



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2023년01월27일  
(11) 등록번호 10-2493075  
(24) 등록일자 2023년01월25일

- (51) 국제특허분류(Int. Cl.)  
G06V 10/42 (2022.01) G06F 17/11 (2006.01)  
G06F 17/16 (2006.01) G06F 18/00 (2023.01)
- (52) CPC특허분류  
G06V 10/431 (2023.01)  
G06F 17/11 (2013.01)
- (21) 출원번호 10-2019-7026847
- (22) 출원일자(국제) 2018년02월19일  
심사청구일자 2021년02월16일
- (85) 번역문제출일자 2019년09월11일
- (65) 공개번호 10-2019-0121323
- (43) 공개일자 2019년10월25일
- (86) 국제출원번호 PCT/EP2018/054032
- (87) 국제공개번호 WO 2018/150026  
국제공개일자 2018년08월23일
- (30) 우선권주장  
17156741.5 2017년02월17일  
유럽특허청(EPO)(EP)
- (56) 선행기술조사문헌  
EP02790130 A1\*  
KR1020130067612 A\*  
\*는 심사관에 의하여 인용된 문헌

- (73) 특허권자  
인텔 코포레이션  
미합중국 캘리포니아 95054 산타클라라 미션 칼리지 블러바드 2200
- (72) 발명자  
리준더스, 크리스티안 에릭  
이탈리아, 00135 로마, 비아 카발레세 8
- (74) 대리인  
제일특허법인(유)

전체 청구항 수 : 총 10 항

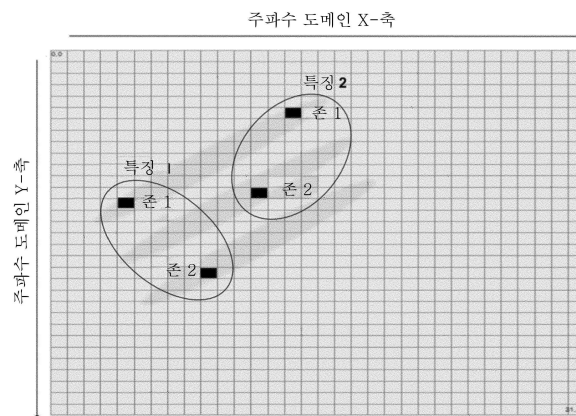
심사관 : 강현일

(54) 발명의 명칭 콘텐츠 검출에 대한 이미지 프로세싱 방법

(57) 요약

이미지 또는 이미지의 디지털 비디오 스트림을 기반으로 하여, 전자 프로세싱 유닛에 의해 실행되는 콘텐츠 검출 방법으로서, 상기 이미지는, 식별될 콘텐츠를 나타내는 주파수 도메인 신호를 강화하고 그 후에 격리시키고, 비디오 스트림 또는 이미지 내의 콘텐츠에 대한 주파수 도메인 노이즈를 감소 또는 무시하기 위한 목적으로, 상기 (뒷면에 계속)

대표도



비디오 스트림의 프레임 시퀀스에 의해 또는 단일 프레임에 의해 정의되는, 콘텐츠 검출 방법은: 상기 비디오 스트림의 해당 프레임 시퀀스 또는 해당 단일 프레임으로부터 디지털 이미지 시퀀스 또는 디지털 이미지를 얻는 단계 - 모든 디지털 이미지는 공간 도메인에서 정의됨; 상기 단일 프레임의 적어도 일 부분 또는 상기 프레임 시퀀스의 적어도 2 개의 프레임을 각각 커버하는 하나 이상의 쌍의 최소 존을 선택하는 단계 - 각 쌍의 최소 존은 특징을 발생시키고, 각 존은 2 개의 공간 데이터 시퀀스에 의해 정의됨; 각 존에 대해, L- 변환의 2D 변형을 통해 상기 2 개의 공간 데이터 시퀀스를 조합하고, 각 존에 대한 주파수 도메인 데이터의 전달 함수, 형상 및 방향을 변화시켜, 상기 특징 각각에 대해 정규화된 복소 벡터를 발생시킴으로써, 선택 특징을 주파수 도메인 데이터로 변환시키는 단계; 콘텐츠 분류기에 대한 입력으로서 사용 가능한, 식별될 콘텐츠의 모델을 정의하기 위해 상기 정규화된 복소 벡터 모두를 조합하는 단계;를 포함한다. 현재의 검출 방법은 더 나은 입력 데이터, 더 빠른 계산, 특히 비디오 스트림에서 시간적 데이터를 캡처하는 모델을 발생시키는 매우 증가된 능력을 제공한다. 이점이 있다.

(52) CPC특허분류

*G06F 17/16* (2013.01)

*G06F 18/213* (2023.01)

*G06F 18/24* (2023.01)

*G06V 20/64* (2023.01)

*G06T 2207/20076* (2013.01)

## 명세서

### 청구범위

#### 청구항 1

이미지 또는 이미지의 디지털 비디오 스트림을 기반으로 하여, 전자 프로세싱 유닛에 의해 실행되는 콘텐츠 검출 방법으로서, 상기 이미지는, 식별될 콘텐츠를 나타내는 주파수 도메인 신호를 강화하고 그 후에 격리시키고, 비디오 스트림 또는 이미지 내의 콘텐츠에 대한 주파수 도메인 노이즈를 감소 또는 무시하기 위한 목적으로, 상기 비디오 스트림의 프레임 시퀀스에 의해 또는 단일 프레임에 의해 정의되는, 콘텐츠 검출 방법에 있어서,

- 상기 비디오 스트림의 해당 프레임 시퀀스 또는 해당 단일 프레임으로부터 디지털 이미지 시퀀스 또는 디지털 이미지를 얻는 단계 - 모든 디지털 이미지는 공간 도메인에서 정의됨;
- 상기 단일 프레임의 적어도 일 부분 또는 상기 프레임 시퀀스의 적어도 2 개의 프레임을 각각 커버하는 하나 이상의 쌍의 최소 존을 선택하는 단계 - 각 쌍의 최소 존은 특징 (feature)을 발생시키고, 각 존은 2 개의 공간 데이터 시퀀스에 의해 정의됨;
- 각 존에 대해, L- 변환의 2D 변형을 통해 상기 2 개의 공간 데이터 시퀀스를 조합하고, 각 존에 대한 주파수 도메인 데이터의 전달 함수, 형상 및 방향을 변화시켜, 상기 특징 각각에 대해 정규화된 복소 벡터를 발생시킴으로써, 상기 특징을 주파수 도메인 데이터로 변환시키는 단계;
- 콘텐츠 분류기에 대한 입력으로서 사용 가능한, 식별될 콘텐츠의 모델을 정의하기 위해 상기 정규화된 복소 벡터 모두를 조합하는 단계;를 포함하는, 콘텐츠 검출 방법.

#### 청구항 2

청구항 1에 있어서,

상기 특징을 주파수 도메인 데이터로 변환시키는 단계는 프레임의 다양한 수/선택으로부터 공간 데이터를 사용하는, 콘텐츠 검출 방법.

#### 청구항 3

청구항 1에 있어서,

상기 최소 존이 식별되는 입력 프레임을 발생시키기 위해 전체 입력 이미지는 검색 로직이 사용되는, 콘텐츠 검출 방법.

#### 청구항 4

청구항 1에 있어서,

상기 최소 존은 국부 해상도를 증가시키기 위해 가능하다면 서로 부분적으로 겹쳐지거나 나란히 위치되어 함께 그룹화되는, 콘텐츠 검출 방법.

#### 청구항 5

청구항 1에 있어서,

상기 특징을 주파수 도메인 데이터로 변환하는 것은 상기 2 개의 공간 데이터 시퀀스에 대하여 병렬로 수행되는, 콘텐츠 검출 방법.

#### 청구항 6

청구항 1에 있어서,

상기 변환시키는 단계에서, 먼저 1D 피켈 계산이 행에 의해 수행되고, 그 후에 결과는 1D 피켈 계산이 열에 의해 수행되는 제 2 단계 동안에 사용되거나, 또는 그 반대로도 행해지는, 콘텐츠 검출 방법.

**청구항 7**

청구항 1에 있어서,  
한 쌍의 각 최소 존에 대해 상이한 타겟 주파수가 선택되는, 콘텐츠 검출 방법.

**청구항 8**

청구항 1에 있어서,  
상기 주파수 도메인 데이터로의 변환 단계를 위한 입력 셀은 변환 계산이 필요한 위치 주위에서만 취해지는, 콘텐츠 검출 방법.

**청구항 9**

청구항 8에 있어서,  
상기 위치의 변환 계산은 상기 위치에서 행 및 열에 대한 1D 출력을 별도로 계산하고 그 후에 이를 단일 값으로 조합하여 취해지는, 콘텐츠 검출 방법.

**청구항 10**

청구항 1에 있어서,  
상기 전달 함수는, 제 1 입력 및 제 2 입력이 상이한 개별 전달 함수 설정을 가지도록 최소 존의 각 입력에 대해 별도로 선택되는, 콘텐츠 검출 방법.

**발명의 설명**

**기술 분야**

- [0001] 본 개시는, 이미지를 디지털화하기 위한 적절한 수단이 구비된 임의의 종류의 디바이스, 가능하게는 일반적인 디바이스에 내장되도록 의도된, 콘텐츠 검출에 대한 방법에 관한 것이다. 방법은 디지털 비디오 스트림에 제한되지는 않지만, 디지털 비디오 스트림에 사용할 시간 모델 (temporal models)을 생성하는데 특히 유용하다.
- [0002] 콘텐츠는 검출하기에 관심이 될 수 있는 임의의 객체가 될 수 있음을 의미한다. 그 후, 콘텐츠 개념은 객체, 즉 이미지를 통해 볼 수 있는 물리적 항목에 제한되는 것이 아니라, 인수 (argument) 또는 종류에 의해 선택된 객체 패밀리, 예를 들어, 폭력 표현, 누드 표시, 스포츠 활동 디스플레이, 군중 얼굴, 차량 식별 및 종류 또는 크기별로 선택, 자율-주행 차량 시스템을 위한 보행자, 자전거 운전자 및 교통 신호 분별, 장소 또는 풍경 인식 등의 이미지일 수도 있다. 다른 것들 중에서 특정 콘텐츠를 검출하는 단계를 포함하는 임의의 필드는 본 발명에 수반될 수 있다.
- [0003] 일반적인 디바이스는 스마트폰, 태블릿, 랩톱 또는 적합한 디지털 비디오 카메라를 갖는 임의의 휴대용 또는 핸드-헬드 디바이스와 같은 쉽게 상업적으로 구매할 수 있는 전자 디바이스가 될 수 있음을 의미한다. 다른 한편으로, 디바이스는 이미지 컴퓨팅을 위해 컴퓨터 또는 서버에 연결된, 네트워크로 구성될 수 있는 하나 이상의 카메라로 나타낼 수 있다. 게다가, 검출 방법은 하드웨어에서 운영되는 소프트웨어를 통해, 저장 메모리, RAM 메모리 및 적어도 마이크로프로세서를 포함하는 적절한 하드웨어에 의해 저장된 디지털 이미지 또는 비디오 상에서, 오프라인 프로세스에서 간단하게 구현될 수 있다.
- [0004] 본 개시는 또한 비디오 스트림 또는 일련의 디지털 이미지를 생성하는 디지털 카메라가 적어도 구비된 디바이스 또는 시스템을 동작시키는 방법에 관한 것으로서, 동작 디바이스 또는 시스템에 의해 액세스 가능한, 메모리 디바이스에 저장된 적어도 하나의 분류기 모델과 관련하여, 디바이스 또는 시스템의 카메라 및 프로세서 둘 다를 통해 객체 검출을 얻는 것이다.

**배경 기술**

- [0005] 이미지 프로세싱의 공학기술 적용에서 근사치는 갈수록 더 더 이상 수용될 수 없게 되었다. 그러한 적용의 예시는 진단에서 및 시뮬레이션을 위한 모델 생성에서 또는 시뮬레이터에서 찾을 수 있다. 신호와 관련하여 높은 수준의 노이즈를 갖는 강한 비-선형, 과도 및 동적 효과를 적절히 캡처해야 한다. 이 적용 프로그램의 예

시는 최소한의 정보량만을 사용하여 객체를 검출하는 것으로 나타난다. 그러한 검출은 부분적으로 폐색된 객체에 대해 또는 촬영 카메라로부터 멀리 떨어진 객체에 대해, 또는 심지어 많은 상이한 표현을 갖는 객체 및 빠르게 움직이는 객체, 또는 마지막으로 이들의 조합에 대해 요구될 수 있다.

- [0006] 이미지 프로세싱에서 콘텐츠 및 객체 검출 기술은 다양한 상황에서 널리 적용되고 있다. 예로서 그리고 제한적인 목적없이, 이러한 인식은 소셜 네트워크 상의 인간 안면 태깅 기술 (human face tagging technology)에서, 손 제스처 인식용 소프트웨어에서, 보행자, 자전거 및 차량의 검출용 자동차 소프트웨어에서, 신체 움직임의 인식용 소프트웨어에서, 증강 현실 및 3D 효과를 갖는 스크린용 인간 얼굴 검출 기술에서, 증강 현실용 객체 인식에서, 머리 방향 또는 눈 방향 추적을 사용한 인터페이스에서, 보안 시스템용 객체 추적 기술에서, 그리고 마지막으로 시선 추적 기술에서 사용된다.
- [0007] 그러한 객체 검출 제품을 위한 특정 하드웨어 솔루션이 여전히 있다; 예를 들어, 시선 추적 및 제스처 인식은 필요한 정확도를 얻기 위해 통합 적외선 프로젝터 또는 스테레오 비전 (stereo vision)용 추가 카메라가 종종 필요한 기술이다.
- [0008] 차세대 상호작용 인터페이스 및 운영 체제로 이동되는 이들 객체 검출 기술 모두는 미래 추세가 분명하다. 그러한 기술을 나타낼 수 있는 디바이스는, 예를 들어, 스마트폰, 태블릿, 상호작용 안경과 같은 착용식 하드웨어 뿐만 아니라, 미래의 가정, 사무실 또는 공공 장소에서 임의의 종류의 상호작용 객체이다. 이들 디바이스는 상호작용 텔레비전 또는 지능형 주택과 같은 특수 용도로 제공될 수 있으며, 자동차 안전 시스템, 건강관리, 광고, 보안 카메라 네트워크, 사물 인터넷 및 많은 다른 가능한 용도로도 사용될 수 있다.
- [0009] 그러한 기술은 임의의 디바이스 또는 네트워크 연결 디바이스에 기본적으로 통합될 수 있고, 이 경우에 재프로그래밍 가능한 하드웨어가 사용되고 비디오 카메라 입력이 부가될 수 있다.
- [0010] 객체 검출 알고리즘의 기능을 전적으로 돕기 위해 디바이스에 추가 하드웨어를 부가함은 비용을 증가시키고 배터리 소모가 더 많음을 암시한다. 그 후에, 소형 하드웨어를 생성하는데 추가 연구 및 개발 비용이 필요하며, 현재의 최신 하드웨어는 종종 여전히 너무 커서 대부분의 가전 디바이스에 통합될 수 없다.
- [0011] 하드웨어 비용 외에도, 광범위하게, 예를 들어, 모바일 하드웨어 플랫폼에 관한 객체 검출 기술의 대규모 사용을 저해하는 것은 필요한 객체 검출 계산이 카메라의 프레임 속도에 맞추기에 너무 느리거나, 그렇게 하기에 이용 가능한 프로세싱 전력을 너무 많이 사용한다는 점이다.
- [0012] 그러므로, 차세대 운영 체제 및 디바이스에 대한 구현이 대량으로 실현되기 전에, 그러한 객체 검출 기술은 우선 최신 기술보다 이미지를 훨씬 빠르게 프로세싱할 수 있는 소프트웨어-전용 솔루션을 필요로 할 것이다.
- [0013] 소프트웨어-전용 솔루션을 향한 이러한 이동은 또한, 점점 더 높은 프레임 속도, 보다 나은 모션 프로세싱, 보다 효과적인 색상 강조, 대비에 대한 보다 예리한 적응, 광 변화에 대한 보다 나은 스마트 적응 및 점점 더 높은 화면 해상도를 제공하는 지속적인 카메라 기술 개선으로 인해 촉진되고 있다. 이러한 추세는 객체 검출을 위한 소프트웨어-전용 솔루션의 유효성을 추가로 증가시킬 것이다.
- [0014] 가능한 적은 프로세싱 전력을 사용하기 위해 콘텐츠 검출에 대한 필요성은 배터리 수명 절약 및 실시간 사용 요건 둘 다에 대해 의도된다. 실시간 객체 검출 알고리즘은 보다 자연스러운 사용자 인터페이스를 갖춘 디바이스를 생성해야 한다. 실시간으로 운영되는 동안 객체 검출 알고리즘은 또한 전경 (foreground)에서 운영되는 주요 프로세스를 제한함 없이, 배경에서 운영될 필요가 있다.
- [0015] 추가로, 유의해야 하는 바와 같이, 입력 이미지 크기가 증가함에 따라 필요한 계산량이 기하급수적으로 증가될 수 있다. 비디오 프레임 속도의 증가는 또한, 다음 비디오 입력 프레임이 도착하기 전에 객체 검출 알고리즘이 계산을 끝내는 시간이 줄어들 수 있음을 의미한다.
- [0016] 그러므로, 점점 더 높은 비디오 프레임 속도 및 증가하는 입력 이미지 품질의 부작용은 최신 기술 객체 검출 알고리즘이 수용 가능한 프로세싱 속도로 돌아가기 위해 입력 이미지를 점점 더 다운 샘플링할 필요가 있고, 이로써 보다 높은 품질의 입력 이미지의 많은 추가 정보가 손실될 것이라는 점이다.
- [0017] 그로 인해 그러한 다운-샘플링은 그러한 고 선명도 이미지를 객체 검출용 입력에서 가지는 장점의 대부분을 무효화한다.
- [0018] 부가적으로, 완전한 이동성 환경에서 콘텐츠 검출 방법이 필요하고, 이는, 상기 방법이 객체와 카메라 사이의 움직임이 상대적으로 적은 요건 없이, 그리고 일정한 조명 조건에 대한 요건 없이, 일상적인 상황에서 사용할

