

- 전체 청구항 수 : 총 15 항

Figure 1 illustrates the architecture of the proposed RGB-T video object segmentation network. The input is an RGB-T video sequence (110). The RGB branch (102) processes the RGB feature map through a series of convolutional layers (Conv1 to Conv4) and an SCA Module (104) repeated N times. The Thermal branch (102) processes the Thermal feature map through a series of convolutional layers (Conv1 to Conv4) and an SCA Module (104) repeated N times. The SCA Module (104) takes inputs from the RGB and Thermal feature maps (102) and the output of the previous SCA module (100). The final output is a segmentation map (114) produced by Conv5 to Conv11. A legend on the right defines the color-coded components: Illuminance Network (green), IN output feature map (dark blue), RGB Feature map (light blue), Thermal Feature map (purple), Self-Attention (yellow), Cross-Attention (orange), Fused Feature map (brown), and N2N(1x1 Convolution) (light yellow). Symbols for Element-wise Sum ($\oplus$), Hadamard Product ($\otimes$), and Concat. Fusion Block ($\boxplus$) are also defined.

(52) CPC특허분류

G06V 10/14 (2023.08)

G06V 10/56 (2023.08)

G06V 10/60 (2023.08)

G06V 10/809 (2023.08)

B60W 2554/4029 (2020.02)

(72) 발명자

김도경

서울특별시 광진구 광나루로20길 34, 706호(화양동)

김태주

서울특별시 광진구 동일로80길 31-5, 102호(중곡동)

김형준

경기도 수원시 영통구 도청로 10, B동 1225호(이의동)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711179335
과제번호	00156345
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성
연구과제명	ICT혁신인재4.0(세종대학교)
기 여 율	1/2
과제수행기관명	세종대학교산학협력단
연구기간	2023.01.01 ~ 2023.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711198854
과제번호	00262891
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	ICTR&D혁신바우처지원사업(예타)
연구과제명	스마트팜 농작업 자동화 로봇을 위한 인공지능 기반정밀 지도 작성 및 농작물 이미

지 분석 기술 개발

기 여 율	1/2
과제수행기관명	주식회사 량데뷰
연구기간	2023.06.01 ~ 2024.05.31

명세서

청구범위

청구항 1

어텐션 기반 조도 환경에 따른 멀티스펙트럴 보행자 검출 장치로서,

프로세서;

상기 프로세서에 연결되는 메모리를 포함하되,

상기 메모리는,

전방에 대한 컬러 영상 및 열화상 영상을 포함하는 멀티스펙트럴 영상을 입력 받아 상기 컬러 영상 및 열화상 영상 각각에 대한 중단 단계의 제1 피처맵을 추출하고,

상기 중단 단계의 제1 피처맵을 셀프 크로스 어텐션 모듈에 입력하여 상기 컬러 영상 및 열화상 영상 각각에 대한 제2 피처맵을 추출하고,

상기 컬러 영상을 입력 받아 입력 영상의 조도 환경에 따라 낮과 밤으로 분류될 신뢰도를 계산하고,

상기 신뢰도를 상기 컬러 영상 및 열화상 영상 각각에 대한 제2 피처맵에 가중합하여 상기 전방에 위치한 보행자를 검출하도록,

상기 프로세서에 의해 실행되는 프로그램 명령어들을 저장한 멀티스펙트럴 보행자 검출 장치.

청구항 2

제1항에 있어서,

상기 프로그램 명령어들은,

복수의 컨볼루션 레이어 중 일부의 제1 컨볼루션 레이어들을 통해 상기 컬러 영상에 대한 중단 단계의 제1-1 피처맵을 추출하고,

상기 복수의 컨볼루션 레이어 중 일부의 제2 컨볼루션 레이어들을 통해 상기 열화상 영상에 대한 중단 단계의 제1-2 피처맵을 추출하는 멀티스펙트럴 보행자 검출 장치.

청구항 3

제2항에 있어서,

상기 프로그램 명령어들은,

상기 제1-1 피처맵과 상기 제1-2 피처맵을 패치 단위로 분할하여 컬러 영상 임베딩 토큰과 열화상 영상 임베딩 토큰으로 변환하고, 상기 컬러 영상 임베딩 토큰과 열화상 영상 임베딩 토큰을 상기 셀프 크로스 어텐션 모듈에 입력하는 멀티스펙트럴 보행자 검출 장치.

청구항 4

제3항에 있어서,

상기 셀프 크로스 어텐션 모듈은,

상기 컬러 영상 임베딩 토큰을 쿼리(Query)로 입력 받는 제1 셀프 어텐션 모듈 및 상기 제1 셀프 어텐션 모듈에서 추출된 값을 쿼리로 입력 받는 제1 크로스 어텐션 모듈을 포함하고,

상기 열화상 영상 임베딩 토큰을 쿼리로 입력 받는 제2 셀프 어텐션 모듈 및 상기 제2 셀프 어텐션 모듈에서 추출된 값을 쿼리로 입력 받는 제2 크로스 어텐션 모듈을 포함하는 멀티스펙트럴 보행자 검출 장치.

청구항 5

제4항에 있어서,

상기 제1 크로스 어텐션 모듈은, 상기 제2 셀프 어텐션 모듈에서 추출된 값을 키(Key) 및 밸류(Value)로 입력 받고,

상기 제2 크로스 어텐션 모듈은, 상기 제1 셀프 어텐션 모듈에서 추출된 값을 키 및 밸류로 입력 받는 멀티스펙트럴 보행자 검출 장치.

