True Color Distributions of Scene Text and Background

Renwu Gao, Shoma Eguchi, Seiichi Uchida
Kyushu University
Fukuoka, Japan
Email: {kou, eguchi}@human.ait.kyushu-u.ac.jp, uchida@ait.kyushu-u.ac.jp

Abstract—Color feature, as one of the low level features, plays important role in image processing, object recognition and other fields. For example, in the task of scene text detection and recognition, lots of methodologies employ features that utilize color contrast of text and the corresponding background for connected component extraction. However, the true distributions of text and its background, in terms of color, is still not examined because it requires an enough number of scene text database with pixel-level labelled text/non-text ground truth. To clarify the relationship between text and its background, in this paper, we aim at investigating the color non-parametric distribution of text and its background using a large database that contains 3018 scene images and 98,600 characters. The results of our experiments show that text and its background can be discriminated by means of color, therefore color feature can be used for scene text detection.

I. INTRODUCTION

Scene text detection is still a difficult task in the field of scene text recognition. Comparing with traditional Optical Character Recognition (OCR), which has been well researched and has already achieved a great progress in the literature of pattern recognition, scene text can be in any kind of scenarios that have any kind of complex and unpredictable context. The illimitation of natural scene makes it very difficult to extract text from its background.

Color feature, as one of the low level features, has been widely used in the task of scene text detection. Since texts in natural scenes are supposed to convey important information to pedestrians (for example, texts in a billboard, or traffic information in the signal board), they are generally designed to be different from their backgrounds to make themselves easier to be noticed, as shown in Fig. 1. Using this characteristics of text, Gao *et al.* [1] introduce a bottom-up visual saliency model utilizing color feature for scene text detection. Yi *et al.* [2] use color feature and clustering method for the extraction of connected components. Different from Yi *et al.*, Ezaki *et al.* [3] use color feature and Fisher Discriminant Ratio (FDR) for connected component extraction. Khan *et al.* [4] propose a novel way of adding color information to shape bag-of-words to guide attention by means of top-down attention map. Weijer *et al.* [5] introduce a color saliency boosting algorithm to exploit the saliency of color edges based on information theory. Shivakumara *et al.* [6] propose a method that is comprised of wavelet decomposition and color features, namely RGB, for text detection in video. Jung *et al.* [7] utilize neural networks to extract texture information on several color bands. Rusinol *et al.* [8] add local color information to the shape context description for perceptual image retrieval. Karatzas *et al.* [9]

Fig. 1. Examples of high color contrast of text and its background.

exploit characteristics of human perception of color differences and then utilize the color differences to segment/detect text in web images. Rigaud *et al.* present a color based approach for comics character retrieval using content-based drawing retrieval and color palette in [10].

Though text features that utilize color have been widely used in the past scene text detection researches, some properties of text are assumed according to some intuitions, or just a small number of observations. The true relationship between color and its background still needs investigation. To clarify the relationship between text and its background, we aim at, in this paper, analyzing the color contrast (the difference of text and its background by means of color) by evaluating color non-parametric distribution using a large database that contains 3018 scene images and 98,600 characters. As far as we know, this is the first time of revealing the true color distribution of scene text and its background with a large enough observations. We use the HSV color space instead of RGB color space in our work. For each channel in HSV color space, we build a histogram to represent the relationship between text and its corresponding background (i.e., three histograms in total). We give a trial of using the color feature for scene text detection to investigate how it works.

II. COLOR DISTRIBUTION EVALUATION

In this section, we will evaluate the non-parametric distributions for each channel in HSV color space, and analyze the relationship of text and its background. However, before the evaluation of the distribution, we have to discuss the definition of the background given the text, because different background definitions may result into different distribution.

A. Definition of Text and Background Color

Text color can be easily defined: we can simply use the average color value of all pixels that belong to the text;
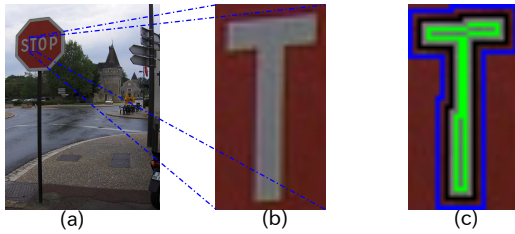
Fig. 2. Illustration of text and background color extraction. (a) Original image. (b) Enlarged part of the blue rectangle of (a). (c) Colors of text and background definition. Blue contour denotes the background pixels, black contour represents edge of text and green contour indicates pixels that used to extraction text color. Note that there is a one-pixel wide gap between each contour.

