Vision-based robot identification and tracking

Dong He¹*, Jun Lu², Qinghe Zhou³, Hongjie Yin⁴

¹,²,³,⁴ School of Software Engineering, Chengdu University of Information Engineering, Chengdu, Sichuan 610225, China

*Corresponding author’s e-mail: 3190702027@stu.cuit.edu.cn

Abstract: In recent years, computer vision technology in the field of science and technology has been stubbornly high, making a lot of scientific and technical personnel involved in the research gate technology. Depth learning, artificial intelligence in recent years, often around our ears, vision technology is occupying a place in the field of artificial intelligence. This paper mainly studies the recognition and localization of robots based on vision. The mobile robot is detected by the background difference method, and the background model is established by the Gaussian background modeling. Recognizing the color and number tags on the robot for matching the robot’s features. Real-time positioning of the robot under the premise of the same background.

1. Introduction

Target object detection is currently a major research topic in the direction of vision. For object recognition and detection, it is mainly the detection of features such as object color, shape contour, spatial relationship and texture [1]. In deep learning, algorithms such as RCNN, Fast CRNN and Faster CRNN, as well as advanced algorithms such as SSD and YOLO that appeared later, have made detection more and more efficient and accurate.

Motion detection is mainly to judge moving objects on time series image frames. Typical moving target detection [2] The current main methods include background difference method, inter-frame difference method, optical flow method, etc. [3]. The main point of the background difference method is background modeling. Among many background modeling methods, Gaussian mixture background modeling stands out. It has good robustness in image noise interference, light changes, shadows, etc. Gaussian mixture model of the image taken by the local division processing, each section is provided with a Gaussian model, each section has a threshold value to binarize the target image segmentation, and more flexibility in the processing of the image.

2. Motion detection

2.1. Background-difference method

For the detection of moving objects it is generally priority camera is fixed or mobile. For fixed camera, we usually use the difference method to detect the object, a background difference method is mainly the background pixel frame changes obtained by subtracting the motion frame, the moving object is determined by a change in pixels.

This motion detection is the detection of moving objects under a fixed camera, so the background difference method is selected for motion detection. This experiment chooses to model the Gaussian mixture background. It has the ability to self-learn the background establishment and can keep up with the new background. Among currently Gaussian mixture background modeling method regarded as a
successful background modeling method of modeling, each of its local component to build a model, each part has its own independent local segmentation threshold [4], Therefore, it is more excellent in the effect of local treatment.

2.2. Gaussian mixture background modeling

Each pixel value $x_t$ we can use K Gaussian distributions represented by each pixel value of the difference between the mean of a Gaussian model $x_t$ thereto is less than 2.5 times the standard deviation of the pixel to be matched with the model, if the matching success is as a background, or foreground. According to such a phenomenon that we can use each of pixels Gaussian mixture model of K Gaussian distributions to represent configuration, i.e.

$$p(X_t) = \sum_{i=1}^{K} \omega_{i,t} \cdot \eta(X_t, \mu_{i,t}, \varphi_{i,t})$$

$$\eta(X_t, \mu_{i,t}, \varphi_{i,t}) = \frac{1}{(2\pi)^{D/2} |\varphi_{i,t}|^{1/2}} \exp \left[ -\frac{1}{2} (X_t - \mu_{i,t})^T \cdot \varphi_{i,t}^{-1} \cdot (X_t - \mu_{i,t}) \right]$$

$\omega_{i,t}$ is the weight, $\mu_{i,t}$ is the average, $\varphi_{i,t}$ is the covariance, $\varphi_{i,t} = \sigma_{i,t}^2 \cdot I$, $I$ is the identity matrix, $\sigma_{i,t}^2$ is the variance.

