Nikola Ikonomov and **Milena Dobreva**
(Bulgarian Academy of Sciences)

THE MAKING OF … DIGITAL BOOKS

Abstract. The paper can be considered as a practical guide in the process of making digital books from printed editions, especially in regard to the digital imaging part. It summarizes the experience gathered at the Digital Humanities Department of the Institute for Mathematics and Informatics in Sofia. The proposed workflow is put together and tested at the Digitization Centre using a professional Book Scanner Omniscan 5000 TT and standard computer workstations and software.

Keywords: digitization, scanning, image processing, image enhancement, OCR

1. Introduction

In the Information Age, knowledge resources have become critical to almost every aspect of human life. But as the volume of information exponentially increases, so does the amount of efforts required to collect, organize, use, publish, and preserve that information.

One aspect of these efforts is the process of making digital books from printed editions. It is a common view that the related technology is straightforward enough and does not require special attention. Scanning a book and compiling a PDF derivative from the images seems a simple and easy surmountable task.

However, things are quite different when speaking in the context of digital libraries (DL) with millions of pages, which should meet strict quality requirements, be optically recognized and published electronically. Geometrical distortions, light and contrast variations, page alignment discrepancies, insufficient resolution for OCR or inadequate file sizes for publishing are only part of the problem. Standard image editing software like Adobe Photoshop or ACDSee can solve most of above mentioned problems, but being not specially conceived for work with digital books turn the process into a time consuming and complicated task.

The making of digital books requires an efficient and sophisticated system consisting of hardware, software and workflow management processes that could fully benefit of the unique potential of the DL [1]. The practice of the Digitization Centre at IMI with various projects has shown undoubtedly the need for such a system that manages the process of metadata creation, scanning, image processing, quality control, and creation of output for web publishing and is flexible enough to simultaneously manage projects with different materials (books, journals, newspapers, manuscripts, maps, photographs, unbound archive materials) and purposes (preservation, web access, print-on-demand). The system should also faultlessly supply digital content to the libraries' repository in order to guarantee its preservation for years to come.

The workflow of the digitization process for the DL is illustrated in Figure 1. As shown it encompasses both automated and manual processes, and is modular to allow for customization according to the different project needs, and integration of new image processing technologies as they develop. Our further detailed considerations deal with that part of the whole workflow, which is related to the mere process of making digital books (blocks marked in red).

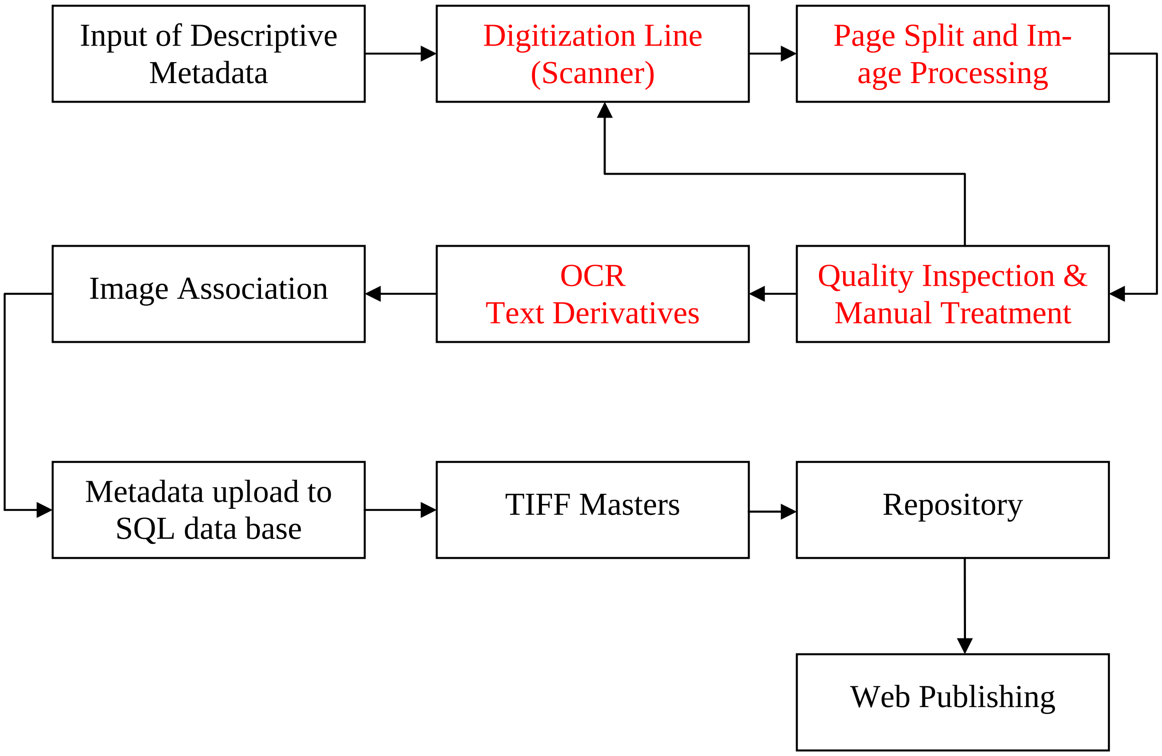

Input of Descriptive Metadata
Digitization Line (Scanner)
Page Split and Image Processing
Image Association
OCR Text Derivatives
Quality Inspection & Manual Treatment
Metadata upload to SQL data base
TIFF Masters
Repository
Web Publishing

