Segmentation of Postal Envelopes for Address Block Location: an approach based on feature selection in wavelet space

David Menoti, Díbio Leandro Borges, Jacques Facon, Alceu de Souza Britto Jr

Pontifical Catholic University of Parana (PUCPR)
Postgraduate Program in Applied Informatics (PPGIA)
Image Science Group
Rua Imaculada Conceição, 1155 – Prado Velho, Curitiba, PR, Brazil

E-mail: \{menoti, dibio, facon, alceu\}@ppgia.pucpr.br

Abstract

This paper presents a segmentation algorithm based on feature selection in wavelet space. The aim is to automatically separate in postal envelopes the regions related to background, stamps, rubber stamps, and the address blocks. First, a typical image of a postal envelope is decomposed using Mallat algorithm and Haar basis. High frequency channel outputs are analyzed to locate salient points in order to separate the background. A statistical hypothesis test is taken to decide upon more consistent regions in order to clean out some noise left. The selected points are projected back to the original gray level image, where the evidence from the wavelet space is used to start a growing process to include the pixels more likely to belong to the regions of stamps, rubber stamps, and written area. Experiments are run using original postal envelopes from the Brazilian Post Office Agency, and here we report results on 440 images with many different layouts and backgrounds.

1. Introduction

Postal automation has been recently integrated into the research agenda of the pattern recognition and computer vision communities, since acquisition and storage of images of envelopes and parcels has become easier and cheaper than a decade ago. However, segmentation of a typical image of a mail piece into background, stamps, and the address blocks is still a challenging problem due also to the large variety of stamps, backgrounds, written text of the address (e.g. handwritten, printed, locations).

Other works in the literature have tackled different aspects of that problem. A survey in document image understanding up to 1994 can be seen in [3]. In [1] a texture segmentation technique, which organizes the wavelet coefficients of an image into a probabilistic graph is presented. Fusion of that information by Hidden Markov modeling is used to refine segmentation hypotheses. A layout page segmentation is presented in [2], and it is based on local feature extraction by wavelet packets, followed by a soft integration process to vote for layout borders detection. One of the few works we found with results on envelopes is in [4], which presents a method to identify regions in envelope images candidates for being the destination address. The technique is a texture segmentation based on Gabor filters. In [5] a method to locate text areas against different backgrounds is shown, which is based on a pseudo-motion technique to identify oscillations on the wavelet coefficients. An interesting work in text detection in document images such as newspapers, photographs, and magazines is shown in [8], where a texture segmentation module uses gaussian derivative filters followed by a non-linear transformation to produce the feature vectors. A method to locate address blocks on images where an arbitrary layout of printed text is known a priori is presented in [9].

We present here a novel approach for segmentation of an image of a postal envelope, it is a general and robust segmentation method not restricted to a particular layout. The method is divided into 5 steps, where salient features are located in wavelet space, and using local statistics from the data, hypotheses about the classes are built and tested for achieving the final segmentation. Results are shown for a great variety of layouts and backgrounds taken directly from sample images.

The rest of this paper is organized as follows. Section 2 describes the segmentation task for postal automation we address in this work. Section 3 shows our approach for this task, which is based on feature selection on wavelet space. Section 4 shows results from an experimental setup we organized using original postal envelopes with different backgrounds. Section 5 points to the conclusions and future directions for this line of work.
2. The Segmentation task for postal automation

A typical image of a postal envelope will have different backgrounds, stamps in many sizes, rubber stamps from the post office, and the address block, which can be handwritten or printed. All of those information over the background usually can appear in a great variety of locations. An image of a postal envelope is shown in Figure 1(a) The segmentation task to be performed for postal automation would be to fully separate the background, and locate the other regions as stamps, rubber stamps, and more important the address blocks for posterior recognition. From our database of postal envelopes, used under contract with the Brazilian Post Agency, we have prepared a ground truth for further evaluation pixel by pixel of the segmentation. Figure 1(b) shows for the envelope in Figure 1(a) what would be the separation a segmentation is aimed at reaching in the end.

3. The approach

The approach we propose can be divided onto 5 main steps: 1) First, the image \( I \) is decomposed into wavelet space, using Mallat decomposition [6] with Haar basis. This is a non-redundant transform which leads to four output channels of features, being LL, LH, HL, HH; a selection of significant values is put into an image called \( I_W \); 2) Identification of salient points, based on the intersection of the output high frequency channels LH, HL; producing \( I_S \); 3) Constrained labeling of sets of salient points, for noise cleaning and background separation, based on a statistical hypothesis test for sets of points; leading to output \( I_C S \); 4) Back projection of selected salient points to original gray level image, and cleaning of loosely connected sets; which produces \( I_{BPCS} \); 5) Contour following from the selected points based on statistical hypothesis testing; indicated as \( I_{FINAL} \). Figure 2 shows a flowchart of the approach. The rationale of each of the 5 steps of the segmentation algorithm is given in the following sections.

