Learning-Based Approach for Postal Envelope Address Block Segmentation

Horacio A. Legal-Ayala
Universidad Nacional de Asunción, Facultad Politécnica,
San Lorenzo, Paraguay
hlegal@pol.una.py

Jacques Facon
Pontifícia Universidade Católica do Paraná,
Programa de Pós-Graduação em Informática Aplicada,
Curitiba, Brazil, 80215-901
facon@ppgia.pucpr.br

and

Benjamín Barán
Universidad Nacional de Asunción, Facultad Politécnica,
San Lorenzo, Paraguay
bbaran@pol.una.py

Abstract

Although nowadays there are working systems for sorting mail in some constraint ways, segmenting gray level images of envelopes is still a difficult problem. This paper presents a learning-based approach for address block segmentation in postal envelopes. The method consists in learning to segment correctly submitting both an image and its ideal version. From this stage, for each pixel and each gray level, it is generated a feature vector and built a classification array, re-utilized at the moment of new image segmentation. Features used in this approach are: pixel gray level, mean, variance, skewness and kurtosis calculated within an adaptive square neighbor. The new image is segmented by means of k-Nearest Neighbor that seeks, for each pixel, the best solution in the classification array. Performed tests were validated using a pixel-to-pixel accuracy measure over a ground truth database of 200 images of complex postal envelopes. Results were compared to other prior approaches for the same database. Success rates achieved for address block (over 98%), stamps, rubber stamps and noise suggest that the features used in the proposed approach improves segmentation results.

Keywords: Artificial Intelligence, Computational Learning, Image Segmentation, k-NN.

1 Introduction

Postal automation is an area of increasing research in the last years, since image acquisition and storage has become easier and cheaper than a decade ago. Also, the evolution of computer processing units made possible the implementation of more complex algorithms. However, location of address blocks in postal envelopes still remains a challenging problem in any image analysis system development for postal automation.

The Brazilian Post Office Agency suggests some rules to be followed for the correct filling of a postal envelope:

- Stamps and rubber stamps must occupy the upper right side of the envelope.
- The address block must be filled in such a way that it does not superimpose to the stamps and rubber stamps.

In several cases the above rules are not followed, impairing any segmentation technique based on the relative position of the composing classes of a Brazilian postal envelope. Besides these practical challenging situations, other problems increase the difficulty in developing efficient postal image segmentation systems, such as:

- Imperfections of the envelopes themselves, due to handling.
- Presence of different kind of drawings, making the background of the envelopes much more complex.
- Problems in acquisition, generating additional borders that do not belong to the postal envelopes.

Several authors have dealt with the problem of locating address blocks in envelopes images. One of the earliest works in address location [11] first applies a digital Laplacian operator on an image to separate light and dark regions. In [3], Jain and Bhattacharjee apply a texture segmentation based on Gabor filters to identify regions in envelope image candidates for being the destination address. In [9] 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. They report tests on 174 mail pieces with success rates of 81% for the Destination Address Block. In [10], another address block location method for working with both machine and hand printed addresses is proposed. The method is based on dividing the input image into blocks where the homogeneity of each block gradient magnitude is measured. In their tests, 1600 machine printed address and 400 hand printed ones were used, reporting over 91% successful location.

Recently, postal envelope segmentation was presented [12], combining a 2-D histogram and morphological clustering. The proposed clustering is based on the watershed transform and morphological grayscale reconstruction filtering. Experiments on a database composed of complex postal envelopes images, with creased background and no fixed position for the handwritten address blocks, postmarks and stamps showed that the method was successful in locating the correct address block in 75%.

In [2], an address block segmentation approach based on fractal dimension was proposed. After computing the local fractal dimension using overlapping square windows of length r, a clustering technique based on K-means was used to label pixels into semantic objects. The evaluation of the efficiency is carried out from a total of 200 postal envelopes images with no fixed position for the address block, postmark and stamp. A ground-truth strategy was used to achieve an objective comparison. Experiments reached a success rate over 97% on average.

In [8], another approach to segment address block location based on feature selection in wavelet space was presented. They run an experimental setup by separating and classifying blocks in the envelopes, and validating results by a pixel to pixel accuracy measure using a ground truth database of 200 original images with different layouts and backgrounds. Success rate for address block location was over 97%.

The method we present in this paper is an approach through computational learning strategy for postal image segmentation; it is a general and robust segmentation method not restricted to a particular envelope layout. The main idea is to use and take advantage of the user's knowledge and his expertise to segment a gray level image into several classes. The method consists of two stages, learning and segmentation. In the first stage, a sample image and its ideal segmented image is submitted in order to extract the relevant features around each pixel neighborhood. The term "ideal image" means the expected image generated by an expert. From this step, one feature vector is generated each gray level. The association of this feature vector set with its respective ideal output is named the *classification array.*

The second stage consists in segmenting new images by seeking the best solution for each pixel in the classification array.

