1

Introduction

The history of retouching images is as old as film photography, where the usage of chemicals, dyes or graded tone paints is common practice in the silver print workflow: after the darkroom processing, the prints are usually manipulated with soft brushes for removing dust spots, hairs or scratches enlarged from the film negative; see Lambrecht and Woodhouse (2013) for common print spotting techniques. The artistic skills in manipulating the final paper print need to take into account both the structure and texture of the image, as well as the film grain. These skills are even more demanding when large areas have to be reconstructed: famous examples include the substitution of people in group photographs with the landscape background for propagandist political purposes or photo montages. The ability to retouch photographs is also a key ingredient in digital images. With the rise of digital photography and photo-editing software, a lot of effort has been made by the mathematical and computer vision communities to develop effective methods and algorithms for digital imaging data. On top of that, the increasing quality of recording reached by camera sensors requires continuous improvement of the algorithms to adapt to modern needs, both in performance and in diminishing the visual impact of a retouch.

What is Inpainting

In the scientific community, the problem of recovering missing data in digital images, initially called *disocclusion* or *amodal completion*, dates back to the early 1990s. It relates to recovering the contours of occluded objects, thus making them appear disoccluded. Early approaches were based on mathematical methods for recovering missing data from a few samples either lying on equally spaced points or at scattered locations, with the usual assumption that the function to be recovered has a certain degree of smoothness making easy its interpolation on unknown locations. However, the inpainting problem is much more than an interpolation problem, embracing a broader class of methods dealing with the non-uniqueness of the solution and its expected plausibility as a natural image.

Real-life data are not smooth in general: for example, images come with edges, jumps of colour or structural discontinuities. Thus, it is necessary to set the mathematical problem in a space allowing for such discontinuities.[1] Moreover, the image texture (e.g. in grass, stones, etc.) makes the reconstruction problem even more challenging, especially if the task is to recreate an image that is visually plausible to the human eye and indistinguishable from the expected original. Additionally, the difficulty of the problem is closely related to the size of

[1] This is called the *space of functions of (special) bounded variation*, Ambrosio et al. (2000).

the region to be reconstructed. If the region is relatively large, a simple propagation of either smooth or texture information is not enough, and a knowledge of the semantics of the full image is necessary. Nowadays the term *inpainting* is commonly accepted for describing the general task of restoring the geometric and texture content in damaged images by filling in the unknown regions (which can be of any size and shape) with plausible visual information. The word "inpainting" first appeared in the seminal work of Bertalmío et al. (2000), borrowed from the dictionary of art conservators dealing very often with the restoration of fragmented missed data in artworks.

Why this Book

Over the last few years, the inpainting research field has become extremely popular in the scientific community. Existing reviews of digital inpainting describe only thematic approaches, e.g. local methods (Schönlieb, 2015) and non-local methods (Guillemot and Le Meur, 2014; Buyssens et al., 2015a), as well as efforts to organise the vast literature on deep learning methods in Elharrouss et al. (2019). As authors, we believe that a comprehensive and detailed monograph covering the whole history of inpainting is still lacking and we are aiming to match this challenge.

In this monograph we offer the reader a timeline of the evolution of inpainting research, with a description of methods selected by their novelty and representing the overall research directions in the field, from early initiatives to the present. In particular, we analyse the theoretical contributions and emphasise the quality of the solutions that were proposed to the novel challenges presented at each stage in this evolution, e.g. the increasing size of the images to be processed, the growing expected quality of seamless reconstruction and the efficiency of the algorithms involved. Additionally, we consider different test cases and image comparisons regarding damages from imaging data arising in the cultural heritage domain: this is another key aspect of this work. Indeed, we also aim to enrich the existing reviews by focusing on the restoration of artworks, specifically the collection of illuminated manuscripts in the Fitzwilliam Museum (Cambridge, UK), ideally to bring back the inpainting research field to the place where it originated.

All things considered, with this monograph we aim to promote new interdisciplinary research opportunities at the intersection between art and science.

