

A Methodology for Writer Identification, With Application to An Important Historical Greek Document

Elias Koukoutsis¹, Athanasios Rafail Mamatsis¹, George Tsavdaridis¹, Eirini Mamatsi¹, Athanasios Delis¹, Dimitris Arabadjis², Constantin Papaodysseus^{1*}

¹National Technical University of Athens
Athens, Greece

Emails or ORCID: e.koukoutsis@ece.ntua.gr, 0000-0001-9927-9918, 0000-0002-7772-5056, eirinimamatsi@gmail.com, th.delis@gmail.com, cpapaod@cs.ntua.gr*

²University of West Attica
Athens, Greece

ORCID: 0000-0002-8438-0471

Abstract—In the present work, a general methodology is introduced for the determination of the writer of a given document, provided that texts of the same hand are available. The approach is successfully applied to an unsigned document that has been spotted in the warehouses of Greek Army History Directorate and led, for the first time, to the conclusion that it has been written by the great Greek Politician Eleftherios Venizelos. The content of the paper is important from the historical point of view.

Keywords—Writer Identification, Curve Fitting, Curve Similarity, Handwriting Style

1. INTRODUCTION

1.1 The goal of the present work

In the Warehouses of the Army History Directorate (AHD), an unsigned document, we will henceforth call UD, has been spotted, and a number of historians wondered if the text belonged to the great Greek Politician Eleftherios Venizelos. A positive answer to this question can be really important, since the content of UD could reveal various, unknown aspects of El. Venizelos' personality, different than those widely accepted.

Consequently, the authors decided to give an, as objective as possible, answer concerning the hand that wrote UD.

1.2 A very short biography of Eleftherios Venizelos

Eleftherios Venizelos was one the greater politicians of modern Greece. He was born in 1864 in Crete. He was elected Prime Minister of Greece for the first time in 1910. When the First World War was declared, E. Venizelos wanted Greece to enter the Coalition of Great Britain, France, USA, Italy, etc., coming into conflict with the King of Greece of this Period, Konstantinos. He almost made a “coup d' etat”, he moved the Greek capital from Athens to Thessaloniki, and led the Greek Army to fight by the side of the Allies. His choice has been justified by the defeat of the German Army. Moreover, before the beginning of the First World War, he decided that Greece should participate in the Balkan wars; Subsequently, after the decisive victory of the Greek army in these wars, the area of the State of Greece had been practically tripled. Moreover, E. Venizelos passed an advanced for his era legislation, pretty favorable to farmers and industry workers. He died in Paris in 1936.

1.3 State of the Art in Writer Identification

Recently, due to the great evolution of computers, identification of the hand that has written a document may be achieved computationally. In fact:

in [1] a technique is proposed which divides a given handwriting into small fragments and considers each fragment as a texture. In [2], the authors suggest computational tools for classification and quantification of calligraphic style, on a statistical basis. [3] use texture-based schemes for Writer Identification. In [4], the Levenshtein edit distance based on Fisher-Wagner algorithm is used to estimate the cost of transforming one handwritten word into another. He et al. [5] employ graphological information for dating the Medieval Paleographic Scale (MPS) dataset. The identification of the writer of handwritten excerpts from their binary images is treated in [6] using two different features' classes. In [7] the minimum number of hands that could have written a set of 16 inscriptions from the “Judahite desert fortress of Arad” has been determined using statistical inference over the distribution of the distances between mixed features representations of letter shapes. [8] uses Fourier descriptors for semi-automatic classification and retrieval of document excerpts. The authors of [9] use computer-vision tools and statistical inference techniques to identify fragments that might originate from the same codex ('joins'). Wolf et al. [10] search for possible joins between catalogued excerpts, using a combination of local descriptors and learning techniques. In [11], a novel junction detection method is introduced and it was applied to writer identification. A new system for writers' classification of medieval manuscripts is presented in [12]; the system is based on features from layout analysis. In [13], Deep Learning schemes and classical machine learning approaches are experimentally compared and the authors reach the conclusion that the DL-schemes outperform or are equivalent to the latter ones.

