

Detecting Text in Natural Scenes Based on a Reduction of Photometric Effects: Problem of Color Invariance

Alain Trémeau¹, Christoph Godau², Sezer Karaoglu², and Damien Muselet¹

¹ Laboratoire Hubert Curien, Batiment E, 18 rue Benoit Lauras, University Jean Monnet, 42000 Saint Etienne, France

{alain.tremeau, daminen.muselet}@univ-st-etienne

² Erasmus Mundus CIMET Master, University Jean Monnet, Batiment B, 18 rue Benoit Lauras, 42000 Saint Etienne, France

Abstract. In this paper, we propose a novel method for detecting and segmenting text layers in complex images. This method is robust against degradations such as shadows, non-uniform illumination, low-contrast, large signal-dependent noise, smear and strain. The proposed method first uses a geodesic transform based on a morphological reconstruction technique to remove dark/light structures connected to the borders of the image and to emphasize on objects in center of the image. Next uses a method based on difference of gamma functions approximated by the Generalized Extreme Value Distribution (GEVD) to find a correct threshold for binarization. The main function of this GEVD is to find the optimum threshold value for image binarization relatively to a significance level. The significance levels are defined in function of the background complexity. In this paper, we show that this method is much simpler than other methods for text binarization and produces better text extraction results on degraded documents and natural scene images.

Keywords: Text binarization, Contrast enhancement, Gamma function, Photometric invariants, Color invariants.

1 Introduction

One of the most challenging tasks for any computer vision system is to recognize the changes in an image which are due to a change in the underlying imaged surfaces from changes which are due to the effects of the scene illumination. The interaction between light and surface is complex and introduces unwanted artifacts into an image [1]. For example, shading, shadows, specularities and inter-reflections, as well as change to local variation in the intensity of color of the illumination all make it more difficult to achieve basic visual tasks such as text extraction (see Fig. 1) or background extraction (see Fig. 2 and 3). Fig. 2 demonstrates that the colors distribution of the background of image (c) of Fig. 1 is not homogeneous and that in the background there is strong chrominance variations. Fig. 3 demonstrates that the colors distribution of the background of image (d) of Fig. 1 is also not homogeneous and that in this region there is greater chrominance variations than in Fig. 1 (c). In order to attenuate

these effects illuminant-invariant models have been proposed. Several studies have shown that these models greatly attenuate most of effects described above. In this paper, in section 2, we show that these models suffer from limitations and do not perform well when addressing complex illumination conditions, such as those illustrated by image of Fig. 1 (d).

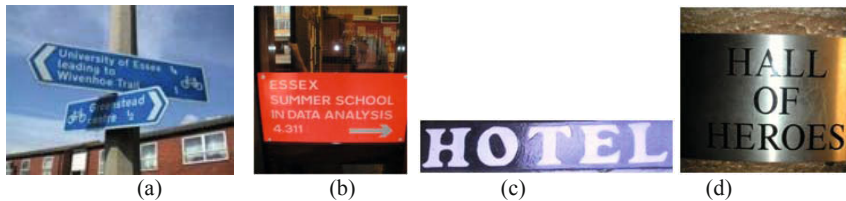
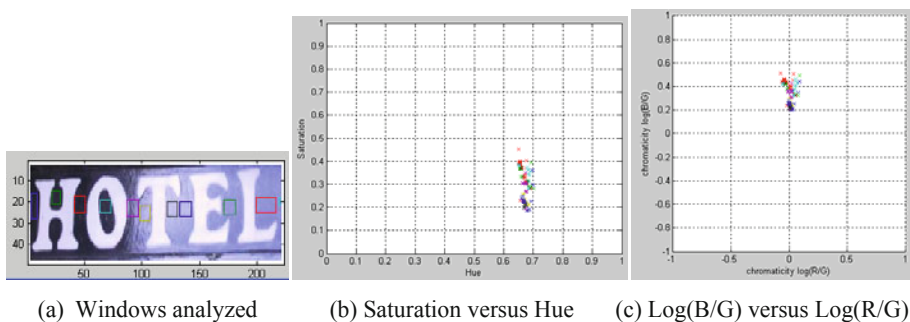


Fig. 1. Color changes due to shading (a), local variation in the intensity of the illumination (b), specularities (c), and specularities and inter-reflections (d)

Another challenging task is to enhance the image so that result is more suitable than original image for specific application such as segmentation. Several image enhancement techniques, often elementary or heuristic methods, have been proposed for improving the visual quality of images. Appropriate choice is greatly influenced by the imaging modality, task at hand and viewing conditions [2]. As example, power-law transformations with a fractional exponent can be used to expand the gray scale range of dark images. Log Transformation can be used for enhancing details in darker regions but at the expense of details in higher-level values, i.e. brighter regions. Histogram equalization can be used to stretch the contrast of an image by redistributing the gray-level values uniformly. In section 3 we show that these models suffer also from limitations and do not perform well when addressing complex illumination conditions, such as those illustrated by image of Fig. 1 (d).



(a) Windows analyzed

(b) Saturation versus Hue

(c) $\text{Log}(B/G)$ versus $\text{Log}(R/G)$

Fig. 2. Color space representation of different windows of image (c) in Fig. 1. Colors associated with the points (in (b) and (c)) represent color of corresponding windows (in (a)). Each window is represented by four points which corresponds to the lowest and highest values computed in this window for the two dimensions considered.

