

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)(11) 공개번호 10-2021-0093484
(43) 공개일자 2021년07월28일

(51) 국제특허분류(Int. Cl.)
 G06N 3/08 (2006.01) G06K 9/00 (2006.01)
 G06N 3/04 (2006.01) G06T 7/11 (2017.01)
 (52) CPC특허분류
 G06N 3/08 (2013.01)
 G06K 9/00288 (2013.01)
 (21) 출원번호 10-2020-0007078
 (22) 출원일자 2020년01월20일
 심사청구일자 2020년01월20일

(71) 출원인
 인하대학교 산학협력단
 인천광역시 미추홀구 인하로 100(용현동, 인하대학교)
 (72) 발명자
 송병철
 서울특별시 양천구 목동서로 38, 125동 307호(목동, 목동신시가지아파트1단지)
 최동윤
 충청남도 서산시 고북면 고북1로 299, 204호
 (74) 대리인
 양정보

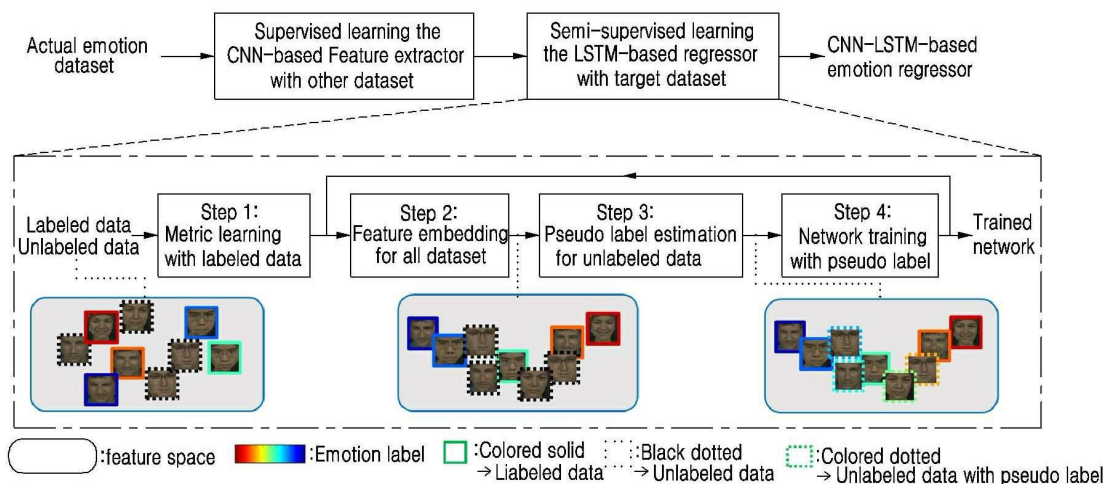
전체 청구항 수 : 총 8 항

(54) 발명의 명칭 실시간 감정인식 어플리케이션을 위한 계층 학습 기반 반지도 학습 알고리즘

(57) 요약

실시간 감정인식 어플리케이션을 위한 계층 학습 기반 반지도 학습 알고리즘이 개시된다. 일 실시예에 따른 감정 인식 시스템에 의해 수행되는 감정 인식용 반지도 학습 방법은, 동영상 데이터가 입력됨을 수신함에 따라 동영상 데이터의 각 프레임으로부터 얼굴 영역을 검출하는 단계; 상기 검출된 얼굴 영역에 대하여 두 개 이상의 학습 네트워크가 결합된 학습 모델을 이용하여 의사 라벨 기반의 반지도 학습을 수행하는 단계; 및 상기 수행된 반지도 학습을 통하여 획득된 학습 결과로서, 상기 검출된 얼굴 영역에 대한 감정 정보를 추정하는 단계를 포함할 수 있다.

대표도



(52) CPC특허분류

G06N 3/0454 (2013.01)

G06T 7/11 (2017.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1415162019
부처명	산업통상자원부
과제관리(전문)기관명	한국산업기술평가관리원
연구사업명	산업핵심기술개발사업
연구과제명	인간 내면상태의 인식 및 이를 이용한 인간친화형 인간-로봇 상호작용 기술 개발
(3차년도)	
기 여 율	1/1
과제수행기관명	인하대학교 산학협력단
연구기간	2019.01.01 ~ 2019.12.31

명세서

청구범위

청구항 1

감정 인식 시스템에 의해 수행되는 감정 인식용 반지도 학습 방법에 있어서,

동영상 데이터가 입력됨을 수신함에 따라 동영상 데이터의 각 프레임으로부터 얼굴 영역을 검출하는 단계;

상기 검출된 얼굴 영역에 대하여 두 개 이상의 학습 네트워크가 결합된 학습 모델을 이용하여 의사 라벨 기반의 반지도 학습을 수행하는 단계; 및

상기 수행된 반지도 학습을 통하여 획득된 학습 결과로서, 상기 검출된 얼굴 영역에 대한 감정 정보를 추정하는 단계

를 포함하는 감정 인식용 반지도 학습 방법.

청구항 2

제1항에 있어서,

상기 두 개 이상의 학습 네트워크가 결합된 학습 모델은,

CNN과 LSTM이 결합된 학습 모델로서, 지도 학습 기반의 CNN 학습과 반지도 학습 기반의 LSTM 학습을 각각 독립적으로 수행하는

감정 인식용 반지도 학습 방법.

청구항 3

제1항에 있어서,

상기 반지도 학습을 수행하는 단계는,

지도 학습 기반의 CNN 학습을 통하여 상기 검출된 얼굴 영역에 대한 특징을 추출하고, 상기 추출된 특징을 반지도 학습 기반의 LSTM에 입력하는 단계

를 포함하는 감정 인식용 반지도 학습 방법.

