Robust Extraction of Text from Camera Images using Colour and Spatial Information Simultaneously

Chowdhury, S. P., Dhar, S., Rafferty, K., Das, A. K., & Chanda, B. (2009). Robust Extraction of Text from Camera Images using Colour and Spatial Information Simultaneously. JOURNAL OF UNIVERSAL COMPUTER SCIENCE, 15(18), 3325-3342. DOI: 10.3217/jucs-015-18-3325

Published in:
JOURNAL OF UNIVERSAL COMPUTER SCIENCE

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Download date:15. Feb. 2017

Journal of Universal Computer Science, vol. 15, no. 18 (2009), 3325-3342
submitted: 26/10/09, accepted: 20/11/09, appeared: 28/12/09 © J.UCS

Robust Extraction of Text from Camera Images using Colour and Spatial Information Simultaneously

Shyama Prosad Chowdhury
(Queen's University Belfast, UK
schowdhury01@qub.ac.uk)

Soumyadeep Dhar
(Videonetics Technology Pvt. Ltd., Kolkata, India
s.dhar.in@gmail.com)

Karen Rafferty
(Queen's University Belfast, UK
k.rafferty@ee.qub.ac.uk)

Amit Kumar Das
(Bengal Engineering and Science University, Shibpur, India
amit@cs.becs.ac.in)

Bhabatosh Chanda
(Indian Statistical Institute, Kolkata, India
chanda@isical.ac.in)

Abstract: The importance and use of text extraction from camera based coloured scene images is rapidly increasing with time. Text within a camera grabbed image can contain a huge amount of meta data about that scene. Such meta data can be useful for identification, indexing and retrieval purposes. While the segmentation and recognition of text from document images is quite successful, detection of coloured scene text is a new challenge for all camera based images. Common problems for text extraction from camera based images are the lack of prior knowledge of any kind of text features such as colour, font, size and orientation as well as the location of the probable text regions. In this paper, we document the development of a fully automatic and extremely robust text segmentation technique that can be used for any type of camera grabbed frame be it single image or video. A new algorithm is proposed which can overcome the current problems of text segmentation. The algorithm exploits text appearance in terms of colour and spatial distribution. When the new text extraction technique was tested on a variety of camera based images it was found to out perform existing techniques (or something similar). The proposed technique also overcomes any problems that can arise due to an unconstraint complex background. The novelty in the works arises from the fact that this is the first time that colour and spatial information are used simultaneously for the purpose of text extraction.

Key Words: Text extraction, text localisation, camera image, video frame, discrete edge boundary

Category: I.4, I.4.6, I.4.8

1 Introduction

The term *document* in no longer confined to scanned pages and any camera based image can be subject to operations like text information extraction (TIE) for applications such as optical character recognition (OCR), image/video indexing, mobile reading system for visually challenged persons etc. TIE is one of the most important aspects of traditional document processing and new challenges have surfaced with camera based coloured scene images. TIE from camera based scene images is a very difficult problem because it is not always possible to precisely define the features of text in a coloured scene image due to the wide variations in possible formats; for example, geometry (location and orientation), colour similarity, font and size. Moreover, camera based images can be subject to numerous possible degradations such as, blur, uneven lighting, low resolution and contrast which makes it more difficult to recognise any text from the background noise.

There are two main areas of research in the area of text retrieval that are popular amongst researchers. One is OCR and other is text segmentation. Recent development in the OCR techniques shows very high accuracy [Agrawal and Doermann 2008, Bieniecki *et al.* 2007, Kunte and Samuel 2007] in text recognition for well segmented characters or words. However its performance can be degraded by the presence of background noise. This is the main reason for the growing demand [Laine and Nevalainen 2006] for good text segmentation techniques. It has to be noted that text segmentation alone can not solve the requirement unless the segmentation information is used by the OCR. In this research the use of any kind of OCR information is strictly avoided which makes the techniques presented here useful as a preprocessing step before using OCR. Indeed the use of different fonts, size, colour etc. are not the major constraint for OCR to perform as required. Moreover, many researchers have done multilingual OCR for identifying characters from the union of the alphabet sets of two different languages [Peng *et al.* 2006, Chau and Hsing 2002, Jawahar and Kumar 2003]. To promote this feature of the OCR it is very much required to have a good text segmentation technique independent of language specific tuning. The set of techniques proposed here are developed without any kind of language specific information.

The accuracy of a text segmentation technique can be measured using a function of false positive and false negative error. The amount of text that has not been recognised counts as a false negative error and the inclusion of background noise is counted a false positive error. False negative error also creates inconsistency in run-time training based OCR. To reduce the false negative error this paper proposes a novel technique of use the spatial and colour information together. In brief this paper is organised as follows. Section 2 describes relevant past research. The techniques proposed in this paper for text extraction are detailed in Section 3. Section 3.1 details new colour information can be extracted from the image for text segmentation. Section 3.2 and 3.3 mainly describe the preprocessing steps required for text segmentation. Here novel methodologies for edge enhancement are proposed that are particular useful for camera based document processing. Section 3.4 details how the spatial information is utilised for text segmentation. The novelty of this research is that colour features alone can not be used to extract the text region. Rather meta data which is later merged with the spatial features to segment the text region. Section 3.5 and 3.6 are mainly for postprocessing which includes regrouping of the text blocks and noise removal. Section 4 contains experimental results followed by a conclusion in Section 5.

