Segmentation of Envelopes and Address Block Location by Salient Features and Hypothesis Testing

DAVID MENOTTI[1,3]
DÍBIO LEANDRO BORGES[2]

PUCPR - Pontifícia Universidade Católia do Paraná
PPGIA - Programa de Pós-Graduação em Informática Aplicada
Laboratório de Análise e Reconhecimento de Documentos
Rua Imaculada Conceição, 1155, 80215-901, Curitiba, PR, Brasil
[1]menotti@ppgia.pucpr.br

UnB - Universidade de Brasília
CIC - Departamento de Ciência da Computação
Campus Universitário Darcy Ribeiro, Asa Norte, 70910-900, Brasília, DF, Brasil
[2]dibio@cic.unb.br

UFMG - Universidade Federal de Minas Gerais
DCC - Departamento de Ciência da Computação
Av. Antônio Carlos, 6627, 31270.010, Belo Horizonte, MG, Brasil
[3]menotti@dcc.ufmg.br

Abstract. Although nowadays there are working systems for sorting mail in some constrained ways, segmenting gray level images of envelopes and locating address blocks in them is still a difficult problem. Pattern Recognition research has contributed greatly to this area since the problem concerns feature design, extraction, recognition, and also the image segmentation if one deals with the original gray level images from the beginning. This paper presents a segmentation and address block location 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. Besides the new features and a growing process controlled by the salient points presented here, a fully comprehensive experimental setup was run by separating and classifying blocks in the envelopes, and validating results by a pixel to pixel accuracy measure using a ground truth database of 2200 images with different layouts and backgrounds. Success rate for address block location reached is over 90%.

Keywords: postal automation, segmentation of envelopes, address block location, wavelet features, features design, statistical hypothesis.

(Received June 16, 2006 / Accepted November 03, 2006)

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 [7]. 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 modelling is used to refine segmentation hypotheses. A layout page segmentation is presented in [4], and it is based on local feature extraction by wavelet packets, followed by a soft integration process to vote for layout borders detection. In [9] 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 [21], 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 [23].

One of the earliest works in address location is [22], where they first apply a digital Laplacian operator on an image to separate light and dark regions. The dark connected regions have their properties computed and tabulated in a way that clusters could be formed by looking at a uniform print style for each dark region. Categories considered were destination address, stamps, and return address. Weighting functions were based mostly on the spatial position of the cluster in the envelope image. Results were given for machine printed addresses in a set of images.

[3] presents a feasibility study for OCR classification of envelope addresses over 1500 address images. The technique they proposed for locating addresses is targeted to machine printed addresses only, and they have heuristically chosen values based on the properties of pixel densities, region shape and location as criteria. First, an image is thresholded for binary, then a horizontal smearing procedure is ran to accentuate regions of high pixel density. Subsampling and grouping of 8 connected blocks followed in order to reduce processing and noise. Features such as size, direction, and position of the blocks together with their empirical chosen values were used to classify what it would be a most likely address destination block. Some results on handwritten envelopes were reported and the results were not satisfactory, since the method was constrained to the heuristic geometrical properties found in their feasibility study.

In [8], the authors present a method to identify regions in envelope images candidates for being the destination address. The technique is a texture segmentation based on Gabor filters. Text and non-text are considered as different textures, and a three (3) cluster problem is formulated using the mean and variance of the Gabor filters applied to the envelope image. A connected component analysis is then run to isolate words and pass them to an OCR for identification. No final figure is actually provided to see where and when the system would fail. Problems where written text could cross other areas, as with the rubber stamps for example, would certainly arise in practical situations.

In [13] a dedicated hardware for postal address block location is presented. The system is designed as a blackboard architecture which invokes image processing and block analysis tools in a rule-based order. Collected statistics of geometric features of address blocks, position, different aspects between machine and hand printed blocks are all used to feed the inference engine. A score is computed based on all the evidences collected and the candidate with the highest score is selected as the Destination Address Block (DAB). They report tests on 174 mail pieces with success rate of 81% for the DAB. In [15] the authors describe an end-to-end method for Delivering Point Codes (DPC), which is meant to analyze a handwritten text of an address block in the USA mail system and interpret it. The core of the proposal works as a constrained domain word classifier, in a Hypothesize and Test paradigm. They have evaluated the proposal using 1600 images, and for the ones which the system found an acceptable parse the error rate was 6% on average. Based on entropy and redundancy of the components in [16] the authors have extended their system for automatic interpretation of USA zip codes using the DPC. Examples given in their paper show greater accuracy than presented in the earlier literature.

An address block location method is proposed in [20] for working with both machine and hand printed addresses. The method is based on dividing the input image into blocks where the homogeneity of each block gradient magnitude is measured. Heuristically given

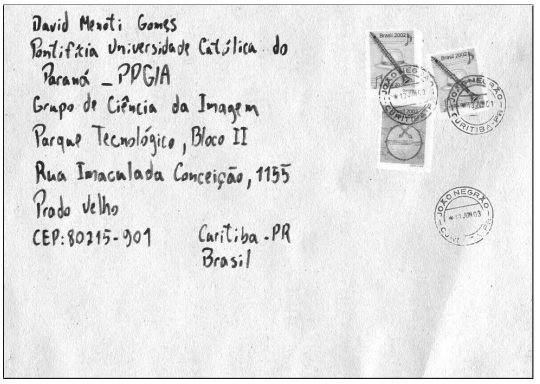
thresholds are used to decide upon the gradient magnitude of a typical address block candidate. In their tests 1600 machine printed addresses and 400 handprinted ones were used, reporting over 91% successful location. The solution appears to work fast for well constrained envelopes, whereby a large separation exists between the image regions since they mentioned a large drawback in the figures if the envelopes have more than one stamp for example.

A clustering technique that works with contour features, i.e. connected components, such as their position, size and width is presented in [6]. Their hypothesis is that a whole address block will belong to just one cluster considering those features. It does work on cleaning and separating address blocks in images where they appear as a more dense region. Address blocks where a zip code, or city destination is written at the lower part of the envelope, or far from the main block seems to be a problem for the technique.