청구항 4

제3항에 있어서,

상기 반지도 학습을 수행하는 단계는,

반지도 학습 기반의 LSTM 학습 과정에서, 트레이닝 데이터셋(dataset) 중 분류된 데이터를 이용하여 메트릭 학습(metric learning)을 통해 LSTM 네트워크를 프리 트레이닝(pre-training)하고, 상기 프리 트레이닝된 LSTM 네트워크를 이용하여 분류된 데이터와 비분류된 데이터에 대한 임베디드된 특징을 추출하고, 상기 추출된 임베디드된 특징의 공간에서 특징간 거리를 이용하여 비분류된 데이터의 의사 라벨을 추정하고, 상기 추정된 의사 라벨과 함께 상기 분류된 데이터와 상기 비분류된 데이터를 학습시키는 단계

를 포함하는 감정 인식용 반지도 학습 방법.

청구항 5

제4항에 있어서,

상기 반지도 학습을 수행하는 단계는,

상기 메트릭 학습에서, 연속성있는 라벨의 데이터셋의 분류된 데이터를 대상으로 앵커 데이터와 이웃 데이터를 배치(batch) 단위로 구성하여 학습을 수행하고, 상기 배치 단위로 구성된 앵커 데이터를 LSTM 네트워크로부터

임베디드된 특징을 출력하고, 상기 출력된 임베디드된 특징들로부터 구성된 앵커 특징 배치를 랜덤하게 서플링하여 이웃 배치를 구성하고, 특징 배치들에 대해 손실(loss)를 계산하는 과정을 수행하는 단계를 포함하는 감정 인식용 반지도 학습 방법.

청구항 6

제4항에 있어서,
상기 반지도 학습을 수행하는 단계는,
상기 비분류된 데이터와 근접한 분류된 데이터의 라벨을 이용하여 비분류된 데이터의 의사 라벨로 활용하는 단계를 포함하는 감정 인식용 반지도 학습 방법.

청구항 7

제4항에 있어서,
상기 반지도 학습을 수행하는 단계는,
상기 비분류된 라벨을 포함하는 비분류된 데이터를 이용하여 네트워크 학습을 수행하며, 상기 추정된 의사 라벨의 불확실성을 측정하여 상기 비분류된 데이터에 대한 트레이닝 손실의 부분을 조정하는 단계를 포함하는 감정 인식용 반지도 학습 방법.

청구항 8

감정 인식 시스템에 있어서,
동영상 데이터가 입력됨을 수신함에 따라 동영상 데이터의 각 프레임으로부터 얼굴 영역을 검출하는 검출부;
상기 검출된 얼굴 영역에 대하여 두 개 이상의 학습 네트워크가 결합된 학습 모델을 이용하여 의사 라벨 기반의 반지도 학습을 수행하는 수행부; 및
상기 수행된 반지도 학습을 통하여 획득된 학습 결과로서, 상기 검출된 얼굴 영역에 대한 감정 정보를 추정하는 추정부를 포함하는 감정 인식 시스템.

발명의 설명

기술 분야

[0001] 아래의 설명은 실시간 감정인식 어플리케이션을 위한 계층 학습 기반 반지도 학습 기술에 관한 것이다.

배경 기술

[0003] 인간과 인공지능의 감정적인 상호작용을 위하여 감정인 기술은 매우 중요한 요소이다. 최근 빅데이터와 딥러닝 기술의 발달과 함께 감정인식 기술 또한 발전되었으며 감정인식 기술의 성능도 크게 향상되었다. 최근 자연스럽게 발생하는 감정을 인식하기 위하여 감정인식 연구에서는 감정의 종류 예(화남, 기쁨, 슬픔, 놀람 등) 등을 분리하는 것이 아닌 valence(긍정/부정), arousal(흥분 정도)를 추정하는 연구가 진행되고 있다. 실제 인간 감정반응을 수집한 데이터에 대하여 label 정보를 지정하는 것은 시간과 비용이 매우 크게 들어가는 작업이다. 감정인식 알고리즘 학습 데이터의 어노테이션 비용(annotation cost)을 절감하기 위하여 반지도 학습(semi-supervised learning)을 적용할 수 있다. 반지도 학습은 라벨(label)이 존재하지 않는 비분류 데이터(unlabeled data)를 동시에 이용하여 라벨이 존재하는 데이터셋만을 이용하는 것보다 학습 성능을 향상할 수 있다.

발명의 내용

해결하려는 과제

- [0005] 감정 인식 데이터셋의 어노테이션 비용(annotation cost)을 절감하기 위해 반지도 학습(SSL, semi-supervised learning)을 제공할 수 있다.
- [0006] 연속성 있는 영역(continuous domain)의 라벨(label)을 추정하는 감정 인식 방법 및 시스템을 제공할 수 있다.
- [0007] 다수의 비분류된 데이터(unlabeled data)를 함께 학습함으로써 학습 성능 및 인식 성능을 향상하는 방법 및 시스템을 제공할 수 있다.
- [0008] 유인가/각성도(valence/arousal)와 같은 연속성 있는 영역(continuous domain)의 정보를 출력하며 기존의 분류(classification)보다 복합적인 감정 인식을 수행하는 방법 및 시스템을 제공할 수 있다.

