Cobotic Induction System for Garment Picking

Fabio Pini[a,*], Olaf Kähler[b], Martina Uray[b], Eugeni Llagostera Saltor[c], and Luigi Biagiotti[a]

(a) "Enzo Ferrari" Engineering Department, University of Modena and Reggio Emilia, Modena, IT
(b) Digital – Institut für Digitale Technologien, Joanneum Research, Graz, AU
(c) Mobile Biometrics SL, Barcelona, SP

Abstract

The growing trend of garment waste highlights the need for effective and affordable processes to enable a more sustainable lifecycle. Collaborative robotic systems that support operators represent a feasible solution, but some limitations remain related to manipulating deformable objects, such as garments. By integrating a versatile gripper system and machine learning models trained to detect each garment, this work suggests a solution to automate garment picking. The first experiments highlight interesting results about effective grasping and promise further productivity improvements.

1 Introduction

The trend in the clothing market points out that global textile production doubled since 2000, with a further expected increase of 63% by 2030, corresponding to 102 million tonnes. Currently, 5.8 million tonnes of textiles are landfilled or incinerated every year, with only 15% of used textiles undergoing recycling pathways [1]. New circular business models and virtuous practices such as reuse, repair, and recycling are increasing in importance to make the fashion industry more sustainable [2]. These processes primarily rely on manual operations that involve humans in basic actions like garment sorting. Manual operations impose limitations with the constant increase in volumes and robotic systems therefore seem like an attractive alternative. Many research activities focus on developing effective robotic solutions to grasp garments from randomly arranged volumes of textile items. Nevertheless, open challenges remain in (i) the ability to grasp garments with different packaging types, i.e., unpackaged or packaged with undamaged/damaged polybags; (ii) recognition of robust grasping locations suitable for a wide range of garment categories; and (iii) picking just one garment at a time, avoiding the grasping of multiple items. To address these challenges, we present a picking approach based on collaborative robots, advanced grippers, and computer vision machine learning models to support the sustainability of the garment lifecycle.

1.1 Handling of deformable objects

Picking actions with robots, i.e. pick-and-place operations, are a standard problem in robotics and have received extensive attention. However, robotic grasping of garments still remains challenging [14] due to their deformable nature, self-occlusion, and the huge variability in fabric types with very different characteristics. In many applications, from production to shipping, garments can be individually wrapped in plastic bags, and suction grippers can efficiently manipulate them. Two interesting examples related to existing applications are proposed by Covariant [4] and Kindred [5], that respectively suggest the use of adaptive single suction cup grippers or vacuum integrated with parallel grippers. However, suction grippers are no longer applicable for returned goods, where the plastic bags may be damaged or missing. In this scenario it is neither clear, what kind of gripper is

[*]Corresponding author - fabio.pini@unimore.it

suited to manage the different packing conditions of the garments, nor how to identify the optimal grasping point. Recent works focused on grasping methods to manipulate garments fabrics, but specifically on starting planar configuration [12, 6]. Further challenges arise if a robot is tasked with picking up and moving clothes as efficiently as possible [3], or if the garments have to be singulated [11], as required for individually inspecting and returning items to e-commerce cycles or recycling and reuse pathways. Similar to DexNet for rigid object manipulation [7, 8], most recent works on robotic manipulation of clothes aim to directly predict the manipulation actions from sensory input using deep learning. Acquiring the training data to develop such methods is a challenge, however, and in certain situations, simulations [10] or virtual reality environments [9] have therefore been used. It emerges that an integrated robotic solution combining gripper features and garment recognition to support the picking is still missing.

1.2 Method outline

To address the aforementioned challenges, we suggest an integrated approach considering both hardware and software. We developed a dedicated platform to implement and assess the proposed approach, and we also use this platform to generate the training data required for continuous improvement of the system. Therefore, the focus is not only on defining an initial solution but also on obtaining real data with a simple, yet efficient fallback approach. The architecture of the suggested approach is described in Section 2. The proposed traditional multi-stage approach already yields excellent results, as we will evaluate in Section 3. Nevertheless, it clearly has some limitations. As a longer-term goal, direct prediction of grasps from recorded images is planned. Thus, the current system can be used to bootstrap the training process, as we will discuss in Section 4.

