(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0072750
(43) 공개일자 2025년05월26일

(51) 국제특허분류(Int. Cl.)
H04N 23/85 (2023.01) *G06N 3/0464* (2023.01)
G06N 3/048 (2023.01) *H04N 23/71* (2023.01)
H04N 23/88 (2023.01) *H04N 25/13* (2023.01)
(52) CPC특허분류
H04N 23/85 (2023.01)
G06N 3/0464 (2023.01)
(21) 출원번호 10-2023-0159707
(22) 출원일자 2023년11월17일
심사청구일자 없음

(71) 출원인
삼성전자주식회사
경기도 수원시 영통구 삼성로 129 (매탄동)
이화여자대학교 산학협력단
서울특별시 서대문구 이화여대길 52 (대현동, 이화여자대학교)
(72) 발명자
김우식
경기도 성남시 분당구 판교백현로 38, 107동 1404호 (백현동, THE SHARP 판교 퍼스트파크)
오혜진
서울특별시 마포구 동교로 68-5, 202호 (합정동, 하나빌)
(뒷면에 계속)
(74) 대리인
리앤목특허법인

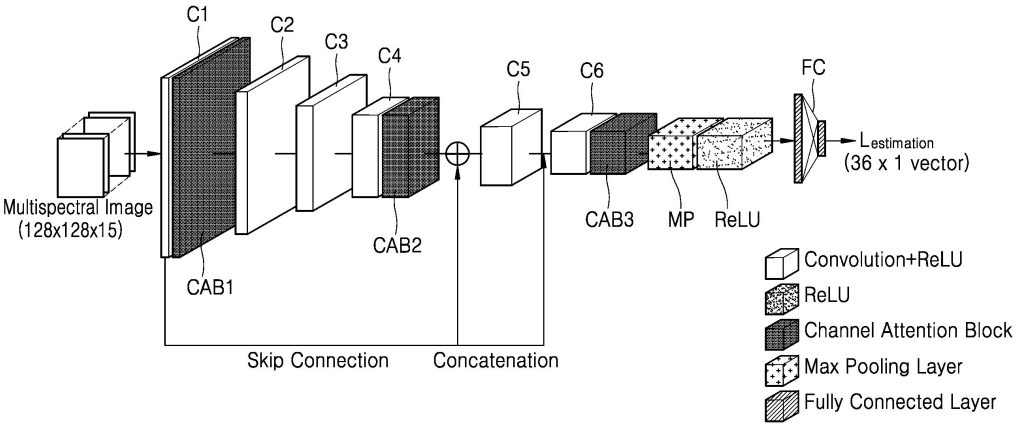
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 **영상 획득 장치 및 영상을 획득하는 방법**

(57) 요 약

영상 획득 장치에 있어서, 4개 이상의 채널들을 통해 영상을 획득하는 멀티스펙트럴 이미지 센서; 및 상기 획득된 영상을 미리 학습된 딥러닝 네트워크에 입력하여 조명 정보를 추정하고, 상기 추정된 조명 정보에 기초하여 상기 획득된 영상에 대한 색변환을 수행하는 프로세서;를 포함하고, 상기 딥러닝 네트워크는, 상기 추정된 조명 정보를 출력하기 위해 상기 채널들 간 채널 상관관계(channel correlation)를 학습하는, 영상 획득 장치.

대 표 도



(52) CPC특허분류
G06N 3/048 (2023.01)
H04N 23/71 (2023.01)
H04N 23/88 (2023.01)
H04N 25/135 (2023.01)
(72) 발명자
강제원
서울특별시 마포구 독막로 145, 104동 904호 (창전동, 서강쌍용예가)
이상윤
경기도 수원시 영통구 삼성로 130(매탄동)
최희진
경기도 수원시 영통구 영통로290번길 25, 514동 1502호 (영통동, 신나무실 주공아파트)
박영경
서울특별시 서초구 잠원로14길 42, 102동 904호(잠원동, 신반포르엘)

명 세 서

청구범위

청구항 1

영상 획득 장치에 있어서,

4개 이상의 채널들을 통해 영상을 획득하는 멀티스펙트럴 이미지 센서; 및

상기 획득된 영상을 미리 학습된 딥러닝 네트워크에 입력하여 조명 정보를 추정하고, 상기 추정된 조명 정보에 기초하여 상기 획득된 영상에 대한 색변환을 수행하는 프로세서;를 포함하고,

상기 딥러닝 네트워크는, 상기 추정된 조명 정보를 출력하기 위해 상기 채널들 간 채널 상관관계(channel correlation)를 학습하는, 영상 획득 장치.

청구항 2

제1항에 있어서,

상기 프로세서는, 상기 딥러닝 네트워크를 구성하는 복수의 레이어들 중 적어도 하나의 레이어에 상기 채널 상관관계를 입력하는, 영상 획득 장치.

청구항 3

제2항에 있어서,

상기 채널 상관관계는, 상기 획득된 영상에 대응하는 제1 채널들 간 제1 채널 상관관계, 및 상기 복수의 레이어들 중 어느 하나의 레이어에서 출력되는 중간 영상에 대응하는 제2 채널들 간 제2 채널 상관관계 중 적어도 하나를 포함하는, 영상 획득 장치.

청구항 4

제1항에 있어서,

상기 딥러닝 네트워크는,

복수의 컨볼루션 레이어들을 포함하는 컨볼루션 블록, 및

상기 복수의 컨볼루션 레이어들 중 적어도 하나의 뒤에 위치하는 채널 어텐션 블록(Channel Attention Block)을 포함하고,

상기 채널 어텐션 블록은 풀링(Pooling) 레이어 및 제1 완전 연결 층(Fully Connected Layer)을 포함하는, 영상 획득 장치.

청구항 5

제4항에 있어서,

상기 프로세서는, 상기 획득된 영상에 대해 전역 풀링(global pooling)을 수행하여 제1 채널 벡터를 생성하고, 상기 생성된 제1 채널 벡터를 상기 제1 완전 연결 층의 출력 결과에 더하는, 영상 획득 장치.

청구항 6

제4항에 있어서,

상기 프로세서는, 상기 복수의 컨볼루션 레이어들 중 제1 컨볼루션 레이어에서 출력되는 중간 영상에 대해 전역 풀링을 수행하여 제2 채널 벡터를 생성하고,

상기 채널 어텐션 블록은 상기 제1 컨볼루션 레이어에 후속하는 제2 컨볼루션 레이어 뒤에 위치하며,

상기 생성된 제2 채널 벡터는 상기 제1 완전 연결 층의 출력 결과에 더해지는, 영상 획득 장치.

청구항 7

제4항에 있어서,

상기 채널 어텐션 블록은, 렐루 함수, 제2 완전 연결 층 및 시그모이드 함수(Sigmoid function) 중 적어도 하나를 더 포함하는, 영상 획득 장치.

청구항 8

제1항에 있어서,

상기 딥러닝 네트워크는, 상기 채널 상관관계를 학습하기 위한 적어도 하나의 셀프 어텐션 레이어를 포함하는, 영상 획득 장치.

청구항 9

제1항에 있어서,

상기 딥러닝 네트워크는, 제1 각 손실 및 제2 각 손실을 포함하는 손실 함수를 이용하여 미리 학습되고,

상기 제1 각 손실은, 상기 딥러닝 네트워크의 학습을 위한 훈련 영상의 제1 정답 조명 정보와 제1 추정 조명 정보에 기초하고,

상기 제2 각 손실은, 상기 제1 정답 조명 정보와 상기 제1 추정 조명 정보를 각각 등색 함수(Color Matching Function)를 이용하여 XYZ 색공간으로 변환한 제2 조명 정보와 제2 추정 조명 정보에 기초하는, 영상 획득 장치.

청구항 10

제1항에 있어서,

상기 획득된 영상이 제1 광원과 제2 광원이 존재하는 환경에서 획득된 혼합 조명 영상인 경우,

상기 프로세서는, 상기 혼합 조명 영상을 적어도 2개의 영역들로 분할하고, 상기 분할된 영역들 중 상기 제1 광원이 우세한(dominant) 제1 영역에 기초하여 제1 조명 정보를 추정하고, 상기 분할된 영역들 중 상기 제2 광원이 우세한 제2 영역에 기초하여 제2 조명 정보를 추정하고, 상기 추정된 제1 조명 정보 및 상기 추정된 제2 조명 정보에 기초하여 상기 혼합 조명 영상의 혼합 조명 정보를 추정하는, 영상 획득 장치.

청구항 11

제1항에 있어서,

상기 프로세서는, 상기 추정된 조명 정보에 대응하는 조명 벡터의 차원을 상기 획득된 영상에 대응하는 채널들의 개수보다 많도록 조정하거나 3개로 조정하는, 영상 획득 장치.

청구항 12

제1항에 있어서,

상기 딥러닝 네트워크는,

상기 획득된 영상에 대응하는 2D 영상 및 채널들에 기초하여 3D 컨볼루션을 수행하기 위한 3D 컨볼루션 레이어 및

상기 3D 컨볼루션 레이어 뒤에 위치하는 스펙트럴 채널 어텐션 블록을 포함하는, 영상 획득 장치.

청구항 13

제12항에 있어서,

상기 딥러닝 네트워크는, 상기 채널 상관관계를 학습하기 위한 스펙트럴 셀프 어텐션 레이어를 더 포함하는, 영상 획득 장치.

청구항 14

영상을 획득하는 방법에 있어서,

멀티스펙트럴 이미지 센서를 이용하여 4개 이상의 채널들을 통해 영상을 획득하는 단계;

상기 획득된 영상을 미리 학습된 딥러닝 네트워크에 입력하여 조명 정보를 추정하는 단계; 및

상기 추정된 조명 정보에 기초하여 상기 획득된 영상에 대한 색변환을 수행하는 단계;를 포함하고,

상기 딥러닝 네트워크는, 상기 추정된 조명 정보를 출력하기 위해 상기 채널들 간 채널 상관관계를 학습하는, 방법.

청구항 15

제14항에 있어서,

상기 조명 정보를 추정하는 단계는,

상기 딥러닝 네트워크에 상기 획득된 영상을 입력하는 단계; 및,

상기 딥러닝 네트워크를 구성하는 복수의 레이어들 중 적어도 하나의 레이어에 상기 채널 상관관계를 입력하는 단계;를 더 포함하는, 방법.

청구항 16

제15항에 있어서,

상기 채널 상관관계는, 상기 획득된 영상에 대응하는 제1 채널들 간 제1 채널 상관관계, 및 상기 복수의 레이어들 중 적어도 하나의 레이어에서 출력되는 중간 영상에 대응하는 제2 채널들 간 제2 채널 상관관계 중 적어도 하나를 포함하는, 방법.

청구항 17

제14항에 있어서,

상기 딥러닝 네트워크는,

복수의 컨볼루션 레이어들을 포함하는 컨볼루션 블록, 및

상기 복수의 컨볼루션 레이어들 중 적어도 하나의 뒤에 위치하는 채널 어텐션 블록을 포함하고,

상기 채널 어텐션 블록은 풀링 레이어 및 제1 완전 연결 층을 포함하는, 방법.

청구항 18

제17항에 있어서,

상기 조명 정보를 추정하는 단계는,

상기 획득된 영상에 대해 전역 풀링을 수행하여 제1 채널 벡터를 생성하는 단계; 및,

상기 생성된 제1 채널 벡터를 상기 제1 완전 연결 층의 출력 결과에 더하는 단계;를 더 포함하는, 방법.

청구항 19

제17항에 있어서,

상기 조명 정보를 추정하는 단계는,

상기 복수의 컨볼루션 레이어들 중 제1 컨볼루션 레이어에서 출력되는 중간 영상에 대해 전역 풀링을 수행하여 제2 채널 벡터를 생성하는 단계; 및

상기 생성된 제2 채널 벡터를 상기 제1 완전 연결 층의 출력 결과에 더하는 단계;를 더 포함하고,

상기 채널 어텐션 블록은 상기 제1 컨볼루션 레이어에 후속하는 제2 컨볼루션 레이어 뒤에 위치하는, 방법.

청구항 20

제14항에 있어서,

상기 조명 정보를 추정하는 단계는,

상기 추정된 조명 정보에 대응하는 조명 벡터의 차원을 상기 획득된 영상에 대응하는 채널들의 개수보다 많도록 조정하거나 3개로 조정하는 단계;를 더 포함하는, 방법.

발명의 설명

기 술 분 야

[0001] 본 개시는 영상을 획득하는 방법 및 영상 획득 장치에 관한다.

배 경 기 술

[0002] 영상 센서는 피사체로부터 입사되는 빛을 수광하고, 수광된 빛을 광전 변환하여 전기적 신호를 생성하는 장치이다.

[0003] 영상 센서는 컬러 표현을 위해, 통상, 적색광, 녹색광, 청색광을 선택적으로 투과시키는 필터 요소들의 어레이로 이루어진 컬러 필터를 사용하며, 각 필터 요소를 투과한 빛의 양을 센싱한 후, 영상 처리를 통해 피사체에 대한 컬러 영상을 형성한다.

[0004] 영상 센서를 통해 센싱된 값은 조명에 의해 영향을 받기 때문에, 카메라로 촬영된 영상의 컬러도 조명에 의해 영향을 받게 된다. 이러한 영향을 없애고 최대한 물체 고유의 색을 촬영할 수 있도록 하는 기술을 화이트밸런스라고 한다.

[0005] 종래의 화이트밸런스 기술은 RGB 영상을 촬영한 후 그 영상 내 정보를 분석하여 화이트밸런스를 수행한다. 이러한 방법은 Gray World Assumption, 즉 영상의 R, G, B 채널별 평균값이 동일하다는 가정 또는 다른 제약 사항이 있기 때문에, 그러한 제약 사항을 만족시키지 못하는 상황에서는 올바로 동작하지 않을 수 있다.

발명의 내용

해결하려는 과제

[0006] 다양한 실시 영상을 획득하는 방법 및 영상 획득 장치를 제공하는데 있다. 본 개시가 이루고자 하는 기술적 과제는 상기된 바와 같은 기술적 과제들로 한정되지 않으며, 이하의 실시 예들로부터 또 다른 기술적 과제들이 유추될 수 있다.

과제의 해결 수단

[0007] 일 측면에 따르면, 영상 획득 장치는 4개 이상의 채널들을 통해 영상을 획득하는 멀티스펙트럴 이미지 센서; 및 상기 획득된 영상을 미리 학습된 딥러닝 네트워크에 입력하여 조명 정보를 추정하고, 상기 추정된 조명 정보에 기초하여 상기 획득된 영상에 대한 색변환을 수행하는 프로세서;를 포함하고, 상기 딥러닝 네트워크는, 상기 추정된 조명 정보를 출력하기 위해 상기 채널들 간 채널 상관관계(channel correlation)를 학습할 수 있다.

[0008] 일 측면에 따르면, 영상을 획득하는 방법은 멀티스펙트럴 이미지 센서를 이용하여 4개 이상의 채널들을 통해 영상을 획득하는 단계; 상기 획득된 영상을 미리 학습된 딥러닝 네트워크에 입력하여 조명 정보를 추정하는 단계; 및 상기 추정된 조명 정보에 기초하여 상기 획득된 영상에 대한 색변환을 수행하는 단계;를 포함하고,상기 딥러닝 네트워크는, 상기 추정된 조명 정보를 출력하기 위해 상기 채널들 간 채널 상관관계를 학습할 수 있다.

도면의 간단한 설명

[0009] 도 1은 일 실시 예에 따른 영상 획득 장치의 블록도이다.

도 2a는 일 실시 예에 따른 RGB 센서의 파장 스펙트럼을 도시한 도면이다.

도 2b는 일 실시 예에 따른 멀티스펙트럴 이미지 센서의 파장 스펙트럼을 도시한 도면이다.

도 2c는 다른 실시 예에 따른 멀티스펙트럴 이미지 센서의 파장 스펙트럼을 도시한 도면이다.

도 3은 일 실시 예에 따른 멀티스펙트럴 이미지 센서의 단면도이다.

도 4는 일 실시 예에 따른 멀티스펙트럴 이미지 센서의 픽셀 배열을 도시한 도면이다.

도 5a는 도 1의 멀티스펙트럴 이미지 센서로부터 획득한 RAW 영상을 도시한 도면이다.

도 5b는 도 1의 멀티스펙트럴 이미지 센서로부터 획득한 디모자이크가 수행된 후의 채널 별 영상을 도시한 도면이다.

도 6은 일 실시 예에 따른 조명 정보를 추정하기 위한 딥러닝 네트워크를 설명하기 위한 도면이다.

도 7은 다른 실시 예에 따른 조명 정보를 추정하기 위한 딥러닝 네트워크를 설명하기 위한 도면이다.

도 8a는 일 실시 예에 따른 도 7의 채널 어텐션 블록을 설명하기 위한 도면이다.

도 8b는 다른 실시 예에 따른 도 7의 채널 어텐션 블록을 설명하기 위한 도면이다.

도 9a는 일 실시 예에 따른 트랜스포머의 구조를 설명하기 위한 도면이다.

도 9b는 도 9a의 멀티헤드 어텐션 레이어의 구조를 설명하기 위한 도면이다.

도 9c는 도 9b의 멀티헤드 어텐션 레이어의 연산 결과를 나타낸 도면이다.

도 10은 또 다른 실시 예에 따른 조명 정보를 추정하기 위한 딥러닝 네트워크를 설명하기 위한 도면이다.

도 11a는 3D 컨볼루션 레이어를 포함하는 딥러닝 네트워크를 설명하기 위한 도면이다.

도 11b는 도 11a의 스펙트럴 채널 어텐션 블록을 설명하기 위한 도면이다.

도 11c는 도 10a의 스펙트럴 셀프 어텐션 블록을 설명하기 위한 도면이다.

도 12a는 일 실시 예에 따른 추가 딥러닝 네트워크를 설명하기 위한 도면이다.

도 12b는 다른 실시 예에 따른 추가 딥러닝 네트워크를 설명하기 위한 도면이다.

도 13은 일 실시 예에 따른 영상을 획득하는 방법을 설명하기 위한 순서도이다.

도 14는 일 실시 예에 따른 전자 장치의 구성을 나타낸 블록도이다.

도 15은 도 14의 전자 장치에 구비되는 카메라 모듈을 도시한 블록도이다.

도 16 내지 도 25는 다양한 실시 예들에 따른 영상 획득 장치가 적용될 전자 장치의 다양한 예시들을 도시한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0010] 본 실시예들에서 사용되는 용어는 본 실시예들에서의 기능을 고려하면서 가능한 현재 널리 사용되는 일반적인 용어들을 선택하였으나, 이는 당 기술분야에 종사하는 기술자의 의도 또는 판례, 새로운 기술의 출현 등에 따라 달라질 수 있다. 또한, 특정한 경우는 임의로 선정된 용어도 있으며, 이 경우 해당 실시예의 설명 부분에서 상세히 그 의미를 기재할 것이다. 따라서, 본 실시예들에서 사용되는 용어는 단순한 용어의 명칭이 아닌, 그 용어가 가지는 의미와 본 실시 예들의 전반에 걸친 내용을 토대로 정의되어야 한다.

[0011] 실시 예들에 대한 설명들에서, 어떤 부분이 다른 부분과 연결되어 있다고 할 때, 이는 직접적으로 연결되어 있는 경우뿐 아니라, 그 중간에 다른 구성요소를 사이에 두고 전기적으로 연결되어 있는 경우도 포함한다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 또한 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.

[0012] 본 실시예들에서 사용되는 "구성된다" 또는 "포함한다" 등의 용어는 명세서 상에 기재된 여러 구성 요소들, 또는 여러 단계들을 반드시 모두 포함하는 것으로 해석되지 않아야 하며, 그 중 일부 구성 요소들 또는 일부 단계들은 포함되지 않을 수도 있고, 또는 추가적인 구성 요소 또는 단계들을 더 포함할 수 있는 것으로 해석되어야 한다.

[0013] 또한, 본 명세서에서 사용되는 '제 1' 또는 '제 2' 등과 같이 서수를 포함하는 용어는 다양한 구성 요소들을 설명하는데 사용할 수 있지만, 상기 구성 요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성 요소를 다른 구성 요소로부터 구별하는 목적으로 사용될 수 있다.

[0014] 하기 실시 예들에 대한 설명은 권리범위를 제한하는 것으로 해석되지 말아야 하며, 해당 기술분야의 당업자가 용이하게 유추할 수 있는 것은 실시 예들의 권리범위에 속하는 것으로 해석되어야 할 것이다. 이하 첨부된 도면들을 참조하면서 오로지 예시를 위한 실시 예들을 상세히 설명하기로 한다.

[0015] **도 1은 일 실시 예에 따른 영상 획득 장치의 블록도이다.** 도 1을 참조하면, 일 실시 예에 따른 영상 획득 장치(100)는 멀티스펙트럴 이미지 센서(200) 및 프로세서(300)를 포함할 수 있다. 다만, 도 1의 영상 획득 장치(100)에는 본 실시 예들과 관련된 구성요소들만이 도시되어 있으며, 이에 제한되는 것은 아니다. 영상 획득 장치(100)의 설계에 따라, 도 1에 도시된 구성 중 일부가 생략되거나 새로운 구성(예: 메모리) 이 더 추가될 수 있음을 본 실시 예들과 관련된 기술분야에서 통상의 지식을 가진 자라면 이해할 수 있다. 이하에서는 영상 획득 장치(100)에 포함된 각 구성요소들이 위치하는 공간을 한정하지 않고, 각 구성요소들의 동작에 대해 설명하도록 한다.

[0016] 멀티스펙트럴 이미지 센서(200)는 다양한 종류의 파장 대역의 광을 센싱하는 센서를 의미할 수 있다. 예를 들어, 멀티스펙트럴 이미지 센서(200)는 RGB 센서보다 더 많은 종류의 파장 대역의 광을 센싱할 수 있다. 도 2a를 참조하면, RGB 센서는 R채널, G채널 및 B채널을 포함할 수 있고, RGB 센서는 3개의 채널들 각각에 대응되는 파장 대역의 광을 센싱할 수 있다. 이와 달리, 도 2b 및 도 2c를 참조하면, 멀티스펙트럴 이미지 센서(200)는 16개의 채널들을 포함하거나 31개의 채널들을 포함할 수 있다. 다만, 멀티스펙트럴 이미지 센서(200)에 포함된 채널들의 개수가 이에 제한되는 것은 아니며, 멀티스펙트럴 이미지 센서(200)는 4개 이상인 임의의 개수의 채널들을 포함할 수 있다.

[0017] 멀티스펙트럴 이미지 센서(200)는 각 채널을 통해 흡수되는 광의 중심 파장, 대역폭 및 투과량을 조정할 수 있다. 예를 들어, 멀티스펙트럴 이미지 센서(200)의 각 채널의 대역폭은 R채널, G채널 및 B채널의 대역폭보다 좁게 설정될 수 있다. 다른 예를 들어, 멀티스펙트럴 이미지 센서(200)의 모든 채널들이 대역폭들을 합한 전체 대역폭은 RGB 센서의 전체 대역폭을 포함하며 이보다 넓게 설정될 수도 있다. 또 다른 예를 들어, 멀티스펙트럴 이미지 센서(200)가 획득하는 영상은 멀티스펙트럴 또는 하이퍼스펙트럴 영상일 수 있다. 멀티스펙트럴 이미지 센서(200)는 복수의 채널들을 통해 가시광선 대역, 적외선 대역 및 자외선 대역을 포함하는 파장 대역을 복수 개로 분할하여 영상을 획득할 수 있다. 멀티스펙트럴 이미지 센서(200)는 가용한 모든 채널을 이용하여 영상을 획득할 수 있으나, 특정 채널을 선별하여 영상을 획득할 수도 있다.

[0018] 프로세서(300)는 영상 획득 장치(10)의 전반적인 동작을 제어한다. 프로세서(300)는 하나의 프로세서 코어(Single Core)를 포함하거나, 복수의 프로세서 코어들(Multi-Core)을 포함할 수 있다. 프로세서(300)는 메모리에 저장된 프로그램들 및/또는 데이터를 처리 또는 실행할 수 있다. 예를 들어, 프로세서(300)는 메모리에 저장된 프로그램들을 실행함으로써 영상 획득 장치(100)의 기능을 제어할 수 있다.

[0019] 도시하진 않았으나 영상 획득 장치(100)는 메모리를 더 포함할 수 있다. 메모리는 영상 획득 장치(100) 내에서 처리되는 각종 데이터를 저장하는 하드웨어로서, 예를 들어, 메모리는 멀티스펙트럴 이미지 센서(200)로부터 획득되는 영상을 저장할 수 있다. 메모리는 라인 단위의 영상을 순차적으로 저장하는 라인 메모리일 수 있으며, 영상 전체를 저장하는 프레임 버퍼일 수도 있다. 또한, 메모리는 영상 획득 장치(100)에 의해 구동될 애플리케이션들, 드라이버들 등을 저장할 수 있다. 메모리는 DRAM(dynamic random access memory), SRAM(static random access memory) 등과 같은 RAM(random access memory), ROM(read-only memory), EEPROM(electrically erasable programmable read-only memory), CD-ROM, 블루레이 또는 다른 광학 디스크 스토리지, HDD(hard disk drive), SSD(solid state drive), 또는 플래시 메모리를 포함할 수 있다. 다만, 이에 제한되는 것은 아니다.

[0020] 메모리는 멀티스펙트럴 이미지 센서(200)의 외부에 위치할 수 있고, 멀티스펙트럴 이미지 센서(200)의 내부에 집적될 수도 있다. 메모리가 멀티스펙트럴 이미지 센서(200)의 내부에 집적될 경우, 메모리는 회로부와 함께 집적될 수 있다. 화소부와 그 외의 부분(즉, 회로부와 메모리)이 각각 하나의 스택이 되어 총 2개의 스택으로 집적될 수 있다. 이 경우 멀티스펙트럴 이미지 센서(200)는 2개의 스택을 포함하는 하나의 칩으로 구성될 수 있다. 다만, 이에 제한되는 것은 아니며, 멀티스펙트럴 이미지 센서(200)는 화소부, 회로부 및 메모리의 세 층을 갖는 3 스택으로 구현될 수도 있다.

