TEXT SEGMENTATION FROM WEB IMAGES USING TWO-LEVEL VARIANCE MAPS

Insook Jung and Il-Seok Oh
Department of Computer and Information science, Chonbuk National University, Jeonju, Korea

Keywords: Text segmentation, Web images, Two-level Variance maps, Text location.

Abstract: Variance map can be used to detect and distinguish texts from the background in images. However previous variance maps work as one level and they revealed a limitation in dealing with diverse size, slant, orientation, translation and color of texts. In particular, they have difficulties in locating texts of large size or texts with severe color gradation due to specific value in mask sizes. We present a method of robustly segmenting text regions in complex web color images using two-level variance maps. The two-level variance maps works hierarchically. The first level finds the approximate locations of text regions using global horizontal and vertical color variances with the specific mask sizes. Then the second level segments each text region using intensity variation with a local new mask size, in which a local new mask size is determined adaptively. By the second process, backgrounds tend to disappear in each region and segmentation can be accurate. Highly promising experimental results have been obtained using the our method in 400 web images.

1 INTRODUCTION

While existing search engines index a page on the text that is readily extracted from its HTML encoding, an increasing amount of the information on the Web is embedded in images. In extreme cases, all of the text on a page might be present solely in image format. Existing Web commercial search engines are limited to indexing the raw ASCII text they find in the HTML-they cannot recover image text. This situation presents new and exciting challenges for the fields of Web document analysis and text information retrieval, as WWW image text is typically rendered in color and at very low spatial resolutions.

Considering traditional Optical Character Recognition (OCR), one may initially think that Web images containing texts present some advantages over scanned documents, such as the lack of digitisation-induced noise and skew. However, the task is considerably difficult for traditional OCR for a number of reasons. First, these (often complex) Web images tend to be of low resolution (just good enough for display and usually 72 dpi) and the font size used for text is very small (about 5~7pt) or very large (about 50~70pt). Such conditions clearly pose a challenge to traditional OCR, which works with 300dpi images and character sizes of usually 10pt. Moreover, Web images tend to have various artefacts (anti-aliasing, colour quantization and lossy compression), colour schemes (multi colour text over multi colour background) and character effects (characters not always on a straight on a line, 3D-effects, shadows, outlines, etc.).

The goal of locating the text in image form can be split into two objectives.

Text segmentation. The image must be segmented first so that regions corresponding to potential character components are separated from the background. A successful segmentation will be one where background and foreground regions are not merged.

Text extraction. It classifies the segmented regions as text/non-text and then character-like components that fulfil criteria of constituting text (e.g., they appear to form a text line) are extracted.

In view of the difficulties posed by the image and text characteristics, it can be appreciated that the segmentation stage is by far the most challenging. The performance at that stage affects quite crucially the degree of success in a subsequent recognition. This paper presents a new approach to segment text;

Jung I. and Oh I. (2009).
TEXT SEGMENTATION FROM WEB IMAGES USING TWO-LEVEL VARIANCE MAPS .
In *Proceedings of the Fourth International Conference on Computer Vision Theory and Applications*, pages 363-370
DOI: 10.5220/0001808303630370

especially in complex Web images (e.g., see Figure 1).

There are two primary methods to segment texts in images: colour representation-based (or region-based) methods and texture-based methods (Jung and Kim, 2004), (Jung *et al.*, 2004).

It argues that the colour representation (commonly used by previous approaches) is not suited to this particular task for Web images. Their method for text segmentation and extraction is based on clustering in the colour space. However, they are not appropriate for low-resolution and various character effects. They depend on the effectiveness of the segmentation method, which should guarantee that a character is segmented into a few connected components (CC) separated from other objects and the background. These methods produce good results for relatively simple images, but fail when more complex images are encountered for the following reasons. These approaches mostly deal with a very small number of colors (they do not work on full-color – e.g., JPEG – images). They also assume a practically constant and uniform color for text (Zhou and Lopresti, 1997), (Antonacpoulos and Delporte, 1999) and fail when this is not the case. In practice, there are many situations where gradient or multi colour text is present (see Figure 1).

