Multi-Modal Garment Sorting and Classification Combining Tactile and Visual Sensing

Serkan Ergun *
Institute of Smart Systems Technologies
University of Klagenfurt
Klagenfurt am Wörthersee, A 9020
serkan.ergun@aau.at

Tobias Mitterer
Institute of Smart Systems Technologies
University of Klagenfurt
Klagenfurt am Wörthersee, A 9020
tobias.mitterer@aau.at

Hubert Zangl†
Institute of Smart Systems Technologies
University of Klagenfurt
Klagenfurt am Wörthersee, A 9020
hubert.zangl@aau.at

Abstract

Automated garment handling in textile recycling remains challenging due to the deformability of textiles, their high shape variability, frequent self-occlusion, and the presence of foreign objects in cluttered heaps. This paper presents a Multi-Modal robotic sorting system that combines semantic visual perception with tactile grasp monitoring. The proposed approach integrates Visual Language Model (VLM) based garment classification, Convolutional Neural Network (CNN) based grasp prediction using RGB-D images, and capacitive tactile fingertips mounted on a parallel gripper to detect grasp success, object loss, and approximate weight during manipulation. The estimated weight serves as a plausibility measure for the visually predicted garment class and as a coarse indicator of garment size. To support safe execution, a Digital Twin implemented in MoveIt2 is used for motion planning and collision avoidance in a synchronized real and virtual environment. A classification accuracy of up to $87.89\,\%$ across six classes was achieved in an experimental robotic sorting scenario including 219 items. Furthermore, the tactile finger sensor is evaluated under wet conditions and in contact with wet textiles to assess robustness, showing reliable sensing behavior even in these challenging scenarios. Overall, the results demonstrate the potential of combining semantic vision and robust tactile sensing for dependable textile sorting in recycling applications.

1 Introduction

Robotic manipulation of deformable objects remains one of the central challenges in automation [1, 2]. Among deformable materials, garments are particularly difficult to handle due to their high variability in shape, frequent self-occlusion, and non-rigid dynamics, especially when presented as unordered heaps. These challenges are further amplified in textile recycling scenarios, where garments may be entangled and mixed with foreign objects such as plastic packaging or metallic accessories. At the same time, upcoming regulations such as the European Union's Digital Product Passport (DPP) for textiles aim to improve material traceability by 2027 [3, 4]. However, legacy garments without

*S.E. and T.M. contributed equally

†H.Z. is also affiliated with the Ubiquitous Sensing Lab, University of Klagenfurt, 9020 Klagenfurt, Austria

The Third Austrian Symposium on AI, Robotics, and Vision (AIROV26).

digital metadata will remain present in recycling streams, requiring perception-driven identification and manipulation.

Recent advances in multi-modal neural networks have introduced VLMs, which enable semantic queries on visual data by combining vision and language representations [5]. Such models allow flexible interpretation of garment attributes and categories from images. Prior work demonstrated the feasibility of integrating VLMs with CNNs for garment classification in robotic sorting scenarios [6]. Nevertheless, these approaches primarily focus on visual perception and do not sufficiently address the physical interaction challenges associated with grasping deformable textiles in cluttered environments.

Reliable manipulation of garments requires robust grasping strategies that account for the compliance and variability of textile materials. In particular, tactile sensing during grasping plays a crucial role in detecting slip events, monitoring grasp stability, and preventing object loss. Measuring normal and shear forces during manipulation enables the robot to adapt its grasp and maintain stable contact with deformable objects.

In this work, we present a robotic textile manipulation system that combines semantic perception with tactile grasping feedback. The system integrates VLMs and CNNs for garment classification while emphasizing robust physical interaction with textiles. The main contribution of this work is that, during grasping, force measurements from the gripper are used to detect shear forces that indicate object loss and to estimate the approximate weight of the grasped textile, thereby providing an additional plausibility check for the visually predicted garment class and enabling coarse inference of the object size. The system further incorporates a Digital Twin using MoveIt2 for motion planning, where reconstructed 3D representations of manipulated textiles are integrated into the planning environment. Particular emphasis is placed on the robustness of the tactile sensing frontend, which is designed to maintain reliable operation even when interacting with challenging materials such as wet textiles.

2 Related Work

Recent advances in robot grasping and classification of textiles show, that general VLMs models perform better on untrained objects like textiles, than specialized CNN networks. An example of a CNN is given by [7] and an example of a VLM is [8]. Multimodal Large Language Model (MLLM) models, like [9] by Alibaba Cloud are computation-heavy, with bigger versions requiring VRAM greater than 100GiB but also support better visual reasoning capabilities.