2 Method Overview

Our picking approach is depicted in Fig. 1. Thanks to a collaborative environment, supervision activities by the human operator are easily implemented through the picking operations These activities are related to restoring of the system after unexpected conditions as well as the management of further data to train the detection model on new sets of garments. More in detail, the latter is related to collection of new images, bounding boxing and labeling to detect new clothes. A dedicated area to collect the garments serves as the source to feed the picking process. Simple boxes (#1) or trolleys are common solutions for collecting the items. A RGB-D Camera (#2) allows collecting digital data of the visible garments, such as 2D pictures and 3D point clouds. A collaborative robot, Cobot (#3), enables safe continuous movements without safety fences. Alternatively, an industrial robot with a dedicated safety system can be used, but this robot must stop if the operator moves inside the robot's working zone. The final important hardware is the gripping system (#4), which allows reliable grasping of clothes with different fabrics and packages.

The software framework that drives the grasping operations follows a traditional approach with a separate object detection module for identifying garments. Once the detection is complete, a grasp point extraction logic determines the position for gripping the garments. Details of the developed system are provided in the subsequent subsections: Hardware and the setup of the prototypical

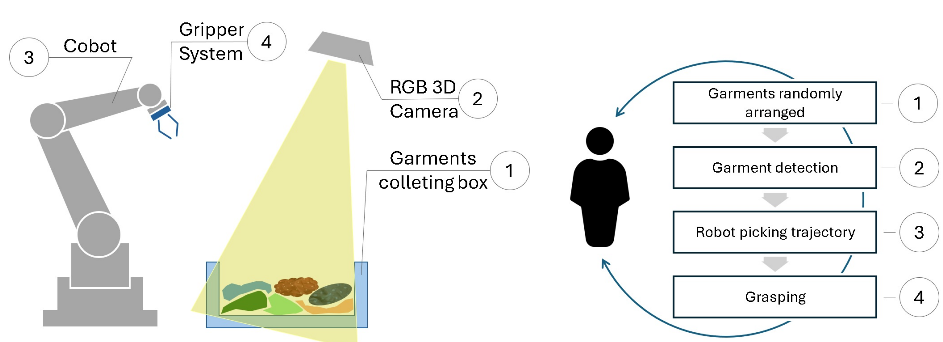

Figure 1: Main hardware components of the cobotic picking system and the main actions involved.

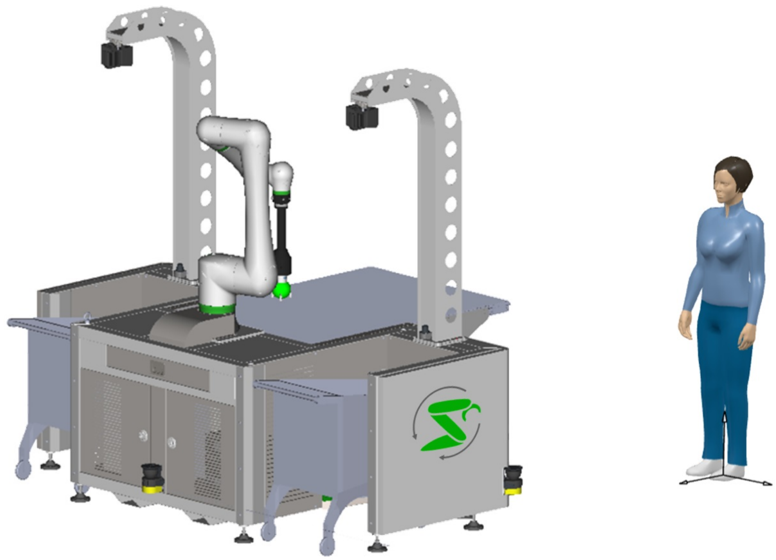

Figure 2: Cobosort system - The prototypical platform for garment picking

platform in Section 2.1; garment detection model in Section 2.2; and grasp point extraction algorithms in Section 2.3.

