

Cigarette box code recognition based on machine vision

Qifan Wen, Xiaofei Ji *

School of Qingdao University, Shandong 266000, China.

* Corresponding author: Xiaofei Ji.

Abstract: This paper addresses this challenge by using machine vision to firstly collect cigarette box image data and process the images with grayscale and binarization, and then extract the characters to be trained by extracting the region of interest and doing threshold segmentation in turn. The SVM classifier was used to train the extracted characters, and finally the characters were recognized in turn, and the recognition effects of different classifiers were compared and analyzed, and it was concluded that the SVM classifier has the best effect and is suitable for enterprise production.

Keywords: Machine Vision; Machine Learning; OCR; Image Processing.

1. Introduction

With the rapid development of China's economy, the pace of iteration and upgrading of the tobacco industry is also accelerating. In order to achieve long-term, stable and healthy development of enterprises, on the inspection line of the incoming tobacco carton, in the past, the domestic production line generally adopted the manual inspection method, relying on the human eye to check the information of the carton markings. The results show that the correct rate of identifying the incoming information of cigarette boxes is greatly affected by human factors, and it is difficult to achieve the complete incoming and backup of cigarette boxes stably, which also has a certain impact on the production efficiency of enterprises. The traditional manual verification of cigarette box information in high-speed production is both labor-intensive and has a high mis-checking rate, which is far from meeting the needs of refined production. This paper constructs an OCR[1] cigarette box code recognition system based on machine vision, which can effectively solve the problems of low efficiency and high error of manual verification of cigarette box information.

2. Technical Approach

2.1. Image pre-processing.

Image pre-processing is the grayscale[1], binarization, expansion erosion and other operations on the acquired image, which is able to meet the requirements of text detection and recognition, the image pre-processing flow chart is shown in the figure 1

2.1.1. Image grayscale

Grayscale is the process of converting a color image into a gray image. In the image, each pixel point is represented by a three-dimensional vector as (r,g,b), which ranges between [0,255]. And (r,g,b) represents the three primary colors in color: red, green, and blue. Because each pixel point corresponds to each vector value of 256 kinds, the value of each pixel point is taken as three times of 256.

The grayscale map responds to the change of brightness of the image, and since in the grayscale image, $r=g=b$, the value interval of the grayscale map is [0,255]. The grayscale method used in this paper is the weighted average method. It refers to the gray value image obtained by taking each pixel point in

the color image and calculating it according to the linear variation of the YUV color model and the rgb color space.

$$f(i, j) = 0.299r(i, j) + 0.587g(i, j) + 0.144b(i, j)$$

2.1.2. Binarization

Binarization is the conversion of a 3-channel image to a 1-channel image by targeting a fixed threshold and setting the gray value of the pixel points that pass that threshold to 0 or 255. The binarization operation can reduce the information not related to the image and greatly improve the processing speed of the image.

2.2. Target area segmentation

In order to obtain the text information to be extracted from the image, it is necessary to segment the text part and divide the image into different parts, i.e., into different parts of interest according to the pixel value compared with the threshold value, and extract those regions of interest in the image.

2.2.1. Region of interest

Once the original image has been captured, the region of interest can be selected as the ROI, which can be any shape, conventionally a rectangle, a circle, an ellipse, a custom one or a specific one derived from image processing. At this point, the selected area cannot yet be called an ROI, it is still only a shape or a range of pixels. To turn this part of the region into a separate image, it also needs to be cropped out of the original image. As shown in the figure, we need to intercept the image of the "finished piece of cigarette logo" field from the original image as the region of interest, in order to facilitate the subsequent training and recognition process, the selection of the above-mentioned area with a rectangular box, Next, we will further process the region of interest: first, we use a threshold segmentation to extract the text. Since the text in the region of interest has different grayscale values from the background, we can dynamically select a better threshold range by adjusting different thresholds to extract the text we need to recognize.

2.2.2. Expansion and corrosion

Expansion[3] and erosion belong to image morphological ground processing and are used to extract meaningful local image details from an image. By changing the pixel morphology of local regions in order to enhance the target for subsequent text training and processing. After the binary image of the target region is extracted by thresholding, the

edges of the region may not be ideal, and the region can be "shrunk" or "expanded" using an erosion or expansion operation. Expansion and corrosion are described in the following order.

1)Corrosion:

Erosion is an operation that "shrinks" the selected area and can be used to eliminate edges and stray spots. The size of the corrupted area is related to the size and shape of the structural element. The principle is to use a custom structural element, such as a rectangle, circle, etc., to perform a sliding operation similar to "filtering" on a binary image, and then compare the pixel points corresponding to the binary image with the pixels of the structural element, and the resulting intersection (by "with " operation) is the corrupted image pixels. The image on the left is the binarized image, and the image on the right is the image that has been "shrunk" by using the middle structural element to corrupt the image. After the erosion operation, the edges of the image area may become smooth, the number of pixels in the area will be reduced, and the connected parts may be broken. Even so, the parts still belong to the same area.

