



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2015-0104126
(43) 공개일자 2015년09월14일

- (51) 국제특허분류(Int. Cl.)
G06K 9/18 (2006.01) **G06F 3/00** (2006.01)
G06K 9/22 (2006.01) **G06T 7/20** (2006.01)
HO4N 1/00 (2006.01)
- (52) CPC특허분류
G06K 9/18 (2013.01)
G06F 3/005 (2013.01)
- (21) 출원번호 10-2015-7020751
- (22) 출원일자(국제) 2013년11월22일
 심사청구일자 **없음**
- (85) 번역문제출일자 2015년07월30일
- (86) 국제출원번호 PCT/US2013/071518
- (87) 국제공개번호 WO 2014/107246
 국제공개일자 2014년07월10일
- (30) 우선권주장
 61/749,248 2013년01월04일 미국(US)
 14/021,337 2013년09월09일 미국(US)
- (71) 출원인
퀄컴 인코포레이티드
 미국 92121-1714 캘리포니아주 샌 디에고 모어하
 우스 드라이브 5775
- (72) 발명자
게르파우초, 미하엘
 미국 92121-1714 캘리포니아주 샌 디에고 모어하
 우스 드라이브 5775
김, 지은
 미국 92121-1714 캘리포니아주 샌 디에고 모어하
 우스 드라이브 5775
 (뒷면에 계속)
- (74) 대리인
특허법인 남엔드남

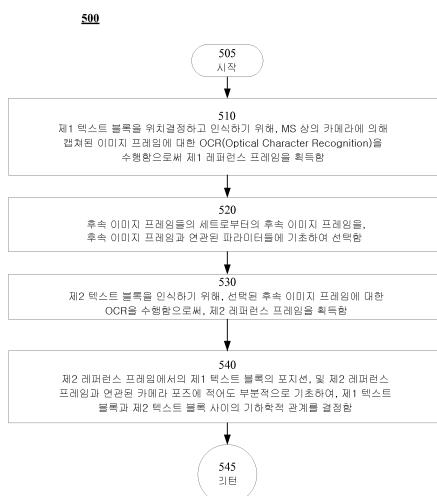
전체 청구항 수 : 총 39 항

(54) 발명의 명칭 모바일 디바이스 기반 텍스트 검출 및 추적

(57) 요약

개시된 실시예들은 모바일 디바이스 기반 텍스트 검출 및 추적에 관한 것이다. 몇몇 실시예들에서, 제1 텍스트 블록을 위치결정하고 인식하기 위해, 카메라에 의해 캡쳐된 이미지 프레임에 대한 OCR(Optical Character Recognition)을 수행함으로써, 제1 레퍼런스 프레임이 획득된다. 후속 이미지 프레임들의 세트로부터의 후속 이미지 프레임은, 선택되는 후속 이미지와 연관된 파라미터들에 기초하여 선택될 수 있고, 제2 텍스트 블록을 인식하기 위해, 선택되는 후속 이미지 프레임에 대한 OCR을 수행함으로써, 제2 레퍼런스 프레임이 획득될 수 있다. 제2 레퍼런스 프레임에서의 제1 텍스트 블록의 포지션, 및 제2 레퍼런스 프레임과 연관된 포즈에 기초하여, 제1 텍스트 블록과 제2 텍스트 블록 사이의 기하학적 관계가 결정된다.

대 표 도 - 도5



(52) CPC특허분류
G06K 9/228 (2013.01)
G06T 7/2033 (2013.01)
HO4N 1/00204 (2013.01)
G06T 2207/30244 (2013.01)

(72) 발명자

니엘센, 페르, 오.

미국 92121-1714 캘리포니아주 샌 디에고 모어하우스 드라이브 5775

로렌스 아속 이니고, 로이

미국 92121-1714 캘리포니아주 샌 디에고 모어하우스 드라이브 5775

챈, 쿠

미국 92121-1714 캘리포니아주 샌 디에고 모어하우스 드라이브 5775

탈론네아우, 로메인

미국 92121-1714 캘리포니아주 샌 디에고 모어하우스 드라이브 5775

명세서

청구범위

청구항 1

모바일 스테이션(MS:mobile station) 상에서의 방법으로서,

제1 텍스트 블록을 위치결정(locate)하고 인식하기 위해, 상기 MS 상의 카메라에 의해 캡쳐된 이미지 프레임에 대한 OCR(Optical Character Recognition)을 수행함으로써, 제1 레퍼런스 프레임을 획득하는 단계;

후속 이미지 프레임들의 세트로부터의 후속 이미지 프레임을, 선택되는 후속 이미지 프레임과 연관된 파라미터들에 기초하여 선택하는 단계;

제2 텍스트 블록을 인식하기 위해, 상기 선택되는 후속 이미지 프레임에 대한 OCR을 수행함으로써, 제2 레퍼런스 프레임을 획득하는 단계; 및

상기 제2 레퍼런스 프레임에서의 상기 제1 텍스트 블록의 포지션, 및 상기 제2 레퍼런스 프레임과 연관된 카메라 포즈에 적어도 부분적으로 기초하여, 상기 제1 텍스트 블록과 상기 제2 텍스트 블록 사이의 기하학적 관계를 결정하는 단계

를 포함하는,

모바일 스테이션(MS) 상에서의 방법.

청구항 2

제 1 항에 있어서,

상기 제1 텍스트 블록과 상기 제2 텍스트 블록 사이의 상기 기하학적 관계에 기초한 시퀀스로 상기 제1 텍스트 블록 및 상기 제2 텍스트 블록을 어셈블링하는 단계

를 더 포함하는,

모바일 스테이션(MS) 상에서의 방법.

청구항 3

제 2 항에 있어서,

상기 제1 텍스트 블록과 상기 제2 텍스트 블록 사이의 상기 기하학적 관계는, 텍스트 블록들이 나타나는 매체와 연관된 레퍼런스의 프레임에 적어도 부분적으로 기초하는,

모바일 스테이션(MS) 상에서의 방법.

청구항 4

제 2 항에 있어서,

상기 제1 텍스트 블록 및 상기 제2 텍스트 블록의 어셈블링된 시퀀스를 텍스트-투-스피치(text-to-speech) 애플리케이션에 입력으로서 제공하는 단계

를 더 포함하는,

모바일 스테이션(MS) 상에서의 방법.

청구항 5

제 1 항에 있어서,

상기 후속 이미지 프레임을 선택하는 단계는,

상기 후속 이미지 프레임들의 세트에 대한 카메라 포즈들을 컴퓨팅하는 단계 – 각각의 카메라 포즈는, 별개의

후속 이미지 프레임과 연관되고, 연관된 후속 이미지 프레임을 상기 제1 레퍼런스 프레임과 정렬하는 것에 적어도 부분적으로 기초하여 결정됨-; 및

컴퓨팅된 카메라 포즈들에 적어도 부분적으로 기초하여, 상기 후속 이미지 프레임들의 세트의 대응하는 이미지 프레임들과 연관된 파라미터들을 결정하는 단계

를 더 포함하는,

모바일 스테이션(MS) 상에서의 방법.

청구항 6

제 5 항에 있어서,

상기 정렬은, ESM(Efficient Second-order Minimization)을 사용하여 수행되는,

모바일 스테이션(MS) 상에서의 방법.

청구항 7

제 6 항에 있어서,

상기 ESM은 상기 연관된 후속 이미지 프레임의 더 낮은 해상도 버전에 대해 동작하는,

모바일 스테이션(MS) 상에서의 방법.

청구항 8

제 5 항에 있어서,

상기 후속 이미지 프레임들의 세트에 대한 카메라 포즈들을 컴퓨팅하는 단계는,

상기 제1 레퍼런스 프레임에서 복수의 특징 지점들을 식별함으로써 획득된 이미지 패치들을 포함하는 추적 타겟을 생성하는 단계;

상기 제1 레퍼런스 프레임과 상기 후속 이미지 프레임 사이의 이미지 패치들의 대응에 기초하여, 상기 세트의 상기 후속 이미지 프레임에서 상기 추적 타겟의 위치를 결정하는 단계; 및

상기 후속 이미지 프레임에서 상기 추적 타겟의 위치에 적어도 부분적으로 기초하여, 상기 후속 이미지 프레임과 연관된 카메라 포즈를 컴퓨팅하는 단계

를 더 포함하는,

모바일 스테이션(MS) 상에서의 방법.

청구항 9

제 8 항에 있어서,

상기 특징 지점들은 상기 제1 레퍼런스 프레임의 내추럴(natural) 특징들에 기초하는,

모바일 스테이션(MS) 상에서의 방법.

청구항 10

제 8 항에 있어서,

개별 특징 지점들에 가중치들이 할당되고, 상기 제1 텍스트 블록 위의 특징 지점들에는 상기 제1 레퍼런스 프레임의 다른 곳에 위치된 특징 지점들에 비해 더 큰 가중치가 할당되는,

모바일 스테이션(MS) 상에서의 방법.

청구항 11

제 8 항에 있어서,

상기 추적 타겟의 생성은, 상기 연관된 후속 이미지 프레임의 상기 제1 레퍼런스 프레임과의 정렬과 실질적으로 동시에 수행되는,

모바일 스테이션(MS) 상에서의 방법.

청구항 12

제 1 항에 있어서,

상기 제1 레퍼런스 프레임과, 상기 후속 이미지 프레임들의 세트는 마커리스(markerless)인,

모바일 스테이션(MS) 상에서의 방법.

청구항 13

제 1 항에 있어서,

상기 파라미터들은,

상기 선택되는 후속 이미지 프레임과 상기 제1 레퍼런스 프레임 사이의 오버랩 영역의 퍼센티지, 또는

상기 선택되는 후속 이미지 프레임에서 보이는 상기 제1 텍스트 블록의 프랙션(fraction) –상기 프랙션은, 상기 선택되는 후속 이미지 프레임에서 상기 제1 텍스트 블록의 보이는 부분을 포함하는 영역 대 상기 제1 텍스트 블록의 총 영역의 비로서 결정됨–, 또는

상기 제1 레퍼런스 프레임에 대한 상기 선택되는 후속 이미지 프레임의 회전 규모, 또는

상기 제1 레퍼런스 프레임에 대한 상기 선택되는 후속 이미지 프레임의 병진 규모(magnitude of translation)

중 적어도 하나를 포함하는,

모바일 스테이션(MS) 상에서의 방법.

청구항 14

제 1 항에 있어서,

상기 카메라 포즈는 6-DOF(6 Degrees of Freedom)로 결정되고, 상기 카메라는 텍스트 블록들을 포함하는 평면 매체에 앞-평행(fronto-parallel)한,

모바일 스테이션(MS) 상에서의 방법.

청구항 15

제 1 항에 있어서,

상기 방법은 증강 현실(AR:Augmented Reality) 애플리케이션에 의해 호출되는,

모바일 스테이션(MS) 상에서의 방법.

청구항 16

제 15 항에 있어서,

가상 오브젝트가 상기 AR 애플리케이션에 의해 상기 제1 텍스트 블록 및 상기 제2 텍스트 블록 위에 배치되는,

모바일 스테이션(MS) 상에서의 방법.

청구항 17

제 16 항에 있어서,

상기 가상 오브젝트는 상기 제1 텍스트 블록 및 상기 제2 텍스트 블록으로부터 번역된 텍스트를 포함하고, 상기 번역된 텍스트는, 상기 제1 텍스트 블록 및 상기 제2 텍스트 블록을 표현하는데 사용된 언어와 상이한 언어인,

모바일 스테이션(MS) 상에서의 방법.

청구항 18

모바일 스테이션(MS)으로서,

제1 이미지 프레임, 및 후속 이미지 프레임들의 세트를 캡쳐하도록 구성된 카메라, 및

상기 카메라에 커플링된 프로세서

를 포함하고, 상기 프로세서는,

제1 텍스트 블록을 위치결정하고 인식하기 위해, 상기 제1 이미지 프레임에 대한 OCR(Optical Character Recognition)을 수행함으로써, 제1 레퍼런스 프레임을 획득하고;

상기 후속 이미지 프레임들의 세트로부터의 후속 이미지 프레임을, 선택되는 후속 이미지 프레임과 연관된 파라미터들에 기초하여 선택하고; 그리고

제2 텍스트 블록을 인식하기 위해, 상기 선택되는 후속 이미지 프레임에 대한 OCR을 수행함으로써, 제2 레퍼런스 프레임을 획득하도록

구성된 워드 인식(word recognition) 모듈, 및

상기 제2 레퍼런스 프레임에서의 상기 제1 텍스트 블록의 포지션, 및 상기 제2 레퍼런스 프레임과 연관된 카메라 포즈에 적어도 부분적으로 기초하여, 상기 제1 텍스트 블록과 상기 제2 텍스트 블록 사이의 기하학적 관계를 결정하도록 구성된 텍스트 어셈블러 모듈

을 포함하는,

모바일 스테이션(MS).

청구항 19

제 18 항에 있어서,

상기 텍스트 어셈블러 모듈은, 상기 제1 텍스트 블록과 상기 제2 텍스트 블록 사이의 상기 기하학적 관계에 기초한 시퀀스로 상기 제1 텍스트 블록 및 상기 제2 텍스트 블록을 어셈블링하도록 추가로 구성되는,

모바일 스테이션(MS).

청구항 20

제 19 항에 있어서,

상기 텍스트 어셈블러 모듈은, 상기 제1 텍스트 블록 및 상기 제2 텍스트 블록의 어셈블링된 시퀀스를 텍스트-투-스피치 애플리케이션에 입력으로서 제공하도록 추가로 구성되는,

모바일 스테이션(MS).

청구항 21

제 18 항에 있어서,

상기 프로세서는 상기 워드 인식 모듈에 동작 가능하게 커플링된 추적 모듈을 더 포함하고, 상기 추적 모듈은,

상기 후속 이미지 프레임들의 세트에 대한 카메라 포즈들을 컴퓨팅하고 –각각의 카메라 포즈는, 별개의 후속 이미지 프레임과 연관되고, 연관된 후속 이미지 프레임을 상기 제1 레퍼런스 프레임과 정렬하는 것에 적어도 부분적으로 기초하여 결정됨–; 그리고

컴퓨팅된 카메라 포즈들에 적어도 부분적으로 기초하여, 상기 후속 이미지 프레임들의 세트의 대응하는 이미지 프레임들과 연관된 파라미터들을 결정하도록

구성된,

모바일 스테이션(MS).

