



**(19) 대한민국특허청(KR)**  
**(12) 공개특허공보(A)**

(11) 공개번호 10-2018-0087994  
(43) 공개일자 2018년08월03일

(51) 국제특허분류(Int. Cl.)  
G06T 7/593 (2017.01) G06N 3/02 (2006.01)  
H04N 13/00 (2018.01)

(52) CPC특허분류  
G06T 7/593 (2017.01)  
G06N 3/02 (2013.01)

(21) 출원번호 10-2017-0012468  
(22) 출원일자 2017년01월26일  
심사청구일자 없음

(71) 출원인  
삼성전자주식회사  
경기도 수원시 영통구 삼성로 129 (매탄동)

(72) 발명자  
이원희  
경기도 화성시 태안로 50, 302동 701호 (병점동, 우남퍼스트빌3차아파트)

장현성  
서울특별시 송파구 올림픽로 435, 212동 1002호 (신천동, 파크리오아파트)  
(뒷면에 계속)

(74) 대리인  
특허법인 무한

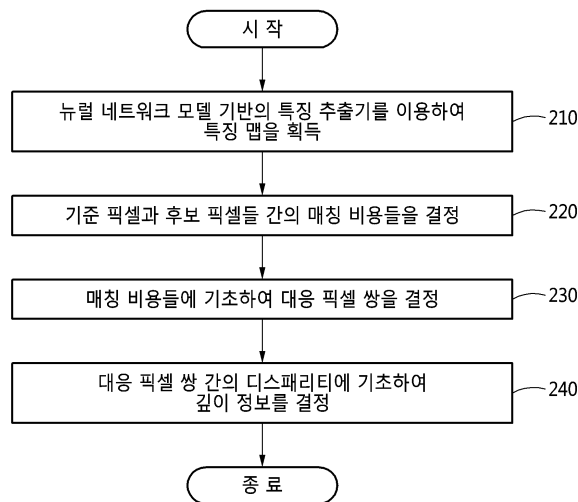
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 스테레오 매칭 방법 및 영상 처리 장치

**(57) 요약**

스테레오 매칭 방법 및 영상 처리 장치가 개시된다. 스테레오 매칭 방법은 뉴럴 네트워크 모델 기반의 특징 추출기를 이용하여 제1 시점 영상에 대한 제1 특징 맵 및 제2 시점 영상에 대한 제2 특징 맵을 획득하는 과정, 제1 특징 맵 및 제2 특징 맵을 이용하여 제1 시점 영상의 기준 픽셀과 제2 시점 영상의 후보 픽셀들 간의 매칭 비용들을 결정하는 과정과 매칭 비용들에 기초하여 제1 시점 영상과 제2 시점 영상 간에 대응 픽셀 쌍을 결정하는 과정을 포함한다.

**대표도** - 도2



(52) CPC특허분류

*H04N 13/128* (2018.05)

*G06T 2207/10012* (2013.01)

*G06T 2207/20084* (2013.01)

*G06T 2207/20228* (2013.01)

(72) 발명자

**성영훈**

경기도 화성시 동탄공원로 21-40, 926동 1702호 (능동, 동탄푸른마을두산위브아파트)

**정경부**

서울특별시 노원구 섭발로 265, 17동 1410호 (중계동, 상아아파트)

## 명세서

### 청구범위

#### 청구항 1

뉴럴 네트워크 모델 기반의 특징 추출기를 이용하여, 제1 시점 영상에 대한 제1 특징 맵 및 제2 시점 영상에 대한 제2 특징 맵을 획득하는 단계;

상기 제1 특징 맵 및 상기 제2 특징 맵을 이용하여 상기 제1 시점 영상의 기준 픽셀과 상기 제2 시점 영상의 후보 픽셀들 간의 매칭 비용들을 결정하는 단계; 및

상기 매칭 비용들에 기초하여 상기 후보 픽셀들 중에서 상기 기준 픽셀에 대응하는 대응 픽셀을 결정하는 단계를 포함하는 스테레오 매칭 방법.

#### 청구항 2

제1항에 있어서,

상기 매칭 비용들을 결정하는 단계는,

상기 기준 픽셀의 특징 벡터와 상기 후보 픽셀들 각각의 특징 벡터 간의 차이에 기초하여 상기 매칭 비용들을 결정하는, 스테레오 매칭 방법.

#### 청구항 3

제1항에 있어서,

상기 대응 픽셀을 결정하는 단계는,

상기 후보 픽셀들 중 가장 작은 값의 매칭 비용을 가지는 후보 픽셀을 상기 대응 픽셀로 결정하는, 스테레오 매칭 방법.

#### 청구항 4

제1항에 있어서,

상기 특징 추출기에는, 상기 제1 시점 영상의 영상 정보가 입력되고,

상기 특징 추출기는, 상기 입력된 영상 정보에 기초하여 상기 제1 특징 맵을 형성하는 특징 벡터를 결정하는, 스테레오 매칭 방법.

#### 청구항 5

제1항에 있어서,

상기 특징 추출기에는, 상기 제2 시점 영상의 영상 정보가 입력되고,

상기 특징 추출기는, 상기 입력된 영상 정보에 기초하여 상기 제2 특징 맵을 형성하는 특징 벡터를 결정하는, 스테레오 매칭 방법.

#### 청구항 6

제1항에 있어서,

상기 제1 특징 맵 및 제2 특징 맵을 획득하는 단계는,

상기 제1 시점 영상의 패치(patch) 영역들에 기반하여 상기 제1 특징 맵을 획득하고, 상기 제2 시점 영상의 패치 영역들에 기반하여 상기 제2 특징 맵을 획득하는, 스테레오 매칭 방법.

#### 청구항 7

제6항에 있어서,

상기 특징 추출기에는, 상기 제1 시점 영상의 상기 기준 픽셀을 포함하는 패치 영역의 정보가 입력되고,

상기 특징 추출기는, 상기 입력된 패치 영역의 정보에 기초하여 상기 기준 픽셀의 특징 벡터를 결정하는, 스테레오 매칭 방법.

#### 청구항 8

제7항에 있어서,

상기 특징 추출기에는 상기 기준 픽셀을 포함하는 패치 영역에 포함된 픽셀들의 휘도 정보(intensity information), 컬러 정보 및 그래디언트 정보(gradient information) 중 적어도 하나가 입력되는, 스테레오 매칭 방법.

#### 청구항 9

제6항에 있어서,

상기 특징 추출기에는, 상기 제2 시점 영상의 후보 픽셀을 포함하는 패치 영역의 정보가 입력되고,

상기 특징 추출기는, 상기 입력된 패치 영역의 정보에 기초하여 상기 후보 픽셀의 특징 벡터를 결정하는, 스테레오 매칭 방법.

#### 청구항 10

제9항에 있어서,

상기 특징 추출기에는, 상기 후보 픽셀을 포함하는 패치 영역에 포함된 픽셀들의 휘도 정보, 컬러 정보 및 그래디언트 정보 중 적어도 하나가 입력되는, 스테레오 매칭 방법.

#### 청구항 11

제1항에 있어서,

상기 제1 특징 맵은,

상기 제1 시점 영상에 포함된 픽셀들의 특징 벡터에 대한 정보를 포함하고,

상기 제2 특징 맵은,

상기 제2 시점 영상에 포함된 픽셀들의 특징 벡터에 대한 정보를 포함하는, 스테레오 매칭 방법.

#### 청구항 12

제1항에 있어서,

상기 후보 픽셀들은,

상기 제2 시점 영상에서 상기 제1 시점 영상의 상기 기준 픽셀에 대응되는 지점을 포함하는 라인 상에 위치하는 픽셀들인, 스테레오 매칭 방법.

#### 청구항 13

제1항에 있어서,

상기 매칭 비용들을 결정하는 단계는,

상기 기준 픽셀과 후보 픽셀 간의 초기 매칭 비용을 결정하는 단계; 및

상기 기준 픽셀에 인접한 다른 기준 픽셀의 매칭 비용 정보에 기초하여 상기 초기 매칭 비용을 조정하는 단계를 포함하는 스테레오 매칭 방법.

#### 청구항 14

제1항에 있어서,

상기 기준 픽셀과 상기 대응 픽셀 간의 디스패리티(disparity)에 기초하여 깊이 정보(depth information)를 결정하는 단계

를 더 포함하는 스테레오 매칭 방법.

**청구항 15**

뉴럴 네트워크 모델 기반의 특징 추출기를 이용하여, 제1 시점 영상에 포함된 기준 픽셀의 제1 특징 벡터 및 제2 시점 영상에 포함된 후보 픽셀들의 제2 특징 벡터를 추출하는 단계; 및

상기 추출된 제1 특징 벡터 및 상기 제2 특징 벡터 간의 벡터 거리에 기초하여 상기 후보 픽셀들 중에서 상기 기준 픽셀에 대응하는 대응 픽셀을 결정하는 단계

를 포함하는 스테레오 매칭 방법.

**청구항 16**

서로 다른 시간에서 획득된 제1 영상 및 제2 영상에서 특징점들을 추출하는 단계;

뉴럴 네트워크 모델 기반의 특징 추출기를 이용하여, 상기 제1 영상의 특징점들에 대한 제1 특징 벡터 및 상기 제2 영상의 특징점들에 대한 제2 특징 벡터를 결정하는 단계;

상기 제1 특징 벡터와 상기 제2 특징 벡터에 기초하여 상기 제1 영상 및 상기 제2 영상에서 서로 대응하는 특징점 쌍을 결정하는 단계; 및

상기 특징점 쌍 간의 위치 차이에 기초하여 변환 파라미터를 추정하는 단계

를 포함하는 변환 파라미터 추정 방법.

**청구항 17**

제16항에 있어서,

상기 특징 추출기에는, 상기 제1 영상 또는 상기 제2 영상의 특징점을 포함하는 패치 영역의 정보가 입력되고,

상기 특징 추출기는 입력된 정보에 기초하여 상기 특징점의 특징 벡터를 결정하는, 변환 파라미터 추정 방법.

**청구항 18**

제1항 내지 제17항 중 어느 하나의 항의 방법을 실행하기 위한 인스트럭션들을 저장하는 컴퓨터 판독 가능한 저장 매체.

