



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2025년02월13일
(11) 등록번호 10-2767441
(24) 등록일자 2025년02월10일

(51) 국제특허분류(Int. Cl.)
G06V 10/764 (2022.01) G01N 21/64 (2006.01)
G06N 20/20 (2019.01) G06N 5/00 (2023.01)
G06N 5/04 (2023.01) G06T 5/20 (2024.01)
G06T 7/00 (2017.01) G06V 10/28 (2022.01)
G06V 10/50 (2022.01) G06V 10/774 (2022.01)
G06V 20/69 (2022.01)

(52) CPC특허분류
G06V 10/765 (2023.08)
G01N 21/6428 (2013.01)

(21) 출원번호 10-2022-7026377

(22) 출원일자(국제) 2021년01월29일
심사청구일자 2024년01월26일

(85) 번역문제출일자 2022년07월28일

(65) 공개번호 10-2022-0134752

(43) 공개일자 2022년10월05일

(86) 국제출원번호 PCT/US2021/015906

(87) 국제공개번호 WO 2021/155291

국제공개일자 2021년08월05일

(30) 우선권주장

62/968,950 2020년01월31일 미국(US)

17/161,595 2021년01월28일 미국(US)

(56) 선행기술조사문헌

WO2018140014 A1*

JP2018005640 A*

Ching Yu Austin Huang외 10명,
“Auto-validation of fluorescent primer
extension genotyping assay using signal
clustering and neural networks”, BMC
Bioinformatics, Vol.5, pp.1-10(2004.04.02.)
1부.*

*는 심사관에 의하여 인용된 문헌

(73) 특허권자

일루미나, 인코포레이티드

미국 캘리포니아 92122 샌디에고 일루미나 웨이
5200

(72) 발명자

레자에이 나그메

미국 캘리포니아주 94062 레드우드 시티 410 토파
즈 스트리트

필리페 크루즈 페드로 미겔

미국 캘리포니아주 92129 샌디에이고 9474 버베인
스트리트

(74) 대리인

특허법인아주김장리

전체 청구항 수 : 총 17 항

심사관 : 정수진

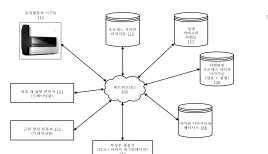
(54) 발명의 명칭 프로세스 사이클 이미지들의 기계 학습 기반 근본 원인 분석

(57) 요약

개시된 기술은 프로세스 사이클들의 성공 또는 실패를 예측하기 위한 프로세스 사이클 이미지들의 분류에 관한 것이다. 개시된 기술은 유전형분석 프로세스에서 이미지 생성 칩 상에 배열된 섹션들의 이미지들을 캡처하고 프로세싱하는 것을 포함한다. 생성 사이클 이미지들의 이미지 설명 특징부들은 분류자들에 대한 입력으로서 생성

(뒷면에 계속)

대표도



되고 주어진다. 트레이닝된 분류자는 성공적인 생성 이미지들을 성공적이지 못한 또는 실패한 생성 이미지들과 분리한다. 실패한 생성 이미지들은 트레이닝된 근본 원인 분류자에 의해 다양한 실패 카테고리들로 추가로 분류된다.

(52) CPC특허분류

G06N 20/20 (2021.08)

G06N 5/01 (2023.01)

G06N 5/04 (2023.01)

G06T 5/20 (2024.01)

G06T 7/0012 (2013.01)

G06V 10/28 (2023.08)

G06V 10/507 (2023.08)

G06V 10/7747 (2023.08)

G06V 20/69 (2022.01)

명세서

청구범위

청구항 1

프로세서를 포함하는 컴퓨팅 디바이스에서 수행되는 유전형분석(genotyping) 프로세스 사이클 이미지에 대한 랜덤 포레스트 분류자(random forest classifier)를 트레이닝시키는 방법으로서,

상기 프로세서가 성공 클래스 및 다수의 실패 클래스에 속하는 프로세스 사이클로부터의 이미지에 대한 라벨링된 트레이닝 예에 액세스하는 단계로서, 각 실패 클래스는 실패의 상이한 근본 원인에 대응하는, 상기 액세스하는 단계;

상기 프로세서가 아이겐 이미지(Eigen image)의 선형 조합에 기초하여 각각의 라벨링된 트레이닝 예에 대한 이미지 설명 특징부를 생성하는 단계;

상기 프로세서가 상기 라벨링된 트레이닝 예의 이미지 설명 특징부를 사용하여 상기 랜덤 포레스트 분류자를 트레이닝시켜 상기 다수의 실패 클래스에 대응하는 특정 특징부를 식별하는 단계; 및

상기 프로세서가 트레이닝된 랜덤 포레스트 분류자의 파라미터를 저장하는 단계

를 포함하는, 방법.

청구항 2

제1항에 있어서,

상기 프로세서가 아이겐 이미지의 베이스(basis)에 액세스하는 단계;

상기 프로세서가 설명된 가변성(variability)의 측정치에 따라 상기 아이겐 이미지의 베이스를 정렬하는 단계;

상기 프로세서가 임계치 초과된 가변성을 점증적으로 설명하는 아이겐 이미지의 상단 정렬된 베이스를 선택하는 단계; 및

상기 프로세서가 상기 아이겐 이미지의 선택된 베이스를 사용하여 상기 프로세스 사이클 이미지를 분석하는 단계

를 추가로 포함하는, 방법.

청구항 3

제1항에 있어서, 상기 성공 클래스 대 상기 다수의 실패 클래스의 1대-나머지(one-vs-the-rest) 결정을 위하여 상기 프로세서가 상기 이미지 설명 특징부를 사용하여 상기 랜덤 포레스트 분류자를 트레이닝시키는 단계를 추가로 포함하는, 방법.

청구항 4

제1항에 있어서, 상기 랜덤 포레스트 분류자는 100 내지 400개의 결정 트리를 포함하는, 방법.

청구항 5

제1항에 있어서, 상기 랜덤 포레스트 분류자는 10 내지 40의 깊이를 갖는, 방법.

청구항 6

제2항에 있어서,

상기 이미지 설명 특징부가 상기 성공 클래스에 속하지 않는다는 결정 시에 상기 프로세서가 제2 랜덤 포레스트 분류자에 액세스하는 단계로서, 상기 제2 랜덤 포레스트 분류자는 이미지를 상기 다수의 실패 클래스 중 하나에 속하는 프로세스 사이클과 구별하기 위해 트레이닝되는, 상기 액세스하는 단계; 및

상기 프로세서가 상기 제2 랜덤 포레스트 분류자를 상기 이미지 설명 특징부에 적용하는 단계로서, 상기 다수의 실패 클래스 각각 대 상기 성공 클래스와 상기 다수의 실패 클래스에 속하지 않는 나머지를 스코어링하는 것과, 생성된 스코어를 사용하여 실패한 프로세스 사이클의 가능성 있는 근본 원인으로 상기 다수의 실패 클래스 사이에서 선택하는 것을 포함하는, 상기 적용하는 단계

를 추가로 포함하는, 방법.

청구항 7

유전형분석 프로세스 사이클 이미지에 대한 랜덤 포레스트 분류자를 트레이닝시키기 위해 컴퓨터 프로그램 명령어들이 저장된 비밀시적 컴퓨터 판독가능 저장 매체로서,

상기 명령어는, 프로세서상에서 실행될 때, 상기 프로세서가,

성공 클래스 및 다수의 실패 클래스에 속하는 프로세스 사이클로부터의 이미지에 대한 라벨링된 트레이닝 예에 액세스하게 하되, 각 실패 클래스는 실패의 상이한 근본 원인에 대응하고;

아이젠 이미지의 선형 조합에 기초하여 각각의 라벨링된 트레이닝 예에 대한 이미지 설명 특징부를 생성하게 하고;

상기 라벨링된 트레이닝 예의 이미지 설명 특징부를 사용하여 상기 랜덤 포레스트 분류자를 트레이닝시켜 상기 다수의 실패 클래스에 대응하는 특정 특징부를 식별하게 하고; 그리고

트레이닝된 랜덤 포레스트 분류자의 파라미터를 저장하게 하는, 비밀시적 컴퓨터 판독가능 저장 매체.

청구항 8

제7항에 있어서, 상기 명령어는 추가로 상기 프로세서가,

상기 성공 클래스 대 상기 다수의 실패 클래스의 1-대-나머지 결정을 위한 상기 이미지 설명 특징부를 사용하여 상기 랜덤 포레스트 분류자를 트레이닝시키게 하는, 비밀시적 컴퓨터 판독가능 저장 매체.

청구항 9

제7항에 있어서, 상기 명령어는,

아이젠 이미지의 베이스스에 액세스하게 하고;

설명된 가변성의 측정치에 따라 상기 아이젠 이미지의 베이스스를 정렬하게 하고;

임계치 초과된 가변성을 점증적으로 설명하는 아이젠 이미지의 상단 정렬된 베이스스를 선택하게 하고; 그리고

상기 아이젠 이미지의 선택된 베이스스를 사용하여 상기 프로세스 사이클 이미지를 분석하게 하도록 추가로 구성되는, 비밀시적 컴퓨터 판독가능 저장 매체.

청구항 10

제7항에 있어서, 상기 랜덤 포레스트 분류자는 100 내지 400개의 결정 트리를 포함하는, 비밀시적 컴퓨터 판독가능 저장 매체.

청구항 11

제7항에 있어서, 상기 랜덤 포레스트 분류자는 10 내지 40의 깊이를 갖는, 비밀시적 컴퓨터 판독가능 저장 매체.

청구항 12

제9항에 있어서, 상기 명령어는 추가로 상기 프로세서가,

상기 이미지 설명 특징부가 상기 성공 클래스에 속하지 않는다는 결정 시에 제2 랜덤 포레스트 분류자에 액세스하게 하도록 구성되되, 상기 제2 랜덤 포레스트 분류자는 이미지를 상기 다수의 실패 클래스 중 하나에 속하는 프로세스 사이클과 구별하기 위해 트레이닝되고; 그리고

상기 제2 랜덤 포레스트 분류자를 상기 이미지 설명 특징부에 적용하게 하도록 구성되되, 상기 프로세서가 상기 다수의 실패 클래스 각각 대 상기 성공 클래스와 상기 다수의 실패 클래스에 속하지 않는 나머지를 스코어링하고, 생성된 스코어를 사용하여 실패한 프로세스 사이클의 가능성 있는 근본 원인으로써 다수의 실패 클래스 사이에서 선택하도록 추가로 구성된 것을 포함하는, 비밀시적 컴퓨터 판독가능 저장 매체.

청구항 13

메모리에 결합된 하나 이상의 프로세서를 포함하는 시스템으로서,

상기 메모리에는 유전형분석 프로세스 사이클 이미지에 대한 랜덤 포레스트 분류자를 트레이닝시키기 위해 컴퓨터 명령어들이 로딩되고, 상기 명령어는, 상기 하나 이상의 프로세서상에서 실행될 때, 상기 하나 이상의 프로세서가,

성공 클래스 및 다수의 실패 클래스에 속하는 프로세스 사이클로부터의 이미지에 대한 라벨링된 트레이닝 예에 액세스하게 하되, 각 실패 클래스는 실패의 상이한 근본 원인에 대응하고;

아이젠 이미지의 선형 조합에 기초하여 각각의 라벨링된 트레이닝 예에 대한 이미지 설명 특징부를 생성하게 하고;

상기 라벨링된 트레이닝 예의 이미지 설명 특징부를 사용하여 상기 랜덤 포레스트 분류자를 트레이닝시켜 상기 다수의 실패 클래스에 대응하는 특정 특징부를 식별하게 하고; 그리고

트레이닝된 랜덤 포레스트 분류자의 파라미터를 저장하게 하는, 시스템.

청구항 14

제13항에 있어서, 상기 명령어는 추가로 상기 하나 이상의 프로세서가,

아이젠 이미지의 베이스스에 액세스하게 하고;

설명된 가변성의 측정치에 따라 상기 아이젠 이미지의 베이스스를 정렬하게 하고;

임계치 초과된 가변성을 점증적으로 설명하는 아이젠 이미지의 상단 정렬된 베이스스를 선택하게 하고; 그리고

상기 아이젠 이미지의 선택된 베이스스를 사용하여 상기 프로세스 사이클 이미지를 분석하게 하는, 시스템.

청구항 15

제13항에 있어서, 상기 랜덤 포레스트 분류자는 100 내지 400개의 결정 트리를 포함하는, 시스템.

청구항 16

제13항에 있어서, 상기 랜덤 포레스트 분류자는 10 내지 40의 깊이를 갖는, 시스템.

청구항 17

제13항에 있어서, 상기 명령어는 추가로 상기 하나 이상의 프로세서가,

상기 이미지 설명 특징부가 상기 성공 클래스에 속하지 않는다는 결정 시에 제2 랜덤 포레스트 분류자에 액세스하게 하되, 상기 제2 랜덤 포레스트 분류자는 이미지를 상기 다수의 실패 클래스 중 하나에 속하는 프로세스 사이클들과 구별하기 위해 트레이닝되고; 그리고

상기 제2 랜덤 포레스트 분류자를 상기 이미지 설명 특징부에 적용하게 하되, 상기 하나 이상의 프로세서가 상기 다수의 실패 클래스 각각 대 상기 성공 클래스와 상기 다수의 실패 클래스에 속하지 않는 나머지를 스코어링하고, 생성된 스코어를 사용하여 실패한 프로세스 사이클의 가능성 있는 근본 원인으로써 다수의 실패 클래스 사이에서 선택하도록 구성된 것을 포함하는, 시스템.

청구항 18

삭제

청구항 19

삭제

청구항 20

삭제

발명의 설명

기술 분야

[0001]

우선권 출원

[0002]

본 출원은 2021년 1월 28일자로 출원되고 발명의 명칭이 "MACHINE LEARNING-BASED ROOT CAUSE ANALYSIS OF PROCESS CYCLE IMAGES"인 미국 특허 출원 제17/161,595호(대리인 문서 번호 ILLM 1026-2/IP-1911-US)의 이익을 주장하며, 이는 2020년 1월 31일자로 출원되고 발명의 명칭이 "MACHINE LEARNING-BASED ROOT CAUSE ANALYSIS OF PROCESS CYCLE IMAGES"인 미국 가특허 출원 제62/968,950호(대리인 문서 번호 ILLM 1026-1/IP-1911-PRV)의 이익을 주장한다. 우선권 출원들은 모든 목적들을 위해 참고로 포함된다.

[0003]

개시된 기술의 분야

[0004]

개시된 기술은 생성 프로세스들의 평가 및 근본 원인 실패 분석을 위한 이미지들의 분류에 관한 것이다.

배경 기술

[0005]

이 섹션에서 논의되는 주제는 단지 이 섹션에서의 언급의 결과만으로 종래 기술이라고 가정되어서는 안 된다. 유사하게, 이 섹션에서 언급되거나 배경기술로서 제공되는 주제와 연관된 문제는 종래기술에서 이전에 인식되었다고 가정되어서는 안 된다. 이 섹션에서의 주제는 단지 상이한 접근법들을 표현하는 것일 뿐이고, 접근법들 자체도 또한 청구되는 기술의 구현예들에 해당할 수 있다.

[0006]

유전형분석(genotyping)은 완료하는 데 며칠이 걸릴 수 있는 프로세스이다. 프로세스는 기계적 프로세싱 오류 및 화학적 프로세싱 오류 둘 모두에 취약하다. 유전형분석을 위한 수집된 샘플들은 추출되어 이미지 생성 칩들의 섹션들 및 영역들에 분포된다. 이어서, 샘플들은 형광 이미지들을 생성하기 위해 다수의 단계들을 통해 화학적으로 프로세싱된다. 프로세스는 분석된 각 섹션에 대한 품질 스코어를 생성한다. 이러한 품질은 저품질 프로세스에서의 실패의 근본 원인에 대한 통찰력을 제공할 수 없다. 일부 경우들에 있어서, 실패한 섹션 이미지는 수용가능한 품질 스코어를 여전히 생성한다.

[0007]

따라서, 생성물 유전형분석 동안 섹션 이미지들을 평가하고 실패 분석의 근본 원인들을 결정하기 위한 새로운 방법들 및 시스템들을 도입할 기회가 생긴다.

도면의 간단한 설명

[0008]

도면에서, 유사한 도면 부호는 일반적으로 상이한 도면들 전체에 걸쳐서 유사한 부분들을 지칭한다. 또한, 도면은 반드시 실적이 아니며, 그 대신에 일반적으로는 개시된 기술의 원리를 예시하는 데 중점을 둔다. 아래의 설명에서, 개시된 기술의 다양한 구현예가 아래의 도면을 참조하여 설명된다.

도 1은 유전형분석 기구들로부터의 프로세스 사이클 이미지들이 분류되고 불량한 이미지들의 근본 원인이 결정되는 시스템의 아키텍처 레벨 개략도를 도시한다.

도 2는 도 1의 특징부 생성기의 서브시스템 컴포넌트들을 도시한다.

도 3은 예시적인 유전형분석 프로세스에 대한 프로세스 단계들을 도시한다.

도 4는 성공적인 프로세스 완료 후 이미지 생성에서 배열된 섹션들의 이미지들을 도시한다.

도 5a 및 도 5b는 유전형분석 프로세스 동안의 혼성화(hybridization) 실패로 인한 실패한 섹션 이미지들의 예들을 도시한다.

도 5c 및 도 5d는 스페이서 시프트 실패들로 인한 실패한 섹션 이미지들의 예들을 도시한다.

도 5e는 오프셋 실패들로 인한 실패한 섹션 이미지들의 예들을 도시한다.

- 도 5f는 표면 마모 실패로 인한 실패한 섹션 이미지들의 예들을 도시한다.
- 도 5g 및 도 5h는 시약 흐름 실패로 인한 실패한 섹션 이미지들의 예들을 도시한다.
- 도 5i는 실패의 소스가 알려지지 않은 실패한 또는 비정상적인 섹션 이미지들의 예들을 도시한다.
- 도 6a 및 도 6b는 주성분 분석(Principal Component Analysis, PCA)을 사용하여 생성된 주성분들을 랭크 정렬함으로써 선택된 96개의 아이겐 이미지(Eigen image) 성분들의 베이스(basis)의 예들이다.
- 도 6c 및 도 6d는 도 6a 및 도 6b에서의 96개의 이미지들로부터 선택된 상단 40개의 아이겐 이미지 성분들의 확대본이다.
- 도 7a는 주성분 분석에 대한 입력을 위한 섹션 이미지들의 이미지 리스케일링 및 평탄화를 도시한다.
- 도 7b는 주성분 분석을 사용한 아이겐 이미지들의 베이스의 생성을 도시한다.
- 도 8a는 섹션들의 영역들의 세기들을 사용한 특징부 생성의 일례를 도시한다.
- 도 8b는 섹션들의 영역들의 세기들을 사용한 특징부 생성의 다른 예를 도시한다.
- 도 9는 1-대-나머지(one-vs-the-rest, OvR) 분류자의 그래픽 예시이다.
- 도 10a는 프로세스 사이클 이미지들을 포함하는 라벨링된 트레이닝 데이터를 사용한 이진(양호 대 불량) 분류자 및 다중클래스(근본 원인) 분류자의 트레이닝을 도시한다.
- 도 10b는 생성 프로세스 사이클 이미지들이 양호 대 불량으로 분류되고, 추가로, 불량한 이미지들의 실패 카테고리 결정되는 2-단계 프로세스를 도시한다.
- 도 11은 개시된 기술을 구현하는 데 사용될 수 있는 컴퓨터 시스템의 단순화된 블록도이다.

발명을 실시하기 위한 구체적인 내용

[0009] 아래의 논의는 어느 당업자라도 개시된 기술을 제조하고 사용할 수 있게 하도록 제시되며, 특정의 응용 및 그의 요건과 관련하여 제공된다. 개시된 구현예에 대한 다양한 변형은 당업자에게 용이하게 명백할 것이며, 본원에서 정의된 일반적인 원리는 개시된 기술의 사상 및 범주로부터 벗어남이 없이 다른 구현예 및 응용에 적용될 수 있다. 따라서, 개시된 기술은 도시된 구현예로 제한되도록 의도된 것이 아니라, 본원에 개시된 원리 및 특징과 일치하는 가장 넓은 범주에 부합되어야 한다.

[0010] **서문**

[0011] 개시된 기술은 생성물 유전형분석의 평가 및 근본 원인 실패 분석을 위한 비전 시스템들 및 이미지 분류를 적용한다. 2개의 별개의 접근법들, 즉 아이겐 이미지들을 수반하는 접근법 및 영역에 의한 임계화(thresholding)에 기초한 다른 접근법이 기술된다. 개시된 기법들 중에 주성분 분석(PCA) 및 비-음수 행렬 인수분해(non-negative matrix factorization, NMF)가 있다. 이미지들에 적용될 수 있는 다른 차원수 감소 기법들은 독립 성분 분석, 사전 학습, 최소 주성분 분석, 요인 분석(factor analysis), 미니 배치 K-평균(mini-batch K-means)을 포함한다. 이미지 분해 및 차원수 감소 기법들의 변형들이 사용될 수 있다. 예를 들어, PCA는 특이 값 분해(singular value decomposition, SVD)를 사용하여 또는 커널 PCA로서 구현될 수 있다. 이러한 기법들로부터의 출력들은 분류자들에 대한 입력들로서 주어진다. 적용되는 분류자들은 랜덤 포레스트(random forest), K-최근접 이웃(K-nearest neighbour, KNN), 다항 로지스틱 회귀(multinomial logistic regression), 서포트 벡터 머신(support vector machine, SVM), 기울기 증대 트리(gradient boosted tree), 나이브 베이즈(Naive Bayes) 등을 포함할 수 있다. 라벨링된 이미지들의 더 큰 본체들이 이용가능해짐에 따라, 콘볼루션 신경 네트워크들, 예컨대 ImageNet가 또한 사용될 수 있다.

[0012] 이러한 생성 프로세스는 기계적 프로세싱 오류 및 화학적 프로세싱 오류 둘 모두에 취약하다. 수집된 샘플들은 추출되고, BeadChip들의 섹션들 및 영역들에 분포되고, 이어서, 형광 이미지들을 생성하기 위해 다수의 단계들을 통해 화학적으로 프로세싱된다. 생성을 모니터링하고 실패 분석을 이행하기 위해 최종 형광 이미지 또는 심지어 중간 형광 이미지들이 분석될 수 있다.

[0013] 대부분의 생성물 분석들은 성공적이다. 현재, 실패한 분석들은 5개의 카테고리 플러스 잔류 실패 카테고리에 피팅되는 것으로 이해된다. 5개의 실패 카테고리들은 혼성화 또는 hyb 실패들, 스페이서 시프트 실패들, 오프

셋 실패들, 표면 마모 실패들 및 시약 흐름 실패들이다. 잔류 카테고리는 혼합된 효과들, 식별되지 않은 원인들 및 약한 신호들로 인한 비정상적인 패턴들이다. 시간상, 특히 근본 원인 분석이 개선된 생성으로 이어짐에 따라, 더 많은 상이한 원인들이 식별될 수 있다.

