Target Segmentation and Attitude Positioning Based on Depth Information

Yi Zhang^{1, a}, Jianchun Dai^{1, b} and Han Zhang^{2, c}

¹School of Advanced Manufacturing Engineering, Chongqing University of Posts and Telecommunications University, Chongqing 400065, China;

²School of automation, Chongqing University of Posts and Telecommunications University, Chongqing 400065, China.

^a657508406@qq.com, ^b1414529969@qq.com, ^c363997565@qq.com

Abstract

Target localization is an important prerequisite for robot to achieve accurate target grasping. In the process of robot grasping, there is a variety of target postures, and the robot needs to adjust the attitude of the end gripper continuously. In order to achieve accurate target grasping, a target pose localization method based on Kinect depth image information is proposed. Firstly, the target two valued image is detected in the experimental space based on the depth threshold method. Secondly, the mathematical morphology combined with Canny edge detection algorithm is proposed to extract the target edge, and the target edge is mapped to the 3D coordinate system according to the depth information. Finally, the discrete points of target edge in the 3D coordinate system are fitted in plane, and the direction vectors and the normal vectors of the target center point are selected to describe the target's attitude. The experimental results show that the method can realize the 3D location of target center, while achieving the target attitude description with small positioning error and full information.

Keywords

Depth Data, Edge Detection, Space Plane Fitting, Attitude Positioning.

1. Introduction

Robot grabbing is the most basic function of robot. It requires that we can achieve target grasping operation. Target localization is an important prerequisite for robot to achieve accurate grasp operation^[1]. Target localization based on vision is one of the key technologies of robot localization crawl, which aims to improve the autonomy and adaptability of the machine, which has wide application prospect in modern industrial, military and other fields, so the study of visual localization based on has a very important significance ^[2].

At present, target positioning technology based on vision mainly uses monocular vision, binocular vision and Kinect. For example, Zhisheng, Zhang and other^[3] propose an absolute location estimation method based on mapping relationship between projection points and their corresponding points in monocular vision system, and combine with error correction algorithm to achieve target location. A. SHARMA and other^[4] proposed a target recognition and location method of the scale invariant feature matrix transform (SIFT) based on the monocular vision. Zhang Yuchen and other^[5] uses a gray correlation template matching algorithm to achieve the target location under the binocular vision system. Gao Ruxin and other^[6] proposed a method of target recognition and location in binocular vision system combined with SURF features. Samia Nefti-Mezianideng and other^[7] use binocular vision to build three-dimensional perception for humanoid robot NAO.

However, the above two kinds of visual positioning methods: monocular vision system has the advantages of simple structure and fast speed, but because of the use of a single camera, the system has a small field of vision, less depth of information and low positioning accuracy. Binocular vision increases information acquisition and improves positioning accuracy, but needs to be calibrated,

complicated and complicated. As a low-cost peripheral device with deep camera, Kinect can directly get the scene depth value. In recent years, it has been widely applied in target vision location by researchers at home and abroad. For example, Dong Jianmin^[8] et al. combined with the color images and the depth image information obtained from Kinect to realize the location of tomatoes to achieve the purpose of automatic harvesting of tomatoes. Figueroa^[9] et al. proposed the use of Kinect for target recognition and positioning, which is used to assist the robotic arm to grab objects. M. Clark and other^[10] calculate the moment information through the contour extracted by Kinect, and realize the centroid positioning of the target according to the projection formula of the camera. Zhang Chen and other^[11] realized the accurate positioning of the point based on the Kinect depth information with the help of human hand instructions. The target objects of Dong Jianmin and Figueroa's research is a sphere or sphere, it is only a geometric center 3D positioning for the target, and there is no problem of positioning the target. However, the object of literature [10] and [11] is non sphere, but the attitude of the target in space is not taken into account in the paper.

Therefore, aiming at the problem of uncertain target attitude in visual location based on Kinect, this paper presents a method of Kinect target attitude positioning based on depth information, using mathematical morphology to extract the target value of two image edge combined with Canny edge detection algorithm, combined with 3D positioning and attitude description goal orientation. The experimental results show that the method can describe the spatial attitude of the target at the same time that the target center point 3D is located.

