Master Erasmus Mundus in
Color in Informatics and Media Technology (CIMET)

Digital Inpainting for Artwork Restoration: Algorithms and Evaluation

Master Thesis Report

Presented by

Alexandra Ioana Oncu Feier

and defended at

Gjøvik University College

Academic Supervisor(s): Prof. Jon Yngve Hardeberg
PhD student Ferdinand Deger

Jury Committee: Prof. Alain Trémeau
Prof. Philippe Colantoni

Digital Inpainting for Artwork Restoration: Algorithms and Evaluation

Alexandra Ioana Oncu Feier

2012/07/15

Abstract

Digital image inpainting refers to a technique used for filling in the missing or corrupted regions of an image using information from the surrounding area. Inpainting techniques have found widespread use in applications that include, but are not limited to error recovery, red-eye removal, multimedia editing. This thesis will discuss inpainting techniques in the context of digital artwork restoration.

Due to an extensive research in the field of digital inpainting and to rapid advances in technology, new and improved inpainting techniques are continuously proposed. Improved digital image inpainting algorithms could provide substantial support for future artwork restoration. However, as the latter is characterised by a high demand in quality, methods for an accurate evaluation of the performance of inpainting algorithms are needed. A literature review showed however, that currently, there is an acknowledged lack of quantitative metrics for image inpainting evaluation. In this thesis the performance of eight representative inpainting algorithms will first be evaluated by means of a psychophysical experiment. Based on the obtained perceptual data, a ranking of the algorithms will be established, that confirms that exemplar based methods generally outperform Partial Differential Equation based methods.

Two novel inpainting quality metrics, proposed in this thesis, eight general image quality metrics and four quality metrics specially developed for inpainting assessment will then be evaluated against human observers. Two types of evaluation will be carried out; one investigates the performance of the metrics over the entire image database considered, and the other assesses the correlation of the metrics for individual images. Results show that none of the considered metrics can adequately predict inpainting quality over the entire image database, and that the performance of the metrics is image-dependent. The two newly proposed metrics outperform some of the existent inpainting quality metrics for the first type of evaluation. However, additional work is needed to find metrics that correlate well with the percept.

Aknowledgements

I would like to take this opportunity to express my sincere gratitude towards my main supervisor, Prof. Jon Yngve Hardeberg from Gjovik University College, who introduced me to the field of digital inpainting. His guidance and encouragement to aim higher and do better have made possible the completion of this work. Next, I would like to thank my second supervisor, PhD student Ferdinand Deger, for his help, assistance and constructive comments.

An unmeasurable amount of thanks goes to Prof. Alain Trémeau and Prof. Philippe Colantoni at Université Jean Monnet for their inspiring questions during my Master Thesis defence, allowing me to discover new directions for further research.

Many thanks goes to PhD Marius Pedersen, member of The Norwegian Colour and Visual Computing Laboratory, for his advice, encouragement and support.

Thanks also goes to all the participants to my experiment that made this study possible. Thank you very much for your time and participation.

I would also like to thank to my CIMET family, who has always been there for me, with a good word and wise counsel. Special thanks goes to Kristina, Marialena, Laksmita and Vamsi, my friends that made days spent together feel like seconds.

Above all, I thank my family. I would not be here if it was not for them. Thank you for always supporting me and always being there for me.

Contents

List of Figures

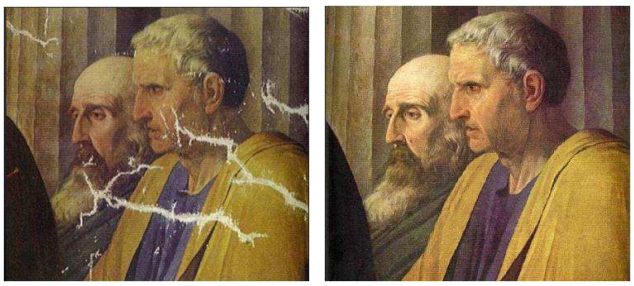
List of Tables

1 Introduction

1.1 Background

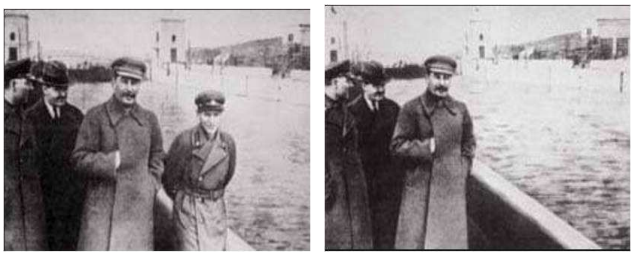
Early artistic creation could not have survived over the years if it was not for the image and art restoration artists, that brought old or damaged paintings back to the original or a close-to-the-original state. Artwork dating back to the Middle Ages was already in need of restoration towards the end of this period of European history, as noted by Emile in [10]. At the beginning of the Renaissance era, the main concern regarding medieval artwork was to bring it *up to date*, which many times required reconstructing missing or deteriorated parts of a canvas, by filling in existent gaps [10] with visually pleasing content. Thus, it can be said that almost as old as art itself is the practice of making modifications to paintings, in such a way that if an observer would look at the modified work of art, without knowing the original, he wouldn't be able to perceive any alteration. This practice is traditionally carried out by restoration experts, such as museum art restorers, and it is commonly known as *retouching* or *inpainting*. Its desired outcome is to make a damaged artwork more discernible, while restoring its unity. An example of a painting that has been the subject of a manually restoration process, that involved inpainting, is given in Figure 1.

Advances of technology made possible the transition from traditional manual retouching methods to digital techniques. In the digital domain, *inpainting* was initially known as *error concealment* and referred to a specialized technique used in the field of telecommunications, in order to fill-in image blocks that were lost during data transmission. As it grew to become a topic of interest in a more general setting than just data transmission, inpainting came to be known as the process performing *image disocclusion*. When considering a damaged image, the affected region, further on referred to as *the gap*, was considered to be an occluding object that needed to be removed, and the image underneath it, the desired restoration result. Other popular terms, that can be found in the related literature, as referring to inpainting, comprise *image completion* and *image fill-in*.



(a) (b)

Figure 1: *Detail of the fresco* Cornelia, mother of the Gracchi *by J. Suvee (Louvre). (a) Painting in an advanced state of deterioration. (b) Manually restored painting. Source: [11]*



(a) (b)

Figure 2: *Inpainting performed by professional artists. Example of the "airbrushing" technique in a photograph. Source: [10]*

The term *digital inpainting* was introduced by Bertalmio *et al.* [5] only later, in the year 2000 .

In the field of artwork restoration, inpainting algorithms can be employed for digital restoration purposes, by reversing the damage (i.e. torn canvas, scratches, stains) in a painting converted to a digital form. Digital inpainting provides the means to restore a painting without touching it, and by doing so, it removes the threat of doing permanent changes with potentially damaging consequences for valuable work of art. Moreover, improved digital inpainting techniques could substitute the time-consuming and costly process of manual restoration.

Extensive research has been carried out in the field of digital inpainting, such that currently, a large number and variety of algorithms exist. Depending on the approach they take towards image completion, digital inpainting algorithms can be grouped into two main categories. Partial differential equation (PDE) based algorithms, like the ones proposed in [11], [7], [4], [6], fill in missing regions in an image by extending lines of equal luminance values, from the source region into the target region, via diffusion. The main drawback of this type of inpainting algorithms consists of introducing post-inpainting blur artifacts, that become more visible when larger areas are inpainted. The second category comprises exemplar-based inpainting algorithms [2] [3] [8]. Methods in this category try to overcome the drawback exhibited by PDE based techniques, by reconstructing large missing image regions from sample textures. In addition to the latter two, another separate category can be considered, consisting of algorithms with high performance in terms of running time [9] [8],[4]. The latter is an important aspect to be taken into consideration when developing commercial software that involves user interaction.

Due to extensive research and rapid technology advancements, new digital image inpainting methods continuously emerge, trying to achieve better performance in terms of quality of the resultant images. Consequently, an accurate method for evaluating the inpainting quality is required, in order to establish if newly developed methods increase the quality of the inpainted images, and can thus meet the high standards imposed in the field of artwork restoration.

1.2 Motivation

Notwithstanding the potentially different approaches, the goal of any digital image inpainting algorithm is to reconstruct the missing or damaged regions in a visually plausible

way. Thus, inpainted images that are significantly different from the original, undamaged images (i.e. reference images), but exhibit visual pleasantness are approved outcomes of the inpainting process. Adding to the latter the fact that a reference image might not always be available for comparison, explains why inpainting quality evaluation is still considered a challenging task, that has been only narrowly researched in the recent years.

Qualitative human comparisons are currently and frequently used in order to quantify the quality of the inpainting. However, despite the accurate results provided, the latter is a resource and time demanding evaluation method. In an attempt to address this issue Mahalingam [12] and Ardis *et al.* [13] propose an objective method for inpainting quality evaluation by using visual-saliency based metrics. However, these metrics are not commonly used by researchers to assess the performance of digital image inpainting techniques.

As opposed to the scarce research in the field of inpainting quality evaluation, a considerable number of metrics for general IQ assessment have been proposed in the related literature. However, the use of this type of metrics for quantifying the quality of inpainted images has been researched only to a limited extent. IQ metrics simulating the human visual system (HVS) and taking into account structural information in an image might be useful for image inpainting quality evaluation.

1.3 Thesis objective

The goal of this research is to qualitatively evaluate the performance of eight representative digital inpainting algorithms [11] [7] [4], [6] [2] [3] [8] [9] and to develop, use and evaluate objective methods for quality assessment in the context of digital inpainting for artwork.

In the above, *develop* refers to introducing new quality metrics, based on existent approaches, so that to achieve a better correlation between calculated quality and perceived quality. Furthermore, *use* refers to applying various metrics for measuring the quality of an inpainted image. Finally, the term *evaluate* is used to denote the assessment of quality metrics by validation against perceptual data.

1.4 Research methods

The qualitative evaluation of the eight selected digital inpainting algorithms will be accomplished by conducting a psychophysical experiment. The obtained perceptual data will be statistically analysed and then used as a basis for establishing a ranking of the algorithms considered.

The evaluation of quality metrics will use quantitative methods for numerical analysis of data, and will be accomplished by comparing the results of the metrics against the results obtained from the psychophysical experiment. The correlation between objective scores and subjective scores will be used as an indicator of the metrics performance.

Thus, the findings of the current research are based on both theoretical (i.e. literature review) and practical (i.e. psychophysical experiment) work.

1.5 Paper submitted to workshop

Based on the findings of the current research, a paper was submitted to the 1st Workshop on Computer VISion for ART Analysis (In conjunction with ECCV 2012). The submitted

paper will be enclosed separately, as an Appendix to this thesis.

1.6 Thesis outline

The remainder of this thesis is organized as follows:

Chapter 2 provides a detailed literature survey of existent digital image inpainting methods. Based on the approach they take, three categories of algorithms will be discussed. The first category refers to PDE based inpainting algorithms, the second group consists of exemplar based algorithms, while the last category is represented by fast inpainting algorithms.

Chapter 3 discusses relevant work on evaluating perceptual quality. Metric specifically designed for inpainting quality evaluation, as well as more general IQ metrics will be introduced in this Chapter.

Chapter 4 introduces two novel quality metrics for digital image inpainting evaluation. Based on visual-saliency maps and structural information in an image, these metrics try to achieve a better correlation with perceived quality than other existent inpainting quality metrics.

Chapter 5 explains the methodology for evaluation of quality metrics. The inpainting algorithms and quality metrics selected for evaluation will be listed along with their main characteristics and the steps for creating the image database used as basis for the evaluation will be included in this Chapter.

Chapter 6 discusses the psychophysical experiment for subjective assessment of inpainting quality. In the first part of this Chapter, the experimental setup will be presented, including references to the test images used, the experimental method chosen, the viewing conditions during the experiment and finally, instructions given to the observers. The second part of this Chapter will present the analysis of perceptual quality of image inpainting algorithm.

Chapter 7 investigates the performance of quality metrics in predicting perceived quality. Two types of evaluation are carried out: one that considers all the images in the database and reports on the overall performance of the metrics, and one that refers to the evaluation of the metrics for each image in the database, individually.

Chapter 8 summarizes the findings of the current research and points out several lines of future research.

The bibliography and appendices are given at last.

2 Digital Inpainting

In this chapter an extensive survey of previous work related to digital inpainting techniques will be presented. The survey includes scientific techniques used mainly in the research environment, but it also includes inpainting methods that were incorporated in commercial software (i.e. Adobe Photoshop CS5.1, G'MIC plug-in to the GNU software GIMP). The decision of including the latter originated from the intent of showing what can be done by non-expert users that are not familiar with the inpainting concept.

The pioneering works by Mumford et al. [14], Masnou and Morel [15], Caselles et al. [16] and Bertalmio *et al.* [17] in the field of digital inpainting were followed by several different approaches that aimed at achieving the same task, of seamlessly filling in any missing regions in an image [**?**]. Despite the significant differences between the existent techniques, in [5] the authors manage to draw the main directions that each method must follow. Accordingly, the four steps that make up the underlying methodology [5]of any inpainting algorithm are given as:

1. "The global picture determines how to fill in the gap, the purpose of inpainting being to restore the unity of the work;
2. The structure of the area surrounding the hole is continued into the gap, contour lines are drawn via the prolongation of those arriving at the boundary of the inpainting domain ;
3. The different regions inside the inpainting domain, as defined by the contour lines, are filled with color, matching those of the boundary of the region to be inpainted;
4. The small details are painted (e.g. little white spots on an otherwise uniformly blue sky): in other words,*texture* is added."

In the related literature, many research papers propose a rough classification of the inpainting methods, dividing them into two categories: methods that are based on partial differential equation (PDE) and examplar based methods. However, this is a too general classification that would not see all the techniques discussed in this thesis properly fit into a category. Thus, instead of it, the following classification, including three different categories will be used:

- Partial Differential Equation (PDE) based inpainting algorithms
- Exemplar based inpainting algorithms
- Fast digital inpainting algorithms

Notation

Before proceeding with the inpainting methods review, a note must be made on the notation that will be used. The omega symbol, Ω, will denote the inpainting domain, also referred to as the hole in the image, while $\delta\Omega$ will stand for its boundary. The source region, representing the area not covered by the mask, or the area complementary to the gap, will be denoted by Φ This notation was first used in [5] and adopted by the

Figure 3: *Unsuccessful inpainting in terms of information propagation direction. (left) Original image, with the inpainting domain represented in white. (right) Inpainted image, that exhibits a wrong direction of propagation. Source: [5]*

following works.

2.1 Partial Differential Equation (PDE) Based Inpainting

Before presenting the inpainting methods falling under this category, a short definition of a Partial Differential Equation (PDE) will be given. A differential equation contains one or more variables, relating the values of the function itself and its derivatives of various orders. Consequently, a PDE is a differential equation that uses partial derivatives - that is a derivative with respect to one variable, with all other variables held constant. An example of a PDE for the function $u(x_1 \dots x_n)$ is:

$$F(x_1, ..., x_n, u, \frac{\delta u}{\delta x_1}, ..., \frac{\delta u}{\delta x_n}, \frac{\delta^2 u}{\delta x_1 \delta x_1}, ..., \frac{\delta^2 u}{\delta x_1 \delta x_n}, ...) = 0. \tag{2.1}$$

2.1.1 Basic PDE based inpainting method

The method that paved the way for modern digital inpainting was a PDE based method, proposed by Bertalmio *et al.* in [5]. Inspired by manual inpainting concepts, this method tries to actually translate the four rules used in manual inpainting and mentioned in the beginning of this section into a mathematical and algorithmic language. The underlying idea of this algorithm is to propagate the geometric and photometric information arriving at the boundary of the occluded area, $\delta\Omega$, into the area Ω itself. This is achieved by extending the *isophote* lines (i.e.contours of equal luminance value in an image, edges, lines of equal gray values) that try to capture the direction of minimal change. An important observation is that the angle of *arrival* of the isophotes at the boundary must be preserved for a successful inpainting (see Figure 3). Thus, the information propagation is carried out in the direction of minimal change, and in addition from $\delta\Omega$ inward, in a smooth way. Essentially an iterative process, the algorithm proposed by Bertalmio *et al.* in [5] constructs a set of images $I_0, I_1 \dots I_n$, where I_0 is the input image, $n \to \infty$ and $\lim_{n\to\infty} I_n = I_R$. Thus, I_R is the output image of the algorithm and thus represents the inpainting result. Each image in the set is an improvement on the previous image. The information propagation process just described can thus be given by the following equation:

$$I^{n+1}(i,j) = I^n(i,j)\Delta t I_t^n(i,j), \forall (i,j) \in \Omega, \tag{2.2}$$

where the superscript n denotes the inpainting iteration *time*, (i,j) are the coordinates of the current pixel, Δt is the rate of change, or improvement and I_t^n is the update of the image I^n. The multiplication of the latter two terms gives the *improvement* observed in image I^{n+1} as compared to the previous version of the image, I^n.

With a user-defined value for Δt, the only unknown parameter in Eq. 2.2 is now the update factor, I_t. Determining the value of the aforementioned parameter represents the

most important part of the method under discussion, setting the PDE-based nature of the algorithm. The latter affirmation builds upon the fact that the update factor must capture both the information that needs to be propagated, and the propagation direction [**?**]. Taking into consideration all of the above, in [5] the authors propose the following equation (the space coordinates (i,j) and the inpainting *time* n have been omitted for simplicity):

$$I_t = \nabla L \cdot \nabla^{\perp} I, where L = \Delta I, \tag{2.3}$$

where ∇L represents a measure of the change in the propagated information, while the second part of Eq. 2.3, $\nabla^{\perp} I$ gives the isophotes direction. As mentioned earlier, this algorithm tries to achieve a smooth propagation, thus for expressing the information being propagated, the authors propose the usage of an image (2D) smoothness estimator, given as a discrete implementation of the Laplacian: $L^n(i,j) := I^n_{xx}(i,j) + I^n_{yy}(i,j)$. In the latter equation the subscripts indicate second derivatives. The idea behind the calculation of the isophotes direction is mainly based on obtaining the gradient vector $\nabla I^n(i,j)$ at pixel location (i,j). As the latter indicates the direction of the largest spatial change, a rotation by 90 degrees of the gradient is needed in order to obtain the direction of least spatial change. For more detailed specifications on the implementation the reader is referred to [5].

Furthermore, the inpainting process described above alternates with an anisotropic diffusion process, whose main role is to provide a correct update/evolution of the directional field used when filling in the hole, during the inpainting process. In order to achieve such a goal, the image restoration loop is built in such a way that after every few steps of inpainting, a few iterations of image diffusion are applied. In [5] the authors set as default values the following: 15 steps of inpainting, followed by 2 steps of diffusion, at an update speed $\Delta t = 0.1$. The total number of steps (both of inpainting and diffusion) can be pre-set by the user, but there is also the possibility of allowing the algorithm to stop when $I^{n+1} = I^n$ (however the authors didn't provide a specific convergence analysis to sustain the latter). In addition to the aforementioned steps, that affect only the information in the inpainting region, an anisotropic diffusion smoothing is applied to the whole image beforehand, in order to reduce the influence of noise when initially estimating the isophotes direction.

One of the alleged drawbacks of the inpainting technique proposed by Bertalmio *et al.* in [5] is represented by the impossibility of reproducing large textured regions, as the diffusion process causes blurring. Another potential drawback is that many inpainting iterations may be needed before the algorithm converges to reach a steady state ($I^{n+1} = I^n$), which can eventually lead to an infinite loop, if the user doesn't provide a total number of inpainting and diffusion steps.

2.1.2 Improved PDE based inpainting methods

Navier-Stokes, fluid dynamics and image and video inpainting

A subsequent work of Bertalmio *et al.* [17] refines the previous approach by resorting to classical fluid dynamics and the Navier-Stokes equations (describing the motion of fluid substances). In this context, the image intensity is the equivalent of the fluid's stream function, as the level lines in the image define the stream lines of the flow. Furthermore, the isophote direction is the equivalent to the fluid's velocity, while the smoothness of the image is comparable, in fluid dynamics terms, to the curl of the fluid's velocity, called

vorticity. Following this approach, inpainting resumes to solving a vorticity transport equation, instead of the transport equation given in 2.3. Such an inpainting method, based on fluid dynamics, has the great benefit that one can exploit the already well-elaborated history of fluid problems and thus can more easily find solutions to existent or arising issues [18].

Filling-in by joint interpolation of vector fields and grey levels

Ballester *et al.* [16] developed a method based on the ideas previously presented by Bertalmio *et al.* in [5], addressing the inpainting problem through the use of second order PDEs. The authors define a formal variational approach (which aims at finding functions that minimize the value of the quantities depending upon them) in order to achieve a simultaneous graylevel and gradient continuation. Consequently, the inpainting problem is actually split into two coupled variational formulations, one corresponding to the isophotes direction (i.e. gradient orientations) and the other corresponding to the gray levels, both of which are solved by using a gradient descent. This technique implies taking steps that are proportional to the negative of the gradient of a specific function, at the current point, in order to find that function's local minimum. Finally, the gradient descent flows obtained, corresponding to the gray-values and gradient orientations respectively, give two coupled second order PDEs that must be solved in order to finalize the inpainting process.

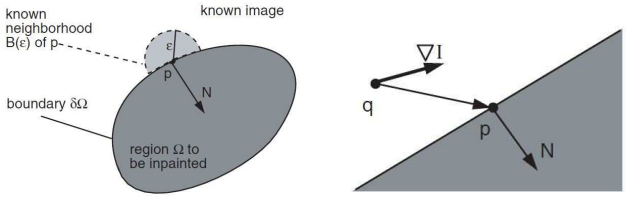
Vector-valued image regularization with PDEs

In [4], Tschumperle and Deriche introduce another PDE based method, which is referred to as vector valued image regularization with PDEs. The regularization of an image is the process through which that specific image is simplified in such a way that interesting features are preserved and unimportant data is removed. In this context, the unimportant data is considered to be *noise*. Given the latter definition of regularization, it becomes obvious that the man application, for which this algorithm was designed, is represented by image denoising. However, because of their capability to produce simplified representations of data, techniques based on image regularization are also useful when dealing with feature (i.e. edges) extraction. Taking advantage of this, in inpainting, the method proposed by Tschumperle and Deriche [4] successfully fills in the gap, by diffusing the boundary pixels until completion of the missing region, in a structure preserving way. The main problem with this approach comes from the fact that being still a PDE-based method, the inpainted region is subject to blur, which results in unsharp edges. Thus, it can be inferred that the technique introduced in [4] will fail in reproducing textured regions, but will successfully fill in small and narrow holes.

The vector-valued image regularization approach proposed by Tschumperle and Deriche and applied for inpainting, was the first high-quality inpainting method to be incorporated in a commercial, interactive image manipulation software [6], as the G'MIC plug-in to the GNU software GIMP. This plug-in actually comes as a sequel to the *GREYCstoration* software, which was exclusively dedicated to image denoising and regularization. Thus, the G'MIC plug-in provides all the features of *GREYCstoration*, to which it adds the inpainting capability.

An image inpainting technique based on the Fast Marching Method

A notable improvement of PDE based inpainting techniques was brought by Telea, with the method discussed in [7]. The proposed algorithm is based on the iterative inpainting method introduced by Bertalmio *et al.* in [5], but compared to it, Telea's algorithm



(a) (b)

Figure 4: *Inpainting algorithm principle as adopted by Telea in [7]. Source: [7]*

provides a faster, non-iterative (single-pass) solution to the inpainting problem. The improvement in speed is due to the fact that the value of each pixel in the gap is modified (i.e. inpainted) only once, thus excluding the main issue arising when using iterative techniques - possibly needing many iterations before converging.

The idea behind the algorithm presented in [7] is to fill in the gap in a fixed order, from the boundary of the missing region inwards, where the value of the pixel to be inpainted is determined by the values of already inpainted pixels (or other known pixel values from the source region, in the case of pixels close to the initial gap boundary). The latter describes an inpainting principle also suggested in [5], [9] and [19]. Letting the pixel to be inpainted (at a certain moment in time) be denoted by p, the area containing known image points around p is called the *neighbourhood of p* and is denoted by $B(\epsilon)$, where ϵ is the radius of $B(\epsilon)$, as shown in Figure 4a. The value of ϵ is set by the user, and it should be small. Having set a value for ϵ, and regarding the pixel p, the algorithm proceeds by considering a first order approximation of the image in point p. Denoting this approximation by $I_q(p)$, with q being a pixel in the known neighbourhood of p, and given the pixel value of point q, $I(q)$ and its gradient, denoted by $\nabla I(q)$, the following holds:

$$I_q(p) = I(q) + \nabla I(q)(p - q). \tag{2.4}$$

For a graphical representation of the variables involved in the above equation the reader is referred to Figure 4b. The next step in the algorithm proposed by Telea in [7] is the actual inpainting step. The value of the pixel p is obtained by summing the estimates $I_q(p)$ corresponding to the pixels in B_ϵ, weighted by a function *w(p,q)*. Consequently, the value of the inpainted pixel p is calculated as:

$$I(p) = \frac{\sum_{q \in B_\epsilon(p)} w(p,q)[I(q) + \nabla I(q)(p - q)]}{\sum_{q \in B_{\epsilon(p)}} w(p,q)}. \tag{2.5}$$

As noted by Telea in [7], the weighting function *w(p,q)* is designed so that it allows for the propagation of sharp image details, as well as smooth regions, into the gap, being defined as:

$$w(p,q) = dir(p,q) \cdot dst(p,q) \cdot lev(p,q), \tag{2.6}$$

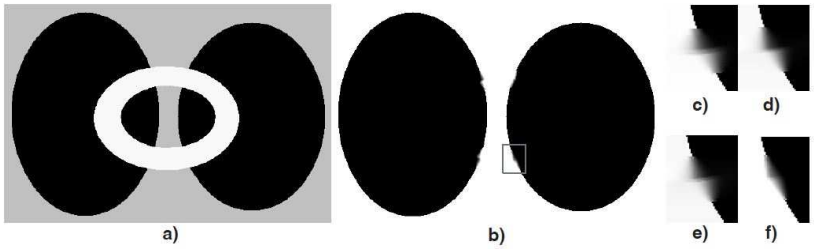


Figure 5: *Influence of the weight function proposed by Telea [7] on the inpainting results. (a) Original image with mask in white. (b) Inpainted image using all three components. Inpainting result with: (c) only* dir *component (d)* dir *and* dst *components (e)* dir *and* lev *components (f)* dir, dst *and* lev *components. Source: [7]*

where:

$$
\begin{aligned}
dir(p,q) &= \frac{p-q}{\|p-q\|} \cdot N(p) \\
dst(p,q) &= \frac{d_0^2}{\|p-q\|}^2 \\
lev(p,q) &= \frac{T_0}{1+|T(p)-T(q)|}.
\end{aligned}
\tag{2.7}
$$

In Eq. 2.7 above, *dir(p,q)* denotes the *directional* component used in order to guarantee a higher contribution of the pixels situated along, or close to the normal direction *N* of the image gradient. Furthermore, *dst(p,q)* denotes the *geometric distance* component, and as its name suggests, sets a higher contribution for the pixels that are geometrically closer to the pixel *p* than for the others. Finally, *lev(p,q)* is the *level set* distance component, making sure that the pixels closer to the isophote line through the considered pixel *p* have a higher contribution to the final value of *p* than the rest of the pixels. In [7],Telea specifies that in practice, the values for the reference distances d_0 and T_0 are set to be equal to 1. Also, an important observation made by the author refers to the importance of the each of the three components, depending on the size of the gap. Accordingly, for missing regions of up to 6 pixels, the *dst* and *lev* components have a weak effect on the resulting inpainted value, whereas for larger gaps, up to 12 pixels, including the two terms has proven to provide better results than those obtained by using only the *dir* component. The inpainting results obtained with various combinations of the three components are shown in Figure 5, where the inpainting mask is wider than 30 pixels. This example makes it clear that the best result is the one provided when all three components are used.