**청구항 19**

프로세서를 포함하고,

상기 프로세서는,

뉴럴 네트워크 모델 기반의 특징 추출기를 이용하여, 제1 시점 영상에 대한 제1 특징 맵 및 제2 시점 영상에 대한 제2 특징 맵을 획득하고,

상기 제1 특징 맵 및 상기 제2 특징 맵을 이용하여 상기 제1 시점 영상의 기준 픽셀과 상기 제2 시점 영상의 후보 픽셀들 간의 매칭 비용들을 결정하며,

상기 매칭 비용들에 기초하여 상기 후보 픽셀들 중에서 상기 기준 픽셀에 대응하는 대응 픽셀을 결정하는, 영상 처리 장치.

**청구항 20**

제19항에 있어서,

상기 프로세서는,

상기 기준 픽셀에 대응하는 특징 벡터와 상기 후보 픽셀들 각각의 특징 벡터 간의 거리들에 기초하여 상기 매칭 비용들을 결정하는, 영상 처리 장치.

### 발명의 설명

#### 기술 분야

[0001] 아래의 실시예들은 영상 처리 기술에 관한 것이다.

#### 배경 기술

[0002]스테레오 매칭(stereo matching)은 2차원 영상으로부터 깊이 정보(depth information)를 얻는 여러 방법들 중 하나다. 스테레오 매칭은 두 장 혹은 그 이상의 영상에서 서로 대응점(correspondence points)을 검출하고, 영상 속 물체의 깊이를 예측한다. 스테레오 영상에는 서로 다른 시점에서 촬영된 동일한 물체가 존재하는데, 양안 시차의 특성에 따라 해당 물체가 카메라로부터 가까이 위치하는 경우에는 스테레오 영상의 영상들 사이에서 큰 변위 차이를 나타내고, 해당 물체가 카메라로부터 멀리 위치하는 경우에는 작은 변위 차이를 나타낸다. 이때, 물체의 깊이는 한 시점의 영상 속 픽셀들과 다른 시점 영상에 있는 대응점들 사이의 거리 차이인 디스패리티(disparity)에 의해 결정된다. 디스패리티를 구할 수 있으면, 디스패리티에 기초하여 물체까지의 깊이(depth)를 계산할 수 있다.

#### 발명의 내용

#### 해결하려는 과제

#### 과제의 해결 수단

- [0003] 일 실시예에 따른 스테레오 매칭 방법은, 뉴럴 네트워크 모델 기반의 특징 추출기를 이용하여, 제1 시점 영상에 대한 제1 특징 맵 및 제2 시점 영상에 대한 제2 특징 맵을 획득하는 단계; 상기 제1 특징 맵 및 상기 제2 특징 맵을 이용하여 상기 제1 시점 영상의 기준 픽셀과 상기 제2 시점 영상의 후보 픽셀들 간의 매칭 비용들을 결정하는 단계; 및 상기 매칭 비용들에 기초하여 상기 후보 픽셀들 중에서 상기 기준 픽셀에 대응하는 대응 픽셀을 결정하는 단계를 포함할 수 있다.
- [0004] 일 실시예에 따른 스테레오 매칭 방법에서, 상기 매칭 비용들을 결정하는 단계는, 상기 기준 픽셀의 특징 벡터와 상기 후보 픽셀들 각각의 특징 벡터 간의 차이에 기초하여 상기 매칭 비용들을 결정할 수 있다.
- [0005] 일 실시예에 따른 스테레오 매칭 방법에서, 상기 대응 픽셀을 결정하는 단계는, 상기 후보 픽셀들 중 가장 작은 값의 매칭 비용을 가지는 후보 픽셀을 상기 대응 픽셀로 결정할 수 있다.
- [0006] 일 실시예에 따른 스테레오 매칭 방법에서, 상기 특징 추출기에는, 상기 제1 시점 영상의 영상 정보가 입력되고, 상기 특징 추출기는, 상기 입력된 영상 정보에 기초하여 상기 제1 특징 맵을 형성하는 특징 벡터를 결정할 수 있다.
- [0007] 일 실시예에 따른 스테레오 매칭 방법에서, 상기 특징 추출기에는, 상기 제2 시점 영상의 영상 정보가 입력되고, 상기 특징 추출기는, 상기 입력된 영상 정보에 기초하여 상기 제2 특징 맵을 형성하는 특징 벡터를 결정할 수 있다.
- [0008] 일 실시예에 따른 스테레오 매칭 방법에서, 상기 제1 특징 맵 및 제2 특징 맵을 획득하는 단계는, 상기 제1 시점 영상의 패치 영역들에 기반하여 상기 제1 특징 맵을 획득하고, 상기 제2 시점 영상의 패치 영역들에 기반하여 상기 제2 특징 맵을 획득할 수 있다.
- [0009] 일 실시예에 따른 스테레오 매칭 방법에서, 상기 특징 추출기에는, 상기 후보 픽셀을 포함하는 패치 영역에 포함된 픽셀들의 휘도 정보, 컬러 정보 및 그레디언트 정보 중 적어도 하나가 입력될 수 있다.
- [0010] 다른 실시예에 따른 스테레오 매칭 방법은, 뉴럴 네트워크 모델 기반의 특징 추출기를 이용하여, 제1 시점 영상에 포함된 기준 픽셀의 제1 특징 벡터 및 제2 시점 영상에 포함된 후보 픽셀들의 제2 특징 벡터를 추출하는 단계; 및 상기 추출된 제1 특징 벡터 및 상기 제2 특징 벡터 간의 벡터 거리에 기초하여 상기 후보 픽셀들 중에서

상기 기준 픽셀에 대응하는 대응 픽셀을 결정하는 단계를 포함할 수 있다.

[0011] 일 실시예에 따른 변환 파라미터 추정 방법은, 서로 다른 시간에서 획득된 제1 영상 및 제2 영상에서 특징점들을 추출하는 단계; 뉴럴 네트워크 모델 기반의 특징 추출기를 이용하여, 상기 제1 영상의 특징점들에 대한 제1 특징 벡터 및 상기 제2 영상의 특징점들에 대한 제2 특징 벡터를 결정하는 단계; 상기 제1 특징 벡터와 상기 제2 특징 벡터에 기초하여 상기 제1 영상 및 상기 제2 영상에서 서로 대응하는 특징점 쌍을 결정하는 단계; 및 상기 특징점 쌍 간의 위치 차이에 기초하여 변환 파라미터를 추정하는 단계를 포함할 수 있다.

[0012] 일 실시예에 따른 변환 파라미터 추정 방법에서, 상기 특징 추출기에는, 상기 제1 영상 또는 상기 제2 영상의 특징점을 포함하는 패치 영역의 정보가 입력되고, 상기 특징 추출기는 입력된 정보에 기초하여 상기 특징점의 특징 벡터를 결정할 수 있다.

[0013] 일 실시예에 따른 영상 처리 장치는, 프로세서를 포함하고, 상기 프로세서는, 뉴럴 네트워크 모델 기반의 특징 추출기를 이용하여, 제1 시점 영상에 대한 제1 특징 맵 및 제2 시점 영상에 대한 제2 특징 맵을 획득하고, 상기 제1 특징 맵 및 상기 제2 특징 맵을 이용하여 상기 제1 시점 영상의 기준 픽셀과 상기 제2 시점 영상의 후보 픽셀들 간의 매칭 비용들을 결정하며, 상기 매칭 비용들에 기초하여 상기 후보 픽셀들 중에서 상기 기준 픽셀에 대응하는 대응 픽셀을 결정할 수 있다.

**도면의 간단한 설명**

[0014] 도 1은 일 실시예에 따른 스테레오 매칭 장치의 동작을 설명하기 위한 도면이다.

도 2는 일 실시예에 따른 스테레오 매칭 방법의 동작을 도시한 흐름도이다.

도 3은 일 실시예에 따른 시점 영상들에서 매칭 비용을 결정하는 일례를 설명하기 위한 도면이다.

도 4는 일 실시예에 따른 뉴럴 네트워크 모델 기반의 특징 추출기의 구조를 설명하기 위한 도면이다.

도 5는 일 실시예에 따른 변환 파라미터 추정 방법의 동작을 도시한 흐름도이다.

도 6은 일 실시예에 따른 변환 파라미터 추정 방법의 일례를 설명하기 위한 도면이다.

도 7은 일 실시예에 따른 영상 처리 장치의 구성을 도시한 도면이다.

도 8은 일 실시예에 따른 특징 추출기의 트레이닝 방법을 설명하기 위한 흐름도이다.

도 9는 일 실시예에 따른 트레이닝 방법의 일례를 설명하기 위한 도면이다.

**발명을 실시하기 위한 구체적인 내용**

[0015] 실시예들에 대한 특정한 구조적 또는 기능적 설명들은 단지 예시를 위한 목적으로 개시된 것으로서, 다양한 형태로 변경되어 실시될 수 있다. 따라서, 실시예들은 특정한 개시형태로 한정되는 것이 아니며, 본 명세서의 범위는 기술적 사상에 포함되는 변경, 균등물, 또는 대체물을 포함한다.

[0016] 제1 또는 제2 등의 용어를 다양한 구성요소들을 설명하는데 사용될 수 있지만, 이런 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 해석되어야 한다. 예를 들어, 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소는 제1 구성요소로도 명명될 수 있다.

[0017] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다.

