

Detecting Gradients in Text Images Using the Hough Transform

Dimosthenis Karatzas

Computer Vision Centre, UAB, Barcelona, Spain

dimos@cvc.uab.es

Abstract

The use of gradients in text images is nowadays quite frequent. Existing segmentation methods encounter serious problems when it comes to modern text images where gradients might appear in the background or the foreground or both at the same time. This paper presents an approach for lightness gradient areas detection based on the Hough Transform. The issues arising are discussed, and results are presented on a dataset comprising Web images, logos and scanned documents.

1. Introduction

The use of colour gradients either as the image background or to emphasise the textual content is nowadays quite frequent. This is especially true for images on Web pages (Figure 1a) or company logos (Figure 1b) where the primary purpose of the designer is to capture the attention of the reader with his design. Previous experience of the author in the domain of Web images has identified the extensive use of gradients as a prominent problem in this class of text images [1]. Gradient areas are also habitually used on paper documents, where, in addition to emphasising and beautifying, they often serve to define semantic entities and structural aspects of the document (e.g. the columns of the form in Figure 1c).

Nevertheless, the detection and segmentation of gradient areas in images has not received any specific attention. This can be partly attributed to the fact that gradients have not been a problem historically due to their sparse use in the past. The problem has traditionally been tackled indirectly by using segmentation techniques based on edge detection [2, 3] or adaptive thresholding [4, 5], which, depending on the context, can separate reasonably well foreground text from a smooth gradient background. Such methods have varying degrees of success, and encounter serious problems when it comes to modern

text images such as the ones displayed below, where gradients might appear in the background or the foreground or both at the same time.

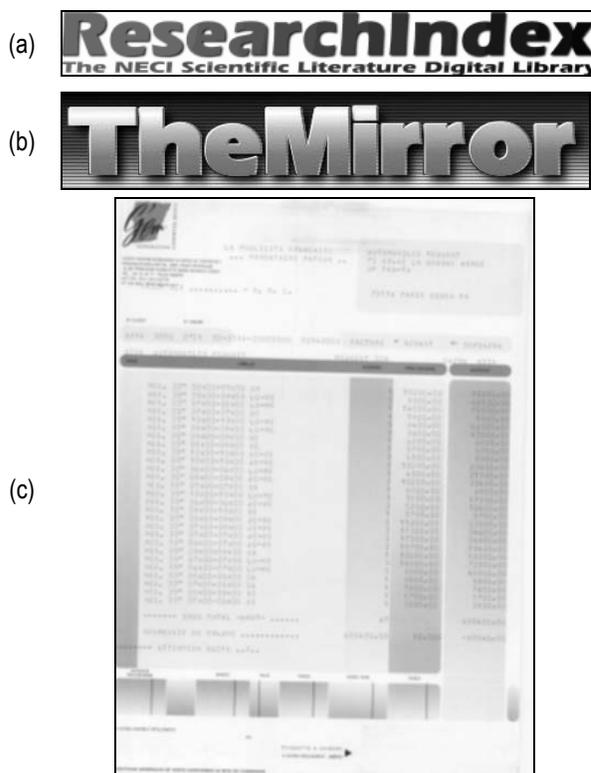


Figure 1. Sample images (a) Web image, (b) Logo, (c) Scanned document.

A purpose-made approach for gradient detection can offer clear advantages over tackling the problem indirectly. For example in many cases it is of interest to group together disjointed areas that belong to the same gradient (e.g. characters of the same word or the columns in a form).

The main idea of the Hough Transform (HT), first introduced by Paul Hough in 1962 [6], is to transform pixel coordinates into a parameter space where clusters or particular configurations identify instances of a

shape under detection. Originally conceived to detect straight lines, the HT was later extended to other parametric models [7] and was finally generalised to any parametric shape [8].

