. DNN 프로세서(124)가 컨텐츠 인식 DNN을 전달받으면 컨텐츠 인식 DNN을 초기화한다. 컨텐츠 인식 DNN은 비디오 화질(또는 품질)의 향상을 위해 프레임 단위로 수행된다. 이후, DNN 프로세서(124)는 다운로드되어 버퍼(126)에 저장된 비디오를 이용해 화질이 개선된 비디오를 생성해 다운로드된 비디오와 교체하여 버퍼(126)에 저장한다. 또는 DNN 프로세서(124)는 다운로드되어 버퍼(126)에 저장된 비디오를 이용해 화질이 개선된 비디오를 생성해 바로 재생할 수도 있다. 이로써 실질적으로 재생이 되는 비디오는 화질이 개선된 비디오가 될 수 있다. 디코딩, 컨텐츠 인식 DNN, 그리고 인코딩은 지연 시간을 최소화하기 위해 파이프라인과 병렬로 처리될 수 있다.

[0020] 클라이언트(120)는 서버(110)로부터 다운로드 받을 비디오에 대한 설정 파일을 다운로드 받고 비디오 및 컨텐츠 인식 DNN을 다운로드 받는 송수신부(미도시)와 상기 설정 파일에 저장된 정보를 이용하여 컨텐츠 인식 DNN의 수행에 필요한 추론 시간을 측정하여, 상기 클라이언트의 성능과 상기 측정된 추론 시간을 이용하여 다운로드 받을 상기 대상을 결정하고, 상기 다운로드 받은 대상이 비디오이면 버퍼에 저장하고, 상기 다운로드 받은 대상이 컨텐츠 인식 DNN의 일부이면 상기 컨텐츠 인식 DNN에 추가하고, 상기 컨텐츠 인식 DNN을 이용해 상기 버퍼에 저장된 비디오의 품질을 개선하여, 상기 품질이 개선된 비디오를 실시간으로 재생하는 제어부(미도시)로 구성될 수 있다.

[0021] 서버(110)와 클라이언트(120)의 구성을 제어부와 송수신부로 나누어 설명했으나, 하나의 구성으로 통합되어 구현될 수 있으며 또는 하나의 구성이 여러 개의 구성으로 나누어 구현될 수도 있다.

[0022] 또한, 도 1에서는 서버와 클라이언트로 나누어 설명하였으나, 서버와 클라이언트 사이에는 프록시(proxy) 서버가 더 포함될 수 있다. 이 경우 프록시 서버는 실시간 비디오 재생을 제외한 도 1에서 설명한 클라이언트의 기능을 수행하고, 클라이언트는 상기 프록시 서버로부터 실시간으로 재생할 수 있는 비디오를 수신하여 재생할 수

있다.

[0023] 도 2는 적응형 비트 전송율을 지원하는 DNN을 나타낸 도면이다.

[0024] 일반적인 DNN은 적응형 비트 전송율을 지원하기 어렵다. 적응형 비트 전송율을 지원하기 위해서는 다중 해상도를 입력으로 할 수 있어야 하며, DNN에 의한 추론이 실시간으로 수행되어야 한다. 즉, 시스템이 DNN을 이용해 비디오의 화질을 개선하고, 클라이언트가 적응형 비트 전송율(adaptive bitrate)을 지원한다면, 서버는 다양한 비트 전송율(또는 해상도)에 따라 DNN을 학습시켜야 한다. 여기서, 비트 전송율과 해상도는 특정한 관계에 있다. 예를 들면, 고해상도의 비디오를 실시간으로 재생하기 위해서는 비트 전송율은 높아야 하나, 저해상도의 비디오를 실시간으로 재생하는 경우에는 비트 전송율이 높을 필요는 없다. 그 외 비디오의 크기(가로, 세로)도 해상도와 비트 전송율에 영향을 미칠 수 있다.

[0025] 도 2는 예를들어 해상도가 240p, 360p, 480p, 720p인 비디오를 입력해 1080p인 비디오를 출력하는 DNN을 나타낸 것이다. 이 때 DNN은 입력 해상도와 관계없이 적용될 블록(즉, 레이어)과 입력 해상도에 따라 다르게 적용될 블록으로 구성될 수 있다. 이와 같이 DNN을 이용해 서로 다른 저해상도의 비디오(또는 이미지)를 학습시켜 고해상도의 비디오를 출력시킬 수 있는 시스템을 MDSR(Multi-scale Deep Super Resolution) 시스템이라 한다.

[0026] 도 3은 본 개시의 일 실시예에 따라 적응형 비트 전송율을 지원하는 DNN을 나타낸 도면이다.

[0027] 도 2와 같은 DNN은 입력 비디오의 해상도와 관계없이 일부 블록을 공유하기 때문에 저장 공간을 줄일 수는 있으나, 입력 비디오의 해상도가 추론시간에 크게 영향을 미칠 수 있다. 예를 들어, 해상도가 720p인 비디오의 추론시간은 해상도가 240p인 비디오의 추론시간보다 약 4.3배 더 걸릴 수 있다. 따라서 해상도가 240p인 비디오는 실시간으로 재생이 가능하나, 해상도가 720p인 비디오는 실시간으로 재생하기 불가능할 수 있다. 만약 고해상도인 비디오를 실시간으로 재생하기 위해 DNN의 구성을 축소하게 되면 해상도가 낮은 비디오의 화질이 좋지 않을 수 있다.