[0018] 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 설명된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함으로써 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0019] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 해당 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 갖는 것으로 해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

- [0020] 이하, 실시예들을 첨부된 도면들을 참조하여 상세하게 설명한다. 첨부 도면을 참조하여 설명함에 있어, 도면 부호에 관계없이 동일한 구성 요소는 동일한 참조 부호를 부여하고, 이에 대한 중복되는 설명은 생략하기로 한다.
- [0021] 도 1은 일 실시예에 따른 스테레오 매칭 장치의 동작을 설명하기 위한 도면이다.
- [0022] 도 1을 참조하면, 일 실시예에 따른 스테레오 매칭 장치(100)는 스테레오 영상(120)(stereo image)을 수신하고, 스테레오 영상(120)에 기반하여 깊이 정보를 결정한다. 스테레오 영상(120)은 두 대 이상의 카메라를 통해 획득된 서로 다른 시점 영상들(예를 들어, 좌측 영상 및 우측 영상)을 포함한다. 시점 영상들은 동일 시간에 서로 다른 위치(또는 시점(viewpoint))에서 촬영된 영상들이다. 스테레오 영상(120)은 예를 들어, 스테레오 카메라에 의해 획득될 수 있다.
- [0023] 다른 실시예에 따르면, 스테레오 매칭 장치(100)는 세 개 이상의 시점 영상들을 포함하는 다시점 영상(multi-view image)에 기초하여 깊이 정보를 결정할 수도 있다. 다만, 이하에서는 설명의 편의를 위해 두 개의 시점 영상들로 구성된 스테레오 영상(120)에 기초하여 실시예들을 설명하도록 한다.
- [0024] 스테레오 매칭 장치(100)는 스테레오 영상(120)에서 서로 대응하는 픽셀 쌍을 찾고, 대응 픽셀 쌍 사이의 위치 차이인 디스패리티(disparity)에 기초하여 객체와 배경의 깊이 정보를 결정할 수 있다. 깊이 정보는 3D 영상을 렌더링하거나 또는 카메라의 시점에서 객체나 배경까지의 거리를 추정하는데 이용될 수 있다. 예를 들어, 차량(vehicle)의 네비게이션 시스템에서 스테레오 카메라를 이용하여 앞에 존재하는 차량 또는 장애물까지의 거리를 추정하는데 깊이 정보가 이용될 수 있다. 다른 예로, 증강 현실(augmented reality; AR) 영상에서는 실존하는 객체와 가상의 객체가 혼합되어 표현될 수 있는데, 실존하는 객체의 깊이 정보를 이용하여 가상의 객체를 표현할 영상 내 위치를 결정할 수 있다. 스테레오 매칭 장치(100)는 하나 이상의 소프트웨어 모듈, 하드웨어 모듈, 또는 이들의 다양한 조합으로 구현될 수 있다.
- [0025] 스테레오 매칭 장치(100)는 스테레오 매칭을 통하여 깊이 정보를 결정한다. 스테레오 매칭은 스테레오 영상(120)에 포함된 각 픽셀들의 특징을 추출하는 과정과 스테레오 영상(120) 간에 추출된 특징을 비교하여 각 픽셀의 대응 픽셀을 찾고, 디스패리티를 예측하는 과정을 포함한다. 예시적으로, 스테레오 매칭 장치(100)는 우측 영상에서 좌측 영상에 포함된 픽셀의 대응 픽셀을 검출하거나 또는 좌측 영상에서 우측 영상에 포함된 픽셀의 대응 픽셀을 검출할 수 있다. 스테레오 매칭 장치(100)는 대응 픽셀을 찾기 위해 픽셀의 휘도(intensity) 정보, 컬러 정보, 그래디언트(gradient) 정보, 또는 이들의 어느 조합을 이용할 수 있다. 스테레오 매칭 장치(100)는 검출된 대응 픽셀들 간의 위치 차이에 기초하여 디스패리티를 결정한다. 이하에서, 좌측 영상 및 우측 영상 중 어느 하나는 제1 시점 영상으로 지칭될 수 있고, 이 때 나머지 하나는 제2 시점 영상으로 지칭될 수 있다.
- [0026] 스테레오 매칭 장치(100)는 뉴럴 네트워크 모델(neural network model) 기반의 특징 추출기(110)를 이용하여 스테레오 매칭을 수행한다. 뉴럴 네트워크 모델로서 여러 개의 레이어들(layers)을 포함하는 깊은 뉴럴 네트워크 모델(deep neural network model)이 이용될 수 있다. 뉴럴 네트워크 모델 기반의 특징 추출기(110)를 이용하는 것에 의해 영상의 특징을 정확하고 빠르게 추정할 수 있다. 뉴럴 네트워크 모델은 생물학의 신경망의 특징을 모방한 통계학적 모델이다. 뉴럴 네트워크 모델은 입력된 정보로부터 바람직한 결과를 출력하는 문제 해결 능력을 가지는데, 이러한 문제 해결 능력을 기르기 위해 시냅스의 결합으로 네트워크를 형성한 인공 뉴런(또는 노드)들을 트레이닝(training)(또는 학습)하는 과정이 수행된다. 트레이닝 과정을 통해 뉴럴 네트워크 모델을 형성하는 파라미터들의 값이 보다 바람직하게 수정될 수 있다.
- [0027] 스테레오 매칭 장치(100)는 뉴럴 네트워크 모델 기반의 특징 추출기(110)를 이용하여 각 시점 영상 속 픽셀들의 특징(또는 특징 벡터)을 추출한 다음, 특징 거리 계산을 통해 대응 픽셀을 결정하는데 이용되는 픽셀 간 유사도를 결정할 수 있다. 스테레오 매칭 장치(100)는 뉴럴 네트워크 모델을 이용함으로써 시점 영상의 특징을 보다 정확하고 빠르게 추출하고, 추출된 특징 간의 유사도는 계산 복잡도가 상대적으로 낮은 특징 거리 계산을 이용하여 결정함으로써, 높은 정확성은 유지하면서 계산 복잡도를 줄일 수 있다. 이에 따라, 스테레오 매칭이 정확하고 빠르게 수행될 수 있다.
- [0028] 이하 도 2 내지 도 4를 통해 스테레오 매칭 장치(100)에 의해 수행되는 스테레오 매칭 방법을 보다 자세히 설명한다.
- [0029] 도 2는 일 실시예에 따른 스테레오 매칭 방법의 동작을 도시한 흐름도이다.



- [0030] 스테레오 매칭 방법은 도 1의 스테레오 매칭 장치(또는 도 7의 영상 처리 장치)에 의해 수행될 수 있다. 도 2를 참조하면, 단계(210)에서 스테레오 매칭 장치는 뉴럴 네트워크 모델 기반의 특징 추출기를 이용하여 서로 다른 시점 영상들에 대한 특징 맵(feature map)들을 획득한다. 예를 들어, 스테레오 매칭 장치는 특징 추출기를 이용하여 제1 시점 영상에 포함된 픽셀들의 제1 특징 벡터 및 제2 시점 영상에 포함된 픽셀들의 제2 특징 벡터를 추출하고, 제1 시점 영상 및 제2 시점 영상 각각에 대한 제1 특징 맵 및 제2 특징 맵을 획득할 수 있다. 제1 특징 맵은 제1 시점 영상에 포함된 픽셀들의 제1 특징 벡터에 대한 정보를 포함하고, 제2 특징 맵은 제2 시점 영상에 포함된 픽셀들의 제2 특징 벡터에 대한 정보를 포함한다.
- [0031] 일 실시예에서, 스테레오 매칭 장치는 영상 단위로 특징 맵을 획득할 수 있다. 특징 추출기에는 제1 시점 영상의 영상 정보가 입력되고, 특징 추출기는 입력된 영상 정보에 기초하여 제1 특징 맵을 형성하는 특징 벡터들을 결정할 수 있다. 제1 시점 영상의 영상 정보는 예를 들어, 제1 시점 영상에 포함된 픽셀들의 휘도 정보, 컬러 정보, 그래디언트 정보, 또는 이들의 어느 조합과 같은 픽셀 정보일 수 있다. 특징 추출기는 제1 시점 영상의 영상 정보가 입력되면, 제1 시점 영상의 각 픽셀에 대응하는 특징 벡터를 결정한다. 예를 들어, 특징 추출기는 제1 시점 영상의 기준 픽셀을 중심으로 하는 일 영역(예를 들어, 패치 영역)에 포함된 픽셀들의 픽셀 정보에 기초하여 기준 픽셀에 대응하는 특징 벡터를 결정할 수 있다. 특징 추출기는 이러한 과정을 다른 픽셀들에도 그대로 수행하여 제1 시점 영상에 포함된 각 픽셀에 대응하는 특징 벡터를 결정할 수 있다. 위 특징 추출기가 제1 특징 맵을 형성하는 특징 벡터들을 결정하는 과정과 유사하게, 특징 추출기에는 제2 시점 영상의 영상 정보가 입력되고, 특징 추출기는 입력된 영상 정보에 기초하여 제2 특징 맵을 형성하는 특징 벡터들을 결정할 수 있다.
- [0032] 다른 실시예에서, 스테레오 매칭 장치는 패치(patch) 영역들에 기반하여 특징 맵을 획득할 수 있다. 패치 영역은 예를 들어, 8X8 픽셀 블록과 같은 복수의 픽셀들을 포함하는 픽셀 영역을 나타낸다. 스테레오 매칭 장치는 제1 시점 영상의 패치 영역들에 기반하여 제1 특징 맵을 획득하고, 제2 시점 영상의 패치 영역들에 기반하여 제2 특징 맵을 획득할 수 있다. 예를 들어, 특징 추출기에는 제1 시점 영상의 기준 픽셀(reference pixel)을 중심으로 하는 패치 영역의 정보가 입력될 수 있고, 특징 추출기는 입력된 패치 영역의 정보에 기초하여 기준 픽셀의 특징 벡터를 결정한다. 이 때, 특징 추출기에는 패치 영역에 포함된 픽셀들의 휘도 정보, 컬러 정보, 그래디언트 정보, 또는 이들의 어느 조합이 입력될 수 있다. 기준 픽셀의 특징이 주변 픽셀들을 고려하여 결정되므로, 기준 픽셀에 대해 결정된 특징이 보다 정확한 값을 가질 확률이 높아질 수 있다. 이와 유사하게, 특징 추출기에는 제2 시점 영상의 후보 픽셀을 중심으로 하는 패치 영역의 정보가 입력될 수 있고, 특징 추출기는 입력된 패치 영역의 정보에 기초하여 후보 픽셀의 특징 벡터를 결정한다. 이 때, 특징 추출기에는 패치 영역에 포함된 픽셀들의 휘도 정보, 컬러 정보, 그래디언트 정보, 또는 이들의 어느 조합이 입력될 수 있다. 후보 픽셀은 제2 시점 영상에서 제1 시점 영상의 기준 픽셀에 대응하는 픽셀을 결정하는데 있어, 비교 대상이 되는 제2 시점 영상의 픽셀이다. 일 실시예에서, 후보 픽셀은 제2 시점 영상에서 제1 시점 영상의 기준 픽셀에 대응되는 지점을 포함하는 라인(line) 상에 위치하는 픽셀일 수 있다.
- [0033] 단계(220) 및 단계(230)에서, 스테레오 매칭 장치는 특징 맵들을 이용하여 시점 영상들 사이에서 서로 대응되는 픽셀 쌍을 결정한다. 스테레오 매칭 장치는 대응되는 픽셀 쌍을 결정하기 위해 특징 벡터들 간의 유사도를 이용하는 데, 이 때 픽셀들 간의 매칭 비용(matching cost)이 유사도로서 이용될 수 있다. 스테레오 매칭 장치는 후보 픽셀들 중에서 최적의 매칭 비용을 가지는 픽셀을 기준 픽셀의 대응 픽셀로 결정할 수 있다.
- [0034] 먼저, 단계(220)에서, 스테레오 매칭 장치는 제1 시점 영상의 제1 특징 맵 및 제2 시점 영상의 제2 특징 맵을 이용하여 제1 시점 영상의 기준 픽셀과 제2 시점 영상의 후보 픽셀들 간의 매칭 비용들을 결정한다. 스테레오 매칭 장치는 기준 픽셀의 특징 벡터와 후보 픽셀들 각각의 특징 벡터 간의 차이(또는 벡터 거리)를 계산하고, 계산된 차이에 기초하여 매칭 비용을 결정할 수 있다. 예시적으로, 스테레오 매칭 장치는 기준 픽셀의 특징 벡터와 후보 픽셀의 특징 벡터 간의 차이로서 특징 벡터들 간의 유클리디언 거리(Euclidean distance)를 계산하고, 계산된 유클리디언 거리에 기초하여 매칭 비용이 결정될 수 있다. 이 때, 스테레오 매칭 장치는 계산된 유클리디언 거리를 정규화(normalization)하거나 유클리디언 거리에 가중치를(weight) 적용할 수 있다. 일 실시예에서, 매칭 비용은 유클리디언 거리의 값에 비례하는 특징을 가질 수 있고, 큰 값의 매칭 비용은 비교 대상인 두 픽셀들의 유사도가 낮다는 것을 의미한다. 반대로, 작은 값의 매칭 비용은 비교 대상인 두 픽셀들의 유사도가 높다는 것을 의미한다.
- [0035] 다른 실시예에 따르면, 스테레오 매칭 장치는 기준 픽셀과 후보 픽셀 간의 매칭 비용이 결정된 이후에 매칭 비용 최적화(matching cost optimization) 또는 매칭 비용 종합(matching cost aggregation)을 수행할 수 있다. 예를 들어, 스테레오 매칭 장치는 기준 픽셀과 후보 픽셀 간의 초기 매칭 비용을 결정하고, 기준 픽셀에 인접한 다른 기준 픽셀의 매칭 비용 정보에 기초하여 초기 매칭 비용을 조정할 수 있다. 이러한 과정을 통해 매칭 비

