



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0066694
(43) 공개일자 2021년06월07일

(51) 국제특허분류(Int. Cl.)
G06K 9/00 (2006.01) G06K 9/46 (2006.01)
G06K 9/62 (2006.01)
(52) CPC특허분류
G06K 9/00335 (2013.01)
G06K 9/46 (2013.01)
(21) 출원번호 10-2020-0096684
(22) 출원일자 2020년08월03일
심사청구일자 2020년08월03일
(30) 우선권주장
1020190155307 2019년11월28일 대한민국(KR)

(71) 출원인
경희대학교 산학협력단
경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스내)
(72) 발명자
이영구
경기도 수원시 영통구 청명북로 33, 431동 1704호(영통동, 청명마을삼성아파트)
우딘 엠디 아제르
경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스)
(74) 대리인
특허법인도담

전체 청구항 수 : 총 10 항

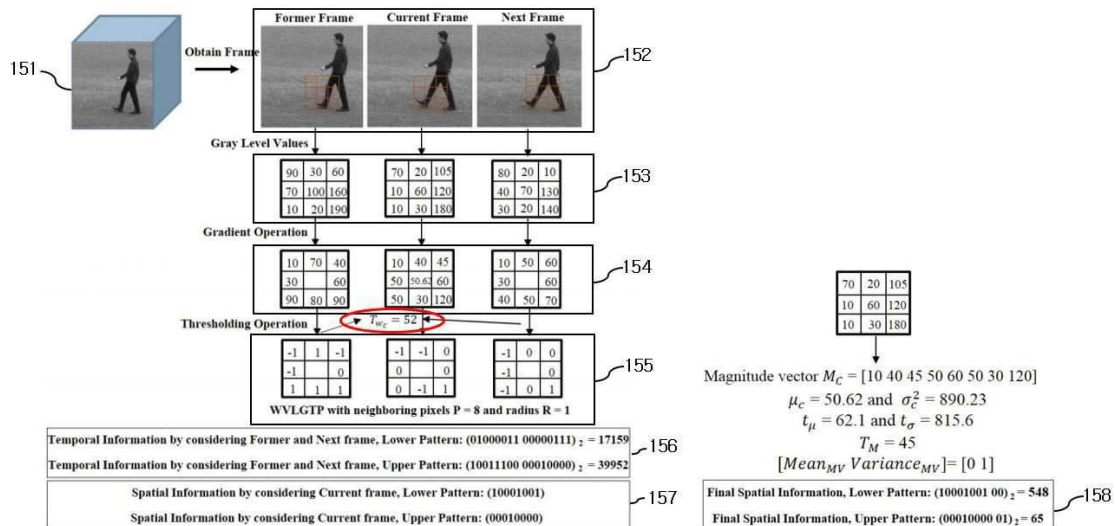
(54) 발명의 명칭 비디오 데이터의 시공간 정보를 활용한 인간 행동 인식 장치

(57) 요약

본 발명의 실시 예에 따른 인간 행동 인식 장치는 인간 행동이 포함된 비디오 데이터를 저장하는 데이터 저장부, 비디오 데이터에서 공간 특징을 추출하여 공간 정보를 생성하는 공간 정보 생성부, 비디오 데이터에서 시간 정보 및 공간 정보를 활용하여 동적 정보를 생성하는 동적 정보 생성부를 포함하며, 동적 정보 생성부는 비디오 데이터

(뒷면에 계속)

대표도 - 도3



터에서 3개의 3x3 크기의 연속된 Former Frame, Current Frame, Next Frame을 추출하고, 연속된 프레임 각각에 대한 픽셀값을 제 1픽셀값으로 하고, 제 1픽셀값에 대해 그라디언트 연산과 임계값 연산을 통해 WVLGTP(Weber's law Volume Local Gradient Ternary Pattern) 생성하며, 생성된 WVLGTP를 통해 시간 정보 및 공간 정보에 대한 특징 벡터를 생성할 수 있고, 동적 정보는 시간 정보에 대한 특징 벡터 및 공간 정보에 대한 특징 벡터인 것을 특징으로 하며 공간 정보와 동적 정보를 패턴 인식을 위한 지도학습 모델인 SVM(Support Vector Machine)알고리즘을 통해 인간 행동을 인식하는 인식부를 더 포함한다.

(52) CPC특허분류

G06K 9/6267 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711102869
과제번호	2016-0-00406-005
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	SW컴퓨팅산업원천기술개발(R&D)
연구과제명	(기반SW-창조씨앗2단계) SIAT형 CCTV 클라우드 플랫폼 기술 개발
기 여 율	1/1
과제수행기관명	숙명여자대학교 산학협력단
연구기간	2020.01.01 ~ 2020.12.31

명세서

청구범위

청구항 1

인간 행동이 포함된 비디오 데이터를 저장하는 데이터 저장부;

비디오 데이터에서 공간 특징을 추출하여 공간 정보를 생성하는 공간 정보 생성부;

비디오 데이터에서 시간 정보 및 공간 정보를 활용하여 동적 정보를 생성하는 동적 정보 생성부;

상기 동적 정보 생성부는 비디오 데이터에서 3개의 3x3 크기의 연속된 Former Frame, Current Frame, Next Frame을 추출하고, 연속된 프레임 각 각에 대한 픽셀값을 제 1픽셀값으로 하고, 제 1픽셀값에 대해 그라디언트 연산과 임계값 연산을 통해 WVLGTP(Weber's law Volume Local Gradient Ternary Pattern) 생성하며, 생성된 WVLGTP를 통해 시간 정보 및 공간 정보에 대한 특징 벡터를 생성하는 것을 포함하며

상기 동적 정보는 시간 정보에 대한 특징 벡터 및 공간 정보에 대한 특징 벡터인 것을 특징으로 하며

공간 정보와 동적 정보를 패턴 인식을 위한 지도학습 모델인 SVM(Support Vector Machine)알고리즘을 통해 인간 행동을 인식하는 인식부;

를 포함하는 인간 행동 인식 장치.

청구항 2

제 1항에 있어서,

상기 그라디언트 연산은

상기 제 1픽셀값의 중심 픽셀값을 기준으로 인접 픽셀값과의 차이의 절대값을 인접 픽셀의 제 2픽셀값으로 하고, 제 1픽셀값의 중심 픽셀값은 그대로 중심 픽셀의 제 2픽셀값으로 하는 인간 행동 인식 장치.

청구항 3

제 2항에 있어서,

상기 중심 픽셀의 제 2픽셀값은 인접 픽셀의 제 2픽셀값의 평균값으로 하는 인간 행동 인식 장치.

청구항 4

제 2항에 있어서,

상기 임계값 연산은

상기 제 2픽셀값에 대해, 웨버의 법칙에 기초한 임계값과의 차이의 절대값을 계산하여, 절대값과 임의의 값을 비교하여 1, -1, 0의 값을 부여한 결과를 제 3픽셀값으로 하는 인간 행동 인식 장치.

청구항 5

제 1항에 있어서,

상기 시간 정보 및 공간 정보에 대한 특징 벡터는

3개의 연속된 프레임에 대해, Current Frame의 제 3픽셀값을 기준으로 공간 정보에 대한 특징 벡터를 생성하고,

Former Frame 과 Next Frame을 기준으로 시간 정보에 대한 특징 벡터를 생성하는 인간 행동 인식 장치.

청구항 6

제 5항에 있어서,

상기 공간 정보에 대한 특징 벡터는

Current Frame의 크기 벡터를 생성하고, 생성된 크기 벡터의 평균과 분산을 활용하여 생성되는 인간 행동 인식 장치.

청구항 7

제 1항에 있어서,

상기 동적 정보는

멀티스케일 WVLGTP를 통해 생성된 시간 정보 및 공간 정보에 대한 특징 벡터이며,

상기 멀티스케일 WVLGTP는

상기 연속된 프레임에 대해 5x5 크기 픽셀값을 추출하고, 평균화 방식 연산과 그라디언트 연산과 임계값 연산을 통해 생성된 인간 행동 인식 장치.

청구항 8

제 7항에 있어서,

상기 평균화 방식 연산은

임의의 픽셀에 대하여, 상기 픽셀의 픽셀값은 상기 픽셀의 근접한 복수개의 픽셀값의 평균을 계산하여 그 평균 값을 상기 픽셀의 픽셀값으로 하는 연산 방식인 인간 행동 인식 장치.

청구항 9

제 7항에 있어서,

상기 동적 정보는

멀티스케일 WVLGTP를 통해 생성된 시간 정보 및 공간 정보에 대한 특징 벡터와 WVLGTP를 통해 생성된 시간 정보 및 공간 정보에 대한 특징 벡터를 하나의 시간 정보 및 공간 정보에 대한 특징 벡터로 하는 인간 행동 인식 장치.

청구항 10

제 6항에 있어서,

상기 크기 벡터는

Current Frame의 제 1 픽셀값에 대해, 중앙 픽셀값을 기준으로 인접 픽셀값과 중앙 픽셀값의 차이의 절대값인 것을 특징으로 하는 인간 행동 인식 장치.

발명의 설명

기술 분야

[0001] 본 발명은 비디오 데이터를 분석하여 인간의 행동을 인식하는 장치에 관한 것으로, 구체적으로 비디오 데이터의 배경 공간 정보 및 비디오 데이터의 피사체 동적 정보를 고려하여 인간 행동을 인식하는 장치에 관한 것이다.

