

University of Bristol



DEPARTMENT OF COMPUTER SCIENCE

On the Recovery of Oriented Documents from Single Images

Paul Clark Majid Mirmehdi

On the Recovery of Oriented Documents from Single Images

P. Clark and M. Mirmehdi

Department of Computer Science, University of Bristol,
Bristol, BS8 1UB, UK.
{pclark, majid}@cs.bris.ac.uk
<http://www.cs.bris.ac.uk/~{pclark, majid}>

Abstract. A method is presented for the fronto-parallel recovery of text documents in images of real scenes. Initially an extension of the standard 2D projection profile, commonly used in document recognition, is introduced to locate the horizontal vanishing point of the text plane. This allows us to segment the lines of text, which are then analysed to reveal the style of justification of the paragraphs. The change in line spacings exhibited due to perspective is then used to recover the vertical vanishing point of the document. We do not assume any knowledge of the focal length of the camera. Finally, a fronto-parallel view is recovered, suitable for OCR or other high-level recognition. We provide results demonstrating the algorithm's performance on documents over a wide range of orientations.

1 Introduction

Optical character recognition (OCR) is a long-standing area of computer vision which in general deals with the problem of recognising text in skew-compensated fronto-parallel images. In preparation to apply OCR to text from images of real scenes, a fronto-parallel view of a segmented region of text must be produced. This is the issue considered in this paper. The extraction of oriented documents from camera images is a new challenge in document processing, made possible by high resolution digital cameras, as well as recent developments and demands in the multimedia environment. There has been little research into the recognition of text in real scenes in which the text is oriented relative to the camera. Processing and compensating for such perspective skew has applications in replacing the document scanner with a point-and-click camera to facilitate non-contact text capture, assisting the disabled and/or visually impaired, wearable computing tasks requiring knowledge of local text, and general automated tasks requiring the ability to read where it is not possible to use a scanner.

Previous work in estimating the orientation of planar surfaces in still images varies in the assumptions made to achieve this. Ribeiro and Hancock [10] and Criminisi and Zisserman [5] both presented methods which used the distortion of repetitive planar texture to estimate the vanishing points of the plane. Affine transformations in power spectra were found along straight lines in [10], and

correlation measures were used in [5] to determine first the orientation of the vanishing line and then its position. Although text has repetitive elements (characters and lines) these elements do not match each other exactly, and sometimes may cover only a small area of the image. Rother [11] attempted to find orthogonal lines in architectural environments, which were assessed relative to the camera geometry. Murino and Foresti [8] used a 3D Hough transform to estimate the orientation of planar shapes with known rectilinear features. Gool et al. [12] and Yip [13] both found the skewed symmetry of 2D shapes which have an axis of symmetry in the plane, allowing for affine recovery. We require recovery from perspective transformation, and as with these latter works we will use a priori information about the 2D shape we are observing. Other than our recent work in [3], the only other work known to the authors on perspective recovery of text is [9]. The author seeks visual clues in the image which correspond to horizontal and vertical features on the document plane. Unfortunately, the vertical vanishing points of the text plane cannot be robustly estimated when only one vertical clue is present. Examples of this situation are when the document is single-column and when paragraphs are not fully justified.

Knowledge of the principal vanishing points of the plane on which text lies is sufficient to recover a fronto-parallel view. We observe that in a paragraph which is oriented relative to the camera, the lines of text all point towards the horizontal vanishing point of the text plane in the image. Also, paragraphs often exhibit some form of justification, either with straight margins on the left and/or right, or if the text is centred, a central vertical line around which the text is aligned. In such cases these vertical lines point toward the vertical vanishing point of the text plane. We have therefore concentrated our work on the recovery of paragraphs with three lines of text or more, with the reasonable assumption that at least some justification exists (left, right, centred or full).

To avoid the problems associated with bottom-up grouping of elements into a paragraph model, in this work we ensure the use of all of the global information about the paragraph at one time. The principle of 2D projection profiles are extended to the problem of locating the horizontal vanishing point by maximising the separation of the lines in the paragraph. The segmented lines of text are then analysed to reveal the style of justification or alignment of the paragraph. Depending on the type of paragraph, either margins or line spacings are used to provide an estimate of the vertical vanishing point. This allows us to recover the perspective skew (or orientation) of the plane of text, and hence generate a fronto-parallel view. The use of line spacings to find the position of the vertical vanishing point makes it possible to recover left, right, and centrally justified paragraphs accurately, a previously unsolved problem. Throughout the work we make no use of the focal length of the camera, hence the techniques are applicable to images taken from cameras with unknown.

