(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0011559
(43) 공개일자 2022년01월28일

(51) 국제특허분류(Int. Cl.)
G06T 7/33 (2017.01) *G06N 3/04* (2006.01)
G06T 3/40 (2006.01) *G06T 7/246* (2017.01)
(52) CPC특허분류
G06T 7/33 (2017.01)
G06N 3/0454 (2013.01)
(21) 출원번호 10-2020-0161385
(22) 출원일자 2020년11월26일
심사청구일자 없음
(30) 우선권주장
1020200090560 2020년07월21일 대한민국(KR)

(71) 출원인
삼성전자주식회사
경기도 수원시 영통구 삼성로 129 (매탄동)
포항공과대학교 산학협력단
경상북도 포항시 남구 청암로 77 (지곡동)
(72) 발명자
한승주
서울특별시 동작구 사당로17길 52, 3동 602호 (사당동, 대림아파트)
조민수
경상북도 포항시 남구 청암로 77, 컴퓨터공학과(지곡동, 포항공과대학교)
(뒷면에 계속)
(74) 대리인
특허법인 무한

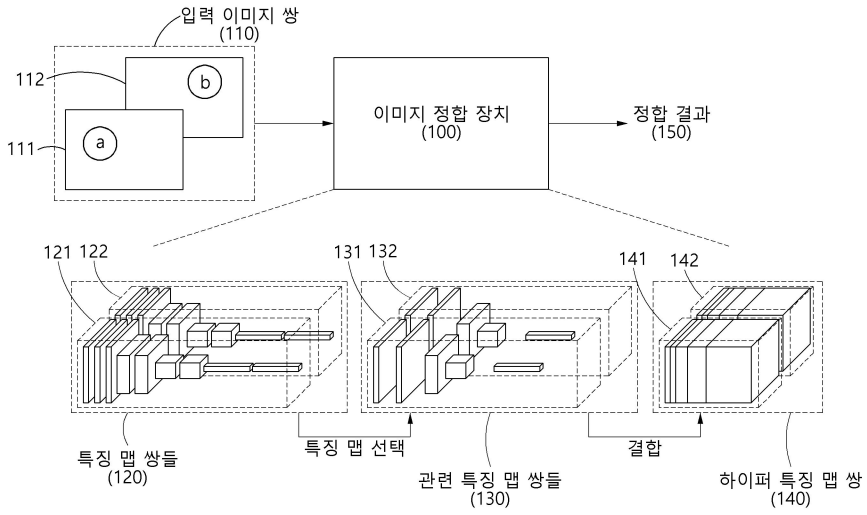
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 **동적 특징 선택에 기반한 이미지 정합 방법 및 장치**

(57) 요 약

동적 특징 선택에 기반한 이미지 정합 방법 및 장치가 제공된다. 일 실시예에 따르면, 그 방법은 복수의 특징 맵 쌍들의 각 특징 맵 쌍의 특성을 고려하여 복수의 특징 맵 쌍들 중에 일부의 특징 맵 쌍들을 선택하고, 선택된 일부의 특징 맵 쌍들에 기초하여 입력 이미지 쌍에 관한 정합 결과를 생성하는 단계들을 포함한다.

대 표 도



(52) CPC특허분류
G06T 3/4007 (2013.01)
G06T 7/246 (2017.01)
G06T 2207/20084 (2013.01)
(72) 발명자
민주홍
경상북도 포항시 남구 지곡로 260, 107동 506호(지곡동, 효자그린아파트)
이종민
전라북도 전주시 덕진구 백동로 47-13, 1층(인후동2가)
박창범
경기도 용인시 기흥구 흥덕중앙로105번길 40, 1501동 205호 (영덕동, 우남퍼스트빌 리젠트)

명 세 서

청구범위

청구항 1

입력 이미지 쌍의 입력에 따른 CNN(convolutional neural network)의 복수의 레이어들의 출력에 대응하는 복수의 특징 맵 쌍들을 획득하는 단계;

상기 복수의 특징 맵 쌍들의 각 특징 맵 쌍의 특성을 고려하여 상기 복수의 특징 맵 쌍들 중에 일부의 특징 맵 쌍들을 선택하는 단계;

상기 선택된 상기 일부의 특징 맵 쌍들에 기초하여 하이퍼 특징 맵 쌍을 생성하는 단계; 및

상기 하이퍼 특징 맵 쌍의 상관관계에 기초하여 상기 입력 이미지 쌍에 관한 정합 결과를 생성하는 단계

를 포함하는 이미지 정합 방법.

청구항 2

제1항에 있어서,

상기 CNN의 상기 복수의 레이어들은 상기 CNN의 적어도 하나의 중간 레이어를 포함하고,

상기 복수의 특징 맵 쌍들은 상기 CNN의 상기 적어도 하나의 중간 레이어에 의해 출력된 적어도 하나의 중간 특징 맵 쌍을 포함하는,

이미지 정합 방법.

청구항 3

제1항에 있어서,

상기 복수의 특징 맵 쌍들을 획득하는 상기 단계는

상기 입력 이미지 쌍의 제1 입력 이미지를 상기 CNN에 입력하여 상기 CNN의 상기 복수의 레이어들에서 출력된 제1 특징 맵들을 획득하는 단계;

상기 입력 이미지 쌍의 제2 입력 이미지를 상기 CNN에 입력하여 상기 CNN의 상기 복수의 레이어들에서 출력된 제2 특징 맵들을 획득하는 단계; 및

상기 CNN의 상기 복수의 레이어들에 따라 상기 제1 특징 맵들 및 상기 제2 특징 맵들을 페어링하여 상기 복수의 특징 맵 쌍들을 결정하는 단계

를 포함하는, 이미지 정합 방법.

청구항 4

제1항에 있어서,

상기 일부의 특징 맵 쌍들을 선택하는 상기 단계는

상기 복수의 특징 맵 쌍들의 각 특징 맵 쌍에 관한 적어도 하나의 뉴럴 네트워크의 출력에 기초하여 상기 복수의 특징 맵 쌍들 중에 상기 일부의 특징 맵 쌍들을 동적으로 선택하는 단계를 포함하는,

이미지 정합 방법.

청구항 5

제4항에 있어서,

상기 적어도 하나의 뉴럴 네트워크는

상기 복수의 특징 맵 쌍들의 각 특징 맵 쌍에 대응하는 채널 방향의 특징 벡터를 관련성 벡터로 인코딩하는 제1

뉴럴 네트워크; 및

채널 방향의 차원을 감소시켜서 상기 복수의 특징 맵 쌍들의 각 특징 맵 쌍을 저-차원의 특징 맵 쌍으로 변환하는 제2 뉴럴 네트워크

를 포함하는, 이미지 정합 방법.

청구항 6

제4항에 있어서,

상기 적어도 하나의 뉴럴 네트워크는

설정 파라미터에 의해 정해진 선택 비율에 기초하여 상기 복수의 특징 맵 쌍들 중에 상기 일부의 특징 맵 쌍들을 선택하도록 트레이닝되는,

이미지 정합 방법.

청구항 7

제1항에 있어서,

상기 일부의 특징 맵 쌍들을 선택하는 상기 단계는

상기 복수의 특징 맵 쌍들의 제1 특징 맵 쌍에 대응하는 제1 특징 벡터를 결정하는 단계;

상기 제1 특징 벡터의 입력에 따른 다층 퍼셉트론의 출력에 대응하는 제1 관련성 벡터를 획득하는 단계; 및

상기 제1 관련성 벡터의 값에 기초하여 상기 제1 특징 맵 쌍을 선택할지 결정하는 단계

를 포함하는, 이미지 정합 방법.

청구항 8

제1항에 있어서,

상기 일부의 특징 맵 쌍들을 선택하는 상기 단계는

상기 복수의 특징 맵 쌍들의 각 특징 맵 쌍의 특성을 고려하기 위해 상기 복수의 특징 맵 쌍들의 각 특징 맵 쌍을 병렬적으로 처리하여 상기 복수의 특징 맵 쌍들 중에 상기 일부의 특징 맵 쌍들을 선택하는 단계를 포함하는,

이미지 정합 방법.

청구항 9

제1항에 있어서,

상기 일부의 특징 맵 쌍들을 선택하는 상기 단계는

설정 파라미터에 의해 정해진 선택 비율에 기초하여 상기 복수의 특징 맵 쌍들 중에 상기 일부의 특징 맵 쌍들을 선택하는 단계를 포함하는,

이미지 정합 방법.

청구항 10

제9항에 있어서,

상기 설정 파라미터는 어플리케이션에 의해 요구되는 속도 및 정확도 중 적어도 하나에 기초하여 결정되는,

이미지 정합 방법.

청구항 11

제1항에 있어서,

상기 하이퍼 특징 맵 쌍을 생성하는 상기 단계는

상기 선택된 상기 일부의 특징 맵 쌍들에 기초한 업샘플링 및 연쇄화를 수행하여 상기 하이퍼 특징 맵 쌍을 생성하는 단계를 포함하는,

이미지 정합 방법.

청구항 12

제1항에 있어서,

상기 하이퍼 특징 맵 쌍을 생성하는 상기 단계는

상기 선택된 상기 일부의 특징 맵 쌍들에 대응하는 저-차원의 특징 맵 쌍들을 결합하여 상기 하이퍼 특징 맵 쌍을 생성하는 단계를 포함하는,

이미지 정합 방법.

청구항 13

제1항 내지 제12항 중 어느 한 항의 방법을 수행하는 명령어들을 포함하는 하나 이상의 프로그램을 저장한 컴퓨터 판독 가능 저장매체.

청구항 14

프로세서; 및

상기 프로세서에서 실행가능한 명령어들을 포함하는 메모리

를 포함하고,

상기 명령어들이 상기 프로세서에서 실행되면, 상기 프로세서는

입력 이미지 쌍의 입력에 따른 CNN(convolutional neural network)의 복수의 레이어들의 출력에 대응하는 복수의 특징 맵 쌍들을 획득하고,

상기 복수의 특징 맵 쌍들의 각 특징 맵 쌍의 특성을 고려하여 상기 복수의 특징 맵 쌍들 중에 일부의 특징 맵 쌍들을 선택하고,

상기 선택된 상기 일부의 특징 맵 쌍들에 기초하여 하이퍼 특징 맵 쌍을 생성하고,

상기 하이퍼 특징 맵 쌍의 상관관계에 기초하여 상기 입력 이미지 쌍에 관한 정합 결과를 생성하는,

이미지 정합 장치.

청구항 15

제14항에 있어서,

상기 CNN의 상기 복수의 레이어들은 상기 CNN의 적어도 하나의 중간 레이어를 포함하고,

상기 복수의 특징 맵 쌍들은 상기 CNN의 상기 적어도 하나의 중간 레이어에 의해 출력된 적어도 하나의 중간 특징 맵 쌍을 포함하는,

이미지 정합 장치.

청구항 16

제14항에 있어서,

상기 프로세서는

상기 복수의 특징 맵 쌍들의 각 특징 맵 쌍에 관한 적어도 하나의 뉴럴 네트워크의 출력에 기초하여 상기 복수의 특징 맵 쌍들 중에 상기 일부의 특징 맵 쌍들을 동적으로 선택하는,

이미지 정합 장치.

청구항 17

제14항에 있어서,

상기 프로세서는

설정 파라미터에 의해 정해진 선택 비율에 기초하여 상기 복수의 특징 맵 쌍들 중에 상기 일부의 특징 맵 쌍들을 선택하는,

이미지 정합 장치.

청구항 18

입력 이미지 쌍의 적어도 하나의 입력 이미지를 생성하는 카메라; 및

상기 입력 이미지 쌍의 입력에 따른 CNN(convolutional neural network)의 복수의 레이어들의 출력에 대응하는 복수의 특징 맵 쌍들을 획득하고, 상기 복수의 특징 맵 쌍들의 각 특징 맵 쌍의 특성을 고려하여 상기 복수의 특징 맵 쌍들 중에 일부의 특징 맵 쌍들을 선택하고, 상기 선택된 상기 일부의 특징 맵 쌍들에 기초하여 하이퍼 특징 맵 쌍을 생성하고, 상기 하이퍼 특징 맵 쌍의 상관관계에 기초하여 상기 입력 이미지 쌍에 관한 정합 결과를 생성하는, 프로세서

를 포함하는 전자 장치.

청구항 19

제18항에 있어서,

상기 CNN의 상기 복수의 레이어들은 상기 CNN의 적어도 하나의 중간 레이어를 포함하고,

상기 복수의 특징 맵 쌍들은 상기 CNN의 상기 적어도 하나의 중간 레이어에 의해 출력된 적어도 하나의 중간 특징 맵 쌍을 포함하는,

전자 장치.

청구항 20

제18항에 있어서,

상기 프로세서는

상기 복수의 특징 맵 쌍들의 각 특징 맵 쌍에 관한 적어도 하나의 뉴럴 네트워크의 출력에 기초하여 상기 복수의 특징 맵 쌍들 중에 상기 일부의 특징 맵 쌍들을 동적으로 선택하는,

전자 장치.

발명의 설명

기 술 분 야

[0001] 아래 실시예들은 동적 특징 선택에 기반한 이미지 정합 방법 및 장치에 관한 것이다.

