OTSU Guided Adaptive Binarization of CAPTCHA Image using Gamma Correction

Cunzhao Shi[†], Yanna Wang[†], Baihua Xiao[†] and Chunheng Wang[†]

[†]The State Key Laboratory of Management and Control for Complex Systems

Institute of Automation, Chinese Academy of Sciences (CASIA), Beijing 100190, China

Email: {cunzhao.shi, wangyanna2013, baihua.xiao, chunheng.wang}@ia.ac.cn

Abstract—Gamma correction, a nonlinear operation, has long been used to code and decode luminance or tristimulus values in video or still image systems [1]. In this paper, we make the following observations: for CAPTCHA images which could not be well binarized using the threshold of OTSU, there exists a gamma corrected image which could be well segmented by the OTSU threshold and the value of the best gamma could be revealed by observing the maximal inter-class variance (MICV) values of different images transformed by different values of gamma. Concretely, we convert the R, G, B channels of the original CAPTCHA image with different gamma values and transform the color images to gray-level images. Each gray-level image could be then segmented by the threshold acquired by OTSU. By linking each gamma value with the corresponding maximal inter-class variance value, we could draw a changing curve of variance values versus gamma. The best gamma could be acquired by finding the point whose related MICV starts to change slowly. Moreover, the polarity of the image could also be revealed by the changing trend of the curve. Experimental results on different categories of CAPTCHA images demonstrate the effectiveness of the observations for binarizing the CAPTCHA images and telling the polarity as well.

I. Introduction

Binarization, which segments the foreground characters from the cluttered background, is the premise for the following character localization and recognition under the conventional Optical Character Recognition (OCR) framework. Over the past decades, many image binarization algorithms [2], [3], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14] have been proposed. The threshold based binarization methods could be roughly classified to two categories: global threshold based methods and local threshold based ones. Among all the global threshold based binarization methods [20], [21], [22], [2], OTSU [2] is one of the most simple but effective methods [19] and has been widely used for a large variety of applications, such as the binarization of scanned books, historical documents, video text, scene text, etc. The core concept of OTSU is to find the threshold that would maximize the inter-class (foreground/background) variance value so as to segment the pixels whose intensity higher than the threshold to foreground and those lower to background. Global threshold based methods perform well on images with simple or uniform background and might fail if the background is cluttered or the foreground has non-uniform intensities. Whereas for local threshold based methods [23], [24], [25], [26], [27], the threshold is local or pixel based. The most representative one is the Niblack's
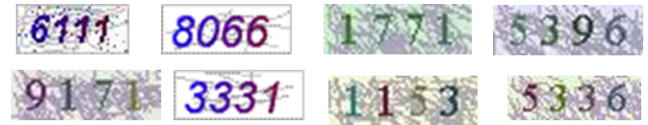

Fig. 1. Some examples of CAPTCHA images.

method [3], which finds the threshold for each pixel using the mean and standard deviation value of a certain window surrounding this pixel. The local threshold based methods perform well on clear document images. However, they are sensitive to the window size and might bring in much noise if the size of the window is not proper or the document is degraded.

Since the threshold based method might fail on images with complex background, rather than binarize the image on the fixed channels, some binarization methods try to compute the threshold on several channels, such as the {R,G,B} or {H,S,V} channels and finally select the best segmentation result from a series of candidate ones. However, when multiple channels are used, how to select the channel with the best segmentation result still remains an unsolved problem. On the other hand, gamma correction has long been used to code and decode luminance or tristimulus values in video or still image systems. Human vision, under common illumination conditions (not pitch black nor blindingly bright), follows an approximate gamma or power function, with greater sensitivity to relative differences between darker tones than between lighter ones. By taking advantage of the non-linear manner in which humans perceive light and color, gamma encoding of images is used to optimize the usage of bits when encoding an image, or bandwidth used to transport an image [1].

Although gamma correction has been widely used to compensate for the input-output characteristic of cathode ray tube (CRT) displays and to encode images and video images [15], [16], [17], [18], few people have evaluated its potential value for using it as a preprocessing method to binarize images. Is it possible that after certain gamma correction, the converted image could be better segmented than the original image using the same binarization method? If there exists the gamma value corrected by which the image could be better segmented, can we find a way to automatically select the value of this gamma?