The rest of the paper is structured as follows. In Section 2 we briefly review our previous work which provides the input to the work described here. Sections 3 to 5 discuss the paragraph model fitting stage, location of the horizontal vanishing point, separation of the lines of text, and estimation of the

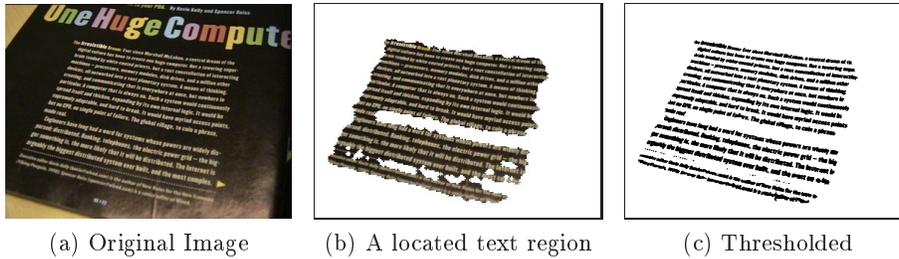


Fig. 1. Preparation of paragraph for planar recovery

vertical vanishing point, followed by some examples of recovered documents. We conclude and consider future work in Section 6.

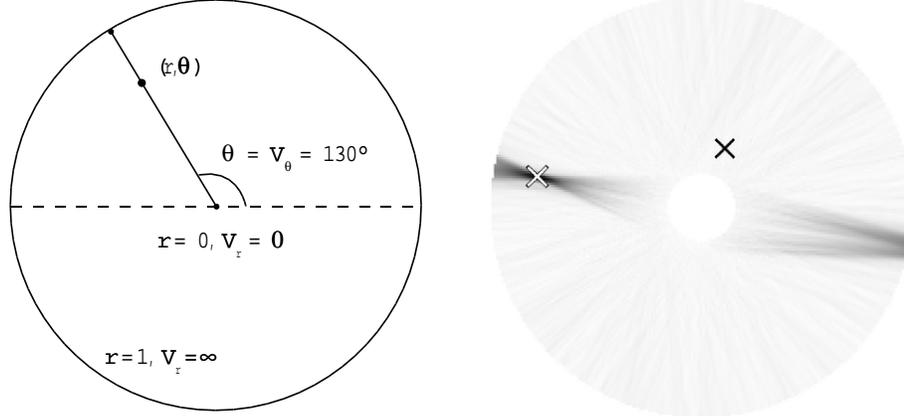
2 Finding Text Regions

In [1] we introduced a text segmentation algorithm which used localised texture measures to train a neural network to classify areas of an image as text or non-text. Figure 1(b) shows a large region of text which was found in Figure 1(a) using this approach. In this work we consider the output of the system presented in [1] and analyse each region individually to recognise the shape of the paragraph, recover the 3D orientation of the text plane, and generate a fronto-parallel view of the text.

In order to analyse the paragraph shape, we first require a classification of the text and background pixels. Since the region provided by the text segmentation algorithm will principally contain text, the background and foreground colours are easily separable through thresholding. We choose the average intensity of the image neighbourhood as an adaptive threshold for each pixel, in order to compensate for any variation in illumination across a text region. The use of partial sums [6] allow us to compute these local thresholds efficiently. To ensure the correct labelling of both 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 of text elements, the group of pixels with the lower proportion is labelled as text, and the other group as background. The example shown in Figure 1(c) demonstrates the correct labelling of some light text on a dark background and is typical of the input into the work presented here.

3 Locating the Horizontal Vanishing Point

In [7], Messelodi and Modena demonstrated a text location method on a database of images of book covers. They employed projection profiles to estimate the skew angle of the located text. A number of potential angles were found from pairs of components in the text, and a projection profile was generated for each angle.



(a) Relationship between search space C and \mathbb{R}^2 (b) Scores for all projection profiles in C generated from Figure 1(c)

Fig. 2. Search space C

They observed that the projection profile with the minimum entropy corresponds to the correct skew angle. This guided 1D search is not directly applicable to our problem, which is to find a *vanishing point* in \mathbb{R}^2 , with two degrees of freedom. In order to search this space, we will generate projection profiles from the point of view of vanishing points, rather than from skew angles.

