 DOI: 10.3217/978-3-85125-869-1-16

Multi-Spectral Segmentation with Synthesized Data for Refuse Sorting

Harald Ganster[1] Alfred Rinnhofer[1] Georg Waltner[2] Christian Payer[2]
Heimo Gursch[3] Christian Oberwinkler[4] Reinhard Meisenbichler[4] Horst Bischof[2]

Abstract— **Refuse sorting is a key technology to increase the recycling rate and reduce the growths of landfills worldwide. However, monitoring and parameterization of sorting facilities is still done in a mostly static fashion. This work combines multi-spectral imaging with deep learning based image recognition to monitor and dynamically optimize processes in sorting facilities. Our solution is capable of monitoring the sorting process remotely avoiding potentially harmful working conditions due to dust, bacteria, and fungal spores. Furthermore, the introduction of objective sorting performance measures enables informed decisions to improve the sorting parameters and react quicker to changes in the refuse composition.**

I. INTRODUCTION

The global refuse production is still on the increase worldwide, since the refuse output increases faster than the recycling rates [9]. The ever-changing refuse composition poses a major challenge to automated sorting in recycling application. This work presents preliminary findings of KI-Waste [5] capturing the refuse composition on conveyor belts in a refuse sorting facility. This is done by multi-spectral imaging and deep learning for semantic segmentation and object recognition on refuse streams at key points in the sorting facility.

II. RELATED WORK

Sorting facilities extract usable fractions with sorting and shredding machines connected by conveyor belts [4]. Image recognition applied to the refuse streams on these conveyor belts is capable of capturing the refuse composition since different substances have different spectral reflection characteristics. Thus, multi-spectral cameras can provide a spectral fingerprint of the material streams on the conveyor belts [12], [13]. A four-channel setup is often used consisting of RGB plus near-infra-red (NIR) cameras [3], [15], [11], [10]. In addition to these two-dimensional (2D) multi-spectral systems, a tree-dimensional (3D) acquisition can capture geometric properties useful in automatic material separation.

The resulting images of the refuse on the conveyor belt are the input for image recognition software identifying predefined refuse categories on a pixel-wise basis. Traditional image recognition techniques based on color and gradient features are typically not able to handle the large variations in appearance and shape occurring in mixed-material streams. Convolutional neural networks (CNNs) [7] have shown great performance on a variety of image recognition tasks including semantic segmentation [8], where a category label is assigned to each pixel of an image. The results of the image recognition are a good basis for predictive maintenance, optimization, automation and self-adaptation of the refuse sorting process [14].

III. IMAGE CAPTURING AND CLASSIFICATION

The high variety in substances and the challenging environmental conditions like dust, dirt, lighting, temperature, and vibrations make the image capturing challenging. We overcome these challenges by employing a line-scan-based multi-spectral system with a light-sectioning method that uses laser-line projection to determine surface profiles. It outputs high-quality four-channel multi-spectral 2D images and 3D registered image data.

The hardware setup is designed so that all capturing devices cover the same acquisition area. Nevertheless, calibration methods are required, registering the captured image data to each other. Finally, all image modalities are transformed into one common coordinate system by geometric mapping, ensuring that each pixel has a direct correspondence between geometric and spectral information.

The image classification segments each image pixel-wise into the predefined refuse categories by state-of-the-art fully-convolutional CNNs with a huge number of trainable parameters. To set these parameters in a meaningful way, CNNs need to be trained with hundreds or thousands of representative ground truth images, where each pixel is correctly annotated with its category.

Creating this ground truth manually requires an enormous labeling effort. Hence, this project uses empty belt images and images of mono-material refuse streams to effortlessly create ground truth labels as shown in Fig. 1 (top row). With this groundtruth, we can synthetically create realistic mixed-material images with known proportions and locations of refuse categories. This way we can generate unlimited amounts of annotated mixed-material images as depicted in Fig. 1 (bottom row).

IV. INITIAL PROJECT RESULTS

We train and evaluate the proposed approach within the DeepLabv3+ [2] framework. Our training consists of two steps. First, we train a binary segmentation model to distinguish belt vs. waste. In the second step, we use this model to generate groundtruth for the known mono-material images

*This work was supported by Land Steiermark within the research program „Zukunftsfonds“(No. 1330).