청구항 6

제5항에 있어서,

상기 제1 크로스 어텐션 모듈에서 추출된 값이 N번 반복하여 상기 제1 셀프 크로스 어텐션 모듈에 입력되고,

상기 제2 크로스 어텐션 모듈에서 추출된 값이 N번 반복하여 상기 제2 셀프 크로스 어텐션 모듈에 입력되는 멀티스펙트럴 보행자 검출 장치.

청구항 7

제6항에 있어서,

상기 셀프 크로스 어텐션 모듈에서 상기 컬러 영상의 피처와 상기 열화상 영상의 피처 간의 상호 공간 종속성과 독립 공간 종속성을 명시적으로 학습하여 상기 컬러 영상과 상기 열화상 영상 각각의 제2-1 피처맵 및 제2-2 피처맵을 추출하는 멀티스펙트럴 보행자 검출 장치.

청구항 8

제1항에 있어서,

상기 조도 네트워크는 복수개 층의 컨볼루션 블록으로 구성되고, 각 층의 컨볼루션 블록은 서로 다른 입출력 채널을 갖는 컨볼루션 레이어로 구성되는 멀티스펙트럴 보행자 검출 장치.

청구항 9

제2항에 있어서,

상기 프로그램 명령어들은,

상기 접합된 제2 피처맵을 상기 복수의 컨볼루션 레이어 중 일부의 컨볼루션 레이어들을 통과시켜 상기 전방에 위치한 보행자를 검출하는 멀티스펙트럴 보행자 검출 장치.

청구항 10

프로세서 및 메모리를 포함하는 장치에서 어텐션 기반으로 조도 환경에 따른 멀티스펙트럴 보행자 검출하는 방법으로서,

전방에 대한 컬러 영상 및 열화상 영상을 포함하는 멀티스펙트럴 영상을 입력 받아 상기 컬러 영상 및 열화상 영상 각각에 대한 중간 단계의 제1 피처맵을 추출하는 단계;

상기 중간 단계의 제1 피처맵을 셀프 크로스 어텐션 모듈에 입력하여 상기 컬러 영상 및 열화상 영상 각각에 대한 제2 피처맵을 추출하는 단계;

상기 컬러 영상을 입력 받아 입력 영상의 조도 환경에 따라 낮과 밤으로 분류될 신뢰도를 계산하는 단계; 및

상기 신뢰도를 상기 컬러 영상 및 열화상 영상 각각에 대한 제2 피처맵에 가중합하여 상기 전방에 위치한 보행자를 검출하는 단계를 포함하는 멀티스펙트럴 보행자 검출 방법.

청구항 11

제10항에 있어서,

상기 제1 피처맵을 추출하는 단계는,

복수의 컨볼루션 레이어 중 일부의 제1 컨볼루션 레이어들을 통해 상기 컬러 영상에 대한 중간 단계의 제1-1 피

처맵을 추출하는 단계; 및

상기 복수의 컨볼루션 레이어 중 일부의 제2 컨볼루션 레이어들을 통해 상기 열화상 영상에 대한 중간 단계의 제1-2 피처맵을 추출하는 단계를 포함하는 멀티스펙트럴 보행자 검출 방법.

청구항 12

제11항에 있어서,

상기 제2 피처맵을 추출하는 단계는,

상기 제1-1 피처맵과 상기 제1-2 피처맵을 패치 단위로 분할하여 컬러 영상 임베딩 토큰과 열화상 영상 임베딩 토큰으로 변환하는 단계; 및

상기 컬러 영상 임베딩 토큰과 열화상 영상 임베딩 토큰을 상기 셀프 크로스 어텐션 모듈에 입력하는 단계를 포함하는 멀티스펙트럴 보행자 검출 방법.

청구항 13

제12항에 있어서,

상기 셀프 크로스 어텐션 모듈은,

상기 컬러 영상 임베딩 토큰을 쿼리로 입력 받는 제1 셀프 어텐션 모듈 및 상기 제1 셀프 어텐션 모듈에서 추출된 값을 쿼리로 입력 받는 제1 크로스 어텐션 모듈을 포함하고,

상기 열화상 영상 임베딩 토큰을 쿼리로 입력 받는 제2 셀프 어텐션 모듈 및 상기 제2 셀프 어텐션 모듈에서 추출된 값을 쿼리로 입력 받는 제2 크로스 어텐션 모듈을 포함하는 멀티스펙트럴 보행자 검출 방법.

청구항 14

제13항에 있어서,

상기 제1 크로스 어텐션 모듈은, 상기 제2 셀프 어텐션 모듈에서 추출된 값을 키 값 및 밸류 값으로 입력 받고,

상기 제2 크로스 어텐션 모듈은, 상기 제1 셀프 어텐션 모듈에서 추출된 값을 키 값 및 밸류 값으로 입력 받는 멀티스펙트럴 보행자 검출 방법.

청구항 15

어텐션 기반 조도 환경에 따른 멀티스펙트럴 보행자 검출 장치로서,

전방에 대한 컬러 영상 및 열화상 영상을 포함하는 멀티스펙트럴 영상을 입력 받아 상기 컬러 영상 및 열화상 영상 각각에 대한 중간 단계의 제1 피처맵을 추출하고, 상기 중간 단계의 제1 피처맵을 셀프 어텐션 모듈과 크로스 어텐션 모듈에 순차적으로 입력하여 상기 컬러 영상 및 열화상 영상 각각에 대한 제2 피처맵을 추출하는 검출 네트워크; 및

상기 컬러 영상을 입력 받아 입력 영상의 조도 환경에 따라 낮과 밤으로 분류될 신뢰도를 계산하는 조도 네트워크를 포함하되,

상기 검출 네트워크는 상기 신뢰도를 상기 컬러 영상 및 열화상 영상 각각에 대한 제2 피처맵에 가중합하여 상기 전방에 위치한 보행자를 검출하는 멀티스펙트럴 보행자 검출 장치.

