OCM 2025

7th INTERNATIONAL CONFERENCE ON

Optical Characterization of Materials

MARCH 26nd – 27rd, 2025
KARLSRUHE | GERMANY

J. BEYERER | T. LÄNGLE | M. HEIZMANN (Eds.)

KIT Scientific Publishing

Jürgen Beyerer | Thomas Längle | Michael Heizmann (eds.)

OCM 2025

7th International Conference on Optical Characterization of Materials

March 26nd – 27rd, 2025
Karlsruhe | Germany

OCM 2025

7th International Conference on Optical Characterization of Materials

March 26nd – 27rd, 2025
Karlsruhe | Germany

edited by
Jürgen Beyerer | Thomas Längle | Michael Heizmann

KIT Scientific Publishing

Organizer

Fraunhofer Institute of Optronics,
System Technologies and Image Exploitation IOSB
c/o Karlsruhe Center for Material Signatures KCM
Fraunhoferstraße 1, 76131 Karlsruhe

The proceedings are also available as an online version
http://dx.doi.org/10.5445/KSP/1000178356

Impressum

Scientific
Publishing

Karlsruher Institut für Technologie (KIT)
KIT Scientific Publishing
Straße am Forum 2
D-76131 Karlsruhe

KIT Scientific Publishing is a registered trademark
of Karlsruhe Institute of Technology.
Reprint using the book cover is not allowed.

www.bibliothek.kit.edu/ksp.php | E-Mail: info@ksp.kit.edu | Shop: www.ksp.kit.edu

Print on Demand 2025 – Printed on FSC-certified paper

ISSN 2510-7240
ISBN 978-3-7315-1408-4
DOI 10.5445/KSP1000178356

Preface

The state of the art in optical characterization of materials is advancing rapidly. New insights into the theoretical foundations of this research field have been gained and exciting practical developments have taken place, both driven by novel applications and innovative sensor technologies that are constantly emerging. The big success of the international conferences on Optical Characterization of Materials in 2013, 2015, 2017, 2019, 2021 and 2023 proves the necessity of a platform to present, discuss and evaluate the latest research results in this interdisciplinary domain. Due to that fact, the international conference on Optical Characterization of Materials (OCM) took place the seventh time in Karlsruhe, Germany from March 26-27, 2025. The aim of this conference was to bring together leading researchers in the domain of Characterization of Materials by spectral characteristics from UV (240 nm) to IR (14 µm), multispectral image analysis, X-ray methods, polarimetry, and microscopy. Typical application areas for these techniques cover the fields of, e.g., food industry, recycling of waste materials, detection of contaminated materials, mining, process industry, and raw materials.

The OCM 2025 was organized by the Karlsruhe Center for Spectral Signatures of Materials (KCM) in cooperation with the German Chapter of the Instrumentation & Measurement Society of IEEE. The Karlsruhe Center for Spectral Signatures of Materials is an association of institutes of the Karlsruhe Institute of Technology (KIT) and the business unit Advanced Sensing of the Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB.

This year again the organizing committee has had the pleasure to evaluate a large amount of contributions. Based on the submissions, we selected 33 papers as posters and talks, a plenary lecture and several practical demonstrations. The present book is based on the conference and contains extended versions of the submitted abstracts.

The editors would like to thank all authors that have contributed to these proceedings as well as the reviewers, who have invested a generous amount of their time to suggest possible improvements of the papers. The help of Lukas Dippon and Jürgen Hock in the preparation of this book is greatly appreciated. Last but not least, we thank the organizing committee of the conference, led by Britta Ost, for their effort in organizing this event. The excellent technical facilities and the friendly staff of the Fraunhofer IOSB greatly contributed to the success of the conference.

March 2025

Jürgen Beyerer
Thomas Längle
Michael Heizmann

Automated data acquisition method for sensor-based real-time material flow characterization of recyclable waste streams using sensor fusion: A case study

Fabian Roth, Aydin Deliktas, Kaushik Rattawa, Alexander Feil, and Kathrin Greiff

RWTH Aachen University, Department of Anthropogenic Material Cycles, Wüllnerstraße 2, 52062 Aachen

