

Evaluation of Binarization Algorithms for Camera-Based Devices

M. Nava-Ortiz, W. Gómez-Flores, A. Díaz-Pérez, and G. Toscano-Pulido

Information Technology Laboratory, CINVESTAV-IPN, Ciudad Victoria, Mexico
mnava@tamps.cinvestav.mx

Abstract. Segmentation is an important step within optical character recognition systems, since the recognition rates depends strongly on the accuracy of binarization techniques. Hence, it is necessary to evaluate different segmentation methods for selecting the most adequate for a specific application. However, when gold patterns are not available for comparing the binarized outputs, the recognition rates of the entire system could be used for assessing the performance. In this article we present the evaluation of five local adaptive binarization methods for digit recognition in water meters by measuring misclassification rates. These methods were studied due to of their simplicity to be implemented in based-camera devices, such as cell phones, with limited hardware capabilities. The obtained results pointed out that Bernsens method achieved the best recognition rates when the normalized central moments are employed as features.

Keywords: local adaptive binarization, optical character recognition, camera-based devices, feature selection.

1 Introduction

Optical Character Recognition (OCR) has been widely studied for decades, specially in document image analysis. In many situations only digit characters are required to be recognized. For instance, a common use of OCR is recognizing numbers in car plates [1,2]. Another useful application developed for helping persons who are blind or have low vision was presented by Shen and Coughlan [3]. The authors implemented an OCR algorithm for camera cell phones that read aloud the numbers presented in LCD/LED displays.

As one can note, the availability of cheap portable devices with suitable computation power makes possible to process images in real time. Also, the built-in cameras in mobile devices (e.g. smart phones) give additional capabilities to implement OCR algorithms.

However, the task of recognizing characters acquired with cameras is not trivial, since uncontrolled environment variables, such as uneven illumination and shadows, leads low quality images. Moreover, compared to general desktop environment, most camera-based mobile devices have limitations of low computing

power and insufficient storage space. Also its CPUs are only integer processor and therefore the floating point operations must be simulated in integer arithmetic operations, degrading processing rate seriously.

The paradigm of image processing can be depicted in terms of five basic steps: (i) image acquisition, (ii) image preprocessing (noise filtering and enhancement), (iii) image segmentation, (iv) feature extraction and selection, and (v) object recognition [7]. For assessing the performance of each step, several authors have been proposed techniques for measuring the quality of their algorithms. For example, Tin and Kamata [4] presented an iterative image enhancement algorithm and developed a new evaluation framework based on objective and subjective quality image measures. Besides, Thulke et al. [5] proposed a general evaluation approach for comparing document segmentation algorithms based on the segments directly. Obviously, these evaluation approaches need to define the range of values that represent the levels of efficiency.

The image segmentation is a critical step within the recognition process. It is convenient to use references to compare the output of the segmentation algorithm for evaluating its performance. However, when the references are not available, the recognition rates of the complete system could be used for evaluating the segmentation performance. Trier and Jain [6] used this approach to evaluate recognition accuracy of an OCR system, by comparing different binarization methods.

Numerical meters, such as wattmeters or water meters, are devices used to measure the amount of some service commonly supplied by a public utility. Generally, the records from these devices are captured manually by employees. However, this procedure may cause wrong readings due to human errors. To overcome this inconvenient, an OCR system could be implemented in camera-based devices handled by public utility employees. In order to deal with the hardware limitations and uncontrolled environment variables, it is necessary to develop adaptive methods that do not demand a lot of computational resources.

In this article we present the evaluation of five local adaptive binarization methods for digit recognition in water meters by measuring misclassification rates. The methodology involves the five steps in image processing paradigm for object recognition. First, the photograms coming from two brands of water meters were cropped to separate each digit to create the image data set. Next, five local adaptive binarization methods were applied to the entire dataset. Thereafter, seven scale invariant moments were calculated from the segmented images. Besides, a feature selection method based on mutual information and intrinsic dimensionality was applied to the space of characteristics. This procedure took the first three moments ranked with maximum-relevance-minimum-redundance (MRmR) criteria [16]. Moreover, the digit recognition was performed by a minimum distance classifier. Finally, we measured the recognition rates for determining which is the most adequate binarization method for this particular application. Additionally, we made a comparison using the selected features and all of them.