수 있음을 의미한다

- [0019] 콘텐츠 검출에 대한 이들 과제를 복합화하는 것은 비디오 스트림의 시간적 (temporal) 데이터에서만 점점 더 보여지고 캡처될 필요가 있는 콘텐츠가 있다는 사실이다. 예를 들어 폭력 검출, 보행자 의도 검출, 보안 카메라의 실시간 제공 (live feed)에서 의심스러운 동작 검출 등이 있다. 이는 비디오 스트림의 둘 이상의 이미지 프레임이 단일 모델에서 상호 참조되어야 함을 의미한다. 현재 방법은 주로 정적 이미지에 관한 트레이닝을 기반으로 한다. 다시 말해, 비디오는 실제로 시간적 데이터를 프로세싱하는 대신, 정적 이미지의 시퀀스로 프로세싱된다. 단일 분류 모델을 프로세싱하기 위해 다수의 프레임을 상호 참조해야 할 때 복잡성과 프로세싱 오버헤드는 명확하게 부가될 것이다.
- [0020] 주어진 알고리즘을 통해 프로세싱 전력의 상대적 사용을 감소시키기 위해 프로세싱 전력의 지속적인 개선에 의존하는 것 역시 효과적이지 않은데, 이는 애플리케이션, 예를 들어, 게임이 최대 프로세싱 전력을 사용하도록 조정되어, 객체 검출과 같은 알고리즘이 배경에서 운영되도록 항상 최소한의 양만 남겨두기 때문이다.
- [0021] 상기를 고려하여, 객체 인식 및 객체 추적 기술을 위한 컴퓨터 비전 문헌에 많은 방법이 개시된다.
- [0022] 콘텐츠 검출에 있어서, 모든 공지된 방법은 디지털 이미지로부터 추출된 여러 종류의 특징을 차례로 사용하되, 이들 특징을 추출하기 위한 복수의 방법을 기반으로 하여, 사용한다. 디지털 이미지로부터 특징을 선택 및 추출하는 방법은 객체 검출 기술의 근원이다. 그러한 선택된 특징을 사용하기 위한 상이한 방법도 알려져 있다: 대부분의 경우, 그러한 특징은 분류기를 트레이닝하거나 분류기에서 알려진 특징의 세트 사이에서 상관되기 위해 사용된다.
- [0023] 객체 검출 방법에 대한 트레이닝 단계는 일반적으로 많은 시간이 걸리는데, 이는 합리적이고 알맞은 모델을 생성하기 위해 상당한 양의 데이터를 입력해야 하기 때문이다.
- [0024] 그러한 트레이닝된 모델을 사용하면, 객체 검출을 위한 일반적인 프로세스는 다음과 같이 요약될 수 있다: 입력 이미지, 특징 추출, 특징 분석 및 최종 객체 분류.
- [0025] 비올라-존스 (Viola-Jones)는 가장 일반적으로 사용되는 객체 검출 프레임워크 중 하나이다. 검출 프레임워크에 의해 사용되는 특징은 보편적으로 단순한 직사각형 구역 내의 이미지 픽셀의 합을 수반하여, 이는 일반적으로 예견된 계산을 약간 단순화하기 위해 그레이-스케일 필터를 통해 변환된다.
- [0026] 그러한 방법은 위크 분류기 (weak classifiers)라고 하고, 허용 가능한 수준의 정확도를 얻기 위해 다수의 특징 (수천개 정도) 또는 심지어 그러한 다수의 특징의 다수의 캐스케이드 (cascade)를 필요로 하며, 이러한 다수의 특징은 여러 검색 박스 및 계산을 다수의 반복의 사용을 수반한다.
- [0027] 비올라-존스와 같은 많은 특징을 사용하는 방법 및 "Adaptive Boosting"의 줄임말인 AdaBoost와 같은 트리형 캐스케이드 (tree-like cascades)를 사용하는 기술은 알고리즘 검출 견고성과 그러한 알고리즘에 필요한 계산 수 사이에 선형 상관 관계가 없기 때문에, 본질적으로 비-선형 성능을 가진다. 검출 성능에서의 각 단계마다 모델의 복잡성이 증가한다. 실시간 적용에서 그러한 알고리즘의 사용을 허용하기 위해, 종래의 디바이스에 의해 상기 알고리즘에 전용화될 수 있는 프로세싱 전력량이 일반적으로 제한되는 반면, 필요한 계산의 양이 알고리즘의 성능의 요구된 수준에 대해 예측될 수 없기 때문에, 그러한 기술은 모델로부터 솔루션 공간의 추출을 필요로 하여, 전체 모델을 사용하기 전에 빠른 검출 추정에 사용된다.
- [0028] 스케일-불변 특징 변환 (Scale-invariant feature transform)의 약자인 SIFT와 같은 다른 방법은 이미지에서 국부 특징의 모델을 검출 및 기술한다.
- [0029] 디지털 이미지에서의 임의의 객체에 대해 객체의 흥미로운 점을 추출하여 그 객체의 모델 기술을 제공할 수 있다.
- [0030] SIFT 특징은 국부적이며 특정 관심 점에서 객체의 외형을 기반으로 하며, 이미지 스케일 및 회전에 대해 변화하지 않는다. 또한 이들은 조명, 노이즈에서의 변화, 및 관점에서의 사소한 변화에도 강력하다. SIFT 특징 세트에 의한 객체 기술도 부분 폐색에 강력하다.
- [0031] 그러나, 그러한 기술은 계산적으로 매우 집중적일 수 있으며, 그러한 기술을 실시간으로 높은 프레임 속도로 작동하게 하려면 종종 특징의 수 및 기능의 복잡성을 제한하여 모델의 유효성을 저해하는 것이 요구된다.
- [0032] 인간 머리 방향을 검출하거나 인간의 얼굴에서 감정을 캡처하는데 종종 사용되는 다른 방법은 디지털화된 프레임으로부터 기하학적 점을 추출한다. 이들은 이들 점과 미리 결정된 2D 또는 3D 형상 모델 사이의 상관 관계

또는 대응 관계에 의존한다. 이들 방법의 예시는 ASM (Active Shape models) 및 AAM (Active Appearance models)로 알려져 있다.

- [0033] ASM 방법은 현재 모델 점 위치에 대한 프로파일을 따라 검색하여 객체 형상의 현재 추정치를 업데이트한다. 실제로, ASM 기술은 형상의 통계 모델에 의해 제한되는 한 세트의 모델 점을 이미지에 매칭시키려고 한다.
- [0034] 대신에, AAM 기술은 이미지 데이터를 샘플링하고 모델과 샘플 간의 차이를 사용하여 외형 모델 파라미터를 업데이트한다. ASM과의 차이점은 AAM이 모델 점의 위치 및 객체의 텍스처 표현 둘 다를 이미지에 매칭시키려고 한다는 점이다.
- [0035] 상기 언급된 방법의 특징은 일반적으로 공간 도메인 내에서 추출된다. 공간 도메인 이미지 정보 내에서 계산만을 사용하는 방법은 실질적인 속도 개선에 대한 범위가 적다.
- [0036] 종래 기술에 속하는 콘텐츠 검출을 위한 방법은 또한, 분류되고 인식될 객체의 차별적 특징을 우선적으로 강조하기 위해 주파수 도메인에서, 소위 필터링 커널, 간단하게 커널을 사용할 수 있다.
- [0037] 그 후, 커널을 사용한 후 추출된 특징은 변환된 이미지에서의 픽셀 합 또는 주파수 도메인에서의 셀 합의 결과일 수 있다.
- [0038] 그러한 방법의 대부분은 주파수 도메인에 균등하게 분배된 사전-정의된 일련의 커널을 사용한다. 이러한 기술의 예시는 가버 (Gabor) 뱅크의 사용이다. 각 커널 설정은 커널에 의해 변환된 상이한 이미지를 공간 도메인에서 초대한다. 그러한 방법으로 추출된 특징은 통상적으로 변환된 이미지 내에서 가장 높은 픽셀 세기 값의 좌표이다. 통상적으로, 약 40 개의 상이한 가버 필터가 그러한 방법에 사용된다.
- [0039] 전체 주파수 도메인 계산이 수행되고, 변환이 공간 도메인으로부터 주파수 도메인으로, 그 후에 반대로 다시 주파수 도메인으로부터 공간 도메인으로 매회 필요하기 때문에, 그러한 방법은 또한 다수의 계산을 필요로 한다.
- [0040] 추가로, 그러한 방법은 종종 계산 속도를 높이는 푸리에 변환의 고속 버전 인 FFT의 변형 중 하나 (DCT는 이러한 것의 개별 버전임)를 사용한다. FFT는 입력 이미지를 재-샘플링해야 하는데, 이는 입력에서 이미지의 픽셀 폭 및 픽셀 높이 둘 다가 2의 거듭 제곱 크기여야하기 때문이다.
- [0041] 많은 객체 인식 방법에서, 추출된 특징이 차별되어야 하고 특정 객체에 대해 인식될 수 있게 하는 추가 분류 단계가 있다. 이러한 양태는 주로 개인이 인식되는 인간 얼굴 인식에 사용된다. 객체 인식에 대한 방법의 예시는 아이겐페이스 (eigenfaces)를 사용한 Principal Component Analysis, Linear Discriminate Analysis, 및 Fisherface 알고리즘을 사용한 Elastic Bunch Graph Matching이다.
- [0042] 상기에서 언급된 것과 유사한 객체 검출용 방법은 객체 추적용 방법과 조합될 수 있다. 객체 추적에서는 스크린 상에서 객체의 움직임 역학이 따른다. 이를 위해 일반적으로 사용되는 방법은 예를 들어 Mean-Shift 알고리즘 및 Template Matching이다. 그러한 알고리즘은 실시간 사용에 대해 우수한 성능을 가진다. 그러나, 그러한 방법은 일반적인 객체 모델을 발생시키지 않고, 대신에 객체의 작은 하위 샘플을 학습시킨다.
- [0043] 상기에서 언급한 바와 같이, 일반적으로 사용되는 콘텐츠 검출 방법은 이미 이미지에서의 통상적인 객체, 예를 들어 카메라 앞에서 포즈를 취하는 사람의 얼굴 검출에 대해 우수한 검출 속도를 달성한다. 그러나, 실제 사용에서는 다수의 조명 조건, 검출될 객체의 더 많은 예기치 않은 버전 및 또한 종종 단지 인식될 물체의 부분적인 가시성이 있을 수 있기 때문에, 검출 업무의 난이도가 크게 증가함으로 인해, 일반적인 디바이스의 실시간 적용에서의 이들 사용은 여전히 제한적이다. 이들 가능한 모든 우발 사건 최종 검출을 달성하기 위해 요청된 방법 검출 트레이닝의 복잡성 및 또한 필요한 견고성을 달성하기 위해 그러한 방법에 필요한 계산의 수로 인해, 일반적으로 사용되는 많은 객체 검출 방법의 사용을 제한한다.
- [0044] 상기는 객체 검출을 위한 이미지의 오프라인 프로세싱에 대한 제한이 아니며, 이 경우 검출을 위한 경과 시간이 주요 제약은 아니다. 그러나, 실시간 적용 에서, 객체 검출은 적어도 비디오 입력 프레임 속도에서 작동해야 한다.
- [0045] 요약하면, 상기의 종래 기술로 기술된 방법은 제약이 없는 실제 실시간 적용에서 여전히 대규모로 사용되지 않는데, 이는 그러한 방법에 의한 현재 프로세싱 전력으로 인해 객체 검출의 견고성 및 속도 사이에서 수용가능한 타협을 달성하는 것이 어렵기 때문이다. 또한, 중요하게, 종래 기술에서 기술된 방법은, 동적 비디오 스트림인 프레임 시퀀스를 상호 참조할 때만 발견될 수 있는 시간적 데이터의 프로세싱이 아니라, 스틸 이미지의 프로세싱을 위해 주로 설계되었다.

[0046] US 2009/0238466은 얼굴 식별에 초점을 둔 객체 인식 방법을 개시하며, 정적 이미지의 예지는 픽셀 카테고리화 및 적합성 테스트 (congruency test)를 통해 결정된다.

**발명의 내용**

**해결하려는 과제**

[0047] 콘텐츠 검출 및 이미지 프로세싱을 위한 본 방법은 데이터가 주파수 도메인으로 변환될 수 있는 경우에 적용될 수 있다. 이미지 데이터를 주파수 도메인으로 변환시킴 없이, 공간 도메인에서 픽셀 정보의 직접적인 분석을 기반으로 하는 방법에는 적용할 수 없다.

[0048] 일반적으로, 본 발명의 방법은 다음과 같이 요약될 수 있다: 다수의 정규화 된 복소 벡터를 가진 모델은 주파수 도메인으로 변환된 희소 데이터 (sparse data)의 표현을 포함한다. 이들 정규화된 벡터는 다수의 이미지 프레임 시퀀스로부터 주파수 도메인에서 추출된 특징으로부터 생성될 수 있으며, 상기 방법을 비디오 스트림에서 시간적 데이터를 캡처해야 하는 모델에 대해 특히 효과적이게 한다.

[0049] 이 과정에서, 다수의 특징이 먼저 선택되고, 이후에 오프라인 학습 단계에서 타겟 주파수와 같은 각 특징에 대한 파라미터, 변환 내의 기본 전달 함수 및 주파수 변환의 방향 및 시퀀스가 최적화되고, 최상의 파라미터 설정이 선택된다. 이들 특징은 다수의 프레임에 걸쳐 펼쳐지게 되어, 비디오 스트림에서 시간적 프로세스를 캡처할 수 있다. 특징 세트는 주파수 도메인에서 정규화된 복소 벡터의 단일 공간으로 조합된다. 이러한 정규화된 복소 벡터 세트는 이로써 비디오 또는 비디오 스트림에서 검출되는 콘텐츠를 위해 주파수 도메인에서 모델을 기술한다.

**과제의 해결 수단**

[0050] 보다 자세하게, 주파수 도메인에서 특징을 추출하고 사용하는 방법은 다음 단계를 포함한다:

[0051] ● 공간 도메인에서의 데이터를 통해 정의된, 비디오 스트림에서 디지털 이미지의 시퀀스, 또는 디지털 이미지를 얻는 단계;

[0052] ● 주파수 도메인으로의 변환을 위해, 주파수 도메인의 일 부분을 커버하고, 가능하게는 비디오 스트림의 시퀀스에서 2 개 이상의 프레임을 커버하는 하나 이상의 쌍의 희소 존을 선택하는 단계;

[0053] ● L-변환의 2D 변형을 사용하여 이미지 데이터의 총 주파수 도메인 데이터의 희소 부분을 주파수 도메인으로 전송하는 단계;

[0054] ● 각 존의 각 입력에 대해 2D L-변환에 대한 타겟 주파수, 전달 함수 및 입력의 형상 및 방향의 최적화를 적용하는 단계;

[0055] ● 각 쌍의 존을 정규화된 복소 벡터인 특징으로 변환하는 단계;

[0056] ● 생성된 모든 벡터를 모든 정규화된 복소 벡터를 포함하는 단일 모델로 함께 조합하는 단계; 및

[0057] ● 분류기에 대한 입력 데이터를 생성하기 위해 정규화된 복소 벡터의 모델을 사용하는 단계.

[0058] 객체 검출 분야의 전문가에게는 디지털 이미지 입력으로 사용할 수 있는 여러 가능성이 있음이 분명할 것이다:

[0059] ● 전체 이미지가 변환된다.

[0060] ● 이미지의 세그먼트가 변환된다.

[0061] 디지털 입력의 크기가 여기에 개시된 청구 범위를 변경하지 않는 것이 또한 명백할 것이다.

[0062] 상기를 고려하여, 본 발명에 따른 이미지 프로세싱 및 콘텐츠 검출에 대한 방법은 첨부된 청구항 1에서 정의된다.

[0063] 부가적인 장점으로 이어지는 방법의 추가 세부 사항은 종속 청구항에서 정의된다.

[0064] 따라서, 본 방법은 콘텐츠 및 이미지 프로세싱 방법의 검출의 프로세싱 속도를 크게 증가시킴과 동시에 그의 정확도를 증가시킬 수 있다. 본 방법은 시간적 모델의 유효성을 개선시키는 데 특히 효과적이다.

[0065] 위에서 설명된 바와 같이, 이미지 프로세싱에서 콘텐츠 검출 기술의 속도를 증가시킬 수 있는 새로운 방법에 대

한 요구가 매우 크다. 이러한 속도 증가는 최신 세대의 모바일 디바이스 및 다른 유사한 디바이스 내에서 백그라운드에서 작동하기에 충분히 빠르게 콘텐츠를 검출해야 한다.