배 경 기 술

[0002] 이미지 정합(image correspondence)을 푸는 선구적인 기술들은 HOG(histogram of oriented gradients)나 SIFT(scale-invariant feature transform)와 같은 핸드-크래프트(hand-crafted) 기반의 기술들을 이용한다. HOG와 SIFT 기반의 특징 벡터는 각 이미지의 국지적인 대상 영역을 그리드(grid)로 나누고, 그리드의 각 셀들의 기울기(gradient) 방향이나 크기에 대한 히스토그램 빈(histogram bin) 값들을 하나의 벡터로 연결하여 구할 수 있다. 그 이후에 나온 기술들은 CNN(convolutional neural network)을 이용하여 이미지 정합을 해결한다. 그 중에 대부분은 이미지 쌍의 국지 영역들 사이의 상관관계(correlation)를 예측하는 방식을 채택하고, 일부는 이미지 정합 문제를 이미지 정렬(image alignment) 문제로 해석하여 이미지 간의 광역 변환 파라미터(global transformation parameter)를 예측하는 방식을 이용한다.

발명의 내용

해결하려는 과제

과제의 해결 수단

[0003] 일 실시예에 따르면, 이미지 정합 방법은 입력 이미지 쌍의 입력에 따른 CNN(convolutional neural network)의 복수의 레이어들의 출력에 대응하는 복수의 특징 맵 쌍들을 획득하는 단계; 상기 복수의 특징 맵 쌍들의 각 특징 맵 쌍의 특성을 고려하여 상기 복수의 특징 맵 쌍들 중에 일부의 특징 맵 쌍들을 선택하는 단계; 상기 선택된 상기 일부의 특징 맵 쌍들에 기초하여 하이퍼 특징 맵 쌍을 생성하는 단계; 및 상기 하이퍼 특징 맵 쌍의 상관관계에 기초하여 상기 입력 이미지 쌍에 관한 정합 결과를 생성하는 단계를 포함한다.

[0004] 상기 CNN의 상기 복수의 레이어들은 상기 CNN의 적어도 하나의 중간 레이어를 포함하고, 상기 복수의 특징 맵 쌍들은 상기 CNN의 상기 적어도 하나의 중간 레이어에 의해 출력된 적어도 하나의 중간 특징 맵 쌍을 포함할 수 있다. 상기 복수의 특징 맵 쌍들을 획득하는 상기 단계는 상기 입력 이미지 쌍의 제1 입력 이미지를 상기 CNN에 입력하여 상기 CNN의 상기 복수의 레이어들에서 출력된 제1 특징 맵들을 획득하는 단계; 상기 입력 이미지 쌍의 제2 입력 이미지를 상기 CNN에 입력하여 상기 CNN의 상기 복수의 레이어들에서 출력된 제2 특징 맵들을 획득하는 단계; 및 상기 CNN의 상기 복수의 레이어들에 따라 상기 제1 특징 맵들 및 상기 제2 특징 맵들을 페어링하여 상기 복수의 특징 맵 쌍들을 결정하는 단계를 포함할 수 있다. 상기 일부의 특징 맵 쌍들을 선택하는 상기 단계는 상기 복수의 특징 맵 쌍들의 각 특징 맵 쌍에 관한 적어도 하나의 뉴럴 네트워크의 출력에 기초하여 상기 복수의 특징 맵 쌍들 중에 상기 일부의 특징 맵 쌍들을 동적으로 선택하는 단계를 포함할 수 있다. 상기 적어도 하나의 뉴럴 네트워크는 상기 복수의 특징 맵 쌍들의 각 특징 맵 쌍에 대응하는 채널 방향의 특징 벡터를 관련성 벡터로 인코딩하는 제1 뉴럴 네트워크; 및 채널 방향의 차원을 감소시켜서 상기 복수의 특징 맵 쌍들의 각 특징 맵 쌍을 저-차원의 특징 맵 쌍으로 변환하는 제2 뉴럴 네트워크를 포함할 수 있다.

[0005] 상기 적어도 하나의 뉴럴 네트워크는 설정 파라미터에 의해 정해진 선택 비율에 기초하여 상기 복수의 특징 맵 쌍들 중에 상기 일부의 특징 맵 쌍들을 선택하도록 트레이닝될 수 있다. 상기 일부의 특징 맵 쌍들을 선택하는 상기 단계는 상기 복수의 특징 맵 쌍들의 제1 특징 맵 쌍에 대응하는 제1 특징 벡터를 결정하는 단계; 상기 제1 특징 벡터의 입력에 따른 다층 퍼셉트론의 출력에 대응하는 제1 관련성 벡터를 획득하는 단계; 및 상기 제1 관련성 벡터의 값에 기초하여 상기 제1 특징 맵 쌍을 선택할지 결정하는 단계를 포함할 수 있다.

[0006] 상기 일부의 특징 맵 쌍들을 선택하는 상기 단계는 상기 복수의 특징 맵 쌍들의 각 특징 맵 쌍의 특성을 고려하기 위해 상기 복수의 특징 맵 쌍들의 각 특징 맵 쌍을 병렬적으로 처리하여 상기 복수의 특징 맵 쌍들 중에 상기 일부의 특징 맵 쌍들을 선택하는 단계를 포함할 수 있다. 상기 일부의 특징 맵 쌍들을 선택하는 상기 단계는 설정 파라미터에 의해 정해진 선택 비율에 기초하여 상기 복수의 특징 맵 쌍들 중에 상기 일부의 특징 맵 쌍들을 선택하는 단계를 포함할 수 있다. 상기 설정 파라미터는 어플리케이션에 의해 요구되는 속도 및 정확도 중 적어도 하나에 기초하여 결정될 수 있다.

[0007] 상기 하이퍼 특징 맵 쌍을 생성하는 상기 단계는 상기 선택된 상기 일부의 특징 맵 쌍들에 기초한 업샘플링 및 연쇄화를 수행하여 상기 하이퍼 특징 맵 쌍을 생성하는 단계를 포함할 수 있다. 상기 하이퍼 특징 맵 쌍을 생성하는 상기 단계는 상기 선택된 상기 일부의 특징 맵 쌍들에 대응하는 저-차원의 특징 맵 쌍들을 결합하여 상기 하이퍼 특징 맵 쌍을 생성하는 단계를 포함할 수 있다.

[0008] 일 실시예에 따르면, 이미지 정합 장치는 프로세서; 및 상기 프로세서에서 실행가능한 명령어들을 포함하는 메모리를 포함하고, 상기 명령어들이 상기 프로세서에서 실행되면, 상기 프로세서는 입력 이미지 쌍의 입력에 따른 CNN(convolutional neural network)의 복수의 레이어들의 출력에 대응하는 복수의 특징 맵 쌍들을 획득하고, 상기 복수의 특징 맵 쌍들의 각 특징 맵 쌍의 특성을 고려하여 상기 복수의 특징 맵 쌍들 중에 일부의 특징 맵 쌍들을 선택하고, 상기 선택된 상기 일부의 특징 맵 쌍들에 기초하여 하이퍼 특징 맵 쌍을 생성하고, 상기 하이퍼 특징 맵 쌍의 상관관계에 기초하여 상기 입력 이미지 쌍에 관한 정합 결과를 생성한다.

[0009] 일 실시예에 따르면, 전자 장치는 입력 이미지 쌍의 적어도 하나의 입력 이미지를 생성하는 카메라; 및 상기 입력 이미지 쌍의 입력에 따른 CNN(convolutional neural network)의 복수의 레이어들의 출력에 대응하는 복수의 특징 맵 쌍들을 획득하고, 상기 복수의 특징 맵 쌍들의 각 특징 맵 쌍의 특성을 고려하여 상기 복수의 특징 맵

쌍들 중에 일부의 특징 맵 쌍들을 선택하고, 상기 선택된 상기 일부의 특징 맵 쌍들에 기초하여 하이퍼 특징 맵 쌍을 생성하고, 상기 하이퍼 특징 맵 쌍의 상관관계에 기초하여 상기 입력 이미지 쌍에 관한 정합 결과를 생성하는, 프로세서를 포함한다.

도면의 간단한 설명

[0010] 도 1은 일 실시예에 따른 입력 이미지 쌍에 관한 전반적인 이미지 정합 과정을 나타낸다.

도 2는 일 실시예에 따른 특징 맵들의 생성 과정을 나타낸다.

도 3은 일 실시예에 따른 동적 특징 선택 동작을 나타낸다.

도 4는 일 실시예에 따른 동적 특징 선택을 위한 뉴럴 네트워크의 추론 및 트레이닝 과정을 나타낸다.

도 5는 일 실시예에 따른 하이퍼 특징 맵의 생성 및 정합 과정을 나타낸다.

도 6은 일 실시예에 따른 뉴럴 네트워크의 트레이닝을 위한 전반적인 구조를 나타낸다.

도 7은 일 실시예에 따른 강한 지도를 통한 트레이닝 과정을 나타낸다.

도 8은 일 실시예에 따른 약한 지도를 통한 트레이닝 과정을 나타낸다.

도 9는 일 실시예에 따른 이미지 정합 장치의 개략적인 구성을 나타낸다.

도 10은 일 실시예에 따른 이미지 정합 방법을 개략적으로 나타낸다.

도 11은 일 실시예에 따른 이미지 정합 장치와 관련된 전자 장치를 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0011] 이하에서, 첨부된 도면을 참조하여 실시예들을 상세하게 설명한다. 그러나, 실시예들에는 다양한 변경이 가해질 수 있어서 특허출원의 권리 범위가 이러한 실시예들에 의해 제한되거나 한정되는 것은 아니다. 실시예들에 대한 모든 변경, 균등물 내지 대체물이 권리 범위에 포함되는 것으로 이해되어야 한다.

[0012] 실시예에서 사용한 용어는 단지 설명을 목적으로 사용된 것으로, 한정하려는 의도로 해석되어서는 안 된다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 명세서 상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0013] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥 상 가지는 의미와 일치하는 의미를 가지는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0014] 또한, 첨부 도면을 참조하여 설명함에 있어, 도면 부호에 관계없이 동일한 구성 요소는 동일한 참조부호를 부여하고 이에 대한 중복되는 설명은 생략하기로 한다. 실시예를 설명함에 있어서 관련된 공지 기술에 대한 구체적인 설명이 실시예의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다.

[0015] 또한, 실시 예의 구성 요소를 설명하는 데 있어서, 제1, 제2, A, B, (a), (b) 등의 용어를 사용할 수 있다. 이러한 용어는 그 구성 요소를 다른 구성 요소와 구별하기 위한 것일 뿐, 그 용어에 의해 해당 구성 요소의 본질이나 차례 또는 순서 등이 한정되지 않는다. 어떤 구성 요소가 다른 구성요소에 "연결", "결합" 또는 "접속"된다고 기재된 경우, 그 구성 요소는 그 다른 구성요소에 직접적으로 연결되거나 접속될 수 있지만, 각 구성 요소 사이에 또 다른 구성 요소가 "연결", "결합" 또는 "접속"될 수도 있다고 이해되어야 할 것이다.

[0016] 어느 하나의 실시 예에 포함된 구성요소와, 공통적인 기능을 포함하는 구성요소는, 다른 실시 예에서 동일한 명칭을 사용하여 설명하기로 한다. 반대되는 기재가 없는 이상, 어느 하나의 실시 예에 기재한 설명은 다른 실시 예에도 적용될 수 있으며, 중복되는 범위에서 구체적인 설명은 생략하기로 한다.

[0017] 도 1은 일 실시예에 따른 입력 이미지 쌍에 관한 전반적인 이미지 정합 과정을 나타낸다. 도 1을 참조하면, 이미지 정합 장치(image correspondence apparatus, 100)는 입력 이미지 쌍(110)에 관한 이미지 정합을 수행하여

정합 결과(150)를 출력할 수 있다. 입력 이미지 쌍(110)은 소스 입력 이미지(111) 및 타겟 입력 이미지(112)를 포함할 수 있고, 정합 결과(150)는 소스 이미지(111) 내의 소스 객체(a)와 타겟 이미지(112) 내의 타겟 객체(b) 간의 정합 관계를 나타낼 수 있다.

[0018] 이미지 정합 장치(100)는 의미론적 정합(semantic correspondence)을 수행할 수 있다. 이미지 정합 장치(100)의 의미론적 이미지 정합을 통해 동한 종류의 카테고리 내의 다른 종류의 객체들이 픽셀 단위로 매칭될 수 있다. 예를 들어, 소스 객체(a)와 타겟 객체(b)가 다른 세부 종류의 고양이에 해당하는 경우, 혹은 소스 객체(a)와 타겟 객체(b)가 다른 기종의 비행기에 해당하는 경우, 정합 결과(150)는 소스 객체(a)와 타겟 객체(b)가 서로 매칭됨을 나타낼 수 있다. 정합 결과(150)는 객체 검출, 객체 추적, 객체 인식과 같은 다양한 응용에 활용될 수 있다.

[0019] 소스 입력 이미지(111) 및 타겟 입력 이미지(112)는 하나의 프레임을 포함하는 스틸 이미지에 해당하거나, 또는 복수의 프레임들을 포함하는 비디오의 어느 하나의 프레임에 해당할 수 있다. 소스 입력 이미지(111) 및 타겟 입력 이미지(112)는 서로 동일한 유형의 이미지이거나 서로 다른 유형의 이미지일 수 있다. 예를 들어, 소스 입력 이미지(111) 및 타겟 입력 이미지(112)는 둘 다 스틸 이미지에 해당할 수 있다. 혹은, 소스 입력 이미지(111)는 스틸 이미지에 해당하고, 타겟 입력 이미지(112)는 비디오의 어느 하나의 프레임에 해당할 수 있다. 아래에서, 소스 입력 이미지(111) 및 타겟 입력 이미지(112) 중에 어느 하나는 제1 입력 이미지로 지칭될 수 있고, 나머지 하나는 제2 입력 이미지로 지칭될 수 있다.