2 Materials and Methods

2.1 Binarization Methods

An image with size $M \times N$ can be represented by a 2D gray-level intensity function $f(x, y)$, within the range 0 to $L - 1$, where L is the maximum number of gray-levels. A binarization method creates bilevel images from $f(x, y)$ by turning all pixels below some threshold to zero (black) and all pixels equal to or greater than that threshold to one (white). This procedure separates the image into two classes: background and foreground. The main problem is how to select the optimal or most adequate threshold. In general, the binarization methods could be classified into two main approaches: global and local adaptive [7]. The former attempts to find a single threshold value for the overall image, whereas the latter computes a threshold value for each pixel regarding its neighborhood information. Due to local methods are able to cope with uneven illumination, we tested five local binarization method depicted bellow. The objective is to evaluate the impact of the each segmentation method in the recognition process. All the algorithms depicted in this article were developed in Matlab 7.10 (The MathWorks Inc., Natick, Mass, USA).

Bernsen's Method. [8] determines the threshold for the pixel (x, y) as the average between the minimum and maximum gray level pixel values in a square $b \times b$ neighborhood centered at (x, y) . However, if the contrast $C(x, y) = I_{max}(x, y) - I_{min}(x, y)$ is lower than a predetermined threshold t , then the pixel (x, y) is labeled as background. The threshold is calculated as follows:

$$T(x, y) = (I_{min} + I_{max}) / 2 . \quad (1)$$

Niblack's Method. [9] obtains the threshold values by calculating the local mean and local standard deviation. The threshold at pixel (x, y) is calculated as:

$$T(x, y) = m(x, y) + k \cdot s(x, y), \quad (2)$$

where $m(x, y)$ and $s(x, y)$, are the mean and standard deviation, respectively, in a local neighborhood of pixel (x, y) . It is necessary define the size of the neighborhood small enough in order to preserve local details, and large enough to suppress noise. The value of k is used to adjust how much of the total print object boundary is taken as a part of the given object.

Sauvola's Method. is an improvement to Niblack's method [10]. It attempts to efficiently reduce the non-homogeneous illumination in image. This algorithm is not sensitive to the value of k parameter. Thus, the threshold at pixel (x, y) is computed as:

$$T(x, y) = m(x, y) + \left[1 + k \cdot \left(\frac{\sigma(x, y)}{R} - 1 \right) \right], \quad (3)$$

where $m(x, y)$ and $s(x, y)$ are the mean and standard deviation, respectively. The R value is the dynamic range of the standard deviation and for this study was set as 8. The k parameter is positive.

Wellner’s Method. [11] applies image smoothing to the input image by using an average filter with a window size $b \times b$. Next, the threshold pixel (x, y) is calculated as:

$$T(x, y) = J(x, y) \cdot \left[1 - \left(\frac{t}{100} \right) \right], \tag{4}$$

where $J(x, y)$ is the filtered image and t is the predetermined threshold that scales each gray value of the filtered image to a lower value.

White’s Method. [12], each gray value at pixel (x, y) is compared with its neighborhood average value, in a square $b \times b$. If pixel (x, y) is significantly darker than its neighborhood mean, then is classified as foreground, otherwise it is classified as background as follows:

$$B(x, y) = \begin{cases} 1 & \text{if } m_{b \times b}(x, y) < I(x, y) \cdot w \\ 0 & \text{otherwise} \end{cases}, \tag{5}$$

where $I(x, y)$ is the original image, $m_{b \times b}(x, y)$ is the local mean, and $w > 1$ is the *bias* value.