In this paper we have extended an early brain resonance magnetic image (RMI), signature and fingerprint image segmentation approach [5, 6] by modifying and optimizing the best set of feature selection and optimizing the best choice of k parameter into Nearest Neighbor strategy. Also, the comparison of the success rate of this learning-based approach to other methods testes over the same database [2, 8, 12] was achieved.

The rest of this paper is organized as follows. Section 2 details our learning-based approach. Section 3 shows results from an experimental setup we organized using original envelopes with different backgrounds and comparing the success rates achieved by different approaches over the same database. Section 4 points to the conclusions and future work.

2 The Approach

2.1. Fundamentals of the Approach and Mathematical Definitions

A digital image I_x is defined in the context of this work as a discrete conjunct of numbers corresponding to each pixel that compounds this image. A gray level image is formed by a discrete conjunct of numbers (pixels) x_i in the interval [0, ... , 255] *(1)*. A binary segmented image is formed by a discrete conjunct of pixels of only two values {0; 255} *(2)*. Thus, we can define:

Gray level image: $x_i \in [0, \ldots, 255]$ *(1)*

Binary segmented image: $x_i \in \{0; 255\}$ *(2)*

being: $x_i=0$, black level; $x_i=255$, white level; and $0<x_i<255$ gray level pixels

N: total number of pixels of a region (image or neighbor).

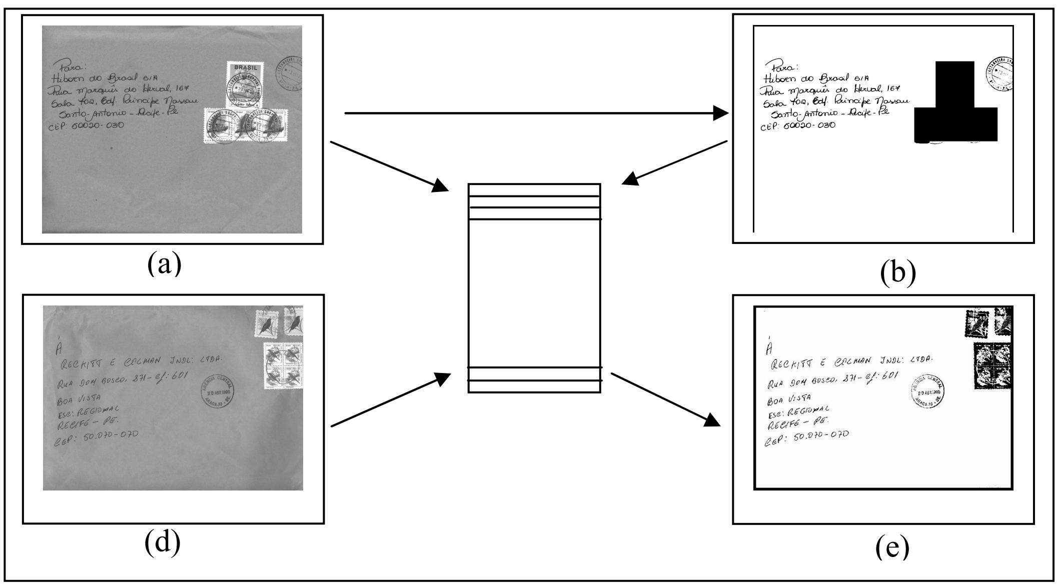


Figure 1. General scheme of the learning-based segmentation approach. (a) Simple input image, (b) Ideal output image, (c) Decision array, (d) New submitted image, (e) New segmented image

A general scheme of the learning-based segmentation approach is depicted in Figure 1. First, a gray level sample input image (a) along with its expected ideal solution (b) is submitted to the learning algorithm. At this stage, feature vectors are extracted for each pixel compounding a decision array (c). In a next stage, new gray level images (d) are submitted to the decision array. After extracting feature vectors for each pixel from the new images, the segmentation algorithm seeks through the k-NN algorithm the best result for every one of these pixels generating segmented images (e).

The nearest neighbor (NN) rule is a learning model based on instances. An early formalization of this rule can be found in [1]. The 1-NN approach can be extended to the k nearest neighbors, or k-NN.

