



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2020년03월23일
(11) 등록번호 10-2092475
(24) 등록일자 2020년03월17일

(51) 국제특허분류(Int. Cl.)
G06K 9/62 (2006.01) G06K 9/00 (2006.01)
G06N 3/08 (2006.01) G10L 17/26 (2013.01)
(52) CPC특허분류
G06K 9/627 (2013.01)
G06K 9/00362 (2013.01)
(21) 출원번호 10-2018-0122900
(22) 출원일자 2018년10월16일
심사청구일자 2018년10월16일
(56) 선행기술조사문헌
KR1020160098581 A*
KR1020180060257 A*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
고한석
서울특별시 용산구 이촌로88길 30, 101동 1402호 (이촌동, 삼성리버스위트아파트)
박상욱
제주특별자치도 제주시 과원로 27, 303동 703호(노형동, 부영아파트)
(74) 대리인
송인호, 윤형근, 최영중, 최관락

전체 청구항 수 : 총 4 항

심사관 : 강현일

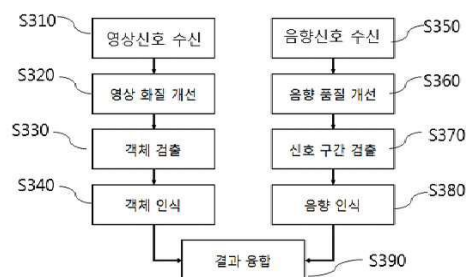
(54) 발명의 명칭 동물 중 식별 방법 및 장치

(57) 요약

본 발명은 시청각 정보를 이용하여 동물 종을 식별하는 방법 및 장치에 관한 것이다.

본 발명의 일실시예에 의한 동물 중 식별방법은 대상 객체에 대한 입력 신호를 수신하는 단계; 상기 입력 신호에 근거하여 영상정보 및 음향정보 처리하는 단계 상기 영상정보 처리 결과 및 상기 음향정보 처리 결과는 클래스별 스코어로 표현됨-; 상기 입력 신호에 대응되는 영상정보 처리 결과 및 음향정보 처리 결과가 존재하는지를 판단하는 단계; 및 상기 판단 결과에 따라 상기 영상정보 처리 결과 및 음향정보 처리 결과를 융합하고, 상기 융합된 처리 결과를 이용하여 상기 대상 객체가 어떤 동물 중 인지를 분류하는 최종 결과 도출 단계를 포함할 수 있다.

대 표 도 - 도3



(52) CPC특허분류

G06N 3/08 (2013.01)

G10L 17/26 (2013.01)

(72) 발명자

고경득

경기도 하남시 대청로59번길 20, 102동 403호(신장동, 더우개마을 통일아파트)

김동현

경상북도 경산시 압량면 청운로 75 오페라 101호

이 발명을 지원한 국가연구개발사업

과제고유번호 E41800021070100000000000000

부처명 환경부

연구관리전문기관 한국환경산업기술원

연구사업명 환경정책기반공공기술개발사업

연구과제명 인공지능 및 기계학습 기반 동물 울음소리를 통한 동물 종 구별 원천 기술개발

기 여 율 1/1

주관기관 고려대학교 산학협력단

연구기간 2018.01.01 ~ 2018.12.31

명세서

청구범위

청구항 1

대상 객체에 대한 입력 신호를 수신하는 단계;

상기 입력 신호에 근거하여 영상정보 및 음향정보 처리하는 단계 상기 영상정보 처리 결과 및 상기 음향정보 처리 결과는 클래스 별 스코어로 표현됨-;

상기 입력 신호에 대응되는 영상정보 처리 결과 및 음향정보 처리 결과가 존재하는지를 판단하는 단계; 및

상기 판단 결과에 따라 상기 영상정보 처리 결과 및 음향정보 처리 결과를 융합하고, 상기 융합된 처리 결과를 이용하여 상기 대상 객체가 어떤 동물 중 인지를 분류하는 최종 결과 도출 단계를 포함하되,

상기 최종 결과 도출 단계는

상기 영상정보 처리 결과 및 음향정보 처리 결과가 모두 존재하는 경우,

상기 영상정보 처리 결과의 스코어 및 상기 음향정보 처리 결과의 스코어를 융합하는 단계를 포함하되,

상기 최종 결과 도출 단계는 하기 수학식 1을 이용하여 도출하는 것을 특징으로 하는 동물 중 식별 방법.

[수학식 1]

$$finalresult = \operatorname{argmax}_c \left(\sum_{n=1}^N w_c^n \tilde{s}_c^n \right)$$

(c와 n은 각각 인식 대상과 시스템에 대한 색인이며, w_c^n 는 n번째 시스템에서 c번째 인식 대상에 대한 신뢰도를

의미하고, s_c^n 는 n번째 시스템에서 c번째 인식 대상에 대한 score를 의미하며, $\tilde{s}_c^n$ 는 s_c^n 를 $\sum_{c=1}^C s_c^n$ 로 나누어 정규화된 score를 의미한다.)

청구항 2

삭제

청구항 3

제1항에 있어서, 상기 최종 결과 도출 단계는

상기 영상정보 처리 결과 및 음향정보 처리 결과 중 어느 하나의 결과만 존재하는 경우,

상기 영상정보 처리 결과의 스코어 및 상기 음향정보 처리 결과의 스코어 wd 어느 하나를 이용하여 상기 대상 객체가 어떤 동물 중 인지를 분류하는 단계를 포함하는 것을 특징으로 하는 동물 중 식별 방법.