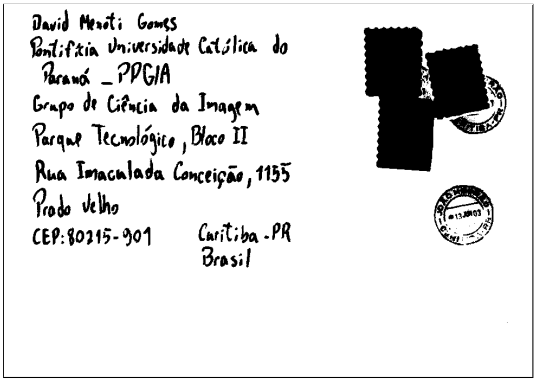
The method we present in this paper is a novel approach for segmentation of an image of a postal envelope, it is a general and robust segmentation method not restricted to a particular layout. Earlier results were presented in [12] and [11]. The method is divided into 5 steps, where salient features 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. In this paper we have extended that approach by modifying the contour growing step, where evidence from the salient points guide the final recovery of the features especially of the address block. Also, in here we run experiments showing the influence of variables of the approach such as window size (i.e. granularity for the parameters estimation) and acceptance probability in the hypothesis testing. The success rate, as it will be shown in the experiments, improved to reach 97% in the address block. Our approach has linear time complexity, i.e. $O(n)$, regarding to the image size. 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 from our database 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 (Correios Brasileiros), 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 expected separation.



(a)



(b)

Figure 1: 1(a) Original image of a postal envelope; 1(b) Aimed segmented image from 1(a), where besides separating the background, knowledge and position of pixels for classes stamp, rubber stamp, and address block ought to be reached.

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 [10] 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 labelling of sets of salient points, for noise cleaning and background separation, based on a statistical hypothesis test for sets of points (depending on the window size k chosen for this step) leading to output I_{CS}; 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.

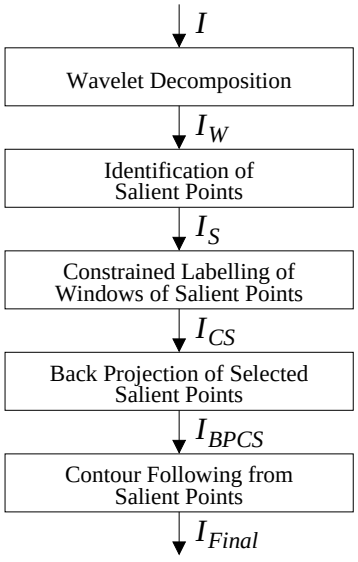


Figure 2: Flowchart of the segmentation approach proposed here, indicating the 5 main steps and their respective outputs. I has dimensions $n \times m$, I_W $n/2 \times m/2$, I_S $n/2 \times m/2$, I_{CS} $n/2k \times m/2k$. I_{BPCS} $n \times m$, and I_{FINAL} $n \times m$.

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 [14][18].

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 [10] 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 property of the wavelet transform [17] it is not necessary to keep all the coefficients depending on the task, i.e. compression, reconstruction, or as it is our case identification and segmentation. Denoising algorithms [2], or especially designed supervised classification schemes [5] would be alternatives for selecting a subset of features for performing further the segmentation. It is our hypothesis in this work that a significance test would be able to select from the wavelet decomposition a suitable and small set of features for segmentation. We designed a test that keep only the $\lambda_1\%$ more significant in the normal distribution sense. Figure 3.1 shows this hypothesis in a typical normal (i.e. Gaussian) distribution, where the $\lambda_1\%$ are kept in this step as the most significant features.

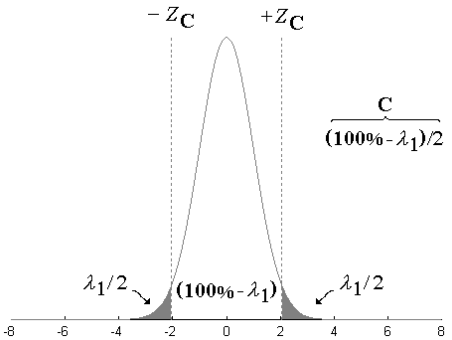


Figure 3: Hypothesis test showing $\lambda_1\%$ coefficients that are kept at this step.

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) in some situations could be also part of these saliency selection, although for the postal automation task they were noisy and did not add much to this procedure, so they were left out. Equation 1 indicates I_S as used in this work.

$$I_S \leftarrow I_{W(LH)} \cap I_{W(HL)} \tag{1}$$

3.3 Constrained labelling 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 $k \times k$ window, k was tested for 4, 8, and 16), 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. Experiments were run for different window sizes over the whole database in order to choose one, and the results are shown further in this paper. For solving the problem of separating evidence more likely to be from background, or the other parts to be segmented, we have designed a special control algorithm for performing statistical significance tests upon the windows of salient points, and it uses the mean μ and standard deviation σ of salient points for each window. The labelling into the two classes of salient regions is constrained at the top by the region with highest mean μ_h, and at the bottom by the one with lowest mean μ_l above zero. If at a significance level $\lambda_2\%$ two windows show no meaningful difference between them Equation 2 must hold, i.e.

$$\frac{|P_1 - P_2|}{\sigma_{(P_1-P_2)}} \leq Z_{\lambda_2\%} \tag{2}$$

where, P_1 and P_2 are the proportions, or means, for windows/regions 1 and 2 respectively; $\sigma_{(P_1-P_2)}$ 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 labelling 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 μ_h and μ_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 $n/2k$ by $m/2k$ 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. Since our intention at this point is to use only the best evidence found from the salient points and propagate it to the original data for finding the contours which support those regions.

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. The final step of the segmentation, starting with I_{BPCS} , from each set of 2×2 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 as follows. The point with the lowest gray level, $I_{SET(lowest)}$, may be selected if the highest value, $I_{SET(highest)}$, falls inside the $\lambda_3\%$ of the image distribution, i.e.

$$I_{SET(highest)} \leq I_{\mu} - Z_{50\%-\lambda_3} \tag{3}$$

If equation 3 holds then $I_{SET(lowest)}$ is selected to be part of the contour. Where, I_μ is the global mean of the image; $Z_{50\%-\lambda_3}$ is the normalized point for probability of $50\% - \lambda_3$; I_σ is the standard deviation of the image. Figure 4 shows those variables in a general distribution for more clarity.

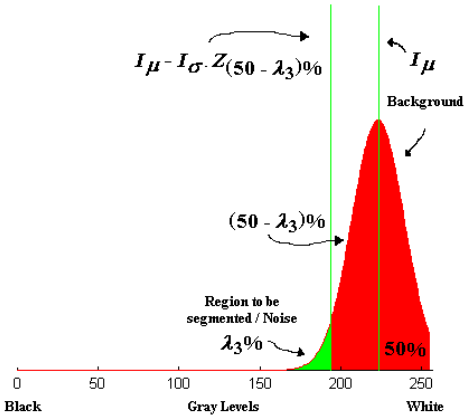


Figure 4: Schematic drawing showing the variables used and their meaning in this step.