The nearest neighbor rule requires the definition of a distance between two elements. Distance used in this work is the Euclidian Distance *(3) (4)*. The value of this distance between two vectors $\underline{x}$ and $\underline{y}$ is given by:

$$D(\underline{x}, \underline{y}) = ||\underline{x} - \underline{y}|| = \sqrt{(\underline{x} - \underline{y})^T (\underline{x} - \underline{y})} \qquad (3)$$

$$= \sqrt{\sum_{i=1}^{n} (\underline{x_i} - \underline{y_i})^2} \qquad (4)$$

2.2 The Learning Stage

The learning stage consists in manipulating two images, the first one in grayscale levels, also called sample, and the second one being its ideal segmented version. Each pair of pixels proceeding from both sample image and the ideal image, along with their respective neighborhoods, is analyzed and a set of features is defined. From this step it is generated one vector of the classification array for each gray level that will be forward re-utilized at the time of the segmentation of the new images. The steps of the learning stage are thus:

- Compute the feature set for each pixel in the sample image.
- Assign to the feature set the corresponding input grayscale level.
- Assign to the feature vector above generated, the ideal output.
- Verify the classification by detecting two pixels with equal feature set and different ideal output. This situation characterizes a learning error. An automatic neighborhood window sizing is then employed. The window size (initially 3x3 pixels) is enlarged and the learning step is re-initiated until these learning errors are eliminated.
- Store the converging learning window size.
- Finally, store the *classification array*.

2.3 Features selection

It is intended to find those features that better define the relationship between one pixel in the sample image and its corresponding one in the ideal segmented image.

Among the wide variety of possibilities of features available in literature, efforts were taken to select those that

(a) better define the relationship between the sample image and the ideal segmented image;

(b) are able to reproduce the learning results in images of different universes; and

(c) reduce the computational cost and learning errors.

The features used in this approach are: gray pixel level (x_i), mean (μ), variance (σ^2 or v), skewness (γ or s) and kurtosis (c). They represent the first *(5)*, second *(6)*, third *(7)* and fourth moments *(8)*, respectively. Corresponding formulas are denoted below.

$$\mu = E(X) = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad (5) \qquad\qquad \sigma^2 = E(X^2) = \frac{1}{N}\sum_{i=1}^{N}[x_i - E(X)]^2 \qquad (6)$$

$$\gamma = \frac{E[(X - E(X)]^3}{N\sigma^3} \qquad (7) \qquad\qquad c = \frac{E[(X - E(X)]^4}{N\sigma^4} \qquad (8)$$

2.4 Adaptive window sizing and classification array construction

A learning process is efficient only if it is able to solve any ambiguity. In order to overcome this challenge, an adaptive square neighborhood was chosen to adjust the learning from the context.

Beginning with a 3x3 window, the learning process computes all features for each pixel. In order to detect learning conflicts, a verification step will detect if there exist two or more input pixels with the same feature set but different ideal outputs. If this happens, the window size is automatically increased (Figure 2). This process is automatically repeated until no learning conflict remains. Concluded this step, the converging learning window size is stored.

The classification array stores, accordingly to the gray levels, the normalized feature vectors obtained for all the pair of pixels of the ideal input and output images, excluding redundancies.

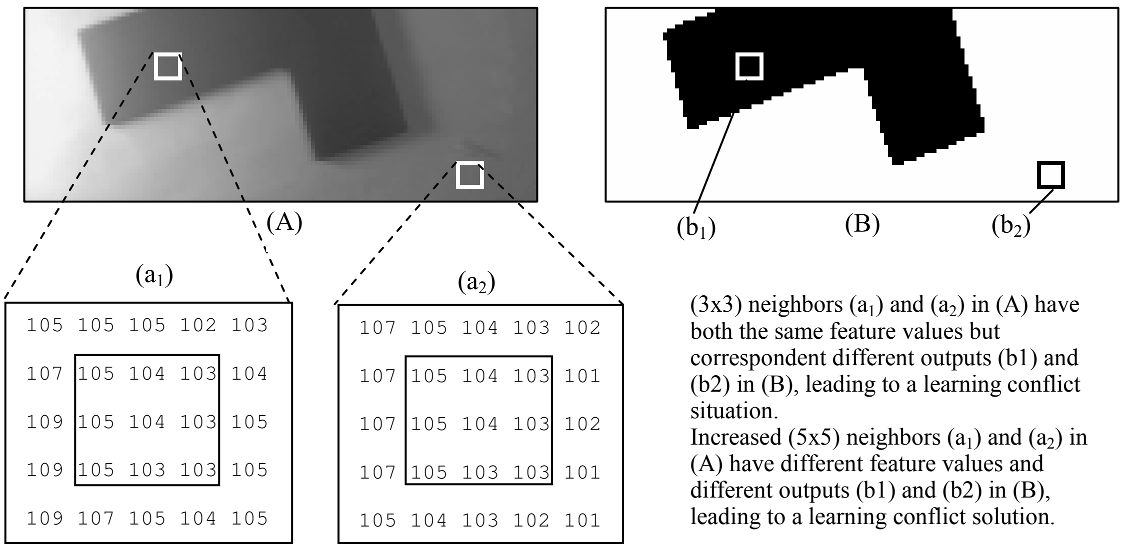


Figure 2. Simplified example of an adaptive window-sizing algorithm

2.4. New images segmentation by learning

After storing the classification array, the process is able to segment other images. The steps of the segmentation stage by learning are:

- Set the classification window size from the stored learning window size.
- Compute the feature set for each pixel in the new image to be segmented.
- Assign the feature set to the corresponding input grayscale level.
- Define the "most likely" pair (feature vector, output) in the classification array. This step is performed using the *k*-Nearest Neighbor strategy.
- Assign to each pixel the output value stored in the classification array.

2.5. Selection of best feature set and best *k* - nearest neighbor parameter

Our novel approach includes an automatically selection of the best feature set and of *k* parameter in Nearest Neighbor strategy. This adaptive feature set selection algorithm consists of the following tasks:

a) Randomly select the pair of sample/ideal images from the database.

b) Define the classification arrays using the combination of the four features and odd values of *k* parameter (typically, from 1 to 15).

c) Perform the cross validation by segmenting the remaining selected images using each classification array.

d) Compare each image segmented and its correspondent ideal one (Ground truth one).

e) Compute the segmented rates.

f) Automatically select the feature set and k parameter that result in the best segmentation rate.

g) Update the selected classification array with the relevant information computed during the learning and adaptive selection stages: window size, feature set and *k*-nearest neighbor parameter.

3. Experimental Results

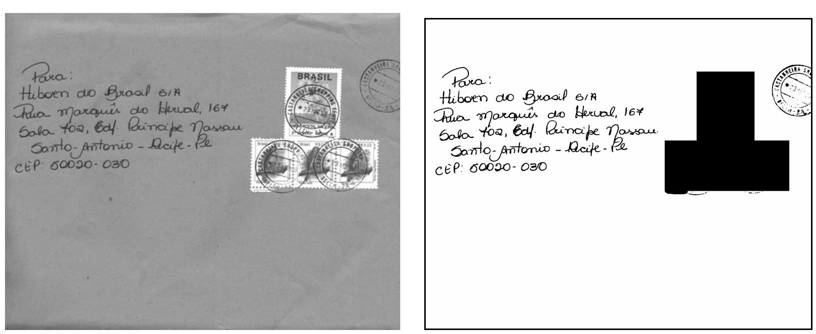
For the experiments we have a database with 55,000 original images of postal envelopes. The tests presented here include 200 originals collected from the same samples used in [2] and [8].

It could be of interest some considerations about the methodology of extraction of the 200 images used in the tests. Our postal envelope database consist of 55,000 original images, being 83.1% of them small pieces of white simple backgrounds and the remaining 16.9% big envelopes with complex backgrounds. The 200 images used in the experiment does not represent statistical samples of the database, but images collected among those from the most complex group.

Ten postal envelopes images were randomly chosen among the 200 ones and submitted to the learning stage.

A ground-truth strategy [7] was employed to evaluate the accuracy of the proposed approach. The ideal result (ground-truth segmentation) regarding each class (address block, postmarks and stamps) has been generated for each envelope image. An example of ground-truth segmentation is illustrated in Figure 3. It is important to emphasize that the ground-truth images have been generated as real as possible following some rules:

- Address blocks appear entirely, with accents, commas, dots, and superfluous words;
- Postmarks appear entirely only if they are overlapped with background. Otherwise, they are omitted;
- Stamps, due to its complexity (which contain complex drawings, dark objects and bright background), appear entirely in a black “rectangle”.



(a) (b)

Figure 3. Example of ground-truth image. (a) Input image, (b) Ideal output image

By comparing identical pixels at the same location in the ground-truth images and segmented ones (true positive), a score of segmentation was computed.

Through a ten fold cross validation strategy [4], there were performed 90 tests for each feature set. After submitting to all the combinations of the four features, it can be observed in Table 1 and Figure 3, that the best segmentation result was achieved with the feature set [mean = “on”, variance = “on”, skewness = “off”, kurtosis = “off”], [mvsc=1100].

Same ten prior postal envelopes were also submitted to the adaptive k-NN selection algorithm, which computed the k Nearest Neighbor parameter that results in the best segmentation rates. Table 2 and Figure 5 show the partial numerical results of the adaptive k-NN parameter selection algorithm. It can be observed that for this specific envelopes image database that average segmentation rates diminish while *k* parameter grows.