배경 기술

[0002] 인간 행동 인식은 비디오 감시, 스포츠 비디오 분석, 영화 검색 등의 다양한 분야로 발전 가능한 분야이다. 인간 행동 인식은 이미지 또는 비디오 데이터에서 인간의 행동이 무엇인지 인식하는 것을 말한다. 인간 행동 인식에서 가장 중요한 기술은 비디오 데이터에서 인간의 복잡한 여러 행동을 구분하고 분류할 수 있는 기술이다. 나아가, 인간의 의복, 모양, 개인 차이에 따른 상이한 인식 결과가 되지 않고 동일한 인식 결과가 도출될 수 있는 기술이 중요하다. 이와 관련하여 인간 행동과 관련된 광범위한 연구가 진행되고 있다.

[0003] 그러나, 이러한 연구에도 불구하고, 여전히 인간 행동 인식과 관련한 많은 문제점이 존재한다. 예를 들어, 비디오 프레임의 길이가 길어짐에 따라 행동 인식의 정확성이 낮아지거나, 복잡한 배경, 피사체와 공간 간의 모호한 경계로 인한 인간 행동 인식의 불명확성 등에 문제가 있다. 또한, 조명 변화 등으로 인해 상이한 인식 결과의 도출도 문제점 중 하나였다.

[0004] 이러한 문제점을 해결하고 인간 행동 인식의 정확도를 향상시키기 위해, 비디오 프레임의 배경 공간 정보와 비디오 데이터에서의 피사체의 동적인 특징을 추출한 동적 정보를 함께 고려하는 방법이 대두되고 있다.

발명의 내용

해결하려는 과제

[0005] 본 발명은 전술한 문제점을 해결하기 위한 것으로서, 비디오 데이터 배경의 공간 특징을 추출하여 생성된 공간 정보를 통해 비디오 데이터의 인간 행동을 인식하는 장치를 제공하는 것을 목적으로 한다.

[0006] 본 발명은 비디오 데이터에서 웨버의 법칙을 활용하여 시간 및 공간 특징을 추출하는 인간 행동 인식 장치를 제공하는 것을 목적으로 한다.

[0007] 본 발명은 비디오 데이터에서 피사체의 동적 정보를 분석하기 위한 WVLGTP 라는 새로운 개념을 도입하여 종래 VLBP 방법을 통한 인간 행동 인식의 정확성을 보완하는 것을 목적으로 한다.

[0008] 본 발명은 비디오 데이터에서 인간 행동을 인식할 때 필요한 동적 정보 생성시 시간 특징과 공간 정보를 구분하여 특징 벡터를 생성하여 인간 행동 인식의 정확성 향상시키는 것을 목적으로 한다.

과제의 해결 수단

[0009] 본 발명의 실시 예에 따른 인간 행동 인식 장치는 인간 행동이 포함된 비디오 데이터를 저장하는 데이터 저장부, 비디오 데이터에서 공간 특징을 추출하여 공간 정보를 생성하는 공간 정보 생성부, 비디오 데이터에서 시간 정보 및 공간 정보를 활용하여 동적 정보를 생성하는 동적 정보 생성부를 포함하며, 동적 정보 생성부는 비디오 데이터에서 3개의 3x3 크기의 연속된 Former Frame, Current Frame, Next Frame을 추출하고, 연속된 프레임 각 각에 대한 픽셀값을 제 1픽셀값으로 하고, 제 1픽셀값에 대해 그라디언트 연산과 임계값 연산을 통해 WVLGTP(Weber's law Volume Local Gradient Ternary Pattern) 생성하며, 생성된 WVLGTP를 통해 시간 정보 및 공간 정보에 대한 특징 벡터를 생성할 수 있고, 동적 정보는 시간 정보에 대한 특징 벡터 및 공간 정보에 대한 특징 벡터인 것을 특징으로 하며 공간 정보와 동적 정보를 패턴 인식을 위한 지도학습 모델인 SVM(Support Vector Machine)알고리즘을 통해 인간 행동을 인식하는 인식부를 더 포함한다.

발명의 효과

[0010] 본 발명의 실시 예를 따르면, 비디오 데이터에서의 인간 행동 인식 시 인간 행동 동작과 배경의 공간 정보를 명확히 구분하여, 인간 행동 인식의 평균 정확성을 향상시킬 수 있다.

[0011] 본 발명의 실시 예를 따르면, 비디오 데이터에서 인간 행동 인식 시 피사체의 시간 특징 및 공간 특징을 구분하여 별개의 특징 벡터를 생성하는 바, 행동 인식 정확성이 향상된다.

[0012] 본 발명의 실시 예를 따르면, 웨버의 법칙을 기반으로 생성된 WVLGTP에 의해 중앙 픽셀값에 관한 잡음 처리를

할 수 있다.

도면의 간단한 설명

- [0013] 도 1은 본 발명의 실시 예에 따른 인간 행동 인식 장치의 블록도를 나타낸 도면이다.
- 도 2는 본 발명의 실시 예에 따른 공간 정보 생성부에서 공간 정보를 생성하는 방법을 설명하기 위한 도면이다.
- 도 3은 본 발명의 실시 예에 따른 동적 정보 생성부에서 동적 정보를 생성하는 방법을 설명하기 위한 도면이다.
- 도 4는 본 발명의 또 다른 일 실시 예에 따른 멀티 스케일 WVLGTP를 생성하는 방법을 설명하기 위한 도면이다.
- 도 5는 본 발명의 실시 예에 따른 멀티 스케일 WVLGTP를 생성할 시 평균화 방식 연산을 설명하기 위한 도면이다.
- 도 6은 본 발명의 실시 예에 따른 인간 행동 인식 장치에 대한 성능을 확인하기 위한 실험 대상 데이터에 관한 도면이다.
- 도 7은 본 발명의 실시 예에 따른 KTH 데이터 셋에 대한 인간 행동 인식 성능 확인 실험 결과값을 나타낸 도면이다.
- 도 8은 본 발명의 실시 예에 따른 UCF 스포츠 액션 데이터 셋에 대한 인간 행동 인식 성능 확인 실험 결과값을 나타낸 도면이다.
- 도 9는 본 발명의 실시 예에 따른 UT-Interaction 데이터 셋에 대한 인간 행동 인식 성능 확인 실험 결과값을 나타낸 도면이다.
- 도 10은 본 발명의 실시 예에 따른 Hollywood2 데이터 셋에 대한 인간 행동 인식 성능 확인 실험 결과값을 나타낸 도면이다.
- 도 11은 본 발명의 실시 예에 따른 UCF-101 데이터 셋에 대한 인간 행동 인식 성능 확인 실험 결과값을 나타낸 도면이다.
- 도 12는 본 발명의 실시 예에 따른 멀티스케일 WVLGTP에 행동 인식 성능 확인 실험 결과값을 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0014] 전술한 목적, 특징 및 장점은 첨부된 도면을 참조하여 상세하게 후술되며, 이에 따라 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 본 발명의 기술적 사상을 용이하게 실시할 수 있을 것이다. 본 발명을 설명함에 있어서 본 발명과 관련된 공지 기술에 대한 구체적인 설명이 본 발명의 요지를 불필요 하게 흐릴 수 있다고 판단되는 경우에는 상세한 설명을 생략한다.
- [0015] 도면에서 동일한 참조부호는 동일 또는 유사한 구성요소를 가리키는 것으로 사용되며, 명세서 및 특허청구의 범위 기재된 모든 조합은 임의의 방식으로 조합될 수 있다. 그리고 다른 식으로 규정하지 않는 한, 단수에 대한 언급은 하나 이상을 포함할 수 있고, 단수 표현에 대한 언급은 또한 복수 표현을 포함할 수 있음이 이해되어야 한다.
- [0016] 본 명세서에서 사용되는 용어는 단지 특정 예시적 실시 예들을 설명할 목적을 가지고 있으며 한정할 의도로 사용되는 것이 아니다. 본 명세서에서 사용된 바와 같은 단수적 표현들은 또한, 해당 문장에서 명확하게 달리 표시하지 않는 한, 복수의 의미를 포함하도록 의도될 수 있다. 용어 "및/또는," "그리고/또는"은 그 관련되어 나열되는 항목들의 모든 조합들 및 어느 하나를 포함한다. 용어 "포함한다", "포함하는", "포함하고 있는", "구비하는", "갖는", "가지고 있는" 등은 내포적 의미를 갖는 바, 이에 따라 이러한 용어들은 그 기재된 특징, 정수, 단계, 동작, 요소, 및/또는 컴포넌트를 특정하며, 하나 이상의 다른 특징, 정수, 단계, 동작, 요소, 컴포넌트, 및/또는 이들의 그룹의 존재 혹은 추가를 배제하지 않는다. 본 명세서에서 설명되는 방법의 단계들, 프로세스들, 동작들은, 구체적으로 그 수행 순서가 확정되는 경우가 아니라면, 이들의 수행을 논의된 혹은 예시된 그러한 특정 순서로 반드시 해야 하는 것으로 해석해서는 안 된다. 추가적인 혹은 대안적인 단계들이 사용될 수 있음을 또한 이해해야 한다.
- [0017] 또한, 각각의 구성요소는 각각 하드웨어 프로세서로 구현될 수 있고, 위 구성요소들이 통합되어 하나의 하드웨어 프로세서로 구현될 수 있으며, 또는 위 구성요소들이 서로 조합되어 복수 개의 하드웨어 프로세서로 구현될

수도 있다.