2 Past Research

Due to the immense potential TIE has for commercial applications, research in this area for camera based colour scene images is being pursued both in academia and industry. There are many reports in the literature regarding different techniques that researchers are pursuing to overcome the current challenges. The survey paper by Jung *et al.* [Jung *et al.* 2004] provides a good overview of different techniques. Generally the TIE techniques can be broadly divided into two classes; i) region based and ii) texture based methods. Region based methods are bottom up approaches where connected components (CC) or edges are found on the basis of their perceptive difference with the background. This is followed by merging the CCs or edges to get text bounding boxes. Some of the CCs and edge based methods are described in [Liu *et al.* 2005, Jain and Yu 1998, Lee and Kankanhali 1995, Messelodi and Modena 1992, Zhong *et al.* 1995]. The second broad class of techniques are known as texture based methods. The basic understanding for texture based methods is that typically text in an image has distinct textural properties that distinguish it from the background. Some of the representative methods are given in [Bae and Kin 1999, Jung 2001, Li *et al.* 2000, Wu and Manmatha 1999]. Different tools used for texture analysis include the Gabor filter, Fast Fourier Transform and wavelets.

Tianqiang *et al.* [Tianqiang *et al.* 2008] proposed a technique to find out the text blocks based on the a 2-D feature vector derived from the image. After getting the fixed blocks they have normalised the feature and used min-max clustering to classify the blocks. Once the classified pixel blocks are clustered they used connected component analysis to find CCs. The did not consider those images where the text is tilted. Moreover this method is not very suitable for detecting the multi font text blocks simultaneously. Colour based text segmentation technique from the camera images is used by Thillou *et al.* [Thillou and Gosselin 2005] and they faced the main problem associated with their method is the lack of the special algorithm. Their method is very much dependent on the colour of the text and the accuracy of text detection degrades with the lesser chromaticity. Ma *et al.* [Ma *et al.* 2007] tracked the text block in video considering that text are fixed in video whereas background is moving. Then they calculated the text stroke features for the text. Their method is very good for retrieval of the subtitle from the video where the position of the probable text, font, colour everything is fixed. However, it is very much dependent on the tuned parameter and may not suitable for still images and videos that contain different variation of texts. A hybrid colour-based segmentation approach has been taken by Fie *et al.* [Fei *et al.* 2008] to segment the text. Colour segmentation is done first followed by a gradient segmentation for extracting the text for a specific application. Their method is not robust for detecting multi font texts as well as extracting text from complex background.

3 Proposed Work

In spite of the source and quality of an image and a lack of knowledge about the language used within the image, humans can easily identify written information from a video or an image. It is quite interesting and important in our research that humans first extract the text and then try to recognise its lingual meaning. In this research, we aim to exploit some of the properties of text which is common among different languages. The major advantage of our work is that it is optical character recognition independent which is not the case for most of the research described in Section 2. Written text has generally some distinct property that can help humans to correctly locate it on camera based images (still or video) [Chowdhury *et al.* 2009]. Even lack of familiarity with the alphabets for different languages does not create a problem for such recognition. A few of the properties are listed.

1. Easily distinguishable text colour that is significantly different from the background.

2. Boundaries of the characters are smooth, that is, there are not many protrusions.

3. For a character, in a single row scan there exits at least one colour band covering the characters stroke-width.

These three observations have led us to develop a new technique for extraction of text from an image. A detailed description of how to segment the colour chains belonging to text within an image is given in Section 3.1. Cameras tend to have variable resolution and different focal points, in some cases the edges of the text within the images may be blurred. To enhance such edges locally an edge enhancement technique was used. This is explained with examples, in Section 3.3. This edge enhancement technique does not depend on the spatial relation of the text. Sometimes this edge enhancement technique can not perform well because of the coarse texture is present in the scene. In the region of the coarse texture, intensity values in all the colour planes demonstrate frequent and periodical variation in small scale. This create frequent gradient changes in the colour planes which does not suit to our edge enhancement technique. To avoid the effect of the coarse texture we have smoothen the coarse texture throughout the image that is described in Section 3.2. Typically, an image should have an coarse textures removed and then the edges should be "de-blurred" (Section 3.3) before the analysis detailed in Section 3.1 is carried out.

The unique approach of our technique is to correlate the colour based segmentation methodology and the spatial distribution based pattern detection which is not prominently reported in any other research. Individually, colour segmentation and spatial distribution based pattern detection are two consecutive processes and the later one is carried out once it is fully segmented from the colour domain to binary domain. In our approach we have merged these two and either segmented the colour image in only the horizontal direction and matched the pattern of this segmented portion (Section 3.4) in the vertical direction or vice-versa. Following this, it is then possible to regroup probable text blocks, this is detailed in Section 3.5. This is then followed by the removal of non text block in Section 3.6. Thus this research work is different from other approaches.

3.1 Colour Chain Segmentation

One very interesting fact about all kinds of written characters invented by human and used by all the people in this world, is that they can be written using a set of strokes by a pointing device which may be pen, pencil or piece of wood. In mathematics the term stroke is a vector having a guided direction. The meaning of stroke is that there should a narrower width perpendicular to the direction of the stroke. Here in this research we are interested about that stroke of the pointing device. If the stroke is in horizontal direction then the stroke width will form along the vertical direction and for vertical stroke, stroke width will be in the horizontal direction. For any angle in between the horizontal and vertical direction the stroke will have both horizontal and vertical components. Here we have tried to find out the possible stroke width of the characters. As there is no prior knowledge of the text location in the image, stroke needs to be searched in the whole image. To link consecutive pixels in the same direction, colour chains are segmented from that image by exploiting the general text properties described earlier. Which are easily distinguishable text colour, smooth boundary of the text and having a stroke-width of the same colour.