In robotics, once an object has been classified, suitable grasping positions need to be determined. Several approaches integrate grasp detection with additional perception tasks [10, 11]. In contrast, the previously discussed VLMs do not inherently provide segmentation or grasp position detection capabilities. Accurate grasp detection is particularly important for objects with complex geometries or deformable materials, such as clothing. In such scenarios, task-specific models like CNNs often achieve superior performance due to their specialized training for grasp prediction. Nevertheless, recent work has explored the use of VLMs that incorporate semantic input to guide grasp planning. For instance, models can interpret commands such as “the tip of the sock” and return corresponding grasp positions [12, 13]. VLMs also differ in that they are either focused on performance, or on usability on edge devices.

After identifying grasp positions, the next step is to grasp an object. During grasping it is important to be able to detect successful grasps by measuring, e.g., normal and shear forces applied on a grasped object [14, 15].

Our method combines VLMs with a CNN-based perception pipeline for semantic classification in a robotic textile-sorting scenario involving densely piled garments. In addition to visual recognition, the system emphasizes robust grasping of deformable objects. Force measurements from the gripper are used to monitor shear forces during manipulation, enabling detection of object loss during grasping. The measured forces also allow for weight estimation of the textile, which serves as a plausibility check for the predicted class and provides a rough indication of the object size. Furthermore, the gripper’s sensor frontend is designed to be robust against direct contact with challenging materials such as wet textiles, ensuring reliable sensing performance.

3 Experimental Setup

The experimental setup and its corresponding Digital Twin (RViz) is illustrated in Figure 1 below. In

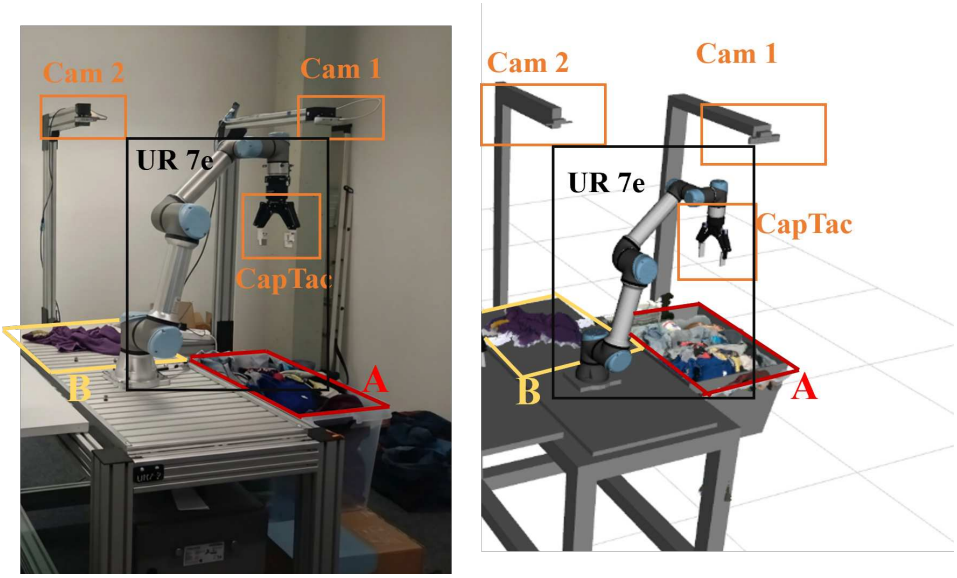


Figure 1: An overview of the experimental setup (left) and the corresponding Digital Twin in RViz (right).