- [0066] 웨어러블 하드웨어와 같은 미래 추세를 예상하여 최신 세대의 모바일 디바이스에서 사용하는 것 외에도, 알고리즘은 주요 개인용 컴퓨터 운영 체제 및 모바일 운영 체제 외부에서 작동할 수 있어야 하므로 프로세서 상에서, 그리고 재-프로그래밍이 가능한 하드웨어, 그 예로 필드-프로그래밍 가능한 게이트 어레이 상에서 프로그램이 가능하다. 또한, 방법은, 멀티-코어 및 강력한 그래픽 프로세싱 유닛 (GPU)과 같은 개인용 컴퓨터 및 모바일 디바이스 상에서 최신 하드웨어 개발을 활용할 수 있는 알고리즘으로 구성되어야 한다.
- [0067] 이와 관련하여, 주파수 도메인에서의 계산의 사용은 자연스럽게 여러 이유로 더 빠른 계산의 필요성에 응답하기에 적합하다. 이 방법은 그러한 병렬 프로세싱을 허용한다. 이 분야의 전문가들은 많은 FFT 변형이 프로세서 상에서 프로그래밍에 적합하지 않다는 것을 잘 알고 있다.
- [0068] 기술된 방법이 비디오 스트림에서 동적 및 시간적 프레임 시퀀스에서만 발견 될 수 있는 모델을 캡처할 필요가 있는 시간적 모델이 생성되어야 할 때 특히 유리한 방법을 강조하는 것이 중요하다. 다시 말해서, 비디오 스트림일 때 캡처될 수 없는 콘텐츠는 스틸 프레임의 시퀀스로만 간주된다.
- [0069] 푸리에의 표준 주파수 도메인 이론은 이 경우 디지털 이미지에서 임의의 신호가 일련의 정현파의 합으로 표현될 수 있다고 말한다. 이미지 프로세싱의 경우에서, 이는 이미지에 걸친 정현파의 밝기 변화이다.
- [0070] 정현파 함수는 다음을 인코딩할 수 있다:
- [0071] ● 공간 주파수
- [0072] ● 크기 (magnitude)
- [0073] ● 위상
- [0074] 공간 주파수는 밝기가 변조되는 공간에 걸친 주파수이다.
- [0075] 정현파의 크기는 그의 대비에, 또는 이미지의 가장 어두운 피크와 가장 밝은 피크 사이의 차이에 대응된다. 위상은 그 파가 원점을 기준으로 어떻게 이동되는지를 나타낸다.
- [0076] 푸리에 변환은 단일 정현파뿐만 아니라 0에서 "나이퀴스트 주파수 (Nyquist frequency)"까지의 공간 주파수 범위를 통한 전체 일련의 정현파를 인코딩하고, 이는 해상도 또는 총 픽셀 수에 관련된 디지털 이미지로 인코딩될 수 있는 가장 높은 공간 주파수를 의미한다.
- [0077] 푸리에 변환은 이미지에 존재하는 모든 공간 주파수를 동시에 인코딩한다.
- [0078] 나이퀴스트 주파수는 이 경우 디지털 이미지에서, 이산 신호 프로세싱 시스템의 샘플링 속도의 1/2이다.
- [0079] 이미지 프로세싱에 사용되는 푸리에 변환 이면의 기본 원리는: 각 픽셀이 모든 주파수에 영향을 미치고 각 주파수가 모든 픽셀에 영향을 미친다는 점이다.
- [0080] 여기에 개시된 방법 이면의 직관은 공간 도메인에서의 상기 대비의 위치가 주파수 도메인에서 정현파에 의해 인코딩된다는 것이다. 공간 도메인에서의 소정의 대비는 주파수 도메인에서 소정의 정현파에 의해서만 영향을 받는다. 공간 도메인에서 대비의 조합으로 정의되는 형상의 경우, 이는 이 형상이 주파수 도메인에서 위치 정보의 고유한 조합으로 캡처됨을 의미한다. 그리고 이는 주파수 도메인에서 위치 정보의 고유한 변경에 의해 움직임이 캡처됨을 의미한다.
- [0081] 그러므로, 특정 객체의 검출을 위해 주파수 도메인에서 정보의 변경 또는 특정 정보의 캡처를 사용할 수 있다.
- [0082] 주파수 도메인에서 작업하면, 전달 함수를 사용한 곱셈 및 행렬을 사용한 다른 유사한 계산이, 계산적으로 더 비싼 두 함수 사이의 컨볼루션인 공간 도메인과 달리, 간단한 성분별 곱셈 (component-wise multiplication)이기 때문에, 계산이 매우 쉬어진다. 그러므로, 주파수 도메인은 객체 검출을 위해 주파수 도메인 정보의 회소하고 작은 부분의 계산적으로 용이한 사용을 허용한다.
- [0083] 그러므로, 주파수 도메인은 객체 검출을 위해 주파수 도메인 정보의 회소하고 작은 부분의 계산적으로 용이한 사용을 허용한다.
- [0084] 주파수 도메인에서 전달 함수, 특징 및 포맷뿐만 아니라, 회소 준은 이하에서 설명되고 상세하게 기술될 것이다.

[0085] 주파수 도메인 계산 이후 이미지를 다시 공간 도메인으로 변환하기 위한 계산을 요구함 없이, 주파수 도메인에 서 완전히 작업하면, 주파수 도메인으로의 변환을 수행하는 수학 선택에 있어 유연성이 역시 부가된다.

**도면의 간단한 설명**

[0086] 본 방법의 추가 특징 및 장점은 첨부 도면을 참조하여 주어진 바람직한 실시 예에 대한 다음의 설명에 의해 보 다 명백해질 것이며, 여기에서:

- 도 1은, 이미지에 대한 공간 도메인에서, 인덱스가 위치를 나타내는 반면 이미지 내의 주파수가 픽셀 세기에 서의 정현파 변화를 나타내며 주파수 도메인 내에서는 정반대가 되는 방법을 보여준다;
- 도 2는 주어진 객체에 대한 공간 도메인에서의 움직임 정보가 주파수 도메인에서의 위치를 코딩하는 파동의 변화에 의해 어떻게 캡처될 것인지를 보여준다;
- 도 3은 주파수 도메인에서 이미지 내의 객체를 분류하기에 충분한 위치 및 형상 정보를 캡처하는데 다수의 파동이 필요한 방법을 보여준다;
- 도 4는 주파수 도메인에서 정현파 정보를 캡처하기 위해 주파수 도메인 정보의 희소 부분만이 필요한 방법을 보여준다;
- 도 5 및 도 6은 사실상 괴첼 (Göertzel)의 전체 2D 구현이 많은 1D 괴첼 계산의 조합일 수 있는 방법 을 보여준다;
- 도 7은 전체 2D 구현이 계산상 이상적이지 않을 방법을 보여준다;
- 도 8은 여전히 이상적이지 않더라도, 전체 2D 구현이 더 빨라질 수 있는 방법을 보여준다;
- 도 9는 인덱스에서 행 및 열에 대한 1D 출력을 별도로 계산하여 그 후 이를 단일 값으로 조합하여 인덱스 결 과를 얻는 경우가 어떻게 효율적인지를 보여준다;
- 도 10은 계산 속도를 높이는 방법을 보여준다; 주파수 도메인으로의 변환을 위한 입력 셀은 값이 필요한 인 텍스의 위치 주위에서만 취해진다;
- 도 11은 준에 대한 입력이 입력으로서 행 및 열에만 제한되지 않고 임의의 자유 형상일 수 있는 방법을 보여 준다;
- 도 12는 임의의 자유 형상일 수 있는 두 개의 입력을 각 준이 가지는 방법을 보여준다;
- 도 13은 두 준이 특징을 형성하는 방법을 보여주고, 상기 특징은 정규화를 통해 두 준 간에 정보 이동을 제 공하는 복소 벡터가 된다;
- 도 14는 각 인덱스가, 타겟 주파수, 변환을 위한 기본 전달 함수 (underlying transfer function) 및 주파 수 도메인 변환의 입력의 방향 및 시퀀스를 정의하는 특정 특징 형상을 가지는 방법을 보여준다;
- 도 15는 주파수 도메인 변환을 위한 입력의 형상이 입력의 각 인덱스를 순차적으로 이동하여 최적화될 수 있 는 방법을 보여준다;
- 도 16은 입력 이미지가 정사각형 및 사각형에 제한되지 않는 임의의 형상일 수 있는 방법을 보여준다;
- 도 17은 준 및 특징이 비디오 스트림에서 프레임 시퀀스에서의 데이터를 상호 참조할 수 있는 방법을 보여준 다;
- 도 18은 각 인덱스의 주파수 도메인 변환 시퀀스를 정의하는 형상이 단일 프레임으로 제한되지 않고, 비디오 스트림에서 다수의 프레임을 상호 참조할 수 있는 것을 보여준다;
- 도 19는 방법이 모델의 주파수 도메인 변환을 생성하는데 사용되는 프레임 시퀀스를 선택할 수 있어 콘텐츠 검출에 실제 시간 데이터가 사용될 수 있도록 하는 자유도 (freedom)를 보여준다;
- 도 20은 주파수 도메인에서의 최적화가 매우 큰 다-차원 노이즈 구역에서 명확한 신호가 되는 것의 매우 큰 이점을 가져, 감독뿐만 아니라 비감독 학습의 현재 타입의 논리를 허용하는 방법을 보여준다;
- 도 21은 기술된 방법에 대한 가능한 최적화 논리의 예시인 흐름도를 보여준다.

**발명을 실시하기 위한 구체적인 내용**

- [0087] 이하에서, 본 발명에 따른 방법의 실시예가 첨부 도면을 참조하여 상세히 기술될 것이다.
- [0088] 명백한 바와 같이, 비디오 스트림, 즉 소정의 비율을 가진 일련의 이미지 프레임의 연속과 관련하여 여기에 기술된 것은 또한 비디오 스트림의 프레임과 동등한 단일 이미지의 임의의 연속에, 그리고 단일 프레임과 동등한 하나의 단일 이미지에 적용된다.
- [0089] 본 발명에 따른 방법의 목적은 식별될 콘텐츠를 나타내는 주파수 도메인 신호를 강화하고 그 후에 격리시키고, 이미지 또는 비디오 스트림 내의 콘텐츠에 대한 주파수 도메인 노이즈를 감소 또는 무시하는 것이다.
- [0090] 첫 번째 단계에서, 검색 로직이 전체 입력 이미지 상에서 사용되어, 이 방법의 계산을 위한 입력 프레임을 발생시킬 수 있다. 검색 로직은 예를 들어 전체 이미지 또는 이미지의 서브세트일 수 있다. 많은 타입의 검색 로직이 가능하지만, 여기에 개시된 방법의 관점으로부터 계산 또는 청구 범위는 변하지 않으며, 단지 변환을 위한 이미지 입력이 변한다는 점이 명백해야 한다. 또한, 단일 프레임이 다수의 계산을 위한 다수의 입력을 가질 수 있음이 명백할 것이며, 상기 다수의 계산 각각은 청구 범위에 기술된 바와 같이 프로세싱된다.
- [0091] 그 후에, 프레임 또는 프레임 시퀀스로부터 취해진 입력 또는 입력들은 주파수 도메인에서 추출된다. 프레임의 주파수 도메인 데이터 내의 데이터는 그 후 콘텐츠를 검출하기 위해 프로세싱된다. 여기에 기술된 방법은 사용된 분류를 열어두고, 대신 이 방법에서 강조된 것은 선택 분류에 사용되는 데이터의 품질 및 타입의 개선이다.
- [0092] 상술된 바와 같이, 기술된 방법은 비디오 스트림의 시간 데이터에 특별하게 또는 심지어 독점적으로 존재하는 객체 검출 및 프로세스에 특히 효과적이다. 검출을 위해 다수의 프레임이 단일 입력으로 조합될 수 있는 방법이 여기에서 기술될 것이다.
- [0093] 검출은, 객체가 검출되었거나 검색 로직의 반복이 이미지 내에서 객체를 찾지 못할 때 비디오 스트림의 다음 프레임 또는 프레임 시퀀스로 이동할 것이다. 분명해야 하는 바와 같이, 검색 로직은 적응적으로 만들어질 수 있고, 이를 기반으로 하여 객체는 사전 프레임에서 발견된다.
- [0094] 공지된 기술에서, 주파수 도메인에서의 이미지 프로세싱은 일반적으로 고속 푸리에 변환 (FFT)의 변형을 사용하여 행해지지만, 본 방법은 FFT 또는 그 변형, 예를 들어 이산 코사인 변환 (DCT)도 사용하지 않고, 이산 푸리에 변환 (DFT)도 사용하지 않는다.
- [0095] 그러나, 종래의 이미지 프로세싱과 본 방법 사이의 차이점을 강조하기 위해, FFT 및 DFT의 일반적인 개요가 여기에 제공된다.
- [0096] FFT는 이미지 분석, 이미지 재구성 및 이미지 압축, 텍스트 인식 등과 같은 광범위한 응용에 사용된다.
- [0097] FFT의 주요 원리는 이산 푸리에 변환 (DFT)에서 따른다. DFT에는 많은 계산이 필요하기 때문에, 프로세스 속도를 높이려는 다른 타입의 변환이 있다. 고속 푸리에 변환 (FFT)이 이들 중 가장 확립된 것이다. DFT를 사용하면, 계산 횟수는  $N^2$ 와 상관되며, 여기서  $N$ 은 입력 매트릭스의 길이이다.
- [0098] FFT 알고리즘은 표준 DFT가 많은 중복 계산 (redundant calculations)을 수반한다는 사실에 의존한다.
- [0099] FFT는 샘플 시퀀스를 서브-시퀀스로 나눔으로써 계산된다; 각 단계에서 사전 단계의 결과를 조합하기 위해  $N/2$  복소수 곱셈이 필요하다.
- [0100]  $\log(N)$  단계가 있기 때문에, FFT를 사용하여  $N$ -포인트 DFT를 평가하는데 필요한 복소수 곱셈의 수는 대략  $N \cdot \log(N)$ 이다.
- [0101] 주파수의 수는 공간 도메인 이미지에서의 픽셀 수에 대응하고, 즉 공간 및 주파수 도메인에서의 이미지는 동일한 크기를 가진다.
- [0102] 상술된 바와 같이, FFT에는 다수의 변형이 있다. FFT는 또한 이미지 프로세싱에 제한을 가진다. 예를 들어, FFT에 대한 입력에 사용된 이미지의 측면은 2의 거듭제곱인 길이를 픽셀 단위로 가져야 한다.
- [0103] 또 다른 제한은 소정의 주파수에 대한 결과가 제공될 수 있기 전에 전체 FFT가 계산되어야 한다는 것이다. 다시 말해, FFT는 희소 계산 (sparse calculations)에 대해 변환될 수 없는데, 이는 단일 주파수에 대한 값이 얻어질 수 있기 전에, 전체 FFT가 계산되어야 하기 때문이다. FFT의 복잡한 구조는 또한 재-프로그래밍 가능한

하드웨어 및 멀티-코어 프로세서 상에서 쉬운 코딩 구현을 허용하지 않는다. 부가적으로, 단일 결과를 얻기 위해서 전체 FFT가 먼저 계산되어야 하기 때문에, 이는 또한 디바이스 상에서 더 많은 메모리 사용을 필요로 한다.

- [0104] 제거된 고속 푸리에 변환과 같은 방법이 제공될 수 있지만, 이들은 속도 및 메모리 사용의 상대적으로 작은 개인을 위해 다량의 복잡한 코드를 필요로 하면서, 여전히 재-프로그래밍 가능한 하드웨어 상에서 구현하기 여전히 어렵다.
- [0105] 공간 도메인에서, 값은 일반적으로 0 내지 255 범위의 픽셀 광 세기이다. 동일한 이미지의 푸리에 도메인 값은 공간 도메인에서의 이미지보다 훨씬 큰 범위를 가진다.
- [0106] 푸리에 변환은 복소수 값의 출력 이미지를 생성하고, 이는 실수 및 허수 부분 또는 크기 및 위상을 갖는 2 개의 이미지로 디스플레이될 수 있다. 이미지 프로세싱에서, 푸리에 변환의 크기만이 종종 디스플레이되는데, 이는 푸리에 변환의 크기가 공간 도메인 이미지의 기하학적 구조에 대한 정보 대부분을 포함하고 있기 때문이다. 그러나, 주파수 도메인에서의 일부 프로세싱 이후에 푸리에 이미지를 정확한 공간 도메인으로 재-변환하기 위해서는, 푸리에 이미지의 크기 및 위상 둘 다를 보존해야 한다.
- [0107] 본 개시에 따른 방법에서, 모든 계산이 주파수 도메인에서의 정보를 독점적으로 사용하는 것이 가능하다.
- [0108] 공간 도메인으로 되돌아가기 위한 모든 정보를 유지할 필요가 없기 때문에 여러 장점이 있다.
- [0109] 첫째, 주파수 도메인으로부터 공간 도메인으로 돌아가는 추가 단계가 없기 때문에 전체 계산 속도가 빨라진다.
- [0110] 둘째, 주파수 도메인 데이터가 정확한 공간 도메인 이미지로 다시 변환될 필요가 없기 때문에, 더 작은 희소 존이 사용될 수 있다. 이는 이미지 품질 및 정보의 큰 손실 없이 이미지 데이터가 공간 도메인으로 다시 변환될 수 있도록 하는 주파수 도메인 데이터를 가질 필요가 없기 때문이다. 주파수 도메인에서 희소 존 자체는 공간 이미지를 재생성하기에 충분한 주파수 도메인 정보를 반드시 포함할 필요는 없다. 그러나 이들은 분류에 충분한 정보를 포함한다.
- [0111] 셋째, 추가 계산은 FFT에서 공통적인 엘리머싱을 제거하기 위해 수행될 수 있고, 또한 다른 계산도 주파수 도메인 내에서 분류에 대한 데이터를 더 잘 준비하기 위해 수행될 수 있다.
- [0112] 넷째, FFT 및 DCT와 같은 방법에 존재하는 다른 제한이 제거된다. 예를 들어, FFT 및 DCT에 대한 주파수 변환은 이미지의 행 및 열을 따라, 그리고 항상 단일 프레임 내에서 행해진다. 이 방법에서, 주파수 도메인 변환의 방향은 임의의 순열일 수 있으며, 전달 함수에 대한 훨씬 더 많은 자유도 및 변환이 비디오 시퀀스의 프레임들 사이에서도 교차할 수 있다.
- [0113] 도 1에서는 이미지에 대한 공간 도메인에서 인덱스가 위치를 나타내는 반면 이미지 내의 주파수가 픽셀 세기에서의 정현파 변화를 나타내는 방법을 나타낸다. 주파수 도메인 내에서는 정반대가 되며, 인덱스는 주파수를 나타내는 반면, 정현파는 위치 데이터를 포함한다.
- [0114] 동일한 방식으로, 주어진 객체에 대한 공간 도메인에서의 움직임 정보는 주파수 도메인에서의 위치를 코딩하는 파동에서의 변화에 의해 캡처될 것이다. 이는 도 2에서 눈 움직임의 이미지와 함께 개략적으로 도시된다.
- [0115] 상기에서 주어진 예시는 물론 개념을 설명하기 위해 단순화되었다. 실제로, 공간 도메인에서 픽셀 세기에서의 변화를 캡처하는 많은 주파수가 이미지를 그리는 데 필요한 것과 동일한 방식으로, 주파수 도메인에서 이미지 내의 객체를 분류하기에 충분한 위치 및 형상 정보를 캡처하는 데는 다수의 파동이 필요하다. 이는 도 3 내에 나타나 있다.
- [0116] 상술된 바와 같이, 주파수 도메인 내의 각 인덱스는 공간 도메인에서의 모든 픽셀에 잠재적으로 영향을 미친다. 그러므로, 공간 도메인으로부터 추출된 특징으로 객체를 분류하는 것에 비해, 주파수 도메인에서 객체를 분류하는 데는 비교적 적은 특징이 요구된다. 여기에 기술된 객체 검색 방법에서, 공간 도메인에서 특정 타입의 형상을 캡처하는 주파수 도메인에서 정보의 최소 부분 조합을 찾는 기술이 개시된다. 이러한 정보의 조합은 특히 또한 시퀀스의 각 스틸 이미지를 자체적으로 고려할 때 발견되지 않는 시간적 및 동적 정보를 캡처하는 것을 목표로 하여, 비디오 스트림에서 프레임 시퀀스일 수 있다.
- [0117] 여기 및 다음의 설명에서, 희소 존은 주파수 도메인의 일부분을 커버하는 정보의 선택인 것으로 의미된다. 각 존은 특정 주파수 도메인 정보로서 보여야 한다. 도 4는 주파수 도메인에서 희소 특징의 레이아웃의 예시를 도시한다.

