DIGITAL AUTOMATED RESTORATION OF MANUSCRIPTS AND ANTIQUE PRINTED BOOKS

Giovanni Ramponi, Filippo Stanco
DEEI – University of Trieste, Trieste, Italy
{ramponi, fstanco}@units.it

William Dello Russo, Simonetta Pelusi
Biblion Centro Studi sul Libro Antico Onlus, Venice, Italy
biblion.venezia@libero.it

Paolo Mauro
Digital Codex, Feltre, Italy
paolo.mauro@digital-codex.it

Abstract – In this paper we describe a new technology for the virtual automated restoration of antique paper affected by foxing and by yellowing deterioration process. The conventional restoration of such objects is both expensive and hard to carry out by non-specialized personnel; this new method ensures fast and cheap results, and can be used also by non-qualified operators.

I - INTRODUCTION

Libraries and other cultural institutions are digitizing increasing quantities of manuscript and printed material, both ancient and modern, but easily damaged or, like newspapers, perishable. Digital reproduction is above all designed to serve the growing community of libraries, museums and archives by providing greater access to their resources as well as helping to preserve the original materials and contents, with images of their original aspect (handwriting, typographic characters, page structure and layout). Digital reproduction has so far had two different aims: access and conservation. But we can single out another purpose: the sharing of contents and their original aspect, in a real condition of perfect readability of both textual and artistic message. Therefore, digitization is intended to create an image as close as possible to the original, ensuring a faithful reproduction of objects, with all their physical extra-textual features (including glosses and all manuscript signs, but also stains, chromatic alterations, and so on), which are part of the object's history and can be of the utmost importance from a philological point of view. Digitization permits an acquisition of maximum information vs. minimum risk. The digitally acquired image is an exact reproduction of the original object and of all included contents.

On the other hand, by digitization and subsequent computer assisted restoration we can try to improve the usability of the object, by obtaining a perfectly restored document image. This procedure would also enable us to reconstruct an excellent virtual copy - for instance, of a single extant copy of an edition, or when a copy is affected by structural or chromatic pathologies. The reader will be enabled to catch all information no more available in the current document. Two more advantages derive from a virtual-type operation of restoration: it allows the work to be read without affecting the original document traumatically or irreversibly, and it allows the greater part of the operations of restoration to be simulated, so as to supply instruments and materials that will help the "official restorer" in planning future work and appraising the final result.

In this paper we describe an innovative technology for automated restoration and imaging of old paper (XIV-XVIII century) affected by foxing and yellowing deterioration process. The restoration of such objects is both expensive and hard to carry out by non-specialized library personnel. This new method ensures fast and cheap results, and can be used also by non-qualified operators. The competitiveness of this method resides in both an economic and a technical plan.

Section 2 of this paper will deal with the technical specifications of the document acquisition process; the algorithms for the automated restoration of the foxing defects and of the yellowing of the paper will be described in Sections 3 and 4, respectively; Section 5 will show a set of experimental results; some conclusions will be drawn in the last section.

II - ACQUISITION PROCESS

For the acquisition a digital photocamera Reflex Kodak DCS PRO SLR/n has been utilized, equipped by a 13,89 MP (4560 x 3048) CMOS sensor; images have been recorded in RAW format (4500×3000 pixels); the size of a file is approximately 15 MB. The standard lenses are AF Micro-Nikkor 60 millimeter f/2.8 D. Digital images are processed and recorded on double CD-ROM or DVD, in three formats, according to the European Union guidelines [1]:

a) Uncompressed TIFF 6.0, 24-bit RGB depth, resolution at least 600 optical dpi if the original dimension is inferior or equal to A4 format; 300 dpi if the original dimension is larger. Such an image is intended for offline conservation and as emergency copy (*master*);

b) JPEG compressed to 300 optical dpi and 24-bit RGB depth;

c) JPEG compressed to 72 dpi, or less, 24-bit RGB depth and a quality factor to be defined according to an easy consultability on local and external net.