[0014] 품질 제어 및 실패 분석에 적용된 제1 이미지 프로세싱 기술은 아이겐 얼굴 분석에 의한 얼굴 인식으로부터 진화된다. 수만 개의 라벨링된 이미지들로부터, 40 내지 100개 이상의 이미지 성분들의 선형 베이스스를 식별하였다. 아이겐 베이스스를 형성하는 한 가지 접근법은 주성분 분석(PCA)에 뒤이어, 설명된 가변성의 측정치에 따른 성분들의 랭크 정렬이었다. 40개의 성분들이 가변성의 대부분을 설명한다는 것을 관찰하였다. 100개의 성분들을 넘어, 추가적인 성분들은 샘플 프로세싱에서 자연적 가변성 또는 잡음의 패턴들을 반영하는 것으로 나타났다. 관련 성분들의 수는 이미지 해상도에 의해 영향을 받을 것으로 예상된다. 여기서, 이미지 생성 칩의 섹션들이 180 x 80 픽셀들의 해상도에서 분석되도록 해상도 감소를 적용하였다. 이는, 성공적인 생성을 성공적이지 못한 생성과 구별하고 이어서 6개의 실패 카테고리들 중에서 실패의 근본 원인들을 분류하기에 충분한 해상도였다. 어떠한 형식적 감도 분석도 적용하지 않았지만, 증가된 계산 비용을 가짐에도 불구하고, 약간 더 낮은 해상도 이미지들이 또한 작용할 것이고 이러한 해상도의 4 내지 22배를 갖는 이미지들이 동일한 방식으로 프로세싱될 수 있다는 것이 예상된다. 아이겐 이미지 분석에 의해 분석될 각각의 이미지는 베이스스 이미지들의 가중 선형 조합으로 표현된다. 정렬된 세트의 베이스스 성분들에 대한 각각의 가중치가 분류자를 트레이닝시키기 위한 특징부로서 사용된다. 예를 들어, 하나의 구현예에서, 라벨링된 이미지들의 성분들에 대한 96개의 가중치들을 사용하여 랜덤 포레스트 분류자들을 트레이닝시켰다. 200개의 트리들 및 20의 깊이를 갖는 랜덤 포레스트 분류자가 잘 작동하였다. 랜덤 포레스트 분류자들에 의해 2개의 태스크들을 수행하였다: 성공적인 생성 이미지들과 성공적이지 못한 생성 이미지들의 분리, 이어서 성공적이지 못한 생성 이미지들의 근본 원인 분석. 성공적인 생성물 런들의 우위(dominance)로 인해 이러한 2-단계 분류를 선택하였지만, 1-단계 분류가 또한 사용될 수 있었다.

[0015] 적용된 제2 이미지 프로세싱 기술은 이미지 영역들의 임계화를 수반하였다. 이미지 생성 칩의 섹션의 생성 이미지가 수 개의 물리적으로 분리된 영역들을 캡처한다. 섹션의 경계를 이루고 섹션의 물리적 영역들을 분리하는 구조들이 생성 이미지에서 가시적이다. 임계화 전략은 활성 영역들을 경계 구조들과 분리하는 것 그리고 이어서 분리된 영역들 사이를 구별하는 것을 수반한다. 선택적으로, 물리적 영역들을 분리하는 구조들은 또한 이미지로부터 필터링될 수 있다. 적어도 활성 영역들은 발광에 대한 임계화의 대상이다. 임계화는 활성 영역의 얼마나 많은 부분이 원하는 신호 강도를 생성하고 있는지를 결정한다. 각각의 활성 영역은 성공 또는 실패에 대한 임계화 후에 평가된다. 이미지 생성 칩의 영역들 및 섹션들 사이에서의 실패들의 패턴은 근본 원인 분류를 위해 추가로 평가될 수 있다.

[0016] 실패한 생성물 런들을 검출하고 근본 원인들을 결정하기 위한 생성 이미지들의 프로세싱은 생성 동안 즉시 수행될 수 있고, 이미지 생성 칩으로부터 결과들이 판독되고 품질에 대해 판단되는 것보다도 더 신속하게 수행될 수 있다. 이러한 이미지 프로세싱은 더 신속하게 행해질 수 있는데, 그 이유는 일 측면 상의 이미지를 20배만큼 축소시키는 것이 계산 요건들을 크게 감소시키고, 감소된 해상도 이미지의 직접 프로세싱이 개별 프로브들에 대한 영역 내의 개별 성장 픽셀들의 상관성을 요구하지 않기 때문이다. 근본 원인 분석의 신속한 터어라운드에는 화학물질 및 프로세싱 시간이 낭비되기 전에 업스트림 프로세스들을 보정하는 데 사용될 수 있다.

[0017] **환경**

[0018] 유전형분석 시스템들에서 실패의 조기 예측을 위한 시스템을 기술한다. 유전형분석은 생물 검정들을 사용하여 개체의 DNA 서열을 검사하고 이를 기준 서열과 비교함으로써 개체의 유전자 구성(유전형)에서의 차이들을 결정하는 프로세스이다. 유전형분석은 연구자들이 단일 뉴클레오타이드 다형성(single nucleotide polymorphism, SNP)들과 같은 유전자 변이체들 및 DNA의 구조적 변화들을 탐구할 수 있게 한다. 시스템은, 일 구현예에 따른 시스템의 아키텍처 레벨의 개략도를 도시한 도 1을 참조하여 기술된다. 도 1이 아키텍처 도면이기 때문에, 설명의 명료성을 개선하기 위해 소정 세부사항들이 의도적으로 생략된다. 도 1의 논의는 하기와 같이 구성된다. 먼저, 도면의 요소들이 기술되고, 뒤이어, 그들의 상호접속이 기술된다. 이어서, 시스템 내의 요소들의 사용이 더 상세히 기술된다.

[0019] 도 1은 시스템(100)을 포함한다. 본 단락은 시스템(100)의 라벨링된 부분들을 명명한다. 도면은 유전형분석 기구들(111), 프로세스 사이클 이미지 데이터베이스(115), 실패 카테고리 라벨 데이터베이스(117), 라벨링된 프로세스 사이클 이미지 데이터베이스(138), 트레이닝된 양호 대 불량 분류자(151), 아이겐 이미지 베이스스 데이터베이스(168), 트레이닝된 근본 원인 분류자(171), 특징부 생성기(185), 및 네트워크(들)(155)를 도시한다.

- [0020] 개시된 기술은 유전형분석 스캐너들 및 유전형분석 플랫폼들로도 지칭되는, 다양한 유전형분석 기구들(111)에 적용된다. 네트워크(들)(155)는 서로 통신하는 유전형분석 기구들(111), 프로세스 사이클 이미지 데이터베이스(115), 실패 카테고리 라벨 데이터베이스(117), 라벨링된 프로세스 사이클 이미지 데이터베이스(138), 트레이닝된 양호 대 불량 분류자(151), 아이겐 이미지 베이스 데이터베이스(168), 트레이닝된 근본 원인 분류자(171), 및 특징부 생성기(185)를 커플링시킨다.
- [0021] 유전형분석 기구들은 ISCAN™ 시스템과 같은 Illumina의 BeadChip 이미징 시스템들을 포함할 수 있다. 기구는 이미지 생성 칩들 상의 맵핑된 위치들에 있는 색선들에 배열된 수백 내지 수백만 개의 비드들의 형광 세기들을 검출할 수 있다. 유전형분석 기구들은 기구의 다양한 태양들, 예를 들어, 레이저 제어, 정밀 기계 제어, 여기 신호들의 검출, 이미지 등록, 이미지 추출, 및 데이터 출력을 제어하는 기구 제어 컴퓨터를 포함할 수 있다. 유전형분석 기구들은 다양한 물리적 환경에서 사용될 수 있고, 다양한 기량 레벨들의 기술자들에 의해 작동될 수 있다. 샘플 준비는 2 내지 3일이 걸릴 수 있고, 샘플들의 수동 및 자동 핸들링을 포함할 수 있다.
- [0022] 도 3의 예시적인 유전형분석 프로세스(300)의 프로세스 단계들을 도시한다. 이러한 예시적인 유전형분석 프로세스는 Illumina의 INFINIUM™ Assay Workflow로 지칭된다. 프로세스는 유전자좌(locus) 멀티플렉싱의 광범위한 레벨들에서 많은 SNP들을 조사하도록 설계된다. 단일 비드 유형 및 이중 색상(예컨대, 적색 및 녹색) 채널 접근법을 사용하여, 프로세스는 샘플당 수백 내지 수백만 개의 SNP들로부터의 유전형분석을 스케일링한다. 프로세스는 DNA 샘플들의 수탁 및 추출로 시작한다. 프로세스는 수백만 개의 SNP 유전자좌들을 검정할 수 있는 200 ng과 같은 비교적 낮은 입력 샘플로 동작할 수 있다. 샘플들이 증폭된다. 증폭 프로세스는 완료하는 데 수 시간 내지 하룻밤(overnight)이 걸릴 수 있다. 증폭된 샘플은 제어된 효소 단편화를 거친다. 이에 뒤이어 알코올 침전 및 재현탁이 이어진다. 이미지 생성 칩은 모세관 관류(flow-through) 챔버에서의 혼성화를 위해 준비된다. 이어서, 샘플들이 준비된 이미지 생성 칩들에 적용되고, 밤새 배양된다. 이러한 밤샘(overnight) 혼성화 동안, 샘플들은 수백만 개에 이르는 비드 유형들에 공유 결합된 유전자좌 특이적 50-mer들에 어닐링된다. 하나의 비드 유형은 SNP 유전자좌마다의 각각의 대립유전자에 대응한다. 대립유전자 특이성은 효소 염기 확장에 뒤이어 형광 염색에 의해 부여된다. 유전형분석 기구 또는 스캐너(예컨대, ISCAN™ 시스템)는 비드들의 형광 세기들을 검출하고, 유전형 호출을 수행한다.
- [0023] 하나의 예에서, 유전형분석의 결과들은 "호출 레이트(Call Rate)"로 불리는 메트릭을 사용하여 제시된다. 이러한 메트릭은 이미지 생성 칩 상에서 정확하게 스캔된 유전형들의 비율을 표현한다. 이미지 생성 칩의 섹션마다 별도의 호출 레이트가 보고된다. 그 결과들을 수용하거나 거절하기 위해 임계치가 사용될 수 있다. 예를 들어, 섹션에 대한 유전형분석 결과들을 수용하기 위해 98% 이상의 호출 레이트가 사용될 수 있다. 98% 미만 또는 98% 초과와 같은 상이한 임계 값이 사용될 수 있다. 섹션에 대한 호출 레이트가 임계치 미만인 경우, 유전형분석 프로세스는 실패한 것으로 간주된다. 유전형분석 프로세스는 수 일에 걸쳐 이어질 수 있고, 따라서, 반복하기에는 고가이다. 유전형분석 프로세스에서의 실패들은 동작 오류들(예컨대, 기계적 또는 핸들링 오류들) 또는 화학적 프로세싱 오류들로 인해 발생할 수 있다.
- [0024] 유전형분석 시스템들은 유전형분석 프로세스의 완료 시에 그들의 호출 레이트들과 함께 이미지 생성 칩 섹션들의 프로세스 사이클 이미지들을 제공할 수 있다. 개시된 기술은 이러한 섹션 이미지들을 프로세싱하여 유전형분석 프로세스가 성공적인지(양호한 섹션 이미지) 아니면 성공적이지 않은지(불량한 또는 실패한 섹션 이미지)를 분류할 수 있다. 개시된 기술은 불량한 또는 실패한 이미지들을 추가로 프로세싱하여 실패의 카테고리를 결정할 수 있다. 현재, 시스템은 하기의 6개의 실패 카테고리들 중 하나의 실패 카테고리에서의 실패한 이미지들을 분류할 수 있다: 혼성화 또는 hyb 실패들, 스페이서 시프트 실패들, 오프셋 실패들, 표면 마모 실패들, 시약 흐름 실패들, 및 혼합 효과들, 미지의 원인들, 약한 신호들 등으로 인한 전체적인 비정상적인 이미지들. 시간상, 특히 근본 원인 분석이 개선된 생성으로 이어짐에 따라, 더 많은 그리고 상이한 원인들이 식별될 수 있다.
- [0025] 이제, 도 1을 참조하여, 시스템(100)의 나머지 컴포넌트들에 대한 설명을 제공한다. 6개의 실패 유형들에 대한 실패 카테고리 라벨들은 실패 카테고리 라벨 데이터베이스(117)에 저장될 수 있다. 라벨링된 프로세스 이미지 사이클들의 트레이닝 데이터세트는 데이터베이스(138)에 저장된다. 라벨링된 트레이닝 예들은 성공적인(양호한) 프로세스 사이클 이미지들 및 성공하지 못한(불량한) 프로세스 사이클 이미지들을 포함할 수 있다. 성공적이지 못한 프로세스 사이클 이미지들은 상기에 열거된 6개의 실패 카테고리들 중 하나의 실패 카테고리에 속하는 것으로 라벨링된다. 하나의 구현예에서, 트레이닝 데이터베이스(138)는 적어도 20,000개의 트레이닝 예들을 포함한다. 트레이닝 데이터베이스의 크기는 증가할 수 있는데, 이는 더 많은 라벨링된 이미지 데이터가 유전형분석 기구들을 사용하여 실험실들로부터 수집되기 때문이다.

[0026] 개시된 기술은 프로세스 사이클 이미지들로부터 특징부들을 추출하기 위한 2개의 독립적인 이미지 프로세싱 기법들을 포함한다. 특징부 생성기(185)는 기계 학습 모델들에 대한 입력을 위해 프로세스 사이클 이미지들로부터 특징부들을 추출하기 위한 2개의 기법들 중 하나의 기법을 적용하는 데 사용될 수 있다. 제1 이미지 프로세싱 기법은 아이겐 얼굴 분석에 의한 얼굴 인식으로부터 진화된다. 40 내지 100개 이상의 이미지 성분들과 같은 비교적 적은 수의 선형 베이스스가 수만 개의 라벨링된 이미지들로부터 식별된다. 아이겐 베이스스를 형성하기 위한 하나의 접근법이 주성분 분석(PCA)이다. 생성 사이클 이미지들은 분류자들에 대한 입력을 위한 베이스스 이미지들의 가중 선형 조합으로서 표현된다. 예를 들어, 하나의 구현예에서, 라벨링된 이미지들의 성분들에 대한 96개의 가중치들이 분류자들을 트레이닝시키는 데 사용된다. 아이겐 이미지들의 베이스스는 데이터베이스(168)에 저장될 수 있다.

[0027] 특징부들을 추출하기 위한 제2 이미지 프로세싱 기법은 섹션 이미지들의 임계화를 수반한다. 이미지 생성 칩의 섹션의 생성 이미지가 수 개의 물리적으로 분리된 영역들을 캡처한다. 섹션의 경계를 이루고 섹션의 물리적 영역들을 분리하는 구조들이 생성 이미지에서 가시적이다. 임계화 기법은 활성 영역의 얼마나 많은 부분이 원하는 신호 강도를 생성하고 있는지를 결정한다. 임계화 기법으로부터의 출력은 양호한 이미지들을 불량한 이미지들과 구별하기 위해 분류자에 대한 입력으로서 주어질 수 있다. 이미지 생성 칩의 영역들 및 섹션들 사이에서의 실패들의 패턴은 근본 원인 분석을 위해 추가로 평가될 수 있다.

[0028] 특징부 생성기(185)에 의해 생성된 생성 이미지들의 이미지 특징부들은 트레이닝된 분류자들(151, 171)에 대한 입력으로서 주어진다. 2개의 유형들의 분류자들이 트레이닝된다. 양호 대 불량 분류자는 성공적인 생성 이미지들 및 성공적이지 못한 생성 이미지들을 예측할 수 있다. 근본 원인 분석 분류자는 성공적이지 못한 이미지들의 실패 카테고리들을 예측할 수 있다. 개시된 기술에 의해 사용되는 분류자들의 하나의 예는 랜덤 포레스트 분류자들을 포함한다. 적용될 수 있는 분류자의 다른 예는 K-최근접 이웃(KNN)들, 다항 로지스틱 회귀, 및 서포트 벡터 머신들을 포함한다. 라벨링된 이미지들의 더 큰 본체들이 이용가능해짐에 따라, 콘볼루션 신경 네트워크(CNN)들이 또한 사용될 수 있다.

[0029] 도 1의 설명을 완료하면, 전송된 시스템(100)의 컴포넌트들은 네트워크(들)(155)와의 통신에 모두 커플링된다. 실제 통신 경로는 공개 및/또는 개인 네트워크를 통한 포인트-대-포인트(point-to-point)일 수 있다. 통신은 다양한 네트워크들, 예컨대, 개인 네트워크들, VPN, MPLS 회로, 또는 인터넷을 통해 발생할 수 있고, 적절한 응용 프로그래밍 인터페이스(application programming interface, API)들 및 데이터 상호교환 포맷들, 예컨대, REST(Representational State Transfer), JSON(JavaScript Object Notation), XML(Extensible Markup Language), SOAP(Simple Object Access Protocol), JMS(Java Message Service), 및/또는 자바 플랫폼 모듈 시스템을 사용할 수 있다. 모든 통신은 암호화될 수 있다. 통신은 대체적으로, 네트워크, 예컨대, LAN(local area network), WAN(wide area network), 전화 네트워크(PSTN(Public Switched Telephone Network), SIP(Session Initiation Protocol), 무선 네트워크, 포인트-대-포인트 네트워크, 방사형 네트워크(star network), 토큰 링 네트워크(token ring network), 허브 네트워크, EDGE, 3G, 4G LTE, Wi-Fi 및 WiMAX와 같은 프로토콜들을 통한 모바일 인터넷을 포함한 인터넷을 통해 이루어진다. 도 1의 엔진들 또는 시스템 컴포넌트들은 다양한 유형들의 컴퓨팅 디바이스들 상에서 실행되는 소프트웨어에 의해 구현된다. 예시적인 디바이스들은 워크스테이션, 서버, 컴퓨팅 클러스터, 블레이드 서버, 및 서버 팜(server farm)이다. 추가적으로, 다양한 인가 및 인증 기법들, 예컨대, 사용자명/패스워드, 공개 인증(Open Authorization, OAuth), 커버로스(Kerberos), 보안된 디지털 인증서들 등이 통신을 보안하는 데 사용될 수 있다.

[0030] **특징부 생성기 - 시스템 컴포넌트들**

[0031] 도 2는 특징부 생성기(185)의 컴포넌트들의 하이-레벨 블록도(200)이다. 이러한 컴포넌트들은 도 11의 설명에서 하기에 제시된 바와 같은 다양한 상이한 컴퓨터 시스템들을 사용하여 컴퓨터 구현된다. 도시된 컴포넌트들은 구현될 때, 병합되거나 추가로 분리될 수 있다. 특징부 생성기(185)는 2개의 이미지 프로세싱 기법들을 구현하는 2개의 하이-레벨 컴포넌트들로 이루어진다: 주성분 분석 또는 PCA 기반 특징부 생성기(235) 및 이미지 세그멘테이션 기반 특징부 생성기(255). PCA 기반 특징부 생성기는 이미지 스케일러(237) 및 아이겐 이미지 베이스스 생성기(239)를 포함한다. 이미지 세그멘테이션 기반 특징부 생성기(255)는 이미지 변환기(257) 및 세기 추출기(259)를 포함한다. 하기의 섹션들에서, 이러한 컴포넌트들의 구현예의 추가 세부사항들을 제시한다.

[0032] **PCA 기반 특징부 생성기**

[0033] 제1 이미지 프로세싱 기법은 아이겐 얼굴 분석에 의한 얼굴 인식으로부터 진화된다. 아이겐 베이스스를 형성하는 것에 대한 하나의 접근법이 주성분 분석(PCA)이다. PCA 기반 특징부 생성기(235)는 크기조절된 프로세스 이

미지들에 PCA를 적용한다. 이미지 스케일러 컴포넌트(237)는 프로세스 사이클 이미지들을 크기조절한다(또는 리스케일링함). 스케일링은 프로세스 이미지들의 크기를 감소시켜, 이들이 아이겐 이미지 베이스 생성기 컴포넌트(239)에 의해 계산 효율적인 방식으로 프로세싱될 수 있도록 한다. 본 발명자들은 하기의 섹션들에서 이러한 컴포넌트들의 세부사항들을 제시한다.

[0034] 이미지 스케일러

[0035] 유전형분석 기구들 또는 스캐너들로부터 얻어진 더 높은 해상도 이미지들은 프로세싱하는 데 더 많은 계산 리소스들을 요구할 수 있다. 유전형분석 스캐너들로부터 얻어진 이미지들은 이미지 스케일러(237)에 의해 크기조절되어, 이미지 생성 칩들의 섹션들의 이미지들이 180 x 80 픽셀들의 감소된 해상도로 분석되도록 한다. 하나의 예에서, 스캐너로부터 얻어진 섹션들의 이미지들은 3600 x 1600 픽셀들의 해상도이고, 이미지들을 크기조절하기 위해 해상도의 20배 감소가 적용된다. 이는, 성공적인 생성 이미지들을 성공적이지 못한 생성 이미지들과 구별하고 이어서 6개의 실패 카테고리들 중에서 실패의 근본 원인들을 분류하기에 충분한 해상도이다. 오리지널 해상도의 4 내지 25배로 리스케일링된 이미지들이 동일한 방식으로 프로세싱될 수 있다.

[0036] 개시된 기술은 생성 이미지들의 크기를 감소시키기 위해 다양한 보간 기법들을 적용할 수 있다. 하나의 구현예에서, 쌍선형 보간(bilinear interpolation)이 섹션 이미지들의 크기를 감소시키는 데 사용된다. 선형 보간은 선형 다항식들을 사용하여 미지의 데이터 포인트들의 이산 세트의 범위를 갖는 새로운 데이터 포인트들을 구성하는 곡선 피팅 방법이다. 쌍선형 보간은 2-차원 그리드 상에서 2개의 변수들(예컨대, x 및 y)의 보간 함수들에 대한 선형 보간의 확장이다. 쌍선형 보간은, 먼저 하나의 방향으로 그리고 이어서 다시 제2 방향으로 선형 보간을 사용하여 수행된다. 각각의 단계가 샘플링된 값들에서 그리고 그 위치에서 선형적이지만, 전체로서의 보간은 선형이 아니라, 오히려 샘플 위치에서 2차(quadratic)이다. 최근접 이웃 보간 및 픽셀 영역 관계를 사용한 리샘플링과 같은, 섹션 이미지들의 크기를 감소시키기 위한(리스케일링하기 위한) 다른 보간 기법들이 또한 사용될 수 있다.