- [0118] 각 특징이 한 쌍의 존으로부터 생성되고, 주파수 도메인 그리드의 크기가 단지 설명을 위한 것이며 나중에 설명될 다른 많은 크기일 수 있음에 유의하여야 한다.
- [0119] 도 4는 공간 도메인에서 위치 및 움직임을 캡처하는 가능한 주파수 도메인 정현파 대비의 예시를 오버레이로도 도시된 것을 보여준다. 도 4가 보여주는 것은, 회소 존이 행하는 것인 공간 도메인에서 움직임 또는 형상의 타입을 검출하기 위해 주파수 도메인 정현파 대비의 일부만 캡처될 필요가 있다는 것이다.
- [0120] 회소 존은 국부적인 해상도를 증가시키기 위해 가능한 서로 부분적으로 겹쳐지거나, 나란히 위치되어 함께 그룹화될 수 있다.
- [0121] 상기 회소 존으로부터 도출된 주파수 값에 대한 계산은 회소 계산으로 나타난다.
- [0122] 이미지를 다시 공간 도메인으로 변환할 필요가 없고, 모든 주파수 도메인 정보가 필요한 것은 아니기 때문에, 이미지를 공간 도메인으로 변환하기 위해 DFT 또는 FFT 이외의 다른 방법을 사용할 가능성도 열려 있다.
- [0123] 본 방법에 따르면, 단일 프레임의 적어도 일부 또는 프레임 시퀀스의 경우 적어도 2 개의 프레임의 시퀀스를 각각 커버하는 하나 이상 쌍의 회소 존이 선택된다.
- [0124] 상기에서 언급된 바와 같이, 각 쌍의 회소 존은 특징 (feature)을 발생시키고, 각 회소 존은 2 개의 공간 데이터 시퀀스에 의해 정의된다.
- [0125] 그 후에, 본 발명에 따르면, 상기 선택된 특징은, 각 회소 존에 대해, L-변환의 2D 변형을 통해 상기 2 개의 공간 데이터 시퀀스를 조합시키고, 각 존에 대한 전달 함수, 형상 및 주파수 도메인 데이터의 방향을 변화시키고, 이로써 상기 특징 각각에 대해 정규화된 복소 벡터를 발생시킴으로써, 주파수 도메인 데이터로 변환된다.
- [0126] 따라서, 변환은 타겟 주파수에 대한 상당한 설계 자유도, 변환에 사용되는 전달 함수, 변환의 입력을 정의하는 루프의 형상 및 방향과 함께, 피켈 알고리즘으로부터 파생된 2-차원 변환과 같은 추가 방법을 사용하여 수행될 수 있다. 본 개시에서 추가로 설명되는 바와 같이, 사용된 방법은 피켈 알고리즘과 매우 상이하므로, 사용된 설명은 L-변환의 2D 변형이라는 것이다.
- [0127] 이전에서 언급된 바와 같이, 이 방법의 장점은 최소한의 메모리를 사용하면서 재-프로그래밍 가능한 프로세서 상에서 또는 GPU 상에서 사용을 위해 보다 유연한 방식으로, 병행하여, 회소적으로 설정될 수 있다는 것이다. 다음에서는, 피켈 변환 뒤의 이론이 먼저 기술된다. 그 이후에, 이미지 프로세싱에서 2D 경우를 확장하고 사용될 수 있는 다양한 설계 옵션으로, 이 방법에 대한 구현이 상세하게 설명된다.
- [0128] 단일 정현파 톤 (tones)의 검출 및 측정에서 스펙트럼 분석이 수행되어야 할 때, 무한 임펄스 응답 (IIR) 필터 구조가 사용된다.
- [0129] 스펙트럼 에너지에 대한 표준 방법은 이산 푸리에 변환 (DFT)이며, 통상적으로 고속 푸리에 변환 (FFT) 또는 이산 코사인 변환 (DCT) 알고리즘을 사용하여 구현된다.
- [0130] 그러나, N-포인트 DFT의 N-빈 중심 주파수의 서브세트에 대해서만 스펙트럼 분석이 필요한 응용이 있다. 1D에서 회소한 FFT 결과를 계산하는데 널리 사용되는 효율적인 기술은 피켈 알고리즘으로, IIR 필터 구현을 사용하여 N 입력 시간 샘플을 기반으로 단일의 복잡한 DFT 스펙트럼 빈 값을 계산한다.
- [0131] 이 프로세스의 가장 일반적인 응용은 단일 연속파 정현파 톤의 존재를 검출하는 것이다. 피켈 알고리즘은 1D 계산이므로 이미지가 2D인 이미지 프로세싱에는 사용됨을 의미하지 않는다.
- [0132] 피켈 알고리즘은 길이가 N인 신호  $\{x[n]\}$ 의 k 성분을 계산하는 아이디어를 기반으로 한다.

$$X[k] = \sum_{n=0}^{N-1} x[n]e^{-j2\pi k \frac{n}{N}}$$

[0133] (식 1)

[0134] 이 식 (1) 우측에  $e^{j2\pi k \frac{N}{N}} = 1$ 을 곱하면 다음과 같다:

$$X[k] = e^{j2\pi k \frac{N}{N}} \sum_{n=0}^{N-1} x[n] e^{-j2\pi k \frac{n}{N}}$$

[0135] (식 2)

[0136] 이는 다음과 같이 기재될 수 있다:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi k \frac{n-N}{N}}$$

[0137] (식 3)

[0138] (3)의 우측은 신호  $\{x[n]\}$  및  $\{h_k[n]\}$ 의 개별 선형 컨볼루션 (discrete linear convolution)으로 볼 수 있고, 이 경우에:

$$h_k[l] = e^{j2\pi k \frac{l}{N}} u[l]$$

[0139] 이다.

[0140] 실제로,  $\{y_k[n]\}$ 이 그 컨볼루션의 결과를 나타내는 경우 다음과 같다:

$$y_k[m] = \sum_{n=-\infty}^{\infty} x[n] h_k[m-n]$$

[0141] (식 4)

[0142] 이는 다음과 같이 기재될 수 있다:

$$y_k[m] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi k \frac{m-n}{N}} u[m-n]$$

[0143] (식 5)

[0144] 컨볼루션은 하나가 5 반전 및 이동된 후 두 함수의 곱의 적분으로 정의된다. 그러한 바와 같이, 이는 특정 종류의 적분 변환이다.

[0145] 컨볼루션 정리는 적합한 조건 하에 컨볼루션의 푸리에 변환이 푸리에 변환의 점별 곱 (pointwise product)이라고 설명한다. 다시 말해, 한 도메인 (예를 들어, 시간 도메인)에서의 컨볼루션은 다른 도메인 (예를 들어, 주파수 도메인)에서의 점별 곱셈과 같다.

[0146] (3)과 (5)를 비교하면, 원하는  $X[k]$ 가 컨볼루션의 N 번째 샘플이 됨은 분명하다:

$$X[k] = y_k[N]$$

[0147] (식 6)

[0148] 이 경우에  $k=0, \dots, N-1$ 이다. 이는 임펄스 응답  $\{h_k[n]\}$ 을 갖는 IIR 선형 시스템의 시간 N에서 출력 샘플로 필요한 값이 얻어질 수 있음을 의미한다.

[0149] 이 시스템의 전달 함수  $H_k(z)$ 가 이제 도출될 것이다; 이는 임펄스 응답의 L-변환이다:

$$H_k(z) = \sum_{n=-\infty}^{\infty} h_k[n] z^{-n}$$

[0150] (식 7)

$$= \sum_{n=-\infty}^{\infty} e^{j2\pi k \frac{n}{N}} u[n] z^{-n} \quad (식 8)$$

$$= \sum_{n=0}^{\infty} e^{j2\pi k \frac{n}{N}} z^{-n} \quad (식 9)$$

$$= \sum_0^{\infty} (e^{j2\pi k \frac{1}{N}} z^{-1})^n \quad (식 10)$$

등비 수열은 수렴형이며 그 합은 전달 함수와 같다:

$$H_k(z) = \frac{1}{1 - e^{j\frac{2\pi k}{N}} z^{-1}} \quad (식 11)$$

이는 다음과 같은 차이 식을 제공한다:

$$y_k[n] = x[n] + e^{j\frac{2\pi k}{N}} y_k[n-1] \quad \text{with } y_k[-1] = 0 \quad (식 12)$$

식 (12)는 복소수에 의한 곱셈을 수반하고 각 복소수 곱셈은 4 개의 실수 곱셈 및 4 개의 실수 더하기를 초래한다.

복소수 곱셈을 피하기 위해, 함수에는 복소공액극점 (complex conjugate pole)이 곱해지고 다음과 같이 단순화 될 수 있다:

$$H_k(z) = \frac{1 - e^{-j\frac{2\pi k}{N}} z^{-1}}{1 - 2 \cos\left(\frac{2\pi k}{N}\right) z^{-1} + z^{-2}} \quad (식 13)$$

이러한 2차 (second order) IIR의 차이 식은 다음과 같다:

$$y_k[n] = x[n] - x[n-1] e^{-j\frac{2\pi k}{N}} + 2 \cos\left(\frac{2\pi k}{N}\right) y_k[n-1] - y_k[n-2] \quad (식 14)$$

그리고 그러한 구조는 상태 변수를 사용하여 기술될 수 있다:

$$s[n] = x[n] - x[n-1] e^{-j\frac{2\pi k}{N}} + 2 \cos\left(\frac{2\pi k}{N}\right) s[n-1] - s[n-2] \quad (식 15)$$

$s[-1] = s[-2] = 0$  을 설정한다.

$$y_k[n] = X(k) = s[n] - s[n-1] e^{-j\frac{2\pi k}{N}} \quad (식 16)$$

$$y_k[n] = s[n] - e^{-j\frac{2\pi k}{N}} s[n-1] \quad (식 17)$$

$$= A - B e^{-j\theta} \quad (식 18)$$

[0169] 
$$=[A - B \cos \theta] + jB \sin \theta$$
 (식 19)

$$A=s[n]$$

$$B=s[n-1]$$

[0170] 
$$\theta = \frac{2\pi k}{N}$$

[0171] ● 피첼 알고리즘은 실제로 단일 1D DFT 계수의 계산을 수행한다. DFT와 비교할 때 여러 장점을 가지며, 이러한 이유로 때때로 1D 응용에서 사용되기도 한다.

[0172] ● 피첼 알고리즘은 전체 스펙트럼이 아닌, 몇 가지 스펙트럼 성분의 값만 필요할 때의 상황에서 유리하다. 특정 오디오 펄스를 가진 버튼 누름의 인식은 한 예시이다. 그러한 경우 알고리즘은 현저하게 빨라질 수 있다.

[0173] ● DFT 성분의 계산에 FFT 알고리즘을 사용하는 효율은 신호 길이 N (N은 2의 거듭 제곱이어야 함)에 의해 크게 결정된다. 반대로, 피첼 알고리즘의 경우에 N은 임의적일 수 있으며 계산 복잡도는 변하지 않는다.

[0174] ● 계산은 임의의 순간에 시작될 수 있으며, FFT의 경우에서와 같이 전체 데이터 블록을 기다릴 필요가 없다. 이로써, 피첼 알고리즘은 메모리 용량 관점에서 덜 까다로울 수 있으며, 매우 낮은 대기 시간 (latency)으로 수행될 수 있다. 그러므로, 피첼 알고리즘은 비트-역순으로 입력 또는 출력 데이터의 임의의 재배열을 필요로 하지 않는다.

[0175] **1D 피첼 알고리즘**

[0176] 1D 피첼에 대한 알고리즘은 매우 기본적인 구조를 가진다. 식 (17)로부터 시작될 수 있다.

[0177] 모든 샘플에서 일부 중간 프로세싱이 행해진다. FFT와 마찬가지로 샘플 블록으로 작업된다.

[0178] 1D 피첼의 계산을 초기화하기 위해서는 여러 설정이 필요하다:

- [0179] 1. 샘플링 속도
- [0180] 2. 블록 크기, N
- [0181] 3. 타겟 주파수

[0182] 샘플링 속도 및 블록 크기가 선택되면, 필요한 상수를 계산하는 5 단계 프로세스가 있다:

[0183] 상수 k, w, cosine, sine 및 coeff는 다음과 같이 정의된다:

$$k=(int) \left( 0.5 + \frac{N*target-freq}{sample-rate} \right)$$

$$w=(2\pi/N) * k$$

$$cosine=\cos w$$

$$sine=\sin w$$

[0184] 
$$coeff=2*cosine$$
 (식 20)

[0185] 샘플당 프로세싱에는 다음과 같은: S0, S1 및 S2의 3 변수가 사용된다. S1은 마지막 반복에서 단순히 S0의 값이다. S2는 2회 반복 단계 이전 (또는 다시 말해, S1 이전의 1회 반복)의 S0 값의 값이다. S1 및 S2는 각 샘플 블록의 시작 부분에서 0으로 초기화되어야 한다.

[0186] 매트릭스 [n x m]의 모든 열 (행)에 대해 다음 3 개의 식이 계산된다:

$$\{ S_0 = coeff * S_1 - S_2 + sample$$

$$S_2 = S_1$$

$$S_1 = S_0$$

$$real=(S_1 - S_2 * cosine)$$

$$imag=(S_2 * sine)$$

$$magnititude^2 = real^2 + imag^2 \}$$

(식 21)

[0187]

[0188]

[0189]

[0190]

[0191]

[0192]

[0193]

[0194]

[0195]

[0196]

[0197]

[0198]

[0199]

이는 1D 피켈 알고리즘의 기본 버전이다. 상술된 바와 같이, 이는 1D DFT와 동일한 결과를 제공한다.

위상 정보를 희생하면서 기본 버전보다 적은 계산이 필요한 1D 피켈 버전도 사용할 수 있고, 이는 변환의 실수부 및 허수부 둘 다를 계산하지 않음을 의미한다. 명확한 바와 같이, 실수부 및 허수부 둘 다를 계산하는 것이 바람직하며 더 빠른 옵션은 프로세서 오버헤드가 매우 엄격하게 제한되는 (capped) 경우에만 해당된다.

더 빠른 버전에서 샘플당 프로세싱은 동일하지만 블록 프로세싱의 끝은 상이하다. 실수 성분 및 허수 성분을 계산하고 그 이후에 이들을 상대 크기 제공으로 변환하는 대신, 실수 성분 및 허수 성분도 계산되는 기본 버전의 단계 없이, 다음과 같이 직접 계산된다:

$$magnititude=S_1^2 + S_2^2 - S_1 * S_2 * coeff$$

(식 22)

**L-변환 버전의 2D 구현**

다시 유의해야 하는 바와 같이, 피켈 알고리즘의 이러한 공통 버전은 1D 계산을 위해 정의된다. 이미지 프로세싱에서, 이러한 계산은 충분하지 않는데, 이는 이미지를 주파수 도메인으로 변환하기 위한 계산이 다음 2 차원: X 및 Y로 행해질 필요가 있기 때문이다. 또한, 피켈의 1D 구현이 1D DFT와 동일하지만, 2D의 경우 이는 참이 아닐 것이다. 따라서, 피켈 알고리즘은 객체 검출 및 이미지 프로세싱에 대한 후보자로 보이지 않는다. 피켈에 대한 또 다른 제한은 특정 신호를 조정하고 최적화할 수 있는 능력이 많지 않다는 것이다.

그러나, 본 개시는 2D 구현으로 이미지를 주파수 도메인으로 변환시키는 것으로, 1D 피켈 알고리즘의 원리로부터 시작하였지만, 여기에서 L-변환의 2D 변형으로서 기술된 2D에서의 완전히 새로운 방법으로 불릴 수 있을 정도로 이들을 변화시키는 방법을 기술했다. 또한, 이러한 방법에서 계산이 공간 도메인으로 돌아갈 필요없이 주파수 도메인에서 완전히 이루어지기 때문에, 2D 계산이 2D DFT와 동일할 필요는 없다.

도 5 및 도 6은 피켈의 전체 2D 구현이 사실상 많은 1D 피켈 계산의 조합일 수 있는 방법을 도시한다.

옵션은 먼저 도 5의 행의 다양한 1D 계산을 행하고, 그 후에 도 6과 같이 모든 1D 피켈 계산이 열에 대해 행해지는 두 번째 단계를 위해 이 결과를 사용하는 것이다. 대안으로, 먼저 열을 먼저 계산한 다음 행을 계산할 수 있다.

여기에 기술된 방법으로 그러한 구현을 사용할 수 있지만, 여러 이유로 선호되는 방식은 아니다. 우선, 행에 대한 계산은 열에 대한 계산이 완료될 때까지 기다려야 하며, 그 반대도 마찬가지이다.

병렬 프로세싱이 불가능하다는 의미이다. 둘째로, 계산은 여전히 정밀하게 최소화하지 않다. 도 7은 이를 보여준다. 도면에서 인덱스 (i, j)에서의 주파수 도메인 값이 필요한 2D 구현에 대하여 필요한 계산이 도시된다. 도 7에서 행이 먼저 계산되고 그 후에 열이 계산되는 옵션이 도시된다. 1D 계산은 먼저 인덱스 i에서 각 행에 대한 값을 계산했다. 이후에, 열에 대한 1D 계산이 수행되어 인덱스 j에서 값을 얻을 수 있다. 계산적으로 이는 이상적이지 않다는 것이 명백할 것이다. 피켈의 2D 구현이, 공간 도메인으로 돌아가서 원본 이미지를 다시 생성할 수 없는 방식으로 데이터를 변화시킬 것이라는 점은 주파수 도메인 변환 분야의 전문가에게도 명백할 것이다. 그러나, 전에 언급된 바와 같이, 이러한 방법은 주파수 도메인 날짜만을 사용하여 데이터의 분류를 기술한다. 그러므로, 여기에 기술된 방법에서의 드라이버는 공간 도메인 데이터인 드라이버 대신에 분류기에 대해 가능한 최상의 입력을 발생시키는 가능한 빠른 계산을 가져야 한다.

다음으로, 본 개시는 속도 및 검출 둘 다의 관점에서, 분류기에 대한 최적화 된 데이터를 갖는 일련의 옵션을

기술할 것이다. 특히, 비디오 스트림에서 시간적 데이터가 어떻게 가장 잘 캡처되는지가 기술될 것이다.