2.1 System description

The hardware in our system consists of a centrally installed Fanuc cobot, the CRX-25iA arm, which is used to grasp garments and drop them off on a conveyor belt for further processing, such as sorting operations. On either side of the cobot, there is a bay for a cart used to feed the garments into the system in a more or less unsorted manner. The prototype solution has two bays to avoid idle times, allowing the cobot to start processing one bay while the operator removes the empty cart from the second bay and brings in a new, full cart. Above each of the bays is a Zivid One Plus depth camera that creates RGB and depth images of the carts and their contents. Both cameras are connected to a central processing unit that analyzes the image data, computes a grasp point, and controls the robot movements. The whole unit is additionally equipped with three safety laser scanners at floor level that identify people or obstacles near the unit and slow down the cobot movements to collaborative speed (250mm/s) to ensure a safe working environment. The picking action is enabled by a dedicated grasping system that integrates a commercial self-centering gripper by OnRobot with dedicated sensors to return data to assess the grasping success rates. Figure 2 depicts the layout of the prototype platform integrating the main aforementioned modules, namely the CoboSort system.

2.2 Garment detection

The garments are detected using a standard object detection approach based on YOLO classification algorithm, specifically the YOLO v5 - XLarge (YOLOv5XL) model structure [13]. Given the RGB images as input, bounding boxes are predicted around the identified garments. For training, a continually growing dataset of images has been used. Table 1 collects the total number of labeled garments used through the preliminary training sessions. *Unpacked garments* is the dataset related to clothes without polybags arranged on planes, while *Organized box* and *recorded box* respectively are datasets referring to clothes inside collecting boxes, in all the three packaging conditions. *Organized trolley* and *Unorganized trolley* are the two configurations that recreate the working condition assumed, where the garments are respectively well-ordered and randomly arranged inside the trolleys. Finally, a first refinement round is performed with e new batch that is a mixture of un/organized garments, named as *Updated trolley*.

Through multiple experiments, the final key training parameters adopted are collected in Tab. 2. Here, *epochs* is the total number of training epochs, assuming that a single epoch represents a full pass

Table 1: Datasets used for preliminary model training

Configuration	Dataset specifications		
	Total	Training	Validation
Unpacked garments	140	112	28
Organized box	500	400	100
Recorded box	38	30	8
Organized trolley	188	151	37
Unorganized trolley	212	170	42
Updated trolley	400	300	100

over the entire dataset. Parameter *batch_size* refers to number of images processed before updating the internal parameters of the model. The *single_cls* defines how to treat the classes in the dataset. The learning rate scheduler selected to manage the convergence of the model is defined by *lr_decay*. Number of epochs to wait without improvement in validation metrics before early stopping the training is managed by *patience*. Finally, *freeze* defines the number of layers to keep to reduce the number of trainable parameters. The table provides the data related to experiment 14, that converged after 45 epochs. The inference was performed on (*i*) a video where the model is neither trained nor validated and (*ii*) a video on which the model is both trained and validated.

Table 2: YOLO model key training settings

Argument	*epochs*	*batch_size*	*single_cls*	*lr_decay*	*patience*	*freeze*
Value	45	16	true	cosine	10	1

Figure 3, on the left side, depicts the two classes defined to train the detector, *no_plastic* and *plastic*, respectively identified by yellow and magenta backgrounds. Nevertheless, in our experiments using a single class (*single_cls* set to true) resulted in better performance. On the right side, the detected objects are shown in a real picture returned by the Zivid camera in the final configuration.

2.3 Grasp Point Extraction

The recorded depth images are aligned with the RGB images and we can directly identify 3D points within the detected 2D bounding boxes. The strategy adopted first sort the bounding boxes by the average height of 3D points contained within. Then, we attempt to extract a grasp point from the topmost box, and in case of failure, revert to the next higher box until a suitable point has been found. For each box, we apply a series of filters: a) to remove invalid points for which no depth could be determined, b) to eliminate points that are close to the edges of the bounding box, c) to eliminate