2.2 Image Acquisition

The dataset includes images captured from two brands of water meters. The NOKIA N80 cell phone was used to acquired the 8-bit images. Moreover, all the images were taken directly from the electronic viewfinder to speed up the image time-processing. As the images presented complete lectures (Fig. 1) , it was necessary to crop the digits for separating each number in categories, that means, 0 to 9. Then, the size of each single image containing a number was 19×30 pixels. The total number of images within the dataset was 1,418.



Fig. 1. Examples of images captured from two brands of water meters

2.3 Image Binarization

The entire dataset was binarized by the five thresholding methods depicted previously: Bernsen [8], Niblack [9], Sauvola [10], Wellner [11], and White [12]. There was not necessary to preprocess the images before segmentation procedure, since the local adaptive methods are able to suppress some amount of noise. Despite

each segmentation method is performed as local adaptive, they depend on the tuning of a global parameter that will be applied indistinctly to overall pixels. For instance, Niblack's and Sauvola's methods depends on k parameter, Bernsen's and Wellner's methods depends on t threshold parameter, and White's method depends on a w bias value. If we see all these variables as a single parameter to be tuned, we can denote all of them as K . We propose to adapt the K value as follows [13]:

$$K = c \cdot \frac{m_g(i, j)\sigma_g(i, j) - m_l(i, j)\sigma_l(i, j)}{\max [m_g(i, j)\sigma_g(i, j), m_l(i, j)\sigma_l(i, j)]}, \quad (6)$$

where $m_g(i, j)$ and $m_l(i, j)$ are the global and local mean values, respectively, and $\sigma_g(i, j)$ and $\sigma_l(i, j)$ are the global and local standard deviation values, respectively. c is a parameter to keep the value of K within the range of threshold values for each pixel in the image regarding the binarization method. For instance, in Bernsen's method a contrast image is computed by subtracting local maximum and minimum gray-levels. Then, if the contrast is below a certain global threshold, then a pixel is said to consist only one class, object or background. Generally, for Bernsen's method the global threshold is set at gray value 15 for 8-bit images. However, this threshold will affect all pixels in the contrast image evenly. Therefore, we proposed to adapt that threshold by using local information where the c parameter keeps the threshold around the gray value 15 and, consequently, the threshold will be adapted for each pixel in the contrast image. This strategy was applied to overall binarization methods, considering the typical global threshold values used for each one. The values for the c parameter were set as follows: Bernsen, 15; Niblack, 0.1; Sauvola, 0.01; Wellner, 0.1; and White, 0.1. Besides, empirical data obtained from experiments varying the window size, reveal that using a window of 5×5 pixels for all methods, produces better results in terms of recognition accuracy.

After performing the digit binarization some undesired regions remain around the binarized number. We made some assumptions about the image to design the strategy for eliminating the noisy regions. First, we assumed that the area of the digit object is greater than any other object in the image. And second, the digit object is centered on the image centroid. Thus, the strategy for eliminating undesired regions involves 3 steps:

1. Label the objects within the image with 4-connectivity.
2. Calculate the centroid of the image.
3. Measure both the area (A) of all labeled regions and the mean Euclidian distance (D) from all pixels within the region to the center of the image. Next, calculate the ratio $(D - A)/(D + A)$ for each region. We assumed that the object with the minimum ratio value correspond to the binarized digit and remains within the image, whereas the other regions were automatically eliminated.

Thereafter, we decided to compare the performance of applying additional post-processing operations to the binary images, just before we extract the features.

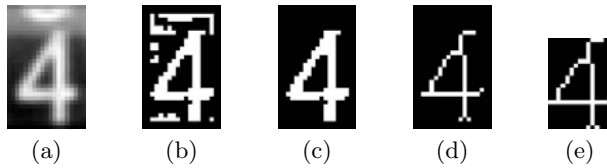


Fig. 2. Segmentation process. (a) Input image, (b) Binarized image by Niblack's method, (c) Binarized image after cleaning strategy, (d) Skeleton of the digit, and (e) Minimum bounding box containing the digit used for classification purposes.

