

<https://doi.org/10.7236/JIIBC.2024.24.1.55>
JIIBC 2024-1-9

빅데이터 기반 미세먼지 이상 탐지 머신러닝 시스템 설계 및 구현

Design and Implementation of Machine Learning System for Fine Dust Anomaly Detection based on Big Data

이재원*, 인치호**

Jae-Won Lee*, Chi-Ho Lin**

요약 본 논문은 빅데이터 기반 미세먼지 이상 탐지 머신러닝 시스템 설계 및 구현을 제안한다. 제안하는 시스템은 빅데이터로 구성된 미세먼지 및 기상 정보를 통해 미세먼지 대기환경지수를 분류하는 시스템이다. 이 시스템은 머신러닝 기반의 대기환경지수 분류 카테고리별 이상치에 따른 이상치 탐지 알고리즘 설계를 통해 미세먼지를 분류한다. 카메라에서 수집된 영상의 심도 데이터는 미세먼지 농도에 따른 영상을 수집한 후 미세먼지 가시마스크를 생성합니다. 그리고 모노 심도 추정 알고리즘을 통한 학습 기반 핑거프린팅 기법으로 모노스코프 카메라에서 수집된 미세먼지의 가시거리를 추론하여 미세먼지 농도를 도출합니다. 본 방법의 실험 및 분석을 위해 미세먼지 농도 데이터와 지역별, 시간별 CCTV 영상 데이터를 매칭하여 학습 데이터를 생성한 후 모델을 생성하여 실제 환경에서 테스트한다.

Abstract In this paper, we propose a design and implementation of big data-based fine dust anomaly detection machine learning system. The proposed is system that classifies the fine dust air quality index through meteorological information composed of fine dust and big data. This system classifies fine dust through the design of an anomaly detection algorithm according to the outliers for each air quality index classification categories based on machine learning. Depth data of the image collected from the camera collects images according to the level of fine dust, and then creates a fine dust visibility mask. And, with a learning-based fingerprinting technique through a mono depth estimation algorithm, the fine dust level is derived by inferring the visibility distance of fine dust collected from the monoscope camera. For experimentation and analysis of this method, after creating learning data by matching the fine dust level data and CCTV image data by region and time, a model is created and tested in a real environment.

Key Words : Big Data, Fine Dust, Anomaly, Machine Learning

*학생회원, 세명대학교 컴퓨터학부

**정회원, 세명대학교 컴퓨터학부

접수일자 2024년 1월 23일, 수정완료 2024년 2월 6일

게재확정일자 2024년 2월 9일

Received: 23 January, 2024 / Revised: 6 February, 2024 /

Accepted: 9 February, 2024

**Corresponding Author: ich410@semyung.ac.kr

School of Computer Science, Semyung University, Korea

1. 서 론

COVID-19가 점차 사그러드는 지금에도 마스크를 사용하는 사람들을 길거리에서 찾기 쉽다. 마스크를 사용하는 사람들의 사용 다양한 이유가 있겠지만 그 중 미세 먼지 때문이라는 이유가 많은 비중을 차지하고 있다. 그만큼 대한민국에서 미세먼지는 일상생활에 많은 영향 미치고 있으며, 이에 대한 문제를 해결하기 위해 다양한 방법들을 제시하고 있다.^[1-4]

대부분 미세먼지에 대한 수치를 측정하기 위해 측정 센서를 공공기관 옥상이나, 초고층 랜드마크에 설치되어 시간 그리고 지역별로 측정되어 해당 정보를 기상청을 비롯한 다양한 매체를 통해 전달된다. 일부 도심지에서는 교차로 신호등에 설치되는 경우도 있으나, 대부분 사람이 미세먼지를 체감하는 높이에서 측정할 수 있는 측정소는 부족한 실정이다.^[4-6]

따라서 사람들이 체감되는 미세먼지 정도와 실제 측정 수치들의 차이를 줄여보기 위한 방법이 필요하며, 좀 더 직관적인 방법이 필요하다.

본 논문에서는 mono depth estimation 기반 미세 먼지 측정 방법을 제안한다. 제안하는 방법은 기존의 미세 먼지 측정 데이터와 이미지 데이터를 활용한 미세 먼지 측정 방법이다. 본 논문 구성은 다음과 같다. 2장에서는 mono depth estimation 기반의 미세 먼지 측정 방법에 대해 설명하며, 3장에서는 제안된 방법의 검증 실험 및 분석에 대해 다룬다. 4장에서는 결론으로 끝마친다.

