



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2024년09월04일  
(11) 등록번호 10-2702448  
(24) 등록일자 2024년08월30일

(51) 국제특허분류(Int. Cl.)  
G06F 18/00 (2023.01) A01K 1/03 (2006.01)  
A01K 29/00 (2006.01) G06V 10/24 (2022.01)  
G06V 10/26 (2022.01) G06V 10/40 (2022.01)  
(52) CPC특허분류  
G06V 20/00 (2023.08)  
A01K 1/031 (2013.01)  
(21) 출원번호 10-2020-7006943  
(22) 출원일자(국제) 2018년08월07일  
심사청구일자 2021년07월20일  
(85) 번역문제출일자 2020년03월09일  
(65) 공개번호 10-2020-0038509  
(43) 공개일자 2020년04월13일  
(86) 국제출원번호 PCT/US2018/045676  
(87) 국제공개번호 WO 2019/032622  
국제공개일자 2019년02월14일  
(30) 우선권주장  
62/542,180 2017년08월07일 미국(US)  
62/661,610 2018년04월23일 미국(US)  
(56) 선행기술조사문헌  
US20160150758 A1\*  
(뒷면에 계속)

(73) 특허권자  
더 잭슨 래브라토리  
미국 메인(우편번호 04609) 바 하버 메인 스트리트 600  
(72) 발명자  
쿠마르, 비벅  
미국 메인 04609 바 하버 메인 스트리트 600  
고이터, 브라이언, 큐.  
미국 메인 04609 바 하버 메인 스트리트 600  
(뒷면에 계속)  
(74) 대리인  
인비전 특허법인

전체 청구항 수 : 총 16 항

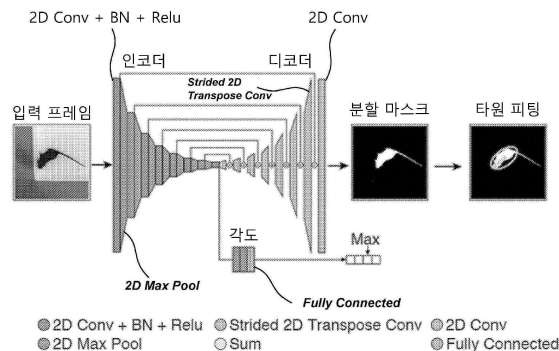
심사관 : 황승희

(54) 발명의 명칭 장기적이고 지속적인 동물 행동 모니터링

(57) 요약

작은 설치류와 같은, 동물들의 행동을 지속적으로 모니터링하기 위한 시스템들 및 방법들이 제공된다. 모니터링은 비디오, 오디오, 및 다른 센서 양식을 포함할 수 있다. 일 실시예에서, 시스템은 카메라, 아레나 디자인, 환경 센서, 및 초음파 센서를 포함할 수 있다. 시스템은 쥐 행동 연구에 적합한 장기적인 모니터링 시스템을 고유하게 제공한다. 모니터링 시스템에 의해 획득된 비디오 데이터와 함께 사용하도록 구성된 신경망 기반의 추적기가 또한 제공된다. 다양한 환경 조건 하에서 유전적으로 다양한 쥐에 대한 성능을 결정하기 위해 3 가지의 상이한 신경망 아키텍처가 테스트되었다. 인코더-디코더 분할 신경망은 최소한의 훈련 데이터로 높은 정확도 및 속도를 달성하는 것으로 관찰되었다. 이 범용 신경망 추적기는 트랜스퍼 학습을 통해 다른 실험 패러다임과 다른 동물들까지 쉽게 확장할 수 있으므로, 생체 행동 연구를 위한 강력하고 일반화 가능한 솔루션을 형성한다.

대표도



(52) CPC특허분류

*A01K 29/005* (2013.01)  
*G06F 18/217* (2023.01)  
*G06F 18/2413* (2023.01)  
*G06F 18/28* (2023.01)  
*G06N 3/04* (2023.01)  
*G06V 10/25* (2023.08)  
*G06V 10/26* (2023.08)  
*G06V 10/454* (2023.08)

(72) 발명자

피터슨, 짐

미국 메인 04609 바 하버 메인 스트리트 600

처칠, 게리

미국 메인 04609 바 하버 메인 스트리트 600

(56) 선행기술조사문헌

US20160232440 A1\*

WO2013170129 A1\*

F. Xu and K. Fujimura, 'Human Detection Using Depth and Gray Images,' In Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS' 03), 2023.07.22. (2023.07.22.)

J. Kim and V. Pavlovic, 'A Shape-Based Approach for Salient Object Detection Using Deep Learning,' ECCV 2016, Part IV, LNCS 9908, pp.455-470, 2016.

\*는 심사관에 의하여 인용된 문헌

## 명세서

### 청구범위

#### 청구항 1

삭제

#### 청구항 2

삭제

#### 청구항 3

삭제

#### 청구항 4

삭제

#### 청구항 5

삭제

#### 청구항 6

삭제

#### 청구항 7

삭제

#### 청구항 8

삭제

#### 청구항 9

복잡한 환경에서의 장기간 동물 추적을 위한 시스템으로서,

적어도 하나의 동물의 관찰을 나타내는 비디오 데이터를 유지하는 데이터 저장 장치(data storage device); 및  
상기 데이터 저장 장치로부터 비디오 데이터를 수신하도록 구성된 프로세서(processor)를 포함하되,

상기 프로세서는:

상기 비디오 데이터로부터 입력 비디오 프레임을 결정하고;

신경망 아키텍처를 사용하여 상기 입력 비디오 프레임을 처리하고;

상기 신경망 아키텍처를 사용하여 상기 입력 비디오 프레임의 처리에 기초하여 상기 적어도 하나의 동물의 타원 묘사를 결정하고 - 상기 타원 묘사는 타원 파라미터들에 의해 정의됨 -; 그리고

상기 적어도 하나의 동물에 대한 상기 타원 묘사에 대응하는 상기 타원 파라미터들에 대한 하나 이상의 값들을 포함하는 데이터를 결정하도록 구성되되,

상기 신경망 아키텍처는 인코더-디코더 분할 네트워크(encoder-decoder segmentation network), 비닝된 분류 네트워크(binned classification network) 및 회귀 네트워크(regression network) 중 적어도 하나를 포함하는, 복잡한 환경에서의 장기간 동물 추적 시스템.

#### 청구항 10

제9항에 있어서,

상기 타원 파라미터들은 평면 내의 상기 동물의 위치, 상기 동물의 장축 길이(major axis length)와 단축 길이(minor axis length), 및 상기 동물의 머리가 향하는 각도(angle) - 상기 각도는 상기 장축의 방향에 대하여 정의됨 - 를 나타내는 좌표인, 복잡한 환경에서의 장기간 동물 추적 시스템.

#### 청구항 11

제9항에 있어서,

상기 신경망 아키텍처는

입력 비디오 프레임으로부터 전경(foreground)-배경(background) 분할된 이미지를 예측하고;

픽셀 단위로, 상기 분할된 이미지에 기초하여 동물이 상기 입력 비디오 프레임에 존재하는지 예측하고;

상기 픽셀 단위 예측에 기초하여 분할 마스크를 출력하고; 그리고

상기 타원 파라미터들을 특징짓는 상기 값들을 결정하기 위해 상기 동물이 존재할 것으로 예측되는 상기 분할 마스크의 부분들을 타원에 맞추도록 구성된 인코더-디코더 분할 네트워크(encoder-decoder segmentation network)인, 복잡한 환경에서의 장기간 동물 추적 시스템.

#### 청구항 12

제11항에 있어서,

상기 인코더-디코더 분할 네트워크는:

상기 입력 비디오 프레임을 특징들의 작은 공간 해상도 세트로 추상화하도록 구성된 특징 인코더(feature encoder);

특징들의 상기 세트를 상기 입력 비디오 프레임과 동일한 형상으로 변환하고, 상기 전경-배경 분할된 이미지를 출력하도록 구성된 특징 디코더(feature decoder); 및

상기 동물의 머리가 향하는 각도를 예측하도록 구성된 각도 예측기(angle predictor)를 포함하는, 복잡한 환경에서의 장기간 동물 추적 시스템.

#### 청구항 13

제9항에 있어서,

상기 신경망 아키텍처는 상기 타원 묘사의 각 타원 파라미터에 대해 가장 가능성 높은 값(most probable value)의 열 지도(heat map)를 예측하도록 구성된 비닝된 분류 네트워크(binned classification network)를 포함하는, 복잡한 환경에서의 장기간 동물 추적 시스템.

#### 청구항 14

제13항에 있어서,

상기 비닝된 분류 네트워크는 상기 입력 비디오 프레임을 작은 공간 해상도 세트로 추상화하도록 구성된 특징 인코더 - 상기 추상화는 상기 열 지도를 생성하는 데 사용됨 - 를 포함하는, 복잡한 환경에서의 장기간 동물 추적 시스템.

#### 청구항 15

제9항에 있어서,

상기 신경망 아키텍처는 입력 비디오 프레임으로부터 특징들을 추출하고, 상기 타원 파라미터들 각각을 특징짓는 상기 값들을 직접 예측하도록 구성된 회귀 네트워크(regression network)를 포함하는, 복잡한 환경에서의 장기간 동물 추적 시스템.

#### 청구항 16



제9항에 있어서,

상기 동물은 설치류(rodent)인, 복잡한 환경에서의 장기간 동물 추적 시스템.

#### 청구항 17

프로그램을 기록한 컴퓨터 판독가능한 기록매체로서, 상기 프로그램은 컴퓨터가:

적어도 하나의 동물의 관찰을 나타내는 비디오 데이터를 수신하도록 하고;

상기 비디오 데이터로부터 입력 비디오 프레임을 수신하도록 하고;

신경망 아키텍처를 사용하여 상기 입력 비디오 프레임을 처리하도록 하고;

상기 신경망 아키텍처를 사용하여 상기 입력 비디오 프레임의 처리에 기초하여 상기 적어도 하나의 동물의 타원 묘사를 결정하도록 하고 - 상기 타원 묘사는 타원 파라미터들에 의해 정의됨 -; 그리고

상기 적어도 하나의 동물에 대한 상기 타원 묘사에 대응하는 상기 타원 파라미터들에 대한 하나 이상의 값들을 포함하는 데이터를 결정하도록 구성되되,

상기 신경망 아키텍처는 인코더-디코더 분할 네트워크(encoder-decoder segmentation network), 비닝된 분류 네트워크(binned classification network) 및 회귀 네트워크(regression network) 중 적어도 하나를 포함하는, 컴퓨터 판독가능한 기록매체.

#### 청구항 18

제17항에 있어서,

상기 타원 파라미터들은 평면 내의 상기 동물의 위치, 상기 동물의 장축 길이(major axis length)와 단축 길이(minor axis length), 및 상기 동물의 머리가 향하는 각도(angle) - 상기 각도는 상기 장축의 방향에 대하여 정의됨 - 를 나타내는 좌표인, 컴퓨터 판독가능한 기록매체.

#### 청구항 19

제17항에 있어서,

상기 신경망 아키텍처는

입력 비디오 프레임으로부터 전경(foreground)-배경(background) 분할된 이미지를 예측하고;

픽셀 단위로, 상기 분할된 이미지에 기초하여 동물이 상기 입력 비디오 프레임에 존재하는지 예측하고;

상기 픽셀 단위 예측에 기초하여 분할 마스크를 출력하고; 그리고

상기 타원 파라미터들을 특징짓는 상기 값들을 결정하기 위해 상기 동물이 존재할 것으로 예측되는 상기 분할 마스크의 부분들을 타원에 맞추도록 구성된 인코더-디코더 분할 네트워크(encoder-decoder segmentation network)인, 컴퓨터 판독가능한 기록매체.

#### 청구항 20

제19항에 있어서,

상기 인코더-디코더 분할 네트워크는:

상기 입력 비디오 프레임을 특징들의 작은 공간 해상도 세트로 추상화하도록 구성된 특징 인코더(feature encoder);

특징들의 상기 세트를 상기 입력 비디오 프레임과 동일한 형상으로 변환하고, 상기 전경-배경 분할된 이미지를 출력하도록 구성된 특징 디코더(feature decoder); 및

상기 동물의 머리가 향하는 각도를 예측하도록 구성된 각도 예측기(angle predictor)를 포함하는, 컴퓨터 판독가능한 기록매체.

#### 청구항 21

제17항에 있어서,

상기 신경망 아키텍처는 상기 타원 묘사의 각 타원 파라미터에 대해 가장 가능성 높은 값(most probable value)의 열 지도(heat map)를 예측하도록 구성된 비닝된 분류 네트워크(binned classification network)를 포함하는, 컴퓨터 판독가능한 기록매체.

#### 청구항 22

제21항에 있어서,

상기 비닝된 분류 네트워크는 상기 입력 비디오 프레임을 작은 공간 해상도 세트로 추상화하도록 구성된 특징 인코더 - 상기 추상화는 상기 열 지도를 생성하는 데 사용됨 - 를 포함하는, 컴퓨터 판독가능한 기록매체.

#### 청구항 23

제17항에 있어서,

상기 신경망 아키텍처는 입력 비디오 프레임으로부터 특징들을 추출하고, 상기 타원 파라미터들 각각을 특징짓는 상기 값들을 직접 예측하도록 구성된 회귀 네트워크(regression network)를 포함하는, 컴퓨터 판독가능한 기록매체.

#### 청구항 24

제17항에 있어서,

상기 동물은 설치류(rodent)인, 컴퓨터 판독가능한 기록매체.

#### 청구항 25

삭제

#### 청구항 26

삭제

#### 청구항 27

삭제

#### 청구항 28

삭제

#### 청구항 29

삭제

#### 청구항 30

삭제

#### 청구항 31

삭제

#### 청구항 32

삭제

#### 청구항 33

삭제

청구항 34

삭제

청구항 35

삭제

청구항 36

삭제

청구항 37

삭제

청구항 38

삭제

청구항 39

삭제

청구항 40

삭제

청구항 41

삭제

청구항 42

삭제

청구항 43

삭제

청구항 44

삭제

청구항 45

삭제

청구항 46

삭제

청구항 47

삭제

청구항 48

삭제

청구항 49

삭제

청구항 50

삭제

청구항 51

삭제

청구항 52

삭제

청구항 53

삭제

청구항 54

삭제

청구항 55

삭제

청구항 56

삭제

청구항 57

삭제

청구항 58

삭제

청구항 59

삭제

청구항 60

삭제

청구항 61

삭제

청구항 62

삭제

## 발명의 설명

## 기술 분야

[0001] 관련 출원에 대한 상호 참조

[0002] 본 출원은 2017년 8월 7일에 제출된 미국 가출원 제62/542,180호("Long-Term and Continuous Animal Behavioral Monitoring") 및 2018년 4월 23일에 제출된 미국 가출원 제62/661,610호("Robust Mouse Tracking In Complex Environments Using Neural Networks")의 이익을 주장하며, 이들 각각의 출원은 모두 참조로 포함된다.

## 배경 기술

[0003] 동물 행동은 내부 또는 외부 자극에 반응한 신경계의 출력으로 이해될 수 있다. 동물들을 정확하게 추적하는 능력은 그들의 행동을 분류하는 프로세스의 일부로서 가치가 있을 수 있다. 예를 들어, 행동의 변화는 노화, 정신 장애, 또는 심지어 대사 질환의 특징이며, 생리학, 신경인지, 및 동물들의 정서적 상태에 미치는 영향에 관한 중요한 정보를 드러낼 수 있다.

## 발명의 내용

[0004] 전통적으로, 동물 행동을 평가하기 위한 실험은, 연구자가 동물과 직접 상호작용하는, 침습적으로 수행되어 왔다. 일 예로서, 연구자는, 쥐와 같은, 동물을 가정 환경(예컨대, 우리(cage))으로부터 제거하고 다른 환경(예컨대, 미로(maze) 또는 다른 장치)으로 옮길 수 있다. 그런 다음 연구자는 새로운 환경 근처에 위치하고 과제 수행 시 동물을 관찰하기 위해 동물을 추적할 수 있다. 그러나, 동물들은 새로운 환경에서 또는 시험을 수행하는 실험자에게 다르게 행동할 수 있다는 것이 인식되었다. 이것은 종종 데이터의 혼란으로 이어지고, 재현 불가능하고 오해의 소지가 있는 결과를 초래한다.

[0005] 행동 모니터링 실험 동안 인간의 간섭을 최소화하기 위해 덜 침습적인 모니터링 기술들이 개발되었다. 일 예로서, 동물 행동을 모니터링하는 데 사용하기 위한 비디오 모니터링이 탐구되었다. 그러나, 비디오 모니터링에는 여전히 문제점들(challenges)이 있다. 일 측면에서, 오랜 기간 동안 연속적으로, 그리고 광범위한 환경 조건 하에서, 높은 시간적 및 공간적 해상도로 비디오 데이터를 캡처하는 능력은 여전히 중요한 장애물이다. 며칠, 몇 주, 및/또는 몇 개월과 같이, 오랜 기간 동안 발생하는 동물의 관찰 연구는 획득 및 저장에 많은 비용이 드는 대량의 데이터를 생성할 수 있다. 다른 측면에서, 충분한 품질의 비디오 데이터를 획득 및 저장하는 능력을 가정하더라도, 인간 연구자들이 오랜 기간 관찰 동안 생성된 많은 양의 비디오 장면을 수동으로 검토하고 그러한 오랜 기간 동안 동물을 추적하는 것은 경제적으로 실현 가능하지 않다. 새로운 약물을 스크리닝하거나 게놈 실험을 수행할 때 필요할 수 있는 것처럼, 관찰 중인 동물들의 수가 증가하면 이런 문제가 확대된다.

[0006] 이 문제를 해결하기 위해, 동물 행동의 캡처된 비디오를 분석하기 위한 컴퓨터 기반 기술들이 개발되었다. 그러나, 종래의 컴퓨터 기반 시스템은 복잡하고 동적인 환경에서 다른 동물들을 정확하게 추적할 수 있는 능력이 없다. 일 예로서, 동물 추적을 위한 종래의 컴퓨터 기반 기술은 한 마리의 동물을 배경(예컨대, 우리(cage) 벽 및/또는 바닥, 물통과 같은 우리(cage) 내의 물체)으로부터 또는 다수의 동물을 서로(one another)로부터 정확하게 구별하는 데 실패할 수 있다. 최선의 경우, 관찰 기간 동안 특정 동물이 정확하게 추적되지 않으면, 귀중한 관찰 데이터가 손실될 수 있다. 최악의 경우, 관찰 기간 동안 특정 동물 또는 그 동물의 일부가 잘못 추적되고 다른 것으로 오인될 경우, 획득된 비디오 데이터로부터 분류된 동작에 오류가 발생할 수 있다. 추적을 용이하게 하기 위해 동물의 털 색상 변경과 같은 기술들이 사용되었지만, 동물의 털 색상을 변경하는 것은 그것의 행동을 변경시킬 수 있다. 결과적으로, 복잡하고 동적인 환경에서 또는 유전자 이중 동물에서 수행되는 종래의 비디오 추적 접근법들은 높은 수준의 사용자 개입을 필요로 하며, 위에서 논의된 비디오 관찰의 이점을 무시한다. 따라서, 대규모 및/또는 장기적인 동물 모니터링 실험은 여전히 불가능하다.

[0007] 신경과학 및 행동이 큰 행동 데이터 및 컴퓨터 윤리의 시대로 이동함에 따라, 장기간에 걸쳐 반자연적(semi-natural) 및 동적 환경에서 동물의 행동 분류를 용이하게 하기 위해 동물 추적을 위한 더 나은 기술이 필요하다.

[0008] 따라서, 신경망을 이용하는 공개 필드에서 동물(예컨대, 쥐)의 강력하고 확장 가능한 추적을 제공할 수 있는 시스템 및 방법이 개발되었다. 일 예로서, 높은 시간적 및 공간적 해상도로 동물 움직임의 비디오 데이터의 획득을 용이하게 하기 위한 시스템 및 방법이 제공된다. 이 비디오 데이터는 오랜 기간 동안 연속적으로, 그리고 광범위한 환경 조건 하에서 연속적으로 캡처될 수 있다.

[0009] 획득된 비디오 데이터는 추적을 위한 컨볼루션 신경망 아키텍처에 대한 입력으로서 사용될 수 있다. 신경망은, 새로운 환경 또는 동물이 제시될 때, 훈련될 때, 신경망이 매우 견고하고 사용자가 개입된 튜닝없이 여러 실험 조건 하에서 추적할 수 있도록 훈련될 수 있다. 이러한 실험 조건의 예는, 상이한 우리 환경뿐만 아니라, 다양한 털 색상, 체형, 및 행동에 관계 없이 상이한 종류의 쥐를 포함할 수 있다. 따라서, 본 개시의 실시예들은 이 질적인 조건 하에서 오랜 시간에 걸쳐 다수의 동물의 행동 모니터링을 용이하게 하기 위해 최소한으로 침습적인 동물 추적을 용이하게 할 수 있다.

[0010] 특정 실시예들에서, 개시된 비디오 관찰 및 동물 추적 기술은 조합하여 사용될 수 있다. 그러나, 이들 기술 각

각은 단독으로 또는 서로 간의 또는 다른 기술과의 임의의 조합으로 사용될 수 있음을 이해할 수 있다.

- [0011] 일 실시예에서, 동물 추적 방법이 제공된다. 상기 방법은 프로세서에 의해, 동물의 관찰을 나타내는 비디오 데이터를 수신하는 단계, 및 상기 프로세서에 의해, 신경망 아키텍처를 실행하는 단계를 포함할 수 있다. 상기 신경망 아키텍처는 상기 비디오 데이터로부터 추출된 입력 비디오 프레임을 수신하고, 상기 입력 비디오 프레임에 기초하여 적어도 하나의 동물의 타원 묘사를 생성하고 - 상기 타원 묘사는 미리 결정된 타원 파라미터들에 의해 결정됨 -, 그리고 상기 적어도 하나의 동물에 대한 상기 미리 결정된 타원 파라미터들을 특징짓는 값들을 포함하는 데이터를 제공하도록 구성될 수 있다.
- [0012] 상기 방법의 다른 실시예에서, 상기 타원 파라미터들은 평면 내의 상기 동물의 위치, 상기 동물의 장축 길이와 단축 길이, 및 상기 동물의 머리가 향하는 각도 - 상기 각도는 상기 장축의 방향에 대하여 정의됨 - 를 나타내는 좌표일 수 있다.
- [0013] 상기 방법의 다른 실시예에서, 상기 신경망 아키텍처는 인코더-디코더 분할 네트워크(encoder-decoder segmentation network)일 수 있다. 상기 인코더-디코더 분할 네트워크는 입력 비디오 프레임으로부터 전경(foreground)-배경(background) 분할된 이미지를 예측하고, 픽셀 단위로, 상기 분할된 이미지에 기초하여 동물이 상기 입력 비디오 프레임에 존재하는지 예측하고, 상기 픽셀 단위 예측에 기초하여 분할 마스크를 출력하고, 그리고 상기 미리 결정된 타원 파라미터들을 특징짓는 상기 값들을 결정하기 위해 상기 동물이 존재할 것으로 예측되는 상기 분할 마스크의 부분들을 타원에 맞추도록 구성될 수 있다.
- [0014] 상기 방법의 다른 실시예에서, 상기 인코더-디코더 분할 네트워크는 특징 인코더(feature encoder), 특징 디코더(feature decoder), 및 각도 예측기(angle predictor)를 포함할 수 있다. 상기 특징 인코더는 상기 입력 비디오 프레임을 특징들의 작은 공간 해상도 세트로 추상화하도록 구성될 수 있다. 상기 특징 디코더는 특징들의 상기 세트를 상기 입력 비디오 프레임과 동일한 형상으로 변환하고, 상기 전경-배경 분할된 이미지를 출력하도록 구성될 수 있다. 상기 각도 예측기는 상기 동물의 머리가 향하는 각도를 예측하도록 구성될 수 있다.
- [0015] 상기 방법의 다른 실시예에서, 상기 신경망 아키텍처는 상기 타원 묘사의 각 타원 파라미터에 대해 가장 가능성 높은 값(most probable value)의 열 지도(heat map)를 예측하도록 구성된 비닝된 분류 네트워크(binned classification network)를 포함할 수 있다.
- [0016] 상기 방법의 다른 실시예에서, 상기 비닝된 분류 네트워크는 상기 입력 비디오 프레임을 작은 공간 해상도 세트로 추상화하도록 구성된 특징 인코더 - 상기 추상화는 상기 열 지도를 생성하는 데 사용됨 - 를 포함할 수 있다.
- [0017] 상기 방법의 다른 실시예에서, 상기 신경망 아키텍처는 입력 비디오 프레임으로부터 특징들을 추출하고, 상기 타원 파라미터들 각각을 특징짓는 상기 값들을 직접 예측하도록 구성된 회귀 네트워크(regression network)를 포함할 수 있다.
- [0018] 상기 방법의 다른 실시예에서, 상기 동물은 설치류(rodent)일 수 있다.
- [0019] 일 실시예에서, 동물 추적 시스템이 제공된다. 상기 시스템은 동물의 관찰을 나타내는 비디오 데이터를 유지하는 데이터 저장 장치(data storage device)를 포함할 수 있다. 상기 시스템은 또한 상기 데이터 저장 장치로부터 비디오 데이터를 수신하고, 신경망 아키텍처를 구현하도록 구성된 프로세서(processor)를 포함할 수 있다. 상기 신경망 아키텍처는 상기 비디오 데이터로부터 추출된 입력 비디오 프레임을 수신하고, 상기 비디오 프레임에 기초하여 적어도 하나의 동물의 타원 묘사를 생성하고 - 상기 타원 묘사는 미리 결정된 타원 파라미터들에 의해 정의됨 -, 그리고 상기 적어도 하나의 동물에 대한 상기 미리 결정된 타원 파라미터들을 특징짓는 값들을 포함하는 데이터를 제공하도록 구성될 수 있다.
- [0020] 상기 시스템의 다른 실시예에서, 상기 타원 파라미터들은 평면 내의 상기 동물의 위치, 상기 동물의 장축 길이와 단축 길이, 및 상기 동물의 머리가 향하는 각도 - 상기 각도는 상기 장축의 방향에 대하여 정의됨 - 를 나타내는 좌표일 수 있다.
- [0021] 상기 시스템의 다른 실시예에서, 상기 신경망 아키텍처는 인코더-디코더 분할 네트워크(encoder-decoder segmentation network)일 수 있다. 상기 인코더-디코더 분할 네트워크는 입력 비디오 프레임으로부터 전경(foreground)-배경(background) 분할된 이미지를 예측하고, 픽셀 단위로, 상기 분할된 이미지에 기초하여 동물이 상기 입력 비디오 프레임에 존재하는지 예측하고, 상기 픽셀 단위 예측에 기초하여 분할 마스크를 출력하고, 그리고 상기 미리 결정된 타원 파라미터들을 특징짓는 상기 값들을 결정하기 위해 상기 동물이 존재할 것으로

예측되는 상기 분할 마스크의 부분들을 타원에 맞추도록 구성될 수 있다.

- [0022] 상기 시스템의 다른 실시예에서, 상기 인코더-디코더 분할 네트워크는 특징 인코더(feature encoder), 특징 디코더(feature decoder), 및 각도 예측기(angle predictor)를 포함할 수 있다. 상기 특징 인코더는 상기 입력 비디오 프레임을 특징들의 작은 공간 해상도 세트로 추상화하도록 구성될 수 있다. 상기 특징 디코더는 특징들의 상기 세트를 상기 입력 비디오 프레임과 동일한 형상으로 변환하고, 상기 전경-배경 분할된 이미지를 출력하도록 구성될 수 있다. 상기 각도 예측기는 상기 동물의 머리가 향하는 각도를 예측하도록 구성될 수 있다.
- [0023] 상기 시스템의 다른 실시예에서, 상기 신경망 아키텍처는 비닝된 분류 네트워크(binned classification network)를 포함할 수 있다. 상기 비닝된 분류 네트워크는 상기 타원 묘사의 각 타원 파라미터에 대해 가장 가능성 높은 값(most probable value)의 열 지도(heat map)를 예측하도록 구성될 수 있다.
- [0024] 상기 시스템의 다른 실시예에서, 상기 비닝된 분류 네트워크는 상기 입력 비디오 프레임을 작은 공간 해상도 세트로 추상화하도록 구성된 특징 인코더 - 상기 추상화는 상기 열 지도를 생성하는 데 사용됨 - 를 포함할 수 있다.
- [0025] 상기 시스템의 다른 실시예에서, 상기 신경망 아키텍처는 입력 비디오 프레임으로부터 특징들을 추출하고, 상기 타원 파라미터들 각각을 특징짓는 상기 값들을 직접 예측하도록 구성된 회귀 네트워크(regression network)를 포함할 수 있다.
- [0026] 상기 시스템의 다른 실시예에서, 상기 동물은 설치류(rodent)일 수 있다.
- [0027] 일 실시예에서, 명령들을 저장하는 비밀시적 컴퓨터 프로그램 제품이 제공된다. 상기 명령들은, 적어도 하나의 컴퓨팅 시스템의 적어도 하나의 데이터 프로세서에 의해 실행될 때, 동물의 관찰을 나타내는 비디오 데이터를 수신하는 단계, 및 신경망 아키텍처를 실행하는 단계를 포함하는 방법을 구현할 수 있다. 상기 신경망 아키텍처는 상기 비디오 데이터로부터 추출된 입력 비디오 프레임을 수신하고, 상기 입력 비디오 프레임에 기초하여 적어도 하나의 동물의 타원 묘사(ellipse description)를 생성하고 - 상기 타원 묘사는 미리 결정된 타원 파라미터들에 의해 정의됨 -, 그리고 상기 적어도 하나의 동물에 대한 상기 미리 결정된 타원 파라미터들을 특징짓는 값들을 포함하는 데이터를 제공하도록 구성될 수 있다.
- [0028] 다른 실시예에서, 상기 타원 파라미터들은 평면 내의 상기 동물의 위치, 상기 동물의 장축 길이와 단축 길이, 및 상기 동물의 머리가 향하는 각도 - 상기 각도는 상기 장축의 방향에 대하여 정의됨 - 를 나타내는 좌표일 수 있다.
- [0029] 다른 실시예에서, 상기 신경망 아키텍처는 인코더-디코더 분할 네트워크(encoder-decoder segmentation network)일 수 있다. 상기 인코더-디코더 분할 네트워크는 입력 비디오 프레임으로부터 전경(foreground)-배경(background) 분할된 이미지를 예측하고, 픽셀 단위로, 상기 분할된 이미지에 기초하여 동물이 상기 입력 비디오 프레임에 존재하는지 예측하고, 상기 픽셀 단위 예측에 기초하여 분할 마스크를 출력하고, 그리고 상기 미리 결정된 타원 파라미터들을 특징짓는 상기 값들을 결정하기 위해 상기 동물이 존재할 것으로 예측되는 상기 분할 마스크의 부분들을 타원에 맞추도록 구성될 수 있다.
- [0030] 다른 실시예에서, 상기 인코더-디코더 분할 네트워크는 특징 인코더(feature encoder), 특징 디코더(feature decoder), 및 각도 예측기(angle predictor)를 포함할 수 있다. 상기 특징 인코더는 상기 입력 비디오 프레임을 특징들의 작은 공간 해상도 세트로 추상화하도록 구성될 수 있다. 상기 특징 디코더는 특징들의 상기 세트를 상기 입력 비디오 프레임과 동일한 형상으로 변환하고, 상기 전경-배경 분할된 이미지를 출력하도록 구성될 수 있다. 상기 각도 예측기는 상기 동물의 머리가 향하는 각도를 예측하도록 구성될 수 있다.
- [0031] 다른 실시예에서, 상기 신경망 아키텍처는 상기 타원 묘사의 각 타원 파라미터에 대해 가장 가능성 높은 값(most probable value)의 열 지도(heat map)를 예측하도록 구성된 비닝된 분류 네트워크(binned classification network)를 포함할 수 있다.
- [0032] 다른 실시예에서, 상기 비닝된 분류 네트워크는 상기 입력 비디오 프레임을 작은 공간 해상도 세트로 추상화하도록 구성된 특징 인코더 - 상기 추상화는 상기 열 지도를 생성하는 데 사용됨 - 를 포함할 수 있다.
- [0033] 다른 실시예에서, 상기 신경망 아키텍처는 입력 비디오 프레임으로부터 특징들을 추출하고, 상기 타원 파라미터들 각각을 특징짓는 상기 값들을 직접 예측하도록 구성된 회귀 네트워크(regression network)를 포함할 수 있다.



