Characterization and Quality Evaluation of Geometric Distortions in Images with Application to Digital Watermarking

Angela D'Angelo

Ph.D Thesis in Information Engineering
University of Siena

UNIVERSITÀ DEGLI STUDI DI SIENA

FACOLTÀ DI INGEGNERIA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

Characterization and quality evaluation of geometric distortions in images with application to digital watermarking

Angela D'Angelo

Ph.D Thesis in Information Engineering

Supervisor

Prof. Mauro Barni

Examination Commitee

Prof. Jean Luc Dugelay

Prof. Alessandro Piva

Prof. Andrea Abrardo

Thesis reviewers

Prof. Patrick Le Callet

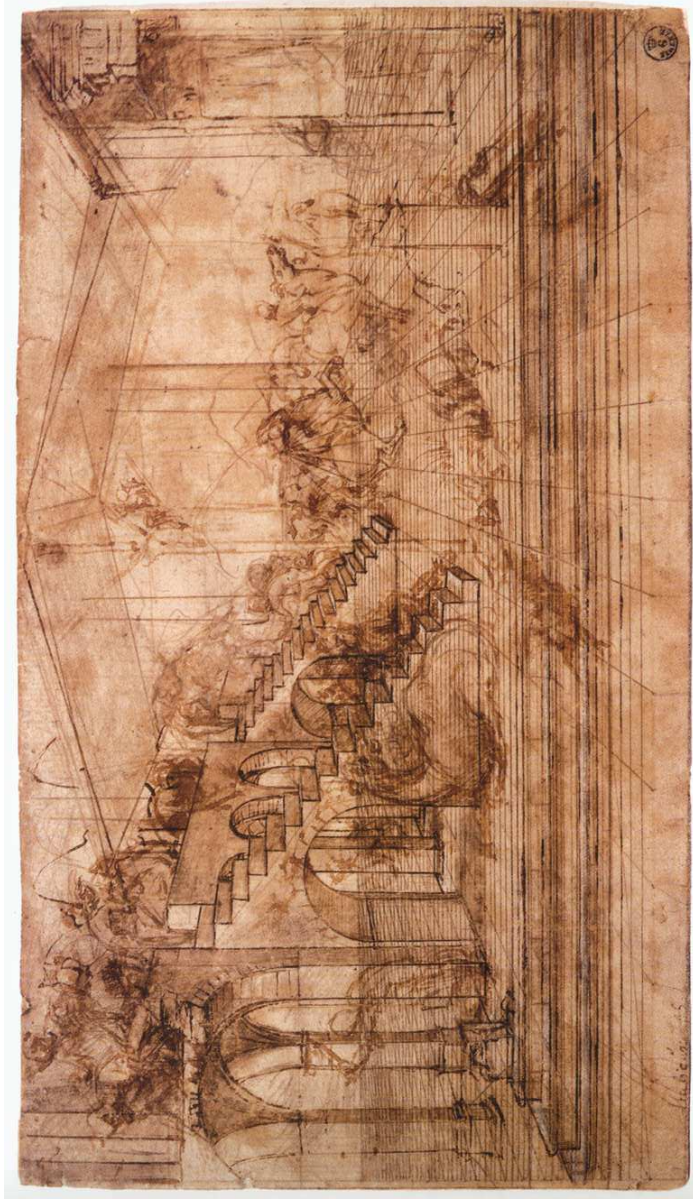Prof. Pedro Comesaña Alfaro

SIENA

MAY 13, 2009

Figure 1: **Leonardo da Vinci**, Perspectival study for *The Adoration of the Magi* (circa 1481)

Contents

i

List of Figures

List of Tables

ix

Acknowledgments

Chapter 1

Introduction

1.1 Motivations and approaches

Digital images are subject to various kinds of distortions that may result in a degradation of visual quality during acquisition, processing, compression, storage, transmission and reproduction. It is therefore necessary for many applications to be able to quantify the image quality degradation that occurs in a system, so that it is possible to control and enhance the quality of the produced images. For this reason, optimizing the performance of digital imaging systems with respect to the capture, display and transmission of visual information represents one of the biggest challenges in the field of image and video processing. For instance, quality assessment tools predict subjective ratings, image compression schemes reduce the visibility of introduced artifacts, watermarking schemes hide invisible information in images, printer half-toning patterns are perceptually optimized and colors rendered as accurately as possible. In all these applications, the characteristics of the Human Visual System (HVS) can be exploited to improve the performance from a visual quality point of view. To achieve this, it is necessary to understand and to build computational models of the HVS model.

The need for accurate vision models has been increasing with the introduction of digital processing of visual information. Let us think for example to television technologies. While traditional analog systems still form the vast majority of tv sets today, production studios, broadcasters and network providers have been installing digital video equipment at an ever-increasing rate. Digital satellite and cable services have been available for quite some time now, and terrestrial digital TV broadcast has been introduced in a number of locations around the world. A similar development can be observed in

photography with the growing popularity of digital cameras.

For conventional analog systems there are well-established performance standards. They rely on particular test signals and measurement procedures to determine signal parameters that can be related to perceived quality with high accuracy. While these parameters are still useful today, their connection with perceived quality has become much more tenuous. Because of compression, digital imaging systems exhibit artifacts that are fundamentally different from those encountered in analog systems. The amount and visibility of these distortions strongly depend on the actual image content. Therefore, traditional measurements are inadequate for the evaluation of these artifacts.

Some examples of very popular distortions that images and videos can undergone during acquisition, processing, compression, storage, transmission and reproduction, and that can affect the quality of the visual information, are noise addition, blurring and JPEG compression. These distortions are very common in image processing applications, for this reason several denoising and deblurring techniques and methods to avoid blocking artifacts due to JPEG compression have been developed.

A different kind of distortions that images can undergone are geometric distortions. The problem of geometric distortions is a well known problem in photographic images as a form of Seidel aberration[1]. Specifically, a geometric distortion in geometric optics is a deviation from rectilinear projection, a projection in which straight lines in a scene remain straight. It is a form of optical aberration and it represents a well known problem, that is the departure of the performance of an optical system from the predictions of paraxial optics[2], and instrument-makers need to correct optical systems to compensate for aberration and to prevent image quality degradation.

Some examples of optical aberrations are shown in Fig. 1.1. In the image on the left, the second and third layers show *barrel distortion*, in which image magnification decreases with distance from the optical axis, while the top layer reveals a more pronounced *pincushion distortion*, in which image

---

[1]http://toothwalker.org/about.html

[2]Paraxial optics is a method of determining the first-order properties of an optical system that assumes all ray angles are small. A paraxial raytrace is linear with respect to ray angles and heights since all paraxial angles are defined to be the tangent of the actual angle (http://toothwalker.org/about.html).

Figure 1.1: *Examples of geometric distortions in photografic images. Credits: Paul van Walree*[4].

magnification increases with the distance from the optical axis. An example of pincushion distortion is also presented in the image on the right.

Apart from the world of digital photography, the problem of the characterization and quality evaluation of geometrically distorted images is important in many applications in different research fields. In the context of medical imaging, for example, a great attention has been paid to study and correct the distortions that affect magnetic resonance imaging in radiotherapy (Petersch et al. 2004). In the same way, geometric distortions introduced by the Landsat imaging system are well-known (Shlien 1979). Remote Sensing data includes two types of geometric errors or distortions, that is, internal external errors. An internal error mainly results from the geometric characteristics or performance of sensors, therefore, it can be corrected systematically if the calibrations data or parameter for correction are given. On the other hand, the external error results from the altitude of the platform and the geometric configuration of the objects. The distortions resulting from the altitude can be corrected systematically only if the variations with respect to the parallel

to the ground can be precisely measured on board. However, measurements on board is usually not enough to obtain a satisfied accuracy, thus a sufficient number of ground control points with known coordinates should be defined to establish a transformation between the image coordinate system and the geographic coordinate system, resulting in a local non linear distortion.

Other popular applications that require the study of geometric distortions and geometric correction are registration of biomedical images that usually requires the application of local and nonlinear transformations (Brown 1992), collusion-secure fingerprinting techniques by random pre-warping (Celik et al. 2004), the problem of recovering 3D models from uncalibrated images of architectural scenes (Cipolla et al. 1999) or digital watermarking in presence of desynchronization attacks (Barni 2005). In this last application the characterization and quality evaluation of geometric distortions is a very important step to develop a class of watermarking systems robust against geometric attacks. The importance of geometric distortions in image watermarking will be further detailed in the next section.

1.2   Application to digital watermarking

During the last decade, the exponential growth of digital technology has caused the transition from the analog to the digital world including multimedia production and distribution, from the authoring tools of content providers to the receivers. Some of the advantages of digital processing and distribution are noise-free transmission, software instead of hardware processing leading to improved reconfigurability of systems (i.e. Software Defined Radio), digital support (i.e. CD, DVD), and so on. Together with the expansion of Internet, multimedia communication is nowadays a fact.

However film and music content owners are still reluctant to release digital contents. A major concern raised by the new digital technologies, in fact, is the ease to reproduce and distribute media contents in an unauthorized way. In analogue form, reproduction and distribution was prevented by unavoidable degradations and material constraints. With digital technology, perfect copies of the digital content can be made at very low cost, rapidly, at large scale, without any limitation on the number of copies and can be distributed easily

e.g. via Internet. Moreover perfect reproduction processes make more difficult the determination of the paternity of a work causing considerable financial loss for media producers and content providers.

Protection of digital contents has relied for a long time on encryption though it is evident that encryption alone is not sufficient to protect digital data. In fact, when a digital content is decrypted to be eventually presented to the consumer, the protection offered by encryption no longer exists and a user may duplicate or manipulate the content. Digital watermarking has consequently been introduced as a complementary mean to enforce intellectual property rights. The watermark is a digital code unremovably, robustly, and imperceptibly embedded in the host data and typically contains information about origin, status, or destination of the data. The role of the embedded information can be manifold. For copyright protection, the watermark information can refer to the rights holder. If this information is detected, it can prevent illicit usage of the content or can be used as a proof of ownership. Another option is to use the watermark as a fingerprint. Using a watermarking scheme, a fingerprint identifying the buyer is embedded in every sold copy. If an illegal is found copy appears, the watermark permits to trace back to the guilty buyer.

While copyright protection is the most prominent application of watermarking, others exist, including data authentication by means of fragile watermarks which are impaired or destroyed by manipulations (Fridrich 2002), embedded transmission of value added services in multimedia data (Swanson et al. 1998), data monitoring, tracking, and indexing obtained by embedding digital labels (Cox et al. 2002).

1.2.1 Requirements of a watermarking scheme

The design of a watermarking system involves several tradeoffs depending on the particular application the system is intended for. Generally, the watermark should be imperceptible (imperceptibility), it should convey as much information as possible (payload), and it should be robust (robustness), especially in copyright protection. Finally, in most watermarking applications, an important requirement is security (security).

Imperceptibility requires that the embedded watermark does not alter

the perceived quality of the host image. For this purpose the characteristics of the Human Visual System (HVS) are usually exploited in the embedding phase to keep the embedded signal invisible.

The **payload** of a watermarking system is defined as the amount of information that the watermark is able to convey and it is strictly depending on the application of the system. Possible requirements range from few hundreds of bits in security-oriented applications, where robustness is a major concern, through several thousands of bits in application like captioning or labeling.

For most watermarking applications, an important requirement is **security**. Watermarking security is defined as "the inability by unauthorized users to access (i.e., to remove, to read, or to write the hidden message) the communication channel" established by a robust watermarking (Cayre et al. 2005). Security deals with intentional attacks whose aims are not only the removal of the watermark, excluding those already encompassed in the robustness category since the watermarking technique is assumed to be robust.

Watermark **robustness** accounts for the capability of the hidden data to survive host signal manipulations, including both non-malicious manipulations, which do not explicitly aim at removing the watermark or at making it unreadable, including digital to analog conversion and digital format conversion, and malicious manipulations, which precisely aim at damaging the hidden information.

In the last years, many watermarking systems have been proposed that are robust against common signal processing manipulations, such as addition of gaussian noise, blur, jpeg compression, histogram equalization, contrast enhancement, etc (Cox et al. 1997b). One of the main challenges in the field of digital watermarking, however, is to design systems robust against geometric attacks, as explained in the next section.

1.2.2   The synchronization issue in watermarking

Synchronization is a very important issue in a communication system. Indeed, a signal bears information only relatively to its original references and modifying the references results in a damage of the conveyed information. Robust communication systems require means to recover from a possible loss of the original signal references.

Many services running on modern digital telecommunication networks require accurate synchronization for correct operation. For example, if switches do not operate with the same rate clocks then slips will occur degrading performance. Telecommunication networks rely on the use of highly accurate Primary Reference Clocks which are distributed network wide using Synchronization Links and Synchronization Supply Units (Bregni et al. 2002).

The synchronization issue is a key point also in digital watermarking. The use of digital watermarking in real applications today, in fact, is impeded by the weakness of current available algorithms against desynchronization attacks. Desynchronization attacks, or geometric attacks in the case of still images, in a watermarking system occur when the watermarked data undergoes a geometric transformation. Achieving robustness of watermarking schemes against geometric transformations, or geometric distortions, is a very difficult task and it has been the subject of quite much research in the last ten years in the watermarking community. In fact, the tolerance of the human visual system to such distortions is surprisingly high and situations where geometric transformations take place are numerous and they can be consequent to malevolent manipulations, aimed at removing the watermark, as well as to usual image-processing manipulations such as scaling images for a Web site, printing and scanning marked documents, changing a digital video's aspect ratio, and cropping an image to extract a region of interest.

The study of geometric distortions in watermarking involves two aspects that are of interest for the watermarking community, namely the watermark desynchronization aspect and the perceptual quality aspect. Regarding the first aspect, global geometric transformations, especially rotation, scaling and translation, have been extensively studied in the watermarking literature given their simple mathematical formulation and diffusion. Though no perfect solution exists to cope with geometric attacks, desynchronzation attacks based on global transformations can be handled in a variety of ways, including exhaustive search (Barni 2005), template-based re-synchronization (Voloshynovskiy et al. 2001), selfsynchronizing watermarks (Delannay and Macq 2000, Kutter 1999) and watermarking in invariant domains (Lin et al. 2001). In all the cases, the proposed solutions rely on the restricted number of parameters specifying the attack. For instance, it is the relatively low cardinality of

the set of possible global geometric attacks that makes the estimation of the geometric transformation applied by the attacker via exhaustive search or template matching possible (computationally feasible).

On the other hand, local geometric attacks are transformations that affect subsets of the image, allowing an attacker to apply distinct operations under different parameters over each image subset. Thus, the space of all possible localized attacks applied to an image (that is, all possible combinations of the basic transformations applied individually to each subset) has a much higher dimension than the space of all possible global attacks applied to the same image. Thus, recovering from localized attacks is much harder than recovering from a global attack.

A possibility to overcome this problem in case of local attacks could be to split the search into a number of local searches. However, in this way, it is likely that the accuracy of the estimation is reduced, given that the estimation would have to rely on a reduced number of samples. In despite the threats they pose, local geometric transformations have received little attention by the watermarking community. In practice, only the Random Bending Attack (RBA) (Petitcolas and Anderson 1999) contained in the Stirmark software [5] has been studied to some extent. The underlying idea is that each location of the sampling grid is displaced by a random amount that is possibly a function of the point coordinates.

The second aspect of interest for the study of geometric distortions in watermarking systems, is related to the understanding of the perceptual quality impact of geometric attacks on the watermarked images since the human observer is the end user in many multimedia applications. Geometric distortions degrade the visual quality of the watermarked data but, like all the other distortions that affect watermarking systems, distortions due to geometric transformations are also bounded by the maximum visual quality degradation the watermarked data can incur before the distorted image looses its commercial value. It is therefore important to be able to measure the annoyance of such distortions and to find the visibility threshold for the HVS for this class of distortions. It is clear that the availability of an objective quality metric capable of dealing with geometric distortions would be of invaluable help in this

---

[5]http://www.petitcolas.net/fabien/watermarking/stirmark/

sense.

Although in multimedia applications there has been an increasing interest in the use of quality measures based on human perception, only few works can be found in the scientific literature dealing with the quality assessment of geometrically distorted images.

Watermarking is not the only field where an image quality metric for geometrically distorted images would be needed. Interested applications include registration of biomedical images that usually requires the application of local and nonlinear transformations, whose strength should be constrained to satisfy certain quality constraints, collusion-secure fingerprinting techniques by random pre-warping, the problem of recovering 3D models from uncalibrated images of architectural scenes. In all the above applications, an objective metric capable of assessing the quality of geometrically distorted images is missing.

1.3 Contributions and outline

The work of this thesis can be seen as a first step towards the characterization and quality evaluation of the class of geometric distortions.

Specifically, we focus our attention on the problem of local geometric transformations. Global transformations, in fact, usually do not affect image quality at all or introduce a degradation proportional to the parameters defining the distortion, for instance the zooming factor could be an indication of the quality degradation of a zoomed image. Moreover in some applications like digital watermarking it is more difficult to deal with local transformations with respect to global distortions for which many watermarking schemes have been proposed during the last years, as explained in the previus section. The issue of the development of watermarking algorithms robust against local geometric transformations, in fact, is still an open problem and it could be useful, for this pourpose, to characterize and to evaluate the quality of a geometrically distorted image.

The first part of the this work focuses on modeling local geometric transformations for still images, while, in the second part, two objective quality metrics for the perceptual evaluation of geometrically distorted images are

introduced.

The contributions of this thesis can be summarized as follows:

- Two new models to describe local geometric distortions from a mathematical point of view are introduced. The C-LPCD model (chapter 3), based on local permutations of the position of the pixels in the image, is a model for which a simple interpretation in terms of Markov chain is possible. The MF model (chapter 4) is based on the theory of Markov Random Fields. This last model has a more complex formulation but is able to generate larger displacement vectors while keeping the distortions invisible. These models represent the first attempt (except for the RBA) to provide a mathematical formulation of local geometric distortions in images.

- The perceptual quality impact of the defined transformations on still images is evaluated through extensive tests defining the class of perceptually admissible distortions. This analysis will be a valuable tool in the field of digital watermarking for the development of watermarking systems robust against geometric attacks.

- The effectiveness of the new models as geometric attacks in watermarking systems is evaluated from different perspectives including perceptual intrusiveness and de-synchronization capabilities, and compared with the classical RBA (chapter 5). The experimental results show that the two new classes of attacks are more powerful than the local geometric attacks proposed so far, and represent a new benchmark for the evaluation of watermarking systems robust against geometric attacks.

- Two objective quality metrics for the perceptual evaluation of geometrically distorted images have been developed. The first objective metric (chapter 7) is based on the theory of Markov Random Fields and on the definition of the potential function of the configuration defining the geometric distortion. It relies on the assumption that the potential function of the configuration defining the distortion is an indication of the perceived image quality.

- The second method is based on image features processed by human vision through the use of Gabor filters (chapter 8). Extended psychovisual experiments have been performed to tune the objective metrics with psychovisual data in order to obtain perceptual metrics and to validate the proposed approaches.

- A multiresolution extension of both the metrics has been introduced to incorporate image details at different resolutions and to obtain a multiresolution extension of the described metrics (chapter 9). The experimental results show good performances of the metrics, that outperform the quality metrics proposed so far for the evaluation of geometric distortions in images.

In order to ensure the reproducibility of the experimental results, the software we used for the experiments, as well as all the image databases, is available on the web site *http://www.dii.unisi.it/∼vipp/*, furthermore a pseudo-code description of the algorithms is provided in the appendix of the single chapters in order to link the software to the global description of the algorithms.

Part I

Characterization of geometric distortions in images

Chapter 2
Geometric distortions in images

2.1 Introduction

The field of image and video processing generally deals with signals that are meant for human consumption, such as images or videos. Digital images or videos may go through many stages of processing before being presented to a human observer, and each stage of processing may introduce distortions that could reduce the quality of the final displayed signal. For example, images and videos are acquired by camera devices that may introduce distortions due to optics, sensor noise, color calibration, exposure control, camera motion etc. After acquisition, the image or video may further be processed by a compression algorithm that reduces the bandwidth requirements for storage or transmission. Compression algorithms are generally designed to achieve greater savings in bandwidth by letting certain distortions happen to the signal without a great loss of details in the media. Similarly, bit errors, which occur while an image is transmitted over a channel, can also introduce distortions. Finally, the display device may introduce some of its own distortions, such as low reproduction resolution, bad calibration etc. The amount of distortion that each of these stages adds depends mostly on economics and/or physical limitations of the devices.

Some examples of popular distortions that images (we focus now on the problem of digital images, but these concepts can be easily extended to videos or audio) may undergone during acquisition, processing, compression, storage, transmission and reproduction, are noise addition, blurring and JPEG compression whose artifacts are shown in Fig. 2.1.

These distortions are very common in image processing applications, for this reason several denoising techniques, aiming at removing noise from the image while preserving useful information as much as possible, and deblurring

(a)

(b)

(c)
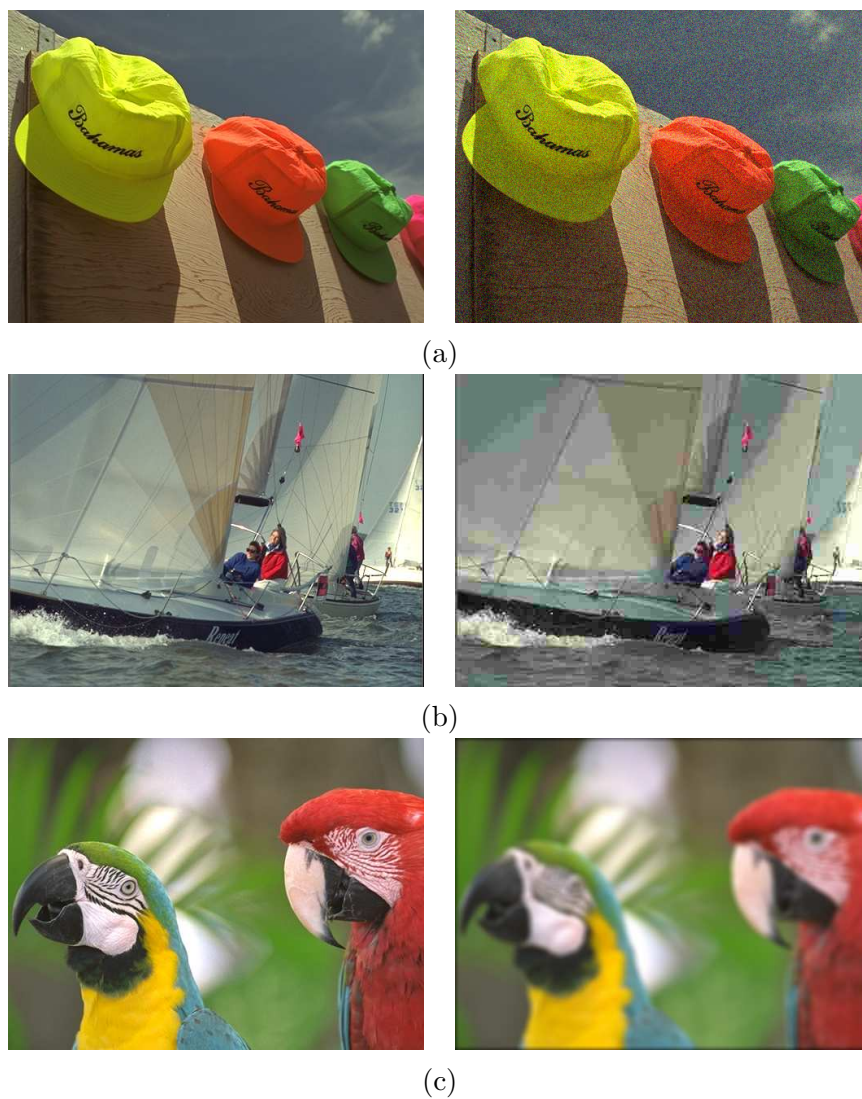
Figure 2.1: *Examples of distortions (the original image is on the left): (a) additive Gaussian noise; (b) JPEG compression; (c) gaussian blur. Credits:Tamper Image Database 2008 (Ponomarenko et al. 2008)*

methods have been developed. In the same way, techniques to avoid annoying blocking artifacts due to JPEG compression have been investigated.

A different kind of distortions that images can undergone is the class of geometric distortions.

In general, a geometric distortion (sometimes referred to as image warping)[1], is the process of digitally manipulating an image in such a way that the shape in the image are distorted. In this process there is a transformation of the position of the pixels in the image.

Even if a strict definition does not exist, it is possible to define a geometric distortion in the following way: let I be the original image, the distorted image Z is obtained (neglecting border effects) by assigning to each pixel $I(x, y)$ a displacement vector $D(x, y) = (D_h(x, y), D_v(x, y))$, where $D_h(x, y)$ and $D_v(x, y)$ are respectively the horizontal and vertical displacements. In the following D is called displacement field.

A geometric distortion can also be defined through the function generating the displacement field: a geometric transformation T is a map from the plane to itself that associates each point (x, y) with a different point in the same coordinate system, as denoted by $(x, y) \rightarrow T(x, y)$.

Depending on how D is produced and how T is defined, it is possible to distinguish between global and local geometric distortions, as explained in the following sections.

2.2 Global geometric distortions

A geometric transformation can be defined by a mapping function that relates the points in the input image to the corresponding points in the output image. The mapping function of a global geometric transformation is defined by a set of operational parameters and applied to all the image pixels; that is, the same operation, under the same parameters, affects all the image pixels. In this section a brief description of some commons geometric distortions is provided, using concepts from the analytical geometry.

An *isometry* is a transformation of the plane that preserves distances. Translations, rotations and reflections are the three basic isometries. The

---

[1]Image warping is the process of digitally manipulating an image such that any shapes portrayed in the image are significantly distorted. Warping is usually associated to morphing, that is a special effect in motion pictures and animations that changes one image into another through a seamless transition (Wolberg 1998).

simplest isometry is *translation*, where all pixel positions are modified by a fixed offset in the horizontal direction D_h and a fixed offset in the vertical direction D_v.

A clockwise *rotation* by an angle of θ radiants can be expressed as:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{2.1}$$

where (x', y') are the coordinates of the new sampling points, and (x, y) the original ones.

Reflection refers to those transformations for which each point of the space is reflected in a plane, the reflection plane or mirror plane, such that all points of this plane, and only these points, are fixed points. In addition, each line and each plane perpendicular to the mirror plane are left invariant as a whole. While reflection is a common term in analytical geometry, in the field of image processing is more popular to deal with *flipping*, that is the reflection of the image across its central line (horizontal or vertical). That line is called axis of reflection.

All other isometries can be defined as a sequence of translations, rotations and reflections.

A *proportional scaling*, usually known as *zooming*, is basically a re-sampling process defined by the function $T : (x, y) \rightarrow (ax, ay)$. In typical applications it is common to deal with zooming factors ranging from -50% to 150%, thanks to effective interpolation techniques which allow to avoid deteriorating image quality in a dramatic way. When the zooming factor in the x direction is different from the zooming factor in the y direction, i.e. when we have $T : (x, y) \rightarrow (ax, by)$ with $a \neq b$, the corresponding transformation is referred to as *affine transformation*. Affine transformations preserve lines and parallelism.

Projective transformations map lines to lines while not necessarily preserving parallelism. It is possible to express these transformations in homogeneous coordinates by an invertible 3×3 matrix. Projective transformations do not

preserve areas or angles but they preserve incidence[2] and cross-ratios[3]. A *perspective transformation* gives a transformation from one plane to another and is a common type of projective transformation.

A *shear* is a transformation preserving horizontal lines and it has the form $(x, y) \rightarrow (x + ry, y)$ where r is the shearing factor.

In typical applications, it is easy to find more complex transformations given by combinations of simple distortions. Rotations, for example, are usually associated to cropping. In *cropping*, that is not exactly a geometric distortion since there is no displacement associated to the pixels, a certain region of interest of the image is extracted while discarding the other portions. Obviously, it is not possible to invert cropping because the cropped image incurs a permanent loss of information about the discarded pixels. Let us think to digital photography and the problem, for example, with taking photos with a baseline exactly on the horizon without a tripod. Often people need to rotate the photo and, after rotation, to crop it. Fig. 2.2 shows an example of an image that has been rotated by 10 degrees counterclockwise, then cropped and zoomed in order to let the distorted image to be of the same size of the original image. The resulting image in Fig. 2.2(d) is a high quality image and it is not possible to notice the distortion without the reference image.

Global distortions are quite popular because they are caused by most common physical or digital handling. Circumstances where such distortions can take place are content editing using image manipulation software, format conversion, print and scan operations, handy-cam video acquisition, etc.

---

[2]In geometry, incidence relations are expressed by statements such as "lies on" between points and lines (as in "point P lies on line L"), and "intersects" (as in "line L1 intersects line L2", in three-dimensional space). That is, they are binary relations describing how subsets meet.

[3]The cross-ratio of a set of four distinct points on the complex plane is given by:

$$CR\left(z_1, z_2, z_3, z_4\right) = \frac{\|z_1 - z_3\| \|z_2 - z_4\|}{\|z_1 - z_4\| \|z_2 - z_3\|}$$
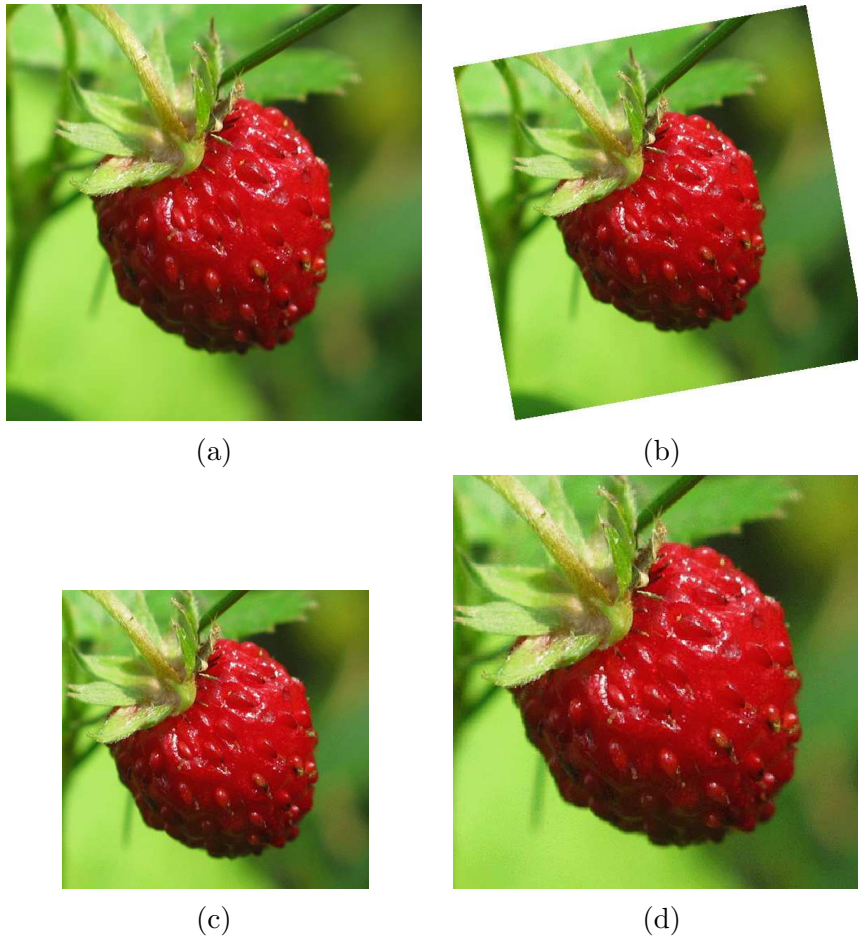
Figure 2.2: *Examples of global geometric distortions: (a) original image, (b) rotated image by 10 degrees, (c) cropped image, (d) zoomed image.*

2.3   Local geometric distortions

Local distortions refer to transformations affecting in different ways the position of the pixels of the same image or affecting only a subset of the image.

The space of all possible local geometric distortions that is possible to apply to an image (that is, all possible combinations of the basic transformations applied individually to each subset) has a much higher dimension

than the space of all possible global distortions. Stated in another way, the number of parameters necessary to describe the modifications resulting from a set of localized distortions in an image is by far larger than the number of parameters necessary to characterize a global geometric distortion.

Fig. 2.3 shows two examples of local geometric distortions applied to the Barbara image (the image on the top) with a different perceived image quality: image Fig. 2.3(b) has a good quality thanks to the smoothness of the displacement field, while image Fig. 2.3(c) presents an annoying distortion affecting heavily image quality.

(a)

(b)   (c)

Figure 2.3: *Local geometric distortions applied to Barbara image (a).*

A possible way to model local geometric transformations is to define a series of global distortions applied to non-overlapping image subsets under distinct parameters (Voloshynovskiy et al. 2001).

Another way to think about local geometric distortions is by letting a random set of parameters distortions to be applied at each individual sampling grid location. This class of distortions is called Random Bending distortions and it dramatically increases the distortions parameter space's dimensionality. Random Bending is very common in digital watermarking. An example of this kind of distortion is provided in the next section.

2.3.1   Stirmark Random Bending distortion(RBA)

In digital watermarking, the most popular example of local geometric distortion is the random warping firstly implemented by the Stirmark software (Petitcolas and Anderson 1999), that is very popular in the digital watermarking community as a benchmark tool to test the performance of watermarking schemes against de-synchronization attacks (as we will see in chapter 5).

In most of the scientific literature, by Random Bending Attack (RBA) the geometric distortion implemented in the Stirmark software is meant, however such a distortion is not a truly local attack since it couples three different geometric transformations applied sequentially, only the last of which corresponds to a local geometric distortion.

The first transformation applied by Stirmark is defined by:

$$\begin{aligned} x' &= t_{10} + t_{11}x + t_{12}y + t_{13}xy \\ y' &= t_{20} + t_{21}x + t_{22}y + t_{23}xy \end{aligned} \qquad (2.2)$$

where x', y' are the new coordinates and x, y the old ones. In practice, this transformation corresponds to moving the four corners of the image into four new positions, and modifying coherently all the other sampling positions. The second step is given by:

$$\begin{aligned} x'' &= x' + d_{max}\sin(y'\frac{\pi}{M}) \\ y'' &= y' + d_{max}\sin(x'\frac{\pi}{N}) \end{aligned} \qquad (2.3)$$

where M and N are the vertical and horizontal dimensions of the image. This transformation applies a displacement which is zero at the border of the image

and maximum (d_{max}) in the center. The third step of the Stirmark geometric attack is expressed as:

$$x''' = x'' + \delta_{max} \sin\left(2\pi f_x x''\right) \sin\left(2\pi f_y y''\right) \text{rand}_x(x'', y'')$$
$$y''' = y'' + \delta_{max} \sin\left(2\pi f_x x''\right) \sin\left(2\pi f_y y''\right) \text{rand}_y(x'', y'') \tag{2.4}$$

where f_x and f_y are two frequencies (usually smaller than $1/20$) that depend on the image size, and $\text{rand}_x(x'', y'')$ and $\text{rand}_y(x'', y'')$ are random numbers in the interval $[1, 2)$. Eq. (2.4) is the only local component of the Stirmark distortion since it introduces a random displacement at every pixel position. The Stirmark software is usually used as a benchmark tool in digital watermarking schemes and the parameters of the model are set in order to obtain unperceptible distortions. In Fig. 2.4 an example of the distortions generated by using this software is shown. For the purpose of visibility a strong and annoying distortion is generated.