How to Read this Book

The book is targeted at both image inpainting experts, serving as a reference book for mathematicians, engineers and computer scientists interested in image processing and image inpainting more specifically, and also art historians who are keen to understand the main concepts behind image inpainting and its various capabilities. In order to achieve this, we have structured the book in a two-layer format. The main text is accessible to our whole target readership. The framed boxes give technical details of particular interest for mathematicians and image inpainting experts.

1.1 Mathematical Setting

In order to describe an image mathematically, we introduce some standard notation to be used in all the following chapters unless otherwise stated. Additional terminology for specific use within each chapter will be discussed at the start of the chapter.

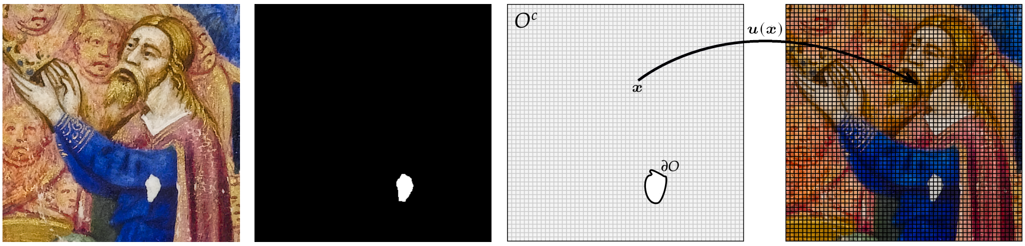
In the continuous setting, where an image is defined over a continuous range of coordinates, let $\Omega \subset \mathbb{R}^2$ be an open, bounded Lipschitz domain, meaning an open and connected set with sufficiently regular boundary. Specifically, at each point on its boundary, Ω can locally be described as the graph of a Lipschitz-continuous function indexed by the variable $\boldsymbol{x} = (x, y)$; its outer normal is denoted by $\boldsymbol{n}(\boldsymbol{x})$. A grey-scale image is defined as a function $u\colon \Omega \to \mathbb{R}$, while a colour image is defined as a multi-valued function $\boldsymbol{u}\colon \Omega \to \mathbb{R}^C$, where C is a positive value indicating the number of colour channels. The *occluded* or *inpainting* domain is the portion O of the imaging domain Ω that contains the region with the damage to be restored, with boundary denoted by ∂O. At the same time, the *intact* domain is denoted by its complement $O^c := \Omega \setminus O$, and the initial image is denoted by $u^\diamond(\boldsymbol{x})$ (the composite of the intact part for all $\boldsymbol{x} \in O^c$ and of the damaged part for all $\boldsymbol{x} \in O$). Thus, the mathematical definition of the *inpainting* problem can be given as the *interpolation* problem aiming to recover the missed values in O, according to the information available in $u^\diamond$.

Since the inpainting problem is mostly addressed using algorithms dealing with real images that, unlike their abstract mathematical representation, lie in a discrete setting, we introduce some specific terminology. In the discrete setting, the domain Ω can be represented by a discrete grid of height H and width W whose vertices, called *pixels*, are equally spaced by a certain grid size $h > 0$. Therefore, a grey-scale image is represented as a matrix of size $H \times W$, with positive entries falling typically in the range $[0, 255]$ (for an 8-bit image) and which are usually unrolled into a vector $\boldsymbol{u}\colon \Omega \to [0, 255]^{HW}$; a colour image of C channels is represented by layered matrices as $\boldsymbol{u}\colon \Omega \to [0, 255]^{H\times W\times C}$ or $\boldsymbol{u}\colon \Omega \to [0, 255]^{HW\times C}$, in the case of unrolled images.