- [0018] 이하, 첨부된 도면을 참조하여 본 발명에 따른 바람직한 실시 예를 상세히 설명하기로 한다.
- [0019] 도 1은 본 발명의 실시 예에 따른 인간 행동 인식 장치의 블록도를 나타낸 도면이다.
- [0020] 인간 행동 인식 장치(100)은 인간 행동이 포함된 비디오 데이터가 저장된 데이터 저장부(110), 데이터 저장부(110)의 저장된 비디오 데이터에서 공간 특징을 추출하여 공간 정보를 생성하는 공간 정보 생성부(130), 비디오 데이터에서 연속된 복수의 프레임을 추출하고, 픽셀값의 기울기 연산을 통하여 동적 정보를 생성하는 동적 정보 생성부(150), 공간 정보 생성부(130)와 생성된 공간 정보와 동적 정보 생성부(150)에서 생성된 동적 정보를 이용하여 SVM(Support Vector Machine) 알고리즘을 통해 인간 행동을 인식하는 인식부(170)을 포함할 수 있다.
- [0021] 데이터 저장부(110)의 저장되는 비디오 데이터는 인간 행동을 포함하는 비디오 데이터일 수 있다. 비디오 데이터는 인간 한 명이 수행하는 행동을 포함한 데이터 또는 복수의 인간이 수행하는 행동을 포함하는 데이터 일 수 있다. 예를 들어, 비디오 데이터는 인간 행동 중 걷기, 달리기, 조깅, 복싱, 박수치기, 댄스, 다이빙, 골프 스윙, 리프팅, 스케이트보드 타기, 축구, 전자기기에 응답하기, 운전, 차에서 내리기, 야구공 잡기, 드럼, 홀라후프 돌리기, 요요 하기 등을 포함할 수 있다. 또한, 복수의 인간이 수행하는 행동은 악수하기, 포옹, 태권도, 가르키기, 주먹 지르기, 밀기, 포옹, 키스, 퍼레이드 등이 포함될 수 있다. 이는 예시에 해당될 뿐, 한정되어 해석되지 아니한다.
- [0022] 이하에서는, 공간 정보 생성부(130), 동적 정보 생성부(150), 인식부(170)에 대한 구체적 설명을 후술하도록 한다.
- [0023] 도 2는 본 발명의 실시 예에 따른 공간 정보 생성부에서 공간 정보를 생성하는 방법을 설명하기 위한 도면이다.
- [0024] 비디오 데이터의 배경 공간 정보는 인간 행동의 많은 동작과 활발히 연관되어 있는 경우가 많다. 인간 행동 인식시, 인간 행동의 동작 정보와 배경의 공간 정보가 명확히 구분될수록 인간 행동 인식 장치의 인식 정확성은 향상될 수 있다.
- [0025] 본 발명의 일 실시 예에 따른, 공간 정보 생성부는 비디오 데이터의 배경의 공간 특징을 추출하여 공간 정보를 생성할 수 있다. 공간 정보 생성부는 데이터 저장부의 저장된 비디오 데이터를 입력 데이터(131)로 사용할 수 있다. 입력 데이터(131)는 개별 RGB 프레임 크기가 299x299x3일 수 있고, 이에 한정되지 않고, 개별 RGB 프레임 크기는 여러 값을 가질 수 있다.
- [0026] 공간 정보 생성부는 비디오 데이터로부터 깊은 공간의 공간적 특징을 추출하기 위해서 Inception-Resnet-v2 네트워크 컨볼루션 모델을 활용할 수 있다. 도 2를 참조하면, Inception-Resnet-v2 네트워크 모델을 활용한 깊은 공간 특징 추출 아키텍처(132)를 확인할 수 있다.
- [0027] 공간 특징 추출 아키텍처는(132)은 Reduction 레이어(135)와 Inception Resnet 레이어(134)의 조합을 포함할 수 있다. 또한, 공간적 특징 추출 아키텍처(132)는 Stem 레이어(133)를, Inception Resnet 레이어(134) 앞에 더 포함할 수 있다. 보다 구체적으로, 공간 특징 추출 아키텍처(132)은 하나의 Stem 레이어(134)와 3개의 Inception Resnet 레이어(134)와 2개의 Reduction 레이어(135)와 순차적으로 하는 구조를 갖을 수 있다. 이러한 레이어 다음에는 평균 풀링 레이어(136)와 완전히 연결된 1000개의 채널이 있는 레이어(137)가 연결되어 공간 특징 추출 아키텍처(132)를 구성할 수 있다. Stem 레이어(133)는 공간 특징 추출 아키텍처(132)에 입력되기 전에 예비 컨볼루션 작업이 포함될 수 있다. Inception Resnet 레이어(134)는 컨볼루션 작업과 함께 Reduction 레이어(135)가 연결될 수 있다. Reduction 레이어(135)는 그리드의 높이와 너비를 조정할 수 있다. Stem 레이어(133)와 Inception Resnet 레이어(134)는 공간적 특징을 추출할 수 있고, 평균 풀링 레이어(136)는 개별 공간 특징 추출의 차원을 줄임과 동시에 조명과 변환 과정에 따라 추출된 공간 특징이 변경되는 것을 방지할 수 있다. 완전히 연결된 1000개의 채널이 있는 레이어(137)는 비선형 함수와 학습 시간을 줄이는 기능을 수행하는 정류 선형 단위 Rectified Linear Unit(ReLU)를 포함할 수 있다.
- [0028] 공간 정보 생성부는 비디오 데이터를 공간 특징 추출 아키텍처(132)에 입력하여, 입력 데이터에 대한 배경의 공간 특징을 추출할 수 있다. 추출된 공간 특징을 집계하여 비디오 데이터에 대한 공간 정보(138)를 생성할 수 있다. 공간 정보(138)은 인간 행동 인식 장치의 인식부에 입력 데이터로 활용될 수 있다. 이하에서는 동적 정보 생성부가 동적 정보를 생성하는 방법에 대한 구체적인 설명을 하도록 한다.
- [0029] 도 3은 본 발명의 실시 예에 따른 동적 정보 생성부에서 동적 정보를 생성하는 방법을 설명하기 위한 도면이다.

- [0030] 동적 정보 생성부는 시공간 특징을 모두 추출하여 동적 정보로 생성할 수 있으며, 이는 WVLGTP(Weber's law based Volume Local Gradient Ternary Pattern) 새로운 개념을 통해 생성될 수 있다.
- [0031] 종래에는 시공간 볼륨 내 픽셀값을 처리하기 위해 VLBP(Volume Local Binary Pattern) 방법을 사용하였다. VLBP(Volume Local Binary Pattern) 방법은 LBP(Local Binary Pattern)와 유사한 방법에 해당된다. 보다 구체적으로 VLBP 방법은 시공간 볼륨 내에서 인접 픽셀의 회색 값을 비교하여 1 또는 0의 이진법으로 값을 부여하는 것을 말한다. 즉, 중앙 픽셀값을 기준으로, 인접 픽셀의 값이 중앙 픽셀값보다 높으면 1를 부여하고, 낮으면 0을 부여하는 방법으로 VLBP를 계산한다.
- [0032] WVLGTP는 VLBP 방법을 보다 발전시킨 개념으로서, 본 발명에서의 동적 정보를 생성하기 위해 도입한 새로운 개념이다. 본 발명의 실시 예에 따르면, 픽셀값에 대한 여러 연산화 과정을 통해 WVLGTP를 활용한 동적 정보가 생성될 수 있다. WVLGTP를 생성하기 위한 여러 연산화 과정에 대한 상세한 설명은 도 3을 참조하여 설명하도록 한다.
- [0033] 도 3에서 동적 정보 생성부는 데이터 저장부의 저장된 비디오 데이터를 입력 데이터(151)로 할 수 있다. 동적 정보 생성부는 입력 데이터(151)에 대해서 연속된 3개의 비디오 프레임(152)을 추출할 수 있다. 연속된 3개의 비디오 프레임은 Former Frame, Current Frame, Next Frame으로 지칭할 수 있으며, 과거, 현재, 미래 프레임으로 지칭할 수 있다.
- [0034] 동적 정보 생성부는 3개의 연속된 프레임에 대한 3x3 크기의 제 1픽셀값(153)을 계산할 수 있다. 제 1픽셀값은 3개의 연속된 프레임의 회색 픽셀값을 말한다. 동적 정보 생성부는 제 1픽셀값에 대해서 그라디언트 연산을 수행할 수 있다. 그라디언트 연산을 수행한 결과값을 제 2픽셀값(154)으로 할 수 있다.
- [0035] 제 2픽셀값(154)을 추출하기 위한 그라디언트 연산은 중심 픽셀의 제 1 픽셀값(153)을 기준으로 인접 픽셀값과의 차이의 절대값을 계산하는 것을 말한다. 동적 정보 생성부는 3개의 연속된 프레임의 제 1픽셀값(153)에 대해 그라디언트 연산을 수행하고, 이에 대한 결과값으로 제 2픽셀값(154)을 생성할 수 있다.
- [0036] 예를 들어, Former Frame의 경우 중심 픽셀값이 100이고, Current Frame의 경우에는 60이며, Next Frame의 경우에는 70인 바, 100, 60, 70 제 1픽셀값을 기준으로 인접 픽셀값과의 차이값의 절대값을 계산한다.
- [0037] 여기서 Current Frame의 중심 픽셀의 제 2픽셀값(154)은 다른 방법으로 계산할 수 있다. 이는 Current Frame의 중심 픽셀의 경우 가장 많은 노이즈를 가지고 있어, 이를 조정할 필요가 있기 때문이다. 또한, 비디오 데이터에서 3개의 연속된 프레임을 추출하는 과정에서 3개의 연속된 프레임의 시간적 차이가 매우 짧기 때문에 Current Frame의 중심 픽셀의 노이즈 처리를 통해, 인간 행동 인식 정확도를 향상시킬 수 있다. 따라서 Current Frame의 중심 픽셀의 제 2 픽셀값(154)은 그라디언트 연산을 통해 산출된 인접 픽셀들의 제 2픽셀값(154)의 평균값으로 대체할 수 있다. 예를 들어, Current Frame의 인접 픽셀의 그라디언트 연산을 통해 추출된 제 2픽셀값은 10, 40, 45, 50, 60, 50, 30, 120 이고, 이 값의 평균값인 50.62를 Current Frame의 중심 픽셀의 제 2픽셀값(154)으로 할 수 있다. Current Frame의 중심 픽셀의 제 2픽셀값(154)은 Current Frame의 중심 픽셀의 제 1픽셀값(153)을 그대로 사용할 수도 있다.
- [0038] 동적 정보 생성부는 추가적으로 중심 픽셀을 효과적으로 처리하기 위한 임계값 연산 방법을 활용할 수 있다. 임계값 연산 방법은 Weber's Law, 즉 웨버의 법칙을 기반으로 할 수 있다. 이에 대한 구체적인 설명을 후술하도록 한다.
- [0039] 기본적으로, 웨버의 법칙은 다음과 같은 수학적 식 1을 통해서 설명할 수 있다.