[0028] 본 개시에서는 이를 해결하기 위해 입력 비디오의 해상도 별로 DNN을 따로 구성할 수 있다. 또한, 각 DNN은 클라이언트의 시간적 변화에 적응할 수 있도록 구성될 수 있다. 예를 들면, 해상도가 240p인 비디오의 DNN(510)은 14개의 레이어(layers)로 구성될 수 있으며, 클라이언트의 시간적인 성능 변화에 따라 8번째 레이어(512)를 실행 후 9번째 레이어(514)와 10번째 레이어(516)를 건너뛰고 11번째 레이어(518)를 실행할 수도 있다. 즉, DNN 추론시간이 비디오를 실시간으로 재생하기에 충분하다면 모든 레이어를 실행할 수 있으나, 그렇지 않다면 일부 레이어의 실행을 생략할 수 있다. 이와 관련하여서는 도 5에서 자세히 설명한다.

[0029] DNN에 관한 정보로는 입력 비디오의 해상도, 레이어의 수, 채널의 수, DNN의 용량, 출력 비디오의 품질 등이 될 수 있으며, 서버는 상기 DNN에 관한 정보를 설정 파일에 저장할 수 있다.

[0030] 도 4는 본 개시의 일 실시예에 따른 컨텐츠 인식 DNN을 간략히 나타낸 도면이다.

[0031] 비디오의 에피소드가 다양하기 때문에 모든 비디오에 잘 동작하는 범용 DNN 모델을 개발하는 것은 현실적으로 불가능하다. 이에 본 개시에서는 비디오의 에피소드(즉, 배경, 환경 등)마다 다른 DNN을 적용하는 컨텐츠 인식 DNN을 이용한다. 다만, 배경, 환경, 등장인물 등이 유사한 에피소드의 경우 동일한 컨텐츠 인식 DNN을 이용할 수도 있다. 입력 비디오의 해상도, 출력 비디오의 품질, 뿐만 아니라 비디오의 에피소드까지 고려한 컨텐츠 인식 DNN을 학습시키는 것은 시간이나 비용측면에서 비효율적일 수 있다. 본 개시에서는 이를 해결하기 위해 가장 많이 이용되는 에피소드를 일반 모델로 하여 학습한 후, 이를 기초로 다른 에피소드를 학습하여 학습 시간이나 비용 등을 줄일 수 있다. 가장 많이 이용되는 에피소드는 하나일 수도 있지만 여러 개일 수도 있다.

[0032] 도 5는 본 개시의 일 실시예에 따른 확장 가능한 컨텐츠 인식 DNN을 나타낸 도면이다.

[0033] 본 개시의 일 실시예에 따른 컨텐츠 인식 DNN은 확장이 가능하다. 상기 컨텐츠 인식 DNN은 필수 구성 요소(510)와 선택적 구성 요소(520)로 나뉠 수 있다. 필수 구성 요소(510)는 반드시 실행되어야 하나, 선택적 구성 요소(520)는 반드시 실행되어야 하는 것은 아니다. 필수 구성 요소(510)는 사전 처리 과정(512)과 사후 처리 과정(514)으로 구성될 수 있다. 선택적 구성 요소(520)는 다중 잔여 블록(522, 524, 526, 528)으로 구성될 수 있다. 또한, 다중 잔여 블록(522, 524, 526, 528)은 각각 2개의 컨볼루션 계층으로 구성될 수 있다. 선택적 구성 요소(520)를 실행하는 경우 재생되는 비디오의 품질은 더 좋아질 수 있다.

[0034] 서버가 상기 확장 가능한 컨텐츠 인식 DNN을 지원하는 경우, 상기 서버는 필수 구성 요소(510)와 선택적 구성 요소(520)의 포함 여부에 따른 모든 경로에 대해 비디오를 학습시켜야 한다. 따라서 상기 경로는 다양할 수 있다. 학습은 임의로 경로를 지정하여 이루어 질 수 있으며 출력되는 비디오와 원본 비디오 간의 오차가 적도록

학습된다. 학습이 완료되면, 상기 서버는 상기 확장 가능한 컨텐츠 인식 DNN을 청크 단위로 분할하여 저장하고, 상기 분할된 확장 가능한 컨텐츠 인식 DNN이 저장된 위치를 설정 파일에 URL로 저장할 수 있다.