this scenario, a UR7e robot is tasked with picking up unsorted, cluttered but clean garments from a basket (zone *A* in Figure 1) and to place them on a separate table for garment class inspection (zone *B* in Figure 1). Afterwards, the garment is picked up and distributed for further processing (outside the scope of this paper). Grasp locations for both zones are determined by using an adapted version of the grasp prediction algorithm developed by Ainetter et al. [10] and extensively tested in [6, 11, 16]. Two Intel Realsense depth cameras from the D400 model family (*Cam 1* and *Cam 2* in Figure 1) provide the necessary depth and color images. The RGB stream of *Cam 2* also provides the frames needed for garment type classification. The robot is equipped with a Robotiq 2F-140 gripper. The fingertips are replaced with CapTac [14] capacitive tactile sensors (a detailed view is shown in Figure 3a). These fingertips are used for two purposes: Determining if a garment was grasped from the basket (zone *A*) or dropped during manipulation and to measure its weight after garment type classification to provide further information on the garment. The detection threshold of CapTac is stated to be $20\,\mathrm{g}$ [13] corresponding to a weight force of $0.4\,\mathrm{N}$ for a two-fingered gripper. The necessary computing power for running the grasp prediction algorithm, robot control script as well as the Digital Twin is provided by a home grade PC equipped with an 11th Gen Intel® Core™ i7-11700KF @ 3.60GHz × 16 CPU with 64 GB of DDR4 RAM paired with a Graphics Processing Unit (GPU) Nvidia GeForce RTX 3060 with 12GiB VRAM. The VLMs are run on an external Nvidia H200X-141C cloud GPU with 144GiB VRAM.

The communication between all hardware modules is handled by ROS 2 Jazzy. A schematic overview showing the connection between all hardware and software components is shown in Figure 2. The real and simulated environment share the same coordinate frames allowing motion planning and static obstacle avoidance to be conducted on the Digital Twin via MoveIt2 before executing movements on the real robot. Furthermore, a combined RGB/depth stream of *Cam 1* is projected into the Digital Twin as a pointcloud, allowing remote observation. A segmented RGB/depth image of a garment in zone *B* can be projected as well, if needed.

The CapTac sensors are connected to the ROS 2 system via the Arrowhead Eclipse framework [17], as capacitive sensors, which provide normal and shear forces measured on multiple channels to the system. On the ROS 2 side, an Arrowhead Consumer is running, which takes the measured forces and publishes them into the ROS 2 system. The Arrowhead system is used for safe and secure transmission of sensor data.

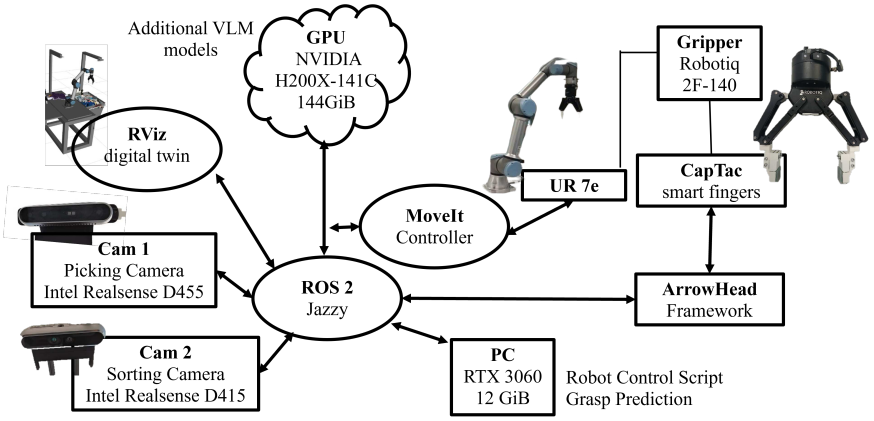
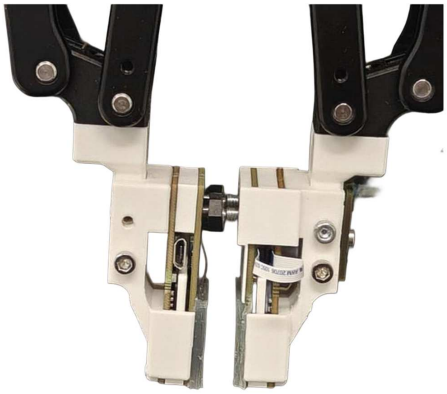


Figure 2: Schematic of the proposed framework, consisting of a UR7e robot equipped with a Robotiq 2F-140 gripper - and CapTac [14] fingertips - a desktop PC equipped with a Nvidia RTX 3060 budget consumer graphics card, one Nvidia H200X professional graphics card and two Intel Realsense cameras for grasp detection and object classification.



(a) Closed gripper, the gap in between the two gripper fingers is used to avoid touch between the sensors, when no object is grasped, so as to better detect an empty grasp. The distance can be manually adapted with the screws on top.

(b) Garment manipulation process, showcased on a sock 1-3, a shirt 4-6 and a trouser 7-9. The first image displays a textile being grasped from the box, the second image displays the textile in the air and the third image displays the textile being placed on the inspection table.