[1]JOANNEUM RESEARCH Forschungsgesellschaft mbH, DIGITAL, harald.ganster@joanneum.at

[2]Graz University of Technology, Institute of Computer Graphics and Vision waltner@icg.tugraz.at

[3]Know-Center GmbH hgursch@know-center.at

[4]Komptech GmbH

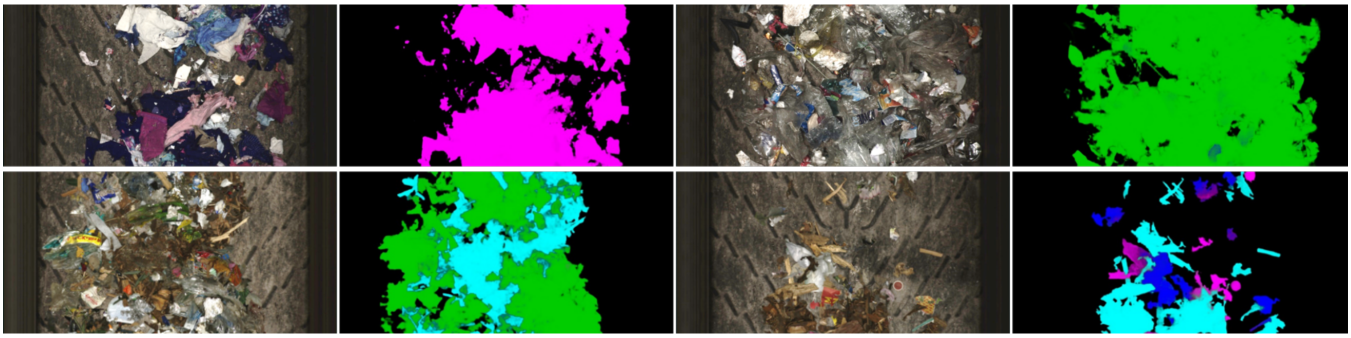

Fig. 1. Mono-material stream samples and binary segmentation belt/waste (top row) and synthetic ground-truth data with mixed stream (bottom row). All images: plastics (green), wood (turquoise), textiles (purple), paper (blue).

and use this groundtruth to train a model on synthetic mixed-material images. Manually labeling only a few waste images is sufficient for the binary network to train an initial model, which is further improved with the automatically generated labels it produces. Having a well working binary model, we use it to label the mono-material images. We then generate 130*k* superpixels [1] of different sizes from 492 mono-material images for our synthetic training regime, holding back 324 for testing. During training of our multi-class model, the synthetic mixed-material images are generated on-the-fly to guarantee a diverse training set. We train the network for 500*k* iterations with batch size 8 and Adam optimizer [6] using an initial learning rate of 0.0001 and a decay of 0.1.

As it is almost impossible to manually annotate mixed-material streams even for a trained person, we limit evaluation of the multi-class model to mono-material recordings. A great accuracy of $84 - 100\%$ can be observed on the refuse fractions *clothes*, *paper*, *plastic* and *wood*. Most of the fractions are very well classified except for *wood* that is partly misclassified as paper, as the confusion matrix in Fig. 2 shows. The reduction of these confusions is topic of an ongoing refinement and validation. In addition, while we cannot measure the performance due to the lack of groundtruth, we can visually observe very promising results produced by our trained CNN model also on real mixed-material streams, as shown in Fig. 3.

	belt	clothes	paper	plastic	wood
belt (24)	100.00	0.00	0.00	0.00	0.00
clothes (75)	0.00	99.88	0.04	0.02	0.06
paper (90)	0.00	1.09	97.46	1.44	0.01
plastic (90)	0.00	2.84	0.13	95.75	1.28
wood (45)	0.00	2.92	10.46	2.35	84.27

Fig. 2. Confusion matrix with pixel-wise accuracies in % for 324 test images. Apart from minor confusions between *wood* and *paper*, the performance of the CNN model is very promising.

Several properties of the visible refuse can be calculated when combining the semantic segmentation output with the 3D surface information, e. g. refuse category distribution, particle size, and the height of specific regions of the image. These properties will then be used for further analysis and refuse processing parameters adjustment.