We use a circular search space C as illustrated in Figure 2(a). Each cell $c = (r, \theta)$, $r \in [0, 1)$ and $\theta \in [0, 2\pi)$, in the space C corresponds to a hypothesised vanishing point $\mathbf{V} = (V_r, V_\theta)$ on the image plane \mathbb{R}^2 , with scalar distance $V_r = r/(1-r)$ from the centre of the image, and angle $V_\theta = \theta$. This maps the infinite plane \mathbb{R}^2 exponentially into the finite search space C . A projection profile of the text is generated for every vanishing point in C , except those lying within the text region itself (the central hole in Figure 2(b)).

A projection profile B is a set of bins $\{B_i, i = 0, \dots, N\}$ into which image pixels are accumulated. In the classical 2D case, to generate the projection profile of a binary image from a particular angle ϕ , each positive pixel \mathbf{p} is assigned to bin B_i , where i is dependent on \mathbf{p} and ϕ according to the following equation:

$$i(\mathbf{p}, \phi) = \frac{1}{2}N + N \frac{\mathbf{p} \cdot \mathbf{U}}{s} \quad (1)$$

where $\mathbf{U} = (\sin \phi, \cos \phi)$ is a normal vector describing the angle of the projection profile, and $s > N$ is the diagonal distance of the image. In this equation, the dot product $\mathbf{p} \cdot \mathbf{U}$ is the position of the pixel along the axis of the projection profile in the image defined by ϕ . Manipulation with s and N is then employed to map from this axis into the range of the bins of the projection profile.

In our case, instead of an angle ϕ , we have a point of projection \mathbf{V} on the image plane, which has two degrees of freedom. Our bins, rather than representing parallel slices of the image along a particular direction, must represent angular

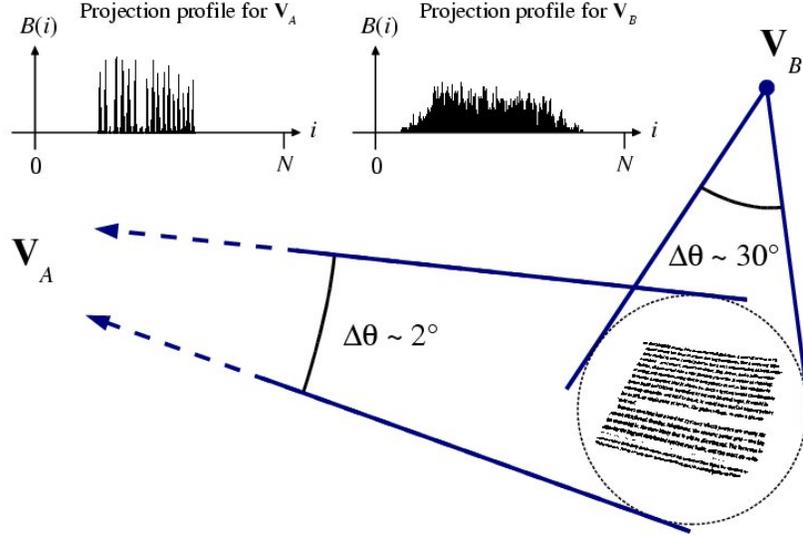


Fig. 3. Two potential vanishing points V_A and V_B , and their projection profiles.

slices projecting from V . Hence, we refine (1) to map from an image pixel p into a bin B_i as follows:

$$i(p, V) = \frac{1}{2}N + N \frac{\text{ang}(V, V - p)}{\Delta\theta} \quad (2)$$

where $\text{ang}(V, V - p)$ is the angle between pixel p and the centre of the image, relative to the vanishing point V , and $\Delta\theta$ is the size of the angular range within which the text is contained, again relative to the vanishing point V . $\Delta\theta$ is obtained from $\Delta\theta = \text{ang}(V + t, V - t)$ where t is a vector perpendicular to V with magnitude equal to the radius of the bounding circle of the text region (shown in Figure 3). Unlike s in (1), it can be seen that $\Delta\theta$ is dependent on the point of projection V . In fact $\Delta\theta \rightarrow 0$ as $V_r \rightarrow \infty$ since more distant vanishing points view the text region through a smaller angular range. The use of t to find $\Delta\theta$ ensures that the angular range over which the text region is being analysed is as closely focused on the text as possible, without allowing any of the text pixels to fall outside the range of the projection profile's bins. This is vital in order for the generated profiles to be comparable, and also beneficial computationally since no bins need to be generated for the angular range $2\pi - \Delta\theta$ which is absent of text.