The HT is used to detect shapes in images, where a shape is described by the coordinates of its contour pixels. If an image is visualised in the 3D space comprising the pixel's coordinates (x , y) and the lightness component (L), a linear gradient will appear as a plane in this 3D space. This is illustrated in Figure 2, where an image is shown along with its 3D representation in the (x , y , L) space.

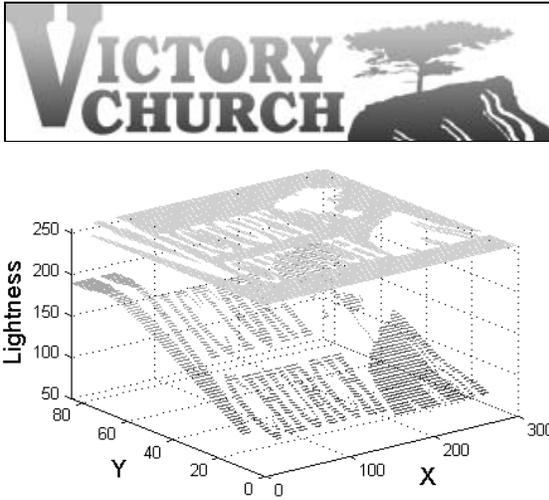


Figure 2. A sample image and its 3D representation in the (x , y , L) space.

This paper proposes an extension of the HT to detect parametric shapes in the combined spatial-lightness space. Here we will show how the HT can be extended to detect gradient planes in the (x , y , L) space and we will discuss the problems arising by mixing spatial and feature (lightness in our case) information.

The main assumption behind this approach is that lightness gradients are approximately linear changes of lightness over distance. The validity of this statement was assessed experimentally over the images of the dataset, and it was established that the vast majority of lightness gradients are indeed linear. Some of the reasons why this should be expected are presented shortly in Section 5.1.

The method described here provides a direct approach to gradient detection and deals with the problem in a global way. The main focus of this paper is linear lightness gradients. The dataset used to evaluate the method comprises Web images, logos and scanned document images.

In the next section we detail the extension of the Hough Transform to the gradient detection problem. Section 3 describes the gradient detection method used. Results are presented in Section 4, followed by a discussion in section 5. Section 6 concludes the paper.

2. Extension of the Hough Transform for Gradient Detection

A 3D plane can be described by its polar coordinates in a similar fashion as a 2D straight line by Eq 1.

$$\rho = x \cdot \cos \theta \cdot \cos \varphi + y \cdot \cos \theta \cdot \sin \varphi + z \cdot \sin \theta \quad \text{Eq. 1}$$

Where θ and φ define the orientation of the normal vector of the plane and ρ is the distance of the plane to the origin, as shown in Figure 3.

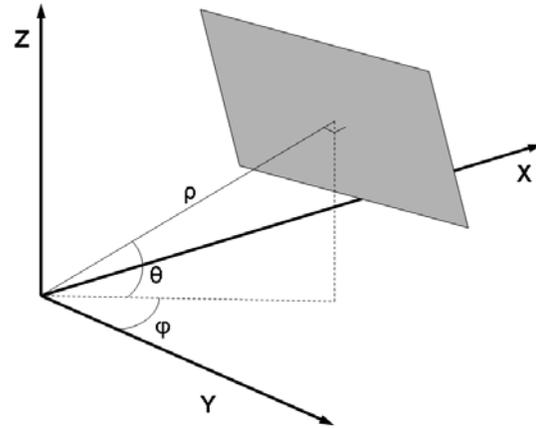


Figure 3. Polar plane parameters.

If we choose x and y to correspond to the spatial pixel coordinates in the image and z to the lightness component (L), lightness gradient areas in the image will be represented by planes in the (x , y , L) Cartesian space. Lightness is calculated here as a weighted average of the RGB components based on Eq. 2 [9].

$$L = 0.2125 \cdot R + 0.7154 \cdot G + 0.0721 \cdot B \quad \text{Eq. 2}$$

If the plane in Figure 3 represents a gradient plane in the (x , y , L) space, then the plane parameter φ , will indicate the spatial direction of the gradient, while the angle θ the rate of change of the gradient.