Figure 3: Detailed view of the gripper (a); Garment manipulation process (b)

4 Experimental Procedure

Upon starting the experimental procedure, a set of random clean garments is thrown randomly in a transparent basket alongside random foreign objects, such as bottles, cans and 3D printed objects from the EGAD training set [18]. The basket does not need to be perfectly aligned for every experimental run. A bounding box is automatically set around the edges of the basket to avoid grasping objects that are outside the basket.

After initial manual setup, the robot control script is started, which operates the robot as a finite-state machine (FSM). Figure 4 shows the flowchart for the states and transition conditions. In the *Init* state,

the robot is moved to a pose outside the field-of-view of both cameras and the CapTac sensors are initialized by recording baseline measurements. Next, the control script moves to *Find Garment*, calling the grasp prediction algorithm with a RGB and depth frame of *Cam 1*. If a potential grasp candidate is found within 5 tries, the program will switch to *Pick up Garment* and uses MoveIt2 within the Digital Twin to plan and validate a trajectory, otherwise the program will shut down. After the grasp pose is reached and the gripper closes, CapTac sensor readings are examined to check for an object between the fingers. In case of success, the sensor readings are monitored until the robot reaches zone *B* for garment classification. While an object is being lifted, a slight shaking motion is applied to the final three joints of the robot to shake off potential by-catch and avoid a drop in an undesired area. The garment is then pulled over the edge of the inspection table to enforce spreading out as good as possible. In case of an unsuccessful grasp or object loss, the robot moves outside the field-of-view of *Cam 1* and the program switches to *Find Garment*. Impressions of the garment handling procedure are shown in Figure 3 for exemplary objects of classes: sock, shirt and trousers.

Once the robot has dropped off the item and moved out of the field-of-view of *Cam 2*, the program switches to *Inspection*. The VLM running on the external GPU receives a RGB image from *Cam 2* and returns the predicted object class; the options being: trousers, shirt, underwear, sock, other (foreign object, or garment outside the aforementioned classes) and empty. We run two models from the Qwen3 family by Alibaba Cloud, which delivered promising results in a recently published benchmark [19], utilizing the locally run Python API of Ollama [20]. A minimal Python code snippet showcasing the usage of Ollama with Python3 is shown in listing 1. After classification, the grasp prediction algorithm is called with a RGB and depth image from *Cam 2*. The object is then lifted and an estimate of the weight is calculated using the CapTac sensors. In the event of multiple objects being grasped from the box, the VLM-powered classification is run again to ensure an empty inspection table. The combined information is then stored and the object is then placed aside for further processing. This information can be used to provide additional information on the object and conducting sanity checks. Some examples are: If an object of class shirt or trousers has a very low weight, it may be infant clothing. If an object of class sock is measured to have a rather high weight, the object class may have been predicted incorrectly, and the garment could be put aside for further manual inspection. The aforementioned garment segmentation scheme, using the same procedure as described in [19] could be used to create further training data for VLM based characterization of garments. Finally, the program will switch to *Reset* and prepare for picking up the next object from the basket.

Listing 1: Minimal Python code example for running a Vision Language Model with Ollama

```
import ollama

response = ollama.chat(
    model='model-name',
    messages=[{"role": "system",
        "content": "You are an expert garment classification
            device."},
        {'role': 'user', 'content': '"Do you spot a clothing item on
            the table? "
    "If yes: Classify them in the classes: "
    "shirt, sock, underwear or trousers. "
    "Do you see something else instead? respond with other. "
    "Is the table empty? respond with empty. "
    "Your response is a single word - either "
    "shirt, sock, underwear, trousers, other or empty"', 'images':
        [fullPathToImages]
            }]
)
```

In parallel, the robustness of CapTac against wet garments was investigated in manual experiments. Sensor readings were recorded for dry and wet garments and empty measurements with dry and manually wet sensor pads.