1.4 The skeleton of the present work

In order to test if UD belongs to El. Venizelos or not, we have applied the following steps:

S1. We have, rather randomly, chosen two documents belonging to the documents collection of the National Research Foundation “Eleftherios K. Venizelos” (NRFEV); we will henceforth call these two documents D^{EV}_1 and D^{EV}_2 . Historically, D^{EV}_1 and D^{EV}_2 have been undoubtedly written by the hand of Eleftherios Venizelos. In addition, we have selected the document “The Greek Constitution” (D^{CON}), written by another hand in the 19th Century, as well as a Greek citizen handwritten letter (D^{GC}), dated in the beginning of the 19th Century.

S2. We have applied pre-processing to all these five documents, as described in Section 2.

S3. A new procedure leading to a similarity criterion between any two realizations of the same alphabet symbol has been developed, presented in Section 3.

S4. A documents' comparison based on the previous S3 procedure, has been developed and applied. This comparison proved that UD belonged to El. Venizelos, practically with certainty.

2. DOCUMENTS PRE-PROCESSING

At each one of the aforementioned five documents, we have applied the subsequent first stage processing:

- A. We have chosen ten (10) alphabet symbols for which we feel a priori that may convey the writing style of a hand-person, namely "α", "γ", "δ", "ε", "θ", "κ", "λ", "μ", "π", "ρ"; we have avoided "easy-to-write" alphabet symbols, like "ι", "ν", "ο", etc., as well as alphabet symbols that are relatively rarely used in Greek language, such as "ζ", "ξ", "φ", etc.
- B. We have extracted the realizations of these 10 alphabet symbols from the aforementioned five documents. The extracted letter was placed into a frame, having at least 10-20 pixels distance from the frame borders. We note that the frame dimensions remain intact throughout the entire writer identification procedure. The extraction has been made by a semi-interactive method, similar to the one presented in [14].
- C. Next, we have applied the image segmentation method introduced in [15], [16], in order to isolate the body of the alphabet symbol realization from its background.
- D. Finally, we have automatically extracted the contour of each letter, using a novel method that will appear in a future publication.

3. COMPARING TWO ARBITRARY REALIZATIONS OF THE SAME ALPHABET SYMBOL

At first, we point out that we consider each allograph as a distinct alphabet symbol; e. g. the Greek letter kappa may be encountered in various handwritten documents either as "u" or as "κ", both being considered as different alphabet symbols in our analysis. Then, in order to compare any two realisations of the same alphabet symbol, say L_1 and L_2 , belonging to the same or different documents, we take the following actions:

- A1. On basis of the analysis introduced in Section 2, we extract the contours of L_1 and L_2 and we symbolize them as C_1 and C_2 respectively. Evidently, contour C_1 consists of, say, N_1 pixels with coordinates (x_i^1, y_i^1) , $i = 1, 2, \dots, N_1$, while C_2 includes N_2 pixels with coordinates (x_i^2, y_i^2) , $i = 1, 2, \dots, N_2$. We evaluate the "centers of mass" K_1 and K_2 of contours C_1 and C_2 .
- A2. We embed both C_1 and C_2 into a properly "abundant" frame F and we let the center of F be the origin O of an orthogonal cartesian system with axes parallel to the sides of F .
Next, we parallel translate both C_1 and C_2 , so that their centers of mass K_1 and K_2 coincide with O .
- A3. We rotate C_1 , with O being the rotation center, by an angle $\phi \in \left\{-\frac{\pi}{4}, \frac{\pi}{4}\right\}$, where the endpoints of the interval of the angle of rotation correspond to the maximum and the minimum inclination of all the hand-written letters we

have encountered. We apply a partition of this interval, using a pretty small angle $\delta\phi = \pi \cdot 10^{-4}$ rad. Clearly, these endpoints, as well as the partition, may dynamically change immediately, depending on the application in hand. As always, the rotation is accomplished by means of the rotation matrix

$$R = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix},$$

and, in this way, we obtain a rotated version of C_1 , say C_1^R , consisting of points (x_i^{R1}, y_i^{R1}) , $i=1, 2, \dots, N_1$.

We re-evaluate the center of mass of C_1^R and we parallel translate the entire rotated contour, so as its new center of mass coincides with O .