[0035] 클라이언트가 상기 확장 가능한 컨텐츠 인식 DNN을 이용하는 경우, 먼저 필수 구성 요소(510)를 다운로드 받고, 상기 클라이언트의 실시간 성능(또는 자원)을 고려하여 상기 선택적 구성 요소(520)를 실행할지 여부 및 실행할 구성(522, 524. 526, 528) 등을 비디오 스트리밍 서비스 중에 결정할 수 있다. 상기 클라이언트는 처리중인 비디오 청크의 재생 시간까지 남은 시간을 계산해 이를 기초로 사용할 수 있는 상기 확장 가능한 컨텐츠 인식 DNN의 최대 레이어 수를 계산할 수 있다. 상기 클라이언트는 룩업(look-up) 테이블을 이용해 레이어 수와 추론 시간을 저장할 수 있다. 또한, 상기 클라이언트는 상기 선택적 구성 요소(520)를 서버로부터 다운받을지 여부에 대해서도 결정할 수 있다. 예를 들면, 클라이언트가 파일을 다운로드 받으면서 비디오 스트리밍 서비스를 이용한다면 상기 클라이언트는 상기 비디오 스트리밍 서비스에 이용할 자원이 충분치 않기 때문에 확장 가능한 컨텐츠 인식 DNN의 필수 구성 요소(510)만을 실행할 수 있다. 그러나, 클라이언트가 오로지 비디오 스트리밍 서비스만 이용한다면 필수 구성 요소(510)뿐만 아니라 선택적 구성 요소(520) 또한 실행할 수 있다. 클라이언트가 확장 가능한 컨텐츠 인식 DNN을 이용한다면 전송 시작시 필수 구성 요소(510)만을 수행함으로써 빠르게 비디오 스트리밍 서비스를 사용자에게 제공할 수 있다. 또한, 상기 클라이언트의 자원을 실시간으로 반영할 수 있기 때문에 사용자는 지연없이 비디오 스트리밍 서비스를 이용할 수 있다.

[0036] 도 6은 본 개시의 일 실시예에 따른 서버의 순서도를 나타낸 도면이다.

[0037] 상기 서버는 다른 장치로부터 비디오를 다운로드 받는다(610). 상기 비디오는 다양한 클라이언트들을 위해 제공되기 위한 것이다.

[0038] 상기 서버는 상기 다운로드 받은 비디오를 다양한 해상도 또는 비트 전송율에 따라 인코딩을 한다(620). 해상도와 비트 전송율은 특정한 관계에 있다. 예를 들면, 고해상도의 비디오를 실시간으로 재생하기 위해서는 비트 전송율은 높아야 하나, 저해상도의 비디오를 실시간으로 재생하는 경우에는 비트 전송율이 높을 필요는 없다. 또한 비디오의 크기(가로, 세로)도 해상도와 비트 전송율에 영향을 미칠 수 있다.

[0039] 상기 서버는 상기 인코딩된 비디오를 청크 단위로 분할한다(630).

[0040] 상기 서버는 상기 인코딩된 비디오를 이용해 컨텐츠 인식 DNN을 학습시킨다(640). 컨텐츠 인식 DNN은 해상도 별로 따로 학습될 수 있으며, 일반 모델을 이용해 학습될 수도 있다. 일반 모델이 없다면 초기화된 모델을 이용할 수도 있다.

[0041] 상기 서버는 상기 인코딩된 비디오의 저장 위치, 상기 컨텐츠 인식 DNN에 대한 정보 등을 설정 파일로 생성한다(650). 학습된 컨텐츠 인식 DNN에 관한 정보, 비트 전송율, 해상도, 인코딩된 비디오를 다운로드 받을 수 있는 URL 등이 설정 파일에 저장될 수 있다. 이때 학습된 컨텐츠 인식 DNN에 관한 정보는 인덱스 값일 수도 있다.

[0042] 도 7은 본 개시의 일 실시예에 따른 클라이언트의 순서도를 나타낸 도면이다.

[0043] 먼저, 상기 클라이언트는 서버로부터 다운로드 받은 비디오에 대한 설정 파일을 다운로드 받는다(710). 상기 설정 파일에는 상기 다운로드 받을 비디오에 관한 정보 외에도 상기 비디오의 품질을 향상시키는데 이용될 컨텐츠 인식 DNN에 대한 정보도 포함되어 있다. 상기 컨텐트 인식 DNN에 대한 정보는 컨텐트 인식 DNN의 인덱스 정보일 수 있다. 만약, 상기 클라이언트가 저장하고 있는 컨텐츠 인식 DNN이 있다면, 상기 서버와 이에 대한 정보를 공유할 수 있다. 이후 상기 클라이언트는 설정 파일에 저장하고 있는 컨텐츠 인식 DNN이 지시되면 상기 비디오만을 다운로드 받을 수 있다.