[0021] 한편, 멀티스펙트럴 이미지 센서(200)에 포함되는 회로부는 프로세서(300)와 동일할 수도 있고, 상이할 수도 있

다. 멀티스펙트럴 이미지 센서(200)에 포함되는 회로부가 프로세서(300)와 동일한 경우, 영상 획득 장치(100)는 온 칩으로 구현되는 멀티스펙트럴 이미지 센서(200) 자체일 수 있다. 또한, 멀티스펙트럴 이미지 센서(200)에 포함되는 회로부가 프로세서(300)와 상이한 경우라도 프로세서(300)가 멀티스펙트럴 이미지 센서(200) 내부에 배치된다면, 영상 획득 장치(100)는 온 칩으로 구현될 수 있다. 다만, 이에 제한되는 것은 아니며, 프로세서(300)는 멀티스펙트럴 이미지 센서(200)의 외부에 별도로 위치할 수도 있다.

[0022] 프로세서(300)는 멀티스펙트럴 이미지 센서(200)의 각 채널에 대응하는 출력 신호인 채널 신호들을 획득할 수 있다. 프로세서(300)는 멀티스펙트럴 이미지 센서(200)에 물리적으로 구비된 소정 개수의 채널들 중 적어도 일부를 선택하고, 선택된 채널들로부터 채널 신호들을 획득할 수 있다. 예를 들어, 프로세서(300)는 멀티스펙트럴 이미지 센서(200)에 물리적으로 구비된 소정 개수의 채널들 모두로부터 채널 신호들을 획득할 수 있다. 또한, 프로세서(300)는 멀티스펙트럴 이미지 센서(200)에 물리적으로 구비된 소정 개수의 채널들 중 일부만을 선택하여 채널 신호들을 획득할 수 있다.

[0023] 프로세서(300)는 멀티스펙트럴 이미지 센서(200)에 물리적으로 구비된 소정 개수의 채널들로부터 획득되는 채널 신호들을 합성 또는 보간함으로써, 소정 개수보다 증가되거나 감소된 개수의 채널 신호들을 획득할 수도 있다. 예를 들어, 프로세서(300)는 멀티스펙트럴 이미지 센서(200)의 화소들 또는 채널들에 대한 비닝(binning)을 수행함으로써, 소정 개수보다 감소된 개수의 채널 신호들을 획득할 수 있다. 또한, 프로세서(300)는 채널 신호들의 보간을 통해 새로운 채널 신호를 생성함으로써, 소정 개수보다 증가된 개수의 채널 신호들을 획득할 수 있다.

[0024] 획득되는 채널 신호들의 개수가 감소되는 경우, 채널 신호들 각각은 광대역(wide band)에 해당할 수 있는바, 신호의 감도가 증가되고 노이즈가 감소될 수 있다. 반대로, 획득되는 채널 신호들의 개수가 증가되는 경우, 채널 신호들 각각의 감도는 다소 감소될 수 있으나, 다수의 채널 신호들에 기초하여 더 정밀한 영상이 획득될 수 있다. 이와 같이, 획득되는 채널 신호들의 개수의 증감에 따른 트레이드 오프(trade off)가 존재하므로, 프로세서(300)는 응용에 따라 적절한 개수의 채널 신호들을 획득할 수 있다.

[0025] 프로세서(300)는 멀티스펙트럴 이미지 센서(200)에 의해 획득된 영상 또는 신호가 메모리에 저장되기 전 또는 후에 기본적인 영상 처리를 수행할 수 있다. 기본적인 영상 처리에는 불량 픽셀 보정(bad pixel correction), 고정 패턴 잡음 보정(fixed pattern noise correction), 크로스토크 감쇄(crosstalk reduction), 재모자이크(remosaicing), 디모자이크(demosaicing), 오색상 감쇄(false color reduction), 노이즈 제거(denoising), 색수차 보정(chromatic aberration correction) 등이 포함될 수 있다.

[0026] 프로세서(300)는 채널 신호들에 대해 디모자이크를 수행함으로써 채널 별 영상을 생성하고, 채널 별 영상에 대한 영상 처리를 수행할 수도 있다. 도 5a 및 도 5b를 참조하면, 멀티스펙트럴 이미지 센서(200)로부터 획득된 RAW 영상(610)과 디모자이크가 수행된 이후의 채널 별 영상(620)이 도시되어 있다. RAW 영상(610)에서 하나의 작은 사각형은 하나의 화소를 나타내고, 사각형 내의 숫자는 채널 번호를 나타낸다. 채널 번호로부터 알 수 있듯이, RAW 영상(610)은 16개의 채널들을 포함하는 멀티스펙트럴 이미지 센서(200)에 의해 획득된 영상일 수 있다. RAW 영상(610)에는 서로 다른 채널에 해당하는 화소들이 모두 포함되어 있으나, 디모자이크를 통해 동일한 채널의 화소들이 취합됨에 따라 채널 별 영상(620)이 생성될 수 있다.

[0027] 프로세서(300)는 4개 이상의 채널들을 통해 멀티스펙트럴 이미지 센서(200)로부터 획득된 영상을 미리 학습된 딥러닝 네트워크에 입력하여 조명 정보를 추정할 수 있다. 본 개시의 조명 정보는 파장에 따라 멀티스펙트럴 이미지 센서로부터 획득하는 채널 신호들의 세기(intensity)에 관한 정보를 의미할 수 있다.

[0028] 멀티스펙트럴 이미지 센서(200)는 4개 이상의 채널들을 통해 멀티스펙트럴 영상을 획득하거나 9개 이상의 채널들을 통해 하이퍼스펙트럴 영상을 획득할 수 있다. 멀티스펙트럴 이미지 센서(200)는 물리적으로 구비된 채널들 전부 또는 일부를 선택하여 채널 신호들을 획득할 수 있으며, 프로세서(300)는 획득된 채널 신호들을 합성 또는 보간함으로써 획득할 채널 신호들의 개수를 조정할 수 있다.

[0029] 예를 들어, 멀티스펙트럴 이미지 센서(200)는 채널들을 통해 RAW 영상(예: 도 5a의 RAW 영상(610))을 획득할 수 있고, 프로세서(300)는 획득한 RAW 영상에 디모자이크를 수행하여 채널 별 영상(예: 도 5b의 채널 별 영상(620))을 생성할 수 있다. 프로세서(300)는 생성된 채널 별 영상을 딥러닝 네트워크에 입력하여 조명 정보를 추정할 수 있다.

[0030] 프로세서(300)는 획득된 영상의 전체 또는 일부를 딥러닝 네트워크에 입력할 수 있다. 프로세서(300)는 획득된 영상의 크기를 조정하는 스케일링을 수행할 수 있으며, 프로세서(300)는 스케일링이 수행되지 않은 영상과 스케

일링이 수행된 영상을 모두 딥러닝 네트워크에 입력할 수 있다. 다만, 이에 제한되는 것은 아니며, 프로세서(300)는 기타 다른 변환이 수행된 영상을 딥러닝 네트워크에 입력할 수도 있다.

[0031] 딥러닝 네트워크에 멀티스펙트럴 이미지 센서(200)에 의해 획득된 영상이 입력됨에 따라, 딥러닝 네트워크는 획득된 영상에 대한 조명 정보를 추정할 수 있다. 딥러닝 네트워크는 지도 학습 방식(supervised learning method)을 이용하여 미리 학습될 수 있다. 예를 들어, 딥러닝 네트워크는 딥러닝 네트워크에 입력된 영상들 각각에 대하여 추정된 조명 정보, 정답(Ground Truth) 조명 정보 및 손실 함수(loss function) 등을 이용하여 학습될 수 있다. 딥러닝 네트워크는 획득된 영상에 대응하는 채널들 간의 채널 상관관계(channel correlation)에 대해 미리 학습될 수 있다. 본 개시에서 채널 상관관계는 멀티스펙트럴 이미지 센서(200)에 의해 획득되는 영상에 대응하는 채널들 각각의 중요도 및/또는 연관성에 관한 정보를 의미할 수 있다.

[0032] 멀티스펙트럴 이미지 센서(200)에 의해 획득되는 영상은 영상의 촬영 시 존재하는 광원을 스펙트럼의 함수로 나타낸 값과 대상체(object)의 표면 반사(surface reflectance)를 스펙트럼의 함수로 나타낸 값의 곱으로 표현될 수 있다. 채널에 입사되는 광원의 세기가 강할수록 해당 채널을 통해 획득되는 신호의 크기는 커질 수 있다. 채널 별로 획득되는 신호의 크기에 기초하여, 채널 별 중요도 및 채널들 간 연관성에 관한 정보가 결정될 수 있다.

[0033] 채널 상관관계에 대해 학습된 딥러닝 네트워크는 딥러닝 네트워크에 입력되는 영상에 대응하는 채널들 간 채널 상관관계에 따라 채널 별로 획득되는 신호들 각각에 상이한 가중치를 부여함으로써 보다 정확한 조명 정보를 추정할 수 있다. 채널 상관관계에 대해 미리 학습된 딥러닝 네트워크의 구조 및 학습 방식에 대하여 도 7에서 구체적으로 후술하도록 한다.

[0034] 프로세서(300)는 딥러닝 네트워크로부터 추정된 조명 정보를 벡터의 형태로 출력할 수 있다. 프로세서(300)는 딥러닝 네트워크로부터 추정된 조명 정보를 추정된 조명 정보에 대응하는 조명 벡터로 출력할 수 있다. 본 개시에서 조명 벡터는 각 채널에 대응하는 채널 신호의 크기를 벡터의 요소들로 구성하는 벡터를 의미할 수 있다. 다만, 프로세서(300)는 조명 벡터를 다른 벡터의 형태로 변환하여 출력할 수도 있다. 예를 들어, 프로세서(300)는 조명 벡터를 XYZ 색공간 상의 XYZ 값을 벡터의 요소들로 구성하는 3차원 벡터로 변환하여 출력하거나, 조명 벡터를 RGB 색공간 상의 RGB 값을 벡터의 요소들로 구성하는 3차원 벡터로 변환하여 출력할 수도 있다. 다만, 이에 제한되는 것은 아니며, 프로세서(300)는 조명 벡터를 추정된 조명 정보에 대응하는 색 온도 값을 벡터의 요소들로 구성하는 벡터 또는 미리 정의된 조명 타입을 나타내는 인덱스 값을 벡터의 요소들로 구성하는 벡터로 출력할 수도 있다.

[0035] 프로세서(300)가 출력하는 조명 벡터의 차원은 딥러닝 네트워크에 입력되는 영상에 대응하는 채널들의 개수와 동일할 수 있다. 예를 들어, 프로세서(300)는 16개의 채널들을 통해 멀티스펙트럴 이미지 센서(200)로부터 획득한 영상을 딥러닝 네트워크에 입력하여 추정된 조명 정보를 16차원을 가지는 조명 벡터로 출력할 수 있다. 다만, 이에 제한되는 것은 아니다. 프로세서(300)는 조명 벡터의 차원을 멀티스펙트럴 이미지 센서(200)에 의해 획득되는 영상에 대응하는 채널들의 개수보다 많도록 조정할 수도 있고, 프로세서(300)는 조명 벡터의 차원을 3으로 조정할 수도 있다.

[0036] 프로세서(300)는 추정된 조명 정보에 기초하여 멀티스펙트럴 이미지 센서(200)에 의해 획득된 영상에 대한 색변환을 수행할 수 있다. 예를 들어, 프로세서(300)는 추정된 조명 정보에 대응하는 조명 벡터의 각 요소들을 등색함수(Color Matching Function)를 이용하여 XYZ 색공간으로 변환함으로써 멀티스펙트럴 이미지 센서(200)에 의해 획득된 영상에 대하여 색변환을 수행할 수 있다. 본 개시에서 색공간은 사람의 눈이 인식하는 색을 공간 좌표 상에 나타낸 것을 의미할 수 있고, 색변환은 각 픽셀에서 컬러 필터를 통과한 후 이미지 센서에 기록된 채널 신호들을 색공간에 해당하는 신호들로 변환하는 것을 의미할 수 있다.

[0037] 프로세서(300)는 추정된 조명 정보에 기초하여 멀티스펙트럴 이미지 센서(200)에 의해 획득된 영상에 대한 자동 화이트 밸런스(Auto White Balance)를 수행할 수 있다. 멀티스펙트럴 이미지 센서(200)에 의해 획득되는 영상은 영상이 촬영된 환경에 존재하는 광원에 의해 영상의 색이 달라지는 현상이 나타날 수 있다. 예를 들어, 흰색 종이를 광원이 존재하는 환경에서 촬영하는 경우 멀티스펙트럴 이미지 센서(200)에 의해 획득되는 영상에서 종이의 색은 흰색으로 구현되지 않을 수 있다. 본 개시에서 화이트 밸런스 보정은 광원에 의해 영상의 색이 달라지는 현상을 없애고 획득한 영상에서 최대한 대상체의 고유의 색이 구현될 수 있도록 하는 보정 방식을 의미할 수 있다.

[0038] 다만, 프로세서(300)가 추정된 조명 정보에 기초하여 색변환 및/또는 자동 화이트 밸런스를 수행하는 대상은 멀

티스펙트럴 이미지 센서(200)에 의해 획득된 영상에 한정되지 않으며, 프로세서(300)는 추정된 조명 정보에 기초하여 다른 이미지 센서에 의해 획득한 영상에 대한 색변환 및/또는 자동 화이트 밸런스를 수행할 수도 있다. 예를 들어, 영상을 획득하는 다른 이미지 센서로서, 스마트폰의 메인 카메라에 구비된 이미지 센서 등이 이용될 수 있다.

[0039] 예를 들어, 영상 획득 장치는 동일한 광원이 존재하는 환경에서 동일한 대상체에 대하여 동시에 멀티스펙트럴 이미지 센서(200)와 RGB 센서에 의해 영상을 획득할 수 있다. 프로세서(300)는 멀티스펙트럴 이미지 센서(200)에 의해 획득된 영상을 딥러닝 네트워크에 입력하여 추정된 조명 정보(이하, 하이퍼 조명 정보라고 지칭함)를 획득할 수 있고, 이와 동시에 RGB 센서에 의해 획득된 영상을 딥러닝 네트워크에 입력하여 추정된 조명 정보(이하, RGB 조명 정보라고 지칭함)를 획득할 수 있다.

[0040] 프로세서(300)는 하이퍼 조명 정보에 대응하는 하이퍼 조명 벡터의 각 요소들을 XYZ 색공간 상의 XYZ 값으로 변환한 후, RGB 조명 정보에 대응하는 RGB 조명 벡터의 각 요소들을 XYZ 값으로 변환된 조명 벡터의 각 요소들로 나눔에 따라, RGB 센서에 의해 획득된 영상에 대한 자동 화이트 밸런스를 수행할 수 있다.

[0041] 이하에서는 조명 정보를 추정하기 위한 딥러닝 네트워크를 설명하기에 앞서, 멀티스펙트럴 이미지 센서(200)의 구성요소들에 대해 먼저 설명하도록 한다.

[0042] **도 3은 일 실시 예에 따른 멀티스펙트럴 이미지 센서의 단면도이다.**

[0043] 도 3에 도시된 멀티스펙트럴 이미지 센서(200)는 예를 들면, CMOS(complementary metal oxide semiconductor) 이미지 센서 또는 CCD(charge coupled device) 이미지 센서를 포함할 수 있다.

[0044] 도 3을 참조하면, 멀티스펙트럴 이미지 센서(200)는 화소 어레이(또는 픽셀 어레이)(65)와, 이 화소 어레이 상에 마련되는 분광 필터(83)를 포함할 수 있다. 여기서, 화소 어레이(65)는. 2차원적으로 배열되는 복수의 화소를 포함할 수 있으며, 분광 필터(83)는 복수의 화소에 대응되게 마련되는 복수의 공진기를 포함할 수 있다. 도 3에는 화소 어레이(65)가 4개의 화소를 포함하고, 분광 필터(83)가 4개의 공진기를 포함하는 경우가 예시적으로 도시되어 있다.

[0045] 화소 어레이(65)의 각 화소는 광전 변환 소자인 포토다이오드(photodiode, 62)와 이 포토다이오드(62)를 구동시키기 위한 구동회로(52)를 포함할 수 있다. 포토다이오드(62)는 반도체 기판(61)에 매립되도록 마련될 수 있다. 반도체 기판(61)으로는 예를 들어 실리콘 기판이 사용될 수 있다. 하지만, 이에 한정되는 것은 아니다. 반도체 기판(61)의 하면(61a)에는 배선층(wiring layer, 51)이 마련될 수 있으며, 이 배선층(51)의 내부에 예를 들어, MOSFET 등과 같은 구동회로(52)가 마련될 수 있다.

[0046] 반도체 기판(61)의 상면(61b)에는 복수의 공진기를 포함하는 분광 필터(83)가 마련되어 있다. 각 공진기는 원하는 특정 파장 영역의 광을 투과시키도록 마련될 수 있다. 각 공진기는 서로 이격되게 마련되는 반사층들과, 반사층들 사이에 마련되는 캐비티를 포함할 수 있다. 반사층들 각각은 예를 들면, 금속 반사층 또는 브래그 반사층을 포함할 수 있다. 각 캐비티는 원하는 특정 파장 영역의 빛을 공진시키도록 마련될 수 있다.

[0047] 분광 필터(83)는 분광 필터(83)를 투과하여 포토다이오드(62) 쪽으로 입사되는 광의 투과도를 향상시키는 하나 이상의 기능층들을 포함할 수 있다. 기능층은 굴절률이 조절된 유전체층 또는 유전체 패턴을 포함할 수 있다. 또한, 기능층은 예를 들면, 반반사층(antireflection layer), 집속 렌즈, 컬러 필터, 단파장 흡수 필터 또는 장파장 차단 필터 등을 포함할 수 있다. 하지만 이는 단지 예시적인 것이다.

[0048] 이하에서, 멀티스펙트럴 이미지 센서(200)의 화소 어레이(또는 픽셀 어레이)에 대해 구체적으로 설명하도록 한다.

[0049] **도 4는 일 실시 예에 따른 멀티스펙트럴 이미지 센서의 픽셀 배열을 도시한 도면이다.**

[0050] 도 4를 참조하면, 픽셀 어레이(4100)에 구비되는 분광 필터(120)는 2차원 형태로 배열되는 복수의 필터 그룹(4110)들을 포함할 수 있다. 여기서, 각 필터 그룹(4110)은 4×4 어레이 형태로 배열되는 16개의 유닛 필터들(F1~F16)을 포함할 수 있다. 다만, 이에 제한되는 것은 아니다. 예를 들어, 각 필터 그룹(12)은 3×3 어레이 형태로 배열되는 9개의 유닛 필터들(F1~F9)을 포함하거나, 각 필터 그룹(12)은 5×5 어레이 형태로 배열되는 25개의 유닛 필터들(F1~F25)을 포함할 수 있으며, 각 필터 그룹(12)은 M×N(여기서, M 및 N은 1 이상인 임의의 정수를 의미함) 어레이 형태로 배열되는 M×N개의 유닛 필터들(~)을 포함할 수 있다.

[0051] 일 예시에 따라, 각 필터 그룹(4110)이 4×4 어레이 형태로 배열되는 경우, 제1 및 제2 유닛 필터(F1, F2)는 자

외선 영역의 중심 파장들(UV1, UV2)을 가질 수 있으며, 제3 내지 제5 유닛 필터(F3~F5)는 청색광 영역의 중심 파장들(B1~B3)을 가질 수 있다. 제6 내지 제11 유닛 필터(F6~F11)는 녹색광 영역의 중심 파장들(G1~G6)을 가질 수 있으며, 제12 내지 제14 유닛 필터(F12~F14)는 적색광 영역의 중심 파장들(R1~R3)을 가질 수 있다. 그리고 제15 및 제16 유닛 필터(F15, F16)는 근적외선 영역의 중심 파장들(NIR1, NIR2)을 가질 수 있다.

[0052] 다른 예시에 따라, 각 필터 그룹(4110)이 3×3 어레이 형태로 배열되는 경우, 제1 및 제2 유닛 필터(F1, F2)는 자외선 영역의 중심 파장들(UV1, UV2)을 가질 수 있으며, 제4, 제5 및 제7 유닛 필터(F4, F5, F7)는 청색광 영역의 중심 파장들(B1~B3)을 가질 수 있다. 제3 및 제6 유닛 필터(F3, F6)는 녹색광 영역의 중심 파장들(G1, G2)을 가질 수 있으며, 제8 및 제9 유닛 필터(F8, F9)는 적색광 영역의 중심 파장들(R1, R2)을 가질 수 있다.

[0053] 또 다른 예시에 따라, 각 필터 그룹(123)은 5×5 어레이 형태로 배열되는 경우, 제1 내지 제3 유닛 필터(F1~F3)는 자외선 영역의 중심 파장들(UV1~UV3)을 가질 수 있으며, 제6, 제7, 제8, 제11, 및 제12 유닛 필터(F6, F7, F8, F11, F12)는 청색광 영역의 중심 파장들(B1~B5)을 가질 수 있다. 제4, 제5, 및 제9 유닛 필터(F4, F5, F9)는 녹색광 영역의 중심 파장들(G1~G3)을 가질 수 있으며, 제10, 제13, 제14, 제15, 제18, 및 제19 유닛 필터(F10, F13, F14, F15, F18, F19)는 적색광 영역의 중심 파장들(R1~R6)을 가질 수 있다. 그리고 제20, 제23, 제24, 및 제25 유닛 필터(F20, F23, F24, F25)는 근적외선 영역의 중심 파장들(NIR1~NIR4)을 가질 수 있다.

[0054] 분광 필터(120)에 구비되는 상술한 유닛 필터들은 두 개의 반사판을 가지는 공진 구조를 가질 수 있으며 공진 구조의 특성에 따라 투과되는 파장 대역이 결정될 수 있다. 반사판의 재질 및 캐비티 내 유전 물질의 재질, 캐비티 두께에 따라 투과 파장 대역이 조절될 수 있다. 이 외에도, 그레이팅을 활용한 구조, DBR을 활용한 구조 등이 유닛 필터에 적용될 수 있다.

[0055] 이 외에도 픽셀 어레이(4100)의 픽셀들은 멀티스펙트럴 이미지 센서(100)의 색 특성에 따라 다양한 방식으로 배열될 수 있다. 이하에서는 상술한 멀티스펙트럴 이미지 센서(100)로부터 획득한 영상에 대한 조명 정보를 추정하기 위한 딥러닝 네트워크의 다양한 실시 예들을 설명하도록 한다.

[0056] **도 6은 일 실시 예에 따른 조명 정보를 추정하기 위한 딥러닝 네트워크를 설명하기 위한 도면이다.**

[0057] 도 6의 딥러닝 네트워크는 도 1 내지 도 5를 참조하여 설명한 영상 획득 장치(100)에서 도 1 내지 도 5를 참조하여 설명한 멀티스펙트럴 이미지 센서(200)로부터 획득된 영상에 대한 조명 정보를 추정하기 위한 딥러닝 네트워크의 일 실시 예이다. 이에 따라, 도 1 내지 도 5를 참조하여 설명한 영상 획득 장치(100)에 관하여 이상에서 설명된 내용은 도 6의 딥러닝 네트워크에도 적용될 수 있다.

[0058] 도 6을 참조하면, 컨볼루션 블록(CB)을 포함하는 딥러닝 네트워크는 멀티스펙트럴 이미지 센서(200)로부터 획득된 영상을 입력 받아 획득된 영상에 대한 조명 정보를 추정할 수 있다. 본 개시의 컨볼루션 블록(CB)은 하나 이상의 컨볼루션 레이어를 포함하는 신경망을 의미할 수 있다. 예를 들어, 컨볼루션 블록(CB)은 5개의 컨볼루션 레이어들(C1, C2, C3, C4, C5), 맥스 풀링 레이어(MP), 렐루 함수(Rectified Linear Unit function)(ReLU) 및 완전 연결 층(FC)를 포함할 수 있다. 컨볼루션 레이어들(C1, C2, C3, C4, C5)은 각각 렐루 함수(ReLU)를 활성화 함수로서 더 포함할 수 있다. 다만, 상술한 컨볼루션 블록(CB)은 일 실시 예에 지나지 않으며, 딥러닝 네트워크의 설계에 따라 다양하게 구성될 수 있음은 물론이다.

[0059] 프로세서(300)는 멀티스펙트럴 이미지 센서(200)에 의해 획득된 영상을 컨볼루션 블록(CB)을 포함하는 딥러닝 네트워크에 입력할 수 있다. 즉, 멀티스펙트럴 이미지 센서(200)에 의해 획득된 영상은 딥러닝 네트워크에 입력되어 딥러닝 네트워크의 첫번째 레이어(예: 도 6의 제1 컨볼루션 레이어(C1))에 입력될 수 있다. 컨볼루션 블록(CB)을 포함하는 딥러닝 네트워크에 입력된 영상인 입력 피쳐 맵(input feature map)은 복수 개의 컨볼루션 레이어들(C1, C2, C3, C4, C5)에 순차적으로 입력되어 컨볼루션 연산이 수행될 수 있다. 컨볼루션 연산의 수행 결과인 출력 피쳐 맵(output feature map)이 맥스 풀링 레이어(MP)에 입력되면 딥러닝 네트워크는 출력 피쳐 맵의 전역(global) 또는 로컬(local) 영역에 대하여 풀링을 수행함으로써 풀링이 수행된 대상 영역에 대한 최대값을 출력할 수 있다. 딥러닝 네트워크는 맥스 풀링 레이어(MP)로부터 출력된 결과에 렐루 함수(ReLU)를 활성화 함수로서 적용하여, 멀티스펙트럴 이미지 센서(200)에 의해 획득된 영상에 대한 조명 정보를 추정할 수 있다.

[0060] 딥러닝 네트워크는 완전 연결 층(FC)을 이용하여 추정된 조명 정보에 대응되는 출력 벡터의 차원을 조정할 수 있다. 예를 들어, 완전 연결 층(FC)의 출력 노드의 개수(C_o)는 딥러닝 네트워크에 입력되는 영상에 대응하는 채널들의 개수(C_i)와 동일하도록 설정되거나 3개로 설정될 수 있다. 다만, 이에 제한되는 것은 아니고, 완전 연결 층(FC)의 출력 노드의 개수(C_o)는 딥러닝 네트워크에 입력되는 영상에 대응하는 채널들의 개수(C_i)에 대하여 미

리 설정된 관계식에 의하여 결정될 수도 있다. 다음은 출력 노드의 개수(C_o)와 딥러닝 네트워크에 입력되는 영상에 대응하는 채널들의 개수(C_i)에 관한 예시적인 관계식들을 나타낸 것이다.