3.1. Wavelet decomposition

A wavelet transform decomposes data into fundamental building blocks. Its basic difference from Fourier decomposition is that the wavelet functions are well localized in time and space, whereas sinusoidal functions used in Fourier transform are not. Since it is possible to design wavelet decompositions with a great variety of basis functions, and also either emphasizing redundancy or eliminating it throughout the levels of decomposition, the literature is plenty of different useful techniques [7]. For our purposes in the segmentation task, a desired decomposition would have to help locate discontinuities in the image most prone to be text, stamps and stamp borders. Mallat decomposition [6] is a decimated scheme which produces as output four sets related to the original image, one for smooth or low frequency data (LL), and three for details, being high frequency with horizontal (LH), vertical (HL), and (HH) diagonal directions. As basis function Haar seemed appropriate for this kind of image application, and it was used. Our first stage then consists of transforming \( I \) into \( I_W \) as a preparation step for the segmentation, using one level of decomposition only. Because of the energy packing [7] property of the wavelet transform it is not necessary to keep all the coefficients, and we keep only the \( \lambda_1 \) % more significant in the normal distribution sense.

3.2. Identification of salient points

The input data for this step is \( I_W \), and our aim here would be to identify evidence for the borders of more consistent regions likely to be either background, text
(address blocks and rubber stamps) or stamp. We call a set of salient points, those points which have strong evidence for being a detail, although those points would only be marks at this time, since there will be other steps further on to check consistency and include more evidence. This identification is done by the intersection of the two vertical and horizontal details channels from \( I_W \), i.e. wherever it is found a presence (point by point) of a horizontal and vertical detail (from LH and HL), that would be a salient point. Diagonal points (HH) are very noisy and do not add much to this procedure, so they were left out.

\[ I_S \leftarrow I_W(LH) \cap I_W(HL) \] (1)

3.3. Constrained labeling of the salient points

The set of salient points \( I_S \) is evidence for texture-like loosely connected components we are trying to segment. Some salient points might appear in a local region with different distribution than the other regions, i.e. basically as a result we want to consider two types of local regions of salient points, one with a high density of presence of salient points (measured in a 8x8 window), and the other with low density. High density ones are more likely to be from connected regions or so, and then the others would be considered coming from noise. Decision of what value is high density, and what is low, can not be made as a fixed one since for each envelope the image pixels and their distribution change at a great extent.

For solving that we have designed a special control algorithm for performing statistical significance tests upon the windows of salient points, and it uses the mean \( \mu \) and standard deviation \( \sigma \) of salient points for each window. The labeling into the two classes of salient regions is constrained at the top by the region with highest mean \( \mu_h \), and at the bottom by the one with lowest mean \( \mu_l \) above zero. If at a significance level \( \lambda_2 \) two windows show no meaningful difference between them equation (2) must hold, i.e.

\[ \left| P_1 - P_2 \right| / \sigma_{P1-P2} \leq Z\lambda_2 \] (2)

Where, \( P_1 \) and \( P_2 \) are the proportions, or means, for windows/regions 1 and 2 respectively; \( \sigma_{P1-P2} \) is the standard deviation of the difference \( P_1 - P_2 \); \( Z\lambda_2 \) is the normalized point for probability at significance level \( \lambda_2 \).

The algorithm starts at both levels, lowest and highest mean, simultaneously, and goes on at deciding and labeling windows/regions either for the new set \( I_{CS} \) or to be left out. If after a run through the image some regions are left unmarked, the values for \( \mu_h \) and \( \mu_l \) are updated considering all the already marked regions, and computing the new means. After this step we are left with the set \( I_{CS} \) which has removed most of the noise and background.

3.4. Back projection of selected points

Since \( I_{CS} \) is an n/16 by m/16 image with the selected salient points, each of its point has four (4) children related to it in the original image \( I \). So, a back projection will create a new image \( I_{BPCS} \) by only picking up the gray levels of the children of \( I_{CS} \), and removing out regions that are not connected to at least one other in the closest radii neighborhood.