과제의 해결 수단

- [0010] 감정 인식 시스템에 의해 수행되는 감정 인식용 반지도 학습 방법은, 동영상 데이터가 입력됨을 수신함에 따라 동영상 데이터의 각 프레임으로부터 얼굴 영역을 검출하는 단계; 상기 검출된 얼굴 영역에 대하여 두 개 이상의 학습 네트워크가 결합된 학습 모델을 이용하여 의사 라벨 기반의 반지도 학습을 수행하는 단계; 및 상기 수행된 반지도 학습을 통하여 획득된 학습 결과로서, 상기 검출된 얼굴 영역에 대한 감정 정보를 추정하는 단계를 포함할 수 있다.
- [0011] 상기 두 개 이상의 학습 네트워크가 결합된 학습 모델은, CNN과 LSTM이 결합된 학습 모델로서, 지도 학습 기반의 CNN 학습과 반지도 학습 기반의 LSTM 학습을 각각 독립적으로 수행할 수 있다.
- [0012] 상기 반지도 학습을 수행하는 단계는, 지도 학습 기반의 CNN 학습을 통하여 상기 검출된 얼굴 영역에 대한 특징을 추출하고, 상기 추출된 특징을 반지도 학습 기반의 LSTM에 입력하는 단계를 포함할 수 있다.
- [0013] 상기 반지도 학습을 수행하는 단계는, 반지도 학습 기반의 LSTM 학습 과정에서, 트레이닝 데이터셋(dataset) 중 분류된 데이터를 이용하여 메트릭 학습(metric learning)을 통해 LSTM 네트워크를 프리 트레이닝(pre-training)하고, 상기 프리 트레이닝된 LSTM 네트워크를 이용하여 분류된 데이터와 비분류된 데이터에 대한 임베디드된 특징을 추출하고, 상기 추출된 임베디드된 특징의 공간에서 특징간 거리를 이용하여 비분류된 데이터의 의사 라벨을 추정하고, 상기 추정된 의사 라벨과 함께 상기 분류된 데이터와 상기 비분류된 데이터를 학습시키는 단계를 포함할 수 있다.
- [0014] 상기 반지도 학습을 수행하는 단계는, 상기 메트릭 학습에서, 연속성있는 라벨의 데이터셋의 분류된 데이터를 대상으로 앵커 데이터와 이웃 데이터를 배치(batch) 단위로 구성하여 학습을 수행하고, 상기 배치 단위로 구성된 앵커 데이터를 LSTM 네트워크로부터 임베디드된 특징을 출력하고, 상기 출력된 임베디드된 특징들로부터 구성된 앵커 특징 배치를 랜덤하게 서플링하여 이웃 배치를 구성하고, 특징 배치들에 대해 손실(loss)를 계산하는 과정을 수행하는 단계를 포함할 수 있다.
- [0015] 상기 반지도 학습을 수행하는 단계는, 상기 비분류된 데이터와 근접한 분류된 데이터의 라벨을 이용하여 비분류된 데이터의 의사 라벨로 활용하는 단계를 포함할 수 있다.
- [0016] 상기 반지도 학습을 수행하는 단계는, 상기 비분류된 라벨을 포함하는 비분류된 데이터를 이용하여 네트워크 학습을 수행하며, 상기 추정된 의사 라벨의 불확실성을 측정하여 상기 비분류된 데이터에 대한 트레이닝 손실의 부분을 조정하는 단계를 포함할 수 있다.
- [0017] 감정 인식 시스템은, 동영상 데이터가 입력됨을 수신함에 따라 동영상 데이터의 각 프레임으로부터 얼굴 영역을 검출하는 검출부; 상기 검출된 얼굴 영역에 대하여 두 개 이상의 학습 네트워크가 결합된 학습 모델을 이용하여 의사 라벨 기반의 반지도 학습을 수행하는 수행부; 및 상기 수행된 반지도 학습을 통하여 획득된 학습 결과로서, 상기 검출된 얼굴 영역에 대한 감정 정보를 추정하는 추정부를 포함할 수 있다.

발명의 효과

- [0019] 감정 인식 데이터셋의 어노테이션 비용(annotation cost)을 감소하기 위한 목적으로 의사 라벨(pseudo label) 기반의 반지도 학습을 제공할 수 있다. 또한, 로그 비율 손실(log-ratio loss)기반의 메트릭 학습(metric learning)을 이용하여 연속성 있는 라벨(continuous label)을 갖는 감정 데이터의 특징 임베딩(feature embedding) 성능을 향상시킬 수 있다. 메트릭 학습 과정을 통하여 의사 라벨 추정의 정확도를 개선하고, 반지도 학습 과정이 네트워크의 성능을 개선할 수 있다. 이를 통하여 감정 인식 데이터셋에 대한 어노테이션 비용을 감소시키는데 기여할 수 있다.

도면의 간단한 설명

- [0021] 도 1은 일 실시예에 따른 감정 인식 시스템의 감정인식 학습 과정을 설명하기 위한 도면이다.
- 도 2는 일 실시예에 따른 감정 인식 시스템에서 CNN-LSTM 기반의 회귀 동작을 설명하기 위한 도면이다.
- 도 3은 일 실시예에 따른 감정 인식 시스템에서 의사 라벨을 추정하는 과정을 설명하기 위한 도면이다.
- 도 4는 일 실시예에 따른 감정 인식 시스템에서 반지도 학습으로 학습한 학습 모델의 결과를 설명하기 위한 도면이다.
- 도 5는 일 실시예에 따른 감정 인식 시스템에서 메트릭 학습을 적용하여 특징을 분석하는 것을 설명하기 위한 예이다.
- 도 6은 일 실시예에 따른 감정 인식 시스템의 구성을 설명하기 위한 블록도이다.
- 도 7은 일 실시예에 따른 감정 인식 시스템에서 감정을 인식하는 방법을 설명하기 위한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0022] 이하, 실시예를 첨부한 도면을 참조하여 상세히 설명한다.
- [0024] 실시예에서는 감정인식 데이터 세트(dataset)의 어노테이션 비용(annotation cost)을 절감하기 위해 반지도 학습(SSL, semi-supervised learning)을 제안하기로 한다. 반지도 학습은 분류된 데이터(labeled data)와 비분류된 데이터(unlabeled data)를 동시에 이용하여 네트워크를 학습하는 방식으로서, 분류된 데이터만을 이용하는 지도 학습(supervised learning(SL))보다 높은 학습 성능을 획득할 수 있다.
- [0025] 도 1은 일 실시예에 따른 감정 인식 시스템의 감정인식 학습 과정을 설명하기 위한 도면이다.
- [0026] 실시예에서는 동영상에서의 연속성 있는 영역(continuous domain)의 감정정보를 회귀(regression)하기 위한 의사 라벨(pseudo label)기반 반지도 학습을 제안한다. 이러한 반지도 학습은 미리 학습된 네트워크를 이용하여 모든 데이터의 임베디드된 특징(embedded feature)를 추출하고, 특징 공간(feature space)에서 비분류된 데이터(unlabeled data)와 분류된 데이터(labeled data)의 관계를 분석하여 비분류된 데이터(unlabeled data)의 의사 라벨(pseudo label)을 추정한다. 그런데 일반적인 방법으로 학습된 네트워크로 데이터들의 특징 공간을 생성하게 되면, 얼굴 특성이 유사한 데이터들 간 상관성이 커지는 현상이 발생할 수 있다. 다시 말해서, 감정의 유사도가 아닌 얼굴의 유사도가 큰 분류된 데이터의 라벨이 비분류된 데이터의 의사 라벨로 추정될 수 있기 때문에 반지도 학습의 성능이 저하될 수 있다. 이에 따라, 서브젝트 ID(subject ID)의 영향을 최소화하면서 특징 임베딩(feature embedding)을 하기 위해 네트워크를 메트릭 학습(metric learning)을 이용하여 프리 트레이닝(pre-training)할 수 있다. 메트릭 학습은 실제의 데이터들이 갖는 시맨틱 정보(semantic information)의 유사도 및 상관 관계를 임베디드된 특징 공간(embedded feature space)에서도 표현할 수 있는 네트워크를 학습하는 방법이다.
- [0027] 감정 인식 시스템은 얼굴 이미지 시퀀스(facial image sequence)에 대한 감정정보를 추정하는 CNN-LSTM기반의 회귀를 학습할 수 있다. 이때, CNN과 LSTM의 학습은 독립적으로 이루어질 수 있다. CNN은 별도의 데이터셋(dataset)으로 학습을 진행하고, 실시예에서 제안하는 반지도 학습(SSL)은 LSTM에만 적용될 수 있다. LSTM에 대한 반지도 학습 과정은 도 2와 같은 과정으로 진행될 수 있다.