Let us define a colour image F as a two dimensional function of three colour values that range from 0 to $(L-1)$. $F^R(i,j)$, $F^G(i,j)$ and $F^B(i,j)$ represents the red, green and blue colour component of i^{th} row and j^{th} column position pixel. The colour euclidean distance between two pixels with (i_1, j_1) and (i_2, j_2) coordinates is $E((i_1, j_1), (i_2, j_2))$;

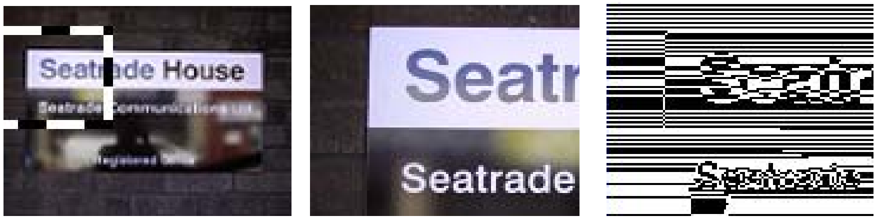
$$E((i_1, j_1), (i_2, j_2)) = \sqrt{\sum_{\forall X,\ X \in \{R,G,B\}} [F^X(i_1, j_1) - F^X(i_2, j_2)]^2}\ . \tag{1}$$

For segmenting the colour chains, the threshold values for checking colour similarity is kept at 30% in each colour plane and the euclidean distance among two colour values is kept at 45% of the farthest possible euclidean distance $(\sqrt{3}L)$. It has to be noted that the colour chains are segmented from the image in both of the horizontal and vertical direction separately. These two results of orthogonal analysis will be merged in Section 3.5. Now consider two consecutive pixels (i_1, j_1) and $(i_1, j_1{+}1)$ along the same row. These two pixels will be assigned to the same chain if the following conditions are satisfied;

$$|F^X(i_1, j_1) - F^X(i_1, j_1+1)| < 0.3L \quad \forall X,\ X \in \{R,G,B\} \tag{2}$$

$$E((i_1, j_1), (i_1, j_1+1)) < 0.45(\sqrt{3}L)\ . \tag{3}$$

A visual representation of colour chains is shown in the Figure 1(c). Consecutive colour chains in the same row have been labelled with alternative black and white runs. Figure 1(a) illustrates the original image with a marked region. Figure 1(b) and (c) are the magnified images of the marked region and colour chains into that region.



(a) (b) (c)

Figure 1: (a) Camera image with a marked region (b) Details of the marked region (c) Details of the colour chain segmentation result in the marked region

3.2 Coarse Texture Removal

In the process of colour chain segmentation some difficulties may be encountered because of smooth transitions at the edge of character components e.g. due to image blurring. An image may suffer from blur, it can also contain coarse texture where coarse texture is defined by the small periodical variation of the intensities in different colour planes. Our proposed edge enhancement works on the gradient and it requires a prominent gradient for enhancing the edges. Coarse textures can create problems when measuring the actual gradient of the object of interest. Thus it is necessary to reduce any coarse texture in the image.

In our technique coarse texture is removed in each of the red, green and blue colour planes of the image. As described earlier, colour chain segmentation is done separately in both of the horizontal or vertical directions, coarse texture also needs to be removed in the same direction. Assuming segmentation in the horizontal direction, let us define the horizontal gradient value function GV_H^X and horizontal gradient sign function GS_H^Xon a particular colour plane X. We may write;

$$GV_H^X(i,j) = F^X(i,j) - F^X(i,j+1) \;\; \forall X, \;\; X \in \{R,G,B\} \tag{4}$$

$$GS_H^X(i,j) = \begin{cases} -1 \; GV_H^X(i,j) < 0 \\ \;\;\, 0 \; GV_H^X(i,j) = 0 \\ \;\;\, 1 \text{ otherwise} \end{cases} \quad \forall X, \;\; X \in \{R,G,B\} \; . \tag{5}$$

To find the coarse texture, we need to determine a consecutive pair of points where the changes of the gradient value is either positive to negative or negative to positive. For coarse texture removed in the horizontal direction, let us assume that such pair of points are based at (i,j_1) and (i,j_2) $(j_1 < j_2)$ where the condition is $GS_H^X(i,j_1) = GS_H^X(i,j_2)$ and at least one of the following two conditions will hold

$$GS_H^X(i,j_1-1) = GS_H^X(i,j_2-1) = -GS_H^X(i,j_1) \tag{6}$$

$$GS_H^X(i,j_1+1) = GS_H^X(i,j_2+1) = -GS_H^X(i,j_1) \; . \tag{7}$$

Consecutive pair means that there should not be any other point (i,j_3) such that $(j1 < j3 < j2)$ for which the above conditions are true. From the well known Roll's theorem, it can be claimed that there exists one point (i,j_4) $(j_1 < j_4 < j_2)$ such that any of the following conditions satisfies

$$GS_H^X(i,j_1) = GS_H^X(i,j_4-1) \neq GS_H^X(i,j_4) \tag{8}$$

$$GS_H^X(i,j_1) = GS_H^X(i,j_4+1) \neq GS_H^X(i,j_4) \; . \tag{9}$$

Then it is possible to determine that there is a coarse texture in between (i,j_1) and (i,j_2) if the equation Eq.10 and either of the equations among Eq.11 and

Eq.12 holds;

$$-0.08L < (GS_H^X(i, j_1) - GS_H^X(i, j_2)) < 0.08L \tag{10}$$

$$-0.08L < (GS_H^X(i, j_1) - GS_H^X(i, j_4)) < 0.08L \tag{11}$$

$$-0.08L < (GS_H^X(i, j_2) - GS_H^X(i, j_4)) < 0.08L \ . \tag{12}$$

If the above conditions satisfy then it may be concluded that the coarse texture is detected in between the points (i, j_1) and (i, j_2). To remove this coarse structure all the intermediate pixels (i, j_3) such that $(j1 < j3 < j2)$ are replaced with linearly interpolated values of $F^X(i, j_1)$ and $F^X(i, j_2)$ based on the distance.

