(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0090642
(43) 공개일자 2023년06월22일

(51) 국제특허분류(Int. Cl.)
G06V 40/16 (2022.01) *G06N 3/04* (2023.01)
G06N 3/08 (2023.01) *G06T 5/50* (2006.01)
G06V 10/46 (2022.01) *G06V 10/48* (2022.01)
G06V 10/62 (2022.01) *G06V 10/74* (2022.01)
(52) CPC특허분류
G06V 40/173 (2022.01)
G06N 3/045 (2023.01)
(21) 출원번호 10-2021-0179580
(22) 출원일자 2021년12월15일
심사청구일자 2021년12월15일

(71) 출원인
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
함범섭
서울특별시 강남구 압구정로61길 37, 72동 506호 (압구정동, 한양아파트)
엄찬호
서울특별시 마포구 백범로 230, 102동 2203호(신공덕동, 브라운스톤 공덕 아파트)
(뒷면에 계속)
(74) 대리인
민영준

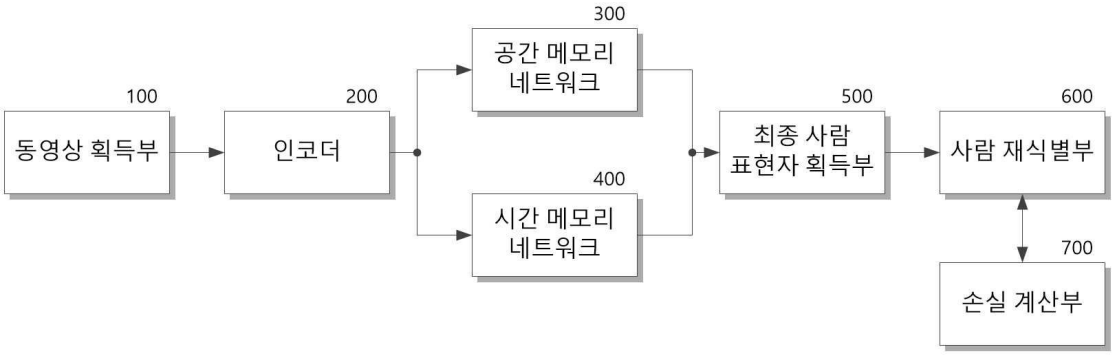
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 **시공간 메모리 네트워크를 활용한 동영상 기반 사람 재식별 장치 및 방법**

(57) 요 약

본 발명은 공간 메모리 네트워크를 이용하여 다수의 프레임 각각에서 공간적 산만 요인을 제거하고, 시간 메모리 네트워크를 이용하여 각 프레임에 대한 중요도를 가중하여 사람 표현자를 추출하여 재식별하므로, 다양한 환경에서 획득된 동영상에서도 정확하게 사람을 재식별할 수 있어 사람 재식별 성능을 크게 향상시킬 수 있는 사람 재식별 장치 및 방법을 제공한다.

대 표 도 - 도2



(52) CPC특허분류
G06N 3/08 (2023.01)
G06T 5/50 (2023.01)
G06V 10/469 (2023.01)
G06V 10/48 (2023.01)
G06V 10/62 (2023.01)
G06V 10/761 (2023.01)
G06T 2207/30201 (2013.01)
(72) 발명자
이건
서울특별시 강서구 화곡로27길 41, 304호(화곡동, 진흥오피스텔)

이중협
서울특별시 서대문구 연희로8길 26, 407호(연희동)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711132436
과제번호	2018M3E3A1057289
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	실종아동등신원확인을위한복합인지기술개발(R&D)(과기정통부)
연구과제명	이종 CCTV 영상에서의 딥러닝 기반 실종자 초동 신원확인 및 추적 시스템
기 여 율	1/1
과제수행기관명	연세대학교 산학협력단
연구기간	2021.01.01 ~ 2021.12.31

명 세 서

청구범위

청구항 1

다수의 프레임으로 구성된 동영상으로 인가받아, 다수의 프레임 각각에 대해 미리 학습된 인공 신경망으로 신경망 연산하여 다수의 공간 쿼리맵, 다수의 시간 쿼리맵 및 다수의 사람 특징맵을 추출하는 인코더;

상기 다수의 공간 쿼리맵 각각에서 채널 방향의 다수의 공간 픽셀 벡터를 추출하고, 공간적 산만 요인에 대한 특징과 대표 표현자가 각각 공간 키와 공간 값으로 서로 매칭되어 미리 저장된 다수의 공간 아이템에서 다수의 공간 키 각각과 다수의 공간 픽셀 벡터 사이의 유사도에 따른 공간 가중치를 대응하는 공간 값에 가중합하여 재배치함으로써 다수의 공간 강조맵을 획득하는 공간 메모리 네트워크;

신경망 연산으로 상기 다수의 시간 쿼리맵의 시간적 변화에 따른 시간 패턴 특징을 나타내는 시간 문맥 벡터를 획득하고, 시간 패턴 특징과 이에 따른 주의도가 각각 시간 키와 시간 값으로 서로 매칭되어 미리 저장된 다수의 시간 아이템에서 상기 다수의 시간 키 각각과 상기 시간 문맥 벡터 사이의 유사도에 따른 시간 가중치를 대응하는 시간 값에 가중합하여 시간 강조 벡터를 획득하는 시간 메모리 네트워크; 및

상기 다수의 사람 특징맵을 대응하는 다수의 공간 강조맵으로 정제하고, 시간 강조 벡터의 대응하는 원소로 강조하여 사람 표현자를 획득하는 사람 표현자 획득부를 포함하는 사람 재식별 장치.

청구항 2

제1항에 있어서, 상기 공간 메모리 네트워크는

상기 다수의 공간 쿼리맵 각각에서 채널 방향의 다수의 공간 픽셀 벡터를 추출하는 공간 벡터 선택부;

각각 상기 공간 키와 상기 공간 값이 서로 매칭되어 포함된 상기 다수의 공간 아이템이 미리 저장된 공간 메모리;

상기 공간 메모리에서 다수의 공간 키를 리드하여, 상기 다수의 공간 픽셀 벡터 각각과 상기 다수의 공간 키 각각 사이의 유사도에 따른 공간 가중치를 계산하는 공간 유사도 계산부; 및

상기 공간 메모리에서 다수의 공간 값을 리드하고, 상기 다수의 공간값 각각에 대응하는 공간 가중치를 가중합하여 다수의 공간 강조 벡터를 획득하고, 획득된 공간 강조 벡터를 대응하는 다수의 공간 픽셀 벡터의 위치에 재배치하여 상기 다수의 공간 쿼리맵에 각각 대응하는 다수의 공간 강조맵을 획득하는 공간 강조맵 획득부를 포함하는 사람 재식별 장치.

청구항 3

제2항에 있어서, 상기 시간 메모리 네트워크는

상기 다수의 시간 쿼리맵 각각에 대해 글로벌 평균 풀링하여, 다수의 시간 쿼리 벡터를 획득하는 시간 쿼리 풀링부;

미리 학습된 인공 신경망으로 구현되어, 상기 다수의 시간 쿼리 벡터에 대해 신경망 연산하여 시간적 변화 특징을 나타내는 상기 시간 문맥 벡터를 획득하는 시간적 문맥 인코딩부;

각각 상기 시간 키와 상기 시간 값이 서로 매칭되어 포함된 상기 다수의 시간 아이템이 미리 저장된 시간 메모리;

상기 시간 메모리에서 다수의 시간 키를 리드하여, 상기 시간 문맥 벡터와 상기 다수의 시간 키 각각 사이의 유사도에 따른 시간 가중치를 계산하는 시간 유사도 계산부; 및

상기 시간 메모리에서 다수의 시간 값을 리드하고, 상기 다수의 시간값 각각에 대응하는 시간 가중치를 가중합하여 상기 시간 강조 벡터를 획득하는 시간 강조 벡터 획득부를 포함하는 사람 재식별 장치.

청구항 4

제3항에 있어서, 상기 시간적 문맥 인코딩부는

LSTM(Long Short Term Memory)으로 구현되는 사람 재식별 장치.

청구항 5

제3항에 있어서, 상기 사람 표현자 획득부는

상기 다수의 사람 특징맵과 상기 다수의 공간 강조맵을 인가받아, 상기 다수의 사람 특징맵 각각에 대해 대응하는 공간 강조맵을 차감하여 다수의 정제 사람 특징맵을 획득하는 공간적 사람 특징맵 정제부;

상기 다수의 정제 사람 특징맵을 인가받아 각각 글로벌 풀링하여 다수의 정제 사람 특징 벡터를 획득하는 정제 사람 특징맵 풀링부; 및

상기 다수의 정제 사람 특징 벡터와 상기 시간 강조 벡터를 인가받고, 상기 다수의 정제 사람 특징 벡터 각각에 상기 시간 강조 벡터의 대응하는 원소를 가중하고 결합하여 상기 사람 표현자를 획득하는 시간적 강조부를 포함하는 사람 재식별 장치.

청구항 6

제3항에 있어서, 상기 사람 재식별 장치는

미리 학습된 인공 신경망으로 구현되어, 상기 사람 표현자를 신경망 연산으로 분류하여 상기 동영상에 포함된 사람에 대한 식별자를 추출하는 사람 재식별부를 더 포함하는 사람 재식별 장치.

청구항 7

제3항에 있어서, 상기 공간 유사도 계산부는

상기 다수의 프레임 중 i번째 프레임에서 획득된 공간 쿼리맵(q^s_i)에서 추출된 k번째 공간 픽셀 벡터($q^s_{i,k}$)와 M개 공간 아이템 중 n번째 공간 아이템의 공간 키(k^s_n) 사이의 유사도에 따른 공간 가중치($a^s_{i,k,n}$)를 수학식

$$a^{s}_{i,k,n} = \frac{\exp\left(\left(\mathbf{q}^{s}_{i,k}\right)^{\mathrm{T}} \mathbf{k}^{s}_{n}\right)}{\sum_{n'=1}^{M} \exp\left(\left(\mathbf{q}^{s}_{i,k}\right)^{\mathrm{T}} \mathbf{k}^{s}_{n'}\right)}$$

에 따라 계산하고,

상기 공간 강조맵 획득부는

상기 공간 가중치($a^s_{i,k,n}$)를 대응하는 n번째 공간 값(v^s_n) 각각에 가중합하여 상기 다수의 공간 강조 벡터($o^s_{i,k}$) 각각을 수학식

$$\mathbf{o}^{s}_{i,k} = \sum_{n=1}^{M} a^{s}_{i,k,n} \mathbf{v}^{s}_{n}$$

에 따라 계산하여 획득하는 사람 재식별 장치.

청구항 8

제3항에 있어서, 상기 시간 유사도 계산부는

상기 다수의 시간 쿼리맵을 글로벌 평균 풀링하고 신경망 연산하여 획득된 상기 시간 문맥 벡터(q^t)와 N개 시간 아이템 중 n번째 시간 아이템의 시간 키(k^t_n) 사이의 유사도에 따른 시간 가중치(a^t_n)를 수학식

$$a_n^{\mathrm{t}} = \frac{\exp((\mathbf{q}^{\mathrm{t}})^{\mathrm{T}}\,\mathbf{k}_n^{\mathrm{t}})}{\sum_{n'=1}^{N}\exp((\mathbf{q}^{\mathrm{t}})^{\mathrm{T}}\,\mathbf{k}_{n'}^{\mathrm{t}})}$$

에 따라 계산하고,

상기 시간 강조맵 획득부는

상기 시간 가중치(a^{t}_{n})를 대응하는 n번째 시간 키(v^{t}_{n}) 각각에 가중합하여 상기 시간 강조 벡터(o^{t})를 수학식

$$\mathbf{o}^{\mathrm{t}} = \sum_{n=1}^{N} a_n^{\mathrm{t}}\,\mathbf{v}_n^{\mathrm{t}}$$

에 따라 계산하여 획득하는 사람 재식별 장치.

청구항 9

제6항에 있어서, 상기 사람 재식별 장치는

학습 시에 구비되어 상기 공간 메모리에 저장된 상기 다수의 공간 아이템과 상기 시간 메모리에 저장된 상기 다수의 시간 아이템의 분포에 따른 메모리 확산 손실과 추출된 사람에 대한 식별자에 따른 식별 손실을 기지정된 교차 엔트로피와 삼중항 손실로 계산하며, 계산된 상기 메모리 확산 손실과 상기 식별 손실의 합으로 계산되는 총 손실을 역전파하는 손실 계산부를 더 포함하는 사람 재식별 장치.

청구항 10

제9항에 있어서, 상기 메모리 확산 손실은

기지정된 크기의 미니 배치 내 다수의 프레임에서 추출된 다수의 공간 쿼리맵(q^{s}_{i}) 또는 다수의 시간 쿼리맵(q^{t}_{i})과 다수의 공간 키(k^{s}) 또는 다수의 시간 키(k^{t}) 각각에 대해 계산된 다수의 공간 가중치(a^{s}_{n}) 또는 시간 가중치(a^{t}_{n})의 최대값과 최소값 사이의 편차로 계산되는 확산 분포가 기지정된 기준 분포(α) 이상이 되도록 수학식

$$\mathcal{L}_S = \sum_{n=1}^{M}[\,min(\mathbf{a}_n^{\mathrm{s}}) - max(\mathbf{a}_n^{\mathrm{s}}) + \alpha\,]_+ + [\,min(\mathbf{a}_n^{\mathrm{t}}) - max(\mathbf{a}_n^{\mathrm{t}}) + \alpha\,]_+$$

에 따라 계산되는 사람 재식별 장치.

