ISSN: 2664-9640

RANSAC Circle Fitting Algorithm Based on K-means Clustering

Haohan Wu[1, 2, 3, 4, a, *], Dong Li[1, 2, 3, 4, b] and Xiaoshi Shi[1, 2, 3, 4, c]

[1]Automation and Information Engineering, Sichuan University of Science & Engineering Zigong, Sichuan, 643000, China,

[2]Artificial intelligence key laboratory of Sichuan province, Zigong, Sichuan, 643000, China,

[3]Enterprise informatization and Internet of things measurement and control technology key laboratory, 643000, China,

[4]Sichuan Provincial Academician (Expert) workstation of Sichuan University of Science and Engineering, 643000, China.

[a]1114137217@qq.com, [b]1017130100@qq.com, [c]340280663@qq.com

Abstract

In the process of engineering image detection feature, due to environmental interference and the target itself influence the accuracy of the circle fitting. A K-means algorithm is used to cluster the correct samples, and they are fitted to circle by the random sampling consistency (RANSAC) algorithm. The algorithm firstly filters out the environmental interference of the sample and the influence of the target by Gaussian filtering, then extracts its contour by Canny algorithm, and then circumscribe the minimum enclosing rectangle according to the characteristics of the sample. The minimum enclosing rectangle is divided into four sample regions and its centroid is obtained. According to K value selection and the principle of distance minimization, four sample regions are merged into K sample region, and K intersection region is generated at the same time. The K intersection region point set is classified into the corresponding sample area by K-means algorithm and calculate their centroid, and the sample area is determined as the correct number of samples by the principle of minimum distance of centroid each other, and the correct sample is fitted circle to using the RANSANC algorithm. The algorithm is tested by C++. The results show that the algorithm can effectively solve the problem of fitting accuracy due to contour difference. The fitting result can be consistent with the ideal result. The algorithm is fast, the fitting precision is high, and it has strong adaptability.

Keywords

Filter; circumscribe; divide; merge; K-means; RANSAC.

1. Preface

In the engineering application of machine vision, target detection in many fields involves detecting or fitting the circular features of the target, such as mechanical, electronic, etc., which will be rounded and fitted. There are many methods for fitting circles, such as Hough circle transformation, least squares method, weighted average method, RANSAC algorithm, etc. The Hough circle transformation algorithm has a large time complexity, a large amount of calculation, and requires a large memory, so the accuracy of detecting a circle is not good. The least squares algorithm is to fit three points in the sample point set according to the basic equation of the circle, find the parameters, and use the principle of distance square sum to find the most suitable fit circle. The traditional RANSAC algorithm is robust but has no iteration limit and high time complexity.

The k-means algorithm is a partition-based unsupervised clustering algorithm. The distance is used as the criterion for the similarity measure between data objects. The smaller the distance between data objects, the higher their similarity, the more likely they are. In the same class cluster. However, it is sensitive to outliers and isolated points. The initial cluster center selection is important and can only be applied to small data sets.

Therefore, an improved K-means algorithm based on region segmentation is proposed to classify the correct samples and error samples, and the RANSAC algorithm is used to fit the correct samples. Firstly, the minimum bounding circumscribing rectangle of the sample is obtained, and the outer rectangle is divided into rectangles. Then the K-means algorithm is used to divide the initial point set into two categories: the point set to be fitted and the set of error points. The RANSAC algorithm performs the fitting.

2. Contour Extraction

In image engineering applications, due to environmental disturbances and the nature of the target itself, it is usually not possible to directly extract the contour of the detection target. Therefore, before extracting the contour, the image needs to be preprocessed first, then the edge extraction algorithm is used to extract the contour. The commonly used edge contour extraction method has a first order differential method: Sobel operator, Roberts operator, Prewitt operator, Canny operator. Second-order differential method: Laplacian operator. Just to extract the image with clear outline of the target and no internal contour, firstly, the original image (Fig. 1) is Gaussian filtered to blur the image, then the maximum inter-class method is used to threshold the image and binarize it. Finally, the Canny operator is used to extract the edge contour of the image as shown in Fig. 2. At the same time, the circumscribed rectangle of all the contours is obtained. By selecting the appropriate aspect ratio, the target contour image is obtained, and the minimum bounding rectangle of the target is obtained.

