A case study on the System for Paleographic Inspections (SPI): challenges and new developments

Fabio AIOLLI [a,1], Arianna CIULA [b]

[a] *Dept. of Pure and Applied Mathematics, University of Padova, Italy*
[b] *Centre for Computing in the Humanities, King's College London, England*

Abstract. Of all the disciplines that study the past through its written heritage, paleography[2] is the one that focuses mainly on the analysis of the script borne by manuscript books. The goal of the paleographical method is to date and to localize books produced by hand through the analysis of their ancient scripts. This chapter describes the System for Paleographic Inspections (SPI) software suite developed at the University of Pisa, how the system has been used by paleographers in their attempts to classify and identify scripts, and how it can be improved further to meet the research needs of paleographers.

Keywords. Paleographic Analysis, Segmentation, Tangent Distance, SPI

Introduction

What allow paleographers to locate handwritten material, being it entire books or tiny snippets of handwriting, within a particular chronological interval and within a specific cultural environment, are many different aspects of the book (e.g. various crafts involved in its manufacture, notes made by scribes or illuminators, subject specific indications, history of provenance and textual tradition, decoration). However, a crucial role is played by the graphic comparison of the manuscripts under study with similar witnesses, which are dated or datable, and localized. It deals with a method based on "formal and stylistic comparison between what has reached us with a defined paternity (or date) and what can supposedly be brought back to the same paternity (or date)".[3] It is this comparison and the particular observation of graphic occurrences in context that enables the paleographer to discover pertinent similarities. It is the same comparative method that reveals unconsciously individual graphic styles, which are indispensable clues to establish the identity or non-identity of hands.

---

[1] Corresponding Author: Fabio Aiolli, Dept. of Pure and Applied Mathematics, Via Trieste 63, 35121 Padova, Italy; E-mail: aioli@math.unipd.it

[2] For an introduction to paleography see [1].

[3] Free translation from the Italian "confronto formale e stilistico fra quanto ci è giunto con paternità (o data) vicina e quanto si presume possa essere ricondotto alla stessa paternità (o data)" in [2], pag. 18.

Consequentially, while carrying on an analysis of old scripts, the relationship between writing styles is crucial for paleographers to infer spatio-temporal locations of documents especially when there is no other clear evidence about the origin of a manuscript. Often, in these cases, experts disagree on which morphological features should be considered relevant and different experts may have different types of expertise and experience.

The System for Paleographic Inspections (SPI) is a software that was developed by Aiolli [3,4] during its Master Degree thesis under the supervision of Prof. Sperduti and Prof. Starita at the Computer Science department of the University of Pisa in the 1999, in collaboration with Dr. Zaccagnini of the department of Medieval Studies at the University of Pisa. The software was the result of a preliminary study of several years. [4]

The aim behind the development of SPI was to provide an adaptive support system, built on rigorous statistical techniques, that, on the basis of old manuscript images, generates models of different writing styles. These models are in turn used to relate 'new' unseen documents with the ones already processed by the system. The idea is to support the paleographical research by facilitating a rigorous analytical approach, where quantitative comparisons can be used to confirm or discard hypotheses concerned with dating and localization.

Recently, SPI has been successfully tested on a concrete corpus of manuscripts held at the public library in Siena (Italy). The software is currently under revision and redevelopment so as to become multiplatform and open source, to provide image processing filters for ancient and damaged manuscript images, to include a markup model to wrap annotations that can be then searched semantically, and to enrich the visualization capability of its internal comparative tools.

In Section 1, a brief survey of the SPI system is given; its three main modules, namely, the segmentation, the learning, and the analysis modules are described in some detail. In Section 2, the application of the system on a real paleographical study performed at the University of Siena is described together with its methodological implications. Finally, in Section 3, we report on current development of the system and suggest new research directions in order to provide paleographers with an even more sophisticated tool.

1. SPI (System for Paleographic Inspections)

SPI consists basically of three modules or tools:

- a *Segmentation Tool* exploiting image processing techniques that is able to extract letters from a digitalized manuscript;
- a *Learning Tool*, where letter-based prototypes based on tangent distance are generated (these models provides an approximate representation of the handwriting style identifiable in a given manuscript image);
- an *Analysis Tool*, where a paleographer can classify and study the relationships between handwriting styles and sets of manuscript images.