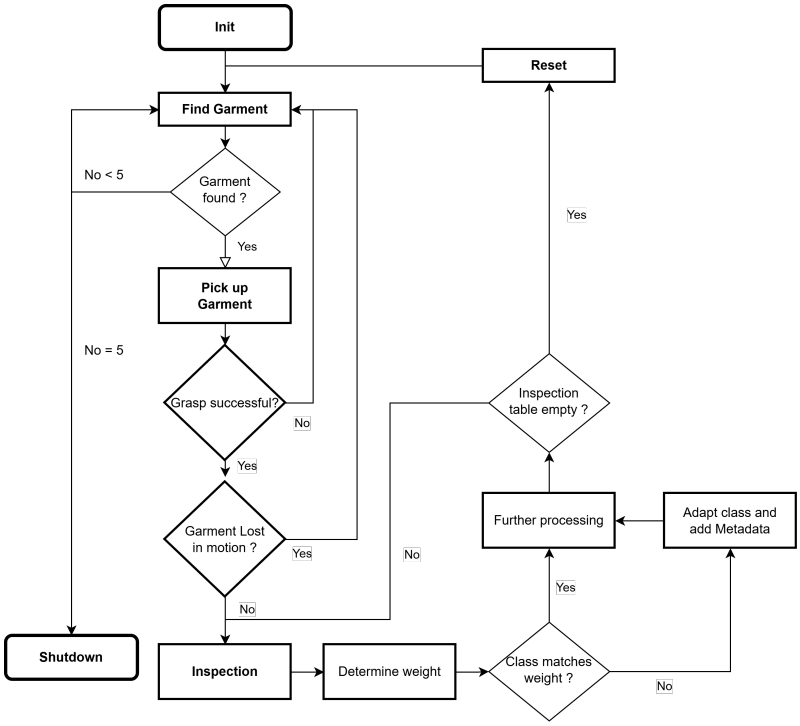


Figure 4: Flowchart of the textile grasping process.

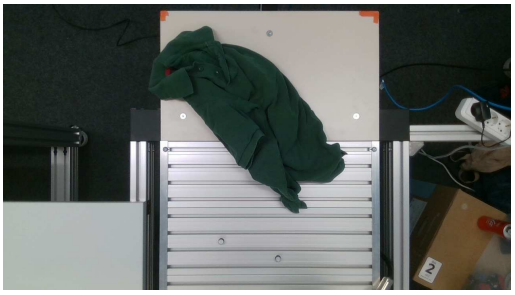
5 Results

5.1 Multi-Modal sensing experiment

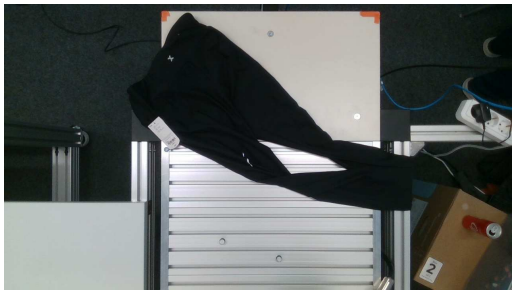
During the automated garment sorting scenario, a total of 219 items (garments and foreign objects) were grasped from the basket, classified on the inspection table and their weight estimated by CapTac. Ground truth data for these experiments were manually collected by human classification and manual weighing using a kitchen grade precision scale. To address robustness of the VLMs, images from objects on the inspection table were uncropped, and some distracting items were placed in the scenery. Two exemplary samples are shown in Figure 5. At two random instances during the experiments, objects were manually removed from the inspection table to check for robustness against hallucinations. Furthermore, two more empty samples were taken after the experiment was conducted. To avoid a misclassification, when two or more garments are present on the inspection table, the grasp prediction algorithm provides the location of the garment on the image. The combination of CapTac with a well established grasp prediction algorithm ensured that no unintentional empty scenes were recorded. The number of samples per object class is shown in Table 1 alongside the accuracy of both investigated models. The confusion matrices for both VLMs are shown in Figure 6. Furthermore, the computation time for each model was recorded and stored alongside the model predictions. The average processing times $\bar{t}$ alongside the 10^{th} percentile P_{10} and 90^{th} percentile P_{90} are shown in the bottom half of Table 1.

A prediction was considered correct if the model's response exactly matched the requested class (ignoring lower/uppercase lettering). Returning a different, but semantically correct class name was considered as wrong. Also, if the response also contained more words than requested, the response was treated as incorrect. While the larger Qwen model provides an accuracy of >97 % for shirts and full accuracy for socks, it lacks in detecting trousers and empty scenes, where it would typically return one of the distracting items nearby. As garments cannot be perfectly presented on a table using only a single robotic arm in reasonable time, the overall accuracy could be improved with adding multi-robot grasping.

