Archaeometry **49**, 4 (2007) 749–764 doi: 10.1111/j.1475-4754.2007.00333.x

IDENTIFYING HANDS ON ANCIENT ATHENIAN INSCRIPTIONS: FIRST STEPS TOWARDS A DIGITAL APPROACH*

S. V. TRACY

American School of Classical Studies at Athens, 54 Souidas Street, GD-106 76, Athens, Greece

C. PAPAODYSSEUS,† P. ROUSSOPOULOS, M. PANAGOPOULOS, D. FRAGOULIS, D. DAFI and TH. PANAGOPOULOS

National Technical University of Athens, School of Computer and Electrical Engineering, Zografou 157 73, Greece

In this paper, a novel methodology is introduced for the identification of the workmen (hands) that carved ancient inscriptions. This methodology employs specific geometric characteristics of each letter and computes the mean value and variance of these characteristics for each one of the available inscriptions separately. Subsequently, we define original decision thresholds that make use of the statistical distribution of the difference of these values in order to attribute an inscription to a given hand. The inscriptions of the hands under consideration have been properly processed and all information extracted, both visual and statistical, is stored in a suitable database. Application of this methodology to nine Athenian inscriptions, some of which contain very similar letters, offered correct, clear-cut hand identification.

KEYWORDS: AUTOMATED LETTER-CUTTER IDENTIFICATION, AUTOMATED INSCRIBER IDENTIFICATION, HANDWRITER IDENTIFICATION, AUTOMATED CLASSIFICATION OF HANDWRITING STYLE, STATISTICAL HANDWRITING RECOGNITION

1 INTRODUCTION—A BRIEF DESCRIPTION OF THE PROBLEM AND THE APPLIED METHODOLOGY

What follows is a first methodological article exploring the possibility of using digital technology to aid the recognition of hands on ancient Athenian inscriptions. In Tracy (1990, 1995, 2003), it has been established that it is actually possible to reliably identify the hands of individual ancient letter-cutters. Nevertheless, employing mathematics, digital image processing and pattern recognition for identifying the hand that formed an inscription may make the process automated and essentially more objective. In fact, as things are at the present time, identifying writers of ancient inscriptions incorporates a significant amount of subjectivity.

The first step towards the development of the methodology introduced in this paper was to set up a test case. Professor S. V. Tracy, as the epigraphist—that is, the specialist in inscriptions among the authors—provided the rest of the team with photographs of stones cut by very prolific workmen who were inscribing decrees in Athens during the second half of the third century BC. In particular, Professor Tracy offered high-quality digital images of the following inscriptions with reference numbers: IG II^2 336, Agora XV 240, Agora XVI 208, IG

*Received 15 August 2005; accepted 10 August 2006
†Corresponding author: email mpanagop@central.ntua.gr

II^2 264, IG II^2 775 and IG II^2 833. These six epigraph images constituted the reference set/database; no more information concerning these inscriptions was provided to the other authors. In addition, the epigraphist furnished the other authors with the images of three inscriptions, those with reference numbers IG II^2 788, part of Agora XV 243 and part of Agora inv. no. 14033, to be classified with the use of the methodology introduced in this paper, again without any additional information. We would like to stress that the mathematics/informatics group was not aware of the corresponding IG numbers of any of the inscriptions. For a description of the lettering and the dossiers of the aforementioned inscriptions, see Tracy (1990, 2003). We can now point out that two of the hands that cut some of the aforementioned inscriptions, Hand 1 (The Cutter of Agora XVI 208) and Hand 2 (IG II^2 833), are so close in style that it is very probable that one was trained by the other and was his apprentice (see Tracy 1990, 2003). On the level of shape of individual letters, which is what is being examined in this paper, the two hands are not at all easy to distinguish.

The next step was to characterize mathematically and rigorously four letters of each hand. The authors discovered that there were clear-cut differences in the chosen mathematical descriptions of these letters, no matter how similar they seemed initially.

The analysis introduced below refers to a problem which is more or less associated with automatic handwriting recognition and writer identification/authentication. Handwritten text recognition has been the object of extensive research, even from as far back as the 1960s, and numerous publications on the subject have appeared (Plamondon and Srihari 2000; Arica and Yarman 2001; Connell and Jain 2002; Schomaker and Bulacu 2004). On the other hand, writer identification seems to be a much more difficult task (Plamondon and Srihari 2000; Franke and Köppen 2001). However, a number of approaches and corresponding publications have appeared aiming at the identification of English and Chinese writers (Said *et al*. 2000; Franke and Köppen 2001; Marti *et al*. 2001; Xianliang *et al*. 2004; Yefeng *et al*. 2004). Nevertheless, to the best of our knowledge, this is the first time that identification of cutters of ancient inscriptions has been attempted. In addition, it is the first time that Greek writer identification has been performed. Finally, the novel methodology introduced in this paper may also offer analytical and objective information about the style of each cutter of ancient inscriptions, its evolution in time, as well as the similarities and dissimilarities of the style of each cutter when compared to the others.

2 A FIRST-STAGE PROCESSING OF THE LETTERS

2.1 An initial shape-based classification of the letters

We have decided to divide the whole class of 24 capital Greek letters into two subclasses, according to whether their contour lines are predominantly rectilinear or curved:

• *Rectilinear class of letters*. The letters whose contour essentially consists of quasi- straight-line segments belong to this class. In this class belong the letters Α, Γ, Δ, Ε, Ζ, Η, Ι, Κ, Λ, Μ, Ν, Ξ, Π, Σ, Τ, Υ and Χ.

• *Curved class of letters*. This class includes those letters for which a substantial part of the contour is noticeably curved. Thus, in this class belong letters Β, Θ, Ο, Ρ, Φ, Ψ and Ω.

We would like to point out that in a very limited number of letters, this distinction may be hand dependent. Thus, for example, one may encounter inscriptors who cut round letters with straight letter-strokes, with the result that diamond-shaped omicrons and cruciform phi’s occasionally occur.

