



Identifying the writer of ancient inscriptions and Byzantine codices. A novel approach [☆]



C. Papaodysseus ^{a,*}, P. Rousopoulos ^b, F. Giannopoulos ^a, S. Zannos ^a, D. Arabadjis ^a, M. Panagopoulos ^c,
E. Kalfa ^a, C. Blackwell ^d, S. Tracy ^e

^a National Technical University of Athens, Department of Electrical and Computer Engineering, 9 Heron Polytechniou, GR-15773 Athens, Greece

^b Technical Institute of Chalkida, Department of Automatic Control, Chalkida, Euboea, Greece

^c Ionian University, Department of Music Studies, Corfu, Greece

^d Furman University, Department of Classics, Greenville, SC, USA

^e Institute for Advanced Study, School of Historical Studies, Princeton, NJ, USA

ARTICLE INFO

Article history:

Received 26 July 2012

Accepted 20 January 2014

Available online 5 February 2014

Keywords:

Writer identification

Dating ancient inscriptions

Dating Byzantine codices

Curve fitting

Curvature

Letters similarity

Pattern recognition for manuscript

classification

ABSTRACT

In this paper, a novel methodology is presented aiming at the automatic identification of the writer of ancient inscriptions and Byzantine codices. This identification can offer unambiguous dating of these ancient manuscripts. The introduced methodology is also applicable to contours of complexes of letters or any class of similar curves. The method presented here initially estimates the normalized curvature at each pixel of a letter contour. Subsequently, it performs pair-wise comparisons of the curvatures sequences that correspond to two realizations of the same alphabet symbol. Then, it introduces a new Proposition that, on the basis of the previous results, offers a closed solution to the problem of matching two equinumerous digital contours in the Least Squares sense. Next, a criterion is employed quantifying the similarity of two realizations of the same alphabet symbol. Finally, a number of statistical criteria are introduced for the automatic identification of the writer of ancient manuscripts. The introduced method did not employ any reference manuscript neither the number of distinct hands who had written the considered set of manuscripts nor any related information whatsoever; it also performs quite efficiently even if a small number of realizations (less than 6) of certain alphabet symbols appear in a tested document. The only a priori knowledge is the alphabet of the language under consideration. We would like to stress that otherwise the method does not depend at all on the language itself. Namely it does not take into account if the alphabet is Latin, Greek, Etruscan, etc. The methodology and the related, developed information system has been applied to 46 ancient inscriptions of the Classical and Hellenistic era and 23 Byzantine codices, offering 100% accurate results, in the sense that the obtained results are in full agreement with prominent scholars in the field of Archaeology, History and Classical Studies.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

1.1. The importance of identifying the writer of ancient inscriptions and manuscripts

The most important source for the science of History is the preserved set of written documents. Concerning Antiquity, this set mainly includes ancient stone inscriptions [1–3] and manuscripts. Dating the content of these documents is absolutely crucial for History and Archaeology. One of the most prominent historians, Professor Christian Habicht has recently written, “Proper historical

use of inscriptions can only be made if they can be dated”. However, writers of ancient inscriptions and manuscripts, as a rule, did not sign or date their documents. On the other hand manuscripts are the handwritten texts on papyrus or parchment that were the vehicles transmitting the literature of the ancient world through the middle ages, until mechanical reproduction by printing presses took over. These documents pose questions similar to those raised by inscriptions, and some new ones. For example, the Homeric Iliad survives mainly through a handful of large, beautiful manuscript volumes. These were all produced in Constantinople during the 10th or 11th century, but made their way to different libraries in Europe: Venice, El Escorial in Spain, London, Geneva, Florence, Rome. Each of these manuscripts, too, contains many different texts – the text of Homer’s poem, and several different commentary texts, as well as shorter notes in the margins and

[☆] This paper has been recommended for acceptance by Daniel Lopresti.

* Corresponding author.

E-mail address: cpapaod@cs.ntua.gr (C. Papaodysseus).

between lines. One main goal of the present paper is to perform quantitative analysis on the scribal hands, so that the relationships among these volumes and their relative dates of production are obtained. Most importantly, the results obtained in this paper might be able to ascertain something of the nature of the source-documents these scribes were working from as they produced these deluxe, comprehensive editions. So far, as a rule, dating the content of ancient inscriptions and manuscripts is a very difficult task and it is based on the scholars' instinct and frequently subjective considerations [1]. Thus, it is not a surprise that ancient documents dating, often causes scientific disputes and disagreements.