For any (x , y , L) triplet representing the coordinates and lightness of a pixel in the image space (by “image space” we refer to the combined spatial coordinates – lightness space), there are infinite planes (or else

infinite lightness gradients) that pass through it. The parameters of all the planes passing through a point in the image space, define a surface in the parameter space (ρ, φ, θ) . We can use Eq. 1 to transform any point to its corresponding 3D surface in the parameter space.

Any set of three points in the image space defines exactly one plane. This is manifested by the three corresponding surfaces in the parameter space having a single intersection point as illustrated in Figure 4.

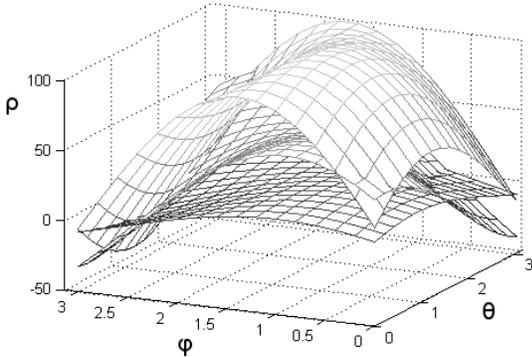


Figure 4. Three surfaces in the parameter space (corresponding to three points in the image space).

If we discretise the (ρ, φ, θ) parameter space, and apply Eq. 1 to all the pixels of the image we will obtain the “Gradient Hough Transform” of the image. A test image and its transformation is shown in Figure 5.

The lightness gradients in the image will appear as peaks (accumulator cells with a high count in the quantised space). Thus, a difficult global detection problem in the image space is now reduced to an easier problem of peak detection in the parameter space.

2.1. Modelling Gradient Areas

The definition of lightness gradient as a plane in the image space is quite generic. Depending on the value of the angle θ , a lightness gradient can describe a constant lightness area, a traditional variable-lightness gradient area, as well as an edge. This is summarised in Table 1.

Table 1. Meaning of parameter θ .

Value of θ	Rate of Change	Type of Area
$\theta=90^\circ$	zero	Constant Colour
$0^\circ < \theta < 90^\circ$	medium	Gradual Change
$\theta=0^\circ$	infinite	Edge

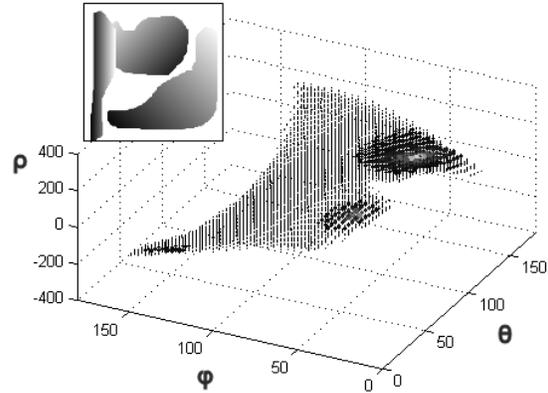


Figure 5. A test image (insert) and its Gradient Hough Transform.

By modelling constant lightness regions as a special gradient case, the method described here can deal with constant colours areas and gradients in the same step, performing this way a simple form of image segmentation. Edges though are very difficult to extract, since although they are described well by the gradient model, they comprise a very small number of pixels and therefore do not produce statistically adequate information to be identified easily in the parameter space. Since we do not wish to deal with edges here, a lower bound of 5 degrees is used for the parameter θ (corresponding to a lightness change from 0 to 255 over ~ 20 pixels). Below this threshold a plane in the image space is considered to represent an edge and is not assessed.

2.2. Issues arising from mixing spatial and lightness information

We have to appreciate at this point that the gradient plane is not really a parametric shape in the image in the traditional sense, instead it is a shape that spans the spatial (x and y coordinates) and feature (lightness in our case) space. This entails a few issues, since the parameters θ and ρ , are actually linking two distinct quantities, and their interpretation is not trivial.

