Computer Life
ISSN: 1819-4818 | Vol. 12, No. 1, 2024

Design of moving object detection algorithm based on RoboMaster

Shugang Liu, Yonghang Zhang, Xuyang Dong

School of Computer Science, North China Electric Power University, Baoding 071003, China

Abstract: RoboMaster is a global high-level robot competition, in which the vision system of infantry robot is studied in this paper. In complex environment, the existing vision detection algorithms have some problems, such as unstable recognition, low detection accuracy and long running time. The improvement and optimization of the algorithm are as follows: image quality detection is carried out when brightness is bright or brightness is abnormal; gamma transform is used in the pre-processing process; maximum class difference method is used in image segmentation; SVM algorithm is used to train the selection of light strips; directional gradient histogram is used to extract features; and PNP algorithm is used to calculate the single-visual distance measurement and camera attitude. The experimental results show that the vision detection algorithm designed in the complex scene can realize the recognition of moving objects.

Keywords: RoboMaster, Motion target detection, SVM, PNP attitude solution, multi-feature.

1. Introduction

The object detection algorithm based on RoboMaster's robot vision system is designed to assist robots such as infantry and heroes to realize the function of automatically aiming at enemy units.

2. Research Background and Significance

RoboMaster is a high-tech robotics competition launched and hosted by DJI Innovation for university students and graduate students around the world. The research object of this paper is the vision system carried by our infantry robot in the competition. In the complex field background environment and the requirement of real-time, the specific area of the enemy mobile robot is recognized stably, accurately and quickly. In complex scenes, when the camera moves or the target itself rotates, the fast-moving target is prone to blur, and the appearance of the moving object will change accordingly, which will affect its detection accuracy, which also brings challenges to the target detection. Traditional target detection algorithm has the advantages of high efficiency, easy interpretation, fast speed and wide application range[1]. Under the condition of making full use of image features, the traditional target detection algorithm can still play a great role in the field of image target detection.

3. Vision System Design

3.1. Hardware Selection

The selection of vision system hardware mainly involves the choice of camera, lens and computing hardware. This system uses Daheng Industrial camera +intelNUC+USB to TTL. Due to the need to have the ability to recognize the fast movement of the target, in order to avoid the large-area blur effect caused by too fast speed, the camera exposure mode is selected global shutter, and the camera acquisition requires a high frame rate.

3.2. Algorithm Flow

Image quality detection

In this paper, the brightness gray detection method is used. In the actual competition process, the subject of the shooting is usually illuminated by the stronger indoor light or under the lower light. No matter the background light of the image is dark or bright, the overall background details of the image are usually not legible, which affects the pre-processing and recognition effect of the actual object.

Image brightness refers to the degree of light and dark of the image. In color images, grayscale values are usually used to represent the brightness of the image. The gray-scale processing of the three components was carried out by weighted average method. The gray image can be obtained by using the following formula to process RGB three-component weighted average.

$$Gray(i,j) = 0.299 * R(i,j) + 0.578 * G(i,j) + 0.114 * B(i,j)$$

18% is the reflectivity of gray in digital images when RGB values are all 128. Therefore, the grayscale value of 128 is used as a reference for moderate brightness.

$$da = \frac{\sum_{i=0}^{N}(x_i - 128)}{N}$$

Let $D = |da|$, $N = W * H$ where da means that the gray value of the image deviates from the mean value of the gray value 128, and W and H represent the width and height of the image respectively.

$$Ma = \frac{\sum_{i=0}^{255}|(i - 128) - da| * Hist[i]}{N}$$

Let $M = |Ma|$, where Ma represents the variance of the image gray value from the gray value 128, and Hist[i] represents the order of magnitude of the gray value i.

Where K is the brightness coefficient, which is used to measure the brightness of the image. If K is less than 1, the brightness of the image is normal; If K is greater than or equal to 1, the brightness of the image is abnormal. Specifically, when the value of K is greater than 1, if the gain da is greater than 0, the image is too bright; Conversely, if da is less than 0, the image is too dark.

Image preprocessing

In order to solve the possible luminance anomaly problem, gamma transform is used for image preprocessing. By transforming the gray information in the image, the algorithm enhances the contrast and brightness information of the image and reduces the unnecessary interference in the image. The transformation formula is shown below:

$$s = cr^{\gamma}, r \in [0,1] \qquad (3.4)$$

Where s and r represent output gray and input gray respectively. In short, if the scene is dark, selecting a gamma value of less than 1 for correction can effectively enhance the brightness and contrast of the image. If the scene is light, you need to select a gamma value greater than 1 for correction.

Object detection algorithm design

The two sides are divided into red and blue colors, and the target colors are significantly different, so the color characteristics of the target are fully utilized for image segmentation, which is realized by OSTU method in this paper. From the principle of the Otsu method, this method is also called the maximum inter-class variance method, because after image binarization segmentation according to the threshold obtained by the Otsu method, the foreground and background images have the maximum inter-class variance.

Fig. 1 Split blue light strip

The specific steps are roughly as follows: Firstly, the first step uses color features to segment the foreground image and mask the original image to obtain an RGB three-channel image with only part of the target region. The second step will be weighted grayscale processing. The third step is to segment the binary image of the target region by using the maximum difference between classes method.

Strip screening

After the above segmentation process, assuming that n luminous regions have been obtained, and both sides of the armor plate are composed of two symmetrical luminous strips, if paired in pairs there are $n(n-1)/2$ pairing possibilities.