수학적 식 1

$$\frac{\Delta I}{I} = K$$

[0040]

- [0041] 여기서 ΔI 는 차이의 현저함을 나타내는 증분의 임계 값을 나타내고, I는 초기 강도, K는 상수를 나타낸다. 즉 웨버의 법칙은 같은 종류의 두 자극에 대해서 두 자극을 구별할 수 있는 최소 차이는 초기 자극의 강도에 비례한다는 것에 관한 것이다. 보다 구체적으로, 초기 자극이 존재하고, 다음 자극을 초기 자극과 구별하기 위해서

는 초기 자극의 강도가 강할수록, 다음 자극의 강도 차이가 매우 커야 한다는 것이다.

[0042] 본 발명의 일 실시 예를 따르면 웨버의 법칙을 이용하여 T_{wc} 값을 계산할 수 있다. T_{wc} 값은 다음과 같은 수학적 2를 통해 설명할 수 있다.

수학적식 2

$$T_{wc} = \bar{g}_{C_c} + \sum_{n=0}^{l-1} \frac{|C_n - C_c|}{C_c}$$

[0043] 여기서, $\frac{|C_n - C_c|}{C_c}$ 은 웨버의 분수 값이다. $\bar{g}_{C_c}$ 는 중심 픽셀의 제 2픽셀값을 말한다. C_n 은 Current Frame의 인접 픽셀의 제 1픽셀값이고, C_c 는 Current Frame의 중심 픽셀의 제 1픽셀값이다.

[0044] 동적 정보 생성부는 그라디언트 연산을 통한 제 2픽셀값과 웨버의 법칙에 기초한 임계값인 T_{wc} 를 기반으로 임계값 연산을 수행할 수 있다. 보다 구체적으로 임계값 연산은 아래와 같은 수학적식 3을 통해서 설명할 수 있다.

수학적식 3

$$WVLGTP_{Fp,R}(x,y) = S(g_{F_n} - T_{wc})$$

$$WVLGTP_{Cp,R}(x,y) = S(g_{C_n} - T_{wc})$$

$$WVLGTP_{Np,R}(x,y) = S(g_{N_n} - T_{wc})$$

$$\text{Where, } S(x) = \begin{cases} 1, & \text{if } x > T \\ -1, & \text{if } x < -T \\ 0, & \text{otherwise} \end{cases}$$

[0045] 임계값 연산을 통해서 추출된 픽셀값은 제 3픽셀값(155)이라 할 수 있다. 여기서 적응형 로컬 임계값인 T의 값은 제 2픽셀값을 3진 패턴으로 변환하기 위해 도출되어야 할 값이다. P와 R은 각각 인접 화소의 수 및 반경을 뜻한다. T 값은 $|g_{C_n} - T_{wc}|$ 의 중앙값으로 계산될 수 있다. 본 발명의 일 실시예에 따르면 T 값은 8로 계산될 수 있다. 임계값 연산을 통해 추출된 제 3픽셀값은 동적 정보 생성부에서 시간 정보에 대한 특징 벡터와 공간 정보에 대한 특징 벡터 생성시 활용된다.

[0046] 종래 VLBP는 3개의 연속된 프레임에 대해서 8개의 주변 픽셀을 가져와 크기가 큰 특징 벡터를 생성할 수 있었다. 그러나 이러한 VLBP로 생성된 특징 벡터는 모호성이 매우 높아 인간 행동 인식의 정확성이 감소하는 원인 중 하나였다. 따라서 본 발명의 일 실시 예에 따르면, 동적 정보 생성부는 동적 정보를 생성할 때, 시간 정보와 공간 정보를 별도로 구분하여 생성할 수 있다.

[0047] 동적 정보 생성부에서 생성하는 시간 정보에 대한 특징 벡터(156)는 Former Frame 과 Next Frame을 Current Frame과 비교함으로써 생성되고, 공간 정보에 대한 특징 벡터(157)은 Current Frame을 기초로 하여 생성될 수 있다.

[0048] 시간 정보에 대한 특징 벡터(156)는 임계값 연산 방법을 통해 생성된 제 3픽셀값(155)에 대해서 Lowe Pattern과 Upper Pattern 연산을 통해 생성될 수 있다. Lowe Pattern 연산은 1과 0은 0의 값으로, -1의 값은 1의 값으로 변환하는 연산을 뜻한다. Upper Pattern 연산은 1은 1로, 0과 -1은 0으로 변환하는 연산을 뜻한다. 시간 정보에 대한 특징 벡터 생성시, Lowe Pattern 연산을 하면, 8개의 인접 픽셀 중 중앙 픽셀에서 가장 상단 픽셀을 기준으로 시계 방향으로 읽어 나가며 생성할 수 있다. 예를 들어, Former Frame의 제 3픽셀값은 $(1-10111-1-1)_3$ 이었고, 이를 Lowe Pattern 연산을 통해 시간 정보에 대한 특징 벡터를 생성하면, $(01000011)_2$ 가 된다. 따라서, 시간 정보에 대한 특징 벡터(156)는 Lowe Pattern과 Upper Pattern 2가지 연산을 통해, 2개의 연산 결과값을

찾을 수 있다.

[0051] 한편, 공간 정보에 대한 특징 벡터(157)는 Lowe Pattern과 Upper Pattern 연산을 Current Frame을 기준으로 계산하여 생성할 수 있다. 예를 들어, Current Frame의 제 3픽셀값이 $(-1001-100-1)_3$ 이었고, 이를 Lowe Pattern 연산을 통해 공간 정보에 대한 특징 벡터를 생성하면, $(10001001)_2$ 가 된다.

[0052] 본 발명의 또 다른 일 실시 예에 따르면, 공간 정보에 대한 특징 벡터 생성시, 크기 벡터라는 보조 개념을 활용할 수 있다. 본 발명의 일 실시 예에 따르면, 크기 벡터를 계산하여 인간 행동 인식의 정확성을 향상시킬 수 있다. 크기 벡터를 그대로 사용하지 아니하고, 크기 정보를 유지하기 위해 크기 벡터의 평균과 분산, 2가지를 활용할 수 있다.

[0053] 크기 벡터의 평균과 분산은 아래와 같은 수학적 4를 통해서 설명할 수 있다.

수학적 4

$$\begin{aligned}\mu_c &= \frac{1}{l} \sum_{n=0}^{l-1} g_{C_n} \\ \sigma_c^2 &= \frac{1}{l} \sum_{n=0}^{l-1} (g_{C_n} - \mu_c)^2 \\ Mean_{MV} &= S_M(\mu_c - t_\mu) \\ Variance_{MV} &= S_M(\sigma_c^2 - t_\sigma) \\ Where, S_M(x) &= \begin{cases} 1, & \text{if } x > T_M \\ -1, & \text{if } x < -T_M \\ 0, & \text{otherwise} \end{cases}\end{aligned}$$

[0054]

[0055] 크기 벡터 M_c 는 Current Frame 픽셀값을 기준으로 인접 픽셀값과 중앙 픽셀값의 차이의 절대값을 사용할 수 있다.

[0056] 예를 들어, 크기 벡터 M_c 는 Current Frame 픽셀값의 중앙값인 60을 기준으로, 70, 20, 105, 10, 120, 10, 30, 180에 대해서 차이의 절대값을 계산하면, 10, 40, 45, 50, 60, 60, 30, 120의 크기 벡터 값을 산출할 수 있다.