Having accumulated projection profiles for all the hypothesised vanishing points using (2), a simple measure of confidence is applied to each projection profile B . The confidence measure was chosen to respond favourably to projection profiles with distinct peaks and troughs. Since straight lines are most clearly distinguishable from the point where they intersect, this horizontal vanishing

point and its neighbourhood will be favoured by the measure. We found the squared-sum of derivatives

$$SSQ(B) = \sum_{i=1}^{N-1} (B_{i+1} - B_i)^2 \quad (3)$$

to respond better than entropy and squared-sum measures, as well as being efficient enough to compute. The confidence of each of the vanishing points with regard to the binarised text in Figure 1(c) is plotted in Figure 2(b), where darker pixels represent a larger squared-sum, and a more likely vanishing point.

To locate the vanishing point accurately, the resolution of the search space must be sufficient to hypothesise a large number of potential vanishing points. (During experiments we found empirically that 10^4 vanishing points was reasonable.) Since each vanishing point examined requires the generation and analysis of a projection profile, a full search of the space, as shown in Figure 2(b), is computationally expensive. However, due to the large scale features of the search space, we introduced an efficient hierarchical approach to the search. An initial scan of the search space at a low resolution is performed, requiring the generation of only a few hundred projection profiles. Adaptive thresholding is then applied to the confidence measures of these projection profiles, to extract the most interesting regions of the search space. (In our experiments, this was taken to be the top scoring 2% of the space.) A full resolution scan is then performed on these interesting regions in the search space, requiring the generation of a few further projection profiles close to the expected solution. Finally, the projection profile with the largest confidence is chosen as the horizontal vanishing point of the text plane. The winning projection profile and an example of a poor projection profile are shown in Figure 3, and marked in Figure 2(b) with a white cross and a black cross respectively. The hierarchical search reduces the processing time on a Sun Enterprise from over two minutes to under ten seconds. In this situation, the derivative measure performs far more accurately than the squared-sum and entropy measures, which can mislead the hierarchical search by also responding favourably to the *vertical* vanishing point of a ‘thin’ document, i.e. a document which has been rotated about its vertical axes.

In order to assess the performance of the algorithm, simulated images such as those in Figure 4 were generated at various orientations ranging from 0° to 90° in both yaw and pitch, resulting in 900 test images. Figure 5 shows the accuracy of recovery of the horizontal vanishing point for these images, calculated as the relative distance of the located vanishing point \mathbf{H} from the ground truth \mathbf{H}_{gt} , given by $-|\mathbf{H} - \mathbf{H}_{gt}|/|\mathbf{H}_{gt}|$. As can be seen, the accuracy of the algorithm’s performance begins to drop as the orientation of the plane approaches 90° in yaw or pitch. In these cases, the document has been rotated so as to be almost orthogonal to the view plane, and hence nearly invisible in the image, explaining the reduction in performance. The slope of the graph at low yaw may be attributed to the finity of the search space C . Since the vanishing points in these situations lie close to infinity, the distances of the located vanishing points can not be precise. Nevertheless, the vanishing point chosen will be in the correct

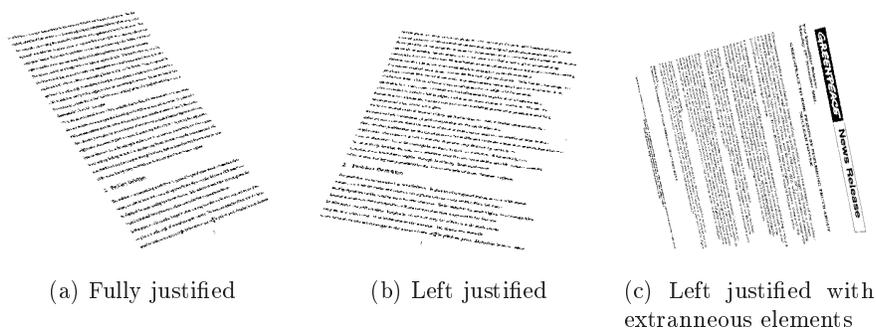


Fig. 4. Examples of simulated images used for performance analysis.

