

딥러닝 및 기계학습 활용 반려견 얼굴 정면판별 방법

김종복¹, 장동화¹, 양가영¹, 권경석¹, 김종곤¹, 이준환^{2*}
¹농촌진흥청 국립축산과학원, ²전북대학교 컴퓨터공학부

Recognition of dog's front face using deep learning and machine learning

Jong-Bok Kim¹, Dong-Hwa Jang¹, Yang Kayoung¹, Kyeong-Seok Kwon¹,
Jung-Kon Kim¹, Joon-Whoan Lee^{2*}

¹National Institute of Animal Science, Rural Development Administration

²Devison of Computer Science and Engineering, Jeonbuk National University

요약 반려견을 키우는 가구 수가 급격하게 증가함에 따라 유기, 유실견도 많이 증가하고 있다. 국내에서는 2014년부터 반려동물 등록제를 시행하고 있지만, 안전성과 실효성 문제로 등록률이 높지 않은 실정이다. 이러한 문제를 해결할 방법으로 반려견 생체인식 기술이 주목을 받고 있다. 생체인식률을 높이기 위해서는 최대한 정면에서 같은 형태로 생체정보 이미지를 수집해야 한다. 하지만 반려견은 사람과 달리 비협조적이기 때문에 생체정보 이미지 수집이 어렵다. 본 논문에서는 반려견 생체인식에 적합한 생체정보 이미지 수집을 위해 실시간 영상에서 반려견 얼굴 방향이 정면인지를 판별하는 방법을 제안한다. 제안 방법은 딥러닝을 활용하여 반려견 눈과 코를 검출하고, 검출된 눈과 코의 상대적 크기와 위치를 통해 5가지의 얼굴 방향 정보를 추출하여 기계학습 분류기로 정면 여부를 판별한다. 2,000개의 반려견 이미지를 분류하여 학습, 검증 및 테스트에 사용하였다. 눈과 코 검출에는 YOLOv3와 YOLOv4를 사용하였고, 분류기는 MLP(Multi-layer Perceptron), RF(Random Forest), SVM(Support Vector Machine)을 사용하였다. YOLOv4와 RF 분류기를 사용하고 제안하는 5가지 얼굴 방향 정보 모두를 적용하였을 때 얼굴 정면 판별 성능이 95.25%로 가장 좋았으며, 실시간 처리도 가능한 것으로 나타났다.

Abstract As pet dogs rapidly increase in number, abandoned and lost dogs are also increasing in number. In Korea, animal registration has been in force since 2014, but the registration rate is not high owing to safety and effectiveness issues. Biometrics is attracting attention as an alternative. In order to increase the recognition rate from biometrics, it is necessary to collect biometric images in the same form as much as possible—from the face. This paper proposes a method to determine whether a dog is facing front or not in a real-time video. The proposed method detects the dog's eyes and nose using deep learning, and extracts five types of directional face information through the relative size and position of the detected face. Then, a machine learning classifier determines whether the dog is facing front or not. We used 2,000 dog images for learning, verification, and testing. YOLOv3 and YOLOv4 were used to detect the eyes and nose, and Multi-layer Perceptron (MLP), Random Forest (RF), and the Support Vector Machine (SVM) were used as classifiers. When YOLOv4 and the RF classifier were used with all five types of the proposed face orientation information, the face recognition rate was best, at 95.25%, and we found that real-time processing is possible.

Keywords : Dog, Front Face, Biometrics, YOLO, Machine Learning

본 논문은 농촌진흥청 연구사업(세부과제번호: PJ01398601, 과제명: 반려견 생체인식 기반 개체식별 기술 개발)과 2020년 농촌진흥청 국립축산과학원 전문연구원 과정 지원사업에 의해 이루어진 것임.

*Corresponding Author : Joon-Whoan Lee(Jeonbuk National University)

email: chlee@jnu.ac.kr

Received October 30, 2020

Revised November 25, 2020

Accepted December 4, 2020

Published December 31, 2020

1. 서론

반려견의 수가 증가함에 따라 반려견 유기, 유실 등 사회문제가 증가하고 있다[1]. 2014년부터 동물 등록제를 전국 확대 시행하고 있지만, 내장형 무선식별장치의 경우 개체 삽입에 대한 거부감이 있고 외장형 식별장치의 경우 미착용 및 분실 등의 문제로 실효성이 떨어진다. 이러한 기존 등록방법의 문제점을 해결해줄 방법으로 안전하고 별도의 장치 착용 없이 개체식별이 가능한 비문, 홍채 등 영상기반의 생체인식기술이 주목받고 있다.

