Color Structure Recovering in Strong Specular Text Regions

Tam Nguyen  
Saigon Technology University  
Ho Chi Minh city, Vietnam  
tam.nguyenth@stu.edu.vn

GueeSang Lee  
Chonnam National University  
Gwangju, Korea, 500-757  
gslee@jnu.ac.kr

Huan Nguyen Nhu  
HCMC University of Science  
Ho Chi Minh city, Vietnam  
1012158@student.hcmus.edu.vn

Thuc Nguyen Dinh  
HCMC University of Science  
Ho Chi Minh city, Vietnam  
ndthuc@fit.hcmus.edu.vn

Abstract—Highlight component is unavoidable in natural text images, which is caused by the effects of degradations, environment conditions, illumination and material of objects during the capture process. There are some techniques which could reduce or remove the specular regions but almost of them are effective only in simple cases which contain small and slight highlight areas. Strong specular zones, especially in the cases of the destroyed color structure are still challenges for the current approaches. In this paper, we introduce an effective method to remove strong highlight components which ruin the color structure in the natural text images. By analyzing tensor saliency map based on saliency magnitude and saturation information, specular and diffuse candidates are determined. Then, diffuse nominees are used to remove highlight components from specular candidates based on the painting technique. Our proposed method is evaluated using the standard dataset of the ICDAR 2005. The promising results have confirmed the effectiveness of the proposed method.

Keywords—Strong highlight; tensor voting; painting; saliency map; color structure

I. INTRODUCTION

Highlight detection and removal is an issue which has received much attention from researchers because of its usefulness. First, the information achieve from highlight detection could be used to determine the light source and solve problems in computer graphics such as 3D object modeling, rendering and geometry. Furthermore, the removal of specular components in images has significant role in increasing the accuracy and performance of segmentation, pattern recognition systems. A typical application is that specular removal in text recognition systems which are applied to natural text images. It is easy to release that text recognition from printed or scanned documents are being scarce and out of date. Instead, text pictures captured from camera, which are affected by a lot of factors such as illumination, environment conditions, and material of objects, become a new trend and attract many science to focus on them. This tendency confirms the gravity of the highlight issues in text natural images. But almost of them still give approaches to remove the effect of highlight regions in text incomprehensiblyand cases of strong specular areas which can change the color structure of text objects have not mentioned yet [1-17].

The authors used multiple images with different viewpoints and polarization angles to extract and remove specular components in [1, 2]. The specular component is separated using color space analysis in [3, 4]. A specular-free (SF) image which is a pseudo diffuse component was introduced in [5-8]. A decomposition technique based on a Mean-Shift Decomposition (MSD), an Eigen-decomposition, can be used to separate two reflection components in the textured surfaces [9]. Yang [10] proposed a method using a Bilateral filter to improve the speed for real-time applications. Shen [11] proposed a real-time algorithm using the intensity ratios between the maximum value and the range of the pixels. Approaches using tensor voting to separate diffuse and specular component in color image are mentioned in [12, 13]. All of them focused on non-text objects in natural images and the highlight regions are relatively small and not too strong.

In text detection and recognition applications, the effect of high-light events has not been taken into account [14]. Some techniques proposed approaches to recognize text in natural images with uneven lighting (shadow, highlight): Gatos et al. [15] used an adaptive binarization and enhancement technique to handle uneven lighting in text images. Du et al. [16] applied entropy-based thresholding on each of R, G, and B channels for caption texts in video sequences. Thillou and Gosselin [17] focused on versatility. They aimed to enhance the complementarily between different metrics in the same color space along with intensity information. These approaches could deal with uneven lighting, but highlight removal issue was not solved completely.

In this paper, we propose an effective method to detect strong highlight regions which could ruin the color structure in natural text images and the way to recover the color formation on areas affected by specular. By analyzing the saliency map achieved after tensor voting process based on saliency magnitude and saturation information of white color, diffuse and specular candidates are determined relatively. These specular nominees are removed and highlight regions are recovered using inpainting technique under the supporting of diffuse candidates.

II. PROPOSED METHOD

A. Specular Detection

1) Pre-processing

The first, a Sobel edge detector is applied to the grayscale image of the input image to detect the positions which belong
to boundary or edge. Since the color of the pixels which are object boundaries is not true color. The edge magnitude, $G(x, y)$ of the Sobel edge is calculated by (1). Pixels having minimum edge magnitude value are considered as non-boundary pixels.

$$G(x, y) = \sqrt{G_x^2(x, y) + G_y^2(x, y)}$$  \hspace{1cm} (1)

All boundary pixels are removed in which non-boundary ones gather into separate color clusters called sampled image. This step reduces the complexity as well as time consuming of method.