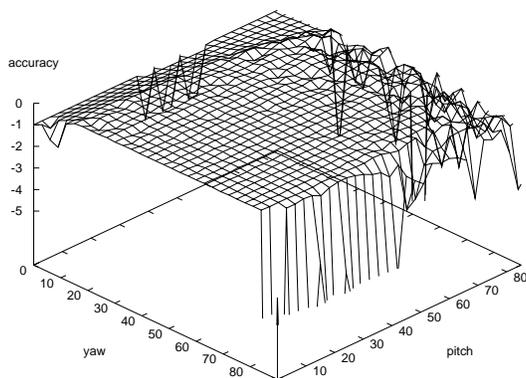


Fig. 5. Accuracy of recovery of the horizontal vanishing point (HVP) for simulated paragraphs at various orientations.

direction, and suitably large so as not to affect further processing. A numerical analysis of the performance is given in Table 2 and discussed at the end of this paper.

4 Determining the Style of Justification

The location of the horizontal vanishing point, and the projection profile of the text from that position, now make it possible to separate the individual lines of text. This will allow the style of justification of the paragraph to be determined, and lead to the location of the vertical vanishing point.

We apply a simple algorithm to the winning projection profile to segment the lines. A *peak* is defined to be any range of angles over which all the projection

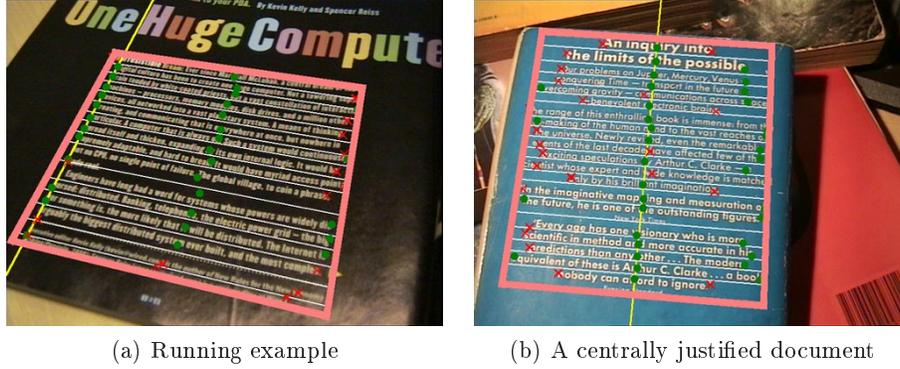


Fig. 6. Summary of the information extracted from two test images. Line segmentation marked in white; points for margin fitting in green (used) and red (rejected outliers); the baseline in yellow; the rectangular frame on the text plane in pink.

profile's bins register more than K pixels, taken as the average height of the interesting part of the projection profile:

$$K = \frac{1}{y - x + 1} \sum_{i=x}^y B_i \quad (4)$$

where x and y are the indices of the first and last non-empty bins respectively. A *trough* is defined to be the range of angles between one peak and the next. The central angle of each trough is used to indicate the separating boundary of two adjacent lines in the paragraph. We project segmenting lines from the vanishing point through each of these central angles. All pixels in the binary image lying between two adjacent segmenting lines are collected together as one line of text. The result of this segmentation is shown in Figure 6. Most lines of text are segmented accurately, although in Figure 6(a) a very short line has been ignored. Noisy pixels, very short lines, and extraneous document elements may become attached to a true text line, or be segmented as a separate line. However, the processing which follows will compensate for this irrelevant data.

Depending on the type of paragraph being recovered, there are now two possible ways to analyse the segmented lines to reveal the vertical vanishing point. If the paragraph is *fully justified*, then the left and right margins of the text are straight, and intersecting these two margin lines will provide us with the vertical vanishing point, and the problem is fully resolved. Alternatively, if the paragraph is *left justified*, *right justified*, or *centred*, a straight line will be visible either on the left margin, on the right margin, or through the centres of the lines. The vanishing point will lie somewhere along this *baseline*. However, the actual position of the vanishing point will be unknown, and must be estimated.