2)Expansion

In contrast to erosion, expansion is an operation that "expands" the selection. The principle is to use a custom structure element to perform a sliding operation similar to "filtering" on the binary image to be processed, then compare the pixel points corresponding to the binary image with the pixels of the structure element, and the resulting merge (with an "or" operation) is the expanded image pixels. The image on the left is the image after binarization, and a circular structural element is used to expand the image, as shown in the middle image, resulting in an "expanded" image, as shown on the right image. After the expansion operation, the edges of image regions may become smooth, the number of pixels in the region will increase, and disconnected parts may be connected, all of which are the opposite of the erosion operation. Even so, the originally disconnected regions still belong to their own regions and will not be merged because of pixel overlap.

3)Open computing

After the expansion operation, the edges of image regions may become smooth, the number of pixels in the region will increase, and disconnected parts may be connected, all of which are the opposite of the erosion operation. Even so, the originally disconnected regions still belong to their own regions and will not be merged because of pixel overlap.

4)Closed operations

The calculation steps of the closed operation are the opposite of the open operation, being expansion first and erosion second. This two-step operation joins elements that look close together, such as voids inside a region or isolated points outside, without any significant change in the appearance or area of the region. In layman's terms, it is similar to a "gap-filling" effect. Unlike the separate expansion operation, the closing operation fills the gaps without thickening the outline of the image edges.

2.3. Character Recognition

HALCON provides some OCR recognition models, and training the models required for the project in the specified project can improve the accuracy of recognition, so the OCR models are trained in this paper.

HALCON supports three kinds of training classifiers, namely BOX, multilayer neural network (MLP) and support

vector machine (SVM). Considering the small number of samples sets and the superiority of classification effect, this paper adopts support vector machine as the training classifier of OCR model.

2.3.1. SVM theory

The basic idea of SVM[4] is binary classification, which aims to find the best separating hyperplane in the feature space so that the training set samples in the training set produce the maximum positive and negative intervals. For the sample set $T=\{(x_1,y_1), (x_2,y_2), (x_3,y_3), (x_4,y_4), \dots, (x_i,y_i)\}$, x_i stands for sample, $y_i \in \{+1, -1\}$, i denotes the i -th sample

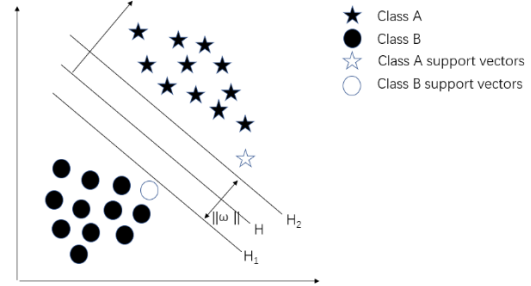


Fig. 1 Schematic diagram of linear support vector machine

In the Figure 1, the circle represents the positive sample, the pentagram represents the negative sample, and H represents the hyperplane, which can be represented by the linear equation $\omega x+b=0$. The equation w represents the normal vector, which is represented by $\omega x+b=+1$ and $\omega x+b=-1$, respectively, and the distance between the two planes is $2/\omega$. The core idea of linear support vector machine is to solve the optimal hyperplane H, which is equivalent to the process of solving for w and b . As shown in Eq.

$$\begin{cases} \min \frac{1}{2} \|\omega\|^2 \\ s.t. y_i(\omega x_i + b) \geq 1, i = 1, 2, \dots, n \end{cases}$$

introduction of Lagrangian operator and build function:

$$J(\omega, b, a) = \frac{1}{2} \omega^T \omega - \sum_{i=1}^n a_i [y_i(\omega x_i + b) - 1]$$

where a_i is greater than 0, then set w and b to 0 and then find the partial derivative, as shown in Eq.

