

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0084011

(43) 공개일자 2025년06월10일

(51) 국제특허분류(Int. Cl.)

G06V 20/70 (2022.01) G06F 40/169 (2020.01)

G06T 11/60 (2006.01) G06V 20/40 (2022.01)

G10L 13/08 (2006.01) G10L 15/26 (2006.01)

(52) CPC특허분류

G06V 20/70 (2023.08)

G06F 40/169 (2020.01)

(21) 출원번호 10-2024-0109934

(22) 출원일자 2024년08월16일

심사청구일자 2024년08월16일

(30) 우선권주장

1020230172181 2023년12월01일 대한민국(KR)

(71) 출원인

이화여자대학교 산학협력단

서울특별시 서대문구 이화여대길 52 (대현동, 이화여자대학교)

(72) 발명자

오유란

서울특별시 마포구 만리재옛2길 14, 101동 501호 (신공덕동, 모던하우스)

김수현

인천광역시 연수구 컨벤시아대로130번길 32, 202동 2401호(송도동, 송도자이하버뷰2단지아파트)

이세민

서울특별시 마포구 포은로 143, 101동 305호(망원동, 뉴성신아파트)

(74) 대리인

김연권

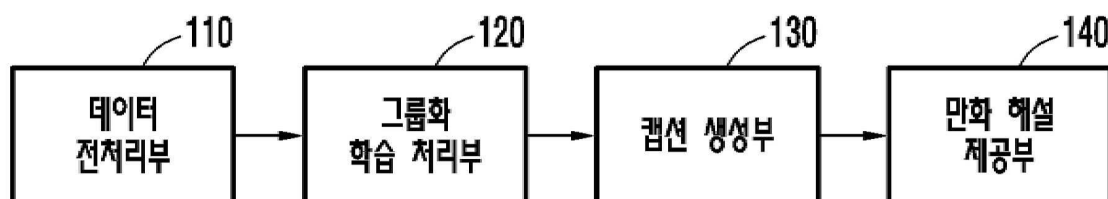
전체 청구항 수 : 총 11 항

(54) 발명의 명칭 만화 이미지들의 그룹화를 통해 만화 해설을 제공하기 위한 학습 모델 및 학습 모델 기반의 만화 해설 제공시스템

(57) 요약

본 발명은 만화 이미지들의 그룹화를 통해 만화 해설을 제공하는 만화 해설 제공장치 및 방법에 관한 것으로, 본 발명의 일실시예에 따른 만화 해설 제공장치는 만화 텍스트들과 만화 이미지들의 관련성을 학습하여 그룹화 기준을 학습하고, 학습된 그룹화 기준에 기반하여 만화 이미지들을 그룹화하여 불필요한 반복을 줄이면서 해당 상황에 대하여 명확한 만화 해설을 제공하는 기술에 관한 것이다.

대표도 - 도1

100

(52) CPC특허분류

G06T 11/60 (2013.01)

G06V 20/46 (2022.01)

G10L 13/08 (2013.01)

G10L 15/26 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711193328
과제번호	2020-0-01460-004
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성
연구과제명	Simulated Reality 콘텐츠 구현을 위한 핵심기술 연구 및 인력양성(센터명:시물레이
선기반 융복합 콘텐츠 연구센터)	
과제수행기관명	이화여자대학교산학협력단
연구기간	2020.07.01 ~ 2027.12.31

명세서

청구범위

청구항 1

비디오 데이터로부터 추출된 이미지 패널 및 텍스트를 포함하는 제1 학습데이터셋, 만화를 구성하는 만화 이미지 패널 및 만화 이미지 패널에서 추출된 텍스트를 포함하는 제2 학습데이터셋 및 만화 이미지 패널의 기 설정된 구간의 반복을 기초로 생성된 비디오화 데이터로부터 추출된 이미지 패널 및 텍스트를 포함하는 제3 학습데이터셋 중 적어도 어느 하나의 학습데이터셋을 수신하고, 수신된 학습데이터셋을 학습 모델에 입력하기 위한 기 설정된 형식의 이미지 변환데이터 및 텍스트 변환 데이터로 전처리하는 데이터 전처리부;

상기 기 설정된 형식의 이미지 변환데이터 및 텍스트 변환 데이터에서 추출된 특징에 기초하여 이미지 패널의 그룹화 기준을 학습하고, 학습된 그룹화 기준에 기초하여 이미지 패널의 그룹화를 처리하는 그룹화 학습처리부; 및

상기 이미지 패널의 그룹화를 통해 생성된 그룹화된 이미지 패널들에 대한 그룹 캡션을 생성하고, 학습된 그룹 캡션 생성 모델을 생성하는 캡션 생성부를 포함하는

학습 모델기반의 만화 해설 제공시스템.

청구항 2

제1항에 있어서,

상기 그룹화 학습 처리부는 이미지 패널에 대하여 입력된 추가 설명에 대한 추가 설명 텍스트들, 이미지 패널에 기반하여 생성되는 묘사 텍스트들과 이미지 패널에서 추출된 텍스트들 간의 관련성을 학습하고, 텍스트들 간의 중복성에 기반하여 그룹화 점수를 산출하고, 상기 산출된 그룹화 점수가 기 설정된 그룹화 기준을 초과하는지 여부에 기반하여 상기 이미지 패널의 그룹화를 처리하는 것을 특징으로 하는

학습 모델기반의 만화 해설 제공시스템.

청구항 3

제1항에 있어서,

상기 데이터 전처리부는

상기 비디오 데이터로부터 N초(N은 1 이상의 정수) 당 1 프레임을 추출하고, 추출된 프레임을 이미지 패널로 선택하고, 추출된 프레임에 포함된 오디오 신호로부터 텍스트를 추출하고,

상기 기 설정된 구간은 최종 학습된 만화 해설을 제공하기 위한 학습 모델의 성능 평가로부터 결정된 M초(M은 N보다 큰 정수)로 설정하는

학습 모델기반의 만화 해설 제공시스템.

청구항 4

제1항에 있어서,

상기 만화 해설 제공부는

상기 그룹 캡션에 대하여 텍스트를 제공하는 시각 정보를 생성하고, 상기 시각 정보에 대응하여 기계음, 합성음 및 효과음 중 적어도 하나를 포함하는 오디오를 청각 정보로 생성하는 만화 해설 제공부를 더 포함하는

학습 모델기반의 만화 해설 제공시스템.

청구항 5

제4항에 있어서,

상기 만화 해설 제공부는 상기 그룹 캡션과 연관된 적어도 하나의 이미지 패널에서 사용자의 몰입도를 위한 서로 다른 음성의 특징을 갖도록 상기 청각 정보를 생성하는 것을 특징으로 하는

학습 모델기반의 만화 해설 제공시스템.

청구항 6

제4항에 있어서,

상기 만화 해설 제공부는 상기 그룹 캡션에서 텍스트를 구성하는 단어, 문장 및 단락 중 적어도 하나를 고려하여 상기 시각 정보 및 상기 청각 정보가 제공되는 시간을 제어하는 것을 특징으로 하는

학습 모델기반의 만화 해설 제공시스템.

청구항 7

데이터 전처리부에서, 만화 이미지들과 만화 텍스트들로 구성된 이미지 데이터를 수신하고, 상기 수신된 이미지 데이터 내 상기 만화 이미지들과 상기 만화 텍스트들을 시간 순서로 정리하며, 상기 정리된 이미지 데이터에서 상기 만화 텍스트들과 상기 만화 이미지들을 상호 분리하는 단계;

그룹화 학습 처리부에서, 상기 만화 텍스트들과 상기 만화 이미지들의 관련성을 학습하여 그룹화 기준을 결정하고, 상기 결정된 그룹화 기준에 기반하여 상기 만화 이미지들을 그룹화하여 적어도 하나의 그룹 이미지를 생성하는 단계;

캡션 생성부에서, 상기 학습된 관련성에 기반하여 상기 생성된 적어도 하나의 그룹 이미지에 대하여 적어도 하나의 그룹 캡션을 생성하는 단계; 및

만화 해설 제공부에서, 상기 생성된 적어도 하나의 그룹 캡션에 대하여 시각 정보 및 청각 정보 중 적어도 하나의 정보로 만화 해설을 제공하는 단계를 포함하는 것을 특징으로 하는

만화 해설 제공방법.