[4] Previous work was devoted to study the applicability of pattern recognition techniques to palaeographic analysis (e.g., [5], [6], [7]).

A very important aspect that has driven the design of SPI was the necessity to extract knowledge while, at the same time, preserving an understandable representation, helpful for human analysis. To this aim, in SPI, a manuscript is viewed as a collection of character images and every object of study in SPI is an image. Another important aspect was the design of a system able to capture the main features characterizing the writing styles. For this reason, the output of the system had to be robust enough to tolerate global transformations of a character (e.g. rotations, scaling, location shifts, etc.) which are not normally considered as relevant features to characterize a writing style. These two last points motivated the use of tangent-distance based techniques which appeared to be particularly suitable for handwritten character/digit recognition tasks (see [8]) allowing a certain degree of invariance to given input transformations. Specifically, tangent distance based methods are used in the learning and classification modules of SPI.

The idea behind tangent distance is to consider, for each pattern, the manifold that represents the set of patterns which can be obtained from the original pattern through a chosen set of transformations. An ideal transformation invariant distance between two patterns would compute the minimum distance between the associated manifolds. However, given the non-linearity of the manifolds, this computation would be infeasible. Thus, in [8] an approach based on the local linear approximation of the manifold is proposed. Finally, the computation of the tangent distance is reduced to the computation of a minimum distance between two low dimensional linear spaces.

In a typical SPI session (see Figure 1) the expert can first extract or segment the characters from a digitalized manuscript image (which was previously stored in the database) by means of the Segmentation Interface of the system. These characters are hence stored in the palaeographic database. Once a sufficient number of characters have been extracted, the palaeographer is able to produce the tangent-distance based models for each letter of interest in the manuscript. This operation is performed by relying on the Learning Interface and the generated models are stored in the palaeographic database. Finally, the palaeographer can analyze the manuscript by using the Analysis tool which offers different ways to look at the relationships between handwriting styles and sets of documents. In the following sections, the three modules of SPI are described in more detail.

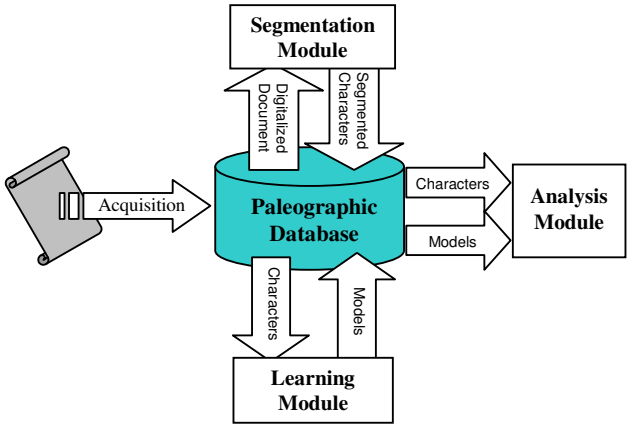

Figure 1. View of the interactions between the three modules of SPI and its paleographic database.

1.1. Segmentation Module

The segmentation module of SPI allows the user to extract single characters (letters or connected letters) from a digitized manuscript image. It is important to note that, since the main objective is to perform a morphological analysis of the extracted characters, it is crucial to capture all relevant morphological features of a pattern. Therefore, one of the critical aspects of such a system is the decision of where to cut the junctions between characters. Moreover, in some cases, two letters may form a single morphological object (e.g. as it is the case for connected letters, a.k.a. ligatures).

To this aim, a semi-automatic segmentation algorithm has been devised [7] which requires a minimal interaction with the user as she is expected to click with the mouse approximately on the center of a given letter. Starting from this selected point, the algorithm computes a box which contains the whole character. This box is explored in order to determine the real extension of the character. At each step in the calculation of this box, new vertical segments are added so as to determine new admissible segmentations. At the end of the process, a final segmentation is obtained: it represents the solution that minimizes the distance from a typical segmentation prototype, which is learned *a priori* and consists of *circularity* and *complexity* features of a typical character within the selected letter class.