- [0034] 다른 실시예에서, 상기 동물은 설치류(rodent)일 수 있다.
- [0035] 일 실시예에서, 시스템이 제공되고, 그것은 아레나(arena) 및 획득 시스템(acquisition system)을 포함할 수 있다. 상기 아레나는 프레임(frame) 및 상기 프레임에 장착된 인클로저(enclosure)를 포함할 수 있다. 상기 인클로저는 동물을 수용하도록 설정될 수 있고, 그것은 상기 인클로저 내부로의 접근을 허용하도록 구성된 도어(door)를 포함할 수 있다. 상기 획득 시스템은 카메라(camera), 광원들의 적어도 두 개의 세트들, 컨트롤러(controller), 및 데이터 저장 장치(data storage device)를 포함할 수 있다. 광원들의 각각의 세트는 서로 파장이 다르고 상기 인클로저에 입사하는 광을 방출하도록 구성될 수 있다. 상기 카메라는 광원들의 상기 세트들 중 적어도 하나에 의해 조명될 때 상기 인클로저의 적어도 일부의 비디오 데이터를 획득하도록 구성될 수 있다. 상기 컨트롤러는 상기 카메라 및 광원들의 상기 세트들과 전기적으로 통신할 수 있다. 상기 컨트롤러는 상기 카메라에 의한 비디오 데이터의 획득 및 광원들의 상기 세트들에 의한 광의 방출을 제어하도록 동작하는 제어 신호를 생성하고, 상기 카메라에 의해 획득된 비디오 데이터를 수신하도록 구성될 수 있다. 상기 데이터 저장 장치는 상기 컨트롤러와 전기적으로 통신할 수 있고, 그것은 상기 컨트롤러로부터 수신한 비디오 데이터를 저장하도록 구성될 수 있다.
- [0036] 상기 시스템의 다른 실시예에서, 상기 인클로저의 적어도 일부는 가시광선에 대해 거의 불투명할 수 있다.
- [0037] 상기 시스템의 다른 실시예에서, 상기 인클로저의 적어도 일부는 가시광선 파장에 대해 거의 불투명한 재료로부터 형성될 수 있다.
- [0038] 상기 시스템의 다른 실시예에서, 상기 인클로저의 적어도 일부는 적외선 파장에 대해 거의 비반사적인(non-reflective) 재료로부터 형성될 수 있다.
- [0039] 상기 시스템의 다른 실시예에서, 상기 인클로저의 적어도 일부는 폴리염화비닐(PVC, polyvinyl chloride) 또는 폴리옥시메틸렌(POM, polyoxymethylene)의 시트(sheet)로부터 형성될 수 있다.
- [0040] 상기 시스템의 다른 실시예에서, 광원들의 제1 세트는 적어도 하나의 가시광선 파장에서 광을 방출하도록 구성된 적어도 하나의 제1 발광체를 포함할 수 있고, 광원들의 제2 세트는 적어도 하나의 적외선 파장에서 광을 방출하도록 구성된 적어도 하나의 제2 발광체를 포함할 수 있다.
- [0041] 상기 시스템의 다른 실시예에서, 상기 적외선 파장은 거의 940 nm일 수 있다.
- [0042] 상기 시스템의 다른 실시예에서, 상기 카메라는 적어도 480x480 픽셀의 해상도로 비디오 데이터를 획득하도록 구성될 수 있다.
- [0043] 상기 시스템의 다른 실시예에서, 상기 카메라는 쥐의 움직임의 주파수보다 큰 프레임 레이트(frame rate)에서 비디오 데이터를 획득하도록 구성될 수 있다.
- [0044] 상기 시스템의 다른 실시예에서, 상기 카메라는 적어도 29 fps(frames per second)의 프레임 레이트로 비디오 데이터를 획득하도록 구성될 수 있다.
- [0045] 상기 시스템의 다른 실시예에서, 상기 카메라는 적어도 8-비트 심도(bit depth)를 갖는 비디오 데이터를 획득하도록 구성될 수 있다.
- [0046] 상기 시스템의 다른 실시예에서, 상기 카메라는 적외선 파장에서 비디오 데이터를 획득하도록 구성될 수 있다.
- [0047] 상기 시스템의 다른 실시예에서, 컨트롤러는 상기 카메라로부터 수신된 비디오 데이터를 압축하도록 구성될 수 있다.
- [0048] 상기 시스템의 다른 실시예에서, 상기 컨트롤러는 분산 기반 배경 감산(variance-based background subtraction)을 사용하는 필터와 함께 MPEG4 코덱을 이용하여 상기 카메라로부터 수신된 비디오 데이터를 압축하도록 구성될 수 있다.
- [0049] 상기 시스템의 다른 실시예에서, 상기 MPEG 코덱 필터는 Q0 HQDN3D일 수 있다.
- [0050] 상기 시스템의 다른 실시예에서, 상기 컨트롤러는 명(light)/암(dark) 순환을 모의하는 스케줄에 따라 상기 인클로저를 조명하도록 상기 제1 광원에 명령하도록 구성될 수 있다.
- [0051] 상기 시스템의 다른 실시예에서, 상기 컨트롤러는 상기 명/암 순환의 명 부분(light portion) 동안 약 50 럭스(lux) 내지 약 800 럭스의 세기를 갖는 가시광선으로 상기 인클로저를 조명하도록 상기 제1 광원에 명령하도록



구성될 수 있다.

- [0052] 상기 시스템의 다른 실시예에서, 상기 컨트롤러는 상기 인클로저의 온도가 적외선 조명에 의해 5℃ 미만으로 상승되도록 적외선으로 상기 인클로저를 조명하도록 상기 제2 광원에 명령하도록 구성될 수 있다.
- [0053] 상기 시스템의 다른 실시예에서, 상기 컨트롤러는 로그 스케일링된 1024 레벨의 광에 따라 상기 인클로저를 조명하도록 상기 제1 광원에 명령하도록 구성될 수 있다.
- [0054] 일 실시예에서, 방법이 제공되고, 그것은 광원들의 적어도 하나의 세트에 의해, 동물을 수용하도록 구성된 인클로저(enclosure)를 조명하는 단계를 포함할 수 있다. 광원들의 각각의 세트는 서로 파장이 다른 광을 방출하도록 구성될 수 있다. 상기 방법은 또한 카메라(camera)에 의해, 광원들의 상기 세트들 중 적어도 하나에 의해 조명되는 상기 인클로저의 적어도 일부의 비디오 데이터를 획득하는 단계를 포함할 수 있다. 상기 방법은 추가적으로 상기 카메라 및 광원들의 상기 세트들과 전기적으로 통신하는 컨트롤러(controller)에 의해, 상기 카메라에 의한 비디오 데이터의 획득 및 광원들의 상기 세트들에 의한 광의 방출을 제어하도록 동작하는 제어 신호를 생성하는 단계를 포함할 수 있다. 상기 방법은 상기 컨트롤러에 의해, 상기 카메라에 의해 획득된 비디오 데이터를 수신하는 단계를 더 포함할 수 있다.
- [0055] 상기 방법의 다른 실시예에서, 상기 인클로저의 적어도 일부는 가시광선에 대해 거의 불투명할 수 있다.
- [0056] 상기 방법의 다른 실시예에서, 상기 인클로저의 적어도 일부는 가시광선 파장에 대해 거의 불투명한 재료로부터 형성될 수 있다.
- [0057] 상기 방법의 다른 실시예에서, 상기 인클로저의 적어도 일부는 적외선 파장에 대해 거의 비반사적인(non-reflective) 재료로부터 형성될 수 있다.
- [0058] 상기 방법의 다른 실시예에서, 상기 인클로저의 적어도 일부는 폴리염화비닐(PVC, polyvinyl chloride) 또는 폴리옥시메틸렌(POM, polyoxymethylene)의 시트(sheet)로부터 형성될 수 있다.
- [0059] 상기 방법의 다른 실시예에서, 광원들의 제1 세트는 적어도 하나의 가시광선 파장에서 광을 방출하도록 구성된 적어도 하나의 제1 발광체를 포함할 수 있고, 광원들의 제2 세트는 적어도 하나의 적외선 파장에서 광을 방출하도록 구성된 적어도 하나의 제2 발광체를 포함할 수 있다.
- [0060] 상기 방법의 다른 실시예에서, 상기 적외선 파장은 거의 940 nm일 수 있다.
- [0061] 상기 방법의 다른 실시예에서, 상기 카메라는 적어도 480x480 픽셀의 해상도로 비디오 데이터를 획득하도록 구성될 수 있다.
- [0062] 상기 방법의 다른 실시예에서, 상기 카메라는 쥐의 움직임의 주파수보다 큰 프레임 레이트(frame rate)에서 비디오 데이터를 획득하도록 구성될 수 있다.
- [0063] 상기 방법의 다른 실시예에서, 상기 카메라는 적어도 29 fps(frames per second)의 프레임 레이트로 비디오 데이터를 획득하도록 구성될 수 있다.
- [0064] 상기 방법의 다른 실시예에서, 상기 카메라는 적어도 8-비트 심도(bit depth)를 갖는 비디오 데이터를 획득하도록 구성될 수 있다.
- [0065] 상기 방법의 다른 실시예에서, 상기 카메라는 적외선 파장에서 비디오 데이터를 획득하도록 구성될 수 있다.
- [0066] 상기 방법의 다른 실시예에서, 컨트롤러는 상기 카메라로부터 수신된 비디오 데이터를 압축하도록 구성될 수 있다.
- [0067] 상기 방법의 다른 실시예에서, 상기 컨트롤러는 분산 기반 배경 감산(variance-based background subtraction)을 사용하는 필터와 함께 MPEG4 코덱을 이용하여 상기 카메라로부터 수신된 비디오 데이터를 압축하도록 구성될 수 있다.
- [0068] 상기 방법의 다른 실시예에서, 상기 MPEG 코덱 필터는 Q0 HQDN3D일 수 있다.
- [0069] 상기 방법의 다른 실시예에서, 상기 컨트롤러는 명(light)/암(dark) 순환을 모의하는 스케줄에 따라 상기 인클로저를 조명하도록 상기 제1 광원에 명령하도록 구성될 수 있다.
- [0070] 상기 방법의 다른 실시예에서, 상기 컨트롤러는 상기 명/암 순환의 명 부분(light portion) 동안 약 50 럭스(lux) 내지 약 800 럭스의 세기를 갖는 가시광선으로 상기 인클로저를 조명하도록 상기 제1 광원에 명령하도록

구성될 수 있다.

[0071] 상기 방법의 다른 실시예에서, 상기 컨트롤러는 상기 인클로저의 온도가 적외선 조명에 의해 5℃ 미만으로 상승 되도록 적외선으로 상기 인클로저를 조명하도록 상기 제2 광원에 명령하도록 구성될 수 있다.

[0072] 상기 방법의 다른 실시예에서, 상기 컨트롤러는 로그 스케일링된 1024 레벨의 광에 따라 상기 인클로저를 조명하도록 상기 제1 광원에 명령하도록 구성될 수 있다.

### 도면의 간단한 설명

[0073] 이들 및 다른 특징들은 첨부 도면과 함께 기재된 다음의 상세한 설명으로부터 보다 쉽게 이해될 것이다.

도 1은 동물 추적을 위한 운영 환경의 하나의 예시적인 실시예를 도시한 흐름도이다.

도 2는 동물 행동 모니터링 시스템의 일 실시예의 개략도이다.

도 3A 내지 도 3F는 도 2의 시스템에 의해 획득된 샘플 프레임들을 도시한 이미지이다 - (A-C) 가시광선, (D-F) 적외선 -.

도 4A 내지 도 4B는 두 개의 카메라 모델에 대한 파장의 함수로서 양자 효율의 플롯이다 - (A) Sentech STC-MC33USB에 대한 상대 응답(Relative Response), (B) Basler acA1300-60gm-NIR의 양자 효율(Quantum Efficiency) -.

도 5는 적외선 롱-패스 필터(IR long-pass filter)에 대한 투명도(transparency)-파장(wavelength) 프로파일의 플롯이다.

도 6A 내지 도 6D는 상이한 압축 기술이 적용된 비디오 프레임의 예시적인 실시예들을 도시한 이미지이다 - (A) 압축 없음, (B) MPEG4 Q0, (C) MPEG4 Q5, (D) MPEG4 Q0 HQDN3D -.

도 7은 도 2의 시스템과 함께 사용하기 적합한 획득 시스템의 구성요소의 실시예를 도시한 도면이다.

도 8A는 검정색 쥐, 회색 쥐, 알비노 쥐, 및 얼룩무늬(piebald) 쥐를 포함하여, 본 개시에 따라 분석된 관찰 환경의 예시적인 실시예들의 개략도이다.

도 8B는 열악한 동물 추적을 일으키는 조건들의 개략도이다.

도 8C는 타원 형태로 대상 추적을 포함하는 쥐 추적의 하나의 예시적인 실시예의 개략도이다.

도 9는 분할 네트워크 아키텍처의 하나의 예시적인 실시예의 개략도이다.

도 10은 비닝된 분류 네트워크 아키텍처의 하나의 예시적인 실시예의 개략도이다.

도 11은 회귀 분류 네트워크 아키텍처의 하나의 예시적인 실시예의 개략도이다.

도 12A는 전경(F) 및 배경(B)에 대한 두 개의 마크들의 배치를 나타내는 그래픽 사용자 인터페이스의 하나의 예시적인 실시예이다.

도 12B는 도 12A의 마킹으로 인한 분할을 나타내는 그래픽 사용자 인터페이스의 하나의 예시적인 실시예이다.

도 13A는 도 9 내지 도 11의 분할, 회귀, 및 비닝된 분류 네트워크들의 실시예에 대한 훈련 곡선들의 플롯이다.

도 13B는 도 9 내지 도 11의 분할, 회귀, 및 비닝된 분류 네트워크들의 실시예에 대한 검증 곡선들의 플롯이다.

도 13C는 도 9의 분할 네트워크 아키텍처의 훈련 및 검증 성능을 도시한 플롯이다.

도 13D는 도 11의 회귀 네트워크 아키텍처의 훈련 및 검증 성능을 도시한 플롯이다.

도 13E는 도 10의 비닝된 분류 네트워크 아키텍처의 훈련 및 검증 성능을 도시한 플롯이다.

도 14A는 본 개시의 실시예들에 따라 상이한 크기의 세트들을 훈련하기 위한 단계의 함수로서 훈련 에러를 도시한 플롯이다.

도 14B는 본 개시의 실시예들에 따라 상이한 크기의 세트들을 훈련하기 위한 단계의 함수로서 검증 에러를 도시한 플롯이다.

도 14C는 훈련 샘플들의 전체 훈련 세트에 대한 단계의 함수로서 훈련 및 검증 에러를 도시한 플롯이다.

도 14D는 10,000(10k)개의 훈련 샘플들 포함하는 훈련 세트에 대한 단계의 함수로서 훈련 및 검증 에러를 도시한 플롯이다.

도 14E는 5,000(5k)개의 훈련 샘플들을 포함하는 훈련 세트에 대한 단계의 함수로서 훈련 및 검증 에러를 도시한 플롯이다.

도 14F는 2,500(2.5k)개의 훈련 샘플들을 포함하는 훈련 세트에 대한 단계의 함수로서 훈련 및 검증 에러를 도시한 플롯이다.

도 14G는 1,000(1k)개의 훈련 샘플들을 포함하는 훈련 세트에 대한 단계의 함수로서 훈련 및 검증 에러를 도시한 플롯이다.

도 14H는 500개의 훈련 샘플들을 포함하는 훈련 세트에 대한 단계의 함수로서 훈련 및 검증 에러를 도시한 플롯이다.

도 15A 내지 도 15D는 각각의 쥐를 서로 구별하는 오버레이된 색상 표시를 갖는 캡처된 비디오 데이터의 프레임들이다 - (A-B) 가시광선 조명, (C-D) 적외선 조명 -.

도 16은 도 9의 분할 네트워크 아키텍처의 성능을 빔 브레이크 시스템과 비교한 플롯이다.

도 17A는 본 개시의 일 실시예 및 Ctrax로부터의 예측을 도시한 플롯이다.

도 17B는 도 9의 분할 네트워크 아키텍처에 의해 결정된 단축 예측의 상대 표준편차의 플롯이다.

도 18A는 도 9의 분할 네트워크 아키텍처에 의해 결정된 유전적으로 다양한 동물들의 대규모 변형률 조사를 위해 추적된 총 거리의 플롯이다.

도 18B는 도 9의 분할 네트워크 아키텍처에 의해 결정된 동적 환경에서 4일에 걸쳐 연속적으로 추적된 6마리의 동물들에서 관찰된 생체주기 활동 패턴의 플롯이다.

도면들이 반드시 축척대로 된 것은 아니라는 점에 유의한다. 도면들은 본 명세서에 개시된 주제의 전형적인 측면들만을 도시하도록 의도되며, 따라서 본 개시의 범위를 제한하는 것으로 간주되어서는 안 된다.

### 발명을 실시하기 위한 구체적인 내용

[0074] 명확성을 위해, 본 명세서에서 행동 모니터링을 용이하게 하기 위한 적어도 한 마리의 동물의 비디오 캡처 및 적어도 한 마리의 동물의 추적을 위한 시스템들 및 대응하는 방법들의 예시적인 실시예들은, 쥐와 같은, 작은 설치류와 관련하여 논의된다. 그러나 개시된 실시예들은 제한없이 다른 동물들을 모니터링 하기 위해 채용 및/또는 적용될 수 있다.

[0075] 도 1은 아레나(200), 획득 시스템(700), 및 신경망 추적기를 구현하도록 구성된 추적 시스템을 포함하는 운영 환경(100)의 하나의 예시적인 실시예를 도시한 개략도이다. 아래에서 더 상세히 논의되는 바와 같이, 적어도 한 마리의 쥐가 아레나(200)에 수용될 수 있다. 쥐와 같은, 적어도 한 마리의 동물의 비디오 데이터가 획득된다. 비디오 데이터는 단독으로 또는, 오디오 또는 환경 파라미터들(예컨대, 온도, 습도, 광도 등)과 같은, 동물 모니터링과 관련된 다른 데이터와 함께 획득될 수 있다. 카메라, 마이크로폰, 조명, 다른 환경 센서, 데이터 저장, 및 데이터 압축의 제어와 같은, 이 데이터를 획득하는 프로세스는 획득 시스템(700)에 의해 수행될 수 있다. 획득된 비디오 데이터는 비디오 데이터에 기초하여 적어도 한 마리의 동물을 추적할 수 있는 컨볼루션 신경망(CNN, convolutional neural network)을 실행할 수 있는 추적 시스템에 입력될 수 있다.

### [0076] I. 비디오 데이터 획득

[0077] 일 실시예에서, 동물들의 움직임을 포함하는 비디오 데이터의 캡처를 위한 시스템들 및 방법들이 제공된다. 아래에서 논의되는 바와 같이, 비디오 데이터는 미리 결정된 기간(예컨대, 적어도 하나의 분, 시간, 일, 주, 월, 년 등)에 걸쳐 연속적으로 획득될 수 있다. 해상도, 프레임 레이트(frame rate), 및 비트 심도(bit depth) 중 적어도 하나 이상을 포함하지만 이에 제한되지 않는, 비디오 데이터의 특성들은 행동 패턴들을 추출하기 위한 후속 분석을 용이하게 하기에 충분할 수 있다. 실제 작동하는 솔루션이 제공되었고, 종래의 비디오 캡처 시스템들보다 강력하고 품질이 높은 것으로 나타났다. 본 개시의 실시예들은 쥐를 눈에 띄게 표시하는 몇 가지 방법들로 테스트된다. 비디오 및 초음파 발성 데이터의 동기화된 획득의 실무 예시들도 제공된다.

[0078] 일 실시예에서, 비디오 모니터링 시스템은 약 4 내지 6주의 기간 동안 동물 모니터링을 위해 배치될 수 있다.

배치(deployment)는 이미지 캡처와 아레나 설계, 챔버(chamber) 설계 개선, 비디오 획득 소프트웨어의 개발, 및 오디오 데이터의 획득, 카메라의 스트레스 테스트, 챔버와 소프트웨어, 및 배치 단계(phase)를 위한 챔버 생산에 관한 결정 중 적어도 하나를 포함할 수 있다. 이들 각각은 아래에 상세히 설명되어 있다. 위에서 언급된 4 내지 6주의 관찰 기간은 예시적인 목적으로 제공되며, 본 개시의 실시예들은, 필요에 따라, 더 길거나 더 짧은 기간 동안 사용될 수 있음이 이해될 수 있다.

#### [0079] a. 아레나 설계

[0080] 고품질의 행동 데이터를 획득하기 위해서는 적절한 아레나 설계가 중요할 수 있다. 이 아레나는 동물의 "집(home)"이고, 환경 교란으로부터의 격리, 적절한 생체주기의 조명, 음식, 물, 및 잠자리(bedding) 중 적어도 하나를 제공하도록 구성될 수 있으며, 그리고 일반적으로 스트레스가 없는 환경이다.

[0081] 행동의 관점에서, 그 영역이 스트레스 및 환경 교란을 최소화하고 자연스러운 행동이 표현되도록 하는 것이 바람직할 수 있다.

[0082] 축산의 관점에서, 아레나가 쥐의 세척, 추가 또는 제거, 및 음식과 물을 첨가하고 제거하는 것을 용이하도록 하는 것이 바람직할 수 있다.

[0083] 수의학적 관점에서, 아레나가 관심있는 행동을 실질적으로 방해하지 않고 환경 조건을 모니터링 하는 것뿐만 아니라 건강 확인 및 의료 제공을 용이하게 하도록 하는 것이 바람직할 수 있다.

[0084] 컴퓨터 비전의 관점에서, 아레나가 실질적인 폐색(occlusion), 왜곡(distortion), 반사(reflection), 및/또는 소음 공해 없이 관심있는 행동의 표현을 실질적으로 방해하지 않으면서 고품질의 비디오 및 오디오의 획득을 용이하게 하도록 하는 것이 바람직할 수 있다.

[0085] 설비의 관점에서, 아레나가 바닥 면적을 실질적으로 최소화하고 분해 또는 재조립할 필요 없이 상대적으로 쉬운 보관을 제공하도록 하는 것이 바람직할 수 있다.

[0086] 따라서, 아레나는 행동, 축산, 컴퓨터, 및 설비의 균형을 제공하도록 구성될 수 있다. 아레나(200)의 하나의 예시적인 실시예가 도 2에 도시되어 있다. 아레나(200)는 인클로저(enclosure)(204)가 장착되는 프레임(frame)(202)을 포함할 수 있다. 인클로저는 인클로저(204) 내부로의 접근을 허용하도록 구성된 도어(door)(206)를 포함할 수 있다. 적어도 하나의 카메라(210) 및/또는 발광체(212)는 프레임(202)에 인접하여 또는 직접적으로 장착될 수 있다(예컨대, 인클로저(204)의 위에).

[0087] 아래에서 상세히 설명되는 바와 같이, 특정 실시예들에서, 발광체(212)는 광원들의 적어도 두 개의 세트들을 포함할 수 있다. 광원들의 각각의 세트는 다른 세트와 파장이 다르고 인클로저(204)에 입사하는 광을 방출하도록 구성된 적어도 하나의 발광체를 포함할 수 있다. 일 예로서, 광원들의 제1 세트는 적어도 하나의 가시광선 파장(예컨대, 약 390nm 내지 약 700nm)에서 광을 방출하도록 구성될 수 있고, 광원들의 제2 세트는 적어도 하나의 적외선 파장(예컨대, 약 700nm 초과 내지 약 1mm)에서 광을 방출하도록 구성될 수 있다.

[0088] 카메라(210) 및/또는 발광체(212)는 사용자 인터페이스(214)와 전기적으로 통신할 수 있다. 사용자 인터페이스(214)는 카메라(210)에 의해 획득된 비디오 데이터를 보도록 구성된 디스플레이일 수 있다. 특정 실시예들에서, 사용자 인터페이스(214)는 카메라(210) 및/또는 발광체(212)의 제어를 위한 적어도 하나의 사용자 인터페이스를 표시하도록 구성된 터치 스크린 디스플레이일 수 있다.

[0089] 대안적으로 또는 추가적으로, 카메라(110), 발광체(212), 및 사용자 인터페이스(214)는 컨트롤러(controller)(216)와 전기적으로 통신할 수 있다. 컨트롤러(216)는 카메라(210)에 의한 비디오 데이터의 획득, 발광체(212)에 의한 광의 방출, 및/또는 사용자 인터페이스(214)에 의한 획득된 비디오 데이터의 표시를 제어하도록 동작하는 제어 신호를 생성하도록 구성될 수 있다. 특정 실시예들에서, 사용자 인터페이스는 선택적으로 생략될 수 있다.

[0090] 컨트롤러(216)는 또한 데이터 저장 장치(data storage device)(220)와 통신할 수 있다. 컨트롤러(216)는 카메라(210)에 의해 획득된 비디오 데이터를 수신하고 획득된 비디오 데이터를 저장하기 위해 데이터 저장 장치(220)에 전송하도록 구성될 수 있다. 적어도 하나의 카메라(210), 발광체(212), 유저 인터페이스(214), 컨트롤러(216), 및 데이터 저장 장치(220) 사이의 통신은 유선 통신 링크, 무선 통신 링크, 및 이들의 조합을 사용하여 수행될 수 있다.

[0091] 아래에서 논의되는 바와 같이, 아레나(200)는 행동, 축산, 컴퓨터, 및 설비의 원하는 균형(balance)을 달성하도록

록, 또한 미리 결정된 기간(예컨대, 약 5개월) 내에 완료를 허용하도록 구성된 오픈 필드 설계를 가질 수 있다.

[0092] 재료

[0093] 특정 실시예들에서, 인클로저(204)를 형성하는 재료의 적어도 일부(예컨대, 인클로저(204)의 하부)는 가시광선 파장에 대해 실질적으로 불투명할 수 있다. 이런 방식으로, 인클로저(204) 내에서 동물에 의해 관찰될 수 있는 시각적 신호들(예컨대, 물체 및/또는 사용자의 움직임)뿐만 아니라, 발광체(212) 이외의 광원들에 의해 방출된 가시광선이 감소 및/또는 거의 제거될 수 있다. 추가적인 실시예들에서, 인클로저(204)를 형성하는 재료는 비디오 데이터의 획득을 용이하게 하는 적외선 파장들에 거의 비반사적(non-reflective)일 수 있다. 인클로저(204)의 벽들의 두께는 기계적 지지를 제공하기에 적합한 범위(예컨대, 약 1/8인치 내지 약 1/4인치) 내에서 선택될 수 있다.

[0094] 일 실시예에서, 인클로저(204)는 폴리염화비닐(PVC, polyvinyl chloride) 또는 폴리옥시메틸렌(POM, polyoxymethylene)으로부터 형성된 발포 시트(foam sheet)를 이용하여 지어질 수 있다. Delrin®(DuPont, Wilmington DE, USA)은 폴리옥시메틸렌(POM)의 일 예이다. 유리하게는, 이러한 발포 시트는 장기적인 동물 모니터링을 위한 아레나(200)에 충분한 다목적성 및 내구성을 부여할 수 있다.

[0095] 일 실시예에서, 프레임(202)은 레그(202a) 및 그 사이로 연장되는 (예컨대, 수평으로) 적어도 하나의 선반(202b)을 포함할 수 있다. 일 예로서, 프레임(202)은 저장 영역으로 이동하기 위한 잠금 휠(locking wheel)을 갖는 미리 결정된 크기의 상업용 선반 시스템일 수 있다. 일 실시예에서, 미리 결정된 크기는 약 2피트 x 2피트 x 6피트(예컨대, Super Erecta Metroseal 3TM, InterMetro Industries Corporation, Wilkes-Barre, PA, USA)일 수 있다. 그러나, 다른 실시예들에서, 상이한 크기의 아레나들이 제한없이 사용될 수 있다.