Feature-set [mvsc]	**Segmentation (% μ±σ)**	**Noise (% μ±σ)**
0001	70.01±15.08	3.52±2.61
0010	66.84±12.44	3.11±2.28
0011	71.25±14.66	3.09±1.67
0100	86.10±12.00	2.23±1.24
0101	85.04±12.23	1.94±1.08
0110	85.35±12.29	1.74±1.01
0111	84.37±12.55	1.76±1.08
1000	79.99±11.75	3.91±2.86
1001	77.86±12.59	2.17±1.41
1010	77.19±13.51	2.09±1.30
1011	76.09±14.09	2.03±1.35
1100	86.50±12.23	1.74±1.15
1101	85.37±12.28	1.47±0.86
1110	85.28±12.14	1.48±0.89
1111	84.03±12.39	1.55±1.04

Table 1. Numerical results of the adaptive feature set selection algorithm. A codification [mvsc] was use for mean (m), variance (v), skewness (s) and kurtosis (c), where 1 means “feature on” and 0 means “feature off”.

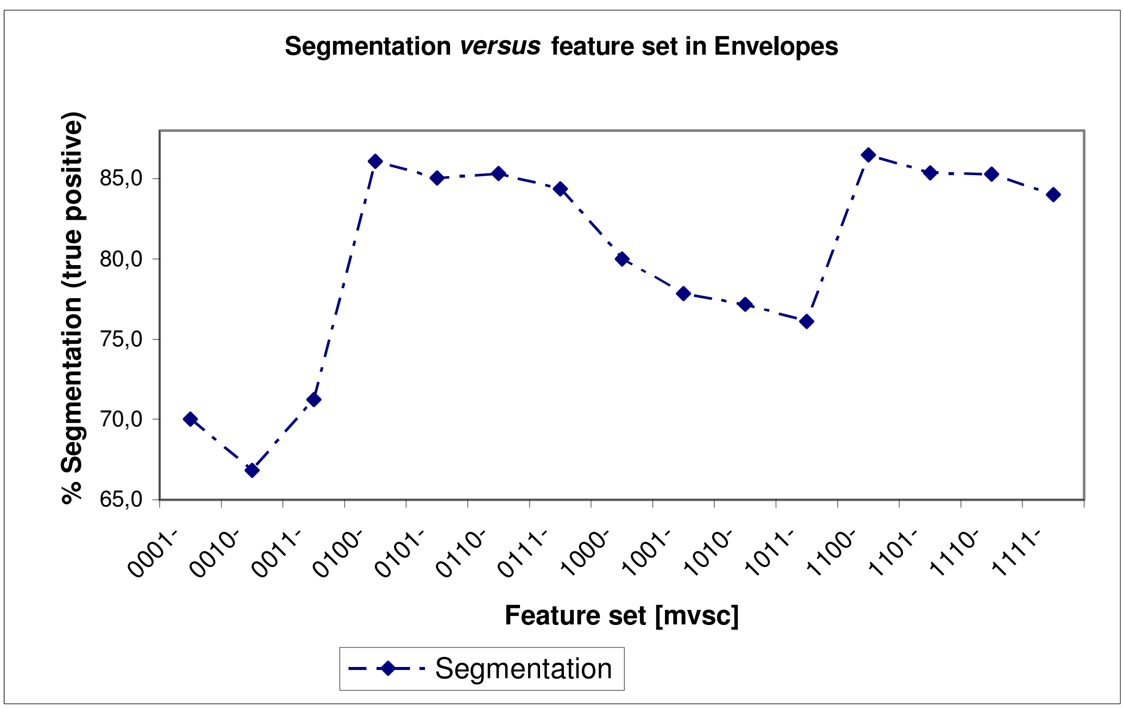


Figure 4 - Numerical results of the adaptive feature set selection. A codification [msvc] was use for mean (m), variance (v), skewness (s) and kurtosis (c), where 1 means “feature on” and 0 means “feature off”.

Feature set: Mean - Variance		
k-NN	Segmentation (% $\mu\pm\sigma$)	Noise (% $\mu\pm\sigma$)
1	80.31±14.3	1.86±1.68
3	78.26±11.6	1.44±1.61
5	74.18±11.9	1.21±1.58
7	71.04±12.2	1.17±1.73
9	69.39±12.2	1.25±1.95
11	68.70±12.1	1.41±2.33
13	68.49±12.1	1.60±2.86
15	68.40±12.0	1.78±3.46