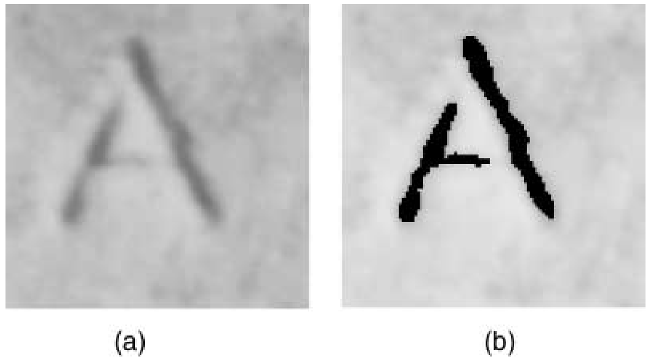

Figure 1 *(a) An original letter A. (b) The result of image segmentation applied to Figure 1 (a).*

2.2 Proper image segmentation to extract the letter

First, we have applied various image segmentation methods (Casey and Lecolinet 1996; Román-Roldán *et al*. 2001; Panagopoulos *et al*. 2004), in order to obtain quite clear-cut and accurate region borders for each letter. A rather simple method that seems to work well is the one that uses each letter's pixel intensity histogram and its lower turning point. All pixels with an intensity lower than this turning point may be considered to belong to the letter (see the application of this method to the letter in Fig. 1 (a), which results in the segmented letter shown in Fig. 1 (b)). This method, together with the letter itself, may generate various artefacts that may be removed by application of proper morphological filters (Panagopoulos *et al*. 2004).

2.3 Letter contour extraction

Each letter contour is obtained with the use of the following quite common method:

• The colour depth of each letter image is decreased from millions of colours to black and white. Thus, the whole letter is black (value '1') and its background is white (value '0').

• The letter contour is extracted. However, no edge detection algorithm could generate the letter contour in the form necessary for the subsequent analysis. In fact, in order for the introduced methodology to be applied, each contour must have the following properties (see Fig. 2): (a) each pixel must have exactly two neighbouring pixels; (b) no isolated pixels or groups of pixels are allowed; and (c) three pixels must not form a compact right (90°) angle. Therefore, suitable software has been developed in order to guarantee this form of the contour.

2.4 Determining the turning points in each letter

We properly approximate the letter contour with slightly overlapping splines, as defined below.

We divide the entire contour of a letter into slightly overlapping subsets of consecutive pixels of length L_s and we calculate the spline; namely, the third-degree polynomial that best fits each such subset. In this way, one obtains an ensemble of consecutive, slightly overlapping splines, S_i, $i = 1, 2, \ldots, N_s$, covering each object (Fig. 3). Note that probable small gaps in the

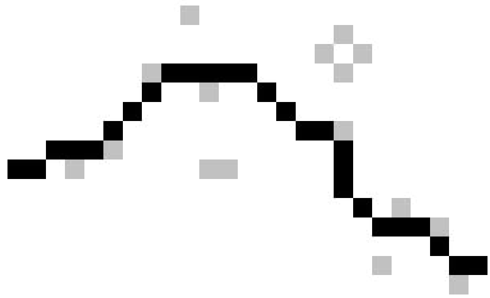

Figure 2 *Proper (in black) and forbidden (in grey) pixel combinations of a letter contour.*

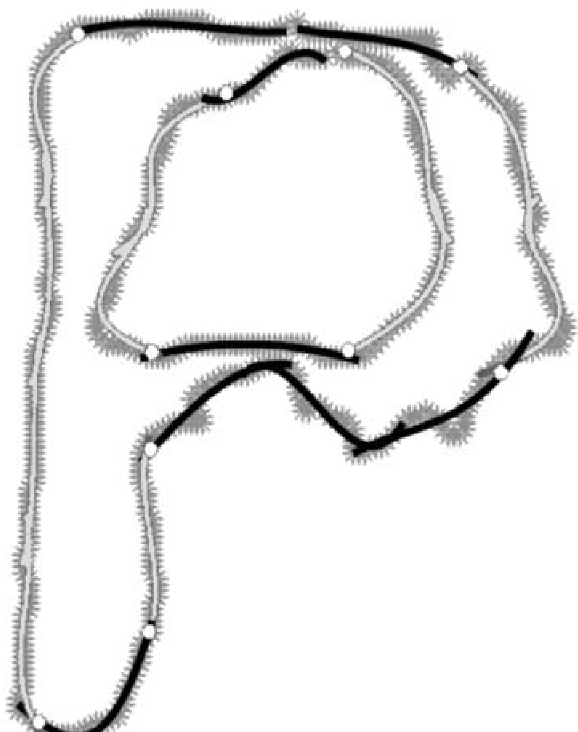

Figure 3 *An ensemble of consecutive, slightly overlapping splines, covering the letter contour depicted in grey. The turning points are indicated by white dots, while splines covering connected contour segments between two successive turning points are indicated by the same colour (light grey or black).*

letter contour due to wear of the surface of the inscription are bridged with this procedure. A good choice for L_s seems to be a small proportion of its maximum dimension, say $\alpha = 0.25$, and for the overlapping region a satisfactory length seems to be $L_O = 0.2L_S$. Note that neighbouring values of L_s and L_O offer quite similar and satisfactory results.

Next, we compute the curvature at an arbitrary point of the set of splines. This offers a very good approximation of the letter contour curvature at this point. When this curvature is greater than a predefined threshold—say, three times the mean value of the whole contour curvature—then we consider that this point is a critical turning point of the contour of the letter in hand. For example, the critical turning points of a letter P are shown in Figure 3.

3 A BRIEF OUTLINE OF THE EMPLOYED METHODOLOGY

In this section, we will briefly describe the basic steps of the methodology that has been employed towards the cutter (writer) identification. The procedure that has been applied is as follows:

(i) Letter images are extracted from high-quality digital images of the inscriptions, via automated image segmentation techniques.

(ii) The contour of each letter is extracted from the aforementioned segmented image, by means of original algorithms specifically developed for the application in hand.

(iii) The critical turning points of each letter's contour are automatically extracted by means of novel dedicated algorithms.

(iv) The shape of the letter is modelled in a piecewise manner by means of least-squares methods.

(v) Various characteristic letter features are determined, mainly in the form of ratios and angles, which are considered to give quantitative measures of the peculiarities of the cutting style of each writer.

(vi) We state statistical hypotheses that lead to estimates of the distributions of the mean values and variances of the populations of the features determined in (v).

(vii) Using the results of (vi), certain quantities are defined, which are employed as decision thresholds for the classification of inscriptions according to the writer.

4 EXTRACTING A SET OF ESSENTIAL FEATURES FROM EACH LETTER

After defining the critical turning points of each letter, we proceed by defining a set of features that we will call 'essential features or characteristics', on which we will base the characterization, processing and all subsequent analysis for each letter separately.

