An Adaptive Zoning Technique for Efficient Word Retrieval Using Dynamic Time Warping

A. Papandreou
Computational Intelligence Laboratory,
Institute of Informatics and Telecommunications,
National Research Center "Demokritos",
153 10 Athens, Greece
alexpap@iit.demokritos.gr

Department of Informatics and Telecommunications,
National and Kapodistrian University of Athens,
Panepistimioupoli,
157 84 Athens, Greece
alxpap@di.uoa.gr

B. Gatos
Computational Intelligence Laboratory,
Institute of Informatics and Telecommunications,
National Research Center "Demokritos",
153 10 Athens, Greece
bgat@iit.demokritos.gr

G. Louloudis
Computational Intelligence Laboratory,
Institute of Informatics and Telecommunications,
National Research Center "Demokritos",
153 10 Athens, Greece
louloud@iit.demokritos.gr

ABSTRACT

Zoning features have been proved one of the most efficient statistical features which provide high speed and low complexity word matching. They are calculated by the density of pixels or pattern characteristics in several zones that the pattern frame is divided. In this paper, an adaptive zoning technique for efficient word retrieval is introduced. The main idea is that the zoning features are extracted after adjusting the horizontal boundaries of the zones with the use of Dynamic Time Warping (DTW). This adjustment is performed by coupling every zone of the query word to the corresponding zone of each candidate match-word with the use of the corresponding warping matrix. This process absorbs the ambiguities between the query and the candidate match words. The proposed word retrieval technique is tested using the pixel density as a characteristic feature in every zone and significant improvement is recorded compared to other state-of-the-art methods.

DATeCH 2014, May 19-20 2014, Madrid, Spain
ACM 978-1-4503-2588-2/14/05.
http://dx.doi.org/10.1145/2595188.2595218

Categories and Subject Descriptors

I.4.6 [**Image Processing**]: Segmentation; I.5.4 [**Pattern Recognition**]: Applications---Text Processing; I.7.5 [**Document and Text Processing**]: Document Capture---Document Analysis.

General Terms

Algorithms, Measurement, Performance, Experimentation.

Keywords

Zoning Features; Word Retrieval; Dynamic Time Warping; Word Matching.

1. INTRODUCTION

Word retrieval applications help locating all the occurrences of a given word image in a set of document images. Measuring the distance between two word images is an important step for all word retrieval applications that involve a word segmentation stage. Existing approaches for word image matching involve statistical, structural and transformation-based features [1-4]. In [1], sets of 1-dimensional structural features are created from the segmented word images which are then compared using Dynamic Time Warping (DTW). The distance between two word images is calculated in [2] using an image dissimilarity measure based on curvature estimation using integral invariants and a windowed Hausdorff distance.

Zoning features have been proved one of the most efficient statistical features which provide high speed and low complexity word matching. They are calculated by the density of pixels or pattern characteristics in several zones that the pattern frame is divided. In particular, standard zoning methods are defined according to an N×M regular grid superimposed on the image body [5]. Recently, zoning features based on pixel density have been combined with word profiles in a hybrid scheme for handwritten word recognition [3] as well for word spotting in historical printed documents [4]. In [6], features based on distances and angles of the skeleton pixels in each zone are used. Both neural networks and fuzzy logic techniques are then used for recognition. The methodology presented in [7] is based on the direction of the contour of the character by computing histograms of chain codes in each zone. In [8], it is observed that when the contour curve is close to zone borders, small variations in the contour curve can lead to large variations in the extracted features. For this reason, zones with fuzzy borders are introduced. Features detected near the zone borders are given fuzzy membership values to two or four zones. In [9], the role of feature membership functions in Voronoi-based zoning methods is investigated. Zoning is considered in [10] as the result of an optimization problem and a genetic algorithm is used to find the optimal zoning that minimizes the value of the cost function associated to the classification. In [11], the idea of adaptive zoning is introduced and the features are extracted after adjusting to the position of every zone based on local pattern information. This adjustment is performed by moving every zone towards the pattern body maximizing the local pixel density around each zone.

