Using Attributes for Word Spotting and Recognition in Polytonic Greek Documents

Giorgos Sfikas[1], Angelos P. Giotis[1,2], Georgios Louloudis[1], Basilis Gatos[1]

[1] Computational Intelligence Laboratory, Institute of Informatics and Telecommunications
National Center for Scientific Research "Demokritos", GR-15310 Agia Paraskevi, Athens, Greece
{sfikas, louloud, bgat}@iit.demokritos.gr
[2]Department of Computer Science and Engineering, University of Ioannina, Greece
agiotis@cs.uoi.gr

***Abstract*—Word spotting and recognition are among the most important applications used today in the field of document processing and text understanding. In word spotting, the goal is to search a scanned document for instances of a specific word. In word recognition, we aim to identify the transcription of the document words. While substantial work in both topics has been published, not all are readily adaptible to scripts other than a specific script and/or language. This is especially true for documents written in the polytonic greek script, a script used to write the greek language during a period that approximately spans two millenia. In this work, we extend the attribute-based model for word spotting and recognition recently presented in [1] for use with polytonic greek documents. To this end, we present three alternative ways to extend the model mechanism to handle the greek alphabet and its various combinations of diacritic marks. We have run numerical experiments over polytonic machine-printed and handwritten documents for word spotting and recognition. The proposed model is shown to outperform other state-of-the-art methods in word spotting trials. Regarding polytonic greek unconstrained handwritten word recognition, to the best of our knowledge, this is the first work to address the problem succesfully.**

I. Introduction

Libraries and individual researchers are interested in using digitised versions of historical documents. To this end, document processing techniques have given important results in the past 20 years. A variety of methods has been proposed to preprocess, binarize, deskew, deslant, and analyze the layout of the scanned image [2]. After segmenting the document up to character level, an OCR technique is traditionally applied to recognize the document text. This approach however has its shortcomings. Handwritten text is known to be more challenging than machine-printed text, as segmentation to the character level is usually problematic, while recognition performance can suffer due to the variety of writing styles.

Word spotting, or searching for a specific word within a document, has been proposed as an alternative to documents that are too hard to recognise with acceptable accuracy, or where performing full recognition is not necessary [2]. Word spotting has been introduced to document processing as an adaptation of a related technique used originally in speech processing.

Historical manuscripts comprise a document category that poses serious challenges for proper processing of the document, and documents written in the polytonic greek script are no exception to this rule. Polytonic greek is a script that has been used to write the greek language throughout various stages of its evolution, since the introduction of the script as a standard in the late antiquity and up until modern times. The polytonic greek script is based on the greek alphabet, and comprises both capital and small versions of the letters, plus the addition of special diacritics that are placed above or below the respective letters [3]. It has been practically the sole script available to write greek until the introduction of monotonic greek -a simplified version of polytonic greek- in 1982. Thus, it can be well understood that a huge amount of both handwritten as well as machine-printed documents exists in polytonic greek.

Not much work in the literature of text understanding is targeted to polytonic greek, albeit the volume and academic importance of many of the available texts. Commercial OCR systems do exist for monotonic, printed greek, but processing of polytonic printed text is known to give poor results [4]. Word spotting techniques that use learning-free, zoning features are proposed in [4], [5]. The elaboration of recognition or spotting techinques for polytonic greek handwritten texts remains a challenge largely unaddressed. One exception to this rule is [6], where an OCR system for early christian greek documents is proposed. The target documents are written in a form of polytonic greek, but the proposed model is finetuned towards the specific writing style and conventions of the given era and context, thus constraining its scope of use.

