Pattern Recognition 74 (2018) 326–339

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

A knowledge based architecture for the virtual restoration of ancient photos

E. Ardizzone, H. Dindo, G. Mazzola*

Dipartimento di Innovazione Industriale e Digitale - Ingegneria Chimica, Gestionale, Informatica, Meccanica (DIID), Università degli Studi di Palermo, Viale delle Scienze building 6, 90128, Palermo, Italy

ARTICLE INFO

Article history:
Received 29 July 2016
Revised 8 September 2017
Accepted 19 September 2017
Available online 21 September 2017

Keywords:
Image restoration
Historical photos
Digitization
Ontology
Knowledge base

ABSTRACT

Historical images are essential documents of the recent past. Nevertheless, time and bad preservation corrupt their physical supports. Digitization can be the solution to extend their "lives", and digital techniques can be used to recover lost information. This task is often difficult and time-consuming, if commercial restoration tools are used for the purpose. A new solution is proposed to help non-expert users in restoring their damaged photos. First, we defined a dual taxonomy for the defects in printed and digitized photos. We represented our restoration domain with an ontology and we created some rules to suggest actions to perform in case of some specific events. Classes and properties of the ontology are included into a knowledge base, that grows dynamically with its use. A prototypal tool and a web application version have been implemented as an interface to the database, and to support non-expert users in the restoration process.

1. Introduction

The art of photography is more than 150 years old, but it absorbed quickly the technological innovations of the following years. Historical photos constitutes a significant part of our Cultural Heritage, as well museums, paintings, archaeological sites. Time and careless preservation methods corrupt physical supports, therefore solutions must be found in order to protect their economic worth and high cultural value. Digitization is the definitive solution to preserve historical images and their content. Digital copies last almost forever, since they can be used and duplicated without losing quality. Furthermore, digital restoration techniques can be used to take images back to their original state. Professional digital restorers often use commercial software tools, like Adobe Photoshop, but they are neither automatic nor user friendly. In fact defects are subjectively detected, and correction methods are selected by user as well. This task is complex, expensive, and acceptable just for very important pictures.

In literature some general approaches that dealt with the problem of the virtual artworks restoration have been proposed. Landon et al. [1] presented a system to acquire and restore deteriorated film negatives, focusing onto the digitization process. Del Mastio et al. [2] developed a tool for processing cultural heritage images, with the goal to protect the digital reproductions of artworks, by means of a watermarking technique, and for the virtual restoration of some typical defects. Cislariu et al. [3] developed a tool for the detection and the restoration of digitized wood paintings. Marionidis et al. [4] presented an integrated tool for damage detection, shape restoration and texture restoration of faces appearing in Byzantine icons. Samko et al. [5] proposed a system to process fragile scrolled historical documents, without the need to physically unravel them, and to recover partially lost information. Hedjam and Cheriet [6] proposed a new approach to acquire images from historical documents by using a multispectral imaging system, and then restore them combining information from the different channels. The previous approach by Moghaddam and Cheriet [7] was based on an adaptive multi-scale binarization algorithm. Bhardwaj et al. [8] presented a new technique for photo defect detection, by sensing the light after it passes through the photo and then thresholding the digitized image to identify defect objects. Zhang et al. [9] proposed a framework, based on an inpainting algorithm, and a shape-from-shading technique, to remove unwanted distortions in digitized image from printed documents. Some state of the art algorithms for the detection and the restoration of specific defects in digitized degraded photos will be discussed in Section 2.

The adopted approach consists in the definition of an ontology to represent objects that are involved in restoration process of digitized degraded old photos. Ontology engineering has been pre-

* Corresponding author.
E-mail addresses: edoardo.ardizzone@unipa.it (E. Ardizzone), haris.dindo@unipa.it (H. Dindo), giuseppe.mazzola@unipa.it, giuseppe.mazzola75@gmail.com (G. Mazzola).

https://doi.org/10.1016/j.patcog.2017.09.031
0031-3203/

viously applied to the image processing, for example, domain in [10], in which a photo annotation ontology is used to help users in querying images, and in [11], for the art image retrieval problem.

Photo restoration can be made automatic by finding a way of representing knowledge in order to formalize the pair "kind of defect-recovering algorithm". Hence, it is necessary to design a scheme which enables to select the proper restoration methodology starting from the interpretation of the defect meta-representation. Such a scheme constitutes an ontology in which the need of restoration, expressed by the image meta-representation, is satisfied by means of experience coming from knowledge. Starting from the ontological representation, a knowledge-based model is created to guide non expert users through the restoration process of the degraded artworks. In particular we analysed photos from Alinari Archives in Florence, which is composed of high resolution, colored or black and white images since 1840.

The proposed approach presents some completely novel tools for digital image restoration:

- while manual restoration process of printed photos is based on the physical-chemical causes of defects, digital restoration algorithms process digital damage features. Our approach let us to correlate damages in printed and digitized photos;
- the restoration process is based on knowledge and then it is able to receive experience about new kinds of defects along with more recent and effective algorithms for restoration: the model constantly grows while it is used;
- the meta-representation of image damages can be also used in finer applications, as content retrieval or the automatic definition of pictures degradation typologies;
- a basis of knowledge is defined to automatically select the best restoration method for an image which is affected by a specific damage.