[0020] 이미지 정합은 객체 인식, 이미지 검색, 3차원 재건, 모션 추정, 깊이 추정, 동작 인식과 같은 수많은 응용을 위한 전제로서 이미지를 이해하는데 중요한 요소이다. 이때, 특징 표현은 이미지 정합에서 중요한 역할을 하며, 최근의 이미지 정합은 CNN(convolutional neural network)에 의존적인 추세이다. 뉴럴 네트워크 기술의 발전으로 인해 이미지들 간의 정합을 설립하기 위한 강력한 특징 표현을 학습하는데 상당한 진전이 있었고, 현재 사실상의 표준은 학습 가능한 아키텍처에서 CNN의 출력을 특징 표현으로 사용하는 것이다. 이러한 CNN 모델은 일반적으로 특정 레이어의 특징(예: 마지막 레이어의 출력)을 사용하고, 매칭될 이미지의 특성에 관계없이 이를 준수한다는 점에서 정적(static)이라고 할 수 있다.

[0021] 이미지 정합 장치(100)는 주어진 이미지의 특성을 고려하여 이미지 정합에 관련된 레이어 혹은, 해당 레이어에서 출력된 특징을 동적(dynamic)으로 선택하여 이미지 정합을 수행할 수 있다. 보다 구체적으로, 이미지 정합 장치(100)는 CNN을 이용하여 입력 이미지 쌍(110)에 대응하는 특징 맵 쌍들(120)을 획득할 수 있다. 예를 들어, 이미지 정합 장치(100)는 소스 입력 이미지(111) 및 타겟 입력 이미지(112)를 CNN에 순차적으로 입력하고, CNN의 출력에 대응하여 소스 특징 맵들(121) 및 타겟 특징 맵들(122)을 순차적으로 획득할 수 있다. 특징 맵 쌍들(120)은 CNN의 중간 레이어를 포함하는 복수의 레이어들의 출력에 대응할 수 있다. 특징 맵 쌍들(120)의 생성 과정은 추후 도 2를 참조하여 보다 상세히 설명한다.

[0022] 이미지 정합 장치(100)는 특징 맵 쌍들(120)의 각 특징 맵 쌍의 특성을 고려하여 특징 맵 쌍들(120) 중에 관련 특징 맵 쌍들(130)을 동적으로 선택할 수 있다. 예를 들어, 관련 특징 맵 쌍들(130)은 특징 맵 쌍들(120) 중에 이미지 정합과 관련된 일부에 해당할 수 있다. 소스 관련 특징 맵들(131)은 소스 특징 맵들(121)으로부터 선택된 것이고, 타겟 관련 특징 맵들(132)은 타겟 특징 맵들(122)으로부터 선택된 것이다. 이미지 정합 장치(100)는 적어도 하나의 뉴럴 네트워크를 이용하여 특징 맵 쌍들(120) 중에 관련 특징 맵 쌍들(130)을 선택할 수 있다. 적어도 하나의 뉴럴 네트워크는 입력 이미지 쌍(110)의 특성, 예를 들어 소스 특징 맵들(121)과 타겟 특징 맵들(122)의 페어-와이즈(pair-wise) 관계에 기초하여 특징 맵 쌍들(120) 중에 관련 특징 맵 쌍들(130)을 선택하도록 미리 트레이닝될 수 있다.

[0023] 뉴럴 네트워크는 딥 러닝에 기반하여 트레이닝된 후, 비선형적 관계에 있는 입력 데이터 및 출력 데이터를 서로 매핑함으로써 트레이닝 목적에 맞는 추론(inference)을 수행해낼 수 있다. 딥 러닝은 빅 데이터 세트로부터 이미지 인식 또는 음성 인식과 같은 문제를 해결하기 위한 기계 학습 기법이다. 딥 러닝은 준비된 트레이닝 데이터를 이용하여 뉴럴 네트워크를 트레이닝하면서 에너지가 최소화되는 지점을 찾아가는 최적화 문제 풀이 과정으로 이해될 수 있다.

[0024] 딥 러닝의 지도식(supervised) 또는 비지도식(unsupervised) 학습을 통해 뉴럴 네트워크의 구조, 혹은 모델에 대응하는 웨이트가 구해질 수 있고, 이러한 웨이트를 통해 입력 데이터 및 출력 데이터가 서로 매핑될 수 있다. 뉴럴 네트워크의 폭과 깊이가 충분히 크면 임의의 함수를 구현할 수 있을 만큼의 용량(capacity)을 가질 수 있다. 뉴럴 네트워크가 적절한 트레이닝 과정을 통해 충분히 많은 양의 트레이닝 데이터를 학습하면 최적의 성능을 달성할 수 있다.

[0025] 뉴럴 네트워크는 '미리' 트레이닝된 것으로 표현될 수 있는데, 여기서 '미리'는 뉴럴 네트워크가 '시작'되기 전을 나타낼 수 있다. 뉴럴 네트워크가 '시작'되었다는 것은 뉴럴 네트워크가 추론을 위한 준비가 된 것을 의미할 수 있다. 예를 들어, 뉴럴 네트워크가 '시작'된 것은 뉴럴 네트워크가 메모리에 로드된 것, 혹은 뉴럴 네트워크가 메모리에 로드된 이후 뉴럴 네트워크에 추론을 위한 입력 데이터가 입력된 것을 포함할 수 있다.

[0026] 이미지 정합 장치(100)는 관련 특징 맵 쌍들(130)에 기초하여 하이퍼 특징 맵 쌍(140)을 생성할 수 있다. 예를 들어, 이미지 정합 장치(100)는 소스 관련 특징 맵들(131)을 결합하여 소스 하이퍼 특징 맵(141)을 생성하고, 타겟 관련 특징 맵들(132)을 결합하여 타겟 하이퍼 특징 맵(142)을 생성할 수 있다. 혹은, 정합 과정의 컴퓨팅 부담을 완화(예: 추론 시간의 단축)하기 위해, 이미지 정합 장치(100)는 소스 관련 특징 맵들(131) 및 타겟 관련 특징 맵들(132) 대신 이들의 저-차원(low-dimension) 버전을 이용하여 하이퍼 특징 맵 쌍(140)을 생성할 수도 있다.

[0027] 하이퍼 특징 맵은 대응 이미지의 픽셀을 하이퍼 픽셀로 표현한다. 하이퍼 특징 맵은 하이퍼 이미지로 지칭될 수도 있다. CNN의 다중 레이어들을 통해 어느 이미지에 대응하는 특징 맵들이 획득되고, 이들이 예를 들어 업샘플링(upsampling) 및 연쇄화(concatenation)를 통해 결합되어, 해당 이미지에 대응하는 하이퍼 이미지가 생성될 수 있다. 이때, 이미지의 각 픽셀은 하이퍼 이미지의 어느 하이퍼 픽셀에 대응할 수 있다. CNN의 개입에 따라 하이퍼 픽셀은 대응 이미지를 더욱 세밀하게 분석할 수 있는 정보를 제공할 수 있다. 하이퍼 이미지의 하이퍼 픽셀과 이미지의 픽셀 간에는 일대일 또는 일대다의 관계가 성립될 수 있다. 예를 들어, 하이퍼 이미지와 대응 이미지의 공간 해상도(spatial resolution)가 동일하다면 하이퍼 픽셀과 이미지 픽셀이 서로 일대일로 매칭될 수 있고, 하이퍼 이미지와 대응 이미지의 공간 해상도가 다르다면 다른 만큼의 일대다의 매칭 관계가 형성될 수 있다.

[0028] 이미지 정합 장치(100)는 소스 하이퍼 특징 맵(141)과 하이퍼 특징 맵(142) 간의 상관관계(correlation)를 계산하여 정합 결과(150)를 생성할 수 있다. 어느 이미지 픽셀에 대응하는 하이퍼 픽셀은 해당 이미지 픽셀을 하이퍼 이미지 내에서 해당 이미지 픽셀의 공간 위치(spatial position)에 대응하는 벡터로 표현하고, 해당 벡터는 CNN의 다중 레이어의 출력들의 결합을 통해 생성되므로, 하이퍼 이미지를 통해 이미지 정합 작업이 보다 정밀하게 수행될 수 있다. 이때, 이미지 정합 장치(100)는 하이퍼 이미지를 생성하는데 CNN의 다중 레이어에서 출력된 특징 맵들을 모두 이용하는 대신 동적 특징 선택을 통해 일부의 관련 특징 맵들만 이용하여, 이미지 정합의 정확도 및/또는 속도를 향상시키거나, 정확도와 속도 사이의 균형을 맞출 수 있다.

[0029] 도 2는 일 실시예에 따른 특징 맵들의 생성 과정을 나타낸다. 도 2를 참조하면, 이미지 정합 장치는 CNN(200)을 이용하여 입력 이미지(205)에 대응하는 특징 맵 세트(220)를 생성할 수 있다. CNN(200)은 복수의 레이어들을 포함할 수 있다. 각 레이어는 컨볼루셔널 레이어 및 풀링(pooling) 레이어를 포함할 수 있고, 이들을 통해 특징 맵을 출력할 수 있다. CNN(200)은 ImageNet과 같은 대규모의 트레이닝 데이터 세트를 통해 특징 추출을 목적으로 미리 트레이닝될 수 있다.

[0030] 예를 들어, CNN(200)은 제1 레이어(201), 제l 레이어(202), 및 제L 레이어(203)를 포함할 수 있다. 제1 레이어(201)는 입력 이미지(205)에서 특징을 추출하여 제1 특징 맵(211)을 출력할 수 있다. 제l 레이어(202)는 제l-1 레이어(미도시)의 출력 특징 맵에서 특징을 추출하여 제l 특징 맵(212)을 출력할 수 있고, 제L 레이어(203)는 제L-1 레이어(미도시)의 출력 특징 맵에서 특징을 추출하여 제L 특징 맵(213)을 출력할 수 있다. CNN(200)의 L개의 레이어들을 통해 L개의 특징 맵들이 출력될 수 있다. L개의 레이어들은 CNN(200)의 중간 레이어를 포함할 수 있다.

[0031] CNN(200)에서 출력된 L개의 특징 맵들은 특징 맵 세트(220)를 구성할 수 있다. 도 1의 소스 특징 맵들(121) 및 타겟 특징 맵들(122)은 각각 특징 맵 세트(220)에 해당할 수 있다. 예를 들어, 도 1의 입력 이미지 쌍(110)의 소스 입력 이미지(111)가 CNN(200)에 입력되어 CNN(200)의 복수의 레이어들에서 소스 특징 맵들(121)이 출력되고, 입력 이미지 쌍(110)의 타겟 입력 이미지(112)가 CNN(200)에 입력되어 CNN(200)의 복수의 레이어들에서 타겟 특징 맵들(122)이 출력될 수 있다. 이후에, CNN(200)의 복수의 레이어들에 따라 소스 특징 맵들(121) 및 타겟 특징 맵들(122)을 페어링하여 도 1의 특징 맵 쌍들(120)이 결정될 수 있다. CNN(200)의 레이어들은 중간 레이어를 포함하므로, 특징 맵 쌍들(120)은 CNN(200)의 중간 레이어에 의해 출력된 중간 특징 맵 쌍을 포함할 수 있다.

[0032] 도 3은 일 실시예에 따른 동적 특징 선택 동작을 나타낸다. 도 3을 참조하면, 이미지 정합 장치는 동적 특징 선택(300)을 통해 특징 맵 쌍들(310) 중에 관련 특징 맵 쌍들(320)을 선택한다.

[0033] 동적 특징 선택(300)은 특징 맵 쌍들(310)의 각 특징 맵 쌍의 개수만큼 수행될 수 있다. 예를 들어, 특징 맵 쌍들(310)에 특징 맵 쌍이 L개 존재하는 경우, 특징 맵 쌍들(310)에 관해 동적 특징 선택(300)이 L번 수행될 수 있다. 이때, L번의 동적 특징 선택(300)은 순차적으로 수행되거나 혹은 병렬적으로 수행될 수 있다. 예를 들어, 이미지 정합 장치는 특징 맵 쌍들(310)의 각 특징 맵 쌍의 특성을 고려하기 위해 특징 맵 쌍들(310)의 각 특징 맵 쌍을 병렬적으로 처리하여 특징 맵 쌍들(310) 중에 관련 특징 맵 쌍들(320)을 선택할 수 있다.

[0034] 특징 맵 쌍들(310)은 특징 맵 쌍(311)을 포함할 수 있다. 이하, 대표적으로 특징 맵 쌍들(310) 중에 특징 맵 쌍(311)에 관한 동적 특징 선택(300)을 설명하겠으나, 아래의 설명은 특징 맵 쌍들(310)의 나머지 특징 맵 쌍들에 관해서도 적용될 수 있다. 이미지 정합 장치는 뉴럴 네트워크(301)를 이용하여 특징 맵 쌍(311)에 관한 동적 특징 선택(300)을 수행할 수 있다. 예를 들어, 이미지 정합 장치는 특징 맵 쌍(311)에 관한 뉴럴 네트워크(301)의 출력에 기초하여 특징 맵 쌍(311)의 선택 여부를 동적으로 결정할 수 있다. 특징 맵 쌍(311)이 선택된 경우 특징 맵 쌍(311)은 관련 특징 맵 쌍들(320)에 포함되고, 특징 맵 쌍(311)이 선택되지 않은 경우 특징 맵 쌍(311)은 관련 특징 맵 쌍들(320)에 포함되지 않는다.