In this paper we present a method for word spotting and recognition of handwritten and machine-printed documents written in polytonic greek. The work presented in this paper is largely based on the state-of-the-art model that was presented in [1] and won the segmentation-based track at the recent H-KWS 2014 word spotting competition [7]. The current work can essentially be seen as an extension of this previous work for the polytonic greek script. In [1], a learning-based model is presented for segmentation-based word spotting. A training set is required, where each word image is to be supplied with a transcription. The word image data is used to create a Fisher Vector (FV) descriptor [8], while the transcription is used to create a histogram-based descriptor that the authors name Pyramidal Histogram of Characters (PHOC). PHOC records essentially the appearance of a specific letter in the transcription, a strategy that builds on the concept of attribute-based models used for natural image understanding in the related literature [9]. The two descriptor sets are used together to learn a projection to a new space and create a new, single

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

descriptor based on the scaled output of a structured SVM. The output fixed-length descriptors can then simply be compared to each other and to descriptors from a test set using the Euclidean distance. The result is efficient word spotting, which when coupled with a lexicon can be used for word recognition. Also, due to the attribute-based structure of the model, words that do not appear in the training set can also be retrieved and/or recognized. In [10], a closely related model to [1] is proposed for text recognition, and a completely analogous transcription descriptor is used in the same spirit with the PHOCs of [1].

The transcription descriptor (PHOC) proposed in [1] is script and language-dependent, as it is comprised of a bin for each script character plus bins for the most likely language bigrams. This means that there can be no comparison between words of different scripts and/or languages and that adaptation to complex scripts such as the polytonic greek script, that also comprise diacritics of various types, is not necessarily straightforward. In this paper, we address the latter issue. We propose and compare three different ways to extend PHOC to polytonic greek, and with it the model of [1]. We test the proposed scheme on word spotting trials, over handwritten as well as machine printed greek documents. The results show superior performance compared to state-of-the-art learning-free methods. We also showcase the model's performance on word recognition, a task which, to the best of our knowledge, has not been yet addressed in the case of handwritten polytonic greek text.

The structure of the remainder of the paper is as follows. In section II, we review the basic components of model, the image and transcription representations, and the model mechanism. In section III, we concisely present the polytonic greek script, its diacritics and particularities, and describe alternative ways to model it in the form of an extended PHOC descriptor. Experimental results on word spotting and recognition are presented in section IV, whereas conclusions are discussed in section V.

II. Base model description

In this section we review the data representation and pipeline of the model introduced in [1]. The basic framework to represent word images is the Fisher Vector description [8]. We assume that our input is already segmented at word level. For each image we extract dense SIFT descriptors [11]. A Gaussian mixture model (GMM) is trained using SIFT descriptors from all input images, and Fisher vectors are calculated for each image as a function of their SIFT description and the gradients of the GMM with respect to its parameters. This results to a fixed-length, highly discriminative representation, that can be seen as an augmented bag of visual words description that encodes higher order statistics. Fisher vectors have been used previously with success in various fields of computer vision [8], [12].

The Fisher vector representation is shown to give good results on word spotting when used as a standalone descriptor on a Query-by-Example (QBE) setting [1]. However, if the presence of a training set of word images is assumed, for which ground truth transcriptions are known, a more discriminative descriptor can be created. The proposed descriptor is based on the concept of attributes, which have recently gained increasing popularity in the machine vision literature [9], [13], [14]. Attributes are semantic properties defined over images and categories, and in effect are used as labels that denote the presence or absence of a specific feature. An image attribute may be defined for example as, "does this image contain a person?", or "is this object brown/shiny/furry?". Attributes also allow for zero-shot learning, where new, unseen instances of images or classes can be correctly processed.

In the context of document processing, attributes can be defined on word images on the basis of appearance or not of a specific letter. The set of attributes each of which is defined as a function of the presence or absence of a specific alphabet letter in the word can be aggregated to a single multivariate binary vector. This descriptor can be duplicated to answer if specific letters are found on the first or second half of the word, and so on for any partition of the word transcription into k equal parts. The subsequent aggregation of higher levels of this set of attributes makes up for the Pyramidal Histogram of Characters (PHOC) representation, where more attributes are added to capture the presence of letter bigrams. In [1], this scheme is applied on the 26 letters of the English language plus bins for digits and the 50 most frequent bigrams of the English language, leading to a 604-variate vector.