영상기반의 생체인식을 위해서는 우선 생체정보 이미지를 획득해야 한다. 이때 가장 중요한 것은 얼굴 방향이 정면일 때 이미지를 획득하는 것이다. 얼굴 방향이 정면일 때 생체정보를 온전하게 수집할 수 있고, 최대한 같은 형태로 이미지를 수집해야 개체 인식률을 높일 수 있기 때문이다. 반려견은 사람과 달리 원하는 때에 정면을 바라봐 주지 않기 때문에 생체정보 이미지 획득 확률을 높이기 위해서는 초당 30프레임 이상의 실시간 영상에서 반려견이 정면을 바라보고 있는지 아닌지를 판단하는 기술이 필요하다. 또한, 얼굴 방향이 정면일 때 그 이미지에서 홍채인식이나 비문 인식에 사용될 눈과 코 영역 분할을 위한 위치 정보를 제공해야 한다. 하지만 아직 반려견 얼굴 방향을 추정하는 연구는 아직 수행된 바 없다.

사람의 얼굴 방향 추정에 대한 연구는 화상회의, 사람과 로봇의 상호작용, 운전자 감시 시스템 등 다양한 응용 분야로 인해 계속 발전해왔다. 얼굴 방향 추정 방법에는 모델기반(Model-based) 방법, 외형기반(Appearance-based) 방법, 매니폴드 임베딩(Manifold Embedding) 방법, 비선형 회귀(Nonlinear Regression) 방법 등이 있다[2]. 모델기반 방법은 얼굴의 기하학적(Geometric) 정보 또는 랜드마크(Landmark)들의 위치 정보로 얼굴 방향을 추정한다[3-7]. 이 방식은 얼굴의 외형과 배경의 변화에 강하고 작은 각도에 좋은 성능을 보이지만 랜드마크 검출성능에 크게 의존하는 특성이 있다[8]. 반면, 외형기반 방법은 이미지를 그대로 사용하여 얼굴 방향을 추정하는 방식으로 얼굴 일부 가림(Occlusions) 및 큰 각도 추정 등에 강한 특성이 있지만, 얼굴 방향과 관련 없는 정보들이 많아 정확성 향상을 위한 많은 연구가 필요하다[9].

객체검출 기술의 발전과 함께 랜드마크를 기반으로 하는 얼굴 방향 추정 방법의 성능도 향상되었다. 또한, 랜드마크 정보로부터 얼굴 방향 추정을 위한 정보를 추출하는 방법도 더욱 효과적으로 발전해오고 있다. J. XIA et

al.(2019)은 얼굴 랜드마크 기반으로 히트맵(Heatmap) 정보를 생성하여 순방향 신경망(Feed-forward Network)을 통해 실제 환경에 강한 얼굴 방향 추정 방법을 제안하였으며[10], A. F. ABATE et al.(2019)은 얼굴 랜드마크 기반으로 학습 없이 얼굴 방향을 추정하는 방법을 제시하였다[11].

반려견의 경우 사람과 달리 털이 많아 개체마다 얼굴 방향과 관련 없는 외형적 변화가 크고, 히트맵(Heatmap)과 같은 정밀한 기하학적 정보를 활용하는 데에도 제약이 있다. 따라서 본 논문에서는 반려견 얼굴 방향 추정에 적합한 정보를 추출하는 방법을 제안하고, 이를 기반으로 반려견 얼굴 정면 여부를 판단하여 얼굴 방향이 정면일 때 생체인식 이미지 분할을 위한 눈과 코의 위치 정보를 제공하는 시스템을 제안하고자 한다.

제안하는 시스템은 이미지에서 딥러닝 알고리즘을 통해 반려견의 눈과 코를 검출하고, 검출된 눈과 코의 상대적 위치와 크기 등으로부터 제안한 방법으로 얼굴 방향 추정을 위한 5가지 정보를 추출하여 기계학습 분류기를 통해 정면인지 아닌지를 판단한다. 얼굴 방향이 정면일 경우, 생체인식에 사용할 눈과 코 이미지 분할을 위해 딥러닝을 통해 검출된 눈과 코의 위치정보를 제공한다. 알고리즘 학습 및 평가를 위해 반려견 사진 2,000장을 수집하여 정면과 비 정면으로 분류하여 사용하였다. 1,600장으로 제안 방법에 사용된 모델을 학습하고 검증하였으며, 400장의 데이터로 반려견 얼굴 정면 판별 테스트를 수행하였다. 실험결과 95.25%의 정확도를 보였다.

본 논문의 나머지 구성은 다음과 같다. 2장에서는 반려견 얼굴 정면 판별을 위해 제안한 방법을 자세하게 설명하고, 3장에서는 제안한 방법에 대한 실험결과를 보인다. 마지막으로 4장에서는 결론 및 향후 연구 방향을 기술한다.

2. 재료 및 방법

Fig. 1은 본 논문에서 제안하는 반려견 얼굴 정면판별 방법의 전체 흐름도를 나타낸다. 딥러닝 알고리즘을 통해 입력되는 반려견 이미지에서 눈과 코를 검출하고 각각의 좌표를 출력한다. 눈과 코의 위치 좌표를 기준으로 얼굴 정면판별에 사용할 정보를 추출한다. 추출된 정보를 기계학습 분류기에 입력하여 반려견이 정면을 보고 있는지 아닌지를 판별한다. 얼굴 방향이 정면인 경우, 생체인식에 사용할 눈과 코 영역을 이미지에서 분할 할 수 있도록

검출된 눈과 코의 위치 정보를 제공한다. 2.1은 실험에 사용한 데이터를 설명하고 2.2에서 2.4는 제안한 알고리즘을 단계별로 자세하게 설명한다.