Abstract In recent years, the development of real-time applications has become increasingly popular in the field of sensor-based systems and recyclable waste streams. One promising method is sensor-based real-time characterization (SBRTC) involving object detection or instance segmentation models as well as specific datasets containing recyclable materials to assess the quality of material flows. Building such models requires image data for training, testing and validation. This process is labor-intensive and prone to error, mainly when conducted manually. Here, we explore two approaches for the acquisition on conveyor belts: In approach I, a rotary encoder (a) and pre-defined time intervals (b) are compared to reduce acquisition gaps and redundancies, thereby improving data quality. The data acquisition was possible with a mean relative acquisition error (MRAE) of about 0% (a) and up to 66% (b). Approach II demonstrates the technical feasibility of object tracking which allows the counting of particles in a real-time video stream by leveraging Kalman Filter (KF), K-Nearest Neighbor (KNN), and Hungarian Maximum Matching (HMM). An accuracy of 99% ± 1% could be achieved. Therefore, this work contributes to novel data acquisition methods for high resolution RGB area images of SBRTC applications to effectively address the challenges of noisy and biased real-world datasets making it easier to perform data splitting.

Keywords Sensor-based real-time characterization, data acquisition, data quality, object tracking, computer vision

1 Introduction

In 2020, the European Union generated a total of 8.6 million tons of non-hazardous, non-ferrous metal waste, of which Germany accounted for 16.6%, ranking as the largest producer [1]. One key factor in reducing the environmentally and economically expensive consumption of primary non-ferrous raw materials is the transition towards a circular economy (CE). This aims at substituting more and more primary raw materials with secondary raw materials [2]. Therefore, reducing, reusing, and recycling has gained increasing importance, and within recycling, the characterization of recyclable waste streams has taken on a key role [3]. For this reason, gaining information about material characteristics is necessary for SBRTC of material flows. In order to leverage SBRTC, extracting frames from real-world video sequences is required to create novel datasets for inline analytics [4].
However, collecting and capturing images is both labor-intensive and time-consuming, particularly when conducted manually [5,6], as data collection, acquisition, and preparation can account for 45% to 90% of the total pre-processing time [7]. Moreover, a significant challenge in extracting frames from video sequences using an RGB area camera can be the issue of spatial and temporal correlation. That is, extracting two consecutive frames can lead to redundancies, such as the repeated occurrence of objects [8]. According to [6], a model's perceived performance can be artificially enhanced positively. This effect can be attributed to the fact that objects extracted twice can be part of both training and validation sets. Once a model is trained, it already knows the object. Moreover, duplicates only in training data can also falsely influence the model performance concerning generalization capabilities, as the model memorizes objects which are 'very similar', 'near duplicate', or 'exact duplicate' [9]. That is, data sparsity and redundancy can negatively impact the model's performance in its generalization capability by skewing the model and introducing a bias. This bias is also known as data leakage [9,10].
An automated data acquisition method for SBRTC of recyclable waste streams using sensor fusion offers the possibility of overcoming these limitations, i.e., avoiding acquisition gaps and redundancies.

Prior research has already demonstrated the suitability of rotary encoders and various sensor types. While in the automotive and manufacturing industries, rotary encoders are used to measure the movement of conveyor belts accurately [11], an optical encoder was used by [12] to start the acquisition of a line scan camera once the conveyor belt reached a stable velocity for the reconstruction of a 3D material transportation status. Another study demonstrated the calculation of the object length on a conveyor belt as it passes the scanline of a LiDAR camera [13]. Furthermore, the combination of RGB and magnetic sensors allows the determination of the rotational movements while utilizing an endoscopic capsule robot [14]. To deal with overlapping images, another study demonstrated a static acquisition method for scrap, where five frames per second were saved [15]. Moreover, rotary encoders allow high reliability and accuracy by converting a rotary motion into a digital or analog signal comprising several pulses per revolution [11]. However, using rotary encoders with RGB area cameras for data acquisition, i.e., preventing duplicate particle occurrences in the dataset by synchronizing camera acquisition framerate and conveyor belt speed, seems largely unexplored in literature. That is, sensor-based systems using optical sensors in combination with a rotary encoder and object tracking represents a relatively unsophisticated technology, thus offering untapped potential to improve dataset creation. For this reason, this study aims at improving the data quality of a data acquisition process by exploring methods to address spatial redundancies in consecutive frames and accurately determine ground truth data (i.e., true particle values) of real-time video sequences. To this end, we address the following research questions:

- RQ1: What is the impact of data acquisition without rotary encoder and how can spatial redundancies be prevented using a rotary encoder in order to improve data quality?
- RQ2: To which extent is it possible to determine the ground truth data of video sequences in real-time?