Estimating the weight of the garments using CapTac faced a few difficulties. Garments with a very low weight (e.g. socks) do not surpass the reliable detection limit for shear force (0.2 N corresponding



(a) Class: "Shirt", Weight:264 g; " qwen3-vl:235b: "Shirt", qwen3-vl:8b: "Shirt", Estimated Weight: 225 g;



(b) Class: "Trousers", Weight: 257 g; qwen3-vl:235b: "Trousers", qwen3-vl:8b: "Trousers", Estimated Weight: 240 g;

Figure 5: Two representative images of garments from zone B. To assess robustness in cluttered environments, several smaller garments were scattered on the ground and additional distracting objects were introduced. Each model's output is shown together with the corresponding ground truth.

Table 1: Performance Benchmark for both investigated Qwen3 models and computation time metrics on a Nvidia H200 in s: average $\bar{t}$, 10^{th} percentile P_{10} and 90^{th} percentile P_{90}

	Overall	Shirt	Sock	Trousers	Underwear	Other	Empty
Image Count	223	38	64	43	12	65	4
Models				Accuracy			
qwen3-vl:235b	87.89 %	97.37 %	100.00 %	60.47 %	83.33 %	93.85 %	25.00 %
qwen3-vl:8b	83.86 %	86.84 %	93.75 %	55.81 %	66.67 %	95.38 %	50.00 %

	qwen3-vl:8b	qwen3-vl:235b
$\bar{t}$	1.595	2.444
P_{10}	0.993	1.739
P_{90}	2.550	3.072

to 20 g per finger - yielding a minimal garment weight of 40 g, assuming equal distribution of shear force), and therefore only the normal force information can be used to detect an object between the fingers. Larger garments, such as adult trousers or jackets, tend to entangle above the sensor pads, and thus produce less shear force. In any case, the weight of the garment will always be underestimated. CapTac require a manual one-time calibration using a tray with known weight that has full contact with both sensor pads and additional precision weights to determine a baseline. An exemplary calibration curve is shown in [14].

Thus, the authors rather propose a qualitative assessment of the shear force in weight classes, as listed in Table 2. Combining garment type and weight class allows to easier differentiate between adult and infant clothing. Garment classes with generally lower weight will then trigger an additional manual sanity check to verify if the VLM prediction is accurate.

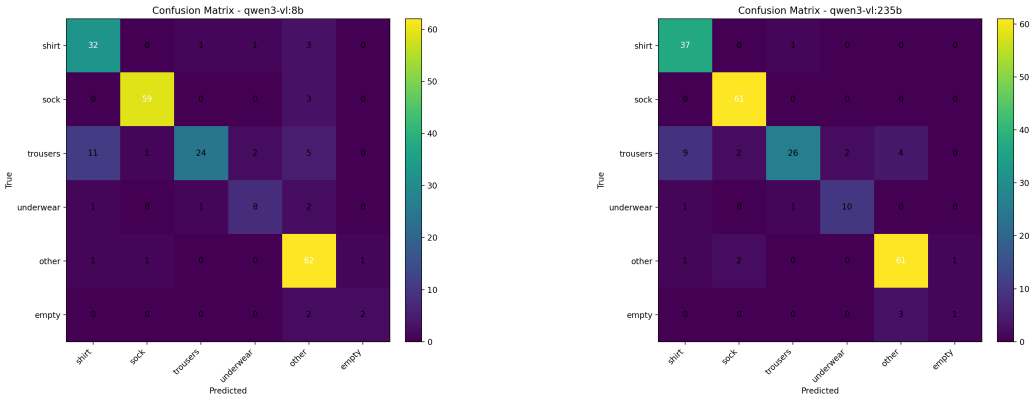
5.2 Robustness Against Liquid Contamination

In an additional set of experiments, the robustness of CapTac against liquid contamination on the sensor pads and wet garments was investigated. The silicone structure of CapTac (Ecoflex) is, by

Table 2: Qualitative assessment of garments using weight classes.

Weight	Shirt	Sock	Trousers	Underwear	Other
low(< 45 g)	toddler/inf.	single/pair	toddler/inf.	ok	ok
mid (45 g $<$150 g)	inf	pair	inf	ok	ok
high (150 g $<$400 g)	adult	s/c	adult	heavy underwear	ok
heavy (≥ 400 g)	s/c	s/c	adult	s/c	ok

Notes: inf.: infants, s/c: sanity check recommended



(a) qwen3-vl_8b

(b) qwen3-vl_235b

Figure 6: Confusion matrices for Qwen3 models: 8b parameters (left) and 235b parameters (right)

nature, highly hydrophobic, however potential effects on the capacitive sensing needed to be verified. During multiple measuring cycles, the sensor readings were recorded while the sensor pad was in a dry and wet state (by manually wetting the pad with a paper towel). No significant change in noise or drift was recorded during these experiments.