The situation where dithered colors are present (especially in GIF images) has received some attention (Zhou and Lopresti, 1998), (Lopresti and Zhou, 2000) but such colours can only be found in a relative small number of Web images. Furthermore, the background may also be complex (in terms of colour) so that the assumption that it is the largest area of (almost) uniform colour in the image (Jain and Yu, 1998) does not necessarily hold.

Unlike the above color representation, the texure based methods employ distinct textual properties of texts compared with their backgrounds. In these methods, the textual properties of an image are often detected by using techniques of Gabor filters, wavelets, spatial variance, etc.

2 OUR APPROACH

2.1 Conventional Variance Map

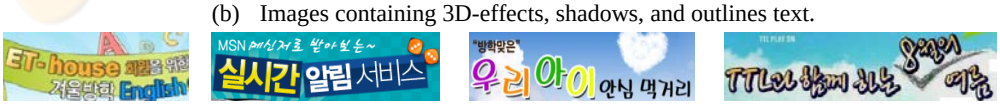
If a document or an image containing texts is viewed at a certain distance far from a person, the person sees a blurred image of the document, but is still able to detect the different blocks of the document. Detection is possible since each block has a specific texture pattern. These patterns correspond to regions of text, regions of graphics and regions of pictures. Thus document image can be segmented into regions of text, and regions of graphics and/or pictures using the texture of low resolution images.

The assumption of the variation map is that each part of a document image has a different texture. Text regions have different textures from that of graphics and pictures. Graphic regions have different textures from that of text and pictures, and the same may be applied to regions with pictures. Under this consideration document segmentation may be considered as a texture segmentation problem.

One of the variations of this method compared to others is that it maps the variance around the pixel to just one value. This value represents the variance of the texture variance around the pixel of interest. Also, in the computation of the variances, a mean value is used instead of the actual value of the pixel at the center of the mask. The texture at a pixel is defined as the average value of the variance of the neighbors of the pixel in four or 8 different directions, vertical, horizontal, left diagonal, and right diagonal. Thus the document image is mapped into the texture image.

(a) Images with multi colour text over multi colour background.

(b) Images containing 3D-effects, shadows, and outlines text.



(c) Images containing text not always on a straight line.

Figure 1: Texts in Web images.

Despite its simplicity, variance maps have shown to be effective and robust. (Mario and Chucon, 1998) proved that a document image can be segmented into regions of text, and regions of graphics and/or pictures using gray-level spatial variation of low resolution images. It was designed to work with free format documents, text in background other than white, skew greater than 10 degrees. Besides it requires less computation than the segmentation methods using the other textures described in other papers.

(Karatzas and Antonacoppoulos, 2006) abandons analysis by the RGB colour clustering and adopts a segmentation method based on analysing differences in color and lightness that is closer to how humans perceive distinct objects. However it only present a method to the topical problem of segmenting characters in colour Web images containing text (headers, titles, banners etc.) and fail when image size is not small or character size is large. (Song *et al*, 2005) proposed an extraction method to detect text regions from the images using the intensity variation and color variance but the method has difficulties in locating large size texts or texts with severe illuminations changes.

In such a global approach, the mask of variation maps was applied to a single value for the entire image. Global mask has a good performance in the case that there is uniform size of texts and width of character is less than width of mask width. However, very often, Web images are exposed to other cases.

Although very effective in text localization, the variation maps have some shortcomings:

(1) difficulties in designing a locality of the mask information (type, value) to satisfactory the wide variations of text size.

(2) cannot ensure accurate location of texts. For example, the resulting variation map suffers from a great amount of background in the case that there is non-uniform size of texts and/or not on the straight line.

2.2 Proposed Method

In this paper, variance maps are used to detect and distinguish texts from the background in web images. We propose a method to deal with local information for masks in variance maps using two levels. Local information may guide the adaptive mask size for the local text region only.