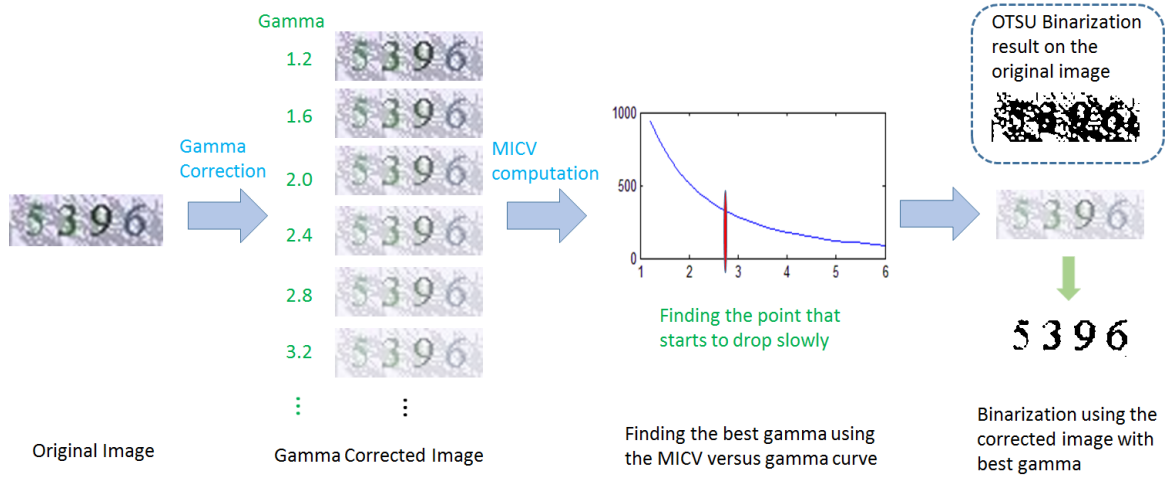In this paper, we explore the possibility of using gamma

Fig. 2. Flowchart of the proposed method.

correction to get better binarization result of CAPTCHA image. We choose CAPTCHA images for our experiment and the reasons are two-fold: first, as we can see from some examples of the CAPTCHA images in Fig. 1, these images have random noise and complex background and thus simple binarization method on the original image is not good enough for foreground/background segmentation; second, if we could get the satisfactory binarization results, these images could be easily recognized by the off-the-shell OCR engine. As for binarization method, we choose the classical and widely used OTSU binarization criterion. Luckily, we find the following observations: 1) there exists a gamma corrected image which could be best segmented by the OTSU threshold and the value of the best gamma could be revealed by observing the maximal inter-class variance (MICV) values of different images transformed by different values of gamma; and 2) the polarity of the original image could be acquired according to the changing trend of the curve of MICV-versus-gamma. We conduct a series of experiments on several different categories of CAPTCHA images with random noise and complex background from different websites. Experimental results demonstrate the effectiveness of our observations for binarizing the CAPTCHA images and telling the polarity as well.

The rest of the paper is organized as follows. Section II describes the proposed method. Experimental results and discussions are given in Section III and conclusions are drawn in Section IV.

II. THE PROPOSED METHOD

In this paper, we propose to make use of gamma correction along with the OTSU binarization criterion (maximize the inter-class variance) to get the best segmentation result for CAPTCHA images. The flowchart of the proposed method is shown in Fig. 2. Given an RGB color image that need to be binarized, we get a series of gamma corrected RGB images with different values of gamma. For each of these

transformed image, we convert the color image to gray-level image using the standard transformation coefficients and compute the maximal inter-class variance (MICV) value of the gray-level image. Then a changing curve of different gamma values versus the corresponding MICV values could be drawn and the polarity of the image could be acquired by getting the changing trend of the curve. We could get the best gamma for binarization using OTSU criterion by finding the point of the curve whose MICV starts to drop slowly. Once we get the best gamma, the final segmentation image could be acquired by binarizing the transformed image with the selected gamma using the OTSU criterion.

In the remaining of this section, first we will give some background knowledge of gamma correction and OTSU binarization criterion, then we will describe the proposed method and finally give the implementation details.