Thus, it was computed the skeleton of the binarized digit and, at last, the image size was adjusted to the minimum bounding box containing the thinned digit. In Fig. 2 is illustrated the segmentation process for Niblack's method using skeletonized digits. Since skeletonization can be made by using mathematical morphology, in specific through successive erosions, the computational requirements are lower than those required by other algorithms.

2.4 Feature Extraction

The feature extraction process consists in computing attributes from the objects of interest to get some quantitative information for differentiating one class of objects from another. The attributes used in this work were the normalized central moments, denoted η_{pq} , which are defined as [14]:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}}, \quad (7)$$

where

$$\gamma = \frac{p+q}{2} + 1,$$

and μ_{pq} correspond to the central moments of order $(p + q)$, and for binary images they are expressed as:

$$\mu_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (x - \bar{x})^p (y - \bar{y})^q, \quad (8)$$

for $p = 0, 1, 2, \dots$ and $q = 0, 1, 2, \dots$, M and N are the width and height of the image, respectively, and (\bar{x}, \bar{y}) is the center of mass of the object. Herein, we used the first 3 orders, resulting in 7 normalized central moments: η_{11} , η_{20} , η_{02} , η_{30} , η_{03} , η_{21} , η_{12} .

2.5 Feature Selection

In many pattern recognition problems, a higher number of used features (or attributes) do not necessarily translate into higher classification accuracy. Therefore, feature selection is the process commonly used for removing irrelevant and

redundant features while maintaining acceptable classification accuracy. An irrelevant feature does not contribute to distinguish data of different classes and can be removed without affecting the classification performance. On the other hand, a redundant feature implies the co-presence of another feature, being both attributes relevant, but the removal of one of them will not affect learning performance [15].

In this article, a feature selection technique based on mutual information and intrinsic dimensionality was tested to reduce the space of attributes, which was developed by our group and whose technical details can be found in [16]. We employed a mutual information scheme based on minimal-redundancy-maximal-relevance (mRMR) criterion to rank the input data. Besides, the intrinsic dimensionality of the feature space was calculated by using principal component analysis (PCA). Thus, when using ranking features algorithms, the intrinsic dimensionality could estimate automatically the number of m features to be introduced into the classifier.

2.6 Digit Recognition

Although it is one of the earliest methods suggested, the minimum distance classifier is still an effective tool in solving the pattern recognition problem. With the minimum distance classifier data belonging to a class, are assumed to be represented by the mean value of this class. Suppose that we define the prototype of each pattern class to be the mean vector of the patterns of that class:

$$\mathbf{m}_j = \frac{1}{N_j} \sum_{\mathbf{x} \in \omega_j} \mathbf{x}_j, \quad j = 1, 2, \dots, W., \quad (9)$$

where W is the number of pattern classes, N_j is the number of pattern vectors from class ω_j and the summation is taken over these vectors. One way to determine the class membership of an unknown pattern vector \mathbf{x} is to assign it to the class of its closest prototype. Using the Euclidian distance to determine closeness reduce the problem to computing the distance measures:

$$D_j(\mathbf{x}) = \|\mathbf{x} - m_j\|, \quad j = 1, 2, \dots, W. \quad (10)$$

We then assign \mathbf{x} to class ω_j if $D_j(\mathbf{x})$ is the smallest distance [7].

3 Results

Each digit class (0 to 9) of the entire dataset was segmented by using the five binarization methods depicted in Section 2.1. Thereafter, normalized central moments were calculated from each single segmented image to create the feature space. Next, feature selection procedure described in Section 2.5 determines automatically that the first 3 ranked normalized central moments are enough for classifying. The results of this stage are presented in Table 1.