청구항 8

제7항에 있어서,

상기 만화 텍스트들과 상기 만화 이미지들의 관련성을 학습하여 그룹화 기준을 결정하고, 상기 결정된 그룹화 기준에 기반하여 상기 만화 이미지들을 그룹화하여 적어도 하나의 그룹 이미지를 생성하는 단계는,

상기 만화 이미지들에 대하여 입력된 추가 설명에 대한 추가 설명 텍스트들, 상기 만화 이미지들에 기반하여 생성되는 묘사 텍스트들과 상기 만화 텍스트들 간의 관련성을 학습하여 텍스트들 간의 중복성에 기반하여 그룹화 점수를 산출하고, 상기 산출된 그룹화 점수가 상기 그룹화 기준에 대하여 초과하는지 여부에 기반하여 상기 적어도 하나의 그룹화 이미지를 생성하는 단계를 포함하는 것을 특징으로 하는

만화 해설 제공방법.

청구항 9

제7항에 있어서,

상기 학습된 관련성에 기반하여 상기 생성된 적어도 하나의 그룹 이미지에 대하여 적어도 하나의 그룹 캡션을 생성하는 단계는,

상기 만화 텍스트들 중 동일한 의미 또는 동일한 형태를 가지는 텍스트를 중복 텍스트로 결정하고, 상기 결정된 중복 텍스트가 등장하는 수를 카운팅하며, 상기 카운팅 수에 기반하여 결정된 텍스트로 구성된 상기 적어도 하나의 그룹 캡션을 생성하는 단계를 포함하는 것을 특징으로 하는

만화 해설 제공방법.

청구항 10

입력되는 복수의 만화 이미지 패널을 학습된 모델에 기초하여 그룹화하고, 그룹화된 이미지 패널들에 대한 그룹

캡션을 제공하는 그룹 캡션 제공부; 및

상기 그룹 캡션에 대하여 시각 정보 및 청각 정보 중 적어도 하나의 정보로 만화 해설을 제공하는 만화 해설 제공부를 포함하고,

상기 학습된 모델은 비디오 데이터로부터 추출된 이미지 패널 및 텍스트를 포함하는 제1 학습데이터셋, 만화를 구성하는 만화 이미지 패널 및 만화 이미지 패널에서 추출된 텍스트를 포함하는 제2 학습데이터셋 및 만화 이미지 패널의 기 설정된 구간의 반복을 기초로 생성된 비디오화 데이터로부터 추출된 이미지 패널 및 텍스트를 포함하는 제3 학습데이터셋 중 적어도 어느 하나의 학습데이터셋에 기초하여 학습된 모델인

학습 모델 기반의 만화 해설 제공시스템.

청구항 11

제10항에 있어서,

상기 만화 해설 제공부는

상기 그룹 캡션에 대하여 텍스트를 제공하는 시각 정보를 생성하고, 상기 시각 정보에 대응하여 기계음, 합성음 및 효과음 중 적어도 하나를 포함하는 오디오를 청각 정보로 생성하는

학습 모델 기반의 만화 해설 제공시스템.

발명의 설명

기술 분야

[0001] 본 발명은 만화 이미지들의 그룹화를 통해 만화 해설을 제공하기 위한 학습 모델 및 만화 해설 제공시스템에 관한 것으로, 만화 텍스트들과 만화 이미지들의 관련성을 학습하여 그룹화 기준을 학습하고, 학습된 그룹화 기준에 기반하여 만화 이미지들을 그룹화하여 불필요한 반복을 줄이면서 해당 상황에 대하여 명확한 만화 해설을 제공하는 기술에 관한 것이다.

배경 기술

[0002] 일반적으로, 만화 콘텐츠들은 전자 기술의 발전에 힘입어 컴퓨터 및 스마트폰과 같은 전자 기기에서 감상할 수 있다.

[0003] 만화는 텍스트와 이미지로 이야기(story)를 전달하는 예술 매체로, 이미지에 대한 설명이 별도로 존재하지 않는 이상 시각장애인들이 해당 매체를 즐기는 것이 어렵다. 사람이 직접 작성하는데, 이 경우 많은 시간과 노력이 소요된다.

[0004] 만화는 텍스트와 결합된 이미지를 통해 순차적으로 배열된 패널에 기반하여 하나의 연속된 흐름을 가진 이미지의 연속성을 가진다.

[0005] 이미지에 대한 기존의 텍스트 설명은 유사한 순차적 패널로 인해 만화에 적용될 때 그 효과가 미비하다.

[0006] 따라서, 개별 이미지에 대하여 이미지 캡션(image caption)을 생성하게 되면 비슷한 내용이 계속 반복되는 문제점이 존재한다.

선행기술문헌

특허문헌

[0007] (특허문헌 0001) 한국공개특허 제10-2023-0069402호, "사용자를 위한 오디오 코믹북 변환 방법, 오디오 코믹북 제공방법 및 상기 방법을 수행하는 사용자 단말 "

(특허문헌 0002) 한국공개특허 제10-2024-0069326호, "딥러닝을 이용한 만화책 편집 시스템 및 그 방법"

(특허문헌 0003) 한국공개특허 제10-2022-0065540호, "웹툰 및 만화의 말풍선 내 텍스트 추출 및 자동배치기술을 활용한 사용자 참여형 번역 및 자동번역시스템"

발명의 내용

해결하려는 과제

- [0008] 본 발명은 만화 텍스트들과 만화 이미지들의 관련성 및 그룹화 기준을 학습하고, 학습된 그룹화 기준에 기반하여 만화 이미지들을 그룹화하여 불필요한 반복을 줄이면서 해당 상황에 대하여 명확한 만화 해설을 제공하는 만화 해설 제공장치 및 방법을 제공하는 것을 목적으로 한다.
- [0009] 본 발명은 만화 장면에 대한 그룹 캡션(caption)을 자동으로 생성하여 만화에 대하여 대체 텍스트를 작성하는 시간과 인력의 소모를 감축시키는 것을 목적으로 한다.
- [0010] 본 발명은 그룹화된 이미지를 연관된 장면을 묶어 구성하고, 그룹화된 이미지들에 대한 그룹 캡션을 제공함에 따라 만화 이미지들의 불필요한 반복을 줄이면서 남녀노소, 장애 유무와 상관없이 모두가 동등하게 불편 없이 이용할 수 있는 만화 해설을 제공하는 것을 목적으로 한다.

