Human Factors in Robots, Drones and Unmanned Systems, Vol. 93, 2023, 29–36
https://doi.org/10.54941/ahfe1003744

An Interactive Learning Framework for Item Ownership Relationship in Service Robots

Yuanda Hu, Yate Ge, Tianyue Yang, and Xiaohua Sun

College of Design and Innovation, Tongji University, Shanghai, China

ABSTRACT

Autonomous agents, including service robots, require adherence to moral values, legal regulations, and social norms to interact effectively with humans. A vital aspect of this is the acquisition of ownership relationships between humans and their carrying items, which leads to practical benefits and a deeper understanding of human social norms. The proposed framework enables the robots to learn item ownership relationships autonomously or through user interaction. The autonomous learning component is based on Human-Object Interaction (HOI) detection, through which the robot acquires knowledge of item ownership by recognizing correlations between human-object interactions. The interactive learning component allows for natural interaction between users and the robot, enabling users to demonstrate item ownership by presenting items to the robot. The learning process has been divided into four stages to address the challenges posed by changing item ownership in real-world scenarios. While many aspects of ownership relationship learning remain unexplored, this research aims to explore and design general approaches to item ownership learning in service robots concerning their applicability and robustness. In future work, we will evaluate the performance of the proposed framework through a case study.

Keywords: Service robots, Social norm, Item ownership relationship, Human-robot-interaction, Human-object interaction

INTRODUCTION

In recent years, the use of autonomous agents has become widespread across various applications, including domestic robots, assistive robots, and self-driving cars. For these agents to effectively interact with humans, it is crucial for them to not only complete their specific tasks but also to understand and abide by moral values, legal regulations, and social norms to achieve human acceptability (Guo et al., 2020). An integral aspect of such competency is acquiring the ownership relationship between humans and their carrying items. The capability of a robot to learn and abide by ownership norms not only leads to practical benefits in environments containing owned objects but also provides a deeper understanding of human social norms and their navigation (Tan et al., 2019).

Item ownership relationship learning refers to the process by which a service robot learns and understands the relationship between an item and its owner. This involves detecting correlations between human-object interactions and recognizing patterns that indicate the ownership of an item. The goal is to enable the robot to understand who owns which items and make decisions based on that knowledge, such as avoiding interactions with personal items or correctly returning an item to its owner. Since the relationship between users and items may change at any time, pre-programming is impossible. Service robots need to acquire and revise the knowledge of the ownership relationship in an adaptive way. Recently, attempts have been made to develop service robots capable of learning and inferring ownership relationships and norms (Wu et al., 2020) (Tan et al., 2019). However, the existing approaches need more robustness due to the integration of multiple models or an inability to handle the complexities of changing item ownership in real-world environments. To overcome the limitations of previous work, we present an interactive learning framework that enables service robots to interact with users to resolve ambiguities and refine learning ownership relationships. This approach allows the robot to ask for confirmation and modification of item attribution, thereby enhancing its understanding of human-object relationships and improving its ability to perform real-world tasks.

This work presents a novel approach for enabling service robots to learn object attribution using computer vision. Specifically, we aim to achieve autonomous learning of item ownership by detecting human-object interaction behaviors and verifying human-object interaction rules. To mitigate the potential integration of multiple models, which may lead to increased system complexity and decreased robustness, we propose the Autonomous Ownership Relationship Learning Framework based on human-object interaction (HOI) detection. Furthermore, we introduce the Interactive Item Ownership Relationship Learning Framework to handle the complexities of changing item ownership in real-world environments and the possibility of misidentification. This framework leverages the Autonomous Ownership Relationship Learning Framework. It enables users to teach the robot new item ownership relationships by simply displaying their items in front of the robot. The proposed framework categorizes the learning process into four stages: initial registration, observation, interactive learning, and deployment, which enables the service robot to learn item attribution autonomously and interactively through interaction with a human teacher.

Autonomous Ownership Relationship Learning Based on Human-Object Interaction

Autonomous ownership relationship learning has been a topic of ongoing research for a while, with several methods available for combining ownership and users. (Tan et al., 2019) proposed an incremental norm learning algorithm to learn and infer ownership relations and norms. The robot is equipped to comprehend its surroundings' ownership context and behave in conformity with the norms. (Wu et al., 2020) integrate multiple models for detecting human-object interaction, recognizing the instance of objects, and identifying the users, enabling autonomous learning of item ownership relationships. Inspired by the methods previously discussed, we propose a method for autonomous ownership relationship learning based on the human-object interaction detection model. This approach efficiently learns ownership relationships by detecting specific correlations between human-object interactions.