2 Materials and Method

In this study, two approaches are investigated: Approach I compares two methods to overcome the issue of spatial and temporal correlation. For approach II, object tracking is leveraged to determine the dataset's ground truth.

2.1 Acquisition principles

The experimental setup for data acquisition comprises several key components, as shown in Fig. 1: An industrial RGB camera (GV-77Q5WP-C-HQ) (a) with a resolution of 4500 by 4500 pixels, a pixel size of 3.2 µm and a color depth of 12-bit to extract color features from a 2D image with a framerate of 42 frames per second (FPS) at full resolution. This results in 0.089 mm per pixel with a recording area of 370.3 mm by 370.3 mm, which is illuminated by a diffuse illumination chamber (Planistar 60-60-Sled-3-VAD-19w-O, 5700K) (b). Images were taken by triggering a rotary sensor (MSK320) (c), which has a maximum angle resolution of 0.006 °. Additionally, a rotary encoder (MR320) (d) is mounted at the motor shaft, which sends up to 2000 impulses per revolution on a single signal line(SIKO Global 2024). Afterwards, an Arduino Nano v3 (e) receives data, which is transmitted to an industrial workstation (f). Processing the received data in (e) requires the circumference of a circle, which includes the radius of the motor shaft (60 mm) and the thickness of the belt (1.8 mm). Following, the total distance moved by the rotary encoder is compared with the total frame length of the recording area (y-axis). Once the rotary encoder has been moved to 370.3 mm, the total distance is set to zero, and the current image is saved. Otherwise, the distance travelled by each impulse is accumulated until the total distance is reached. Consequently, the recording software requests a periodic update from the Arduino Nano v3, repeating the acquisition process until the conveyor belt speed reaches zero.

2.2 Experiment I: Spatial and temporal correlation (RQ1)

Extracting frames from a video sequence requires addressing the challenge of spatial and temporal correlation. According to [16], spatial and temporal correlation is fundamental to analyzing object movement in

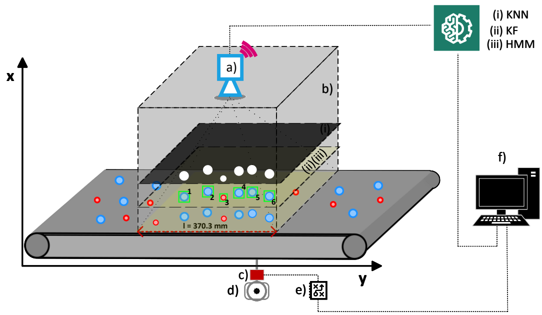

Figure 1: General structure of the experimental setup: (a) rgb camera, (b) illumination chamber, (c) rotary sensor, (d) rotary encoder, (e) Arduino Nano v3, f) industrial workstation and (i) K-Nearest Neighbor (KNN) background subtraction, (ii) Kalman Filter (KF), (iii) Hungarian Maximum Matching (HMM).

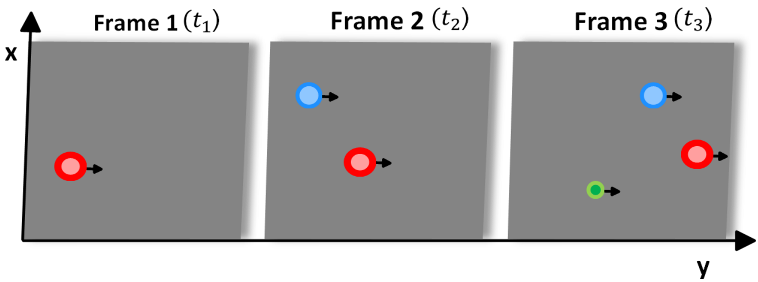

Figure 2: Consecutive frames over time (t_n) of a video sequence with particles in $t_1 =' red'$, $t_2 =' blue'$, $t_3 =' green'$.

a video sequence as objects change their positions dynamically. In this context, the spatial correlation is used to describe the location of objects within a single frame. In contrast, the temporal correlation describes the relationship between consecutive frames in a video sequence. As shown in Fig. 2, the position of the objects 'red', 'green, and 'blue' changes over time across consecutive frames.

