HAL
open science

Gamma correction acceleration for real-time text extraction from complex colored images

Rostom Kachouri, Christian Medina Armas, Mohamed Akil

To cite this version:

Rostom Kachouri, Christian Medina Armas, Mohamed Akil. Gamma correction acceleration for real-time text extraction from complex colored images. Image Processing (ICIP) 2015, Sep 2015, Quebec City, QC, Canada. 10.1109/ICIP.2015.7350854 . hal-01305887

HAL Id: hal-01305887

https://hal.science/hal-01305887

Submitted on 21 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

GAMMA CORRECTION ACCELERATION FOR REAL-TIME TEXT EXTRACTION FROM COMPLEX COLORED IMAGES

Rostom Kachouri, Christian Medina Armas and Mohamed Akil

Université Paris-Est, Laboratoire d'informatique Gaspard-Monge, A3SI, ESIEE Paris, CNRS, France

ABSTRACT

Text extraction from complex colored images involves the suppression of unwanted background while keeping text features. Imaging devices are almost omnipresent and the unrestricted conditions of the images present new challenges for real-time OCR systems. The recently proposed Gamma Correction Method [1] is a robust and good quality method for text extraction in complex colored images. However it requires a large amount of computing resources and is not well suited for real-time applications. In this paper we propose an efficient acceleration of the GCM to drastically reduce its execution-time, while preserving the text extraction quality. Experimental results on ICDAR dataset show that our approach is effective and can reach a speedup of up to 11.430.

***Index Terms*—** OCR, Scene Text, Complex Images, Text Extraction, Gamma Correction, Real-time

1. INTRODUCTION

Optical character recognition (OCR) consists of several steps to extract information from text contained in digital images. The OCR process can be divided in segmentation, binarization and recognition. The segmentation procedure isolates text features in the image, the binarization outputs a binary image of the detected segment and the recognition step recognizes each individual character present in the binary image. The domain of application for OCR is wide and image sources abundant. The ubiquitous presence of devices capable of producing digital images demands quicker and more efficient text extraction algorithms for applications like traffic sign interpretation, automatic object inspection, detection of vehicle license plates, document retrieving, hand written text interpretation, video indexing among others. Examples of sources of text extraction images are digital reproductions of books, maps, documents, manuscripts, newspapers and document collections.

Scenes captured by imaging devices are diverse in characterization and can be grouped in three main categories: Document text images, caption text images and scene text images. Text extraction in these cases are widely studied and rely generally on edge detection, morphological filters, Wavelet transform, thresholding, DCT based high pass filter among others [2].

However in scene text images, where text appears within the imaged environment, the extraction of text becomes more challenging. This is due to the unrestricted conditions encountered in colored scenes and text presented in multiple sizes, fonts, distortions, partial occlusions within complex backgrounds. Existing methods related to this problems are based on applying adaptive color reduction (ACR), Principal Component Analyzer (PCA) and Self-Organized Feature Map (SOFM) [3]. Zhan et al. uses multiscale wavelet features and SVM classifier [4]. Other methods propose Globally Matched Wavelet Filters (GWM) with fisher classifiers [5] and a method insensitive to orientation, size and noise [6]. Zhang et al. proposed a method based on link energies [7].

Among the most recent text extraction methods for scene text images Sumathi et al. proposes the Gamma correction Method (GCM) [1]. This method consists of a procedure to partially or completely suppress the heterogeneous background whilst preserving text features. The quality of the results of the GCM proves that it is well suited for demanding applications. On the other hand, the computing resources required by the GCM are considerable and its integration to a real-time OCR system presents difficulties. In this paper we propose an acceleration to the GCM that reduces the execution time and needed resources but conserves the quality of the results.

In the following, section 2 deals with the GCM, then we describe the proposed acceleration process in section 3. In section 4 we present the obtained experimental results and finally we conclude in section 5.