4.1 A set of features for the rectilinear class of letters

For simplicity, brevity and easy reference, we will base the subsequent analysis on letter A, which is an excellent representative of the rectilinear class. Note that in the inscriptions considered, the number of alphas ranges from 15 to 76, with a mean value of approximately 37 letters per inscription.

For this analysis, it is necessary to clarify that when we approximate a chain of pixels of a letter contour—say, the chain A_1A_2 in Figure 4 (a)—with a straight-line segment in the least-squares sense, we choose the independent variable to be x if $\Delta_x \geq \Delta_y$, and y otherwise, where $\Delta_x = (x_{max} - x_{min})$ and $\Delta_y = (y_{max} - y_{min})$. Thus, for all letters A of the type similar to the one depicted in Figures 1 and 4, we define the following set of essential characteristics. Note that special care has been taken to ensure independence of these characteristics, probably associated with the corresponding freedom of the artist in carving the letter. Evidently, independence of the degrees of freedom in carving the letter does not ensure stochastic independence. The

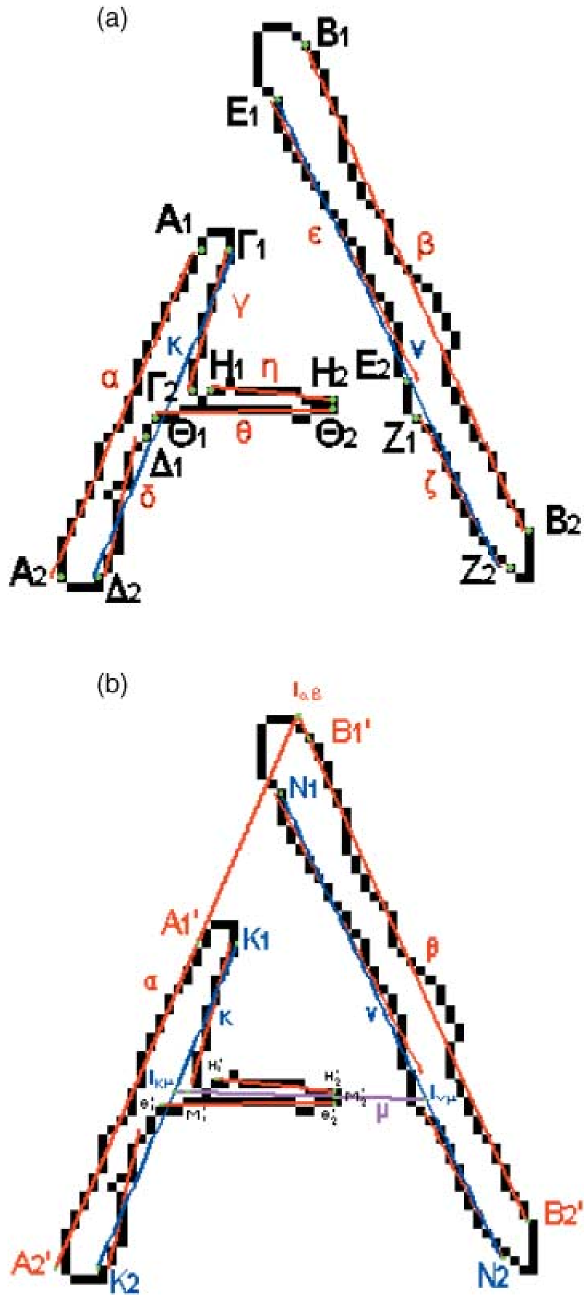

Figure 4 *(a) The contour of letter A of Figure 1, together with LS line segments necessary for extracting the essential characteristics of this letter. (b) A depiction of the additional geometrical objects necessary for extracting the essential characteristics of letter A.*

latter is a plausible assumption, the validity of which will be verified by experimental results as the inscription database grows. However, correlation tests applied to the available data do not contradict this hypothesis.

The set of chosen characteristics concerning letter A is described below (see Fig. 4).

C_1^A: We consider the chain of contour pixels of the letter in hand starting at A_1 and ending at A_2 and we compute the straight-line segment, say α, that best approximates this chain in the least-squares sense. Moreover, we consider chain B_1B_2, consisting of the contour of pixels of the letter in hand starting at B_1 and ending at B_2. Once more, we compute the straight-line segment, say β, that best approximates this chain B_1B_2 in the least-squares sense.

A first essential characteristic for letter A is the angle that the straight-line segments α and β form. To avoid confusion, for the angle computation we consider α and β to be vectors starting at the points corresponding to A_1 and B_1, respectively. In other words, we let $\varphi_{\alpha\beta} = \angle(\vec{\alpha},\vec{\beta})$ be the first characteristic of letter A.

C_2^A: In an analogous manner, we consider straight-line segments γ and δ best approximating chains $\Gamma_1\Gamma_2$ and $\Delta_1\Delta_2$, respectively, in the least-squares sense. Proceeding as above, we define the angle $\varphi_{\gamma\varepsilon} = \angle(\vec{\gamma},\vec{\delta})$.

C_3^A: The third characteristic of letter A aims at checking how parallel the sides of the left leg of A are. In practice, we consider chains $\Gamma_1\Gamma_2$ and $\Delta_1\Delta_2$ together, and we compute the straight-line segment κ that best approximates the union of these two chains in the least-squares sense. Then, the third essential feature is defined to be the angle $\varphi_{\alpha\kappa} = \angle(\vec{\alpha},\vec{\kappa})$.

C_4^A: In an analogous manner, we check whether or not the two sides of the right leg of A are parallel by concatenating chains E_1E_2 and Z_1Z_2, computing the least-squares line segment ν that best approximates the union of these two chains and defining the angle $\varphi_{\beta\nu} = \angle(\vec{\beta},\vec{\nu})$.

C_5^A: Proceeding similarly, we define the fifth characteristic to be the angle $\varphi_{\eta\theta} = \angle(\vec{\eta},\vec{\theta})$, where η and θ are the straight-line segments best approximating the chains of pixels H_1H_2 and $\Theta_1\Theta_2$, respectively.