[0200] 이들은, 다른 것 중에서, 다음과 같다:

- [0201] ● 2 개의 존을 사용하는 다양한 각각의 특징 사용;
- [0202] ● 특징 및 존에 의해 커버되는 비디오 스트림의 시퀀스에서 프레임 수를 선택;
- [0203] ● 각 존에 대해 상이한 타겟 주파수 선택;
- [0204] ● 각 존에 대해 2 개의 입력을 가짐 - 상기 입력 각각은 주파수 도메인 변환임;
- [0205] ● 변환의 실수 부와 허수 부 둘 다에 최적화될 수 있는 각 입력을 위해 가변 코어 필터를 가짐;
- [0206] ● 각 인덱스의 주파수 도메인 변환을 정의하는 루프에 대한 입력의 가변 시퀀스 및 형상을 가짐;
- [0207] ● 각 특징의 특징 쌍을 사용하여 각 특징에 대해 정규화된 복소 벡터 발생; 및
- [0208] ● 마지막으로, 모든 정규화된 복소 벡터를 단일 포맷으로 서로 조합시킴.

[0209] 계산은 도 7에서의 예시보다 더 최소화될 수 있다. 한 가지 방식은 도 8에서 도시된 바와 같은데, 이 경우 주파수 도메인으로 변환하기 위한 입력 셀은 값이 필요한 인덱스의 위치에서만 취해진다. 그러나, 이는 여전히 행이 열의 결과를 기다리거나 그 반대로 하는 것을 필요로 한다. 셀 양의 길이가 입력 파라미터가 되어 특징들 사이에 더 많은 차별화를 허용하는 한편, 분류되는 객체의 세부 사항도 캡처할 수 있다는 것이 유리할 것이다.

[0210] 보다 효과적인 계산이 도 9에 도시된다. 여기에서 인덱스 결과는 인덱스에서 행 및 열에 대한 1D 출력을 별도로 계산하고, 그 후에 이를 단일 값으로 조합하여 취해진다. 속도에서 개인과는 별도로, 이 방식에서 가장 큰 이점은 행에 대한 1D 또는 열에 대한 1D가 먼저 계산되는 경우 차이가 없어, 행 및 열에 대한 값을 병렬로 계산될 수 있다는 것이다. 도 10에 도시된 바와 같이 입력 길이를 제한하여, 훨씬 큰 속도 계인을 달성할 수 있고, 이 경우 단일 행 및 열 입력만 계산하면 된다.

[0211] 도 11에 도시된 바와 같이, 주어진 존 인덱스에서 결과를 얻기 위해 따라온 2 개의 입력이 행 및 열, 또는 심지어 인접한 셀을 따라갈 필요조차 없다고 고려하는 경우, 주파수 도메인 입력 데이터로 분류기를 트레이닝시킬 수 있는 자유도의 양이 훨씬 더 커진다.

[0212] 다음 설명에서, 도면을 보다 개략적으로 유지하기 위해 존당 단일 입력이 종종 도시된다. 그러나, 도 12에 도시된 바와 같이, 각 존에 대해 두 개의 입력이 있음이 강조되어야 한다.

[0213] 각 입력에 대해 주파수 도메인 변환이 행해지게 되어 실수와 허수가 제공된다. 언급된 바와 같이, 각 존은 2 개의 입력을 가지며, 이하에서 이들을 단일 정규화된 복소수 값으로 조합하는 방식이 기술된다. 이는 또한 도 13에 도시된다.

[0214] 우선, 존 1의 2 개의 실수 (real) 입력이 조합된다:

$$Real(Zone1_{Feature_i}) = \sqrt{Real(Input1_{Zone1})^2 + Imag(Input1_{Zone1})^2} + \sqrt{Real(Input2_{Zone1})^2 + Imag(Input2_{Zone1})^2}$$

(식 23)

[0216] 존 1의 두 개의 허수 (imaginary) 입력이 위상을 제공하기 위해 조합된다:

$$Imag(Zone1_{Feature_i}) = \left/ \frac{Imag(Input1_{Zone1})}{Real(Input1_{Zone1})} \right. + \left/ \frac{Imag(Input2_{Zone1})}{Real(Input2_{Zone1})} \right.$$

(식 24)

[0218] 존 2의 두 개의 실수 입력 및 허수 입력에 대해 동일한 프로세스가 반복된다:

$$Real(Zone2_{Feature_i}) = \sqrt{Real(Input1_{Zone2})^2 + Imag(Input1_{Zone2})^2} + \sqrt{Real(Input2_{Zone2})^2 + Imag(Input2_{Zone2})^2}$$

(식 25)

$$Imag (Zone2_{Feature_i}) = \frac{Imag(Input1_{Zone2})}{Real(Input1_{Zone2})} + \frac{Imag(Input2_{Zone2})}{Real(Input2_{Zone2})}$$

[0220] (식 26)

[0221] 다음으로, 존 1 및 존 2의 실수 값에 대한 결과가 정규화로 조합된다:

$$Real (Feature_i) = \left| \frac{Real (Zone2) - Real (Zone1)}{Real (Zone1) + Real (Zone2)} \right|$$

[0222] (식 27)

[0223] 이는 또한 존 1 및 존 2의 허수 값에 대해서도 행해진다:

$$Imag (Feature_i) = \left| \frac{Imag (Zone2) - Imag (Zone1)}{Imag (Zone1) + Imag (Zone2)} \right|$$

[0224] (식 28)

[0225] 이러한 방식으로 1 개의 특징을 형성하는 각 존 쌍은 도 13에 도시된 바와 같이 정규화된 복소 벡터를 제공한다:

$$\vec{V}_{Feature_i} = Real(Feature_i) + jImag(Feature_i)$$

[0226] (식 29)

[0227] 다수의 그러한 정규화된 복소 벡터로 모델이 구성될 수 있다:

$$\vec{V}_{Feature} = \{ \vec{V}_{Feature_1}, \vec{V}_{Feature_2}, \dots, \vec{V}_{Feature_n} \}$$

[0228] (식 30)

[0229] 이는 이러한 포맷의 다수의 정규화된 복소 벡터이고, 상기 복소 벡터는 여기에 기술된 방법이 분류기에 대한 새로운 타입의 입력으로 제공되는 입력이다. 이러한 포맷이 확률 이론 및 양자 물리학의 모든 수학이 분류에 적용되도록 함은 기술분야의 전문가에게 명백할 것이다.

[0230] 공간 도메인에서의 픽셀을 분류기에 대한 입력으로서 가지는 것에 비해, 인덱스에서 값을 얻는데 필요한 주파수 도메인 계산의 수가 기술된 방법에서 크게 감소될 것이다. 이러한 방식으로 얻은 값은 여전히 주파수 도메인에서의 형상 정보와 강한 상관관계가 있으면서, 또한 많은 제어가 가능해져 앨리어싱 및 주기 신호와 같은 효과를 줄일 수 있다. 하나의 주요 목표가 각 인덱스에서 고유한 결과를 가지는 것이기 때문에, 이들 효과의 감소는 중요하다. 여기서 이 방법에 대해 공간 도메인에서 이미지를 재구성하기 위해 모든 데이터를 가질 필요는 없다는 것을 다시 주목해야 한다. 목표는 검출되는 객체의 위치 및 움직임을 희소하게 인코딩하는 주파수 도메인 정보를 캡처하는 것이다.

[0231] 생성된 주파수 도메인 공간이, 공간 도메인에서 픽셀 입력을 직접 사용하는 것과 비교할 때, 분류기가 프로세싱해야 하는 파라미터의 양을 매우 크게 감소시킴은 명백하다. 기술된 방법이 비디오 스트림에서 프레임 시퀀스에서만 검출될 수 있는 시간 정보를 캡처하기 위해 사용될 때 이러한 이점은 지속적으로 증폭된다.

[0232] 행 또는 열을 따라 있을 필요는 없는, 그러나 자유로운 각 인덱스를 따라 1D 계산으로 돌아가면, 다음과 같이 다시 작성할 수 있다:

$$k = (Int) \left( 0.5 - \frac{N - Target\_frequency}{Sample\_rate} \right)$$

[0233] (식 31)

$$\omega = \left( \frac{2\pi k}{N} \right)$$

[0234] (식 32)

$$\text{coeff} = 2 \cdot \text{digital\_filter}$$

[0235] (식 33)

$$\text{digital\_filter} = \sin(A\omega + b\pi) + j \sin\left(A\omega + (2b + 1) \frac{\pi}{2}\right)$$

[0236] (식 34)

$$S_0 = \text{coeff} \cdot S_1 - S_2 + \text{sample}$$

[0237] (식 35)

$$S_1 = S_0$$

$$S_2 = S_1$$

[0238]

$$\text{Real} = (S_1 - S_2 \cdot \sin(A\omega + b))$$

[0239] (식 36)

$$\text{Imag} = \left( S_2 \cdot \sin\left(A\omega + b \frac{\pi}{2}\right) \right)$$

[0240] (식 37)

[0241] 이 방법에서 주파수 도메인 변환의 실수 부와 허수 부 둘 다에 대해, 파라미터 a 및 b를 사용하여 전달 함수가 어떻게 조정 가능한지 살펴본다.

[0242] 주목해야 하는 바와 같이, 디지털 필터로 동작되는 이들 전달 함수는 존의 각 입력에 대해 개별적으로 선택될 수 있는 옵션을 가지며, 이는 제 1 입력 및 제 2 입력이 상이한 개별 전달 함수 설정 (discrete transfer function settings)을 가질 수 있음을 의미한다.

[0243] 앞에서 언급한 바와 같이, 이 방법의 이점 중 하나는 분류기에 대한 입력으로 사용되기 전에 주파수 도메인 데이터를 정리할 수 있는 대량의 자유도를 조정하기 위해 많은 옵션 역시 기술되어 있다는 것이다.

[0244] 이는 도 14에 도시된다. 각 입력에 대해 코어 루프에 사용되는 타겟 주파수, 입력 방향 및 입력 시퀀스, 및 코어 루프 내에서 사용되는 전달 함수를 자유롭게 선택하는 방법을 알 수 있다.

[0245] 도 15는 최적화 위상에서 코어 내부 루프에 대한 입력의 방향 및 시퀀스를 조정할 수 있는 방법을 도시한다.

[0246] 이 시점에서 이 방법의 계산은 L-변환 (식 7)을 사용하여 곱셈 알고리즘을 생성하는 이론과는 매우 상이함이 명백할 것이다. 또한, L-변환에 연결된 Z-변환과는 실질적으로 상이하다. 설명을 위해, 이 방법에서 사용된 변환을 여기에서 L-변환의 2D 변형이라고 한다.

[0247] 이 방법을 사용하면, 특징을 포함하는 인덱스에서 주파수 도메인 값만이 계산될 필요가 있음이 통상의 기술자에게 명백할 것이다. FFT를 사용하는 경우, FFT를 최소하게 계산할 수 없기 때문에, 주파수 도메인에서 모든 값을 계산해야 한다. 이 방법에는 FFT와 같은 이미지 입력 크기의 제한이 없다는 점을 다시 강조하는 것이 또한 중요하다. 실제로, 도 16에 도시된 바와 같이, 정사각형 또는 직사각형 입력 이미지 크기를 가지는 것에 대한 제한은 없다.

[0248] 다음으로, 비디오 프레임의 시퀀스로부터 시간 정보를 추출하는 방법의 적용이 기술된다.

[0249] 도 17은 특징을 형성하는 각 존 쌍이 반드시 비디오 시퀀스의 동일한 프레임에 있을 필요가 없는 방법을 도시한다. 도 18은 코어 내부 루프의 각 입력의 방향 및 시퀀스가 어떻게 비디오 프레임 시퀀스의 단일 프레임으로 제한되지 않는 경로를 취할 수 있는지를 도시한다. 픽셀이 공간 도메인으로부터 입력으로서 직접 사용되는 경우, 비디오 시퀀스의 프레임들 사이의 픽셀을 상호 참조할 때 해결해야 할 순열이 여기에 기술된 방법에 비해 훨씬 느리고 트레이닝하기가 더 어렵다는 것이 전문가에게는 명백할 것이다.

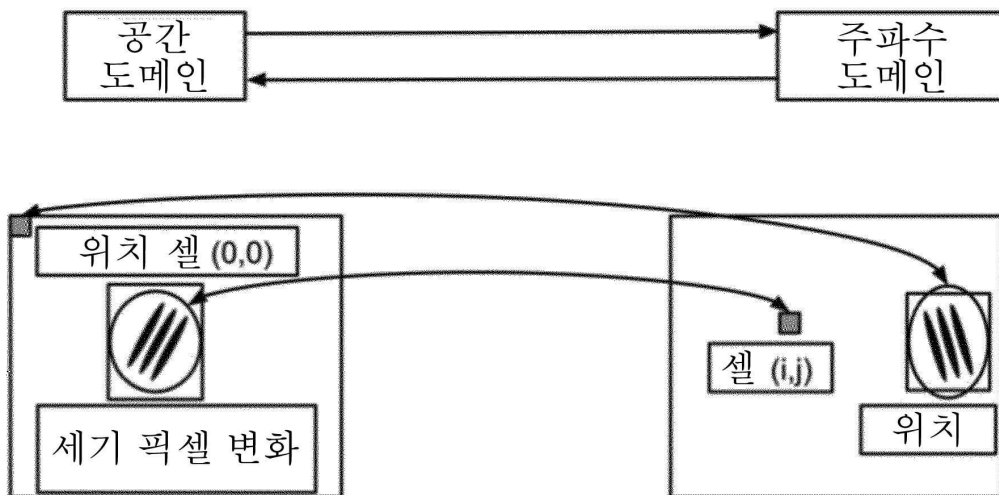
[0250] 도 19는 본 방법이 모델의 주파수 도메인 변환을 발생시키기 위해 사용된 프레임 시퀀스의 선택을 허용하여 컨테츠의 검출을 위해 실제 시간 데이터가 사용되도록 하는 자유도를 도시한다;

[0251] 또한, 도 20에서 주파수 도메인에서의 작업이 픽셀의 공간 도메인에 비해 어떻게 큰 이점을 가지는지 알 수 있다. 공간 도메인에서 각 픽셀은, 픽셀의 장점에 대해 추론할 가능성이 거의 없이, 0 내지 255 사이의 입력을 가지며, 대신에 주파수 도메인에서 주파수 도메인에서의 검색 공간은 신호가 분명히 두드러지게 나타내는 다량의 노이즈로 대부분 구성된다. 따라서, 여기에 기술된 방법은 또한 신호 품질에 기초한 보다 효과적인 최적화 로직에 대한 가능성을 가지며, 이는 잠재적으로 감독되지 않은 방식으로 행해질 수도 있다.

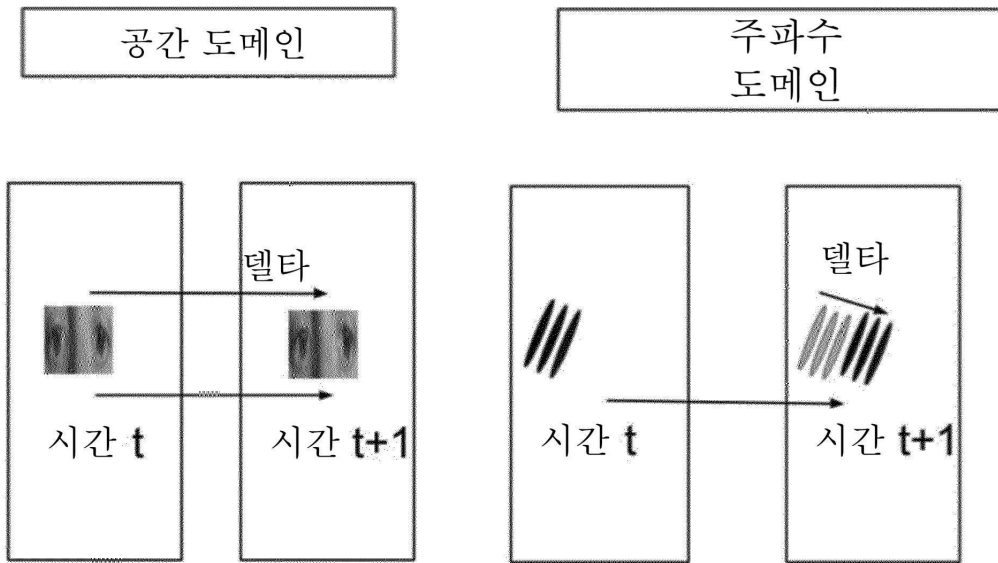
[0252] 도 21은 비디오 프레임의 시퀀스에서 시간적 효과를 검출하는 모델을 생성하는 방법을 사용할 때 기술된 방법에 사용될 수 있는 트레이닝의 가능한 상위 레벨 다이어그램을 도시한다. 도 21은 시퀀스에 사용된 프레임 수가 가변적이고, 그 후에 도 14에도 도시된 바와 같이 각 특징의 각 인덱스에 대해 타겟 주파수가 최적화되고 기본 루프의 시퀀스가 선택되며 루프 내에서 사용되는 기본 전달 함수가 최적화되는 방법을 도시한다. 이는 각 루프의 실수 부 및 허수 부 둘 다에 대해 행해진다. 최적화의 출력은 다수의 정규화된 복소 벡터를 갖는 포맷이 될 것이며, 이는 비디오 프레임의 시퀀스에서 시간적 효과의 검출 및 분류를 위해 특히 효과적인 입력으로서 사용될 수 있지만, 방법은 그러한 구현에 제한되지 않는다.