[0044] 상기 클라이언트는 상기 설정 파일에 저장된 정보를 이용하여 컨텐츠 인식 DNN의 수행에 필요한 추론 시간을 측정한다(720). 상기 추론 시간을 측정하기 위해 모든 옵션에 대한 상기 설정 파일에 저장된 컨텐츠 인식 DNN을 다운로드할 수 있다. 그러나, 이 경우 불필요하게 많은 자원과 시간을 낭비할 수 있다. 따라서, 본 개시의 일 실시예에서 상기 클라이언트는 상기 컨텐츠 인식 DNN을 모두 다운로드 하지 않고 상기 설정 파일에 저장된 옵션에 대한 정보를 이용해 임의로 컨텐츠 인식 DNN을 구성할 수 있다. 즉, 실제 사용할 컨텐츠 인식 DNN을 다운로드 받지 않고, 상기 설정 파일에 저장된 입력 비디오의 해상도, 품질 레벨, 레이어의 수, 및 채널의 수 등을 이용해 임의로 컨텐츠 인식 DNN을 구성하여 컨텐츠 인식 DNN의 수행에 필요한 추론 시간을 측정할 수 있다. 이렇게 하는 경우, 상기 클라이언트가 4개의 DNN 옵션을 테스트하는데 소요되는 시간은 일반적인 비디오 청크의 재생 시간보다 적기 때문에 상기 클라이언트는 두 번째 비디오 청크를 재생하기 전에 어떤 옵션의 컨텐츠 인식 DNN을 사용할지 결정할 수 있다. 선택적으로, 상기 클라이언트는 DNN 옵션별로 수행하는데 필요한 추론 시간을

미리 저장하고 이를 이용해 사용할 컨텐츠 인식 DNN을 선택할 수도 있다.

[0045] 상기 클라이언트는 상기 클라이언트의 성능과 상기 측정된 추론 시간을 이용하여 다운로드 받을 대상을 결정한다(730). 상기 클라이언트는 통합된 ABR 알고리즘(integrated adaptive bitrate algorithm)을 이용해 이후에 재생될 비디오를 다운로드 받거나 컨텐츠 인식 DNN을 다운로드 받을 수 있다. 상기 통합된 ABR 알고리즘은 직접 최적화하는 보강학습(Reinforcement Learning) 프레임 워크를 사용하고 A3C(asynchronous advantage actor-critic)을 딥러닝 알고리즘으로 선택한다. 구체적으로, 관측으로부터 전략(또는 정책)을 배우고 다운로드된 컨텐츠 인식 DNN의 비율, 컨텐츠 인식 DNN으로 인한 품질 향상, 네트워크 처리량 샘플 및 버퍼 점유율과 같은 원시 관측을 결정에 매핑할 수 있다.

[0046] 보강학습에서 에이전트는 환경과 상호 작용을 한다. 반복되는 t마다 에이전트는 환경에서 상태(s_t)를 관찰한 후 작업(a_t)을 수행한다. 이후, 환경은 보상(r_t)을 생성하고 상태를 s_{t+1}로 갱신한다. 정책(π)은 주어진 상태(s_t)에서 행동(a_t)을 취할 확률을 제공하는 함수로 다음과 같이 정의될 수 있다.

[0047] $\pi(s_t, a_t): \rightarrow [0, 1]$

[0048] 목표는 미래의 할인 보상의 합($\sum_{t=0}^{\infty} \gamma^t r_t$)을 최대화하는 정책($\pi$)을 학습한다. 여기서, $\gamma \in (0,1]$은 미래 보상에 대한 할인율이다.

[0049] 또한, 작업의 셋($\{a_t\}$)는 컨텐츠 인식 DNN 청크를 다운로드할지 또는 특정 비트 전송율의 비디오 청크를 다운로드할지 여부로 지정할 수 있다. 보상(r_t)는 비트 전송율 유틸리티, 재버퍼링 시간 및 선택된 비트 전송율의 평활도의 함수로 QoE 측정항목이 될 수 있다. 상태(s_t)는 다운로드 받을 남은 컨텐츠 인식 DNN 청크의 개수, 처리량 측정치 및 클라이언트 측정치 (예를 들어, 버퍼 점유율, 과거의 비트 전송율)를 포함할 수 있다. 이후, 컨텐츠 인식 DNN 다운로드 및 품질 향상을 반영하여 환경 보상 및 상태를 업데이트한다. 컨텐츠 인식 DNN 다운로드는 '남은 컨텐츠 인식 DNN 청크의 개수'를 감소시켜 상태를 업데이트할 수 있다. 다운로드된 각 비디오 청크에 대한 보상은 컨텐츠 인식 DNN 기반 품질 향상에 반영하도록 업데이트될 수 있다. 상기 품질 향상은 다운로드된 컨텐츠 인식 DNN의 일부분에 대한 함수일 수 있다. 특히 컨텐츠 인식 DNN이 제공하는 평균 품질 향상에 따른 QoE의 비트율 유틸리티 구성 요소를 향상시킬 수 있다. 표 1은 본 개시에 따른 상태(s_t)를 나타낸다.