Up until this point, the way a single pixel is inpainted using Telea's method has been discussed. In order to fill in the whole missing region, an iterative process is required, applying the Eq. 2.5 to each of the pixels in the gap. When doing so, the initial boundary delimiting the gap will advance inside the gap, becoming thus smaller and smaller, until the whole missing region has been filled in. An important aspect of the inpainting process is related to the filling order, as it was also mentioned earlier. Establishing an increasing distance from the boundary $\delta\Omega$ inpainting order guarantees that the closest regions to the known pixels (contained within the source area) are filled in first, mimicking thus the technique employed in the process of manual inpainting [5]. In order to implement the latter, Telea uses the *Fast Marching Method* (FMM), which ensures that the pixels are filled

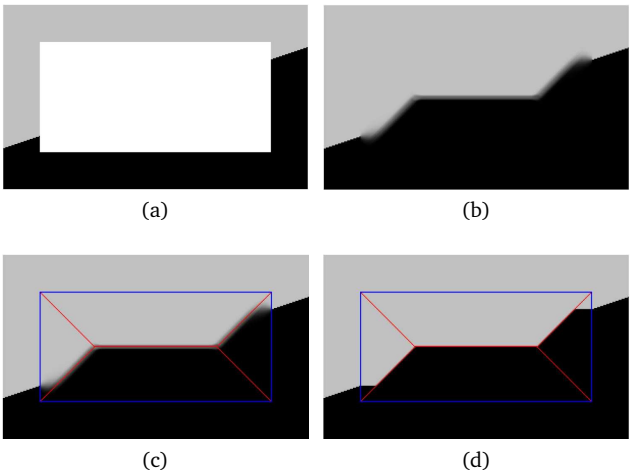


Figure 6: *Influence of the gap's geometry on the inpainting results obtained with the method proposed by Telea [7]. (a) Synthetic digital image with mask in white (b) Inpainting result for* $\epsilon = 6$px. *Note that the straight line is incorrectly reconstructed. (c) The inpainting order (increasing distance from the boundary) and the geometry of the inpainting domain determine the way the gap is filled in. The information propagation follows the structure of the gap's skeleton, given in red. (d) Same effect as in* c, *but less blur due to a smaller* ϵ *value. Source: [7]*

in in the correct order, based on their distance to the boundary (i.e. pixels situated near the known region are always filled in first). The FMM is a numerical method typically used for solving the following Eikonal equation:

$$|\nabla T| = 1 on \Omega, with T = 0 on \delta\Omega, \tag{2.8}$$

where T is the distance map of the pixels belonging to the inpainting domain Ω, to the boundary $\delta\Omega$. For more details on how to implement the *Fast Marching Method* the reader is referred to [7], which gives the complete pseudocode.

Similarly to the previously presented PDE based inpainting methods, the limitation of the algorithm introduced by Telea is given by the considerable amount of blur in the inpainted region, when the gap is wider than 10 pixels, as noted in [20]. Another drawback of this method is that the direction of propagation for the inpainting information is fully determined by the geometry of the gap Ω. The latter statement is exemplified in Figure 6, where the images on the second row provide a graphical explanation of the effect observed in Figure 6b.

Summary While there are many PDE based inpainting methods that have been developed, all of them are, in general, suitable only for filling in small, non-textured areas. The reason leading to the latter affirmation is that PDE based inpainting techniques make use only of local information, thus when the gap is large, or when it contains texture, the linear structure inside the gap can't be reconnected and blur artifacts can be noticed, resulting thus in a visual perception of the resultant inpainted region that is generally bad [21].

(a) (b)

Figure 7: *Filling large gaps in an image by using the method introduced by Criminisi* et al. *in [2]. (a) Original image (b) The area corresponding to the person in the foreground has been removed from the original image and replaced with information from the background. Source: [2]*

2.2 Exemplar Based Inpainting

As it was shown in Section 2.1, PDE based inpainting algorithms are not sufficient for faithfully reconstructing textured images, nor images with large missing areas. Thus, when inpainting is done with an image restauration purpose in mind, more complex techniques are required, as paintings are composed of both structures (i.e. primal sketches) and textures (i.e. regions with homogeneous patterns). Because of this characteristic of paintings (and natural images, in a more general manner), a technique that is strictly designed for texture synthesis will not perform well, either. Exemplar-based inpainting methods can overcome this drawback, being able to provide reasonably good quality results, even for large gaps, by combining the isophote driven inpainting with texture synthesis [22].

The idea behind examplar-based image completion refers to the use of a set of image blocks (or exemplars) that can be extracted either from the actual image that needs to be inpainted, or from another image that belongs to a set of representative images. However, it is most common, in the context of inpainting, to fill in the gap with the best matching parts of the same image [20]. The order in which the gap is filled in has a great importance, as it may contribute to the minimization of artifacts.

2.2.1 Basic exemplar based inpainting method

One of the algorithms frequently given as a reference when discussing examplar based inpainting methods is the one proposed by Criminisi *et al.* in [2], an algorithm that has been designed for removing large objects from an image, while filling in the gaps with visually plausible information from the background, as the authors state. An example of the algorithm's performance is given in Figure 7.

The method proposed by Criminisi *et al.* is based on two remarks. The first one states that exemplar based synthesis is enough for successfully filling in gaps, while preserving and extending the linear image structure (i.e. isophotes). That is because the very essence of exemplar based algorithms is to find patches belonging to the complementary area

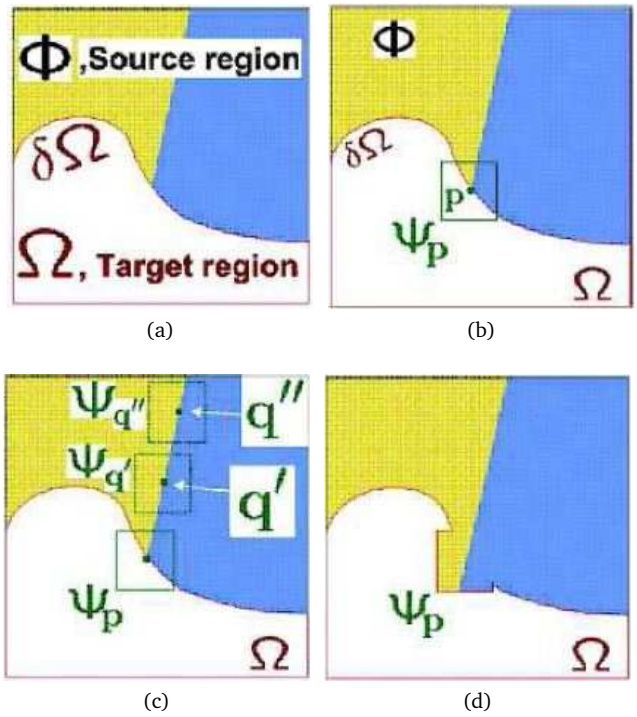


Figure 8: *Visualisation of the exemplar based inpainting process. (a) Original image showing the source and target regions as well as the boundary contour (b) Patch that was given the highest priority. (c) Candidate patches* $\Psi_{q'}$ *and* $\Psi_{q''}$ *(d) The patch* Ψ_p *is filled in with the best matching patch. Source: [2]*

of the gap, that contain similar information to the known parts of the patch that is to be filled. As an example, in Figure 8b the information in the upper right corner of the patch p will be used for finding similar patches in the rest of the image. The second remark is that the gap filling order is a critical aspect of the algorithm. The quality of the inpainted image greatly depends on the order in which patches are filled in. A priority based mechanism is proposed by the authors, estimating the importance of each of the pixels situated on the boundary of the inpainted area. According to this mechanism, the pixels that lie on the path of edges are given a higher priority and are thus filled in sooner than the rest of the boundary pixels. Figure 8b shows a high priority pixel p, belonging to a patch Ψ_p, represented as a square surrounding it and lying on the gap's boundary contour $\delta\Omega$. After finding the pixel with the highest priority, the algorithm searches for the best matching patch in the source region, ϕ. This process is exemplified in Figure 8c, where candidates for filling in the patch determined by the pixel p are shown as well. Furthermore, the best match patch chosen from the candidates is used to fill in the patch Psi_p, as shown in Figure 8d.

Another important aspect that contributes to the overall quality of the inpainting is the size of the patches. Their size must be carefully chosen, according to the underlying characteristics of the image that inpainting is applied to [12]. Currently there is no way to automatically adjust the size of the patches, so the user must do it manually. This is actually one of the drawbacks of the algorithm discussed here, as too small patches will

(a)

Figure 9: *Example of an image posing challenges for exemplar based inpainting. There are not enough valid source patches, unless patch size considered for inpainting would be very small, resulting thus in a repetitive, noticeable pattern. Source: [23]*

result in a repeated, noticeable pattern and an increased running time, while patches that are too large will increase the possibility of choosing a *bad* patch, that contains unwanted information [23]. Another disadvantage that is valid for any of the exemplar based methods is that it will provide bad results in cases when the inpainted region is spread out along most of the image. The problem in this case is that there are not enough valid source patches that can be used for determining possible candidates. One such example is given in Figure 9.

As noted in [12], the algorithm also encounters some difficulties in reconstructing curved structures. The authors underline another possible drawback, as the bias produced by selecting incorrect patches when using the priority based filling mechanism. The incorrect information contained within these patches will be propagated further on, hijacking the entire inpainting process.

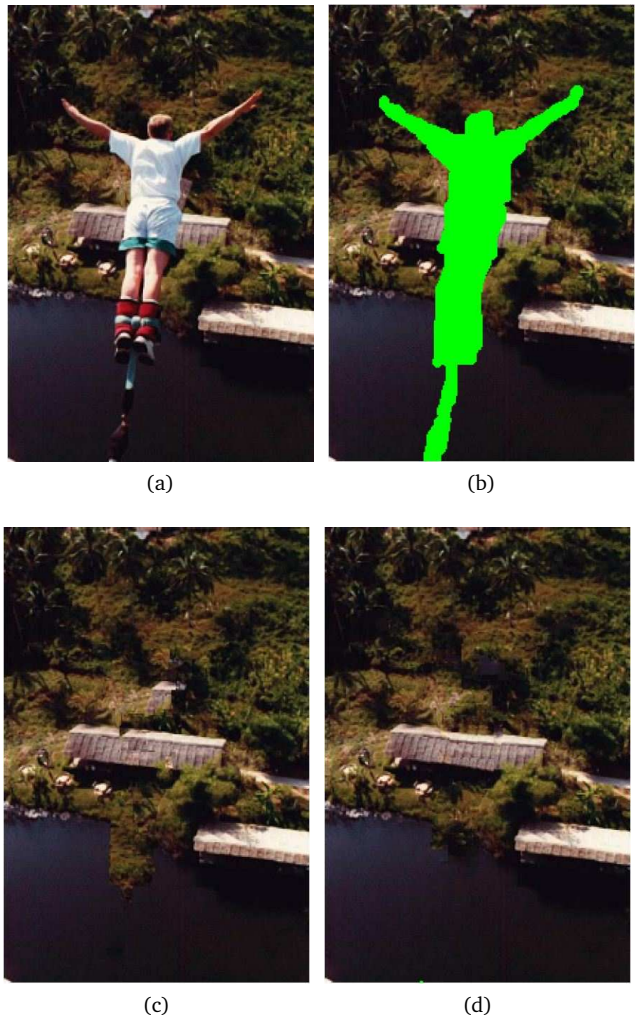
However, for a wide range of images with good texture and structure replication, the results of obtained with the method proposed by Criminisi *et al.* [2] proved to be quite impressive. Unlike PDE based algorithms, that inevitably produce blur in the inpainted region, the exemplar based method preserves texture and structure information.

2.2.2 Improved exemplar based inpainting methods

Image inpainting based on local optimization

Another exemplar based inpainting algorithm was proposed by Zhou and Kelly in [3], as an extension of the method by Criminisi *et al.* [2]. The former uses the same priority based mechanism as in [2] in order to select the boundary pixel with the highest priority. However, instead of determining the candidates based on the similarity to the patch that needs to be filled in, Zhou and Kelly propose a method that aims at preserving the local consistency of the inpainted region. Given the chosen high priority pixel p, the algorithm follows the next steps:

1. Search the source region for candidate patches of the patch Ψ_p (determined by the boundary pixel p) - it gives a list X_p of candidate patches for p;
2. Search the source region for the best n matching patches of each neighbour of the pixel p - it gives a list $X_{x_{\bar{p}}}$ of candidate patches for p's neighbourhood pixels;
3. Find the optimal patch to use by maximising the joint distribution probability given



(a) (b) (c) (d)

Figure 10: *Exemplar based inpainting results by Zhou and Kelly. (a) Original image (b) Original image with overlapped mask (c) Inpainting result with the algorithm by Criminisi* et al. *[2] (d) Inpainting result with the algorithm by Zhou and Kelly [3]. Source: [3]*

by:

$$P(\Psi_p, x_p, x_{\bar{p}}) = P(\Psi_p|x_p)P(x_p|x_{\bar{p}}), P(x_{\bar{p}}) \tag{2.9}$$

where Ψ_p denotes the patch to be filled in (see also Figure 8b), x_p is a candidate patch for Ψ_p, with $x_p \in X_p$, and $x_{\bar{p}}$ is a candidate patch for the neighbourhood pixels of p, with $x_{\bar{p}} \in X_{x_{\bar{p}}}$. In the above equation, the term $P(\Psi_p|x_p)$ reflects the similarity of a candidate patch x_p to the patch Ψ_p, while the second right-sided term $P(x_p|x_{\bar{p}})$ represents the similarity of a candidate patch for the pixel p to a candidate of a neighbour pixel of p. Finally, the last term, $P(x_{\bar{p}})$, establishes the *importance* (priority) of the candidate patch for the neighbourhood pixel $\bar{p}$.

4. Select the patch that gives the maximum probability and use it to fill in the patch Ψ_p.

The algorithm proposed by Zhou and Kelly in [3] provides, generally, better results than Criminisi *et al.*'s method [2], avoiding over-smoothing and thus preserving and

Figure 11: *Exemplar based inpainting using a database of millions of images. (a) Original image. (b) Image with mask (c) Nearest semantic matches for the original image (d) Inpainted image. Source: [24]*

extending image detail, textures and structures. A better local image consistency can be noticed in Figure 10, where the resultant image obtained with the algorithm by Criminisi *et al.* is shown next to the results of Zhou and Kelly.

Scene completion using millions of photographs

A different approach to exemplar based inpainting was taken by Hays *et al.* in [24]. While previous algorithms search for matching patches in the same image that is to be inpainted. This is based on the assumption that the image contains all the information that is necessary to fill in the gap. In [24] the authors state that a more natural and semantically rich completion can be achieved by using a database of millions of images as possible sources for candidate patches. The method proposed by Hays *et al.* uses a Gist image descriptor to characterize each image in the database. Using this information to search through the entire database, the nearest semantic match is determined for the input image (the original image, in need of inpainting). After a matching image is found, the information contained within the area that corresponds to the inpainting mask is seamlessly blended into the original image.The inpainting process steps are illustrated in Figure 11, along with the resultant image.

2.3 Fast Digital Inpainting

The size of the gap that needs to be filled in determines, generally, the time required by an inpainting algorithm to complete. Most of the techniques discussed in the above subsections can take from minutes to hours until the missing region is fully inpainted. An exception from the latter are Tschumperle and Deriche method, that has been incorporated into the GIMP software, and Telea's noniterative algorithm [7]. As interactive user applications require a quite short response time, a new category of fast inpainting algorithms has emerged.

Oliveira et al. proposed in 2001 a fast, iterative, kernel based inpainting method, being probably the easiest, currently existing, algorithm to implement. Essentially, nothing more than a convolution of the gap with a Gaussian kernel, Oliveira et al.'s algorithm diffuses known image information into the gap, by using one of the following weighted average kernels, that only consider the contribution of the neighbouring pixels (the kernel's centre has a weight equal to zero):

$$\begin{bmatrix} a & b & a \\ b & 0 & b \\ a & b & a \end{bmatrix} \quad \begin{bmatrix} c & c & c \\ c & 0 & c \\ c & c & c \end{bmatrix} \tag{2.10}$$

where $a = 0.073235, b = 0.176765, c = 0.125$. The number of diffusion iterations can be set by the user, or the algorithm can stop when there are no more pixels whose value

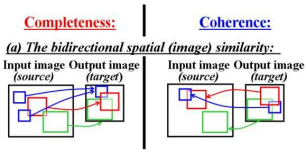


Figure 12: *Bidirectional similarity terms: completeness and coherence. Images source: [25]*

changed by more than a certain threshold as a result of the previous iteration.

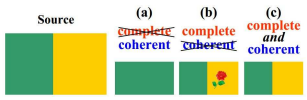
The whole inpainting process described by Oliveira et al. in [9] was designed to fill in small regions, as it can't reproduce structural, nor textural information, resulting in inpainted areas with blur artifacts. Thus in image restoration it could be used only for repairing small scratches or small gaps (i.e. stains). Its strength though, is represented by its simplicity, which makes it very easy to understand and, more importantly, to implement. Despite its simplicity however, the results provided (given the constraint of small sized gaps) are not to be disregarded. Moreover, the little time required to complete and its ability to preserve general color information are also noted as advantages of Oliveira et al.'s method in [23].

The PatchMatch algorithm presented in [8] was designed with the goal of interactive performance in mind, as it is incorporated in the image editing Adobe Photoshop CS 5.1. software. The algorithm presented in [8] can be used for image retargeting, inpainting, as well as image reshuffling. However, as in this work we are interested only in inpainting, the general algorithm proposed by Barnes et al. will not be discussed here, but the reader is referred to [8] for further details. Instead, the specific method employed by the authors for image completion will be presented next.

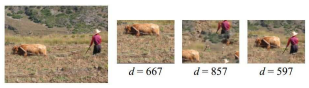
Based on the bidirectional similarity synthesis approach initially proposed by Simakov et al. in [25], the inpainting algorithm denotes the missing region (the gap) by T and the source region (the rest of the image) as S. Given the two, the algorithm proceeds by defining the following bidirectional measure:

$$d_{BDS}(S,T) = \frac{1}{N_S}\sum_{s\in S} \min_{t\in T} D(s,t) + \frac{1}{N_T}\sum_{t\in T} \min_{s\in S} D(t,s) \tag{2.11}$$

where the first term is referred as the *completeness term* and the second as the *coherence term*. While the first "measures the deviation of the target T from 'completeness' w.r.t. S" (), the latter "measures the deviation of the target T from 'coherence' w.r.t. S" [25]. In equation 2.11, N_S and N_T denote the number of patches in the source region, S, and the target region, T, that is to be inpainted. Furthermore, in the same equation, D is a measure of the distance between two patches, and is chosen to be the sum of squared differences (SSD) of pixel values corresponding to the patches s and t in the CIE $L^*a^*b^*$ color space. Thus, for every patch $t \in T$ the algorithm searches for the most similar patch $s \in S$, and once found, it calculates the distance D(,) between the two patches. That gives the 'coherence' term. In order to account for 'completeness', the algorithm takes every patch s, belonging to the source region S, and searches for best matches in the region T, corresponding to the gap. These two processes are schematically shown in Figure 12, where the overlapping resulting from the fact that patches are taken at multiple scales can also be observed.



(a)



(b)

Figure 13: *Examples of images fulfilling the 'completeness' and/or 'coherence' requirements. Column **a** shows examples of coherent images with respect to the source image indicated on the corresponding row, far left, but not complete (all patches within the reproduced image belong to the source image). Column **b** shows examples of complete but not coherent images (all patches in the source image are contained in the reproduced image - at different scales). Finally, column **c** shows examples of images that are complete and coherent. For the last row image the value of the measure computed in equation 2.14, with equal weights, is also given. Images source: [25]*

Considering the following notations:

$$d_{complete}(S,T) = \frac{1}{N_S}\sum_{s\in S} min_{t\in T} D(s,t) \; d_{cohere}(S,T) = \frac{1}{N_T}\sum_{t\in T} min_{s\in S} D(t,s) \tag{2.12}$$

the equation 2.11 becomes:

$$d(S,T) = d_{complete}(S,T) + d_{cohere}(S,T) \tag{2.13}$$

In [25] the authors propose weighting the two terms in equation 2.13, so as to be able to adjust the importance of each of the two, according to the application for which the bidirectional similarity measure is used for. Consequently, the equation in 2.13 becomes:

$$d(S,T) = \alpha d_{complete}(S,T) + (1-\alpha) d_{cohere}(S,T) \tag{2.14}$$

For image completion tasks, Simakov et al. set the value of α to be equal to zero in equation 2.14, ending up to be an image completion algorithm similar to the one proposed by Wexler et al. in [26] . As the *completeness term*, $d_{complete}$, is used for checking if all patches of the source region S (possibly at different scales) have been preserved in the target region T (see Figure 13), it becomes obvious that for inpainting this term is unnecessary, motivating thus the weight choice. As also shown in Figure 13, the coherence term d_{cohere}, on the other hand, checks if the region T contains any patches that have not originated from the source region S (i.e. creating visual artifacts), holding thus a great importance for the goal of inpainting. Hereby, the given weight for the coherence term is set to one, resulting in the following final version of the inpainting equation:

$$d(S,T) = d_{cohere}(S,T) \tag{2.15}$$

Furthermore, the inpainting problem is solved by filling in the missing region T with those patches that optimize the similarity measure of equation 2.15 with respect to the source region S. Thus, the output of the algorithm will be the filled region T_{output}, defined as it follows:

$$T_{output} = argmin_T d(S,T) \tag{2.16}$$

(a) (b) (c)

Figure 14: *Patch Match algorithm inpainting result. (a) 'Soria Moria' painting by Theodor Kittelsen (b) Original image with mask (c) Inpainted image.*

(a) (b) (c)

Figure 15: *Patch Match algorithm inpainting result. (a) 'Women portrait' painting by Misu Popp - private collection, XIXth century (b) Original image with mask (c) Inpainted image.*

Despite the improvement in the running time, the authors of [8] admit that the task of image completion of large gaps still represents a challenge, as the inpainting of structured content (i.e. straight line crossing the gap) can lead to inconsistencies even with the method they propose. The reader is referred to Table .... for a comparison of running times between different algorithms, showing that the PatchMatch algorithm (time) performance is among the best. Figure 14 shows an example of the inpainting result on an image with less structure and more texture. It can be observed how well Barnes et al.'s algorithm performs on this type of images. On the other hand, in Figure 15 the missing region contains an almost straight line (the women's shoulder line) that should be faithfully reconstructed through inpainting. It it obvious, though, from Figure 15c, that the Patch Match algorithm is incapable of doing so, replacing instead the missing information with patches from the source area (thus achieving coherence) in an undesirable way.

Bornemann and Marz proposed in [6] a noniterative, fast-marching inpainting algorithm that has proven (see Table ....) to have a high time performance (translated as little time required to complete). The authors set as their goal the development of an inpainting method that matches "the high level of quality of the methods presented by Bertalmio et al. (2000) and Tschumperle (2006), while being considerably faster" - [6]. In order to achieve this goal, they propose an inpainting algorithm actually based on the work of Bertalmio in [5] and inspired by the fast, non-iterative inpainting method pro-

posed by Telea in [7]. Acknowledging the drawbacks of Telea's algorithm, Bornemann and Marz try to resolve them, at least partially, by implementing an algorithm that modifies the weight function used for the method proposed by Telea in [7] (see the 'Partial Differential Equation (PDE) based inpainting' section for details), while trying to keep its fast performance. In doing so, the authors of [6] thrive to obtain the formal stationary (i.e. not changing over time) state of Bertalmio et al.'s technique.

By writing the Equation 2.3 in its equivalent form given by:

$$I_t(x,y) = -\nabla^{\perp}\Delta I(x,y) \cdot \nabla I(x,y) \tag{2.17}$$

Bornemann and Marz show that the transport equation proposed by Bertalmio et al. in [5] is essentially a transport equation for the image value I(x,y) in the direction of the level lines, and not for the smoothness L of the image I along the isophotes, as it was claimed. The latter affirmation results from the very simple fact that it's I_t and not L_t that appears on the left hand side of the Equation 2.17, thus being assigned a new value. The same can be observed from the formal version of equation 2.17, in a stationary state, which is given as:

$$\vec{c} \cdot \nabla I(x,y) = 0, \qquad where \quad \vec{c} = \nabla^{\perp}\Delta I(x,y), \tag{2.18}$$

Once again, from the above equation one can infer that the image values are transported along the direction given by the continuations of edges arriving at the boundary of the inpainting domain, inside the gap. As shown previously, when discussing Bertalmio et al.'s method, their choice of propagating information along the isophote's direction and the diffusion process interleaved with the inpainting steps results in blurred regions and introduces 'peculiar transport patterns' ([**?**]), as the one shown in Figure 6b. Therefore, Bornemann and Marz propose the use of a coherence direction instead of Bertalmio et al.'s edge-oriented transport direction.

To sum up the above, the fast inpainting method presented in [6] relies on the following two main concepts:

1. Modifying Telea's weight function (in [7])
2. Replacing the edge-oriented transport direction proposed by Bertalmio et al. in [5] by the coherence direction.