[0035] 뉴럴 네트워크(301)는 특징 맵 쌍(311)의 특성을 고려하여 특징 맵 쌍(311)의 선택 여부를 결정하기 위한 관련성 벡터(relevance vector)를 출력할 수 있다. 이때, 뉴럴 네트워크(301)는 적절한 관련성 벡터를 출력하도록 미리 트레이닝될 수 있다. 예를 들어, 뉴럴 네트워크(301)는 강한 지도 또는 약한 지도를 통해 미리 트레이닝될 수 있다. 뉴럴 네트워크(301)의 트레이닝은 추후 도 6 내지 도 8을 참조하여 보다 상세히 설명한다.

[0036] 도 4는 일 실시예에 따른 동적 특징 선택을 위한 뉴럴 네트워크의 추론 및 트레이닝 과정을 나타낸다. 도 4를 참조하면, 이미지 정합 장치는 동적 특징 선택(400)을 통해 L개의 특징 맵 쌍 $\{(\mathbf{b}_l, \mathbf{b}'_l)\}_{l=0}^{L-1}$ 의 선택 여부를 결정한다. CNN을 통해 추출된 L개의 특징 맵들을 포함하는 두 특징 맵 세트를 각각 B와 B'로 표시하고, 이중에 어느 하나의 특징 맵 쌍은 $(\mathbf{b}_l, \mathbf{b}'_l)$로 표시될 수 있다.

[0037] 동적 특징 선택(400)은 특징 선택을 위한 제1 분기(branch, 401) 및 특징 변환(feature transformation)을 위한 제2 분기(402)를 포함할 수 있다. 제1 분기(401)는 특징 맵 쌍 $(\mathbf{b}_l, \mathbf{b}'_l)$에 대응하는 채널 방향의 특징 벡터(411)를 관련성 벡터(relevance vector, 421)로 인코딩하는 제1 뉴럴 네트워크(420)를 포함하고, 관련성 벡터(421)에 기초하여 특징 선택이 이루어질 수 있다. 보다 구체적으로, 제1 분기(401)는 l번째 특징 맵 쌍 $(\mathbf{b}_l, \mathbf{b}'_l)$을 입력으로 받고, 특징 맵 쌍 $(\mathbf{b}_l, \mathbf{b}'_l)$에 관한 벡터화(410)를 수행하여 특징 맵 쌍 $(\mathbf{b}_l, \mathbf{b}'_l)$에 대응하는 특징 벡터(411)를 생성한다. 예를 들어, 벡터화(410)는 글로벌 평균 풀링(global average pooling) 및 덧셈을 포함할 수 있다. 특징 맵 쌍 $(\mathbf{b}_l, \mathbf{b}'_l)$에 관한 글로벌 평균 풀링을 통해 특징 맵 쌍 $(\mathbf{b}_l, \mathbf{b}'_l)$의 채널-와이즈 통계(channel-wise statistic)를 나타내는 벡터 쌍이 생성될 수 있고, 벡터 쌍의 각 벡터를 엘리먼트-와이즈(element-wise)로 더하여 특징 벡터(411)가 생성될 수 있다. 특징 벡터(411)의 사이즈는 c_l이라고 가정한다.

[0038] 제1 뉴럴 네트워크(420)는 특징 벡터(411)의 입력에 따라 관련성 벡터(421)를 출력할 수 있다. 제1 뉴럴 네트워크(420)는 완전-연결 레이어(fully-connected layer)들을 포함하는 다중 레이어 퍼셉트론(multi-layer perceptron, MLP)일 수 있다. 예를 들어, 제1 뉴럴 네트워크(420)는 2개의 완전-연결 레이어들을 포함할 수 있고, 완전-연결 레이어들은 ReLU 비-선형성(non-linearity)을 가질 수 있다. 일 실시예에 따르면, 관련성 벡터(421)의 사이즈는 2일 수 있고, 관련성 벡터(421)의 엘리먼트들은 l번째 레이어를 선택할지 스킵할지('on' 또는 'off') 결정하기 위한 스코어를 나타낼 수 있다. 이러한 선택 여부의 결정은 관련성 벡터(421)에 argmax(450)를 적용하여 간단히 획득될 수 있지만, argmax(450)는 미분할 수 없기 때문에 이러한 간단한 결정은 역전파(backpropagation)를 불가능하게 한다.

[0039] 일 실시예에 따르면, 이미지 정합 장치는 동적 특징 선택(400)을 트레이닝 가능하고 효과적으로 만들기 위해, 검밸-맥스 트릭(Gumbel-max trick) 및 연속적 완화(continuous relaxation)를 이용할 수 있다. 노이즈 벡터(422)는 독립 항등 분포(independent and identically distributed, iid) 검벨 랜덤 노이즈의 시퀀스라고 가정될 수 있고, z로 표기될 수 있다. Y는 K-클래스의 카테고리 분포(categorical distribution) u의 이산 확률 변수(discrete random variable)라고 할 수 있다. 예를 들어, $p(Y = y) \propto u_y$이고, $y \in \{0, ..., K-1\}$일 수 있다.

[0040] 이에 따라, Y를 $y = \arg\max_{k \in \{0,...,K-1\}} (\log u_k + z_k)$로 샘플링하는 것을 검밸-맥스 트릭을 사용하여

재-파라미터화(re-parameterize)할 수 있다. 미분 가능한 방식으로 argmax(450)를 근사화하기 위해, 검벨-맥스 트릭의 연속적 완화를 통해 argmax(450)는 소프트맥스(440)로 대체할 수 있다. 이산 랜덤 샘플 y를 원-핫 벡터(one-hot vector) $\mathbf{y}$로 표현하면, 검벨-소프트맥스로부터의 샘플은 $\hat{\mathbf{y}} = \mathrm{softmax}((\log \mathbf{u} + \mathbf{z})/\tau)$로 나타낼 수 있다. 여기서 τ는 소프트맥스의 온도(temperature)를 나타낸다.

[0041] 일 실시예에 따르면, 이산 확률 변수는 베르누이 분포(Bernoulli distribution), 예를 들어 $y \in \{0,1\}$를 따를 수 있다. 또한, 관련성 벡터(421)는 'on' 및 'off'에 대한 로그 확률 분포, 예를 들어 $\log \mathbf{u} = \mathbf{r}$를 나타낼 수 있다. 여기서, r_l은 관련성 벡터(421)를 나타낸다. 이 경우, 소프트맥스(440)의 출력은 수학식 1의 형태를 가질 수 있다.

수학식 1

[0042] $$\hat{\mathbf{y}}_l = \mathrm{softmax}(\mathbf{r}_l + \mathbf{z}_l)$$

[0043] 수학식 1에서 z_l은 iid 검벨 랜덤 샘플들의 쌍이다. 예를 들어, 소프트맥스 온도 $\hat{0}$는 1로 설정될 수 있다. 이러한 재-파라미터화는 역전파를 통한 제1 뉴럴 네트워크(420) 및 제2 뉴럴 네트워크(430)의 트레이닝을 가능하게 할 수 있다.

[0044] 다음으로, 제2 분기(402)는 제2 뉴럴 네트워크(430)를 통해 채널 방향의 차원을 감소시켜서 특징 맵 쌍 $(\mathbf{b}_l, \mathbf{b}'_l)$을 저-차원의 특징 맵 쌍 $(\bar{\mathbf{b}}_l, \bar{\mathbf{b}}'_l)$으로 변환(transform)할 수 있다. 일 실시예에 따르면, 제2 뉴럴 네트워크(430)는 1*1 컨볼루션을 통해 위치-별 선형 변환(position-wise linear transformation)을 수행할 수 있다. 또한, 1*1 컨볼루셔널 레이어는 ReLU 비-선형성을 가질 수 있다. 제2 뉴럴 네트워크(430)는 1*1 컨볼루션을 통해 특징 맵 쌍 $(\mathbf{b}_l, \mathbf{b}'_l)$의 차원을 1/ρ 만큼 줄일 수 있다. 따라서, 제2 분기(402)를 통해 각각 h_l X w_l X c_l 크기의 특징 맵 쌍 $(\mathbf{b}_l, \mathbf{b}'_l)$은 보다 간결하고 효과적인 표현으로서 h_l X w_l X c_l/ρ 크기의 저-차원의 특징 맵 쌍 $(\bar{\mathbf{b}}_l, \bar{\mathbf{b}}'_l)$으로 변환될 수 있다. 일 실시예에 따르면, 특징 맵 쌍 $(\mathbf{b}_l, \mathbf{b}'_l)$의 제1 분기(401)를 통해 해당 특징 맵 쌍 $(\mathbf{b}_l, \mathbf{b}'_l)$이 스킵되는 것으로 결정되면, 제2 분기를 통한 해당 특징 맵 쌍 $(\mathbf{b}_l, \mathbf{b}'_l)$의 변환도 함께 스킵되어, 계산 비용을 줄일 수 있다.

[0045] 도 4에서 실선 화살표는 추론을 위한 순방향 경로(forward path)를 나타내고, 점선 화살표는 트레이닝을 위한 역방향 경로(backward path)를 나타낸다. 일 실시예에 따르면, 트레이닝을 위해 검벨-소프트맥스 추정기(estimator)의 직선(straight-through) 버전이 사용될 수 있다. 순방향 경로는 argmax(450)에 의한 이산 샘플로 진행될 수 있고, 역방향 패스는 수학식 1의 소프트맥스(440) 이완의 기울기(gradient)를 계산할 수 있다.

[0046] 순방향 경로에서, argmax(450)의 이산 결정 y에 따라 저-차원의 특징 맵 쌍 $(\bar{\mathbf{b}}_l, \bar{\mathbf{b}}'_l)$에 1('on') 또는 0('off')이 곱해진다. argmax(450)는 순방향 경로에서 이산 결정 y를 내리고, 역방향 경로의 연속적 완화는 기울기가 소프트맥스(440)의 출력 $\hat{\mathbf{y}}$를 통해 전파될 수 있게 한다. 따라서, argmax(450)의 결정에 관계없이 특징 변환 및 관련성 추정의 두 분기들(401, 402)이 효과적으로 업데이트될 수 있다. 이러한 확률적 선택은 랜덤 노이즈를 통해 샘플의 다양성을 증가시키고 트레이닝에서 모드 붕괴를 방지할 수 있고, 시그모이드(sigmoid)를 사용하는 것과 같은 소프트 게이팅에 비해 정확도와 속도 측면에서 우수한 성능을 가질 수 있다.

[0047] 도 5는 일 실시예에 따른 하이퍼 특징 맵의 생성 및 정합 과정을 나타낸다. 도 5를 참조하면, 이미지 정합 장치는 관련 특징 맵 쌍들(510)에 기초한 상관관계 예측(500)을 통해 입력 이미지 쌍에 관한 정합 결과(550)를 생성한다. 상술된 것처럼, 관련 특징 맵 쌍들(510)은 동적 특징 선택을 통해 선택될 수 있다. 관련 특징 맵 쌍들(510)은 도 4의 특징 맵 쌍 $(\mathbf{b}_l, \mathbf{b}'_l)$을 포함하거나, 혹은 저-차원의 특징 맵 쌍 $(\bar{\mathbf{b}}_l, \bar{\mathbf{b}}'_l)$을 포함할 수 있다.

[0048] 이미지 정합 장치는 관련 특징 맵 쌍들(510)에 기초하여 하이퍼 특징 맵 쌍(530)을 생성할 수 있다. 예를 들어, 이미지 정합 장치는 관련 특징 맵 쌍들(510)에 기초한 결합(520)을 통해 하이퍼 특징 맵 쌍(530)을 생성할 수 있다. 결합(520)은 관련 특징 맵 쌍들(510)에 기초한 업샘플링(upsampling) 및 연쇄화(concatenation)를 수행하는 것을 포함할 수 있다. 이때, 이미지 정합 장치는 특징 맵 쌍 $(\mathbf{b}_l, \mathbf{b}'_l)$을 서로 결합하거나, 혹은 저-차원의 특징 맵 쌍 $(\bar{\mathbf{b}}_l, \bar{\mathbf{b}}'_l)$을 서로 결합하여 하이퍼 특징 맵 쌍(530)을 생성할 수 있다. 이하, 저-차원의 특징 맵 쌍 $(\bar{\mathbf{b}}_l, \bar{\mathbf{b}}'_l)$이 결합되는 경우가 설명되지만, 아래의 설명은 특징 맵 쌍 $(\mathbf{b}_l, \mathbf{b}'_l)$이 결합되는 경우에도 적용될 수 있다.

[0049] 하이퍼 특징 맵 쌍(530)의 각 하이퍼 특징 맵은 수학식 2와 같이 나타낼 수 있다.