1.2. The goal of the present work: A new approach to dating ancient documents via the identification of their writer

The idea emerged among the authors to achieve dating of ancient documents by means of automatic identification of their writer. In fact, as far as ancient inscriptions are concerned, their writers carved the stones as a career; in other words, writing inscriptions was a profession in antiquity. This has the following important consequences: (a) The number of different hands who cut the stones to write a document, was particularly limited; thus, one may expect that the tenths of thousands of inscriptions unearthed in Attica may belong to few hundreds of different writers. (b) If one succeeds in attributing a set of inscriptions to a writer, then it is very easy to determine his career period, since, as a rule, the content of at least one inscription will reveal it. Evidently, the greater the number of the inscriptions correctly attributed to this hand, the greater is the probability to date his career. (c) If a new inscription is found and/or studied, the content of which cannot be dated, if the developed methodology attributes it to a writer, then the content of the inscription gains a date immediately, which is clearly the time period during which the writer was active professionally. We note that the working careers of most ancient writers covered about 20–25 years.

Similar remarks hold true for those who preserved ancient texts by copying them manually. For example, as has already been pointed out, a few hands preserved Homers' poems by copying them in Constantinople during the 10th or 11th century; a few other writers added their comments in the manuscripts. If one may by automatic means identify the various hands that had written text in these documents, one immediately obtains a chronological classification as well as additional most valuable information as stated in the previous Section 1.1.

Therefore, the importance of developing an ensemble of methods and a related information system that performs correct writer identification is absolutely essential. The authors' research group has already made a first attempt towards this direction with very fruitful results, in connection with ancient inscriptions [1–3]. There is, however, substantial novelty in the present work and manuscript, namely: (1) The method presented here is essentially different than the methods introduced in [3,2]. We would like to emphasize that, as described in [4], when a new independent statistical method is developed for the solution of a problem including stochastic processes, then one may be asymptotically certain for the results offered in common with the already existing methods. In other words, every new approach that confirms already existing results offered by other statistical methods increases the degree of confidence that these results are correct. (2) The methods developed by the authors and presented in [3,2], work really well, when there is a sufficient number (usually >5) of realizations for each alphabet symbol on a tested inscription. Without this limitation, the method introduced here seems to work quite well even if a very small number (1–5) of realizations exist on an inscription. As it will become evident from the subsequent analysis, the comparison of two documents in order to test if they have been written

by the same writer, is exhaustive; in other words, all pairs of realizations of the same alphabet symbol in the two compared documents are processed. Thus, for example, suppose that one has to decide if two manuscripts D_1 and D_2 have been written by the same writer or not. Suppose, moreover, that there are four (4) realizations of a certain alphabet symbol L in D_1 and five (5) realizations of L in D_2 . Then, the method introduced here, generates $5 \times 5 = 25$ and $4 \times 5 = 20$ proper quantities, which may offer a quite reliable relative answer, after the proposed statistical processing. The same hold true for documents D_1 and D_2 including two (2) and ten (10) realizations of an arbitrary L , respectively, etc. (see Section 7.3). The previous work of the authors deals with ancient inscriptions; the present work tackles both the problem of ancient inscriptions, as well as the problem of Byzantine codices. The essential difference between these two cases will become evident in the main text of the manuscript. (3) The method proposed here is exhaustive in association with all realizations appearing on a document.

In any case, the authors are dealing with the problem of identifying the writer of documents, for which no prior information whatsoever is known. In other words, there are no reference documents at all. Equivalently, absolutely no training set of documents were available for the development of the introduced system, nor any supporting database. We would like to stress that the method presented here tries to accomplish writer identification based on individual letter realizations and/or repeated complexes of letters. We emphasize that the introduced analysis is based on the individual characters or blocks of characters, extracted as described in Section 3. Strictly speaking, the introduced approach is not text-independent. However, the only prerequisite for the application of the entire set of actions for writer identification, which is presented here, is knowledge of the set of the fundamental symbols. Equivalently, in order to apply the introduced method is sufficient to be able to distinguish the different symbols of an alphabet, say the Greek or Latin or of an earlier such set of symbols, as the Babylonian, the Etruscan one, etc. In fact, the only reason for knowing how to distinguish distinct letters or complexes of letters is the need to make comparisons only between realizations of the same symbols or complexes of symbols. Beyond that, no text-dependency at all exists in the introduced method.

Concerning the approach introduced here, we would like to make the following two remarks: (a) It is assumed that the main text appearing on a single stone or on a single papyrus has been written by the same hand. (b) Every single papyrus consists of the main text as well as a number of comments written on the boards of each page by either the person copying the original text or by scholars/researchers afterwards. In the present work, we completely ignore the comments. However, the authors intend to attempt classification of the comments in the near future.

2. A literature survey and intrinsic difficulties of the problem in hand

2.1. State of the art in writer identification and related subjects

In the last years the research topic of automated writer identification and verification is very active, mainly concerning hand written text. Concerning on-line writer identification, Ref. [5] proposes a method for dynamic writer identification that uses the relation between static and dynamic information in a handwritten text. In [6] a system for on-line writer identification from data captured from a whiteboard is presented. The developed system employs Gaussian Mixture Models in order to process the text-independent features from the whiteboard handwriting, while the data selected from different writers are used to train a

Download English Version:

<https://daneshyari.com/en/article/525681>

Download Persian Version:

<https://daneshyari.com/article/525681>

[Daneshyari.com](https://daneshyari.com)