Human-Object Interaction (HOI) detection aims to identify the relationships and locations of humans and objects within an image. The process involves analyzing the input image and predicting the interactions between people and the objects around them, which is accomplished by predicting the triplet <human, predicate, object> (Antoun & Asmar, 2023). HOI detection can be classified into two main categories: two-stage and one-stage. The two-stage approach first employs an object detection model to locate the human and object pair and then utilizes a neural network to extract the features for predicting the interaction (Gao et al., 2018) (Bansal et al., 2020) (Gao et al., 2020). The one-stage methods extract features and detect HOI triplets simultaneously by localizing the interaction using an interaction point or an anchor box, then predicting the action and the interacting pair bounding boxes (Liao et al., 2020) (Kim et al., 2020) (Wang et al., 2020). Figure 1 shows the visualization results of HOI detection (see Figure 1).

In ownership relationship learning, utilizing HOI detection to identify the interaction between people and objects eliminates the need to incorporate multiple models, such as object detection, action recognition, and person identification. This reduces system complexity by avoiding the intricate task of reconciling the outputs from different models. Furthermore, HOI can effectively recognize a broader range of human-object interactions, not just actions involving handheld objects.

The proposed autonomous ownership relationship learning framework utilizes the HOI detection model to identify ownership attribution by detecting eligible human-object interaction events, such as drinking from a cup or wearing a piece of clothing. The framework comprises three key components: an HOI detection model, a Re-ID model, and an ownership relationship recognition model. The Re-ID model focuses on matching individuals of interest across camera views and non-overlapping scenes, enabling the service robot to recognize specific individuals in various environments (Ye et al., 2022). The ownership relationship recognition model is a generic image recognition model. The model is trained using images of items that users interact with. These images are labeled based on the user's identity, as outputted by the Re-ID model. This training process aims to enable the ownership relationship recognition model to accurately identify the ownership of object items based on the interaction between the user and the object.

Figure 1: Sample annotations of HICO-DET (Chao et al., 2018).

The framework first utilizes an HOI model to detect human-object interactions in the image frames. Subsequently, the HOI model detects the bounding boxes of the human and object in the image, which are then cropped and processed as separate entities. The human image crop is input into the Re-ID model to obtain the user's identity, which serves as a label for the object image crop. This results in a newly created image sample labeled with item ownership added to the item ownership relationship recognition dataset. Finally, the item ownership relationship recognition model is retrained or fine-tuned to incorporate the newly learned ownership relationship (see Figure 2). The framework is designed to learn item attribution based on specific user behavior effectively and only requires separate training of the item ownership recognition model, thereby reducing the coupling of the system. The HOI detection and close-world Re-ID models can use existing models pre-trained on public datasets, further simplifying the system.

Interactive Item Ownership Relationship Learning Framework

Autonomous detection alone is insufficient to address the complexities of changing item ownership and misidentification in real-world scenarios. In order to improve such systems, we identified the following requirements as drivers of our research for improving such systems: applicability and robustness. To address these challenges, we propose an interactive object ownership relationship learning framework, which facilitates natural human-robot interaction through the user-initiated presentation of objects for the robot to learn from. This framework integrates autonomous and interactive ownership relationship learning to construct a framework for home service robots to learn item ownership relationships. The learning process is divided into four phases (see Figure 3).

Registration Phase

During the Initial Registration phase, the user must initiate the categorization of personal items for detection and provide images and relevant information

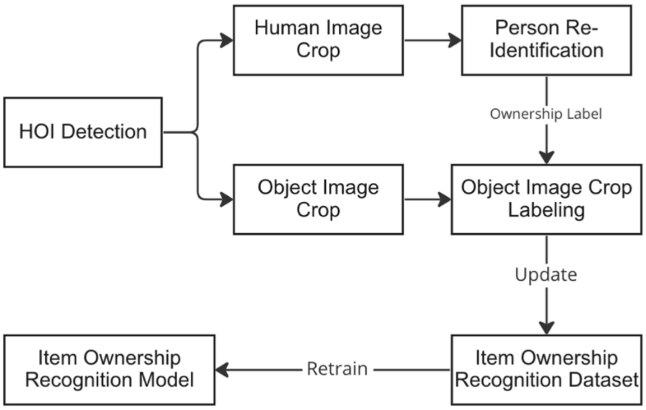

Figure 2: Autonomous ownership relationship learning framework based on human-object interaction (HOI).