표 1

[0050]

type	상태(S_t)
DNN status	남은 컨텐츠 인식 DNN 청크의 개수
Network status	과거 N개의 청크에 대한 처리량
	과거 N개의 청크를 다운로드 받는데 걸리는 시간
Client status	재생할 비디오를 저장할 버퍼 점유율
	다음 비디오 청크의 크기(sizes)
Video status	최근 비디오 청크의 비트 전송율
	남은 비디오 청크의 개수

[0051] 보강학습에서는 정책을 나타내는 배우(actor)와 정책(policy)의 성능을 평가하는 데 사용되는 비평가라는 두 가지 신경 근사를 가지고 있다. 본 개시에서는 배우와 비평 네트워크를 교육시키기 위해 policy gradient method을 이용한다. 에이전트는 먼저 현재 정책($\pi_\Theta(s_t, a_t)$)에 따라 궤적을 생성한다. 여기서, Θ는 배우의 신경망의 매개 변수(또는 가중치)를 나타낼 수 있다. 비판 네트워크는 이러한 궤적을 관찰하고 상태 s_t에서 시작하여 정책 π_Θ 를 따르는 동작 z를 취하는 것과 관련하여 예상되는 보상인 동작 값 함수 $Q^{\pi_\theta}(s_t, a_t)$를 추정하는 방법을 학습한다.

[0052] $$\theta \leftarrow \theta + \alpha \sum_t \nabla_\theta \log \pi_\theta(s_t, a_t)(Q^{\pi_\theta}(s_t, a_t) - V^{\pi_\theta}(s_t)),$$

[0053] 여기서 $V^{\pi_\theta}(s_t)$ 는 상태 s_t에서 시작하는 π_θ 의 총 예상 보상을 나타내는 값 함수이고, α는 학습 속도이다. 본 개시에 따른 보강학습에서 보상은 컨텐츠 인식 DNN이 제공하는 평균 QoE 향상을 반영하기 때문에 비평 네트워크는 업데이트된 총 보상을 예측하는 방법을 학습한다. 이를 통해 배우(actor)는 QoE를 최대화하기 위해 비디오와 DNN 다운로드의 균형을 유지하는 정책을 학습할 수 있다.

[0054] 다시 도 7을 참조하여, 상기 클라이언트는 상기 다운로드 받은 대상이 비디오 청크이면 버퍼에 저장하고, 상기 다운로드 받은 대상이 컨텐츠 인식 DNN 청크이면 상기 컨텐츠 인식 DNN에 추가한다(740).

[0055] 상기 클라이언트는 다운로드 받은 비디오 청크를 컨텐츠 인식 DNN을 이용해 화질을 개선한다(750). 상기 클라이언트는 DNN을 수행하기 위한 전용 DNN 프로세서를 이용할 수 있다.

[0056] 상기 클라이언트는 화질이 개선된 비디오 청크를 실시간으로 재생한다(760).

[0057] 도 6과 도 7에서는 각 과정을 순차적으로 실행하는 것으로 기재하고 있으나, 이는 본 발명의 일 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것이다. 다시 말해, 본 발명의 일 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 발명의 일 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 도 6과 도 7에 기재된 순서를 변경하여 실행하거나 과정들 중 하나 이상의 과정을 병렬적으로 실행하는 것으로 다양하게 수정 및 변형하여 적용 가능할 것이므로, 도 6과 도 7은 시계열적인 순서로 한정되는 것은 아니다.

[0058] 한편, 도 6과 도 7에 도시된 과정들은 컴퓨터로 읽을 수 있는 기록매체에 컴퓨터가 읽을 수 있는 코드로서 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 즉, 컴퓨터가 읽을 수 있는 기록매체는 마그네틱 저장매체(예를 들면, 롬, 플로피 디스크, 하드디스크 등), 광학적 판독 매체(예를 들면, 시디롬, 디브이디 등) 및 캐리어 웨이브(예를 들면, 인터넷을 통한 전송)와 같은 저장매체를 포함한다. 또한 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어 분산방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수 있다.

[0059] 도 8은 9가지 비디오 에피소드에 대해 본 개시의 일 실시예와 종래 기술의 평균 QoE를 비교해 도시한 도면이다.

[0060] 구체적으로, 해상도 1080p이고 길이가 5분보다 긴 9가지의 비디오 에피소드를 이용해 학습을 위해 80%의 무작위 표본을 사용해 10시간 이상 학습시켰다. Pensieve는 심층 보강학습을 이용하여 QoE를 최대화한 기술이며, MPC는 QoE를 최대화하는 비트 전송율을 선택하기 위해 다음 5개의 청크에 대한 버퍼 점유율과 처리량 예측을 이용한 기술이다. 또한, BOLA는 버퍼 점유율을 기반으로 Lyapunov 최적화한 기술이다.