In the discrete setting and in order to preserve the *fidelity* information on the intact domain, we will make use of matrix notation to distinguish between the inpainting mask and the intact mask. The *inpainting mask* denotes the pixels belonging to the inpainting domain and is represented by a matrix $\mathbf{M}$ which is the same size $H \times W$ as the image and which is associated with the inpainting domain O via the indicator function

$$\mathbf{M}(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \boldsymbol{x} \in O, \\ 0 & \text{if } \boldsymbol{x} \in O^c. \end{cases}$$

The *intact mask* is denoted by the indicator function on the intact domain O^c, i.e. $\mathbf{1} - \mathbf{M}(\boldsymbol{x})$, where $\mathbf{1}$ denotes the matrix of size $H \times W$ with all entries equal to 1. For the unrolled version of the image, we will also take advantage of a diagonal projection matrix $\mathbf{P}$ and its complement $\mathbf{I} - \mathbf{P}$, where $\mathbf{I}$ is the identity matrix, both of size $HW \times HW$ and acting on the unrolled version of each colour channel of the rectangular image. The main diagonal of $\mathbf{P}$ consists of the unrolled entries of $\mathbf{M}(\boldsymbol{x})$ and, by means of $\mathbf{P}$, we can identify the intact part of the vector image as $\boldsymbol{u}^\diamond = \mathbf{P}\boldsymbol{u}$, of size $HW \times 1$ (for grey-scale images) or $HW \times C$ (for coloured images). An illustration of this terminology is given in Figure 1.1.



(a) Discrete image $\boldsymbol{u}$ (before unrolling) (401×401 px, $C = 3$)

(b) Mask $\mathbf{M}$: the inpainting domain O in white, O^c in black and $O \cup O^c = \Omega$.

(c) A visual diagram: (left) the image space, showing the boundary ∂O of the inpainting domain O; (right) the image representation. The image coordinate $\boldsymbol{x}$ is mapped to the image pixel colour via the discrete function $\boldsymbol{u}(\boldsymbol{x})\colon \Omega \to [0, 255]^3$.

Figure 1.1 Illustration of the notation used in this text. Coronation of the Virgin (detail) from MS 153, fol. 15r. © Fitzwilliam Museum, Cambridge, UK.

1.2 Overview of Inpainting Approaches

The classic mathematical approaches for the inpainting problem are essentially two: *geometry-oriented* and *texture-oriented* methods (also called *exemplar-based* methods). With the increasing computational power of computers and graphics processing units (GPUs), as well as the availability of large datasets, these approaches have been overtaken by the new era of *deep learning* techniques, where the geometric and texture approaches are eventually combined, plus other possible clues such as semantic information.

Geometry-oriented approaches (see Chapter 2) are *local* methods, which are influenced only by points neighbouring the inpainting domain: thus, they are designed to propagate local information, mainly level lines (contours of equal brightness, or isophotes), from the boundary of the inpainting domain to its inside. They were very popular in the mid 2000s and produce good results for tiny scratches that need to be filled in (but are not suited for propagation of the texture).

Texture-oriented approaches (see Chapter 3) are *non-local* methods, which originated as an application of the texture synthesis problem: they are able to recover information from the whole image. This is different from the approach based on the influence of local neighbours in geometric-oriented methods. Texture-oriented approaches are also referred to as *exemplar-based* or *non-parametric* methods, since they look for similar examples in the intact part of the image for realistic reconstruction. They peaked in popularity in the mid 2010s and do provide excellent results in many situations where there is limited data available. However, they only perform well in an inpainting task with repetitive patterns or more complex textures if enough examples are present in the same image. For this reason a good *similarity measure* between small portions within the image, called *patches*, needs to be defined and customised. Therefore, texture-based methods are not able to generate reasonable new geometries.

From the beginning of 2010 onwards, a novel class of inpainting methods based on neural networks and deep learning (DL) strategies appeared in the literature (see Chapter 4). Those approaches aim to leverage the information found in large-scale datasets, thus combining local, non-local and high-level information. While early deep learning approaches used reconstruction distances on small inpainting regions, perceptual objectives were soon introduced, frequently in a two-step approach where structure and texture are incorporated.