[0096] **b. 데이터 획득**

[0097] 비디오 획득 시스템은 카메라(210), 발광체(212), 사용자 인터페이스(214), 컨트롤러(216), 및 데이터 저장 장치(220)를 포함할 수 있다. 비디오 획득 시스템은, 비디오 압축 및 저장뿐만 아니라, 비디오 획득의 프레임 레이트, 비트 심도, 각 프레임의 해상도, 및 적외선 범위의 스펙트럼 감도 중 적어도 하나를 포함하지만 이에 제한되지 않는, 성능 특성들의 미리 결정된 균형을 갖는 것이 사용될 수 있다. 아래에서 논의되는 바와 같이, 이들 파라미터들은 품질을 최대화하고 데이터의 양을 최소화하기 위해 최적화될 수 있다.

[0098] 일 실시예에서, 카메라(210)는 약 640x480 픽셀의 해상도, 약 29 fps(frames per second), 및 약 8 비트 심도 중 적어도 하나를 갖는 비디오 데이터를 획득할 수 있다. 이들 비디오 획득 파라미터들을 이용하면, 약 33GB/hr의 비압축된(uncompressed) 비디오 데이터가 생성될 수 있다. 일 예로서, 카메라(210)는 Sentech USB2 카메라(Sensor Technologies America, Inc., Carrollton, TX, USA)일 수 있다. 도 3A 내지 도 3F는 가시광선(도 3A 내지 도 3C) 및 적외선(도 3D 내지 도 3F)을 이용하는 비디오 획득 시스템의 실시예로부터 획득된 샘플 프레임들을 도시한다.

[0099] 수집된 비디오 데이터는, 아래에서 논의되는 바와 같이, 카메라(210) 및/또는 컨트롤러(216)에 의해 압축될 수 있다.

[0100] 다른 실시예에서, 비디오 획득 시스템은 획득한 비디오 데이터의 해상도를 거의 두 배(예컨대, 약 960x960 픽셀)로 높이도록 구성될 수 있다. Sentech USB보다 높은 해상도를 갖는 4개의 추가적인 카메라가, 아래에 표시된 것과 같이, 조사되었다.



[0101] 표 1 - 예시적인 카메라 및 선택된 특성

카메라	인터페이스	해상도	최대 프레임 레이트	최대 비트 심도	셔터	양자 효율 @ 940nm	비압축 데이터
Sentech STC-MB33USB	USB3	640x480	60	12	글로벌	6%	9.2MB/s @30fps, 8bit
Basler acA1300-60gm-NIR	GigE	1280x1024	60	10	글로벌	21%	39.3MB/s @30fps, 8bit
Raspberry Pi NoIR V2	Serial	1280x720 1640x1232	90 30	8	글로벌	NA	H264 Lossy Compressed, Max 3.1MB/s
PtGrey BFLY-U3-23S6M-C	USB3	1920x1200	41	12	글로벌	5%	69.1MB/s @30fps, 8bit
Basler daA1600-60um	USB3	1600x1200	60	12	글로벌	7%	57.6MB/s @30fps, 8bit

[0102]

[0103]

이들 카메라들은 가격, 해상도, 최대 프레임 레이트, 비트 심도, 및 양자 효율이 달라질 수 있다.

[0104]

비디오 획득 시스템의 실시예들은 약 30fps 및 약 8비트 심도에서 흑백으로 비디오 데이터를 수집하도록 구성될 수 있다. Shannon-Nyquist 정리에 따르면, 프레임 레이트는 관심있는 이벤트 빈도의 적어도 두 배가 되어야 한다(예: Shannon, 1949 참조). 쥐의 행동은 털 손질(grooming)의 경우 수 Hz에서 털기(whisking)의 경우 20Hz까지 다양할 수 있다(예: Deschenes 외, 2012; Kalueff 외, 2010; Wiltischko 외, 2015 참조). 털 손질은 최대 약 7Hz에서 발생하는 것으로 관찰되었고, 따라서, 쥐의 움직임의 주파수보다 큰 프레임 레이트(예컨대, 29fps)로 비디오를 녹화하는 것이 대부분의 쥐의 행동들을 관찰하기에 적합한 것으로 간주된다. 그러나, 적외선 범위에서는 카메라의 감도가 급격히 떨어질 수 있다. 이러한 대비(contrast) 손실은 적외선 광의 레벨을 증가시킴으로써 극복될 수 있지만, 적외선 광의 세기를 증가시키는 것은 환경 온도의 증가로 이어질 수 있다.

[0105]

#### 조명

[0106]

위에서 언급한 바와 같이, 발광체(212)는 백색 가시광선 및 적외선과 같은, 적어도 하나의 유형의 광을 방출하도록 구성될 수 있다. 가시광선은 조명에 이용될 수 있고, 명(light)/암(dark) 순환 및 조정 가능한 세기를 제공하도록 (예컨대, 컨트롤러(216)에 의해) 프로그래밍 될 수 있다. 광의 주기를 조절하는 능력은 동물이 야생에서 접하는 태양광의 모의를 가능하게 한다. 명암 주기의 길이는 계절을 모의하기 위해 조절될 수 있고, 시차 지연(생체주기의 진행 및 지연) 실험을 모의하기 위해 광 변환(light shift)이 수행될 수 있다. 또한, 특정 동물들의 불안을 유발하기 위해 높은 광이 사용될 수 있고, 상이한 탐색적 행동을 이끌어내기 위해 낮은 광이 사용될 수 있다. 따라서 광의 세기뿐만 아니라 명/암 길이를 일시적으로 제어하는 능력은 적절한 행동 실험에 중요하다.

[0107]

특정 실시예들에서, 컨트롤러(216)는 명/암 순환의 명 부분(light portion) 동안 약 50럭스 내지 약 80럭스의 세기를 갖는 가시광선으로 인클로저(204)를 조명하도록 가시광선 광원에 명령하도록 구성될 수 있다. 선택된 광의 세기는 관찰되는 행동의 유형에 따라 달라질 수 있다. 일 측면에서, 관찰을 위해 쥐에 의한 탐색 활동을 부추기기 위해 상대적으로 낮은 세기(예컨대, 약 200럭스 내지 약 300럭스)가 사용될 수 있다.

[0108]

특정 실시예들에서, 거의 모든 비디오 데이터는 적외선 롱-패스 필터(IR long-pass filter)를 이용하여 적외선 범위에서 카메라(210)에 의해 획득될 수 있다. 적외선 롱-패스 필터는 카메라(210)로 입력되는 거의 모든 가시광선을 제거할 수 있다. 유리하게는, 적외선은 낮 또는 밤 시간에 상관없이 인클로저(104)의 균일한 조명을 제공할 수 있다.

[0109]

두 개의 적외선 파장 - 850nm 및 940nm LED - 이 평가되었다. 850nm 광은 육안으로 볼 수 있는 선명한 붉은 빛깔을 나타내고, 동물에 대한 낮은 광 노출로 이어질 수 있다. 그러나, 이러한 희미한 광은 쥐의 기분 변화로 이어질 수 있다. 따라서, 녹화(recording)를 위해 940nm 광이 선택된다.

[0110]

940nm 파장에서의 녹화는 매우 낮은 양자 수율(quantum yield)을 가질 수 있으며, 이는 높은 이득(gain)으로 인

해 거칠게 보이는 이미지로서 나타날 수 있다. 따라서, 적외선 조명으로 인해 인클로저(204)의 온도를 실질적으로 올리지 않고 얻을 수 있는 최대 광 레벨을 식별하기 위하여 상이한 카메라를 이용하는 다양한 적외선 조명 레벨이 평가되었다. 특정 실시예들에서, 인클로저(204)의 온도는 약 5℃ 이하(예컨대, 약 3℃ 이하)로 증가될 수 있다.

[0111] 또한, Basler acA1300-60gm-NIR 카메라가 평가되었다. 이 카메라는, 도 4A 내지 도 4B에 표시된 바와 같이, 표 1에 열거된 다른 카메라들과 비교하여 거의 3~4배의 스펙트럼 감도를 갖는다. 도 4A는 대표 예시로서 Sentech 카메라의 스펙트럼 감도를 상대 응답(relative response)의 관점에서 도시하고, 도 4B는 Basler 카메라의 스펙트럼 감도를 양자 효율(quantum efficiency)의 관점에서 도시한다. 양자 효율은 센서에 부딪치는 광자(photon)와 비교할 때 방출되는 전자의 측정값이다. 상대 응답은 0에서 1의 스케일로 나타낸 양자 효율이다. 940nm 파장은 참고를 위해 도 4A 내지 도 4B에 수직선으로 추가적으로 도시되어 있다.

[0112] 발광체(212)에 의해 제공되는 가시광선 주기는 컨트롤러(216) 또는 발광체(212)와 통신하는 다른 장치에 의해 제어될 수 있다. 특정 실시예들에서, 컨트롤러(216)는 조명 제어 보드(Phenome Technologies, Skokie, IL)를 포함할 수 있다. 이 보드는 RS485 인터페이스를 통해 제어될 수 있는 로그 스케일링된 1024 레벨의 광을 가지며, 그것은 새벽(dawn)/황혼(dusk) 이벤트가 가능하다. 아래에서 더 상세히 논의되는 바와 같이, 가시광선의 제어는 컨트롤러(216)에 의해 실행되는 제어 소프트웨어에 통합될 수 있다.

[0113] 필터

[0114] 위에서 언급된 바와 같이, 선택적으로, 비디오 데이터 획득 동안 거의 모든 가시광선이 카메라(210)에 도달하는 것을 차단하기 위해, 적외선 롱-패스 필터가 사용될 수 있다. 일 예로서, 물리적 적외선 롱-패스 필터가 카메라(들)(110)과 함께 사용될 수 있다. 이 구성은 아레나(200)에서의 명 또는 암 단계에 상관없이 실질적으로 균일한 광을 제공할 수 있다.

[0115] 개시된 시스템들 및 방법들의 실시예들에서 사용하기에 잠재적으로 적합한 필터 프로파일들이 도 5에 도시되어 있다(예컨대, 적외선 통과 필터(092 및 093)). 적외선을 차단하는 적외선 컷 필터(486)가 비교를 위해 도시되어 있다. RG-850(유리, Edmunds Optics) 및 43-939(플라스틱, 레이저 절단 가능, Edmunds Optics)에 대한 추가적인 프로파일들도 적합할 수 있다.

[0116] 렌즈

[0117] 일 실시예에서, 카메라 렌즈는 1/3" 3.5-8mm f1.4 (CS 마운트)일 수 있다. 이 렌즈는 도 3A 내지 도 3B에 보이는 이미지를 생성할 수 있다. C 마운트 렌즈의 유사한 렌즈도 사용될 수 있다.

[0118] 비디오 압축

[0119] 압축을 무시하면, 로(raw) 비디오 데이터는 카메라(210)에 의해 약 1MB/frame, 약 30MB/second, 약 108GB/Hour, 약 2.6TB/day의 비율로 생성될 수 있다. 저장 방법 선택할 때, 다양한 목표가 고려될 수 있다. 비디오의 맥락(context)에 따라, 장기간 저장 전에 비디오의 특정 요소들을 제거하는 것이 중요한 옵션이 될 수 있다. 추가적으로, 장기간 저장을 고려할 때 (예컨대, 컨트롤러(216)에 의해) 필터 또는 다른 형태의 처리를 적용하는 것이 바람직할 수 있다. 그러나, 원본(original) 또는 로(raw) 비디오 데이터를 저장하는 것은 처리 방법이 추후에 변경될 때 유용한 해결책이 될 수 있다. 비디오 압축 테스트의 일 예가 아래에 설명된다.

[0120] 약 480x480의 픽셀 해상도, 약 29fps, 및 약 8비트/픽셀에서 약 100분 동안 수집된 비디오 데이터에 대하여 여러 압축 표준들이 평가되었다. 로(raw) 비디오에서 테스트 된 두 개의 무손실 포맷은 Dirac 및 H264이다. H264는 약간 더 작은 파일 크기를 갖지만 코드 변환에 약간 더 많은 시간이 걸린다. Dirac은 다른 포맷으로의 후속 코드 변환으로 폭넓게 지원될 수 있다.

[0121] H264와 밀접한 관련이 있고, 비트 레이트를 양호하게 제어하는 것으로 알려져 있기 때문에, MPEG4 손실 포맷도 평가되었다. 비트 레이트를 설정하는 방법에는 두 가지가 있다. 첫 번째는 전체 인코딩된 비디오에서 일정한 고정 비트 레이트를 설정하는 것이고, 두 번째는 원본 비디오와의 편차에 기초하여 가변 비트 레이트를 설정하는 것이다. MPEG4 인코더를 사용하는 ffmpeg 내에서, 가변 비트 레이트를 설정하는 것은 품질 값(0-31에서 0은 거의 무손실임)의 선택을 통해 쉽게 달성될 수 있다.

[0122] 도 6A 내지 도 6D에서 세 가지의 상이한 이미지 압축 방법들이 원본(로(raw)) 캡처 비디오 프레임과 비교된다. 원본 이미지는 도 4A에 도시되어 있다. 다른 세 가지 방법들은 도 4B의 원본과 픽셀의 차이로 나타나고 압축의 효과만을 보여준다. 즉, 압축된 이미지가 원본과 얼마나 다른 가를 보여준다. 따라서, 작은 차이가 더 좋고, 더

큰 압축율이 더 좋다. 도 4B에 도시된 바와 같이, Q0 필터를 갖는 MPEG4 코덱에 따라 수행된 압축은 1/17의 압축비를 나타낸다.

[0123] 게시된 실시예들에 따라 수집된 비디오 데이터는 quality 0 파라미터(Q0 필터, 도 4B; Q0 HQDN3D 필터)를 사용할 때 원본으로부터 변경된 (세기의 최대 4% 증가 또는 감소) 거의 0.01%의 픽셀을 갖는다. 이것은 프레임 당 약 25픽셀을 차지한다. 이 픽셀의 대부분은 그림자의 경계에 위치한다. 이 작은 이미지의 변화는 카메라(210) 자체를 방해하는 노이즈의 스케일을 따른다는 것이 이해될 수 있다. 이들은 종종 압축하는 동안 주의를 기울이지 않을 때 나타나는 덩어리 픽셀화 된 아티팩트에 익숙하다.

[0124] 이들 포맷 외에, 개별 사용자의 데이터 세트를 수용하기 위해 다른 적합한 무손실 포맷들이 생성될 수 있다. 이들 중 두 가지는 FMF(fly movie format) 코덱 및 UFMF(micro fly movie format) 코덱이다. 이들 포맷들의 목적은 외부 정보를 최소화하고 추적을 위한 가독성을 최적화하는 것이다. 이들 포맷들은 무손실이고 정적 백그라운드 모델에서 작동하기 때문에, 필터링 되지 않은 센서 노이즈는 어떠한 실질적인 데이터 압축도 허용하지 않는다. 이 압축 평가의 결과가 표 2에 예시되어 있다.

[0125] 표 2 - 압축 테스트

코덱	필터	파일 크기 (bytes)	비고
Raw	None	41,911,246,848	
Dirac	None	15,568,246,208	무손실
H264	None	14,840,999,602	무손실
MPEG4	Q0	2,471,731,352	손실
Dirac	HQDN3D	10,113,632,202	필터됨
H264	HQDN3D	8,008,906,680	필터됨
MPEG4	HQDN3D Q0	429,522,590	손실 필터됨
FMF	None	41,720,668,160	무손실
UFMF	None	41,722,662,912	무손실

[0126]

[0127] 데이터 압축을 위한 코덱의 선택에 더하여, 이미지의 배경 노이즈의 감소가 또한 바람직할 수 있다. 배경 노이즈는 모든 카메라에 내재되어 있으며 종종, 이미지 내에서 기준 노이즈를 나타내는, 어두운 노이즈로 표시된다.

[0128] 더 긴 노출 시간, 더 넓은 조리개, 및 이득(gain) 감소를 포함하여 이 노이즈를 제거하는 많은 방법들이 있다. 그러나, 이러한 방법들이 실험에 직접적인 영향을 미치는 경우 실행 가능한 옵션이 아니다. 따라서, ffmpeg의 HQDN3D 필터가 사용될 수 있으며, 이는 시간 및 공간 정보를 사용하여 작은 변동을 제거한다.

[0129] 도 6B 내지 도 6D에 나타난 바와 같이, HQDN3D 필터는 획득된 비디오 데이터의 파일 크기의 현저한 감소(예컨대, 원본 비디오 데이터의 파일 크기와 비교하여 거의 100배 더 작음)를 제공하는 것으로 관찰된다. HQDN3D 필터로 MPEG4 코덱을 사용하여 압축한 후, 결과적인 평균 비트 레이트는 압축된 비디오의 약 0.34GB/hr 일 수 있다. 또한, 실질적으로 모든 정보 손실은 센서 노이즈(귀가 없는 상태에서 획득한 비디오)에서 생성된 것보다 수십 배 적은 것으로 실험적으로 검증되었다. 이러한 유형의 노이즈 제거는 압축성을 크게 향상시킨다.

[0130] 예상치 못하게, HQDN3D 필터는 컨볼루션 신경망(CNN)에 의해 수행되는 추적 성능의 현저한 증가를 제공하는 것으로 밝혀졌으며, 이는 아래에서 상세히 논의된다. 이론에 구속되지 않고, HQDN3D 필터가 분산-기반 배경 감산 방법이기 때문에 이 성능 향상이 달성되는 것으로 여겨진다. 분산이 낮을수록, 전경(foreground)을 식별하기 쉽고 고품질의 추적을 생성한다.

[0131] **초음파 오디오 획득**

[0132] 쥐는 사회적 의사소통, 교배, 침략, 및 새끼 양육을 위해 초음파 범위에서 발성할 수 있다(예: Grimsley 외, 2011). 후각 및 촉각 신호와 함께, 이 발성은 쥐의 의사소통의 가장 중요한 형태 중 하나일 수 있다. 쥐에게서 시험되지는 않았지만, 인간에게는, 음성 및 발성의 변화(노인성 음성)는 사춘기 및 노화와 같은 전이를 정의할 수 있다(예: Decoster and Bebruyne, 1997; Martins 외, 2014; Mueller, 1997 참조)

[0133] 따라서, 아래에서 상세히 논의되는 바와 같이, 아레나(200)의 실시예들은 적어도 하나의 마이크로폰(222)을 더 포함할 수 있다. 마이크로폰(222)은 프레임(222)에 장착될 수 있고, 인클로저(204)에 위치된 동물로부터 오디오



데이터를 획득하도록 구성될 수 있다. 동기화된 데이터 수집은 마이크로폰 어레이의 형태의 마이크로폰(222)의 사용에 의해 파일럿 될 수 있다. 마이크로폰(222)의 이 구성은 어떤 쥐가 발생하고 있는 지를 정확하게 지시할 수 있게 한다. 쥐 그룹 중 어떤 쥐가 발생하는 지를 추가적으로 결정하는 능력은 최근 마이크로폰 어레이를 이용하여 입증되었다(예: Heckman 외, 2017; Neunuebel 외, 2015 참조).

[0134] 데이터 수집 설정은 Neunuebel 외와 유사하게 제공될 수 있다. 아레나의 측면에 소리를 캡처할 수 있는 네 개의 마이크로폰이 배치될 수 있다. 비디오 데이터와 통합될 때, 발생하는 쥐는 최대 가능성 방법을 이용하여 식별될 수 있다(예: Zhang 외, 2008 참조).

#### [0135] **환경 센서**

[0136] 일 실시예에서, 아레나(200)는 온도, 습도, 및/또는 광 세기(예컨대, 가시광선 및/또는 적외선)와 같은 적어도 하나의 환경 파라미터를 측정하도록 구성된 적어도 하나의 환경 센서(224)를 더 포함할 수 있다. 특정 실시예들에서, 환경 센서(224)는 적어도 두 개의 환경 파라미터들을 측정하도록 통합 및 구성될 수 있다(예: Phenome Technologies, Skokie, IL). 환경 센서(224)는 광 레벨과 함께 매일 온도 및 습도 데이터를 수집하기 위해 컨트롤러(216)와 전기적으로 통신할 수 있다. 수집된 환경 데이터는 조명 활동뿐만 아니라 최소 및 최대 온도를 나타내는 사용자 인터페이스에 표시하기 위해 출력될 수 있다(아래의 제어 소프트웨어에 관한 설명 참조).

#### [0137] **소프트웨어 제어 시스템**

[0138] 소프트웨어 제어 시스템은 데이터 획득 및 광 제어를 위해 컨트롤러(216)에 의해 실행될 수 있다. 소프트웨어 제어 시스템은 비디오, 오디오/초음파, 및 환경 데이터를 해당 타임 스탬프와 함께 독립적인 수집을 제공하도록 구성될 수 있다. 데이터는 임의의 미리 결정된 시간 주기동안(예컨대, 적어도 하나의 초, 분, 시간, 일, 년 등), 중단없이, 이러한 방식으로 수집될 수 있다. 이것은 추후에 획득된 비디오, 오디오/초음파, 및/또는 환경 데이터를 분석 또는 프레젠테이션을 위해 모으거나 동기화할 수 있도록 한다.

#### [0139] **운영체제**

[0140] 운영체제의 선택은 다양한 센서를 위한 드라이버의 이용 가능성에 의해 구동될 수 있다. 예를 들어, Avisoft Ultrasonic 마이크로폰 드라이버는 Windows 운영체제만 호환된다. 그러나 선택은 다음에 영향을 줄 수 있다.

[0141] · 프로세스 간 통신 - 프로세스 간 통신 옵션은 기본 운영체제의 영향을 받는다. 마찬가지로, 운영체제는 스레드 간의 통신 선택에 영향을 주지만, QT와 같은 크로스 플랫폼 프레임워크에서 개발하면 이 문제를 해결할 수 있다.

[0142] · 시스템 클럭에 대한 액세스 - 고해상도 시스템 클럭에 액세스하는 방법은, 아래에서 더 상세히 논의되는 바와 같이, 운영체제마다 다르다.

#### [0143] **하드웨어 옵션**

[0144] 특정 실시예들에서, 제어 시스템은 컨트롤러(216)에 의해 단일 보드 컴퓨터의 형태로 구현될 수 있다. 지속적인 운영을 위해 매우 강인한 MilSpec/산업용 컴퓨터를 포함하여 사용 가능한 여러 가지 옵션이 있다.

#### [0145] **외부 클럭 vs 시스템 클럭**

[0146] 시스템에 외부 클럭을 도입하지 않고서도 시스템 클럭으로부터 적절한 실시간 클럭 값을 이용할 수 있다. Posix 시스템에서 clock\_gettime(CLOCK\_MONOTONIC, ...) 함수는 초(s)와 나노초(ns)를 반환할 수 있다. 클럭의 해상도는 clock\_getres() 함수를 이용하여 쿼리될 수 있다. 제어 시스템의 실시예들의 클럭 해상도는 약 33 밀리초(ms)의 프레임 주기보다 작은 것이 바람직할 수 있다. 일 실시예에서, 시스템 클럭은 Unix 시스템이다.

[0147] GetTickCount64() 시스템 함수는 시스템이 시작된 이후 밀리초(ms)의 수를 얻기 위해 개발되었다. 이 타이머의 예상되는 해상도는 약 10 내지 16 밀리초(ms)이다. 이는 값 줄 바꿈을 확인하고 설명하는 데 유리할 수 있지만, 이는 clock\_gettime() 시스템 콜과 동일한 목적으로 사용된다.

[0148] Macintosh 컴퓨터에서, 시스템 클럭에 대한 유사한 액세스가 있다. 다음의 코드 스니펫(code snippet)은 마이크로초 미만의 해상도를 평가 및 관찰하였다.

[0149] `clock_serv_t cclock;`

[0150] `mach_timespec_t mts;`

- [0151] *host\_get\_clock\_service(mach\_host\_self(),SYSTEM\_CLOCK,&cclock);*
- [0152] *clock\_get\_time(cclock, &mts);*
- [0153] 임의의 운영체제에서, 시간을 반환하는 시스템 호출은 조정이 주기적으로 이루어질 때 때때로 뒤로 이동할 수 있다. 일 실시예에서, 단조 증가하는 시스템 클럭이 사용될 수 있다. GetTickCount64(), clock\_gettime(), 및 clock\_get\_time()은 모두 이 기준을 만족시킬 수 있다.
- [0154] 비디오 파일 크기
- [0155] 카메라 벤더 소프트웨어가 적절한 타임 스탬프 출력 파일을 적당한 크기로 자동으로 분할하여 저장할 가능성은 낮다. 컨트롤러(116)의 실시예들은 카메라(110)에서 각 프레임을 읽고 수집된 비디오 데이터를 간단한 형태로 제공하는, 중단 없는 방법으로 데이터를 수집하는 것이 바람직하다. 예를 들어, 컨트롤러(116)는 타임 스탬프 헤더 또는 프레임 사이의 타임 스탬프와 함께, 로(raw) 포맷으로 파일 당 약 10분의 비디오 프레임을 데이터 저장 장치(120)에 제공하도록 구성될 수 있다. 그러면 각 파일은 2GB 미만이 된다.
- [0156] 제어 시스템 아키텍처
- [0157] 도 7은 획득 시스템(700)의 구성을 도시하는 블록도이다. 특정 실시예들에서, 획득 시스템(700)은 컨트롤러(216)에 의해 실행될 수 있다. 각 블록은 별도의 프로세스 또는 실행 스레드를 나타낸다.
- [0158] 컨트롤러 프로세스
- [0159] 제어 프로세스는 다른 프로세스 또는 스레드를 시작 및 중지하도록 구성될 수 있다. 제어 프로세스는 또한 획득 시스템(700)에 사용자 인터페이스를 제공하도록 구성될 수 있다. 제어 프로세스는 활동의 로그를 저장하고 획득 동안 발생한 에러(예컨대, 로그)를 추적하도록 구성될 수 있다. 제어 프로세스는 또한 비활성화된 프로세서 또는 스레드를 재시작 하도록 구성될 수 있다.
- [0160] 구성요소들 간의 통신 방법은 시스템 운영 체제가 선택된 이후 결정될 수 있다. 제어 프로세스에 대한 사용자 인터페이스는 명령 행 인터페이스이거나 그래픽 인터페이스일 수 있다. 그래픽 인터페이스는 운영체제로부터의 독립성을 제공하는 QT와 같은, 휴대용 프레임워크에 구축될 수 있다.
- [0161] 비디오 획득 프로세스
- [0162] 비디오 획득 프로세스는 카메라(210)와 직접 통신하도록 구성될 수 있으며, 타임 스탬프 된 프레임을 데이터 저장 장치(220)에 저장한다. 비디오 획득 프로세스는 높은 우선순위로 실행될 수 있으며, 프레임 손실 가능성을 최소화한다. 비디오 획득 프로세스는, 프레임 사이의 처리를 최소화하면서, 상대적으로 간단하게 유지될 수 있다. 비디오 획득 프로세스는 또한 최소의 효과적인 셔터 속도로 적절한 노출을 보장하기 위하여 발광체(212)에 의해 방출된 적외선 조명을 제어하도록 구성될 수 있다.
- [0163] 오디오 획득 프로세스
- [0164] 별도의 오디오 획득 프로세스가 적절한 타임 스탬프를 갖는 초음파 오디오 데이터를 획득하도록 구성될 수 있다. 일 실시예에서, 오디오 시스템은 인클로저(204)와 오디오 통신하도록 배치된 마이크로폰(222) 어레이를 포함할 수 있다. 특정 실시예들에서, 적어도 하나의 마이크로폰(222)은 인클로저(204) 내에 배치될 수 있다. 마이크로폰 어레이의 마이크로폰은 다음의 능력 중 적어도 하나를 포함할 수 있다: 약 500kHz의 샘플링 주파수, 약 16비트의 ADC 해상도, 약 10kHz 내지 약 210kHz의 주파수 범위, 및 210kHz의 8차 안티 앨리어싱 필터. 일 예로서, 마이크로폰 어레이의 마이크로폰은 Pettersson M550 마이크로폰(Pettersson Elektronik AB, Upsala Sweden) 또는 이의 기능적 등가물을 포함할 수 있다. 위에서 논의된 바와 같이, 마이크로폰(222)에 의해 캡처된 오디오 데이터는 타임 스탬프되고, 분석을 위해 컨트롤러(216)에 및/또는 저장을 위해 데이터 저장 장치(220)에 제공될 수 있다.
- [0165] 환경 데이터 획득 프로세스
- [0166] 별도의 환경 데이터 획득 프로세스가, 온도, 습도, 및 광 세기와 같은, 환경 데이터를 수집하도록 구성될 수 있다. 환경 데이터는 저주파수(예컨대, 약 0.01Hz - 0.1Hz)로 수집될 수 있다. 환경 데이터는 데이터 저장 장치(220)에 의해 레코드마다 타임스탬프와 함께 (예컨대, 적어도 하나의 CSV 파일로서) 저장될 수 있다.
- [0167] 조명 제어 프로세스