청구항 11

다수의 프레임으로 구성된 동영상으로 인가받아, 다수의 프레임 각각에 대해 미리 학습된 인공 신경망으로 신경망 연산하여 다수의 공간 쿼리맵, 다수의 시간 쿼리맵 및 다수의 사람 특징맵을 추출하는 단계;

상기 다수의 공간 쿼리맵 각각에서 채널 방향의 다수의 공간 픽셀 벡터를 추출하고, 공간적 산만 요인에 대한 특징과 대표 표현자가 각각 공간 키와 공간 값으로 서로 매칭되어 미리 저장된 다수의 공간 아이템에서 다수의 공간 키 각각과 다수의 공간 픽셀 벡터 사이의 유사도에 따른 공간 가중치를 대응하는 공간 값에 가중합하여 재배치함으로써 다수의 공간 강조맵을 획득하는 단계;

인공 신경망을 이용하여 신경망 연산으로 상기 다수의 시간 쿼리맵의 시간적 변화에 따른 시간 패턴 특징을 나타내는 시간 문맥 벡터를 획득하고, 시간 패턴 특징과 이에 따른 주의도가 각각 시간 키와 시간 값으로 서로 매칭되어 미리 저장된 다수의 시간 아이템에서 상기 다수의 시간 키 각각과 상기 시간 문맥 벡터 사이의 유사도에 따른 시간 가중치를 대응하는 시간 값에 가중합하여 시간 강조 벡터를 획득하는 단계; 및

상기 다수의 사람 특징맵을 대응하는 다수의 공간 강조맵으로 정제하고, 시간 강조 벡터의 대응하는 원소로 강조하여 사람 표현자를 획득하는 단계를 포함하는 사람 재식별 방법.

청구항 12

제11항에 있어서, 상기 공간 강조맵을 획득하는 단계는

상기 다수의 공간 쿼리맵 각각에서 채널 방향의 다수의 공간 픽셀 벡터를 추출하는 단계;

각각 상기 공간 키와 상기 공간 값이 서로 매칭되어 포함된 상기 다수의 공간 아이템이 미리 저장된 공간 메모리에서 다수의 공간 키를 리드하여, 상기 다수의 공간 픽셀 벡터 각각과 상기 다수의 공간 키 각각 사이의 유사도에 따른 공간 가중치를 계산하는 단계; 및

상기 공간 메모리에서 다수의 공간 값을 리드하고, 상기 다수의 공간값 각각에 대응하는 공간 가중치를 가중합하여 다수의 공간 강조 벡터를 획득하고, 상기 다수의 공간 쿼리맵에 각각 대응하는 다수의 공간 강조맵을 획득하기 위해 획득된 공간 강조 벡터를 대응하는 다수의 공간 픽셀 벡터의 위치에 재배치하는 단계를 포함하는 사람 재식별 방법.

청구항 13

제12항에 있어서, 상기 시간 강조 벡터를 획득하는 단계는

상기 다수의 시간 쿼리맵 각각에 대해 글로벌 평균 풀링하여, 다수의 시간 쿼리 벡터를 획득하는 단계;

미리 학습된 인공 신경망을 이용하여, 상기 다수의 시간 쿼리 벡터에 대해 신경망 연산하여 시간적 변화 특정을 나타내는 상기 시간 문맥 벡터를 획득하는 단계;

각각 상기 시간 키와 상기 시간 값이 서로 매칭되어 포함된 상기 다수의 시간 아이템이 미리 저장된 시간 메모리에서 다수의 시간 키를 리드하여, 상기 시간 문맥 벡터와 상기 다수의 시간 키 각각 사이의 유사도에 따른 시간 가중치를 계산하는 단계; 및

상기 시간 메모리에서 다수의 시간 값을 리드하고, 상기 시간 강조 벡터를 획득하기 위해 상기 다수의 시간값 각각에 대응하는 시간 가중치를 가중합하는 단계를 포함하는 사람 재식별 방법.

청구항 14

제13항에 있어서, 상기 시간 문맥 벡터를 획득하는 단계는

LSTM(Long Short Term Memory)을 이용하여 신경망 연산하는 사람 재식별 방법.

청구항 15

제13항에 있어서, 상기 사람 표현자를 획득하는 단계는

상기 다수의 사람 특징맵과 상기 다수의 공간 강조맵을 인가받아, 상기 다수의 사람 특징맵 각각에 대해 대응하는 공간 강조맵을 차감하여 다수의 정제 사람 특징맵을 획득하는 단계;

상기 다수의 정제 사람 특징맵을 인가받아 각각 글로벌 풀링하여 다수의 정제 사람 특징 벡터를 획득하는 단계; 및

상기 다수의 정제 사람 특징 벡터와 상기 시간 강조 벡터를 인가받고, 상기 사람 표현자를 획득하기 위해 상기 다수의 정제 사람 특징 벡터 각각에 상기 시간 강조 벡터의 대응하는 원소를 가중하고 결합하는 단계를 포함하는 사람 재식별 방법.

청구항 16

제13항에 있어서, 상기 사람 재식별 방법은

미리 학습된 인공 신경망을 이용하여 상기 사람 표현자를 신경망 연산으로 분류하여 상기 동영상에 포함된 사람에 대한 식별자를 추출하는 단계를 더 포함하는 사람 재식별 방법.

청구항 17

제13항에 있어서, 상기 공간 가중치를 계산하는 단계는

상기 다수의 프레임 중 i번째 프레임에서 획득된 공간 쿼리맵(q^{s}_{i})에서 추출된 k번째 공간 픽셀 벡터($q^{s}_{i,k}$)와 M개 공간 아이템 중 n번째 공간 아이템의 공간 키(k^{s}_{n}) 사이의 유사도에 따른 공간 가중치($a^{s}_{i,k,n}$)를 수학식

$$a^{s}_{i,k,n} = \frac{\exp((\mathbf{q}^{s}_{i,k})^{T}\,\mathbf{k}^{s}_{n})}{\sum_{n'=1}^{M}\exp((\mathbf{q}^{s}_{i,k})^{T}\,\mathbf{k}^{s}_{n'})}$$

에 따라 계산하고,

상기 공간 픽셀 벡터의 위치에 재배치하는 단계는

상기 공간 가중치($a^{s}_{i,k,n}$)를 대응하는 n번째 공간 값(v^{s}_{n}) 각각에 가중합하여 상기 다수의 공간 강조 벡터($o^{s}_{i,k}$) 각각을 수학식

$$\mathbf{o}^{s}_{i,k} = \sum_{n=1}^{M} a^{s}_{i,k,n}\mathbf{v}^{s}_{n}$$

에 따라 계산하여 획득하는 사람 재식별 방법.

청구항 18

제13항에 있어서, 상기 시간 가중치를 계산하는 단계는

상기 다수의 시간 쿼리맵을 글로벌 평균 풀링하고 신경망 연산하여 획득된 상기 시간 문맥 벡터(q^{t})와 N개 시간 아이템 중 n번째 시간 아이템의 시간 키(k^{t}_{n}) 사이의 유사도에 따른 시간 가중치(a^{t}_{n})를 수학식

$$a^{t}_{n} = \frac{\exp((\mathbf{q}^{t})^{T}\,\mathbf{k}^{t}_{n})}{\sum_{n'=1}^{N}\exp((\mathbf{q}^{t})^{T}\,\mathbf{k}^{t}_{n'})}$$

에 따라 계산하고,

상기 시간 가중치를 가중합하는 단계는

상기 시간 가중치(a^{t}_{n})를 대응하는 n번째 시간 키(v^{t}_{n}) 각각에 가중합하여 상기 시간 강조 벡터(o^{t})를 수학식

$$\mathbf{o}^{t} = \sum_{n=1}^{N} a^{t}_{n}\mathbf{v}^{t}_{n}$$

에 따라 계산하여 획득하는 사람 재식별 방법.

청구항 19

제16항에 있어서, 상기 사람 재식별 방법은

상기 공간 메모리에 저장된 상기 다수의 공간 아이템과 상기 시간 메모리에 저장된 상기 다수의 시간 아이템의 분포에 따른 메모리 확산 손실과 추출된 사람에 대한 식별자에 따른 식별 손실을 기지정된 교차 엔트로피와 삼중항 손실로 계산하며, 계산된 상기 메모리 확산 손실과 상기 식별 손실의 합으로 계산되는 총 손실을 역전파하는 학습 단계를 더 포함하는 사람 재식별 방법.

청구항 20

제19항에 있어서, 상기 메모리 확산 손실은

기지정된 크기의 미니 배치 내 다수의 프레임에서 추출된 다수의 공간 쿼리맵(q^s_i) 또는 다수의 시간 쿼리맵(q^t_i)과 다수의 공간 키(k^s) 또는 다수의 시간 키(k^t) 각각에 대해 계산된 다수의 공간 가중치(a^s_n) 또는 시간 가중치(a^t_n)의 최대값과 최소값 사이의 편차로 계산되는 확산 분포가 기지정된 기준 분포(α) 이상이 되도록 수학식

$$\mathcal{L}_S = \sum_{n=1}^{M} [\, min(\mathbf{a}^s_n) - max(\mathbf{a}^s_n) + \alpha \,]_+ + [\, min(\mathbf{a}^t_n) - max(\mathbf{a}^t_n) + \alpha \,]_+$$

에 따라 계산되는 사람 재식별 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 사람 재식별 장치 및 방법에 관한 것으로, 시공간 메모리 네트워크를 활용한 동영상 기반 사람 재식별 장치 및 방법에 관한 것이다.

배 경 기 술

[0002] 사람 재식별(Person Re-identification: reID 라고 함)은 다양한 카메라에서 촬영된 이미지 또는 동영상에서 관심 대상인 쿼리에 해당하는 사람을 탐색하는 것을 목표로 한다. 이와 같은 사람 재식별 기술은 최근 CNN(Convolutional Neural Networks)과 같은 인공 신경망을 적용함에 따라 급격하게 발전되어 다양한 분야에 이용되고 있다.

[0003] 사람 재식별 기술 중에서도 동영상 기반 사람 재식별은 CCTV 등의 여러 카메라에서 촬영된 동영상 집합에서 쿼리된 사람과 동일한 사람이 포함된 비디오를 검색하는 것을 목표로 한다. 동영상 기반 사람 재식별은 배경 클러터(background clutter)나 프레임에 의한 부분 폐색(partial occlusions over frame)과 같이 공간적 및 시간적으로 사람의 특징을 정확하게 도출하기 어렵게 하는 산만 요소들로 인해 이미지 기반 사람 재식별에 비해 정확한 재식별이 어렵다는 문제가 있다.

[0004] 도 1은 동영상 기반 사람 재식별에서 공간적 및 시간적 산만 요인들을 설명하기 위한 도면이다.

[0005] 도 1에서 (a)는 공간적 산만 요인을 설명하기 위한 도면으로, (a)에서는 서로 다른 3가지 공간에서 촬영된 동영상에서 별도로 추출된 프레임 이미지로서 각 공간에 존재하는 공간적 산만 요인을 이미지의 하단에 도시하였다. (a)의 좌측, 가운데 및 우측 이미지에서는 하단에 도시된 바와 같이, 각각 배경으로 포함된 운동장(Playfield)과 가로등(Street light) 및 콘크리트 포장(concrete pavers) 등이 공간적 산만 요소로 작용함을 알 수 있다.

[0006] 따라서 획득된 동영상에서 공통적으로 표시되는 공간적 산만 요인을 제거할 수 있다면 사람 특징을 더욱 정확하게 도출할 수 있으며, 도출된 사람 특징을 기반으로 사람을 정확하게 재식별할 수 있다.

[0007] 한편, (b)는 시간적 산만 요인을 설명하기 위한 도면으로, (b)에서는 동영상의 연속하는 프레임 중 일부를 발췌한 이미지를 나타내고, 각 이미지 하단에서는 각 프레임에 따른 시간적 중요도를 나타낸다.

[0008] (b)에서 좌측에서는 다수의 프레임에서 시간의 변화에 따른 변화가 크지 않고, 사람이 명확하게 나타나므로 어느 프레임에서도 동등한 수준의 사람 특징을 추출할 수 있다. 따라서 각 프레임에 대한 시간적 주의를 동등하게 하여 동영상에서 사람을 식별할 수 있다.

[0009] 반면, 가운데에서는 시간의 흐름에 따라 각 프레임에 포함된 사람의 크기가 작아질뿐만 아니라 다른 사람들이나 장애물로 인해 폐색이 발생하며, 이로 인해 시간적으로 초기 프레임에서는 높은 수준으로 특징이 추출될 수 있으나, 시간이 흐름에 따라 추출되는 특징 수준이 낮아지게 된다. 즉 시작적으로 이후 프레임이 시간적 산만 요인이 될 수 있다. 따라서 시간적으로 초기 프레임에 대한 주의도를 높게 하고, 이후 시간의 흐름에 따라 각 프

레임에서의 주의도를 낮게 하는 것이 바람직하다.