C_6^A: To define the sixth characteristic, we let $(x_1^{\mathrm{H}}, y_1^{\mathrm{H}})$, $(x_2^{\mathrm{H}}, y_2^{\mathrm{H}})$ be the coordinates of the first pixel H_1 and the last pixel H_2 of the chain H_1H_2. In addition, we let the equation of η be $y = a_\eta x + b_\eta$, in which case η starts at point $\mathrm{H}_1' : (x_1^{\mathrm{H}}, a_\eta x_1^{\mathrm{H}} + b_\eta)$ and ends at point $\mathrm{H}_2' : (x_2^{\mathrm{H}}, a_\eta x_2^{\mathrm{H}} + b_\eta)$. Similarly, we consider that straight-line segment θ starts at point $\Theta_1' : (x_1^{\Theta}, a_\theta x_1^{\Theta} + b_\theta)$ and ends at point $\Theta_2' : (x_2^{\Theta}, a_\theta x_2^{\Theta} + b_\theta)$. Now, we define points

$$\mathrm{M}_1' : \left(\frac{x_1^{\mathrm{H}} + x_1^{\Theta}}{2}, \frac{a_\eta x_1^{\mathrm{H}} + b_\eta + a_\theta x_1^{\Theta} + b_\theta}{2}\right), \mathrm{M}_2' : \left(\frac{x_2^{\mathrm{H}} + x_2^{\Theta}}{2}, \frac{a_\eta x_2^{\mathrm{H}} + b_\eta + a_\theta x_2^{\Theta} + b_\theta}{2}\right),$$

where, clearly, M_1' is the middle of $\mathrm{H}_1'\Theta_1'$ and M_2' is the middle of $\mathrm{H}_2'\Theta_2'$. Subsequently, we define the straight-line segment μ starting at M_1' and ending at M_2'. Finally, we compute the angle between κ and μ, namely $\varphi_{\kappa\mu} = \angle(\vec{\kappa},\vec{\mu})$, and consider it to be the sixth characteristic.

C_7^A: Next, we consider the straight line μ^ε containing segment μ and we find the point of intersection of μ^ε and κ—say, $\mathrm{I}_{\kappa\mu}$. Now, if $\mathrm{K}_1 : (x_1^{\Gamma}, a_\kappa x_1^{\Gamma} + b_\kappa)$ and $\mathrm{K}_2 : (x_2^{\Delta}, a_\kappa x_2^{\Delta} + b_\kappa)$ are the beginning and the end of κ, respectively, then we define the seventh characteristic of letter A to be the fraction

$$f_{\kappa\mu} = \frac{\mathrm{K}_2\mathrm{I}_{\kappa\mu}}{\mathrm{I}_{\kappa\mu}\mathrm{K}_1}.$$

C_8^A: Using exactly the same method as before, we define the eighth characteristic to be the fraction

$$f_{\nu\mu} = \frac{\mathrm{N}_2\mathrm{I}_{\nu\mu}}{\mathrm{I}_{\nu\mu}\mathrm{N}_1},$$

where $\mathrm{I}_{\nu\mu}$ is the intersection of μ^{ε} and ν, while $\mathrm{N}_1 : (x_1^{\mathrm{E}}, a_\nu x_1^{\mathrm{E}} + b_\nu)$ and $\mathrm{N}_2 : (x_2^{\mathrm{Z}}, a_\nu x_2^{\mathrm{Z}} + b_\nu)$ are the beginning and the end of ν, respectively.
C_9^A: The ninth feature is defined to be $g_{\nu\mu} = |\,\nu\,|/|\,\mu\,|$, where $|\,\nu\,|$ and $|\,\mu\,|$ denote the length of the straight-line segments ν and μ, respectively.
C_{10}^A: We let the 10th characteristic be $g_{\kappa\nu} = |\,\kappa\,|/|\,\nu\,|$.
C_{11}^A: Similarly, we define $g_{\alpha\beta} = |\,\alpha\,|/|\,\beta\,|$.
C_{12}^A: In order to define the 12th characteristic, we first let α^{ε} and β^{ε} be the straight lines to which segments α and β belong. Moreover, we let $\mathrm{I}_{\alpha\beta}$ be the point of intersection of lines α^{ε} and β^{ε}. Then, we define the criterion

$$f_{\alpha} = \frac{\mathrm{I}_{\alpha\beta}\mathrm{A}_2'}{\mathrm{A}_1'\mathrm{A}_2'}.$$

C_{13}^A: Similarly, we define the feature

$$f_{\beta} = \frac{\mathrm{I}_{\alpha\beta}\mathrm{B}_2'}{\mathrm{B}_1'\mathrm{B}_2'}.$$

C_{14}^A: We consider the domain D_1 enclosed by the simple curve consisting of the straight-line segment $\mathrm{A}_1\Gamma_1$, the chain of contour pixels $\Gamma_1\Gamma_2$, the straight-line segment $\Gamma_2\Delta_1$, the chain of contour pixels $\Delta_1\Delta_2$, the straight-line segment $\Delta_2\mathrm{A}_2$ and finally the chain of contour pixels $\mathrm{A}_2\mathrm{A}_1$. We also consider the domain D_2 enclosed by the simple curve consisting of the straight-line segment $\mathrm{B}_1\mathrm{E}_1$, the chain of contour pixels $\mathrm{E}_1\mathrm{E}_2$, the straight-line segment $\mathrm{E}_2\mathrm{Z}_1$, the chain of contour pixels $\mathrm{Z}_1\mathrm{Z}_2$, the straight-line segment $\mathrm{Z}_2\mathrm{B}_2$ and finally the chain of contour pixels $\mathrm{B}_2\mathrm{B}_1$. If $E(D_1)$ and $E(D_2)$ are the areas of domains D_1 and D_2 respectively, then we define the 14th criterion to be

$$e_{1,2} = \frac{E(D_1)}{E(D_2)}.$$

C_{15}^A: Using exactly the same method as before, we define the 15th characteristic to be

$$e_{1,3} = \frac{E(D_1)}{E(D_3)},$$

where $E(D_3)$ is the area of domain D_3 enclosed by the simple curve consisting of the straight-line segment $\Gamma_2\mathrm{E}_2$, the straight-line segment $\mathrm{E}_2\mathrm{Z}_1$, the straight-line segment $\mathrm{Z}_1\Delta_1$ and finally the straight-line segment $\Delta_1\Gamma_2$.

We would like to point out that addition of independent characteristics always offers new information (Kokolakis 1981).