The two-level variance maps works hierarchically. The first level variance finds the approximate locations of text regions using horizontal and vertical color variances with the specific mask sizes to ensure extraction of large size texts as well as small size ones. At the first level, we can increase the recall rate using color variance map with the specific mask size to approximately segment text like region and apply CC analysis (CCA). Then it segments text components in these regions using local thresholds, in each of which a new mask size is determined adaptively. As a second level, the automatic and non heuristic gray variance map using the new mask size is applied to each region. By the second process, backgrounds tend to disappear in each region and segmentation can be accurate. In second level, we can also increase the precision rate using intensity variation map with adaptive mask size to find accurate text like regions. This technique has been widely used an image analysis because it has a better performance in segmenting the objects from an image that contains spatially uneven texture features.

Our method is to improve the variance map to overcome shortcomings in previous variance maps and then to accurately locate complicated cases such as multicolor and/or various sizes of texts and texts arranged on not a straight line.

Figure 2 shows an overview of our extraction method.

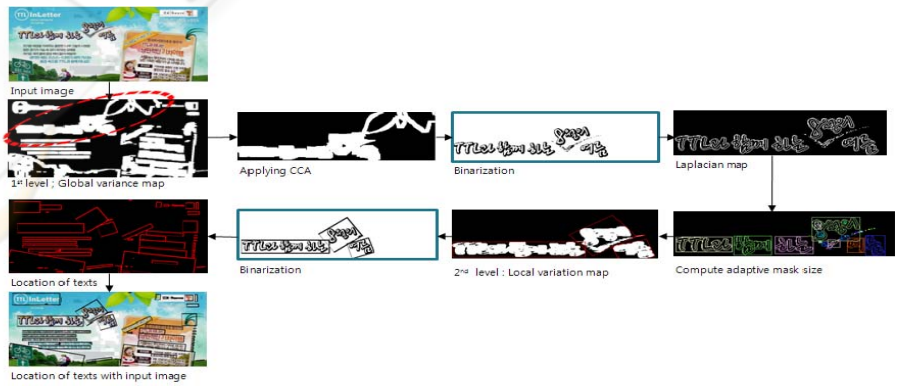


Figure 2: Overview of two-level variance maps.

First, the global variation of the input image is computed to segment text regions in the input image. This is because readable text usually appears with sufficient differences with the background. Applying CCA, every CC of the variance image is binarized using a local threshold. Each CC is processed seperately afterwards. This image is passed through a Laplacian filter which is useful to extract variances which occur in texts of CC region. Next, for every CC, determine the new mask value based on the laplacian map and then apply local variance map with the new mask value to segment text region accurately. Finished when the local variation is computed and applied to all the image area (The result after this step will be an image having two level variation map).

Our approach is to segment the image into text/no-text regions as best as possible, and then let the OCR system do the detailed refinement. In other words, we would like to find all text areas and as few spurious non-text areas as possible, without actually classifying the characters.

Section 3 provides a detailed description of our method and section 4 shows the experiment results. Conclusion is given in Section 5.

3 DETAILED ALGORITHM

3.1 Parameterization of Variance Map

We find variance image by applying each pixel with masks (Jung et al, 2008). Then we transform the variance image into a binary image. The pixels that were assigned to 0 are regarded as background and excluded in subsequent process. At this processing, we have various maps that depend on mask shape, mask size, and combination between masks if multi masks are used. Let's define them in terms of parameters as follows.

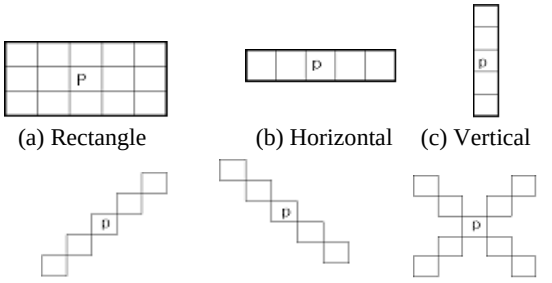
$\theta=(m,s,p)$

m=subset of {rectangle, vertical form, horizontal form, cross, diagonals from the upper-right and upper-left corner, diagonal cross,.....}

s= mask size, $(3 \leq s)$

p={and, or, ...}

The possible shapes, m, of the mask are displayed in Figure 4. They can have different sizes according to shapes. Size s must be greater than at least 3 to obtain its variances. Size s is defined as the number of pixels in the mask.