청구항 4

삭제

청구항 5

제1항에 있어서, 상기 수학식 1에서의 w_c^n 은 하기 수학식 2에 의해 산출되는 것을 특징으로 하는 동물 중 식별 방법.

[수학적식 2]

$$w_c^n = P(G=c|O=c) = \frac{P(G=c, O=c)}{\sum_g P(G=g, O=c)}$$

(여기서, G는 실제 값을 의미하고 O는 인식 시스템에서 얻은 인식 결과를 의미한다.)

청구항 6

제1항에 있어서, 상기 영상정보 및 음향정보 처리 단계는

CNN(Convolutional Neural Network) 기법을 이용하는 처리하는 단계를 포함하는 것을 특징으로 하는 동물 종 식별 방법.

발명의 설명

기술 분야

[0001] 본 발명은 동물 종 식별 방법 및 장치에 관한 것으로, 보다 상세하게는 시청각 정보를 이용하여 동물 종을 식별하는 방법 및 장치에 관한 것이다.

배경 기술

[0002] UN을 포함한 각 국가에서는 국내뿐만 아니라 국제적으로 동물들의 생산, 유통, 이동 등의 과정에서 발생할 수 있는 모든 리스크를 통제하기 위해서, 효과적이면서도 신뢰성이 높은 동물개체 추적(tracking) 및 식별 시스템에 관한 제도를 운영하고자 노력하고 있으며, 최근 들어서는 전통적인 방법 이외에 발달된 정보화 기술을 통해서 보다 나은 시스템을 구축하고자 다양한 시도 및 연구를 진행하고 있다. 관련된 선행문헌으로 공개특허 공보 제10-2014-0138103호가 있다.

[0003] 또한, 관련된 기술로 영상 인식 기반의 동물을 감지하는 방법이 있다. 이는 미리 설정해 둔 감시영역을 복수의 센서를 이용하여 동물들이 있는지 감지할 수 있는 기술이다. 즉, 동물들이 감시영역에 들어올 경우 이를 감지하고 하우징 등의 경고장치를 작동시킬 수 있다.

[0004] 그러나 상기 기술은 감시영역에 동물이 들어오는 것을 감지할 수 있다. 그러나 해당 동물이 동물인지 아닌지는 판별이 가능하지만, 어떠한 종인지는 구분이 불가능하다.

[0005] 따라서 다양한 동물 종에 대해 보다 정확하고 편리하게 구별할 수 있는 기술에 대한 연구가 필요한 실정이다.

발명의 내용

해결하려는 과제

[0006] 본 발명의 목적은 대상 객체에 대한 시청각 정보를 이용하여 인공지능 방식으로 동물 종을 식별할 수 있는 동물 종 식별 방법을 제공하는 데 있다.

과제의 해결 수단

[0007] 상기 목적을 달성하기 위해 본 발명의 일실시예에 의하면, 대상 객체에 대한 입력 신호를 수신하는 단계; 상기 입력 신호에 근거하여 영상정보 및 음향정보 처리하는 단계 상기 영상정보 처리 결과 및 상기 음향정보 처리 결과는 클래스 별 스코어로 표현됨-; 상기 입력 신호에 대응되는 영상정보 처리 결과 및 음향정보 처리 결과가 존재하는지를 판단하는 단계; 및 상기 판단 결과에 따라 상기 영상정보 처리 결과 및 음향정보 처리 결과를 융합하고, 상기 융합된 처리 결과를 이용하여 상기 대상 객체가 어떤 동물 종 인지를 분류하는 최종 결과 도출 단계를 포함하는 동물 종 식별 방법이 개시된다.

발명의 효과

[0008] 본 발명의 일실시예에 의한 동물 종 식별 방법은 대상 객체의 시청각 정보를 이용하여 보다 편리하게 동물 종을 식별할 수 있다.

[0009] 본 발명의 일실시예에 의하면, 인공지능 방식으로 동물 종을 식별하기에 자생 동물을 조사할 수 있는 인프라 구

축 및 부족한 전문가의 인력을 보완할 수 있다.

도면의 간단한 설명

[0010] 도 1은 본 발명의 일실시예와 관련된 동물 중 식별 방법을 설명하기 위한 블록도이다.

도 2는 본 발명의 일실시예와 관련된 동물 중 식별 장치를 나타낸다.

도 3은 본 발명의 일실시예와 관련된 동물 중 식별 방법을 나타내는 흐름도이다.

도 4는 본 발명의 일실시예와 관련된 대기 광학 모델 구조를 설명하기 위한 도면이다.

도 5는 본 발명의 일실시예와 관련된 영상에서 비, 눈 제거 방법을 설명하기 위한 도면이다.

도 6은 본 발명의 일실시예와 관련된 음향 신호 인식을 위한 CNN 구조도이다.

도 7은 본 발명의 일실시예와 관련된 동물 중 식별 방법에서 영상정보와 음향정보를 융합하는 방법을 나타내는 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0011] 이하, 본 발명의 일실시예와 관련된 동물 중 식별 방법 및 장치에 대해 도면을 참조하여 설명하도록 하겠다.

[0012] 본 명세서에서 사용되는 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "구성된다" 또는 "포함한다" 등의 용어는 명세서상에 기재된 여러 구성 요소들, 또는 여러 단계들을 반드시 모두 포함하는 것으로 해석되지 않아야 하며, 그 중 일부 구성 요소들 또는 일부 단계들은 포함되지 않을 수도 있고, 또는 추가적인 구성 요소 또는 단계들을 더 포함할 수 있는 것으로 해석되어야 한다.