Fig.1. Workflow of the digitization process for the DL

The process of making digital books can be divided into three basic steps [2]:

- Scanning the printed book and storing of masters for digital preservation purposes;
- Processing of all acquired images;
- Creation of the output file which will be used for online access (PDF, DjVu, plain text, multi-page TIFF, etc).

2. Scanning the book

We will not discuss the type of scanner that should be used - conventional flat bed or planetary camera. We would only mention briefly some severe limitations of flat bed scanners:

- The book must be placed face down on the glass, picked up, the page turned, and placed down again. That repeated operation is not good especially for old and valuable books. In addition, the book cannot be put exactly on the same place used for the previous scanning, thus resulting in different top-, bottom-, left- and right margins.
- The book should be pressed down on its spine to flatten the pages, thus damaging it. Despite pressing down, the page still "curves down" at the centre of the binding, distorting the text which is near the bound edge and producing a dark shadow down the centre of the book.
- A normal flat bed scanner can't cope with anything larger than A3, while many books, drawings, geographical maps, and documents are larger than that.

In contrast to flat bed scanner by planetary cameras the book sits on the desktop or on an adjustable cradle in its normal open position. It doesn't need to be repeatedly handled, except for turning over the pages. Some book scanners are equipped with a glass plate that presses flat the scanning surface.

Concerning the scanning parameters here are some basic recommendations:

- Use always TIFF as an output file format of the master image. JPEG files produce undesired blurring effects in the output text.
- Do not scan in “Black and White” attracted by the acceptable size of the output files. Such images are not prone to more of the post processing techniques, thus producing outcomes with irreversibly deteriorated quality parameters. If forced to do so by some reasons, use resolution of 600 dpi.
- Use instead “Gray Scale” (possibly 8 bits) for text, tables, drawings, or “Colour” (possibly 32 bits) if the page contains illustrations. The resolution should be at least 400 dpi.
- Store the “raw scans” on DVD or other digital media to avoid accidental data loss.

3. Image processing

In our practical work we are using three dedicated programs for batch processing of scanned images, and namely:

- Book Restorer[1]
- Scan Cromsator (free)[2]
- RasterID[3]

All three products have similar parameters and technical specifications. We will not enter details concerning the special features of each program or try to compare them, but will rather focus on the possibilities they offer to facilitate the making of digital books.

The image processing is a complicated and time consuming task especially when using standard software (Adobe Photoshop, ACDSee, etc.). Dedicated facilitate greatly the process, but are not a heal-all remedy. Depending on the quality of the scanned book (color of the paper, contrast of the text, quality of the printing), the output images differ from each other. Applying default or automatic settings instead of manual tuning of the various program parameters gives satisfying but rather not best results. Nevertheless, we will limit our considerations to the case of automatic program settings in batch processing mode, which is a reasonable compromise between quality, complexity, and required time.

The image processing stage includes the following basic procedures [3]:

3.1 Page splitting, cropping and margin equalization . Usually, scanned images contain both left-hand and right-hand pages and should be split into two separate page images. Furthermore, scans have different margins in all directions - left, right, top, and bottom. Scrolling through a digital book compiled from such scans on a computer screen accentuates notably their unevenness and divergence.

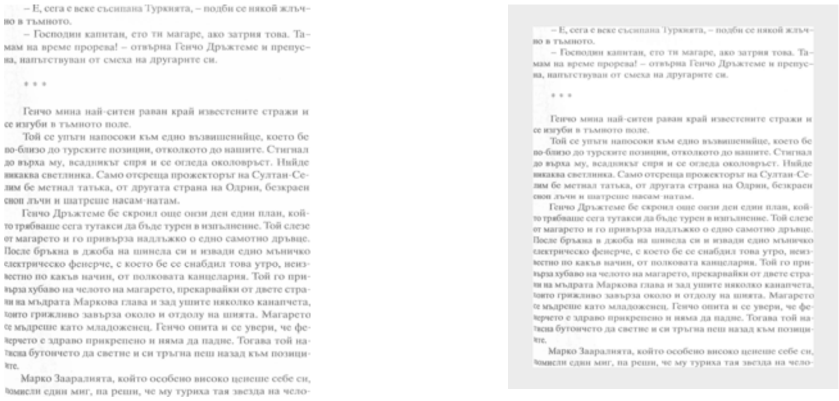
Cropping algorithms are used to separate double pages into two independent files and/or to eliminate margins. A completely automated process crops any black space or other “noise” around the fore-edge of the book, and splits the image into a clean left and right page. While effective, this process is not perfect, and operators may need to perform some additional cropping at a later stage.