용의 정확도를 높이고, 잘못된 정보에 의한 영향을 줄일 수 있다.

[0036] 단계(230)에서, 스테레오 매칭 장치는 결정된 매칭 비용들에 기초하여 제1 시점 영상과 제2 시점 영상 사이에서 대응 픽셀 쌍을 결정한다. 스테레오 매칭 장치는 제2 시점 영상의 후보 픽셀들 중에서 최적의 매칭 비용을 가지는 후보 픽셀을 기준 픽셀의 대응 픽셀로 결정한다. 예를 들어, 스테레오 매칭 장치는 후보 픽셀들 중 가장 작은 값의 매칭 비용을 가지는 후보 픽셀을 기준 픽셀의 대응 픽셀로 결정할 수 있다.

[0037] 위에 설명된 단계(220)과 단계(230)를 제1 시점 영상에 포함된 모든 패치 영역들에 대해 수행하면, 제1 시점 영상과 제2 시점 영상 사이에서 서로 대응되는 픽셀 쌍들이 결정될 수 있다.

[0038] 단계(240)에서, 스테레오 매칭 장치는 대응 픽셀 쌍 간의 디스패리티에 기초하여 깊이 정보를 결정한다. 일 실시예에서, 스테레오 매칭 장치는 기준 픽셀과 기준 픽셀에 대응하는 대응 픽셀 간의 디스패리티에 기초하여 깊이 정보를 결정하고, 깊이 정보에 기반하여 깊이 맵을 생성할 수 있다. 이 때, 제1 시점 영상에 포함된 기준 픽셀들과 제2 시점 영상의 대응 픽셀들 간의 디스패리티 정보를 포함하는 디스패리티 맵이 이용될 수도 있다.

[0039] 예를 들어, 스테레오 매칭 장치는 다음의 수학적 식 1에 기초하여 디스패리티로부터 깊이 값을 결정할 수 있다.

**수학적 식 1**

$$Z = \frac{B \times f \times s}{d}$$

[0040]

[0041] 여기서, Z 는 깊이 값이고, B 는 제1 시점 영상과 제2 시점 영상을 촬영한 두 카메라 간의 거리인 베이스라인(baseline)이다. F 는 카메라의 초점 거리(focal length)를 나타내고, d 는 디스패리티를 나타낸다. s 는 깊이 값의 크기를 결정하기 위한 스케일링 요소로, 미리 정의된 값을 가질 수 있다.

[0042] 도 3은 일 실시예에 따른 시점 영상들에서 매칭 비용을 결정하는 일례를 설명하기 위한 도면이다.

[0043] 도 3을 참조하면, 스테레오 매칭 장치가 뉴럴 네트워크 모델 기반의 특징 추출기(330)를 이용하여 시점 영상들의 특징을 추출하고 매칭 비용을 결정하는 과정이 도시된다. 제1 시점 영상(310)에서 픽셀(312)을 포함하는 패치 영역(315)에 대한 정보가 특징 추출기(330)에 입력되고, 특징 추출기(330)는 픽셀(312)에 대응하는 특징 벡터를 출력한다. 일 실시예에서, 제1 시점 영상(310)에 포함된 각 픽셀을 중심으로 하는 패치 영역에 포함된 픽셀들의 휘도 정보, 컬러 정보, 그레디언트 정보, 또는 이들의 조합과 같은 픽셀 정보가 특징 추출기(330)에 입력될 수 있다. 특징 추출기(330)는 입력된 패치 영역의 정보에 기초하여 패치 영역에 대응하는 특징 벡터를 출력하도록 미리 학습된다. 특징 추출기(330)가 제1 시점 영상(310)에 포함된 각 픽셀의 패치 영역들에 특징 벡터를 결정하면, 제1 시점 영상(310)에 대응하는 제1 특징 맵(340)을 획득할 수 있다. 제1 특징 맵(340)은 제1 시점 영상(310)에 포함된 픽셀들의 특징 값들을 포함한다.

[0044] 이와 유사하게, 제2 시점 영상(320)에서 픽셀(322)을 포함하는 패치 영역(324)에 대한 정보가 특징 추출기(330)에 입력되고, 특징 추출기(330)는 픽셀(322)에 대응하는 특징 벡터를 출력한다. 이와 같은 과정을 제2 시점 영상(320)에 포함된 전체 픽셀들에 대해 수행하면 제2 시점 영상(320)에 대응하는 제2 특징 맵(350)을 획득할 수 있다. 제2 특징 맵(350)은 제2 시점 영상(320)에 포함된 픽셀들의 특징 값들을 포함한다.

[0045] 일 실시예에서, 제1 시점 영상(310) 및 제2 시점 영상(320)의 너비(width)가 W 이고, 영상의 높이(height)가 H

라고 가정하면, 제1 시점 영상(310) 및 제2 시점 영상(320)에 포함된 픽셀의 총 개수는 (W × H) 개가 된다.

스테레오 매칭 장치는 특징 추출기(330)에 (2 × W × H) 개의 패치 영역을 통과시키면 제1 특징 맵(340) 및 제2 특징 맵(350)을 획득할 수 있다.

[0046] 다른 실시예에 따르면, 패치 영역의 정보가 특징 추출기(330)에 입력되는 것이 아니라, 영상 전체 영역의 정보가 특징 추출기(330)에 입력될 수 있다. 예를 들어, 제1 시점 영상(310)에 포함된 전체 픽셀들의 픽셀 정보가

특징 추출기(330)에 입력될 수 있고, 특징 추출기(330)는 입력된 영상 정보에 기초하여 제1 특징 맵(340)을 형성하는 특징 벡터를 결정할 수 있다. 이와 유사하게, 제2 시점 영상(320)에 포함된 전체 픽셀들의 픽셀 정보가 특징 추출기(330)에 입력되고, 특징 추출기(330)는 입력된 영상 정보에 기초하여 제2 특징 맵(350)을 형성하는 특징 벡터를 결정할 수 있다.

[0047] 위와 같은 과정을 통해, 스테레오 매칭 장치는 시점 영상들(310, 320)에 대응하는 특징 맵들(340, 350)을 생성하고, 특징 맵들(340, 350)을 이용하여 제1 시점 영상(310)과 제2 시점 영상(320) 사이에서 서로 대응되는 픽셀 쌍을 결정할 수 있다. 예를 들어, 스테레오 매칭 장치는 제1 시점 영상(310)의 픽셀의 특징 벡터와 제2 시점 영상(320)의 픽셀의 특징 벡터를 비교하여 매칭 비용을 결정(360)할 수 있다. 매칭 비용은 비교되는 픽셀들 간의 유사도를 추정하는데 이용될 수 있다. 일 실시예에서, 스테레오 매칭 장치는 제1 시점 영상(310)의 픽셀(315)과 제2 시점 영상(320)의 픽셀(322) 간의 유사도를 추정하기 위해 픽셀(315)의 특징 벡터(345)와 픽셀(322)의 특징 벡터(352) 간의 유클리디언 거리인 L2 (또는 L1) 거리를 이용할 수 있다.

[0048] 매칭 비용을 결정(360)할 때, 스테레오 매칭 장치는 제1 시점 영상(310)에서 기준이 되는 기준 픽셀과 기준 픽셀에 대응하는 제2 시점 영상(320)의 각 후보 픽셀들 간의 매칭 비용을 결정할 수 있다. 일 실시예에서, 픽셀(315)을 기준 픽셀이라고 가정하면, 후보 픽셀들은 제2 시점 영상(320)의 픽셀들 중 픽셀(315)의 위치에 대응하는 지점을 포함하는 라인 상에 있는 픽셀들 중에서 결정될 수 있다. 예를 들어, 제2 시점 영상(320)에서 픽셀(322)과 픽셀(326)을 연결하는 라인 상에 존재하는 픽셀들이 후보 픽셀들이 될 수 있다. 픽셀(326)의 특징 벡터(354) 또한 픽셀(326)을 중심으로 하는 패치 영역(328)에 대한 정보가 특징 추출기(330)에 입력되어 결정될 수 있다.