[0061] [수학식 1]

[0062] $$C_o = k * C_i$$

[0063] [수학식 2]

[0064] $$C_o = k * (C_i - a) + b$$

[0065] 수학식 1과 수학식 2에서 k는 1, 2, 4, 쯘 와 같은 2의 지수 승으로 결정될 수 있고, 수학식 2에서 a는 멀티스펙트럴 이미지 센서(200)에 의해 획득된 영상에 대응되는 채널들 중 딥러닝 네트워크에 입력되지 않는 채널들의 개수를 의미할 수 있으며, b는 추가적으로 생성된 딥러닝 네트워크로부터 출력되는 채널들의 개수를 의미할 수 있다.

[0066] 본 개시의 실시 예에 따른 영상 획득 장치는 딥러닝 네트워크로부터 추정되는 조명 정보에 대응하는 조명 벡터의 차원을 딥러닝 네트워크에 입력되는 영상에 대응하는 채널의 개수보다 많도록 조정함으로써, 조명 정보를 보다 정확하게 추정할 수 있고, 그에 따라 멀티스펙트럴 이미지 센서(200)로부터 획득되는 영상에 대한 자동 화이트 밸런스를 보다 효과적으로 수행할 수 있다.

[0067] 이하에서는 멀티스펙트럴 이미지 센서(100)로부터 획득한 영상에 대한 조명 정보를 추정하기 위한 딥러닝 네트워크로서, 채널 상관관계에 대해 미리 학습된 딥러닝 네트워크의 구조 및 학습 방식에 대하여 설명하도록 한다.

[0068] **도 7은 다른 실시 예에 따른 조명 정보를 추정하기 위한 딥러닝 네트워크를 설명하기 위한 도면이다.** 도 7의 딥러닝 네트워크는 도 6의 딥러닝 네트워크에서 딥러닝 네트워크에 채널 어텐션 블록(CAB)이 더 추가된 점만이 상이할 수 있으며, 이하에서 중복되는 설명은 생략하도록 한다.

[0069] 도 7을 참조하면, 채널 어텐션 블록(CAB)을 포함하는 딥러닝 네트워크는 멀티스펙트럴 이미지 센서(200)로부터 획득된 영상을 입력 받아 획득된 영상에 대한 조명 정보를 추정할 수 있다. 본 개시의 채널 어텐션 블록(CAB)은 딥러닝 네트워크에 입력되는 영상에 대응하는 채널들 간의 채널 상관관계를 학습하기 위한 신경망을 의미할 수 있다.

[0070] 채널 어텐션 블록(CAB)은 컨볼루션 레이어들(C1, C2, C3, C4, C5) 중 적어도 하나의 컨볼루션 레이어의 뒤에 위치하여, 채널 어텐션 블록(CAB)에 입력되는 영상에 대응하는 채널 신호들의 세기에 관한 정보를 해당 컨볼루션 레이어에 입력할 수 있다.

[0071] 채널 어텐션 블록(CAB)은 딥러닝 네트워크의 첫번째 레이어(예: 제1 컨볼루션 레이어(C1)) 뒤에 위치하여 멀티스펙트럴 이미지 센서(200)로부터 획득된 영상에 대응하는 채널 신호들의 세기에 관한 정보를 첫번째 레이어에 입력할 수 있다. 예를 들어, 제1 채널 어텐션 블록(CAB1)은 제1 컨볼루션 레이어(C1)의 뒤에 위치하여 제1 채널 어텐션 블록(CAB1)에 입력되는 영상에 대응하는 채널 신호들의 세기에 관한 정보를 제1 컨볼루션 레이어(C1)에 입력할 수 있다.

[0072] 다만, 이에 한정되는 것은 아니며, 채널 어텐션 블록(CAB)은 딥러닝 네트워크의 첫번째 레이어에 후속하는 중간 레이어(예: 제4 컨볼루션 레이어(C4), 제6 컨볼루션 레이어(C6))의 뒤에 위치하여 중간 레이어보다 앞서는 레이어(예: 제3 컨볼루션 레이어(C3), 제5 컨볼루션 레이어(C5))에서 출력되는 중간 영상에 대응하는 채널 신호들의 세기에 관한 정보를 중간 레이어에 입력할 수 있다. 예를 들어, 제2 채널 어텐션 블록(CAB2)은 제4 컨볼루션 레이어(C4)의 뒤에 위치하여 제2 채널 어텐션 블록(CAB2)에 입력되는 영상에 대응하는 채널 신호들의 세기에 관한 정보를 제4 컨볼루션 레이어(C4)에 입력할 수 있고, 제3 채널 어텐션 블록(CAB1)은 제6 컨볼루션 레이어(C6)의 뒤에 위치하여 제3 채널 어텐션 블록(CAB3)에 입력되는 영상에 대응하는 채널 신호들의 세기에 관한 정보를 제6 컨볼루션 레이어(C6)에 입력할 수 있다.

[0073] 본 개시의 실시 예에 따른 영상 획득 장치는 채널 어텐션 블록(CAB)을 포함하는 딥러닝 네트워크를 이용하여 멀티스펙트럴 이미지 센서(200)로부터 획득한 영상에 대응하는 채널들 간의 채널 상관관계에 대하여 미리 학습할 수 있으므로, 멀티스펙트럴 이미지 센서(200)로부터 획득한 영상에 대한 조명 정보를 보다 정확하게 추정할 수 있다.

[0074] 채널 어텐션 블록(CAB)을 포함하는 딥러닝 네트워크는 지도 학습(supervised learning)을 이용하여 학습될 수

있다. 예를 들어, 채널 어텐션 블록(CAB)을 포함하는 딥러닝 네트워크는 역전파 알고리즘(backpropagation algorithm)을 이용하여 딥러닝 네트워크의 파라미터를 최적화할 수 있다. 채널 어텐션 블록(CAB)을 포함하는 딥러닝 네트워크는 정답(Ground Truth) 조명 정보(L_{GT})와 추정 조명 정보(L_{est}) 간의 오차에 비례하여 딥러닝 네트워크의 처리 방향과 반대 방향으로 딥러닝 네트워크의 출력 및 은닉층의 가중치(weight)를 순서대로 갱신하는 방식으로 학습될 수 있다. 본 개시에서 정답 조명 정보(L_{GT})는 딥러닝 네트워크의 학습을 위한 훈련 영상에 대응하는 참값인 조명 정보를 의미할 수 있고, 추정 조명 정보(L_{est})는 훈련 영상을 딥러닝 네트워크에 입력하여 추정되는 조명 정보를 의미할 수 있다.

[0075] 채널 어텐션 블록(CAB)을 포함하는 딥러닝 네트워크는 각 손실(Angular Error, AE)을 포함하는 손실 함수(loss function)를 이용하여 학습될 수 있다. 각 손실은 벡터들 간의 각도를 측정한 것으로, 벡터의 크기 자체가 아닌 벡터들 간의 상대적인 방향(orientation)의 차이만을 고려할 수 있다. 딥러닝 네트워크가 추정하는 조명 정보는 이미지 센서로부터 획득되는 채널 신호들의 세기에 관한 정보이므로, 딥러닝 네트워크는 채널 별 채널 신호들의 세기의 상대적인 차이를 학습하는 것이 중요할 수 있다. 딥러닝 네트워크는 각 손실에 대한 역전파 알고리즘을 이용하여 학습됨에 따라 학습 효율이 향상될 수 있고, 이러한 딥러닝 네트워크가 구비된 영상 획득 장치는 멀티스펙트럴 이미지 센서(200)로부터 획득한 영상에 대한 조명 정보를 보다 정확하게 추정할 수 있다.

[0076] 채널 어텐션 블록(CAB)을 포함하는 딥러닝 네트워크는 하이퍼 각 손실(AE_{hyper})과 XYZ 각 손실(AE_{XYZ})을 이용하여 학습될 수 있다. 예를 들어, 딥러닝 네트워크는 하기의 수학식 3과 같이 하이퍼 각 손실(AE_{hyper})과 XYZ 각 손실(AE_{XYZ})을 일정한 비율로 가중 합(weighted sum)하여 손실함수로서 사용할 수 있다.

[0077] [수학식 3]

[0078] $$\ell = \alpha\ \mathrm{AE_{hyper}} + (1-\alpha)\ \mathrm{AE_{XYZ}}$$

[0079] 수학식 3에서 ℓ은 손실함수를 의미하고, α는 실험적으로 정해지는 값이다.

[0080] 하이퍼 각 손실(AE_{hyper})은 하이퍼 정답 조명 정보(L_{GT_hyper})와 하이퍼 추정 조명 정보(L_{est_hyper}) 사이의 각도를 의미하고, XYZ 각 손실(AE_{XYZ})은 XYZ 정답 조명 정보(L_{GT_hyper})와 XYZ 추정 조명 정보(L_{est_hyper}) 사이의 각도를 의미하는 것으로, 하이퍼 각 손실(AE_{hyper})과 XYZ 각 손실(AE_{XYZ})은 다음의 수학식 4와 수학식 5에 의해 각각 결정될 수 있다.

[0081] [수학식 4]

[0082] $$\mathrm{AE_{hyper}} = \mathrm{acos}\left(\frac{\mathrm{L_{GT_hyper}} \cdot \mathrm{L_{est_hyper}}}{\|\mathrm{L_{GT_hyper}}\|_2 \cdot \|\mathrm{L_{est_hyper}}\|_2}\right)$$

[0083] [수학식 5]

[0084] $$\mathrm{AE_{XYZ}} = \mathrm{acos}\left(\frac{\mathrm{L_{GT_XYZ}} \cdot \mathrm{L_{est_XYZ}}}{\|\mathrm{L_{GT_XYZ}}\|_2 \cdot \|\mathrm{L_{est_XYZ}}\|_2}\right)$$

[0085] 본 개시에서 하이퍼 각 손실(AE_{hyper})은 하이퍼 정답 조명 정보(L_{GT_hyper})와 하이퍼 추정 조명 정보(L_{est_hyper}) 사이의 각도를 의미할 수 있고, XYZ 각 손실(AE_{XYZ})은 XYZ 정답 조명 정보(L_{GT_hyper})와 XYZ 추정 조명 정보(L_{est_hyper}) 사이의 각도를 의미할 수 있다.

[0086] 하이퍼 정답 조명 정보(L_{GT_hyper})는 멀티스펙트럴 이미지 센서(200)와 관련된 정답 조명 정보(L_{GT})로서, 멀티스펙트럴 이미지 센서(200)의 각 채널에 대응하는 참값인 채널 신호의 크기를 벡터의 요소들로 구성하는 조명 벡터로 나타낼 수 있다. 하이퍼 추정 조명 정보(L_{est_hyper})는 멀티스펙트럴 이미지 센서(200)와 관련된 추정 조

명 정보(L_{est})로서, 멀티스펙트럴 이미지 센서(200)의 각 채널에 대응하는 추정되는 채널 신호의 크기를 벡터의 요소들로 구성하는 조명 벡터로 나타낼 수 있다.

[0087] XYZ 정답 조명 정보(L_{GT_hyper})는 하이퍼 정답 조명 정보(L_{GT_hyper})를 등색 함수(Color Matching Function)를 이용하여 XYZ 색공간으로 변환한 조명 정보를 의미할 수 있고, XYZ 추정 조명 정보(L_{est_hyper}) 는 하이퍼 추정 조명 정보(L_{est_hyper}) 를 등색 함수(Color Matching Function)를 이용하여 XYZ 색공간으로 변환한 조명 정보를 의미할 수 있다.

[0088] 딥러닝 네트워크는 하이퍼 각 손실(AE_{hyper})과 XYZ 각 손실(AE_{XYZ})을 모두 이용하여 학습되므로, XYZ 색공간으로 변환된 조명 정보보다 고차원인 멀티스펙트럴 이미지 센서(200)에 의해 측정되는 조명 정보에 대해서도 충분히 학습될 수 있다. 이에 따라, 본 개시의 실시 예에 따른 영상 획득 장치는 하이퍼 각 손실(AE_{hyper})과 XYZ 각 손실(AE_{XYZ})을 모두 이용하여 학습된 딥러닝 네트워크를 이용하여 조명 정보를 추정함으로써, 멀티스펙트럴 이미지 센서(200)로부터 획득한 영상에 대한 조명 정보를 보다 정확하게 추정할 수 있다.

[0089] 또한, 딥러닝 네트워크는 하이퍼 각 손실(AE_{hyper})과 XYZ 각 손실(AE_{XYZ})을 일정한 비율로 가중 합하여 손실함수로서 사용하여 학습되므로, 멀티스펙트럴 이미지 센서(200)에 의해 측정되는 조명 정보와 XYZ 색공간으로 변환된 조명 정보의 관계성에 대해서도 학습될 수 있다. 이에 따라, 본 개시의 실시 예에 따른 영상 획득 장치는 하이퍼 각 손실(AE_{hyper})과 XYZ 각 손실(AE_{XYZ})을 일정한 비율로 가중 합하여 손실함수로서 사용하여 학습된 딥러닝 네트워크를 이용하여 조명 정보를 추정함으로써, XYZ 색공간으로 변환된 조명 정보 또한 보다 정확하게 추정할 수 있다.

[0090] 영상 획득 장치(100)에 의해 영상이 촬영된 환경에는 하나 이상의 광원이 존재할 수 있다. 멀티스펙트럴 이미지 센서(200)에 의해 획득되는 영상이 서로 다른 제1 광원과 제2 광원이 존재하는 환경에서 획득된 혼합 조명 영상인 경우, 영상 획득 장치(100)는 채널 어텐션 블록(CAB)을 2개 이상 포함된 딥러닝 네트워크를 이용할 수 있다.

[0091] 프로세서(300)는 멀티스펙트럴 이미지 센서(200)에 의해 획득되는 혼합 조명 영상을 2개 이상의 영역들로 분할할 수 있다. 예를 들어, 프로세서(300)는 혼합 조명 영상을 제1 광원이 우세한 영역과 제2 광원이 우세한 영역으로 분할할 수 있다. 본 개시에서 제1 광원이 우세한 영역이라는 표현은 하나 이상의 광원이 존재하는 환경에서 제1 광원이 그 외 광원들보다 지배적으로 강한 세기를 가지는 영역을 의미할 수 있으며, 이하에서 해당 표현은 동일한 의미로 사용될 수 있다.

[0092] 프로세서(300)는 혼합 조명 영상 중 제1 광원이 우세한 영역에 대하여 딥러닝 네트워크를 이용하여 제1 광원에 대한 조명 정보를 추정하고, 혼합 조명 영상 중 제2 광원이 우세한 영역에 대하여 딥러닝 네트워크를 이용하여 제2 광원에 대한 조명 정보를 추정할 수 있다. 프로세서(300)는 제1 광원에 대한 조명 정보에 대응하는 제1 조명 벡터와 제2 광원에 대한 조명 정보에 대응하는 제2 조명 벡터를 일정한 비율로 가중 합할 수 있으며, 이 때 가중 계수는 0 이상이며 가중 계수들의 합은 1 이 되도록 설정될 수 있다.

[0093] 채널 어텐션 블록(CAB)을 포함하는 딥러닝 네트워크는 어느 하나의 컨볼루션 레이어(예: 제1 컨볼루션 레이어(C1))에 입력되는 영상을 어느 하나의 컨볼루션 레이어에 후속하는 다른 하나의 컨볼루션 레이어(예: 제5 컨볼루션 레이어(C5))에 입력하기 위한 스킵 커넥션(skip connection)을 포함할 수 있다. 예를 들어, 딥러닝 네트워크는 어느 하나의 컨볼루션 레이어(예: 제1 컨볼루션 레이어(C1))에 입력되는 영상을 픽셀 단위로 어느 하나의 컨볼루션 레이어에 후속하는 다른 하나의 컨볼루션 레이어(예: 제5 컨볼루션 레이어(C5))에 입력할 수 있다. 다른 예를 들어, 딥러닝 네트워크는 어느 하나의 컨볼루션 레이어(예: 제1 컨볼루션 레이어(C1))에 입력되는 영상을 채널 단위로 어느 하나의 컨볼루션 레이어에 후속하는 다른 하나의 컨볼루션 레이어(예: 제5 컨볼루션 레이어(C5))에 입력할 수 있다.

[0094] 딥러닝 네트워크의 입력단에 가깝게 위치한 컨볼루션 레이어에 입력되는 영상을 스킵 커넥션에 이용할수록, 딥러닝 네트워크는 딥러닝 네트워크에 입력되는 영상에 대한 조명 정보를 보다 정확하게 추정할 수 있다. 예를 들어, 제1 컨볼루션 레이어(C1)에 입력되는 영상은 딥러닝 네트워크에 입력되는 영상 자체일 수 있다. 반면, 제2 컨볼루션 레이어(C2)에 입력되는 영상은 딥러닝 네트워크에 입력되는 영상이 제1 컨볼루션 레이어(C1)에 입력된 후 제1 컨볼루션 레이어(C1)로부터 출력된 중간 영상일 수 있다. 이에 따라, 제1 컨볼루션 레이어(C1)에 입력되

는 영상은 제2 컨볼루션 레이어(C2)에 입력되는 영상보다 경계선, 색상 정보와 같은 미시적 정보를 더 많이 포함할 수 있다. 본 실시 예에 따른 영상 획득 장치(100)는 스킵 커넥션을 포함하는 딥러닝 네트워크를 이용함에 따라, 멀티스펙트럴 이미지 센서(200)에 의해 획득되는 영상에 대한 조명 정보를 보다 정확하게 추정할 수 있다.

[0095] 이하에서는, 채널 어텐션 블록의 구조에 대하여 보다 구체적으로 설명하도록 한다.

[0096] **도 8a는 일 실시 예에 따른 도 7의 채널 어텐션 블록을 설명하기 위한 도면이다.**

[0097] 도 8a를 참조하면, 채널 어텐션 블록(CAB)은 컨볼루션 레이어(conv) 뒤에 위치하고, 컨볼루션 레이어(conv)에서 출력된 출력 피쳐 맵은 채널 어텐션 블록(CAB)에 입력될 수 있다.

[0098] 채널 어텐션 블록(CAB)은 풀링 레이어 및 제1 완전 연결 층을 포함할 수 있다.

[0099] 컨볼루션 레이어(conv)로부터 출력된 컨볼루션 연산의 수행 결과인 출력 피쳐 맵이 풀링 레이어에 입력될 수 있고, 이에 따라, 딥러닝 네트워크는 출력 피쳐 맵의 전역 또는 로컬에 대하여 맥스 풀링 또는 평균 풀링을 수행함으로써 풀링이 수행된 대상 영역에 대한 최대값 또는 평균값을 출력할 수 있다.

[0100] 딥러닝 네트워크는 제1 완전 연결 층을 이용하여 맥스 풀링 레이어로부터 출력된 결과에 대응하는 채널들의 개수가 채널 어텐션 블록(CAB)을 포함하는 딥러닝 네트워크에 입력되는 영상에 대응하는 채널들의 개수와 동일하도록 조정할 수 있다.

[0101] 프로세서는 멀티스펙트럴 이미지 센서(200)로부터 획득된 영상에 전역 맥스 풀링 또는 전역 평균 풀링을 수행하여 멀티스펙트럴 이미지 센서(200)로부터 획득된 영상의 전역에 대한 최대값 또는 평균값을 벡터의 요소로 구성하는 제1 채널 벡터를 생성할 수 있다. 딥러닝 네트워크는 생성된 제1 채널 벡터를 채널 어텐션 블록(CAB)의 제1 완전 연결 층의 출력 결과에 더하여 컨볼루션 레이어(conv)로부터 출력된 컨볼루션 연산의 수행 결과인 출력 피쳐 맵과 컨볼루션 연산이 수행되도록 할 수 있다.

[0102] 생성되는 제1 채널 벡터의 차원은 멀티스펙트럴 이미지 센서(200)로부터 획득된 영상에 대응하는 채널의 개수와 동일할 수 있으나, 이에 제한되지 않으며, 멀티스펙트럴 이미지 센서(200)로부터 획득된 영상에 대응하는 채널의 개수보다 많도록 조정될 수 있다. 딥러닝 네트워크는 완전 연결 층을 이용하여 제1 채널 벡터의 차원을 조정할 수 있다. 딥러닝 네트워크가 완전 연결 층을 이용하여 벡터의 차원을 조정하는 방법은 도 6에서 상술한 컨볼루션 블록(CB)을 포함하는 딥러닝 네트워크가 완전 연결 층(FC)을 이용하여 추정된 조명 정보에 대응되는 출력 벡터의 차원을 조정하는 방법이 그대로 적용될 수 있다.

[0103] 딥러닝 네트워크가 멀티스펙트럴 이미지 센서(200)로부터 획득된 영상에 전역 맥스 풀링 또는 전역 평균 풀링을 수행하여 생성한 제1 채널 벡터를 컨볼루션 레이어로부터 출력되는 출력 피쳐 맵에 입력함으로써, 딥러닝 네트워크는 딥러닝 네트워크에 입력되는 영상에 대응하는 채널들 간 채널 상관관계에 따라 채널 별로 획득되는 신호들 각각에 상이한 가중치를 부여할 수 있다. 이에 따라, 본 개시의 실시 예에 따른 영상 획득 장치는 멀티스펙트럴 이미지 센서(200)로부터 획득한 영상에 대한 조명 정보를 보다 정확하게 추정할 수 있다.

[0104] 특히, 딥러닝 네트워크가 멀티스펙트럴 이미지 센서(200)로부터 획득된 영상에 전역 평균 풀링을 수행하여 생성한 제1 채널 벡터를 컨볼루션 레이어(conv)로부터 출력되는 출력 피쳐 맵에 입력함으로써, 본 개시의 실시 예에 따른 영상 획득 장치는 파장 대역이 좁은 광원이 존재하는 환경에서 촬영된 영상에 대해서도 보다 정확하게 조명 정보를 추정할 수 있다.

[0105] 다른 예를 들어, 채널 어텐션 블록(CAB)은 렐루 함수, 제2 완전 연결 층 및 시그모이드 함수를 추가로 더 포함할 수도 있다. 딥러닝 네트워크는 채널 어텐션 블록(CAB)의 제1 완전 연결 층의 출력 결과에 생성된 제1 채널 벡터를 더한 출력 결과에 렐루 함수, 제2 완전 연결 층 및 시그모이드 함수를 순차적으로 적용할 수 있고, 그로 인해 생성된 출력 결과가 컨볼루션 레이어(conv)로부터 출력된 컨볼루션 연산의 수행 결과인 출력 피쳐 맵과 컨볼루션 연산이 수행되도록 할 수 있다.

[0106] **도 8b는 다른 실시 예에 따른 도 7의 채널 어텐션 블록을 설명하기 위한 도면이다.** 도 8b의 채널 어텐션 블록는 도 8a의 채널 어텐션 블록과 제1 완전 연결 층의 출력 결과에 더해지는 채널 벡터만이 상이할 수 있으며, 이하에서 중복되는 설명은 생략하도록 한다.

[0107] 도 8b에 도시된 컨볼루션 레이어(conv)는 채널 어텐션 블록(CAB)으로부터 출력되는 결과가 입력되는 컨볼루션 레이어보다 앞서 위치한 컨볼루션 레이어로서, 중간 컨볼루션 레이어(예: 도 6의 제4 컨볼루션 레이어(C4), 제6

컨볼루션 레이어(C6))들 중 하나일 수 있다. 이하에서는 도 8b에 도시된 컨볼루션 레이어를 중간 컨볼루션 레이어로 지칭하기로 한다. 프로세서는 중간 컨볼루션 레이어로부터 출력되는 중간 영상에 전역 맥스 풀링 또는 전역 평균 풀링을 수행하여 중간 영상의 전역 또는 로컬에 대한 최대값 또는 평균값을 벡터의 요소로 구성하는 제2 채널 벡터를 생성할 수 있다.

[0108] 예를 들어, 딥러닝 네트워크는 생성된 제2 채널 벡터를 채널 어텐션 블록(CAB)의 제1 완전 연결 층의 출력 결과에 더하여 중간 컨볼루션 레이어로부터 출력된 컨볼루션 연산의 수행 결과인 출력 피처 맵과 컨볼루션 연산이 수행되도록 할 수 있다.

[0109] 생성되는 제2 채널 벡터의 차원은 중간 컨볼루션 레이어로부터 출력된 출력 피처 맵에 대응하는 채널의 개수와 동일할 수 있으나, 이에 제한되지 않으며, 중간 컨볼루션 레이어로부터 출력된 출력 피처 맵에 대응하는 채널의 개수보다 많도록 조정될 수 있다. 딥러닝 네트워크는 완전 연결 층을 이용하여 제2 채널 벡터의 차원을 조정할 수 있다. 딥러닝 네트워크가 완전 연결 층을 이용하여 벡터의 차원을 조정하는 방법은 도 6에서 상술한 컨볼루션 블록(CB)을 포함하는 딥러닝 네트워크가 완전 연결 층(FC)을 이용하여 추정된 조명 정보에 대응되는 출력 벡터의 차원을 조정하는 방법이 그대로 적용될 수 있다.

[0110] 딥러닝 네트워크가 중간 컨볼루션 레이어로부터 출력된 출력 피처 맵에 전역 맥스 풀링, 전역 평균 풀링, 로컬 맥스 풀링 또는 로컬 평균 풀링을 수행하여 생성한 제2 채널 벡터를 중간 컨볼루션 레이어로부터 출력되는 출력 피처 맵에 입력함으로써, 딥러닝 네트워크는 딥러닝 네트워크에 입력되는 영상에 대응하는 채널들 간 채널 상관관계에 따라 채널 별로 획득되는 신호들 각각에 상이한 가중치를 부여할 수 있다. 이에 따라, 본 개시의 실시예에 따른 영상 획득 장치는 멀티스펙트럴 이미지 센서(200)로부터 획득한 영상에 대한 조명 정보를 보다 정확하게 추정할 수 있다.