청구항 22

제 21 항에 있어서,

상기 추적 모듈은 ESM(Efficient Second-order Minimization)을 사용하여 상기 정렬을 수행하도록 추가로 구성되는,

모바일 스테이션(MS).

청구항 23

제 22 항에 있어서,

상기 ESM은 상기 연관된 후속 이미지 프레임의 더 낮은 해상도 버전에 대해 동작하는,

모바일 스테이션(MS).

청구항 24

제 21 항에 있어서,

상기 후속 이미지 프레임들의 세트에 대한 카메라 포즈들을 컴퓨팅하기 위해, 상기 추적 모듈은,

상기 제1 레퍼런스 프레임에서 복수의 특징 지점들을 식별함으로써 획득된 이미지 패치들을 포함하는 추적 타겟을 생성하고;

상기 제1 레퍼런스 프레임과 상기 후속 이미지 프레임 사이의 이미지 패치들의 대응에 기초하여, 상기 세트의 상기 후속 이미지 프레임에서 상기 추적 타겟의 위치를 결정하고; 그리고

상기 후속 이미지 프레임에서 상기 추적 타겟의 위치에 적어도 부분적으로 기초하여, 상기 후속 이미지 프레임과 연관된 카메라 포즈를 컴퓨팅하도록

추가로 구성되는,

모바일 스테이션(MS).

청구항 25

제 24 항에 있어서,

상기 특징 지점들은 상기 제1 레퍼런스 프레임의 내추럴 특징들에 기초하는,

모바일 스테이션(MS).

청구항 26

제 24 항에 있어서,

상기 추적 모듈은, 상기 제1 텍스트 블록 위의 특징 지점들에는 상기 제1 레퍼런스 프레임의 다른 곳에 위치된 특징 지점들에 비해 더 큰 가중치가 할당되도록, 개별 특징 지점들에 가중치들을 할당하도록 구성되는,

모바일 스테이션(MS).

청구항 27

제 24 항에 있어서,

상기 추적 모듈은, 상기 연관된 후속 이미지 프레임의 상기 제1 레퍼런스 프레임과의 정렬과 실질적으로 동시에 상기 추적 타겟을 생성하도록 구성되는,

모바일 스테이션(MS).

청구항 28

제 18 항에 있어서,

상기 카메라에 의해 캡쳐되는 상기 제1 레퍼런스 프레임 및 상기 후속 이미지 프레임들의 세트는 마커리스인,

모바일 스테이션(MS).

청구항 29

제 18 항에 있어서,

상기 파라미터들은,

상기 선택되는 후속 이미지 프레임과 상기 제1 레퍼런스 프레임 사이의 오버랩 영역의 퍼센티지, 또는

상기 선택되는 후속 이미지 프레임에서 보이는 상기 제1 텍스트 블록의 프랙션 – 상기 프랙션은, 상기 선택되는 후속 이미지 프레임에서 상기 제1 텍스트 블록의 보이는 부분을 포함하는 영역 대 상기 제1 텍스트 블록의 총 영역의 비로서 결정됨 –, 또는

상기 제1 레퍼런스 프레임에 대한 상기 선택되는 후속 이미지 프레임의 회전 규모, 또는

상기 제1 레퍼런스 프레임에 대한 상기 선택되는 후속 이미지 프레임의 병진 규모

중 적어도 하나를 포함하는,

모바일 스테이션(MS).

청구항 30

장치로서,

이미지 프레임들의 시퀀스를 캡처하기 위한 이미징 수단;

제1 텍스트 블록을 위치결정하고 인식하기 위해, 상기 이미지 프레임들의 시퀀스의 이미지 프레임에 대한 OCR(Optical Character Recognition)을 수행함으로써, 제1 레퍼런스 프레임을 획득하기 위한 수단;

상기 이미지 프레임들의 시퀀스로부터 후속 이미지 프레임을 선택하기 위한 수단 –상기 선택은, 선택되는 후속 이미지 프레임과 연관된 파라미터들에 기초함–;

제2 텍스트 블록을 인식하기 위해, 상기 선택되는 후속 이미지 프레임에 대한 OCR을 수행함으로써, 제2 레퍼런스 프레임을 획득하기 위한 수단; 및

상기 제2 레퍼런스 프레임에서의 상기 제1 텍스트 블록의 포지션, 및 상기 제2 레퍼런스 프레임과 연관된 상기 이미징 수단의 포즈에 적어도 부분적으로 기초하여, 상기 제1 텍스트 블록과 상기 제2 텍스트 블록 사이의 기하학적 관계를 결정하기 위한 수단

을 포함하는,

장치.

청구항 31

제 30 항에 있어서,

상기 제1 텍스트 블록과 상기 제2 텍스트 블록 사이의 상기 기하학적 관계에 기초한 시퀀스로 상기 제1 텍스트 블록 및 상기 제2 텍스트 블록을 어셈블링하기 위한 수단

을 더 포함하는,

장치.

청구항 32

제 31 항에 있어서,

상기 제1 텍스트 블록 및 상기 제2 텍스트 블록의 어셈블링된 시퀀스를 텍스트-투-스피치 애플리케이션에 입력으로서 제공하기 위한 수단

을 더 포함하는,

장치.

청구항 33

제 30 항에 있어서,

상기 후속 이미지 프레임을 선택하기 위한 수단은,

상기 이미지 프레임들의 시퀀스의 이미지 프레임들에 대한 상기 이미징 수단의 포즈들을 컴퓨팅하기 위한 수단 – 상기 이미징 수단의 각각의 컴퓨팅된 포즈는, 별개의 이미지 프레임과 연관되고, 적어도 부분적으로, 연관된 이미지 프레임을 상기 제1 레퍼런스 프레임과 정렬함으로써 결정됨 –; 및

상기 이미징 수단의 상기 컴퓨팅된 포즈들에 적어도 부분적으로 기초하여, 상기 이미지 프레임들의 시퀀스의 대응하는 이미지 프레임들과 연관된 파라미터들을 결정하기 위한 수단

을 포함하는,

장치.

청구항 34

제 33 항에 있어서,

상기 이미징 수단의 포즈들을 컴퓨팅하기 위한 수단은,

상기 제1 레퍼런스 프레임에서 복수의 특징 지점들을 식별함으로써 획득된 이미지 패치들을 포함하는 추적 타겟을 생성하기 위한 수단;

상기 제1 레퍼런스 프레임과 후속 이미지 프레임 사이의 이미지 패치들의 대응에 기초하여, 상기 이미지 프레임들의 시퀀스의 상기 후속 이미지 프레임에서 상기 추적 타겟의 위치를 결정하기 위한 수단; 및

상기 후속 이미지 프레임에서 상기 추적 타겟의 위치에 적어도 부분적으로 기초하여, 상기 후속 이미지 프레임과 연관된 카메라 포즈를 컴퓨팅하기 위한 수단

을 포함하는,

장치.

청구항 35

제 34 항에 있어서,

개별 특징 지점들에 가중치들이 할당되고, 상기 제1 텍스트 블록 위의 특징 지점들에는 상기 제1 레퍼런스 프레임의 다른 곳에 위치된 특징 지점들에 비해 더 큰 가중치가 할당되는,

장치.

청구항 36

제 34 항에 있어서,

상기 추적 타겟을 생성하기 위한 수단은 상기 연관된 이미지 프레임의 상기 제1 레퍼런스 프레임과의 정렬과 실질적으로 동시에 동작하는,

장치.

청구항 37

제 30 항에 있어서,

상기 이미징 수단에 의해 캡쳐되는 상기 이미지 프레임들의 시퀀스의 이미지 프레임들은 마커리스인,

장치.

청구항 38

제 30 항에 있어서,

상기 파라미터들은,

상기 선택되는 후속 이미지 프레임과 상기 제1 레퍼런스 프레임 사이의 오버랩 영역의 퍼센티지, 또는

상기 선택되는 후속 이미지 프레임에서 보이는 상기 제1 텍스트 블록의 프랙션, 또는

상기 제1 레퍼런스 프레임에 대한 상기 선택되는 후속 이미지 프레임의 회전 규모, 또는

상기 제1 레퍼런스 프레임에 대한 상기 선택되는 후속 이미지 프레임의 병진 규모

중 적어도 하나를 포함하는,

장치.

청구항 39

비-일시적 컴퓨터-관독가능 매체로서,

프로세서에 의해 실행될 때, 모바일 스테이션(MS) 상에서의 방법을 수행하는 명령들

을 포함하고, 상기 방법은,

제1 텍스트 블록을 위치결정하고 인식하기 위해, 상기 MS 상의 카메라에 의해 캡쳐된 이미지 프레임에 대한 OCR(Optical Character Recognition)을 수행함으로써, 제1 레퍼런스 프레임을 획득하는 단계;

후속 이미지 프레임들의 세트로부터 후속 이미지 프레임을 선택하는 단계 – 상기 선택은, 선택되는 후속 이미지 프레임과 연관된 파라미터들에 기초함 –;

제2 텍스트 블록을 인식하기 위해, 상기 선택되는 후속 이미지 프레임에 대한 OCR을 수행함으로써, 제2 레퍼런스 프레임을 획득하는 단계; 및

상기 제2 레퍼런스 프레임에서의 상기 제1 텍스트 블록의 포지션, 및 상기 제2 레퍼런스 프레임과 연관된 카메라 포즈에 적어도 부분적으로 기초하여, 상기 제1 텍스트 블록과 상기 제2 텍스트 블록 사이의 기하학적 관계를 결정하는 단계

를 포함하는,

비-일시적 컴퓨터-관독가능 매체.

발명의 설명

기술 분야

관련 출원들에 대한 상호 참조

[0001] 본 출원은, 2013년 9월 9일자로 출원되고 "Mobile Device Based Text Tracking and Detection"로 명명된 미국 출원 번호 14/021,337에 대한 이득 및 우선권을 주장하고, 이어서, 이 미국 출원은, 2013년 1월 4일자로 출원되고 "Mobile Device Based Text Tracking and Detection"로 명명된 미국 가출원 번호 61/749,248에 대한 이득 및 우선권을 주장하며, 이 출원들 둘 다가 인용에 의해 그들 전체가 본원에 통합된다.

[0002] 본원에 개시되는 발명의 요지는 텍스트 검출 및 추적에 관한 것이다.

배경 기술

[0003] 점점 더, 스마트폰들과 같은 모바일 디바이스들은 다양한 다른 타입들의 센서들과 함께 카메라들과 같은 하나 또는 그 초과의 광학 센서들을 통합한다. 모바일 디바이스들 상의 카메라들은 종종, 스크린 이미지 및 비디오 이미지 둘 다 가능하다. 그러나, 텍스트의 스캐닝을 수행하는 애플리케이션들에 대한 모바일 디바이스들의 사용은 많은 과제들 때문에 제한되었다. 특정 과제들의 예들은 사용자 손 움직임들로부터의 모션 블러(motion blur), 스캐닝되는 텍스트에 대한 카메라의 방향의 변화(variation)들에 의해 유발되는 관점에서의 왜곡들, 및 모바일 디바이스 광학 서브시스템 제한들을 포함한다. 예컨대, 모바일 디바이스 카메라에 의해 찍힌 비디오 이미지들은 스캐너에 비해 저 해상도를 가질 수 있고, 부가하여, 모션 블러 및 관점 왜곡에 시달릴 수 있다. 신

속한 무결정성 텍스트 검출 및 추적을 가능하게 하는 것은, 텍스트 검출 및 추적 애플리케이션의 출력을 활용할 수 있는 넓은 범위의 애플리케이션들을 가능하게 함으로써, 모바일 디바이스들의 유용성을 개선할 수 있다.

[0005] [0004] 그러므로, 텍스트 스캐닝, 검출 및/또는 인식, 및/또는 추적을 가능하게 하기 위한 장치, 시스템들 및 방법들에 대한 필요가 존재한다.

발명의 내용

[0006] [0005] 일 양상에서, 모바일 스테이션(MS:Mobile Station) 상의 방법은, 제1 텍스트 블록을 위치결정(locate)하고 인식하기 위해, MS 상의 카메라에 의해 캡쳐된 이미지 프레임에 대한 OCR(Optical Character Recognition)을 수행함으로써, 제1 레퍼런스 프레임을 획득하는 단계; 후속 이미지 프레임들의 세트로부터의 후속 이미지 프레임을, 선택되는 후속 이미지 프레임과 연관된 파라미터들에 기초하여 선택하는 단계; 제2 텍스트 블록을 인식하기 위해, 상기 선택되는 후속 이미지 프레임에 대한 OCR을 수행함으로써, 제2 레퍼런스 프레임을 획득하는 단계; 및 제2 레퍼런스 프레임에서의 제1 텍스트 블록의 위치, 및 제2 레퍼런스 프레임과 연관된 카메라 포즈에 적어도 부분적으로 기초하여, 제1 텍스트 블록과 제2 텍스트 블록 사이의 기하학적 관계를 결정하는 단계를 포함할 수 있다.

[0007] [0006] 다른 양상에서, MS는, 제1 이미지 프레임, 및 후속 이미지 프레임들의 세트를 캡쳐하도록 구성된 카메라, 및 카메라에 커플링된 프로세서를 포함할 수 있다. 프로세서는, 제1 텍스트 블록을 위치결정하고 인식하기 위해, 제1 이미지 프레임에 대한 OCR(Optical Character Recognition)을 수행함으로써, 제1 레퍼런스 프레임을 획득하고; 후속 이미지 프레임들의 세트로부터의 후속 이미지 프레임을, 선택되는 후속 이미지 프레임과 연관된 파라미터들에 기초하여 선택하고; 그리고 제2 텍스트 블록을 인식하기 위해, 상기 선택되는 후속 이미지 프레임에 대한 OCR을 수행함으로써, 제2 레퍼런스 프레임을 획득하도록 구성된 워드 인식(word recognition) 모듈을 포함할 수 있다. 프로세서는, 제2 레퍼런스 프레임에서의 제1 텍스트 블록의 위치, 및 제2 레퍼런스 프레임과 연관된 카메라 포즈에 적어도 부분적으로 기초하여, 제1 텍스트 블록과 제2 텍스트 블록 사이의 기하학적 관계를 결정하도록 구성된 텍스트 어셈블리 모듈을 더 포함할 수 있다.

