Toolset for Image Virtual Recovery of Distorted Cultural Heritage Objects

David Asatryan
Institute for Informatics and Automation Problems of NAS RA
Russian-Armenian University,
Yerevan, Armenia
e-mail: dasat@iiap.sci.am

Grigor Sazhumyan
Institute for Informatics and Automation Problems of NAS RA
Yerevan, Armenia
e-mail: grigorsazhumyan@gmail.com

***Abstract*—this paper is devoted to a brief presentation of software tools designed for processing, analyzing, and restoring images of distorted cultural heritage objects. In particular, programs of segmentation, binarization, and virtual restoration were used, with the help of which it is possible to solve various problems of extracting the necessary information about an object from an image. Examples of restored images of objects distorted by cracks, various noises, and artifacts, as well as an example of image separation into background, text, and drawing are considered.**

***Keywords*— Ancient manuscript, distortion, segmentation, binarization, restoration.**

I. Introduction

Computer processing of images of historical, ancient documents, and other objects of cultural heritage of peoples is of special interest in the field of digital image processing. Historical documents, as a rule, contain important, sometimes unique, information about the historical events of antiquity, so an unambiguous and accurate perception of this information is necessary. However, historical documents undergo thermal, physical, and other types of influence during a long storage process, as a result of which, part of the information contained in it is distorted, and complete restoration of the content of the document becomes impossible. In such cases, it becomes expedient to apply, and even develop new computer technologies and mathematical methods for processing the images of such documents.

Over the past decades, teams of scientists from the Institute of Informatics and Automation Problems and the Russian-Armenian University have developed software modules, systems, and complexes for the efficient processing of images of damaged or distorted cultural heritage objects. Taking the opportunity to use the rich material available in the Matenadaran and other repositories of the ancient sources of the cultural heritage of Armenia, the developed software tools were used for structural analysis and improvement of the quality of distorted images, as well as virtual restoration of damaged objects.

The main method underlying the proposed procedures is image segmentation with various modifications, accompanied by the evaluation of certain informative parameters. At the same time, the toolkit provides a variety of possibilities for manipulating the structure and pixels of segments, including the exchange of the contents of various parts of the image.

This work is of a review nature, and the details of the applied methods are not given in the article, since these methods are published in the cited articles.

II. Applied Algoritms and Procedures

Let us briefly describe the main functions of individual tools.

Segmentation. The number of pixel intensity quantization levels is fixed, and full segmentation is performed so that all pixels of each segment correspond to the same quantization level, and each pixel has neighbors only from the same set of pixels [1]. In this case, all segments are saved as independent files, and a number of statistical characteristics of the distribution of pixels in the segment are stored. There is also a function to select the parameters of any segment for transmission to any other segment. This allows, in particular, recoloring the image.

Fig. 1. An example of complete segmentation of a color image at two levels of quantization. One of the received segments is shown

https://doi.org/10.51408/csit2023_55

Fig. 1 shows the result of full-color image segmentation with two quantization levels for each RGB color component and an example of selecting a separate segment.

Note that, despite the minimum number of quantization levels, the result of segmentation is quite good.

Virtual restoration of a heritage object. This operation calls virtual since it is not the original distorted object of cultural heritage that is restored, but only its image. This is an important feature, which can be used not only by amateurs; but also by professional restorers.

The concept of virtual recovery is as follows. Any part of the image, considered as damaged, consists of a set of certain segments, the content of which, in the opinion of the operator, has undergone a change in relation to the segments of the original. According to this concept, virtual image restoration is reduced to replacing the content of some segments with the content of others with suitable properties. Without going into details of the techniques that can be found in [2, 3], we will give examples of virtual recovery.