Figure 1. Original image

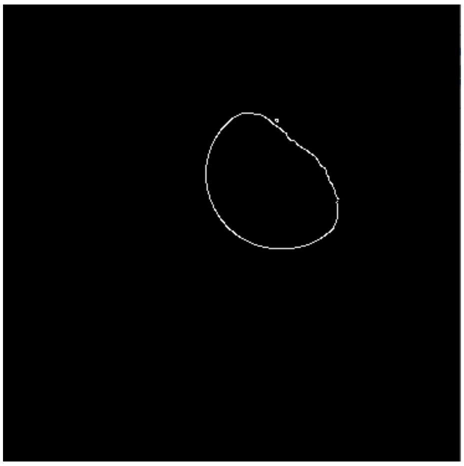

Figure 2. Canny image

2.1. Page Numbers

Do not number your paper: All manuscripts must be in English, also the table and figure texts, otherwise we cannot publish your paper. Please keep a second copy of your manuscript in your office. When receiving the paper, we assume that the corresponding authors grant us the copyright to use the paper for the book or journal in question. Should authors use tables or figures from other Publications, they must ask the corresponding publishers to grant them the right to publish this material in their paper. Use italic for emphasizing a word or phrase. Do not use boldface typing or capital letters except for section headings (cf. remarks on section headings, below).

ISSN: 2664-9640

$$c_2 = a_2 + b_2. \tag{1}$$

3. Improve the K-means Algorithm

The traditional K-means algorithm divides all the points in the sample from the initial cluster center into classes according to the corresponding mapping, and then repeatedly calculates the cluster center and updates it until the set condition is reached or the maximum number of iterations is reached. The loop gets the classification result. As the sample increases, the algorithm time will increase. Therefore, the sample point set is filtered first, then the sample is divided into 8 regions, then the region is redistributed and classified according to the specified algorithm, and finally the sample is divided into Two sets of ensemble set and error point set, and SANSAC algorithm circle fitting of the fitted point set, the sample point set is now recorded as P.

First: establish a minimum bounding circumscribing rectangle model as shown in Figure 3.

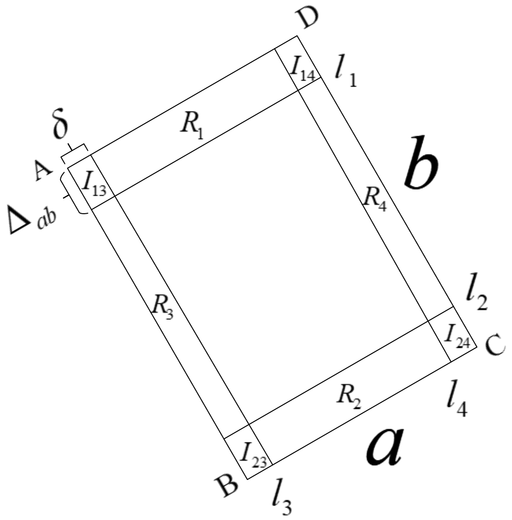

Figure 3. External rectangle model

Suppose the width of the rectangle is a, the length is b, and $a<b$. Since the outer rectangle of the circle is a square, so the distance difference Δ_{ab} between the length and width of the rectangle is obtained as $b-a$. If two lines l_1, l_2 are parallel to AD and BC, and the distance from AD and BC is a, the rectangular regions R_3, R_4 can be obtained. At the same time, make two rectangular fields R_1, R_2 parallel to AB and CD, and the distance from AB and CD is both δ. The value of δ needs to be set manually, so that the left and right points of the sample are exactly included in the fields R_1, R_2. Therefore, the point set P is divided into four rectangular regions: R_1, R_2, R_3, R_4, and each rectangular region contains two intersecting regions in R_{13}, R_{14}, R_{24}, R_{23}. The point set of each rectangular region is recorded as $\{P_i\}_{i=1}^4$.

Secondly, calculate the centroid $\{u_i\}_{i=1}^4$ and the sample P centroid $\bar{u}$ of the four regions R_1, R_2, R_3, R_4, as follows:

$$\bar{u}_x = \frac{\sum v_{xj}}{\sum N_i}, \bar{u}_y = \frac{\sum v_{yj}}{\sum N_i} \tag{2}$$

$$\bar{u}_x = \frac{\sum v_{xj}}{\sum N_i}, \bar{u}_y = \frac{\sum v_{yj}}{\sum N_i} \tag{3}$$

N_i indicates the size of the subset $\{P_i\}_{i=1}^4$, V_{xj} indicates the x axis coordinate corresponding to the j th point in the subset $\{P_i\}_{i=1}^4$, and V_{yj} indicates the y axis coordinate corresponding to the j th point in the subset.

The value of K $(K \le 4,\ K \in Z)$ is determined according to the characteristics of the sample, and the number of initial cluster centers is obtained.