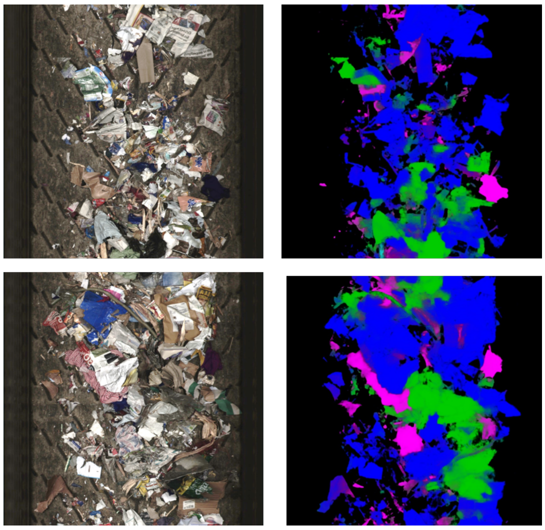

Fig. 3. Semantic segmentation results with plastics (green), wood (turquoise), textiles (purple), paper (blue).

V. Conclusion & Outlook

The project is currently in its first phase focusing on hardware design, interface definition, and data collections. The ground-truth generation strategy of using single refuse categories as starting point to synthesise realistic refuse mixtures proofed to be extremely valuable and brought a tremendous speed-up in necessary data generation, which can also have an impact on other similar deep learning applications. This data collection is the basis for all future work in the project including improvements of the camera and lighting setup, training of image recognition models, domain-specific adaption and improvements of the image recognition models, validation of the image recognition results, and all further analysis and optimizations.

Initial results of semantic segmentation and refuse classification already showed the feasibility of the approach, which will be further refined during the ongoing project and applied to other machinery on the refuse processing chain as well as to other sorting facilities in the future.

REFERENCES

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC Superpixels Compared to State-of-the-art Superpixel Methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 34, no. 11, pp. 2274–2282, 2012.

[2] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," in *Proceedings of the European Conference on Computer Vision 2018 (ECCV 2018)*, ser. Lecture Notes in Computer Science. Cham, Switzerland: Springer International Publishing, 2018, pp. 833–851.

[3] J. F. Deprez, "Hyperspectral Analysis For Precision Optical Sorting," 2017, Proceedings of the Conference on Hyperspectral Imaging in Industry (CHII 2017).

[4] S. P. Gundupalli, S. Hait, and A. Thakur, "A Review on Automated Sorting of Source-Separated Municipal Solid Waste for Recycling," *Waste Management*, vol. 60, pp. 56–74, 2017.

[5] H. Gursch, H. Ganster, A. Rinnhofer, G. Waltner, C. Payer, C. Oberwinkler, R. Meisenbichler, and R. Kern, "KI-Waste - Combining Image Recognition and Time Series Analysis in Refuse Sorting," in *Mensch und Computer 2021 - Workshopband*, C. Wienrich, P. Wintersberger, and B. Weyers, Eds. Bonn, Germany: Gesellschaft für Informatik e.V., 2021, pp. 1–4. [Online]. Available: https://dl.gi.de/handle/20.500.12116/37354

[6] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[8] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015)*. New York, USA: IEEE, June 2015, pp. 3431–3440.

[9] A. Minelgaitė and G. Liobikienė, "Waste Problem in European Union and its Influence on Waste management Behaviours," *Science of The Total Environment*, vol. 667, pp. 86–93, June 2019.

[10] J. Piao, Y. Chen, and H. Shin, "A New Deep Learning Based Multi-Spectral Image Fusion Method," *Entropy*, vol. 21, no. 6, pp. 1–16, June 2019.

[11] P. Prayagi, "Prism Based Multi-Sensor Technology for Multispectral Imaging Applications," 2017, Proceedings of the Conference on Hyperspectral Imaging in Industry (CHII 2017).

[12] A. Rinnhofer, "Combination of Multispectral and Multisensory Data," 2019, Fraunhofer Vision Technology Day, INNOVATIVE TECHNOLOGIES FOR INDUSTRIAL QUALITY ASSURANCE WITH IMAGE PROCESSING.

[13] M. Sackewitz, *Leitfaden zur hyperspektralen Bildverarbeitung*. Stuttgart, Germany: Fraunhofer Verlag, 2019.

[14] R. Sarc, A. Curtis, L. Kandlbauer, K. Khodier, K. E. Lorber, and R. Pomberger, "Digitalisation and Intelligent Robotics in Value Chain of Circular Economy Oriented Waste Management - A Review," *Waste Management*, vol. 95, pp. 476–492, July 2019.

[15] P. Wollmann, "Hyperspectral Imaging for Surface Inspections," 2017, Proceedings of the Conference on Hyperspectral Imaging in Industry (CHII 2017).