[0037] 아이겐 이미지 베이스 생성기

[0038] 분류자들에 대한 입력 특징부들을 생성하기 위해 섹션 이미지들에 적용되는 제1 이미지 프로세싱 기법은 아이겐 얼굴 분석에 의한 얼굴 인식으로부터 진화된다. 수만 개의 라벨링된 이미지들로부터, 40 내지 100개 이상의 이미지 성분들의 선형 베이스가 식별된다. 아이겐 이미지들의 베이스를 형성하는 것에 대한 하나의 접근법이 주성분 분석(PCA)이다. 벡터 공간 V 내의 요소들(벡터들)의 세트 B는, V의 모든 요소가 B의 요소들의 선형 조합으로서 고유한 방식으로 기록될 수 있는 경우, 베이스로 불린다. 동등하게, B는 그의 요소들이 선형적으로 독립적인 경우의 베이스이고, V의 모든 요소는 B의 요소들의 선형 조합이다. 벡터 공간은 수 개의 베이스들을 가질 수 있다. 그러나, 모든 베이스들은 벡터 공간의 디멘션(dimension)으로 불리는 동일한 수의 요소들을 갖는다. 본 발명의 기술에서, 벡터 공간의 베이스는 아이겐 이미지들이다.

[0039] PCA는 종종, d-차원 데이터세트의 디멘션들을 k-차원 서브공간에 투영함으로써 그를 감소시키는 데 사용되며, 여기서 $k < d$ 이다. 예를 들어, 본 발명의 트레이닝 데이터베이스 내의 크기조절된 라벨링된 이미지는 디멘션 $d=14,400$ -차원 공간(180 x 80 픽셀들)의 벡터를 설명한다. 다시 말해, 이미지는 14,400-차원 공간 내의 포인트이다. 아이겐 공간 기반 접근법들은 더 낮은 디멘션 특징부 벡터들로 이미지 벡터들을 근사화한다. 이러한 기법 뒤의 주요 추정은, 이미지 공간(특징부 벡터들에 의해 주어짐)이 이미지 공간(이미지 내의 픽셀들의 개수에 의해 주어짐)보다 더 낮은 디멘션을 갖고 이미지들의 인식이 이러한 감소된 공간에서 수행될 수 있다는 것이다. 전체 구성에서 유사한, 이미지 생성 칩들의 섹션들의 이미지들은 이러한 거대한 공간에 랜덤으로 분포되지 않을 것이고, 따라서, 비교적 낮은 차원 서브공간에 의해 설명될 수 있다. PCA 기법은 전체 이미지 공간 내의 섹션 이미지들의 분포를 가장 잘 설명하는 벡터들을 발견한다. 이러한 벡터들은 "이미지 공간"으로도 지칭되는 이미지들의 서브공간을 정의한다. 본 발명의 구현예에서, 각각의 벡터는 180 x 80 픽셀 이미지를 설명하고, 트레이닝 데이터에서의 이미지들의 선형 조합이다. 하기의 텍스트에서, 주성분 분석(PCA)이 아이겐 이미지들의 베이스를 생성하는 데 어떻게 사용될 수 있는지의 세부사항들을 제시한다.

[0040] 라벨링된 트레이닝 이미지들의 PCA 기반 분석은 하기의 5개의 단계들을 포함할 수 있다.

[0041] 단계 1: 다차원 상관된 데이터에의 액세스

[0042] PCA의 적용에서의 제1 단계는 고차원 데이터에 액세스하는 것이다. 하나의 예에서, 트레이닝 데이터로서 20,000개의 라벨링된 이미지들을 사용하였다. 각각의 이미지를 180 x 80 픽셀 해상도로 크기조절하였고, 픽셀 당 하나씩 14,400-차원 공간 내의 포인트로서 표현하였다. 이러한 기법은 상기에서 특정된 것보다 더 높은 해

상도 또는 더 낮은 해상도의 이미지들을 핸들링할 수 있다. 트레이닝 데이터세트의 크기는 실험실들로부터 더 많은 라벨링된 이미지들을 수집함에 따라 증가할 것으로 예상된다.

[0043] 단계 2: 데이터의 표준화

[0044] 표준화(또는 Z-스코어 정규화)는 특징부들을 리스케일링하여 그들이 0과 동일한 평균 또는 $\mu = 0$ 을 갖는 가우스 분포 및 1과 동일한 평균 또는 $\sigma = 1$ 로부터의 표준 편차의 속성들을 갖도록 하는 프로세스이다. 서로 유사한 범위들을 갖는 특징부들을 구축하기 위해 표준화가 수행된다. 이미지의 표준 스코어는, 이미지로부터 평균(이미지)을 감산하고 그 결과를 표준 편차로 제곱함으로써 계산될 수 있다. PCA가 축들을 따르는 분산을 최대화하는 특징부 서브공간을 산출하기 때문에, 그것은 데이터를 표준화하여 그것이 축들을 가로질러 중심에 있도록 하는 것을 돕는다.

[0045] 단계 3: 공분산 행렬의 계산

[0046] 공분산 행렬은 각각의 요소가 2개의 특징부들 사이의 공분산을 표현하는 d-차원 공간의 $d \times d$ 행렬이다. 2개의 특징부들의 공분산은 그들의 함께 변화하는 경향을 측정한다. 변화량은 특징부의 평균으로부터의 그의 평균의 제곱 편차의 평균이다. 공분산은 특징부 값들의 평균으로부터의 그들의 편차들의 곱들의 평균이다. 특징부 k 및 특징부 j를 고려한다. $\{x(1, j), x(2, j), \dots, x(i, j)\}$ 를 특징부 j의 i개의 예들의 세트라고 하고, $\{x(1, k), x(2, k), \dots, x(i, k)\}$ 를 특징부 k의 i개의 예들의 세트라고 한다. 유사하게, \bar{x}_j 를 특징부 j의 평균으로 하고, \bar{x}_k 를 특징부 k의 평균으로 한다. 특징부 j와 특징부 k의 공분산은 하기와 같이 계산된다:

$$\sigma_{jk} = \frac{1}{n-1} \sum_{i=1}^n (x(i, j) - \bar{x}_j)(x(i, k) - \bar{x}_k) \quad (1)$$

[0047] 하기의 행렬 수학적식을 통해 공분산 행렬의 계산을 표현할 수 있다:
 [0048]

$$\Sigma = \frac{1}{n-1} ((\mathbf{X} - \bar{\mathbf{x}})^T (\mathbf{X} - \bar{\mathbf{x}})) \quad (2)$$

[0049] 여기서, 평균 벡터는 하기와 같이 표현될 수 있다:
 [0050]

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i.$$

[0051] 평균 벡터는 이러한 벡터의 각각의 값이 트레이닝 데이터세트에서 특징부 열(column)의 샘플 평균을 표현하는 d-차원 벡터이다. 공분산 값 σ_{jk} 는 $-(\sigma_{ij})(\sigma_{ik})$, 즉, 역 선형 상관관계 내지 $+(\sigma_{ij})(\sigma_{ik})$ 선형 상관관계에서 변화할 수 있다. 2개의 특징부들 사이에 어떠한 의존성도 없을 때, σ_{jk} 의 값은 0이다.
 [0052]

[0053] 단계 4: 아이겐 벡터들 및 아이겐 값들의 계산

[0054] 공분산 행렬의 아이겐 벡터들 및 아이겐 값들은 PCA의 코어를 표현한다. 아이겐 벡터들(또는 주성분들)은 새로운 특징부 공간의 방향들을 결정하고, 아이겐 값들은 그들의 크기들을 결정한다. 다시 말해, 아이겐 값들은 새로운 특징부 공간의 축들을 따르는 데이터의 분산을 설명한다. 아이겐 분해는 행렬의 아이겐 벡터들 및 아이겐 값들을 사용하여 그 행렬을 표현하는 것에 의한 행렬 인수분해 방법이다. 아이겐 벡터는 선형 변환이 그것에 적용될 때 오로지 스칼라만큼만 변화하는 벡터로서 정의된다. A가 선형 변환을 표현하는 행렬이고, v가 아이겐 벡터이고 λ 가 대응하는 아이겐 값인 경우, 그것은 $Av = \lambda v$ 로서 표현될 수 있다. 정사각형 행렬은 그것이 디멘션들을 갖는 만큼 많은 아이겐 벡터들을 가질 수 있다. 모든 아이겐 벡터들을 행렬 V의 열들로서 그리고 대응하는 아이겐 값들을 대각선 행렬 L의 엔트리들로서 표현하는 경우, 상기의 수학적식은 $AV = VL$ 로서 표현될 수 있다. 공분산 행렬의 경우, 모든 아이겐 벡터들은 서로 직교하고 새로운 특징부 공간의 주성분들이다.
 [0055]

[0055] 단계 5: 설명된 분산을 사용한 아이겐 이미지들에 대한 베이스스의 선택

[0056] 상기의 단계는 특징부 공간의 디멘션과 동일한 본 발명의 구현예의 경우에 14,400개의 주성분들을 생성할 수 있다. 아이겐 쌍은 아이겐 벡터 및 스칼라 아이겐 값으로 이루어진다. 아이겐 값들에 기초하여 아이겐 쌍들을

분류할 수 있고, "설명된 분산"으로 지칭되는 메트릭을 사용하여 아이겐 이미지들의 베이스스를 생성할 수 있다. 설명된 분산은 얼마나 많은 정보(또는 분산)가 각각의 주성분에 기인할 수 있는지를 나타낸다. 설명된 측정 값들의 결과들을 2차원 그래프 상에 플로팅할 수 있다. 분류된 주성분들은 x-축을 따라 표현된다. 누적 설명된 분산을 나타내는 그래프가 플로팅될 수 있다. 분산의 주요 부분을 표현하는 처음 m개의 성분들이 선택될 수 있다.

[0057] 본 발명의 구현예에서, 처음 40개의 성분들은 설명된 분산의 높은 백분율을 표현하였고, 따라서, 처음 40개의 주성분들을 선택하여 본 발명의 새로운 특징부 공간의 베이스스들을 형성하였다. 다른 구현예들에서, 아이겐 이미지들의 베이스스들을 생성하기 위해, 25 내지 100개의 주성분들 또는 최대 256 또는 512개의 주성분들에 이르는 100개 초과 주성분들이 선택될 수 있다. 아이겐 이미지 분석에 의해 분석될 각각의 생성 이미지는 베이스스 이미지들의 가중 선형 조합으로서 표현된다. 정렬된 세트의 베이스스 성분들의 각각의 가중치가 분류자를 트레이닝시키기 위한 특징부로서 사용된다. 예를 들어, 하나의 구현예에서, 라벨링된 이미지들의 성분들에 대한 96개의 가중치들을 분류자를 트레이닝시키는 데 사용하였다.

[0058] 개시된 기술은 다른 이미지 분해 및 차원수 감소 기법들을 사용할 수 있다. 예를 들어, 이미지들의 완전한 표현들을 학습하는 PCA에 비교되는 바와 같은, 이미지들의 부분 기반 표현을 학습하는 비-음수 행렬 인수분해(NMF). PCA와 달리, NMF는 이미지들의 부분들을 닳은 베이스스 이미지들의 세트로 이미지들을 표현하도록 학습한다. NMF는 행렬 X를 2개의 행렬들 W 및 H로 인수분해하며, 이때 3개의 행렬들 모두가 비-음수 요소들을 갖는 속성이 있다. 행렬 X가, 각각 p개(예컨대, 14,400개)의 디멘션들을 갖는 n개의 데이터 포인트들(예컨대, 이미지 생성 칩들 상의 섹션들의 이미지들)이 있도록 셋업된다고 가정한다. 따라서, 행렬 X는 p개의 행들 및 n개의 열들을 갖는다. 본 발명자들은 p개의 디멘션들을 r개의 디멘션들로 감소시키거나 또는 다시 말해서, 랭크 r 근사치를 생성하기를 원한다. NMF는 행렬 X를 2개의 행렬들의 곱으로서 근사화한다: W(p개의 행들 및 r개의 열들) 및 H(r개의 행들 및 n개의 열들).

[0059] 행렬 W의 해석은 각각의 열이 베이스스 요소라는 것이다. 베이스스 요소라는 것은, n개의 오리지널 데이터 포인트들(또는 이미지들)에 존재하는 일부 성분을 의미한다. 이들은 구축 블록들이며, 이들로부터 오리지널 데이터 포인트들 또는 이미지들 모두에 대한 근사치들을 재구성할 수 있다. 행렬 H의 해석은 각각의 열이 베이스스 행렬 W 내의 데이터 포인트의 좌표들을 제공한다는 것이다. 다시 말해, 그것은 행렬 W 내의 구축 블록들의 선형 조합으로부터 오리지널 데이터 포인트에 대한 근사치를 어떻게 재구성할지를 알려준다. 얼굴 이미지들의 경우, 행렬 W 내의 베이스스 요소들(또는 베이스스 이미지들)은 눈, 코, 입술 등과 같은 특징부들을 포함할 수 있다. 행렬 H의 열들은 어느 특징부들이 어느 이미지에 존재하는지를 나타낸다.

[0060] **이미지 세그먼테이션 기반 특징부 생성기**

[0061] 프로세스 사이클 이미지들로부터 특징부들을 추출하기 위한 제2 이미지 프로세싱 기법은 이미지 영역들의 임계화에 기초한다. 이미지 세그먼테이션 기반 특징부 생성기(255)는, 이미지 세그먼트화기(segmentor)(257)를 사용하여 이미지 생성 칩의 섹션들의 이미지들을 먼저 세그먼트화하고 이어서 섹션 이미지의 관심 활성 영역들 또는 부위들의 세기를 추출함으로써 임계화를 적용한다. 임계화는 활성 영역의 얼마나 많은 부분이 원하는 신호 강도를 생성하고 있는지를 결정한다.

[0062] 이미지 생성 칩은 행들 및 열들로 구성된 다수의 섹션들, 예컨대 24, 48, 96개 이상의 섹션들을 포함할 수 있다. 이러한 설계는 많은 샘플들(섹션당 하나씩)이 병렬로 프로세싱될 수 있으므로 하나의 프로세스 사이클에서 다수의 샘플들의 프로세싱을 가능하게 한다. 일정 섹션이 다른 섹션들과 물리적으로 분리되어, 샘플들이 서로 혼합되지 않도록 한다. 추가적으로, 일정 섹션은 "슬롯들"로 지칭되는 다수의 병렬 영역들로 구성될 수 있다. 따라서, 섹션들과 슬롯들의 경계들에서의 구조들은 유전형분석 스캐너들로부터의 프로세스 사이클 이미지들에서 가시적이다. 이미지 특징부들의 추출을 위해 섹션 이미지들을 변환하기 위한 기법들을 구현할 수 있는 이미지 세그먼테이션 기반 특징부 생성기(255)의 2개의 컴포넌트들의 세부사항들을 하기에 제시한다.

[0063] **이미지 변환기**

[0064] 이미지 변환기(257)는 관심 부위들로부터의 세기들을 추출하기 위해 섹션 이미지들을 준비하기 위한 일련의 이미지 변환 기법들을 적용한다. 하나의 구현예에서, 이미지 변환 및 세기 추출의 이러한 프로세스는 하기의 5개의 단계들의 일부 또는 전부에 의해 수행된다. 이미지 변환은 섹션의 그레이스케일 이미지를, 흑색 픽셀들 및 밝은 픽셀들로 이루어진 이진 이미지로 변환한다. 그레이스케일 이미지 및 이진 이미지의 활성 영역들의 평균 세기 값들은 이미지를 정상적인(양호한) 이미지 또는 비정상적인(불량한) 이미지로 분류하기 위한 분류자에 대

한 입력 특징부들로서 주어진다. 하기의 텍스트에서, 그레이스케일 이미지를 이진 이미지로 변환하기 위해 임계화를 적용하는 것을 포함하는 이미지 변환 단계들의 세부사항들을 제시한다. 프로세스 단계들은 잡음을 제거하기 위해 필터들을 적용하는 단계를 포함한다.

[0065] 이미지 변환 프로세스에서의 제1 단계는 섹션들의 프로세스 사이클 이미지들에 쌍방향 필터들을 적용하는 것이다. 쌍방향 필터는 이미지들을, 에지들을 보존하면서 평활화하는 기법이다. 이는 각각의 픽셀의 세기를 그의 이웃 픽셀들로부터의 세기 값들의 가중 평균으로 대체한다. 각각의 이웃은 원거리 픽셀들에 페널티를 부과하는 공간 성분 및 상이한 세기를 갖는 픽셀들에 페널티를 부과하는 범위 성분에 의해 가중된다. 성분들 둘 모두의 조합은 인근의 유사한 픽셀들만이 최종 결과에 기여하는 것을 보장한다. 따라서, 쌍방향 필터는 이미지를, 그의 불연속부들 또는 에지들을 보존하면서 평활화하는 효율적인 방식이다. 중간값(median) 필터 및 이방성 확산과 같은 다른 필터들이 사용될 수 있다.

[0066] 이미지 변환의 제2 단계는 단계 1로부터의 출력 이미지들에 임계화를 적용하는 것일 수 있다. 하나의 구현예에서, 세기들의 히스토그램을 사용하고 임계치를 검색하여 어둡 세기 클래스 및 밝은 세기 클래스에 배정된 픽셀들 사이의 그레이스케일 분산의 가중 합을 최대화하는 Otsu의 방법(Otsu, N., 1979, "A threshold selection method from gray-level histograms", IEEE Transactions on Systems, Man, and Cybernetics, Volume 9, Issue 1)을 적용한다. Otsu의 방법은 클래스간 분산을 최대화하려고 시도한다. 기본 아이디어는, 잘 임계화된 클래스들이 그들의 픽셀들의 세기 값들과 관련하여 구별되어야 하고, 역으로, 그들의 세기 값들과 관련하여 클래스들 사이의 최상의 분리를 제공하는 임계치가 최상의 임계치일 것이라는 것이다. 덧붙여, Otsu의 방법은, 그것이, 용이하게 얻을 수 있는 1-차원 어레이인 이미지의 히스토그램 상에서 수행되는 계산들에 전적으로 기초한다는 속성을 갖는다. 추가 세부사항들의 경우, 문헌[Section 10.3.3 of Gonzalez and Woods, "Digital Image Processing", 3rd Edition]을 참조한다.

[0067] 이미지 변환의 제3 단계는 스펙클-유사 잡음(speckle-like noise)을 제거하기 위한 잡음 감소 가우스 블러 필터의 적용이다. 잡음은 작은 스펙클들로 프로세스 사이클 이미지들을 오염시킬 수 있다. 가우스 필터링은 중심 위치까지의 공간 거리에 따라 감소하는 가중치를 갖는 인접 위치들의 세기의 가중 평균이다.

[0068] 이미지 변환의 제4 단계는 이미지 모폴로지(morphology) 동작들을 포함한다. 제3 단계로부터의 이진 출력 이미지들은 이미지들 내의 구멍들을 충전하도록 모폴로지 변환에 의해 프로세싱된다. 구멍은 전경 픽셀들(1들에 의해 표현됨)의 연결된 경계에 의해 둘러싸인 배경 부위(0들에 의해 표현됨)로서 정의될 수 있다. 2개의 기본 이미지 모폴로지 동작들은 "침식(erosion)" 및 "확장(dilation)"이다. 침식 동작에서, 이진 이미지 위에서 커널이 슬라이딩한다(또는 이동함). 이진 이미지 내의 픽셀(1 또는 0 중 어느 하나)은 커널 아래의 모든 픽셀들이 1들인 경우에 1로 간주된다. 그렇지 않은 경우, 그것은 침식된다(0으로 변경됨). 침식 동작은 이진 이미지에서 격리된 1들을 제거하는 데 있어서 유용하다. 그러나, 침식은 에지들을 침식시킴으로써 1들의 클러스터들을 또한 수축시킨다. 확장 동작은 침식의 반대이다. 이러한 동작에서, 커널이 이진 이미지 위에서 슬라이딩할 때, 커널 아래의 적어도 하나의 픽셀의 값이 1인 경우, 커널에 의해 중첩된 이진 이미지 영역 내의 모든 픽셀들의 값들은 1로 변경된다. 확장 동작이 이진 이미지에 적용되고 뒤이어 침식 동작이 이어진 경우, 효과는 1들의 클러스터들 내부의 작은 구멍들(이미지에서 0들에 의해 표현됨)의 폐쇄이다. 이러한 단계로부터의 출력은 이러한 이미지 변환 기법의 제5 단계를 수행하는 세기 추출기 컴포넌트(259)에 대한 입력으로서 제공된다.

[0069] 세기 추출기

[0070] 세기 추출기(259)는 섹션들과 슬롯들의 경계들에서 구조들을 필터링함으로써 섹션 이미지들을 활성 영역들 또는 세그먼트들로 분할한다. 세기 추출기는 섹션 이미지들을 8 내지 최대 17개 이상의 활성 영역들로 분할하기 위해 상이한 세그먼테이션들을 적용할 수 있다. 섹션 이미지 내의 영역들의 예들은 4개의 슬롯들, 4개의 코너들, 코너들 사이의 4개의 에지들, 및 섹션과 슬롯들의 경계들에서의 다양한 수직 및 수평 라인들을 포함한다. 이어서, 활성 영역들을 분리하는 공지의 구조들에 대응하는 영역들이 이미지로부터 제거된다. 나머지 활성 영역들에 대한 이미지 부분들은 세기 추출기(259)에 의해 프로세싱된다. 세기 값들이 추출되고, 변환된 이미지 및 대응하는 변환되지 않은 이미지의 각각의 활성 영역에 대해 평균된다. 예를 들어, 세기 값들이 변환된 이미지의 17개의 활성 영역들로부터 추출되는 경우, 세기 추출기는 또한, 변환되지 않은 이미지의 동일한 17개의 활성 영역들로부터 세기 값들을 추출한다. 따라서, 섹션 이미지마다 총 34개의 특징부들이 추출된다.

[0071] 이진 이미지들의 경우, 활성 영역의 평균 세기는 1 내지 0일 수 있다. 예를 들어, 흑색 픽셀의 세기가 0이고 밝은(또는 블랭크) 픽셀의 세기가 1이라고 간주한다. 활성 영역 내의 모든 픽셀들이 흑색인 경우, 활성 영역의 평균 세기는 0일 것이다. 유사하게, 활성 영역 내의 모든 픽셀들이 밝은 경우, 그 영역의 세기는 1일 것이다.

정상적인 이미지들 내의 활성 영역들은 이진 이미지들에서 블랭크 또는 밝은 것으로 나타나는 반면, 흑색 픽셀들은 비정상적인 이미지들을 표현한다. 그레이스케일 이미지 내의 대응하는 활성 영역들의 평균 세기들이 또한 추출된다. 그레이스케일 이미지 및 변환된 이진 이미지 둘 모두로부터의 활성 영역들의 평균 세기들은 양호 대 불량 분류자에 대한 입력으로서 주어진다. 하나의 구현예에서, 분류자로부터의 분류 신뢰도 스코어는 이미지를 정상적인(양호한) 이미지 또는 비정상적인(불량한) 이미지로 분류하기 위해 임계치와 비교된다. 임계 값의 일례가 80%이다. 임계치의 더 높은 값은 더 많은 이미지들이 비정상적인 것으로 분류되는 결과를 초래할 수 있다.

