Audio-visual 멀티모달 정보 기반의 비정상 활성 돼지 탐지 시스템

채희찬[1], 이준희[1], 이종욱[2], 정용화[2], 박대희[2]
[1]고려대학교 컴퓨터정보학과
[2]고려대학교 컴퓨터융합소프트웨어학과
chay219@korea.ac.kr, watrtdc@korea.ac.kr, eastwest9@korea.ac.kr, ychungy@korea.ac.kr, dhpark@korea.ac.kr

Abnormal Active Pig Detection System using Audio-visual Multimodal Information

Heechan Chae[1], Junhee Lee[1], Jonguk Lee[2], Yonghwa Chung[2], Daihee Park[2]
[1]Dept. of Computer and Information Science, Korea University
[2]Dept. of Computer and Convergence Software, Korea University

요 약

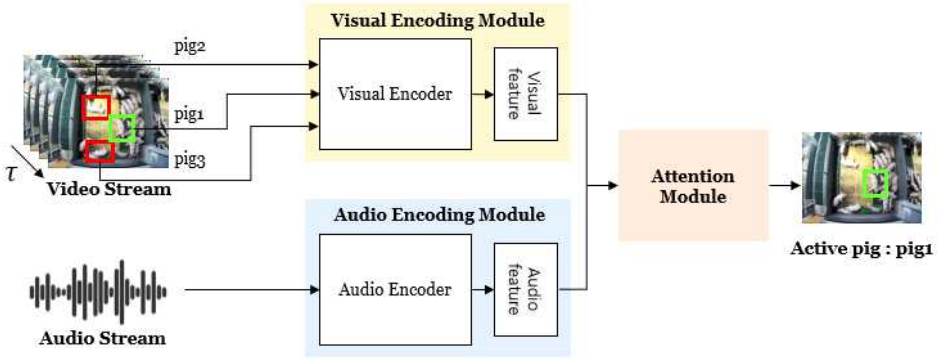
양돈을 관리하는 데에 있어 비정상 개체를 식별하고 사전에 추적하거나 격리할 수 있는 양돈업 시스템을 구축하는 것은 효율적인 돈사관리를 위한 필수 요소이다. 그러나 돈사내의 이상 상황을 탐지하는 연구는 보고되었지만, 이상 상황이 발생한 돼지를 특정하여 식별하는 연구는 찾아보기 힘들다. 따라서, 본 연구에서는 소리를 활용하여 이상 상황이 발생함을 탐지한 후 영상을 활용하여 소리를 낸 특정 돼지를 식별할 수 있는 시스템을 제안한다. 해당 시스템의 주요 알고리즘은 활성 화자 탐지 문제에서 착안하여 이를 돈사에 맞게 적용하여, 비정상 소리를 내는 활성 돼지를 식별 가능하도록 구현하였다. 제안한 방법론은 모의 실험을 통해 돈사 내의 이상 상황이 발생한 돼지를 식별할 수 있음을 확인하였다.

1. 서론

‘한돈농가 생산성 향상방안’[1]에 따르면 우리나라 양돈농가의 돼지 이유 후 육성률은 86.6%로, 상위권을 차지하는 네덜란드 97.7%, 덴마크 96.7%, 미국 95.1% 등에 비해 약 10%가량의 큰 차이로 낮은 수치를 보인다. 결과적으로 이러한 차이는 농가 사이의 막대한 수익 차를 야기한다. 이와 같은 농가 간 육성률 차이를 유발하는 가장 큰 원인 중 하나는 ‘질병예방’ 측면에 있다[1]. 이와 같은 돼지 질병의 경우 해당 질병을 조기에 탐지하면 다른 돼지들이 동일한 질병에 걸릴 위험을 상당히 낮출 수 있지만, 조기 탐지를 위한 자동 시스템이나 인력이 부족한 것이 국내 양돈농가의 현실이다. 따라서, 선진 농가들과의 격차를 줄이기 위해서는 질병 등이 발병한 비정상 개체를 조기에 자동으로 탐지할 수 있는 시스템을 개발하는 것이 필수적이다.