[0008] [0007] 추가 양상에서, 장치는, 이미지 프레임들의 시퀀스를 캡쳐하기 위한 이미징 수단; 제1 텍스트 블록을 위치결정하고 인식하기 위해, 이미지 프레임들의 시퀀스의 이미지 프레임에 대한 OCR(Optical Character Recognition)을 수행함으로써, 제1 레퍼런스 프레임을 획득하기 위한 수단; 이미지 프레임들의 시퀀스로부터 후속 이미지 프레임을 선택하기 위한 수단 -선택은, 선택되는 후속 이미지 프레임과 연관된 파라미터들에 기초함-; 제2 텍스트 블록을 인식하기 위해, 선택되는 후속 이미지 프레임에 대한 OCR을 수행함으로써, 제2 레퍼런스 프레임을 획득하기 위한 수단; 및 제2 레퍼런스 프레임에서의 제1 텍스트 블록의 위치, 및 제2 레퍼런스 프레임과 연관된 이미징 수단의 포즈에 적어도 부분적으로 기초하여, 제1 텍스트 블록과 제2 텍스트 블록 사이의 기하학적 관계를 결정하기 위한 수단을 포함할 수 있다.

[0009] [0008] 다른 양상에서, 비-일시적 컴퓨터-판독가능 매체는, 프로세서에 의해 실행될 때, 모바일 스테이션(MS) 상의 방법을 수행할 수 있는 명령들을 포함할 수 있고, 방법은, 제1 텍스트 블록을 위치결정하고 인식하기 위해, MS 상의 카메라에 의해 캡쳐된 이미지 프레임에 대한 OCR(Optical Character Recognition)을 수행함으로써, 제1 레퍼런스 프레임을 획득하는 단계; 후속 이미지 프레임들의 세트로부터 후속 이미지 프레임을 선택하는 단계 -선택은, 선택되는 후속 이미지 프레임과 연관된 파라미터들에 기초함-; 제2 텍스트 블록을 인식하기 위해, 선택되는 후속 이미지 프레임에 대한 OCR을 수행함으로써, 제2 레퍼런스 프레임을 획득하는 단계; 및 제2 레퍼런스 프레임에서의 제1 텍스트 블록의 위치, 및 제2 레퍼런스 프레임과 연관된 카메라 포즈에 적어도 부분적으로 기초하여, 제1 텍스트 블록과 제2 텍스트 블록 사이의 기하학적 관계를 결정하는 단계를 포함할 수 있다.

[0010] [0009] 또한, 개시되는 실시예들은 소프트웨어, 펌웨어, 그리고 비 일시적 컴퓨터 판독가능 미디어 또는 컴퓨터 판독가능 메모리를 사용하여 프로세서들에 의해 생성, 저장, 액세스, 판독 또는 수정되는 프로그램 명령들에 관한 것이다.

[0011] [0010] 이러한 그리고 다른 실시예들은 하기의 도면들에 대하여 아래에서 추가로 설명된다. 하기의 상세한 설명으로부터 다른 양상들이 당업자들에 쉽게 명백해질 수 있음이 이해되며, 다양한 양상들이 예시로서 도시 및 설명된다. 도면들 및 상세한 설명은 사실상 예시적인 것으로서 간주되어야 하고, 제한적인 것으로서 간주되지 않아야 한다.

도면의 간단한 설명

[0012]

- [0011] 도 1은 텍스트 추적 및 검출을 수행할 수 있는 예시적 MS의 블록도를 도시한다.
- [0012] 도 2는 표면 상의 워드들에 대한 카메라의 포지션들, 및 텍스트 블록들의 포지션을 추적 및/또는 유지하는데 사용될 수 있는 좌표계들을 도시한다.
- [0013] 도 3은 개시되는 실시예들과 일치하는 텍스트 검출 및 추적의 예시적 방법에 대한 흐름도를 도시한다.
- [0014] 도 4는 개시되는 실시예들과 일치하는 텍스트 추적을 위한 시스템의 예시적 아키텍처를 도시한다.
- [0015] 도 5는 개시되는 실시예들과 일치하는, 모바일 스테이션 상에서 수행될 수 있는 예시적 방법에 대한 흐름도를 도시한다.

발명을 실시하기 위한 구체적인 내용

[0013]

[0016] 본원에서 개시되는 실시예들은, 단지 예로서, 도면들을 참조하여 설명될 것이다. 첨부된 도면들과 관련하여 아래에서 진술되는 상세한 설명은 본 개시물의 다양한 양상들의 설명으로서 의도되며, 본 개시물이 실시될 수 있는 유일한 양상들을 표현하도록 의도되지 않는다. 본 개시물에서 설명되는 각각의 양상은 단지, 본 개시물의 예 또는 예시로서 제공되고, 반드시 다른 양상들보다 바람직하거나 또는 유리한 것으로서 이해되어서는 안 된다. 상세한 설명은 본 개시물의 완전한 이해를 제공하는 목적을 위해 특정 세부사항들을 포함한다. 그러나, 이러한 특정 세부사항들 없이 본 개시물이 실시될 수 있음이 당업자들에게 명백할 것이다. 몇몇 실례들에서는, 본 개시물의 개념들을 모호하게 하는 것을 방지하기 위하여, 잘 알려진 구조들 및 디바이스들은 블록도 형태로 도시된다. 두문자어들 및 다른 설명적 용어는 단지 편의 및 명확성을 위해 사용될 수 있고, 본 개시물의 범위를 제한하도록 의도되지 않는다.

[0014]

[0017] 모바일 디바이스 카메라들은 종종, 그들의 적용 가능성을 제한하는 단점들에 시달린다. 예컨대, 저 카메라 해상도 및 다른 광학 서브시스템 제한들이 종래의 스캐닝 및 OCR(Optical Character Recognition) 관련 애플리케이션들에서 모바일 디바이스 카메라들의 유용성을 제한할 수 있다. 사용자 손 떨림으로부터의 모션 블러, 및 스캐닝 프로세스 동안의 카메라 틸트 변화들로부터의 관점 왜곡들이, 종래의 텍스트 스캐닝 및 OCR 애플리케이션들에서 모바일 디바이스 카메라들을 사용하는 것에 부가적인 장애들을 야기한다. 부가하여, 스캐닝 동안 정확한 텍스트 캡쳐 및 더 큰 상대 텍스트 크기를 보장하기 위해, 카메라가 종종 텍스트에 더 가까이 유지되고, 이로써 단일 이미지 또는 프레임에서 캡쳐될 수 있는 레터(letter)들 또는 워드들의 수가 제한된다. 부가하여, 종래의 OCR이 상당한 계산적 오버헤드를 초래할 수 있기 때문에, 프레임 단위로 프레임에 대한 OCR의 사용은, 수용할 수 없는 지역들을 도입시킴으로써 많은 애플리케이션들에 대해 비효율적이고 비실용적일 수 있다.

[0015]

[0018] 본원에 개시되는 장치, 시스템들 및 방법들을 포함하는 실시예들은 효율적이고 연속적인 텍스트 스캐닝, 추적, 및 인식을 가능하게 한다. 본원에서 사용되는 바와 같이, 용어 모바일 스테이션(MS)은 셀룰러 폰들, 게임 디바이스들, 이미징 디바이스들, 모바일 폰들 또는 다른 무선 통신 디바이스들, PCS(personal communication system) 디바이스들, PND(personal navigation device)들, PIM(Personal Information Manager)들, PDA(Personal Digital Assistant)들, 랩톱들 등을 포함하는 모바일 디바이스들을 폭넓게 지칭하는데 사용된다.

[0016]

[0019] 추가로, "모바일 스테이션"은 통신 엘리먼트들 및/또는 기능을 생략할 수 있다. 또한, 용어 "모바일 스테이션"은, 네트워크에 연결되거나 또는 무선으로든 또는 유선 연결을 통해서든 다른 디바이스와 다른 방식으로 통신하도록 구성되지 않을 수 있는 디바이스들로 의도된다. 예컨대, 본원에서 설명되는 실시예들은 독립형 디바이스들, 예컨대 게임 디바이스, 핸드 헬드 카메라, 또는 네트워크에 연결되거나 또는 무선으로든 또는 유선 연결을 통해서든 다른 디바이스와 다른 방식으로 통신하도록 구성되지 않을 수 있는 다른 디바이스에서 구현될 수 있다.

[0017]

[0020] 몇몇 실시예들에서, 모바일 스테이션은, 무선 통신 및/또는 내비게이션 신호들을 수신할 수 있고, 그리고 단거리 무선, 적외선, 및/또는 유선 연결들을 통해서를 포함하여 다른 디바이스들과 통신할 수 있는 모바일 디바이스의 형태를 취할 수 있다. 또한, "모바일 스테이션"은, 무선 통신 디바이스들, 컴퓨터들, 랩톱들 등을 비롯해, 예컨대 인터넷, Wi-Fi, 또는 다른 네트워크를 통해 서버와 통신할 수 있는 모든 디바이스들을 포함하도록 의도된다.

[0018]

[0021] 도 1은 예시적 MS(100)의 블록도를 도시한다. 도 1에 도시된 바와 같이, MS(100)는 카메라(들)(110), 관성 측정 유닛(IMU: Inertial Measurement Unit)(130), 디스플레이(145), 프로세서(들)(150), 메모리(160) 및/

또는 트랜시버(170)를 포함할 수 있고, 이들은 연결들(120)을 통해 동작 가능하게 커플링될 수 있다. 연결들(120)은 버스들, 라인들, 광섬유들, 링크들 등, 또는 이들의 어떤 결합을 포함할 수 있다.

[0019] [0022] 트랜시버(170)는, 예컨대, 하나 또는 그 초과의 타입들의 무선 통신 네트워크들을 통해 하나 또는 그 초과의 신호들을 송신하도록 인에이블링된 송신기, 및 하나 또는 그 초과의 타입들의 무선 통신 네트워크들을 통해 송신된 하나 또는 그 초과의 신호들을 수신하기 위한 수신기를 포함할 수 있다. 트랜시버들(170)은 다양한 기술들, 예컨대, 표준들 중 IEEE 802.11 패밀리에 기초할 수 있는 WLAN(Wireless Local Area Network)들 또는 Wi-Fi 네트워크들, 블루투스, NFC(Near Field Communication), WWAN(Wireless Wide Area Network)들, 예컨대 LTE, WiMAX 등에 기초하여 무선 네트워크들과의 통신을 허용할 수 있지만, 이들로 제한되는 것은 아니다. 몇몇 실시예들에서, MS(100)는 유선 네트워크들을 통해 통신하기 위한 하나 또는 그 초과의 포트들을 또한 포함할 수 있다. 몇몇 실시예들에서, MS(100) 상의 하나 또는 그 초과의 다른 포트들 및/또는 트랜시버(170)는 생략될 수 있다.

[0020] [0023] 카메라(들)(110)는 캡쳐된 이미지들을 프로세서(150)에 송신할 수 있다. 몇몇 실시예들에서, 카메라들(110)은 전방 카메라(front-facing camera) 및/또는 후방 카메라(rear-facing camera)를 포함할 수 있고, 그리고/또는 CMOS/광학 센서들을 또한 통합할 수 있다. 일 실시예에서, 전방 카메라가 디바이스의 정상적인 사용자 동작 동안에 사용자를 향할 수 있는 반면에, 후방 카메라는 디바이스의 정상적인 동작 동안에 사용자를 등질 수 있다. 몇몇 실시예들에서, 카메라들(110)은 스틸 이미지 및 비디오 이미지 둘 다를 캡쳐할 수 있다. 카메라(들)(110)는 비디오 이미지들, 및/또는 스틸 이미지들의 시퀀스를 프로세서들(150)에 송신할 수 있다. 일 실시예에서, 카메라들(110)에 의해 캡쳐된 이미지들은 미가공의 압축되지 않은 포맷으로 있을 수 있고, 그리고 프로세싱되고 그리고/또는 메모리(160)에 저장되기에 앞서 압축될 수 있다. 몇몇 실시예들에서, 이미지 압축은, 프로세서들(150)에 의해 무손실 또는 손실 압축 기술들을 사용하여 수행될 수 있다.

[0021] [0024] 몇몇 실시예들에서, 프로세서들(150)은 또한 IMU(130)로부터 입력들을 수신할 수 있다. 몇몇 실시예들에서, IMU(130)는 3-축 가속도계(들), 3-축 자이로스코프(들), 및/또는 자력계(들)를 포함할 수 있다. IMU(130)는 속도, 방향, 및/또는 다른 포지션 관련 정보를 프로세서(150)에 제공할 수 있다. 몇몇 실시예들에서, IMU(130)는 카메라들(110)에 의한 각각의 이미지 프레임의 캡쳐와 동기화된 측정 정보를 출력할 수 있다.

[0022] [0025] MS(100)에 포함된 모든 모듈들이 도 1에 도시된 것은 아니다. 예시적 MS(100)는 또한, 다양한 방법들에서 본 개시물과 일치하는 방식으로, 예컨대, 도시된 기능 블록들 중 하나 또는 그 초과를 부가, 결합, 또는 생략함으로써 수정될 수 있다. 예컨대, 몇몇 구성들에서, MS(100)는 IMU(130)를 포함하지 않을 수 있다. 추가로, 특정 예시적 구현들에서, MS(100)는 주변광 센서, 마이크로폰들, 음향 센서들, 초음파 센서들 등과 같은 다양한 다른 센서들(미도시)을 포함할 수 있다. 몇몇 실시예들에서, MS(100)는, 카메라(110)에 의해 캡쳐된 이미지 프레임들의 라이브 뷰잉(live viewing)을 허용할 수 있는 디스플레이를 포함할 수 있다. 몇몇 실시예들에서, MS(100)의 일부분들은 하나 또는 그 초과의 칩셋들 등의 형태를 취할 수 있다.