Figure 2.4: *Example of distortions generated using the Stirmark software (the original image is on the left).*

Chapter 3

The LPCD model

3.1 Introduction

In the previous chapter, a definition of the class of geometric distortions has been given together with a classification of local and global geometric distortions. As already explained, in some applications like digital watermarking, it is more difficult to deal with local transformations than with global distortions against which many watermarking schemes have been proposed during the last years. The development of watermarking algorithms robust against local geometric transformations, instead, is still an open problem and for this reason it could be interesting to characterize and analyze this class of distortions. Furthermore no general mathematical model to describe local geometric distortions can be found in the literature, except for the Stirmark Random Bendind transformation (see 2.3.1).

The goal of the next two chapters is twofold: first, to introduce two new classes of local geometric distortions based on precise and rather simple mathematical models; second, to measure the impact of the parameters of the models from a perceptual point of view. In this chapter we introduced the so-called LPCD model (Local Permutation with Cancelation and Duplication). Specifically different mathematical formulations of the model are provided following perceptual constraints, in order to let the model of the distortion be useful in different applications. mIn most of the applications, in fact, we are interested only in distortions that do not affect too much the image quality. It should be interesting to have a mathematical formulation of geometric distortions that allow to control the impact of image quality by means of a reduced set of parameters.

For this reason we carried out both objective and subjective tests to characterize the perceptual impact of the artifacts introduced by the distortion.

<center>(a)   (b)</center>

Figure 3.1: *Sample images of size SxS (S = 512): (a) Lena; (b) Duomo.*

The maximum distortion that can be applied before the distortion becomes visible is also measured leading to the definition of the perceptually admissible subset of the possible distortions that can be applied to an image.

3.2   Theoretical background

In the following subsections some classes of geometric transformations are introduced. Each class is analyzed by means of visual inspection using the images in Fig.3.1. The two images have different characteristics: one is the Lena standard image that is mostly smooth, the other one, the Duomo image, is much more structured.

3.2.1   Block-based Local Permutation

A generic, very simple, local distortion can be described by a permutation of the position of the pixels in the image. Of course this kind of distortion introduces an annoying degradation. A way to overcome this problem could be to fix a maximum displacement of the position of the pixels, i.e. to perform block-based local permutations.

(a)   (b)

Figure 3.2: *Block-based Local Permutation ($Dim = 5$): (a) Lena; (b) Duomo.*

This model consists in partitioning the $S \times S$ original image in $\frac{S}{Dim} \times \frac{S}{Dim}$ blocks and obtaining the distorted image by allowing random permutations within each block.

To be specific, let $Z(i,j)^{(k)}$ be a generic pixel of the distorted image Z belonging to the k-th block (with $k = \left\{ 1, 2, ..., \left(\frac{S}{Dim} \right)^2 \right\}$), then $Z(i,j)^{(k)} = Y \pi_k(i,j)^{(k)}$, where Y is the original image and $\pi_k(i,j)$ is a random permutation of the indices belonging to the k-th block.

Increasing the value of Dim allows to consider a larger number of transformations but, at the same time, affects the image quality leading to increasingly annoying artifacts.

Fig. 3.2 shows the effect of this kind of distortion on the two sample images (considering $Dim = 5$). It is easy to deduce that the transformations generated with this model are not perceptually admissible, even when a small value of Dim is used. For this reason a new model is introduced.

3.2.2   Local Permutation with Cancelation and Duplication

By focusing on the 1D-case, let $y = \{y(1), y(2)....y(n)\}$ be a generic signal and let $z = \{z(1), z(2)....z(n)\}$ be the distorted version of y. The Local Permutation with Cancelation and Duplication model (LPCD model) states that $z(i) = y(i + \Delta_i)$ where Δ_i is a sequences of i.i.d random variables uniformly distributed in a predefined interval $I = [-\Delta, \Delta]$. For simplicity we assume that Δ_i can take only integer values in I. This way, the values assumed by the samples of z are chosen among those of y. The above model yields an interesting interpretation of the distorted signal z. To introduce it, it is convenient to describe the LPCD model as a channel $W(z|y)$ defined as follows (neglecting edge effects):

$$W(z|y) = \prod_{i=1}^{n} W(z(i)|y_{i-\Delta}^{i+\Delta}), \qquad (3.1)$$

where: (i) y_i^j, for $i \leq j$ denotes $(y(i), y(i + 1), \ldots, y(j))$ (a similar notation applies to z), and (ii) :

$$W(z(i)|y_{i-\Delta}^{i+\Delta}) = \frac{1}{2\Delta + 1} \sum_{k=-\Delta}^{\Delta} \mathbf{1}\{z(i) = y(i - k)\}, \qquad (3.2)$$

where $\mathbf{1}\{z(i) = y(i - k)\}$ denotes the indicator function[1] of the event $\{z(i) = y(i - k)\}$. According to the above equation the LPCD channel $W(z(i)|y_{i-\Delta}^{i+\Delta})$ assigns the same probability, $1/(2\Delta+1)$, and independently, to all possible values of $k \in \{-\Delta, -\Delta + 1, \ldots, \Delta\}$ and picks $z(i) = y(i - k)$. However *any* other probability assignment $W(z(i)|y_{i-\Delta}^{i+\Delta})$ is allowed. Likewise, the probability law of y does not need to be known (except the fact that it is memoryless). An equivalent representation of this model is obtained by defining $u(i) = y_{i-\Delta}^{i+\Delta}$. Here, if $|y(i)|$ are i.i.d., then $|u(i)|$ is a first–order Markov process. Also, the

---

[1]An indicator function is a function defined on a set X that indicates membership of an element in a subset A of X. The indicator function of a subset A of a set X is a function $1_A : X \rightarrow \{0, 1\}$ defined as:

$$1_A(x) = \left\{ \begin{array}{ll} 1 & \text{if} \;\; x \in A \\ 0 & \text{if} \;\; x \notin A \end{array} \right\}$$

channel W from $u = (u(1), \ldots, u(n))$ to y is obviously memoryless according to (3.1). Thus, z is governed by a hidden Markov process:

$$Q(z) = \sum_u \prod_{i=1}^{n} [P(u(i)|u(i-1))W(z(i)|u(i))]. \qquad (3.3)$$

The above interpretation of the LPCD model may open the way to the definition of optimum embedding and detection strategies along the same lines described in (Merhav 2005).

To extend the 1D-LPCD model to the two-dimensional case, if $Z(i,j)$ is a generic pixel of the distorted image Z, let

$$Z(i,j) = Y(i + \Delta_h(i,j), j + \Delta_v(i,j)), \qquad (3.4)$$

where Y is the original image and $\Delta_h(i,j)$ and $\Delta_v(i,j)$ are i.i.d. integer random variables uniformly distributed in the interval $[-\Delta, \Delta]$.

Fig. 3.3 shows the sample images distorted with the LPCD model with $\Delta = 2$. Is possible to note that this distortion, even if it is not yet a perceptually admissible distortion, does not present the annoying block artifacts visible in Fig. 3.2, this is due to the overlapping of the windows of the possible displacements of neighboring pixels.

3.2.3 Constrained LPCD

An important limitation of the LPCD model is the lack of memory. This is surely a problem from a perceptual point of view: with no constraints on the smoothness of the displacement field there is no guarantee that the set of LPCD distortions is perceptually admissible even by considering very small values of Δ (as shown in Fig. 3.3).

One way to overcome the limitation of the LPCD model, and to obtain better results from a perceptual point of view, is to require that the sample order, in the 1D case, is preserved (thus introducing memory in the system). In practice, the displacement of each element i of the distorted sequence z is conditioned on the displacement of the element $i-1$ of the same sequence. In formulas, $z(i) = y(i + \Delta_i)$ where Δ_i is a sequence of i.i.d integer random variables uniformly distributed in the interval $I = [\max(-\Delta, \Delta_{i-1} - 1), \Delta]$.

(a)   (b)

Figure 3.3: *Local Permutation with Cancelation and Duplication (LPCD) with $\Delta = 2$: (a) Lena; (b) Duomo.*
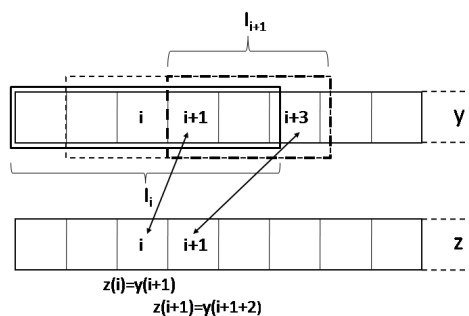
Figure 3.4: *Constrained LPCD with $\Delta = 2$ (one-dimensional case).*

In the sequel we will refer to this new class of DAs as C-LPCD (Constrained LPCD).

Fig. 3.4 illustrates the behavior of the C-LPCD model in the 1D case, with $\Delta = 2$. We know that $z(i) = y(i + \Delta_i)$, let us assume that Δ_i is chosen in the interval $I_i = [-2, 2]$ (the solid line box) and that $\Delta_i = 1$, it means that $z(i) = y(i + 1)$. At the next step we know that $z(i + 1) = y(i + 1 + \Delta_{i+1})$

where Δ_{i+1}, due to the position of the pixel $z(i)$, must be chosen in the interval $I_{i+1} = [0, 2]$ (the bold dotted line box). The interval I_{i+1} is smaller than I_i because the position of the element $i + 1$ cannot precede that of the element i. For example, Δ_{i+1} could be equal to 2 yielding $z(i + 1) = y(i + 3)$ (as indicated by the arrow on the right).

To extend the C-LPCD model to the 2-D case we let $Z(i, j)$ be equal to

$$Y\left(i + \Delta_h\left(i, j\right), j + \Delta_v\left(i, j\right)\right) \tag{3.5}$$

where $\Delta_h\left(i, j\right)$ is randomly chosen, to preserve the horizontal sample order, in:

$$I_h = \left[\max\left(-\Delta, +\Delta_h\left(i - 1, j\right) - 1, +\Delta_h\left(i - 1, j - 1\right) - 1\right), \Delta\right],$$

and $\Delta_v\left(i, j\right)$ is randomly chosen, to preserve the vertical sample order, in:

$$I_v = \left[\max\left(-\Delta, +\Delta_v\left(i, j - 1\right) - 1, +\Delta_v\left(i - 1, j - 1\right) - 1\right), \Delta\right].$$

In other words, the horizontal and vertical displacements of the pixel (i, j) are limited by the horizontal and vertical displacements of the pixels $(i - 1, j), (i, j - 1)$ and $(i - 1, j - 1)$. Specifically, the lower bound of the intervals I_h and I_v of each pixel (i, j) is determined by the positions of the pixels $(i - 1, j), (i, j - 1)$ and $(i - 1, j - 1)$ while the upper bound depends only on Δ.

Fig. 3.5 shows the Duomo and the Lena images distorted with the C-LPCD model (considering $\Delta = 2$).

After a visual inspection conducted on the images we can deduce that the constrained LPCD is a more perceptually admissible model than the previous ones.

Another simpler way to extend the C-LPCD model to the 2-D case, is to apply the 1-D model to each row of the image, obtaining the horizontal displacements, and then to each column of the image, obtaining the vertical displacements. This way to generate the distortions, however, returns worst results from a perceptual point of view. In the rest of this chapter, by 2-D C-LPCD the model described by Eq. (3.5) is meant.

(a)   (b)

Figure 3.5: *Constrained LPCD (C-LPCD) with $\Delta = 2$: (a)Lena; (b)Duomo.*

3.2.4   Multiresolution C-LPCD model

To further improve the C-LPCD and LPCD models making the distortion less perceptible, we considered a multiresolution version of these distortions, whereby the transformations are applied at different resolutions to obtain the global displacement field: a low resolution displacement field is first generated, then a full size displacement field is built by means of bilinear interpolation. The full resolution field is applied to the original image to produce the distorted image.

More specifically the multiresolution models consist of two steps. Let $S \times S$ be the size of the image (for sake of simplicity we assume S is a power of 2). To apply the LPCD (or C-LPCD) model at the $L-$th level of resolution two displacement fields $\delta_h(i,j)$ and $\delta_v(i,j)$ with size $\frac{S}{2^L} \times \frac{S}{2^L}$ are generated. Then the full resolution fields $\Delta_h(i,j)$ and $\Delta_v(i,j)$ are built by means of bilinear interpolation. In formulas, we first interpolate the field along the horizontal

direction

$$\Delta'_h(i,j) = \delta_h\left(\left\lfloor\frac{i}{2^L}\right\rfloor, \left\lfloor\frac{j}{2^L}\right\rfloor\right) \cdot \left[1 - \left(i - \left\lfloor\frac{i}{2^L}\right\rfloor\right)\right] +$$
$$\delta_h\left(\left\lceil\frac{i}{2^L}\right\rceil, \left\lfloor\frac{j}{2^L}\right\rfloor\right) \cdot \left[i - \left\lfloor\frac{i}{2^L}\right\rfloor\right] \tag{3.6}$$

$$\Delta''_h(i,j) = \delta_h\left(\left\lfloor\frac{i}{2^L}\right\rfloor, \left\lceil\frac{j}{2^L}\right\rceil\right) \cdot \left[1 - \left(i - \left\lfloor\frac{i}{2^L}\right\rfloor\right)\right] +$$
$$\delta_h\left(\left\lceil\frac{i}{2^L}\right\rceil, \left\lceil\frac{j}{2^L}\right\rceil\right) \cdot \left[i - \left\lfloor\frac{i}{2^L}\right\rfloor\right]. \tag{3.7}$$

Then we interpolate vertically the values produced by the horizontal interpolation, obtaining the final, full resolution, field Δ_h (the same procedure is used for Δ_v):

$$\Delta_h(i,j) = \Delta'_h(i,j) \cdot \left[1 - \left(j - \left\lfloor\frac{j}{2^L}\right\rfloor\right)\right] + \Delta''_h(i,j) \cdot \left[j - \left\lfloor\frac{j}{2^L}\right\rfloor\right] \tag{3.8}$$

Note that in this way non-integer displacement values are introduced[2]. The full resolution displacement fields Δ_h and Δ_v are used to generate the warped image Z as follows:

$$Z(i,j) = Y(i - \Delta_h(i,j), j - \Delta_v(i,j)). \tag{3.9}$$

As opposed to the original version of LPCD and C-LPCD, however, the presence non-integer displacements is now possible due to the interpolation rule defined by Eq. (3.6-3.8). To account for this possibility, whenever the displacement vector points to non integer coordinates of the original image, the gray level of the distorted image $Z(i,j)$ is computed by means of bilinear interpolation. While the above interpolation does not have a significant impact on the visual quality of the distorted image, the possible introduction of new gray levels that were not present in the original image, complicates the LPCD and C-LPCD models, by making impossible, for instance, to describe the distorted signal as a hidden markov process (as we did in section 3.2.2).

The pseudo-code description of the multiresolution version of LPCD model is provided by Algorithms 1 and 2 in the Appendix of this chapter.

---

[2]It is still possible to obtain integer displacements by applying a nearest neighbor interpolation instead of a bilinear one (of course at the expense of the smoothness of the displacement field).

3.3   A Markov chain interpretation of the C-LPCD model

After a visual inspection conducted on a set of images distorted with the MC-LPCD and C-LPCD models, we observed that rather surprisingly changing the value of Δ_{max} does not seem to change the perceptual quality of the images. This can be explained by resorting to the theory of Markov Chains.

Let us go back to the constrained LPCD model described in Section 3.2.3, and let us focus on the 1-D case, it is possible to design a Markov chain whose states correspond to the possible sizes of the interval

$$I = \left[\max\left(-\Delta, -\Delta_{i-1} - 1\right), \Delta\right]$$

representing the set of elements of the original sequence among which the element of the distorted sequence is selected. By choosing for example $\Delta = 2$, the state space S is $\{1, 2, 3, 4\}$ where the state S_i corresponds to a size of I equal to $i + 1$, i.e. state 2 corresponds to the set $I = \{0, 1, 2\}$ of size 3, while state 4 corresponds to $I = \{-2, -1, 0, 1, 2\}$ of size 5 and it is the maximum size that I can have (with $\Delta = 2$). The 4×4 transition matrix of this Markov Chain, representing all the possible one-step transition probabilities among the states of the chain, is:

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{2}{5} \end{bmatrix}$$

and the corresponding graph is reported in Fig. 3.6. To exemplify the above concepts, let us assume to be in state 3 at a generic step i, corresponding to $I = \{-1, 0, 1, 2\}$. By looking at the graph, it means that, at step $i + 1$ we have one fourth of probability to be in state 1, one fourth to be in state 2 and the same probability to be in state 3 or 4.

In a more general case, given Δ, the maximum size of I is equal to $N = 2\Delta + 1$ and the transition matrix of size $2\Delta \times 2\Delta$ is:
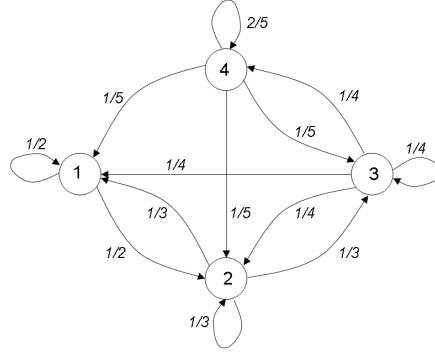
Figure 3.6: *Graph of the Markov chain describing the C-LPCD model, for $\Delta = 2$.*

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & .. & .. & .. & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & .. & .. & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & .. & 0 \\ .. & .. & .. & .. & .. & .. & .. \\ \frac{1}{2\Delta+1} & \frac{1}{2\Delta+1} & \frac{1}{2\Delta+1} & .. & .. & .. & \frac{2}{2\Delta+1} \end{bmatrix} \qquad (3.10)$$

where each element of the matrix p_{ij} is the transition probability of going from state i to state j. From Markov chain theory (Norris 1998) we know that if a Markov chain is aperiodic and irreducible then there exists a unique limit distribution that can be found by solving the following system:

$$\begin{cases} \pi_j = \sum_{i \in S} \pi_i p_{ij} & \forall j \in S \\ \sum_{j \in S} \pi_j = 1 \end{cases} \qquad (3.11)$$

where $\pi_i = P\{X = i\}$ is the limit probability to be in state i after n steps.

It is easy to show that in our case the chain is irreducible and aperiodic thus it is possible to find a unique limit distribution by solving the system in Eq. (3.11) that now is equivalent to:

$$
\begin{cases}
\pi_1 = \sum\limits_{i=1}^{2\Delta} \pi_i \cdot \frac{1}{i+1} \\
\pi_j = \sum\limits_{i=j-1}^{2\Delta} \pi_i \cdot \frac{1}{i+1} \quad \text{with j going from 2 to } 2\Delta - 1 \\
\pi_{2\Delta} = \pi_{2\Delta-1} \cdot \frac{1}{2\Delta} + \pi_{2\Delta} \cdot \frac{2}{2\Delta+1} \\
\sum\limits_{j=1}^{2\Delta} \pi_j = 1
\end{cases}
$$

After some algebraic manipulations it is possible to find the limit probability distribution of states as expressed in the following equation:

$$
\begin{cases}
\pi_1 = \frac{1}{\sum\limits_{k=0}^{2\Delta-2} \frac{1}{k!} + \frac{2\Delta+1}{(2\Delta)!}} \\
\pi_j = \frac{1}{(j-1)!}\pi_1 \quad \text{with j going from 2 to } 2\Delta - 1 \\
\pi_{2\Delta} = \frac{2\Delta+1}{(2\Delta)!}\pi_1
\end{cases}
\qquad (3.12)
$$

Fig. 3.7.a shows the limit probability of the states (i.e, the probability to be in state i after n steps when $n \to \infty$) as a function of N ($N = 2\Delta + 1$). Interestingly, regardless of the value of N and therefore of Δ, states 1 and 2 are by far the most probable states ($\pi_1 = \pi_2$ except when $\Delta = 1$). This is due to the second equation in (3.12) and explains why changing the value of Δ does not change the perceptual quality of the image. In fact even by increasing the value of Δ states 1 and 2 are the most probable states, but state 2 corresponds to the interval $I = \{0, 1, 2\}$ of size 3 so even increasing the value of N the model tends to prefer small displacements. Furthermore it is easy to verify that the states reach their limit probability distribution already for very small values of n.

Of course, by increasing the value of N the limit probabilities of the states decreases because the number of possible states increases, in particular the minimum value of π_1 is approximately equal to 0.3679, as was found by solving the following limit:

$$
\lim_{N \to \infty} \pi_1 = \lim_{N \to \infty} \frac{1}{\sum\limits_{k=0}^{2N-2} \frac{1}{k!} + \frac{2N+1}{(2N)!}} = \frac{1}{e} \approx 0.3679
$$
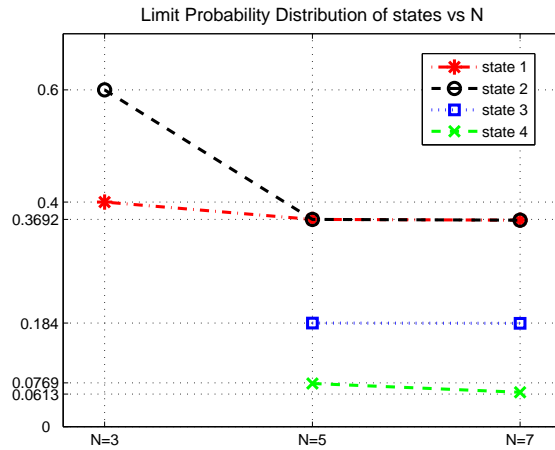
(a)
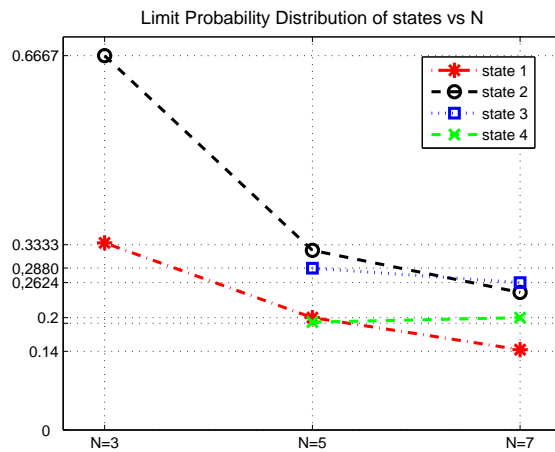
(b)

Figure 3.7: *Limit probability distribution of states versus* Δ *(N = 2Δ + 1): (a) original Markov chain ; (b) modified Markov chain.*

Interestingly, by observing Fig. 3.7.a it is possible to note that this limit is already reached for small values of N. Since $\pi_1 = \pi_2$, it is clear why changing the value of Δ does not change the perceptual quality of the image.

To avoid this undesirable effect and to allow the model to generate a larger

variety of displacement fields, we modified the Markov chain described by the graph in Fig. 3.4 by changing the transition probabilities among the states in order to give a greater probability to those transitions that allow a larger interval I. A way to do this is to assign the same probability (equal to $\frac{1}{2\Delta+1}$) to those transitions that cause a decrease of the size of I, corresponding to the elements i, j with $i = 1, .., \Delta$ and $j = 1, .., i$ of the transition matrix, and to assign all the remaining probabilities, equal to $1 - \sum_{j=1}^{i} p_{ij}$, to the transition corresponding to the element i, j with $i = 1, .., \Delta$ and $j = i + 1$, i.e. those transitions whose effect is to enlarge the interval I. The corresponding transition matrix becomes:

$$P = \begin{bmatrix} \frac{1}{2\Delta+1} & \frac{2\Delta}{2\Delta+1} & 0 & .. & .. & .. & 0 \\ \frac{1}{2\Delta+1} & \frac{1}{2\Delta+1} & \frac{2\Delta-1}{2\Delta+1} & 0 & .. & .. & 0 \\ \frac{1}{2\Delta+1} & \frac{1}{2\Delta+1} & \frac{1}{2\Delta+1} & \frac{2\Delta-2}{2\Delta+1} & 0 & .. & 0 \\ .. & .. & .. & .. & .. & .. & .. \\ \frac{1}{2\Delta+1} & \frac{1}{2\Delta+1} & \frac{1}{2\Delta+1} & .. & .. & .. & \frac{2}{2\Delta+1} \end{bmatrix} \tag{3.13}$$

and the system to be solved to the limit distribution is:

$$\begin{cases} \pi_1 = \frac{1}{2\Delta+1} \sum_{i=1}^{2\Delta} \pi_i \\ \pi_j = \frac{1}{2\Delta+1} \sum_{i=j}^{2\Delta} \pi_i + \frac{2\Delta+2-j}{2\Delta+1} \cdot \pi_{j-1} & \text{with j going from 2 to } 2\Delta - 1 \\ \pi_{2\Delta} = \frac{2}{2\Delta+1}(\pi_{2\Delta-1} + \pi_{2\Delta}) \\ \sum_{j=1}^{2\Delta} \pi_j = 1 \end{cases}$$

$$\tag{3.14}$$

Fig. 3.7.b shows the limit probability distribution of states versus Δ_{max} of the new Markov chain found in a numerical way.

By comparing Fig. 3.7.a and 3.7.b, it is evident that with the modified Markov Chain is possible to obtain larger displacement fields because regardless of the value of Δ all the states have almost the same limit probabilities.

Fig. 3.8 shows the sample images distorted with MC-LPCD applied at different levels of resolution using the modified Markov chain applied to the 2D case: in (a) the level of resolution is 3, (b) is the case of $L = 4$, and in figure (c) the level of resolution is 6 (in all the cases $\Delta = 2$).

By observing the last figures it is clear that this model provides a better way to incorporate perceptual considerations. In particular the image quality increases, from a perceptual point of view, if the MC-LPCD model is applied to a lower level of resolution but, in the meantime, the number of possible distortions decreases.

In Fig. 3.9, two examples of displacement fields generated with the LPCD model with $L = 6$ and $N = 5$ (3.9.a) and the C-LPCD model with $L = 5$ and $N = 5$ (3.9.b) are given: as expected, by applying the model to a lower level of resolution, it is possible to obtain a more uniform field (for the purpose of visibility the total displacement field is cropped and only one vector every sixteen samples is depicted in the figure).

3.4 Perceptual analysis

Given the good performance of the MC-LPCD model (the modified version presented in Section 3.3), we carried out some objective and subjective tests in order to evaluate the perceptual impact of the artifacts introduced by this model of distortions. To do so, let us observe that from a perceptual point of view MC-LPCD has a different behavior for different values of N and for different levels of resolution L.

The goal of the tests was to establish the sensitivity of the visual system to the geometric distortions introduced by the model as a function of the control parameters N and L. In this way we were able to identify the range of variation of the control parameters that do not affect image quality.

For objective test, we used the PSNR measurement and state-of-the-art metrics such as the Universal Quality Index (Wang and Bovik 2002b), the SSIM-index (Wang, Bovik, Sheikh and Simoncelli 2004) and the RST based metric developed by Setyawan et al.(Setyawan et al. 2003) (an explanation of this metric is provided in Sec. 6.4.3). It is worth mentioning that despite the great research effort, there is still a lack of objective visual quality metrics

(a)

(b)

(c)

Figure 3.8: *MC-LPCD applied at different resolution levels: (a) $L = 3$, (b) $L = 4$,(c) $L = 6$.*
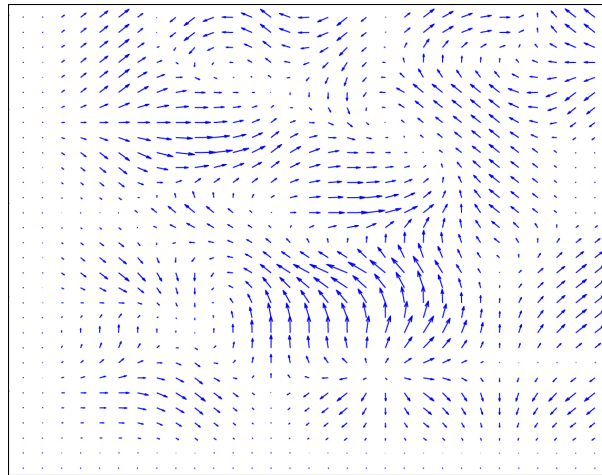
(a)
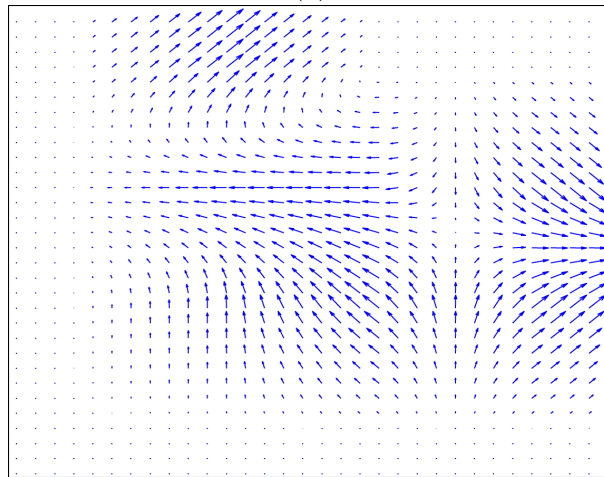
(b)

Figure 3.9: *Examples of displacement fields generated with LPCD DA's: (a) C-LPCD with L = 5 and N = 5; (b) C-LPCD with L = 6 and N = 5.*

suitable for geometric distortions and the few metrics proposed so far are ineffective for most of the distortions. For this reason in the following two metrics for the evaluation of this kind of distortions will be introduced.

In the subjective test we applied the Two Alternatives Forced Choice

(2AFC) test(MADIGAN and WILLIAMS 1987): the users were asked to compare two images at time, the original image and the distorted image, and to indicate which one was the original. We decided to use a single stimulus test because we are interested in finding the perceptually admissible distortions[3].

The source image database used in both tests is shown in Fig. 3.10. It includes sixteen gray scale images, 512×512 pixel in size, and was derived from a set of source images that reflects adequate diversity in image contents. The images, in fact, included pictures of faces, houses, natural scenes and images without any specific object of interest. Some images have high activity, while some do not have much structures and are mostly smooth.

We chose to distort the source images through the MC-LPCD model using different distortion types obtained by changing the dimension of N ($N = 5$, $N = 7$, and $N = 9$) and the level of resolution ($L = 6$, $L = 5$, $L = 4$, $L = 3$, $L = 2$). Specifically we produced thirteen distorted versions for each image (all the possible combinations of N and L except $N = 9$ $L = 2$ that always generates a visible distortions and $N = 9$ $L = 6$ because the level of resolution is smaller than N), for a total of 208 images.

3.4.1   The objective test

Many image quality assessment algorithms have been proposed in the scientific literature and many of them provide good results with certain kinds of distortions (e.g., JPEG compression), but the effectiveness of these metrics degrades when they are applied to a set of images distorted geometrically. However, just for completeness, we present the results we obtained by applying some of these metrics to the images used in the subjective test. For a given set of parameters, the comparison of the subjective test described in the following subsection with the objective quality measures will allow to establish the difficulty of objective metrics in predicting the perceived amount of geometrical distortions present in an image.

The results of the objective test are shown in Fig. 3.11. We do not show the results obtained with the Universal Quality Index because SSIM index is an improvement of it.

---

[3]An explanation of subjective tests classification and procedures is provided in Sec. 6.4

Figure 3.10: *Source image dataset used for the 2AFC test.*

We also applied the RST based metric developed by Setyawan et. al, but the results we obtained are not meaningful. In our case, in fact, this metric is not able to predict the visual quality of the images because it does not find a local RST or affine transform (neither in a small interval) approximating the geometric distortion introduced by the MC-LPCD model.

3.4.2 The two alternative forced-choice test

The subjective test we used is the 2AFC test. Two stimuli are presented at each trial. One of these stimuli is the original image; the other is a dis-
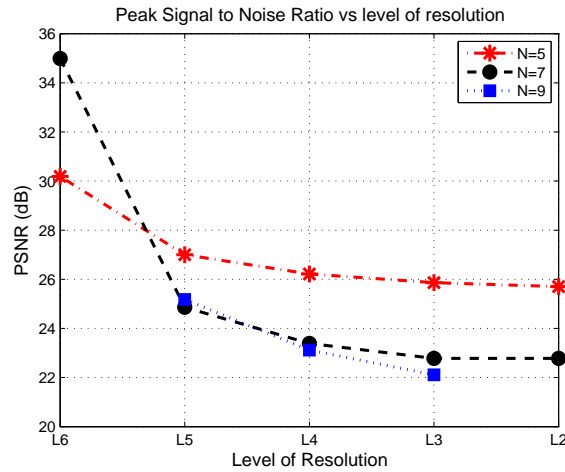
(a)
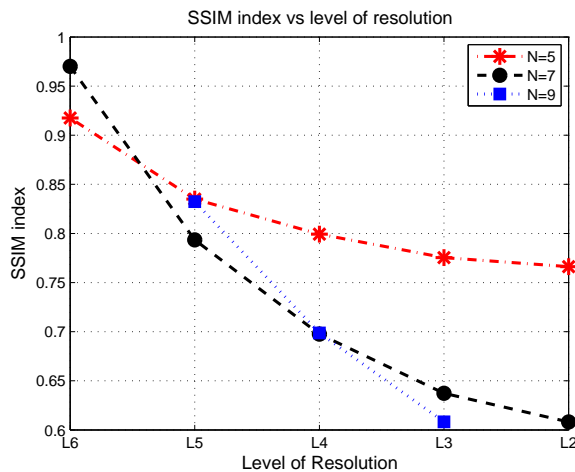
(b)

Figure 3.11: *Plot of objective metrics versus the level of resolution used in the MC-LPCD model: (a) Peak Signal to Noise Ratio; (b) Structural Similarity based IMage index.*

torted version of the image. The observer is asked to select the original image. Procedures for such experiment have been designed by following the ITU-T Recommendation P.910 (Int 1996), which suggests standard viewing

conditions, criteria for the selection of observers and test material, assessment procedures, and data analysis methods.

The experiments were conducted by using the VP800 video card of the Cambridge Research Systems together with a high resolution digital monitor Mitsubishi DiamondPro 2070 with the external adaptor ViSaGe 71.02.00D2[4]. To have a correct color representation a luminosity calibration was previously carried out through a videocamera ColorCAL[4].

The tests involved a panel of fifteen subjects, all naives with respect to image quality assessment methods and image impairments. Each subject was individually briefed about the goal of the experiment, and given a demonstration of the test. Subjects were shown images in a random order, the randomization was different for each subject. The test was performed in a dark room in free viewing conditions.

In order to analyze the results obtained with the subjective test an hypothesis test was conducted. This test tells us whether the subjective test, based on the number of sample points used, allows to make a statistically sound conclusion about the visibility of distortions introduced with the MC-LPCD model. The hypothesis test is based on a probability test: we tested the hypothesis H_0 that the probability $p = P(\mathcal{A})$ of an event $\mathcal{A}$ equals a given constant p_0, using as data the number k of successes of $\mathcal{A}$ in n trials. Specifically we tested the hypothesis that the probability p that the users choose the original image in the 2AFC test (event $\mathcal{A}$) is equal to $p_0 = \frac{1}{2}$ (it means that the original image and the distorted image are perceptually indistinguishable) using as observable data the number of times that the users decided for the original images in the 240 total comparisons (16 images $\times$ 15 users). The alternative hypothesis is that $P(\mathcal{A}) < \frac{1}{2}$ and we use $\alpha = 0.05$ as significativity level (Papoulis 1984) .