이와 같은 필요성에 의해, 양돈농가의 생산성에 영향을 줄 수 있는 돼지의 이상행동을 파악하려는 연구들이 이미 상당수 보고되었다. 그중에서도 영상에 비해 빠른 프로세스가 가능한 소리를 활용하여 돼지의 이상 상황을 탐지하려는 연구들이 다양하게 진행되었다[2-4]. Chung 등[2]은 기침 소리를 SVDD(Support Vector Data Description) 및 SRC(Sparse Representation Classifier) 알고리즘에 적용하여 돼지의 질병을 구분하였으며, Exadaktylos 등[3]은 AlexNet을 활용하여 아픈 돼지의 기침 여부를 식별하는 방법을 제안하였다. 하지만 돼지의 소리 정보만 활용한 기존 연구들은 단순히 이상 상황의 발생 여부만 판단할 수 있을 뿐 이상행동을 보이는 돼지 개체를 특정할 수는 없다. 따라서, 이상 상황 발생 시에 세밀한 개체 관리는 불가능하다.

반면에, 이러한 한계점을 보완하기 위해 이상 상황의 개별 돼지를 식별하기 위한 연구도 진행되었다. Kim 등[5]은 소리 및 영상 정보를 모두 사용하여 기침하는 돼지를 자동으로 식별하는 방법을 제안했다. 이 방법은 구체적으로, 평균 음높이를 이용하여 기침 소리를 먼저 인식한 후, MHI(Motion History Image) 기반의 분석을 수행하여 기침을 식별한다. 그러나 평균적인 음높이는 주변 소음에 민감하다는 한계가 있다. 또한, MHI 방식은 돼지가 기침하는 순간에 움직이는 돼지의 수가 많은 경우, 기



(그림 1) 활성 돼지 탐지 시스템 구조.

침하는 개체를 감지하기 어려운 한계점도 존재한다.

본 연구는 앞선 관련 연구들의 단점을 보완하고, 비정상적인 행동을 보이는 개별 돼지 식별이 가능한 새로운 접근 방식을 제안한다. 해당 방법은 소리 및 영상 정보를 활용해 화자를 찾는 활성 화자 탐지(Active Speaker Detection)[6] 연구에서 제시한 미세한 얼굴 움직임과 소리의 연관성을 포착하여 활성 화자를 탐지한 연구의 아이디어를 기반으로 하며, 이를 돈사 환경에 맞도록 재설계하였다. 즉, 비정상 행동 시에 발생하는 돼지의 미세한 움직임과 소리의 연관성을 포착한 후, 비정상적인 행동을 보이는 활성 돼지를 탐지하는 시스템을 제안한다. 결과적으로 이를 통해 주변 소음과 여러 움직임이 존재하는 상황 속에서도 비정상 활성 돼지를 강인하게 탐지할 수 있는 시스템을 구축하고자 한다.

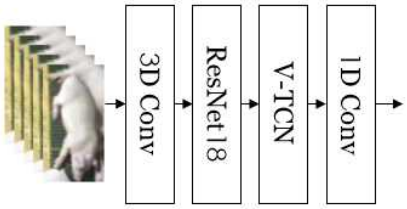
2. 활성 돼지 탐지 시스템

본 논문에서 제안하는 이상 상황이 발생한 활성 돼지 탐지 시스템의 구조는 그림 1과 같으며, 영상 특징을 추출하는 시각 인코딩 모듈, 소리 특징을 추출하는 청각 인코딩 모듈, 시청각 특징을 종합하여 활성 돼지의 프레임을 판단하는 어텐션 모듈 3단계로 구성된다. 개별 돼지와 소리가 각각 한 쌍을 이루어 인코딩 모듈의 입력으로 사용되고, 어텐션 모듈을 거쳐 활성화 돼지(비정상 소리를 낸 돼지) 여부를 프레임별로 나타내는 벡터가 시스템의 최종 출력으로 나오게 된다. 이때, 활성 돼지 탐지를 위하여, TalkNet[7] 기반의 활성 화자 탐지 모델을 사용한다.

