SAPIENZA
UNIVERSITÀ DI ROMA

Study of the subjective effects of blur on the vision of natural images: an abstract, physical parametric model for Image Quality Assessment.

Department of Information Engineering, Electronics and Telecommunications

Dottorato di Ricerca in
Information and Communication Technology – XXXIV Ciclo

Candidate

Paolo Giannitrapani
ID number 784554

Thesis Advisor

Prof. Elio D. Di Claudio

27 May 2022

Study of the subjective effects of blur on the vision of natural images: an abstract, physical parametric model for Image Quality Assessment.
Ph.D. thesis. Sapienza – University of Rome

This thesis has been typeset by LaTeX and the Sapthesis class.

Version: June 15, 2022

Author's email: paolo.giannitrapani@uniroma1.it

To my parents

Abstract

Looking at a link between blur and visual discomfort, in the present thesis, blur is viewed as a cause of a cognitive loss, and the discomfort as the immediate consequence of this loss. Among the basic cognitive functions of the Human Visual System (HVS), detection, recognition, and coarse localization functions are strongly conditioned by the individual experience. Conversely, it seems plausible that the fine localization function is committed to stabler and inter-subjective functions of the HVS. After a preliminary discussion of the operators and the ML model used (Part II), the approach presented in Part III of this thesis starts from postulating that, in the absence of vision problems, the HVS performs the fine localization of the observed objects with the best accuracy allowed by its physical macro-structure. This is a fundamental assumption, because it is known from the estimation theory that the maximum accuracy attainable when measuring the fine position of patterns in background noise is obtained by the Fisher Information about positional parameters. In fact, the Fisher Information inverse yields the minimum estimation variance. The proposed approach is based on an abstract, functional model of the Receptive Fields (RF) of the HVS, referred to as Virtual Receptive Field (VRF) and it is tuned to statistical features of natural scenes. It is a complex-valued operator, orientation-selective both in the space domain and in the spatial frequency domain. The role of the VRF model is to extract the Positional Fisher Information (PFI) as a measure of the pattern localizability loss. In the Image Quality Assessment (IQA) Full Reference (FR) environment, subjective assessments refer to the retinal image and lead to the MOS/DMOS values (Difference of Mean Opinion Score). The quality calculated by the IQA metrics is objective and refers to the image reproduced on the display. A parametric scoring function maps these metrics onto the MOS/DMOS values and depends critically on the Viewing Distance (VD) of the subject from the monitor in which the image is reproduced. When objective quality estimates for different VDs are required, as in the case of auditoria, cinemas, classrooms, a re-training procedure must be repeated for each different VDs. In the final part of this thesis (Part IV), the problem of VD is dealt with from a theoretical point of view and a model of the scoring function is defined for the case of blurred images where image degradation substantially depends on the VD. Starting from a Fisher Information loss model applied to the Gaussian distortion case in natural images, we see that the VD is estimated from the data themselves. Several maps are given with the aim of obtaining a DMOS prediction at different distances starting from the data available for a specific distance, without performing new experiments. Moreover, the theoretical results are verified on some most popular IQA FR methods and the problem of VD correction is generalized to the other distortions. Finally, the impact of isolated, long, strong, unidirectional edges on early vision is shown. As for the VD correction, an a-priori linear estimator is presented. It does not require rectification through a re-training procedure. Useful maps for detecting the position and the intensity of the PFI losses in an image are given, and the isoluminance colors allow to highlight strong and isolated edges, maintaining a constant intensity at the same edge level. We have an easy visual feedback on the images themselves to see where the greatest loss of information and the greatest discomfort due to blur are.

Contents

Part I

Background Material

Chapter 1

Introduction

1.1 Image Quality Assessment environment

Image Quality Assessment (IQA) is performed using *subjective* as well as *objective* methods.

The *subjective* quality of an image is defined as an average of the quality scores assigned to it by typical subjects, usually expressed in the MOS (Mean Opinion Score) scale. Likewise, the DMOS (Difference of Mean Opinion Score) scale measures the subjective degree of quality loss of an image with respect to its perfect quality version.

Subjective methods are unpractical whenever the image quality must be assessed in real time and repeatedly, such as in broadcasting. In these cases, algorithmic methods are employed to calculate *objective quality indices* as *estimates* or *predictions* of the subjective quality.

On the other hand, objective methods make use of a set of examples constituted by empirical MOS/DMOS ratings, expressed by a pool of volunteer subjects that visually compare a set of assorted, undistorted images with their degraded versions. These images, annotated with their empirical MOS/DMOS values, form an IQA *database*.

Several objective quality assessment methods and corresponding indices have been proposed in the last decades, suited for different operative protocols [97, 81, 19]. Methods using a complete representation of the undistorted and degraded images, are referred to as *Full Reference* (FR) IQA methods. For each set of images with artifacts to be evaluated, the FR environment requires the complete set of pristine images. Methods using incomplete representations are referred to as *Reduced Reference* (RR) methods. RR requires only some features extracted from the pristine image set. Finally, methods using only degraded images are referred to as *No Reference* (NR) methods. In this case, no pristine image sets are required for the metrics calculation. This thesis is concerned with FR methods.

In this thesis, we focus on four independent databases. The LIVE IQA Database Release 2 (LIVE DBR2) [81] reports the quality ratings of 779 distorted versions of 29 reference images (included 145 blurred images) from about 23 subjects. Ratings of subjective quality loss with respect to reference images were expressed on a DMOS scale ranging from 0 (perfect quality) to 100 (bad quality) using a *double*

stimulus strategy. The Tampere Image Database 2013 (TID2013) [75] contains 3000 distorted images, including 125 blurred images. Quality ratings were collected in five independent labs and on the internet using more than 300 subjects. They were asked to select the *best image between two distorted images* in direct comparison to the reference image. The average quality scores were expressed on a MOS scale ranging from 0 (bad quality) to 9 (perfect quality). The Computational and Subjective Image Quality Database (CSIQ) [54] contains 30 reference images and 866 distorted versions, including 150 blurred images. The database includes 5000 ratings of 25 subjects, and the ratings, obtained by *comparative ratings between different images*, are reported in DMOS units. The LIVE Multiply Distorted Image Quality Database (LIVE MD) [48] contains 15 reference images and 405 distorted images, including 45 blurred images, whose quality was rated by 37 subjects. The study was conducted using a *single stimulus with hidden reference* strategy, using DMOS scores. All databases contain *natural images*, i.e., images representing *natural scenes*, except the image i25 of the TID2013 database.

1.2 Image Quality Assessment methods

One of the best performing methods in terms of linearity and r.m.s.e. is the Visual Information Fidelity (VIF) [80, 81]. It is a natural image statistics based method and it exploits Information Theory. Belonging to the broader class of the Information Fidelity Criterion (IFC), VIF models the whole system as source (natural image), channel (which implements the distortion process) and receiver (projection of the image on the retina), and computes the mutual information between the wavelet sub-bands of the reference and distorted images. The distortion model is obtained by attenuating the source through the channel and adding Gaussian noise. VIF is defined as the ratio of the test image information and the reference image information, and the score of the metric is obtained from the sum of the mutual information of all sub-bands.

The Multi-Scale Structural Similarity Index (MS-SSIM) [95] and the Feature Similarity Index (FSIM) [109] are particular case of the wider class of the Structural Similarity Index (SSIM). SSIM compares the reference and the distorted images thorough the the mean intensity value (luminance), the standard deviation (contrast) and the structure of the images in terms of correlation, once the means are removed and the variances are normalized. These three operations are performed in the spatial domain and a sliding window operator combines them in a local quality score. The overall index is given by the mean of the local scores. To take into account the different Viewing Distance (VD) a multi-scale version (MS-SSIM) uses five scales, down-sampling the images by a factor of 2. On the other hand, FSIM uses phase congruency (similarity) and the gradient magnitude (contrast) as image features. As we will explain in the final part of this thesis, phase congruency implicitly takes into account the different VD, so, also FSIM has a coarse VD correction.

Another structural similarity based method is the Gradient Magnitude Similarity Deviation (GMSD) [108] that uses the gradient magnitude of the reference and distorted images to have a local index on the map and the standard deviation as pooling operator to obtain the overall quality index.

Still in use is the Peak Signal to Noise Ratio (PSNR), based on a logarithmic form of the Mean Squared Error (MSE). Although it represents the energy of the error signal, PSNR has a very poor correlation with perceptual image quality coming from MOS/DMOS values.

The subsequent discussion of this thesis will have experimental results that will be compared with the IQA methods presented above, but a comprehensive discussion of the IQA methods is in [2].

1.3 Framework

In the search for a simple, linear model, the first step is to choose suitable operators for the purpose (Part II). In the first part of this thesis, the entire theoretical part of the Circular Harmonic Functions (CHF) [77, 65, 6, 63] was presented and useful libraries to use these operators on images have been built. it has been observed how the Laguerre-Gauss and Hermite-Gauss transforms reproduce the behavior of the Human Vision System (HVS) well, and it was defined the Virtual Receptive Field (VRF) model with a very simple formula that maintains its shape and linearity even in the passage from space to frequency. The ML model in the Laguerre-Gauss domain and the role of Positional Fisher Information (PFI) [69] have been reproposed. Assuming, in fact, that the HVS performs the fine localization of the observed objects with the best accuracy, from estimation theory, the maximum accuracy attainable when measuring the fine position of patterns in background noise is pointed out by the Fisher Information about positional parameters and the Fisher Information inverse yields the minimum estimation variance. Therefore, this work focuses on how blur discomfort depends on the undesired loss of the PFI about observed patterns.

In Part III, a new theoretical model for blur has been defined, based on the loss of PFI, and the Scaled Blur Discomfort Index (SBDI) is defined. An experimental verification of the estimator with the databases available in the literature has been presented and an interesting application that allows us to calculate the loss of a generic Optical Transfer Function (OPT) has been provided, i.e., the spatial frequency response of the optical system, from the observed object to the retina, through an isotropic Gaussian blur, within certain working ranges (the overall OTF of a vision system, including the human eye, is a combination of the OTFs of cascaded subsystems including the OTF of correcting lenses, the OTF of an imaging system, the OTF of a display system, etc., and under the hypothesis of linearity, the overall OTF is the product of the single OTFs). Moreover, two applications are reported: the first one is about the equivalence of a sinc blur with a Gaussian blur, the second one seeks the equivalence between an astigmatic Gaussian blur and the isotropic Gaussian blur.

In Part IV, the VD and the role of the contents of the images are presented, from an early vision point of view. When we talk about linearity, we cannot ignore an in-depth study of the VD used in the experiments that led to the DMOS data (Difference of Mean Opinion Score, the ratings of subjective quality loss with respect to reference images). In some IQA methods, the images are re-scaled and filtered to emulate different VDs, and the corresponding metrics are combined with the criterion of optimizing the performance, but in the theoretical model, however, they

lose linearity. As a practical application, a DMOS map has been provided, to allow to calculate new DMOS values for the same blur level by changing the distance parameter.