In this paper, an improved version of adapting zoning is proposed for word image matching. The use of DTW is introduced in order to adjust the horizontal boundaries of the zones of the two word images. According to the proposed approach, this adjustment is performed by coupling every zone of the query word to the corresponding zone of each candidate match-word with the corresponding warping matrix. This process absorbs the ambiguities between the query and the candidate words.

In Section 2, the proposed adaptive zoning technique is described. Furthermore, word retrieval with the use of the pixel density in every zone is extensively tested. As it is shown in Section 3, a significant improvement is recorded when the zoning features are used in the proposed adaptive way. Conclusions and future work plans are given in Section 4.

2. THE PROPOSED ADAPTIVE ZONING TECHNIQUE

2.1 Adjusting the Vertical Zones

As a first step, the vertical zones of the query word must be defined. Let a $N \times 1$ regular grid be superimposed on the pattern image. If the binarized pattern image $I(x,y) \in \{0,1\}$ has an overall size of $Ix \times Iy$ and every grid window is of size $K \times Iy$ then $Ix = N \times K$. Following the above mentioned zoning procedure, the coordinates of every vertical zone are defined as follows:

$$q_n^{x1} = (n-1)K \tag{1}$$

$$q_n^{x2} = nK - 1 \tag{2}$$

where q_n^{x1} and q_n^{x2}, denote respectively the beginning and the end of the n vertical zone where $n = 1..N$. In Figure 1 the partitioning of a query image into $N = 10$ zones is presented.

Figure 1. The zones of a query image for $N = 10$.

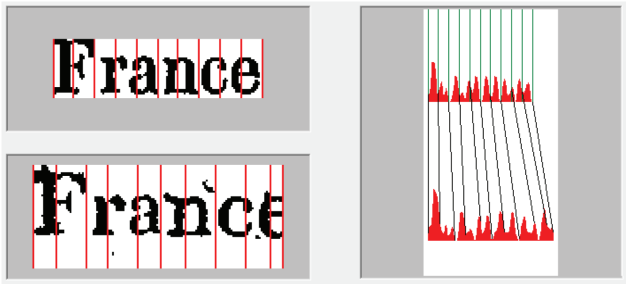
At a next step, a set of features v_i [1], where $i = 1 \dots 4$, of the query word image is calculated. The features include i) the vertical projection profile, ii) the upper profile, iii) the lower profile and iv) the black to white transitions. Then, a linear scaling is applied in order to transform the calculated features normalized between [0,1]. Finally, they are combined in a query characteristic feature vector sequence V. The same procedure is followed in order to produce the features p_i of the candidate word image and its characteristic feature vector sequence P. Once both feature sequences V and P are calculated, the DTW algorithm is applied in order to compute the minimum cost and the corresponding alignment path.

The role of the DTW algorithm is to measure the similarity between two sequences which may have variable lengths. After applying the DTW algorithm, the feature vectors are aligned along a common, warped time axis. The cost of the alignment is given by the sum of distances $d(V,P)$ of each aligned vector pair along the corresponding alignment path. The distance similarity measure $d(V,P)$ employed is the squared Euclidean distance:

$$d(V,P) = \sum_{i=1}^{4} (v_i - p_i)^2 \tag{3}$$

The DTW cost $DTW(V,P)$ and the corresponding warping matrix between a query word and a candidate word is defined as the minimum alignment cost which is calculated using the principles of dynamic programming [1]. This cost is normalized with respect to the length of the previously calculated warping path.

Consequently, according to the estimated warping path the coordinates of every vertical zone q_n^{x1}, q_n^{x2} belonging to the query word image are matched with the coordinates of the candidate word image c_n^{x1}, c_n^{x2} (see Figure 2).



(a) (b)

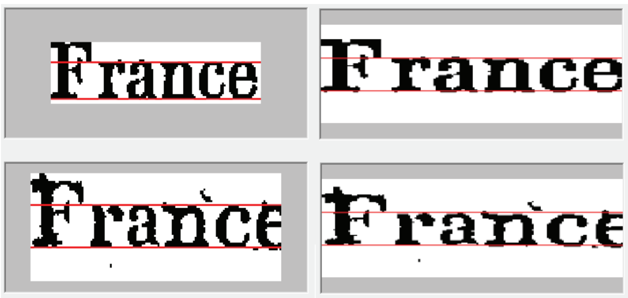
Figure 2. The coordinates of the vertical zones of the query are matched with the corresponding coordinates of the candidate match word image. (a) mapping on the word images and (b) on their vertical Histograms.