발명의 설명

기술 분야

[0001] 본 발명은 어텐션 기반 조도 환경에 따른 멀티스펙트럴 보행자 검출 방법 및 장치에 관한 것이다.

배경 기술

[0002] 인공 지능과 컴퓨터 비전 기술의 발달로, 딥러닝 기반 도로 위 보행자 검출 연구는 자율주행 자동차와 운전자

보조 시스템 구축에 활용되고 있다.

[0003] 기존 딥러닝 기반 보행자 검출 연구는 주로 컬러 영상만을 사용하는 단일 영상 기반으로 설계되었지만, 밤과 같은 저조도 환경 혹은 역광과 같은 조도 변화에 강인하지 못하다는 단점이 있다. 이러한 단점을 보완하고자 멀티스펙트럴(컬러 + 열화상) 영상 기반 퓨전 알고리즘이 제안되었으며, 컬러 영상과 열화상 영상 각각의 장점을 활용하여 단일 영상 기반 네트워크 대비 밤과 같은 저조도 환경 및 환경 변화에 대한 강인성을 보여 높은 검출 성능을 보인다.

[0004] 하지만 해당 연구들은 합성곱 신경망(CNN) 기반으로 고정된 크기의 커널을 사용하여 컬러 영상 피처와 열화상 영상 피처의 장거리 공간 종속성(long-range dependencies)을 커널 크기의 한정된 공간 내에서만 고려할 수 있다는 단점이 존재한다.

[0005] 해당 합성곱 신경망 기반 네트워크의 한계를 극복하고자 최근 트랜스포머(Transformer)의 셀프 어텐션(Self-attention) 기반 퓨전 네트워크인 CFT(Cross-Modality Fusion Transformer)가 제안되었으며 기존 합성곱 신경망 기반 네트워크 대비 향상된 검출 성능을 보인다. 하지만 CFT는 컬러 영상의 피처와 열화상 영상의 피처를 채널 축으로 결합한 피처에 대해 셀프 어텐션을 적용하여 결합된 피처에 한해서만 장거리 공간 종속성을 고려할 수 있으며 컬러 영상과 열화상 영상 각각의 피처에 대해 장거리 공간 종속성을 고려하지 않는다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) KR 등록특허 10-2354862

발명의 내용

해결하려는 과제

[0007] 상기한 종래기술의 문제점을 해결하기 위해, 본 발명은 컬러 영상 모달리티와 열화상 영상 모달리티의 피처맵 간의 상호(inter-modality) 공간 종속성과 각각 모달리티 내에서의 독립(intra-modality) 공간 종속성을 동시에 고려할 수 있는 어텐션 기반 조도 환경에 따른 멀티스펙트럴 보행자 검출 방법 및 장치를 제안하고자 한다.

과제의 해결 수단

[0008] 상기한 바와 같은 목적을 달성하기 위하여, 본 발명의 일 실시예에 따르면, 어텐션 기반 조도 환경에 따른 멀티스펙트럴 보행자 검출 장치로서, 프로세서; 상기 프로세서에 연결되는 메모리를 포함하되, 상기 메모리는, 전방에 대한 컬러 영상 및 열화상 영상을 포함하는 멀티스펙트럴 영상을 입력 받아 상기 컬러 영상 및 열화상 영상 각각에 대한 중간 단계의 제1 피처맵을 추출하고, 상기 중간 단계의 제1 피처맵을 셀프 크로스 어텐션 모듈에 입력하여 상기 컬러 영상 및 열화상 영상 각각에 대한 제2 피처맵을 추출하고, 상기 컬러 영상을 입력 받아 입력 영상의 조도 환경에 따라 낮과 밤으로 분류될 신뢰도를 계산하고, 상기 신뢰도를 상기 컬러 영상 및 열화상 영상 각각에 대한 제2 피처맵에 가중합하여 상기 전방에 위치한 보행자를 검출하도록, 상기 프로세서에 의해 실행되는 프로그램 명령어들을 저장한 멀티스펙트럴 보행자 검출 장치가 제공된다.

[0009] 상기 프로그램 명령어들은, 복수의 컨볼루션 레이어 중 일부의 제1 컨볼루션 레이어들을 통해 상기 컬러 영상에 대한 중간 단계의 제1-1 피처맵을 추출하고, 상기 복수의 컨볼루션 레이어 중 일부의 제2 컨볼루션 레이어들을 통해 상기 열화상 영상에 대한 중간 단계의 제1-2 피처맵을 추출할 수 있다.

[0010] 상기 프로그램 명령어들은, 상기 제1-1 피처맵과 상기 제1-2 피처맵을 패치 단위로 분할하여 컬러 영상 임베딩 토큰과 열화상 영상 임베딩 토큰으로 변환하고, 상기 컬러 영상 임베딩 토큰과 열화상 영상 임베딩 토큰을 상기 셀프 크로스 어텐션 모듈에 입력할 수 있다.