- [0168] 조명 제어 프로세스는 쥐에게 주간/야간 순환을 제공하기 위해 발광체(212)에 의해 방출된 가시광선을 제어하도록 구성될 수 있다. 일 실시예에서, 위에서 논의된 바와 같이, 적외선에만 응답하는, 카메라(210)는 실질적으로 모든 가시광선을 필터링하도록 구성될 수 있고, 가시광선은 적외선을 생성하지 않도록 필터링 될 수 있어서, 이 프로세스는 비디오 캡처에 영향을 미치지 않을 수 있다.
- [0169] *비디오 편집 프로세스*
- [0170] 비디오 편집 프로세스는 획득된 비디오 데이터를 미리 결정된 압축으로 미리 결정된 포맷으로 재-패키징(re-package)하도록 구성될 수 있다. 이 프로세스는 프레임을 손실할 가능성을 최소화하기 위해 비디오 획득과 분리될 수 있다. 이 비디오 편집 프로세스는 우선순위가 낮은 백그라운드 작업으로 또는 데이터 획득이 완료된 이후에 실행될 수 있다.
- [0171] *와치독 프로세스*
- [0172] 와치독 프로세스는 데이터 획득 프로세스의 상태를 모니터링하도록 구성될 수 있다. 일 예로서, 그것은 문제를 (예컨대, 로그로) 기록하고 필요에 따라 재시작을 트리거할 수 있다. 와치독 프로세스는 감시 중인 구성요소로부터 "하트 비트(heartbeat)"를 들을 수도 있다. 일반적으로, 하트 비트는 시스템(700)의 구성요소가 정상적으로 작동하고 있음을 확인하는 컨트롤러(216)로 전송된 신호일 수 있다. 일 세로서, 시스템(700)의 구성요소가 작동을 멈추면, 이 구성요소에 의해 전송된 하트 비트의 부족이 컨트롤러(216)에 의해 검출될 수 있다. 이 검출에 이어서, 컨트롤러(216)는 이벤트를 기록하고 경보가 표시되도록 할 수 있다. 이러한 경보는 오디오 경보 및 시각적 경보(예컨대, 조명, 수문자 표시 등)를 포함할 수 있지만, 이에 제한되지 않는다. 그러한 경보에 대한적으로 또는 부가하여, 컨트롤러(216)는 재초기화 신호 또는 사이클링 파워를 전송하는 것과 같은, 구성요소의 동작을 재시작하려는 시도를 할 수 있다. 시스템(700)의 구성요소들과 컨트롤러(216) 사이의 통신 방법은 운영체제의 선택에 의존할 수 있다.
- [0173] *쥐 마킹(marking)*
- [0174] 특정 실시예들에서, 쥐는 추적을 용이하기 하기 위해 마킹될 수 있다. 그러나, 아래에서 더 상세히 논의되는 바와 같이, 마킹을 생략될 수 있고, 추적은 다른 기술에 의해 용이하게 될 수 있다.
- [0175] 시각적 식별을 위해 쥐를 마킹하는 것은 해결하기 쉽지 않은 여러 파라미터가 있다. 일 실시예에서, 마킹은 쥐 자체에게는 보이지 않는 방식으로 장기간(몇 주)으로 쥐에 대해 수행될 수 있고, 따라서 쥐 의사소통 및 행동에 대한 영향을 최소화한다. 일 예로서, 정상적인 쥐의 관찰 범위에서 보이지 않는 장기간의 적외선 민감성 마커가 사용될 수 있다.
- [0176] 대안적인 실시예에서, 쥐 털은 사람의 모발 색상 및 모발 표백제를 이용하여 마킹될 수 있다. 이 접근법은 몇 주 동안 쥐를 명확하게 식별할 수 있고 행동 실험에 성공적으로 이용될 수 있다(예: Ohayon 외, 2013 참조). 그러나, 모발 마킹 공정은 본 쥐 모니터링 시스템에 허용되지 않는 공정인 쥐의 마취를 필요로 한다. 마취는 생리학을 변화시키고, 염색약 자체는 종종 쥐의 행동을 변화시키는 자극제일 수 있다. 각각의 DO 쥐는 독특하기 때문에, 이것은 염료/마취 x 유전자형 효과를 초래하고 알려지지 않은 변수(들)을 도입할 수 있다.
- [0177] 적외선 염료 기반의 마커를 이용하는 추가적인 대안적인 방법들 또한 사용되고 최적화될 수 있다.
- [0178] 추가적인 실시예에서, 털 면도는 마킹의 형태로서 쥐의 뒷면에 패턴을 생성하기 위해 사용될 수 있다.
- [0179] *데이터 저장*
- [0180] 개발 단계 동안, 총 2TB 미만의 데이터가 요구될 수 있다. 이들 데이터는 다양한 카메라 및 압축 방법의 샘플로(raw) 및 압축된 비디오를 포함할 수 있다. 따라서, 스트레스 테스트 동안 통합된 USV 및 비디오 데이터의 데이터 전송뿐만 아니라 장기간 7-10 일 길이의 비디오 데이터가 달성될 수 있다. 비디오의 크기는 선택된 압축 표준에 따라 감소될 수 있다. 샘플 데이터 저장 추정치는 아래에 제공된다.
- [0181] 테스팅:
- [0182] 1 아레나
- [0183] 최대 5 카메라
- [0184] 비디오의 기간: 각각 ~1-2시간

- [0185] 총 ~10GB, 하이 엔드.
- [0186] 스트레스 테스트:
- [0187] 1 아레나
- [0188] 1 카메라
- [0189] 비디오의 기간: 14일
- [0190] 해상도: 2x (960x960)
- [0191] 총 ~2TB
- [0192] 생산:
- [0193] 120 총 실행 (12-16 아레나, 그룹 실행 당 80마리의 동물, 엇갈린 실험)
- [0194] 기간 (각) 7일
- [0195] 해상도: 2x (960x960)
- [0196] 32.25TB
- [0197] **II. 동물 추적**
- [0198] 쥐와 같은, 동물의 비디오 추적은 복잡하고 동적인 환경에서 또는 종래의 동물 모니터링 시스템에서 유전자 이종 동물과 함께 수행될 수 없으며, 사용자의 높은 수준의 참여없이 대규모 실험을 수행할 수 없다. 아래에서 논의되는 바와 같이, 종래의 시스템 및 방법을 이용하여 여러 환경에서 다수의 상이한 쥐 품종을 추적하고 이러한 시스템 및 방법이 대규모 실험의 데이터 세트에 적합하지 않음을 입증하려고 시도한다.
- [0199] 검정색, 야구티, 알비노, 회색, 갈색, 누드, 및 얼룩무늬를 포함하는 다양한 털 색상의 쥐를 포함하는 예시적인 데이터 세트가 분석에 사용되었다. 모든 동물은, 아래에 요약된, JAX-IACUC로부터 승인된 프로토콜에 따라 테스트되었다. 생후 8주에서 14주 사이의 쥐가 실험되었다. 데이터 세트는 총 1702시간의 59개 품종에 대한 1857개 비디오를 포함한다.
- [0200] 모든 동물은 Jackson Laboratory 생산 콜로니로부터 얻었다. 생후 8 내지 14주의 성인 쥐는 Jackson Laboratory Institutional Animal Care and Use Committee의 가이드라인으로부터 승인된 프로토콜에 따라 행동으로 테스트되었다. 오픈 필드 행동 분석은 Kumar, 2011에 설명된 대로 수행되었다. 요컨대, 그룹-보관된 쥐는 비디오 녹화 시작 전에 30-45분 동안 시험실에서 칭량되고(weigh) 순응된다. 첫 55분 동안 활동한 데이터가 여기에 표시된다. 이용 가능한 경우, 8마리의 수컷 및 8마리의 암컷이 각각의 근친계(inbred strain) 및 F1 등질계(isogenic strain)로부터 테스트되었다.
- [0201] 일 측면에서, 흰색 배경을 갖는 동일한 오픈 필드 장치(예컨대, 아레나(200))에서 다수의 동물을 추적하는 것이 바람직할 수 있다. 비디오 획득 시스템에 의해 획득된 전체 프레임 및 잘라낸 비디오 이미지의 예들이 도 8A, 1행(전체 프레임) 및 2행(잘라낸 프레임)에 도시되어 있다. 이상적인 및 실제 추적 프레임의 예들은 다양한 유적 배경에 대해 각 환경에 표시된다(도 8A, 3행[이상적인 추적] 및 4행[실제 추적]).
- [0202] 다른 측면에서, 음식과 물컵이 포함된 아레나(200)의 실시예 및 Jackson Laboratory에서 Knockout Mouse Project (KOMP2)와 같은, 도전적인 환경에서 행동의 비디오 분석을 수행하는 것이 바람직하다(각각, 도 8, 5열 및 6열).
- [0203] 24시간 장치에서, 쥐는 백지 잠자리(bedding) 및 음식/물컵과 함께 아레나(200)에 수용되었다. 쥐를 아레나(200)에 유지하고, 발광체(212)에 의해 방출된 적외선을 이용하여 밝고 어두운 조건에서 연속적인 녹화가 수행되었다. 잠자리 및 음식 컵은 쥐에 의해 이동되고 발광체(212)에 의해 방출된 가시광선은 명/암 순환을 모의하기 위해 매일의 경과에 따라 변화되었다.
- [0204] KOMP2 프로젝트는 5년에 걸친 데이터를 수집하였고, 빔 브레이크 시스템으로는 식별될 수 없는 보행 영향을 검출하기 위해 추가 분석 양식으로서 비디오 기반 녹화를 수행하는 것이 요구되었다. 보행 분석에서, 동물의 움직임이 분석된다. 동물이 비정상적인 보행을 갖는 경우, 굴격, 근육 및 또는 신경과 같은 이상(abnormality)이 유도될 수 있다. KOMP2 프로젝트는 쥐가 모든 면에 적외선 빔이 있는 투명한 폴리카보네이트 상자에 위치하는 빔 브레이크 시스템을 이용한다. 매트릭스의 바닥은 역시 폴리카보네이트이며, 기본 벤치 표면은 어두운 회색이다.



두 테이블의 교차점에 배치된 일부 상자는, 조인트를 남길 수 있으며, 조명 오버헤드(예컨대, LED 광)은 모든 상자에서 고유한 눈부심을 유발할 수 있다.

- [0205] 일 측면에서, 비디오의 이 데이터 세트를 추적하는 것은 백그라운드 감산 및 블롭(blob) 검출 휴리스틱스를 이용하는 현대적인 오픈 소스 추적 툴인 Ctrax를 이용하여 시도되었다. Ctrax는 5개의 메트릭, 장축 및 단축, 위 중심의 x 및 y 위치, 및 동물의 방향으로 프레임마다 쥐를 추상화한다(Branson, 2009). 이 소프트웨어는 MOG2 백그라운드 감산 모델을 사용한다. 여기서 소프트웨어는 백그라운드 감산에 사용하기 위해 비디오 백그라운드의 평균 및 변동을 모두 추정한다. Ctrax는 타원을 피팅하기 위해 예측된 전경의 모양을 이용한다.
- [0206] 다른 측면에서, 이 비디오 데이터 세트를 추적하는 것은 독점적 추적 알고리즘을 이용하는 상업적으로 이용 가능한 추적 소프트웨어인 LimeLight를 이용하여 시도되었다. LimeLight는 분류 및 감지를 위해 단일 키 프레임 배경 모델을 이용한다. 쥐가 감지되면, LimeLight는 독점 알고리즘을 이용하여 쥐를 질량 중심으로 추상화한다.
- [0207] 데이터 세트는 이러한 종래의 분석 시스템에 대한 중요한 도전을 포함한다. 예를 들어, 쥐 털 색상과 환경의 조합은 Ctrax와 LimeLight로 다루기가 어려웠다. 일반적으로, 흰색 배경에서 어두운 쥐(예컨대, 검정색, 아고티)와 같이 대비가 높은 환경에서는 추적 결과가 양호하다. 그러나 흰색 배경에서 밝은 색상의 쥐(예컨대, 알비노, 회색 또는 얼룩무늬 쥐)와 같이 대비가 적은 환경에서는 결과가 좋지 않다. 흰색 오픈 필드의 검정색 쥐는 높은 전경-배경 대비를 달성하므로 실제 추적은 이상적인 것과 거의 일치한다. 회색 쥐는 아레나 벽과 시각적으로 유사하므로 종종 벽에서 양육하는 동안 코를 제거한다. 알비노 쥐는 아레나 자체의 배경과 유사하며 추적 중에 종종 발견되지 않는다. 얼룩무늬 쥐는 패턴화된 털 색상으로 인해 절반으로 절단된다. 각 비디오에 대해 Ctrax를 최적화하고 미세 조정하려는 시도가 있었지만, 3행(이상적인 추적)과 비교할 때 도 8A, 4행(실제 추적)의 실제 추적 결과에 표시된 바와 같이, 여전히 많은 불량 추적 프레임이 관찰된다. 불량 추적 프레임을 폐기하면 치우친 샘플링 및 생물학적 해석이 왜곡될 수 있으므로 바람직하지 않다.
- [0208] 24시간 및 KOMP2 환경과 같은, 추적에 덜 이상적인 환경 때문에 이들 예러가 증가하는 것으로 관찰되었다. 더욱이, 예러 분포는 무작위가 아니었다. 예를 들어, 쥐가 도 8, 4행(실제 추적)에 도시된 바와 같이, 모서리, 벽 근처, 또는 음식 컵 위에 있을 때 추적이 매우 부정확하지만 동물이 중앙에 있을 때는 덜 부정확한 것으로 밝혀졌다. 24시간 환경에서 음식 컵을 아레나에 위치시키는 것은 쥐가 올라갈 때 추적 문제를 발생시킨다. KOMP2와 같이, 반사 표면이 있는 아레나에서도 추적 알고리즘에 예러가 발생한다.
- [0209] 불량 추적의 원인에 대한 추가적인 조사에서, 대부분의 경우, 부적절한 추정은 배경으로부터 쥐의 불량한 분할 때문인 것으로 밝혀졌다. 이것은 대비가 좋지 않아 쥐를 전경에서 제거했거나 배경이 전경에 포함된 경우를 포함한다. 전통적으로, 이러한 장애물 중 일부는 최적화된 비디오 데이터 수집을 위한 환경을 변경함으로써 해결되었다. 예를 들어, 알비노 쥐를 추적하기 위해, 오픈 필드의 배경 색을 검정색으로 변경하고 대비를 증가시킬 수 있다. 그러나, 환경의 색상이 쥐 및 인간의 행동에 영향을 미치고 그러한 조작이 실험 결과를 잠재적으로 혼란시킬 수 있기 때문에(Valdez, 1994; Kuleshkaya, 2014), 이러한 환경 변화는 현재 상황에 적합하지 않다. 또한, 이러한 솔루션은 24시간 데이터 수집 시스템 또는 KOMP2 아레나에서 얼룩무늬 쥐에 대해 작동하지 않을 수 있다.
- [0210] Ctrax가 단일 배경 모델 알고리즘을 사용하지 때문에, 다른 배경 모델이 추적 결과를 개선할 수 있는지 여부를 결정하기 위한 테스트가 수행되었다. 26개의 상이한 분할 알고리즘(Sobral, 2013)이 테스트되었으며, 도 8B에 도시된 것과 같이, 이러한 전통적인 알고리즘 각각은 특정 상황에서 잘 작동하고 다른 곳에서는 실패하는 것이 발견되었다. 추적을 위한 배경 감산 접근법에 의존하는 것을 포함하여 동물 추적에 사용 가능한 다른 시스템들 및 방법. 26개의 배경 감산 방법들이 모두 실패했기 때문에, Ctrax 및 LimeLight의 결과는 이러한 다른 기술들을 대표한다고 생각된다. 이러한 분할 알고리즘은 부적절한 분할로 인해 실패한다고 생각된다.
- [0211] 따라서, 비디오 데이터의 분석을 위한 많은 추적 솔루션이 존재하지만, 종래 솔루션의 대표적인 예를 갖는 고충실도 쥐 추적을 달성하기 위해 적절한 쥐 분할의 근본적인 문제를 극복하기 위한 시도는 성공하지 못했다. 쥐 분할의 근본적인 문제를 적절하게 해결하는 것은 없으며, 일반적으로 적절한 분할을 달성하기 위해 환경 최적화에 의존하므로 잠재적인 혼란을 일으킨다.
- [0212] 또한, 배경 감산 알고리즘 파라미터를 미세 튜닝하기 위한 시간 비용이 엄청날 수 있다. 예를 들어, 24시간 설정으로부터 데이터를 추적할 때, 쥐가 한 자세로 장시간 자고 있을 때, 쥐는 배경 모델의 일부가 되고 추적될 수 없었다. 일반적인 감독은 고품질의 추적 결과를 얻기 위해 숙련된 사용자에게 비디오의 각 시간 마다 5분의 상호작용이 필요하다. 소규모 및 제한된 실험에서는 이 수준의 사용자 상호작용이 다루기 쉽지만, 대규모 및 장

기 실험에서는 추적 성능을 감독하기 위해 많은 시간이 소요된다.

[0213] 본 개시의 실시예들은 이러한 어려움을 극복하고 쥐와 같은, 동물을 포함하는 비디오 데이터의 분석에 적합한 강력한 자세대 추적기를 구축한다. 아래에서 상세히 논의되는 바와 같이, 복잡하고 동적인 환경 조건 하에서 고성능을 달성하고, 털 색상의 유전학에 무관하여, 사용자에 의한 지속적인 미세 조정이 필요하지 않은 인공지능경망이 사용된다.

[0214] 컨볼루션 신경망은 여러 레벨의 추상화로 데이터 표현을 학습하는 다수의 프로세싱 레이어를 포함하는 계산 모델이다. 이들 방법들은 음성 인식, 시각적 객체 인식, 객체 탐지, 및 약물 발견 및 유전체학과 같은 다른 영역에서 최첨단 기술을 극적으로 향상시켰다(LeCun, 2015). 한 가지 장점으로, 적합한 하이퍼파라미터를 갖춘 효율적인 네트워크가 개발되면, 신경망은 단순히 적절한 훈련 데이터를 추가함으로써 다른 작업으로 쉽게 확장될 수 있다. 따라서, 개시된 실시예들은 쥐 추적을 위한 고도로 일반화 가능한 솔루션을 제공한다.

#### [0215] 신경망 아키텍처

[0216] 시각적 추적 문제를 해결하기 위해 3개의 1차 네트워크 아키텍처가 개발되었다. 일 실시예에서, 객체 추적은, 도 8C에 도시된 바와 같이, 분할 마스크에 기초하여 쥐의 타원 묘사의 형태를 취할 수 있다. Branson 2005 참조. 대안적인 실시예들에서, 타원 이외의 형상이 사용될 수 있다.

[0217] 선택적으로, 필요한 경우, 좌표를 결정하는 것을 돕기 위해 랜드마크(예컨대, 인클로저(204)의 코너)가 비디오 프레임에서 검출될 수 있다. 다른 측면에서, 변수들은 쥐의 장축 길이 및 단축 길이, 및 장축의 벡터 각도의 사인 및 코사인을 더 포함할 수 있다. 각도는 장축의 방향에 대해 정의될 수 있다. 장축은 비디오 프레임의 좌표계에서 동물의 머리 끝(예컨대, 코)으로부터 동물의 몸 끝(예를 들어, 동물의 꼬리가 몸에서 연장되는 지점)까지 연장될 수 있다. 명확성을 위해, 잘라낸 프레임이 신경망에 입력되는 것으로 도시되어 있지만 실제 입력은 표시되지 않은 전체 프레임이다.

[0218] 신경망 아키텍처를 이용하는 타원 파라미터들을 결정하기 위한 예시적인 시스템들 및 방법들이 아래에서 상세히 논의된다. 다른 파라미터들이, 적절하게, 개시된 실시예들에 의해 이용되고 결정될 수 있음을 이해할 수 있다.

[0219] 일 실시예에서, 첫 번째 아키텍처는 인코더-디코더 분할 네트워크이다. 도 9에 도시된 바와 같이, 이 네트워크는 주어진 입력 프레임으로부터 전경-배경 분할된 이미지를 예측하고, 쥐가 있는지 없는지를 픽셀 단위로 예측할 수 있으며, 출력은 분할 마스크이다.

[0220] 이 첫 번째 아키텍처는 입력을 작은 공간 해상도 세트의 특징(예컨대, 480x480과 대조적으로 5x5)으로 추상화하도록 구성된 특징 인코더를 포함한다. 학습하기 위해 많은 수의 파라미터가 신경망에 할당된다. 학습은 지도학습에 의해 수행될 수 있으며, 여기서 신경망에는 예가 제시되고 파라미터를 튜닝하여 정확한 예측이 이루어진다. 최종 모델 정의 및 훈련 하이퍼파라미터에 대한 자세한 설명은 아래의 표 3에 나타났다.

[0221] 표 3 - 훈련 하이퍼파라미터

모델	파라미터	값
모두	Rotation	$\pm 2.5$ deg.
	Translation	$\pm 5.0$ px.
	Additive Noise	$\mu=0.0, \sigma=5.0$
	Brightness	$\pm 5\%$
	Contrast	$\pm 5\%$
인코더-디코더 분할 네트워크	Learning Rate	0.00001
	Batch Size	50
회귀 네트워크	Learning Rate	0.00001
	Batch Size	5
비닝된 분류 네트워크	Learning Rate	0.001
	Batch Size	25

[0222]

[0223] 특징 인코더 다음에 특징의 작은 공간 해상도 세트를 원래의 입력 이미지와 동일한 형상으로 다시 변환하도록 구성된 특징 디코더가 뒤따른다. 즉, 신경망에서 학습된 파라미터는 특징 인코딩 동작을 반대로 한다.

- [0224] 타원이 향하고 있는 기본 방향을 예측하기 위해 3개의 완전연결층(fully connected layer)이 인코딩된 특징에 추가된다. 완전연결층은 신경망 층을 지칭할 수 있으며, 여기서 주어진 층의 각 숫자에는 다른 매개변수(예컨대, 학습 가능한 매개변수)가 곱해지고 함께 합쳐져 새로운 층에서 단일 값을 생성한다. 이 특징 디코더는 전경-배경 분할된 이미지를 생성하도록 훈련될 수 있다.
- [0225] 네트워크(인코더)의 전반부는 2D 컨볼루션층, 이어서 배치 정규화, ReLu 활성화, 및 2D 맥스 풀링 층을 이용한다. 자세한 내용은 Goodfellow, 2016에서 확인할 수 있다.
- [0226] 모든 풀링 층이 사용된 후 두 배가 되는 8의 시작 필터 크기가 사용된다. 사용된 커널은 2D 컨볼루션 층의 경우 5x5이고, 맥스 풀링 층의 경우 2x2이다. 입력 비디오의 형상은 480x480x1(예컨대, 단색)이며, 이러한 반복되는 층 6개 이후에 결과 형상은 15x15x128(예컨대, 128색)이다.
- [0227] 대안적인 실시예들에서, 3x3과 같은 다른 형상의 풀링 층이 사용될 수 있다. 반복되는 층은 층의 반복되는 구조를 나타낸다. 신경망은 각 층에 대해 상이한 매개변수를 학습하고, 층들이 적층된다. 6개의 반복된 층이 위에서 논의되었지만, 더 많거나 더 적은 수의 반복된 층이 사용될 수 있다.
- [0228] 또 다른 2D 컨볼루션 층(커널 5x5, 2x 필터)이 적용되고 이어서 3x3의 다른 커널 및 3의 보폭을 갖는 2D 맥스 풀이 적용된다. 15x15 공간 형상은 3의 인자를 사용하여 더 많이 줄어들 수 있다. 전형적인 맥스 풀은 커널 2x2 보폭 2이며, 각 2x2 그리드는 최대값을 선택하고 1개의 값을 생성한다. 이들 설정은 3x3 그리드에서 최대값을 선택한다.
- [0229] 마지막 2D 컨볼루션 층은 5x5x512의 형상을 갖는 특징 병목 현상을 생성하기 위해 적용된다. 특징 병목 현상은 인코딩된 특징 세트, 모든 행렬 연산으로부터 출력된 실제 행렬 값을 지칭한다. 학습 알고리즘은 인코딩된 특징 세트가 제대로 수행되도록 훈련된 작업에 가장 인코딩된 특징 세트를 최적화한다. 이 기능 병목 현상은 분할 디코더 및 각도 예측기 모두에 전달된다.
- [0230] 분할 디코더는 계단형 트랜스포즈 2D 컨볼루션 층을 사용하여 인코더를 역전시키고 합산 접합을 통해 미리 다운샘플링된 활성화를 전달한다. 이 디코더는 ReLu 활성화를 사용하지 않는 점에 유의해야 한다. 미리 다운샘플링된 활성화 및 합산 접합은 스킵 연결이라고도 한다. 디코딩하는 동안 층에서의 특징들은 인코더 층과 동일한 형태와 일치하기 때문에, 네트워크는 더 나은 것을 인코딩하거나 인코딩 상태 동안의 것을 유지하는 것을 선택할 수 있다.
- [0231] 층들이 480x480x8 형상으로 되돌아간 후, 깊이를 2개의 흑백 이미지(배경 예측 및 전경 예측)로 병합하기 위해 1x1의 커널 크기로 추가 컨볼루션이 적용된다. 최종 출력은 480x480x2(2 색)이다. 제1 색은 배경을 나타내도록 지정된다. 제2 색은 전경을 나타내도록 지정된다. 각 픽셀에 따라, 둘 중 더 큰 것은 네트워크가 입력 픽셀이라고 생각하는 것이다. 아래에서 논의되는 것처럼, Softmax 연산은 누적 확률이 1이 되도록 이러한 색상의 배율을 다시 조정한다.
- [0232] 이어서, Softmax가 이 깊이에 걸쳐 적용된다. Softmax는 그룹으로의 분류 또는 binmin의 형태이다. Softmax에 관한 추가 정보는 Goodfellow, 2016에서 확인할 수 있다.
- [0233] 특징 병목 현상으로부터, 각도 예측도 생성된다. 이는 배치 정규화 및 ReLu 활성화(커널 크기 5x5, 특징 깊이 128 및 64)로 2개의 2D 컨볼루션 층을 적용함으로써 달성된다. 여기에서, 하나의 완전연결층이 평평해지고 4개의 뉴런 모양을 만들어 쥐의 머리가 향하는 사분면을 예측하는 데 사용된다. 배치 정규화, ReLu 활성화, 및 평탄화에 관한 자세한 내용은 Goodfellow, 2016에서 확인할 수 있다.
- [0234] 각도는 분할 마스크에 의해 예측되기 때문에, 정확한 방향( $\pm 180^\circ$ )만 선택될 필요가 있다. 즉, 타원이 예측되지 때문에, 하나의 장축만 있다. 장축의 한쪽 끝은 쥐의 머리 방향이다. 쥐는 머리-꼬리 축을 따라 더 길다고 가능하다. 따라서, 한 방향은  $+180^\circ$  (머리)이고 다른 방향은  $-180^\circ$  (꼬리)이다. 인코더-디코더 신경망 아키텍처가 선택할 수 있는 4개의 가능한 방향은 극좌표 그리드에서 45-135, 135-225, 225-315, 및 315-45도이다.
- [0235] 이들 경계들은 각도 예측에서 불연속을 피하기 위해 선택되었다. 특히, 위에서 논의된 바와 같이, 각도 예측은 장축의 벡터 각도의 사인 및 코사인의 예측이며, atan2 함수를 사용한다. atan2 함수에는 ( $180^\circ$  에서) 불연속성이 있으며 선택된 경계는 이러한 불연속성을 피한다.
- [0236] 네트워크가 분할 마스크를 생성할 수, 타원 피팅 알고리즘이 Branson, 2009에 설명된 바와 같이 추적을 위해 적용될 수 있다. Branson은 이들 계산에 가중된 샘플 평균 및 분산을 사용하지만, 분할 신경망은 개선을 설명하는 상황에 대한 불변성을 유지한다. 캐스트되는 그림자는 배경 감산 알고리즘에 의해 생성된 분류 마스크에 에러를

추가할 수 있다. 신경망은 이러한 문제가 전혀 없다는 것을 학습한다. 추가적으로, 가중 및 비가중 샘플 수단과 분산의 사용 간에 유의미한 차이가 관찰되지 않는다. 가중 및 비가중 접근법으로부터 예측된 타원 피팅 파라미터는 개시된 신경망의 실시예들에 의해 예측된 마스크를 사용하여 크게 다르지 않다.

[0237] 분할 마스크가 주어지면 픽셀 위치의 샘플 평균이 중심 위치를 나타내도록 계산된다.

$$\mu_{x,y} = \frac{1}{N} \sum_i^N p_i \quad (1)$$

[0238] 유사하게, 픽셀 위치의 샘플 분산은 장축 길이(a), 단축 길이(b), 및 각도( $\theta$ )를 나타내도록 계산된다.

$$\sigma = \frac{1}{N} \sum_i^N (p - \mu_{x,y})(p - \mu_{x,y})^T \quad (2)$$

[0241] 축 길이와 각도를 얻기 위해, 고유값(eigenvalue) 분해 방정식을 풀어야 한다.

$$\sigma = U^T D U, \quad U = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \quad D = \begin{pmatrix} \frac{a}{2} & 0 \\ 0 & \frac{b}{2} \end{pmatrix}^2 \quad (3)$$

$$a = 2\sqrt{D_{11}}, \quad b = 2\sqrt{D_{22}}, \quad \theta = \text{atan}(U_{12}, U_{21}) \quad (4)$$

[0244] 두 번째 네트워크 아키텍처는 비닝된 분류 네트워크이다. 비닝된 분류 네트워크 아키텍처의 구조는, 도 10에 도시된 바와 같이, 각 타원 피팅 파라미터에 대해 가장 가능한 값의 열 지도를 예측할 수 있다.

[0245] 이 네트워크 아키텍처는 입력 이미지를 작은 공간 해상도로 추상화하는 특징 인코더로 시작한다. 회귀 예측의 대부분은 경계 상자(예컨대, 정사각형 또는 직사각형)를 통해 솔루션을 인식하지만 타원은 단순히 하나의 추가 파라미터인 각도를 추가한다. 각도는  $360^\circ$  와  $0^\circ$  에서 등가의 반복 시리즈이므로 각도 파라미터는 사인 및 코사인 성분으로 변환된다. 이로 인해 네트워크에서 총 6개의 파라미터가 회귀된다. 이 네트워크의 전반부는 문제 해결과 관련된 일련의 특징을 인코딩한다.

[0246] 인코딩된 특징들은 특징을 나타내는 행렬(어레이)을 단일 벡터로 변환함으로써 평탄화된다. 평탄화된 인코딩된 특징들은 그 후 출력의 원하는 해상도에 의해 결정되는 추가적인 완전연결층에 연결된다. 예를 들어, 쥐의 X 좌표 위치의 경우,  $480 \times 480$  픽셀 이미지의 각 x 열마다 한 개의 빈(bin)씩, 480개의 빈이 있다.

[0247] 네트워크가 실행될 때, 각각의 열 지도에서 가장 큰 값이 가장 가능성 높은 값으로 선택된다. 각각의 원하는 출력 파라미터는 인코딩된 특징들에 연결된 훈련 가능한 완전연결층들의 독립적인 세트로서 실행될 수 있다.