[0013] 도 1은 본 발명의 일실시예와 관련된 동물 중 식별 방법을 설명하기 위한 블록도이다.

[0014] 도시된 바와 같이, 식별서버(100)와 동물 중 식별장치(200)는 네트워크 망을 통해 연결될 수 있다. 네트워크 망은 매체를 통해 통신을 수행하는 망을 말하는 것으로, 유선 통신망, 컴퓨터 네트워크 및 무선 통신망 중 적어도 하나를 포함할 수 있다.

[0015] 상기 식별서버(100)는 상기 동물 중 식별장치(200)로부터 특정 동물에 대한 움직임 등을 촬영한 영상정보 및 특정 동물의 울음소리 등을 녹음한 음향정보를 수신할 수 있다. 상기 수신된 정보에 근거하여 상기 식별서버(100)는 상기 특정 동물이 어떤 동물 종인지를 식별하고, 그 결과를 상기 동물 중 식별장치(200)로 전송할 수 있다.

[0016] 상기 식별서버(100)는 통신부(110), 영상 처리부(120), 음향 처리부(130), 융합부(140) 및 최종 분류부(150)를 포함할 수 있다.

[0017] 상기 통신부(110)는 상기 동물 중 식별장치(200)에 입력된 특정 동물에 대한 영상신호 및 음향신호 등을 수신하고, 수신한 입력신호를 기반으로 최종 분류된 특정 동물에 대한 식별 결과를 상기 동물 중 식별장치(200)로 전송할 수 있다.

[0018] 상기 영상 처리부(120)는 수신된 특정 동물에 대한 영상신호를 기반으로 객체를 인식할 수 있다.

[0019] 상기 음향 처리부(130)는 수신된 특정 동물에 대한 음향신호를 기반으로 상기 특정 동물의 음향을 인식할 수 있다.

[0020] 상기 융합부(140)는 상기 영상 처리부(120)에서 처리된 영상정보 처리 결과와 상기 음향 처리부(130)에서 처리된 음향정보 처리 결과를 융합할 수 있다.

[0021] 상기 최종 분류부(150)는 상기 융합부(140)에서 융합한 결과에 근거하여 특정 동물이 어떤 동물 종인지 최종적으로 분류할 수 있다.

[0022] 도 2는 본 발명의 일실시예와 관련된 동물 중 식별 장치를 나타낸다. 이하, 실시예에서는 설명되는 동물 중 식별장치는 이동형 장치를 나타내나, 고정형 장치로 구현될 수도 있다.

[0023] 도 2(a)는 일방향에서 바라본 동물 중 식별 장치(200)의 사시도이고, 도 2(b)는 도 2(a) 상태의 동물 중 식별장치(200)를 윗면과 아랫면이 바뀌도록 뒤집어 놓은 상태의 사시도이고, 도 2(c)는 도 2(a) 상태의 동물 중 식별장치(200)를 앞측면과 뒤측면이 바뀌도록 위치시킨 상태의 사시도이고, 도 2(d)는 도 2(b) 상태의 동물 중 식별

장치(200)를 앞측면과 뒤측면이 바뀌도록 위치시킨 상태의 사시도이다.

- [0024] 상기 동물 중 식별장치(200)는 터치스크린(1), 제1카메라(2), 제2카메라(3), 마이크(4), 방열공(5), 전원스위치(6), 데이터단자(7)를 포함할 수 있다.
- [0025] 상기 터치스크린(1)은 사용자의 명령을 입력 받음과 동시에 상기 식별서버(100)로부터 수신한 인공지능 식별방법의 분석결과 및 관련 정보를 출력할 수 있다.
- [0026] 상기 제1카메라(2) 및 제2카메라(3)는 인공지능 식별방법으로 분석될 영상 또는 이미지 신호를 촬영할 수 있다. 상기 제1카메라(2) 및 제2카메라(3)는 동일한 종류의 카메라일 수도 있고, 서로 다른 종류의 카메라일 수도 있다. 예를 들어, 상기 제1카메라(2) 및 제2카메라(3) 중 어느 하나가 CCD 카메라이고, 다른 하나는 적외선 카메라일 수 있다.
- [0027] 마이크(4)는 인공지능 식별방법으로 분석될 음향 신호를 입력받을 수 있다. 상기 마이크(4)는 복수 개의 마이크가 소정 간격을 이루도록 배열될 수 있다. 예를 들어, 도시된 바와 같이 네 개의 마이크로 구성되며, 각 마이크의 간격이 2.5cm로 마이크가 배열될 수 있다.
- [0028] 상기 방열공(5)은 동물 중 식별장치(200)가 구동됨에 따라 발생하는 열을 방출할 수 있다.
- [0029] 상기 전원스위치(6)는 동물 중 식별장치(200)의 전원을 온 시키거나 오프시킬 수 있다.
- [0030] 상기 데이터단자(7)는 기기에 내장된 저장메모리에 접근할 수 있는 단자로써, 이를 통해 데이터를 업로드/다운로드 할 수 있다.
- [0031] 상기 동물 중 식별 장치(200)는 제1카메라(2) 또는 제2카메라(3)로부터 대상 객체(또는 대상 동물)의 영상신호를 입력받고, 마이크(4)로부터 대상 객체(또는 대상 동물)의 음향신호를 입력받을 수 있다. 상기 입력받은 영상신호 및 음향신호는 식별서버(100)로 전송될 수 있다. 상기 식별서버(100)는 수신한 영상신호 및 음향신호를 각각 처리하고, 각각 처리된 결과를 융합하여 상기 대상 객체가 어떤 동물 종인지를 최종적으로 분류할 수 있다.
- [0032] 도 3은 본 발명의 일실시예와 관련된 동물 중 식별 방법을 나타내는 흐름도이다.
- [0033] 상기 식별서버(100)의 통신부(110)는 동물 중 식별장치(200)에 입력된 영상신호를 수신할 수 있다(S310).
- [0034] 상기 영상 처리부(120)는 수신한 영상신호를 처리하여 영상정보 처리 결과를 도출할 수 있다.
- [0035] 먼저, 상기 영상 처리부(120)는 영상 화질 개선을 수행할 수 있다(S320). 화질 개선을 위해 수신한 영상신호에서 안개/연기를 제거하는 방법이 사용될 수 있다.
- [0036] 도 4는 본 발명의 일실시예와 관련된 대기 광학 모델 구조를 설명하기 위한 도면이다.
- [0037] 도 4에서 I는 관측 데이터, x는 영상 내 픽셀 좌표, J는 안개가 제거된 영상, t는 전달량, A는 전역 대기 산란광을 의미한다. 이러한 대기 광학 모델에 바탕한 안개/연기 제거를 위한 방법은 지역적 대기 산란광 추정 단계, 지역적 전달량 추정 단계, 대기 산란광 및 전달량 정련 단계로 구성된다.
- [0038] 지역적 대기 산란광 추정에서는 입력 영상을 gray 영상으로 변환하고, MxM 블록으로 분할한 뒤, 수학적 식 1과 같이 최소값을 찾는다.

