Isolated Character Recognition using Projections of Oriented Gradients

George Retsinas[1,2], Basilis Gatos[1], Nikolaos Stamatopoulos[1] and Georgios Louloudis[1]

[1] Computational Intelligence Laboratory, Institute of Informatics and Telecommunications
National Center for Scientific Research ”Demokritos”
GR-15310 Athens, Greece
{georgeretsi,bgat,nstam,louloud}@iit.demokritos.gr

[2] School of Electrical and Computer Engineering
National Technical University of Athens
GR-15773 Athens, Greece

***Abstract*—In this paper, we present a new approach for off-line isolated character recognition. The proposed method relies upon the application of a projection-based feature extraction stage, which resembles the Radon transform, on both the original image and a set of generated images corresponding to different gradient orientations of the original image. For the classification stage, Support Vectors Machines (SVM) are used. The proposed method is evaluated using one typewritten (GRPOLY-DB - Historical Greek) and two handwritten (CIL - Greek, CEDAR - English) publicly available databases. Experimental results prove the efficiency of the proposed method compared to several state-of-the-art techniques.**

***Keywords*—Character Recognition, Feature Extraction, Projection-based Features, Radon Transform**

I. Introduction

Recognition of handwritten and machine-printed characters has been an active research area for decades. A widely used approach in isolated character recognition systems is to follow a two step schema: a) feature extraction and b) classification. Selection of an appropriate feature extraction method is of great importance for achieving high recognition performance. The main goal of feature extraction techniques is to extract representative information from the original image in a compact and robust way. In order to improve the effectiveness and efficiency of the subsequent classification stage, the selected features should minimize the within class pattern variability, while enhancing the separability of the classes.

In the literature, feature extraction methods for character recognition can be categorized broadly into two groups according to different types of features: **a)** statistical and **b)** structural. Statistical features are derived from statistical distribution of the foreground pixels and can be sub-categorized as follows: i) zoning, where the image is divided into several zones ([1],[2]), ii) projections [3], iii) moment-based features, such as Legendre or Zernike moments [4], and iv) crossings and distances [5]. Structural features are based on topological and geometrical properties of the character, such as maxima and minima, reference lines, ascenders, descenders, cusps above and below a threshold, cross points, branch points, strokes and their direction [6].

Image zoning is a widespread and efficient feature extraction technique in which a regular or semi-regular [7] grid is superimposed on the image in order to extract local features. The extracted features are drawn from a variety of choices, such as pixel density, local gradient based on contour, distances and angles of the skeleton pixels, etc. The main drawback of a zoning methodology lies in the imposed locality of the grid, which may lead to poorly correlated features between images of the same class. This problem is addressed in some extent in [2], where the constraint of a fixed grid is relaxed and adaptive topologies are described. A slightly different approach is adopted in [8], where the image is iteratively subdivided according to a center point. Conversely, a more global approach is adopted by projection-based features. Projections were one of the first successful techniques in character recognition [3], using vertical and horizontal projections of the image. However, crucial local information may be lost, especially for such a sparse sampling of the initial image while using vertical and horizontal projections.

In this paper, a novel feature extraction technique for character recognition is presented. The proposed approach attempts to overcome the problem of high variation, mainly in handwriting, which is accentuated by the use of local features. In contrast to using a grid-based technique for local features extraction, a more global interpretation of the image is adopted. Specifically, the proposed feature extraction is a projection-based technique, which resembles the Radon transform [9]. In addition, an attempt is made to incorporate popular ideas from state-of-the-art feature extraction methods, such as Histogram of Oriented Gradients (HOG) [10], based on the fact that the gradient information is proven very useful for a variety of object recognition problems. Consequently, instead of features derived from local gradient, a projection-based feature extraction for each gradient orientation, which is represented as a binary image, is adopted.

The remainder of the paper is organized as follows. In Sec-

978-1-4799-1805-8/15/$31.00 ©2015 IEEE

tion II, the proposed feature extraction technique is presented, as well as a visualization of the resulting descriptor. In Section III, the used datasets are described and the experimental results are discussed. Finally, conclusions and future directions are drawn in Section IV.

II. Feature Extraction

The proposed feature extraction technique, Projections of Oriented Gradients (*POG*), consists of two stages:

A) Computation of the oriented gradients of the image.