For each new pixel value $x_t$, compare it with the existing K Gaussian distribution models. If $|x_t - \mu_{i,t-1}| \leq 2.5\sigma_{i,t-1}$ is satisfied, the pixel is the background Point, otherwise it is the prospect. If the Gaussian distribution, the weights in the model are updated according to formula 3,

$$\omega_{i,t} = (1 - \alpha)\omega_{i,t-1} + \alpha M_{i,t}$$

If a match is Gaussian distribution, the parameters of the Gaussian model that matches the current update in accordance with Formula 4 Formula 5:

$$\mu_t = (1 - \gamma)\mu_{t-1} + \gamma X_t$$

$$\sigma_{t}^2 = (1 - \gamma)\sigma_{t-1}^2 + \gamma (X_t - \mu_t)^T(X_t - \mu_t)$$

$\gamma = \frac{\alpha}{\omega_{i,t}}$ is the update rate, If the Gaussian model is not matched, just update the weight $\omega_{i,t}$;

According $\frac{\omega_{i,t}}{\sigma_{i,t-1}}$ descending order of the ratio of K Gaussian distributions of Gaussian mixture model of each pixel thereof. Then we choose the first B Gaussian distribution as the background model

$$B = \arg_b \min (\sum_{k=1}^{b} \omega_k > T)$$

$$X_{bg} = \sum_{t=1}^{n} \omega_{i,t} \mu_{i,t}$$

T is the threshold, This experiment setting $T=0.85$, $X_{bg}$ is the background;

The component description of the image in the HSV space is more in line with the human eye’s observation of image Saturation and Hue, so it is selected Eliminate shadows in HSV space [5][6].

$$S(x,y) = \begin{cases} 
1, & \frac{\alpha \leq I_v(x,y)}{B_v(x,y)} \leq \beta \\
|I_s(x,y) - B_s(x,y)| & \leq T_s \\
|I_H(x,y) - B_H(x,y)| & \leq T_R \\
0, & \text{else}
\end{cases}$$

$I_v(x,y) , I_s(x,y) , I_H(x,y)$ is the components of V,S,H in the current frame,
\(B_{0}(x, y), B_{s}(x, y), B_{u}(x, y)\) is the components of V, S, H in the background frame, \(\alpha, \beta\) is the threshold to control the brightness, \(\alpha, \beta \sim (0, 1)\). \(T_{s}, T_{r}\) is the threshold of Hue and Saturation.

3. Recognition

3.1. Color recognition

The image displayed by the computer is formed by permutation and combination of each pixel. The current display system adopts the RGB color model [7] for color display, as shown in the RGB space model represented by the Cartesian coordinate system in Figure 1:

![Figure 1 RGB model](image1.png)

Since the RGB color space with the human eye color vision information received does not match, it does not make a very good response to specific kind of color, it will cause a lot of inconvenience in the processing of the image analysis, image processing will therefore usually RGB color space converted into other color spaces. The HSV space [8] from the dimension can effectively reflect various specific color information, and therefore we need to convert the image pixels from the RGB space to HSV space. As shown in Figure 2. First picture from RGB space to HSV space, and then setting the threshold value for each color, the image color lookup based on the threshold of the color, and then converts the image into a binary image. Find the area to be color image segmentation By the outline query. This time, the traditional method of threshold segmentation is used, mainly because of its simple calculation and high efficiency. Segment the original image by color area according to the contour area, and perform a color query again. If the completeness of the query area is greater than 85%, the query is successful.

3.2. Digital recognition

Digital recognition is to segment each digit of the picture to be tested, and match the similarity of the segmented picture with the template. The corresponding template with the greatest similarity is the recognized digit. First, the image is binarized, and then scanned to find the sum of each column. When scanning to the far left of the first number, its value is 1, the sum of its columns is not 0, and the value of the current column is recorded as left. Then continue scanning. When the rightmost contour of the number is scanned, the sum of the columns is 0, and the value of the current column is recorded as right, and the width of the number is: right – left. Similarly, record the top and bottom of the upper and lower parts, and the height of the number is: top – bottom. The numbers have been divided. Adjust the segmented image to the same pixel size as the template to facilitate comparison. Finally, compare the pixels of the template and the segmentation map, and accumulate the sum of the same points: same, the sum of different points: different, the similarity is: \(\frac{\text{same}}{\text{same} + \text{different}}\). Comparing each template, the template number with the greatest similarity is the identified number.
Table 1 Similarity and accuracy of digital recognition