수학적 식 1

[0039]
$$T_k^{block} = \min_{y \in block} T_k^{block}(y)$$

[0040] 이때, L은 나뉜 개별 블록을 의미하며, y는 블록 내 픽셀 좌표를 의미한다.

[0041] Quad-tree subdivision을 통해 $\| (rx, gx, bx) - (1, 1, 1) \|$ 값을 최소화하는 하안 산란광 값 Allow_threshold를 추정한다. 하안 산란광 값을 통해 최종적으로 지역적 산란광을 수학적 식 2를 통해 추정한다.

수학식 2

$$\mathbf{A}(\mathbf{x}) = \begin{cases} \mathbf{A}_{local}(\mathbf{x}), & \|\mathbf{1} - \mathbf{A}_{local}(\mathbf{x})\| \leq \rho \cdot \|\mathbf{1} - \mathbf{A}_{low_threshold}\| \\ \mathbf{A}_{low_threshold}, & otherwise \end{cases}$$

[0042]

[0043] 전달량 추정 단계에서 전달량은 영상 엔트로피, 정보 신뢰성, 다이내믹 레인지의 함수로 이루어진 목적함수, 하기 수학식 3을 최대화 하는 값으로 추정한다.

수학식 3

$$\begin{aligned} f_{objective}(t) &= f_{entropy}(t) \cdot f_{fidelity}(t) \cdot f_{range}(t), \\ f_{entropy}(t) &= -\sum_{i=0}^{255} \frac{h_i(t)}{N} \log \frac{h_i(t)}{N}, \\ f_{fidelity}(t) &= \min_{c \in \{r, g, b\}} s^c(t), \quad s^c(\mathbf{x}) = \frac{1}{N} \sum_{\mathbf{x}=1}^N \delta(\mathbf{x}), \quad \delta(\mathbf{x}) = \begin{cases} 1, & 0 \leq J^c(\mathbf{x}) \leq 255 \\ 0, & otherwise \end{cases} \\ f_{range}(t) &= \max(J^{gray}(t)) - \min(J^{gray}(t)) \end{aligned}$$

[0044]

[0045] 여기서, N은 전체 화소의 개수, hi는 i 화소값(intensity)의 개수이다.

[0046] 안개 및 연기의 경우, 거리에 따라 그 농도가 다르기 때문에, 목적 함수 (수학식 3)를 블록에 따라 최대화하여 전달량을 추정하며, 최종적으로 k번째 블록의 추정된 전달량은 하기 수학식 4와 같다.

수학식 4

$$t_k^{block} = \arg \max_{t \in \{0.01 \leq t \leq 1\}} f_{objective}(t),$$

[0047]

[0048] 지역 블록 단위로 추정된 산란광 및 전달량을 정련하는 방법으로 본 발명에서는 WLS (Weighted Least Squares) 최적화 방법을 통해 경계 영역에서 발생하는 블록 결함 (block artifact)을 정련한다. WLS 기반 정련 방법은 목적 함수 (하기 수학식 5)를 최소화하는 해를 구한다.