도면

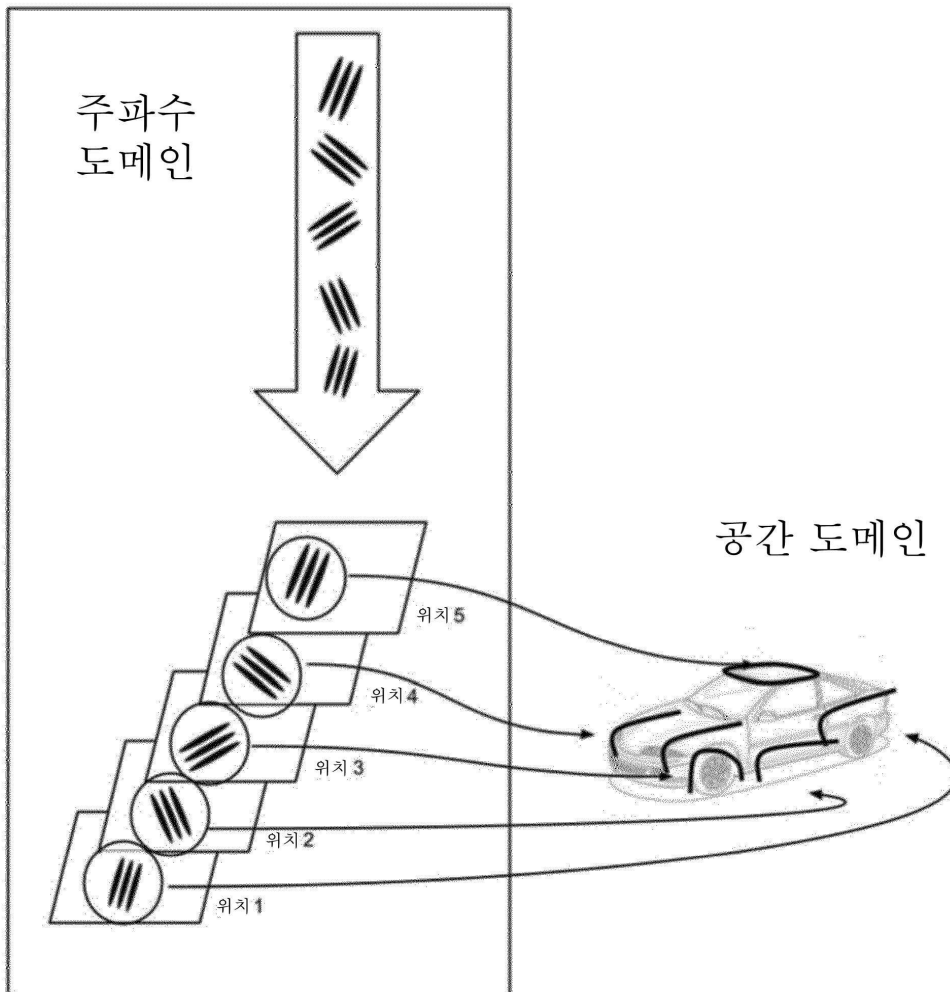
도면1



도면2

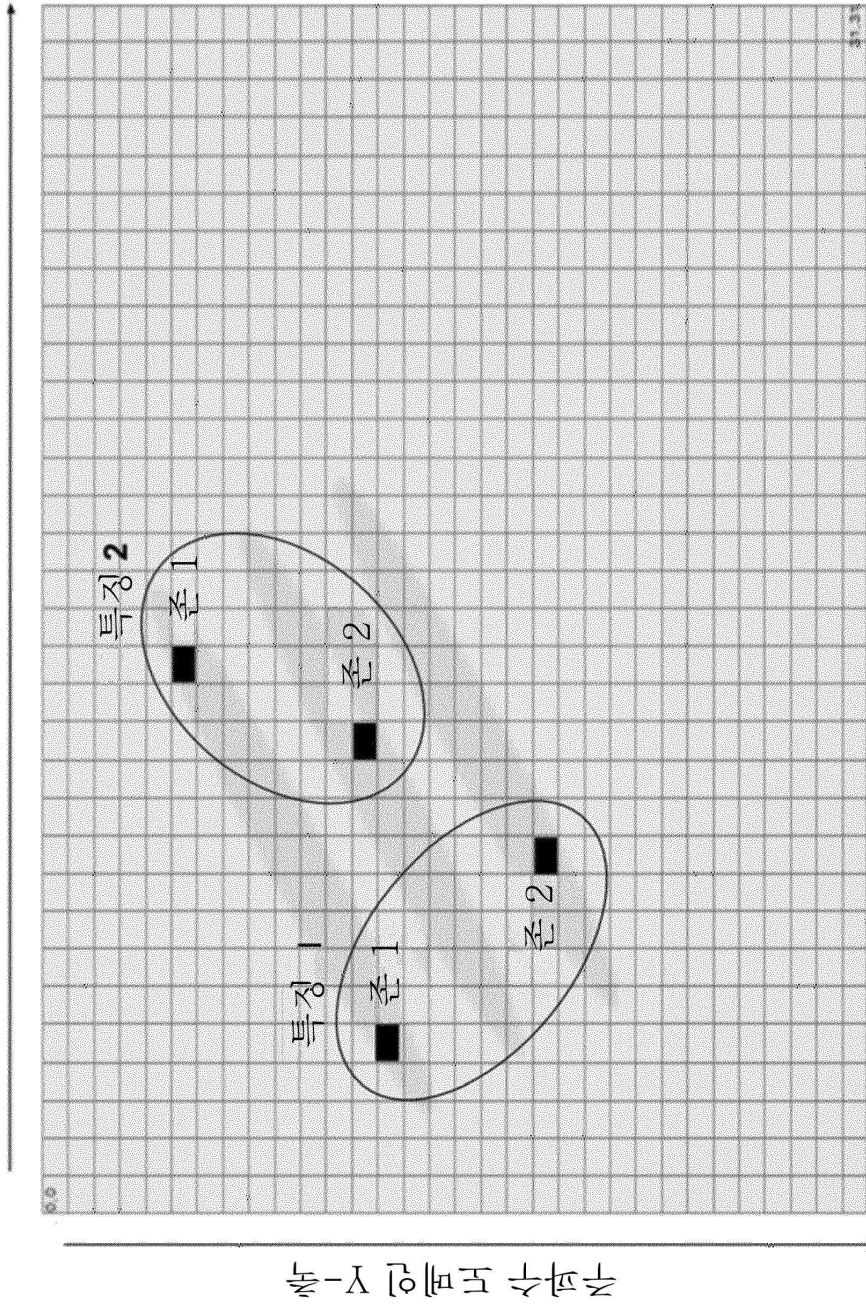


도면3

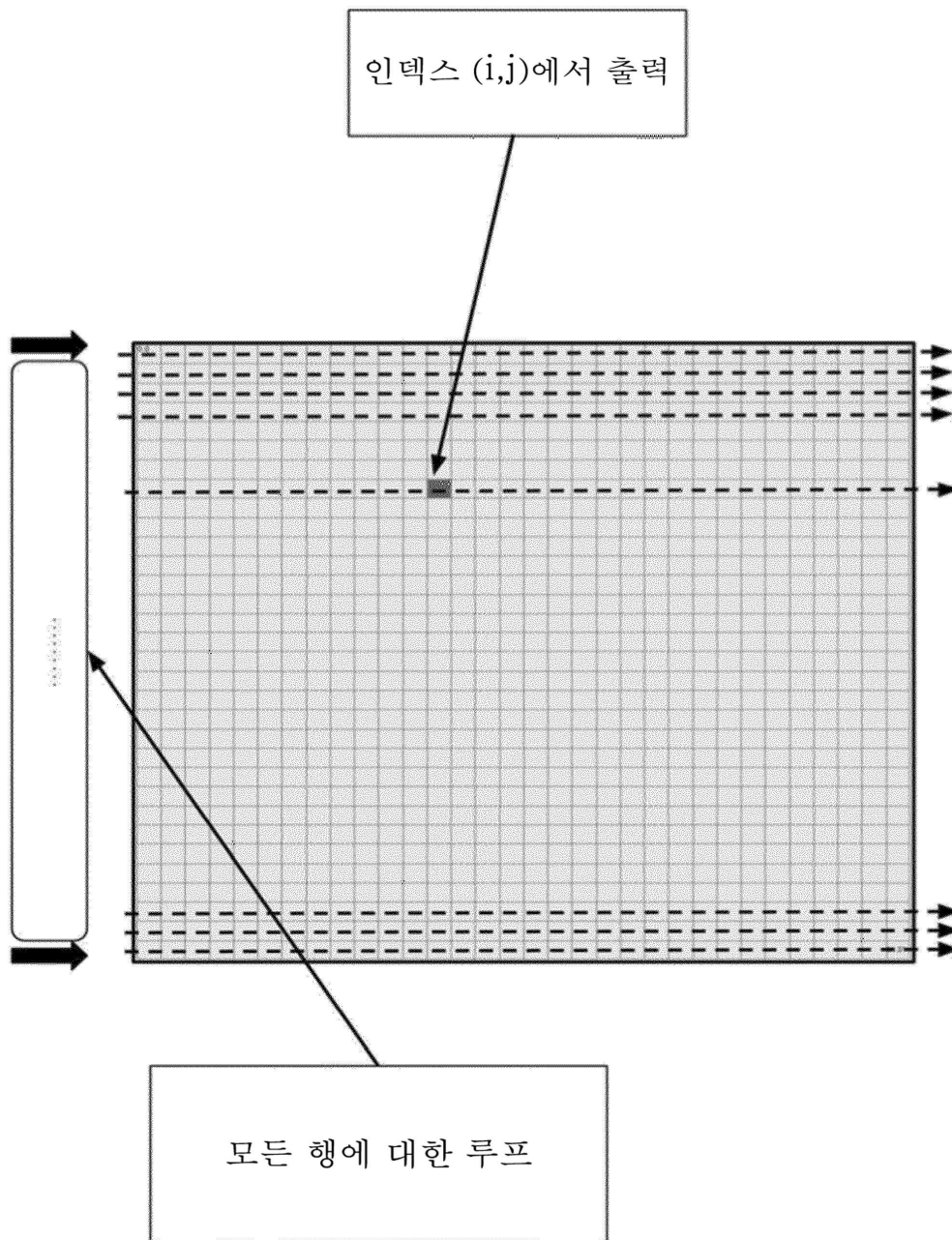


도면4

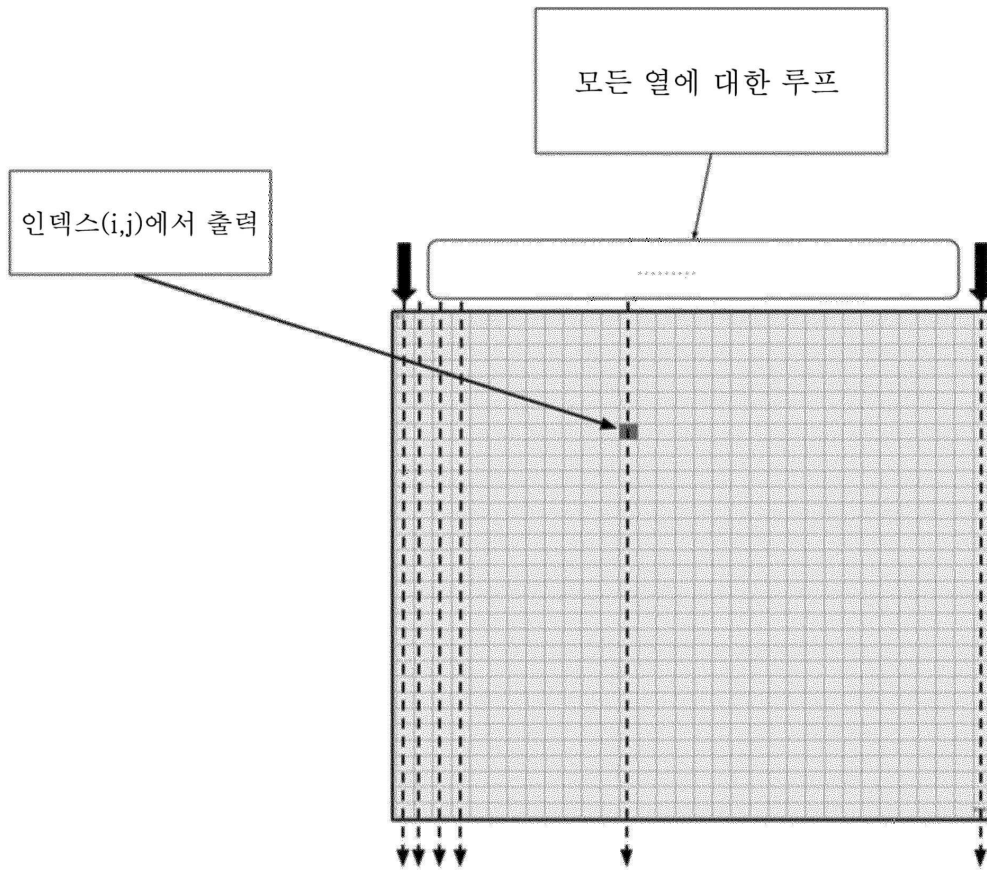
주파수 도메인 X-축



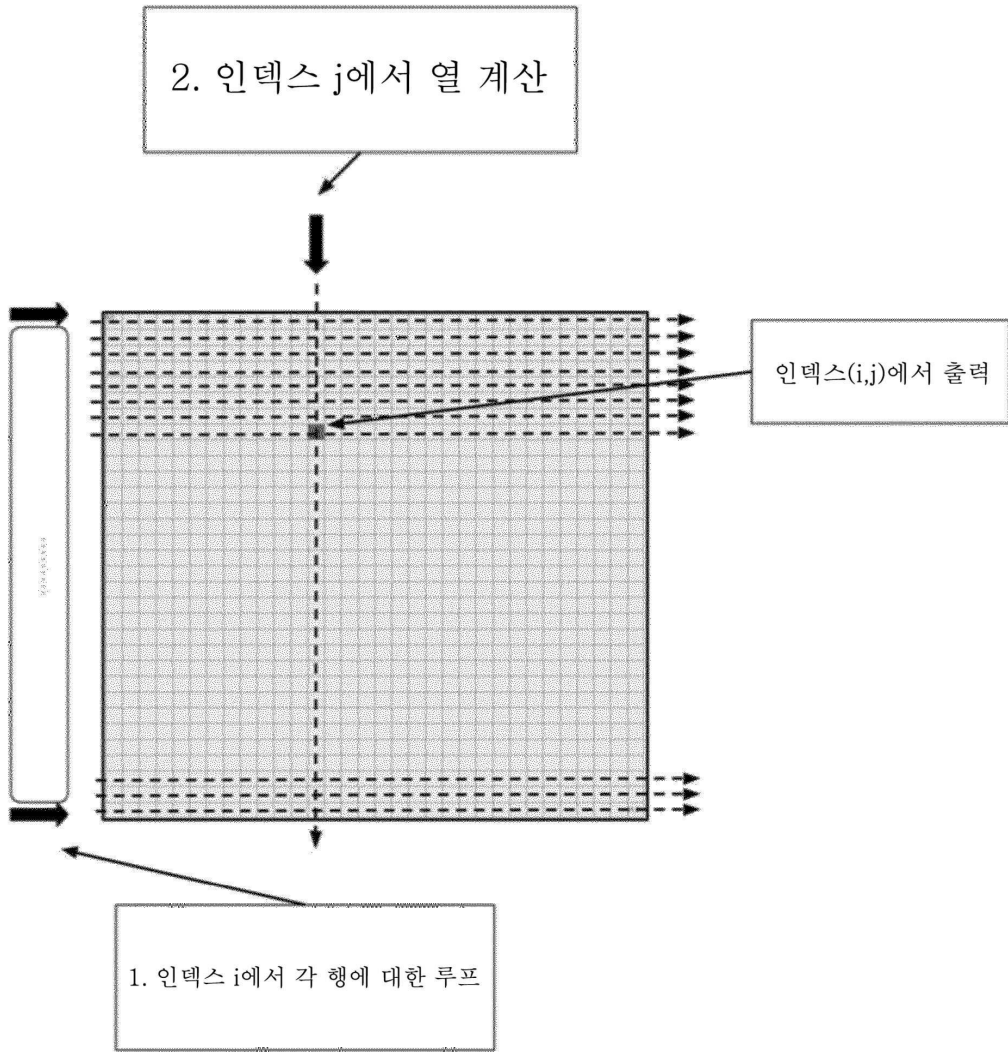
도면5



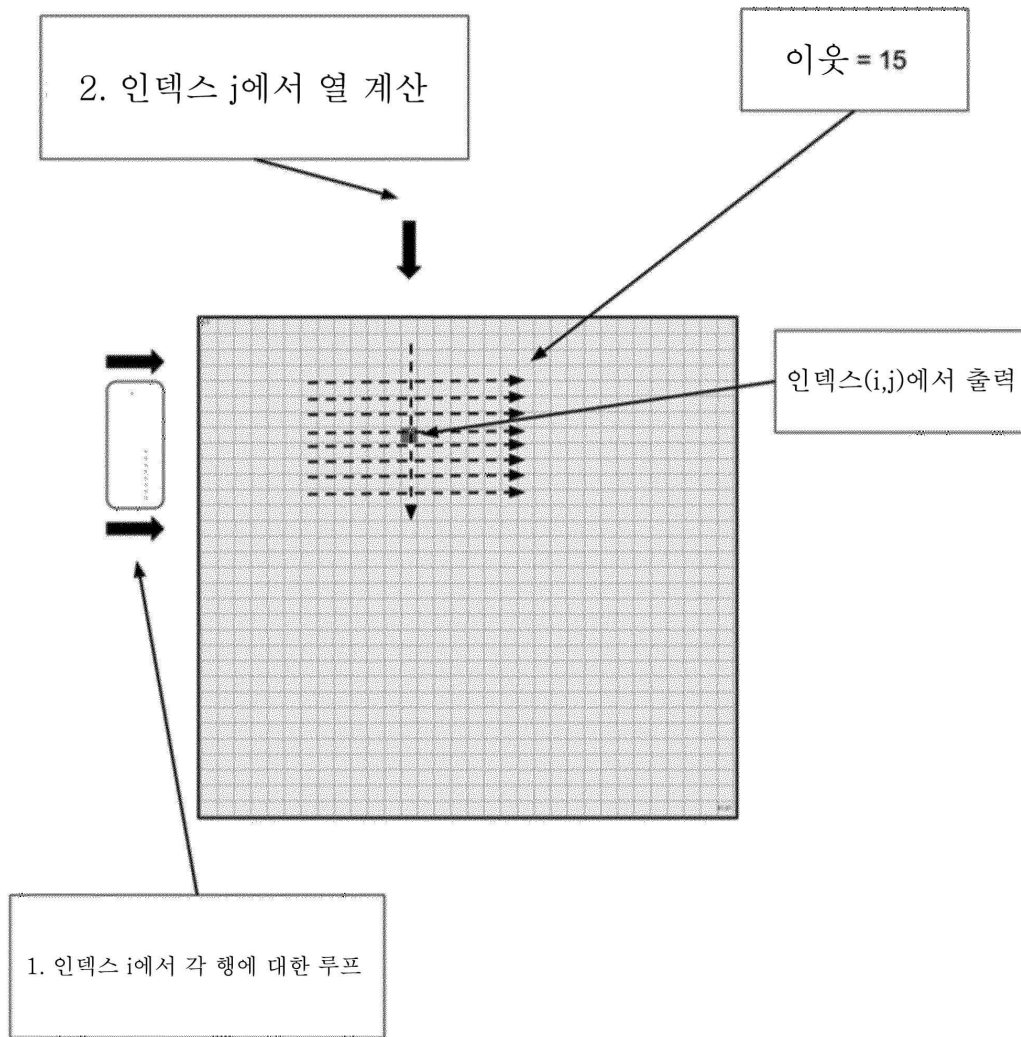
도면6



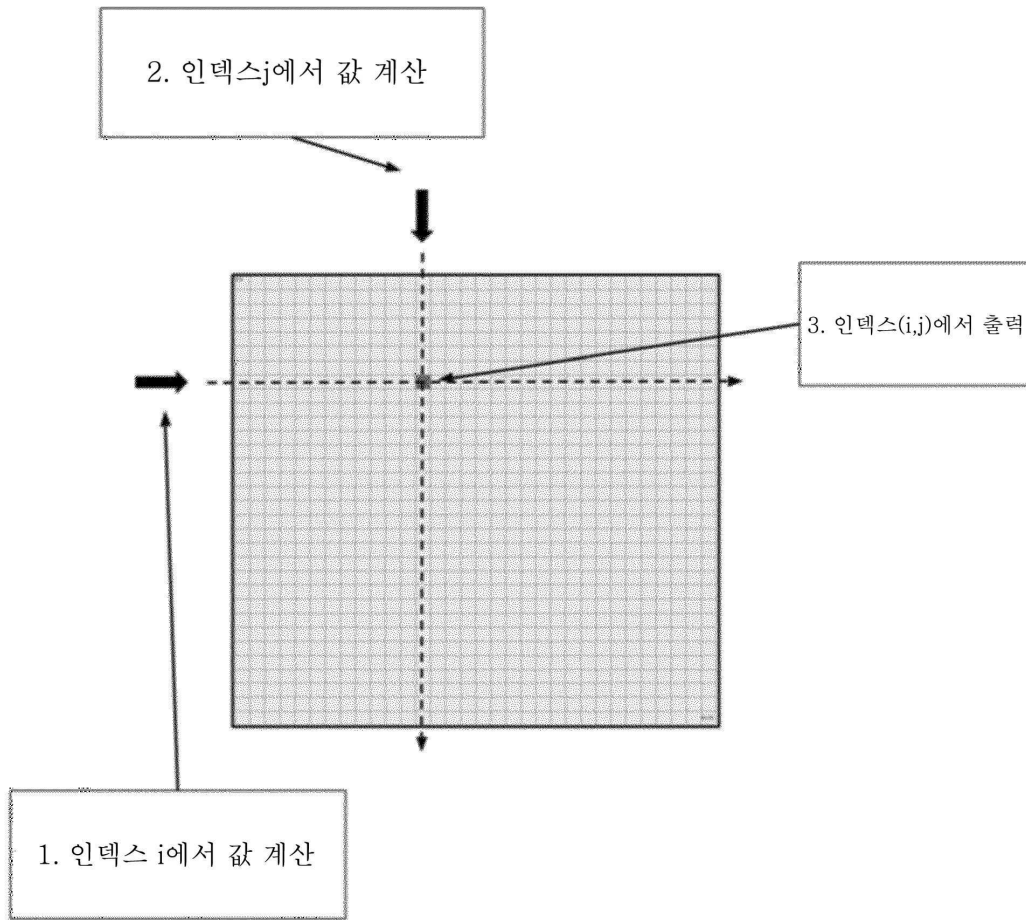