II. 빅데이터 기반 미세먼지 이상 탐지 머신러닝 시스템 설계 및 구현

제안한 방법은 이미지를 통한 미세먼지 측정 방법이다. 이 방법은 이미지의 깊이 데이터를 통해 미세먼지 수치를 측정한다. 하지만 미세먼지의 크기는 너무 작기 때문에 저성능 카메라에서는 데이터를 수집되기 어렵다.^[7-9] 따라서 고안한 방법은 Mono depth estimation 알고리즘을 활용하는 방법이다. 먼저 Mono depth estimation을 통해 모노 스코프 카메라에서도 깊이 데이터를 도출하는 깊이 정보 모델을 생성한다. 그리고 기상청과 및 도시교통정보센서로부터 대기정보와 영상정보를 활용하며 제안하는 알고리즘을 학습하여 미세먼지 측정 모델을 생성한다. 미세먼지 측정 모델은 미세먼지 수치인 물질 농도(C)와 대기정보의 가시거리($L_v(km)$),

실험적 정수(A)를 이용하여 이미지 특징벡터로부터 가시거리와 실험적 정수를 추정한다. 식 (1)을 통해 가시거리를 도출할 수 있다.

$$L_v(km) = \frac{A \times 10^3}{C} \quad (1)$$

도출된 가시거리($L_v(km)$)는 실제 데이터 정보와 오차를 통해 머신러닝을 학습한다. 그림 1은 제안하는 방법을 도식화한 그림이다.

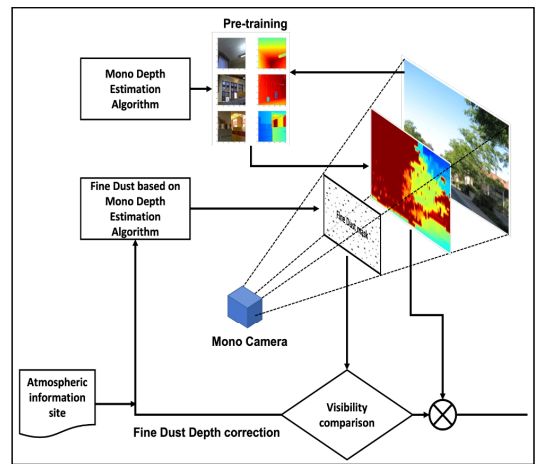


그림 1. mono depth estimation 기반 미세먼지 측정 방법 프로세스 여기에 그림 제목을 입력하세요
Fig. 1. Fine dust measurement method process based on mono depth estimation

미세먼지 영상은 안개와 같은 환경 및 실제 미세먼지 환경에서 취득된 영상이다. 미세먼지는 대기 중인 입자로부터 산란한 빛과 물체로부터 반사된 빛에 의해 취득된다. 입자의 밀도가 클수록 빛의 산란이 강해지고 영상 대조도(Contrast)와 채도(Saturation)가 감소한다. 이러한 환경에서 영상 깊이 추정은 단안 깊이 추정 알고리즘을 활용하며, 사전에 듀얼 영상 센서를 이용한 영상 깊이 데이터를 통해 영상 깊이 추정 모델을 생성할 수 있다. 그러나, 단안 영상 센서를 이용한 본 연구에 맞지 않아 Dark Channel Prior[8] 알고리즘을 이용한 영상 깊이 추정을 수행한다. 이를 위한 수식은 식(2)과 같이 나타낸다.

$$I(x) = J(x)t(x) + A(1 - t(x)) \quad (2)$$

$I(x)$ 는 실제 카메라에 찍히는 이미지로 안개로 인한 헤이지 영상, $J(x)$ 는 깨끗한 영상, $t(x)$ 는 대기 투과율, 그리고 A 는 대기 산란광 수치로 정의된다. 식 (2) 에서 $J(x)$ 와 $t(x)$ 을 구해 깨끗한 영상 추출할 수 있는 모델 생성할 수 있다. 그림 2은 Auto Encoder 기반의 $J(x)$ 와 $t(x)$ 추출 네트워크이다.