The margin equalization (Fig. 2) includes automatic detection of page borders, removal of all parts located outside this zone and adding security margins, defined by the user. In some cases, especially when the text occupies only a pert of a particular page, cropping should be done manually in order to avoid unwanted changes in the image size.

[1] http://www.i2s-bookscanner.com/produits.asp?gamme=1011&sX_Menu_selectedID=left_1011_GEN

[2] http://www.djvu-soft.narod.ru/kromsator/

[3] http://www.csoft.com/index.cfm?menuid=105



– Е, сега е веке съсипана Туркията, – подби се някой жлъчно в тъмното.

– Господин капитан, ето ти магаре, ако затрия това. Тамам на време прорева! – отвърна Генчо Дръжтеме и препусна, напътствуван от смеха на другарите си.

* * *

Генчо мина най-ситен раван край известените стражи и се изгуби в тъмното поле.

Той се упъти напосоки към едно възвишенийце, което бе по-близо до турските позиции, отколкото до нашите. Стигнал до върха му, всадникът спря и се огледа околовръст. Никъде никаква светлинка. Само отсреща прожекторът на Султан-Селим бе метнал татька, от другата страна на Одрин, безкраен сноп лъчи и шатреше насам-натам.

Генчо Дръжтеме бе скроил още онзи ден един план, който трябваше сега тутакси да бъде турен в изпълнение. Той слезе от магарето и го привърза надлъжко о едно самотно дръвце. После бръкна в джоба на шинела си и извади едно мъничко електрическо фенерче, с което бе се снабдил това утро, неизвестно по какъв начин, от полковата канцелария. Той го привърза хубаво на челото на магарето, прекарвайки от двете страни на мъдрата Маркова глава и зад ушите няколко канапчета, които грижливо завърза около и отдолу на шията. Магарето се мъдреше като младоженец. Генчо опита и се увери, че фенерчето е здраво прикрепено и няма да падне. Тогава той натисна бутончето да светне и си тръгна пеш назад към позициите.

Марко Заралията, който особено високо ценеше себе си, помисли един миг, па реши, че му туриха тая звезда на чело-

– Е, сега е веке съсипана Туркията, – подби се някой жлъчно в тъмното.

– Господин капитан, ето ти магаре, ако затрия това. Тамам на време прорева! – отвърна Генчо Дръжтеме и препусна, напътствуван от смеха на другарите си.

* * *

Генчо мина най-ситен раван край известените стражи и се изгуби в тъмното поле.

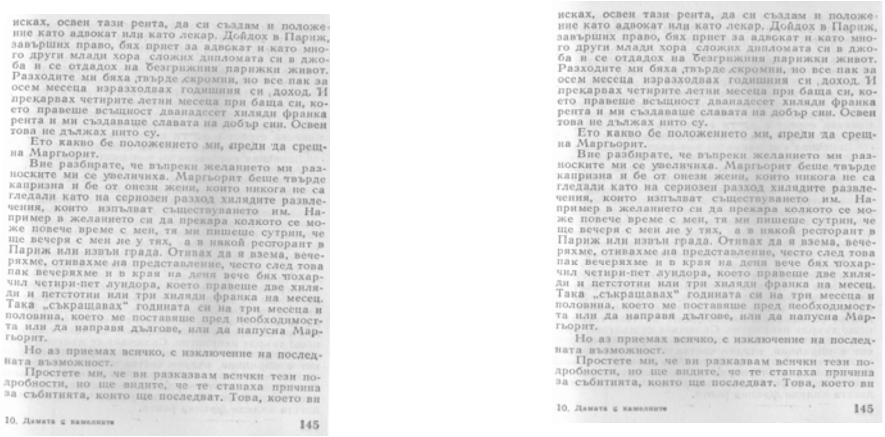
Той се упъти напосоки към едно възвишенийце, което бе по-близо до турските позиции, отколкото до нашите. Стигнал до върха му, всадникът спря и се огледа околовръст. Никъде никаква светлинка. Само отсреща прожекторът на Султан-Селим бе метнал татька, от другата страна на Одрин, безкраен сноп лъчи и шатреше насам-натам.