도면7



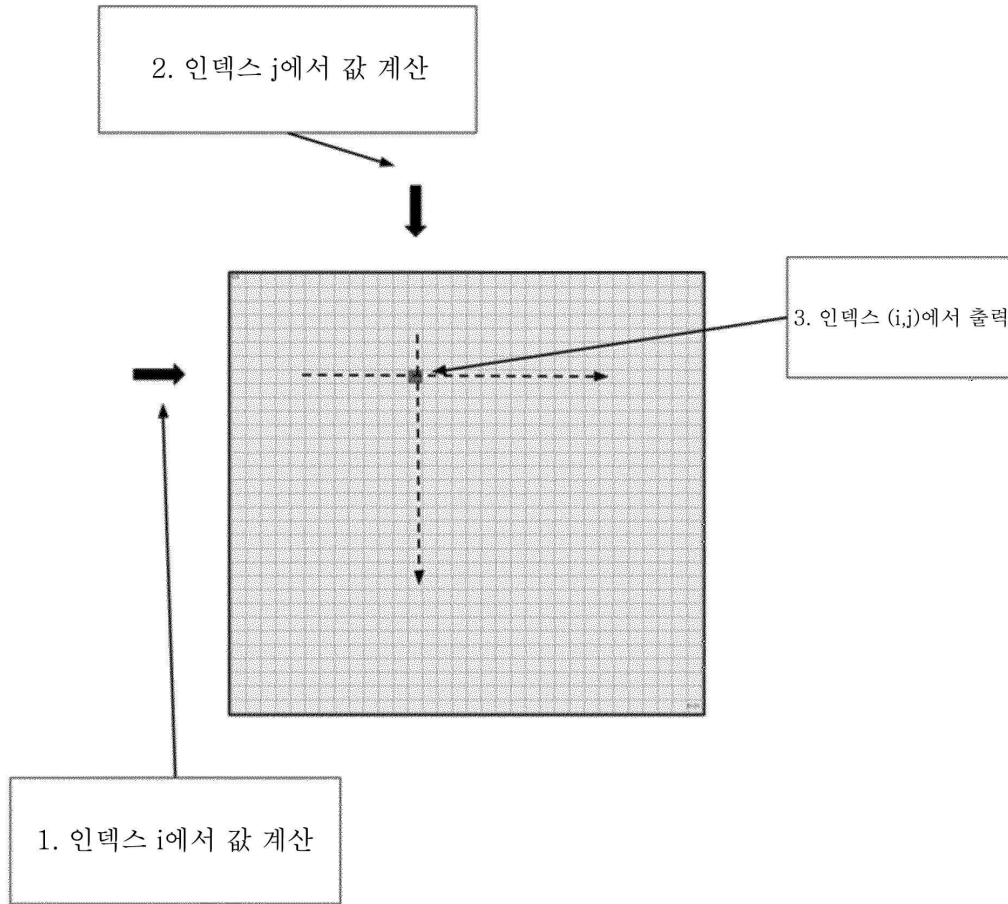
도면8



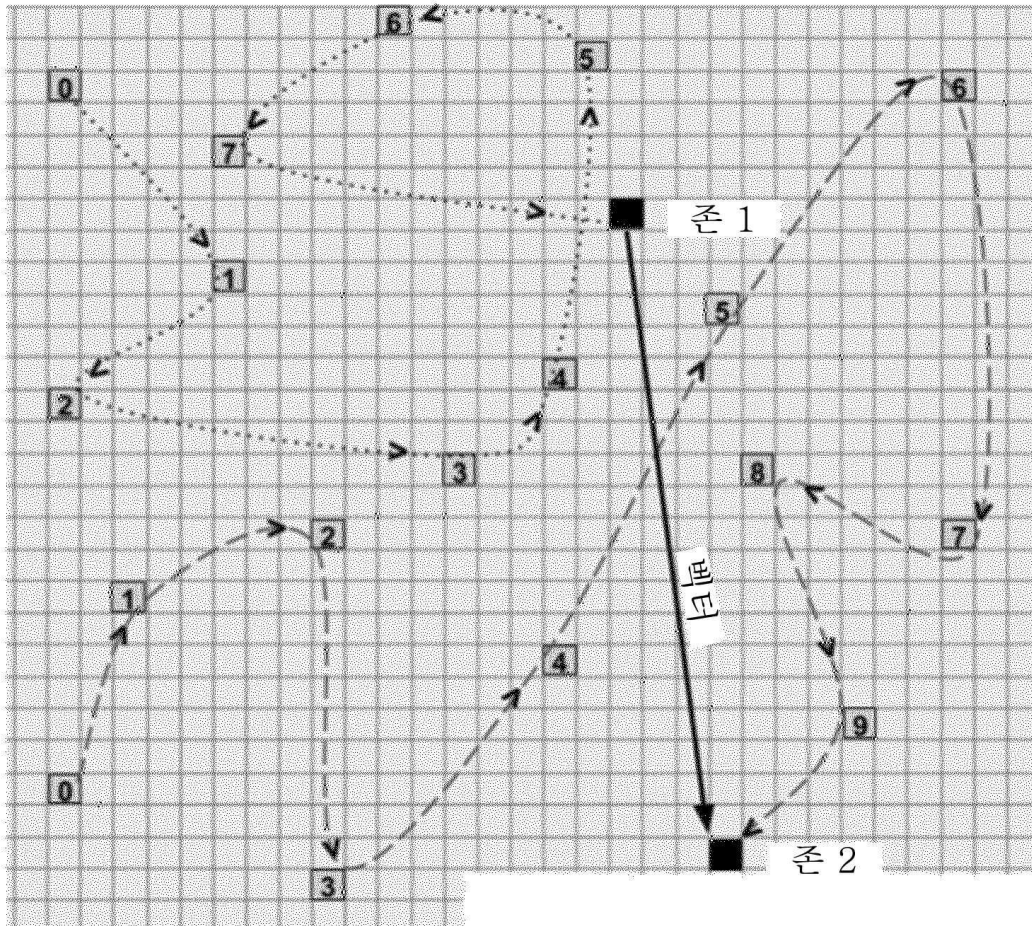
도면9



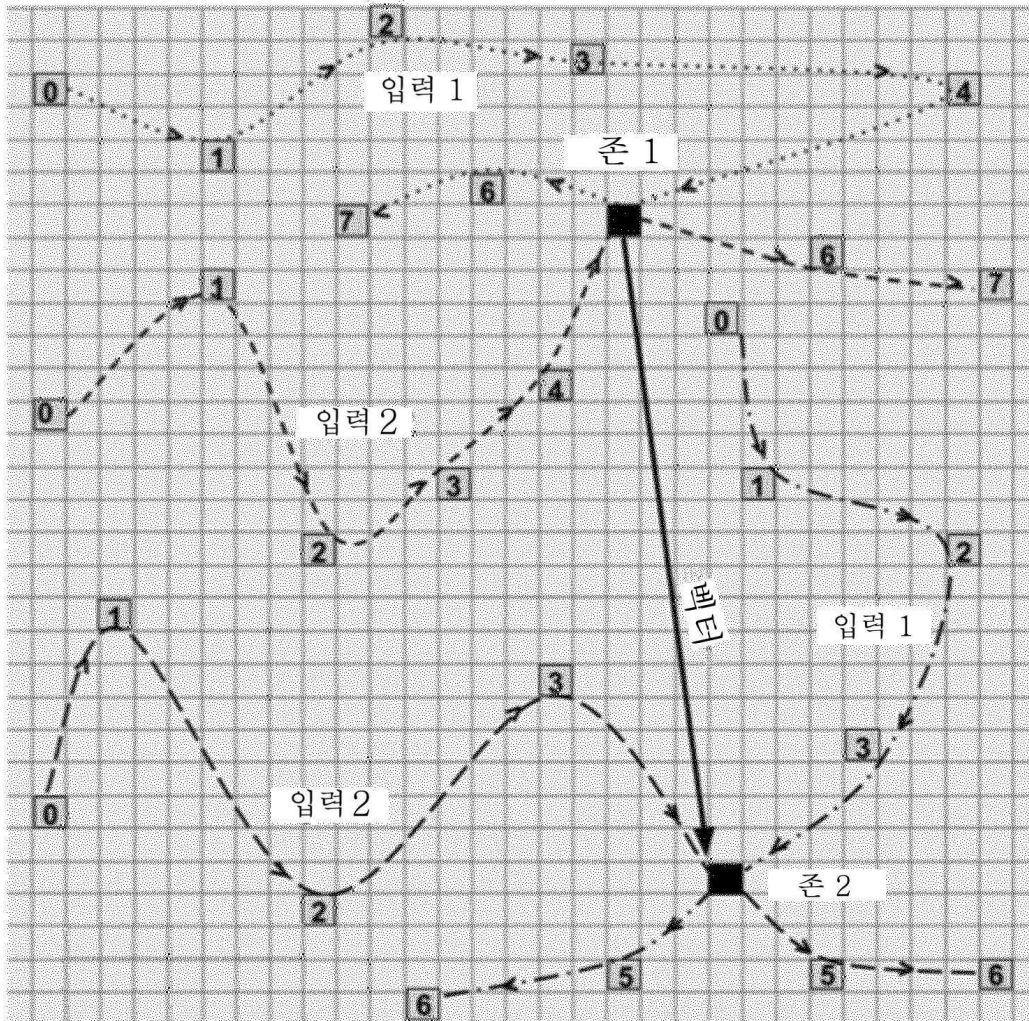
도면10



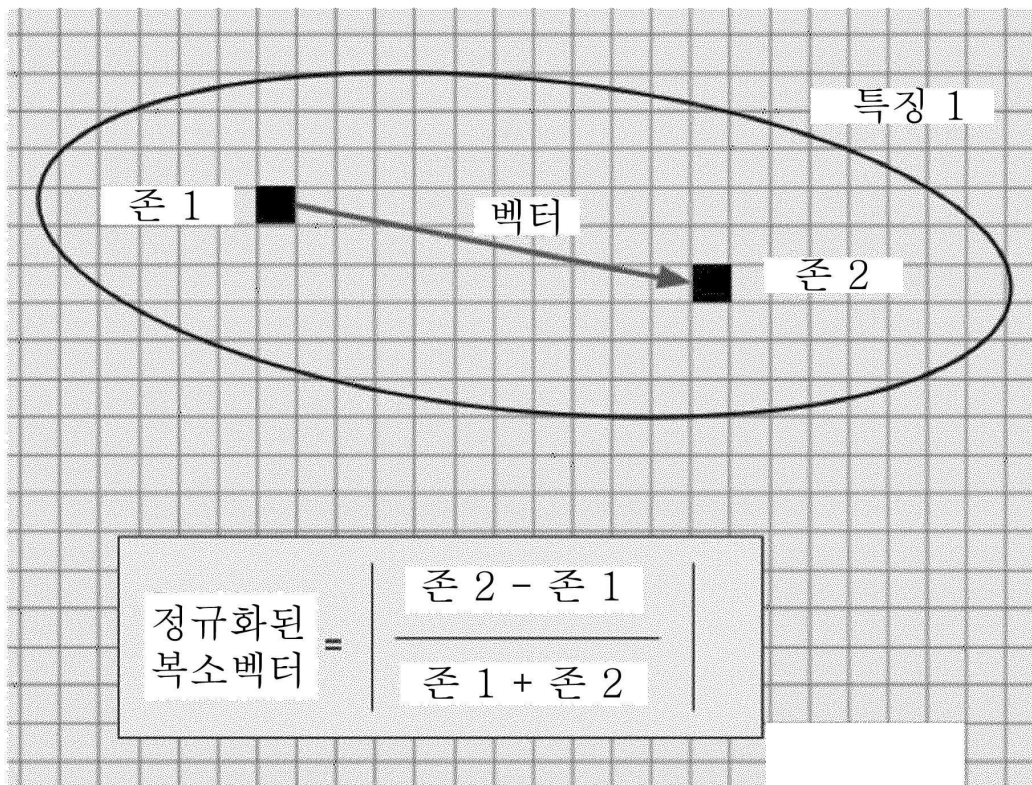
도면11



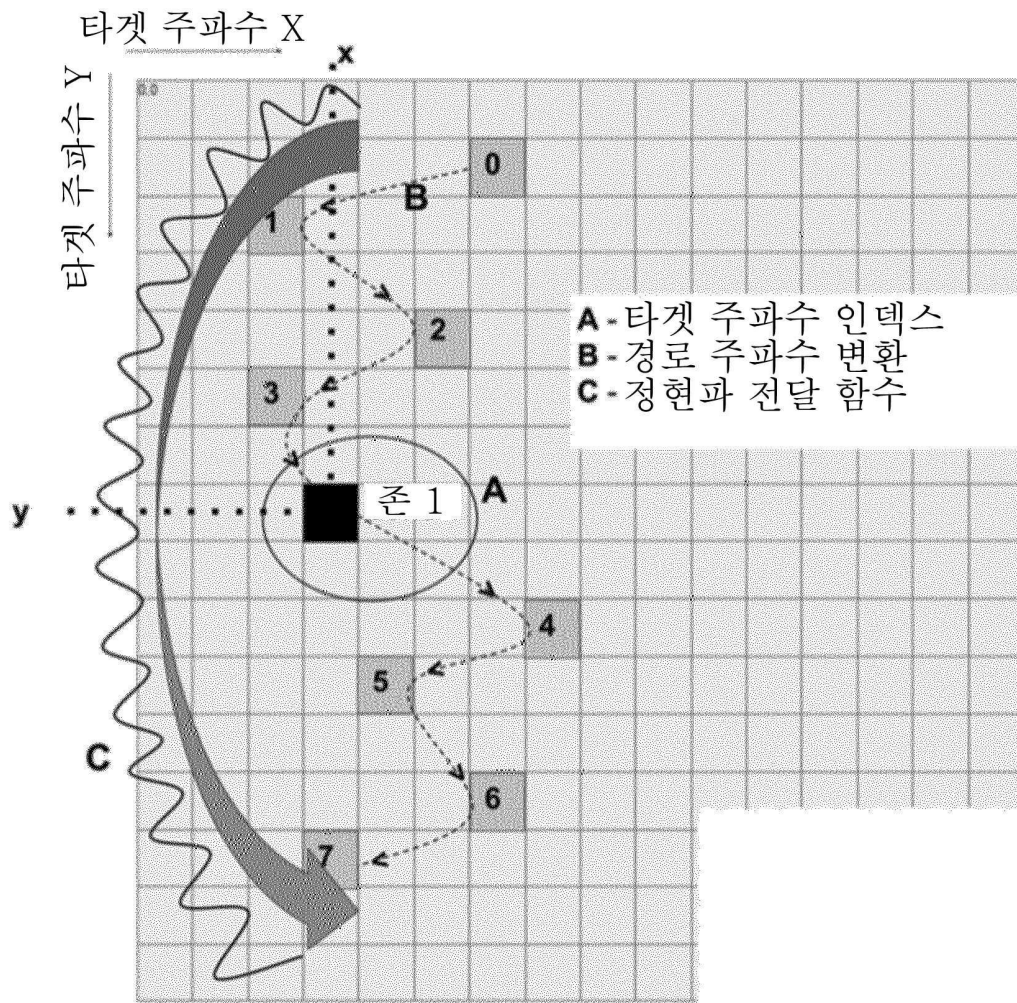
도면12



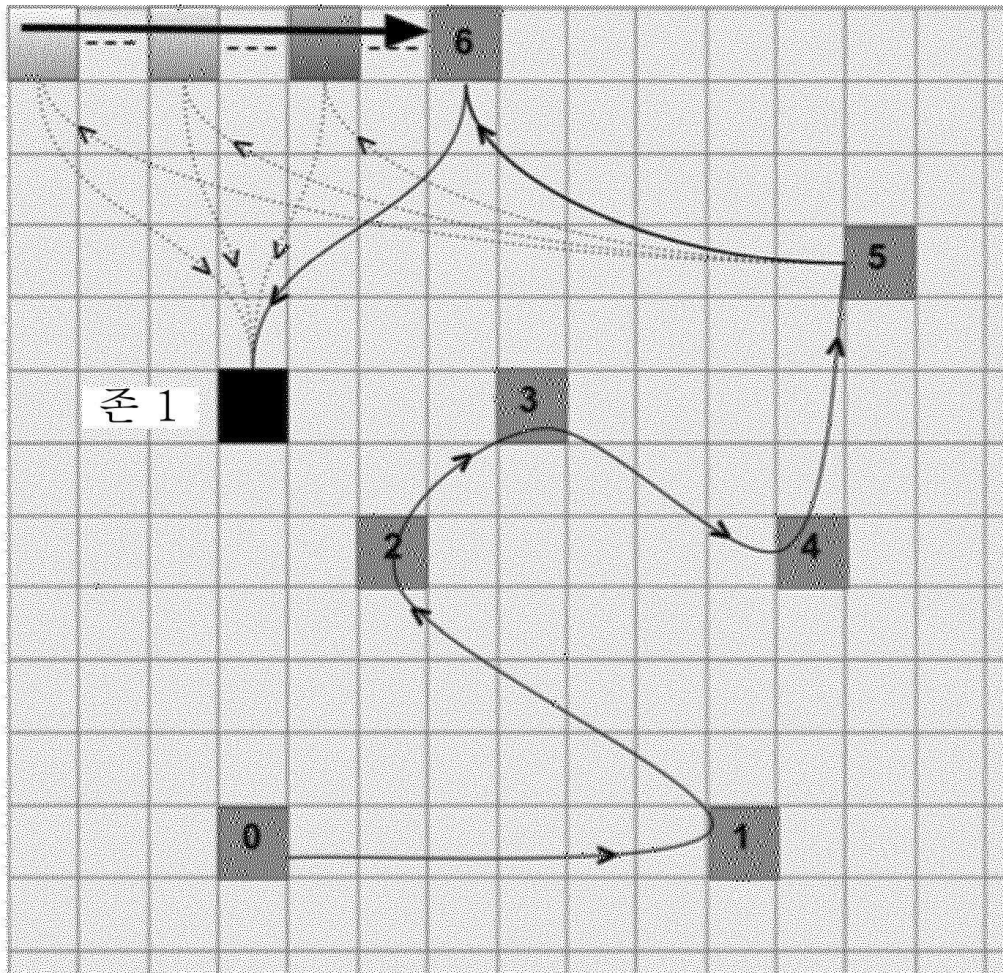
도면13



도면14