- A4. We apply scaling to C_1^R by a factor $\lambda \in \{\lambda_{min}, \lambda_{max}\}$, where λ_{min} and λ_{max} correspond to the minimum and maximum of the same alphabet symbol encountered in the studied documents. In the application in hand, a fully satisfactory choice is $\lambda_{min} = 0.5$ and $\lambda_{max} = 1.7$; evidentially these limits may dynamically change immediately if another application requires that. Consequently, the coordinates of each point (x_i^{R1}, y_i^{R1}) , $i=1, 2, \dots, N_1$, are multiplied by a factor λ , thus generating curve C_1^{RS} , which includes the ensemble of points $(x_i^{RS1}, y_i^{RS1}) = \lambda(x_i^{R1}, y_i^{R1})$, $i=1, 2, \dots, N_1$. Once again, we re-compute the center of mass of these points and we parallel translate C_1^{RS} , so that this new center of mass coincides with the origin O .

We would like to emphasize that, in the case where $\lambda < 1$, we render points (x_i^{RS1}, y_i^{RS1}) unique pixels. In the case where $\lambda > 1$, we once more render points (x_i^{RS1}, y_i^{RS1}) unique pixels, but, now, if a pixel is missing in the chain that forms C_1^{RS} , then we generate it by a simple linear interpolation; we note that a cubic interpolation also works pretty well, especially in the case where two or even more consecutive pixels are missing. It goes without saying that we maintain the very same symbol C_1^{RS} for the finally obtained chain of pixels.

- A5. We parallel translate the entire digital curve C_1^{RS} by δx along the x -axis and by δy along the y -axis, where $\delta x \in \{\delta x_{min}, \delta x_{max}\}$ and $\delta y \in \{\delta y_{min}, \delta y_{max}\}$. For the application in hand

$$\delta x_{min} = \delta y_{min} = -11, \quad \delta x_{max} = \delta y_{max} = 11,$$

while we have partitioned the latter intervals, using a step of 0.5. In this way we obtain the Rotated, Scaled and parallel Translated version of letter contour C_1 , for which we use the obvious symbol $C_1^{RST}(\phi, \lambda, \delta x, \delta y)$.

- A6. For each quadruple $(\phi, \lambda, \delta x, \delta y)$, we evaluate the area enclosed by $C_1^{RST}(\phi, \lambda, \delta x, \delta y)$. We have already computed the area enclosed by C_2 . In order to avoid mixing pixels' co-ordinates with positive and negative values, we parallel translate C_2 and $C_1^{RST}(\phi, \lambda, \delta x, \delta y)$ to first quadrant of the frame F , in order that their centers of mass coincide with the center of this quadrant. Then, we compute the following similarity criterion between realizations' L_1 and L_2 :

Let $D(C_2)$ be the domain enclosed by contour C_2 , i. e. the body of realization L_2 ; similarly let $D(C_1^{RST}(\phi, \lambda, \delta x, \delta y))$

be the domain delimited by contour $C_1^{\text{RST}}(\varphi, \lambda, \delta x, \delta y)$, as defined in action A5 before. Moreover, we symbolize with

$$E^I((D(C_2) \cap D(C_1^{\text{RST}}(\varphi, \lambda, \delta x, \delta y))))$$

the area of the intersection of domains $D(C_2)$ and $D(C_1^{\text{RST}}(\varphi, \lambda, \delta x, \delta y))$ and with

$$E^U((D(C_2) \cup D(C_1^{\text{RST}}(\varphi, \lambda, \delta x, \delta y))))$$

the area of the intersection of these two domains.

Then, the similarity between L_1 and L_2 is decided by using the quantity

$$SC(\varphi, \lambda, \delta x, \delta y) = \frac{E^I(D(C_2) \cap D(C_1^{\text{RST}}(\varphi, \lambda, \delta x, \delta y)))}{E^U(D(C_2) \cup D(C_1^{\text{RST}}(\varphi, \lambda, \delta x, \delta y)))}$$

and by obtaining each maximum value, say SC_{max} among all quadruples $(\varphi, \lambda, \delta x, \delta y)$. One may, in a rather straightforward manner, conclude that $SC_{\text{max}} = 1$ holds, if



Fig. 1 Demonstration of the similarity criterion between two realizations of ‘ρ’. The intersection of the two realizations, is depicted in black, while their union is shown with all degradations of gray, except the white one.

and only if realizations are identical, excluded the three aforementioned affine transformations, i. e. rotation, scaling and parallel translation (see Fig. 1). On the contrary, the smaller SC_{max} ; the greater is the dissimilarity of L_1 and L_2 .