[0111] 특히, 딥러닝 네트워크가 중간 컨볼루션 레이어로부터 출력된 출력 피처 맵에 전역 평균 풀링 또는 로컬 평균 풀링을 수행하여 생성한 제2 채널 벡터를 중간 컨볼루션 레이어로부터 출력되는 출력 피처 맵에 입력함으로써, 본 개시의 실시 예에 따른 영상 획득 장치는 파장 대역이 좁은 광원이 존재하는 환경에서 촬영된 영상에 대해서도 보다 정확하게 조명 정보를 추정할 수 있다.

[0112] **도 9a는 일 실시 예에 따른 트랜스포머의 구조를 설명하기 위한 도면이다.**

[0113] 도 9a를 참조하면, 일 실시 예에 따른 트랜스포머(900)는 인코더(901) 및 디코더(910)를 포함할 수 있다. 인코더(901) 및/또는 디코더(910)는 일련의(a series of) 하부 모델(sub-model)들을 포함하는 하나 이상의 모듈들로 구현될 수 있다.

[0114] 인코더(901)는 N이 정수일 때 N 개의 층(902)들을 포함할 수 있다. 각 층(902)은 하부 모델로 지칭될 수 있고 , 멀티헤드 어텐션(multi-head attention) 레이어(903)와 피드포워드(feed-forward) 레이어(904)를 포함할 수 있다. 잔차 연결(residual connection)(905)이 멀티헤드 어텐션(multi-head attention) 레이어(903)와 피드포워드(feed-forward) 서브 레이어(904)의 둘레에 각각 사용될 수 있으며, 계층 정규화(layer normalization)(906)가 후속된다. 디코더(912)의 출력은 인풋 임베딩(907)에 임베딩될 수 있고, 위치 인코딩(positional encoding)(908)의 입력들의 임베딩된 표현에 가산될 수 있다.

[0115] 디코더(910)도 N 개의 층(912)들을 포함할 수 있다. 각 층(912)은 하부 모델로 지칭될 수 있고, 마스킹된(masked) 멀티헤드 어텐션 레이어(911), 멀티헤드 어텐션(multi-head attention) 레이어(913)와 피드포워드(feed-forward) 레이어(914)를 포함할 수 있다. 잔차 연결(residual connection)(915)이 멀티헤드 어텐션(multi-head attention) 서브 레이어(913)와 피드포워드(feed-forward) 레이어(914)의 둘레에 각각 사용될 수 있으며, 계층 정규화(layer normalization)(916)가 후속된다. 마스킹된(masked) 멀티헤드 어텐션 레이어(911)는 후속 위치들에 어텐딩(attending to)하는 것을 방지할 수 있다. 디코더(912)의 출력은 아웃풋 임베딩(917)에 임베딩될 수 있고, 위치 인코딩(918)에서 한 위치만큼 위치적으로 편이(positionally offset)되어, 위치(i)에 대한 예측들은 i 미만인 위치들에서 기지의(known) 출력에만 좌우될 수 있다.

[0116] 디코더(910)의 출력은 선형 분류기(linear classifier) 레이어(919)에 대한 입력이 될 수 있고, 이어서 소프트맥스 레이어(920)에 입력되어, 출력 시퀀스의 다음 토큰을 예측하는데 사용될 수 있는 확률들을 출력할 수 있다. 트랜스포머(900)는 고정(static) 가중치(weight)들을 사용하지 않는 몇 개의 행렬 곱셈 연산들을 포함할 수 있다.

[0117] **도 9b는 도 9a의 멀티헤드 어텐션 레이어의 구조를 설명하기 위한 도면이고, 도 9c는 도 9b의 멀티헤드 어텐션**

레이어의 연산 결과를 나타낸 도면이다.

[0118] 도 9b를 참조하면, 도 9b에 도시된 멀티헤드 어텐션 레이어는 도 9a에 도시된 인코더(901)의 멀티헤드 어텐션 레이어(903)와 디코더(910)의 멀티헤드 어텐션 레이어(913)일 수 있다. 멀티헤드 어텐션 레이어는 h가 정수일 때, 행렬 V, K, 및 Q의 h 개의 선형 투영(linear projection)(930)들과 h개의 스케일드(scaled) 도트곱(dot-product) 어텐션 레이어(931)들과 연결(concatenation) 레이어(932) 및 선형 분류기 계층(933)을 포함할 수 있다.

[0119] 인코더(901)의 멀티헤드 어텐션 레이어(903)와 디코더(910)의 멀티헤드 어텐션 레이어(913)는 V, K, 및 Q의 선형 투영들과 병렬화(parallelize)될 수 있다. V는 다시 시퀀스 내의 모든 단어들의 벡터 표현들의 값의 행렬(또는, 피쳐 맵(feature map))이 될 수 있고, K는 시퀀스 내의 모든 단어들의 벡터 표현들인 모든 키(key)들의 행렬(또는, 피쳐 맵)이 될 수 있고, Q는 시퀀스 내의 한 단어의 벡터 표현인 쿼리(querry)를 포함하는 행렬(또는, 피쳐 맵)이 될 수 있다. 병렬화는 트랜스포머(900)가 V, K, 및 Q의 다른 표현들로부터 유용하게 학습할 수 있도록 할 수 있다. 선형 표현들은 V, K, 및 Q에 트랜스포머(900)가 훈련을 통해 학습한 가중치 행렬(W)을 곱하여 형성될 수 있다. 행렬 V, K, 및 Q들은 행렬 V, K, 및 Q들이 인코더(901)에 있는지 디코더(910)에 있는지 또는 인코더(901)와 디코더(910) 사이에 있는지에 따라 트랜스포머(900) 내의 어텐션 모듈들의 위치들이 각각 다를 수 있으므로, 인코더 입력 시퀀스의 전체 또는 부분이 어텐딩될 수 있다. 인코더(901)와 디코더(910)를 연결하는 멀티헤드 어텐션 모듈은 주어진 위치까지, 인코더 입력 시퀀스를 디코더 입력 시퀀스와 함께 고려할 수 있다.

[0120] 인코더(901)의 점별 피드포워드 레이어(904)와 디코더(910)의 점별 피드포워드 레이어(914)는 각각 인코더(901)의 멀티헤드 어텐션 레이어(903)의 후방과 디코더(910)의 멀티헤드 어텐션 레이어(913)의 후방에 배치될 수 있다. 인코더(901)의 피드포워드 레이어(904)와 디코더(910)의 피드포워드 레이어(914)는 각각 주어진 시퀀스로부터의 각 요소로부터의 요소에 대해 분리된 동일한 선형 변환을 제공할 수 있도록 각 위치에 대해 동일한 매개변수들을 포함할 수 있다

[0121] **도 10은 또 다른 실시 예에 따른 조명 정보를 추정하기 위한 딥러닝 네트워크를 설명하기 위한 도면이다.** 도 10은 도 7의 딥러닝 네트워크에서 딥러닝 네트워크에 셀프 어텐션 레이어(SA) 및 완전 연결층(FC)이 더 추가된 점만이 상이할 수 있으며, 이하에서 중복되는 설명은 생략하도록 한다.

[0122] 본 개시의 셀프 어텐션 레이어(SA)은 딥러닝 네트워크에 입력되는 영상에 대응하는 채널들 간의 채널 상관관계를 학습하기 위하기 위하여 신경망으로, 도 9a에 도시된 트랜스포머(900)의 일부 레이어를 포함하는 신경망을 의미할 수 있다.

[0123] 도 10을 참조하면, 셀프 어텐션 레이어(SA)를 포함하는 딥러닝 네트워크는 멀티스펙트럴 이미지 센서(200)로부터 획득된 영상을 입력 받아 획득된 영상에 대한 조명 정보를 추정할 수 있다. 프로세서(300)는 멀티스펙트럴 이미지 센서(200)에 의해 획득된 영상을 제1 컨볼루션 레이어(C1) 및 셀프 어텐션 레이어(SA) 각각에 입력할 수 있다. 즉, 멀티스펙트럴 이미지 센서(200)에 의해 획득된 영상은 딥러닝 네트워크에 입력되어 제1 컨볼루션 레이어(C1) 및 셀프 어텐션 레이어(SA)에 입력될 수 있다.

[0124] 프로세서(300)는 멀티스펙트럴 이미지 센서(200)에 의해 획득된 영상을 분할하여 셀프 어텐션 레이어(SA)에 입력할 수 있다. 셀프 어텐션 레이어(SA)에 입력되는 영상의 크기가 W x H인 경우, 프로세서(300)는 영상을 N x N의 크기를 갖는 영상으로 분할하거나 영상을 채널 별로 분할할 수 있고, 프로세서(300)는 분할된 영상을 셀프 어텐션 레이어(SA)에 입력할 수 있다. 채널 별로 분할된 영상이 셀프 어텐션 레이어(SA)에 입력되는 경우, 프로세서(300)는 채널들을 통해 획득된 영상들 전부를 셀프 어텐션 레이어(SA)에 입력하거나 일부 채널들을 통해 입력된 영상들 만을 셀프 어텐션 레이어(SA)에 입력할 수 있다. 프로세서(900)는 채널들을 통해 획득된 영상들 각각을 개별적으로 셀프 어텐션 레이어(SA)에 입력하거나 채널들을 통해 획득된 영상들을 그룹핑한 후 그룹핑된 영상들을 셀프 어텐션 레이어(SA)에 입력할 수도 있다. 예를 들어, 프로세서(900)는 딥러닝 네트워크에 입력된 영상에 대응되는 채널의 개수가 C_i인 경우, N개의 채널을 통해 획득되는 영상들을 채널 별로 적층하고 그룹핑하여 C_i/N 개의 채널을 통해 획득되는 영상들만을 셀프 어텐션 레이어(SA)에 입력할 수 있다. 프로세서(900)는 일부 채널들을 통해 획득된 영상들을 셀프 어텐션 레이어(SA)에 입력하지 않거나 셀프 어텐션 레이어(SA)에 트랜스포머(900)의 인코더(901)가 포함된 경우 인코더(901)에서 일부 채널들을 통해 획득된 영상들을 마스킹하여, 이후 셀프 어텐션을 수행하는 과정에서는 사용되지 않도록 할 수 있다. 프로세서(300)는 획득된 채널 신호들을 합성 또는 보간하여 획득할 채널 신호들의 개수를 조정함으로써 셀프 어텐션 레이어(SA)에 입력되는 영상에 대

응되는 채널의 개수를 조정할 수 있다.일 예시에 따른 셀프 어텐션 레이어(SA)는 도 9a의 트랜스포머(900)를 포함할 수 있다. 트랜스포머(900)는 입력된 영상을 인코더(901)의 멀티헤드 어텐션 레이어(930)에 입력하여 채널별로 키, 쿼리, 값을 연산할 수 있다. 트랜스포머(900)는 연산된 키, 쿼리, 값을 스케일드 도트곱 어텐션 레이어(931)에 입력함에 따라, 스코어 함수를 획득할 수 있다. 본 개시에서 스코어 함수라는 표현은 채널들 각각이 조명 정보를 획득하고자 하는 채널에 대하여 미치는 영향의 정도를 의미할 수 있으며, 이하에서 해당 표현은 동일한 의미로 사용될 수 있다. 스코어 함수는 다음의 수학식 6에 의해 결정될 수 있다.

[0125] [수학식 6]

[0126] $$\text{score function}(q,k) = \frac{q * k}{\sqrt{d^k}}$$

[0127] 딥러닝 네트워크는 특정 채널을 통해 획득되는 영상에 대응되는 값에 스코어 함수를 곱하거나 특정 채널 자체에 스코어 함수를 곱한 후 특정 채널을 통해 영상을 획득함으로써, 멀티스펙트럴 이미지 센서(200)에 의해 획득되는 영상에 대한 채널 상관관계를 학습할 수 있다. 본 실시 예에 따른 영상 획득 장치(100)는 채널 상관관계를 미리 학습한 딥러닝 네트워크를 이용하여 멀티스펙트럴 이미지 센서(200)에 의해 획득되는 영상에 대한 조명 정보를 보다 정확하게 추정할 수 있다.다른 예시에 따른 셀프 어텐션 레이어(SA)는 도 9a의 트랜스포머(900)의 인코더(901)만 포함할 수 있다. 즉, 셀프 어텐션 레이어(SA)는 인코더(901)의 멀티헤드 어텐션 레이어(903)와 피드포워드 레이어(904)를 포함할 수 있다. 다른 예시에 따른 셀프 어텐션 레이어(SA)는 입력된 영상을 인코더(901)의 멀티헤드 어텐션 레이어(930)에 입력하여 스코어 함수를 획득할 수 있다. 딥러닝 네트워크는 특정 채널을 통해 획득되는 영상에 대응되는 값에 스코어 함수를 곱하거나 특정 채널 자체에 스코어 함수를 곱한 후 특정 채널을 통해 영상을 획득함으로써, 멀티스펙트럴 이미지 센서(200)에 의해 획득되는 영상에 대한 채널 상관관계를 학습할 수 있다. 본 실시 예에 따른 영상 획득 장치(100)는 채널 상관관계를 미리 학습한 딥러닝 네트워크를 이용하여 멀티스펙트럴 이미지 센서(200)에 의해 획득되는 영상에 대한 조명 정보를 보다 정확하게 추정할 수 있다.또 다른 예시에 따른 셀프 어텐션 레이어(SA)는 전역 평균 풀링 레이어를 포함할 수 있다. 셀프 어텐션 레이어(SA)는 딥러닝네트워크에 입력되는 영상을 전역 평균 풀링 레이어에 입력한 후, 평균 풀링 레이어로부터 출력되는 결과에 도 9a의 트랜스포머(900)에 사용되는 스코어 함수를 곱함으로써, 멀티스펙트럴 이미지 센서(200)에 의해 획득되는 영상에 대한 채널 상관관계를 학습할 수 있다. 본 실시 예에 따른 영상 획득 장치(100)는 채널 상관관계를 미리 학습한 딥러닝 네트워크를 이용하여 멀티스펙트럴 이미지 센서(200)에 의해 획득되는 영상에 대한 조명 정보를 보다 정확하게 추정할 수 있다.

[0128] 셀프 어텐션 레이어(SA)로부터 출력되는 결과는 완전 연결 층들(FC1, FC2, FC3) 각각에 입력된 후 채널 어텐션 블록(CAB)을 포함하는 딥러닝 네트워크(예: 도 7의 딥러닝 네트워크)의 채널 어텐션 블록들(CAB1, CAB2, CAB3) 각각에 입력될 수 있다. 셀프 어텐션 레이어(SA)를 포함하는 딥러닝 네트워크는 셀프 어텐션 레이어(SA) 및 채널 어텐션 블록들(CAB1, CAB2, CAB3)을 이용하여 보다 정확하게 채널 상관관계를 학습할 수 있으며, 이에 따라, 본 실시 예에 따른 영상 획득 장치(100)는 멀티스펙트럴 이미지 센서(200)에 의해 획득되는 영상에 대한 조명 정보를 보다 정확하게 추정할 수 있다.

[0129] **도 11a는 3D 컨볼루션 레이어를 포함하는 딥러닝 네트워크를 설명하기 위한 도면이다. 도 11b는 도 11a의 스펙트럴 채널 어텐션 블록을 설명하기 위한 도면이고, 도 11c는 도 10a의 스펙트럴 셀프 어텐션 블록을 설명하기 위한 도면이다.**

[0130] 도 11a를 참조하면, 딥러닝 네트워크는 하나 이상의 공간 어텐션 블록(SAB), 3D 컨볼루션 레이어(3D conv), 3D 풀링 레이어 및 완전 연결층을 포함할 수 있다. 본 개시의 공간 어텐션 블록(SAB)은 3D 컨볼루션을 수행하기 위한 3D 컨볼루션 레이어(3D conv) 및 3D 컨볼루션 레이어(3D conv)의 뒤에 위치하는 스펙트럴 채널 어텐션 블록(SCAB)를 포함하는 신경망을 의미할 수 있다. 본 개시에서 3D(3-dimension) 컨볼루션은 3개의 파라미터에 대한 컨볼루션 연산을 수행하는 것을 의미하며, 본 개시의 3D 컨볼루션 레이어는 2D 영상의 높이와 폭 및 스펙트럼 채널에 관한 정보에 기초하여 컨볼루션 연산을 수행할 수 있다. 이 때, 스펙트럼 채널에 관한 정보는 채널의 크기 내지 인덱스를 포함할 수 있으나, 이에 한정되는 것은 아니다. 3D 컨볼루션 레이어(3D conv)을 포함하는 딥러닝 네트워크는 채널 상관관계를 보다 정확하게 추정할 수 있다.

[0131] 도 11b의 스펙트럴 채널 어텐션 블록(SCAB)을 포함하는 딥러닝 네트워크는 도 7의 채널 어텐션 블록(CAB)을 포함하는 딥러닝 네트워크와, 도 11b의 스펙트럴 채널 어텐션 블록(SCAB)과 도 7의 채널 어텐션 블록(CAB)에 입력되는 영상이 다르다는 점만이 상이할 수 있으며, 이하에서 중복되는 설명은 생략하기로 한다. 즉, 도 11b의 스펙트럴 채널 어텐션 블록(SCAB)는 3D 컨볼루션 레이어(3D conv)의 뒤에 위치하는 반면, 도 7의 채널 어텐션 블

록(CAB)은 컨볼루션 레이어(conv)의 뒤에 위치하는 점이 서로 상이할 수 있다.

[0132] 3D 컨볼루션 레이어(3D conv)와 스펙트럴 채널 어텐션 블록(SCAB)을 포함하는 딥러닝 네트워크는 멀티스펙트럴 이미지 센서(200)에 의해 획득된 영상에 대응하는 채널들 간 채널 상관관계에 대해 미리 학습되므로, 본 실시 예에 따른 영상 획득 장치(100)는 멀티스펙트럴 이미지 센서(200)로부터 획득한 영상에 대한 조명 정보를 보다 정확하게 추정할 수 있다.

[0133] 도 11a의 공간 어텐션 블록(SAB)은 스펙트럴 셀프 어텐션 블록(SSAB)을 더 포함할 수 있고, 스펙트럴 셀프 어텐션 블록(SSAB)은 3D 컨볼루션 레이어(3D conv) 및 3D 컨볼루션 레이어(3D conv)의 뒤에 위치하는 스펙트럴 셀프 어텐션 레이어(SSA)를 포함할 수 있다. 3D 컨볼루션 레이어(3D conv)로부터 출력되는 출력 피쳐 큐브(cube)는 스펙트럴 셀프 어텐션 레이어(SSA)에 입력되어 쿼리(Q), 키(K), 값(V)을 연산할 수 있다. 스펙트럴 셀프 어텐션 레이어(SSA)는 연산된 쿼리(Q)와 키(K)를 도트곱하고 소프트 맥스(SoftMax) 함수를 적용하여 확률에 관한 함수로 변환된 스코어 함수를 출력할 수 있다. 출력된 스코어 함수는 연산된 값(V)와 곱해져 후속하는 3D 컨볼루션 레이어(3D conv)에 입력되는 입력 피쳐 큐브를 획득할 수 있으므로, 멀티스펙트럴 이미지 센서(200)에 의해 획득되는 영상에 대한 채널 상관관계를 학습할 수 있다. 도 11a의 딥러닝 네트워크는 멀티스펙트럴 이미지 센서(200)에 의해 획득된 영상이 연속하는 2개의 공간 어텐션 블록(SAB), 3D 컨볼루션 레이어(3D conv), 3D 풀링 레이어 및 완전 연결층을 거침에 따라 멀티스펙트럴 이미지 센서(200)에 의해 획득되는 영상에 대한 조명 정보를 추정할 수 있다. 본 실시 예에 따른 영상 획득 장치(100)는 채널 상관관계를 미리 학습한 딥러닝 네트워크를 이용하여 멀티스펙트럴 이미지 센서(200)에 의해 획득되는 영상에 대한 조명 정보를 보다 정확하게 추정할 수 있다.

[0134] **도 12a는 일 실시 예에 따른 추가 딥러닝 네트워크를 설명하기 위한 도면이고, 도 12b는 다른 실시 예에 따른 추가 딥러닝 네트워크를 설명하기 위한 도면이다.** 도 12a 및 도 12b을 참조하면, 도 12a 및 도 12b에 도시된 조명 추정 모델은 도 6에 도시된 컨볼루션 블록(CB)을 포함하는 딥러닝 네트워크를 나타낸 것이다. 도 6을 참조하여 설명한 컨볼루션 블록(CB)을 포함하는 딥러닝 네트워크에 관한 설명은 도 12a 및 도 12b의 조명 추정 모델에도 동일하게 적용될 수 있으며, 도 12a 및 도 12b의 추가 딥러닝 네트워크는 조명 추정 모델을 이용하여 멀티스펙트럴 이미지 센서(200)에 의해 획득된 영상에 대한 조명 정보를 추정할 수 있다.

[0135] 도 12a의 추가 딥러닝 네트워크는 추정된 조명 정보에 대응하는 조명 벡터의 차원과 동일한 개수의 노드를 가지는 완전 연결 층을 더 포함하여 추가 딥러닝 네트워크로부터 출력되는 출력 결과가 하이퍼 정답 조명 정보에 근접할 수 있도록 학습된 네트워크일 수 있다.

[0136] 도 12b의 추가 딥러닝 네트워크는 3개의 노드를 가지는 완전 연결 층을 더 포함하여 추가 딥러닝 네트워크로부터 출력되는 출력 결과가 XYZ 정답 조명 정보에 근접할 수 있도록 학습된 네트워크일 수 있다.

[0137] 프로세서는 도 12a의 추가 딥러닝 네트워크로부터 출력되는 출력 결과를 등색 함수를 이용하여 XYZ 색공간으로 변환할 수 있고, 도 12b의 추가 딥러닝 네트워크의 출력 결과의 각 요소들을 변환된 도 12a의 추가 딥러닝 네트워크의 출력 결과의 각 요소들로 나눔으로써 멀티스펙트럴 이미지 센서(200)에 의해 획득된 영상에 대한 자동 화이트 밸런스를 수행할 수 있다.

[0138] 본 실시 예의 영상 획득 장치는 도 12a의 추가 딥러닝 네트워크와 도 12b의 추가 딥러닝 네트워크를 이용함으로써, 컨볼루션 블록(CB)을 포함하는 딥러닝 네트워크로부터 추정되는 조명 정보보다 정답 조명 정보에 근접한 조명 정보를 출력할 수 있다. 이에 따라, 본 실시 예의 영상 획득 장치는 멀티스펙트럴 이미지 센서(200)에 의해 획득된 영상에 대하여 보다 정확한 자동 화이트 밸런스를 수행할 수 있다.

[0139] **도 13은 일 실시 예에 따른 영상을 획득하는 방법을 설명하기 위한 순서도이다.**

[0140] 도 13을 참조하면, 영상을 획득하는 방법은 도 1 내지 도 12를 참조하여 설명한 영상 획득 장치(100)에서 처리되는 단계들로 구성된다. 이에 따라, 도 1 내지 도 12를 참조하여 설명한 영상 획득 장치(100)에 관하여 이상에서 설명된 내용은 도 13의 영상을 획득하는 방법에도 적용될 수 있다.

[0141] 일 실시 예에 따른 영상을 획득하는 방법은 단계 1110에서 멀티스펙트럴 이미지 센서(200)를 이용하여 4개 이상의 채널들을 통해 영상을 획득함에 따라 개시될 수 있다

[0142] 영상 획득 장치(100)는 멀티스펙트럴 이미지 센서(200)의 각 채널에 대응하는 출력 신호인 채널 신호들을 획득할 수 있다. 영상 획득 장치(100)는 멀티스펙트럴 이미지 센서(200)에 물리적으로 구비된 소정 개수의 채널들 중 적어도 일부를 선택하고, 선택된 채널들로부터 채널 신호들을 획득할 수 있다. 예를 들어, 영상 획득 장치

(100)는 멀티스펙트럴 이미지 센서(200)에 물리적으로 구비된 소정 개수의 채널들 모두로부터 채널 신호들을 획득할 수 있다. 또한, 영상 획득 장치(100)는 멀티스펙트럴 이미지 센서(200)에 물리적으로 구비된 소정 개수의 채널들 중 일부만을 선택하여 채널 신호들을 획득할 수 있다.

[0143] 영상 획득 장치(100)는 멀티스펙트럴 이미지 센서(200)에 물리적으로 구비된 소정 개수의 채널들로부터 획득되는 채널 신호들을 합성 또는 보간함으로써, 소정 개수보다 증가되거나 감소된 개수의 채널 신호들을 획득할 수도 있다. 예를 들어, 영상 획득 장치(100)는 멀티스펙트럴 이미지 센서(200)의 4개 이상의 채널들을 통해 멀티스펙트럴 영상을 획득하거나 9개 이상의 채널들을 통해 하이퍼스펙트럴 영상을 획득할 수 있다. 영상 획득 장치(100)는 멀티스펙트럴 이미지 센서(200)에 물리적으로 구비된 채널들 전부 또는 일부를 선택하여 채널 신호들을 획득할 수 있으며, 영상 획득 장치(100)는 획득된 채널 신호들을 합성 또는 보간함으로써 획득할 채널 신호들의 개수를 조정할 수 있다.

[0144] 단계 1120에서 영상 획득 장치(100)는 채널들 간 채널 상관관계를 미리 학습한 딥러닝 네트워크를 이용하여 조명 정보를 추정할 수 있다. 영상 획득 장치(100)는 멀티스펙트럴 이미지 센서(200)로부터 획득된 영상을 채널들 간 채널 상관관계를 미리 학습한 딥러닝 네트워크에 입력할 수 있다.

[0145] 영상 획득 장치(100)의 딥러닝 네트워크는 지도 학습 방식(supervised learning method)을 이용하여 미리 학습될 수 있다. 예를 들어, 딥러닝 네트워크는 딥러닝 네트워크에 입력된 영상들 각각에 대하여 추정된 조명 정보, 정답(Ground Truth) 조명 정보 및 손실 함수(loss function) 등을 이용하여 학습될 수 있다. 영상 획득 장치(100)의 딥러닝 네트워크는 획득된 영상에 대응하는 채널들 간의 채널 상관관계(channel correlation)에 대해 미리 학습될 수 있다.