수학식 5

$$\begin{aligned} \sum_{\mathbf{x}} \left((t_{\mathbf{x}} - \tilde{t}_{\mathbf{x}})^2 + \lambda \left(w_{a,\mathbf{x}}(h) \left(\frac{\partial t}{\partial a} \right)_{\mathbf{x}}^2 + w_{b,\mathbf{x}}(h) \left(\frac{\partial t}{\partial b} \right)_{\mathbf{x}}^2 \right) \right), \\ w_{a,\mathbf{x}}(h) = \left(\left| \frac{\partial h}{\partial a} \right|_{\mathbf{x}}^{\alpha} + \varepsilon \right)^{-1}, \quad w_{b,\mathbf{x}}(h) = \left(\left| \frac{\partial h}{\partial b} \right|_{\mathbf{x}}^{\alpha} + \varepsilon \right)^{-1}, \end{aligned}$$

[0049]

[0050] 이때, $\tilde{t}$ 는 수학식 4를 통해 추정된 결과, t는 정련될 결과 영상이다. h는 그레이로 변환된 안개 영상을 통해 모폴로지 처리된 영상이며, λ 는 스무딩 조절 변수이다. a, b는 각 영상의 가로 및 세로 방향을 나타내며 a는 가

중치 조정 변수이다. ε 는 분모가 0으로 나누어지는 것을 방지하기 위한 미세 상수 값이다. 수학식 5의 미분을 통한 해는 하기 수학식 6와 같다.

수학식 6

$$t = (\mathbf{I} + \lambda(\mathbf{D}_a^T \mathbf{A}_a \mathbf{D}_a + \mathbf{D}_b^T \mathbf{A}_b \mathbf{D}_b))^{-1} \tilde{t},$$

[0051]

상기 수학식 6에서 A는 각 가로, 세로 방향에 대해 식 11의 가중치 값이 대각 성분에 포함된 행렬이며 D는 1차 미분 행렬 연산자이다.

[0052]

최종적으로 추정된 지역 산란광 및 전달량을 통해 안개가 제거된 영상을 하기 수학식 7을 통해 복원한다.

[0053]

수학식 7

$$\mathbf{J}(\mathbf{x}) = \frac{\mathbf{I}(\mathbf{x}) - \mathbf{A}(\mathbf{x})}{t(\mathbf{x})} + \mathbf{A}(\mathbf{x})$$

[0054]

또한, 화질 개선을 위해 수신한 영상신호에서 안개/연기를 제거하는 방법이 사용될 수 있다.

[0055]

도 5는 본 발명의 일실시예와 관련된 영상에서 비, 눈 제거 방법을 설명하기 위한 도면이다.

[0056]

제1카메라(2) 또는 제2카메라(3)를 통해 연속적으로 관측한 이미지 데이터에서 객체의 이동성을 바탕으로 움직이는 객체를 검출하고 움직임이 없는 배경을 제거한다. 검출된 영역의 크기를 바탕으로 눈/비에 대한 객체를 선정한다. 선정된 후보군에 대해 Histogram of Orientation (HOG)에 기반하여 눈/비 객체를 판단한다. 최종 검출된 객체를 눈/비에 대한 객체로 판단하고 관측데이터에서 제거한다.

[0057]

상기와 같이 복원된 영상을 통해 영상 처리부(120)는 객체를 검출하고, 검출된 객체를 인식할 수 있다(S330, S3340).

[0058]

상기 객체 검출은 화질이 개선된 이미지에서 객체들의 위치와 영역을 인식하는 단계이고, 객체 인식은 검출된 객체가 무엇인지 인식하는 단계이다. 이 두 단계를 위해 CNN(Convolutional Neural Network)에 기반한 RCNN(Region Convolutional Neural Network)이 적용된다. 상기 RCNN은 CNN과 같이 이미지를 고정된 크기의 filter를 이용하여 convolution 연산 결과를 바탕으로 객체의 위치와 객체가 무엇인지를 인식한다. 최종 output layer에서 영상 신호에 대한 클래스(class)별 스코어(score)가 산출된다.

[0059]

한편, CNN은 convolutional layer와 pooling layer, fully-connected layer로 구성된다.

[0060]

상기 convolutional layer에서는 2차원 데이터에 고정된 크기의 2차원 필터를 슬라이딩 하며, convolution 연산을 수행한다. 이때, 복수의 필터를 사용할 수 있으며, 이 경우에는 3차원 필터를 사용하여 convolution을 수행하며, 그 결과는 3차원으로 표현된다. 일반적인 CNN에서는 다수의 필터를 적용하여 관측데이터를 다각도로 분석하여 결과를 추출한다.

[0061]

상기 Pooling layer는 convolutional layer 이후, 결과는 “가로x세로x필터 수”의 차원을 갖는다. 일반적인 2차원 데이터에서 인근 영역의 정보는 서로 유사성이 높아 인근 영역에서 대표값을 선정함으로써 차원을 줄일 수 있다. 이러한 과정이 pooling layer에서 수행된다.

[0062]

상기 fully-connected layer는 관측데이터가 복수의 convolution-pooling layer를 따라 전파된 후, 마지막 output layer에 연결하기 위한 마지막 layer를 의미한다. convolution - pooling 연산을 마무리한 결과 도출되는 3차원 데이터를 1차원 데이터로 변형한 이후, output layer와 완전히 연결된 NN를 의미한다.

[0063]

한편, NN는 1차원 관측 데이터를 입력받아 전파되는 반면, CNN은 2차원 관측 데이터를 입력받는다. 관측 데이터가 영상/이미지의 경우 관측 데이터를 바로 CNN에 입력하여 전파할 수 있지만, 음향인 경우 1차원 데이터를 2차원 데이터로 변형해야 한다. 일반적으로, 1차원 음향 데이터는 Short Time Fourier Transform (STFT)를 적용하여 spectrogram으로 변환하고, 저주파에 민감한 인간의 청각 특성을 반영한 Mel-filter를 적용하고 log scale로 변형함으로써, 2차원 log-Mel spectrum이 변환되어 CNN에 입력된다.