4.2 Extracting a set of features for the curved class of letters

The analysis of the curved letters will be made by referring to O, since the techniques for extracting essential features for O can be easily extended so as to extract characteristics from the curved parts of the other letters of the class. Note that in the inscriptions of the reference

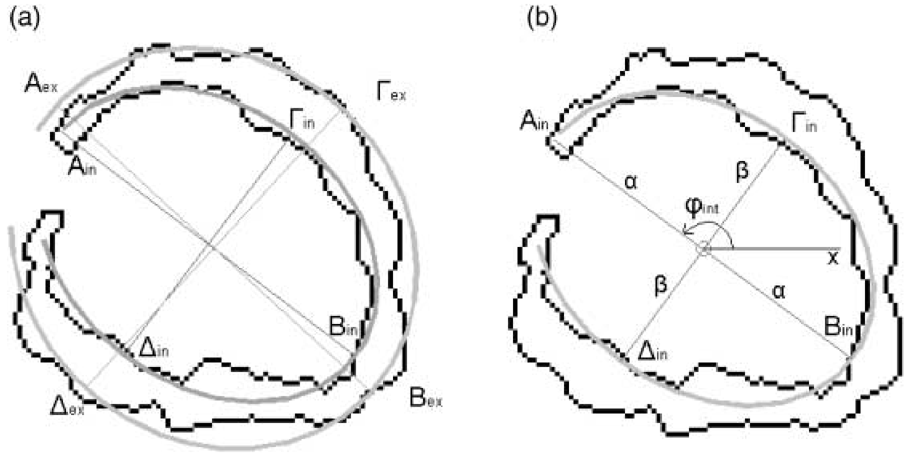

Figure 5 *(a) An approximation of the internal and external contours of letter O by ellipses. (b) The semi-axes and the rotation angle of the ellipsis approximating the internal contour of letter O.*

database, the number of omicrons ranged from 14 to 154, with a mean value of approximately 49 letters per inscription.

One approach is to approximate the internal and external boundaries of each letter O by an ellipsis or another proper geometrical figure (see Fig. 5). The choice of the ellipsis or another figure is made by using the number of turning points spotted in Section 2.4. Suppose that we are dealing with a set of hands whose letters 'O' are well approximated by a set of ellipses, which seems to be the most usual case. In practice, to achieve the approximation of letter O contours with ellipses, we proceed as follows.

First, we note that the general equation of an ellipsis is, in polar coordinates, $\overrightarrow{r^M}(t) = (x_0 + a\cos(t))\vec{i} + (y_0 + b\cos(t))\vec{j}$, where x_0, y_0 are the coordinates of the ellipsis centre and $2a$, $2b$ are the corresponding lengths of the ellipsis axes (see Figs 5 (a) and 5 (b)). Note that the ellipses of the aforementioned type have their axes parallel to the coordinate axes. To obtain the general form of the rotated ellipsis, one must multiply these equations by the rotation matrix

$$\begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix},$$

where φ is the rotation angle.

Next, consider the internal and external contours of a letter O and their digitized images, consisting of N^O_{int} and N^O_{ext} pixels, respectively, described by the sequence of vectors $\overrightarrow{r_i^{\text{int}}}$, $i = 1, 2, \ldots, N^O_{\text{int}}$, $\overrightarrow{r_j^{\text{ext}}}$, $j = 1, 2, \ldots, N^O_{\text{ext}}$ starting at a reference centre and pointing to each pixel centre, as shown in Fig. 5 (a). Suppose that one wants to test whether this curve is the successful result of a writer's attempt to simulate an ellipsis described by the parametric vector equation $\overrightarrow{r^E}(t \mid \Pi)$, where t is the independent variable, Π is the ellipsis set of parameters and superscript E denotes an ellipsis. It is an immediate consequence of the ellipsis polar parametric equation that $\Pi = \{x_0, y_0, a, b, \varphi_0\}$.

Subsequently, we compute the optimal set of parameters Π^{O} and the corresponding sequence of values of the independent variable t_i, $i = 1, 2, \ldots, N^O$, so that $\overrightarrow{r^E}(t_i \mid \Pi^{\text{O}})$ best fits

$\overrightarrow{r_i^{\mathrm{int}}}$ or $\overrightarrow{r_j^{\mathrm{ext}}}$ separately according to a chosen norm L—say, the Euclidean. Algorithms to achieve this are the well-known conjugate gradient and/or the easier to implement Nelder–Mead method. They both start from a tentative set of values of Π and let Π converge to Π^{O} so that—say, for the internal boundary of O—norm $L_{\mathrm{int}} = \sum_{i=1}^{N_{\mathrm{int}}^{O}} (\overrightarrow{r_i^{\mathrm{int}}} - \overrightarrow{r^E}(t_i \mid \Pi))^2$ is minimized.

After applying this method to all available letters O in all of the considered inscriptions, all O internal and external contours have been optimally approximated by ellipses (see Fig. 5 (a)). The average error of this approximation—namely, the minimal values of norms L_{int} and L_{ext}—is particularly low, at around two pixels.

Using these approximations of the O internal and external contours, we have set the following characteristics for letter O (see Figs 5 (a) and 5 (b)).

C_1^O: The ratio $a_{\mathrm{int}}/b_{\mathrm{int}}$ of the long and short axes of the ellipsis approximating the letter O internal contour.

C_2^O: The ratio $a_{\mathrm{ext}}/b_{\mathrm{ext}}$ of the long and short axes of the ellipsis approximating the letter O external contour.

C_3^O: The ratio $a_{\mathrm{ext}}/a_{\mathrm{int}}$ of the two longer axes of the ellipses approximating the letter O external and internal contours, respectively.

C_4^O: The angle ϕ_{int} in $[0, \pi)$ the longer axis of the ellipsis approximating the internal contour forms with the x coordinate axis.

C_5^O: The angle ϕ_{ext} in $[0, \pi)$ the longer axis of the ellipsis approximating the external contour forms with the x coordinate axis.

5 DECIDING THAT A GIVEN INSCRIPTION CORRESPONDS TO A SPECIFIC WRITER: APPROACH 1, BASED ON MEAN VALUES

The whole analysis starts with the processing of a set of inscriptions Σ_1 generated by an arbitrary hand, say H_1, as verified by an expert of the field. Consider an arbitrary letter appearing in these inscriptions and a specific characteristic of this letter. Thus, for example and easy reference, consider letter A and any one of its 15 characteristics introduced in Section 4, say C_i. The peculiarities of each such characteristic can be obtained in a straightforward manner with the use of the confidence intervals of its mean value and standard deviation.

In this section, we define novel decision quantities/thresholds p_k^A based on the mean values of each characteristic C_i, for classifying an unknown inscription I_u to a proper hand already existing in the database. First, we make the statistical assumption that all chosen characteristics come from a normal distribution. Although this is an empirical assumption, we have applied the Kolmogorov test and in many cases the χ^2 test to find if there is sufficient evidence that contradicts it. The performed statistical tests did not reject this hypothesis. Specifically, they do not offer evidence that would prohibit application of the methodology that will be described below ($a = 0.01$).