The new weight function used in [6] is given by:

$$w(p,q) = (\frac{\pi}{2})^2 \frac{\mu}{|p-q|} exp\left(-\frac{\mu^2}{2\epsilon^2}|\vec{c}^{\perp}(p)\cdot(p-q)|^2\right) \tag{2.19}$$

In Equation 2.19 p and q denote the pixel that needs to be filled in and a known pixel belonging to its neighbourhood, respectively. ϵ gives the radius of the neighbourhood, and is set by the user, while the vector $\vec{c}$ gives the coherence direction. The latter is defined as the normalized eigenvector to the minimal eigenvalue of $J_\rho(\nabla u_\sigma)$, where J_ρ is a structure tensor of the image $I(x,y)$. A structure tensor is a symmetric positive semidefinite 2x2 matrix and is computed as it follows:

$$J_\rho(\nabla u_\sigma) = K_\rho * (\nabla u_\sigma \otimes \nabla u_\sigma), \qquad u_\sigma = K_\sigma * u \tag{2.20}$$

In the above equation K is a Gaussian kernel, its indices σ and ρ denoting its variance. In other words, the two parameters, set by the user, ρ and σ refer to the scale on which details influence the calculation of the transport direction. Furthermore, in equation 2.20,

the $*$ symbol refers to a convolution operation. The only unknown in equation 2.19 is now the μ parameter, which can be computed as:

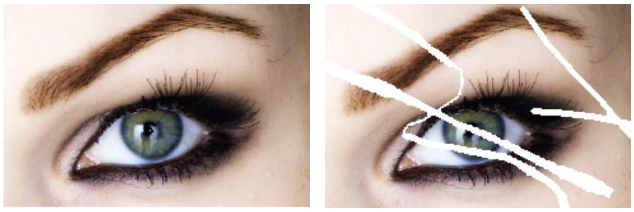
$$\mu(x) = \begin{cases} 1 & if \lambda_1(p) = \lambda_2(p) \\ 1 + \kappa exp\left(\frac{-\delta^4_{quant}}{(\lambda_2(p) - \lambda_1(p))^2}\right) & otherwise \end{cases}$$

Finally, in equation 2.3 the parameter denoted by $\kappa$is a sharpness parameter, set by the user, and λ_1 and λ_2 are the eigenvalues of J_ρ, with $0 \leq \lambda_1(p) \leq \lambda_2(p)$,where p is the pixel to be inpainted. Denoting by $\vec{w}_1(p)$ and $\vec{w}_2(p)$ the normalized eigenvectors corresponding to the two eigenvalues λ_1 and λ_2, and knowing that the coherence direction $\vec{c}$ is given by the normalized eigenvector to the minimal eigenvalue of $J_\rho(\nabla u_\sigma)$, which is $\vec{w}_1$, it results that:

$$\vec{c} = \vec{w}_1 \tag{2.21}$$

As the weight function is now defined, the algorithm proposed by Bornemann and Marz proceeds with the inpainting of the pixel p, whose value is modified according to the expression given in Equation 2.5, where instead of using the weight function given in equation 2.6, the new weight function in equation 2.19 is used. Except for the change in the weighting function, the inpainting process is similar to the one described in Telea's paper [7].

The image obtained by employing Bornemann and Marz's inpainting technique, compared to the one obtained after applying Telea's method is presented in Figure 16. The differences between the two resulting images are easily noticed, as in Figure 16c the inpainted area is visibly more blurred than the image in Figure 16d, where virtually no diffusion can be noticed. The authors of [6] note even that an (nonexpert) observer could actually consider the image in Figure 16d as being the original one, if not presented with the one in Figure 16a. Disregarding the visually improved results, Bornemann and Marz's algorithm increases the computing time, from 0.08 sec. with Telea's method, to 0.4 sec. However, compared to other inpainting algorithms, the method proposed in [6] is still a fast one (see Table.... for the processing time of other algorithms).



(a) (b)

(c) (d)

Figure 16: *Comparison of inpainting results: Bornemann and Marz vs. Telea methods. (a) Original image (b) Original image with mask (c) Result of Telea's inpainting method [7] (d) Result of Bornemann and Marz inpainting method*

3 Survey of Quality Metrics

Crosby [27] gives a general definition of quality, that can relate to any field of study, by saying that quality is the compliance to the imposed requirements. Later on, with the introduction and expansion of photography and television, a more specialized term came to be known, and widely used, namely the *image quality*. Jansen defined image quality [28], as being the extent to which an image is both natural and useful. In this context, the naturalness is given by the degree of similarity between the internal representation of the considered image and the actual (real life) representation of the scene in the image, while the usefulness is given by the internal representation's precision. Other similar definitions have been given in [29], [30], adapting the general definition of quality to the field of digital imaging.

In order to evaluate the quality of an image, many Image Quality (IQ) metrics have been proposed, considering various applications (i.e. video, images) and different distortions (i.e. compression, noise). The diversity of the existent image quality metrics can be explained by the general tendency that tries to obtain a good correlation with perceived image quality [1]. However, in the specialized field of digital inpainting, research has been more concentrated around the development of new algorithms, or improvement of the existent ones, than on the quality evaluation of the inpainted images, as highlighted by Mahalingam in his work, [12].

This Chapter will discuss methods that have been extensively used for the evaluation of inpainting quality, as well as more recently proposed methods having the same goal. An overview of general image quality metrics that could be applied for inpainting quality evaluation is given in the second part of this Chapter.

3.1 Overview of inpainting quality metrics

Early inpainting algorithms (i.e. Bertalmio *et al.* [5]) have been specifically designed for filling in small missing or damaged regions in digital images and later in vintage films. Their objective was, at that time, to reconstruct the image such that the result would appear to be as close to the original (undamaged) image as possible. Consequently, many researchers evaluated the quality of the images inpainted with their newly proposed method by employing simple objective metrics, such as the Mean Squared Error (MSE) or the Peak Signal to Noise Ratio (PSNR). One such example is given in [9], where Oliveira *et al.*, the authors of this fast inpainting technique discussed in Section 2.3, evaluate the performance of their algorithm by computing the MSE of the inpainted image, in the RGB color space (compute the MSE separately for each of the three color channels, R, G and B). However, because the MSE and PSNR are essentially image fidelity metrics, they do not perform well in characterizing perceptual image quality [12].

As the domain of digital image inpainting spread, to include applications that require a larger area to be inpainted, the objective of the inpainting process also changed, from trying to achieve high fidelity with the original, not degraded, image, to making the inpainted area as visually pleasing as possible. In other words, the modifications brought to the original image through inpainting should not be discernible by human observers

[12]. Given this new objective of inpainting, it becomes obvious that evaluation methods previously used (i.e. MSE, PSNR) could no longer provide a good assessment of the inpainting quality. Thus, new inpainting evaluation methods were required, methods that would take into account the subjective nature of the human visual system..

One such method that accommodates the new objective of the inpainting process is represented by the visual comparison technique, which, as the name suggests, consists of nothing more than putting two (or more) images together followed by a visual comparison made by a single person (i.e. author). As a common practice, one of the considered images is a reference image, which, most of the times, is an image inpainted by using a well-known inpainting method (i.e. Criminisi *et al.* [2], Bertalmio *et al.* [5]). It is clear that this evaluation method reflects the personal opinion of the author, being thus a highly unreliable and subjective evaluation method.

Hays et al [24] were the first to perform a psychophysical experiment, with human observers, that required the participants to rate the quality of the inpainted images presented. Based on the perceptual results obtained from the experiment, they conclude that the method they propose outperforms the exemplar based inpainting method presented by Criminisi *et al.* in [2].

As psychophysical experiments are both time and resource demanding, objective methods that evaluate the quality of inpainting without the involvement of human observers are usually preferred. With this in mind, Ardis *et al.* [13] and Mahalingam [12] proposed visual saliency-based metrics, that will be discussed in the following sections.

3.1.1 ASVS and DN

Ardis *et al.* [13] acknowledge the lack of research work in the field of inpainting evaluation and propose two metrics that relate the visual saliency map of an image with the perceived quality of the corresponding image. The saliency maps used for computing the metric values are generated by using version 3.1 of the *iLab Neuromorphic Vision Toolkit (iNVT)*, at 1:16 discretization of scale-4 (noiseless) expected visual cortex stimulation with 0.1 ms observation cutoff, 4 orientation channels, 3 center scales (2 to 4) and 2 center-surround channels (3, 4) [13].

The first metric proposed by Ardis *et al.* is the *Average Squared Visual Salience (ASVS)*, computed as:

$$ASVS(I) = (1/\|\Omega\|)(\sum_{\Omega}(S'(p))^2). \tag{3.1}$$

In the above equation *I* is the image considered for inpainting, *p* is a pixel belonging to the inpainted region, Ω, and *S'(p)* is the saliency value corresponding to it. By computing the normalized sum of the squared salience values of the inpainted pixels, the ASVS metric is a no-reference metric, which means that it does not require the use of an original (not degraded) image. This holds a great importance for artwork restoration, where there might be cases when there is no information about how a work of art originally looked like.

The ASVS metric takes into consideration only the inpainted area, as inferred from Eq. 3.1. Thus, this metric accounts for in-region artifacting, a class of observable artifacting defined by Ardis *et al.* in [31], which occurs when the pixels in the inpainted region draw more attention than they are suppose to. As noted in [31] the in-region artifacting is, most of the times, related to distinct coloration or structure, created after inpainting, that can not be observed anywhere else in the image. Consequently, this newly introduced

color or structure will result in high saliency values, and thus higher ASVS score.

Another class of artifacting specific to inpainting applications is represented by the out-region artifacting. As its name suggests, this class encompasses cases where the inpainting process modifies the flow of attention outside the inpainting region. The latter usually happens when an inpainting algorithm can not successfully continue a locally repeating color or structure, from outside the gap, inside it. The discontinuation thus introduced indirectly draws the attention towards the inpainted area, by decreasing the flow of attention drawn by otherwise salient areas and increasing the attention in the neighbourhood area of the inpainting domain. The out-region artifacting can be computed by using the following equation:

$$\mathrm{out-region}(I) = \frac{1}{|\Theta|} \sum_{p \in \Theta} (S'(p) - S(p))^2, \tag{3.2}$$

where *I* is the image to be inpainted, and *S'(p)* and *S(p)*are the saliency map values corresponding to a pixel p belonging to the inpainting domain in the modified image and original, unmodified image, respectively.

In order to account for out-region artifacting, Ardis *et al.* propose another metric, namely the *Degree of Noticeability (DN)*, which takes into consideration both classes of artifacting, being calculated as:

$$\mathrm{DN}(I) = \frac{|\Omega|}{|\Omega| + |\Theta|} \mathrm{in-region}(I) + \frac{|\Omega|}{|\Omega| + |\Theta|} \mathrm{out-region}(I), \tag{3.3}$$

where *in-region(I) = ASVS(I)*. A s with the *Average Squared Visual Salience*, higher scores obtained for DN generally indicate a poor inpainting performance.

In [31] the authors perform an psychophysical experiment for the evaluation of the proposed metrics and conclude that the two saliency-based metrics "correctly (but imperfectly) measure the phenomena that they are intended to (i.e. increased notice of the inpainted region, significant attention change)" [31]. The authors recommend the use of the ASVS metric for the evaluation of inpainting when the fidelity of the resultant image to the original is not a matter of importance. A high value for the ASVS metric can be interpreted as an indicator of highly visible artifacts. The DN metric, on the other hand, should be used for evaluating the inpainting quality of images required for applications that demand that the original flow of attention be kept after inpainting. Higher values for DN may indicate the presence of artifacts that interfere with this goal.

Although the subjective evaluation against which the two metrics are evaluated shows a good correlation between perceived and calculated quality, the findings of Ardis *et al.*'s study needed to be confirmed by further experiments, as it only considered a total of five observers.

3.1.2 GD_{in} and GD_{out}

In [12], Mahalingam proposes another two visual saliency-based metrics, resembling the approach presented earlier by Ardis *et al.* Mahalingam defines a normalized gaze density measure that uses the original (not deteriorated) image as a reference, and shows that if there is any change in the saliency map corresponding to the inpainted image, then this change is related to the perceptual quality of the inpainting.

As in [13], the evaluation metrics proposed in [12]are also computed separately, for the region corresponding to the inpainting domain, and outside it respectively, resulting

in the following two equations:

$$GD_{in} = \sum_{p \in \Omega} S'(p) \quad and \quad GD_{out} = \sum_{p \in \Theta} S'(p), \tag{3.4}$$

where Ω represents the inpainting domain and Θ the complementary region. Furthermore, S' is the saliency map corresponding to the inpainted image, which gives $S'(p)$ as the saliency map value corresponding to pixel p. In the above equation, the two metrics compute the gaze density inside and outside the inpainted area of modified images. In order to be able to compare the values provided by these metrics for different images, a normalization is required. Consequently, the following two ratios are calculated:

$$\bar{GD}_{in} = \frac{GD_{in}^{inp}}{GD_{in}^{unmod}} \quad and \quad \bar{GD}_{out} = \frac{GD_{out}^{inp}}{GD_{out}^{unmod}}, \tag{3.5}$$

where GD_{in}^{unmod} and GD_{out}^{unmod} are calculated similarly to Eq. 3.4, by replacing $S'(p)$ with $S(p)$, where S is the saliency map of the original, unmodified image. For simplicity, from this point further, whenever discussing the GD_{in} and GD_{out} inpainting evaluation metrics, this simple notation will be used, referring to the normalized version of the metrics. When computing the GD_{in} and GD_{out} scores for an inpainted image, a value closer to unity generally indicates that the flow of attention after inpainting does not deviate from normal viewing behaviour.

In his work, Mahalingam performs an eye tracking experiment in order to obtain the saliency maps for the images considered. However, since inpainting evaluation methods that are fully automatic are preferred to those that imply the participation of human observers, in this thesis the *SaliencyToolbox* version 2.2 [32] is used to generate the saliency maps. The *SaliencyToolbox* is a reimplementation of the *iNVT* toolkit developed by Laurent Itti's lab at the USC. This toolbox is more compact, easier to understand and experiment with.

3.2 Overview of image quality metrics

As mentioned earlier, objective metrics for image quality assessment provide a replacement for the resource-intensive perceptual experiments. It was shown in Section 3.1 that in the field of inpainting, the focus of research has not been put on the development of objective metrics, but on the development and improvement of inpainting methods. Thus, in the remainder of this section a number of objective metrics that have been more extensively used for general image quality evaluation will be discussed. These metrics will be further on applied to inpainted images in order to determine whether they can provide similar, or better results than inpainting-specific objective metrics.

3.2.1 MSE and PSNR

The Mean Squared Error (MSE) and the Peak Signal to Noise Ratio (PSNR) are two mathematically based metrics that have been widely used due to the fact that they are easy to implement and thus convenient to use for optimization purposes [1].

The MSE computes the cumulative squared error between two images: one reference (i.e. original) image and a modified version of it. The equation for calculating the MSE is given as:

$$MSE = \frac{1}{mn} \sum_{y=0}^{m-1} \sum_{x=0}^{n-1} [I_O(x,y) - I_R(x,y)]^2, \tag{3.6}$$

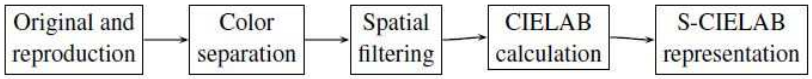


Figure 17: S-CIELAB flowchart
S-CIELAB flowchart. Source: [1]

where I_O denotes the original (reference) image and I_R its reproduction (modified version). x and y indicate a pixel location and m and n give the number of rows and columns of the input images, respectively. A lower value for the MSE indicates a lower error, thus better quality of the reproduction.

The PSNR computes the peak signal to noise ratio between two images and gives a value in decibels as a result. The higher the value obtained for PSNR, the better the quality of the reproduction. First step in computing the PSNR is to obtain the MSE value for the two images. Then, the PSNR is calculated as:

$$PSNR = 10 \log_{10} \frac{R^2}{MSE}, \tag{3.7}$$

where R is the largest value a pixel can take. For the images considered in this thesis, stored as 8-bit unsigned integer data type, $R = 255$.

The MSE and PSNR were computed for color images by converting the images to the YCbCr color space, which separates the intensity (luma) channel from the color channels. This choice was made because the human eye is more sensitive to intensity information. Taken this into consideration, the MSE and PSNR are calculated on the Y channel.

Despite their easy calculation, it has been proven that these two metrics do not correlate well with perceived IQ [33]. However, the MSE and PSNR will be considered for further comparison, as they have been widely used for inpainting quality assessment (due to a lack of an effective objective metric).

3.2.2 S-CIELAB

The S-CIELAB metric was proposed by Zhang and Wandell in [34] as a spatial extension of the CIELAB color formula. Thus, this metric aims to simulate the blurring of the human visual system by including a spatial filtering step in the quality prediction workflow. As shown in Figure 17 the spatial filtering step is followed by the basic CIELAB color difference calculation between the two images.

In the color separation stage, represented in Figure 17, the RGB input images are converted into the CIE-XYZ color space and then into the opponent color space which separated the luminance information ($O_1 channel$) from the red-green ($O_2 channel$) and blue-yellow information ($O_3 channel$) respectively. The spatial filtering is then applied to each channel, by using a 2D separable spatial kernel. The resultant image is then converted back to CIE-XYZ color space and from that to the CIELAB color space. The next step consists of applying the ΔE^*_{ab} formula for calculating the pixel-wise color difference between the two input images. Finally, the value for the S-CIELAB metric that represents the overall image difference is given by the average over the pixels color difference. The lower the value for the S-CIELAB metric, the better the quality of the reproduction.

The S-CIELAB image quality metric has been initially designed for the assessment of halftone image quality, but as shown in [1], it has also been widely used for computing the overall IQ for a large number of distortions. Moreover, the S-CIELAB is frequently

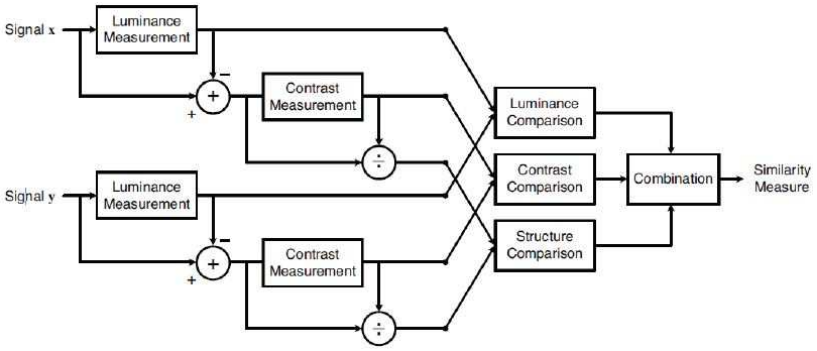


Figure 18: *SSIM flowchart for a local window. Source: [36]*

considered as a reference metric when evaluating different metrics against each other.

3.2.3 $\mathrm{SSIM}_{\mathrm{IPT}}$

The $\mathrm{SSIM}_{\mathrm{IPT}}$ metric was introduced by Bonnier *et al.* in [35] and represents the color version of the Structural Similarity Index (SSIM) proposed by Wang *et al.* in [36]. The idea behind this metric is to perform separate comparisons of the luminance, contrast and structure information between local windows in the reference and distorted images and then combine the results of these comparisons in order to obtain the value for the SSIM. Finally, the SSIM value corresponding to the overall quality of the distorted image is obtained by averaging the previously calculated SSIM values, corresponding to local windows in the two images. The workflow for the computation of the similarity index is given in Figure 18, as proposed in [36]. $\mathtt{Signal}_x$ and $\mathtt{Signal}_y$ denote local windows for which the similarity measure is computed. According to Wang *et al.*, the structural information within an image is given by the attributes that describe the structure of the objects in that image, independently of the average contrast and luminance.

3.2.4 VSNR

The VSNR denotes a wavelet-based Visual Signal-to-Noise Ratio which was proposed by Chandler and Hemami in [33], and which aims at quantifying the visual fidelity of distorted images. The metric takes into consideration low level as well as mid-level features of the HVS. The low level properties such as contrast sensitivity and visual masking are employed in the wavelet domain and compared to the detection threshold. If detected distortions are below the threshold, the reproduced image is considered to have perfect visual fidelity, corresponding to a value for VSNR which would be equal to infinity, and thus no further analysis is required. However, if above the visual detection threshold, the distortions are considered as suprathreshold (visible) and the low level HVS feature of perceived contrast of the distortions and mid-level property represented by the global precedence are taken into consideration as alternative indicators of structural degradation. In this context, the global precedence property of the HVS refers to the proportionality of contrast distortions across spatial frequency. According to the above, the value for the VSNR metric corresponding to two images is given by a combination of perceived contrast of the distortions and discontinuity of global precedence.

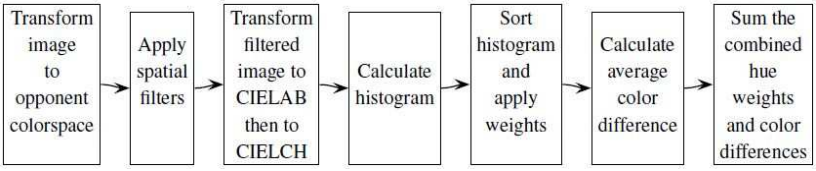


Figure 19: *SHAME and SHAME-II flowchart. Source: [1]*

3.2.5 SHAME and SHAME-II

Pedersen and Hardeberg [37] proposed two new IQ metrics that consider regions of interest in an image and assign assigned different weights to different pixels. The Spatial Hue Angle MEtric (SHAME) and SHAME-II metrics take into consideration the fact that when evaluating quality, the human perception depends on the viewing conditions (i.e the distance an image is viewed from) and thus it tries to simulate the human visual system (HVS) in order to be able to provide results that are better correlated with the perceived quality.

The SHAME metrics are both based on the Hue Angle algorithm proposed by Hong and Luo in [38], but each of them uses a different spatial filtering method. The hue angle algorithm is essentially a full-reference image difference metric implemented for color images and using the CIELAB ΔE^*_{ab} color difference formula proposed in [39]. The four main aspects that the algorithm is based upon are summarized by Hong and Luo as it follows:

- "Pixels or areas of high significance can be identified, and suitable weights can be assigned to these.
- Pixels in larger areas of the same color should be given a higher weight than those in smaller areas.
- Larger color difference between the pixels should get higher weights.
- Hue is an important color perception for discriminating colors within the context."

The two spatial filtering methods included in the SHAME(II) metric computation, as proposed by Pedersen and Hardeberg [37], refer to a spatial filtering method adopted from S-CIELAB (for SHAME) and a spatial filtering adopted from Johnson and Fairchild in [40] (for SHAME-II), respectively. While the idea behind the S-CIELAB spatial filtering implementation has been previously presented, the second method proposed by Pedersen and Hardebeerg implements the spatial filters using Contrast Sensitivity Functions in the frequency domain, as opposed to the spatial domain (i.e. S-CIELAB). In [37] the authors show that metrics that use spatial filtering in the frequency domain generally outperform those that are based on spatial filtering in the spatial domain.

In Figure 19 the workflow of the SHAME and SHAME-II metrics is shown, where one can note that after converting the input images to the opponent color space, a spatial filtering is applied to the converted input images, and further on, the obtained filtered images are used as input to the hue angle algorithm. By adding spatial filtering to the hue angle measure, the SHAME and SHAME-II metrics present the following key features:

- Higher weights are allocated to pixels belonging to large areas of same color.
- Simulates spatial properties of HVS.

- Distortions that can not be detected are neglected.
- Can be applied to different types of distortions; not restricted to color patches.
- The final result is given as a single, numerical value that can be easily interpreted.

3.2.6 ABF

The Adaptive Bilateral Filter proposed by Wang and Hardeberg in [41] is essentially an image difference metric. This metric uses an edge preserving bilateral filter, which blurs the image while preserving the existent edges, providing thus better results than those obtained when using Contrast Sensitivity Functions (CSFs) - as is the case of S-CIELAB metric. Before applying the bilateral filter, the images are converted to the CIELAB color space where the ΔE^*_{ab} formula is applied in order to obtain the color difference. It can be noticed from these two last steps that the workflow for computing the ABF value is similar to the one corresponding to S-CIELAB metric, the difference between the two consisting in the filtering step.

When evaluating the image quality, a lower value for the ABF will indicate a better quality of the reconstructed image.

4 Proposed Metrics: BorSal and StructBorSal

As seen in Section 3.1, visual saliency based metrics proposed by Mahalingam [12] and Ardis et al. [13] attempt to predict inpainting quality by separately considering the regions inside and outside the gap, respectively. Building upon this previous related work, one full-reference metric, denoted by Border Saliency (BorSal), is proposed in this thesis. The novelty here consists of considering only one border region around the gap, that contains information from both the inpainting domain and its complementary area. The choice for the region border was inspired by the work of Ardis et al. [13], in which the authors state that, if existent, out-region artifacting affects the flow of attention outside the inpainting area, but concentrated around its neighbourhood. Consequently, investigating the saliency map values corresponding to a border region around the gap should be able to accurately capture this change of attention flow, without the need of examining the saliency map of the whole image. Furthermore, the BorSal metric is computed as a normalized gaze density measure, similar to the GD_{in} and GD_{out} metrics proposed by Mahalingam in [12]:

$$\mathrm{BorSal}(I) = \frac{\sum_{p \in \mathrm{Border}} S'(p)}{\sum_{p \in \mathrm{Border}} S(p)}, \tag{4.1}$$

where S' and S are the saliency maps of the inpainted and reference image, respectively and p is a pixel in the region corresponding to the Border region. In this thesis and for further evaluation of the metric, the size of the border region was chosen to be equal to six pixels, containing information from both outside and inside the gap, and was obtained by applying morphological operations (i.e. erosion, dilation) to the inpainting domain.

Using the already defined BorSal metric, a second inpainting quality measure is proposed, by combining the BorSal metric with the $SSIM_{IPT}$ measure introduced by [35]. The latter takes into account structural information in the whole image and thus can indicate the presence of artifacting due to poor structure reconstruction, while the BorSal metric accounts for changes in the flow of attention, post inpainting. The new metric will be denoted by StructBorSal and its value calculated by adding up the value obtained according to Equation 4.1 for BorSal, with the $SSIM_{IPT}$ score, calculated for the entire image.

5 Evaluation of Inpainting Algorithms and Quality Metrics

As shown in Chapter 2, a large number and variety of inpainting algorithms currently exist. To try and cover all these methods would be infeasible. Thus, this thesis only looks at a selection of algorithms, that take different approaches in their shared goal of filling in missing parts of an image. The performance evaluation of these algorithms requires an image database, created according to the instructions given in the following section. The methodology for evaluation of inpainting algorithms and quality metrics will be discussed in the last part of this Chapter.

5.1 Inpainting algorithms

Eight inpainting algorithms, from those discussed in Chapter 2, have been selected for further evaluation. Table 1 lists these algorithms, where *Year* is the head of the column containing the year of the publication proposing the algorithm, the column *Author(s)* gives the name of the author(s) and *Type* indicates the category the algorithm belongs to, depending on the approach it takes. Furthermore, in the *Gap size* column, the symbols *L, M* and *S* denote the size of the gap for which the algorithm provides the best results. In the same table, *Structure* and *Texture* indicate whether an algorithm is capable of reproducing structure and texture, respectively, whereas the *Blur* column indicates the existence of blurring post-inpainting artifacts.

