

Identifying Tasking for Service Robots with Change Detection*

E. Martinson

Abstract—Mobile manipulators show incredible promise as domestic service robots – interacting with a wide variety of objects using increasingly affordable hardware. But although perception, manipulation, and mobility have advanced, there remain fundamental challenges in making the robots more useful. How can a robot proactively identify tasks that it can complete while supporting individual human preferences for how a home should be configured? We propose the use of change detection. Change affords action. Only those objects that have been interacted with need to be considered for tasking. Other objects, even if in non-standard positions in the house, can be ignored. This work combines open-vocabulary based change detection methods with large language models to select the key phrases to search for and identify tasks to be completed.

I. INTRODUCTION

The floorcare robots of today are only the first generation of domestic robots. Even so, they support busy families, improve general quality of life, and help elderly people stay in their own homes. Unfortunately, houses are unstructured environments with too many variables – many of which are unimportant for most tasks – so robotics companies have struggled to generalize other applications to homes, even within the narrower domain of elder care. Change detection can reduce this complexity by focusing attention on only those parts of the environment important to available robotic tasks. Take the example of a mobile manipulator that can detect and put away different categories of objects. Current best approaches to tasking are to treat every object in a predefined area as an object to be put away (e.g. floor [1], table [2], etc). Only picking up objects from a single area does not usually justify the cost of the robot, even in many healthcare applications. And current methods do not generalize well to all areas of the house as there will be too many exceptions. What if a person does not want some objects to be picked up, despite being in an unusual spot according to the Large Language Model (LLM)? Similarly, where should the object be put away? Change detection can be a vital part of such decision making. It is not the only driver people use for deciding to act, but when someone detects change, they make conscious choices about whether to do anything about it.

Previous work in change detection for indoor environments [3] has combined Neural Radiance Field (NeRF) [4] modeling with CLIP-segmentation [5] to search for change associated with open-vocabulary utterances (i.e. “unexpected mess”). This system has the advantage of being adaptable to arbitrary search queries without any need for further training, but it still requires human direction about what to search for. Large

Language Models can fill the gap. Processing only limited numbers of images with context about the room and/or personal needs, an LLM can identify search targets. By comparing those search targets to previous runs using change detection, clean-up tasks can be identified that are not restricted to fixed areas and which take into account knowledge of individual cleaning preferences.

The remainder of this paper is organized as follows. First, the meaningful change detection system and initial results finding unexpected messes is discussed. Second, extensions of this system using an LLM are identified along with some demonstrated environments. Finally, we elaborate on future work with home care robots using change detection.

II. CHANGE DETECTION

Change detection has been a challenge of interest to computer vision and robotics for several decades; one which has grown significantly in interest and scope in recent years thanks to deep learning. Some of the earliest works were on land use, using long range cameras and other remote sensors to automatically capture new construction [6], but more recently there has been significant interest in street level change detection [7] supporting autonomous car development, underwater change [8], and now even indoor environments [9]. Within each of these different domains there are at least two common challenges: accounting for differences in viewpoint when estimating change, and discarding normal, environmental change (e.g. trees that have lost leaves [7], and snow covered ground [10]) from meaningful differences.

The most well explored solution to both problems is to train a neural network with manually labeled examples of change [11] [8] [12]. Another approach is to rely on object detection, building semantic maps [13], using semantic segmentation [12], or even pre-training a captioning system with common examples of the target environment [9]. The problem with all of these approaches is that they require a large amount of hand labeled data to work, making them difficult to deploy to highly individual living spaces like homes – at least not without a lot of effort before deployment.

Recognizing arbitrary change without additional training would significantly increase deployment flexibility. To this end, a multi-step prototype that combines NeRF models with CLIP-segmentation was developed to detect change and deployed this system on the Stretch 2 mobile robot [3]. This prototype involves first building a NeRF model of a baseline “clean” environment. Then, after a subsequent run, newly captured images are registered into the “clean” space and the NeRF model is used to build synthetic images from the same viewpoint. Resulting images are closer in alignment than is possible with nearest neighbor methods (Figure 1) given

*Research supported by Lawrence Technological University.