3.3 Edge Enhancement

Prior to colour chain segmentation it is necessary to perform edge enhancement to be performed in the image before. This is really a preprocessing operation to ensure there are to make the sharp transitions at the edges of any text or object within the image which will aid the detection. Popular edge enhancement techniques like the sobel filter enhances the edge using the intensity information of a pixel eight neighbour pixel values. In our case this is not sufficient as edges are continuously changing with a small gradient. In some applications the stabilised inverse diffusion equation is used by Pollak *et al.* [Palloak *et al.* 2000] can be used to enhance the edges. This algorithm is good for images where the number of different regions are known a priory, which is also not applicable in this research.

Thus we propose a two pass enhancement technique for the enhancement of the edges. In the first pass, all the fluctuating but prominent gradient transitions will be enhanced. This is known as light edge enhancement (LEE) and the objective is to find a set of consecutive candidate points and enhance the edge between them. LEE needs to be operated in the same direction that we want to segment the colour image. In LEE, the first step is to obtain the set of candidate points thus, it is necessary to consider all points in the raster scan direction.

To obtain a set of candidate points, start from a position (i, j) and store the initial gradient sign value in S^R, S^G, S^B. Continue to check the gradient until there is no other point (k, l) consist of at least one colour plane Y $(Y \in \{R, G, B\})$, for which $|F^Y(i, j) - F^Y(k, l)| \geq 0.3\%L$. At any point (p,q) the gradient checking condition for continuing accumulation in the candidate points is $GS^X(p, q) = S^X \quad or \quad |GV^X(p, q)| < 0.05L$ where $(X \in \{R, G, B\})$. Once this condition is not fulfilled, LEE stops and starts a new search from the next point. On availability of a set of successful and consecutive candidate points we find the point which is in the middle of the total colour distribution among the candidate points in the Y plane which has the maximum change in colour value between the maximum and minimum points. This middle point will come within the maximum and minimum colour valued points in Y plane. Using this point as a gate we change all the colour values in minimum side with the minimum value and with the maximum value in maximum side. Thus the edge information is significantly enhanced. Figure 2(a) describes where LEE is required and Figure 2(b) illustrates the outcome of LEE.

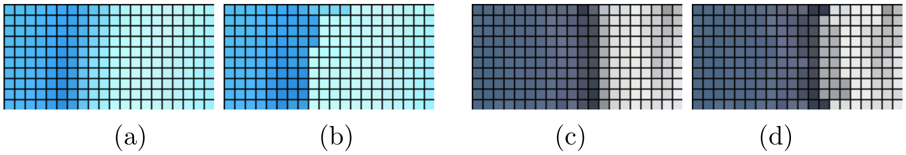


Figure 2: (a) (b) Input and result of LEE and (c) (d) Input and result of HEE

Next, on completion of the LEE, heavy edge enhancement (HEE) will be carried out. The main difference between HEE and LEE is that HEE needs a strictly positive or negative gradient throughout the candidate pixels for all three colour planes. The condition for accumulating the candidate points for HEE is $GS^X(p,q) = S^X$ where ($X \in \{R, G, B\}$). After finding the candidate points, enhancement of the edge will be achieved using the same technique implemented for the LEE. It is also necessary to find the colour mean point among the colour distribution in the candidate points. The purpose of HEE is to find the consecutive points where the gradient is strictly of same sign for a colour plane. All three colour planes need to follow this condition. Figure 2(c) describes where HEE is required and Figure 2(d) is the outcome after HEE.

These edge enhancement techniques are used as a preprocessing step on the camera images. An added advantage of this edge enhancement technique is that it is also useful to remove or reduce motion blur from camera image sequences. That allows us to treat images from the still camera and frames from the video images in the same manner.

3.4 Spatial Distribution Based Pattern Detection

After the colour chain segmentation is performed on the enhanced edge image, the next step is to link the chains vertically to get a two dimensional spatial colour component. A number of features that have been observed on the vertical distribution of the horizontal chains for a particular character are listed.

1. In a single row there exits at least one colour chain.

2. Horizontal chains cover the horizontal continuity of the characters.

3. The size of the chains that are positioned vertically one after another, has mainly three types of relation for different reasons such as:
   (a) Almost of same size; This can be related to the same stroke width of a character or the same length of a protrusion.
   (b) Continuous change in size; The main reason behind this is circular bending in the character.
   (c) Sudden change in size; this only occurs when there are protrusions that come out of the character
4. The average colour values of two chains that are positioned one after another vertically are almost the same unless written with high contrast colour mixtures; this is not very common.