3.5. Contour following based on statistical hypothesis testing

At this stage the image \( I_{BPCS} \) is our selected evidence for all the points, i.e. pixels, likely to belong to either address blocks, stamps, or rubber stamps, since the most of the background is its complementary image against the original data. However, this evidence has to be used properly in order to find in the rest of the image only the pixels at gray level that are coherent to the \( I_{BPCS} \) image, and to the idea that local regions must share similar attributes. At this stage we designed the final step of the segmentation, starting with \( I_{BPCS} \), from each set of 2x2 pixels \( (I_{SET(i,j)}, I_{SET(i+1,j)}, I_{SET(i,j+1)}, I_{SET(i+1,j+1)}) \) originated from a salient point a decision about which point would be finally selected is made as follows. The point with the lowest gray level may be selected if the second lowest value falls inside the \( \lambda_3\% \) of the image distribution, i.e.

\[ \text{If } I_{SET(2nd\ lowest)} \leq I_\mu - Z\lambda_3 \% \cdot I_\sigma \quad \text{Then } I_{SET(\ lowest)} \text{ is selected} \] (3)

Where,

- \( I_\mu \) is the global mean of the image;
- \( Z\lambda_3 \% \) is the normalized point for probability of \( \lambda_3 \% \);
- \( I_\sigma \) is the standard deviation of the image.

For deciding upon the other pixels in the image, starting from \( I_{SET(\ lowest)} \), a contour following algorithm includes a neighboring pixel only if the following holds.

Let

\[ \varepsilon = \left| \text{Max} (\text{Mean} (I_{SET(lowest)}, I_{SET(2nd\ lowest)}), I_{NEIGHB}) - I_{NEIGHB} \right| / \text{Max} (\text{Mean} (I_{SET(lowest)}, I_{SET(2nd\ lowest)}), I_{NEIGHB}) \] (4)

\[ \beta = \left| I_{SET(lowest)}, I_{SET(2nd\ lowest)} - I_{NEIGHB} \right| / \text{Max} (I_{SET(lowest)}, I_{NEIGHB}) \] (5)

If \( \beta \leq \varepsilon \) (6)

Then \( I_{NEIGHB} \) is included as part of the contour for the final segmentation, and the same is applied to all the other points updating \( I_{SET} \) as the new included element. If (6) does not hold, the algorithm goes on to other selected points. Function Max() takes the maximum of the values, and Mean() computes the average between the elements. The rationale for this is that we only include the points with evidence, and the thresholds are automatically set from each image. So, the algorithm can run in different
4. Experimental results

For the experiments we have a database with original images of postal envelopes, and we sampled different backgrounds and included for testing.

The tests presented here include 440 images, being 40 originals, each then was synthesized using 11 different backgrounds collected from samples in the database. Figure 3 shows 4 images of envelopes from the 440 testing set used. Figures 4, 5, 6, and 7 shows respectively the output results $I_S, I_{CS}, I_{BPCS}, I_{FINAL}$ for the segmentation with $\lambda_1 = 66\%, \lambda_2 = 90\%, \lambda_3 = 80\%$.

Independently of the layout and background in the input images (Figure 3) it can be seen that the segmentation recovered all the address blocks, stamps, rubber stamps, and background with great success.

Table 1. Average final results with identification of regions (pixel by pixel accuracy) for the images tested

<table>
<thead>
<tr>
<th>Region Class</th>
<th>Accuracy by pixel ($\mu \pm \sigma$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Address Block</td>
<td>85.36% ± 6.49%</td>
</tr>
<tr>
<td>Noise (Background)</td>
<td>0.52% ± 0.33%</td>
</tr>
</tbody>
</table>

Since no further treatment is applied in the resulting images $I_{FINAL}$ such as filtering or closing morphological procedures, those results are significant and promising. Results show that the method is robust to different layouts, sizing of text, and backgrounds.
5. Conclusions and future work

We presented a novel approach for segmentation of postal envelopes that can recover with great accuracy distinctive elements such as the address blocks and their locations. The method is based on using features in wavelet space to identify salient points in the image, and to produce consistent hypotheses about the regions and their information with pixel accuracy. A set of 440 images showing different layouts and backgrounds was tested, and recovered address blocks reached over 85% success rate on average. Future paths for works opened up with this research and results are to model the optimized response curves for parameters in the 5 steps such as significant thresholds for salient points and statistical hypotheses, and to increase even more the database for postal envelopes, evaluate the segmented outputs in handwritten recognition systems for full automatic classification. Those lines of work are being undertaken by our group.

Acknowledgments

We would like to acknowledge support for this research from PUCPR, CAPES/MEC, CNPq/MCT, Fundação Araucária and the Brazilian Post Office Agency (Correios Brasileiros).

References