A. Related Background Knowledge

1) Gamma Correction: Gamma correction, gamma nonlinearity, gamma encoding, or often simply gamma, is the name of a nonlinear operation used to code and decode luminance or tristimulus values in video or still image systems [1]. The simplest form of Gamma correction could be defined by power-law expression:

$$V_{out} = AV_{in}^{\gamma} \tag{1}$$

where A is a constant and the input and output values are non-negative real values; in the common case of $A = 1$, inputs and outputs are typically in the range [0-1]. A gamma value $\gamma < 1$ is sometimes called an encoding gamma, and the process of encoding with this compressive power-law nonlinearity is called gamma compression; conversely a gamma value $\gamma > 1$ is called a decoding gamma and the application of the expansive power-law nonlinearity is called gamma expansion.

2) OTSU Binarization Criterion: OTSU aims to find the threshold that would maximize the inter-class (foreground/background) variance value. Let the pixels of a given

image be represented in L gray levels $[1, 2, ..., L]$. The number of pixels at level i is denoted by n_i and the total number of pixels by $N = n_1 + n_2 + ... + n_L$. The normalized gray-level histogram is regarded as a probability distribution:

$$p_i = n_i/N, \qquad p_i >= 0, \sum_{i=1}^{L} p_i = 1. \qquad (2)$$

Suppose we could dichotomize the pixels into two classes C_0 and C_1 (background and objects, or vice versa) by a threshold at level k; C_0 denotes pixels with levels $[1, ..., k]$, and C_1 denotes pixels with levels $[k+1, ..., L]$. Then the probabilities of class occurrence and the class mean levels, respectively, are given by:

$$w_0 = Pr(C_0) = \sum_{i=1}^{k} p_i \qquad (3)$$

$$w_1 = Pr(C_1) = \sum_{i=k+1}^{L} p_i \qquad (4)$$

and

$$\mu_0 = \sum_{i=1}^{k} iPr(i|C_0) = \sum_{i=1}^{k} ip_i/w_0 \qquad (5)$$

$$\mu_1 = \sum_{i=k+1}^{L} iPr(i|C_1) = \sum_{i=k+1}^{l} ip_i/w_1 \qquad (6)$$

The total mean level of the original image is

$$\mu_T = \sum_{i=1}^{L} ip_i \qquad (7)$$

The following relation could be verified for any choice of k:

$$w_0\mu_0 + w_1\mu_1 = \mu_T, \qquad w_0 + w_1 = 1. \qquad (8)$$

The best threshold could then be computed by maximizing the between-class variance:

$$\sigma_B^2 = w_0(\mu_0 - \mu_T)^2 + w_1(\mu_1 - \mu_T)^2 = w_0w_1(\mu_0 - \mu_1)^2 \qquad (9)$$

To get the best threshold, we compute the σ_B using all the gray-level intensities as the threshold and find the one which has the largest σ_B.

B. Image Correction with Different Gamma

The R,G,B channels of the original color image are transformed separately with gamma correction. As the input of the gamma correction should be in the range [0-1], we normalize all the intensities of pixels to [0-1] and after the correction, the intensities are transformed back to the range [0-255]. Different from Eq. 1, we use $V_{out} = AV_{in}^{1/\gamma}$ as our correction equation. Suppose the normalized R, G, B values for pixel i are R_i, G_i and B_i respectively, the intensity value $Gray_i$ of i of the transformed gray-level image after gamma correction could be computed as follows:

$$Gray_i = \boldsymbol{\alpha}(1)(R_i^{1/\gamma} * 255) + \boldsymbol{\alpha}(2)(G_i^{1/\gamma} * 255) \\ + \boldsymbol{\alpha}(3)(B_i^{1/\gamma} * 255) \qquad (10)$$
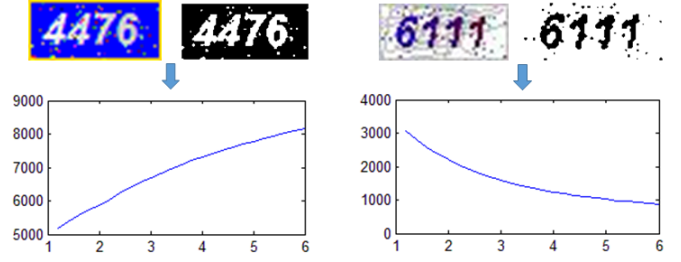

Fig. 3. The MICV-versus-gamma curves of images with different polarities.

where

$$\boldsymbol{\alpha} = [0.2989, 0.5870, 0.1140] \qquad (11)$$

$\boldsymbol{\alpha}$ is the standard coefficients to convert the RGB color image to gray-level one. We vary gamma from 1.2 to 6 with step of 0.2 and a series of corrected images could be acquired. After correction with different values, the contrast of the color image and the gray-level image changes, leading to different binarization result. Although the binarization result of the original image might not be satisfactory, there exist some images whose binarization results using the same OTSU criterion are better. Here we come to the question of how to automatically choose the satisfactory result from all the possible ones, which would be detailed in the following section.