Based on these observations, horizontal chains are linked vertically to from a two dimensional object. Let us number the chains chronologically 1 to N and define three functions LC, RC and WC which represents the left most and right most column positions and width of a chain. The average red, green and blue colour values of the k^{th} $(1 \leq k \leq N)$ chain is denoted as $A^R(k)$, $A^G(k)$ and $A^B(k)$. The function MIN selects the minimum value among a set of values. For two vertical neighbour chains a and b $(1 \leq a, b \leq N)$, the presence of a vertical overlap can be established by either of $(LC(a) \leq LC(b) \leq RC(a))$ or $(LC(b) \leq LC(a) \leq RC(b))$. If there is an overlap then the amount of vertical overlap of a and b chains can be calculated by a function OV where,

$$OV(a,b) = MIN((RC(a) - LC(b)), (RC(b) - LC(a))) \ . \tag{13}$$

If there is no vertical overlap among a and b then $OV(a,b) = 0$. The euclidean colour distance function of the average colour of the two chains (a and b) is defined as EC; where

$$EC(a,b) = \sqrt{\sum_{\forall X,\ X \in \{R,G,B\}} [A^X(a) - A^X(b)]^2} \ . \tag{14}$$

To get a set of groups representing the characters fully or partially, we have imposed two sets of conditions to link the selected chains vertically. Selection of the chains are made on the basis of their widths. Chains having width more than the width of largest possible character (WLC) are rejected. For any selected chain c $(1 \leq c \leq N)$, it is necessary that $WC(c) \leq WLC$ where, $WC(c)$ is the width of the chain c. This is to remove noise from the output and it is justified as this whole process is carried out without having any resolution information. Two further sets of conditions can be used to satisfy the observations 3a and 3b mentioned in this section. If a and b $(1 \leq a, b \leq N)$ have to satisfy the first condition then there has to be a significant vertical overlap between them; such that

$$OV(a,b) > 0.8MIN(WC(a), WC(b)). \qquad (15)$$

The second set of condition (Eq.16, Eq.17 and Eq.18) is used so that the flattened part of the characters can be identified, for example to allow for bending. Here, in addition to the first observation (1) with more relaxation, we have used the average colour distance among the chains to exploit the observation number 4.

$$OV(a,b) > 0.5MIN(WC(a), WC(b)) \qquad (16)$$

$$|A^X(a) - A^X(b)| < 0.3L \;\; \forall X, \; X \in \{R, G, B\} \qquad (17)$$

$$EC(a,b) < 0.45(\sqrt{3}L) \qquad (18)$$

The output from this process is a set of candidate groups that contains vertically linked chains. It is now necessary to restrict the number of candidate groups before further processing is carried out. Any candidate group which has a smaller number of chains than the smallest possible character (HSC) are ignored. Now restriction of the candidature is the number of chains present in that group. For each group the number of chains present are determined. This is given by the function $GC(m)$ where m is the group number. Thus the criterion to be a candidate group is $GC(m) \geq HSC$.

3.5 Regrouping of the Probable Text Blocks

In this part of work, all the candidate groups are used to form a two dimensional matrix I which has the same dimension as the input image. For any position q, if it is part of any chain belonging to a candidate group then we assign $I(q) = 1$ else $I(q) = 0$. Now using the methods described in Sections 3.1, 3.2 and 3.3, vertically instead of horizontally a new set of vertical chains are obtained. Let us define P as the number of vertical chains. Then let us determine the topmost position $TC(p)$ and bottom most position $BC(p)$ for each chain where p is the chain number and $1 \leq p \leq P$.

Now with a vertical ortho-raster scan on I we find all the gaps on the series of $1s$ or all the protrusions on the series of $0s$ having length less than HSC. All the gaps of 0 are filled with 1 and all the protrusions are removed by 0. This step is important to remove miss detection or some false extra detection among the probable text bands.

The initial groups of chains GC will now be distorted and we need to regroup these using the matrix I as a template. Using another matrix J of same size as the input image, at any point q, $J(q) = 1$ if

$$XC = q \mid \exists XC, \quad XC \in \{LC(s), RC(s), TC(t), BC(t)\} \qquad (19)$$

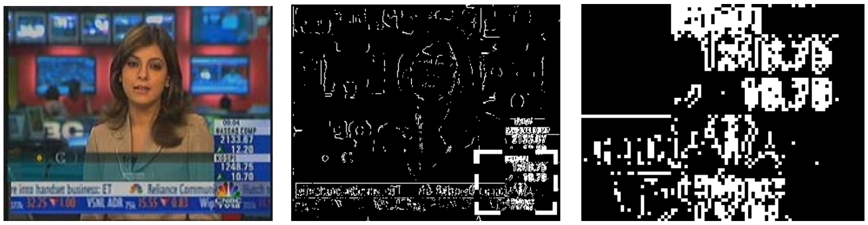
where, $(1 \leq s \leq N)$ and $(1 \leq t \leq P)$ otherwise $J(q) = 0$. Thus the matrix J will contain all the transition points for both of the horizontal and vertical chains. In Figure 3(a) an image is shown whilst its j matrix is illustrated in Figure 3(b). Here the original image has a mark region for further description. If there are multiple characters forming a text side by side then there are some additional characteristics for that group of characters that can be observed such as:

1. If the characters are thin and the density of the characters are high then the density of the number of chains in a particular region increases.

2. If there are many characters side by side then there exists repetitions of the chains whose average colour value is similar to one another.

3. Spatial characteristics of the chains due to inter character space is almost similar to that of characters.

It is possible to regroup the chains using the connectivity obtained from the matrix I and chain density of transition points (1) present the matrix J. In the new set of the groups, all the connected points having 1 in matrix I and having similar density of 1 in matrix J, will get grouped. Different densities of the transition points are shown in Figure 3(c) which is the magnified J matrix of the region marked in Figure 3(b). Let us consider that U number of new groups are formed and k^{th} $(1 \leq k \leq U)$ new group has density of transitions points (obtained from matrix J) $TD(k)$. All the groups have some spatial characteristics based on the two dimensional spread of the member points. Denote the height and width of the spread of member points and existence of the k^{th} group as $HG(k)$ and $WG(k)$ and $EF(k)$. Then in the initial state $EF(k) = 1$ $(\forall k,\ 1 \leq k \leq U)$.