The Fisher vector representation of the word images and the PHOC representation of the word transcriptions are subsequently combined to create a single, more discriminative descriptor. For each variate i of the PHOC vector, a Support Vector Machine (SVM) [15] is trained using all Fisher vectors as inputs, labelled according to attribute i. The model parameters for each SVM are saved and can be then used to calculate attribute outputs for unseen data. Such data are typically non-training set data, for which their Fisher vector can be computed since it depends on image data, while their PHOC vector cannot be computed since the transcription is unknown. The output of the Structured SVM model parameters given some Fisher vector gives an output attribute vector that has the same dimensionality as the PHOC vector.

Summing things up, for every training point n we would have a Fisher vector representation f_n, a PHOC binary representation p_n, and an attribute vector representation a_n. For non-training points only the FV representation f_n and attribute vector a_n is available. The attribute vector can be used as a valid feature vector and can be compared against other attribute vectors simply by calculating their Euclidean distance, making Query by Example (QBE) word spotting possible. Also, comparing the PHOC representation of a query against attribute vectors is also possible since both vectors are of the same dimensionality, allowing for Query by String (QBS) [1]. However, in both cases it is desirable to apply a notion of scaling or calibration over the PHOC and attribute vectors, since (a) PHOC vectors and attribute vectors are not necessarily comparable in principle, even if of the same dimensionality, (b) vector variates are not necessarily commensurate, since training of each element of the Structured SVM is done independently from others, leading some of the outputs to possibly dominate over the others and (c) the inter-bin correlation is not taken into account. In the light of this, Canonical Correlation Analysis (CCA) [15] can be applied with the PHOC vectors p_n and attribute vectors a_n as its input views. In CCA, a projection is calculated for each view that maximizes correlation between vectors in the projected space. Formally, we are looking for projection vectors w_p, w_a that maximize

$\arg\max_{w_p,w_a} \frac{w_p^T C_{pa} w_a}{\sqrt{w_a^T C_{aa} w_a}\sqrt{w_p^T C_{pp} w_p}}$ where C_{aa}, C_{pp}, C_{ap} are respectively sample covariance matrices between vectors in the set of attribute descriptors, vectors in the set of PHOC descriptors, and cross-covariance between the two latter sets. In practice, we are looking to combine a series of k optimal orthogonal projection vectors $w_{a1}, \cdots, w_{ak}, w_{p1}, \cdots, w_{pk}$ to project our views to a k-dimensional space. It can be shown that the required projection vectors are given by identifying vectors $w_{a1}, \cdots, w_{ak}$ with the k leading eigenvectors of matrix $C_{aa}^{-1} C_{ap} C_{pp}^{-1} C_{pa}$, and vectors $w_{p1}, \cdots, w_{pk}$ with the k leading eigenvectors of matrix $C_{pp}^{-1} C_{pa} C_{aa}^{-1} C_{ap}$. Embedding inputs to an appropriate feature space before applying CCA is equivalent to a kernel version of CCA (KCCA) and has given the best experimental results (with a random Fourier feature mapping [16], corresponding to a Gaussian kernel embedding [1]).

III. Polytonic word description

A. Polytonic greek script

The polytonic greek script has been introduced in the late antiquity to write the greek language [3]. It is comprised of the standard 24 greek letters, in upper-case and lower-case versions of the characters. Also, a number of diacritics are used. These have originally been introduced in the script with the rationale of aiding the reader with proper pronunciation of the words, while in later phases of evolution of the greek language they have retained largely only an orthographic and etymological value. These diacritics are the smooth and rough breathing, the accute, grave and circumflex accent, the subscript and the diaeresis. A visual example of these diacritics can be seen in figure 1. These diacritics can have a combined appearance on

Diacritic type			Usage examples	
Breathings	Smooth	᾽	ἀ	ἀποκάλυψη
	Rough	῾	ἁ	ἱστορία
Accents	Acute	́	ά	πάτερ
	Grave	̀	ὰ	τὸν
	Circumflex	͂	ᾶ	κλασικοῦ
Subscript		ͅ	ᾳ	χριστῷ
Diaeresis		¨	ϋ	λαΐκὸς

Fig. 1. Polytonic greek diacritics.

the same character or on the same word, according to a certain set of grammatical rules. Further discussion of these rules is out of the scope of this paper.