2. TEXT EXTRACTION BASED ON GAMMA CORRECTION METHOD (GCM) [7]

The gamma transform, expressed in equation 1, is an operator that depends on the parameter γ. Where $\gamma < 1$ results in a whiter image and $\gamma > 1$ results in a darker image than the original. The original image is described by the instance of $\gamma = 1$.

$$q(i,j) = p(i,j)^{\frac{1}{\gamma}} \tag{1}$$

The GCM [1] transforms the input image to gamma-transformed images using a range of γ from 0.1 to 10.0 with increments of 0.1, resulting in 100 different gamma values. Thus a total of 100 gamma-transformed images have to be computed and for each image three features are calculated as described in below.

Two textural features derived from the Gray-Level Co-Occurrence Matrix (GLCM) are used to characterize each gamma-transformed image. Haralick [8] proposed in 1973 the GLCM matrix and a set of 14 textural features while working in the surface classification of aerial and early satellite images. The GCM only uses two textural features: Energy (equation 2), also called Angular Second Moment (ASM), which measures the smoothness of the image and Contrast (equation 3), which measures local gray level variation.

$$Energy = \frac{1}{4}\Sigma_\alpha \left(\Sigma_{i=1}^{N_g} \Sigma_{j=1}^{N_g} \left(\frac{p(i,j)}{R} \right)^2 \right) \quad (2)$$

$$Contrast = \frac{1}{4}\Sigma_\alpha \left(\Sigma_{n=0}^{N_g-1} n^2 \left[\Sigma_{|i-j|=n} \left(\frac{p(i,j)}{R} \right) \right] \right) \quad (3)$$

Where $p(i,j)$ is the value of the reference pixel, α is an angle value and R is a normalization term. The number of possible Gray-level values N_g in the image is usually 256. The texture features should be equal for an image A and a rotated version of A called B. In order to comply with this requirement, texture features are interpreted using the average of the features computed from four possible GLCM α angles belonging to the set $\{0, 45, 90, 135\}$.

The threshold value is calculated using Otsu's Threshold [9]. The algorithm aims to separate the foreground and background pixels by minimizing the intra-class variance σ_i separated by a threshold value t defined in equation 4.

$$\sigma_w^2(t) = w_1(t)\sigma_1^2(t) + w_2(t)\sigma_2^2(t) \quad (4)$$

This value is computed directly from the gamma-modified image. Table 1 shows an example of the computed values. A set of 3 rules estimates which gamma-transformed image will suppress the largest amount of background while keeping the text features. The background suppression and binarization is done by thresholding the selected image using its corresponding Otsu's Threshold value. All gamma values are considered to have equal probability of being selected by the GCM algorithm as the best gamma-transformed image to threshold. The selected image can be identified by its unique gamma value, this value is called selected gamma value, SGV. figure 1 illustrates the GCM result from a sample image that belongs to the ICDAR data set [10]. However, the resources needed to compute the 100 gamma transformed images and the Co-occurrence Matrix features are very large and the overall execution time to select the SGV is excessively high. For example the execution time of GCM on the input image in figure 1*(a)* is 2.13019 seconds, which is not well suited for a real-time system but has an effective result, shown in figure 1*(c)*.

Gamma	Energy	Contrast	Otsu's Threshold
0.1	0.005508	20.2623	0.902344
0.2	0.002021	67.5451	0.820313
0.3	0.001264	127.710	0.750000
...			
9.8	0.463231	361.558	0.332031
9.9	0.474122	359.967	0.328125
10.0	0.485401	358.094	0.339844

Table 1. First and last three values of features from the gamma transformed images computed by the GCM from the input image presented in figure 1*(a)*.

(a) (b) (c)

Fig. 1. GCM Results: (a) Original image, (b) Gamma Transformed Image with SGV = 5.5, (c) Output image [1].

In the following section we propose an acceleration to the GCM that will allow to keep the quality of the results and dramatically reduce the computation time.

3. ACCELERATED GCM (AGCM)

We propose an acceleration of the previously described Gamma Correction Method, that we will call in the following AGCM. The AGCM uses the same input and evaluation parameters used by the GCM. Instead of analyzing the parameters of the whole set as individual values with the same probability of being the selected gamma value, the accelerated method predicts a sub-range of values where the selected gamma is expected to be found by the GCM. The AGCM uses a set rules derived from a behavioral analysis of parameter values. This approach was inspired from the fact that similar images will lead necessarily to similar evaluation parameters and eventually to the same SGV.