Генчо Дръжтеме бе скроил още онзи ден един план, който трябваше сега тутакси да бъде турен в изпълнение. Той слезе от магарето и го привърза надлъжко о едно самотно дръвце. После бръкна в джоба на шинела си и извади едно мъничко електрическо фенерче, с което бе се снабдил това утро, неизвестно по какъв начин, от полковата канцелария. Той го привърза хубаво на челото на магарето, прекарвайки от двете страни на мъдрата Маркова глава и зад ушите няколко канапчета, които грижливо завърза около и отдолу на шията. Магарето се мъдреше като младоженец. Генчо опита и се увери, че фенерчето е здраво прикрепено и няма да падне. Тогава той натисна бутончето да светне и си тръгна пеш назад към позициите.

Марко Заралията, който особено високо ценеше себе си, помисли един миг, па реши, че му туриха тая звезда на чело-

Fig.2. Margin equalization

3.2 Deskew. Deskewing is the process of removing skew from images. Skew is an artefact that can occur in scans because of the camera being misaligned, imperfections in the scanning or surface, or simply because the paper was not placed completely flat when scanned. This technique, also called auto straighten, is the automatic rotation of an image such that the text is vertically aligned. The rotation can be done in reference to the left or right borders of the page or to the whole text block. In manual mode the rotation angle and the direction are set arbitrarily (Fig. 3).



исках, освен тази рента, да си създам и положение като адвокат или като лекар. Дойдох в Париж, завърших право, бях приет за адвокат и като много други млади хора сложих дипломата си в джоба и се отдадох на безгрижния парижки живот. Разходите ми бяха твърде скромни, но все пак за осем месеца изразходвах годишния си доход. И прекарвах четирите летни месеца при баща си, което правеше всъщност дванадесет хиляди франка рента и ми създаваше славата на добър син. Освен това не дължах нито су.

Ето какво бе положението ми, преди да срещна Маргьорит.

Вие разбирате, че въпреки желанието ми разноските ми се увеличиха. Маргьорит беше твърде капризна и бе от онези жени, които никога не са гледали като на сериозен разход хилядите развлечения, които изпълват съществуването им. Например в желанието си да прекара колкото се може повече време с мен, тя ми пишеше сутрин, че ще вечеря с мен не у тях, а в някой ресторант в Париж или извън града. Отивах да я взема, вечеряхме, отивахме на представление, често след това пак вечеряхме и в края на деня вече бях похарчил четири-пет луидора, което правеше две хиляди и петстотин или три хиляди франка на месец. Така „съкращавах" годината си на три месеца и половина, което ме поставяше пред необходимостта или да направя дългове, или да напусна Маргьорит.

Но аз приемах всичко, с изключение на последната възможност.

Простете ми, че ви разказвам всички тези подробности, но ще видите, че те станаха причина за събитията, които ще последват. Това, което ви

10. Дамата с камелиите 145

исках, освен тази рента, да си създам и положение като адвокат или като лекар. Дойдох в Париж, завърших право, бях приет за адвокат и като много други млади хора сложих дипломата си в джоба и се отдадох на безгрижния парижки живот. Разходите ми бяха твърде скромни, но все пак за осем месеца изразходвах годишния си доход. И прекарвах четирите летни месеца при баща си, което правеше всъщност дванадесет хиляди франка рента и ми създаваше славата на добър син. Освен това не дължах нито су.

Ето какво бе положението ми, преди да срещна Маргьорит.

Вие разбирате, че въпреки желанието ми разноските ми се увеличиха. Маргьорит беше твърде капризна и бе от онези жени, които никога не са гледали като на сериозен разход хилядите развлечения, които изпълват съществуването им. Например в желанието си да прекара колкото се може повече време с мен, тя ми пишеше сутрин, че ще вечеря с мен не у тях, а в някой ресторант в Париж или извън града. Отивах да я взема, вечеряхме, отивахме на представление, често след това пак вечеряхме и в края на деня вече бях похарчил четири-пет луидора, което правеше две хиляди и петстотин или три хиляди франка на месец. Така „съкращавах" годината си на три месеца и половина, което ме поставяше пред необходимостта или да направя дългове, или да напусна Маргьорит.

Но аз приемах всичко, с изключение на последната възможност.

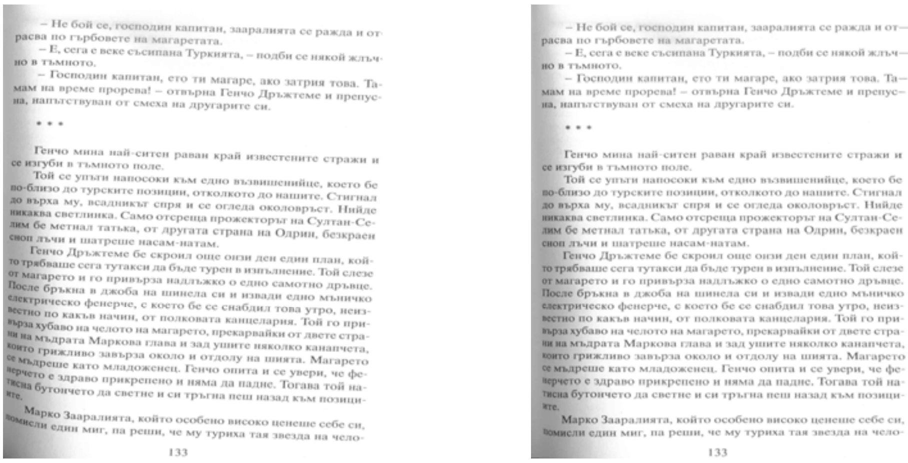
Простете ми, че ви разказвам всички тези подробности, но ще видите, че те станаха причина за събитията, които ще последват. Това, което ви