[0011] 상기 셀프 크로스 어텐션 모듈은, 상기 컬러 영상 임베딩 토큰을 쿼리(Query)로 입력 받는 제1 셀프 어텐션 모듈 및 상기 제1 셀프 어텐션 모듈에서 추출된 값을 쿼리로 입력 받는 제1 크로스 어텐션 모듈을 포함하고, 상기 열화상 영상 임베딩 토큰을 쿼리로 입력 받는 제2 셀프 어텐션 모듈 및 상기 제2 셀프 어텐션 모듈에서 추출된 값을 쿼리로 입력 받는 제2 크로스 어텐션 모듈을 포함할 수 있다.

- [0012] 상기 제1 크로스 어텐션 모듈은, 상기 제2 셀프 어텐션 모듈에서 추출된 값을 키(Key) 및 밸류(Value)로 입력 받고, 상기 제2 크로스 어텐션 모듈은, 상기 제1 셀프 어텐션 모듈에서 추출된 값을 키 및 밸류로 입력 받을 수 있다.
- [0013] 상기 제1 크로스 어텐션 모듈에서 추출된 값이 N번 반복하여 상기 제1 셀프 크로스 어텐션 모듈에 입력되고, 상기 제2 크로스 어텐션 모듈에서 추출된 값이 N번 반복하여 상기 제2 셀프 크로스 어텐션 모듈에 입력될 수 있다.
- [0014] 상기 셀프 크로스 어텐션 모듈에서 상기 컬러 영상의 피쳐와 상기 열화상 영상의 피쳐 간의 상호 공간 종속성과 독립 공간 종속성을 명시적으로 학습하여 상기 컬러 영상과 상기 열화상 영상 각각의 제2-1 피쳐맵 및 제2-2 피쳐맵을 추출할 수 있다.
- [0015] 상기 조도 네트워크는 복수개 층의 컨볼루션 블록으로 구성되고, 각 층의 컨볼루션 블록은 서로 다른 입출력 채널을 갖는 컨볼루션 레이어로 구성될 수 있다.
- [0016] 상기 프로그램 명령어들은, 상기 접합된 제2 피쳐맵을 상기 복수의 컨볼루션 레이어 중 일부의 컨볼루션 레이어들을 통과시켜 상기 전방에 위치한 보행자를 검출할 수 있다.
- [0017] 본 발명의 다른 측면에 따르면, 프로세서 및 메모리를 포함하는 장치에서 어텐션 기반으로 조도 환경에 따른 멀티스펙트럴 보행자 검출하는 방법으로서, 전방에 대한 컬러 영상 및 열화상 영상을 포함하는 멀티스펙트럴 영상을 입력 받아 상기 컬러 영상 및 열화상 영상 각각에 대한 중단 단계의 제1 피쳐맵을 추출하는 단계; 상기 중단 단계의 제1 피쳐맵을 셀프 크로스 어텐션 모듈에 입력하여 상기 컬러 영상 및 열화상 영상 각각에 대한 제2 피쳐맵을 추출하는 단계; 상기 컬러 영상을 입력 받아 입력 영상의 조도 환경에 따라 낮과 밤으로 분류될 신뢰도를 계산하는 단계; 및 상기 신뢰도를 상기 컬러 영상 및 열화상 영상 각각에 대한 제2 피쳐맵에 가중합하여 상기 전방에 위치한 보행자를 검출하는 단계를 포함하는 멀티스펙트럴 보행자 검출 방법이 제공된다.
- [0018] 본 발명의 또 다른 측면에 따르면, 어텐션 기반 조도 환경에 따른 멀티스펙트럴 보행자 검출 장치로서, 전방에 대한 컬러 영상 및 열화상 영상을 포함하는 멀티스펙트럴 영상을 입력 받아 상기 컬러 영상 및 열화상 영상 각각에 대한 중단 단계의 제1 피쳐맵을 추출하고, 상기 중단 단계의 제1 피쳐맵을 셀프 어텐션 모듈과 크로스 어텐션 모듈에 순차적으로 입력하여 상기 컬러 영상 및 열화상 영상 각각에 대한 제2 피쳐맵을 추출하는 검출 네트워크; 및 상기 컬러 영상을 입력 받아 입력 영상의 조도 환경에 따라 낮과 밤으로 분류될 신뢰도를 계산하는 조도 네트워크를 포함하되, 상기 검출 네트워크는 상기 신뢰도를 상기 컬러 영상 및 열화상 영상 각각에 대한 제2 피쳐맵에 가중합하여 상기 전방에 위치한 보행자를 검출하는 멀티스펙트럴 보행자 검출 장치가 제공된다.

발명의 효과

- [0019] 본 발명에 따르면, 멀티스펙트럴 영상의 상호 공간 종속성을 명시적으로 학습하며 동시에 영상의 조도 환경을 고려한 융합 방식을 통해 효과적인 보행자 탐지가 가능한 장점이 있다.

도면의 간단한 설명

- [0020] 도 1은 본 실시예에 따른 셀프-크로스 어텐션 기반 멀티스펙트럴 보행자 검출 네트워크의 구조를 도시한 도면이다.
- 도 2는 본 실시예에 따른 셀프 크로스 어텐션 모듈의 상세 구성을 도시한 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0021] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.
- [0022] 본 명세서에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요