B) Extraction of a projection-based descriptor for each image gradient as well as for the initial image.

Before the computation of oriented gradients, a preprocessing step for noise reduction is applied. This is accomplished using a two-dimensional median filter (3×3 neighborhood) on the original image.

A. Gradient Orientation

This stage is essential for capturing informative details of the image, such as the change in the direction of the contour, through the representation of the gradient orientation as binary images.

The computation of the directional gradients G_x and G_y of the image $I(x,y)$, along x-axis (horizontal) and y-axis (vertical) respectively, is performed using the following filter kernels: $[-1\,0\,1]$ and $[-1\,0\,1]^T$. In order to compute the gradient orientation at each pixel, a transformation into polar coordinates is performed through Eq. 1,2. Only the orientation $\angle G$ is of interest and a wrapping is performed so that the orientation values lie on the interval $[0, 180°)$ instead of $[0, 360°)$ ("unsigned" gradient as in [10]).

$$|G(x,y)| = \sqrt{(G_x^2(x,y) + G_y^2(x,y))} \tag{1}$$

$$\angle G(x,y) = arctan\Big(\frac{G_y(x,y)}{G_x(x,y)}\Big) \tag{2}$$

The possible orientations, due to the selected gradient filter, are four (0°, 45°, 90° and 135°) and thus four binary images, that represent the gradient orientation, are constructed: $G_0 = (\angle G(x,y) = 0°)$, $G_{45} = (\angle G(x,y) = 45°)$, $G_{90} = (\angle G(x,y) = 90°)$ and $G_{135} = (\angle G(x,y) = 135°)$.

B. Projection-Based Descriptor

The basic concept of the projection-based feature extraction is to decompose the binary image into several projections under selected angles, imitating the Radon transform. The projections are n_θ in total, sampled every 180° / n_θ , i.e. the angles of the projections are $\theta_k = k\frac{180°}{n_\theta}$, $k \in [0, n_\theta - 1]$. Let I_x, I_y denote the size of the image, (x_i, y_i) denote every foreground pixel of the image ($i = 1, \ldots, N$), (x_c, y_c) denote the center of the image and θ denote the chosen angle of projection. The projection p_i of each pixel into the selected angle vector is computed according to Eq. 3.

$$p_i = (x_i - x_c)cos(\theta) + (y_i - y_c)sin(\theta), \quad i \in [1, N] \tag{3}$$

Finally, in order to obtain a K-length vector that represents the projection, a soft binning technique (K bins) is applied for p_i in the interval $[-\frac{1}{2}\sqrt{I_x^2 + I_y^2}, \frac{1}{2}\sqrt{I_x^2 + I_y^2}]$.

Our goal is to eliminate the redundancy and the excessive variability of the image. However, the resulting matrix of the Radon transform usually is rather large to be a useful descriptor. Furthermore, possible variations of the same character may translate into slight distortion of the projections or even spikes due to noise. Thus, each projection subjects to a smoothing-like procedure in order to reduce unnecessary variations. The most descriptive information corresponds to smoothed regions of relatively high pixel concentration, or equivalently to the low frequency coefficients of the Fourier Transform of the projection. Therefore, after computing the Discrete Fourier Transform coefficients c_j , $j \in [0, K-1]$ of the projection vector, only the first $n_c + 1$ are used to form the feature vector while the remaining are discarded. The aforementioned steps are depicted in Fig. 1 for a specific angle ($\theta = 45°$).

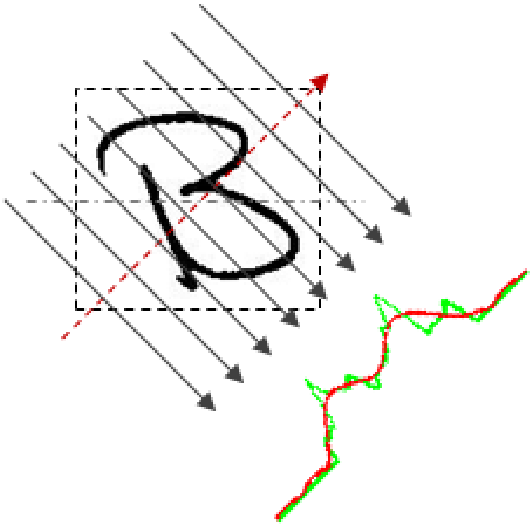


Fig. 1: Visualization of a projection ($\theta = 45°$) and its smoothing (low-frequency components)

Subsequently, a normalization with regard to the number of pixels in each image is applied by dividing each Fourier coefficient by N (foreground pixels), or equivalently by c_0, while discarding c_0. The complex feature vector for each projection is computed according to the following equation.