[0049] 스테레오 매칭 장치는 하나의 기준 픽셀과 여러 후보 픽셀들 간의 매칭 비용들을 각각 계산할 수 있다. 매칭 비용의 계산 결과에 기초하여 매칭 비용 볼륨(370, matching cost volume)이 계산될 수 있다. 일 예에서, 매칭

비용 볼륨(370)은 기준 픽셀과 후보 픽셀들 간의 매칭 비용을 엘리먼트로 가지는  $(D \times W \times H)$  차원의 행렬 구조로 표현할 수 있다. 여기서,  $W$  는 제1 시점 영상(310) 및 제2 시점 영상(320)의 너비이고,  $H$ 는 제1 시점 영상(310) 및 제2 시점 영상(320)의 높이이다.  $D$ 는 탐색 범위를 가리키고, 이는 기준 픽셀과 비교되는 후보 픽셀들

의 개수에 대응된다. 스테레오 매칭 장치는  $(D \times W \times H)$  번의 매칭 비용 계산을 통해 매칭 비용 볼륨(370)을 획득할 수 있다. 일 예에서, 매칭 비용 볼륨(370)에서 매칭 비용(372)은 픽셀(312)의 특징 벡터(345)와 픽셀(322)의 특징 벡터(352)에 기초하여 결정된 매칭 비용이고, 매칭 비용(374)은 픽셀(312)의 특징 벡터(345)와 픽셀(326)의 특징 벡터(354)에 기초하여 결정된 매칭 비용이다. 매칭 비용 볼륨(370)에서 매칭 비용(372)와 매칭 비용(374) 사이에는 픽셀(322)와 픽셀(326) 사이에 존재하는 각 후보 픽셀들과 기준 픽셀인 픽셀(312) 간의 매칭 비용들이 엘리먼트로서 존재한다.

[0050] 매칭 비용 볼륨(370)은 매칭 비용 최적화 또는 매칭 비용 중합을 통해 수정될 수 있다. 예를 들어, 매칭 비용의 최적화 과정은 매칭 비용의 노이즈를 제거하기 위해 픽셀들의 매칭 비용의 연속성을 확인하여, 동일 영역에 있는 픽셀들끼리 유사한 매칭 비용을 갖도록 조정하는 과정을 포함할 수 있다. 매칭 비용 최적화 또는 매칭 비용 중합의 과정은 전체 매칭 비용에 대해 수행되거나 또는 국부적으로 수행될 수도 있다.

[0051] 스테레오 매칭 장치는 후보 픽셀들의 매칭 비용들 중 최적의 매칭 비용을 가지는 후보 픽셀을 기준 픽셀에 대응하는 대응 픽셀로 결정한다. 스테레오 매칭 장치는 각 기준 픽셀에 대해 가장 낮은 매칭 비용을 가지는 후보 픽셀을 대응 픽셀로 결정할 수 있다. 스테레오 매칭 장치는 대응 픽셀 간의 위치 차이인 디스패리티와 카메라 파라미터에 기초하여 깊이 정보를 결정할 수 있다. 여기서, 카메라 파라미터는 예를 들어, 카메라의 초점 거리, 카메라 픽셀 피치(camera pixel pitch), 카메라의 베이스라인 등을 포함할 수 있다. 이렇게 계산된 깊이 정보에 기초하여 깊이 맵이 획득될 수 있다.

[0052] 일 실시예에 따르면, 뉴럴 네트워크 모델 기반의 특징 추출기를 이용하여 시점 영상들에서 특징들을 추출하고, 뉴럴 네트워크 모델을 이용하는 것 없이 간단한 연산을 통해 매칭 비용을 결정함으로써 높은 정확성은 유지하면서 계산 복잡도를 줄일 수 있다. 이에 따라, 스테레오 매칭을 보다 빠르게 수행할 수 있으며 필요한 자원을 줄일 수 있다.

[0053] 도 4는 일 실시예에 따른 뉴럴 네트워크 모델 기반의 특징 추출기의 구조를 설명하기 위한 도면이다.

- [0054] 도 4를 참조하면, 특징 추출기(430)에는 시점 영상(410)의 일 픽셀(422)을 중심으로 하는 패치 영역(424)의 정보가 입력되고, 특징 추출기(430)는 패치 영역(424)의 정보에 기초하여 픽셀(422)에 대응하는 특징 벡터  $f$  를 출력한다. 특징 추출기(430)를 형성하는 뉴럴 네트워크 모델은 컨볼루션 레이어(convolution layer)와 렐루(ReLU, Rectified Linear Unit) 레이어가 반복되어 연결된 구조를 가질 수 있다. 예를 들어, 뉴럴 네트워크 모델에서 제1 컨볼루션 레이어와 제1 렐루 레이어가 연결되고, 제1 렐루 레이어에 제2 컨볼루션 레이어가 연결되며, 제2 컨볼루션 레이어에 제2 렐루 레이어가 연결되는 구조를 가질 수 있다. 컨볼루션 레이어는 입력된 값에 대해 컨볼루션 필터링을 수행하고, 렐루 레이어는 입력된 값에서 음수 값을 버리고 양수 값을 다음 레이어로 전달한다.
- [0055] 도 1 내지 도 4에서 설명한 뉴럴 네트워크 모델 기반의 특징 추출기는 서로 다른 시간에서 획득한 영상들을 이용하여 카메라 이동 파라미터 또는 포즈 변환 파라미터와 같은 변환 파라미터를 추정할 때에도 이용될 수도 있다. 해당 영상들 사이에는 시간 간격이 존재하고, 해당 시간 간격 동안 카메라 또는 객체가 이동할 수 있기 때문에 따른 변환 파라미터를 추정할 필요가 있다. 특징 추출기를 이용하여 변환 파라미터를 추정하는 실시예에 대해서는 도 5 및 도 6을 참조하여 설명한다.
- [0056] 도 5는 일 실시예에 따른 변환 파라미터 추정 방법의 동작을 도시한 흐름도이다. 도 5를 참조하면, 변환 파라미터 추정 방법은 도 7의 영상 처리 장치에 의해 수행될 수 있다.
- [0057] 단계(510)에서, 영상 처리 장치는 서로 다른 시간에서 획득된 제1 영상 및 제2 영상에서 특징점(feature point)들을 추출한다. 예를 들어, 제1 영상은 제1 시간에서 획득된 영상이고, 제2 영상은 제1 시간의 다음 시간인 제2 시간에서 획득된 영상일 수 있다. 영상 처리 장치는 각 제1 영상 및 제2 영상 내 픽셀의 휘도 정보, 컬러 정보 또는 그레디언트 정보 등에 기초하여 에지(edge) 영역 및 코너(corner) 영역을 특징점으로서 추출할 수 있다. 일 실시예에서, 영상 처리 장치는 Harris corner 기법, SIFT(Scale Invariant Feature Transform) 기법, 또는 FAST(Features from Accelerated Segment Test) 기법 등을 이용하여 제1 영상 및 제2 영상으로부터 특징점을 추출할 수 있다. 예를 들어, 영상 처리 장치는 SIFT 기법을 이용하여 영상의 국부적(local)인 특징을 추출할 수 있다. SIFT는 특징점 주변의 국부적인 그레디언트 분포 특성을 표현하는 특징으로, 특징점 주변의 패치 영역을 4x4 블록으로 분할하고, 각 블록에 포함된 픽셀들의 그레디언트 방향(gradient orientation)과 그레디언트 크기(gradient magnitude)에 대한 히스토그램(histogram)을 결정한 후, 해당 히스토그램의 빈(bin) 값들을 일렬로 연결한 128 차원의 벡터이다.
- [0058] 단계(520)에서, 영상 처리 장치는 뉴럴 네트워크 모델 기반의 특징 추출기를 이용하여, 제1 영상의 특징점들에 대한 제1 특징 벡터 및 제2 영상의 특징점들에 대한 제2 특징 벡터를 결정한다. 특징 추출기에는 제1 영상의 특징점을 포함하는 패치 영역에 대한 정보가 입력되고, 특징 추출기는 입력된 정보에 기초하여 제1 영상의 특징점에 대응하는 특징 벡터를 결정한다. 이와 유사하게, 특징 추출기에는 제2 영상의 특징점을 포함하는 패치 영역에 대한 정보가 입력되고, 특징 추출기는 입력된 정보에 기초하여 제2 영상의 특징점에 대응하는 특징 벡터를 결정한다.
- [0059] 단계(530)에서, 영상 처리 장치는 제1 특징 벡터와 제2 특징 벡터에 기초하여 제1 영상 및 제2 영상에서 서로 대응하는 특징점 쌍을 결정한다. 영상 처리 장치는 제1 특징 벡터 및 제2 특징 벡터를 이용하여 제2 영상의 특징점들 중 제1 영상의 특징점에 가장 가까운 특징점을 결정할 수 있다. 예를 들어, 영상 처리 장치는 위에서 설명한 것과 같이, 제1 특징 벡터와 제2 특징 벡터 간의 벡터 거리(예를 들어, 유클리디언 거리)를 계산하고, 계산된 벡터 거리에 기초하여 제1 영상과 제2 영상에서 서로 대응하는 특징점 쌍을 결정할 수 있다. 영상 처리 장치는 제2 영상의 후보 특징점들 중 제1 영상의 기준 특징점과 벡터 거리가 가장 작은 후보 특징점을 기준 특징점에 대응하는 특징점으로 결정할 수 있다.
- [0060] 다른 실시예에 따르면, 영상 처리 장치는 오 정합 제거(outlier removal) 과정을 추가적으로 수행할 수 있다. 오 정합 제거는 일차적으로 결정된 특징점 쌍에서 서로 정합되지 않을 것 같은 특징점 쌍을 구별하고, 구별된 특징점 쌍은 고려하지 않음으로써 특징점 매칭의 정확성을 높이기 위한 기법이다. 예를 들어, 영상 처리 장치는 후보 특징점과 기준 특징점 간의 벡터 거리가 다른 후보 특징점들의 벡터 거리들보다는 작으나 임계값보다 큰 값을 가지는 경우에는 해당 후보 특징점과 기준 특징점은 서로 대응하지 않는 것으로 결정할 수 있다.
- [0061] 단계(540)에서, 영상 처리 장치는 대응되는 특징점 쌍 간의 위치 차이에 기초하여 변환 파라미터를 추정한다. 영상 처리 장치는 제1 영상과 제2 영상 사이에서 서로 대응하는 것으로 결정된 특징점 쌍들의 위치 정보에 기초하여 변환 행렬을 추정할 수 있다. 특징점 쌍들의 위치 정보를 행렬로 표현하고, 이에 행렬 연산을 적용하는

것에 의해 제1 영상과 제2 영상 간의 변환 관계를 추정할 수 있는 변환 행렬이 계산될 수 있다.