3.1. Behavioral study of the parameters used in the GCM

A behavioral study of the energy, contrast and Otsus threshold curves and the selected gamma value was conducted and allowed the identification of important facts which are the basis to build the relations between parameter behaviors and sub-ranges of the gamma values.

Figure 2 shows plots of the entire set of contrast, energy and Otsu's threshold values computed by the GCM. Notice

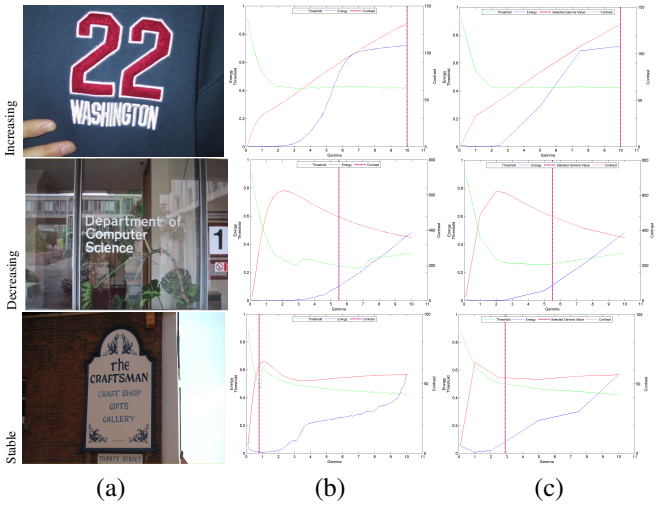


Fig. 2. Image classifications and the corresponding feature curves: (a) Original images, (b) Continuous feature curves, (c) Sampled feature curves. Contrast curve (Red), Energy (blue), Otsus's threshold (green) and SGV (vertical magenta line).

that the energy and threshold curves are plotted according to the left *y-axis* scale and the contrast curve is plotted according to the right *y-axis* scale. The Continuous curves in figure 2 show the plot with the 100 values computed by the GCM.

Depending on the shape of the contrast curve each image was classified as either increasing, decreasing or stable. The presence of peaks and abrupt changes in the curve were also taken in account. Abrupt changes are called discontinuities since they represent a dramatic change in the visual perception of the resulting image. These curves show in a clear way the behavior of the features and allow us to extract relations between the curve and the SGV, not by numerical criteria but rather by the shape of the curve. These relations are presented in the following sub-section.

3.2. Derived behavioral rules

As mentioned above, the contrast curve is the main factor (first order feature) of influence on the gamma value region. Other factors (second order features) that must be evaluated in a second stage are high energy values and Otsu's Threshold peaks. We present a description of the main behaviors and their relation to the SGV.

The contrast curve has three main behaviors: Always increasing, decreasing or stable. An always increasing curve (see figure 2*(a)*) has an optimal gamma value of 10.0 and are generally images without many variations or features. A decreasing curve has, in general, the optimal gamma value near its peak point. The peak of a decreasing curve is always located in the range of γ = [0.9 - 3.5]. (see figure2*(b)*). A stable curve (see figure 2*(c)*) denotes an image that is more influenced by second order features than by the contrast curve behavior.

The energy curve is always increasing. Energy is usually low at low-gamma values and has a tendency to increase at the end of the curve. Initial and final energy values that do not comply with the previous statement influence in the estimated gamma value. A flat curve is the result of a text image or alike.

The threshold curve is always decreasing since the increase in gamma renders the image darker (lower pixel values). A peak in the curve is direct evidence of the transition of a significant sized block from bright to dark values.

The relations of curve behaviors and SGV can be used to predict a sub-range of gamma values that contain the SGV. In addition, a sampled curve can produce a simplified, but reliable, version of the feature curves while avoiding the calculation of values that are not candidates for the SGV. In the following we explain the computation and use of this sampled curve.

3.3. Curve Sampling

We propose to compute the parameter values (contrast, energy and Otsu's threshold) at particular sampling points. The sampled gamma value points are: γ = [0.1, 1.0, 2.0, 2.5, 5.0, 7.5, 10.0]. These values were chosen from the regions that contain the gamma values that define the behavior of the parameter curves so that the entire curve can be reliably classified.