Table 2. Numerical results of the adaptive k-NN parameter selection algorithm

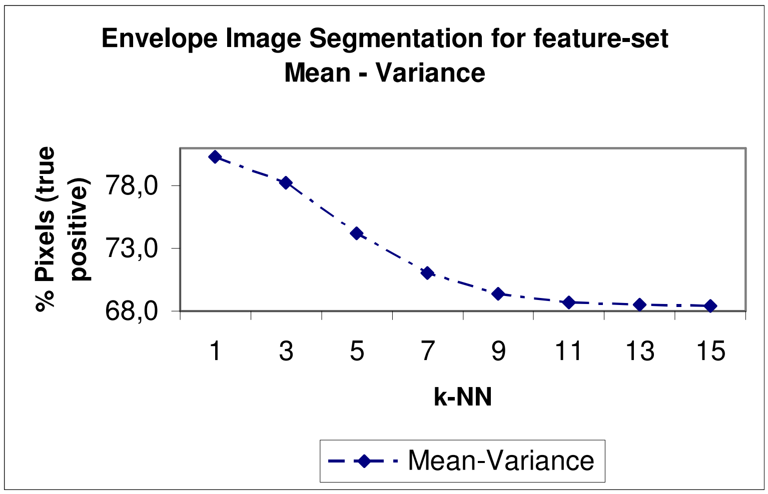


Figure 5. Numerical results of the adaptive k-NN parameter selection algorithm

Thus, the best feature set and k parameter computed by the automatic selection algorithm for this specific tested database were:

- Feature set: mean and variance;
- k-NN parameter*:* 1**.**

Once all the parameters computed, and the classification array built, the segmentation stage was performed over the 200 images.

Table 3 gives the average success accuracy considered pixel by pixel for the address block, stamps, rubber stamps and the noise left in the background. Also, a comparison between this results and prior approaches over the same database is shown.

<table>
<tr><th>Region class</th><th colspan="4">Accuracy pixel by pixel (% $\mu\pm\sigma$)</th></tr>
<tr><th></th><th>Learning-based approach</th><th>2-D histogram approach</th><th>Fractal dimension approach</th><th>Wavelet approach</th></tr>
<tr><td>Address block</td><td>98.62 ± 06.91</td><td rowspan="2">75</td><td>97.24 ± 13.64</td><td>97.13 ± 05.86</td></tr>
<tr><td>Rubber stamp</td><td>95.36 ± 13.64</td><td>91.89 ± 17.22</td><td>88.12 ± 19.36</td></tr>
<tr><td>Stamp</td><td>87.04 ± 19.67</td><td>25</td><td>66.34 ± 16.90</td><td>25.44 ± 19.29</td></tr>
<tr><td>Noise</td><td>02.06 ± 01.32</td><td>N/A</td><td>N/A</td><td>00.58 ± 00.88</td></tr>
</table>

Table 3. Average results with identification of regions (pixel by pixel accuracy) for the 200 envelopes images tested and comparison with prior approaches.

It can be observed that the proposed learning-based approach achieves superior address block class segmentation rates (98.62%) compared to 2-D histogram approach [12] (75%), best solution of the fractal dimension approach [2] (97.24%) and wavelet approach [8] (97.13%).

Success rates achieved for address block, stamps, rubber stamps and noise suggest that the features used in the proposed learning-based approach improves the segmentation results.

Figure 6 shows the learning sample/ideal pair and one envelope segmentation result. Addressee names were hidden on purpose. Observe the segmentation objectives of the learning stage (sample / ideal image pair (a) and (b)):

- Gray level background must be segmented as white class.
- Addressee information must be segmented as black class.
- Wrinkled parts of the envelope must not interfere with the segmentation process.
- Rubber stamps must be segmented as black class.
- All stamps must be segmented as black class, even background (white) areas.

After submitting a new image (c) to the algorithm, segmented image by learning (d) achieved the proposed objectives with an error of 1.08% (noise).

4. Conclusions and Future Work

In this paper we have shown a modified learning-based approach for segmentation of postal envelopes and address block location. Success rates ($\mu \pm \sigma$) achieved for address block (98.62% ± 06.91%), stamps (87.04% ± 19.67%), rubber stamps (95.36% ± 13.64%) and noise (02.06% ± 01.32%) suggest that the features used in the proposed approach improves the segmentation results. The improvement of the segmentation rates achieved in this approach relies on the adaptive feature set and *k* Nearest Neighbor parameter selection. It must be remarked that the learning and parameter selection stage is performed once for the database and correspondent values are stored in the classification array built. The main advantages of this approach are the facility to faithfully reproduce the objectives of the user, and it does not require the use of heuristic parameters neither the interaction of a specialist user after the learning process.