1.2. Learning Module

The learning module constitutes the core of the system. Given a set of letters of the same type belonging to a given manuscript, it produces a model summarizing their morphological features. This model has to be understandable by human experts and give them the possibility to analyze the specific features and the variability of a script style. To this end, learning techniques based on tangent-distance are used as they are particularly suited to this context. The algorithm we actually use is based on the work in [9] and consists of generating models (a centroid and a set of tangent vectors) which, similarly to PCA (Principal Component Analysis), represent linear subspace approximations of the space generated by the letters under consideration.

An interesting aspect of this kind of models is that the centroid graphically represents the features which are shared by the different letters, while the tangents represent the type of deviations in the writing style with respect to these common features. An example of a centroid of such a model, and the set of characters used to generate it, can be found in Figure 2.

1.3. Analysis Module

Once the learning phase is terminated, the generated models can be used both to classify handwriting styles identified in 'new' manuscript images and to perform a study of the morphological features of a given manuscript image. This can be done with the following tools available in SPI.

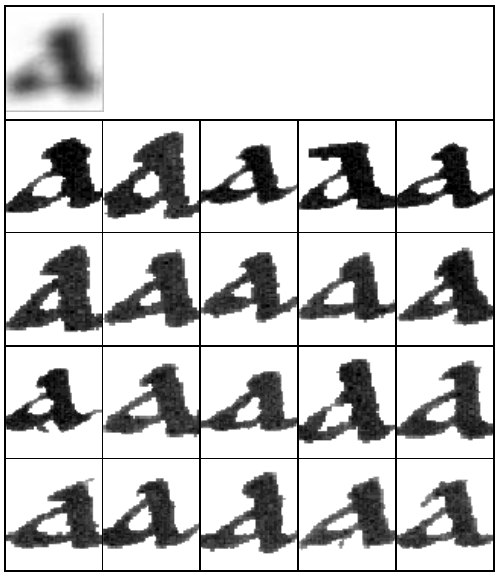

Figure 2. Model centroid (top left corner) generated from 20 letters a extracted from the digital images of the codex FIII3 (BCI, Siena).

1.3.1. Classification

Given a new manuscript image, one of the tools which is available within SPI allows the user to retrieve its closest (from a morphological point of view) stored manuscript. The most similar script is determined by a nearest-neighbor algorithm. In particular, the system compares the characters extracted from the new document with all the models of the same letter class stored into the system. The comparison is done using tangent-distance. A similarity score is then given to each of these documents based on the number of characters for which the model associated to this document was the closest .

1.3.2. Model Morphing

Another interesting tool provided within SPI allows the user to visualize the span of the tangent model by showing which transformations are characteristic of the letters found in a given script (see Figure 4). This tool highlights the morphological modifications a writing style is subject to by visualizing the morphing effect on the centroid obtained by moving it within the associated tangent space model.

1.3.3. Complete Diagram

Another tool builds up and visualizes the complete diagram of the pair-wise tangent distances between documents stored into the system. Given a prototype, the other prototypes are then ordered on the basis of their similarity to the model under study (see Figure 5).

1.3.4. Dendrogram

The last internal tool allows a global analysis of all the models of a given letter, by drawing a *dendrogram* based on the Average Linkage cluster algorithm. The dendrogram is a binary tree where the information about the similarity between clusters of models is visualized in the form of a topologic similarity in a hierarchical tree (see Figure 6).

2. Application of SPI to a Paleographic Corpus

To test the real support that the software is able to offer to the palaeographical analysis, a corpus of manuscripts was selected and digitized according to specific criteria as described in [10,11,12,13,14,15][5] and briefly reported below. Subsequently, the relevant letters and ligatures were segmented to generate the corresponding models. As part of this work, some qualitative interpretations have been associated to the quantitative models.

2.1. Manuscript Corpus

The corpus consisted of samples from around forty codices[6] held at the *Biblioteca Comunale degli Intronati* in Siena (Italy), written in Caroline minuscule and pre-Gothic script, datable between the tenth and the twelfth centuries. The reason for such a choice was twofold.

