Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 3, Issue 3, 2025

A NEW REAL-WORLD HAZY IMAGE DATASET FOR IMAGE ENHANCEMENT AND RECOGNITION

Sanaullah Memon[*1], Rafaqat Hussain Arain[2], Ghulam Ali Mallah[2], Sagar Lohana[3], Farheen Mirza[4]

[*1]*Department of Information Technology, Shaheed Benazir Bhutto University Shaheed Benazirabad, Pakistan.*
[2]*Institute of Computer Science, Shah Abdul Latif University Khairpur, Pakistan.*
[3]*Department of Computer Science, Shaheed Zulfikar Ali Bhutto Institute of Science and Technology (SZABIST) University Gharo Campus, Pakistan.*
[4]*School of Computing and Artificial Intelligence, Southwest Jiaotong University Chengdu, P.R. China*

[*1]sanaullah.memon_nf@sbbusba.edu.pk, [2]rafaqat.arain@salu.edu.pk, [2]ghulam.ali@salu.edu.pk, [3]sagar.lohana@ghr.szabist.edu.pk, [4]farheenmirza@my.swjtu.edu.cn

DOI: https://doi.org/10.5281/zenodo.15123637

Keywords
Real-world dataset, Computer vision, Image dehazing, Machine learning, Image enhancement, Image recognition

Article History
Received on 23February 2025
Accepted on 23March 2025
Published on 30 March 2025

Corresponding Author: *

Abstract
Many computer vision tasks are benefited significantly from deep learning. However, research in the field of single image dehazing is still needed. The unclear image boundaries and dense haze degrade the visible quality of haze-free images. For measuring the efficiency of techniques, many deep learning-based approaches are evaluated on the real-world hazy image datasets for more credibility. Moreover, this study discusses on creation of a dataset that includes real-world hazy images. The dataset is valuable for computer vision and image processing research. It offers training resource for deep neural networks and machine learning models to handle more hazy circumstances.

1. Introduction

There has been growing interest in the computer vision and image processing fields for research from last two decades [1]. It is imperative to measure the visual quality of image employing deep learning-based approach for single image dehazing. The real-world dataset gives more credibility to proposed approaches. The prompt adoption of a systematic strategy ensures the dataset diversity across weather, time, and locations. For generating a real-world image dataset, utilizing top-notch cameras is necessary for capturing clear images [2]. After collecting the hazy images, resizing with image editing software ensures consistency for efficient data processing [3][4]. The comprehensive understanding and modeling of the atmospheric light are necessary for effective dehazing [5]. The creation of hazy dataset requires estimation of the transmission map and atmospheric light for each hazy image, provides useful details for training and testing dehazing algorithms. The specifications of the hazy dataset of real-world images are shown in Table 1.

Table 1 Specifications of hazy dataset of real-world images

Subject	Computer Science
Specific subject area	Image Processing , Machine Learning
Data format	Raw images having jpeg format
Type of data	Hazy images of real-world objects
Data collection	The images are captured in outdoor hazy weather conditions between 7am to 6pm employing OPPO-Reno 6 and resized to 620 x 460 pixels.
Data source location	City: Khairpur, Sukkur, Hyderabad, and Karachi. Province: Sindh Country: Pakistan