In the theoretical calculation of the SBDI estimator we always measure the Fisher information loss, for each detail or cluster of details. The difference with real cases is that the eye does not always see every object (isolated gradients that are too small and not identifiable) or the displacement due to the loss of the Fisher Information is so complex that noise is replaced by other noise (this is the case of textures) and, therefore, our metric is too precise compared to what actually happens in the HVS. In order to overcome this limit, a correction of the theoretical estimator has been proposed, taking into account mainly the strong and isolated gradients identified starting from a threshold that distinguishes them from textures and weak edges, which we will call *natural vision threshold*, with reference to the index found in Part III in the study of natural images affected by blur. The performances of this new estimator, which always maintains a linearity of the model, are presented and compared with the existing estimators in the literature that require neural rectification. A very useful application for detecting the position and intensity of the losses in an image has been provided: a *certainty map* that allows us to have immediate visual feedback of the discomfort due to blur, highlighting the most critical portions in degraded images.

Chapter 2

Blur Perception and Viewing Distance

2.1 Blur and Visual Discomfort

Among the various sources of non-clinical visual discomfort [20, 38], blur caused by refractive errors is perhaps the most common one.

The discomfort associated with blur is often explained as a consequence of the concentration of the spatial energy spectrum of the perceived image into some bands, or as a byproduct of the discrepancy of this spectrum from the expected spectrum of natural images [70, 106]. Alternative explanations addressed the mismatch of the spatial patterns with the expected ones [50, 99, 4]. An in-depth account of previous studies and mathematical models about the blur phenomenon is provided in [102].

Looking at a possible physical source of blur discomfort, three hypotheses are examined in [71]. The first hypothesis is that discomfort is stimulated by the weak response of the accommodation system. A second hypothesis, somewhat related to the first one, is that discomfort arises because the "micro-fluctuations" observed in the accommodation feedback signal become ineffective [17, 67, 59, 18]. A third hypothesis maintains that, when an image is correctly projected onto the retina, the receptive fields produce a parsimonious, *sparse* representation of this image [66]. The spatial spread caused by blur excites more receptors, producing a metabolic overload [49].

In the present approach, blur is viewed as a cause of a cognitive loss, and the discomfort as the immediate consequence of this loss. It is argued that, among the basic cognitive functions of the Human Visual System (HVS), detection, recognition, and coarse localization functions are strongly conditioned by the individual experience. Conversely, it seems plausible that the fine localization function is committed to stabler and inter-subjective functions of the HVS.

Based on the above consideration, the present approach starts from postulating that, under normal conditions, the HVS performs the fine localization of the observed objects with the best accuracy allowed by its physical macro-structure, given the characteristics of the environment and man's interaction.

This assumption is fundamental, because it is known from the estimation theory that the maximum accuracy attainable when measuring the fine position of patterns in

background noise is deduced from the Fisher Information about positional parameters. In fact, the Fisher Information inverse yields the minimum estimation variance [91].

Therefore, the focus here is on how the discomfort of blurring depends on the unwanted loss of PFI on observed patterns.

The present cognitive approach is agnostic as to whether the discomfort is related to accommodation frustration or metabolic unbalances. On the other hand, it is compliant with the fact that blur discomfort concerns the regions of visual interest [87] and that blur is not always undesired or detrimental [85]. For instance, blur is sometimes a wanted effect in photography and microscopy.

Previous analyses of the blur perception phenomenon were mainly oriented to the study of the visual acuity, employing specific stimuli localized either in space, such as edges, lines, crosses, or in the spatial frequency domain, such as sinusoidal gratings, or even in both domains, such as Gabor wavelets [103].

The model presented here is oriented to the evaluation of the visual impact of blur in the vision of natural scenes. To this purpose, a generic image projected on the retina is viewed as an element of the *random set of natural images*, characterized by stable statistical features.

The proposed approach is based on an *abstract*, *functional* model of the receptive fields (RF) of the HVS, allowing for a direct computation of the PFI.

For analytical convenience, blur is modeled as a Gaussian shaped isotropic blur. Then, the analysis is generalized to other types of blur, invoking a criterion of informational equivalence with respect to isotropic Gaussian blur under the PFI paradigm.

In Part III, the model-based discomfort predictions were first compared to empirical data about the *subjective* quality loss of blurred images, which is argued to be strictly related to blur visual discomfort. These data are available in organized databases containing the results of experimental sessions finalized to IQA, conducted for multimedia industry purposes [97, 10, 39]. Moreover, the model is applied to blurred images annotated with ratings of visual discomfort. The results of these experiments confirm the validity of the approach.

2.2 Full Reference environment and Viewing Distance

The typical scheme of IQA methods first consists of the local extraction of some features suggested by similarity criteria or visual information loss among reference and distorted images. This stage is followed by a pooling over the entire image of the selected features, whose result is usually referred to as IQA *metric*.

The final stage is the application to the metric of a *parametric scoring function* that map the metric values onto the MOS/DMOS scores.

The parameters are determined by interpolating the metric on the empirical MOS/DMOS data contained in the IQA databases. The VQEG [93, 94] suggests a logistic function using from three to five parameters. In [42] the scoring function is a S-curve function regulated by three parameters.

The scoring function first depends on the database images, which are representative of the images of interest for a specific application. Most popular databases contain *natural images*, i.e., images captured by optical cameras, usually experienced

by subjects in their everyday life.

For a given images set, the scoring function depends critically on the experimental settings and protocols employed in a database, and, specifically, on the VD [55]. In fact, subjective MOS/DMOS scores are obtained by the images perceived by the *retina*, which vary with the VD, while the objective quality is calculated from the images reproduced on the *display*.

Unfortunately, the existing scoring functions lack an interpretation in terms of the VD. By consequence, whenever the VD employed in an application differs from the one used in experiments, it is necessary to modify the scoring function. So, if objective quality estimates for different VDs are required, as in the case of auditoria, cinemas, classrooms, etc., the re-training procedure must be repeated for each different VDs.

In some IQA methods, namely the Multi-Scale Structural Similarity (MS-SSIM) [98] and the Feature Similarity Index (FSIM) [109] methods, the dependence of their scoring function on the VD is attenuated by mixing a set of metrics calculated at different resolutions and encompassing the VD of addressed databases. Extending this approach, in [35] a preprocessing stage is applied to different IQA methods. Images are re-scaled and filtered to emulate different VDs, and the corresponding metrics are combined with the criterion of optimizing the statistical performance.

All these methods do not elude the crucial necessity of specific experiments for the blind determination of the right VD parameters. Noticeably, the VD has been recently included as a learnable parameter in a CNN based NR IQA method [10].

In this work, following a quite opposite approach, the problem is tackled using a theoretical viewpoint, based on principles and reasonable hypotheses. The essential scope is to define an effective model of the scoring function explicitly related to the VD.

To this purpose, the case of *blurred images* is adopted as a significant *benchmark*. In fact, in this case, the image degradation substantially depends on the VD, being scarcely influenced by additional environmental factors, such as room illumination or display contrast ratio. In addition, modeling the effects of blur is mathematically feasible, and relatively easy in the case of natural images.

The analysis conducted here stems from the overarching principle that the Human Vision System (HVS) is an optimal device for fine pattern localization, given the eye macro-structure. This feature is especially important for stereoscopic vision. The outstanding consequence, drawn from information theory, is that the variance of fine localization of structures in observed images is objective, and calculable as a function of the degree of blur. It is approximated by the inverse of the Fisher Information about the position of observed patterns.

Based on this principle and on well-known spectral properties of *natural* images, the scoring function mapping the degree of *Gaussian blur* to the subjective quality loss can be expressed in a closed form, where it was ascertained that this function fits well empirical DMOS data provided of independent, well known IQA databases.

In Part IV, an essential account of these theoretical results is provided, with some extensions about the role of the VD. One remarkable feature of the theoretical model, referred to as the *canonical model* for IQA FR of blurred images, is that the VD is estimated from the data themselves. Moreover, the applicability of the theoretical results to some most popular IQA FR methods is verified, with the scope

of applying them to the DMOS prediction at various distances starting from the data available for a specific distance, without performing new experiments.

Chapter 3

From Gabor Functions to Harmonic Angular Functions

The discovery of spatial frequency and orientation tuned *channels* started the modern spatial vision [13, 14] and, later, the existence of these channels was confirmed by different indipendent studies, including signal mixture and adaptation experiments [7, 34]. The postulate of independent spatial frequency and orientation channels allows predicting detection thresholds for any signal pattern from the knowledge of the Fourier spectrum of the stimulus and the sensitivity to single sinusoidal gratings of different frequencies, i.e., the contrast sensitivity function. In the image processing, an example of operator is the log-Gabor, which corresponds to a Gaussian shape in log-frequency and in orientation. A log-Gabor is directly and completely defined by its preferred spatial frequency and orientation and its bandwidth in each dimension. In general, Gabor-filters are maximally localized jointly in space and frequency, have a monotonically and smoothly decreasing response for frequencies and orientations moving away from the preferred parameters, and no response to uniform fields. These are all desirable properties for a subband decomposition. Linear filters are also tuned to the phase of the stimulus as simple cells in primary visual cortex are [22], even if, psychophysical performance seems not to depend on absolute phase [79].

In the field of the 2-dimensional anisotropic visual filters, a two-dimensional Gabor representation is an attractive framework for a unified theory of visual cortical representation, because such filters simultaneously capture all the fundamental properties of linear neural receptive fields in the visual cortex: spatial localization, spatial frequency selectivity, and orientation selectivity. In particular, an optimal two-dimensional Gabor filter can achieve very sharp spatial resolution but poor spectral resolution in one orientation, while simultaneously achieving very sharp spectral resolution but poor spatial resolution in the perpendicular orientation. Thus, none of these *standard* models, when quantified, have polar separable two-dimensional spectra; rather, these models (including generalized two-dimensional Gabor filters) possess different orientation characteristics at different spatial frequencies and vice versa. Any simple concatenation of any number of spatial differential operators in a (neural) filter hierarchy which begins with an isotropic two-dimensional filter as its first stage can only produce two-dimensional filters whose orientation selectivity

and spatial frequency selectivity are independent (i.e., filters with polar separable two-dimensional spectra). A neurophysiologically relevant corollary is that such filters with polar separable two-dimensional spectra would have the same orientation tuning curve for single thin bars as for sine wave gratings [23].

The Harmonic Angular Functions (HAFs) are polar-separable, and they have a polar-separable two-dimensional Fourier transform. Compared to the two-dimensional Gabor functions, they possess the same *shape orientation tuning curve for sine wave gratings as for thin bar* (property 6 of [23]) allowing for two-dimensional analysis both in space and in the spatial frequency domain [22], without involving the use multiple channels, as usual [79]. The HAFs were introduced for digital image analysis in [44, 43] leading to the definition of the Complex Harmonic Wavelets [46].