First of all, the palaeographical studies on the Caroline script, especially on the chosen manuscripts, were dated and in need of further analysis. Despite the vastness of handwritten witnesses belonging to this chronological interval, this material has been neglected as far as the study of both its graphical aspects and textual characteristics is concerned. In addition, scholars have been concentrating their efforts on the study of local centres, and in doing so have acquired specialized knowledge of only some writing environments active in this period. In particular, central Italy (where the manuscripts under study are currently stored), with the exception of the area characterised by the Caroline type known as *romanesca*, is still waiting for a deep historical, philological and palaeographical analysis of its manuscript heritage.

The second reason is a more pragmatic one. Besides being an appealing subject of study, Siena and its library were a feasible target for this test. The codices were accessible, and therefore could be selected, observed directly and synoptically when necessary, and captured digitally.

---

[5] This work was carried out as part of the second author's PhD thesis [15] on Manuscript and Book Studies (Scienze del Libro) at the University of Siena-Arezzo completed in June 2005.

[6] Within the corpus in question, the typical manuscript book or codex is a high size codex. It consists of parchment or membrane sheets gathered together so as to form quires, typically of eight leaves, where the lines are ruled with a hard point or, in very few cases, lead point. Usually, numbered quires and catchwords are also present. The original foliation coexists today with one or more modern foliation or pagination. Due to the process of restoration accomplished mainly in the nineteenth century, the majority of these codices do not feature the original binding anymore. The painted initials and titles, or rubrications, are more or less distinctive. In some particularly elaborated cases, historiated initials and miniatures are also present. The corpus includes, above all, liturgical texts (see [14] for further details). The main textual unit is usually written by more than one scribe per codex, either full page or in two columns, and it may be annotated or corrected by contemporary or later glosses.

The past studies of art history and palaeography (in particular [16] and [17]) have not been able to identify a *scriptorium* as a common source for such a variety of codices. Due to the diversity of the writing style and to the rather rare information on the copying activity in the relevant monasteries, several codices have been classified as generally belonging to the area of central Italy. Hence, the general paleography of the corpus (based on [18,17,18,16,20]) preceding the experimentation carried out with SPI could be described as following:

- Extreme of not formalized Caroline minuscule (IX-X centuries);
- 1st group (XI century);
- 2nd group (XII first half);
- 3rd group (XII second half);
- Extreme of incipient Gothic script (XIIex-XIIIin).

2.2. Digitization Criteria

As far as the digitization process regards, for every codex belonging to the corpus, at least four leaves were captured for every style that seemed to respond to a certain uniformity of script. More pages were scanned when evidently inhomogeneous hands occurred in the same codex, so as to maintain the proportion of four pages per style.

The images with resolution of 300 dpi were archived in TIFF format on CD-ROMs as general backup for preservation purposes. As required by the image processing settings within SPI, the image files were subsequently converted into bitmaps with two levels of quantization. The digital leaves were then cropped and loaded into the database as sections (corresponding to columns of handwriting when possible).

It has to be noted that the digitization process is not systematically recorded or supported within SPI. Indeed, when initially developed, SPI was not conceived as an image management tool; therefore, besides the image file system, no metadata are required and can be introduced while loading the image files. However, so as to facilitate the work of paleographers and document the digitization process in more transparent terms, the new development of the application will allow for the image metadata to be stored, so as to track the digitization criteria and any manipulation of the images preceding any further automatic processing.

2.3. Segmentation

Once the images were loaded, the segmentation was performed page by page, starting from the letter *a*, then *d*, *f*, *g*, *m*, *n*, *p,* and *u*. It would have been ideal to perform the segmentation and the following phases for the whole handwriting system, for the ligatures and, eventually, for the punctuation. However, this was not possible both for limitations of time and because of the reduced alphabetical grid from which to select characters within SPI.

Indeed, the menu of available letterforms to be segmented is another area where new development will be concentrated on in the near future, so as to offer paleographers extended and, eventually, extensible handwriting system menus with the provision of a separate module where the application can be trained to 'learn' new typical letters and ligatures for segmentation purposes.