With the hypothesis test we derived the maximum admissible distortion that can be applied before the distortion becomes visible using the MC-LPCD model. Specifically, the results of the hypothesis test are shown in table 3.1. For each level of resolution and for each typology of images, we found the maximum value of N that can be used while keeping the distortion invisible (the hypothesis $p_0 = \frac{1}{2}$ is accepted). The empty boxes correspond to cases in

---

[4]http://www.crsltd.com/catalog/visage/index.html

	L=6	L=5	L=4	L=3	L=2
faces	N=7	N=5	–	–	–
houses	N=7	N=5	–	–	–
nature	N=9	N=7	N=5	–	–
others	N=7	N=5	–	–	–

Table 3.1: *Maximum admissible distortion that can be applied to the images before the distortion becomes visible using the MC-LPCD model.*

which it was not possible to find a value of N ensuring an invisible distortion. By looking at the results we note that, as expected, in images that do not have many structures, like natural images, the distortions are less visible. Furthermore the results in table 3.1 show the difficulty of the objective metrics in predicting the degradation introduced by the model in the images: for example the plot $N = 5$ in Fig. 3.11.a shows a loss of quality going from $L = 6$ to $L = 5$ while the subjective test assures the invisibility of the distortion in both cases. In the same way the loss of about 5 dB in the PSRN plot between the configurations $N = 5$ $L = 6$ and $N = 7$ $L = 6$ is not meaningful according table 3.1.

The LPCD model proposed in this chapter is a first attempt to have a mathematical formulation of local geometric distortions that allow to control the impact of image quality by means of a reduced set of parameters. The main advantages of this model are the simple formulation and in the same time the possibility to have a great number of perceptually admissible distortions, that is particular interesting in the field of digital watermarking. In fact, the larger is the DA space, the more difficult will be to recover the synchronization between the embedded and the detector, both in terms of complexity and accuracy. This will be explained in details in chapter 5.

Appendix

Algorithm 1.*Multiresolution LPCD model*

1: Read image to be distorted Y, read size of the window Δ, read level of resolution L

2: $\text{dim} = \frac{\text{size(image)}}{2^L}$ {size of the low resolution displacement field}

3: Initialize matrices δ_h and δ_v of horizontal and vertical displacement fields to 0

4: **for** $i = 1 : \text{dim}$ **do**

5:     **for** $j = 1 : \text{dim}$ **do**

6:         **if** $(i < \Delta + 1)$ or $(j < \Delta + 1)$ **then**

7:             $\delta_h(i,j)$ and $\delta_v(i,j)$ are randomly chosen in $[-(\min(i,j)-1); (\min(i,j)-1)]$

8:         **else if** $(i > \text{dim} - \Delta)$ or $(j > \text{dim} - \Delta)$ **then**

9:             $\delta_h(i,j)$ and $\delta_v(i,j)$ are randomly chosen in $[-(\text{dim}-\max(i,j)); (\text{dim}-\max(i,j))]$

10:         **else**

11:             $\delta_h(i,j)$ and $\delta_v(i,j)$ are randomly chosen in $[-\Delta; \Delta]$

12:         **end if**

13:     **end for**

14: **end for**

15: Resize the displacement fields given by δ_h and δ_v to the image size through bicubic interpolation provided by the matlab function *imresize*{to obtain the high resolution displacement fields Δ_h an Δ_v}

16: **for** $i = 1 : \text{size(image)}$ **do**

17:     **for** $j = 1 : \text{size(image)}$ **do**

18:         $Z(i,j) = Y(i + \Delta_h(i,j), j + \Delta_v(i,j))$ {Apply the displacement fields to the image, to obtain the distorted image Z, by means of bicubic interpolation}

19:     **end for**

20: **end for**

Algorithm 2.*Multiresolution Constrained LPCD model (modified version)*

1: Read image to be distorted, read size of the window Δ, read level of resolution L

2: dim $= \dfrac{\text{size(image)}}{2^L}$ {size of the low resolution displacement field}

3: Initialize matrices δ_h and δ_v of horizontal and vertical displacement fields to 0

4: **for** $i = 1 : \text{dim}$ **do**

5:    **for** $j = 1 : \text{dim}$ **do**

6:       **if** $(i < \Delta + 1)$ or $(j < \Delta + 1)$ **then**

7:          $\delta_h(i,j)$ and $\delta_v(i,j)$ are randomly chosen in $[-(\min(i,j)-1);(\min(i,j)-1)]$

8:       **else if** $(i > \text{dim} - \Delta)$ or $(j > \text{dim} - \Delta)$ **then**

9:          $\delta_h(i,j)$ and $\delta_v(i,j)$ are randomly chosen in $[-(\text{dim}-\max(i,j));(\text{dim}-\max(i,j))]$

10:       **else**

11:          $\delta_h(i,j)$ is chosen in $Ix = [\max(\Delta, \delta_h(i-1,j)-1), \Delta]$ with a distribution vector $P = [1 - \frac{\text{size}(Ix)-1}{\Delta}; \frac{1}{\Delta}; ..; \frac{1}{\Delta}]$

12:          $\delta_v(i,j)$ is chosen in $Iy = [\max(\Delta, \delta_v(i-1,j)-1), \Delta]$ with a distribution vector $P = [1 - \frac{\text{size}(Iy)-1}{\Delta}; \frac{1}{\Delta}; ..; \frac{1}{\Delta}]$

13:       **end if**

14:    **end for**

15: **end for**

16: Resize the displacement fields given by δ_h and δ_v to the image size through bicubic interpolation provided by the matlab function *imresize*{to obtain the high resolution displacement fields Δ_h an Δ_v}

17: **for** $i = 1 : \text{size(image)}$ **do**

18:    **for** $j = 1 : \text{size(image)}$ **do**

19:      $Z(i,j) = Y(i + \Delta_h(i,j), j + \Delta_v(i,j))$ {Apply the displacement fields to the image, to obtain the distorted image Z, by means of bicubic interpolation}

20:    **end for**

21: **end for**

Chapter 4

A model based on Markov Random Fields

4.1   Introduction

The previous chapter introduced a model to describe perceptually admissible geometric distortions. The idea behind the LPCD model is the assumption that the displacement field of the pixel at location (i, j) is conditioned to those of the pixels at locations $(i-1, j), (i, j-1)$ and $(i-1, j-1)$. The sensitivity of the human visual system to geometrical transformation as a function of the control parameters L and N was then established through a 2AFC test, that permitted to identify the range of these parameters producing distortions that do not affect image quality.

One problem with the C-LPCD class of geometric distortions is that it does not take into account the two-dimensional nature of images, since it is based on a one-dimensional Markov chain. To overcome this limitation, in this chapter we introduce a new class of geometric distortions based on the theory of Markov Random Fields (MRF). In the sequel we will refer to this class of distortions as MF distortions.

4.2   Theoretical background

The theory of Markov Random Fields is a branch of probability theory for analyzing the spatial or contextual dependencies of physical phenomena. The foundations of this theory may be found in statistical physics of magnetic materials (Ising models, spin glasses, etc..) and also in solids and crystals, where the molecules are arranged in a lattice structure and there are interactions with close neighbors (for example, Debye's theory for the vibration of atoms in a lattice is based on a model of quantum harmonic oscillators with coupling among nearest neighbors (Caputo and Niemann 2001)). MRF are often used

in image processing applications, because they provide a model for describing the correlation among neighboring pixels (Li 1995).

4.2.1 Model description

Many vision problems can be posed as labeling problems in which the solution is a set of labels assigned to image pixels or features. A labeling problem is specified in terms of a set of sites and a set of labels. Let $\mathcal{S} = \{1, ..., m\}$ be a discrete set of m sites in which $1, ..., m$ are indices (a site often represents a point or a region in the Euclidean space such as an image pixel or an image feature). A label is an event that may happen to a site. Let $\mathcal{L} = \{l_1, ..., l_n\}$ be a set of labels. The labeling problem is to assign a label from $\mathcal{L}$ to each of the sites in $\mathcal{S}$. In the terminology of random fields, a labeling is called a configuration.

The sites in $\mathcal{S}$ are related to one another via a neighborhood system. A neighborhood system for $\mathcal{S}$ is defined as $N = \{N_i | i \in \mathcal{S}\}$ where N_i is the set of sites neighboring i. The neighboring relationship has the following properties:

1. a site is not a neighbor of itself: $i \notin N_i$

2. the neighboring relationship is mutual: $i \in N_{i'} \Leftrightarrow i' \in N_i$

If $\mathcal{S}$ is a regular lattice, the neighboring set of i is often defined as the set of nearby sites within a radius of r:

$$N_i = \{i' \in \mathcal{S} | [dist(i, i')]^2 \leq r, i' \neq i\}$$

Once introduced a set $\mathcal{S}$ and a neighborhood system N, it is possible to define a clique c for $(\mathcal{S}, N)$ as a subset of sites in $\mathcal{S}$. A clique could consist either of a single site $c = \{i\}$ (single-site clique), or of a pair of neighboring sites $c = \{i, i'\}$ (pair-sites cliques), or of a triple of neighboring sites $c = \{i, i', i''\}$ (triple-sites cliques), and so on.

The collections of single-site, pair-site and triple-site cliques will be denoted by $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{C}_3$, respectively, where:

$$\mathcal{C}_1 = \{i | i \in \mathcal{S}\}$$

$$\mathcal{C}_2 = \{\{i, i'\} | i' \in \mathcal{N}_i, i \in \mathcal{S}\}$$

$$\mathcal{C}_3 = \{\{i, i', i''\} | i, i', i'' \in \mathcal{S} \text{ are neighbors to one another}\}$$

The collection of all cliques for $(\mathcal{S}, N)$ is denoted by $\mathcal{C}$.

Fig. 4.1 shows a first order neighborhood system, also called a 4-neighborhood system, with the four corresponding pair-sites cliques. The **x** symbol denotes the considered site and the letters indicate its neighbors.
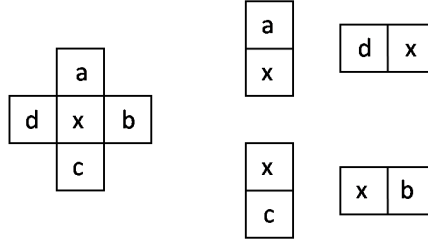
Figure 4.1: *Structure of a first order neighborhood system and corresponding pair-sites cliques.*

A random field $F = \{F_1, F_2, ..., F_m\}$ is a family of random variables defined on a set $\mathcal{S}$, in which each random variable F_i takes a value f_i in a set of labels $\mathcal{L}$.

F is said to be a Markov Random Field (MRF) on $\mathcal{S}$ with respect to a neighborhood system N if and only if the two following conditions are satisfied:

$$P(f) > 0, \quad \forall f \in \mathcal{L}^m \qquad \text{(positivity)} \qquad (4.1)$$

$$P(f_i | f_{S-\{i\}}) = P(f_i | f_{N_i}), \quad \forall i \in \mathcal{S} \quad \text{(Markov property)} \qquad (4.2)$$

where $f = \{f_1, ..., f_m\}$ is a configuration of F (corresponding to a realization of the field), $P(f)$ is the joint probability $P(F_1 = f_1, ..., F_m = f_m)$ of the joint event $F = f$, i.e., it measures the probability of the occurrence of a

particular configuration, and

$$f_{N_i} = \{f_{i'}, i' \in N_i\} \tag{4.3}$$

denotes the set of values at the sites neighboring i. The positivity is due to technical reasons, since it is a necessary condition if we want the Hammersley-Clifford theorem (see below) to hold (Besag 1986).

To exploit the characteristics of a Markov Random Fields in a practical way, we need to refer to the Hammersley-Clifford theorem (Li 1995) for which F is an MRF on $\mathcal{S}$ with respect to N if and only if F is a Gibbs Random Field (GRF) on $\mathcal{S}$ with respect to N, i.e., the probability distribution of a MRF has the form of a Gibbs distribution:

$$P(f) = \frac{e^{-\frac{1}{T}U(f)}}{Z} \tag{4.4}$$

where Z is a normalizing constant called the partition function , T is a constant called the temperature and $U(f)$ is the energy function. The energy function

$$U(f) = \sum_{c \in \mathcal{C}} V_c(f) \tag{4.5}$$

is a sum of cliques potentials, $V_c(f)$, over all possible cliques $\mathcal{C}$. Thus the value of $V_c(f)$ depends on the local configuration on the clique c. The practical value of the theorem is that it provides a simple way of specifying the joint probability $P(f)$. Since $P(f)$ measures the probability of the occurrence of a particular configuration we know that the most probable configurations are those with lowest energies.

It is possible to model a geometric distortion with a random field F defined on the set S of the image pixels. The value assumed by each random variable represents the displacement associated to a particular pixel. Specifically, for each pixel we have two values for the two directions x and y. For this reason each variable F_i is assigned a displacement vector $\mathbf{f}_i = (f_{i_h}, f_{i_v}) \in \mathcal{L} \times \mathcal{L}$.

As we said, a MRF is uniquely determined once the Gibbs distribution and the neighborhood system are defined. In the approach proposed here, for each pixel (x, y) only four neighbors of first order and the corresponding four pair-site cliques, as described by Fig. 4.1, are considered. The potential function we used is a bivariate normal distribution expressed by:

$$V_{((x,y)(\tilde{x},\tilde{y}))}\left(x,y\right)=\frac{1}{2\pi\sigma_x\sigma_y}\exp\left\{-\left[\frac{\left(f_h-f_{\tilde{h}}\right)^2}{2\sigma_x^2}+\frac{\left(f_v-f_{\tilde{v}}\right)^2}{2\sigma_x^2}\right]\right\}\qquad(4.6)$$

where f_h and f_v are the components of the displacement vector $\mathbf{f}_{(x,y)}$ associated to the pixel (x,y), $(\tilde{x},\tilde{y})$ is a point belonging to the 4-neighborhood of (x,y), the pair $((x,y)(\tilde{x},\tilde{y}))$ describes the pair site clique, $f_{\tilde{h}}$ and $f_{\tilde{v}}$ are the components of the displacement vector $\mathbf{f}_{(\tilde{x},\tilde{y})}$ associated to the pixel $(\tilde{x},\tilde{y})$ and σ_x and σ_y are the two components of the standard deviation vector σ (these values are defined based on perceptual constraints).

A typical application of MRF in the image processing field is to recover the original version of an image (or a motion vector field) by relying on a noisy version of the image. By assuming that the original image can be described by means of a MRF, the above problem is formulated as a Maximum a Posteriori estimation problem. Thanks to the Hammersley-Clifford theorem, this corresponds to an energy minimization problem that is usually solved by applying an iterative relaxation algorithm to the noisy version of the image (Besag 1986). The problem we have to face here, however, is slightly different. We simply want to generate a displacement field according to the Gibbs probability distribution defined by Eq. (4.4) and the particular potential function expressed in (4.6).

To do so, the displacement field is initialized by assigning to each pixel (x,y) in the image a displacement vector $\mathbf{f}_{(x,y)}$ generated randomly (and independently on the other pixels) in the the space given by $\mathcal{L}\times\mathcal{L}$ with $\mathcal{L}=\{f\in\mathbb{Z}:-c\leq f\leq c\}$ (the value of c is determined by relying on perceptual considerations). This initial random field is treated as a noisy version of an underlying displacement field obeying the MF model. The MF field is then obtained by applying an iterative smoothing algorithm to the randomly generated field. More specifically, the technique we used visits all the points of the displacement field and updates their values through the Iterated Conditional Mode (ICM) algorithm detailed in (Besag 1986). Specifically, when the ICM algorithm starts, all the points of the displacement field are randomly visited and their displacement vectors updated by trying to minimize the potential

function in Eq. (4.6). Specifically, a local minimum is sought by letting

$$\mathbf{f}_{(x,y)}\text{opt} = \underset{\mathbf{f} \in (\mathcal{L} \times \mathcal{L})}{\arg\min} V_{((x,y)(\tilde{x},\tilde{y}))}(x,y) \qquad (4.7)$$

Note that in the above equation the displacements of the pixels in the neighborhood of (x,y) are fixed, hence resulting in a local minimization of the Gibbs potential. After that each pixel is visited and the corresponding displacement updated, a new iteration starts. The algorithm ends when no new modification is introduced for a whole iteration, which is usually the case after 7-8 iterations.

As for the LPCD model, we considered a multiresolution version of the MF model, where the full resolution version of the displacement field is built by interpolating the displacement field obtained by applying the MF model at a resolution level L. In Fig. 4.2.a and 4.2.b two examples of displacement fields generated with the MF model are shown, using respectively the parameters $L = 6$ $\sigma = (1,1)$ $c = 6$ and $L = 4$ $\sigma = (7,7)$ $c = 18$. Comparing the displacement vectors generated by the MRF model with those shown in Fig. (3.9), we can observe that with the MF model it is possible to obtain larger displacement vectors than with the LPCD (due to the high value of the c parameter), while keeping the distortion invisible thanks to the ability of the iterative conditional mode to generate a very smooth field, as we can see from Fig. 4.2. A pseudo-code description of the MF model is provided by algorithms 1, 2 and 3 in the appendix.

4.3   Perceptual analysis

In order to evaluate the potentiality of the MF model, the perceptual impact of the distortion it generates must be taken into account. From a perceptual point of view, the MF model has a different behavior for different values of L, σ and c, in particular the image quality increases if the distortions are generated at a lower level of resolution but, in the meantime, the number of possible distortions decreases.

After a visual inspection conducted on a set of images, we found, for each level of resolution, the maximum value of the σ components and c that can be used while keeping the distortion invisible. Specifically, we found that, in case
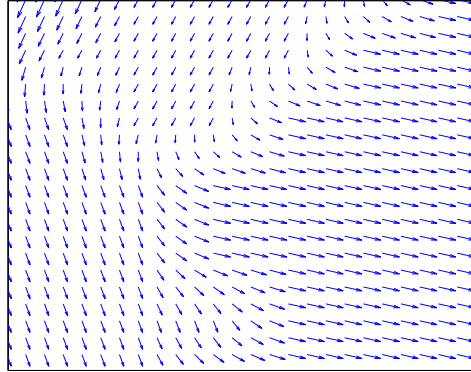
(a)
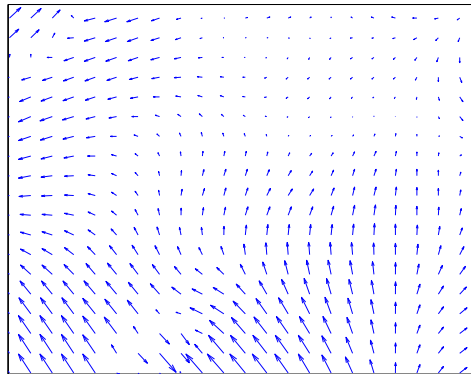
(b)

Figure 4.2: *Examples of displacement fields generated with MF model: (a) MF model with $L = 6$, $\sigma = 1$ and $c = 6$; (b) MF model with $L = 4$, $\sigma = 7$ and $c = 18$.*

of images of size 512×512, the larger perceptually admissible displacements are obtained by using $L = 6$ $\sigma = 1$ $c = 6$, $L = 5$ $\sigma = 3$ $c = 8$ and $L = 4$ $\sigma = 7$ $c = 18$ ($\sigma = \sigma_x = \sigma_y$).

In Fig. 4.3 two examples of images distorted with the MF model applied at different levels of resolution are shown: in the Barbara image the MRF

(a)   (b)

(c)   (d)

Figure 4.3: *Example of two images distorted with the MF model: (a) original Barbara image; (b) distorted Barbara image with $L = 6$; (c) original Lena image; (d) distorted Lena image with $L = 4$*

is applied at a lower level of resolution ($L = 6$), while in the Lena image the distortion is generated at a higher level of resolution ($L = 4$). In both cases by comparing the original image (on the left) with the distorted one (on the right), we can notice a slightly perceptible distortion, that is however not

annoying due to the smoothness constraints of the field (the distortion is not visible if only the distorted image is provided so that the comparison with the original image is not possible).

The advantage brought by MRF theory, as obvious through the perceptual analysis, is that, by letting the displacement field of a generic point (x, y) of the image depend on the displacement fields of the other points of its neighborhood, we can automatically impose that the resulting displacement field is smooth enough to avoid annoying geometrical distortions. This allows to obtain larger displacement fields than the ones obtained with the LPCD model. The main drawback of the MF model is its complex formulation that is not possible to described in a closed form.

Appendix

Algorithm 1.*MF model*

1: Read image to be distorted, read level of resolution L, read standard deviation σ, read c

2: dim $= \dfrac{\text{size(image)}}{2^L}$ {size of the low resolution displacement fields δ_h and δ_v}

3: Initialize matrices δ_h and δ_v with random values in the interval $[-c, c]$

4: diff$_h = \delta_h$

5: diff$_v = \delta_v$

6: **while** diff$_h$ and diff$_v$ are $\neq 0$ **do**

7:    temp$_h = \delta_h$

8:    temp$_v = \delta_v$

9:    row=randperm(dim);

10:    col=randperm(dim);

11:    **for** $k = 1 : \text{dim}$ **do**

12:      **for** $h = 1 : \text{dim}$ **do**

13:        i=col(1,k);

14:        j=row(1,h)

15:        $[sx, sy] = V_{\text{opt}}(i, j, \delta_h, \delta_v, \sigma, \text{dim})$ {Find the optimum displacements sx and sy, i.e. the ones minimizing the potential function}

16:        $\delta_h(i, j) = sx$

17:        $\delta_v(i, j) = sy$

18:      **end for**

19:    **end for**

20:    diff$_h = \delta_h - \text{temp}_h$

21:    diff$_v = \delta_v - \text{temp}_v$

22: **end while**

23: Resize the displacement fields given by δ_h and δ_v to the image size through bicubic interpolation provided by the matlab function *imresize*{to obtain the high resolution displacement fields Δ_h an Δ_v}

24: **for** $i = 1 : \text{size(image)}$ **do**

25:    **for** $j = 1 : \text{size(image)}$ **do**

26:     $Z(i,j) = Y(i + \Delta_h(i,j), j + \Delta_v(i,j))$ {Apply the displacement fields to the image, to obtain the distorted image Z, by means of bicubic interpolation}

27:     **end for**

28: **end for**

Algorithm 2. *Function* $V_{opt}(i, j, \delta_h, \delta_v, \sigma, dim)$

1: Read position of the pixel (i,j), matrices of displacement fields δ_h and δ_v, standard deviation σ

2: $sx_{\text{temp}} = \delta_h(i,j)$

3: $sy_{\text{temp}} = \delta_v(i,j)$

4: $V_{\text{init}} = \text{Gibbs}(i, j, sx, sy, \delta_h, \delta_v)$ {Initial potential}

5: **for** $sx = -i + 1 : \dim - i$ **do**

6:     **for** $sy = -j + 1 : \dim - j$ **do**

7:         $V_{\text{temp}} = \text{Gibbs}(i, j, sx, sy, \delta_h, \delta_v)$

8:         **if** $V_{\text{temp}} < V_{\text{init}}$ **then**

9:             $V_{\text{init}} = V_{\text{temp}}$

10:             $sx_{\text{temp}} = sx$

11:             $sy_{\text{temp}} = sy$

12:         **end if**

13:     **end for**

14: **end for**

15: $sx = sx_{\text{temp}}$

16: $sy = sy_{\text{temp}}$

17: **return** sx and sy

Algorithm 3. *Potential funtion* $Gibbs(i, j, sx, sy, \delta_h, \delta_v)$

1: Read position of the pixel (i,j), displacements sx and sy, matrices of displacement fields δ_h and δ_v

2: $N(i,j) = [(i - 1, j); (i + 1, j); (i, j - 1); (i, j + 1)]$ {$N(i,j)$ is a first order neighborhood system associated with the pixel (i,j)}

3: $V_{((i,j),(\tilde{i},\tilde{j}))} = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left\{-\left[\frac{\left(sx-\delta_h(\tilde{i},\tilde{j})\right)^2}{2\sigma_x^2} + \frac{\left(sy-\delta_v(\tilde{i},\tilde{j})\right)^2}{2\sigma_y^2}\right]\right\}$

4: Potential $= \displaystyle\sum_{(\tilde{i},\tilde{j})\in N(i,j)} V_{((i,j),(\tilde{i},\tilde{j}))}$

5: **return** Potential

Chapter 5

Application to digital watermarking

5.1 Introduction

In the previous chapters, two new classes of local geometric distortions have been introduced and analyzed by means of visual inspection. We now evaluate the effectiveness of the new models as desynchronization attacks against watermarking systems.

In the last years the problem of watermarking in presence of geometric distortions has received an increasing attention from the watermarking community, due to the central role that such distortions play in watermarking theory. As a matter of fact, the application of a geometric distortion to a watermarked image causes a de-synchronization between the watermark embedder and the detector that in most cases prevents the correct extraction of the watermark (Barni and Bartolini 2004). All the more, that in most cases the geometric distortion is less annoying than other kind of distortions like noise addition, blurring or lossy compression. This is especially true for local or spatially varying distortions, for which an exhaustive search of the watermark is unfeasible due to the huge size of the search space.

While global geometric transformations, especially rotation, scaling and translation, have been extensively studied in the watermarking literature given their simplicity and diffusion, local geometric transformations have received little attention by the watermarking community. In practice, only the RBA contained in the Stirmark software has been studied to some extent (see 2.3.1). A first step to solve the problems with local geometric attacks is the characterization of the class of perceptually admissible distortions, defined as the class of geometric distortions whose effect can not be perceived, or is judged acceptable, by a human observer.

In this chapter we evaluate the desynchronization capabilities of the mod-

els introduced in the previous chapters and we compare them with the RBA introduced by the Stirmark software. For the evaluation of the desynchronization capabilities of LPCD and MF models, the perceptual impact of the DAs must be taken into account, since this is the only factor limiting the choice of the attacking strategy. For this purpose the perceptual intrusiveness analysis of the models, carried out in the previous chapters, will be useful.

In the sequel by RBA we will mean only the transformation expressed by Eq. (2.4). This can be obtained by using the Stirmark software and setting to 0 the b, d, i and o parameters (respectively the bending factor, the maximum variation of a pixel value, the maximum distance a corner can move inwards and outwards), and leaving R (the randomisation factor) to the default value of 0.1 (in order to follow perceptual constraints).

5.2   Cardinality evaluation of DA's

A measure of the difficulty of coping with a given type of DA is given by the cardinality of the attack class. In fact, the larger the DA space, the more difficult will be to recover the synchronization between the embedder and the detector, both in terms of complexity and accuracy. As a matter of fact, it is possible to show (Merhav 2005, Barni 2005) that as long as the cardinality of the DAs is subexponential, the exhaustive search of the watermark results in asymptotically optimum watermark detection with no loss of accuracy with regard to false detection probability. By contrast, when the size of the DA is exponential, simply considering all the possible distortions may not be a feasible solution both from the point of view of computational complexity and detection accuracy (Merhav 2005). For this purpose it could be interesting to evaluate the cardinality of the LPCD and MF models.

5.2.1   Cardinality evaluation of LPCD model

In order to evaluate the cardinality of this class of attacks, the perceptual impact of LPCD and C-LPCD must be taken into account.

In chapter 3, both subjective and objective tests were performed to establish the sensitivity of the human visual system to the geometric distortions introduced by the LPCD model as a function of the control parameters N

and L. This way we were able to identify the range of values of the control parameters that do not affect image quality: for each level of resolution, the maximum value of N that can be used while keeping the distortion invisible was found. For instance, in the case of images of size 512×512, the maximum admissible geometric distortions are obtained by using $L = 6$, $N = 5$ for the LPCD model and $L = 5$, $N = 5$ or $L = 6$, $N = 7$ for the C-LPCD model.

We can now use the above considerations to estimate the cardinality of the class of LPCD DAs. For the LPCD model, the number of possible admissible geometric distortions is simply equal, neglecting the boundary effects, to $(N^{\frac{S}{2L} \times \frac{S}{2L}})^2$, where S is the size of the image. Then if we consider a 512×512 image, and we take into account the perceptual analysis in chapter 3, we obtain 2.93×10^{89} different attacked images.

With regard to the C-LPCD model, we need to refer again to the theory of Markov chains. Let us consider the one-dimensional case and the graph of the Markov chain describing the C-LPCD model. It is possible to construct the adjacency matrix A of zeroes and ones, where $A_{i,j} = 1$ if in the graph there is an edge going from node i to node j and zero otherwise. The number of paths of length n that start from node i and end into node j is given by the (i,j) entry of the matrix A^n. The exponential growth rate of the number of paths of length n in the graph is $e^{n \ln \lambda_{max}}$ where λ_{max} is the largest eigenvalue of A. In the C-LPCD case, the practical values of n are not very large, for instance for a 512×512 image, with $L = 5$, we have $n = 16$, then we can easily compute the matrix A^n and derive the exact size of the C-LPCD class of attacks. Specifically, by considering the two-dimensional extension of C-LPCD obtained by applying the one-dimensional C-LPCD DA first by rows and then by columns, we obtain the results reported in table 5.1.

	LPCD $L6 - N5$	C-LPCD $L5 - N5$	C-LPCD $L6 - N7$
cardinality	2.93×10^{89}	1.54×10^{265}	1.54×10^{84}
2^{nH}	2.93×10^{89}	4.76×10^{114}	8.53×10^{30}

Table 5.1: *Cardinality evaluation of the LPCD attacks: in the first row the number of possible distortions is reported, the second row refers to the number of typical sequences.*

With the above approach, we were able to count all the distortions that can be generated with the C-LPCD model. Nevertheless, as explained in the previous subsection, the occurrence of a particular distortion configuration depends on the Markov chain transition matrix and is not constant for all the configurations. Thus, for a more appropriate evaluation of the cardinality of C-LPCD DAs, we need to refer to the entropy rate of the corresponding Markov chain. In this context, the following result from information theory (Cover and Thomas 1991) is useful: let $\{X_i\}$ be a stationary Markov chain with stationary distribution μ and transition matrix P, then the entropy rate is

$$H(\mathcal{X}) = - \sum_{ij} \mu_i p_{ij} \log p_{ij}. \tag{5.1}$$

The knowledge of the entropy rate of the Markov chain and the Asymptotic Equipartition Property (AEP) [1] (Cover and Thomas 1991) help us to find the number of possible distortions that can be generated with a so defined Markov chain, since it asymptotically corresponds to the number of typical sequences[2], i.e., 2^{nH}. After some algebraic manipulations, we find that in the case of C-LPCD with $N = 5$ and $L = 5$, $H(\mathcal{X})$ is approximately equal to 1.4881 bits and the number of typical sequences that is possible to obtain is $2^{256 \cdot 1.4881} \simeq 4.76 \cdot 10^{114}$. In the same way, in the case of C-LPCD with $N = 7$ and $L = 6$, is possible to have $2^{64 \cdot 1.6055} \simeq 8.53 \cdot 10^{30}$ typical sequences. By looking at table 5.1, we can see that, as we expected, the cardinality of C-LPCD evaluated by considering the entropy rate of the Markov chain (second row) is much smaller than the number of possible distortions (first row).

---

[1] AEP is a theorem from information theory that is a simple consequence of the weak law of large numbers. It states that if a set of values $X_1, X_2, ..., X_n$ is drawn independently from a random variable X distributed according to $P(x)$, then the joint probability $P(X_1, ..., X_n)$ satisfies

$$-\frac{1}{n} \log_2 P\left(X_1, X_2, ..., X_n\right) \to H\left(X\right)$$

where $H(X)$ is the entropy of the random variable X.

[2] If a sequence $x_1, ..., x_n$ is drawn from an i.i.d. distribution X defined over a finite alphabet $\mathcal{X}$, then the typical set $A_\epsilon^{(n)}$, that is the set of typical sequences, is defined as those sequences which satisfy:

$$2^{-n(H(X)+\varepsilon)} \leq p\left(x_1, x_2, ..., x_n\right) \leq 2^{-n(H(X)-\varepsilon)}$$

where $H(X)$ is the entropy of the random variable X.

We conclude this section by observing that the size of both the LPCD and the C-LPCD DAs exhibit an exponential growth, with the constrained model resulting in a higher growth rate. For this reason, both classes of attacks are likely to make watermark detection rather difficult.

5.2.2   Cardinality evaluation of MF model

The second class to be analyzed is the MF model introduced in chapter 4. Regarding the cardinality evaluation of this new class of DAs, in principle all the displacement fields are allowed, with the most annoying distortions corresponding to very low probabilities (and thus very large Gibbs potential). In order to evaluate the cardinality of the MF-DA class, then, a first step would be to calculate the entropy rate of the field. However this is a prohibitive task given that no technique is known to calculate the entropy rate of even the simplest MRFs.

5.2.3   Cardinality evaluation of RBA

Regarding the cardinality evaluation of the Stirmark RBA we need to refer to the "rand" factor in Eq. (2.4) which introduces a random displacement in the interval $[1, 2)$ at every pixel position. Even if it is not possible to compute the cardinality in an exact way, by using, for example, a quantization step of 10^{-1}, for a $S \times S$ image the corresponding cardinality will be $(10^{S \times S})^2$. The cardinality of the RBA is very large, however the quantization step used for the computation is a too small value to identify two different distortions.

5.3   De-synchronization properties of the various DAs

In this section, we evaluate the de-synchronization capability of the various classes of attacks. To do so, two very simple watermarking algorithms were implemented and the ability of the various DAs to inhibit watermark detection evaluated. The source image database used for the experiments includes the six standard images shown in Fig. 5.1.

The tested algorithms include:

- Blind additive Spread Spectrum in the frequency domain (BSS-F);

(a)

(b)
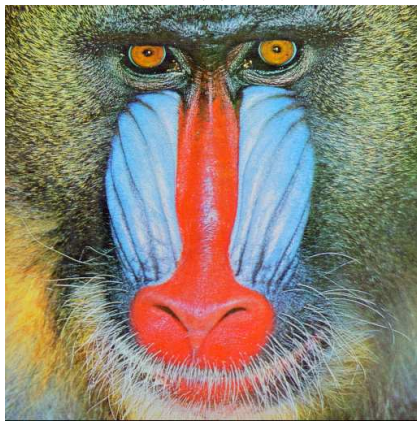
(c)

(d)

(e)

(f)

Figure 5.1: *Source image database: (a) lena, (b) barbara, (c) baboon, (d) peppers, (e) goldhill, (f) boats.*

- Blind additive Spread Spectrum in the wavelet domain (BSS-W).

In both the systems the watermark consists of a sequence of n_b bits

$$X = \{x(1), x(2), ..., x(n_b)\}$$

each value $x(i)$ being a random scalar that is either 0 or 1, with equal probability.

In the BSS-F algorithm the watermark is inserted in the middle frequency coefficients of the full frame DCT domain. The DCT of the original image is computed, the frequency coefficients are reordered into a zig-zag scan and the first $L + M$ coefficients are selected to generate a vector $W = \{t(1), t(2), ..., t(L), t(L + 1), ..., t(L + M)\}$. Then, in order to obtain a trade-off between perceptual invisibility and robustness, the lowest L coefficients are skipped and the watermark X is embedded in the last M coefficients $T = \{t(L + 1), ..., t(L + M)\}$, to obtain a new vector $T' = \{t'(L + 1), ..., t'(L + M)\}$ according to the following rule:

$$\begin{aligned} T' = T + k\text{PN} \quad &\text{if} \quad \text{bit} = 0 \\ T' = T - k\text{PN} \quad &\text{if} \quad \text{bit} = 1 \end{aligned} \tag{5.2}$$

where k is the embedding strength and PN is a uniformly distributed pseudo-random sequence of 1 and -1. Eq. (5.2) refers to the embedding of one bit, the extension to multiple bits consists of applying Eq. (5.2) for each bit considering each time a different subset of T and a different PN sequence (a more detailed description of the watermark embedding is given by the pseudo-codes in appendix).

In watermark detection the DCT is applied to the watermarked (and possibly attacked) image, the DCT coefficients are reordered into a zig-zag scan, and the coefficients from the $(L + 1)th$ to the $(L + M)th$ are selected to generate a vector $T^* = \{t(L + 1), ..., t(L + M)\}$. For each bit the correlation between the corresponding subset of the T^* vector and the PN sequence is evaluated and compared to a threshold (equal to 0) to recover the embedded bit.