In its current state, the sensor (with its electronics) is not waterproof, and liquid contamination can only be mitigated on the sensor pads and the front face of the sensors.

6 Discussion, Conclusion and Outlook

This paper presented a Multi-Modal robotic sorting system that combines VLM based garment classification with a CNN based grasp prediction using RGB-D images, and capacitive tactile fingertips mounted on a parallel gripper to detect grasp success, object loss, and approximate weight during manipulation. The estimated weight serves as a plausibility measure for the visually predicted garment class and as a coarse indicator of garment size. A Digital Twin implemented in RViz uses MoveIt2 for motion planning and collision avoidance in a synchronized real and virtual environment.

A classification accuracy of up to $87.89\,\%$ across six classes was achieved in an experimental robotic sorting scenario including 219 items. Garment manipulation is handled by a single robotic arm, which does not allow optimal garment placement on the inspection table (zone `B`), practically reducing the accuracy for garment classes of larger sizes. Socks, due to their smaller size and distinctive shape, gave the highest accuracies. Furthermore, the tactile finger sensor is evaluated under wet conditions and in contact with wet textiles to assess robustness, showing reliable sensing behavior even in these challenging scenarios. Overall, the results demonstrate the potential of combining semantic vision and robust tactile sensing for dependable textile sorting in recycling applications.

The present state of the proposed experimental setup offers multiple options for future improvement. Adding a second manipulator for simultaneous two-arm manipulation to spread out garments can improve the classification accuracy and opens the door for visual fault inspection. A weighted combination with other VLMs might increase the accuracy as well and serve as a second source for classification estimates. The current state of the weight measurement can be improved by optimizing the gripper shape, to avoid tangling of garments.

Acknowledgments and Disclosure of Funding

This work has received funding from the ”Austrian Research Promotion Agency” (FFG) within the AdapTex project under grant number 899044 and by the European Commission, through the European H2020 research and innovation programme, KDT Joint Undertaking, and National Funding Authorities from 10 involved countries – including Hungary – under the research project Arrowhead fPVN with Grant Agreement no. 101111977.

References

[1] R. Herguedas, G. López-Nicolás, R. Aragüés, and C. Sagüés, "Survey on multi-robot manipulation of deformable objects," in *2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2019, pp. 977–984.

[2] A. Billard and D. Kragic, "Trends and challenges in robot manipulation," *Science*, vol. 364, no. 6446, p. eaat8414, 2019. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.aat8414

[3] European Parliamentary Research Service, "Digital product passport for the textile sector," 2024. [Online]. Available: {https://www.europarl.europa.eu/RegData/etudes/STUD/2024/757808/EPRS_STU(2024)757808_EN.pdf}

[4] Directorate-General for Environment, "COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS - EU Strategy for Sustainable and Circular Textiles," 2022. [Online]. Available: {https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52022DC0141}

[5] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning transferable visual models from natural language supervision," in *Proceedings of the 38th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, M. Meila and T. Zhang, Eds., vol. 139. PMLR, 2021, pp. 8748–8763. [Online]. Available: https://proceedings.mlr.press/v139/radford21a.html

[6] S. Ergun, T. Mitterer, and H. Zangl, "Towards automated handling and sorting of garments combining visual language models and convolutional neural networks," *Proceedings of the Austrian Robotics Workshop 2025*, pp. 25–30, 2025. [Online]. Available: https://www.fh-salzburg.ac.at/fileadmin/fhs_daten/departments/information-technologies/documents/ARW2025_Proceedings_final_kl.pdf

[7] M. Suchi, T. Patten, D. Fischinger, and M. Vincze, "Easylabel: a semi-automatic pixel-wise object annotation tool for creating robotic rgb-d datasets," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 6678–6684.