[0146] 채널 상관관계에 대해 학습된 딥러닝 네트워크는 딥러닝 네트워크에 입력되는 영상에 대응하는 채널들 간 채널 상관관계에 따라 채널 별로 획득되는 신호들 각각에 상이한 가중치를 부여함으로써 보다 정확한 조명 정보를 추정할 수 있다.

[0147] 예를 들어, 채널 상관관계에 대해 학습된 딥러닝 네트워크는 채널 어텐션 블록(CAB)을 포함하는 딥러닝 네트워크일 수 있다. 영상 획득 장치(100)는 채널 어텐션 블록(CAB)을 포함하는 딥러닝 네트워크를 이용하여 멀티스펙트럴 이미지 센서(200)로부터 획득한 영상에 대응하는 채널들 간의 채널 상관관계에 대하여 미리 학습할 수 있으므로, 멀티스펙트럴 이미지 센서(200)로부터 획득한 영상에 대한 조명 정보를 보다 정확하게 추정할 수 있다.

[0148] 채널 어텐션 블록(CAB)을 포함하는 딥러닝 네트워크는 지도 학습(supervised learning)을 이용하여 학습될 수 있다. 예를 들어, 채널 어텐션 블록(CAB)을 포함하는 딥러닝 네트워크는 역전파 알고리즘(backpropagation algorithm)을 이용하여 딥러닝 네트워크의 파라미터를 최적화할 수 있다. 채널 어텐션 블록(CAB)을 포함하는 딥러닝 네트워크는 각 손실(Angular Error, AE)을 포함하는 손실 함수(loss function)를 이용하여 학습될 수 있다. 딥러닝 네트워크는 각 손실에 대한 역전파 알고리즘을 이용하여 학습됨에 따라 학습 효율이 향상될 수 있고, 이러한 딥러닝 네트워크가 구비된 영상 획득 장치(100)는 멀티스펙트럴 이미지 센서(200)로부터 획득한 영상에 대한 조명 정보를 보다 정확하게 추정할 수 있다.

[0149] 채널 어텐션 블록(CAB)을 포함하는 딥러닝 네트워크는 하이퍼 각 손실(AE_{hyper})과 XYZ 각 손실(AE_{XYZ})을 이용하여 학습될 수 있다. 예를 들어, 딥러닝 네트워크는 하기의 수학식 3과 같이 하이퍼 각 손실(AE_{hyper})과 XYZ 각 손실(AE_{XYZ})을 일정한 비율로 가중 합(weighted sum)하여 손실함수로서 사용할 수 있다.

[0150] 딥러닝 네트워크는 하이퍼 각 손실(AE_{hyper})과 XYZ 각 손실(AE_{XYZ})을 모두 이용하여 학습되므로, XYZ 색공간으로 변환된 조명 정보보다 고차원인 멀티스펙트럴 이미지 센서(200)에 의해 측정되는 조명 정보에 대해서도 충분히 학습될 수 있다. 이에 따라, 본 개시의 실시 예에 따른 영상 획득 장치는 하이퍼 각 손실(AE_{hyper})과 XYZ 각 손실(AE_{XYZ})을 모두 이용하여 학습된 딥러닝 네트워크를 이용하여 조명 정보를 추정함으로써, 멀티스펙트럴 이미지 센서(200)로부터 획득한 영상에 대한 조명 정보를 보다 정확하게 추정할 수 있다.

[0151] 또한, 딥러닝 네트워크는 하이퍼 각 손실(AE_{hyper})과 XYZ 각 손실(AE_{XYZ})을 일정한 비율로 가중 합하여 손실함수로서 사용하여 학습되므로, 멀티스펙트럴 이미지 센서(200)에 의해 측정되는 조명 정보와 XYZ 색공간으로 변환된 조명 정보의 관계성에 대해서도 학습될 수 있다. 이에 따라, 본 개시의 실시 예에 따른 영상 획득 장치

는 하이퍼 각 손실(AE_{hyper})과 XYZ 각 손실(AE_{XYZ})을 일정한 비율로 가중 합하여 손실함수로서 사용하여 학습된 딥러닝 네트워크를 이용하여 조명 정보를 추정함으로써, XYZ 색공간으로 변환된 조명 정보 또한 보다 정확하게 추정할 수 있다.

[0152] 영상 획득 장치(100)에 의해 영상이 촬영된 환경에는 하나 이상의 광원이 존재할 수 있다. 멀티스펙트럴 이미지 센서(200)에 의해 획득되는 영상이 서로 다른 제1 광원과 제2 광원이 존재하는 환경에서 획득된 혼합 조명 영상인 경우, 영상 획득 장치(100)는 채널 어텐션 블록(CAB)을 2개 이상 포함된 딥러닝 네트워크를 이용할 수 있다.

[0153] 영상 획득 장치(100)는 멀티스펙트럴 이미지 센서(200)에 의해 획득되는 혼합 조명 영상을 제1 광원이 우세한 영역과 제2 광원이 우세한 영역으로 분할할 수 있다. 영상 획득 장치(100)는 혼합 조명 영상 중 제1 광원이 우세한 영역에 대하여 딥러닝 네트워크를 이용하여 제1 광원에 대한 조명 정보를 추정하고, 혼합 조명 영상 중 제2 광원이 우세한 영역에 대하여 딥러닝 네트워크를 이용하여 제2 광원에 대한 조명 정보를 추정할 수 있다. 영상 획득 장치(100)는 제1 광원에 대한 조명 정보에 대응하는 제1 조명 벡터와 제2 광원에 대한 조명 정보에 대응하는 제2 조명 벡터를 일정한 비율로 가중 합할 수 있으며, 이 때 가중 계수는 0 이상이며 가중 계수들의 합은 1 이 되도록 설정될 수 있다.

[0154] 영상 획득 장치(100)는 채널 어텐션 블록(CAB)은 풀링 레이어 및 제1 완전 연결 층을 포함할 수 있다. 멀티스펙트럴 이미지 센서(200)로부터 획득된 영상에 전역 맥스 풀링 또는 전역 평균 풀링을 수행하여 멀티스펙트럴 이미지 센서(200)로부터 획득된 영상의 전역에 대한 최대값 또는 평균값을 벡터의 요소로 구성하는 제1 채널 벡터를 생성할 수 있다.

[0155] 딥러닝 네트워크는 생성된 제1 채널 벡터를 채널 어텐션 블록(CAB)의 제1 완전 연결 층의 출력 결과에 더하여 컨볼루션 레이어(conv)로부터 출력된 컨볼루션 연산의 수행 결과인 출력 피쳐 맵과 컨볼루션 연산이 수행되도록 할 수 있다.

[0156] 생성되는 제1 채널 벡터의 차원은 멀티스펙트럴 이미지 센서(200)로부터 획득된 영상에 대응하는 채널의 개수와 동일할 수 있으나, 이에 제한되지 않으며, 멀티스펙트럴 이미지 센서(200)로부터 획득된 영상에 대응하는 채널의 개수보다 많도록 조정될 수 있다. 딥러닝 네트워크는 완전 연결 층을 이용하여 제1 채널 벡터의 차원을 조정할 수 있다.

[0157] 딥러닝 네트워크가 멀티스펙트럴 이미지 센서(200)로부터 획득된 영상에 전역 맥스 풀링 또는 전역 평균 풀링을 수행하여 생성한 제1 채널 벡터를 컨볼루션 레이어로부터 출력되는 출력 피처 맵에 입력함으로써, 딥러닝 네트워크는 딥러닝 네트워크에 입력되는 영상에 대응하는 채널들 간 채널 상관관계에 따라 채널 별로 획득되는 신호들 각각에 상이한 가중치를 부여할 수 있다. 이에 따라, 영상 획득 장치(100)는 멀티스펙트럴 이미지 센서(200)로부터 획득한 영상에 대한 조명 정보를 보다 정확하게 추정할 수 있다.

[0158] 특히, 딥러닝 네트워크가 멀티스펙트럴 이미지 센서(200)로부터 획득된 영상에 전역 평균 풀링을 수행하여 생성한 제1 채널 벡터를 컨볼루션 레이어(conv)로부터 출력되는 출력 피처 맵에 입력함으로써, 영상 획득 장치(100)는 파장 대역이 좁은 광원이 존재하는 환경에서 촬영된 영상에 대해서도 보다 정확하게 조명 정보를 추정할 수 있다.

[0159] 영상 획득 장치(100)는 중간 컨볼루션 레이어로부터 출력되는 중간 영상에 전역 맥스 풀링 또는 전역 평균 풀링을 수행하여 중간 영상의 전역 또는 로컬에 대한 최대값 또는 평균값을 벡터의 요소로 구성하는 제2 채널 벡터를 생성할 수 있다.

[0160] 생성되는 제2 채널 벡터의 차원은 중간 컨볼루션 레이어로부터 출력된 출력 피쳐 맵에 대응하는 채널의 개수와 동일할 수 있으나, 이에 제한되지 않으며, 중간 컨볼루션 레이어로부터 출력된 출력 피쳐 맵에 대응하는 채널의 개수보다 많도록 조정될 수 있다. 딥러닝 네트워크는 완전 연결 층을 이용하여 제2 채널 벡터의 차원을 조정할 수 있다.

[0161] 딥러닝 네트워크가 중간 컨볼루션 레이어로부터 출력된 출력 피쳐 맵에 전역 맥스 풀링, 전역 평균 풀링, 로컬 맥스 풀링 또는 로컬 평균 풀링을 수행하여 생성한 제2 채널 벡터를 중간 컨볼루션 레이어로부터 출력되는 출력 피쳐 맵에 입력함으로써, 딥러닝 네트워크는 딥러닝 네트워크에 입력되는 영상에 대응하는 채널들 간 채널 상관관계에 따라 채널 별로 획득되는 신호들 각각에 상이한 가중치를 부여할 수 있다. 이에 따라, 영상 획득 장치(100)는 멀티스펙트럴 이미지 센서(200)로부터 획득한 영상에 대한 조명 정보를 보다 정확하게 추정할 수 있다.

[0162] 특히, 딥러닝 네트워크가 중간 컨볼루션 레이어로부터 출력된 출력 피쳐 맵에 전역 평균 풀링 또는 로컬 평균

풀링을 수행하여 생성한 제2 채널 벡터를 중간 컨볼루션 레이어로부터 출력되는 출력 피처 맵에 입력함으로써, 영상 획득 장치(100)는 파장 대역이 좁은 광원이 존재하는 환경에서 촬영된 영상에 대해서도 보다 정확하게 조명 정보를 추정할 수 있다.

[0163] 영상 획득 장치(100)는 딥러닝 네트워크로부터 추정된 조명 정보를 벡터의 형태로 출력할 수 있다. 영상 획득 장치(100)는 딥러닝 네트워크로부터 추정된 조명 정보를 추정된 조명 정보에 대응하는 조명 벡터로 출력할 수 있다.

[0164] 영상 획득 장치(100)가 출력하는 조명 벡터의 차원은 딥러닝 네트워크에 입력되는 영상에 대응하는 채널들의 개수와 동일할 수 있다. 다만, 이에 제한되는 것은 아니다. 영상 획득 장치(100)는 조명 벡터의 차원을 멀티스펙트럴 이미지 센서(200)에 의해 획득되는 영상에 대응하는 채널들의 개수보다 많도록 조정할 수도 있고, 영상 획득 장치(100)는 조명 벡터의 차원을 3으로 조정할 수도 있다.

[0165] 단계 1130에서 영상 획득 장치(100)는 추정된 조명 정보에 기초하여 획득된 영상에 대한 색변환을 수행할 수 있다.

[0166] 영상 획득 장치(100)는 추정된 조명 정보에 대응하는 조명 벡터의 각 요소들을 등색 함수(Color Matching Function)를 이용하여 XYZ 색공간으로 변환함으로써 멀티스펙트럴 이미지 센서(200)에 의해 획득된 영상에 대하여 색변환을 수행할 수 있다.

[0167] 영상 획득 장치(100)는 추정된 조명 정보에 기초하여 멀티스펙트럴 이미지 센서(200)에 의해 획득된 영상에 대한 자동 화이트 밸런스(Auto White Balance)를 수행할 수 있다.

[0168] 다만, 영상 획득 장치(100)가 추정된 조명 정보에 기초하여 색변환 및/또는 자동 화이트 밸런스를 수행하는 대상은 멀티스펙트럴 이미지 센서(200)에 의해 획득된 영상에 한정되지 않으며, 영상 획득 장치(100)는 추정된 조명 정보에 기초하여 다른 이미지 센서에 의해 획득한 영상에 대한 색변환 및/또는 자동 화이트 밸런스를 수행할 수도 있다.

[0169] 영상 획득 장치(100)는 멀티스펙트럴 이미지 센서(200)에 의해 획득된 영상을 딥러닝 네트워크에 입력하여 추정된 조명 정보(이하, 하이퍼 조명 정보라고 지칭함)를 획득할 수 있고, 이와 동시에 RGB 센서에 의해 획득된 영상을 딥러닝 네트워크에 입력하여 추정된 조명 정보(이하, RGB 조명 정보라고 지칭함)를 획득할 수 있다.

[0170] 영상 획득 장치(100)는 하이퍼 조명 정보에 대응하는 하이퍼 조명 벡터의 각 요소들을 XYZ 색공간 상의 XYZ 값으로 변환한 후, RGB 조명 정보에 대응하는 RGB 조명 벡터의 각 요소들을 XYZ 값으로 변환된 조명 벡터의 각 요소들로 나눔에 따라, RGB 센서에 의해 획득된 영상에 대한 자동 화이트 밸런스를 수행할 수 있다.

[0171] **도 14는 일 실시 예에 따른 전자 장치의 구성을 나타낸 블록도이다.**

[0172] 도 14는 실시 예에 따른 전자 장치의 개략적인 구조를 보이는 블록도이다. 도 14를 참조하면, 네트워크 환경(ED00)에서 전자 장치(ED01)는 제1 네트워크(ED98)(근거리 무선 통신 네트워크 등)를 통하여 다른 전자 장치(ED02)와 통신하거나, 또는 제2 네트워크(ED99)(원거리 무선 통신 네트워크 등)를 통하여 또 다른 전자 장치(ED04) 및/또는 서버(ED08)와 통신할 수 있다. 전자 장치(ED01)는 서버(ED08)를 통하여 전자 장치(ED04)와 통신할 수 있다. 전자 장치(ED01)는 프로세서(ED20), 메모리(ED30), 입력 장치(ED50), 음향 출력 장치(ED55), 표시 장치(ED60), 오디오 모듈(ED70), 센서 모듈(ED76), 인터페이스(ED77), 햅틱 모듈(ED79), 카메라 모듈(ED80), 전력 관리 모듈(ED88), 배터리(ED89), 통신 모듈(ED90), 가입자 식별 모듈(ED96), 및/또는 안테나 모듈(ED97)을 포함할 수 있다. 전자 장치(ED01)에는, 이 구성요소들 중 일부(표시 장치(ED60) 등)가 생략되거나, 다른 구성요소가 추가될 수 있다. 이 구성요소들 중 일부는 하나의 통합된 회로로 구현될 수 있다. 예를 들면, 센서 모듈(ED76)(지문 센서, 홍채 센서, 조도 센서 등)은 표시 장치(ED60)(디스플레이 등)에 임베디드되어 구현될 수 있다. 또한, 이미지센서(1000)에 분광 기능이 포함될 경우, 센서 모듈의 일부 기능(컬러 센서, 조도 센서)이 별도의 센서 모듈이 아닌 이미지센서(1000) 자체에서 구현될 수 있다.

[0173] 프로세서(ED20)는, 소프트웨어(프로그램(ED40) 등)를 실행하여 프로세서(ED20)에 연결된 전자 장치(ED01) 중 하나 또는 복수개의 다른 구성요소들(하드웨어, 소프트웨어 구성요소 등)을 제어할 수 있고, 다양한 데이터 처리 또는 연산을 수행할 수 있다. 데이터 처리 또는 연산의 일부로, 프로세서(ED20)는 다른 구성요소(센서 모듈(ED76), 통신 모듈(ED90) 등)로부터 수신된 명령 및/또는 데이터를 휘발성 메모리(ED32)에 로드하고, 휘발성 메모리(ED32)에 저장된 명령 및/또는 데이터를 처리하고, 결과 데이터를 비휘발성 메모리(ED34)에 저장할 수 있다. 프로세서(ED20)는 메인 프로세서(ED21)(중앙 처리 장치, 어플리케이션 프로세서 등) 및 이와 독립적으로

또는 함께 운영 가능한 보조 프로세서(ED23)(그래픽 처리 장치, 이미지 시그널 프로세서, 센서 허브 프로세서, 커뮤니케이션 프로세서 등)를 포함할 수 있다. 보조 프로세서(ED23)는 메인 프로세서(ED21)보다 전력을 작게 사용하고, 특화된 기능을 수행할 수 있다.

[0174] 보조 프로세서(ED23)는, 메인 프로세서(ED21)가 인액티브 상태(슬립 상태)에 있는 동안 메인 프로세서(ED21)를 대신하여, 또는 메인 프로세서(ED21)가 액티브 상태(어플리케이션 실행 상태)에 있는 동안 메인 프로세서(ED21)와 함께, 전자 장치(ED01)의 구성요소들 중 일부 구성요소(표시 장치(ED60), 센서 모듈(ED76), 통신 모듈(ED90) 등)와 관련된 기능 및/또는 상태를 제어할 수 있다. 보조 프로세서(ED23)(이미지 시그널 프로세서, 커뮤니케이션 프로세서 등)는 기능적으로 관련 있는 다른 구성 요소(카메라 모듈(ED80), 통신 모듈(ED90) 등)의 일부로서 구현될 수도 있다.

[0175] 메모리(ED30)는, 전자 장치(ED01)의 구성요소(프로세서(ED20), 센서모듈(ED76) 등)가 필요로 하는 다양한 데이터를 저장할 수 있다. 데이터는, 예를 들어, 소프트웨어(프로그램(ED40) 등) 및, 이와 관련된 명령에 대한 입력 데이터 및/또는 출력 데이터를 포함할 수 있다. 메모리(ED30)는, 휘발성 메모리(ED32) 및/또는 비휘발성 메모리(ED34)를 포함할 수 있다. 비휘발성 메모리(ED32)는 전자 장치(ED01) 내에 고정 장착된 내장 메모리(ED36)와 탈착 가능한 외장 메모리(ED38)를 포함할 수 있다.

[0176] 프로그램(ED40)은 메모리(ED30)에 소프트웨어로 저장될 수 있으며, 운영 체제(ED42), 미들 웨어(ED44) 및/또는 어플리케이션(ED46)을 포함할 수 있다.

[0177] 입력 장치(ED50)는, 전자 장치(ED01)의 구성요소(프로세서(ED20) 등)에 사용될 명령 및/또는 데이터를 전자 장치(ED01)의 외부(사용자 등)로부터 수신할 수 있다. 입력 장치(ED50)는, 마이크, 마우스, 키보드, 및/또는 디지털 펜(스타일러스 펜 등)을 포함할 수 있다.

[0178] 음향 출력 장치(ED55)는 음향 신호를 전자 장치(ED01)의 외부로 출력할 수 있다. 음향 출력 장치(ED55)는, 스피커 및/또는 리시버를 포함할 수 있다. 스피커는 멀티미디어 재생 또는 녹음 재생과 같이 일반적인 용도로 사용될 수 있고, 리시버는 착신 전화를 수신하기 위해 사용될 수 있다. 리시버는 스피커의 일부로 결합되어 있거나 또는 독립된 별도의 장치로 구현될 수 있다.

[0179] 표시 장치(ED60)는 전자 장치(ED01)의 외부로 정보를 시각적으로 제공할 수 있다. 표시 장치(ED60)는, 디스플레이, 홀로그램 장치, 또는 프로젝터 및 해당 장치를 제어하기 위한 제어 회로를 포함할 수 있다. 표시 장치(ED60)는 터치를 감지하도록 설정된 터치 회로(Touch Circuitry), 및/또는 터치에 의해 발생되는 힘의 세기를 측정하도록 설정된 센서 회로(압력 센서 등)를 포함할 수 있다.

[0180] 오디오 모듈(ED70)은 소리를 전기 신호로 변환시키거나, 반대로 전기 신호를 소리로 변환시킬 수 있다. 오디오 모듈(ED70)은, 입력 장치(ED50)를 통해 소리를 획득하거나, 음향 출력 장치(ED55), 및/또는 전자 장치(ED01)와 직접 또는 무선으로 연결된 다른 전자 장치(전자 장치(ED02) 등)의 스피커 및/또는 헤드폰을 통해 소리를 출력할 수 있다.

[0181] 센서 모듈(ED76)은 전자 장치(ED01)의 작동 상태(전력, 온도 등), 또는 외부의 환경 상태(사용자 상태 등)를 감지하고, 감지된 상태에 대응하는 전기 신호 및/또는 데이터 값을 생성할 수 있다. 센서 모듈(ED76)은, 제스처 센서, 자이로 센서, 기압 센서, 마그네틱 센서, 가속도 센서, 그립 센서, 근접 센서, 컬러 센서, IR(Infrared) 센서, 생체 센서, 온도 센서, 습도 센서, 및/또는 조도 센서를 포함할 수 있다.

[0182] 인터페이스(ED77)는 전자 장치(ED01)가 다른 전자 장치(전자 장치(ED02) 등)와 직접 또는 무선으로 연결되기 위해 사용될 수 있는 하나 또는 복수의 지정된 프로토콜들을 지원할 수 있다. 인터페이스(ED77)는, HDMI(High Definition Multimedia Interface), USB(Universal Serial Bus) 인터페이스, SD카드 인터페이스, 및/또는 오디오 인터페이스를 포함할 수 있다.

[0183] 연결 단자(ED78)는, 전자 장치(ED01)가 다른 전자 장치(전자 장치(ED02) 등)와 물리적으로 연결될 수 있는 커넥터를 포함할 수 있다. 연결 단자(ED78)는, HDMI 커넥터, USB 커넥터, SD 카드 커넥터, 및/또는 오디오 커넥터(헤드폰 커넥터 등)를 포함할 수 있

[0184] 햅틱 모듈(ED79)은 전기적 신호를 사용자가 촉각 또는 운동 감각을 통해서 인지할 수 있는 기계적인 자극(진동, 움직임 등) 또는 전기적인 자극으로 변환할 수 있다. 햅틱 모듈(ED79)은, 모터, 압전 소자, 및/또는 전기 자극 장치를 포함할 수 있다.

[0185] 카메라 모듈(ED80)은 정지 영상 및 동영상을 촬영할 수 있다. 카메라 모듈(ED80)은 상술한 영상 획득 장치

(1000)를 포함할 수 있고, 추가적인 렌즈 어셈블리, 이미지 시그널 프로세서들, 및/또는 플래시들을 포함할 수 있다. 카메라 모듈(ED80)에 포함된 렌즈 어셈블리는 이미지 촬영의 대상인 피사체로부터 방출되는 빛을 수집할 수 있다.

[0186] 전력 관리 모듈(ED88)은 전자 장치(ED01)에 공급되는 전력을 관리할 수 있다. 전력 관리 모듈(ED88)은, PMIC(Power Management Integrated Circuit)의 일부로서 구현될 수 있다.

[0187] 배터리(ED89)는 전자 장치(ED01)의 구성 요소에 전력을 공급할 수 있다. 배터리(ED89)는, 재충전 불가능한 1차 전지, 재충전 가능한 2차 전지 및/또는 연료 전지를 포함할 수 있다.

[0188] 통신 모듈(ED90)은 전자 장치(ED01)와 다른 전자 장치(전자 장치(ED02), 전자 장치(ED04), 서버(ED08) 등)간의 직접(유선) 통신 채널 및/또는 무선 통신 채널의 수립, 및 수립된 통신 채널을 통한 통신 수행을 지원할 수 있다. 통신 모듈(ED90)은 프로세서(ED20)(어플리케이션 프로세서 등)와 독립적으로 운영되고, 직접 통신 및/또는 무선 통신을 지원하는 하나 또는 복수의 커뮤니케이션 프로세서들을 포함할 수 있다. 통신 모듈(ED90)은 무선 통신 모듈(ED92)(셀룰러 통신 모듈, 근거리 무선 통신 모듈, GNSS(Global Navigation Satellite System 등) 통신 모듈) 및/또는 유선 통신 모듈(ED94)(LAN(Local Area Network) 통신 모듈, 전력선 통신 모듈 등)을 포함할 수 있다. 이들 통신 모듈 중 해당하는 통신 모듈은 제1 네트워크(ED98)(블루투스, WiFi Direct 또는 IrDA(Infrared Data Association) 같은 근거리 통신 네트워크) 또는 제2 네트워크(ED99)(셀룰러 네트워크, 인터넷, 또는 컴퓨터 네트워크(LAN, WAN 등)와 같은 원거리 통신 네트워크)를 통하여 다른 전자 장치와 통신할 수 있다. 이런 여러 종류의 통신 모듈들은 하나의 구성 요소(단일 칩 등)로 통합되거나, 또는 서로 별도의 복수의 구성 요소들(복수 칩들)로 구현될 수 있다. 무선 통신 모듈(ED92)은 가입자 식별 모듈(ED96)에 저장된 가입자 정보(국제 모바일 가입자 식별자(IMSI) 등)를 이용하여 제1 네트워크(ED98) 및/또는 제2 네트워크(ED99)와 같은 통신 네트워크 내에서 전자 장치(ED01)를 확인 및 인증할 수 있다.

[0189] 안테나 모듈(ED97)은 신호 및/또는 전력을 외부(다른 전자 장치 등)로 송신하거나 외부로부터 수신할 수 있다. 안테나는 기판(PCB 등) 위에 형성된 도전성 패턴으로 이루어진 방사체를 포함할 수 있다. 안테나 모듈(ED97)은 하나 또는 복수의 안테나들을 포함할 수 있다. 복수의 안테나가 포함된 경우, 통신 모듈(ED90)에 의해 복수의 안테나들 중에서 제1 네트워크(ED98) 및/또는 제2 네트워크(ED99)와 같은 통신 네트워크에서 사용되는 통신 방식에 적합한 안테나가 선택될 수 있다. 선택된 안테나를 통하여 통신 모듈(ED90)과 다른 전자 장치 간에 신호 및/또는 전력이 송신되거나 수신될 수 있다. 안테나 외에 다른 부품(RFIC 등)이 안테나 모듈(ED97)의 일부로 포함될 수 있다.