The correlation is evaluated in the following way:

$$r\left(A,B\right) = \sum_{i=1}^{n}\left(A(i) - \mu(A)\right)\left(B(i) - \mu(A)\right) \qquad (5.3)$$

(where A and B are two vectors of same size n and μ is the mean operator) and the decision rule states that:

$$\begin{aligned} \text{bit} &= 0 \quad \text{if } r > 0 \\ \text{bit} &= 1 \quad \text{if } r < 0 \end{aligned} \qquad (5.4)$$

In the BSS-W watermarking system the watermark is added to the DWT coefficients of the three largest detail (i.e. LH, HL, HH) subbands of the image. The embedding and decoding functions are implemented in the same way of the previous system but the watermark is inserted in the wavelet coefficients obtained with a one step wavelet decomposition. A more detailed description of the two watermarking systems is given by the pseudo-codes 1,2, 3 and 4 in the appendix.

The six standard images were watermarked with the systems described above with different payloads and then attacked with RBA and the two new classes of attacks. Each image is attacked with a different realization of the field. In table 5.2 the values of the parameters used for the experiments are shown. Fig. 5.2 and 5.3 show the ability of the RBA and that of the two new DAs to inhibit correct decoding. The average of the bit error rate obtained for the six images is plotted versus different values of the payload for both the watermarking systems.

For both the systems, the RBA attack is not able to prevent the correct watermark decoding, in fact the RBA plot is not visible in the figures because the bit error rate is always equal to zero. A more powerful class of DAs is the LPCD that in both the systems gives a quite high bit error rate. The MF-DA always results in a very high bit error rate even if we apply the attack to a lower level of resolution.

For completeness a pool of popular watermarking algorithms found in literature is also considered and the ability of the various DAs to inhibit watermark detection is evaluated by using the same database. While the algorithms tested before are multibit watermarking systems,we evaluate now the desynchronization capability by using 1-bit watermarking systems.

	Parameter	Value
	b	0
	d	0
Stirmark	i	0
	o	0
	R	0.1
MF-DA	c	$\frac{\dim}{2}$
	k	5
DCT system	L	25000
	M	16000
DWT system	k	2

Table 5.2: *Value of the parameters used for the experiments*
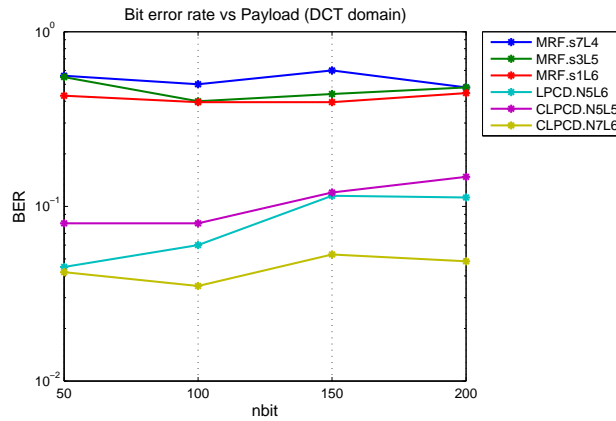
Figure 5.2: *De-synchronization capabilities of the various DAs against the DCT domain system*

The tested algorithms include:

- Non Blind Spread Spectrum in the DCT domain (non-blind SS DCT) (Cox et al. 1997a);

- Blind Spread Spectrum in the frequency domain (SS-DFT)(Barni, Bar-
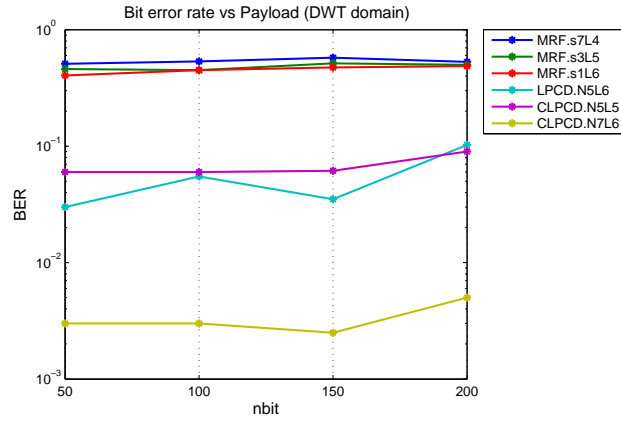
Figure 5.3: *De-synchronization capabilities of the various DAs against the DWT domain system*

    tolini, De Rosa and Piva 2001);

- Blind Spread Spectrum in the wavelet domain (SS-DWT)(Barni, Bartolini and Piva 2001);

- Dirty trellis watermarking (Dirty trellis)(Miller et al. 2004);

- Orthogonal dirty paper coding (Orthogonal dirty paper)(Abrardo and Barni 2005);

Two versions of the Blind Spread Spectrum in the frequency domain algorithm were considered, one with no synchronization measures (Barni, Bartolini, De Rosa and Piva 2001), and one equipped with template-based resynchronization (Piva et al. 1985). All the algorithms are blind techniques except for the system by Cox et al., however in this case the availability of the original image at the detector is not exploited to recover the synchronization.

Tables 5.3, 5.4 and 5.5 show the ability of the Stirmark RBA and of the two new DAs to inhibit the correct watermark detection. The six standard images were watermarked with the systems described above and then attacked with the Random bending and the two new classes of attacks. Each entry of the tables gives the number of images in which the various systems were able to recover the watermark.

	Stirmark
SS-DFT no synchronization	6
SS-DFT with synchronization	3
SS-DWT	6
Dirty trellis	6
Orthogonal dirty paper	6

Table 5.3: *Number of correctly detected watermarks after the Stirmark random bending attack.*

	LPCD N5-L6	CLPCD N7-L6	CLPCD N5-L5
SS-DFT no synchronization	1	2	0
SS-DFT with synchronization	0	1	0
SS-DWT	6	6	6
Dirty trellis	4	5	3
Orthogonal dirty paper	1	2	0

Table 5.4: *Number of correctly detected watermarks after the LCDP attacks.*

By looking at table 5.3, corresponding to the Stirmark random bending attack, we can see how, in most of the cases, the RBA attack is not able to prevent watermark detection, hence calling for a more powerful class of DAs. The bad performance of the SS-DFT algorithm with template-based synchronization can be explained by noting that such an algorithm accounts only for global transformations, hence it is not suited to resist local geometric attacks.

Table 5.4 describes the de-synchronization capabilities of the LPCD DAs:

	MRF L6-$\sigma 1$	MRF L5-$\sigma 3$	MRF L4-$\sigma 7$
SS-DFT no synchronization	6	0	0
SS-DFT with synchronization	0	0	0
SS-DWT	2	0	0
Dirty trellis	0	0	0
Orthogonal dirty paper	0	0	0

Table 5.5: *Number of correctly detected watermarks after the MF-DA attacks.*

except for the SS in the wavelet domain system, most of the algorithms are not able to detect all the watermarks. The problems are accentuated with MF-DA, as it can be seen from 5.5, in fact, such an attack almost always result in the loss of the watermark.

	non-blind SS DCT
Stirmark	0.9943
LPCD N5-L6	0.9803
CLPCD N7-L6	0.9906
CLPCD N5-L5	0.9734
MRF L6-$\sigma 1$	0.8274
MRF L5-$\sigma 3$	0.6639
MRF L4-$\sigma 7$	0.6304

Table 5.6: *De-synchronization capabilities of the various DAs against non-blind SS in the frequency domain.*

Finally, table 5.6 shows the output of the correlation coefficient detector adopted by the non blind system by Cox et al. (Cox et al. 1997a), for all the de-synchronization attacks. Once again, we can observe that the MF-DA exhibits the best desynchronization capabilities.

5.4 Concluding remarks

In this chapter we used the geometric models introduced in the previous chapters to extend the class of local geometric attacks so to allow for more powerful attacks with respect to classical RBA. The effectiveness of the new classes of DAs is evaluated from different perspectives including perceptual intrusiveness and de-synchronization efficacy. The experimental results showed that the two new classes of attacks are more powerful than the local geometric attacks proposed so far. This can be seen as an initial effort towards the characterization of the whole class of perceptually admissible DAs, a necessary step for the theoretical analysis of the ultimate performance reachable in the presence of watermark de-synchronization and for the development of a new class of watermarking systems that can effectively cope with them.

This chapter ends the first part of the thesis whose purpose is to characterize local geometric transformations for still images. After the definition of geometric transformations and a classification of global geometric distortions and local geometric distortions, the aim of the following chapters has been to introduce two new models to describe, from a mathematical point of view, local geometric distortions and evaluate, through subjective tests, the perceptual quality impact of the defined transformations on still images. In this chapter, the proposed models have been used as desynchronization attacks in watermarking systems and compared with the classical RBA.

The proposed analysis of local geometric distortions can be an important step in many applications. The problem of the characterization of geometrically distorted images and the perceptual impact of geometric transformations in images, in fact, is a challenge in different research fields like registration of biomedical imaging or digital photography, and the interest in this topic is more and more increasing with the advent of 3DTV and 3D displays.

It is clear that the availability of an objective quality metric capable of dealing with geometric distortions would be of invaluable help in this sense. This will be the focus of the second part of the thesis.

Appendix

Algorithm 1.*DCT domain watermarking: embedding*

1: Read image to be watermarked, length of the watermark n_b, energy of the
    watermark k, seed *key*, L, M

2: Generate a random n_b long message

3: Perform full frame DCT

4: Reorder the DCT coefficients into a zig-zag scan

5: Select the coefficients: $T_L^{L+M} = \{t(L), t(L+1), ..., t(L+M)\}$ {middle
    frequency coefficients to be watermarked}

6: **for** bit=1:n_b **do**

7:    Generate an antipodal PN sequence of length $lbit = M/n_b$

8:    $a = (\text{bit} - 1) * \text{lbit} + 1$ and $b = (\text{bit} - 1) * \text{lbit} + \text{lbit}$

9:    **if** bit=0 **then**

10:      $\hat{T}_a^b = T_a^b + k\text{PN}$

11:    **else**

12:      $\hat{T}_a^b = T_a^b - k\text{PN}$

13:    **end if**

14: **end for**

15: Reinsert the vector $\hat{T}$ in the zig-zag scan

16: Perform inverse scan

17: Perform inverse full frame DCT

18: Save watermarked image and message

Algorithm 2.*DCT domain watermarking: decoding*

1: Read watermarked image, seed key, length of the watermark n_b and load
    inserted message {needed to evaluate bit error rate}

2: Perform full frame DCT transform

3: Reorder the DCT coefficients into a zig-zag scan

4: Select the coefficients: $T^*{}_L^{L+M} = \{t(L), t(L+1), ..., t(L+M)\}$ {middle
    frequency watermarked coefficients}

5: **for** bit=1:n_b **do**

6:    Generate an antipodal PN sequence of length $lbit = M/n_b$

7:     Compute the correlation as expressed in Eq. (5.3) between PN and T^{*b}_a where $a = (\text{bit} - 1) * \text{lbit} + 1$ and $b = (\text{bit} - 1) * \text{lbit} + \text{lbit}$

8: **end for**

9: **for** bit=1:n_b **do**

10:     **if** correlation(bit) > 0 **then**

11:         extracted_message(bit)=0

12:     **else**

13:         extracted_message(bit)=1

14:     **end if**

15: **end for**

16: **return** Bit Error Rate

Algorithm 3. *DWT domain watermarking: embedding*

1: Read image to be watermarked, length of the watermark n_b, energy of the watermark k, seed key

2: Generate a random n_b long message

3: Perform a one step wavelet decomposition using Haar filter

4: Reorder the LH, HL and HH components into a vector T

5: **for** bit=1:n_b **do**

6:     Generate an antipodal PN sequence of length $lbit = \text{size}(T)/n_b$

7:     $a = (\text{bit} - 1) * \text{lbit} + 1$ and $b = (\text{bit} - 1) * \text{lbit} + \text{lbit}$

8:     **if** bit=0 **then**

9:         $\hat{T}^b_a = T^b_a + k\text{PN}$

10:     **else**

11:         $\hat{T}^b_a = T^b_a - k\text{PN}$

12:     **end if**

13: **end for**

14: Perform a one step inverse wavelet decomposition using Haar filter

15: Save watermarked image and message

Algorithm 4. *DWT domain watermarking: decoding*

1: Read watermarked image, seed *key*, length of the watermark n_b and load
   inserted message {needed to evaluate bit error rate}
2: Perform a one step wavelet decomposition using Haar filter
3: Reorder the LH, HL and HH components into a vector T^*
4: **for** bit=1:n_b **do**
5:     Generate an antipodal PN sequence of length $lbit = \text{size}(T)/n_b$
6:     Compute the correlation as expressed in Eq. (5.3) between PN and T^{*b}_a
       where $a = (\text{bit} - 1) * \text{lbit} + 1$ and $b = (\text{bit} - 1) * \text{lbit} + \text{lbit}$
7: **end for**
8: **for** bit=1:n_b **do**
9:     **if** correlation(bit)> 0 **then**
10:         extracted_message(bit)=0
11:     **else**
12:         extracted_message(bit)=1
13:     **end if**
14: **end for**
15: **return** Bit Error Rate

Part II

Quality assessment of geometric distortions in images

Chapter 6

Image quality assessment

6.1 Introduction

In the first part of this thesis, the problem of the characterization of geometrically distorted images has been described by introducing two new models to describe local geometric distortions from a mathematical point of view. The perceptual impact of geometric transformations in images have also been analyzed.

In addition to image watermarking, the topic of geometric distortions is a challenge in different research fields and whose interest is increasing with the advent of 3DTV and 3D displays. Todays 3D displays, in fact, enable multiple-viewpoint autostereoscopic experiences with significantly improved visual quality over previous generation displays and the resizing of stereoscopic images for display adaptation is one of the main challenges. The analysis on the issue of 3D image resizing from geometrical perspectives is a focal point in this contest.

In the same way, the conversion of existing 2D videos to 3D videos is necessary for multimedia applications and is obtained through the rendering of stereoscopic images that are free of geometric distortions. Geometric distortion calibration is also a key problem in multi-projector display, where immersive multi-projector virtual environment based on PC cluster has become a research hotspot in virtual reality applications.

It is clear, from these considerations, how the availability of an objective quality metric for the assessment of geometric distortions would be of invaluable help. The development of a metric for the evaluation of geometric distortion in images is the goal of the second part of the thesis. For this purpose it is necessary to review the procedures for the design and then the validation of a quality assessment method.

Quality assessment is a central issue in the design, implementation, and performance testing of digital systems. Generally digital signal processing systems deal with visual information that are meant for human consumption. In the same way, digital images may go through different stages of processing before being presented to a human observer, and each stage of processing may introduce distortions that could reduce the quality of the final display. The goal of objective quality assessment is to design metrics that can automatically predict quality, i.e. the way humans perceive such distortions.

The benchmark for any kind of visual quality assessment are subjective experiments, since a human observer is the end user in many multimedia applications. Standard procedures for subjective experiments have been established in the last years as valuable research tools in the image and video processing field for a better understanding of how humans judge quality and the perceived distortions.

In many image processing applications, in fact, the limitations and the characteristics of the human visual system can be exploited to improve the performance from a visual quality point of view. By taking into account the properties and limitations of HVS, for example, images can be more efficiently compressed, colors more accurately reproduced, prints better rendered, image compression schemes can reduce the visibility of introduced artifacts, and watermarking schemes can hide more robustly information in images. In the same way, by exploiting HVS characteristics, it is possible to develop quality assessment tools that are able to predict in an automatic way subjective ratings. An image quality measure that is based on human vision model, in fact, seems to be most appropriate for predicting user acceptance and for system optimization. To achieve these goals it is necessary to build a computational model of the HVS.

In the following section a short overview of the basic characteristics of the HVS is provided. Section 6.3 describes the design and implementation of subjective tests to evaluate the performance of a perception-based image processing tool. In Section 6.4 a brief review of popular image quality metrics, together with a description of quality assessment methods for geometric distortions, is provided. Finally, in section 6.5, the standard procedures for the evaluation of the performance of an objective quality metric is described.

6.2 Human Visual System

The Human Visual System has been studied for several years. One of the majorn concern for this research field is that most visual properties of the HVS are not intuitive. Even when they have been characterized by psychophysical experiments, physiological evidence is the only way to understand the phenomenon completely. This section gives a short introduction to the main physiological concepts of the HVS that could also serve for its modeling. For a more detailed review of vision physiology, the reader is referred to (Wandell 1996).

The physiology of human vision includes the eyes and the retina, where vision is initiated, as well as the visual pathways and the visual cortex, where high-level perception takes place. The eyes represent the first stage of the HVS, they are the interface between the input signal (the visual information) and the processing unit (the brain). They can be understood as a complicated camera continually in motion, allowing accommodation to different light levels and to objects at various distances. Fig. 6.1 shows a section of the eye.
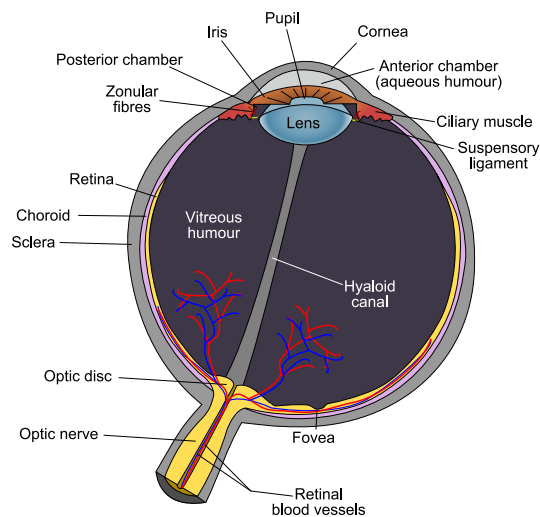

Figure 6.1: *Schematic diagram of the vertebrate eye. Credit: Wikipedia article "Eye".*

At the back of the eyes lies the retina, a dense layer of interconnected

neurons that sample and process the visual information. The retina encodes the visual information before transmitting it along the optical nerve, which is a channel with limited capacity. The ratio ($\approx 100:1$) between the number of receptors in the retina and the number of fibers in the optical nerve implies already at this stage a compression of the visual information.

The primary function of the retina is the sampling of the optical signal by photoreceptors. There are two kinds of photoreceptors, rods and cones. Rods are sensitive to low levels of luminosity. They are long and slender and contain the same pigment, the photochemical rhodopsin. There are about 100 million rods per eye, uniformly distributed in the fovea, which has a diameter of about 0.5 mm. The adaptation of our vision system to the dark is obtained by slowly increasing the amount of rhodopsins. Since rods are almost non-existent in the center of the visual field, their contributions are generally neglected in image processing applications.

Cones do not respond in low light conditions and are responsible for fine details and color vision. They can be classified as L-, M- and S-cones according to their sensitivity to long (red cones), medium (green cones) and short (blue cones) wavelengths, respectively. The cones do not provide detailed spectral information, but a weighted summation over the different sensitivity spectra. This means that three values should be sufficient to reproduce human color distinction capabilities, which leads to the description of color by tri-stimulus values. Color can be formalized as a three-dimensional vector space, which makes computations with color values possible. This idea has been used by the CIE (Commission International de l'Eclairage)[1] to define several colorimetric functions like RGB and XYZ. The number of cones in the retina is much smaller than the number of rods and it is estimated around 5 millions, most concentrated in the fovea, there are more red and green cones than blue cones and this fact can explain the lower sensitivity of HVS to the blue component.

Human color perception is not directly related to the cone responses, but rather to their differences.These are represented by an achromatic channel and two opponent-color channels, which code red-green and blue-yellow color differences. In image processing this coding is exploited in several color spaces such as YCbCr, where Y is the luminance channel and Cb and Cr are the

---

[1]http://members.eunet.at/cie/

color-difference channels.

The HVS and especially the retina are able to adapt their sensitivity to the input signal. This allows to handle a wide range of light intensities with a small number of quantization levels. This mechanisms for adaptation include the iris, which controls the size of the pupillary aperture and thus the retinal illumination, the photoreceptors, and the ganglion cells. These adaptations greatly influence the perception of color and luminosity contrast, for that reason, the CIE has formalized a color space called L.a.b., a non-linear opponent colors space adapted to the light source.

The neurons in the retina realize a spatio-temporal filtering of the visual signal through their synaptic interactions. This filtering is quite complex and not yet completely understood, however, its influence on perception is very high and, for this reason, improvements to the L.a.b. color space were proposed to model the filtering properties of the retina.

Two main pathways have been identified at the output of the retina, the magnocellular pathway that carries blurred spatial information of luminance at high speeds, which is important for reflex actions, and the parvocellular pathway that carries spatial detail and color information, which is important for conscious perception.

In the visual cortex, many cells are tuned to specific stimulus properties such as orientation, form, color, spatio-temporal frequency, stereo information, or motion, and decompose the visual information accordingly. Several areas can be distinguished in the visual cortex, among them area V1 (also known as primary visual cortex), which receives the input from the retina, area V2, which processes color, form and stereo, area V4, which also processes color, and area MT, which handles movement and stereo vision. Although these cortical areas have been identified according to their functional role, this role is not explicit yet.

Despite our current knowledge of the HVS, its complexity makes it impossible to construct a complete physiological model. Consequently HVS models used in image processing are usually behavioral and are based on psychophysical studies.

6.3   Subjective measurements methods

In order to be able to design reliable visual quality metrics, it is necessary to understand what "quality" means to the viewer. The viewer's enjoyment when watching video depends on many factors. One of the most important is of course content and material. Research has shown that perceived image and video quality depends on viewing distance, display size, resolution, brightness, contrast, sharpness, colorfulness, naturalness and other factors (Ahumada Jr 1993). It is also important to note that there is often a difference between *fidelity* (the accurate reproduction of the original on the display) and perceived *quality*. Sharp images with high contrast are usually more appealing to the average viewer. Likewise, subjects prefer slightly more colorful and saturated images despite realizing that they look somewhat unnatural (de Ridder et al. 1995).

Subjective tests provide the foundations for building vision models and they are the only true benchmark for evaluating the performance of perception-based image processing tools. For this reason, the measurement of the perceived quality of images requires the use of subjective scaling methods. The condition for such measurements to be meaningful is that there exists a relation between the physical characteristics of the image or video sequence presented to the subjects in a test, and the magnitude and nature of the sensation caused by the stimulus.

Different applications require different testing procedures. Subjective assessment of visual quality has been formalized in ITU-T Recommendation P.910 (Int 1996), which suggests standard viewing conditions, criteria for the selection of observers and test material, assessment procedures, as well as data analysis methods.

6.3.1   Classification of the methods

Two kinds of decision tasks can be distinguished, namely *adjustment* and *judgment* (Pelli and Farell 1995). In the former, the observer is given a classification and provides a stimulus, while in the latter, the observer is given a stimulus and provides a classification. Adjustment tasks include setting the amplitude of a stimulus, canceling a distortion, or matching a stimulus to a

given one. Judgment tasks on the other hand include yes/no decisions, forced choices between two alternatives, and magnitude estimation on a rating scale.

A number of experimental methods have been validated in the last years for different purposes. In the following, a description of the three most commonly used methods is given, as described in the recommendation. The final choice of one of these methods for a particular application depends on several factors, such as the context, the purpose and when in the development process the test is to be performed.

Absolute Category Rating (ACR). The Absolute Category Rating test (also called Single Stimulus method) is a category judgement where the images are presented one at a time and are rated independently on a category scale. After each presentation the subjects are asked to evaluate the quality of the image shown by using the following five-level category scale:

5) Excellent

4) Good

3) Fair

2) Poor

1) Bad

If higher discriminative power is required, a nine-level scale may also be used.

The time pattern for the stimulus presentation is illustrated in Fig. 6.2. For all the subjective tests the voting time should be less than or equal to 10 seconds, depending upon the voting mechanism used. The presentation time may be reduced or increased according to the content of the test material.

Degradation Category Rating (DCR). The Degradation Category Rating (also called the Double Stimulus Impairment Scale method) implies that the images are presented in pairs. The subjects are asked to rate the impairment of the second stimulus in relation to the reference one.

The following five-level scale for rating the impairment should be used:

5) Imperceptible

4) Perceptible but not annoying

3) Slightly annoying
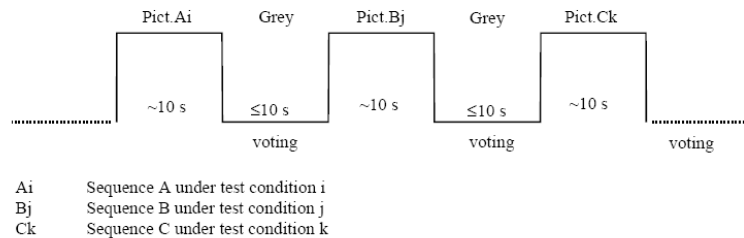
Figure 6.2: *Stimulus presentation in the ACR method.*

2) Annoying

1) Very annoying

The time pattern for the stimulus presentation is illustrated by Fig. 6.3.
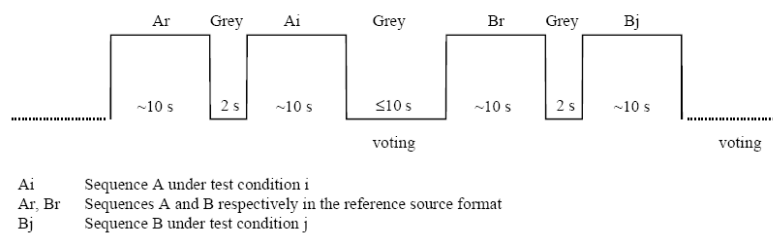
Figure 6.3: *Stimulus presentation in the DCR method.*

Pair Comparison method (PC) (or Two Alternative Forced Choice test (2AFC)). The method of Pair Comparisons implies that the distorted images are presented in pairs. All the images under tests are generally combined in all the possible $n(n-1)$ combinations and after each pair a judgment is made on which element in the pair is preferred for that particular scenario.

The time pattern for the stimulus presentation is illustrated by Fig. 6.4.

An important issue in choosing a test method is the fundamental difference between methods that use explicit references (e.g. DCR) and methods that

Ai, Aj   Sequence A under i[th] and j[th] test condition respectively
Bk, B1   Sequnce B under k[th] and 1th test condition respectively
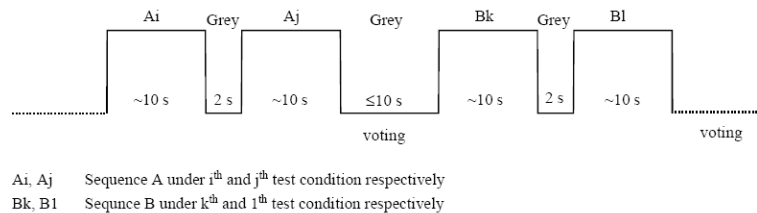
Figure 6.4: *Stimulus presentation in the PC method.*

do not use any explicit reference (e.g. ACR and PC). This second class of method does not test transparency or fidelity, while the DCR method should be used when testing the fidelity of transmission with respect to the source signal. This is an important factor, for example, in the evaluation of high quality systems.

ACR is easy and fast to implement and the presentation of the stimuli is similar to that of the common use of the systems. Thus, ACR is well-suited for qualification tests.

The principal merit of the PC method is its high discriminatory power, which is of particular value when several of the test items are nearly equal in quality. However, when a large number of items are to be evaluated in the same test, the procedure based on the PC method tends to be lengthy. In such a case an ACR or DCR test may be carried out first with a limited number of observers, followed by a PC test on those items which have been selected in the first test.

6.3.2   Experiment requirements

The results of quality assessments often depend on several factors as the actual image quality, the total quality range of the test conditions, the experience and expectations of the assessors, etc. In order to control some of these effects, a number of test conditions can be added and used as references. A description of reference conditions and procedures to produce them is given in the following.

Equipment and display configuration.   Typical viewing conditions to be used in video and image quality assessment are described in Table 6.1, where H is the picture height.

Parameter	Settings
Viewing distance	1-8 H
Peak luminance of the screen	100-200 cd/m
Ratio of luminance of inactive screen to peak luminance	≤ 0.05
Ratio of the luminance of the screen, when displaying only black level in a dark room, to that corresponding to peak white	≤ 0.1
Ratio of luminance of background behind picture monitor to peak luminance of picture	≤ 0.2
Background room illumination	≤ 20 lux

Table 6.1: *Viewing conditions*

The values shown in the table are typically used default values. Some of these values, as for example the viewing distance, can be modified depending on the type of application and the goal of the experiments.

Concerning the displaying format, it is preferable to use the whole screen for displaying the images. If they are displayed on a window of the screen, the colour of the background in the screen should be grey level corresponding to Y=U=V=128.

All the experiments described in this thesis were conducted following the conditions described in table 6.1. The tests were performed in a dark room by using the VP800 video card of the Cambridge Research Systems together with a high resolution 21-inch digital monitor Mitsubishi DiamondPro with the external adaptor ViSaGe 71.02.00D2.14 [2]. To have a correct color representation a luminosity calibration was previously carried out through a video-camera ColorCAL.14.

---

[2]http://www.crsltd.com/catalog/visage/index.html

Human subjects, training and testing. The number of subjects in a viewing test should range from 4 to 40. In general, at least 15 observers should participate in the experiment. They should have a good vision, and be naives with respect to image quality assessment methods and image impairments.

Before starting the experiment, a scenario of the intended application of the system under test should be given to the subjects. In addition, a description of the type of assessment, the opinion scale and the presentation of the stimuli is given in written form.

The experiment should follow a five-stage procedure: (1) oral instructions, (2) training, (3) practice trials, (4) experimental trials, (5) interview. In the first stage, the subjects are verbally given instructions and made familiar with the task and the graphic interface. In the training the original models and the distorted models are shown to establish the range for the impairment scale. The practice trials stage are used to familiarize subjects with the experimentation. The range and type of impairments are presented in preliminary trials, which contain images and distortions other than those used in the real tests. In the experimental stage, the subject has to give a score to indicate how much the distortion is evident. Finally, in the interview stage, the test subjects are asked for a qualitative description of the perceived distortions.

6.3.3 Processing Subjective Data

A subjective experiment produces a large number of data. In order to obtain effective information from these numbers, some statistical processing is needed. Generally, the following methods are used (depending on the applications):

1. The **probability of detection** (PD) of an impairment is estimated by counting the number of subjects who detect the impairment and dividing it by the total number of subjects.

2. The **Mean Opinion Score** (MOS) is calculated by averaging the scores over all observers for each test sequence:

$$\text{MOS} = \frac{1}{L}\sum_{l=0}^{L} Score(l)$$

where *Score* is the score reported by the l-th subject and L is the total number of subjects.

3. The **screening** is used to eliminate subject whose rating is outside the upper and lower boundaries computed for the test model. It reduces the variability of the gathered data . Regarding this point, a variety of opinions can be found concerning what is an unacceptable level of skewness (the symmetry of a distribution) and kurtosis (the clustering of scores toward the center of a distribution) for a particular variable (Sheikh et al. 2006). A typical approach in image quality assessment, is to find the outliers by evaluating the kurtosis coefficient defined in the following way:

$$k = \frac{E\left(x - \mu\right)^4}{\sigma^4}$$

where μ is the mean of the distribution of the scores x provided by a subject, σ is the standard deviation and $E()$ is the expected value. Subjects with a kurtosis coefficient greater than two times the standard deviation of the distribution, are considered outlier-prone and are discarded from the distribution.

6.4   Visual quality assessment

The best way of assessing the quality of an image is through subjective evaluation because in most cases human eyes are the ultimate receivers. The Mean Opinion Score has been used for many years; however, the MOS is tedious, quite expensive in terms of time (preparation and running) and human resources. Furthermore the subjective results depend on several external factors such as the observers background, motivation, etc.

The goal of objective image quality assessment, as already explained, is to design a quality measure based on the human vision model that can predict the perceived image quality automatically. The objective measure should give a numerical value quantifying the dissatisfaction of a typical human viewer when he/she observes the reproduced image in place of the original.

Objective quality metrics can be classified according to the availability of the original image.

- *Full reference metrics* (FR). The evaluation system has access to the original media. Some examples of full reference metrics include the works in (Daly 1993), (Lubin 1993), (Watson et al. 2001), (Wolf et al. 1991), and (Winkler 2005). The Structural SIMilarity (SSIM) index is a widely used method for measuring the similarity between two images (Wang, Bovik, Sheikh and Simoncelli 2004). A complete survey of the available FR quality metrics is presented in (Winkler 1999).

- *Reduced reference metrics* (RR). The evaluation system has access to a small amount of side information regarding the original media. In general, certain features or physical measures are extracted from the reference and transmitted to the receiver as side information to help evaluate the quality of the video. Metrics in this class may be less accurate than the FR metrics, but they are also less complex, and make real-time implementations more affordable. Some examples include the works in (Webster et al. 1913), (Bretillon et al. 1999) and (Carnec et al. 2008).

- *No-Reference metrics* (NR). The evaluation system has no reference to any side information regarding the original media. This kind of metrics is the most promising in the context of video broadcast, since the original images or video are in practice not accessible to end users. Requiring the reference video or even limited information about it becomes a serious impediment in many real-time applications. Designing effective NR metrics is a big challenge. Most of the proposed NR metrics estimate annoyance by detecting and estimating the strength of commonly found artifact signals. For example, the metrics in (Wang et al. 2000) and (Wu and Yuen 1997) estimate quality based on blockiness measurements, while the metric in (Caviedes and Jung n.d.) takes into account measurements of 5 types of artifacts.

The problem of assessing the perceived quality of digital images and video is far to be solved even if many methods have been proposed in the last

decades. The Video Quality Expert Group[3] (VQEG) has conducted two phases of tests on objective models (Corriveau and Webster 1999), however, no objective quality metric is widely accepted. One of the main difficulties in designing an efficient image quality assessment method is that the perceived quality is not necessarily equivalent to fidelity and most quality metrics are actually fidelity metrics based on the comparison of the distorted image with a reference image and neglect these phenomena.

Focusing on gray level images, the simplest and most widely used full-reference quality metric is the Mean Squared Error (MSE), computed by averaging the squared intensity differences of the distorted and reference image pixels, along with the related quantity of Peak Signal-to-Noise Ratio (PSNR), as follows:

$$MSE = \frac{1}{mn} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} \|I(x,y) - Z(x,y)\|^2 \qquad (6.1)$$

$$PSNR = 20 \log \left(\frac{\max(I)}{\sqrt{MSE}} \right) \qquad (6.2)$$

where $\max(I)$ is the maximum possible pixel value of the image (when the pixels are represented using 8 bits per sample, this is 255), Z is the image to be evaluated and m and n are the width and the height of the two images.