소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

- [0023] 또한, 각 도면을 참조하여 설명하는 실시예의 구성 요소가 해당 실시예에만 제한적으로 적용되는 것은 아니며, 본 발명의 기술적 사상이 유지되는 범위 내에서 다른 실시예에 포함되도록 구현될 수 있으며, 또한 별도의 설명이 생략될지라도 복수의 실시예가 통합된 하나의 실시예로 다시 구현될 수도 있음은 당연하다.
- [0024] 또한, 첨부 도면을 참조하여 설명함에 있어, 도면 부호에 관계없이 동일한 구성 요소는 동일하거나 관련된 참조 부호를 부여하고 이에 대한 중복되는 설명은 생략하기로 한다. 본 발명을 설명함에 있어서 관련된 공지 기술에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다.
- [0026] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.
- [0027] 본 명세서에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 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.
- [0028] 또한, 각 도면을 참조하여 설명하는 실시예의 구성 요소가 해당 실시예에만 제한적으로 적용되는 것은 아니며, 본 발명의 기술적 사상이 유지되는 범위 내에서 다른 실시예에 포함되도록 구현될 수 있으며, 또한 별도의 설명이 생략될지라도 복수의 실시예가 통합된 하나의 실시예로 다시 구현될 수도 있음은 당연하다.
- [0029] 또한, 첨부 도면을 참조하여 설명함에 있어, 도면 부호에 관계없이 동일한 구성 요소는 동일하거나 관련된 참조 부호를 부여하고 이에 대한 중복되는 설명은 생략하기로 한다. 본 발명을 설명함에 있어서 관련된 공지 기술에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다.
- [0031] 본 실시예는 전방에 관한 컬러 영상 모달리티와 열화상 영상 모달리티의 피쳐맵 간의 상호(inter-modality) 공간 종속성과 각각 모달리티 내에서의 독립(intra-modality) 공간 종속성을 동시에 고려하고자 트랜스포머에서 제안된 셀프 어텐션(Self-attention)과 이를 응용하여 설계한 크로스 어텐션(Cross-attention)을 기반으로 하는 셀프-크로스 어텐션 기반 멀티스펙트럴 보행자 검출 네트워크를 제안한다.
- [0032] 본 실시예에 따른 네트워크는 멀티스펙트럴 각 모달리티 피쳐에 대해 독립적으로 셀프 어텐션 연산을 적용하여 단일 모달리티 내에서의 독립 공간 종속성을 학습한다. 이후 셀프 어텐션 연산이 적용된 컬러 영상 피쳐와 열화상 영상 피쳐에 대해 크로스 어텐션 연산을 적용하여 두 피쳐맵 간의 상호 공간 종속성을 학습하는 방식을 사용한다.
- [0033] 또한 검출 네트워크 외에도 컬러 영상의 조도 환경에 따라 입력 영상을 낮과 밤으로 분류하는 서브 네트워크(조도 네트워크, Illuminance-Network; IN)를 제안한다. IN은 입력된 컬러 영상에 대해 낮과 밤의 신뢰도 점수를 각각 컬러 영상 피쳐와 열화상 영상 피쳐에 가중합 하는 방식을 사용하여 검출 네트워크가 영상의 조도 특성을 활용할 수 있도록 한다.
- [0034] 도 1은 본 실시예에 따른 셀프-크로스 어텐션 기반 멀티스펙트럴 보행자 검출 네트워크의 구조를 도시한 도면이다.
- [0035] 본 실시예에 따른 멀티스펙트럴 보행자 검출은 프로세서 및 메모리를 포함하는 장치에서 수행될 수 있다.
- [0036] 여기서, 프로세서는 컴퓨터 프로그램을 실행할 수 있는 CPU(central processing unit)나 그 밖에 가상 머신 등을 포함할 수 있다.
- [0037] 메모리는 고정식 하드 드라이브나 착탈식 저장 장치와 같은 불휘발성 저장 장치를 포함할 수 있다. 착탈식 저장 장치는 콤팩트 플래시 유닛, USB 메모리 스틱 등을 포함할 수 있다. 메모리는 각종 랜덤 액세스 메모리와 같은 휘발성 메모리도 포함할 수 있으며, 컴퓨터 판독 가능한 기록매체로 정의될 수 있다.