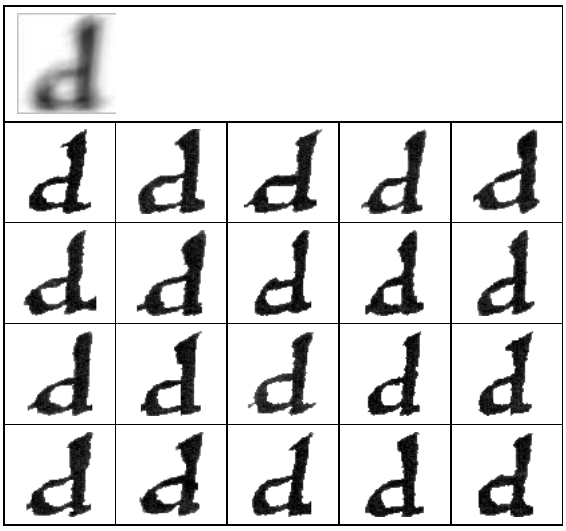

Figure 3. Model centroid (top left corner) generated from 20 letters *d* extracted from the digital images of the codex FIII3 (BCI, Siena). The model was described as following: the rather upright ascender of the letter *d* in FIII3 may present different serifs at its top: a club-shaped serif, which is leftwards cut and sometimes, as in the *b*, is decorated by a leftwards-line serif, or a wedge-shaped serif. The exit of the ascender is slightly curved towards the right to form a tiny hook. The lobe is regular in shape.

2.4. Results

As a result of the experimentation with SPI, some detailed descriptions of the morphological variants that occurred in the Caroline script in central Tuscany between the tenth and the twelfth centuries were produced. Although these descriptions derive from the application of the traditional palaeographical method (the terminology is, for instance, entirely borrowed from the palaeographical literature) they are intertwined with a new methodological approach based on the digital models generated by SPI. The captions of Figures 3 and 4 provide examples of such analysis.

As explained above, thanks to the so called tangent-distance based algorithms, the models can be compared numerically with the use of different tools internal to the system (see Figures 5,6,7).

The comparative analysis of the specific occurrences of the Caroline minuscule under study was also carried out outside the tool by drawing some tables, where graphical samples and textual descriptions could be associated and interpreted, so as to reflect on the regional evolution of the script (see Table 1).

Indeed, the overall research has produced some results in terms of the analysis of the handwriting and of the clustering of the manuscripts on the basis of their morphological similarity. This resulted in the reorganization of the paleography of the corpus as following:

- Caroline of IX-X centuries;
- Caroline *tipizzata* of XI ex-first half of XII century;
- Caroline as imitation of Roman Caroline of XI ex-third quarter of XII century;

- Caroline with stylistic interferences of XII med.;
- Caroline of beginning of canon of second quarter-half of XII century;
- “formed” Caroline from second half of XII century;
- Scripts from XIIex-XIIIin (incipient Gothic, round script, Gothic script).