- [0028] 감정 인식 시스템은 먼저 전체 트레이닝 데이터셋(training dataset) 중 분류된 데이터(labeled data)만을 이용하여 메트릭 학습을 통해 LSTM 네트워크를 프리 트레이닝할 수 있다(step 1). 다음으로, 프리 트레이닝된 LSTM 네트워크를 이용하여 분류된 데이터와 비분류된 데이터에 대한 임베디드된 특징을 추출할 수 있다(step 2). 임베디드된 특징 공간에서 특징들 간의 거리(distance)를 이용하여 비분류된 데이터의 의사 라벨(pseudo label)을 추정하고(step 3), 분류된 데이터(labeled data)와 함께 네트워크를 학습할 수 있다(step 4). 의사 라벨 추정(Pseudo label estimation) 및 네트워크 학습에 해당하는 step 2 내지 step 4의 과정이 반복될 수 있다.
- [0029] 도 2를 참고하면, CNN-LSTM기반의 네트워크는 CNN과 LSTM이 결합된 형태로서 동영상 분석이나 감정 인식 분야에서 널리 이용되고 있다. 감정 인식 시스템은 얼굴 동영상으로부터 각성도(arousal) 및 유인가(valence) 정보를 인식하기 위하여 CNN-LSTM 기반의 회귀를 이용할 수 있다. 감정 인식 시스템은 입력 동영상의 각 프레임으로부터 얼굴 영역을 검출할 수 있다. 다음으로 CNN-based feature extractor를 이용하여 각 얼굴 이미지의 특징을 추출하고, LSTM이 특징 시퀀스를 연속적으로 처리하여 최종적인 감정 정보를 회귀할 수 있다.
- [0030] 구체적으로, 감정 인식 시스템은 연속성 있는 라벨을 위한 메트릭 학습을 수행할 수 있다. 감정 인식 시스템은 임베디드된 특징 공간에서 비분류된 데이터 $\mathbf{x}_U$ 의 임베디드된 특징 $\mathbf{f}_{x_U}$ 와 유클리드 거리(Euclidean distance)가 가장 작은 분류된 데이터의 특징 $\mathbf{f}_{x_L}$ 을 추출하고, 대응하는 데이터의 라벨 정보에 기반하여 의사 라벨을 추정할 수 있다. 의사 라벨 추정의 정확도를 향상시키기 위하여 특징 임베딩의 성능이 매우 중요하다. 감정 인식 시스템은 메트릭 학습을 이용하여 특징 임베딩의 성능을 향상시킬 수 있다. step 1을 참고하면, 감정 인식 시스템은 연속성 있는 라벨의 데이터셋의 분류된 데이터만을 대상으로 메트릭 학습을 수행할 수 있다. 감정 인식 시스템은 앵커 데이터(anchor data) $\mathbf{x}_a$, 이웃 데이터(neighbor data) $\mathbf{x}_b, \mathbf{x}_c$ 를 배치(batch) 단위로 구성하여 메트릭 학습을 진행할 수 있다. 먼저, 다양한 데이터로 구성된 미니 배치(mini-batch) $\mathbf{x}_f$ 를 앵커 데이터 $\mathbf{x}_a$ 로 정의하고, 네트워크로부터 임베디드된 특징을 출력할 수 있다. 임베디드된 특징들로부터 구성된 앵커 특징 배치 $\mathbf{f}_a$ 를 랜덤하게 셔플링(shuffling)하여 이웃 배치(neighbor batch) $\mathbf{f}_b$ 를 구성할 수 있다. 동일한 방법으로 이웃 데이터 $\mathbf{x}_c$ 에 대한 이웃 배치 $\mathbf{f}_c$ 도 함께 생성할 수 있다. 실시예에서는 이웃 데이터에 대한 특징 추출 과정을 생략하기 때문에 학습 속도에 대한 이득을 가질 수 있다. 상기와 같이 구성된 특징 배치(feature batch)들에 대해 수학적 1에 의해 손실(loss)가 계산될 수 있다.
- [0031] 수학적 1:
- [0032]
$$l_{ml}(\mathbf{x}, \mathbf{y}) = \alpha \cdot l_{mse}(\hat{\mathbf{y}}, \mathbf{y}) + (1 - \alpha) \cdot l_{lr}(\mathbf{x}_a, \mathbf{x}_b, \mathbf{x}_c)$$
- [0033] 첫번째 항의 $l_{mse}(\hat{\mathbf{y}}, \mathbf{y})$ 은 입력 데이터에 대한 출력과 라벨 데이터의 MSE(mean of squared errors)에 기반한 지도 학습 손실(supervised learning loss)이다. 두번째 항은 비특허문헌 1 "Sungyeon Kim, Minkyoo Seo, Ivan Laptev, Minsu Cho, and Suha Kwak. 2019. Deep Metric Learning Beyond Binary Supervision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2288-2297"의 로그 비율 손실(log-ratio loss)에 기반한다. α 는 MSE loss와 log-ratio loss의 비율을 조절하는 파라미터이다. 예를 들면, 0.5로 지정하여 동일한 비율로 학습할 수 있다. 감정 인식 시스템은 메트릭 학습을 완료한 CNN-LSTM 네트워크를 통해 분류된 데이터와 비분류된 데이터의 임베디드된 특징들을 추출할 수 있다(step 2).
- [0034] 감정 인식 시스템은 의사 라벨을 추정하기 위한 지도 학습 과정을 수행할 수 있다. step 3에서, 감정 인식 시스템은 비분류된 데이터의 의사 라벨을 추정할 수 있다. 종래의 의사 라벨 추정 방법들 중 비특허문헌 2 "Ahmet Iscen, Giorgos Tolias, Yannis Avrithis, and Ondrej Chum. 2019. Label propagation for deep semi-supervised learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 5070-5079."에 state-of the-art 기법이 제시된 바 있다. 이러한 state-of the-art 기법은 one-hot 벡터의 형태인 분류된 데이터의 라벨 정보가 비분류된 데이터의 제로 벡터(zero vector)에 전파(propagation)되는 방식이다. 그러나, 회귀 작업(regression task)에서는 라벨 정보를 스코어 벡터(score vector) 형태로 정의할 수 없기 때문에 비특허문헌 2의 의사 라벨 추정 방식을 이용할 수 없다. 실시예에서는 임베디드된 특징 공간에서 분류된 데이터와 비분류된 데이터의 특징 거리를 이용하여 회귀 작업(regression task)에서도 의사 라벨을 추정할