The digital camera is mounted on a stative for professional reproduction IFF Advanced Reproduction System with column IFF 1210 BL SuperRepro or Repro with a base of 60×67 cm. Beyond to normal vertical movement, such columns also permit a horizontal movement. The lighting system is composed by two Starlight or superior illuminators with continuous light: this equipment can emit a light to 5400° Kelvin, but does not generate heat because special Neon tubes are used and therefore there is no risk to damage manuscripts and old books.

III - FOXING

The term *foxing* describes the red-brown spots (color of a fox) that may form on the surface of the paper of antique books. The causes of foxing are not completely understood [2,3,4,5]; it is often triggered by high humidity and temperature extremes in places where books and documents are stored; other causes are certain ingredients found in older inks, and metals such as iron. Physical restoration based on aqueous methods is often insufficient, and the use of a laser with suitable wavelength has been proposed [6]. In our virtual restoration approach, foxing is automatically detected and removed from a digital version of a page (block scheme in Fig.1).

Foxing detection. Since the foxing damage is formed by a set of reddish-brown spots, we extract their position analyzing the C_r chrominance matrix. For this reason, the input image is first converted into the YC_bC_r color space via a conventional transformation. Foxed pixels have the highest values in C_r; we label as foxed all the pixels whose C_r component has a value greater than $Max(C_r) - s$, with s an user-selected value. The matrix *Fox* is a map where the coordinates of foxed pixels are represented as a *1* value.

Detail image extraction. First of all, we produce a smoothed image Y_{lp} from the original luminance values in Y. This image is obtained applying n times a simple *rational filter* (RF) [7] over Y. The RF attenuates small image variations while preserving edges. Different versions of this operator can be devised; in the one we have selected, $Y_{lp}(i,j)$ is obtained according to:

$$Y_{lp}(i,j) = Y(i,j) + \frac{Y(i-1,j) + Y(i+1,j) - 2Y(i,j)}{k(Y(i-1,j) - Y(i+1,j))^2 + A} + \frac{Y(i,j-1) + Y(i,j+1) - 2Y(i,j)}{k(Y(i,j-1) - Y(i,j+1))^2 + A} \quad (1)$$

A detail image Y_{hp} is hence obtained as follows:

$$Y_{hp}(i,j) = Y(i,j) - Y_{lp}(i,j) + K \qquad (2)$$

where K is a shifting parameter which is set as follows. Let t be the threshold output by Otsu's method [8]. The values of Y that do not belong to foxing areas ($Fox(i,j) = 0$) and are not part of the text ($Y(i,j) > t$) are stored in the set M; then,

$$K = mean[M] = mean[\{Y(i,j) \mid Fox(i,j) = 0 \wedge Y(i,j) > t\}] \qquad (3)$$

The detail image Y_{hp} will be used in the *merging* step.

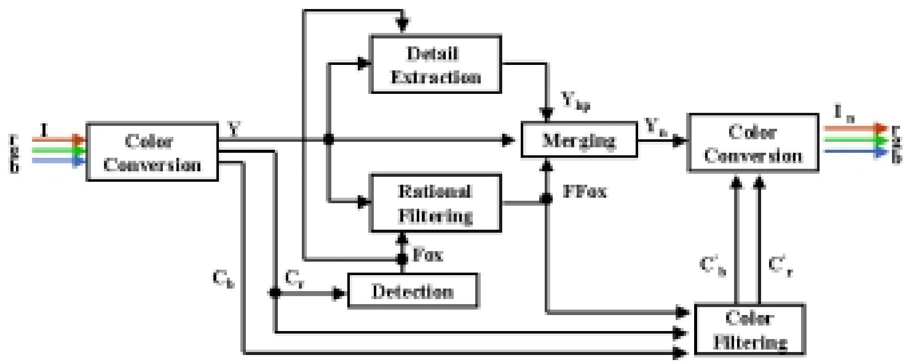


Figure 1: Block scheme of the de-foxing algorithm.

Processing of foxed areas. The binary image *Fox* is the map of foxed areas. As expressed above it has value *1* for the pixels that belong to the foxing, and *0* otherwise. In this step we create a new map that is no longer a binary image but shows smooth transitions between *1* and *0*. The slope of the transition is an direct function of the slope of the edges in the original image. To build this new map *FFox* we use the above RF applied over *Fox* using the image *Y* as an edge sensor. More precisely, we use *n* times Eq. 1 over *Fox* where in the denominator *Fox* is substituted by *Y*.