Part II

Orthogonal Operators and ML Model for Image Processing

Chapter 4

Object decomposition into weighted sums of orthogonal basis functions

4.1 Normalized Hermite Functions and Cartesian Shapelets Basis Functions

High precision image processing and analysis is a fundamental requirement in many fields of applications. Circular Harmonic Functions (CHF) are widely used in the analysis of a wide range of textures, recreating synthetic objects that match very well with the targets. It has also been shown that they reproduce the behavior of the Human Visual System (HVS) with applications such as restoring blurry moving text images, blurry sharp images, and blurry natural images. [15, 72].

We now refer to a linear decomposition of the object observed in a series of basis functions with a set of weighted Hermite polynomials corresponding to a distorted gaussians [77]. Unlike the wavelet transform which decomposes the image into a sum of basis functions of different scales but with a set shape, the set of Hermite polynomials decompose the image into a series of compact disjoint objects of arbitrary shapes. The normalized Hermite functions ϕ_n have the orthogonality property:

$$\int_{-\infty}^{+\infty} \phi_n(x)\,\phi_m(x)dx = \delta_{mn} \tag{4.1}$$

where δ_{mn} is the usual Kronecker δ, which is one when $m = n$ and zero otherwise. ϕ_n can be thought of as shape perturbations around the Gaussian $\phi_0(x)$, so the Shapelets dimensionless basis functions are defined as distorted Gaussians:

$$\phi_n(x) \equiv \left[2^n \pi^{1/2} n!\right]^{-1/2} H_n(x)\, e^{-x^2/2} \tag{4.2}$$

where $H_n(x)$ is the Hermite polynomials of order n, with $n \geq 0$:

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2} \tag{4.3}$$

The first few terms of the Hermite polynomials are:

$$\begin{aligned} H_0(x) &= 1 \\ H_1(x) &= 2x \\ H_2(x) &= 4x^2 - 2 \\ H_3(x) &= 8x^3 - 12x \\ H_4(x) &= 16x^4 - 48x^2 + 12 \end{aligned} \tag{4.4}$$

with the useful recurrences and derivatives properties:

$$\begin{aligned} H_{n+1}(x) &= 2xH_n(x) - 2nH_{n-1}(x) \\ H_n'(x) &= 2nH_{n-1}(x) \\ H_n(x) &= (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2} = e^{x^2/2} \left(x - \frac{d}{dx}\right) e^{-x^2/2} \end{aligned} \tag{4.5}$$

In Fig. 4.1 are plotted the first terms of the Hermite polynomials.

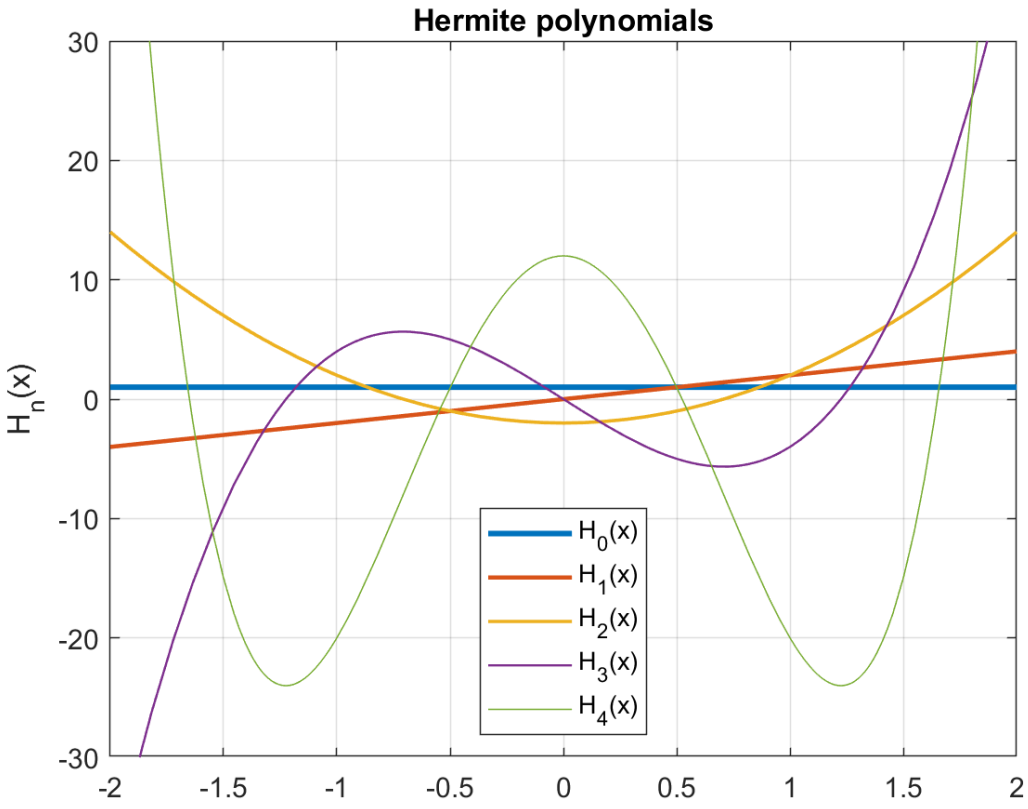

Figure 4.1. The first five terms of the Hermite polynomials.

The Hermite functions can be efficiently computed using a three-term recursion

Table 4.1. Normalization factor $d_n = \pi^{1/4} 2^{n/2} \sqrt{n!}$ increasing n.

n	d_n	$1/d_n$
0	1.33	0.75
1	1.88	0.53
2	3.77	0.27
3	9.22	0.11
4	26.09	0.04

relation from two starting values:

$$\begin{aligned} \phi_0(x) &\equiv \pi^{-1/4} e^{-(1/2)x^2} \\ \phi_1(x) &\equiv \pi^{-1/4} \sqrt{2} x e^{-(1/2)x^2} \\ \phi_{n+1}(x) &\equiv \sqrt{\frac{2}{n+1}} x \phi_n(x) - \sqrt{\frac{n}{n+1}} \phi_{n-1}(x) \end{aligned} \tag{4.6}$$

The unnormalized Hermite polynomials $H_n(x)$ are ill-conditioned in numerical applications but have coefficients which are integers. The relationship between normalized and unnormalized Hermite functions is given in Tab.4.1.

Recall the first five expansion terms:

$$\begin{aligned} \phi_0(x) &\equiv \pi^{-1/4} e^{-(1/2)x^2} \\ \phi_1(x) &\equiv \frac{\sqrt{2}}{\pi^{1/4}} x e^{-(1/2)x^2} \\ \phi_2(x) &\equiv \left[\sqrt{2}x^2 - \frac{1}{\sqrt{2}}\right] \pi^{-1/4} e^{-(1/2)x^2} \\ \phi_3(x) &\equiv \left[\sqrt{4/3}x^3 - \sqrt{3}x\right] \pi^{-1/4} e^{-(1/2)x^2} \\ \phi_4(x) &\equiv \left[\sqrt{2/3}x^4 - \sqrt{6}x^2 + \sqrt{6}/4\right] \pi^{-1/4} e^{-(1/2)x^2} \end{aligned} \tag{4.7}$$

The Hermite functions satisfy recursion relations which are very useful in applications such as

$$x\phi_n(x) = \sqrt{(n+1)/2}\,\phi_{n+1}(x) + \sqrt{n/2}\,\phi_{n-1}(x) \tag{4.8}$$

In Fig. 4.2 are plotted the first terms of the Shapelets polynomials.

So, in a more general form, $\phi_n(x)$ are rescaled to the dimensional basis functions:

$$B_n(x;\beta) \equiv \beta^{-1/2} \phi_n\left(\beta^{-1} x\right) \tag{4.9}$$

where β is the scale close to the size of the object. Applying these functions on the one-dimensional signal $S(x)$, it can be expressed as:

$$S(x) = \sum_n S_n B_n(x;\beta) \tag{4.10}$$

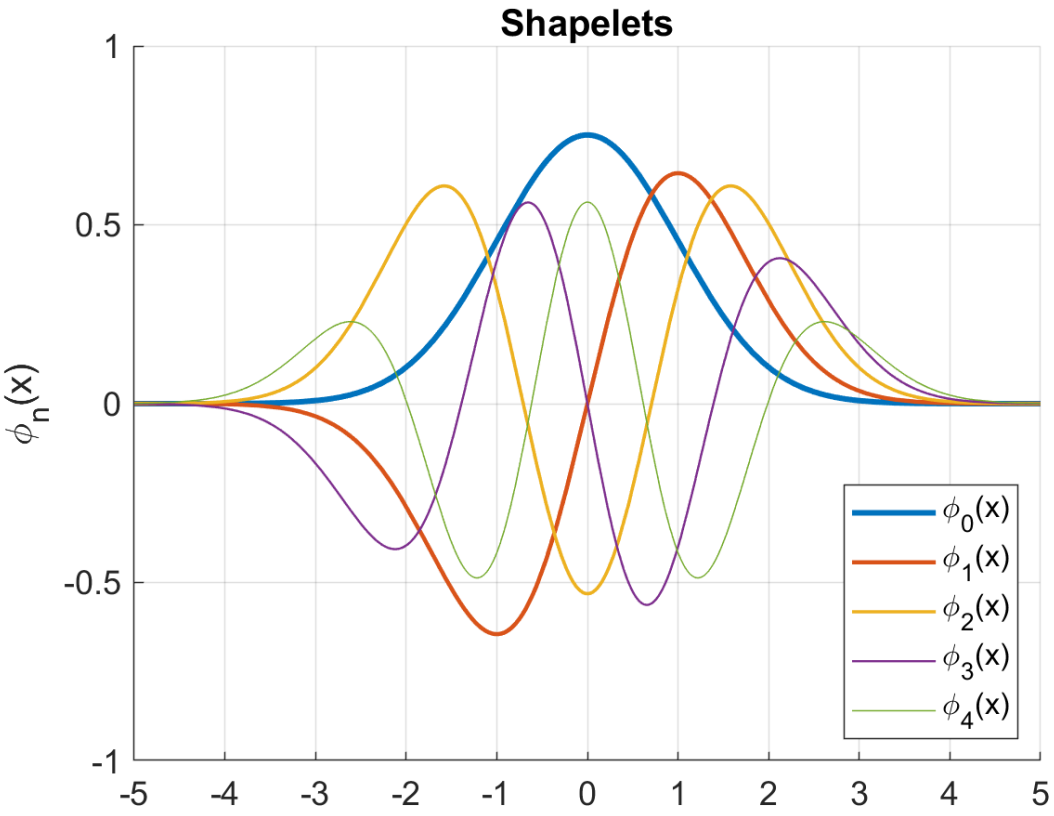

Figure 4.2. The first five terms of the Laguerre polynomials.