As discussed in Section 1, one method for extracting frames from a video sequence is to select the sixth frame (static), while another approach involves using an encoder (dynamic). For this reason, three speed levels were initialized at 39 FPS, i.e. $0.1\frac{m}{s}$, $0.15\frac{m}{s}$ and $0.24\frac{m}{s}$ (see Table 1), whereby the speed interval was set in incremental steps. The time interval $\mathrm{t_{woe}}$ for the static method is a fixed time interval that remains constant at different speed levels. It was calculated using an r.p.m. counter (PCE-DT 66) and the frame length of 370.3 mm. Each n-frame ($\mathrm{f_{woe}}$) was captured, e.g. a frame was extracted after 3.7 seconds at $0.1\frac{m}{s}$, which equates to one image being captured every 144.42 frames. Using an encoder involves the dynamic time interval $\mathrm{t_{we}}$, which varies at different speed levels or intervals. That is, every frame ($\mathrm{f_{we}}$) is extracted depending on the conveyor belt's speed.

Speed Level	**Speed Interval**	**Time Interval**		**Frame Interval**	
$[\frac{m}{s}]$	$[\frac{m}{s}]$	t_{woe}	t_{we}	f_{woe}	f_{we}
0.10	0.10	3.70 s	3.70 s	144.42	144.42
0.10	0.15	3.70 s	2.47 s	144.42	96.28
0.10	0.24	3.70 s	1.54 s	144.42	60.17
0.15	0.10	2.47 s	3.70 s	96.28	144.42
0.15	0.15	2.47 s	2.47 s	96.28	96.28
0.15	0.24	2.47 s	1.54 s	96.28	60.17
0.24	0.10	1.54 s	3.70 s	60.17	144.42
0.24	0.15	1.54 s	2.47 s	60.17	96.28
0.24	0.24	1.54 s	1.54 s	60.17	60.17

Table 1: Speed levels and intervals at 39 FPS with the corresponding time interval (t_{woe};t_{we}) and frame interval (f_{woe};f_{we}).

To evaluate the acquisition performance, 50 particles (n_p) were recorded at pre-defined speed and time intervals. Afterwards, the total number of particles (c_p) within the recorded images was manually counted to calculate the mean relative acquisition error (MRAE), which can be denoted as Eq. (1):

$$\mathrm{MRAE}(n, c) = \frac{1}{N} \sum_{p=1}^{N} \frac{|n_p - c_p|}{|n_p|} \tag{1}$$

Where N is the number of the speed intervals.

2.3 Experiment II: Object tracking

In order to acquire ground truth particles at different speed levels (see Table 2) from a monolayer input stream, object tracking was utilized by leveraging three distinct algorithms, including KNN, KF and HMM (see Fig. 1) : (1) KNN is a pixel-level background subtraction method that determines the background by analysing the nearest neighbours in a pixel's "short-term-long-term" history [17]. In other words, the number of frames (or pixel values) used to model the background is determined by the history of the model (i). The model's sensitivity to changes is controlled by the Dist2Threshold (ii). Both parameters were set to 60 (i) and depending on the speed level: 200, 300, 320, and 340, respectively (ii). (2) KF is a statistical method for estimating the future position of an object based on its previous and current position, including both process noise covariance (PNC) (i) and measurement noise covariance (MNC) (ii). While (i) quantifies the degree of uncertainty associated with the behaviour and movements of the tracked object, (ii) represents the uncertainty inherent in the observations, capturing the inaccuracies and noise present in the measurement process [18]. In enhancing KF's responsiveness to changes at higher speed levels, (i) a scaling factor of 0.9 was set using a 4x4 matrix. Furthermore, a 2x2 matrix with a scaling factor of 0.5 and 0.3, respectively (ii), was defined to prevent missing changes in fast-moving objects. (3) In order to overcome the assignment problem in consecutive frames, HMM leverages inputs from (1) and (2) by calculating a cost matrix with a maximum distance of 70, thus assigning unique identifiers to each of the objects [19].

3 Results and Discussion

3.1 Experiment I: Spatial and temporal correlation (RQ1)

The data acquisition process could achieve a mean relative acquisition error of 27.3% at 0.1 $\frac{m}{s}$, 24.7% at 0.15 $\frac{m}{s}$, and 66% at 0.24 $\frac{m}{s}$. Furthermore, the relative acquisition error at 0.1 $\frac{m}{s}$, at 0.15 $\frac{m}{s}$ and 0.24 $\frac{m}{s}$ is zero, as the acquisition rate relates to the conveyor belt speed. An increase in conveyor belt speed from 0.1 $\frac{m}{s}$ to 0.15 $\frac{m}{s}$ and 0.24 $\frac{m}{s}$ with the same acquisition rate results in a relative acquisition error of 34% (48%).

