

A Dual Taxonomy for Defects in Digitized Historical Photos

E. Ardizzone, A. De Polo*, H. Dindo, G. Mazzola, C. Nanni*

*Dipartimento di Ingegneria Informatica (DINFO) dell'Università di Palermo.
Viale delle Scienze – Ed. 6, 90128, Palermo, Italy
ardizzon@unipa.it, dindo@csai.unipa.it, mazzola@csai.unipa.it*

**Fratelli Alinari Museum Collections, (RMFA),
Largo Fratelli Alinari, 15, 50123, Florence, Italy.
andrea@alinari.it, restauro@alinari.it*

Abstract

Old photos may be affected by several types of defects. Manual restorers use their own taxonomy to classify damages by which a photo is affected, in order to apply the proper restoration techniques for a specific defect. Once a photo is digitally acquired, defects become part of the image, and their aspect change. This paper wants to be a first attempt to correlate real defects of printed photos, and digital defects of their digitized versions. A dual taxonomy is proposed, for real and digital defects, and used to classify an image dataset, for a posteriori comparative study. Furthermore, a set of digital features is analyzed for digitized images, to identify which of them could be useful for an automatic inspection method.

1. Introduction and previous works

The art of photography is more than 150 years old, but it absorbed quickly technological innovations of the following years. Methods, cameras, techniques changed and improved, and so supports changed, from physical (paper) to digital ones. Nevertheless, the economic worth and high cultural value of old documents induced the use of digital techniques to protect and preserve them as “cultural objects”. Old photographic prints may present several types of defects, caused by inaccurate handling and/or store of the original image, or by chemical factors, or by decomposition of the support. If the knowledge of a degradation origin is essential for defect analysis on the physical support, different defects may look similar once the document is digitized, and could be described and removed by similar underlying processes. Several works rely on the damage analysis and restoration of digitized/digital pictures: cracks and craquelures [1],

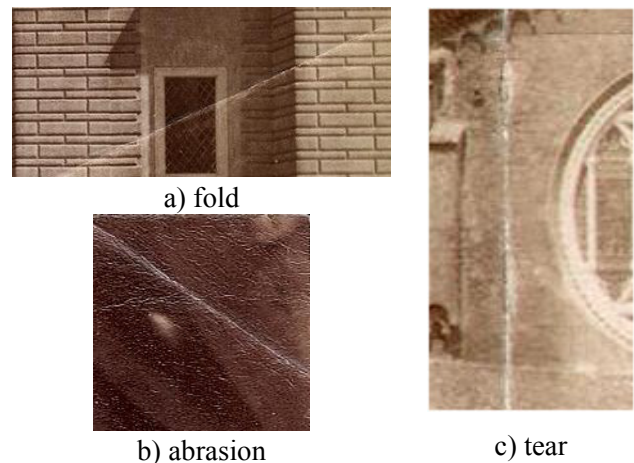


Figure 1 A comparison between three defects, which have similar digital aspects. According to their origin they are classified as three different types of defect.

water blotches and foxing [2], fading [3], scratches in photos and movies [4][5]. PrestoSpace project [6] focused on defects in audiovisual collections.

A first interesting attempt to classify defects in old photos is proposed in [7], but it was incomplete and it didn't focus on the digital aspect of the defects. In [8] we proposed our origin-based defect taxonomy, but it wasn't either based on digital features (shape, color, texture, etc.), which should be analyzed, rather than considering the origin. For example, an observer won't be able to discriminate an abrasion from a tear, if their digital versions have similar aspect (see fig. 1).

In this paper we present a new dual taxonomy, in order to find distinctions and correlations between defects in printed and digitized photos. Furthermore, for the digital taxonomy, we analyzed relationships between each type of damages and a subset of MPEG-7 visual descriptors.

Table 1 Real defect taxonomy			
Biological Alterations		14	abrasions
1	infections	15	tears
2	infestations	16	lacunas
3	other	17	cracks
Physical Alterations		18	Pr. of adhesive
4	garbage	19	other
5	dust	Chemical Alterations	
6	fingerprints	20	spots
7	stains	21	fading
8	folds	22	yellowing
9	craquelures	23	silver mirroring
10	lifting	24	sulfuration
11	deformations	25	foxing
12	bending	26	other
13	marks		

2. Dual taxonomy

This section presents the proposed dual taxonomy. Section 2.1 deals with “real” defects of old printed photos, while section 2.2 analyzes defects according to their digital features.