2. Target segmentation based on depth threshold

The Kinect camera includes a color VGA camera and a depth sensor that gets the scene depth image data while obtaining the scene color images. The depth data stream is made up of deep image frames. Each pixel in the depth image contains a specific distance information from the perspective of the Kinect infrared camera, the distance from the nearest plane to the plane at the particular point coordinates (unit: mm). The depth data can be obtained by calling API. According to the depth value, the experimental space is set between distances and points. If and only when objects appear in the experimental space area, it will be detected to prevent the interference of complex background and foreground. The experimental space is shown in Figure 1.



Fig. 1 Experimental space

In the experimental space, the depth value of each pixel is greater than d_1 and less than d_2 , according to the API provided by kinect SDK set the RGB value of the pixel point of the image is (255,255,255), namely, the pixel is white; and when the depth value is not in the experimental space, set the image pixel RGB (0,0,0), which is the the pixel is black. Through the above operation steps, the two value image of the target can be detected, which is expressed in matrix. At the same time, according to the data structure provided by the deep data flow, the high 13 bits of each pixel represent the depth value of the pixel point, and the corresponding relation between each pixel and depth value is represented by matrix.

3. Target attitude location

The attitude orientation of the target is first to detect the edge of the two value image of the target. This paper uses mathematical morphology and Canny edge detection algorithm to get the edge matrix A' of the target two value image. On this basis, we use the equal interval search method to get the minimum outer moment of the target edge, the smallest outer part of the moment in the outer cutting moment, and then determine the center coordinates O'(x', y') of the target in the two valued image by the four vertices of the outer cutting moment. Combining the matrix depth matrix B and A' matrix edge matrix, the target edge is mapped to the three-dimensional space coordinate system according to the relationship between the pixel depth value. The discrete point of the edge space is fitted by plane, and the minimum external moment and center point are projected to the plane, and the target 3D location O''(x', y', z) is finally determined. The plane normal vector and the internal vector of the plane which through the point O'' are used to describe the attitude of the target. In combination with 3D positioning and attitude description, the attitude orientation of the target is finally realized.



Fig. 2 System flowchart

3.1 Two valued morphological filters

Image filtering is the noise suppression of the target image, which is an indispensable operation for image preprocessing. The commonly used image filters include mean filter, median filter, Gauss filter, morphological filter and so on. The mean filter and median filter take the mean or median value instead of the original pixel value in the template. Gauss filter uses the weighted average of the pixel value itself and the field to replace the original pixel value, resulting in edge blur at the same time of denoising. The two valued morphological filter in morphological filter uses the basic corrosion and expansion operation to preprocess the image, which does not change the two value of the original image, and is simple and effective for the two value image processing.

The mathematical morphology method is a nonlinear filtering method, morphological filter by selecting image features smaller structural elements and digital image interaction to achieve, according to different purposes can choose different types of structural elements, the size and shape of the corresponding morphological transform, is widely used in image denoising, image segmentation, edge detection etc. [12]. The morphology of two valued images is the basis of mathematical morphology. It has four basic transformations: expansion (Dilation), corrosion (Erosion), open (Open), and closed (Close).

Expansion: refers to the convolution of image p and core e (structural elements) to find the local maximum value, that is, to calculate the maximum value of pixels in the area covered by kernel E, and assign the maximum value to the pixel designated by reference point. Let z be a translation amount,

the symmetric set of *E* on the origin is $\hat{E} = \{-b : b \in E\}$, the expansion of *P* by *E* is recorded as $P \oplus E$, and \oplus is the expansion operator, then the set definition of the expansion is:

$$P \oplus E = \left\{ z \mid \left[\left(\hat{E} \right)_{z} \cap P \right] \neq \Phi \right\}$$
(1)

The expansion of the expression is the first of the *E* to map the far point and then to move the *z*. The expansion of *P* by *E* is that the *P* has at least one common non zero element after all \hat{E} translates to *z*. The expansion process is shown in Figure 3.