Speed level	Speed interval	KF		KNN		HMM
$[\frac{m}{s}]$	$[\frac{m}{s}]$	PNC	MNC	History	Dist2 Threshold	max Distance
0.10	0.10	$I(4 \times 4) \cdot 0.9$	$I(2 \times 2) \cdot 0.5$	60	320	70
0.10	0.15	$I(4 \times 4) \cdot 0.9$	$I(2 \times 2) \cdot 0.5$	60	300	70
0.10	0.24	$I(4 \times 4) \cdot 0.9$	$I(2 \times 2) \cdot 0.5$	60	320	70
0.15	0.10	$I(4 \times 4) \cdot 0.9$	$I(2 \times 2) \cdot 0.3$	60	340	70
0.15	0.15	$I(4 \times 4) \cdot 0.9$	$I(2 \times 2) \cdot 0.3$	60	340	70
0.15	0.24	$I(4 \times 4) \cdot 0.9$	$I(2 \times 2) \cdot 0.3$	60	340	70
0.24	0.10	$I(4 \times 4) \cdot 0.9$	$I(2 \times 2) \cdot 0.3$	60	200	70
0.24	0.15	$I(4 \times 4) \cdot 0.9$	$I(2 \times 2) \cdot 0.3$	60	200	70
0.24	0.24	$I(4 \times 4) \cdot 0.9$	$I(2 \times 2) \cdot 0.3$	60	200	70

Table 2: Object tracking parameters for (i) KF, (ii) KNN and (iii) HMM based on speed levels and intervals.

Starting with a speed level of 0.15 $\frac{m}{s}$ while decreasing (increasing) the conveyor belt speed to 0.1 $\frac{m}{s}$ (0.24 $\frac{m}{s}$) leads to a relative acquisition error of 38% (36%). The relative acquisition error at the speed level of 0.24 $\frac{m}{s}$ is 64% when the speed is decreased to 0.15 $\frac{m}{s}$ and 134% when it is decreased further to 0.1 $\frac{m}{s}$. As shown in Figure 3a, the number of particles counted at a speed level of 0.24 $\frac{m}{s}$, with a speed interval of 0.1 $\frac{m}{s}$, is 117. That is, the acquisition rate depends on the conveyor belt speed, which remains constant at this specific speed level and interval. Furthermore, an increase in the conveyor belt speed at a constant acquisition rate, for example, from 0.1 to 0.24 $\frac{m}{s}$, results in 48% of the particles being missed, whereas a decrease in the conveyor belt speed from 0.24 to 0.1 $\frac{m}{s}$ results in more than double the number of particles. Consequently, this demonstrates that pre-defined time intervals probably do not enhance model performance, even when employing the best deep-learning models. In contrast, the use of a rotary encoder couples the temporal component to the conveyor belt speed, resulting in a relative acquisition error of zero for each speed level. That is, the spatial correlation shows neither redundant nor missing particles, whereas edge cases cannot be avoided, as there are 1 at 0.1 $\frac{m}{s}$ and 1 at 0.24 $\frac{m}{s}$. Nevertheless, this enables precise control of the data acquisition

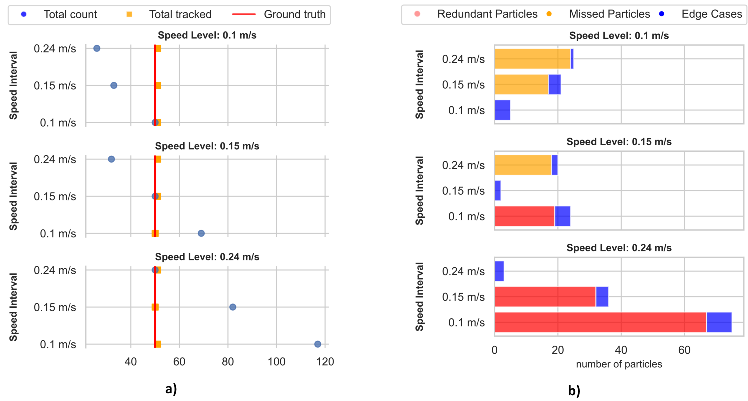

Figure 3: Acquisition without encoder using time intervals (a) Manual counted vs. tracked particles in consecutive frames and (b) redundant and missed particles with edge cases at different speed levels: $0.1\frac{m}{s}$, $0.15\frac{m}{s}$ and $0.24\frac{m}{s}$.

system through automated fine-tuning of the acquisition rate based on the conveyor belt speed. This ensures data integrity of datasets used for deep-learning models, thereby helping to create datasets that are free from bias and noise.