2.1 Real defect taxonomy

Fig 2 shows a typical annotation file of a manual restorer for a damaged printed photo. Table 1 reports the taxonomy used by expert manual restorers of the Fratelli Alinari Museum Collections in Florence. 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:

2.1.1 Mechanical (physical) damages

Usually originated by inaccurate handling and/or store of the original image, can be further divided into:

- **Deposited matter:** different materials adhere to the surface creating small spots that cover the original image. Some examples are:
 - *Dirt & Dust:* may be hygroscopic, absorbing water or humidity, causing molds, hydrolysis and acidification;
 - *Fingerprints:* the finger impression modifies the emulsion structure; typical shape;
- **Stains:** spots originated by water or humidity; they alter the color and the physical-chemical structure of the phototype components; the presence of water may swell the surface of the support;

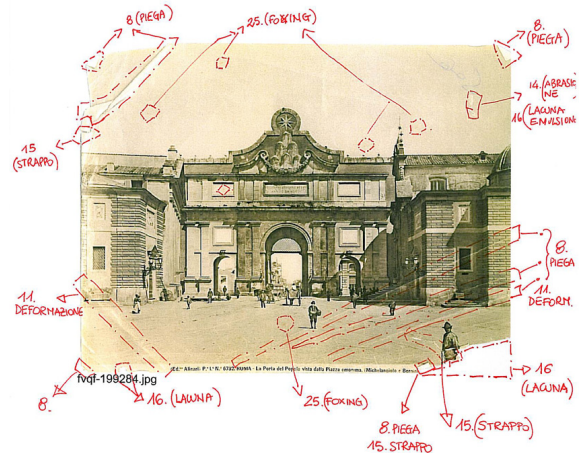


Figure 2 An example of an annotation file, written (in Italian) by one of the manual restorers of the Fratelli Alinari Museum Collection (courtesy)

- **Physical alteration of images:** often lead to a complete loss of information and should be removed by specialized techniques. Typical examples are:
 - *Cracks:* may be very large; do not exhibit a dominant orientation; however, each crack has its own direction;
 - *Folds:* caused by improper human intervention; a bend in a material which may cause a crack or break in the emulsion or support;
 - *Craquelures:* micro-fractures of the support of the photo, usually branched; orientation depends on that of the paper fibers of the support;
 - *Abrasions:* lack in the emulsion of the support, caused by friction with other part of the photo or with some external tool (i.e. scratches);
 - *Tears:* caused by improper human intervention; a rip of the support, with or without paper loss;
- **Deformations:** originated by an inappropriate conservation of original images; often caused by excessive humidity and/or temperature and corrupt the way the gelatin is fixed to the support; the effect is a deformation of the planarity of the support
 - *Lifting:* the local or diffuse partial detachment of the structural components of a photo from the support;
 - *Bending:* curving of the support of the photo;
- **Human retouches:** deliberate human retouches that irremediably alter the image; some examples:
 - *Lacunas:* a portion of the photo, or of one of the components, that is totally lost;
 - *Marks:* impression that modifies the surface or the structure of the phototype components;
 - *Presence of adhesive:* caused by scotch tapes or glue, which degradation alters the color of the support or of the image (to amber-ochre yellow).

2.1.2 Chemical damages

Chemically originated, may be further divided into:

- **Spots:** local or punctual alteration of the original colouring of the image. Examples are:
 - *Foxing:* chemical reactions between the print and some micro-organisms; appears as reddish-brown spots, with irregular edges, with a darker kernel and a lighter surrounding area;
 - *Silver mirroring:* bluish-silver colored spots; more evident in the darkest areas of the image; caused by bad quality of the wrappers or pollution;
 - *Sulfuration:* dark-gray blue spots due to chemical reaction between sulphur and metal components of the support; caused by pollution or incomplete washing phase during the photo development;
- **Tonal and color balance:** originated by an excessive exposure of original photos to light; some examples:
 - *Fading:* overall whitening of the image; contrast and details of the image are lost;
 - *Yellowing:* alteration in the image chromaticity, which tends to yellow; may be also caused by improper handling or preservation.