B. Extending PHOC

In adapting PHOC to work with polytonic greek, our basic problem is what would be the most efficient way to integrate the use of polytonic diacritics in the word transcription representation. To this end, we propose three possible alternatives. We dub these (i) Atonic PHOC (A-PHOC), (ii) Polytonic Header PHOC (PH-PHOC) and (iii) Mixed Bin PHOC (MB-PHOC). The difference of each one to the other is to the number and meaning of the bins used for the descriptor.

In Atonic PHOC we use 24 bins for letters at the base level of the descriptor. Each one corresponds to a single letter of the greek alphabet. All letters can appear in two forms, capital or lowercase; capital and lowercase letters are therefore merged to the same bin, making the model case-insensitive. The letter sigma (Σ, σ) is an exception to this rule, as it can appear in one extra form besides its capital and lowercase forms, that of the final sigma (ς). This is also merged on the same bin with the other forms of the letter. We also add bins for numerical digits and bigrams. The 50 most frequent bigrams of the greek language are added at level 2, and the rest of the bins are iterated at levels 2, 3, 4 and 5. Level 1 histograms are excluded altogether. The most frequent bigrams are extracted by processing a corpus of 34 million greek words (corpus "C", [17]). The total number of bins of Atonic PHOC therefore sums up to $(2+3+4+5)*(24+10)+2*50 = 576$ bins. Polytonic diacritics are ignored altogether, so a letter with no diacritics uses the same bin as the same letter with any diacritics added to the letter. In the sense that the descriptor bins correspond to letters + digits + bigrams, A-PHOC can be understood as the conceptually closest to, or the most straightforward adaptation of the standard PHOC descriptor of [1] for greek.

In Polytonic header PHOC we add information about polytonic diacritics, in the form of a number of extra bins added to what we described as Atonic PHOC. These are 7 bins, each one corresponding to the appearance or not of a diacritic in the word. We do not reiterate their use to higher levels of the PHOC pyramid. This choice makes the descriptor non-spatially aware when it comes to polytonic diacritics, at a gain of smaller descriptor length. The rationale of this choice is founded on the fact that many of the diacritics are gramatically constrained to appear at fixed positions at the beginning or the end of the word, with some rare exceptions. The Polytonic header PHOC is $576 + 7 = 583$ bins long.

With the third proposed alternative, Mixed Bin PHOC, we consider each letter, with all combinations of diacritics, as a separate case. Our histogram of base does not comprise 24 bins as in A-PHOC and PH-PHOC, but one bin for each combination of letter and polytonic diacritic. For example, α with no diacritics is assigned to a different bin than α with a smooth breathing, while α with a smooth breathing and an acute accent is assigned to a third bin, and so on. This sums up to 128 extra bins to the already existing 24 for the plain versions of letters. The size of the descriptor totals to $(2+3+4+5)*(128+24+10)+2*50 = 2368$ bins.

IV. Experimental results

We have run numerical experiments over databases of polytonic greek documents, both machine-printed and handwritten [18]. For these documents, the original text as well as a binarized version of all document pages and word-level segmentations along with ground truth transcriptions for each word is available [19]. We have available one set of handwritten documents and two sets of machine-printed documents. We shall refer to these sets in this work as *Memoirs*, *Gazette* and *Proceedings* [1] . Excerpts from these sets can be seen in figures 2 and 3. Our handwritten set, *Memoirs*, consists of 46 pages segmented into 4941 word images. The text is written by a single author in the late 19^{th} century, and is part of the memoirs of Sophia Trikoupi, sister of the important greek prime minister Charilaos Trikoupis. The machine printed text *Gazette* consists of 5 pages segmented into 5004 word images. This text contains pages taken from the official journal of the

[1]These sets are referred to as GRPOLY-DB-Handwritten, GRPOLY-DB-MachinePrinted-A and GRPOLY-DB-MachinePrinted-B respectively in their original publication in [18].