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

$$\varepsilon = \frac{|I_{SET(highest)} - I_{neighbour}|}{Max(I_{SET(highest)}, I_{neighbour})}$$

$$\beta = \frac{|I_{SET(lowest)} - I_{neighbour}|}{Max(I_{SET(lowest)}, I_{neighbour})}$$

$$(\beta \leq \varepsilon) \wedge (I_{neighbour} \leq I_\mu - Z_{50\%-\lambda_3} \cdot I_\sigma) \quad (4)$$

If Equation 4 holds then $I_{neighbour}$ 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 Equation 4 does not hold, the algorithm goes on to other selected points. Functions $Max()$ takes the maximum of the value 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 images and backgrounds without any manual intervention.

4 Post-Processing for Classifying the Candidate Segmented Blocks

At this stage the images are already segmented from all the kinds of different backgrounds, and for the full address block location it is necessary to classify the candidate segmented blocks into either stamps, rubber stamps, or the address block itself. We call this stage post-processing. This process is divided into 5 main steps: 1) Getting the size of run-length automatically; 2) Creating bounding boxes in the segmented image using the image of run-length process; 3) Merging boxes that possibly belong to the same lines; 4) Classifying the boxes in order to separate address block boxes from the others. 5) Verifying if the zip code of the postal envelope belong to one of the identified address boxes. Figure 5 shows a flowchart of this post-processing stage.

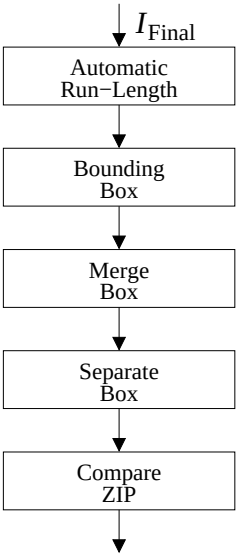


Figure 5: Flowchart of the post-processing, indicating the 5 main steps.

4.1 Automatic Run-length

Firstly, the different regions segmented from the background from the last stage would need to be grouped automatically for consideration of text, handwritten or typewritten, geometrical contiguous figures, and so on, independently from the resolution of the image. A known algorithm for helping with this is run-length [19], and in here we have designed an automatic run-length for the postal envelopes by adjusting horizontal and vertical lengths in the proportion 1:3 (vertical:horizontal). This is invariant to rotation, and although slanted boxes could also be found and corrected

it was not necessary to implement this feature in the automatic run-length algorithm here.

The algorithm consists of an iterative process in order to find a parameter for a run-length to merge nearby elements, and consequently delimiting their bounding boxes. In each iteration the run-length algorithm is run using an aspect ratio of 1:3 (vertical:horizontal), which was set empirically by analyzing the envelopes database. After the run-length algorithm is completed, a labelling process is performed over the binary image, and the number of labels found is saved for further use. This iterative process stops whenever the condition in Equation 6 holds.

$$Ratio \leftarrow \frac{|nLabLast - nLabCurr|}{|nRLLast - nRLCurr|} \quad (5)$$

$$Ratio >= Threshold \quad (6)$$

where $nLabCurr$ and $nLabLast$ represent the number of labels respectively of current and last iterations for $nRLCurr$ and $nRLLast$ run-length parameters. Algorithm 1 shows a pseudocode of it.

Algorithm 1 Automatic Run-Length

$nRLCurr \leftarrow 0$;
$nLabelCurr \leftarrow LabelProcess(ImgIn, 0)$;
repeat
  $nRLLast \leftarrow nRLCurr$;
  $nLabelLast \leftarrow nRLCurr$;
  $nRLCurr \leftarrow nRLCurr + RLstep$;
  $ImgOut \leftarrow RunLength(ImgIn, nRLCurr)$;
  $nLabelCurr \leftarrow LabelProcess(ImgOut, nRLCurr * fH * fV)$;
  $Ratio \leftarrow \frac{|nLabLast - nLabCurr|}{|nRLLast - nRLCurr|}$;
until $Ratio < Threshold$
$nRL \leftarrow nRLLast$
return (nRL)

In our experiments the parameter $RLstep$ was set as 2, and fH and fV, the horizontal and vertical factors, were set to 3 and 1, respectively.

4.2 Bounding Box

In this step, bounding boxes are found in the image generated by a run-length parameter nRL, obtained in the last stage with the labelling process. Small components are removed based on their dimensions and parameters used in the first step (i.e. $Height \times Width < nRL^2 \times fH \times fV$). This rule was generated under the hypothesis that boxes lower than $nRL^2 \times fH \times fV$ are isolated objects or noise, which would not belong to the aimed segmented objects (address block, stamps and rubber stamps).

4.3 Merge Box

In order to generate less, but significant components, words in the same lines are merged into single components. This is achieved by sorting the bounding boxes by $Height$ (in descending order) such that small boxes (with small height) could be merged first. Thus, avoiding that larger boxes merge with small boxes, and more small boxes were left out. The algorithm developed for this is shown in Algorithm 2.

Algorithm 2 Merge Box

for each box l **do**
  for each box m of $l + 1$ up to the end **do**
    if Intersection between l and m **then**
      merge them in only one
      remove l and m from list
      put the new box in its sorted position
    end if
  end for
end for

Algorithm 3 Intersection

if $IntersecV \geq 40\%$ and $IntersecNV \leq 50\%$ and $max(Heightl, Heightm) < \mu_H + 2 * \sigma_H$ and $DistH < 2 * (Heightl, Heightm)/2$ **then**
  true
else
  false;
end if

The function $Intersection$ is showed in Algorithm 3 and $IntersecV$, $IntersecNV$ and $DistH$ are calculated as showed on Figure 6. $DistH$ is defined as the horizontal distance between two boxes. $IntersecV$ and $IntersecNV$ are defined as Vertical and Not Vertical Intersection between two boxes, and are calculated as:

$$IntersecV = n_V/d_V$$
$$IntersevNV = n_{NV}/d_{NV}$$
$$d_V = min(d_l, d_m)$$
$$n_{NV} = d_{NV} - n_V$$

where $DistH$, n_V, d_l, d_m and d_{NV} are computed as indicated in Figure 6.

Some example boxes are shown in Figure 7.