수 있다.

[0035] 감정 인식 시스템은 임베디드된 특징 공간에서 의사 라벨을 추정할 수 있다. 임베디드된 특징 공간에서 거리가 가까운 특징들은 유사한 라벨 정보를 갖는 것이 일반적으로 바람직하다고 가정하기로 한다. 실시예에서 제안된 반지도 학습의 step 2를 통해 모든 데이터들이 가정에 맞게 임베디드 되었다고 볼 수 있다. 예를 들면, 비분류된 데이터의 실제 라벨은 특징 공간에서 가장 근접한 분류된 데이터의 라벨과 유사할 것이다. 이에 따라 비분류된 데이터와 근접한 분류된 데이터의 라벨을 이용하여 비분류된 데이터의 의사 라벨로 활용할 수 있다. 감정 인식 시스템은 step 2에서 획득된 분류된 데이터와 비분류된 데이터의 임베디드된 특징을 $f_{x_{L,i}}$, $f_{x_{U,j}}$ 라고 하자. 여기서, i와 j는 분류된 데이터와 비분류된 데이터의 인덱스를 의미한다.

[0036] 도 3을 참고하면, 감정 인식 시스템은 비분류된 데이터의 임베디드된 특징 $f_{x_{U,j}}$ 과 모든 가능한 분류된 데이터의 임베디드된 특징 $f_{x_{L,i}}$ 과의 유클리드 거리를 계산할 수 있다. 도 3에 도시된 바와 같이, 원의 색깔은 라벨 정보를 의미한다. 감정 인식 시스템은 비분류된 데이터의 임베디드된 특징 $f_{x_{U,j}}$ 과 모든 가능한 분류된 데이터의 임베디드된 특징 $f_{x_{L,i}}$ 과의 유클리드 거리를 계산함에 따라 거리가 가까운 K(K는 자연수)개의 가장 가까운(nearest) 이웃 $f_{x_{NN,k}}$ 을 탐색할 수 있다. $f_{x_{NN,k}}$ 는 k(k는 자연수)번째 NN을 의미한다. 감정 인식 시스템은 비분류된 데이터의 임베디드된 특징 $f_{x_{U,j}}$ 과 가장 가까운 이웃 $f_{x_{NN,k}}$ 간 거리에 기반한 거리 가중치 $w_{j,k}$ 를 계산할 수 있다.

[0037] 수학적식 2:

$$w_{j,k} = e^{-\sigma D(f_{x_{U,j}}, f_{x_{NN,k}})} / \sum_{k=1}^K e^{-\sigma D(f_{x_{U,j}}, f_{x_{NN,k}})}$$

[0038] 여기서, D는 유클리드 거리를 의미한다. 예를 들면, $\sigma=1$, $K=5$ 로 설정될 수 있다. 최종적으로 가장 가까운 이웃들에 대응하는 라벨 정보 $y_{NN,k}$ 들의 가중합(weighted sum)을 통해 $f_{x_{U,j}}$ 의 의사 라벨 $y'_{U,j}$ 가 수학적식 3과 같이 추정될 수 있다.

[0040] 수학적식 3:

$$y'_{U,j} = \sum_{k=1}^K w_{j,k} \cdot y_{NN,k}$$

[0041] 감정 인식 시스템은 의사 라벨과 함께 네트워크를 학습시킬 수 있다. 이때, 의사 라벨 추정을 통하여 모든 비분류된 데이터의 라벨 정보가 확보되었다고 가정할 수 있다. 감정 인식 시스템은 분류된 데이터뿐만 아니라 의사 라벨을 갖는 비분류된 데이터를 이용하여 일반적인 지도 학습처럼 네트워크를 학습할 수 있다(step 4). 그러나, 상대적으로 소량의 분류된 데이터만으로 학습된 네트워크를 통해 의사 라벨들이 추정되었기 때문에 의사 라벨의 정확도는 떨어질 수 있다.