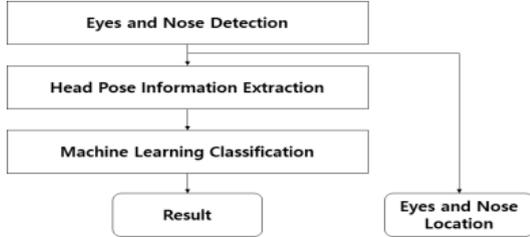


Fig. 1. Block diagram of proposed method.

2.1 데이터 세트

Table 1. Overview of dog data set

Data Set		Front	Not Front	total
Training	Maltese	201	61	262
	Poodle	181	87	268
	Shitzu	229	59	288
	Yorkshire	122	120	242
	Pomeranian	130	90	220
	Sub Total	863	417	1,280
Validation	Maltese	50	16	66
	Poodle	45	22	67
	Shitzu	58	15	73
	Yorkshire	30	30	60
	Pomeranian	32	22	54
	Sub Total	215	105	320
Test	Maltese	63	19	82
	Poodle	56	27	83
	Shihtzu	72	19	91
	Yorkshire	38	38	76
	Pomeranian	40	28	68
	Sub Total	269	131	400
Total	1,347	653	2,000	

국내에서 가장 많이 양육되는 반려견 대표 품종인 말티즈, 푸들, 시츄, 요코셔테리어, 포메라니안 이미지를 수집하고, 정면과 비 정면으로 분류하였다. 두 눈과 코 중 하나라도 보이지 않는 이미지는 눈과 코 검출 후 개수를 확인하여 쉽게 정면이 아닌 것으로 판별할 수 있다. 따라서 두 개의 눈과 코가 모두 보이는 이미지만 수집하였다.

Table 1은 실험에 사용된 데이터를 상세하게 보여준다. 총 2,000장의 사진 중 정면은 1,347개이고 비 정면

은 653개로 구성되어 있으며, 이를 학습데이터 1,280개, 검증데이터 320개, 평가데이터 400개로 분류하였다. 데이터는 tzutalin GitHub에서 배포하고 있는 공개소스 프로그램인 LabelImg[12]를 사용하여 반려견 눈과 코에 대해 라벨링을 하였다. 라벨링이 완료되면 각 이미지에 대해 클래스 번호, 경계상자(Bounding Box)의 정규화된 좌푯값(x, y, w, h)이 텍스트 파일로 저장된다.

2.2 딥러닝 활용 반려견 눈과 코 검출

얼굴 특징들의 상대적인 위치와 크기로 얼굴 방향을 판단하는 기하학적 정보를 활용하는 방법에서는 각 특징의 위치를 얼마나 정확하게 찾는지가 매우 중요하다. 또한, 비협조적인 반려견으로부터 생체인식에 적합한 생체 정보 이미지 수집 확률을 높이기 위해서는 초당 30프레임 이상의 실시간 영상처리가 가능해야 한다. 따라서 본 논문에서는 반려견 눈과 코 검출을 위해 실시간 처리가 가능하며 정확도도 우수한 YOLOv3[13]과 YOLOv4[14] 알고리즘을 선택하였다. YOLO(You Only Look Once)는 하나의 합성곱 신경망(Convolutional Network)이 이미지를 한번 보는 것으로 객체의 위치검출(Localization)과 분류(Classification)를 동시에 수행하여 실시간 처리에 최적화된 딥러닝 알고리즘이다. 이 알고리즘은 계속 수정되고 발전해오면서 여러 형태로 존재하며, YOLOv3와 YOLOv4는 가장 최근에 공개된 모델이다. 객체 검출 알고리즘의 구조는 Fig. 2처럼 Backbone, Neck, Head (Dense Prediction) 세 부분으로 구성된다. Backbone은 특징지도(Feature Map) 추출을 위해 활용되는 이미 학습된 네트워크이고, Neck은 Backbone의 다른 단계에서 다른 특징지도를 추출하기 위해 Backbone과 Head 사이에 존재하는 계층(Layer)이며, Head는 실제로 객체검출을 담당하는 부분이다. YOLOv3는 darknet53을 Backbone으로 적용하고 Feature Pyramid Network(FPN)을 Neck으로 사용하며, Head는 경계상자에 객체가 있는지를 판단하기 위한 점수와 그 객체가 어떤 클래스(Class)인지 분류하기 위한 클래스별 확률을 동시에 예측하는 구조로 되어있다. YOLOv4는 최신 딥러닝 기술을 접목하여 YOLOv3를 개선한 모델이며, CSPdarknet53을 Backbone에 적용하고, Spatial Pyramid Pooling (SPP)과 Path Aggregation Network(PAN)을 Neck으로 사용하며, Head는 YOLOv3의 구조를 가진다.