10. Дамата с камелиите 145

Fig.3. Deskew

3.3 Geometric correction . A geometric correction is used to eliminate flaws due to the curve of books. In digitization through camera-based systems, simple imaging setups often produce geometric distortions in the resultant 2D images because of the non-planar geometric shapes of certain documents such as thick bound books, rolled, folded or crumpled materials, etc. Arbitrarily warped documents can be successfully restored by flattening a 3D scan of the document.

The geometrical correction also makes it possible to correct the effect of crushing of the characters and to preserve original distances. This procedure is also carried out thanks to processes making it possible to reconstitute the 3D shape of the book (Fig.4).



– Не бой се, господин капитан, заaралията се ражда и отрасва по гърбовете на магаретата.

– Е, сега е веке съсипана Туркията, – подби се някой жлъчно в тъмното.

– Господин капитан, ето ти магаре, ако затрия това. Тамам на време прорева! – отвърна Генчо Дръжтеме и препусна, напътствуван от смеха на другарите си.

* * *

Генчо мина най-ситен раван край известените стражи и се изгуби в тъмното поле.

Той се упъти напосоки към едно възвишенийце, което бе по-близо до турските позиции, отколкото до нашите. Стигнал до върха му, всадникът спря и се огледа околовръст. Нийде никаква светлинка. Само отсреща прожекторът на Султан-Селим бе метнал татька, от другата страна на Одрин, безкраен сноп лъчи и шатреше насам-натам.

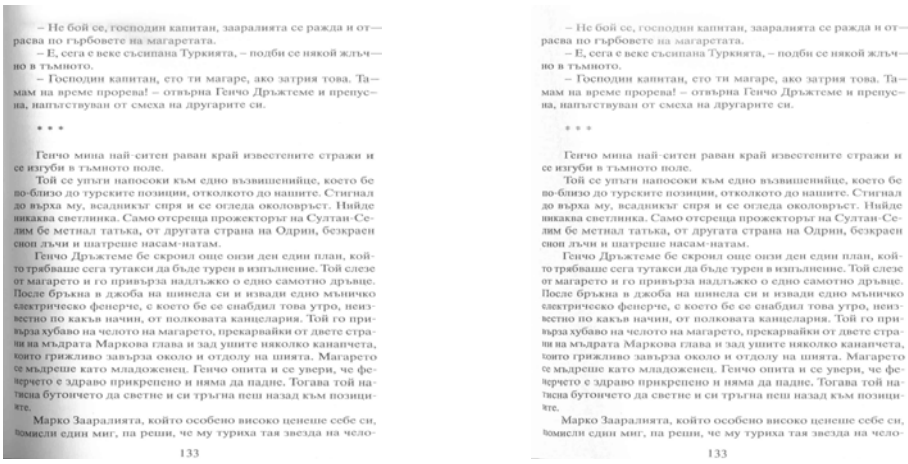
Генчо Дръжтеме бе скроил още онзи ден един план, който трябваше сега тутакси да бъде турен в изпълнение. Той слезе от магарето и го привърза надлъжко о едно самотно дръвце. После бръкна в джоба на шинела си и извади едно мъничко електрическо фенерче, с което бе се снабдил това утро, неизвестно по какъв начин, от полковата канцелария. Той го привърза хубаво на челото на магарето, прекарвайки от двете страни на мъдрата Маркова глава и зад ушите няколко канапчета, които грижливо завърза около и отдолу на шията. Магарето се мъдреше като младоженец. Генчо опита и се увери, че фенерчето е здраво прикрепено и няма да падне. Тогава той натисна бутончето да светне и си тръгна пеш назад към позициите.

Марко Заaралията, който особено високо ценеше себе си, помисли един миг, па реши, че му туриха тая звезда на чело-

133

Fig.4. Geometric correction

3.4 Lighting correction. Lighting correction corrects light variation produced by surface relief or book curvature. When a thick book is scanned, the shadow of the binding will appear on the image. This technique allows obtaining light uniformity and eliminates such shadows, whether vertical or horizontal (Fig. 5).



– Не бой се, господин капитан, заaралията се ражда и отрасва по гърбовете на магаретата.