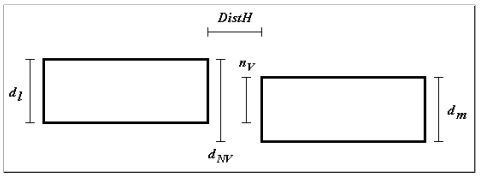


Figure 6: Measures between two boxes used in Merge Box step.

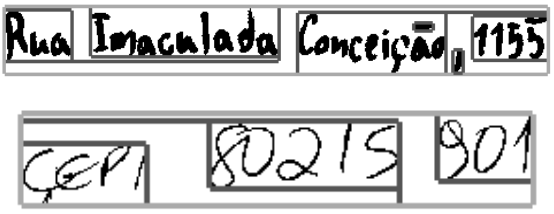


Figure 7: Some snapshots on images showing the merge step between box.

4.4 Separate Box

At this stage all the significative boxes are labelled and decision upon which classes they belong to is needed. For such an aim we have designed a set of rules based on $Density$ (i.e. number of pixels inside bounding box), $Area$ ($Height \times Width$) and $AspectRatio$ (i.e. $Height/Width$), as shown in Algorithms 4 and 5, where fD, fR and fA, are computed as given in Algorithms 6, 7, 8. Thus, a box is not classified as possible address block candidate if any of the rules given in Algorithms 4 and 5 is verified.

Algorithm 4 Discard Rule 1

if $Density \leq Dmin$ or $Ratio \geq Rmax$ **then**
the box is discarded
end if

Algorithm 5 Discard Rule 2

if $Density \geq fD$ and $AspectRatio \leq fR$ and $Area \geq fA$ **then**
the box is discarded.
end if

where the values are set empirically as $Rmin = fH/2$, $Rmax = fH$, $Rct = 7.0$, $Rshift = 3.0$, $Rrange = Rmax - Rmin$, $Dmin = 0.15$, $Dmax = 0.40$, $DShift = 0.15$, $DCt = 0.15$, $Drange = Dmax - Dmin$.

Algorithm 6 Density - fD

if $AspectRatio \leq Rmax$ **then**
$fD = Dshift$
else if $Rmin < AspectRatio < Rmax$ **then**
$fD = Dct/2 * (0.5) * atanh(1/Rvar * (Ratio - Rmax)) + Dshift$
else
$fD = 0$
end if

Algorithm 7 Aspect Ratio - fR

if $Density \leq Dmax$ **then**
$fR = \infty$
else if $Dmin < Density < Dmax$ **then**
$fR = Rct/4 * (0.5) * atanh(1/Dvar * (Density - Dmin)) + Rhift$
else
$fR = Rshift$
end if

Algorithm 8 Area - fA

if $Density \geq Dmax$ or $AspectRatio \leq Rmin$ **then**
$fA = 0$
else
$fA = \mu_{Area}$
end if

4.5 Compare ZIP

According to the variables and parameters set on the other steps of the post-processing, the final stage for providing a zip candidate consists of testing if the ZIP box have area intersection greater than a Threshold (90%) with any other selected box. If this condition holds then ZIP is considered to be found.

Next section provides explanation and details of the experiments run for a large and original database of envelope images, and consideration upon the algorithms and solutions proposed here will be taken forward.

4.6 Time complexity

Here we present a time complexity analysis of the post-processing stage. The iterative Automatic Run-Length step has time complexity (TC) $O(c.n)$, where c is the number of iterations executed and n is the size image; The Merge Box step has TC $O(nRL^2)$; The Bounding and Separate Box steps are linear with the number of labels, i.e. they has TC $O(nRL)$; The Compare ZIP step is constant, i.e. $O(1)$. That is, it required a single comparison.

5 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 2200 images, being 200 originals, plus 2000 that were synthesized using 10 different backgrounds collected from samples in the database.

In order to find suitable values upon our tests for the variables k (size of window from I_S), and λ_2, which are selective choices in the method, we have run experiments varying k for 4, 8, and 16, and λ_2 for 80%, 90%, and 95%. We comment on the best values found for all sets in the following.

Figure 8 shows four images of envelopes from the 2200 testing set used. Changes in background due to illumination or handling affect the image information for the segmentation algorithm. Also sizing of letters, stamps, positions of those features vary at great extent. Despite these the segmentation would have to tackle those problems. Figure 9 shows the images I_S related to the originals shown here, and $\lambda_1 = 58\%$. The rationale for having at this size is to include more candidates for saliency at this stage, since the other steps in the algorithm will be more selective.

Figure 10 brings the I_{CS} images enlarged for better visualization, and it can be seen that even with the second hypothesis testing being more restrict ($\lambda_2 = 80\%$) only strong evidence is left, and the more loosely salient points will be further cleaned at the next stage.

Figure 11 shows only two images of I_{BPCS} because they are harder to visualize at this resolution, since they have few points as constrained salient points. This step cleaned almost completely the weak evidence (i.e. not connected salient points), and it did not loose the important points as the final step will show.

Figure 12 shows four images of a small region of an envelope as sequential results starting from a I_{BPCS} and ending at I_{FINAL}. It shows clearly that the salient points recovered from I_{BPCS} are good and enough evidence for finding the important elements in the end. The images are in gray level, since it is how the algorithm works in the final step.

Independently of the layout and background in the input images (Figure 8) it can be seen that the segmentation recovered all the address blocks, stamps, rubber stamps, and background with great success. Figure 14 shows final results recovered by the algorithm for the 4 images as example inputs. Instead of evaluating this output with bounding boxes, or submitting them to a recognition algorithm, we have built ground truth images (pixel by pixel) for all the 2200 in our test set. Figure 13 shows how this test is performed envelope by envelope. Table 1 gives the average success accuracy considered pixel by pixel for the address blocks, stamp, rubber stamp, and the noise left in the background. The most important figures are the ones related to the address block and noise, since for practical purposes the classes of stamp, and rubber stamp would be discarded in the end and it would not harm the result. The stamp accuracy is actually low in the end because the algorithm works tuned mostly to the address block. Best values for the window size k is 8, and for $\lambda_2 = 80\%$ (Table 1).

Results achieved success rate over 97% (Table 1) for $k = 8$, $\lambda_1 = 58\%$, $\lambda_2 = 80\%$, and $\lambda_3 = 80\%$. Since no further treatment is applied in the resulting images I_{FINAL} such as filtering or closuring morphological procedures, those results are significant and promising.