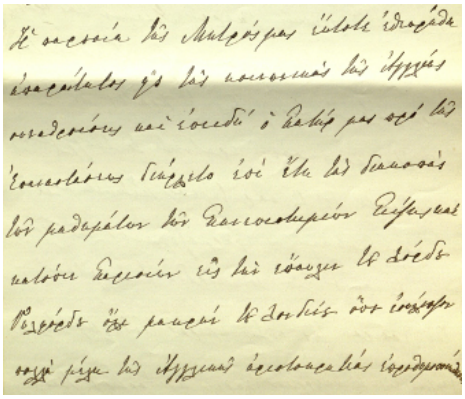
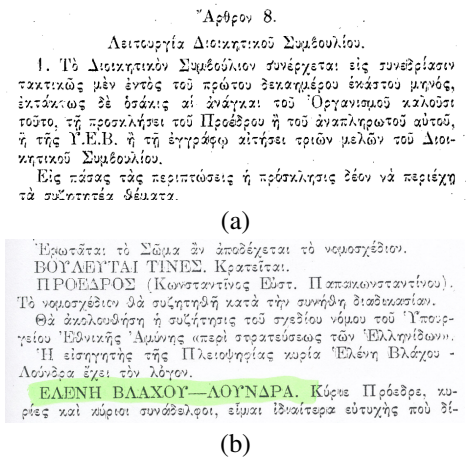


Fig. 2. Handwritten polytonic text sample, "*Memoirs*". Excerpt from the memoirs of Sophia Trikoupi (1838-1916).

greek government describing laws and edicts, published from the mid-19^{th} to the mid-20^{th} century. The machine printed text *Proceedings* is made up of 33 pages segmented into 26783 word images and records various speeches delivered in the greek parliament within almost the same time period as the dataset *Gazette*. All texts are therefore of historical value, and written in the polytonic greek script.



"Αρθρον 8.

Λειτουργία Διοικητικοῦ Συμβουλίου.

1. Τὸ Διοικητικὸν Συμβούλιον συνέρχεται εἰς συνεδρίασιν τακτικῶς μὲν ἐντὸς τοῦ πρώτου δεκαημέρου ἑκάστου μηνός, ἐκτάκτως δὲ ὁσάκις αἱ ἀνάγκαι τοῦ Ὀργανισμοῦ καλοῦσι τοῦτο, τῇ προσκλήσει τοῦ Προέδρου ἢ τοῦ ἀναπληρωτοῦ αὐτοῦ, ἢ τῆς Υ.Ε.Β. ἢ τῇ ἐγγράφῳ αἰτήσει τριῶν μελῶν τοῦ Διοικητικοῦ Συμβουλίου.

Εἰς πάσας τὰς περιπτώσεις ἡ πρόσκλησις δέον νὰ περιέχῃ τὰ συζητητέα θέματα.

(a)

Ἐρωτᾶται τὸ Σῶμα ἂν ἀποδέχεται τὸ νομοσχέδιον.

ΒΟΥΛΕΥΤΑΙ ΤΙΝΕΣ. Κρατεῖται.

ΠΡΟΕΔΡΟΣ (Κωνσταντῖνος Εὐστ. Παπακωνσταντίνου). Τὸ νομοσχέδιον θὰ συζητηθῇ κατὰ τὴν συνήθη διαδικασίαν.