[0064]

- [0065] 상기 식별서버(100)의 통신부(110)는 동물 중 식별장치(200)에 입력된 음향신호를 수신할 수 있다(S350).
- [0066] 상기 음향 처리부(130)는 수신한 음향신호를 처리하여 음향정보 처리 결과를 도출할 수 있다.
- [0067] 먼저, 상기 음향 처리부(120)는 음향 품질 개선을 수행할 수 있다(S360).
- [0068] 상기 마이크(4)를 통해 입력된 음향신호에서 잡음을 줄이기 위한 방법으로 빔포밍 방법이 적용된다. 빔포밍(미도시)은 특정 방향에서 입력된 신호를 추출하는 방법으로 음향신호 입력부, 위상 차 계산부, 잡음 성분 추출부, 신호 크기 추측부, 잡음제거 계수 산출부로 구성될 수 있다.
- [0069] 상기 음향신호 입력부는 마이크(4)에 입력된 신호를 디지털로 변환하는 부분으로 개별 마이크에서 입력된 신호를 이하 제1~4신호라고 명명한다. 음향 신호의 방향에 따라 제1~4신호의 위상에 차이가 발생하는 특성을 이용하여, 방향을 추측하기 위해, 위상 차 계산부에서 제1~4신호의 위상 차이를 계산한다. 잡음 성분 추출부는 계산된 위상 차이 이외의 신호를 잡음으로 고려하여 이를 분리한다. 신호 크기 추측부는 제1~4신호에서 추출된 잡음을 제거함으로써, 신호 크기 성분을 추측한다. 끝으로, 잡음제거 계수 산출부는 추출된 잡음과 신호크기 성분의 비율을 잡음제거 계수로 결정하고 제1~4신호에서 이를 빼내어 관측 신호를 개선한다.
- [0070] 상기 음향 처리부(130)는 상기 개선된 관측 신호에서 필요한 신호 구간을 검출할 수 있다(S370).
- [0071] 상기 음향 처리부(130)는 음향 품질이 개선된 음향 신호에서 신호 구간을 검출하기 위해 변형된 Double Fourier Transform Voice Activity Detection (DFVAD) 알고리즘을 적용할 수 있다. 음향 품질이 개선된 1차원 음향 신호에 Short Time Fourier Transform (STFT)를 적용하여 2차원 이미지(spectrogram)로 변환한 후, 각 주파수 축에서 다시 한번 Fast Fourier Transform (FFT)를 적용하여 각 주파수의 시간적 변화에 대한 정보를 추출한다. 각 주파수에서 관심 대상(동물 개체군)의 울음소리 주파수 대역 별 정보를 바탕으로 최종적으로 신호의 시작지점과 끝지점을 검출함으로써 신호 구간을 판단한다.
- [0072] 그리고 음향 처리부(130)는 검출된 신호 구간을 통해 음향을 인식한다(S380).
- [0073] 상기 음향 처리부(130)는 검출된 음향 신호를 인식하기 위해, CNN 기법을 적용할 수 있다. 예를 들어, 검출된 신호 구간의 음향 신호에 STFT를 적용하여 spectrogram으로 변환한 후, Mel-filter를 적용하고, log scale로 변환하여 log-mel spectrum을 생성한다. 이때, log-mel spectrum이 음향 신호 인식을 위한 CNN으로 입력되어, convolution, pooling, fully-connected layer를 통과하여 최종 output layer에서 음향 신호에 대한 class별 score가 산출된다.
- [0074] 도 6은 본 발명의 일실시예와 관련된 음향 신호 인식을 위한 CNN 구조도이다.
- [0075] 도시된 5개의 convolutional layer 와 2개의 fully-connected layer로 구성된다. 각각의 convolution layer에서 filter의 크기는 3×3으로 모두 동일하고, 첫 번째, 두 번째, 그리고 마지막 convolutional layer 이후에 2×2 max-pooling 을 적용했다. 그림에서 숫자는 CNN을 통해 전파되면서 데이터의 차원 변화를 나타낸다. 최종 output layer의 차원은 인식 대상의 수에 의해 결정되며, 도 6에서는 총 9종을 인식하는 CNN구조의 예시를 보여 주며, 인식 대상의 수가 증가할 경우, 최종 output layer의 차원이 증가하며, 내부 구조도 변형될 수 있다.
- [0076] 상기 융합부(140)는 상기 영상 처리부(120) 및 음향 처리부(130)에서 처리된 결과를 융합할 수 있다(S390).
- [0077] 도 7은 본 발명의 일실시예와 관련된 동물 중 식별 방법에서 영상정보와 음향정보를 융합하는 방법을 나타내는 흐름도이다.
- [0078] 상기 식별서버(100)의 융합부(140)는 상기 영상 처리부(120)에서 처리된 영상정보 처리 결과 및 음향 처리부(130)에서 처리된 음향정보 처리 결과가 모두 존재하는지를 판단할 수 있다(S710). 상기 영상정보 처리 결과 및 상기 음향정보 처리 결과는 클래스 별 스코어로 표현될 수 있다.
- [0079] 만약, 두 가지 결과가 모두 존재한다면, 상기 융합부(140)는 영상정보 처리 결과에 대한 스코어와 음향정보 처리 결과에 대한 스코어를 융합할 수 있다(S720).
- [0080] 그러나 두 가지 결과 중 어느 하나만 존재하는 경우는 결과가 도출된 정보 처리 결과에 의존하여 최종 결과가 결정된다.
- [0081] 예를 들어, 영상정보 처리 결과만 존재하는 경우는 영상정보를 호출하여 대상 객체가 어떤 동물 종인지를 최종 분류할 수 있다(S730, S740).