To minimize the effect of lighting in the highlight detection and removal process, sampled image is converted into L*a*b color space [19] and only the a-component and b-component are used as a feature vector to encode into a tensor.

2) Specular and Diffuse Candidates Estimating
   a) Tensor Voting

After encoding process is complete, we achieve a tensor space in which each tensor corresponds to a feature vector in the image. Voting process is implemented in this space. The tensors are gradually deformed as a result of the accumulation of votes cast from other neighboring tensors by a ball voting field. The detail of these processes can be seen in [18].

![Fig. 1.](image)

Fig. 1. The original image (a), diffuse (black color) and specular (white color) candidates image (d), saliency map (c), saliency map without noise and highlight tensors (diffuse reflectance distribution ) (d).

Each tensor is presented by two eigenvectors $(e_1, e_2)$ and two corresponding eigen-values $(\lambda_1, \lambda_2)$ which control orientation and saliency of tensor respectively [18]. To determine the specular candidates, we need the information about saliency magnitude that only saliency of tensor is taken in account to analysis and process. Therefore, the saliency map is reproduced by sum of all eigenvalues $(\lambda_1 + \lambda_2)$ of each tensor. As can be seen, Figure 1(c) show the 3-D saliency map of input image in Figure 1(a). Two axes on the horizontal plane illustrate the color space (256 × 256). The vertical axis shows the value of $\lambda_1 + \lambda_2$. Small pink circles describe local maxima of saliency map.

b) Saliency Map Achieving and Analyzing

We observe that saliencies of tensors in the color space are grouped into areas. Each area and local maxima present a corresponding color region and its centroid respectively in the input image. Moreover, in voting process, the saliency of tensors increases proportionality to the number of neighbor tensors inside the voting range. Therefore, a big local maximum in the saliency map implies to alarge corresponding color region is in the input image.

To remove the specular tensor from saliency map, there are three assumes as following:

- The saturation value of specular-only or monochromatic pixels is identical to zero.
- Highlight regions are very strong so that the color of these areas become white quietly.
- Highlight regions are inside objects, not cross over two or more ones or a little bit.

With all local maxima achieved from saliency map, we remove peaks having small saliency and small saturation compared to white color. In proposed method, the thresholds for saliency and saturation values are 10 and 0.2 respectively. The rest of peaks are denoted as the color of centroids of the diffuse regions in the input image, called dominant colors. From these peaks, diffuse and specular candidates in the input image could be discovered using dissimilarity values.

Let n be the number of dominant colors. Let $S = \{S^u, 1 \leq u \leq n\}$ be the set of dominant colors. For each pixel $X(i, j)$ in the $L + a + b$ image, we calculate the dissimilarity to all dominant colors using the Euclidean distance:

$$D_u = \sqrt{(S_u^a - X_a)^2 + (S_u^b - X_b)^2}$$  \hspace{1cm} (2)

Where $X_a, X_b$ are a and b channel values of pixel X, respectively. Because the dissimilarity of specular pixels is high, while that of diffuse pixels is low, for every $S^u \in S$, all minimum dissimilarity values are extracted and considered as the diffuse pixels’ dissimilarity. Then, diffuse candidates are pixels having the minimum RGB value being smaller than that value of average of diffuse pixels’ minimum RGB value. By way of illustration, the Figure 1(b) shows the diffuse and specular candidates of the text image. The black color represents to diffuse candidates and white one to specular candidates.

B. Specular Removal

1) Reflection Model

The dichromatic reflection model is used to describe the reflection of most inhomogeneous materials [19]. The model assumes that the light reflected from an object is a linear combination of diffuse and specular colors:

$$J = J^d + J^s$$  \hspace{1cm} (3)

where $J$ is the reflected light color captured by an RGB camera, $J^d$ and $J^s$ are the diffuse and specular components, respectively.

We assume that the RGB response values are proportional to the intensity of the light incident on the sensor. Let $J_c(x)$ be the color of channel $c$ ($c = 1, 2, 3$, or equivalently, red, green,
blue) at pixel x. The component color value at x can be formulated as [19]:

\[ I_{c}(x) = \alpha(x)I_{b,c} + 255\beta(x) \tag{4} \]

where \( \alpha(x) \) and \( \beta(x) \) denote the geometrical factors of these two reflections at pixel x, respectively. \( I_{b,c} \) denotes the intrinsic body color of the material.