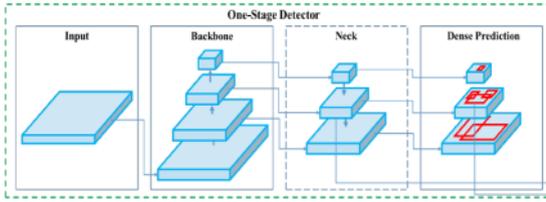


Fig. 2. Structure of object detection[14]

2.3 반려견 얼굴 정면판별을 위한 정보 추출

본 논문에서는 Fig. 3(a)에서처럼 반려견의 얼굴 움직임의 방향을 Yaw, Pitch, Roll로 나타낸다. Yaw, Pitch, Roll은 각각 Z축 기준 회전, X축 기준 회전, Y축 기준 회전을 의미한다. 반려견 얼굴에서 Roll 방향의 틀어짐은 생체정보 이미지의 변형에 크게 영향을 주지 않고, 두 눈을 연결하는 선의 틀어짐 각도를 구해 쉽게 보정할 수 있다. 따라서 얼굴 정면 여부는 Yaw와 Pitch 방향 틀어짐으로만 판단한다.

사람의 얼굴 방향 추정에 사용된 기하학적 방법을 응용하여 Yaw와 Pitch 방향 틀어짐에 대한 정보를 제공하는 다섯 가지 요소를 추출하고 A~E로 표시한다. 첫 번째 (A)는 Fig. 3(b)에서처럼 두 눈과 코의 중심점을 연결하는 삼각형의 3개의 내각 값(a, b, c)이고, 이는 Yaw와

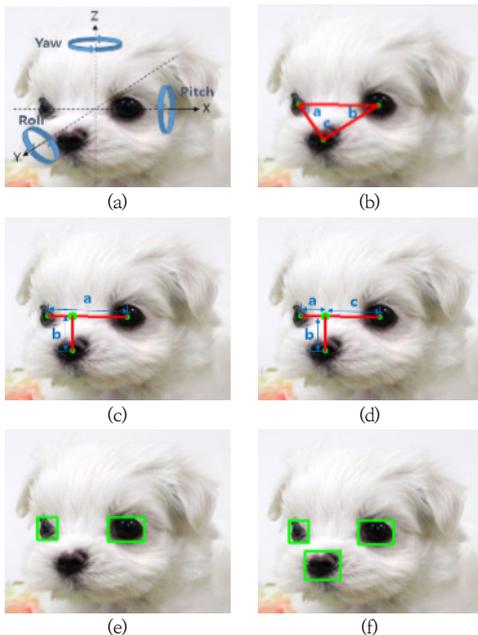


Fig. 3. Information extraction related to head pose. (a) 3 degrees of freedom of a dog head, (b)~(f) Proposed geometry information to estimate dog's front face.

Pitch 방향에 대한 정보를 제공한다. 두 번째(B)는 Fig. 3(c)에서처럼 두 눈의 중심점을 연결하는 선의 길이(a)와 그 선에서 수직으로 코와 연결된 선의 길이(b)의 비율 (b/a)이고, Pitch 방향에 대한 정보를 제공한다. 세 번째 (C)는 Fig. 3(d)에서처럼 두 눈 중심점을 연결하는 선 (a+c)과 그 선에서 수직으로 코와 연결된 선(b)이 만나는 지점에서 왼쪽 눈까지의 길이(a)와 오른쪽 눈까지의 길이 (c)의 비율(c/a)이고, Yaw 방향의 정보를 제공한다. 네 번째(D)는 Fig. 3(e)에서처럼 검출된 왼쪽 눈과 오른쪽 눈의 경계상자 면적 비율이고, Yaw 방향의 정보를 제공한다. 마지막 다섯 번째(E)는 Fig. 3(f)에서처럼 검출된 눈과 코의 경계상자에서 코 면적 대비 왼쪽 눈의 면적과 오른쪽 눈의 면적 비율이고, 이 또한 Yaw 방향의 정보를 제공한다.

2.4 기계학습 기반 반려견 얼굴 정면 여부 판별

추출한 5가지 반려견 얼굴 방향 정보(A~E)를 활용하여 얼굴 방향이 정면인지를 판별하기 위해 딥러닝 방법인 Multi-Layer Perceptron(MLP)과 대표적인 기계학습 분류기인 Random Forest(RF), Support Vector Machine(SVM)을 제안한 방법에 적용하였다. MLP 분류기는 입력층(Input Layer), 은닉층(Hidden Layer), 출력층(Output Layer)으로 구성되는 딥러닝 알고리즘이다. MLP 분류기 성능에 영향을 주는 요소에는 은닉층의 수, 각 층의 노드(Node) 수, 활성화 함수(Activation), 최적화 알고리즘(Optimizer), L2 일반화 파라미터 (Alpha), 학습속도(Learning Rate)가 있다. RF 분류기는 나무 형태로 데이터를 분류하는 의사결정 나무 (Decision Tree) 분류기를 같은 형태로 여러 개 사용하고 투표를 통해 최종적으로 분류하는 알고리즘이다. RF 분류기 성능에 영향을 주는 요소에는 의사결정 나무의 수(Estimator), 최대깊이(Max Depth), 하위 노드 분리를 위한 최소 샘플 수(Min Samples Split), 노드가 되기 위한 최소 샘플 수(Min Samples Leaf) 등이 있다. SVM 분류기는 근접한 샘플들 사이의 거리를 최대화하는 결정 경계(Decision Boundary)를 찾아 데이터를 분류하는 알고리즘이다. SVM 분류기 성능에 영향을 주는 요소에는 커널(Kernel), 일반화 파라미터(C), 감마(Gamma)가 있다.