Fig. 3. Examples of scene images of our database. (a) The original images. (b) The corresponding ground truth images of (a).

however, for the definition of background color, it deeply depends on the definition of the background. Since the definition of background can be arbitrary, background color definition becomes ambiguous. Generally, according to the intuition, for a given text, pixels that have the same color and wrap this text can be considered as the background. For example, in Fig. 1(d), the signboard saying "STOP" should be considered as the background of text "STOP". However this kind of definition requires detecting "carriers" (where the texts are written) first. Unfortunately, "carrier" detection could be another difficult task to be researched. In this paper, we give a simpler definition of the background: pixels that are only one pixel away from the text edge. Correspondingly, text color should be re-defined as the average value of pixels that belong to text and one pixel away from the text edge, as shown by Fig. 2. Note that for a given text color, there always is the corresponding background color.

B. Non-parametric Distribution

Since we aim at clarifying the color contrast between text and its background, we should not first assume the distribution (e.g., Gaussian distribution) and then optimize the parameters. This is because text and its background color distribution could be any kind of shape. In other words, prior knowledge should not be assumed. Instead of parametric distribution, we evaluate the distribution based on non-parametric statistics of text and its background color. Once the text and the corresponding background color are defined, we can then evaluate the non-parametric color distribution. For each channel in HSV space, we plot the text color (as the x axis) against the background color (as the y axis). The color of the plot indicates the density of points being in a coordinate. The darker color represents the existence of more points.

III. Observation and Analysis of Actual Color Contrast

In order to investigate the *actual* relationship between text and its corresponding background, we use a big enough database for our experiments. The database was prepared by our laboratory and contains 3018 scene images with 96,800 characters totally (we removed small characters, and 8601 characters are used in our experiment). For each scene image, the corresponding ground truth image was manually labelled in pixel level. Fig. 3 shows some examples of scene images and the corresponding ground truth.

Fig. 4 shows the results of the non-parametric distribution evaluation. Roughly speaking, text color has different distribution against to the corresponding background color in value and saturation channels of HSV color space; in hue channel, text and background have similar distribution. This indicates that color feature can be employed as a cue for the task of scene text detection.

A. Value

Fig. 4(a) gives the histogram of the text value channel against to the background value channel. From this figure we can know that there are few texts having the same value as the corresponding background (as shown by the dark blue part of the diagonal): in natural scenes, text is either bright with dark background (e.g., Fig. 1(a)), or dark with bright background (e.g., Fig. 1(b)). Both combinations make the text easier to be seen (i.e., salient). Comparing the top-left part of Fig. 4(a) with the bottom-right part of it, we can conclude that, statistically, there are more dark texts with bright background than bright texts with dark background in real world.

B. Saturation

Fig. 4(b) shows the saturation histogram. Its left and bottom parts show that saturate texts seems always to be wrapped by the unsaturated background (the bottom part); and the unsaturated texts followed by saturate background (the left part). These kind of combinations can also make text more attractive. In HSV color space, saturation becomes unstable when the value channel is low (at the small side of the cone in HSV color space, saturation length, in terms of radius, becomes shorter and shorter). This also explains why saturation mainly focus on the bottom-left part of Fig. 4(b), providing that most scene texts are dark with bright background according to the left part of Fig. 4(a). This becomes a problem when evaluating the distribution of saturation channel, therefore we exclude texts that the color of either text itself or its background is lower than 0.5 in value channel. Fig. 5(a) shows the result of saturation channel distribution after removing low value texts. The left part of Fig. 5(a) indicates that, for unsaturated bright texts, background can be any degree of saturation. In other words, if scene text is white, background can be any kind of color. The bottom of Fig. 5(a) means that, for saturate bright text, background can only be some kind of color.
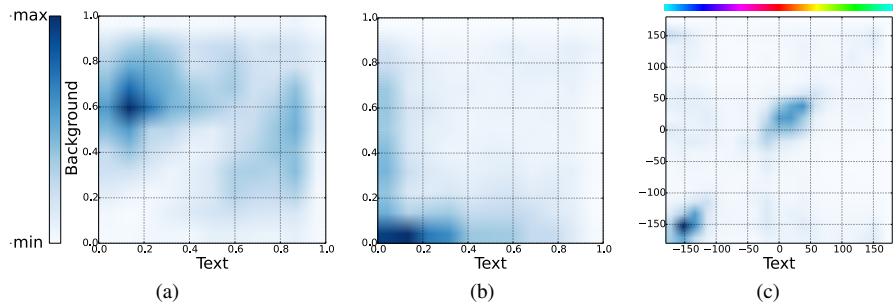
Fig. 4. Non-parametric distribution of each channel in HSV color space. (a) Distribution in value channel. (b) Distribution in saturation channel. (c) Distribution in hue channel.
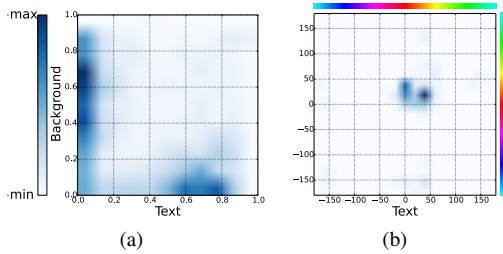