[0248] Resnet V2 50, Resnet V2 101, Resnet V2 200, Inception V3, Inception V4, VGG, 및 Alexnet을 포함하여, 다양한 사전 구축된 특징 검출기가 테스트되었다. 특징 검출기는 입력 이미지에 작동하는 컨볼루션을 나타낸다. 이러한 사전 구축된 특징 검출기에 더해, 맞춤 네트워크의 다양한 어레이도 조사되었다. Resnet V2 200이 이 조사에서 가장 잘 수행하는 것으로 관찰되었다.

[0249] 마지막 아키텍처는, 도 11에 도시된, 회귀 네트워크이다. 일 예로서, 그것은 입력 비디오 프레임을 가져오고, Resnet200 CNN을 통해 특징을 추출하고 타원 피팅을 위한 6개의 파라미터들을 직접 예측한다. 연속적인 각 값(타원 피팅을 위한 6개)은 무한대의 범위를 가질 수 있다. 네트워크는 적절한 값의 범위를 학습해야 한다. 이런 방식으로, 추적 타원을 기술하기 위해 숫자로 나타낸 타원 값들이 입력 이미지로부터 직접 예측된다. 즉, 파라미터들을 직접 예측하는 대신에, 회귀 네트워크는 비닝 가능한 값의 선택으로부터 가장 가능성 높은 값을 선택한다.

[0250] 다른 신경망 아키텍처들은 다른 방식으로 동작한다. 인코더-디코더 신경망 아키텍처는 각 픽셀이 쥐이거나 쥐가 아닐 확률을 출력한다. 비닝된 분류 신경망 아키텍처는 쥐의 위치를 설명하는 빈(bin)을 출력한다. 각 파라미터에 대한 클래스는 미리 결정되어 있으며 (인코더-디코더 또는 비닝된) 네트워크는 각 클래스에 대한 확률을 간



단히 출력해야 한다.

- [0251] 회귀 네트워크 아키텍처는 입력을 작은 공간 해상도로 추상화하는 특징 인코더로 시작한다. 이전의 아키텍처와 대조적으로, 회귀 신경망 네트워크 훈련은, 평균 제곱 오차 손실 함수와 대조적으로, 교차 엔트로피 손실 함수에 의존한다.
- [0252] 메모리 제한으로 인해, 커스텀 VGG-유사 네트워크만이 감소된 특징 차원으로 테스트되었다. 최고 성능의 네트워크는 2개의 2D 컨볼루션 층에 이어 2D 맥스 풀링 층으로 구성되었다. 사용된 커널은 2D 컨볼루션 층의 경우 3x3의 형상이고 2D 맥스 풀링 층의 경우 2x2의 형상이다. 이 2개의 컨볼루션 더하기 맥스 풀 시퀀스는 15x15x256의 형상을 만들기 위해 5회 반복된다.
- [0253] 이 층은 평탄화되고 각 출력에 대해 완전연결층에 연결된다. 각 출력의 형상은 원하는 해상도와 예측 범위에 의해 결정된다. d일 예로서, 이들 인코딩된 특징들은 6의 출력 형상을 생성하기 위해 평탄화되고 완전연결층에 연결되는데, 이는 네트워크가 타원에 피팅하도록 예측하도록 요구되는 값의 수이다. 테스트 목적으로, 중앙 위치만 관찰되고 전체 이미지 범위(0-480)로 훈련되었다. 각도 예측과 같은, 추가적인 출력은 추가 출력 벡터로서 쉽게 추가될 수 있다. 다양한 최신 기능 인코더가 테스트 되었고, 이 네트워크에 대해 여기에서 논의되는 데이터는 이 아키텍처에 대해 최상의 성능을 달성한(He, 2016) 200개의 층이 있는 Resnet V2로부터 가져왔다.
- [0254] **훈련 데이터 세트**
- [0255] 네트워크 아키텍처를 테스트하기 위해, 다수의 스트레인 및 환경에 걸쳐 16,234개의 훈련 이미지 및 568개의 개별 검증 이미지의 훈련 데이터 세트가, 아래에 논의된 바와 같이, OpenCV 기반 라벨링(labeling) 인터페이스를 사용하여 생성되었다. 이 라벨링 인터페이스는 타원 피팅뿐만 아니라 전경 및 배경의 빠른 라벨링을 허용하며, 전송 학습을 통해 네트워크를 새로운 실험 조건에 적응시키기 위해 훈련 데이터를 신속하게 생성하는 데 사용될 수 있다.
- [0256] 대화형 워터셰드(watershed) 기반의 분할 및 윤곽 기반 타원 피팅을 생성하기 위해 OpenCV 라이브러리가 사용되었다. 이 소프트웨어를 사용하여, 사용자는, 도 12A에 도시된 바와 같이, 포인트를 전경(예컨대, 쥐; F)으로 마킹하기 위해 좌클릭을 하고 다른 포인트를 배경(B)으로 마킹하기 위해 우클릭을 한다. 키스트로크 시, 워터셰드 알고리즘은, 도 12B에 도시된 바와 같이, 분할 및 타원을 예측하기 위해 실행된다. 사용자가 예측된 분할 및 타원을 편집할 필요가 있는 경우, 더 많은 영역에 레이블을 지정하고 워터셰드를 다시 실행할 수 있다.
- [0257] 예측이 신경망의 사용자(예컨대, 연구자)에 의해 선택된 미리 결정된 오차 허용범위 내에 있을 때, 사용자는 타원의 방향을 선택한다. 상, 하, 좌, 우의 네 가지 기본 방향 중 하나를 선택하여 이 작업을 수행한다. 정확한 각도는 타원 피팅 알고리즘에 의해 선택되므로, 사용자는  $\pm 90$ 도 방향만 식별하면 된다. 방향이 선택되면, 모든 관련 데이터가 저장되고 레이블을 지정할 새 프레임이 사용자에게 표시된다.
- [0258] 주석이 달린 데이터 세트의 목적은 쥐에 대한 양호한 타원 피팅 추적 데이터를 식별하는 것이다. 데이터를 라벨링하는 동안, 타원 피팅은 타원이 쥐의 몸통 중앙에 오도록 장축 가장자리가 쥐의 코에 거의 닿도록 최적화된다. 종종, 더 좋은 타원 피팅을 제공하기 위해 분류 마스크로부터 꼬리가 제거된다.
- [0259] 추론을 위해 네트워크를 훈련시키기 위해, 3개의 주석이 달린 훈련 세트가 생성된다. 각 데이터 세트는 참조 프레임(입력), 분할 마스크, 및 타원 피팅을 포함한다. 훈련 세트 각각은 상이한 환경에서 죄를 추적하기 위해 생성되었다.
- [0260] 첫 번째 환경은 16,802개의 주석이 달린 프레임을 포함하는 일정한 흰색 배경을 갖는 오픈 필드였다. 첫 16,000 프레임은 24개의 동일한 설정 중 하나에서 획득한 65개의 개별 비디오에서 레이블이 지정되었다. 네트워크에 대한 첫 훈련 후, 주석이 달린 데이터에 포함되지 않은 특수한 상황에서 네트워크 성능이 좋지 않은 것으로 관찰되었다. 중간 점프, 이상한 자세, 및 아레나에서의 배로 사려는 일반적으로 실패로 관찰되었다. 이러한 실패는 성능을 더욱 일반화하고 향상시키기 위해 식별되고 올바르게 라벨링되며 주석이 달린 훈련 세트에 통합되었다.
- [0261] 두 번째 환경은 2개의 별개의 조명 조건(주간 가시광선 조명 및 야간 적외선 조명) 하에서 ALPHA-dri bedding 및 음식 컵을 갖는 표준 오픈 필드였다. 이 데이터 세트에서 4일 동안 6개의 설정에서 총 2,192개의 프레임에 주석이 달렸다. 주석이 달린 프레임 중 916개는 야간 조명에서, 주석이 달린 프레임 중 1,276개는 주간 조명에서 가져왔다.
- [0262] 최종 주석이 달린 데이터 세트는 KOMP 데이터 세트에 대한 Opto-M4 오픈 필드 케이지를 사용하여 생성되었다. 이 데이터 세트는 1083개의 레이블된 프레임이 포함한다. 이 레이블들은 모두 서로 다른 비디오(비디오 당 1 프

레이프로 레이블 됨)와 8개의 상이한 설정에서 샘플링 되었다.

[0263] **신경망 훈련**

[0264] **a) 훈련 데이터 세트 증강**

[0265] 이 훈련 데이터 세트는 대비, 밝기, 회전에서 작은 임의의 변화를 적용하여 입력 데이터의 작은 변동에 네트워크를 강하게 함으로써 훈련 동안 8배 증강되었다. 신경망이 훈련 데이터 세트를 암기하는 것을 방지하기 위해 증강이 수행되며, 이는 데이터 세트(유효)에 없는 예에서 성능이 저하될 수 있다. 자세한 내용은 Krizhevsky, 2012에서 확인할 수 있다.

[0266] 훈련 세트 증강은 Alexnet (Krizhevsky, 2012) 이후 신경망 훈련의 중요한 측면이었다. 우수한 정규화 성능을 달성하기 위해 소수의 훈련 세트 증강이 사용된다. 이 데이터는 조감도에서 나온 것이므로, 동일한 훈련 세트 크기를 즉시 8배 증가시키기 위해 수평, 수직, 및 대각선 반사를 적용하는 것이 간단하다. 또한 런타임에는, 전체 프레임에 대해 작은 회전 및 변환이 적용된다. 회전 증강 값은 균일한 분포에서 샘플링된다. 마지막으로, 노이즈, 밝기, 및 대비 증강도 프레임에 적용될 수 있다. 이들 증강에 사용된 난수 값은 정규 분포에서 선택된다.

[0267] **b) 훈련 학습률 및 배치 크기**

[0268] 훈련 학습률 및 배치 크기는 각각의 네트워크 훈련에 대해 독립적으로 선택되었다. Resnet V2 200과 같은, 더 큰 네트워크는 480x480의 입력 크기에서 배치 크기에 대한 메모리 제한에 부딪칠 수 있지만 그리드 검색 접근법을 사용하여 양호한 학습률 및 배치 크기를 실험적으로 식별했다. 이러한 네트워크를 훈련시키기 위해 선택된 하이퍼파라미터는 위의 표 3에 설명되어 있다.

[0269] **모델**

[0270] 모델은 Tensorflow v1.0에서 구축, 훈련, 및 테스트 되었다. 제시된 훈련 벤치마크는 NVIDIA® Tesla® P100 GPU 아키텍처에서 수행되었다.

[0271] 하이퍼파라미터는 여러 훈련 반복을 통해 훈련되었다. 네트워크의 첫 번째 훈련 후, 훈련 데이터에 제시되지 않은 특수한 상황에서 네트워크의 성능이 좋지 않은 것으로 관찰되었다. 중간 점프, 이상한 자세, 및 아레나에서의 배뇨 사례는 일반적으로 실패로 관찰되었다. 이러한 어려운 프레임은 성능을 더욱 향상시키기 위해 훈련 데이터 세트에 식별 및 통합되었다. 최종 모델 정의 및 훈련 하이퍼파라미터에 대한 자세한 설명은 위의 표 3에 설명되어 있다.

[0272] 3개의 네트워크 모두에 의해 표시되는 훈련 및 검증 손실 곡선의 플롯이, 각각, 도 13A 내지 도 13E에 도시되어 있다. 전반적으로, 훈련 및 검증 손실 곡선은 3개의 네트워크 모두 1에서 2 사이의 픽셀의 평균 오차의 성능으로 훈련함을 나타낸다. 예상치 못하게, 비닝된 분류 네트워크는 불안정한 손실 곡선을 보여, 과적합 및 검증에 대한 일반화가 불량함을 나타낸다(도 13B, 도 13E). 회귀 아키텍처는 1.2 px의 유효성 검사 오류로 수렴되어 유효성 검사보다 우수한 훈련 성능을 보여준다(도 13A, 13B, 13D). 그러나, 최상의 결과를 제공한 특징 추출기인 Resnet V2 200은 200 개가 넘는 층과 62.7백만 개의 파라미터를 가진 크고 깊은 네트워크이며 프레임 당 처리 시간(33.6ms)이 상당히 길어진다. 다른 사전 구축된 범용 네트워크(Zoph, 2017)는 더 빠른 컴퓨팅 시간의 균형을 유지하면서 비슷하거나 더 나쁜 성능을 달성할 수 있다. 따라서, 회귀 네트워크는 정확하지만 계산 비용이 많이 드는 솔루션이다.

[0273] 도 13A, 도 13B, 도 13C에 더 도시된 바와 같이, 인코더-디코더 분할 아키텍처는 0.9px의 검증 에러로 수렴되었다. 분할 아키텍처는 성능이 좋을 뿐만 아니라 프레임 당 평균 5-6ms의 평균 처리시간에서 GPU 계산에 효율적이다. 비디오 데이터는, 서버 등급 GPU인, NVIDIA® Tesla® P100에서는 최대 200fps(6.7X 리얼타임)로 처리될 수 있고, 소비자 등급 GPU인, NVIDIA® Titan XP에서는 최대 125fps(4.2X 리얼타임)로 처리될 수 있다. 이 높은 처리 속도는 구조가 깊이가 18층이고 10.6 백만 개의 파라미터 때문일 수 있다.

[0274] 훈련 세트 크기는 또한 인코더-디코더 분할 네트워크 아키텍처에 대한 양호한 네트워크 성능을 위해 필요한 레이블된 훈련 데이터의 상대적인 스케일을 식별하기 위해 벤치마킹 되었다. 이 벤치마크는 훈련 세트의 서브 세트(예컨대, 10000, 5000, 2500, 1000, 및 500)를 섞고 무작위로 샘플링하여 테스트되었다. 서브 샘플링된 각 훈련 세트는 훈련되었고 동일한 검증 세트와 비교되었다. 이 벤치마킹의 결과는 도 14A 내지 도 14H에 도시되어 있다.

[0275] 일반적으로, 훈련 곡선은 구별할 수 없는 것으로 보인다(도 14A). 즉, 훈련 크기 세트는 훈련 세트 에러율의 성

능 변화를 나타내지 않는다(도 14A). 놀랍게도, 검증 성능은 2,500개의 훈련 샘플을 초과하는 동일한 값으로 수렴하지만 에러는 1,000개의 훈련 샘플보다 증가된다(도 14B). 추가로 도시된 바와 같이, 검증 정확도는 2,500개 이상의 훈련 샘플동안(도 14C 내지 도 14F) 훈련 정확도보다 우수한 반면, 검증 정확도는 1,000에서 훈련 정확도만 매칭함으로써 약한 일반화의 징후를 나타내기 시작한다(도 14G). 분기 및 증가하는 검증 에러율(도 14H)에 의해 예시된 바와 같이, 500개의 훈련 샘플만을 사용하는 것이 명백하게 오버 트레이닝 된다. 이는 훈련 세트가 더 이상 네트워크가 일반화 될 수 있을 만큼 충분히 크지 않다는 것을 의미한다. 따라서, 2,500개의 주석이 달린 이미지로만 훈련된 네트워크에서 좋은 결과를 얻을 수 있으며, 라벨링 인터페이스로 생성하는 데 약 3시간이 걸린다. 따라서 정확한 수의 훈련 샘플이 궁극적으로 시각적 문제의 어려움에 의존하지만 권장되는 시작점은 약 2,500개의 훈련 샘플이다.

[0276] 개시된 실시예들에 따라 추적된 쥐를 예시하는 예시적인 비디오 프레임이 가시광선에서 도 15A 내지 도 15B 및 적외선에서 도 15C 내지 도 15D에 도시되어 있다. 도시된 바와 같이, 개별 쥐의 공간 범위는 픽셀 별로 컬러 코딩된다.

[0277] 계산 효율, 정확성, 훈련 안정성, 및 적은 수의 필요한 훈련데이터가 주어지면, 다른 접근법들과 비교하기 위해 전체 비디오에 대한 쥐의 위치를 예측하기 위해 인코더-디코더 분할 아키텍처가 선택되었다.

[0278] 신경망 기반 추적의 품질은 상이한 텍 생성 및 데이터 수집 환경을 갖는 쥐로부터의 전체 비디오를 추론하고(도 8a) 추적 품질을 시각적으로 평가함으로써 평가되었다. 신경망 기반 추적은 또한 독립적인 추적 양식인 KOMP2 빔 브레이크 시스템과 비교되었다(도 8a, 6열).

[0279] **실험 아레나**

[0280] **a) 오픈 필드 아레나**

[0281] 아레나(200)의 실시예는 오픈 필드 아레나로서 사용되었다. 오픈 필드 아레나의 크기는 52cm x 52cm이다. 바닥은 흰색 PVC 플라스틱이고 벽은 회색 PVC 플라스틱이다. 청소 유지 보수를 돕기 위해 모든 내부 모서리에 흰색 2.54cm 모따기(chamfer)가 추가되었다. 조명은 LED Light Ring(모델: F&V R300)에 의해 제공된다. 각각의 아레나에서 600럭스의 광을 생성하기 위해 Light Ring을 보정했다.

[0282] **b) 24시간 모니터링 오픈 필드 아레나**

[0283] 오픈 필드 아레나는 수일 간의 테스트를 위해 보강되었다. 발광체(212)는 표준 12:12LD 순환으로 설정된 오버헤드 LED 조명의 형태이다. ALPHA-dri는 잠자리를 위해 아레나에 배치되었다. 음식과 물을 제공하기 위해 단일 Diet Gel 76A 음식 컵을 아레나에 배치했다. 이 영양 공급원은 고갈되면 모니터링되고 교체되었다. 각 매트릭스에는 주간에는 250 럭스, 야간에는 약 500럭스 미만으로 조명되었다. 야간 동안 비디오를 녹화하기 위해, 발광체(212)는 적외선 LED(940nm)를 포함하였다.

[0284] **c) KOMP 오픈 필드 아레나**

[0285] 커스텀 아레나와 더불어, 개시된 시스템들 및 방법들의 실시예들은 또한 시판되는 시스템에서 벤치마킹 되었다. Opto-M4 Open Field Cage는 투명한 플라스틱 벽을 사용하여 제작되었다. 따라서 결과적인 반사로 인해 시각적 추적이 매우 어려워진다. 우리(cage)는 42cm x 42cm로 측정된다. 이 아레나의 조명은 100-200 럭스의 LED 조명으로 수행되었다.

[0286] **비디오 획득**

[0287] 모든 비디오 데이터는 도 2 및 도 7과 관련하여 논의된 비디오 획득 시스템의 실시예에 의해 획득되었다. 비디오 데이터는 Sentech 카메라(모델: STC-MB33USB) 및 컴퓨터 렌즈(모델: T3Z2910CS-IR)을 이용하여 640x480 픽셀 해상도, 8비트 흑백 심도, 및 약 29fps(예컨대, 29.9fps)에서 획득되었다. 노출 시간과 이득은 190/255의 목표 밝기를 사용하여 디지털 방식으로 제어되었다. 조리개를 가장 넓게 조정하여 더 낮은 아날로그 이득을 사용하여 목표 밝기를 달성했다. 이는 베이스 라인 노이즈의 증폭을 줄인다. "raw video" 코덱과 "pal8" 픽셀 형식을 사용하여 파일을 로컬 하드 드라이브에 임시 저장했다. 약 2시간 동안 분석을 수행하여 약 50GB의 로(raw) 비디오 파일을 생성했다. 발새, ffmpeg 소프트웨어를 사용하여 480x480 픽셀 자르기, 노이즈 제거 필터를 적용하고, mpeg4 코덱(품질을 최대로 설정)을 사용하여 압축하여 약 600MB의 압축된 비디오 크기를 생성했다.

[0288] 카메라(210)는 원근 왜곡을 완화시키기 위해 선반(202b) 위로 거의 100cm 떨어진 프레임(202)에 장착되었다. 8px/cm의 줌을 달성하기 위해 줌과 초점이 수동으로 설정되었다. 이 해상도는 아레나 경계에서 사용되지 않는

픽셀을 최소화하고 쥐 마다 약 800 픽셀의 영역을 생성한다. KOMP 아레나는 약간 작지만 동일한 8px/cm 대상 줌이 사용되었다.

[0289] 2002개의 비디오(총 700시간)는 인코더-디코더 분할 신경망을 사용하여 KOMP2 데이터 세트로부터 추적되었고 그 결과는 도 8에 도시되어 있다. 이들 데이터는 20분 오픈 필드 분석에서 테스트 된 C57BL/6NJ 배경에서 232개의 녹아웃 라인을 포함한다. 각 KOMP2 아레나는 투명한 매트릭스로 인해 배경이 약간 다르기 때문에, 추적 성능을 8개의 테스트 챔버(평균  $n=250$ , 도 16) 각각 및 모든 박스와 비교했다. KOMP2에 의해 사용된 모든 8개의 시험 챔버에 걸쳐 2개의 접근법 사이의 오픈 필드에서 이동된 총 거리 사이에 매우 높은 상관 관계가 관찰되었다 ( $R=96.9\%$ ). 이 경향(빨간색 화살표)으로부터 두 가지 동물이 높은 불일치로 관찰되었다. 비디오를 관찰한 결과 두 동물 모두에 이상한 자세가 나타났으며, 한 쪽에는 뒤통거림이 있고, 다른 쪽에는 구부러진 보행이 있었다. 뒤통거리고 구부러진 보행은 비정상적인 빔 브레이크를 초래하여 빔 브레이크 시스템으로부터 잘못된 전체 거리 이동 측정을 일으키는 것으로 여겨진다. 이 예는 신경망의 장점 중 하나를 강조하며, 이는 동물의 자세에 영향을 받지 않는다.

[0290] 훈련된 분할 신경망의 성능은 또한 도 8A와 관련하여 위에서 논의된 다양한 테스트 환경 및 털 색상으로부터의 광범위한 비디오 선택에 걸쳐 Ctrax와 비교되었다. Ctrax 비교는 여러 가지 이유로 동기가 부여된다. 일 측면에서, Ctrax는 많은 추적 설정의 미세 조정을 허용하는 최고의 종래의 추적기 중 하나로 간주된다.

[0291] Ctrax는 추적 능력을 최적화하기 위해 다양한 설정을 포함한다(Branson, 2009). 이 소프트웨어의 저자는 좋은 추적을 보장하기 위해 아레나가 특정 기준에 따라 설정되도록 강력히 권장한다. 본 명세서에서 논의된 대부분의 테스트(예컨대, 흰색 배경의 알비노 쥐)에서, Ctrax가 잘 수행되도록 설계되지 않은 환경이 사용된다. 그럼에도 불구하고, 잘 조정된 파라미터를 사용하면 여전히 우수한 성능을 달성할 수 있다. Ctrax는 조작할 수 있는 설정이 많이 때문에 추적 성능을 높이기 위해 많은 시간이 소요된다. 개시된 환경에서 쥐를 추적하기 위해 Ctrax를 설정하기 위한 프로토콜은 다음과 같다.

[0292] 첫 번째 동작에서, 배경 모델이 생성된다. Ctrax의 핵심은 배경 감산을 기반으로 하므로 강력한 배경 모델이 기능에 필수적이다. 쥐가 움직일 때 모델이 최적으로 작동한다. 배경 모델을 생성하기 위해 마우스가 명확하게 이동하고 해당 섹션에서 프레임을 샘플링하는 비디오 세그먼트가 필요하다. 이렇게 하면 쥐가 배경 모델에 포함되지 않는다. 이 접근법은 쥐가 자주 움직이지 않고 일반적으로 배경 모델에 통합되지 때문에 24시간 데이터에서 Ctrax의 추적 성능을 크게 향상시킨다.

[0293] 두 번째 동작은 배경 감산에 대한 설정을 설정하는 것이다. 여기서, 배경 밝기 정규화 방법은 Std 범위가 254.9에서 255.0으로 사용된다. 노출 및 털 색상의 약간의 변화가 성능에 영향을 미치므로 쥐 세그먼트를 분할하는 데 적용되는 임계값은 비디오 별로 조정된다. 이러한 임계값을 조정하기 위해 적절한 시작 값 세트가 적용되고 일반적으로 우수한 성능을 보장하기 위해 비디오가 검토된다. 특정 실시예들에서, 모든 비디오는 전형적으로 그립자로 인해 추적하기 가장 어려운 프레임이기 때문에, 벽에서 쥐 양육의 예를 점검할 수 있다. 또한 타원을 피팅하기 위해 쥐의 꼬리를 제어할 뿐만 아니라 환경에서 드문 변화를 제거하기 위해 형태 필터링을 적용할 수 있다. 개방 반경 4 및 폐쇄 반경 5를 사용하였다.

[0294] 추가 동작에서, Ctrax가 가능하게 하는 다양한 추적 파라미터는 관찰이 실제로 쥐가 되도록 하기 위해 수동으로 조정된다. 시간을 고려하여, 이들 파라미터를 한번 잘 조정된 후 추적된 다른 모든 쥐에 사용하였다. 비디오의 성능이 눈에 띄게 좋지 않으면 일반 설정이 조정되어 성능이 향상되었다. 형상 파라미터의 경우, 2개의 표준 편차에 기초한 경계는 개별 검정색 쥐 비디오로부터 결정되었다. 특정 쥐가 분할 단계에서 제대로 수행되지 않을 것으로 예상되므로 최소값이 더 낮아졌다. 이를 통해 Ctrax는 쥐 전체를 분할할 수는 없지만 쥐의 좋은 위치를 찾을 수 있다. 셋업이 모두 8개의 동일한 줌을 가지며 테스트된 쥐는 일반적으로 동일한 모양이므로 이 방법은 잘 작동한다. 실험 설정은 아레나에서 하나의 쥐만 추적하기 때문에 모션 설정은 매우 관대하다. 관찰 파라미터에서 "최소 영역 무시"가 주로 사용되며, 이는 큰 탐지를 걸러낸다. 여기에서 2500 보다 큰 탐지가 필터링된다. 후시 탭에서 "스푸리어스 감지 수정" 설정은 500 프레임보다 짧은 감지를 제거하는 데 사용된다.

[0295] 동물들이 전체 시간 동안 계속해서 자고 있는, 24시간 장치로부터의 비디오는 Ctrax가 유효한 배경 모델을 만들 수 없었기 때문에 비교에서 수동으로 생략되었다. Ctrax와 신경망 사이에서 이동한 총 거리의 누적 상대 오차가 계산되었고 도 17A에 도시되어 있다. 비디오에서 1분마다 신경망과 Ctrax의 거리 이동 예측을 비교한다. 이 측정 항목은 각 쥐의 질량 중심 추적 정확도를 측정한다. 검정색, 회색, 및 얼룩무늬 쥐를 추적한 결과 4% 미만의 에러가 발생했다. 그러나 알비노(14%), 24시간 아레나(주황색 27%), 및 KOMP2(파란색 10%)에서 더 높은 수준의 에러가 나타났다(도 17A). 신경망 추적기가 없으면 트랙 알비노, KOMP2, 또는 24시간 데이터는 적절히 추적될



수 없다.