[0010] 그리고 오른쪽에서는 시간적으로 초기 프레임에는 손잡이 등으로 인한 폐색이 발생된 반면, 이후 프레임에서는 폐색이 해소되었다. 이 경우, 시간적으로 이후 획득된 프레임에서 더 높은 수준의 사람 특징이 추출될 수 있으며, 따라서 시간적으로 초기 프레임에 대한 주의도를 낮게 하고, 이후 프레임에서의 주의도를 높게 하는 것이

[0011] 즉 동영상의 다수의 프레임에서 중요한 프레임을 식별할 수 있다면, 이 또한 정확한 사람 특징을 도출할 수 있도록 하여, 사람 재식별 성능을 향상시킬 수 있다.

[0012] 다만 이와 같은 공간적 시간적 산만 요인은 매우 다양하게 나타나므로, 동영상에서 공간적 시간적 산만 요인을 고려하여 사람 재식별을 수행하기 어렵하는 한계가 있다.

선행기술문헌

특허문헌

[0013] (특허문헌 0001) 한국 등록 특허 제10-2225022호 (2021.03.03 등록)

발명의 내용

해결하려는 과제

[0014] 본 발명의 목적은 다양한 환경에서 획득된 동영상에서도 정확하게 사람을 재식별할 수 있는 사람 재식별 장치 및 방법을 제공하는데 있다.

[0015] 본 발명의 다른 목적은 공간 메모리 네트워크를 이용하여 다수의 프레임 각각에서 공간적 산만 요인을 제거하고, 시간 메모리 네트워크를 이용하여 각 프레임에 대한 중요도를 가중하여 정확한 사람 표현자를 추출하여 재식별할 수 있는 사람 재식별 장치 및 방법을 제공하는데 있다.

과제의 해결 수단

[0016] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른 사람 재식별 장치는 다수의 프레임으로 구성된 동영상으로 인가받아, 다수의 프레임 각각에 대해 미리 학습된 인공 신경망으로 신경망 연산하여 다수의 공간 쿼리맵, 다수의 시간 쿼리맵 및 다수의 사람 특징맵을 추출하는 인코더; 상기 다수의 공간 쿼리맵 각각에서 채널 방향의 다수의 공간 픽셀 벡터를 추출하고, 공간적 산만 요인에 대한 특징과 대표 표현자가 각각 공간 키와 공간 값으로 서로 매칭되어 미리 저장된 다수의 공간 아이템에서 다수의 공간 키 각각과 다수의 공간 픽셀 벡터 사이의 유사도에 따른 공간 가중치를 대응하는 공간 값에 가중합하여 재배치함으로써 다수의 공간 강조맵을 획득하는 공간 메모리 네트워크; 신경망 연산으로 상기 다수의 시간 쿼리맵의 시간적 변화에 따른 시간 패턴 특징을 나타내는 시간 문맥 벡터를 획득하고, 시간 패턴 특징과 이에 따른 주의도가 각각 시간 키와 시간 값으로 서로 매칭되어 미리 저장된 다수의 시간 아이템에서 상기 다수의 시간 키 각각과 상기 시간 문맥 벡터 사이의 유사도에 따른 시간 가중치를 대응하는 시간 값에 가중합하여 시간 강조 벡터를 획득하는 시간 메모리 네트워크; 및 상기 다수의 사람 특징맵을 대응하는 다수의 공간 강조맵으로 정제하고, 시간 강조 벡터의 대응하는 원소로 강조하여 사람 표현자를 획득하는 사람 표현자 획득부를 포함한다.

[0017] 상기 공간 메모리 네트워크는 상기 다수의 공간 쿼리맵 각각에서 채널 방향의 다수의 공간 픽셀 벡터를 추출하는 공간 벡터 선택부; 각각 상기 공간 키와 상기 공간 값이 서로 매칭되어 포함된 상기 다수의 공간 아이템이 미리 저장된 공간 메모리; 상기 공간 메모리에서 다수의 공간 키를 리드하여, 상기 다수의 공간 픽셀 벡터 각각과 상기 다수의 공간 키 각각 사이의 유사도에 따른 공간 가중치를 계산하는 공간 유사도 계산부; 및 상기 공간 메모리에서 다수의 공간 값을 리드하고, 상기 다수의 공간값 각각에 대응하는 공간 가중치를 가중합하여 다수의 공간 강조 벡터를 획득하고, 획득된 공간 강조 벡터를 대응하는 다수의 공간 픽셀 벡터의 위치에 재배치하여 상기 다수의 공간 쿼리맵에 각각 대응하는 다수의 공간 강조맵을 획득하는 공간 강조맵 획득부를 포함할 수 있다.

[0018] 상기 시간 메모리 네트워크는 상기 다수의 시간 쿼리맵 각각에 대해 글로벌 평균 풀링하여, 다수의 시간 쿼리 벡터를 획득하는 시간 쿼리 풀링부; 미리 학습된 인공 신경망으로 구현되어, 상기 다수의 시간 쿼리 벡터에 대해 신경망 연산하여 시간적 변화 특정을 나타내는 상기 시간 문맥 벡터를 획득하는 시간적 문맥 인코딩부; 각각 상기 시간 키와 상기 시간 값이 서로 매칭되어 포함된 상기 다수의 시간 아이템이 미리 저장된 시간 메모리; 상

기 시간 메모리에서 다수의 시간 키를 리드하여, 상기 시간 문맥 벡터와 상기 다수의 시간 키 각각 사이의 유사도에 따른 시간 가중치를 계산하는 시간 유사도 계산부; 및 상기 시간 메모리에서 다수의 시간 값을 리드하고, 상기 다수의 시간값 각각에 대응하는 시간 가중치를 가중합하여 상기 시간 강조 벡터를 획득하는 시간 강조 벡터 획득부를 포함할 수 있다.

[0019] 상기 사람 표현자 획득부는 상기 다수의 사람 특징맵과 상기 다수의 공간 강조맵을 인가받아, 상기 다수의 사람 특징맵 각각에 대해 대응하는 공간 강조맵을 차감하여 다수의 정제 사람 특징맵을 획득하는 공간적 사람 특징맵 정제부; 상기 다수의 정제 사람 특징맵을 인가받아 각각 글로벌 풀링하여 다수의 정제 사람 특징 벡터를 획득하는 정제 사람 특징맵 풀링부; 및 상기 다수의 정제 사람 특징 벡터와 상기 시간 강조 벡터를 인가받고, 상기 다수의 정제 사람 특징 벡터 각각에 상기 시간 강조 벡터의 대응하는 원소를 가중하고 결합하여 상기 사람 표현자를 획득하는 시간적 강조부를 포함할 수 있다.

[0020] 상기 사람 재식별 장치는 미리 학습된 인공 신경망으로 구현되어, 상기 사람 표현자를 신경망 연산으로 분류하여 상기 동영상에 포함된 사람에 대한 식별자를 추출하는 사람 재식별부를 더 포함할 수 있다.

[0021] 상기 사람 재식별 장치는 학습 시에 구비되어 상기 공간 메모리에 저장된 상기 다수의 공간 아이템과 상기 시간 메모리에 저장된 상기 다수의 시간 아이템의 분포에 따른 메모리 확산 손실과 추출된 사람에 대한 식별자에 따른 식별 손실을 기지정된 교차 엔트로피와 삼중항 손실로 계산하며, 계산된 상기 메모리 확산 손실과 상기 식별 손실의 합으로 계산되는 총 손실을 역전파하는 손실 계산부를 더 포함할 수 있다.

[0022] 상기 목적을 달성하기 위한 본 발명의 다른 실시예에 따른 사람 재식별 방법은 다수의 프레임으로 구성된 동영상으로 인가받아, 다수의 프레임 각각에 대해 미리 학습된 인공 신경망으로 신경망 연산하여 다수의 공간 쿼리맵, 다수의 시간 쿼리맵 및 다수의 사람 특징맵을 추출하는 단계; 상기 다수의 공간 쿼리맵 각각에서 채널 방향의 다수의 공간 픽셀 벡터를 추출하고, 공간적 산만 요인에 대한 특징과 대표 표현자가 각각 공간 키와 공간 값으로 서로 매칭되어 미리 저장된 다수의 공간 아이템에서 다수의 공간 키 각각과 다수의 공간 픽셀 벡터 사이의 유사도에 따른 공간 가중치를 대응하는 공간 값에 가중합하여 재배치함으로써 다수의 공간 강조맵을 획득하는 단계; 인공 신경망을 이용하여 신경망 연산으로 상기 다수의 시간 쿼리맵의 시간적 변화에 따른 시간 패턴 특징을 나타내는 시간 문맥 벡터를 획득하고, 시간 패턴 특징과 이에 따른 주의도가 각각 시간 키와 시간 값으로 서로 매칭되어 미리 저장된 다수의 시간 아이템에서 상기 다수의 시간 키 각각과 상기 시간 문맥 벡터 사이의 유사도에 따른 시간 가중치를 대응하는 시간 값에 가중합하여 시간 강조 벡터를 획득하는 단계; 및 상기 다수의 사람 특징맵을 대응하는 다수의 공간 강조맵으로 정제하고, 시간 강조 벡터의 대응하는 원소로 강조하여 사람 표현자를 획득하는 단계를 포함한다.

발명의 효과

[0023] 따라서, 본 발명의 실시예에 따른 사람 재식별 장치 및 방법은 공간 메모리 네트워크를 이용하여 다수의 프레임 각각에서 공간적 산만 요인을 제거하고, 시간 메모리 네트워크를 이용하여 각 프레임에 대한 중요도를 가중하여 사람 표현자를 추출하여 재식별하므로, 다양한 환경에서 획득된 동영상에서도 정확하게 사람을 재식별할 수 있어 사람 재식별 성능을 크게 향상시킬 수 있다.

도면의 간단한 설명

[0024] 도 1은 동영상 기반 사람 재식별에서 공간적 및 시간적 산만 요인들을 설명하기 위한 도면이다.

도 2는 본 발명의 일 실시예에 따른 사람 재식별 장치의 개략적 구성을 나타낸다.

도 3은 도 2의 인코더의 동작을 설명하기 위한 도면이다.

도 4 및 도 5는 도 2의 공간 메모리 네트워크의 개략적 동작을 설명하기 위한 도면이다.

도 6은 도 2의 시간 메모리 네트워크의 개략적 동작을 설명하기 위한 도면이다.

도 7은 도 2의 사람 표현자 획득부의 동작을 설명하기 위한 도면이다.

도 8은 도 2의 공간 메모리 네트워크의 상세 구성의 일 예를 나타낸다.

도 9는 도 8에 도시된 공간 메모리 네트워크의 상세 동작을 설명하기 위한 도면이다.

도 10은 도 2의 시간 메모리 네트워크의 상세 구성의 일 예를 나타낸다.

도 11은 도 10에 도시된 시간 메모리 네트워크의 상세 동작을 설명하기 위한 도면이다.

도 12는 도 2의 사람 표현자 획득부의 상세 구성의 일 예를 나타낸다.

도 13은 도 12의 공간적 사람 표현자 정제부의 동작을 설명하기 위한 도면이다.

도 14는 도 12의 정제 사람 표현자 풀링부 및 시간적 강조부의 동작을 설명하기 위한 도면이다.

도 15는 메모리 확산 손실을 설명하기 위한 도면이다.

도 16은 본 발명의 일 실시예에 따른 사람 재식별 방법을 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0025] 본 발명과 본 발명의 동작상의 이점 및 본 발명의 실시에 의하여 달성되는 목적을 충분히 이해하기 위해서는 본 발명의 바람직한 실시예를 예시하는 첨부 도면 및 첨부 도면에 기재된 내용을 참조하여야만 한다.

[0026] 이하, 첨부한 도면을 참조하여 본 발명의 바람직한 실시예를 설명함으로써, 본 발명을 상세히 설명한다. 그러나, 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며, 설명하는 실시예에 한정되는 것이 아니다. 그리고, 본 발명을 명확하게 설명하기 위하여 설명과 관계없는 부분은 생략되며, 도면의 동일한 참조부호는 동일한 부재임을 나타낸다.

[0027] 명세서 전체에서, 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. 또한, 명세서에 기재된 "...부", "...기", "모듈", "블록" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.

[0028] 도 2는 본 발명의 일 실시예에 따른 사람 재식별 장치의 개략적 구성을 나타내고, 도 3은 도 2의 인코더의 동작을 설명하기 위한 도면이며, 도 4 및 도 5는 도 2의 공간 메모리 네트워크의 개략적 동작을 설명하기 위한 도면이다. 그리고 도 6은 도 2의 시간 메모리 네트워크의 개략적 동작을 설명하기 위한 도면이고, 도 7은 도 2의 사람 표현자 획득부의 동작을 설명하기 위한 도면이다.

[0029] 도 2를 참조하면, 본 실시예에 따른 사람 재식별 장치는 동영상 획득부(100), 인코더(200), 공간 메모리 네트워크(300), 시간 메모리 네트워크(400), 사람 표현자 획득부(500) 및 사람 재식별부(600)를 포함할 수 있다.