Color image segmentation is also a challenging task as solutions have to be effective against image shadows, illumination variations and highlights. Several approaches based on the computation of image invariants that are robust to photometric effects have been proposed in the literature [3-5]. Unfortunately, there are too many color invariant models in the literature, making the selection of the best model and its combination with local image structures (e.g. color derivatives) quite difficult to produce optimal results [6]. In [7], Gevers et al. survey the possible solutions available to the practitioner. In specific applications, shadow, shading, illumination and highlight edges have to be identified and processed separately from geometrical edges such as corners, and T-junctions. To address this issue, Gevers et al. proposed to compute local differential structures and color invariants in a multidimensional feature space to detect salient image structures (i.e. edges) on the basis of their physical nature in [7]. In [8] the authors proposed a classification of edges into five classes, namely object edges, reflectance edges, illumination/shadow edges, specular edges, and occlusion edges to enhance the performance of the segmentation solution utilized. Shadow segmentation is of particular importance in applications such as video object extraction and tracking. Several research proposals have been developed in an attempt to detect a particular class of shadows in video images, namely moving cast shadows, based on the shadow's spectral and geometric properties [9]. The problem is that cast shadow models cannot be effectively used to detect other classes of shadows, such as self shadows or shadows in diffuse penumbra [9] suggesting that existing shadow segmentations solutions could be further improved using invariant color features. The main challenge in color image segmentation is since a decade the fusion of low level image features so that image content would be better described and processed. Several researches provided some solutions to combine color derivatives features and color invariant features, color features and other low level features (e.g. color and texture, color and shape [7]), low-level features and high-level features (e.g. from graph representation [10]). However, none of the proposed solutions appear to provide the expected performance to segment complex color images unlike the human visual system which is able to take into account the semantic contents of images. Of course if some a priori information or knowledge about the segmentation task is incorporated in the process that will optimise the algorithm results.

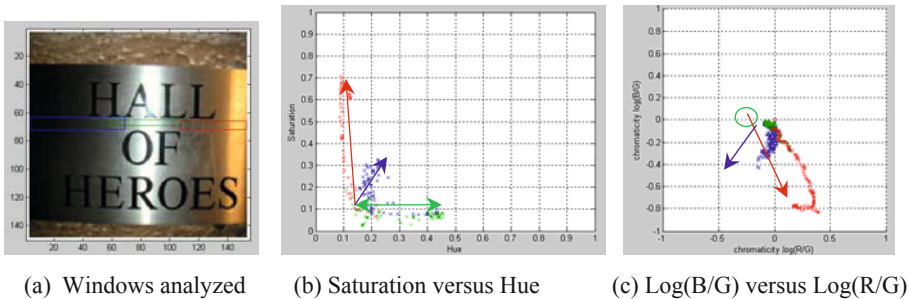


Fig. 3. Color space representation of different areas of image of Fig. 1 (d). Colors associated with the points (in (b) and (c)) represent color of corresponding windows (in (a)). Each window is represented by several points computed by sampling of this window. The arrows represent color changes from left to right in the corresponding windows in (a).

Most of existing text segmentation approaches assume that text layers are of uniform color and fail when this is not the case. The background may also be multicolor consequently the assumption according with it is the largest area of (almost) uniform color in the image does not necessarily hold [11]. Lastly, most of existing text segmentation approaches assume that there is a high contrast between text and background in the image this is unfortunately not always the case in real images. Many approaches assume also that in segmenting the highest peak in the lightness histogram we can deduce if text layers are of lower or a higher lightness than the background region, this information may be helpful to segment text layers, but this is once again not always the case in real images.

In this paper we demonstrate that none illuminant-invariant model is sufficiently robust to complex photometric effects to solve the issue of text detection in complex natural scenes. To solve this issue, in a second paper [12], we propose to use another strategy, more robust to photometric effects, based on the computation of the difference of gamma functions to detect text layers in complex scenes.

2 Color Spaces Invariants to Photometric Effects

2.1 Illuminant-Invariant Models

A first approach to compute illuminant-invariant consists to use reflection models (e.g. Lambertian or dichromatic reflectance) but these reflection models are too restricted to model real-world scenes, such as the scene illustrated by Fig. 3, in which different reflectance mechanisms hold simultaneously. Different photometric invariance models based on dichromatic reflection models have been proposed [13-14]. These models are invariant for different types of lighting variations, i.e. light intensity (LI) or light color (LC) change and/or light intensity or light color shift (see Table 1) but are not invariant to strong or complex lighting changes such as those illustrated by image of Fig. 1 (d).

Table 1. Invariance of color models for different types of lighting changes [14]. Invariance is indicated with '+' and lack of invariance with '-'

Color space	LI change	LI shift	LI change and shift	LC change	LC change and shift
Value (V)	-	-	-	-	-
Saturation (S)	-	+	+	-	-
Hue (H)	+	+	+	-	-
$\log(R/G)$, $\log(B/G)$	+	+	+	+	+

Fig. 4 and Table 2 show the non-invariance of H, S and $\log(B/G)$ color descriptors for a simple real-world scene. These "invariants" fail to attenuate a strong intensity shift due to a smooth specular reflection. Likewise, Fig. 5 and Table 3 show the non-invariance of H, S, $\log(B/G)$ and $\log(B/G)$ color descriptors for a more complex real-world scene.

In order to attenuate photometric effects for text segmentation in natural images Jim et al. proposed in [15] to decompose the image studied into chromatic and achromatic regions in the HSI color space using a decision function computed in RGB

space. Since the HSI space and the decision function used are not invariant to photometric effects this method is not relevant for complex images such as image of Fig. 1 (d). Furthermore, this method is based on the hypothesis that text layers have homogeneous values in the chromaticity image but this is not always the case for complex images, such as in Fig. 6. In Fig. 6 there are two color changes, one from left to center due to a white specular highlight, and a second from center to right due to a second (orange) specular highlight. Karatzas et al. proposed also to decompose the image studied into chromatic and achromatic regions in the HLS color space using a decision function based on saturation and lightness values. The problem with this strategy is that the exact levels of saturation and lightness for which color should be considered achromatic are not straightforward to set [11]. To simplify the problem either empirical or heuristic methods are generally used.

