Graph based line segmentation on cluttered handwritten manuscripts

Fredrik Wahlberg  Anders Brun
Centre for Image Analysis, Uppsala University and Swedish University of Agricultural Sciences
{fredrik.wahlberg,anders.brun}@cb.uu.se

Abstract

We propose a two phase line segmentation method for handwritten pre-modern densely written manuscripts. The proposed method combines the robustness of projection based methods with the flexibility of graph based methods. The result are cut-outs of the image containing each text line. Overlapping characters, help lines and degradation can create foreground elements spanning several lines that are hard to separate. We treat the problem of finding a cut through the text line separation as a graph optimization problem, which allows for flexible separation of entangled components.

The proposed method has been tested on two medieval sources with satisfying results. A comparison to similar methods, using standard metrics, is presented.

1. Introduction

In handwritten text recognition, line segmentation is a crucial step. The target usage of our algorithm is large scale automatic segmentation of handwritings in several script types. The proposed algorithm has been successfully applied in word spotting[8] but has not been described in detail or evaluated in previous publications.

1.1. Source material

Our sources are two medieval manuscripts, labelled C61 and C64, from the Uppsala University Library. They were chosen since they are representative of a large number of manuscripts in the library’s possession. C61 contains the Revelations of Saint Bridget of Sweden in Old Swedish. C64 is a handbook in Latin for priests called Summula de ministris et sacramentis ecclesiasticis. The quality of the manuscript’s pages vary but all images are of high quality (high resolution, truthful colours, geometric consistency etc.), as shown in figure 1.

For the sake of this paper, we assume that layout analysis of the manuscript pages have identified bounding boxes containing text. However, some notes between lines together with a number of enlarged initial characters remain.

2. Previous work

Several methods have been described for text line and word segmentation in handwritten sources, see for instance [5, 7].

A common approach is to use projections to find text lines and single words[6, 1]. A projection profile is created by finding the sum of foreground pixels after binarization in each horizontal image line. In this profile, high sums indicate text lines. This method is very robust but also lacks in exactness if the line separation is small.
A way of using graph theory to group connected components (characters) is described by [4]. They let connected components be nodes in a graph where the a graph cut corresponds to a text line. An advantage to using graph methods is the ease of which their algorithm finds severely curved lines. A disadvantage of graph-based methods is the lack of robustness (e.g. if spacing between text lines is smaller than between words the grouping might span different text lines).

A less common approach is to try to segment the image into the areas where each line resides (i.e. trying to find the separations between lines instead of connected components belonging to a line ). In [2], seam carving is used to find lines and separate the image areas they reside in. A horizontal seam is defined as a path through the image containing one pixel in each column and where each seam pixel must be vertically adjacent (e.g. ±2 pixels) to its neighbours. The cost of a seam is the sum of the pixel values it contains. The minimum cost horizontal seams through the grey level distance transform of the image then corresponds to the positions of text lines. Similarly maximum cost seams correspond to separations between lines.

We have used the seam carving method for comparison below. However, finding the lines on our sources was found to not be stable enough for large scale unsupervised segmentation. Hence, to get a fair comparison, instead of finding the lines using the original method we used the estimates given by piecewise projection. Our modified seam carving segmentation then only finds the cuts between lines using the original method.

3. Proposed method

In the proposed method, the line segmentation is performed in two steps. First an estimation of the positions of the text line separations is performed using a projection based method. This results in piecewise linear cuts separating the text lines. In the second step the cuts through the line separations are refined using a graph based method. The combination of the two methods give the robustness of projection based methods but with the flexibility and accuracy of graph based methods.

3.1. Segmentation using piecewise projection profiles

To create a projection profile curve the page is binarized and all the pixels in each horizontal line are summed. Valleys in the profile curve correspond to line separations. To mitigate the problem with slanting or irregular lines (where valleys overlap) the page can be split vertically into several strips. A projection profile is then found on each strip. Piecewise linear cuts through text line separations are then grown by matching valleys between strips. In most cases empty areas or illuminated characters will be found in the strips towards the page edges so matching valleys in neighbouring strips have been found starting at the centre strip. An illustration is shown in figure 2 with both a projection profile and the resulting segmentation.

By using many strips, lines can be slanted to the degree that one line’s end can be below the following line’s start which would not be possible without using strips. A piecewise projection profile using 15 splits has shown itself to give good results on our sources. A Gaussian filter was applied to the projection profiles to make finding valleys more robust. The width of the filter was the line spacing (found to be approximately the double median height of the connected components on the page).

3.2. Refinement of cuts using graphs

The initial estimation of the line cuts need to be refined so as not to cut through characters and separate overlapping characters where possible. This can be achieved by treating the segmentation as a graph optimization problem.

The line separation cut should keep a certain distance from the text if possible to allow for binarization errors (e.g. due to degradation), hence there should be a
penalty for cutting too close to the character.