In the case of θ , as mentioned before its physical meaning is linked to the rate of change of lightness on the direction of the gradient. It has to be stressed though that the interpretation of the rate of change depends on the spatial resolution of an image. An example will help clarify what this means. Suppose an image is resized to double the size of the original. The spatial dimensions of any gradient in the image will also double up in such a case, while the lightness values of the pixels will remain unaffected. In terms of any gradient in the image, that would mean that the

same change in lightness will now take place over double the distance it did in the original image, effectively producing a different angle θ for the corresponding gradient plane.

This does not create many serious problems, as the linearity of the gradient is not affected, and the method will still be able to identify it. Nevertheless, it has to be stressed that the bound set above for the parameter θ , as well as the bin size selected, specify lightness change over pixel-based distance. An easy way to overcome the problem, especially in relation to scanned documents, would be to normalise the bounds and bin size of θ based on the DPI of the image, so they are relative to lightness change over some metric distance. In practice the difference this makes is not important for most of the cases.

What is more difficult to deal with is the parameter ρ . The parameter ρ defines the distance of the plane to the origin. Most importantly the bin size we select for this parameter defines the “tolerance” for (x, y, L) triplets to be counted in a particular accumulator cell. It is difficult to assign a physical meaning to this parameter, as the meaning is actually depending on the value of θ .

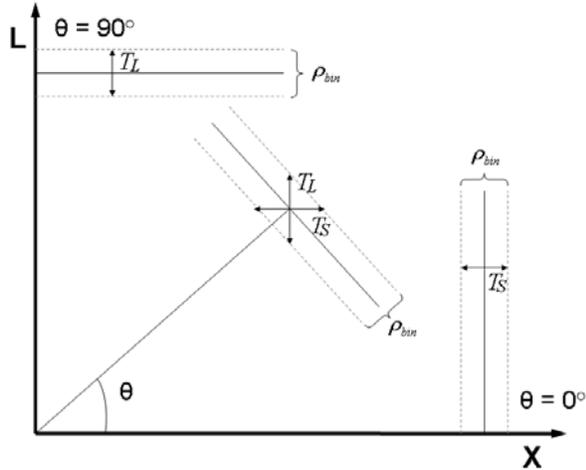


Figure 6. Effect of constant ρ bin size.

Figure 6 shows a 2D example of the above. As seen in the figure, when θ is 90° the ρ bin size refers to lightness tolerance (T_L) only. When θ is 0° , the ρ bin size refers to spatial tolerance (T_S) only. In any intermediate case, the ρ bin size refers to a combination of the two. What this means from a practical point of view is that choosing a bin size for ρ which works well for near-constant gradients will not work well for steep gradients and vice versa.

The way to deal with this problem is to make the ρ bin size relative to the angle θ . There are many

different ways to achieve this. In our case we chose to define a lightness-only tolerance and vary the size of the ρ bin size so that the lightness-only tolerance remains constant. This is displayed in Figure 7.

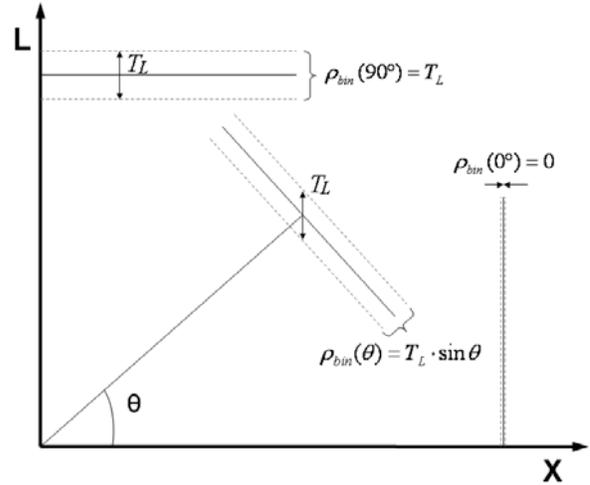


Figure 7. Effect of variable ρ bin size.