2.1.3 Biological damages

Deterioration of the support due to the attack of living organism. May be further divided in:

- **Infections:** biological attacks (by fungi, molds or bacteria) to the emulsion or the support; they look like colored spots (yellow, brown, violet, black);
- **Infestations:** losses and staining caused by insects or small rodent feeding on the photographic materials.

2.2 Digital defect taxonomy

Within the proposed digital taxonomy, a damage can be **global**, if it affects every pixel in the image, **diffuse**, if distributed overall the image, without affecting all the pixels, and **local**, if it is located in a limited portion of the image. According to their digital aspect, defects may be classified as:

- **Spots:** local defects. Fig. 3.a. Underlying information is lost and must be fully substituted. No specific colors. More or less rounded shaped.
- **Semi-Transparent Spots:** local defects. Fig. 3.b. Similar to spots, residual information can be recovered with restoration techniques.
- **Scratches:** local defects. Fig. 3.l. Thin lines, with a preferential direction. Usually lighter than the rest of the image. Can have a darker kernel. Possible limited changing in width and slope.
- **Foxing:** local/diffuse defects. Fig. 3.d. Covering or semi-transparent spots. It is composed by a dark red- brown kernel, surrounded by a lighter red area, which is usually semi-transparent.

- **Folds:** local defects. Fig. 3.g. Located near the edges of the photo. It is composed by a lighter central area (the verso side) and darker edges, depending on the acquisition operation.

- **Cracks:** local defects. Fig. 3.e. Undefined orientation. In some cases they can have branches. Cracks are usually composed by a darker kernel surrounded by a lighter area.

- **Deformations (lifting):** local defects. Fig. 3.f. Due to the digital acquisition of a non planar support. They look like the negative of a branched crack.

- **Blotches:** diffuse defects. Fig. 3.c. Semi-transparent spots which can be seen all over the image. Usually lighter than the rest of the image.

- **Whitening:** global defect. Fig. 3.k. Overall fading.

- **Yellowing:** global defect. Fig. 3.j. Distortion in the chromaticity (to yellow) of the whole image.

- **Lacking color:** local defect. Fig. 3.m. Undefined shape (some similar to scratches, some other to spots). Usually white, because of the exposition of the color of the support. Information is totally lost.

- **Lacking Portions:** local defect. Fig. 3.h. Usually black, but it depends on the acquisition condition. In most cases lacking portions have jaggy edges.

- **Handwritings:** writings and scrawls. Fig. 3.i. Complex curve lines, darker than the rest of the image.

Table 2 lists the digital defects within our testing dataset (about 220 images from the Alinari Photographic Archives in Florence). Table 2 shows also a-posteriori analysis of the relationships between the two proposed taxonomies, we used to annotate our dataset. Note that there is no 1-1 relationship between the defects in the two taxonomies. For example, “digital” spots can be caused by the digital acquisition of defects such as garbage, abrasions or chemical spots. On the other hand, real photos, which had been manually annotated as affected by abrasion, according to the digital defect taxonomy are classified as spots, lacking color, or scratches. It is clear that digital and manual restorers often cannot be in agree about the classification of a defect in an image. This work wants to be the first attempt to build an useful tool to let digital and manual restorers to draw nigh their different points of view.

3. Description ability

We used our database, which has been manually annotated using the digital taxonomy, to test the description ability of a subset of the standard digital descriptors. The purpose is to find which are the most suitable descriptors for each kind of damage. Actually, three color (Color Coherence Vector Dominant Color and Color Structure) and two texture (Edge Histogram