과제의 해결 수단

- [0011] 본 발명의 일실시예에 따른 학습 모델기반의 만화 해설 제공시스템은 비디오 데이터로부터 추출된 이미지 패널 및 텍스트를 포함하는 제1 학습데이터셋, 만화를 구성하는 만화 이미지 패널 및 만화 이미지 패널에서 추출된 텍스트를 포함하는 제2 학습데이터셋 및 만화 이미지 패널의 기 설정된 구간의 반복을 기초로 생성된 비디오화 데이터로부터 추출된 이미지 패널 및 텍스트를 포함하는 제3 학습데이터셋 중 적어도 어느 하나의 학습데이터셋을 수신하고, 수신된 학습데이터셋을 학습 모델에 입력하기 위한 기 설정된 형식의 이미지 변환데이터 및 텍스트 변환 데이터로 전처리하는 데이터 전처리부와, 상기 기 설정된 형식의 이미지 변환데이터 및 텍스트 변환데이터에서 추출된 특징에 기초하여 이미지 패널의 그룹화 기준을 학습하고, 학습된 그룹화 기준에 기초하여 이미지 패널의 그룹화를 처리하는 그룹화 학습 처리부 및 상기 이미지 패널의 그룹화를 통해 생성된 그룹화된 이미지 패널들에 대한 그룹 캡션을 생성하고, 학습된 그룹 캡션 생성 모델을 생성하는 캡션 생성부를 포함한다.
- [0012] 상기 그룹화 학습 처리부는 이미지 패널에 대하여 입력된 추가 설명에 대한 추가 설명 텍스트들, 이미지 패널에 기반하여 생성되는 묘사 텍스트들과 이미지 패널에서 추출된 텍스트들 간의 관련성을 학습하고, 텍스트들 간의 중복성에 기반하여 그룹화 점수를 산출하고, 상기 산출된 그룹화 점수가 기 설정된 그룹화 기준을 초과하는지 여부에 기반하여 상기 이미지 패널의 그룹화를 처리할 수 있다.
- [0013] 상기 데이터 전처리부는 상기 비디오 데이터로부터 N초(N은 1 이상의 정수) 당 1 프레임을 추출하고, 추출된 프레임을 이미지 패널로 선택하고, 추출된 프레임에 포함된 오디오 신호로부터 텍스트를 추출하고, 상기 기 설정된 구간은 최종 학습된 만화 해설을 제공하기 위한 학습 모델의 성능 평가로부터 결정된 M초(M은 N보다 큰 정수)로 설정할 수 있다.
- [0014] 학습 모델 기반의 만화해설 제공 시스템은 상기 그룹 캡션에 대하여 텍스트를 제공하는 시각 정보를 생성하고, 상기 시각 정보에 대응하여 기계음, 합성음 및 효과음 중 적어도 하나를 포함하는 오디오를 청각 정보로 생성하는 만화 해설 제공부를 더 포함할 수 있다.
- [0015] 상기 만화 해설 제공부는 상기 그룹 캡션과 연관된 적어도 하나의 이미지 패널에서 사용자의 몰입도를 위한 서로 다른 음성의 특징을 갖도록 상기 청각 정보를 생성할 수 있다.
- [0016] 상기 만화 해설 제공부는 상기 그룹 캡션에서 텍스트를 구성하는 단어, 문장 및 단락 중 적어도 하나를 고려하여 상기 시각 정보 및 상기 청각 정보가 제공되는 시간을 제어할 수 있다.
- [0017] 본 발명의 일실시예에 따르면 만화 해설 제공방법은 데이터 전처리부에서, 만화 이미지들과 만화 텍스트들로 구성된 이미지 데이터를 수신하고, 상기 수신된 이미지 데이터 내 상기 만화 이미지들과 상기 만화 텍스트들을 시간 순서로 정리하며, 상기 정리된 이미지 데이터에서 상기 만화 텍스트들과 상기 만화 이미지들을 상호 분리하는 단계, 그룹화 학습 처리부에서, 상기 만화 텍스트들과 상기 만화 이미지들의 관련성을 학습하여 그룹화 기준을 결정하고, 상기 결정된 그룹화 기준에 기반하여 상기 만화 이미지들을 그룹화하여 적어도 하나의 그룹 이미지를 생성하는 단계, 캡션 생성부에서, 상기 학습된 관련성에 기반하여 상기 생성된 적어도 하나의 그룹 이미지에 대하여 적어도 하나의 그룹 캡션을 생성하는 단계 및 상기 생성된 적어도 하나의 그룹 캡션에 대하여 시각 정보 및 청각 정보 중 적어도 하나의 정보로 만화 해설을 제공하는 단계를 포함할 수 있다.
- [0018] 상기 만화 텍스트들과 상기 만화 이미지들의 관련성을 학습하여 그룹화 기준을 결정하고, 상기 결정된 그룹화

기준에 기반하여 상기 만화 이미지들을 그룹화하여 적어도 하나의 그룹 이미지를 생성하는 단계는, 상기 만화 이미지들에 대하여 입력된 주석 텍스트들과 상기 만화 이미지들에 기반하여 생성되는 묘사 텍스트들과 상기 만화 텍스트들 간의 관련성을 학습하여 텍스트들 간의 중복성에 기반하여 그룹화 점수를 산출하고, 상기 산출된 그룹화 점수가 상기 그룹화 기준에 대하여 초과하는지 여부에 기반하여 상기 적어도 하나의 그룹화 이미지를 생성하는 단계를 포함할 수 있다.

- [0019] 상기 학습된 관련성에 기반하여 상기 생성된 적어도 하나의 그룹 이미지에 대하여 적어도 하나의 그룹 캡션을 생성하는 단계는, 상기 만화 텍스트들 중 동일한 의미 또는 동일한 형태를 가지는 텍스트를 중복 텍스트로 결정하고, 상기 결정된 중복 텍스트가 등장하는 수를 카운팅하며, 상기 카운팅 수에 기반하여 결정된 텍스트로 구성된 상기 적어도 하나의 그룹 캡션을 생성하는 단계를 포함할 수 있다.

발명의 효과

- [0020] 본 발명은 만화 텍스트들과 만화 이미지들의 관련성을 학습하여 그룹화 기준을 결정하고, 결정된 그룹화 기준에 기반하여 만화 이미지들을 그룹화하여 불필요한 반복을 줄이면서 해당 상황에 대하여 명확한 만화 해설을 제공하는 만화 해설 제공장치 및 방법을 제공할 수 있다.
- [0021] 본 발명은 만화 장면에 대한 그룹 캡션(caption)을 자동으로 생성하여 만화에 대하여 대체 텍스트를 작성하는 시간과 인력의 소모를 감축시킬 수 있다.
- [0022] 본 발명은 연관된 장면을 묶어 그룹화된 이미지를 구성하고, 그룹화된 이미지들에 대한 그룹 캡션을 제공함에 따라 만화 이미지들의 불필요한 반복을 줄이면서 남녀노소, 장애 유무와 상관없이 모두가 동등하게 불편 없이 이용할 수 있는 만화 해설을 제공할 수 있다.

도면의 간단한 설명

- [0023] 도 1 및 도 2는 본 발명의 일실시예에 따른 만화 해설 제공을 위한 학습 모델의 구축 및 학습 모델 기반의 만화 해설 제공 시스템을 설명하는 도면이다.
- 도 3은 본 발명의 일실시예에 따른 만화 이미지 그룹화 및 그룹 캡션을 제공하는 절차 또는 학습 과정을 설명하는 도면이다.
- 도 4는 본 발명의 일실시예에 따른 그룹 캡션을 위한 보충 데이터를 이용하는 구성 또는 학습 과정을 설명하는 도면이다.
- 도 5a 및 도 5b는 본 발명의 일실시예에 따른 만화 해설 제공방법에서 그룹 이미지의 생성 기준을 설명하는 도면이다.
- 도 6은 본 발명의 일실시예에 따른 만화 해설 제공방법을 설명하는 도면이다.
- 도 7은 본 발명의 실시예에 따른 학습 모델 기반의 만화 해설 제공 시스템을 설명하는 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0024] 이하, 본 문서의 다양한 실시 예들이 첨부된 도면을 참조하여 기재된다.
- [0025] 실시 예 및 이에 사용된 용어들은 본 문서에 기재된 기술을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 해당 실시 예의 다양한 변경, 균등물, 및/또는 대체물을 포함하는 것으로 이해되어야 한다.
- [0026] 하기에서 다양한 실시 예들을 설명에 있어 관련된 공지 기능 또는 구성에 대한 구체적인 설명이 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략할 것이다.
- [0027] 그리고 후술되는 용어들은 다양한 실시 예들에서의 기능을 고려하여 정의된 용어들로서 이는 사용자, 운용자의 의도 또는 관례 등에 따라 달라질 수 있다. 그러므로 그 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다.
- [0028] 도면의 설명과 관련하여, 유사한 구성요소에 대해서는 유사한 참조 부호가 사용될 수 있다.
- [0029] 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함할 수 있다.
- [0030] 본 문서에서, "A 또는 B" 또는 "A 및/또는 B 중 적어도 하나" 등의 표현은 함께 나열된 항목들의 모든 가능한

조합을 포함할 수 있다.