Table 1.1 *Summary of inpainting approaches: local (L), non-local (NL) and deep learning (DL)*

Type	Ch.	Unit	Mathematics	Inpainting domain	Drawbacks	Source of data
L	2	pixel	fluid dynamics PDE, variational methods	tiny for textured or large for cartooned images, regular geometry	no texture reconstructed	single image, the boundary of the inpainting domain
NL	3	patch	variational methods, minimisation of patch similarity functions	relatively large with irregular geometry	subject to availability of the content to be replicated	single image, the intact part
DL	4	—	neural networks, minimisation of loss functions	size variable (better visual results for smaller size), irregular geometry	dataset size and type, neural training time, possibly unrealistic results	multiple images, content learned from very large datasets

On the other hand, deep learning methods depend very much on the dataset on which they are trained, as these methods aim to learn underlying visual priors from data, including geometric, smoothness, self-similarity patterns, global perceptual coherency and image semantics. Note that, as in the inpainting literature, throughout the book the terms inpainting region, inpainting hole or just hole are used interchangeably.

We summarise the main differences between such methods in Table 1.1.

Terminology

Local and non-local approaches can also be classified with respect to the domain in which the inpainting task is performed. For example, inpainting methods dealing with missed wavelet or Fourier coefficients can be considered as *transformed domain* methods; they differ from the standard manipulation of intensity values in the imaging domain. In order to highlight inpainting approaches in a transformed domain, as they are transverse to both local and non-local methods we add a star $*$ in Tables 2.1 and 3.1. Notice that deep learning approaches can also be seen as methods belonging to the transformed domain, since most tasks are based on learning the underlying feature space.

Note on Evaluation

In general, there exist no unique solutions to the inpainting problem. Therefore evaluation of the effectiveness of each approach is purely descriptive. This is the reason why, in the case of art restoration, it is important to collaborate with specialised restorers in order to code methods that are close to their needs. It is important to highlight the fact that very often a damaged region causes a *distraction* when viewing images. Therefore, the general principle of inpainting can be translated as finding a good retouching that minimises the disruption to a general audience. If any methodology results in an artefact that does not affect the overall quality of the image and can still improve upon the input, then the artefact is recognised as a *feature* of that particular approach.

1.3 The Gestalt Laws

All inpainting methods have been inspired, explicitly or implicitly, by the laws of the *Gestalt* psychology theory, a German word translated as "form" or "shape". In this section we summarise the key principles of the theory, which are extensively described in Metzger (1975), and how they relate to the inpainting problem. We refer the reader to Desolneux et al. (2008, Chapter 2) for more technical insights.

The Gestalt theory was developed in 1923 at the Berlin School of Experimental Psychology by Max Wertheimer (1880–1943), Kurt Koffka (1886–1941) and Wolfgang Köhler (1887–1967): see Wertheimer (1923), Koffka (1935) and Köhler (1929). The idea of the Gestalt theory can be summarised in the words of K. Koffka, "the whole is something other than the sum of its parts", claiming that the human mind is able to recognise an object thanks to the priority of the total perception of the object over the perception of its individual parts. The Gestalt theory is based on *grouping* and *governing* laws: the former can be applied independently and recursively to the "atom elements", or previously formed objects, to build a total image (*gestalt*); the latter are principles dealing with the collaboration and conflictions of the elements. The local to global grouping laws are also described in the work of Kanizsa (1980) and are called there *primary processes* (they are also called *basic grouping principles* or *local principles*). These laws are very close to the list introduced by M. Wertheimer and are described as follows.

- **Proximity**: Objects with proximity are close enough to be grouped together (this can be related to the mathematical properties of *connected sets* or *clusters*).
- **Similarity**: Objects sharing features (e.g. colours or texture) are said to have similarity and are grouped together.
- **Continuity of direction**: Objects aligned with others are said to have continuity of direction and are grouped together.
- **Amodal completion**: In amodal completion, relatable[2] curves (curves that can be related) interrupted by the creation of T-junctions are grouped together as part of the same object (*principle of good continuation*).
- **Closure**: Closure occurs when a figure is perceived as complete and unified even when parts of it are missing or incomplete.
- **Constant width**: Two parallel curves, which can be said to be of constant width, identify boundaries of the same object.
- **Tendency to convexity**: A convex curve (not necessarily closed) identifies the boundary of a convex object.
- **Symmetry**: Objects symmetric around a central point belong to the same group.
- **Common motion** or **fate**: Objects moving along the same smooth path or behaving with a similar trend are said to have a common motion or fate and can be grouped together.
- **Past experience**: Past experience influences our visual perception and tendency to group objects.