수학식 2

[0050] $$\left[\zeta(\bar{\mathbf{b}}_{s_1}), \zeta(\bar{\mathbf{b}}_{s_2}), \ldots, \zeta(\bar{\mathbf{b}}_{s_N})\right]$$

[0051] 수학식 2에서 ζ는 입력 특징 맵 $\bar{b}_{sn}$을 공간적으로 업샘플하는 함수를 나타낸다. 예를 들어, ζ는 입력 특징 맵 $\bar{b}_{sn}$을 기본 특징 맵(base feature map) b_0의 사이즈, 또는 입력 이미지의 일정 비율(예: 입력 이미지의 1/4배)로 업샘플할 수 있다. $S = \{s_1, s_2, \ldots, s_N\}$는 선택된 특징(또는, 레이어)의 인덱스의 집합이다. N은 선택된 레이어의 수를 나타낸다. N은 온전히 동적 특징 선택을 통해 결정될 수 있다. 모든 레이어가 오프(off)된다면 $S = \{0\}$으로 설정되고, 기본 특징 맵이 사용될 수 있다.

[0052] 하이퍼 특징 맵 쌍(530)의 각 하이퍼 특징 맵은 H 및 H'으로 나타낼 수 있다. 하이퍼 특징 맵의 각 공간적 위치 p는 대응 이미지 좌표 및 하이퍼 픽셀 특징과 연관될 수 있다. 예를 들어, 위치 p의 이미지 좌표는 x_p로 표시되고, 대응 특징은 f_p로 표시되고, $\mathbf{f}_p = \mathbf{H}(\mathbf{x}_p)$로 나타낼 수 있다. 하이퍼 특징 맵에서 위치 p의 하이퍼 픽셀은 $\mathbf{h}_p = (\mathbf{x}_p, \mathbf{f}_p)$로 정의될 수 있다. 소스 입력 이미지 및 타겟 입력 이미지가 주어지면, $\mathcal{H}$ 및 $\mathcal{H}'$의 두 세트의 하이퍼 픽셀들이 획득될 수 있다.

[0053] 이미지 정합 장치는 하이퍼 특징 맵 쌍(530)에 기초한 매칭(540)을 수행하여 하이퍼 특징 맵 쌍(530)의 상관관계(correlation)를 결정하고, 하이퍼 특징 맵 쌍(530)의 상관관계에 기초하여 입력 이미지 쌍에 관한 정합 결과(550)를 생성할 수 있다. 일 실시예에 따르면, 이미지 정합 장치는 매칭(540)에 기하학적 일관성을 반영하기 위해 확률적 허프 매칭(probabilistic Hough matching, PHM)을 수행할 수 있다. PHM는 기하학적 일관성을 강화하기 위해 허프 공간 투표(Hough space voting)를 통해 외관(appearance)의 유사성을 재-가중(re-weight)할 수 있다.

[0054] 두 세트의 하이퍼 픽셀들은 $\mathcal{D} = (\mathcal{H}, \mathcal{H}')$로 표현될 수 있고, $\mathcal{H}$ 및 $\mathcal{H}'$의 엘리먼트들은 h 및 h'로 표현될 수 있고, h 및 h'의 매치는 $m = (\mathbf{h}, \mathbf{h}')$로 표현될 수 있다. 두 하이퍼 픽셀 사이의 가능한 오프셋(다시 말해, 이미지 변환(image transfer))을 나타내는 허프 공간 X가 주어지면, 매치 m에 대한 신뢰도 $p(m|\mathcal{D})$는 수학식 3과 같이 계산될 수 있다.

수학식 3

[0055] $$p(m|\mathcal{D}) \propto p(m_{\mathrm{a}}) \sum_{\mathbf{x}\in\mathcal{X}} p(m_{\mathrm{g}}|\mathbf{x}) \sum_{m\in\mathcal{H}\times\mathcal{H}'} p(m_{\mathrm{a}}) p(m_{\mathrm{g}}|\mathbf{x})$$

[0056] 수학식 3에서 $p(m_{\mathrm{a}})$는 외관 매치에 대한 신뢰도를 나타낸다. $p(m_{\mathrm{g}}|\mathbf{x})$는 오프셋 x를 사용한 기하학적 매치

에 대한 신뢰도이며, m에 의해 유도된 오프셋이 x에 얼마나 가까운지 측정할 수 있다. 모든 매치에 대해 허프 공간 X를 공유함으로써, PHM은 우수한 실증적 성능으로 매치 신뢰도를 효율적으로 계산할 수 있다. 외관 매치 신뢰도는 수학식 4과 같이 하이퍼 픽셀 특징을 이용하여 계산될 수 있다.

수학식 4

[0057]
$$p(m_{\mathrm{a}}) \propto \mathrm{ReLU}\left(\frac{\mathbf{f}_p \cdot \mathbf{f}'_p}{\|\mathbf{f}_p\| \left\|\mathbf{f}'_p\right\|}\right)^2$$

[0058] 수학식 4에서 제곱은 매치 신뢰도가 작아지는 것을 억제하는 효과가 있다. PHM의 출력은 $|\mathcal{H}| \times |\mathcal{H}'|$의 상관 행렬(correlation matrix)일 수 있고, 해당 상관 행렬은 C로 표시될 수 있다. 일 실시예에 따르면, 해당 상관 행렬에서 잡음 상관 값(noisy correlation value)을 억제하기 위해 소프트 상호 인접 이웃 필터링(soft mutual nearest neighbor filtering)이 수행될 수 있다.

[0059] 이미지 정합 장치는 상관 행렬 C에 기초하여 각 소스 하이퍼 픽셀 $\mathbf{h}_i$에 가장 높은 상관 관계를 가진 타겟 하이퍼 픽셀 $\hat{\mathbf{h}}'_j$를 할당하여 각 하이퍼 픽셀의 정합(correspondence)을 설정하고, 각 하이퍼 픽셀의 정합에 따라 정합 결과(550)를 생성할 수 있다. 대부분의 경우 기본 특징 맵의 해상도(예: ResNet-101을 백본으로 사용하는 입력 이미지의 1/4)는 상대적으로 높고, 하이퍼 이미지의 공간 해상도는 기본 특징 맵의 해상도와 동일하므로, 하이퍼 픽셀 정합은 유사-밀집 매치(quasi-dense match)를 생성할 수 있다.

[0060] 소스 입력 이미지의 키포인트 p_m이 주어지면, 이미지 정합 장치는 p_m의 가장 가까운 하이퍼 픽셀 정합을 이용해 p_m을 변환하여 타겟 입력 이미지에서 정합 위치 $\hat{\mathbf{p}}'_m$를 예측할 수 있다. 이때, 키포인트 p_m의 이웃 하이퍼 픽셀의 모든 정합이 수집되고, 각 정합에 따른 개별 변환의 기하학적 평균이 $\hat{\mathbf{p}}'_m$의 최종 예측을 위해 사용될 수 있다. 이러한 통계 기반의 키포인트 변환은 개별 변환 시 발생할 수 있는 잘못-로컬화된(mis-localized) 예측을 개선하여 정확도를 향상시킬 수 있다.

[0061] 도 6은 일 실시예에 따른 뉴럴 네트워크의 트레이닝을 위한 전반적인 구조를 나타낸다. 도 6을 참조하면, 상관관계 예측(610)에 기초하여 예측된 상관 행렬 C가 생성될 수 있다. 예를 들어, 상관관계 예측(610)에 따른 예측 결과에 로-와이즈 소프트맥스(row-wise softmax)를 적용하여 예측된 상관 행렬 C가 생성될 수 있다. 상관관계 예측(610)은 도 5의 상관관계 예측(500)에 대응할 수 있다.

[0062] 이후에, 예측된 상관 행렬 C 및 진실(ground-truth) 상관 행렬 G에 기초하여 트레이닝 손실 $\mathcal{L}$이 계산될 수 있고, 트레이닝 손실 $\mathcal{L}$에 기초하여 동적 특징 선택을 위한 뉴럴 네트워크가 트레이닝될 수 있다. 예를 들어, 도 4의 제1 뉴럴 네트워크(420) 및 제2 뉴럴 네트워크(430)가 역방향 경로를 통해 트레이닝될 수 있다.

[0063] 일 실시예에 따르면, 트레이닝 목표에 수학식 5에 따른 선택 비율에 관한 제약이 추가될 수 있다.

수학식 5

[0064]
$$\mathcal{L}_{\mathrm{sel}} = \sum_{l=0}^{L-1} (\bar{z}_l - \mu)^2$$

[0065] 수학식 5에서 $\bar{z}_l$은 l번째 레이어가 선택된 미니-배치 내에서 이미지 쌍의 일부이고, μ는 선택 비율에 대한 설정 파라미터이다. 예를 들어, μ는 0.3, 0.5, 1.0 등으로 설정될 수 있다. 동적 특징 선택을 위한 뉴럴 네트워크는 설정 파라미터에 의해 정해진 선택 비율에 기초하여 관련 특징 맵 쌍들을 선택하도록 제어될 수 있다. 예를 들어, μ가 0.3으로 설정된 경우, 동적 특징 선택을 통해 특징 맵 쌍들의 30%만 관련 특징 맵 쌍들로 선택될 수 있다. 설정 파라미터는 어플리케이션에 의해 요구되는 속도 및 정확도 중 적어도 하나에 기초하여 결정될 수 있다. 예를 들어, 어플리케이션의 우선순위가 빠른 속도에 있는지, 혹은 정확도에 있는지에 따라 설정 파라미터가 조절될 수 있다. 레이어 선택 손실 $\mathcal{L}_{\mathrm{sel}}$은 선택의 다양성을 증가시켜 트레이닝을 향상시키고, 속

도와 정확도 사이에서 균형을 맞출 수 있다.

[0066] 일 실시예에 따르면, 트레이닝 목표에 대응하는 트레이닝 손실 $\mathcal{L}$은 수학식 6과 같이 매칭 손실 $\mathcal{L}_{\text{match}}$과 레이어 선택 손실 $\mathcal{L}_{\text{sel}}$의 조합으로 정의될 수 있다.

수학식 6

[0067] $$\mathcal{L} = \mathcal{L}_{\text{match}} + \mathcal{L}_{\text{sel}}$$

[0068] 수학식 6의 선택 손실 $\mathcal{L}_{\text{sel}}$은 수학식 5를 통해 결정될 수 있고, 매칭 손실 $\mathcal{L}_{\text{match}}$은 아래에서 설명되는 도 7의 강한 지도 또는 도 8의 약한 지도를 통해 결정될 수 있다.

[0069] 도 7은 일 실시예에 따른 강한 지도를 통한 트레이닝 과정을 나타낸다. 강한 지도에 따른 트레이닝을 위해, 각 트레이닝 이미지 쌍의 키포인트 매치에 관한 주석(annotation)이 제공된다. 예를 들어, 각 이미지 쌍에 관해 수학식 7에 따른 좌표 쌍 세트가 주석으로 제공될 수 있다. 수학식 7에서 M은 매치 주석의 개수이다.

수학식 7

[0070] $$\mathcal{M} = \{(\mathbf{p}_m, \mathbf{p}'_m)\}_{m=1}^{M}$$

[0071] 뉴럴 네트워크의 출력을 진실 주석(ground-truth annotation)과 비교하기 위해, 주석은 이산 상관 행렬(discrete correlation matrix)의 형태로 변환될 수 있다. 우선, 트레이닝 이미지 쌍(710, 720)의 각 좌표 쌍 $(\mathbf{p}_m, \mathbf{p}'_m)$에 대응하는 하이퍼 특징 맵 쌍(711, 721)의 위치 인덱스 쌍 (k_m, k'_m)가 식별된다. 식별된 매치 인덱스 쌍 세트 $\{(k_m, k'_m)\}_{m=1}^{M}$가 주어지면, 진실 행렬(ground-truth matrix) G의 m번째 행에 k'_m의 원-핫 벡터 표현을 할당하여 $\mathbf{G} \in \{0,1\}^{M \times |\mathcal{H}'|}$가 구성될 수 있다. 한편으로, $\hat{\mathbf{C}}$의 m번째 행에 C의 k_m번째 행을 할당하여 $\hat{\mathbf{C}} \in \mathbb{R}^{M \times |\mathcal{H}'|}$가 구성될 수 있다. $\hat{\mathbf{C}}$은 0의 평균(mean) 및 단위 분산을 갖도록 정규화(normalize)될 수 있고, 이후에 행렬 $\hat{\mathbf{C}}$의 각 행에 소프트맥스가 적용될 수 있다. 도 7에 $\hat{\mathbf{C}}$ 및 G의 구성이 나타나 있다.

[0072] 이제 $\hat{\mathbf{C}}$과 G 사이의 대응 행(corresponding row)이 카테고리 확률 분포(categorical probability distribution)로서 비교될 수 있다. 수학식 8에 따라, $\hat{\mathbf{C}}$과 G 사이의 교차-엔트로피 손실(730)을 구한 뒤, 교차-엔트로피 손실(730)에 대한 평균화(740) 작업을 수행하여, 강한 지도에 따른 매칭 손실 $\mathcal{L}_{\text{match}}$이 계산될 수 있다.

수학식 8

[0073] $$\mathcal{L}_{\text{match}} = -\frac{1}{M} \sum_{m=1}^{M} \omega_m \sum_{j=1}^{|\mathcal{H}'|} \mathbf{G}_{mj} \log \hat{\mathbf{C}}_{mj}$$

[0074] 수학식 8에서 ω_m은 m번째 키포인트에 대한 중요도 가중치이며, 수학식 9와 같이 정의될 수 있다.