Therefore, the real-time and accuracy of the system's target detection will be reduced, and how to screen out the real light strip becomes the core problem. The following geometric features are selected: the aspect ratio of the minimum surrounding rotating rectangle of the luminous region, the absolute difference of the Angle of the minimum surrounding rotating rectangle of the luminous region, the absolute difference of the area of the minimum surrounding rotating rectangle of the two luminous regions, and the Euclidean distance of the center point of the luminous region in the pixel coordinate system. By establishing the eigenvector containing these eigenvalues, the target armor plate area is determined jointly. SVM algorithm is used for training and classification recognition is completed.

Number recognition based on HOG+SVM

HOG (Histogram of Oriented Gradient), which belongs to texture feature, is called histogram of oriented gradient. The way to form the feature is to calculate and count the gradient histogram of the local area of the image, which can describe the edge feature of the image well[2]. The process of HOG feature extraction is shown in the figure

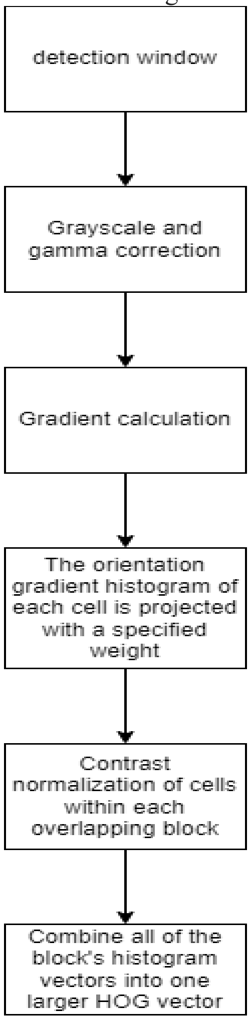

Fig. 2 Feature extraction process

Solution of monocular distance measurement and attitude based on PnP algorithm

PnP(Perspective-n-Point) is a common method for solving 3D to 2D point pair motion. It can be used to estimate the pose of the target relative to the camera. When n 3D space points and their projected positions on the 2D image are known, the PnP method can be used to estimate the pose of the target[3].

In order to solve the problem of 3D to 2D point pose estimation, P3P algorithm can be used. The algorithm can be solved by only 3 pairs of matching points, which greatly reduces the data requirement and calculation cost. Compared with other methods, P3P algorithm has the advantages of convenient calculation and fast calculation speed, and is suitable for solving the camera pose problem quickly and accurately. The P3P problem is shown in Figure 3.

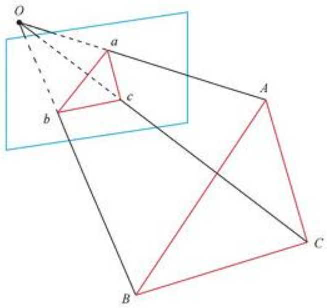

Fig. 3 P3P problem schematic

In the world coordinate system, 3D points are a, b, c, and 2D points are a, b, c, where the point represented by the lowercase letter is the projection of the uppercase letter on the camera imaging plane. The following similarity exists as shown in the figure above:

$$\triangle Oab \backsim \triangle OAB, \triangle Obc \backsim \triangle OBC, \triangle Oac \backsim \triangle OAC$$

For the triangle △OAC, △OAB,△OBC can be obtained using the law of cosine:

$$\begin{cases} OA^2 + OC^2 - 2OAOCcos\langle a, c\rangle = AC^2 \\ OA^2 + OB^2 - 2OAOBcos\langle a, b\rangle = AB^2 \\ OB^2 + OC^2 - 2OBOCcos\langle b, c\rangle = BC^2 \end{cases} \quad (3.5)$$

Let $x = OA/OC, y = OB/OC, v = AB^2/OC^2, uv = BC^2/OC^2, wv = AC^2/OC^2$, after a series of substitutions, transformations, and simultaneous equations can be obtained as follows:

$$\begin{cases} (1-u)y^2 - ux^2 - ycos\langle b, c\rangle + 2uxycos\langle a, b\rangle + 1 = 0 \\ (1-w)x^2 - wy^2 - xcos\langle a, c\rangle + 2wxycos\langle a, b\rangle + 1 = 0 \end{cases} \quad (3.6)$$

The Angle value cos<a,b> ,cos<b,c>,cos<a,c>in the above equations can be calculated by the internal parameters of the camera, and then according to the 3D-3D point pairs, the camera motion pose rotation matrix R and the translation vector t can be solved.

4. Conclusion

This paper uses the infantry robot vision system in robomaster as the research object, and designs a set of visual algorithms for image processing and tracking recognition under complex environment scenarios. This paper fully analyzes the environmental complexity of the robot, uses the morphological characteristics and number of the enemy armor plate, and proves the feasibility and robustness of the algorithm through theoretical analysis and experimental verification.

References

[1] Kalantari S, Navarathna R, Dean D, et al. Visual front-end wars: Viola-Jones face detector vs Fourier Lucas-Kanade[C]// Proceedings of the 12th International Conference on Auditory-Visual Speech Processing. Inria, 2013: 163-168.

[2] Wang Q ,Li J ,Lu C , et al. HOG-based feature extraction and identification for raised or indented characters pressed on metal labels[C]//International Research Association of Information and Computer Science.Proceedings of the International Conference on Control,Mechatronics and Automation Technology(ICCMAT 2014).CRC Press/Balkema,2014:4..

[3] Fischler M A . A paradigm for model fitting with applications to image analysis and

[4] automated cartography[J]. Communications of the ACM, 1981, 24.