ROBOTIC CLEANING AND MAINTENANCE TECHNIQUES FOR SMART HOMES

Y. Li
X. Wie

Household robots are accelerating the evolution of smart-home living, and effective cleaning depends heavily on reliable perception and navigation. This study improves image-recognition capabilities within a robotic cleaning system by focusing on its image-processing module. The proposed approach refines image preprocessing and feature extraction to suppress noise and enhance robustness, and integrates regional stereo-matching constraints to achieve more accurate region-level correspondence. Dynamic obstacles are tracked in real time using a SURF–KLT scheme, and a Greedy strategy is then applied to pinpoint moving targets, lowering collision risk and increasing cleaning coverage. Experimental results show an image-matching accuracy of 99.7%, with an mAP@0.5 of 0.932 and stable training precision–recall behavior. In practical cleaning tests, the robot successfully identified 23 pieces of household waste and computed a weighted total score of 39 to support optimal path planning.

Index Terms — region stereo matching constraints, SURF-KLT algorithm, Greedy algorithm, dynamic tracking

https://doi.org/10.66033/judsc2025-205

INTRODUCTION

With the rapid advancement of social sciences and emerging technologies, alongside the growing demand for higher living standards, intelligent residences have gradually evolved into a prominent trend as an innovative integration of advanced technology and living environments [1, 2]. Intelligent residences employ a wide range of sophisticated technologies and devices to create living spaces that are efficient, safe, convenient, and comfortable. In the domain of smart homes, technological applications now extend far beyond basic device interconnectivity. Contemporary smart home systems achieve seamless communication and intelligent coordination among household devices through the integration of advanced Internet of Things (IoT) technologies.

For instance, smart thermostat systems can automatically regulate heating and cooling based on indoor and outdoor temperature, humidity levels, and user preferences, thereby ensuring both energy efficiency and thermal comfort [3, 4]. Similarly, intelligent kitchen systems offer personalized services such as ingredient management and cooking recommendations by integrating smart appliances, including refrigerators and ovens, significantly enhancing convenience and safety [5, 6, 7]. Moreover, modern security systems have evolved beyond conventional video surveillance and access control, incorporating advanced technologies such as facial recognition and behavioral analysis, which substantially improve overall residential security [8, 9].

Despite these advancements, residential cleaning—an essential component of smart living—has primarily relied on conventional sweeping robots. Currently, cleaning robots have achieved a global penetration rate of approximately 38%, whereas maintenance robots remain relatively uncommon [10, 11]. In addition, cleaning tasks in smart homes are typically performed two to four times per week. While traditional intelligent robots can effectively replace manual labor for routine cleaning, they often struggle with complex scenarios such as irregular furniture layouts, special wall or floor materials, localized deep-cleaning requirements, and sudden cleaning demands [12, 13, 14, 15]. Consequently, the targeted development of advanced robotic cleaning and maintenance technologies is crucial for enhancing the overall functionality of smart residential environments.

This study explores the application of image recognition techniques and region-based stereo matching methods to improve the performance of residential cleaning and maintenance robots. Leveraging digital image processing, a quantitative evaluation of residential cleanliness is achieved through modules for preprocessing, feature extraction, and data analysis. To address the challenges associated with parallax computation in stereo matching, a multi-constraint fusion strategy is proposed, integrating epipolar geometry and region similarity measures to improve matching efficiency. Furthermore, the SURF algorithm is employed for rapid feature point detection, while the KLT algorithm is utilized for feature matching. In combination with a Greedy algorithm, these methods enable accurate spatial localization of the cleaning targets.

ANALYSIS OF ROBOTIC RESIDENTIAL CLEANING AND MAINTENANCE TECHNOLOGIES

Image Recognition Technology

In a vision-driven residential cleaning robot, the perception component is largely built upon digital image processing (DIP). DIP can be understood as the computational manipulation and interpretation of images to support operations such as enhancement, feature representation, statistical evaluation, and recognition. In this context, the goal is to strengthen useful visual information, derive discriminative descriptors, and ultimately infer an objective level of household cleanliness.