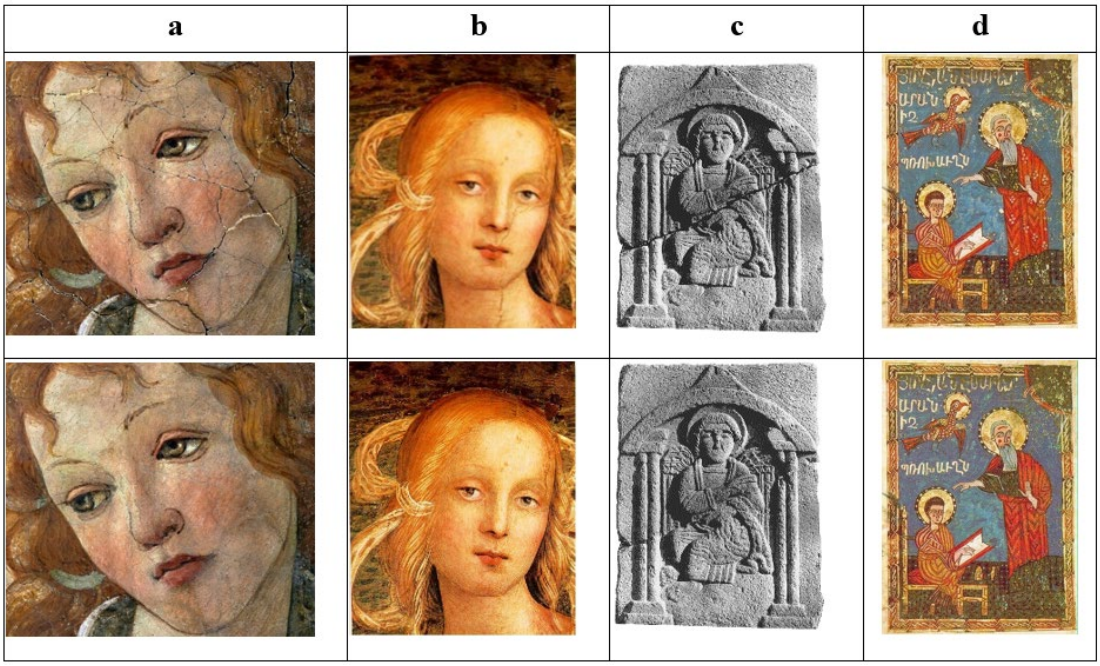

Fig. 2. Examples of virtual restoration of various distorted objects. a, b, and c - elimination of cracks, d - elimination of extraneous fragments of the image

The use of tools for the virtual restoration of damaged works of art is of particular interest. Below are examples illustrating the effectiveness of the proposed procedures.

Fig. 3. Virtual restoration of damaged art works of Armenian artists Abeghyan and Minas (fragment of a fresco)

Image binarization. Binarization, as a special case of segmentation, is one of the old but widely used methods of image processing.

Two types of binarization are distinguished in the literature: global (or thresholding) and local (or adaptive).

Global binarization involves choosing a threshold (or thresholds) applied to the entire image, found by optimizing a certain quality criterion.

Adaptive binarization is based on dividing the image into blocks using sliding scanning and independent binarization of the central pixel of each block separately. In this case, the threshold value is determined taking into account the characteristics of both this pixel and its neighbors. Many approaches and algorithms for adaptive binarization and their applications have been proposed in the literature [4].

Binarization is successfully used in the processing of ancient handwritten documents. At the same time, when binarizing images of noisy and distorted ancient handwritten or printed texts, the quality of binarization is often characterized by the degree of readability of the text or understanding of the content of the document.

Let us give examples of global and adaptive binarization.

It is known that global binarization cannot guarantee acceptable quality in all cases, since many images have a non-uniform pixel intensity distribution and contain differently distorted areas. In [5], a comparative analysis of various methods of global binarization was carried out.

Let us give examples of global and adaptive binarization.

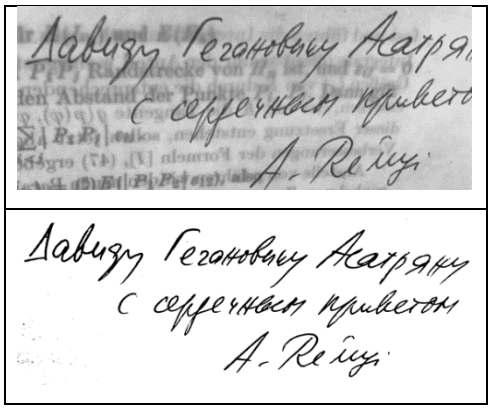

Fig. 4. An example of global binarization

It is known that global binarization cannot guarantee acceptable quality in all cases, since many images have a non-uniform pixel intensity distribution and contain differently distorted areas. In [5], a comparative analysis of various methods of global binarization was carried out.

However, in difficult cases, when there is a sufficiently strong background inhomogeneity, no global binarization algorithm works with sufficient efficiency. In such cases, adaptive algorithms are used, many of which have been proposed in the literature. Let us point out, for example, the well-known algorithms of Nibleck and Sauvola [6, 7], which also have various modifications with improved quality. To illustrate the above statement, let us give an example of adaptive binarization of an image with an uneven background, and for comparison, we will also give a variant of binarization by the Otsu method [8] (Fig. 5).

Thus, the use of adaptive binarization procedures allows for solving the main problems of processing ancient handwritten documents related to the accuracy of the reproduction of their content.

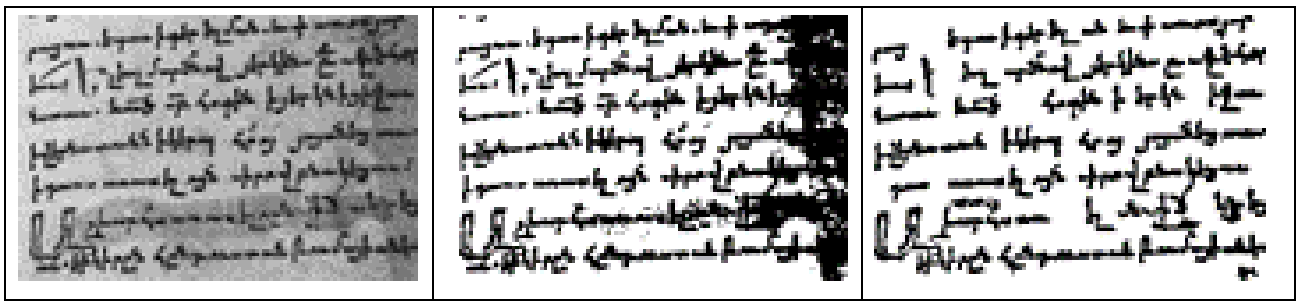

Fig. 5. An example of adaptive binarization using the quality improvement procedure

