

반려견 영상 실시간 행동 인식 시스템

봉정근, 조민아, 하유성, 황준원, 원일용
 서울호서전문학교 ICT 융합보안과
 jgbong0306@gmail.com, oiajmn@gmail.com, evernal2@naver.com,
 angelhjw91@gmail.com, clccclcc@shoseo.ac.kr

A Real-Time System for Recognizing Companion Dog Behavior Through Video

Jung-Geun Bong, Min-A Jo, Yu-Seong Ha, Jun-Won Hwang, IL-Yong Weon
 Dept. of Cyber Investigation, Seoul-Hoseo Technical College

요 약

본 논문은 기존의 웨어러블 센서 방식이 아닌 영상으로 반려견의 행동을 분석하는 연구에 대한 것이다. 제안한 시스템은 영상에서 반려견의 영역을 탐지하고, 탐지된 이미지에서 반려견의 관절 좌표를 추출하여 행동을 판단하는 방식이다. 모든 프레임에 대해 처리하지 않고, 일정 주기 단위로 영상을 처리해 실시간성을 확보하였다. 제안한 시스템의 유용성은 실험으로 검증하였으며, 유의미한 실험 결과를 얻을 수 있었다.

1. 서론

KB경영연구소의 '2023 한국 반려동물보고서'에 따르면, 2022년 말 기준 552만 가구가 반려동물을 키우고 있다고 응답하였으며, 이는 2020년 말 조사(536만 가구) 대비 2.8% 증가한 수치이다[1].

반려동물에 대한 인식이 높아지면서 반려동물의 행동을 이해하고, 모니터링하는 것에 관한 관심이 증가하고 있다. 기존에는 웨어러블 센서를 활용한 행동 인식 방식이 주로 연구되고 있으나[2], 이러한 방식은 반려동물에게 불편함이나 고통을 줄 수 있을 뿐만 아니라 센서의 위치가 변하면 정확하지 않은 데이터를 수집할 수 있다는 단점이 있다.

본 연구는 비접촉식 카메라만을 활용하여 반려견을 탐지하고 행동을 분석하는 방법을 다룬다. 핵심적인 행동 인식 알고리즘은 영상에서 대표적인 관절들의 좌표를 측정하고 이를 활용하여 각도를 계산하고 학습하는 과정을 포함하고 있다. 알고리즘의 처리 속도 향상을 위해 일정 간격의 영상 프레임만을 대상으로 추출하였다.

2장에서는 제안된 시스템에 대한 상세한 설명이, 3장에서는 실험과 결과에 관한 내용이 다루어지며, 결론 및 향후 과제는 4장에서 다루어진다.

2. 제안 시스템

본 논문에서 제안하는 시스템 구조는 그림 1과

같다. 이 시스템은 먼저, 실시간으로 입력된 영상에서 일정 주기로 프레임을 선택하고 해당 프레임에서 반려견 객체를 탐지하여 특정 영역을 정의한다. 이후, 탐지된 영역에서 관절의 좌표를 추출하고, 이 좌표를 활용하여 특정 관절 간의 각도를 계산한다. 마지막으로, 이렇게 전처리된 데이터를 시계열 형태로 변환하여 반려견의 행동을 분류한다.

Algorithm 1 Video Processing Pseudocode with Rolling Window

```

Open Video
Specify video source
Initialize frame_number to 1
Initialize empty list: angle_data
Initialize window_size
while frame_number ≤ total_frames_in_video do
  frame ← ReadFrame(video, frame_number)
  if frame is not null then
    joint_angles ← ConvertFrameToAngles(frame)
    Append joint_angles to angle_data
    if length(angle_data) > window_size then
      Remove the oldest element from angle_data
    end if
    if length(angle_data) = window_size then
      ProcessClassificationResult(AnalyzeAngles(angle_data))
    end if
    frame_number ← frame_number + 1
  else
    HandleEndOfVideo()
  end if
end while
Close the video

```

(그림 1) 탐지 알고리즘

3. 실험 및 결과

본 실험은 AMD Ryzen 5 3600XT CPU, 64GB 메모리, NVIDIA RTX 3060 그래픽 카드를 갖춘 환경에서 수행되었다.

3.1 실험 데이터 생성

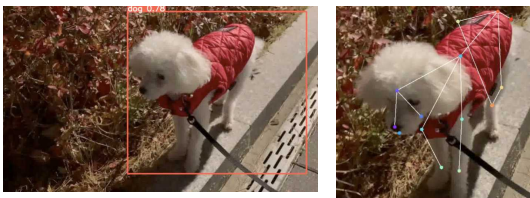
행동 분류 학습을 위해 AI Hub의 ‘반려동물 구분을 위한 동물 영상[3]’의 이미지와 라벨을 활용하였다. 이 데이터 세트는 총 13가지 동작으로 구성되어 있으며, 행동별 데이터 분포는 <표 1>과 같다.

| 동작 | 건 | 동작 | 건 |
|-------------|--------|----------|--------|
| BodyLower | 72,572 | Mounting | 4,743 |
| BodyScratch | 14,397 | Sit | 72,018 |
| BodyShake | 13,802 | Tailing | 32,530 |
| FeetUp | 31,269 | Taillow | 7,602 |
| FootUp | 48,006 | Turn | 19,601 |
| Heading | 17,414 | WalkRun | 81,601 |
| Lying | 29,249 | | |

<표 1> 행동 통계

반려견의 관절 좌표 추출을 위해 15개의 관절 좌표를 사용하였으며, 유의미한 결과를 얻기 위해 임의로 선택된 18개의 사잇각을 활용하였다. 이는 네 다리의 움직임과 머리와 허리, 꼬리와 같은 부분의 움직임을 계산하기 위한 것이다.

YOLOv8 모델을 활용하여 반려견 객체를 감지한 후 해당 영역의 크기를 조정한다. 그런 다음, DLC 모델을 적용하여 반려견의 관절 좌표를 추출한다 (그림 2 참고). 추출된 관절 좌표를 사용하여 각 다리와 몸체 사이의 18개의 사잇각을 계산하며, 이러한 관절 사잇각 데이터를 시계열로 처리하여 학습 데이터를 생성한다. (그림 3 참고)



(그림 2) YOLO 및 DLC 적용 예시

| | |
|-----------|---|
| WalkRun | [149.28008429 -305.19198162 46.91553281 205.01110911 202.43918319 158.93300548 -127.75682496 -237.32846348 -63.79205884 -161.9669525 48.62695199 250.56877042 70.9003564 66.86263461 17.62617239 17.23433191 -124.19109809 -14.70065483] |
| BodyLower | [99.50920726 169.08785134 -272.84819243 -261.70386815 -230.21294395 114.4130431 273.85658755 -61.19749679 -348.61174574 196.21992033 -299.42349183 53.30923401 -58.63487201 25.3501483 -86.96522218 174.69222833 -266.54764509 -27.93105392] |
| Sit | [82.23995791 172.84472425 -262.72250609 200.63843534 124.60326738 113.46911488 -66.17709643 -77.75114796 -340.62053065 192.01849831 -283.54513163 194.74562774 -56.42468561 44.06591245 -77.95509784 175.13677744 -301.81706041 -24.68717373] |

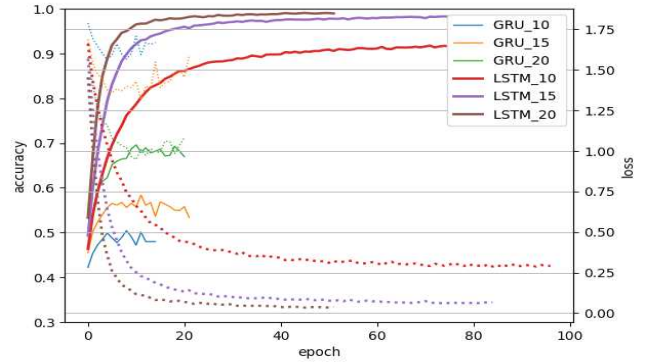
(그림 3) 학습 데이터 샘플

시계열 데이터 생성을 위해 초당 5프레임의 일정한 간격으로 데이터를 추출하고, 각 시퀀스는 1초 이상의 데이터 크기를 갖도록 한다. 최종 시퀀스의 길이는 실험을 통해 결정할 예정이다.

3.2 실험 결과 및 분석

본 실험에서는 LSTM과 GRU와 같은 모델을 활용하였으며, 시계열의 길이를 변화시켜가며 실험하

였다. 실험 결과는 그림 4와 같다.



(그림 4) 실험 결과

실험 결과, 시계열의 길이가 증가함에 따라 정확도도 함께 상승하였다. LSTM이 GRU보다 더 높은 정확도를 나타내었기 때문에, LSTM이 복잡한 패턴과 시퀀스를 학습하고 처리하는데 더 적합한 모델임을 결론지을 수 있다. 더불어, 데이터 입력부터 결과 출력까지의 처리 시간이 평균 0.038 sec로 초당 5프레임의 데이터만을 추출하여 분류를 진행하기에 충분히 실시간 처리가 가능했다.

4. 결론 및 향후 과제

최근 반려동물을 키우는 반려인들이 증가함에 따라 반려동물에 관한 관심 또한 증가했다. 본 논문에서는 기존의 웨어러블 센서 방식 대신 영상을 통해 반려견을 탐지하고 행동을 분석하는 시스템을 제안하였다. 실험을 통해 실제로 반려견의 행동을 분석할 수 있다는 것을 입증하였으며, 더불어 시스템의 실행 시간도 매우 짧아 실시간으로 분석할 수 있음을 확인하였다.

앞으로는 다양한 동물들의 행동을 인식할 수 있는 범용적인 모델을 개발하는 것이 필요하다. 또한, 실시간 분석을 넘어 미래의 행동을 예측하는 연구가 필요해 보인다.

참고문헌

[1] KB경영연구소, 2023 한국 반려동물 보고서, <https://www.kbfg.com/kbresearch/report/reportView.do?reportId=2000396>

[2] A. Hussain, S. Ali, Abdullah and H. -C. Kim, "Activity Detection for the Wellbeing of Dogs Using Wearable Sensors Based on Deep Learning," in IEEE Access, vol. 10, pp. 53153-53163, 2022, doi: 10.1109/ACCESS.2022.3174813.

[3] AI Hub, 반려동물 구분을 위한 동물 영상, <https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=59>