One of the most important elements in the address block is the zip code itself, since it usually gives a detailed geographic location for the address. We say usually because the zip code has different coding schemes and levels of detail dependent of each country. As a post-processing stage we ran a zip code identification block for the original images we have in our database. Brazilian zip codes use 8 numerals (i.e. digits) and no letters, where the first five are separated by the other three commonly by a sign "−". Figure 5 shows a flowchart of this post-processing. Zip code block confirmation/identification were successful in 90.72% of the cases, using the rules developed in Section IV here and without use of OCR or digit by digit recognition, so it confirms to us that the solution is promising for industry consideration and deployment.

Table 1: Average results with identification of regions (pixel by pixel accuracy) for the images tested, $\lambda_2 = 80\%$.

Region Class $k = 8$	Accuracy pixel by pixel ($\mu \pm \sigma$)
Address Block	97.72% ± 4.86%
Stamp	32.34% ± 20.72%
Rubber Stamp	92.42% ± 16.00%
Noise (Background)	0.16% ± 0.40%

6 Conclusions and future work

In this paper we have shown a novel approach for segmentation of postal envelopes and address block location that can recover with great accuracy distinctive elements such as the address blocks and their locations. A post-processing stage identifies in the address blocks

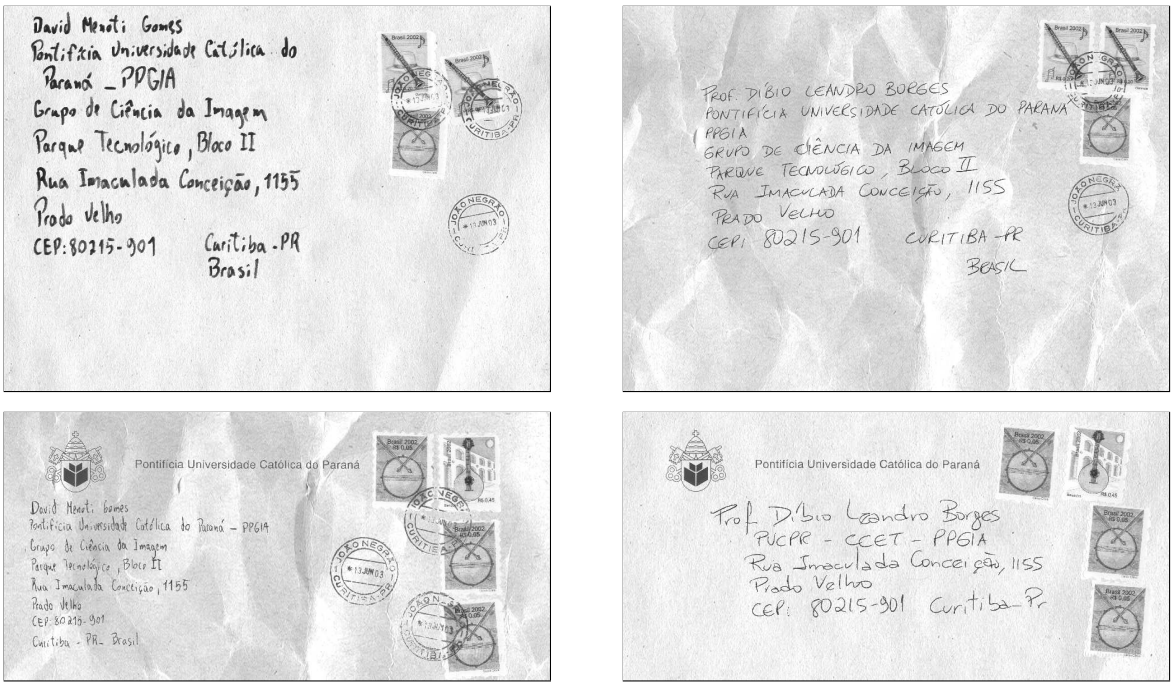


Figure 8: Four (4) different images of postal envelopes used in the experiments, taking from a total of 2200 images.

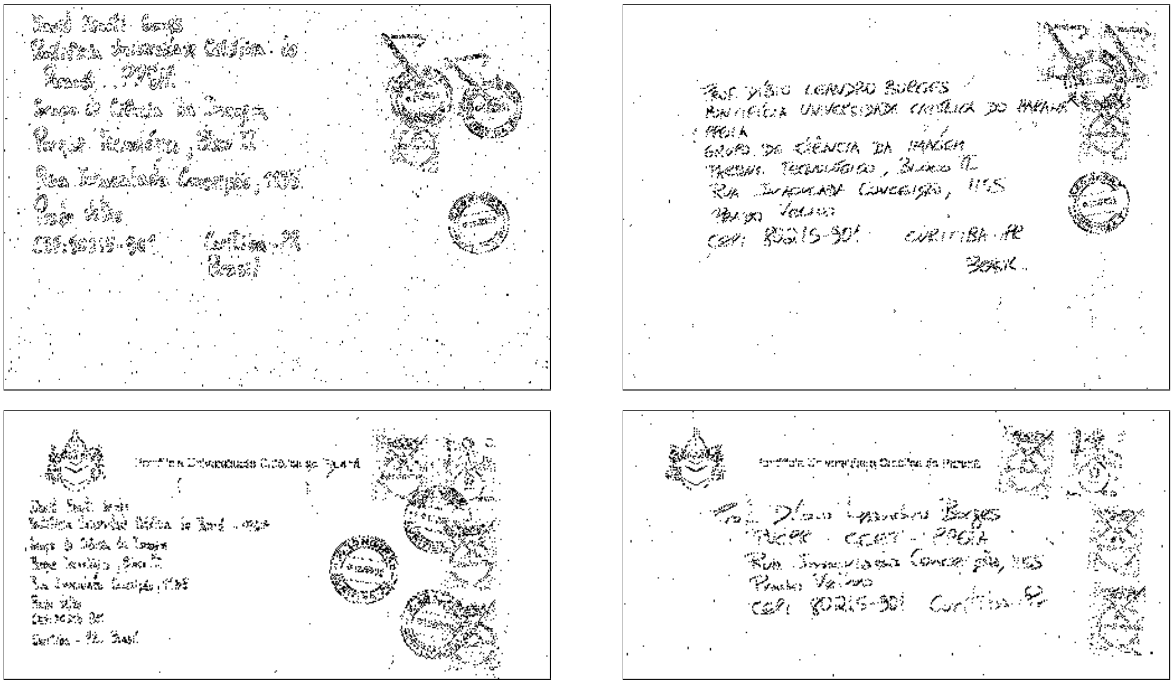


Figure 9: Images I_S of salient points for 4 different envelopes with $\lambda_1 = 58\%$.