(a) (b) (c)

Figure 3: (a) Camera image (b) Horizontal and vertical transition points with a marked region (c) Transition points detail of the marked region

All the groups are then checked and a group x $(1 \leq x \leq U)$ will be removed if any of the values in $HG(x)$ or $WG(x)$ is small and have a value lower than

$TD(x)$. This is done by exploiting the observation made on 1 and 3. Removal of a group x is noted by making $EF(k) = 0$. Here the threshold values are obtained easily if we consider that in most of the cases the aspect ratio of the character bounding box varies within 0.5 *to* 2. And it is obvious that for any of the horizontal or vertical scan line there should be at least two transition points to define the boundary.

3.6 Removal of Non Text Blocks using Repetition Feature

Here our aim is to exploit the observation 2 made on Section 3.5. Towards that goal we use each of the groups x $(1 \leq x \leq U)$ which are the outcome Section 3.5 and not removed yet $(EF(x) = 1)$. Using the points inside that group as mask, we are recalculating horizontal chains as we have done previously on Section 3.1 on the edge enhanced image obtained from Section 3.3. Let us number the chains those come out from a particular group chronologically 1 to Q. Denote average red, green and blue colour values of the k^{th} $(1 \leq k \leq Q)$ chain by $A^R(k)$, $A^G(k)$ and $A^B(k)$ as it was done previously in Section 3.4. Also denote width of the chain and euclidean colour distance function of the average colour of two chains by WC, EC consecutively. Now in each pair of the horizontal chains a and b $(1 \leq a, b \leq Q)$ come in single row scan, check whether the average colour distance follow these similarity conditions,

$$|A^X(a) - A^X(b)| < 0.05L \quad \forall X,\ X \in \{R, G, B\} \tag{20}$$

$$EC(a, b) < 0.07(\sqrt{3}L)\ . \tag{21}$$

For a group x $(1 \leq x \leq U)$ that contains all the chains y which are satisfying the above mentioned condition; add there width $WC(y)$ into CN. Let the sum of the all chains TW $(TW = \sum WC(y)\ \forall y, 1 \leq y \leq Q)$. Now remove the group x $(1 \leq x \leq U)$ for which the following conditions are satisfied

$$\frac{HG(x)}{WG(x)} < 0.5 \tag{22}$$

$$\frac{CN}{TW} < 0.5\ . \tag{23}$$

Removal of a group x will make the value of $EF(x)$ to 0. All the remaining groups y for which $EF(y) = 1$ will be considered for the intended text block of the image.

4 Experimental Results

In order to test the developed algorithms, they were applied to approximately 400 images. Some of these images were taken from the ICDAR'03 Robust Reading competition data set ([ICDAR 2003]). In addition we used samples from our own collection of images and video frames. The images utilised were varied in nature and as such this presented a very robust test for the newly developed algorithms. That is the text within the sample images were varied in terms of colour, size, font and orientation. It was found that for the sampled images, we achieved a 94% success rate of correctly identifying text within the image. In the other 6% of cases, the text within the images was incorrectly identified as background. This generally occurred when the text within the image was faint and as such it was very difficult to determine a prominent gradient at the edges of the text.

In addition, some incorrect identification of text was also evidenced. In these cases, part of the background to the image was identified as text. This occurred in 4% of the images, however in the other 96% the background was correctly classified. In most of the cases of misdetection a logo that is present in the image is detected as text.

Image	Number of	Execution time (s)		
size (pixel)	tested images	mean	mode	variance
110592 (384 x 288)	25	0.3	0.3	0.1
307200 (640 x 480)	172	0.8	0.7	0.5
1228800 (1280 x 960)	52	2.7	3.3	2.0
1920000 (1600 x 1200)	70	3.7	4.7	2.9

Table 1: Statistical analysis on the execution time

For each of the test images, the execution time of the algorithm was also recorded. The results are shown in Table 1. It can be seen that the time for the complete analysis per image is low and therefore suitable for many applications. In this experiment the code was written in the C language and processed using a standard PC with Core 2 Duo 2.2GHz CPU and 2GB RAM. It is observed that execution time linearly increases as the size of the image under analysis also increases. For a small image of approximately 100K pixels, the execution time is as low as 0.3s whereas it takes almost 8.8s to process an image of size