2) Color Structure recovering

Let \( I_{c}(x) \) be the color of the specular pixel x. The color \( I_{sf,c}(x) \) of the corresponding specular-free pixel is calculated by:

\[ I_{sf,c}(x) = I_{c}(x) - I_{\min}(x) \tag{5} \]

where \( I_{r}(x), I_{g}(x), I_{b}(x) \) are the three color values of red, green, and blue of pixel x.

\[ I_{\min}(x) = \min(I_{r}(x), I_{g}(x), I_{b}(x)) \]

\[ = \alpha(x) \min(I_{b,r}, I_{b,g}, I_{b,b}) + 255\beta(x) \tag{6} \]

By combining Eq. (5) and (6), we have:

\[ I_{sf,c}(x) = \alpha(x)(I_{b,c} - I_{\min}) \tag{7} \]

Let \( I_{d,c}(x) \) is the diffuse amount of pixel x, according to Eq. (4), \( I_{d,c}(x) = \alpha(x)I_{b,c} \). We have:

\[ I_{d,c}(x) = I_{c}(x) - 255\beta(x) \]

\[ = I_{sf,c}(x) + I_{\min}(x) - 255\beta(x) \]

\[ = I_{sf,c}(x) + \alpha(x)I_{b,\min}(x) + 255\beta(x) - 255\beta \tag{8} \]

As can be seen in the Eq. (7) and (8), because the color of highlight regions is white, the values of \( \alpha(x) \) and \( \beta(x) \) are approximately 0 and 1 respectively. As a consequence, if we remove the highlight component from original pixels, the color of specular candidates become black completely. The color structure of clusters has not recovered yet. To solve this problem, the inpainting technique is used, which is similar to Texture Synthesis Based method [20]. The color value of each specular pixel is replaced by the average value of neighbor diffuse pixels. In the detail, the region having the number of diffuse pixels is bigger; the recovered color of the specular pixel is more affected by the color of that region.

III. EXPERIMENTS

A. Database

To investigate the performance and the robustness of the proposed method, the available public database ICDAR 2005 is used as standard resource for experiments. Moreover, we use a popular mobile phone (Galaxy S3) to collect our own dataset from real world, which contains images with different levels of highlight regions. We compare the proposed method to two other well-known approaches which are very reliable in this field. Their source codes are available on the website[21,
We also use segmentation results from original text image and our proposed method to confirm again about the substance of highlight removal in natural text images.

### Evaluation

With 100 strong highlight images selected from ICDAR 2005 and our own database, the specular and diffuse component images from three methods are extracted and compared together.

Two images with the highlight regions being very strong are shown in the Figure 2. In these cases, the highlight areas are inside of objects and they destroy the color of objects. The Tan’s and Shen’s results can determine the specular regions acceptably but the diffuse image cannot recover the original color for texts in these regions. The color of these areas becomes dark after highlight removal process. In the proposed method, thanks to the recovery mechanism based on the voting information, we can restore the original color of highlighted text areas expectedly.

The more challenge cases is presented in the Figure 3. This figure includes two images with the highlight regions being very strong, large and cross the objects and background. These images still are obstacles to Tan’s and Shen’s methods. In our case, the result is still acceptable.

With the same segmentation algorithm, we applied to original image and the image achieved from the proposed method to segment image into color clusters. By comparing the number of clusters and the quantity of results, we can realize that the proposed method getting the better results. Aim to fair, we used the segmentation method from Toan [18] which can automatic to cluster color regions in the input images without supports from users.

Figure 4 shows the segmentation results of the original images and of the highlight removed images using the proposed method. With the parameters (sigma = 15, distance = 30) of Toan’s method [18], when applying to original images including the strong highlight regions, the number of segmentation is larger than the number of correct color regions in the original image. In general, the highlight regions are segmented into a cluster in the segmentation result. Meanwhile, by dint of using the highlight removed images produced using the proposed method, the segmentation result is correct.

### IV. Conclusions

In this paper, we introduced an effective approach to remove strong highlight regions in natural text images. By analysing the saliency map achieved after tensor voting process, the dominant colors which represent to color regions in the original image are determined to separate the diffuse and specular regions based on the saturation information of white color and saliency value. Specular candidates are removed and the color of highlight regions is recovered using painting technique combined to reflection model. The proposed method showed expected results in the case of strong highlight regions which destroy the color structure of objects in the image. In the future work, we continue to improve the performance of our proposed method to images with the highlight regions across two or more objects. With our estimation, the proposed method...
method will be an important pre-process for current text recognition applications.

ACKNOWLEDGEMENT

This research was supported partly by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by MEST(2014-024950) and by the technological innovation R&D Program of SMBA(S2173771). The research was also supported by NAFOSTED (the National Foundation for Science and Technology Development in Vietnam). We would like to give thanks to bigDLab@FIT.UoS for constructive comments.

REFERENCES