C. Best Gamma Selection for Binarization

According to our observation, the polarity of the CAPTCHA image could be revealed by the trend of the curve of the MICV-versus-gamma. As shown in Fig. 3, we find that if the MICV value decreases with the increasing value of gamma, the polarity of the image is '0' (the foreground text is black "0" and background is white "255"), and if the MICV value also increases, the polarity is 1 (the foreground is white and background is black). To check the polarity, we first compute the MICV value $S_{1.2}$ of the transformed image with $\gamma = 1.2$ and then compute the MICV S_6 of the corrected image with $\gamma = 6$. The polarity is decided as follows:

$$polarity = \begin{cases} 0 & \text{if } S_{1.2} \geqq S_6 \\ 1 & \text{if } S_{1.2} < S_6 \end{cases} \qquad (12)$$

After getting the polarity, we invert the images if the polarity is '1'. To get the best gamma, we use a simple criterion to find the point of the curve that starts to change slowly. The following criterion is used: if the MICV difference between two successive corrected images exceeds the average gap, we regard it as normal, and if the MICV difference are lower than the average level, we regard the first point that is lower than the average as the point that starts to change slowly. We first compute the average changing score S_{avg} by dividing the differential of $S_{1.2}$ and S_6 with 24 equal partitions. S_{avg} is used as a threshold to decide whether the MICV value of gamma corrected images starts to change slowly. If S_{avg} is smaller than T_g (a predefined threshold, set to 15 in the experiment), the color image needs no gamma correction and
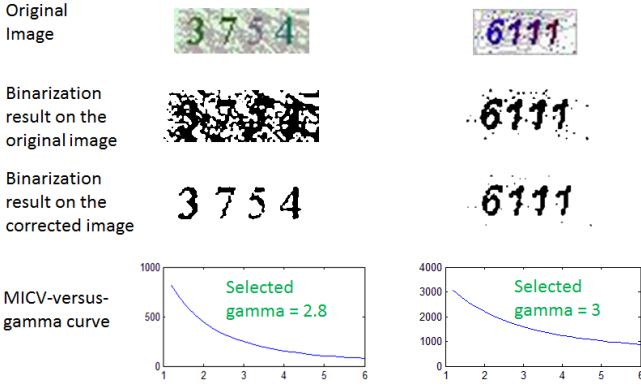
Fig. 4. Comparative binarization results with and without gamma correction.

Fig. 5. Some samples from each website of our dataset.

the OTSU binarization result on the original image is the final segmentation result. If S_{avg} is larger than T_g, we sequentially compute the MICV of the corrected image with incremental values of gamma and once the MICV difference between two consecutive gamma corrected images is smaller than S_{avg}, we regard this gamma as the proper one and stop the following computation. The matlab-like code of our algorithm is listed in Algorithm. 1.

Algorithm 1: Best gamma selection algorithm for binarization

1: $\gamma = [1.2 : 0.2 : 6]$
2: Compute $S_{avg} = (S_{1.2} - S_6 + 24)/24$
3: **If** $S_{avg} < T_g$
4: $gamma = 1$
5: **Else**
6:    **For** $i = 2 : k$
7:      **If** $S_{i-1} - S_i < S_{avg}$
8:        $gamma = \gamma(i-1)$
9:        break;
10:      **Else**
11:        continue;
12:      **end if**
13:    **end for**
14: **end if**

From the above algorithm we can see that only a subset of the potential gamma corrected images need to be computed. To get the polarity, we only need to compute $S_{1.2}$ and S_6. For best gamma selection, once we get the point whose MICV starts to change slowly, the following computation would be terminated. Fig. 4 gives two examples, whose best gamma values are 2.8 and 3.0 respectively. In the experiment, we find that the best gamma for most CAMPTHA images are smaller than 3.