[0190] 구성요소들 중 일부는 주변 기기들간 통신 방식(버스, GPIO(General Purpose Input and Output), SPI(Serial Peripheral Interface), MIPI(Mobile Industry Processor Interface) 등)을 통해 서로 연결되고 신호(명령, 데이터 등)를 상호 교환할 수 있다.

[0191] 명령 또는 데이터는 제2 네트워크(ED99)에 연결된 서버(ED08)를 통해서 전자 장치(ED01)와 외부의 전자 장치(ED04)간에 송신 또는 수신될 수 있다. 다른 전자 장치들(ED02, ED04)은 전자 장치(ED01)와 동일한 또는 다른 종류의 장치일 수 있다. 전자 장치(ED01)에서 실행되는 동작들의 전부 또는 일부는 다른 전자 장치들(ED02, ED04, ED08) 중 하나 또는 복수의 장치들에서 실행될 수 있다. 예를 들면, 전자 장치(ED01)가 어떤 기능이나 서비스를 수행해야 할 때, 기능 또는 서비스를 자체적으로 실행시키는 대신에 하나 또는 복수의 다른 전자 장치들에게 그 기능 또는 그 서비스의 일부 또는 전체를 수행하라고 요청할 수 있다. 요청을 수신한 하나 또는 복수의 다른 전자 장치들은 요청과 관련된 추가 기능 또는 서비스를 실행하고, 그 실행의 결과를 전자 장치(ED01)로 전달할 수 있다. 이를 위하여, 클라우드 컴퓨팅, 분산 컴퓨팅, 및/또는 클라이언트-서버 컴퓨팅 기술이 이용될 수 있다.

[0192] **도 15는 도 14의 전자 장치에 구비되는 카메라 모듈을 도시한 블록도이다.**

[0193] 도 24은 도 9의 전자 장치에 구비되는 카메라 모듈(ED80)을 개략적으로 예시한 블록도이다. 카메라 모듈(ED80)은 전술한 영상 획득 장치(1000)를 포함할 수 있고, 또는 이로부터 변형된 구조를 가질 수 있다. 도 24을 참조하면, 카메라 모듈(ED80)은 렌즈 어셈블리(CM10), 플래시(CM20), 이미지 센서(CM30), 이미지 스태빌라이저(CM40), 메모리(CM50)(버퍼 메모리 등), 및/또는 이미지 시그널 프로세서(CM60)를 포함할 수 있다.

[0194] 이미지 센서(CM30)는 도 1 내지 도 5를 참조하여 설명한 이미지 센서로서, 톤카운팅을 이용한다. 따라서, 이미지센서의 픽셀 크기를 최소화하면서 HDR와 높은 프레임 레이트를 구현할 수 있다. 실시 예에 따른 이미지 센서의 픽셀은 SPAD를 포함하고, 픽셀에 입사된 포톤에 따라 생성된 전기 신호를 기초로 펄스 신호를 생성하고, 펄

스 신호의 갯수를 카운팅하여 출력하는 카운터를 포함하고, 펄스 신호의 주기는, 이미지 센서의 주변 조도에 따라 변조될 수 있다. 따라서, 이미지 센서(CM30)는 픽셀의 크기를 줄일 수 있고, 프레임 레이트를 향상시킬 수 있다.

[0195] 렌즈 어셈블리(CM10)는 이미지 촬영의 대상인 피사체로부터 방출되는 빛을 수집할 수 있다. 카메라 모듈(ED80)은 복수의 렌즈 어셈블리(CM10)들을 포함할 수도 있으며, 이런 경우, 카메라 모듈(ED80)은, 듀얼 카메라, 360도 카메라, 또는 구형 카메라(Spherical Camera)가 될 수 있다. 복수의 렌즈 어셈블리(CM10)들 중 일부는 동일한 렌즈 속성(화각, 초점 거리, 자동 초점, F 넘버(F Number), 광학 줌 등)을 갖거나, 또는 다른 렌즈 속성들을 가질 수 있다. 렌즈 어셈블리(CM10)는, 광각 렌즈 또는 망원 렌즈를 포함할 수 있다.

[0196] 렌즈 어셈블리(CM10)는 이미지 센서(CM30)에 구비되는 이미지 센서가 피사체의 광학 상을 형성할 수 있도록 구성될 수 있고 및/또는 포커스 제어될 수 있다.

[0197] 플래시(CM20)는 피사체로부터 방출 또는 반사되는 빛을 강화하기 위하여 사용되는 빛을 방출할 수 있다. 플래시(CM20)는 하나 또는 복수의 발광 다이오드들(RGB(Red-Green-Blue) LED, White LED, Infrared LED, Ultraviolet LED 등), 및/또는 Xenon Lamp를 포함할 수 있다.

[0198] 이미지 스태빌라이저(CM40)는 카메라 모듈(ED80) 또는 이를 포함하는 전자 장치(CM01)의 움직임에 반응하여, 렌즈 어셈블리(CM10)에 포함된 하나 또는 복수개의 렌즈 또는 이미지센서(1000)를 특정한 방향으로 움직이거나 이미지센서(1000)의 동작 특성을 제어(리드 아웃(Read-Out) 타이밍의 조정 등)하여 움직임에 의한 부정적인 영향이 보상되도록 할 수 있다. 이미지 스태빌라이저(CM40)는 카메라 모듈(ED80)의 내부 또는 외부에 배치된 자이로 센서(미도시) 또는 가속도 센서(미도시)를 이용하여 카메라 모듈(ED80) 또는 전자 장치(ED01)의 움직임을 감지할 수 있다. 이미지 스태빌라이저(CM40)는, 광학식으로 구현될 수도 있다.

[0199] 메모리(CM50)는 이미지센서(1000)를 통하여 획득된 이미지의 일부 또는 전체 데이터가 다음 이미지 처리 작업을 위하여 저장할 수 있다. 예를 들어, 복수의 이미지들이 고속으로 획득되는 경우, 획득된 원본 데이터(Bayer-Patterned 데이터, 고해상도 데이터 등)는 메모리(CM50)에 저장하고, 저해상도 이미지만을 디스플레이 해준 후, 선택된(사용자 선택 등) 이미지의 원본 데이터가 이미지 시그널 프로세서(CM60)로 전달되도록 하는데 사용될 수 있다. 메모리(CM50)는 전자 장치(ED01)의 메모리(ED30)로 통합되어 있거나, 또는 독립적으로 운영되는 별도의 메모리로 구성될 수 있다.

[0200] 이미지 시그널 프로세서(CM60)는 이미지 센서(CM30)를 통하여 획득된 이미지 또는 메모리(CM50)에 저장된 이미지 데이터에 대하여 이미지 처리들을 수행할 수 있다.

[0201] 이미지 처리들은, 이외에도, 깊이 지도(Depth Map) 생성, 3차원 모델링, 파노라마 생성, 특징점 추출, 이미지 합성, 및/또는 이미지 보상(노이즈 감소, 해상도 조정, 밝기 조정, 블러링(Blurring), 샤프닝(Sharpening), 소프트닝(Softening) 등)을 포함할 수 있다. 이미지 시그널 프로세서(CM60)는 카메라 모듈(ED80)에 포함된 구성 요소들(이미지 센서(CM30) 등)에 대한 제어(노출 시간 제어, 또는 리드 아웃 타이밍 제어 등)를 수행할 수 있다. 이미지 시그널 프로세서(CM60)에 의해 처리된 이미지는 추가 처리를 위하여 메모리(CM50)에 다시 저장 되거나 카메라 모듈(ED80)의 외부 구성 요소(메모리(ED30), 표시 장치(ED60), 전자 장치(ED02), 전자 장치(ED04), 서버(ED08) 등)로 제공될 수 있다. 이미지 시그널 프로세서(CM60)는 프로세서(ED20)에 통합되거나, 프로세서(ED20)와 독립적으로 운영되는 별도의 프로세서로 구성될 수 있다. 이미지 시그널 프로세서(CM60)가 프로세서(ED20)와 별도의 프로세서로 구성된 경우, 이미지 시그널 프로세서(CM60)에 의해 처리된 이미지는 프로세서(ED20)에 의하여 추가의 이미지 처리를 거친 후 표시 장치(ED60)를 통해 표시될 수 있다.

[0202] 전자 장치(ED01)는 각각 다른 속성 또는 기능을 가진 복수의 카메라 모듈(ED80)들을 포함할 수 있다. 이런 경우, 복수의 카메라 모듈(ED80)들 중 하나는 광각 카메라이고, 다른 하나는 망원 카메라일 수 있다. 유사하게, 복수의 카메라 모듈(ED80)들 중 하나는 전면 카메라이고, 다른 하나는 후면 카메라일 수 있다.

[0203] **도 16 내지 도 25은 다양한 실시 예들에 따른 영상 획득 장치가 적용될 전자 장치의 다양한 예시들을 도시한 도면이다.**

[0204] 실시 예들에 따른 영상 획득 장치는 도 16에 도시된 모바일 폰 또는 스마트 폰(5100m), 도 17에 도시된 태블릿 또는 스마트 태블릿(5200), 도 18에 도시된 디지털 카메라 또는 캠코더(5300), 도 19에 도시된 노트북 컴퓨터(5400)에 또는 도 24에 도시된 텔레비전 또는 스마트 텔레비전(5500) 등에 적용될 수 있다. 예를 들어, 스마트 폰(5100m) 또는 스마트 태블릿(5200)은 고해상 이미지센서가 각각 탑재된 복수의 고해상 카메라를 포함할 수 있다. 고해상 카메라들을 이용하여 영상 내 피사체들의 깊이 정보를 추출하거나, 영상의 아웃포커싱을

조절하거나, 영상 내 피사체들을 자동으로 식별할 수 있다.

[0205] 또한, 영상 획득 장치(1000)는 도 21에 도시된 스마트 냉장고(5600), 도 23에 도시된 보안 카메라(5700), 도 24에 도시된 로봇(5800), 도 25에 도시된 의료용 카메라(5900) 등에 적용될 수 있다. 예를 들어, 스마트 냉장고(5600)는 영상 획득 장치(1000)를 이용하여 냉장고 내에 있는 음식을 자동으로 인식하고, 특정 음식의 존재 여부, 입고 또는 출고된 음식의 종류 등을 스마트폰을 통해 사용자에게 알려줄 수 있다. 보안 카메라(5700)는 초고해상도 영상을 제공할 수 있으며 높은 감도를 이용하여 어두운 환경에서도 영상 내의 사물 또는 사람을 인식 가능하게 할 수 있다. 로봇(5800)은 사람이 직접 접근할 수 없는 재해 또는 산업 현장에서 투입되어 고해상도 영상을 제공할 수 있다. 의료용 카메라(5900)는 진단 또는 수술을 위한 고해상도 영상을 제공할 수 있으며 시야를 동적으로 조절할 수 있다.

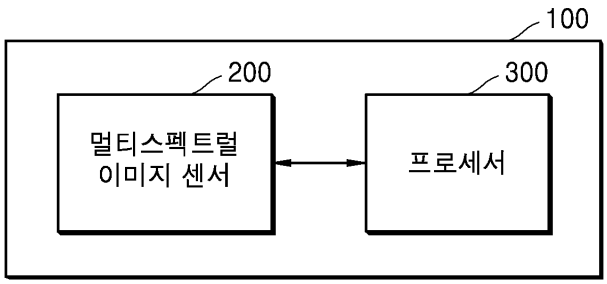
[0206] 또한, 영상 획득 장치(1000)는 도 25에 도시된 바와 같이 차량(6000)에 적용될 수 있다. 차량(6000)은 다양한 위치에 배치된 복수의 차량용 카메라(6010, 6020, 6030, 6040)를 포함할 수 있으며. 각각의 차량용 카메라(6010, 6020, 6030, 6040)는 실시 예에 따른 영상 획득 장치를 포함할 수 있다. 차량(6000)은 복수의 차량용 카메라(6010, 6020, 6030, 6040)를 이용하여 차량(6000) 내부 또는 주변에 대한 다양한 정보를 운전자에게 제공할 수 있으며, 영상 내의 사물 또는 사람을 자동으로 인식하여 자율 주행에 필요한 정보를 제공할 수 있다.

[0207] 한편, 전술한 방법은 그 방법을 실행하는 명령어들을 포함하는 하나 이상의 프로그램이 기록된 컴퓨터로 읽을 수 있는 비일시적(non-transitory) 기록 매체에 기록될 수 있다. 컴퓨터로 읽을 수 있는 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 명령어의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

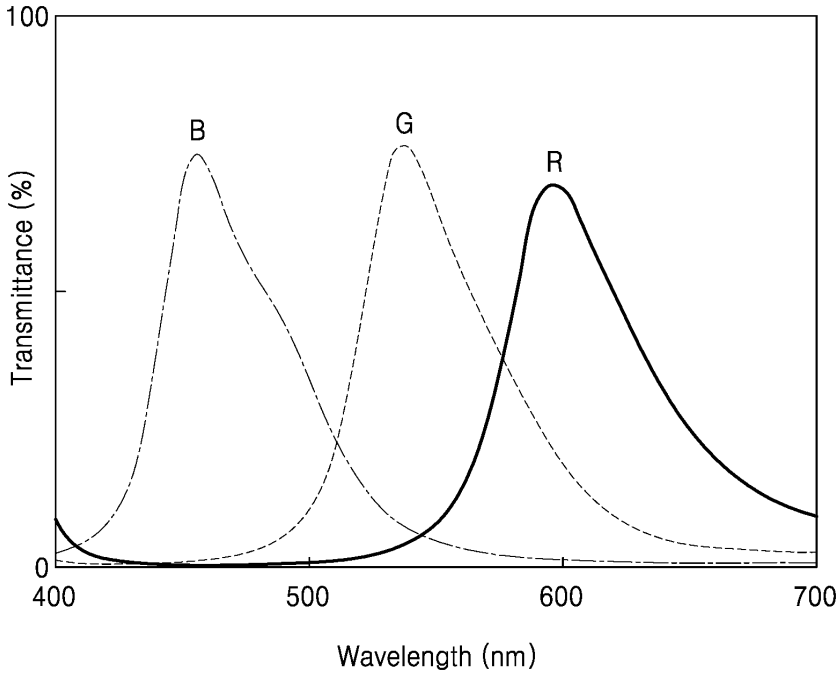
[0208] 이상에서 실시예들에 대하여 상세하게 설명하였지만 본 개시의 권리범위는 이에 한정되는 것은 아니고 다음의 청구범위에서 정의하고 있는 본 개시의 기본 개념을 이용한 당업자의 여러 변형 및 개량 형태 또한 본 개시의 권리범위에 속한다.

도면

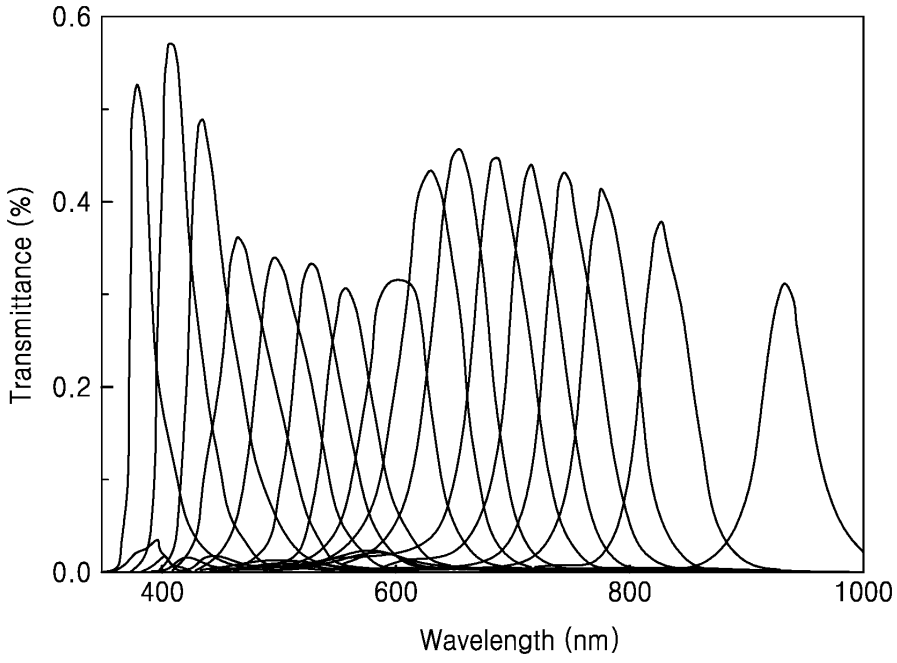
도면1



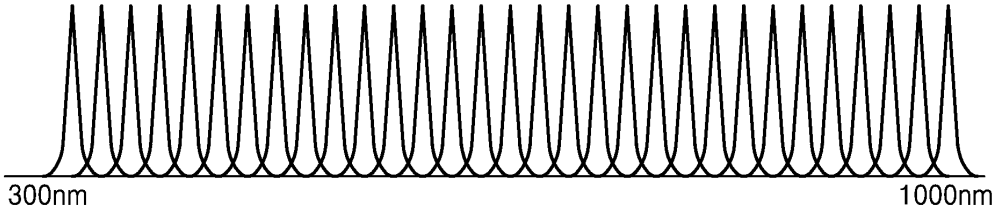
도면2a



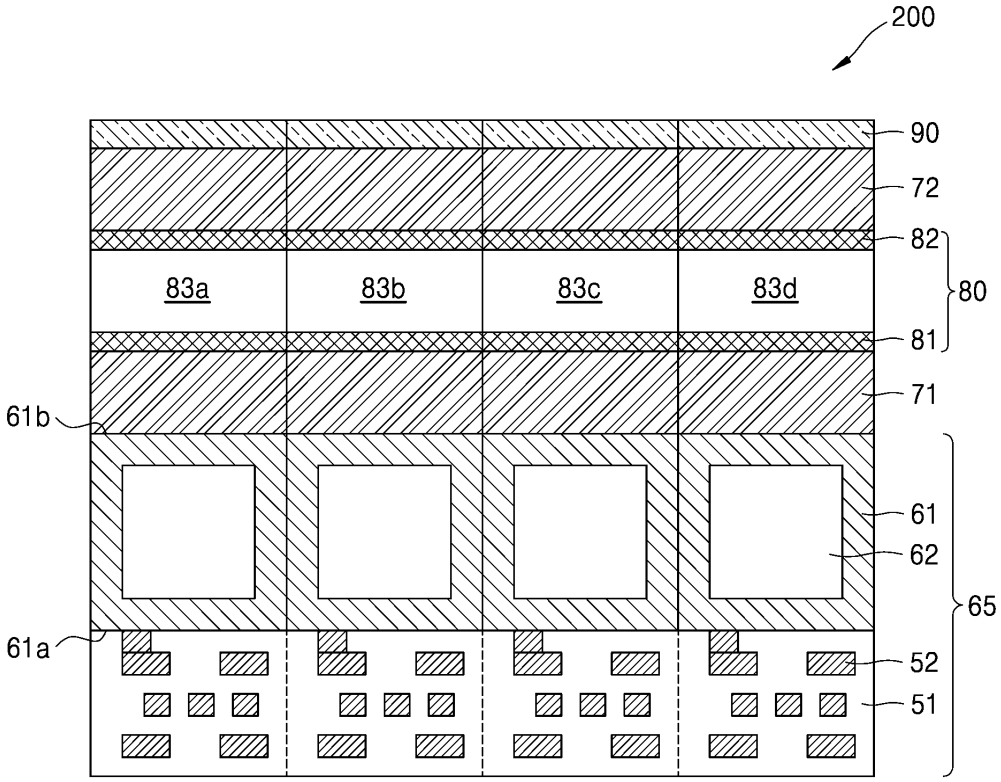
도면2b



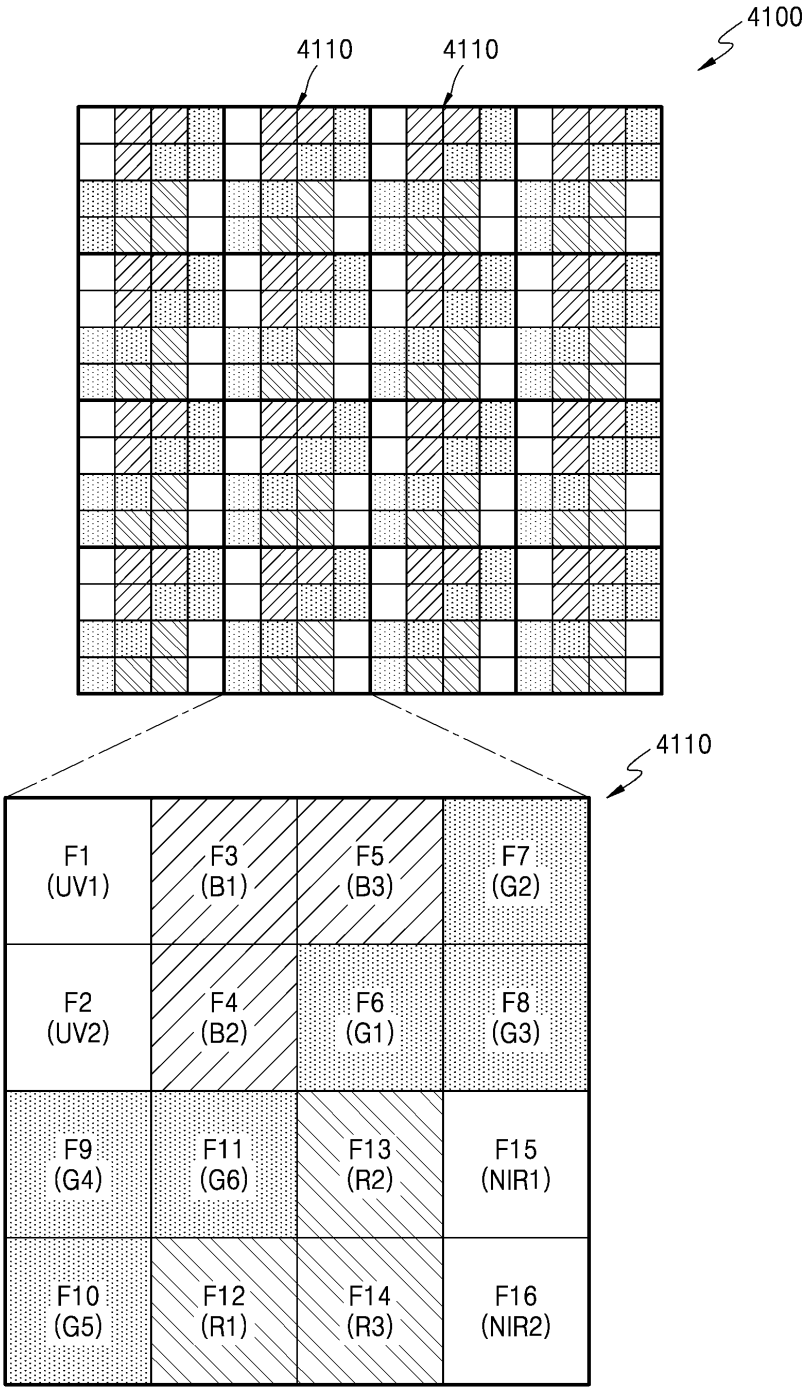
도면2c



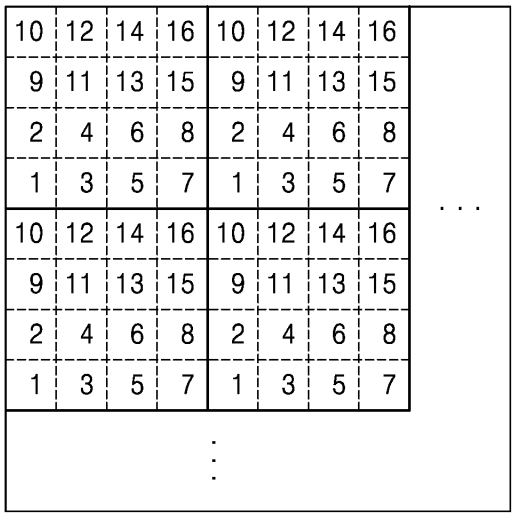
도면3



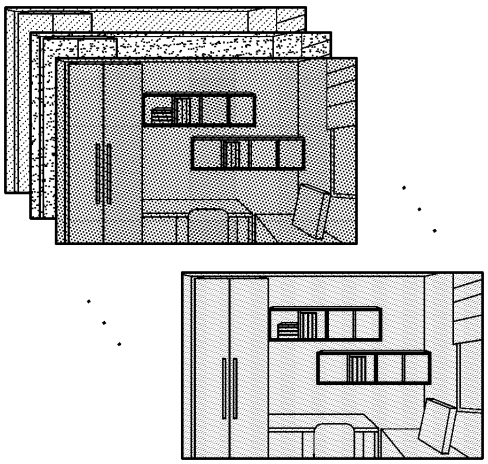
도면4


4100
4110
4110
4110
F1 (UV1)
F3 (B1)
F5 (B3)
F7 (G2)
F2 (UV2)
F4 (B2)
F6 (G1)
F8 (G3)
F9 (G4)
F11 (G6)
F13 (R2)
F15 (NIR1)
F10 (G5)
F12 (R1)
F14 (R3)
F16 (NIR2)