[0042] 감정 인식 시스템은 의사 라벨의 불확실성(uncertainty)을 측정하여 비분류된 데이터에 대한 트레이닝 손실(training loss)의 부분(portion)을 조절할 수 있다. 일반적으로 비분류된 데이터 특징과 가장 가까운 분류된 데이터의 특징이 가까울수록 의사 라벨의 정확도가 높다. 이러한 현상에 근거하여 가장 짧은 거리가 증가할수록 트레이닝 손실에서 비분류된 데이터의 부분을 감소시킬 수 있다. 구체적으로, 가장 짧은 거리에 근거하여 수학적식 4, 수학적식 5와 같이 의사 라벨에 대한 신뢰도 ω_j 를 계산할 수 있다. 예를 들면, $\lambda=15$ 로 설정될 수 있다.

$$\omega_j = e^{-\lambda \cdot d_{min,j}}$$

[0044] 수학적식 4:

$$d_{\min,j} = \min_{k \in \mathbb{K}} D(f_{x_{U,j}}, f_{x_{L,k}}), \mathbb{K} = \{1, 2, \dots, K\}$$

수학식 5:

최종적으로, 분류된 데이터와 비분류된 데이터를 이용하여 네트워크를 학습시키는 트레이닝 손실은 수학식 6과 같다.

수학식 6:

$$l_{pl}(X_L, X_U, Y_L, Y'_U) = \sum_{i=1}^l l_{mse}(\hat{y}, y_{L,i}) + \sum_{j=l+1}^m \omega_j \cdot l_{mse}(\hat{y}, y'_{U,j})$$

X_L 와 X_U 는 분류된 데이터셋과 비분류된 데이터셋을 의미하고, Y_L 와 Y'_U 는 각 데이터셋의 라벨 셋(label set)과 의사 라벨 셋(pseudo label set)이다. 수학식 6에서 첫번째 항은 분류된 데이터에 대한 MSE 손실이고, 두번째 항은 비분류된 데이터에 대한 MSE 손실이다. ω_j 를 통하여 의사 라벨의 불확실성에 따라 트레이닝 손실에 반영되는 가중치가 조절될 수 있다.

감정 인식 시스템은 비분류된 데이터와 분류된 데이터를 함께 이용하여 네트워크를 학습하면, 초기의 프리-트레이닝 모델보다 학습 성능을 개선시킬 수 있다. 감정 인식 시스템은 이러한 과정을 반복적으로 수행하여 최적화된 의사 라벨과 네트워크의 파라미터를 획득할 수 있다. 다시 말해서, 도 2와 같이 step 2 내지 step 4까지의 과정들이 반복적으로 수행될 수 있다.

도 4는 일 실시예에 따른 감정 인식 시스템에서 반지도 학습으로 학습한 학습 모델의 결과를 설명하기 위한 도면이다.

도 4는 시간에 따른 valence 추정 결과의 예를 나타낸 것이다. SL100%와 관련된 선은 전체 데이터셋 중 모든 라벨 정보를 활용하여 지도 학습을 수행한 결과이고, SL10%와 관련된 선은 전체 데이터셋 중 10%의 분류된 데이터를 이용하여 지도 학습을 수행한 결과이고, SSL10%/90%와 관련된 데이터는 전체 학습 데이터셋 중 10%의 분류된 데이터와 90%의 비분류된 데이터를 이용하여 제안하는 반지도 학습으로 학습한 모델의 결과이다. 이를 통하여 반지도 학습의 결과가 100% 지도 학습의 결과와 유사해짐을 확인할 수 있다.

표 1은 MAHNOB-HCI dataset에서 labeled data의 비율에 따른 지도학습(SL, supervised learning) 및 반지도 학습(SSL, semi-supervised learning)의 RMSE(root mean square error) 비교, ML(metric learning)의 유무에 따라 성능을 비교한 것이다.

표 1:

Method	Portion of labeled data				
	100%	40%	30%	20%	10%
SL only	0.0448	0.0465	0.0466	0.0480	0.0498
SL w/ ML	-	0.0464	0.0466	0.0475	0.0487
SSL w/o ML	-	0.0457	0.0454	0.0468	0.0475
SSL w/ ML	-	0.0456	0.0455	0.0460	0.0468

도 5는 일 실시예에 따른 감정 인식 시스템에서 메트릭 학습을 적용하여 특징을 분석하는 것을 설명하기 위한 예이다.

도 5(a), 5(b)는 메트릭 학습을 적용하지 않은 상태에서의 특징 공간을 분석한 것으로, 도 5(a)에서 각 점의 색은 감정의 긍정 또는 부정인 valence 값을 의미하고, 도 5(b)에서 각 점의 색은 얼굴 영상에 대한 ID 정보를 의미한다.

또한, 도 5(c), 5(d)는 메트릭 학습을 적용하였을 때의 특징 공간을 분석한 것으로, 도 5(c)에서 각 점의 색은 감정의 긍정 또는 부정인 valence 값을 의미하고, 도 5(d)에서 각 점의 색은 얼굴 영상에 대한 ID 정보를 의미한다.