[8] Minderer, Matthias and Gritsenko, Alexey and Stone, Austin and Neumann, Maxim and Weissenborn, Dirk and Dosovitskiy, Alexey and Mahendran, Aravindh and Arnab, Anurag and Dehghani, Mostafa and Shen, Zhuoran and Wang, Xiao and Zhai, Xiaohua and Kipf, Thomas and Houlsby, Neil, "Simple Open-Vocabulary Object Detection," in *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part X*. Berlin, Heidelberg: Springer-Verlag, 2022, p. 728–755. [Online]. Available: https://doi.org/10.1007/978-3-031-20080-9_42

[9] S. Bai, Y. Cai, R. Chen, K. Chen, X. Chen, Z. Cheng, L. Deng, W. Ding, C. Gao, C. Ge, W. Ge, Z. Guo, Q. Huang, J. Huang, F. Huang, B. Hui, S. Jiang, Z. Li, M. Li, M. Li, K. Li, Z. Lin, J. Lin, X. Liu, J. Liu, C. Liu, Y. Liu, D. Liu, S. Liu, D. Lu, R. Luo, C. Lv, R. Men, L. Meng, X. Ren, X. Ren, S. Song, Y. Sun, J. Tang, J. Tu, J. Wan, P. Wang, P. Wang, Q. Wang, Y. Wang, T. Xie, Y. Xu, H. Xu, J. Xu, Z. Yang, M. Yang, J. Yang, A. Yang, B. Yu, F. Zhang, H. Zhang, X. Zhang, B. Zheng, H. Zhong, J. Zhou, F. Zhou, J. Zhou, Y. Zhu, and K. Zhu, "Qwen3-vl technical report," *arXiv preprint arXiv:2511.21631*, 2025.

[10] S. Ainetter and F. Fraundorfer, "End-to-end trainable deep neural network for robotic grasp detection and semantic segmentation from rgb," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 13 452–13 458.

[11] S. Ergun, T. Mitterer, S. Khan, N. Anandan, R. B. Mishra, J. Kosel, and H. Zangl, "Wireless capacitive tactile sensor arrays for sensitive/delicate robot grasping," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023, pp. 10 777–10 784.

[12] META, "segment anything website," https://segment-anything.com/, accessed: 2025-06-04.

[13] N. Carion, L. Gustafson, Y.-T. Hu, S. Debnath, R. Hu, D. Suris, C. Ryali, K. V. Alwala, H. Khedr, A. Huang, J. Lei, T. Ma, B. Guo, A. Kalla, M. Marks, J. Greer, M. Wang, P. Sun, R. Rädle, T. Afouras, E. Mavroudi, K. Xu, T.-H. Wu, Y. Zhou, L. Momeni, R. Hazra, S. Ding, S. Vaze, F. Porcher, F. Li, S. Li, A. Kamath, H. K. Cheng, P. Dollár, N. Ravi, K. Saenko, P. Zhang, and C. Feichtenhofer, "Sam 3: Segment anything with concepts," 2025. [Online]. Available: https://arxiv.org/abs/2511.16719

[14] S. Ergun, R. B. Mishra, N. Ananadan, T. Mitterer, V. Mattoli, and H. Zangl, "Captac: Robust capacitive sensing for distributed force mapping in parallel robotic grasping," *IEEE Robotics and Automation Letters*, vol. 11, no. 3, pp. 3668–3675, 2026.

[15] X. Zhang, T. Yang, D. Zhang, and N. F. Lepora, "Tacpalm: A soft gripper with a biomimetic optical tactile palm for stable precise grasping," *IEEE Sens. J.*, vol. 24, no. 22, pp. 38 402–38 416, 2024.

[16] S. Ergun, T. Mitterer, and H. Zangl, "A hybrid approach towards automated textile sorting," *e+i Elektrotechnik und Informationstechnik*, vol. 142, no. 6, pp. 360–370, Nov 2025. [Online]. Available: https://doi.org/10.1007/s00502-025-01340-2

[17] P. Varga, F. Blomstedt, L. L. Ferreira, J. Eliasson, M. Johansson, J. Delsing, and I. Martnez de Soria, "Making system of systems interoperable the core components of the arrowhead framework," *J. Netw. Comput. Appl.*, vol. 81, no. C, p. 85–95, Mar. 2017.

[18] D. Morrison, P. Corke, and J. Leitner, "Egad! an evolved grasping analysis dataset for diversity and reproducibility in robotic manipulation," *IEEE Robot. Autom. Lett.*, vol. 5, no. 3, pp. 4368–4375, 2020.

[19] S. Ergun, T. Mitterer, and H. Zangl, "Digital twin driven textile classification and foreign object recognition in automated sorting systems," 2026. [Online]. Available: https://arxiv.org/abs/2603.05230

[20] Ollama, "Ollama website," https://ollama.com/, accessed: 2026-02-02.