- [0031] "제1," "제2," "첫째," 또는 "둘째," 등의 표현들은 해당 구성요소들을, 순서 또는 중요도에 상관없이 수식할 수 있고, 한 구성요소를 다른 구성요소와 구분하기 위해 사용될 뿐 해당 구성요소들을 한정하지 않는다.
- [0032] 어떤(예: 제1) 구성요소가 다른(예: 제2) 구성요소에 "(기능적으로 또는 통신적으로) 연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 상기 어떤 구성요소가 상기 다른 구성요소에 직접적으로 연결되거나, 다른 구성요소(예: 제3 구성요소)를 통하여 연결될 수 있다.
- [0033] 본 명세서에서, "~하도록 구성된(또는 설정된)(configured to)"은 상황에 따라, 예를 들면, 하드웨어적 또는 소프트웨어적으로 "~에 적합한," "~하는 능력을 가지는," "~하도록 변경된," "~하도록 만들어진," "~를 할 수 있는," 또는 "~하도록 설계된"과 상호 호환적으로(interchangeably) 사용될 수 있다.
- [0034] 어떤 상황에서는, "~하도록 구성된 장치"라는 표현은, 그 장치가 다른 장치 또는 부품들과 함께 "~할 수 있는" 것을 의미할 수 있다.
- [0035] 예를 들면, 문구 "A, B, 및 C를 수행하도록 구성된(또는 설정된) 프로세서"는 해당 동작을 수행하기 위한 전용 프로세서(예: 임베디드 프로세서), 또는 메모리 장치에 저장된 하나 이상의 소프트웨어 프로그램들을 실행함으로써, 해당 동작들을 수행할 수 있는 범용 프로세서(예: CPU 또는 application processor)를 의미할 수 있다.
- [0036] 또한, '또는'이라는 용어는 배타적 논리합 'exclusive or' 이기보다는 포함적인 논리합 'inclusive or' 를 의미한다.
- [0037] 즉, 달리 언급되지 않는 한 또는 문맥으로부터 명확하지 않는 한, 'x가 a 또는 b를 이용한다'라는 표현은 포함적인 자연 순열들(natural inclusive permutations) 중 어느 하나를 의미한다.
- [0038] 이하 사용되는 '..부', '..기' 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어, 또는, 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.
- [0040] 도 1 및 도 2는 본 발명의 일실시예에 따른 만화 해설 제공을 위한 학습 모델의 구축 및 학습 모델 기반의 만화 해설 제공 시스템을 설명하는 도면이다.
- [0041] 도 1은 본 발명의 일실시예에 따른 만화 해설 제공을 위한 학습 모델의 구축장치의 구성 요소를 예시한다.
- [0042] 도 1을 참고하면, 본 발명의 일실시예에 따른 학습 모델의 구축 장치(100)는 데이터 전처리부(110), 그룹화 학습 처리부(120), 캡션 생성부(130) 및 만화 해설 제공부(140)를 포함할 수 있다.
- [0043] 본 발명의 일실시예에 따른 만화 해설 제공장치(100)는 개별 만화 이미지에 대하여 캡션을 생성하면 비슷한 내용이 반복적으로 생성되는 문제점을 해결하기 위해서 유사도가 높은 만화 이미지 장면들을 그룹화하고 이에 대한 그룹 캡션을 생성하여 불필요한 반복이 없고, 해당 상황에 대한 설명을 생성 및 제공할 수 있다.
- [0044] 이하, 본 명세서에서 개별 만화 이미지는 '만화 이미지 패널'로 칭해질 수 있다. '만화 이미지 패널'은 만화의 한 컷에 해당하는 이미지를 의미할 수 있다. '만화 이미지 패널'은 의성어, 의태어, 말 풍선 등의 텍스트를 포함할 수 있다. 또한, 본 명세서에서 '만화 이미지 패널'과 함께 '이미지 패널'이란 용어를 사용할 수 있다. '이미지 패널'은 만화와 같이 하나의 컷 또는 장면으로 간주할 수 있는 이미지 데이터를 의미하거나 '만화 이미지 패널'을 포함하는 개념으로 사용될 수 있다. 예를 들어, 비디오 데이터에서 추출된 하나의 프레임은 설명의 편의상 '이미지 패널'로 칭할 수 있다. 한편, 본 명세서에서 '만화 데이터'는 만화 이미지 패널과 만화 이미지 패널로부터 추출된 텍스트 데이터를 모두 포괄하는 의미로 사용될 수 있다.
- [0045] 복수의 만화 이미지 패널을 그룹화하기 위한 학습데이터셋은 학습 모델 구축을 위해 충분하게 제공되지 않을 수 있다. 따라서, 본 발명의 실시예에 따른 학습 모델 구축은 비디오 데이터를 이용하여 학습 모델을 구축할 수 있다.
- [0046] 비디오 데이터는 만화의 특성을 포함하지 않을 수 있다. 따라서, 비디오 데이터와 함께 만화 이미지 패널을 학습 데이터셋으로 사용할 수도 있다. 한편, 학습 모델 자체가 비디오 데이터의 'Dense Video Captioning'을 위한 모델인 경우 만화 이미지 패널을 변형 없이 학습 데이터셋으로 사용하기에 적합하지 않을 수 있다. 이러한 문제를 해소하고자, 본 발명의 실시예에 따른 학습 모델의 구축 장치(100)는 만화 이미지 패널을 시간 순서로 정리하고, 비디오 데이터와의 특성 차이를 최소화하기 위해 기 설정된 구간동안 만화 이미지 패널을 반복 또는

유지하는 비디오화된 데이터를 생성하고, 만화 이미지 패널의 기 설정된 구간의 반복을 기초로 생성된 비디오화 데이터로부터 학습 데이터셋을 추출하는 방법을 함께 사용할 수 있다.

- [0047] 예를 들어, 비디오화된 데이터는 5개의 연속된 1번에서 5번의 만화 이미지 패널이 존재할 때, 1번~5번 만화 이미지 패널 각각을 특정 기간(duration) 동안 유지하는 방식으로 생성될 수 있다. 이때, 기 설정된 구간은 최종 구축된 모델의 성능 평가를 통해 결정될 수 있다.
- [0048] 본 발명의 실시예에 따른 학습 데이터셋은 비디오 데이터와 만화 이미지 패널의 특성 차이를 최소화한 학습데이터셋으로 구성될 수 있다. 이때, 비디오 데이터로부터 N초(N은 1 이상의 정수) 당 1 프레임을 추출하고, 추출된 프레임을 이미지 패널로 선택하고, 추출된 프레임에 포함된 오디오 신호로부터 텍스트를 추출하고, 상기 기 설정된 구간은 최종 학습된 만화 해설을 제공하기 위한 학습 모델의 성능 평가로부터 결정된 M초(M은 N보다 큰 정수)로 설정될 수 있다. 예를 들어, N은 1이고, M은 4일 수 있다.
- [0049] 이때, 학습 모델의 성능 평가는 BLEU(Bilingual Evaluation Understudy) 점수 및 METEOR(Metric for Evaluation of Translation with Explicit Ordering) 점수 중 어느 하나의 점수에서 높은 점수로 결정할 수 있다.
- [0050] BLEU 점수는 기계 번역 결과와 사람이 직접 번역한 결과가 얼마나 유사한지 비교하여 번역에 대한 성능을 측정한 결과일 수 있다.
- [0051] METEOR 점수는 BLEU 점수의 단점을 보완하여 성능을 측정하는 방법일 수 있다.
- [0052] METEOR 점수는 기계 번역 및 자연어 처리 작업에서 사용되는 다목적 평가 메트릭이다.
- [0053] 단어 순서와 어휘의 다양성을 고려하고, 정확성, 충실도, 순서 정확성 및 어휘 다양성을 평가하여 종합적인 평가 메트릭을 제공한다.
- [0054] 데이터 전처리부(110)은 비디오 데이터로부터 추출된 이미지 패널 및 텍스트를 포함하는 제1 학습데이터셋, 만화를 구성하는 만화 이미지 패널 및 만화 이미지 패널에서 추출된 텍스트를 포함하는 제2 학습데이터셋 및 만화 이미지 패널의 기 설정된 구간의 반복을 기초로 생성된 비디오화 데이터로부터 추출된 이미지 패널 및 텍스트를 포함하는 제3 학습데이터셋 중 적어도 어느 하나의 학습데이터셋을 수신하고, 수신된 학습데이터셋을 학습 모델에 입력하기 위한 기 설정된 형식의 이미지 변환데이터 및 텍스트 변환 데이터로 변환할 수 있다.
- [0055] 여기서, 제1 학습데이터셋 내지 제3 학습데이터셋에 포함되는 텍스트는 예를 들어, 이미지 패널 내에 말풍선 형태의 텍스트, 주석(annotation)에 해당하는 텍스트, 사용자에게 의해 생성된 텍스트 또는 장면을 추가적으로 설명하거나 묘사하는 텍스트 중 적어도 어느 하나를 포함할 수 있다.
- [0056] 이미지 변환데이터와 텍스트 변환데이터는 자연수로 셀 수 있는 형태이며, 1:1 혹은 1:다 형태의 쌍으로 대응될 수 있다. 예를 들어, 복수의 이미지 변환데이터와 하나의 텍스트 변환데이터가 서로 대응될 수 있으며, 대응 정보는 이미지 변환데이터 및 텍스트 변환데이터 중 어느 하나에 포함된다.
- [0057] 그룹화 학습 처리부(120)는 기 설정된 형식의 이미지 변환데이터 및 텍스트 변환 데이터에서 추출된 특징에 기초하여 이미지 패널의 그룹화 기준을 학습하고, 학습된 그룹화 기준에 기초하여 이미지 패널의 그룹화를 처리한다.
- [0058] 예를 들어, 그룹화 학습 처리부(120)는 복수의 이미지 변환 데이터들 간의 관련성(혹은 유사도), 복수의 텍스트 변환 데이터들 간의 관련성(혹은 중복성) 중 적어도 하나에 기초하여 그룹화를 위한 학습, 또는 그룹화를 진행할 수 있다.
- [0059] 다른 예를 들어, 그룹화 학습 처리부(120)는 만화 이미지들에 대하여 입력된 주석 텍스트들과 만화 이미지들에 기반하여 생성되는 묘사 텍스트들과 만화 텍스트들 간의 관련성을 학습하여 텍스트들 간의 중복성에 기반하여 그룹화 점수를 산출하고, 산출된 그룹화 점수가 그룹화 기준에 대하여 초과하는지 여부에 기반하여 그룹화를 수행할 수 있다.
- [0060] 한편, 그룹화 학습 처리부(120)는 상술한 기 설정된 형식의 이미지 변환데이터 및 텍스트 변환데이터에서 추출된 특징에 기초하여 그룹화된 이미지들에 대한 그룹 이미지를 생성할 수 있다.
- [0061] 여기서, '그룹 이미지'는 그룹화된 복수의 이미지들을 대표하는 이미지일 수 있다. 예를 들어, 만화 이미지 패널들이 제1 이미지 내지 제5 이미지로 구성된 경우에 관련성이 있는 제1 이미지 내지 제3 이미지에 대하여 그룹