[0023] [0026] 프로세서들(150)은 하드웨어, 펌웨어, 및 소프트웨어의 결합을 사용하여 구현될 수 있다. 프로세서들(150)은, 텍스트 검출 및 추적, OCR(Optical Character Recognition) 및/또는 컴퓨터 비전 및 이미지 프로세싱에 관련된 컴퓨팅 프로세서 또는 프로세스의 적어도 일부분을 수행하도록 구성 가능한 하나 또는 그 초과의 회로들을 표현할 수 있다. 프로세서들(150)은 메모리(160)로부터 명령들 및/또는 데이터를 리트리빙할 수 있다. 프로세서들(150)은 하나 또는 그 초과의 GPU(Graphical Processing Unit)들, ASIC(Application Specific Integrated Circuit)들, DSP(Digital Signal Processor)들, DSPD(Digital Signal Processing Device)들, PLD(Programmable Logic Device)들, FPGA(Field Programmable Gate Array)들, 제어기들, 마이크로제어기들, 마이크로프로세서들, 임베딩된 프로세서 코어들, 및 본원에 설명되는 기능들을 수행하도록 설계된 다른 전자 유닛들, 또는 이들의 결합을 사용하여 구현될 수 있다.

[0024] [0027] 메모리(160)는 프로세서들(150) 내부에 그리고/또는 프로세서들(150) 외부에 구현될 수 있다. 본원에서 사용되는 바와 같이, 용어 "메모리"는 임의의 타입의 장기, 단기, 휘발성, 비휘발성, 또는 다른 메모리를 지칭하고, 그리고 임의의 특정 타입의 메모리 또는 메모리들의 수, 또는 메모리가 저장되는 물리적 미디어의 타입으로 제한되지 않아야 한다. 몇몇 실시예들에서, 메모리(160)는 이미지 프로세싱, 텍스트 검출 및 추적, 및 프로세서(150)에 의해 수행되는 다른 작업들을 가능하게 하기 위한 코드를 보유할 수 있다. 예컨대, 메모리(160)는 데이터, 캡쳐된 스틸 이미지들, 비디오 프레임들, 프로그램 결과들, 뿐만 아니라 IMU(130) 및 다른 센서들에 의해 제공되는 데이터를 보유할 수 있다. 일반적으로, 메모리(160)는 임의의 데이터 스토리지 메커니즘을 표현할 수 있다. 메모리(160)는, 예컨대, 주 메모리 및/또는 보조 메모리를 포함할 수 있다. 주 메모리는, 예컨대,

핸덤 액세스 메모리, 읽기 전용 메모리 등을 포함할 수 있다. 도 1에서 프로세서들(150)과 별개인 것으로서 예시되지만, 주 메모리의 일부 또는 전부가 프로세서들(150) 내에 제공될 수 있거나 또는 그렇지 않으면 프로세서들(150)과 코로케이트(co-locate)되고 그리고/또는 프로세서들(150)에 커플링될 수 있음이 이해되어야 한다.

[0025] [0028] 보조 메모리는, 예컨대, 주 메모리와 동일한 또는 유사한 타입의 메모리, 및/또는 하나 또는 그 초과의 데이터 스토리지 디바이스들 또는 시스템들, 이를테면, 예컨대, 플래시/USB 메모리 드라이브들, 메모리 카드 드라이브들, 디스크 드라이브들, 광학 디스크 드라이브들, 테이프 드라이브들, 고체 상태 메모리 드라이브들 등을 포함할 수 있다. 특정 구현들에서, 보조 메모리는 MS(100)에 커플링되는 탈착 가능한 미디어 드라이브(미도시)의 비-일시적 컴퓨터-판독가능 매체를 동작 가능하게 수용할 수 있거나 또는 그렇지 않으면 이 비-일시적 컴퓨터-판독가능 매체에 커플링되도록 구성 가능할 수 있다. 몇몇 실시예들에서, 비-일시적 컴퓨터 판독가능 매체는 메모리(160) 및/또는 프로세서(150)의 일부를 형성할 수 있다.

[0026] [0029] MS(100)는 카메라(110)에 의해 캡처된 이미지들을 프로세싱하기 위한 소프트웨어를 또한 포함할 수 있다. 몇몇 실시예들에서, 프로세서(150)는, 캡처된 이미지들의 특징들을 검출 및 해석하기 위해 카메라(110)에 의해 캡처된 하나 또는 그 초과의 이미지들을 프로세싱하기 위한 소프트웨어를 실행할 수 있다. 예컨대, 프로세서(150)는, 이미지들의 텍스트를 인식하기 위해 이미지들에 대한 OCR(Optical Character Recognition)을 수행할 수 있다.

[0027] [0030] 몇몇 실시예들에서, 프로세서(150)는, 이미지를 정합하고 카메라 포즈를 결정하기 위해 내추럴 특징 추적 기술들을 사용할 수 있다. 컴퓨터 비전에서, 검출은, 캡처된 이미지 프레임에서 타겟 오브젝트를 로컬화하고 오브젝트에 대한 카메라 포즈를 컴퓨팅하는 프로세스를 지칭한다. 추적은, 이미지 프레임들의 시간상 시퀀스에 걸쳐 오브젝트에 대한 카메라 포즈 추정을 지칭한다. 특징 추적에서는, 특징들이 한 번 검출될 수 있고, 그 다음, 이전 정보 또는 카메라 포즈에 관한 가정들을 사용하여 후속 이미지들에서 직접 추적될 수 있다. 추적할 때, 연속적인 프레임들 사이에서 타겟 오브젝트 포지션들의 차이는 통상적으로 작다. 카메라의 포즈는 레퍼런스의 프레임에 대한 카메라의 포지션 및 방향을 지칭한다. 정합, 이미지 정합 또는 이미지 정렬은, 상이한 데이터 세트들을 단일 좌표계로 변환 및/또는 통합하는 프로세스를 지칭한다. 따라서, 예컨대, 이미지 쌍 사이의 픽셀-단위 차이들을 최소화하면서, 두 개의 이미지들의 좌표계들을 정렬하는 기하학적 변환을 결정하는데 이미지 정합 또는 정렬이 사용될 수 있다. 이미지 정합에서, 정렬은 임의적인 또는 비교적 큰 베이스라인을 갖는 두 개의 이미지들을 수반할 수 있다.

[0028] [0031] 통상적으로, 특징-기반 추적 방법들에서는, 오브젝트들을 표현하는데 특징들이 사용될 수 있다. 많은 종래의 특징-기반 추적 방법들에서는, 추적이 인공 외부 마커들에 기초하고, 그 다음, 로컬화 및 추적을 위해 이 인공 외부 마커들이 사용된다. 그러나, 관련 없는 마커들을 사용하는 종래의 방법들은 도로명 게시판들, 오브젝트들 상의 라벨들, 또는 다양한 미디어 상의 텍스트와 같은, 자연 세계의 오브젝트들이 스캐닝될 때에는 비 실용적이다. 그러므로, 본원에 개시되는 실시예들은 추적을 수행하기 위해 내재적 특징들, 예컨대 텍스트-기반 특징들, 텍스트가 나타나는 매체 상의 특징들, 텍스처들, 및 캡처된 이미지 프레임들의 다른 독특한 양상들을 사용할 수 있다. 관련 없는 마커들을 사용하지 않는 추적은 또한 "마커리스(markerless)" 추적으로 지칭된다. 본원에서 개시되는 실시예들은 마커리스 추적을 가능하게 한다.

[0029] [0032] 특징-기반 오브젝트 추적은 특징 추출 및 특징 대응(feature correspondence)을 포함할 수 있다. 특징 추출이 이미지에 있는 특징들의 검출을 지칭하는 반면에, 특징 대응은 두 개의 이미지들의 특징들 사이의 대응을 결정 또는 매칭하는 프로세스이다. 그러므로, 상이한 이미징 조건들(예컨대, 스케일, 조명, 방향 등) 하에서 다수의 뷔들보다 두드러질 공간이 있는 두드러진 특징들의 사용은 특징 대응 결정을 가능하게 한다.

[0030] [0033] 따라서, 특징 기반 추적 방법들에서는, 이미지 프레임에서 하나 또는 그 초과의 오브젝트들이 식별될 수 있고, 레퍼런스의 하나 또는 그 초과의 프레임들에 대한 카메라의 포즈를 결정하기 위해 다수의 이미지 프레임들에 걸쳐 오브젝트들이 추적될 수 있다. 추적되는 오브젝트(들)는 타겟 또는 추적 타겟을 구성할 수 있다. 몇몇 실시예들에서, 이미지의 오브젝트(들)로부터 추출된 또는 그와 연관된 특징 지점들을 사용하여 타겟이 추적될 수 있다. 예컨대, 이미지 프레임이 다수의 오브젝트들을 포함할 수 있고, 관심 지점들은 이미지 프레임, 이미지 프레임의 개별 오브젝트들 및/또는 오브젝트들의 그룹과 연관될 수 있다. 용어 "내추럴"은, 어떠한 인공의 또는 외부의 장면 마커들도 없이, 캡처된 이미지 프레임에서 정상적으로 발생하는 특징 지점들을 지칭하는 데 사용된다.

[0031] [0034] 몇몇 실시예들에서, 개시되는 방법들은 실시간으로 또는 거의 실시간으로 수행될 수 있다. 몇몇 실시예들에서, 초기 현재 이미지 프레임을 사용하여 텍스트 인식 프로세스가 시작될 수 있다. 이미지 프레임의 맥락

에서 본원에서 사용되는 바와 같은 용어 "현재"는, 카메라(110)에 의해 캡쳐되고 사용자에 의한 동시 뷰잉을 위해 이용 가능한 이미지 프레임을 지칭한다. 용어들 "초기", "이전" 및 "후속"은, 서로에 대한 또는 제3 이미지 프레임에 대한 이미지 프레임들의 시간상 포지션들을 지칭한다. 몇몇 실시예들에서, 텍스트가 예컨대 OCR 프로세스를 통해 식별 및 인식되었던 초기 현재 이미지 프레임이 이미지 타겟 생성 프로세스를 위한 기초로서 저장 및 사용될 수 있다. 텍스트가 식별 및 인식되었던 이미지 프레임이 레퍼런스 프레임으로 불린다.

[0032] [0035] 몇몇 실시예들에서, 이미지 타겟 생성 프로세스는 그 다음, 레퍼런스 프레임에 기초하여 추적 모델을 형성하는데 사용될 수 있다. 몇몇 실시예들에서, 텍스트를 포함하는 레퍼런스 프레임 및 하나 또는 그 초과의 후속 이미지 프레임들의 다양한 내추럴 특징들을 사용하여 텍스트를 추적하는데 SLAM(Simultaneous Localization And Mapping) 기술들, 예컨대 VSLAM(Visual SLAM)이 사용될 수 있다. 몇몇 실시예들에서, 텍스트를 추적하는데 현재 이미지와 레퍼런스 프레임 사이의 특징 대응이 사용될 수 있다. 몇몇 실시예들에서, PTAM(Parallel Tracking and Mapping)과 같은 SLAM 변형들이 추적을 위해 사용될 수 있다.

[0033] [0036] VSLAM에서는, 특징들의 포지션들 및 카메라의 모션 둘 다를 결정하기 위해 다수의 비디오 프레임들에 걸쳐 이미지 패치들의 현저한 내추럴 특징들이 추적된다. SLAM 기반 기술들은, 환경의 3-D(3-Dimensional) 구조물을 결정하면서 동시에, 카메라(110)의 방향(예컨대, 각방향 롤(angular orientation roll), 피치(pitch) 및 요(yaw)) 및 3D 포지션(예컨대, x, y, z 좌표들)을 포함하는 6-DOF(6-Degrees of Freedom) 카메라 포즈의 추적을 허용한다.

[0034] [0037] 몇몇 실시예들에서, 이미지 프레임에서 코너 지점들, 에지들 등과 같은 특징들을 식별하는데 SIFT(Scale Invariant Feature Transform) 기술들이 사용될 수 있다. SIFT-기반 디스크립터들, 및 다양한 계산적으로 효율적인 수정된 SIFT 디스크립터들을 포함하는 이들의 변형들은, 비교적 큰 뷰포인트 변화들에도 불구하고, 견고한 추적 및 이미지를 사이의 특징 대응을 가능하게 한다. 몇몇 실시예들에서, 이미지 프레임의 키포인트들이 식별될 수 있고, 이미지 특징들을 추적하기 위해 개별 키포인트들에 대응하는 로컬 디스크립터들이 형성 및 사용될 수 있다. "패치" 또는 "이미지 패치"는, 추적을 위해 사용될 수 있는 특징 지점 주위의 픽셀들의 구역이다.

[0035] [0038] 위에서 개요가 설명된 기술들은 단지 예들이고, 몇몇 실시예들에서, 특징 검출을 위해 다양한 다른 기술들이 사용될 수 있다. 예컨대, 당업자에게 명백할 바와 같이, "FAST(Features from Accelerated Segment Test)" 또는 그 변형들, "SURF(Speeded-Up Robust Features)" 또는 그 변형들, 하이브리드 지점 및/또는 에지 검출 기술들 등에 기초한 기술들이 적절한 수정들로 특징 검출을 위해 사용될 수 있다.

[0036] [0039] 추가로, 몇몇 실시예들에서, 레퍼런스 이미지 프레임과 하나 또는 그 초과의 후속 이미지 프레임들 사이의 상대 모션을 컴퓨팅하는데 이미지 정렬 기술이 사용될 수 있다. 이미지 정렬 프로세스는 오브젝트 모션을 추정하기 위해 모션 파라미터들에 대한 레퍼런스 이미지 프레임과 현재 후속 이미지 프레임 사이의 정렬의 척도(measure)를 컴퓨팅한다. 예컨대, 레퍼런스 프레임과 현재 후속 프레임 사이의 차이를 반복적으로 최소화함으로써 레퍼런스 프레임에 대한 현재 후속 프레임의 정렬 및/또는 상대 모션을 컴퓨팅하는데 ESM(Efficient Second-order Minimization)이 사용될 수 있다.