E. Martinson is with the Department of Math and Computer Science, Lawrence Technological University, Southfield, MI 48075 (e-mail: emartinso@ltu.edu)



Figure 1. The “after” image is captured during a second run through the environment. The “before” image is the NeRF reconstruction of the same viewpoint. The “Nearest Image” is presented to show the closest image captured during the original baseline run. The final, “positive change” image is change detected using the query, “unexpected mess”.

normal variations in robot paths. CLIP-segmentation is applied to both images, the results are then subtracted and clustered to identify regions with significant change. Additional filtering removes known false positives based on depth images.

This method was evaluated originally in the context of a security scenario. The robot was looking for potential security threats over 8 runs, half of which had no interesting change to observe. Remaining runs had missing objects, scattered papers, ransacked bookshelves, and suspicious packages. Performance was analyzed compared to a segmentation only method (no change detection) and a nearest neighbor image matching method (VG). Using NeRF to generate the same view significantly improved precision and specificity while recall stayed about 50% (Table 1).

Algorithm	Precision	Recall	Specificity
Segmentation Only	0.24	0.77	0.48
VG	0.33	0.38	0.82
NeRF	0.78	0.54	0.97

III. SETTING GOALS WITH AN LLM

The existing change detection prototype only searches for predetermined queries. For a service robot, a hand designed query like “unexpected mess” may catch a significant number of the objects that need to be put away, but it will not catch everything – nor will it provide enough information about how to correct said change. The current approach also needs human input to identify objects that should not be put away (e.g. the human does not want books to be moved). An LLM can help with both tasks – setting change detection goals to identify what tasks can be done, and integrating with human

preferences to identify what help is actually needed/wanted (Figure 2).

Setting Detection Goals – Change detection works by comparing before and after images with one or more semantic queries. With the existing publication, we demonstrated queries such as “unexpected mess” and “scattered papers” in a security setting. These queries were intended to be handcrafted by a security guard for the specific needs of the facility. Service robotics, however, is inherently dynamic. It varies from home to home, and from day-to-day. A system is needed that does not require constant manual adjustment to be effective. An LLM can identify such daily detection tasks.

Assuming the robot can identify a few representative views of the environment, perhaps through farthest point sampling methods, the best approach is to upload an image to a multi-modal LLM. In Figure 3, Microsoft CoPilot/ChatGPT was queried with an image and asked to list potential objects the system could clean or tidy up. The LLM correctly identifies multiple objects that could be cleaned including: papers and keyboard. Change detection successfully finds those objects in comparison to a baseline run. The LLM also detects a number of other objects that are successfully discarded by change detection: cloth, mouse, phone, keyboard, etc.

LLM Based Task Selection – The combination of LLM-based query selection and change detection will identify objects that have been moved. As demonstrated by TidyBot [1], the LLM can also identify regions of interest for putting the object away. These regions can be further narrowed by utilizing information about negative change (i.e. missing objects) in the region of interest. But is every such detected change actually interesting? Definitely not.

There are at least two ways in which the LLM can help with task selection. First, the LLM can incorporate user preferences.



Query:
A cleaning robot is being sent to clean this environment. The robot can pick up and put things away. What tasks should it look at performing? List out the objects that the robot can potentially help clean using JSON as {'tasks': [<list of tasks>], 'objects':[{'type': string, 'location': string}]}

CoPilot/ChatGPT Response:

```
{
  "tasks": [
    "Pick up and organize papers on the desk",
    "Put away the tablet",
    "Remove the cloth from the chair",
    "Organize items on the desk"
  ],
  "objects": [
    {
      "type": "paper",
      "location": "desk"
    },
    {
      "type": "tablet",
      "location": "desk"
    },
    {
      "type": "cloth",
      "location": "chair"
    },
    {
      "type": "keyboard",
      "location": "desk"
    },
    {
      "type": "mouse",
      "location": "desk"
    },
    {
      "type": "phone",
      "location": "desk"
    }
  ]
}
```

Figure 2. The LLM is asked to identify potential objects to search for in the environment from a sample image. These can be compared to baseline using change detection.