Also $B_n(x;\beta)$ function has the orthonormality property:

$$\int_{-\infty}^{+\infty} B_n(x;\beta)\, B_m(x;\beta)dx = \delta_{mn} \tag{4.11}$$

Thus, we can solve the coefficients as follow:

$$S_n = \int_{-\infty}^{+\infty} S(x)\, B_n(x;\beta)dx \tag{4.12}$$

The series $S(x)$ converges quickly if the object is enough compact and if β and the origin are respectively close to the size and the location of the object. The dimensional basis functions $B_n(x;\beta)$ increase the oscillation extension with the order n and decrease the scale factor, maintaining β constant.

From a computational point of view, looking at 4.8, the first derivative can be computed by the following recurrence:

$$\frac{d\phi_n(x)}{dx} = -\sqrt{\frac{n+1}{2}}\phi_{n+1}(x) + \sqrt{\frac{n}{2}}\phi_{n-1}(x) \tag{4.13}$$

or using this alternative form:

$$\frac{d\phi_n(x)}{dx} = -x\phi_n(x) + \sqrt{2n}\phi_{n-1}(x) \tag{4.14}$$

both with $n \geq 0$, using as initialization:

$$\frac{d\phi_0(x)}{dx} = -\frac{1}{\sqrt{2}}\phi_1(x) = -x\phi_0(x) = -x\pi^{-1/4}e^{-(1/2)x^2} \tag{4.15}$$

Anyway, in practical application is preferred the following explicit formula for the derivative of $\phi_1(x)$:

$$\frac{d\phi_1(x)}{dx} = \sqrt{2}\left[1-x^2\right]\phi_0(x) = \sqrt{2}\pi^{-1/4}e^{-(1/2)x^2}\left[1-x^2\right] \tag{4.16}$$

The one-dimensional formalism is easily extended to image domain. Each two-dimensional basis function is the tensor product of two one-dimensional basis functions. Defining the dimensionless functions:

$$\phi_{\mathbf{n}}(\mathbf{x}) \equiv \phi_{n_1}(x_1)\,\phi_{n_2}(x_2) \tag{4.17}$$

where $\mathbf{x} = (x1, x2)$ and $\mathbf{n} = (n1, n2)$. Considering again of as perturbations around the 2-dimensional Gaussian $\phi_{00}(x)$, the two-dimensional basis function become:

$$B_{n_{x_1}n_{x_2}}(x_1; x_2; \beta) = B_{n_{x_1}}(x_1; \beta)\,B_{n_{x_2}}(x_2; \beta) \tag{4.18}$$

It is possible to prove that they maintain the orthogonality property:

$$\int_{-\infty}^{+\infty} B_{\mathbf{n}}(\mathbf{x}; \beta)\,B_{\mathbf{m}}(\mathbf{x}; \beta)d^2x = \delta_{m_1n_1}\delta_{m_2n_2} \tag{4.19}$$

Finally, the image $I(\mathbf{x})$ can be decomposed in this orthonormal, complete basis as follows:

$$I(\mathbf{x}) = \sum_{n_1,n_2=0}^{+\infty} I_{\mathbf{n}}B_{\mathbf{n}}(\mathbf{x}; \beta) \tag{4.20}$$

with coefficients:

$$I_{\mathbf{n}} = \int_{-\infty}^{+\infty} I(\mathbf{x})\,B_{\mathbf{n}}(\mathbf{x}; \beta)d^2x \tag{4.21}$$

In Fig. 4.3 are plotted the overall 3D plots of the Hermite polynomials.

4.2 Quantum Harmonic Oscillator approach

Recall the previous formula for the Shapelets dimensionless basis functions:

$$\phi_n(x) \equiv \left[2^n\pi^{1/2}n!\right]^{-1/2} H_n(x)\,e^{-x^2/2} \tag{4.22}$$

We see that these are also the eigenfunctions of the Quantum Harmonic Oscillator (QHO) [77], so for our calculus we can use the formalism of the Quantum Mechanics. The Fourier transform follow from the succeeding considerations. The dimensionless 1D equation of the QHO looks like:

$$\hat{H}\Phi = \left(\hat{x}^2 + \hat{p}^2\right)\Phi = E\Phi \tag{4.23}$$

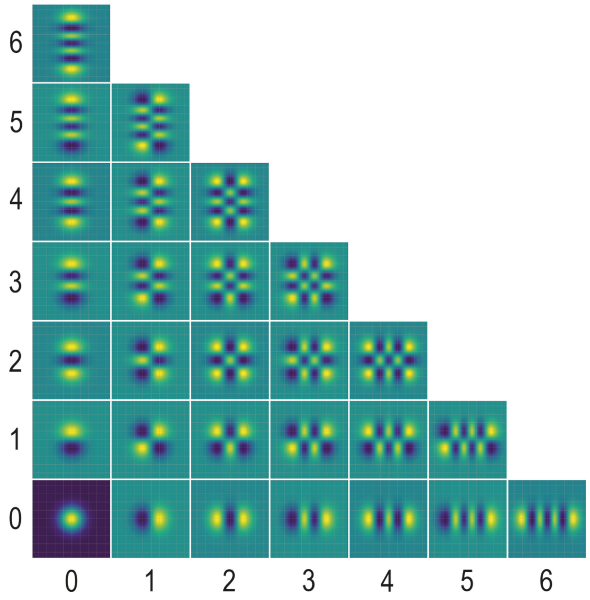

Figure 4.3. The overall 3D plots of the Hermite polynomials.

In position space, the base kets used are the position kets $\hat{x}|\ x'\rangle = x'|\ x'\rangle$, which has the orthonormal property $\langle x''|x'\rangle = \delta(x'' - x')$. Considering the kets basis $|\ \alpha\rangle = \int dx'|\ x'\rangle\langle x'|\alpha\rangle$, we have for any physical states $\langle\beta|\alpha\rangle = \int dx'\phi^*_\beta(x')\phi_\alpha(x')$, where $\phi_\alpha(x')$ is equal to $\langle x'|\alpha\rangle$. Similar consideration for momentum space, with the base kets $\hat{p}|\ p'\rangle = p'|\ p'\rangle$, the orthonormal property $\langle p''|p'\rangle = \delta(p'' - p')$ and the expansion in the base kets $|\ \alpha\rangle = \int dp'|\ p'\rangle\langle p'|\alpha\rangle$. At this point, using momentum as generator of translations:

$$\hat{p}|\ \alpha\rangle = \int dx'|\ x'\rangle\left(-i\hbar\frac{\partial}{\partial x'}\langle x'|\alpha\rangle\right) \tag{4.24}$$

$$\langle x'|\hat{p}|\alpha\rangle = -i\hbar\frac{\partial}{\partial x'}\langle x'|\alpha\rangle \tag{4.25}$$

(we avoid the proof) and consequently:

$$\langle x'|\hat{p}|p'\rangle = p'\langle x'|p'\rangle = -i\hbar\frac{\partial}{\partial x'}\langle x'|p'\rangle \tag{4.26}$$

and finally, we have the relation between position and momentum space:

$$\langle x'|p'\rangle = Ae^{\frac{ip'x'}{\hbar}} \tag{4.27}$$

where the normalization imposes that the term A is equal to $\frac{1}{\sqrt{2\pi}}$ and hence:

$$|\ \Phi\rangle = \int dp|\ p\rangle\langle p|\Phi\rangle = \int dp|\ p\rangle\phi(p) \tag{4.28}$$

$$\phi(x) = \langle x|\Phi\rangle = \int dp\langle x|p\rangle\phi(p) = \int dp\frac{1}{\sqrt{2\pi}}e^{\frac{ip'x'}{\hbar}}\phi(p) \tag{4.29}$$

So, from the symmetry between x and p, we deduce that the Fourier transform of the basis function should be, up to a multiplicative constant, equal to the same basis function and we have:

$$\widetilde{\phi}(k) = i^n \phi_n(x) \tag{4.30}$$

In the image case (bidimensional), the Fourier transform of an image consists of the components of a vector with shapelets coefficients and a scale parameter β, multiplied with factors of i, and with scale $1/\beta$.

Looking at the recursive formulas 4.13-4.16, the similarity of the results of the operations of differentiation and of multiplication by x make it possible to define so-called *raising* and *lowering* operators. The motive for these names is self-explanatory in that the raising operator $\mathfrak{R}$, when applied to a Hermite function of degree n, gives a result which is proportional to the Hermite function of the next highest degree. Similarly, the lowering operator $\mathcal{L}$ reduces the degree of a Hermite function by one. For the raising operator we have:

$$\mathfrak{R} \equiv \left(\frac{d}{dx} - x\right) \tag{4.31}$$

$$\mathfrak{R}\phi_n = -\sqrt{2(n+1)}\phi_{n+1} \tag{4.32}$$

and for the lowering operator:

$$\mathcal{L} \equiv \left(\frac{d}{dx} + x\right) \tag{4.33}$$

$$\mathcal{L}\phi_n = \sqrt{2n}\phi_{n-1} \tag{4.34}$$

Thus, the Hermite eigenoperator

$$\mathfrak{H} \equiv \left(\frac{d^2}{dx^2} - x^2\right) \tag{4.35}$$

$$\mathfrak{H}\phi_n = -(2n+1)\phi_n \tag{4.36}$$

can be written in terms of the raising and lower operators as

$$\mathfrak{H} \equiv \frac{1}{2}(\mathfrak{R}\mathcal{L} + \mathcal{L}\mathfrak{R}) \tag{4.37}$$

The operations of differentiation and multiplication by x can also be expressed in terms of the raising and lowering operators:

$$x\phi_n = \frac{1}{2}(\mathcal{L} - \mathfrak{R})\phi_n = \sqrt{n/2}\phi_{n-1} + \sqrt{(n+1)/2}\phi_{n+1} \tag{4.38}$$

$$\frac{d}{dx}\phi_n = \frac{1}{2}(\mathcal{L} + \mathfrak{R}) = \sqrt{n/2}\phi_{n-1} - \sqrt{(n+1)/2}\phi_{n+1} \tag{4.39}$$

For our practical interest, we can use the lowering and raising operators defined as:

$$\hat{a} \equiv \frac{1}{\sqrt{2}}(\hat{x} + i\hat{p}) \tag{4.40}$$

$$\hat{a}^\dagger \equiv \frac{1}{\sqrt{2}}\left(\hat{x} - i\hat{p}\right) \tag{4.41}$$

They commute as $\left[\hat{a}, \hat{a}^\dagger\right] = 1$ and are related to the basis function as $\hat{a}\phi_n = \sqrt{n}\phi_{n-1}$, $\hat{a}^\dagger\phi_n = \sqrt{n+1}\phi_{n+1}$. The Hamiltonian can then be rewritten as $\hat{H} = \hat{N} + 1/2$, where the number operator $\hat{N} \equiv \hat{a}^\dagger\hat{a}$ has the property that $\hat{N}\phi_n = n\phi_n$. The dimensional basis functions are the eigenfunctions of the Hamiltonian:

$$\hat{H}_\beta = \frac{1}{2}\left[\beta^{-2}\hat{x}^2 + \beta^2\hat{p}^2\right] \tag{4.42}$$

With the bra-ket notation, the dimensional basis functions are then given by

$$B_n\left(x;\beta\right) = \langle x|n;\beta\rangle \tag{4.43}$$

since $\hat{H}_\beta\,|n;\beta\rangle = (n + \frac{1}{2})\,|n;\beta\rangle$.

4.3 From cartesian to polar basis functions – A first evidence

Exploiting the analogy with Quantum Mechanics, we give a first proof of the transition from a Cartesian base to a polar base [77]. The Cartesian basis functions discussed above are separable in the Cartesian coordinates x_1 and x_2. It is also useful to construct basis functions which are separable in the polar coordinates x and φ. These are eigenstates of the Hamiltonian $\hat{H}$ and of the angular momentum $\hat{L}$ simultaneously. Defining the left and right lowering operators as

$$\hat{a}_l = \frac{1}{\sqrt{2}}\left(\hat{a}_1 + i\hat{a}_2\right) \tag{4.44}$$

$$\hat{a}_r = \frac{1}{\sqrt{2}}\left(\hat{a}_1 - i\hat{a}_2\right) \tag{4.45}$$

We have the non-vanishing commutators between these operators if:

$$\left[\hat{a}_l, \hat{a}_l^\dagger\right] = \left[\hat{a}_r, \hat{a}_r^\dagger\right] = 1 \tag{4.46}$$

where $\hat{a}_l^\dagger$ and $\hat{a}_r^\dagger$ are the raising operators. In terms of raising and lowering operators, the Hamiltonian and the angular momentum operators

$$\hat{H} = \frac{1}{2}\left[\hat{x}_1^2 + \hat{x}_2^2 + \hat{p}_1^2 + \hat{p}_2^2\right] \tag{4.47}$$

$$\hat{L} = \hat{x}_1\hat{p}_2 - \hat{x}_2\hat{p}_1 = i\left(\hat{a}_1\hat{a}_2^\dagger - \hat{a}_1^\dagger\hat{a}_2\right) \tag{4.48}$$

can be written as

$$\hat{H} = \hat{N}_r + \hat{N}_l + 1 \tag{4.49}$$

$$\hat{L} = \hat{N}_r - \hat{N}_l \tag{4.50}$$

where $\hat{N}_l = \hat{a}_l^\dagger\hat{a}_l$, $\hat{N}_r = \hat{a}_r^\dagger\hat{a}_r$, so the operators $\hat{a}_l^\dagger$, $\hat{a}_r^\dagger$, $\hat{a}_l$, and $\hat{a}_r$ can thus be thought of as creating and destroying left- and right-handed quanta. Since the operators

$\hat{N}_r$ and $\hat{N}_l$ form a complete set of commuting observables, their eigenstates $\mid n_l, n_r\rangle$ provide a complete basis for our function space. By applying the raising operators several times on the ground state, we can build the quantum states:

$$\mid n_r, n_l\rangle = \frac{(\hat{a}_r^\dagger)^{n_r}(\hat{a}_l^\dagger)^{n_l}}{\sqrt{n_r! n_l!}} \mid 0, 0\rangle \tag{4.51}$$

and it is easy to verify that

$$\hat{H}\,|\,n_l, n_r\rangle = (n_r + n_l)|\; n_l, n_r\rangle \tag{4.52}$$

$$\hat{L}\,|\,n_l, n_r\rangle = (n_r - n_l)|\; n_l, n_r\rangle \tag{4.53}$$

then, we can recall the states in terms of their energy and angular momentum numbers, $n = n_r + n_l$ and $m = n_r - n_l$ as

$$\mid n, m\rangle = \mid n_l = \tfrac{1}{2}(n-m), n_r = \tfrac{1}{2}(n+m)\rangle \tag{4.54}$$

The angular momentum quantum number takes on the $n+1$ values given by $m = -n,\ -n+2,\ \ldots,\ n-2,\ n$. Using the x-representation for $\hat{a}_l^\dagger$ and $\hat{a}_r^\dagger$, we can see that the basis function $\chi_{n_l,n_r}(x,\varphi) \equiv \langle x|n_l, n_r\rangle$ for the angular momentum states are given by

$$\chi_{n_l,n_r}(x,\varphi) = [\pi n_l! n_r!]^{-1/2}\, H_{n_l,n_r}(x)\, e^{-x^2/2} e^{i(n_r - n_l)\varphi} \tag{4.55}$$

where $H_{n_l,n_r}(x)$ are the polar Hermite polynomials. The useful recursive function can be used to compute them as follow:

$$\frac{l-k}{x} H_{k,l}(x) = l H_{k,l-1}(x) - k H_{k-1,l}(x) \tag{4.56}$$

initialized with $H_{0,0}(x) = 1$, with the diagonal polynomials computed as

$$H_{k,k}(x) = H_{k+1,k-1}(x) - x^{-1} H_{k,k-1}(x) \tag{4.57}$$

They are symmetric, i.e., $H_{k,l}(x) = H_{l,k}(x)$, and their derivative obey

$$H'_{k,l}(x) = k H_{k-1,l}(x) + l H_{k,l-1}(x) = 2x H_{k,l}(x) - H_{k+1,l}(x) - H_{k,l+1}(x) \tag{4.58}$$

Following, the first few terms of the polar Hermite polynomials:

$$\begin{aligned}
&H_{0,0}(x) = 1\\
&H_{0,1}(x) = H_{1,0}(x) = x\\
&H_{0,2}(x) = H_{2,0}(x) = x^2\\
&H_{1,1}(x) = x^2 - 1\\
&H_{0,3}(x) = H_{3,0}(x) = x^3\\
&H_{1,2}(x) = H_{2,1}(x) = x^3 - 2x\\
&H_{0,4}(x) = H_{4,0}(x) = x^4\\
&H_{1,3}(x) = H_{3,1}(x) = x^4 - 3x^2\\
&H_{2,2}(x) = x^4 - 4x^2 + 2
\end{aligned} \tag{4.59}$$

Dimensional polar basis functions can be constructed as

$$A_{n_l,n_r}(x,\varphi;\beta) = \beta^{-1}\chi_{n_l,n_r}\left(\beta^{-1}x,\varphi\right) \tag{4.60}$$

and it preserves the orthonormal property:

$$\begin{gathered}\int_0^{2\pi} d\varphi \int_0^{+\infty} dx x A_{n_l,n_r}(x,\varphi;\beta)\, A_{n_l\prime,n_r\prime}(x,\varphi;\beta) = \\ \langle n_l,n_r;\beta|n_l\prime,n_r\prime;\beta\rangle = \delta_{n_l,n_l\prime}\delta_{n_r,n_r\prime}\end{gathered} \tag{4.61}$$

From the equations 4.44, 4.45 and 4.51 we can obtain the following relation between cartesian and polar bases:

$$\begin{gathered}\langle n_1,n_2|n_l,n_r\rangle = 2^{-(1/2)(n_r+n_l)} i^{n_r-n_l}\left[\frac{n_1!n_2!}{n_r!n_l!}\right]^{1/2}\delta_{n_1+n_2,n_r+n_l}\times \\ \times\sum_{n_r\prime=0}^{n_r}\sum_{n_l\prime=0}^{n_l} i^{n_l\prime-n_r\prime}\binom{n_r}{n_r\prime}\binom{n_l}{n_l\prime}\delta_{n_r\prime+n_l\prime,n_1}\end{gathered} \tag{4.62}$$

and the result is that only the states with $n_1+n_2=n_r+n_l$ are mixed.The first angular momentum states $|\, n,m\rangle$ in cartesian states are:

$$\begin{aligned}&|n=0\ ,\ m=0\rangle = |0\ ,\ 0\rangle \\ &|n=1\ ,\ m=1\rangle = \frac{1}{\sqrt{2}}\left(|1\ ,\ 0\rangle + i\,|0\ ,\ 1\rangle\right) \\ &|n=1\ ,\ m=-1\rangle = \frac{1}{\sqrt{2}}\left(|1\ ,\ 0\rangle - i\,|0\ ,\ 1\rangle\right) \\ &|n=2\ ,\ m=2\rangle = \frac{1}{2}\left(|2\ ,\ 0\rangle + i\sqrt{2}\,|1\ ,\ 1\rangle - |0\ ,\ 2\rangle\right) \\ &|n=2\ ,\ m=0\rangle = \frac{1}{\sqrt{2}}\left(|2\ ,\ 0\rangle + |0\ ,\ 2\rangle\right) \\ &|n=2\ ,\ m=-2\rangle = \frac{1}{2}\left(|2\ ,\ 0\rangle - i\sqrt{2}\,|1\ ,\ 1\rangle - |0\ ,\ 2\rangle\right)\end{aligned} \tag{4.63}$$

Because the polar shapelet states are eigenstates of the angular momentum, they have simple rotational properties. Indeed, under a finite rotation by an angle ρ, the polar states transform as

$$|\, n_r,n_l\rangle\prime = e^{-i\rho\hat{L}}|\, n_r,n_l\rangle = e^{-i\rho(n_r-n_l)}|\, n_r,n_l\rangle \tag{4.64}$$

In this basis, finite rotations thus correspond only to a phase factor. It is therefore a simple matter to rotate an arbitrary function $f(x)$. First, we decompose it into polar shapelet coefficients $f_{n_r,n_l} = \langle n_r,n_l;\beta|f\rangle$, with an appropriate shapelet scale β. The coefficients of the rotated function are given by

$$f'_{n_r,n_l} = e^{-i\rho(n_r-n_l)} f_{n_r,n_l} \tag{4.65}$$

By contrast, operating a finite rotation in the cartesian basis requires an infinite number of applications of the $\hat{R} = -i\hat{L}$ operator. On the other hand, convolutions do not have simple analytical expressions in the polar basis, as they do in the cartesian basis.

4.4 Fourier Transform of the Hermite Functions

The Hermite functions are (to within a factor of i^n) their own Fourier transform. Consequently, one can sometimes easily obtain the Hermite series of a function $f(y)$ by calculating the Hermite expansion of its Fourier transform $F(k)$ and then multiplying the coefficients by i^n. A function that is a truncated series of Hermite functions can be expressed as:

$$f_N(x) = \sum_{j=0}^{N} a_j \phi_j(x) \tag{4.66}$$

where $\phi_j(x) \equiv H_j(x) e^{-x^2/2}$. If we define the forward and inverse transforms via

$$F(k) \equiv (2\pi)^{-1/2} \int_{-\infty}^{+\infty} e^{ikx} f(x) dx \tag{4.67}$$

$$f(x) \equiv (2\pi)^{-1/2} \int_{-\infty}^{+\infty} e^{-ikx} F(k) dk \tag{4.68}$$

then

$$i^n \phi_n(k) \equiv (2\pi)^{-1/2} \int_{-\infty}^{+\infty} e^{ikx} \phi_n(x) dx \tag{4.69}$$

Thus, the Hermite functions are, to within i^n, their own Fourier transforms. For the Hermite functions of degree $4n$, $i^{4n} = 1$ so that $\phi_{4n}(x)$ is *self-reciprocal* with respect to the Fourier Transform. Conversely, a function which is Fourier Transform self-reciprocal has a Hermite series expansion involving only the Hermite functions whose degree is an integer multiple of four.

4.5 Complex polar-separable functions

Shapelets are a complete, orthonormal set of 2D basis functions [65, 6, 63]. A linear combination of these functions can be used to model any image, in a similar way to Fourier or wavelet synthesis. The surface brightness $f(x)$ of an object can be written as

$$f(\mathbf{x}) = \sum_{n=0}^{+\infty} \sum_{m=-n}^{n} a_{nm} \chi_{nm}(\mathbf{x} - \mathbf{x}_c; \beta) \tag{4.70}$$

where β is a scale parameter, and $\mathbf{x}_c$ is the position of the center of the basis functions. These functions are defined only for combinations of n and m where both are even or both are odd. In the field of optical systems, it is customary to define the set of mode functions as

$$\Phi_p^m(r, \theta) = \left[\frac{(2 - \delta_{0m})\, p!}{\pi \beta^2 (p + |m|)!} \right]^{1/2} x^{|m|/2} e^{-x/2} L_p^{|m|}(x)\, e^{im\theta} \tag{4.71}$$

where $x = \frac{r^2}{\beta^2}$ and $L_p^{|m|}(x)$ are the standard associated Laguerre polynomials. The functions Φ_p^m are conventionally called the Gaussian beam *pm*-modes or simply the *pm*-modes and form an orthonormal set over the infinite Euclidean plane. The radial

index can take the values $p = 0, 1, 2, \ldots, \infty$, and the azimuthal index m can take the values $m = 0, \pm 1, \pm 2, \ldots, \pm\infty$.

$$L_p^q(x) \equiv \frac{x^{-q}e^x}{p!}\frac{d^p}{dx^p}\left(x^{p+q}e^{-x}\right) \tag{4.72}$$

with the useful recurrences:

$$\begin{aligned} &L_0^q(x) = 1 \\ &L_1^q(x) = 1 - x + q \\ &L_p^q(x) = \left(2 + \frac{q-1-x}{p}\right) L_{p-1}^q(x) - \\ &-\left(1 + \frac{q-1}{p}\right) L_{p-2}^q(x) = L_p^{q+1}(x) - L_{p-1}^{q+1}(x) = \sum_{i=0}^{p} L_i^{q-1}(x) \end{aligned} \tag{4.73}$$

and derivatives properties:

$$\frac{dL_p^q(x)}{dx} = x^{-1}\left[pL_p^q(x) - (p+q)L_{p-1}^q(x)\right] = -L_{p-1}^{q+1}(x) \tag{4.74}$$

$$\frac{dL_p^0(x)}{dx} = \frac{dL_{p-1}^0(x)}{dx} - L_{p-1}^0(x) = -\sum_{i=0}^{p-1} L_i^0(x) \tag{4.75}$$

In Fig. 4.4 are plotted the first terms of the Laguerre polynomials.

Using the convention in [65], and remember the analogy with the Quantum Mechanics, it is more usual to label the modes in terms of the energy quantum number $n = 2p + |m|$ and m, rather than p and m. The eigenfunctions can be defined as:

$$\begin{aligned} \chi_{n,m}(r,\theta;\beta) &= \frac{-1^{(n-|m|)/2}}{\beta^{|m|+1}}\times \\ \times\left[\frac{((n-|m|)/2)!}{\pi((n+|m|)/2)!}\right]^{1/2} &r^{|m|}L_{(n-|m|)/2}^{|m|}\left(\frac{r^2}{\beta^2}\right)e^{-\frac{r^2}{2\beta^2}}e^{-im\theta} \end{aligned} \tag{4.76}$$

where the index n describes the radial oscillations and the index m describes the order of rotational symmetry.

As in the case of Cartesian shapelets, the polar shapelet basis functions are orthonormal:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \chi_{n,m}^*(r,\theta;\beta)\,\chi_{n\prime,m\prime}(r,\theta;\beta)\,r dr d\theta = \delta_{n,n\prime}\delta_{m,m\prime} \tag{4.77}$$

and complete:

$$\sum_{n=0}^{+\infty}\sum_{m=-n}^{n} \chi_{n,m}(r,\theta;\beta)\,\chi_{n,m}(r\prime,\theta\prime;\beta) = \delta(r-r')\,\delta(\theta-\theta') \tag{4.78}$$

where δ is the Kronecker delta and the asterisk denotes complex conjugation.

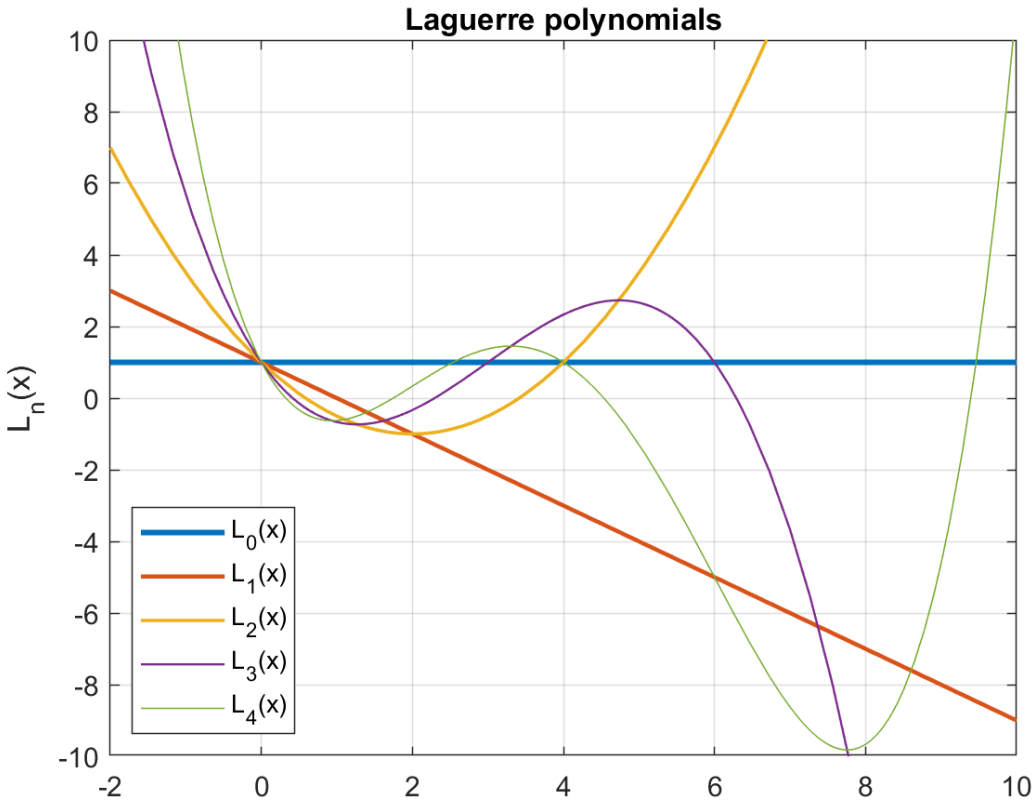

Figure 4.4. The first five terms of the Laguerre polynomials.

For reasonable choices of the centroid x_c and scale size β, the image shape information is contained within only the first few shapelet coefficients. The two-dimensional Gauss-Laguerre continuous mode functions are the simultaneous eigenstates of energy and angular momentum for the two-dimensional isotropic quantum harmonic oscillator. The main advantage of using the Gauss-Laguerre polynomials is the high quality of the image reconstruction for each type of shape. In addition, we have excellent data compression by simply setting the minimum and maximum physical scale of interest. The deviation of the coefficients of order greater than n contains a small amount of information at high spatial frequencies, which, however, are corrupted by noise. In the reconstruction, however, we must pay attention to the possible artifacts created by the truncation. Orthonormality ensures that the shapelet coefficients $a_{n,m}$ are given by

$$a_{n,m} = \int_0^{2\pi} d\theta \int_0^{+\infty} r dr f\left(r,\theta\right) \chi_{n,m}\left(r,\theta;\beta\right) \tag{4.79}$$

These are Gaussian-weighted multipole moments of the surface brightness.

Thus, in polar coordinates, we have the brightness of $f(r,\ \theta)$ [64] equal to:

$$f\left(r,\theta\right) = \sum_{n=0}^{+\infty} \sum_{m=-n}^{n} a_{n,m} \chi_{n,m}\left(r,\theta;\beta\right) \tag{4.80}$$

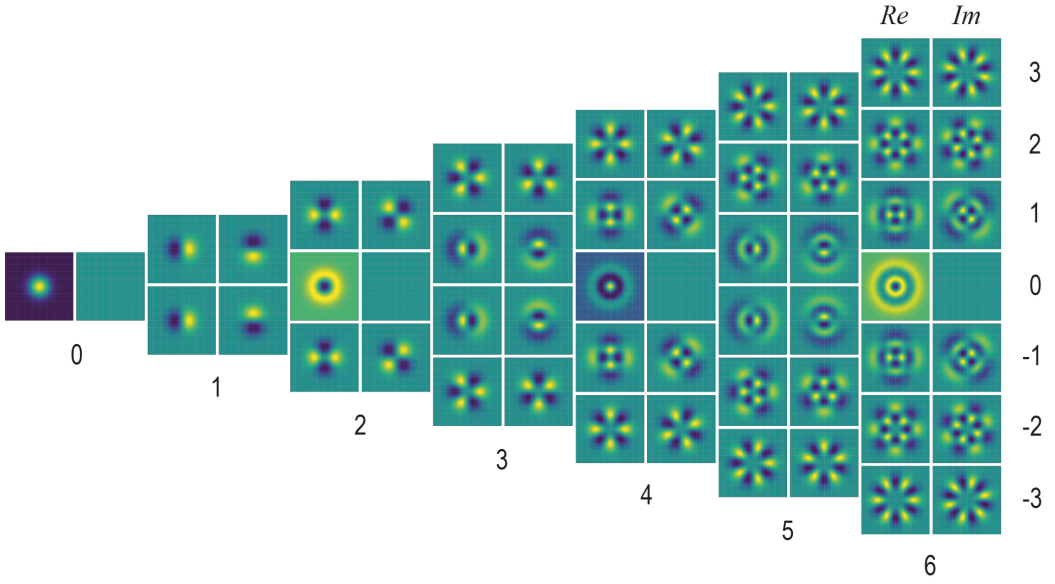

Figure 4.5. The overall 3D plots of the Laguerre polynomials. Each pair of columns represents the real and imaginary components of the signal: the basis functions are only defined if n and m are both even or both odd. The polar shapelet basis functions and coefficients are also complex numbers. However, the constraint that a combined image should be a wholly real function introduces degeneracies, $a_{n,m} = a^*_{n,-m}$ and the coefficients with $m = 0$ are wholly real.

These are successive perturbations around a Gaussian of width β, parameterized by indices n and m. The mathematics of a shapelets is somewhat analogous to Fourier synthesis, but with a compact support well-suited to the modelling of localized image shapes. For example, a shapelet decomposition can similarly be truncated to eliminate the highly oscillatory basis functions that correspond noise in the original image. A polar shapelet decomposition conveniently separates components of an image that are intuitively different (see Fig. 4.5). The index n describes the total number of oscillations (spatial frequency) and also the size (radius) of the basis function. The index m describes the degree of rotational symmetry of the basis functions. Basis functions with $m = 0$ are rotationally invariant. A circularly symmetric object contains power only in these states; its flux and radial profile are defined by the relative values of its $m = 0$ coefficients. Basis functions with $|m| = 1$ are invariant only under rotations of 360°. These coefficients encode an object's centroid: their real and imaginary parts correspond to displacements in the x and y directions. Alternatively, their moduli correspond to an absolute distance, and their phases indicate a direction. Basis functions with $|m| = 2$ are invariant under rotations of 180°, and become negative versions of themselves under rotations of 90°. These are precisely the properties of an ellipse. Indeed, an object's Gaussian-weighted ellipticity $e \equiv e_1 + ie_2$ is simply given by $f_{2,\pm 2}$. Its unweighted ellipticity is a combination of all of the $|m| = 2$ shapelet coefficients. Ellipticity estimators can also incorporate coefficients with $|m| = 6,\ 10,\ 14,\ ...$, because their basis functions also contain at least the necessary symmetries. All linear transformations can be described in shapelet space by the mixing of power between a few adjacent shapelet

coefficients. An initially circular object may contain power in all of its $m = 0$ coefficients. After a small shear, it also contains power in its $|m| = 2$ coefficients: the combination of circularly-symmetric plus quadrupole states produces an ellipse. Using shapelets, every shapelet coefficient with $|m| = 2$ can provide a statistically independent ellipticity estimator. Multiple shear estimators are very useful. Firstly, they can act as a consistency check to examine measurement errors within each object. They can also be combined to increase S/N. For example, it is possible to take a linear combination of $|m| = 2$ coefficients that is independent of the choice of β. However, the most successful estimator involves a *multiple multipole* combination of $|m| = 2,\ 6,\ 10, ...$ shapelet coefficients.

This simple expression results from the fact that only the $m = 0$ basis functions are invariant under rotations. These are given by

$$\chi_{n,0}(r;\beta) = \frac{-1^{n/2}}{\beta\sqrt{\pi}} L^0_{n/2}\left(\frac{r^2}{\beta^2}\right) e^{-\frac{r^2}{2\beta^2}} \tag{4.81}$$

We expand the first few rotationally invariant basis functions:

$$\begin{aligned}
\chi_{0,0}(r;\beta) &= \frac{1}{\beta\sqrt{\pi}} e^{-\frac{r^2}{2\beta^2}} \\
\chi_{2,0}(r;\beta) &= \frac{-1}{\beta\sqrt{\pi}}\left(1 - \frac{r^2}{\beta^2}\right) e^{-\frac{r^2}{2\beta^2}} \\
\chi_{4,0}(r;\beta) &= \frac{1}{\beta\sqrt{\pi}}\left(1 - 2\frac{r^2}{\beta^2} + \frac{1}{2}\frac{r^4}{\beta^4}\right) e^{-\frac{r^2}{2\beta^2}} \\
\chi_{6,0}(r;\beta) &= \frac{-1}{\beta\sqrt{\pi}}\left(1 - 3\frac{r^2}{\beta^2} + \frac{3}{2}\frac{r^4}{\beta^4} - \frac{1}{6}\frac{r^6}{\beta^6}\right) e^{-\frac{r^2}{2\beta^2}} \\
\chi_{8,0}(r;\beta) &= \frac{1}{\beta\sqrt{\pi}}\left(1 - 4\frac{r^2}{\beta^2} + \frac{6}{2}\frac{r^4}{\beta^4} - \frac{4}{6}\frac{r^6}{\beta^6} + \frac{1}{24}\frac{r^8}{\beta^8}\right) e^{-\frac{r^2}{2\beta^2}} \\
\chi_{10,0}(r;\beta) &= \frac{-1}{\beta\sqrt{\pi}}\left(1 - 5\frac{r^2}{\beta^2} + \frac{10}{2}\frac{r^4}{\beta^4} - \frac{10}{6}\frac{r^6}{\beta^6} + \frac{5}{24}\frac{r^8}{\beta^8} - \frac{1}{120}\frac{r^{10}}{\beta^{10}}\right) e^{-\frac{r^2}{2\beta^2}}
\end{aligned} \tag{4.82}$$

Thus, we developed the formalism of *polar shapelets*, in which an image is decomposed into components with explicit rotational symmetries. Whilst the original Cartesian shapelets remain useful in certain situations, the polar shapelets, which are separable in r and θ, frequently provide a more elegant and intuitive form.

In Fig. 4.5 are plotted the overall 3D plots of the Laguerre polynomials.

Cartesian shapelets are real functions, but polar shapelet basis functions $\chi_{n,m}$ and coefficients $a_{n,m}$ are both complex. However, their symmetries

$$\chi_{n,-m}(r,\theta;\beta) = \chi^*_{n,m}(r,\theta;\beta) = \chi_{n,m}(r,-\theta;\beta) \tag{4.83}$$

simplify matters somewhat if we are concerned only with the representation of real functions $f(x)$, such as the surface brightness of an image. Equations 4.77 and 4.83 imply that $f(x)$ is real if and only if

$$a_{n,-m} = a^*_{n,m} \tag{4.84}$$

Coefficients with $m = 0$ are thus wholly real. All polar shapelet coefficients are paired with their complex conjugate on the other side of the line $m = 0$. Therefore, even though the polar shapelet coefficients $a_{n,m}$ are generally complex, the number of independent parameters in the shapelet decomposition of a real function is conserved from the Cartesian case. A set of Cartesian shapelet coefficients a_{n_1,n_2} with $n_1 + n_2 \leq n_{max}$ can be transformed, into polar shapelet representation with $n \leq n_{max}$, using

$$a_{n,m} = 2^{-n/2} i^m \left[\frac{n_1! n_2!}{((n+m)/2)!\,((n-m)/2)!} \right]^{1/2} \delta_{n_1+n_2,n} \times \\ \times \sum_{n_r\prime=0}^{n_r} \sum_{n_l\prime=0}^{n_l} i^{m\prime} \binom{(n+m)/2}{n_r\prime} \binom{(n-m)/2}{n_l\prime} \delta_{n_l\prime+n_r\prime,n_1} a_{n_1,n_2} \tag{4.85}$$

4.6 Fourier Transform of the Laguerre-Gauss Functions

The Harmonic Angular Filter (HAF) are complex-valued and polar-separable functions formed by the product of a radial isotropic profile with an azimuthal harmonic function. Considering only the first-order component of a local expansion over the orthogonal family of the Laguerre Gauss (LG) functions [66, 79, 46] it is possible to define the Virtual Receptive Field model (VRF, whose concept will be better extended in Part III) with a very simple formulas:

$$h(r, \varphi) = \frac{r}{2\pi s_G^2} e^{-\frac{r^2}{2{s_G}^2}} e^{j\varphi} \tag{4.86}$$

where $r = \sqrt{x_1^2 + x_2^2}$, $\varphi = tg^{-1} \dfrac{x_2}{x_1}$ and s_G is a parameter that defines the spatial resolution as the grid in the background in Fig. 4.6.

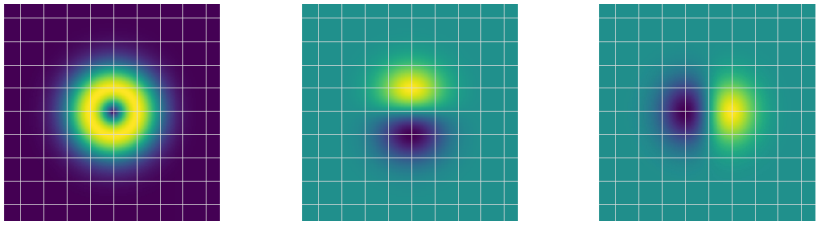

Figure 4.6. (left) $|h(x_1, x_2)|$, (center) $Re\{h(x_1, x_2)\}$, (right) $Im\{h(x_1, x_2)\}$. They represent, respectively, the magnitude of the Point Spread Function (PSF) of the VRF, its real and imaginary parts, referred to the 1 arcmin ideal retina grid.

With reference to the spatial frequency domain, defined by the horizontal and vertical frequencies f_1,f_2, and to the polar coordinates $\rho = \sqrt{f_1^2 + f_2^2}$ and $\vartheta = arctan\left(\dfrac{f_2}{f_1}\right)$, where the ρ [22] is referred to as the radial frequency and ϑ is the azimuth, the Fourier spectrum $H(\rho, \vartheta)$ is defined as

$$H(\rho, \vartheta) = j2\pi \left(\rho e^{-S_G^2 \rho^2} \cdot e^{j\vartheta} \right) \tag{4.87}$$

where j is the imaginary unit. It represents the Virtual Neural Transfer Function (VNTF, see Part III) and it indicates how much the complex amplitude of a single harmonic image component is modified by the VRF. This spectrum is polar separable, i.e., it is the product of a function of the radial frequency, and a function of the azimuth, as shown in Fig. 4.7.

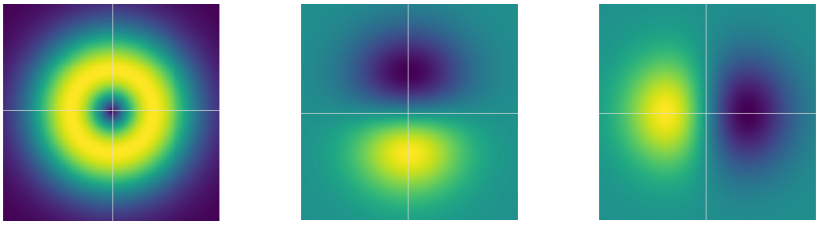

Figure 4.7. (left) $|H(f_1, f_2)|$, (center) $Re\{H(f_1, f_2)\}$, (right) $Im\{H(f_1, f_2)\}$. They represent, respectively, the magnitude of the VNTF, its real and imaginary parts in the spatial frequency plane. The vertical and horizontal spatial frequencies span the $(-30, 30)$ *cycles/degree* range.

Chapter 5

ML estimator and Positional Fisher Information

5.1 Pattern localization

Pattern localization is one of the most critical issues of artificial vision. Classical applications of pattern recognition and localization are found in robotics (with applications in the field of automatic inspection in industrial production), as well as automatic driving. Accurate pattern localization also plays a vital role for images registration in many relevant applications, such as image mosaicing, video compression and enhancement, multi-spectral overlap, etc. On the other hand, the exponential growth of large scientific image databases in different domains (e.g., remote sensing, medical imaging, genomics, climate modeling, astrophysics, etc.) demands for effective and accurate pattern recognition techniques to automate the exploration of data. The search of images in large archives distributed on the web is increasing in the world of Internet. Therefore, Content-Based Image Retrieval (CBIR) constitutes one of the most relevant challenges in the area of multimedia image processing.

