

## PAPER

# Adaptive Binarization for Vehicle State Images Based on Contrast Preserving Decolorization and Major Cluster Estimation

Ye TIAN<sup>†a)</sup> and Mei HAN<sup>††b)</sup>, *Nonmembers*

**SUMMARY** A new adaptive binarization method is proposed for the vehicle state images obtained from the intelligent operation and maintenance system of rail transit. The method can check the corresponding vehicle status information in the intelligent operation and maintenance system of rail transit more quickly and effectively, track and monitor the vehicle operation status in real time, and improve the emergency response ability of the system. The advantages of the proposed method mainly include two points. For decolorization, we use the method of contrast preserving decolorization [1] obtain the appropriate ratio of R, G, and B for the grayscale of the RGB image which can retain the color information of the vehicle state images background to the maximum, and maintain the contrast between the foreground and the background. In terms of threshold selection, the mean value and standard deviation of gray value corresponding to multi-color background of vehicle state images are obtained by using major cluster estimation [2], and the adaptive threshold is determined by the 2 sigma principle for binarization, which can extract text, identifier and other target information effectively. The experimental results show that, regarding the vehicle state images with rich background color information, this method is better than the traditional binarization methods, such as the global threshold Otsu algorithm [3] and the local threshold Sauvola algorithm [4], [5] based on threshold, Mean-Shift algorithm [6], K-Means algorithm [7] and Fuzzy C Means [8] algorithm based on statistical learning. As an image preprocessing scheme for intelligent rail transit data verification, the method can improve the accuracy of text and identifier recognition effectively by verifying the optical character recognition through a data set containing images of different vehicle statuses.

**key words:** image binarization, contrast preserving decolorization, major cluster estimation, clustering, optical character recognition

## 1. Introduction

As one of the most common steps in digital image processing, image binarization is an extremely important part of image preprocessing in optical character recognition and text analysis and recognition. The background area in the image can be separated by image binarization, which can quickly reduce the interference of useless data and highlight the target area of text and identifier. The accuracy of the result has a decisive effect on text and identifier recognition.

To optimize the effect of image binarization, many methods have been proposed one after another. Traditional threshold calculation methods mainly include global

threshold method and local threshold method. The representative algorithms of the global threshold method are Otsu algorithm [3], Histogram Peaks algorithm [9]–[11] and so on. The global threshold method is simple and applicable to the situation where the background information is single and the histogram of the gray value is bimodal. When the background is complex, the local threshold method is more accurate. Local threshold method includes Niblack algorithm [12], Sauvola algorithm [4], [5] and improved algorithms [13], Bernsen algorithm [14] and LMM algorithm [15] and so on. The local threshold method needs to consider the local different characteristics of the image. According to the size of the sliding window, the best threshold of the local area of the image to be analyzed is determined and then binarization is performed. In this way, the foreground will not be misjudged as background pixel, resulting in the partial missing of foreground text and identifier.

The main goal of image binarization is to distinguish between foreground and background, therefore the problem can be transformed into clustering or classification. The representative clustering methods based on statistical learning mainly include mean shift algorithm [6], K-means algorithm [7], fuzzy c-means algorithm [8]. These clustering algorithms are sensitive to parameter initialization, which affects the stability and accuracy of clustering results and is prone to local convergence.

The color information around the text and identifier is particularly rich and unfixed in the vehicle state images obtained from the intelligent operation and maintenance system of rail transit. For different images, if the RGB ratio is fixed for grayscale, the contrast between the foreground and the background cannot be guaranteed. It is easy to misjudge the foreground as the background by image binarization. To solve this problem, this study adopts the method of contrast preserving decolorization [1] to adjust the ratio of R, G and B for each image adaptively, and grayscale while preserving the contrast of background and foreground. After getting the grayed image, the major cluster estimation method [2] in our previous study is used to get the mean and standard deviation of the background gray value accurately, avoiding the influence of the initial value setting and other parameters selection on the clustering algorithm [6]–[8]. The adaptive binarization threshold is determined by the 2 sigma principle which can be used in histogram multimodal scene. The scheme is used for image binary processing, and then each text, identifier in the vehicle state images are identified. The

Manuscript received October 14, 2021.

Manuscript publicized December 7, 2021.

<sup>†</sup>The author is with the Data and Intelligence R&D Center, Zhuzhou CRRC Times Electric Co., Ltd., Zhuzhou, 412001 China.

<sup>††</sup>The author is with the School of Electrical and Information Engineering, Hunan University of Technology, Zhuzhou, 412007 China.

a) E-mail: tianye@csrzc.com

b) E-mail: hanmei@hut.edu.cn

DOI: 10.1587/transinf.2021EDP7218

recognition accuracy of each category is significantly improved.

The structure of this paper is as follows. Section 2 introduces the characteristics of vehicle state images and the discussion of processing route in detail. Section 3 introduces the principle of global threshold method and local threshold method. In Sect. 4, an adaptive image binarization method based on the combination of contrast preserving decolorization and major cluster estimation is proposed, and the comparison results with the traditional algorithm in real data are provided. Section 5 introduces the recognition of text and identifiers in intelligent operation and maintenance data verification, and compares the impact on recognition rate of text and identifiers with different image binarization methods. Section 6 draws the conclusions of this paper and the prospect of further work.

## 2. Vehicle State Images

### 2.1 Rail Transit Intelligent Operation and Maintenance System

On trains or subways, there were various monitoring devices monitoring the operating status of each device separately, which directly collected the sequence signals of vehicle status for processing. However, these monitoring devices are independent of each other, difficult to transmit signals from vehicle to ground. The data management is not standardized. The efficiency of auxiliary fault emergency treatment and status tracking and hidden danger mining is low. To solve these problems, the rail transit intelligent operation and maintenance system [16] has been set up on the trains or subways of the China Railway Administration to realize the functions of on-board data collection and integration, pre-processing, data analysis, centralized recording, vehicle-to-ground transmission, ground management and so on. The system collects and integrates the operation status signals of all vehicle devices, proofreads the clock, and feeds them back to the display interface of the system instantaneously. The staffs can monitor the operating status of the vehicle devices through the display interface, give immediate warning of possible accidents and improve emergency response capability.