2.1 시각 인코딩 모듈

시각 인코딩 모듈에서는 원본 영상에서 돼지 객체의 부분 영상을 입력으로 받아들인 후, 해당 돼지 객체의 움직임 정보가 임베딩된 특징 벡터를 생성한다.

그림 2는 시각 인코더의 구조를 보여준다. 우선 입력 영상은 3D 컨볼루션 계층과 ResNet18을 거치면서 영상의 시각적 정보를 추출한다. 이후 시계열 데이터에서 좋은 성능을 보였던 TCN(Temporal Convolutional Network)모델을 영상 데이터에 맞게 변형시킨 V-TCN(Video-TCN)[7]을 사용해 영상의 시간 정보를 추출한다. 마지막으로 소리 특징과 크기를 맞추기 위해 1D 컨볼루션 계층을 통해 특징 크기를 조정한다.



(그림 2) 시각 인코더 구조.

2.2 청각 인코딩 모듈

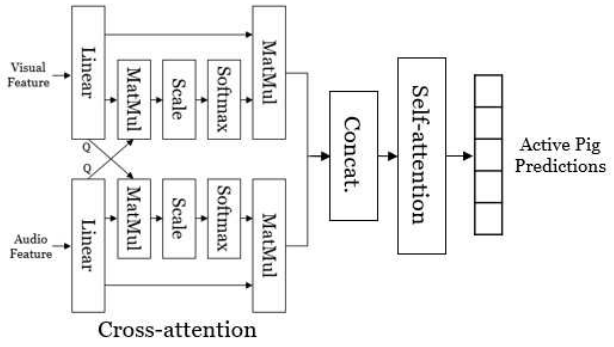
청각 인코딩 모듈에서는 소리 스트림에서 돼지의 비정상 소리 특징을 포착할 수 있도록 임베딩된 특징 벡터를 추출한다.

청각 인코더의 구조는 SE(Squeeze-and-Excitation) 모듈이 포함된 ResNet34 네트워크이다. 입력으로는 MFCC(Mel-Frequency Cepstral Coefficients)로 변환된 소리 정보를 사용하고, 출력의 경우 시각 인코더와 같은 크기의 특징을 추출하도록 한다.

2.3 어텐션 모듈

그림 3은 어텐션 모듈의 전체 구조를 보여준다. 각각의 인코더에서 개별로 생성된 시각 특징과 청각 특징의 시간적 동기화를 맞추기 위해 교차 어텐션 네트워크를 사용한다. 교차 어텐션 네트워크는 두

개의 기본적인 셀프 어텐션 네트워크로 이루어져 있고, 선형 계층 이후 두 개의 어텐션 쿼리(Q)를 교환하는 형태를 가진다. 이후 합쳐진 시청각 특징이 최종 셀프 어텐션 네트워크의 입력으로 사용되고, 셀프 어텐션을 통해 비정상 행동이 활성화되는 시점을 포착한다. 마지막으로, 각 프레임당 입력된 돼지가 활성화 돼지인지를 판단하는 확률값이 출력되고, 특정 임계값을 기준으로 돼지의 활성화 시점 프레임을 특정한다.



(그림 3) 어텐션 모듈 구조.

3. 실험 및 결과분석

본 절에서는 양돈농가의 영상에서 비정상 활성 돼지를 탐지하는 제안 방법의 성능을 평가한다. 실험을 위해 비정상 소리는 기침소리, 비명소리로 설정했다. 실험 환경으로 GeForce RTX 2070 8G GPU를 사용했으며, Pytorch 1.0.0을 사용해 학습 및 테스트를 진행하였다.

3.1 데이터

실험을 위해 세종시에 위치한 실제 양돈농가의 한 돈방에서 약 3일간 촬영된 영상 데이터를 사용했다. 영상 취득을 위해 돈방 바닥으로부터 약 3m 높이의 천장에 카메라를 설치하였고, 돈방 내 20마리 돼지를 대상으로 640×480의 해상도와 10FPS로 설정된 RGB 영상을 소리와 함께 수집했다. 이후 학습 및 테스트에 사용하기 위해 돼지의 비정상 소리의 시작 및 종료 시점을 기준으로 앞·뒤 0.5초의 여유를 두고 원본 영상을 잘라 영상 클립을 만들었다. 그 결과 1~3초 사이의 길이를 가지는 기침 소리 33개, 비명 소리 104개인 총 137개의 영상 클립을 수집했다.