with 90.72% success the zip code blocks only by the geometric and heuristic consistency, and it delivers handwritten blocks labelled as zip codes to a possible OCR for industry application further. Although those figures

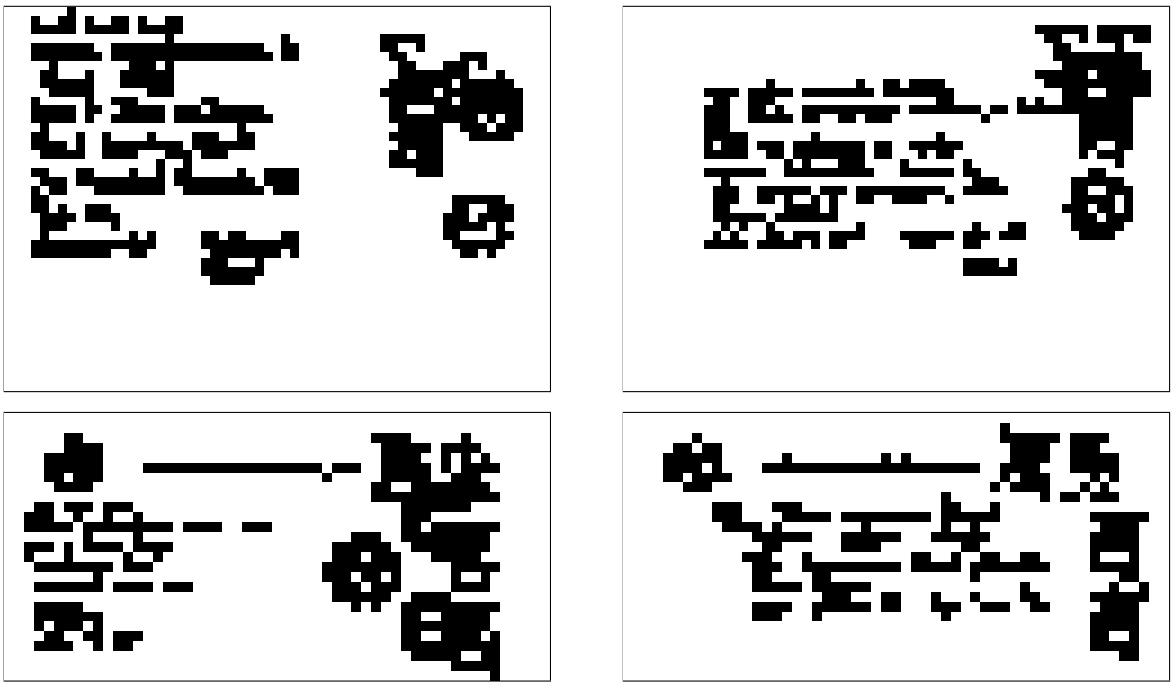


Figure 10: Images I_{CS} of constrained salient points for 4 different envelopes with $\lambda_2 = 80\%$.

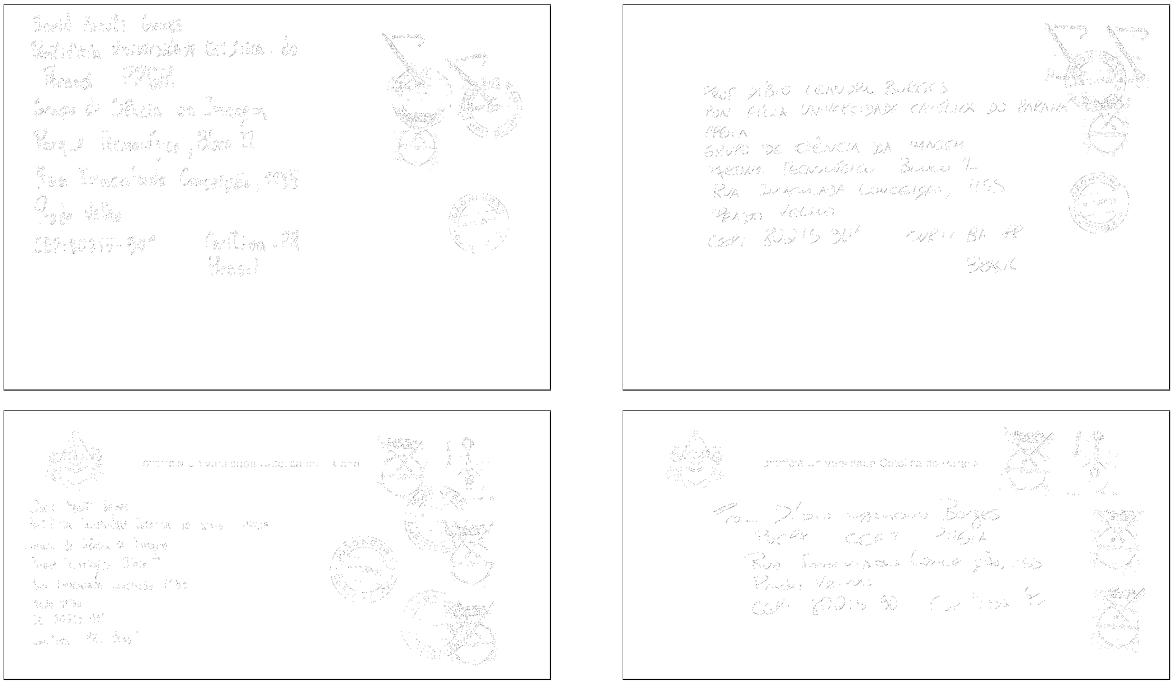


Figure 11: Images I_{BPCS} of the projected constrained salient points for 4 different envelopes.

could be considered with room for improvement, and of course it should be possible, we want to point out that the database is made of unstructured and with a great variety of difficult images, and indeed identifying more

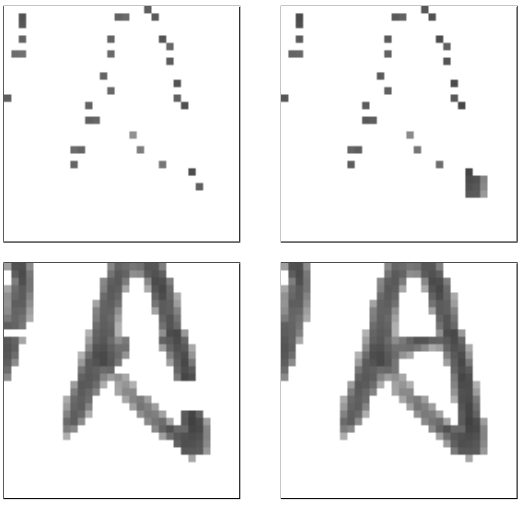


Figure 12: A sequence of a region of an image, showing at greater resolution results coming from the contour following algorithm started with I_{BPCS} (top left image), and ending with I_{FINAL} (bottom right image).

than 90% successful zips cuts the complexity and time for further processing of a practical system a lot.