Table. 2는 분류기별 성능 최적화를 위해 조정할 하이퍼 파라미터(Hyper-parameter)를 보여준다. 성능에 영향을 많이 주는 변수들로 선정하였다. 가장 성능이 좋은 조합을 찾기 위해 Grid Search 알고리즘을 사용하였

으며[15], 성능 검증에는 최근 보편적으로 사용되는 K-겹 교차검증 방법을 사용하였다[16]. 본 논문에서는 K를 5로 설정하였다.

Table 2. Different machine learning(ML) classifiers with hyper-parameters to be tuned by grid search

Classifiers	Hyper-parameters
MLP	1) Number of hidden layers : 1, 2, 3, 4, 5 2) Number of nodes at a layer : 16, 32, 64 3) Activation : tanh, relu 4) Optimizer : SGD, adam 5) Alpha : 0.0001, 0.001, 0.01, 0.1, 0.9 6) Learning rate : constant, adaptive
RF	1) Number of estimators : 100, 500, 1000 2) Max depth : 5, 8, 15, 25, 30 3) min samples split : 2, 5, 10, 15, 100 4) min samples leaf : 1, 2, 5, 10
SVM	1) Kernel : rbf 2) C : 0.1, 1, 10, 100, 1000 3) Gamma : 1, 0.1, 0.01, 0.001, 0.0001

3. 실험결과

제안 방법은 크게 반려견 눈과 코를 검출하는 과정, 검출된 눈과 코의 위치를 기반으로 얼굴 방향 정보를 추출하는 과정, 그리고 최종적으로 얼굴 정면 여부 판단하는 과정으로 구성된다. 실험결과는 이 세 가지 과정으로 구분하여 자세히 설명한다.

3.1 반려견 눈과 코 검출 실험결과

YOLOv3와 YOLOv4 모델의 입력 이미지 크기를 416x416과 608x608 두 가지로 설정하여 총 4개의 모델을 사용하였다. 1,280개의 데이터로 6,000회 반복 학습하였으며, 학습 중 4회 반복마다 320개의 데이터로 검증하여 성능이 가장 좋은 학습가중치를 최종모델로 선정하였다. 최종모델을 400개의 데이터로 테스트하였다. 알고리즘은 YOLOv4 개발자가 AlexeyAB darknet GitHub[17]에서 배포하고 있는 공개소스 프로그램을 사용하였다. Backbone 가중치 및 하이퍼 파라미터는 각 모델의 기본값을 사용하였고, 성능 평가는 Intersection of Union (IOU)이 0.5일 때의 Mean Average Precision(mAP)를 사용하였다.

Table 3은 검출성능과 속도 측정 결과를 보여준다. 4개의 모델 모두 검증데이터와 평가데이터에 대한 성능의 차이가 크지 않았다. 이는 과적합(Overfitting) 없이 학

습이 성공적으로 수행된 것을 보여준다. 실험에 사용한 4개 모델 모두 98% 이상의 높은 검출성능을 보였으며, 이미지 크기가 416x416일 때가 608x608보다 더 좋은 성능을 보였다. YOLOv4(416x416)가 99.30%로 가장 높은 검출 정확도를 보였다. 검출시간은 i7-8700 CPU, 64GB 램 PC에서 측정하고 분석하였다. 4개의 모델 중 가장 처리속도가 늦은 YOLOv4(608x608)도 400장의 이미지 처리에 6.84ms밖에 걸리지 않았으며, 이는 이미지 당 평균 17.1ms의 속도로 초당 58프레임 이상 처리가 가능한 수준이다. 따라서 4가지 모델 모두 초당 30프레임 이상의 실시간 처리가 가능한 것으로 분석되었다.

Table 3. Detection performance of dog's eyes and nose

Model	Input Size (Pixel)	mAP(@IOU=0.50)		Detection Time(ms)
		Valid	Test	
YOLOv4	608x608	99.76	98.41	6.84
	416x416	99.66	99.30	6.83
YOLOv3	608x608	98.85	98.07	5.66
	416x416	99.60	98.59	5.61

3.2 얼굴 방향 정보 및 최적 분류모델 선정

본 논문에서 제시한 5가지 얼굴 방향 정보(A~E)의 유효성을 검증하고 각 기계학습 분류기의 최적 모델을 찾는 실험을 수행하였다. 1,600개의 학습데이터와 검증데이터에 대해 사람이 라벨링한 경계박스의 좌표정보를 기준으로 5개의 얼굴 방향 정보(A~E)를 추출하고, Yaw와 Pitch 정보를 동시에 제공하는 A 정보에 B~E 정보를 추가해가며 5개의 정보그룹(Group1~5)을 만들었다. 그다음 Grid Search 알고리즘과 K겹 교차검증 방법으로 각 분류기의 최적 성능을 분석하였다. 얼굴 방향 정보 추출 알고리즘과 분류기 모델, 그리고 검증 알고리즘은 파이썬(Python)과 기계학습 라이브러리인 케라스(Keras), Scikit-learn[18]을 사용하여 구현하였다.