2. Literature Review

The development of real-world hazy image datasets is essential for advancing image dehazing techniques and enhancing recognition systems. High-quality datasets enable the training and evaluation of algorithms under authentic conditions, ensuring their applicability in practical scenarios. This literature review highlights several significant contributions in this domain. Islam et al. [6] introduced HazeSpace2M dataset. This dataset is designed to improve dehazing through haze type classification. It encompasses diverse scenes with 10 haze intensity levels, featuring fog, cloud, and environmental haze. The dataset facilitates a novel approach where haze type classification precedes specialized dehazing, yielding improvements in image clarity. Benchmarking with state-of-the-art models described the challenging nature of HazeSpace2M, underscoring its significance in multimedia processing research. The real-world hazy image datasets such as I-HAZE[7]and O-HAZE[8]were introduced to evaluate single image dehazing techniques under controlled conditions. Paired hazy and haze-free images are provided by both datasets, enabling supervised deep learning approaches. However, their limited image pair sizes makes them less effective for large-scale deep learning training, leading to the development of newer datasets like NH-HAZE. The NH-HAZE dataset [9] addresses the limitations of synthetic datasets by providing 55 pairs of real hazy and corresponding haze-free images with non-homogeneous haze distribution. The images are captured using a professional haze generator to imitate real hazy conditions. It serves as a benchmark for evaluating single image dehazing techniques, emphasizing the complexity of non-uniform haze scenarios. The RTTS dataset [10] was introduced in the RESIDE dataset collection to bridge the gap between synthetic and real-world hazy image evaluation. It highlighted the performance degradation of object detection models in hazy conditions and the importance of dehazing. It lacks paired ground truth images compared to NH-HAZE but remains useful for real-world validation. The HazeRD dataset wasintroduced by Zhang et al.[11]. It focuses on outdoor scenes with realistic haze conditions. HazeRD offers 14 haze-free RGB images of real outdoor scenes paired with corresponding depth maps. It provides a MATLAB function to generate hazy images under different parameters, facilitating benchmarking of dehazing algorithms under physically realistic conditions. The RealBlur dataset was developed by Rim et al. [12]. It consists of real-world blurred images and corresponding sharp images, captured using a synchronized camera to ensure geometric alignment. While primarily aimed at deblurring research, RealBlur's methodology in acquiring real-world data offers valuable insights applicable to the collection of authentic hazy images for dehazing studies. S-HAZE[13]offers real-world hazy images and their ground truth images with varying haze densities, categorized based on the presence of sky regions. This categorization aids in assessing the performance of dehazing algorithms across different scene compositions. Low-Light Real-world indoor images dataset [1] comprises real-world indoor images captured under low-light conditions with varying haze densities. This dataset includes depth maps and infrared images, facilitating the development and evaluation of dehazing methods in complex indoor environments.

3. Value of the Data

- The dataset contains real-world hazy images which are collected at different weather conditions and haze levels.
- The dataset can be utilized for developing and improving image processing and deep learning techniques, leading to advancements in computer vision field.
- The dataset is supportive for training models to recognize and enhance images affected by haze.
- The dataset will help researchers in assessing their current deep learning models, trained on different datasets.
- The surveillance cameras might face challenges due to the adverse weather environment. Developing algorithms that diminish the effects of haze can significantly improve efficiency of surveillance system.

4. Data Description

The real-world dataset consists of 2000 hazy images. A systematic annotation scheme is created, spotting images as hazy and potentially sorting them by the level of haze as shown in Table 2. The dataset is organized into subfolders based on the locations and weather conditions. The appropriate file format such as JPEG for image storage is chosen and ensures they are consistent throughout the dataset. The detailed documentation about the dataset is maintained, including the data collection process, camera settings, and any post-processing applied to the images. The sample of real-world hazy images is shown in Figure 1.

Table 2 Details of real-world hazy image dataset

Level of haze	Locations			
	City	Village	Highway	Total
Thin haze	150	440	210	800
Thick haze	215	560	425	1200
			Grand Total	2000

Figure 1 Sample of real-world hazy images

5. Data Collection

Capturing hazy images of diverse hazy weather conditions and preparing a private dataset of hazy images involves a systematic approach. We travel to different locations across various cities of Pakistan, including Karachi, Hyderabad, Sukkur, and Khairpur to capture hazy images in diverse weather conditions. These locations are visited at different times of the day between 7 am to 6 pm. The high-quality camera features with adjustable settings of OPPO-Reno 6 are utilized to capture images in hazy conditions effectively. The focus is set manually to emphasize the subject while keeping the haze visible in the background. The exposure is reduced slightly to maintain details in brighter areas of the haze. The front view (0° angle) is used for balanced framing, ensuring the subject aligns with the light source for a glowing effect. The consistent camera settings and capture parameters are maintained to ensure that the images of the dataset are comparable and usable for research. The essential weather information for each image is recorded such as visibility, relative humidity, temperature, and other atmospheric conditions. The process of dataset creation is shown in Figure 2.