수학식 9

[0075]
$$\omega_m = \begin{cases} \left(\left\|\hat{\mathbf{p}}'_m - \mathbf{p}'_m\right\| / \delta_{\text{thres}}\right)^2 & \text{if} \quad \left\|\hat{\mathbf{p}}'_m - \mathbf{p}'_m\right\| < \delta_{\text{thres}} \\ 1 & \text{otherwise.} \end{cases}$$

[0076] 수학식 9에 따르면 예측된 키포인트 $\hat{\mathbf{p}}'_m$과 타겟 키포인트 $\mathbf{p}'_m$ 사이의 유클리언 거리(Euclidean distance)가 임계 거리 δ_{thres}보다 작은 경우, 키포인트 가중치 ω_m은 대응하는 교차-엔트로피 항의 효과를 감소시켜 트레이닝을 도울 수 있다. 일 실시예에 따르며, 강한 지도 학습은 합성 쌍(synthetic pair)을 사용하는 자가-지도 학습(self-supervised learning)에도 사용될 수 있다. 이러한 자가-지도 학습은 일반적으로 일반화 성능에 대한 지도 비용을 상쇄할 수 있다.

[0077] 도 8은 일 실시예에 따른 약한 지도를 통한 트레이닝 과정을 나타낸다. 강한 지도에 따른 트레이닝과 달리, 약한 지도에 따른 트레이닝의 경우 각 이미지 쌍에 대해 이미지 수준의 레이블만 제공된다. 예를 들어, 동일한 카테고리의 객체를 포함하는 양의 이미지 쌍(positive image pair, 810)에 대해 양의 레이블이 제공될 수 있고, 다른 카테고리의 객체를 포함하는 음의 이미지 쌍(negative image pair, 820)에 대해 음의 레이블이 제공될 수 있다. 또한, 양의 이미지 쌍(810)의 상관 행렬은 $\mathbf{C}_+$로 표시되고, 음의 이미지 쌍(820)의 상관 행렬은 $\mathbf{C}_-$로 표시될 수 있다. 이때, $\mathbf{C} \in \mathbb{R}^{|\mathcal{H}| \times |\mathcal{H}'|}$에 대해, 상관 엔트로피(correlation entropy, 840) $s(\mathbf{C})$는 수학식 10과 같이 정의될 수 있다.

수학식 10

[0078]
$$s(\mathbf{C}) = -\frac{1}{|\mathcal{H}|} \sum_{i=1}^{|\mathcal{H}|} \sum_{j=1}^{|\mathcal{H}'|} \phi(\mathbf{C})_{ij} \log \phi(\mathbf{C})_{ij}$$

[0079] 수학식 8에서 $\phi(\cdot)$는 정규화(830)에 대응하며, 예를 들어 로-와이즈 L1-정규화(L1-normalization)에 해당할 수 있다. 상관 엔트로피가 높을수록 두 이미지 사이의 정합이 덜 구별(distinctive)될 수 있다. 도 8에 도시된 것처럼, 양의 이미지 쌍(810)에 보다 뚜렷한 정합이 포함할 가능성이 높기 때문에, 양의 이미지 쌍(810)에는 낮은 엔트로피가 부여되고, 음의 이미지 쌍(820)에는 높은 엔트로피가 부여될 수 있다. 이에 따라, 상관 엔트로피(840)에 관한 평균화(850)가 수행될 수 있다. 따라서, 약한 지도의 매칭 손실 $\mathcal{L}_{\text{match}}$은 수학식 11과 같이 나타낼 수 있다.

수학식 11

[0080]
$$\mathcal{L}_{\text{match}} = \frac{s(\mathbf{C}_+) + s(\mathbf{C}_+^\top)}{s(\mathbf{C}_-) + s(\mathbf{C}_-^\top)}$$

[0081] 트레이닝 방식에 따라 수학식 8 또는 수학식 11의 매칭 손실 $\mathcal{L}_{\text{match}}$이 정의되고, 이를 이용하여 동적 특징 선택을 위한 뉴럴 네트워크가 트레이닝될 수 있다. 이러한 트레이닝을 통해 뉴럴 네트워크는 이미지 정합에 최적의 특징 맵들의 조합을 구성할 수 있다. 이때, 수학식 6과 같이 트레이닝 손실 $\mathcal{L}$의 정의를 위해 선택 손실 $\mathcal{L}_{\text{sel}}$이 더 고려될 수 있다. 선택 손실 $\mathcal{L}_{\text{sel}}$을 통해 컴퓨팅 부담이 완화(예: 추론 시간의 단축)될 수 있고, 경우에 따라 정합 정확도가 높아지는 결과가 도출될 수도 있다.

[0082] 도 9는 일 실시예에 따른 이미지 정합 장치의 개략적인 구성을 나타낸다. 도 9을 참조하면, 장치(900)는 프로세서(910) 및 메모리(920)를 포함한다. 메모리(920)는 프로세서(910)에 연결되고, 프로세서(910)에 의해 실행

가능한 명령어들, 프로세서(910)가 연산할 데이터 또는 프로세서(910)에 의해 처리된 데이터를 저장할 수 있다. 메모리(920)는 비일시적인 컴퓨터 판독가능 매체, 예컨대 고속 랜덤 액세스 메모리 및/또는 비휘발성 컴퓨터 판독가능 저장 매체(예컨대, 하나 이상의 디스크 저장 장치, 플래쉬 메모리 장치, 또는 기타 비휘발성 솔리드 스테이트 메모리 장치)를 포함할 수 있다.

[0083] 프로세서(910)는 도 1 내지 도 8, 도 10 및 도 11을 참조하여 설명된 하나 이상의 동작을 실행하기 위한 명령어들을 실행할 수 있다. 프로세서(910)는 입력 이미지 쌍의 입력에 따른 CNN의 복수의 레이어들의 출력에 대응하는 복수의 특징 맵 쌍들을 획득하고, 복수의 특징 맵 쌍들의 각 특징 맵 쌍의 특성을 고려하여 복수의 특징 맵 쌍들 중에 일부의 특징 맵 쌍들을 선택하고, 선택된 일부의 특징 맵 쌍들에 기초하여 하이퍼 특징 맵 쌍을 생성하고, 하이퍼 특징 맵 쌍의 상관관계에 기초하여 입력 이미지 쌍에 관한 정합 결과를 생성할 수 있다.

[0084] 또한, 프로세서(910)는 정합 결과를 이용한 연계 동작들을 수행할 수 있다. 예를 들어, 연계 동작은 객체 검출, 객체 추적, 객체 인식, 이미지 검색, 3차원 재건, 모션 추정, 깊이 추정, 동작 인식과 같은 다양한 응용을 포함할 수 있다. 예를 들어, 객체 추적의 경우 대상 객체 이미지와 현재 비디오 프레임의 정밀한 정합 계산이 핵심적인 요소이다. 동작 인식을 위해, 연속적인 프레임들 사이의 효과적인 정합은 중요한 모션 정보로 사용된다. 또한, 실시예들은 최근의 다른 작업에 널리 사용될 수 있다. 예를 들어, 이미지 정합에서 사용되는 4차원 CNN은 비디오 프레임에서 광학적 흐름을 생성하는 모델을 설계하는데 필요하다. 따라서, 실시예들에 따른 이미지 정합 동작은 다른 문제와 모델에 광범위하게 확장되고 적용될 수 있다.

[0085] 도 10은 일 실시예에 따른 이미지 정합 방법을 개략적으로 나타낸다. 도 10을 참조하면, 이미지 정합 장치는 단계(1010)에서 입력 이미지 쌍의 입력에 따른 CNN의 복수의 레이어들의 출력에 대응하는 복수의 특징 맵 쌍들을 획득하고, 단계(1020)에서 복수의 특징 맵 쌍들의 각 특징 맵 쌍의 특성을 고려하여 복수의 특징 맵 쌍들 중에 일부의 특징 맵 쌍들을 선택하고, 단계(1030)에서 선택된 일부의 특징 맵 쌍들에 기초하여 하이퍼 특징 맵 쌍을 생성하고, 단계(1040)에서 하이퍼 특징 맵 쌍의 상관관계에 기초하여 입력 이미지 쌍에 관한 정합 결과를 생성한다. 그 밖에, 이미지 정합 방법에는 도 1 내지 도 9 및 도 11에 관한 설명이 적용될 수 있다.

[0086] 도 11은 일 실시예에 따른 이미지 정합 장치와 관련된 전자 장치를 나타낸다. 도 11를 참조하면, 전자 장치(1100)는 입력 이미지를 획득하고, 획득된 입력 이미지에 관한 시각적 정합(visual correspondence)을 수행할 수 있다. 또한, 전자 장치(1100)는 정합 결과를 이용한 연계 동작들을 수행할 수 있다. 예를 들어, 연계 동작은 객체 검출, 객체 추적, 객체 인식, 이미지 검색, 3차원 재건, 모션 추정, 깊이 추정, 동작 인식과 같은 다양한 응용을 포함할 수 있다. 전자 장치(1100)는 도 1의 이미지 정합 장치(100)를 구조적 및/또는 기능적으로 포함할 수 있다.

[0087] 전자 장치(1100)는 프로세서(1110), 메모리(1120), 카메라(1130), 저장 장치(1140), 입력 장치(1150), 출력 장치(1160) 및 네트워크 인터페이스(1170)를 포함할 수 있으며, 이들은 통신 버스(1180)를 통해 서로 통신할 수 있다. 예를 들어, 전자 장치(1100)는 이동 전화, 스마트 폰, PDA, 넷북, 태블릿 컴퓨터, 랩톱 컴퓨터 등과 같은 모바일 장치, 스마트 워치, 스마트 밴드, 스마트 안경 등과 같은 웨어러블 디바이스, 데스크탑, 서버 등과 같은 컴퓨팅 장치, 텔레비전, 스마트 텔레비전, 냉장고 등과 같은 가전 제품, 도어 락 등과 같은 보안 장치, 스마트 차량 등과 같은 차량의 적어도 일부로 구현될 수 있다.

[0088] 프로세서(1110)는 전자 장치(1100) 내에서 실행하기 위한 기능 및 명령어들을 실행한다. 예를 들어, 프로세서(1110)는 메모리(1120) 또는 저장 장치(1140)에 저장된 명령어들을 처리할 수 있다. 프로세서(1110)는 도 1 내지 도 10을 통하여 설명된 하나 이상의 동작을 수행할 수 있다.

[0089] 메모리(1120)는 이미지 정합을 위한 데이터를 저장한다. 메모리(1120)는 컴퓨터 판독가능한 저장 매체 또는 컴퓨터 판독가능한 저장 장치를 포함할 수 있다. 메모리(1120)는 프로세서(1110)에 의해 실행하기 위한 명령어들을 저장할 수 있고, 전자 장치(1100)에 의해 소프트웨어 및/또는 애플리케이션이 실행되는 동안 관련 정보를 저장할 수 있다.

[0090] 카메라(1130)는 사진 및/또는 비디오를 촬영할 수 있다. 일 실시예에 따르면, 카메라(1130)는 입력 이미지 쌍의 적어도 하나의 입력 이미지를 생성할 수 있다. 이때, 카메라(113)에 의해 생성된 이미지는 타겟 이미지로 사용될 수 있다. 예를 들어, 카메라(1130)를 통해 타겟 이미지들의 시퀀스가 생성될 수 있고, 주어진 소스 이미지 내의 소스 객체의 대응 객체(예: 동일한 카테고리의 객체)를 나타내는 대응 점이 시퀀스의 각 타겟 이미지에서 검출될 수 있다. 이러한 시퀀스의 생성 및 대응 점의 검출은 실시간으로 수행될 수 있다.

[0091] 저장 장치(1140)는 컴퓨터 판독가능한 저장 매체 또는 컴퓨터 판독가능한 저장 장치를 포함한다. 일 실시예에

따르면, 저장 장치(1140)는 메모리(1120)보다 더 많은 양의 정보를 저장하고, 정보를 장기간 저장할 수 있다. 예를 들어, 저장 장치(1140)는 자기 하드 디스크, 광 디스크, 플래쉬 메모리, 플로피 디스크 또는 이 기술 분야에서 알려진 다른 형태의 비휘발성 메모리를 포함할 수 있다.

[0092] 입력 장치(1150)는 키보드 및 마우스를 통한 전통적인 입력 방식, 및 터치 입력, 음성 입력, 및 이미지 입력과 같은 새로운 입력 방식을 통해 사용자로부터 입력을 수신할 수 있다. 예를 들어, 입력 장치(1150)는 키보드, 마우스, 터치 스크린, 마이크로폰, 또는 사용자로부터 입력을 검출하고, 검출된 입력을 전자 장치(1100)에 전달할 수 있는 임의의 다른 장치를 포함할 수 있다.

[0093] 출력 장치(1160)는 시각적, 청각적 또는 촉각적인 채널을 통해 사용자에게 전자 장치(1100)의 출력을 제공할 수 있다. 출력 장치(1160)는 예를 들어, 디스플레이, 터치 스크린, 스피커, 진동 발생 장치 또는 사용자에게 출력을 제공할 수 있는 임의의 다른 장치를 포함할 수 있다. 네트워크 인터페이스(1170)는 유선 또는 무선 네트워크를 통해 외부 장치와 통신할 수 있다.