Table 4는 얼굴 방향 정보에 따른 각 분류기 최고 성능 평균과 편차의 결과를 보여준다. MLP, RF, SVM 분류기 모두 Group5에서 가장 좋은 성능을 보여주어, 본 논문에서 제시한 5개의 얼굴 방향 정보 모두 얼굴 정면 판별에 도움이 되는 것으로 나타났다.

Table 5는 하이퍼 파라미터의 여러 조합 중 분류기별 가장 좋은 성능을 발휘하는 조합을 보여준다. 분류기별로 가장 좋은 성능을 보여주는 하이퍼 파라미터를 최적 모델로 선정하였다. MLP 분류기는 은닉층 5개, 각 층의 노

드 수 32/32/64/32/32, 활성화 함수 tanh, 최적화 알고리즘 SGD, L2 일반화 파라미터 0.0001, 학습속도 Adaptive일 때 97.38%로 가장 좋은 성능을 나타냈다. RF 분류기는 의사결정 나무 분류기 수 100개, 최대깊이 25층, 하위 노드 분리 최소 샘플 수 5개, 노드 최소 샘플 수 1개일 때 97.5%로 가장 좋은 성능을 나타냈다. SVM 은 rbf 커널을 사용하였으며, C가 1000이고 Gamma 0.0001일 때 97.25%로 가장 좋은 성능을 보여주었다.

3.3 반려견 얼굴 정면판별 실험결과

3.1에서 학습되고 최적 모델로 선정된 YOLOv3와 YOLOv4를 통해 400개의 평가데이터에서 반려견의 눈과 코를 검출하였다. 검출된 좌표정보와 사람이 라벨링한 좌표정보를 사용하여 본 논문에서 제안한 5가지 얼굴 방향 정보를 추출하였다. 추출된 얼굴 방향 정보를 3.2에서 선정된 분류기별 최적 모델에 적용하여 얼굴 정면판별 실험을 수행하였다.

400개의 평가 이미지 중 두 눈과 코 중 하나 이상을 검출하지 못한 경우는 YOLOv4(608x608)가 7개, YOLOv4(416x416)가 4개, YOLOv3(608x608)가 14

개, YOLOv3(416x416)가 8개였다. 이 중 실제 정면 이미지는 YOLOv3(608x608)에서 1개 있었고, 나머지는 정면이 아닌 이미지였다. 일반적으로 두 눈과 코 중 하나 이상을 검출하지 못하는 경우는 정면이 아닌 때문에 본 논문에서는 해당 이미지들을 정면이 아닌 것으로 분류하고 나머지 이미지는 얼굴 방향 정보를 추출하여 기계학습 분류기에 입력하였다.

Table 6은 기계학습 분류기별 얼굴 정면판별 성능을 보여준다. 사람이 라벨링한 정보에 대해서는 SVM 분류기가 97.25%의 정확도로 가장 높았다. YOLO 모델에서 검출된 정보에 대해서는 RF 분류기가 4개의 모델 모두에서 가장 정확도가 높았다. 그중 가장 성능이 좋은 조합은 YOLOv4(416x416)와 RF 분류기를 사용한 경우로 95.20%의 정확도를 보였다.

눈과 코 중 하나 이상을 검출하지 못한 경우를 정면이 아닌 것으로 분류하여 전체 평가데이터에 대한 정면판별 성능을 보면, 실제 정면인 이미지에서 두 눈과 코 중 하나 이상을 검출하지 못한 YOLOv3(608x608)의 경우만 정면판별 정확도가 떨어졌고 나머지 경우는 정면판별 정확도가 향상되었다. 가장 좋은 성능을 보여주는 검출기와

Table 4. Hyper-parameters tuning results for each group of head pose information

Group of head pose Information	Classifier Accuracy(%)		
	MLP	RF	SVM
Group1 (A)	97.25 ± 0.011	95.38 ± 0.017	96.81 ± 0.014
Group2 (A+B)	97.31 ± 0.008	95.56 ± 0.015	96.81 ± 0.014
Group3 (A+B+C)	97.25 ± 0.008	97.13 ± 0.010	96.81 ± 0.014
Group4 (A+B+C+D)	97.31 ± 0.008	97.25 ± 0.007	96.81 ± 0.014
Group5 (A+B+C+D+E)	97.38 ± 0.009	97.50 ± 0.005	97.25 ± 0.007