[0072] **프로세스 사이클 이미지들**

[0073] 이제, 이미지 생성 칩들 상의 섹션들의 성공적인 생성 이미지들 및 성공적이지 못한 생성 이미지들의 예들을 제시한다. 도 4는 이미지 생성 칩 상의 24개의 섹션들의 생성 이미지들의 예시(400)이다. 섹션들은 12개의 행들 및 2개의 열들로 배열된다. 각각의 섹션은 4개의 슬롯들을 갖는다. 예시(400)는 성공적인 생성 사이클의 섹션 이미지들을 보여준다. 48, 96개 이상의 섹션들을 포함하는 것과 같은 다른 구성의 섹션들을 갖는 이미지 생성 칩들이 또한 사용될 수 있다. 하기의 도면들에서, 성공적이지 못한 생성 사이클들의 섹션 이미지들의 예들을 제시한다. 생성 프로세스는 동작 프로세싱 오류 및 화학적 프로세싱 오류 둘 모두에 취약하다. 동작 결함들은 기계적 또는 샘플 핸들링 문제들로 인해 야기될 수 있다. 화학적 프로세싱 오류들은 샘플들 또는 샘플들의 화학적 프로세싱에서의 문제들에 의해 야기될 수 있다. 개시된 기술은 동작 프로세싱 오류 및 화학적 프로세싱 오류 둘 모두로 인해 발생하는 불량한 프로세스 이미지 사이클들을 분류하려고 시도한다.

[0074] 도 5a는 성공적이지 못한 생성 사이클로부터의 섹션 이미지의 일례(510)를 도시한다. 도 5a의 이미지 생성 칩의 제2 열 및 제7 행에 있는 섹션(512)의 이미지는 하단 절반 부분에서 어렵게 착색되고 상단 부분에서 약간 밝게 착색된다. 이러한 실패의 원인은 혼성화 프로세스와 연계된다. 따라서, 섹션의 실패한 이미지는 "Hyb" 실패로 라벨링된다. 혼성화 실패들은 또한, 이미지 생성 칩들 상에서 샘플 준비 프로세스 동안 샘플들을 핸들링하는 로봇들의 실패들로 인해 발생할 수 있다. 이러한 섹션에 대한 호출 레이트는 도면에 도시된 바와 같이 98 퍼센트 임계치 미만이다. 일부 경우들에 있어서, 유전형분석 기구들로부터의 섹션에 대한 호출 레이트는 통과 임계치를 초과할 수 있고, 그때까지도, 섹션 이미지는 혼성화 오류로 인해 실패할 수 있다.

[0075] 예시(510)에서, 행 11 및 열 2에 있는 섹션(514)의 이미지는 우측 벽 상의 어두운 착색 부위를 갖는 것이 주목될 수 있다. 이는 또한 프로세싱 문제를 나타낼 수 있지만, 이러한 이미지의 전체 호출 레이트는 통과 임계치를 초과하고, 그것은 실패한 이미지로서 라벨링되지 않는다. 섹션 상에 충분한 리던던시의 샘플들이 있는데, 이들로 인해 분명한 실패를 갖는 섹션들의 작은 영역들이 무시될 수 있고 결과들에 오류들을 야기하지 않을 수 있다. 예를 들어, 하나의 예에서, 스캐너는 10의 리던던시를 갖는 섹션 상의 약 700K개의 프로브들로부터의 형광을 판독한다. 따라서, 호출 레이트는 약 7백만 개의 프로브들의 판독물에 기초한다. 도 5b의 예시(515)에서 혼성화 실패들의 추가 예들을 제시한다. 파선 경계들에서의 이미지 생성 칩 상의 4개의 섹션들은 혼성화 실패로 인한 섹션들의 불량한 생성 이미지들을 보여준다. 이러한 4개의 섹션들에 대한 호출 레이트 값들은 통과 임계치를 초과하지만, 이러한 섹션들의 이미지들은 혼성화 오류로 인해 실패한 것으로 라벨링된다는 것에 주목한다.

[0076] 도 5c는 스페이스 시프트 실패로 인한 성공적이지 못한 프로세싱을 보여주는 9개의 섹션 이미지들의 예시(520)를 제시한다. 샘플들이 이미지 생성 칩 상의 섹션들 상에 준비될 때, 어렵게 착색된 마커가 섹션들 주위에 배치된다. 스페이스는 각각의 섹션 내의 샘플들을 이웃 섹션들 내의 다른 샘플들과 분리한다. 마커가 정확하게 배치되지 않은 경우, 그것은 이미지 신호의 일부를 차단할 수 있다. 오프셋 오류는 도 5c에 도시된 바와 같이, 다수의 이웃하는 섹션들에 걸쳐 발생할 수 있다. 이러한 도면에서 9개의 섹션들의 상단 부분들은 어렵게 착색된 것으로 나타난다. 좌측에서 우측으로 이동함에 따라 섹션들의 상단 부분의 어두운 부분이 증가한다. 스페이스 시프트 문제는 동작 오류인데, 이는 그것이 이미지 생성 칩 상의 샘플들의 준비 동안 실험실 기술자들에 의한 마커의 부정확한 배치에 의해 유발되기 때문이다. 도 5d는 스페이스 시프트 실패로 인한 섹션들의 실패한 이미지들의 3개의 더 많은 예들을 제시한다. 박스(525)는 섹션 이미지들의 상단 부분들이 상단 우측으로부터 상단 좌측으로 폭이 증가하면서 어렵게 착색됨에 따라, 스페이스 시프트 실패를 갖는 5개의 섹션 이미지들을 보여준다. 박스(527)는 섹션들의 하단 부분들에서 스페이스 시프트 문제로 인한 실패한 프로세스를 나타내는 2개의 섹션 이미지들을 보여준다. 유사하게, 박스(529)는 스페이스 시프트 문제로 인해 실패한 2개의 섹션들의 이미지들을 보여준다.

[0077] 도 5e는 오프셋 실패에 의해 야기되는 성공적이지 못한 프로세싱으로 인한 섹션들의 실패한 이미지들의 일례를

도시한다. 오프셋 실패에서, 이미지 생성 칩 상의 섹션들의 이미지들은 일 측면으로 시프트된다. 예를 들어, 예시(530)에서, 이미지 생성 칩 상의 모든 섹션 이미지들이 좌측을 향해 시프트되고, 따라서, 좌측 에지 상의 이미지 생성 칩의 어둡게 착색된 외부 경계는 이미지로부터 컷오프된다. 오프셋 실패들은 칩 캐리어 상의 이미지 생성 칩의 스캐너 오정렬 또는 오배치(misplacement)와 같은 스캐닝 오류들에 의해 야기될 수 있다.

[0078] 도 5f는 표면 마모 실패로 인한 실패한 섹션 이미지들의 예들을 도시한다. 표면 마모는 제조 프로세스 동안 또는 섹션들 상의 샘플들의 준비 동안 이미지 생성 칩에서의 섹션들의 표면 상의 스크래치들에 의해 야기된다. 스크래치들은 예시(535)에서 보여지는 바와 같이, 섹션들의 이미지들 상의 라인들로서 가시적이다. 호출 레이 트 값들이 좌측의 파선 박스 내의 3개의 섹션들에 대한 통과 임계치를 초과함에도 불구하고, 이미지들은 표면 마모 실패로 인해 실패한 것으로 라벨링된다는 것에 주목한다.

[0079] 도 5g는 시약 흐름 실패로 인한 실패한 섹션 이미지들의 예시(540)를 도시한다. 박스(542) 내의 10개의 섹션 이미지들은 시약 흐름 실패로 인한 실패한 이미지들로서 라벨링된다. 섹션 이미지들은 부적절한 시약 흐름에 의해 야기되는 성공적이지 못한 프로세스로 인해 실패하였다. 유전형분석 프로세스 동안, 시약은 일 측면으로부터 이미지 생성 칩에 도입된다. 시약은 이미지 생성 칩의 일 단부로부터 반대편 단부를 향해 흐르고, 모든 섹션들을 완전히 커버한다. 때때로, 시약의 흐름에 문제가 있고, 그것은 모든 섹션들에 균일하게 전파되지 않는다. 이러한 경우, 시약은 충분한 양의 시약이 섹션을 커버하지 않을 때 건조해질 수 있다. 부적절한 시약 흐름은 일부 섹션들로부터의 방출된 신호의 강도를 감소시킬 수 있는데, 이는 형광 염료가 모든 섹션들에 걸쳐 균일하게 분포되지 않아서, 이에 따라, 이미지 품질에 영향을 미칠 수 있기 때문이다. 시약 흐름 실패로 인한 실패한 이미지들은 성공적인 프로세스 사이클을 표현하는 섹션 이미지들에 비해 색상이 더 어둡게 나타날 수 있다. 도 5h는 예시(545)에서 시약 흐름 실패로 인한 실패한 섹션 이미지들의 추가 예들을 도시한다. 시약 흐름 실패는 도 5g 및 도 5h에 도시된 바와 같이 이미지 생성 칩의 일정 부위 내의 다수의 이웃 섹션들에 영향을 미칠 수 있다.

[0080] 도 5i는 미지의 이유들로 인한 실패한 이미지들의 예들을 도시한다. 실패한 섹션 이미지들은 "비정상적인" 것으로 라벨링된다. 실패들의 비정상 클래스의 실패한 이미지들은 혼합된 또는 식별되지 않은 원인들 및 약한 신호들로 인한 것일 수 있다. 섹션들의 이미지들의 예시(550)는 또한, 이미지 생성 칩의 상단 좌측 상의 섹션에 대한 스페이서 실패의 일례를 보여준다. 좌측 상단 위치(행 1 및 열 2) 상의 이미지 섹션은 스페이서 실패로 라벨링된다. 실패한 섹션 이미지의 상단 부분은 어둡게 착색된다는 것을 알 수 있다. 상단의 어둡게 착색된 부위의 부분은 섹션 이미지의 우측 코너로부터 좌측 코너로 증가한다.

[0081] **주성분 분석 기반 특징부 생성**

[0082] 이제, 얼굴 인식 분야에서 아이젠 얼굴들로 지칭되는 아이젠 이미지들의 예들을 제시한다. 수만 개의 라벨링된 이미지들로부터, 40 내지 100개 이상의 이미지 성분들의 선형 베이스가 식별된다. 도 6a 및 도 6b는 주성분 분석(PCA)을 적용함으로써 얻어진 96개의 아이젠 이미지들(610, 611)의 일례를 도시한다. 96개의 아이젠 이미지들은 상기에 제시된 바와 같은 설명된 가변성의 측정치에 따른 성분들의 랭크 정렬에 기초하여 선택된다. 도 6c 및 도 6d는 도 6a 및 도 6b의 96개의 아이젠 이미지들로부터의 상단 40개의 랭크된 아이젠 이미지들(620, 621)을 도시한다. 하나의 구현예에서, 40개의 성분들이 가변성의 대부분을 설명했다는 것을 관찰하였다. 선택된 추가적인 성분들은 샘플 프로세싱에서 자연적 가변성 또는 잡음의 패턴들을 반영하는 것으로 나타났다.

[0083] 이제, PCA를 사용한 아이젠 이미지들의 베이스의 생성 및 차원수 감소를 기술한다. 제1 단계는, 섹션들의 이미지들의 해상도를 감소시키고 PCA에 입력하기 위한 감소된 이미지들을 준비하는 것이다. 도 7a(710으로 라벨링됨)는 20배 리스케일링되어 크기 180 x 80 픽셀들의 감소된 섹션 이미지들을 생성하는 3600 x 1600 픽셀 해상도의 섹션 이미지들을 도시한다. 리스케일링된 섹션 이미지들은 평탄화된다. 생성된 평탄화된 리스케일링된 섹션 이미지들은 1-차원 어레이, 즉 각각 14,400 x 1 픽셀들이다. 평탄화된 리스케일링된 섹션 이미지들은 상기에 설명된 바와 같이 표준화되어, 이에 따라, 도 7b(740으로 라벨링됨)에 도시된 바와 같은 표준화된 평탄화된 리스케일링된 섹션 이미지들을 생성하는데, 이들은 PCA에 대한 입력으로서 주어진다. 따라서, PCA는 14,400개의 주성분들 또는 아이젠 이미지들을 생성한다. 각각의 입력 이미지는 14,400-차원 공간 내의 벡터이다. 이어서, 설명된 분산을 사용하여 주성분들 또는 아이젠 이미지들을 랭크 정렬하고, 베이스, 예를 들어, 40 내지 100개의 성분들의 베이스를 생성한다. 성분들은 선형 공간의 베이스를 형성한다.

[0084] **이미지 세그먼테이션 기반 특징부 생성**

[0085] 섹션의 이미지들로부터 특징부들을 생성하기 위한 제2 이미지 프로세싱 기법은 이미지 영역들 또는 세그먼트들

의 임계화를 수반한다. 도 8a는 섹션 이미지에 적용된 세그멘테이션 기반 특징부 생성의 일례(810)를 도시한다. 예시(812)는 이미지 생성 칩의 섹션의 생성 이미지이다. 상대편 이진 이미지(814)를 생성하기 위해 이러한 그레이스케일 생성 이미지(812)를 변환하기 위한 이미지 변환을 적용한다. 하나의 구현예에서, 이미지 변환기(257)를 참조하여 상기에서 제시된 5개의 단계들의 일부 또는 전부가 수행되어 그레이스케일 이미지를 이진 이미지로 변환할 수 있다. 이진 이미지(814) 내의 흑색 픽셀들은 비정상적인 또는 불량한 이미지 픽셀들을 나타내는 반면, 밝은 픽셀들은 정상적인 또는 양호한 이미지 픽셀들을 나타낸다.

[0086] 우측 상의 예시(816)는 섹션의 다양한 영역들 및 이러한 영역들 주위의 경계들 또는 라인들을 나타내는 섹션의 예시적인 개략도이다. 형광 신호의 세기가 기록되는 영역들은 관심 활성 영역들 또는 부위들로도 지칭된다. 예를 들어, 섹션 개략도(816)는 상단으로부터 하단까지 서로 평행하게 이어지는 4개의 슬롯들의 활성 영역들을 나타낸다. 활성 영역들이 아닌 섹션 이미지의 영역들은 이미지로부터 필터링된다. 예를 들어, 슬롯들의 경계들 또는 경계선들을 나타내는 수직 라인들에 의해 서로 분리된 슬롯들의 경계 영역들. 유사하게, 섹션 이미지의 4개의 측면들 상의 경계들이 필터링될 수 있다. 세그멘테이션 기법은 섹션 이미지들을 4 내지 20개 이상의 세그먼트들 또는 활성 영역들로 분할할 수 있다. 임계화는 활성 영역의 얼마나 많은 부분이 원하는 신호 강도를 생성하고 있는지를 결정한다.

[0087] 활성 영역들의 수는 이미지마다 생성된 특징부들의 수를 결정한다. 예를 들어, 섹션 이미지가 8개의 활성 영역들로 세그먼트화되는 경우, 변환된 이미지의 8개의 활성 영역들로부터의 이미지 세기 및 변환 전의 오리진널 섹션 이미지의 동일한 8개의 활성 영역들로부터의 이미지 세기 값들이 분류자에 대한 입력으로서 주어진다. 따라서, 이러한 예에서, 섹션 이미지당 총 16개의 특징부들이 분류자에 주어질 것이다. 활성 영역으로부터의 신호 강도의 평균 세기는 분류자에 대한 입력으로서 사용될 수 있다. 예를 들어, 섹션 이미지가 8개의 활성 영역들로 세그먼트화되는 경우, 이러한 8개의 활성 영역들의 평균 세기가 그레이스케일 이미지 및 이진 이미지 둘 모두에 대해 계산된다. 이러한 16개의 세기 값들은 섹션 이미지를 양호 대 불량으로 분류하기 위해 분류자에 대한 입력으로 주어진다. 이미지를 이미지당 4, 12, 17개 이상의 세그먼트들과 같은 더 적은 또는 더 많은 세그먼트들로 분할하는 다른 세그멘테이션 스킴들이 사용될 수 있다. 랜덤 포레스트 분류자에 대한 입력으로 주어지는 경우, 특징부들의 서브셋이 각각의 결정 트리에 대해 랜덤으로 선택된다. 결정 트리는 이미지를 정상적인 것으로 또는 비정상적인 것으로 표결(vote)한다. 랜덤 포레스트에서의 대부분의 표결들은 이미지를 분류하는 데 사용된다. 하나의 구현예에서, 랜덤 포레스트 분류자 내의 트리들의 수의 값은 200 내지 500의 범위에 있고, 모델의 깊이의 값은 5 내지 40의 범위에 있다. 이미지 생성 칩의 영역들 및 섹션들 사이에서의 실패들의 패턴들은 근본 원인 분류를 위해 추가로 평가될 수 있다.

[0088] 도 8b는 섹션 이미지들의 3개의 쌍들(822, 824, 826)의 예시(820)를 도시한다. 각각의 쌍 내의 좌측 이미지는 사전변환된 그레이스케일 섹션 이미지이고, 각각의 쌍 내의 우측 이미지는 도 2를 참조하여 기술된 바와 같은 이미지 변환들을 적용한 후의 프로세싱된 이진 이미지이다. 제1 이미지 쌍(822)은 성공적인 유전형분석 프로세스의 생성 이미지이다. 제2 이미지 쌍(824)은 혼성화(또는 hyb) 실패로 인한 실패한 생성 이미지의 생성 이미지이다. 제3 이미지 쌍(826)은 표면 마모 문제로 인한 실패한 이미지의 것이다.

[0089] **1-대-나머지(OvR) 분류**

[0090] 도 9는 1-대-나머지 분류자를 실행하는 그래픽 예시들(900)을 도시한다. 그래프들은 좌측 그래프(931)에 도시된 바와 같이, 3개의 클래스들(정사각형들, 원형들, 및 삼각형들)에 속하는 샘플들로 이루어진 데이터셋에 대해 1-대-나머지(OvR) 분류자를 실행하는 예들을 보여준다. 상단 우측 그래프(915)에 도시된 바와 같은 제1 초평면(hyperplane)(916)은 실측 자료(ground truth)로서 정사각형 클래스에 대한 초평면 결정을 보여준다. 초평면(916)은 나머지 데이터 포인트들(원형들 및 삼각형들)로부터 정사각형 클래스의 데이터 포인트들을 파티셔닝한다. 유사하게, 그래프들(936, 955)은 각각, 초평면들(937, 956)을 통해 데이터 내의 다른 클래스들로부터 원형 클래스 및 삼각형 클래스 내의 데이터 포인트들을 각각 파티셔닝한다. 초평면의 위치는 가중 벡터에 의해 결정된다. 트레이닝 알고리즘은 일반화를 위해 실측 자료 클래스로부터 초평면의 마진을 최대화하려고 시도하지만, 그것은 하나 이상의 데이터 포인트들의 부정확한 분류를 초래할 수 있다. 양호 클래스에 속하는 프로세스 사이클들로부터의 섹션 이미지들을, 다수의 불량 클래스들에 속하는 이미지들과 구별하기 위해 OvR 분류를 적용한다.

[0091] **랜덤 포레스트 분류자들**

[0092] 개시된 기술은 양호한 또는 정상적인 이미지들로부터의 이미지들을, 다수의 실패 클래스들에 속하는 불량한 또는 비정상적인 이미지들과 구별하기 위해 다양한 분류자들을 적용할 수 있다. 적용된 분류자들은 랜덤 포레스트

트, k-최근접 이웃들, 다항 로지스틱 회귀, 및 서포트 벡터 머신들을 포함한다. 일례로서 랜덤 포레스트 분류자를 사용한 개시된 기술의 구현을 제시한다.

[0093] 랜덤 포레스트 분류자(랜덤 결정 포레스트로도 지칭됨)는 앙상블 기계 학습 기법이다. 앙상블된 기법들 또는 알고리즘들은 객체들을 분류하기 위한 동일한 또는 상이한 종류의 하나 이상의 기법을 조합한다. 랜덤 포레스트 분류자는 앙상블로서 동작하는 다수의 결정 트리들로 이루어진다. 랜덤 포레스트 내의 각각의 개별 결정 트리는 베이스 분류자로서 작용하고, 클래스 예측을 출력한다. 대부분의 표결들을 갖는 클래스는 랜덤 포레스트 모델의 예측이 된다. 랜덤 포레스트들 뒤의 기본 개념은 위원회(committee)로서 동작하는 많은 수의 상대적으로 상관되지 않은 모델들(결정 트리들)이 개별 구성 모델들 중 임의의 것을 능가할 것이라는 것이다.

[0094] 개시된 기술은 2-단계형 분류 프로세스에서 랜덤 포레스트 분류자들을 적용한다. 제1 트레이닝된 랜덤 포레스트 분류자는 성공적인 생성 이미지들을 성공적이지 못한 생성 이미지들과 분리하는 태스크를 수행한다. 제2 트레이닝된 랜덤 포레스트 분류자는 성공적이지 못한 이미지의 실패 클래스를 예측함으로써 성공적이지 못한 생성 이미지들의 근본 원인 분석의 태스크를 수행한다. 성공적인 생성물 런들의 우위(dominance)로 인해 이러한 2-단계 분류를 선택하였지만, 1-단계 분류가 또한 사용될 수 있다. 2-단계 접근법을 선택하는 다른 이유는, 그것이 이미지를 정상적인 또는 성공적인 생성 이미지 대 비정상적인 또는 실패한 생성 이미지로서 분류하기 위한 감도 임계치를 제어할 수 있게 한다는 것이다. 제1 단계 분류에서 임계치를 증가시켜, 이에 따라, 분류자가 더 많은 생성 이미지들을 실패한 이미지들로 분류하게 할 수 있다. 이어서, 이러한 실패한 이미지들은 실패 클래스를 식별함으로써 근본 원인 분석을 위해 제2 단계 분류자에 의해 프로세싱된다.

[0095] **랜덤 포레스트 분류자들의 트레이닝**

[0096] 도 10a는 예시(1000)에 도시된 바와 같이 2개의 랜덤 포레스트 분류자들의 트레이닝을 설명한다. 트레이닝 데이터는 도 1에 도시된 바와 같은 트레이닝 데이터베이스(138)에 저장된 라벨링된 프로세스 사이클 이미지들에 대한 입력 특징부들을 포함한다. 분류자들의 하나의 예시적인 트레이닝에서, 섹션들의 20,000개의 라벨링된 생성 이미지들을 사용하였다. 라벨링된 이미지들은 성공적인 생성 사이클들로부터의 양호한 이미지들 및 성공적이지 못한 생성 사이클들로부터의 실패한 이미지들 둘 모두를 포함한다. 트레이닝 데이터베이스(138)의 크기는, 유전형분석 프로세스를 수행하는 실험실들로부터 더 많은 라벨링된 생성 이미지들이 수신됨에 따라 성장할 것이다.