[0057] 크기 벡터 M_c 값이 구해지면, 크기 벡터 값의 평균과 그 분산을 계산할 수 있다. 예를 들어, 본 발명의 일 실시 예에서는 크기 벡터의 평균은 50.62, 크기 벡터의 분산은 890.23으로 계산될 수 있다. 여기서 t_μ 와 t_σ 값은 Current Frame 서브 프레임의 크기 벡터의 대한 평균값, 분산값에 해당된다. 또한, T_M 값은 크기 벡터를 3진법의 수로 변환하기 위한 값이다. T_M 은 Current Frame에서 중앙 픽셀값을 기준으로 인접 픽셀값과의 차이값의 절대값의 평균값을 말한다. 본 발명에서 T_M 값은 45가 된다.

[0058] 공간 정보에 대한 특징 벡터 생성시, 크기 벡터 값의 평균과 분산에 관해서도 임계값 연산 과정이 필요하다. 임계값 연산은 크기 벡터의 평균값과 t_μ 값의 차이값을 T_M 값과 비교하여 1, -1, 0의 값을 부여하는 것을 말한다. 또한 크기 벡터의 분산값과 t_σ 값의 차이값을 T_M 값과 비교하여 1, -1, 0의 값을 부여한다. 예를 들어, 본 발명에서는 크기 벡터의 평균값이 50.62이고, t_μ 값은 62.1 인바, 그 차이값은 -11.48이 되고, -45와 45 값 사이이므로 0의 값이 부여된다. 또한 크기 벡터의 분산값은 890.23이고, t_σ 값은 815.6 이므로, 그 차이값은 74.63이 되고, 45 값보다 크므로 1의 값이 부여된다. 따라서, 크기 벡터의 임계값 연산 과정을 거친 결과는 $(01)_3$ 이 된다.

[0059] 동적 정보 생성부는 크기 벡터의 임계값 연산 과정을 거친 결과를 가지고 공간 정보에 대한 특징 벡터 생성시 Lowe Pattern과 Upper Pattern 연산 활용할 수 있다. 예를 들어, Lowe Pattern 연산을 통해 공간 정보에 대한

특징 벡터를 생성하면, $(10001001)_2$ 가 되고, 여기서 크기 벡터의 평균과 분산을 고려한 Lowe Pattern 연산을 하면 $(10001001\ 00)_2$ 이 된다. 동적 정보 생성부는 크기 벡터의 평균과 분산을 고려하여 최종 공간 정보에 대한 특징 벡터(158)를 생성할 수 있다.

[0060] 동적 정보 생성부는 비디오 데이터에서 추출된 3개의 연속된 프레임에 대해서 복수 개의 시간 정보에 대한 특징 벡터, 공간 정보에 대한 특징 벡터를 생성하고, 생성된 모든 벡터들을 종합하여 동적 정보를 생성할 수 있다. 이하에서는 본 발명의 또 다른 일 실시 예에 따른 멀티 스케일 WVLGTP를 생성하는 방법을 설명하도록 한다.

[0061] 도 4는 본 발명의 또 다른 일 실시 예에 따른 멀티 스케일 WVLGTP를 생성하는 방법을 설명하기 위한 도면이다. WVLGTP 생성시, 중앙 픽셀에서 거리가 1인 인접 픽셀에 대해서만 WVLGTP를 생성할 수 있다. 즉 3x3 크기의 비디오 프레임 픽셀에 대해서 WVLGTP를 생성할 수 있다. 본 발명의 다른 일 실시 예에 따르면, 중앙 픽셀에서 거리가 2인 인접 픽셀에 대해서 멀티스케일 WVLGTP를 생성할 수 있다. 즉, 5x5 크기의 비디오 프레임 픽셀에 대해서 멀티스케일 WVLGTP를 생성할 수 있다. 멀티스케일 WVLGTP 생성시 비디오 프레임의 픽셀 크기는 한정되어 해석되지 아니하며 5x5 크기 보다 더 큰 크기를 갖을 수 있다. 멀티스케일 WVLGTP가 생성되면 도 3과 중복되는 방법을 통해, 공간 정보에 대한 특징 벡터를 생성할 수 있고, 생성된 특징 벡터를 집계하여 동적 정보를 생성할 수 있다. 동적 정보를 생성하는 상세한 설명은 도 3과 중복되는 바 설명은 생략하도록 한다.

[0062] 본 발명의 또 다른 일 실시 예에 따르면 동적 정보 생성시 멀티스케일 WVLGTP를 통한 공간 정보에 대한 특징 벡터와 WVLGTP로 생성된 공간 정보에 대한 특징 벡터를 집계하여 동적 정보로 생성할 수도 있다. 보다 구체적으로 복수의 크기를 갖는 프레임을 활용하여 WVLGTP를 2개 이상 생성할 수 있고, 생성된 각각의 WVLGTP를 통해 공간 정보에 대한 특징 벡터를 생성할 수 있다. 생성된 복수개의 공간 정보에 대한 특징 벡터를 모두 집계하여 동적 정보로 생성할 수 있다.

[0063] 도 4에서 멀티 스케일 WVLGTP를 생성할 시 도 3과 중복되는 방법에 대한 설명은 생략하도록 한다. 도 4를 참조하면, 동적 정보 생성부는 데이터 저장부의 저장된 비디오 데이터를 입력 데이터(161)로 할 수 있다. 동적 정보 생성부는 입력 데이터(161)에 대해서 3개의 연속된 비디오 프레임(162)을 추출할 수 있다.

[0064] 3개의 연속된 프레임에 대해서 중앙 픽셀에서의 거리가 2인, 5x5 크기 인접 픽셀에 대해서 도 3에서 설명한 방법과 동일한 방법으로 그라디언트 연산, 임계값 연산 등을 통해 제 2 픽셀값, 제 3 픽셀값을 생성할 수 있다. 마찬가지로 도 3에서 설명한 방법과 동일한 방법으로 시간 정보에 대한 특징 벡터 및 공간 정보에 대한 특징 벡터(163)를 생성할 수 있다.

[0065] 다만, 중앙 픽셀에서 거리가 2인 인접 픽셀에 대해 멀티 스케일 WVLGTP 생성시, 추가적인 단계를 거칠 필요가 있다.

[0066] 이러한 단계는 그라디언트 연산을 수행하기 전에 수행하는 단계로서 평균화 방식 연산을 말한다. 평균화 방식 연산을 통해 비디오 프레임의 조명의 변화로 인해 인간 행동 인식의 결과값이 모호해지는 것을 방지할 수 있다. 구체적인 평균화 방식 연산은 도 5를 참조하여 설명하도록 한다.

[0067] 도 5는 본 발명의 실시 예에 따른 멀티 스케일 WVLGTP를 생성할 시 평균화 방식 연산을 설명하기 위한 도면이다.

[0068] 멀티 스케일 WVLGTP 생성시 비디오 프레임(164)에 대해서 평균화 방식 연산 과정을 추가적으로 진행할 수 있다. 중앙 픽셀을 중심으로 인접 픽셀까지 거리가 3인 픽셀을 모두 모아 픽셀값(165)를 계산한다.

[0069] 픽셀값(165)에서 중앙 픽셀을 중심으로 인접 픽셀까지 거리가 1인 픽셀에 경우에는 그 픽셀값을 그대로 적용한다. 그 결과 인접 픽셀까지 거리가 1인 경우에는 그대로 값을 적용한 결과 픽셀값(166)을 생성할 수 있다.

[0070] 인접 픽셀까지 거리가 2인 경우에는, 인접 픽셀 하나에 대해서 3개의 양, 옆, 위, 아래의 픽셀의 평균을 결과값으로 하여 픽셀값(167)을 생성한다. 예를 들어, 픽셀값(167)의 오른쪽 상단의 149 값은 152, 140, 156의 평균값을 계산한 그 결과값을 픽셀값으로 생성한 것이다.

[0071] 인접 픽셀까지 거리가 3인 경우에는, 인접 픽셀 하나에 대해서 5개의 픽셀값에 대한 평균을 결과값(167)으로 생성한다. 예를 들어, 픽셀값(168)에서 오른쪽 상단의 140의 값은 148, 129, 128, 137, 156의 값에 대한 평균값을 계산한 그 결과값을 픽셀값으로 생성한 것이다.

[0072] 이러한 평균화 방식 연산을 통해서 비디오 프레임의 조명 변화로 인한 결과값의 모호성을 방지할 수 있다.