- [0062] 도 6은 일 실시예에 따른 변환 파라미터 추정 방법의 일례를 설명하기 위한 도면이다.
- [0063] 도 6을 참조하면, 영상 처리 장치는 제1 영상(610)에서 제1 특징점들(612, 616)을 추출하고, 제2 영상(620)에서 제2 특징점들(622, 626)을 추출한다. 영상 처리 장치는 다양한 특징점 추출 기법을 이용하여 특징점을 추출할 수 있다. 이후에, 뉴럴 네트워크 모델 기반의 특징 추출기(630)에 특징점을 포함하는 패치 영역에 대한 정보가 입력되고, 특징 추출기(630)는 해당 특징점에 대응하는 특징 벡터를 출력한다. 예를 들어, 특징 추출기(630)에는 제1 영상(610)의 제1 특징점(612)을 중심으로 하는 패치 영역(614)에 대한 정보로서, 패치 영역(614)에 포함된 픽셀들의 휘도 정보, 컬러 정보, 그레디언트 정보 또는 이들의 어느 조합이 입력될 수 있고, 특징 추출기(630)는 제1 특징점(612)에 대응하는 특징 벡터를 출력할 수 있다. 특징 추출기(630)는 제1 영상(610)에서 추출된 각 제1 특징점에 대해 이와 같은 과정을 반복적으로 수행하여 제1 영상(610)의 각 제1 특징점들에 대응하는 특징 벡터를 결정할 수 있다.
- [0064] 위 과정과 유사하게, 특징 추출기(630)에는 제2 영상(620)의 제2 특징점(622)을 중심으로 하는 패치 영역(624)에 대한 정보가 입력되고, 특징 추출기(630)는 제2 특징점(622)에 대응하는 특징 벡터를 출력할 수 있다. 특징 추출기(630)는 제2 영상(620)에서 추출된 각 제2 특징점에 대해 이와 같은 과정을 반복적으로 수행하여 제2 영상(620)의 각 제2 특징점들에 대응하는 특징 벡터를 결정할 수 있다.
- [0065] 이후에, 영상 처리 장치는 제1 특징점들의 특징 벡터와 제2 특징점들의 특징 벡터를 비교(640)하여, 서로 대응하는 특징점 쌍을 결정한다. 일 실시예에서, 영상 처리 장치는 특징 벡터들 간의 벡터 거리를 계산하고, 벡터 거리가 가장 작은 특징점 쌍을 제1 영상(610)과 제2 영상(620)에서 서로 대응하는 특징점 쌍으로 결정할 수 있다.
- [0066] 영상 처리 장치는 특징점 쌍을 결정한 이후에 오 정합 제거 과정을 추가적으로 수행할 수 있다. 일 실시예에서, 영상 처리 장치는 제1 영상(610)과 제2 영상(620) 사이에서 결정된 특징점 쌍들 중 벡터 거리가 임계 값 이상이 되는 특징점 쌍은 대응 특징점 쌍 집합에서 제외시킬 수 있다. 예를 들어, 특징점(612)의 특징 벡터와 특징점(624)의 특징 벡터 간의 벡터 거리는 임계값보다 작아 대응 특징점 쌍 집합에 계속 포함될 수 있으나, 특징점(616)의 특징 벡터와 특징점(626)의 특징 벡터 간의 벡터 거리는 임계값보다 커 대응 특징점 쌍 집합에서 제거될 수 있다.
- [0067] 영상 처리 장치는 위 오 정합 제거 과정이 수행된 이후에 대응 특징점 쌍 집합에 포함된 대응 특징점 쌍 간의 위치 차이에 기초하여 변환 파라미터를 추정(650)할 수 있다. 추정된 변환 파라미터는 영상을 보정하거나 변환하는데 이용될 수 있다.
- [0068] 도 7은 일 실시예에 따른 영상 처리 장치의 구성을 도시한 도면이다.
- [0069] 일 실시예에 따른 영상 처리 장치는 센서(710), 프로세서(720) 및 메모리(730)를 포함한다. 센서(710), 프로세서(720) 및 메모리(730)는 통신 버스(740)를 통해 서로 통신할 수 있다.
- [0070] 센서(710)는 영상을 촬영할 수 있다. 일 실시예에서, 센서(710)는 스테레오 영상을 촬영할 수 있고, 제1 시점 영상을 촬영하기 위한 제1 센서 및 제2 시점 영상을 촬영하기 위한 제2 센서를 포함할 수 있다. 제1 센서 및 제2 센서는 예를 들어, 이미지 센서, 근접 센서 또는 적외선 센서일 수 있다. 센서(710)는 잘 알려진 방식(예를 들어, 광학 이미지를 전기 신호로 변환하는 방식 등)으로 스테레오 영상을 촬영할 수 있다. 센서(710)는 촬영된 컬러 영상, 깊이 영상 및 적외선 영상 중 적어도 하나를 프로세서(720) 및 메모리(730) 중 적어도 하나로 전달할 수 있다.
- [0071] 프로세서(720)는 상술한 스테레오 매칭 또는 변환 파라미터 추정에 관한 동작을 처리할 수 있다. 일 실시예에서, 프로세서(720)는 뉴럴 네트워크 모델 기반의 특징 추출기를 이용하여 스테레오 영상을 특징 맵들로 변환하고, 특징 맵에 포함된 특징 정보에 기초하여 픽셀들 간의 매칭 비용을 계산할 수 있다. 프로세서(720)는 매칭 비용에 기초하여 스테레오 영상에서 서로 대응하는 대응 픽셀 쌍을 결정하고, 대응 픽셀 쌍 간의 디스패리티에 기초하여 깊이 정보를 추정할 수 있다.
- [0072] 다른 실시예에서, 프로세서(720)는 서로 다른 시간에서 촬영된 영상들에서 특징점들을 추출하고, 특징 추출기를 이용하여 특징점들에 대응하는 특징 벡터들을 결정할 수 있다. 이후에, 프로세서(720)는 영상들 간의 특징 벡터들을 비교하여 영상들에서 서로 대응하는 특징점 쌍을 결정하고, 결정된 대응 특징점 쌍들의 위치 정보에 기초하여 변환 파라미터를 추정할 수 있다.

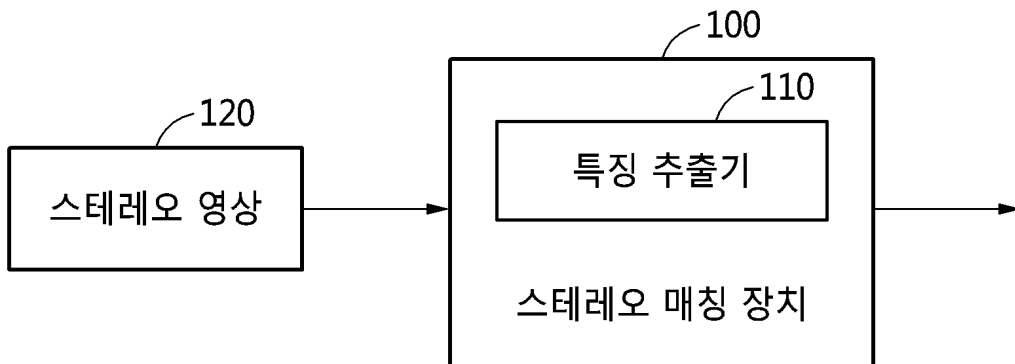
- [0073] 이 밖에도, 프로세서(720)는 도 1 내지 도 6을 통해 전술한 적어도 하나의 동작을 수행할 수 있으며, 보다 상세한 설명은 생략한다. 프로세서(720)는 명령어들 또는 프로그램들을 실행하거나, 영상 처리 장치를 제어할 수 있다.
- [0074] 메모리(730)는 상술한 스테레오 매칭 또는 변환 파라미터의 추정에서 이용되는 정보 및 결과 정보를 저장할 수 있다. 또한, 메모리(730)는 컴퓨터에서 읽을 수 있는 명령어들을 저장할 수 있다. 메모리(730)에 저장된 명령어들이 프로세서(720)에서 실행되면, 프로세서(720)는 상술된 하나 이상의 동작을 처리할 수 있다.
- [0075] 영상 처리 장치는 입/출력 장치(도면 미 표시)를 통하여 사용자 입력을 수신하거나 영상 및 처리 결과를 출력할 수 있다. 또한, 영상 처리 장치는 통신 장치(도면 미 표시)를 통해 외부 장치(예를 들어, 퍼스널 컴퓨터 또는 네트워크)에 연결되고, 외부 장치와 데이터를 교환할 수 있다.
- [0076] 도 8은 일 실시예에 따른 특징 추출기의 트레이닝 방법을 설명하기 위한 흐름도이다.
- [0077] 도 8의 트레이닝 방법은 트레이닝 장치에 의해 수행될 수 있고, 상술된 실시예들에서 이용되는 특징 추출기를 트레이닝하기 위해 수행될 수 있다. 트레이닝은 특징 추출기가 보다 바람직한 결과를 출력하도록 특징 추출기를 구성하는 뉴럴 네트워크 모델의 파라미터를 조정하는 과정을 포함한다. 일 실시예에서, 트레이닝 방법은 오프라인 상에서 시점 영상들 간의 대응 패치의 상관 관계를 이용하여 특징 추출기를 트레이닝할 수 있다.
- [0078] 도 8을 참조하면, 단계(810)에서 트레이닝 장치는 트레이닝을 위한 스테레오 영상에서 패치 영역들을 결정한다. 트레이닝 장치는 스테레오 영상에 포함된 제1 시점 영상 및 제2 시점 영상에서 각 픽셀을 중심으로 하는 패치 영역들을 결정할 수 있다. 단계(820)에서, 트레이닝 장치는 뉴럴 네트워크 모델 기반의 특징 추출기를 이용하여 패치 영역들에 대응하는 특징 벡터들을 결정한다. 특징 추출기에는 패치 영역에 대한 정보가 입력되고, 특징 추출기는 패치 영역의 중심 픽셀에 대한 특징 벡터를 출력할 수 있다.
- [0079] 단계(830)에서, 트레이닝 장치는 결정된 특징 벡터들에 기초하여 특징 추출기를 트레이닝한다. 트레이닝 과정에서는, 제1 시점 영상과 제2 시점 영상 사이에서 서로 대응하는 패치 영역을 이미 알고 있기 때문에, 트레이닝 장치는 특징 추출기가 위 서로 대응하는 패치 영역에 대해서는 서로 유사한 값의 특징 벡터를 출력하도록 뉴럴 네트워크 모델의 파라미터들을 업데이트할 수 있다. 또한, 트레이닝 장치는 제1 시점 영상과 제2 시점 영상 사이에서 서로 대응하지 않는 패치 영역에 대해서는 특징 벡터 간의 차이가 커지도록 뉴럴 네트워크 모델의 파라미터들을 업데이트할 수 있다. 업데이트 과정은 뉴럴 네트워크 모델을 구성하는 인공 뉴런들 간의 연결 가중치를 조정하는 과정을 포함한다.
- [0080] 일 실시예에서, 트레이닝 장치는 손실 역 전파 학습(back-propagation learning)을 통해 인공 뉴런들 사이의 연결 가중치를 조정할 수 있다. 손실 역 전파 학습은, 주어진 트레이닝 데이터에 대해 전방 계산(forward computation)을 통하여 손실을 추정한 후, 뉴럴 네트워크 모델의 출력 레이어에서 시작하여 하위 레이어로 향하는 역 방향으로 추정한 손실을 전파하면서, 손실을 줄이는 방향으로 인공 뉴런들 간의 연결 가중치를 조정하는 방법이다. 트레이닝 장치는 현재 설정된 연결 가중치들이 얼마나 최적에 가까운지를 측정하기 위한 목적 함수(objective function)를 정의하고, 목적 함수의 결과가 만족할 만한 수준에 도달할 때까지 연결 가중치들을 계속 조정할 수 있다.
- [0081] 도 9는 일 실시예에 따른 트레이닝 방법의 일례를 설명하기 위한 도면이다.
- [0082] 도 9를 참조하면, 일 실시예에 따른 트레이닝 장치는 삼중 뉴럴 네트워크 모델(triplet neural network model)을 이용하여 특징 추출기(930)를 트레이닝할 수 있다. 이 때, 특징 추출기(930)를 형성하는 뉴럴 네트워크 모델은 총 3 개의 뉴럴 네트워크 모델과 파라미터를 공유하는 삼중 네트워크로 구성될 수 있다.
- [0083] 일 실시예에서, 제1 뉴럴 네트워크 모델에는 제1 시점 영상(910)에 포함된 기준 픽셀(912)을 중심으로 하는 기준 패치 영역(914)에 대한 정보가 입력될 수 있다. 제2 뉴럴 네트워크 모델에는 제2 시점 영상(920)에 포함된 참 후보 픽셀(true candidate pixel, 922)을 중심으로 하는 참 후보 패치 영역(924)에 대한 정보가 입력되고, 제3 뉴럴 네트워크 모델에는 제2 시점 영상(920)에 포함된 거짓 후보 픽셀(false candidate pixel, 926)을 중심으로 하는 거짓 후보 패치 영역(928)에 대한 정보가 입력된다. 참 후보 패치 영역(924)는 참 디스패리티(true disparity)를 결정하기 위한 것으로, 참 정합(inlier)을 위한 패치 영역이다. 이와 반대로, 거짓 후보 패치 영역(928)은 거짓 디스패리티(false disparity)를 효과적으로 구별하기 위한 것으로, 오 정합(outlier)을 구별하기 위한 패치 영역이다. 예를 들어, 128개의 후보 패치 영역들이 존재하는 경우, 1개의 후보 패치 영역이 참 정합을 위한 패치 영역이 되고, 나머지 127개의 후보 패치 영역들은 오 정합을 구별하기 위한 패치 영역