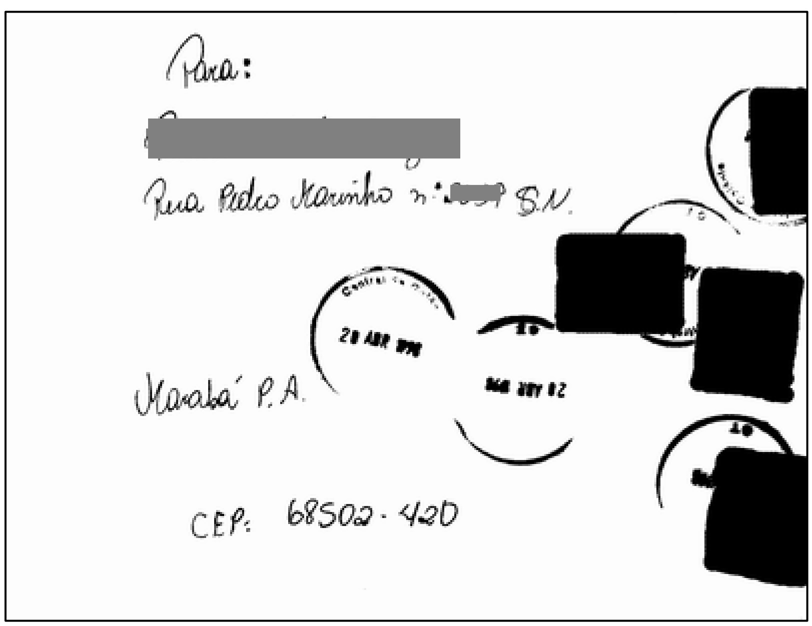
After these encouraging results, authors began to work over different strategies in order to achieve better segmentation rates and diminish noise. Undergoing research includes:

- Image post-processing.
- Use of new features such as: Laplacian, Fourier coefficients and Zernike moments.
- Implementation of more efficient *k* Nearest Neighbor searches.
- Implementation of other similarities and metrics tools.
- Experiments over other images databases, such as: faces, Guttenberg's Bible, MRI, fingerprint and signatures.

Results of these works are planned to be published in the near future.

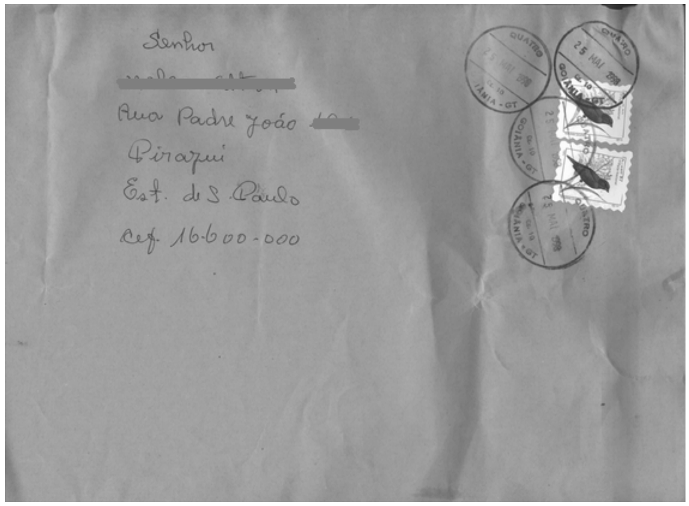

Para:
[redacted]
Rua Pedro Marinho n.º [redacted] S.N.
20 ABR 1998
Marabá P.A.
CEP: 68502-420

(a)

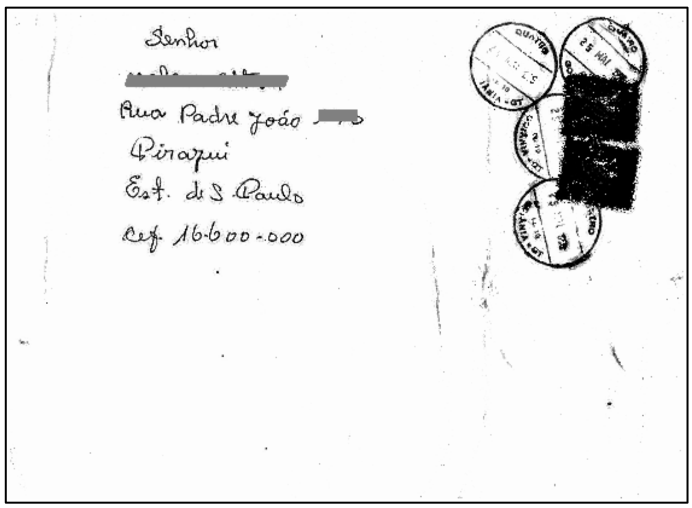

Para:
[redacted]
Rua Pedro Marinho n.º [redacted] S.N.
20 ABR 1998
Marabá P.A.
CEP: 68502-420

(b)

(c)



(d)



(e)

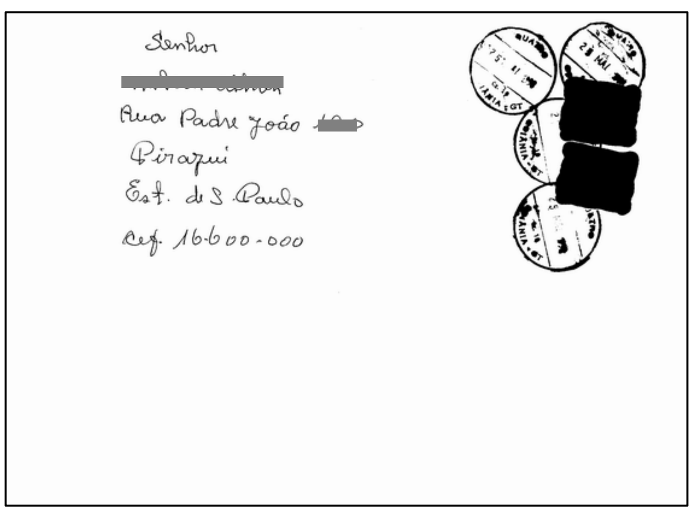


Figure 6. Learning sample/ideal pair and one envelope segmentation result.

(a) Gray level image sample (b) Ideal image (c) New gray level image
(d) Segmented image by learning approach (e) Ground truth image
Segmentation by classes – Address block: 95.77%, Rubber stamp: 96.67%, Stamp: 84.01% - Noise: 1.08%

References

[1]Cover, T.M. and Hart P.E. Nearest Neighbor Pattern Classification. *IEEE T. Information Theory*, Vol. IT-13, no. 1, pp. 21-27, 1967.

[2] Eiterer L.F., Facon, J. and Menoti, D. Segmentation of Envelope Address Blocks through Fractal-Based Approach*, 5th Iberoamerican Meeting on Optics (RIAO) and 8th Latin American Meeting on Optics, Lasers and Their Applications (OPTILAS), Proceedings SPIE*, Vol. 5622, 2004, pp. 1265-1270.

[3] Jain, A. and Bhattacharjee, S. Address block location on envelopes using Gabor filters. *Pattern Recognition*, Vol. 25, No. 12, 1992, pp. 1459-1477.

[4] Kohavi, R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, IJCAI 95, Montréal, Québec, Canada, Vol. 2, 1995, pp. 1137-1145.

[5] Legal-Ayala, H.A. and Facon, J. Automatic Segmentation of Brain MRI Through Learning by Example*, Proceedings of 11th International Conference on Image Processing (ICIP)*, IEEE Signal Processing Society, Singapore, Vol. 2, 2004, pp. 917-920.

[6] Legal-Ayala, H.A. and Facon, J. Comparative Results of Image Segmentation By Learning in Different Types of Images*. Proceedings of ICISP'2003 - IAPR International Conference on Image and Signal Processing*, Agadir, Morocco, Vol. 2, 2003, pp 381-388.

[7] Legal-Ayala, H.A. Uma Abordagem de Segmentação de Imagens por Aprendizagem, *PhD Thesis, PUCPR*, Curitiba, 2004, 151p. In http://www.ppgia.pucpr.br/teses/TesePPGIa-HoracioLegal-2004.pdf [Apr, 23, 2007].

[8] Menoti, D.; Facon, J.; Borges, D.L.; Brito Jr., A. Segmentação de Envelopes Postais para Localização do Bloco Endereço: uma abordagem baseada em seleção de características no espaço Wavelet. *In: Workshop of Thesis and Dissertation in SIBGRAPI 2004,* Anais Eletrônicos do WTD-SIBGRAPI2004, Curitiba, Brasil., 2004. pp. 1-8.

[9] Palumbo, P., Srihari, S. and Soh J. Postal address block location in real time. *IEEE Computer*, July 1992, pp. 34-42.

[10] Wolf, M., Niemann, H., and Schmidt. W. Fast address block location on handwritten and machine printed mail-piece images, *ICDAR´1997 - Fourth IEEE-IAPR International Conference on Document Analysis and Recognition*, Ulm, Germany, Aug. 1997, pp.753-757.

[11] Yeh, P., Antoy, S., Litcher, A. and Rosenfeld, A. Address Location on envelopes. *Pattern Recognition*, Vol. 20, No. 2, 1987, pp. 213-227.

[12] Yonekura, E. and Facon, J. Postal envelope segmentation by 2-d histogram clustering through watershed transform, *ICDAR 2003 - 7th International Conference on Document Analysis and Recognition*, 2003, pp. 338-342.