Initially, we must determine the structure of the paragraph, i.e. its style of justification. We collect the left end, the centroid, and the right end of each of

Condition	Type of paragraph
$E_L \simeq E_C \simeq E_R$	Fully justified.
$\min(E_L, E_C, E_R) = E_L$	Left justified.
$\min(E_L, E_C, E_R) = E_R$	Right justified.
$\min(E_L, E_C, E_R) = E_C$	Centrally justified.

Table 1. Classifying the type of paragraph

the segmented lines, to form three sets of points P_L, P_C, P_R respectively. Since we anticipate some justification in the paragraph, we will expect a straight line to fit well through at least one of these sets of points, representing the left or right margin, or the centre line of the paragraph. To establish the line of best fit for each set of points, we use a RANSAC (random sampling consensus, [4]) algorithm to reject outliers caused for example by short lines, equations or headings. Given a set of points P , the line of best fit through a potential fit $F = \{\mathbf{p}_i, i = 1, \dots, M\} \subseteq P$ passes through \mathbf{c} , the average of the points, at an angle ψ found by minimising the following error function:

$$E_F(\psi) = \frac{1}{M^5} \sum_{i=1}^M ((\mathbf{p}_i - \mathbf{c}) \cdot \mathbf{n})^2 \quad (5)$$

where $\mathbf{n} = (-\sin \psi, \cos \psi)$ is the normal to the line, M^2 normalises the sum, and a further M^3 rewards the fit for using a large number of points. Hence for the three sets of points P_L, P_C, P_R we obtain three lines of best fit F_L, F_C, F_R with their respective errors E_L, E_C, E_R . It is now possible to classify the style of justification of the paragraph using the rules in Table 1. Figure 6(a) shows the line F_C passing through the left margin of the paragraph. In this case $E_L < E_C$ and $E_L < E_R$, hence the second condition in Table 1 is satisfied and the paragraph is correctly identified as being left justified.

For fully justified paragraphs, the recovery of the vertical vanishing point is trivial, and may be achieved by intersecting the left and right margins of the paragraphs, the results of which are shown later in Table 2 and Figure 8(a). However, for a left justified, right justified or centralised paragraph, we can retrieve only one baseline. The other two fitted lines will have significant errors due to the jagged margin(s). In these situations, a different method must be used to determine the position on the baseline at which the vanishing point lies.

5 Locating the Vertical Vanishing Point

Like train tracks which merge together as they approach the horizon, the spacings exhibited between adjacent lines of text in the image will vary relative to their distance from the camera. This change in spacing can be used to determine the angle at which the document is tilted, and hence the vertical vanishing point of the text plane.

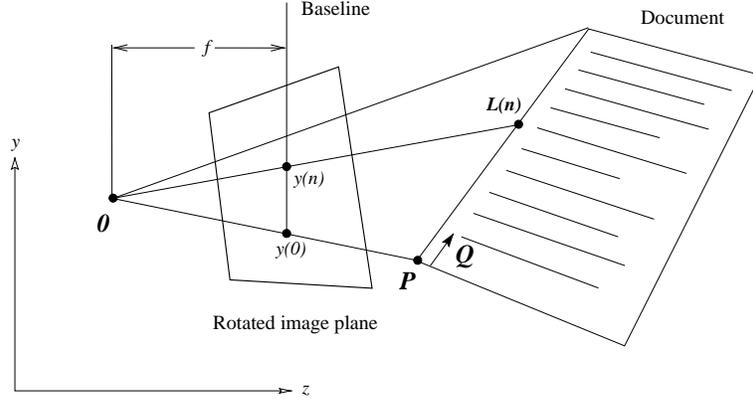


Fig. 7. The geometry involved in line spacings.

By rotating the image plane to place the baseline vertically, we may disregard the x -coordinates and deal solely in the y, z plane, as shown in Figure 7. Here, the bottom of the paragraph is positioned at \mathbf{P} with lines occurring at even spacings of distance \mathbf{Q} . Therefore the n th line from the bottom of the paragraph will appear at

$$\mathbf{L}(n) = \mathbf{P} + n\mathbf{Q} \quad (6)$$

and will project to the point in the image plane

$$y(n) = f \frac{\mathbf{L}(n)_y}{\mathbf{L}(n)_z} = f \frac{\mathbf{P}_y + n\mathbf{Q}_y}{\mathbf{P}_z + n\mathbf{Q}_z} \quad (7)$$