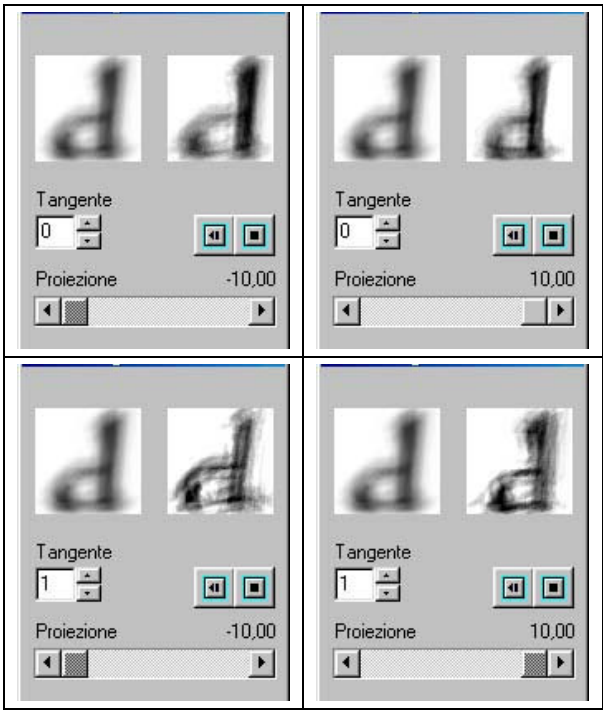

Figure 4. Example of tangent analysis for the letter *d* of codex FIII3, BCI, Siena. A detailed description was carried out model by model as following. The first tangent (top left corner) shows a variation in the module of the lobe, wider or smaller.. The club-serif serif of the ascender is cut leftwards and the hook at its bottom may function as linkage to the following letter. The third and fourth tangent show that the body of the letter can vary in the slant of the lobe, higher or flatter, and of the shaft, slightly rightwards or upright. When the overall slant is slightly rightwards, the serif is thinner and simpler, otherwise is evidently club-shaped. Linkages on both side of the body are visible. Note that when the slant of the *d* is upright, a linkage is present on the left of the letter. This means that it probably links with the previous letter. On the other hand, when the slant is rightwards, a linkage is positioned on the right of the letter, towards the following character. Possibly, the slant is determined by the speed of the hand that is already proceeding toward the letters to be written.

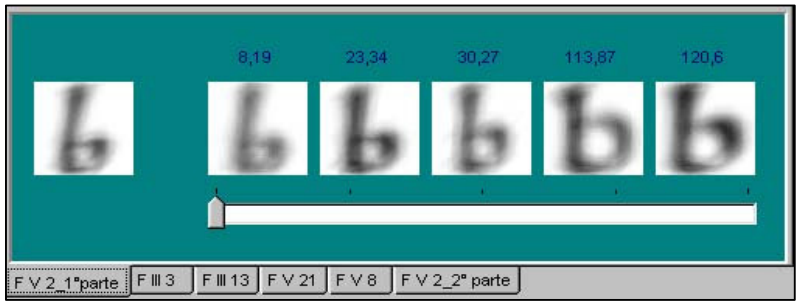

Figure 5. Example of the use of the diagram tool within SPI to compare various model of the letter *b* in quantitative terms.

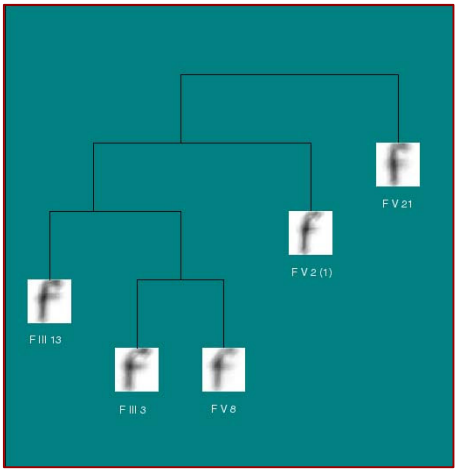

Figure 6. Example of the use of the dendrogram tool within SPI to compare various models of the letter *f* in quantitative terms.

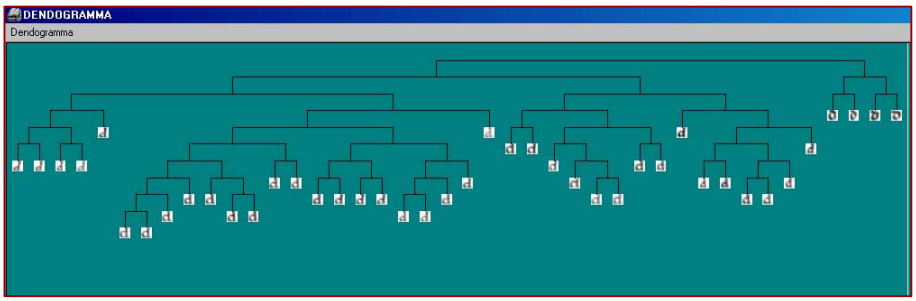

Figure 7. Example of the use of the dendrogram tool within SPI to compare models of the letter *d* for the entire corpus in quantitative terms.

Table 1. Example of the categorization of some digital models of the letter *g*.