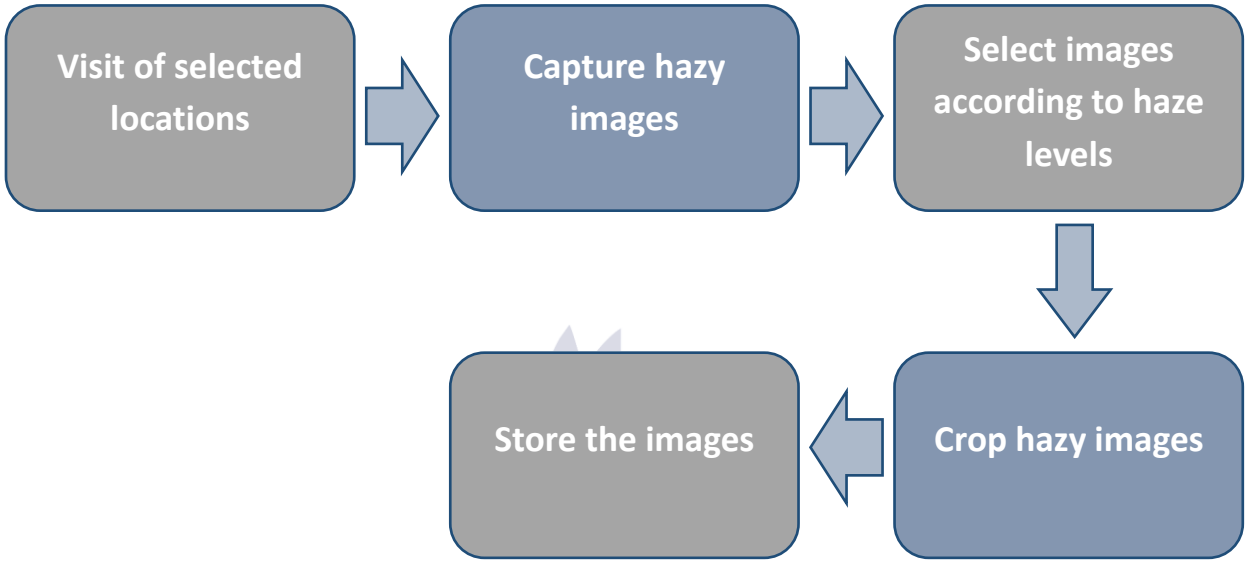

Figure 2 Process of dataset creation

6. Challenges

The following numerous challenges are faced while capturing hazy images from foggy weather conditions.

- It is challenging to capture sharp and clear images in dense fog. The visibility is severely limited that can result in images with poor contrast and reduced details.
- The amount of available light is reduced by fog, impacting the exposure setting required for photography. Adjusting settings can be tricky to balance for low light while avoiding overexposure.
- The quality of the captured images is impacted by hazy conditions, which may obscure minor details and textures.
- Fog can generate undesirable lens flare and lower contrast in images. It can be challenging to control without special equipment.
- Fog may be accompanied by moisture, which could damage camera equipment. To avoid damage, it is crucial to shield cameras and lenses from moisture and humidity.
- It could be difficult to construct visually appealing and relevant images when the foggy surroundings limit the visibility of the subjects and objects.

The challenges must be managed with the aid of dedicated tools. Moreover, dehazing algorithms can be applied in post-processing to reduce haze and enhance visual quality of image.

7. Conclusion

The dataset of real-world hazy images was generated using the controlled experimental setup. Using an OPPO-Reno 6 camera in foggy conditions, the outdoor images were captured and scaled to 620x460 pixels using python programming language. The dataset adds realistic levels of haze using hazy images

taken under various weather situations. Hazy images have a variety of visibility problems, which makes it possible to thoroughly assess image dehazing techniques. This dataset is valuable resource for computer vision and image processing research, particularly in the area of haze removal techniques.