where f is the focal length of the camera. Without losing the nature of the projection, we may scale the scene about the focal point in order to set \mathbf{P}_z to f , hence modelling the paragraph as if it touched the image plane. In this case, $\mathbf{P}_y = y(0)$, and we may rewrite (7) as:

$$y(n) = U \frac{1 + nV}{1 + nW} \quad (8)$$

with $U = y(0)$ and only two unknowns, $V = \mathbf{Q}_y/\mathbf{P}_y$ and $W = \mathbf{Q}_z/\mathbf{P}_z$. The cancelling of the focal length f in this way means that the technique is applicable to images for which the internal parameters of the original camera are unknown. By projecting the centroids of the lines of text located in the image from the horizontal vanishing point onto the baseline, estimates for $y(n)$ may be obtained. However, since it is common for documents to also contain lines of text which are not part of an evenly spaced paragraph, and for extraneous elements to enter

the data, the n th line found in the image may lose correspondence with the n th line in the paragraph model. To fit a curve of position $y(n)$ against line number n would be unwise in this situation. It is therefore preferable to fit the curve of *line spacing* Y_n against *position* X_n , defined as:

$$Y_n = y(n + 1) - y(n) \quad (9)$$

$$X_n = y(n) \quad (10)$$

In this case any odd lines will appear as isolated outliers in line spacing, but will not propagate through the remaining points. By substituting (8) into the definition of line spacing (9), the curve of Y in terms of X may be written in two parts:

$$Y(X) = U \left(\frac{1 + (n(X) + 1)V}{1 + (n(X) + 1)W} - \frac{1 + n(X)V}{1 + n(X)W} \right) \quad (11)$$

with $n(X)$ derived by a similar substitution of (8) into the definition of line position (10) and rearranging to:

$$n(X) = - \frac{U - X}{UV - XW} \quad (12)$$

Initial parameters V and W are chosen for line fitting using a simple estimate for the error optimisation. However, due to the complexity of 11 and 12, many false minima exist, and one of these may be converged upon during optimisation. Therefore, to refine the parameters, an initial fit is made with an approximation of (11):

$$Y(X) = \frac{UV}{1 + n(X)W} \quad (13)$$

This ensures that parameters close to the desired minima are obtained before the final fitting. Once optimised, V and W are plugged into (8) to find the altitude of the horizon $y(\infty) = UV/W$. By reversing the rotation made earlier to bring the baseline upright, this point will correspond to the location of the vertical vanishing point in the original image.

Figure 8 shows the accuracy of recovery of the vertical vanishing point using the methods described. In Figure 8(a) it can be seen that, as expected, intersecting the left and right margins of a fully justified paragraph gives a good estimate of the vertical vanishing point. Figure 8(b) shows the accuracy when the line spacings are taken into account. The method provides comparative results for all of the simulated images except those documents oriented beyond 80° in pitch, where the algorithm begins to fail. As with the horizontal vanishing point in Section 3, this may be explained by the orientation of the document becoming nearly perpendicular to the image plane. At such an extreme tilt, even if the lines of text are separated correctly, their proximity in the image means there is little accuracy in position and spacing for the curve fitting. In practice, documents at such extreme angles cannot practically be read or used by