- [0296] 그림자가 예측에 포함되는 경우와 같이, 전경 분할 예측이 부정확할 때, 타원 피팅은 쥐의 자세를 정확하게 나타내는 것이 아니라는 것이 관찰되었다. 이 경우 질량 중심 추적이 허용 가능하더라도 타원 피팅 자체는 매우 가변적이었다.
- [0297] JAABA(Kabra, 2013)와 같은, 행동 인식을 위한 최신 기계 학습 소프트웨어는 행동의 분류를 위해 이러한 특징을 이용한다. 타원 추적의 분산은 단축의 상대 표준편차를 통해 양자화 되었으며 도 17B에 도시되어 있다. 추적이 정확할 때 행동 분석에서 표현된 넓은 범위의 자세를 통해 개별 쥐의 너비가 유사하게 유지되기 때문에 이 메트릭은 모든 실험실 쥐에서 가장 적은 분산을 보여준다. 총 이동거리의 낮은 누적 상대 에러(도 17B)에도 불구하고, 회색 및 얼룩무늬 쥐(도 17A)에서 높은 추적 변화가 관찰되었다. 알비노 및 KOMP2 추적에 대해 단축의 상대 표준편차가 높을 것으로 예상된다. 따라서, 질량 중심 추적과 타원 피팅의 분산 모두에서, 신경망 추적기는 종래의 추적기보다 성능이 우수하다.
- [0298] 매우 정확한 추적기로서 인코더-디코더 분할 신경망을 확립한 후에, 그 성능이 2개의 큰 행동 데이터 세트로 추가로 테스트 되었다. 오픈 필드 비디오 데이터는 모든 다양한 색상, 얼룩무늬, 누드, 및 비만 쥐를 포함하여, 57개 품종에 걸쳐 1845 마리의 쥐(1691 시간)로 생성되었다. 이 데이터 세트에는 47개의 근친 교배, 11개의 등질계 F1 쥐 품종이 포함되어 있으며, Bogue, 2018의 쥐 현상 데이터베이스에 따라 생성된 가장 큰 오픈 필드 데이터 세트이다.
- [0299] 총 이동거리에 대한 추적 결과가 도 18A에 도시되어 있다. 점은 품종의 개체를 나타내며, 박스는 평균  $\pm$  표준편차를 나타낸다. 사용자 튜닝 없이 단일 훈련된 네트워크를 사용하여, 모든 쥐가 높은 정확도로 추적되었다. 대부분의 품종으로부터 쥐를 추적하여 충실도를 육안으로 확인하고 우수한 성능을 관찰하였다. 관찰된 활동 표현형은 이전에 공개된 쥐 오픈 필드 행동의 데이터 세트와 일치한다.
- [0300] 4개의 C57BL/6J 및 2개의 BTBR T<sub>Itpr3tf</sub>/J 쥐에 대해 수집된 24시간 비디오 데이터를 추적하기 위해 동일한 신경망이 사용되었다(도 8A, 5열). 이 쥐는 잠자리, 음식, 및 물컵과 함께 수용되었고, 그 동안 음식은 위치를 바꾸고 12:12의 명암 조건에서 변화시켰다. 가시광선과 적외선을 사용하여 비디오 데이터가 녹화되었다. 동일한 네트워크를 사용하여 이러한 조건 하에서 모든 동물에 대한 활동이 추적되었고, 매우 양호한 성능이 밝고 어두운 조건에서 관찰되었다.
- [0301] 결과는 도 18B에 도시되어 있으며, 명 조건을 나타내는 8개의 광점 및 암 조건을 나타내는 암점을 각각 갖는다. 예상한 바와 같이, 어두운 단계 동안 활동 리듬이 높은 수준의 운동 활동으로 관찰되었다.
- [0302] 요컨대, 복잡한 환경에서 동물의 비디오 기반 추적은 동물 행동 분야에서 오랫동안 도전되었다(Egnor, 2016). 현재의 최신 시스템은 동물 분할의 근본적인 문제를 다루지 않으며 정확한 추적을 위해 전경과 배경 사이의 시각적 대비에 크게 의존한다. 결과적으로, 사용자는 최적의 결과를 얻기 위해 환경을 제한해야 한다.
- [0303] 복잡하고 동적인 환경에서 기능할 수 있는 현대 신경망 기반 추적기 및 대응하는 사용 방법이 여기에 설명된다. 추적, 전경 및 배경 분할의 근본적인 문제는 훈련 가능한 신경망의 사용을 통해 해결된다. 상이한 세 가지 아키텍처를 테스트한 결과 인코더-디코더 분할 네트워크가 높은 수준의 정확도와 기능을 높은 속도(6X 실시간 이상)로 달성한다.
- [0304] 2,500개의 이미지를 라벨링함으로써, 거의 3시간이 걸리는, 사용자가 특정 환경에 대한 새로운 네트워크를 훈련시킬 수 있는 라벨링 인터페이스가 추가로 제공된다.
- [0305] 개시된 훈련된 신경망은 2개의 종래 솔루션과 비교 되었으며 복잡한 환경에서 이들을 훨씬 능가하는 것으로 밝혀졌다. 배경 감산 접근법을 사용하는 선반 시스템에서 유사한 결과가 예상된다. 실제로, 26개의 상이한 배경 감산 접근법을 테스트할 때, 특정 상황에서 각각 실패한 것으로 관찰되었다. 그러나, 단일 신경망 아키텍처는 미세 조정이나 사용자 입력 없이 여러 환경에서 쥐의 모든 털 색상에 대해 기능할 수 있다. 이 기계 학습 접근법은 사용자 입력을 최소화하면서 동적인 환경 조건에서 장기 추적을 가능하게 하여, 행동 연구를 위한 차세대 추적 아키텍처의 기초를 형성한다.
- [0306] 본 명세서에 기술된 제어 시스템의 적어도 하나의 측면들 또는 특징들은 디지털 전자회로, 집적회로, 특별히 설계된 주문형 집적회로(ASIC), 필드 프로그램 가능 게이트 어레이(FPGA), 컴퓨터 하드웨어, 펌웨어, 소프트웨어, 및/또는 이들의 조합에 의해 실현될 수 있다. 이러한 다양한 측면들 또는 특징들은 적어도 하나의 프로그램 가능한 프로세서를 포함하는 프로그램 가능한 시스템상에서 실행 가능하고 및/또는 해석 가능한 적어도 하나의 컴

퓨터 프로그램에서의 구현을 포함할 수 있으며, 이는 특수 또는 범용일 수 있고, 저장 시스템, 적어도 하나의 입력 장치, 및 적어도 하나의 출력 장치로부터/에 데이터 및 명령을 수신/전송하기 위해 결합될 수 있다. 프로그램 가능한 시스템 또는 컴퓨터 시스템은 클라이언트 및 서버를 포함할 수 있다. 클라이언트 및 서버는 일반적으로 서로 멀리 떨어져 있으며, 일반적으로 통신 네트워크를 통해 상호작용한다. 클라이언트와 서버의 관계는 각 컴퓨터에서 실행되고 서로 클라이언트-서버 관계를 갖는 컴퓨터 프로그램으로 인해 발생한다.

[0307] 프로그램, 소프트웨어, 소프트웨어 애플리케이션, 애플리케이션, 컴포넌트, 또는 코드로도 지칭될 수 있는, 컴퓨터 프로그램은 프로그램 가능한 프로세서를 위한 기계 명령어를 포함하고, 고수준 절차 언어, 객체 지향 프로그래밍 언어, 기능적 프로그래밍 언어, 논리적 프로그래밍 언어, 및/또는 어셈블리/기계언어로 구현될 수 있다. 본 명세서에서 사용된 것과 같이, 용어 "기계 판독 가능한 매체"는, 기계 명령어 및/또는 데이터를, 기계 판독 가능한 신호로서 기계 명령어들을 수신하는 기계 판독 가능한 매체를 포함하는, 프로그램 가능한 프로세서에 제공하기 위해 사용되는, 예를 들어 자기 디스크, 광학 디스크, 메모리, 및 프로그램 가능한 로직 디바이스(PLDs)와 같은, 임의의 컴퓨터 프로그램 제품, 장치, 및/또는 디바이스를 지칭한다. 용어 "기계 판독 가능한 신호"는 기계 명령 및/또는 데이터를 프로그램 가능한 프로세서에 제공하는데 사용되는 임의의 신호를 지칭한다. 예를 들어 비일시적 솔리드 스테이트 메모리 또는 자기 하드 드라이브 또는 임의의 등가 저장 매체와 같은, 기계 판독 가능한 매체는 그러한 명령어를 비일시적으로 저장할 수 있다. 기계 판독 가능한 매체는, 예를 들어 적어도 하나의 물리적 프로세서 코어와 연관된 프로세서 캐시 또는 다른 랜덤 액세스 메모리와 같은, 일시적인 방식으로 그러한 기계 명령어들을 대안적으로 또는 추가적으로 저장할 수 있다.

[0308] 사용자와의 상호작용을 제공하기 위해, 본 명세서에 설명된 주제의 적어도 하나의 측면들 또는 특징들은, 예를 들어, 사용자에게 정보를 표시하기 위한 음극선관(CRT) 또는 액정 디스플레이(LCD) 또는 LED 모니터와 같은, 디스플레이 장치, 및 사용자가 컴퓨터에 입력을 제공할 수 있는 키보드와 예컨대, 마우스, 트랙볼 등 포인팅 장치를 갖는 컴퓨터에서 구현될 수 있다. 사용자와의 상호 작용을 제공하기 위해 다른 종류의 장치 역시 사용될 수 있다. 예를 들어, 사용자에게 제공되는 피드백은, 예를 들어 시각적 피드백, 청각적 피드백, 또는 촉각 피드백과 같은, 임의의 형태의 감각 피드백일 수 있고; 사용자로부터의 입력은 음향, 음성, 또는 촉각 입력을 포함하지만, 이에 제한되지 않는, 임의의 형태로 수신될 수 있다. 다른 가능한 입력 장치로는 터치 스크린, 또는 단일 또는 멀티 포인트 저항성 또는 용량성 트랙패드, 음성 인식 하드웨어 및 소프트웨어, 광학 스캐너, 광학 포인터, 디지털 이미지 캡처 장치 및 관련된 해석 소프트웨어 등과 같은 기타 터치 감지 장치가 포함된다.

[0309] 본 출원 전체에 걸쳐 인용된, 예를 들어 발행 또는 승인된 특허 또는 그에 상응하는 것을 포함하는 특허 문서, 특허출원공보, 및 비특허문헌 또는 기타 소스 자료와 같은, 모든 참고문헌 각각의 참고문헌이 본 출원의 개시 내용과 적어도 부분적으로 일치하지 않는 정도로, 개별적으로 참조로 인용되어 있지만, 그 전문에 본원에 참조로 포함된다. 예를 들어, 부분적으로 일치하지 않는 참조문헌은 참조의 부분적으로 일치하지 않는 부분을 제외하고 참조로 통합된다.

[0310] 마쿠쉬 그룹, 또는 다른 드롭이 본 명세서에서 사용될 때, 그룹의 모든 개별 멤버 및 그룹의 가능한 모든 조합 및 하위 조합은 본 개시에 개별적으로 포함되도록 의도된다.

[0311] 본 명세서에서 사용된 바와 같이, 단수 형태 "a", "an", 및 "the"는 문맥상 명백하게 다르게 지시하지 않는 한 복수 참조를 포함한다. 따라서, 예를 들어, "세포"에 대한 언급은 복수의 이러한 세포들 및 당업자에게 공지된 등가물을 포함한다. 또한, 용어 "a"(또는 "an"), "하나 이상" 및 "적어도 하나"는 본 명세서에서 상호 교환적으로 사용될 수 있다.

[0312] 본 명세서에서 사용된 바와 같이, 용어 "구성되는"은 "포함하는", "갖는", "함유하는", 및 "특징된"과 동의어이며 각각은 상호 교환적으로 사용될 수 있다. 이들 용어 각각은 추가로 포함되거나 개방형이며, 언급되지 않은 추가 요소 또는 방법 단계를 배제하지 않는다.

[0313] 본 명세서에서 사용된 바와 같이, 용어 "~로 이루어진"은 청구항 구성요소에 명시되지 않은 임의의 요소, 단계, 또는 성분을 배제한다.

[0314] 본 명세서에서 사용된 바와 같이, 용어 "본질적으로 구성되는"은 청구범위의 기본 및 신규 특성에 실질적으로 영향을 미치지 않는 물질 또는 단계를 배제하지 않는다. 본 명세서의 각 예에서, "포함하는", "본질적으로 구성되는", 및 "구성되는"이라는 용어는 다른 두 용어 중 하나로 대체될 수 있다.

[0315] 본 명세서에 예시적으로 설명된 실시예들은 본 명세서에 구체적으로 개시되지 않은 임의의 요소 또는 요소들, 제한 또는 제한들이 없는 상태에서 적절하게 실시될 수 있다.

- [0316] "임의의 청구항 XX-YY의(여기서 XX 및 YY는 청구항 번호를 나타냄)"의 표현은 대안적인 형태로 다중 종속항을 제공하도록 의도되며, 일부 실시예들에서, "청구항 XX-YY 중 어느 하나"의 표현과 상호 교환될 수 있다.
- [0317] 다르게 정의되지 않는 한, 본 명세서에서 사용된 모든 기술적 및 과학적 용어는 개시된 실시예들이 속하는 기술 분야의 당업자에 의해 일반적으로 이해되는 것과 동일한 의미를 갖는다.
- [0318] 본 명세서에서, 예를 들어, 온도 범위, 시간 범위, 조성 범위, 또는 농도 범위와 같은, 범위가 주어질 때마다, 모든 범위 및 하위 범위뿐만 아니라 주어진 범위에 포함된 모든 개별값들도 본 개시에 포함되도록 의도된다. 본 명세서에 사용된 바와 같이, 범위는 구체적인 범위의 종점값으로서 제공된 값을 포함한다. 예를 들어, 1 내지 100의 범위는 구체적으로 1 및 100의 종점값을 포함한다. 본 명세서의 설명에 포함된 범위 또는 하위 범위의 임의의 하위 범위 또는 개별값은 본 명세서의 청구범위에서 제외될 수 있음을 이해할 것이다.
- [0319] 위의 설명 및 청구범위에서, "적어도 하나" 또는 "하나 이상"과 같은 문구 다음에 구성요소 또는 특징의 결합된 목록이 나타날 수 있다. 용어 "및/또는"은 또한 둘 이상의 구성요소 또는 특징의 목록에서 나타날 수 있다. 사용된 문맥에 의해 달리 암시적으로 또는 명시적으로 모순되지 않는 한, 이러한 문구는 열거된 구성요소 또는 특징 중 어느 하나를 개별적으로 또는 언급된 구성요소 또는 특징 중 임의의 다른 인용된 구성요소 또는 특징과 조합하여 의미하는 것으로 의도된다. 예를 들어, 문구 "A 및 B 중 적어도 하나"; "A 및 B 중 하나 이상"; "A 및/또는 B"는 각각 "A 단독, B 단독 또는 A 및 B 함께"를 의미하는 것으로 의도된다. 세 개 이상의 항목을 포함하는 목록에 대해서도 유사한 해석이 의도된다. 예를 들어, 문구 "A, B, 및 C 중 적어도 하나"; "A, B, 및 C 중 하나 이상"; "A, B, 및/또는 C"를 각각 "A 단독, B 단독, C 단독, A 및 B 함께, A 및 C 함께, B 및 C 함께 또는 A 및 B 및 C 함께"를 의미하는 것으로 의도된다. 또한, 위에서 또는 청구항에서 "기반"이라는 용어의 사용은 인용되지 않은 특징 또는 요소가 또한 허용될 수 있도록 "적어도 부분적으로 기초한"을 의미하는 것으로 의도된다.
- [0320] 본 명세서에 사용된 용어 및 표현은 제한이 아닌 설명의 용어로 사용되며, 도시되고 기술된 특징 또는 그 일부를 등가로 배제하려는 그러한 용어 및 표현을 사용하려는 의도는 없으며, 청구된 실시예의 범위 내에서 다양한 변형이 가능하다는 것이 인식된다. 따라서, 본 출원은 바람직한 실시예들, 예시적인 실시예들, 및 선택적 특징들에 대한 논의를 포함할 수 있지만, 여기에 개시된 개념들의 수정 및 변형은 당업자에 의해 이루어질 수 있음을 이해해야 한다. 이러한 수정 및 변형은 첨부된 청구범위에 의해 정의된 바와 같이 개시된 실시예의 범위 내에 있는 것으로 간주된다. 본 명세서에 제공된 특정 실시예는 본 개시의 유용한 실시예의 예이며, 본 명세서에 기재된 장치, 장치 구성 요소, 및 방법 단계의 많은 변형을 이용하여 수행될 수 있음이 당업자에게 명백할 것이다. 당업자에게 명백한 바와 같이, 본 방법에 유용한 방법 및 장치는 다수의 선택적인 조성물 및 처리 요소 및 단계를 포함할 수 있다.
- [0321] 본 개시의 실시예들은 그 사상 또는 본질적인 특성을 벗어나지 않고 다른 특성 형태로 구현될 수 있다. 그러므로 전술한 실시예들은 모든 면에서 본 명세서에 기술된 주제를 제한하기 보다는 예시적인 것으로 간주되어야 한다.
- [0322] 아래에 열거된 각각의 참고문헌은 그 전문이 본 명세서에 참조로 포함된다.
- [0323] Bargmann, C., Newsome, W., Anderson, A., Brown, E., Deisseroth, K., Donoghue, J., MacLeish, P., Marder, E., Normann, R., and Sanes, J., BRAIN 2025: a scientific vision. Brain Research Through Advancing Innovative Neurotechnologies (BRAIN) Working Group Report to the Advisory Committee to the Director, NIH (2014).
- [0324] Bogue, M.A. et al., "Mouse Phenome Database: an integrative database and analysis suite for curated empirical phenotype data from laboratory mice," *Nucleic Acids Res*, 46, D843-D850 (2018).
- [0325] Branson, K., Robie, A.A., Bender, J., Perona, P. & Dickinson, M.H., "High-throughput ethomics in large groups of *Drosophila*," *Nature methods*, 6(6), 451-457 (2009).
- [0326] Branson, K. and Belongie, S., "Tracking Multiple Mouse Contours (Without Too Many Samples)," 2005 IEEE Computer Society Conference on Computer Vision and Pattern (CVPR'05), June 20-25, San Diego, CA, pp. 1039-1046 (2005).
- [0327] Decoster, W., and Debruyne, F., "The ageing voice: changes in fundamental frequency, waveform stability and spectrum," *Acta Otorhinolaryngol Belg* 51, pp. 105-112 (1997).

- [0328] Deschenes, M., Moore, J., and Kleinfeld, D., "Sniffing and whisking in rodents," *Curr Opin Neurobiol*, 22, pp. 243-250 (2012).
- [0329] Egnor, S.E. and Branson, K., "Computational Analysis of Behavior," *Annu Rev Neurosci*, 39, pp. 217-236 (2016).
- [0330] Gomez-Marin, A., Paton, J.J., Kampff, A.R., Costa, R.M., and Mainen, Z.F., "Big behavioral data: psychology, ethology and the foundations of neuroscience," *Nat Neurosci*, 17, pp. 1455-1462 (2014).
- [0331] Goodfellow, I., Bengio, Y., and Courville, A. Deep Learning, MIT Press (2016)
- [0332] Grimsley, J.M., Monaghan, J.J., and Wenstrup, J.J., "Development of social vocalizations in mice," *PLoS One*, 6, e17460 (2011).
- [0333] He, K., Zhang, X., Ren, S. & Sun, J. in European Conference on Computer Vision Springer, pp. 630-645 (2016).
- [0334] Heckman, J.J., Proville, R., Heckman, G.J., Azarfar, A., Celikel, T., and Englitz, B., "High-precision spatial localization of mouse vocalizations during social interaction," *Sci Rep*, 7(7), 3017 (2017).
- [0335] Kabra, M., Robie, A.A., Rivera-Alba, M., Branson, S. & Branson, K., "JAABA: interactive machine learning for automatic annotation of animal behavior," *Nat Methods*, 10, pp. 64-67 (2013).
- [0336] Kalueff, A.V., LaPorte, J.L., and Bergner, C.L., Neurobiology of grooming behavior Cambridge; New York: Cambridge University Press (2010).
- [0337] Krizhevsky, A., Sutskever, I., and Hinton, G.E., "ImageNet classification with deep convolutional neural networks," Advances in neural information processing systems 25 (NIPS 2012).
- [0338] Kuleshkaya, N. & Voikar, V., "Assessment of mouse anxiety-like behavior in the light-dark box and open-field arena: role of equipment and procedure," *Physiol Behav* 133, 30-38 (2014).
- [0339] Kumar, V, Kim, K., Chryshanthi, J. Thomas, L.C., Hong, H., and Takahashi, J.S., "Second-generation high-throughput forward genetic screen in mice to isolate subtle behavioral mutants," *Proceedings of the National Academy of Sciences*, 108 (Supplement 3), pp. 15557-15564 (2011).
- [0340] LeCun, Y., Bengio, Y. & Hinton, G., "Deep learning," *Nature* 521, 436-444 (2015).
- [0341] Martins, R.H., Goncalvez, T.M., Pessin, A.B., and Branco, A., "Aging voice: presbyphonia," *Aging Clin Exp Res*, 26, pp. 1-5 (2014).
- [0342] Mueller, P.B., "The aging voice," *Semin Speech Lang*, 18(2), pp. 159-169 (1997).
- [0343] Neunuebel, J.P., Taylor, A.L., Arthur, B.J., and Egnor, S.E., "Female mice ultrasonically interact with males during courtship displays," *Elife*, 4:e06203, (2015).
- [0344] Ohayon, S., Avni, O., Taylor, A.L., Perona, P., and Roian Egnor, S.E., "Automated multi-day tracking of marked mice for the analysis of social behaviour," *J. Neurosci. Methods*, 219(1), pp. 10-19 (2013).
- [0345] Shannon, C.E., "Communication in the presence of noise," *Proceedings of the IRE*, 37(1), pp. 10-21 (1949).
- [0346] Sobral, A., "BGSLibrary: An OpenCV C++ Background Subtraction Library," in IX Workshop de Visão Computacional (WVC 2013), Vol. 2 7 (2013).
- [0347] Valdez, P. & Mehrabian, A., "Effects of color on emotions," *Journal of experimental psychology: General*, 123(4), pp. 394-409 (1994).
- [0348] Wiltshko, A.B., Johnson, M.J., Iurilli, G., Peterson, R.E., Katon, J.M., Pashkovski, S.L., Abaira, V.E., Adams, R.P., and Datta, S.R., "Mapping Sub-Second Structure in Mouse Behavior," *Neuron*, 88, pp. 1121-1135 (2015).

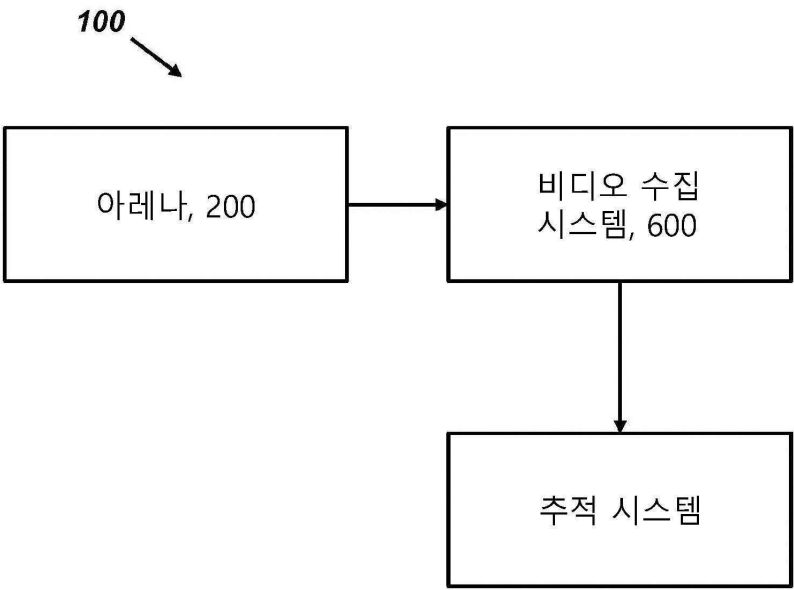


[0349] Zhang, C., Florêncio, D., Ba, D.E., and Zhang, Z., "Maximum likelihood sound source localization and beamforming for directional microphone arrays in distributed meetings," *IEEE Transactions on Multimedia*, 10, pp. 538-548 (2008).

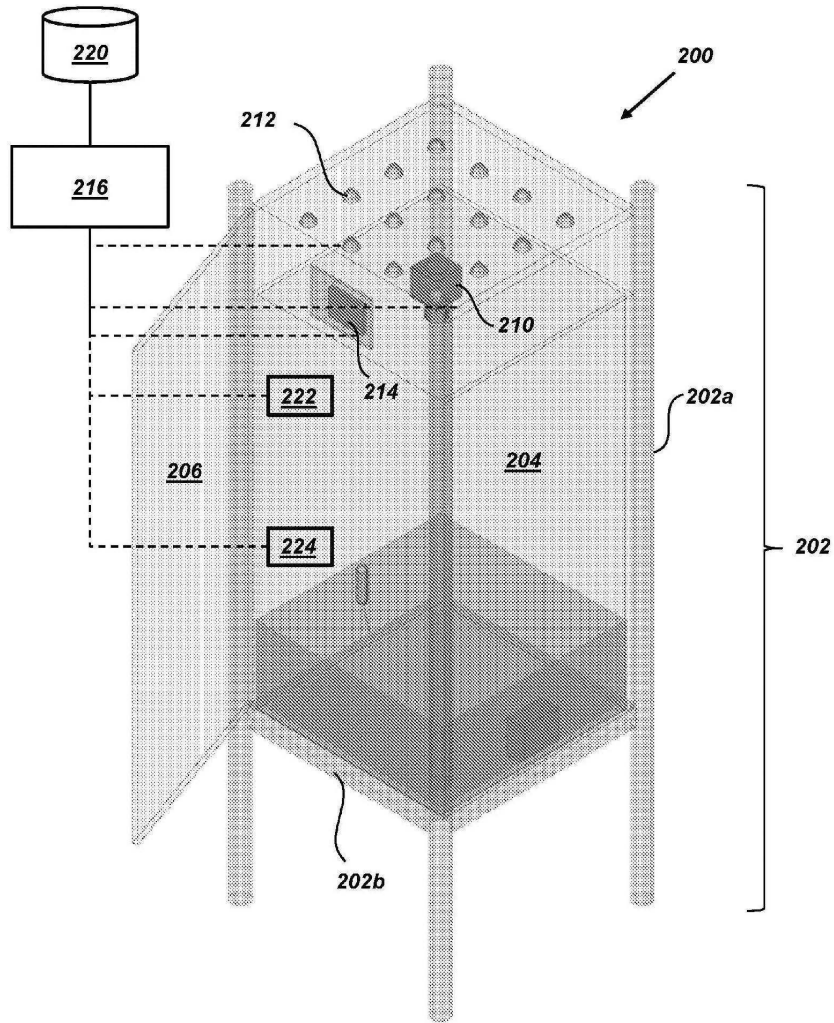
[0350] Zoph, B., Vasudevan, V., Shlens, J. & Le, Q.V., "Learning transferable architectures for scalable image recognition," *arXiv preprint arXiv:1707.07012* (2017).