[0094] 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. 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. 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 또는 복수의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

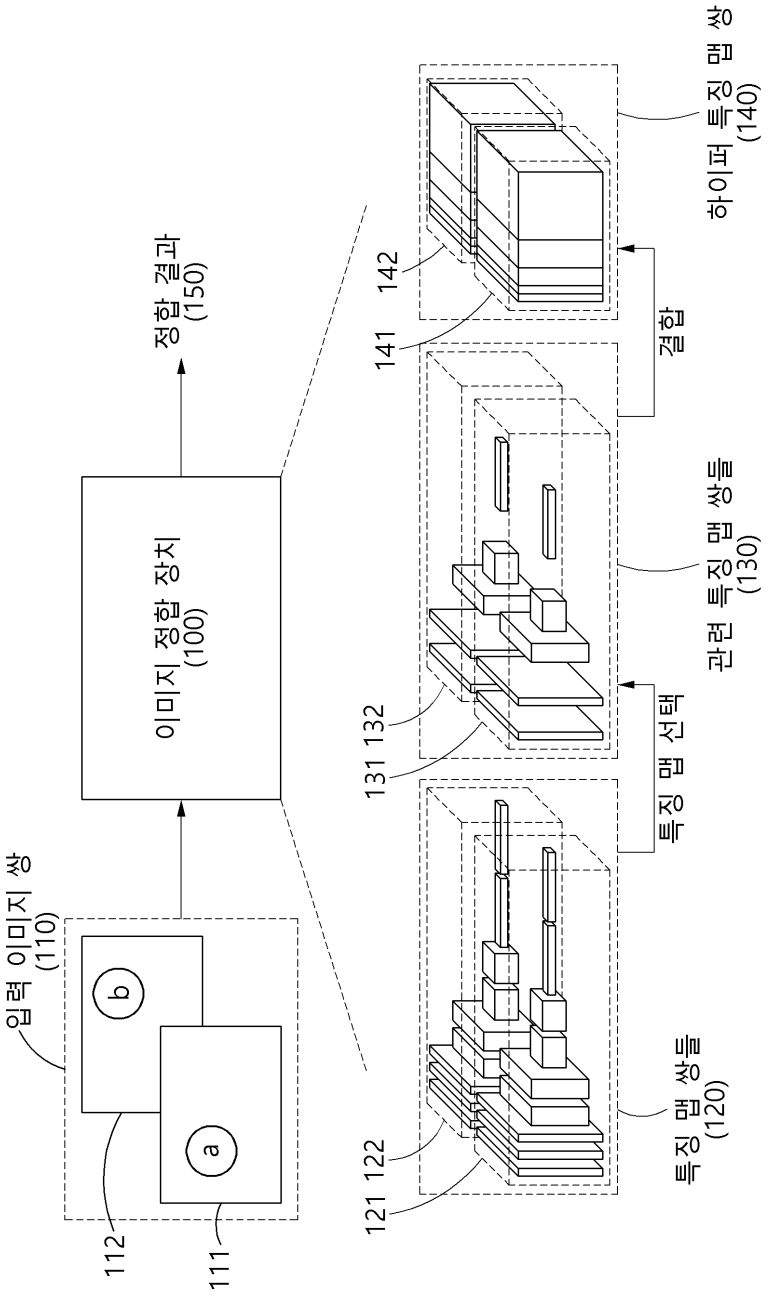
[0095] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 또는 복수의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치, 또는 전송되는 신호 파(signal wave)에 영구적으로, 또는 일시적으로 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 또는 복수의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

[0096] 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.

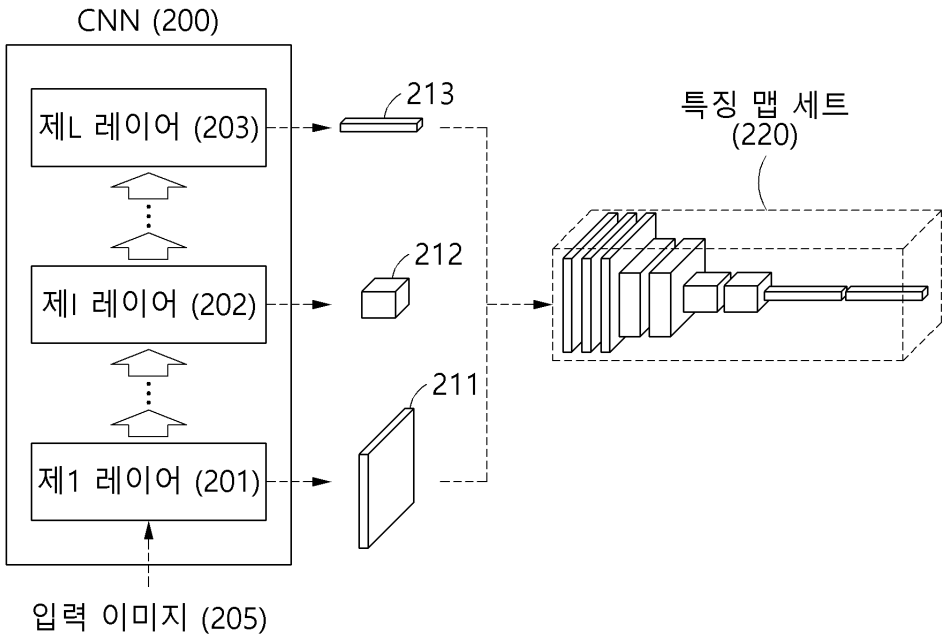
[0097] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 청구범위의 범위에 속한다.

도면

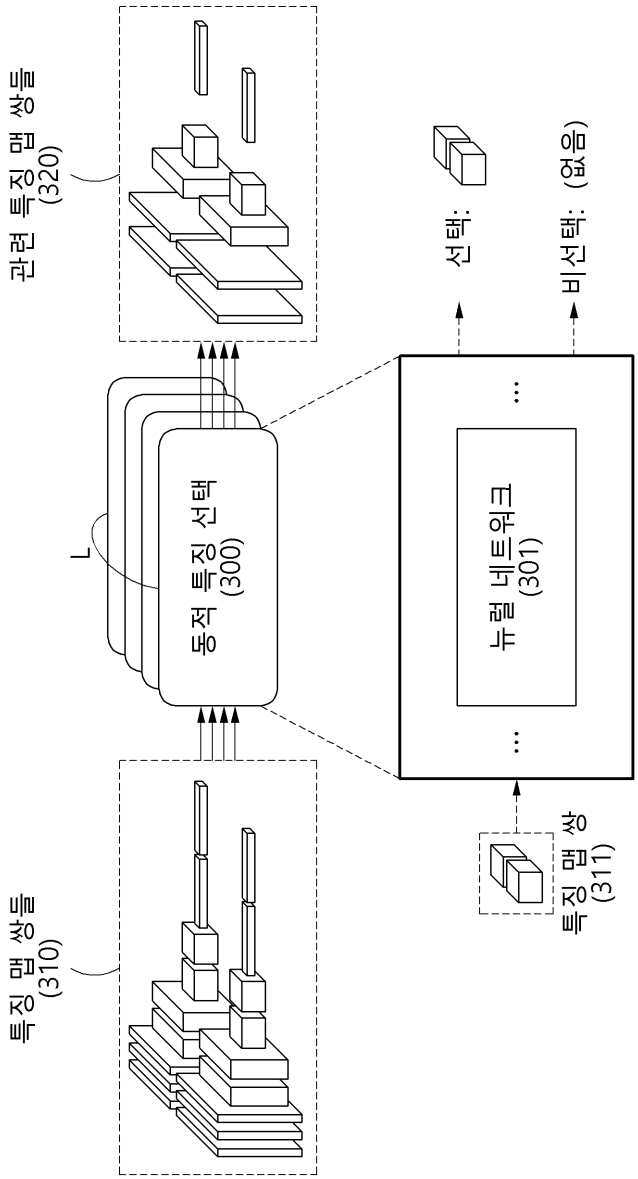
도면1


입력 이미지 쌍
(110)
b
a
112
111
이미지 정합 장치
(100)
정합 결과
(150)
121 122
131 132
141 142
특징 맵 쌍들
(120)
특징 맵 선택
관련 특징 맵 쌍들
(130)
결합
하이퍼 특징 맵 쌍
(140)

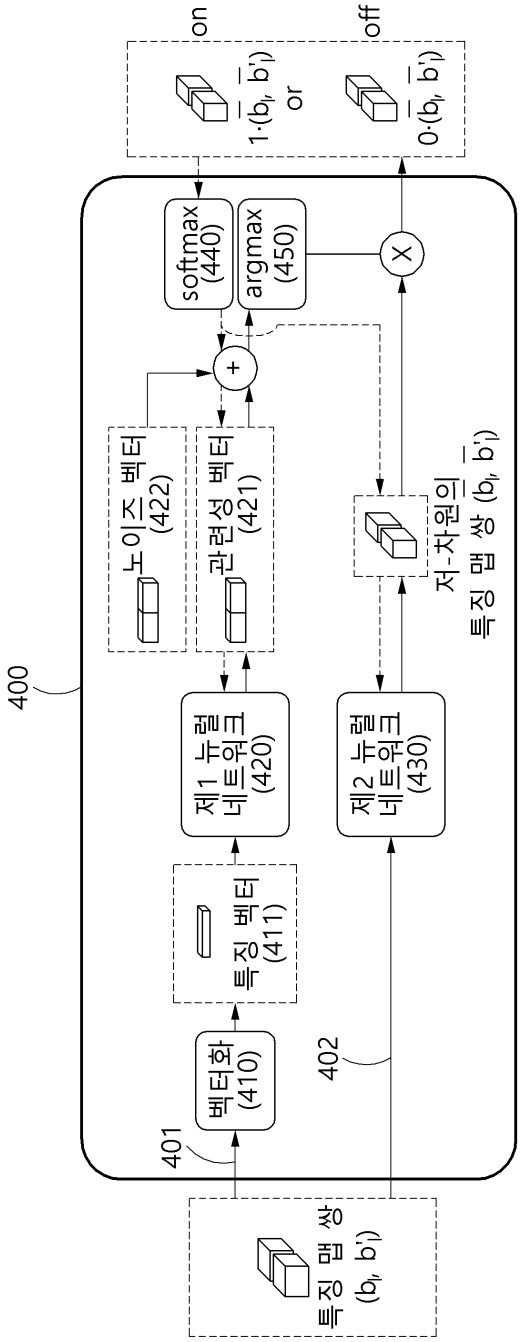
도면2


CNN (200)
제L 레이어 (203)
제I 레이어 (202)
제1 레이어 (201)
입력 이미지 (205)
213
212
211
특징 맵 세트
(220)

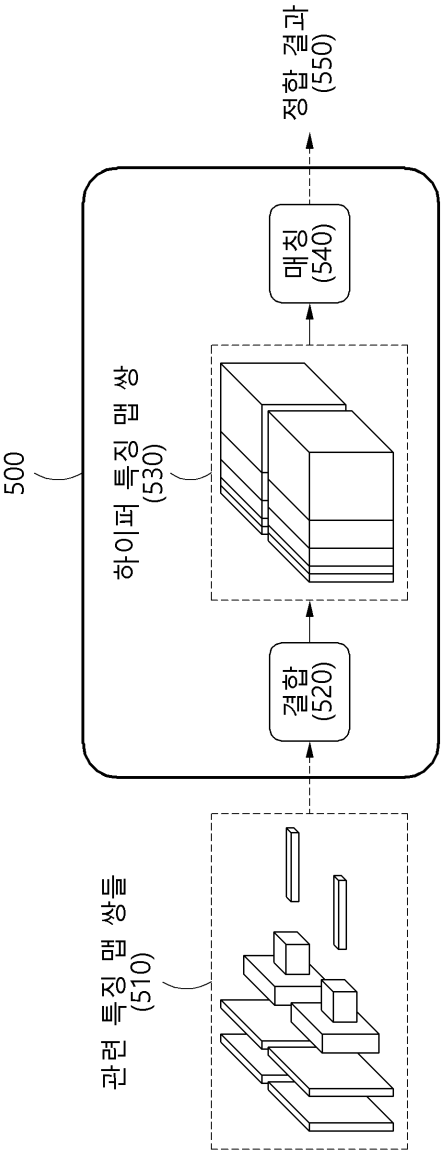
도면3


관련 특징 맵 쌍들
(320)
선택: 1
비선택: 0 (없음)
동적 특징 선택
(300)
L
뉴럴 네트워크
(301)
특징 맵 쌍
(311)
특징 맵 쌍들
(310)

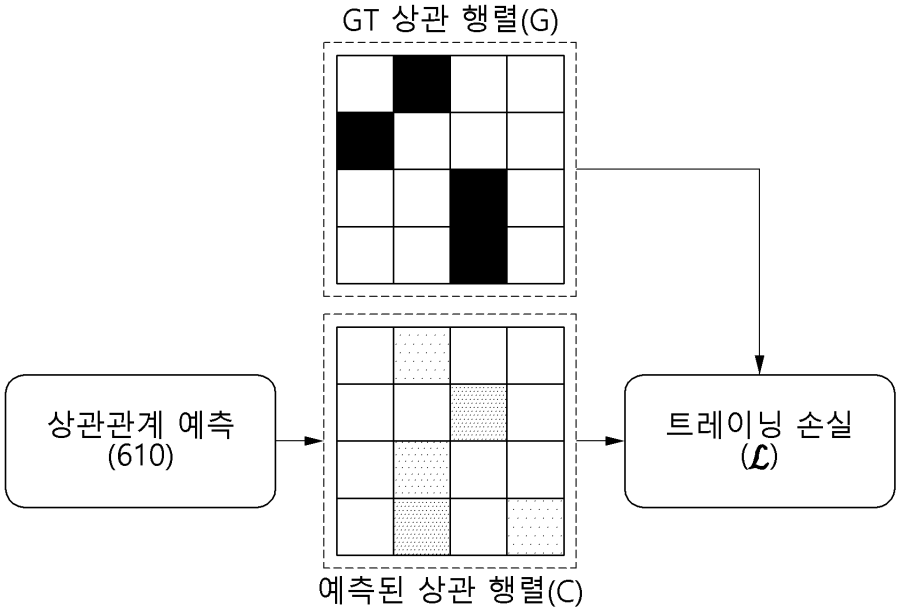
도면4


on
off
softmax (440)
argmax (450)
노이즈 벡터 (422)
관련성 벡터 (421)
저-차원의 특징 맵 쌍
제1 뉴럴 네트워크 (420)
제2 뉴럴 네트워크 (430)
특징 벡터 (411)
벡터화 (410)
특징 맵 쌍
400
401
402