Here, we call the screen shot of the display interface of the rail transit intelligent operation and maintenance system the vehicle state images. The vehicle state images reflect the instantaneous status signals of all operating device of the train or subway. The staffs can combine all the vehicle status signals from the images more intuitively to make vehicle failure warning judgments. Once an abnormal situation is found, the timing signal of the corresponding vehicle device needs to be retrieved for further investigation. Therefore, we need to recognize the character information (text and identifier information) of the fault area from the vehicle state images, and then extract the timing signal of the corresponding device from the on-board database according to the character information for further confirmation.

### 2.2 Characteristics of Vehicle State Images

The vehicle state images obtained from the intelligent operation and maintenance system of rail transit are mainly discussed in this paper. Each type of train or subway has different design forms of vehicle state images as shown in Fig. 1.

In these vehicle state images, we need to pay attention to three kinds of important vehicle state information: region of interest (ROI) corresponding to door, ROI corresponding to train, ROI corresponding to passengers. Taking the upper-left image in Fig. 1 as an example, we pickup the three types of state information respectively in advance to obtain the ROI images as shown in Fig. 9 (a), Fig. 10 (a) and Fig. 11 (a). In ROI images, the Chinese character on the most left side mainly indicates which specific vehicle state the information on the right side corresponds to. We only need to extract the character information in the right box shown in Fig. 7 (a). The combination of these characters and background colors corresponds to the state of one of the vehicle's components or modes of operation. In ROI Images, the target character information we need to recognize is black or white, separated by boxes with different background colors. For the all-black box in Fig. 11 (b), the binarization interference can be solved later by judging that there is no character information in the box.

### 2.3 Points in Dealing with Vehicle State Images

Since most of the target characters on the right side of the ROI images are either black or white, taking black as an illustration (invert the color when white), it is very important to extract the target black characters from the non-solid or solid gradient color background accurately.

Suwa [17] proposed a color-mixing correction method that converts the color of postprints on any background color into that on white for scanned document images. This method is based on two important conditions. First, the background is solid color, and the distribution of a single-color corresponding to background becomes Gaussian in RGB color space because of the heat noise of the system. Next, the distribution of a uniform color pattern corresponding to foreground becomes line/curve-shaped because of the blurring at its boundary. Unless prior information such as the type and distribution of background colors of the target region is known, it is not suitable for the scene that color with highly different properties existing together in a target field.

Edge-based methods usually make use of the edge information of the character strokes of text, and then extract the text in further processing steps. Kasar [18] performed an edge-based connected component analysis and estimated the threshold for each edge connected component with the help of the foreground and the background pixels. In [19], the union of edge information on R, G and B channel was used to generate an edge image. The representative colors

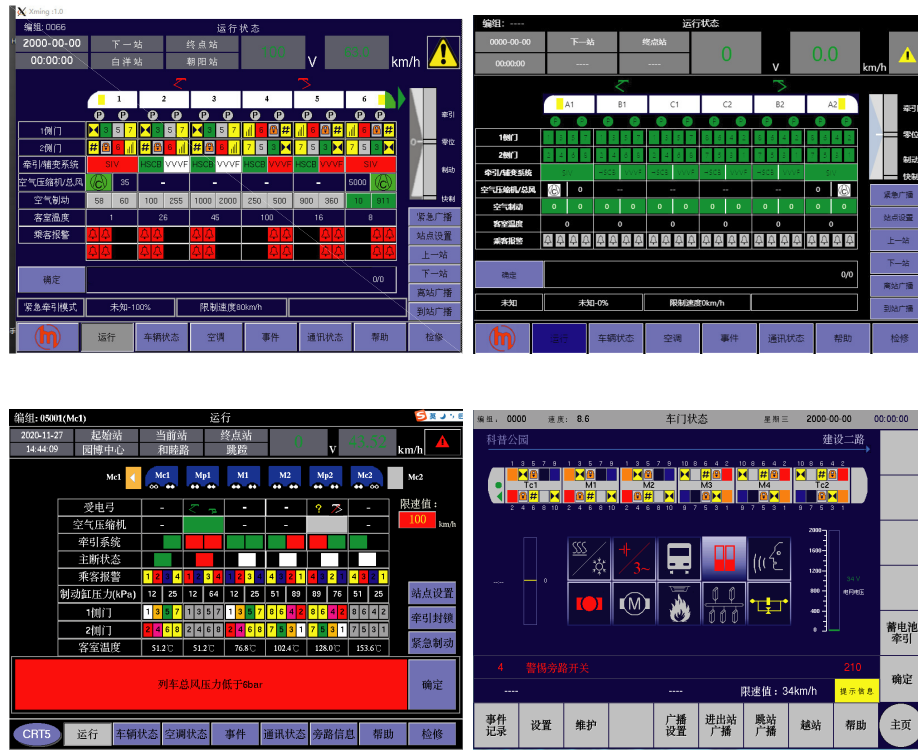


Fig. 1 Vehicle state images.

in CIE  $L^*a^*b^*$  space were obtained along the normal directions of the edge contour. These representative colors served as the initialization of K-means clustering. In each colour cluster, connected component labelling was utilized and several constraints were defined to remove the non-text components. The final binarization for each component is achieved by the threshold estimation similar to that in [18]. This kind of method firstly distinguishes the color types in the image through the text contour information, and then extracts the text information through the color clustering algorithm based on K-means. Therefore, edge detection and color clustering directly affect the accuracy of the algorithm.

In addition, the binarization method of natural scene text segmentation in Zeng [20] based on color clustering assumes that the pixels in the image can be divided into three categories: textual foreground, background and noise in RGB color space. The color segmentation effect of this method depends on the accuracy of two 3-means clustering operations with different distance measurements and the chromaticity differences between textual foreground and background in RGB color space.

However, in the ROI images for this paper, there are background pixel colors that are close to the chromaticity of the target character pixels. Although Zeng [20] uses K-means clustering algorithm twice to solve the impact of classification center initialization, which is better than Kasar [19], the K-means clustering algorithm is based on the assumption that the RGB values of each color cluster are distributed in a high-latitude sphere. Similar to Suwa [17], it is also not suitable for ROI images where the background

pixels have various colors and are difficult to segment by sphere.