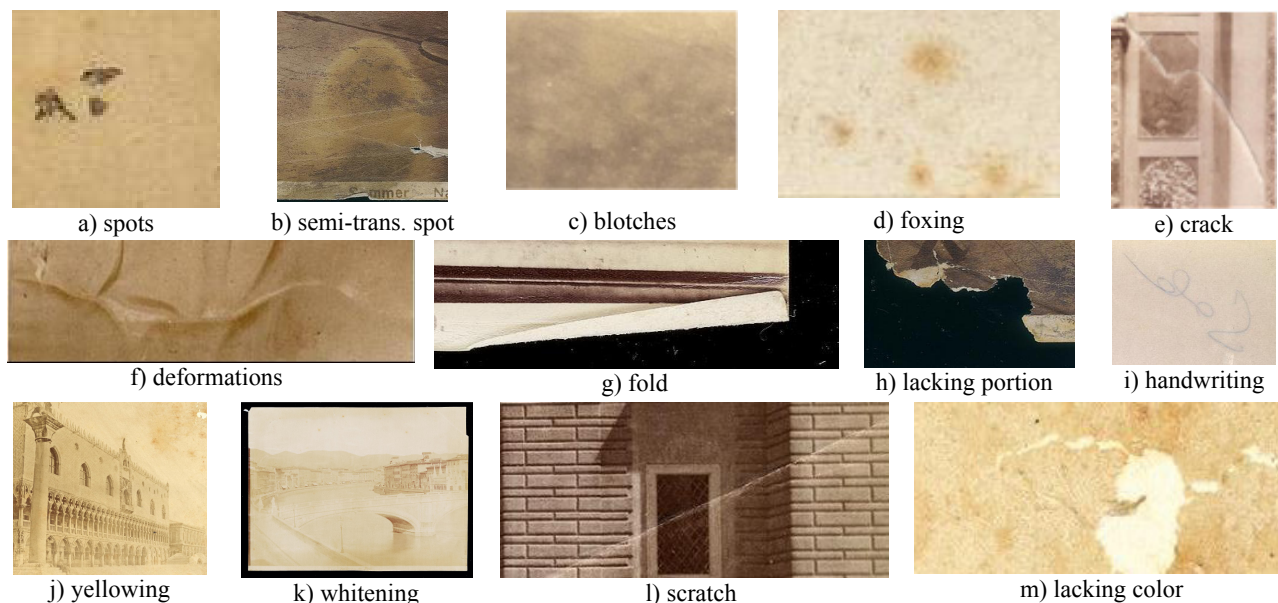


Figure 3 some examples of defects classified according to the digital taxonomy (courtesy of Fratelli Alinari Museum Collections (RFMA) in Florence).

and Homogenous Texture) descriptors have been tested (see [9][10]). No shape descriptors have been yet used for testing. For each descriptor we implemented a simple content-based image retrieval (CBIR) application, to compare a damaged image to our set of annotated images. We used Euclidean distance between descriptor values as metrics for comparison. That is, given a image, we compare its content, according to the selected descriptor, to those of all the images in the dataset. The damage of the most similar image, is assigned to the tested image and then compared to the manual annotation. In our experiments, the whole dataset has been used for testing. Table 3 shows our experimental results. First of all, color descriptors work better than texture ones. In particular we observed that between color descriptors, Dominant Color (DC) gives best results, for almost all the defects. Color Structure (CSD) works better if used to describe covering and semi-transparent spots. With respect of texture descriptors, Edge Histogram (EHD) better describes line-like defects (scratches above all), while Homogenous Texture Descriptor (HTD) is appropriate for spot-like (spots, foxing, etc.). Global (color) defects cannot be described by texture.

Note that table 3 would suggest that some defects, could be classified as the same damage, because they can be described using the same descriptor set. Furthermore, relationships between descriptors and damages, and the development of damage specific ones, will be an essential factor in designing an automatic classification method.

4. Conclusions

Manual and digital restorers don't speak the same language. The first ones watch for the origin of a defect, in order to manually remove it from the support, using the appropriate techniques. Digital restorers must concern with the digital features of the image, because they use digital techniques to describe, detect and restore defects. That's why a defect taxonomy based on the origin (physical, chemical, etc.) of defect in printed photos, cannot be used to catalogue defects of their digitized versions. Our work wants to be a first attempt to show the differences and the analogies of the two approaches to the problem. A larger dataset is needed to try to derive a relationship between digital damages and their causes in the original printed photo.

This paper presented also a preliminary study on the relationships between MPEG-7 visual descriptors and digital defects of the proposed taxonomy. Further descriptors must be tested, and further experiments made, to complete this analysis. Our final purpose is to implement an automatic method to classify digital defects of damaged photos.

Acknowledgements

This work has been partially funded by the MIUR (Italian Ministry of Education, University and Research) project FIRB 2003.

References

- [1] F. S. Abas, and K. Martinez, "Craquelure Analysis for Content-Based Retrieval", In 14th International Conf. on Digital Signal Processing, July 1-3, 2002, Santorini, Greece.