Background-text-drawing separation in the document. As a part of the toolset under consideration, a software tool that allows selecting text and a drawing from a document, as well as a part of the background that represents the existing artifacts [9]. An algorithm based on complete segmentation and subsequent binarization with the calculation of the sizes of the resulting "black" segments and the localization of segments with certain sizes is proposed. The intergroup scattering of pixel intensities was used as a criterion, similar to the criterion used in discriminant analysis. Let us show an example.

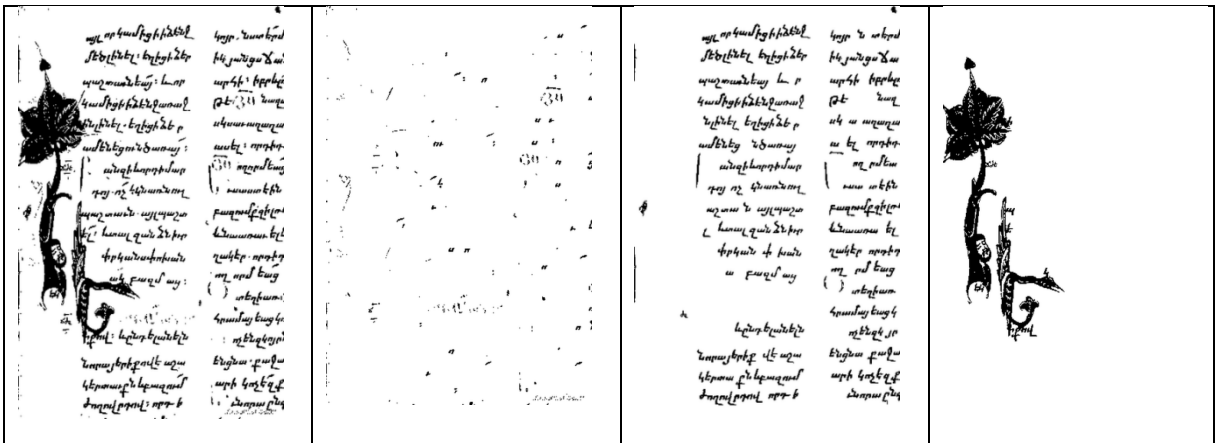

Fig. 6. An example of splitting a document (a) into three components: artifacts (b), text part (c), and drawing (d)

III. CONCLUSIONS

This paper illustrates the use of software tools developed by the authors for the processing, analysis, and virtual restoration of various cultural heritage objects through their images. Examples of objects with cracks, noise, with artifacts of various natures, as well as an example of splitting a document into artifacts, a text part, and a picture are given. The effectiveness of the developed software tools for processing historical documents and other objects of cultural heritage is shown.

REFERENCES

[1] Д. Г. Асатрян, Г. С. Сажумян "Об одном методе пороговой локальной сегментации изображения". *Mathematical Problems of Computer Science*, vol. 26, pp. 15-20, 2006.

[2] Д. Г. Асатрян, Г. С. Сажумян "О методе виртуального восстановления памятника культуры по его изображению". *Труды международной молодёжной конференции «Информационные технологии – 2005»,* Ереван, с. 346-350, 2005.

[3] Д. Г. Асатрян, Г. С. Сажумян "Метод когерентной сегментации и его приложение к восстановлению повреждённых изображений". *Вестник ГИУА, сер. Моделирование, оптимизация, управление*, т. 2(9), с. 15-21, 2006.

[4] Jagroop Kaur and Rajiv Mahajan, "A review of degraded document image binarization techniques", *International Journal of Advanced Research in Computer and Communication Engineering,* vol. 3, no. 5, pp. 6581-6586, 2014.

[5] D. Asatryan, M. Haroutunian, G. Sazhumyan, A. Kupriyanov, R. Paringer and D. Kirsh, "Comparative Quality Analysis of Image Global Binarization Procedures", *Proceedings of IX International Conference on Information Technology and Nanotechnology (ITNT)*, Samara, Russian Federation, pp. 1-5, 2023.

[6] W. Niblack, "An Introduction to Digital Image Processing", *Englewood Cliffs, N.J.:Prenlice Hall* pp. 115-116, 1956.

[7] J. Sauvola and M. Pietikäinen, "Adaptive document image binarization", *Pattern Recognition*., vol. 33, no. 2, pp. 225–236, 2000.

[8] N. Otsu. "A threshold selection method from gray-level histograms", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no 1, pp. 62–66, 1979.

[9] David Asatryan, Grigor Sazhumyan, Lusine Aznauryan, "Novel approach to Background-Text_Non-Text Separation in Ancient Degraded Document Images", *Proceedings of Int. Conf. "Computer Science and Information Technologies*, pp. 349-352, 2017.