- [0073] 도 1로 다시 돌아가, 인간 행동 인식 장치에서 인식부에 대한 설명을 하도록 한다.
- [0074] 도 1에서, 인식부(170)은 공간 정보 생성부(130)에서 생성된 공간 정보와 동적 정보 생성부(150)에서 생성된 동적 정보를 SVM(Support Vector Machine) 알고리즘을 통해 인간 행동을 예측할 수 있다. SVM 알고리즘은 특징 벡터를 입력 데이터로 하여 입력 데이터의 특징을 학습하고, 특징에 관해 데이터를 분류하는 지도학습 모델에 일종에 해당된다. SVM은 RBF 커널 함수와 함께 비선형 SVM을 이용하여 특징 벡터의 동작을 분류할 수 있다. 또한 5배 교차검증을 수행하기 위해 C와 Y가 매개 변수로 사용될 수 있다. 이하에서는 도 6 내지 도 12를 참조하여 본 발명의 인간 행동 인식 장치에 대한 성능 확인 실험 설계 및 그 결과에 대해서 설명하도록 한다.
- [0075] 도 6은 본 발명의 실시 예에 따른 인간 행동 인식 장치에 대한 성능을 확인하기 위한 실험 대상 데이터에 관한 도면이다.
- [0076] 본 발명의 인간 행동 인식 장치에 대한 성능을 확인하기 위해, KTH 데이터셋, UCF 스포츠 액션 데이터셋, UT-Interaction 데이터 셋, Hollywood2 데이터셋, UCF-101 데이터셋을 이용할 수 있다.
- [0077] UCF-101 데이터 셋은 많은 수의 비디오 데이터로 구성되고, UT-Interaction 데이터 셋은 두 사람 간의 상호 작용을 나타내는 비디오 데이터로 구성된다. Hollywood2 데이터 셋은 KTH 데이터 셋 같은 단순한 동작이 아니라 복잡한 활동을 나타내는 데이터로 구성된다. 또한 Hollywood2 데이터 셋은 각 비디오에 눈에 띄는 카메라 동작과 빠른 장면 변경이 포함되어 있어 인간 행동을 인식하기 가장 어려운 데이터 셋에 해당된다.
- [0078] 도 6(a)는 KTH 데이터 셋의 샘플 데이터에 해당된다. KTH 데이터 셋은 걷기, 달리기, 조깅, 권투, 박수 등 인간 25명이 수행하는 100개의 시퀀스가 *?*포함된다.
- [0079] 도 6(b)는 UCF 스포츠 액션 데이터 셋의 샘플 데이터에 해당된다. UCF 스포츠 액션 데이터 셋은 해상도가 150인 비디오 데이터를 포함할 수 있다. UCF 스포츠 액션 데이터 셋은 걷기, 달리기, 발길질, 리프팅, 다이빙, 골프 스윙, 승마, 스케이트 보드, 스윙 사이드 및 스윙 벤치 등을 포함할 수 있다. UCF 스포츠 액션 데이터 셋은 다양한 관점에서 촬영한 비디오 데이터 및 많은 카메라 움직임을 포함하는 비디오 데이터로 구성될 수 있다.
- [0080] 도 6(c)는 UT-Interaction 데이터 셋의 샘플 데이터에 해당된다. UT-Interaction 데이터 셋은 720 × 480의 해상도를 가진 120 개의 비디오로 구성된다. UT-Interaction 데이터 셋은 악수, 포옹, 발로 차기, 밀기, 가리 키기 및 펀치 등의 6가지 액션에 관한 비디오 데이터를 포함할 수 있다. 또한, UT-Interaction 데이터 셋의 비디오 데이터는 15가지의 다른 옷 조건을 가진 사람들이 6가지 액션을 수행하는 데이터를 포함하고 있다.
- [0081] 도 6(d)는 Hollywood2 데이터 셋의 샘플 데이터에 해당된다. Hollywood2 데이터 셋은 12가지 인간 행동을 가진 3669개의 *?*비디오 데이터로 구성된다.
- [0082] Hollywood2 데이터 셋은 전화 응답, 자동차 운전, 식사, 싸우기, 차에서 나오기, 악수, 포옹, 키스, 뛰기, 앉기, 일어나기, 일어나기 등의 액션 데이터를 포함할 수 있다. Hollywood2 데이터 셋은 빠른 카메라 움직임과 빠른 장면 전환 관련 비디오 데이터를 포함하고 있다.
- [0083] 도 6(e)는 UCF-101 데이터 셋의 샘플 데이터에 해당된다. UCF-101 데이터 셋은 13320개의 비디오 데이터로 구성된 가장 큰 액션 데이터 셋에 해당된다. UCF-101 데이터 셋은 101개의 액션 동작을 포함하고, 비디오 데이터는 YouTube에서 가져올 수 있다. UCF-101 데이터 셋은 4개 내지 7개의 동작 시퀀스를 다루는 25개의 그룹으로 구분될 수 있다.
- [0084] 본 발명의 인간 행동 인식 장치의 성능을 확인하기 위한 실험 설계를 위해 5개 데이터 모두에 대해 비디오 데이터의 70%를 인식부의 지도학습 모델인 SVM 알고리즘을 훈련시켰다. 비디오 데이터의 나머지 30%에 대하여 본 발명의 인간 행동 인식 장치의 행동 인식 성능을 확인하였다. 또한 5배의 교차 검증을 통해, 장치의 성능을 확인하였다.
- [0085] 본 발명의 인간 행동 인식 장치의 성능을 확인하기 위해, 실험 진행하는 동안 비디오 데이터의 프레임 수는 동일한 시간 간격으로 고정시켜 수행할 수 있다. 비디오 프레임 수는 35로 설정하여 샘플링 할 수 있다. 이하에서는 본 발명의 인간 행동 인식 장치 성능에 관한 실험 결과를 설명하도록 한다.
- [0086] 도 6은 본 발명의 실시 예에 따른 인간 행동 인식 장치에 대한 성능을 확인하기 위한 실험 대상 데이터에 관한 도면이다.
- [0087] 본 발명의 인간 행동 인식 장치에 대한 성능을 확인하기 위해, KTH 데이터셋, UCF 스포츠 액션 데이터셋, UT-

Interaction 데이터 셋, Hollywood2 데이터셋, UCF-101 데이터셋을 이용할 수 있다.

- [0088] UCF-101 데이터 셋은 많은 수의 비디오 데이터로 구성되고, UT-Interaction 데이터 셋은 두 사람 간의 상호 작용을 나타내는 비디오 데이터로 구성된다. Hollywood2 데이터 셋은 KTH 데이터 셋 같은 단순한 동작이 아니라 복잡한 활동을 나타내는 데이터로 구성된다. 또한 Hollywood2 데이터 셋은 각 비디오에 눈에 띄는 카메라 동작과 빠른 장면 변경이 포함되어 있어 인간 행동을 인식하기 가장 어려운 데이터 셋에 해당된다.
- [0089] 도 6(a)는 KTH 데이터 셋의 샘플 데이터에 해당된다. KTH 데이터 셋은 걷기, 달리기, 조깅, 권투, 박수 등 인간 25명이 수행하는 100개의 시퀀스가 *?*포함된다.
- [0090] 도 6(b)는 UCF 스포츠 액션 데이터 셋의 샘플 데이터에 해당된다. UCF 스포츠 액션 데이터 셋은 해상도가 150인 비디오 데이터를 포함할 수 있다. UCF 스포츠 액션 데이터 셋은 걷기, 달리기, 발길질, 리프팅, 다이빙, 골프 스윙, 승마, 스케이트 보드, 스윙 사이드 및 스윙 벤치 등을 포함할 수 있다. UCF 스포츠 액션 데이터 셋은 다양한 관점에서 촬영한 비디오 데이터 및 많은 카메라 움직임을 포함하는 비디오 데이터로 구성될 수 있다.
- [0091] 도 6(c)는 UT-Interaction 데이터 셋의 샘플 데이터에 해당된다. UT-Interaction 데이터 셋은 720×480의 해상도를 가진 120 개의 비디오로 구성된다. UT-Interaction 데이터 셋은 악수, 포옹, 발로 차기, 밀기, 가리 키기 및 펀치 등의 6가지 액션에 관한 비디오 데이터를 포함할 수 있다. 또한, UT-Interaction 데이터 셋의 비디오 데이터는 15가지의 다른 옷 조건을 가진 사람들이 6가지 액션을 수행하는 데이터를 포함하고 있다.
- [0092] 도 6(d)는 Hollywood2 데이터 셋의 샘플 데이터에 해당된다. Hollywood2 데이터 셋은 12가지 인간 행동을 가진 3669개의 비디오 데이터로 구성된다.
- [0093] Hollywood2 데이터 셋은 전화 응답, 자동차 운전, 식사, 싸우기, 차에서 나오기, 악수, 포옹, 키스, 뛰기, 앉기, 일어나기, 일어나기 등의 액션 데이터를 포함할 수 있다. Hollywood2 데이터 셋은 빠른 카메라 움직임과 빠른 장면 전환 관련 비디오 데이터를 포함하고 있다.
- [0094] 도 6(e)는 UCF-101 데이터 셋의 샘플 데이터에 해당된다. UCF-101 데이터 셋은 13320개의 비디오 데이터로 구성된 가장 큰 액션 데이터 셋에 해당된다. UCF-101 데이터 셋은 101개의 액션 동작을 포함하고, 비디오 데이터는 YouTube에서 가져올 수 있다. UCF-101 데이터 셋은 4개 내지 7개의 동작 시퀀스를 다루는 25개의 그룹으로 구분될 수 있다.
- [0095] 본 발명의 인간 행동 인식 장치의 성능을 확인하기 위한 실험 설계를 위해 5개 데이터 모두에 대해 비디오 데이터의 70%를 인식부의 지도학습 모델인 SVM 알고리즘을 훈련시켰다. 비디오 데이터의 나머지 30%에 대하여 본 발명의 인간 행동 인식 장치의 행동 인식 성능을 확인하였다. 또한 5배의 교차 검증을 통해, 장치의 성능을 확인하였다.
- [0096] 본 발명의 인간 행동 인식 장치의 성능을 확인하기 위해, 실험 진행하는 동안 비디오 데이터의 프레임 수는 동일한 시간 간격으로 고정시켜 수행할 수 있다. 비디오 프레임 수는 35로 설정하여 샘플링 할 수 있다. 이하에서는 본 발명의 인간 행동 인식 장치 성능에 관한 실험 결과를 설명하도록 한다.
- [0097] 도 7은 본 발명의 실시 예에 따른 KTH 데이터 셋에 대한 인간 행동 인식 성능 확인 실험 결과값을 나타낸 도면이다. 도 7을 참조하면, 본 발명의 인간 행동 인식 장치는 KTH 데이터 셋의 카테고리 중 박수치기 부분에서 가장 높은 정확도를 보여준다. 도 7는 KTH 데이터 셋에서 Inception-Resnet-v2 네트워크 컨볼루션 모델을 활용하여 공간 정보만을 분석한 모델, WVLGTP를 활용한 동적 정보를 분석한 모델, 공간 정보와 WVLGTP를 활용한 동적 정보를 모두 분석한 모델에 대한 각각의 평균 행동 인식 정확도를 나타낸다. 3가지 모델에 대한 평균 인식 정확도는 각각 94.9%, 94.4%, 96.5%로 확인되었다.
- [0098] 도 7을 참조하면, KTH 데이터 셋에 대한 인간 행동 인식 관련 다른 모델과 본 발명의 모델을 비교한 결과를 확인할 수 있다. 공간 정보와 WVLGTP를 활용한 동적 정보를 모두 분석한 모델의 경우 기존 VLBP 접근 방법보다 정확도가 향상됨을 확인할 수 있다. 본 발명의 Inception-Resnet-v2 네트워크 컨볼루션 모델을 이용하여 공간 정보와 WVLGTP 활용한 동적 정보를 분석한 모델은 VLBP, LBP-TOP, MBP, ALMD, Extended LBP-TOP, LTP-TOP, 3D Gradient LBP 보다 더 나은 행동 인식 정확도를 갖는다. 본 발명의 인간 행동 인식 장치의 인식 정확도가 96.5%임에 반해, Inception-Resnet-v2 네트워크 컨볼루션 모델을 이용한 iDT, VLBP, ALMD는 각각 95.7%, 90.1%, 93.7% 정확도를 갖는 바, 상대적으로 우수한 성능을 갖음을 확인할 수 있다.
- [0099] 도 8은 본 발명의 실시 예에 따른 UCF 스포츠 액션 데이터 셋에 대한 인간 행동 인식 성능 확인 실험 결과값을 나타낸 도면이다. 도 8은 UCF 스포츠 액션 데이터 셋에서 Inception-Resnet-v2 네트워크 컨볼루션 모델을 활용