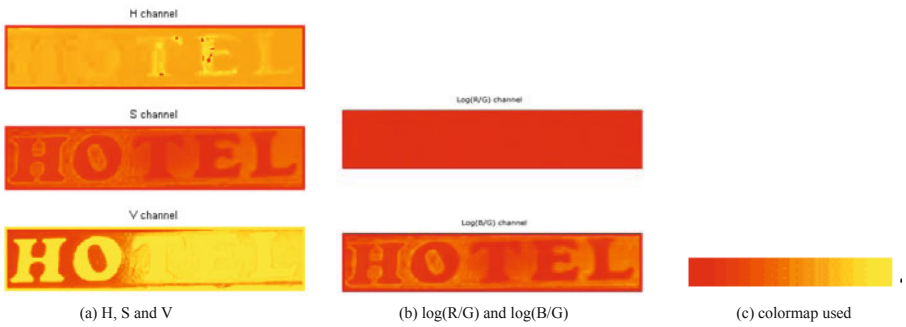


Fig. 4. Representation of H, S, V, log(R/G) and log(B/G) components computed for image of Fig. 1 (c). The V channel shows a LI shift in the background from left to right ($\Delta V_{\max 2}=0.71$). On each component, reddish colors and yellowish colors correspond to values closed to 0 and 1, respectively.

Table 2. Invariance of color models / LI change in image (c) of Fig. 1. Color component are normalised in the interval [0, 1]. Invariance is indicated with ‘+’ and lack of invariance with ‘-’. ΔX_{ext} and $\Delta X_{\text{max}1}$ represents, respectively, the maximum of difference between two consecutive areas of the background (from left to right) and the maximal difference between two areas of the background, computed from mean values calculated for each letter for color component X. $\Delta X_{\text{max}2}$ and $\Delta X_{\text{max}3}$ represents, respectively, the maximum of differences between two consecutive letters (from left to right) and the maximum of differences between two letters, computed from mean values calculated for each letter for color component X.

Color space	Background	Letters
Saturation (S)	- $\Delta S_{\text{ext}}=0.16$, $\Delta S_{\text{max}1} = \mathbf{0.61}$	- $\Delta S_{\text{cl}}=0.04$, $\Delta S_{\text{max}2} = \mathbf{0.12}$
Hue (H)	- $\Delta H_{\text{ext}}=0.25$, $\Delta H_{\text{max}1} = \mathbf{0.30}$	- $\Delta H_{\text{max}2}=0.10$, $\Delta H_{\text{max}3} = \mathbf{0.13}$
log(R/G)	+ $\Delta \log(R/G)_{\text{ext}}=0.03$, $\Delta \log(R/G)_{\text{max}1} = \mathbf{0.04}$	+ $\Delta \log(R/G)_{\text{max}2}=0.01$, $\Delta \log(R/G)_{\text{max}3} = \mathbf{0.01}$
log(B/G)	- $\Delta \log(B/G)_{\text{ext}}=0.19$, $\Delta \log(B/G)_{\text{max}1} = \mathbf{0.20}$	- $\Delta \log(B/G)_{\text{max}2}=0.01$, $\Delta \log(B/G)_{\text{max}3} = \mathbf{0.16}$

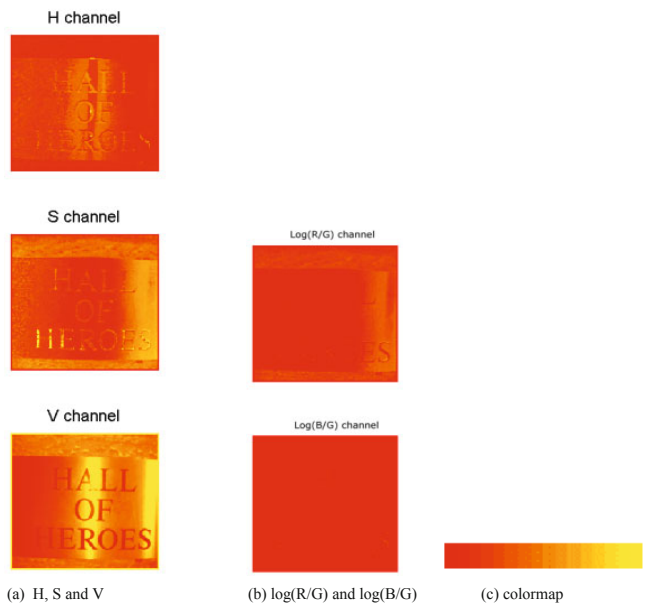


Fig. 5. Representation of H, S, V, log(R/G) and log(B/G) components computed for image (d) of Fig. 1. The V channel shows two LI shifts in the background ($\Delta V_{\max 2} = 0.94, 0.92$ resp.). The first LI shift (i.e. at center) corresponds also to a LC shift on the H channel meanwhile the second LI shift (i.e. at right) corresponds to a LC shift on the S channel. On each component, reddish colors and yellowish colors correspond to values closed to 0 and 1, respectively.

Table 3. Invariance of color models / LI and LC change and shift in image (d) of Fig. 1. Color component are normalized in the interval [0, 1]. Invariance is indicated with ‘+’ and lack of invariance with ‘-’. ΔX_{ext} and $\Delta X_{\max 1}$ represents, respectively, the maximum of difference between two consecutive areas of the background (from left to right) and the maximal difference between two areas of the background, computed from mean values calculated for each letter for color component X. $\Delta X_{\max 2}$ and $\Delta X_{\max 3}$ represents, respectively, the maximum of differences between two consecutive letters (from left to right) and the maximum of differences between two letters, computed from mean values calculated for each letter for color component X.