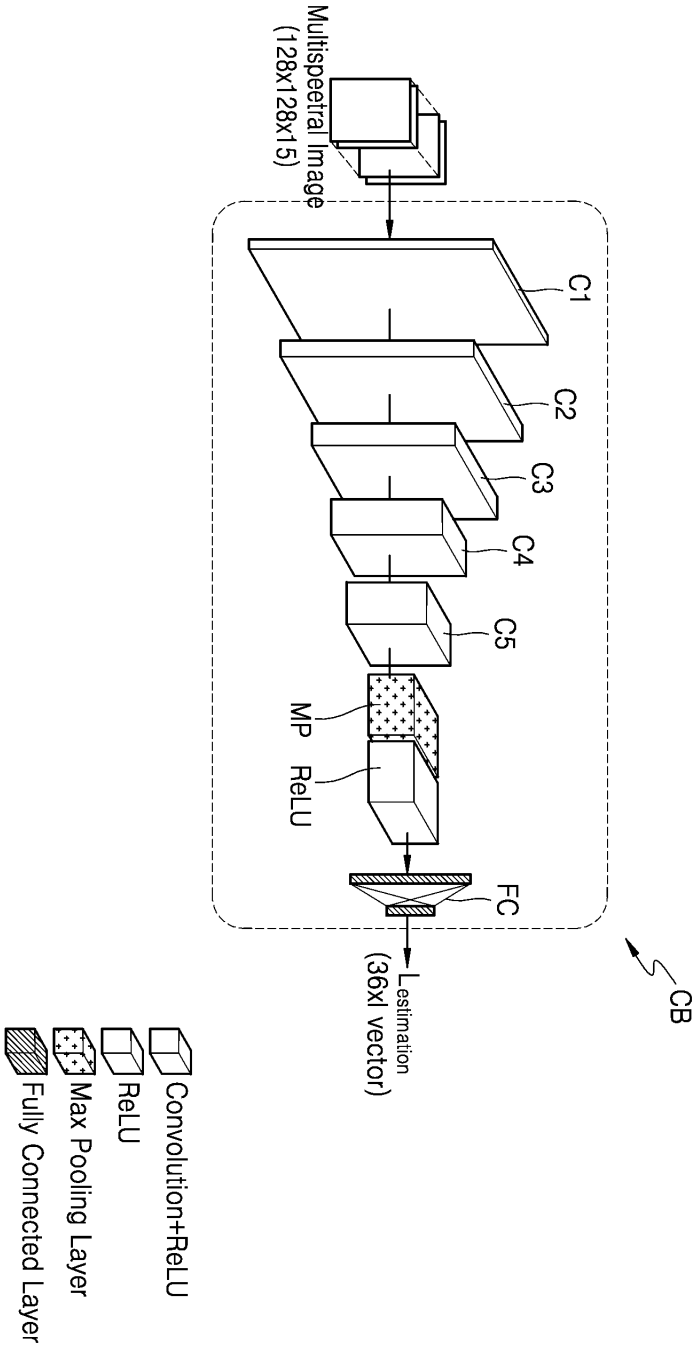
도면5a


10 12 14 16 10 12 14 16
9 11 13 15 9 11 13 15
2 4 6 8 2 4 6 8
1 3 5 7 1 3 5 7
10 12 14 16 10 12 14 16
9 11 13 15 9 11 13 15
2 4 6 8 2 4 6 8
1 3 5 7 1 3 5 7

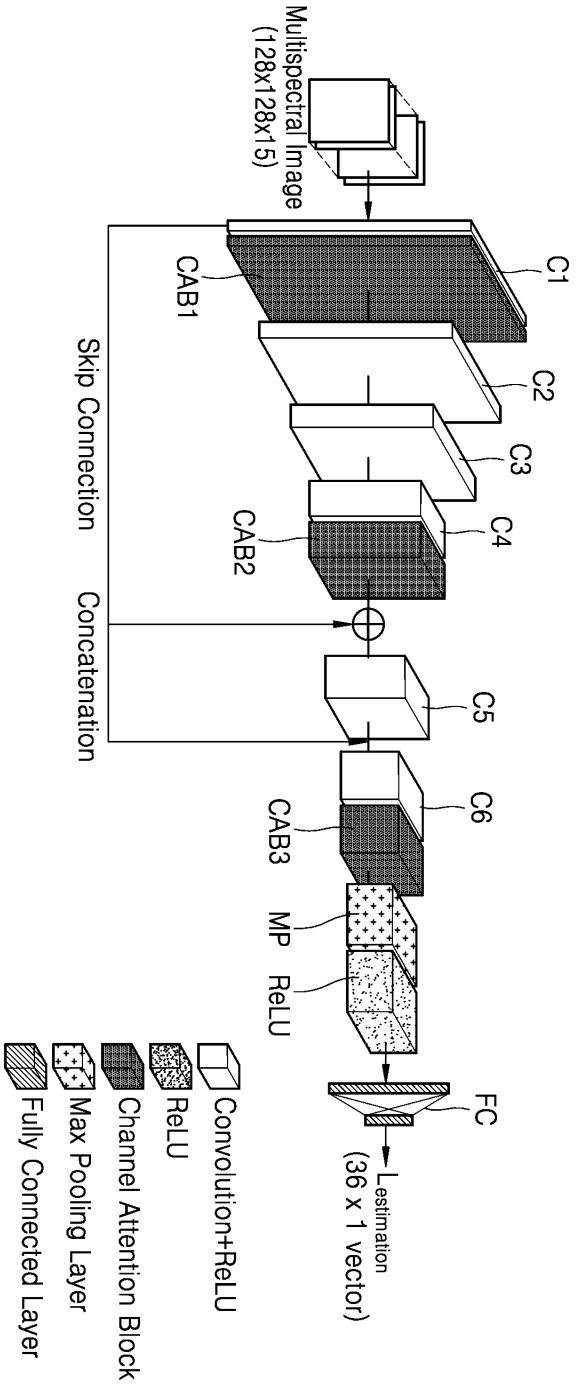
도면5b



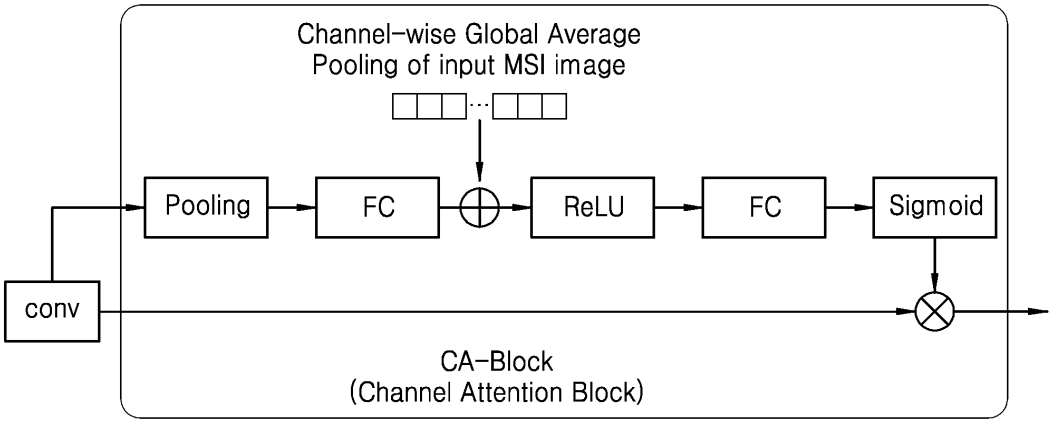
도면6


Multispectral Image
(128x128x15)
C1
C2
C3
C4
C5
MP
ReLU
FC
Lestimation
(36x1 vector)
CB
Convolution+ReLU
ReLU
Max Pooling Layer
Fully Connected Layer

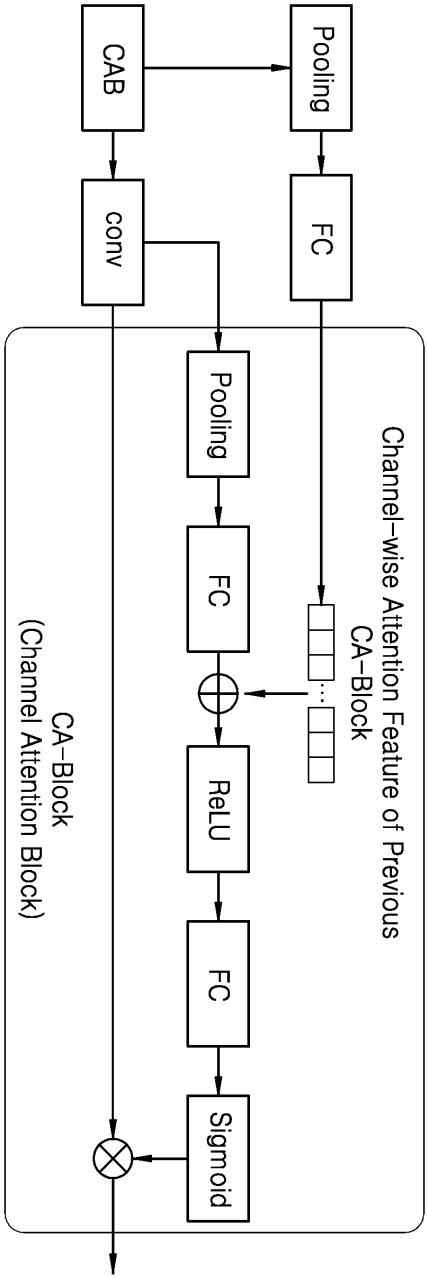
도면7


Multispectral Image
(128x128x15)
C1
CAB1
C2
C3
C4
CAB2
Skip Connection
Concatenation
C5
C6
CAB3
MP
ReLU
FC
Estimation
(36 x 1 vector)
Convolution+ReLU
ReLU
Channel Attention Block
Max Pooling Layer
Fully Connected Layer

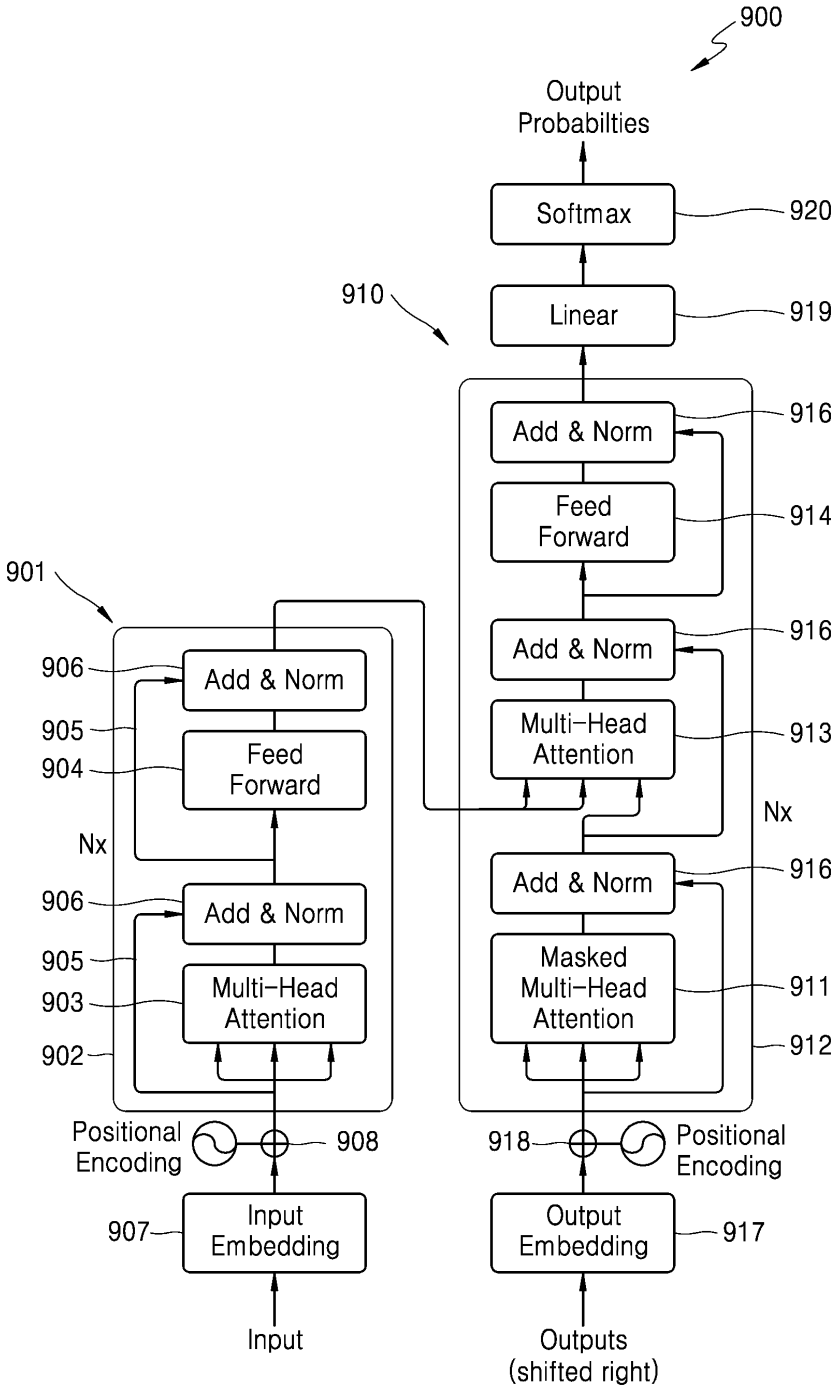
도면8a


Channel-wise Global Average
Pooling of input MSI image
Pooling
FC
ReLU
FC
Sigmoid
conv
CA-Block
(Channel Attention Block)

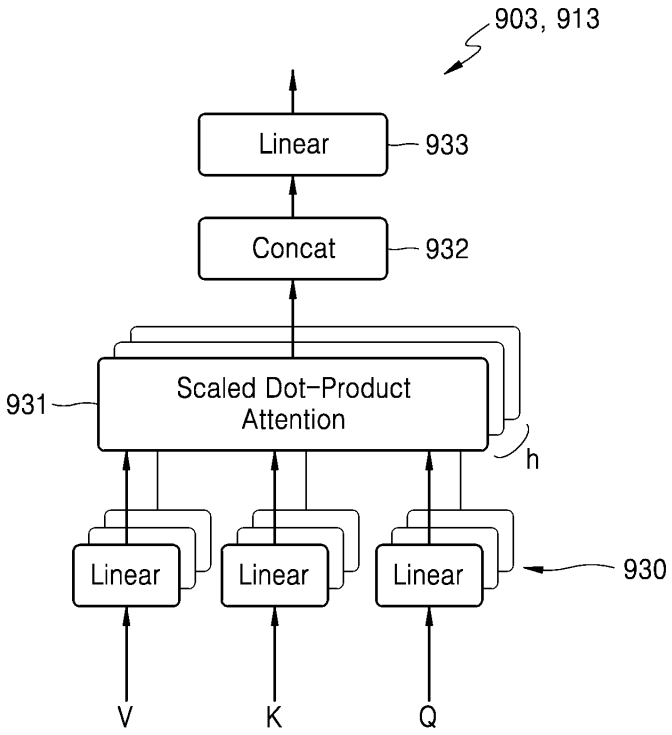
도면8b


CAB
Pooling
conv
FC
Pooling
FC
ReLU
FC
Sigmoid
CA-Block
(Channel Attention Block)
Channel-wise Attention Feature of Previous
CA-Block

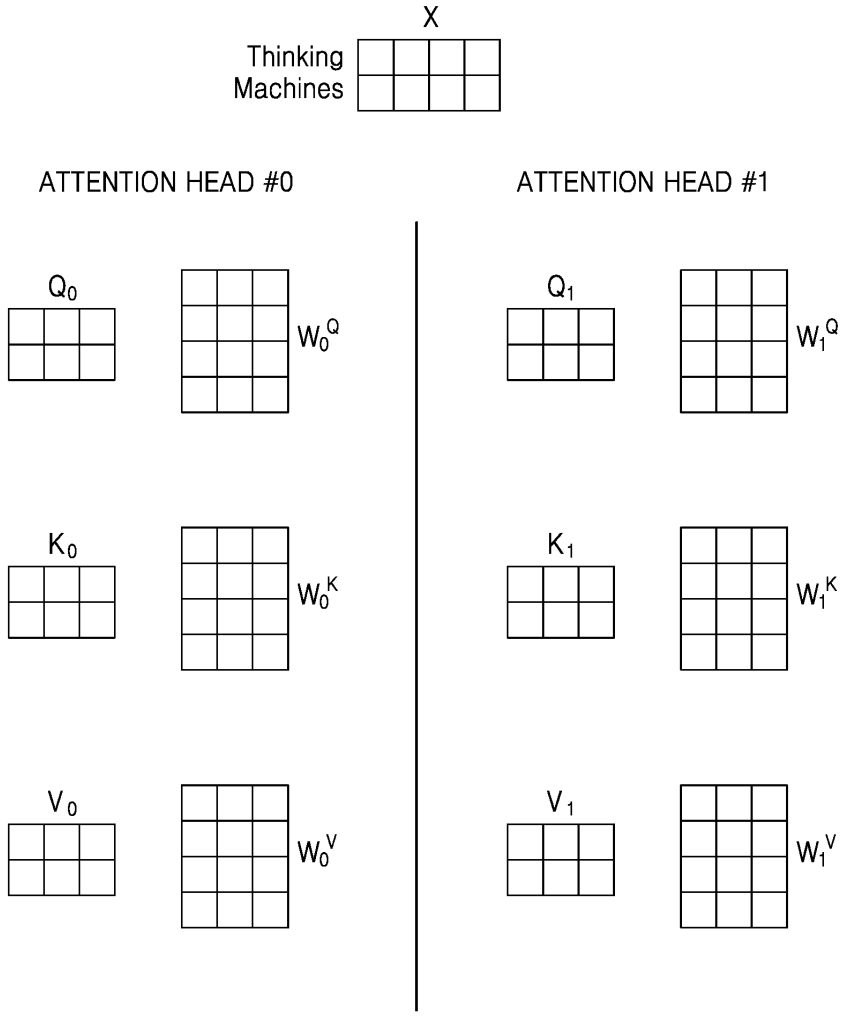
도면9a


900
Output Probabilties
Softmax
920
910
Linear
919
Add & Norm
916
Feed Forward
914
901
Add & Norm
916
906
Add & Norm
905
Multi-Head Attention
913
904
Feed Forward
Nx
Nx
Add & Norm
916
906
Add & Norm
Masked Multi-Head Attention
911
905
903
Multi-Head Attention
912
902
Positional Encoding
908
918
Positional Encoding
Input Embedding
907
Output Embedding
917
Input
Outputs (shifted right)

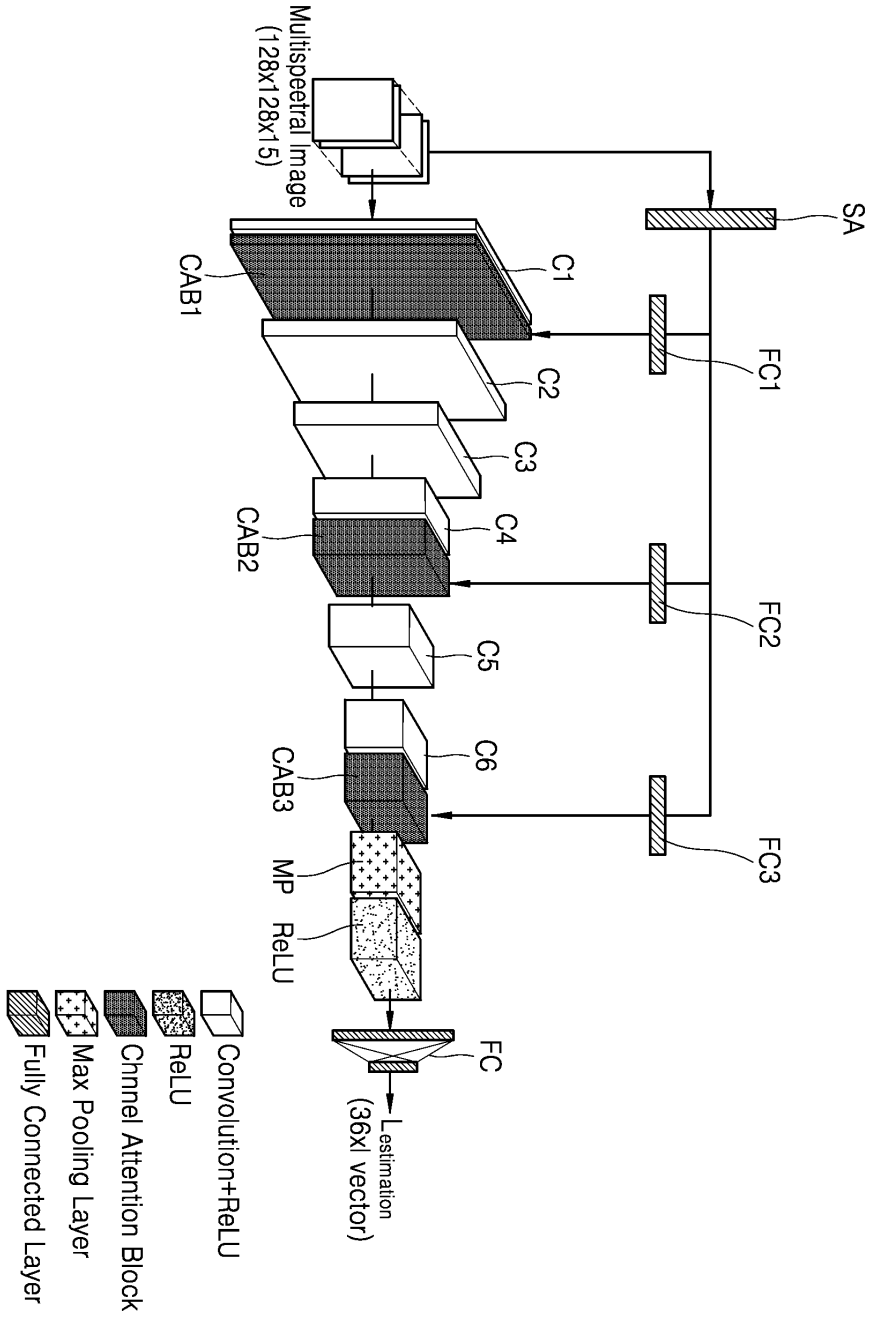
도면9b


903, 913
Linear
933
Concat
932
931
Scaled Dot-Product Attention
h
Linear
Linear
Linear
930
V
K
Q

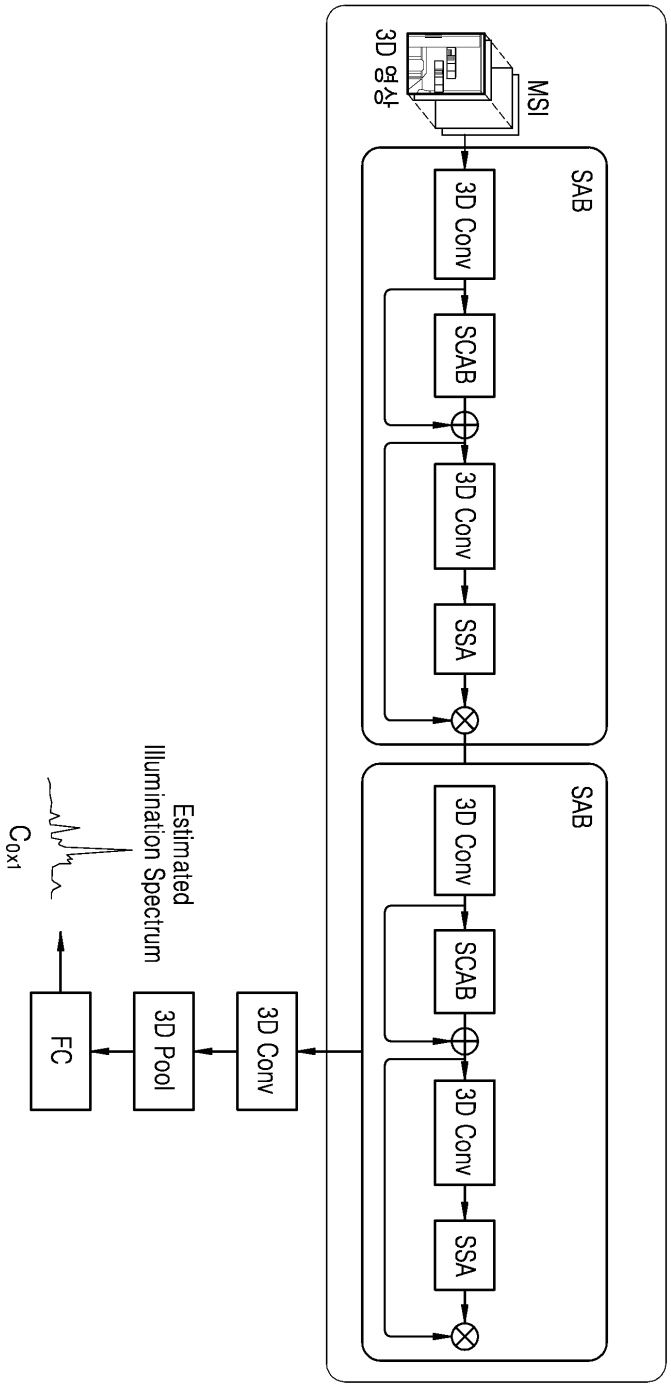
도면9c


X
Thinking
Machines
ATTENTION HEAD #0
ATTENTION HEAD #1
Q_0
W_0^Q
Q_1
W_1^Q
K_0
W_0^K
K_1
W_1^K
V_0
W_0^V
V_1
W_1^V

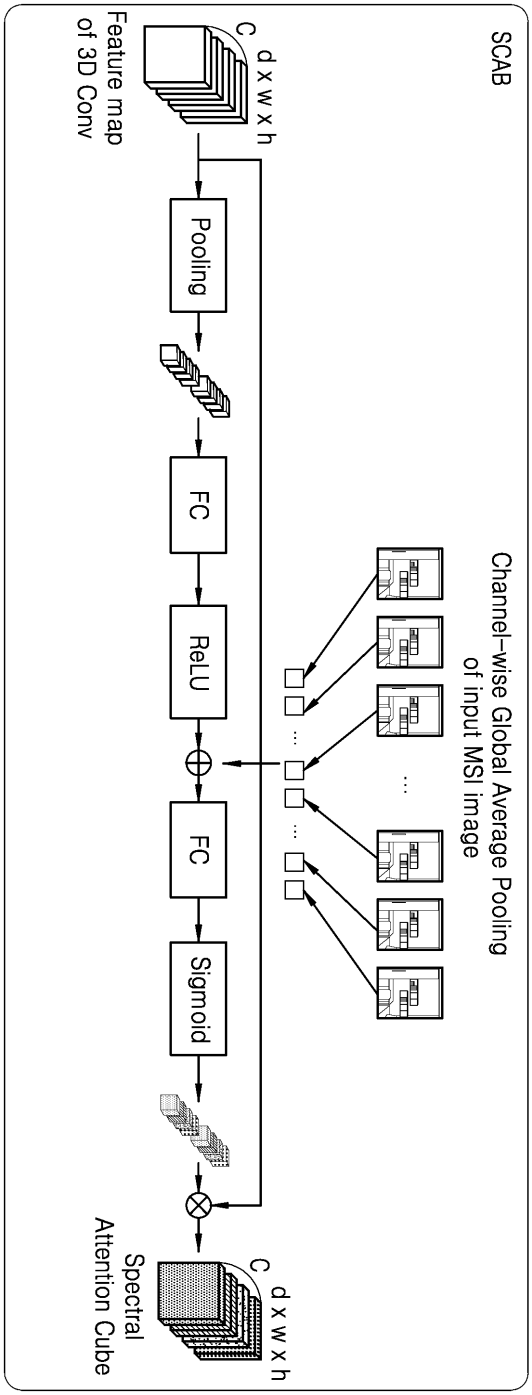
도면10


Multispectral Image
(128x128x15)
SA
C1
CAB1
FC1
C2
C3
C4
CAB2
FC2
C5
C6
CAB3
FC3
MP
ReLU
FC
Lestimation
(36x1 vector)
Convolution+ReLU
ReLU
Chnnel Attention Block
Max Pooling Layer
Fully Connected Layer