As mentioned in Desolneux et al. (2008), *global gestalt* and *partial gestalt* are the results of different synthesis processes: for example, a square on its own is *global*, being the result of recursive *partial* groupings, e.g. edges and corners. On top of the basic laws, the Gestalt

[2] The notion of *relatability* was introduced by Kellman and Shipley (1991) and was proved experimentally in Rubin (2001) to play a global role in cueing the first occlusion-perception stages; see Oliver et al. (2016) for a computational model for amodal completion based on relatability.

theory requires the *global principles* of *inheritance*, describing how the parts share the features of the whole, and *Prägnanz*, a German word meaning "pithiness" and describing perceptual cognition on the assumption that the human brain acquires information in a regular, orderly, symmetrical and simple way. In the Gestalt theory the independent basic laws can collaborate or compete, thus generating *conflicts* or *masking* phenomena and possibly leading to incompatible objects, especially when no law is strong enough to overcome the others. Kanizsa (1979) also contributed, with the *connectivity principle*, meaning that human perception tends to prefer disjoint aligned objects to be connected; see Chan and Shen (2001b, Figures 1 and 2).

In more recent years, Gestalt theory has impacted on the inpainting problem described by the mathematical and computer vision research communities. Indeed, the results of inpainting models are visually evaluated by their perceived coherence with respect to the intact part of the image; this is called the *principle of visual reconstruction* in Gombrich (2006).

1.4 Mathematical Interpretation of Visual Perception

Over the recent years, a different class of approaches to the inpainting problem has gained attention, owing to its connections with the neurophysiology of vision (Citti and Sarti, 2006; Duits and Franken, 2010; Bekkers et al., 2014; Barbieri et al., 2014; Citti and Sarti, 2014; Sarti and Citti, 2014; Prandi and Gauthier, 2018; Franceschiello et al., 2017; Boscain et al., 2018; Bertalmío et al., 2019; Bertalmío et al., 2020; Smets et al., 2020; Bertalmío et al., 2021) and the perception of objects described by the Gestalt theory in Section 1.3. This class of methods, deserving a separate discussion in this section, takes the name of *cortical-inspired* approaches; this stresses their connections with the hypercolumn architecture of the first layer of the visual cortex (V1), the first protagonist of visual perception tautology. As observed by Hubel and Wiesel (1968) in their Nobel-prize-winning work, discovery neurons in V1 are sensitive to both the spatial location and the local orientation of the visual stimulus observed. From a mathematical point of view, this discovery intrinsically changes the classical modelling of images and inpainting operators.

Namely, while standard variational and partial differential equation (PDE) approaches (see Chapter 2) regard images as discrete or continuous objects defined on bidimensional domains Ω of $\mathbb{R}^2$, so that the local information is all contained at point $\boldsymbol{x} = (x, y)$, cortical-inspired models enlarge the space to an extra dimension, so that the cortical domain becomes $\mathcal{M} := \Omega \times [0, \pi)$ and the local independent variables are thus the pair $(\boldsymbol{x}, \theta(\boldsymbol{x}))$, where θ denotes the local orientation at point $\boldsymbol{x}$. The space of positions and orientations $\mathcal{M}$ can thus be considered as a reference space to which images can be "lifted" appropriately using tools from harmonic analysis and wavelet theory (Bekkers et al., 2014; Prandi and Gauthier, 2018). The space $\mathcal{M}$ can now be used as a reference framework to define new functional spaces, variational models and evolutionary PDEs which can be applied, for instance, to solve the inpainting problem. We remark that the associated metric and induced geometry in the space $\mathcal{M}$ are profoundly different from the standard Euclidean case, which changes both qualitatively and quantitatively the image reconstruction properties of these models. The fact that at each point $(\boldsymbol{x}, \theta) \in \mathcal{M}$ the local orientation information is encoded in the variable θ, indeed makes possible the definition of a new, strongly anisotropic, (sub-Riemannian) metric which can be used to define linear and nonlinear diffusion operators (such as respectively sub-Riemannian Laplacians, Boscain et al., 2018, or total-variation and mean-curvature flows, Citti et al., 2016 respectively) promoting at each point diffusion along the local direction