4. DECIDING IF TWO DOCUMENTS HAVE BEEN WRITTEN BY THE SAME HAND

Suppose that two (2) distinct documents, say D_1 and D_2 , are given for testing if they are written by the same hand or not. To accomplish this task, we take two steps:

First, we decide if all realizations of a single alphabet symbol, which, generically we will call ‘ρ’, have been written by the same hand in both documents or not.

Second, we repeat the first step of a considerable number of alphabet symbols, like the ones chosen in Section 2.

On the basis of the previous two steps, we express a general likelihood concerning the writer(s) of D_1 and D_2 . The approach is a substantial modification and extension of the one introduced in [14], [17], [18] and in connection with Athenian inscriptions.

4.1 Deciding if all realizations of a single alphabet symbol have been written by the same hand

We consider all ‘ρ’ realizations appearing in D_1 , say $\rho_1^{D_1}, \rho_2^{D_1}, \dots, \rho_{N_1}^{D_1}$. We let the contour of the arbitrary realization $\rho_i^{D_1}$, say $C\rho_i^{D_1}$, play the role of the “fixed” curve

C_2 and the contour of realization $\rho_j^{D_1}$, $i \neq j$, say $C\rho_j^{D_1}$ play the role of C_1 that will undertake the affine transformations defined in Section 3. We apply to these two contours the entire approach introduced in Section 3, thus obtaining the similarity criterion $SC_{\text{max}}(D_1, i; D_1, j)$, where the first pair in the parenthesis before the semicolon refers to $C\rho_i^{D_1}$, while the second pair after the semicolon to $C\rho_j^{D_1}$. The obtained bundle of optimally fit ‘ρ’ realizations, in this way, are shown in Fig. 2.

We apply this procedure for all distinct realizations of ‘ρ’ in D_1 . In this way, we end up with $M^{D_1} = \frac{N_1(N_1-1)}{2}$ values that express the degree of similarity of any two distinct realizations of ‘ρ’ in D_1 . We symbolize the mean value of $SC_{\text{max}}(D_1, i; D_1, j)$ with $mv_{D_1}^{D_1}$ and their standard deviation with $S_{D_1}^{D_1}$, where the superscript in both symbols stands for the ensemble of fixed curves, while the subscript the ensemble of the transformed curves.

We repeat the aforementioned procedure, but this time we compare contours $C\rho_i^{D_1}$ with all N_2 contours $C\rho_j^{D_2}$ in document D_2 . Consequently, we obtain $M_{D_2}^{D_1} = N_1 N_2$ values $SC_{\text{max}}(D_1, i; D_2, j)$ that convey the degree of similarity of any two realizations of ‘ρ’, where the fixed one $C\rho_i^{D_1}$ belongs to D_1 , while the transformed one $C\rho_j^{D_2}$ belongs to D_2 . The corresponding obtained bundle of optimally fit ‘ρ’ realizations is shown in Fig. 3. We symbolize the mean value of $SC_{\text{max}}(D_1, i; D_2, j)$ with $mv_{D_2}^{D_1}$ and their standard deviation with $S_{D_2}^{D_1}$. Then, we apply the subsequent statistical criterion:

First, we compute quantity

$$Q_p(D_1, D_2) = \frac{mv_{D_1}^{D_1} - mv_{D_2}^{D_1}}{\sqrt{\frac{(S_{D_1}^{D_1})^2}{M^{D_1}} + \frac{(S_{D_2}^{D_1})^2}{M_{D_2}^{D_1}}}}$$

that follows a Student distribution with $D = N_1 + N_2 - 1$ degrees of freedom. For alphabet symbol ‘ρ’ and if we consider that D_1 is the document “discovered” in AHD, while D_2 is the document of NRFEV, which has been definitely written by E. Venizelos, then it follows that the hypothesis that all realizations of alphabet symbol ‘ρ’ in the latter two documents have been written by the same hand, cannot be rejected practically with certainty.

4.2 Deciding if the document of AHD has been written by Eleftherios Venizelos

By repeating the previous process for the remaining nine (9) letters, too, referred to in Section 2, we have deduced that the hypothesis that the document spotted in AHD and the one of NRGEV have been written by the same hand, cannot be rejected with even greater degree of certainty. On the contrary, the hypothesis that the Document of AHD had been written by the same hand with the one that wrote the “Greek Constitution” is practically rejected. This is graphically manifested in Fig. 4, where all realizations of ‘ρ’ in the document of AHD (depicted in blue) as well as those of the “Greek Constitution” are optimally fit to a single ‘ρ’ realization of AHD document. Very similar results hold in connection with the document written by the Greek citizen in the beginning of the 20th Century.