Let T_L be the lightness tolerance that we want to maintain, then the bin size will be given by Eq. 3.

$$\rho_{bin} = T_L \cdot \sin \theta \quad \text{Eq. 3}$$

The bounds and bin sizes used to quantise the parameter space are summarised in Table 2 below.

Table 2. Parameter space quantisation thresholds.

Parameter	Minimum	Maximum	Bin Size
ρ	$-\rho_{max}$	$+\rho_{max}$	$10 \cdot \sin \theta$
φ	0°	179°	4
θ	5°	175°	2

ρ_{max} is defined by Eq. 4, where *Width* and *Height* are the dimensions of the image and L_{min} and L_{max} the minimum and maximum levels of L (in our case 0 and 255).

$$\rho_{max} = \sqrt{Width^2 + Height^2 + (L_{max} - L_{min})^2} \quad \text{Eq. 4}$$

The quantised parameter space is visualized in Figure 8 below.

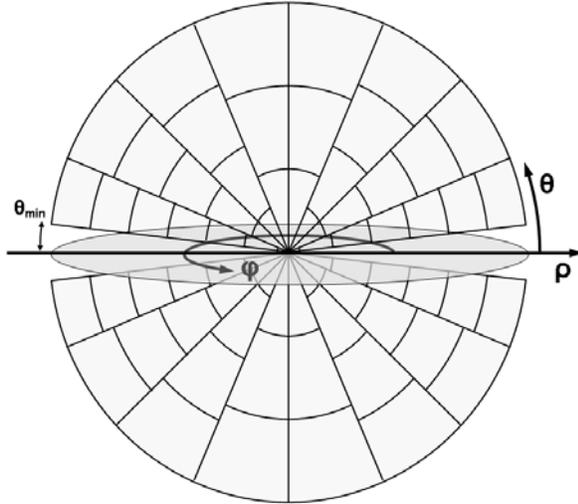


Figure 8. Quantised parameter space, with variable ρ bin size.

3. The Method

The method used to detect lightness gradient areas in the image is explained here. It is structured in four steps, namely Pre-processing, Transformation, Gradient extraction and Post-processing.

3.1. Pre-processing

Generally no pre-processing is needed for the method to produce satisfactory results. Nevertheless, there are certain issues that can affect the performance of the algorithm and could be easily addressed before applying the Gradient Hough Transform. The main problems are colour half-toning (typically associated with scanned documents), and compression artefacts as illustrated in Figure 9. To tackle these issues, a 3×3 or 5×5 (depending on the resolution) smoothing operation vastly improves the performance of the subsequent processes.

3.2. Transformation

Initially, all the (x, y, L) points of the image are transformed to the corresponding surfaces in the parameter space using Eq. 1. The parameter space is quantised as described above, using variable quantization for the ρ bin size. Depending on the image it might be preferable to use only a subset of the image pixels (for example for scanned documents in high resolution, sample one every n pixels). In this paper, for Web images and logos all the pixels were sampled, while for scanned documents at 300 DPI one every 5 pixels was sampled.