We assume the character regions spanning between two text lines are of the two types, touching and overlapping characters. Overlapping characters are hard to separate without a model of the character’s ideal shape. Touching characters on the other hand only slightly overlap, i.e. only a few pixels overlap or touch. Hence, if a cut through a character is unavoidable it should be as short as possible.

The page image I is binarized using Otsu’s method where all pixels belonging to the foreground (e.g. characters) are set to belong to the class F. $D_f$ and $D_b$ are the binary image distance transforms on the foreground ($F$) and background ($B = F^C$) respectively. To identify a specific pixel $I$, $D_f$ and $D_b$ can be indexed as $(i, j)$. Each pixel is represented by a graph node that is 8-connected to its neighbours with weighted (non-symmetric) edges. The costs for following an edge to a node representing the pixel at position $(i, j)$ are given by equation 1.

$$\text{EdgeCost}(i, j) = \begin{cases} 
D_f(i, j) \cdot 2, & I(i, j) \in F \\
1 - \frac{D_b(i, j)}{10}, & I(i, j) \notin F \land D_b(i, j) < 10 \\
0.01, & I(i, j) \notin F \land D_b(i, j) \geq 10 
\end{cases} \tag{1}$$

The new line separation cut is the minimum cost path found using Dijkstra’s algorithm. As start and goal points for the new cut, the earlier estimation of the separation’s endpoints are used. An example of the cost at each pixel and minimum cost path is shown in figure 4. An example of the resulting segmentation is show in figure 3.

4. Evaluation

We have developed a ground truth dataset from C61 (20 pages, 466 lines) and C64 (20 pages, 460 lines) for the purpose evaluation. The piecewise projection method is first evaluated on its own. Then we used the modified seam carving method and the proposed method to refine it’s results. The test scores can be seen in table 1.

4.1. Performance measure

As performance measure we have chosen the measures used in the line segmentation competitions at IC-DAR 2009 and ICHFR 2010[3]. Detection rate (DR) and recognition accuracy (RA) are defined by equation 2. $N$ and $M$ are the number of pixels belonging to the ground truth and the segmentation results respectively and $T$ the one-to-one matches.

$$DR = \frac{T}{N}, \quad RA = \frac{T}{M} \tag{2}$$

These measures were originally used for evaluating binary images where there is no ambiguity in which pixels belong to the foreground. We have chosen to modify the definition of $M$ as the number of foreground pixels belonging to the automatically segmented line and the ground truth of any line. This assumption can be made since background areas would otherwise be counted as foreground.

For easier comparison, a combined (single number) performance metric [3] is shown in equation 3.

$$FM = \frac{2DR \cdot RA}{DR + RA} \tag{3}$$
Table 1. Comparison between piecewise projection, modified seam carving and the proposed method for sources C61 and C64. All values are in percent. The proposed method is consistently better. Differences are significant concerning text information though small in percent (sec. 5).

<table>
<thead>
<tr>
<th></th>
<th>C61 DR</th>
<th>C61 RA</th>
<th>C61 FM</th>
<th>C64 DR</th>
<th>C64 RA</th>
<th>C64 FM</th>
</tr>
</thead>
<tbody>
<tr>
<td>PWP</td>
<td>97.8</td>
<td>98.0</td>
<td>97.9</td>
<td>96.4</td>
<td>96.8</td>
<td>96.6</td>
</tr>
<tr>
<td>Seams</td>
<td>98.2</td>
<td>98.4</td>
<td>98.3</td>
<td>97.6</td>
<td>98.0</td>
<td>97.8</td>
</tr>
<tr>
<td>Prop.</td>
<td>98.7</td>
<td>98.9</td>
<td>98.8</td>
<td>98.5</td>
<td>98.8</td>
<td>98.7</td>
</tr>
</tbody>
</table>

5. Conclusion

We have proposed a two step method for robust line segmentation of densely written pre-modern manuscripts. The proposed method combines the advantages of projection based and graph based segmentation approaches. The disadvantages of each approach is mitigated by the other. As the numbers in table 1 indicates, the proposed method is consistently better.

After manual analysis of the remaining error of the modified seam carving and proposed method, it was found to mainly stem from misclassification of dots and strokes above characters. These errors will hardly show in the comparison metric but are crucial for the text meaning. In Swedish for example, several characters are only distinguished from others by the dots above the line. This can be addressed by incorporating stronger prior knowledge in the segmentation model. In the case of the modified seam carving method some error also arises from its inability to make complex cuts e.g. when the cut must go straight down. An example is shown in figure 5. For both methods some error also stems from the classification of overlapping characters where no good separation is possible. This is a model error, the methods we have investigated all assume that a pixel belongs to a single line.

6. Acknowledgements

We would like to thank Per Cullhed at the Uppsala University Library for providing source material; Ewert Bengtsson, Mats Dahllöf and Lasses Mårtensson for support and valuable input; Abdelkadir Asi of Ben-Gurion University for his help with understanding the line segmentation method based on seam carving.

References