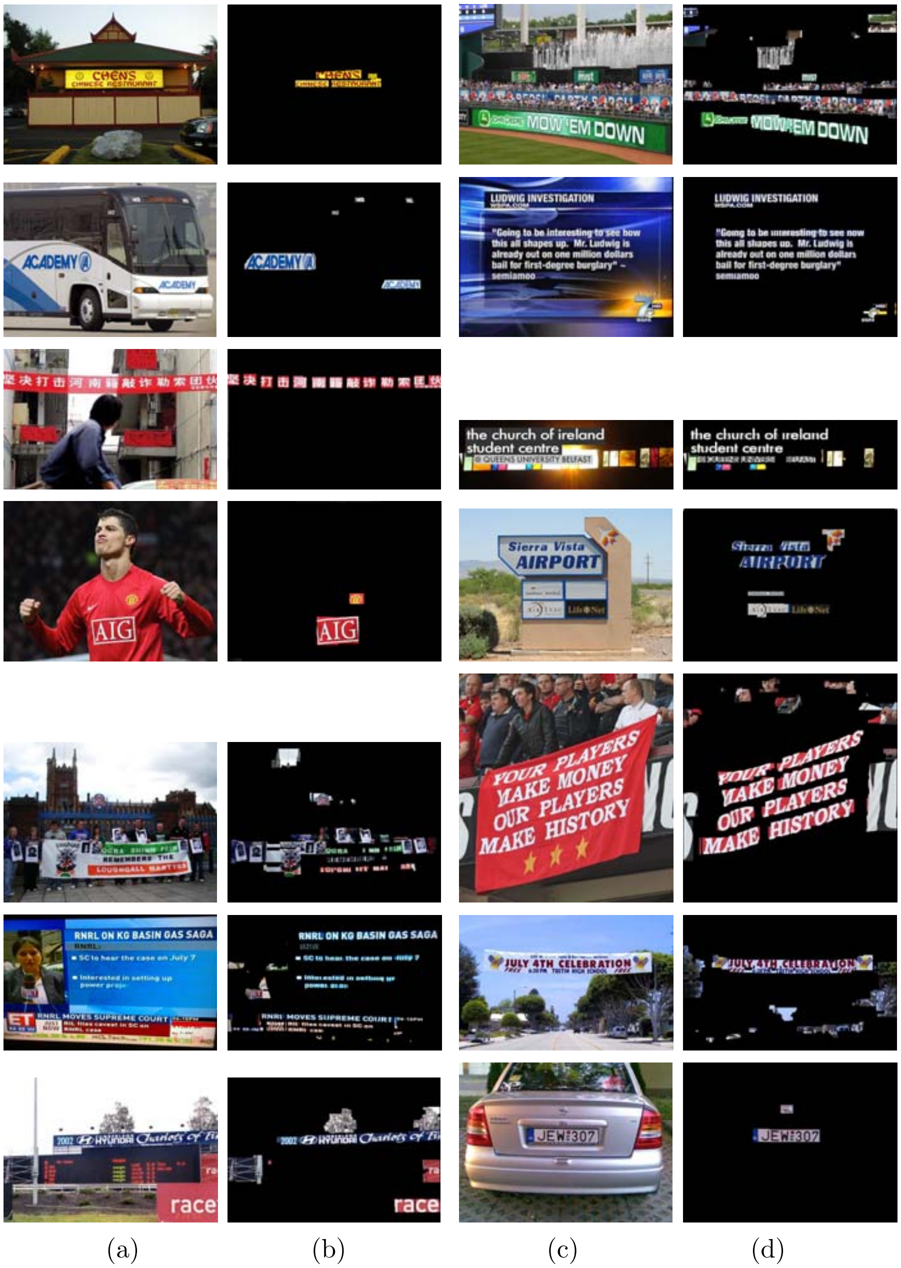

CHEN'S
CHINESE RESTAURANT
MOW 'EM DOWN
ACADEMY
LUDWIG INVESTIGATION
WSPA.COM
"Going to be interesting to see how this all shapes up. Mr. Ludwig is already out on one million dollars bail for first-degree burglary" - semiamoo
坚决打击河南籍敲诈勒索团伙
the church of ireland
student centre
QUEENS UNIVERSITY BELFAST
AIG
Sierra Vista
AIRPORT
Life Net
YOUR PLAYERS
MAKE MONEY
OUR PLAYERS
MAKE HISTORY
REMEMBERS THE
LOUGHGALL MARTYRS
RNRL ON KG BASIN GAS SAGA
SC to hear the case on July 7
Interested in setting up power projs
RNRL MOVES SUPREME COURT
JULY 4TH CELEBRATION
2002 HYUNDAI Chariots of Fire
race
JEW 307
(a)
(b)
(c)
(d)

Figure 4: (a) (c) Still and video camera image (b) (d) Detected text regions

$\sim$ 3000K pixels. It is also observed that execution time increases for a complex scene where frequent changes in colour are present in the image. A sample of the end results are illustrated in Figure 4. Overall, these results are extremely encouraging and illustrate the fact that when colour and spatial information are simultaneously considered together, they provide a robust and fast way of extracting text from colour images.

5 Conclusion

This paper presents a novel technique for extracting text from colour camera based images. A number of processes are applied to a give image. Initially any coarse texture is smoothed out. Then LEE and HEE algorithms are applied to the images so that any edges within the image are more prominent. The coarse texture removal and edge enhancement are simply pre-processing operations that are applied to the images before the text segmentation and greatly aid in text segmentation. Then colour and spatial information within the image is used to group and detect probable text within this images. It may be noted that the thresholds used in the algorithm are not very rigid and sensitive. Reasonable variation of the threshold values are possible without appreciable degradation in the final results.

These algorithms represent an extension to the state of the art since they use both spatial and colour information simultaneously to correctly identify text within an image. With currently available text segmentation techniques it is a challenge to extract text which can be very different in terms of colour, size, font and orientation. However, we have shown through the application of our novel techniques to 400 images, that they work very robustly and effectively when extracting different variations of text. Indeed a 94% success rate was achieved for text extraction from images. In the 6% of cases that text was not correctly extracted, the text was usually very faint in nature, meaning that it would be extremely difficult for any type of detection technique to recognise it. Some misidentification of text also occurred in that in 4% of cases, background information was wrongly classified as text. Again, this mostly happened whenever there was a logo in the background and this was then classified as text. In the future we will look at resolving this misidentification. However, currently we have developed text extraction algorithms that work very effectively when trying to extract text of different colour, size, font and orientation from camera based images. Finally, we claim that the novelty of this work is the blending of the colour feature and spatial distribution together for text extraction. We have also used a good edge enhancement technique well suited to different types of scene analysis.