[0037] [0040] 추적 모델 및/또는 이미지 정렬 프로세스들은, 인식된 텍스트의 포지션, 및 레퍼런스 프레임에 대한 현재 및/또는 후속 이미지 프레임의 포즈를 각각 산출할 수 있다.

[0038] [0041] 몇몇 실시예들에서, 그 다음, 추적 모델로부터 획득된, 인식된 텍스트의 포지션, 및/또는 이미지 정렬 프로세스들로부터 획득된, 레퍼런스 프레임에 대한 현재 후속 이미지 프레임의 포즈를 사용하여, 패치 추적기가 시작 및 초기화될 수 있다. 용어 이미지 패치는 일반적으로, 이미지에 있는 특징 지점 주위의 구역에 대응하는 이미지의 일부분을 지칭한다. 패치 추적 알고리즘은, 텍스트에 대한 카메라 포즈를 계산하기 위해, 레퍼런스 이미지 패치와 현재 이미지로부터 추출된 현재 이미지 패치 사이의 특징 지점들을 매칭시킬 수 있다.

[0039] [0042] 예컨대, 많은 증강 현실(AR:Augmented Reality) 애플리케이션들에서는, 텍스트 또는 그래픽 엘리먼트들과 같은 하나 또는 그 초과의 가상 오브젝트들이 라이브 카메라 뷰 위에 중첩될 수 있고 디스플레이(145) 상에서 디스플레이될 수 있다. 몇몇 실시예들에서, 카메라(110)에 의해 캡쳐된 라이브 이미지들, AR 이미지들, GUI(Graphical User Interface)들, 프로그램 출력 등을 디스플레이하는데 디스플레이(145)가 사용될 수 있다. 디스플레이(145)는 또한 스크린으로 지칭된다. 가상 오브젝트들의 정확한 정합은, 가상 오브젝트를 렌더링할 때 디스플레이(145) 상의 라이브 비디오 화상에서 가상 텍스트 및/또는 그래픽스 오브젝트들의 정확한 배치를 보장할 수 있다. 따라서, 몇몇 실시예들에서, 텍스트 블록들이 인식될 수 있고 텍스트 인식 이후 하나 또는 그

초과의 애플리케이션들에 의해 프로세싱될 수 있도록, 카메라(110)에 의해 스캐닝되고 있는 텍스트가 추적될 수 있다.

[0040] [0043] 용어 "텍스트 블록"은 캐릭터(character)들의 시퀀스, 예컨대, 워드, 워드들의 시퀀스, 및/또는 텍스트의 하나 또는 그 초과의 라인들을 지칭하는데 사용되지만, 이들로 제한되는 것은 아니다. 용어 "캐릭터"는 일반적으로, 레터들, 숫자(number)들, 및 언어와 관계없이 다양한 다른 심볼들을 지칭하는데 사용된다. 몇몇 실시예들에서, 텍스트 검출 및 추적의 결과들에 기초하여 하나 또는 그 초과의 가상 오브젝트들이 텍스트 블록들 위에 오버레이될 수 있다. 예컨대, AR 현장(in-situ) 텍스트 번역 애플리케이션에서는, 매체 상의 캐릭터들이 OCR 프로세스를 통해 추적 및 인식될 수 있고, 다른 언어로 번역될 수 있으며, 그리고 본래 텍스트가 번역된 텍스트를 표현하는 캐릭터들과 오버레이될 수 있고, 이로써 번역된 텍스트의 라이브 카메라 뷰가 가능하게 된다. 예컨대, AR 애플리케이션들을 실행할 때, 프로세서(150)는 가상 오브젝트를 라이브 카메라 뷰 안에 배치할 수 있고, 카메라의 포즈에 대응하도록, 가상 오브젝트는 캡쳐된 이미지에서 회전 및/또는 변위될 수 있다.

[0041] [0044] 몇몇 실시예들에서, 이미지에 대한 MS(100) 및/또는 카메라(110)의 포즈를 결정하는데 이미지의, 하나 또는 그 초과의 내추럴 및/또는 텍스트 기반 특징들이 사용될 수 있다. 몇몇 실시예들에서, 이미지의 특정한 특징 지점들, 예컨대 높은 대비 차이(contrast difference)를 갖는 지점들, 에지들, 및/또는 코너 지점들이 프레임에서 검출될 수 있고, 프레임들에 걸쳐 카메라(110)의 포즈를 추적하는데 사용될 수 있다. 몇몇 실시예들에서, 프로세서들(150)은 특징 지점들에 관련된 핵심-관련 정보를 저장할 수 있다. 예컨대, 특징 지점 핵심들과 연관된 루미넌스 그래디언트 값들 및/또는 루미넌스 그래디언트 값들의 함수들이 메모리(160)에 저장될 수 있다. 몇몇 실시예들에서, 추적은, 인식된 텍스트의 이미지들에서 발생하는 코너 지점들과 같은 내추럴 특징들에 부분적으로 기초할 수 있다.

[0042] [0045] 몇몇 실시예들에서, 이미지에 대한 카메라(110)의 포즈는, IMU(130)로부터의 입력에 부분적으로 기초하여 결정 또는 정정될 수 있다. 몇몇 실시예들에서, 카메라 포즈 결정을 돋기 위해 그리고/또는 개선하기 위해, 카메라(110)의 알려진 고유한 파라미터들 및 특성들, 예컨대 렌즈의 초점 길이, 카메라 초점 거리 등, 및/또는 다른 센서들로부터의 입력이 IMU 입력 및/또는 다른 텍스트 추적 방식들과 함께 사용될 수 있다.

[0043] [0046] 몇몇 실시예들에서, 텍스트 추적은, 엘리먼트들이 현재 스캐닝되고 있는 상태에서, 이전에 스캐닝된 텍스트 블록들의 기하학적 포지션들 사이의 상관(correlation)을 가능하게 할 수 있다. 텍스트 추적은 스캐닝된 텍스트 블록들 사이의 기하학적 관계들을 결정하는데 사용될 수 있고, 이로써 스캐닝된 이미지/프레임 시퀀스로부터 스캐닝된 텍스트의 구조의 재구성이 허용된다.

[0044] [0047] 도 2는 매체 상의 워드들에 대한 카메라(110)의 포지션들, 및 텍스트 블록들의 포지션을 추적 및/또는 유지하는데 사용될 수 있는 좌표계들을 도시한다. 예컨대, 카메라(110)는 카메라 포지션(230-1)으로부터 텍스트 블록-1(210-1)을 포함하는 초기 이미지 프레임을 캡쳐할 수 있다. 카메라 포지션(230-1)에서의 카메라(110)의 시야는 점선들(250) 내의 구역이다.

[0045] [0048] 몇몇 실시예들에서, 텍스트 블록(210-1)의 하나 또는 그 초과의 캐릭터들 또는 워드들을 식별하기 위해 그리고 레퍼런스(275)의 프레임에 대한 텍스트 블록(210-1)의 초기 포즈(260)를 결정하기 위해, OCR을 사용함으로써, 카메라 포지션(230-1)에서 캡쳐된 이미지 프레임이 프로세싱될 수 있다. 몇몇 실시예들에서, 카메라 포지션(230-1)에서의 레퍼런스(275)의 프레임은 직교 축들의 세트를 포함할 수 있는데, 축들 중 하나 또는 그 초과가 MS(100)의 몸체 및/또는 카메라(110)의 평면과 정렬될 수 있다. 예컨대, 일 실시예에서, 카메라 포지션(230-1)에서 캡쳐된 이미지 프레임이 레퍼런스 이미지 프레임으로서 저장 및 사용될 수 있다.

[0046] [0049] 몇몇 실시예들에서, 텍스트 블록들(210)의 포즈는 또한, "페이지 좌표계"를 표현할 수 있는 레퍼런스(285)의 프레임을 사용하여 유지될 수 있다. 예컨대, 사용자가 텍스트 블록들(210)이 제시되는 매체에 앞-평행(fronto-parallel)하게 카메라를 잡고 있는 상황에서는, 별개의 경계 박스가 각각의 텍스트 블록(210)과 연관될 수 있다. 몇몇 실시예들에서, 인식된 텍스트 블록(210-1)의 원쪽 하단 지점 및/또는 텍스트 블록(210-1)과 연관된 경계 박스가 페이지 좌표계의 원점으로서 셋팅될 수 있고, 텍스트 블록(210-1)의 베이스라인 및/또는 연관된 경계 박스의 베이스라인이 x-축으로서 셋팅될 수 있다. y-축은 카메라 평면에서 x-축에 수직인 것으로서 정의될 수 있다. 몇몇 실시예들에서, 레퍼런스(285)의 프레임은, 텍스트 블록(210-1)이 인식되는 제1 이미지 프레임에 기초하여 정의될 수 있고, 후속 이미지 프레임들에 대한 페이지 좌표계로서의 역할을 할 수 있다.

[0047] [0050] 몇몇 실시예들에서, 추적 모델은, 레퍼런스 프레임으로부터 시작해, 레퍼런스 이미지 프레임 및 후속 프레임들의 하나 또는 그 초과의 내추럴 특징들 및/또는 특징 지점들을 사용하여 형성될 수 있다. 일단 추적 모

델이 형성되었다면, 추적 모델은, 후속 프레임들에서 인식된 텍스트의 포지션을 결정하는데 사용될 수 있다. 예컨대, 카메라(110)는, 스캐닝 프로세스 동안, 레퍼런스(280)의 프레임 및 점선들(255) 내의 구역으로 도시된 시야를 갖는 카메라 포지션(230-2)으로 이동할 수 있다.

[0048] [0051] 도 2에 도시된 바와 같이, 카메라 포지션(230-2)은, 텍스트 블록들(210)이 제시되는 매체에 대한 카메라(110)의 포지션 및 방향의 변화를 반영한다. 몇몇 실시예들에서, 레퍼런스(275)의 프레임에 대한 카메라(110)의 6-DOF 포즈의 추정치를 획득하는데 내추럴 특징 추적이 부분적으로 사용될 수 있다. 레퍼런스(280)의 프레임에서 인식된 텍스트 블록(210-1)의 포지션을 결정하는데, 레퍼런스(275)의 프레임에서 카메라(110)의 추정된 6-DOF 포즈가 사용될 수 있다.

[0049] [0052] 추가로, 몇몇 실시예들에서, 텍스트 블록 1(210-1)에서 인식된 캐릭터들 또는 그 일부분에 부분적으로 기초하여, 카메라 포지션(230-1)의 레퍼런스 프레임과 카메라 포지션(230-2)의 레퍼런스 프레임 사이의 이미지의 상대 모션을 컴퓨팅하는데 이미지 정렬 기술이 사용될 수 있다. 예컨대, ESM과 같은 이미지 정렬 기술들은, 인식된 텍스트 블록(210-1)의 알려진 추정된 포지션을 사용함으로써 카메라 포지션(230-1)에서 획득된 레퍼런스 이미지에 대한, 카메라 포지션(230-2)에서의 현재 이미지의 모션을 결정하는데 사용될 수 있다.

[0050] [0053] 그 다음, (레퍼런스 프레임이 이전에 생성되었다면) 이전 레퍼런스 프레임 및/또는 이미지 정렬 프로세스에 의해 측정된 이미지의 상대 모션을 사용하여 폐치 추적기가 초기화될 수 있다. 폐치 추적기는, 텍스트에 대한 개선된 카메라 포즈를 계산하기 위해, 레퍼런스 이미지 폐치와 현재 이미지로부터 추출된 현재 이미지 폐치 사이의 특징 지점들을 매칭할 수 있다.

[0051] [0054] 몇몇 실시예들에서, 예컨대, 포지션(230-1)에서의 이미지 프레임과 카메라 포지션(230-2)에서 캡쳐된 이미지 프레임 사이에 불충분한 오버랩이 있을 때(또는 레퍼런스 프레임이 아직 생성되지 않았을 때), 그 다음, 텍스트 블록(210-1) 및 텍스트 블록(210-2)의 하나 또는 그 초과의 캐릭터들 또는 워드들을 식별하기 위해, OCR을 사용함으로써, 포지션(230-2)에서 캡쳐된 이미지 프레임이 프로세싱될 수 있다. 레퍼런스(280)의 프레임에 대한 텍스트 블록(210-2)의 포즈(270)는, 텍스트 블록 1(210-1)의 알려진 포즈(265)에 기초하여 그리고 ESM 또는 다른 이미지 정렬 기술에 의해 결정된 바와 같이 포지션들(230-1 및 230-2) 사이의 카메라(110)의 상대 모션 및 방향에 기초하여 결정될 수 있다. 몇몇 실시예들에서, ESM은 후속 이미지 프레임의 더 낮은 해상도 버전에 대해 동작할 수 있다. 추가로, 레퍼런스(285)의 프레임 면에서 텍스트 블록(210-2)의 포즈를 결정하는데 포지션들(230-1 및 230-2) 사이의 카메라(110)의 컴퓨팅된 상대 모션 및 방향이 또한 사용될 수 있다. 따라서, 제1 텍스트 블록과 제2 텍스트 블록 사이의 기하학적 관계를 결정하는데 제2 텍스트 블록에 대한 카메라의 포지션이 사용될 수 있다. 따라서, 몇몇 실시예들에서, 텍스트 블록들(210)의 포지션들은 두 개의 방식들: (i) 레퍼런스(275 및/또는 280)의 프레임과 같은 카메라 좌표계 내에서의 상대 포즈 및 (ii) 페이지 좌표계(285) 내에서의 상대 포즈로 표현될 수 있다.