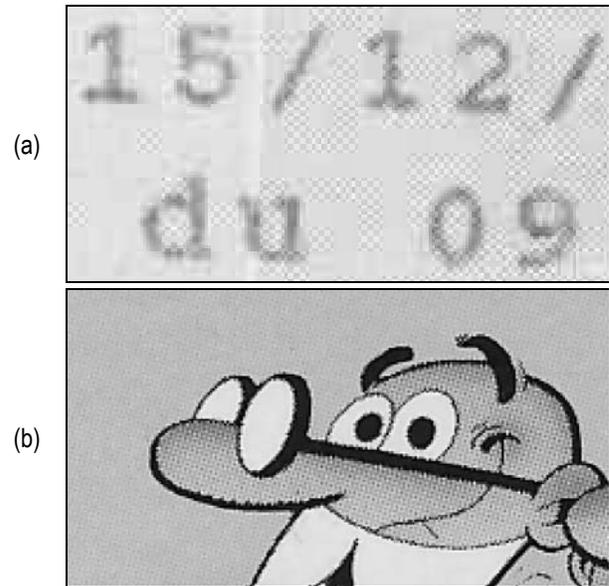


Figure 9. (a) Compression and (b) Half-toning artefacts.

3.3. Gradient Extraction

Potential gradients are extracted by a simple search in the accumulator array for the cell with the highest count. For each potential gradient identified, the associated image points are found and their corresponding votes removed from the accumulator array before the search continues for the next potential gradient. The result of this stage is a ranked list (in terms of counts) of potential gradients in the image.

3.4. Post-processing

A typical problem associated with the HT, arising from the discretisation of the parameter space, is that due to the presence of noise and distortion, true peaks are splitted between several accumulator cells. The best approach to address this issue would be to perform some kind of clustering in the 3D space before identifying potential gradients, but this is computationally expensive. Instead, here we perform some post-processing, where potential gradients already extracted which share similar properties are grouped together. The resulting groups of potential gradients are called “Meta-gradients”. The steps we follow are described below.

Step 1. Starting from the gradient with the highest count, gradients with similar parameters are grouped together into a Meta-gradient. Three thresholds are used for this grouping, which indicate the maximum

allowed difference for each of the parameters (ρ , ϕ , θ) between the participating gradients. These are set to $Tr=6^\circ$, $Tf=6^\circ$ and $Tr=15 \cdot \sin(\theta_{Meta-Gradient})$.

Step 2. The parameters of each resulting Meta-gradient are recalculated as the weighted sum of the parameters of the participating gradients. The weights used are the counts (amount of pixels) of the participating gradients.

Step 4. Pixels are labelled based on their lightness difference with each Meta-gradient defined gradient plane. For a point defined by the triplet (x , y , L) and a Meta-gradient with parameters (ρ , ϕ , θ), the Lightness distance from the plane is given by Eq. 5.

$$D = L - \frac{(\rho - x \cdot \cos \theta \cdot \cos \phi - y \cdot \cos \theta \cdot \sin \phi)}{\sin \theta} \quad \text{Eq. 5}$$

4. Evaluation

The method was tested on a dataset of images comprising Web images and logos available in the public domain, and scanned documents kindly provided by ITESOFT and from the PRIMa Layout Analysis Data Set. The dataset comprises a total of 22 images which contain 63 gradient areas. The method was able to correctly segment 87.3% of the gradient areas. The detailed results are shown in Table 3.

Table 3. Evaluation of the method

	Web Images	Logos	Scanned Documents	Total
<i>number of gradients</i>	12	18	33	63
<i>identified gradients</i>	11	16	28	55
<i>Performance</i>	91.7%	88.9%	85.7%	87.3%

Some illustrative examples are shown in Figures 10-14 that appear at the end of the paper. The first few gradients extracted are shown next to each original image (gradient areas appear in black).

5. Discussion

As in most HT implementations, a tricky point is the interpretation of the transformation, especially when working in a 3D parameter space, which is computationally expensive.

There are a few pre-processing steps that can be performed to make the gradient clusters more separable in the parameter space. One such step is to perform histogram stretching in the image before the Gradient HT. Histogram stretching preserves the

linearity of the gradient, while it was shown to help to differentiate very gentle background gradients from constant colour areas.

In terms of post-processing, the results indicate that substantial improvement can be achieved by examining the topology of the extracted gradients in parallel to their parametric representation. The spatial proximity or connectedness of gradient areas in the image could be used in post-processing to facilitate the decision making of combining them into a single Meta-gradient. Also, depending on the image in question (especially for large scanned documents), performing opening and closing morphological operations on the extracted gradient areas might improve the results substantially.