[0097] 하나의 구현예에서, 랜덤 포레스트 분류자들을 트레이닝시키기 위해 라벨링된 생성 이미지들의 성분들의 96개의 가중치들을 사용하였다. 200개의 결정 트리들 및 20의 깊이를 갖는 랜덤 포레스트 분류자가 잘 작동하였다. 200 내지 500개의 결정 트리들의 범위 및 10 내지 40의 깊이의 범위를 갖는 랜덤 포레스트 분류자들이 이러한 구현예에 대한 양호한 결과들을 제공할 것으로 예상된다는 것이 이해된다. 랜덤화된 검색 교차 검증을 사용하여 하이퍼파라미터들을 조정하였다. 깊이에 대한 검색 범위는 5 내지 150이었고, 트리들의 수에 대한 검색 범위는 100 내지 500이었다. 트리들의 수를 증가시키는 것은 모델의 성능을 증가시킬 수 있지만, 그것은 또한, 트레이닝을 위해 요구되는 시간을 증가시킬 수 있다. 20,000개의 생성 사이클 이미지들에 대한 특징부들을 포함하는 트레이닝 데이터베이스(1001)는 양호 대 불량 분류자(151)로서 라벨링되는 이진 분류자를 트레이닝시키는 데 사용된다. 동일한 트레이닝 데이터베이스가, 실패 클래스를 예측하기 위해 근본 원인 분류자(171)를 트레이닝시키는 데 사용될 수 있다. 근본 원인 분류자(171)는 도 10a에 도시된 바와 같이, 불량한 또는 실패한 생성 이미지들만으로 이루어진 트레이닝 데이터베이스(1021) 상에서 트레이닝된다.

[0098] 결정 트리들을 오버피팅되기 쉽다. 이러한 문제를 극복하기 위해, 배깅(bagging) 기법이 랜덤 포레스트 내의 결정 트리들을 트레이닝시키는 데 사용된다. 배깅은 부트스트랩(bootstrap) 및 집적(aggregation) 기법들의 조합이다. 부트스트랩에서, 트레이닝 동안, 본 발명의 트레이닝 데이터베이스로부터 행들의 샘플을 취하고, 이를 사용하여 랜덤 포레스트 내의 각각의 결정 트리를 트레이닝시킨다. 예를 들어, 선택된 행들에 대한 특징부들의 서브셋이 결정 트리 1의 트레이닝에 사용될 수 있다. 따라서, 결정 트리 1에 대한 트레이닝 데이터는 행 샘플 1과 열 샘플 1 또는 RS1+CS1로 지칭될 수 있다. 열들 또는 특징부들은 랜덤으로 선택될 수 있다. 랜덤 포레스트 내의 결정 트리 2 및 후속 결정 트리들은 트레이닝 데이터의 서브셋을 사용함으로써 유사한 방식으로 트레이닝된다. 대체물로 결정 트리들에 대한 트레이닝 데이터가 생성되는데, 즉, 동일한 행 데이터가 다수의 결정 트리들의 트레이닝에 사용될 수 있다는 것에 주목한다.

[0099] 배깅 기법의 제2 부분은 제조 동안 적용되는 집적 부분이다. 각각의 결정 트리는 각각의 클래스에 대한 분류를 출력한다. 이진 분류의 경우, 그것은 1 또는 0일 수 있다. 랜덤 포레스트의 출력은, 랜덤 포레스트의 출력으로서 선택된 대부분의 표결을 갖는 랜덤 포레스트 내의 결정 트리들의 출력들의 집적이다. 다수의 결정 트리들

로부터의 표결들을 사용함으로써, 랜덤 포레스트는 결정 트리들의 결과들에서 높은 분산을 감소시켜, 이에 따라, 양호한 예측 결과들을 초래한다. 행 및 열 샘플링을 사용하여 개별 결정 트리들을 트레이닝시킴으로써, 각각의 결정 트리는 선택된 특징부들을 갖는 트레이닝 기록들과 관련한 전문가가 된다.

[0100] 트레이닝 동안, 랜덤 포레스트의 출력은 실측 자료 라벨들과 비교되고, 예측 오류가 계산된다. 역방향 전파 동안, 96개의 성분들(또는 아이겐 이미지들)의 가중치들이 조정되어, 예측 오류가 감소되도록 한다. 성분들 또는 아이겐 이미지들의 수는 설명된 분산 측정치를 사용하여 주성분 분석(PCA)의 출력으로부터 선택된 성분들의 수에 의존한다. 이진 분류 동안, 양호 대 불량 분류자는 트레이닝 데이터로부터의 이미지 설명 특징부들을 사용하고, 양호 클래스(또는 정상적인 것으로 라벨링된 이미지들) 대 다수의 불량 클래스들(6개의 실패 클래스들 중 하나의 실패 클래스로 라벨링된 이미지들)의 1-대-나머지(OvR) 분류를 적용한다. 트레이닝된 랜덤 포레스트 분류자의 파라미터들(예컨대, 성분들의 가중치들)은 추론 동안 생성 사이클 이미지들의 양호 대 불량 분류에 사용하기 위해 저장된다.

[0101] 근본 원인 분류자(171)의 트레이닝은 유사한 방식으로 수행된다. 트레이닝 데이터베이스(1021)는 다수의 실패 클래스들에 속하는 불량한 프로세스 사이클들로부터의 라벨링된 프로세스 사이클 이미지들로부터의 특징부들을 포함한다. 랜덤 포레스트 분류자(171)는 각각의 실패 클래스 대 나머지 라벨링된 트레이닝 예들의 1-대-나머지(OvR) 분류에 대한 이미지 설명 특징부들을 사용하여 트레이닝된다.

[0102] **랜덤 포레스트 분류자들을 사용한 분류**

[0103] 이제, 트레이닝된 분류자들(151, 171)을 사용한 생성 이미지들의 분류를 설명한다. 도 10b는 제1 단계에서 양호 대 불량 분류자(151)를 사용한 그리고 제2 단계에서 근본 원인 분류자(171)를 사용한 생성 이미지들의 2-단계 분류(1080)를 도시한다. 프로세스는 1 내지 9로 라벨링된 프로세스 흐름 단계들의 시퀀스를 사용하여 제시된다. 프로세스는 양호 대 불량 분류자(151)로 라벨링된 트레이닝된 랜덤 포레스트 분류자에 액세스함으로써 단계 1에서 시작한다. 데이터베이스(1030)에 저장된 생성 이미지들의 입력 특징부들은 분류자(151)에 대한 입력으로서 제공된다. 분류자는 성공적인 프로세스 사이클에 속하는 양호한 이미지들을 실패한 프로세스 사이클들에 속하는 불량 이미지들과 구별한다. 불량한 이미지들은, 예를 들어, 다수의 실패 클래스들에 속하고, 각각의 이미지는 전술된 바와 같이 6개의 실패 클래스들 중 하나에 속할 수 있다. 트레이닝된 분류자는 생성 이미지를 분석하기 위한 아이겐 이미지들의 베이스에 액세스한다. 트레이닝된 분류자는 아이겐 이미지들의 선형 조합에 기초하여 생성 이미지에 대한 이미지 설명 특징부들을 생성한다. 아이겐 이미지들의 가중치들은 전술된 바와 같이, 분류자의 트레이닝 동안 학습된다.

[0104] 1-대-나머지 분류를 적용하기 때문에, 랜덤 포레스트 분류자 내의 모든 결정 트리들은 각각의 클래스에 대한 출력, 즉, 이미지가 7개의 클래스들(하나의 양호 클래스 및 6개의 실패 클래스들) 중 하나의 클래스에 속하는지 여부를 예측한다. 따라서, 랜덤 포레스트 내의 각각의 결정 트리는 7개의 확률 값들, 즉 클래스당 하나의 값을 출력할 것이다. 결정 트리들로부터의 결과들이 집계되고, 대부분의 표결이 이미지를 양호한 것 또는 불량한 것으로 예측하는 데 사용된다. 예를 들어, 랜덤 포레스트 내의 결정 트리들의 50% 초과가 이미지를 양호한 것으로 분류하는 경우, 이미지는 성공적인 생성 사이클에 속하는 양호한 이미지로서 분류된다. 분류자의 감도는, 예를 들어, 임계치를 더 높게 설정함으로써 조정될 수 있는데, 이는 더 많은 이미지들이 불량한 것으로 분류되는 결과를 초래할 것이다. 프로세스 단계 2에서, 분류자(151)로부터의 출력이 체크된다. 이미지가 양호한 이미지로 분류되는 경우(단계 3), 프로세스는 종료된다(단계 4). 그렇지 않은 경우, 이미지가 실패한 프로세스 사이클을 나타내는 불량한 이미지로 분류되는 경우(단계 5), 시스템은 근본 원인 분류자(171)를 호출한다(단계 6).

[0105] 근본 원인 분류자는 불량한 이미지의 실패의 클래스를 결정하기 위해 2-단계 프로세스의 제2 단계에 적용된다. 프로세스는, 불량한 이미지에 대한 생성 이미지 입력 특징부에 액세스하고(단계 7), 트레이닝된 근본 원인 분류자(171)에 대한 입력 특징부들을 제공함으로써(단계 8), 제2 단계에서 계속된다. 근본 원인 분류자(171) 내의 각각의 결정 트리는 1-대-나머지 분류를 적용함으로써 입력 이미지 특징부들에 대해 표결한다. 이러한 경우, 분류는 이미지가 6개의 실패 클래스들 중 하나에 속하는지 아니면 나머지 5개의 실패 클래스들에 속하는지를 결정한다. 각각의 결정 트리는 각각의 클래스에 대한 분류를 제공한다. 결정 트리들로부터의 대부분의 표결은 이미지의 실패 클래스를 결정한다(단계 9).

[0106] 다른 분류자들을 사용하여, 양호한 섹션 이미지들 대 불량한 섹션 이미지들을 분류하고 근본 원인 분석을 수행할 수 있다. 예를 들어, 개시된 기술은 섹션 이미지들을 분류하기 위해 K-최근접 이웃(k-NN 또는 KNN) 알고리즘을 적용할 수 있다. k-NN 알고리즘은 유사한 예들(또는 본 발명의 구현예에서의 섹션 이미지들)이 근접하게

존재한다고 가정한다. k-NN 알고리즘은 데이터 포인트들 또는 이미지들 사이의 거리를 계산함으로써 유사성(근접성 또는 긴밀성(closeness)으로도 지칭됨)의 아이디어를 캡처한다. 이러한 목적을 위해 직선 거리(또는 유클리드 거리)가 흔히 사용된다. k-NN 분류에서, 출력은 클래스 멤버십, 예를 들어, 양호한 이미지 클래스 또는 불량한 이미지 클래스이다. 이미지는 그의 이웃들의 복수의 표결들에 의해 분류되며, 이때 객체는 그의 K개의 최근접 이웃들 사이에서 가장 흔한 클래스에 배정된다. k의 값은 양의 정수이다.

[0107] 데이터에 대한 k의 우측 값을 선택하기 위해, k-NN 알고리즘을 상이한 k의 값들로 수 회 실행하며, 그것이 이전에 보이지 않았던 데이터를 제공받을 때 정확하게 예측들을 행하는 알고리즘의 능력을 유지하면서 직면하는 오류들의 수를 감소시키는 k의 값을 선택한다. k의 값을 1로 설정한다고 가정한다. 이는 부정확한 예측들을 초래할 수 있다. 데이터 포인트들의 2개의 클러스터를 갖는다고 고려한다: 양호한 이미지 및 불량한 이미지. 그러나, 많은 양호한 이미지 데이터 포인트들에 의해 둘러싸인 질의 예(query example)를 갖는 경우, 그것은 양호한 이미지 데이터 포인트들의 클러스터 내에도 있는 하나의 불량한 이미지 데이터 포인트에 가깝다. k=1이면, k-NN은 질의 예가 불량한 이미지라고 부정확하게 예측한다. k의 값을 증가시킴에 따라, k-NN 알고리즘의 예측은 (분류에서의) 대부분의 표결 및 (회귀에서의) 평균화로 인해 더 안정적이 된다. 따라서, 알고리즘은 k의 소정 값에 이르기까지 더 정확한 예측들을 행할 가능성이 더 크다. k의 값이 증가함에 따라, 오류들의 증가하는 수를 관찰하기 시작한다. 6 내지 50 범위의 k의 값은 작동할 것으로 예상된다.

[0108] 개시된 기술에 의해 트레이닝되고 적용될 수 있는 다른 분류자들의 예들은 다항 로지스틱 회귀, 서포트 벡터 머신(SVM)들, 기울기 증대 트리들, 나이브 베이즈 등을 포함한다. 하기의 3개의 기준들을 사용하여 분류자들의 성능을 평가하였다: 트레이닝 시간, 정확도, 및 결과들의 해석능력(interpretability). 랜덤 포레스트 분류자는 다른 분류자들보다 더 양호하게 수행하였다. 하기의 텍스트에서 다른 분류자들을 간략하게 제시한다.

[0109] 서포트 벡터 머신 분류자는 또한, 랜덤 포레스트 분류자와 동등하게 잘 수행하였다. SVM 분류자는 양호 클래스에 대한 특징부 벡터와 다수의 불량 클래스들에 대한 특징부 벡터들 사이에 초평면을 위치시킨다. 개시된 기술은 다항 로지스틱 회귀를 트레이닝시키는 것을 포함할 수 있다. 다항 회귀 모델은 상이한 가능한 결과들(다중 클래스 분류)의 확률들을 예측하기 위해 트레이닝될 수 있다. 모델은 출력이 단정적일 때 사용된다. 따라서, 모델은 이미지가 양호 클래스에 속하는지 아니면 다수의 불량 클래스들 중 하나에 속하는지를 예측하기 위해 트레이닝될 수 있다. 로지스틱 회귀 분류자의 성능은 랜덤 포레스트 및 SVM 분류자들보다 못하였다. 개시된 기술은 결정 트리들과 같은 예측 모델들의 앙상블인 기울기 증대 모델을 트레이닝시키는 것을 포함할 수 있다. 모델은 음수 기울기 방향으로 향하는 함수를 반복적으로 선택함으로써 함수 공간보다 비용 함수를 최적화하려고 시도한다. 예를 들어, 모델은 트레이닝 데이터세트보다 평균 제곱 오차를 최소화하기 위해 트레이닝될 수 있다. 기울기 증대 모델은 다른 분류자들에 비해 더 많은 트레이닝 시간을 요구하였다. 개시된 기술은 특정 특징부의 값이 임의의 다른 특징부의 값과는 독립적이라고 가정하는 나이브 베이즈 분류자를 트레이닝시키는 것을 포함할 수 있다. 나이브 베이즈 분류자는 특징부들 각각이 일레가 클래스에 속하는 확률에 독립적으로 기여하는 것으로 간주한다. 나이브 베이즈 분류자는 양호 클래스 대 다수의 불량 클래스들로 이미지들을 분류하기 위해 트레이닝될 수 있다.

[0110] **특정 구현예들**

[0111] 개시된 기술은 유전형분석 프로세스의 평가 및 근본 원인 분석을 위한 이미지 분류를 적용한다. 랜덤 포레스트 분류자들에 의해 하기의 2개의 태스크들이 수행된다: 성공적인 생성 이미지들과 성공적이지 못한(또는 실패한) 생성 이미지들의 분리, 이어서 성공적이지 못한 이미지들의 근본 원인 분석. 개시된 기술은 유전형분석 프로세스 기구들로부터의 프로세스 사이클 이미지들에 적용될 수 있다. 유전형분석 프로세스로부터의 이미지들을 분류하기 위해 개시된 기술이 설명되지만, 분류는 프로세스 동안 또는 프로세스의 종료 시 이미지 생성 칩들 상에 위치한 샘플들의 이미지들을 생성하는 다른 유형들의 프로세스들로부터의 이미지들에 적용될 수 있다.

[0112] 성공적인 생성 이미지들 및 성공적이지 못한 생성 이미지들의 분류를 먼저 제시한다. 개시된 기술의 하나의 구현예에서, 유전형분석 프로세스 사이클 이미지들을 분류하기 위한 랜덤 포레스트 분류자를 트레이닝시키기 위한 방법이 설명된다. 분류자를 트레이닝시키는 방법은 성공(또는 양호 또는 정상) 클래스 및 다수의 실패(또는 불량 또는 비정상) 클래스들에 속하는 프로세스 사이클들로부터의 이미지들에 대한 라벨링된 트레이닝 예들에 액세스하는 단계를 포함한다. 본 방법은 이미지들을 분석하기 위한 아이젠 이미지들의 베이스에 액세스하는 단계를 포함할 수 있다. 본 방법은 아이젠 이미지들의 선형 조합에 기초하여 각각의 라벨링된 트레이닝 예에 대한 이미지 설명 특징부들을 생성하는 단계를 포함한다. 본 방법은 라벨링된 트레이닝 예들의 성공 클래스 대 다수의 실패 클래스들의 1-대-나머지 결정을 위한 이미지 설명 특징부들을 사용하여 랜덤 포레스트 분류자를 트

레이닝시키는 단계를 포함한다. 본 방법은 라벨링된 트레이닝 예들의 성공 클래스 대 다수의 실패 클래스들의 1대-나머지 결정을 위한 다른 분류자들을 트레이닝시키는 단계를 포함할 수 있다. 예를 들어, 본 방법은 서포트 벡터 머신(SVM) 분류자를 트레이닝시키는 단계를 포함할 수 있다. SVM 분류자는 성공 클래스에 대한 특징부 벡터와 나머지 클래스들(또는 본 발명의 경우에서의 다수의 실패 클래스들)에 대한 특징부 벡터들 사이에 초평면을 위치시킨다. 본 방법은 다항 로지스틱 회귀를 트레이닝시키는 단계를 포함할 수 있다. 다항 회귀 모델은 상이한 가능한 결과들(다중클래스 분류)의 확률들을 예측하기 위해 트레이닝될 수 있다. 모델은 출력에 단정적일 때 사용된다. 따라서, 모델은 이미지가 성공 클래스에 속하는지 아니면 다수의 실패 클래스들 중 하나에 속하는지를 예측하기 위해 트레이닝될 수 있다. 본 방법은 결정 트리들과 같은 예측 모델들의 앙상블인 기울기 증대 모델을 트레이닝시키는 단계를 포함할 수 있다. 모델은 음수 기울기 방향으로 향하는 함수를 반복적으로 선택함으로써 함수 공간보다 비용 함수를 최적화하려고 시도한다. 예를 들어, 모델은 트레이닝 데이터세트보다 평균 제곱 오차를 최소화하기 위해 트레이닝될 수 있다. 본 방법은 특정 특징부의 값이 임의의 다른 특징부의 값과는 독립적이라고 가정하는 나이브 베이즈 분류자를 트레이닝시키는 단계를 포함할 수 있다. 나이브 베이즈 분류자는 특징부들 각각이 일레가 클래스에 속하는 확률에 독립적으로 기여하는 것으로 간주한다. 나이브 베이즈 분류자는 성공 클래스 대 다수의 실패 클래스들로 이미지들을 분류하기 위해 트레이닝될 수 있다. 트레이닝된 분류자의 파라미터들은 프로세스 사이클 이미지들의 성공 대 실패 분류들의 생성에 사용하기 위해 저장된다. 진술된 분류자들 중 임의의 것을 사용하여 생성 이미지들을 분류하는 것은 실패한 생성물 런들을 검출할 수 있다. 분류자들은 유전형분석 프로세스의 품질에 관해 운용자들에게 빠른 피드백을 제공할 수 있다. 분류자들로부터의 피드백은 프로세싱 시간 및 가치있는 샘플들의 낭비를 감소시킬 수 있는 업스트림 프로세스를 보정하는데 사용될 수 있다.

[0113] 유전형분석 프로세스 사이클 이미지들에 대한 분류자를 트레이닝시키는 방법은 이미지들을 분석하기 위한 아이젠 이미지들의 베이스를 생성하는 단계를 추가로 포함할 수 있다. 본 방법은 아이젠 이미지들의 베이스에 액세스하는 단계를 포함할 수 있다. 아이젠 이미지들의 베이스는 설명된 가변성의 측정치에 따라 정렬된다. 임계치 초과된 가변성을 점증적으로 설명하는 아이젠 이미지들의 상단 정렬된 베이스가 프로세스 사이클 이미지들을 분석하기 위해 선택된다. 본 방법은 아이젠 이미지의 선택된 베이스를 사용하여 프로세스 사이클 이미지들을 분석하는 단계를 포함할 수 있다. 랜덤 포레스트 분류자는 100 내지 400개의 결정 트리들을 포함할 수 있다. 랜덤 포레스트 분류자의 깊이는 10 내지 40일 수 있다.

[0114] 생성 구현예에서, 본 방법은 생성 프로세스 사이클 이미지들을 성공한 것 또는 실패한 것으로 분류한다. 본 방법은 성공 클래스에 속하는 프로세스 사이클들로부터의 이미지들을 다수의 실패 클래스들에 속하는 이미지들과 구별하기 위해 트레이닝된 랜덤 포레스트 분류자에 액세스하는 단계를 포함한다. 본 방법은 생성 이미지를 분석하기 위한 아이젠 이미지들의 베이스에 액세스하는 단계 및 아이젠 이미지들의 선형 조합에 기초하여 생성 이미지에 대한 이미지 설명 특징부들을 생성하는 단계를 포함할 수 있다. 본 방법은 트레이닝된 랜덤 포레스트 분류자를 성공 클래스 대 다수의 실패 클래스들의 1대-나머지 결정을 위한 이미지 설명 특징부들에 적용한다. 본 방법은 실패한 프로세스 사이클 결정에 기초하여 프로세스 사이클을 반복하기 위해 사용자 인터페이스 상에 통지를 제공하는 단계를 포함한다.

[0115] 이제, 성공적이지 못한(또는 실패한) 이미지들의 근본 원인 분석을 제시한다. 개시된 기술의 하나의 구현예에서, 프로세스 사이클 이미지들 또는 유전형분석 프로세스 사이클 이미지들에 대한 근본 원인 랜덤 포레스트 분류자를 트레이닝시키기 위한 방법이 설명된다. 분류자를 트레이닝시키는 방법은 다수의 실패 클래스들에 속하는 실패한 프로세스 사이클들로부터의 이미지들에 대한 라벨링된 트레이닝 예들에 액세스하는 단계를 포함한다. 본 방법은 또한, 이미지들을 분석하기 위한 아이젠 이미지들의 베이스에 액세스하는 단계 및 아이젠 이미지들의 선형 조합에 기초하여 각각의 라벨링된 트레이닝 예에 대한 이미지 설명 특징부들을 생성하는 단계를 포함한다. 본 방법은 라벨링된 트레이닝 예들의 각각의 실패 클래스 대 나머지 클래스들의 1대-나머지 결정을 위한 이미지 설명 특징부들을 사용하여 랜덤 포레스트 분류자를 트레이닝시키는 단계를 포함한다. 트레이닝된 랜덤 포레스트 분류자의 파라미터들은 실패한 프로세스 사이클들로부터의 프로세스 사이클 이미지들의 근본 원인 분류들의 생성에 사용하기 위해 저장된다.

[0116] 근본 원인 분류자를 트레이닝시키는 방법은 이미지들을 분석하기 위한 아이젠 이미지들의 베이스를 생성하는 단계를 추가로 포함할 수 있다. 아이젠 이미지들의 베이스는 설명된 가변성의 측정치에 따라 정렬된다. 임계치 초과된 가변성을 점증적으로 설명하는 아이젠 이미지들의 상단 정렬된 베이스가 프로세스 사이클 이미지들을 분석하기 위해 선택된다. 랜덤 포레스트 분류자는 100 내지 400개의 결정 트리들을 포함할 수 있다. 랜덤 포레스트 분류자의 깊이는 10 내지 40일 수 있다.