Table 2 Digital defect taxonomy. For each digital damage it is shown the percentage of images in the dataset, affected by that digital damage, which has been annotated with a real damage, identified by the numerical code of Table 1 .

DIGITAL DEFECTS	TYPE	REAL DEFECTS
Spots	Local	4(6%), 6(3%), 14(3%), 18(3%), 19(30%), 20(46%), 23(3%), 25(6%)
Semi-Transparent Spots	Local	20
Scratches	Local	8(10%), 11(15%), 14(50%), 15(15%), 17(10%)
Foxing	Local/Diffuse	25
Folds	Local	14
Cracks	Local	15(75%),11(25%)
Deformations	Local	11
Blotches	Diffuse	7(5%), 20(35%), 23(60%)
Fading	Global	21
Yellowing	Global	22
Lacking Color	Local	14(43%),16(57%)
Lacking Portions	Local	8(5%), 11(13%), 16(77%), 18(5%)
Handwritings	Local	19

Table 3 Description Ability. CC=percentage of correctly classified defects, with respect of the number of that defect in the dataset. FP=percentage of false positives, with respect to the whole dataset. Last five rows are not statistically significant, because of the limited number of samples, but shown for completion.

DEFECT	DESCRIPTORS									
	CC(def)%					FP(tot)%				
	CCV	DC	CSD	HTD	EHD	CCV	DC	CSD	HTD	EHD
spot	50	55	60,5	60	7	6	4	8	8	3,5
semi-transparent spot	37,5	57	55,5	37,5	37,5	1,5	1,5	2,5	3,5	3
blotches	78	91	65	27	4,5	4	2,5	4	7	9
scratch	45	62	35	19	47,5	4	3	6	8	13,5
foxing	89	89	76	80	29	5	2	7	11	3
lacking portion	70	73	41	13,5	0	2	4	5	8	3
lacking color	70	73,5	17,5	16	10,5	2	3,5	7	4	8
yellowing	43	71,5	57	0	85,5	1	2	1,5	2	25,5
fold	100	100	50	0	0	0,5	0,5	0	0,5	0
crack	50	67	50	0	33	1	1	1	2	5
deformation	0	0	33	0	0	2	2	1	0,5	2,5
whitening	0	0	50	0	0	2	1	0,5	1,5	1,5
handwriting	67	67	0	0	0	1	1	1	2	1

[2] F. Stanco, L. Tenze and G. Ramponi, "Virtual Restoration of Vintage Photographic Prints Affected by Foxing and Water Blotches", J. of Electronic Imaging, vol.14, no.4, Oct.-Dec. 2005.

[3] G. Ramponi and V. Bruni, "Virtual Restoration of Faded Photographic Prints", Proc. EUSIPCO-06, 14th European Signal Processing Conf., 4-8 Sept. 2006, Florence, Italy.

[4] V. Bruni, D. Vitulano, "A Generalized Model for Scratch Detection", IEEE Transactions on Image Processing, Vol. 13, No. 1, January 2004.

[5] E. Ardizzone, H. Dindo, O. Gambino, G. Mazzola, "Scratches Removal in Digitised Aerial Photos Concerning Sicilian Territory", 14th Int. Conf. on Systems, Signals and Image Processing, Maribor, Slovenia 27-30 June 2007.

[6] <http://prestospace.org>.

[7] A De Polo, F. Stanco, G Ramponi, "Towards the Automated Restoration of Old Photographic Prints: A Survey", In Proc. of IEEE EUROCON 2003, pp. 370-374, Ljubljana, Slovenia, September, 2003.

[8] E. Ardizzone, H. Dindo, U. Maniscalco, G. Mazzola, "Damages of Digitized Historical Images as Objects for Content Based Applications," in Proc. EUSIPCO 2006, 4-8 Sept. 2006, Florence, Italy.

[9] T. Sikora, "The MPEG-7 Visual standard for content description - an overview". IEEE Trans. on Circuits and Systems for Video Technology, special issue on MPEG-7, vol.11 pp:696--702, June 2001.

[10] G. Pass, R. Zabih and J. Miller, "Comparing Images Using Color Coherence Vectors", ACM Conference on Multimedia, Boston, Massachusetts, Nov. 1996, pp. 65-74.