Technically, MSE measures the difference between two images, whereas PSNR measures image fidelity, i.e. how closely an image resembles a reference image. The popularity of these two metrics is due to the fact that computing MSE and PSNR is very easy and fast. Because they are based on a pixel-by-pixel comparison of images, however, they only have a limited, approximate relationship with the distortion or quality perceived by human observers. In certain situations the subjective image quality can be improved by adding noise and thereby reducing the PSNR. Dithering of color images with reduced color depth, which adds noise to the image to remove the perceived banding caused by the color quantization, is a common example of this. Furthermore, the visibility of distortions depends to a great extent on the image content, a property known as masking (as we will see in 6.4.1). Distortions are often

---

[3]http://www.its.bldrdoc.gov/vqeg/

much more disturbing in relatively smooth areas of an image than in textured regions with a lot of activity, an effect not taken into account by pixel-based metrics. Therefore the perceived quality of images with the same PSNR can actually be very different (as we can see by looking at image in Fig. 6.5).
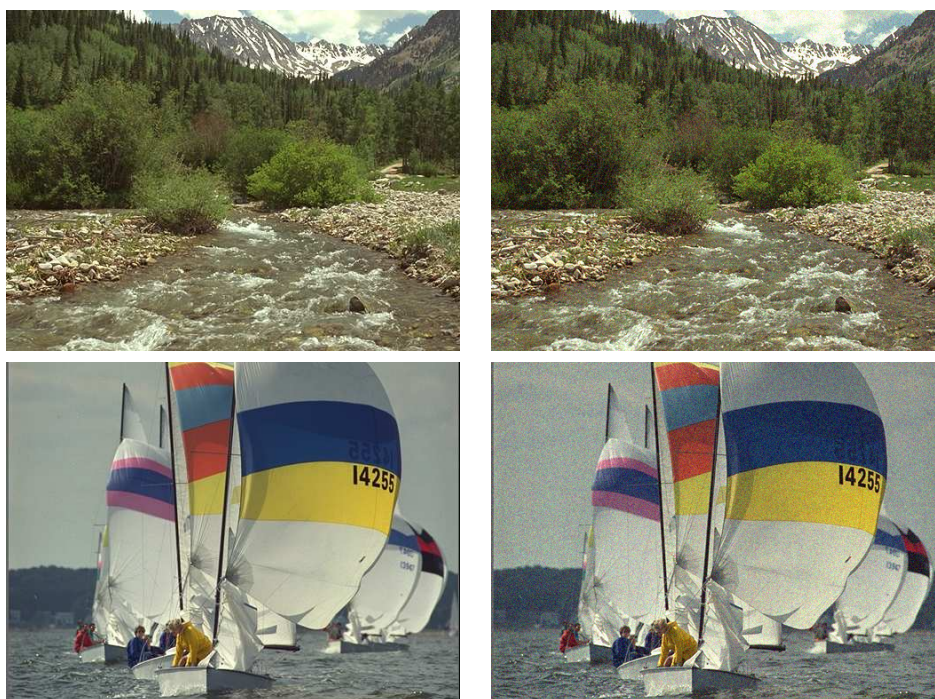
Figure 6.5: *Two images with identical PSNR (the original images are on the left). The noise is much less visible in the image on the top due to strong masking, an effect PSNR does not take into account. Credit: Tamper Image Database 2008 (Ponomarenko et al. 2008)*

A number of additional pixel-based metrics have been proposed. It was found that although some of these metrics can predict subjective ratings quite successfully for a given compression technique or a specific type of distortion, they are not reliable for evaluations across different kinds of distortions.

In the last three decades, a great effort has been paid to develop quality assessment methods that take into account the characteristics of the human visual system. It is a matter of fact that the visual assessment task involves

high complex psychophysical mechanisms, however, the HVS is too complex to be completely understood with current psychophysical knowledge, thus current HVS models rely on some simplified assumptions and they may not provide fully reliable results. In the following a brief review of the state of art approaches of the quality assessment of image distortions is provided.

6.4.1   Measures based on Error sensitivity

HVS-models account for a number of psychophysical effects that are typically implemented in a sequential process as shown in Fig. 6.6, called error sensitivity approach. The underlying principle of this approach is that perceptual quality is best estimated by quantifying the visibility of errors. This is essentially accomplished by simulating the functional properties of early stages of the HVS, as characterized by both psychophysical and physiological experiments.
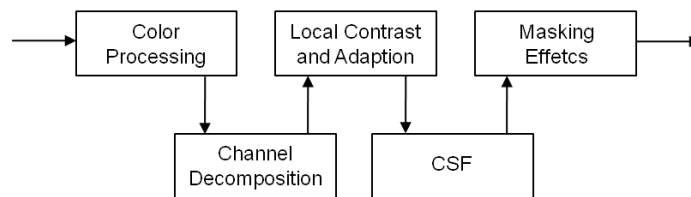
Figure 6.6: *Block-diagram of a typical HVS-model*

The first stage in the processing chain of HVS-models concerns the transformation into an adequate perceptual color space, usually based on opponent colors. After this step the image is represented by one achromatic and two chromatic channels carrying color difference information.

Regarding the multi-channel decomposition, it is widely accepted that the HVS bases its perception on multiple channels that are tuned to different ranges of spatial frequencies and orientations. This behavior is well matched by a multi-resolution filter bank or a wavelet decomposition.

The response of the HVS depends much less on the absolute luminance

than on the relation of its local variations to the surrounding background, a property known as Weber-Fechner law (Shen 2003). Contrast is a measure of this relative variation, which is commonly used in vision models. It is very difficult to model human contrast perception in complex images, because it varies with the local image content. Furthermore, the adaptation to a specific luminance level or color can influence the perceived contrast.

One of the most important issues in HVS-modeling concerns the decreasing sensitivity to higher spatial frequencies. This phenomenon is parameterized by the contrast sensitivity function (CSF).

The CSF tells us how sensitive we are to the various frequencies of visual stimuli. A contrast sensitivity assessment procedure consists of presenting the observer with a sine-wave grating target of a given spatial frequency (i.e., the number of sinusoidal luminance cycles per degree of visual angle). The contrast of the target grating is then varied while the observer's contrast detection threshold is determined. Typically, contrast thresholds of this sort are collected using vertically oriented sine-wave gratings varying in spatial frequency from 0.5 (very wide) to 32 (very narrow) cycles per degree of visual angle.

The contrast sensitivity function proposed by Manos and Sakrison(Mannos and Sakrison 1974) is:

$$A\left(f\right) = 2.6\left(0.0192 + 0.114f\right)\exp^{-(0.114f)^{1.1}} \tag{6.3}$$

where f is the spatial frequency of the visual stimuli given in cycles/degree. The function has a peak of value 1 aproximately at $f = 8.0$ cycles/degree, and is meaningless for frequencies above 60 cycles/degree. Fig. 6.7 shows the contrast sensitivity function $A(f)$.

A correct modeling of the CSF is especially difficult for color images. Typically, separability between color and pattern sensitivity is assumed, so that a separate CSF for each channel of the color space needs to be determined and implemented. The human contrast sensitivity also depends on the temporal frequency of the stimuli. Similar to the spatial CSF, the temporal CSF has a low-pass or slightly band-pass shape. The interaction between spatial and temporal frequencies can be described by spatio-temporal contrast sensitivity functions, which are commonly used in vision models for video (Robson 1966).
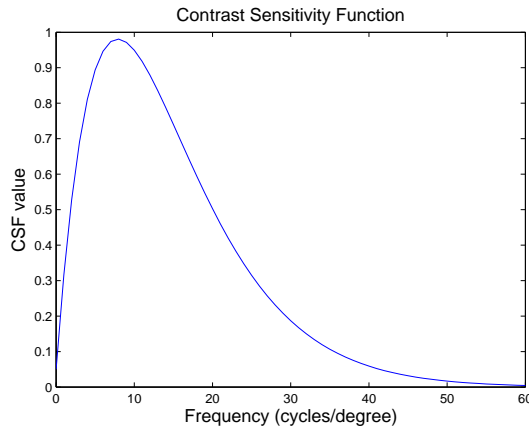
Figure 6.7: *Contrast sensitivity function*

The last step in Fig. 6.6 regards the Masking effect. Masking occurs when a stimulus that is visible by itself cannot be detected due to the presence of another. Masking explains why similar distortions are disturbing in certain regions of an image while they are hardly noticeable elsewhere. The popular terms contrast masking, edge masking, and texture masking are widely used to describe masking due to strong local contrast, edges, and local activity, respectively.

Most of the image quality metrics proposed so far are based on a error sensitivity approach exploiting all or some of the processes described in Fig. 6.6. This bottom-up approach to the problem has found nearly universal acceptance, however it has many limitations. The HVS is a complex and highly nonlinear system, but most models of early vision are based on linear or quasi-linear operators that have been characterized using restricted and simplistic stimuli. Thus, error-sensitivity approaches must rely on a number of strong assumptions and generalizations. A brief summary of these assumptions is provided here.

- *Definition of Quality.* The most fundamental problem with the traditional approach is the definition of image quality. In particular, it is not clear that error visibility should be equated with loss of quality, as some

distortions may be clearly visible but not so objectionable. The study in (Silverstein and Farrell 1996) also suggested that the correlation between image fidelity and image quality is moderate.

- *Suprathreshold Problem.* The psychophysical experiments that underlies many error sensitivity models are specifically designed to estimate the threshold at which a stimulus is visible. These measured threshold values are then used to define visual error sensitivity measures, such as the CSF. However, very few psychophysical studies indicate whether such near-threshold models can be generalized to characterize perceptual distortions significantly larger than threshold levels, as is the case in a majority of image processing situations.

- *Complexity of natural images.* Most psychophysical experiments are conducted using relatively simple patterns, such as spots, bars, or sinusoidal gratings. For example, the CSF is typically obtained from threshold experiments using global sinusoidal images. Actually such patterns are much simpler than real world images, which can be thought of as a superposition of a much larger number of simple patterns.

- *Cognitive Interaction Problem.* It is widely known that cognitive understanding and interactive visual processing (e.g., eye movements) influence the perceived quality of images. For example, a human observer will give different quality scores to the same image if he/she is provided with different instructions. Most image quality metrics do not consider these effects, as they are difficult to quantify and not well understood.

6.4.2 Measures based on structural distortions

In order to overcome the limitations of the error sensitivity based approach, a new framework for the design of image quality measures has been proposed in literature, based on the assumption that the human visual system is highly adapted to extract structural information from the viewing field. It follows that a measure of structural information change can provide a good approximation to perceived image distortion.

The first metric introduced in this context is the Universal Quality Index (UQI) (Wang and Bovik 2002a), that is designed by modeling any image

distortion as a combination of three factors: loss of correlation, luminance distortion, and contrast distortion. The UQI between the original image $\mathbf{x} = \{x_i | i = 1, 2, ..., N\}$ and the test image $\mathbf{y} = \{y_i | i = 1, 2, ..., N\}$ is defined in the following way:

$$UQI = \frac{4\text{cov}(x,y)\mu_x\mu_y}{\left(\sigma_x^2 + \sigma_y^2\right)\left[\mu_x^2 + \mu_y^2\right]} \tag{6.4}$$

where

$$
\begin{aligned}
\mu_x &= \frac{1}{N}\sum_{i=1}^{N} x_i \\
\sigma_x^2 &= \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \mu_x\right)^2 \\
\text{cov}(x,y) &= \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \mu_x\right)\left(y_i - \mu_y\right)
\end{aligned}
\tag{6.5}
$$

The dynamic range of UQI is $[-1, 1]$, with the best value 1 achieved if and only if $x_i = y_i$ for all $i = 1, 2, ..., N$.

An improvement of the UQI is the SSIM (Structural SIMilarity) index (Wang, Bovik, Sheikh and Simoncelli 2004), a full-reference method for measuring the similarity between two images comparing local patterns of pixel intensities that have been normalized for luminance and contrast. If the error sensitivity approach estimates perceived errors to quantify image degradations, this approach considers image degradations as perceived changes in structural information variation.

The SSIM index between two images $\mathbf{x}$ and $\mathbf{y}$ is defined in the following way:

$$\text{SSIM}\left(\mathbf{x}, \mathbf{y}\right) = \frac{\left(2\mu_x\mu_y + c_1\right)\left(2\text{cov}_{xy} + c_2\right)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)} \tag{6.6}$$

where $c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$, L is the dynamic range of the pixel-values and k_1 and k_2 are two constants equal to 0.01 and 0.03 respectively.

For image quality assessment, it is useful to apply the SSIM index locally rather than globally. First, image statistical features are usually highly spatially nonstationary. Second, image distortions may also be space-variant. Third, at typical viewing distances, only a local area in the image can be

perceived with high resolution by the human observer at one time instance. And finally, localized quality measurement can provide a spatially varying quality map of the image, which delivers more information about the quality degradation of the image and may be useful in some applications. Thus, in the SSIM index a 11×11 circular-symmetric Gaussian weighting function $\mathbf{w} = \{w_i | i = 1, 2, ..., N\}$ is used, with standard deviation of 1.5 samples and normalized to unit sum. The estimates of local statistics in Eq. (6.5) are then modified accordingly as:

$$
\begin{aligned}
\mu_x &= \sum_{i=1}^{N} w_i x_i \\
\sigma_x^2 &= \left(\sum_{i=1}^{N} w_i \left(x_i - \mu_x \right)^2 \right)^{\frac{1}{2}} \\
\mathrm{cov}(x, y) &= \sum_{i=1}^{N} w_i \left(x_i - \mu_x \right) \left(y_i - \mu_y \right)
\end{aligned}
\tag{6.7}
$$

With such a windowing approach, the quality maps exhibit a locally isotropic property. To evaluate the overall image quality the authors use a mean SSIM index (MSSIM) :

$$
\mathrm{MSSIM}\left(\mathbf{x}, \mathbf{y}\right) = \frac{1}{M} \sum_{j=1}^{M} \mathrm{SSIM}\left(\mathbf{x_j}, \mathbf{y_j}\right)
\tag{6.8}
$$

where $\mathbf{x}$ and $\mathbf{y}$ are the reference and the distorted images, respectively; $\mathbf{x_j}$ and $\mathbf{y_j}$ are the image contents at the j-th local window; and M is the number of local windows of the image.

This paradigm is a top-down approach, mimicking the hypothesized functionality of the overall HVS. As the authors stated, this approach avoids the suprathreshold problem mentioned in the previous section because it does not rely on threshold psychophysics to quantify the perceived distortions. On the other hand, the cognitive interaction problem is also reduced because probing the structures of the objects being observed is thought of as the purpose of the entire process of visual observation, including high level and interactive processes. Third, the problems of natural image complexity are also avoided to some extent because this philosophy does not attempt to predict image quality by accumulating the errors associated with psychophysically under-

stood simple patterns. Instead, it proposes to evaluate the structural changes between two complex-structured signals directly.

A recently introduced image quality criterion belonging to the approaches based on structural distortions, and based on the comparison between the structural information extracted from the distorted image and from the original image, is the C4 criterion (Carnec et al. 2003). This criterion is fully generic (i.e., not designed for predefined distortion types or for particular images types) and based on a rather elaborate model of the human visual system. This model describes the organization and operation of many stages of vision, from the eye to the ventral and dorsal pathways in the visual cortex. The novelty of this quality criterion relies on the extraction, from an image represented in a perceptual space, of visual features that can be compared to those used by the HVS. The similarity metric computes the objective quality score of a distorted image by comparing the features extracted from this image to features extracted from its reference image (i.e., non-distorted). The features extracted from the reference image constitute a reduced reference which, in a transmission context with data compression, can be computed at the sender side and transmitted in addition to the compressed image data so that the quality of the decompressed image can be objectively assessed at the receiver side.

The authors show in the paper the high correlation between produced objective quality scores and subjective ones and better performance than full reference metrics.

6.4.3   Quality assessment of geometrically distorted images

As already explained, many image quality metrics have been proposed in the last decades exploiting the characteristics of the HVS and some of them have been shown to behave consistently when applied to certain kinds of distortions (e.g. JPEG compression, gaussian noise, median filtering, etc.). However the effectiveness of these metric degrades when they are applied to images affected by geometric distortions because most of them rely on a pixel-by-pixel relationship between the original and the distorted images, as we can see by looking the images in Fig. 6.8 (Weighted PSNR is an improved version of the PSNR: it uses the CSF to weight spatial frequency of error image).
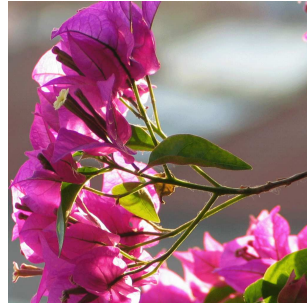
PSNR = 13.1178 dB
SSIM index = 0.58
WPSNR =16.6692 dB
UQI = 0.9222

(a)
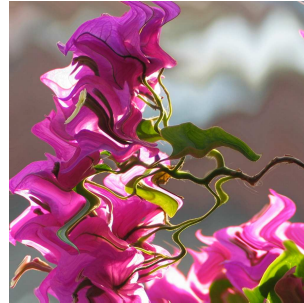
PSNR = 14.2464 dB
SSIM index = 0.6249
WPSNR = 16.6692
UQI = 0.9440

(b)

PSNR = 24.4421 dB
SSIM index = 0.8010
WPSNR = 22.1319 dB
UQI = 0.9925

(c)

PSNR = 19.0849 dB
SSIM index = 0.1656
WPSNR = 21.9162 dB
UQI = 0.9969

(d)

Figure 6.8: *Popular state of the art metric (PSNR, SSIM index, Weighted PSNR and Universal Quality index(Wang and Bovik 2002a)) are used to evaluate different kinds of distortions applied to the same image: (a) good quality local geometric distortion, (b) annoying local geometric distortion, (c) blur, (d) gaussian noise addition*

To the best of our knowledge, only few works can be found in literature regarding the problem of the assessment of geometrically distorted images.

A simple measure, proposed by Licks et al. (Licks et al. 2003) and Baüml et al.(Baüml et al. 2002), is based on the evaluation of the variance of the sampling grid jitter. However this method does not take into account the spectral features of the jitter on which the perception of the transformation depends. This means that two jitter noises with completely different spectral characteristic but with the same variance will be evaluated in the same way.

The method proposed by Desurmont et al. (Desurmont et al. 2004), called Mean Scaling of the Geometric Transformation (MSGT), is based on the average gradient of the sampling grid transformation. This method does not work in presence of geometric distortions based on local permutations of the position of the pixels in the image, as the LPCD model (see chapter 3).

Setyawan et al. (Setyawan et al. 2003) proposed an objective quality measurement scheme based on the use of simple transformation models, for example RST (rotation, scaling and translation) or affine transformation, to locally approximate the underlying complex geometric transform. The approach is based on the assumption that a complex geometric transformation applied on a global scale can be approximated by a simpler transformation applied on a local scale and the visual quality of the overall distortion is determined by the degree of homogeneity of the local geometric transformations. The less homogenous the distortion, the worse the visual quality will be. This method is very expensive from a computational point view due to the optimization involved in its computation. In their paper the authors set to 32 pixels the minimum block size: the choice of this parameter is a trade-off between the precision and the reliability of the approximation, as block sizes that are too small will make the approximation less reliable. As for the MSGT method, this metric does not work in presence of geometric distortions based on the LPCD model.

6.5   Evaluation of the effectiveness of quality metric

The goal of objective quality assessment is to design algorithms whose quality prediction is in good agreement with subjective scores from human observers,

thus, the only benchmark for evaluating the performance of perception-based quality metrics is the comparison with the Mean Opinion Score.

The relationship between an objective metric and MOS does not need to be linear as subjective testing can have nonlinear quality rating compression at the extremes of the test range. It is not the linearity of the relationship that is critical, but the stability of the relationship. To remove any nonlinearities due to the subjective rating process and to facilitate comparison of the models in a common analysis space, the relationship between each models predictions and the subjective ratings is estimated by using a nonlinear regression. Once the nonlinear transformation has been applied, the objective models prediction performance is evaluated by computing various attributes. There are, in fact, a number of attributes that characterize a visual quality metric in terms of its prediction performance with respect to subjective ratings. These attributes are accuracy, monotonicity, and consistency[4], which are defined as follows:

- **Accuracy** is the ability of a metric to predict subjective ratings with minimum average error and can be determined by means of the *Pearson correlation coefficient*. For a set of N data pairs (x_i, y_i), it is defined as follows:

$$r_P = \frac{\sum (x_i - \bar{x}) (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$$

  where $\bar{x}$ and $\bar{y}$ are the means of the respective data sets. The Pearson correlation coefficient must be evaluated between the Mean Opinion Scores provided by the subjects (x_i) and the predicted values of the MOS (y_i) after the nonlinear regression and it ranges from 0 to 1. Another parameter of the accuracy of the model is given by the RMSE (root MSE) between the MOS and the predicted MOS (MOSp).

- **Monotonicity** measures if the increase (or decrease) in one variable is associated with increase (decrease) in the other variable, independently of the magnitude of the increase (decrease). Ideally, differences of a metric's rating between two sequences should always have the same sign

---

[4]http://ftp.crc.ca/test/pub/crc/vqeg/phase1_obj_test_plan.rtf

of the differences between the corresponding subjective ratings. The degree of monotonicity can be quantified by the *Spearman rank-order correlation coefficient*, which is defined as follow:

$$r_S = \frac{\sum \left(\chi_i - \bar{\chi}\right)\left(\gamma_i - \bar{\gamma}\right)}{\sqrt{\sum \left(\chi_i - \bar{\chi}\right)^2}\sqrt{\sum \left(\gamma_i - \bar{\gamma}\right)^2}}$$

where χ_i is the rank of x_i (MOS) and γ_i is the rank of y_i (MOSp) in the ordered data series; $\bar{\chi}$ and $\bar{\gamma}$ are the respective midranks.

- The **consistency** of a metric prediction can be evaluated by measuring the number of outliers. An outlier is defined as a data point (x_i, y_i) for which the prediction error is greater than a certain threshold, usually twice the interval of confidence at 95%.

  The *outlier ratio* is then simply defined as the number of outliers N_O determined in this fashion in relation to the number of data points N:

$$r_O = \frac{N_O}{N}$$

6.6   Concluding remarks

Quality assessment is a central issue in the design, implementation, and performance testing of digital system and the benchmark for any kind of visual quality assessment are subjective experiments. Standard procedures for subjective experiments have been established in the last years as valuable research tools in the image and video processing field for a better understanding of how humans judge quality and the perceived distortions. These procedures, together with some popular approaches to objective quality metrics, have been reviewed in this chapter and will be used throughout this second part of the thesis.

The goal of the following chapters, in fact, will be to describe two objective quality metrics for the evaluation of local geometric distortions in images. The procedures for subjective experiments described in this chapter will be applied to the evaluation of the proposed metrics.

Chapter 7
A quality metric based on MRF

7.1   Introduction

In this chapter we propose a method for objectively assessing the perceptual quality of geometrically distorted images. Our approach is based on the theory of Markov Random Fields. The idea is that the potential function of the Markov Random Field describing the distortion gives an indication of the degradation of the distorted image.

In the first part of the chapter we describe the mathematical model underlying the proposed method, the second part regards the design and implementation of two user-tests to measure human perception of geometric distortions in images. Section 7.3 describes the metric design and in section 7.4 a new subjective test is performed to evaluate the performance of the new metric.

7.2   Theoretical background

In chapter 4 the theory of Markov Random Field was used to introduce a new class of geometric distortions. The idea was to model a geometric distortion with a random field F defined on the set S of the image pixels. The value assumed by each random variable represented the displacement associated to a particular pixel. Specifically, for each pixel, there were two values for the two directions x and y. A displacement field was generated according to the Gibbs probability distribution defined by the Hammersley-Clifford theorem and the particular potential function expressed in (4.6).

The advantage brought by MRF theory is that by letting the displacement of a generic point (x, y) depend on the displacemens of the other points of its neighborhood, we can automatically impose that the resulting displacement field is smooth enough to avoid annoying geometrical distortions. The

idea behind the proposed metric relies on this statement, that is, the potential function of the Markov Random Field describing the distortion gives an indication of the perceptual degradation of the distorted image. It is in fact possible to observe that minimizing an ad hoc potential function we can obtain an invisible distortion that, for the Hammersley-Clifford theorem, corresponds to a highly probable configuration. Thus, once defined a potential function, given a particular distortion and the corresponding configuration, we can evaluate the probability of the occurrence of that configuration by using Eq. (4.4): the more probable the configuration, the more invisible the distortion. The potential function we used to evaluate the degradation introduced by a geometric distortion is the one defined in Eq. (7.1) (it is the same bivariate normal distribution used in chapter 4 with $\sigma_x = \sigma_y = 1$):

$$V_{((x,y)(\tilde{x},\tilde{y}))}(x,y) = \frac{1}{2\pi} \exp\left\{ -\left[\frac{\left(f_h - f_{\tilde{h}}\right)^2}{2} + \frac{\left(f_v - f_{\tilde{v}}\right)^2}{2} \right] \right\} \qquad (7.1)$$

where f_h and f_v are the components of the displacement vector $\mathbf{f}_{(x,y)}$ associated to the pixel (x,y), $(\tilde{x}, \tilde{y})$ is a point belonging to N, that is the 4-neighborhood of (x,y), $f_{\tilde{h}}$ and $f_{\tilde{v}}$ are the components of the displacement vector $\mathbf{f}_{(\tilde{x},\tilde{y})}$ associated to the pixel $(\tilde{x}, \tilde{y})$.

The function described in Eq. (7.1) returns the potential associated to each pair-site clique given by the pixels $((x,y)(\tilde{x},\tilde{y}))$. To compute the score associated to each pixel, quantifying the perceived distortion in that pixel of the image, we need to summarize the potential given by each pair-site clique of the selected neighborhood system. The neighborhood system we used is the first order neighborhood system described by Fig. 4.1 and the overall quality score will be described by the following equation:

$$\text{MF} = \sum_{(x,y) \in I} \sum_{(\tilde{x},\tilde{y}) \in N} V_{((\tilde{x},\tilde{y})(x,y))}(x,y) \qquad (7.2)$$

7.3 Metric design

In this section we design an image quality metric by using the Gibbs potential expressed in Eq. (7.1), we build a function that, given the quality

score expressed by Eq. (7.2), returns a numerical value that quantifies the image quality. As a first step we need to measure the perceived quality of geometrically distorted images trough subjective scaling methods.

The subjective scaling method we used is the ACR method (as described in Sec. 6.4).

Before starting the experiment, a scenario of the intended application of the system under test was explained to the subjects. In addition, a description of the type of assessment, the opinion scale and the presentation of the stimuli was given. The range and type of impairments were presented in preliminary trials, which contained images other than those used in the real tests.

The source image database used for the test included twelve gray scale images, 512×512 pixel in size, and was derived from a set of source images that reflects adequate diversity in image contents. The images, in fact, included pictures of faces, houses and natural scenes. Some images have high activity, while some do not contain many structures and are mostly smooth. The source image database used in the test is shown in Fig. 7.1.

To automatically generate the distortions to be applied to the images we used the Constrained LPCD model (chapter 3) and the MF model (chapter 4) with different parameters in order to obtain different kinds of distortions going from invisible distortions to very annoying distortions, for a total of fourteen different distortions for each image.

The results of the subjective tests are described in Fig. 7.2 which shows the scatter plot of the MOS versus the quality score evaluated by using Eq. (7.2). Specifically, Fig. 7.2.(a) shows the scatter plot for all the 168 images while the following three graphs present the results obtained for each class of images.

We used a Weibull function (the solid plot in the figure) in a fitting procedure to provide a nonlinear mapping between the objective/subjective score as described by the following equation:

$$y = -4 \left(1 - e^{-\left(\frac{x}{b}\right)^k} \right) + 5 \qquad (7.3)$$

where b and k are parameters whose value is reported in table 8.3. The Weibull function expressed in Eq. (8.9) describes our metric: for each potential (the x value in the equation) it returns a numerical score, going from 1 to
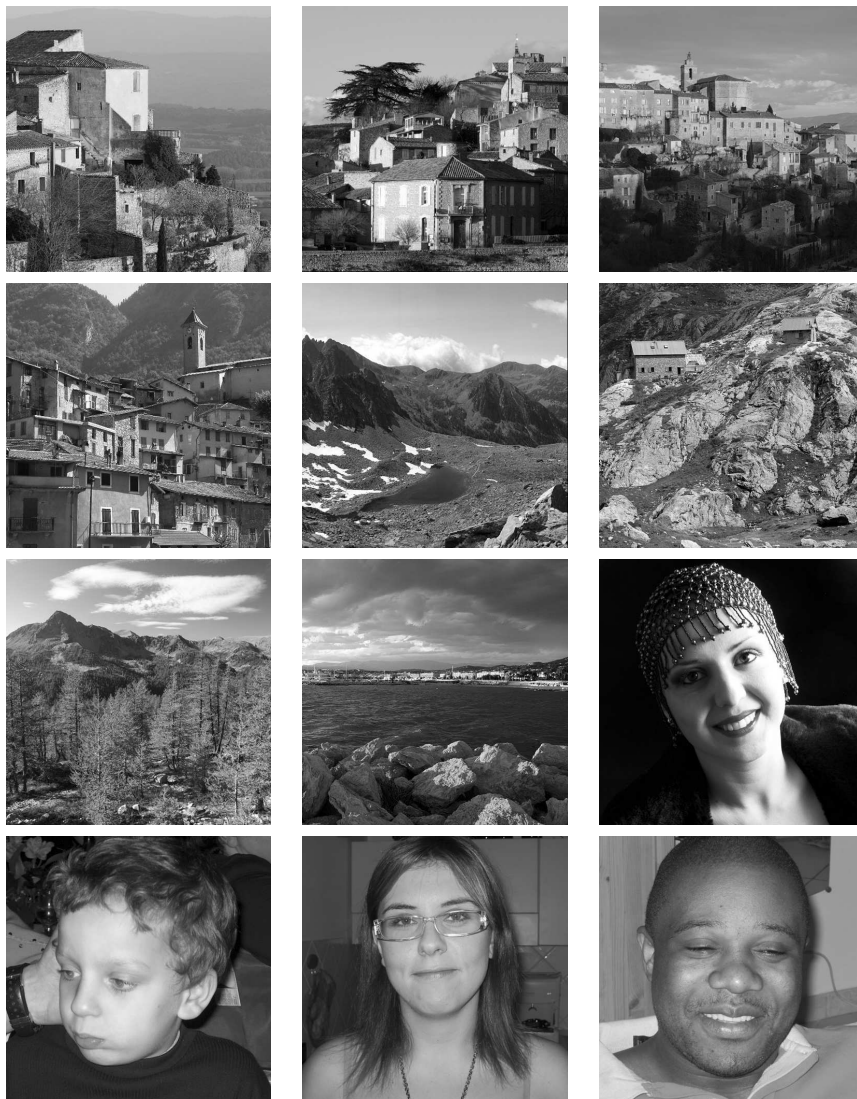
Figure 7.1: *Source image dataset used for the ACR test.*

5, quantifying the dissatisfaction of the viewer observing the distorted image
(with 1 corresponding to a bad image quality and 5 to an excellent image
quality). Thus, given a generic distortion, we evaluate the potential for the

Figure 7.2: *Scatter plot of the Mean Opinion Score versus Potential: (a) all images; (b) house images; (c) natural images; (d) face images.*

corresponding configuration by using Eq. (7.1) and Eq. (7.2), then we obtain the corresponding objective metric by applying to the quality score the fitting function given by Eq. (7.3). The values of the b and k parameters for each class of image are shown in table 7.1, it is possible to use the values of the first column for a generic image.

7.4 Metric validation

After the definition of the metric, we need to validate it. Validation is an important step towards successful development of practical image quality mea-

	All images	House images	Natural images	Face images
k	0.5425	0.6174	0.4693	0.6401
b	0.0504	0.0175	0.06883	0.04527

Table 7.1: *Value of the parameters used for the Weibull fitting function*

surement systems and the most standard form of validation is to compare objective quality measures with ratings by human subjects on an extensive database of images.

7.4.1 Subjective Test

To validate the proposed metric, a new subjective test was designed and performed. Once again we used the ACR test following the procedures explained in the previous section. A new dataset of twelve images (shown in Fig. 7.3) was used according to the class of images explained above and new ten different distortions for each image were generated by using the same models C-LPCD and MRF. The tests involved a panel of other fifteen subjects, all naives with respect to image quality assessment methods and image impairments.

Fig. 7.4 shows the results we obtained from the psychovisual experiment: Fig. 7.4.(a) shows the scatter plot of the Mean Opinion Score versus the Objective metric evaluated as explained in the previous section for all the 120 images while the following three graphs present the results obtained for each class of images.

In order to provide quantitative measures on the performance of the proposed model, we follow the performance evaluation procedures employed in the VQEG and explained in the previous chapter.

To remove any nonlinearities due to the subjective rating process and to facilitate comparison of the models in a common analysis space, the relationship between objective data and the subjective ratings was estimated by using a nonlinear regression, the Weibull functions described by the solid plots in Fig. 7.4. By looking at the plots in the figure it is possible to observe that, after the nonlinear regression, the relationship between the objective and sub-

Figure 7.3: *Source image dataset used for the validation test.*

jective data is almost linear thanks to the fitting procedure that we used for the metric design and described in the previous section, however we applied the nonlinear transformation to follow the standard performance evaluation
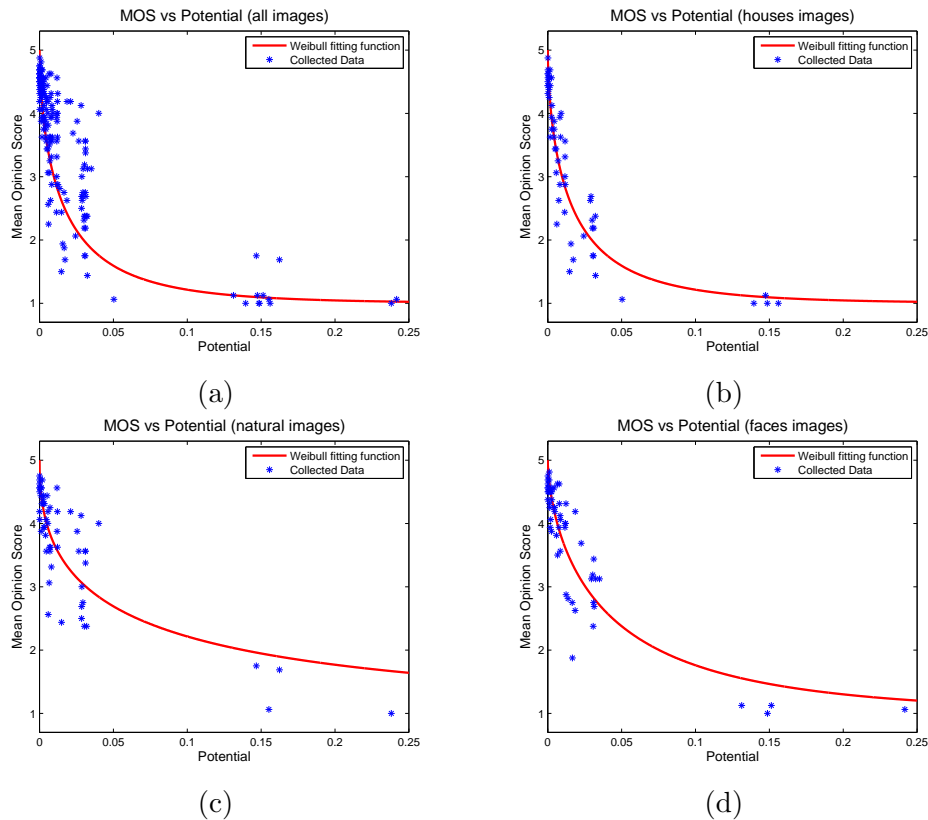
Figure 7.4: *Scatter plot of the Mean Opinion Score versus Objective metric: (a) all images; (b) house images; (c) natural images; (d) face images*

procedures.

The result of the performance evaluation of the proposed algorithm, as described in Sec. 6.5, is shown in table 7.2.

By looking at the table we can see that the outlier ratio is always equal to zero, meaning that the model maintains prediction accuracy over the range of image sequences, and both the Pearson and the Spearman coefficient are quite high revealing a good prediction accuracy and monotonicity of the model. Applying the model per class of images instead to all the images together allows to obtain a little improvement of the performance of the objective metric but it is not so relevant considering the disadvantage of having different objective

	All images	House images	Natural images	Face images
Pearson	0.8179	0.8993	0.8142	0.8601
RMSE	0.5708	0.4603	0.5304	0.4902
Spearman	0.8292	0.8851	0.8359	0.8112
Outlier ratio	0.1403	0.0413	0.1103	0.0803

Table 7.2: *Performance of the objective model*

metrics for different classes of images. By looking at the values in table 7.2 or the scatter plots in Fig. 7.4, we can state that the correlation between the proposed objective quality measure and the ratings by human subjects is evident but additional work is needed to further improve the performance of the proposed metric.