[0030] 우선 동영상 획득부(100)는 탐색 대상이 되는 타겟 사람이 포함되었는지 여부가 확인되어야 하는 동영상을 획득한다. 여기서 동영상은 CCTV와 같은 고정식 카메라뿐만 아니라 차량의 블랙 박스나 스마트폰과 같이 이동식 카메라에서 획득된 영상일 수도 있으며 촬영 장소나 환경에 제한되지 않으며, 시간적으로 연속하는 다수의 프레임(F_1 ~ F_L)으로 구성될 수 있다.

[0031] 인코더(200)는 미리 학습된 인공 신경망으로 구현되어, 동영상 획득부(100)에서 획득된 동영상의 다수의 프레임(F_1 ~ F_L)을 각각 인가받고, 인가된 각 프레임($F_i|_{i=1}^{L}$)에 대해 학습된 방식에 따라 신경망 연산하여 공간 쿼리맵(q^s_i), 시간 쿼리맵(q^t_i) 및 사람 특징맵(f^o_i)을 획득한다. 즉 인코더(200)는 도 3에 도시된 바와 같이, 다수의 프레임(F_1 ~ F_L) 각각을 인가받아 학습된 방식에 따라 신경망 연산하여, 각 프레임(F_i)에 대응하는 공간 쿼리맵(q^s_i), 시간 쿼리맵(q^t_i) 및 사람 특징맵(f^o_i)을 획득할 수 있다.

[0032] 여기서 인코더(200)는 인가된 프레임(F_i)에 대해 신경망 연산하여 공통 특징을 추출하여 공통 특징맵을 획득하는 백본 레이어(미도시)와 공통 특징맵에 대해 각각 서로 다르게 신경망 연산하여 공간 쿼리맵(q^s_i), 시간 쿼리맵(q^t_i) 및 사람 특징맵(f^o_i) 중 대응하는 맵을 획득하는 3개의 헤드 레이어(미도시)를 포함하여 구성될 수 있다.

[0033] 공간 메모리 네트워크(300)는 공간 메모리(미도시)와 인공 신경망을 포함하여 구성되고, 공간 메모리에는 공간 키(k^s)와 공간 값(v^s)이 서로 매칭되어 포함된 M개의 공간 아이템이 저장된다. 공간 아이템에서 공간 키(k^s)는 동영상의 각 프레임에서 가로등, 나무, 콘크리트 포장 등과 같은 공간적 산만 요인을 검출하기 위해 다양한 산만 요인에 각각 대응하는 프로토타입 특징으로 구조적 특징을 나타내고, 공간 값(v^s)은 매칭된 공간 키(k^s)에 대

응하는 산만 요인의 주요 특징을 나타내는 대표 표현자이다. 그리고 서로 매칭되어 공간 아이템에 포함되는 공간 키(k^s_n, $n \in \{1, 2, \ldots, M\}$)와 공간 값($v^s_n$)은 다양한 산만 요인 중 하나의 산만 요인에 대응하도록 학습 시에 미리 획득되어 저장된다.

[0034] 공간 메모리 네트워크(300)는 도 4에 도시된 바와 같이, 인코더(200)로부터 다수의 공간 쿼리맵(q^s_i)을 각각 인가받고, 인가된 공간 쿼리맵(q^s_i)의 다수의 픽셀 벡터와 공간 메모리(미도시)에 저장된 다수의 공간 아이템의 공간 키(k^s_n) 각각 사이의 유사도에 따른 가중치를 각 공간 키(k^s_n)에 매칭된 공간 값(v^s_n)에 가중합하여 재배치함으로써 공간 강조맵(o^s_i)을 획득한다.

[0035] 여기서 다수의 공간 강조맵(o^s_i) 각각은 각 공간 쿼리맵(q^s_i)에 대응하여 개별적으로 획득된다. 즉 공간 강조맵(o^s_i)은 프레임(F_i) 개수에 대응하는 개수로 획득되며, 이에 각각 동일한 프레임(F_i)에서 추출된 사람 특징맵(f^o_i)에 포함된 산만 요소들의 특징을 표현하는 것으로 볼 수 있다.

[0036] 따라서 공간 메모리 네트워크(300)에서 획득된 다수의 공간 강조맵(o^s) 각각은 도 4 및 5에 도시된 바와 같이, 다수의 사람 특징맵(f^o) 중 대응하는 사람 특징맵(f^o_i)에서 불필요한 산만 요소를 제거하여 사람 특징이 강조되어 나타나도록 정제하기 위해 이용될 수 있다.

[0037] 한편, 시간 메모리 네트워크(400)는 시간 메모리(미도시)와 인공 신경망을 포함하여 구성되고, 시간 메모리에는 시간 키(k^t)와 시간 값(v^t)이 서로 매칭되어 포함된 N개의 시간 아이템이 저장된다. 여기서 시간 키(k^t)는 동영상의 연속하는 다수의 프레임에서의 변화 패턴을 검출하기 위해 미리 지정된 패턴 특징이고, 시간 값(v^t)은 매칭된 시간 키(k^s)에 대응하는 시간적 주의값을 나타낸다.

[0038] 시간 메모리 네트워크(400)는 도 6에 도시된 바와 같이, 인코더(200)로부터 다수의 시간 쿼리맵(q^t_i)을 인가받고, 신경망 연산을 수행하여 인가된 다수의 시간 쿼리맵(q^t_i)의 시간적 변화에 따른 시간적 문맥(temporal context) 정보를 나타내는 시간 문맥 벡터를 추출한다. 그리고 추출된 시간 문맥 벡터와 시간 메모리(미도시)에 저장된 다수의 시간 아이템의 시간 키(k^t_n) 각각 사이의 유사도에 따른 가중치를 각 시간 키(k^t_n)에 매칭된 시간 값(v^s_n)에 가중합하여 시간 강조 벡터(o^t)를 획득한다.

[0039] 시간 강조 벡터(o^t)는 다수의 시간 쿼리맵(q^t_i)에 대해 하나로 획득되며, 시간 강조 벡터(o^t)의 각 원소는 동영상의 다수의 프레임(F_i) 각각에 대한 주의도를 나타낸다.

[0040] 따라서 시간 메모리 네트워크(400)에서 획득된 시간 강조 벡터(o^t)의 각 원소는 도 7에 도시된 바와 같이, 다수의 공간 강조맵(o^s)에 의해 사람 특징이 강조되도록 정제된 사람 특징맵(f^s_i)에 각각에 대한 시간적 주의 수준을 조절하기 위한 가중치로 이용될 수 있다.

[0041] 사람 표현자 획득부(500)는 도 4 및 도 5에 도시된 바와 같이, 인코더(200)로부터 다수의 사람 특징맵(f^s_i)을 인가받고 공간 메모리 네트워크(300)로부터 다수의 공간 강조맵(o^s)을 인가받아, 다수의 사람 특징맵(f^s_i) 각각에서 대응하는 공간 강조맵(o^s_i)을 기지정된 방식으로 차감하여 공간적 산만 요인이 제거되어 정제된 다수의 정제 특징맵(f^s_i)을 획득한다. 그리고 도 7에 도시된 바와 같이, 다수의 정제 특징맵(f^s_i) 각각에 시간 강조 벡터(o^t)의 대응하는 원소를 가중하고 결합하여 공간적 산만 요인이 배제되었을 뿐만 아니라 시간적 강조 위치가 지정된 사람 표현자를 획득한다.

[0042] 그리고 사람 재식별부(600) 또한 인공 신경망으로 구현되어 사람 표현자 획득부(500)에서 획득된 사람 표현자를 미리 학습된 방식에 따라 신경망 연산으로 분류하여 사람 표현자가 나타내는 사람에 대한 식별자를 획득한다. 사람 재식별부(600)는 획득된 식별자가 탐색되어야 하는 타겟 사람에 대한 식별자인지 여부를 판별하여 타겟 사람을 재식별할 수 있다.

[0043] 결과적으로 본 실시예에 따른 사람 재식별 장치는 공간 메모리 네트워크(300)를 이용하여 동영상의 다수 프레임(F_1 ~ F_L) 각각에서 사람 이외의 산만 요인을 제거하고, 시간 메모리 네트워크(400)를 이용하여 다수 프레임(F_1 ~ F_L)에 대한 주의도를 서로 다르게 조절하여 매우 정확하게 사람에 대한 사람 표현자인 특징만을 추출할 수 있도록 한다. 따라서 다양한 환경에서 획득된 동영상에서도 정확하게 사람을 재식별할 수 있어 사람 재식별 성능을 크게 향상시킬 수 있다.

[0044] 도 8은 도 2의 공간 메모리 네트워크의 상세 구성의 일 예를 나타내고, 도 9는 도 8에 도시된 공간 메모리 네트워크의 상세 동작을 설명하기 위한 도면이다.

[0045] 도 8 및 도 9을 참조하면, 공간 메모리 네트워크(300)는 공간 벡터 선택부(310), 공간 메모리(320), 공간 유사도 계산부(330) 및 공간 강조맵 획득부(340)를 포함할 수 있다.

[0046] 공간 벡터 선택부(310)는 인코더(200)로부터 다수의 공간 쿼리맵(q^s_i)을 인가받는다. 다수의 공간 쿼리맵(q^s_i) 각각은 채널 수(D) × 픽셀 수(K)의 크기로 획득될 수 있으며, 이에 공간 벡터 선택부(310)는 인가된 다수의 공간 쿼리맵(q^s_i) 각각의 K개의 픽셀(k ∈ {1, 2, …, K}, K = 높이(H) × 폭(W)) 각각에서 채널 방향의 벡터인 공간 픽셀 벡터($\mathbf{q}^s_{i,k} \in \mathbb{R}^D$)를 추출하여 공간 유사도 계산부(330)로 전달한다.

[0047] 한편 공간 메모리(320)에는 상기한 바와 같이, M개의 공간 아이템이 저장되고, M개의 공간 아이템 각각에는 공간 키(k^s)와 공간 값(v^s)이 서로 매칭되어 포함된다. 공간 키(k^s)와 공간 값(v^s) 각각은 공간 픽셀 벡터($q^s_{i,k}$)와 동일한 길이(D)를 갖는 벡터($\mathbf{k}^s \in \mathbb{R}^{D\times M}$, $\mathbf{v}^s \in \mathbb{R}^{D\times M}$)이다. 여기서 공간 키($k^s$)는 학습 시에 산만 요인에 따른 픽셀의 구조적 특징이 누적된 프로토타입 특징으로 획득되고, 공간 값(v^s)은 공간 키(k^s)의 클래스를 픽셀 크기로 대표하여 표현할 수 있는 대표 표현자를 나타낸다. 즉 공간 메모리(320)에는 다양한 공간적 산만 요인에 대한 프로토타입 특징과 대표 표현자가 공간 아이템의 키(k^s)와 값(v^s)으로 서로 매칭되어 저장된다.

[0048] 공간 유사도 계산부(330)는 공간 벡터 선택부(310)에서 추출된 공간 픽셀 벡터($q^s_{i,k}$)를 인가받아 공간 메모리(320)에 저장된 M개의 공간 키(k^s) 각각과의 코사인 유사도를 계산하고 소프트맥스(softmax) 함수 등으로 정규화하여 각 공간 픽셀 벡터($q^s_{i,k}$)가 M개의 공간 키(k^s) 중 n번째 공간 키(k^s_n)에 매칭될 확률을 나타내는 공간 가중치($a^s_{i,k,n}$)를 수학식 1과 같이 계산한다.

수학식 1

[0049]
$$a^s_{i,k,n} = \frac{\exp\left((\mathbf{q}^s_{i,k})^{\mathrm{T}}\,\mathbf{k}^s_n\right)}{\sum_{n'=1}^{M}\exp\left((\mathbf{q}^s_{i,k})^{\mathrm{T}}\,\mathbf{k}^s_{n'}\right)}$$

[0050] 즉 공간 가중치($a^s_{i,k,n}$)는 M개의 공간 아이템 중 n번째 공간 아이템의 공간 키(k^s_n)에 해당하는 장면 디테일이 i번째 프레임(F_i)의 k번째 픽셀에 존재할 가능성을 나타낸다. 따라서 공간 유사도 계산부(330)에서는 각 공간 픽셀 벡터($q^s_{i,k}$)가 M개의 공간 키(k^s) 각각과 매칭될 확률을 나타내는 M개의 공간 가중치($a^s_{i,k,n}$)가 획득한다.

[0051] 그리고 공간 강조맵 획득부(340)는 공간 메모리(320)에서 M개의 공간 키(k^s) 각각에 매칭된 M개의 공간 값(v^s)을

리드하고, 리드된 M개의 공간 값(v^s) 각각에 수학식 2와 같이 공간 픽셀 벡터($q^s_{i,k}$)에 대해 획득된 M개의 공간 가중치($a^s_{i,k,n}$) 각각을 가중합하여 공간 강조 벡터($o^s_{i,k}$)를 획득한다.