(a) Rectangle (b) Horizontal (c) Vertical

(d) Right-upper (e) Left-upper (f) Diagonal cross

Figure 3: Mask shape *m*.

|m| is the number of masks used. It is necessary to combine |m| maps into one when |m| is larger than 1. We need combination operations are required. Usually bitwise 'and' or bitwise 'or' is used. In this paper, we use the following parameters.

m= {rectangle, vertical, horizontal}
s=3x21 rectangle, 17x3 rectangle,
1x21 horizontal, 17x1 vertical
p={ and }

In the RGB colour space, the variance at each pixel (x,y) in applying the horizontal mask is defined as follows. The variances by the other masks are defined in the similar way.

Horizontal mask operation:

$\sigma_R^h(x,y)$ = Variance of a pixel (x,y) in applying the horizontal mask to the pixel in the R plane (with 21 neighbour pixels)

$\sigma_G^h(x,y)$ = Variance of a pixel (x,y) in applying the horizontal mask to the pixel in the G plane (with 21 neighbour pixels)

$\sigma_B^h(x,y)$ = Variance of a pixel (x,y) in applying the horizontal mask to the pixel in the B plane (with 21 neighbour pixels)

$$\sigma^h(x,y)=1/3*(\sigma_R^h(x,y)+\sigma_G^h(x,y)+\sigma_B^h(x,y)) \quad (1)$$

Figure 4 shows the horizontal variance image $\sigma^h(x,y)$ and vertical variance image $\sigma^v(x,y)$ that are obtained in this way.

(a) (b) (c)

Figure 4: Binary horizontal and vertical variance images.(a) – Input image; (b) – Horizontal variance image; (c) – Vertical variance image.

Next we transform the two variance images to a variance image T(x,y) using a threshold value t and a combination operation 'and'.

$$T(x,y)=1,\ \sigma^{h}(x,y)>t \text{ and } \sigma^{v}(x,y)>t$$
$$0, \text{ otherwise}$$

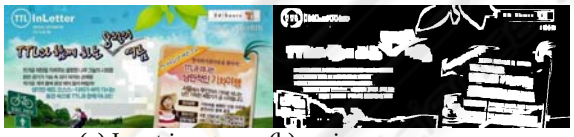
Figure 5 displays the variance image obtained from the two variance maps.

Figure 5: Variance map.

In applying colour variances, there can be many varieties according to the parameters used.

3.2 First Level: Global Variance Step

In our method, we first compute the global horizontal and vertical variations using the equation (1) with 3x21 horizontal and 19x3 vertical masks. The mask sizes are determined by experiments of previous studies. At this step, we obtain a rough estimation of text regions. Our intention is to proceed to an initial segmentation of text (foreground) and non-text (background) regions that will provide us a superset of the correct set of foreground pixels. This is refined at a later step, so-called local variance map. Figure 6 shows the image of the variance map.



(a) Input image (b) variance map

Figure 6: Global variance map.

We see that the text areas are high in variances (white pixels in Figure 6 (b)).

To remove further the noisy regions, we also apply morphological operations. The operations have structuring elements with sizes 2x5 for dilation and 3x3 for erosion to emphasize more on horizontal texts. Morphological closing, opening and opening are applied. Figure 7(a) shows the results of the morphological operations, and Figure 7(b) shows the connected components (CC) which are randomly colored.

Figure 7: (a) Morphological operations applied (left), and (b) CC (right).

3.2.1 CC Analysis

As shown in Figure 7 (b), every CC usually contains texts of the image, however, in a sub-optimal fashion: some CC spans more than one line and/or column of text, others contain no text, while in many the background makes up a large portion of the pixels. Fortunately, these shortcomings can overcome by the next step, local variation map.