D. Implementation Details

To get the corrected images with different values of gamma, we use the look-up table strategy to improve computation efficiency. We pre-compute the gamma corrected results of

gray-level intensities [0-255] and restore the corrected results in the look-up table. In the table, each intensity has 25 corresponding corrected values with gamma varying from 1.2 to 6 at a step of 0.2. Thus, the look-up table is a $256 * 25$ matrix, each row of which corresponds to the corrected values of a certain intensity. Suppose the image has m rows and n columns, the computation cost for correcting one image is only $m * n$ look-up table operations.

III. EXPERIMENTS

In this section, we will evaluate the feasibility of using our observations to binarize CAPTCHA images downloaded from various websites. Both the effect of the binarization and polarity will be evaluated and the pros and cons will be discussed.

A. Datasets

We downloaded CPATCHA images from 10 websites. The testset contains 1000 images from each website, leading to 10000 CAPTCHA images in total. Some of the images from the dataset are shown in Fig. 5. As we can see, these images have different polarities, random speckle noise, non-uniform foreground color and cluttered background. However, most of these images could be easily recognized by off-the-shelf OCR engine if satisfactory binarization results could be given.

B. Evaluation Protocols

Since we do not have the ground truth of the binarization image, we could not use the pixel-level evaluation protocol. We use the character recognition accuracy (CRR) to evaluate the binarization performance. We define CRR as:

$$CRR = N_c/N \qquad (13)$$

where N_c refers to the number of correctly recognized characters and N refers to the number of total characters in the dataset. Given the binarization image, we use the simple

(a) Images with hollow characters

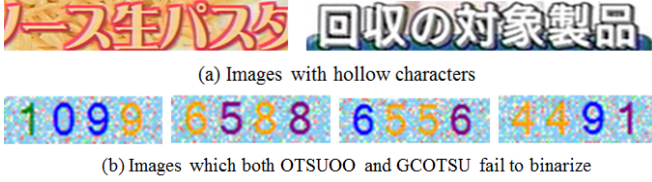(b) Images which both OTSUOO and GCOTSU fail to binarize

Fig. 6. Some failure examples of polarity estimation and binarization.

projection-based segmentation method to get each single characters and use our OCR engine which needs a binary input to recognize each character. As one false recognition results of the CAPTCHA image will lead to the failure of verification, apart from the evaluation of the single character recognition performance CRR, we also evaluate the recognition accuracy of the whole CAPTCHA image–text recognition accuracy (TRR):

$$TRR = N_{tc}/N_t \qquad (14)$$

where N_{tc} and N_t refer to the number of correctly recognized CAPTCHA images and that of total CAPTCHA images respectively.

We evaluate the performance of the polarity estimation method by polarity classification accuracy (PCR), which is defined as:

$$PCR = N_{tpc}/N_t \qquad (15)$$

where N_{tpc} represents the number of CAPTCHA images whose polarities are correctly estimated.

C. Evaluation of Polarity Estimation

We use the trend of MICV-versus-gamma curve to estimate the polarity. We estimate the polarity by comparing the value of $S_{1.2}$ and S_6 of the CAPTCHA images and the PCR on the testset is as high as 99.8%, demonstrating the effectiveness of the proposed polarity estimation method. However, our observation could only deal with images with uniform polarity, which means if the characters in the image have multiple polarities, our method will fail. Moreover, for images with hollow characters, our polarity estimation method might also fail. Fig. 6(a) shows some images with hollow characters whose polarities the proposed method fails to estimate.

D. Evaluation of Binarization via Gamma Correction and OTSU

As we use the OTSU binarization criterion for selecting the best gamma, we compare the proposed method (GCOTSU) with OTSU on the original image (OTSUOO). For OTSUOO, the color images are converted to a single grayscale image using the standard coefficients. As OTSUOO could not decide the polarity of the image, we manually reverse the image if the polarity is '1' so that the binarized images could be recognized by the OCR engine. For GCOTSU, since it could tell the polarity of the image, we directly use the binarization results for recognition. The CRR and TRR of the proposed methods as well as those of OTSU on the images from different websites are shown in Table I and Table II respectively.