5.1. A short discussion on linearity

As mentioned in the introduction, the main assumption behind this approach is that lightness gradients are approximately linear changes of lightness over distance. Although, the validity of this statement was verified experimentally it is useful to discuss shortly why this is expected for the types of images contained in the dataset.

There are two broad categories of images that we are dealing with here, on one hand we have computer generated images such as logos and Web images, and on the other hand scanned documents. In terms of computer generated images, the answer to the linearity of the gradient is twofold. First, it is a fact that most software packages create gradients by linearly interpolating between two user-selected RGB colours. This means that the RGB values change linearly through the gradient. Computing the lightness using Eq. 2 preserves this linearity. There is a reason why a linear RGB interpolation is typically used by software. The conversion of RGB values to luminance on a display follows a non-linear power-law relationship with a typical gamma between 1.9 and 2.2. Luminance on the other hand relates to lightness (perceived intensity) through a cube-root law [10, 11]. The two effects cancel each other out to a great extent, and the result this has is that a linear change in RGB produces a lightness change which is almost perceptually linear. Therefore software packages reasonably make use of a fast linear interpolation to create something perceptually plausible.

Translating colour from any input device to any output device entails gamut mapping. When documents are printed and eventually scanned, gamut mapping is performed in each step of the process. Gamut mapping algorithms, with very few exceptions, have a linear response to the lightness component

[12, 13, 14]. Therefore printing and scanning a document generally preserves the linearity of the lightness gradients.

That is not to say that all gradients are linear. There are cases where gradients are not linear (e.g. certain natural processes can create non-linear effects). In most of the cases these are described well by a power-law and can be linearised easily by using the logarithm of the encoded lightness. In more complex cases, the gradient parametric model will have to be changed.

6. Conclusion

In this paper we presented a method for gradient detection in text images, based on an extension of the Hough Transform, and the modelling of lightness gradients as planes in the combined spatial – feature space. Preliminary results reported indicate that this method can be useful to a wider range of images.

Due to the modelling of lightness gradients, the algorithm is able to deal in a single step with gradient changes and constant colour areas, thus it provides the means for simple image segmentation. A property of the HT is that it works even when the shape under detection is partially occluded. In our case this is a very useful property both for the detection of gradient background “occluded” by text and for the extraction of individual characters created from the same gradient (e.g. in Figure 10).

The case was made for lightness gradients, but the technique discussed here is applicable to any type of linear gradients. Although the method is easily transferable to other channels, colour gradients are typically not a direct fusion of linear gradients in their individual components, and they necessitate more than merely the combination of individual results. The extension of the technique to colour gradients is identified as future work.

7. Acknowledgements

This work was partially supported by the Spanish projects TIN2006-15694-C02-02 and CONSOLIDER INGENIO 2010 (CSD2007-00018) and the fellowship 2006 BP-B1 00046. Images of scanned documents were kindly provided by ITESOFT (www.itesoft.fr) and the PRImA Layout Analysis Data Set (www.primaresearch.org). In addition, thanks must go to Carlo Gatta for his help in explaining why lightness gradients are indeed expected to be linear and to Belén Gómez, Josep Lladós and Ernest Valveny for providing useful insight.