Table 5. Best-tuned hyper-parameters using grid search

ML Models	Hyper-parameters	Group of head pose information				
		Group5	Group4	Group3	Group2	Group1
MLP	1) Number of hidden layers	5	4	5	4	5
	2) Number of nodes at a layer	32/32/64/32/32	16/16/16/32	32/64/64/64/16	16/64/32/32	32/64/16/64/32
	3) Activation	tanh	tanh	tanh	tanh	tanh
	4) Optimizer	sgd	adam	adam	sgd	sgd
	5) Alpha	0.0001	0.01	0.01	0.001	0.0001
	6) Learning rate	adaptive	-	-	adaptive	adaptive
RF	1) Number of estimators	100	1000	100	500	500
	2) Max depth	25	25	25	25	15
	3) min samples split	5	15	20	5	2
	4) min samples leaf	1	1	1	1	1
SVM	1) Kernel : rbf	rbf	rbf	rbf	rbf	rbf
	2) C	1000	1000	100	100	100
	3) Gamma	0.001	0.001	0.001	0.001	0.001

분류기의 조합은 YOLOv4(608x608), YOLOv4(416x416) 및 YOLOv3(416x416)를 RF 분류기와 사용되었을 때였고, 정확도는 95.25%로 같게 나타났다. 하지만 반려견 눈과 코 검출률이 떨어지는 것은 잠재적으로 성능 저하를 가져올 수 있어 검출률이 가장 좋은 YOLOv4(416x416)와 RF 분류기를 사용한 시스템이 가장 좋은 성능을 보여준다고 볼 수 있다.

처리시간은 i7-8700 CPU, 64GB 램 PC에서 측정하고 분석하였다. 400장의 평가데이터에서 얼굴 방향 정보를 추출하고 분류하는데 소요된 총 시간은 MLP 분류기를 사용한 경우 4.246ms, RF 분류기는 4.277ms, SVM은 4.230ms로 측정되었다. 시간이 가장 많이 소요되는 RF 분류기를 적용한 경우 얼굴정보 추출부터 분류까지 이미지 한 장당 평균 10.7ms 가 소요되었다. 이는 초당 93.46개의 이미지를 처리할 수 있는 속도로 초당 30프레임 이상의 실시간 처리가 가능한 것으로 분석되었다.

Table 6. Recognition of front head with ML classifiers

Detection method	Number of Image	Accuracy(%)			Winner
		MLP	RF	SVM	
Human Labelling	400	95.75	96.00	97.25	SVM
YOLOv4 (608x608)	393	94.15	95.17	93.13	RF
	400*	94.25	95.25	93.25	
YOLOv4 (416x416)	396	94.70	95.20	94.44	RF
	400*	94.75	95.25	94.50	
YOLOv3 (608x608)	386	93.52	95.08	93.52	RF
	400*	93.50	95.00	93.50	
YOLOv3 (416x416)	392	94.13	95.15	93.88	RF
	400*	94.25	95.25	94.00	

* Results including the cases missing at least one eye or nose by YOLOv3 and YOLOv4

4. 결론

본 논문에서는 반려견 생체정보 수집을 위해 실시간 영상에서 반려견 얼굴 방향이 정면 여부를 판별하는 방법을 제안하였다. 반려견 얼굴 정면 여부 판별을 수행하는 목적에 맞게 실제 환경에서 작은 각도의 변화도 잘 감지할 수 있고 반려견 얼굴이 정면일 때 생체정보 인식에 사용할 눈과 코의 위치 정보를 쉽게 제공할 수 있는 구조로 설계되었다.

제안한 방법은 YOLOv4를 통하여 실시간 영상에서 반려견 눈과 코를 검출하고, 검출된 눈과 코의 상대적 크

기와 위치로 5가지의 얼굴 방향 정보를 추출하여 RF 분류기를 통해 정면 여부를 판별한다. 눈과 코 검출성과 속도, 5가지 얼굴 방향 정보의 효과 정보 추출 속도, 그리고 최종판별을 위한 분류기의 성능과 속도에 대해 각각 실험을 통해 검증하였다.

본 논문에서 제안한 방법은 반려견 비문, 홍채 등 생체인식을 위한 이미지 정보수집에 활용될 수 있을 것으로 기대한다. 반려견 생체인식에 적합한 생체정보 수집을 위해서는 제안한 얼굴 정면판별과 더불어 생체정보 이미지의 크기, 선명도 및 빛 반사 여부 등 생체인식에 적합한지를 판단하는 추가적인 방법이 필요하다. 향후 얼굴 정면판별 성능 향상을 위한 연구와 이를 활용한 생체정보 수집방법 연구를 수행할 것이다.