The color-based text segmentation method is not suitable for processing the ROI images in this paper and the requirements of real-time. Since the target character color is black, in the histogram of the gray value distribution of ROI images, the gray value corresponding to the target character black (solid color or gradient color) will be close to 0 on the most left side, and the parts corresponding to other colors will be distributed on the right. In considering the conditions above, We choose to grayscale the ROI images and then select the appropriate threshold adaptively for binarization.

### 3. Typical Image Binarization Algorithms

#### 3.1 Global Threshold Method

For an image, select a binarization threshold, traverse each pixel of the image and compare it with the selected threshold, which is the global threshold method. The global threshold method is easy to implement, fast and practical. The global Otsu algorithm [3] is introduced as a representative.

According to the characteristics of image gray value, the Otsu algorithm [3] obtains the optimal threshold value when the variance between classes is the largest, dividing the image into two categories: foreground and background. The image gray value is divided into  $[1, 2, \dots, L]$  gray levels, where  $n_i$  is the number of pixels of each gray level and

the total number of image pixels is  $N = \sum_{i=1}^L n_i$ . The probability of the occurrence of grayscale level is

$$p_i = \frac{n_i}{N}. \quad (1)$$

Assuming the threshold is  $t$ , the image can be divided into foreground (target information) and background according to the threshold value, which is represented by cluster  $C_1$  and  $C_2$  respectively.

The gray level of  $C_1$  is  $[1, t]$  and the gray level of  $C_2$  is  $[t + 1, L]$ . Thus the probability that a pixel is divided into foreground and background are respectively as

$$p_1(t) = P_r(C_1) = \sum_{i=1}^t p_i, \quad (2)$$

$$p_2(t) = P_r(C_2) = \sum_{i=t+1}^L p_i = 1 - p_1(t). \quad (3)$$

The mean pixel gray values of foreground  $C_1$  and background  $C_2$  are respectively as

$$\mu_1(t) = \sum_{i=1}^t i P_r(i|C_1) = \sum_{i=1}^t \frac{i p_i}{p_1(t)} = \frac{1}{p_1(t)} \sum_{i=1}^t i p_i, \quad (4)$$

$$\mu_2(t) = \sum_{i=t+1}^L i P_r(i|C_2) = \frac{1}{p_2(t)} \sum_{i=t+1}^L i p_i. \quad (5)$$

The average gray value of the whole image can be expressed as

$$\mu = \sum_{i=1}^L i p_i. \quad (6)$$

The between-class variance is

$$\sigma^2(t) = \mu_1(t)[\mu_1(t) - \mu]^2 + \mu_2(t)[\mu_2(t) - \mu]^2. \quad (7)$$

Then the optimal binarization threshold is

$$\sigma^2(\text{Threshold}) = \max(\sigma^2(t)), \quad 1 \leq t \leq L. \quad (8)$$

The global Otsu algorithm [3] mainly uses the normalized histogram to select the segmentation threshold which is simple to operate and widely used in image processing.

### 3.2 Local Threshold Method

Local threshold method considers the different characteristics of the local region of the image, and selects different optimal threshold in different regions for segmentation. This algorithm needs to determine the local region of the image to be processed according to the size of the sliding window. Therefore the size of the sliding window has a great influence on the effect of the local threshold method. Here is a brief introduction with as a representative. The local threshold Sauvola algorithm [5] is introduced briefly as a representative. The algorithm is improved based on Niblack algorithm [12]. Taking the pixel  $(x, y)$  as center, the corresponding threshold value of the point is

$$\text{Threshold}(x, y) = m(x, y) \left[ 1 - \frac{k(1 - s(x, y))}{R} \right], \quad (9)$$

where  $m(x, y)$  represents the average gray value of all pixels in the sliding window centered on pixel  $(x, y)$ , and  $s(x, y)$  is the standard deviation. Here  $k$  represents the correction coefficient, which needs to be adjusted according to the actual image, and the value range is  $(0, 1)$ .  $R$  is the dynamic range of the standard deviation. We usually use  $R = 128$  with 8-bit gray level images. In local threshold Sauvola algorithm [5], the size of the sliding window and the correction coefficient  $k$  need to be preset in advance, which has a great influence on the calculation efficiency of the algorithm and the result of binarization. The algorithm and the global Otsu algorithm [3] will be discussed in detail in Sect. 4.

## 4. Proposed Method and Its Experimental Verification

### 4.1 Grayscale and Contrast Preserving Decolorization [1]

Decolorization is the process of converting a color image into a gray image. According to the original color information of the image, the ratio of R, G and B is adjusted adaptively, and the original color contrast of the image is kept as much as possible with limited gray scale range. The contrast preserving decolorization [1] is introduced to perform image grayscale. The method adopts bimodal distribution to constrain the spatial pixel difference, and selects the appropriate gray level to maintain the original contrast automatically alleviating the strict order constraint of color mapping based on human vision system.

Contrast preserving decolorization [1] adopts bimodal distribution to select color order automatically relaxing the constraint of original color order, thus the objective energy function  $E(g)$  can be obtained as

$$E(g) = - \sum_{(p,q) \in \mathcal{P}} \ln \{ \mathcal{N}_\sigma(\Delta g_{p,q} + \delta_{p,q}) + \mathcal{N}_\sigma(\Delta g_{p,q} - \delta_{p,q}) \}, \quad (10)$$

where the gray value difference of pixel point  $p$  and pixel point  $q$  in the set of pixel points  $\mathcal{P}$  is  $\Delta g_{p,q} = g_p - g_q$ . The Gaussian distribution  $\mathcal{N}_\sigma$  has a single mode peaked at  $\delta(p, q)$  and the standard deviation  $\sigma$  is usually set as 0.05. In RGB space, the module length of  $\delta_{p,q}$  corresponds to the Euclidean distance of pixel pair  $(p, q)$ , namely

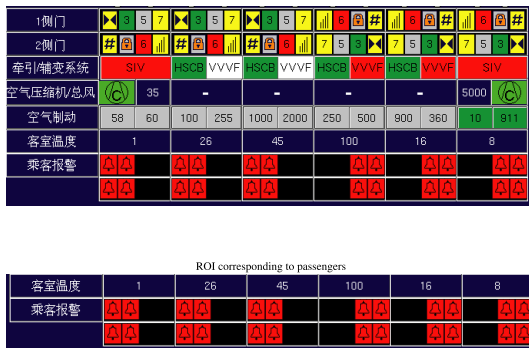
$$|\delta_{p,q}| = \sqrt{(R_p - R_q)^2 + (G_p - G_q)^2 + (B_p - B_q)^2}. \quad (11)$$