7.4.2 Comparison with other metrics

For sake of completeness, we present the results we obtained by applying some of these metrics to the images used in the subjective test. For objective testing, we used the PSNR measurement and the SSIM-index.

Fig. 7.5 shows the scatter plot of the Mean Opinion Score versus PSNR and SSIM index for all the images and for classes of images. As expected, there is no evident correlation between the two objective metrics and the users response.

7.5 Concluding remarks

In this chapter we have proposed a method of objectively assessing the perceptual quality of images distorted with local geometric distortions. The proposed metric is based on the assumption that the potential function of the configuration defining the distortion gives an indication of the image quality.

The experimental results show quite good performances of the metric, either applying the scheme to a generic image or applying the algorithm to a specific class of images.
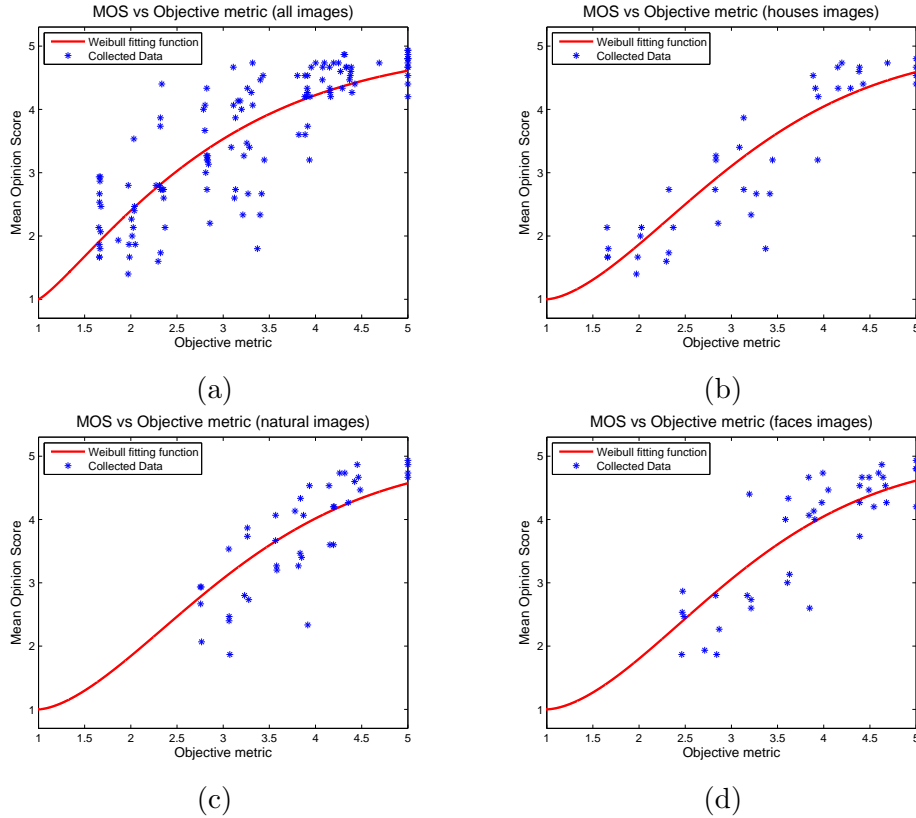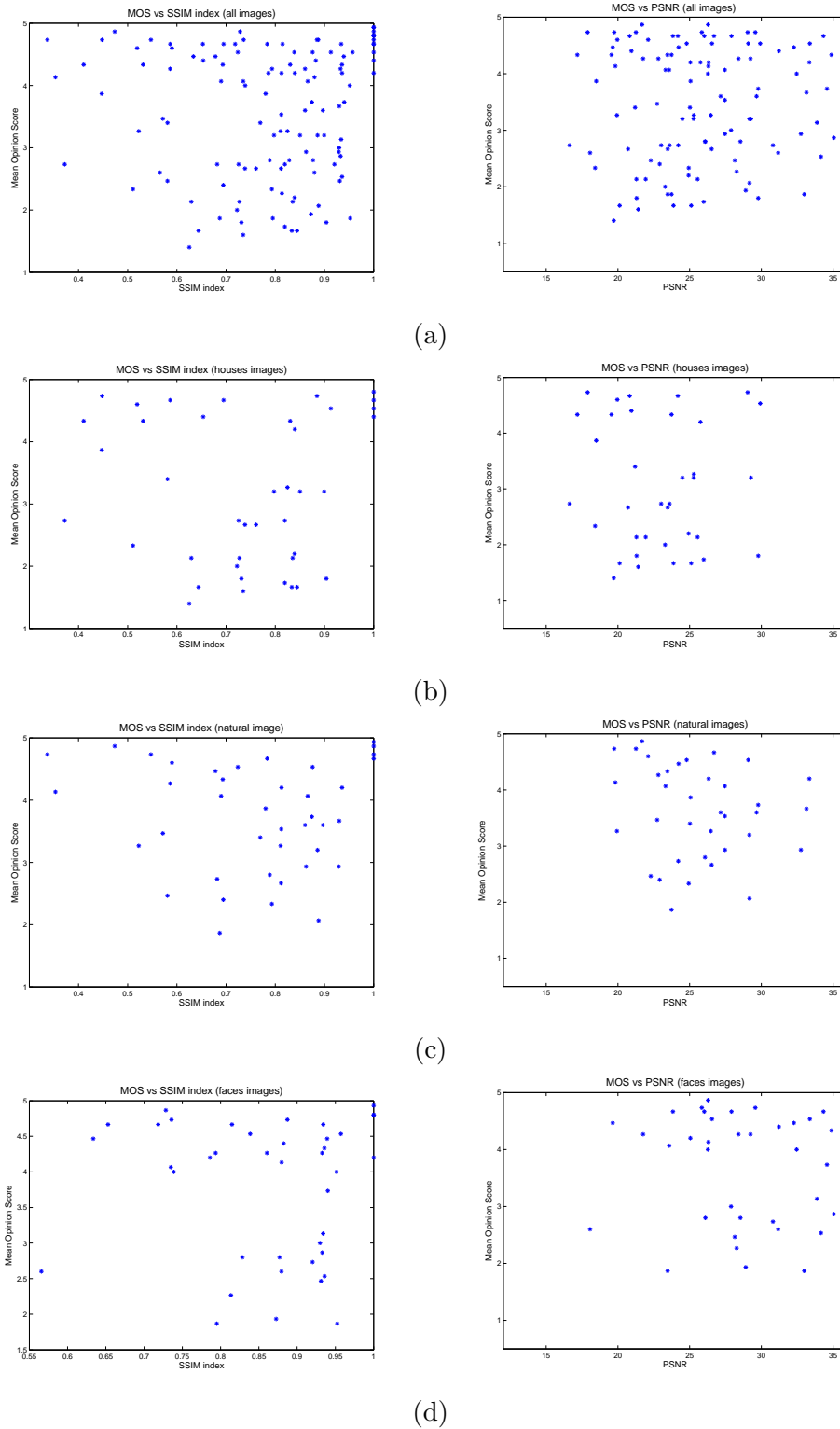
(a)

(b)

(c)

(d)

Figure 7.5: *Scatter plot of the Mean Opinion Score versus PSNR and SSIM index: (a) all images; (b) house images; (c) natural images; (d) face images*

Figure 7.6: *Two distorted images generated using the same displacement field: the image on the right is more annoying.*

The limitation of the proposed metric (common to all the metrics proposed so far for the evaluation of geometric distortions in images), is that it relies only on the displacement field defining the distortion without taking into account the characteristics of the images, that is, the same distortion applied to different images returns the same value of the objective metric. This limitation can be easily understood by looking at the images in Fig. 7.6. These are two distorted images generated by using the same displacement field. In this example all the metrics discussed above would return the same score, however the visual quality of the two distorted images is drastically different since the distortion in the image on the right is more annoying. Consequently, an efficient image-quality measure would need to consider the structural information in the image signals and this is the purpose of the next two chapters.

Chapter 8
Gabor filters based quality metric

8.1   Introduction

In the previous chapter we proposed a method of objectively assessing the perceptual quality of images distorted with local geometric distortions based on the theory of Markov Random Fields. The limitation of that metric is that it does not take into account the characteristics of the images, while an efficient image-quality measure would need to be able to consider the structural information contained in the image. To overcome this problem, in this chapter a new metric is proposed, based on image features processed by human vision. The proposed approach is a full-reference image quality metric based on the use of Gabor filters that have received considerable attention because the characteristics of certain cells in the visual cortex of some mammals can be approximated by these filters. The novelty of the proposed technique with respect to the MF metric and other metrics proposed in the scientific literature, is that it considers both the displacement field describing the distortion and the structure of the image.

The basic idea of our approach is that human vision is sensitive to structures in images, thus a measurement of structural distortions should be a good approximation of the perceived image distortion. In order to provide such a measurement, we need to analyze how the displacement field describing the distortion affects the structure of the image from a perceptual point of view. In our approach we identify image structures with edges and bars and we use Gabor filters to detect them since these filters provide better results with respect to classical features detectors such as Robert Cross or Sobel operators. The use of Gabor filters is also justified by the consideration that their shapes are quite similar to receptive fields found in V4 (Gallant et al. 1993), a visual cortical area of the primary visual cortex. Finally we link bars and edges in

the image to the displacement field describing the geometric distortions.

In doing so, we adopt, as well for the MF metric, a *full-reference* approach since we assume that the displacement field describing the distortion is known. If the original image is available, the displacement field can be estimated by many available techniques like image registration, optical flow or motion estimation. Note that is some cases, the estimation of the displacement field may not be necessary, since the the purpose of the system is just to build a displacement field whose application does not degrade significantly the perceived image quality. This is the case, for instance, of image registration applications, where the displacement field used to register the image at hand on a target image is the final outcome of the registration process, or in digital watermarking, where the goal of the attacker is to build a displacement field that does not impair image quality, but it is strong enough to de-synchronize the detector.

This chapter is divided into two parts. In the first part we describe the Gabor filters on which the metric relies, and, in Sec. 8.3, we present the mathematical background behind the idea of the proposed method. The second part of the chapter regards the design and implementation of two user-tests to measure human perception of geometric distortions in images. The first set of experiments, explained in Sec. 8.4, has been performed to set the parameters of the model. The second set is used to cross-validate (Sec. 8.5), with a different dataset, the performance of the new metric.

8.2   Gabor filters

In this section we briefly summarize the main characteristics of Gabor filters and their application to the problem of edge extraction.

8.2.1   Gabor functions

Two-dimensional Gabor functions were firstly proposed by Daugman (Daugman 1985) to model the spatial summation properties of the receptive fields of simple cells in the visual cortex. These filters have been shown to posses optimal localization properties in both spatial and frequency domain and thus are well suited for image processing applications. Gabor filters have been used

in many applications, like texture segmentation (Jain and Farrokhnia 1991), target detection (Bhanu et al. 1997), fractal dimension management (Super and Bovik 1991), document analysis (Ma and Doermann 2003), edge and bar detection (Mehrotra et al. 1992), retina identification(Zewail et al. 2003), image coding and image representation (Lee 1996).

A Gabor filter can be viewed as a sinusoidal plane wave of particular frequency and orientation, known as *carrier*, modulated by a 2D Gaussian shaped function, known as *envelope*. It is generally described by the following equation:

$$g_{\lambda,\theta,\varphi,\sigma,\gamma}(x,y) = e^{\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)} \cos\left(2\pi\frac{x'}{\lambda} + \varphi\right) \qquad (8.1)$$

where:

$$x' = x\cos\theta + y\sin\theta$$
$$y' = -x\sin\theta + y\cos\theta$$

A Gabor filter is uniquely defined once the $\lambda, \theta, \varphi, \sigma$ parameters are known. λ is the wavelength of the cosine factor of the Gabor filter kernel, θ specifies the orientation of the normal to the parallel stripes of a Gabor function, φ is the phase offset and the aspect ratio γ specifies the ellipticity of the support of the Gabor function.

In the following a brief description of the parameters of the filter and the effect of the parameters on the shape of the filter kernel is provided.

Wavelength λ. This is the wavelength of the cosine factor of the Gabor filter kernel. Its value is specified in pixels and valid values are usually smaller than one fifth of the input image size. Fig. 8.1 shows the Gabor filter kernels for different values of λ.

Orientation θ. This parameter specifies the orientation of the normal to the parallel stripes of a Gabor function. Its value is specified in degrees. Fig. 8.2 shows the Gabor filter kernels for different values of θ.

Figure 8.1: *The images above show Gabor filter kernels with different values of the wavelength parameter (5, 10 and 15, from left to right, respectively). The values of the other parameters are the following: orientation 0, phase offset 0, aspect ratio 0.5, and bandwidth 1.*

Figure 8.2: *The images above show Gabor filter kernels with different values of the orientation parameter (0, 45 and 90, from left to right, respectively). The values of the other parameters are the following: wavelength 10, phase offset 0, aspect ratio 0.5, and bandwidth 1.*

Phase offset φ. The phase offset φ in the argument of the cosine factor of the Gabor function is specified in degrees. Valid values are real numbers between -180 and 180. The values 0 and 180 correspond to symmetric functions, while -90 and 90 correspond to anti-symmetric functions. All other cases correspond to asymmetric functions. Fig. 8.3 shows the Gabor filter kernels for different values of φ.

Aspect ratio γ. This parameter specifies the ellipticity of the support of the Gabor function. For $\gamma = 1$, the support is circular. For $\gamma < 1$ the support is elongated in orientation of the parallel stripes of the function. Default value is $\gamma = 0.5$. Fig. 8.4 shows the Gabor filter kernels for different values of γ.
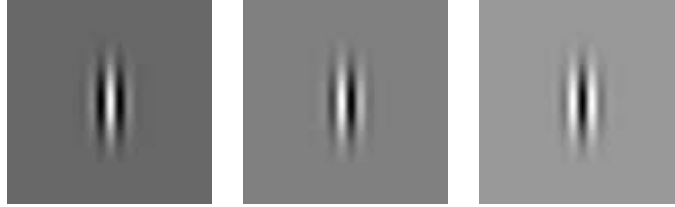
Figure 8.3: *The images above show Gabor filter kernels with different values of the phase offset parameter (0, 180 and −90 degrees, from left to right, respectively). The values of the other parameters are the following: wavelength 10, orientation 0, aspect ratio 0.5, and bandwidth 1.*

Figure 8.4: *The images above show Gabor filter kernels with different values of the aspect ratio parameter (0.5 and 1, from left to right, respectively). The values of the other parameters are the following: wavelength 10, orientation 0, phase offset 0, and bandwidth 1.*

Bandwidth b. The half-response spatial frequency bandwidth b (in octaves) of a Gabor filter is related to the ratio $\frac{\sigma}{\lambda}$, where σ and λ are the standard deviation of the Gaussian factor of the Gabor function and the preferred wavelength, respectively, as follows:

$$b = \log_2 \frac{\frac{\sigma}{\lambda}\pi + \sqrt{\frac{\ln 2}{2}}}{\frac{\sigma}{\lambda}\pi - \sqrt{\frac{\ln 2}{2}}}, \quad \frac{\sigma}{\lambda} = \frac{1}{\pi}\sqrt{\frac{\ln 2}{2}}\frac{2^b + 1}{2^b - 1} \tag{8.2}$$

The value of σ cannot be specified directly. It can only be changed through the bandwidth b. The bandwidth value must be specified as a real positive number. Default is 1, in which case σ and λ are connected as follows: $\sigma = 0.56\lambda$. The smaller the bandwidth, the larger are σ, the support of the Gabor function and the number of visible parallel excitatory and inhibitory stripes.

Fig. 8.5 shows the Gabor filter kernels for different values of b.

Figure 8.5: *The images above show Gabor filter kernels with different values of the bandwidth parameter (0.5, 1, and 2, from left to right, respectively). The values of the other parameters are the following: wavelength 10, orientation 0, phase offset 0, and aspect ratio 0.5.*

8.2.2   Gabor features for edge detection

In recent years, several researchers developed computational models based on Gabor filters for a variety of low level vision tasks and, among these, for edge detection. Edge detection is a fundamental operation in image processing and computer vision. Consequently, much research has been done in this area and a considerable body of literature has been accumulated.

In this context, typical features detectors such as the Roberts Cross or Sobel operators (Torre et al. 1984), which measure the gray-scale gradient component in a given orientation, cannot discriminate between edges or bars while this is possible by using symmetric and antisymmetric Gabor filters. Gabor filters, with respect to other filters, can also solve the problem of integrating the filter responses across different scales.

For $\varphi = 90$ degrees (or -90) the filter in Eq. (8.1) deploies an antisymmetric Gabor function and gives a maximum response at an edge. A 2D Gabor kernel for the edge detection in images can be mathematically defined as:

$$\text{GaborE}_{\lambda,\theta,\sigma,\gamma}\left(x,y\right) = e^{\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)} \sin\left(2\pi\frac{x'}{\lambda}\right) \qquad (8.3)$$

A symmetric Gabor function ($\varphi = 0$ or 180 degrees) can be used for bar detection. The corresponding equation is:

$$\text{GaborB}_{\lambda,\theta,\sigma,\gamma}(x,y) = \text{Bar}_{\lambda,\theta,\sigma,\gamma}(x,y) - \overline{\text{Bar}}_{\lambda,\theta,\sigma,\gamma} \tag{8.4}$$

where

$$\text{Bar}_{\lambda,\theta,\sigma,\gamma}(x,y) = e^{\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)} \cos\left(2\pi\frac{x'}{\lambda}\right) \tag{8.5}$$

and $\overline{\text{Bar}}_{\lambda,\theta,\sigma,\gamma}$ is the mean value of the function defined in Eq. (8.5).

Fig. 8.6 shows the filters for the edge and bar extraction in the vertical direction, with the following parameters: $\theta = 0$ degrees, $\lambda = 10$ pixels, $\gamma = 0.5$ and $b = 1$.
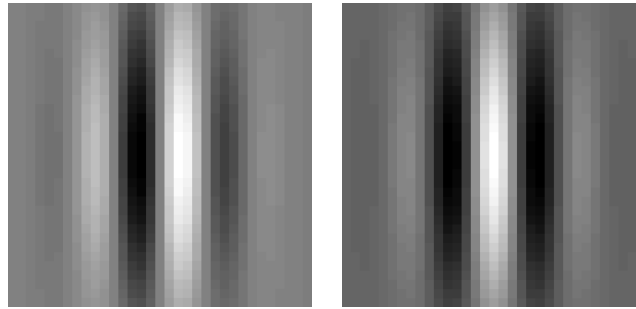

Figure 8.6: *Filters for the edges and bars detection with $\gamma = 0.5, b = 1, \theta = 0, \lambda = 10$: (a) GaborE$_{\lambda,\theta,\sigma,\gamma}$, (b) GaborB$_{\lambda,\theta,\sigma,\gamma}$.*

The filter on the left is an antisymmetric Gabor function and gives a maximum response at an edge, while in the filter on the right, a symmetric Gabor function: the maximum is shifted from the edge, there are actually two maxima, one corresponding to the left and the other to the right of the edge. This symmetric filter gives a maximum response in the presence of a bar.

8.3 Overall architecture of the proposed metric

Images are highly structured since their pixels exhibit strong dependencies, especially when they are spatially close, and these dependencies carry important information about the structure of the objects in the visual scene. It is well known that one of the main functions of the HVS when looking at an

image is to extract structural information from the viewing field, therefore a measurement of structural distortions should be a good approximation of perceived image distortion. Based on this idea, the overall architecture of the proposed metric is described in Fig. 8.7.
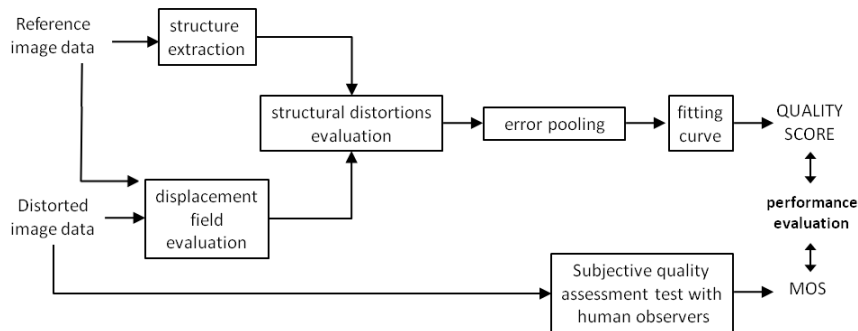

Figure 8.7: *Full process of the objective quality criterion*

Given an original image and the corresponding distorted version, in order to evaluate the perceived image quality, the displacement field is first estimated (*displacement field evaluation*)[1]. At the same time a set of features describing the structures contained in the image are extracted (*structure extraction*). The impact of the displacement field on the extracted structural features is then evaluated (*structural distortions evaluation*) providing a local quality score. The local scores are then pooled to quantify the overall perceived image quality (*error pooling*). Finally a fitting curve is applied with the twofold aim of accounting for saturation effects typical of human quality judgement and to obtain an objective quality score with the same range as the subjective scores (*fitting curve*).

The objective quality score may then be compared to the MOS, that is the scores given by human observers (*subjective quality assessment test with human observers*).

By inspecting the scheme reported in Fig. 8.7, the full-reference nature of the proposed metric clearly comes out. In particular, the availability of the

---

[1]As we already noted, in some applications the displacement may be known in advance, hence this step can be skipped.

original and the distorted images is exploited to recover the displacement field describing the distortion. Note however, that the choice of the algorithm used to estimate the displacement field is not fixed, hence leaving the possibility of tailoring such a step to the application at hand. For this reason, and to develop a metric that is independent of the effectiveness of the displacement estimation step, in the next two sections we assume that the displacement field generating the distortion is known. We will evaluate the impact of this step experimentally, in Sec. 8.5.3.

8.3.1   The main idea

Before describing the details of all the steps described in Fig. 8.7, we explain the main rationale behind the proposed metric, i.e. the approach we chose to measure how the displacement field modifies the structural information contained in the image.

Psychophysical studies show that human vision is sensitive to edges and bars in images, and structures of objects in images are typically outlined by edges and bars (Wandell 1996). Hence, we expect that a measure that links the geometric distortions to the presence of edges and bars in the image is likely to provide an adequate measure of image quality. Based on these considerations, the idea behind our metric is that a geometric distortion causes a degradation of the structure of the objects in the visual scene when the displacement field describing the distortion is orthogonal to the direction of the image bars and edges.

In the following this idea is exemplified by means of examples applied to the synthetic image shown in Fig. 8.8. This image is well suited to show our approach due to the presence of several bars and edges.

For the sake of simplicity let us consider a geometric distortion described by a horizontal displacement field (the vertical displacement field is set to zero), that is each pixel in the image is assigned a displacement vector with an horizontal direction and a specific magnitude. Let us now evaluate the perceived degradation when applying this displacement field to images with differently oriented features. Applying the distortion to an image with several vertical features (Fig. 8.9.(a)) results in a very annoying distortion (Fig. 8.9.(b)), while the distortion is almost imperceptible, except for the border
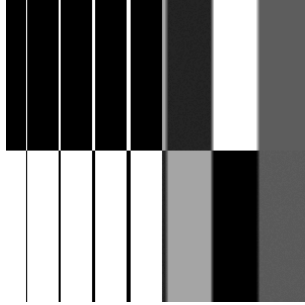
Figure 8.8: *Sample synthetic image.*

effects (Fig. 8.9.(d)) when the displacement field is applied to an image containing only horizontal structures (Fig. 8.9.(c)).

The above considerations guided us in the definition of the *structure extraction* and the *structural distortions evaluation* modules in Fig. 8.7.

8.3.2 Structural feature extraction

As described in Sec. 8.2, we decided to use Gabor filters to extract bar and edges information from the images and to use these features to evaluate the perceptibility of the distortions.

The parameters we used for the design of the filters are the following: $\gamma = 0.5$ and $b = 1$ (these are the default values in typical applications of Gabor filters). In the next section we will see how to set the value of θ and λ.

Once defined the parameters of the filters, and fixed a particular θ, we use the functions described in equations (8.3) and (8.4) to filter the original image and to find edges and bars in the direction orthogonal to θ. The filtering process is described by the following equation:

$$I_f = \sqrt{I_{f,\text{bar}}^2(x,y) + I_{f,\text{edge}}^2(x,y)} \tag{8.6}$$

where I_f is the filtered image and $I_{f,\text{edge}}$ and $I_{f,\text{bar}}$ are obtained by convolving the original image with the Gabor filters described respectively by equations (8.3) and (8.4).
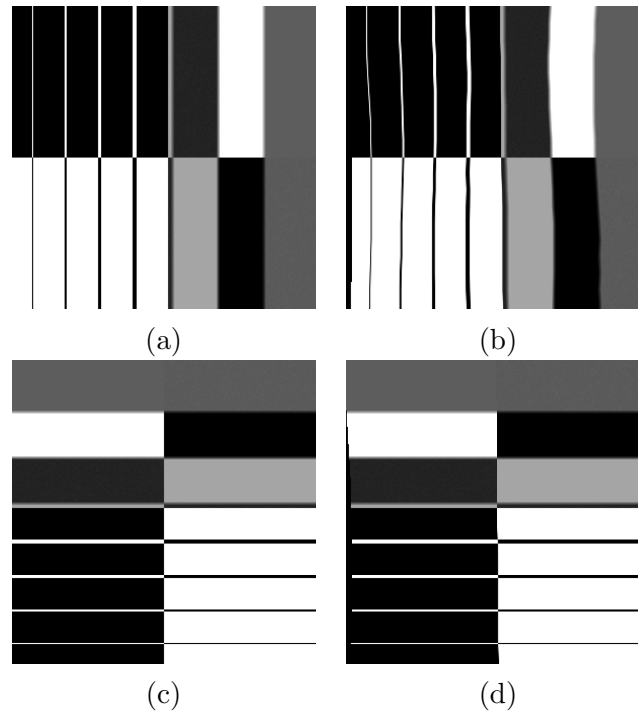
Figure 8.9: *Example of perceptibility of a geometric distortion: (a) original image; (b) application of the horizontal displacement field on the original image; (c) rotated image ; (d) application of the horizontal displacement field on the rotated image.*

In the scientific literature, the outputs of a symmetric and an antisymmetric kernel filter at each image point are usually combined into a single quantity as described by equation 8.6. This quantity is called the Gabor energy and is related to the model of a specific type of orientation selective neuron in the primary visual cortex called the complex cell (Grigorescu et al. 2002).

Fig. 8.10 gives an example of the original synthetic image (see Fig. 8.8) filtered by using equation (8.6) in the two directions $\theta = \frac{\pi}{2}$ and $\theta = 0$.

8.3.3   Local distortion computation

To find the quality score associated to each pixel quantifying the perceivable degradation of the image at that pixel, we need to link the edges and bars of
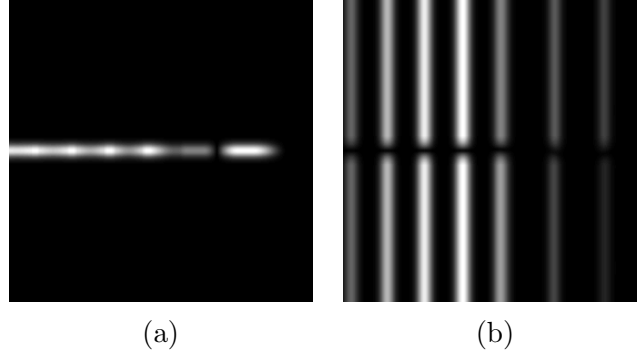
Figure 8.10: *Example of filtered image: (a) filtered image I_f with $\theta = \frac{\pi}{2}$; (b) filtered image I_f with $\theta = 0$.*

the image with the displacement field at the corresponding location. Specifically, as we explained at the beginning of the section, we are interested in the displacement field D_θ (the projection of D along θ), that is orthogonal to the bars and edges of the image, and we have to consider the gradient of D_θ with respect to the direction orthogonal to θ, since we assume that this gradient gives us an indication of the degradation of the structures in the image and thus of the perceptual degradation of the image. Following the above argument, the local quality score associated to each pixel is defined by the following equation:

$$\text{Obj}(x,y) = \sum_\theta \left(I_{f,\theta}(x,y)\right)^\alpha \left(\frac{\partial D_\theta}{\partial d_\theta^\perp}(x,y)\right)^\beta \qquad (8.7)$$

where $I_{f,\theta}$ is the filtered image described by equation (8.6) in the θ direction, α and β are two constants to be fixed, and the notation $\frac{\partial D_\theta}{\partial d_\theta^\perp}$ indicates the gradient of the displacement field in the θ direction with respect to the direction orthogonal to θ (applying the gradient of the displacement field we can associate a score equal to zero to translations that do not affect image quality and we can evaluate the smoothness of the displacement field, since very smooth displacements return small values of the local gradients). The exponents α and β have been inserted to allow a different weighting of the two components of the metric, namely the displacement gradient and the contrast of image structures.

For example, for $\theta = 0$, $I_{f,\theta}$ contains vertical edges and bars, like in Fig. 8.10(b); D_θ is the displacement field in the horizontal direction, $d_\theta^\perp$ is in the vertical direction and $\partial D_\theta / \partial d_\theta^\perp$ is the gradient of the horizontal displacement D_θ in the vertical direction.

The summation over θ in equation (8.7) is needed to extract the salient features along different orientations. Typical applications of Gabor filters require the exploitation of at least 4 different orientations: 0, $\frac{\pi}{4}$, $\frac{\pi}{2}$, $\frac{3\pi}{4}$.

Be aware that the amplitude of the response of the Gabor filter depends on the contrast, as we can see by looking at the images in Fig. 8.10, thus, using this approach, the perceived distortion will depend on the contrast too. This is an important point since it is well known that human vision is more sensitive to high contrast areas of the image. However, there is no evidence that the relationship between the contrast and the perceived distortion is linear, rather, with higher probability, it is a sigmoid function to take into account the saturation effects typical of human senses. In our metric, such a saturation effect is taken into account by means of a fitting function whose goal is to fit the human perception of geometric artifacts with the values provided by the objective metric.

8.3.4 Pooling

In order to pass from the local scores to an overall metric, that is an overall score that quantifies the perceived distortion globally, we included and Error Pooling step. Pooling refers to the task of arriving at a single measurement of quality by starting from local artifacts. The main problem of pooling is that it is not quite understood how the HVS performs pooling, though is quite obvious that pooling involves cognition, where a perceptible distortion may be more annoying in some areas of the scene (such as human faces) than in others. Most quality assessment metrics use Minkowski relation to pool the error signals from the different frequency and orientation selective streams, as well as across spatial coordinates, to arrive at a single fidelity measurement.

In our approach, the global quality score is computed by using the Min-

kowsky relation, as follows:

$$\text{Score} = \left(\sum_{x,y} |\text{Obj}\,(x,y)|^p \right)^{\frac{1}{p}} \tag{8.8}$$

where p is typically a constant between 1 and 4 and whose value was set experimentally as shown in the next section.

All the steps described so far are exemplified in Fig. 8.11 with reference to the synthetic image given in Fig. 8.8. The figure has been obtained by using two orientations ($\theta = 0$ and $\theta = 90$) and the following parameters: $\gamma = 0.5$, $\lambda = 10$, $\alpha = 1$, $\beta = 1$. Fig. 8.11 shows an example of distortion, the filtered images obtained using equation (9.3) and the pixels scores evaluated using equation (8.7) for two separate values of θ ($\theta = 0$ and $\theta = 90$). Fig. 8.11.(f) shows the total score associated to each pixel (found by using equation (8.7)) for both the orientations $\theta = 0$ and $\theta = 90$. By looking at Fig. 8.11.(f) and 8.11.(a) we can see that the local scores are higher at locations with a higher perceived degradation.

8.3.5   Fitting curve

To produce an objective quality score with the same range as the objective scores, and to account for the saturation effect typical of the HVS, a fitting curve is applied to the global quality score obtained trough the error pooling stage. The purpose of the fitting function is to associate the values given by the objective metric to the subjective scores provided by the subjects and this step is always necessary to take into account the saturation effect typical of human senses. Moreover we need to take into account, as explained in the previous section, the non linear relationship between the contrast and the perceived distortion. Through the fitting procedure, a match between the human perception of geometric artifacts and the values provided by the objective metric defined by Eq. (8.8) has to be established.

Many authors transform the value V produced by their criterion into an objective quality score Obj by using the following non-linear function whose parameters are optimized on the tested database:

$$Obj = \frac{a}{1 + \exp\left(b * (V - c)\right)}.$$

(a)   (b)

(c)   (d)

(e)   (f)

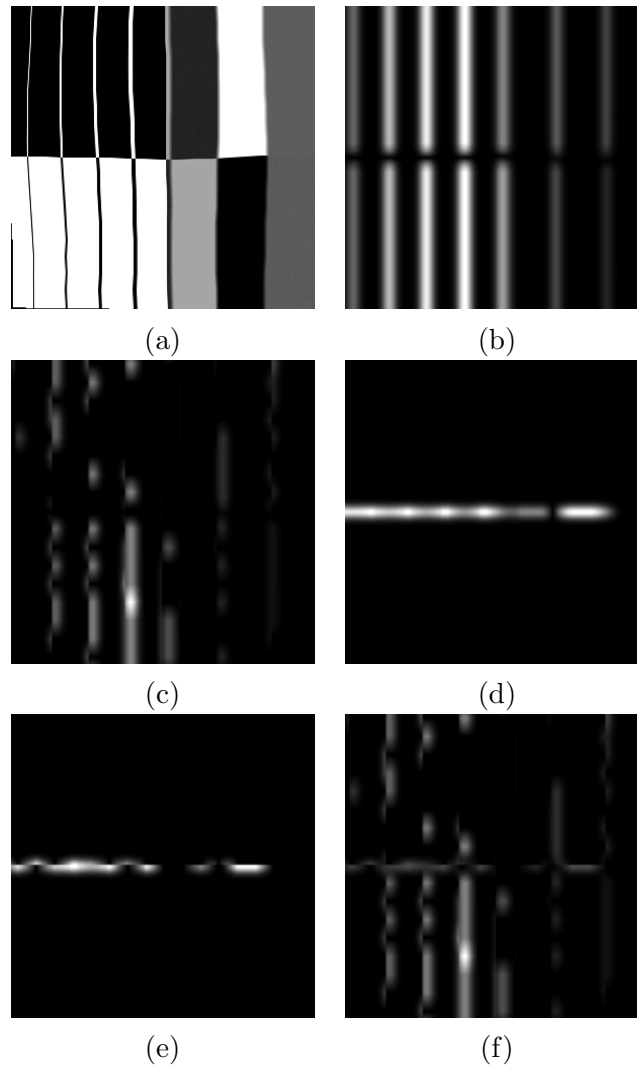**Figure 8.11**: *Example of the proposed approach: (a) distorted image; (b) filtered image $I_f(x, y)$ with $\theta = 0$; (c) pixel scores $\mathrm{Obj}(x, y)$ along the direction $\theta = 0$; (d) filtered image $I_f(x, y)$ with $\theta = 90$; (e) pixel scores $\mathrm{Obj}(x, y)$ along the direction $\theta = 90$; (f) total pixel scores.*

We decided to use the Weibull cumulative distribution function defined as follows:

$$Obj = -4 \left(1 - e^{-\left(\frac{V}{c}\right)^{k}} \right) + 5 \tag{8.9}$$

where c and k are parameters to be estimated by fitting the objective metric values to the subjective data, and V is the global quality score described by equation (8.8). Thanks to the fitting procedure it is possible to obtain a perceptual quality score, going from 1 to 5. We opted for this function because it provides the best fit for our data among the commonly used curves, i.e., Gaussian, logistic and Weibull curves. To estimate the b and k parameters, we used a nonlinear least squares data fitting by the Gauss-Newton method on the tested database, as explained in the next section.