[0082] 또한, 음향정보 처리 결과만 존재하는 경우는 음향정보를 호출하여 대상 객체가 어떤 동물 종인지를 최종 분류할 수 있다(S730, S750).

[0083] 한편, 두 가지 결과가 모두 존재할 경우, 영상/음향 score 융합 단계에서 각각의 score를 융합하여 수학식 8에 따라 최종 결과를 도출한다.

수학식 8

[0084]

$$finalresult = \operatorname{argmax}_c \left(\sum_{n=1}^N w_c^n \tilde{s}_c^n \right)$$

[0085] c와 n은 각각 인식 대상과 시스템에 대한 색인이며, w_c^n 는 n번째 시스템에서 c번째 인식 대상에 대한 신뢰도를 의미하고, s_c^n 는 n번째 시스템에서 c번째 인식 대상에 대한 score를 의미하며, $\tilde{s}_c^n$ 는 s_c^n 를 $\sum_{c=1}^C s_c^n$ 로 나누어 정규화된 score를 의미한다.

[0086] 상기 수학식 8은 각 시스템에서 각각의 인식 대상에 대한 score에 신뢰도를 부여하여 더했을 경우, 가장 높은 score를 가지는 인식 대상으로 인식 결과를 도출한다.

[0087] 각 시스템에서 도출된 결과에 대한 신뢰도 w_c^n 은 하기 수학식 9와 같이 산출된다.

수학식 9

[0088]

$$w_c^n = P(G=c|O=c) = \frac{P(G=c, O=c)}{\sum_g P(G=g, O=c)}$$

[0089] 이때, G는 실제 값을 의미하고 O는 인식 시스템에서 얻은 인식 결과를 의미한다. 즉, w_c^n 은 인식 시스템의 결과가 c일 때, 실제 값도 c일 조건부확률 $P(G=c|O=c)$ 를 의미하며, 이는 bayesian rule에 의해 joint 확률로 표현할 수 있다. 이는, 어느 한 시스템에서 특정 인식 대상으로 오인식되는 경우가 적을 경우 높은 신뢰도를 부여하고, 그렇지 않을 경우 낮은 신뢰도를 부여한다.

[0090] 전술한 바와 같이, 본 발명의 일실시예에 의한 본 발명의 일실시예에 의한 동물 종 식별 방법은 대상 객체의 시청각 정보를 이용하여 보다 편리하게 동물 종을 식별할 수 있다.

[0091] 본 발명의 일실시예에 의하면, 인공지능 방식으로 동물 종을 식별하기에 자생 동물을 조사할 수 있는 인프라 구축 및 부족한 전문가의 인력을 보완할 수 있다.

[0092] 상술한 동물 종 식별방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터로 판독 가능한 기록 매체에 기록될 수 있다. 이때, 컴퓨터로 판독 가능한 기록매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 한편, 기록매체에 기록되는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지된 것일 수도 있다.

[0093] 컴퓨터로 판독 가능한 기록매체에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(Magnetic Media), CD-ROM, DVD와 같은 광기록 매체(Optical Media), 플롭티컬 디스크(Floptical Disk)와 같은 자기-광 매체(Magneto-Optical Media), 및 롬(ROM), 램(RAM), 플래시 메모리, SSD (Solid State Drive)와 같은 메모리 저장장치 등 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다.

[0094] 한편, 이러한 기록매체는 프로그램 명령, 데이터 구조 등을 지정하는 신호를 전송하는 반송파를 포함하는 광 또는 금속선, 도파관 등의 전송 매체일 수도 있다.

[0095] 또한, 프로그램 명령에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 상술한 하드웨어 장치는 본 발명의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

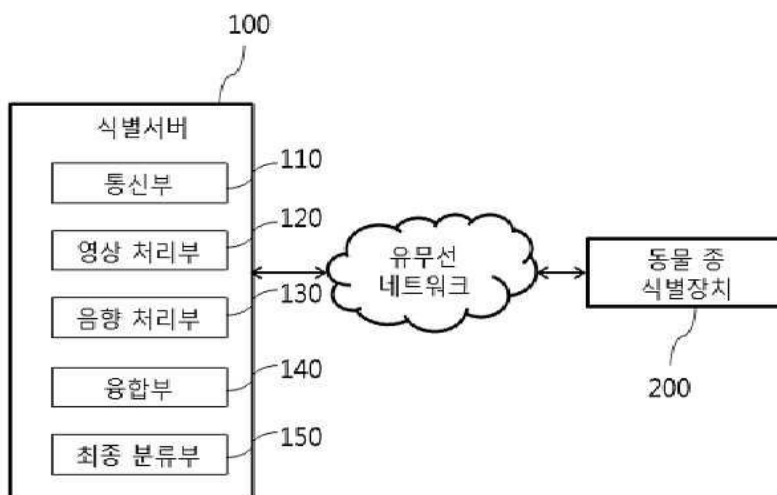
[0096] 상기와 같이 설명된 동물 종 식별 방법 및 장치는 상기 설명된 실시예들의 구성과 방법이 한정되게 적용될 수 있는 것이 아니라, 상기 실시예들은 다양한 변형이 이루어질 수 있도록 각 실시예들의 전부 또는 일부가 선택적으로 조합되어 구성될 수도 있다.