References

[Agrawal and Doermann 2008] M. Agrawal and D. Doermann. Re-targetable ocr with intelligent character segmentation. In *The Eighth IAPR International Workshop on Document Analysis Systems, DAS '08.*, pages 183 – 190, 16 - 19 September 2008.

[Bae and Kin 1999] B. T. C. Y. Bae and T. Y. Kim. Automatic text extraction in digital videos using fft and neural network. In *Proceedings of IEEE International Fuzzy Systems Conference*, volume 2, pages 112–1115, 1999.

[Bieniecki *et al.* 2007] W. Bieniecki, S. Grabowski, and W. Rozenberg. Image preprocessing for improving ocr accuracy. In *Proc. of International Conference on Perspective Technologies and Methods in MEMS Design, MEMSTECH 2007*, pages 75 – 80, 23-26 May 2007 2007.

[Chau and Hsing 2002] R. Chau and Y. Hsing. Multilingual text categorisation for global knowledge discovery using fuzzy techniques. In *IEEE International Conference on Artificial Intelligence Systems, ICAIS 2002*, pages 82–86, 2002.

[Chowdhury *et al.* 2009] S. Chowdhury, S. Dhar, A. Das, B. Chanda, and K. McMenemy. Robust extraction of text from camera images. In *Proceedings of the 10th International Conference on Document Analysis and Recognition, ICDAR 2009*, pages 1280 – 1284, Barcelona, Spain, 26 - 29 July 2009.

[Fei *et al.* 2008] X. Fei, M. Dong, and F. Wong. Digital camera based visual-to-textual translation system. In *Proc. of 5th International Conference on Visual Information Engineering, VIE 2008.*, pages 323 – 326, 29 July - 1 August 2008.

[ICDAR 2003] I. 03. Robust reading competition data set. Active on January 2009.

[Jain and Yu 1998] A. K. Jain and B. Yu. Automatic text location in images and video frames. *Pattern Recognition*, 31(12):2055–2076, 1998.

[Jawahar and Kumar 2003] S. S. R. K. C. V. Jawahar, M. N. S. S. K. Pavan Kumar. A bilingual ocr for hindi-telugu documents and its applications. In *Seventh International Conference on Document Analysis and Recognition, ICDAR'03*, volume 1, page 408, 2003.

[Jung 2001] K. Jung. Neural network based text location in colour images. *Pattern Recognition Letters*, 22 (14):1503–1515, December 2001.

[Jung *et al.* 2004] K. Jung, K. I. Kim, and A. K. Jain. Text information extraction in images and video: A survey. *Pattern Recognition*, 37:977–997, 2004.

[Kunte and Samuel 2007] R.S.Kunte and R. Samuel. An ocr system for printed kannada text using two - stage multi-network classification approach employing wavelet features. In *International Conference on Conference on Computational Intelligence and Multimedia Applications*, volume 2, pages 349 – 353, 13 - 15 December 2007.

[Laine and Nevalainen 2006] M. Laine and O. Nevalainen. A standalone ocr system for mobile cameraphones. In *IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications*, pages 1 – 5, 11-14 September 2006.

[Lee and Kankanhali 1995] C. M. Lee and A. Kankanhalli. Automatic extraction of characters in complex in complex images. *Int. J. of Patter Recognition and AI*, 9(1):67–82, 1995.

[Li *et al.* 2000] H. Li, D. Doermann, and O. Kia. Automatic text detection and traking in digital video. *IEEE Image Processing*, 9 (1):147–155, January 2000.

[Liu *et al.* 2005] Q. Liu, C. Jung, S. Kim, Y. Moon and J. Kim. Text localization based on edge-pca and svm in images. In *Samsung Technology Conference*, 2005.

[Ma *et al.* 2007] R. Ma, W. Hu, Q. Huang, J. Wang, T. Wang, and Y. Zhang. Robust text stroke extraction from video. In *IEEE International Conference on Multimedia and Expo*, pages 1391 – 1394, 2-5 July 2007.

[Messelodi and Modena 1992] S. Messelodi and C. M. Modena. Automatic identification and skew estimation of text lines in real scene images. *Pattern Recognition*, 32:791–810, 1992.

[Palloak *et al.* 2000] I. Pollak, A. S. Wilsky, and H. Krim. Image segmentation and edge enhancement with stabilized inverse diffusion. *IEEE Tr. on IP*, 9(2), February 2000.

[Peng *et al.* 2006] L. Peng, C. Liu, X. Ding, and H. Wang. Multilingual document recognition research and its application in china. In *Second International Conference on Document Image Analysis for Libraries, DIAL06.*, page 132, 27-28 April 2006.

[Thillou and Gosselin 2005] C. Thillou, C, and B. Gosselin. Color text extraction from camera-based images: the impact of the choice of the clustering distance. In *Proc.of Eighth International Conference on Document Analysis and Recognition*, volume 1, pages 312 – 316, 29 August - 1 September 2005.

[Tianqiang *et al.* 2008] P. Tianqiang, T. Pohuang, and L. Bicheng. A robust video text extraction method based on text traversing line and stroke connectivity. In *Proc. of 9th International Conference on Signal Processing, ICSP 2008*, pages 1002 – 1005, 26-29 October 2008.

[Wu and Manmatha 1999] V. Wu and R. Manmatha. Textfinder: An automatic system to detect and recognize text in frames. *IEEE Pattern Analysis and Machine Intelligence*, 21 (11):1224–1229, November 1999.

[Zhong *et al.* 1995] Y. Zhong, K. Karu, and A. K. Jain. Locating texts in complex color images. *Pattern recognition*, 28(10):1523–1535, 1995.