Group 1										
(i)				(ii)			(iii)			
FIII3	FV12	FI10 A	GI4	GVII12	FX4	FII13	FV2 1°	FIII13		
(iv)					(v)		(vi)			
FV8		FIV24			FI6		FVIII16			
Group 2					Group 3					
FIX37					HVII10					

2.5. Methodological Implications

The application of SPI software to the paleographic corpus described above proved useful to augment a qualitative paradigm of analysis with the creation of graphical models. These are quantitative in nature, but visibly available on the screen and therefore useful to show the variations of each handwriting style, while being at the same time open to a further description and interpretation.

The paleographical comparative method is therefore enriched and changed. Indeed, from the beginning, a terminological coherence is needed to describe the models and to compare them. Ideally, to make the most of the tool, the corpus needs to be structured under multiple criteria (e.g. date, type of script, signatures etc.).

As explained expensively in this chapter, SPI computes some features expressed directly by the letterforms. Indeed, the morphological parameters gathered from the patterns in exam can be calculated and recorded as part of the digital representation. These added computational parameters widen the classical system of paleographical characteristics and facilitate a terminological categorization based on the graphical models. The interpretations are forced to be anchored to digital representations and their correspondent numerical values. The result is a quantitative approach with considerable representational and descriptive power.

3. Under Development

Although started as a pilot project some years ago now, the development of such an application remains pioneering for two main reasons:

1. its interdisciplinary approach (the rigor of advanced computing technologies has to support the ambiguities and nuances of research methods in the humanities); [7]
2. the challenges it poses both to pattern recognition techniques and image processing.

After having tested SPI on a concrete corpus of manuscripts, the software is currently under revision and redevelopment so as to:

1. become multiplatform (it was originally developed to run only on Windows 98) and open source;
2. provide image processing filters for ancient and damaged manuscript images;
3. include state-of-the-art large margin learning algorithms having the same properties of invariance to given transformation and easy interpretability of the results as in tangent-distance based algorithms [22].
4. include a markup model to wrap annotations so as to provide for semantic search;
5. visualize simultaneously and compare automatically diagrams and dendrograms;
6. make it multilingual and update the graphical interface.

Some of the points above have been mentioned in the previous sections of this chapter. However, point 4 in particular deserves some further discussion.

As stated above, the process of defining a nomenclature to abstract from and 'reduce' the variance of individual manuscripts is part of the traditional palaeographical method and was very much facilitated by the use of SPI, where every single graphical instance is decomposed into models that can be described one by one. However, beyond verbose descriptions, a 'descriptive protocol' for such an annotation is not structured or configurable within the tool.

Furthermore, while using SPI the paleographer is encouraged to concentrate on the minute features of the letterforms and on their models. This process certainly enhances the rigor of the analysis, but at the same time it does not take into consideration the overall appearance of the script and its immediate context. Indeed, apart from the segmentation process, where the cropping of characters is performed within the digital image-page or within one of its sections, the rest of the analysis focuses on individual letters and therefore it leaves aside the setting of the script (e.g. while the user is focusing on the analysis of the models, the collocation of letters with respect to the pair of lines of writing is not made available anymore by the system). So, ideally, not only it would be desirable to be able to record some structured descriptions of the models (e.g. by using a semantic textual encoding scheme)[8] and attach them to the graphical prototypes, but it would also be desirable if the models could be contextualized and brought back visually to where they were derived from.

---

[7] On the subject of humanities computing and the challenges of modelling interpretations see [21].

[8] A customization of the Text Encoding Initiative Guidelines [23] could be attempted for this purpose. On the issue of stroke-level markup see [24].

References

[1] B. Bischoff. Latin Paleography. Antiquity and the Middle Ages, trans. By Dáibhí O Cróinín and David Ganz. Cambridge: Cambridge University Press, 1990.

[2] P. Supino Martini. "Sul metodo paleografico: formulazione di problemi per una discussione." *Scrittura e Civiltà* 19 (1995): 5-29.

[3] F. Aiolli. *Sviluppo di un Sistema Basato sulla Distanza Tangente per l'Analisi Morfologica di Caratteri Paleografici*, Università degli Studi di Pisa, Italy 1999.