Table 1. Selected features based on mutual information and intrinsic dimensionality

Method Name	Bernsen	Niblack	Sauvola	Wellner	White
Selected features	$\eta_{12}, \eta_{02}, \eta_{11}$	$\eta_{02}, \eta_{12}, \eta_{11}$	$\eta_{02}, \eta_{12}, \eta_{30}$	$\eta_{12}, \eta_{02}, \eta_{30}$	$\eta_{02}, \eta_{21}, \eta_{30}$

For the classification stage, we compare the approach of using just the selected characteristics and all of them, separately. Thus, the technique of cross-validation divided randomly both, the selected and the whole dataset, in 70% for training and 30% for testing. This procedure was performed 500 times for each segmentation method. Table 2 and Table 3 show the percentage of digit classes correctly classified for each binarization method, when using feature selection and when employed the complete set of features.

Table 2. Percentage of recognition rates for each digit class for the five binarization methods, using the selected moments. The results are the mean values of 500 cross-validation runs.

Method	% Recognition rate in each class									Mean	σ	CV	
	0	1	2	3	4	5	6	7	8				9
Bernsen	85.4	91.6	83.6	94.3	91.8	85.2	78.1	79.8	81.1	92.1	86.3	5.8	0.07
Niblack	83.4	79.8	75.0	83.5	87.5	79.1	80.0	74.1	84.4	85.8	81.3	4.4	0.05
Sauvola	81.3	86.7	80.7	86.1	81.6	74.7	72.4	73.6	86.7	88.2	81.2	5.9	0.07
Wellner	86.2	72.8	68.8	87.6	76.7	69.0	79.7	72.3	85.0	88.8	78.7	7.8	0.10
White	77.7	97.0	82.4	89.1	85.3	75.2	70.3	88.1	76.7	79.2	82.1	7.9	0.10

These results pointed out that the Bernsen's method achieved the best recognition rates in both cases, by using feature selection and considering the whole feature space, with a total mean values of $86.3 \pm 5.8\%$ and $90.0 \pm 9.6\%$, respectively. This suggests that Bernsen's method is capable to preserve better the digits attributes than the other four methods, when the normalized central moments are employed. Besides, we used the coefficient of variation (CV) for measuring the dispersion of probability distribution for each binarization technique. One can note that the results obtained with feature selection presented lower variance for all segmentation methods than when using the entire feature space.

An additional computational issue considered to compare the binarization methods was the consuming time processing. Thus, we obtained the elapsed time during the binarization operations without considering the cleaning strategy. For these experiments we used a PC with dual core AMD CPU operated in 2.10 Ghz and Linux OS. The Table 4 shows the sum of running times for each binarization method considering all the images in our dataset. Now, through this comparison it is noticeable that Bernsen's method, that obtained the best

Table 3. Percentage of recognition rates for each digit class for the five binarization methods, using all the moments. The results are the mean values of 500 cross-validation runs.

Method	% Recognition rate in each class									Mean	σ	CV	
	0	1	2	3	4	5	6	7	8				9
Bernsen	70.1	96.2	92.7	99.0	98.4	84.0	83.1	82.5	96.9	96.9	90.0	9.6	0.11
Niblack	69.8	83.8	80.3	89.0	97.0	69.5	77.9	84.9	97.2	93.7	84.3	10.1	0.12
Sauvola	68.9	86.4	87.8	86.8	95.1	75.2	78.9	80.6	97.9	92.5	85.0	9.1	0.11
Wellner	68.9	77.9	77.0	87.4	95.9	68.8	71.4	82.3	97.1	91.9	81.9	10.8	0.13
White	76.0	99.8	81.9	93.6	96.4	85.0	84.2	90.8	95.1	96.4	89.9	7.7	0.09

accuracy, used the largest time to make the recognition. Since our future goal is to implement an OCR system in a mobile platform with limited hardware capabilities, White's method could be used instead of Bernsen's method as it has an adequate execution time and produces the second better accuracy results.