들이 된다.

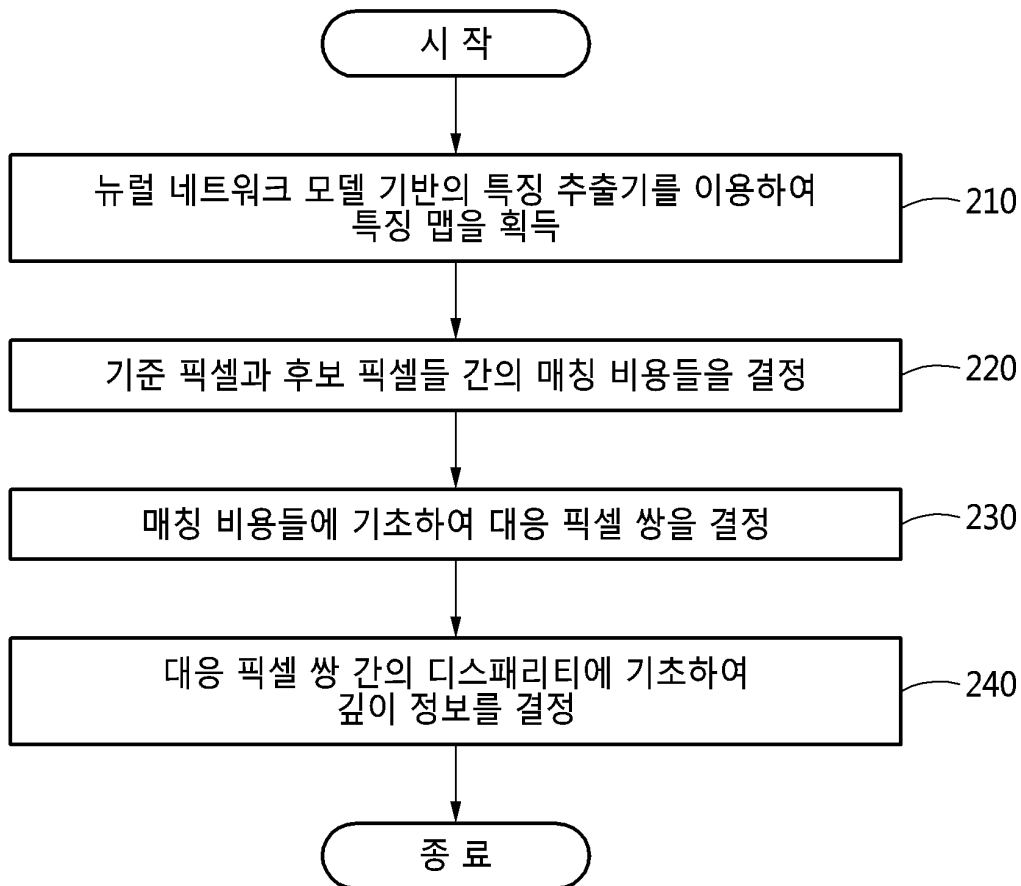
- [0084] 각 뉴럴 네트워크 모델로부터 출력된 특징을 각각  $f_1$ ,  $f_2$ ,  $f_3$  라고 지칭한다. 트레이닝 장치는  $f_1$  과  $f_2$  간의 특징 벡터 거리인  $d_1$  을 계산하고,  $f_1$  과  $f_3$  간의 특징 벡터 거리인  $d_2$  를 계산한다. 트레이닝 장치는  $d_1$  이 0에 가까운 값을 가지되,  $d_2$  는 높은 값을 가지도록 특징 추출기(930)를 구성하는 뉴럴 네트워크 모델을 트레이닝할 수 있다.
- [0085] 도 9에 도시된 삼중 뉴럴 네트워크 모델을 통해 특징 추출기(930)의 특징 추출 과정을 트레이닝하면, 기준 패치 영역에 실제 대응하는 참(또는 인라이어) 패치 영역에 대한 매칭 비용을 거짓(또는 아웃라이어) 패치 영역에 대한 매칭 비용보다 낮은 값을 가지도록 특징 추출기(930)를 트레이닝할 수 있어 효과적이다.
- [0086] 일 실시예에 따르면, 도 1 내지 도 7의 실시예에 포함된 특징 추출 과정에서는 트레이닝된 특징 추출기(930)에 포함된 3개의 뉴럴 네트워크 모델들 중에서 하나의 뉴럴 네트워크 모델이 이용될 수 있다.
- [0087] 이상에서 설명된 실시예들은 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치, 방법 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate 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 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.
- [0088] 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(embodiment)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.
- [0089] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 컴퓨터 판독 가능 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 상기된 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.
- [0090] 이상과 같이 실시예들이 비록 한정된 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기를 기초로 다양한 기술적 수정 및 변형을 적용할 수 있다. 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.