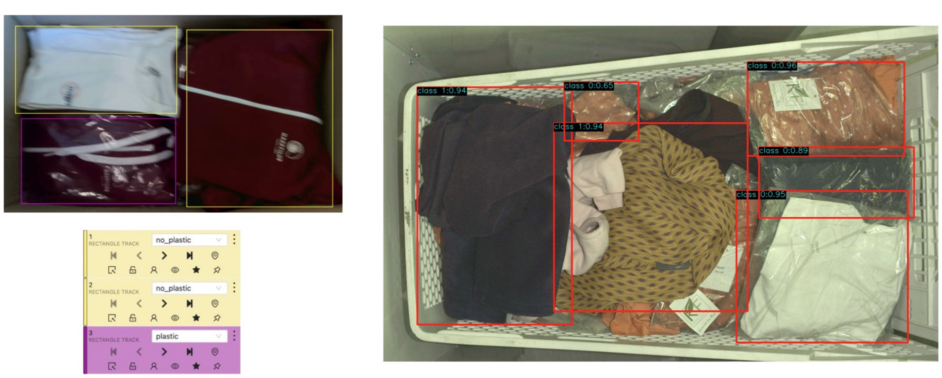

Figure 3: Labeling process for garment pictures with the two classes defined, left side. On the right, trained model applied to a real collection of garments to be picked up from a cart.

potential outliers at the topmost and bottommost percentiles of 3D points, d) to ensure a minimum distance from the edges of other detected bounding boxes, e) to ensure a minimum distance from the most recent grasp point if that grasp failed, and f) to avoid collisions with the empty cart. Finally, we select a point on the 3D surface that is closest to the center of gravity of the remaining points.

3 Experimental validation and evaluation

Experimental tests have been performed on the CoboSort system. The carts have been randomly filled with garments in three different conditions, i.e. unpacked and packed with or without polybags. Initial tests highlight that the most common source of double grasping is the detection model. Figure 4 depicts two examples related to wrong detections.
To improve the accuracy of the model, such as to detect a new set of clothes of different kinds or with novel fabrics and colors, recorded images are continuously stored. These images are then sent for annotation and added to the training process to ensure a wide variety and a representative dataset for practical applications.
Further additional 800 annotated pieces of clothes has been added in the training. Even if the accuracy of the model was very high, this improvement was requested to manage corner cases, such as two similar clothes placed side by side, as depicted in Fig. 4. As a result, the performances of last training with YOLOv5XL after 45 epochs for *plastic* and *no_plastic* model are presented through the quality metrics collected on Tab. 3. *Precision* measures the accuracy of the predictions, so it is the ratio of true positives (*TP*) out of all positive predictions, that is the sum of true and false positive (*TP+FP*). *Recall* measures how good all the positives are identified, so it it is the ratio of all *TP* out of the number of positive instances, that is calculated as the sum of true positives and false negative results (*TP+FN*). Mean Average Precision (*mAP*) is related to the mean of the area under *Precision/Recall* curve. It represents a trade-off between *Precision* and *Recall* where the *FP* and *FN* are evaluated through Intersection over Union (*IoU*) threshold. So, *mAP@0.5* and *mAP@0.5:0.95* relates to a *Precision/Recall* curve for a fixed *IoU=0.5* and *IoU* ranges from to 0.5 to 0.95, respectively.

Table 3: Performances values of last training with YOLOv5XL

Precision	*Recall*	*mAP@0.5*	*mAP@0.5:0.95*
0.9816	0.98256	0.97925	0.85183

A stress test of the system on 10 trials for a total of 650 items returned 7 instances of double grasping, with a success rate of 99%. The average time to complete the picking operation is 13 seconds. Further analysis indicates that the operating speed is mostly determined by the picking trajectory, which is currently composed of three main points: The cobot home position or the waiting point, the grasping point, and the release point. Furthermore, two waypoints on the robot cart are also inserted to avoid obstacles while the cobot is moving the gripper inside the cart volume, as shown in Fig. 5.