The gray value is

$$g = \omega_r R + \omega_g G + \omega_b B, \quad (12)$$

where  $R$ ,  $G$  and  $B$  are three-channel input of RGB, and  $\omega_r$ ,  $\omega_g$  and  $\omega_b$  are the corresponding proportions of color components of R, G, B. The proportion satisfies the conditions:

$$\omega_r \geq 0, \quad \omega_g \geq 0, \quad \omega_b \geq 0, \quad (13)$$



**Fig. 2** An example of vehicle state images and ROI corresponding to passengers.

$$\omega_r + \omega_g + \omega_b = 1. \quad (14)$$

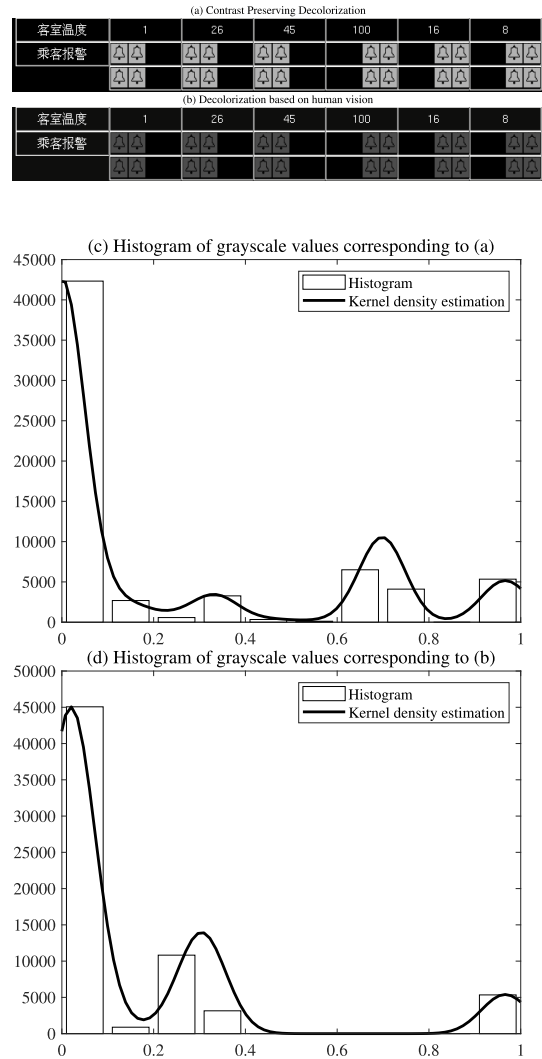
when the objective energy function  $E(g)$  reach the minimum, the value of  $\omega_r$ ,  $\omega_g$  and  $\omega_b$  is the most suitable grayscale proportioning proportion, which can preserve the original color contrast of the image to the utmost extent.

Figure 2 shows the vehicle state images and ROI corresponding to passengers obtained from the intelligent operation and maintenance system of rail transit. Taking image of ROI corresponding to passengers as an example, the images are grayed with fixed ratio based on human vision and decolorized by contrast preserving decolorization [1] respectively. The sensitivity of human eyes to green is the highest, while blue is the lowest. Take the weighted average for RGB three channel input as follow, the fixed proportion gray value can be calculated as:

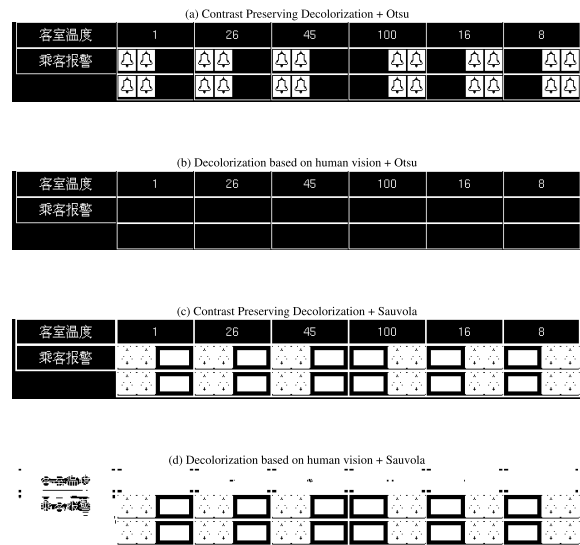
$$g = 0.299R + 0.587G + 0.114B. \quad (15)$$

The grayscale results of ROI corresponding to passengers and corresponding gray frequency histogram in the two methods are shown in Fig. 3. In Fig. 3 (a), the contrast between the bell in the passenger alarm area and the background is preserved, so the bell is clearly visible to see. However, in Fig. 3 (b), the difference between the bell and the background is small, it is difficult to distinguish the bell. Kernels density estimation [20] is performed for the gray values. It can be seen that in the ROI corresponding to passengers, after contrast preserving decolorization [1], the black part representing the bell and the red part of the background have a large peak interval between the two parts, as shown in Fig. 3 (c). The binarization threshold can be easily determined. However, in Fig. 3 (d), after grayed with fixed ratio based on human vision, the peaks corresponding to the two parts of the bell and the background are very close. It can be expected that the traditional image binarization method is difficult to get the appropriate binarization threshold.

Here, the global Otsu algorithm [3] and the local Sauvola algorithm [5] are respectively used to binarize Fig. 3 (a) and Fig. 3 (b), and the binarization results are shown in Fig. 4. The parameter setting in the local Sauvola algorithm is as follows:  $11 \times 11$  domain, the correction



**Fig. 3** Comparison between contrast preserving decolorization and decolorization based on human vision.



**Fig. 4** Comparison between Otsu [3] and Sauvola [5] after contrast preserving decolorization and decolorization based on human vision.

coefficient  $k = 0.9$ , and standard deviation dynamic range  $R = 128$ . It is obvious that the result of binarization on the basis of contrast preserving colorization [1] is obviously better than that of fixed ratio grayscale based on human vision. Moreover, in this case, for ROI corresponding to passengers, the binarization results of the global Otsu algorithm [3] can preserve the target information of the bell better compared with the local Sauvola algorithm [5]. We can conclude that when the image is grayscale, it is necessary to preserve the color contrast of the image which will affect the effect of image binarization directly.