부호의 설명

[0097] 100: 식별서버
110: 통신부
120: 영상 처리부
130: 음향 처리부
140: 융합부
150: 최종 분류부
200: 동물 종 식별장치

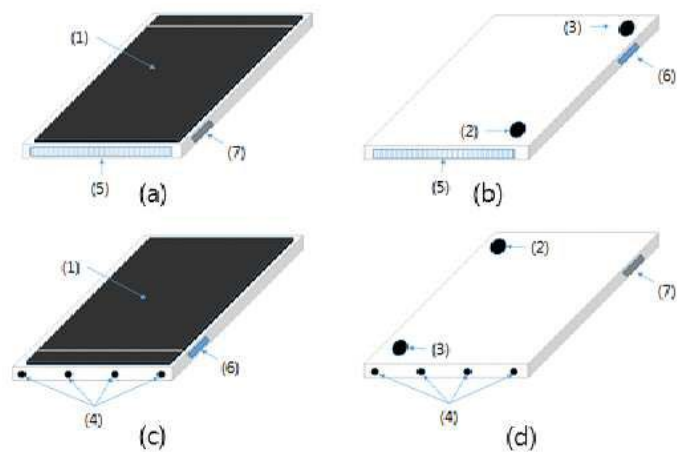
도면

도면1

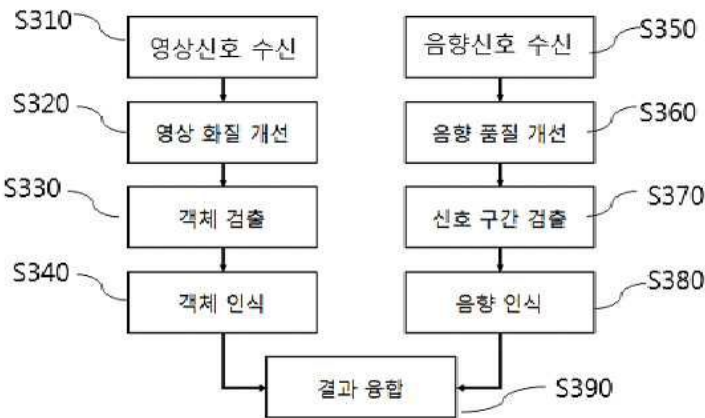


도면2

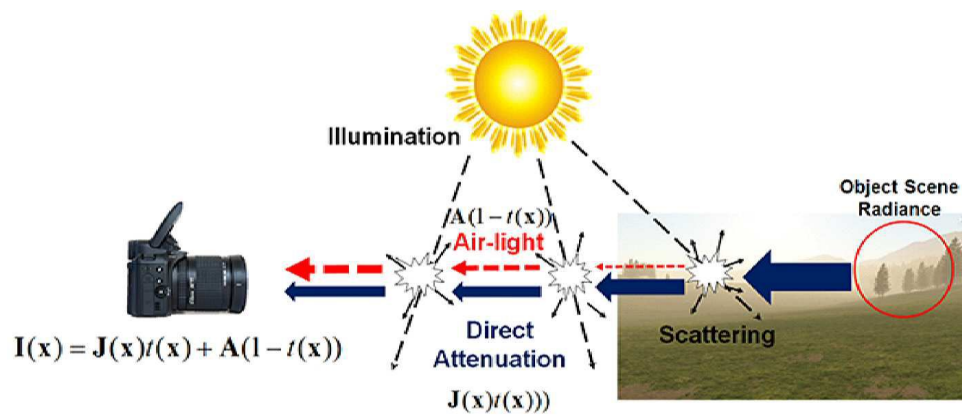
200



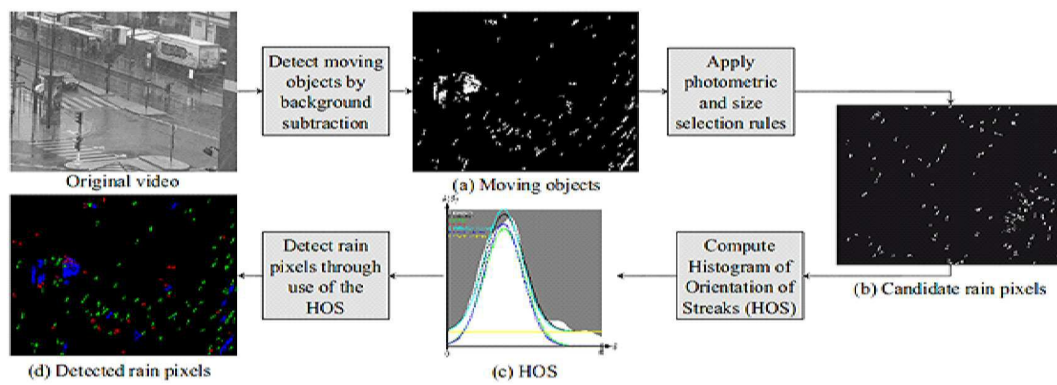
도면3



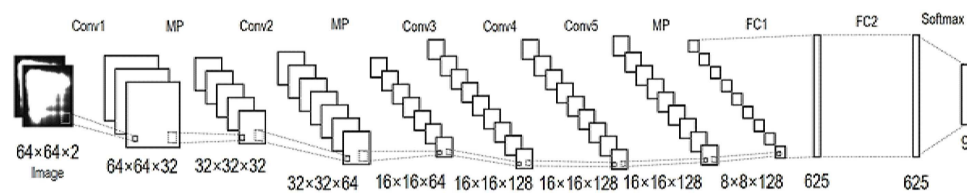
도면4



도면5



도면6



도면7