수학식 2

[0052]
$$\mathbf{o}^{\mathrm{s}}_{i,k} = \sum_{n=1}^{M} a^{\mathrm{s}}_{i,k,n}\mathbf{v}^{\mathrm{s}}_{n}$$

[0053] 공간 강조 벡터($o^s_{i,k}$)는 각 공간 픽셀 벡터($q^s_{i,k}$)에 대응하여 획득되었으므로, 공간 강조맵 획득부(340)는 대응하는 공간 픽셀 벡터($q^s_{i,k}$)의 위치에 따라 공간 강조 벡터($o^s_{i,k}$)를 재배치하여 공간 강조맵(o^s_i)을 획득한다.

[0054] 여기서 공간 강조맵(o^s_i)은 각 공간 쿼리맵(q^s_i)에서 획득되므로, 동영상의 프레임 수(L)와 동일한 개수로 획득된다.

[0055] 따라서 공간 메모리 네트워크(300)는 동영상의 프레임 수(L)에 대응하여 입력되는 L개의 공간 쿼리맵(q^s_i)의 각 픽셀이 공간 메모리(320)에 저장된 M개의 공간 키(k^s) 각각과 매칭될 확률을 가중치로 계산하여, 대응하는 공간 값(v^s)에 가중합하여 공간 쿼리맵(q^s_i)의 각 픽셀에 포함된 산만 요인들의 대표 특징을 획득한다.

[0056] 도 10은 도 2의 시간 메모리 네트워크의 상세 구성의 일 예를 나타내고, 도 11은 도 10에 도시된 시간 메모리 네트워크의 상세 동작을 설명하기 위한 도면이다.

[0057] 도 10 및 도 11을 참조하면, 시간 메모리 네트워크(400)는 시간 쿼리 풀링부(410), 시간적 문맥 인코딩부(420), 시간 메모리(430), 시간 유사도 계산부(440) 및 시간 강조 벡터 획득부(450)를 포함할 수 있다.

[0058] 시간 쿼리 풀링부(410)는 인코더(200)로부터 다수의 공간 쿼리맵(q^s_i)을 각각 인가받는다. 여기서 다수의 시간 쿼리맵(q^t_i) 또한 채널 수(D) × 픽셀 수(K)의 크기로 획득될 수 있다. 시간 쿼리 풀링부(410)는 다수의 시간 쿼리맵(q^t_i) 각각을 글로벌 평균 풀링(Global Average Pooling: 이하 GAP)하여 다수의 시간 쿼리맵(q^t_i) 각각에 대응하는 다수의 시간 쿼리 벡터를 획득한다. 여기서 다수의 시간 쿼리 벡터는 도 11에 도시된 바와 같이, 시간 쿼리맵(q^t_i)의 채널 수(D)에 대응하는 길이를 갖는다.

[0059] 시간적 문맥 인코딩부(420)는 미리 학습된 인공 신경망으로 구현되어 다수의 시간 쿼리맵(q^t_i)에 대해 획득된 다수의 시간 쿼리 벡터를 순차적으로 인가받아 신경망 연산하여 다수의 시간 쿼리 벡터의 시간적 변화 특성을 나타내는 시간 문맥 특성을 추출함으로써, 시간 문맥 벡터(q^t)를 획득한다. 여기서 시간적 문맥 인코딩부(420)는 인공 신경망 중에서 시계열 데이터의 분석을 위해 주로 이용되고 있는 LSTM(Long Short Term Memory)으로 구현될 수 있다. 이 경우 시간 문맥 벡터(q^t)는 수학식 3과 같이 표현될 수 있다.

수학식 3

[0060]
$$\mathbf{q}^{\mathrm{t}} = \mathrm{LSTM}([\mathrm{GAP}(\mathbf{q}^{\mathrm{t}}_1), \mathrm{GAP}(\mathbf{q}^{\mathrm{t}}_2), ..., \mathrm{GAP}(\mathbf{q}^{\mathrm{t}}_L)])$$

[0061] 그리고 시간 문맥 벡터(q^t) 또한 시간 쿼리맵(q^t_i)의 채널 수(D)에 대응하는 길이($\mathbf{q}^{\mathrm{t}} \in \mathbb{R}^D$)를 가질 수 있다.

[0062] 한편, 시간 메모리(430)에는 N개의 시간 아이템이 저장되고, N개의 시간 아이템 각각에는 시간 키(k^t)와 시간 값(v^t)이 서로 매칭되어 포함된다. N개의 시간 키(k^t)는 시간 문맥 벡터(q^t)와 동일한 길이(D)를 갖는 벡터($\mathbf{k}^t \in \mathbb{R}^{D\times N}$)이고, N개의 시간 값($v^t$)은 프레임 수(L)에 따른 길이를 갖는 벡터($\mathbf{v}^t \in \mathbb{R}^{L\times N}$)이다. 시간 키($k^t$)는 학습 시에 각 프레임의 중요도에 따른 시간적 패턴 특징이 구분된 다양한 학습 영상에서 추출된 시간 문맥 벡터(q^t)가 누적 저장된 값이고, 시간 값(v^t)은 시간적 중요도에 따라 시간 키(k^s)에 대응하도록 누적 저장된 시간적 주의값을 나타낸다.

[0063] 시간 유사도 계산부(440)는 시간적 문맥 인코딩부(420)에서 획득된 시간 문맥 벡터(q^t)와 시간 메모리(430)에 저장된 N개의 시간 키(k^t) 각각 사이의 코사인 유사도를 계산하고 정규화하여 시간 문맥 벡터(q^t)가 N개의 시간 키(k^t) 중 n번째 시간 키(k^t_n)에 매칭될 확률을 나타내는 시간 가중치(a^t_n)를 수학식 4에 따라 계산한다.

수학식 4

[0064]
$$a_n^t = \frac{\exp((\mathbf{q}^t)^{\mathrm{T}}\,\mathbf{k}_n^t)}{\sum_{n'=1}^{N}\exp((\mathbf{q}^t)^{\mathrm{T}}\,\mathbf{k}_{n'}^t)}$$

[0065] 즉 시간 가중치(a^t_n)는 시간 문맥 벡터(q^t)가 N개의 시간 아이템 중 n번째 시간 아이템의 시간 키(k^t_n)에 따른 시간 패턴 특징과 유사할 확률을 나타낸다.

[0066] 그리고 시간 강조 벡터 획득부(450)는 시간 메모리(430)에서 N개의 시간 키(k^t) 각각에 매칭된 N개의 시간 값(v^s)을 리드하고, 수학식 5와 같이 N개의 시간 값(v^s_n) 각각에 N개의 시간 가중치(a^t_n)를 가중합하여, 시간 강조 벡터(o^t)를 획득한다.

수학식 5

[0067]
$$\mathbf{o}^t = \sum_{n=1}^{N} a_n^t \mathbf{v}_n^t$$

[0068] 여기서 시간 강조 벡터(o^t)는 시간 값(v^s_n)과 마찬가지로 프레임 수(L)에 따른 길이($\mathbf{o}^t \in \mathbb{R}^L$)로 획득되며, L개의 프레임($F_1$ ~ F_L) 전체에 대응하여 하나의 벡터만이 획된다.

[0069] 따라서 시간 메모리 네트워크(400)는 동영상의 L개의 프레임에서 추출된 L개의 시간 쿼리맵(q^t_i) 각각의 특징을 나타내는 L개의 시간 쿼리 벡터를 획득하고, L개의 시간 쿼리 벡터로부터 L개의 프레임의 시간적 패턴 특징을 나타내는 시간 문맥 벡터(q^t)를 추출하며, 추출된 시간 문맥 벡터(q^t)가 시간 메모리(430)에 저장된 N개의 시간 키(k^t) 각각과 매칭될 확률을 가중치로 계산하여, 대응하는 시간 값(v^t)에 가중합하여 L개의 프레임 각각에 대한 주의 수준을 결정한다.

[0070] 도 12는 도 2의 사람 표현자 획득부의 상세 구성의 일 예를 나타내고, 도 13은 도 12의 공간적 사람 표현자 정제부의 동작을 설명하기 위한 도면이며, 도 14는 도 12의 정제 사람 표현자 풀링부 및 시간적 강조부의 동작을 설명하기 위한 도면이다.

[0071] 도 12를 참조하면, 사람 표현자 획득부(500)는 공간적 사람 특징맵 정제부(510), 정제 사람 특징맵 풀링부(520) 및 시간적 강조부(530)를 포함할 수 있다.

[0072] 공간적 사람 특징맵 정제부(510)는 인코더(200)로부터 사람 특징맵(f^o_i)을 인가받고, 도 13에 도시된 바와 같이, 인가된 사람 특징맵(f^o_i)에서 공간 메모리 네트워크(300)에서 획득된 공간 강조맵(o^s_i)을 차감하여 정제 사람 특징맵(f^s_i)을 획득한다. 상기한 바와 같이, 공간 강조맵(o^s_i)은 해당 프레임(F_i)에서 다수의 산만 요인의 특징을 픽셀별 유사도에 따라 가중합하여 획득된 산만 요인 특징앱이다. 따라서 사람 특징맵(f^o_i)에서 공간 강조맵(o^s_i)을 차감하게 되면, 사람 특징맵(f^o_i)에서 산만 요인이 제거되어 사람에 대한 특징으로 정제된 정제 사람 특징맵(f^s_i)이 획득된다.

[0073] 공간적 사람 특징맵 정제부(510)는 도 5에 도시된 바와 같이, 프레임 수(L)에 대응하는 개수로 인가되는 사람 특징맵(f^o_i) 각각에 대해 대응하는 공간 강조맵(o^s_i)을 각각 차감하여 L개의 정제 사람 특징맵(f^s_i)을 획득할 수 있다.

[0074] 이때 공간적 사람 특징맵 정제부(510)는 사람 특징맵(f^o_i)에서 공간 강조맵(o^s_i)을 직접적으로 차감하지 않고, 배치 정규화(batch normalization: BN) 계층을 이용하여 수학식 6과 같이 공간 강조맵(o^s_i)이 잔차(residual) 형식으로 차감되도록 하여 정제 사람 특징맵(f^s_i)을 획득할 수 있다.

수학식 6

[0075] $$\mathbf{f}^{\mathrm{s}}_{i,k} = \mathbf{f}^{\mathrm{o}}_{i,k} - \mathrm{BN}(\mathbf{o}^{\mathrm{s}}_{i,k})$$

[0076] 이는 인코더(200)에서 출력되는 사람 특징맵(f^o_i)과 공간 메모리(320)의 공간 키 사이의 분포 간격이 조절되도록 하기 위함이다.

[0077] 한편, 정제 사람 특징맵 풀링부(520)는 도 14에 도시된 바와 같이, 공간적 사람 특징맵 정제부(510)에서 획득된 L개의 정제 사람 특징맵(f^s_i) 각각에 대해 시간 쿼리 풀링부(410)와 유사하게 글로벌 평균 풀링(GAP)하여 L개의 정제 사람 특징 벡터를 획득한다.

[0078] 그리고 시간적 강조부(530)는 L개의 정제 사람 특징 벡터 각각에 대해 시간 메모리 네트워크(400)에서 획득된 시간 강조 벡터(o^t)에서 대응하는 원소를 곱하여 가중하여 사람 표현자(f^t)를 획득한다. 즉 시간적 강조부(530)는 도 7에 도시된 바와 같이, L개의 정제 사람 특징맵(f^s_i)에서 추출된 L개의 정제 사람 특징 벡터 각각에 시간 강조 벡터(o^t)의 대응하는 원소를 가중하여 시간적 주의가 강조된 사람 표현자(f^t)를 획득한다.

[0079] 이때 사람 표현자(f^t)는 정규화되어 수학식 7과 같이 획득될 수 있다.

수학식 7

[0080] $$\mathbf{f}^{\mathrm{t}} = \sum_{i=1}^{L} \hat{o}^{\mathrm{t}}_i \, \mathrm{GAP}(\mathbf{f}^{\mathrm{s}}_i)$$

[0081] 여기서 $\hat{o}_i^t$는 정규화된 시간 강조 벡터로서 $\hat{o}_i^t = \exp(o_i^t) / \sum_{i'=1}^{L} \exp(o_{i'}^t)$로 계산된다.

[0082] 따라서 사람 표현자 획득부(500)는 인코더(200)에서 단순하게 신경망 연산하여 획득된 다수의 사람 특징맵(f^o_i)에 대해 공간 메모리 네트워크(300)에서 획득된 공간 강조맵(o^s_i)을 차감하여 사람을 제외한 공간적 산만 요인들을 배제하고, 시간 메모리 네트워크(400)에서 획득된 시간 강조 벡터(o^t)에 따라 시간에 따른 주의도를 가중함으로써, 다수의 프레임(F_1 ~ F_L)에서 공간적으로나 시간적으로 사람의 특징이 가장 잘 표출된 사람 표현자(f^t)를 획득한다. 따라서 이후 사람 재식별부(600)가 사람 표현자(f^t)로부터 사람을 정확하게 식별할 수 있도록 한다.

[0083] 한편, 본 실시예의 사람 재식별 장치는 다수의 인공 신경망을 포함하여 구성되며, 이에 실제 이용하기 이전 인공 신경망을 학습해야 한다. 따라서 사람 재식별 장치는 도 2에 도시된 바와 같이, 인공 신경망을 학습시키고, 공간 메모리(320)와 시간 메모리(430)에 저장되는 다수의 공간 아이템과 다수의 시간 아이템을 업데이트하기 위한 손실 계산부(700)를 더 포함할 수 있다.