The adjustment of the vertical zones between a query word and a candidate word results in vertical zones of variable width at the candidate word. In that way, an optimal adaption of the candidate word to the query word is achieved.

2.2 Size Normalization and Horizontal Zones

For the horizontal adjustment of the query and the candidate word images, a normalization step is applied resulting respectively in images $Q[W \times H]$ and $C[W \times H]$ of width W and height H. The positioning of each word in its size-normalized image $W \times H$ is accomplished by placing the baseline areas of each word in the center of the matrix i.e. the upper word baseline at vertical offset $H/3$ and the lower word baseline at vertical offset $2 \times H/3$. The baseline detection procedure is described in [12]. A similar approach is also used in [11]. A word size-normalization example is presented in Figure 3 for the query and a candidate word.

After normalizing the two words, the estimated coordinates of the vertical zones $q_n^{x1}, q_n^{x2}, cm_n^{x1}, cm_n^{x2}$ are also normalized to $q_n^{x1'}, q_n^{x2'}, cm_n^{x1'}, cm_n^{x2'}$ with the use of the corresponding transformation matrix, respectively to the new width, W, of the normalized images.



(a) (b)

Figure 3. Word size normalization example: (a) original, (b) size-normalized image.

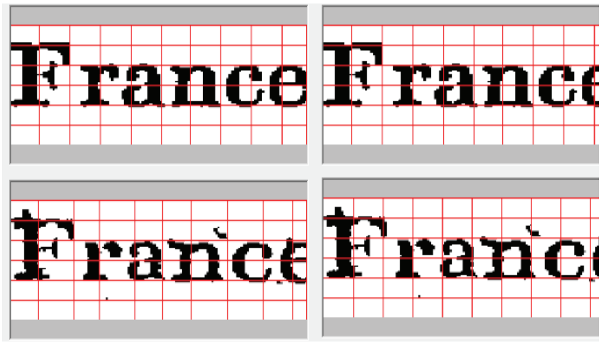
For the definition of the horizontal zones, a $1 \times M$ regular grid is superimposed on the normalized images. The coordinates of every horizontal zone are defined by the following equations:

$$q_m^{y1'} = (m-1)\Lambda \qquad (4)$$

$$q_m^{y2'} = m\Lambda - 1 \qquad (5)$$

where $\boldsymbol{q_m^{y1'}}$ and $\boldsymbol{q_m^{y2'}}$ correspond respectively to the beginning and the end of the horizontal zones, $\boldsymbol{\Lambda}$ is the height of the zones and $\mathbf{m = 1..M}$. Concerning the coordinates of horizontal zones of the candidate word $\mathbf{c_m^{y1'}}$ and $\mathbf{c_m^{y2'}}$ it should be noted that they are equal to $\boldsymbol{q_m^{y1'}}$ and $\boldsymbol{q_m^{y2'}}$ since the normalization procedure results in images having the core region located in the same area. In Figure 4 the zones of the query and a candidate word for $\mathbf{M = 6}$ are demonstrated.

Figure 4. The zones of a query image for $M = 6$.



(a) (b)

Figure 5. Comparison of the Query word and Candidate match word grids cut in (a) the proposed adaptive way and in (b) the classical way.

After the definition of both, horizontal and vertical zones a final $N \times M$ grid is defined which is superimposed on the query word as well as on the candidate word. Both query and candidate match grids are demonstrated in comparison in Figure 5.