Fig. 5. Color distribution with condition. (a) Distribution of saturation channel under the condition of value channel greater than 0.5. (b) Distribution of hue channel under the condition of both value and saturation channels greater than 0.5.

C. Hue

Before looking into the distribution of hue, let us discuss the transformation from RGB to hue, which is significant for the later explanation of the distribution. As we all know, hue is defined to describe how much a given color is different from the primary color (i.e., red, green, blue and yellow). This means that for a hue value, there are a set of colors that correspond to it. For example, hue equaling to 0 corresponds to colors satisfying the condition that red channel is greater than green channel, which equals to blue channel. Precisely, white color (255, 255, 255) and red color (255, 0, 0) share the same hue value (hue=0), because according to the formula, $hue = 60 \times \frac{green-blue}{red-\min\{green,blue\}}$, when green channel equals to blue channel, as long as the red channel is higher, hue equals to zero.

With that in mind, let us look into Fig. 4(c), which illustrates the hue histogram. It tells us that text and its background have similar degree of difference to primary colors, shown in terms of the diagnose of the distribution. This sounds impossible, because, for example, red text in red background makes text even invisible. However, as we have discussed above, white color (and other colors the satisfying a certain condition) can be transformed into the same hue as red color. This means the red part of the color bar in Fig. 4(c) represents not only the red color in RGB, but also also other colors (e.g., white color). Providing that most scene texts are bright in dark background (Fig. 4(a)), most "red" color texts in hue distribution are black in red color background, or red color text in white color background, as shown by Fig. 6(a). Correspondingly, most "blue" color texts are black in blue color background, or blue color text in white color background
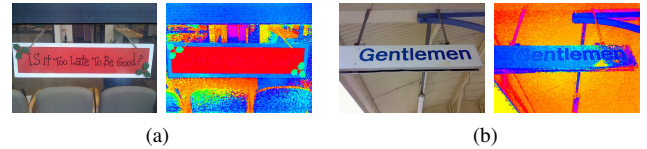

Fig. 6. Examples of different color having the similar hue value. (a) "red" text in red background. (b) blue text in "blue" background. The left column are the original images, the right ones are the RGB color images converted from hue channel by setting value and saturation to 1.0.

(see Fig. 6(b)). Though, from the symmetry, it seems there are few differences (i.e., low contrast) between text and its background, hue can still be considered as a useful cue for text detection: if candidate and its background has high contrast in hue channel, it is unlikely to be text.

For the same reason as the case of saturation, when value and saturation are low, hue becomes unstable. We removed texts that either value channel or saturation channel is lower than 0.5 and obtained Fig. 5(b) which indicates almost all the bright saturate texts in natural scene are close to red primary color, and the corresponding backgrounds can have primary color of red and yellow.

IV. ACTUAL COLOR CONTRAST CHARACTERISTICS REALLY USEFUL FOR SCENE TEXT DETECTION ?

Three experiments were done in order to investigate how color feature works:

1) The same 12 shape features as described in [12] are utilized.
2) Hue, saturation and value channels of text and its background are attached to shape features, by which the color of text and its background is compared when detecting text.
3) Tendencies (in terms of density of the non-parametric distribution) of the text and its background are obtained by looking up Fig. 4 after getting HSV values, and then attached to the shape features. Though the use of hue tendency (Fig. 4(c)) may be unstable, since we aim at revealing the effect of color feature, including positive and negative ones, hue tendency was used.