If $K=4$, the four rectangular area centroids $\{u_i\}_{i=1}^4$ are directly used as the cluster center; when $K<4$, the distance d_{ij} between the centroids of the rectangular areas needs to be calculated:

$$d_{ij} = \sqrt{\left(u_{xi} - u_{xj}\right)^2 + \left(u_{yi} - u_{yi}\right)^2} \quad i,j = 1,2,3,4(i \ne j) \tag{4}$$

If $K=3$, the four rectangular regions are merged into three regions, and the minimum value of d_{ij} is selected, and the corresponding R_i, R_j regions are merged into one region, and then the region centroid is also calculated by Equation 2, and the other two regions remain unchanged. And renumber the area, and find the centroid $\{u_i\}_{i=1}^k$ of the corresponding area and the centroid $\{u_{ij}\}_{i=1,j=2}^k$ after the area is merged.

If $K=2$, four regions are converted into two regions, and there is a problem that three regions are merged into one region, or two regions are merged into one region. Set the appropriate threshold T_u, as long as $d_{ij} < T_u$ is met, then the R_i, R_j regions are merged until condition $d_{ij} > T_u$ is met. Then recalculate the area centroid and renumber the area, and find the centroid $\{u_i\}_{i=1}^k$ of the corresponding area and the center of mass $\{u_{ij}\}_{i=1,j=2}^k$ after the area is merged.

Finally, the improved K-means algorithm is used to cluster the set of fitting points. The specific steps are as follows:

1) Initializing the cluster center by using the regional centroid obtained by the above method;

2) Calculate the distance from each point in the intersection area to the centroid $\bar{u}$ of the sample point set P after the area is merged.

$$d_{p_t}^{(n)} = \sqrt{\left(p_{xt} - \bar{u}_x^{(n)}\right)^2 + \left(p_{yt} - \bar{u}_y^{(n)}\right)^2}, \ t = 1,2,\cdots,s_{ij} \tag{5}$$

Where s_{ij} is the number of intersection area point sets.

And calculate the distance from the centroid of the intersection region to the centroid of the sample point set P after the region fusion, as follows:

$$d_{pt}^{(n)} = \sqrt{\left(u_{xi}^{(n)} - \bar{u}_x^{(n)}\right)^2 + \left(u_{yi}^{(n)} - \bar{u}_y^{(n)}\right)^2}, \quad i = 1,2,\ldots,k \tag{6}$$

For high program execution efficiency, run time period, the formula is simplified to:

$$d_{p_t}^{(n)} = \left|p_{xt}^{(n)} - \bar{u}_x^{(n)}\right| + \left|p_{yt}^{(n)} - \bar{u}_y^{(n)}\right| \tag{7}$$

$$d_{u_i}^{(n)} = \left|u_{xi}^{(n)} - \bar{u}_x^{(n)}\right| + \left|u_{yi}^{(n)} - \bar{u}_y^{(n)}\right| \tag{8}$$

Find the distance from each point p_t in the intersection of the fused area to the other area to the center of mass $\left\{u_i^{(n)}\right\}_{i=1}^{k}$ of their area d_r,

$$d_r = \arg\min\left(d_{p_t}^{(n)} - d_{u_i}^{(n)}\right) \tag{9}$$

When d_r is the smallest, the point is classified into the i th area.

3) Until all points in the intersection area are classified, and then recalculate the centroid $\left\{u_i^{(n)}\right\}_{i=1}^{k}$ and termination conditions of each area;

$$\varepsilon \ge \left\{u_i^{(n)}\right\}_{i=1}^{k} - \left\{u_i^{(n-1)}\right\}_{i=1}^{k} \tag{10}$$

4) Update the cluster center with the centroid of step 3 until the condition is met or the maximum number of iterations is reached, and the clustering is completed, otherwise steps 2 and 3 are repeated.

4. RANSAC algorithm

RANSAC achieves its goal by iteratively selecting a random set of data in the data. The selected subset is assumed to be an intra-site point and verified by the following method:

1. Select the point set P obtained by the clustering algorithm as the initial set;
2. Randomly select 3 points, and calculate the parameter center $C(x,y)$ and radius r of the model conforming to the three points according to the basic formula $x^2 + y^2 + Dx + Ey + F = 0$ of the circle;
3. Set the threshold T, calculate the distance d_i^t from each point of the point set P to the center $C(x,y)$ of the step 2, calculate the distance $\left|d_i^t - r\right|$ from each point to the circle of step 2, determine whether the distance is less than the threshold T, and if it is less than, record it as an intra-point. Otherwise it is an outlier;
4. Iterate the steps 2, 3 for the inner point set until the number of iterations is greater than the set number of iterations or the parameters before and after the inner point set is unchanged, and record it as the best model of the circle, otherwise continue the iteration;

And then we can calculate the center and radius of the circle based on the best model.