화된 경우, 그룹화된 제1 내지 제3 이미지에 대한 그룹 이미지가 결정될 수 있다. 이 경우, 그룹이미지는 제1 내지 제3 이미지 중 어느 하나일 수도 있고, 제1 내지 제3 이미지에 기초하여 새롭게 생성(예를 들어, 제1 내지 제3 이미지를 썸네일과 같이 축소하고, 축소된 썸네일을 병합하여 하나의 이미지를 생성)될 수도 있다.

- [0062] 캡션 생성부(130)는 이미지 패널의 그룹화를 통해 생성된 그룹화된 이미지 패널들에 대한 그룹 캡션을 생성하는, 그룹 캡션 생성 모델을 학습시킬 수 있다.
- [0063] 캡션 생성부(130)는 그룹화를 진행한 후 생성된 그룹화된 이미지들의 특징을 추출하고, 추출된 특징에 기초하여 캡션을 생성할 수 있고, 이때 생성형 언어 모델을 이용하여 캡션을 완성할 수 있다.
- [0064] 또한, 캡션 생성부(130)는 기 학습된 관련성에 기반하여 기 생성된 적어도 하나의 그룹 이미지에 대하여 적어도 하나의 그룹 캡션을 생성할 수 있다.
- [0065] 본 발명의 일실시예에 따르면 만화 해설 제공부(140)는 기 생성된 적어도 하나의 그룹 캡션에 대하여 시각 정보 및 청각 정보 중 적어도 하나의 정보로 만화 해설을 제공할 수 있다.
- [0066] 예를 들어, 만화 해설 제공부(140)는 적어도 하나의 그룹 캡션에 대하여 텍스트를 제공하는 시각 정보를 생성하고, 텍스트에 대응하여 기계음, 합성음 및 효과음 중 적어도 하나를 포함하는 오디오를 상기 청각 정보로 생성할 수 있다.
- [0067] 일례로, 만화 해설 제공부(140)는 적어도 하나의 그룹 캡션과 연관된 적어도 하나의 그룹 이미지에서 사용자의 몰입도를 위한 서로 다른 음성의 특징을 갖도록 청각 정보를 생성할 수 있다.
- [0068] 본 발명의 일실시예에 따르면 만화 해설 제공부(140)는 적어도 하나의 그룹 캡션에서 텍스트를 구성하는 단어, 문장 및 단락 중 적어도 하나를 고려하여 상기 시각 정보 및 상기 청각 정보가 제공되는 시간을 제어할 수 있다.
- [0069] 본 발명의 일실시예에 따른 만화 해설 제공장치(100)는 연속된 흐름을 가진 만화 이미지들에 대하여 연관된 장면을 묶어서 그룹 캡션을 생성하기 때문에 불필요한 반복 없이 정리된 설명을 만화해설로 제공하여 보는 사람과 듣는 사람으로 하여금 지루함 없이 만화를 즐기는데 도움을 준다.
- [0070] 따라서, 본 발명은 만화 텍스트들과 만화 이미지들의 관련성을 학습하여 그룹화 기준을 결정하고, 결정된 그룹화 기준에 기반하여 만화 이미지들을 그룹화하여 불필요한 반복을 줄이면서 해당 상황에 대하여 명확한 만화 해설을 제공하는 만화 해설 제공장치 및 방법을 제공할 수 있다.
- [0071] 도 2는 본 발명의 일실시예에 따른 만화 해설 제공장치의 AI(artificial intelligence) 기반 그룹 캡션을 생성하는 구조를 보충 설명한다.
- [0072] 도 2를 참고하면, 본 발명의 일실시예에 따른 학습 모델 기반의 만화 해설 제공시스템 (200)은 데이터 전처리부, 그룹화 학습 처리부 및 캡션 생성부로 구성될 수 있다.
- [0073] 데이터 전처리부는 이미지 패널(201)과 이미지 패널에 대응하는 텍스트(203)에 대하여 데이터 전처리를 수행한다.
- [0074] 이때, 이미지 패널(201)은 제1 학습데이터셋의 이미지 패널, 제2 학습 데이터셋의 만화 이미지 패널, 제3 학습 데이터셋의 이미지 패널 중 적어도 어느 하나일 수 있다. 또한, 텍스트(203)는 제1 학습데이터셋의 텍스트, 제2 학습데이터 셋의 텍스트 및 제3 학습 데이터 셋의 텍스트 중 적어도 어느 하나일 수 있다.
- [0075] 데이터 전처리부는 이미지 패널(201)을 학습 모델에 입력하기 위한 기 설정된 형식의 이미지 변환 데이터(202)로 변환한다.
- [0076] 데이터 전처리부는 이미지 패널에 관련된 텍스트(203)를 학습 모델에 입력하기 위한 기 설정된 형식의 텍스트 변환 데이터(204)로 변환한다.
- [0077] 이때, 이미지 변환 데이터(202)는 예를 들어, 특정 사이즈의 특징 벡터이거나 널(Null) 데이터가 포함된 규격화된 데이터일 수 있다.
- [0078] 이때, 텍스트(203)는 이미지 패널(201)로부터 추출되는 텍스트를 나타낼 수 있고, 텍스트는 말풍선 내 작성된 텍스트, 효과를 나타내는 텍스트를 포함할 수 있다.
- [0079] 그룹화 학습 처리부는 인코딩부(210)와 이미지 그룹화부(220)를 포함하고, 인코딩부(210)는 셀프 어텐션층(self

attention layer), 바이모달(bi-modal) 어텐션층 및 연결층(fully-connected layer)이 만화 이미지와 만화 텍스트 각각에 대하여 이용되어서 인코딩을 지원한다.