Θὰ ἀκολουθήσῃ ἡ συζήτησις τοῦ σχεδίου νόμου τοῦ Ὑπουργείου Ἐθνικῆς Ἀμύνης «περὶ στρατεύσεως τῶν Ἑλληνίδων».

Ἡ εἰσηγήτρια τῆς Πλειοψηφίας κυρία Ἑλένη Βλάχου - Λούνδρα ἔχει τὸν λόγον.

ΕΛΕΝΗ ΒΛΑΧΟΥ—ΛΟΥΝΔΡΑ. Κύριε Πρόεδρε, κυρίες καὶ κύριοι συνάδελφοι, εἶμαι ἰδιαίτερα εὐτυχὴς ποὺ βί-

(b)

Fig. 3. Machine-printed polytonic text samples. (a) "*Gazette*". Excerpt from the official journal of the greek government. (b) "*Proceedings*". Excerpt from the proceedings of the greek parliament.

We have tested the proposed models in word spotting trials as well as in word recognition. In the word spotting scenario, we evaluate methods using the Mean Average Precision (MAP) benchmark [7] over a selected set of queries taken from each database. We choose the set of queries in each case on the basis of word length and appearance frequency, following [7]. For the handwritten *Memoirs* set we choose all words that have more than 5 letters and 4 instances as queries, for a total of 21 queries. For the machine-printed *Proceedings* we choose all words that have more than 6 letters and 5 instances as queries, for a total of 103 queries respectively. Results for all instances of each query class were averaged to calculate the total MAP in the QBE word spotting tests.

We have used two different evaluation settings in our word spotting experiments. In the first setting, we train our model using a part of the handwritten *Memoirs* dataset and use the rest as our test set (the test set comprises 2000 words [2]). Only queries that are situated in the test set are used as queries, and are matched against only word images within the test set. The 21 selected query classes correspond to 50 query word image instances. In the second spotting setting, we train and test on different texts. We train our model using the full *Gazette* set and test on the *Proceedings* set. In this manner, the model capability to generalize its training over a different test set is also evaluated. While it would be interesting to use a similar setting for handwritten texts also, unfortunately only one corpus of handwritten polytonic greek was available to us (*Memoirs*) at the time of writing of this paper. Note also that the 103 selected query classes of *Proceedings* correspond to 959 query word image instances.

TABLE I. QBE WORD SPOTTING RESULTS (MAP%).

Method	*Memoirs*	*Gazette*/*Proceedings*
A-PHOC	81.8%	52.5%
PH-PHOC	85.2%	56.6%
MB-PHOC	**96.6%**	**74.4%**
Adaptive zoning	60.8%	57.8%
Profiles+DTW	69.6%	62.0%

Results for our Query-by-Example (QBE) word spotting trials can be seen in table I. We also show results comparing the proposed schemes against two state-of-the-art methods learning-free methods, adaptive zoning [20] and profile features with Dynamic Time Warping (DTW) [21]. Concerning the tests over *Memoirs*, all proposed schemes outperform considerably the learning-free methods. This is not the case with the *Gazette*/*Proceedings* scenario, where only MB-PHOC gives better results than the learning-free methods. This is not surprising, since the first scenario uses training and test sets taken from the same base document of the same writer. Also, performance variance has shown to be high in this latter scenario, with parts of the text corresponding to excellent results, while others corresponding to very mediocre results. We must assume that this variance is related to the similarity of the font in the given part of *Proceedings*, with the fonts used in *Gazette*, which has been used for training. In all cases, PH-PHOC is better than A-PHOC, giving a difference of about 4% with A-PHOC, validating the utility of adding the polytonic header bins to the descriptor, at almost the same cost in terms of training time. MB-PHOC on the other hand outperforms all other methods in all cases with a considerable difference of about $11 - 12\%$ from the second winner.

