Image Analysis on the Archimedes Palimpsest

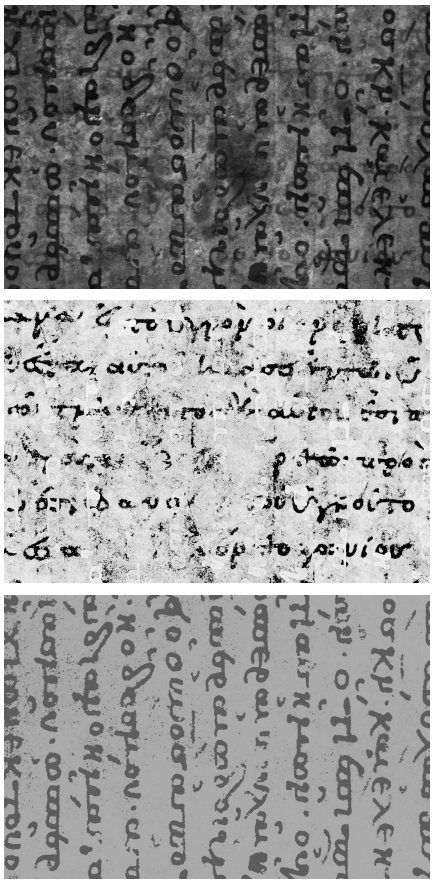
by Anna Tonazzini, Luigi Bedini and Emanuele Salerno

Image-processing procedures developed at ISTI-CNR have been applied to recover some of the hidden writing in the Archimedes palimpsest, an ancient manuscript in which faint remnants of several treatises by the great philosopher and mathematician are partially hidden under a more recent text.

Digital image analysis techniques developed at ISTI-CNR in the framework of the European CRAFT project IsyReaDeT were presented at the Archimedes Symposium in Baltimore, 1-4 April. This important initiative was sponsored by the owner of the Archimedes palimpsest. Since this is the earliest copy available and could reveal many as yet uninvestigated aspects of Archimedes' thinking, the goal is to develop techniques to recover as much of the ancient text as possible.

The manuscript has an incredibly tortuous history: it was first used in Costantinople during the tenth century to copy the works of Archimedes. It was then erased and reused for a Byzantine lithurgic book (an Euchologion) during the twelfth century, probably in southern Italy. This book then travelled through the Holy Land before returning to Costantinople, where it was rediscovered in 1846, in 1899 and again in 1906 when the well-known philologist Johan Heiberg published what he was able to read of the Archimedes text. After the First World War, the palimpsest disappeared, only to emerge yet again in 1998, when an anonymous collector bought it at a Christie's auction for USD 2,000,000. One of the treatises contained in the document is the only existing copy of The Method of Mechanical Theorems, in which Archimedes gives an account of how a philosopher can reach a result with the help of physical intuition, before formulating a rigorous proof. Many points of this work are now seen as anticipating modern calculus, as developed by Newton and Leibniz almost twenty centuries later. Another important component is the Stomachion, which had long been considered lost, and whose content was not known before 1906 (only a fragment of an Arabic translation was available). It now seems that this treatise deals with combinatorics, another largely anticipatory subject for mathematics in the third century B.C.

The palimpsest is now at the Walters Art Museum in Baltimore, which is in charge of its conservation and study. The book has been taken apart, its pages have been partially restored and all the material is now being imaged using various techniques aimed at extracting as much information as possible from the original text. Digital image analysis techniques can play a particularly important role in enhancing the possibility of reading the text. The techniques developed in the IsyReaDeT project have proven particularly useful for this task.



Top: One of the available multi-spectral images of the palimpsest.
Middle and bottom: the separated Archimedes text and overwritten text, obtained by our decorrelation techniques.
Copyright: the Owner of the Archimedes Palimpsest.
Image capture: Rochester Institute of Technology and Johns Hopkins University.

IsyReaDeT began in January 2003 with funding for two years from the European Commission. Its objective was to develop a low-cost system with which to digitize, restore and archive ancient degraded texts. This system is based on a multi-spectral camera with dedicated restoration software. The project consortium includes three research institutions and six small enterprises active in the conservation and restoration of works of art and historical documents.

There is a considerable risk that in the future we may lose many governmental, historical and commercial documents through gradual decay. The digital imaging of important documents can ensure their conservation, enhance their readability and make it possible to acquire new information using non-visible wavelengths and digital image processing techniques. OCR processing for automatic transcription and indexing can facilitate access to digital archives and the retrieval of information.