– Е, сега е веке съсипана Туркията, – подби се някой жлъчно в тъмното.

– Господин капитан, ето ти магаре, ако затрия това. Тамам на време прорева! – отвърна Генчо Дръжтеме и препусна, напътствуван от смеха на другарите си.

* * *

Генчо мина най-ситен раван край известените стражи и се изгуби в тъмното поле.

Той се упъти напосоки към едно възвишенийце, което бе по-близо до турските позиции, отколкото до нашите. Стигнал до върха му, всадникът спря и се огледа околовръст. Нийде никаква светлинка. Само отсреща прожекторът на Султан-Селим бе метнал татька, от другата страна на Одрин, безкраен сноп лъчи и шатреше насам-натам.

Генчо Дръжтеме бе скроил още онзи ден един план, който трябваше сега тутакси да бъде турен в изпълнение. Той слезе от магарето и го привърза надлъжко о едно самотно дръвце. После бръкна в джоба на шинела си и извади едно мъничко електрическо фенерче, с което бе се снабдил това утро, неизвестно по какъв начин, от полковата канцелария. Той го привърза хубаво на челото на магарето, прекарвайки от двете страни на мъдрата Маркова глава и зад ушите няколко канапчета, които грижливо завърза около и отдолу на шията. Магарето се мъдреше като младоженец. Генчо опита и се увери, че фенерчето е здраво прикрепено и няма да падне. Тогава той натисна бутончето да светне и си тръгна пеш назад към позициите.

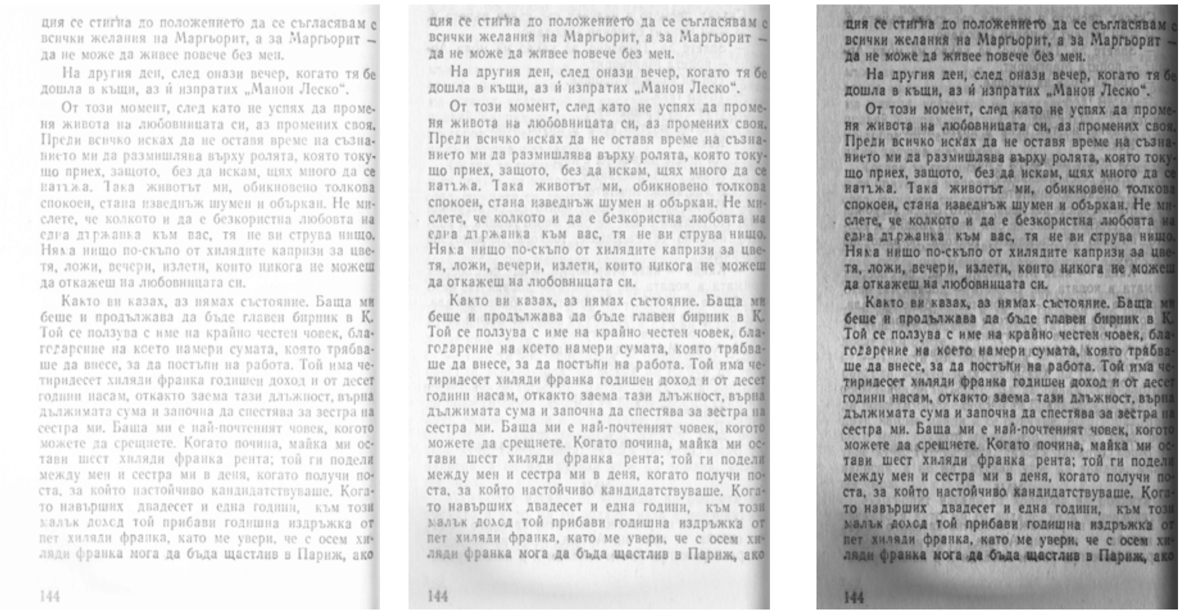
Марко Заaралията, който особено високо ценеше себе си, помисли един миг, па реши, че му туриха тая звезда на чело-

133

Fig.5.Lighting corrections

3.5. Brightness and contrast adjustments. These adjustments can be done either automatically using default values, or manually by means of standard adjustment sliders in preview mode. Most dedicated programs offer also another possibility to tweak the brightness and contrast of the images, thus enhancing their quality and readability. For that purpose histogram adjustment techniques are used, which redistribute the colours over all the possible values. The histogram window displays the distribution of intensities in the image's luminosity. It runs from pure black on the left to pure white on the right. By looking at the graph, one can determine how to adjust the image in order to optimize its contrast and bring out more detail without loss of important information.

A well-exposed image has just the right amount of light in the scene to properly illuminate the subject, so that images are neither too dark nor too light. A good picture has a good exposure "spread" over the total range of pixels in the image (Fig.6).



ция се стигна до положението да се съгласявам с всички желания на Маргьорит, а за Маргьорит — да не може да живее повече без мен.