Our claim here is twofold, as a proposal of a novel segmentation approach and as a practical and well tested system for use in postal automation. 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 2200 images showing different layouts and backgrounds was tested, and recovered address blocks reached over 97% success rate on average. A trade-off response curve is expected considering the parameters in the 5 steps, since the method works from first detecting evidence for features as salient points in wavelet space and in the end decides upon consistency and includes final points in gray level image space.

Future paths for works opened up with this research and results are to deploy such a solution in a specially designed hardware for this purpose, and to evaluate extension for different parcels and postal envelopes. Those lines of work are being undertaken by our group.

Acknowledgments

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

References

[1] Choi, H. and Baraniuk, R. Multiscale image segmentation using wavelet-domain hidden markov models. *IEEE Transactions on Image Processing*, 10(9):1309–1321, 2001.

[2] Donoho, D. De-noising by soft thresholding. *IEEE Transactions on Information Theory*, 41:613–627, 1995.

[3] Downton, A. and Leedham, C. Preprocessing and presorting of envelope images for automatic sorting using ocr. *Pattern Recognition*, 23(3):347–362, 1990.

[4] Etemad, K., Doermann, D., and Chellappa, R. Page segmentation using decision integration and wavelet packets. In *IEEE International Conference on Pattern Recognition*, pages 345–349, Jerusalem, Israel, 1994.

[5] Ferreira, C. B. R. and Borges, D. L. Analysis of mammogram classification using a wavelet transform decomposition. *Pattern Recognition Letters*, 24:973–982, 2003.

[6] Govindaraju, V. and Tulyakov, S. Postal address block location by contour clustering. In *ICDAR'2003 IEEE International Conference on Document Analysis and Recognition*, pages 1126–1130, Edinburgh, UK, 2003.

[7] Haralick, R. M. Document image understanding: Geometric and logical layout. In *CVPR94: IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 385–390, 1994.

[8] Jain, A. and Bhattacharjee, S. Address block location on envelopes using gabor filters. *Pattern Recognition*, 25(12):1459–1477, 1992.

[9] Jin, N. and Tang, Y. Y. Text area localization under complex-background using wavelet decomposition. In *ICDAR'2001 IEEE International Conference on Document Analysis and Recognition*, pages 1126–1130, Seattle, Washington, Sept. 2001.

[10] Mallat, S. A theory of multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Machine Intell.*, 11(7):674–693, 1989.

[11] Menotti, D., Borges, D. L., Facon, J., and de Souza Britto-Jr, A. Salient features and hypothesis testing: Evaluating a novel approach

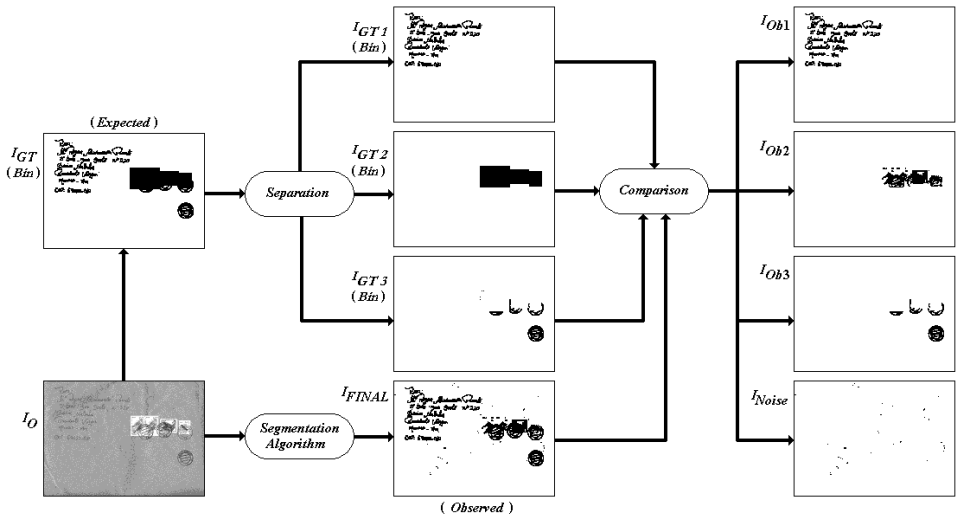


Figure 13: Ground truth images and pixel to pixel testing scheme for the postal images.