Fig. 3 Schematic diagram of expansion process

Corrosion: it can be seen as an inflated inverse or reverse process. The corrosion operation is to move the structural elements from left to right, from top to bottom in the image to be processed, and take the central point of the structural element as the central point of the operation to check whether the pixels around the image coincide completely with the structural elements. If the detection does not overlap fully, the center point pixel is marked 0. The corrosion of *P* by *E* is recorded as $P\Theta E$, and the set definition of the corrosion is as follows:

$$P\Theta E = \left\{ z \mid (E)_z \subseteq P \right\}$$
⁽²⁾

The result of the corrosion of the *P* by *E* is the set of point *z* that has been included in the P after the *E* is moved to *z*, and the diagram of the corrosion process is shown in Figure 4.



Fig. 4 Schematic diagram of corrosion process

Open and Close: set *P* is the original image, *E* is the structure element image, then the set *P* is open operation by the structure element *E*, which is recorded as $P \circ E$, and its formula is defined as:

$$P \circ E = (P \Theta E) \oplus E \tag{3}$$

From equation (3) can be seen, the open operation of P by E is the corrosion of P by E and then expansion, is a collection of E within the P translation by the union, defined by the set theory is expressed as follows:

$$P \circ E = \bigcup \{ (E)_z \mid (E)_z \subseteq P \}$$

$$\tag{4}$$

The set *P* is closed operation by the structural element *E*, and is recorded as $P \bullet E$, and its formula is defined as:

$$P \bullet E = (P \oplus E) \Theta E \tag{5}$$

From the formula (5), it can be seen that the close operation of P by E is the expansion of P by E and then corrosion, The comparison diagram of the effect of open and closed operations is shown in Figure 5.

Opening operation can smooth the image contour, weaken the narrow part, remove fine highlights. The closed operation is contrary to the open operation. It can fill the narrow and concave parts in the background, fill holes and fill gaps in the outline, so that the contour of the image is smoothed. Therefore, based on the characteristics of two valued morphological open and close operation with good two valued image filtering, a two valued morphological filter is proposed for image preprocessing.



Fig. 5 Contrast diagram of open and close operation

3.2 Edge detection based on two value morphological filter

The common edge detection operators are operators such as Robert, Laplace, Sobel and so on. These operators are simple in structure and fast in implementation, but are sensitive to noise and poor in adaptability. Canny operator is an edge detection operator based on optimization algorithm. It outperforms other traditional edge detection operators in dealing with Gauss white noise. However, the traditional Canny operator uses Gauss filtering to de noise, which will cause edge blur at the same time of noise elimination. In this paper, two value morphological filters are used to replace the traditional Gauss filter, and the edge detection algorithm is combined with the Canny edge detection algorithm to extract the edge. The steps of the edge detection are as follows:

(1) the two value morphological filter is used to smooth the two value images of targets, and the noise is suppressed and the holes are filled;

(2) calculate the gradient and direction of the image after the preprocessing;

(3) the non maximum suppression of the gradient is carried out.

(4) an iterative algorithm is used to find the best high and low threshold.

(5) the double threshold algorithm is used to detect and connect the edges.

(6) the edge extraction results are saved into the A' matrix.

Through the above steps, the edge matrix of the target in the two value image can be obtained. The equal interval search method is applied to detect the minimum external moment in the outer edge of the target edge. The four vertex coordinates $P_1(x_1, y_1)$, $P_2(x_2, y_2)$, $P_2(x_2, y_2)$ and $P_4(x_4, y_4)$ of the minimum outer moment can be determined, and the center coordinates O'(x', y') of the target in the two valued image can be determined. Combined with the edge matrix A' and the depth information matrix B, the target edge is 3D mapped, and the 3D point cloud matrix A'' of the target edge is obtained. In the 3D coordinate system, the set of discrete points of the target edge can be obtained.

3.3 Target 3D location and attitude description

Through the edge detection and three-dimensional mapping operation, we can get the object's three-dimensional point cloud matrix A''. In order to further determine the target's 3D location and correctly describe the target's posture, we choose the target edge space discrete point fitting plane as the further research object. The advantage of this research object is that it can reflect the plane at the edge of the target well, and it can be used as the projection plane of the center point of the two valued image and the smallest outer moment vertex, and the computation is small and the efficiency is high.

Because the discrete points of the target edge are not in the same plane, in order to further study the processing, this paper does the plane fitting of the edge discrete points. The plane fitting process is as follows: set the plane equation as 0 = ax+by+cz+d, calculate the *x*, *y*, *z* mean of the edge point:

$$\overline{x} = \sum_{i=1}^{n} x_i / n, \, \overline{y} = \sum_{i=1}^{n} x_i / n, \, \overline{z} = \sum_{i=1}^{n} x_i / n$$
(9)

Then the variance of x, y, z and their cross covariance are calculated:

$$\begin{cases} S_{x} = \sum_{i=1}^{n} (x_{i} - \overline{x})^{2} / n & \begin{cases} S_{xy} = \sum_{i=1}^{n} x_{i} y_{i} / n - \overline{x} \ \overline{y} \\ S_{y} = \sum_{i=1}^{n} (y_{i} - \overline{y})^{2} / n & \end{cases} \begin{cases} S_{yz} = \sum_{i=1}^{n} y_{i} z_{i} / n - \overline{y} \ \overline{z} \\ S_{zx} = \sum_{i=1}^{n} (z_{i} - \overline{z})^{2} / n & \end{cases} \begin{cases} S_{zx} = \sum_{i=1}^{n} z_{i} x_{i} / n - \overline{z} \ \overline{x} \end{cases}$$
(10)

Let the matrix T be the covariance matrix of x, y, z, then it is related:

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{S}_{x} & \boldsymbol{S}_{xy} & \boldsymbol{S}_{zx} \\ \boldsymbol{S}_{xy} & \boldsymbol{S}_{y} & \boldsymbol{S}_{yz} \\ \boldsymbol{S}_{zx} & \boldsymbol{S}_{yz} & \boldsymbol{S}_{z} \end{bmatrix}$$
(11)

The eigenvectors corresponding to the minimum eigenvalue of the matrix T are the coefficients of the plane equation, and the eigenvalues obtained by the matrix T are A, B and C, and then $D = d = -(A\overline{x} + B\overline{y} + C\overline{z})$, so that the fitted plane equation can be obtained:

$$0 = Ax + By + Cz + D \tag{12}$$

The 2D center point of the target is located at the projection point of the fitting plane to achieve the 3D positioning of the target. Meanwhile, the plane normal vector \vec{n} and the target radial vector \vec{r} are obtained, and the space attitude positioning of the target is realized by combining the 3D positioning and the gesture description.

4. Experiment and result analysis

In the experiment, Kinect depth data format DepthImage- Format.Re- solution640x480Fps30 is set, namely m=640, n=480, the experimental space is set to $d_1 = 800mm$ and $d_2 = 1500mm$, the experimental environment is Windows7 64 bit operating system, I3 processor and 4GB RAM, using Matlab R2012b for experimental simulation.Select a cross shaped element with a size of 3×3, and use two consecutive two valued morphological closing operations to preprocess two value images. The results of each step are shown in Figure 6.



4.1 Target 3D positioning experiment

Ten spatial location 3D positioning experiments are carried out on the target model by using this method and the location method in the literature [13]. And ten coordinate records for each space position. The Kinect depth camera is the origin of the coordinate, the horizontal direction is the X axis, in the camera and perpendicular to the Y axis, perpendicular to the plane direction of the Kinect imaging plane is the Z axis. The experimental results are shown in table 1.

Tab T Experimental results of target 5D positioning									
No.	Actual coordinates(cm)			This method(cm)			Document [13](cm)		
	x	у	z	x	у	z	x	у	z
1	12.3	-5.8	82.3	13.2	-6.1	83.1	12.9	-5.2	82.5
2	20.1	17.6	108.2	19.6	17.4	108.8	20.4	17.2	107.5
3	-8.6	14.5	96.4	-9.1	14.1	97.1	-9.6	13.8	96.1
4	-16.7	9.2	112.8	-16.3	9.5	112.5	-17.4	10.3	112.2
5	23.7	-8.7	93.7	23.0	-8.3	93.1	24.5	-8.5	94.5
6	9.4	14.8	126.4	9.6	14.2	126.2	9.8	14.5	125.3
7	16.7	18.5	137.2	17.7	17.9	137.6	16.2	19.1	136.5
8	4.1	-13.8	100.5	3.8	-13.0	101.0	4.3	-13.7	100.1
9	-6.7	10.4	89.3	-7.2	10.7	88.3	-7.0	11.1	90.5
10	13.2	5.7	143.1	13.5	6.1	142.5	12.5	5.2	142.3

Tab 1 Experimental results of target 3D positioning

Analysis of 3D localization experiment results, in the 10 experiment, the maximum error of this method is 1cm, and the variance of the absolute error of each coordinate value is 0.335, 0.215, 0.374, respectively, and the maximum error is 1.2cm in document[13], the coordinate values of the absolute error variance were 0.353, 0.346, 0.498. Based on the above analysis, the location accuracy of 3D positioning in this paper is better than that of document [13], especially in target depth detection, which is more stable than [13]. The positioning is stable and there is no singular point.

4.2 Attitude positioning experiment

In the experiment, 12 sets of specific target models are drawn up. The attitude of the target is described by this method. The experimental results are shown in Table 2. From the comparison of the experimental data, it can be seen that the error of the other values is less than 0.06, except that the individual error is within 0.16.

NO.	Actual target attitude (\vec{n}, \vec{r})	Experimental target attitude (\vec{n}, \vec{r})					
1	(0,0,1),(1,0,0)	(-0.0391, 0.1053, 0.9937), (0.9856, -0.1597, 0.0557)					
2	(0,0,1),(0.71,0.71,0)	(0.1103, 0.0930, 0.9895), (0.6981, 0.7014, -0.1437)					
3	(0,0,1),(0,1,0)	(-0.0752, 0.1361, 0.9878), (0.0253, 0.9906, -0.1346)					
4	(0,0,1), (-0.71,0.71,0)	(-0.0873, -0.1089, 0.9902), (-0.7021, 0.7119, 0.0164)					
5	(-0.71, 0, 0.71), (0.71, 0, 0.71)	(-0.7274, -0.0610, 0.6835), (0.6982, -0.0658, 0.7372)					
6	(-0.5, -0.5, 0.71), (0.5, 0.5, 0.71)	(-0.5372, -0.4958, 0.6823), (0.4920, 0.4729, 0.7310)					
7	(0, -0.71, 0.71), (0, 0.71, 0.71)	(0.1051,-0.6893,0.7168),(-0.0518,0.7160,0.6961)					
8	(0.5, -0.5, 0.71), (-0.5, 0.5, 0.71)	(0.4897, -0.4903, 0.7210), (-0.5207, 0.4988, 0.6929)					
9	(0.71, 0, 0.71), (0.71, 0, -0.71)	(0.6915, 0.0200, 0.7221), (0.7411, -0.0196, -0.7092)					
10	(0.5, 0.5, 0.71), (0.5, 0.5, -0.71)	(0.4812, 0.4793, 0.7340), (0.5384, 0.4905, -0.6853)					
11	(0, 0.71, 0.71), (0, 0.71, -0.71)	(0.0681, 0.6900, 0.7206), (-0.0603, 0.7238, -0.6874)					
12	(-0.5,0.5,0.71),(-0.5,0.5,-0.71)	(-0.5044, 0.5301, 0.6816), (-0.4809, 0.4832, -0.7316)					

Table 2 Experimental results of target attitude orientation

5. Conclusion

This paper presents a Kinect target attitude positioning method based on depth image information acquisition, preliminary target of two value image through the depth threshold is set, the two value object edge extraction method combined with Canny morphological edge detection, edge generation of discrete points in three-dimensional coordinates and plane fitting of the target center point 3D positioning in the projection plane fitting, selection of vector and description of the target attitude. The experimental results show that the localization method achieves the 3D location of the target, and the location error is 1cm. The location is stable and no singularity. At the same time, the target pose is described. In order to improve the positioning accuracy of 3D and the accuracy of attitude description, the error correction will be the next research focus because of the measurement error of Kinect hardware itself.