The sampled curves can be compared in figure 2 *Sampled curves* against the *Continuous curves*. We can see that the extracted feature behavior is completely preserved.

3.4. Gamma Sub-Range Prediction

The relations provided by the behavioral study can be applied in order to predict the gamma value sub-range. These relations are formalized in the flow chart presented in figure 3. The flow chart has 6 possible branches numbered from Return 1 to Return 6. These returns can be grouped in 3 main branches. the contrast curve is always increasing: Return 1. The contrast curve is always decreasing: Return 2 and 3. The contrast curve is stable: Return 4, 5 and 6.

By examining the results from the GCM we can notice that certain type of images always get a particular selected gamma value. For example images that present a smooth background or an image that is mostly dark with a small light feature will always lead to a γ_{sel} = 10.0, while images with a clear text feature and a heterogeneous background usually have a selected gamma value that belongs to the middle section of all the considered values but its never located in the upper or lower bounds of the range.

In particular cases the estimated gamma value by the AGCM may be different from the GCM computed gamma (See figure 2, stable case). The experiments presented in the following section prove that this difference between the two SGV allow to improve the quality of the obtained results.

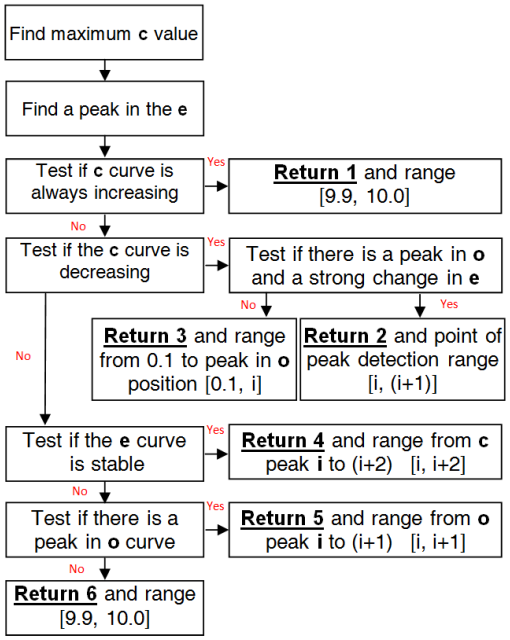


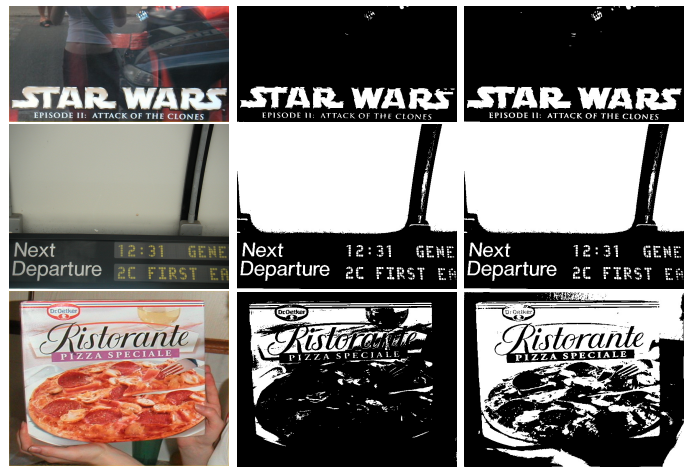
Fig. 3. Flow chart to predict the gamma values sub-range. Where **c** = contrast, **e** = energy, **o** = Otsu's Threshold.

4. RESULTS

For evaluation purposes, the ICDAR Reading Text in Scene Images dataset [10] was used as input to benchmark the AGCM. The set contains 113 complex color images with text. The GCM and AGCM were coded using C++ and executed under Windows 7 running in an Intel® i3 CPU at 3.07 GHz.

The gamma sub-range prediction of the AGCM dramatically improves the execution-time of the method while preserving the quality of the GCM results. As expected, the AGCM output contains the same quality of the GCM results, making it robust and reliable (figure 4, first and second images). The behavioral rules of the sub-range prediction algorithm can sometimes lead to a better binarization than the GCM. An example of this case is presented in the third image of figure 4. This is mainly due to the analysis of peaks in the energy and threshold curves.