The overall DIP pipeline used for cleanliness assessment can be summarized as follows:

1. **Preprocessing:** Apply common enhancement and filtering operations to suppress noise and reduce unwanted visual interference while emphasizing relevant structures.
2. **Feature representation:** Extract informative patterns from the processed images and convert them into measurable feature vectors.
3. **Analytical evaluation:** Examine the feature vectors to form decision rules and numerical criteria that support subsequent recognition.
4. **Recognition and decision:** Use the established criteria to identify the cleanliness condition of the residential environment.

This staged procedure enables automated, repeatable judgment of indoor conditions and provides the perception basis for intelligent robotic actions.

Regional Stereo Matching

Fundamental Constraints in Stereo Matching

Stereo matching aims to compute reliable correspondences efficiently; therefore, both precision and runtime are central concerns. In real acquisition settings, however, the imaging process is affected by multiple sources of disturbance, including sensor noise, optical distortion, illumination fluctuation, occlusion, repeated surface patterns, and disparity jumps near depth edges. Because the same 3D scene point can project differently across two cameras, correspondence may become ambiguous and may lead to incorrect one-to-many matches.

To control this complexity, stereo vision typically employs a collection of constraints derived from geometry and appearance. These rules narrow the search domain, simplify computation, and reduce mismatches. Commonly adopted constraints include:

1. **Epipolar-line constraint:** Given a point in one image, its correspondence in the other image must lie on the associated epipolar line. Once the epipolar geometry is known, correspondence search is reduced from a two-dimensional scan to a one-dimensional scan along the line, which accelerates matching and lowers the error rate. Since physical mounting inaccuracies and lens effects prevent perfect row alignment in practice, rectification based on calibration results is normally required before matching.
2. **Uniqueness constraint:** Each feature is assumed to have a single best correspondence. This assumption is valid in many scenes but may break down in texture-repetitive or texture-poor regions.
3. **Smoothness (continuity) constraint:** Except at object boundaries, disparity tends to vary gradually across most surfaces. Hence, neighboring pixels on the same surface usually correspond to neighboring pixels in the paired image.
4. **Photometric compatibility constraint:** Although lighting changes and viewpoint differences may alter intensity values, corresponding neighborhoods are expected to remain broadly similar. Therefore, regional similarity is preferred over single-pixel comparison, and this idea forms the core of region-based matching.
5. **Left–right consistency constraint:** If a point in the left image matches a point in the right image, reversing the reference images should yield the same pair. This property is useful for locating occlusions and rejecting unreliable matches.

6. **Ordering constraint:** Along the same epipolar line, the relative sequence of corresponding features is preserved between the two images, preventing crossing correspondences.
7. **Disparity-range constraint:** Since disparity is inversely related to depth, restricting disparity to a plausible interval limits the search range and significantly decreases computational cost.

Mechanism of Region-Based Stereo Matching

Region-based matching determines correspondences by comparing intensity patterns in local neighborhoods. This strategy is effective in areas with rich texture, but it often produces higher mismatch rates in occluded zones or regions with weak texture.

Let $P(x,y)$ be the candidate point in the reference (left) image. A square window of size $(2n+1)\times(2n+1)$ is centered at $P(x,y)$. Along the corresponding epipolar line in the right image, candidate windows of the same size are examined within a predefined horizontal search interval $[x_1,x_2]$. A similarity (or cost) function is computed for each candidate, and the position that yields the strongest similarity (or minimum cost) is selected as the match.

In this work, the absolute gray-level difference is used as the elementary cost. A clipping threshold T is introduced to reduce the influence of outliers during accumulation:

$$C(x,y)=\begin{cases}C(x,y), & C(x,y)<T,\\ T, & C(x,y)\geq T.\end{cases} \tag{1}$$

The window cost for disparity d is then formed by summing the local differences:

$$C(x,y,d)=\sum_{i=-n}^{n}\sum_{j=-n}^{n}\left|L(x+i,y+j)-R(x+d+i,y+j)\right|, \tag{2}$$

where $L(\cdot)$ and $R(\cdot)$ denote the left and right images, respectively, and (i,j) represent window offsets. The disparity producing the smallest accumulated cost within the search range is taken as the optimal match.