[0084] 손실 계산부(700)는 인공 신경망의 학습 시에만 구비되고 학습이 종료되면 제거될 수 있다.

[0085] 본 실시예에서 손실 계산부(700)는 메모리 확산 손실(L_S)과 식별 손실(L_{ID})을 계산하고, 계산된 메모리 확산 손실(L_S)과 식별 손실(L_{ID})의 합을 총 손실(L_{total})로 계산하여 역전파함으로써, 인공 신경망을 학습시키고, 공간 메모리(320)와 시간 메모리(430)에 저장되는 다수의 공간 아이템과 다수의 시간 아이템을 업데이트할 수 있다.

[0086] 여기서 메모리 확산 손실(L_S)은 다수의 공간 쿼리맵(q^s_i)이 공간 메모리(320)에 저장된 다수의 공간 키(k^s) 중 특정 공간 키(k^s_n)에 편향되어 유사하게 나타나는 경우와 다수의 시간 쿼리맵(q^t_i)이 시간 메모리(430)에 저장된 다수의 시간 키(k^t) 중 특정 시간 키(k^t_n)에 과도하게 편향되어 유사하게 나타나는 것을 방지하기 위해 설정되는 손실이다. 본 실시예의 재식별 장치에서는 사람에 대한 식별자(ID)를 제외하면 공간 아이템과 시간 아이템을 업데이트하기 위한 ㅂ려도의 레이블이 없으므로 비지도 학습 방식으로 공간 아이템과 시간 아이템의 키와 값이 업데이트되어야 한다. 이때 공간 쿼리맵(q^s_i)과 시간 쿼리맵(q^t_i)이 다수의 공간 키(k^s)와 다수의 시간 키(k^t) 중 어떤 공간 키 또는 시간 키에 대응해야 하는지 또한 판별할 수 없다.

[0087] 이에 본 실시예에서 손실 계산부(700)는 각 쿼리 맵(q^s_i, q^t_i)과 각 키(k^s_n, k^t_n) 사이의 유사도가 기지정된 기준 분포(α) 이상으로 확산되어 분포되도록 학습을 수행한다.

[0088] 도 15는 메모리 확산 손실을 설명하기 위한 도면이다.

[0089] 도 15에서는 다수의 공간 쿼리맵(q^s_i) 또는 다수의 시간 쿼리맵(q^t_i)과 다수의 공간 키(k^s) 또는 다수의 시간 키(k^t) 각각 사이의 유사도에 따른 공간 가중치(a^s_n) 또는 시간 가중치(a^t_n)를 나타낸다. 그리고 (a)는 메모리 확산 손실(L_S)이 고려되지 않은 경우를 나타내고, (b)는 메모리 확산 손실(L_S)이 고려되어 학습된 경우를 나타낸다.

[0090] (a)에 도시바와 같이, 메모리 확산 손실(L_S)이 고려되지 않은 경우, 다수의 공간 쿼리맵(q^s_i) 또는 다수의 시간 쿼리맵(q^t_i)이 특정 공간 키(k^s_n) 또는 특정 시간 키(k^t_n)에 대해서만 높은 유사도를 가져 공간 가중치(a^s_n) 또는 시간 가중치(a^t_n)가 크게 나타나는 반면, 다른 공간 키(k^s) 또는 시간 키(k^t)에 대해서 모두 유사도가 낮게 나올 수 있다.

[0091] 이러한 경우, 사람 재식별 장치는 공간적으로는 특정 산만 요인만을 집중적으로 제거하거고, 시간적으로는 특정 프레임에만 과도하게 집중하여 사람 표현자를 추출하게 되어 다수의 공간 키(k^s)가 다양한 산만 요인을 나타내지 못하게 하거나, 다수의 시간 키(k^t)가 다양한 시간 패턴을 나타내지 못하게 할 수 있다.

[0092] 이에 메모리 확산 손실(L_S)은 기지정된 크기의 미니 배치 내의 다수의 공간 쿼리맵(q^s_i) 또는 다수의 시간 쿼리맵(q^t_i)과 다수의 공간 키(k^s) 또는 다수의 시간 키(k^t) 각각에 대해 계산된 다수의 공간 가중치(a^s_n) 또는 시간 가중치(a^t_n)의 최대값과 최소값 사이의 편차로 계산되는 확산 분포가 기지정된 기준 분포(α) 이상이 되도록 수학식 8과 같이 설정될 수 있다.

수학식 8

[0093]
$$\mathcal{L}_S = \sum_{n=1}^{M} [\, min(\mathbf{a}^s_n) - max(\mathbf{a}^s_n) + \alpha \,]_+ + [\, min(\mathbf{a}^t_n) - max(\mathbf{a}^t_n) + \alpha \,]_+$$

[0094] 수학식 8에 따라 계산되는 메모리 확산 손실(L_S)이 고려되는 경우, 도 15의 (b)에 도시된 바와 같이, 다수의 공간 쿼리맵(q^s_i) 또는 다수의 시간 쿼리맵(q^t_i) 각각은 다수의 공간 키(k^s) 또는 다수의 시간 키(k^t) 각각에 대해 확산된 분포의 유사도를 갖게 되며, 이에 다양한 산만 요소를 제거할 수 있을 뿐만 아니라 다양한 시간적 패턴의 영상들에 대한 시간적 주의를 추출할 수 있게 된다.

[0095] 한편, 식별 손실(L_{ID})은 기존의 사람 재식별 장치를 위한 학습 시에도 이용되는 손실로서, 교차 엔트로피 및 삼중항 손실로 계산될 수 있으며, 이는 공지된 기술이므로 여기서는 상세하게 설명하지 않는다.

[0096] 도 16은 본 발명의 일 실시예에 따른 사람 재식별 방법을 나타낸다.

[0097] 도 2 내지 도 15를 참조하여, 도 16의 사람 재식별 방법을 설명하면, 우선 타겟 사람이 포함되었는지 여부가 확인되어야 하는 다수의 프레임으로 구성된 동영상을 획득한다(S10).

[0098] 그리고 미리 학습된 인공 신경망을 이용하여, 각 프레임(F_i)에 대해 신경망 연산으로 인코딩하여 각 프레임(F_i)에 대응하는 공간 쿼리맵(q^s_i), 시간 쿼리맵(q^t_i) 및 사람 특징맵(f^o_i)을 획득한다(S20). 이때 공통의 백본 레이어가 인가된 프레임(F_i)에 대해 신경망 연산하여 공통 특징맵을 획득하고, 3개의 헤드 레이어가 공통 특징맵에 대해 서로 다르게 신경망 연산하여 공간 쿼리맵(q^s_i), 시간 쿼리맵(q^t_i) 및 사람 특징맵(f^o_i)을 획득할 수 있다.

[0099] 각 프레임(F_i)에 대한 공간 쿼리맵(q^s_i), 시간 쿼리맵(q^t_i) 및 사람 특징맵(f^o_i)이 획득되면, 이중 공간 쿼리맵(q^s_i)에 포함된 산만 요인 성분에 대한 대표 특징을 공간 메모리(320)에 미리 저장된 공간 아이템을 기반으로 획득하는 공간적 산만 요인 추출 단계를 수행한다(S30).

[0100] 공간적 산만 요인 추출 단계(S30)에서는 우선 획득된 공간 쿼리맵(q^s_i)에서 채널 방향의 다수의 공간 픽셀 벡터($q^s_{i,k}$)를 각각 선택한다(S31). 그리고 각각 공간적 산만 요인에 대한 프로토타입 특징과 대표 표현자가 공간 키(k^s)와 공간 값(v^s)으로 서로 매칭되어 포함된 M개의 공간 아이템이 저장된 공간 메모리(320)에서 다수의 공간 키(k^s)를 리드한다(S32). 이에 선택된 공간 픽셀 벡터($q^s_{i,k}$) 각각과의 코사인 유사도를 계산하고 정규화하여, 각 공간 픽셀 벡터($q^s_{i,k}$)가 M개의 공간 키(k^s) 중 n번째 공간 키(k^s_n)에 매칭될 확률을 나타내는 공간 가중치($a^s_{i,k,n}$)를 계산한다(S33).

[0101] 공간 가중치($a^s_{i,k,n}$)가 계산되면, 공간 메모리(320)에서 다수의 공간 값(v^s)을 리드하고, 리드된 다수의 공간 값(v^s)에 대응하는 공간 가중치($a^s_{i,k,n}$)를 가중합하여 공간 강조 벡터($o^s_{i,k}$)를 획득한다(S34).

[0102] 다수의 공간 픽셀 벡터($q^s_{i,k}$)에 대한 다수의 공간 강조 벡터($o^s_{i,k}$)가 획득되면, 다수의 공간 강조 벡터($o^s_{i,k}$)를 재배치하여 각 프레임(F_i)에서 추출된 공간 쿼리맵(q^s_i)에 대응하는 공간 강조맵(o^s_i)을 획득한다(S35).

[0103] 그리고 공간 강조맵(o^s_i)이 획득되지 않은 나머지 프레임이 존재하는지 판별한다(S36). 나머지 프레임이 존재하는 것으로 판별되면, 해당 프레임을 인코딩하여 공간 쿼리맵(q^s_i)을 획득하고(S20), 획득된 공간 쿼리맵(q^s_i)을 이용하여 다시 공간적 산만 요인 추출 단계(S30)를 수행한다. 여기서 프레임 수는 동영상에 포함된 모든 프레임일 수도 있으나, 기지정된 개수(L)로 선택된 제한된 개수의 프레임일 수도 있다.

[0104] 반면, 모든 프레임에 대한 공간 강조맵(o^s_i)이 획득된 것으로 판별되면, 다수의 프레임(F_1 ~ F_L)에서 주의해야하는 프레임을 선택하기 위한 시간적 주의 단계를 수행한다(S40).

[0105] 시간적 주의 단계(S40)에서는 먼저 다수의 프레임(F_1 ~ F_L) 각각에서 추출된 다수의 공간 쿼리맵(q^s_1 ~ q^s_L) 각각에 대해 글로벌 평균 풀링(GAP)을 수행하여 다수의 시간 쿼리 벡터를 획득한다(S41). 그리고 다수의 시간 쿼리 벡터에 대해 미리 학습된 인공 신경망으로 신경망 연산하여 다수의 시간 쿼리 벡터의 시간적 변화 특성을 나타내는 벡터 형태의 시간 문맥 벡터(q^t)를 획득한다(S42).

[0106] 시간 문맥 벡터(q^t)가 획득되면, 각각 각 프레임의 중요도에 따른 시간적 패턴 특징과 시간적 패턴 특징에 따른 시간적 주의도가 시간 키(k^t)와 시간 값(v^t)으로 서로 매칭되어 포함된 N개의 시간 아이템이 저장된 시간 메모리(430)에서 다수의 시간 키(k^t)를 리드한다(S43). 그리고 시간 문맥 벡터(q^t)와 다수의 시간 키(k^t) 각각과의 코사인 유사도를 계산하고 정규화하여, 시간 문맥 벡터(q^t)와가 N개의 시간 키(k^t) 각각에 매칭될 확률을 나타내는 시간 가중치(a^t_n)를 계산한다(S44). 그리고 시간 메모리(430)에서 다수의 시간 값(v^t)을 리드하고, 리드된 다수의 시간 값(v^t) 각각에 시간 가중치(a^t_n)를 가중합하여 시간 강조 벡터(o^t)를 획득한다(S45).

[0107] 다수의 공간 쿼리맵(q^s_i)에 대응하여 다수의 공간 강조맵(o^s_i)이 획득되면, 다수의 사람 특징맵(f^o_i) 각각에 대응하는 공간 강조맵(o^s_i)을 차감하여 정제함으로써, 사람 특징맵(f^o_i)에서 산만 요인이 제거된 다수의 정제 사람 특징맵(f^s_i)을 획득한다(S51). 이후, 다수의 정제 사람 특징맵(f^s_i) 각각에 대해 글로벌 평균 풀링하여 다수의 정제 사람 특징 벡터를 획득한다(S52). 그리고 획득된 다수의 정제 사람 특징 벡터에 시간 강조 벡터(o^t)의 대응하는 각 원소를 가중하여, 시간적 주의가 강조된 사람 표현자(f^t)를 획득한다(S53).

[0108] 산만 요인이 배제되고 시간적으로 강조된 사람 표현자(f^t)가 획득되면, 미리 학습된 인공 신경망으로 획득된 사람 표현자(f^t)에 대해 신경망 연산하여 사람 표현자(f^t)에 대응하는 식별자를 추출하고, 지정된 타겟 사람과 식별자가 일치하는지 여부를 판별함으로써 사람을 재식별한다(S60).

[0109] 본 발명에 따른 방법은 컴퓨터에서 실행시키기 위한 매체에 저장된 컴퓨터 프로그램으로 구현될 수 있다. 여기서 컴퓨터 판독가능 매체는 컴퓨터에 의해 액세스 될 수 있는 임의의 가용 매체일 수 있고, 또한 컴퓨터 저장 매체를 모두 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함하며, ROM(판독 전용 메모리), RAM(랜덤 액세스 메모리), CD(컴팩트 디스크)-ROM, DVD(디지털 비디오 디스크)-ROM, 자기 테이프, 플로피 디스크, 광데이터 저장장치 등을 포함할 수 있다.