3.2 Experiment II: Object tracking (RQ2)

Additionally, object tracking provides the capability to count particles in consecutive frames within the acquisition system, independently of whether it is controlled by a rotary encoder. Consequently, the combination of KF, KNN, HMM and a rotary encoder ensures the real-time creation of datasets with ground truth information, achieving a count accuracy of almost 100%. As shown in Fig 3.a, the system can track particles with a deviation of ± 1 particles (=total tracked) of the total ground truth. This dual capability serves to reduce labor time and enhance the validation of the data acquired by using knowledge of the ground truth to address class imbalance in classification problems. It also enables targeted strategies to effectively balance different classes. However, increasing the conveyor belt speed above 0.24 $\frac{m}{s}$ requires enhancing the responsiveness of the KF to changes by increasing the PNC.

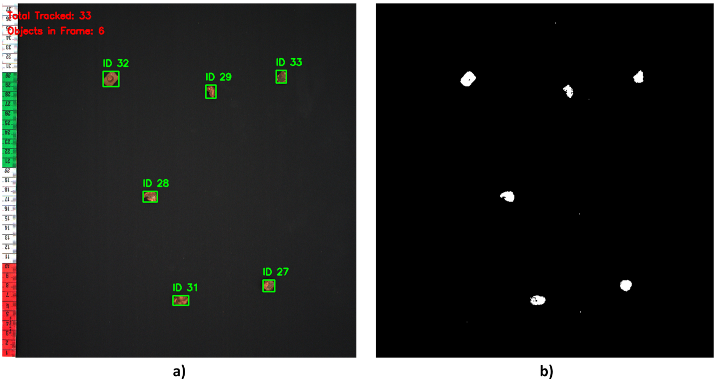

Figure 4: a) object tracking: unique identifiers for each particle by b) extracting particles from background using background subtraction.

Consequently, a reduction in MNC is essential to prevent the system from missing rapid changes. As shown in Fig. 4, background subtraction allows particles to be separated from the background. In this context, a lower Dist2Threshold provides the capability to track fast-moving objects, while the history remains resilient to higher conveyor belt speeds (see Table 2).

4 Conclusion and Outlook

In this study, an automated data acquisition method for dataset creation was investigated by leveraging two approaches. The first approach allows for addressing the issue of temporal and spatial correlation to deal with data sparsity and redundancy by avoiding noisy and biased data. The mean relative acquisition error was observed to be 27.3% at 0.1 $\frac{m}{s}$, 24.7% at 0.15 $\frac{m}{s}$, and 66% at 0.24 $\frac{m}{s}$ when extracting frames using a predefined time interval. In contrast, employing a rotary encoder achieved an impressive accuracy of 100%. On the other hand, object tracking can track particles in real-time to determine the dataset's ground truth with an accuracy of 99% ± 1% to allow for enhanced data validation by leveraging ground truth knowledge to tackle class imbalance in classification problems, ultimately leading to more reliable and robust deep-learning models. It remains to be seen how performance changes at higher speed levels. Furthermore, a data acquisition system that considers dynamic temporal correlation has the potential to enhance the data quality of high resolution RGB area image datasets, thus allowing for higher generalization capabilities as a trained model does not memorize duplicated objects (i.e., skewing the model). Consequently, the capacity for more effective material assessment is enabled for sensor-based real-time characterization while reducing labor time and biased training, testing, and validation data during pre-processing. Further research is recommended to explore the effects of higher conveyor belt speeds, which is necessary for scaling up to technical or industrial scale aimed at acquiring novel datasets from various recyclable waste streams. Additionally, decoupling of the conveyor belt speed from a rotary encoder could enable the design of an RGB-based speed measurement system for flexible integration in industrial applications.

Acknowledgements

We would like to thank REMONDIS SmartRec GmbH for providing the non-ferrous metal sample material and technical assistance.