- [0080] 인코딩부(210)는 텍스트 변환 데이터(206)에 대한 학습 결과를 캡션 생성부(230)의 디코딩부와 공유할 수 있다.
- [0081] 이미지 그룹화부(220)는 셀프 어텐션층(self attention layer), 바이모달(bi-modal) 어텐션층 및 연결층(fully-connected layer)로부터 출력되는 데이터를 헤드(head)층과 공통풀층(common pool)과 분류 및 선택(sort and select)층으로 학습 데이터를 처리하여 이미지 그룹화를 학습하고, 이미지 그룹화를 처리할 수 있다. 이때, 이미지 그룹화의 학습은 몇 개의 이미지 패널을 그룹화할 것인지, 또는 어느 이미지 패널들을 그룹화할 것인지를 포함할 수 있다.
- [0082] 예를 들어, 이미지의 그룹화는 이미지 패널(201)를 구성하는 복수의 만화 이미지를 학습 결과에 기반하여 수행될 수 있다.
- [0083] 한편, 이미지 그룹화부(220)를 통해 출력된 결과는 라벨링된 정답데이터와 비교될 수 있고, 비교 결과는 학습을 위한 파라미터로 활용될 수 있다.
- [0084] 캡션 생성부(230)는 디코딩부와 생성부로 구성되어서 캡션을 입력 받으면 그룹화 학습 처리부와 연동하여서 캡션을 생성하기 위한 생성부가 제1 그룹 캡션(231)을 생성한다.
- [0085] 캡션 생성부(230)는 반복적인 학습 및 언어 추가를 통해서 제1 그룹 캡션(231)을 기초로 제2 캡션(205)을 생성할 수 있고, 제2 캡션은 추가적인 학습을 위한 학습 데이터셋 형식(206)으로 변환될 수 있다. 다시 말해서, 캡션 생성부(230)는 생성형 언어 모델을 통해서 첫 번째 생성한 캡션(231)에 적합한 단어를 추가하고 인코딩부(210)로부터 제공되는 텍스트 변환 데이터(206)에 대한 학습 결과를 함께 학습하여 최종적인 그룹 캡션을 생성할 수 있다. 이때, 캡션 생성부(230)의 최종적인 그룹 캡션은 라벨링된 정답데이터와 비교될 수 있고, 비교 결과는 학습을 위한 파라미터로 활용될 수 있다.
- [0086] 디코딩부는 셀프 어텐션층(self attention layer), 바이모달(bi-modal) 어텐션층, 브릿지층(bridge layer) 및 연결층(fully-connected layer)로 구성된다.
- [0087] 캡션을 생성하기 위한 생성부는 연결층(fully-connected layer)으로 디코딩부와 연결되어서 소프트 맥스 함수를 이용하여 점수에 기반하여 그룹 캡션을 결정하여 생성 및 출력할 수 있다.
- [0089] 도 3은 본 발명의 일실시예에 따른 만화 이미지 그룹화 및 그룹 캡션을 제공하는 절차 또는 학습 과정을 설명하는 도면이다.
- [0090] 도 3에 도시된 예시는 본 발명의 실시예에 따라 학습된 모델에 의해 그룹 캡션을 제공하는 결과물 또는 학습 과정의 예시일 수 있다.
- [0091] 도 3을 참고하면, 본 발명의 일실시예에 따른 만화 해설 제공방법은 이미지 데이터(300)에서 시간 순서(sequence)에 따라서 만화 이미지들을 클리핑(clipping)한다.
- [0092] 만화의 한장면은 만화 이미지 패널과 만화 텍스트로 구성되고, 이미지 데이터(300)로부터 추출되는 정렬 이미지(310)에는 만화 이미지 패널과 만화 텍스트가 결합되어 있다. 또한, 정렬 이미지(310)는 학습 대상이 되는 복수의 만화 이미지가 시간 순서로 정렬되어서 보여진다.
- [0093] 만화 이미지 패널과 만화 텍스트는 이미지 데이터(300)가 시간 순서에 따라서 그룹화되는데 활용될 수 있다.
- [0094] 정렬 이미지(310)에서는 시간 순서로 제1 이미지(1) 내지 제5 이미지(5)가 예시된다. 제1 내지 2 이미지가 그룹화된 경우, 해당 그룹화된 이미지들에 대한 그룹 캡션으로서 제1 그룹 캡션(320)이 제공되는 예가 도시된다. 또한 제1 내지 5 이미지가 그룹화된 경우, 해당 그룹화된 이미지들에 대한 그룹 캡션으로서 제2 그룹 캡션(321)이 제공되는 예가 도시된다.
- [0095] 예를 들어, 제1 그룹 캡션(320) 및 제2 그룹 캡션(321)은 도 2에서 설명된 그룹 캡션 생성부에 의해서 생성되는 결과일 수 있다.
- [0096] 일 실시예에서 그룹화된 이미지는 제1 이미지(1) 내지 제5 이미지(5) 중에서 학습 결과에 기반하여 시간 순서에 기반하여 묶음 형태의 이미지로 나타날 수 있다.

- [0097] 또한, 제1 그룹 캡션(320)에서 제공하는 "(1-2)"와 같이 정렬 이미지(310)에 기반하여 그룹화를 수행할 수도 있고, "(1-2)"에 기반하여 정렬 이미지(310) 중에서 제1 이미지(1)와 제2 이미지(2)가 그룹화된 이미지가 그룹화 이미지로 해석될 수 있다.
- [0098] 제1 그룹 캡션(320)에서 텍스트 문장은 제1 이미지(1)와 제2 이미지(2)의 그룹화에 대한 설명일 수 있다.
- [0099] 또한, 제2 그룹 캡션(321)에서 제공하는 "(1-5)"와 같이 정렬 이미지(310)에 기반하여 그룹화를 수행할 수도 있고, "(1-5)"에 기반하여 정렬 이미지(310) 중에서 제1 이미지(1) 내지 제5 이미지(5)가 그룹화된 이미지가 그룹화 이미지로 해석될 수 있다.
- [0100] 제2 그룹 캡션(321)에서 텍스트 문장은 제1 이미지(1) 내지 제5 이미지(5)의 그룹화에 대한 설명일 수 있다.
- [0102] 도 4는 본 발명의 일실시예에 따른 그룹 캡션을 위한 보충 데이터를 이용하는 구성 또는 학습 과정을 설명하는 도면이다.
- [0103] 도 4를 참고하면, 본 발명의 일실시예에 따른 만화 해설 제공방법에서 그룹 캡션을 위한 보충 데이터를 이용하여 추가 설명 작성 화면(400)을 예시한다.
- [0104] 추가 설명 작성 화면(400)은 만화 이미지(402)에 대한 설명에 대하여 사용자가 지시사항(401)을 따라 추가 설명(403)을 추가적으로 작성하여 그룹 캡션이 만화 이미지에 대하여 제공하는 설명의 정확도를 증가시키는 구성으로 활용될 수 있음을 보여준다.
- [0105] 예를 들어, 만화 이미지(402) 및 추가 설명(403)은 학습을 위한 입력 데이터 또는 정답 데이터로 활용될 수 있다.
- [0106] 예를 들어, 그룹화 학습 처리부는 만화 이미지들에 대하여 입력된 추가 설명(403)에 대한 텍스트와 만화 이미지들에 기반하여 생성되는 텍스트와 만화 텍스트 간의 관련성을 학습하여 텍스트 간의 중복성에 기반하여 그룹화 점수를 산출하고, 산출된 그룹화 점수가 그룹화 기준에 대하여 초과하는지 여부에 기반하여 그룹화를 수행할 수 있다.
- [0107] 예를 들어, 추가 설명(403)은 도 1에 대한 설명에서 기재된 주석 텍스트의 예시일 수 있다.
- [0108] 예를 들어, "집"이라는 텍스트가 추가 설명, 장면을 묘사하는 텍스트 및 만화 텍스트에 동시에 존재하는 경우에 중복성이 인정되어서 중복성에 대한 그룹화 점수가 카운팅될 수 있다.
- [0110] 도 5a 및 도 5b는 본 발명의 일실시예에 따른 만화 해설 제공방법에서 그룹 캡션 및/또는 그룹 이미지의 생성 기준을 설명하는 도면이다.
- [0111] 도 5a는 본 발명의 일실시예에 따른 만화 해설 제공방법에서 그룹 캡션 및/또는 그룹 이미지의 생성 기준으로서 밀도 조건(dense)을 긍정적으로 판단하기 위한 조건을 예시한다.
- [0112] 도 5a를 참고하면, 그래프(500)는 간결함(conciseness), 명확한 설명(clear explanation), 원활한 연결성(smooth connectivity) 등이 확인됨을 예시한다.
- [0113] 도 5b는 본 발명의 일실시예에 따른 만화 해설 제공방법에서 그룹 캡션 및/또는 그룹 이미지의 생성 기준으로서 기준선(baseline)을 부정적으로 판단하기 위한 조건을 예시한다.
- [0114] 도 5b를 참고하면, 그래프(510)는 반복 문장(repetitive sentence), 불충분한 설명(insufficient description), 많은 설명(too much explanation) 및 연결성 부족(lack of connectivity) 등이 고려될 수 있음을 보여준다.
- [0115] 추가적으로, 사용자 설정에 기반한 사용자 평가가 그룹 캡션 및/또는 그룹 이미지 생성에 학습 내용으로 추가될 수 있다.
- [0116] 캐릭터의 외모와 감정 상태, 장면이 변경될 때 위치에 대한 간략한 언급과 시각화하기 쉬운 요소등이 고려될 수 있다.
- [0117] 다시 말해, 본 발명의 일실시예에 따른 만화 해설 제공방법 밀도(dense)가 높은 캡션 설명으로 그룹 캡션을 생

성하여 간결함, 문장연결성 및 반복적이지 않은 그룹 캡션을 생성하여 시각적 정보와 청각적 정보로 고밀도 만화 이미지 캡션을 그룹 캡션으로 제공할 수 있다.