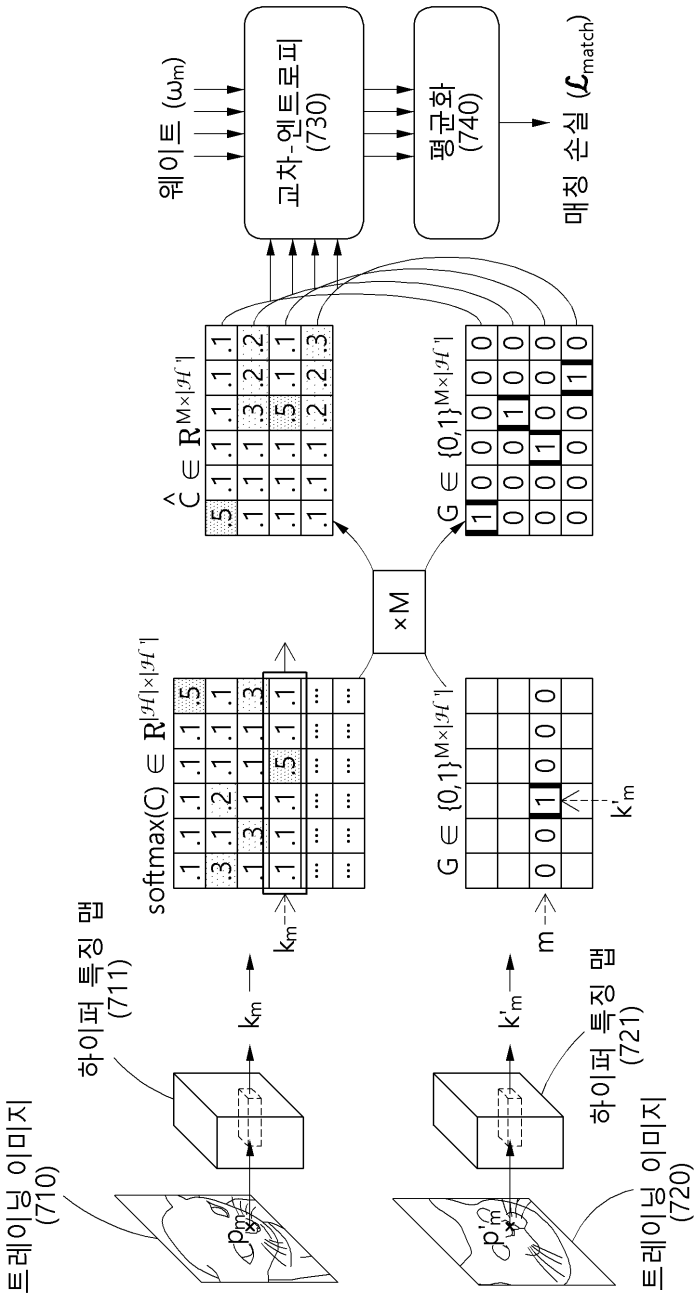
도면5



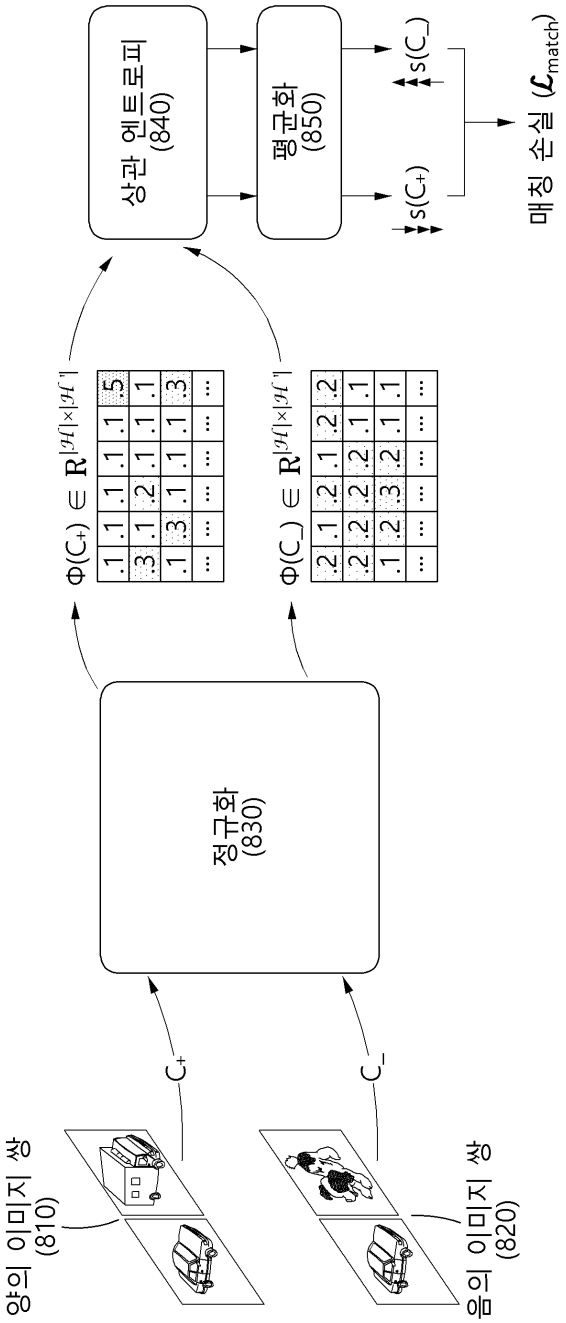
도면6



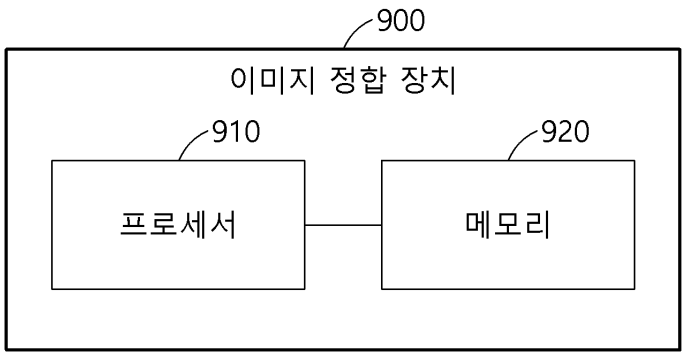
도면7


웨이트 (ω_m)
교차-엔트로피 (730)
평균화 (740)
매칭 손실 ($\mathcal{L}_{match}$)
$\hat{C} \in \mathbb{R}^{M \times |\mathcal{H}'|}$
$G \in \{0,1\}^{M \times |\mathcal{H}'|}$
×M
softmax(C) $\in \mathbb{R}^{|\mathcal{H}| \times |\mathcal{H}'|}$
$G \in \{0,1\}^{M \times |\mathcal{H}'|}$
k_m
k'_m
m
하이퍼 특징 맵 (711)
하이퍼 특징 맵 (721)
트레이닝 이미지 (710)
트레이닝 이미지 (720)

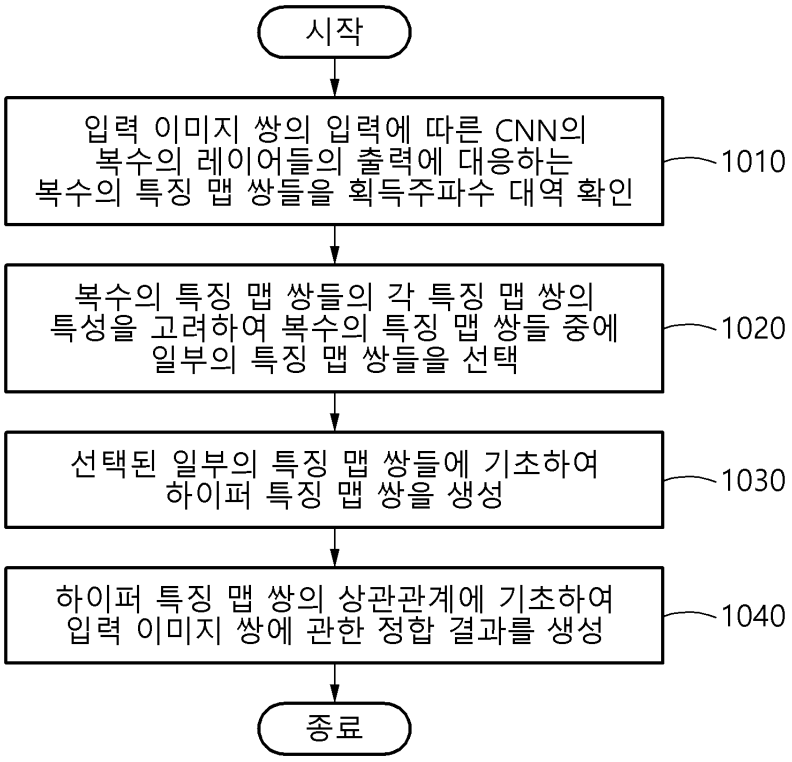
도면8


양의 이미지 쌍 (810)
음의 이미지 쌍 (820)
C+
C_
정규화 (830)
Φ(C+) ∈ R^{|H|×|H'|}
Φ(C_) ∈ R^{|H|×|H'|}
상관 엔트로피 (840)
평균화 (850)
s(C+)
s(C_)
매칭 손실 (ℒmatch)

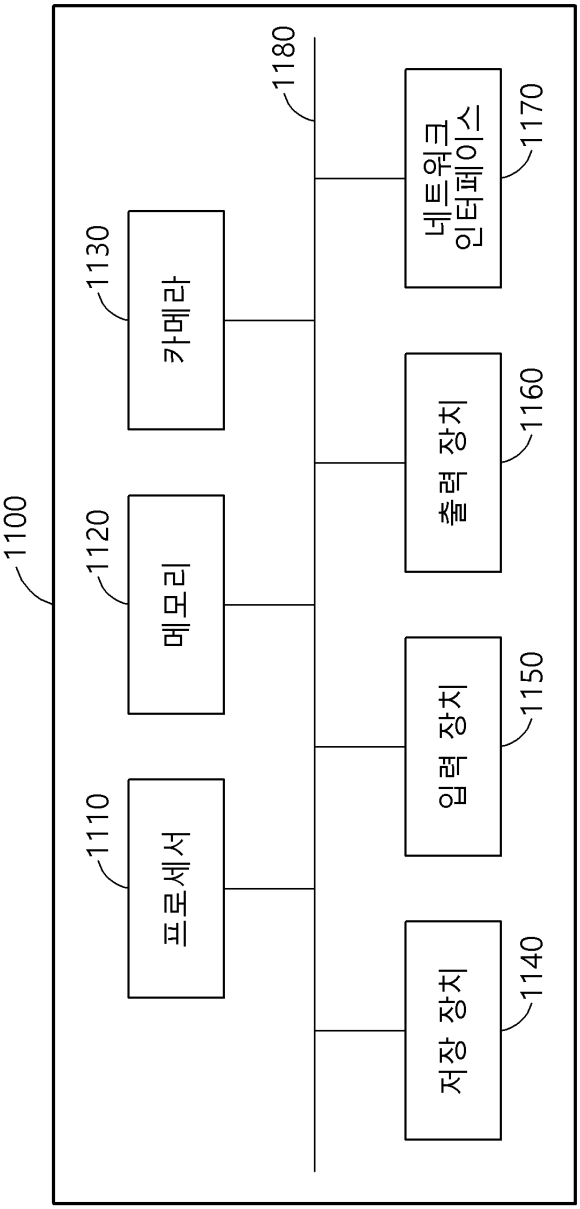
도면9


900
이미지 정합 장치
910
프로세서
920
메모리

도면10


시작
입력 이미지 쌍의 입력에 따른 CNN의 복수의 레이어들의 출력에 대응하는 복수의 특징 맵 쌍들을 획득주파수 대역 확인
1010
복수의 특징 맵 쌍들의 각 특징 맵 쌍의 특성을 고려하여 복수의 특징 맵 쌍들 중에 일부의 특징 맵 쌍들을 선택
1020
선택된 일부의 특징 맵 쌍들에 기초하여 하이퍼 특징 맵 쌍을 생성
1030
하이퍼 특징 맵 쌍의 상관관계에 기초하여 입력 이미지 쌍에 관한 정합 결과를 생성
1040
종료

도면11


1100
1110
프로세서
1120
메모리
1130
카메라
1140
저장 장치
1150
입력 장치
1160
출력 장치
1170
네트워크 인터페이스
1180