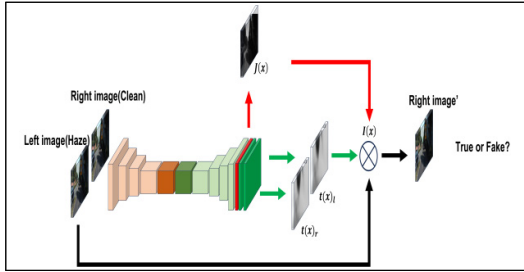


그림 2. 미세먼지 제거 및 단안 깊이 추정 네트워크
 Fig. 2. Fine dust removal and monocular depth estimation network

III. 실험 결과

제안하는 방법을 검증하기 위해 먼저 DIODE:고밀도 실내 및 실외의 깊이 데이터 세트를 사용하여 사전에 mono depth estimation을 학습 모델을 생성한다. 그 결과 그림 3과 같다.

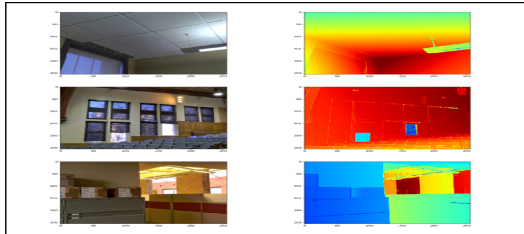


그림 3. mono depth estimation 학습 모델 결과
 Fig. 3. Mono depth estimation prediction model results

그 다음, 대기정보와 도시 교통정보 센터의 영상 정보에 기반한 미세먼지 이미지 데이터셋을 통해 미세먼지 측정 모델을 학습하였습니다. 학습 과정에서 모델의 정확도는 점차 증가하였으며, 물질 농도의 loss는 0.62, 실험적 정수의 loss는 0.75로 나타났습니다. 정확도는 각각 0.65와 0.77을 달성하였습니다. 이러한 결과는 모델이 학습을 통해 어떻게 개선되어 가는지를 시각적으로 보여주며, 최종적으로는 실제 미세먼지 농도의 AQI 지

수와 모델의 예측 값을 비교하여 모델의 예측 성능을 평가하였습니다. 이 결과는 모델이 실제 세계의 미세먼지 농도를 얼마나 정확하게 예측할 수 있는지를 보여주는 지표로 사용됩니다.

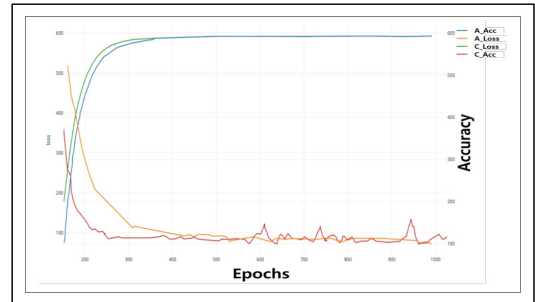


그림 4. 미세먼지 측정 모델 학습 그래프
 Fig. 4. Fine dust measurement model training graph

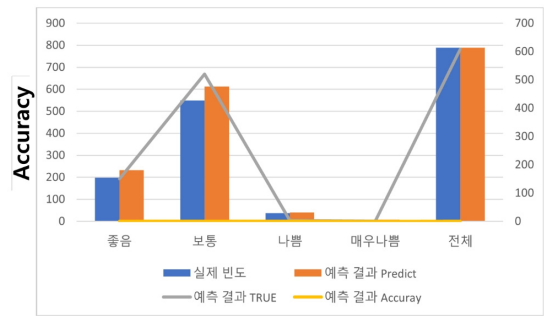


그림 5. 예측 결과 비교
 Fig. 5. Comparison of prediction results

IV. 결론

본 논문에서는 빅데이터 기반 미세먼지 이상 탐지 머신러닝 시스템 설계 및 구현을 제안하였다. 제안하는 방법은 기존의 미세먼지 센서 데이터가 실제 체감하는 수치와 오차 발생하는 문제를 해결위한 방법이다. 이 방법은 기상청과 도시교통정보센터로부터 대기 정보와 영상 정보를 활용하였으며, 영상 데이터에서 가시거리를 도출하는 것으로 미세먼지 농도를 추정하는 방법이다. 가시 거리를 도출하기 위해 mono depth estimation 알고리즘을 활용하였다. 이 알고리즘은 두 개의 모델을 생성한다. 하나는 깊이 정보를 추출하기 위한 모델, 또 하나는 가시 거리를 계산하기 위한 실험적 정수와 물질 농도를 머신러닝 학습으로 추정하는 모델이다. 실험적 정수 및