- [0038] 본 실시예에 따른 메모리에는 셀프-크로스 어텐션 기반으로 멀티스펙트럴 보행자 검출을 위한 프로그램 명령어들이 저장된다.
- [0039] 이하에서는 설명하는 멀티스펙트럴 보행자 검출 과정은 상기한 프로그램 명령어들에 의해 수행되는 과정으로 이해되어야 할 것이다.
- [0040] 도 1을 참조하면, 본 실시예에 따른 네트워크는 컬러 영상과 열화상 영상을 입력으로 하는 검출 네트워크(100)와 컬러 영상을 입력으로 하는 조도 네트워크(102)를 포함한다.
- [0041] 본 실시예에 따른 검출 네트워크는 VGG-16을 백본 네트워크로 할 수 있으며, 복수의 컨볼루션 레이어(Conv1 내지 Conv11) 및 셀프 크로스 어텐션 모듈(Self-Cross Attention; SCA, 104)을 포함한다.
- [0042] 본 실시예에 따른 복수의 컨볼루션 레이어는 컬러 영상을 입력으로 하여 컬러 영상에 대한 중간 단계의 피쳐맵을 출력하는 제1 컨볼루션 레이어들(110) 및 열화상 영상을 입력으로 하여 열화상 영상에 대한 중간 단계의 피쳐맵을 출력하는 제2 컨볼루션 레이어들(112)을 포함할 수 있다.
- [0043] 여기서, 중간 단계의 피쳐맵은 제1 피쳐맵으로 정의될 수 있고, 컬러 영상에 대한 중간 단계의 피쳐맵은 제1-1 피쳐맵, 열화상 영상에 대한 중간 단계의 피쳐맵은 제1-2 피쳐맵으로 표기한다.
- [0044] 본 실시예에 따르면, 제1 및 제2 컨볼루션 레이어들(110, 112)의 말단 컨볼루션 레이어(Conv4)가 출력하는 제1 피쳐맵을 본 실시예에 따른 셀프-크로스 어텐션 모듈(Self-Cross Attention; SCA, 104)에 통과시킨다.
- [0045] 본 실시예에 따르면, 제1-1 피쳐맵 및 제1-2 피쳐맵 각각이 셀프 크로스 어텐션 모듈에 입력된다.
- [0046] 조도 네트워크(102)는 입력 영상을 조도 환경에 따라 낮과 밤으로 분류될 신뢰도를 구한다. 이후 신뢰도를 가중치로 두어 SCA(104)를 통과한 피쳐맵(제2 피쳐맵)과 가중합하여 집합하는 방식을 사용한다.
- [0047] 본 실시예에 따른 SCA(104)는 컬러 영상 피쳐와 열화상 영상 피쳐 간의 상호 공간 종속성과 독립 공간 종속성을 명시적으로 학습하는 것을 목표로 한다.
- [0048] 도 2는 본 실시예에 따른 셀프 크로스 어텐션 모듈의 상세 구성을 도시한 도면이다.
- [0049] 도 2를 참조하면, SCA(104)는 컬러 영상 임베딩 토큰과 열화상 영상 임베딩 토큰을 입력으로 받는다. 임베딩 토큰을 생성하는 과정은 CFT와 동일하며, 제1 컨볼루션 레이어(110)들의 말단 컨볼루션 레이어로부터 추출한 컬러 영상에 대한 제1-1 피쳐맵과 열화상 영상에 대한 제1-2 피쳐맵을 패치 단위로 분할한 후 채널 축으로 평탄화하는 방식(flatten)으로 생성한다.
- [0050] 도 1 및 아래의 수학적 식 1 내지 3에 표현된 컬러 영상 임베딩 토큰 R과 열화상 영상 임베딩 토큰 T는 셀프 어텐션 모듈 τ_i^{self} 와 크로스 어텐션 모듈 τ_i^{cross} 의 결합으로 구성된 SCA(104)를 N번 반복적으로 통과하며 상호 공간 종속성과 독립 공간 종속성을 모델 내에서 학습한다.

수학적 식 1

$$R'_i = \begin{cases} \tau_i^{self}(R + PE, R + PE) & (i = 1) \\ \tau_i^{self}(R_i, R_i) & (i > 1) \end{cases}$$

[0051]

수학적 식 2

$$T'_i = \begin{cases} \tau_i^{self}(T + PE, T + PE) & (i = 1) \\ \tau_i^{self}(T_i, T_i) & (i > 1) \end{cases}$$

[0052]

- [0053] 여기서, i는 어텐션 모듈의 인덱스를 의미하며, 셀프 어텐션 τ_i^{self} 와 크로스 어텐션 τ_i^{cross} 은 동일하게 2

개의 입력값을 받는다.

[0054] 본 실시예에 따르면, 셀프 셀프 크로스 어텐션 모듈은, 컬러 영상에 관한 제1 셀프 어텐션 모듈(200)과 제1 크로스 어텐션 모듈(202) 및 열화상 영상에 관한 제2 셀프 어텐션 모듈(210) 및 제2 크로스 어텐션 모듈(212)을 포함할 수 있다.

[0055] 제1 셀프 어텐션 모듈(200)은 컬러 영상 임베딩 토큰을 첫 번째 입력값인 쿼리(Query)로 입력 받으며, 동일한 토큰을 두 번째 입력값인 키(key)와 벨류(value)로 입력 받는다.

[0056] 또한, 제2 셀프 어텐션 모듈(210)은 열화상 영상 임베딩 토큰을 첫 번째 입력값인 쿼리로 입력 받으며, 동일한 토큰을 두 번째 입력값인 키와 벨류로 입력 받는다.

[0057] 상기한 바와 같이, 각 셀프 어텐션 모듈에서 첫 번째 입력값은 쿼리로 사용되고, 두 번째 입력값은 키와 벨류로 사용된다. 구체적으로 $i=1$ 일 때 임베딩 토큰 R과 T에 위치 정보를 갖는 인코딩 벡터(Positional Encoding; PE)를 더한 값을 τ_i^{self} 의 입력으로 사용한다.

[0058] $i > 1$ 일 때는 τ_i^{cross} 를 통과하고 나온 출력을 다시 입력으로 하며, 이에 대해서는 이하에서 다시 설명한다.

[0059] i 번째 셀프 어텐션 모듈(200, 212)을 통과하고 나온 R'_i 와 T'_i 은 동일한 인덱스 i 를 가지는 크로스 어텐션 모듈(202, 212) τ_i^{cross} 의 입력값으로 사용되며, 아래와 같이 수식화된다.

수학식 3

$$\begin{cases} R_{i+1} = \tau_i^{cross}(R'_i, T'_i) \\ T_{i+1} = \tau_i^{cross}(T'_i, R'_i) \end{cases}$$

[0060]

[0061] 앞선 셀프 어텐션 모듈에서의 과정과 유사하게 셀프 어텐션 모듈을 통과하고 나온 R'_i 와 T'_i 은 크로스 어텐션 모듈 τ_i^{cross} 의 입력값으로 사용되며, 이 때 첫 번째 입력값은 쿼리로 사용되고 두 번째 입력값은 키와 벨류로 사용된다.

[0062] 다만 차이점은 앞서 셀프 어텐션에서는 두 개의 동일한 모달리티에 관한 입력값을 받았다면, 크로스 어텐션에서는 입력값이 서로 다른 모달리티에 관한 입력값을 받는다.