Acknowledgements

Research and application project of automatic palletizing (cage) system for cigarette sorting package robot (YCYC-17-HT-2015232)

References

- [1] K. Wang, Y. H. Liu and L. Li, "A Simple and Parallel Algorithm for Real-Time Robot Localization by Fusing Monocular Vision and Odometry/AHRS Sensors," in IEEE/ASME Transactions on Mechatronics, vol. 19, no. 4, pp. 1447-1457, Aug. 2014.
- [2] Zhang Zhiguo. "Research on Positioning System Based on Monocular Vision" [D]. Huazhong University of Science and Technology, 2009.
- [3] R. G. Lins, S. N. Givigi and P. R. G. Kurka, "Vision-Based Measurement for Localization of Objects in 3-D for Robotic Applications," in IEEE Transactions on Instrumentation and Measurement, vol. 64, no. 11, pp. 2950-2958, Nov. 2015.
- [4] C. Liu, H. Qiao, J. Su and P. Zhang, "Vision-Based 3-D Grasping of 3-D Objects With a Simple 2-D Gripper," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 44, no. 5, pp. 605-620, May 2014.
- [5] J. Han, L. Shao, D. Xu and J. Shotton, "Enhanced Computer Vision With Microsoft Kinect Sensor: A Review,"[J] in IEEE Transactions on Cybernetics, vol. 43, no. 5, pp. 1318-1334, Oct. 2013.
- [6] Z. Zhang, "Microsoft Kinect Sensor and Its Effect,"[J] in IEEE MultiMedia, vol. 19, no. 2, pp. 4-10, Feb. 2012.
- [7] Guo B,Sun J,Wei Y,et al. "Kinect identity:Technology and experience," [J], Computer, 2011,44(4):94-96.
- [8] Dong Jianmin, Chen Weihai, Yue Haosong, et al. "Automatic Recognition and Localization of Tomatoes Based on Kinect Vision System" [J]. Chinese Journal of agricultural chemistry, 2014, 35 (4):169-173.
- [9] Figueroa J,Contreras L,Pacheco A,et al.Development of an Object Recognition and Location System Using the Microsof KinectTM Sensor[M]/RoboCup 2011:Robot Soccer World Cup XV.2012:440-449.
- [10] M. Clark, D. Feldpausch and G. S. Tewolde, "Microsoft kinect sensor for real-time color tracking robot," IEEE International Conference on Electro/Information Technology, Milwaukee, WI, 2014, pp. 416-421.
- [11]Zhang Chen, Zhang Zhao, Lv Zhaohui. "Target Location of Hand Indication Based on Kinect Depth Data," [J]. Journal of Communication University of China (NATURAL SCIENCE EDITION), 2016, (03): 46-50.
- [12] Wang Xiaojun, Liu Xumin, Guan Yong. "Image Edge Detection Algorithm Based on Improved Canny Operator," [J]. Computer Engineering, 2012,38 (14): 196-198+202. [2017-10-03].

- [13] Duan Hongyan, Shao Hao, Zhang Shuzhen, Zhang Xiaoyu, Wang Xiaohong. "An Improved Algorithm for Image Edge Detection Based on Canny Operator," [J]. Journal of Shanghai Jiao Tong University, 2016,50 (12): 1861-1865.
- [14] Zhou Zhen, Du Shanshan. "Target Location and Recognition Based on Kinect Depth Image," [J]. mechanical manufacturing and automation, 2016, (04): 173-176.
- [15] P. Dollár and C. L. Zitnick, "Fast Edge Detection Using Structured Forests," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 8, pp. 1558-1570, Aug. 1 2015.
- [16] T. M. Iversen and D. Kraft, "Generation of synthetic Kinect depth images based on empirical noise model," in *Electronics Letters*, vol. 53, no. 13, pp. 856-858, 6 22 2017.
- [17] Lin Haibo, Zhou Ting, Zhang Yi. "Unmarked Gesture Segmentation and 3D Localization Based on Depth Information," [J]. Journal of Huazhong University of Science and Technology (NATURAL SCIENCE EDITION), 2015, (S1): 62-65.
- [18]C. Lee, W. Chao, S. Lee, J. Hone, A. Molnar and S. H. Hong, "A Low-Power Edge Detection Image Sensor Based on Parallel Digital Pulse Computation," in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 62, no. 11, pp. 1043-1047, Nov. 2015.