- [0117] 생성 구현예에서, 실패한 프로세스 사이클 이미지로부터의 근본 원인 분석의 방법은 실패한 생성 이미지를 분석하기 위한 아이젠 이미지들의 베이스에 액세스하는 단계 및 아이젠 이미지들의 선형 조합에 기초하여 실패한 생성 이미지에 대한 이미지 설명 특징부들을 생성하는 단계를 포함한다. 본 방법은 이미지들을 다수의 실패 클래스들 중 하나에 속하는 프로세스 사이클들과 구별하기 위해 트레이닝된 랜덤 포레스트 분류자에 액세스하는 단계를 포함한다. 본 방법은 랜덤 포레스트 분류자를 다수의 실패 클래스들 중에서 1대-나머지 결정을 위한 이미지 설명 특징부들에 적용하는 단계 - 이는 다수의 실패 클래스들 각각 대 나머지를 스코어링하는 단계 및 생성된 스코어들을 사용하여 불량한 프로세스 사이클의 가능성있는 근본 원인으로 다수의 실패 클래스들 사이에서 선택하는 단계를 포함함 - 를 포함한다.
- [0118] 개시된 기술은 다른 특징부 생성 및 차원수 감소 기법들을 사용하여 분류자들에 대한 입력을 생성할 수 있다. 그러한 기법들의 예들은 비-음수 행렬 인수분해(NMF), 독립 성분 분석, 사전 학습, 희소 주성분 분석, 요인 분석, 미니 배치 K-평균을 포함한다. 이미지 분해 및 차원수 감소 기법들의 변형들이 사용될 수 있다. 개시된 기술은 또한, 프로세스 사이클 이미지들을 분류하기 위해 랜덤 포레스트 분류자들 이외의 분류자들을 적용할 수 있다. 적용되는 분류자들은 랜덤 포레스트, K-최근접 이웃(KNN), 다항 로지스틱 회귀, 서포트 벡터 머신(SVM), 기울기 증대 트리, 나이브 베이즈 등을 포함할 수 있다. 라벨링된 이미지들의 더 큰 본체들이 이용가능해짐에 따라, 콘볼루션 신경 네트워크들, 예컨대 ImageNet가 또한 사용될 수 있다.
- [0119] 하나의 구현예에서, 단일 방법이 단-대-단 방식으로 근본 원인 분류자 및 성공 대 실패 분류자를 구현할 수 있다. 입력 이미지들은 성공적인 이미지들을 실패한 이미지들로부터 분리하는 제1 분류자에 대한 입력으로서 주어진다. 본 방법은 이미지 설명 특징부들이 성공 클래스에 속하지 않는다는 결정 시에 제2 랜덤 포레스트 분류자에 액세스하는 단계를 포함한다. 제2 랜덤 포레스트 분류자는 이미지들을 다수의 실패 클래스들 중 하나에 속하는 프로세스 사이클들과 구별하기 위해 트레이닝된다. 본 방법은 제2 랜덤 포레스트 분류자를 다수의 실패 클래스들 사이의 1대-나머지 결정을 위한 이미지 설명 특징부들에 적용하는 단계를 포함한다. 본 방법은 또한, 다수의 실패 클래스들 각각 대 나머지를 스코어링하는 단계 및 생성된 스코어들을 사용하여 불량한 프로세스 사이클의 가능성있는 근본 원인으로 다수의 실패 클래스들 사이에서 선택하는 단계를 포함한다.
- [0120] 개시된 기술의 다른 구현예는 프로세스 사이클 이미지들을 분류하기 위한 섹션들의 이미지들의 영역들의 임계화를 사용한다. 본 방법은 그레이스케일 생성 사이클 이미지 내의 픽셀들의 세기들로부터, 이미지 픽셀들을 밝음 클래스 및 어둠 클래스로 분류하여 그레이스케일 이미지를 이진 이미지로 변환하기 위한 임계치를 결정하고 적용하는 단계를 포함한다. 본 방법은 밝은 이미지 픽셀 분류 및 어두운 이미지 픽셀 분류를 생성하는 단계를 포함할 수 있다. 본 방법은 그레이스케일 이미지와 변환된 이진 이미지를, 이미지 내의 시스템적인 예상 잡음(systematic expected noise)을 신호와 분리하는 8개 이상의 영역들로 세그먼트화하는 단계를 포함한다. 본 방법은 그레이스케일 이미지 및 변환된 이진 이미지 내의 영역들에 대한 평균 세기 값들을 계산하는 단계를 포함한다. 본 방법은 그레이스케일 및 이진 이미지들 내의 영역들에 대한 평균 세기 값들에 기초하여, 프로세스 사이클들로부터의 이미지들을 분류하기 위해 트레이닝된 랜덤 포레스트 분류자에 액세스하는 단계를 포함한다. 이미지들은 성공 클래스 또는 실패 클래스에 속하는 것으로 분류되고, 신뢰도 스코어가 분류를 위해 생성된다. 본 방법은 트레이닝된 랜덤 포레스트 분류자에 의해 생성된 신뢰도 스코어를 임계치와 비교하는 단계 및 이미지를 성공한 것 또는 실패한 것으로 분류하는 단계를 포함할 수 있다.
- [0121] 하나의 구현예에서, 상기에 제시된 방법은 생성 사이클 이미지를 변환하기 전에 생성 사이클 이미지에서 예지들을 보존하기 위해 쌍방향 필터를 적용하는 단계를 추가로 포함한다.
- [0122] 하나의 구현예에서, 본 방법은 생성 사이클 이미지를 변환한 후에 이진 이미지로부터 스펙클-유사 잡음을 제거하기 위해 가우스 블러 필터를 적용하는 단계를 추가로 포함한다.
- [0123] 하나의 구현예에서, 본 방법은 생성 사이클 이미지를 변환한 후에 이진 이미지 내의 구멍들을 충전하기 위해 모폴로지 동작들을 적용하는 단계를 추가로 포함하며, 여기서 모폴로지 동작들은 침식 및 확장을 포함한다.
- [0124] 위에서 설명한 컴퓨터 구현 방법은 컴퓨터 하드웨어를 포함하는 시스템에서 실행될 수 있다. 컴퓨터 구현 시스템은 위에 설명된 방법 중 하나 이상을 실행할 수 있다. 컴퓨터 구현 시스템은 시스템에 의해 구현된 방법에 적용되는 바로 위에 또는 본 출원 전체에 걸쳐 설명된 방법의 임의의 특징을 통합할 수 있다. 간결함을 위해, 시스템 기능의 대안적 조합은 개별적으로 열거되지 않는다. 시스템들, 방법들, 및 제조 물품들에 적용가능한 특징부들은 기본 특징부들의 각각의 법정 클래스 세트에 대해 반복되지 않는다. 독자는 이 섹션에서 식별된 특징부가 다른 법정 클래스의 기본 특징부와 어떻게 쉽게 조합될 수 있는지 이해할 것이다.

- [0125] 방법이라기 보다는 제조 물품으로서, 비밀시적 컴퓨터 판독가능 매체(CRM)에는 프로세서에 의해 실행가능한 프로그램 명령어가 로딩될 수 있다. 프로그램 명령어들은, 실행될 때, 전송된 컴퓨터 구현 방법들 중 하나 이상을 구현한다. 대안적으로, 프로그램 명령어는 비밀시적 CRM 상에 로딩될 수 있고, 적절한 하드웨어와 조합될 때, 개시된 방법들을 실시하는 하나 이상의 컴퓨터 구현 시스템들의 컴포넌트가 될 수 있다.
- [0126] 방법 구현예에 대한 이 특정 구현 섹션에서 논의된 특징부들 각각은 CRM 및 시스템 구현예들에 동등하게 적용된다. 상기에 나타난 바와 같이, 모든 방법 특징부들은 간결함을 위해 본 명세서에서 반복되지 않으며, 참조로 반복되는 것으로 간주되어야 한다.
- [0127] **조항들**
- [0128] 1. 유전형분석 프로세스 사이클 이미지들에 대한 랜덤 포레스트 분류자를 트레이닝시키는 방법으로서,
- [0129] 성공 클래스 및 다수의 실패 클래스들에 속하는 프로세스 사이클들로부터의 이미지들에 대한 라벨링된 트레이닝 예들에 액세스하는 단계;
- [0130] 아이겐 이미지들의 선형 조합에 기초하여 각각의 라벨링된 트레이닝 예에 대한 이미지 설명 특징부들을 생성하는 단계;
- [0131] 라벨링된 트레이닝 예들의 이미지 설명 특징부들을 사용하여 랜덤 포레스트 분류자를 트레이닝시키는 단계; 및
- [0132] 트레이닝된 랜덤 포레스트 분류자의 파라미터들을 저장하는 단계를 포함하는, 방법.
- [0133] 2. 항목 1에 있어서,
- [0134] 아이겐 이미지들의 베이스스에 액세스하는 단계;
- [0135] 설명된 가변성의 측정치에 따라 아이겐 이미지들의 베이스스를 정렬하는 단계;
- [0136] 임계치 초과와 가변성을 점증적으로 설명하는 아이겐 이미지들의 상단 정렬된 베이스스를 선택하는 단계; 및
- [0137] 아이겐 이미지들의 선택된 베이스스를 사용하여 프로세스 사이클 이미지들을 분석하는 단계를 추가로 포함하는, 방법.
- [0138] 3. 항목 1에 있어서, 성공 클래스 대 다수의 실패 클래스들의 1-대-나머지 결정을 위한 이미지 설명 특징부들을 사용하여 랜덤 포레스트 분류자를 트레이닝시키는 단계를 추가로 포함하는, 방법.
- [0139] 4. 항목 1에 있어서, 랜덤 포레스트 분류자는 100 내지 400개의 결정 트리들을 포함하는, 방법.
- [0140] 5. 항목 1에 있어서, 랜덤 포레스트 분류자는 10 내지 40의 깊이를 갖는, 방법.
- [0141] 6. 유전형분석 프로세스 사이클 이미지들을 분류하는 방법으로서,
- [0142] 성공 클래스에 속하는 프로세스 사이클들로부터의 이미지들을 다수의 실패 클래스들에 속하는 이미지들과 구별하기 위해 트레이닝된 랜덤 포레스트 분류자에 액세스하는 단계;
- [0143] 아이겐 이미지들의 선형 조합에 기초하여 생성 이미지에 대한 이미지 설명 특징부들을 생성하는 단계;
- [0144] 랜덤 포레스트 분류자를 이미지 설명 특징부들에 적용하는 단계; 및
- [0145] 실패한 프로세스 사이클 결정에 기초하여 프로세스 사이클을 반복하기 위해 사용자 인터페이스 상에 통지를 제공하는 단계를 포함하는, 방법.
- [0146] 7. 항목 6에 있어서, 성공 클래스 대 다수의 실패 클래스들의 1-대-나머지 결정을 위한 이미지 설명 특징부들을 사용하여 랜덤 포레스트 분류자를 트레이닝시키는 단계를 추가로 포함하는, 방법.
- [0147] 8. 유전형분석 프로세스 사이클 이미지들에 대한 근본 원인 랜덤 포레스트 분류자를 트레이닝시키는 방법으로서,
- [0148] 다수의 실패 클래스들에 속하는 실패한 프로세스 사이클들로부터의 이미지들에 대한 라벨링된 트레이닝 예들에 액세스하는 단계;
- [0149] 아이겐 이미지들의 선형 조합에 기초하여 각각의 라벨링된 트레이닝 예에 대한 이미지 설명 특징부들을 생성하는 단계;

- [0150] 라벨링된 트레이닝 예들의 이미지 설명 특징부들을 사용하여 랜덤 포레스트 분류자를 트레이닝시키는 단계; 및
- [0151] 트레이닝된 랜덤 포레스트 분류자의 파라미터들을 저장하는 단계를 포함하는, 방법.
- [0152] 9. 항목 8에 있어서, 성공 클래스 대 다수의 실패 클래스들의 1대-나머지 결정을 위한 이미지 설명 특징부들을 사용하여 랜덤 포레스트 분류자를 트레이닝시키는 단계를 추가로 포함하는, 방법.
- [0153] 10. 실패한 프로세스 사이클 이미지로부터의 근본 원인 분석 방법으로서,
- [0154] 실패한 생성 이미지를 분석하기 위한 아이젠 이미지들의 베이스에 액세스하는 단계;
- [0155] 아이젠 이미지들의 선형 조합에 기초하여 실패한 생성 이미지에 대한 이미지 설명 특징부들을 생성하는 단계;
- [0156] 이미지들을 다수의 실패 클래스들 중 하나에 속하는 프로세스 사이클들과 구별하기 위해 트레이닝되는 랜덤 포레스트 분류자에 액세스하는 단계; 및
- [0157] 랜덤 포레스트 분류자를 다수의 실패 클래스들 중에서 1대-나머지 결정을 위한 이미지 설명 특징부들에 적용하는 단계 - 이는 다수의 실패 클래스들 각각 대 나머지를 스코어링하는 단계 및 생성된 스코어들을 사용하여 실패한 프로세스 사이클의 가능성있는 근본 원인으로 다수의 실패 클래스들 사이에서 선택하는 단계를 포함함 - 를 포함하는, 방법.
- [0158] 11. 항목 2에 있어서,
- [0159] 이미지 설명 특징부들이 성공 클래스에 속하지 않는다는 결정 시에 제2 랜덤 포레스트 분류자에 액세스하는 단계 - 제2 랜덤 포레스트 분류자는 이미지들을 다수의 실패 클래스들 중 하나에 속하는 프로세스 사이클들과 구별하기 위해 트레이닝됨 -; 및
- [0160] 제2 랜덤 포레스트 분류자를 이미지 설명 특징부들에 적용하는 단계 - 이는 다수의 실패 클래스들 각각 대 나머지를 스코어링하는 단계 및 생성된 스코어들을 사용하여 실패한 프로세스 사이클의 가능성있는 근본 원인으로 다수의 실패 클래스들 사이에서 선택하는 단계를 포함함 - 를 추가로 포함하는, 방법.
- [0161] 12. 유전형분석 프로세스 사이클 이미지들을 분류하는 방법으로서,
- [0162] 그레이스케일 생성 사이클 이미지 내의 픽셀들의 세기들로부터, 이미지 픽셀들을 밝음 클래스 및 어둠 클래스로 분류하여 그레이스케일 이미지를 이진 이미지로 변환하기 위한 임계치를 결정하고 적용하는 단계;
- [0163] 그레이스케일 이미지 및 변환된 이진 이미지를, 이미지 내의 시스템적인 예상 잡음을 신호와 분리하는 8개 이상의 영역들로 세그먼트화하는 단계;
- [0164] 그레이스케일 이미지 및 변환된 이진 이미지 내의 영역들에 대한 평균 세기 값들을 계산하는 단계; 및
- [0165] 그레이스케일 및 이진 이미지들 내의 영역들에 대한 평균 세기 값들에 기초하여, 프로세스 사이클들로부터의 이미지들을 성공 클래스 또는 실패 클래스에 속하는 것으로 분류하기 위해 트레이닝된 랜덤 포레스트 분류자에 액세스하고, 트레이닝된 랜덤 포레스트 분류자에 의해 생성된 신뢰도 스코어를 임계치와 비교하고, 이미지를 성공 또는 실패로 분류하는 단계를 포함하는, 방법.
- [0166] 13. 항목 12에 있어서,
- [0167] 생성 사이클 이미지를 변환하기 전에 생성 사이클 이미지에서 에지들을 보존하기 위해 쌍방향 필터를 적용하는 단계를 추가로 포함하는, 방법.
- [0168] 14. 항목 12에 있어서,
- [0169] 생성 사이클 이미지를 변환한 후에 이진 이미지로부터 스펙클-유사 잡음을 제거하기 위해 가우스 블러 필터를 적용하는 단계를 추가로 포함하는, 방법.
- [0170] 15. 항목 12에 있어서,
- [0171] 생성 사이클 이미지를 변환한 후에 이진 이미지 내의 구멍들을 충전하기 위해 모폴로지 동작들을 적용하는 단계 - 모폴로지 동작들은 침식 및 확장을 포함함 - 를 추가로 포함하는, 방법.
- [0172] 16. 유전형분석 프로세스 사이클 이미지들에 대한 랜덤 포레스트 분류자를 트레이닝시키기 위해 컴퓨터 프로그램 명령어들이 저장된 비일시적 컴퓨터 판독가능 저장 매체로서, 명령어들은, 프로세서 상에서 실행될 때,

- [0173] 성공 클래스 및 다수의 실패 클래스들에 속하는 프로세스 사이클들로부터의 이미지들에 대한 라벨링된 트레이닝 예들에 액세스하는 단계;
- [0174] 아이겐 이미지들의 선형 조합에 기초하여 각각의 라벨링된 트레이닝 예에 대한 이미지 설명 특징부들을 생성하는 단계;
- [0175] 라벨링된 트레이닝 예들의 이미지 설명 특징부들을 사용하여 랜덤 포레스트 분류자를 트레이닝시키는 단계; 및
- [0176] 트레이닝된 랜덤 포레스트 분류자의 파라미터들을 저장하는 단계를 포함하는 방법을 구현하는, 비일시적 컴퓨터 판독가능 저장 매체.
- [0177] 17. 항목 16에 있어서,
- [0178] 성공 클래스 대 다수의 실패 클래스들의 1-대-나머지 결정을 위한 이미지 설명 특징부들을 사용하여 랜덤 포레스트 분류자를 트레이닝시키는 단계를 추가로 포함하는 방법을 추가로 구현하는, 비일시적 컴퓨터 판독가능 저장 매체.
- [0179] 18. 항목 16에 있어서,
- [0180] 아이겐 이미지들의 베이스스에 액세스하는 단계;
- [0181] 설명된 가변성의 측정치에 따라 아이겐 이미지들의 베이스스를 정렬하는 단계;
- [0182] 임계치 초과된 가변성을 점증적으로 설명하는 아이겐 이미지들의 상단 정렬된 베이스스를 선택하는 단계; 및
- [0183] 아이겐 이미지들의 선택된 베이스스를 사용하여 프로세스 사이클 이미지들을 분석하는 단계를 추가로 포함하는 방법을 구현하는, 비일시적 컴퓨터 판독가능 저장 매체.
- [0184] 19. 항목 16에 있어서, 랜덤 포레스트 분류자는 100 내지 400개의 결정 트리들을 포함하는, 비일시적 컴퓨터 판독가능 저장 매체.
- [0185] 20. 항목 16에 있어서, 랜덤 포레스트 분류자는 10 내지 40의 깊이를 갖는, 비일시적 컴퓨터 판독가능 저장 매체.
- [0186] 21. 항목 18에 있어서,
- [0187] 이미지 설명 특징부들이 성공 클래스에 속하지 않는다는 결정 시에 제2 랜덤 포레스트 분류자에 액세스하는 단계 - 제2 랜덤 포레스트 분류자는 이미지들을 다수의 실패 클래스들 중 하나에 속하는 프로세스 사이클들과 구별하기 위해 트레이닝됨 -; 및
- [0188] 제2 랜덤 포레스트 분류자를 이미지 설명 특징부들에 적용하는 단계 - 이는 다수의 실패 클래스들 각각 대 나머지를 스코어링하는 단계 및 생성된 스코어들을 사용하여 실패한 프로세스 사이클의 가능성있는 근본 원인으로서 다수의 실패 클래스들 사이에서 선택하는 단계를 포함함 - 를 추가로 포함하는 방법을 구현하는, 비일시적 컴퓨터 판독가능 저장 매체.
- [0189] 22. 유전형분석 프로세스 사이클 이미지들을 분류하기 위해 컴퓨터 프로그램 명령어들이 저장된 비일시적 컴퓨터 판독가능 저장 매체로서, 명령어들은, 프로세서 상에서 실행될 때,
- [0190] 그레이스케일 생성 사이클 이미지 내의 픽셀들의 세기들로부터, 이미지 픽셀들을 밝음 클래스 및 어둠 클래스로 분류하여 그레이스케일 이미지를 이진 이미지로 변환하기 위한 임계치를 결정하고 적용하는 단계;
- [0191] 그레이스케일 이미지 및 변환된 이진 이미지를, 이미지 내의 시스템적인 예상 잡음을 신호와 분리하는 8개 이상의 영역들로 세그먼트화하는 단계;
- [0192] 그레이스케일 이미지 및 변환된 이진 이미지 내의 영역들에 대한 평균 세기 값들을 계산하는 단계; 및
- [0193] 그레이스케일 및 이진 이미지들 내의 영역들에 대한 평균 세기 값들에 기초하여, 프로세스 사이클들로부터의 이미지들을 성공 클래스 또는 실패 클래스에 속하는 것으로 분류하기 위해 트레이닝된 랜덤 포레스트 분류자에 액세스하고, 트레이닝된 랜덤 포레스트 분류자에 의해 생성된 신뢰도 스코어를 임계치와 비교하고, 이미지를 성공 또는 실패로 분류하는 단계를 포함하는 방법을 구현하는, 비일시적 컴퓨터 판독가능 저장 매체.
- [0194] 23. 항목 22에 있어서,