- [0059] 도 6은 일 실시예에 따른 감정 인식 시스템의 구성을 설명하기 위한 블록도이고, 도 7은 일 실시예에 따른 감정 인식 시스템에서 감정을 인식하는 방법을 설명하기 위한 흐름도이다.
- [0060] 감정 인식 시스템(600)에 포함된 프로세서는 검출부(610), 수행부(620) 및 추정부(630)를 포함할 수 있다. 이러한 프로세서 및 프로세서의 구성요소들은 도 7의 감정 인식 방법이 포함하는 단계들(710 내지 730)을 수행하도록 감정 인식 시스템을 제어할 수 있다. 이때, 프로세서 및 프로세서의 구성요소들은 메모리가 포함하는 운영체제의 코드와 적어도 하나의 프로그램의 코드에 따른 명령(instruction)을 실행하도록 구현될 수 있다. 여기서, 프로세서의 구성요소들은 감정 인식 시스템에 저장된 프로그램 코드가 제공하는 제어 명령에 따라 프로세서에 의해 수행되는 서로 다른 기능들(different functions)의 표현들일 수 있다.
- [0061] 프로세서는 방법을 위한 프로그램의 파일에 저장된 프로그램 코드를 메모리에 로딩할 수 있다. 예를 들면, 감정 인식 시스템(600)에서 프로그램이 실행되면, 프로세서는 운영체제의 제어에 따라 프로그램의 파일로부터 프로그램 코드를 메모리에 로딩하도록 감정 인식 시스템을 제어할 수 있다.
- [0062] 단계(710)에서 검출부(610)는 동영상 데이터가 입력됨을 수신함에 따라 동영상 데이터의 각 프레임으로부터 얼굴 영역을 검출할 수 있다.
- [0063] 단계(720)에서 수행부(620)는 검출된 얼굴 영역에 대하여 두 개 이상의 학습 네트워크가 결합된 학습 모델을 이용하여 의사 라벨 기반의 반지도 학습을 수행할 수 있다. 이때, 학습 모델은 CNN과 LSTM이 결합된 학습 모델로서, 지도 학습 기반의 CNN 학습과 반지도 학습 기반의 LSTM 학습을 각각 독립적으로 수행할 수 있다. 수행부(620)는 지도 학습 기반의 CNN 학습을 통하여 상기 검출된 얼굴 영역에 대한 특징을 추출하고, 추출된 특징을 반지도 학습 기반의 LSTM에 입력할 수 있다. 수행부(620)는 반지도 학습 기반의 LSTM 학습 과정에서, 트레이닝 데이터셋(dataset) 중 분류된 데이터를 이용하여 메트릭 학습(metric learning)을 통해 LSTM 네트워크를 프리 트레이닝(pre-training)하고, 프리 트레이닝된 LSTM 네트워크를 이용하여 분류된 데이터와 비분류된 데이터에 대한 임베디드된 특징을 추출하고, 추출된 임베디드된 특징의 공간에서 특징간 거리를 이용하여 비분류된 데이터의 의사 라벨을 추정하고, 추정된 의사 라벨과 함께 분류된 데이터와 비분류된 데이터를 학습시킬 수 있다. 수행부(620)는 메트릭 학습에서, 연속성있는 라벨의 데이터셋의 분류된 데이터를 대상으로 앵커 데이터와 이웃 데이터를 배치(batch) 단위로 구성하여 학습을 수행하고, 배치 단위로 구성된 앵커 데이터를 LSTM 네트워크로부터 임베디드된 특징을 출력하고, 출력된 임베디드된 특징들로부터 구성된 앵커 특징 배치를 랜덤하게 서플링하여 이웃 배치를 구성하고, 특징 배치들에 대해 손실(loss)를 계산하는 과정을 수행할 수 있다. 수행부(620)는 비분류된 데이터와 근접한 분류된 데이터의 라벨을 이용하여 비분류된 데이터의 의사 라벨로 활용할 수 있다. 수행부(620)는 비분류된 라벨을 포함하는 비분류된 데이터를 이용하여 네트워크 학습을 수행하며, 추정된 의사 라벨의 불확실성을 측정하여 비분류된 데이터에 대한 트레이닝 손실의 부분을 조정할 수 있다.
- [0064] 단계(730)에서 추정부(630)는 수행된 반지도 학습을 통하여 획득된 학습 결과로서, 검출된 얼굴 영역에 대한 감정 정보를 추정할 수 있다.
- [0065] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.
- [0066] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치에 구체화(embody)될 수 있다. 소프트웨어는 네트워크로

연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

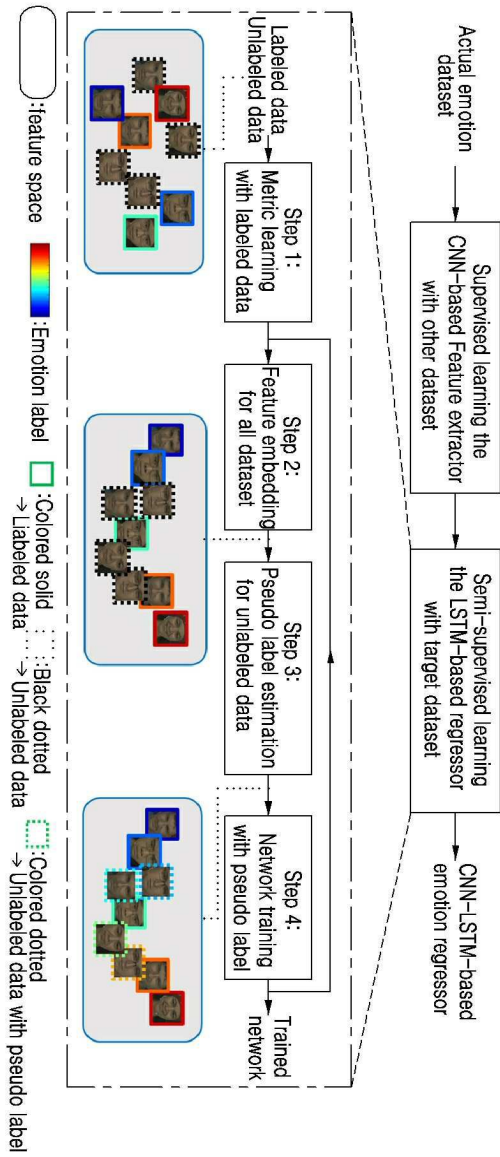
[0067] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic 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.

[0068] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 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.