도면

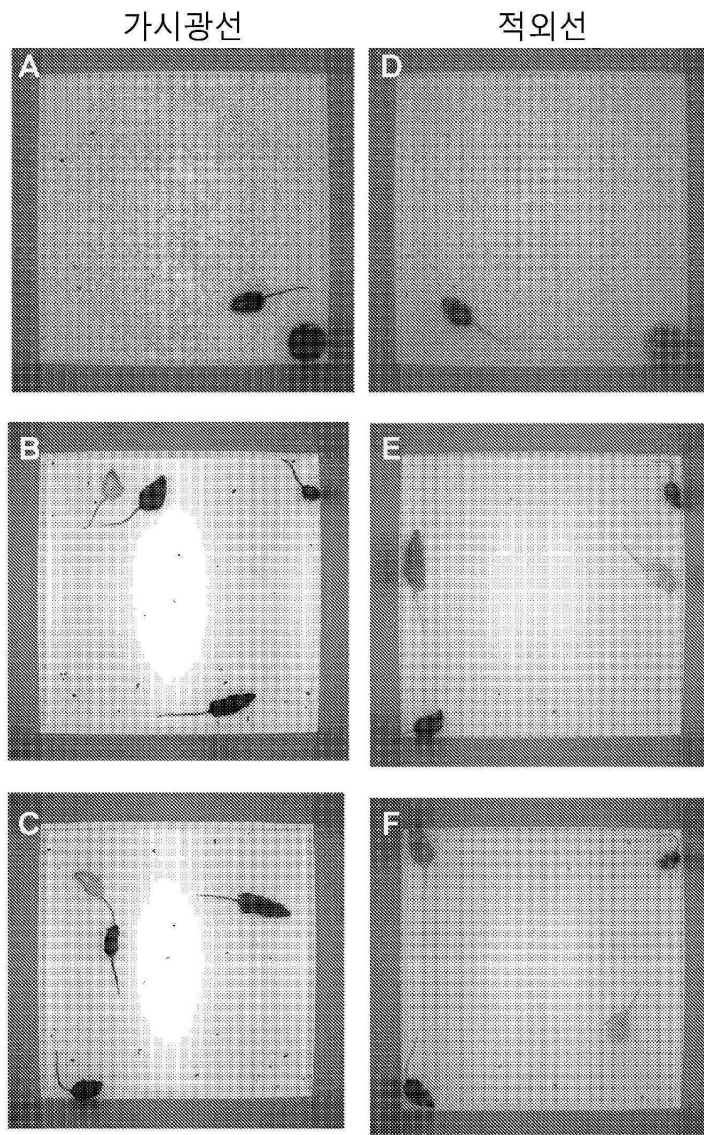
도면1



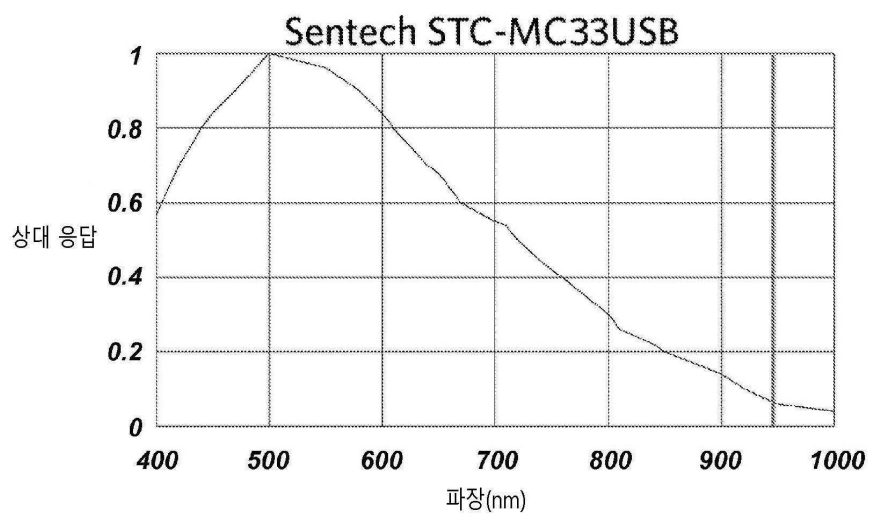
도면2



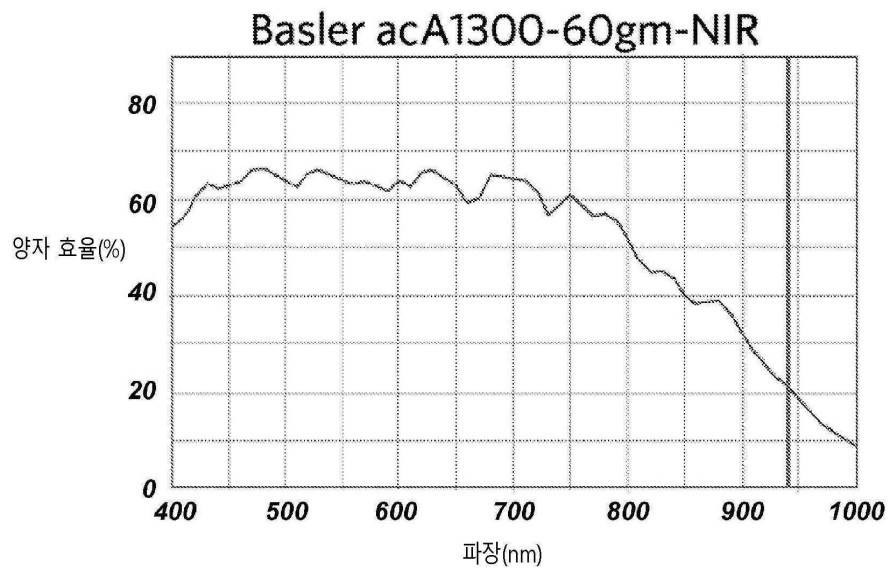
도면3



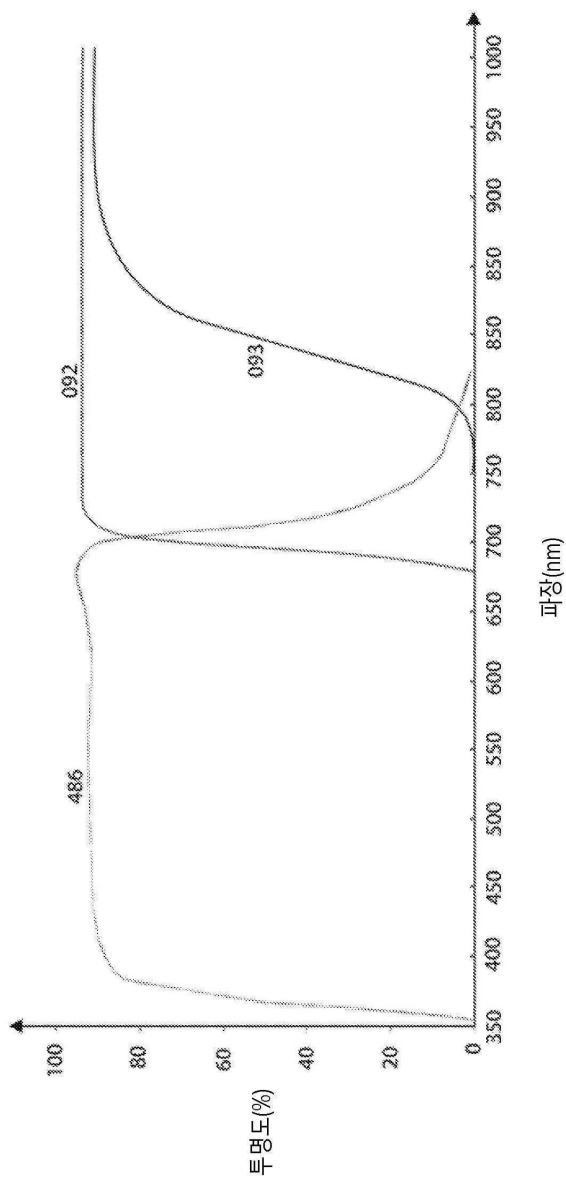
도면4a



도면4b

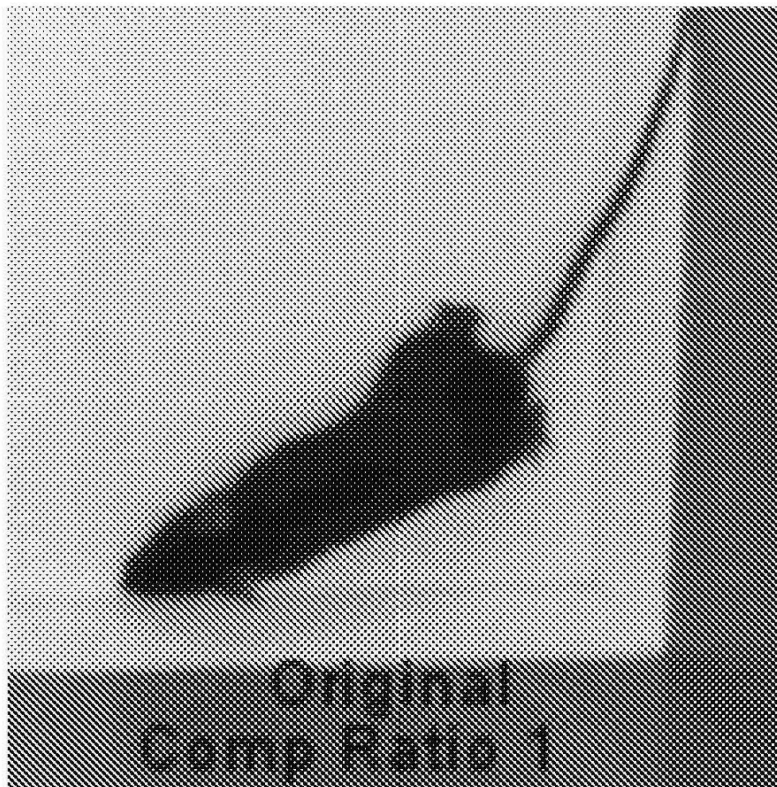


도면5

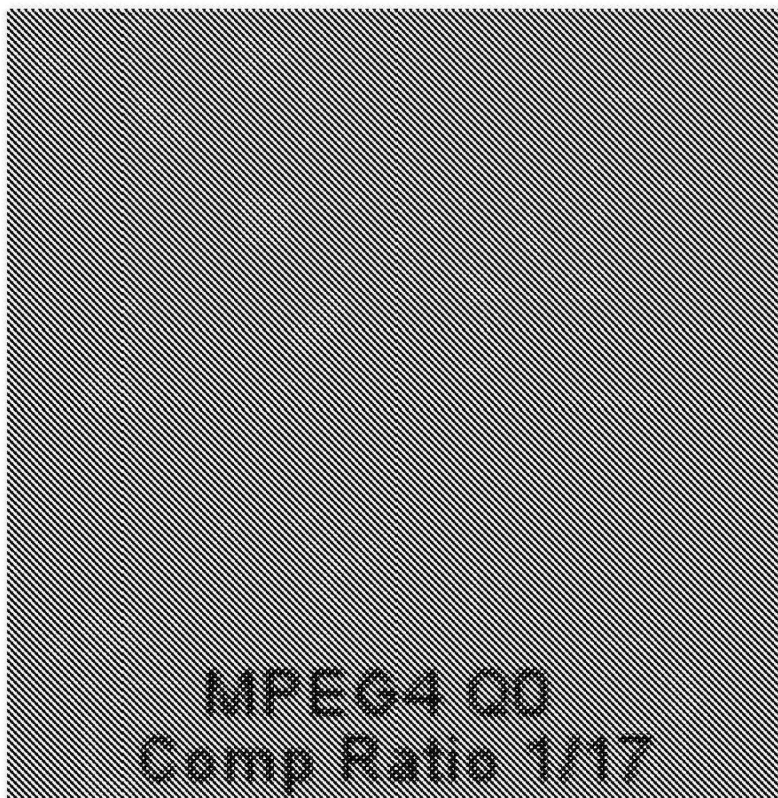




도면6a

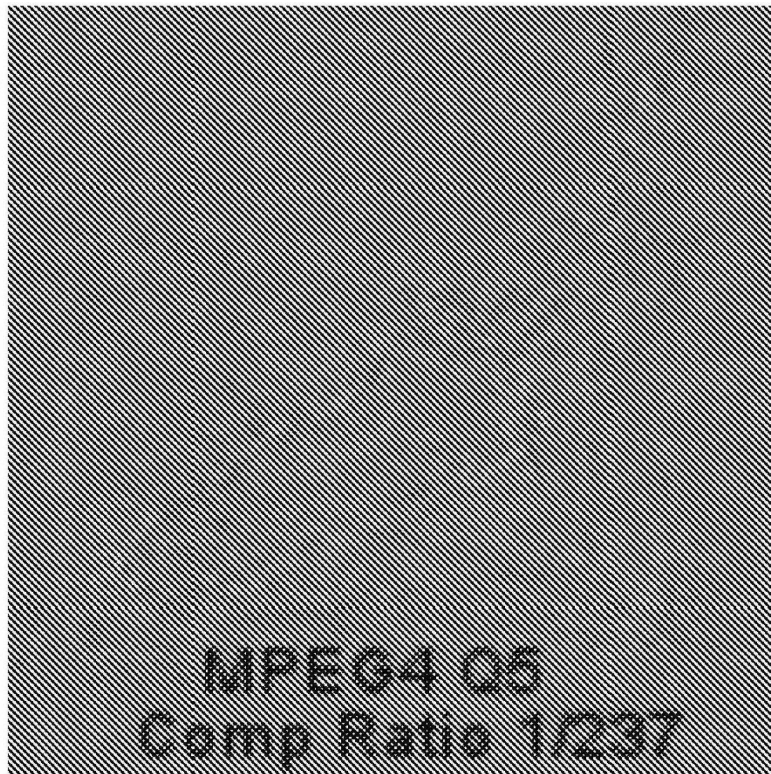


도면6b

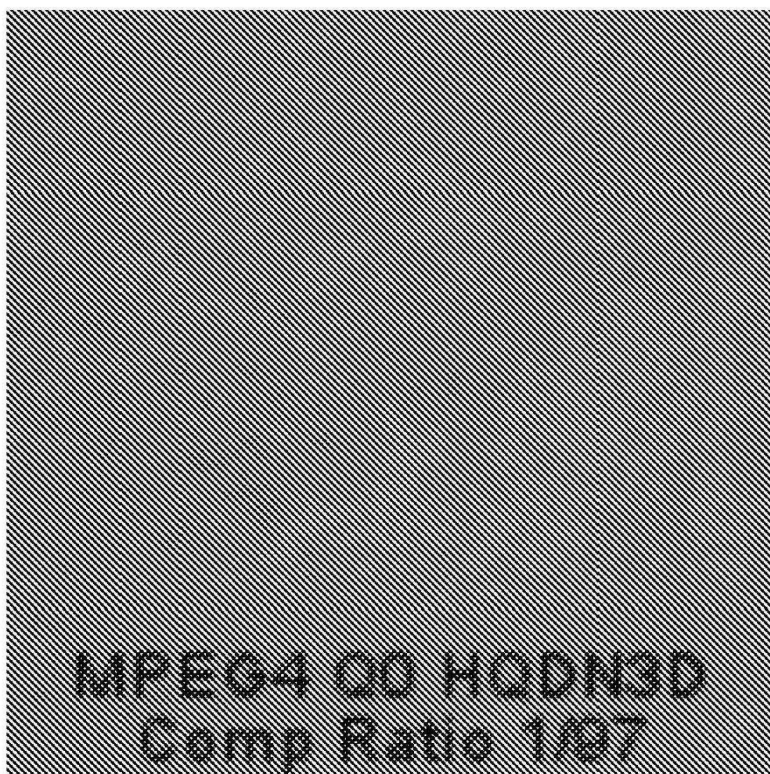




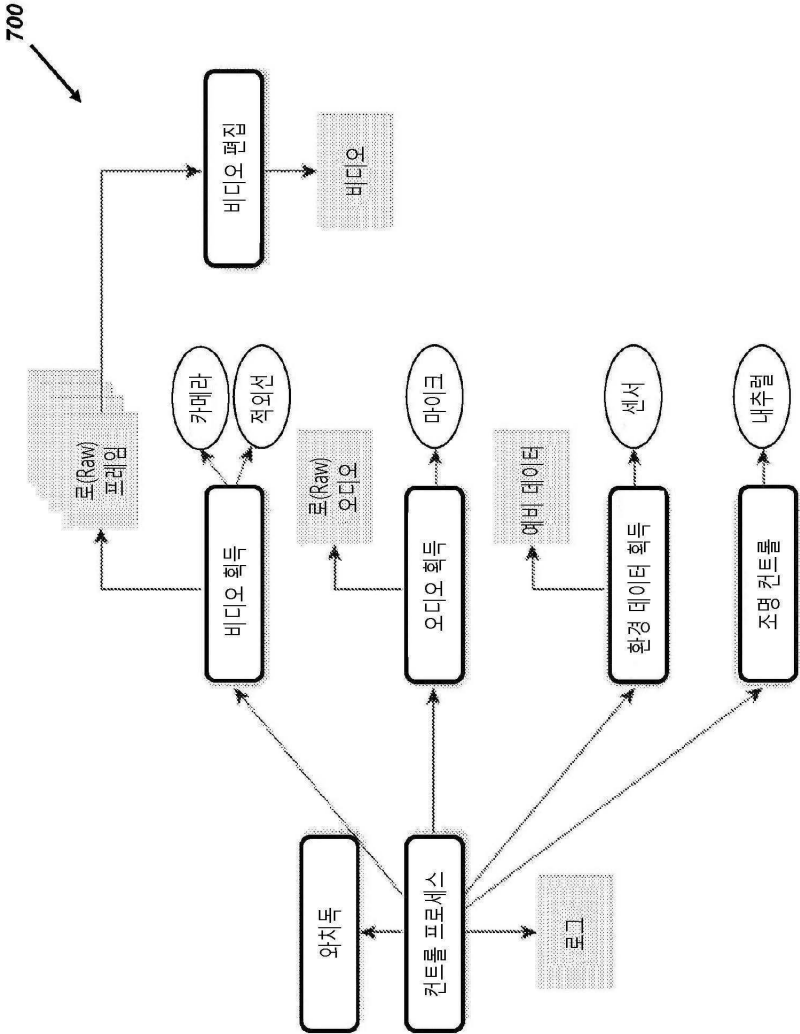
도면6c



도면6d

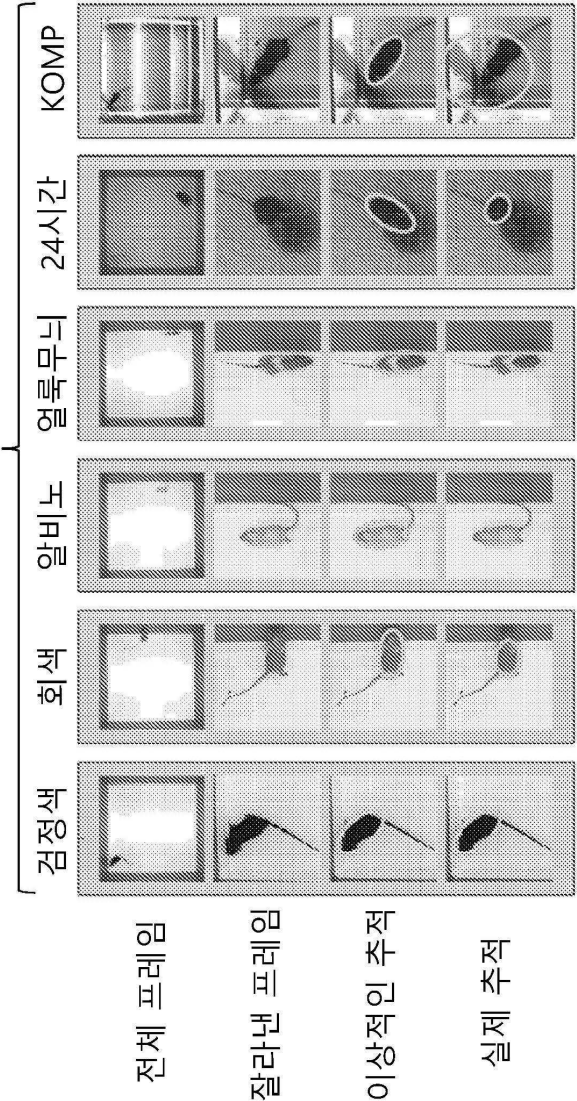


도면7

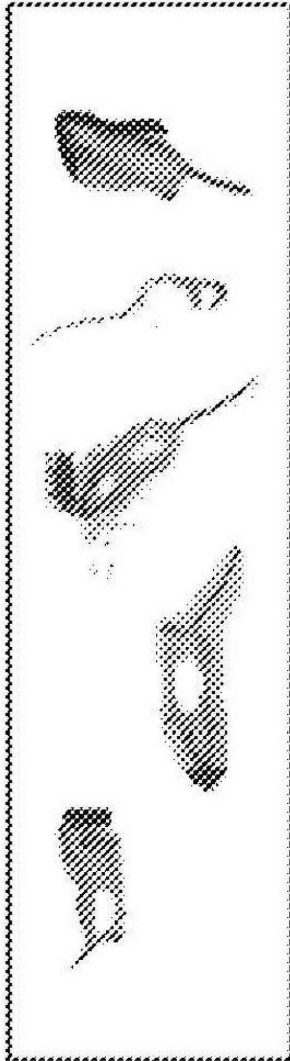




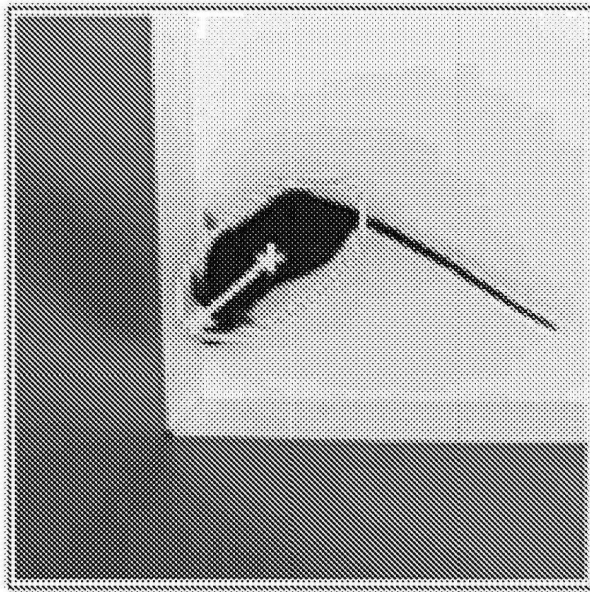
도면8a



도면8b

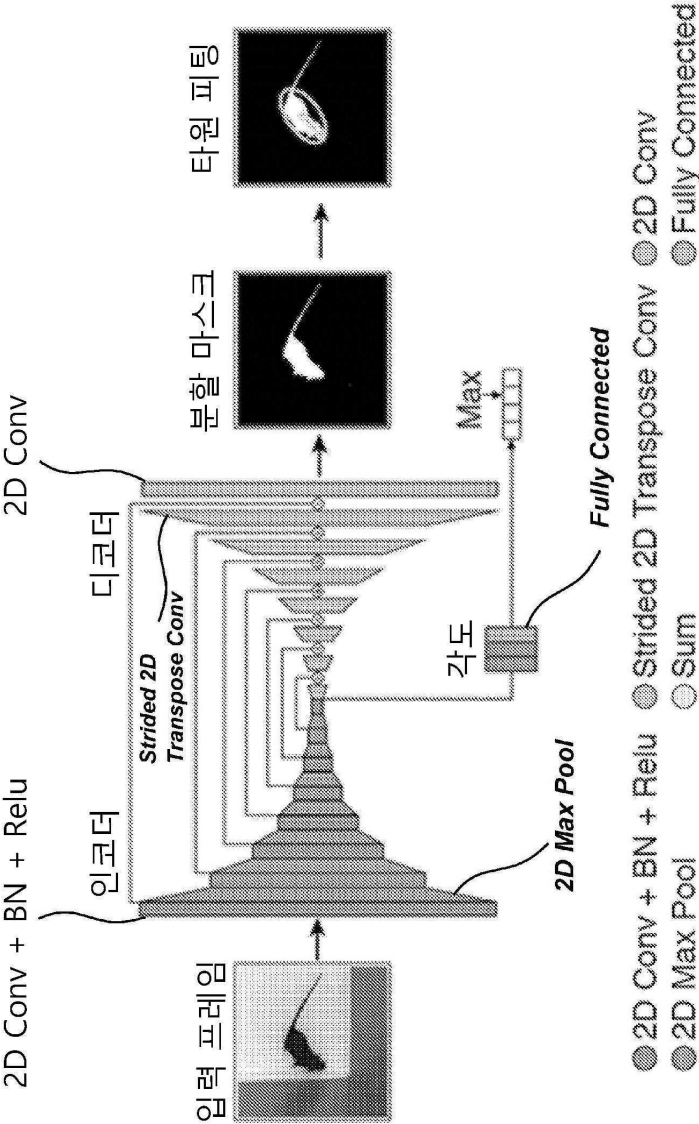


도면8c



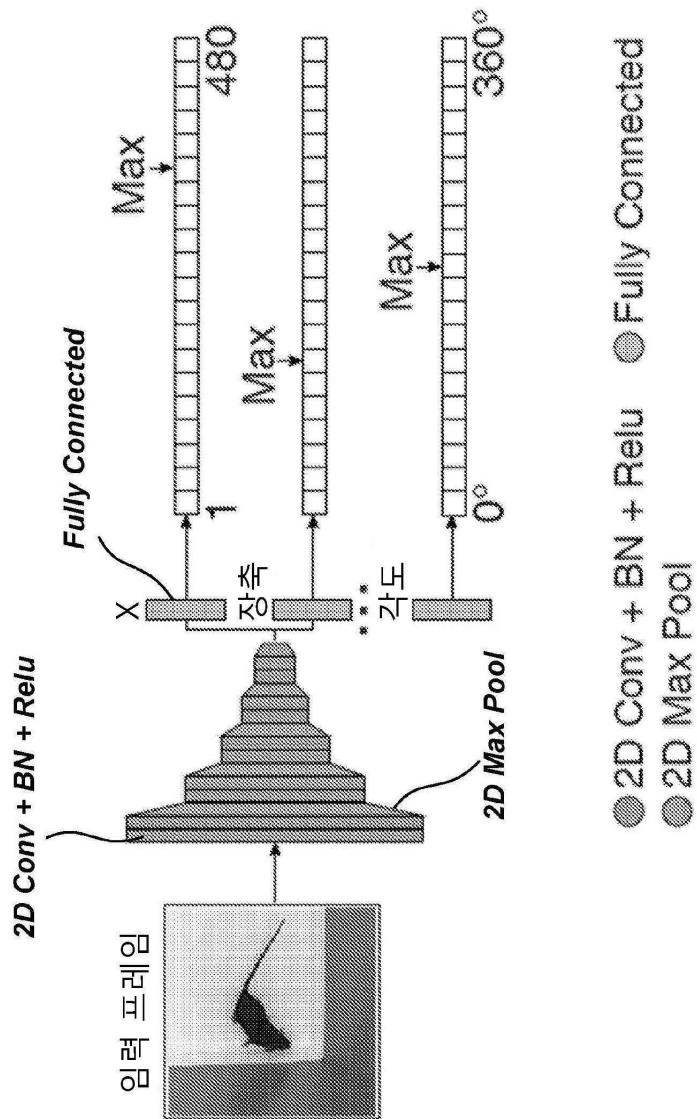
x좌표, y좌표, 장축, 단축, 각도

도면9

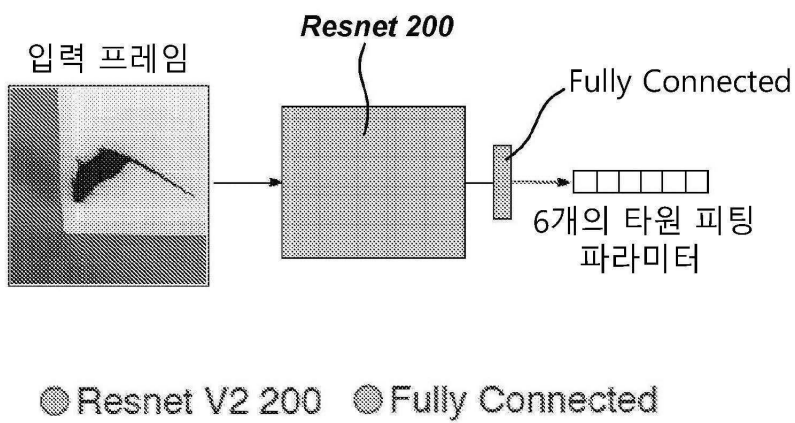




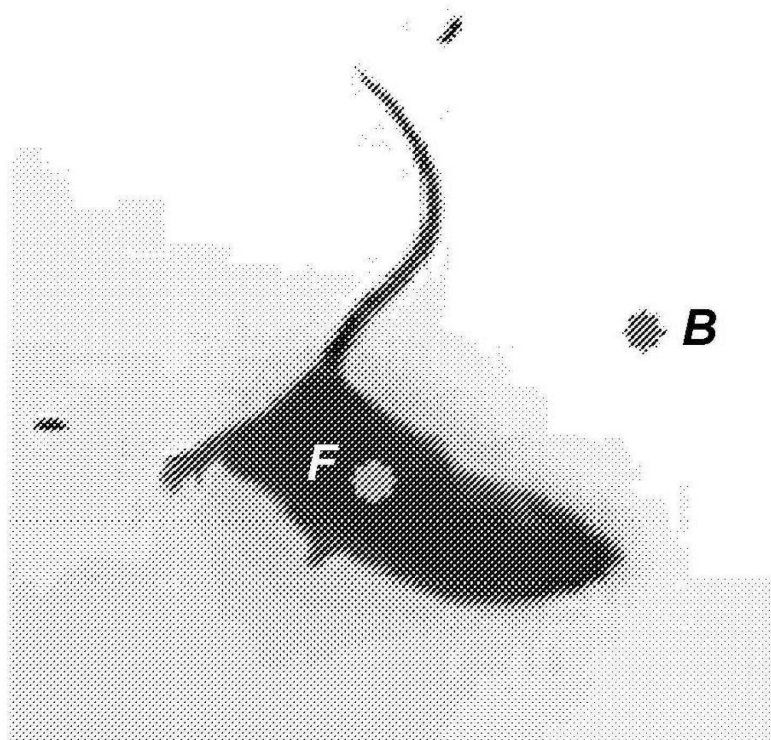
도면10



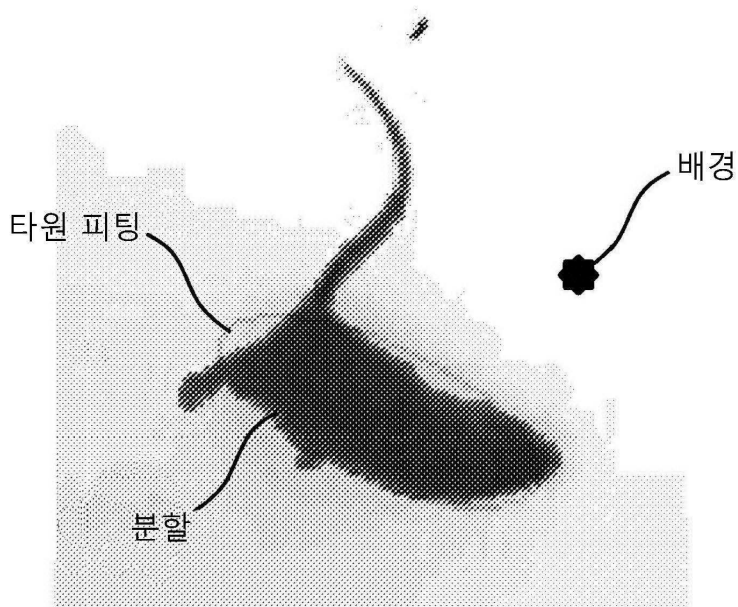
도면11



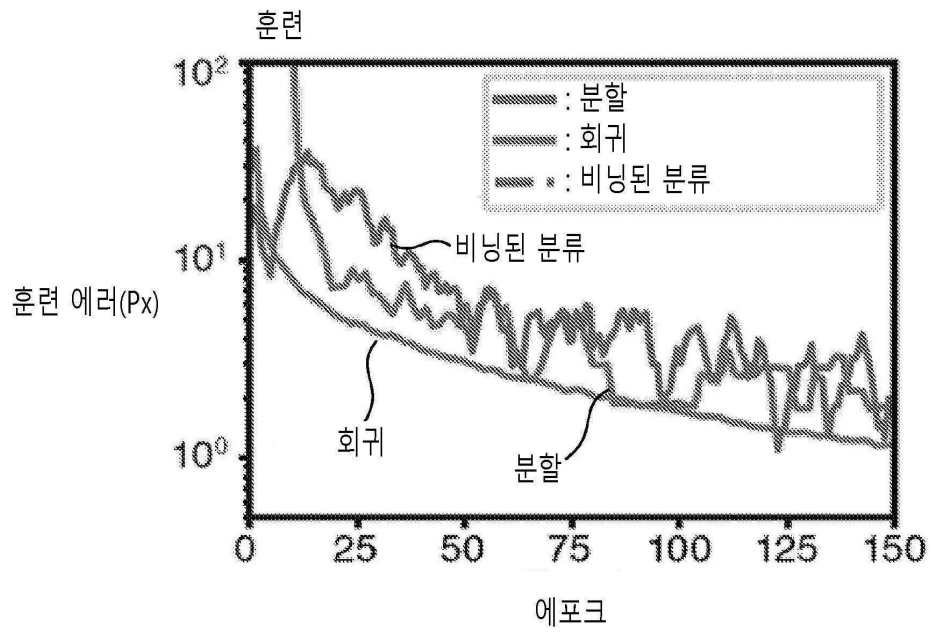
도면12a



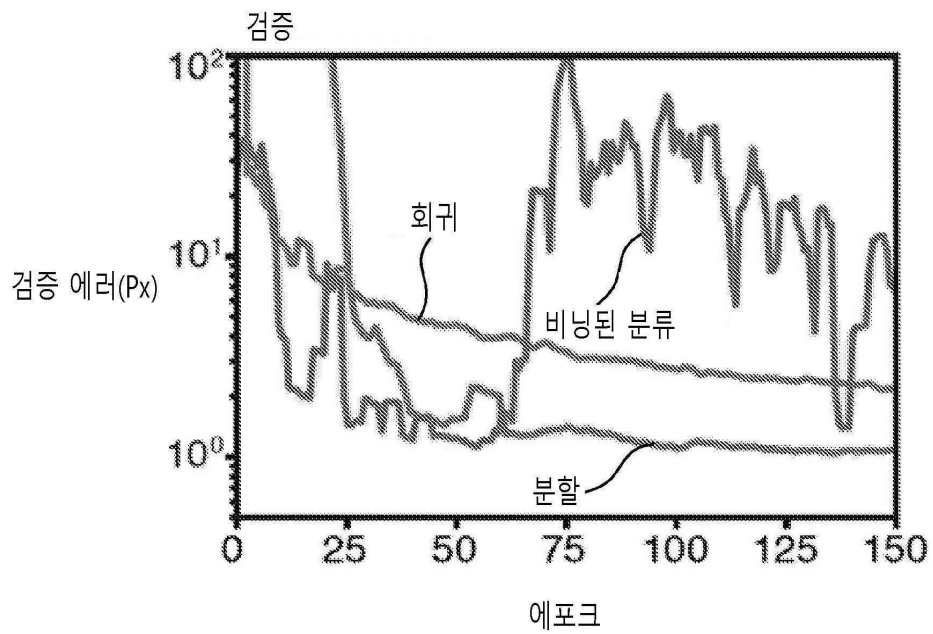
도면12b