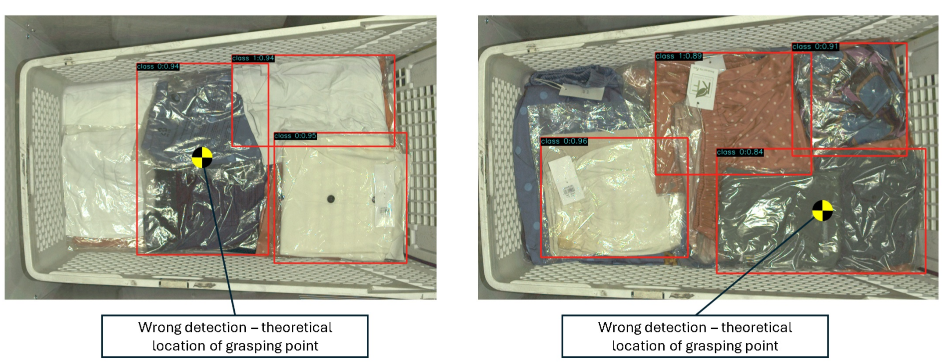

Figure 4: Two examples related to wrong detections. Grasping of multiple items is returned due to the picking targets returned.

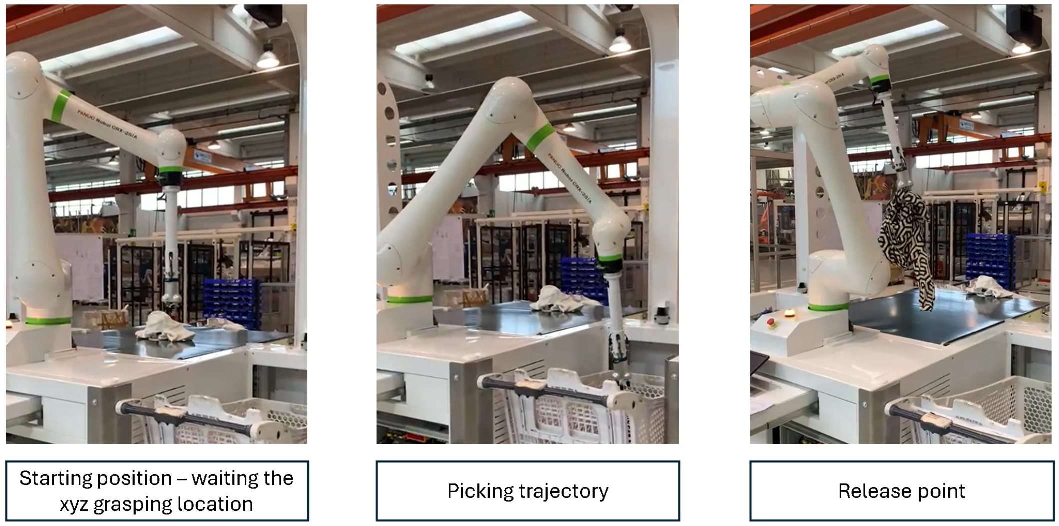

Figure 5: Picking trajectory; the three main points that are involved in the cobot movements

4 Conclusions and Future Plans

We proposed a system for picking and isolating garments from a box or cart. It is designed to work together with human operators in a safe manner, thus lowering the barrier of adopting the system into exiting processes. The designed vision system relies on RGB-D data, uses machine learning for identifying garments and a grasp point extraction logic working on top of the detections. In our evaluation, this simple system already achieves outstanding results with only about 1% of missed or double grasps. This performance level opens the way to further optimizations to integrate the system into existing workflows, and represents a step forward toward a change of fashion industry processes towards more sustainable circular business models.

The grasping still faces problems in some rare cases and we plan to address these by directly predicting grasp positions for the robot from the recorded image data. To achieve this, we need to collect annotated training data and establish a basic dataset. In existing works this is often done using simulation environments, but it is notoriously difficult to realistically simulate fabrics. Therefore, we also view our current approach as a method to automatically record training data and bootstrap a new training process. Once a significant number of more or less successful grasps have been observed in real scenarios, we expect that direct regression models for predicting grasp poses will work better and take over. Continuing the collection of samples on the fly will enable real-time performance metrics and continual improvements of the deployed model, as well as adaptation to unseen and unanticipated scenarios that were not covered in the training data.

Acknowledgments and Disclosure of Funding

This work resulted from the research activities by CoboSort, a cofunded project by EIT Manufacturing, an initiative of the European Institute of Innovation and Technology (EIT), a body of the European Union - Grant ID 23460. More information about the project can be found at: https://www.cobosort.unimore.it/.