하여 공간 정보만을 분석한 모델, WVLGTP를 활용한 동적 정보를 분석한 모델, 공간 정보와 WVLGTP를 활용한 동적 정보를 모두 분석한 모델에 대한 각각의 평균 행동 인식 정확도를 나타낸다. 3가지 모델에 대한 평균 인식 정확도는 각각 92.9%, 93.3%, 94.6%에 해당됨을 알 수 있다. UCF 스포츠 액션 데이터 셋에서 본 발명의 인간 행동 인식 장치는 다이빙, 승마, 스윙 벤치 카테고리에서 인식 정확도가 높은 것을 알 수 있다.

[0100] 도 8을 참조하면, UCF 스포츠 액션 데이터 셋에 대한 인간 행동 인식 관련 다른 모델과 본 발명의 모델을 비교한 결과를 확인할 수 있다. 본 발명의 공간 정보와 WVLGTP를 활용한 동적 정보를 모두 분석한 모델이 SFT 모델과 MAE 모델보다 더 나은 행동 인식 정확도를 갖음을 알 수 있다. UCF 스포츠 액션 데이터 셋에 대해, 본 발명의 공간 정보와 WVLGTP를 활용한 동적 정보를 모두 분석한 모델은 94.6%의 인식 정확도를 갖는다.

[0101] 도 9는 본 발명의 실시 예에 따른 UT-Interaction 데이터 셋에 대한 인간 행동 인식 성능 확인 실험 결과값을 나타낸 도면이다. 도 9는 UT-Interaction 데이터 셋에서 Inception-Resnet-v2 네트워크 컨볼루션 모델을 활용하여 공간 정보만을 분석한 모델, WVLGTP를 활용한 동적 정보를 분석한 모델, 공간 정보와 WVLGTP를 활용한 동적 정보를 모두 분석한 모델에 대한 각각의 평균 행동 인식 정확도를 나타낸다. 3가지 모델에 대한 평균 인식 정확도는 각각 96.7%, 97.7%, 98.7%에 해당됨을 알 수 있다.

[0102] 도 9를 참조하면, UT-Interaction 데이터 셋에 대한 인간 행동 인식 관련 다른 모델과 본 발명의 모델을 비교한 결과를 확인할 수 있다. 본 발명의 공간 정보와 WVLGTP를 활용한 동적 정보를 모두 분석한 모델의 경우 VLBP, MBP, ALMD, 3D Gradient LBP 보다 더 나은 행동 인식 정확도를 갖는다. UT-Interaction 데이터 셋에 대해, 본 발명의 본 발명의 공간 정보와 WVLGTP를 활용한 동적 정보를 모두 분석한 모델은 98.67%의 인식 정확도를 갖는다.

[0103] 도 10은 본 발명의 실시 예에 따른 Hollywood2 데이터 셋에 대한 인간 행동 인식 성능 확인 실험 결과값을 나타낸 도면이다. 도 10은 Hollywood2 데이터 셋에서 UCF 스포츠 액션 데이터 셋에서 Inception-Resnet-v2 네트워크 컨볼루션 모델을 활용하여 공간 정보만을 분석한 모델, WVLGTP를 활용한 동적 정보를 분석한 모델, 공간 정보와 WVLGTP를 활용한 동적 정보를 모두 분석한 모델에 대한 각각의 평균 행동 인식 정확도를 나타낸다. 3가지 모델에 대한 평균 인식 정확도는 각각 67.04%, 66.74%, 70.3%에 해당됨을 알 수 있다. Hollywood2 데이터 셋에서 본 발명의 인간 행동 인식 장치는 달리기, 키스, 일어나기 카테고리에서 인식 정확도가 높은 것을 알 수 있다.

[0104] 도 10을 참조하면, Hollywood2 데이터 셋에 대한 인간 행동 인식 관련 다른 모델과 본 발명의 모델을 비교한 결과를 확인할 수 있다. Hollywood2 데이터 셋에서 가장 높은 정확도를 갖는 것은 Inception-Resnet-v2 네트워크 컨볼루션 모델을 이용한 iDT 모델인 것으로 확인되었다. Hollywood2 데이터 셋에 대해, 본 발명의 공간 정보와 WVLGTP를 활용한 동적 정보를 모두 분석한 모델은 70.3%의 인식 정확도를 갖는다.

[0105] 도 11은 본 발명의 실시 예에 따른 UCF-101 데이터 셋에 대한 인간 행동 인식 성능 확인 실험 결과값을 나타낸 도면이다. 도 11을 참조하면, UCF-101 데이터 셋에 대한 인간 행동 인식 관련 다른 모델과 본 발명의 모델을 비교한 결과를 확인할 수 있다. 공간 정보와 WVLGTP를 활용한 동적 정보를 모두 분석한 모델의 경우 VLBP, MBP, ALMD, 3D Gradient LBP 보다 더 나은 정확도를 갖는다. 또한 Inception-Resnet-v2 네트워크 컨볼루션 모델을 이용한 iDT 모델에 비해서도 2.2% 더 나은 정확도를 갖는다. 본 발명의 공간 정보와 WVLGTP를 활용한 동적 정보를 모두 분석한 모델은 UCF-101 데이터 셋에 대해 94.9 인식 정확도를 갖음을 알 수 있다.

[0106] 도 12는 본 발명의 실시 예에 따른 멀티 스케일 WVLGTP에 행동 인식 성능 확인 실험 결과값을 나타낸 도면이다. 도 12를 참조하면, 모든 데이터 셋에 대해 중앙 픽셀로부터 거리가 1인 인접 픽셀 8개와 중앙 픽셀로부터 거리가 2인 인접 픽셀 16개 고려하여 멀티 스케일 WVLGTP 생성한 경우 가장 좋은 인식 정확도를 갖음을 알 수 있다. 즉, 비디오 프레임에서 3x3 크기 픽셀에 대한 WVLGTP와 5x5 크기 픽셀에 대한 멀티스케일 WVLGTP를 활용하여 동적 정보를 생성한 모델이 평균 인식도가 가장 높음을 알 수 있다.

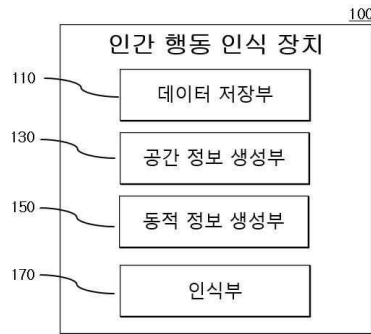
[0107] 진술한 실험 결과를 종합하면, 인간 행동 인식 관련하여 Inception-Resnet-v2 네트워크 컨볼루션 모델을 이용한 공간 정보와 WVLGTP 모델을 이용한 동적 정보를 활용한 모델에서 인간 행동 인식 정확도 향상이 있음을 확인할 수 있다. 나아가, 3x3 크기 픽셀에 대한 WVLGTP와 5x5 크기 픽셀에 대한 멀티스케일 WVLGTP를 활용하여 동적 정보를 생성한 모델이 3x3 크기 픽셀에 대한 WVLGTP를 통해 동적 정보를 분석한 모델 보다 더 향상된 정확도를 갖음을 알 수 있다.