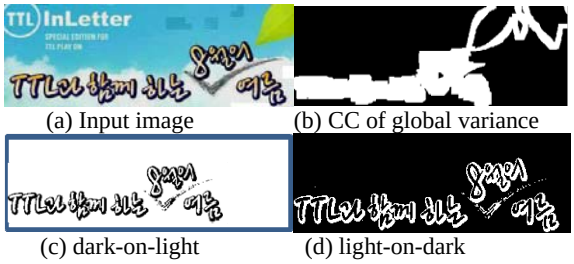
3.3 Second level: Local Variance Step

3.3.1 Binarization

We generate a bounding box around each CC region on the variance map in the first level. Once the bounding box of each CC is obtained, each region in the bounding box is binarized using Otsu thresholding. This step produces binary text regions to be used as inputs to the local variance map (It is possible to run local color image directly but we obtain substantially worse performance if we do so).

Every CC region provided by the variance map is expected to contain only text, implying the background and foreground should be easily separable through thresholding.

To ensure the correct labelling of both the dark-on-light and light-on-dark text, the proportion of pixels which fall above and below the thresholds is considered. Since in a block of text there is always a larger area of background than text elements, the group of pixels with the lower proportion is labelled as text, and the other group as background. The example is shown in Figure 8.



(a) Input image (b) CC of global variance

(c) dark-on-light (d) light-on-dark

Figure 8: Local binary image.

3.3.2 Laplacian Map and CC Analysis

In our approach, the local binary image is passed through a Laplacian filter which is useful in applying the local variation map. The Laplacian edge detector produces closed edge contours because edge strength is not considered, so even the slightest, most gradual intensity transition produces differences of pixel values. Next, CCA is generated based on the Laplacian image.

CCA detects character candidates and enables us to estimate their size and the spacing between them. These estimates will be used to apply *local variance map with adaptive mask* for texts into regions directly.

We show results for the Laplacian map and CCs in figure 9 for the text region shown in Figure 8.

Figure 9: Laplacian map and CCs.

3.3.3 Estimation of Adaptive Mask

In applying the local variance map, the key-problem is how to determine the mask type and size.

Determination of a adaptive mask value is very important and perhaps the most sensitive part of any image segment scheme of variation maps because a wrong value of mask may result in being dropped some texts information (an object can be considered as part of background and vice versa).

The well-known local adaptive method uses mean and standard deviation to compute mask value over a CC text region (e.g., bounding box of every CC in Figure 9).

We use the mean and the standard deviation along with widths of all bounding boxes in every CC to compute another mask value for the CC region. In other words, we compute the adaptive mask value to apply the local variance map, which uses adaptive contribution of mean and standard deviation in determining local mask value.

3.3.4 Local Variance Map

Local variance map is an adaptive one in which a mask value is determined over a CC region. Local method performs better in case of nonlinearly aligned texts and wide variation of text sizes. The segmentation quality is dependent on the above adaptive mask value.

We apply intensity variance map to the CC region presented by the Laplacian filter with adaptive mask value. Our method compute the local horizontal variation only using equation (1) with the mean and the standard deviation as a mask value for every CC region.

We show results for the **local variance map** and red bounding boxes of CCs in Figure 10.

Figure 10: Local variation map and CCs.

4 EXPERIMENTAL RESULTS

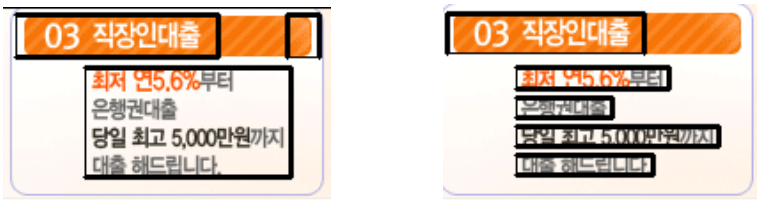
The proposed method was tested using 400 Web images which were selected randomly from WWW. Some of them contain extremely large or small text, non-homogeneously-colored text and/or non-horizontal layouts. No assumptions are made about the size of input images. All images had almost low resolution and the sizes of characters varied from 4pt to 72pt.

Based on visual criteria, the proposed method outperforms the global variation map. False alarms are currently ignored. We have shown a few of them in this paper. We have also segmented the same images using (Song *et al.*, 2005) method in order to provide comparison. Each experiment is performed both of (Song *et al.*, 2005) method and our method. The (Song *et al.*, 2005) method used the different mask sizes (3x21 horizontal and 19x3 vertical masks).