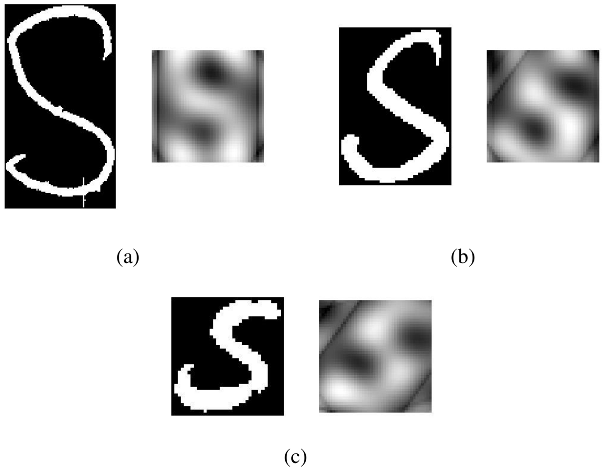
$$f_j = c_j / N\,, \quad j = 1, \ldots, n_c \tag{4}$$

The final feature vector is a concatenation of the real and imaginary parts of the complex feature vector for each projection. It is clear that a selection of a subset of the Fourier coefficients results to projection length independence and, consequently, to image size independence.

It should be noted that the described procedure is equivalent to a circular low-pass filter in 2-D frequency domain via the Projection-Slice theorem [11]. Specifically, each of the computed projections corresponds to a slice in the frequency domain via the Fourier transform of the projection, thus providing a more accurate sampling in low-frequencies comparing to two-dimensional Discrete Fourier Transform. Additionally, due to the applied lowpass filtering, only a few (n_θ) projections (angular sampling) suffice for representing the low-frequency domain.

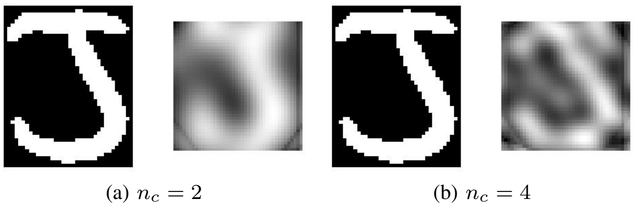
C. Reconstruction And Visualization

A useful property of the described method is the approximate reconstruction of the original image via the inverse Radon transform, after interpolating each projection to a specific length using the inverse Fourier transform. Therefore, a normalized approximation of the image is generated, which provides a visualization of the preserved information. Fig. 2 depicts the visualization of the described feature vector for three instances of the same class. It can be observed that the within-class variations seem to be reduced. Additionally, Fig. 3 shows the impact of the number of Fourier coefficients. It is obvious that an increase in the number of coefficients, corresponds to a more detailed representation, though the additional details may be undesirable (noise or redundant variations of the same class).



(a) (b) (c)

Fig. 2: Examples of reconstruction for different instances of the same class ($n_\theta = 6$ & $n_c = 2$)



(a) $n_c = 2$ (b) $n_c = 4$

Fig. 3: Examples of reconstruction using different number of projection coefficients ($n_\theta = 6$)

D. Final Descriptor

The final feature vector is the concatenation of the projection-based descriptor applied in the initial binary image as well as in each of the four gradient orientation images (G_0,G_{45},G_{90} and G_{135}), as it is depicted in Fig. 4. Overall, the length of the descriptor is: 5 (images) × n_θ (projections) × $2n_c$ (Fourier coefficients). It should be noted that the proposed technique is not restricted to binary images since every substep can be easily generalized to grayscale images.

III. Experiments

A. Data Setup

For our experiments two handwritten character databases were used, the CIL Database [13] and the CEDAR character Database CD-ROM-1 [14], as well as the database GRPOLY-DB [15] of machine printed Historical Greek characters, publicly available at [16].