[0063] 본 실시예에 따르면, 컬러 영상에 관한 제1 크로스 어텐션 모듈(202)은, 컬러 영상에 관한 제1 셀프 어텐션 모듈(200)에서 추출된 값을 쿼리로 입력 받고, 열화상 영상에 관한 제2 셀프 어텐션 모듈(212)에서 추출된 값을 키 및 벨류로 입력 받는다.

[0064] 또한, 제2 크로스 어텐션 모듈(212)은, 열화상에 관한 제2 셀프 어텐션 모듈(210)에서 추출된 값을 쿼리로 입력 받고, 컬러 영상에 관한 제1 셀프 어텐션 모듈(200)에서 추출된 값을 키 및 벨류로 입력 받는다.

[0065] 컬러 영상에 관한 크로스 어텐션 모듈의 첫 번째 입력값은 컬러 영상에 관한 R'_i 이고, 두 번째 입력값은 열화상 영상에 관한 T'_i 가 된다.

[0066] 또한, 열화상 영상에 관한 크로스 어텐션 모듈의 첫 번째 입력값은 열화상 영상에 관한 T'_i 이고, 두 번째 입력값은 컬러 영상에 관한 R'_i 이 된다.

[0067] 해당 차이점으로 인해 셀프 어텐션 모듈에서는 독립 공간 종속성을, 크로스 어텐션 모듈에서는 상호 공간 종속성을 보다 명시적으로 학습한다.

[0068] 크로스 어텐션 모듈 τ_i^{cross} 을 통과한 R_{i+1} 과 T_{i+1} 는 다음 인덱스 $i+1$ 의 셀프 어텐션 τ_{i+1}^{self} 의 입력으로 사용되며 위의 과정은 N번 반복된다. 이때 N은 GMFlow를 참고하여 메모리 사용량과 연산 속도 간의 상충 관계를 고려하여 최적의 값인 2로 설정될 수 있다. 아래의 수학적식 4 내지 5의 M은 조도 네트워크이다.

수학적식 4

$$F_{hard} = \begin{cases} F_R & (M(I_R) \geq 0.5) \\ F_T & (M(I_R) < 0.5) \end{cases}$$

수학적식 5

$$F_{soft} = F_R \cdot M(I_R) + F_T \cdot (1 - M(I_R))$$

[0071] M은 연속적인 조도 환경에 대한 가중치를 입력 영상에 부여하여 조도 변화에 강인성을 보이게끔 제안된다.

[0072] M은 세 층의 컨볼루션 블록으로, 한 층의 컨볼루션 블록은 서로 다른 입출력 채널을 갖는 컨볼루션 레이어로 구성된다.

[0073] M의 컨볼루션 레이어는 순서대로 64, 128, 256 채널의 피쳐맵을 생성하며, 열화상 영상 대비 색상 정보가 풍부한 3채널의 컬러 영상 I_R 을 입력을 받아 낮과 밤에 대한 분류 신뢰도인 $M(I_R)$ 과 $1 - M(I_R)$ 을 구한다.

[0074] M은 16 배치 사이즈에서 학습률 0.0001의 최적화 함수 Adam을 사용하여 3 에포크 학습하였으며, KAIST 멀티스펙트럴 보행자 데이터 셋에서 낮과 밤 영상 각각 3000장을 선별하여 구축한 데이터 셋을 통해 사전 학습하여 사용하였다.

[0075] 수학적식 4 내지 5의 F_R , F_T 는 각각 멀티스펙트럴 네트워크에서 상기한 제1 및 제2 컨볼루션 레이어들 (110,112)의 말단 컨볼루션 레이어(Conv4_3)를 통과한 컬러 영상과 열화상 영상의 피쳐맵을 의미한다.

[0076] 본 실시예에 따르면, 표 1의 조도 환경별 검출 성능 실험 비교를 토대로 영상 접합(Concatenate)을 F_R 과 F_T 는 $M(I_R)$ 과 $1 - M(I_R)$ 에 접합하는 방식을 사용한다.

표 1

	All	Day	Night
SSD-R	37.37	34.49	44.05
SSD-T	33.57	40.46	19.56

[0078] 영상 접합은 하드 라벨(Hard-Label)과 소프트 라벨(Soft-Label)을 사용한 방식으로 나눌 수 있다.

[0079] 수학적식 4의 하드 라벨을 사용한 영상 접합 방식은 $M(I_R)$ 의 임계값 0.5를 기준으로 임계값이 넘는 경우 접합하는 피쳐맵으로 F_R 을 선택하며, 넘지 않는 경우 F_T 를 선택하는 것으로 F_{hard} 를 정한다.

[0080] 반대로 소프트 라벨을 사용한 영상 접합은 $M(I_R)$ 과 $1 - M(I_R)$ 을 F_R 과 F_T 의 가중치로 두어 각각의 피

처맵에 가중합하는 방식을 사용하여 F_{soft} 를 정한다.

[0081] 이후 M을 통해 나온 피쳐맵은 SCA(104)를 통과한 피쳐맵과 접합한 다음, 피쳐맵의 채널을 1x1 합성곱 신경망(NiN)을 통해 1/3 비율로 줄인다.

[0082] 본 실시예에서는 표 2의 하드 라벨과 소프트 라벨의 영상 접합 방식에 따른 검출 성능 비교를 토대로 24시간의 연속된 조도 변화의 영상 도메인을 낮과 밤으로 이진 분류하였을 때 발생할 수 있는 양자화 오차로 인한 검출 네트워크의 오차를 줄이고자 소프트 라벨을 사용한 영상 접합 방식 F_{soft} 를 사용한다.