#### 4.2 Threshold Selection and Major Cluster Estimation [2]

In Sect. 4.1, the importance of preserving contrast has been verified in grayscale.

After contrast preserving decolorization, if the proportion of target black pixels in the whole pixels is high, the gray value of the image will show a bimodal distribution, and the bimodal has a certain interval, which can distinguish the target black pixels from the background better.

But when there are only a few target black pixels in the image, the bimodal distribution is not obvious. If traditional image binarization methods such as global Otsu algorithm [3] are used, the target black pixels and black-like pixels (blue and red) will be confused and cannot be distinguished effectively. In this case, since the target black pixels account for a small proportion, we can regard them as the outliers, while the background pixels can be regarded as a major cluster. Assuming that the major cluster follows a Gaussian probability distribution, if we can estimate the gray value of background pixels, namely, the mean  $\mu_g$  and standard deviation  $\sigma_g$  of main cluster, the 2 sigma principle can be used to determine the range of gray value of main cluster pixels. Tian and Yokota [2] proposed a new mean-shift to estimate the major cluster by updating the Gaussian kernel based on the assumption that the outliers do not affect the sample mode of the major cluster. The major cluster estimation method [2] can estimate the mean and standard deviation of the background gray value accurately when outliers exist. Therefore, the binarization threshold can be selected adaptively based on the main cluster estimation [2]. As the target pixel is black, it is on the left side of the background pixel of the main cluster in the gray value histogram. Considering the allowable error limit of 5%, the binarization threshold can be set as  $\mu_g - 2\sigma_g$  shown as dotted line in Fig. 8 (a), in which the solid line represents major estimation, and the rectangle represents histogram of gray values.

When the proportion of target black pixels in all pixels is high, the result of main cluster estimation will be affected which will correspond to the target black pixels, as shown in Fig. 6 (a). At this time, the mean and standard deviation of the target black pixel are estimated, and the binarization threshold should be set to  $\mu_g + 2\sigma_g$ . Since the estimated value of the mean of the target black pixel  $\mu_g$  must be close to 0, the binarization threshold can be set to  $-(\mu_g - 2\sigma_g)$  equivalently, as shown by the dotted line in Fig. 6 (a).

Therefore, whether the mean and standard deviation obtained by the main cluster estimation [2] correspond to the target black pixel or the background pixel, the binarization threshold can be set as  $|\mu_g - 2\sigma_g|$ , which can segment the target black pixel and the background pixel well.

In view of the discussions above, here we propose a threshold selection method based on major cluster estimation [2]. The algorithm is summarized as presented below:

1. Set the mean value  $\mu_g$  and standard deviation  $\sigma_g$  of the gray value of all pixels in the image as the initial values of the mean  $\hat{\mu}_N$  and standard deviation  $\hat{\sigma}_N$  of the gray value of the major cluster respectively as:

$$\hat{\mu}_N \leftarrow \mu_g, \quad (16)$$

$$\hat{\sigma}_N \leftarrow \sigma_g. \quad (17)$$

Where the gray value set of all pixels in the image is  $\{g_1, g_2, \dots, g_N\}$ .

2. Take a Gaussian distribution with mean  $\mu_W$  and standard deviation  $\sigma_W$  as the kernel function  $p(g; \mu_W, \sigma_W) = e^{-\frac{\|g - \mu_W\|^2}{2\sigma_W^2}}$ . Here, the mean  $\mu_W$  and the standard deviation  $\sigma_W$  are given respectively by the estimated value  $\hat{\mu}_N$  of the mean and the estimated value  $\hat{\sigma}_N$  of the standard deviation of the major cluster:

$$\mu_W \leftarrow \hat{\mu}_N, \quad (18)$$

$$\sigma_W \leftarrow r\hat{\sigma}_N. \quad (19)$$

Where  $r$  is the scale factor and the empirical value is 0.75. Here, the mean  $\mu_W$  and the standard deviation  $\sigma_W$  of the Gaussian kernel are not estimators, although they change when the kernel updates.

3. We can get the mean  $\mu_g$  and standard deviation  $\sigma_g$  of all pixels in the image through such a Gaussian kernel  $p(g; \mu_W, \sigma_W)$ :

$$\mu_g = \sum_{n=1}^N a_n g_n, \quad (20)$$

$$\sigma_g = \sqrt{\sum_{n=1}^N a_n (g_n - \mu_g)^2}. \quad (21)$$

Where  $a_n$  is the weight of the gray value  $g_n$ :

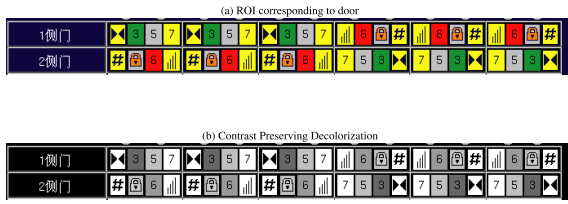
$$a_n = \frac{p(g_n; \mu_W, \sigma_W)}{\sum_{k=1}^N p(g_k; \mu_W, \sigma_W)}. \quad (22)$$

4. The mean value  $\hat{\mu}_N$  and standard deviation  $\hat{\sigma}_N$  of the major cluster are updated by the following equations:

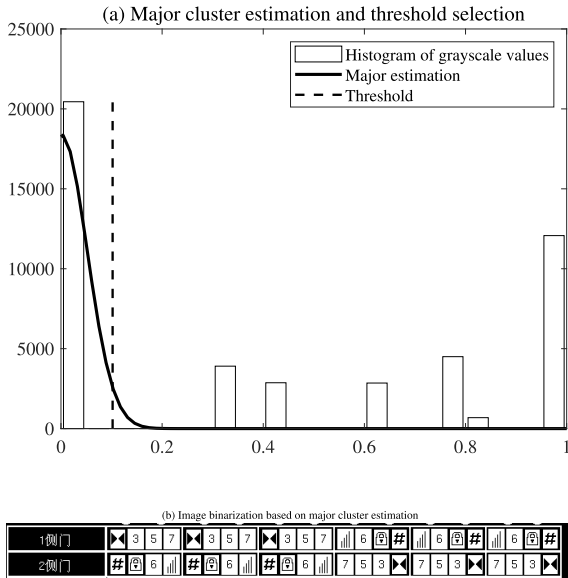
$$\hat{\mu}_N \leftarrow \mu_g, \quad (23)$$

$$\hat{\sigma}_N \leftarrow \sqrt{\frac{\sigma_W^2 \sigma_g^2}{\sigma_W^2 - \sigma_g^2}}. \quad (24)$$