Color space	Background	Letters
Saturation (S)	- $\Delta S_{\text{ext}} = 0.39$, $\Delta S_{\max 1} = \mathbf{0.60}$	- $\Delta S_{\text{ext}} = 0.36$, $\Delta S_{\max 2} = \mathbf{0.55}$
Hue (H)	- $\Delta H_{\text{ext}} = 0.25$, $\Delta H_{\max 1} = \mathbf{0.32}$	- $\Delta H_{\max 2} = 0.24$, $\Delta H_{\max 3} = \mathbf{0.32}$
log(R/G)	- $\Delta \log(R/G)_{\text{ext}} = 0.26$, $\Delta \log(R/G)_{\max 1} = \mathbf{0.42}$	- $\Delta \log(R/G)_{\max 2} = 0.2$, $\Delta \log(R/G)_{\max 3} = \mathbf{0.31}$
log(B/G)	- $\Delta \log(B/G)_{\text{ext}} = 0.40$, $\Delta \log(B/G)_{\max 1} = \mathbf{0.77}$	- $\Delta \log(B/G)_{\max 2} = 0.49$, $\Delta \log(B/G)_{\max 3} = \mathbf{0.87}$

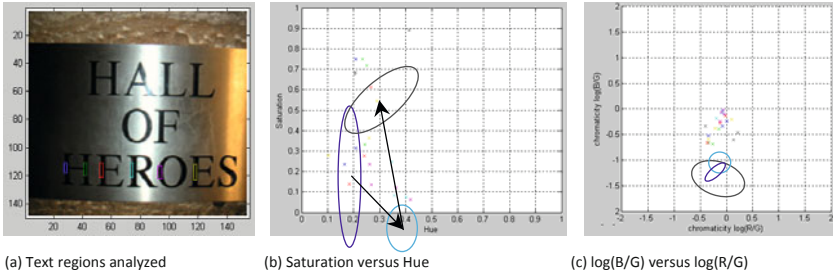


Fig. 6. Color space representation of several text regions of image (d) in Fig. 1. Colors associated with the points (in (b) and (c)) represent color of corresponding windows (in (a)). Each window is represented by four points which corresponds to the lowest and highest values computed in this window for the two dimensions considered. The color distribution of the blue window (at left on (a)), of the cyan window (at the center) and of the grey window (at right, on the S) are bounded by ellipses in (b) and (c). The arrows represent color changes between these ellipses that is to say from left to center and from center to right in (a).

2.2 Logvinenko's Model

In order to attenuate photometric effects such as shading, shadows, specularities and inter-reflections, as well as change to local variation in the intensity of color of the illumination a new approach has been recently proposed by Logvinenko in 2009 [16]. The main idea of this approach is to consider that the set of possible colors of reflecting objects defines a volume called the object-color solid [17]. The object-color solid depends on the spectral power distribution of the illuminant and the color space being used. Logvinenko has therefore proposed a new object-color space that defines a complete color atlas that is invariant to illumination [16] and describes all colors in the object-color solid under any illuminant. However, Logvinenko's existing implementation for calculating the proposed color descriptors is computationally expensive and does not work for all types of illuminants. Fig. 7 shows the calculated color descriptors for a scene viewed under different lighting sources (images from the Barnard's Database [18]) and illustrates the perceptual correlates of the descriptors ($\alpha\delta\lambda$) as described by Logvinenko. Purity (α) describes the grayness of a color, namely, the relative distance to the gray center. Hence, both black and white have a high purity, so for example the dark areas of the image show as white in Fig. 7 (b). The spectral bandwidth (δ) (see Fig. 7 (c)) correlates with blackness and whiteness, but becomes meaningless for low purities (gray colors). Central wavelength (λ) (see Fig. 7 (d)) is correlated to hue.

A possible application of $\alpha\delta\lambda$ color descriptors is to predict the effect of an illumination change. Since $\alpha\delta\lambda$ coordinates determine a metameric reflectance spectrum, the corresponding sensor response for any given illuminant can be calculated according to Logvinenko's model. Fig. 7 shows the result obtained by calculating the $\alpha\delta\lambda$ color descriptors using different illuminants. Although the $\alpha\delta\lambda$ space is invariant to illumination, the $\alpha\delta\lambda$ color descriptors themselves can change with the illumination, since each $\alpha\delta\lambda$ triplet describes a class of metamers, and metamerism depends on the illuminant. To analyze the effect of each illumination change on $\alpha\delta\lambda$ color descriptors, we have computed the histogram of each component next

computed distance histograms (see Tables 4 to 12) based on bin-by-bin dissimilarity metrics such as the χ^2 divergence, the Kullback-Leibner divergence and the Jeffrey divergence [19]. Fig. 8 shows the descriptors for images book3 and Macbeth under different intensities, further illustrating the perceptual correlates (images from the Barnard's Database [18]).

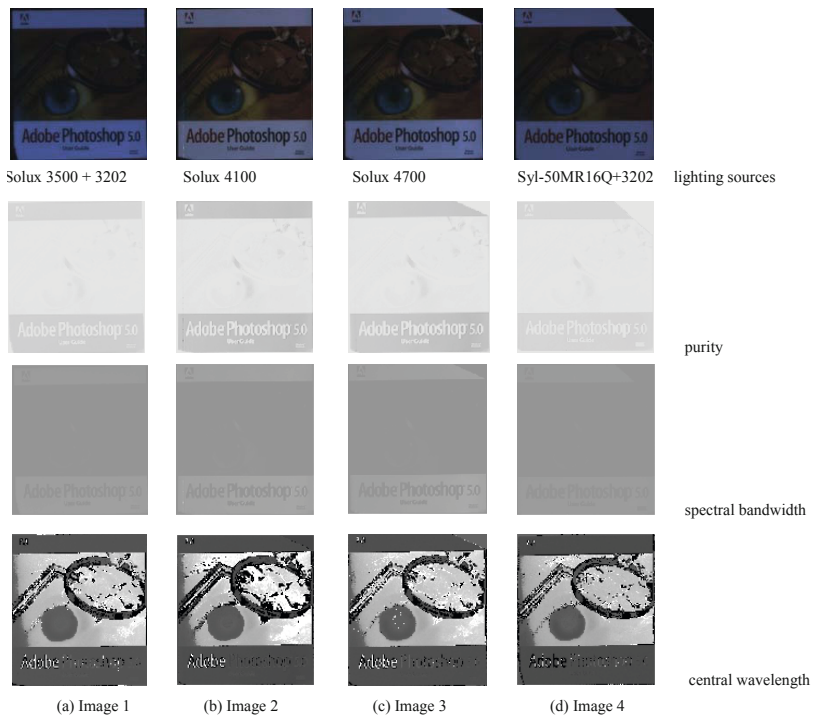


Fig. 7. Image book3 under different lighting sources

Table 4. Dissimilarity distances (Ki2 normalized) computed from histogram of Purity component