[0108] 본 명세서와 도면에 개시된 본 발명의 실시 예들은 본 발명의 기술 내용을 쉽게 설명하고 본 발명의 이해를 돕기 위해 특정 예를 제시한 것뿐이며, 본 발명의 범위를 한정하고자 하는 것은 아니다. 여기에 개시된 실시 예들

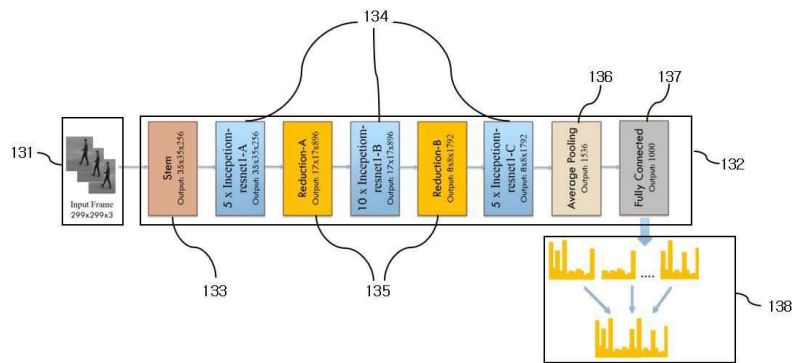
이외에도 본 발명의 기술적 사상에 바탕을 둔 다른 변형 예들이 실시 가능하다는 것은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 자명한 것이다.

도면

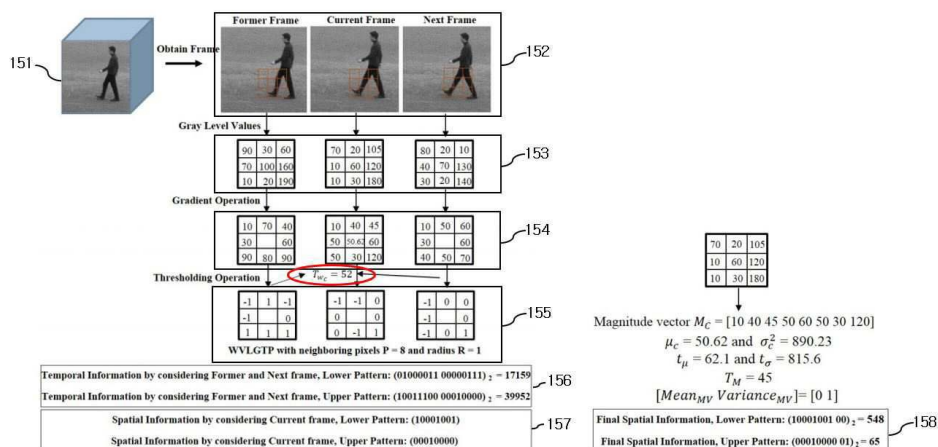
도면1



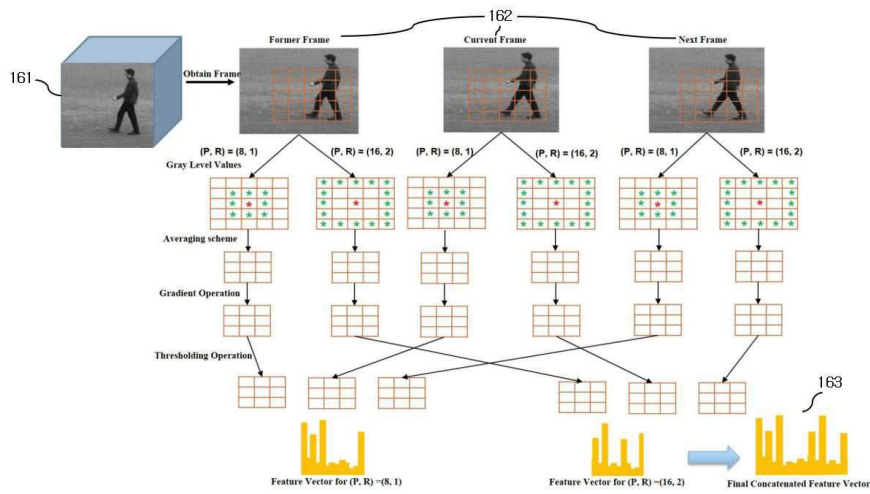
도면2



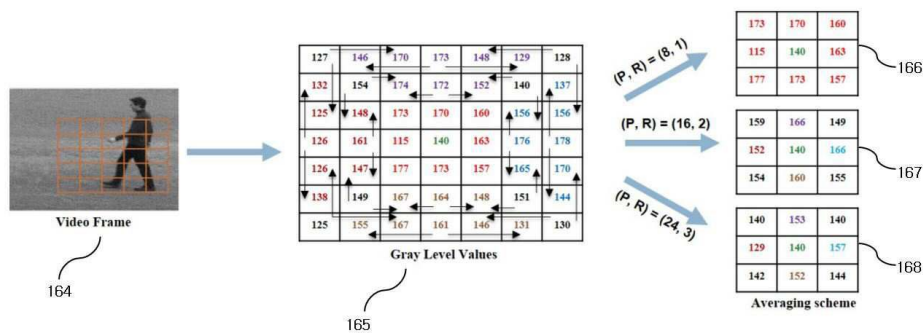
도면3



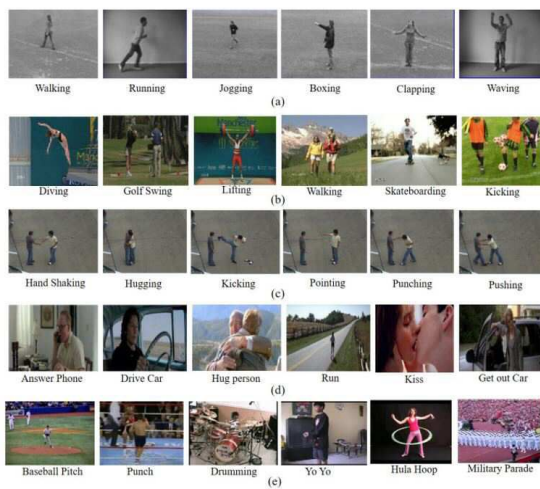
도면4



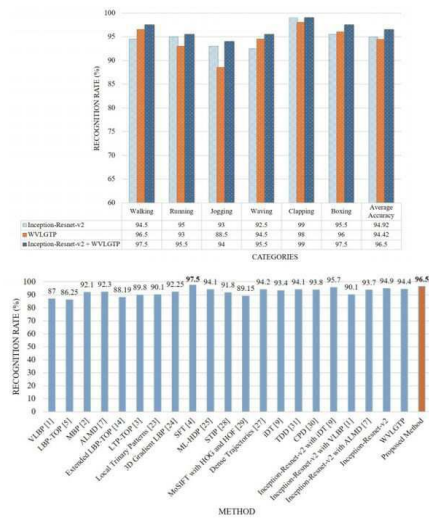
도면5



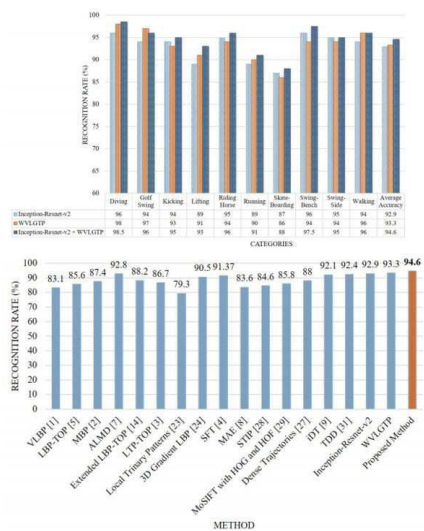
도면6



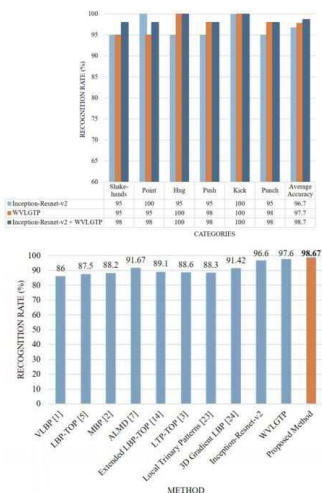
도면7



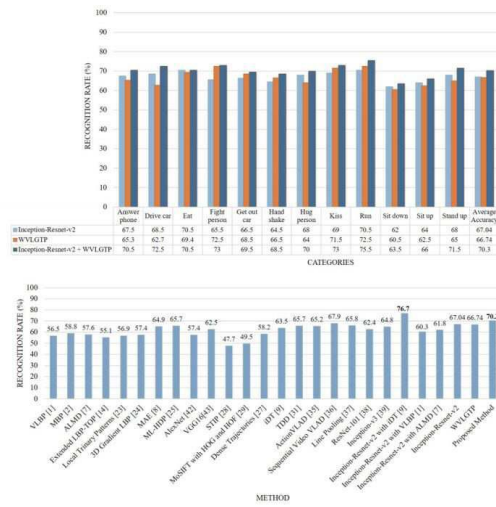
도면8



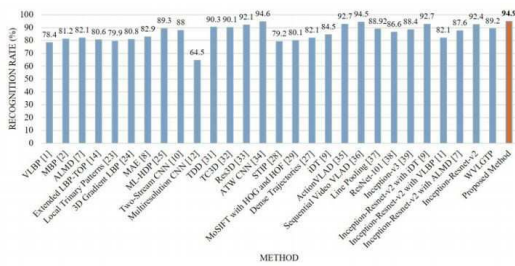
도면9



도면10



도면11



도면12