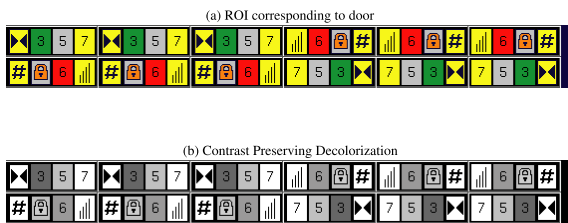
5. If the variation of the values of these estimators are



**Fig. 5** ROI corresponding to door and result of contrast preserving decolorization.



**Fig. 6** Threshold selection and image binarization result based on major cluster estimation corresponding to Fig. 5 (b).



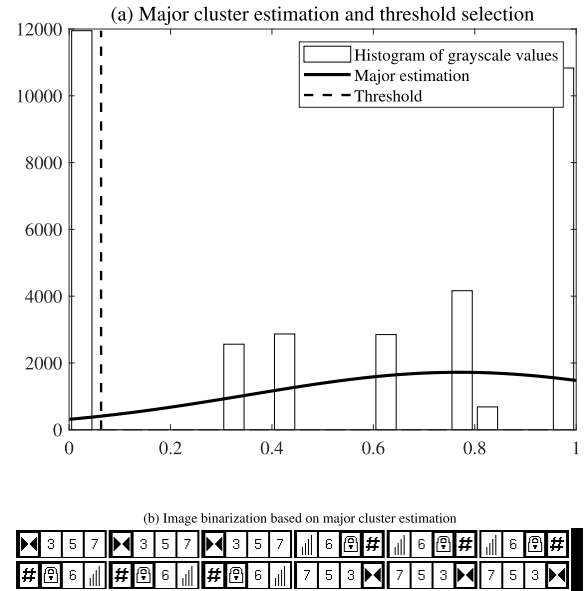
**Fig. 7** Text and identifier of ROI corresponding to door and result of contrast preserving decolorization.

equal to or less than the predetermined fixed value, then the update process is terminated. Otherwise, return to 2 and repeat the iteration.

6. Within the allowable error limit of 5%, the gray value binarization threshold is set as:

$$\text{Threshold} = |\hat{\mu}_N - 2\hat{\sigma}_N|. \quad (25)$$

As shown in Fig. 5 and Fig. 7, taking the ROI corresponding to door as an example, on the basis of contrast preserving decolorization results, we calculate the major cluster estimation of the main part of the image gray values, and then determine the threshold value for image binarization. The result of image binarization is shown in Fig. 6 (b) and Fig. 8 (b). Figure 5 shows full image of ROI corresponding



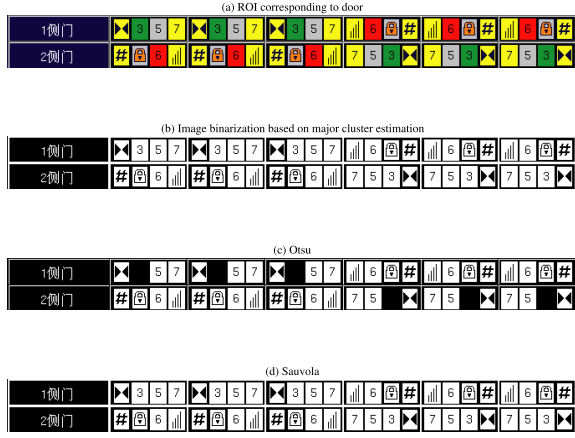
**Fig. 8** Threshold selection and image binarization result based on major cluster estimation corresponding to Fig. 7 (b).

to door, Fig. 7 shows the image of text and identifiers of the ROI. In Fig. 6 (a), from the major cluster estimation result, we can see the mean value is  $-0.0016$  and standard deviation is  $0.0515$ . The major cluster estimation result is in the left area of the histogram, which means the main cluster of gray values is black target pixels. Here, threshold should be  $\mu + 2\sigma = -0.0016 + 2 \times 0.0515 = 0.1014$ . However, since the mean  $\mu$  approaches 0, and the standard deviation must be greater than or equal to 0,  $\mu + 2\sigma \approx -\mu + 2\sigma = -(\mu - 2\sigma) = |\mu - 2\sigma|$  shown as Eq. (25). Thus, for Fig. 6 (a), according to Eq. (25), threshold =  $|\mu - 2\sigma| = -(-0.0016 - 2 \times 0.0515) = 0.1046$ , which is close to  $0.1014$ . In Fig. 8 (a), the mean value is  $0.7597$  and standard deviation is  $0.3512$ . Due to the reduction of black or black-like target information, the gray value of the image is mainly concentrated on the right side of the histogram. According to Eq. (25), Threshold =  $\mu - 2\sigma = 0.7597 - 2 \times 0.3512 = 0.0573$ , the selected threshold can distinguish text, identifier and other target information well.