[4] F. Aiolli, M. Simi, D. Sona, A. Sperduti, A. Starita, G. Zaccagnini. "SPI: a System for Paleographic Inspections." *AIIA Notizie* (1999).

[5] F. Masulli, D. Sona, A. Sperduti, A. Starita, G. Zaccagnini. "A System for the Automatic Morphological Analysis of Medieval Manuscripts." *Journal of Forensic Document Examination*, Vol. 9, 1996.

[6] D. Sona. *Distanza Tangente per l'Analisi Morfologica in Campo Paleografico*. Università degli Studi di Pisa, Pisa, Italy, 1996.

[7] G. Iommi. *Segmentazione di Caratteri nelle Scrittura Librarie Medievali.* Università degli Studi di Pisa, Pisa, Italy, 1997.

[8] P.Y Simard, Y. Le Cun, J. Denker. "Efficient pattern Recognition using a new transformation distance." *Proceedings of Advances in Neural Information Processing Systems*, volume 5, San Mateo CA, 1993, pp. 50-58

[9] T. Hastie, P.Y. Simard, E. Sackinger. "Learning prototype models for tangent distance". In G. Teasauro, D. S. Touretzky, and T. K. Leen, eds., *Advances in Neural Information Processing Systems*, volume 7, MIT Press, 1995, pp. 999-1006.

[10] A. Ciula. *A research project. The application of SPI Software to the Corpus of Manuscripts held in Siena*. King's College London, London, UK, 2004.

[11] A. Ciula. Digital paleography. Poster presented at *Digital Resources for the Humanities* conference, University of Newcastle (UK), September 2004: <http://drh2004.ncl.ac.uk/abstract.php?abstract=247> (Accessed October 2008).

[12] A. Ciula. "Modelli digitali di scrittura carolina." *Gazette du livre médiéval* 45 (autumn 2004): 27-38.

[13] A. Ciula. "Digital paleography: using the digital representation of medieval script to support paleographic analysis." *Digital Medievalist* 1 (spring 2005): <http://www.digitalmedievalist.org/article.cfm?RecID=2> (Accessed October 2008).

[14] A. Ciula. *Paleografia e Informatica. L'applicazione del software SPI al corpus di manoscritti senesi*. Università degli Studi di Siena-Arezzo, Siena, Italy, 2005.

[15] A. Ciula. L'applicazione del software SPI ai codici senesi. In *Poesìa Medieval*, Vitalino Valcàrcel Martìnez and Carlos Pérez Gonzàles eds., Collecciòn Beltenebros 12, Burgos: Fundaciòn Instituto Castellano y Leonés de la Lengua, 2005, pp. 305-325.

[16] E.B. Garrison. *Studies in the history of mediaeval Italian painting*. I-IV vols. Firenze: L' impronta, 1984.

[17] K. Berg. *Studies in Tuscan Twelfth Century Illumination*. Oslo: Universitetforlaget, 1968.

[18] L Avitabile, M.C. Di Franco, V. Jemolo. "Censimento dei codici dei secoli X-XII." *Studi medievali* 11.2 (1970): 1075-1101.

[19] G.M. Cao, T. Catallo, M. Curandai, E. Di Mattia, P. E. Fornaciari, E. Peruzzi, and F. Santi, comp. *Catalogo dei manoscritti filosofici nelle biblioteche italiane*. VIII: 101- 134. Firenze: Olschki, 1996.

[20] B. Klange Addabbo, *Codici miniati della Biblioteca comunale degli Intronati di Siena*, Siena: Edisiena, 1987.

[21] W. McCarty. *Humanities Computing*. Basingstoke: Palgrave MacMillan, 2005.

[22] F. Aiolli, A. Sperduti. "A re-weighting strategy for improving margin." *AI Journal*, 137 (2002): 197-216.

[23] TEI Consortium, eds. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 1.1.0. Last updated on 4th July 2008. TEI Consortium. <http://www.tei-c.org/Guidelines/P5/> (Accessed October 2008).

[24] M. Terras, P. Robertson. "Downs and Acrosses: Textual Markup on a Stroke Level." *Literary and Linguistic Computing* 19.3 (2004): 397-414.