| Digital | experiment/times | Misidentification/times | Similarity/% | Accuracy/% |
|---------|-----------------|-------------------------|--------------|------------|
| 0       | 100             | 6                       | 88.4         | 94         |
| 1       | 100             | 9                       | 82.6         | 91         |
| 2       | 100             | 10                      | 79.8         | 90         |
| 3       | 100             | 12                      | 72.5         | 88         |
| 4       | 100             | 15                      | 81.4         | 85         |
| 5       | 100             | 12                      | 77.4         | 88         |
| 6       | 100             | 13                      | 80.8         | 87         |
| 7       | 100             | 7                       | 85.2         | 93         |
| 8       | 100             | 12                      | 76.4         | 88         |
| 9       | 100             | 12                      | 81.6         | 88         |

4. Experimental conclusion

The motion detection effect of the experiment based on Gaussian background modeling is shown in Figure 3. The shadow processing effect is shown in Figure 4. The contour of the moving target is shown in Figure 5.

The experimental flowchart is shown in Figure 6:
Image preprocessing

Whether the video has been read

Yes

Get pictures in frames

No

Whether the picture has been obtained

Yes

Image preprocessing

RGB is converted to HSV

Color segmentation

Color recognition

Background modeling

Moving object detection

Contour query

Coordinate calculation

Annotation of original drawing

No

End

The color recognition of the experiment is shown in Figure 7. The figure is the recognition of the red area. In the display effect of the binary graph, it can be clearly seen that the red area is recognized.

The experiment on digit recognition is shown in Figure 8. This experiment only recognizes one digit, so the image segmentation of a row of digits is chosen.

The overall experimental results shown in figure 9, for identifying the color of the robot, identification number, a coordinate positioning substantially complete.
5. Conclusion
This experiment is based on the recognition and detection of moving robots based on vision. In the motion detection, the background difference method is used for moving target detection, and the mixed Gaussian background modeling is used. While the background is updated, the background difference method is retained at the same time. The effect of recognizing objects. The color recognition selection is converted into HSV space for recognition, and the number recognition is realized by the method of comparing the similarity of the template. The method is simple and effective, but the shortcomings can only compare similar numbers and cannot achieve adaptive learning. The establishment of coordinates is based on the constant background environment for positioning. The disadvantage of this experiment is that the training set is for this experiment, and more models need to be trained in the follow-up.

Acknowledgments
This article is supported by the key research and development project of Sichuan Provincial Department of Science and Technology (No. 21ZDYF0055).
This article is supported by the Key R&D Support Program of Chengdu Science and Technology Bureau (No. 2019-YF05-01387-SN).

References
[1] Jirui Wang, Cailiang Yan. Summarization of Image Feature Extraction Methods[J]. Journal of Jishou University (Natural Science Edition),2011,32(05):43-47.
[2] Linhao Li, Qinghua Hu, Xin Li 0005. Moving Object Detection in Video via Hierarchical Modeling and Alternating Optimization. [J]. IEEE Transactions on Image Processing, 2019, 28(4):2021-2036.
[3] Dongmei Zahng, Jie Wu, Piding Li. A Survey of Moving Target Detection Algorithms Based on Machine Vision[J]. Intelligent Computers and Applications,2020,10(03):192-195+201.
[4] Haitao Wang, Jingjing Hou. Pig motion tracking based on Gaussian mixture model fusion background reduction[J]. Zhejiang Agricultural Sciences,2020,61(02):329-332.
[5] Jing Li, Jin Che, Jinyan Zhu. Moving target detection and shadow removal based on color space conversion[J].TV Technology,2013,37(13):172-175.
[6] Qing Xu, Zhu Xu, Shitong Wang, Yongzhao Zhan. Moving target detection based on HSV adaptive Gaussian mixture model[J]. Computer Science,2010,37(10):254-256+290.
[7] Ling Mo. Research on image segmentation based on RGB color space[J]. Computer Science,2016,43(S1):168-170.
[8] Mengxia Hong, Shaohua Liang. Color image segmentation based on RGB color space[J]. Computer Knowledge and Technology,2020,16(34):225-227.