$\theta(\boldsymbol{x})$ only. Notice that this framework is then deeply different from the classical Riemannian framework, where directional information is typically encoded either by means of higher-order information or by a preliminary local-orientation estimation performed by means of structure tensors (Weickert, 1998).

The connections between the architecture of V1 and its sub-Riemannian mathematical modelling have been recently investigated in a series of papers (Bertalmío et al., 2019; Bertalmío et al., 2020, 2021). In these papers such cortical-inspired modelling has been used to extend well-known reference models describing neuronal interactions in V1, classically known in the literature as Wilson–Cowan equations. These models were first proposed in Wilson and Cowan (1972) to describe the evolution of a local neuronal state at a given instant $t > 0$ and local coordinate $(\boldsymbol{x}, \theta(\boldsymbol{x}))$ in the space of position and orientations $\mathcal{M}$. They have been widely applied to a variety of studies on physical transmission, diffusion and interaction phenomena of stimuli in the visual cortex over the last 40 years (see e.g. Chow and Karimipanah, 2020 for a review).

An extension of these models favouring *local histogram equalisation* (LHE) was introduced first in the context of image processing (Bertalmío et al., 2007) and has been recently connected with the description of visual perception bias (illusions) on the basis of both local brightness (contrast) and local orientation changes (Bertalmío, 2014; Bertalmío et al., 2019; Bertalmío et al., 2020). Interestingly, it is possible to show mathematically that the standard Wilson–Cowan dynamics do not obey *any* variational principle (Bertalmío et al., 2020, 2021); thus they form a “suboptimal” model as they conflict with the principle of redundancy minimisation or *efficient coding* popularly used in the study of the neurophysiology of vision (Olshausen and Field, 2000). On the other hand, the LHE variation of the Wilson–Cowan model can indeed be shown to rely on a variational structure, whose energy functional models the local neuronal interactions in the image as well as attachment to the given stimulus and local average information. The existence of a variational counterpart of the LHE model is not only interesting from a mathematical point of view, but also because it is an important characteristic which relates to the well-known notion of *efficient representation principle* introduced by Attneave (1954) and Barlow (2012). This principle states that neural responses aim to optimise the available biological resources by adapting to the statistics of the images that the individual typically encounters, so that visual information can be encoded in the most efficient way. In vision science, the efficient representation principle has correctly predicted a number of neural processing aspects and phenomena such as photoreceptor-response histogram equalisation and the fact that the receptive fields of cortical cells have a Gabor-function form (Atick, 1992; Daugman, 1985; Olshausen and Field, 2000).

It is thus natural to combine the cortical-inspired sub-Riemannian framework described above with Wilson–Cowan and LHE modelling so as to provide a fully bio-inspired mathematical framework for studying visual perception. This modelling was recently applied by Bertalmío et al. (2020, 2021) to the study of brightness- and contrast-dependent visual illusions. Their work showed that the existence or lack of an underlying variational principle is crucial for the possibility or impossibility of reproducing the visual perception bias generated by the visual system when it is exposed to illusion-inducing stimuli. In particular, Bertalmío et al. (2021) showed that, as far as orientation-dependent illusions are concerned, both inpainting-type and a perception-type behaviours can be reproduced by these models,