Figure 11 illustrates the results of segmenting Web images using (Song *et al.*, 2005) method (left) and our method (right) and the improvement can be seen at right sides of the resulting images.

Our method has removed the background as much as possible while not disturbing any text area. This is an improvement from the (Song *et al.*, 2005) method. By looking at results of the second level variance step, we have observed that by applying adaptive mask sizes, we get the unnecessary pixels eliminated from the image background in a better way while preserving characters and vice versa. In this way, we have found adaptive local mask values to be appropriate for this kind of Web images.

In the experiments, we found that fragmentation often appears at those text lines that are isolated in both horizontal and vertical orientations. Because headlines vary in size greatly, some headline components are erroneously segmented into body of text components. The fragmentation rate of headline regions is higher than that of body text regions.



(a) Image 1.

(b) Image 2.

(c) Image 3.

(d) Image 4.

(e) Image 5.

Figure 11: The images of different segmentation results, where the segmented results are contained in the rectangles: (left) for the (Song and Kim, 2005) method; (right) for our proposed method.

5 CONCLUSIONS

In this paper, we propose a local adaptive approach of variation maps to segment texts in Web images. The proposed method is less sensitive to parameters by user and can deal with segmentations where shadows, non-uniform character sizes, low resolution and skew occur. After the local approach, our method demonstrated superior performance on Web images using visual criteria.

Also, our method has the additional advantage that it can be applied directly to the line segment without requiring de-skew algorithm.

Further research will be focus on developing the text or non-text classifier and character segmentations.

ACKNOWLEDGEMENTS

This research was financially supported by the Ministry of Education, Science Technology (MEST) and Korea Industrial Technology Foundation (KOTEF) through the Human Resource Training Project for Regional Innovation.

REFERENCES

Jung, I. S., Ham, D. S., and Oh, I. S., 2008. Empirical Evaluation of Color Variance Method for Text Retrieval from Web Image**s,** *In Proceeding of the 19th Workshop on Image Processing and Image Understanding (IPIU'08).*

Zhou. J., and Lopresti, D., 1997, Extracting Text from WWW Images, *Proceedings of the 4th International Conference on Document Analysis and Recognition (ICDAR'97)*, Ulm, Germany, August.

Antonacopoulos, A., and Delporte, F., 1999, Automated Interpretation of Visual Representations: Extracting textual Information from WWW Images, *Visual Representations and Interpretations*, R. Paton and I. Neilson (eds.), Springer, London.

Zhou. J., Lopresti, D., and Tasdizen, T., 1998, Finding Text in Color Images, *proceedings of the IS&T/SPIE Symposium on Electronic Imaging*, San Jose, California, pp. 130-140.

Lopresti, A. D., and Zhou, J., 2000, Locating and Recognizing Text in WWW Images, *Information Retrieval*, vol. 2, pp. 177-206.

Jain, A. K., and Yu, B., 1998, Automatic Text Location in Images and Video Frames, *Pattern Recognition*, vol. 31, no. 12, pp. 2055–2076.

Karatzas. D., and Antonacoppoulos, A., 2006, *Colour* Text Segmentation in Web Images Based on Human Perception, *Image and Vision computing* , 2006.

Song, Y. J., Kim, K. C., Choi, Y. W., Byun, H. R., Kim, S. H., Chi, S. Y., Jang, D. K., and Chung, Y. K., 2005, Text Region Extraction and Text Segmentation on Camera Captured Document Style Images, *Proceedings of the Eight International Conference on Document Analysis and Recognition (ICDAR'05).*

Mario, I., and Chucon, M., 1998, *Document segmentation using texture variance and low resolution images*, Image Analysis and Interpretation, IEEE Southwest Symposium on.

Jung, K. C. and Han, J. H., 2004, Hybrid approach to efficient text extraction in complex color images, *Pattern Recognition Letters*, vol. 25, pp. 679–699.

Jung, K. C., Kim, K. I., and Jain, A. K., 2004, Text Information Extraction in Images and Video: A Survey, *Pattern Recognition*, vol. 37, no. 5, pp. 977–997.