[0061] QoE에 대한 측정항목으로는 QoE_{lin}, QoE_{log}, 및 QoE_{hd}를 이용한다. QoE_{lin}는 선형 비트 전송율 유틸리티, QoE_{log}는 감소하는 한계 유틸리티를 나타내는 log 비트 전송율, 그리고 QoE_{hd}는 고화질 비디오를 선호하는 정도를 나타낸다. 도 8에서 오차 막대는 평균으로부터의 표준 편차를 나타낸다. 본 개시의 일 실시예에 따른 NAS는 세 가지 QoE 측정 항목 모두에서 모든 비디오 에피소드에 대해 가장 높은 QoE를 나타낸다. NAS는 지속적으로 Pensieve보다 QoE_{lin}는 43.08%, QoE_{log}는 36.26%, QoE_{hd}는 42.57% 우수함을 보이고 있다. QoE_{lin}을 사용하면 NAS가 Pensieve보다 평균 43.08%성능이 우수하지만 Pensieve는 MPC보다 19.31% 향상된 성능을 보인다. BOLA와 비교하여도 NAS는 92.28% 향상됨을 알 수 있다. 장면 복잡성, 압축 아티팩트, 시간 중복성과 같은 많은 요소가 DNN 성능에 영향을 주기 때문에 QoE 개선은 Pensieve보다 비디오 에피소드에 따라 21.89%(뷰티)에서 76.04%(음악)까지 다양하였다.

[0062] 도 9는 본 개시의 일 실시예와 종래 기술의 누적된 QoE를 비교해 도시한 도면이다.

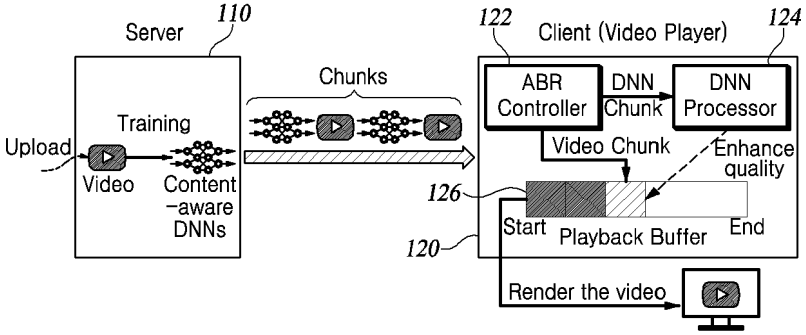
[0063] 도 9는 도 8의 9개의 비디오 에피소드 중 중간값을 나타낸 게임 에피소드를 이용하여 103개 이상의 네트워크 트레이스에 대한 QoE의 누적 분포를 도시한 것이다. NAS는 모든 네트워크 조건에서 이점을 제공한다. 일예로, NAS는 Pensieve보다 QoE_{lin}중간값이 58.55 % 향상되었다. 참고로, Pensive는 주로 비트 전송율 유틸리티를 사용하여 재버퍼링을 줄임으로써 MPC에 비해 QoE 이득을 제공한다. 반대로 NAS는 클라이언트 측의 계산을 사용하기 때문에 이러한 절충을 나타내지 않는다. 다른 비디오 에피소드를 도시하지는 않았지만 비슷한 경향을 나타낸다.

[0064] 이상의 설명은 본 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 다양한 수정 및 변형이 가능할 것이다. 따라서, 본 실시예들은 본 실시예의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위

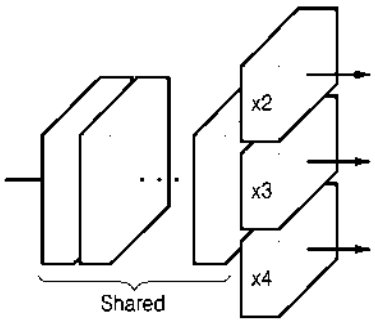
한 것이고, 이러한 실시예에 의하여 본 실시예의 기술 사상의 범위가 한정되는 것은 아니다. 본 실시예의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 실시예의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