References

[1] EU strategy for sustainable and circular textiles. https://environment.ec.europa.eu/document/download/74126c90-5cbf-46d0-ab6b-60878644b395_en?filename=COM_2022_141_1_EN_ACT_part1_v8.pdf, 2022. [Online; accessed 11-February-2024].

[2] The refashion circular design strategy — changing the way we design and manufacture clothes. *Design Studies*, 88:101205, 2023. ISSN 0142-694X. doi: https://doi.org/10.1016/j.destud.2023.101205.

[3] Aviv Adler, Ayah Ahmad, Shengyin Wang, Wisdom C Agboh, Edith Llontop, Tianshuang Qiu, Jeffrey Ichnowski, Mehmet Dogar, Thomas Kollar, Richard Cheng, et al. The teenager's problem: Efficient garment decluttering with grasp optimization. *arXiv preprint arXiv:2310.16951*, 2023.

[4] Covariant. Covartiant website. https://covariant.ai/insights/how-ai-powered-robots-solve-apparel-picking-challenges/. [Online; accessed on March-2024].

[5] Kindred. Kindred website. https://www.kindred.ai/sort. [Online; accessed on March-2024].

[6] Yuqi Liu, Junqiang Su, Xinyu Li, and Guoqing Jin. A systematic automated grasping approach for automatic manipulation of fabric with soft robot grippers. *Industrial Robot*, 50(4):623 – 632, 2023. doi: https://doi.org/10.1108/IR-07-2022-0173.

[7] Jeffrey Mahler, Florian T Pokorny, Brian Hou, Melrose Roderick, Michael Laskey, Mathieu Aubry, Kai Kohlhoff, Torsten Kröger, James Kuffner, and Ken Goldberg. Dex-net 1.0: A cloud-based network of 3d objects for robust grasp planning using a multi-armed bandit model with correlated rewards. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 1957–1964. IEEE, 2016. doi: https://doi.org/10.1109/ICRA.2016.7487342.

[8] Jeffrey Mahler, Matthew Matl, Vishal Satish, Michael Danielczuk, Bill DeRose, Stephen McKinley, and Ken Goldberg. Learning ambidextrous robot grasping policies. *Science Robotics*, 4(26):eaau4984, 2019. doi: https://doi.org/10.1126/scirobotics.aau4984.

[9] Marco Moletta, Maciej K Wozniak, Michael C Welle, and Danica Kragic. A virtual reality framework for human-robot collaboration in cloth folding. *arXiv e-prints*, pages arXiv–2305, 2023.

[10] Daniel Seita, Aditya Ganapathi, Ryan Hoque, Minho Hwang, Edward Cen, Ajay Kumar Tanwani, Ashwin Balakrishna, Brijen Thananjeyan, Jeffrey Ichnowski, Nawid Jamali, et al. Deep imitation learning of sequential fabric smoothing from an algorithmic supervisor. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9651–9658, 2020. doi: https://doi.org/10.1109/IROS45743.2020.9341608.

[11] Sashank Tirumala, Thomas Weng, Daniel Seita, Oliver Kroemer, Zeynep Temel, and David Held. Learning to singulate layers of cloth using tactile feedback. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7773–7780. IEEE, 2022. doi: https://doi.org/10.1109/IROS47612.2022.9981341.

[12] Jianping Wang, Jinzhu Shen, Xiaofeng Yao, and Fan Zhang. Research progress of automatic grasping methods for garment fabrics. *International Journal of Clothing Science and Technology*, 35(6):997 – 1022, 2023. doi: https://doi.org/10.1108/IJCST-05-2023-0068.

[13] Chang Yeong-Hwa and Zhang Ya-Ying. Deep learning for clothing style recognition using yolov5. *Micromachines*, 13(10):1678, 2022. doi: https://doi.org/10.3390/mi13101678.

[14] Jihong Zhu, Andrea Cherubini, Claire Dune, David Navarro-Alarcon, Farshid Alambeigi, Dmitry Berenson, Fanny Ficuciello, Kensuke Harada, Jens Kober, Xiang Li, Jia Pan, Wenzhen Yuan, and Michael Gienger. Challenges and outlook in robotic manipulation of deformable objects. *IEEE Robotics and Automation Magazine*, 29(3):67–77, 2022. doi: https://doi.org/10.1109/MRA.2022.3147415.