5.1 The notation employed and the feature mean value distribution

Suppose that we have in our database a set of inscriptions Σ_k, whose cutters are already known; thus, for example, suppose that Σ_1 was carved by H_1, Σ_2 was carved by H_2, and so on. Assume, now, that a new inscription I_u is given, of which the identity of the hand that inscribed it is unknown. The purpose of this project is to achieve an identification of the hand that carved I_u

with maximum possible confidence. In order to accomplish that, a first step is to attribute to each characteristic C_i of all letters appearing in I_u a number of quantities that will allow us to decide if this feature C_i pertains to one of the hands whose inscriptions already exist in the database. A first method we have chosen to calculate these decision thresholds employs the distribution of the mean values of C_i. In fact, suppose that inscription I_u has been generated by a certain hand H_u, who carves the arbitrary feature C_i of a letter, say A, with mean value $\mu_{i,u}$ and variance $\sigma_{i,u}^2$, which, clearly, is not *a priori* known. Suppose, moreover, that in inscription I_u there are N_A^u realizations of the specific letter. Consequently, one computes the mean value $\bar{X}_{i,u}$ and the variance $S_{i,u}^2$ of the specific letter characteristic in inscription I_u. Now, we state once more that the set of inscriptions Σ_k carved by hand H_k give rise to a set of N_A^k values for the specific letter characteristic having mean value $\bar{X}_{i,k}$ and variance $S_{i,k}^2$. According to the adopted hypotheses, these letter characteristic values of Σ_k come from a normal population with generally unknown mean values $\mu_{i,k}$ and variances $\sigma_{i,k}^2$.

Indeed, in order to compare the mean values of the *i*th characteristic of the specific letter in the inscriptions I_u and Σ_k, we employ quantity

$$t_{i,k,u} = \frac{(\bar{X}_{i,k} - \bar{X}_{i,u}) - (\mu_{i,k} - \mu_{i,u})}{\sqrt{\dfrac{S_{i,k}^2}{N_A^k} + \dfrac{S_{i,u}^2}{N_A^u}}}, \tag{5.1}$$

where the three subscripts in $t_{i,k,u}$ are as follows: the first one denotes the *i*th characteristic of the letter (e.g., for letter A, $i = 1, 2, \ldots, 15$), the second one denotes the reference hand and the third one denotes the unknown hand; note again that we have N_A^k letters A (observations) in the Σ_k inscription, and N_A^u in the unknown inscription, where subscript A in N_A^k, N_A^u denotes the letter under consideration.

Quantity $t_{i,k,u}$ follows a Student distribution with $d_{i,k,u}$ degrees of freedom, where: (a) if the two population variances are equal, $d_{i,k,u} = N_A^k + N_A^u - 2$; and (b) for unequal population variances,

$$d_{i,k,u} = (S_{i,k}^2/N_A^k + S_{i,u}^2/N_A^u)^2 \Bigg/ \left[\frac{(S_{i,k}^2/N_A^k)^2}{N_A^k - 1} + \frac{(S_{i,u}^2/N_A^u)^2}{N_A^u - 1}\right]$$

is computed dynamically. Although there is not enough evidence to contradict the equality of variances, we prefer to use (b) for safety reasons. In any case, both methods offer very similar decision results.

5.2 Decision thresholds

Next, in order to verify whether the two sets of values of the specific *i*th letter characteristic of Σ_k and I_u come from the same population, in the aforementioned equation (5.1) we let $\mu_{i,k} = \mu_{i,u}$, in which case the value of the quantity $t_{i,k,u}$ is known. Let this value be $t_{i,k}$, where, once more, the first subscript denotes the *i*th characteristic of the letter, while the second one denotes the reference hand; clearly, subscript u has been suppressed, since the assumption $\mu_{i,k} = \mu_{i,u}$ has been momentarily adopted. Using this value, we may establish a first approach in order to decide whether the unknown inscription I_u has been carved by a 'known' hand H_k, as far as characteristic C_i of the specific letter is concerned. In fact, let $f_k(t_{i,k})$ be the value of the probability density function of the Student distribution with $d_{i,k,u}$ degrees of freedom, at point

$t_{i,k}$. Then, by definition, the probability that $t_{i,k,u}$ lies in the interval $[t_{i,k}, t_{i,k} + dt]$, given that $\mu_{i,k} = \mu_{i,u}$, is $f_k(t_{i,k})dt$.

Next, suppose that all N_C characteristics of letter A have a mean value coming from the population of the corresponding mean values of hand H_k; equivalently, suppose that $\mu_{i,k} = \mu_{i,u}$ for every $i = 1, 2, \ldots, N_C$.

Hence, if we examine N_C characteristics of the specific letter, then since the hypothesis of their statistical independence has been adopted, the overall probability that we simultaneously obtain the N_C values of all $t_{i,k,u}$ lying in the intervals $[t_{1,k}, t_{1,k} + dt], [t_{2,k}, t_{2,k} + dt], \ldots, [t_{N_C,k}, t_{N_C,k} + dt]$, under the hypothesis that $\mu_{i,k} = \mu_{i,u}$, $i = 1, 2, \ldots, N_C$, is the product $\delta p_k^A = \prod_{i=1}^{N_C} (f_k(t_{i,k})dt)$.

Finally, suppose that we want to decide which hand, out of all the known hands H_k, most probably cut the specific letter of the unknown inscription. This amounts to determining the probability that the mean values of the letter characteristics in I_u pertain to the corresponding mean values of hands H_k, $k = 1, 2, \ldots, N_H$, given that one of these hands has indeed carved the letters. We believe that good and robust decision thresholds relating to these probabilities are the quantities defined below:

$$p_k^A = probability((\mu_{i,u} = \mu_{i,k}) \mid (\bar{X}_{i,u}, S_{i,u}^2), \forall \quad i = 1, 2, \ldots, N_C).$$

According to the previous analysis, and if we make the plausible assumption that all $\mu_{i,u}$ are equiprobable in a certain interval, this decision quantity can be estimated as follows.