8. References

- [1] D. Karatzas, A. Antonacopoulos, “Colour Text Segmentation in Web Images Based on Human Perception”, *Image and Vision Computing*, 25(5), pp. 564-577, 2007
- [2] J. Canny, “A computational approach to edge detection”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6), pp 355-363, 1996
- [3] D. Wang, “A Multiscale Gradient Algorithm for Image Segmentation using Watersheds”, *Pattern Recognition*, 30(12), pp. 2043-2052, 1997
- [4] W. Niblack, *An Introduction to Digital Image Processing*. Prentice-Hall, London, 1986
- [5] B. Gatos, I. Pratikakis, and S.J. Perantonis, "An Adaptive Binarization Technique for Low Quality Historical Documents", *Proceedings of the 6th IAPR Workshop on Document Analysis System (DAS 2004)*, Springer LNCS (3163), pp 102-113, 2004
- [6] P.V.C. Hough, “Method and means for recognising complex patterns”, Technical report, 1962, US Patent No. 3069654.
- [7] R.O. Duda and P.E. Hart, “Use of the Hough Transform to Detect Lines and Curves in Pictures”, *Communications of the ACM*, 15, pp. 11-15, 1972
- [8] D.H. Ballard, “Generalizing the Hough transform to detect arbitrary shapes”, *Pattern Recognition*, 13(2), pp. 111-122, 1981
- [9] ITU-R Recommendation BT.709, Basic Parameter Values for the HDTV Standard for the Studio and for International Programme Exchange (1990), [formerly CCIR Rec. 709] (Geneva: ITU, 1990)
- [10] G. Wyszecki and W.S. Stiles, “Color Science: Concepts and Methods, Quantitative Data and Formulae”, Second Edition, John Wiley & Sons, 1982
- [11] C.A. Poynton, “A Technical Introduction to Digital Video”, John Wiley & Sons, 1996
- [12] J. Morovic and M.R. Luo, “The fundamentals of gamut mapping: A survey,” *J. Imag. Sci. Technol.*, vol. 45, no. 3, pp. 283–290, 2001.
- [13] International Color Consortium, Specification ICC.1:2001-12. File Format for Color Profiles (Ver. 4.0.0) 2001.
- [14] L. Neumann and A. Neumann, “Gamut clipping and mapping based on the coloroid system,” in *Proc. CGIV 2nd Eur. Conf. Color in Graphics, Imaging, and Vision*, pp. 548–555, 2004



Figure 10. A Web image and the two largest gradients extracted.

Figure 11. A Logo and the two largest gradients extracted.

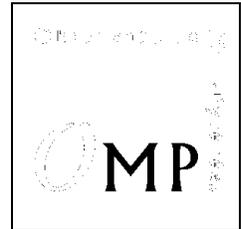
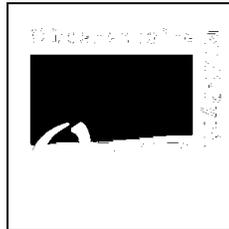


Figure 12. A Logo and the four largest gradients extracted. It is often the case that some pixels can be assigned to more than one gradient (here the white part of the blue gradient background). In such cases spatial information could improve the results.

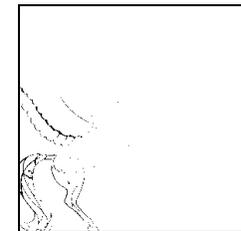
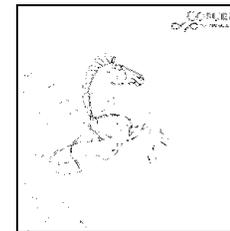
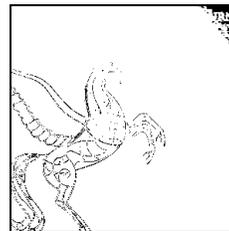
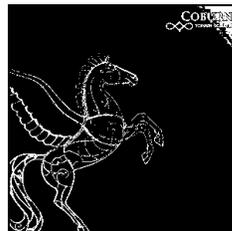


Figure 13. A Logo and the four largest gradients extracted.

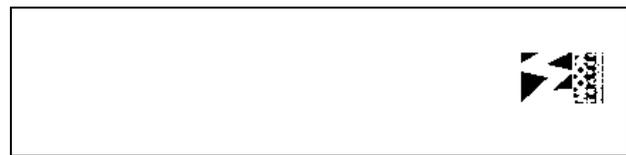


Figure 14. A scanned book cover and the three largest gradients extracted.