Table 1: Selected inpainting algorithms

Year	Author(s)	Type	Gap size	Structure	Texture	Blur
2000	Bertalmio *et al.* [5]	PDE based	S	Yes	No	Yes
2004	Telea [7]	PDE based	S - M	Yes	No	Yes
2005	Tschumperle and Deriche [4]	PDE based	S - M	Yes	No	Yes
2004	Criminisi *et al.* [2]	Exemplar based	M - L	Yes	Yes	No
2010	Zhou and Kelly [3]	Exemplar based	M - L	Yes	Yes	No
2001	Oliveira *et al.* [9]	Fast (convolution)	S	No	No	Yes
2007	Bornemann and Marz [6]	Fast (PDE based)	M - L	Yes	No	Yes
2009	Barnes *et al.* [8]	Fast (Exemplar based)	M - L	Yes	Yes	No

Mai and Chen developed the Image Inpainter tool [23] that implements a number of algorithms, allowing a user to accomplish the task of image inpainting. This tool, is publicly available for download from [42]. The algorithms implemented as part of the Image Inpainter tool exhibit different approaches towards digital image inpainting and have distinct properties, making them suited for different types of image inpainting tasks. This tool has been used in this thesis for accessing the C++ implementations of the methods proposed by Bertalmio *et al.* [5], Telea [7], Criminisi *et al.* [2], Zhou and Kelly [3] and Oliveira *et al.* [9]. The Image Inpainter tool allows the user to tune the parameters specific to each of the aforementioned inpainting methods, so that to obtain best possible quality for the inpainted images. However, setting the correct parameters is a subjective process, highly dependent on the user and on the input image. Currently, there exist no objective method for choosing the set of parameters providing the best results. Thus,

researchers in the field of digital inpainting are required to experiment with different sets of parameters and perform a visual inspection of the inpainted images in order to establish which parameters determine the best result. The latter has been confirmed by researcher Aurelie Bugeau (Associate Professor at Video Analysis and Indexing research group in Image and Sound Department of LaBRI), as a result of an open discussion on this topic.

The commercial software Adobe Photoshop CS5.1 provides a content-aware filling tool, as the implementation of the inpainting method proposed by Barnes *et al.* in [8]. Optimised for user interaction scenarios, two of its main advantages are represented by the redundant parameter tuning and the increased performance in terms of running speed (Table 13).

The inpainting algorithm introduced by Tschumperle and Deriche in [4] has been implemented in the G'MIC plug-in for GIMP 2.6 and is publicly available for download from [43]. The Region Inpainting filter allows parameter tuning and completes the task of reconstructing missing or corrupted content of image regions in a very short amount of time (see Table 9).

Finally, Bornemann and Marz, the authors of the inpainting algorithm proposed in [**?**], provide a Windows executable of the associated code, with its Matlab driver. The latter is publicly available for academic purposes and can be downloaded from [44]. The provided implementation allows parameter tuning.

Experiments using the above implementations were performed on a PC with an Intel Core 2 Duo 2.66 GHz CPU with 3.25 usable RAM.

5.2 Quality metrics

This Section will provide a summary of the 14 metrics that were selected for further evaluation in terms of their correlation with the percept. As seen in Chapter 3, these metrics can be divided into two main groups, one referring to metrics specifically designed for inpainting quality evaluation, and one that consists of metrics developed for general image quality evaluation.

Table 2 lists the inpainting-specific metrics whose performance will be assessed as part of this thesis. In this table, the column *Year* specifies the year of the publication in which the metric was proposed, *Name* gives the abbreviated name of the metric and *Author(s)* its authors name. Furthermore, the column *Type* indicates the type of the metric with respect to the use of a reference image, while the last column, *Affected region*, specifies the region (with respect to the gap) that is considered when calculating the score of the metric. The Matlab implementation of these metrics followed the specifications given by the authors of [13] and [12].

Table 2: Selected inpainting-specific quality metrics

Year	Name	Author(s)	Type	Affected region
2009	DN	Ardis and Singhal [13]	Full reference	in-region and out-region
2009	ASVS	Ardis and Singhal [13]	No reference	in-region
2010	GD_{in}	Mahalingam [12]	Full reference	in-region
2010	GD_{out}	Mahalingam [12]	Full reference	out-region

Table 3 lists the eight general image quality evaluation metrics. As in the case above, *Year*, *Name* and *Author(s)* give the year of publication, name of the metric and name

of the author(s), respectively. The column *Type* indicates the metric's type, as given by its author(s) and *HVS* contains information on whether the metric simulates the human visual system (HVS) or not. Finally, *Image* indicates the type of the images for which the metric is applied (i.e. color or grayscale). All the metrics in Table 3 are full-reference metrics.

Table 3: Selected general image quality metrics. Adapted from [1]

Year	Name	Author	Type	HVS	Image
-	MSE	-	Image difference	No	Gray
-	PSNR	-	Image difference	No	Gray
1996	S-CIELAB	Zhang and Wandell [34]	Image difference	Yes	Color
2004	$SSIM_{IPT}$	Bonnier *et al.* [**?**]	Image difference	No	Color
2007	VSNR	Chandler and Hemami [33]	Image fidelity	Yes	Gray
2009	SHAME	Pedersen and Hardeberg [37]	Image quality	Yes	Color
2009	SHAME-II	Pedersen and Hardeberg [37]	Image quality	Yes	Color
2009	ABF	Wang and Hardeberg [41]	Image difference	Yes	Color

In Table 3 the type of the metrics is given as either *image quality*, *image difference* or *image fidelity*. The former refers to metrics used for predicting image quality, while the goal of the second type of metrics is to predict the perceived magnitude of differences between two images (usually *original* and *reproduction*) [1]. Finally, image fidelity metrics are generally used for predicting the visibility of image reproduction errors. In this thesis, however, the metrics presented in Table 3 will all be referred to as (general) image quality metrics.

In addition to the four inpainting-specific metric and the eight general image quality metrics, the two metrics proposed in this thesis, BorSal and StructBorSal, will be considered for further evaluation.

5.3 Image database

Generally, when an inpainting algorithm is considered for evaluation, a common practice in the related literature is to use random images and to simulate occlusions. One such example is given in Figure 20, where the original image is presented on the right and the image with a simulated occlusion, in white, is shown on the left. Instead of simulating the scratches on images, this master thesis proposes the use of images corresponding to real damaged paintings. This attempt to perform actual digital restoration poses some issues that will be discussed in the following paragraphs.

The issues that arise derive mainly from the fact that the objective evaluation methods considered for comparison in this thesis make use of a reference image - that shows how the painting and its colours looked in their original, unimpaired state (referring to the previous example of the simulated occlusion, the reference image would be the one in Figure 20a). Whereas the common practice of simulating the missing information is straightforward, giving access to the original, unmodified image which can eventually serve as a reference image when evaluating the performance of a specific algorithm, the choice made in this thesis, of using images of real damaged paintings, makes this comparison to the reference image more difficult, if not impossible. The reason for the aforementioned is that the images provided by Ruven Pillay at the Centre for Research

(a) (b)

Figure 20: *Simulated occlusions as input to inpainting algorithms. (a) original image (b) simulated occlusion in white*

Figure 21: *The Virgin of the Light c.1730 by Gio Nicola Buhagiar (1698-1752) – Detail of the painting during cleaning.*

and Restoration of the Museums of France[1] and by restorer Radu Tataru at the Brasov Art Museum in Romania, correspond mainly to pairs of images before and after the restoration. While at first sight it may not be obvious why this represents a drawback for the general purpose of this thesis, the next paragraph will provide more details on this matter.

The process of bringing a painting to its original state, as when it just came out of the artist's studio, or at least to an as-close-as-possible state, is a complex process, involving a number of treatments that must consider the auxiliary frame (if existent), the canvas support, the ground layer, the paint layer, the surface and finally, the decorative frame. From discussions with professional restorer Radu Tataru, resulted that one of the most important treatments applied to a work of art refers to its cleaning. This sub-specialized process aims at reversing the chemical and/or physical phenomena like dirt, dust, and light exposure, that substantially degrade the appearance of a painting over time. Figure 21 emphasises the considerable importance of this step in restoring colours and details that were previously hidden by a layer of old varnish. In this Figure the opaque area is the one that is still covered by the varnish layer.

[1] the Centre for Research and Restoration of the Museums of France (C2RMF Centre de recherche et de restauration des musees de France) is the national research centre in France responsible for the documentation, conservation and restoration of the items held in the collections of more than 1,200 museums across France.

(a) (b)

Figure 22: *'Ex-Voto: Madonna and Child with Saints' painting by Borg [?]. 17th century, oil on canvas - before and after restoration.*

However, it is before applying any treatment to the damaged painting, that the work of art usually goes through a visual examination under raking light (paintings are illuminated from a light source situated at an oblique angle with respect to their surface). This type of examination is preferred for works of art because the increased illumination of the raised paint surfaces facing the light source together with the shadows on the surfaces that face away from the light source reveal the surface texture of a painting. Thus, by using raking light to evidentiate the effects of impasto[2] and any degradations of the canvas (i.e. craquelure, paint cupping, uneven tension in a canvas), conservators are able to judge aspects of the condition of a painting. Moreover, photographing the painting under raking light is an important step in the documentation required before the actual restoration process begins. Figure 22 presents an instance of the 17th century 'Ex-Voto: Madonna and Child with Saints' oil painting, under normal light, before and after the restoration, while Figure 23, below, shows details of the same painting, before restoration, with various defects exposed under a raking light.

Figure 23 shows details of a painting in need of restoration, but not necessarily inpainting, as there are no significant missing regions being noticed. These images have a purely documentary function, as they illustrate in an obvious manner the effects of raking light on a painting, as well as the change in color determined by the removal of the old varnish layer. Thus, this images will not be used further in this thesis. On the other hand, in Figure 24 an image of the self-portrait by Margarete Depner (XX century) provided by Radu Tataru and considered as a test image in the experimental setup is presented, both before and after the restoration process. Details of this painting are presented in Figure 25 and show the use of the impasto technique for the realization of this work of art, while putting emphasis on the areas of the painting surface that need to be inpainted.

After performing inpainting on the image in Figure 24a, the results obained with three different algorithms (Criminisi's examplar based inpainting, Local optimization, Photo-

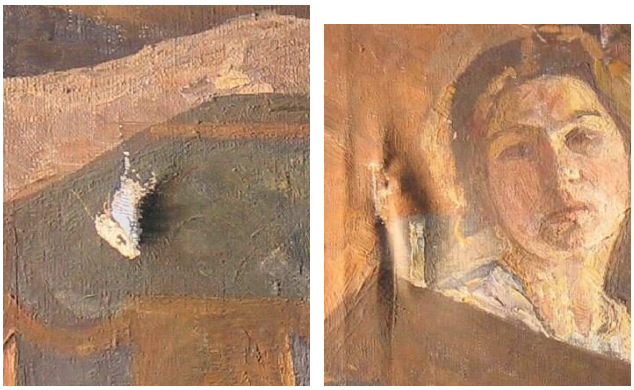
[2]painting technique, where paint is laid on the canvas very thickly, usually thickly enough for the brush strokes to be visible

(a) (b) (c)

Figure 23: *Detail of 'Ex-Voto: Madonna and Child with Saints' painting with defects in the canvas tension, seam and perforations evidentiated by inspection under raking light.*

(a) (b)

Figure 24: *Self-portrait by Margarete Depner - before and after manual restoration. Image courtesy of restorer Radu Tataru at Brasov Art Museum.*



(a) (b)

Figure 25: *Details of 'Self-portrait' by Margarete Depner. Inspection under raking light. Image courtesy of Radu Tataru*

(a) (b) (c)

Figure 26: *Digital image inpainting results with different algorithms. (a) Criminisi - examplar based inpainting (b) Zhou and Kelly - inpainting based on local optimization (c) Photoshop CS 5.1.*

shop CS 5.1) are presented in Figure 26. Having the resultant inpainted images, the next logical step would include their quality evaluation. However, before proceeding to that, there is still one requirement to be fulfilled. As stated earlier, the objective inpainting evaluation methods that are considered in this study make use of a reference image in order to establish the quality of the result provided by an algorithm. In order to provide accurate results this reference image should meet one essential condition: relating to the digitally inpainted image, the reference image should reflect changes (if any) only in the area corresponding to the region that was selected for inpainting in the deteriorated image. In the case under discussion the reference image is taken to be the image in Figure 24b, as it is the image that was provided together with the degraded image, illustrating the state of the painting after the restoration. Hypothetically this image should make a good reference image, as it corresponds to the manually restored self-portrait, and thus including an inpainting step where the missing areas are filled in. However, when inspecting the digitally inpainted images in Figure 26 and the reference image in Figure 24b, it becomes clear that the differences between any two images forming a pair *reference - digitally inpainted*, go beyond the area selected for inpainting, and this is mainly because the manually restored version of the painting was subjected to various treatments (i.e. cleaning), not only inpainting. The differences in the lighting conditions used for the two images and the old varnish layer covering the deteriorated painting affect the overall appearance of the images, contributing to the fading and discoloration of the non-restored image when compared to the restored version. Thus, the assumed reference image in Figure 24b does not meet the previously imposed condition of complete identity outside the inpainted area.

Another issue that may arise when considering images of real deteriorated paintings as input to inpainting algorithms is related to the presence of specular reflection under raking light. Figure 27 illustrates this issue, showing a pair of images corresponding to the state of a painting before and after the restoration, together with the digitally inpainted version of the image. The wide spread area of specular reflection is an important

(a) (b) (c)

Figure 27: *Icon of Tsarevich Dmitry (detail) (a) Original degraded painting (b) Manually restored version (c) Digitally inpainted version*

aspect in this case, as it contributes to how the holes in the image will be reconstructed. In Figure 27c the red rectangles indicate two areas that have been inpainted using information collected from areas with specular reflection. Consequently, instead of eliminating the specular reflection (so that to obtain an image that is closer to the reference image - the manually restored image), by applying inpainting directly on the image in Figure 27a more regions of specular reflection are created. Except for the propagation of specularity, as in the previous study case of the self-portrait, differences in color can also be observed between the reference image and the digitally inpainted one.

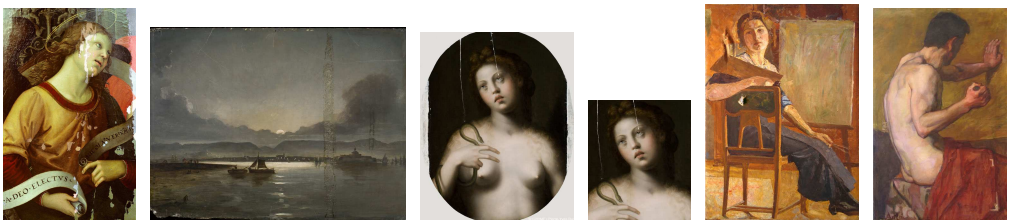
For all the above mentioned reasons and in order to be able to accurately evaluate the inpainting algorithms, a compromise has been done regarding the images used as input. Instead of completely simulating defects, as done in other works on related topics, this master thesis proposes the simulation of the steps preceding the inpainting in the process of manual restoration (i.e. cleaning the dust, old varnish, etc.), so that to bring the original degraded image to a state that is as close as possible to the state of the final restored image. The result of this simulation would be an image that replicates the regions containing missing or corrupted information from the degraded image but that has identical content with the manually restored image in the region outside the area to be inpainted. Knowing that inpainting is the last phase of the restoration process of a painting, the proposed simulation is the equivalent of taking one step back from reaching the final state in the manual restoration. Figure 28 illustrates an example of such a simulation, showing the original degraded image and its corresponding restored version. In Figure 28c an image that was subjected to modifications as the ones proposed above is presented. It can be noticed that the appearance of this image is similar to the one in Figure 28b, while it contains the water stain and other defects characteristic to the degraded image in Figure 28a.

5.4 Test images

For evaluating the performance of the inpainting algorithms presented in Section 5.1, six different images (Figure 29), varying in size and resolution, as seen in Table 4, have been considered. These images, labelled as *angel*, *boat*, *cleopatra*, *detail*, *lady* and *man*, respec-

(a) (b) (c)

Figure 28: *Preparing an input image for inpainting. Image courtesy of Ruven Pillay at C2RMF. (a) Original degraded painting (b) Manually restored version (c) Modified image containing artefacts from the degraded image while maintaining the appearance of the manually restored image*



(a) (b) (c) (d) (e) (f)

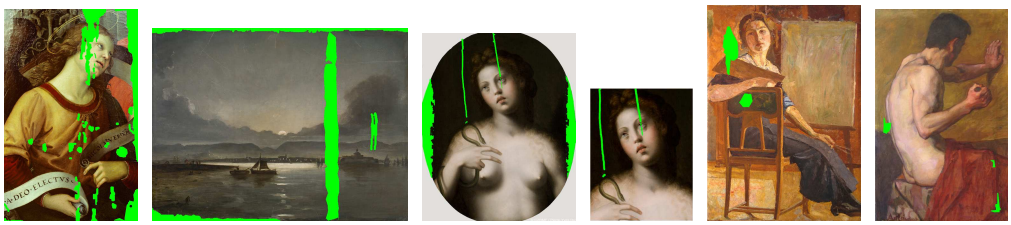
Figure 29: *Modified digitally acquired paintings corresponding to: (a)* Ange *by Raphael (*angel *) (b)* Vue de Drontheim *by Peder Balke (*boat*) (c)* Cleopatra *(*cleopatra*) (d)* Detail of Cleopatra *(*detail*) (e)* Self-portrait *by Margarete Depner (*lady*) (f)* Invention of painting *by Ariton (*man*). (a-d) - courtesy of R. Pillay at C2RMF, (e-f) courtesy of R. Tataru at Brasov Art Museum (Romania)*

tively, were obtained using the procedure described in Section 5.3. Figure 30 shows the six deteriorated images, with the inpainting domain marked in green. As it can be observed, the size of the region to be inpainted differs from image to image, varying from narrow gaps (i.e. detail) to larger gaps (i.e. boat). Furthermore, the selected images contain regions of missing information that require either structure or texture reconstruction, or both. An example for the latter is given in Figure 31.

For each of the six deteriorated images, nine reproductions were considered as test images in a psychophysical experiment. Out of these nine reproductions, eight were digitally inpainted images, obtained as output of the inpainting algorithms presented in Section 5.1. In addition to the latter eight, the digitally acquired image of the manually inpainted work of art was also considered as a test image and included in the experiment. Figure 32 shows an example of a set of 10 images, consisting of one deteriorated version of a painting (*angel*) and its nine reproductions (i.e. inpainted images). The whole set of test images included in the experiment, reproductions of *angel*, *boat*, *cleopatra*,*detail*, *lady*, *man*, respectively, can be seen in Appendix A. The parameters used in order to obtain the test images are also included in Appendix A.

Table 4: Resolution and size of images considered for inpainting algorithms evaluation

Image	Resolution(dpi)	Size (pixels)
angel	72	580x921
boat	72	800x601
cleopatra	300	550x670
detail	300	341x398
lady	762	550x771
man	180	400x635



(a) (b) (c) (d) (e) (f)

Figure 30: *Deteriorated images as input for inpainting algorithms, with overlapped inpainting masks. (a) angel (b) boat (c) cleopatra (d) detail (e) lady (f) man*

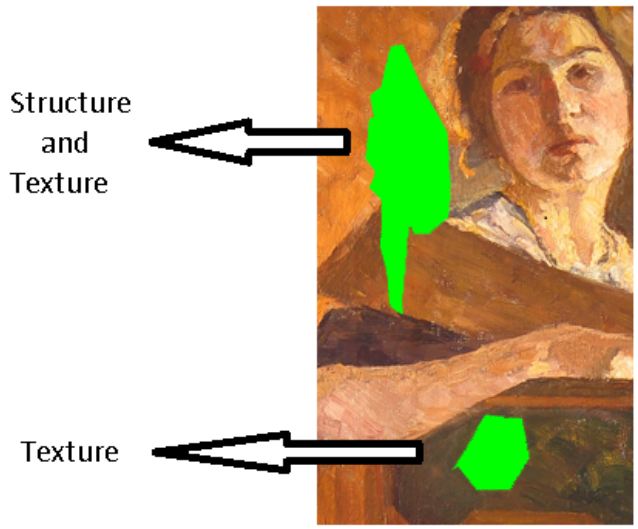


Figure 31: *Deteriorated image requiring structure and texture reconstruction. Inpainting mask is shown overlapped, in green*

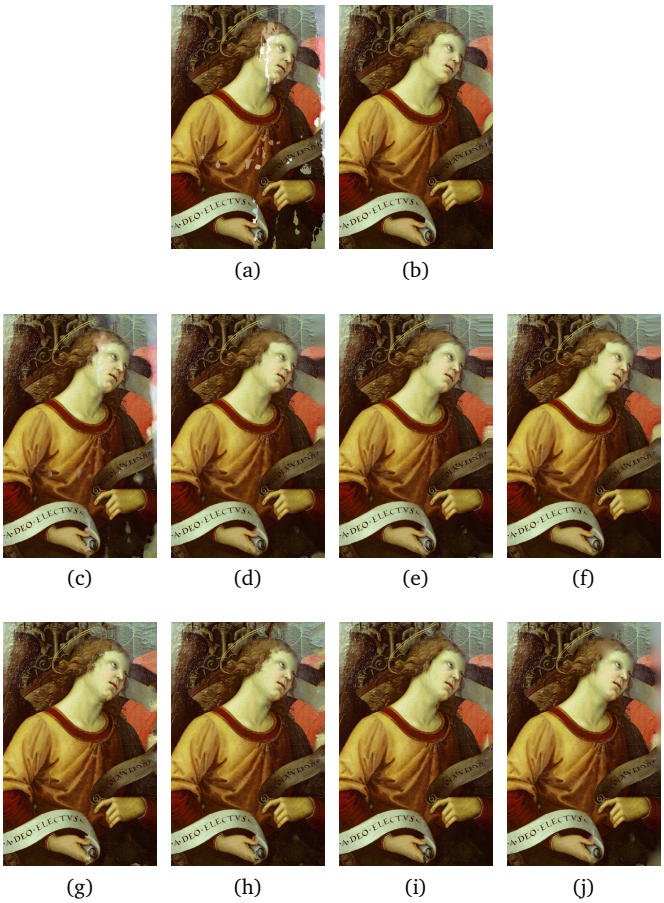


Figure 32: *Image set included in the test images. (a) Modified digitally acquired paintings corresponding to: Retable de Saint Nicolas de Tolentino, "Ange" by Raphael ('angel') - image courtesy of Ruven Pillay at C2RMF. Inpainted images corresponding to: (b) manually restored (c) Bertalmio et al. [11] (d) Telea [7] (e) Tschumperle and Deriche [4] (f) Bornemann and Marz [6] (g) Criminisi et al. [2] (h) Zhou and Kelly [3] (i) Barnes et al. [8] (j) Oliveira et al. [9]*

5.5 Evaluation methodology

In order to establish the difference in quality of the resultant inpainted images, in the field of digital inpainting subjective evaluation is commonly used, as it is a precise way to quantify quality, but also because of the lack of a reliable quantitative metric. Thus, in an attempt to establish a ranking of the considered algorithms in terms of perceived quality of the inpainted images,and considering the database described above, a psychophysical experiment will be carried out, according to the specifications given in Chapter **??**. The raw perceptual data obtained will be statistically analysed in order to determine the ranking of the inpainting algorithms.

For the second part of this study, which refers to the evaluation of objective quality assessment methods, the image database will consist of only digitally inpainted images (a total of 48 images), thus excluding images corresponding to manual inpainting. For this set of test images, the values given by quality metrics are computed, resulting, for each image, in one score that describes its quality. Evaluating the metrics consists then in comparing the obtained objective results, with the subjective scores given by observers to the same images. Consequently, the performance of the metrics is quantified by the correlation between the two. Higher correlation values indicate higher performance of the metrics. In this thesis two standard correlation measures will be used:

- The Pearson product-moment correlation coefficient (PCC) [45], which finds the linear relationship between two variables using the formula:

$$r = \frac{\sum_{i=1}^{N} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{N} (X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{N} (Y_i - \bar{Y})^2}}, \tag{5.1}$$

  where $r \in [-1, 1]$, N denotes the number of samples and X and Y are variables associated with subjective (i.e. z-scores from observers) and objective (i.e metrics) results respectively.

- Spearman's rank correlation coefficient [45] which is defined as the PCC between the ranked variables. Accordingly, the SCC is calculated as in Equation 5.1, where X and Y denote the rank of the subjective and objective results, respectively. This formula allows for tied ranks.

Following the methodology above, the performance of the considered metrics is evaluated over the entire image database and then in an image-wise manner, for each of the six different images.

6 Psychophysical Experiment for Subjective Quality Assessment

A psychophysical experiment was performed as part of the work for this thesis and with the goal of subjective quality assessment of the inpainting results. Moreover, the subjective evaluation is of great significance for the current work, as the data thus obtained will be used further on for validating objective methods of quality assessment. The latter provide the means to ensure the correlation of the metrics with perceived quality. This Chapter will discuss the experimental setup and the psychophysical results obtained.

6.1 Experimental setup

6.1.1 Test images

For the psychophysical experiment performed as part of this thesis, the image database described in Section 5.4 was used, counting a total of 54 images. As mentioned earlier, images corresponding to both, manually and digitally inpainted works of art are included in this database. The digitally acquired images of manually inpainted paintings have been included as test images with the aim of verifying the reliability of the observers, as recommended in Rec. ITU-R BT.500-13 [46]. The latter is done by checking the behaviour of the participants when degraded/manually-inpainted image pairs are presented.

6.1.2 Experimental method

The experimental method for the subjective quality assessment was chosen to be the Mean Opinion Score (MOS) [47]. This method was initially designed for audio quality evaluation, but it has been widely used for image quality evaluation [48]. The MOS values are based on subjective data obtained from the experiment. Participants were presented with a pair of two images at a time. The pair consisted of a degraded image and one of its reproductions (i.e. inpainted version). To each observer pairs were shown in a different and random order. Given a pair of images, the participants were asked to judge the overall IQ of the inpainted image using the ITU-R five point quality scale [46], labelled with the adjectives: *Excellent*, *Good*, *Fair*, *Poor*, *Bad*. In order to be able to analyse the subjective data obtained, each of the five adjectives in the descriptive quality scale had an equivalent numerical value, or score (not shown to the observers). Accordingly, *Excellent* corresponded to a *5* score and *Poor* to a *1* score. The MOS is then obtained for each reproduction by computing the arithmetic mean of the individual scores given by participants:

$$\mathrm{MOS} = \frac{1}{n}\sum_{i=1}^{n} S_i. \tag{6.1}$$

where n denotes the number of observers, and S_i the score given by the i*th*observer to the reproduction under consideration.

A common practice when analysing perceptual data is to transform collected data into interval scale data, by converting MOS values to z-scores. A z-score calculated for an image represents the distance between its MOS and the mean score of the entire set of

images, in units of the standard deviation. Accordingly, negative z-scores will be obtained for MOS values below the mean, and positive when above. The formula for calculating a z-score is:

$$\bar{X} = \frac{x - \mu}{\sigma}, \tag{6.2}$$

where x denotes the MOS value of an image, μ represents the mean of the MOS values for the entire image set and σ the standard deviation.

MOS values and z-score are usually given with a 95% confidence interval (CI). For a set of images with a known estimate of the mean MOS or z-score, denoted by μ, and known standard deviation, σ, the 95% CIs are calculated using the formula [49]:

$$\mu \pm 1.96\frac{\sigma}{\sqrt{N}}, \tag{6.3}$$

where N indicates the size of the sample from which μ was obtained. The CI are used to verify if two estimates μ_1 and μ_2 are significantly different from each other at the 95% level. The latter is true when the confidence intervals of the two estimates do not overlap, and false otherwise.

6.1.3 Viewing conditions

The experiment was carried out in an uncontrolled environment, as a web-based experiment. One of the most important factors that was considered when deciding to conduct this type of experiment was the ease of acquisition of a large number of participants for achieving high statistical power, making it thus possible to draw meaningful conclusions from the experiment. Furthermore, experimenting around the clock and the feasibility of reaching different categories of participants (i.e. experts in image quality, professional restorers, naive observers) were two other aspects that contributed to the final decision. Finally, research on this topic has shown that the difference between results obtained from controlled and uncontrolled experiments is negligible [50], [51].

6.1.4 Instructions given to the observers

The observers were given clear instructions, focused on the overall quality rating of the inpainted images. Accordingly, before commencing the experiment the observers were told they will be presented with pairs of two images, as it follows:

- *One image that corresponds to the original state of a damaged painting, with visible artifacts (i.e. alterations, defects)*
- *One image that corresponds to a restored version of the same painting, after the existent artifacts have been repaired*

Figure 33 shows an example of such a pair presented during the experiment. The left image shows the degraded image and the right the inpainted image. The instructions also stated that this positioning of the two images will be kept as such, throughout the entire experiment. Given a pair of images, as above, observers were instructed to evaluate the overall quality of the right-sided, inpainted image, using a scale from *Excellent* to *Poor*, where *Excellent* represented the most pleasing reproduction:

- Excellent
- Good
- Fair

Figure 33:
Exemple of pair of images presented during the experiment. (left) Damaged painting (right) Inpainted image, using the algorithm proposed by Oliveira *et al.* [9]

- Poor
- Bad

Observers were asked to complete the experiment, which consisted of viewing and rating a total of 54 pairs of images. An estimation of how long the participation in the experiment will take (around 20 minutes) was also given to the observers.

In order to avoid having multiple submissions from the same observer, before commencing the experiment, observers were asked to pass a step involving collecting user information (i.e. initials, age, sex, country of residence, professional restorer or not, expert in image quality or not).

6.2 Psychophysical results

6.2.1 Observers

A total of 91 observers participated in the psychophysical experiment for inpainting quality rating. Before proceeding with the interpretation of the obtained results and data analysis, in Rec. ITU-R BT.500-13 [46] a procedure for screening the observers is recommended. This procedure was implemented for this thesis according to the specifications given in the aforementioned technical report. Following, the mathematical form of the screening procedure is given, while the reader is referred to [46] for more detailed in-

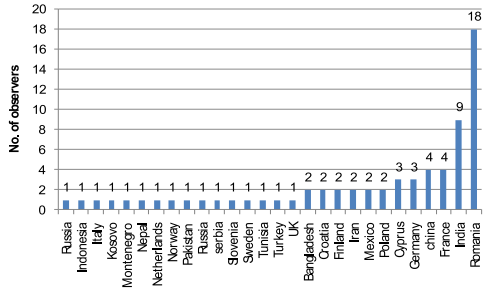


Figure 34:
Demographic distribution of observers based on geographical location

formation.

$$
\begin{aligned}
&\bar{u_{jkr}}(mean), s_{jkr}(standarddeviation), \beta_{2jkr}(kurtosiscoefficient)\\
&\beta_{2jkr} = \frac{m_4}{(m_2)^2}, \quad with \quad m_x = \frac{\sum_{i=1}^{N}(u_{ijkr} - \bar{u_{ijkr}})^x}{N}\\
&foreachobserver, i, calculate P_i and Q_i :\\
&if 2 \leq \beta_{2jkr} \leq 4 then :\\
&if u_{ijkr} \geq \bar{u_{jkr}} + 2s_{jkr} \quad then P_i = P_i + 1\\
&if u_{ijkr} \leq \bar{u_{jkr}} - 2s_{jkr} \quad then Q_i = Q_i + 1\\
&else :\\
&if u_{ijkr} \geq \bar{u_{jkr}} + \sqrt{20}s_{jkr} \quad then P_i = P_i + 1\\
&if u_{ijkr} \leq \bar{u_{jkr}} - \sqrt{20}s_{jkr} \quad then Q_i = Q_i + 1\\
&\text{If} \quad \frac{P_i + Q_i}{JKR} > 0.05 \quad and \quad |\frac{P_i - Q_i}{P_i + Q_i}| < 0.3 \quad then reject observer i,\\
&where :\\
&j, k, r = 1, 1, 1 to J, K, R\\
&J : number of test conditions (J = 9)\\
&K : number of test images (K = 6)\\
&R : number of repetitions (R = 1)\\
&L : number of test presentations; L = JKR (L = 54)
\end{aligned}
\tag{6.4}
$$

In addition to the screening procedure described above, observers were rejected if failing to complete the experiment or if their scores indicated Fair, Poor or Bad quality for manually inpainted images, or if failing to complete the experiment. Consequently, from a total of 91, results of 22 participants were rejected and only 69 considered for further evaluation. A demographic distribution based on the geographical location of the observers is shown in Figure 34, while Figure 35 indicates the number of participants belonging to three different categories, based on the experience in judging or evaluating images.

6.2.2 Perceptual results

This Section will present an analysis of the perceptual results obtained from the psychophysical experiment for inpainting quality evaluation. As stated earlier, three different categories of observers participated in this study (i.e. *naive*, experts in image quality

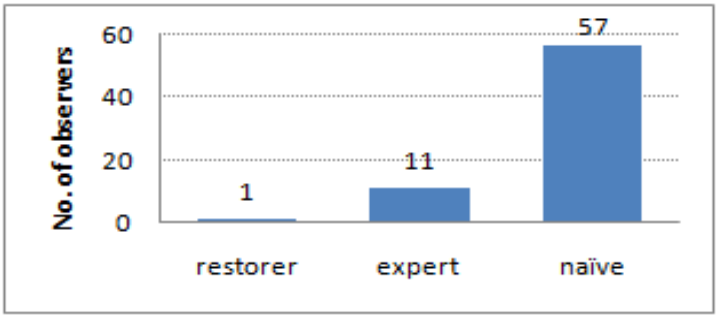


Figure 35: Demographic distribution of observers based on expertise and background

and professional restorers). Thus, the evaluation has been divided into four cases, one for each category of observers and one that refers to the general qualitative evaluation of inpainting methods and that takes into consideration all 69 participants to the experiment.

As expected, for all the four considered evaluation phases, the manual inpainting method received the highest scores, indicating the best perceived quality of the results, among the studied methods. Moreover, the high mean value calculated for it for all four phases, indicates a large agreement among the participants about the high performance of this method. This was an expected outcome, as manual inpainting usually is done by consulting with professional restorers that have access to additional sources of inspiration (i.e. sketches of the original, undamaged painting) for reconstructing the work of art. Reconstructing texture and structure in the gaps of a painting is flawlessly done by following the indications of the additional material or, when not available, based on human intuition. Obviously, artifacts like blurring of the inpainted area do not appear when performing manual inpainting. Other artifacts like misalignments of linear structures are less likely to occur for this type of inpainting.

Case 1: naive observers The ranking of the inpainting methods according to naive observers is given in Figure 36. The algorithm proposed by Barnes *et al.* [8] and implemented in the commercial software *Adobe Photoshop CS.5.1*. received the highest scores among the digital inpainting methods, indicating perceived quality close to *Good*. This method performed significantly better than the rest of the considered algorithms, whose perceived quality of the results is close to *Fair*. The following three best algorithms, that were proposed by Tschumperle and Deriche [4], Zhou and Kelly [3] and Criminisi *et al.*[2]), did not prove to have a significantly different performance from each other. The former uses a vector valued regularization PDE, requiring careful parameter tuning. However, as it has been implemented in the commercial software GNU The GIMP, the parameter tuning has been optimized for user interactivity and thus increased performance. The method proposed by Zhou and Kelly [3] is an improved version of Criminisi *et al.*'s exemplar based method [2]. Allegedly, better results are expected from Zhou and Kelly's method, as it aims at preserving local consistency of the inpainted region and avoids its over-smoothing. In the obtained results, better performance can be observed for Zhou and Kelly's method, when compared to Criminisi *et al.*'s method, but not significantly different. Bornemann and Marz's PDE based method [6] builds on Telea's [7] and Bertalmio *et al.*'s method [5] while trying to minimise the effects of the diffusion process, resulting in less visible blur artifacts. The latter is reflected by the obtained results,

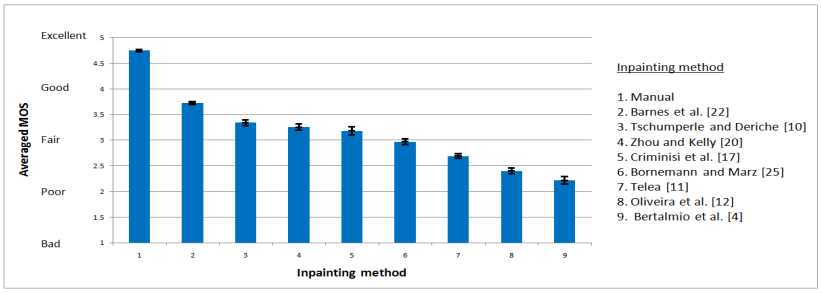


Figure 36:
Averaged MOS based on the results obtained from 57 naive observers, with 95% CI

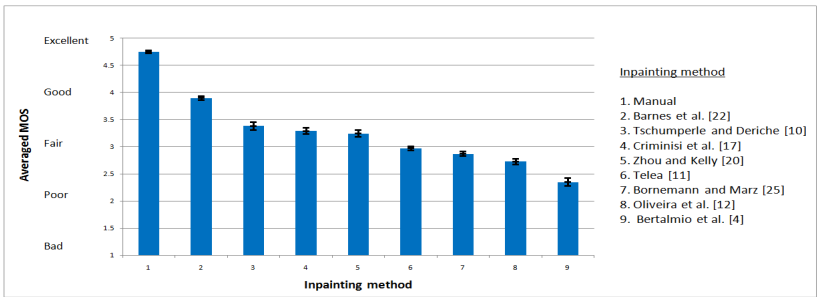


Figure 37:
Averaged MOS based on the results obtained from 11 expert observers, with 95% CI

as they indicate a significantly better performance for the method in [6] when compared to methods proposed in [5] and [7]. Oliveira *et al.*'s inpainting algorithm [9] is rated by naive observers as the second worse in terms of inpainting quality, being followed by Bertalmio *et al.*'s method. This is an expected outcome, as the latter is a basic PDE based method that produces noticeable blurred areas and fails in reproducing texture, while Oliveira *et al.*'s method is able to preserve general color information but can't reproduce structure nor texture in the gap.

Case 2: expert observers The ranking of the inpainting methods according to observers that are experts in image quality evaluation is given in Figure 36. Most of the observations made for naive observers also hold as true in this phase. Again, the average MOS value obtained for the method proposed by Barnes *et al.* [8] indicates the best and significantly different performance among digital inpainting methods. The performance of the similar approaches of Criminisi *et al.* [2] and Zhou and Kelly [3] is not significantly different, as also inferred from the previous evaluation case, but their order in the ranking of the inpainting methods is now interchanged, as compared to the ranking according to naive observers. The same is valid for the methods proposed by Telea [7] and Bornemann and Marz [6]. However, in this case, the results obtained indicate that the quality of the inpainted images for these two methods do not differ in a significant manner, as inferred by their overlapping CIs in Figure 37.

Case 3: professional restorer The averaged MOS values corresponding to the rating given by one professional restorer that participated to the psychophysical experiment are given in Figure 38. In this case, the CIs associated to the obtained results are noticeable large. That is because the length of a CI is determined, among other factors, by the

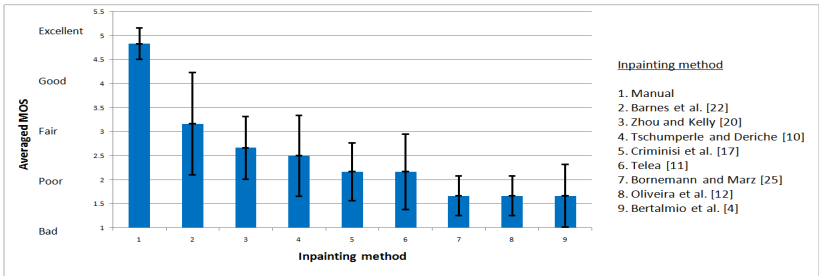


Figure 38:
Averaged MOS (per inpainting method) based on the scores given by a professional restorer, with 95% CI. Based on this data a clear ranking of the inpainting methods is impossible

size of the sample used in the estimation procedure. The largest the size, the smallest the CI. Accordingly, the length of the CIs for data in Figure 38 is determined by the number of professional restorers taking part in the experiment, which is equal to one, resulting thus in large CIs. Because of this, the statistical power of the analysed data is very low. However, few conclusions can be drawn based on this data. The method by Barnes *et al.* [8] obtained the highest averaged MOS value, but the CI associated to it indicates that its performance is only significantly better when compared to methods proposed by Bornemann and Marz [6] and Oliveira *et al.* [9]. Worthy of noticing is that according to the professional restorer, images inpainted with Zhou and Kelly's algorithm [3] outperforms (as given by the mean MOS) Tschumperle and Deriche's method [4], which has been ranked as second best among digital inpainting algorithms by observers in the other two categories.

Case 4: general evaluation As the inspection of the obtained results for each of the three categories of observers shows no significant difference and in order to obtain a higher statistical power, when evaluating the performance of quality metrics against in the following section, objective data will be validated against perceptual data based on the total number of observers that participated in the experiment. Thus, the scores shown in Figure 39 are the average scores of the judgements made by the 69 observers. Based on this data, a ranking of the inpainting methods can be established, indicating that PDE-based methods are, generally, outperformed by exemplar-based methods. An exception to the latter statement is the algorithm proposed by Tschumperle and Deriche [4], which is not significantly different from the exemplar based methods proposed by Zhou and Kelly [3] and Criminisi *et al.* [2], as also shown earlier. The overlapping CIs of these three methods indicate the low visual difference between the inpainted images obtained by the corresponding algorithms and implies a difficult task for the observers to judge image quality.

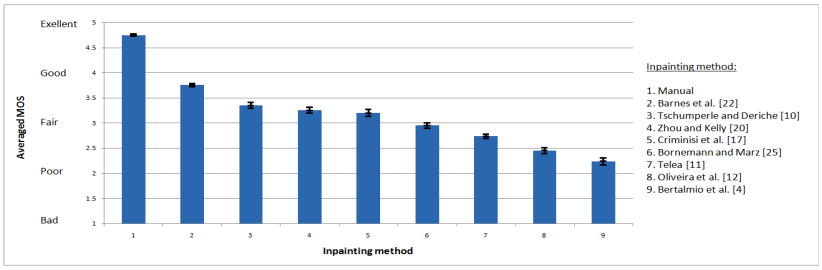


Figure 39:
Averaged MOS from observers based on 69 observers and 6 images, with 95% CI. The algorithm proposed by Barnes *et al.* [8] is rated as the best among digital inpainting methods

7 Objective Quality Evaluation: Results and Discussion

Following the methodology presented in Section 5.5, the performance of a set of quality metrics will be evaluated in this Chapter. The set includes eight general image quality metrics and four inpainting quality metrics, discussed in Chapter 3, and the two novel inpainting quality metrics introduced in Chapter 4.

It will be reminded here that the evaluation methodology refers to statistically analysing the perceptual data represented by ratings given by 69 observers to 48 digitally inpainted images (obtained by excluding from the initial database the six manually inpainted images). Based on raw perceptual data, the Mean Opinion Score (MOS) is calculated for each image in the database and then converted to a corresponding z-score. For each image and each metric, an objective score will be calculated, describing the quality of the reproduction. Finally, the Pearson product-moment (PCC) [45] coefficient and Spearman's rank correlation coefficient (SCC) [45] between the z-scores and the objective scores are calculated in order to evaluate the performance of the metrics considered. The metric that correlates the best with the human observers will be considered the most suitable metric.

In order to achieve an extensive evaluation of the metrics, the investigation of their performance was carried out in two stages. First, the quality of the metrics over the entire image database is evaluated, followed by an image-wise evaluation, when the selected metrics are evaluated on individual images in the database.

7.1 Overall performance of the metrics

In order to evaluate the overall performance of a metric, the correlation between the observers z-scores and the metric raw scores corresponding to the 48 images in the database is calculated. The obtained results, presented in Table 5, indicate that all the considered metrics have a low Pearson correlation, and thus, can not accurately predict perceived image quality. The two novel inpainting quality metrics proposed in this thesis, BorSal and StructBorSal, have a very low correlation. However, they achieve a better performance than the GD_{in} and GD_{out} inpainting quality metrics proposed by Mahalingam [12]. Furthermore, the structural similarity and border saliency based metric, StructBorSal, outperforms the ASVS metric [13] for the considered image database. The DN saliency based metric [13], also designed for inpainting quality evaluation, provides the highest correlation among all metrics, but with a value equal to -0.36 it still indicates a low performance over the entire image database.

The considered metrics perform similar to the case above in terms of rank order correlation, as indicated by the Spearman Correlation Coefficient given in Table 5.

Figure 40 shows z-scores obtained for the DN metric [13] plotted against z-scores from the observers. It can be noticed that data points are very spread, explaining thus the low overall correlation between calculated quality (z-scores for DN) and perceived quality (z-scores for observers). The red line in Figure 40 is a linear regression line determined by all data points. Additionally, on the same Figure, black linear regression lines determined by data points corresponding to individual images have been added.

Table 5: Performance of the metrics over the entire image database. The score for best performing metric is highlighted in bold font

	MSE	PSNR	DN	ASVS	GD_{in}	GD_{out}	S-CIELAB	SHAME	SHAME-II	$SSIM_{IPT}$	ABF	VSNR	BorSal	StructBorSal
PCC	-0.13	0.28	**-0.36**	-0.11	-0.01	-0.01	-0.26	-0.15	-0.25	0.22	-0.32	0.06	0.06	0.12
SCC	-0.28	0.28	**-0.39**	-0.11	0.07	-0.06	-0.19	-0.17	-0.19	0.31	-0.27	-0.02	0.08	0.13

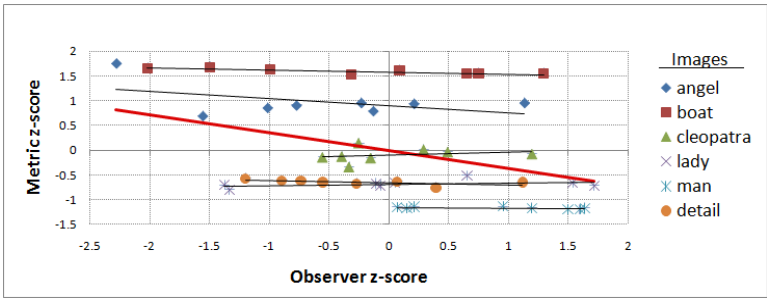


Figure 40:
Observer z-score plotted against DN [13] z-score for all images in the database. PCC = -0.36. The red linear regression line is determined by all data points; the black linear regression lines is determined by data points corresponding to individual images

By doing so, it is possible to observe that the DN metric provides a good fit for data corresponding to individual images.

Figure 41, Figure 42 and Figure 43 depict the relation between z-scores for the GD_{out}, BorSal and StructBorSal metrics, respectively, and observers z-scores. As in the previous case, where the performance of the DN metric was studied, data points associated to the latter three are also widely spread, resulting in low correlation values. This is a trend observed for all considered metrics and can be explained by the scale differences between the images in the database.

Thus, it can be concluded that, for the considered image database, the objective evaluation methods selected for this study can not accurately predict perceived overall quality. Worth noticing is that the best performance was obtained for a saliency based metric, that was specifically designed for inpainting quality evaluation, and that among these type of metrics, the second best performance was observed for one of the metrics proposed in this thesis (StructBorSal).

7.2 Image-wise evaluation

Research in the field of image quality evaluation has shown that the performance of some metrics can be influenced by characteristics of the images for which they are applied. Thus, this section will investigate the performance of the selected metrics with respect to individual images. The correlation between objective and subjective z-scores was again used as a performance measure. Accordingly, data in Table 6 corresponds to the PCC and SCC for the 14 metrics considered in this study, applied to the six test images used. For each image (i.e. *boat*, *clopatra*, *angel*, *lady*, *man*, *detail*) the values listed in the columns of Table 6 indicate the correlation between objective and subjective z-scores corresponding to the eight reproductions (i.e. inpainted images inpainted with the eight algorithms).

The obtained results indicate a great variation between the scores obtained for the 14 metrics corresponding to a single image (column data in Table 6), but also between

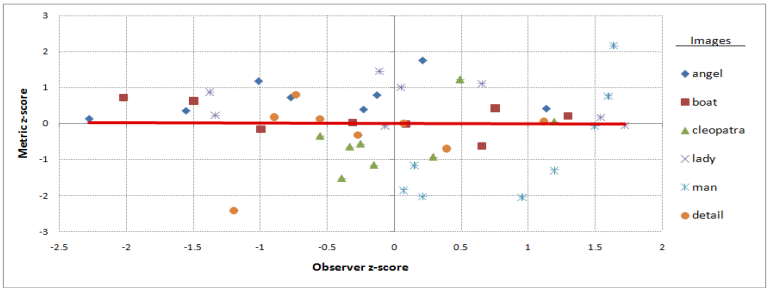


Figure 41:
Observer z-score plotted against GD_{out} [13] z-score for all images in the database, with a linear regression line. PCC = -0.01. The GD_{out} metric shows the lowest correlation with perceived quality, similar to GD_{in}

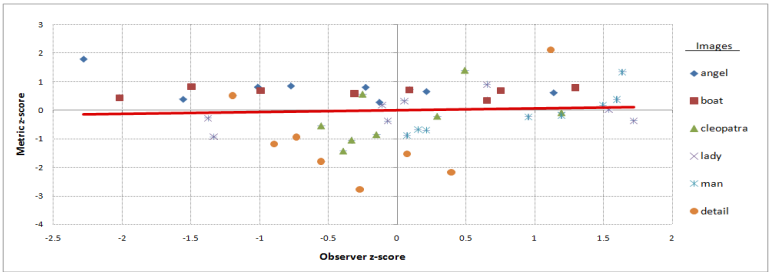


Figure 42:
Observer z-score plotted against BorSal z-score for all images in the database, with a linear regression line. PCC = 0.06. The PCC indicates a poor performance of the BorSal metric, but higher than GD_{in} and GD_{out} [12]

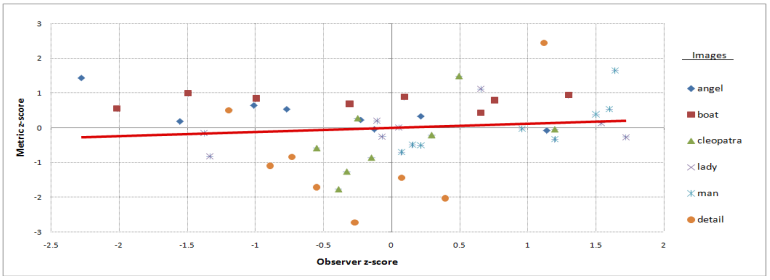


Figure 43:
Observer z-score plotted against StructBorSal z-score for all images in the database, with a linear regression line. PCC = 0.12. The StructBorSal metric outperforms the basic BorSal metric, the saliency based metrics GD_{in}, GD_{out} [12] and the ASVS [13] metric

Table 6: Comparison of metrics performance image-wise. The score for best performing metric is highlighted in bold font for each image and each correlation coefficient

Metric	boat		cleopatra		angel		lady		man		detail	
	PCC	SPP	PCC	SPP	PCC	SPP	PCC	SPP	PCC	SPP	PCC	SPP
MSE	-0.63	-0.31	-0.48	-0.67	-0.37	0.33	0.09	0.17	0.57	0.28	**-0.98**	-0.93
PSNR	0.62	0.31	0.53	-0.67	0.20	0.33	-0.13	0.17	-0.37	-0.57	0.90	0.98
DN	**-0.81**	**-0.66**	0.25	0.43	-0.46	0.11	0.39	0.29	-0.54	-0.46	-0.59	-0.77
ASVS	-0.77	-0.64	0.51	0.50	-0.48	0.07	0.39	0.29	0.88	0.92	0.27	-0.19
GD_{in}	-0.60	-0.64	0.49	0.43	-0.52	-0.02	**0.51**	**0.60**	**0.89**	0.93	0.27	-0.19
GD_{out}	-0.48	-0.36	0.53	0.43	0.34	0.55	-0.28	-0.29	0.76	0.71	0.26	-0.14
SCIELAB	-0.32	-0.21	**-0.62**	**-0.88**	-0.29	0.10	-0.16	-0.12	-0.06	-0.19	-0.90	**-1**
SHAME	-0.43	-0.29	-0.30	-0.40	-0.55	-0.14	-0.14	-0.24	-0.15	-0.31	-0.84	-0.99
SHAME-II	-0.35	-0.40	-0.42	-0.69	-0.55	-0.14	-0.15	0	0.09	-0.14	-0.89	-0.79
$SSIM_{IPT}$	-0.19	-0.26	0.58	0.76	-0.60	-0.45	0.03	-0.19	-0.16	-0.19	0.87	0.95
ABF	-0.39	-0.43	-0.58	0.76	-0.51	-0.14	-0.22	-0.55	-0.09	-0.19	-0.90	-0.71
VSNR	0.49	0.60	0.20	0.14	0.70	-0.45	0	0.12	-0.66	-0.81	-0.82	-0.93
BorSal	0.09	0.07	0.48	0.64	-0.57	-0.50	0.37	0.29	0.87	**0.98**	0.27	-0.19
StructBorSal	0.05	0.07	0.54	0.64	**-0.77**	**-0.67**	0.40	0.19	0.82	0.95	0.31	-0.19

scores of the same metric for different images (row data in Table 6). However, for four out of six images (i.e. *boat*, *angel*, *lady*, *man*) results show that inpainting specific metrics perform better than general IQ metrics. The DN [13], GD_{in} [12], BorSal and StructBorSal metrics provide good correlation with perceived quality, whereas only the S-CIELAB [34] and MSE among the IQ metrics indicate a better performance, in the specific cases of *cleopatra* and *detail* images, respectively. S-CIELAB is a metric that accounts for the significant color differences between the inpainted images and the references.