도면

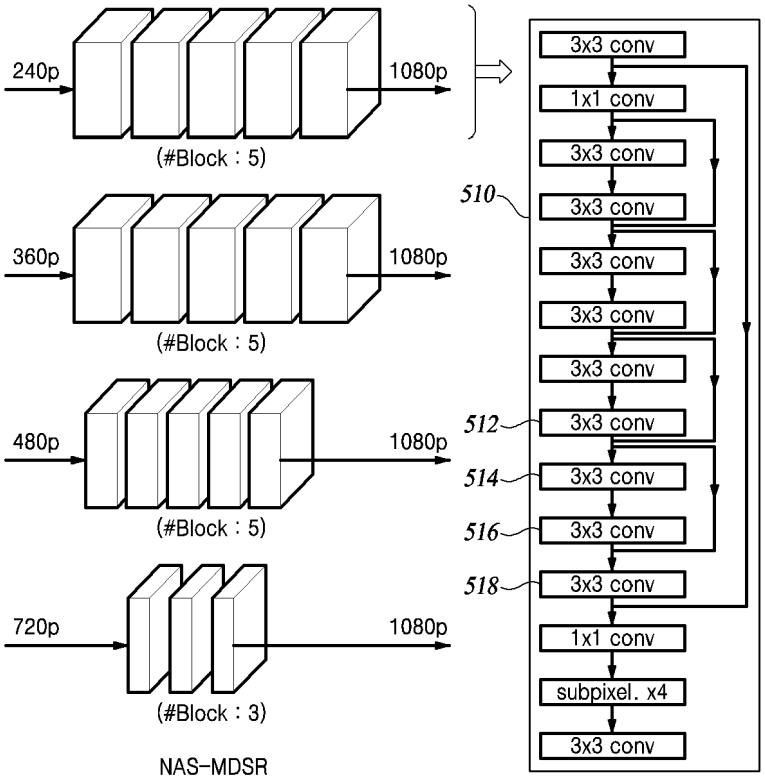
도면1



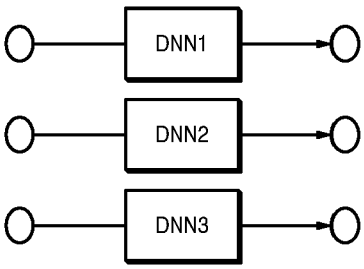
도면2



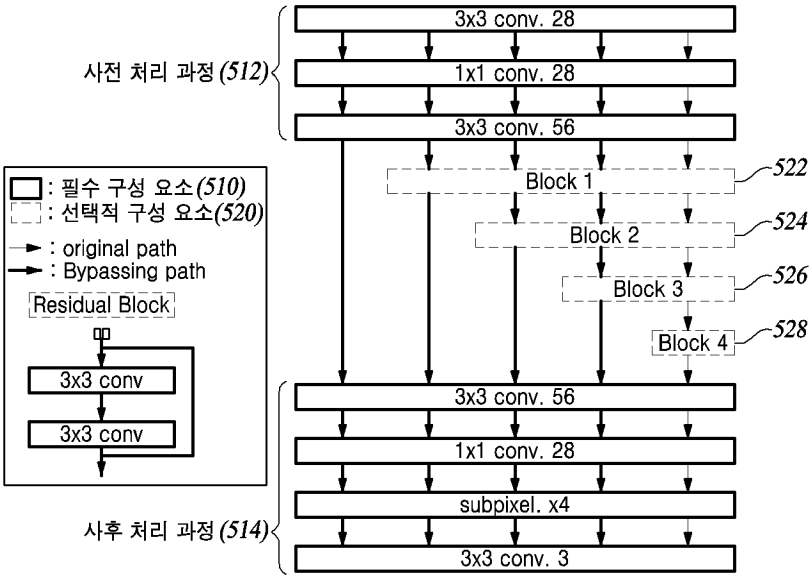
도면3


240p
1080p
(#Block : 5)
360p
1080p
(#Block : 5)
480p
1080p
(#Block : 5)
720p
1080p
(#Block : 3)
NAS-MDSR
510
512
514
516
518
3x3 conv
1x1 conv
3x3 conv
3x3 conv
3x3 conv
3x3 conv
3x3 conv
3x3 conv
3x3 conv
3x3 conv
1x1 conv
subpixel. x4
3x3 conv

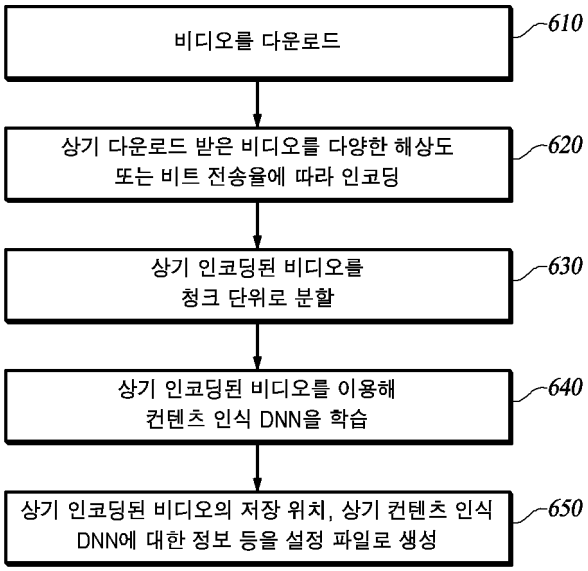
도면4


DNN1
DNN2
DNN3

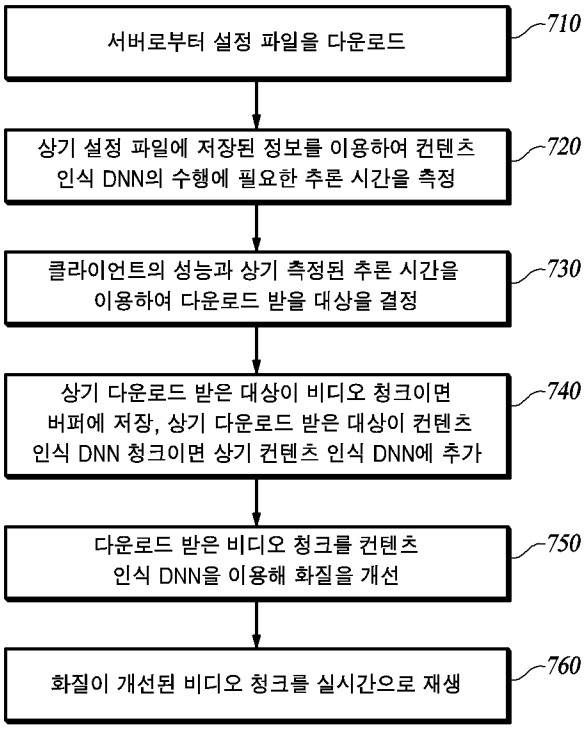
도면5


3x3 conv. 28
1x1 conv. 28
3x3 conv. 56
사전 처리 과정 (512)
Block 1
522
Block 2
524
Block 3
526
Block 4
528
: 필수 구성 요소 (510)
: 선택적 구성 요소 (520)
: original path
: Bypassing path
Residual Block
3x3 conv
3x3 conv
3x3 conv. 56
1x1 conv. 28
subpixel. x4
3x3 conv. 3
사후 처리 과정 (514)