Images	Im1	Im2	Im3	Im4	Im5	Im7
Im1		0.60	0.49	0.27	1.33	
Im2			0.19	0.62	1.70	
Im3				0.55	1.63	
Im4					1.44	
Im5						
Im6						1.46

Table 6. Dissimilarity distances (Jeffrey) computed from histogram of Purity component

Images	Im1	Im2	Im3	Im4	Im5	Im7
Im1		0.38	0.31	0.16	0.89	
Im2			0.10	0.40	1.11	
Im3				0.36	1.07	
Im4					0.96	
Im5						
Im6						0.96

Table 5. Dissimilarity distances (Kullback-Leibler) computed from histogram of Purity component

Images	lm1	lm2	lm3	lm4	lm5	lm7
lm1		0.91	0.78	0.52	0.07	
lm2			0.19	0.10	0.56	
lm3				0.69	0.31	
lm4					0.01	
lm5						
lm6						2.50

Table 7. Dissimilarity distances (Ki2 normalized) computed from histogram of Bandwidth component

Images	lm1	lm2	lm3	lm4	lm5	lm7
lm1		0.53	0.37	0.24	1.13	
lm2			0.09	0.61	0.89	
lm3				0.55	0.93	
lm4					1.15	
lm5						
lm6						1.09

Table 8. Dissimilarity distances (Kullback-Leibler) computed from histogram of Bandwidth component

Images	lm1	lm2	lm3	lm4	lm5	lm7
lm1		0.76	0.48	0.81	1.55	
lm2			0.09	0.02	1.30	
lm3				0.17	1.23	
lm4					1.60	
lm5						
lm6						1.73

Table 9. Dissimilarity distances (Jeffrey) computed from histogram of Bandwidth component

Images	lm1	lm2	lm3	lm4	lm5	lm7
lm1		0.33	0.23	0.16	0.67	
lm2			0.05	0.40	0.53	
lm3				0.36	0.54	
lm4					0.69	
lm5						
lm6						0.70

Table 10. Dissimilarity distances (Ki2 normalized) computed from histogram of Central wavelength component

Images	lm1	lm2	lm3	lm4	lm5	lm7
lm1		0.79	0.74	0.66	0.61	
lm2			0.64	0.73	0.80	
lm3				0.14	0.51	
lm4					0.62	
lm5						
lm6						0.76

Table 11. Dissimilarity distances (Kullback-Leibler) computed from histogram of Central wavelength component

Images	lm1	lm2	lm3	lm4	lm5	lm7
lm1		1.63	1.29	1.25	0.72	
lm2			1.15	1.27	1.31	
lm3				0.15	0.63	
lm4					0.80	
lm5						
lm6						1.09

Table 12. Dissimilarity distances (Jeffrey) computed from histogram of Central wavelength component

Images	lm1	lm2	lm3	lm4	lm5	lm7
lm1		0.49	0.45	0.41	0.35	
lm2			0.40	0.45	0.49	
lm3				0.07	0.28	
lm4					0.35	
lm5						
lm6						0.49

The less images are dissimilar the lower the distance between these histograms is (the three distances considered are bounded by the min value 0). To illustrate our purpose, let us consider images of Fig. 8 and 9. The lowest dissimilarity distances (see Tables 4 to 12) are between images of Fig. 7 (b) and Fig. 7 (c) and between images of

Fig. 7 (c) and Fig. 7 (d). The lowest value for the purity, bandwidth and central wavelength are respectively equal to 0.10, 0.04 and 0.40 for images of Fig. 7 (b) and Fig. 7 (c), and to 0.36, 0.17 and 0.13 for images of Fig. 7 (c) and Fig. 7 (d). These values are not constant. Whatever the set of illuminants considered and the image studied (images of the Barnard's Database [18]) distance between histograms computed are also not constant. This shows that purity, bandwidth and central wavelength components are not invariant to visually perceptible light color changes (see also the first example given in Fig. 7).

Let us now consider images of Fig. 8. The lowest dissimilarity distances (see Tables 4 to 12) between images of Fig. 8 (a) and Fig. 8 (b) and between images of Fig. 8 (c) and Fig. 8 (d) are respectively, for α , δ and λ color descriptors, equal to 0.31, 0.54 and 0.28 for images of Fig. 8 (a) and Fig. 8 (b), and to 0.96, 0.70 and 0.49 for images of Fig. 7 (c) and Fig. 7 (d). These values linked to light intensity changes are not constant and are higher than for illumination changes. Whatever the image studied (images of the Barnard's Database [18]) and light intensity changes, distance between histograms computed are also not constant. This shows that purity, bandwidth and central wavelength components are not invariant to visually perceptible light intensity changes. Our results show that central wavelength (λ) values are less sensitive to light intensity changes than spectral bandwidth (δ) and purity (α), as the former descriptor is correlated to hue meanwhile spectral bandwidth correlates with blackness and whiteness and purity correlates with greenness.