Although region-based matching is computationally intensive because each pixel requires window aggregation, the approach remains attractive due to its straightforward structure and its suitability for optimization toward real-time performance. Moreover, many pipelines initially output integer-valued disparity; when higher precision is required, interpolation, curve fitting, or filtering techniques can be applied to refine the result to subpixel accuracy.

Motion Target Tracking Using SURF–KLT

To robustly estimate the position of cleaning targets across consecutive frames, this study combines SURF-based feature detection with KLT-based feature tracking, and then applies a Greedy strategy to finalize the target bounding region. First, SURF is employed to quickly locate stable interest points with strong robustness. Next, these points are tracked frame-to-frame using the KLT method to maintain temporal stability. Finally, the tracked feature distribution is used to infer the target location and size.

SURF identifies feature points by constructing an integral image, building a scale-space representation, and locating extrema via the Hessian matrix followed by non-maximum suppression. For a point (x,y) at scale σ, the Hessian matrix is written as

$$H(x,y,\sigma) = \begin{bmatrix} L_{xx}(x,y,\sigma) & L_{xy}(x,y,\sigma) \\ L_{xy}(x,y,\sigma) & L_{yy}(x,y,\sigma) \end{bmatrix}. \tag{3}$$

To accelerate computation, the second-order derivatives are approximated using box filters. The Hessian determinant is commonly approximated by

$$\det(H) \approx D_{xx}D_{yy} - 0.9D_{xy}^2. \tag{4}$$

After extracting a suitable set of interest points, the KLT algorithm is applied for matching/tracking with low computational overhead. KLT is based on optimal estimation and uses the sum of squared differences (SSD) between local patches in two consecutive frames. If $I(x,y,t)$ is the image at time t, brightness constancy yields

$$I(x+dx, y+dy, t+dt) = I(x,y,t), \tag{5}$$

where $\mathbf{d} = (dx, dy)$ represents the displacement. The KLT tracker seeks the displacement that minimizes the SSD over a small window, typically through iterative refinement (e.g., Newton iterations) until the desired accuracy is reached.

Once multiple feature points are successfully tracked, the target region is estimated from their spatial distribution. Specifically, a bounding rectangle that encloses the tracked feature set is computed to obtain the target width w and height h. A Greedy procedure is then used to refine the rectangle parameters so that the resulting target localization is stable and consistent across frames.

EVALUATION OF ROBOT CLEANING AND MAINTENANCE PERFORMANCE BASED ON TRACKING AND MATCHING ALGORITHMS

Algorithm Performance Evaluation

Matching Error Analysis

Following the integration of region-based stereo matching techniques and the SURF–KLT framework to enhance both regional correspondence accuracy and motion tracking robustness, it is necessary to assess the effectiveness of the proposed algorithm using standard performance indicators. Commonly adopted classification metrics include accuracy, recall, and the F1-score. In this study, algorithmic performance is primarily evaluated using the confusion matrix, precision–recall (PR) curves, and related statistical indicators.

A dataset consisting of 5000 images representing typical residential waste objects was constructed for experimental validation. Among these, 20% of the images were randomly selected as the testing subset to evaluate the matching performance. The confusion matrix illustrated in Figure 1 summarizes the classification results, where rows denote the predicted categories and columns indicate the ground-truth labels. The diagonal elements represent correct classifications, while the off-diagonal elements correspond to misclassifications.

Out of 1000 test samples, 997 were correctly categorized into their respective seven waste classes, and only three samples were incorrectly assigned. This corresponds to a matching accuracy of 99.7%, indicating that the introduction of the proposed matching strategy significantly improves the robot's ability to identify household waste accurately.

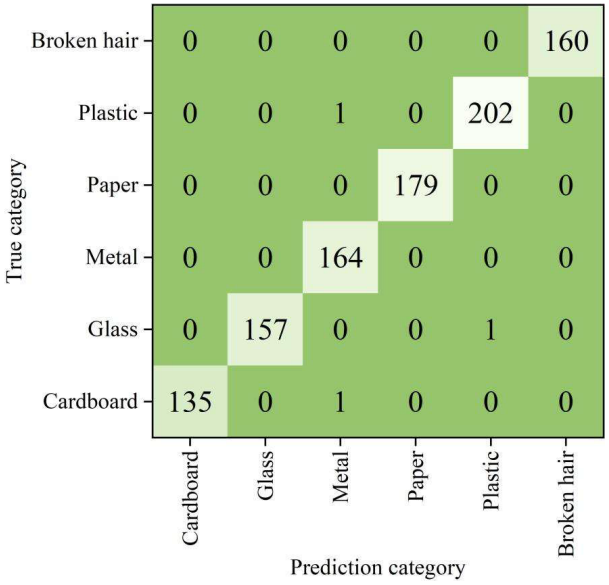

Figure 1: Algorithm matching result

Performance Comparison Before and After Optimization

Since multiple algorithms were combined to enhance the image processing module, a step-by-step comparison was conducted to evaluate their individual and collective contributions. Specifically, four configurations were examined: (i) SURF only, (ii) SURF combined with KLT, (iii) SURF–KLT combined with the Greedy strategy, and (iv) SURF–KLT with Greedy optimization and stereo matching constraints.

Using 40% of the dataset as the training set, the corresponding mAP@0.5 curves are shown in Figure 2. After 150 training iterations, the mAP@0.5 values reached 0.755, 0.814, 0.893, and 0.932, respectively. The monotonic increase across these configurations demonstrates that the collaborative use of multiple algorithms substantially enhances the perception and tracking capability of the robotic system, thereby meeting the precision requirements for waste detection and cleaning tasks.

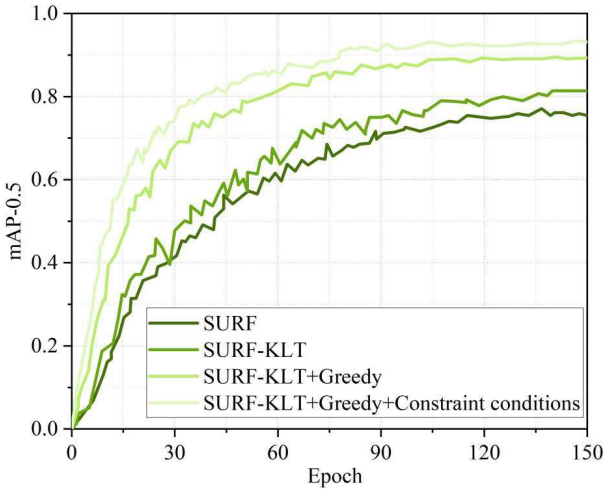

Figure 2: mAP-0.5 curves of different algorithms

Furthermore, Figure 3 illustrates the precision and recall trends of the proposed method incorporating SURF–

KLT, Greedy optimization, and stereo constraints. The relatively small oscillations indicate stable convergence. Both precision and recall gradually increase and stabilize within the range of 0.85–0.87 after approximately 120 iterations, confirming the reliability and consistency of the training process.

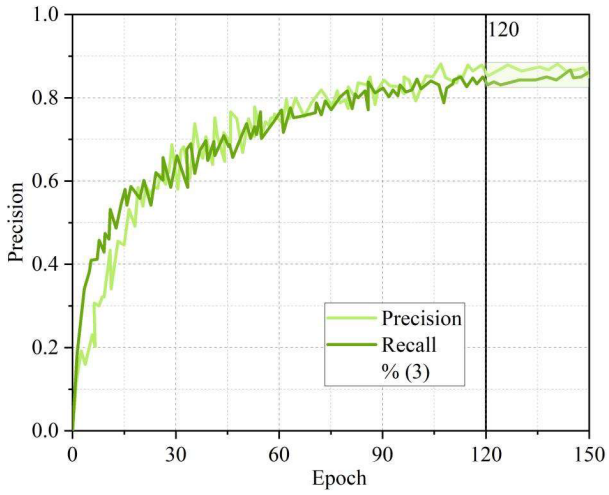

Figure 3: Fluctuations of Precision and Recall of paper method

Practical Implementation of Robotic Cleaning

Cleanliness Evaluation

The trained algorithm was embedded into the visual perception module of the cleaning robot for real-world testing. A residential area of 70 m^2 was selected as the experimental environment. The robot dynamically identified and categorized seven types of waste items distributed throughout the residence.

According to the adopted classification scheme, cardboard and paper were labeled as organic waste, metals and plastics as inorganic waste, scattered hair as uncollected sweeping waste, and glass fragments as solid residue. Each waste category was assigned a corresponding weight to compute a comprehensive cleanliness index. When this index exceeds a threshold value of 20, the robot initiates deep-cleaning operations.

Different cleaning strategies were adopted for different waste types. Plastic, metal, and paper waste were handled by sweeping, hair was removed using suction, and solid residues such as glass were collected mechanically. Table 1 summarizes the detected waste distribution and the corresponding weighted values. A total of 23 waste items were detected, resulting in an aggregated score of 39, which exceeds the predefined threshold and triggers deep cleaning.

Optimal Path Planning

After identifying the waste distribution, the robot plans an optimal cleaning trajectory based on the detected categories and their spatial positions. This planning process minimizes redundant traversal, avoids obstacles such as furniture and corners, and reduces collision risks.

As illustrated in Figure 4, the robot starts at the origin $(0,0)$ and moves sequentially to the target zones. It first reaches $(25,5)$ to remove scattered hair, then proceeds to $(20,30)$ and $(20,50)$ to collect inorganic waste.

Table 1: Quantity of Garbage in Residences and its Weighted Total Value

Types of Garbage	Quantity of Garbage	Weighted Total Value
Inorganic waste (metals, plastics)		
Small	5	7
Middle	1	3
Big	4	6
Organic waste (cardboard, paper)		
Small	6	8
Middle	3	5
Big	1	3
Uncollected cleaning waste (broken hair)	2	4
Solid residue (glass)	1	3
Total	**23**	**39**

Next, it moves to $(30,50)$ to retrieve solid residues, followed by $(50,40)$ for organic waste sweeping. After completing the cleaning process, the robot navigates to $(70,70)$ to enter a standby state.

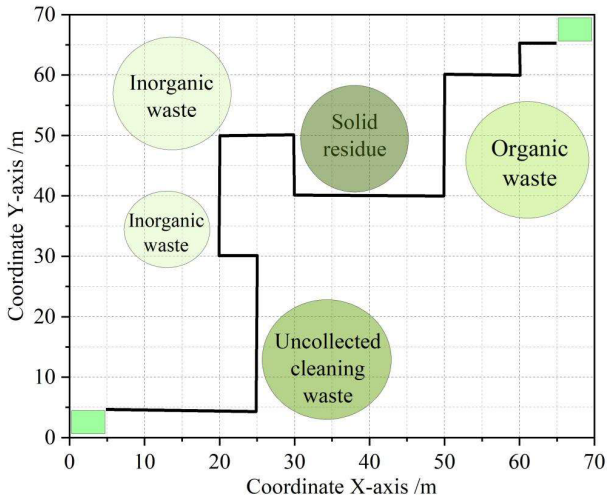

Figure 4: Optimal cleaning path

This optimized route minimizes repeated area coverage and ensures efficient task completion within the shortest possible time, thereby demonstrating the effectiveness of the proposed perception and planning framework.

CONCLUSION

This study presents a multi-algorithm fusion framework that significantly enhances the visual perception and operational performance of residential cleaning robots. By integrating stereo matching constraints with SURF–KLT-based tracking and Greedy optimization, the proposed method achieves a classification accuracy of 99.7% on a dataset of 1000 waste images. The combined framework attains an mAP@0.5 of 0.932, while precision and recall stabilize between 0.85 and 0.87 after 120 training iterations, satisfying the accuracy

requirements for intelligent cleaning tasks.

In a real-world environment of 70 m^2, the robot accurately localized 23 waste regions, computed a weighted cleanliness index, and generated an efficient cleaning path based on dynamic target tracking. These results confirm that the proposed method effectively improves both environmental perception and task execution efficiency.

Future work will extend this approach to larger residential spaces to investigate scalability, explore system limitations, and further enhance the robustness and adaptability of robotic cleaning and maintenance systems.

REFERENCES

[1] Tutkun, N., Burgio, A., Jasinski, M., Leonowicz, Z., & Jasinska, E. (2021). Intelligent scheduling of smart home appliances based on demand response considering the cost and peak-to-average ratio in residential homes. Energies, 14(24), 8510.

[2] Hui, T. K., Sherratt, R. S., & Sánchez, D. D. (2017). Major requirements for building Smart Homes in Smart Cities based on Internet of Things technologies. Future Generation Computer Systems, 76, 358-369.

[3] Sung, W. T., & Hsiao, S. J. (2020). The application of thermal comfort control based on Smart House System of IoT. Measurement, 149, 106997.

[4] Fontes, F., Antão, R., Mota, A., & Pedreiras, P. (2021). Improving the ambient temperature control performance in smart homes and buildings. Sensors, 21(2), 423.

[5] Yu, Y., & Sung, T. J. (2023). A value-based view of the smart PSS adoption: A study of smart kitchen appliances. Service Business, 17(2), 499-527.

[6] Moyeenudin, H. M., Bindu, G., & Anandan, R. (2021). Hyper-personalization of mobile applications for cloud kitchen operations. In Intelligent Computing and Innovation on Data Science: Proceedings of ICTIDS 2021 (pp. 247-255). Springer Singapore.

[7] Nugroho, F., & Pantjawati, A. B. (2018, July). Automation and monitoring smart kitchen based on Internet of Things (IoT). In IOP Conference Series: Materials Science and Engineering (Vol. 384, p. 012007). IOP Publishing.

[8] Dhobale, M. R., Biradar, R. Y., Pawar, R. R., & Awatade, S. A. (2020). Smart home security system using Iot, face recognition and raspberry Pi. International Journal of Computer Applications, 176(13), 45-47.

[9] Taiwo, O., Ezugwu, A. E., Oyelade, O. N., & Almutairi, M. S. (2022). Enhanced intelligent smart home control and security system based on deep learning model. Wireless communications and mobile computing, 2022(1), 9307961.

[10] Kim, J., Mishra, A. K., Limosani, R., Scafuro, M., Cauli, N., Santos-Victor, J., ... & Cavallo, F. (2019). Control strategies for cleaning robots in domestic applications: A comprehensive review. International Journal of Advanced Robotic Systems, 16(4), 1729881419857432.

[11] Naptsoksch, B. (2022). Smart robot using in smart homes. Wasit Journal of Computer and Mathematics Science, 1(4), 55-59.

[12] Hussin, M., Jalani, J., Powdzi, M., Rejab, S., & Ishak, M. K. (2024). Smart Robot Cleaner Using Internet of Things. Journal of Advanced Research in Applied Sciences and Engineering Technology, 46, 175-186.

[13] Vorotnikov, S. A., Nikitin, N. I., & Ceccarelli, M. (2015). A Robotic System for Inspection and Repair of Small Diameter Pipelines. Science & Education of Bauman MSTU/Nauka i Obrazovanie of Bauman MSTU, (2).

[14] Ramalingam, B., Yin, J., Rajesh Elara, M., Tamilselvam, Y. K., Mohan Rayguru, M., Muthugala, M. V. J., & Félix Gómez, B. (2020). A human support robot for the cleaning and maintenance of door handles using a deep-learning framework. Sensors, 20(12), 3543.

[15] Zhang, K. (2024). Technological State and Optimization Analysis of High-Rise Glass Curtain Wall Cleaning Robots. Applied and Computational Engineering, 102, 148-154.

Y. Li, The Chinese university of Hong Kong

X. Wie, The Chinese university of Hong Kong

Manuscript Published; 23 March 2025.