На другия ден, след онази вечер, когато тя бе дошла в къщи, аз й изпратих „Манон Леско“.

От този момент, след като не успях да променя живота на любовницата си, аз промених своя. Преди всичко исках да не оставя време на съзнанието ми да размишлява върху ролята, която току-що приех, защото, без да искам, щях много да се натъжа. Така животът ми, обикновено толкова спокоен, стана изведнъж шумен и объркан. Не мислете, че колкото и да е безкористна любовта на една държанка към вас, тя не ви струва нищо. Няма нищо по-скъпо от хилядите капризи за цветя, ложи, вечери, излети, които никога не можеш да откажеш на любовницата си.

Както ви казах, аз нямах състояние. Баща ми беше и продължава да бъде главен бирник в К. Той се ползува с име на крайно честен човек, благодарение на което намери сумата, която трябваше да внесе, за да постъпи на работа. Той има четиридесет хиляди франка годишен доход и от десет години насам, откакто заема тази длъжност, върна дължимата сума и започна да спестява за зестра на сестра ми. Баща ми е най-почтеният човек, когото можете да срещнете. Когато почина, майка ми остави шест хиляди франка рента; той ги подели между мен и сестра ми в деня, когато получи поста, за който настойчиво кандидатстваше. Когато навърших двадесет и една години, към този малък доход той прибави годишна издръжка от пет хиляди франка, като ме увери, че с осем хиляди франка мога да бъда щастлив в Париж, ако

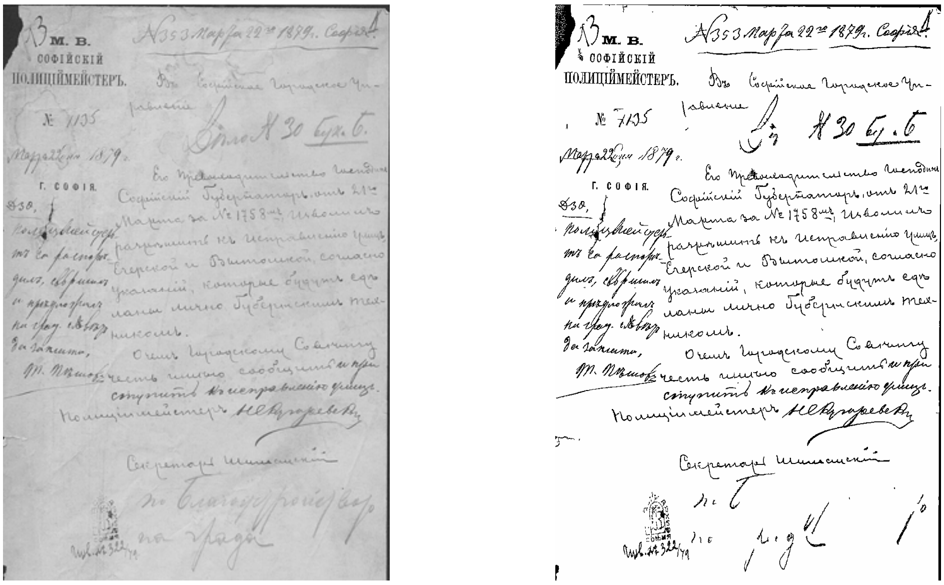
144

Overexposed Normal Underexposed

Fig. 6.Brightness and contrast adjustment

3.6 Filters. All dedicated programs are provided with different filters, which can radically improve the quality of the images. Usually these are standard filters such as blur, sharpen, average, thickening, thinning. Filters are applied on images depending on the readability of the scanned text. Missing parts of some letters are recovered, for example, by using a thickening filter.

3.7 Image Binarization. Image binarization converts an image of up to 256 gray levels to a black and white image (Fig. 7). Frequently, binarization is used as a preliminary step before OCR. In fact, most OCR programs, available on the market work only on black & white images.



13 М. В.
СОФІЙСКІЙ
ПОЛИЦІЙМЕЙСТЕРЪ.
№ 7135
Марта 22 дня 1879 г.
Г. СОФІЯ.

№ 353 Марта 22-го 1879 г. София

Въ Софійское Городское Управленіе

№ 30 бух. б

Его Превосходительство Господинъ Софійскій Губернаторъ, отъ 21-го Марта за № 1758-мъ, изволилъ разрѣшить къ исправленію улицъ Егерской и Витошкой, согласно указаній, которые будутъ сдѣланы лично Губернскимъ Техникомъ.

О чемъ Городскому Совѣту честь имѣю сообщить и приступить къ исправленію улицъ.