The CIL database comprises samples of 56 Greek handwritten characters written by 125 Greek writers. Every writer contributed 5 samples of each letter, thus resulting to a database of 625 variations per letter and an overall of 35000 isolated and labeled characters. Due to the similarity of several classes, especially due to size invariance of the methodology, a total of 10 pairs of classes were merged, randomly selecting 625 characters from each pair, as in [8]. The resulting 46 classes consist of 28750 characters, among which 23000 characters were used for training and the remaining 5750 characters for testing.

The CEDAR database consists of samples of 52 English handwritten characters, where 19145 characters were used for training and 2183 characters for testing. For the CEDAR database, four scenarios are distinguished for our experiments:

- 26 classes of uppercase characters
- 26 classes of lowercase characters
- 35 classes after merging similar classes, as in [8]
- 52 classes without merging

The GRPOLY-DB machine-printed dataset consists of isolated Greek polytonic characters extracted from the Hellenic parliament session proceedings. The Greek polytonic system has a variety of diacritic marks leading to a large number of classes. Only characters belonging to classes with at least 30 instances were selected, resulting to 125 classes in total. Two different scenarios were defined. According to the first scenario (SC-1), all instances were used (143051 instances) while at the second scenario (SC-2), only 30 randomly selected instances per class were used (3750 instances). In each scenario a 5-fold cross-validation was applied in order to evaluate the recognition rate.

B. Compared Methods

The state-of-the art techniques that were used in order to compare the efficiency of the proposed methodology are briefly described below:

- Adaptive zoning features method [7] (AdWin). According to this method, features are extracted after adjusting the

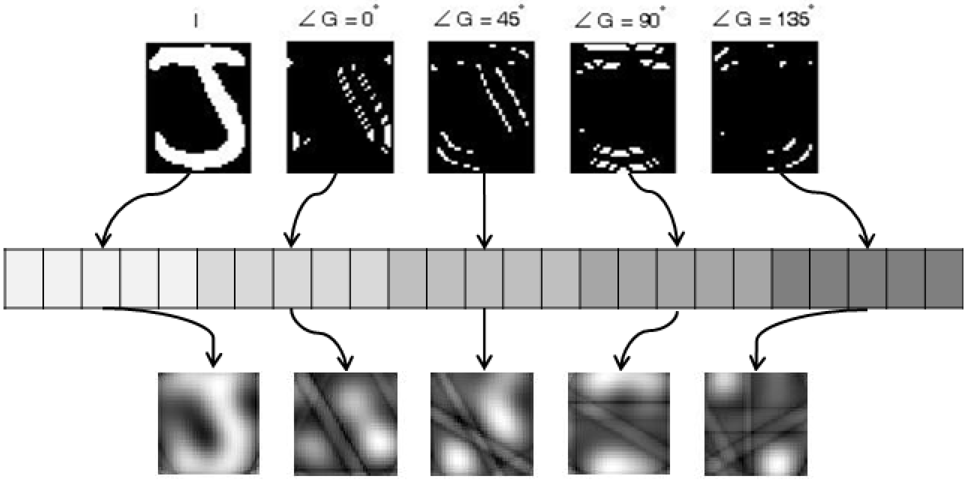


Fig. 4: Final descriptor as the concatenation of features of the initial image and its oriented gradients and their corresponding visualizations

position of every zone based on local pattern information, i.e. the maximization of the local pixel density around each zone. The classification for this feature descriptor is performed using a kNN classifier ($k = 3$).

- Histograms of Oriented Gradients method [10] (HOG). For each zone formed from a regular grid, a histogram of the local gradient orientation is constructed, where each bin corresponds to the frequency of a specific orientation in the zone. For binary images and the central difference gradient operator the orientations are limited, as it has been described previously. Multiclass SVM with RBF kernel is used as a classifier.
- Foreground sub-sampling method [8]. This method iteratively subdivides the image into four sub-images with approximately equal number of foreground pixels. The feature vector consists of the coordinates of the division points for a selected level of granularity (iteration steps). Further improvement to recognition rates is achieved by introducing a two-stage classification scheme (Vam2) comparing to a single step SVM classification (Vam1), where the information of different granularities is utilized.