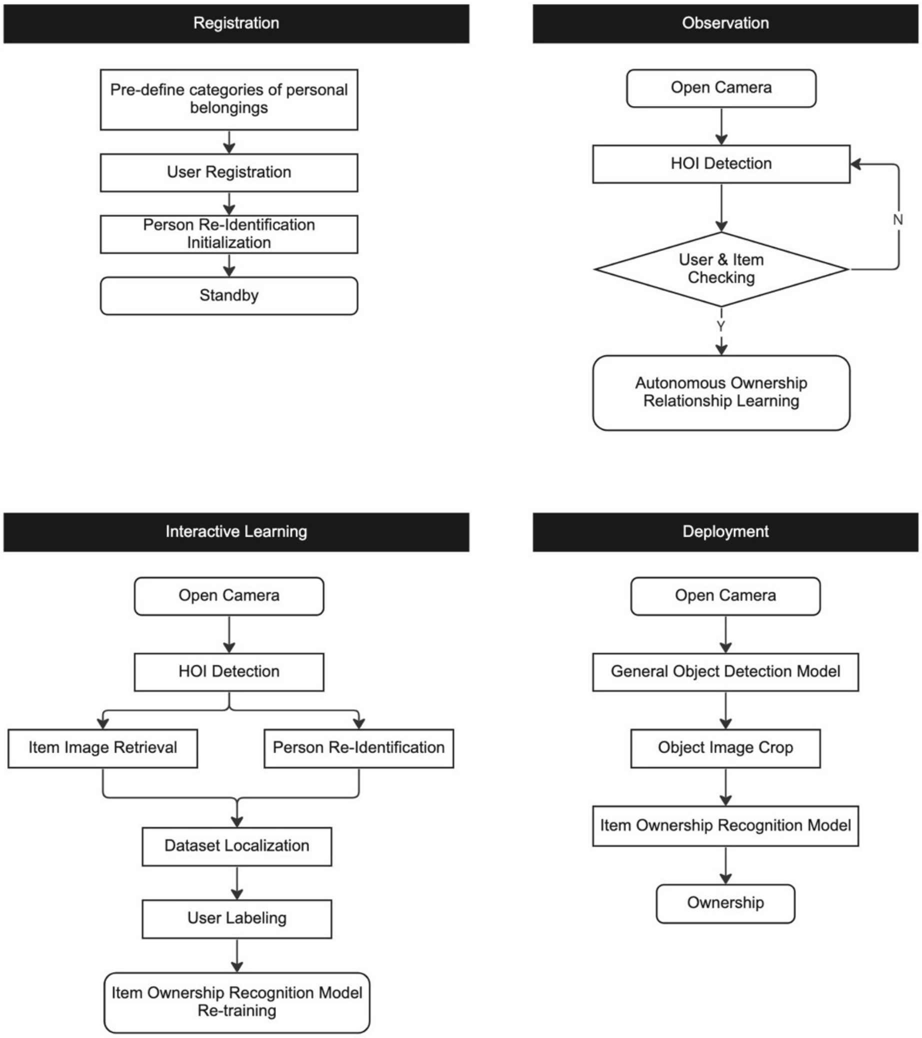

Figure 3: The process of learning item ownership relationships based on the proposed interactive learning framework.

of family members for proper initialization of the Person Re-Identification (Re-ID) Model. This process enables the HOI detection model to exclude unauthorized individuals and public items while focusing on learning and recognizing private objects such as toothbrushes, cups, clothing, etc. By manually specifying these rules, the service robot can effectively mitigate misidentification and improper attribution of items.

Observation Phase

During the observation phase, the robot primarily operates based on the proposed autonomous ownership relationship learning framework for patrolling within the domestic environment. During the detection of Human-Object Interaction (HOI), if the robot observes a human-object interaction that aligns with the registration information, the autonomous ownership relationship learning framework generates a new sample of image with ownership information to advance the understanding of the ownership relationship.

Interactive Learning Phase

For improving the effectiveness and quality of service-oriented social robots in their interactions with humans, fostering natural human-robot interaction is crucial. The proposed Autonomous Ownership Relationship Learning (AORL) approach leverages interactive robot learning methods, such as (Maiettini et al., 2017) and (Lombardi et al., 2022), to allow the robot to train an ownership recognition model in a natural way. During the interactive learning phase, the item's owner can act as the robot's teacher, demonstrating the item and interacting with the robot to update its attribution. This data annotation approach facilitates effective human-robot interaction and enhances the naturalness and robustness of the robot's learning process for a specific task.

To initiate the learning process, the user presents the item for the robot to learn its ownership. The robot locates the item in the dataset and retrieves information about its category and current ownership. Person re-identification is utilized to determine the label of the person to whom the item belongs, and image retrieval is employed to resolve ambiguity caused by similar items. The robot then proactively queries the user regarding the item's new ownership, updates the ownership attribute labels of all item samples in the dataset, and retrains the model. This approach enables robots to request user input to resolve ambiguities in real-world tasks and enhances their understanding of human-item relationships.

Deployment Phase

The framework enables the robot to detect items belonging to the user in a household environment. To do so, the framework employs a generic object detection model to detect the item, then inputs an image clip into the item ownership recognition model to obtain the item's ownership label.