물질 농도를 통해 가지거리를 계산하고 실제 가지거리와 비교하여 최적의 파라미터를 도출하여 예측을 수행하였다. 가지거리 예측 모델을 통해 도출된 정보를 이용하여 AQI 지수 실험 결과 나쁨과 매우 나쁨 수치에 대한 예측이 8%, 0%를 보였으며, 이는 학습 데이터 중 해당하는 수치의 양이 부족하여 생긴 문제로 판단된다. 또한 mono depth estimation 알고리즘을 활용하였으나, 정확한 깊이 정보가 아닌 이유로 실제 외부 환경 및 다양한 변수에 대한 영향을 받는 불규칙적인 특성이 보여 정확한 예측의 한계가 보였다. 따라서 LiDAR 센서와 대기정보에 따른 빛의 산란 계산을 통한 가지거리 및 깊이 정보 데이터를 추출하는 알고리즘이 필요함을 확인할 수 있었다. 이를 통해 본 논문의 결과보다 예측 성능을 향상시킬 수 있을 것이며, 향후 다양한 알고리즘을 통한 예측 모델 성능 평가를 통해 제안한 방법의 최적화를 진행할 것이다.

References

- [1] G. W. Evans, "Air Pollution and Human Behavior," Journal of Social Issues, Vol. 37, No. 1, pp. 95-125, 1981.
- [2] M. S. Seo, "The Impact of Particulate Matter on Economic Activity," The Korean Women Economists Association, Vol. 12, No. 1, pp.75-100, Jun. 2015.
- [3] M. Beak, and H. S. An "Empirical Leisure Environment Satisfaction Evaluation of Public Institution Employees in Innocity," International JOURNAL OF CONTENTS, Vol. 19, No. 2, pp.368-378, Feb. 2019.
- [4] Godard, C., Aodha, O. M., and Brostow, G. J., "Unsuper-vised monocular depth estimation with left-right consistency.", CVPR. 2017
- [5] Karsch, K., Liu, C., and Kang, S., "Depth extraction from video using nonparametric sampling.", IEEE transactions on pattern analysis and machine intelligence, Vol. 36 No. 11, 2014
- [6] Lei Shao, Shuai, Yang, Hongli Liu, and Ji Li, "Research on Location Method of Climbing Robot based on Gyroscope", 2018 IEEE International Conference on Mechatronics and Automation (ICMA), pp. 238-242, 2018.
- [7] K. H. Kim, M. K. Cho, C. Y. Park, J. H. Kim, S. H. Kim, Y. H. Sun and J. Y. Kim, "Research on Unmanned Aerial Vehicle Mobility Model based on Reinforcement Learning". The Journal of The Institute

of Internet, Broadcasting and Communication (IIBC), Vol. 23, No. 6, pp.33-39, Dec. 31, 2023.

- [8] J. W. Kang, H. K. Sung, and K. H. Choi " Feature Matching based Training Data Construction Method for Moving Object Detection from Drone Images". Journal of the Korea Academia-Industrialcooperation Society. Vol. 24, No. 10 pp. 508-517, 2023
- [9] Hojin Ha, " Deep Learning-based Bitrate and Video Quality Prediction Model for Scalable Video Transmission", Journal of KIIT. Vol. 21, No. 12, pp. 163-170, Dec. 31, 2023.

저 자 소 개

이 경 민(학생회원)



- 2016년 : 세명대학교 컴퓨터학부 졸업
- 2016년 ~ 2019 : 세명대학교 일반대학원석사과정(컴퓨터학과 전공)
- 관심분야 : SoC CAD, 임베디드 시스템 신경, 머신러닝

인 치 호(정회원)



- 1985년 : 한양대학교 공과대학 전자공학과 졸업(공학사).
- 1987년 : 한양대학교 대학원(공학석사).
- 1996년 : 한양대학교 대학원(공학박사).
- 1992년 ~ 현재 : 세명대학교 컴퓨터학부 교수.
- 관심분야 : SoC CAD, ASIC 설계, CAD 알고리즘, AI SoC 설계, 지능형 컴퓨팅 및 임베디드 시스템, 지능형 교통시스템 설계,