2.3 Feature Extraction

The most common type of zoning features is calculated using the density of pixels. Features based on pixel density are calculated directly from the size-normalized images Q and C. The density, dQ_{nm}, of the (n,m) window of Q is calculated as follows:

$$dQ_{nm} = \frac{1}{(q_n^{x2'} - q_n^{x1'}) \times \Lambda} \sum_{x=q_n^{x1'}}^{q_n^{x2'}} \sum_{y=q_m^{y1'}}^{q_m^{y2'}} Q(x,y) \qquad (6)$$

Respectively, the density, dC_{nm}, of the (n,m) window of C is computed as follows:

$$dC_{nm} = \frac{1}{(c_n^{x2'} - c_n^{x1'}) \times \Lambda} \sum_{x=c_n^{x1'}}^{c_n^{x2'}} \sum_{y=c_m^{y1'}}^{c_m^{y2'}} C(x,y) \tag{7}$$

The total number of features based on pixel density is $N \times M$ and all features range between 0 and 1.

Additionally, due to the fact that the vertical zones of the candidate word image have variable width, the density of each window (n,m) will be weighted to their respective width. The following equation defines the final density:

$$dC'_{nm} = dC_{nm} \times \frac{(c_n^{x2'} - c_n^{x1'})}{W} \tag{8}$$

The calculation of the distance, $Dist$, of the query and the candidate word image is defined as:

$$\mathrm{Dist} = \frac{1}{\mathrm{N} \times \mathrm{M}} \sum_{x=1}^{\mathrm{N}} \sum_{y=1}^{\mathrm{M}} (\mathrm{dQ_{nm}} - \mathrm{dC'_{nm}})^2 \tag{9}$$

Finally, the distance Dist of the two words is multiplied by the distance $\mathrm{d}(\mathrm{V},\mathrm{P})$ provided by the DTW algorithm, which is also a similarity measure. In more detail, the distance D of the proposed adaptive method is:

$$\mathrm{D} = \mathrm{Dist} \times \mathrm{d}(V,P) \tag{10}$$

3. EXPERIMENTAL RESULTS

In this Section, we present the experimental results of the proposed adaptive zoning method in comparison with several state-of-the-art word matching algorithms. A set of 46197 words coming from 153 pages of a historical French book [13] from project IMPACT [14] was used. The word segmentation together with the ASCII ground-truth was manually produced. The same set was used for evaluation also in [2], [11] and [15]. In order to compare in the same basis and demonstrate solely the efficiency of the proposed adaptive zoning technique, the same normalization parameters were adopted. According to that, all words were normalized to a 300×90 matrix ($\mathrm{W} = 300, \mathrm{H} = 90$) following the procedure described in Section II. Furthermore, every word is divided into 30×9 zones ($\mathrm{N} = 30, \mathrm{M} = 9$).

Five instances of the words "France", "Louis", "famille", "mort" and "justice" were randomly selected, thus yielding a total of 25 queries (see also [11]).

In order to demonstrate the robustness of the proposed algorithm regardless of the ambiguities between similar word images, we tried to tackle a more generic and difficult problem than the one described in [11]. In more detail, we no longer consider the problem from the word-matching perspective. Instead, it is considered as a word retrieval problem. To this end, we do not consider as correct only the word images that correspond to the exact same transcript as the query but we also regard as correct (a) word instances having their first letter in capital, (b) instances that are followed by punctuation (",", ".", ";", "?", etc.) and (c) instances that are fully written in capital letters. In that way, we try to address a more realistic and difficult problem, that has a greater impact on the individual user of a word retrieval application, since the user is trying to spot words of interest regardless their positioning in the sentence and not just the words having exactly the same transcript. Samples of the instances that were included in the correct answers are demonstrated in Figure 6.

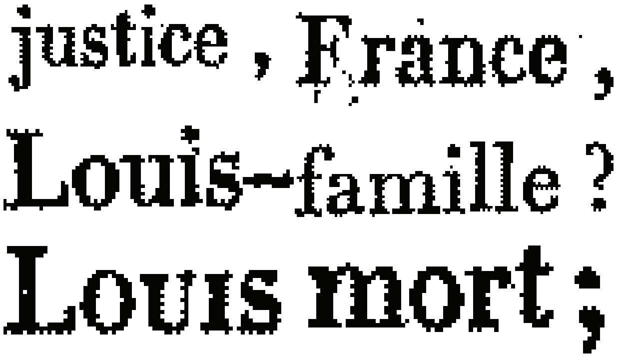


Figure 6. Samples of the instances that were included in the correct answers.