References

- [1] Ministry of Agriculture, Food and Rural Affairs, Rescue and protection of 100,000 lost and abandoned animals in 2017. Press Release, Ministry of Agriculture, Korea, pp. 4-5.
- [2] E. Murphy-Chutorian and M. M. Trivedi, "Head pose estimation in computer vision: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 4, pp. 607-626, Apr. 2009.
- [3] H. Wilson, F. Wilkinson, L. Lin, and M. Castillo, "Perception of Head Orientation," *Vision Research*, vol. 40, no. 5, pp. 459-472, 2000.
- [4] A. Nikolaidis and I. Pitas, "Facial Feature Extraction and Pose Determination," *Pattern Recognition*, vol. 33, no. 11, pp. 1783-1791, 2000
- [5] A. Gee and R. Cipolla, "Determining the Gaze of Faces in Images," *Image and Vision Computing*, vol. 12, no. 10, pp. 639-647, 1994.
- [6] T. Horprasert, Y. Yacoob, and L. Davis, "Computing 3-D Head Orientation from a Monocular Image Sequence," *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition*, pp. 242-247, 1996.
- [7] J.-G. Wang and E. Sung, "EM Enhancement of 3D Head Pose Estimated by Point at Infinity," *Image and Vision Computing*, vol. 25, no. 12, pp. 1864-1874, 2007.
- [8] V. Drouard, R. Horaud, A. Deleforge, S. Ba, and G. Evangelidis, "Robust head-pose estimation based on partially-latent mixture of linear regressions," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 14281440, Mar. 2017.
- [9] N. Ruiz, E. Chong, and J. M. Rehg, "Fine-grained head pose estimation without keypoints," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun.

2018, pp. 2074-2083.

- [10] J. Xia, L. Cao, and J. Liao, "Head pose estimation in the wild assisted by facial landmarks based on convolutional neural networks", IEEE Access, vol. 7, Apr. 2019.
- [11] A. F. Abate, P. Barra, C. Bisogni, M. Nappi, and S. Ricciardi, "Near real-time three axis head pose estimation without training", IEEE Access, vol. 7, May. 2019
- [12] Tzutalin LabelImg GitHub [Internet]. Available : <https://github.com/tzutalin/labelImg/>
- [13] J. Redmon, A. Farhadi, "Yolov3: An incremental improvement", arXiv preprint arXiv:1804.02767, 2018.
- [14] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4:Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.
- [15] D. Krstajic, L. J. Buturovic, D. E. Leahy, and S. Thomas, "Cross-validation pitfalls when selecting and assessing regression and classification models," J. Cheminformatics, vol. 6, no. 1, p. 10, Dec. 2014. DOI: <https://doi.org/10.1186/1758-2946-6-10>
- [16] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," Statist. Surv., vol. 4, pp. 40-79, Jul. 2010. DOI: <https://doi.org/10.1214/09-SS054>
- [17] AlexeyAB darknet GitHub [Internet]. Available : <https://github.com/AlexeyAB/darknet/>
- [18] "Scikit-learn: Machine Learning in Python", Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

김 중 복(Jong-Bok Kim)

[정회원]



- 2006년 9월 : UCLA Electrical Eng.(공학석사)
- 2006년 10월 ~ 2018년 1월 : LG이노텍 책임연구원
- 2018년 2월 ~ 현재 : 농촌진흥청 국립축산과학원 연구관

<관심분야>

축산 스마트팜, 반려동물 ICT

장 동 화(Dong-Hwa Jang)

[정회원]



- 2020년 2월 : 전북대학교 농업기계공학과 (공학석사)
- 2020년 1월 ~ 현재 : 농촌진흥청 국립축산과학원 축산환경과 전문연구원

<관심분야>

축산 스마트팜, 영상처리

양 가 영(Yang Kayoung)

[정회원]



- 2012년 2월 : 강원대학교 동물자원학 (농학석사)
- 2016년 8월 : 강원대학교 축산학 동물시스템과학 (농학박사)
- 2017년 2월 ~ 현재 : 농촌진흥청 국립축산과학원 전문연구원

<관심분야>

동물복지, 동물행동, 정밀축산

권 경 석(Kyeong-Seok Kwon)

[정회원]



- 2010년 8월 : 서울대학교 지역시스템공학 (공학석사)
- 2016년 8월 : 서울대학교 지역시스템공학 (공학박사)
- 2017년 2월 ~ 현재 : 농촌진흥청 국립축산과학원 연구사

<관심분야>

시설환경, 대기환경, 전산유체역학

김 중 곤(Jung-Kon Kim)

[정회원]



- 2007년 2월 : 조선대학교 생물신소재학과 (환경미생물학 전공, 공학박사)
- 2008년 10월 ~ 2009년 10월 : Iowa State University (Post-doc.)

- 2009년 11월 ~ 2012년 12월 : 농촌진흥청 국립식량과학원 (Post-doc.)
- 2013년 1월 현재 : 농촌진흥청 국립축산과학원 축산환경과 농업 연구사

<관심분야>

가축분뇨처리, 바이오에너지

이 준 환(Joon-Whoan Lee)

[정회원]



- 1982년 2월 : 한국과학기술원 전자공학과 (공학석사)
- 1990년 8월 : 미국 미주리대학교 전기 및 컴퓨터공학과 (공학박사)
- 2013년 8월 ~ 2014년 8월 캐나다 SFU 객원교수
- 1985년 10월 ~ 현재 : 전북대학교 교수

<관심분야>

영상 및 오디오 신호처리, 인공지능