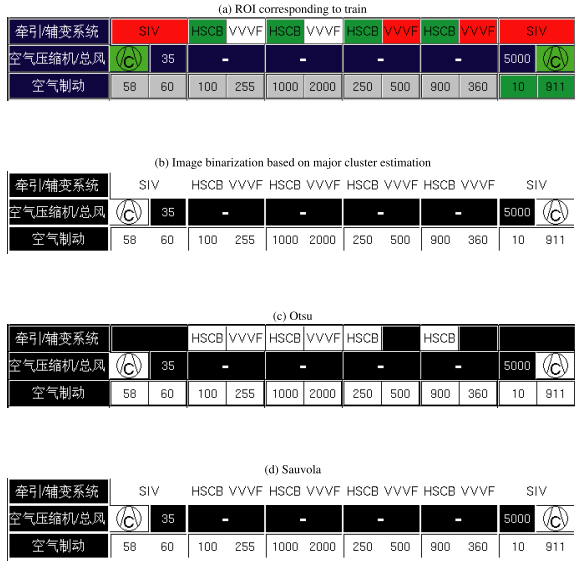
### 4.3 Comparison between Proposed Method and Typical Image Binarization Algorithms

Taking the ROI corresponding to door, train and passengers as examples, after contrast preserving decolorization, the results are binarized with major cluster estimation [2], the global Ostu algorithm [3] and the local Sauvola algorithm [5] separately. The parameter setting in the local Sauvola algorithm is as follows:  $11 \times 11$  domain, the correction coefficient  $k = 0.9$ , and standard deviation dynamic range  $R = 128$ . The comparison of binarization results is shown in Fig. 9, Fig. 10, and Fig. 11, respectively corresponding to the ROI of door, train and passengers.

In Fig. 9, for ROI corresponding to door, the global Ostu algorithm [3] cannot distinguish the number 3 from



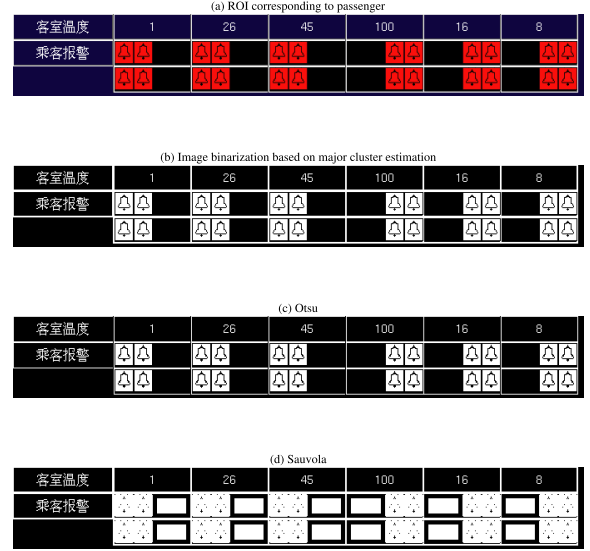
**Fig. 9** ROI corresponding to door and the binarization results with different methods after contrast preserving decolorization.



**Fig. 10** ROI corresponding to train and the binarization results with different methods after contrast preserving decolorization.

the green background, while the other two methods can clearly extract the target information of text and identifier. The same situation occurs in ROI corresponding to vehicle in Fig. 10. The global Otsu algorithm [3] cannot distinguish text information from red background, such as SIV and VVVF. However, in ROI corresponding to passengers in Fig. 11, the major cluster estimation [2] and the global Otsu algorithm [3] can separate the bell information from the red background, while the local Sauvola algorithm [5] can not extract the bell information. In this case, the local Sauvola algorithm [5] needs to adjust the setting value of the correction coefficient  $k$  reversely according to the binarization result, which is difficult to be adaptive and has little significance in practical application.

It can be seen from the comparison results that the selection of binarization threshold using the major cluster estimation [2] does not depend on the preset parameters,



**Fig. 11** ROI corresponding to passengers and the binarization results with different methods after contrast preserving decolorization.



**Fig. 12** Binarized image samples with target information including text (number) and identifier.

and can completely extract target information such as the text and identifier of each ROI in the vehicle state images, which is superior to other traditional binarization threshold selection methods and has wider practicability.

## 5. Text and Identifier Recognition in Intelligent Operation and Maintenance Data Checking

On the basis of contrast preserving decolorization [1], a series of binary images including text and identifier can be obtained by using the image binarization method described above from the vehicle state images. Since the sample with classification is single, we use the SVM method [22] to classify the training samples, train the model and test the test set. The binarized image sample is shown in Fig. 12.

Since the size of small identifier in each vehicle state image is inconsistent, it needs to be resized to a unified size before classification. In order to cover all kinds of small identifiers in vehicle state images, ensure certain classification accuracy and reduce the time of training and classification, the sample images are unified into  $28 \times 28$  size and tiled in a row making sure that the number of samples of each category is not less than 50. Classify the training samples with the SVM [22], and calculate the maximum interval hyperplanes of different classes, in which the linear kernel function is adopted as the kernel function. Taking digital information as an example, the tiled row vector is shown in Fig. 12.

We apply different image binarization methods as im-

**Table 1** Comparison results of recognition accuracy based on different image binarization methods.

Vehicle state images \ Binarization method	Otsu [3]	Sauvola [5]	Proposed method
1	84.55%	85.79%	<b>90.02%</b>
2	94.01%	92.37%	<b>96.88%</b>
3	96.72%	96.45%	<b>99.26%</b>
4	68.66%	69.43%	<b>71.13%</b>

age preprocessing schemes, and analyze the recognition accuracy of text and identifier in the four vehicle state images shown in Fig. 1. The comparison results of the average accuracy of all categories are shown in Table 1. The recognition accuracy is the average value of TPR (True Positive Rate) corresponding to each identifier. From Table 1, we can see that the quality of image binarization results will directly affect the recognition accuracy of subsequent text and identifiers. The proposed image binarization method based on contrast preserving decolorization [1] and major cluster estimation [2] can completely extract the information of text and identifier in the vehicle state images, and is also the best in recognition accuracy of the subsequent optical character, which is higher than the global Otsu algorithm [3] and local Sauvola algorithm [5]. The parameter setting in the local Sauvola algorithm is as follows:  $11 \times 11$  domain, the correction coefficient  $k = 0.9$ , and standard deviation dynamic range  $R = 128$ .

## 6. Discussion

From the perspective of image grayscale and binarization, this paper proposes an image binarization method based on contrast preserving decolorization [1] and major cluster estimation [2]. The method can effectively extract the target information such as text and identifier from the background with rich color information, thus effectively improving the recognition accuracy of text and identifier. The experimental results confirm the effectiveness of the approach. But the limit of this method is clear that the target pixel in the image must be black or white (solid or gradient color). In view of this point, it limits the universality of the proposal method in this paper.