CONCLUSION

This paper proposes an Item Ownership Relationships Learning Framework for service robots. This framework allows the service robots to learn item ownership relationships through autonomous or interactive means. The autonomous component of the framework is based on Human-Object Interaction (HOI) detection, which enables the robot to comprehend item ownership by identifying correlations between human-object interactions. To manage the complexities of changing item ownership in real-world scenarios, we have structured the item ownership relationship learning process into four stages. The proposed framework facilitates natural interaction between users and the robot, allowing users to demonstrate item ownership by actively presenting items to the robot for learning purposes.

However, many other ownership relationships learning capabilities still need to be explored. The current system only learns item ownership relationships by detecting human-object interactions and cannot infer ownership norms such as recognizing changes in ownership. Additionally, the system lacks the capability to distinguish items that appear similar, as it cannot locate the item in real time. While the comprehensive understanding of ownership remains challenging, this study aims to investigate and design a general approach for item ownership relationship learning in service robots based on human-object interaction. As part of future work, a case study will be conducted to assess the performance of the proposed item ownership learning framework and further investigate its applicability and robustness.

REFERENCES

Antoun, M., & Asmar, D. (2023). Human object interaction detection: Design and survey. Image and Vision Computing, 130, 104617. https://doi.org/10.1016/j.imavis.2022.104617

Bansal, A., Rambhatla, S. S., Shrivastava, A., & Chellappa, R. (2020). Detecting Human-Object Interactions via Functional Generalization. Proceedings of the AAAI Conference on Artificial Intelligence, 34(07), 10460–10469. https://doi.org/10.1609/aaai.v34i07.6616

Chao, Y., Liu, Y., Liu, X., Zeng, H., & Deng, J. (2018). Learning to Detect Human-Object Interactions. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), 381–389. https://doi.org/10.1109/WACV.2018.00048

Gao, C., Xu, J., Zou, Y., & Huang, J.-B. (2020). DRG: Dual Relation Graph for Human-Object Interaction Detection. In A. Vedaldi, H. Bischof, T. Brox, & J.-M. Frahm (Eds.), Computer Vision – ECCV 2020 (pp. 696–712). Springer International Publishing. https://doi.org/10.1007/978-3-030-58610-2_41

Gao, C., Zou, Y., & Huang, J.-B. (2018). iCAN: Instance-Centric Attention Network for Human-Object Interaction Detection (arXiv:1808.10437). arXiv. https://doi.org/10.48550/arXiv.1808.10437

Guo, Y., Wang, B., Hughes, D., Lewis, M., & Sycara, K. (2020). Designing Context-Sensitive Norm Inverse Reinforcement Learning Framework for Norm-Compliant Autonomous Agents. 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), 618–625. https://doi.org/10.1109/RO-MAN47096.2020.9223344

Kim, B., Choi, T., Kang, J., & Kim, H. J. (2020). UnionDet: Union-Level Detector Towards Real-Time Human-Object Interaction Detection. In A. Vedaldi, H. Bischof, T. Brox, & J.-M. Frahm (Eds.), Computer Vision – ECCV 2020 (pp. 498–514). Springer International Publishing. https://doi.org/10.1007/978-3-030-58555-6_30

Liao, Y., Liu, S., Wang, F., Chen, Y., Qian, C., & Feng, J. (2020). PPDM: Parallel Point Detection and Matching for Real-Time Human-Object Interaction Detection. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 479–487. https://doi.org/10.1109/CVPR42600.2020.00056

Lombardi, M., Maiettini, E., Tikhanoff, V., & Natale, L. (2022). iCub Knows Where You Look: Exploiting Social Cues for Interactive Object Detection Learning. 2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids), 480–487. https://doi.org/10.1109/Humanoids53995.2022.10000163

Maiettini, E., Pasquale, G., Rosasco, L., & Natale, L. (2017). Interactive data collection for deep learning object detectors on humanoid robots. 2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids), 862–868. https://doi.org/10.1109/HUMANOIDS.2017.8246973

Tan, Z.-X., Brawer, J., & Scassellati, B. (2019). That's mine! Learning ownership relations and norms for robots. Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, 8058–8065. https://doi.org/10.1609/aaai.v33i01.33018058

Wang, T., Yang, T., Danelljan, M., Khan, F. S., Zhang, X., & Sun, J. (2020). Learning Human-Object Interaction Detection Using Interaction Points. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4115–4124. https://doi.org/10.1109/CVPR42600.2020.00417

Wu, H., Chen, Z.-W., Tian, G.-H., Ma, Q., & Jiao, M.-L. (2020). Item Ownership Relationship Semantic Learning Strategy for Personalized Service Robot. International Journal of Automation and Computing, 17(3), 390–402. https://doi.org/10.1007/s11633-019-1206-7

Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L., & Hoi, S. C. H. (2022). Deep Learning for Person Re-Identification: A Survey and Outlook. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(6), 2872–2893. https://doi.org/10.1109/TPAMI.2021.3054775