$$\begin{cases} \frac{\partial J}{\partial \omega} = 0 \\ \frac{\partial J}{\partial b} = 0 \end{cases}$$

solving the above equation yields the following equation

$$\begin{cases} \omega = \sum_{i=1}^n a_i y_i x_i \\ \sum_{i=1}^n a_i y_i = 0 \end{cases}$$

the pairwise problem in the quadratic programming problem can be obtained by sorting J according to the above results

$$\begin{cases} \max Q(a) = J(\omega, b, a) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_i a_j y_i y_j x_i^T x_j \\ s.t. \sum_{i=1}^n a_i y_i = 0, a_i \geq 0 \end{cases}$$

the maximization of $Q()$ is related to the sample data, and the optimal Lagrangian operator a^* can be obtained by solving Eq. (3.7), when the optimal solutions of ω and b are:

$$\begin{cases} \omega^* = \sum_{i=1}^n a_i^* y_i x_i \\ b^* = y_i - \omega^* = \sum_{i=1}^n a_i^* y_i (x_i \boxtimes x_j) \end{cases}$$

the constructed hyperplane H^* is:

$$\omega^* x + b^* = 0$$

the decision function is:

$$f(x) = \text{sgn}[\sum_{i=1}^n a_i^* y_i (x_i \boxtimes x) + b^*]$$

when $f(x) > 0$, the sample is positive; when $f(x) < 0$, the sample is negative. The basic idea of nonlinear support vector machines is similar to that of linear support vector machines, with the difference that it requires the use of kernel functions that map the original feature space to the high-dimensional feature space, and then perform the optimal hyperplane for the samples in the high-dimensional feature space the principle is shown in Figure 2.

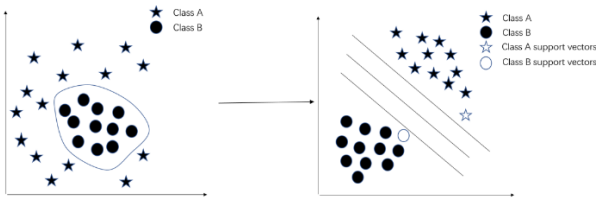


Fig. 2 Mapping of original feature space to higher dimensional feature space

2.3.2. Training classifier and recognition process

The steps for training the classifier and using the classifier for cigarette box character recognition are as follows:

Step 1: Prepare the character image and segment the single character image, identify and normalize the size of the character image, and save it to the corresponding identification folder.

Step 2: Create .trf training file, traverse the folder and add the character images to the training file using the `append_ocr_trainf` statement file.

Step 3: Creating classifiers using the `create_ocr_class_svm` statement.

Step 4: Training the classifier using the `trainf_ocr_class_svm` statement.

Step 5: Use the `write_ocr_class_mlp` statement to save the .omc file obtained from the training, so that the OCR model training is complete.

Step 6: Sorting character images using the `sort_region` statement is achieved with the arguments `row`, `true` and `character` character regions in top-to-bottom, left-to-right order.

Step 7: Read the .omc file obtained from training using the `read_ocr_class_svm` statement.

Step 8: Use `do_ocr_multi_class_svm` for character region recognition.

3. Experimental results and analysis

In this paper, we use 150 cigarette box images as the training set, and each image contains about 124 characters. We extract the characters on each cigarette box separately and use the classifier for training. Then we use 50 cigarette box images as the test set to verify the effect of this classifier, and the experimental results are shown in the following table and the recognition results are shown in Figure 8

Table 1. Experimental results of classifier

Classifier	Number of recognized characters	Correct identification	Average confidence level	Average recognition time
GMM	124	115	81.58%	90ms
KNN	124	103	78.61%	86ms
SVM	124	120	97.74%	78ms

We used three different classifiers: SVM, KNN and GMM to experiment and test the data, and found that the recognition rate and processing time of SVM is the highest and the fastest, the recognition rate can reach 96.77%, the average confidence is 97.74%, and the processing time of a single image is 78ms, which is obviously better than other classifiers and meets the requirements of industrial production.

4. Summary

In this paper, we preprocessed the smoke box data, including grayscale variation and binarization. Then the region of interest is extracted, and operations such as threshold segmentation and expansion erosion are done on the region of interest to lay the foundation for the subsequent text training and recognition. Then the principle of SVM and the process of training recognition are introduced, and experiments are conducted on three classifiers, KNN, SVM and GMM, respectively. After experimental analysis, the SVM classifier has the best effect and can meet the needs of enterprise production.

References

- [1] Mori S , Suen C Y , Yamamoto K . Historical review of OCR research and development[J]. IEEE Computer Society Press, 1995. Zhenpei Li, Ping Li, Xinyu Guo. Three GDI+ based image grayscale implementation methods [J]. Computer Technology and Development, 2009, 19(7):4.
- [2] Sauvola J ,Pietik?inen M.Adaptive document image binarization[J]. Pattern Recognition, 2000, 33(2):225-236.
- [3] Xin Qi. " Digital signal processing technology based on expansion corrosion algorithm." Automation Technology and Applications 10(2010):4.
- [4] Chang C C , Lin C J . LIBSVM: A library for support vector machines[J]. ACM Transactions on Intelligent Systems and Technology, 2007, 2(3, article 27).