Note that all experiments employed Niblack [11] for image segmentation and random forest algorithm for text detection.
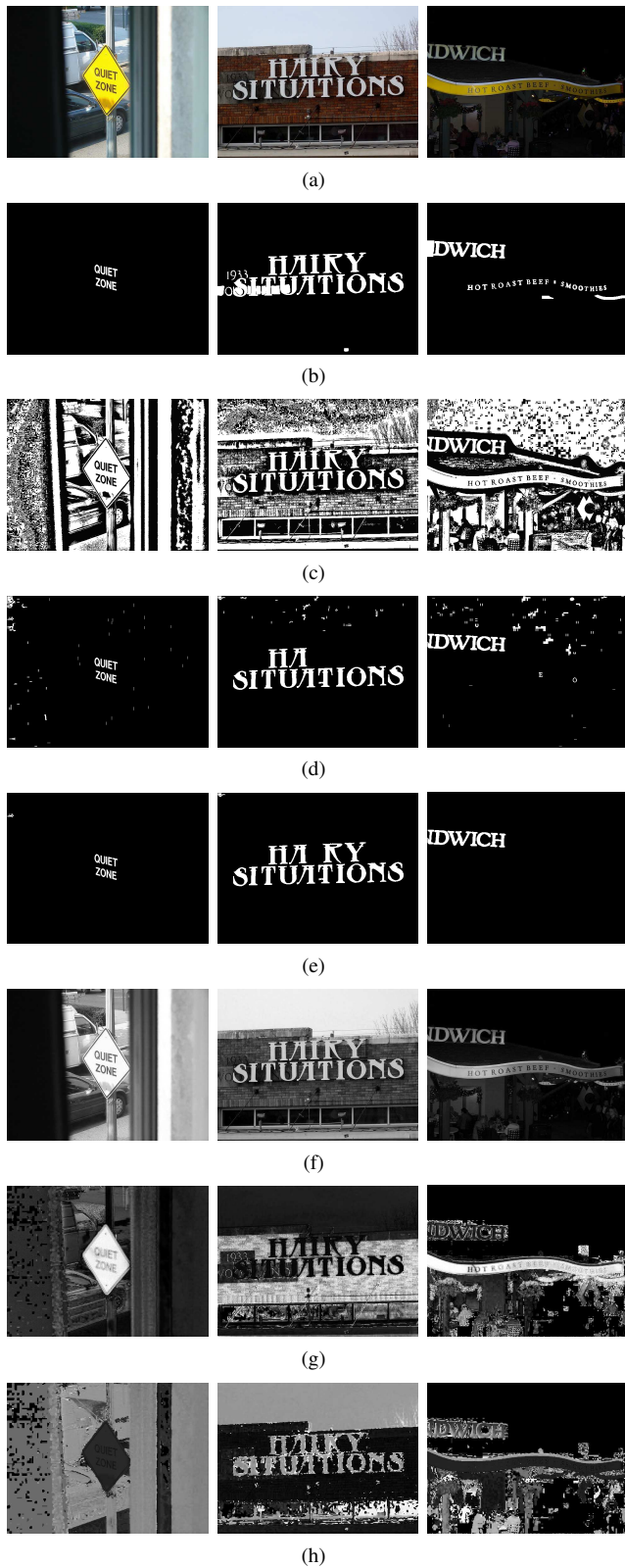
Fig. 7. Positive effect examples of using color feature in terms of density distribution. (a) Original images. (b) Ground truth images of (a). (c) Niblack segmentation results. (d) Text detection results using only shape feature. (e) Text detection results using shape and color features. (f) Value channel images of (a). (g) Saturation channel images. (h) Hue channel images. Note images in (f), (g) and (h) are normalized to [0, 255].
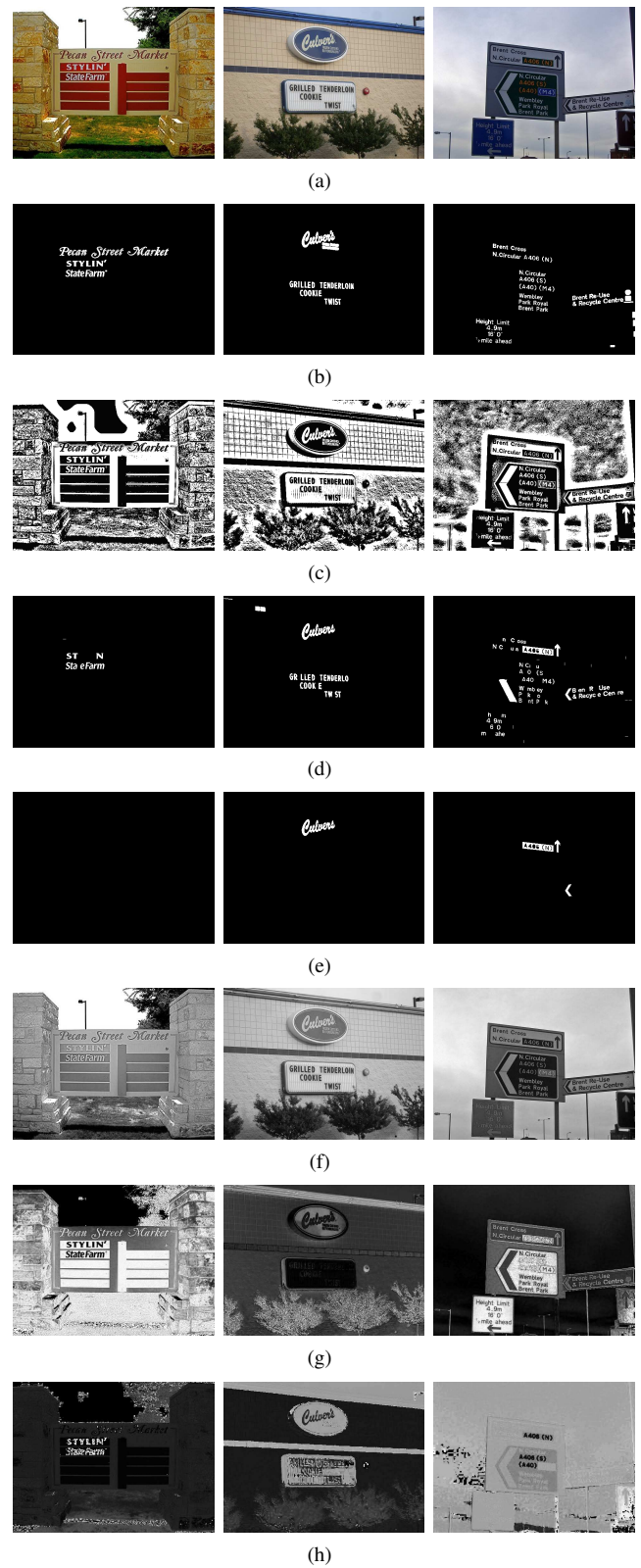