In the *detail* image, the inpainted area is narrow, thus, post-inpainting changes in saliency will not be significant, yielding low correlation values for the inpainting-specific metrics. The small post-inpainting change in saliency is due to the fact that, if existent, artifacts will be more difficult to spot in a very narrow area. Moreover, most of the algorithms can successfully fill in small gaps, resulting thus in small perceived differences in the quality of the reproductions of the *detail* image. General IQ metrics, based on color differences (i.e. S-CIELAB, ABF, SHAME, SHAME-II) or contrast thresholds (i.e. VSNR) can better assess these differences, outperforming inpainting-specific metrics, which rely on visual saliency maps.

Considering the great variance between metrics and between images, graphically illustrated in Figure 44 and Figure 45, it can be concluded that both inpainting-specific and general IQ metrics are image-dependent and thus their performance will depend on the characteristics of the image for which they are applied. One metric can not be used to predict perceived quality, regardless of the considered image. Several metrics that evaluate different aspects and focused on detection of possible post-inpainting artifacts (i.e. poor structure reconstruction, colour differences, blur, *etc.*) could be used for the same image, in order to obtain better correlation between calculated quality and perceived quality.

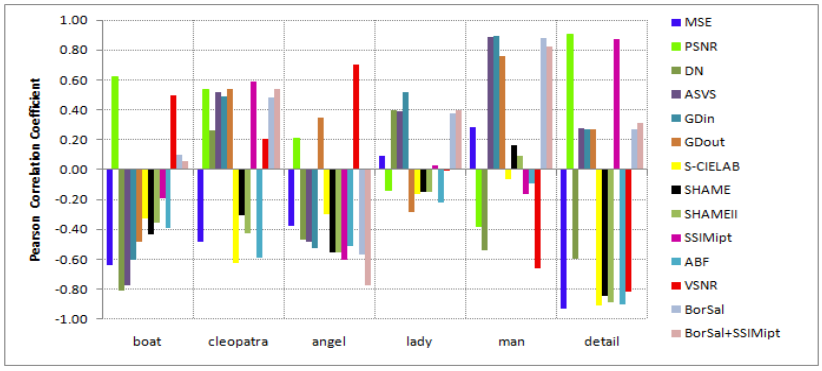


Figure 44: Pearson correlation indicating image-wise performance of the metrics

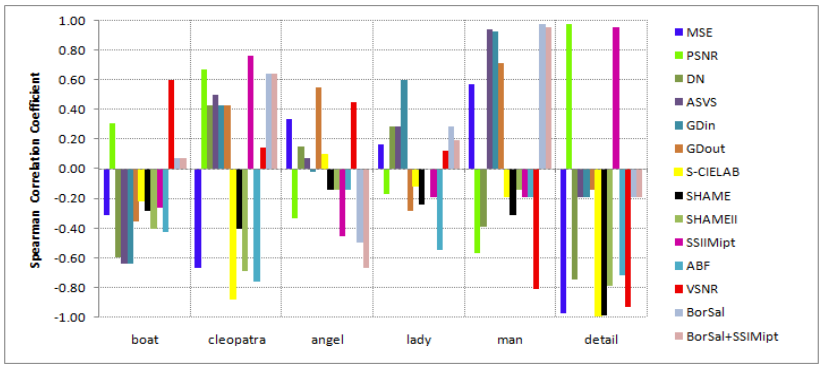


Figure 45:
Spearman correlation indicating image-wise performance of the metrics in terms of rank order

8 Conclusion

The goal of the current research was to qualitatively evaluate the performance of a set of representative digital inpainting algorithms and to develop, use and evaluate objective methods for quality assessment in the context of digital inpainting for artwork.

Following a literature review of existing digital image inpainting algorithms, given in Chapter 2, a set of eight algorithms have been selected for further evaluation. These algorithms were based on different approaches, and included basic methods, as well as more recent and improved methods. evaluating perceptual quality.

It was shown that for the evaluation of inpainting algorithms, researchers frequently resort to human visual comparisons. This method of evaluation gives a precise way to quantify inpainting quality, but it is time consuming and resource demanding. Thus, objective methods for quality evaluation are preferred. Chapter 3 included an overview of existing objective methods for evaluating the performance of inpainting algorithms, showing that a method that could accurately predict perceived inpainting quality was missing. Consequently, state of the art metrics from the more extensively researched field of image quality evaluation were also considered for further inpainting evaluation and presented in Chapter 3. Furthermore, in an attempt to compensate for the lack of inpainting-specific quality metrics, two novel metrics were proposed, extending and improving existent metrics.

Chapter 5 introduced the evaluation methodology and a new technique adopted for creating the image database used as a basis for obtaining test images. The database consisted of modified digitally acquired images of real damaged paintings. The applicability of digital inpainting for artwork restoration was thus demonstrated.

Following the evaluation methodology in Chapter 5, a psychophysical experiment was conducted, which asked human observers to judge the quality of the inpainting for a set of 54 images. The analysis of the obtained perceptual data showed no significant differences among the three categories of observers considered (i.e. naive, expert, professional restorer). Based on the ratings given by the observers to inpainted images, a ranking of the eight inpainting algorithms was established. This ranking showed that, generally, exemplar based methods outperform PDE based methods. The latter verifies the theoretical analysis of digital inpainting methods, as exemplar based inpainting algorithms are able to reconstruct both texture and structure in an image, as opposed to PDE based algorithms, which are based on diffusion and produce blur.

In Chapter 7 two novel and four existent inpainting-specific metrics, along with eight general IQ metrics, previously presented in Chapter 3, were evaluated against experimental data from human observers. To our knowledge, this evaluation is one of the most extensive carried out in the literature, with respect to inpainting quality. The obtained results showed that none of the considered metrics correlated well with perceived overall inpainting quality. However, the scores obtained for the two metrics proposed in this thesis indicate a better performance when compared to some of the other existent inpainting-specific quality metrics. An image-wise evaluation of the metrics was also conducted, showing that all metrics are image dependent. Results indicate that

inpainting-specific metrics outperform image quality metrics when applied for images with small-sized gaps or for images that don't require complex structure reconstruction.

In conclusion, the current research has proposed the use of digital image inpainting for artwork restoration purposes and investigated the performance of a set of representative algorithms in order to determine which of them provides the best results. New metrics have been proposed for the objective evaluation of inpainting quality and their correlation with perceived quality compared to other existent inpainting-specific and general IQ metrics.

9 Future work

In the course of the research carried out for this thesis a number of possible directions for further research have been identified. Relating to different parts of this thesis, they can be divided into two separate sections.

9.1 Digital inpainting

Despite the extensive research carried out in the field of digital image inpainting, there is still room for improvements. Currently, algorithms consider binary masks for the damaged area of an image. However, in the field of artwork restoration, some underlying information may still exist in the degraded area and should not be completely disregarded, by using a binary mask. Algorithms able to perform inpainting from partial degradation would represent an important step towards achieving digitally inpainted images closer to the manually restored works of art.

In the recent years some attempts have been made to extend the inpainting methodology from two dimensions to three dimensions. This would allow developing applications for restoration of damaged monuments and historical artifacts. However, more research is needed in this direction before notable results to be observed.

Experimenting with high resolution images should also be considered for further research. The main issue when dealing with high resolution images as input for inpainting algorithms refers to a significant increase in running time (up to several hours). However, most of the existent algorithms were not designed for high resolution images. Thus, developing new algorithms, able to inpaint high resolution images would pottentially increase the quality of the inpainting.

9.2 Inpainting quality evaluation

Regarding the evaluation of inpainting quality, further work should be focused on finding a metric that does not depend on a reference image and that correlates better with perceived quality.

Conducting a psychophysical experiment that considers more recently proposed inpainting algorithms, as well as quality metrics, could also significantly contribute to a better understanding of the deficiencies of current methods. Identifying these deficiencies would make space for further improvements.

Bibliography

[1] Pedersen, M. 2011. Image quality metrics for the evaluation of printing workflows.

[2] Criminisi, A., Pérez, P., & Toyama, K. September 2004. Region filling and object removal by exemplar-based image inpainting. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 13(9), 1200–12.

[3] Zhou, J. & Kelly, A. R. 2010. Image inpainting based on local optimization. *International Conference on Patteren Recongnition (ICPR)*.

[4] Tschumperle, D. & Deriche, R. 2005. Vector-valued image regularization with PDE's : A common framework for different applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (27(4)), 506–517.

[5] Bertalmio, M., Sapiro, G., Caselles, V., & C.Ballester. 2000. Image inpainting. In *Proceedings of the 27th annual conference on Computergraphics and interactive techniques, SIGGRAPH '00*, 417–424. ACM Press/Addison-Wesley Publishing Co.

[6] Bornemann, F. & März, T. July 2007. Fast Image Inpainting Based on Coherence Transport. *J. Math. Imaging Vision*, 28(3), 259–278.

[7] Telea, A. 2004. An Image Inpainting Technique Based on the Fast Marching Method. *J. Graphics Tools*, 9(1), 23–24.

[8] Barnes, C., Shechtman, E., Finkelstein, A., & Goldman, D. B. 2009. Patchmatch: a randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics (TOG)*, 28(3).

[9] Oliveira, M. M., Bowen, B., McKenna, R., & Chang, Y. S. 2001. Fast Digital Image Inpainting. In *Proceedings of the International Conference on Visualization, Imaging and Image Processing (VIIP 2001*, 261–266. ACTA Press.

[10] Emile, M. 1976. *The Restorer's Handbook of Easel Painting*. Van Nostrand Reinhold.

[11] Bertalmio, M. 2001. Processing of Flat and Non-Flat Image Information on Arbitrary Manifolds Using Partial Differential Equations.

[12] Mahalingam, V. V. *Digital inpainting algorithms and evaluation*. PhD thesis, University of Kentucky, 2010.

[13] Ardis, P. & Singhal, A. 2009. Visual salience metrics for image inpainting. In *SPIE*, number 7257.

[14] Nitzberg, M., M. D. S. 1993. T.: Filtering, segmentation and depth (lecture notes in computer science vol. 662). springer-verlag berlin heidelberg.

[15] Caselles, V., Masnou, S., Morel, J. M., & Sbert, C. 1998. Image interpolation.

[16] C. Ballester, M. Bertalmio, V. C. G. S. J. V. 2001. Filling-in by joint interpolation of vector fields and grey levels. *IEEE Transactions on Image Processing*, 10(8), 1200–1211.

[17] Bertalmio, M., Bertozzi, A. L., & Sapiro, G. 2001. Navier-stokes, fluid dynamics, and image and video inpainting. *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, 355–362.

[18] Schonlieb, C. B. 2009. Modern PDE Techniques for Image Inpainting.

[19] Chan, T. & Shen, J. Mathematical Models for Local Deterministic Inpaintings. Technical report, UCLA, 2000.

[20] Khaitan, A. Automatic Image Inpainting. Technical report, The ICFAI University Dehradun, 2011.

[21] Sangeetha, K., Sengottuvelan, P., & Balamurugan, E. 2011. Comparative Analysis and Evaluation of Image Imprinting Algorithms. *Journal of Information Engineering and Applications*, 1(5), 506–517.

[22] Sangeetha, K., Sengottuvelan, P., & Balamurugan, E. 2011. A Comparative Analysis of Exemplar Based Image Inpainting Algorithms. *European Journal of Scientific Research*, 60(3), 316–325.

[23] Mai, W. & Chen, X. Image Inpainter. Technical report, The University of New South Wales, Sydney, Australia, 2010.

[24] Hays, J. & Efros, A. A. 2007. Scene completion using millions of photographs. *Proceeding of the ACM SIGGRAPH*, 26.

[25] Simakov, D., Caspi, Y., Shechtman, E., & Irani, M. 2008. Summarizing visual data using bidirectional similarity.

[26] Wexler, Y., Shechtman, E., & Irani, M. march 2007. Space-time completion of video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(3), 463 –476.

[27] Crosby, P. B. Sep. 1982. The puzzlement of quality.

[28] Janssen, T. J. 1999. Computational Image Quality.

[29] Jacobson, R. E. 1995. An evaluation of image quality metrics.

[30] Yendrikhovskij, S. N. 1998. Color reproduction and the naturalness constraint.

[31] Ardis, P. A., Brown, C. M., & Singhal, A. 2010. Inpainting quality assessment. *J. Electronic Imaging*.

[32] Walther, D. *Interactions of visual attention and object recognition: computational modeling, algorithms, and psychophysics*. PhD thesis, California Institute of Technology, Pasadena, CA,, 2006.

[33] Chandler, D. M. & Hemami, S. S. Sep 2007. VSNR: A wavelet-based visual signal to noise ratio for natural imges. (16(9):2284-2298).

[34] Zhang, X. & A.Wandell, B. 1996. A spatial extension of cielab for digital color image reproduction. 731–734.

[35] Bonnier, N., Schmitt, F., Brettel, H., & Berche, S. Nov. 2006. Evaluation of spatial gamut mapping algorithms. 56–61.

[36] Z.Wang, Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. (13(4):600-612).

[37] Pedersen, M. & Hardeberg, J. Y. 2009. Computational color imaging. chapter A New Spatial Hue Angle Metric for Perceptual Image Difference, 81–90. Springer-Verlag, Berlin, Heidelberg.

[38] Hong, G. & Luo, M. R. 2006. New algorithm for calculating perceived colour difference of images. (54(2):86-91).

[39] CIE. *Colorimetry. Technical Report 15*, 2004.

[40] Johnson, G. M. & Fairchild, M. D. Nov. 2001. Darwinism of color image difference models. 108–112.

[41] Z.Wang & Hardeberg, J. Y. Nov. 2009. An adaptive bilateral filter for predicting color image difference. 27–31.

[42] Mai, W. & Chen, X. 2010. Image inpainter, project for comp9517 at unsw.

[43] Tschumperle, D. & Deriche, R. 2008. G'MIC, GREYC's Magic Image Converter.

[44] Bornemann, F. & Marz, T. 2007. Fast image inpainting based on coherence transport.

[45] Kendall, M. G., Stuart, A., & Ord, J. K. 1991. Kendall's advanced theory of statistics:classical inference and relationship.

[46] 2012. Methodology for the subjective assessment of the quality of television pictures.

[47] International Telecommunication Union. 1996. Recommendation itu-r p.800:methods for subjective determination of transmission quality.

[48] Ninassi, A., Meur, O. L., Callet, P. L., & Barba, D. March 2008. On the performance of human visual system based image quality assessment metric using wavelet domain. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 6806 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*.

[49] Milton, J. S. & Arnold, J. C. *Introduction to Probability and Statistics*. 1990.

[50] Moroney, N. 2003. Unconstrained web-based color naming experiment, in color imaging: Device-dependent color, color hardcopy and graphic arts. In *VIII, Reiner Eschbach, Gabriel Marcu, Editors, Proc. of the SPIE*.

[51] Sprow, I., Baranczuk, Z., Stamm, T., & Zolliker, P. 2009. Web-based psychometric evaluation of image quality. 7242.

A Parameters Choice and Test Images

Table 7, Table 8, Table 9, Table 10, Table 11, Table 12, Table 13 and Table 14, respectively, list the parameters used in order to obtain the test images included in the pshychophysical experiment for evaluation of digital inpainting algorithms. These parameters were determined through a manual parameter-tuning process.

Fig. 46, Fig. 47, Fig. 48, Fig. 49, Fig. 50, Fig. 51, Fig. 52 and Fig. 53 illustrate the results obtained with the selected parameters, for each of the eight inpainting algorithms.

Table 7: Parameters chosen for test images by Criminisi *et al.* [2]

Image	Iterations	Patch Size	Boundary Radius	Running time
angel	1000	9	51	22'01"
boat	1000	9	33	9'12"
cleopatra	1000	9	30	9'41"
lady	1000	5	15	1'30"
man	1000	5	27	13"
detail	1000	9	30	7"

Table 8: Parameters chosen for test images by Zhou and Kelly [3]

Image	Iterations	Patch Size	Boundary Radius	Running time
angel	500	9	30	20'
boat	1000	9	54	134'14"
cleopatra	1000	9	30	29'47"
lady	1000	7	33	2'40"
man	1000	5	27	1'25"
detail	1000	9	30	8'29"

Table 9: Parameters chosen for test images obtained with the method proposed by Tschumperle and Deriche [4]

Image	Global iterations	Local iterations	Gradient smoothness	Tensor smoothness	Time step	Running time
angel	30	30	7.42	0.11	4.55	1'16"
boat	30	1	14.08	1.26	0.31	34"
cleopatra	30	2	0.18	2.2	15.9	33"
lady	30	1	9.64	2.69	0.72	8"
man	30	16	9.64	2.15	0.72	6"
detail	30	2	0.18	2.2	15.9	9"

Table 10: Parameters chosen for test images obtained with the method proposed by Bertalmio *et al.* [5]

Image	Inpainting steps	Diffusion steps	Climb rate	Climb decay	Running time
angel	15	4	0.02	0.989	3'01"
boat	50	4	0.02	0.989	1'52"
cleopatra	50	1	0.02	0.989	2'17"
lady	50	2	0.02	0.989	1'20"
man	50	4	0.02	0.989	1'22"
detail	50	1	0.02	0.989	1'57"

Table 11: Parameters chosen for test images obtained with the method proposed by Bornemann and Marz [6]

Image	Averaging radius	Sharpness parameter	Standard deviation for pre-smoothing	Standard deviation for post-smoothing	Running time
angel	10	10	1.4	5	8"
boat	15	25	1.4	5	7"
cleopatra	10	25	1.4	5	4"
lady	15	25	1.8	5	2"
man	15	25	1.8	5	3"
detail	10	25	1.4	5	3"

Table 12: Parameters chosen for test images obtained with the method proposed by Telea [7]

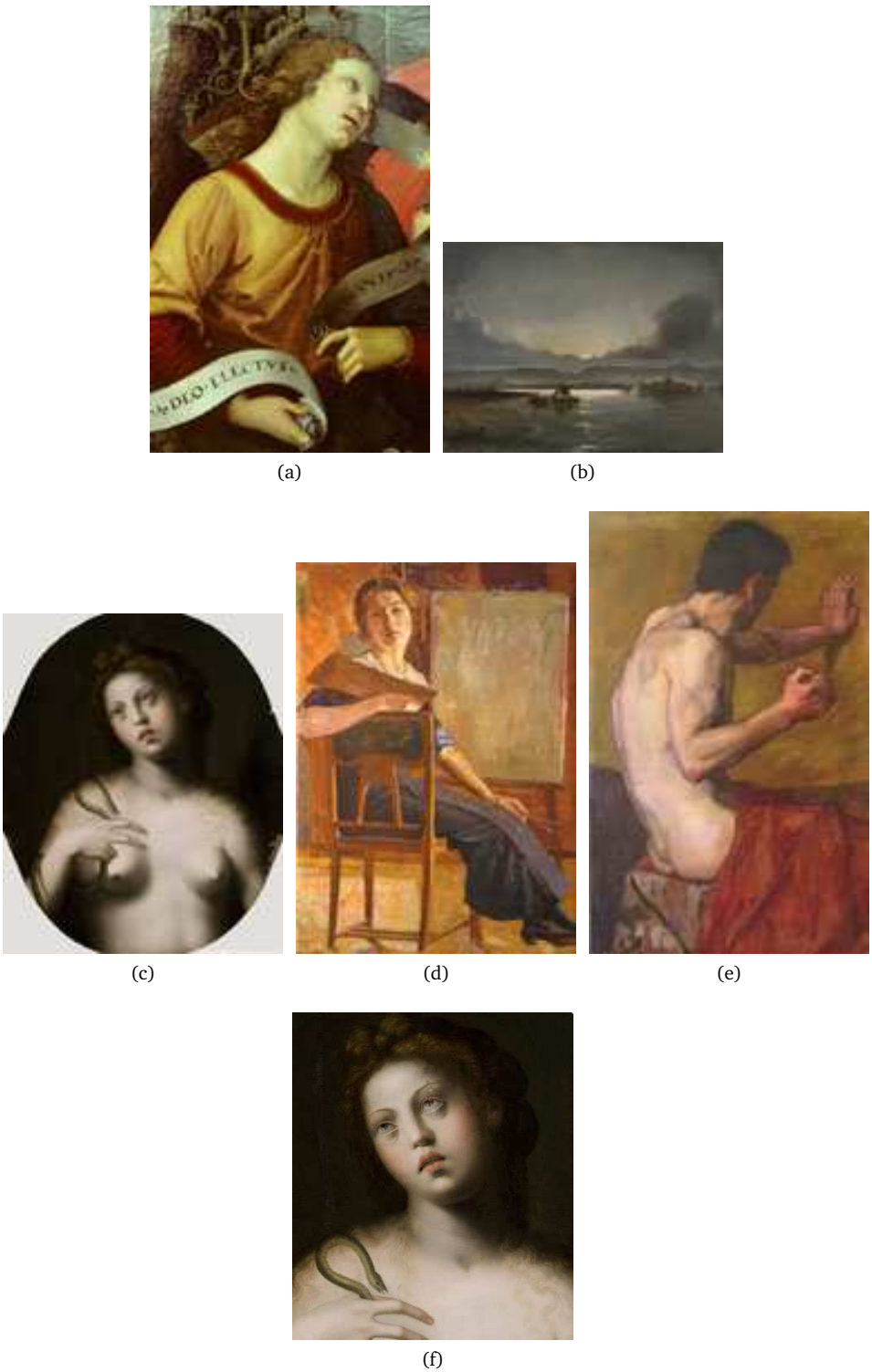
Image	Radius	Running time
angel	10	3"
boat	10	3"
cleopatra	2	1"
lady	3	1"
man	5	1"
detail	2	1"

Table 13: Parameters chosen for test images obtained with the method proposed by Barnes *et al.* [8]

Image	Running time
angel	8"
boat	5"
cleopatra	6"
lady	3"
man	3"
detail	2"

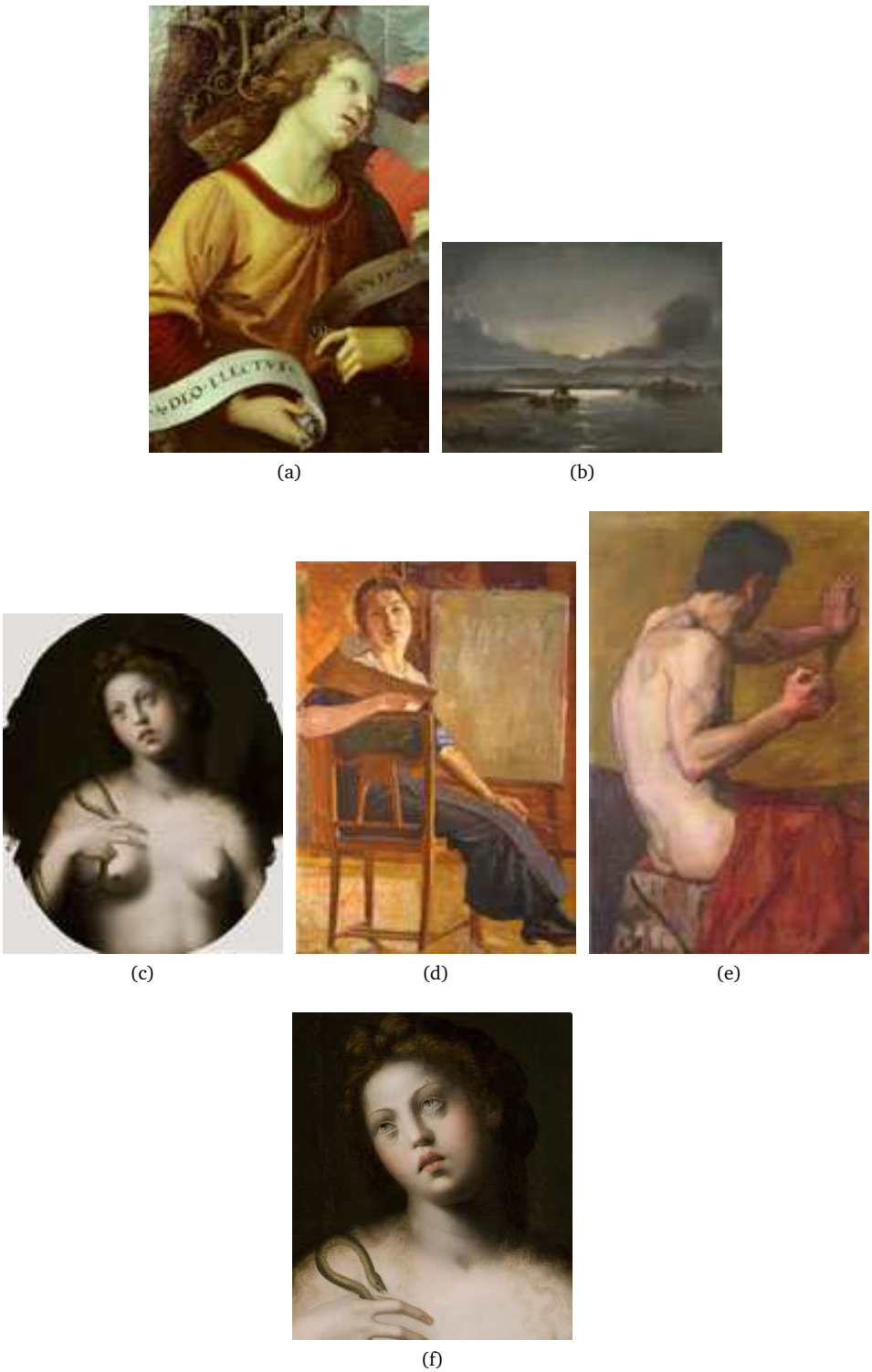
Table 14: Parameters chosen for test images obtained with the method proposed by Oliveira *et al.* [9]

Image	Kernel	Iterations	Running time
angel	1	52	1'
boat	1	200	17"
cleopatra	1	200	1'01"
lady	2	24	7"
man	1	200	3"
detail	1	200	2"