Consequently, after expanding the correct answers by including all the above mentioned instances that are considered as ground-truth, the total number of instances of the words "France", "Louis", "famille", "mort" and "justice" are respectively 82, 162, 68, 68 and 70.

Let n_inst be the total number of instances of a word in the ground truth and n_corr the number of correct instances of the word in the first n_inst retrieved instances. The word retrieval performance can be calculated as follows:

$$Word\ Retrieval\ Performance = \frac{n\_corr}{n\_inst} \tag{11}$$

In Table 1, the retrieval performance achieved by the proposed adaptive method is presented in comparison with the performance of (a) the classical zoning using the same normalization and number of zones as described in [11], (b) the adaptive zoning described in [11] and (c) DTW with the four features that were used also in the proposed method (vertical projection profile, upper profile, lower profile, black to white transitions) for each word query.

Table 1. Comparative Results of Different Methodologies

Word Query	*Different Methodologies*			
	Classical Zoning	Adaptive Zoning	DTW	Proposed Zoning
France	51.71%	53.66%	56.58%	**79.75%**
Louis	91.36%	**93.33%**	71.73%	91.60%
famille	53.53%	59.41%	68.82%	**89.70%**
mort	66.18%	66.18%	64.41%	**82.64%**
justice	43.43%	54.00%	71.14%	**80.57%**
Overall Results	67.16%	70.76%	68.09%	**86.08%**

As it can be derived from Table I the classical and the adaptive zoning techniques [11] were unable to confront with the more generic and difficult problem of word retrieval since they didn't absorb adequately the ambiguities between the word images. Similarly, DTW, though flexible in the variations, it doesn't have the descriptive power of a zoning technique which takes advantage of the local information of the word image and consequently fails to tackle the difficult problem of word retrieval. On the other hand, the proposed method achieved a word retrieval performance of 86% and proved to be descriptive and yet flexible in handling the ambiguities among different instances of the same word.

Also it is interesting the fact that due to the coupling of the DTW the proposed algorithm would be able to retrieve also words of the same root when the right weights would be applied in the vertical zones that correspond to the endings of the query and the candidate words. A visual, qualitative, estimation of this potential is demonstrated in Figure 7.

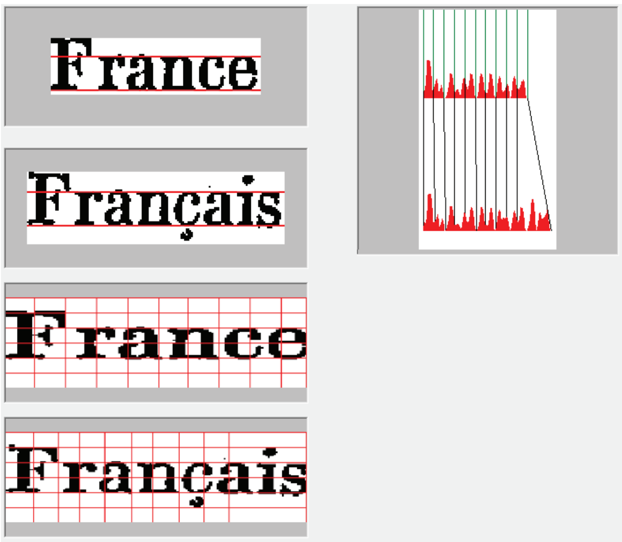


Figure 7. Qualitative estimation of potential use of the proposed algorithm to retrieve words of the same root.

As it is shown in Figure 7, if the windows of the most right vertical zone were weighted to contribute the least in the calculation of the distance between the query and the candidate word images, then "francais" could be retrieved from "France". This is a potential use of the proposed adaptive zoning algorithm and it is feasible due to the DTW coupling of the zones.