2. Damage taxonomy

Old images may present a huge variety of damages, due to several different factors. Some defects may lead to a complete loss of information, while others deteriorate the overall appearance of images. Mostly, damages are originated by inappropriate environmental conditions (temperature, humidity, lighting), inaccurate handling (dirt, image protection, cracks), human intervention (stamps, writings, restorations) and chemical factors (reactions with micro-organisms). While the origin of image defects on the physical support (whether positive or negative) is an important issue for a manual restoration activity, several defects appear similar once images are digitally scanned (Fig. 1) and can be described and removed by similar underlying processes. Several works studied specific defects of digitized/digital pictures: cracks and craquelures [12–14], water blotches and foxing [15], semi-transparent blotches [16], fading [17,18], scratches in photos and movies [19,20], ink sprays and scratches [21], color [22,23]).

A first interesting attempt to classify defects in old photos was proposed in [24], but it was incomplete and it did not focus on the digital aspect of the defects. In [25] we proposed our origin-based defect taxonomy, but it was not either based on digital features (shape, color, texture, etc.). Finally, an observer will not be able to discriminate an abrasion from a tear, if their digital versions have similar aspect (see Fig. 1).

Fig. 2 shows how manual restorers annotate their printed documents: each type of defect is identified by a standard numerical code, and each defect is removed using the most appropriate approach. The need of a digital damage taxonomy arises from the aim to associate damages to the most appropriate "digital" detection and restoration algorithms. To analyze the damages of a digital image, instead of studying their origin, digital features (shape, color, texture, etc.) must be analyzed. An automatic inspection method won't be able to discriminate an abrasion from a tear, if their digital versions have similar features (see Fig. 1). Then we proposed [26] two different taxonomies, to distinguish defects in printed photos and in their digital versions, and to compare features of the two defect sets.

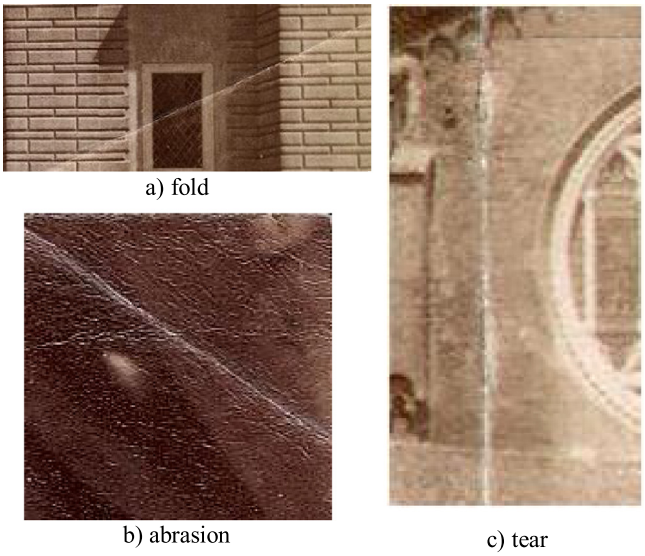

Fig. 1. A comparison between three defects, which have similar digital aspects. In the printed version, according to the physical-chemical origin, they are classified as three different types of defect.

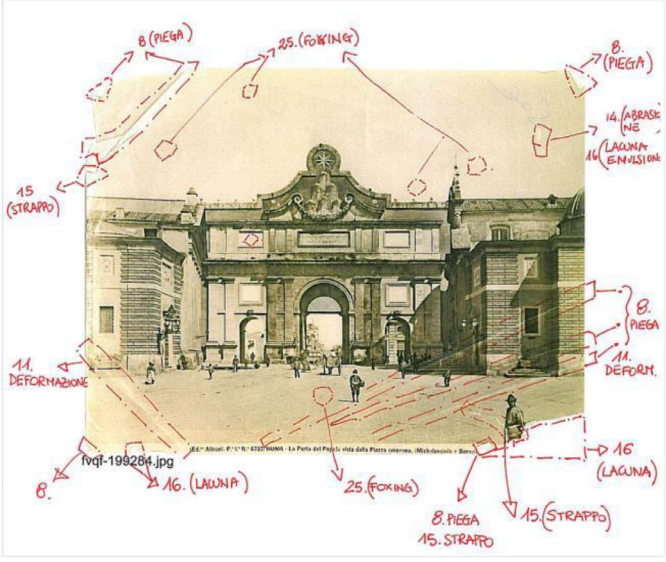

Fig. 2. An example of an annotation file, written (in Italian) by a manual restorer from the Fratelli Alinari Museum Collection (courtesy).

2.1. Real defect taxonomy

Table 1 shows the taxonomy used by expert manual restorers of the Fratelli Alinari Museum Collections in Florence to annotate their photos. Each type of defect is labeled by a standard numerical code, which univocally identifies it (see also Fig. 2). According to their origin, real defects of old photographic prints can be divided in different sets: