International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 6605-6613

https://doi.org/10.70517/ijhsa464560

Research on Robotic Cleaning and Maintenance Technology for Smart Houses

Yan Li[1,*]
[1] Anhui Vocational and Technical College, Hefei, Anhui, 230011, China
Corresponding authors: (e-mail: liyan225588@126.com).

Abstract The application of residential robots promotes the development of smart home. This paper optimizes the image recognition technology in the robot cleaning system. This paper takes the image processing module of the robot cleaning system as the core, optimizes the image preprocessing and feature extraction algorithm, and reduces the noise interference. The regional stereo matching constraints are fused to realize the accurate matching of regional images. Real-time tracking of dynamic obstacles is realized by the SURF-KLT algorithm, after which the Greedy algorithm is used to accurately locate the moving target, reduce the risk of robot collision and improve the cleaning coverage rate. The results show that the image matching accuracy of the method in this paper reaches 99.7%. And the mAP-0.5 value is as high as 0.932, and the fluctuation of training precision and recall is smooth. In the cleaning practice, the robot is able to detect the 23 garbage present in the residence and calculate its weighted total value as 39 to plan the optimal cleaning path.

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

I. Introduction

Currently, with the continuous development of social science and technology and the continuous pursuit of people's expectations for quality of life, intelligent residence, as an innovative model that integrates advanced technology and living space, has gradually become a trend [1], [2]. Intelligent residence utilizes a variety of advanced technologies and equipment to make the residence an efficient, safe, convenient and comfortable living environment. From the field of smart homes, the deep application of technology has gone far beyond simple interconnection of devices. Modern smart home systems realize seamless communication and intelligent linkage between home devices by integrating advanced IoT technologies. For example, smart thermostat systems can automatically adjust air conditioning or heating according to indoor and outdoor temperatures, humidity, and occupant preferences to achieve an energy-efficient and comfortable living environment [3], [4]. Intelligent kitchens provide personalized services and monitoring such as ingredient management and cooking suggestions through the integration of smart refrigerators and ovens, which greatly improves the convenience and safety of life [5]-[7]. Intelligent security systems are not limited to traditional video surveillance and access control management, but also incorporate advanced technologies such as face recognition and behavioral analysis, which effectively enhance the level of community security [8], [9].

However, cleaning, as one of the focuses of smart homes, is mostly worked by ordinary sweeping robots. Among them, cleaning robots are utilized all over the world with a penetration rate of 38%, while maintenance robots are relatively rare [10], [11]. In addition, smart residential cleaning tasks are generally carried out 2-4 times per week, and traditional intelligent robots can replace the human hand to complete the daily simple cleaning work, but in the face of irregular furniture, special materials of the wall or floor, special cleaning, sudden cleaning needs and other factors, such robots are difficult to meet [12]-[15]. Therefore, targeted development of robotic cleaning and maintenance technology in smart houses is of great significance to the functional improvement of smart houses.

This paper investigates the role of image recognition technology and area stereo matching methods in enhancing the effectiveness of robots for residential cleaning and maintenance. Based on digital image processing technology, quantitative assessment of residential cleanliness is realized through preprocessing, feature extraction and data analysis modules. Aiming at the parallax computation difficulty in stereo matching, a multi-constraint fusion strategy is proposed, combining polar geometry and region similarity metrics to enhance the matching speed. SURF algorithm is introduced for fast detection of feature points, KLT algorithm completes the feature matching, and combined with Greedy algorithm realizes the determination of the position of the

II. Analysis of robotic residential cleaning and maintenance technologies

II. A.Image Recognition Technology

The image processing module in the computer vision-based robotic cleaning system mainly relies on digital image processing technology. Digital image processing (DIP) is the use of computers for image preprocessing, feature extraction, data analysis, image recognition and other operations, to achieve the purpose of enhancing the image, extracting image features and identifying feature images, to achieve the judgment of the degree of residential cleanliness. Digital image processing technology mainly contains the following aspects:

(1) Image pre-processing: using basic image enhancement methods to reduce the impact on the image and strengthen the image features.

(2) Feature extraction: extract the image information after preprocessing, and dataize the image features.

(3) Data analysis: analyze the feature data to provide a judgment standard for image recognition.

(4) Image Recognition: Apply the established criteria to recognize the cleanliness of the residence.

II. B.Regional stereo matching

II. B. 1) Basic constraints in stereo matching

The key to stereo matching research is the accuracy of the results and the speed of operation. There are many disturbing factors in the imaging process of an image, such as noise, lens aberrations, illumination variations, occlusion, repetitive textures, and parallax discontinuities. The pixels of the same point in the physical space projected in different cameras are somewhat different, which causes the matching to appear as one-to-many, i.e., mis-matching. In stereo vision, some basic constraint rules are proposed, based on which the complexity of the matching problem can be simplified, the search range of matching can be reduced, and thus the rate of matching can be improved and the probability of mis-matching can be reduced. The following is an introduction to the basic constraints that are commonly used:

(1) Outer pole line constraint

The epipolar line is the intersection of the observation point P with the plane established by the optical centers O_l and O_r and the imaging plane of the left and right cameras. Figure 1 shows the epipolar line constraint process. The epipolar line constraint means that given a feature point within an image, its corresponding point must exist on the corresponding polar line of another image. This means that as long as the epipolar line of a stereo image pair can be determined, the matching of corresponding points between stereo image pairs is converted from a two-dimensional search of images to a one-dimensional search on the polar lines. This not only speeds up the computation, but also reduces the probability of errors. Ideally, in a stereo vision system, the epipolar line where the image point is located in the image coincides with the pixel row where it is located. In practice, however, it is difficult to achieve complete row alignment for stereo image pairs due to physical mounting, distortions, etc. Therefore, the results of camera calibration need to be utilized for correction to align the rows of the left and right images before matching.

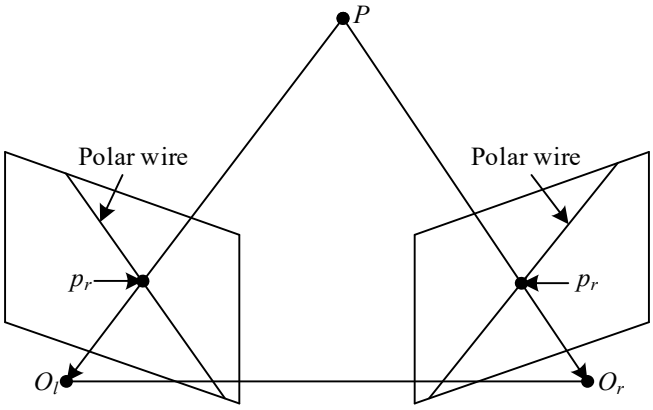

Figure 1: Outer polar line constraint

(2) Uniqueness constraint

This constraint states that there can only be one best match corresponding to each feature point. This holds in most cases, but this constraint fails when there are duplicate textures or untextured objects in the scene.

(3) Continuity constraint

Since most of the object surfaces are uniformly continuous and only the edge regions are fractured, the image parallax should satisfy the continuity property beyond the edges. Under this constraint, two neighboring points on the same object are also adjacent to the corresponding two points in the image plane.

(4) Compatibility constraint

The compatibility constraint is used to describe whether the corresponding point information is similar or not. Due to the interference of light source, shooting angle and other reasons, the brightness of the corresponding points in the image will have some differences, but basically similar. It is difficult to match the corresponding points by the grayscale of a single pixel, and it is necessary to combine with the neighboring pixels to determine the compatibility constraint, which is the core constraint in the regional stereo matching.

(5) Left-right consistency constraint

If the left image is used as a reference, one of its image points A corresponds to B within the right image; conversely, using the right image as a reference, the image point B corresponds to the point A necessarily in the left image. This constraint can be used for occluded region detection.

(6) Order constraint

Under this constraint, one pair of epipolar lines within the left and right images, with feature points in order on each line, have their matching points in the same order on the other line.

(7) Parallax constraint

From the geometric relationship is easy to obtain, the camera on the object and the camera distance and the corresponding parallax inversely proportional to the relationship. By setting the parallax constraint, the nearest objects that can be detected by stereo vision are limited, thus greatly reducing the matching search area.

II. B. 2) Principles of regional stereo matching

The region stereo matching algorithm judges feature points based on their neighborhood gray scale correlation, which works well in the texture information-rich part, but has a high rate of mis-matching in the low texture and occlusion part. For a feature point in the reference image, a sliding window is established with it as the center, and a window of the same size is established by searching on the epipolar line of the corresponding image, using the neighborhood gray scale of the feature point to characterize its properties. Through the appropriate similarity measure function, the difference of the corresponding pixels in the superimposed sliding window is counted to determine the degree of correlation of the corresponding points, and the image point with the highest correlation within the search range is selected as the best match. Figure 2 shows the principle of region matching.

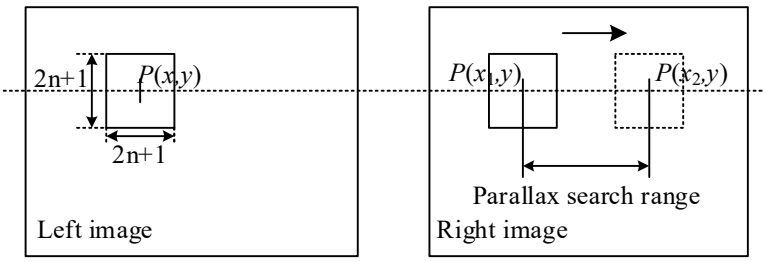

Figure 2: Principle of Regional Matching

For the corrected left and right images, their pixel rows are strictly aligned and co-linear with the epipolar line. Take the left image as the reference, set the image point $P(x,y)$ as the point to be matched, construct a sliding window of $(2n+1)*(2n+1)$ with P as the center, and search for the matching point within the same row in the right image, and set the search range as $x_1 \sim x_2$. Calculate the correlation between $P(x,y)$ and $P(x_i,y)(x_1 < x_i < x_2)$, again constructing a window of the same size centered on $P(x_i,y)$. A similarity measure function is established as a primitive for computation, and this result is called the original matching cost. Here the absolute value of the gray difference AD is used as the matching operator, and in practice a gray threshold can be set to reduce the interference of mis-matching in the superposition process, as shown in equation (1).

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

For the matching window built with $P(x,y)$ and $P(x_i,y)$ as the center, the original matching cost of the corresponding pixel in the window is accumulated to get the matching cost sum characterizing the pixel correlation as shown in equation (2).

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

where i, j are the offsets of the coordinates; d is the size of the parallax to be matched; $C(x,y,d)$ is the superimposed matching cost.

After obtaining all the superimposed matching costs in the search region $x_1 \sim x_2$, the pixel corresponding to the smallest value is the best matching point. As can be seen from the matching principle, region matching for each point needs to calculate the superposition matching cost, the calculation is very large, and only consider the correlation of pixels in a limited area around the point, so the accuracy is limited, but its advantage is a simple structure, easy to improve in order to realize in real time.

Most of the stereo matching algorithms firstly get the result of discrete parallax, which can meet the usage requirements when there is no strict stipulation on the accuracy, but the integer parallax can't meet the needs in the applications with strict accuracy requirements. After obtaining the original parallax, it is necessary to take corresponding methods, such as curve fitting, image filtering, etc. to interpolate the integer parallax to achieve sub-pixel accuracy.

II. C.Motion Target Tracking with SURF-KLT

In this paper, the SURF algorithm is firstly applied to extract feature points with high robustness, then the KLT matching algorithm is applied to stabilize the tracking of the feature points in the subsequent frames, and finally the Greedy algorithm is combined to realize the final position of the target. Figure 3 shows the tracking flow.

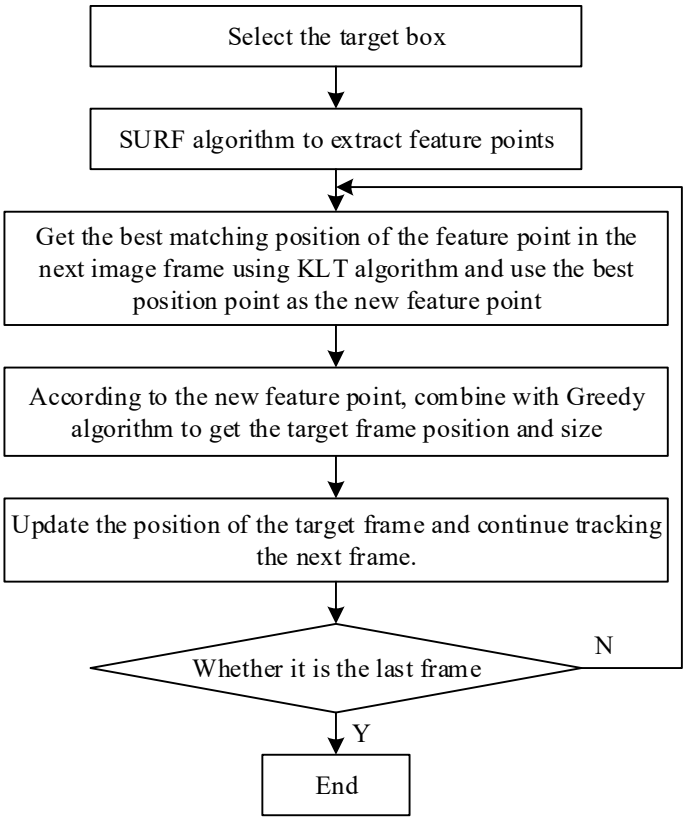

Figure 3: Target tracking Process

Accurately extracting feature points is the basis for realizing target tracking. In this paper, the SURF algorithm, which can extract robust feature points quickly, is adopted to extract feature points, the general principle of which is to establish an integral image and construct a scale space, locate the feature points by means of Hessian matrix and non-maximal value suppression, and generate the corresponding feature point descriptors. Given a point (x,y) in the integral image $f(x,y)$, its Hessian matrix is defined in the following equation:

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

In practice $L(x,y,\sigma)$ is approximated using a box filter, which replaces the convolution template with a simple rectangle to speed up the computation of the convolution.The determinant of the Hessian matrix can be approximated as shown in equation (4):

$$\det\left(H_{approx}\right)=D_{xx}D_{yy}-\left(0.9D_{xy}\right)^2 \tag{4}$$

To represent the corner point response value at point x. All pixel points in the image are traversed to form a response image for feature point detection at a given scale. Each sampled point is then compared with its neighboring points for non-maximal value suppression in the stereo neighborhood of $4\times4\times4$. Finally, interpolation operations are performed in the scale space and image space to obtain stable feature point locations and scale values.

After extracting the appropriate amount of robust feature points using SURF algorithm, just to get their positions and corner response values, this paper uses the less time-consuming KLT algorithm for feature matching.KLT algorithm is a tracking algorithm based on optimal estimation, using the sum of squared gray differences (SSD) between image frames as a metric. The method performs optimal estimation by the similarity metric between the local information of the small window around each feature point, which leads to the best matching position. $I(x,y,t)$ denotes the image frame at the moment of t, and $I\ (x,y,t+dt)$ denotes the image frame at the moment of $t+dt$, and their positions satisfy the following relation:

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

The motion quantity $d=(dx,dy)$ is called the offset of the point $X(x,y)$. The KLT algorithm is the one that transforms the solution of the motion parameter into finding the d that minimizes the SSD function (denoted by ε) when it can be minimized, which also finds the new position $(x+dx,y+dy)$.

Considering the general case, there Eq. (6) holds:

$$B(X+d)=A(X)+n(X) \tag{6}$$

$n(X)$ is the noise due to illumination changes in time dt. Squaring $n(X)$ and integrating over the whole window gives the SSD function of the window image:

$$\varepsilon=\iint_W[n(X)]^2\omega(X)dX=\iint_W\left[B(X+d)-A(X)\right]^2\omega(X)dX \tag{7}$$

where $\omega(X)$ is the weighting function, which is usually taken to be 1. In general, d is negligibly small compared to X, so $B(X+d)$ can be Taylor-expanded and only the linear term is intercepted, and then d is derived according to Eq. (6), which is ultimately obtained by simplification:

$$Zd=e \tag{8}$$

where $Z=\iint_W g(X)g^T(X)dX$, $g(X)=[g_x\ g_y]^T$, where g_x and g_y are the modal values of the gradient in the x, y directions respectively. Newton iterations are performed for each point using Eq. (8) until a certain accuracy is satisfied the image point tracking can be realized, and the optimal matching is found to find the final solution:

$$d_{k+1}=d_k+Z^{-1}\cdot\left[\iint_W[A(X)-B(X)]g(X)\omega(X)dX\right] \tag{9}$$

where d denotes the translation of the center of the feature window, and d_k denotes the value of d calculated by the k th Newton iteration method.

After obtaining the position information of a number of feature points after correct matching, it is also necessary to determine the optimal position and shape of the tracked target from this in order to realize the stable and reliable tracking of the target. Firstly, according to the position distribution of the optimal feature points obtained by KLT matching, the position of the rectangular box containing all feature points is determined, which is the width w and

height h of the target area. In order to accurately reflect the actual size of the target, Greedy algorithm is used to realize the target position determination.

III. Analysis of the effectiveness of robot cleaning and maintenance based on tracking matching algorithm

III. A. Algorithm Performance Analysis

III. A. 1) Algorithm Matching Error Detection

After the comprehensive introduction of region stereo matching algorithms such as SURF-KLT to optimize the region matching level and motion target tracking ability of the image processing module, several performance metrics are needed to evaluate the algorithm's relevant performance, and the commonly used classification performance metrics are the accuracy rate, the recall rate and the F1 value. In this paper, the algorithm performance level is mainly evaluated by confusion matrix, PR curve and other indicators.

A dataset containing 5000 images of common garbage in smart houses is chosen as the dataset for the study of this paper. 1/5 of them are randomly selected as test images to test the matching performance of the algorithm. Figure 4 presents the algorithm matching results through the confusion matrix. The rows of the confusion matrix represent the categories predicted by the algorithm, and the columns represent the true categories. The diagonal values indicate the number of correctly matched spam images, while the off-diagonal values indicate the number of misclassified spam images. Out of 1000 spam images matched, 997 images were correctly matched to 7 categories and only 3 images were misclassified to other categories. As far as the test data is concerned, the matching accuracy is as high as 99.7%, which proves that the robot's matching accuracy for residential garbage has improved after the introduction of the matching algorithm to improve the image processing module.

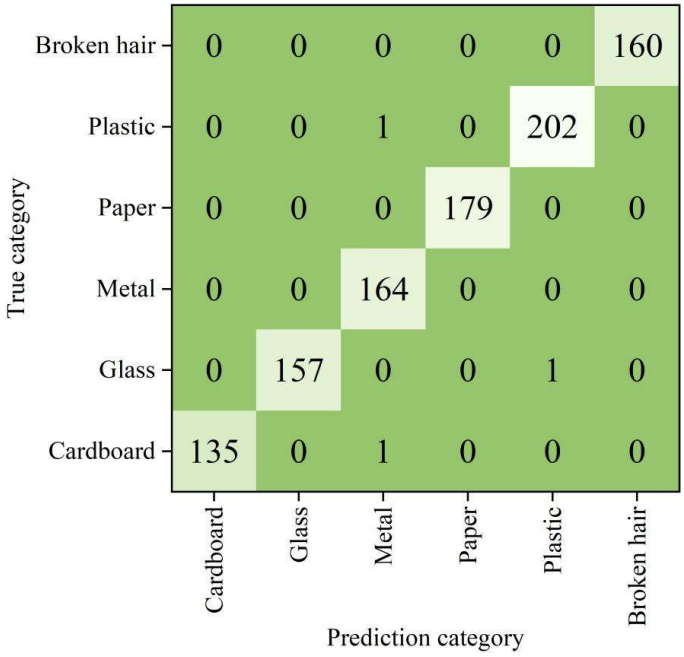

Figure 4: Algorithm matching result

III. A. 2) Performance comparison before and after algorithm introduction

Since this paper introduces several algorithms to comprehensively optimize the image processing module, this section will gradually analyze the optimization effect of the four types of algorithms on the image processing model by introducing only the SURF algorithm, introducing the SURF-KLT algorithm, introducing the SURF-KLT and combining with the Greedy algorithm, and introducing the SURF-KLT and combining with the Greedy algorithm and stereo matching constraints, to judge whether this paper's optimization method is effective. Continuing through the random 2/5 images in the dataset constructed in the previous paper as the training set, Figure 5 shows the mAP-0.5 curves of different algorithms. After 150 iterations, the mAP-0.5 value of introducing only SURF algorithm is 0.755, the mAP-0.5 value of introducing SURF-KLT algorithm is 0.814, the mAP-0.5 value of introducing SURF-KLT algorithm is 0.893, and the mAP-0.5 value of introducing SURF-KLT and combining Greedy algorithm and stereo matching constraint is 0.932. Since 0.755<0.814<0.893<0.932, the combined use of various types of algorithms to

achieve target tracking and the addition of the regional stereo matching constraints results in a large increase in the robot's image processing capability, which can satisfy the accuracy requirements for trash matching and cleaning.

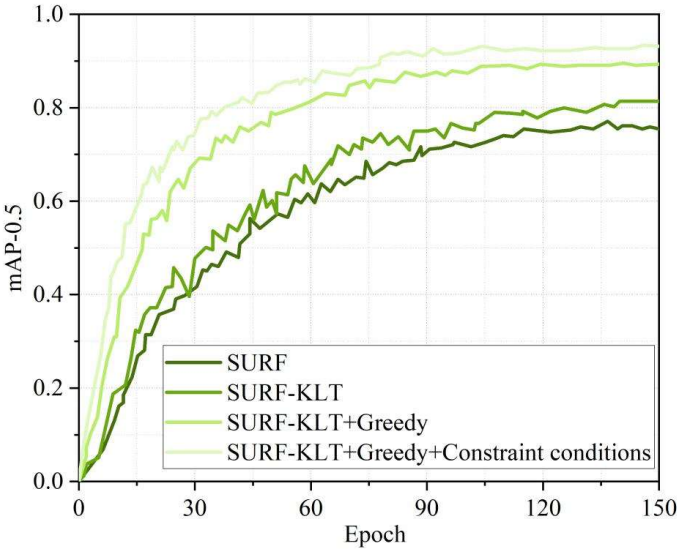

Figure 5: mAP-0.5 curves of different algorithms

Fig. 6 shows the fluctuation of precision and recall of this paper's method with SURF-KLT+Greedy and stereo matching constraints in the training set. Smaller fluctuations indicate better training results. The training precision and recall of this paper's method increases steadily, and basically stabilizes between 0.85-0.87 around 120 times, with no obvious fluctuation throughout the training, which is a good performance.

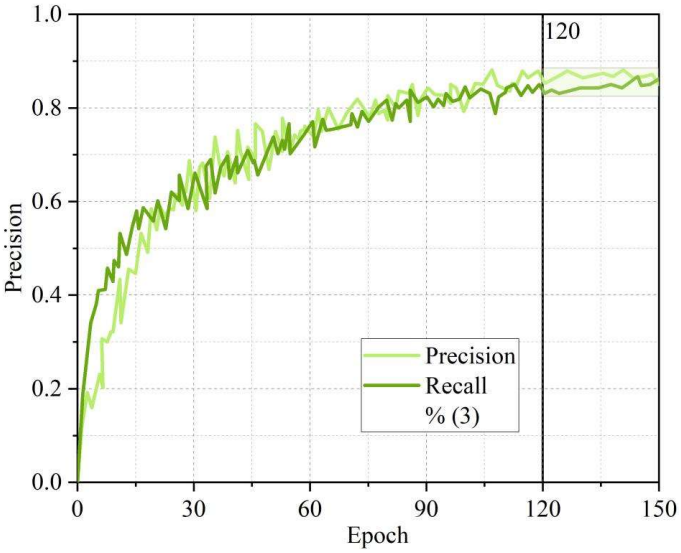

Figure 6: Fluctuations of Precision and Recall of paper method

III. B. Robotic cleaning practices

III. B. 1) Cleanliness assessment

The trained algorithm is applied to the image processing module of the robotic cleaning system for residential cleaning practice. A 70m² sized residence was used as the overall cleaning area and the robot was placed to dynamically categorize and match the seven types of garbage present in the residence. According to the waste classification method, cardboard and paper were categorized as organic waste, metal and plastic as inorganic waste, shredded hair as uncollected sweeping waste, and glass as solid residue. Weights were calculated for each type of waste to weight the overall cleanliness index. When the cleanliness index is higher than 20, the robot initiates path planning for spot cleaning of garbage and categorization of different garbage according to the type of garbage and

the location of the area where the garbage is located. For plastic, metal and paper trash can be cleaned directly by sweeping, broken hair needs to be removed by suction, while solid residues such as glass are directly picked up physically. Table 1 shows the amount of trash in the residence and its weighted total value. As detected by the robot, a total of 23 pieces of trash were present in the experimental residence, with a total weighted value of 39, which is much larger than 20 and requires deep cleaning.

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

III. B. 2) Optimal path planning

The robot determines that the residence needs to be cleaned based on matching the type of trash detected. And combined with the dynamic tracking results, the optimal reference path is planned to complete the cleaning task efficiently and accurately, and reduce the possibility of colliding with furniture, corners, etc. Fig. 7 shows the optimal cleaning path obtained by planning. The robot starts from the position of (0,0) and moves to the right to the position of (25,5) to suction the uncollected cleaning waste (broken hair); then it moves to the positions of (20,30) and (20,50) to clean the inorganic waste (metal, plastic) in the 2 zones; then it moves to the position of (30,50) to pick up the solid residue (glass); then it moves to the position of (50,40) to sweep organic waste (cardboard, paper). After cleaning, the robot moves to the position (70,70) for hibernation. From the optimal cleaning path, the robot needs to reach fewer areas repeatedly, and can complete residential cleaning and maintenance in the fastest way in the least amount of time. This also shows that the algorithm in this paper is effective in optimizing the image processing module.

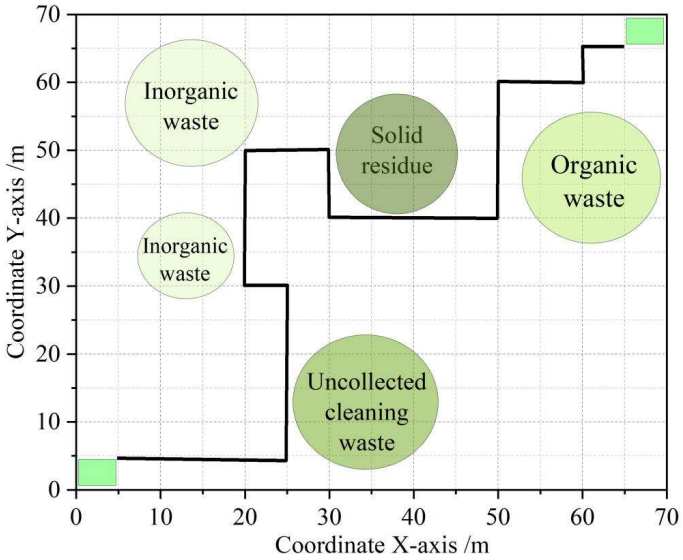

Figure 7: Optimal cleaning path

IV. Conclusion

In this paper, we optimize the robot image recognition capability through the fusion of multi-stereo matching algorithms to significantly improve the environment perception and task execution quality of residential cleaning robots. The correct matching rate is as high as 99.7% in 1000 litter image matching. The mAP-0.5 value of the multi-

algorithm fusion in this paper reaches 0.932, and the precision and recall are stabilized in the range of 0.85-0.87 after 120 iterations, which meets the precision requirements for trash image matching and cleaning. It accurately locates 23 areas where garbage is located in a 70m²-sized residence, and plans an efficient cleaning path by calculating the cleaning weights to determine the need for deep cleaning. In the future, cleaning experiments can be carried out in residences of larger size to judge the upper limit of robotic cleaning and maintenance assisted by the method of this paper, and to find the possibility of breakthrough.

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.