5. Experimental Analysis

Figure 4 uses the Hough transform to detect the circle. The center and radius of the test result are very different from the actual center radius. The fitted circle can only be included in the sample contour. The local boundary distortion has a great influence on the accuracy of the sample fitting. The difference is not applicable to the graph with the flattened shape of the sample; Fig. 5 is the result of fitting the circle by the least squares method, which is better than the Hough fitting, but the result is not much contact with the original sample, and the fitting precision is small. Figure 6 shows the results of the RANSAC algorithm fitting. The results are good, but the fitting circle still has a small deviation from the actual sample contour. Figure 7 is the result of the algorithm fitting. The fitting circle mostly follows the original sample curve and occurs in the contour. In the case of local distortion, the accuracy of the fitting is high and the accuracy is good.

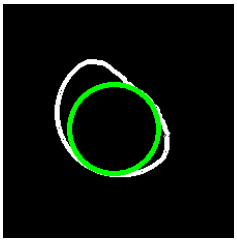

Figure 4. Hough

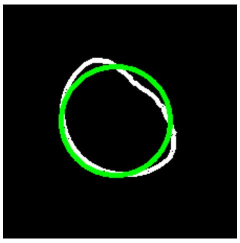

Figure 5. Least squares method

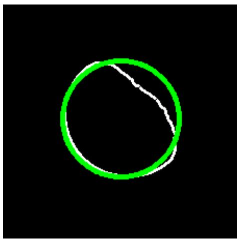

Figure 6. RANSAC

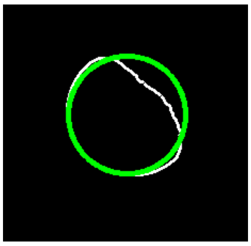

Figure 7. K-means+RANSAC

6. Conclusion

Aiming at the problem of circle fitting accuracy caused by image circle contour distortion, a RANSAC circle fitting algorithm based on region segmentation K-means clustering is proposed. Compared with other methods, the results show that the method has a segmentation in the circular contour. It has a good fitting effect under the condition of deformation, etc. It successfully solves the fitting problem caused by the contour, has fast response, strong anti-interference ability and high fitting precision, and is suitable for industrial testing.

Acknowledgements

The Project of Sichuan provincial science and Technology Department (Grant No. 2017JY0338,2019YJ0477,2018GZDZX0043).

Sichuan University of Science and Engineering talent introduction project (2017RCL53).

Enterprise informatization and Internet of things measurement and control technology key laboratory project of Sichuan provincial university (2018WZY01).

The Project of Sichuan Provincial Academician (Expert) workstation of Sichuan University of Science and Engineering(2018YSGZZ04).

References

[1] G.W.Xu, M.Liao. Several ways to fit a circle, Journal of Wuhan Polytechnic University,vol. 4(2002) 104-106.

[2] G.D. Sun, L.J.Yang, Y.Zhang. Clustering circle fitting algorithm based on three-point iteration, Manufacturing automation,vol.37(2015)No. 13,76-78.

[3] Rudolf Scitovski, Tomislav Marošević, Multiple circle detection based on center-based clustering, Pattern Recognition Letters, vol. 52(2015) 9-16.

[4] X.Y.Chen, Q.Wang,B.L.Li. Improved Hough transform detection circle method, Computer system application,vol. 24(2016)No. 08,197-201.

ISSN: 2664-9640

[5] X.Wang, Y.D.Zhang, L.Wang. An Optimization Method for Random Hough Transform Detection Circle, Measurement and Control Technology, vol. 35(2016)No. 6, 112-116.

[6] J.L.Cui, H.Gong, Y.D.Wang, R.M.Jia, K.Xiao. Optimization and Application Research of Ellipse Fitting Based on Algebraic Distance, Computer Science, vol. 41(2014)88-90.

[7] KANATANIK, RANGRARAKAN P, Hyper least squares fitting of circles and ellipses, Computational Statistics & Data Analysis, vol.55(2011), No. 6, 2197-2208.

[8] Y.Li, B.Qiao, Y.P.Lu, Circle fitting algorithm based on clustering method, computer application, vol. 33(2013),206-208.

[9] A.M.Bagirov, J.Ugon, D.Webb, Fast modified global k-means algorithm for incremental cluster construction, Pattern Recognition,vol.44(2011)