Merging. In this step we create a new luminance image where the foxing stains are restored. It is formed by merging the shifted highpass image in the foxing areas, and the original luminance value in the remaining parts of the image. This is achieved via the linear combination:

$$Y_n(i,j) = Y_{hp}(i,j) FFox(i,j) + Y(i,j)(1 - FFox(i,j)) \qquad (4)$$

Color filtering. As expressed above, the most important characteristic of foxing is its typical color. This means that also the chrominance matrices need to be treated. Our algorithm corrects the foxed areas in the matrices C_r and C_b using the following rule:

$$C_x^{'}(i,j) = \begin{cases} median_x & if \quad FFox(i,j) > 0 \\ C_x(i,j) & elsewhere \end{cases} \qquad (5)$$

where $median_x = median\{C_x(i,j) \mid FFox(i,j) = 0\}$ with $x = \{b, r\}$. The obtained matrices $C_b^{'}$, and $C_r^{'}$ are used to determine the *RGB* components of the final image I_f.

IV - PAGE ENHANCEMENT

We describe in this section a procedure for reducing the yellowing of the paper (Fig. 3a) due to cellulose oxidation catalyzed by metals [6]. The method we propose makes the background aspect of the paper more homogeneous, reducing the humidity-induced local alterations; at the same time, parts of the characters which are discolored are restored. Our algorithm divides the luminance histogram in two parts: text and background. Then, the background is adjusted in color. For clarity, a block scheme of the procedure is reported in Fig. 2.

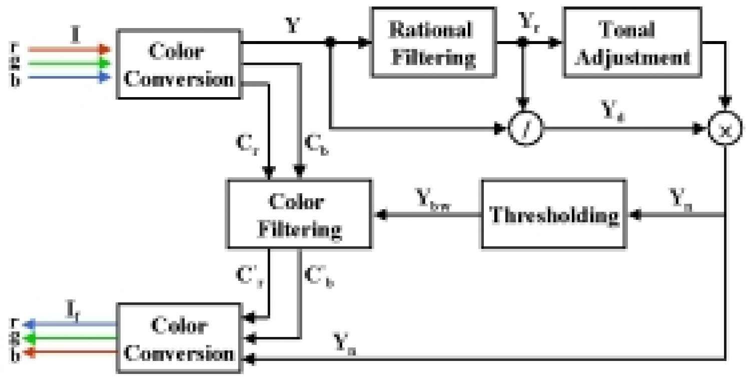


Figure 2: Block scheme of the paper enhancement algorithm.

Like in the previous section, the input image is converted to the YC_bC_r color space. The algorithm operates first on the luminance image Y, and it modifies its values to better reproduce the structure of the original image. Then, using the structure learned in Y, it modifies the chrominance matrices C_b and C_r.

Tonal Adjustment. The luminance component of the input image is processed by an RF analogous to the one used for the de-foxing of the image; an image Y_r results, in which the largest edges are well-marked and the values inside a region are made homogeneous. We want to modify the values of Y_r in order to enhance the contrast between the text and the background. To this purpose we adjust the tonal range of the image using a nonlinear mapping having a sigmoidal shape; a proper selection of the slope of its central portion yields the required effect [9]. Details in the page are lost in Y_r but are preserved in our procedure by evaluating the pixel-by-pixel ratio between the original and filtered images:

$$Y_d(i,j) = Y(i,j) / Y_r(i,j) \qquad (6)$$

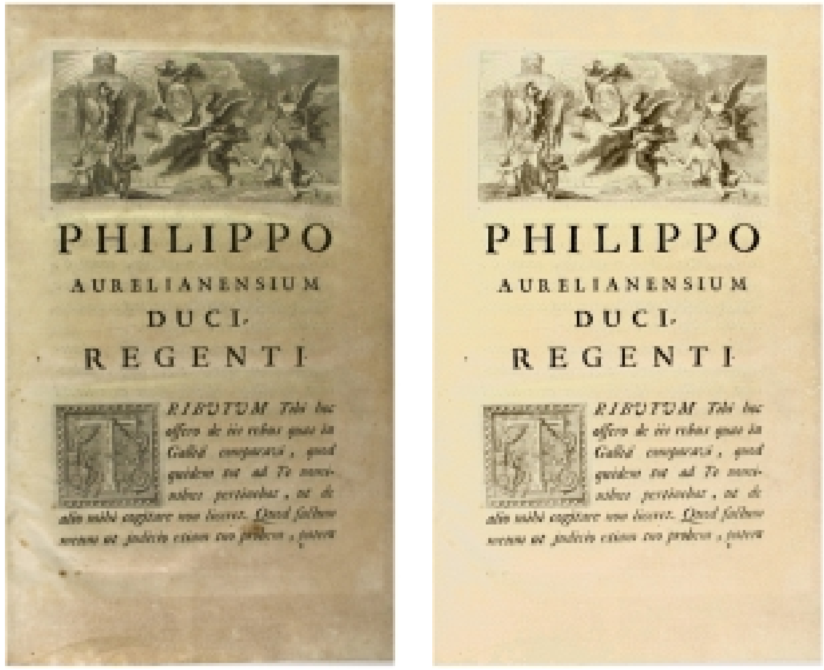
The image Y_d is multiplied back at the end of the tonal adjustment process, as indicated in Fig. 2, to obtain Y_n and the result is used to adjust the chrominance matrices. Y_n is also used in the final *RGB* color space conversion rather than the original Y.

Thresholding. In the image Y_n the algorithm automatically searches for a threshold that divides the luminance histogram in two separate classes. We use again Otsu's method of threshold selection. Since the histogram of Y_n shows two separate peaks depending on the tones correction described above, the threshold t selected by Otsu's method is effective. The image Y_{bw} is the binary version of Y_n; its pixels take a value of one only in the positions in which $Y_n(i,j) \geq t$.

Color correction. Basing on the information derived from the luminance, we then correct the chrominance matrices. We modify only the chrominance values of the pixels that belong to the background, and leave unchanged all the remaining pixels. The binary image Y_{bw} is helpful for this operation. If we let $M_x = \{C_x(i,j) \mid Y_{bw}(i,j) = 1\}$, with $x = \{r, b\}$, the new chrominance matrices are defined as follows:

$$C'_x(i,j) = \begin{cases} median(M_x) & if \quad Y_{bw}(i,j) = 1 \\ C_x(i,j) & elsewhere \end{cases} \qquad (7)$$

The final image I_f is obtained converting the new Y_n, C'_b, and C'_r matrices into *RGB* components.



(a) **(b)**

Figure 3*: (a) An example of background with a variation of colors; (b) The restored image.*

V - EXPERIMENTAL RESULTS

Both algorithms proposed in this paper work without user intervention. Both use the same RF; in our experiments the number of iterations *n* is 10; typical values of the parameters, for 8-bit images, are *A=5* and *k=0.03*. The foxing restoration algorithm described in Section 2 needs only a parameter *s* that represents the number of color levels that will be considered as belonging to foxing. We set *s=6*. For the page enhancement algorithm in Section 3, it is necessary to choose the equation for the curve used during the tonal adjustment. This curve is described by a 5-th degree polynomial with null first derivative in zero and in one, and null second and third derivative in one: $y(x) = -4x^5 + 15x^4 - 20x^3 + 10x^2$.