8.4   Metric tuning

In the previous section we described the overall architecture of the metric. We now must tune all the parameters defining the metric to subjective data and to validate the final metric obtained in this way. To this aim, two sets of subjective experiments were carried out with different purposes. The first set of experiments, that is presented in this section, was performed to set the parameters of the model $\theta, \lambda, \alpha, \beta, p$ (the other parameters are the default values of the Gabor filters: $\gamma = 0.5$ and $b = 1$). The second set of experiments was conducted, as we will see in the next section, to validate the proposed metric.

8.4.1   The image database

Image Content

The image database used for the first test included fifteen gray scale high quality images, 512×512 pixel in size. The images of the database include pictures of faces, houses and landscape scenes. Some images have high activity, while some do not have much structures and are mostly smooth. The source image database used in the test is shown in Fig. 8.12.

Figure 8.12: *Source image dataset used for the ACR test.*

Image Distortion type

To automatically generate the local geometric distortions to be applied to the images and to have a broad range of image impairments we used the Constrained LPCD model and the Markov Random Field (MRF) model with different parameters in order to obtain different kinds of distortions going from invisible distortions to very annoying distortions, for a total of nine different distortions for each image and a total of 135 different images to be evaluated. We decided to use these models because they allow to generate a wide range of distortions, by changing the parameters of the models.

8.4.2   Test methodology

We need to measure the perceived quality of geometrically distorted images trough subjective scaling methods. The subjective scaling method we used is the Absolute Category Rating (ACR) method(Int 1996), with a five-level category scale: 1) Bad; 2) Poor; 3) Fair; 4) Good; 5) Excellent. The time pattern for the stimulus presentation is set in the following way: the voting time is equal to 5 seconds and the images presentation time is 8 seconds.

Equipment and display configuration

The experiments were conducted in a dark room by using the VP800 video card of the Cambridge Research Systems together with a high resolution 21-inch digital monitor Mitsubishi DiamondPro 2070 with the external adaptor ViSaGe 71.02.00D2.14[2]. To have a correct color representation a luminosity calibration was previously carried out through a videocamera ColorCAL.14. Subjects watched the images at a distance of 4 times the height of the image displayed on the monitor. All the parameters regarding the test environment have been set consistently with the international standard as explained in Sec. 6.3.2.

---

[2]http://www.crsltd.com/catalog/visage/index.html

Human subjects, training and testing

The tests involved a panel of sixteen subjects with a good vision, all naives with respect to image quality assessment methods and image impairments. Subjects were shown images in a random order, the randomization was different for each subject.

8.4.3 Processing of data

The subjective scores must be analyzed with statistical techniques (Int 1996) to yield results which summarize the performance of the metric. The averaged score values (MOS), that is the arithmetic mean of all the individual scores, are considered as the amount of distortions that anyone can perceive on a particular image. However, impairment is measured according to a certain scale, as explained before, and such a scale may vary from person to person. For this reason, we used standard methods based on Kurtosis coefficient to screen the judgments provided by the subjects and eliminate viewers with extreme scores.

As explained before, the goal of the first experiment was to match human perception of geometric artifacts with the values provided by the objective metric and to set the parameters of the metric so to maximize this matching.

We tested the objective metric with the following values of the parameters: $\alpha \in \{\frac{1}{2}, 1, 2, 3\}$, $\beta \in \{\frac{1}{2}, 1, 2, 3\}$ and $p \in \{1, 2, 3, 4\}$. Regarding the number of orientations to be used, we tested the proposed approach with different numbers of orientations going from 2 (with $\vartheta \in \{0, \frac{\pi}{2}\}$) to 8 (with $\vartheta \in \{0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}, \frac{5\pi}{8}, \frac{3\pi}{4}, \frac{7\pi}{8}\}$).

In order to find the optimum value of the parameters maximizing the match between the quality score and the human responses, we followed the metric performance evaluation procedures employed in the Video VQEG Phase I FR-TV test(Corriveau and Webster 1999) and described in Sec. 6.5.

To remove any nonlinearities due to the subjective rating process and to facilitate comparison of the models in a common analysis space, the relationship between objective data and the subjective ratings was estimated by using a nonlinear regression, the Weibull functions described by the following equation:

$$y = 4\left(1 - e^{-\left(\frac{x-1}{c1}\right)^{k1}}\right) + 1 \tag{8.10}$$

Once the nonlinear transformation was applied, the metric attributes were evaluated by using different performance metrics, applied to the fitted values, that is the Person correlation coefficient, the RMSE between the MOS and the MOS predicted, the Sperman correlation coefficient and the Outlier Ratio. We found that the values that maximize the correlation between the objective metric and the subjective scores are the following: $\alpha = 1$, $\beta = 3$ and $p = 1$. The value of p shows that in the case of geometric distortions the Minkowsky equation is not useful to understand the visual perception mechanism of the human brain, thus the overall score associated to each distortion, will be simply given by the sum of all pixel scores.

Regarding the number of orientations to be used in Eq. (8.7), we observed that adding more than 2 orientations does not increase the correlation between the objective metric and the MOS data, for this reason we decided to use only two orientations: 0, $\frac{\pi}{2}$. This result can be explained by considering that in real word-images, including both natural landscapes and man-made environments, vertical and horizontal orientations are more frequent than obliques ones (Torralba and Oliva 2003). Furthermore it is well known in neurophysiological studies of visual pathways, that the performance for a large variety of perceptual tasks is superior for stimuli aligned in horizontal or vertical orientations, as compared to stimuli in oblique orientations. This phenomena is called the *Oblique Effect* (Appelle 1972).

In table 8.1 the performance of the proposed quality metric, for different number of orientations θ and different values of p in the Minkowski error pooling, is shown. These are four of all the possible configurations we tested to find the optimum values of the parameters used in the metric. As already said, by looking at the table we can observe that increasing the number of orientations does not improve the performance of the metric while changing the value of p in the Minkowski error pooling results in lower correlation coefficients.

One more parameter need to be evaluated in this experiment, that is the wavelenght λ.

Because real-world images contain distinct features at various resolutions,

	Pearson	RMSE	Spearman	Outlier
$\alpha = 1, \beta = 3, \vartheta = 2, p = 1$	0.8419	0.5928	0.8182	0.1333
$\alpha = 1, \beta = 3, \vartheta = 2, p = 4$	0.7181	0.7945	0.6706	0.1852
$\alpha = 1, \beta = 3, \vartheta = 8, p = 1$	0.8435	0.6108	0.8100	0.1300
$\alpha = 1, \beta = 3, \vartheta = 8, p = 4$	0.6636	0.8547	0.6372	0.2519

Table 8.1: *Performance of the proposed perceptual metric as a function of θ and p*

efficient feature extraction may require the filtering process across several scales. In our approach this consideration corresponds to find the correct value of σ (the standard deviation of the Gaussian factor of the Gabor filter) for each image. Using an adaptive value of σ, linking the structure of the image to the scene scale, seems to be a cumbersome task, thus we decided to find the correct scale to be used for each class of images assuming that images belonging to the same class can be described at the same level of resolution. We followed two different approaches to evaluate the performance of the proposed metric either considering all the images together (fixing an unique value of the image scale) or applying the metric to different classes of images. For this purpose, we tested the objective metric with different values of λ (as we said σ and λ are connected as follows: $\sigma = 0.56\lambda$) and we found, for each category of images, the value of the wavelenght of the cosine factor of the Gabor filter that minimizes the error in the fitting procedure ($\lambda = 9$ for all images together, $\lambda = 10$ for the class of house images, $\lambda = 6$ for landscape images, $\lambda = 8$ for face images). This approach can be justified by considering that natural images can be classified in basic-level categories (Tversky and Hemenway 1983) and images belonging to the same categories share some common features and statistics (Torralba and Oliva 2003) including image scale. In fact, statistics of natural images have been found to follow particular regularities, specifically different categories exhibit different orientations and spatial frequency distributions, captured in the averaged power spectra. These results are usually applied to the problem of scenes categorization in the area of computer vision.

The results of the subjective test, using the values found for $\theta, \lambda, \alpha, \beta, p$, are described in Fig. 8.13 which shows the scatter plot of the Mean Opinion

	All images	House images	Landscape images	Face images
λ	9	10	6	8

Table 8.2: *Value of the wavelength for each class of images*
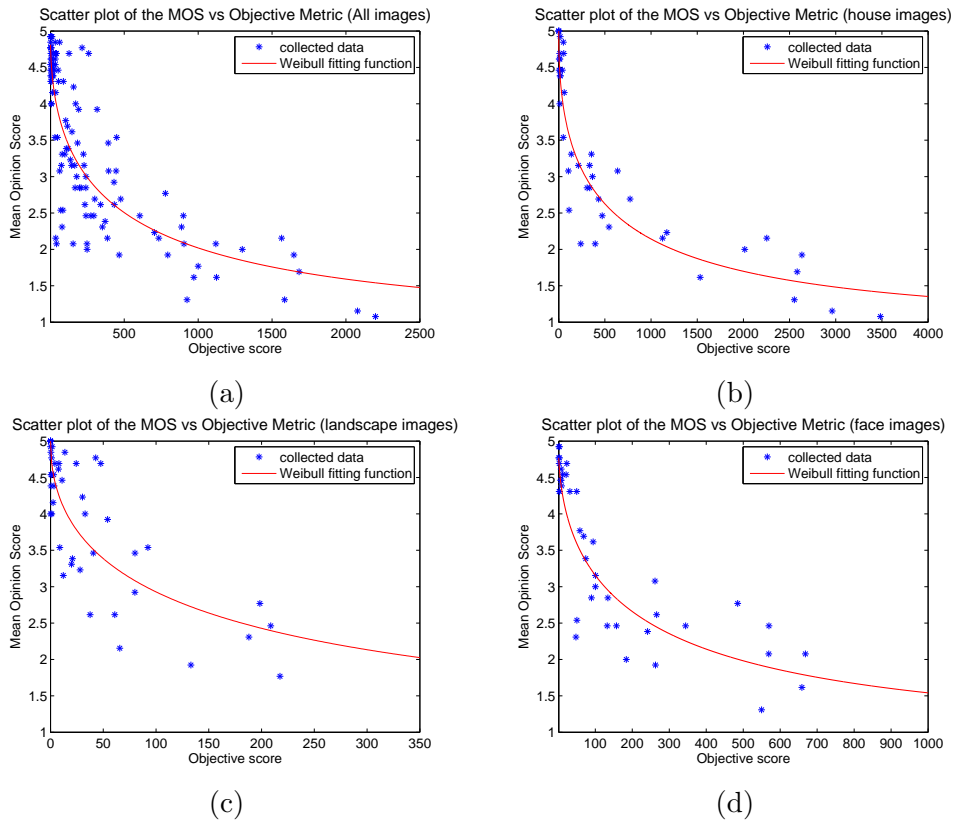

(a)

(b)

(c)

(d)

Figure 8.13: *Scatter plot of the Mean Opinion Score versus Objective Metric: (a) all images; (b) house images; (c) landscape images; (d) faces images.*

Score versus the objective metric evaluated by using Eq. (8.8). Specifically, Fig. 8.13.(a) shows the scatter plot for all the 135 images while the subsequent graphs present the results obtained for each class of images.

The equation of the overall score associated to each distortion, using the values for $\theta, \lambda, \alpha, \beta, p$ found with the subjective test, is given by the following

formula:

$$\text{Score} = \sum_{x,y} \sum_{\theta \in S} \text{I}_{\text{f},\theta}(\text{x},\text{y}) \left(\frac{\partial \text{D}_\theta}{\partial \text{d}_\theta^\perp}(\text{x},\text{y}) \right)^3 \tag{8.11}$$

where $S = \{0, \frac{\pi}{2}\}$ and the final perceptual metric is described by the following equation:

$$\text{GaborMetric} = -4 \left(1 - e^{-\left(\frac{\text{Score}}{c} \right)^k} \right) + 5 \tag{8.12}$$

where c and k are parameters whose value is reported in table 8.3. The Weibull function in Eq. (8.12) describes our metric: for each distortion it returns a numerical score, going from 1 to 5, quantifying the dissatisfaction of the viewer observing the distorted image.

	All images	House images	Landscape images	Face images
c	524.58	623.5	188.4	257
k	0.4838	0.4771	0.4979	0.5101

Table 8.3: *Value of the parameters used for the Weibull fitting function*

8.5 Metric validation

After the definition of the metric, we need to cross-validate it. Cross-validation is an important step towards successful development of practical image quality measurement systems and the most standard form of validation is to compare objective quality measures with ratings by human subjects on an extensive database of images.

8.5.1 Subjective test

To cross-validate the proposed metric a new subjective test was designed and performed. Once again we used the ACR test following the procedures explained in the previous section. A new dataset of fifteen images, shown in Fig. 8.14, was built according to the class of images explained previously and

new nine distortions for each image were generated by using the same models C-LPCD and MF. The tests involved a panel of other sixteen subjects.

The results of the test are shown in Fig. 8.15 that describes the scatter plot of the MOS versus the perceptual metric described by equation (8.12) for all the images and for classes of images. The interval of confidence (IC) at 95% has been added to the plots to have an idea of the accuracy of the MOS value (for the purpose of visibility the IC has not been added to the plot of all images).

The performance of the proposed metric, through the performance evaluation methods described in the previous section, are shown in table 8.4.

	All images	House images	Landscape images	Face images
Pearson	0.8322	0.8808	0.7465	0.9054
RMSE	0.5858	0.5163	0.5370	0.4421
Spearman	0.8482	0.8969	0.7778	0.8772
Outlier	0.1407	0.0667	0.1111	0.0222

Table 8.4: *Performance of the proposed perceptual metric*

By referring to this table and by looking at the scatter plot in Fig. 8.15, the following considerations are in order:

- Both the Pearson and the Spearman coefficient are quite high for all the classes of images revealing a good prediction accuracy and monotonicity of the model. The outlier ratio is quite low revealing that the metric maintains prediction accuracy over the range of image sequences.

- Applying the model per class of images instead than to all the images together brings a little improvement of the objective metric for the class of house and face images but it is not so relevant considering the disadvantage of having different objective metrics for different classes of images.

- The performance of the proposed metric slightly decreases for the class of landscape images. This result can be explained by observing that it

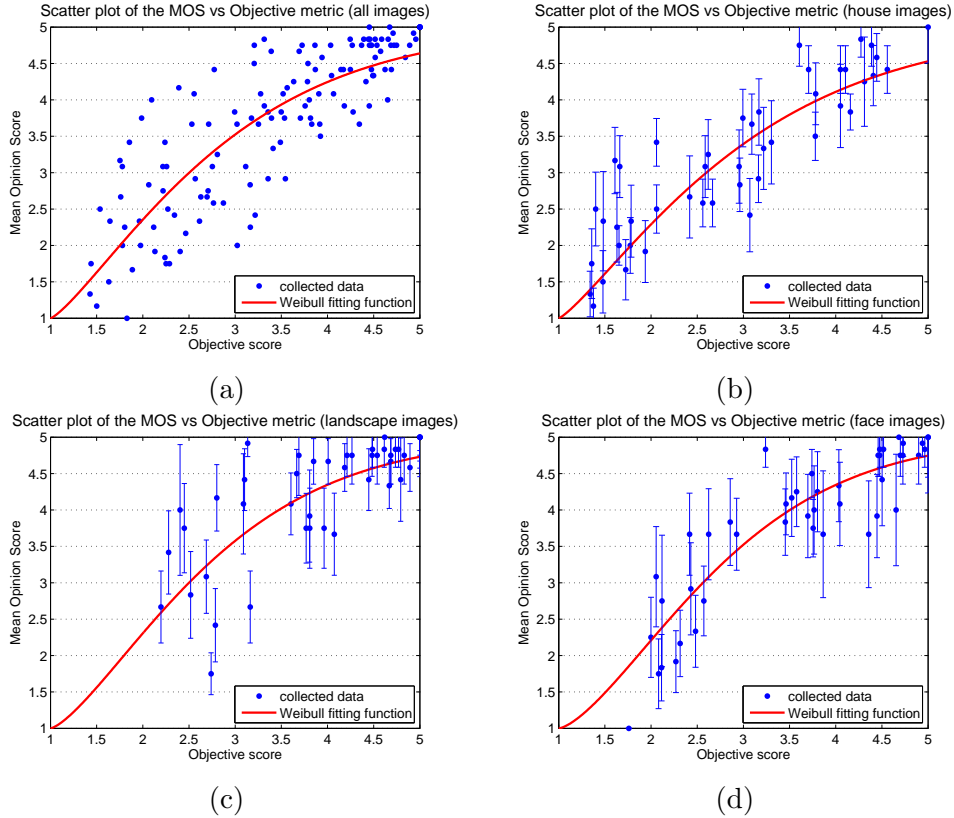**Figure 8.14**: *Source image dataset used for the ACR test.*

Figure 8.15: *Scatter plot of the Mean Opinion Score versus Gabor metric: (a) all images; (b) house images; (c) landscape images; (d) face images.*

is more difficult to perceive the distortions in this kind of images due to the absence of edges and bars.

- After the nonlinear regression the relationship between the objective and subjective data is almost linear thanks to the fitting procedure that we used for the metric design and described in the previous section.

8.5.2   Comparison with other metrics

We now provide a comparison of the proposed technique with other metrics in the literature. Specifically we considered the PSNR measurement and the

SSIM-index (Wang, Bovik, Sheikh and Simoncelli 2004), that are two widely used full-reference quality metrics thanks to their simple formulation and computation, the C4 metric (Carnec et al. 2008) that is a reduced-reference metric, the Gibbs metric (D'Angelo et al. 2008) and the metric based on the variance of the jitter noise (Licks et al. 2003), that are specifically designed to deal with geometric distortions.

PSNR and SSIM index

We expect that statistical measures based on pixel-wise comparisons between the original and the distorted image, like PSNR and SSIM index, cannot work in presence of geometric distortions. However, just for completeness, Fig. 8.16(a) and Fig. 8.16(b) show the scatter plot of the MOS versus PSNR and SSIM index, respectively, for all the images (similar results hold for classes of images). As we expected there is no evident correlation between the two objective metrics and the users response.

The overall quality value defined by the SSIM index is simply obtained as the average of the local quality map.

Depending on the application, it is also possible to compute a weighted average of the different samples in the SSIM index map.

Optimal weighting assignment is difficult because many psychological aspects are involved. For example, region-of-interest image processing systems may give different weights to different segmented regions in the image. As another example, it has been observed that different image textures attract human fixation with varying degrees and can be employed to define the weights.

A weighted version of the SSIM index, as described in (Wang, Lu and Bovik 2004), was also tested and the scatter plot of the Mean Opinion Score versus WSSIM (weighted SSIM) is shown in Fig. 8.16(c). As we expected there is no evident correlation between the improved version of the SSIM index and the users response.

Recently an improvement of the SSIM index has been proposed (Wang and Simoncelli 2005) based on the complex wavelet domain, which is insensitive to luminance change, contrast change, spatial translation, scaling and rotation when small relative to the size of the wavelet filters. The advantage of this metric, with respect to the SSIM index, is that it can evaluate the similarity
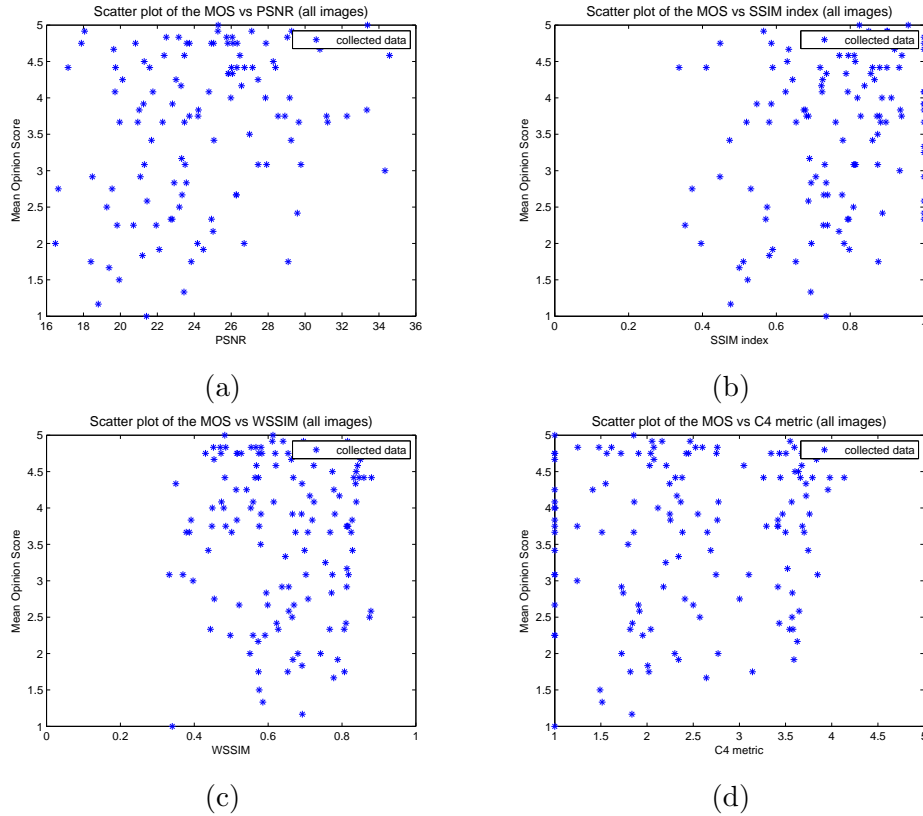
Figure 8.16: *Scatter plot of the Mean Opinion Score versus PSNR (a), SSIM index (b), Weighted SSIM (c) and C4 metric (d) for all images*

of two images without a precise registration process thanks to the translation insensitive property. However it relies, like the SSIM index or the PSNR, on a pixel to pixel comparison thus it can not be applied to the case of geometric distortion that is the purpose of our approach.

C4 metric

Fig. 8.16(d) shows the scatter plot of the Mean Opinion Score versus C4 metric (described in Sec. 6.4.2) for all the images. It is possible to observe that there is no evident correlation between the objective metric and the users response.

Jitter noise variance

The measure proposed by Licks et al. (Licks et al. 2003), focuses the attention on the case where the jitter function $g(t)$ describing the geometric distortion is such that its sample values are equal to $g(n2T) = J = (j1, j2)$ where $j1$, and $j2$, are i.i.d. samples drawn from the marginal distribution $f_J(j) \sim N(0, \sigma_J^2)$. The metric is given by the variance of the jitter noise $\sigma_{n_J}^2$ as follows:

$$E\left[n_J^2\right] \cong E\left[\left(\nabla_Z^T J\right)^2\right] = \sigma_J^2 \sigma_{\nabla_z}^2$$

where ∇_Z is the gradient of the distorted image (readers may refer to (Licks et al. 2003) for details).

Fig. 8.17 shows the scatter plot of the Mean Opinion Score versus the jiiter noise variance for all the images. Although this measure is specifically designed to deal with geometric distortions we can observe that there is no correlation between the objective metric and the MOS.
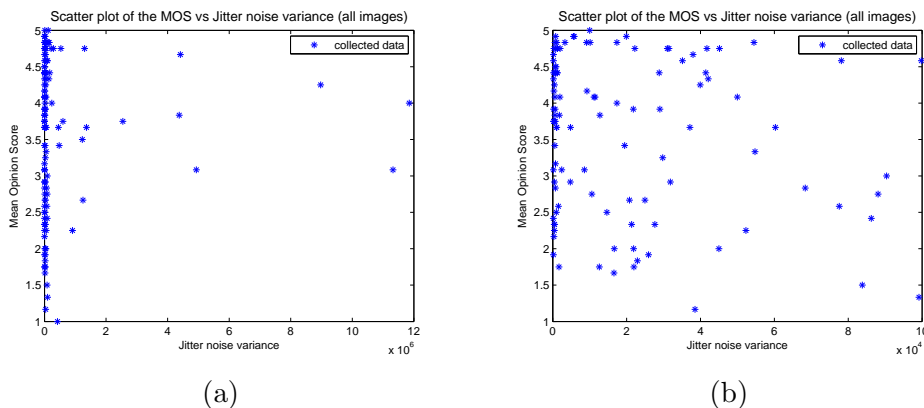

(a) (b)

Figure 8.17: *Scatter plot of the Mean Opinion Score versus Jitter noise variance (on the left there is the total plot, on the right a zoomed version of it).*

Markov Field metric

We tested the MF metric described in chapter 8.

The results of the test are shown in Fig. 8.18 that describes the scatter plot of the MOS versus the MF metric for all the images and for classes of

images. The interval of confidence (IC) at 95% has been added to the plots to have an idea of the accuracy of the MOS value.

The performances of the MF metric, through the standard performance evaluation methods described in the previous section, are shown in table 8.5.
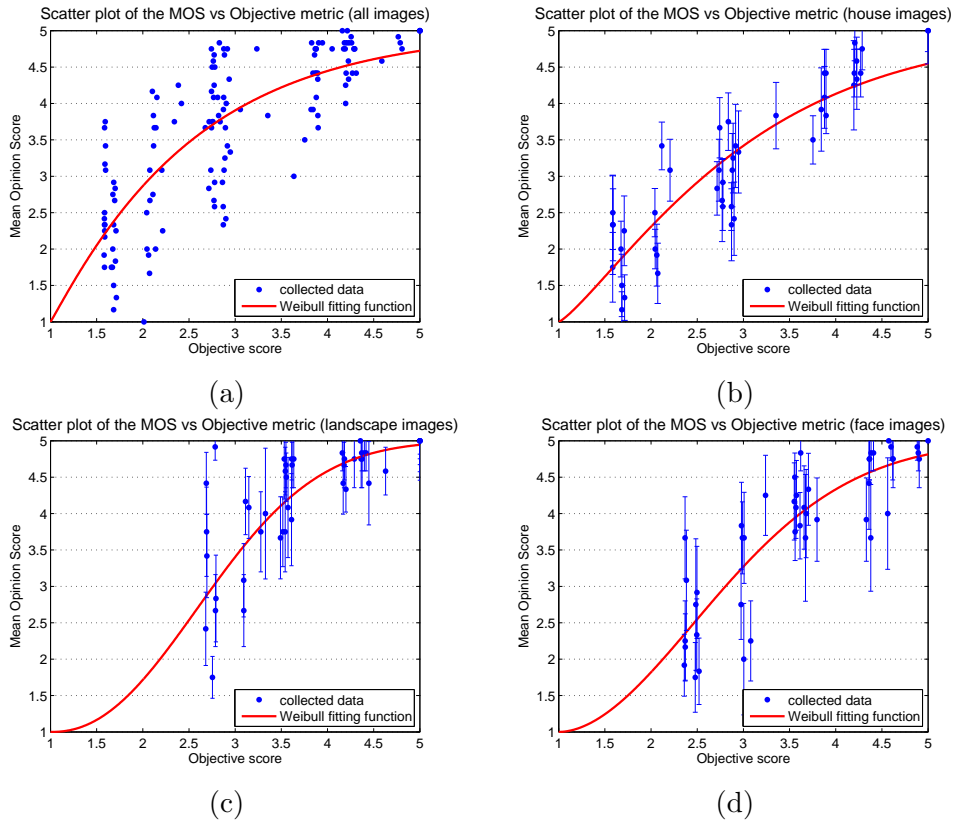

(a)

(b)

(c)

(d)

Figure 8.18: *Scatter plot of the Mean Opinion Score versus Gibbs metric: (a) all images; (b) house images; (c) landscape images; (d) face images.*

By looking at the table we can see that Gibbs metric maintains prediction accuracy over the range of image sequences, and both the Pearson and the Spearman coefficient are quite high revealing a quite good prediction accuracy and monotonicity of the model, except for the class of landscape images, for which the objective metric is not able to predict the human responses.

By comparing the Gabor-based approach with the Gibbs metric we can

	All images	House images	Landscape images	Face images
Pearson	0.7910	0.8766	0.5018	0.8547
RMSE	0.6455	0.4792	0.5412	0.5098
Spearman	0.8054	0.8993	0.4842	0.8447
Outlier	0.2148	0.1433	0.1111	0.0667

Table 8.5: *Performance of the Gibbs metric*

observe the significant improvement brought by the Gabor-based metric. The value of both the correlation coefficients improves approximately from 0.80 to 0.84 for all the images together. There is a significantly improvement for the class of landscape and face images, while there is no improvement (the correlation coefficient is almost the same) for the class of house images.

8.5.3 Impact of the displacement field evaluation

So far we have assumed that the displacement field defining the geometric distortion is known. The reason for such an assumption was that we wanted to test the validity of the main ideas behind our metric, regardless of the accuracy of the displacement field estimation. We now evaluate the performance of the proposed metric when a particular displacement estimation algorithm is adopted. The algorithm we used for the displacement field evaluation, given the original reference images and the distorted ones, is the well known image registration technique described in (Periaswamy and Farid 2006). This is a general purpose registration algorithm designed for medical images/volumes. Given a source image (the distorted image) and a target image (the original image), it automatically estimates a smooth warp field that brings the source image into register with the target image. This approach is built upon a differential multi-scale framework, allowing to capture both large and small scale transformations. The warp field found by applying this registration technique to the image dataset was used to find the corresponding quality scores with the proposed approach, as described in the previous sections.

The results of the tests are shown in Fig. 8.19 and the efficiency of the quality metric in the presence of the particular displacement field estimation

technique we used is evaluated in table 8.6. As we can see, the performance of the metric decreases only slightly with respect to the results shown in table 8.4.

	All images	House images	Landscape images	Face images
Pearson	0.7814	0.8815	0.7865	0.8358
RMSE	0.6888	0.5158	0.5053	0.6545
Spearman	0.7937	0.8875	0.7594	0.8260
Outlier	0.1926	0.1111	0.1111	0.1778

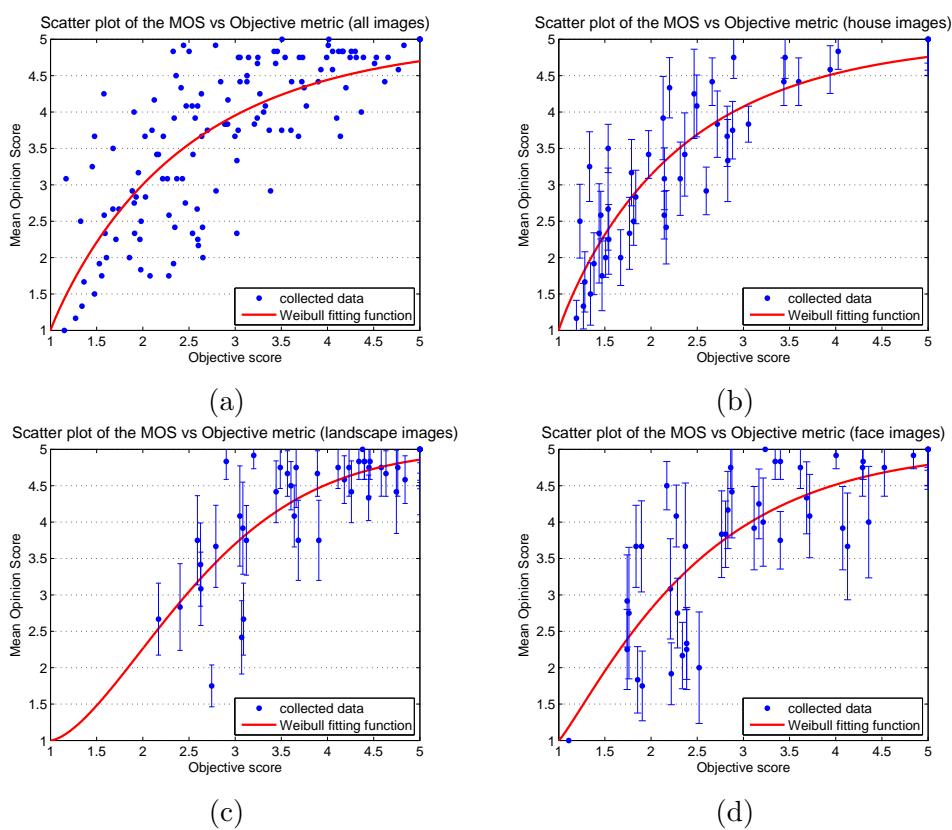**Table 8.6**: *Performance of the proposed metric in the presence displacement field estimation*

Figure 8.19: *Scatter plot of the Mean Opinion Score versus Gabor metric after displacement field evaluation: (a) all images; (b) house images; (c) landscape images; (d) face images.*

Chapter 9

Multiresolution quality metrics

9.1 Introduction

In the previous chapters we have introduced two different quality metrics and the experimental results have shown their efficiency. However, both the MF and Gabor-based metrics are described in a single-scale approach. This is a drawback of the methods because the right scale depends on viewing conditions (e.g., display resolution and viewing distance). Moreover, since real-world images contain distinct features at various resolutions, in order to manage the level of detail of the image and to accomodate a wide range of viewing contexts, we need to cast the metrics in a multiresolution framework.

The goal of this chapter is to improve the performance of the MF metric and the Gabor-based metric by introducing a multiresolution framework. Furthermore, because the MF metric does not take into account the characteristics of the images, we modified it by introducing a perceptibility map to weight the potential function in order to link the displacement field generating the distortion to the characteristic of the image at the same location.

9.2 Theoretical approach

The perceivability of image details depends on the sampling density of the image signal, the distance from the image plane to the observer, and the perceptual capability of the observers. In practice, the subjective evaluation of a given image varies when these factors vary. Single-scale methods as described in the previous sections may be appropriate only for specific settings. Multi-scale methods are convenient way to incorporate image details at different resolutions.

In this section we present the mathematical background behind the multi-resolution approach proposed for the two metrics. Specifically, in Sec. 9.2.1, the multiresolution extension of the feature extraction based on Gabor filters described in Sec. 8.2, is provided. Structural features are extracted from each level of resolution of the image that accurately represent the multiple scales of processing in the human visual system. In Sec. 9.2.2, the multi-scale approach of the Gabor-based metric is described: to obtain an overall quality measure the results obtained by applying the proposed metric at each level of resolution are accumulated.

The multi-scale approach of the MF metric is described in Sec. 9.2.3, which also introduces a perceptibility map in the MF metric to take into account the structure of the image. The perceptibility map is given by the multiresolution feature extraction through the use of Gabor filters as described in Sec. 9.2.1.

9.2.1   Gabor filters

In the previous chapter we have seen how to use Gabor filters to extract bar and edges information from the images and then to apply these features to evaluate the perceptibility of the distortions. In the following a multiresolution extension of the features extraction is provided.

As we have alreday seen in chapter 8, a 2D Gabor kernel can be mathematically defined as:

$$\mathrm{f}_{\lambda,\theta,\sigma,\gamma,\varphi}\left(x,y\right) = e^{\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)}\cos\left(2\pi\frac{x'}{\lambda}+\varphi\right) \qquad (9.1)$$

where:

$$x' = x\cos\theta + y\sin\theta$$
$$y' = -x\sin\theta + y\cos\theta$$

An efficient feature extraction requires the filtering process across several scales. In our approach this consideration corresponds to find the correct value of σ (or λ) for each level of resolution L. We tested different values of σ for each level of resolution and we set experimentally the value of λ to $\frac{10}{2^L}$ pixels.