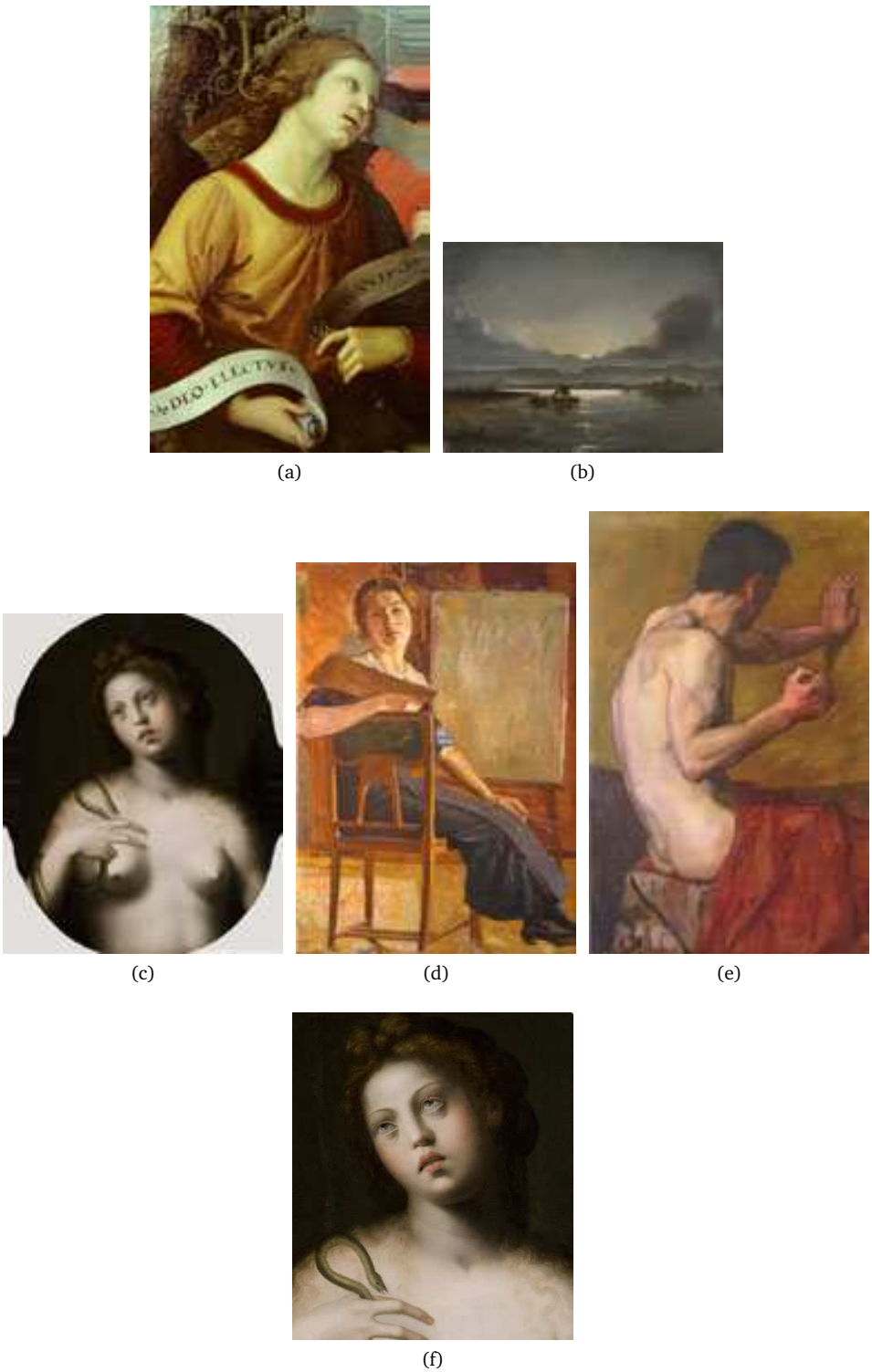
(a) (b) (c) (d) (e) (f)

Figure 46: *Test images as output of the algorithm introduced by Criminisi* et al. *[2]. See Table 7 for associated parameters. (a) angel (b) boat (c) cleopatra (d) lady (e) man (f) detail*



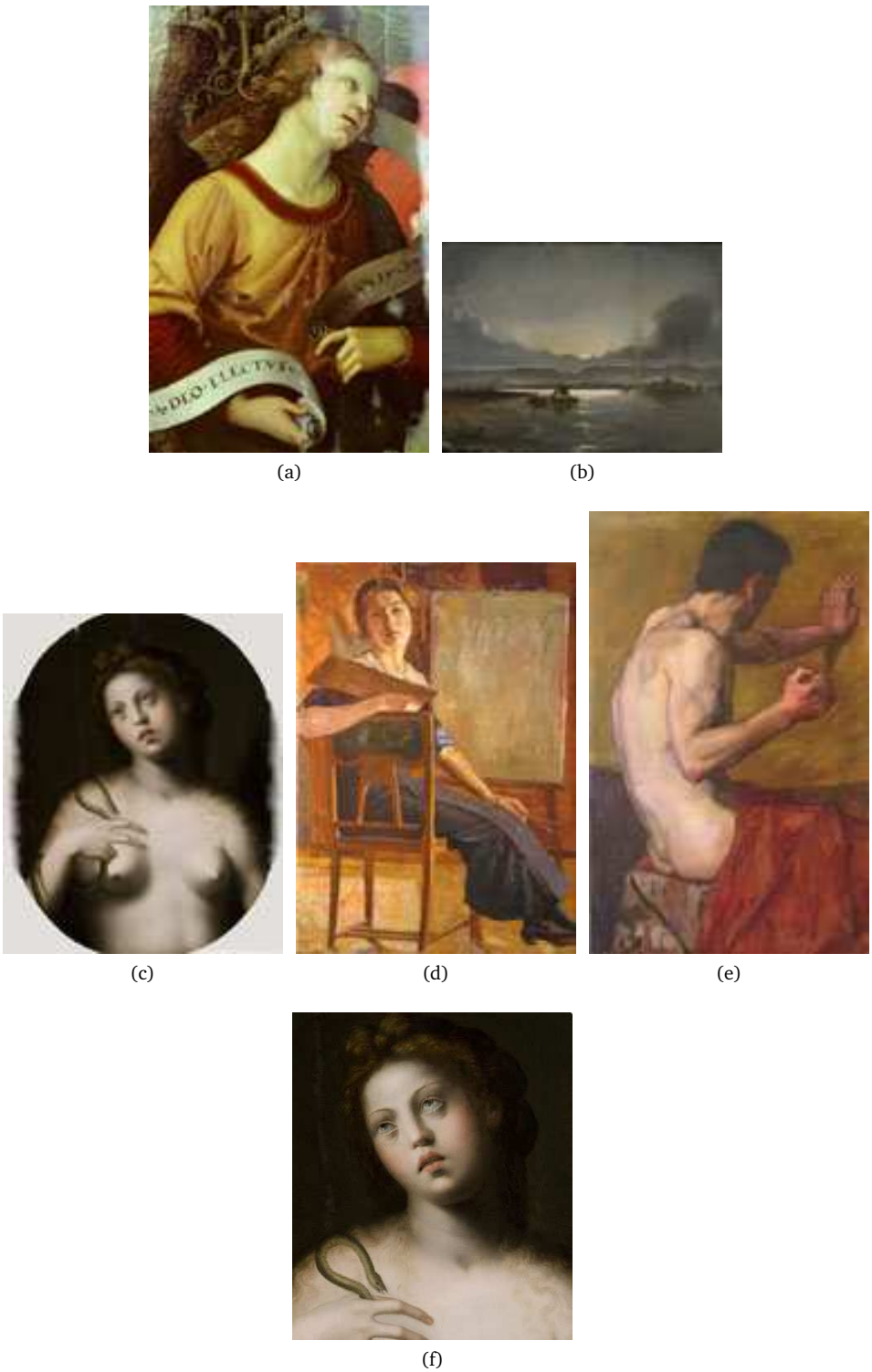
(a) (b) (c) (d) (e) (f)

Figure 47: *Test images as output of the algorithm introduced by Zhou and Kelly [3]. See Table 8 for associated parameters.(a) angel (b) boat (c) cleopatra (d) lady (e) man (f) detail*



(a) (b) (c) (d) (e) (f)

Figure 48: *Test images as output of the algorithm introduced by Tschumperle and Deriche [4]. See Table 9 for associated parameters.(a) angel (b) boat (c) cleopatra (d) lady (e) man (f) detail*



(a) (b) (c) (d) (e) (f)

Figure 49: *Test images as output of the algorithm introduced by Bertalmio* et al. *[5]. See Table 10 for associated parameters.(a) angel (b) boat (c) cleopatra (d) lady (e) man (f) detail*

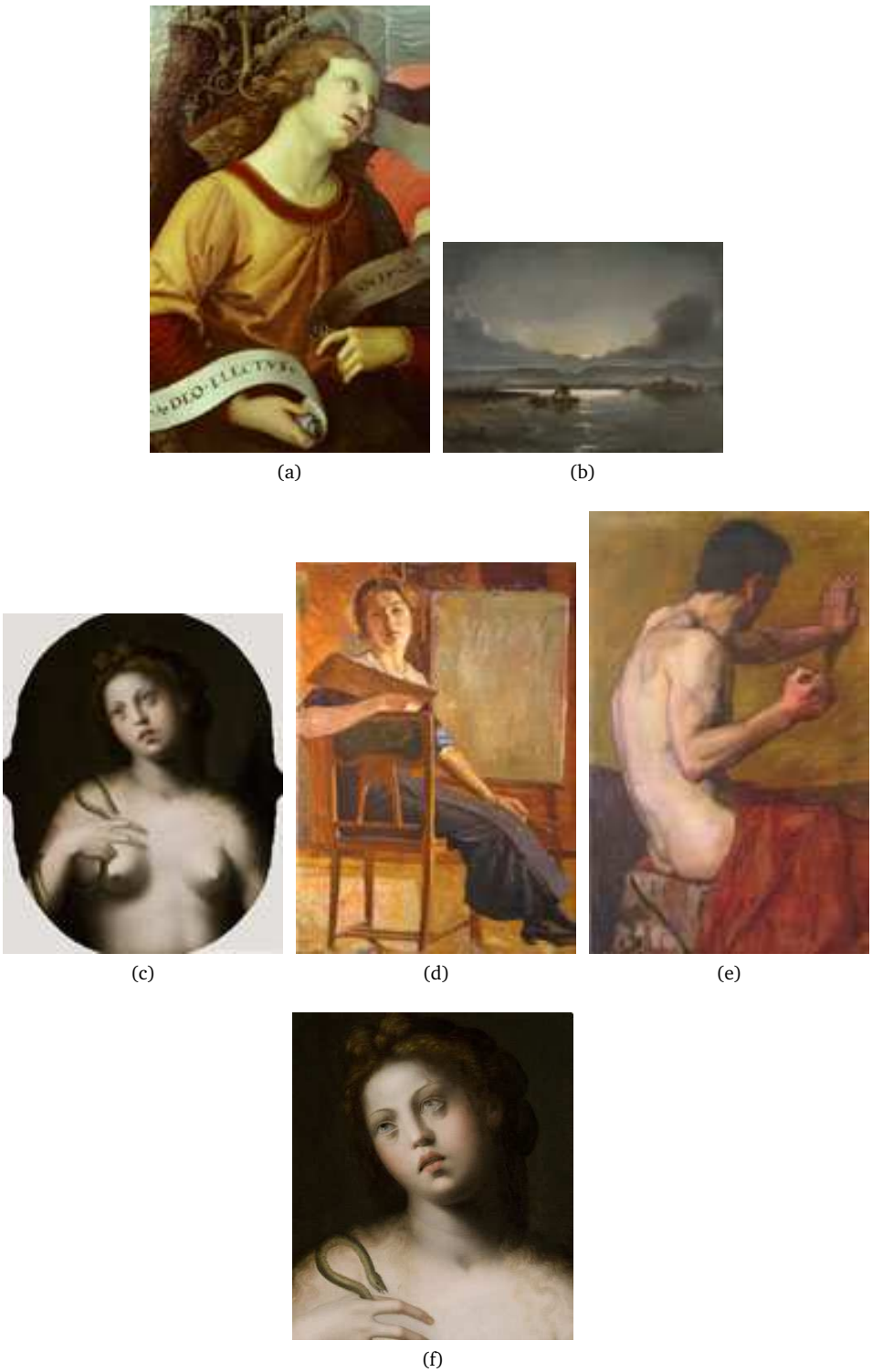
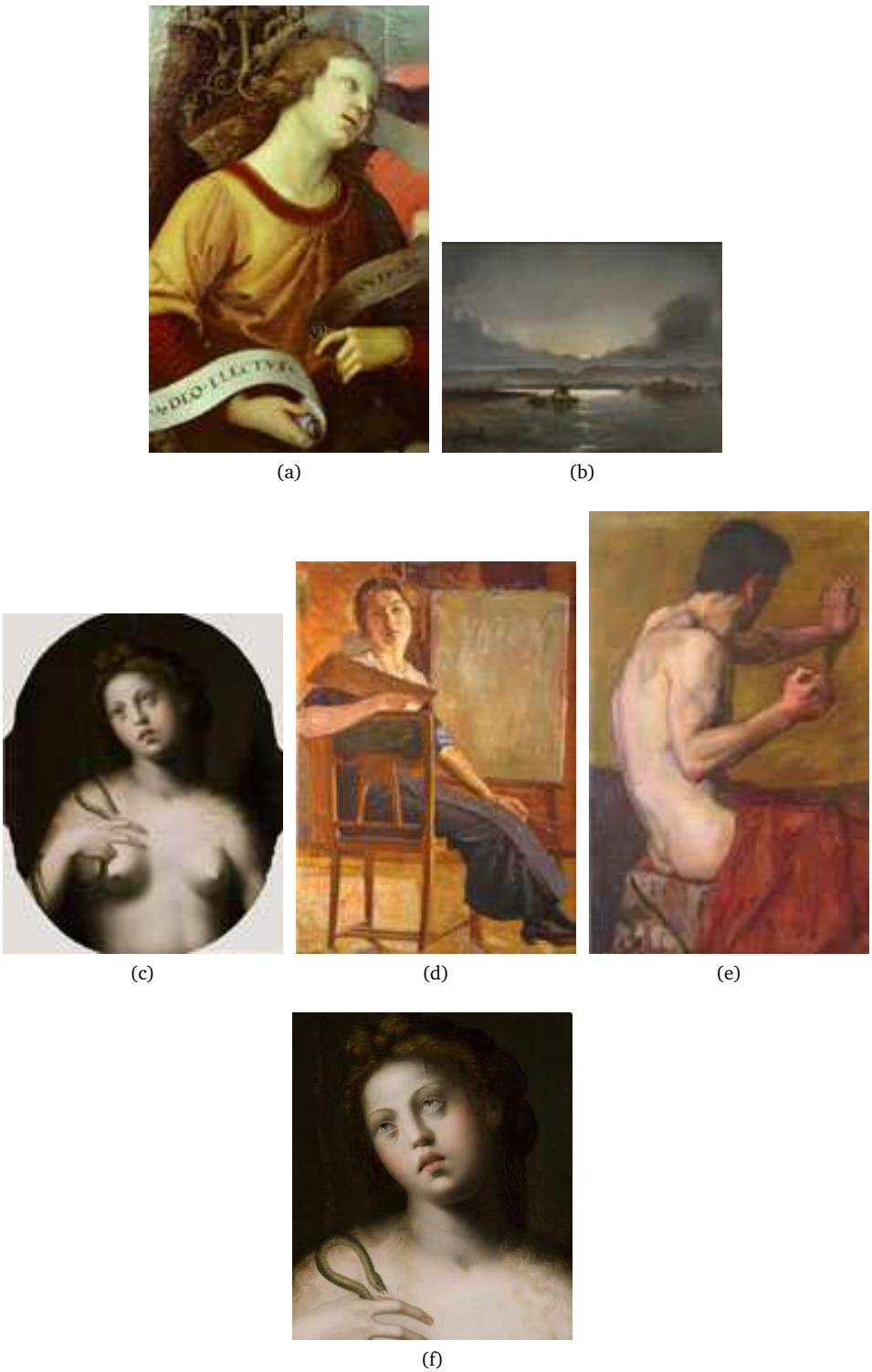
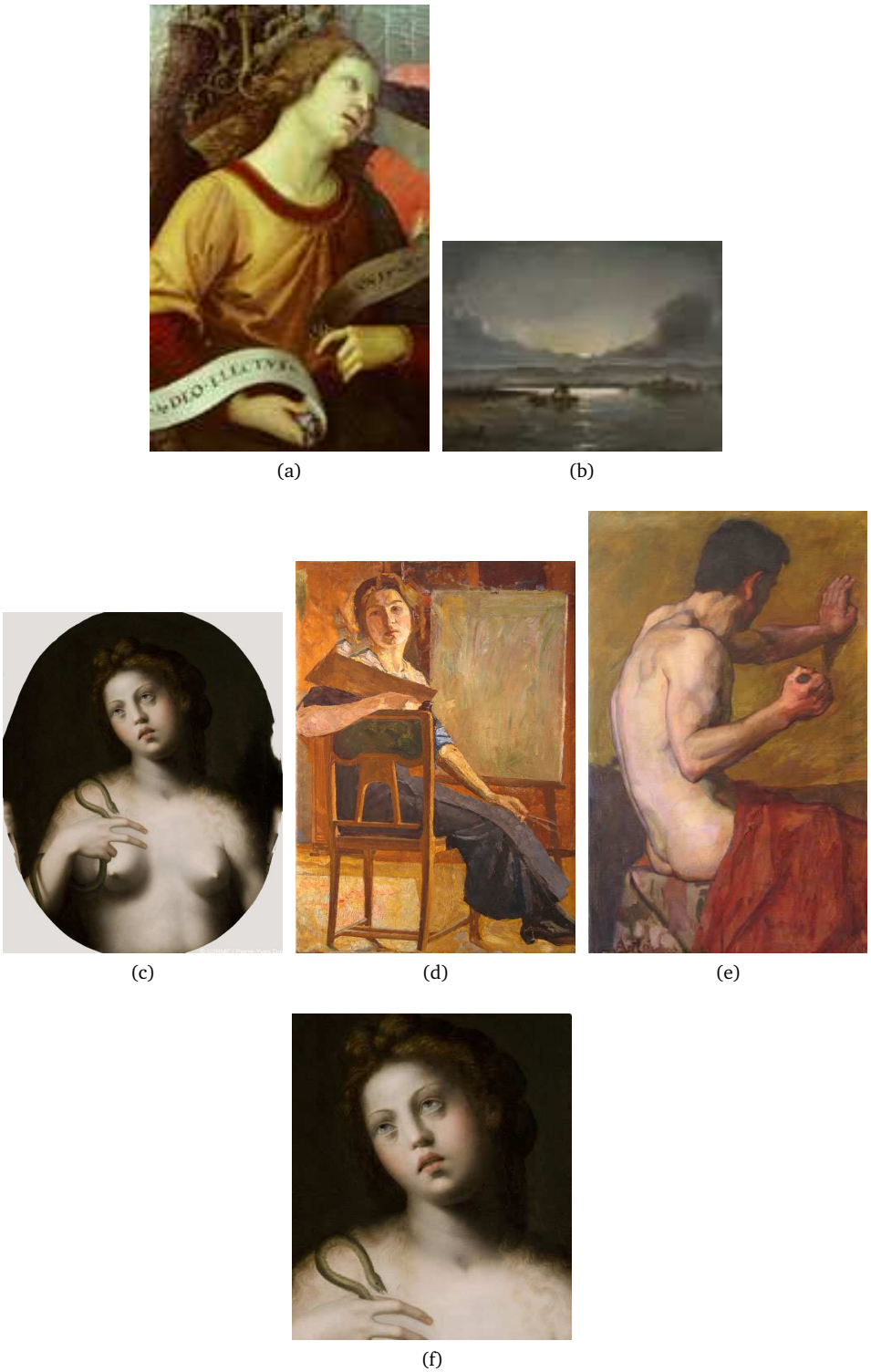


Figure 50: *Test images as output of the algorithm introduced by Bornemann and Marz [6]. See Table 11 for associated parameters. (a) angel (b) boat (c) cleopatra (d) lady (e) man (f) detail*



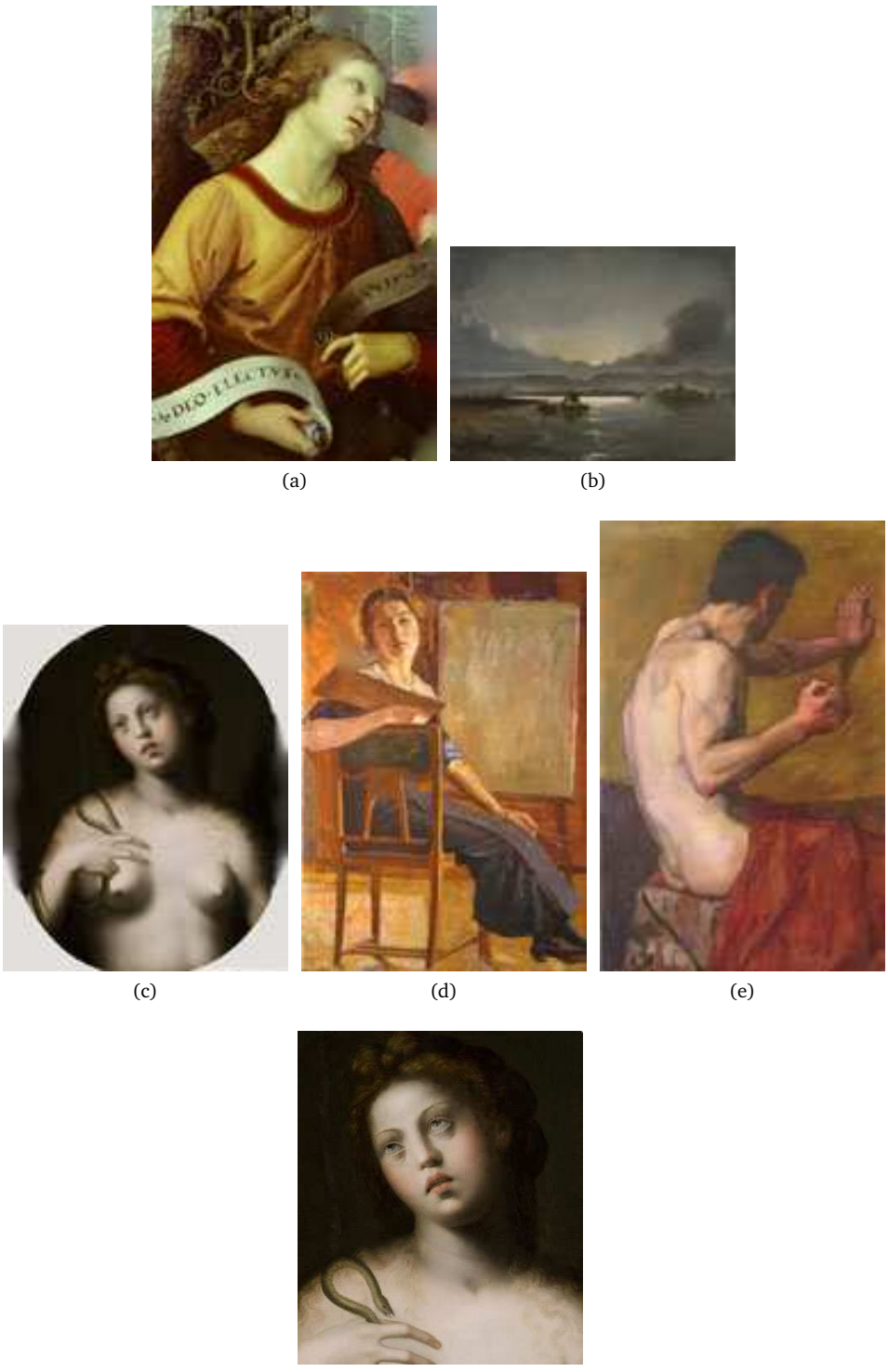
(a) (b) (c) (d) (e) (f)

Figure 51: *Test images as output of the algorithm introduced by Telea [7]. See Table 12 for associated parameters.(a) angel (b) boat (c) cleopatra (d) lady (e) man (f) detail*



(a) (b) (c) (d) (e) (f)

Figure 52: *Test images as output of the algorithm introduced by Barnes* et al. *[8]. See Table 13 for associated parameters.(a) angel (b) boat (c) cleopatra (d) lady (e) man (f) detail*



(a) (b) (c) (d) (e) (f)

Figure 53: *Test images as output of the algorithm introduced by Oliveira* et al. *[9]. See Table 14 for associated parameters. (a) angel (b) boat (c) cleopatra (d) lady (e) man (f) detail*

B Technical Paper

Evaluation of digital inpainting quality in the context of artwork restoration

Anonymous ECCV submission

***Abstract*—Improved digital image inpainting algorithms could provide substantial support for future artwork restoration. However, currently, there is an acknowledged lack of quantitative metrics for image inpainting evaluation. In this paper the performance of eight inpainting algorithms is first evaluated by means of a psychophysical experiment. The ranking of the algorithms thus obtained confirms that exemplar based methods generally outperform PDE based methods. Two novel inpainting quality metrics, proposed in this paper, eight general image quality metrics and four inpainting-specific metrics are then evaluated by validation against the perceptual data. Results show that no metric can adequately predict inpainting quality over the entire image database, and that the performance of the metrics is image-dependent.**

***Index Terms*—inpainting, psychophysical experiment, inpainting quality metrics, image quality metrics.**

I. Introduction

DIGITAL inpainting refers to techniques used to reconstruct areas of missing information in an image, by filling the gaps with visually plausible content. In the field of artwork restoration, inpainting algorithms can be employed for digital restoration, by reversing the damage (i.e. torn canvas, scratches, stains) in a painting converted to a digital form. Digital inpainting algorithms can be grouped into two main categories. Partial differential equation (PDE) based algorithms [1], [2], [3], [4] fill in gaps by extending isophote lines from the source region into the target region via diffusion. Their drawback consist of introducing blur artifacts that become more visible when inpainting larger areas. Exemplar-based inpainting algorithms [5], [6],[7] overcome this drawback by reconstructing large image regions from sample textures. Some approaches try to achieve better performance in terms of running time [8],[7],[3].

As the goal of inpainting is to reconstruct the damaged regions in a visually plausible way and a reference image might not always be available for comparison, inpainting quality evaluation is a challenging task, that has been only narrowly researched. Mahalingam [9] and Ardis *et al.* [10] propose the use of visual-saliency based metrics. However, these metrics are not commonly used by researchers to assess new inpainting techniques. Instead, qualitative human comparisons are currently and frequently used. Other image quality (IQ) metrics simulating the human visual system (HVS) and taking into account structural information in an image might be useful in the field of image inpainting.