- [0119] 도 6은 본 발명의 일실시예에 따른 만화 해설 제공방법을 설명하는 도면이다.
- [0120] 도 6은 본 발명의 일실시예에 따른 만화 해설 제공방법이 관련 장면을 묶어 만화 해설을 생성 및 제공하는 절차를 예시한다.
- [0121] 도 6을 참고하면, 단계(S601)에서 본 발명의 일실시예에 따른 만화 해설 제공방법은 만화 데이터를 수신 및 전처리한다.
- [0122] 즉, 본 발명의 일실시예에 따른 만화 해설 제공방법은 만화 이미지들과 만화 텍스트들로 구성된 이미지 데이터를 수신하고, 수신된 이미지 데이터 내 만화 이미지들과 만화 텍스트들을 시간 순서로 정리하며, 정리된 이미지 데이터에서 만화 텍스트들과 만화 이미지들을 상호 분리할 수 있다.
- [0123] 단계(S602)에서 본 발명의 일실시예에 따른 만화 해설 제공방법은 그룹화 기준 결정 및 그룹 이미지를 생성한다.
- [0124] 즉, 본 발명의 일실시예에 따른 만화 해설 제공방법은 만화 텍스트들과 만화 이미지들의 관련성을 학습하여 그룹화 기준을 결정하고, 결정된 그룹화 기준에 기반하여 만화 이미지들을 그룹화할 수 있다.
- [0125] 단계(S603)에서 본 발명의 일실시예에 따른 만화 해설 제공방법은 적어도 하나의 그룹 캡션을 생성한다.
- [0126] 즉, 본 발명의 일실시예에 따른 만화 해설 제공방법은 학습된 관련성에 기반하여 적어도 하나의 그룹 캡션을 생성할 수 있다.
- [0127] 단계(S604)에서 본 발명의 일실시예에 따른 만화 해설 제공방법은 그룹 캡션에 기반한 만화 해설을 제공한다.
- [0128] 즉, 본 발명의 일실시예에 따른 만화 해설 제공방법은 생성된 적어도 하나의 그룹 캡션에 대하여 시각 정보 및 청각 정보 중 적어도 하나의 정보로 만화 해설을 제공할 수 있다.
- [0129] 따라서, 본 발명은 연관된 장면을 묶어 그룹화를 구성하고, 그룹화된 이미지들에 대한 그룹 캡션을 제공함에 따라 만화 이미지들의 불필요한 반복을 줄이면서 남녀노소, 장애 유무와 상관없이 모두가 동등하게 불편 없이 이용할 수 있는 만화 해설을 제공할 수 있다.
- [0131] 도 7은 본 발명의 실시예에 따른 학습 모델 기반의 만화 해설 제공 시스템을 설명하는 도면이다.
- [0132] 도 7을 참조하면, 학습 모델 기반의 만화 해설 제공 시스템(700)은 그룹 캡션 제공부(710) 및 만화 해설 제공부(720)를 포함한다.
- [0133] 그룹 캡션 제공부(710)는 만화 데이터 또는 만화 이미지 패널을 입력 받고 그룹 캡션을 생성한다.
- [0134] 그룹 캡션 제공부(710)는 입력되는 복수의 만화 이미지 패널을 학습된 모델에 기초하여 그룹화하고, 그룹화된 이미지 패널에 대한 그룹 캡션을 제공할 수 있다.
- [0135] 만화 해설 제공부(720)는 그룹 캡션에 대하여 시각 정보 및 청각 정보 중 적어도 하나의 정보로 만화 해설을 제공할 수 있다.
- [0136] 학습 모델 기반의 만화 해설 제공 시스템(700)은 학습 모델 제공부(260)를 더 포함할 수 있다.
- [0137] 이때, 학습 모델 제공부(260)는 도 1 내지 도 5에서 설명된 학습된 모델을 제공할 수 있고, 그룹 캡션 제공부(710)는 학습 모델 제공부(260)와의 통신을 통해서 그룹화된 이미지 패널에 대한 그룹 캡션을 제공할 수 있다. 또한, 도 1 내지 도 5에서 설명된 학습된 모델은 그룹 캡션 제공부(710)내에 구비될 수 도 있다.
- [0139] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPA(field programmable array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령

(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 컨트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0140] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치, 또는 전송되는 신호 파(signal wave)에 영구적으로, 또는 일시적으로 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분포되어서, 분포된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

[0141] 이상과 같이 실시예들이 비록 한정된 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

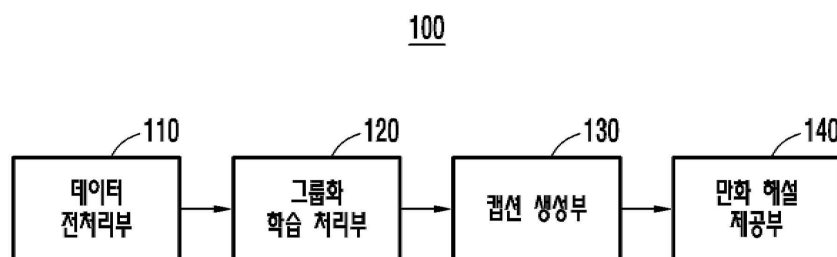
[0142] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

부호의 설명

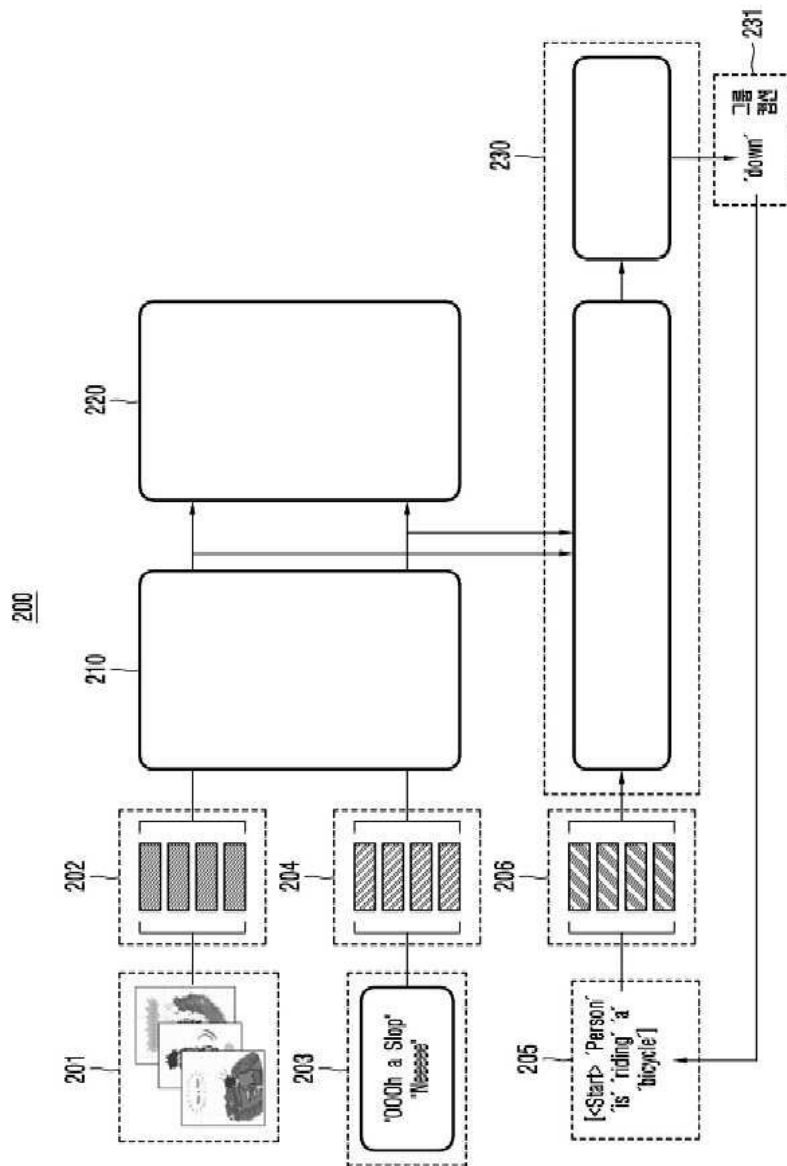
[0143] 100: 만화 해설 제공을 위한 학습 모델 구축 장치
110: 데이터 전처리부 120: 그룹화 학습 처리부
130: 캡션 생성부 140: 만화 해설 제공부