We have run an experiment to test the robustness of MB-PHOC when used with a specific type of entering string queries. We have run Query-by-String (QBS) spotting trials on the handwritten *Memoirs*. We used the same criterion to select string queries as the one used for QBE. We compared two alternative scenarios for running a QBS query: the first scenario assumes the same set of queries as the ones in the QBE test, that is using the frequency/length criterion. The second scenario assumes the same queries but omitting all diacritics from them. These tests correspond to a scenario where the end-user of the document retrieval system would be unsure of the correct diacritics to use for his query. MB-PHOC has given a MAP of 81.3% in the first scenario against 79.6% in the second scenario, proving to be quite robust despite the

[2]Indices of words used for training, test and validation respectively: $1 - 2000$, $2001 - 4000$, $4001 - 4941$.

fact that bins of the same character with and without diacritics are, implementation-wise, unrelated.

We have also run word recognition trials on the handwritten dataset. To perform recognition, we use a lexicon that is made up of the transcriptions appearing in the full dataset. Each PHOC descriptor of the lexicon words, after embedding and projecting to the CCA space, is compared to each of the words of the test set. This corresponds to a lexicon size of 2141 unique words. The representation of the lexicon word that gives the smallest Euclidean distance is selected as the recognized transcription. We used the same training/test set settings for the trials performed for word spotting, and the MB-PHOC model. Recognition accuracy is evaluated over the test set words, measured as the percentage of correctly recognized words. In *Memoirs*, this amounts to 2000 total words in the test set, out of which 1518 have been correctly recognized and 482 missed, for a correct word ratio of 75.9%. Note that we count as a miss the difference of one or more letters or diacritics between the recognized word and the ground truth. An example of recognition result can be seen in fig.4.

ἡ παρουσία τῆς Μητρός μας ἔκτοτε ἐθεωρήθη
ἀπαραίτητος εἰς τὰς κοινωνικὰς τῆς Ἀγγλίας
συναθροίσεις καὶ ἐπειδὴ ὁ Πατήρ μας πρὸ τῆς
Ἐπαναστάσεως διήρχετο ἐπὶ ἔτη τὰς διακοπὰς
τῶν μαθημάτων τῶν Πανεπιστημίων Πατρὸς καὶ
κατόπιν Παρισίων εἰς τὴν ἐποχὴν τοῦ λόρδου
Γυλφόρδου, ὄχι μακρὰν τοῦ Λονδίνου ὅπου ἐσύχναζον
πολλὰ μέλη τῆς Ἀγγλικῆς ἀριστοκράτης ὡρισμένην

Fig. 4. Word recognition result for the excerpt from "*Memoirs*" shown in fig.2. All words are recognized correctly except for the words coloured red. Note that in this figure we manually placed the recognition result for each segmented word image in a line-to-line and word-to-word correspondence with fig.2 to ease comparison of results and ground truth by the reader.

V. Conclusion

This paper has addressed the problem of word spotting and recognition of polytonic greek texts. We have proposed three different ways to adapt the attribute-based model of [1] for polytonic greek, which correspond to three different transcription representations. Experiments have shown that including information about polytonic diacritics always gives better results. A-PHOC is the most naive adaptation of [1], and closest to the original PHOC descriptor in the sense that it uses bins for letters, digits and bigrams, completely disregarding polytonic diacritics. The PH-PHOC representation includes polytonic information using a short information header, which is low-cost and character-independent. PH-PHOC outperformed A-PHOC at the price of only a few extra variates added in the descriptor. The last proposal, MB-PHOC, includes feature vector variates that correspond to all valid combinations of letter and diacritic, and has been shown to be universally the most efficient choice, albeit with a high computational cost in the training phase. It has also shown to be robust in the case that a query string comprising no diacritics is used. This latter scenario may be very relevant today, if one takes into account the declining familiarity of users of modern greek with the polytonic greek script. Finally, we have used our model adaptation for recognition. To the best of our knowledge, this is the first work to succesfully address this problem for unconstrained handwritten polytonic greek texts.