A 2D Gabor kernel for a given level of resolution L in the θ direction is given by the following equation:

$$f_{L,\theta,\varphi}(x,y) = e^{\left(-2^{2L}\frac{x'^2 + 0.25y'^2}{62,72}\right)} \cos\left(2^L \pi \frac{x'}{5} + \varphi\right) \qquad (9.2)$$

Once defined the level of resolution L and a particular orientation θ, we use the function described in Eq. (9.2) to filter the original image and to find edges and bars in the direction orthogonal to θ. The filtering function we used is described by the following equation:

$$I_{f_{L,\theta}}(x,y) = \sqrt{I^2_{f_{L,\theta,\mathrm{bar}}}(x,y) + I^2_{f_{L,\theta,\mathrm{edge}}}(x,y)} \qquad (9.3)$$

where $I_{f_{L,\theta}}$ is the filtered image along θ and $I_{f_{L,\theta,\mathrm{edge}}}$ and $I_{f_{L,\theta,\mathrm{bar}}}$ are the original image convolved with the Gabor filters described by Eq. (9.2) with $\varphi = -90$ and $\varphi = 0$ respectively.

An example of an image filtered with Eq. (9.3) in the direction $\theta = 0$, across different levels of resolutions, is shown in Fig. 9.1.

The overall filtered image along different orientations and for a given level of resolution L, will be given by:

$$I_{f_L}(x,y) = \sum_{\theta \in S} I_{f_{L,\theta}}(x,y) \qquad (9.4)$$

where $S = \{0, \frac{\pi}{2}\}$[1].

In the next subsection we will see how to apply these considerations to build the MF and Gabor-based metrics in a multiresolution framework.

9.2.2 Multiresolution Gabor-based metric

To obtain a multiresolution version of the Gabor-based metric introduced in chapter 8, we need first of all to resize both the image and the displacement fields by means of bicubic interpolation from the original dimension $S \times S$

---

[1]The performance for a large variety of perceptual tasks is superior for stimuli aligned in horizontal or vertical orientations as compared to stimuli in oblique orientations (please refer to chapter 8 for details).

Figure 9.1: *Filtered image across three different scales and with $\theta = 0$ (the original image is on the right, the filtered image is on the left).*

to $\frac{S}{2^L} \times \frac{S}{2^L}$ and then to apply, for each level of resolution L, the following formula:

$$\text{Gb} = \sum_{x,y} \sum_{\theta \in S} I_{f_{L,\theta}}(x,y) \left(\frac{\partial D_\theta}{\partial d_\theta^\perp}(x,y) \right)^3 \tag{9.5}$$

where $I_{f_{L,\theta}}(x,y)$ is the filtered image described by Eq. (9.3) in the θ direction, $S = \{0, \frac{\pi}{2}\}$ and the notation $\frac{\partial D_\theta}{\partial d_\theta^\perp}$ indicates the gradient of the displacement field in the θ direction with respect to the direction orthogonal to θ.

The multiresolution version of the Gabor-based metric is simply obtained by summarizing the contribution given by each level of resolution, in the

following way:

$$\text{M-Gb} = \sum_{L=0}^{\log\left(\frac{S}{2^6}\right)} 2^L \left(\sum_{x,y} \sum_{\theta \in S} I_{f_{L,\theta}}(x,y) \left(\frac{\partial D_\theta}{\partial d_\theta^\perp}(x,y) \right)^3 \right) \qquad (9.6)$$

The upper bound of L is given by visual constraints in order to avoid losing too many details of the image.

9.2.3 Multiresolution MF metric

The idea behind the MF metric is that the potential function of the Markov Random Field describing the distortion gives an indication of the perceptual degradation of the distorted image. The potential function we used to evaluate the degradation introduced by a geometric distortion is a bivariate normal distribution and the score associate to each pixel in the image, quantifying the perceived distortion in that pixel, is given by the following equation:

$$V_{((x,y)(\tilde{x},\tilde{y}))}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left\{ -\left[\frac{(f_h - f_{\tilde{h}})^2}{2\sigma_x^2} + \frac{(f_v - f_{\tilde{v}})^2}{2\sigma_x^2} \right] \right\} \qquad (9.7)$$

In chapter 7, we set $\sigma_x = \sigma_y = 1$. This parameter is linked to the resolution of the image and, in order to obtain a multiresolution version of the potential function, in this chapter we set $\sigma_x = \sigma_y = \frac{8}{2^L}$, where L is the level of resolution. The new equation for the potential function, for a given level of resolution L, is described by:

$$V_{((x,y)(\tilde{x},\tilde{y}))}(x,y) = \frac{2^L \exp\left\{ -\left[2^{2L} \frac{(f_h - f_{\tilde{h}})^2 + (f_v - f_{\tilde{v}})^2}{128} \right] \right\}}{128\pi} \qquad (9.8)$$

To apply the metric at different levels of resolution we need first of all to resize both the image and the displacement field by means of bicubic interpolation from the original dimension $S \times S$ (for simplicity we consider a square image) to $\frac{S}{2^L} \times \frac{S}{2^L}$, and then to apply Eq. (9.8) to evaluate the potential function for each pair-site clique of pixels.

As explained in the introduction of this section, we improve the performances of the MF metric by introducing both the multiresolution framework

and a perceptibility mask using the Gabor filters. The perceptibility map is obtained by filtering the image along different orientations as described by Eq. (9.4).

The final metric, that is the multiresolution version of the MF metric, is obtained by summarizing the contributions for each level of resolution, as described by the following equation:

$$\text{M-MF} = \sum_{L=0}^{\log\left(\frac{S}{2^6}\right)} \sum_{(x,y)\in I} \left[\left(\sum_{(\tilde{x},\tilde{y})\in N} V_{((\tilde{x},\tilde{y})(x,y))}\left(x,y\right) \right) I_{f_L}\left(x,y\right) \right] \qquad (9.9)$$

9.3 Experimental results

To produce an objective quality score with the same range as the objective scores, the Weibull fitting function described in Eq. (8.9) is applied to the global quality scores given by Eq. (9.6) and Eq. (9.9).

Two sets of subjective experiments were carried out with different purposes. The first set of experiments was performed to set the parameters of the Weibull fitting function, the second set of experiments was conducted to validate the proposed metric.

The image database used for the test is the one shown in Fig. 8.12. To generate the local geometric distortions we used the Constrained LPCD distortion and the Markov Field distortion, with ten different distortions for each image for a total of 150 distortions. The subjective scaling method we used is the ACR method.

To estimate the parameters c and k of the Weibull function we used a nonlinear least squares data fitting by the Gauss-Newton method and we found $c = 1.629$ and $k = 0.9136$ for the Multiresolution Markov Field metric, and $c = 0.3357$ and $k = 0.4968$ for the Multiresolution Gabor based metric.

To validate the multiresolution versions of the metrics another subjective test was designed and performed. We used the ACR test with a different database of fifteen images (the images in Fig. 8.14) and ten new different distortions for each image were generated.

The results of the test are shown in Fig. 9.2, for the Multiresolution MF metric, and in Fig. 9.3, for the Multiresolution Gabor-based metric.
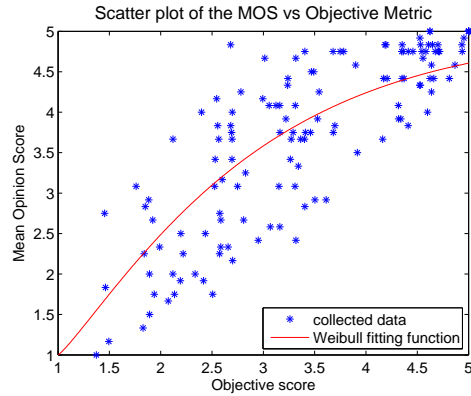
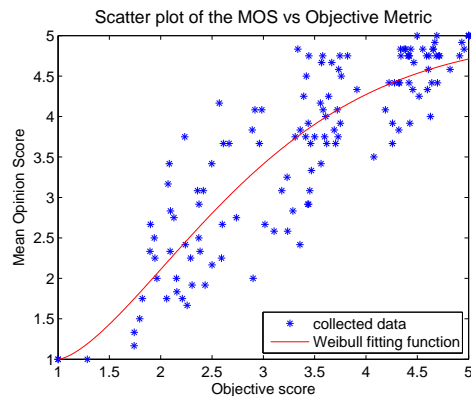**Figure 9.2**: *Scatter plot of the Mean Opinion Score versus Multiresolution MF metric*

Figure 9.3: *Scatter plot of the Mean Opinion Score versus Multiresolution Gabor-based metric*

The correlation between the objective data and the users response is evident in both the plots, anyway, in order to provide quantitative measures on the performance of the proposed models, we followed the performance evaluation procedures employed in the VQEG Phase I FR-TV test and explained in Sec. 6.5. The result of the performance evaluation of the proposed algorithms is shown in table 9.1.

By referring to this table and by looking at the scatter plots in Fig. 9.2

	Pearson	RMSE MOS - MOSp	Spearman	Outlier ratio
M-MF	0.8264	0.5850	0.8298	0.1506
M-Gb	0.8833	0.4945	0.8792	0.0947

Table 9.1: *Performance of the proposed metrics*

and 9.3, we can observe that the outlier ratio have very low value, meaning that the metrics maintain prediction accuracy over the range of image sequences, and both the Pearson and the Spearman coefficients are quite high revealing a good prediction accuracy and monotonicity of both the models. The Multiresolution Gabor-based metric provides better performance than the Multiresolution MF metric. The results of the performance evaluation of the single scale metrics, describe in chapters 7 and 8, are shown in table 9.2.

	Pearson	RMSE MOS - MOSp	Spearman	Outlier ratio
MF metric	0.7910	0.6455	0.8054	0.2148
Gb metric	0.8322	0.5858	0.8482	0.1407

Table 9.2: *Performance of the MF metric and the Gb metric*

By looking at table 9.2 and comparing these results with the ones described in table 9.1 we can observe that, as we expected, the multiresolution framework improves the performances of both the metrics, though the improvement as not strong as expected.

Let us conclude this chapter, and the second part of this thesis, by stating that the improvement from single-scale to multi-scale methods observed in our tests suggests the usefulness of this novel approach. However, this approach is still rather crude. It is necessary to develop it into a more systematic approach by using, for example, different methods to apply the multiresolution framework. Some ideas could be the use of pyramids or wavelet decompositions.

Chapter 10

Conclusions

10.1   Summary and contributions

The work of this thesis can be seen as a first step towards the characterization and quality evaluation of the class of local geometric distortions.

In the last years the problem of evaluating the perceptual impact of geometric distortions in images has received an increasing attention from the watermarking community, due to the central role that such distortions play in watermarking theory. As a matter of fact, the application of a geometric distortion to a watermarked image causes a de-synchronization between the watermark embedder a detector that in most cases prevents the correct extraction of the watermark. All the more, that in most cases the geometric distortion is less annoying than other kind of distortions like noise addition, blurring or lossy compression. This is especially true for local or spatially varying distortions, for which an exhaustive search of the watermark is unfeasible due to the huge size of the search space. A first step to solve the problems with geometric attacks is the characterization of the class of perceptually admissible distortions, defined as the class of geometric distortions whose effect can not be perceived, or is judged acceptable, by a human observer. This requires the development of models to treat the distortions from a mathematical point of view. In this context, the first part of the thesis focuses on modeling local geometric transformations from a mathematical point of view. After the definition of geometric transformations and a classification of global geometric distortions and local geometric distortions, the aim of the subsequent chapters was twofold:

1. to introduce two new models to mathematically describe local geometric distortions and to evaluate, through extensive subjective tests, the perceptual quality impact of the defined transformations on still images.

2. to evaluate the effectiveness of the new models as desynchronization attacks in watermarking systems and compare them with the classical RBA.

The proposed models are the LPCD model, based on local permutations of the position of the pixels in the image with some constraints to obtain perceptually admissible distortions, and the MF model, based on the theory of Markov Random Fields and the idea that the potential function of the configuration defining the distortion is an indication of the perceptual degradation of the image. While the LPCD model is quite simple from a mathematical point of view and this allows to give a Markov chain interpretation of the model and permits to evaluate the cardinality of the distortions that is possible to generate with this model, the MF model, is able to generate larger displacement vectors than the LPCD, while keeping the distortion invisible thanks to the ability of the iterative conditional mode to generate a very smooth field. The characteristics of the proposed models make them suitable for different situations depending on the particular application.

Extensive psychovisual experiments were performed to evaluate the perceptual quality impact of the proposed models and to set the parameters of the models in order to be able to generate invisible distortions.

The de-synchronization capabilities of both the models has been also evaluated and compared to the random bending introduced by Stirmark. The experimental results showed that the two new classes of attacks are more powerful than the local geometric attacks proposed so far.

The analysis carried out in the first part of the thesis can be seen as an initial effort towards the characterization of the whole class of perceptually admissible DAs, a necessary step for the theoretical analysis of the ultimate performance reachable in the presence of watermark de-synchronization and for the development of a new class of watermarking systems that can effectively cope with them.

Watermarking is not the only field where an analysis of geometric distortion in images would be useful. Interested applications include registration of biomedical images that usually requires the application of local and nonlinear transformations, whose strength should be constrained to satisfy certain quality constraints; collusion-secure fingerprinting techniques by random pre-

warping; the problem of recovering 3D models from uncalibrated images of architectural scenes. It is clear that the availability of an objective quality metric capable of dealing with geometric distortions would be of invaluable help in this sense.

In all the above applications, an objective metric capable of assessing the quality of geometrically distorted images is missing. In fact, only few works can be found in literature regarding the problem of the assessment of geometrically distorted images. Thus, in the second part of the thesis, two objective quality metrics for the perceptual evaluation of geometrically distorted images have been introduced.

The first objective metric is based on the theory of Markov Random Fields and on the definition of the potential function of the configuration defining the geometric distortion. It relies on the assumption that the potential function gives an indication of the degradation of the image quality. The limitation of this work is that it only relies on the displacement field defining the distortion without taking into account the characteristics of the images, that is, the same distortion applied to different images returns the same value of the objective metric while the visual quality could be drastically different.

An efficient image-quality measure would need to consider the structural information in the image. To overcome this problem, a new method based on image features processed by human vision has been proposed. The proposed metric is based on the use of Gabor filters to extract the characteristic features of the images and, in the meantime, on the evaluation of the displacement field to take into account the structure of the image and the smoothness of the distortion.

Finally, a multiresolution extension of both the metrics has been introduced to incorporate image details at different resolutions and to obtain a multiresolution extension of the metrics. Though the correlation with human observation is not perfect, the obtained results show that the performance of the metrics developed within the thesis greatly outperforms all the metric proposed so far, be them general purpose metrics or metrics explicitly designed to work with geometric distortions.

10.2   Tracks for future works

This section discusses some of the extensions that may be made to improve the work of this thesis. Regarding the first part of the thesis, future works could be oriented to the application of the results we obtained to the development of watermarking systems robust against de-synchronization attacks. This goal can be formalized in different steps.

- The first step consists in providing theoretical results on the performance limits of watermarking systems under the attacks proposed in this thesis. Specifically, by focusing on 1-bit watermarking, the watermarking problem could be modeled as a joint embedding and detection problem, where the embedder and the detector co-operate to reliably hide a message within the host signal. Finding an optimum solution to this problem is too difficult, however, by imposing some general conditions on the form of the detector, and by considering the set of admissible attacks, an optimum watermarking strategy could be derived. An accurate analysis of the performance could also be obtained, as opposed to current available results based on the loose approximation provided by the union bound assumption.

- The development of new algorithms robust against DAs would stem directly from the theoretical analysis carried out in the previous step. That means to turn into practice the optimal algorithms devised by theory, especially with regard to universal watermarking algorithms that do not rely on a particular statistical model of the host data.

- A second path could be followed to develop new algorithms: to find an efficient way of implementing the exhaustive search strategy that has been shown to be optimum under some general conditions. To this aim we could borrow from the methods usually adopted in video coding for motion estimation, where several efficient algorithms to speed the estimation of motion vectors have been proposed, or from image registration techniques, whereby the to-be-registered image is warped in such a way to maximize the detector answer (e.g. the mutual information or the correlation coefficient). In such a phase, a constraint must be imposed

on the smoothness of the displacement field used for the registration, that ensures that the geometric distortion used during the registration is perceptually admissible. This can be done either by constraining the maximization over the set of distortions for which the perceptual degradation is below a certain limit, or by adding a smoothness term to the objective function. In both cases, the metrics proposed in this thesis will be used to measure the perceptual degradation introduced when a given displacement field is applied to the image.

Concerning the second part of the thesis, although the subjective testing has confirmed the good performance of the objective metrics we developed, there are many areas which require further research.

- To improve the performance of the Gabor based metric is necessary to apply higher level perceptual factors in the proposed techniques. In fact HVS consists of two major components: an early vision model and a visual attention model which indicates regions of interest of a scene through the use of Importance Maps. The attention model takes into account several factors which are known to influence visual attention and eye movements and these factors must be considered to include higher level visual processing in human vision that is not captured by the low level processing of edges and bars and to obtain a more accurate indication of picture quality.

- Human observers are very sensitive to bilateral symmetry in visual patterns. Rapid detection of symmetry may facilitate early visual processes, such as figure ground segmentation and contribute to later processes, such as recognition of objects from novel viewpoints. For this reason, many works on symmetry detection can be found in literature. Geometric distortions can result in a lost of the symmetry of the image and thus in a perceived degradation of the image quality. It could be interesting to integrate a symmetry map in the proposed metrics to take into account this aspect of the HVS[1].

---

[1]Most of the scientific papers dealing with symmetry evaluation apply the model to faces images for which the detection of symmetry is quite simple. The extension to this approach to real world images is a not trivial task and require extensive analysis

- The proposed metrics use only the luminance channel, so the evaluation was carried out on monochromatic images. An extension of the metrics to color images is necessary since real applications use all three color channels. In fact, even if it is well known that colors in images influence the perceived image quality, no results are available on geometrically distorted images. The presence of this kind of distortions in color images must be investigated with extensive psychovisual tests.

- Both the metrics are designed to deal with local geometric distortions. Global geometric distortions, in fact, usually do not affect image quality at all or introduce a degradation that is easily quantifiable linking, for example, the perceived image quality to the parameters of the models. Anyway it could be interesting to incorporate also global geometric distortions into the proposed metrics and to have image quality measures invariant to RST transformations.

- Conclusions drawn using a testbed on specific images do not a priori generalize to other types of content. Additional tests could give better indications of how well the metrics generalize to different contents.

- The presented quality evaluation methods were all based on full reference methods. The development of no reference methods considering the distortion perception still remains an interesting and unexplored topic.

Publications list

This thesis has led to the following publications:

Journal papers

1. A. D'Angelo, M. Barni and Neri Merhav, "Stochastic Image Warping for Improved Watermark De-Synchronization", in *EURASIP Journal on Information Security*, vol. 2008, Article ID 345184, 14 pages, 2008. doi: 10.1155/2008/345184.

2. A. D'Angelo, M. Barni and L. Zhaoping, "A full reference quality metric for geometrically distorted images", submitted to *Transaction on Image Processing*.

Conferences papers

1. A. D'Angelo, G. Menegaz and M. Barni, "Perceptual quality evaluation of geometric distortions in images", in *"Human Vision and Electronic Imaging XII", B. E. Rogowitz; T. N. Pappas; S. J. Daly, eds.*, Proc. of SPIE vol. 6492, 28 January-1 February 2007. S. Jose (CA).

2. M. Barni, A. D'Angelo and Neri Merhav, "Expanding the class of watermark desynchronization attacks", in *Proceedings of 9th ACM Multimedia and Security Workshop*, 20-21 September, Dallas, Texas (USA).

3. A. D'Angelo, M. Pacitto and M. Barni, "A psychovisual experiment on the use of Gibbs potential for the quality assessment of geometrically

distorted images", in *"Human Vision and Electronic Imaging XIII", B. E. Rogowitz; T. N. Pappas; S. J. Daly, eds.*, Proc. of SPIE vol. 6806, February 2008. S. Jose (CA).

4. A. D'Angelo and M. Barni, "A structural method for quality evaluation of desynchronization attacks in image watermarking", in *Proceedings of MMSP 2008- International Workshop on Multimedia Signal Processing*, October 2008, Cairns, Australia. Best student paper award.

5. A. D'Angelo and M. Barni, "Multiresolution quality evaluation of geometrically distorted images", *Proceedings of VPQM 2009- Fourth International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, Janaury 2009, Scottsdale, Arizona (USA).

Bibliography

Abrardo, A. and Barni, M.: 2005, Informed watermarking by means of orthogonal and quasi-orthogonal dirty paper coding, *IEEE Transactions on Signal Processing* **53**(2), 824–833.

Ahumada Jr, A.: 1993, Computational image quality metrics: A review, *SID Digest* **24**, 305–308.

Appelle, S.: 1972, Perception and discrimination as a function of stimulus orientation: the oblique effect in man and animals, *Psychology Bull* **78**(4), 266–278.

Barni, M.: 2005, Effectiveness of exhaustive search and template matching against watermark desynchronization, *IEEE Signal Processing Letters* **12**(2), 158–161.

Barni, M. and Bartolini, F.: 2004, *Watermarking systems engineering: enabling digital assets security and other applications*, CRC Press.

Barni, M., Bartolini, F., De Rosa, A. and Piva, A.: 2001, A new decoder for the optimum recovery of non-additive watermarks, *IEEE Trans. Image Processing* **10**(5), 755–766.

Barni, M., Bartolini, F. and Piva, A.: 2001, Improved wavelet-based watermarking through pixel-wise masking, **10**(5), 783–791.

Bauml, R., Eggers, J. and Huber, J.: 2002, A Channel Model for Watermarks Subject to Desynchronization Attacks, *4th International ITG Conference on Source and Channel Coding, Berlin, Germany, January 28* **30**, 28–30.

Besag, J.: 1986, On the statistical analysis of dirty pictures, *Journal of the Royal Statistical Society. Series B. Methodological* **48**(3), 259–302.

Bhanu, B., Dudgeon, D., Zelnio, E., Rosenfeld, A., Casasent, D. and Reed, I.: 1997, Guest Editorial Introduction To The Special Issue On Automatic Target Detection And Recognition, *IEEE Transactions on Image Processing* **6**(1), 1–6.

Bregni, S., Wiley, J. and Sons, I.: 2002, *Synchronization of digital telecommunications networks*, Wiley New York.

Bretillon, P., Baina, J., Jourlin, M. and Goudezeune, G.: 1999, Method for image quality monitoring on digital television networks, *Proceedings of SPIE*, Vol. 3845, p. 298.

Brown, L. G.: 1992, A survey of image registration techniques, *ACM Computing Surveys* **24**(4), 325–376.

Caputo, B. and Niemann, H.: 2001, From markov random fields to associative memories and back: spin-glass markov random fields, *Proc. of IEEE workshop on Statistical and Computational Theories of Vision, Vancouver*.

Carnec, M., Le Callet, P. and Barba, D.: 2003, An image quality assessment method based on perception of structural information, *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, Vol. 3.

Carnec, M., Le Callet, P. and Barba, D.: 2008, Objective quality assessment of color images based on a generic perceptual reduced reference, *Signal Processing: Image Communication* .

Caviedes, J. and Jung, J.: n.d., No-reference metric for a video quality control loop, *5th World Multiconf. Systemics, Cybernetics and Informatics, SCI2001* .

Cayre, F., Fontaine, C. and Furon, T.: 2005, Watermarking security: theory and practice, *IEEE Transactions on Signal Processing* **53**(10 Part 2), 3976–3987.

Celik, M., Sharma, G. and Tekalp, A.: 2004, Collusion-resilient fingerprinting by random pre-warping, *Signal Processing Letters, IEEE* **11**(10), 831–835.

Cipolla, R., Drummond, T. and Robertson, D.: 1999, Camera calibration from vanishing points in images of architectural scenes, *Proceedings of British Machine Vision Conference* **2**(382-391), 2.

Corriveau, P. and Webster, A.: 1999, VQEG evaluation of objective methods of video quality assessment, *SMPTE journal(1976)* **108**(9), 645–648.

Cover, T. and Thomas, J.: 1991, *Elements of information theory*, Wiley New York.

Cox, I. J., Kilian, J., Leighton, T. and Shamoon, T.: 1997a, Secure spread spectrum watermarking for multimedia, *IEEE Trans. Image Processing* **6**(12), 1673–1687.

Cox, I., Kilian, J., Leighton, F. and Shamoon, T.: 1997b, Secure spread spectrum watermarking for multimedia, *IEEE transactions on image processing* **6**(12), 1673–1687.

Cox, I., Miller, M., Bloom, J. and Honsinger, C.: 2002, Digital Watermarking, *Journal of Electronic Imaging* **11**, 414.

Daly, S.: 1993, The visible differences predictor: An algorithm for the assessment of image fidelity, *Digital images and human vision* **11**.

D'Angelo, A., Pacitto, M. and Barni, M.: 2008, A Psychovisual Experiment on the Use of Gibbs Potential for the Quality Assessment of Geometrically Distorted Image, *Proc. of SPIE Human Vision and Electronic Imaging XIII*, Vol. 6492.

Daugman, J.: 1985, Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, *Optical Society of America, Journal, A: Optics and Image Science* **2**, 1160–1169.

de Ridder, H., Blommaert, F. and Fedorovskaya, E.: 1995, Naturalness and image quality: Chroma and hue variation in color images of natural scenes, *Proceedings of SPIE*, Vol. 2411, p. 51.

Delannay, D. and Macq, B.: 2000, Generalized 2d cyclic patterns for secret watermark generation, *Proc. ICIP2000, Int. Conf. Image processing*, Vol. 2, pp. 72–79.

Desurmont, X., Delaigle, J. and Macq, B.: 2004, Characterization of geometric distortions attacks in robust watermarking, *Proceedings of SPIE* **5306**, 870–878.

Fridrich, J.: 2002, Security of fragile authentication watermarks with localization, *Proc. SPIE Photonic West* **4675**, 691–700.

Gallant, J., Braun, J. and Van Essen, D.: 1993, Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex, *Science* **259**(5091), 100–103.

Grigorescu, S., Petkov, N. and Kruizinga, P.: 2002, Comparison of texture features based on Gabor filters, *Image Processing, IEEE Transactions on* **11**(10), 1160–1167.

Int: 1996, *Subjective Video Quality Assesment Methods for Multimedia Applications Recommendation P.910*.

Jain, A. and Farrokhnia, F.: 1991, Unsupervised texture segmentation using Gabor filters, *Pattern Recognition* **24**(12), 1167–1186.

Kutter, M.: 1999, Watermarking resisting to translation, rotation, and scaling, *Proc. of SPIE, Int. Conf. on Multimedia Systems and Applications*, Vol. 3528, pp. 423–431.

Lee, T.: 1996, Image representation using 2D Gabor wavelets, *IEEE Transactions on pattern analysis and machine intelligence* **18**(10), 959–971.

Li, S.: 1995, *Markov random field modeling in computer vision*, Springer-Verlag London, UK.

Licks, V., Ourique, F., Jordan, R. and Perez-Gonzalez, F.: 2003, The effect of the random jitter attack on the bit error rate performance of spatial domain image watermarking, *Image Processing, 2003. Proceedings. 2003 International Conference on* **2**.

Lin, C. Y., Wu, M., Bloom, J. A., Cox, I. J., Miller, M. L. and Lui, Y. M.: 2001, Rotation, scale, and translation resilient watermarking for images, *IEEE Transactions on Image Processing* **10**(5), 767–782.

Lubin, J.: 1993, The use of psychophysical data and models in the analysis of display system performance.

Ma, H. and Doermann, D.: 2003, Gabor filter based multi-class classifier for scanned document images, *Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on*, pp. 968–972.

MADIGAN, R. and WILLIAMS, D.: 1987, Maximum-likelihood psychometric procedures in two-alternative forced-choice: evaluation and recommendations, *Perception & psychophysics* **42**(3), 240–249.

Mannos, J. and Sakrison, D.: 1974, The effects of a visual fidelity criterion of the encoding of images, *Information Theory, IEEE Transactions on* **20**(4), 525–536.

Mehrotra, R., Namuduri, K. and Ranganathan, N.: 1992, Gabor filter-based edge detection, *Pattern Recognition* **25**(12), 1479–1494.

Merhav, N.: 2005, An information-theoretic view of watermark embedding-detection and geometric attacks, *Proceedings of WaCha, 2005, First Wavila Challenge*, Barcelona, Spain.

Miller, M., Doerr, G. and Cox, I.: 2004, Applying informed coding and embedding to design a robust high-capacity watermark, *IEEE Transactions on Image Processing* **13**(6), 792–807.

Norris, J. R.: 1998, *Markov Chains*, Cambridge University Press, Cambridge.

Papoulis, A.: 1984, Probability, random variables and stochastic processes, *New York: McGraw-Hill, 1984, 2nd ed.* .

Pelli, D. and Farell, B.: 1995, Psychophysical methods, *Handbook of optics* **1**, 29–1.

Periaswamy, S. and Farid, H.: 2006, Medical image registration with partial data, *Medical Image Analysis* **10**(3), 452–464.

Petersch, B., Bogner, J., Fransson, A., Lorang, T. and Pötter, R.: 2004, Effects of geometric distortion in 0.2 T MRI on radiotherapy treatment planning of prostate cancer, *Radiotherapy and Oncology* **71**(1), 55–64.

Petitcolas, F. A. P. and Anderson, R. J.: 1999, Evaluation of copyright marking systems, *Proc. ICMCS99, Int. Conf. Multimedia Comput. and Syst.*, Vol. 1, Firenze, Italy, pp. 574–579.

Piva, A., Barni, B., Bartolini, F., Cappellini, V., De Rosa, A. and Orlandi, M.: 1985, Improving DFT Watermarking Robustness Through Optimum Detection and Synchronization, *GMD Report* **85**, 65–69.

Ponomarenko, N., Carli, M., Lukin, V., Egiazarian, K., Astola, J. and Battisti, F.: 2008, Color Image Database for Evaluation of Image Quality Metrics, *Proceedings of International Workshop on Multimedia Signal Processing* pp. 403–408.

Robson, J.: 1966, Spatial and temporal contrast-sensitivity functions of the visual system, *Journal of the Optical Society of America* **56**(8), 1141–1142.

Setyawan, I., Delannay, D., Macq, B. and Lagendijk, R.: 2003, Perceptual quality evaluation of geometrically distorted images using relevant geometric transformation modeling, *Proceedings of SPIE, Security and Watermarking of Multimedia Contents V* **5020**, 85–94.

Sheikh, H., Sabir, M. and Bovik, A.: 2006, A statistical evaluation of recent full reference image quality assessment algorithms, *IEEE Transactions on image processing* **15**(11), 3440.

Shen, J.: 2003, On the foundations of vision modeling I. Webers law and Weberized TV restoration, *Physica D: Nonlinear Phenomena* **175**(3-4), 241–251.

Shlien, S.: 1979, Geometric correction, registration, and resampling of Landsat imagery, *Canadian Journal of Remote Sensing* **5**, 74–89.

Silverstein, D. and Farrell, J.: 1996, The relationship between image fidelity and image quality, *Image Processing, 1996. Proceedings., International Conference on*, Vol. 1.

Super, B. and Bovik, A.: 1991, Localized measurement of image fractal dimension using Gabor filters, *Journal of Visual Communication and Image Representation* **2**(2), 114–128.

Swanson, M., Kobayashi, M., Tewfik, A., Inc, C. and Minneapolis, M.: 1998, Multimedia data-embedding and watermarking technologies, *Proceedings of the IEEE* **86**(6), 1064–1087.

Torralba, A. and Oliva, A.: 2003, Statistics of natural image categories, *Network: Computation in Neural Systems* **14**(3), 391–412.

Torre, V., Poggio, T., of Technology, M. I. and Laboratory, A. I.: 1984, *On edge detection*, Massachusetts Institute of Technology, Artificial Intelligence Laboratory.

Tversky, B. and Hemenway, K.: 1983, Categories of enviromental scenes, *Cognitive Psychology* **15**(1), 121–149.

Voloshynovskiy, S., Deguillaume, F. and Pun, T.: 2001, Multibit digital watermarking robust against local non linear geometrical distortions, *Proc. Image Processing, 2001. Int. Conf.*, Vol. 3.

Wandell, B.: 1996, *Foundations of Vision*, Sinauer Associates.

Wang, Z. and Bovik, A.: 2002a, A universal image quality index, *Signal Processing Letters, IEEE* **9**(3), 81–84.

Wang, Z. and Bovik, A. C.: 2002b, A universal image quality index, *IEEE Signal Processing Letters* **20**.

Wang, Z., Bovik, A. C., Sheikh, H. R. and Simoncelli, E. P.: 2004, Image quality assessment: from error visibility to structural similarity, *IEEE Transaction on Image Processing* **13**.

Wang, Z., Bovik, A. and Evans, B.: 2000, Blind measurement of blocking artifacts in images, *Proc. IEEE Int. Conf. Image Proc*, Vol. 3, pp. 981–984.

Wang, Z., Lu, L. and Bovik, A.: 2004, Video quality assessment based on structural distortion measurement, *Signal Processing: Image Communication* **19**(2), 121–132.

Wang, Z. and Simoncelli, E.: 2005, Translation Insensitive Image Similarity in Complex Wavelet Domain, *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on* **2**.

Watson, A., Hu, J. and McGowan III, J.: 2001, Digital video quality metric based on human vision, *Journal of Electronic Imaging* **10**, 20.

Webster, A., Jones, C., Pinson, M., Voran, S. and Wolf, S.: 1913, An objective video quality assessment system based on human perception, *SPIE Human Vision, Visual Processing, and Digital Display IV*, pp. 15–26.

Winkler, S.: 1999, Issues in vision modeling for perceptual video quality assessment, *Signal Processing* **78**(2), 231–252.

Winkler, S.: 2005, *Digital video quality: vision models and metrics*, Wiley.

Wolberg, G.: 1998, Image morphing: a survey, *The Visual Computer* **14**(8), 360–372.

Wolf, S., Pinson, M., Voran, S. and Webster, A.: 1991, Objective quality assessment of digitally transmitted video, *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, 1991.*, pp. 477–482.

Wu, H. and Yuen, M.: 1997, A generalized block-edge impairment metric for video coding, *IEEE Signal Processing Letters* **4**(11), 317–320.

Zewail, R., Seil, A., Hamdy, N. and Saeb, M.: 2003, Iris Identification Based on Log-Gabor Filtering, *Proceedings of the IEEE Midwest Symposium on Circuits, Systems, and Computers* .

Index

Digital images are subject to various kinds of distortions that may result in a degradation of visual quality during acquisition, processing, compression, storage, transmission and reproduction. It is therefore necessary for many applications to be able to quantify the image quality degradation that occurs in a system, so that it is possible to control and enhance the quality of the images it produces. For this reason, optimizing the performance of digital imaging systems with respect to the capture, display and transmission of visual information represents one of the biggest challenges in the field of image and video processing. For instance, quality assessment tools predict subjective ratings, image compression schemes reduce the visibility of introduced artifacts, watermarking schemes hide invisible information in images, printer half-toning patterns are perceptually optimized and colors rendered as possible accurately. In all these applications, the characteristics of the Human Visual System (HVS) can be exploited to improve the performance from a visual quality point of view. To achieve this, it is necessary to understand and to build computational models of the HVS model. The need for accurate vision models has been increasing with the introduction of digital processing of visual information….

The Visual Information Processing and Protection (VIPP) group is an informal research group belonging to the Laboratory of Telematics and Tele-communications of the Department of Information Engineering of the University of Siena. The research interests of the VIPP group span the whole area of visual information processing, including image and video processing, with a particular emphasis on the protection of visual information.