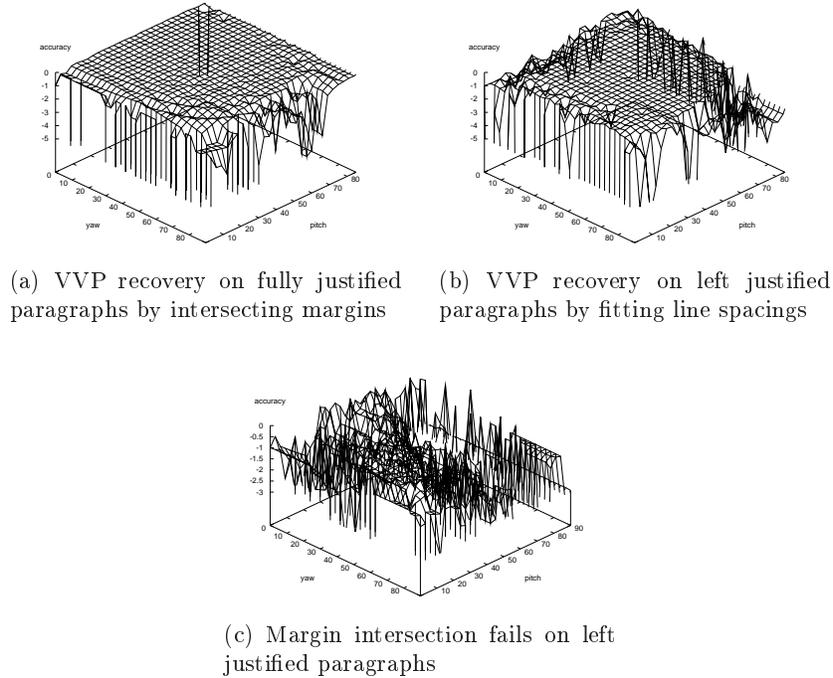


Fig. 8. Accuracy of recovery of vertical vanishing point (VVP) on simulated paragraphs at various orientations.

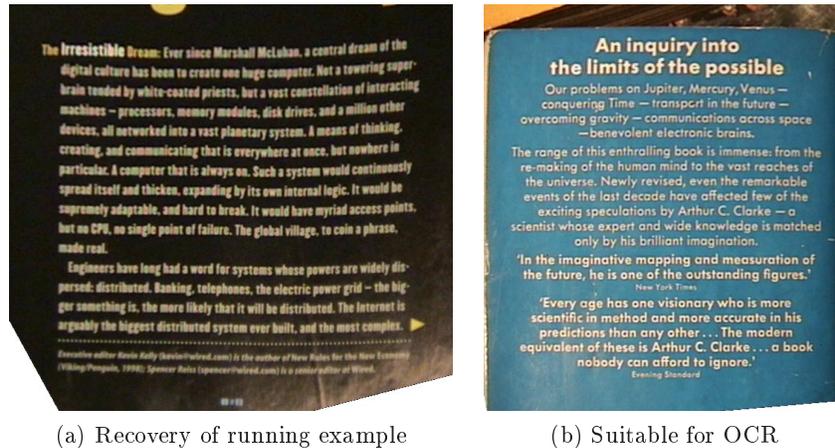
OCR once recovered, hence this failure does not concern us. The advantage of the line spacings method is that it provides consistent results for paragraphs which are not fully justified. In contrast, the poor performance of the margins method when dealing with documents which are not fully justified can be seen in Figure 8(c).

The results for these experiments, and the location of the horizontal vanishing point in Section 3, are shown numerically in Table 2. The vanishing point (VP) error is calculated as the relative distance of the vanishing point from its ground truth, as described in Section 3. The angular error is derived from the final determined orientation of the horizontal and vertical vectors of the text plane. It can be seen that the accuracy of location of the vertical vanishing point is reasonable for both the margin intersection and the line spacings method. As the last row of Table 2 shows, intersecting margins is not suitable for documents with jagged edges.

Having found the vanishing points of the plane, we may project two lines from each to describe the left and right margins and the top and bottom limits of the paragraph. These lines are intersected to form a quadrilateral enclosing the text, as shown in Figure 6. This quadrilateral is then used to recover a fronto-parallel viewpoint of the paragraph of text.

	VP error	Angular error
HVP using projection profiles	0.129	2.16°
VVP using margin intersection	0.0785	2.08°
VVP using line spacings	0.133	3.30°
VVP using margin intersection on left justified paragraphs	1.23	24.5°

Table 2. Average error for the various methods over 10° to 80° in yaw and pitch.



(a) Recovery of running example

(b) Suitable for OCR

Fig. 9. Fronto-parallel recovery of example documents in Figure 6.

The recovered page for the running example may be seen in Figure 9, alongside the second example introduced earlier. Further examples in Figure 10 show the recovery of different styles of paragraphs with left justified and centrally aligned text. Figure 10(a) shows the recovery of a segmented region of a book cover. Despite text of different sizes, and other image noise such as the specularities, the document is recovered robustly. Figure 10(b) shows a centrally justified paragraph which has been recovered at high resolution and is easily readable. In Figure 10(c) a left justified document was correctly identified and recovered, despite occlusion of part of the paragraph. When we applied commercial OCR software to the image in Figure 9(b), 87% of the characters and 94% of the words were recognised correctly.

6 Discussion

We have presented a novel approach to the fronto-parallel recovery of a wide range of documents under perspective transformation in a single image, without knowledge of the focal length of the camera. Projection profiles from hypothesised vanishing points are used to robustly recover the horizontal vanishing point