This paper will evaluate the performance of eight representative inpainting algorithms (1-8) by means of a psychophysical experiment. The obtained perceptual data will be used to establish a ranking of the inpainting methods, described in Section 2. Based on the same data, the correlation between a selection of existing IQ metrics and perceived quality will be investigated. Furthermore, two novel metrics will be introduced in this paper and included in the evaluation.

II. Psychophysical Experiment for Subjective Rating

A. Image Database

A common practice when evaluating the performance of inpainting algorithms is to use predefined inpainting regions. In this paper, modified digitally acquired images of real damaged paintings will be used as test images. In the manual restoration process there are a number of steps that proceed the filling in of missing areas. Instead of completely simulating gaps, this paper proposes the simulation of these steps, using the original degraded and manually restored versions of a painting. The resultant image replicates the regions containing missing or corrupted information from the digitally acquired image but has identical content with the manually restored image in the region outside the area to be inpainted. Figure 1 shows an example of a modified digitally acquired painting, as a result of the simulation process.

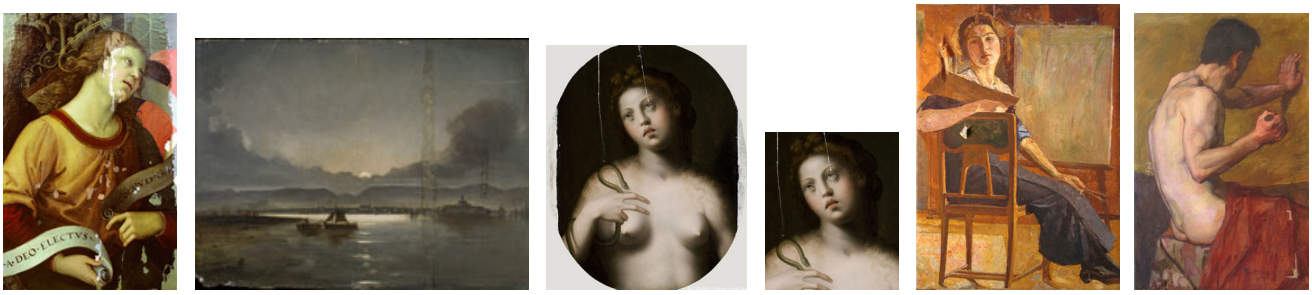
Six test images (Figure 2) have been chosen for the psychophysical experiment. For each of them, eight inpainted images corresponding to the algorithms proposed by Bertalmio *et al.* [1], Telea [2], Tschumperle and Deriche [3], Bornemann and März [4], Criminisi *et al.* [5], Zhou and Kelly [6], Barnes *et al.* [7] and Oliveira *et al.* [8] have been included in the database. Additionally, for each of the test images, the digitally acquired image of the manually restored painting (further on referred to as *manually inpainted image*) has been considered. The manually inpainted images have been included to verify the reliability of the observers, by checking their behaviour when degraded/manually inpainted image pairs are presented. Finally, the image database used for the experiment consisted of 54 images.

B. Experimental Setup

The experiment was carried out as a web-based experiment. Observers were presented with a pair of two images at a time and asked to judge the overall quality of the inpainted image using the ITU-R five grade quality scale [11] labelled with the adjectives: *Excellent*, *Good*, *Fair*, *Poor*, *Bad*. Figure 3 shows an example of such an image pair presented during the experiment. On the left the degraded image is shown and on the right the inpainted image. This positioning was

(a) (b) (c)

Fig. 1: *Preparing an image for inpainting. (a) Original degraded painting (b) Manually restored version (c) Modified image with artefacts (i.e. water stain) from the degraded image and similar appearance to the manually restored image. (a-b) - image courtesy of R. Pillay at C2RMF*



(a) (b) (c) (d) (e) (f)

Fig. 2: Modified digitally acquired paintings corresponding to: (a) "Ange" by Raphael (*angel*) (b) "Vue de Drontheim" by Peder Balke (*boat*) (c) "Cléopâtre" - anonymous (*cleopatra*) (d) Detail of "Cléopâtre" (*detail*) (e) "Self-portrait" by Margarete Depner (*lady*) (f) "Invention of painting" by Ariton (*man*). (a-d) - image courtesy of R. Pillay at C2RMF, (e,f) image courtesy of R. Tataru at Brasov Art Museum (Romania)

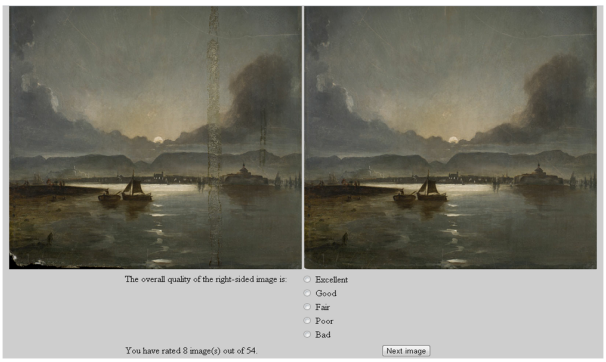


Fig. 3: Exemple of pair of images presented during the experiment. *(left)* Damaged painting *(right)* Inpainted version with the method proposed by Oliveira *et al.* [8]

kept throughout the whole experiment. Participants were asked to complete the experiment, which consisted of viewing and rating a total of 54 pairs of images. The screening of the observers was carried out by implementing the procedure recommended in Rec. ITU-R BT.500-13 [11]. Furthermore, observers were rejected if their scores indicated *Fair*, *Poor* or *Bad* quality for manually inpainted images, or if failing to complete the experiment. Consequently, from a total of 91, results of 22 participants were rejected and only 69 considered for further evaluation.

C. Psychophysical Results

Perceptual data obtained from the experiment was converted to z-scores, indicating the performance of the considered inpainting methods. Figure 4 gives a graphical representation of the obtained results. As expected, the manual inpainting method received the highest score, indicating the best perceived quality among the studied methods. Moreover, the high z-score value associated to it is an indicator of the large consensus among the participants about the high performance of this method. Among the analysed digital inpainting methods, the algorithm proposed by Barnes *et al.* [7] has the highest score. Worth noticing is the low visual difference between the inpainted images obtained by the inpainting algorithms proposed by Criminisi *et al.* [5], Zhou and Kelly [6] and Tschumperle and Deriche [3]. The latter is inferred from the overlapping confidence intervals corresponding to the three methods, and implies a difficult task for the observers to judge IQ.

Based on the obtained perceptual data, a ranking of the inpainting methods can be established (Figure 4). PDE-based methods are, generally, outperformed by exemplar-based methods. An exception is inpainting algorithm by Tschumperle and Deriche [3], which uses a vector valued regularization PDE.

III. Objective Quality Evaluation

Before discussing methods for objective quality evaluation, the notation convention must be defined. The area of missing information in an image will be referred to as the *gap* and

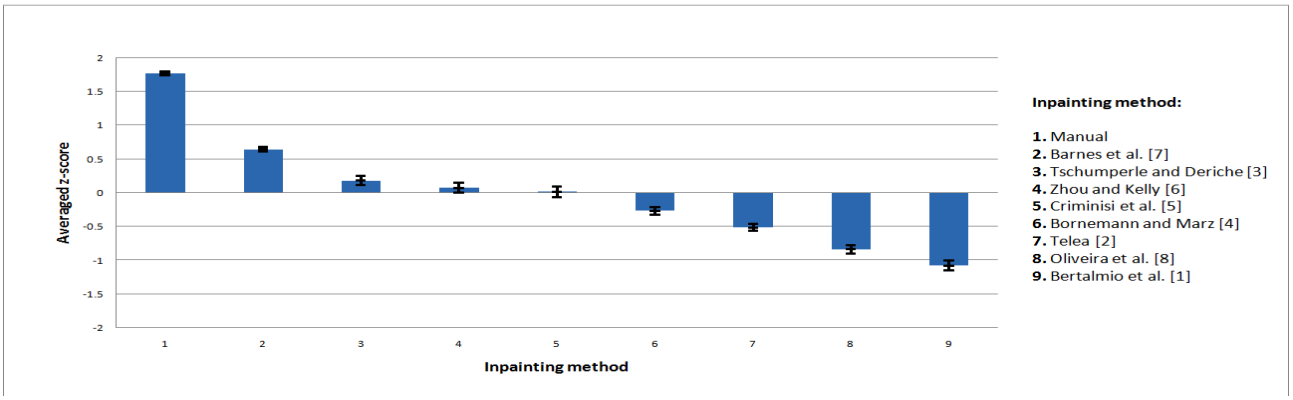


Fig. 4: Z-scores from observers based on 69 observers and 6 images, with 95% CI. The algorithm proposed by Barnes *et al.* [7] is rated as the best among digital inpainting methods

will be denoted by Ω. Its complementary area, referred to as the *source* region, will be denoted by Θ. Furthermore, when discussing saliency based metrics, the pre-inpainting saliency map intensity corresponding to a particular pixel, p, will be denoted by $S(p)$. Similarly, the corresponding post-inpainting saliency map intensity will be denoted by $S'(p)$.

A. ASVS and DN

Ardis *et al.* [10] define two types of observable artifacting in an inpainted image, referred to as *in-region* and *out-region*. The former accounts for artifacts belonging to the gap, Ω, while the *out-region* artifacting considers the complementary area, Θ. In order to quantify the quality of the reproduction and based on the two artifact classes, Ardis *et al.* propose two metrics, the *Average Squared Visual Saience (ASVS)* and the *Degree of Noticeability (DN)*, that relate the visual saliency map of an image with the perceived quality of the same image.

The first metric, ASVS, needs no reference image and equates to in-region artifacting. Thus, $in_{region} = ASVS$, where ASVS is computed as:

$$ASVS = (1/||\Omega||)(\sum_{p\in\Omega}(S'(p))^2). \quad (1)$$

In-region artifacting occurs when the inpainted pixels cause a modification in the flow of attention specific to an image, by increasing the saliency of the inpainted region. Ardis *et al.* [10] relate in-region artifacting to distinct colouration or structure that is introduced after inpainting, that can not be observed elsewhere in the image.

Out-region artifacting occurs when an inpainting algorithm fails to extend a locally repeating colour or structure inside the gap. The latter causes a decreased flow of attention for otherwise salient areas and increase of attention in the neighbouring area of the inpainting domain. Out-region artifacting is computed as $out_{region} = (\sum_{p\in\Theta}(S'(p) - S(p))^2)/|\Theta|$. Combining in- and out-region artifacting, the *Degree of Noticeability (DN)* metric is computed as:

$$DN = \frac{|\Omega|}{|\Omega|+|\Theta|} in_{region} + \frac{|\Theta|}{|\Omega|+|\Theta|} out_{region}. \quad (2)$$

As suggested by Ardis *et al.*, higher scores for ASVS and DN can be interpreted as an indicator of highly visible artifacts and thus a poor inpainting performance. The psychophysical study conducted by the authors proposing the metrics shows a good correlation between perceived and calculated quality. However, their findings required confirmation by further larger scale experiments, as they use only five observers.

B. $\mathrm{GD_{in}}$ and $\mathrm{GD_{out}}$

Mahalingam [9] proposes two visual saliency-based metrics for quantifying inpainting quality. He shows that if there is any change in the saliency maps corresponding to the inpainted and original image, this change is related to the perceptual quality of the inpainting.

According to Mahalingam [9] the gaze density within and outside the gap in an inpainted image is computed as:

$$GD_{in} = \sum_{p\in\Omega} S'(p), \quad and \quad GD_{out} = \sum_{p\in\Theta} S'(p). \quad (3)$$

The gaze density measures given in Equation 3 need to be normalized before indicating the presence of artifacts:

$$\overline{GD}_{in} = \frac{\sum_{p\in\Omega} S'(p)}{\sum_{p\in\Omega} S(p)}, \quad and \quad \overline{GD}_{out} = \frac{\sum_{p\in\Theta} S'(p)}{\sum_{p\in\Theta} S(p)}. \quad (4)$$

For simplicity, further discussion referring to the normalized metrics will use the notation $\mathrm{GD_{in}}$ and $\mathrm{GD_{out}}$.

Mahalingam [9] uses saliency maps generated from an eye tracking experiment. In this paper the SaliencyToolbox version 2.2 developed by Walther [12] is used to generate the saliency maps.

C. Proposed Metrics: BorSal and StructBorSal

Previous work by Mahalingam [9] and Ardis *et al.* [10] considers separately in- and out-region artifacting. In the latter, the authors show that out-region artifacting changes the flow of attention in the area outside the gap, but concentrated around the gap's neighbourhood. Hence, the saliency map values corresponding to a border region around the gap should be able to accurately capture the saliency change. This paper introduces the Border Saliency (BorSal) metric, which accounts for both

in- and out-region artifacting by considering a Border region that extends three pixels inside and outside the gap. The BorSal metric is computed as a normalized gaze density measure, similarly to the $\mathrm{GD_{in}}$ and $\mathrm{GD_{out}}$ metrics [9]:

$$BorSal = \frac{\sum_{p \in Border} S'(p)}{\sum_{p \in Border} S(p)}. \quad (5)$$

The second inpainting quality evaluation metric proposed, denoted by StructBorSal, combines the BorSal metric with the $\mathrm{SSIM_{IPT}}$ measure [13]:

$$StructBorSal = BorSal + SSIM_{IPT}. \quad (6)$$

D. Image Quality Metrics

In addition to the inpainting quality evaluation metrics discussed in the previous Section, eight other metrics from different categories (i.e. image difference, image fidelity, image quality) have been selected for evaluation in terms of correlation with the percept. A brief introduction of the selected metrics will be given here.

- MSE and PSNR: image difference metrics applied for grayscale images. Calculate the Mean Squared Difference and Peak Signal to Noise Ratio. Mathematically based, easily to implement and previously used for quantifying inpainting quality.
- S-CIELAB [14]: image difference metric applied for colour images. Frequently considered a reference when evaluating IQ metrics, having wide acceptance. Simulates the HVS.
- SSIM-IPT [13]: IQ metric applied for grayscale images. Colour version of SSIM, takes structural information in an image into account and works on local neighbourhoods.
- VSNR [15]: image fidelity metric applied for grayscale images. Based on contrast filtering and simulating the HVS.
- SHAME and SHAME-II [16]: IQ metrics applied for colour images. Based on the hue angle algorithm [17] and the $\mathrm{S\text{-}CIELAB_J}$ [18] metric. They latter differs from S-CIELAB only in terms of spatial filtering, which makes also the difference between SHAME and SHAME-II; Both metrics weight the output based on colour differences and region-of-interest. Simulate the HVS.
- ABF [19]: image difference metric applied for colour images. Implementation of bilateral filtering that preserves edges and simulates the HVS.

IV. Evaluation of Quality Metrics

The objective quality metrics presented in the previous section need to be evaluated against the results obtained from the psychophysical experiment in order to ensure the correspondence with perceived quality. The evaluation methodology refers to statistically analysing the ratings given by observers and corresponding to the 48 digitally inpainted images. Based on raw perceptual data, the Mean Opinion Score (MOS) is calculated for each image in the database and then converted to a corresponding z-score. The Pearson product-moment (PCC) [20] and Spearman's rank correlation coefficient (SCC) [20] between the z-scores and the objective scores (i.e. results from the IQ metrics) are calculated in order to evaluate the performance of the metrics considered.

A. Overall Performance

The overall evaluation of a metric is done by calculating the correlation between the observers z-scores and the metric raw scores over the entire image database. The obtained results, presented in Table I, indicate that all the considered metrics have a low correlation with the perceived overall quality. The two newly proposed metrics, BorSal and StructBorSal have a very low correlation. However, they achieve a better performance than the $\mathrm{GD_{in}}$ and $\mathrm{GD_{out}}$ [9] metrics. The DN metric [10], designed for inpainting quality evaluation, provides the highest correlation among all metrics, but with a value equal to -0.36 it still indicates a low performance over the entire image database.

A visual inspection of the relation between observer and metric z-scores depicted in Figure 5 shows very spread data points, due to scale differences between images, resulting thus in a low overall correlation. It can be concluded that, for the considered image database, the objective evaluation methods can not accurately predict perceived overall IQ. However, it is worth noticing that the DN metric provides a better fit for individual images, as shown in Figure 5.

B. Image-wise Evaluation

Hardeberg *et al.* [21] relate the performance of different metrics to certain characteristics of an image. This motivates the choice to evaluate the performance of metrics with respect to individual images. Table II gives the PCC and SCC for the 14 metrics considered in this study, applied to the six test images used.

Data in Table II shows a great variation between the scores obtained for the 14 metrics corresponding to a single image, but also between scores of the same metric for different images. However, for four out of six images (i.e. *boat*, *angel*, *lady*, *man*) results show that inpainting specific metrics perform better than IQ metrics. The DN [10], $\mathrm{GD_{in}}$ [9], BorSal and StructBorSal metrics provide good correlation with perceived quality, whereas only the S-CIELAB [14] and MSE among the IQ metrics indicate a better performance, in the specific cases of *cleopatra* and *detail* images. Considering the above, it can be concluded that both the inpainting-specific and IQ metrics are image-dependent and thus their performance will depend on the characteristics of the image for which they are applied. Several metrics that evaluate different aspects (i.e. structure reconstruction, colour, blur, *etc.*) could be used for the same image, for better results.

V. Conclusion

In this paper, eight different digital inpainting algorithms were qualitatively evaluated by conducting a psychophysical experiment. The obtained subjective data determined a ranking of the inpainting algorithms, showing that exemplar based

TABLE I: Performance of the metrics over the entire image database

	MSE	PSNR	DN	ASVS	GD_{in}	GD_{out}	S-CIELAB	SHAME	SHAME-II	$SSIM_{IPT}$	ABF	VSNR	BorSal	StructBorSal
PCC	-0.13	0.28	**-0.36**	-0.11	-0.01	-0.01	-0.26	-0.15	-0.25	0.22	-0.32	0.06	0.06	0.12
SCC	-0.28	0.28	**-0.39**	-0.11	0.07	-0.06	-0.19	-0.17	-0.19	0.31	-0.27	-0.02	0.08	0.13

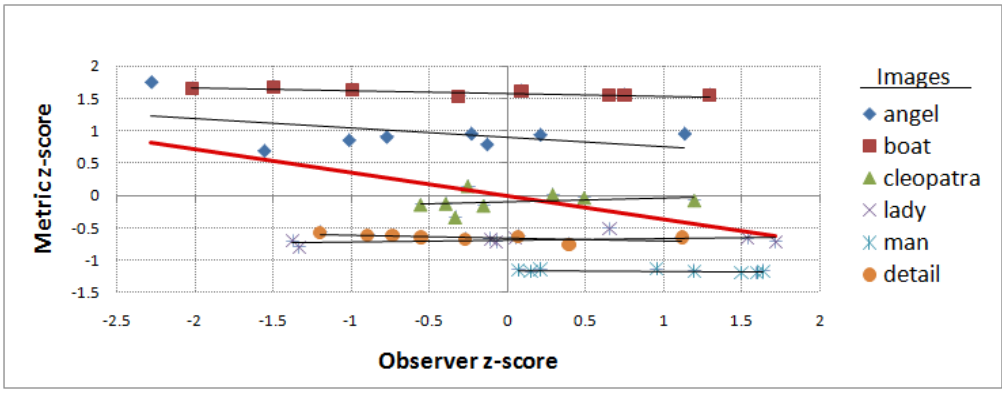


Fig. 5: Observer z-score plotted against DN [10] z-score for all images in the dataset. PCC = -0.36. The red linear regression line fits all data points; the black linear regression lines fit data points corresponding to individual images

TABLE II: Comparison of metrics performance image-wise. The score for best performing metric is highlighted in bold font for each image

Metric	boat		cleopatra		angel		lady		man		detail	
	PCC	SPP	PCC	SPP	PCC	SPP	PCC	SPP	PCC	SPP	PCC	SPP
MSE	-0.63	-0.31	-0.48	-0.67	-0.37	0.33	0.09	0.17	0.57	0.28	**-0.98**	-0.93
PSNR	0.62	0.31	0.53	-0.67	0.20	0.33	-0.13	0.17	-0.37	-0.57	0.90	0.98
DN	**-0.81**	**-0.66**	0.25	0.43	-0.46	0.11	0.39	0.29	-0.54	-0.46	-0.59	-0.77
ASVS	-0.77	-0.64	0.51	0.50	-0.48	0.07	0.39	0.29	0.88	0.92	0.27	-0.19
GD_{in}	-0.60	-0.64	0.49	0.43	-0.52	-0.02	**0.51**	**0.60**	**0.89**	0.93	0.27	-0.19
GD_{out}	-0.48	-0.36	0.53	0.43	0.34	0.55	-0.28	-0.29	0.76	0.71	0.26	-0.14
SCIELAB	-0.32	-0.21	**-0.62**	**-0.88**	-0.29	0.10	-0.16	-0.12	-0.06	-0.19	-0.90	**-1**
SHAME	-0.43	-0.29	-0.30	-0.40	-0.55	-0.14	-0.14	-0.24	-0.15	-0.31	-0.84	-0.99
SHAME-II	-0.35	-0.40	-0.42	-0.69	-0.55	-0.14	-0.15	0	0.09	-0.14	-0.89	-0.79
$SSIM_{IPT}$	-0.19	-0.26	0.58	0.76	-0.60	-0.45	0.03	-0.19	-0.16	-0.19	0.87	0.95
ABF	-0.39	-0.43	-0.58	0.76	-0.51	-0.14	-0.22	-0.55	-0.09	-0.19	-0.90	-0.71
VSNR	0.49	0.60	0.20	0.14	0.70	-0.45	0	0.12	-0.66	-0.81	-0.82	-0.93
BorSal	0.09	0.07	0.48	0.64	-0.57	-0.50	0.37	0.29	0.87	**0.98**	0.27	-0.19
StructBorSal	0.05	0.07	0.54	0.64	**-0.77**	**-0.67**	0.40	0.19	0.82	0.95	0.31	-0.19

methods generally outperform PDE based methods. The latter verifies the theoretical analysis of digital inpainting methods, as exemplar based inpainting algorithms can reconstruct both texture and structure in an image, as opposed to PDE based algorithms.

Furthermore, extending the work of Ardis *et al.* [10] and Mahalingam [9], two inpainting-specific quality metrics have been proposed. Along with four other inpainting-specific and eight image quality metrics, they were considered for performance assessment, measured as the degree of correlation with the percept, against a database of 48 images. To our knowledge, this evaluation is one of the most extensive carried out in the literature, with respect to inpainting quality. The obtained results show that none of the metrics have a high performance over the whole image database, but certain metrics perform well for specific images. Results indicate that inpainting specific metrics outperform image quality metrics when applied for images with small-sized gaps (i.e. *man*, *lady*), or for images that don't require complex structure reconstruction (i.e. *boat*).

Future work will include an expanded psychophysical experiment that will consider more recently proposed inpainting algorithms and a larger number of inpainted images, in order to reconfirm initial findings. Extensive research should be conducted with the aim of developing a no-reference inpainting quality metric.

REFERENCES

[1] M. Bertalmio, "Processing of flat and non-flat image information on arbitrary manifolds using partial differential equations," 2001.
[2] A. Telea, "An image inpainting technique based on the fast marching method," *J. Graphics Tools*, vol. 9, no. 1, pp. 23–24, 2004.

[3] D. Tschumperle and R. Deriche, "Vector-valued image regularization with pde's : A common framework for different applications." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 27(4), pp. 506–517, 2005.

[4] F. Bornemann and T. März, "Fast image inpainting based on coherence transport," *J. Math. Imaging Vision*, vol. 28, no. 3, pp. 259–278, 2007. [Online]. Available: http://dx.doi.org/10.1007/s10851-007-0017-6

[5] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting." *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 13, no. 9, pp. 1200–12, Sep. 2004.

[6] J. Zhou and A. R. Kelly, "Image inpainting based on local optimization," *International Conference on Patteren Recognition (ICPR)*, 2010.

[7] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "Patchmatch: a randomized correspondence algorithm for structural image editing," *ACM Transactions on Graphics (TOG)*, vol. 28, no. 3, 2009.

[8] M. M. Oliveira, B. Bowen, R. Mckenna, and Y. sung Chang, "Fast digital image inpainting," in *Proceedings of the International Conference on Visualization, Imaging and Image Processing (VIIP 2001*. ACTA Press, 2001, pp. 261–266.

[9] V. V. Mahalingam, "Digital inpainting algorithms and evaluation," Ph.D. dissertation, University of Kentucky, 2010.

[10] P. Ardis and A. Singhal, "Visual salience metrics for image inpainting," in *SPIE*, no. 7257, 2009.

[11] ITU-R, "Rec. ITU-R BT.500-13. Methodology for the subjective assessment of the quality of television pictures," 2012.

[12] D. Walther, "Interactions of visual attention and object recognition: computational modeling, algorithms, and psychophysics," Ph.D. dissertation, California Institute of Technology, Pasadena, CA,, 2006.

[13] N. Bonnier, F. Schmitt, H. Brettel, and S. Berche, "Evaluation of spatial gamut mapping algorithms," *Color Imaging Conference*, pp. 56–61, Nov. 2006.

[14] X. Zhang and B.A.Wandell, "A spatial extension of CIELAB for digital color image reproduction," *Soc. Inform. Display 96 Digest*, pp. 731–734, 1996.

[15] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal to noise ratio for natural imges," *IEEE Transactions on Image Processing*, no. 16(9):2284-2298, Sep 2007.

[16] M. Pedersen and J. Y. Hardeberg, "A new spatial hue angle metric for perceptual image difference," *Computational Color Imaging, volume 5646 of Lecture Notes in Computer Science*, pp. 81–90, Mar. 2009.

[17] G. Hong and M. R. Luo, "New algorithm for calculating perceived colour difference of images," *Imaging Science Journal, 54(2):86-91*, no. 54(2):86-91, 2006.

[18] G. M. Johnson and M. D. Fairchild, "Darwinism of color image difference models," *Color Imaging Conference*, pp. 108–112, Nov. 2001.

[19] Z. Wang and J. Y. Hardeberg, "An adaptive bilateral filter for predicting color image difference," *Color Imaging Conference*, pp. 27–31, Nov. 2009.

[20] M. G. Kendall, A. Stuart, and J. K. Ord, "Kendall's Advanced Theory of Statistics: Classical inference and relationship, volume 2. A Hodder Arnold Publication, 5 edition," 1991.

[21] J. Y. Hardeberg, E. Bando, and M. Pedersen, "Evaluating colour image difference metrics for gamut-mapped images," *Coloration Technology, 124(4):243253*, no. 124(4):243-253, Aug 2008.