5. CONCLUSION

In the present work, a quantitative and almost unambiguous answer is given concerning the hand that has written an important document dragged up in the warehouse of the AHD. Towards this goal, the authors presented a novel a general method, here, which employs an exhaustive comparison of many realizations of the same alphabet symbol appearing in various texts. In fact, this comparison is based on a) rendering each realization a prototype (fixed) letter, b) applying the affine transformations, rotation, scaling and parallel translation to all other realizations of the same alphabet symbol and c) testing the similarity of all aforementioned pairs of realizations, via a concrete criterion. The methodology proved that the document spotted in AHD belongs to the great Greek politician Eleftherios Venizelos, with a degree of confidence greater than 99.99% or practically with certainty.



Fig. 2 All realizations of ‘ρ’ appearing in document D_1 , optimally fit to a single corresponding realization depicted by a wider line.

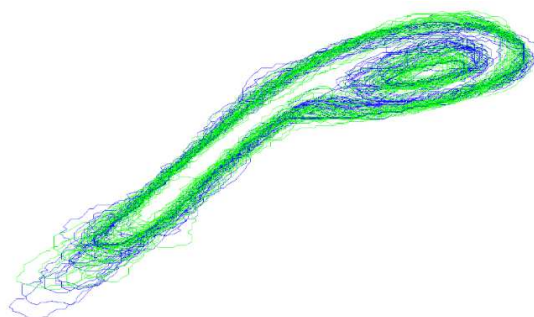


Fig. 3 The two bundles of all realizations of alphabet symbol ‘ρ’, appearing in documents D_1 (shown in blue) and D_2 (shown in green). It is evident that all these contours belong to the same population, demonstrating that the document of AHD has been written by Eleftherios Venizelos.

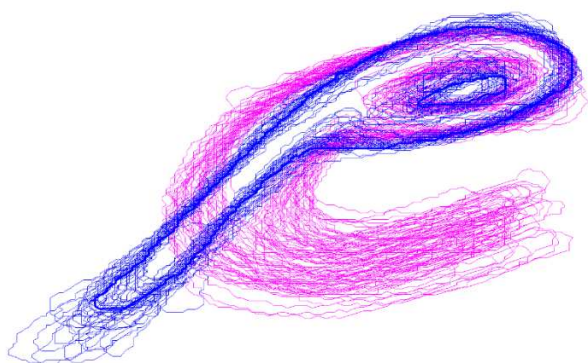


Fig. 4 The two bundles of all realizations of alphabet symbol ‘ρ’, appearing in the document of AHD (shown in blue) and of the ‘Greek Constitution’ (shown in magenta). The dissimilarity of these two bundles indicate that these two documents have been written by different hands.