Fig. 8. Negative effect examples of using color feature in terms of density distribution. (a) Original images. (b) Ground truth images of (a). (c) Niblack segmentation results. (d) Text detection results using only shape feature. (e) Text detection results using shape and color features. (f) Value channel images of (a). (g) Saturation channel images. (h) Hue channel images. Note images in (f), (g) and (h) are normalized to [0, 255].
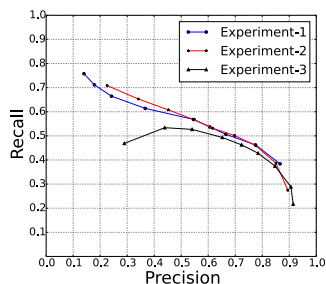
Fig. 9. Performance evaluation of the three experiments. *Experiment-1* used only shape features, *Experiment-2* used shape feature and HSV values of text and its background, and *Experiment-3* used shape feature and the tendencies of text and its background.

Fig. 7 shows some positive effect examples of using color feature in terms of density distribution. False alarmed "texts" using the same features as in [12] (Fig. 7(d)) have been removed successfully comparing to results with color feature added (Fig. 7(e)). If we look into the value channel of feature image (Fig. 7(f)) we can known that, all of these removed text candidates are low in value channel with low value in background; however, according to Fig 4(a), there are few dark texts with dark background in natural scenes. This means that those text candidates have low probability of being text. If we focus on the second column, we can know that, comparing the forth row with the fifth row, more texts are retrieved by adding color feature. Dark text candidate with bright background is more likely to be text in accordance with the conclusion that most scene texts are dark with bright background.