The application of the adaptive binarization method in degenerated document image, such as solving the influence of uneven illumination, character blur, background color leakage penetration, paper aging, paper crease and other problems, needs to be compared and confirmed in the future. In addition, due to the particularity of the industry, there is no one-to-one corresponding standard binary image data set with truth value image. For the sake of engineering, the comparison of binarization algorithms based on deep learning neural network [23], [24] is not discussed in this paper. The basic idea of this algorithm is to learn the characteristics of image binarization directly from the data set. It is necessary to prepare a massive data set. According to the characteristics of the data set and the desired artificial synthesis results, select the appropriate neural network structure, get the required model through training, and then

use the model to process the image data. Such algorithms require huge data sets to get the desired model. Due to the fact that the truth value images are difficult to obtain, and the model relies on data sets seriously, the generalization ability is not good. These issues will be discussed in-depth in the future.

## References

- [1] C. Lu, L. Xu, and J. Jia, "Real-time contrast preserving decolorization," SIGGRAPH ASIA Posters, Article No.16, 2012.
- [2] Y. Tian and Y. Yokota, "Estimating the major cluster by mean-shift with updating kernel," *Mathematics*, vol.7, no.9, 771, pp.1–25, 2019.
- [3] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man Cybern.*, vol.9, no.1, pp.62–66, 1979.
- [4] G. Lazzara and T. Géraud, "Efficient multiscale Sauvola's binarization," *Int. J. Doc. Anal. Recognit.*, vol.17, pp.105–123, 2014.
- [5] J. Sauvola and M. Pietikäinen, "Adaptive document image binarization," *Pattern Recognition*, vol.33, no.2, pp.225–236, 2000.
- [6] Y.A. Ghassebeh and F. Rudzicz, "Modified mean shift algorithm," *IET Image Processing*, vol.12, no.12, pp.2172–2177, 2018.
- [7] J.A. Hartigan and M.A. Wong, "A  $K$ -means clustering algorithm," *Applied Statistics*, vol.28, no.1, pp.100–108, 1979.
- [8] K.-S. Chuang, H.-L. Tzeng, S. Chen, J. Wu, and T.-J. Chen, "Fuzzy c-means clustering with spatial information for image segmentation," *Computerized Medical Imaging and Graphics*, vol.30, no.1, pp.9–15, 2006.
- [9] Y. Chen and L. Wang, "Broken and degraded document images binarization," *Neurocomputing*, vol.237, pp.272–280, 2017.
- [10] P. Lech and K. Okarma, "Fast histogram based image binarization using the Monte Carlo threshold estimation," *ICCVG 2014, Computer Vision and Graphics, Lecture Notes in Computer Science*, vol.8671, pp.382–390, Springer, Cham, 2014.
- [11] C.A. Glasbey, "An analysis of histogram-based thresholding algorithms," *CVGIP: Graphical Models and Image Process*, vol.55, no.6, pp.532–537, 1993.
- [12] O.A. Samorodova and A.V. Samorodov, "Fast implementation of the Niblack binarization algorithm for microscope image segmentation," *Pattern Recognition and Image Analysis*, vol.26, pp.548–551, 2016.
- [13] B. Bataineh, S.N.H.S. Abdullah, and K. Omar, "An adaptive local binarization method for document images based on a novel thresholding method and dynamic windows," *Pattern Recognition Letters*, vol.32, no.14, pp.1805–1813, 2011.
- [14] J. Bernsen, "Dynamic thresholding of gray-level images," *Proc. International Conference on Pattern Recognition*, Paris, 1986.
- [15] B. Su, S. Lu, and C.L. Tan, "Binarization of historical document images using the local maximum and minimum," *IAPR International Workshop on Document Analysis Systems, DAS 2010, Boston, Massachusetts, USA. DBLO*, pp.159–166, 2010.
- [16] M. He, H. Li, and Y. Duan, "Research on railway intelligent operation and maintenance and its system architecture," *2019 6th International Conference on Dependable Systems and Their Applications (DSA)*, pp.488–490, 2020.
- [17] M. Suwa, "Color-mixing correction of overlapped colors in scanner images," *Proc. ICDAR 2011*, pp.217–221, 2011.
- [18] T. Kasar, J. Kumar, and A.G. Ramakrishnan, "Font and background color independent text binarization," *2nd CBDAR*, pp.3–9, 2007.
- [19] T. Kasar and A.G. Ramakrishnan, "COCOCLUST: Contour-based color clustering for robust binarization of colored text," *Proc. 3rd CBDAR*, pp.11–17, 2009.
- [20] C. Zeng, W. Jia, and X. He, "An algorithm for colour-based natural scene text segmentation," *M. Iwamura and F. Shafait (eds.), CBDAR 2011, LNCS 7139*, pp.58–68, 2012.
- [21] B.W. Silverman, *Density Estimation for Statistics and Data Analy-*

- sis, Chapman and Hall, New York, USA, 1986.
- [22] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," *Journal of Machine Learning Research*, vol.2, no.1, pp.45–66, 2001.
  - [23] G. Meng, K. Yuan, Y. Wu, S. Xiang, and C. Pan, "Deep networks for degraded document image binarization through pyramid reconstruction," *ICDAR*, pp.727–732, 2017.
  - [24] Q.N. Vo, S.H. Kim, H.J. Yang, and G. Lee, "Binarization of degraded document images based on hierarchical deep supervised network," *Pattern Recognition*, vol.74, pp.568–586, 2018.



**Ye Tian** was born in Hubei, China, on October 9, 1990. He received B.S. degree in Applied Mathematics from Southeast University, Nanjing, China, in 2012, M.S. and Dr.Eng. degrees in Information Science from Gifu University, Gifu, Japan, in 2016 and 2019, respectively. He is now working in Data and Intelligence R&D Center, Zhuzhou CRRC Times Electric Co., Ltd., Hunan, China. His interests are in image processing, signal processing, information theory, and their applications to rail transit.



**Mei Han** was born in Hubei, China, on December 27, 1989. She received B.S. degree in Communication Engineering from Wuhan University of Technology, Wuhan, China, in 2012, M.S. and Dr.Eng. degrees in Electronics and Information Engineering from Gifu University, Gifu, Japan, in 2016 and 2019, respectively. She is now a lecturer at School of Electrical and Information Engineering, Hunan University of Technology, Hunan, China. Her interests are in low-power VLSI design and image processing.