[0052] [0055] 도 3은 개시되는 실시예들과 일치하는 텍스트 검출 및 추적의 예시적 방법(300)에 대한 흐름도를 도시한다. 도 3을 참조하면, 몇몇 실시예들에서, 방법(300)은 MS(100) 상에서 실행되는 컴퓨터 비전 및/또는 AR 애플리케이션에 의해 호출될 수 있다. 몇몇 실시예들에서, 예시적 방법(300)은 305에서 시작할 수 있고, 305에서는, 다양한 초기화 및 하우스키핑 루틴들이 수행될 수 있다. 다음 차례로, 310에서는, 제1 또는 다음 차례의 스캐닝되는 이미지 프레임이 획득될 수 있다. 예컨대, 사용자는, 페이지와 같은 비교적 평편한 표면 상의 텍스트에 걸쳐 MS(100) 상의 카메라(110)를 스와이핑(swiping)함으로써, 텍스트를 스캐닝하고 그리고/또는 이미지 프레임들을 캡쳐할 수 있고, 이는 스캐닝된 텍스트에 대응하는 하나 또는 그 초과의 이미지들 또는 프레임들을 생성한다. 예컨대, 스캐닝 프로세스 동안, 사용자는 여러 이미지 프레임들의 시간상 시퀀스를 포함하는 비디오를 캡쳐할 수 있는데, 즉 캡쳐된 프레임들은 비디오 시퀀스의 일부를 형성할 수 있다.

[0053] [0056] 320에서는, 325에서 OCR 프로세스를 개시하기 위해, 트리거 이벤트(320에서의 "예")가 사용될 수 있다. 트리거들은, 예컨대, 카메라 뷰에서 이전에 스캐닝되지 않았던 영역의 검출을 포함할 수 있다. 몇몇 실시예들에서, 제1 추적 타겟 및/또는 이전 레퍼런스 이미지에 대한 제2의 스캐닝된 이미지 프레임의 일부분 사이의 오버랩의 부족(lack)으로부터, 스캐닝되지 않은 영역이 검출될 수 있다. 예컨대, 오버랩의 부족이 임계치를 초과하는 경우, 325에서의 OCR이 트리거링될 수 있다. 예컨대, 320에서, 현재 이미지 프레임과 제1 레퍼런스 프레임 사이의 오버랩 영역의 퍼센티지가 임계치 퍼센티지 미만일 경우, 방법은 325로 진행할 수 있다.

[0054] [0057] 몇몇 실시예들에서, 제1 레퍼런스 프레임과 후속 이미지 프레임 사이의 상대 모션이 임계치를 초과할 때, 325에서 OCR이 트리거링될 수 있다. 몇몇 실시예들에서, 제1 레퍼런스 프레임과 후속 이미지 프레임 사이의 상대 모션은, 후속 이미지 프레임에서 이전에 인식된 텍스트 블록의 포지션에 부분적으로 기초하여 그리고/

또는 ESM 또는 다른 이미지 정렬 기술을 사용하여 컴퓨팅될 수 있다. 예컨대, 제1 레퍼런스 프레임에 대한 현재 이미지 프레임의 회전 규모(magnitude), 및/또는 제1 레퍼런스 프레임에 대한 현재 이미지 프레임의 병진 (translation) 규모가 어떤 임계치를 초과하는 경우, 325에서 OCR이 트리거링될 수 있다. 몇몇 실시예들에서, 레퍼런스 프레임이 없을 경우, 예컨대, 새로운 이미지가 스캐닝되고 있을 때 그리고/또는 텍스트 스캐닝 또는 AR 애플리케이션이 처음으로 시작될 때, 그러면, 325에서 OCR이 자동으로 트리거링될 수 있다.

[0055] [0058] 일 실시예에서, 최종 레퍼런스 프레임에 대한 "새로운" 카메라 이미지 내의 영역은 새로운 영역에서 추적되는 특징들의 부재에 기초하여 결정될 수 있다. 몇몇 실시예들에서, 새로운 영역이 이미지 프레임의 임계치 퍼센티지 또는 프랙션(fraction)을 초과할 때마다, 방법은 325로 진행할 수 있다. 다른 실시예에서, 레퍼런스 카메라 프레임들이 페이지 좌표계(285) 상으로 투영(projecting)될 수 있고 현재 카메라 포즈에 기초하여 정렬될 수 있으며, 페이지 평면 상의 "새로운" 영역이 임계치 영역, 또는 현재 레퍼런스 프레임에 의해 커버되는 페이지 평면 상 영역의 퍼센티지를 초과할 때마다, 325에서의 새로운 텍스트/OCR 인식 이벤트가 트리거링될 수 있다.

[0056] [0059] 추가 실시예에서, 이전에 인식된 텍스트 블록 또는 최종 추적된 워드들 중 하나의 어떤 프랙션이 현재 카메라 프레임의 바깥에 있을 때, 325가 트리거링될 수 있다. 예컨대, 일 실시예에서, 현재 이미지 프레임에서 보이는 제1 텍스트 블록의 프랙션이 어떤 임계치 미만일 때, 325가 트리거링될 수 있다. 제1 텍스트 블록의 보이는 프랙션은, 현재 이미지 프레임에서 제1 텍스트 블록의 보이는 부분을 포함하는 영역 대 제1 레퍼런스 프레임에서 제1 텍스트 블록의 총 영역의 비로서 결정될 수 있다.

[0057] [0060] 몇몇 실시예들에서, 어떠한 트리거도 검출되지 않을 경우(320에서 "아니오"), 알고리즘은 355로 진행할 수 있는데, 355에서는, 추적 타겟의 생성을 시작/지속하기 위한, 또는 기존 타겟을 사용하여 텍스트 추적 프로세스를 계속하기 위한 결정이 이루어질 수 있다.

[0058] [0061] 320에서 트리거 이벤트가 검출되는 경우(320에서 "예"), 325에서는, 이미지 프레임에서 스캐닝된 텍스트에 대응하는 새로운 워드들/캐릭터들을 발견하고 인식하기 위해 이미지 프레임이 프로세싱될 수 있다. 예컨대, 이미지 프레임을 프로세싱하기 위해, 예시적 방법(300)과 연관된 OCR 애플리케이션이 호출될 수 있다. 몇몇 실시예들에서, 325에서의 OCR 애플리케이션이 텍스트 블록, 예컨대 텍스트 블록들(210-1 및/또는 210-2)에 대해 동작할 수 있다. 몇몇 실시예들에서, 텍스트 블록의 크기는 다양한 파라미터들, 예컨대 이미지 프레임에서 보이는 텍스트의 양, 이용 가능한 메모리(160), OCR 애플리케이션의 스피드, AR 애플리케이션의 타입, 프로세서들(150)의 개수 및 스피드 등에 기초하여 결정될 수 있다. 몇몇 실시예들에서, OCR 애플리케이션은 아웃 포커스(out-of-focus) 이미지를 및 사용자 손 떨림들로부터의 모션 블러를 보상할 수 있다.

[0059] [0062] 330에서는, OCR 프로세스가 성공했는지가 결정될 수 있다. OCR 프로세스가 성공적이지 않았을 경우(330에서 "아니오"), 방법은 310으로 리턴할 수 있는데, 310에서는, 다음 차례의 이미지 프레임이 획득된다. 몇몇 실시예들에서, 오류에 대한 시각적 또는 청각적 표시가, MS(100)를 이용하여 텍스트를 스캐닝하는 사용자에게 제공될 수 있다. 몇몇 실시예들에서, 여러 이미지들/프레임들에 걸쳐 OCR 애플리케이션에 의해 어떠한 워드들도 검출되지 않을 경우, 카메라(110)의 자동초점이 트리거링될 수 있다.

[0060] [0063] 몇몇 실시예들에서, OCR 프로세스가 성공했을 경우(330에서 "예"), 335에서는, 특정 카메라 이미지/프레임 – 이 특정 카메라 이미지/프레임에 대해, OCR이 330에서 성공적으로 수행되었고, 이 특정 카메라 이미지/프레임이 인식된 텍스트 블록을 포함함 – 이 레퍼런스 프레임으로서 저장될 수 있다. 몇몇 실시예들에서, 레퍼런스 프레임은 메모리(160)에 저장될 수 있다. 몇몇 실시예들에서, 추적 시스템의 다른 모듈들을 초기화하는데 레퍼런스 프레임이 이후에 활용될 수 있다.

[0061] [0064] 340에서는, 현재 카메라 포지션에 대한 새로운 워드들 및/또는 텍스트 블록들의 포즈가 결정될 수 있다. 예컨대, 레퍼런스(280)의 프레임에 대한 텍스트 블록(210-2)의 상대 포지션이 도출될 수 있다. 몇몇 실시예들에서, 부분적으로, 적어도 하나의 다른 이전에 인식된 텍스트 블록, 예컨대 텍스트 블록 1(210-1)을 검출함으로써, 새로운 텍스트 블록(210-2)의 포즈가 결정될 수 있다. 예컨대, 몇몇 실시예들에서, 텍스트 블록(210-1)에서 인식된 캐릭터들에 부분적으로 기초하여, 이미지의 상대 모션을 컴퓨팅하는데 ESM과 같은 이미지 정렬 기술이 사용될 수 있다. 몇몇 실시예들에서, ESM은 후속 이미지 프레임의 더 낮은 해상도 버전에 대해 동작할 수 있다. 따라서, 현재 카메라 포지션에 대한 이전에 인식된 텍스트 블록의 포즈가 정렬에 기초하여 결정될 수 있다. 따라서, 현재 카메라 포지션에 대한 새롭게 인식된 워드들의 포즈는, 이전에 인식된 텍스트 블록들의 결정된 포지션들에 기초하여 결정될 수 있다. 따라서, 제1 텍스트 블록과 제2 텍스트 블록 사이의 기하학적 관계를 결정하는데 제2 텍스트 블록에 대한 카메라의 포지션이 사용될 수 있다. 알려진 그리고 새롭게 인식된 텍스트

블록들의 결정된 포즈들에 기초하여, 페이지 좌표계(285)에서 새로운 워드들의 포즈들이 도출될 수 있고, 이로써 스캐닝된 이미지/프레임 시퀀스로부터 스캐닝된 텍스트의 구조의 재구성이 가능하게 된다.

[0062] [0065] 다음 차례로, 345에서는, 새로운 워드들이 그들의 포즈들에 따라 텍스트 표현에 삽입될 수 있다. 예컨대, 일 실시예에서, 각각의 인식된 텍스트 블록은, 페이지 좌표계(285)의 좌표들 및 경계 박스와 연관될 수 있고, 페이지 좌표계(285)에서의 인식된 텍스트 블록의 좌표들에 기초하여 텍스트 표현에 삽입될 수 있다. 그 다음, 방법은 375로 진행할 수 있다. 몇몇 실시예들에서, 텍스트 표현을 저장하는데 적절한 데이터 구조가 사용될 수 있다.

[0063] [0066] 몇몇 실시예들에서, 어떠한 트리거도 검출되지 않을 경우(320에서 "아니오"), 알고리즘은 355로 진행할 수 있는데, 355에서는, 내추럴 특징 타겟이 구성되었는지 그리고 완전한지의 여부의 결정이 이루어진다. 내추럴 특징 타겟이 구성되었고 완전한 경우(355에서 "예"), 방법은 370으로 진행한다.

[0064] [0067] 추적 타겟, 예컨대 내추럴 특징 추적 타겟이 구성되지 않았고 불완전하다면(355에서 "아니오"), 360에서, 추적 타겟(예컨대, 내추럴 특징 추적 타겟)을 생성하는 프로세스가 시작 또는 계속될 수 있다. 몇몇 실시예들에서, 내추럴 특징 추적 타겟의 생성은, 하나 또는 그 초과의 이미지 프레임들의 텍스트 기반 특징들을 비롯한 하나 또는 그 초과의 내추럴 특징들을 추출하는 것을 포함할 수 있다. 이미지에 대해 추출된 관심 지점들의 개수가 어떤 임계치를 초과하는 것으로 또는 어떤 원하는 신뢰 레벨을 충족하는 것으로 결정될 때, 타겟의 구성은 완전한 것으로 여겨질 수 있다. 몇몇 실시예들에서, 이러한 내추럴 특징 추적 타겟의 생성은 하나보다 많은 카메라 프레임을 취할 수 있는데, 그 이유는 식별되는 특징들의 개수를 증가시키는 것이 검출된 특징들을 사용하는 이후의 추적의 정확성을 증가시키기 때문이다. 몇몇 실시예들에서, 내추럴 특징 추적 타겟의 생성은 방법(300)의 다른 엘리먼트들과 동시에 수행될 수 있다.

[0065] [0068] 몇몇 실시예들에서, 이미지의 특정한 특징 지점들, 예컨대 높은 대비 차이를 갖는 지점들 및/또는 코너 지점들이 프레임에서 검출될 수 있고, 프레임들에 걸쳐 카메라(110)의 포즈를 추적하는데 사용될 수 있다. 몇몇 실시예들에서, 프로세서들(150)은 특징 지점들에 관련된 픽셀-관련 정보를 저장할 수 있다. 예컨대, 특징 지점 픽셀들과 연관된 그래디언트 값들 및/또는 그래디언트 값들의 함수들이 메모리(160)에 저장될 수 있다. 몇몇 실시예들에서, 레퍼런스 프레임의 키포인트들의 특징 벡터들이 특징 데이터베이스에 저장될 수 있다. 몇몇 실시예들에서, 내추럴 특징 추적은, 개별 가중치들을 복수의 특징 지점들에 할당하는 가중 방식을 사용할 수 있다. 몇몇 실시예들에서, 각각의 특징 지점에 대한 가중치는, 하나 또는 그 초과의 텍스트 블록들에 대한 특징 지점의 포지션에 부분적으로 기초할 수 있다. 예컨대, 가중 방식은 텍스트 블록 상에 센터링될 수 있고, 텍스트 블록 상에 위치된 특징들을 카메라 이미지 프레임의 다른 특징들보다 우선할 수 있다.