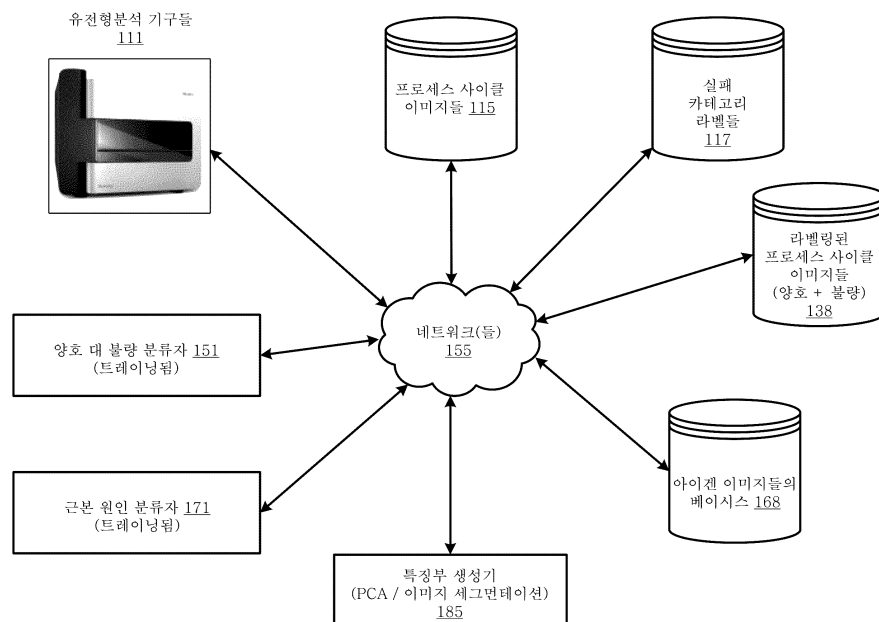
- [0195] 생성 사이클 이미지를 변환하기 전에 생성 사이클 이미지에서 에지들을 보존하기 위해 쌍방향 필터를 적용하는 단계를 추가로 포함하는 방법을 구현하는, 비일시적 컴퓨터 판독가능 저장 매체.
- [0196] 24. 메모리에 커플링된 하나 이상의 프로세서들을 포함하는 시스템으로서, 메모리에는 프로세스 사이클 이미지들에 대한 랜덤 포레스트 분류자를 트레이닝시키기 위한 컴퓨터 명령어들이 로딩되어, 프로세서들 상에서 실행될 때, 항목 16의 명령어들을 구현하는, 시스템.
- [0197] 25. 항목 24에 있어서,
- [0198] 아이겐 이미지들의 베이스스에 액세스하는 것;
- [0199] 설명된 가변성의 측정치에 따라 아이겐 이미지들의 베이스스를 정렬하는 것;
- [0200] 임계치 초과와 가변성을 점증적으로 설명하는 아이겐 이미지들의 상단 정렬된 베이스스를 선택하는 것; 및
- [0201] 아이겐 이미지들의 선택된 베이스스를 사용하여 프로세스 사이클 이미지들을 분석하는 것을 포함하는 액션들을 추가로 구현하는, 시스템.
- [0202] 26. 항목 24에 있어서, 랜덤 포레스트 분류자는 100 내지 400개의 결정 트리들을 포함하는, 시스템.
- [0203] 27. 항목 24에 있어서, 랜덤 포레스트 분류자는 10 내지 40의 깊이를 갖는, 시스템.
- [0204] 28. 항목 25에 있어서,
- [0205] 이미지 설명 특징부들이 성공 클래스에 속하지 않는다는 결정 시에 제2 랜덤 포레스트 분류자에 액세스하는 것 - 제2 랜덤 포레스트 분류자는 이미지들을 다수의 실패 클래스들 중 하나에 속하는 프로세스 사이클들과 구별하기 위해 트레이닝됨 -; 및
- [0206] 제2 랜덤 포레스트 분류자를 이미지 설명 특징부들에 적용하는 것 - 이는 다수의 실패 클래스들 각각 대 나머지를 스코어링하는 것 및 생성된 스코어들을 사용하여 실패한 프로세스 사이클의 가능성있는 근본 원인으로 다수의 실패 클래스들 사이에서 선택하는 것을 포함함 - 을 포함하는 액션들을 추가로 구현하는, 시스템.
- [0207] 29. 메모리에 커플링된 하나 이상의 프로세서들을 포함하는 시스템으로서, 메모리에는 프로세스 사이클 이미지들을 분류하기 위한 컴퓨터 명령어들이 로딩되어, 프로세서들 상에서 실행될 때, 항목 22의 명령어들을 구현하는, 시스템.
- [0208] **컴퓨터 시스템**
- [0209] 도 11은 개시된 기술을 구현하는 데 사용될 수 있는 컴퓨터 시스템(1100)의 단순화된 블록도이다. 컴퓨터 시스템은 전형적으로, 버스 서브시스템(1155)을 통해 다수의 주변 디바이스와 통신하는 적어도 하나의 프로세서(1172)를 포함한다. 이러한 주변 디바이스들은, 예를 들어 메모리 서브시스템(1122) 및 파일 저장 서브시스템(1136)을 포함하는 저장 서브시스템(1110), 사용자 인터페이스 입력 디바이스들(1138), 사용자 인터페이스 출력 디바이스들(1176), 및 네트워크 인터페이스 서브시스템(1174)을 포함할 수 있다. 입력 및 출력 디바이스는 컴퓨터 시스템과의 사용자 상호작용을 허용한다. 네트워크 인터페이스 서브시스템은 인터페이스를 다른 컴퓨터 시스템의 대응하는 인터페이스 디바이스에 대한 인터페이스를 포함하는 외부 네트워크에 제공한다.
- [0210] 하나의 구현예에서, 실패한(또는 불량한) 이미지들을 분류하기 위한 근본 원인 분류자(171)는 저장 서브시스템 및 사용자 인터페이스 입력 디바이스들에 통신 가능하게 연결된다.
- [0211] 사용자 인터페이스 입력 디바이스들(1138)은 키보드; 마우스, 트랙볼, 터치패드, 또는 그래픽 태블릿과 같은 포인팅 디바이스들; 스캐너; 디스플레이 내에 통합된 터치 스크린; 음성 인식 시스템들 및 마이크론들과 같은 오디오 입력 디바이스들; 및 다른 유형들의 입력 디바이스들을 포함할 수 있다. 일반적으로, 용어 "입력 디바이스"의 사용은 정보를 컴퓨터 시스템에 입력하기 위한 모든 가능한 유형의 디바이스 및 방식을 포함하도록 의도된다.
- [0212] 사용자 인터페이스 출력 디바이스들(1176)은 디스플레이 서브시스템, 프린터, 팩스 기계, 또는 오디오 출력 디바이스들과 같은 비시각적 디스플레이들을 포함할 수 있다. 디스플레이 서브시스템은 음극선관(CRT), 액정 디스플레이(LCD)와 같은 평면 디바이스, 프로젝션 디바이스, 또는 가시적인 이미지를 생성하기 위한 일부 다른 메커니즘을 포함할 수 있다. 디스플레이 서브시스템은 또한 오디오 출력 디바이스와 같은 비시각적 디스플레이를 제공할 수 있다. 대체적으로, "출력 디바이스"라는 용어의 사용은 정보를 컴퓨터 시스템으로부터 사용자에게

또는 다른 기계 또는 컴퓨터 시스템에 출력하기 위한 모든 가능한 유형들의 디바이스들 및 방식들을 포함하도록 의도된다.

- [0213] 저장 서브시스템(1110)은 본 명세서에 기술된 모듈들 및 방법들 중 일부 또는 전부의 기능을 제공하는 프로그래밍 및 데이터 구성들을 저장한다. 이들 소프트웨어 모듈은 일반적으로 프로세서에 의해 단독으로 또는 다른 프로세서와 조합으로 실행된다.
- [0214] 저장 서브시스템에 사용되는 메모리는 프로그램 실행 동안 명령어 및 데이터의 저장을 위한 메인 랜덤 액세스 메모리(RAM)(1132) 및 고정 명령어가 저장되는 판독 전용 메모리(ROM)(1134)를 포함하는 다수의 메모리를 포함할 수 있다. 파일 저장 서브시스템(1136)은 프로그램 및 데이터 파일들을 위한 영구 저장소를 제공할 수 있고, 하드 디스크 드라이브, 연관된 착탈식 매체와 함께 플로피 디스크 드라이브, CD-ROM 드라이브, 광학 드라이브, 또는 착탈식 매체 카트리지를 포함할 수 있다. 소정 구현예의 기능을 구현하는 모듈은 저장 서브시스템 내의 파일 저장 서브시스템 또는 프로세서에 의해 액세스 가능한 다른 기계에 저장될 수 있다.
- [0215] 버스 서브시스템(1155)은 컴퓨터 시스템의 다양한 컴포넌트 및 서브시스템이 의도된 대로 서로 통신하게 하기 위한 메커니즘을 제공한다. 버스 서브시스템이 개략적으로 단일 버스로서 도시되어 있지만, 버스 서브시스템의 대안적 구현에는 다수의 버스를 사용할 수 있다.
- [0216] 컴퓨터 시스템 자체는 개인용 컴퓨터, 휴대용 컴퓨터, 워크스테이션, 컴퓨터 단말기, 네트워크 컴퓨터, 텔레비전, 메인프레임, 서버 팜, 느슨하게 네트워킹된 컴퓨터들의 광범위하게 분포된 세트, 또는 임의의 다른 데이터 프로세싱 시스템 또는 사용자 디바이스를 포함한 다양한 유형이 될 수 있다. 끊임없이 변화하는 컴퓨터와 네트워크의 특성으로 인해, 도 11에 도시된 컴퓨터 시스템의 설명은 개시된 기술을 예시하기 위한 특정 예로서만 의도된다. 도 11에 도시된 컴퓨터 시스템보다 더 많은 또는 더 적은 컴포넌트들을 갖는 컴퓨터 시스템의 많은 다른 구성이 가능하다.
- [0217] 컴퓨터 시스템(1100)은 GPU들 또는 FPGA들(1178)을 포함한다. 그것은 또한, Google Cloud Platform, Xilinx, 및 Cirrascale과 같은 기계 학습 클라우드 플랫폼들에 의해 호스팅되는 기계 학습 프로세서들을 포함할 수 있다. 딥 러닝 프로세서의 예는 Google의 Tensor Processing Unit(TPU), GX4 Rackmount 시리즈, GX8 Rackmount 시리즈, NVIDIA DGX-1, Microsoft의 Stratix V FPGA, Graphcore의 IPU(Intelligent Processor Unit), Qualcomm의 Zeroth 플랫폼과 Snapdragon 프로세서, NVIDIA의 Volta, NVIDIA의 DRIVE PX, NVIDIA의 JETSON TX1/TX2 MODULE, Intel의 Nirvana, Movidius VPU, Fujitsu DPI, ARM의 DynamicIQ, IBM TrueNorth와 같은 랙마운트 솔루션 등을 포함한다.

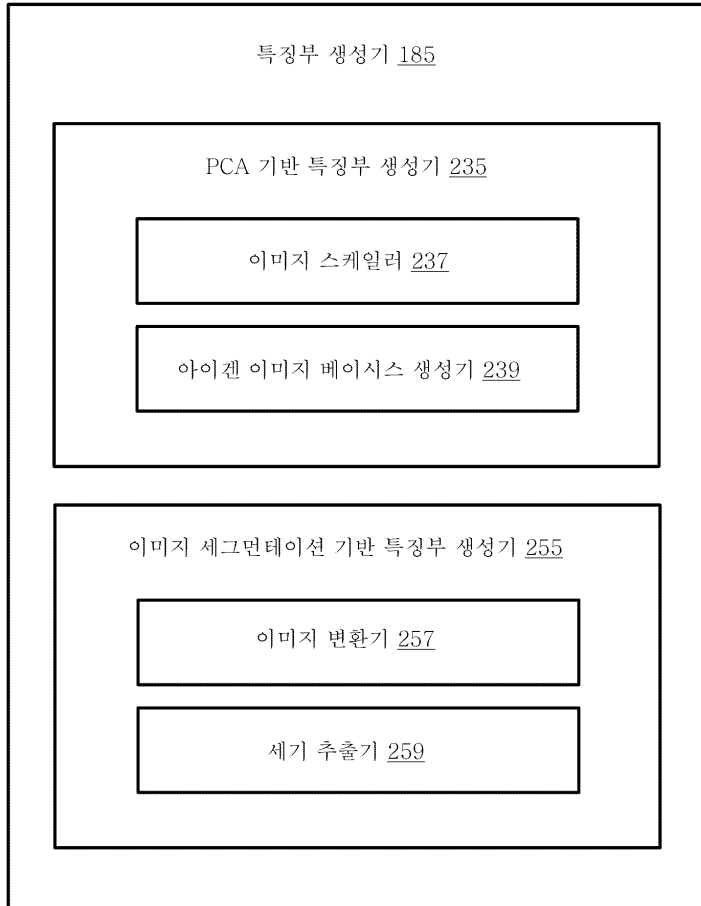
도면

도면1



도면2

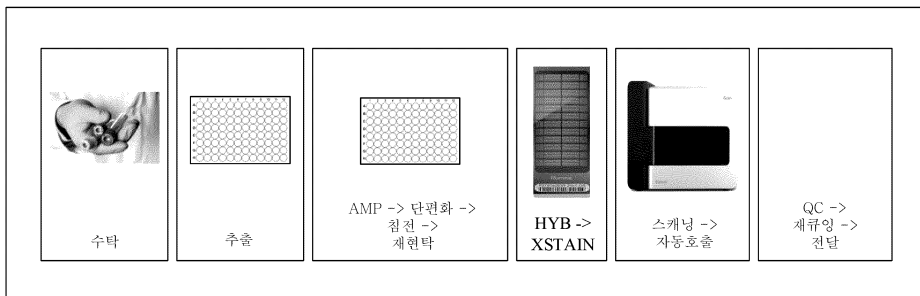
200



도면3

300

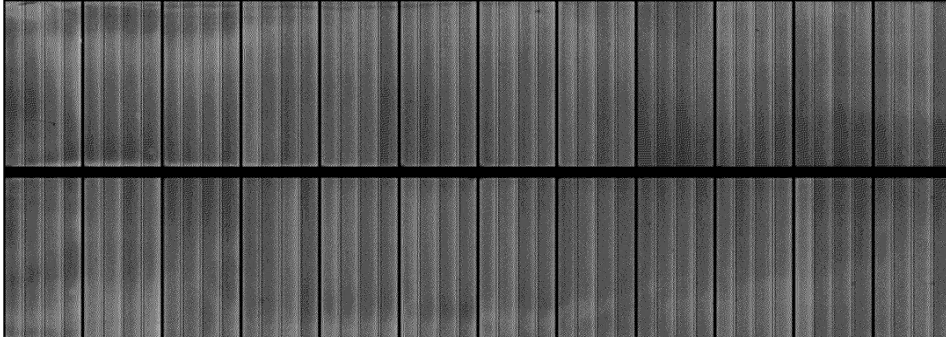
유전형분석 프로세스 단계들



도면4

400




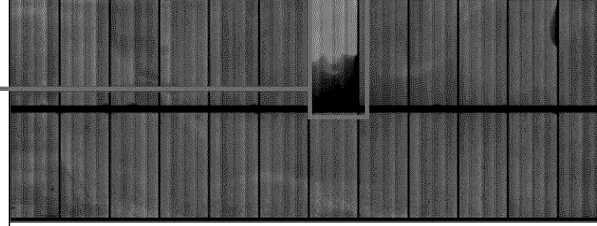
양호한 프로세스 사이클 이미지들 - 성공적인 유전형분석

99.912	99.917	98.91	99.902	99.893	99.913	99.915	99.884	98.833	99.875	99.633	99.692
R01C02	R02C02	R03C02	R04C02	R05C02	R06C02	R07C02	R08C02	R09C02	R10C02	R11C02	R12C02
											
R01C01	R02C01	R03C01	R04C01	R05C01	R06C01	R07C01	R08C01	R09C01	R10C01	R11C01	R12C01
99.9	99.914	199.767	99.862	99.913	99.922	99.884	99.621	99.79	9.843	99.814	99.797

도면5a

510

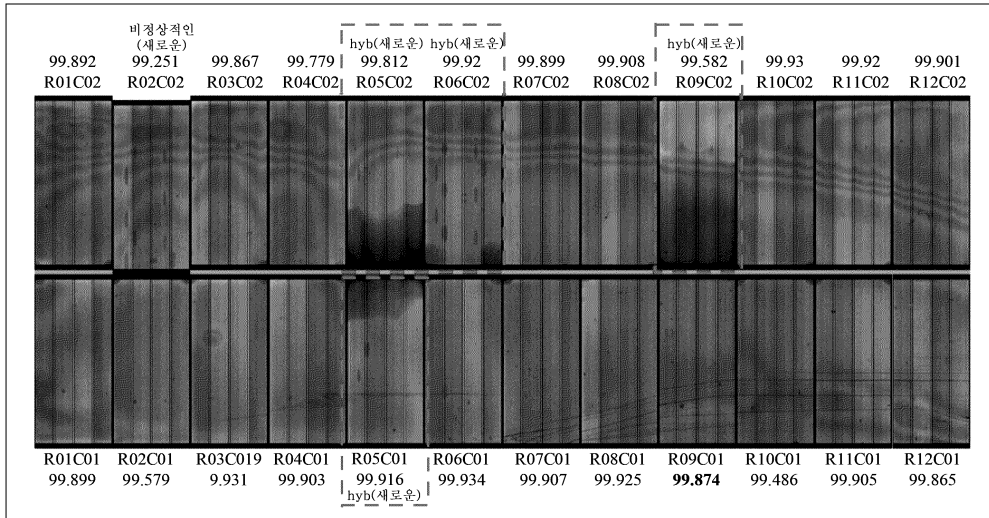
불량한 프로세스 사이클 이미지 - 혼성화(Hyb) 실패(1/2)

														
AMP -> 단편화 -> 침전 -> 재현탁	HYB => XSTAIN	스캐닝 -> 자동호출												
			99.923	99.945	99.888	99.889	99.759	99.778	97.545	98.945	99.789	99.9	99.187	99.473
			R01C02	R02C02	R03C02	R04C02	R05C02	R06C02	R07C02	R08C02	R09C02	R10C02	R11C02	R12C02
														
			R01C01	R02C01	R03C01	R04C01	R05C01	R06C01	R07C01	R08C01	R09C01	R10C01	R11C01	R12C01
			99.737	99.505	99.954	99.949	99.871	99.915	99.897	99.541	99.939	99.928	99.799	99.739

도면5b

515

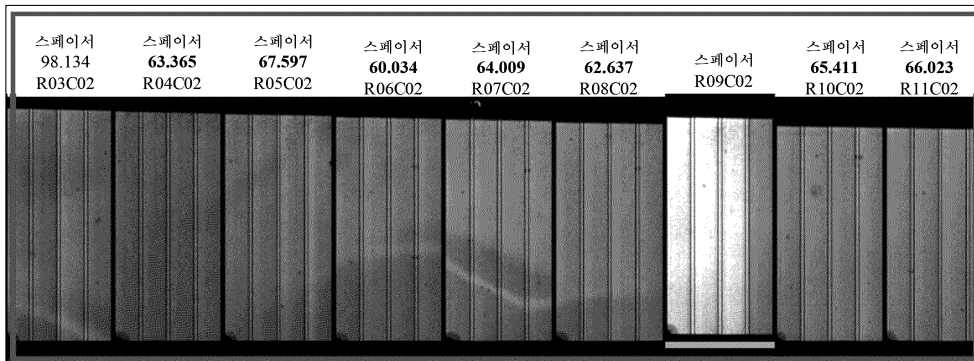
불량한 프로세스 사이클 이미지 - 혼성화(Hyb) 실패(2/2)



도면5c

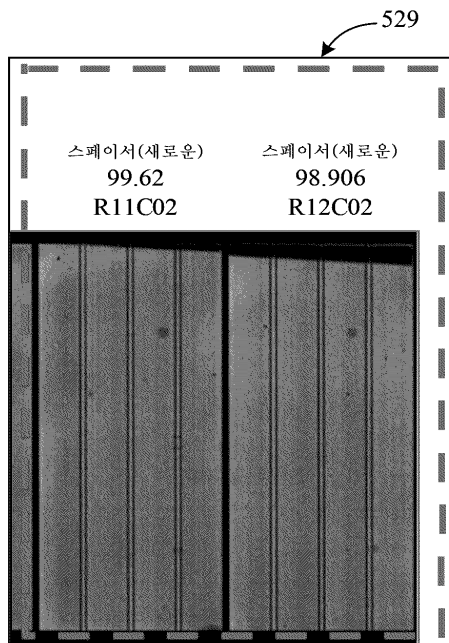
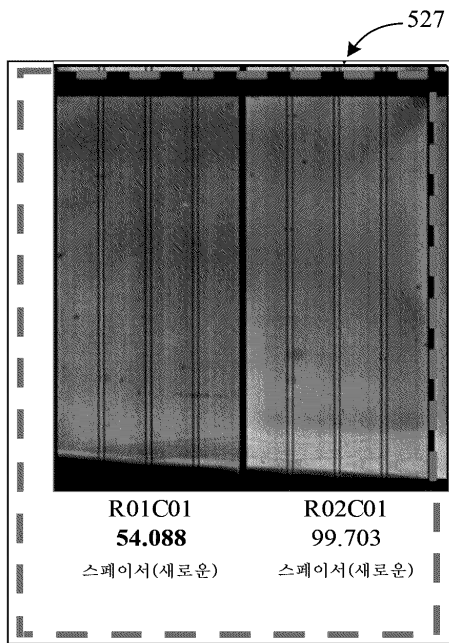
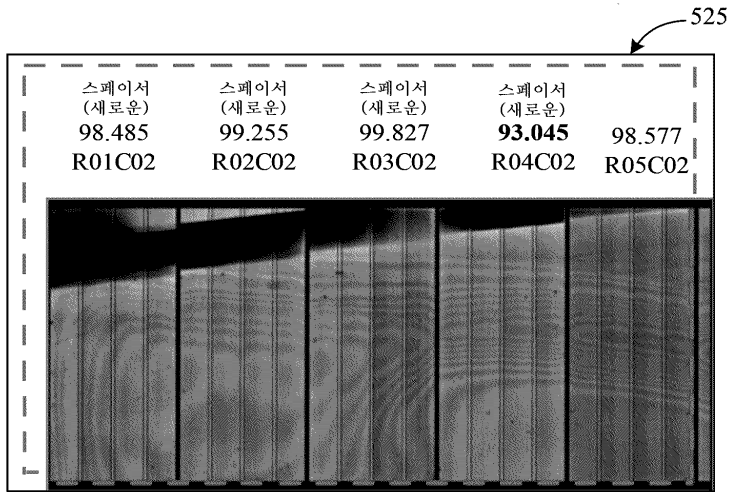
520

불량한 프로세스 사이클 이미지들 - 스페이서 시프트 실패(1/2)



도면5d

불량한 프로세스 사이클 이미지들 - 스페이서 시프트 실패(2/2)



도면5e

530

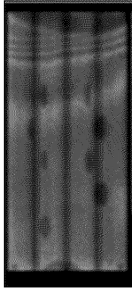
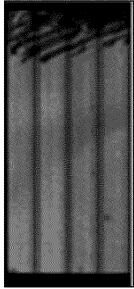
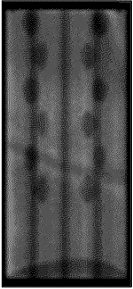
불량한 프로세스 사이클 이미지들 - 오프셋 실패

오프셋(새로운) 96.996	오프셋(새로운) 96.969	오프셋(새로운) 96.944	오프셋(새로운) 96.962	오프셋(새로운) 97.077	오프셋(새로운) 97.277	오프셋(새로운) 97.304	오프셋(새로운) 97.531	오프셋(새로운) 97.522	오프셋(새로운) 97.492	오프셋(새로운) 97.625	오프셋(새로운) 97.251
R01C02	R02C02	R03C02	R04C02	R05C02	R06C02	R07C02	R08C02	R09C02	R10C02	R11C02	R12C02
R01C01	R02C01	R03C01	R04C01	R05C01	R06C01	R07C01	R08C01	R09C01	R10C01	R11C01	R12C01
오프셋(새로운) 97.246	오프셋(새로운) 96.976	오프셋(새로운) 97.177	오프셋(새로운) 97.311	오프셋(새로운) 97.503	오프셋(새로운) 97.652	오프셋(새로운) 97.792	오프셋(새로운) 97.905	오프셋(새로운) 97.876	오프셋(새로운) 97.044	오프셋(새로운) 98.005	오프셋(새로운) 97.906