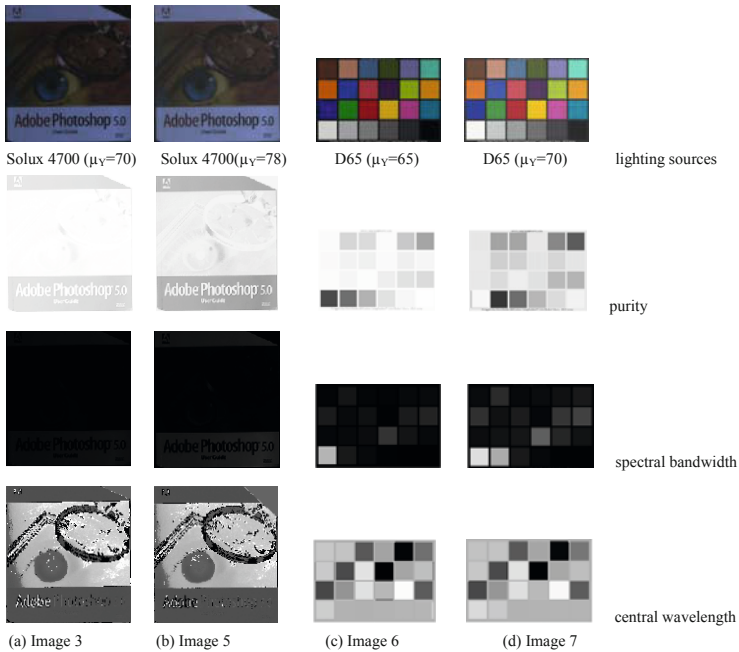


Fig. 8. Images book3 and Macbeth under different intensities



Fig. 9. Representation of α , δ and λ components computed for image (d) in Fig. 1

3 Image Enhancement Methods

3.1 Grey Levels Transformation Methods

In order to better attenuate photometric effects such as shading, shadows, specularities and inter-reflections, as well as change to local variation in the intensity of color of the illumination another approach consists to apply an image enhancement method [20]. The principle objective of image enhancement is to process an image so that results are more suitable than original for specific application [2]. The main task, in our case of study, is to optimize intra-class similarity and inter-class similarity such as every grey-level class (i.e. text regions and background regions) can be well separated.

Different image enhancement algorithms can be used to improve the appearance of an image such as its contrast in order to make the image interpretation, understanding, and analysis easier. Various contrast enhancement algorithms have been developed to modify the appearance of images by highlighting certain features while suppressing others. A widely used approach for contrast enhancement is based on the use of a power law response equation such as follows (see Fig. 10):

$$s = cr^\gamma \tag{1}$$

Generally c and γ are positive constants. r and s represent respectively the input and output intensity levels (see [21]).

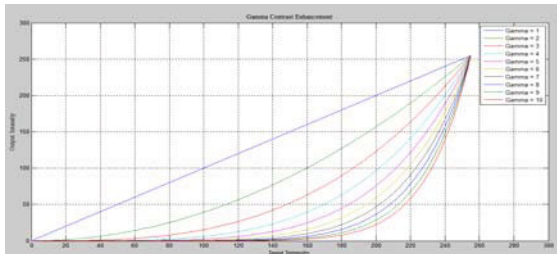


Fig. 10. Influence of the parameter gamma on the contrast of the output image

Another function, the piecewise linear transformation (see Fig. 11) can also be used to modify the distribution of the grey levels of any 1D component such as the saturation or the brightness (see Fig. 12 and 13). In Fig. 12 there are two classes: the background (brighter region) and the text layer (darker regions). The best results that we have experimentally obtained with a piecewise linear transformation are given in Fig. 12 (b). As the most illuminant-invariant components for darker regions (i.e. text layer) are S and $\log(B/G)$ components (see Table 2) we apply this transformation on these components. As we can see on Fig. 12 results are not perfect as on the original image (on the most illuminant-invariant components) the inter-class similarity is quite high, e.g. some pixels of the background have the same saturation value than the text layer). In Fig. 13 there are three classes: the background (brighter and darker regions) and the text layer (intermediate regions). The best results that we have experimentally obtained with a piecewise linear transformation are given in Fig. 13 (b). As the most illuminant-invariant components for text regions are V and $\log(R/G)$ components (see Table 3) and as the gray scale range of $\log(R/G)$ component is too low (see Fig. 5 (b)), we apply this transformation only the V component. As we can see on Fig. 13

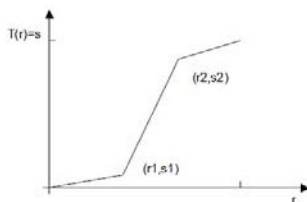


Fig. 11. Example of piecewise linear transformation function used for reducing details in darker regions and brighter regions but at the expense of enhancing details in intermediate regions

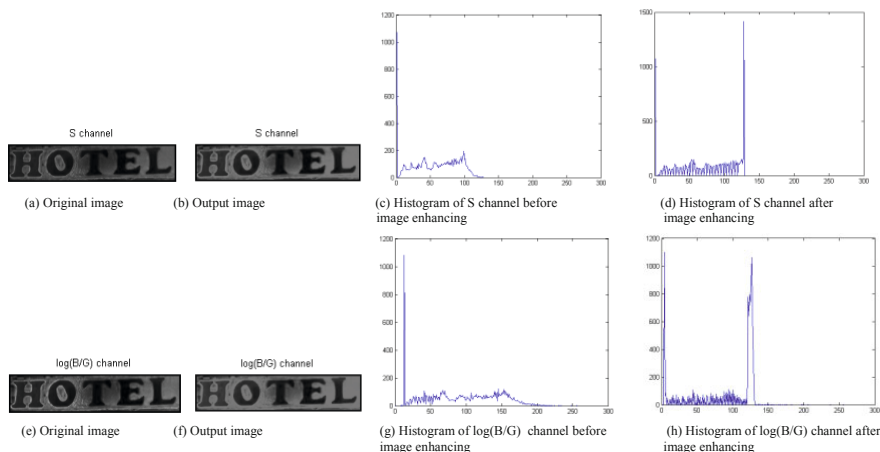


Fig. 12. Examples of image enhancing obtained from a piecewise linear transformation. This transformation is used for enhancing contrast between darker regions (i.e. the text layer) and brighter regions (i.e. the background) without enhancing too much details in intermediate regions. For S we have used the values $(r1=15, s1=5)$ and $(r2=75, s2=120)$ and for $\log(B/G)$ the values $(r1=90, s1=120)$ and $(r2=110, s2=130)$.

results are not very good as on the original image the inter-class similarity is too high as some pixels of the background have the same V value than the text). We can note also on Fig. 6 that for darker regions and brighter regions the intra-class similarity is higher on log(R/G) and H components than for log(B/G) and S components for text regions, but as the inter-class similarity of the background and text regions (see Fig. 3) is lower than the intra-class similarity of each of these two classes we cannot attenuate as desired photometric effects.