[0110] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다.

[0111] 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 청구범위의 기술적 사상에 의해 정해져야 할 것이다.

부호의 설명

[0112] 100: 동영상 획득부 200: 인코더

300: 공간 메모리 네트워크 310: 공간 벡터 선택부

320: 공간 메모리 330: 공간 유사도 계산부

340: 공간 강조맵 획득부 400: 시간 메모리 네트워크

410: 시간 쿼리 풀링부 420: 시간적 문맥 인코딩부

430: 시간 메모리 440: 시간 유사도 계산부

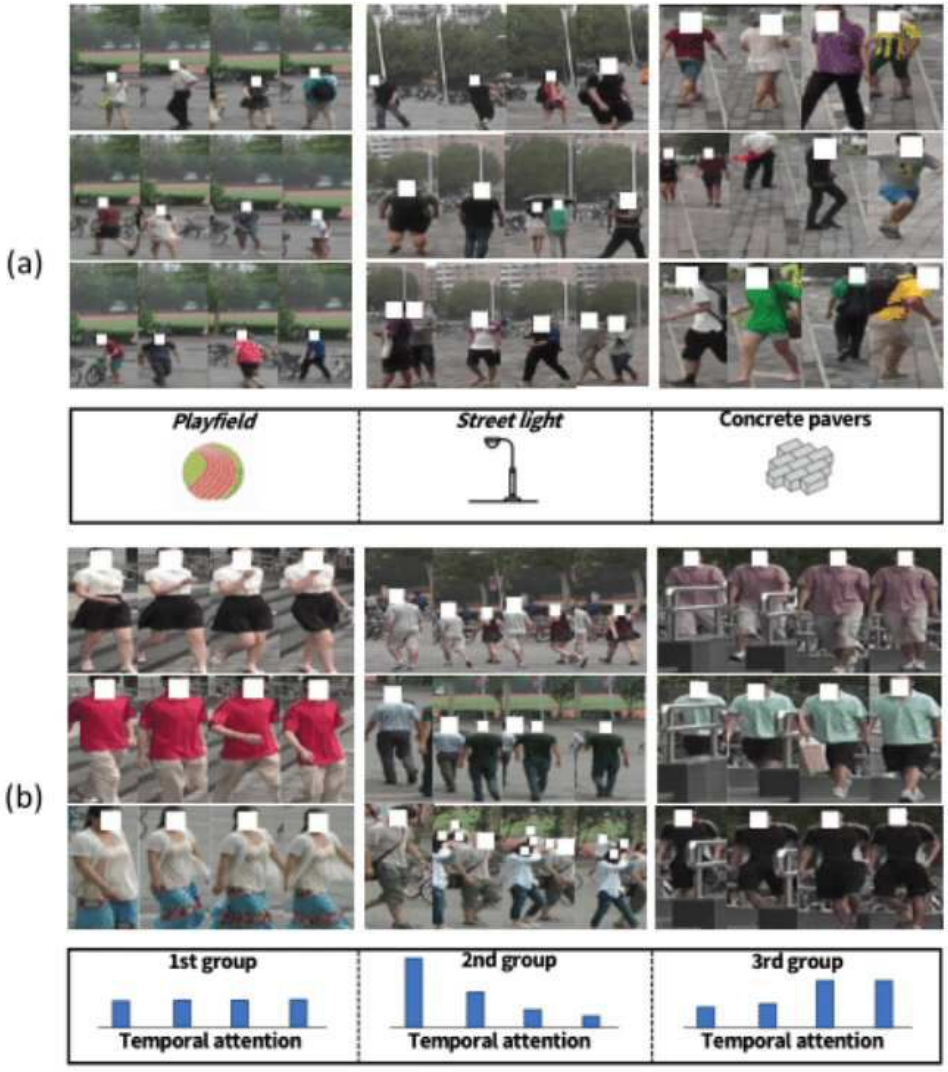
450: 시간 강조 벡터 획득부 500: 사람 표현자 획득부

510: 공간적 사람 특징맵 정제부 520: 정제 사람 특징맵 풀링부

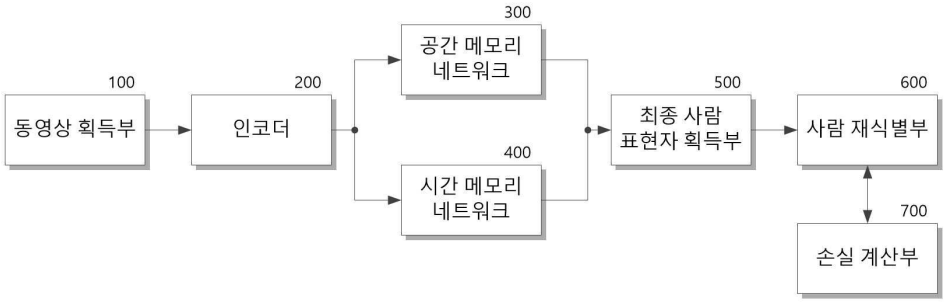
530: 시간적 강조부 600: 사람 재식별부

도면

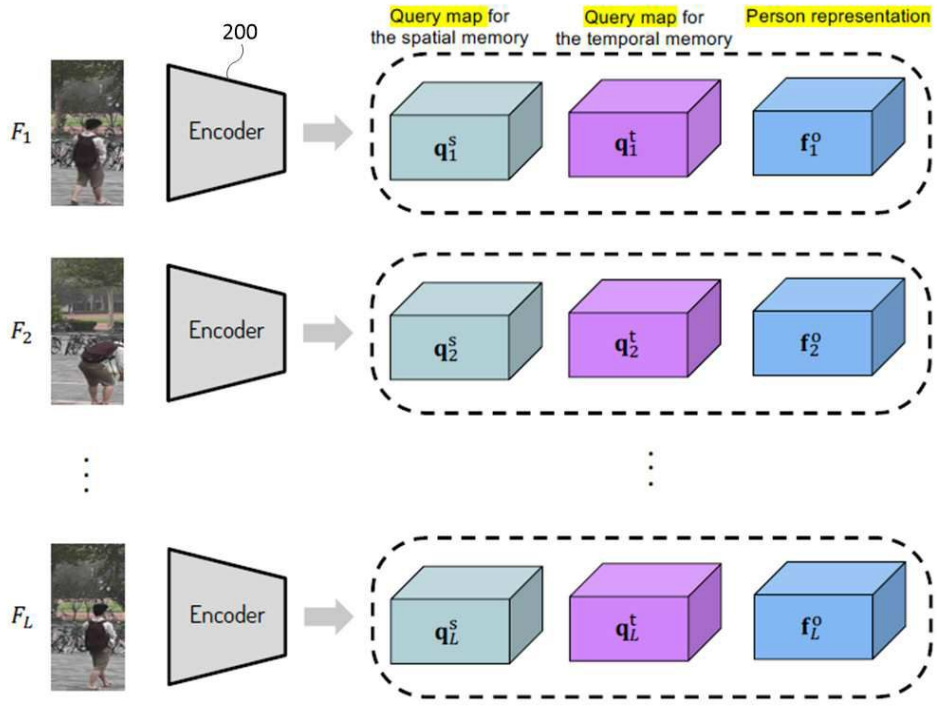
도면1



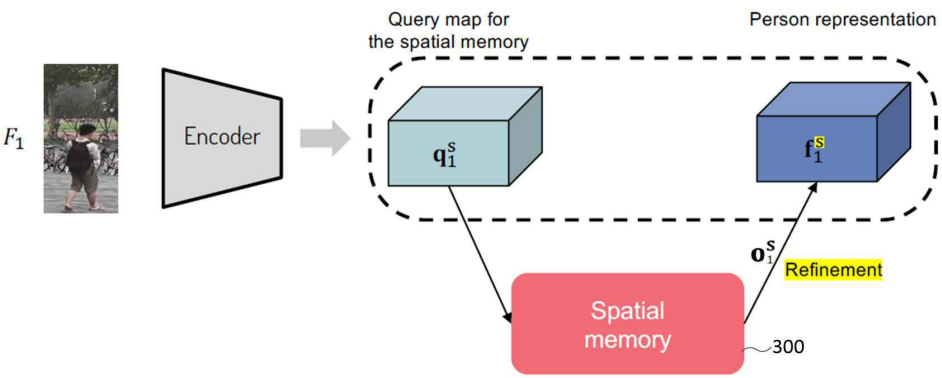
도면2


100
동영상 획득부
200
인코더
300
공간 메모리 네트워크
400
시간 메모리 네트워크
500
최종 사람 표현자 획득부
600
사람 재식별부
700
손실 계산부

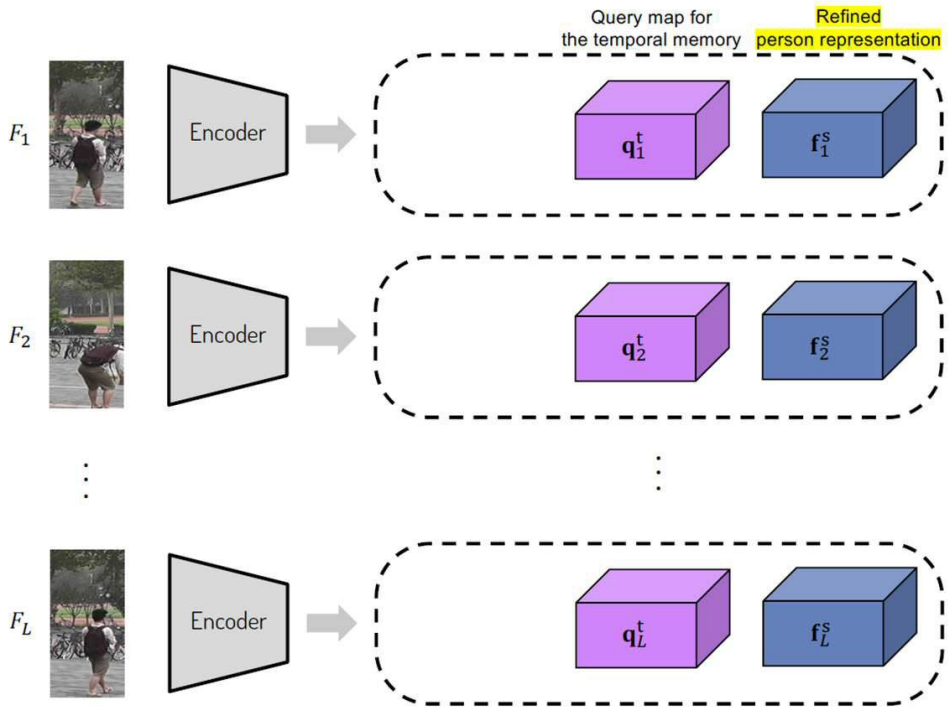
도면3


200
Encoder
Query map for the spatial memory
Query map for the temporal memory
Person representation
F_1
F_2
F_L
$\mathbf{q}_1^s$
$\mathbf{q}_1^t$
$\mathbf{f}_1^o$
$\mathbf{q}_2^s$
$\mathbf{q}_2^t$
$\mathbf{f}_2^o$
$\mathbf{q}_L^s$
$\mathbf{q}_L^t$
$\mathbf{f}_L^o$

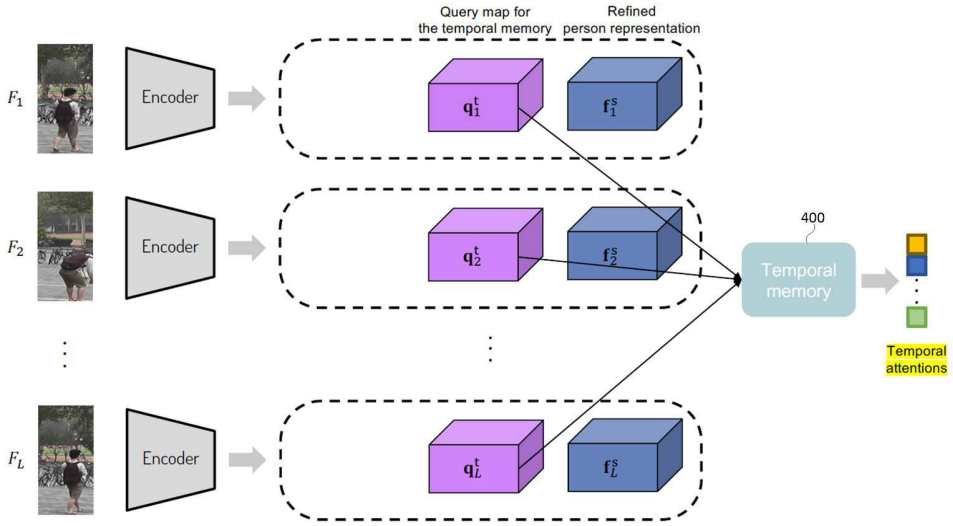
도면4


Query map for the spatial memory
Person representation
Encoder
F_1
$\mathbf{q}_1^s$
$\mathbf{f}_1^s$
$\mathbf{o}_1^s$
Refinement
Spatial memory
300