Figure 5 presents the superposed execution-times of the GCM and AGCM over the ICDAR image set [10]. The execution-time of the AGCM can not be generalized in a fraction of the GCM's execution-time because the gamma sub-range prediction size is variable from 0.1 (Return 1 and 6) to 2.0 (Return 4). This makes that the number of gamma-modified images that have to be calculated and features analyzed is variable, while the GCM always analyses 100 images. In spite of this variability, the execution-time of AGCM presents an average speedup $s = T_{GCM}/T_{AGCM}$ of 4.456 and in best cases a speedup of 11.430 (from the dataset execution-times presented in figure 5). These results are significant and encouraging for the integration of the AGCM method to a real-time system.



(a) Original (b) GCM (c) AGCM

Fig. 4. Comparison of the GCM and AGCM results.

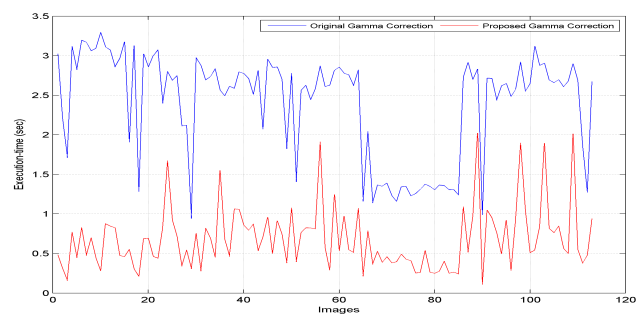


Fig. 5. Comparison of the GCM (blue) and the AGCM (red) execution-times for a set of 113 images.

5. CONCLUSION

This paper has introduced the GCM as a robust method for the segmentation and binarization of text features in complex color images. We presented a behavioral study of the curves generated by the analyzed textural features and an effective sample method that retains the behavior of the curves. Together they reduce dramatically the execution-time of the entire process while keeping the good results of the original method. This improvement renders the AGCM more suitable for its integration in real-time applications. Further improvement of the execution time can be done by a parallelized version of the algorithm.

6. REFERENCES

[1] C.P. Sumathi and G. Gayathri Devi, "Automatic text extraction from complex colored images using gamma correction method," *Journal of Computer Science*, vol. 4, pp. 705–715, October 2014.

[2] C.P. Sumathi, T. Santhanam, and G. Gayathri Devi, "Survey on various apporoaches of text extraction in images," *Internationa Journal of Computer Science & Engineering Survey*, vol. 3, pp. 803–806, August 2012.

[3] C. Strouthopoulos, N. Papamarkos, and A.E. Atsalakis, "Text extraction in complex color documents," *Pattern Recognition*, vol. 35, pp. 1743–1758, August 2002.

[4] Y. Zhan, W. Wang, and W. Gao, "A robust split-and-merge text segmentation approac for images.," *Proceedings of th 18th Internation al conference on Pattern Recongnition, (ICPR 06), IEEE Xplore Press*, vol. DOI: 10.1109/ICPR.2006.169, pp. 1002–1005, 2006.

[5] S. Kumar, r. Gupta, N. Khanna, S. Chaudhury, and S.D. Joshi, "Text extraction and document image segmentation usign matched wavelets and mrf model," *IEEE Trans. Image Process*, vol. DOI:10.1109/TIP.2007.900098, pp. 16:2117–2128, 2007.

[6] G. Babu, R. Mohan, P. Srimaiyee, and A Srikrishna, "Text extraction from heterogeneous images using mathematical morphology," *J. Theoretical Applied Inform. Technol.*, vol. 16, pp. 39–47, 2010.

[7] J. Zhang and R. Kasturi, "Text detection using edge gradient and graph spectrum," *Proceedings of the 20th International Conference on Pattern Recognition (ICPR), Aug 23-26, IEEE Xplore Press, Istambul*, vol. DOI:101109/ICPR.2010.968, pp. 3979–3982, 2010.

[8] R.M. Haralick, K. Shanmugam, and T. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-3, pp. 610– 621, November 1973.

[9] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-9, pp. 62–66, January 1979.

[10] ICDAR Dataset, "Text in scene images," *http://robustreading.opendfki.de/wiki/SceneText*.