4. CONCLUSIONS AND FUTURE WORK

In this paper, a novel and efficient adaptive zoning method was presented. The idea of adjusting the horizontal boundaries of the zones with the use of DTW was introduced and this was proved to result in significantly higher word retrieval performance. The proposed adaptive zoning method was tested in a generic and difficult task of word retrieval and achieved a performance of 86%. This demonstrates the fact that it is a descriptive method with the advantage of handling the ambiguities among different instances of the same word. The proposed adaptive zoning method outperforms the classical and a state-of-the-art adaptive zoning method as well as the DTW algorithm.

Future work includes the investigation of the modifications needed and the potential of the proposed method to handle handwritten text. It would be also interesting to have a quantitative measurement on the success of this adaptive zoning method to retrieve words of the same conceptual root but in different gender or with different endings.

5. ACKNOWLEDGMENT

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 600707 – tranScriptorium.

6. REFERENCES

[1] T.M. Rath and R. Manmatha, "Word image matching using dynamic time warping", Proc. IEEE Computer Society Conference on Vol. 2, pp. 521-527, 2003.

[2] S. Colutto and B. Gatos, "Efficient Word Recognition Using A Pixel-Based Dissimilarity Measure", Proc. 11th International Conference on Document Analysis and Recognition (ICDAR), pp. 1110-1114, 2011.

[3] B. Gatos, I. Pratikakis and S.J. Perantonis, "Hybrid Off-Line Cursive Handwritten Word Recognition", Proc. 18th International Conference on Pattern Recognition (ICPR), pp. 998-1001, 2006.

[4] T. Konidaris, B. Gatos, K. Ntzios, I. Pratikakis, S. Theodoridis and S. J. Perantonis, "Keyword-Guided Word Spotting in Historical Printed Documents Using Synthetic Data and User feedback", International Journal on Document Analysis and Recognition (IJDAR), special issue on historical documents, Vol. 9, No. 2-4, pp. 167-177, 2007.

[5] M. Bokser , "Omnidocument Technologies", Proc. IEEE, Vol. 80, pp. 1066-1078, 1992.

[6] M. Hanmandlua, K.R. Murali Mohanb, S. Chakrabortyc, S. Goyald and D. Roy Choudhurye, "Unconstrained handwritten character recognition based on fuzzy logic", Pattern Recognition, Vol. 36, pp. 603–623, 2003.

[7] K. M. Mohiuddin and J. Mao, "A Comprehensive Study of Different Classifiers for Hand-printed Character Recognition", Pattern Recognition, Practice IV, pp. 437- 448, 1994.

[8] V.L. Lajish, "Handwritten Character Recognition using Perceptual Fuzzy-Zoning and Class Modular Neural Networks", Proc. 4th International Conference on Innovations in Information Technology, Innovations '07, DOI: 10.1109/IIT.2007.4430497, 2007.

[9] S. Impedovo, A. Ferrante, R. Modugno and G. Pirlo, "Feature Membership Functions in Voronoi-Based Zoning", AI*IA 2009: Emergent Perspectives in Artificial Intelligence Lecture Notes in Computer Science Volume 5883, pp 202-211, 2009.

[10] S. Impedovo, M. G. Lucchese, G. Pirlo, "Optimal zoning design by genetic algorithms", IEEE Transactions on Systems, Man, and Cybernetics, Part A 36(5), pp. 833-846, 2006.

[11] B. Gatos, A.L. Kesidis and A. Papandreou,"Adaptive Zoning Features for Character and Word Recognition", Proc. 11th International Conference on Document Analysis and Recognition (ICDAR), pp. 1160-1164, 2011.

[12] A. Papandreou and B. Gatos, "Slant estimation and core-region detection for handwritten Latin words", Pattern Recognition Letters, Available online 29 August 2012, ISSN 0167-8655, 10.1016/j.patrec. 2012.08.005.

[13] Le Dernier fils de France, ou le Duc de Normandie, fils de Louis XVI et de Marie-Antoinette, par A. Bibliothèque nationale de France,1838.

[14] http://www.impact-project.eu/

[15] G. Louloudis, A.L. Kesidis and B. Gatos, "Efficient word retrieval using a multiple ranking combination scheme", Proc. 10th IAPR International Workshop on Document Analysis Systems (DAS), pp. 379-383, 2012.