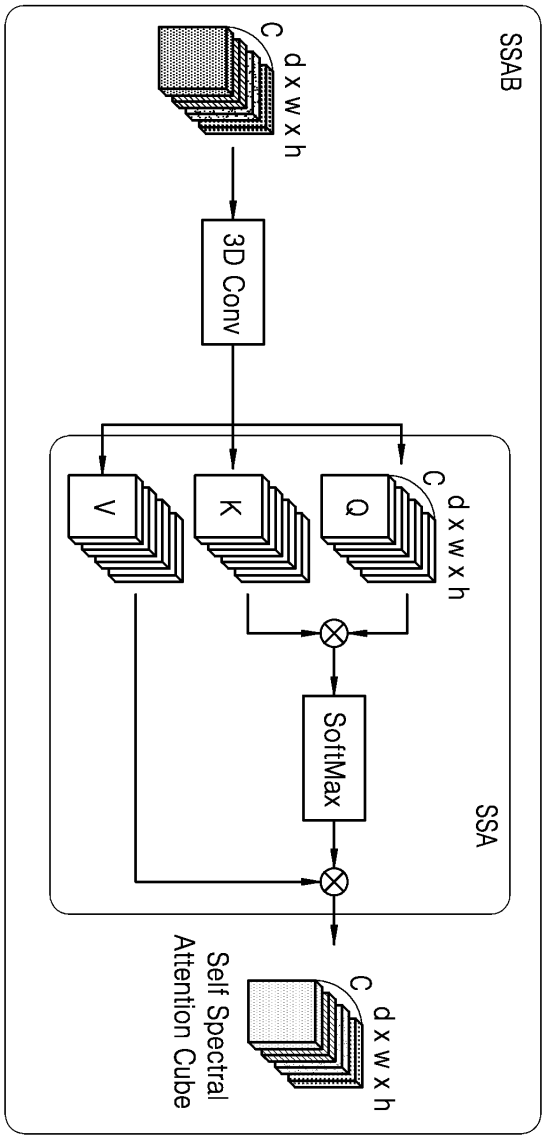
도면11a


3D 영상
MSI
SAB
3D Conv
SCAB
3D Conv
SSA
SAB
3D Conv
SCAB
3D Conv
SSA
3D Conv
3D Pool
FC
Estimated
Illumination Spectrum
C_{0x1}

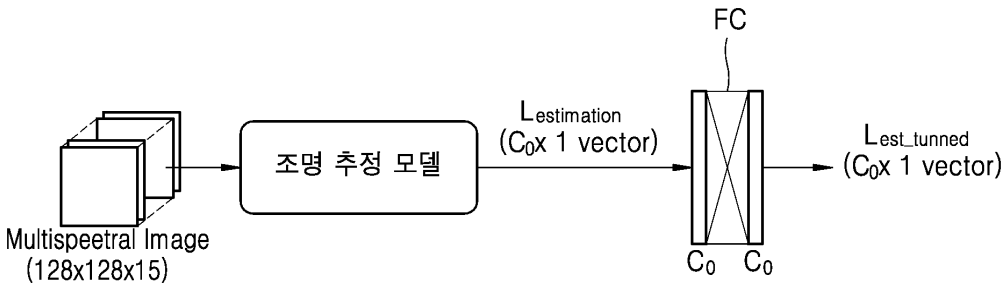
도면11b


SCAB
Feature map of 3D Conv
C
d x w x h
Pooling
FC
ReLU
FC
Sigmoid
Channel-wise Global Average Pooling of input MSI image
Spectral Attention Cube
C
d x w x h

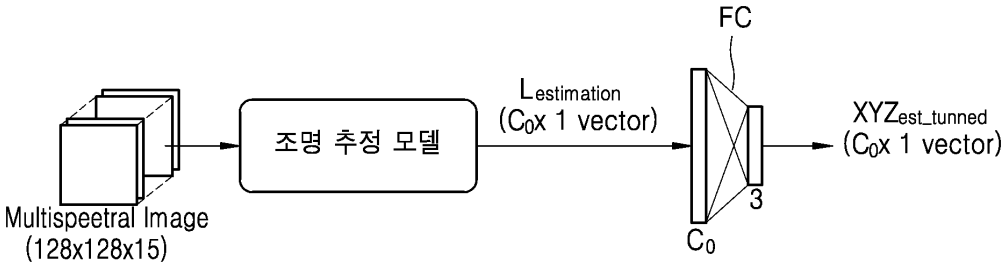
도면11c


SSAB
C
d x w x h
3D Conv
V
K
Q
C
d x w x h
SoftMax
SSA
Self Spectral
Attention Cube
C
d x w x h

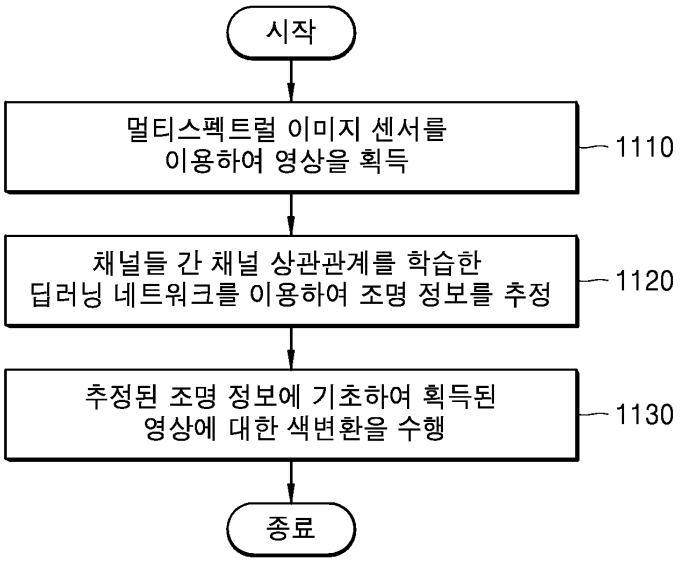
도면12a


FC
조명 추정 모델
Lestimation
(C0x 1 vector)
Lest_tunned
(C0x 1 vector)
Multispeetral Image
(128x128x15)
C0
C0

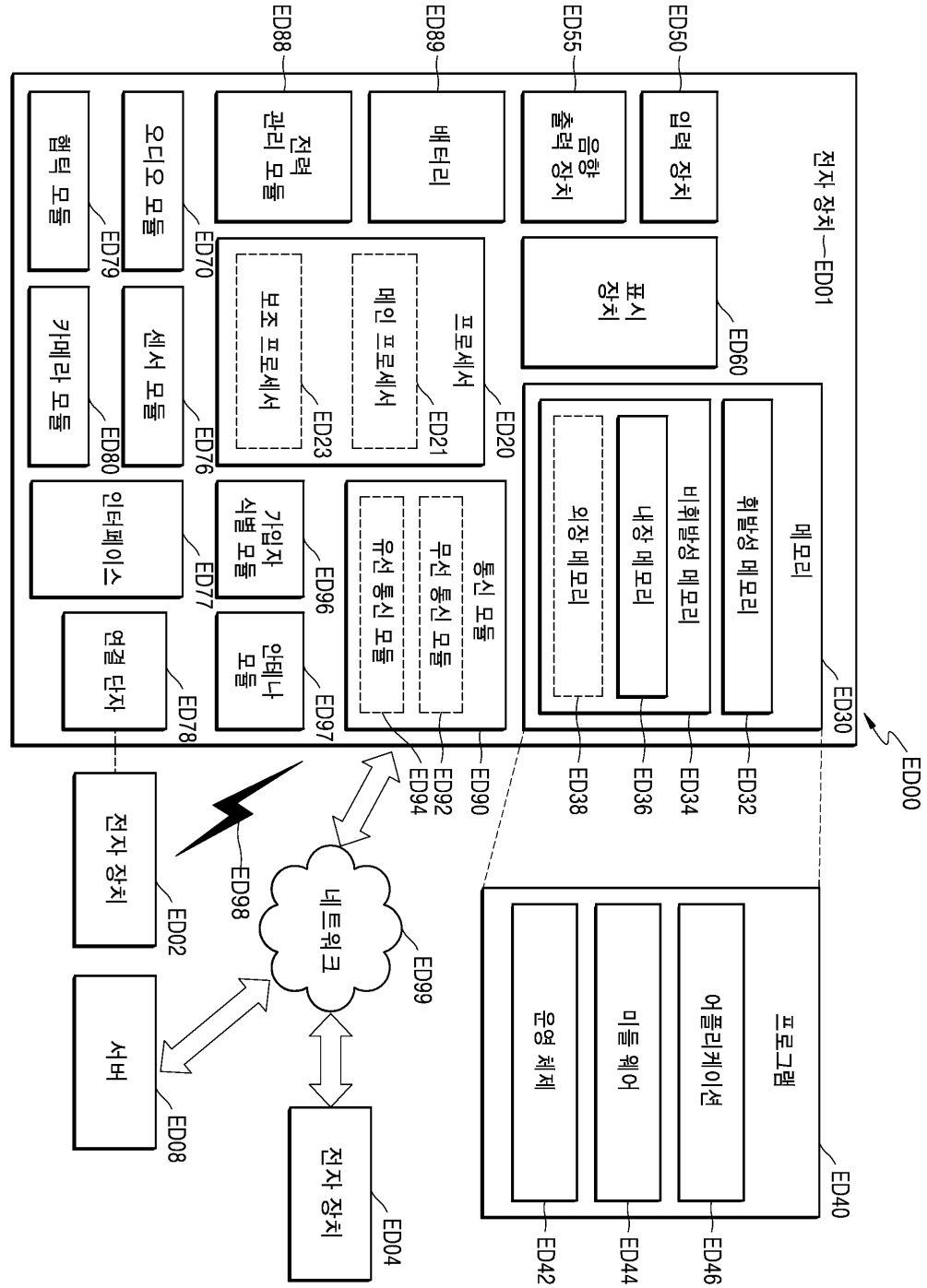
도면12b


Multispeetral Image
(128x128x15)
조명 추정 모델
Lestimation
(C0x 1 vector)
FC
3
C0
XYZest_tunned
(C0x 1 vector)

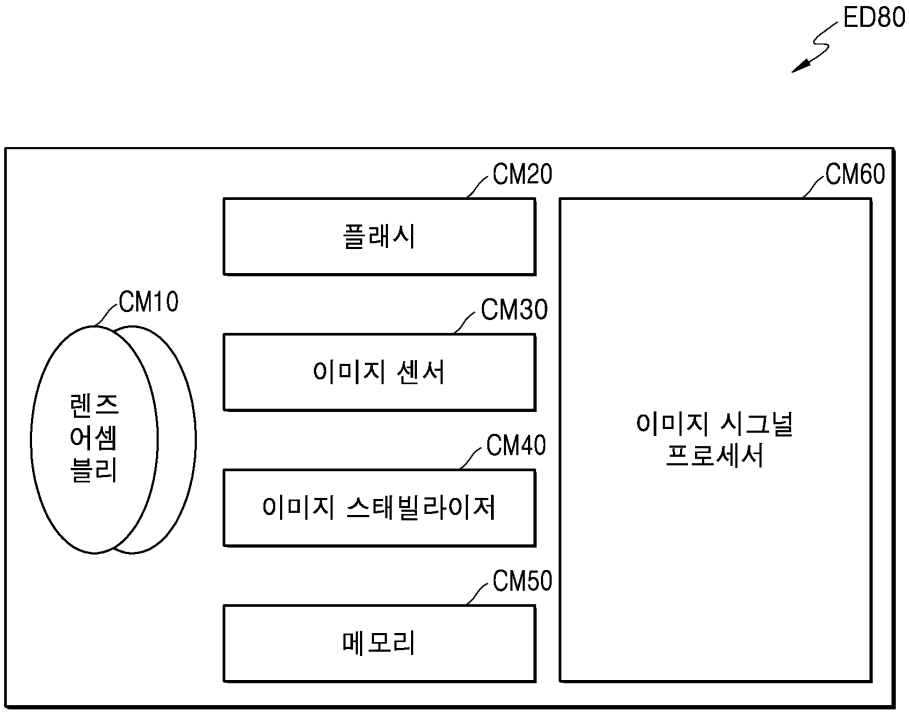
도면13


시작
멀티스펙트럴 이미지 센서를
이용하여 영상을 획득
1110
채널들 간 채널 상관관계를 학습한
딥러닝 네트워크를 이용하여 조명 정보를 추정
1120
추정된 조명 정보에 기초하여 획득된
영상에 대한 색변환을 수행
1130
종료

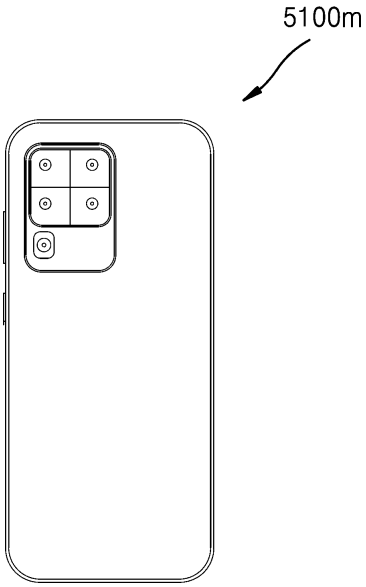
도면14


ED00
전자 장치—ED01
ED50
입력 장치
ED55
음향 출력 장치
ED88
전력 관리 모듈
ED89
배터리
ED60
표시 장치
ED30
메모리
휘발성 메모리
비휘발성 메모리
내장 메모리
외장 메모리
ED32
ED34
ED36
ED38
ED20
프로세서
메인 프로세서
보조 프로세서
ED21
ED23
ED90
통신 모듈
무선 통신 모듈
유선 통신 모듈
ED92
ED94
ED97
안테나 모듈
ED96
가입자 식별 모듈
ED78
연결 단자
ED77
인터페이스
ED70
오디오 모듈
ED76
센서 모듈
ED80
카메라 모듈
ED79
햅틱 모듈
ED40
프로그램
어플리케이션
미들 웨어
운영 체제
ED46
ED44
ED42
ED98
ED99
네트워크
ED02
전자 장치
ED08
서버
ED04
전자 장치

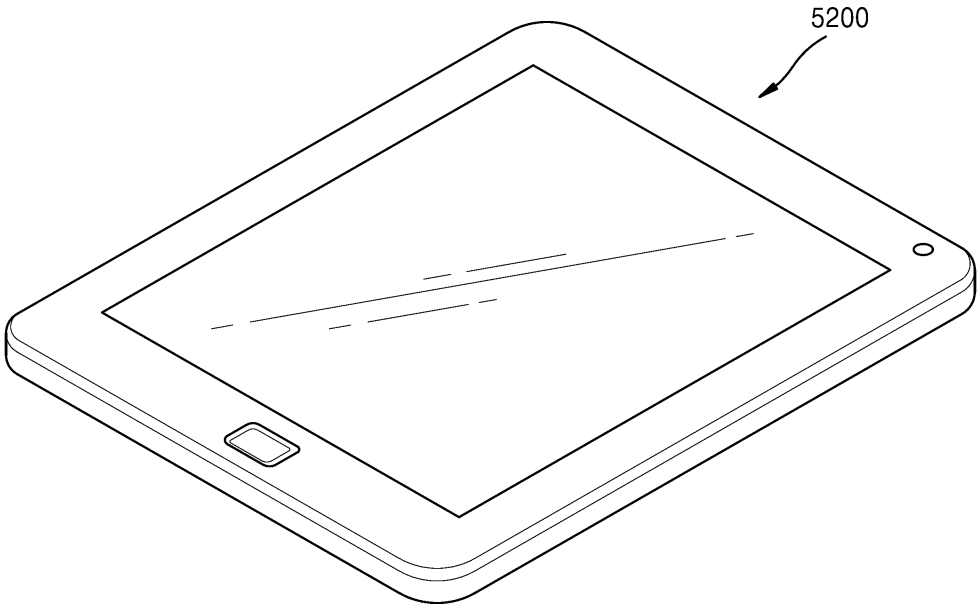
도면15


ED80
CM20
CM60
플래시
CM10
CM30
이미지 센서
렌즈
어셈
블리
이미지 시그널
프로세서
CM40
이미지 스태빌라이저
CM50
메모리

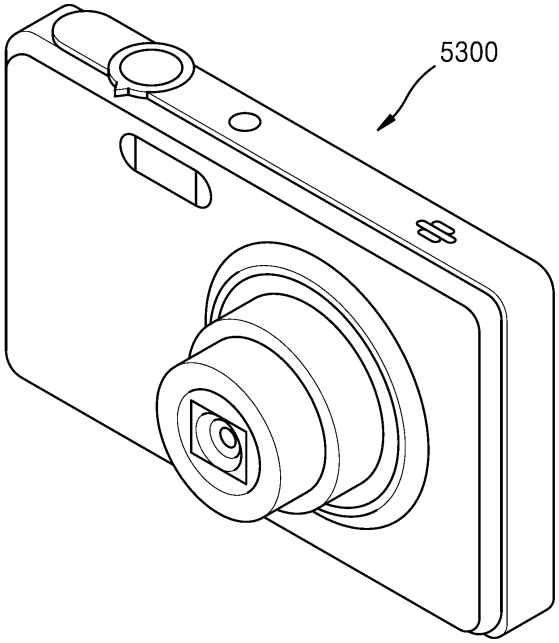
도면16


5100m

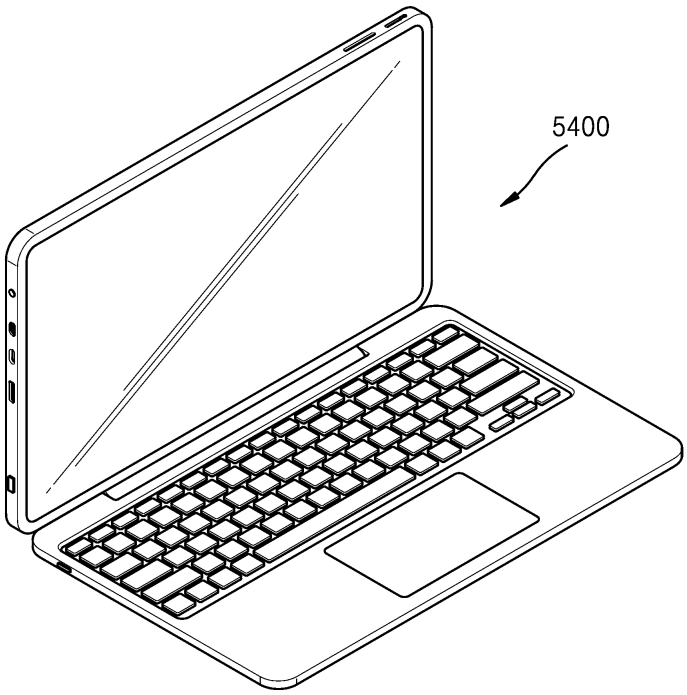
도면17



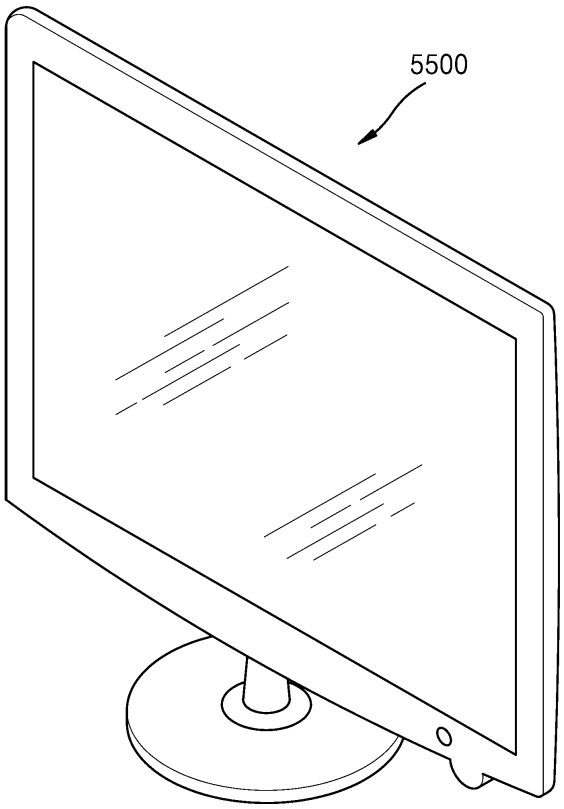
도면18



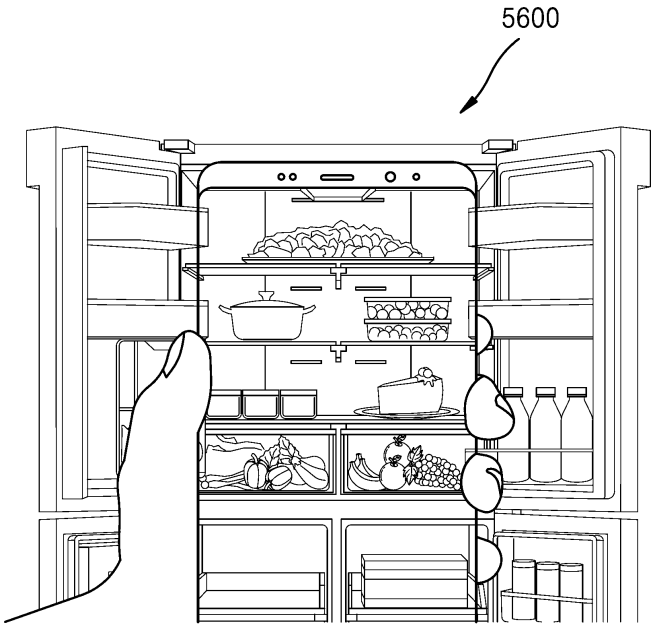
도면19



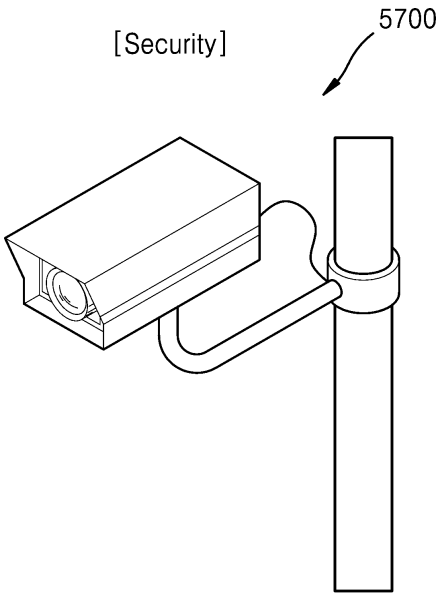
도면20



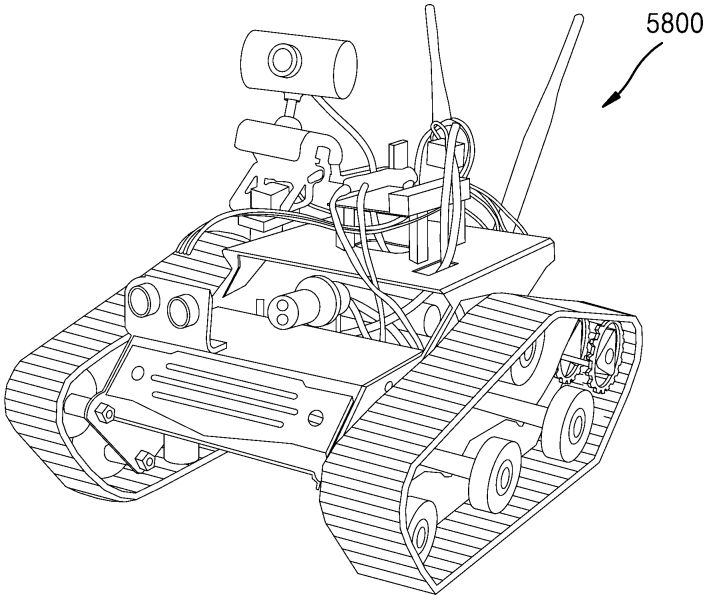
도면21



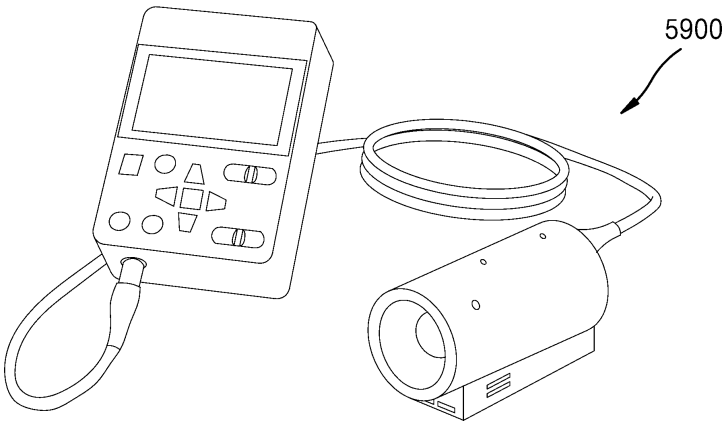
도면22



도면23



도면24



도면25