Even though a floorcare robot can vacuum the floor, people will often mark off areas they do not want vacuumed – whether it is fear of damaging something, a desire to remain undisturbed, or something else. Incorporation of user preferences for suppressing daily tasks is just as important for other service robots. Maybe the person living there does not want the robot to take care of the dishes, or clean up blankets that they are likely to use again in the near future? Maybe someone needs more support with certain tasks (e.g. picking up objects from the floor) due to medical conditions? If a user considers some changes to be inconsequential and/or intrusive, or else needs more help with certain classes of tasks for any reason, the LLM can be directed to incorporate those user preferences to the query along with the sets of detected objects that have changed.

There is also a question of separating important tasks from meaningless tasks directly via the LLM. Take the coffee cup in the image above. Along with a list of everything that has been detected to have changed, the LLM can be asked to estimate the importance of each task. For the picture above, we informed the LLM of the user’s desire to ignore all but the biggest messes on the desk, and asked the LLM to rank the importance of each of the detected objects. CoPilot/ChatGPT ranked the objects as follows: Paper=30%, Mouse=20%, Phone=20%, Cloth=80%, identifying Cloth as the biggest target, but also differentiating Paper from other desk targets due to relative size.

IV. DISCUSSION

The possibility of using change detection in combination with an LLM to generate tasking for a mobile service robot is exciting. This work has demonstrated to date: (1) the ability to detect change using a combination of NeRF reconstructions and CLIP-segmentation; (2) how an LLM can generate queries for this approach with context about the task at hand; and (3) identified next steps for using an LLM to combine detected changes and user preferences into tasking for a service robot to support a wide variety of people in their homes. In the near future, we aim to implement a working prototype and evaluate its effectiveness in multiple environments.

Although this paper has focused particularly on providing services for a wide range of people in their homes, we should not dismiss the potential value of change detection for remote patient monitoring activities as well. Identifying which objects are not put away on a regular basis, or tracking the usage of the robot’s services could be indicators of worsening physical or mental conditions:

1. Occupational therapists, for instance, have long associated object placement with range of motion or musculoskeletal issues, and even suggest a range of home modifications to support long-term conditions [14].
2. Depression has been previously connected with “household chaos” or messy rooms [15] [16].
3. Alzheimers and dementia are also correlated with increased disorder, and specific problems like objects being left in unusual places or lost entirely [17] [18].

Change detection remains a long way from supporting remote patient monitoring tasks, but the value for both patient monitoring and general tasking of service robots is significant. Homes are too difficult and individual for logical if/then style methodologies to succeed on a large scale. But change affords action. And this is an important missing piece for deploying service robots that really help people.

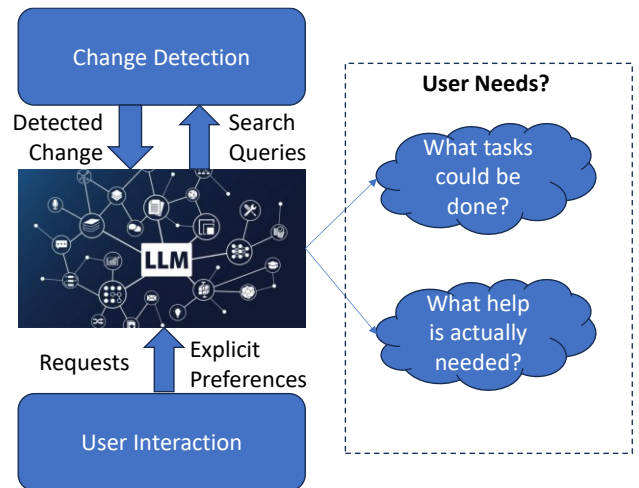


Figure 3 – The LLM can provide search queries to guide change detection in response to human preferences and requests – answering what tasks need to be done and what help is actually needed.