Within IsyReaDeT, ISTI in particular is working on a mathematical description of the typical forms of degradation that affect documents, and the study of restoration techniques exploiting multi-spectral views. The forms of degradation considered are caused by complex background textures, bleed-through, show-through and spots. From multiple views of the document (colour or multi-spectral image data), statistical approaches can be used to separate, extract and classify the different patterns. The data

model we assume is a linear mixture of different classes, each characterized by a different reflectivity spectrum. Each class is considered to have an average value in each component (channel) of the data, and these values form an unknown 'mixing' matrix. Each pixel in an image contains the contribution from the local intensities of all the classes, multiplied by the relevant mixing elements. Since the mixing matrix is unknown, we must rely on 'blind' techniques to recover the different classes. The Principal Component Analysis approach exploits the eigenvalue decomposition of the data covariance matrix to produce mutually orthogonal outputs characterized by maximum variance in each principal direction. A further step is the so-called Independent Component Analysis technique (ICA), in which, besides decorrelation, mutual independence between the output channels is required. The possibility of extracting interfering patterns from the data can be particularly useful in enhancing barely visible classes, as in the case of underwritten text in palimpsests. IsyReaDeT was thus invited to try these techniques on multispectral images of the Archimedes palimpsest.

Links:
http://www.isyreadet.net/home.htm

Please contact:
Anna Tonazzini, ISTI-CNR, Italy
E-mail: anna.tonazzini@isti.cnr.it

Managing Medical Images

by César J. Acuña, Esperanza Marcos, Valeria de Castro and J. A. Hernández

Due to the increasing number of digital images used in medical diagnosis, health centres need new and better applications with which to effectively manage such information. It is also necessary to develop an infrastructure for integrating and sharing that information between medical image repositories. Researchers from the Rey Juan Carlos University of Madrid are working on this problem.

Recently, the use of digital images for medical diagnosis has increased considerably. New and better applications are therefore needed in order to effectively manage large amounts of medical information.

DICOM is currently the standard for Digital Imaging and COmmunications in Medicine. However, despite DICOM being the most widely accepted standard, it can only handle interchange and communications among medical imaging equipment and other specific software tools called PACS (Picture Archive and Communications Systems). PACS tools are developed for several providers that use proprietary formats to store and manage medical information. Drawbacks such as these make the integration and use of DICOM difficult in a wider context such as the Web. In addition, another image storage format, known as Analyze, is widely used. This is a proprietary format part of the Analyze Software.

As a further complication, the standard for information exchange and data transportation between multiple applications is XML. For these reasons, the Database and Web Engineering Kybele Research Group and the Electronic Technology, Bioengineering and Medical Image Research Group at the Rey Juan Carlos University, together with some private and public medical centres in Madrid (Spain) are currently working on two main research topics. These are aimed at making digital image information more easily exchangeable in a wide context.

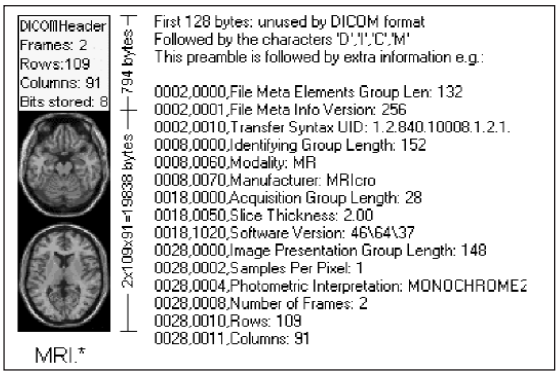


Figure 1: A DICOM file, composed of the data header and the image.

The first of these is the design and development of a Web Information System (WIS) with two main aims. The first of these is that DICOM files will be represented using XML, which will improve the interoperability and integration of medical information in a broader context like the Web. On the other hand, it will facilitate the integrated organisation, query and retrieval of DICOM and Analyze files by means of an XML database.

The second topic is the design and implementation of a Web Integration Architecture that will integrate different instances of the WIS introduced in the previous topic. The main aim of this architecture is to assist retrieval of integrated information from medical image repositories located in geographically diverse medical centres.

The WIS is currently under development. Taking as a reference point the architectures for Web applications development proposed by .NET and J2EE, the WIS architecture was structured in three layers as shown in Figure 2. These are as follows:

- *Presentation Layer:* low-cost image delivery and display are necessary in most networked hospitals, so a Web-based user interface is the most effective solution in this case. To develop the Web-based user interface we have chosen ASP.Net.