TABLE I
CRR OF GCOTSU AND OTSUOO ON DIFFERENT WEBSITES (%).

Website	OTSUOO	GCOTSU
1	100	100
2	98.2	99.3
3	97.3	99.95
4	97.85	98.9
5	90.53	93.7
6	99.95	100
7	93.4	94.5
8	93.6	94.1
9	78.5	81.3
10	95.2	99.8
Average	94.45	96.16

TABLE II
TRR OF GCOTSU AND OTSUOO ON DIFFERENT WEBSITES (%).

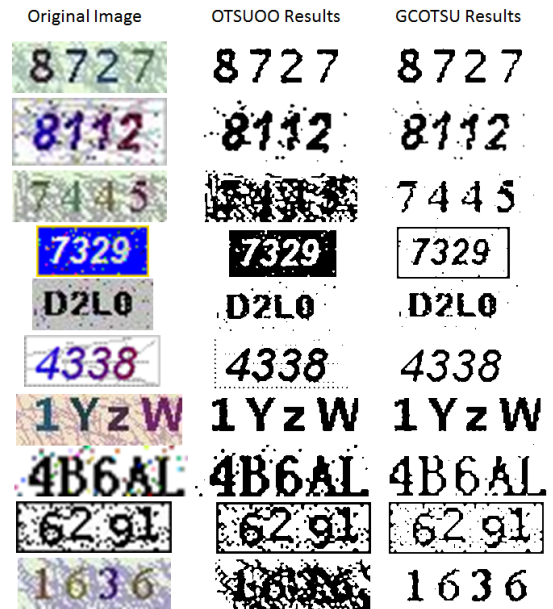| Website	OTSUOO	GCOTSU
1	100	100
2	94.4	96.8
3	96.8	99.8
4	92.2	95.6
5	84.6	90.4
6	99.8	100
7	90.2	91.3
8	88.3	92.5
9	73.1	79.2
10	94.8	98.6
Average	91.4	94.4

Fig. 7. Comparison binarization results of OTSUOO and the proposed GCOTSU.

The results demonstrate that the proposed GCOTSU out-performs OTSUOO on all the datasets both for CRR and TRR. As OTSUOO only binarize the original images, which might have various noise and complex background, the binarization results might be disappointing, making the following projection based character segmentation and recognition very difficult. Whereas for the proposed GCOTSU, although we also choose OTSU criterion to calculate the threshold, we transform the original images with gamma correction and the certain corrected images might have less noise and cleaner background. By selecting the proper gamma to correct the original image, the binarization result is satisfactory enough for the following segmentation and recognition. Fig. 7 shows some comparison binarization results of GCOTSU and OTSUOO from the datasets. As we can see, for images that could be well binarized by the OTSUOO, the GCOTSU could also give quite satisfactory results, whereas for images that OTSUOO fails to binarize, the GCOTSU could still remove the noise and give binarization results that are more suitable for the following character recognition.

There are some CAPTCHA images which OTSUOO and GCOTSU both fail to binarize. Fig. 6(b) shows some failure examples. As we can see, the colors of the characters vary a lot and thus one global threshold might fail to segment characters with different colors. Moreover, the performance of both the polarity estimation and the binarization methods is unstable on scene text images. The reason might lie in the fact that scene text images have various lighting condition as well as complex background, thus making the global threshold not enough to give satisfactory binarization result.

IV. Conclusion

In this paper, we make two observations: 1) gamma correction along with the OTSU binarization criterion could be used to get better segmentation results for CAPTCHA images and the best gamma could be acquired by finding the point of the MICV-versus-gamma curve that starts to change slowly; and 2) the polarity of the CAPTCHA images could be revealed by the changing trend of the curve. The experimental results demonstrate the effectiveness of the observations for binarizing the CAPTCHA images and telling the polarity as well. The proposed observations only use the simple OTSU criterion to help select the best gamma, which might fail on some cameral-captured scene text images. In the future, we would try to explore other criterions together with gamma correction which could not only binarize CAPTCHA images, but also cope with scene text images.

References

[1] Digital video and HDTV: Algorithms and Interfaces, Charles A. Poynton, Morgan Kaufmann. pp. 260, 630.

[2] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, pp. 285–296, 1975.

[3] W. Niblack, *An introduction to digital image processing.* Strandberg Publishing Company, 1985.