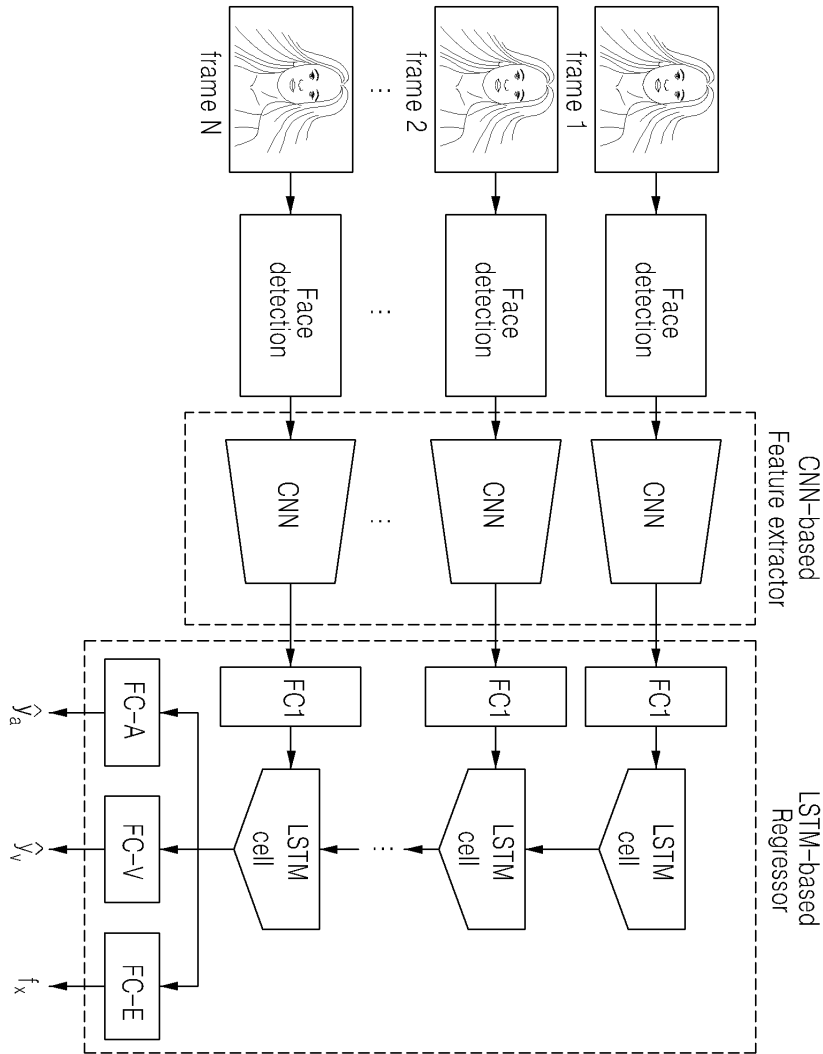
[0069] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

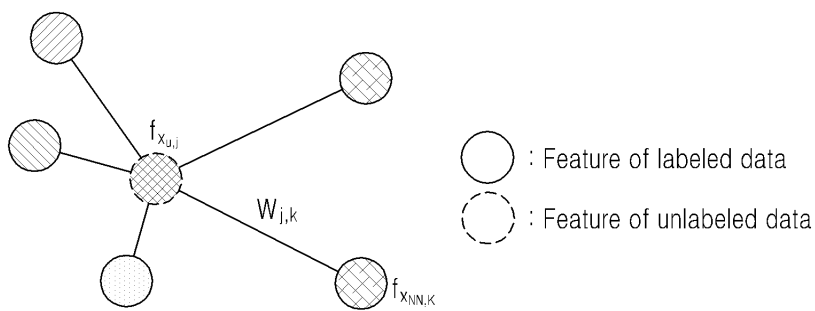
도면1



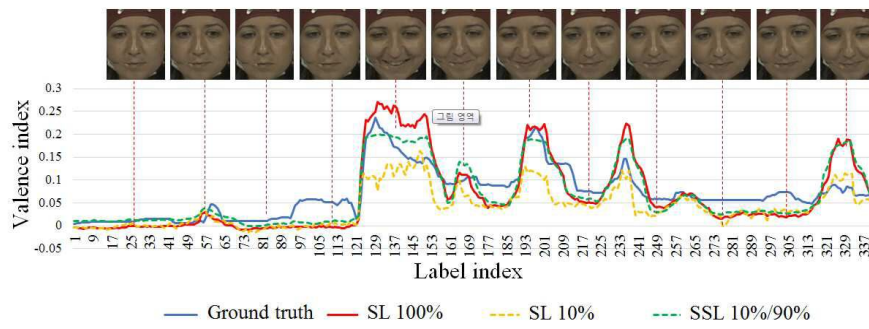
도면2



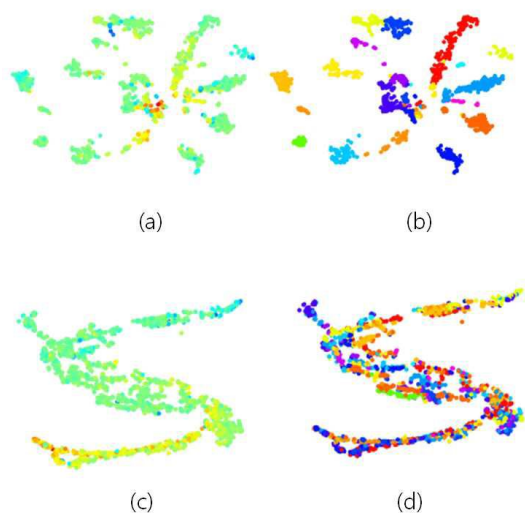
도면3



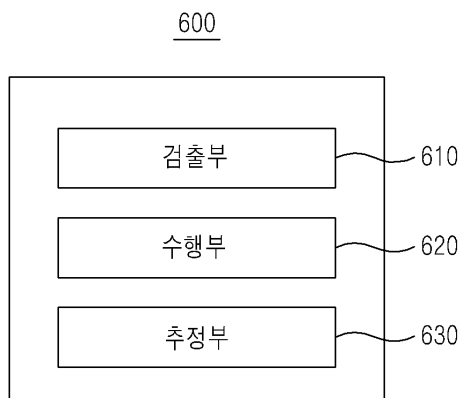
도면4



도면5



도면6



도면7