References

1. Eurostat, "Generation of waste by waste category, hazardousness and nace rev. 2 activity." [Online]. Available: https://ec.europa.eu/eurostat/databrowser/view/env_wasgen__custom_12053946/default/table?lang=de
2. European Commission, "Eur-lex - 52020dc0098 - en - eur-lex," 2020. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1583933814386&uri=COM:2020:98:FIN
3. N. Kroell, X. Chen, K. Greiff, and A. Feil, "Optical sensors and machine learning algorithms in sensor-based material flow characterization for mechanical recycling processes: A systematic literature review," *Waste management (New York, N.Y.)*, vol. 149, pp. 259–290, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0956053X22002628

4. G. Lanza, B. Haefner, L. Schild, D. Berger, N. Eschner, R. Wagner, and M. Zaiß, "In-line measurement technology and quality control," in *Metrology*, ser. Precision Manufacturing, W. Gao, Ed. Singapore: Springer, 2019, pp. 399–433.
5. M. Elasri, O. Elharrouss, S. Al-Maadeed, and H. Tairi, "Image generation: A review," *Neural Processing Letters*, vol. 54, no. 5, pp. 4609–4646, 2022.
6. K. Li, M. I. Fathan, K. Patel, T. Zhang, C. Zhong, A. Bansal, A. Rastogi, J. S. Wang, and G. Wang, "Colonoscopy polyp detection and classification: Dataset creation and comparative evaluations," *PloS one*, vol. 16, no. 8, p. e0255809, 2021.
7. S. E. Whang, Y. Roh, H. Song, and J.-G. Lee, "Data collection and quality challenges in deep learning: a data-centric ai perspective," *The VLDB Journal*, vol. 32, no. 4, pp. 791–813, 2023.
8. H. Zhu, H. Wei, B. Li, X. Yuan, and N. Kehtarnavaz, "A review of video object detection: Datasets, metrics and methods," *Applied Sciences*, vol. 10, no. 21, p. 7834, 2020.
9. B. Barz and J. Denzler, "Do we train on test data? purging cifar of near-duplicates," *Journal of Imaging*, vol. 6, no. 6, 2020.
10. S. Shahinfar, P. Meek, and G. Falzon, ""how many images do i need?" understanding how sample size per class affects deep learning model performance metrics for balanced designs in autonomous wildlife monitoring," *Ecological Informatics*, vol. 57, p. 101085, 2020.
11. F. Paredes, C. Herrojo, and F. Martín, "Position sensors for industrial applications based on electromagnetic encoders," *Sensors*, vol. 21, no. 8, 2021.
12. Z.-L. Li, Q. Wu, F. Zeng, and Z.-S. Yue, "3d reconstruction of material transportation status on belt conveyors," in *2015 12th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*. IEEE, 2015, pp. 168–172.
13. A. Del Bimbo, R. Cucchiara, S. Sclaroff, G. M. Farinella, T. Mei, M. Bertini, H. J. Escalante, and R. Vezzani, Eds., *Pattern Recognition. ICPR International Workshops and Challenges*, ser. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2021.
14. M. Turan, J. Shabbir, H. Araujo, E. Konukoglu, and M. Sitti, "A deep learning based fusion of rgb camera information and magnetic localization information for endoscopic capsule robots," *International journal of intelligent robotics and applications*, vol. 1, no. 4, pp. 442–450, 2017.
15. M. Schäfer, U. Faltings, and B. Glaser, "Does - a multimodal dataset for supervised and unsupervised analysis of steel scrap," *Scientific data*, vol. 10, no. 1, p. 780, 2023.

16. J. Zhang and Y. Peng, “Video captioning with object-aware spatio-temporal correlation and aggregation,” *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 2020.

17. Z. Zivkovic and F. van der Heijden, “Efficient adaptive density estimation per image pixel for the task of background subtraction,” *Pattern Recognition Letters*, vol. 27, no. 7, pp. 773–780, 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865505003521

18. M. Khodarahmi and V. Maihami, “A review on kalman filter models,” *Archives of Computational Methods in Engineering*, vol. 30, no. 1, pp. 727–747, 2023. [Online]. Available: https://link.springer.com/article/10.1007/s11831-022-09815-7

19. A. Frank, “On kuhn’s hungarian method—a tribute from hungary,” *Naval Research Logistics (NRL)*, vol. 52, no. 1, pp. 2–5, 2005.