REFERENCES

- [1] Y. Hannad, I. Siddiqi, and M. E. Y. El Kettani, “Writer identification using texture descriptors of handwritten fragments,” *Expert Syst. Appl.*, vol. 47, pp. 14–22, Apr. 2016, doi: 10.1016/j.eswa.2015.11.002.
- [2] X. Zhang and G. Nagy, “Computational method for calligraphic style representation and classification,” *J. Electron. Imaging*, vol. 24, no. 5, p. 053003, Sep. 2015, doi: 10.1117/1.JEI.24.5.053003.
- [3] A. J. Newell and L. D. Griffin, “Writer identification using oriented Basic Image Features and the Delta encoding,” *Pattern Recognit.*, vol. 47, no. 6, pp. 2255–2265, Jun. 2014, doi: 10.1016/j.patcog.2013.11.029.
- [4] A. Bensefia and T. Paquet, “Writer verification based on a single handwriting word samples,” *EURASIP J. Image Video Process.*, vol. 2016, no. 1, p. 34, Oct. 2016, doi: 10.1186/s13640-016-0139-0.
- [5] S. He, P. Samara, J. Burgers, and L. Schomaker, “Historical manuscript dating based on temporal pattern codebook,” *Comput. Vis. Image Underst.*, vol. 152, pp. 167–175, Nov. 2016, doi: 10.1016/j.cviu.2016.08.008.
- [6] S. He and L. Schomaker, “Writer identification using curvature-free features,” *Pattern Recognit.*, vol. 63, pp. 451–464, Mar. 2017, doi: 10.1016/j.patcog.2016.09.044.
- [7] S. Faigenbaum-Golovin *et al.*, “Algorithmic handwriting analysis of Judah’s military correspondence sheds light on composition of biblical texts,” *Proc. Natl. Acad. Sci.*, vol. 113, no. 17, pp. 4664–4669, Apr. 2016, doi: 10.1073/pnas.1522200113.
- [8] M. Diem, F. Kleber, S. Fiel, and R. Sablatnig, “Semi-automated document image clustering and retrieval,” in *Document Recognition and Retrieval XXI*, Mar. 2014, vol. 9021, pp. 206–215. doi: 10.1117/12.2043010.
- [9] R. Shweka, Y. Choueka, L. Wolf, and N. Dershowitz, “Automatic extraction of catalog data from digital images of historical manuscripts,” *Lit. Linguist. Comput.*, vol. 28, no. 2, pp. 315–330, Jun. 2013, doi: 10.1093/lit/fqt007.
- [10] L. Wolf *et al.*, “Identifying Join Candidates in the Cairo Genizah,” *Int. J. Comput. Vis.*, vol. 94, no. 1, pp. 118–135, Aug. 2011, doi: 10.1007/s11263-010-0389-8.
- [11] S. He, M. Wiering, and L. Schomaker, “Junction detection in handwritten documents and its application to writer identification,” *Pattern Recognit.*, vol. 48, no. 12, pp. 4036–4048, Dec. 2015, doi: 10.1016/j.patcog.2015.05.022.
- [12] C. De Stefano, M. Maniaci, F. Fontanella, and A. Scotto di Freca, “Reliable writer identification in medieval manuscripts through page layout features: The ‘Avila’ Bible case,” *Eng. Appl. Artif. Intell.*, vol. 72, pp. 99–110, Jun. 2018, doi: 10.1016/j.engappai.2018.03.023.
- [13] N. D. Cilia, C. De Stefano, F. Fontanella, C. Marrocco, M. Molinara, and A. S. di Freca, “An Experimental Comparison between Deep Learning and Classical Machine Learning Approaches for Writer Identification in Medieval Documents,” *J. Imaging*, vol. 6, no. 9, Art. no. 9, Sep. 2020, doi: 10.3390/jimaging6090089.
- [14] S. V. Tracy *et al.*, “Identifying Hands on Ancient Athenian Inscriptions: First Steps Towards a Digital Approach*,” *Archaeometry*, vol. 49, no. 4, pp. 749–764, 2007, doi: 10.1111/j.1475-4754.2007.00333.x.
- [15] C. Papaodysseus, M. Exarhos, M. Panagopoulos, P. Rousopoulos, C. Triantafillou, and T. Panagopoulos, “Image and Pattern Analysis of 1650 B.C. Wall Paintings and Reconstruction,” *IEEE Trans. Syst. Man Cybern. - Part Syst. Hum.*, vol. 38, no. 4, Art. no. 4, Jul. 2008, doi: 10.1109/TSMCA.2008.923078.
- [16] D. Arabadjis *et al.*, “A Novel Information System for the Automatic Reconstruction of Highly Fragmented Objects with Application to the Reassembly of Prehistoric Wall Paintings and Vessels,” in *10th International Symposium on the Conservation of Monuments in the Mediterranean Basin: Natural and Anthropogenic Hazards and Sustainable Preservation*, M. Kouli, F. Zezza, and D. Kouli, Eds. Cham: Springer International Publishing, 2018, pp. 571–578. doi: 10.1007/978-3-319-78093-1_62.
- [17] S. V. Tracy and C. Papaodysseus, “The Study of Hands on Greek Inscriptions: The Need for a Digital Approach,” *Am. J. Archaeol.*, vol. 113, no. 1, pp. 99–102, 2009.
- [18] M. Panagopoulos, C. Papaodysseus, P. Rousopoulos, D. Dafi, and S. Tracy, “Automatic Writer Identification of Ancient Greek Inscriptions,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 8, pp. 1404–1414, Aug. 2009, doi: 10.1109/TPAMI.2008.201.