[0066] [0069] 몇몇 실시예들에서, 추출된 특징들은 카메라 평면에 평행한 평면 상에 놓이는 것으로 가정될 수 있다. 예컨대, 텍스트 블록들이 평면의, 거의 평면의 또는 비교적 평면한 표면(예컨대, 잡지 페이지, 책, 신문, 제품 패키지 또는 도로명 계시판) 상에 나타나고 카메라가 텍스트에 앞-평행하게 유지되고 있는 –이는, 더 나은 텍스트 인식을 가능하게 함 – 실시예들에서는, 모델을 형성하는데 내추럴 특징 추적이 사용될 수 있고, 이 모델에서는, 추출된 특징들이 카메라 평면에 평행한 평면 상에 놓이는 것으로 가정될 수 있다. 몇몇 실시예들에서, 평면의 타겟의 경우, 방법은 변하는 조명 조건들 및 높은 견고성을 갖는 6-DOF로 타겟을 추적할 수 있다.

[0067] [0070] 몇몇 실시예들에서, 360의 완료 이후, 또는 내추럴 특징 추적 타겟이 존재하고 완전한 경우, 방법은 370으로 진행할 수 있다. 370에서는, 카메라 포지션(230-1)에서 획득된 레퍼런스 이미지에 대한 230-2에서의 현재 이미지의 모션을 결정하는데 ESM과 같은 이미지 정렬 기술들이 사용될 수 있다. 몇몇 실시예들에서, ESM과 같은 이미지 정렬/정합 기술들은, 텍스트 인식을 트리거링하는데 사용되었던 레퍼런스 이미지에 대한 카메라의 상대 모션의 측정을 가능하게 한다. 몇몇 실시예들에서, 카메라에 의해 캡쳐된 이미지들의 더 낮은 해상도 버전들이 ESM으로의 입력으로서 사용될 수 있고, 시간이 일정한 방식으로 프레임 간 회전(inter-frame rotation)이 컴퓨팅될 수 있다.

[0068] [0071] 375에서는, 레퍼런스 프레임에 대한, 현재 프레임의 인식된 워드들의 새로운 또는 개선된 포즈들이 컴퓨팅될 수 있다. 몇몇 실시예들에서, 패치 추적기가 시작될 수 있고, 기준의 또는 새롭게 생성된 이미지 타겟을 이용하여 초기화될 수 있으며, 이미지 정렬(370)에 의해 컴퓨팅된 현재 포즈를 이용하여 초기화될 수 있다. 몇몇 실시예들에서, 패치 추적기는, 입력으로서 (i) 추적되는 특징들을 갖는 레퍼런스 프레임 및/또는 (ii) 레퍼런스 이미지에 대한 카메라의 상대 모션을 수신할 수 있다. 몇몇 실시예들에서, 패치 추적기는 레퍼런스 프레임에 대한, 현재 프레임의 인식된 워드들의 포즈들을 컴퓨팅할 수 있다. 예컨대, 패치 추적기는, 텍스트에 대한 카메라 포지션을 계산하기 위해, 레퍼런스 이미지 패치와 현재 이미지로부터 추출된 현재 이미지 패치 사이

에 특정 지점들을 매칭할 수 있다. 예컨대, 페치 추적기는 레퍼런스(280)의 프레임에 대한 텍스트 블록 1(210-1)의 포즈(265)를 결정할 수 있다. 몇몇 실시예들에서, 레퍼런스의 키포인트들의 저장된 특정 벡터들이 현재 프레임의 대응하는 특정 벡터와 매칭될 수 있다. 따라서, 예컨대, 좌표계(275) 및/또는 페이지 좌표계(285)에 기초하여, 카메라에 대한 새로운 워드들의 포즈들이 결정될 수 있다.

[0069] [0072] 380에서는, 사용자가 디스플레이된 카메라 이미지의 현재 텍스트 표현에 관련된 퍼드백을 제공받을 수 있다. 예컨대, 인식된 텍스트는 강조될 수 있거나, 폰트 변화를 겪을 수 있거나, 판독될 수 있거나, 번역되고 그리고 스캐닝된 텍스트 위에 오버레이될 수 있거나, 또는 다른 방식으로 디스플레이를 통해 사용자에게 표현될 수 있다. 몇몇 실시예들에서, 보는 또는 들리는 퍼드백이 제공될 수 있다. 예컨대, 인식된 워드들은 텍스트-투-스피치(text-to-speech) 애플리케이션에 의해 판독될 수 있다. 몇몇 실시예들에서, 오류가 발생할 때 퍼드백이 또한 제공될 수 있다.

[0070] [0073] 몇몇 실시예들에서, 385에서는, 입력되고 있는 부가 프레임들이 있을 경우(385에서 "예"), 방법은 다른 반복을 시작하기 위해 310으로 리턴한다. 어떠한 부가 프레임들도 없을 경우(385에서 "아니오"), 프로세스는 390에서 종료한다.

[0071] [0074] 몇몇 실시예들에서, 방법(300)은 AR 애플리케이션에 의해 호출될 수 있다. 몇몇 실시예들에서, 부가 이미지 프레임들이 캡쳐되고 있는 동안, 이미지들로부터 포즈 정보를 결정하기 위한 계산들이 백그라운드에서(즉, 백그라운드 계산 프로세스로서) 수행될 수 있다. 몇몇 실시예들에서, SLAM(simultaneous localization and mapping) 프로시저 방법을 사용하여 카메라 이동 또는 포즈 조정의 결정이 수행될 수 있다. 예컨대, 텍스트의 부분을 추적하고 그리고/또는 카메라에 대한 텍스트 또는 텍스트의 다른 부분의 상대 포지션을 결정하기 위하여, 블록들(360 및/또는 375) 중 하나 또는 그 초과에서 SLAM 프로시저가 사용될 수 있다.

[0072] [0075] 몇몇 실시예들에서, 방법(300)은 MS(100) 상의 카메라(110), 프로세서(150), 메모리(160) 및/또는 IMU(130)의 어떤 결합을 사용하여 수행될 수 있다. 다른 실시예들에서, 방법(300)의 일부분들이 서버 상에서 또는 MS(100)에 무선으로 커플링된 다른 컴퓨팅 상에서 수행될 수 있다. 몇몇 실시예들에서, 서버가 MS(100)로부터 이미지를 수신하고, 이미지를 프로세싱하고, 그리고 예시적 방법(300)을 호출하는 애플리케이션에 의해 지정된 모바일 스테이션들의 세트에 프로세싱의 결과들을 송신할 수 있다. 예컨대, 현장 텍스트 번역 시나리오에서는, 사용자가 텍스트를 스캐닝할 수 있고, 이 텍스트는, OCR을 사용하여 검출되고, 추적되고, 다른 언어로 번역되고 그 다음 텍스트 추적에 기초하여 본래 텍스트 위에 오버레이되고, 그리고 여러 사용자들이 "라이브 카메라" 뷰에서 번역된 텍스트를 보고/공유하도록 허용하기 위해, 하나 또는 그 초과의 사용자들에게 송신될 수 있다. 다른 실시예에서, 텍스트는, MS(100)를 사용하여 스캐닝되고, 예시적 방법(300)을 사용하여 검출 및 추적되며, 텍스트-투-스피치 애플리케이션이 MS(100) 상의 또는 그에 커플링된 스피커를 사용하여 텍스트를 판독할 수 있다. 예컨대, 텍스트-투-스피치 애플리케이션은, 시각장애인들에 의해, 또는 텍스트가 너무 작아서 쉽게 읽을 수 없는 상황들에서, 또는 아직 읽도록 학습 받지 않았을 수 있는 사람들에 의해 사용될 수 있다. 몇몇 실시예들에서, 내추럴 특징 추적의 사용은, 종이의 빈 시트 상에 텍스트의 짧은 워드들까지도 추적하는 것을 가능하게 할 수 있다. 추가로, 텍스처링된 재료의 사용은, 텍스트 추적 시 더 큰 견고성을 허용할 수 있다.

[0073] [0076] 도 4는 개시되는 실시예들과 일치하는 텍스트 추적을 위한 시스템의 예시적 아키텍처(400)를 도시한다. 예시적 아키텍처(400)의 설명은 예로서만 제공되며, 당업자에게 명백할 바와 같이, 아키텍처에 대한 다양한 수정들이 개시되는 실시예들과 일치하는 방식으로 이루어질 수 있다. 예컨대, 도 4에 도시된 하나 또는 그 초과의 블록들 및 모듈들의 기능들이 다양한 방식들로 결합, 재분배 및/또는 수정될 수 있다.

[0074] [0077] 몇몇 실시예들에서, MS(100) 상의 하드웨어, 펌웨어 및/또는 소프트웨어 또는 이들의 어떤 결합을 사용하여 워드 인식 모듈(410), 워드 추적 모듈(420), 텍스트 어셈블러(430) 및 퍼드백 비주얼라이저(460) 모듈이 구현될 수 있다. 몇몇 실시예들에서, 위의 모듈들 중 하나 또는 그 초과의 일부분들이, 메모리(160)에 저장된, 그리고/또는 MS(100)에 커플링된 컴퓨터-판독가능 미디어에서 구현된 코드를 사용하여 프로세서들(150) 상에서 실행될 수 있다.

[0075] [0078] 몇몇 실시예들에서, 시스템(400)은 워드 인식 모듈(410), 워드 추적 모듈(420), 텍스트 어셈블러(430) 및 퍼드백 비주얼라이저(460)를 포함할 수 있다. 몇몇 실시예들에서, 워드 인식 모듈(410)은 OCR 애플리케이션을 포함할 수 있고, OCR 애플리케이션은 카메라(110)에 의해 캡쳐된 이미지들의 텍스트 블록들을 인식할 수 있다. 추가로, 몇몇 실시예들에서, 워드 인식 모듈(410)은 새롭게 인식된 텍스트 블록들의 포즈들을 결정할 수 있다. 몇몇 실시예들에서, 페이지 좌표계(285)를 사용하여, 새롭게 인식된 텍스트 블록들의 포지션들이 결정될 수 있다. 예컨대, 일 실시예에서, 워드 인식 모듈(410)은, 카메라(110)에 의해 캡쳐된 이미지들의 텍스트 블록

들을 인식하기 위해 프로세서(150) 상에서 실행될 수 있는 OCR 서브루틴을 포함할 수 있고, 인식된 텍스트 블록들을 메모리(160)에 저장할 수 있다.

[0076] 몇몇 실시예들에서, 워드 인식 모듈(410)은 사전들, LDA(Linear Discriminant Analysis) 기반 분류기들, 및 유한 상태 트랜스듀서들을 포함할 수 있다. LDA 분류기들은, 이미지 프레임들에서 광선 값들 및 분포들을 사용함으로써 OCR이 캐릭터들을 인식하는 것을 가능하게 하는 다변수 통계 분류 기술들을 사용할 수 있다. 유한 상태 트랜스듀서들은 OCR 정정 시스템에서 인식기 오류들을 정정하는데 사용될 수 있다. 몇몇 실시예들에서, 유한 상태 트랜스듀서들은 또한, 잘려진 캐릭터들, 두문자어들, 부분 워드들 등에 의해 도입되는 모호성들을 해결하는데 사용될 수 있다. 새롭게 인식된 텍스트 블록들 및 새롭게 인식된 텍스트 블록들의 포즈들을 갖는 이미지 프레임들이 워드 인식 모듈(410)에 의해 텍스트 어셈블러(430) 및 워드 추적 모듈(420)에 송신될 수 있다. 몇몇 실시예들에서, 방법(300)의 320, 325, 330 및 340의 일부분들이 워드 인식 모듈(410)에 의해 수행될 수 있다.

[0077] 몇몇 실시예들에서, 텍스트 어셈블러(430)는, 이전에 인식된 텍스트 블록들 또는 워드들에 대한 새롭게 인식된 텍스트 블록들 또는 워드들의 포지션들을 결정하기 위해, 새롭게 인식된 텍스트 블록들 또는 워드들의 포즈들을 사용할 수 있다. 그 다음, 몇몇 실시예들에서, 새롭게 인식된 워드들이 그들의 상대 포지션들에 기초하여, 현재 인식된 텍스트 블록들의 표현(450)에 배치될 수 있다. 예컨대, 현재 인식된 텍스트 블록들의 표현(450)은 링크된 목록 또는 어레이의 형태를 취할 수 있고, 새롭게 인식된 워드는 상대 포지션들에 기초하여 목록에 첨부 또는 삽입될 수 있다. 예컨대, 프로세서(150) 상에서 실행되는 애플리케이션 또는 서브루틴을 포함할 수 있는 텍스트 어셈블러 모듈(430)은 새롭게 인식된 텍스트 블록들 또는 워드들의 포즈들을 사용하여, 이전에 인식된 텍스트 블록들 또는 워드들에 대한 그들의 포지션들을 결정할 수 있고 그리고 현재 인식된 텍스트 블록들의 표현(450)을 메모리(160)에 저장할 수 있다. 몇몇 실시예들에서, 방법(300)의 340 및/또는 345의 일부분들이 텍스트 어셈블러(430)에 의해 수행될 수 있다.

[0078] 몇몇 실시예들에서, 추적 모듈(420)은 레퍼런스 프레임을 생성하기 위해 새롭게 인식된 텍스트 블록들의 포즈들을 사용할 수 있다. 몇몇 실시예들에서, 추적 모듈(420)은 또한, 레퍼런스 이미지 프레임 및 하나 또는 그 초과의 후속 프레임들에 기초하여 추적 타겟을 생성할 수 있다. 몇몇 실시예들에서, 추적 모듈(420)은 폐치 추적기를 포함할 수 있고, 폐치 추적기는 포즈 추정치를 개선하기 위해 워드 인식 모듈(410)에 의해 컴퓨팅된 포즈들을 사용할 수 있다. 몇몇 실시예들에서, 추적 모듈(420)은 초기 포즈를 획득하기 위해 ESM을 사용할 수 있고, 이 초기 포즈는, 폐치 추적기를 초기화하고 개선된 포즈 추정치를 획득하는데 사용될 수 있다. 몇몇 실시예들에서, 추적 모듈(420)은 현재 추적되는 워드들의 목록 및 그들의 개개의 포즈들(440)을 함께 유지할 수 있다. 몇몇 실시예들에서, 355, 360, 370 및 375의 일부분들이 추적 모듈(420)에 의해 수행될 수 있다. 일 실시예에서, 추적 모듈(420)은 프로세서(150) 상에서 실행되는 애플리케이션 또는 서브루틴의 형태를 취할 수 있다.