Fig. 8, in contrast, shows some negative effect examples. In Fig. 8, several texts are missed because they have high contrast in hue channel, which, according to Fig. 4(c), has low probability of being texts. For the rest texts that are excluded in the third column, they are missed because of either bright text with dark background in the value channel or dark saturate text in unsaturated background.

Fig. 9 shows the evaluation result of three experiments, from which we know the second experiment gave the best performance, while the third one performed worst. The reason for the poor performance in the third experiment may be the instability of hue tendency. Comparing the three experiments we can conclude that, by adding color feature, components are more correctly removed than retrieved. Note that we only evaluated true text and its background distribution, leaving non-text and its background distribution unknown. While detecting scene text, candidates having the same tendency of text and its background still can be rejected because they may be more similar to the tendency of non-text and its background, resulting in low recall of text; candidates having very different tendency from text, regardless of the tendency of non-text, we can conclude that they are more likely to be non-text, resulting in high detection precision. In other words, having the similar tendency to text is a prerequisite of component being detected as text.

V. Conclusion

In this paper, the first time of exposing the *actual* color distribution of scene text and its background based on a

large enough observations is given. We evaluate three non-parametric distributions of color in HSV, according to which we can know that text and its background have high contrast in saturation and value channels, but very low contrast in hue channel. The value channel distribution indicates that text and its background has high contrast in terms of brightness. More ever, there are more dark scene texts in bright background than bright texts in dark background in real world. Saturation channel distribution shows that saturate text is generally wrapped by unsaturated background, and unsaturated text is followed by saturate background. Hue channel distribution indicates that the primary colors of text and its background in natural scenes are closer to red and blue colors, and are barely close to purple and green. The trial result of scene text detection tells us that color are more useful for noise removal than text retrieval.

References

[1] R. Gao, S. Uchida, A. Shahab, F. Shafait and V. Frinken, Visual Saliency Models for Text Detection in Real World, PLoS ONE, vol.9, no.12, pp.e114539, 2014.

[2] C. Yi and Y. Tian, Text String Detection from Natural Scenes by Structured-Based Partition and Grouping, IEEE Transactions on Image Processing, vol.20 no.9, pp.2594-2605, 2011.

[3] N. Ezaki, K. Kiyota, M. Bulacu and I. Schomaker, Improved Text Detection Methods for Camera-Based Text Reading System for Blind Persons, International Conference on Document Analysis and Recognition (ICDAR2005), Seoul, pp.257-251, 2005.

[4] F.S. Khan, J. Weijer and M. Vanrell, Top-down Color Attention for Object Recognition, IEEE 12th International Conference on Computer Vision, p.979-986, Sept. 2009.

[5] J. Weijer, T.Gevers and A.Bagdanvo, Boosting Color Saliency in Image Feature Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.28, no.7, pp.150-156, 2006.

[6] P. Shivakumara, T.Q. Phan and C.L. Tan, New Wavelet and Color Features for Text Detection in Video, 20th International Conference on Pattern Recognition (ICPR), pp.3996-3999.

[7] K. Jung, Neural Network-based text location in color images, Pattern Recognition Letters, vol.22, no.14, 2001.

[8] M. Rusinol, F. Noorbakhsh, D. Karatzas, E. Valveny and J. Lldos, Perceptual Image Retrieval by Adding Color Information to the Shape Context Descriptor, 20th International Conference on Pattern Recognition (ICPR), pp.1594-1597, 2010.

[9] D. karatzas, A. Antonacopoulos, Color Text Segmentation in Web Image Based On Human Perception, Image and Vision Computing, vol.25, pp.564-577, 2007.

[10] C. Rigaud, J.C. Burie, J.M. Ogier and D. Karatzas, Color Descriptor for Content-Based Drawing Retrieval, 2014 11th IAPR International Workshop on Document Analysis System (DAS), pp.267-271, 2014.

[11] W. Niblack, An Introduction to Image Processing, Prentice-Hall, Englewood Cliffs, NJ, p.115-116, 1986.

[12] F. Qi, K. Zhu, M. Kimachi, Y. Wu and T. Aziwa, Using AdaBoost to Detect and Segment Characters from Natural Scene, Camera-Based Document Analysis and Recognition, ICDAR Workshop, 2005.