표 2

	All	Day	Night
SSD-SCA	21.17	26.39	10.71
Hard-Label	20.89	25.39	11.86
Ours(Soft-Label)	18.81	22.96	10.39

[0083]

[0084] 다시 도 1을 참조하면, 상기한 셀프 크로스 어텐션 모듈(104)에서 컬러 영상의 피쳐와 상기 열화상 영상의 피쳐 간의 상호 공간 종속성과 독립 공간 종속성을 명시적으로 학습하여 컬러 영상과 열화상 영상 각각의 제2-1 피쳐맵 및 제2-2 피쳐맵을 추출하며, 복수개 층의 컨볼루션 블록으로 구성되는 조도 네트워크(102)에서 계산된 신뢰도를 통해 제2-1 피쳐맵 및 제2-2 피쳐맵을 가중합하여 검출 네트워크(100)의 복수의 컨볼루션 레이어 중 나머지 컨볼루션 레이어들(Conv5 내지 Conv11, 114)을 통과시켜 전방에 위치한 보행자를 검출하게 된다.

[0085] 이하에서는 본 실시예에 따른 손실 함수를 상세하게 설명한다.

[0086] 본 실시예에서는 수학식 6과 같은 SSD 손실 함수를 사용한다.

수학식 6

$$\mathcal{L}(x, c, l, g) = \frac{1}{N} \left(\mathcal{L}_{conf}(x, c) + \mathcal{L}_{loc}(x, l, g) \right)$$

[0087]

[0088] $\mathcal{L}(x, c, l, g)$ 은 두 손실 함수 $\mathcal{L}_{conf}$ 와 $\mathcal{L}_{loc}$ 의 합을 정답 영역 박스와 같은 클래스로 분류된 영역 예측 박스의 수인 N으로 나누어 전체 손실 함수로 정의한다.

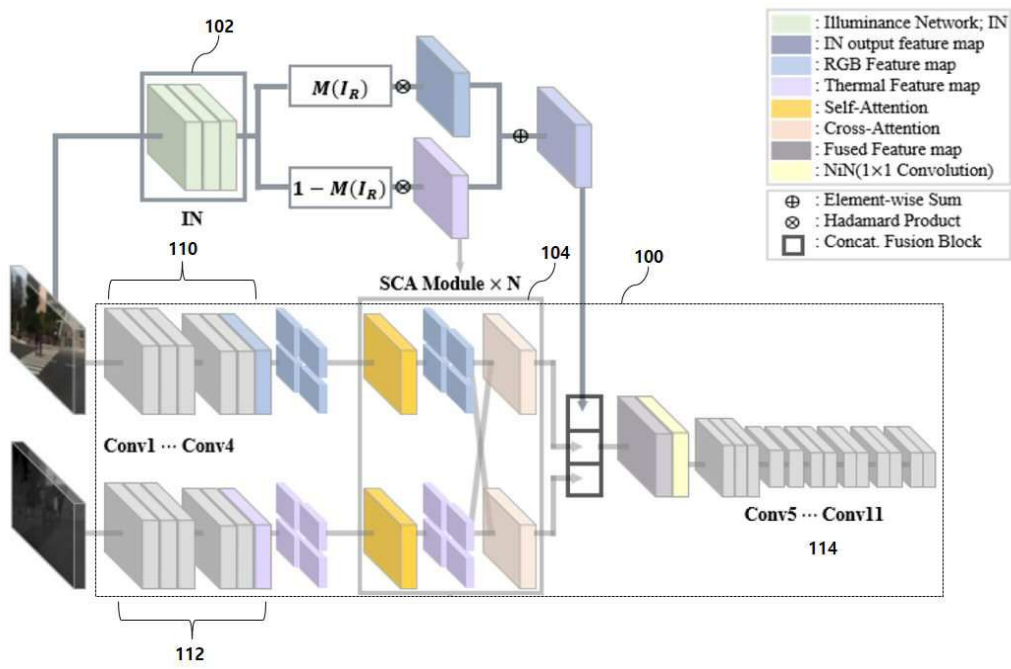
[0089] $\mathcal{L}_{conf}$ 는 보행자 클래스와 보행자가 아닌 배경 클래스의 두 클래스를 예측하는 신뢰도 c에 대한 크로스 엔트로피 손실 함수이며, 영역 예측 박스와 정답 영역 박스 간 클래스 유사도 x가 임계값을 넘는 경우 두 박스는 같은 클래스로 매칭된다. $\mathcal{L}_{conf}$ 는 매칭된 박스와 매칭되지 않은 박스 모두에 대해 손실 값을 계산한다.

[0090] $\mathcal{L}_{loc}$ 은 예측 영역 박스 l와 정답 영역 박스 g가 얼마나 가깝게 위치하고 있는지를 나타내는 Smooth L1손실 함수이다. $\mathcal{L}_{loc}$ 은 영역 예측 박스의 크기로 정규화 된 정답 영역 박스의 오프셋과 예측 영역 박스의 오프셋 차이를 이용하여 손실 값을 계산한다.

[0091] 상기한 본 발명의 실시예는 예시의 목적을 위해 개시된 것이고, 본 발명에 대한 통상의 지식을 가지는 당업자라면 본 발명의 사상과 범위 안에서 다양한 수정, 변경, 부가가 가능할 것이며, 이러한 수정, 변경 및 부가는 하기의 특허청구범위에 속하는 것으로 보아야 할 것이다.

도면

도면1



도면2