Полиціймейстеръ [illegible]

Секретарь [illegible]

№ 30, Полицмейстерът се распорежда, [illegible] и предложи на град. [illegible] да [illegible], Т. [illegible]

по благоустройству на града

Инв. № 322/79

Fig. 7. Image Binarization

The simplest way to use image binarization is to choose a threshold value, and classify all pixels with values above this threshold as white, and all other pixels as black. The problem then is how to select the correct threshold. In many cases, finding one threshold compatible to the entire image is very difficult, and in many cases even impossible. Therefore, adaptive image binarization is needed where an optimal threshold is chosen for each image area.

3.8 Despeckle. Despeckle corrects the binarization effects or scanning interference. The Despeckle filter removes noise from images without blurring edges. It attempts to detect complex areas and leave these intact while smoothing areas where noise will be noticeable. The Despeckle filter smoothes areas in which noise is noticeable while leaving complex areas untouched. The effect is that grain or other noise is reduced without severely affecting edges.

The standard deviation of each pixel and its neighbours is calculated to determine if the area is one of high complexity or low complexity. If the complexity is lower than the threshold, then the area is smoothed using a simple mean filter.

3.9 Image resizing. Image resizing modifies the resolution, size or proportions of an image. Typically this step is used to obtain uniform images with acceptable sizes depending on the required parameters of the particular task.

3.10. Quality control. Once the images are processed, they are further transferred to a quality control workstation where operators perform visual inspection of every page and check their conformity to specified quality standards [4]. If missing pages are detected, or the quality specifications are not met, a re-scan procedure is scheduled. The quality control workstations are equipped with additional image editing software packages to allow operators to perform more sophisticated manual operations or to insert items, obtained by another scanner (for example colour illustrations).

4. Creation of the output file

Once the quality control operators confirm that scanned images meet specified quality standards, there are two possible options to make a book:

- By creating an output file (PDF, DjVu) from images.
- By converting images to text using OCR.

4.1. Creation of files from images. This is the fastest and the simplest way to make a digital book, which doesn’t require further processing steps. The only but serious problem, associated with this option, is the huge size of the output file, which makes it many cases unusable for wide electronic publishing. Nevertheless, the creation of files from images is the only possible option when digitizing manuscripts, rare books, handwritten documents, etc., where character recognition techniques are not applicable.

The unfavorable effect of the file size can be partially compensated by converting the TIFF images into some compressed formats (JPEG, DjVu), thus reducing considerably their size by retaining a satisfactory quality level.

Particularly effective can be the use of the DjVu format which typically achieves compression ratios about 5 to 10 times better than existing methods such as JPEG and GIF. This makes the size of high-quality scanned pages comparable to an average HTML page. DjVu image viewing software never decompresses the entire image, but instead keeps the image in memory in a compact form, and decompresses the piece displayed on the screen in real time as the user opens the image. Thus, images as large as 2,500 pixels by 3,300 pixels (a standard page image at 300 dpi) can be downloaded and displayed on very low-end PCs. One of the main technologies behind DjVu is the ability to separate an image into a background layer (i.e., paper texture and pictures) and foreground layer (text and line drawings). By separating the text from the backgrounds, DjVu can keep the text at high resolution (thereby preserving the sharp edges and maximizing legibility), while at the same time compressing the backgrounds and pictures at lower resolution with a wavelet-based compression technique.

4.2. Creation of output files using OCR. The conversion of the scanned images to text is done automatically in batch mode by means of OCR software packages (ABBYY Fine Reader[4], Omnipage[5]). The precision of the process is strongly dependent on the quality of the original printed page and scanned images and the software settings. In most cases an additional manual editing step of the text is required in order to correct process resultant errors. After OCR is complete, derivative files (such as searchable PDF, DOC or ASCII text) are created.

5. Conclusion

The methodology and the workflow described in this paper can be used as a step by step guide in the process of making digital books from printed editions. They are based on the experience gathered at the digital Humanities Department of the Institute for Mathematics and Informatics in Sofia. Although compiled and tested with specific hardware and software platforms, they both can be applied for a wide spectrum of practical implementations. The authors will be grateful for any useful feedback and suggestions.

Acknowledgments. We would like to acknowledge the support of the project DIGMAP (Discovering our Past World with Digitised Maps, Programme eContentplus – Project: ECP-2005-CULT-038042)

Notes and References

[1] http://www-sul.stanford.edu/depts/diroff/DLStatement20030723.pdf Architecture and Workflow of the Digitization Lab.
[2] Scan and Share v1.07, http://www.djvu-soft.narod.ru/scan/scan_and_share_1_07.htm .
[3] http://www.rod-neep.co.uk/books/production/scan/ The book Scanning & Digitizing process.
[4] Y. Q. Liu, *Best practices, standards and techniques for digitizing library materials: a snapshot of library digitization practices in the USA*, Online Information Review 28 (5) (2004).

nikonomov@gmail.com, milena.dobreva@strath.ac.uk

[4] http://finereader.abbyy.com/
[5] http://www.nuance.com/omnipage/