도면5f

535

불량한 프로세스 사이클 이미지들 - 표면 마모 실패

표면 마모 (새로운) 99.647	표면 마모 (새로운) 98.714	표면 마모 (새로운) 99.839	Ab 78%	Ab 76%	Ab 78%
R08C02	R09C02	R10C02			

도면5g

540

542 불량한 프로세스 사이클 이미지들 - 시약 흐름 실패(1/2)

시약 흐름	시약 흐름	시약 흐름	시약 흐름	시약 흐름							
83.005	83.319	81.665	84.123	97.686	99.278	99.814	99.893	99.93	99.924	99.908	99.83
R01C02	R02C02	R03C02	R04C02	R05C02	R06C02	R07C02	R08C02	R09C02	R10C02	R11C02	R12C02
R01C01	R02C01	R03C01	R04C01	R05C01	R06C01	R07C01	R08C01	R09C01	R10C01	R11C01	R12C01
84.504	82.063	85.746	93.387	98.046	99.295	99.873	99.886	99.916	99.911	99.882	99.88
시약 흐름	시약 흐름	시약 흐름	시약 흐름	시약 흐름							

도면5h

545

불량한 프로세스 사이클 이미지들 - 시약 흐름 실패(2/2)

시약 흐름	시약 흐름	시약 흐름	시약 흐름	시약 흐름	시약 흐름						
93.81	92.745	91.066	94.844	97.972	99.216	99.837	99.879	99.95	99.944	99.943	99.94
R01C02	R02C02	R03C02	R04C02	R05C02	R06C02	R07C02	R08C02	R09C02	R10C02	R11C02	R12C02
R01C01	R02C01	R03C01	R04C01	R05C01	R06C01	R07C01	R08C01	R09C01	R10C01	R11C01	R12C01
95.56	92.513	94.56	96.486	99.311	99.719	99.906	99.898	99.947	99.92	99.79	99.939
시약 흐름	시약 흐름	시약 흐름	시약 흐름	시약 흐름							

도면5i

550

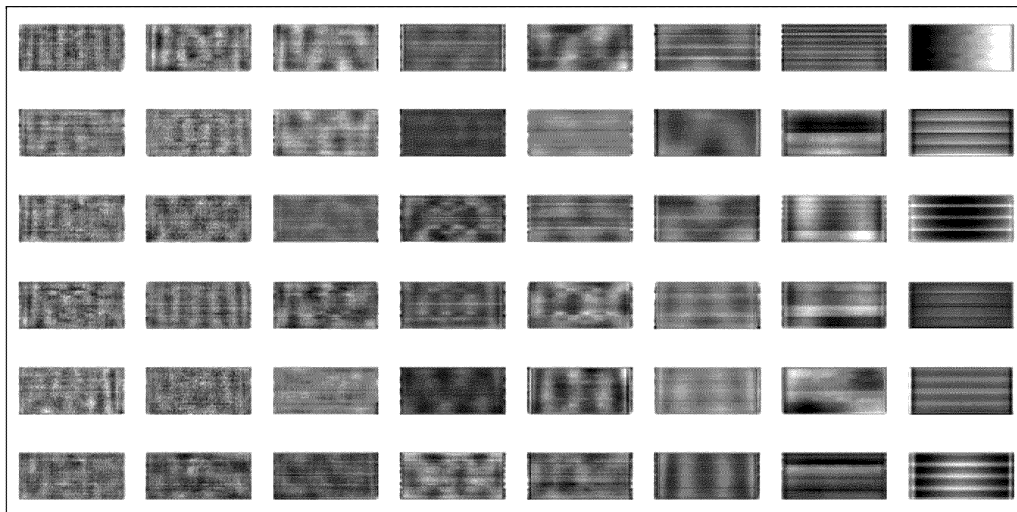
불량한 프로세스 사이클 이미지들 - 비정상적인 / 미지의

스페이스 (새로운)	비정상적인 (새로운)	비정상적인 (새로운)	비정상적인 (새로운)				비정상적인 (새로운)				
32.96	61.393	56.765	70.97	58.146	73.054	81.537	54.686	30.02	82.235	81.37	94.497
R01C02	R02C02	R03C02	R04C02	R05C02	R06C02	R07C02	R08C02	R09C02	R10C02	R11C02	R12C02
R01C01	R02C01	R03C01	R04C01	R05C01	R06C01	R07C01	R08C01	R09C01	R10C01	R11C01	R12C01
99.68	45.306	98.571	37.199	96.85	86.335	97.217	61.317	98.885	31.979	59.851	98.194
	비정상적인 (새로운)		비정상적인 (새로운)				비정상적인 (새로운)		비정상적인 (새로운)	비정상적인 (새로운)	

도면6a

610

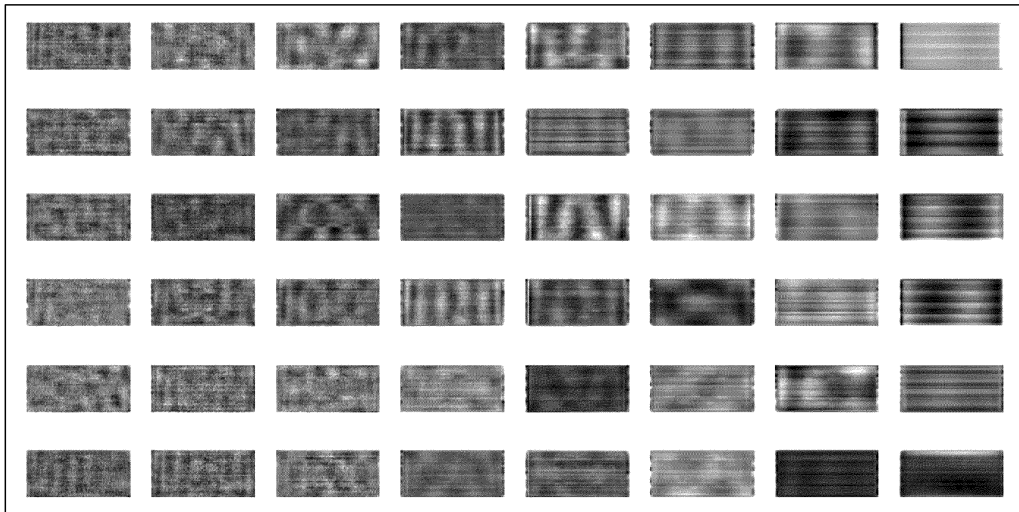
주성분 분석 - 96개의 아이겐 이미지들의 예



도면6b

611

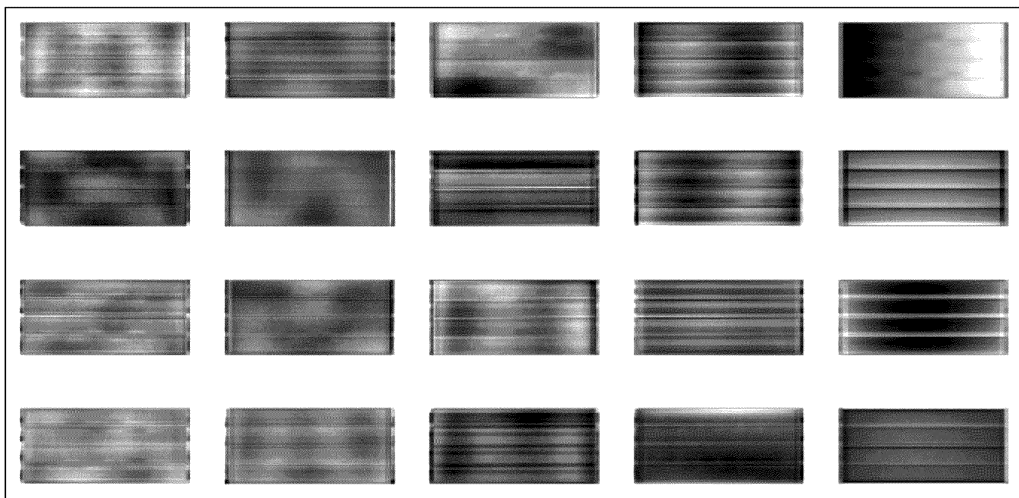
주성분 분석 - 96개의 아이겐 이미지의 예



도면6c

620

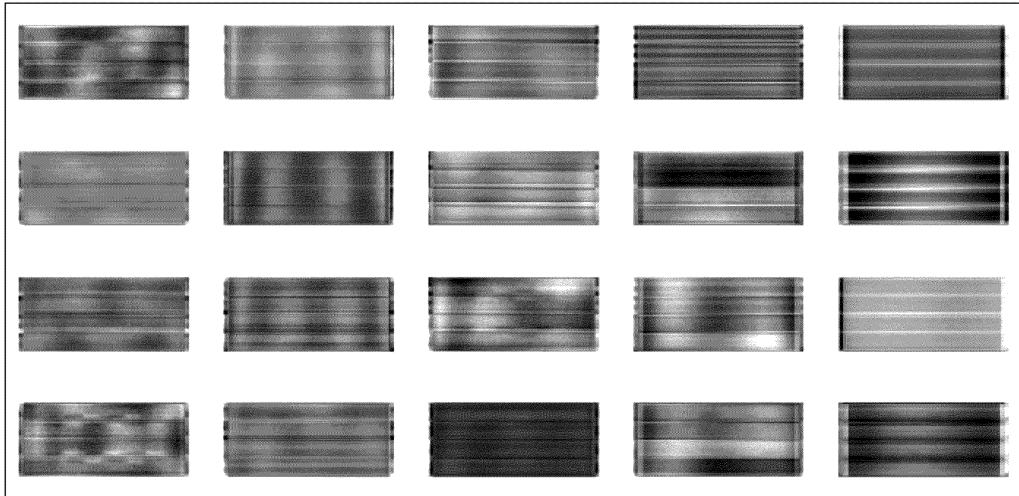
주성분 분석 - 상단 40개의 아이겐 이미지들



도면6d

주성분 분석 - 상단 40개의 아이겐 이미지들

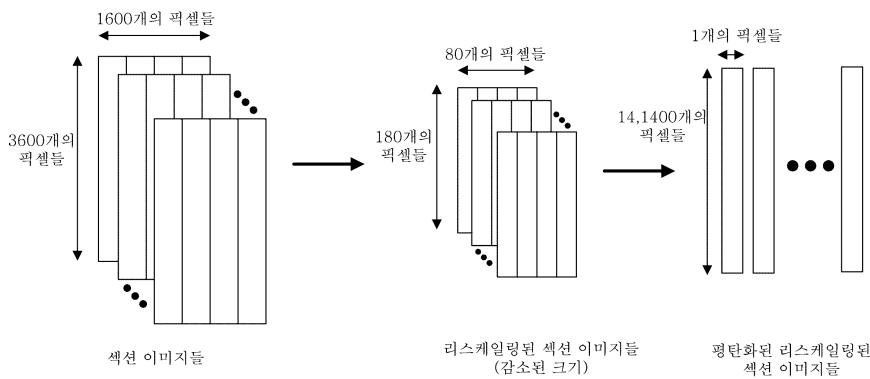
621



도면7a

710

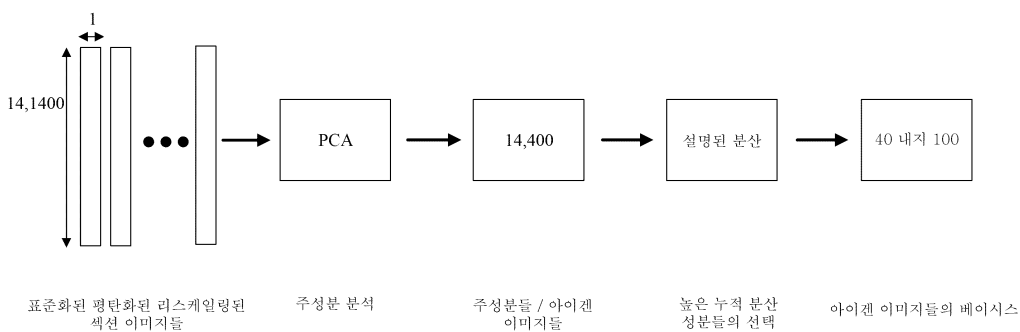
차원수 감소 및 아이겐 이미지들의 베이스의 생성(1/2)



도면7b

740

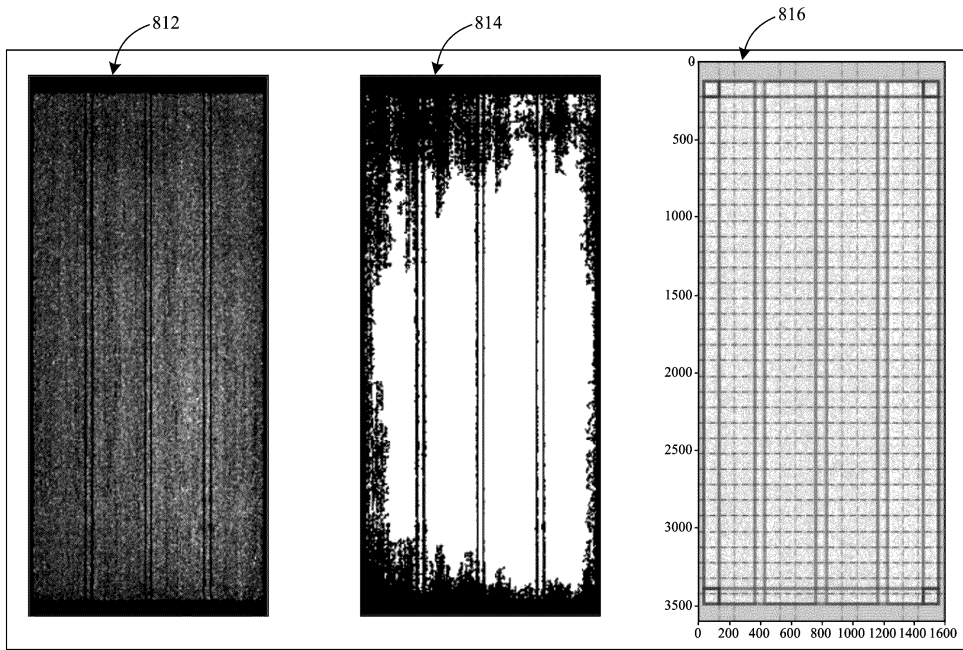
차원수 감소 및 아이겐 이미지들의 베이스의 생성(2/2)



도면8a

프로세스 사이클 이미지들의 영역들의 세기들을 사용한 특징부 생성

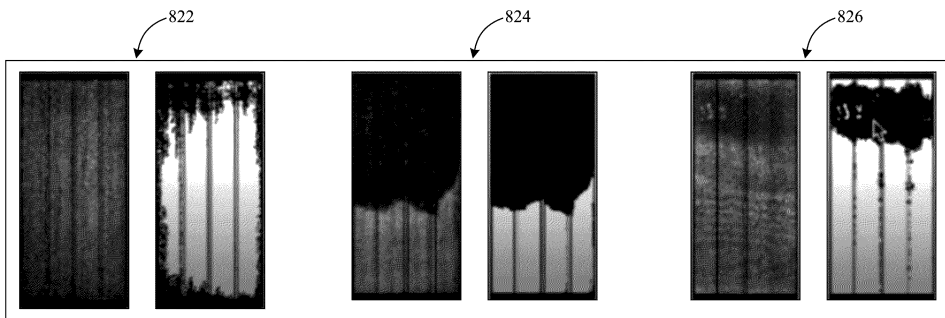
810



도면8b

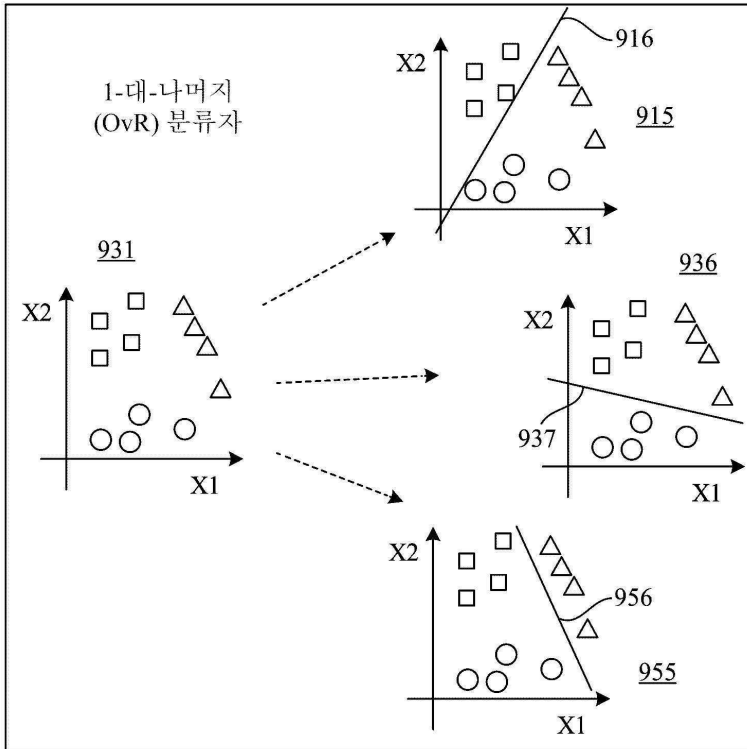
프로세스 사이클 이미지들의 영역들의 세기들을 사용한 특징부 생성

820



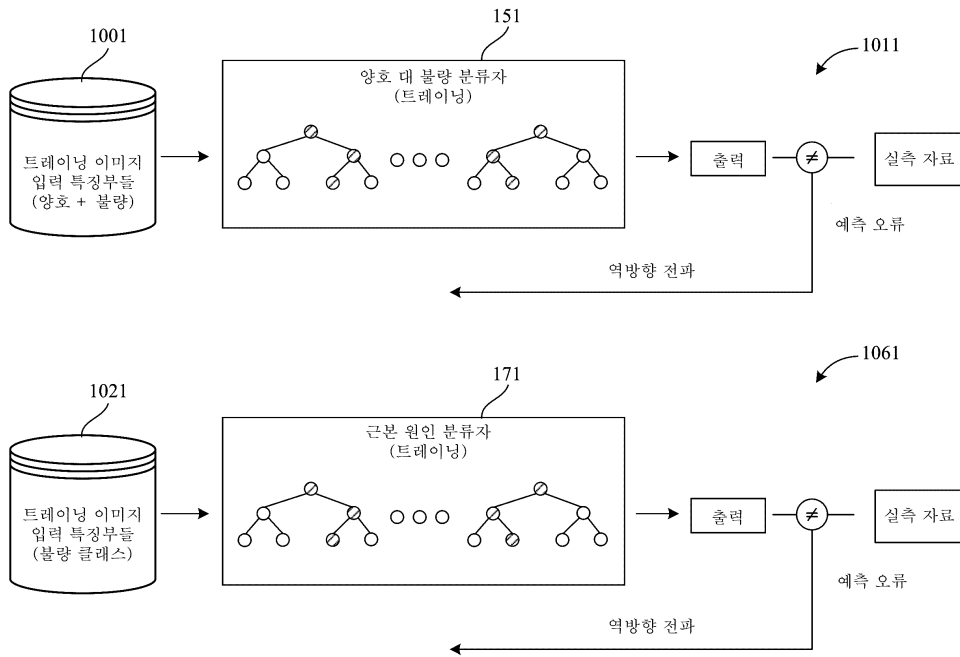
도면9

900



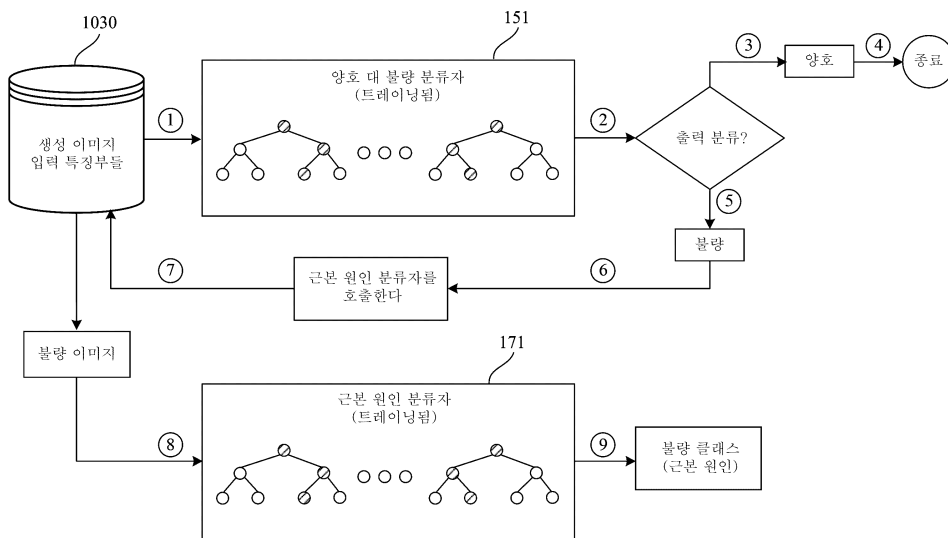
(종래 기술)

도면10a



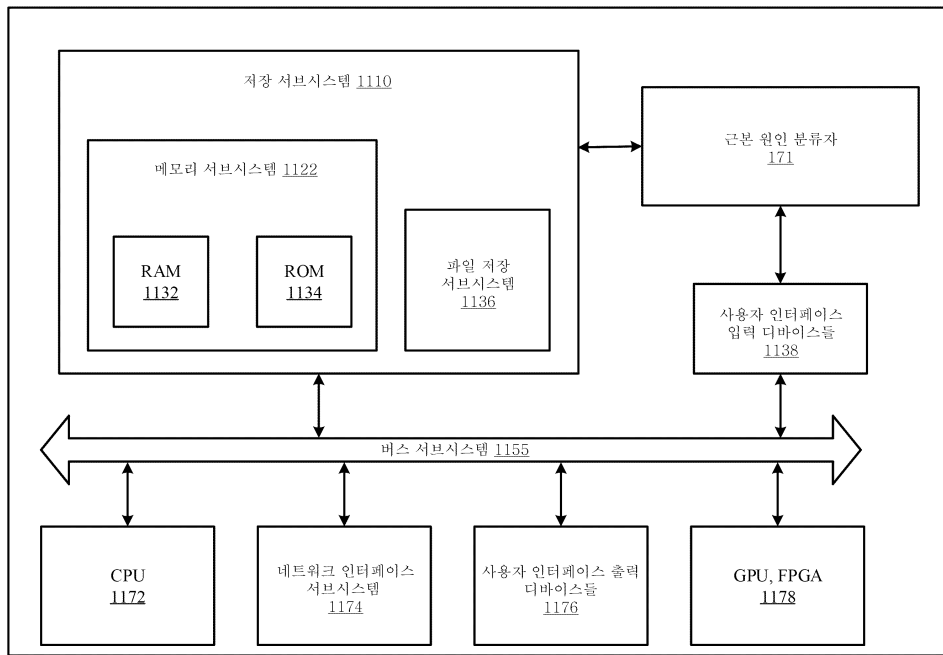
1000

도면10b



1080

도면11



1100