C. Parameter Selection

The number of projections is not an essential parameter due to the lowpass filter equivalence of the image descriptor, thus, in order to avoid oversampling, we choose $n_\theta = 6$, or equivalently, we extract a projection every 30°.

Classification is performed using Support Vector Machines (SVM) [12], which is a machine learning technique for binary classification problems. The multiclass classification problem is reduced to multiple binary classification problems, applying the *one-versus-one* scheme. Specifically, a SVM is trained for each pair of classes and the classification is performed by voting.

In order to evaluate the efficiency of the descriptor while changing the number of the complex Fourier coefficients (n_c) , a linear SVM classifier is applied for the CEDAR (only for the merged scenario), CIL and GRPOLY-DB (only SC-2) datasets. The recognition results are depicted in Fig. 5. It is obvious that for both handwritten character databases the optimal choice is $n_c = 3$ (POG3). A further increase of the coefficients corresponds to an introduction of unnecessary details, hence increasing the within class variation and consequently hindering the classification step. For the machine-printed database, an increased number of complex coefficients is optimal ($n_c = 6$ - POG6), compared to the previous experiments. This is to be expected due to the large number of classes (125) as well as the minor differences of a variety of polytonic Greek diacritic marks (for example $\grave{a}, \acute{a}, \tilde{a}$). Furthermore, the variance between each class is limited, thus increasing the coefficients does not correspond to a significant decrease in the recognition rate as in the CEDAR and CIL databases.

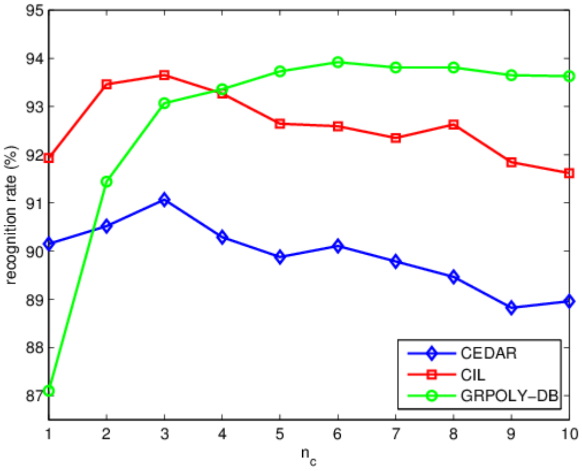


Fig. 5: Recognition rate for different number of Fourier coefficients

The final classifier is a multi-class (one-vs-one) SVM with RBF (Radial Basis Function) kernel, aiming to further increase the performance. A grid search was performed in order to find the optimal values for the parameters γ and C using cross-validation. For our experiments, the selected values were 0.05 and 8 for γ and C respectively.

D. Recognition Results

Table I shows the overall comparison results for the CEDAR database (4 scenarios), while Table II shows the recognition results for the CIL database. In both databases the proposed method (POG3) clearly outperforms the other techniques with only one classification step. Only the two-stage classification scheme (Vam2) in [8] performs better in the CEDAR database for the blended scenarios (both lowercase and uppercase characters), although has slightly worse performance in the remaining scenarios and in the CIL database. Nonetheless, the described technique in [8], as a feature descriptor along with the typical one-step classification (Vam1), displays a significant decrease in the recognition efficiency over the proposed one.

TABLE I: Recognition results for the CEDAR database

	AdWin	**HOG**	**Vam1**	**Vam2**	**POG3**
Uppercase (26 cl.)	85.30%	94.66%	93.78%	95.90%	96.04%
Lowercase (26 cl.)	83.70%	91.17%	89.79%	93.50%	93.68%
All (52 cl.)	69.90%	81.68%	78.42%	85.11%	82.73%
Merged (35 cl.)	82.46%	91.8%	90.70%	94.73%	93.59%

TABLE II: Recognition results for the CIL database

AdWin	**HOG**	**Vam1**	**Vam2**	**POG3**
87.97%	94.02%	92.53%	95.63%	96.14%

TABLE III: Recognition results for the GRPOLY-DB database

	AdWin	**HOG**	**POG3**	**POG6**
SC-1	97.71%	98.37%	98.49%	98.60%
SC-2	88.69%	92.00%	93.67%	94.35%

The recognition results for the two scenarios of the GRPOLY-DB database are shown in Table III. Specifically, we evaluate both the cases for $n_c = 3$ (POG3), which was the optimal for the other databases and the resulting descriptor is fairly small (180 features), and for $n_c = 6$ (POG6), which was chosen as the optimal for this database, even though it is a four times larger descriptor. As with the experimental results in the two previous databases, the proposed technique exhibits superior performance, while the POG6 descriptor shows better performance compared to POG3 (especially in SC-2, which is a more qualitative scenario) at the cost of the length of the feature vector.