Table 4. Execution times of the binarization methods under study

Method Name	Bernsen	Niblack	Sauvola	Wellner	White
Execution time (seg)	16.46	7.04	6.55	6.40	6.04

4 Conclusion

In this article we evaluated five binarization methods in terms of recognition rates. These methods were studied because of their simplicity to be implemented in based-camera devices with limited hardware capabilities. Thus, the objective was to investigate which method is able to preserve important information about the digit nature. We used seven normalized central moments to differentiate quantitatively among digits. Also, there was employed a feature selection technique (for reducing the data dimensionality) as well as the entire feature set. Despite classification rates apparently are better when using all features, the feature selection reduces the problem dimensionality and kept a more stable recognition response with an acceptable misclassification rate by using only 3 moments from the original 7 moments. The representation of an object is not trivial, and for future work we are planning to use other kind of attributes for reaching better recognition rates (>90%) by using the same minimum distance classifier. The final objective of our investigation is to implement an OCR system in a camera-based cell phone, with limited hardware capabilities, for recognizing digits of water meters.

References

1. Anagnostopoulos, C.N.E., Anagnostopoulos, I.E., Loumos, V., Kayafas, E.: A License Plate-Recognition Algorithm for Intelligent Transportation System Applications. *IEEE Transactions on Intelligent Transportation Systems*, 377–392 (2006)
2. Ji-yin, Z., Rui-rui, Z., Min, L., Yin, L.: License Plate Recognition Based on Genetic Algorithm. In: *International Conference on Computer Science and Software Engineering*, pp. 965–968 (2008)
3. Coughlan, J.H.S.: Reading lcd led displays with a camera cell phone. In: *Conference on Computer Vision and Pattern Recognition Workshop (CVPRW 2006)*, pp. 119–119. *IEEE Computer Society, Washington, DC, USA* (2006)
4. Tian, L., Kamata, S.: An iterative image enhancement algorithm and a new evaluation framework. In: *IEEE International Symposium on Industrial Electronics (ISIE 2008)*, pp. 992–997 (2008)
5. Thulke, M., Margner, V., Dengel, A.: Quality evaluation of document segmentation results. In: *Proceedings of the Fifth International Conference on Document Analysis and Recognition, ICDAR 1999* (1999)
6. Trier, O.D., Jain, A.K.: Goal-directed evaluation of binarization methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1191–1201 (1995)
7. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*. Prentice-Hall, New Jersey (2002)
8. J. Bernsen.: Dynamic thresholding of gray-level images. In: *Proc. Eighth Int. Conf. Pattern Recognition*, pp. 1251–1255 (1986)
9. Niblack, W.: *An Introduction to Digital Image Processing*, pp. 115–116. Prentice Hall, Englewood Cliffs (1986)
10. Sauvola, J., Pietikinen, M.: Adaptive document image binarization. *Pattern Recognition* 33(2), 225–236 (2000)
11. Wellner, P. D.: *Adaptive Thresholding for the DigitalDesk*. Technical Report EPC-1993-110, Rank Xerox Ltd. (1993)
12. White, J.M., Rohrer, G.D.: Image thresholding for optical character recognition and other applications requiring character image extraction. *IBM J. Research and Development* 27(4), 400–411 (1983)
13. Rais, N.B., Hanif, M.S., Taj, I.A.: Adaptive thresholding technique for document image analysis. In: *Proc: 8th International Multitopic Conference*, pp. 61–66 (2004)
14. Huang, Z., Leng, J.: Analysis of hu’s moment invariants on image scaling and rotation. In: *2nd International Conference on Computer Engineering and Technology (ICCET 2010)*, vol. 7, pp. 476–480. *IEEE Computer Society, Los Alamitos* (2010)
15. Liu, H., Motoda, H.: *Computational Methods of Feature Selection*. Taylor & Francis, Boca Raton (2008)
16. Gómez, W., Leija, L., Díaz-Pérez, A.: Mutual Information and Intrinsic Dimensionality for Feature Selection. In: *7th International Conference on Electrical Engineering, Computing Sciences and Automatic Control (CCE 2010)*, Tuxtla Gutiérrez, Chiapas, September 8–10, pp. 339–344 (2010)