3.2 Morphological Reconstruction Based on Geodesic Transform

In order to suppress lighter objects (e.g. text layers) than their surroundings and connected to border of the image, another strategy consists to use a morphological reconstruction transform based on geodesic dilation.

According to Soille [22] geodesic dilation of a bounded image always converges after a finite number of iterations (i.e. until the proliferation or shrinking of the marker image is totally impeded by the mask image). For this reason geodesic dilation is considered as a powerful morphological reconstruction scheme. The reconstruction by dilation $R_g^{\partial}(f)$ of a mask image (g) from a marker image (f) is defined as the geodesic dilation of (f) with respect to (g) iterated until stability as follows (see Fig. 14):

$$R_g^{\partial}(f) = \partial_g^{(i)}(f) \tag{2}$$

The stability is reached at the iteration i when: $\partial_g^{(i)}(f) = \partial_g^{(i+1)}(f)$. This reconstruction is constrained by the following conditions that both (f) and (g) images must have

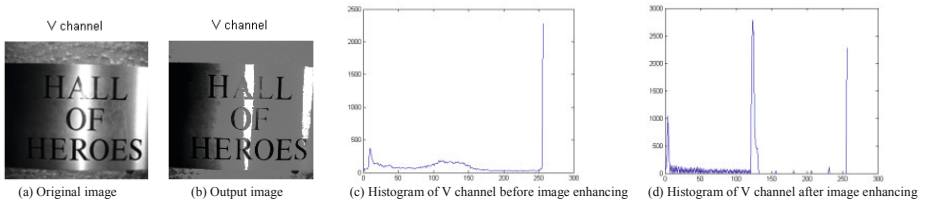


Fig. 13. Examples of image enhancing obtained from a piecewise linear transformation. This transformation is used for enhancing the contrast between intermediate regions (i.e. the text) and darker and brighter regions (i.e. the background) without enhancing too much details in background regions. In this example ($r_1=15, s_1=5$), ($r_2=75, s_2=120$), ($r_3=240, s_3=130$) for V.

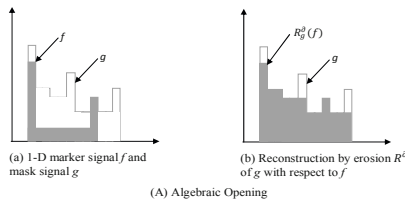


Fig. 14. Algebraic opening for a 1-D signal

the same definition domain (i.e. $D_f = D_g$) and $f \leq g$. This reconstruction transform presents several properties: it is increasing ($g_1 \leq g_2 \Rightarrow R_{g_1}^\partial(f) \leq R_{g_2}^\partial(f)$), anti-extensive ($R_g^\partial(f) \leq g$), and idem-potent ($R_g^\partial(R_g^\partial(f)) = R_g^\partial(f)$). This reconstruction transform corresponds to an algebraic closing of the mask image. The connected opening transformation, $\gamma_x(g)$ of a mask image (g) can be defined as:

$$\gamma_x(g) = R_g^\partial(f_x) \tag{3}$$

where the marker image f_x equals to zero everywhere except as x which has a value equal to that of the image (g) at the same position.

According to Soille [22] the connected opening transformation can be used to extract connected image objects having higher intensity values than their surrounding when we chose the mask image zero everywhere, except for the point x which has a value equal to that of the image (g) at the same position (see Fig. 15).

In order to suppress lighter objects than their surroundings and connected to border of the image, we choose the marker image zero everywhere except the border of the image. At the border of the image we chose the pixel value of marker the same as mask pixel value at the same position. Once we get the connectivity information with the help of morphological reconstruction based on geodesic transform, we suppress these lighter objects connected to image border. After this preprocess step most of the non-text regions are reduced and kept only most probable text layer candidates which leads us to emphasize more on region of interest of the image (see Fig. 15 (b)). Especially in our experiments we have seen that this process reduce the background intensity variations and enhance the text layers of the image.

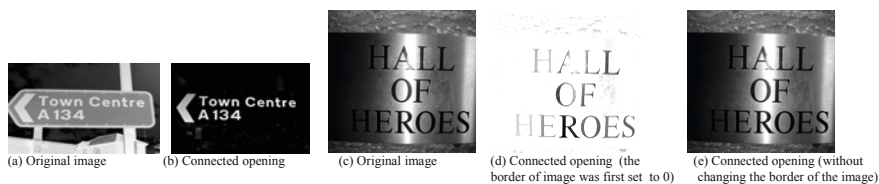


Fig. 15. (b) Connected image objects having higher intensity values than their surrounding can be extracted by the connected opening transform. (d) Connected image objects having darker intensity values than their surrounding can be extracted by the connected closing transform.

In order to suppress darker objects (e.g. text layers) than their surroundings and connected to border of the image the connected closing transformation can be used. The first shortcoming of this morphological transformation and of the former (i.e. closing and opening) is that we must first estimate if the background is lighter or darker than the text layers, i.e. we must first extract the background of the image. The second shortcoming of these two transformations is that they work quite fine when text layers are only darker or whiter than the background but do not perform well when text layers are darker and whiter than their surrounding local background in the image. Lastly, these transforms do not work well when the border of the image has the same intensity than text layers, such as in image (d) in Fig. 1. That why, to enhance this

image we set its borders to zero before applying the connected closing transform (see Fig. 15 (d)) otherwise this transform is inefficient (see Fig. 15 (e)).