V. REFERENCES

- [1] J. Wu, R. Antonova, A. Kan, M. Lepert, A. Zeng, S. Song, J. Bohg, S. Rusinkiewicz and T. Funkhouser, "TidyBot: Personalized Robot Assistance with Large Language Models," *Autonomous Robots*, vol. 47, pp. 1087-1102, 2023.
- [2] F. Reister, M. Grotz and T. Asfour, "Combining Navigation and Manipulation Costs for Time-Efficient Robot Placement in Mobile Manipulation Tasks," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, 2022.
- [3] E. Martinson and P. Lauren, "Meaningful Change Detection in Indoor Environments Using CLIP Models and NeRF-based Image Synthesis," in *accepted to Ubiquitous Robots*, New York, 2024.
- [4] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi and R. Ng, "Nerf: Representing scenes as neural radiance fields," *Communications of the ACM*, vol. 65, no. 1, pp. 99-106, 2021.
- [5] T. Lüddecke and A. S. Ecker, "Image Segmentation Using Text and Image Prompts," in *Conf on Computer Vision and Pattern Recognition*, New Orleans, 2022.
- [6] L. Khelifi and M. Mignotte, "Deep Learning for Change Detection in Remote Sensing Images: Comprehensive Review and Meta-Analysis," *IEEE Access*, vol. 8, 2020.
- [7] K. Sakurada and T. Okatani, "Change Detection from a Street Image Pair using CNN Features and Superpixel Segmentation," in *British Machine Vision Conference*, Swansea, 2015.
- [8] M. Mandal and S. K. Vipparthi, "An Empirical Review of Deep Learning Frameworks for Change Detection: Model Design, Experimental Frameworks, Challenges and Research Needs," *IEEE Trans on Intelligent Transportation Systems*, vol. 23, no. 7, 2022.
- [9] Y. Obinata, K. Kawaharazuka, N. Kanazawa, N. Yamaguchi, N. Tsukamoto, I. Yanokura, S. Kitagawa, K. Shinjo, K. Okada and M. Inaba, "Semantic Scene Difference Detection in Daily Life Patrolling by Mobile Robots using Pre-Trained Large-Scale Vision-Language Model," in *Intelligent Robotics and Systems*, Detroit, 2023.
- [10] P. F. Alcantarilla, S. Stent, G. Ros, R. Arroyo and R. Gherardi, "Street-view change detection with deconvolutional networks," *Autonomous Robots*, vol. 42, pp. 1301-1322, 2018.
- [11] A. Varghese, J. Gubbi, A. Ramaswamy and B. P. "ChangeNet: A Deep Learning Architecture for Visual," in *ECCV*, Munich, 2018.
- [12] G.-H. Wanga, B.-B. Gaob and C. Wang, "How to Reduce Change Detection to Semantic Segmentation," *Pattern Recognition*, vol. 138, no. 109384, 2023.
- [13] L. Göncz and A. L. Majdik, "Object-Based Change Detection Algorithm with a Spatial AI," *Sensors*, vol. 22, no. 6342, 2022.
- [14] D. E. Jachan, U. Müller-Werdan, N. A. Lahmann and S. Strube-Lahmann, "Smart@home – supporting safety and mobility of elderly and care dependent people in their own homes through the use of technical assistance systems and conventional mobility supporting tools: a cross-sectional survey," *BMC Geriatrics*, vol. 21, no. 205, 2021.
- [15] S. Marsh, R. Dobson and R. Maddison, "The relationship between household chaos and child, parent, and family outcomes: a systematic scoping review," *BMC Public Health*, vol. 20, 2020.
- [16] C. Pattemore and S. Pineda, "The Link Between Messy Rooms and Depression," PsychCentral, 6 May 2022. [Online]. Available: <https://psychcentral.com/depression/messy-room-depression>.
- [17] L. McGarrigle, S. Howlett, H. Wong, J. Stanley and K. Rockwood, "Characterizing the symptom of misplacing objects in people with dementia: findings from an online tracking tool," *International Psychogeriatrics*, vol. 31, no. 11, pp. 1635-1641, 2019.
- [18] Alzheimers Association, "Dementia and hiding, hoarding or losing things," [Online]. Available: <https://www.alzheimers.org.uk/about-dementia/symptoms-and-diagnosis/symptoms/hiding-hoarding-losing>. [Accessed 26 February 2025].