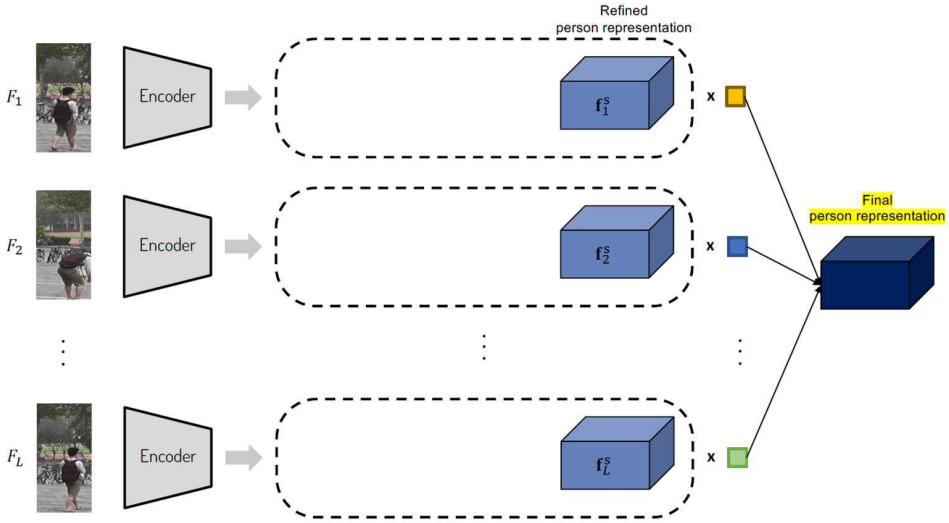
도면5


Query map for
the temporal memory
Refined
person representation
F_1
Encoder
$\mathbf{q}_1^t$
$\mathbf{f}_1^s$
F_2
Encoder
$\mathbf{q}_2^t$
$\mathbf{f}_2^s$
F_L
Encoder
$\mathbf{q}_L^t$
$\mathbf{f}_L^s$

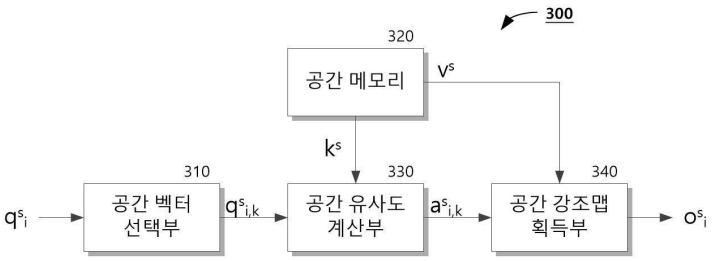
도면6


Query map for
the temporal memory
Refined
person representation
F_1
Encoder
$\mathbf{q}_1^t$
$\mathbf{f}_1^s$
F_2
Encoder
$\mathbf{q}_2^t$
$\mathbf{f}_2^s$
400
Temporal
memory
Temporal
attentions
F_L
Encoder
$\mathbf{q}_L^t$
$\mathbf{f}_L^s$

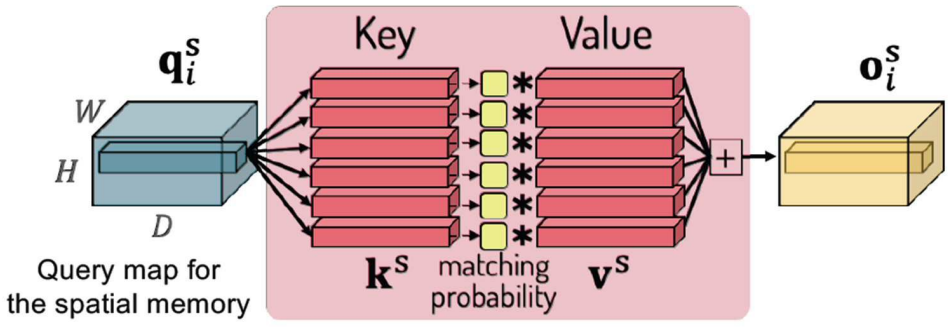
도면7



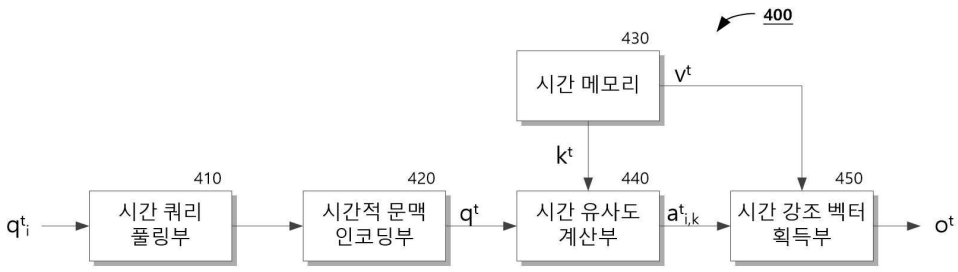
도면8



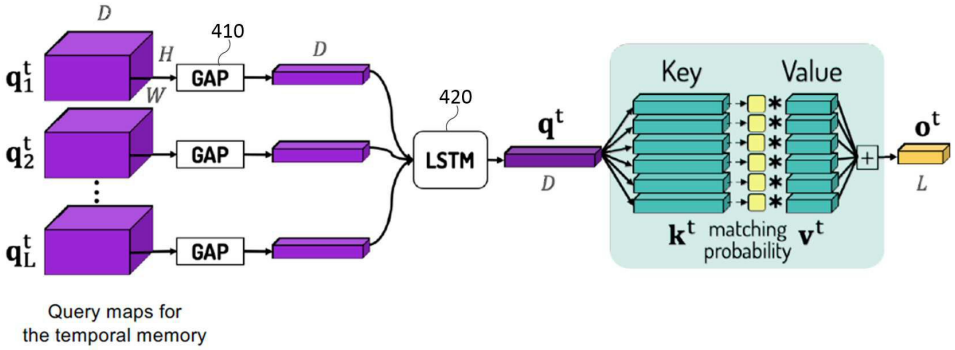
도면9



도면10

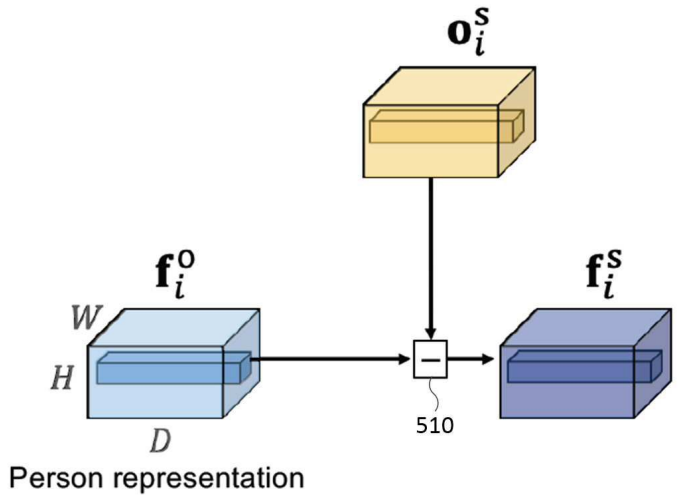


도면11

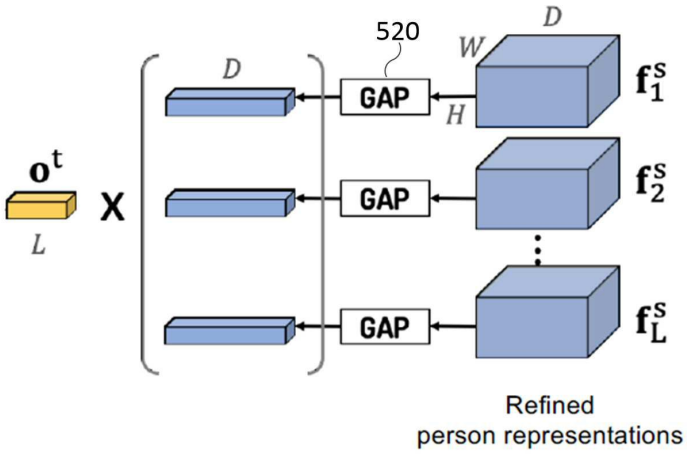


도면12

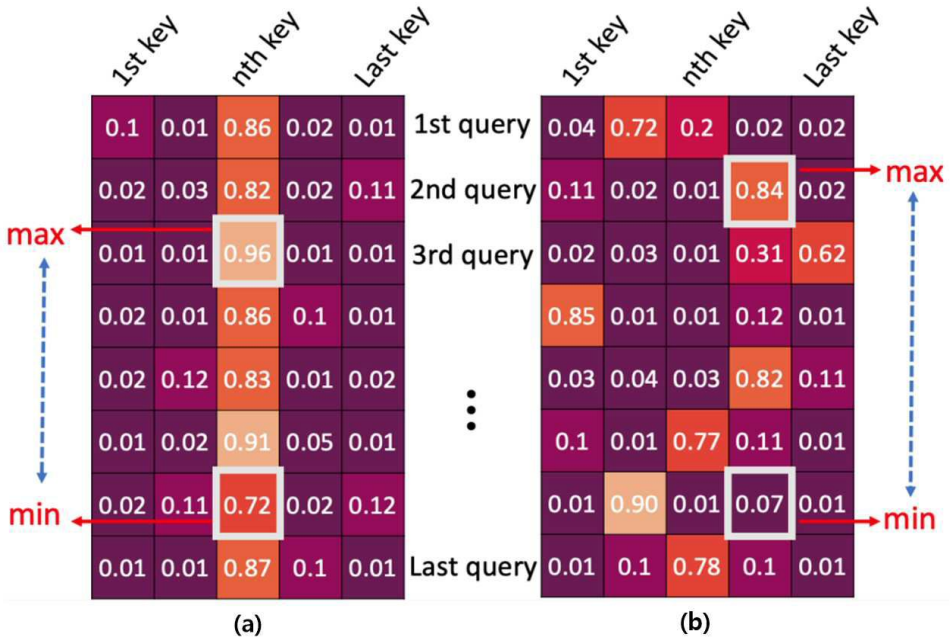
도면13



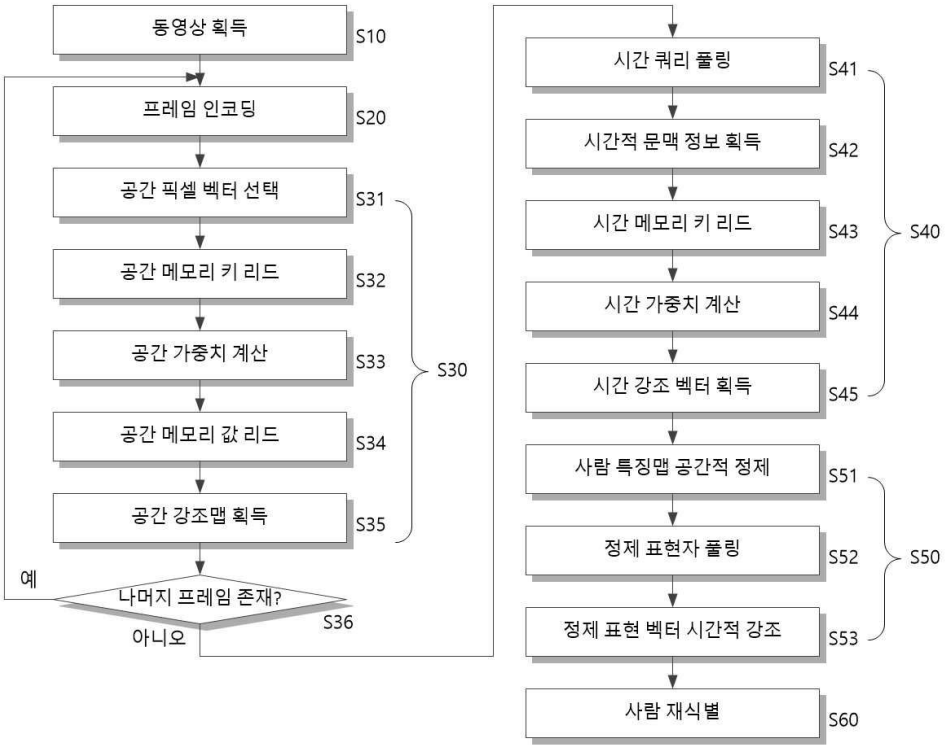
도면14


520
D
W
H
D
GAP
GAP
GAP
f_1^s
f_2^s
f_L^s
o^t
X
L
Refined
person representations

도면15


1st key
nth key
Last key
1st query
2nd query
3rd query
Last query
max
min
(a)
(b)

도면16


동영상 획득
S10
프레임 인코딩
S20
공간 픽셀 벡터 선택
S31
공간 메모리 키 리드
S32
공간 가중치 계산
S33
공간 메모리 값 리드
S34
공간 강조맵 획득
S35
S30
예
나머지 프레임 존재?
S36
아니오
시간 쿼리 풀링
S41
시간적 문맥 정보 획득
S42
시간 메모리 키 리드
S43
시간 가중치 계산
S44
시간 강조 벡터 획득
S45
S40
사람 특징맵 공간적 정제
S51
정제 표현자 풀링
S52
정제 표현 벡터 시간적 강조
S53
S50
사람 재식별
S60