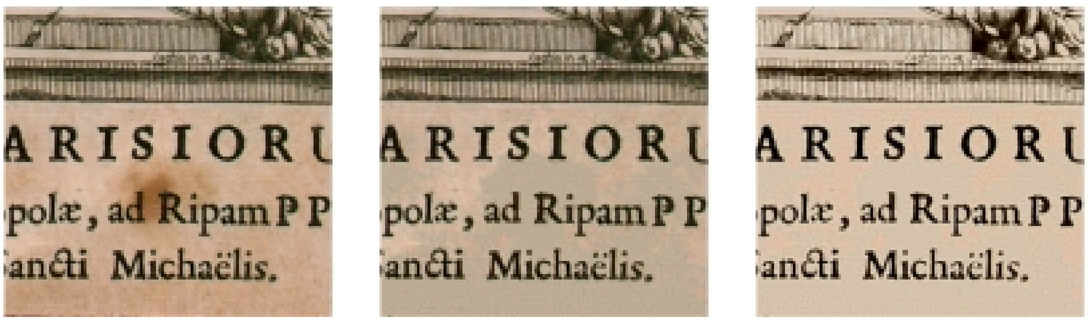
The images used in our experiments are real damaged documents, and the equivalent incorrupt images do not exist. Therefore, the goodness of the results can only be estimated qualitatively examining the results. Figs. 3b and 4 report restoration results. In Fig. 4b the characteristic foxing color present in Fig. 4a is eliminated, even if the background color is not perfectly restored. To improve the visual quality, the paper enhancement algorithm proposed above is used (Fig. 4c). Similar considerations can be done for the image reported in Fig. 3b.

We stress that the original background color is unknown. In order to avoid introducing distortions in the chromatic content, we choose to assign the damaged pixels a color extracted from the original range. As expressed in the previous sections, this choice is performed over the chrominance matrices C_b and C_r instead of the RGB color space; hence the chromatic aberrations are reduced.

VI - CONCLUSIONS

Two algorithms have been introduced in this paper for the de-foxing and the restoration of the pages of antique manuscripts and printed books. Some preliminary experiments show the efficacy of the proposed methods: after processing, the visual quality of the acquired image has improved, but the typical characteristics of the document are still preserved.

The above described research project is of interest in two application areas. The first one concerns a new use of digital retrieval techniques for images and texts consultation. Thanks to digital imaging, direct visualisation of a restored document as a study object on a computer screen is possible. The second one could be seen as an enlargement of research and development concerning the implementation of the "electronic library", mostly in the field of the necessity of contrasting a well known digital information obsolescence risk, as a complement of the procedures of old damaged books image files migration, and-or conversion [10]. Potential users of this method therefore are philologists, librarians, archivists and, in general, scholars in the field of old books and paper documents, and institutions involved in the creation and preservation of digital book images.



(a) (b) (c)

Figure 4: (a) Detail of an image with foxing; (b) Image in (a) after foxing restoration; (c) Image in (b) after background enhancement.

ACKNOWLEDGEMENTS

The original documents used in the experiments belong to the Redentoristi Library at S. Maria della Consolazione in Venice, Italy.

This work has been partially supported by a grant of the Regione Friuli Venezia Giulia, and is conducted in the framework of the SCHEMA NoE (IST-2001-32795).

References

[1] Minerva Consortium, *Good Practice Handbook. Version 1.2*, edited by the Minerva Working Group 6: Identification of good practices and competence centres, November 2003.
[2] www.pbs.org/wgbh/pages/roadshow/speak/foxing.html
[3] palimpsest.stanford.edu/don/dt/dt1434.html
[4] www.lowyonline.com/spr01_foxing.html
[5] www.octavo.com/collections/essays/foxing.html
[6] Z. Kollia, E. Sarantopoulou, A.C. Cefalas, S. Kobe, and Z. Samardzija, "Nanometric size control and treatment of historic paper manuscript and prints with laser light at 157 nm". *Appl. Phys., A, Mater. Sci. Process. Vol.* 79, pp. 379-382, 2004.
[7] G. Ramponi, "The Rational Filter for Image Smoothing", *IEEE Signal Processing Letters*, vol. 3, no. 3, pp. 63-65, March 1996.
[8] N. Otsu "A Threshold selection method from gray-scale histogram", *IEEE Trans. Syst., Man, Cyber.*,SMC-8, pp.62-66, 1978.
[9] F. Stanco, G. Ramponi and L. Tenze, "A Method for Improving the Visual Quality of Digitized Antique Books", *Proc. 7th COST 276 Workshop*, Ankara (Turkey), 4-5 Nov. 2004.
[10] G. W. Lawrence, et al., *Risk Management of Digital Information: A File Format Investigation*, Council on Library and Information Resources, Washington, D.C., June 2000.