Acknowledgment

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

References

[1] J. Almazán, A. Gordo, A. Fornés, and E. Valveny, "Word spotting and recognition with embedded attributes," *IEEE TPAMI*, vol. 36, no. 12, pp. 2552–2566, Dec 2014.

[2] D. Doermann, K. Tombre *et al.*, *Handbook of Document Image Processing and Recognition*, 2014.

[3] G. Horrocks, *Greek: A History of the Language and its Speakers*. John Wiley & Sons, 2009.

[4] T. Konidaris, B. Gatos, K. Ntzios, I. Pratikakis, S. Theodoridis, and S. Perantonis, "Keyword-guided word spotting in historical printed documents using synthetic data and user feedback," *IJDAR*, vol. 9, no. 2-4, pp. 167–177, 2007.

[5] A. Kesidis, E. Galiotou, B. Gatos, A. Lampropoulos, I. Pratikakis, I. Manolessou, and A. Ralli, "Accessing the content of greek historical documents," in *Proceedings of The Third Workshop on Analytics for Noisy Unstructured Text Data*. ACM, 2009, pp. 55–62.

[6] K. Ntzios, B. Gatos, I. Pratikakis, T. Konidaris, and S. J. Perantonis, "An old greek handwritten ocr system based on an efficient segmentation-free approach," *IJDAR*, vol. 9, no. 2-4, pp. 179–192, 2007.

[7] I. Pratikakis, K. Zagoris, B. Gatos, G. Louloudis, and N. Stamatopoulos, "ICFHR 2014 competition on handwritten keyword spotting (H-KWS 2014)," in *2014 International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 2014, pp. 814–819.

[8] J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek, "Image classification with the fisher vector: Theory and practice," *IJCV*, vol. 105, no. 3, pp. 222–245, 2013.

[9] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth, "Describing objects by their attributes," in *Proceedings of IEEE CVPR*, 2009, pp. 1778–1785.

[10] J. Rodriguez and F. Perronnin, "Label embedding for text recognition," in *Proceedings of the British Machine Vision Conference*. BMVA Press, 2013.

[11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[12] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Fisher vector faces in the wild," in *Proceedings of the British Machine Vision Conference*, vol. 1, no. 2, 2013, p. 7.

[13] M. Douze, A. Ramisa, and C. Schmid, "Combining attributes and fisher vectors for efficient image retrieval," in *Proceedings of IEEE CVPR*, 2011, pp. 745–752.

[14] M. Danelljan, F. Shahbaz Khan, M. Felsberg, and J. Van de Weijer, "Adaptive color attributes for real-time visual tracking," in *Proceedings of IEEE CVPR*, 2014.

[15] T. Hastie, R. Tibshirani, J. Friedman, T. Hastie, J. Friedman, and R. Tibshirani, *The elements of statistical learning*. Springer, 2009, vol. 2, no. 1.

[16] A. Rahimi and B. Recht, "Random features for large-scale kernel machines," in *Advances in Neural Information Processing Systems*, 2007, pp. 1177–1184.

[17] A. Protopapas, M. Tzakosta, A. Chalamandaris, and P. Tsiakoulis, "IPLR: An online resource for Greek word-level and sublexical information," *Language resources and evaluation*, vol. 46, no. 3, pp. 449–459, 2012.

[18] B. Gatos, N. Stamatopoulos, G. Louloudis, G. Sfikas, G. Retsinas, F. Simistira, V. Papavassiliou, and V. Katsouros, "GRPOLY-DB: An old greek polytonic document image database," in *Proceedings of ICDAR*, 2015.

[19] [Online]. Available: http://www.iit.demokritos.gr/~nstam/GRPOLY-DB

[20] B. Gatos, A. L. Kesidis, and A. Papandreou, "Adaptive zoning features for character and word recognition," in *Proceedings of ICDAR*, 2011, pp. 1160–1164.

[21] T. Rath and R. Manmatha, "Word spotting for historical documents," *IJDAR*, vol. 9, no. 2-4, pp. 139–152, 2007.