[0079] 몇몇 실시예들에서, 피드백 비주얼라이저(460)는, 워드들, 및/또는 텍스트 블록들 위에 오버레이된 증강 오브젝트들을 디스플레이하기 위해, 현재 인식된 텍스트의 표현(450) 그리고 현재 추적되는 워드들의 목록 및 그에 따른 그들의 개개의 포즈들(440)을 사용할 수 있다. 예컨대, 증강 오브젝트들은 인식된 텍스트의 변역을 포함할 수 있다. 몇몇 실시예들에서, 인식된 텍스트는 MS(100) 상의 디스플레이 상에서 디스플레이될 수 있고, 텍스트-투-보이스 애플리케이션을 사용하여 MS(100) 상의 또는 그에 커플링된 스피커를 통해 판독될 수 있다. 몇몇 실시예들에서, 380의 일부분들이 피드백 비주얼라이저(460)에 의해 수행될 수 있다. 예컨대, 일 실시예에서, 피드백 비주얼라이저(460)는 프로세서(150) 상에서 실행되는 애플리케이션 또는 서브루틴의 형태를 취할 수 있고, 디스플레이(145)를 사용하여 워드들, 및/또는 텍스트 블록들 위에 오버레이된 증강 오브젝트들을 디스플레이할 수 있다.

[0080] [0083] 도 5는 개시되는 실시예들과 일치하는, 모바일 스테이션 상에서 수행될 수 있는 예시적 방법(500)에 대한 흐름도를 도시한다. 예컨대, 몇몇 실시예들에서, 방법은 텍스트 검출 및 추적 동안에 수행될 수 있다.

[0081] [0084] 도 5를 참조하면, 몇몇 실시예들에서, 방법(500)은 MS(100) 상에서 실행되는 AR 애플리케이션 및/또는 컴퓨터 비전에 의해 호출될 수 있다. 몇몇 실시예들에서, 예시적 방법(500)은 505에서 호출될 수 있는데, 505에서는, 다양한 초기화 및 하우스키핑 루틴들이 수행될 수 있다.

[0082] [0085] 다음 차례로, 510에서는, 제1 텍스트 블록을 위치결정하고 인식하기 위해, 모바일 스테이션들, 예컨대 MS(100) 상의 카메라에 의해 캡처된 이미지 프레임에 대한 OCR(Optical Character Recognition)을 수행함으로써 제1 레퍼런스 프레임이 획득될 수 있다. 예컨대, 몇몇 실시예들에서, 510의 일부분들은 워드 인식 모듈

(410)의 OCR 애플리케이션에 의해 수행될 수 있다.

[0083]

[0086] 520에서는, 후속 이미지 프레임들의 세트로부터의 후속 이미지 프레임은, 선택되는 후속 이미지 프레임과 연관된 파라미터들에 기초하여 선택될 수 있다. 예컨대, 몇몇 실시예들에서, 520의 일부분들이 워드 인식 모듈(410)에 의해 수행될 수 있다. 몇몇 실시예들에서, 예컨대, 후속 이미지 프레임의 선택 이전에, 후속 이미지 프레임들의 세트에 대한 카메라 포즈들이 컴퓨팅될 수 있다. 예컨대, 몇몇 실시예들에서, 카메라 포즈들의 계산의 일부분들이 추적 모듈(420)에 의해 수행될 수 있다. 몇몇 실시예들에서, 각각의 카메라 포즈는 별개의 후속 이미지 프레임과 연관되고, 그리고/또는 연관된 후속 이미지 프레임을 제1 레퍼런스 프레임과 정렬하는 것에 적어도 부분적으로 기초하여 결정된다. 선택되는 후속 이미지 프레임이 후속 이미지 프레임들의 세트로부터 나오기 때문에, 이 프로세스 동안에, 선택되는 후속 이미지 프레임에 대한 카메라 포즈가 컴퓨팅될 수 있다. 몇몇 실시예들에서, 후속 이미지 프레임들의 세트의 이미지 프레임들과 연관된 파라미터들은, 그러한 이미지 프레임들과 연관된 컴퓨팅된 카메라 포즈들에 적어도 부분적으로 기초하여 결정된다. 선택되는 후속 이미지 프레임이 후속 이미지 프레임들의 세트로부터 나오기 때문에, 이 프로세스 동안에, 선택되는 후속 이미지 프레임과 연관된 파라미터들이 결정될 수 있다. 몇몇 실시예들에서, 파라미터들의 결정의 일부분들이 추적 모듈(420)에 의해 수행될 수 있다. 몇몇 실시예들에서, 파라미터들은 하기의 각각의 파라미터 중 하나 또는 그 조과 또는 이들의 결합을 포함하거나 또는 가질 수 있다: 선택되는 후속 이미지 프레임과 제1 레퍼런스 프레임 사이의 오버랩 영역의 퍼센티지; 선택되는 후속 이미지 프레임에서 보이는 제1 텍스트 블록의 프랙션 —프랙션은, 선택되는 후속 이미지 프레임에서 제1 텍스트 블록의 보이는 부분을 포함하는 영역 대 제1 텍스트 블록의 총 영역의 비로서 결정됨—; 제1 레퍼런스 프레임에 대한 선택되는 후속 이미지 프레임의 회전 규모, 및 제1 레퍼런스 프레임에 대한 선택되는 후속 이미지 프레임의 병진 규모.

[0084]

[0087] 530에서는, 제2 텍스트 블록을 인식하기 위해, 선택되는 후속 이미지 프레임에 대한 OCR을 수행함으로써, 제2 레퍼런스 프레임이 획득될 수 있다. 예컨대, 몇몇 실시예들에서, 530의 일부분들이 워드 인식 모듈(410)의 OCR 애플리케이션에 의해 수행될 수 있다.

[0085]

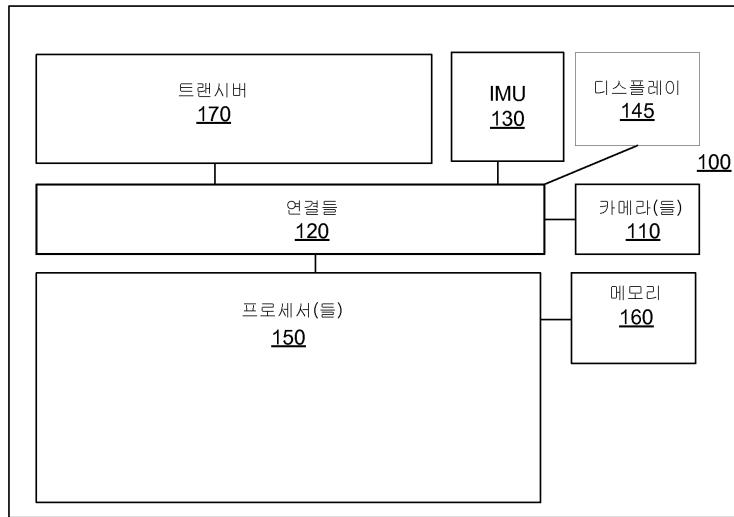
[0088] 다음 차례로, 540에서는, 제1 텍스트 블록과 제2 텍스트 블록 사이의 기하학적 관계가, 제2 레퍼런스 프레임에서의 제1 텍스트 블록의 포지션, 및 제2 레퍼런스 프레임과 연관된 카메라 포즈에 적어도 부분적으로 기초하여 결정될 수 있다. 예컨대, 몇몇 실시예들에서, 540의 일부분들은 텍스트 어셈블러 모듈(430)에 의해 수행될 수 있다. 545에서는, 제어가 콜링(calling) 프로그램 및/또는 애플리케이션, 예컨대 컴퓨터 비전 및/또는 AR 애플리케이션에 리턴될 수 있다.

[0086]

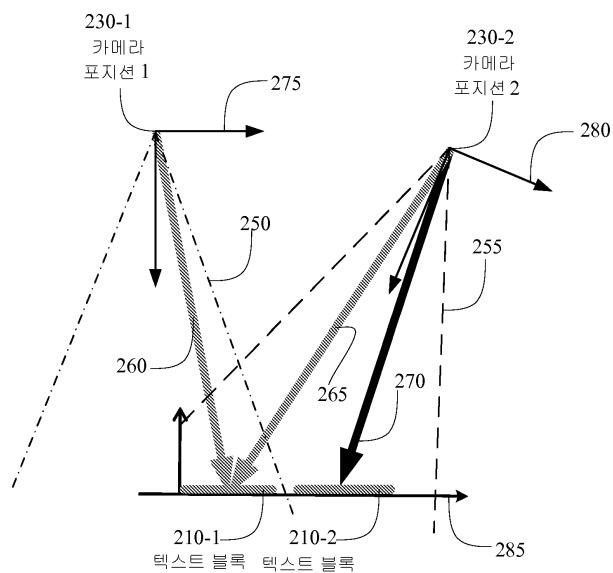
[0089] 개시된 양상들의 앞선 설명은 당업자가 본 개시물과 일치하는 다양한 실시예들을 만들거나 또는 사용하는 것을 가능하게 하도록 제공된다. 이러한 양상들에 대한 다양한 수정들이 당업자들에게 쉽게 명백할 것이며, 본 개시물의 사상 또는 범위로부터 벗어남 없이, 본원에 정의되는 일반 원칙들은 다른 양상들에 적용될 수 있다.

도면

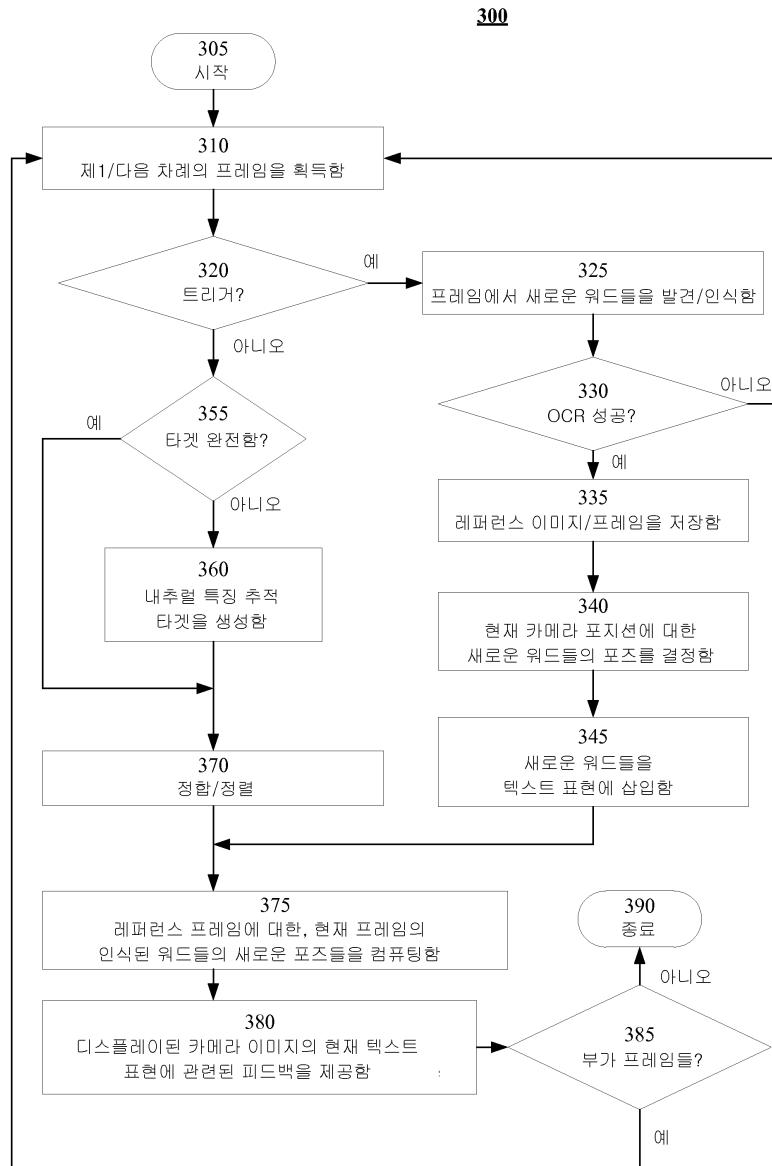
도면1



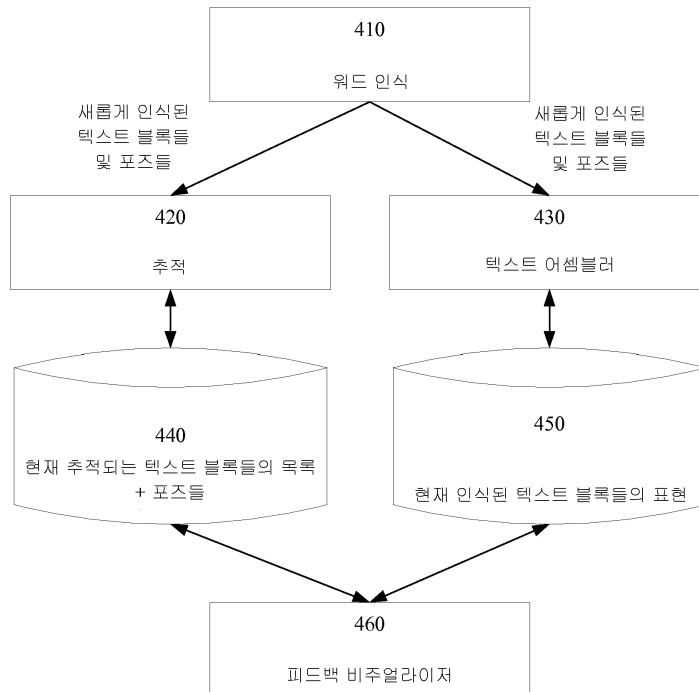
도면2



도면3



도면4



도면5

500