[4] Frequently Questioned Answers about Gamma, Charles Poynton, 2010.

[5] J Sauvola, T Seppanen, S Haapakoski, M Pietikainen, Adaptive document binarization, Proceedings of the Fourth International Conference on Document Analysis and Recognition, pp. 147-152, 1997,IEEE.

[6] T. Sato, T. Kanade, E. K. Hughes, and M. A. Smith, Video OCR for digital news archives, Proc. IEEE Int. Workshop on Content-Based Access of Image and Video Database (CAVID'98), pp. 52 - 60 , 1998.

[7] X. Chen and A. Yuille, Detecting and Reading Text in Natural Scenes, Proc. Int'l Conf. Computer Vision and Pattern Recognition, pp. II:366-373, 2004.

[8] M. R. Lyu , J. Song and M. Cai, A comprehensive method for multilingual video text detection, localization, and extraction, IEEE Trans. Circuit and Systems for Video Technology, vol. 15, num. 2, pp. 243-255, 2005.

[9] ØD Trier, T Taxt,Evaluation of binarization methods for document images, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 17, num. 3, pp. 312-315, 1995.

[10] B Gatos, I Pratikakis, SJ Perantonis, Adaptive degraded document image binarization, Pattern recognition, vol. 39, num. 3, pp. 317-327, 2006.

[11] Y Liu, SN Srihari, Document image binarization based on texture features, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, num. 5, pp. 540-544, 1997.

[12] B Gatos, K Ntirogiannis, I Pratikakis, ICDAR 2009 Document Image Binarization Contest (DIBCO 2009), International Conference on Document Analysis and Recognition (ICDAR), vol. 9, pp. 1375-1382, 2009.

[13] Y Yang, H Yan, An adaptive logical method for binarization of degraded document images, Pattern Recognition, vol. 33, num. 5, pp. 787-807, 2000.

[14] NR Howe, A laplacian energy for document binarization, International Conference on Document Analysis and Recognition (ICDAR), pp. 6-10, 2011.

[15] H Farid, Blind inverse gamma correction, IEEE Transactions on Image Processing, vol. 10, num. 10, pp. 1428-1433, 2001.

[16] MJ Liaw, HH Yang, YR Shen, Automatic gamma correction system for displays, US Patent 6,593,934, 2003.

[17] PM Lee, HY Chen, Adjustable gamma correction circuit for TFT LCD, IEEE International Symposium on Circuits and Systems (ISCAS), pp. 780-783, 2005.

[18] J Kim, Color correction device for correcting color distortion and gamma characteristic, US Patent 5,949,496, 1999.

[19] OD Trier, AK Jain, Goal-directed evaluation of binarization methods, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 17, num. 12, pp. 1191-1201, 1995.

[20] AS Abutaleb, Automatic thresholding of gray-level pictures using two-dimensional entropy, Computer vision, graphics, and image processing, vol. 47, num. 1, pp. 22-32, 1989.

[21] J.N. Kapur, P.K. Sahoo, and A.K.C. Wong, A new method for gray-level picture thresholding using the entropy of the histogram, Computer Vision, Graphics and Image Processing, vol. 29, pp. 273-285, 1985.

[22] J. Kittler and J. Illingworth., Minimum error thresholding, Pattern Recognition, vol. 19, no. 1, pip. 41-47, 1986.

[23] J. Bernsen, Dynamic thresholding of grey-level images, Proc. Eighth Intl Conj Pattern Recognition, pp. 1,251-1,255, Paris, 1986.

[24] C.K. Chow and T. Kaneko, Automatic detection of the left ventricle from cineangiograms, Computers and Biomedical Research, vol. 5, pp. 388410, 1972.

[25] Y. Nakagawa and A. Rosenfeld, Some experiments on variable thresholding, Pattern Recognition, vol. 11, no. 3, pp. 191-204, 1979.

[26] L. Eikvil, T. Taxt, and K. Moen, A fast adaptive method for binarization of document images, Proc. First Int'l Con Document Analysis and Recognition, pp. 435-443, Saint-Malo, France, 1991.

[27] K.V. Mardia and T.J. Hainsworth, A spatial thresholding method for image segmentation, IEEE Trans. Partern Analysis and Machine Intelligence, vol. IO, no. 6, pp. 919-927, 1988.