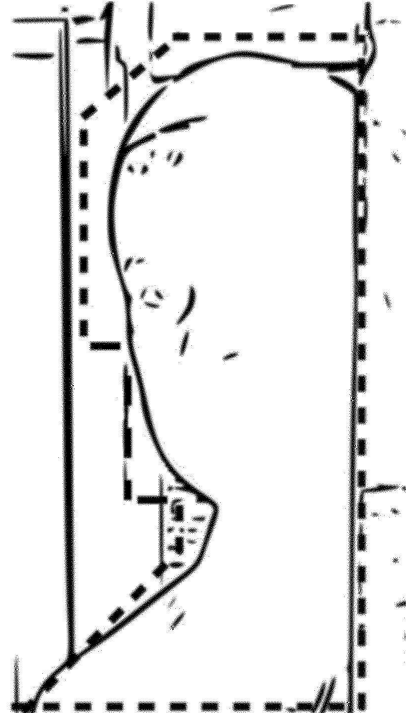
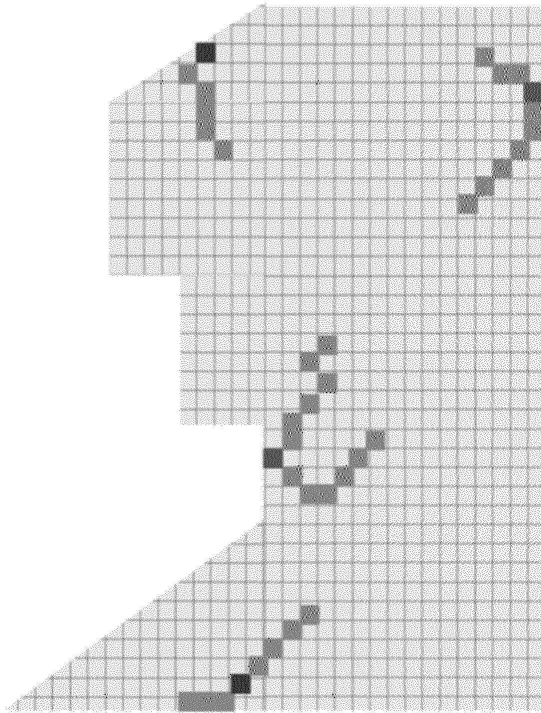


도면15

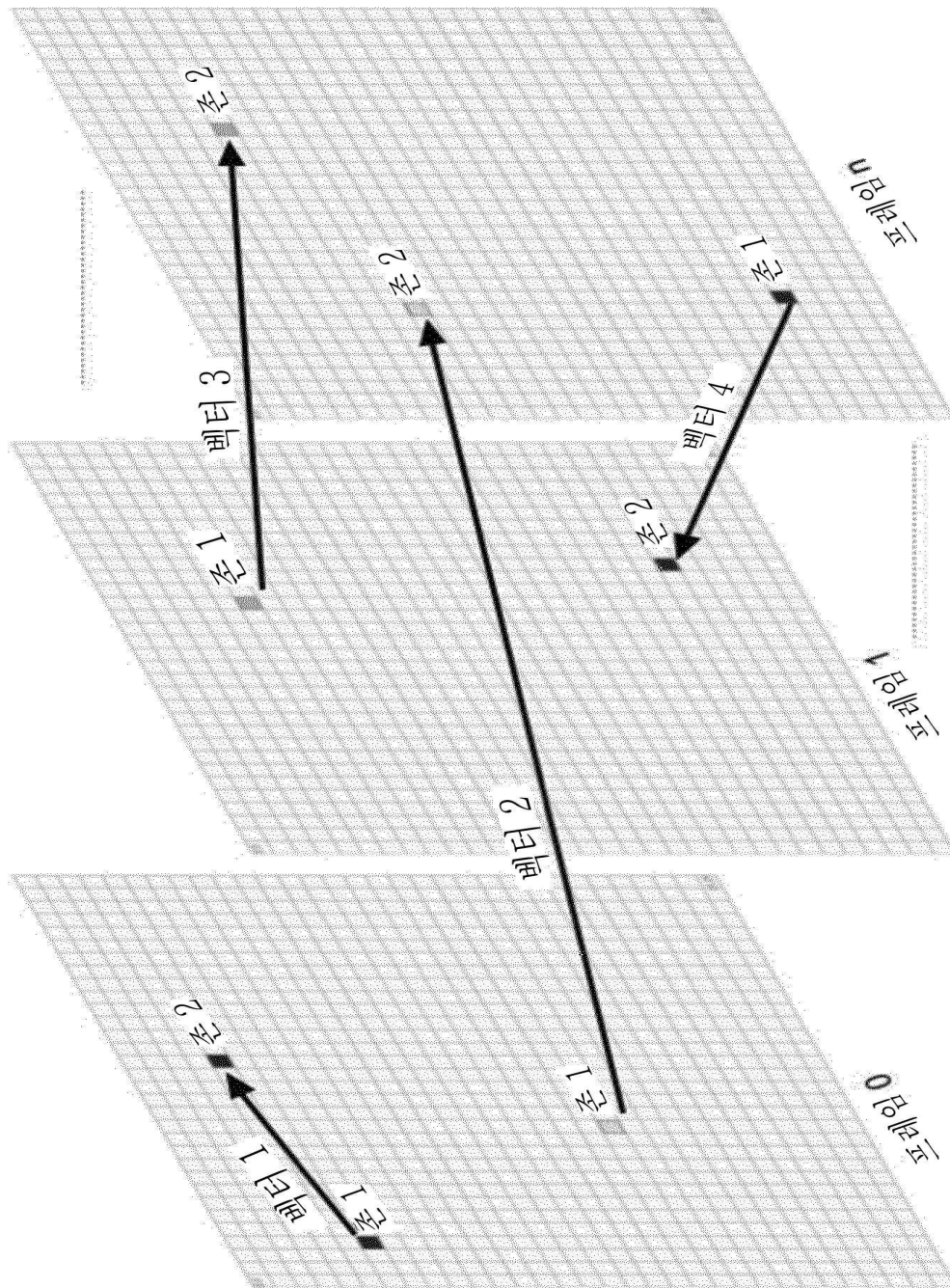


도면16

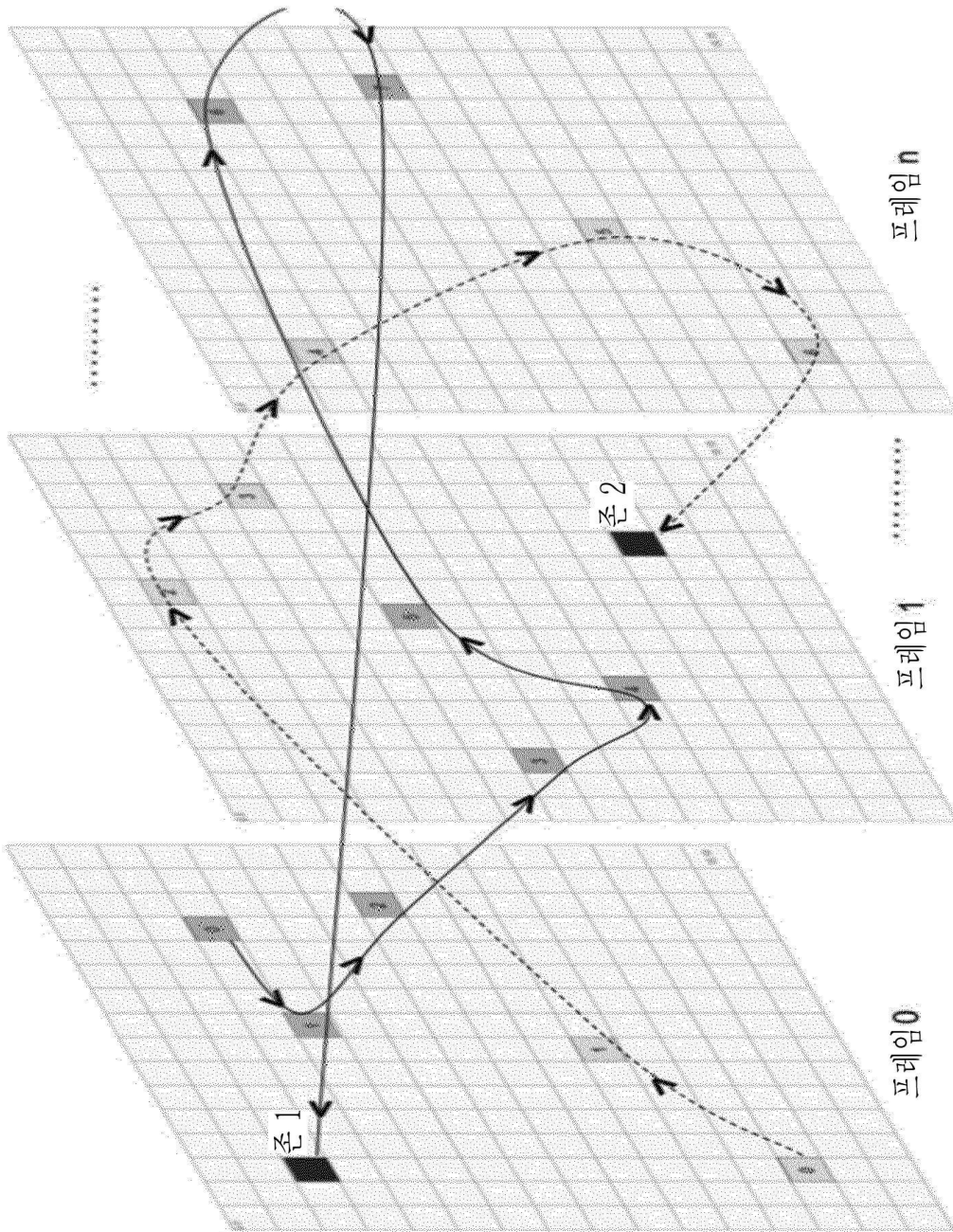
임의의 이미지 형상은 입력으로서 사용될 수 있음



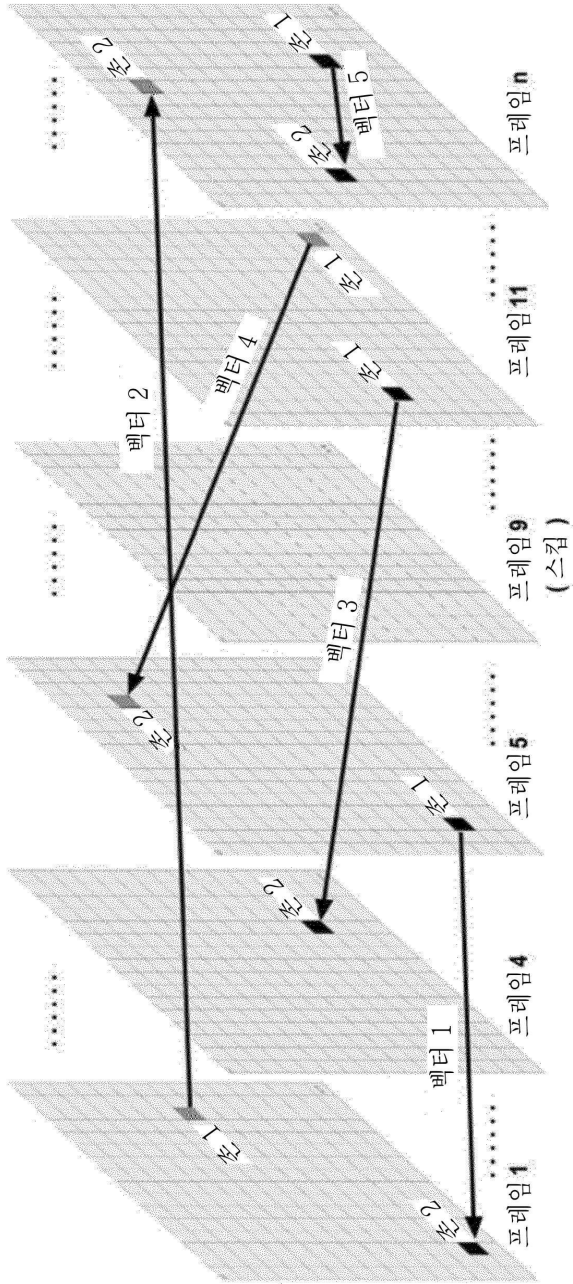
도면17



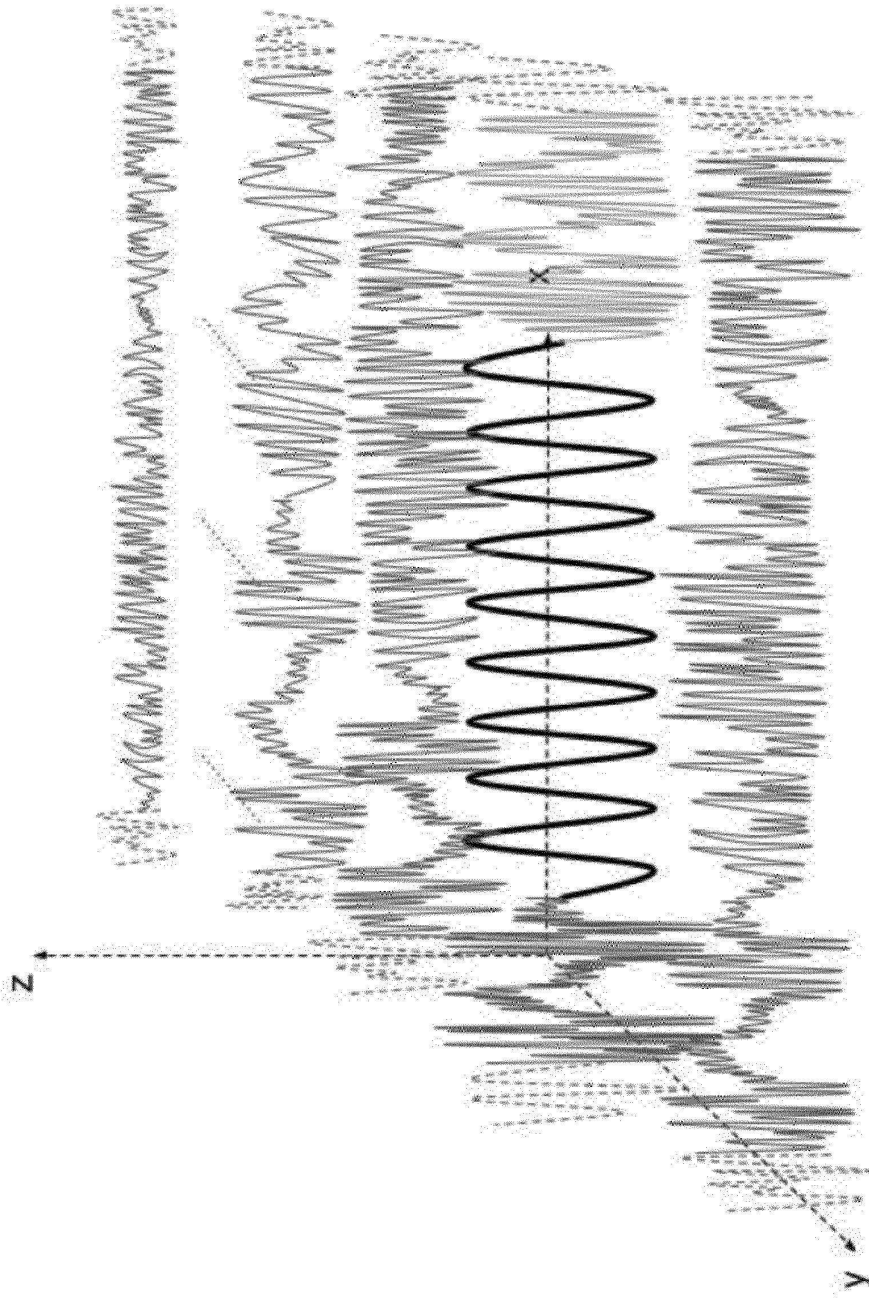
도면18



도면19



도면20



도면21