We define the events E_k, for all available hands $k = 1, 2, \ldots, N_H$, via $E_k = \{t_{i,k,u}$ *belongs to* $[t_{i,k}, t_{i,k} + dt]$, *given that* $\mu_{i,u} = \mu_{i,k}$ *for all* $i = 1, 2, \ldots, N_C\}$, in which case,

$$p_k^A = \frac{probability\ that\ a\ particular\ event\ E_k\ occurs}{probability\ that\ any\ of\ the\ N_H\ events\ E_k\ occurs}.$$

Eventually,

$$p_k^A = \frac{\delta p_k^A}{\sum_{\ell=1}^{N_H} \delta p_\ell^A} = \frac{\prod_{i=1}^{N_C} (f_k(t_{i,k}))}{\sum_{\ell=1}^{N_H} \prod_{i=1}^{N_C} (f_\ell(t_{i,\ell}))}.$$

We would like to point out that the hypothesis of equiprobable $\mu_{i,k}$ is fully compatible with the archaeologist's feeling. However, as the inscription database grows, if a different distribution for the mean value characteristics occurs, then the above formula will be appropriately weighted using Bayes' law, while the rest of the methodology will remain intact.

Finally, the decision as to which hand formed the unknown inscription is made by comparing the p_k^A values: if, for a given hand, the corresponding decision threshold is greater than the others, then the unknown inscription is attributed to this writer. In the experiments performed in connection with the inscriptions referred to in the introduction, one decision threshold is always overwhelmingly greater than the other thresholds, thus offering a clear-cut correct decision. Thus, for example, for the most difficult case, according to the archaeologist, for the classification of the inscription IG II^2 788, the non-matching inscriptions generated decision thresholds p_k^A for letter A of the order of 10^{-41} or less, while, clearly, the correct hand offered a decision threshold of the order $1–10^{-41}$.

6 DECIDING THAT A GIVEN INSCRIPTION CORRESPONDS TO A SPECIFIC WRITER: APPROACH 2, BASED ON VARIANCES

In this approach, we employ the variance instead of the mean value to obtain a decision threshold that a letter of the unknown description I_u has been carved by the hand with the set of inscriptions Σ_k. Without any loss of generality, we once more describe the approach for the letter A. Thus, we point out that quantity

$$F_{i,k,u} = \frac{S_{i,k}^2/\sigma_{i,k}^2}{S_{i,u}^2/\sigma_{i,u}^2}$$

follows Snedecor's F distribution, with $(N_A^k - 1, N_A^u - 1)$ degrees of freedom. Now, the assumption that inscriptions Σ_k and I_u belong to the same hand implies that $\sigma_{i,k} = \sigma_{i,u}$, for the arbitrary ith characteristic of the considered letter. In this case, quantity $F_{i,k,u} = S_{i,k}^2/S_{i,u}^2$ is known and we let its value be $F_{i,k}$. We employ this value to establish another approach in order to decide whether or not the unknown inscription I_u has been carved by hand H_k, as far as characteristic C_i of the specific letter is concerned. In fact, let $u_k(F_{i,k})$ be the value of the probability density function of the Snedecor distribution with $(N_A^k - 1, N_A^u - 1)$ degrees of freedom, at point $F_{i,k}$.

Then, once more, the probability that $F_{i,k,u}$ lies in the interval $[F_{i,k}, F_{i,k} + dF]$, given that $\sigma_{i,k} = \sigma_{i,u}$, is $u_k(F_{i,k})dF$.

Hence, if we examine N_C characteristics of the letter, and once more take into account their assumed statistical independence, then the overall probability that we simultaneously obtain the N_C values of all $F_{i,k,u}$ lying in the corresponding intervals $[F_{i,k}, F_{i,k} + dF]$, under the hypothesis that $\sigma_{i,k} = \sigma_{i,u}$, $i = 1, 2, \ldots, N_C$, is the product $\delta q_k^A = \Pi_{i=1}^{N_C} (u_k(F_{i,k})dF)$.

Our final goal is to identify the hand out of all N_H hands that most probably inscribed the letter under consideration in inscription I_u, given that one of these hands has indeed carved the letters. We once more believe that good and robust decision thresholds relating to these variances are the quantities:

$$q_k^A = probability((\sigma_{i,u} = \sigma_{i,k}) \mid (\bar{X}_{i,u}, S_{i,u}^2), \qquad \forall\ i = 1, 2, \ldots, N_C),$$

According to the previous analysis, this quantity can be estimated as follows. We define the events G_k, via

$$G_k = \{F_{i,k,u}\ belongs\ to\ [F_{i,k}, F_{i,k} + dF],\ given\ that\ \sigma_{i,u} = \sigma_{i,k}\ for\ all\ i = 1, \ldots, N_C\},$$

in which case,

$$q_k^A = \frac{probability\ that\ a\ particular\ event\ G_k\ occurs}{probability\ that\ any\ of\ the\ N_H\ events\ G_k\ occurs}.$$

In writing the above formula, we have considered a uniform distribution of the corresponding variances in their domains. As in Approach 1, the methodology will remain unaltered in the case that the expanded database gives rise to a different distribution for the variances, the only change being an appropriate weighting in the above formula according to Bayes' law.

Finally, the employed decision thresholds concerning variances are given by

$$q_k^A = \frac{\delta q_k^A}{\sum_{\ell=1}^{N_H} \delta q_\ell^A} = \frac{\prod_{i=1}^{N_C} (u_k(F_{i,k}))}{\sum_{\ell=1}^{N_H} \prod_{i=1}^{N_C} (u_\ell(F_{i,\ell}))}.$$

Next, as in Approach 1 again, we compare the q_k^A values: if for a given hand the corresponding decision threshold is greater than the others, then the unknown inscription is attributed to this writer. In the experiments performed in connection with the inscriptions referred to in Section 1, one decision threshold is always overwhelmingly greater than the other thresholds, thus offering a clear-cut decision. Thus, for example, for the most difficult case, according to the archaeologist, for the classification of the inscription IG II^2 788, the non-matching inscriptions generated decision thresholds q_k^A for letter A of the order of 10^{-12} or less, while, clearly, the correct hand offered a decision threshold of the order 1–10^{-12}.