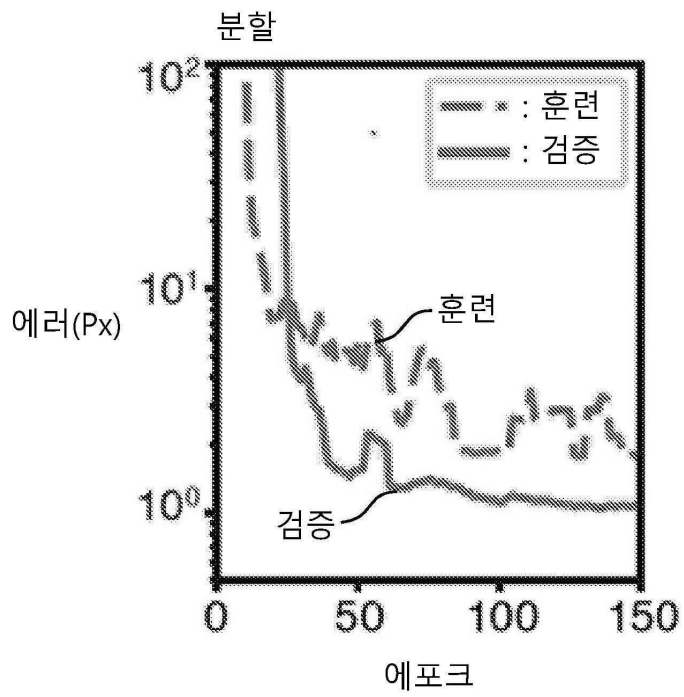
도면13a



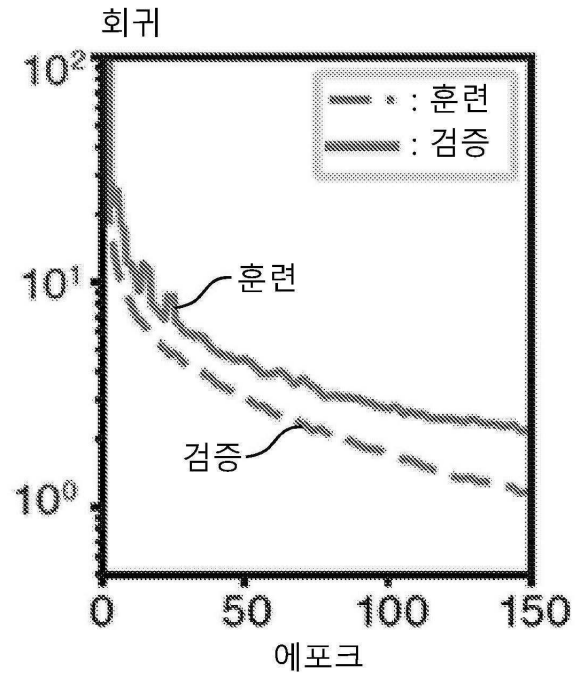
도면13b



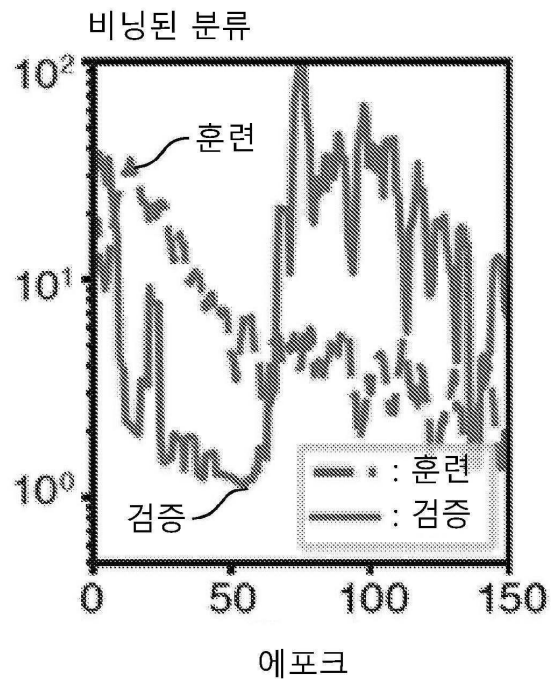
도면13c



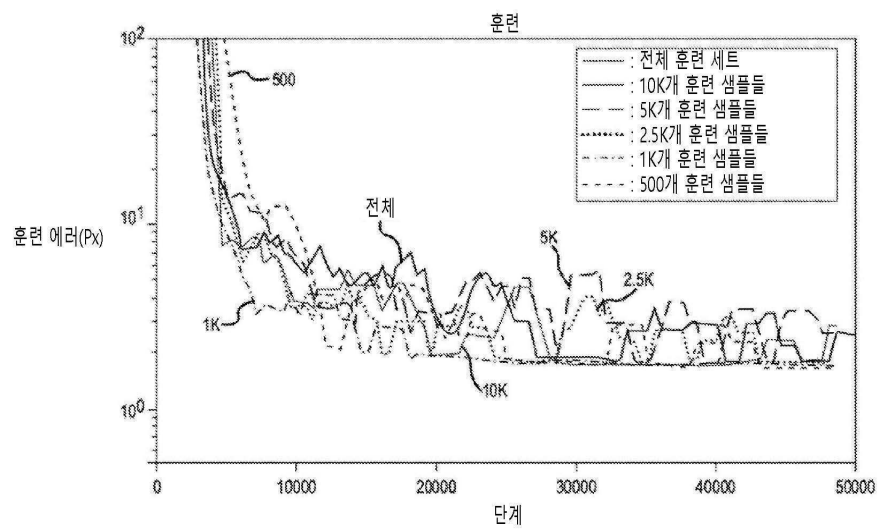
도면13d



도면13e

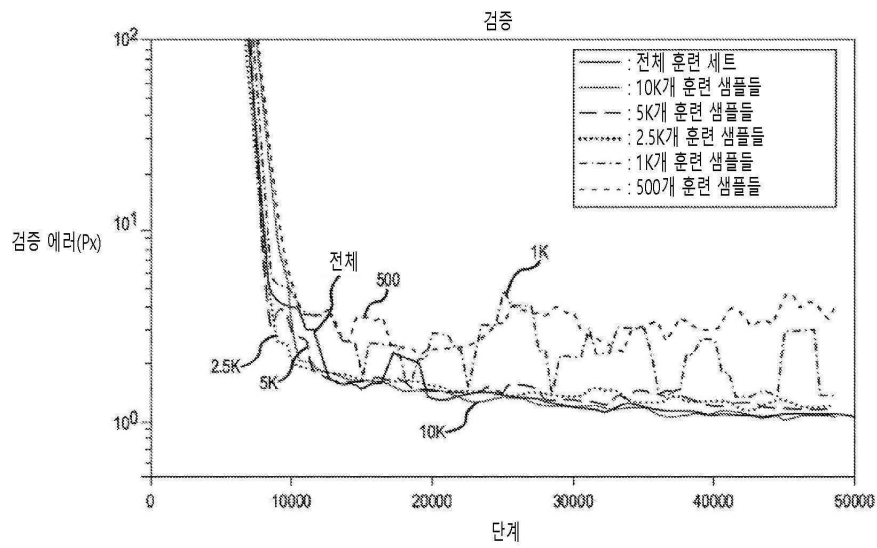


도면14a

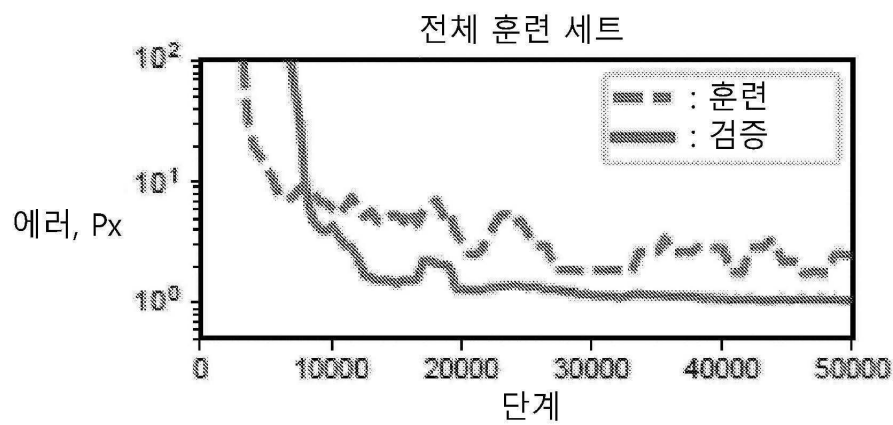




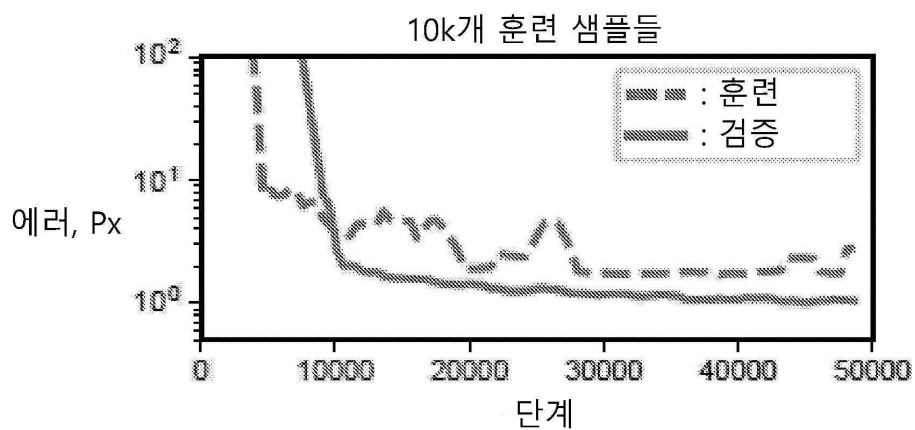
도면14b



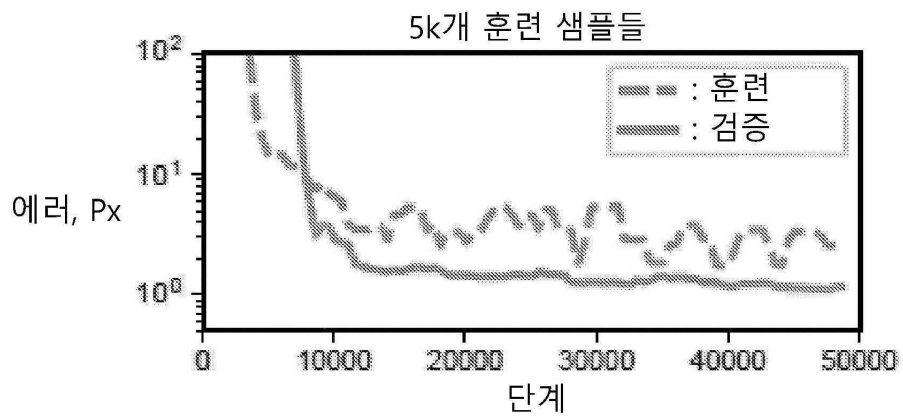
도면14c



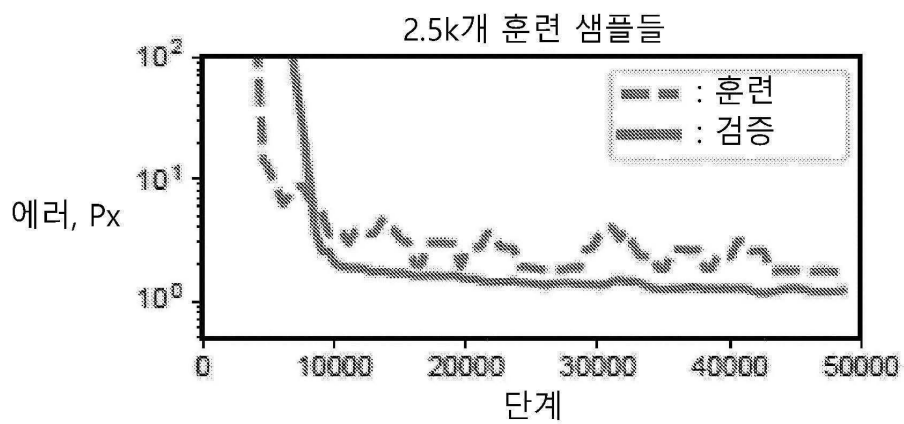
도면14d



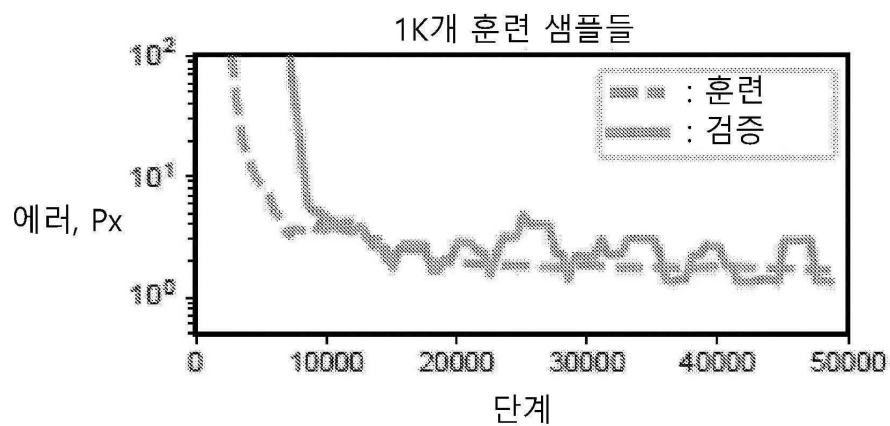
도면14e



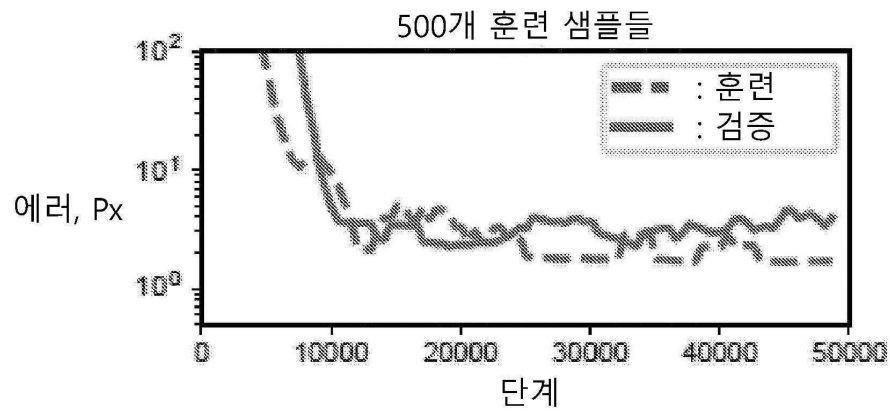
도면14f



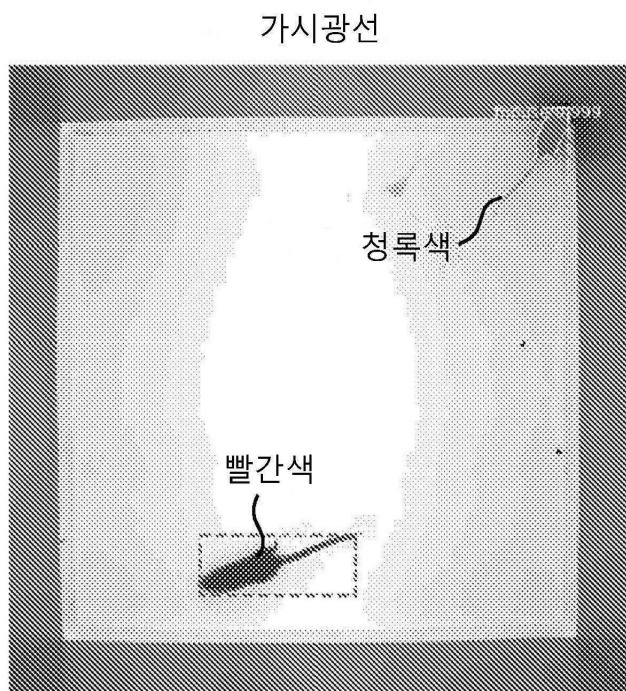
도면14g



도면14h

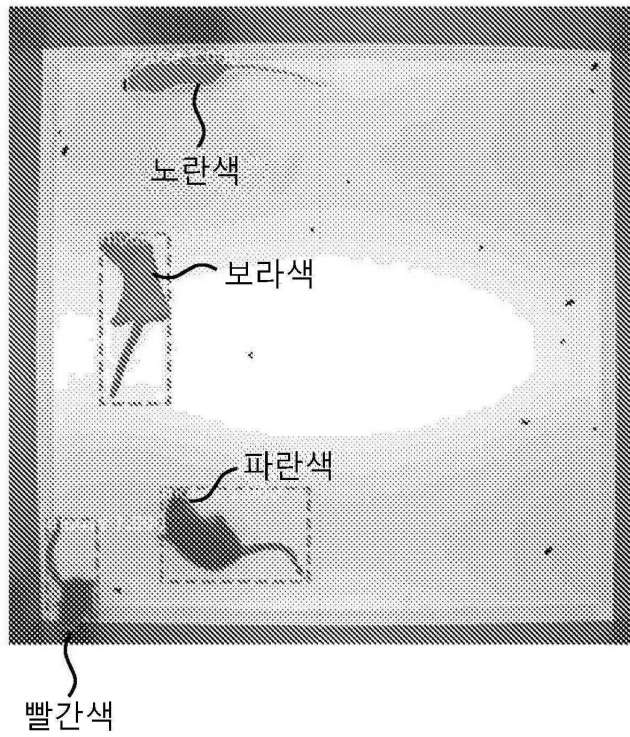


도면15a



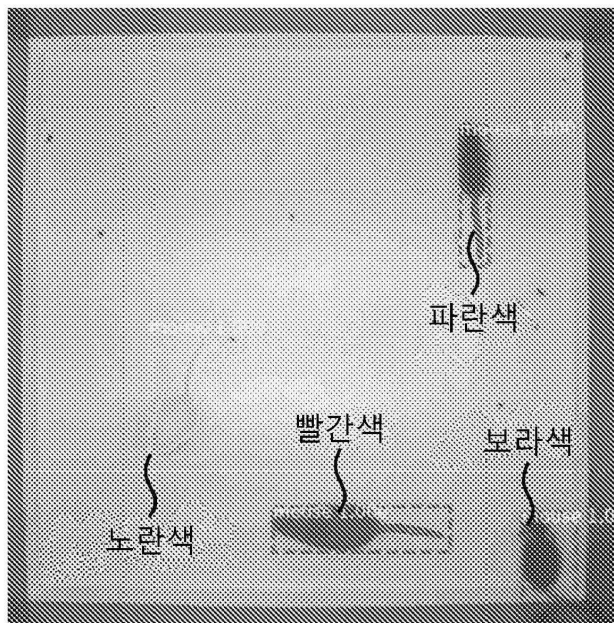


도면15b

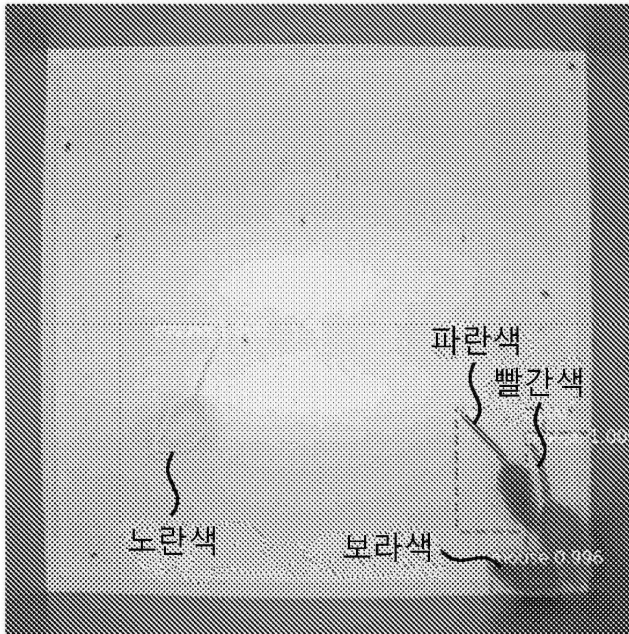


도면15c

적외선

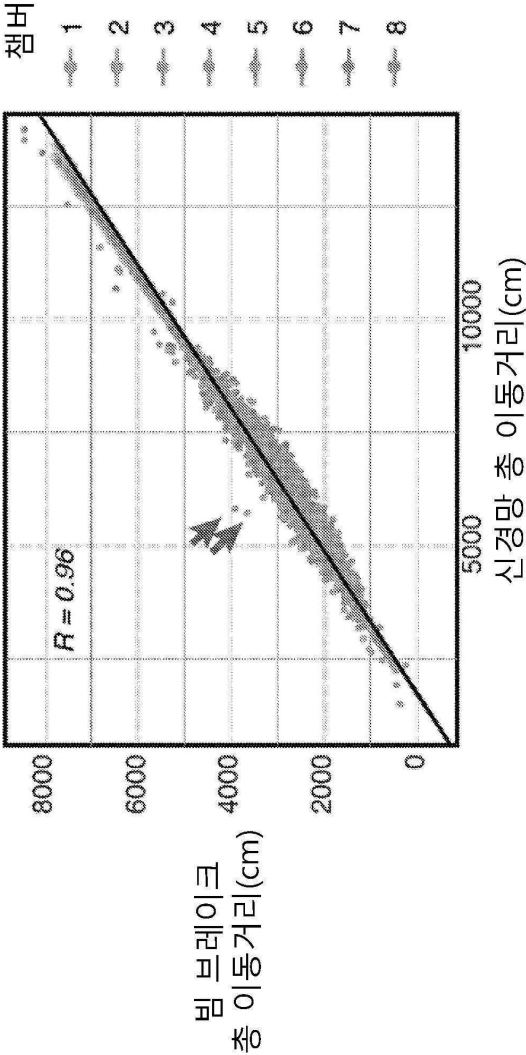


도면15d

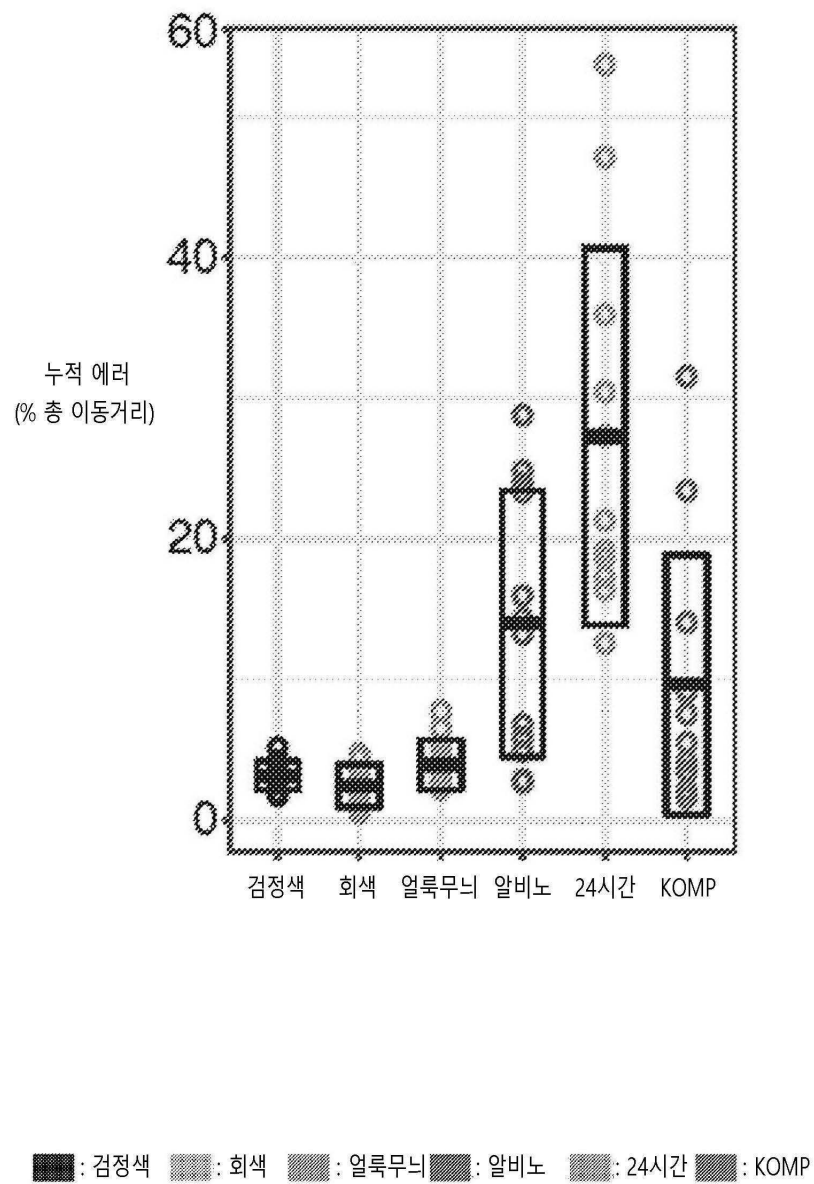




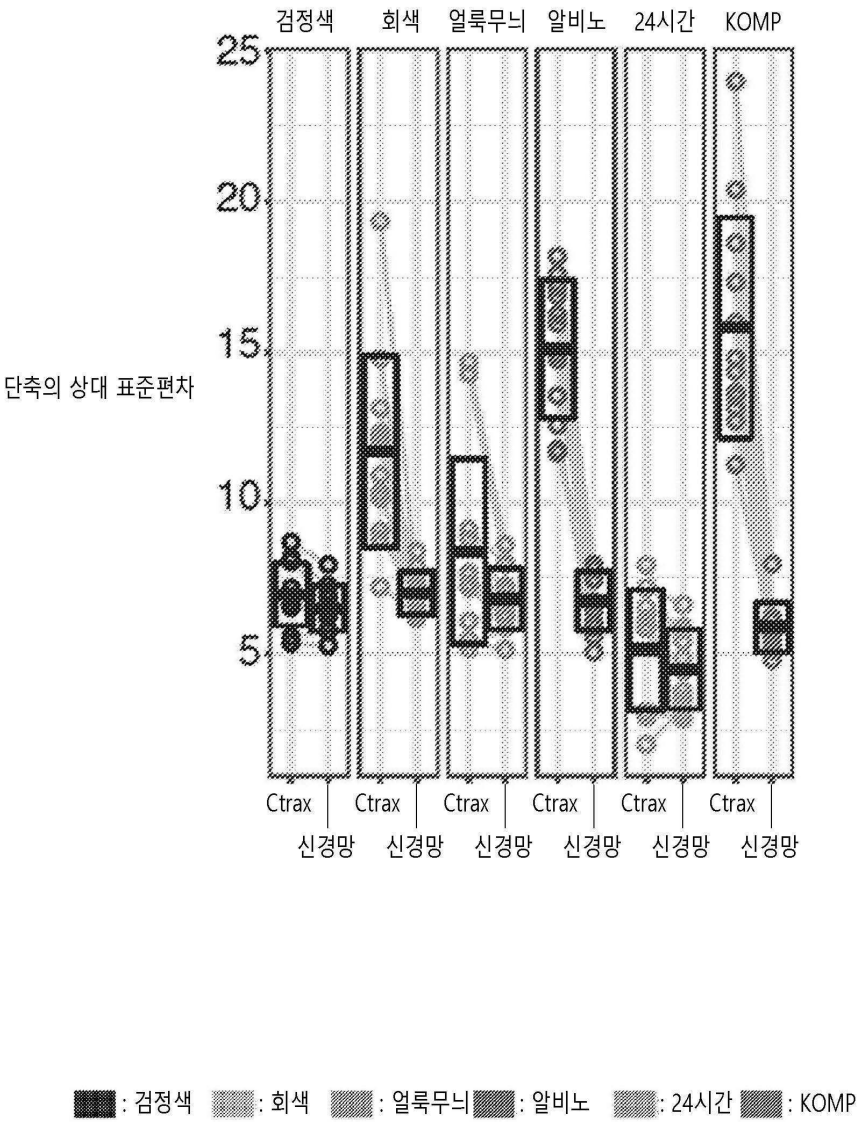
도면16



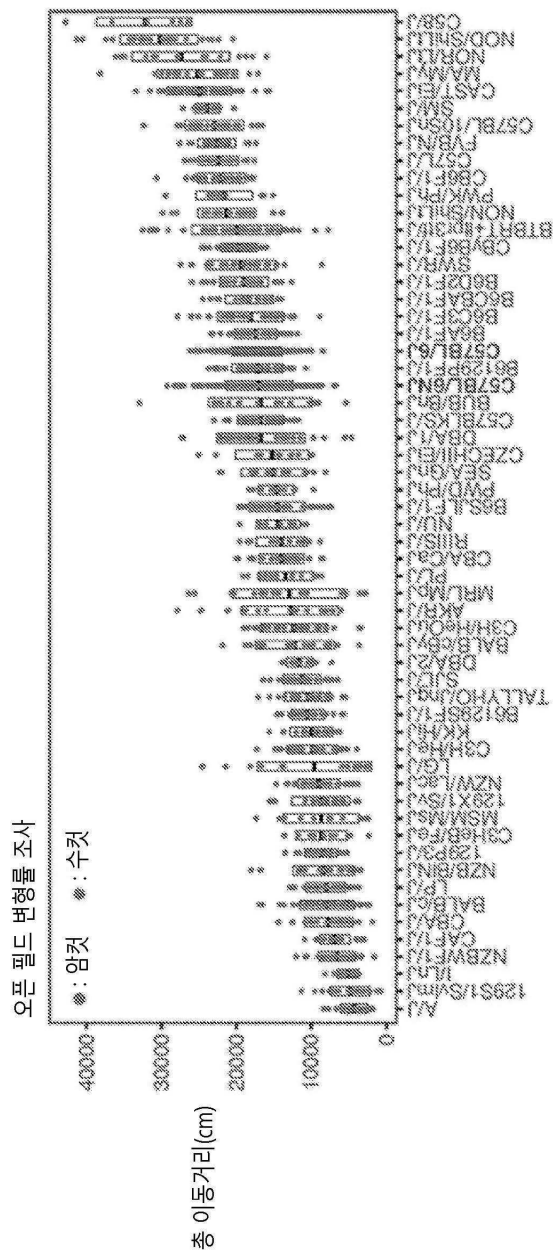
도면17a



도면17b



도면 18a





도면18b