Data Availability

Data will be provided on demand.

REFERENCES

[1] A. Filin, A. Kopylov, and I. Gracheva, "A Single Image Dehazing Dataset with Low-Light Real-World Indoor Images, Depth maps and Infrared Images," Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci., vol. 48, no. 2/W3-2023, pp. 53–57, 2023, doi: 10.5194/isprs-archives-XLVIII-2-W3-2023-53-2023.

[2] Z. Zheng and X. Jia, "4K-HAZE: A Dehazing Benchmark with 4K Resolution Hazy and Haze-Free Images," pp. 9–11, 2023, [Online]. Available: http://arxiv.org/abs/2303.15848

[3] H. Zhang, V. Sindagi, and V. M. Patel, "Joint Transmission Map Estimation and Dehazing Using Deep Networks," IEEE Trans. Circuits Syst. Video Technol., vol. 30, no. 7, pp. 1975–1986, 2020, doi: 10.1109/TCSVT.2019.2912145.

[4] H. Lu, Y. Li, S. Nakashima, and S. Serikawa, "Single image dehazing through improved atmospheric light estimation," Multimed. Tools Appl., vol. 75, no. 24, pp. 17081–17096, 2016, doi: 10.1007/s11042-015-2977-7.

[5] A. Kar, S. K. Dhara, D. Sen, and P. K. Biswas, "Progressive Update Guided Interdependent Networks for Single Image Dehazing," J. IEEE Trans. Artif. Intell., pp. 1–29, 2023, [Online]. Available: http://arxiv.org/abs/2008.01701

[6] I. M. Tanvir, R. Nasir, A. Saeed, M. Saqib, B. Sambit, and M. Khan, "HazeSpace2M : A Dataset for Haze Aware Single Image Dehazing," 2024.

[7] C. O. Ancuti, C. Ancuti, R. Timofte, and C. De Vleeschouwer, "I-HAZE: A dehazing benchmark with real hazy and haze-free indoor images," IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work., vol. 2018-June, pp. 867–875, 2018, doi: 10.1109/CVPRW.2018.00119.

[8] C. O. Ancuti, C. Ancuti, R. Timofte, and C. De Vleeschouwer, "O-HAZE: A dehazing benchmark with real hazy and haze-free outdoor images," IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work., vol. 2018-June, pp. 867–875, 2018, doi: 10.1109/CVPRW.2018.00119.

[9] C. O. Ancuti, C. Ancuti, and R. Timofte, "NH-HAZE: An image dehazing benchmark with non-homogeneous hazy and haze-free images," IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work., vol. 2020-June, pp. 1798–1805, 2020, doi: 10.1109/CVPRW50498.2020.00230.

[10] B. Li et al., "Benchmarking Single-Image Dehazing and beyond," IEEE Trans. Image Process., vol. 28, no. 1, pp. 492–505, 2019, doi: 10.1109/TIP.2018.2867951.

[11] Y. Zhang, L. Ding, and G. Sharma, "HazeRD: An outdoor scene dataset and benchmark for single image dehazing," Proc. - Int. Conf. Image Process. ICIP, vol. 2017-Septe, pp. 3205–3209, 2017, doi: 10.1109/ICIP.2017.8296874.

[12] J. Rim, H. Lee, J. Won, and S. Cho, "Real-World Blur Dataset for Learning and Benchmarking Deblurring Algorithms," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 12370 LNCS, pp. 184–201, 2020, doi: 10.1007/978-3-030-58595-2_12.

[13] S. Banerjee and S. S. Chaudhuri, "S-HAZE: dataset consisting of real world ground truth and hazy images of vaying haze density with no sky, little sky, and large sky regions" vol. 7, pp. 96–101, 2019.

.