7 DECIDING THAT A GIVEN INSCRIPTION CORRESPONDS TO A SPECIFIC WRITER: APPROACH 3, A JOINT APPROACH

Eventually, one can define a third criterion employing both the mean value and the variance in order to identify the hand, among the known ones, that carved I_u. In a sense, this approach is complete, since we have adopted the hypothesis that the considered letter characteristics follow normal distributions, a hypothesis that has not been rejected by the available data, and, therefore, these distributions are completely defined by their mean values and the variances. We notice that the fact that the same hand carved both Σ_k and I_u implies both $\mu_{i,k} = \mu_{i,u}$ and $\sigma_{i,k} = \sigma_{i,u}$ for all $i = 1, 2, \ldots, N_C$. Moreover, since we have adopted the assumption that each characteristic follows a normal distribution, the two events $\mu_{i,k} = \mu_{i,u}$ and $\sigma_{i,k} = \sigma_{i,u}$ are independent. Therefore, after taking into consideration the analysis in Approaches 1 and 2, decision thresholds π_k^A can be expressed as follows:

$$\pi_k^A = \frac{\delta\pi_k^A}{\sum_{\ell=1}^{N_H} \delta\pi_\ell^A},$$

where $\delta\pi_k^A = \delta p_k^A \delta q_k^A = \Pi_{i=1}^{N_C} f_k(t_{i,k}) u_k(F_{i,k}) dtdF$ is the probability that both events E_k and G_k occur simultaneously. Finally,

$$\pi_k^A = \frac{\prod_{i=1}^{N_C} f_k(t_{i,k}) u_k(F_{i,k})}{\sum_{\ell=1}^{N_H} \prod_{i=1}^{N_C} f_\ell(t_{i,\ell}) u_\ell(F_{i,\ell})}.$$

Once more, a decision is made by means of the greater π_k^A value, and the hand identification in the performed experiments is clearly unambiguous: the decision threshold corresponding to the correct identification obtains a value of the order 1–10^{-53}, while the thresholds for the non-matching inscriptions are of the order of 10^{-53} or less.

We would like to point out that analogous results were obtained for the tau's and rho's. Clearly, the above methodology can be immediately extended so that we can form threshold quantities for deciding whether an arbitrary number of N_L letters appearing in the unknown inscription I_u have been cut by a specific hand H_k.

8 CONCLUSION

In this paper, a methodology for the automated identification of inscribers who carved Athenian inscriptions is presented. Thus, the shapes of the letters, automatically extracted from their images, are processed and sets of mathematical characteristics are obtained for each letter separately. Subsequently, statistical processing of these characteristics' values is performed and, in this way, a number of approaches is introduced to decide which hand of those belonging to the system's database most probably carved an unknown inscription.

A first test of the introduced methodology refers to six inscriptions, with IG numbers IG II^2 336, Agora XV 240, Agora XVI 208, IG II^2 264, IG II^2 775 and IG II^2 833, as mentioned in Section 1. After processing a set of letters appearing in these inscriptions, the expert archaeologist furnished images of three new inscriptions with IG numbers IG II^2 788, part of Agora XV 243 and part of Agora inv. no. 14033, all cut by unknown hands. We emphasize once more that no additional information had been provided to the other authors and, in particular, the IG numbers of the epigraphs to be classified were disclosed after hand identification had taken place. Application of the introduced methodology led to the conclusion that the unidentified inscriptions had been carved as follows:

(1) The hand that cut IG II^2 833 had also carved part of Agora inv. no. 14033.
(2) The hand that cut Agora XV 208 had also carved IG II^2 788.
(3) The hand that cut Agora XV 240 had also carved part of Agora XV 243.

The epigraphist confirmed that all identifications were absolutely correct. We would like to emphasize that the introduced methodology offered clear-cut decisions about the hand that cut the three unknown inscriptions; in a forthcoming paper, we will also consider the probability that an unknown inscription does not belong to the available database.

Evidently, the number of letters examined, as well as the number of inscriptions studied, needs to be increased and the related results will be presented in future papers. In addition, a large number of rigorous test cases must be examined before we can conclude that we have a successful digital method for identifying the hands that cut ancient inscriptions. However, the present results are sufficiently promising and sufficiently important that we want to share them with colleagues at this stage.

REFERENCES

Arica, N., and Yarman-Vural, F. T., 2001, An overview of character recognition focused on off-line handwriting, *IEEE Transactions on Systems, Man and Cybernetics, Part C*, **31**(2), 216–33.

Casey, R. G., and Lecolinet, E., 1996, A survey of methods and strategies in character segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(7), 690–760.

Connell, S. D., and Jain, A. K., 2002, Writer adaptation for online handwriting recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**(3), 329–46.

Franke, K., and Köppen, M., 2001, A computer-based system to support forensic studies on handwritten documents, *International Journal of Document Analysis and Recognition*, **3**(4), 218–31.

Kokolakis, G. E., 1981, Bayesian classification and classification performance for independent distributions, *IEEE Transactions on Information Theory*, **IT-27**, 419–21.

Marti, U.-V., Messerli, R., and Bunke, H., 2001, Writer identification using text line based features, in *Proceedings of the Sixth International Conference on Document Analysis and Recognition (ICDAR '01)*, 101–5, IEEE Computer Society, Washington, DC.

Panagopoulos, Th., Papaodysseus, C., Exarhos, M., Triantafillou, C., Roussopoulos, G., and Roussopoulos, P., 2004, Prehistoric wall-paintings reconstruction using image pattern analysis and curve fitting, *WSEAS Transactions on Electronics*, **1**(1), 108–13.

Plamondon, R., and Srihari, S. N., 2000, On-line and off-line handwriting recognition: a comprehensive survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(1), 63–84.

Román-Roldán, R., Gómez-Lopera, J. F., Atae-Allah, C., Martínez-Aroza, J., and Luque-Escamilla, P. L., 2001, A measure of quality for evaluating methods of segmentation and edge detection, *Pattern Recognition*, **34**, 969–80.

Said, H., Tan, T., and Baker, K., 2000, Writer identification based on handwriting, *Pattern Recognition*, **33**(1), 133–48.

Schomaker, L., and Bulacu, M., 2004, Automatic writer identification using connected-component contours and edge-based features of uppercase Western script, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **26**(6), 787–98.

Tracy, S. V., 1990, *Attic letter-cutters of 229 to 86 B.C.*, University of California Press, Berkeley.

Tracy, S. V., 1995, *Athenian democracy in transition: Attic letter-cutters of 340 to 290 B.C.*, University of California Press, Berkeley.

Tracy, S. V., 2003, *Athens and Macedon: Attic letter-cutters of 300 to 229 B.C.*, University of California Press, Berkeley.

Xianliang Wang and Xiaoging Ding, 2004, An effective writer verification algorithm using negative samples, in *Ninth International Workshop on Frontiers in Handwriting Recognition*, 509–603.

Yefeng Zheng, Huiping Li and Doermann, D., 2004, Machine printed text and handwriting identification in noisy document images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **26**(3), 337–53.