도면

도면1

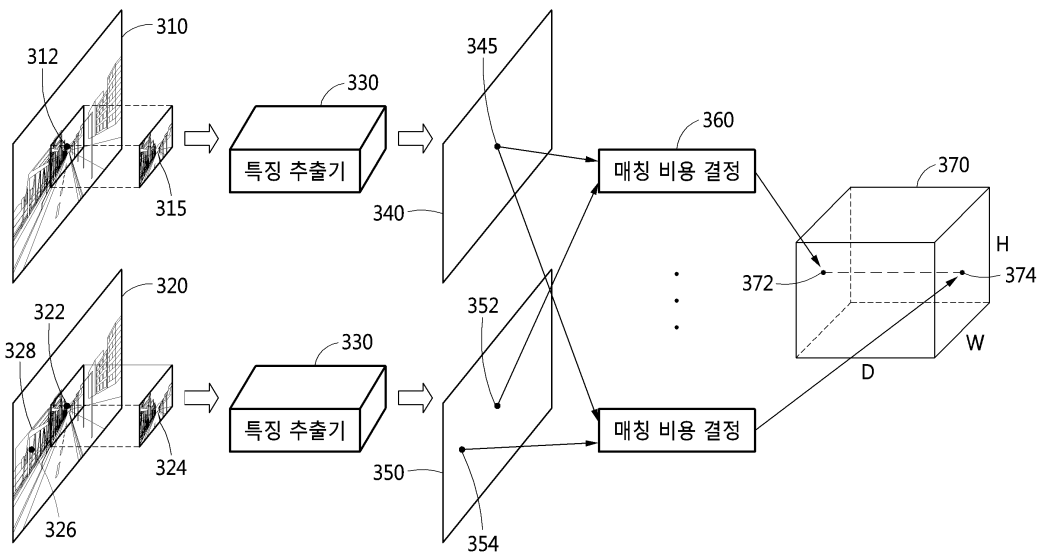


도면2

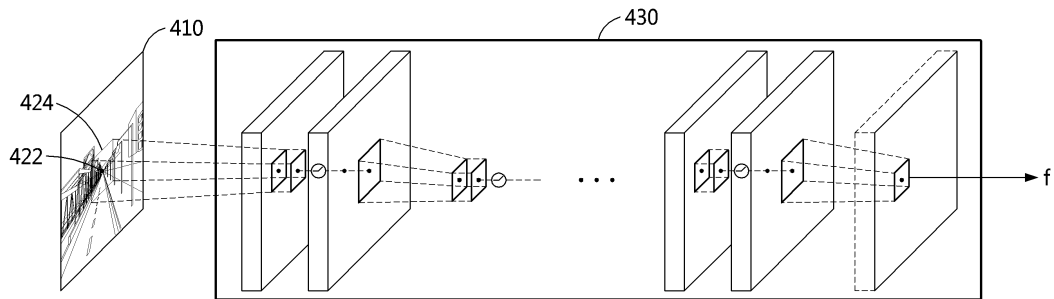




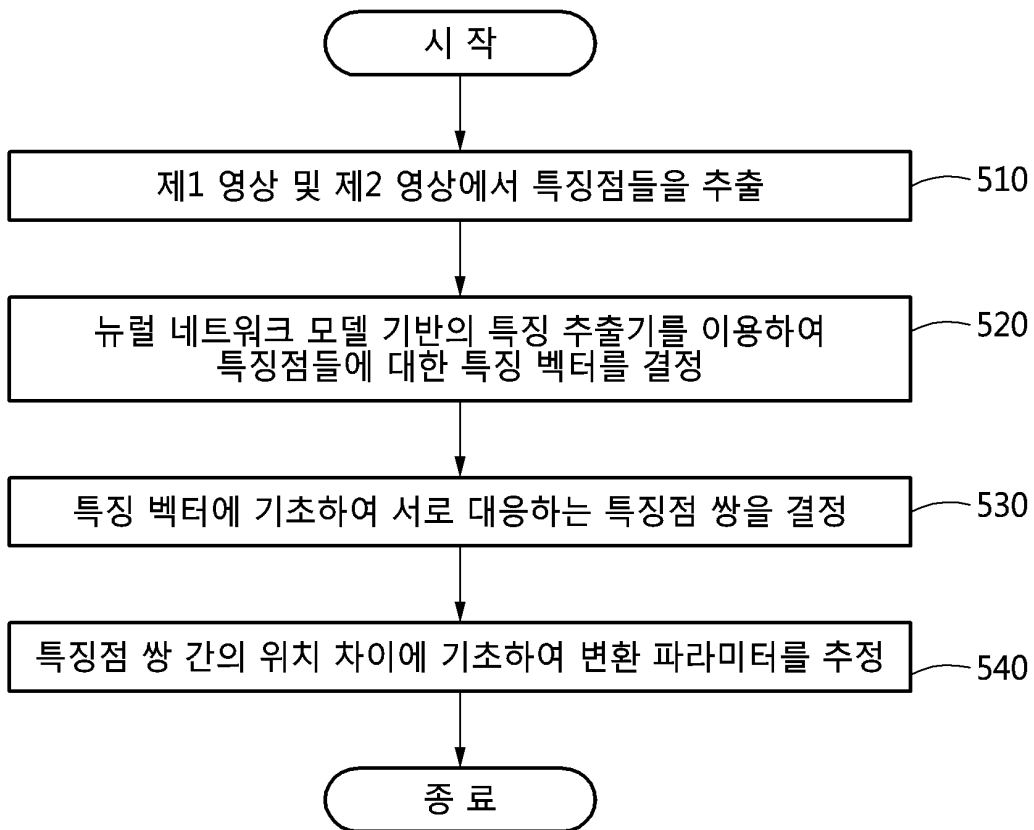
도면3



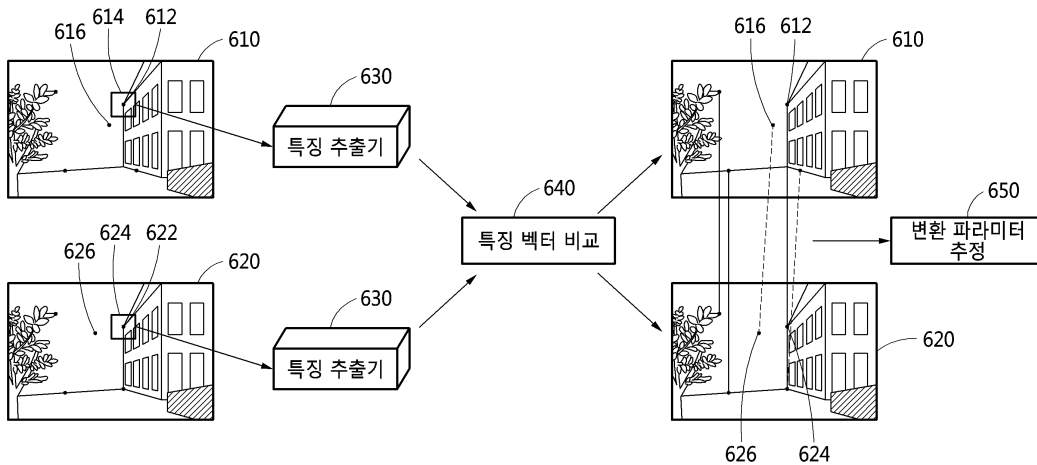
도면4



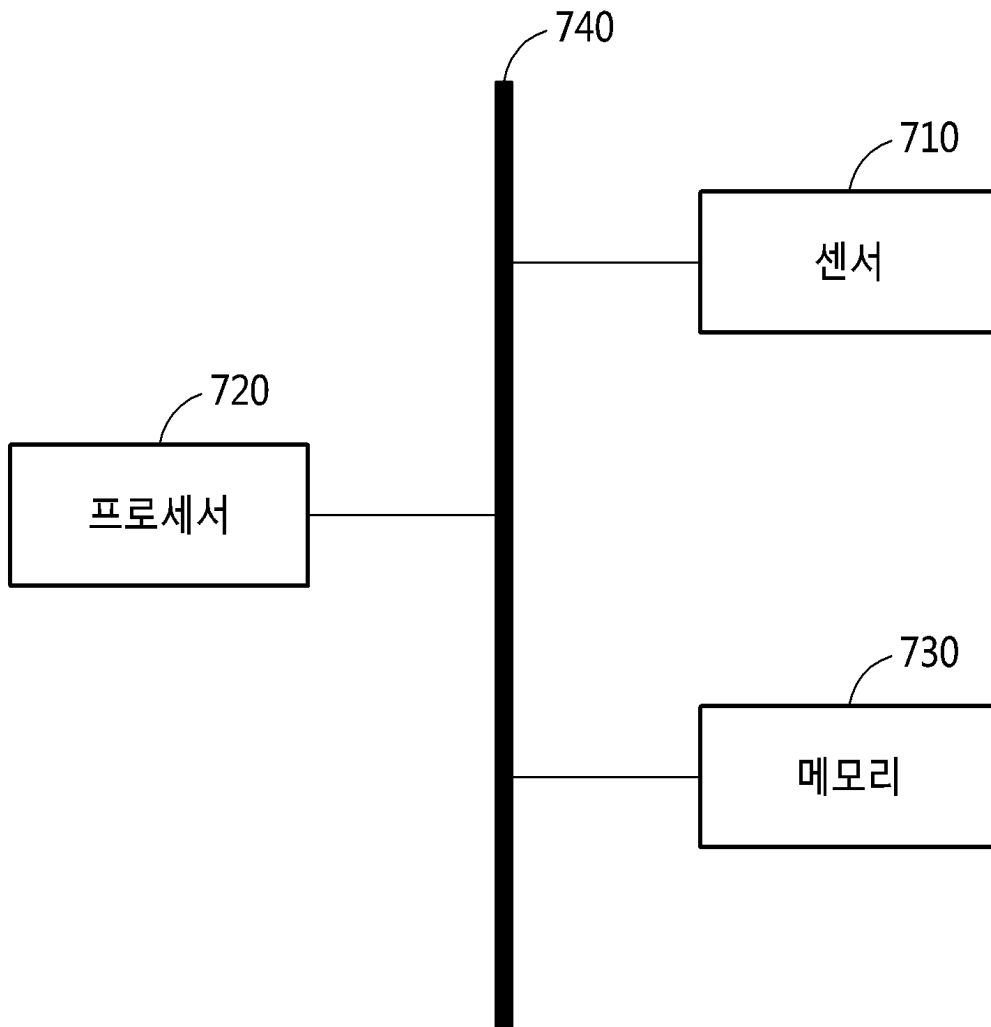
도면5



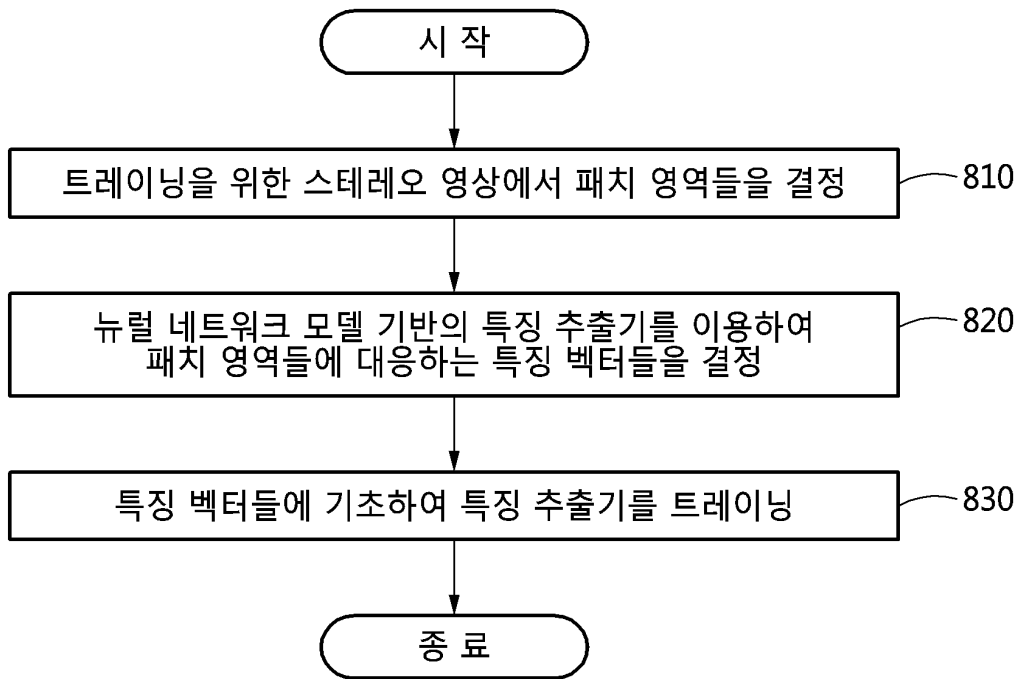
도면6



도면7



도면8



도면9