4 Conclusion

In this paper we have shown that even if illuminant-invariant models greatly attenuate most of photometric effects these models suffer from limitations when addressing complex illumination conditions. We have also shown that even if most of image enhancement methods can greatly attenuate most of photometric effects these methods suffer from limitations which makes them inoperative when addressing complex illumination conditions. To face this issue we propose to use an image enhancement method based on morphological reconstruction through geodesic transform. This method is used to remove objects connected to borders which are lighter than their surroundings. The idea behind this method is that in images/videos of natural-scenes, the most interesting information area at the center of the image not at its periphery. In this context, we consider as noise any region which does not belong to text layers and which are lighter than their surroundings. In other words, noise corresponds to regions which do not belong to the background region and which are not necessarily connected to image borders. Thanks to the proposed method most of the noise in the image is removed which facilitates further processing steps.

Another approach used in several studies consists firstly to use Gaussian mixture models (GMMs) with spatial connectivity information to model color distributions of text layers and background regions [23], next to reduce local variations in the corresponding regions. We do not believe that the GMMs are effective to model complex images, such as that illustrated by image (d) in Fig. 1. For complex backgrounds there may be several classes that overlap, i.e. when the intra-class variances are high and the inter-class variances are low as in Fig. 3. Furthermore, there may be an overlap between classes, i.e. between the background and the text layers, such as in image (d) in Fig. 1 where the color of the background is very close in the left part of the figure to that of text layers. In such cases, characters may not be segmented as separate objects of background. Such a strategy is equivalent to segment the image first.

To solve the issue of shadowing, reflection and uneven illumination, we have shown in another paper (see [12]) that the Generalized Extreme Value Distribution (GEVD) is a very relevant model to approximate differences of gamma functions. Indeed GEVD is capable of finding proper extreme values based on image statistics allowing us to deal with extreme conditions like shadows, high illuminations and reflections. To solve the problem of low contrast between text and background, we have shown in paper [12] that the difference of gamma functions is a very relevant model to enhance contrast between text layers and background regions while reducing noise, using substantially lesser complex processes than other well-known approaches. Lastly, we have also demonstrated that the propose method is robust to photometric effects.

References

1. Finlayson, G.D., Hordley, S.D., Lu, C., Drew, M.S.: On the Removal of Shadows From Images. *IEEE Pattern Analysis and Machine Intelligence (PAMI)* 28(1), 59–68 (2006)
2. Maini, R., Aggarwal, H.: A Comprehensive Review of Image Enhancement Techniques. *Journal of Computing* 2(3), 8–13 (2010)

3. van de Weijer, J., Gevers, T., Geusebroek, J.M.: Edge and corner detection by photometric quasi-invariants. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 27(4), 625–630 (2005)
4. Li, B., Xue, X., Fan, J.: A robust incremental learning framework for accurate skin region segmentation in color images. *Pattern Recognition* 40(12), 3621–3632 (2007)
5. Moreno-Noguer, F., Sanfeliu, A., Samaras, D.: Integration of deformable contours and a multiple hypotheses Fischer color model for robust tracking in varying illuminant environments. *Image and Vision Computing* 25, 285–296 (2007)
6. Trémeau, A., Tominaga, S., Plataniotis, K.: Color in Image and Video Processing: most recent trends and future research directions. *EURASIP Journal on Image and Video Processing* 2008, article ID 581371, 26 pages (2008)
7. Gevers, T.: Chapter 9: Color feature detection. In: *Color Image Processing: Methods and Applications Book*, pp. 203–226. CRC Press, Boca Raton (2007)
8. Koschan, A., Abidi, M.: Detection and classification of edges in color images. *IEEE Signal Processing Magazine*, 64–73 (2005)
9. Salvador, E., Cavallaro, A., Ebrahimi, T.: Cast shadow segmentation using invariant color features. *Computer Vision and Image Understanding* 95, 238–259 (2004)
10. Dong, G., Xie, M.: Color clustering and learning for image segmentation based on neural networks. *IEEE Trans. on Neural Networks* 16, 925–936 (2005)
11. Karatzas, D., Antonacopoulos, A.: Colour text segmentation in web images based on human perception. *Image and Vision Computing* 25, 564–577 (2007)
12. Trémeau, A., Fernando, B., Karaoglu, S., Muselet, D.: Detecting text in natural scenes based on a reduction of photometric effects: problem of text detection. In: *Proceedings of CCIW 2011. LNCS*, vol. 6626, pp. 217–233. Springer, Heidelberg (2011)
13. van de Sande, K.E.A., Gevers, T., Snoek, C.G.M.: Evaluation of color descriptors for object and scene recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 453–464 (2008)
14. Álvarez, J.M., Gevers, T., López, A.M.: Learning Photometric Invariance for Object Detection. *Int. J. Comput. Vis.* 90, 45–61 (2010)
15. Lim, J., Park, J., Medioni, G.G.: Text segmentation in color images using tensor voting. *Image and Vision Computing* 25, 671–685 (2007)
16. Logvinenko, A.D.: An object-color space. *J. Vis.* 9(11), 1–23 (2009)
17. Wyszecki, G., Stiles, W.S.: *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd edn. Wiley-Interscience, Hoboken (August 2000)
18. Barnard, K.: *Practical Colour Constancy*, Phd thesis, Simon Fraser University, School of Computing (1999), http://kobus.ca/research/data/objects_under_different_lights/index.html
19. Rubner, Y., Puzicha, J., Tomasi, C., Buhmann, J.M.: Empirical Evaluation of Dissimilarity Measures for Color and Texture. In: *International Conference on Computer Vision*, vol. 2, p. 1165 (1999)
20. Chen, S., Beghdadi, A.: Natural enhancement of color image. *Eurasip Journal on Image and Video Processing*, 19 pages (2010), doi:10.1155/2010/175203
21. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*, 2nd edn. Pearson Education, London (2008) ISBN-13: 978-0135052679
22. Soille, P.: *Morphological Image Analysis: Principles and Applications*, pp. 182–198. Springer, Heidelberg (2003)
23. Ye, Q., Gao, W., Huang, Q.: Automatic text segmentation from complex background. In: *IEEE Int. Conf. on Image Processing*, vol. 5, pp. 2905–2908 (2004)