도면

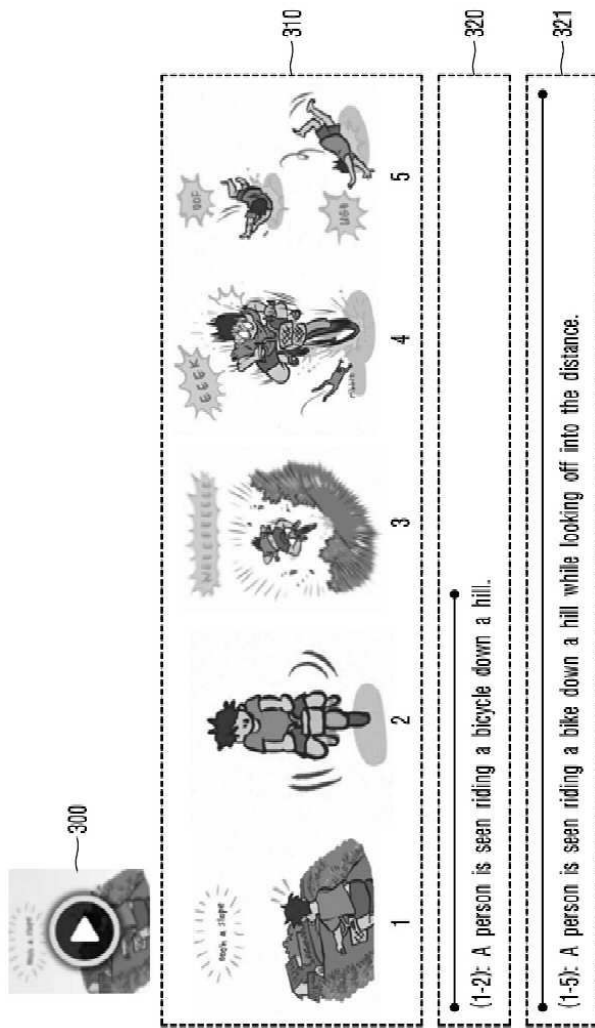
도면1



도면2



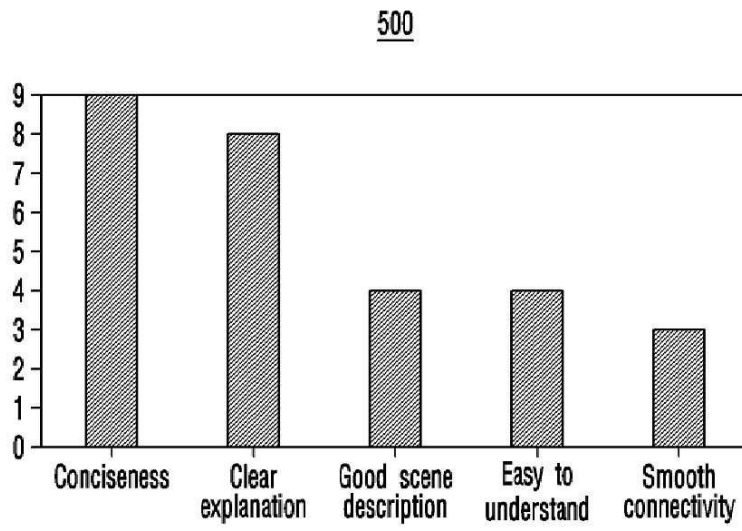
도면3



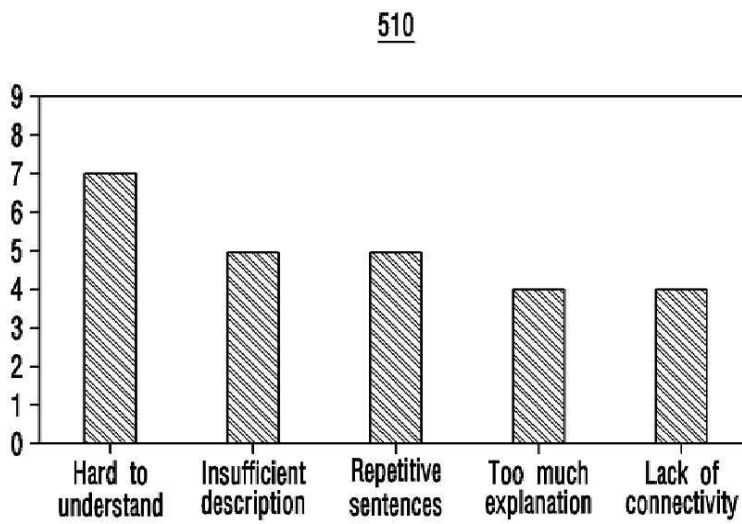
도면4

400	401	402	403
지시사항 1. Please describe the visual content 'of the comic image' such as actors(e.g., people, animal), their actions and expressions(e.g. walking, looking happy) and the environment where these actions take place(e.g., gym, car). *Please use a minimum of 8 words; *Please keep your description concise overall; 2. After describe the comic image below, then click 'Submit' button for the get NTurk code;	만화 이미지 	추가 설명 3. Please provide the description of the comic first, and then indicate the scene numbers in parentheses, ex. (1-5) Ex: He looks sad(1-5) 4. You must write at least two sentences. description Submit	

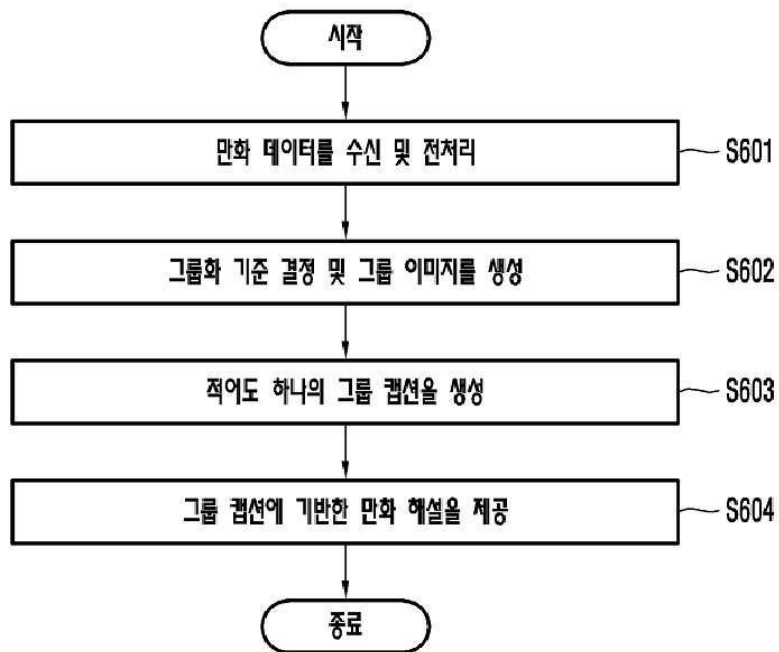
도면5a



도면5b



도면6



도면7