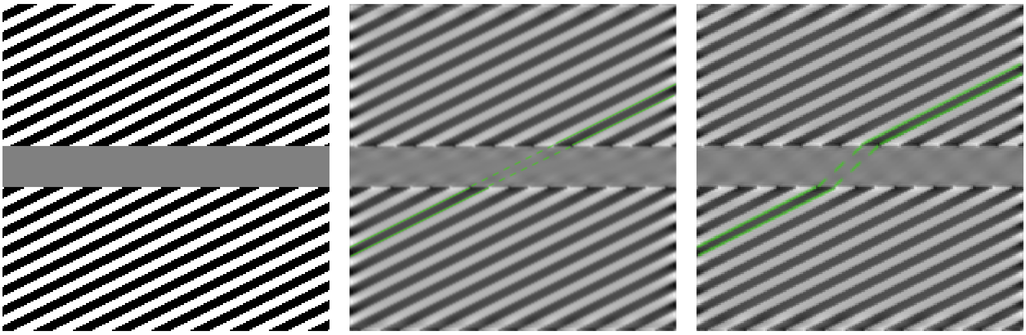


Figure 1.2 (a) A variation of the Poggendorff illusion (Weintraub and Krantz, 1971) for a grating. The grey central surface induces a misalignment of the background lines. The completion inside the middle grey bar can be (b) geometric (inpainting-type) or (c) illusory (perception-type). Image courtesy of Luca Bertalmío et al. (2021).

depending on the choice of model parameters, thus suggesting the existence of parameter thresholds (such as the radius of the neighbourhood used to model local interactions) which separate collinear perception from falsely aligned perception; see Figure 1.2.

1.5 The Fitzwilliam Museum Dataset

The dataset for this monograph has been drawn from the work of two interrelated research projects at the Fitzwilliam Museum, namely Cambridge Illuminations[3] (CI) and Manuscript Illumination: Non-Invasive Analysis, Research and Expertise[4] (MINIARE). The first project aims to research and document 4000 western illuminated manuscripts dating from the sixth to the sixteenth centuries in the Fitzwilliam and Cambridge colleges, and to publish them in print and online. The second employs cutting-edge non-invasive analytical techniques such as near-infrared imaging, spectroscopic analyses and optical microscopy to identify the materials and techniques used by manuscript illuminators.[5] Both projects have imaging at their heart. In the case of CI, images illustrate and evidence the research findings outlined in the published descriptions, while, for MINIARE, technical images – e.g. ultraviolet and infrared images, elemental maps as well as micrographs – are vital tools in the analytical protocol and investigative armoury of the project. The dataset is therefore immensely rich, but there are further reasons – historic and ethical – for selecting images of medieval manuscripts as the focus of this mathematical enquiry. It has been argued that illuminated manuscripts contain between their covers the largest and best preserved repository of medieval painting in existence, less subject to destruction or damage by natural forces or ideological violence than paintings and sculpture (Panayotova, 2016). Moreover, where such manuscripts have been damaged, the thinness of their paint layers and their small scale means that current conservation practice tends to not retouch areas of loss – in contrast with work on panel and

[3] www.fitzmuseum.cam.ac.uk/research/cambridgeilluminations.
[4] www.fitzmuseum.cam.ac.uk/research/miniare.
[5] The full analytical protocol is described in detail in Panayotova and Ricciardi (2016); an up-to-date survey of analytical methods for working with manuscripts is Ricciardi and Patterson (2020).

easel paintings. Indeed, privileging the authenticity of the work over the subjective judgement of the conservator means that damage – in so far as it reveals aspects of an object's construction and biography - is evaluated and stabilised, but not reversed or concealed (Cheese and Honold, 2016). In line with the non-invasive approach of MINIARE and the non-interventive approach of conservators, mathematical tools offer a new way of working with images rather than objects, to reverse damages virtually and open up a conversation with curators, conservators and the public about the processes and problems of restoration. They also offer a "reliable surrogate" for the research and teaching of medieval manuscripts (Calatroni et al., 2018).

A detailed description of manuscripts that have been virtually restored can be found in the Appendix.