IV. Conclusion

In this paper a new feature extraction method for isolated character recognition was presented based on projections of oriented gradients. The proposed methodology is scale invariant and thus no preprocessing step for normalization purposes is needed. As the experimental results indicate, the proposed feature extraction technique followed by a SVM classifier outperforms state-of-the-art techniques and has comparable results to techniques with a two-stage classification scheme. A possible future extension of the presented work, will be the substitution of the classification step with a more sophisticated one, either adding a second step to adjust the level of details, as in [8], or introducing a fusion technique in order to combine the proposed technique with a supplementary one, e.g. a zoning-based technique that focuses on local information.

Acknowledgments

This work has been supported by the OldDocPro project (ID 4717) funded by the GSRT .

References

[1] 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.

[2] D. Impedovo and G. Pirlo, “Zoning methods for handwritten character recognition: A survey”, Pattern Recognition, Vol. 47, Issue 3, pp. 969-981, 2014.

[3] A. L. Koerich and P. R. Kalva, “Unconstrained handwritten character recognition using metaclasses of characters”, IEEE International Conference on Image Processing (ICIP ’05), Vol. 2, pp. 542-545, 2005.

[4] A. Khotanzd and Y. H. Hong, “Invariant Image Recognition by Zernike Moments”, IEEE Transactions on Pattern Analysis Machine Intelligence, Vol. 12, Issue 5, pp. 489-497, 1990.

[5] J. H. Kim, K. K. Kim and C. Y. Suen, “Hybrid Schemes Of Homogeneous and Heterogeneous Classifiers for Cursive Word Recognition”, Proc. 7th International Workshop on Frontiers in Handwritten Recognition, Amsterdam, pp 433-442, 2000.

[6] N. Arica and F. Yarman-Vural, “An Overview of Character Recognition Focused on Off-line Handwriting”, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, Vol. 31, Issue 2, pp. 216-233, 2001.

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

[8] G. Vamvakas, B. Gatos and S. J. Perantonis, “Handwritten character recognition through two-stage foreground sub-sampling”, Pattern Recognition, Vol. 43, Issue 8, pp. 2807-2816, 2010.

[9] G. Beylkin, “Discrete radon transform”, Transactions on Acoustics, Speech and Signal Processing, IEEE , Vol. 35, Issue 2, pp. 162-172, 1987.

[10] N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection”, Proc. IEEE Conference Computer Vision and Pattern Recognition, Vol. 2, pp. 886-893, 2005.

[11] D. E. Dudgeon and R. M. Mersereau, “Multidimensional Digital Signal Prosecing”, Prentice-Hall, 1984.

[12] C. Cortes and V. Vapnik, “Support-vector network”, Machine Learning, vol. 20, pp. 273-297, 1997.

[13] G. Vamvakas, B. Gatos, S. Petridis and N. Stamatopoulos, “An Efficient Feature Extraction and Dimensionality Reduction Scheme for Isolated Greek Handwritten Character Recognition”, 9th International Conference on Document Analysis and Recognition, Brazil, pp. 1073-1077, 2007.

[14] J.J. Hull, “A database for handwritten text recognition research”, IEEE Transactions on Pattern Analysis Machine Intelligence, Vol. 16, Issue 5, pp. 550-554, 1994.

[15] B. Gatos, N. Stamatopoulos, G. Louloudis, G. Sfikas, G. Retsinas, V. Papavassiliou, F. Simistira and V. Katsouros, “GRPOLY-DB: An Old Greek Polytonic Document Image Database”, 13th International Conference on Document Analysis and Recognition, Tunisia, 2015.

[16] http://www.iit.demokritos.gr/~nstam/GRPOLY-DB