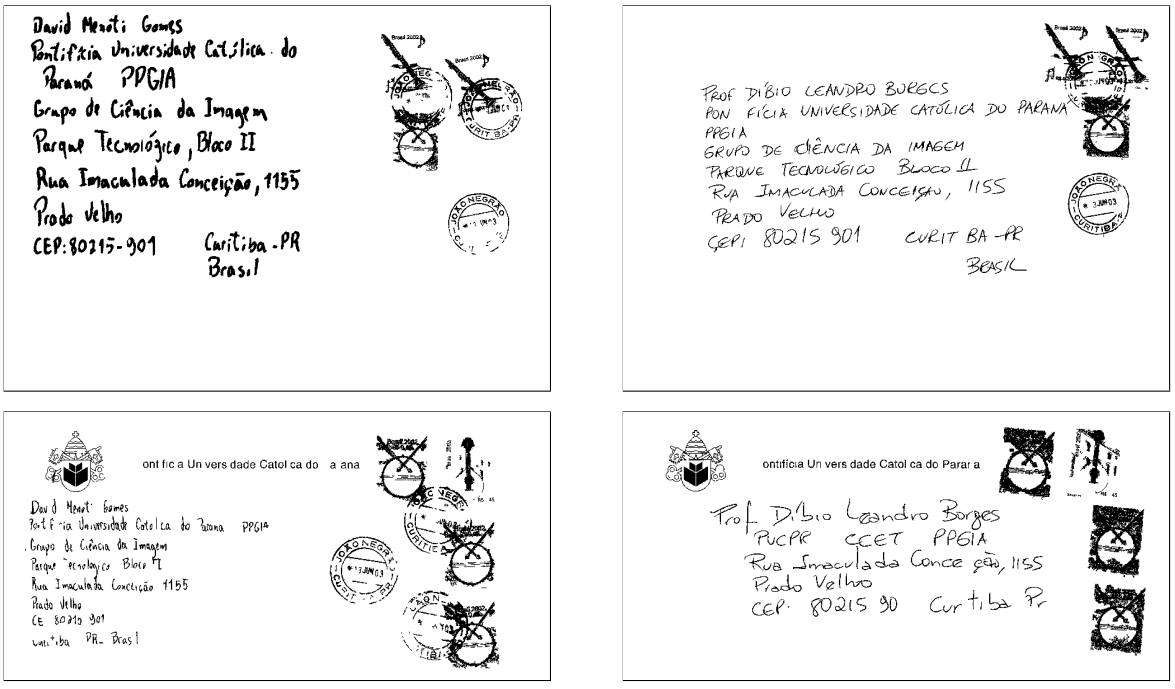


Figure 14: Final images I_{FINAL} with reconstructed address blocks, stamps, and rubber stamps with background removed, for 4 different envelopes, with $\lambda_3 = 80\%$.

for segmentation and address block location. In *DIAR'2003 IEEE Document Image Analysis and Retrieval Workshop, part of CVPR'2003*, pages 41–48, Madison, USA, 2003.

[12] Menotti, D., Borges, D. L., Facon, J., and de Souza Britto-Jr, A. Segmentation of postal envelopes for address block location: an approach based on feature selection in wavelet space. In *ICDAR'2003 IEEE International Conference on Document Analysis and Recognition*, pages 699–703, Edinburgh, Scotland, 2003.

[13] Palumbo, P. W., Srihari, S. N., Soh, J., Sridhar, R.,

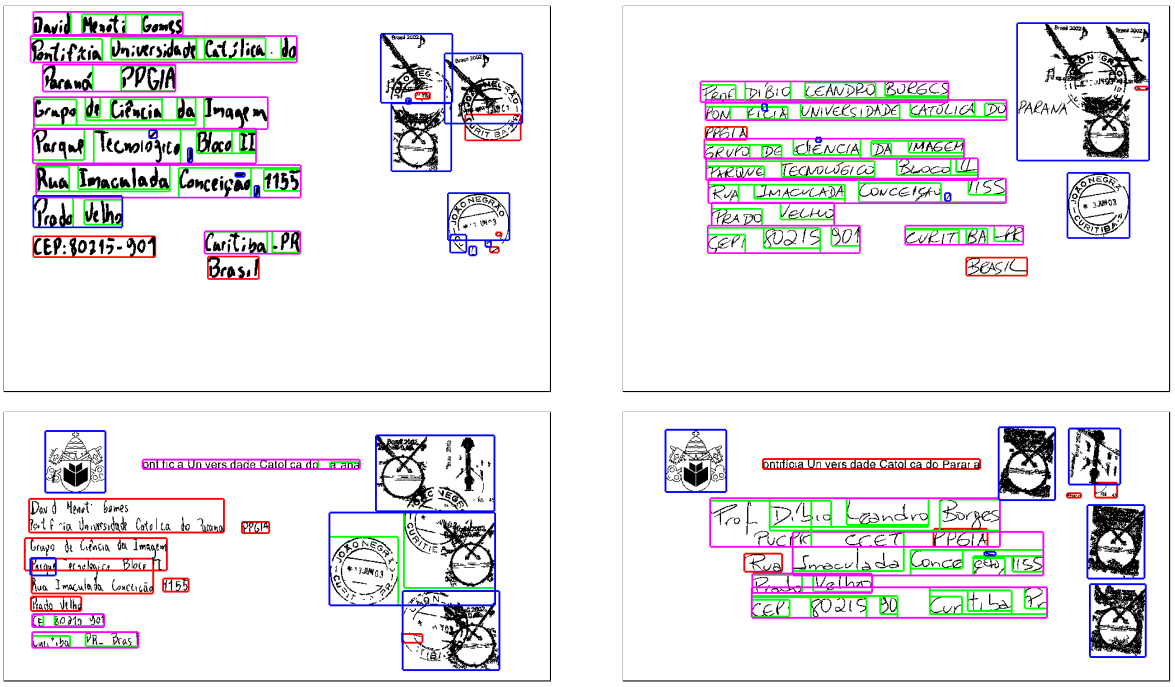


Figure 15: Four (4) different images of postal envelopes used in the experiments.

and Demjanenko, V. Postal address block location in real time. *IEEE Computer*, 25(7):34–42, July 1992.

[14] Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. *Numerical Recipes in C*. Cambridge University Press, UK, 2nd edition, 1996.

[15] Srihari, S., Govindaraju, V., and Shekhawat, A. Interpretation of handwritten addresses in US mailstream. In *ICDAR'1993 IEEE International Conference on Document Analysis and Recognition*, pages 291–294, Tsukuba, Japan, august 1993.

[16] Srihari, S., Yang, W., and Govindaraju, V. Information theoretic analysis of postal address fields for automatic address interpretation. In *ICDAR'1999 IEEE International Conference on Document Analysis and Recognition*, pages 309–312, Bangalore, India., august 1999.

[17] Starck, J., Murtagh, F., and Bijaoui, A. *Image and Data Analysis: the Multiscale Approach*. Cambridge University Press, UK, 1996.

[18] Strang, G. and Nguyen, T. *Wavelets and Filter Banks*. Wellesley-Cambridge Press, Massachusetts, MA, 1997.

[19] Wang, C., Palumbo, P., and Srihari, S. Performance of a system to locate address blocks on mail pieces. In *AAAI National Conference on Artificial Intelligence*, pages 837–843, 1988.

[20] Wolf, M., Niemann, H., and Schmidt, W. Fast address block location on handwritten and machine printed mail-piece images. In *ICDAR'1997 IEEE International Conference on Document Analysis and Recognition*, pages 753–757, Ulm, Germany, Aug. 1997.

[21] Wu, V., Mammatha, R., and Riseman, E. Automatic text detection and recognition. In *DARPA Image Understanding Workshop*, pages 707–702, 1997.

[22] Yeh, P., Antoy, S., Litcher, A., and Rosenfeld, A. Address location on envelopes. *Pattern Recognition*, 20(2):213–227, 1987.

[23] Yu, B., Jain, A. K., and Mohiuddin, M. Address block location on complex mail pieces. In *ICDAR'1997 IEEE International Conference on Document Analysis and Recognition*, pages 897–901, Ulm, Germany, Aug. 1997.