Taxonomy of Damage

Though they are relatively well preserved, many manuscript illuminations have suffered various forms of damage over their long histories. Broadly speaking, damage can be classified as either accidental or intentional, though the two categories may overlap. Accidental damage includes the inevitable changes brought about by the chemical process of ageing or exposure to environmental change, such as colour change or craquelure (cracking of the paint surface). It also includes mishaps and disaster, such as damage by fire or water. Intentional damage, perhaps more frequently encountered than we might expect, includes deliberately inflicted removals (scratches, scrapes, erasures) and additions (over-paintings, additions) for reasons of censorship, ideological commitment, the harvesting of rare materials, re-purposing of materials, the signalling of a change of ownership or commercial gain. In both cases, the geometry of the damaged domain is varied in nature, ranging from tiny lines to changes which destroy the legibility of the image; see Figure 1.3.

1.6 Plan of the Book

The inpainting methods described in this monograph are grouped into thematic chapters, according to their mathematical approach. These can be summarised as follows.

- In Chapter 2 we introduce the foundation of inpainting strategies by describing *local* methods; these started in early 1990 and became increasingly popular around the year 2000. These approaches are based on the local diffusion of imaging intensity values into the damaged domain by means of partial differential equations and variational techniques that can preserve the coherent geometric structures and produce a smooth version of the expected fill-in.
- In Chapter 3 we detail *non-local* inpainting methods, also called *exemplar-based*, which were predominant in the first decade of the 2000s. These approaches are based on finding good examples (or patches), within the same image, to be copied and pasted onto the damaged domain in order to reconstruct the textured information.
- In Chapter 4 we illustrate inpainting approaches that use recently developed techniques in *machine* and *deep learning*, which appeared in the literature from 2010 onwards. These methods aim to leverage the relevant information not only within a single image but also on large datasets.
- In Chapter 5 we describe inpainting approaches specifically conceived within the context of cultural heritage data, and their challenges. We illustrate specific cases of effective inpainting strategies in real scenarios.

(a) Craquelure (accidental)

(b) Pigment degradation (accidental)

(c) Water damage (accidental)

(d) Over-painting – censorship (intentional)

(e) Scratches (intentional)

(f) Over-painting – re-purposing (intentional)

Figure 1.3 Types of damage and their taxonomy. Details from: (a) MS Marlay 18.iii, (b) MS 39-1950, fol. 147r, (c) MS McClean 173, fol. 51r, (d) MS 159, fol. 4r, (e) MS 330.iii and (f) MS 62, fol. 16v. All images © Fitzwilliam Museum, Cambridge, UK.

At the end of Chapters 2, 3 and 4 we compare inpainting algorithms for which we have found a working implementation; these were tested on reference images we identified in our dataset. The format of the references list provides a glimpse of the time frame of the research work described in this book and is summarised in Tables 2.1, 3.1 and 4.1.

Note on Experiments

In this book we discuss applications of different inpainting methods to the specific cultural-heritage-imaging dataset described in Section 1.5. Some of these methods were developed for the virtual retouching of natural images and with the goal of being included in photo/video-editing software. Very often, people working in cultural heritage conservation deal with damaged objects that in fact are rarely restored, as they are too fragile to be handled. Therefore, virtual restoration seems a promising alternative way for preserving such artworks. None of the inpainting methods tested preprocess the damaged images, i.e. they do not require any initialisation of the inpainting domain, which is indicated by the white area of the masks provided for each test. Different initialisation choices are certainly possible, e.g. using a random initialisation of the damaged area, but these are avoided in this book in order not to bias the final result. Throughout the book it is assumed that the inpainting domain is manually defined by the end-user, unless explicitly stated otherwise. Our experiments, with no quantitative evaluation owing to the lack of ground truth solutions, were performed in Chapters 2, 3 and 5 on a standard MacBook Pro (13-inch, 2019), 2.4 GHz Intel Core i5 quad-core, 16 GB 2133 MHz LPDDR3, while in Chapter 4 we used Ubuntu 16.04 with NVIDIA Quadro P6000 GPU.