수집된 클립을 대상으로 활성 돼지 탐지를 위한 annotation 작업을 수작업으로 진행했다. 해당 영상에서, 비정상 소리를 내는 돼지와 추가로 랜덤하게 2마리를 선정해 총 3마리 후보군 돼지를 annotation 대상으로 선별하였으며, 모든 프레임의 해당 돼지들에게 활성/비활성 바운딩 박스로 annotation을 진행하였다. 후보군 돼지를 3마리로 특정하는 이유는 활성 화자 탐지 연구에서 최대 3명의 화자 후보군만을 실험에 사용하는 것에서 비롯했고, 본 시스템의 성능과 참조 모델의 성능을 비교하고자 함에 있다.

3.2 데이터 증강

학습을 진행하기에 수집된 137개의 클립 수는 현저히 적다. 따라서 수집된 데이터 클립을 대상으로 데이터 증강을 수행하였다. 증강 작업은 영상 데이터의 모든 프레임에 vertical, horizontal, diagonal image flip을 적용하여, 원본 영상 대비 총 4배수 데이터를 추가로 생성했다. 다음으로 소리 데이터에는 SNR(Signal-to-Noise Ratio) 10, 20dB을 적용하여 원본 대비 총 3배수 소리 데이터를 생성 완료했다. 해당 작업 수행 결과, 기존 데이터의 총 12배인 1,644개의 데이터를 추가 생성했으며, 이를 활용하여 학습 1,586개(기침:364, 비명:1,222), 테스트 195개(기침:65, 비명:130)의 클립으로 실험을 진행했다.

3.3 실험 및 결과

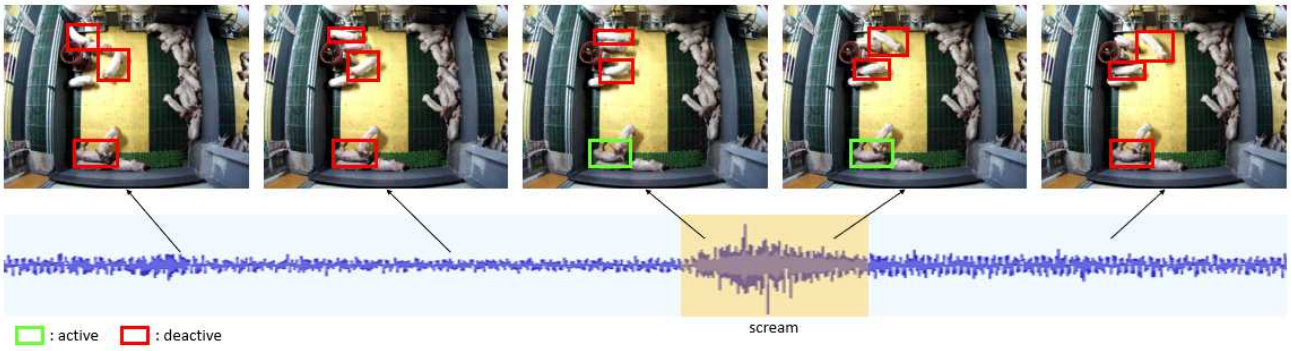
비정상 활성 돼지 탐지 시스템의 학습에는 Adam 옵티마이저를 사용해 50 에포크를 기준으로 학습을 진행했다. 초기 학습률은 10^{-4}이고, 매 에포크마다 5%씩 감소한다. 해당 학습에 입력되는 소리는 MFCC이며, 차원은 13이다. 또한, 돼지 영상의 크기는 112×112로 조정되어 입력 데이터로 사용된다. 인코더를 통해 출력되는 소리 및 영상 특징의 크기는 128로 설정했다.

<표 1> 데이터별 활성 돼지 탐지 시스템 성능 비교표

	Original data	Augmented data
mAP	0.715	**0.796**

표 1은 본 시스템의 정량적 결과를 보여주는 표이다. 평가 지표로는 활성 화자 문제에서 사용하는 mAP를 사용했다[8]. 원본 데이터만을 사용했을 때 mAP는 0.715의 상대적으로 낮은 성능을 기록했지만, 데이터 증강을 활용하여 데이터를 확장하여 실험을 수행한 결과 활성 돼지 탐지 성능이 0.796로 향상된 결과를 보였다. 이와 같은 mAP 성능 수치는 양돈 농가에서도 활성 돼지 탐지 가능성을 보여주는 결과라고 판단된다.

그림 4는 본 실험의 정성적 결과이다. 모니터링



(그림 4) 활성 돼지 탐지 정성적 결과.

되고 있는 돈사에서 '비명' 소리가 발생한 부분을 탐지했으며, 소리가 발생한 동일 시간대의 영상 프레임에서 다수의 움직이는 돼지들 속에서 활성 돼지를 정확하게 찾아내고 있음을 보여준다(녹색 선으로 표현된 돼지가 활성 돼지이다).

4. 결론

본 논문은 소리와 영상 정보가 모두 제공되는 모니터링 영상에서, 이상 상황이 발생한 활성 돼지를 식별하는 새로운 시스템을 제안하였다. 제안된 시스템을 검증하기 위해 실제 돈사에 녹음이 가능한 카메라를 설치한 후, 돼지의 정상 및 비정상 행동이 발생하는 영상을 수집하였다. 본 시스템의 결과 비정상 행동을 보인 특정 돼지를 탐지할 수 있음을 확인하였으며, 양돈농가에 적용 가능성을 확인하였다. 향후 본 시스템에 객체 탐지, 추적, 소리 탐지 등의 최신 기술들을 접목해 더욱 고도화된 시스템을 구축하고자 한다.

감사의 글

본 연구는 정부(교육부)의 재원으로 한국연구재단 (NRF-2020R1I1A3070835 and NRF2021R1I1A3049475) 사업의 지원을 받아 수행된 연구임.

참고문헌

[1] 한돈농가 생산성(MSY) 향상을 위한 연구, "http://data.han-don.com/f_index.php/menu4/view1/notice_no/3689/page/2", 2018.

[2] Y. Chung, S. Oh, J. Lee, D. Park, H. Chang, and S. Kim, "Automatic Detection and Recognition of Pig Wasting Diseases using Sound Data in Audio Surveillance Systems," Sensors, Vol. 13, No. 10, pp. 12929-12942, 2013.

[3] V. Exadaktylos, M. Silva, J. M. Aerts, C. J. Taylor, and D. Berckmans, "Real-time Recognition of Sick Pig Cough Sounds," Computers and Electronics in Agriculture, Vol. 63, No. 2, pp. 207-214, 2018.

[4] 최용주, 이종욱, 박대희, 정용화, "질감 분석과 CNN을 이용한 잡음에 강인한 돼지 호흡기 질병 식별", 정보처리학회논문지 소프트웨어 및 데이터 공학, Vol. 7, No. 3, pp. 91-98, 2018.

[5] H. Kim, J. Sab, B. Nohc, J. Leed, Y. Chung, and D. Park, "Automatic Identification of a Coughing Animal using Audio and Video Data," In Proceedings of the Fourth International Conference on Information Science and Cloud Computing, Guangzhou, China, pp. 18-19, 2015.

[6] AVA Challenge - Google Research, "https://research.google.com/ava/challenge.html", 2021.

[7] R. Tao, Z. Pan, R K.. Das, X. Qian, M. Z. Shou, and H. Li, "Is Someone Speaking? Exploring Long-term Temporal Features for Audio-Visual Active Speaker Detection," In Proceedings of the 29th ACM International Conference on Multimedia, pp. 3927-3935, 2021.

[8] Active Speaker Detection Evaluation Metric, "http://activity-net.org/challenges/2021/tasks/guest_ava.html", 2021.