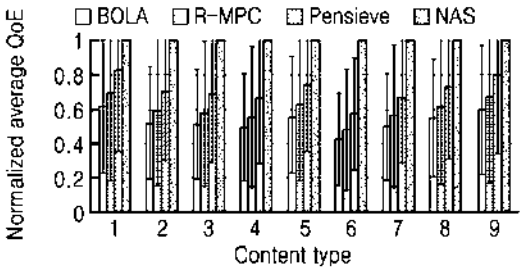
도면6


비디오를 다운로드
610
상기 다운로드 받은 비디오를 다양한 해상도
또는 비트 전송율에 따라 인코딩
620
상기 인코딩된 비디오를
청크 단위로 분할
630
상기 인코딩된 비디오를 이용해
컨텐츠 인식 DNN을 학습
640
상기 인코딩된 비디오의 저장 위치, 상기 컨텐츠 인식
DNN에 대한 정보 등을 설정 파일로 생성
650

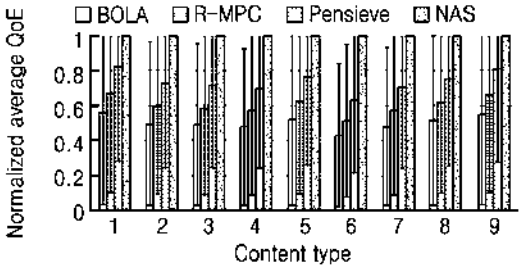
도면7


서버로부터 설정 파일을 다운로드
710
상기 설정 파일에 저장된 정보를 이용하여 컨텐츠 인식 DNN의 수행에 필요한 추론 시간을 측정
720
클라이언트의 성능과 상기 측정된 추론 시간을 이용하여 다운로드 받을 대상을 결정
730
상기 다운로드 받은 대상이 비디오 청크이면 버퍼에 저장, 상기 다운로드 받은 대상이 컨텐츠 인식 DNN 청크이면 상기 컨텐츠 인식 DNN에 추가
740
다운로드 받은 비디오 청크를 컨텐츠 인식 DNN을 이용해 화질을 개선
750
화질이 개선된 비디오 청크를 실시간으로 재생
760

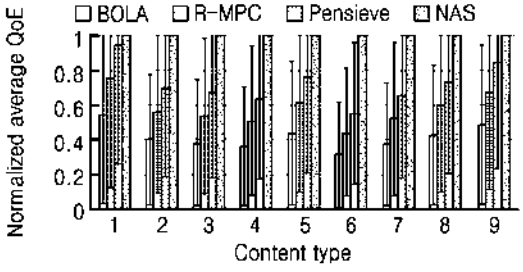
도면8



(a) QoE_{lin}

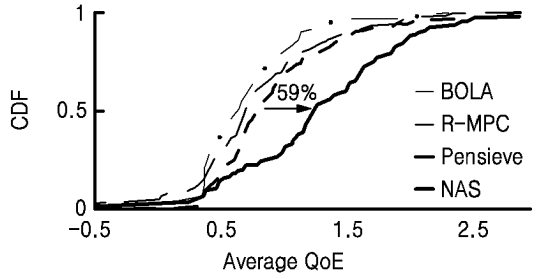


(b) QoE_{log}

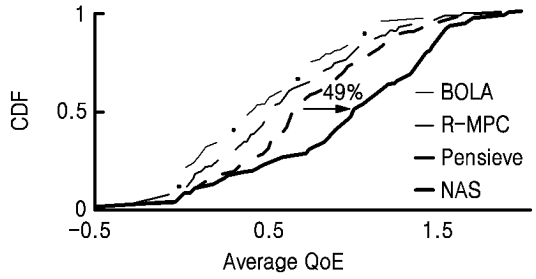


(c) QoE_{hd}

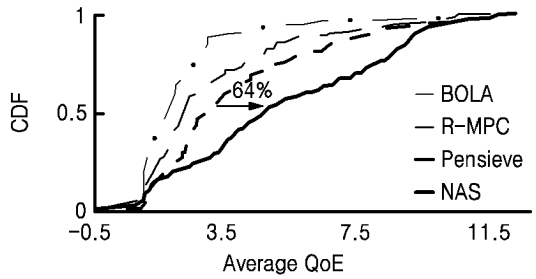
도면9



(a) QoE_{lin}



(b) QoE_{log}



(c) QoE_{hd}