The proposed method is suited for a large variety of applications, such as image archive exploration, data mining, map registration for telesensing and medical imaging, optical telemetry. Besides, it has been employed for motion estimation in video sequences that were required to be robust to rotation and zoom. A simpler form for computer-assisted reassembly of fragmented frescoes in the context of a research program for artwork recovery has been also experimented.

5.2 ML estimate in the Gauss-Laguerre domain

A set of consecutive points (pixels) of an image are by definition independent on the plane and equally spaced. It is possible to work on single spots and then combine the results for the study of an entire image $I(\mathbf{x})$ [69]. We consider $I(\mathbf{x})$ a noisy version, translated into position $\mathbf{b} = [b1, b2]^T$, rotated by φ and scaled with respect to parameter a, of the reference image $\widetilde{I}(\mathbf{x})$, so we have:

$$w\left(R_\varphi\left(\mathbf{x}-\mathbf{b}\right)\right) I\left(\mathbf{x}\right) = w\left(R_\varphi\left(\mathbf{x}-\mathbf{b}\right)\right) \widetilde{I}\left(R_\varphi\left(\frac{\mathbf{x}-\mathbf{b}}{a}\right)\right) + v\left(\mathbf{x}\right) \tag{5.1}$$

where $R_\varphi = \begin{bmatrix} cos(\varphi) & sin(\varphi) \\ -sin(\varphi) & cos(\varphi) \end{bmatrix}$ is a rotation operator and $v(\mathbf{x})$ is the white circular Gaussian, zero expected value, random, added noise with power density spectrum equal to $(N_0/4)$. To estimate the parameters $\mathbf{b}$, φ and a, we need a window $w(\mathbf{x})$ to isolate the objects to be identified in the image. For each portion of the corrupted image analyzed, the similar portion of the reference image is rotated and scaled to search for corresponding patterns, by varying the size of the window according to the variation of the scale factor. The conditional probability density of $I(\mathbf{x})$ w.r.t. $\mathbf{b}$, φ and a corresponds to the p.d.f. of $v(\mathbf{x})$:

$$p\left(I\left(\mathbf{x}\right)/a,\varphi,\mathbf{b}\right)=p_v\left(w\left(R_\varphi\left(\mathbf{x}-\mathbf{b}\right)\right)\left[I\left(\mathbf{x}\right)-\widetilde{I}\left(R_\varphi\left(\frac{\mathbf{x}-\mathbf{b}}{a}\right)\right)\right]\right) \tag{5.2}$$

The maximum likelihood functional for our two-dimensional study is the following (we have assumed $v(\mathbf{x})$ as an implementation of a random Gaussian process with zero expected value):

$$\ln\Lambda\left(I\left(\mathbf{x}\right);a,\varphi,\mathbf{b}\right)=-\frac{2}{N_0}\iint\left|w\left(R_\varphi\left(\mathbf{x}-\mathbf{b}\right)\right)\right|^2\left|I\left(\mathbf{x}\right)-\widetilde{I}\left(R_\varphi\left(\frac{\mathbf{x}-\mathbf{b}}{a}\right)\right)\right|^2dx \tag{5.3}$$

The scale factor with respect to the size of the window used is a critical point because extracted patterns could be introduced to the analyzed area of the reference image or, on the contrary, patterns of interest could be eliminated, even in the case of correct positioning and correct orientation. Recalling the Gauss-Laguerre polynomials seen in the previous chapter $\Phi_p^q(r,\theta)$, it is possible to represent locally each type of angular harmonic by means of its series expansion with respect to this orthonormal basis:

$$\Phi_p^q\left(r,\theta\right)=\left[\frac{\left(2-\delta_{0q}\right)p!}{\pi a^2\left(p+|q|\right)!}\right]^{1/2}\left(\frac{r^2}{a^2}\right)^{|q|/2}e^{-\left(r^2/a^2\right)/2}L_p^{|q|}\left(\frac{r^2}{a^2}\right)e^{iq\theta} \tag{5.4}$$

where we have replaced β with a (scale parameter) and $x=\dfrac{r^2}{a^2}$. Using a symmetrical circular Gaussian window $w(\mathbf{x})=V(\mathbf{x}/\sigma)$, with $V(\mathbf{x})=e^{-\pi|\mathbf{x}|^2}$, we can avoid the explicit scale parameter a as follow:

$$\Phi_p^q\left(r,\theta\right)=(-1)^p\,2^{(|q|+1)/2}\pi^{|q|/2}\left[\frac{p!}{(p+|q|)!}\right]^{1/2}r^{|q|}e^{-\pi r^2}L_p^{|q|}\left(2\pi r^2\right)e^{iq\theta} \tag{5.5}$$

The generalized Laguerre polynomials $L_p^q(x)$ are defined by the Rodrigues' formula:

$$L_p^q\left(x\right)\equiv\frac{x^{-q}e^x}{p!}\frac{d^p}{dx^p}\left(x^{p+q}e^{-x}\right) \tag{5.6}$$

Thus, each image $I(\mathbf{x})$ can be represented by its local expansion with respect to point $\mathbf{b}$, in terms of the orthonormal basis $\Phi_p^q(r,\theta)$:

$$I\left(\mathbf{x}\right)V\left(\frac{\mathbf{x}-\mathbf{b}}{\sigma}\right)=\sum_q\sum_p D_{q,p}\left(\mathbf{b}\right)\frac{1}{\sigma}\Phi_p^q\left(\frac{|\mathbf{x}-\mathbf{b}|}{\sigma},\theta\left(\mathbf{x}-\mathbf{b}\right)\right) \tag{5.7}$$

where $D_{q,p}(\mathbf{b})$ are the expansion coefficients defined as:

$$D_{q,p}(\mathbf{b}) = \left\langle I(\mathbf{x})\, V\left(\frac{\mathbf{x}-\mathbf{b}}{\sigma}\right),\ \frac{1}{\sigma}\Phi_p^q\left(\frac{|\mathbf{x}-\mathbf{b}|}{\sigma}, \theta(\mathbf{x}-\mathbf{b})\right)\right\rangle = I(\mathbf{b}) * V\left(\frac{-\mathbf{b}}{\sigma}\right)\left[\frac{1}{\sigma}\Phi_p^q\left(\frac{|-\mathbf{b}|}{\sigma}, \theta(-\mathbf{b})\right)\right] \tag{5.8}$$

We have, therefore, the Gauss-Laguerre transform of image $I(\mathbf{x})$, which is by fact the polar form of the Hermite Transform. Images are represented by means of special Circular Harmonic Functions (CHF), called Gauss-Laguerre (GL) functions. In general, CHF's are complex, polar separable functions with harmonic angular shape, suited for generating rotation invariant signatures closely related to moment invariants. GL functions have the important isomorphic property with respect to their Fourier transforms and, since they can be easily obtained from the differentiation of isotropic Gaussian functions, they are widely used in computer vision. Applying the GLT to the image $\widetilde{I}(\mathbf{x})$ we have:

$$V\left(\frac{R_\varphi(\mathbf{x}-\mathbf{b})}{\sigma}\right)\widetilde{I}\left(R_\varphi\left(\frac{\mathbf{x}-\mathbf{b}}{a}\right)\right) = \sum_q\sum_p \eta_{q,p}(a)\, e^{-iq\varphi}\frac{1}{\sigma}\Phi_p^q\left(\frac{|\mathbf{x}-\mathbf{b}|}{\sigma}, \theta(\mathbf{x}-\mathbf{b})\right) \tag{5.9}$$

with $\eta_{q,p}(a) = \left\langle \widetilde{I}\left(\frac{\mathbf{x}}{a}\right) V\left(\frac{\mathbf{x}}{\sigma}\right),\ \frac{1}{\sigma}\Phi_p^q\left(\frac{|\mathbf{x}|}{\sigma}, \theta(\mathbf{x})\right)\right\rangle$. By replacing the last expressions found in the likelihood function and considering that the Gaussian window is invariant to rotations, we have the new expression for the likelihood functional:

$$\ln\Lambda(I(\mathbf{x}); a, \varphi, \mathbf{b}) = -\frac{2}{N_0}\sum_{q=0}^{+\infty}\sum_{p=0}^{+\infty}\left|D_{q,p}(\mathbf{b}) - \eta_{q,p}(a)\, e^{-iq\varphi}\right|^2 \tag{5.10}$$

We note a dependence of the coefficients $\eta_{q,p}$ on the scale parameter a. If the function $\widetilde{I}(\mathbf{x})\, V\left(\frac{\mathbf{x}}{\sigma}\right)^{-1} V\left(\frac{a\mathbf{x}}{\sigma}\right)$ is square-integrable as:

$$C_{q,p} = \left\langle \widetilde{I}(\mathbf{x})\, V\left(\frac{\mathbf{x}}{\sigma}\right),\ \frac{1}{\sigma}\Phi_p^q\left(\frac{|\mathbf{x}|}{\sigma}, \theta(\mathbf{x})\right)\right\rangle \tag{5.11}$$

then, the GLT coefficients of the scaled image can be determined starting from new coefficients $C_{q,p}$ of the non-scaled image. We thus obtain a new expression for the coefficients $\eta_{q,p}$:

$$\eta_{q,p}(a) = \sum_{l=p}^{+\infty} B(a; q, p, l) C_{q,l} \tag{5.12}$$

where

$$B(a; q, p, l) = (-1)^{l-p}\sqrt{\frac{(q+l)!l!}{(q+p)!p!}}\frac{a^{-q-2p}}{(l-p)!}\left(1-\frac{1}{a^2}\right)^{l-p} \tag{5.13}$$

with $l \geq p$, and the likelihood functional becomes independent of rotation and scale factor a:

$$\ln\Lambda(I(\mathbf{x}); a, \varphi, \mathbf{b}) = -\frac{2}{N_0}\sum_{q=0}^{+\infty}\sum_{p=0}^{+\infty}\left|D_{q,p}(\mathbf{b}) - \sum_{l=p}^{+\infty} B(a; q, p, l)\, C_{q,l} e^{-iq\varphi}\right|^2 \tag{5.14}$$

At this point, the maximum value with respect to scale and orientation is:

$$GLLM\left(\mathbf{b}\right)=\max_{a,\varphi}\left\{\ln\Lambda\left(I\left(\mathbf{x}\right);a,\varphi,\mathbf{b}\right)\right\} \tag{5.15}$$

that we can call *Gauss-Laguerre Likelihood Map* (GLLM), whose absolute maximum indicates the pattern location along with the associate scale and orientation. The essential point of this method is that the true ML solution can be approximated by GLT truncation, and that few GLT coefficients are most often sufficient to obtain unambiguous and reliable estimates.

5.3 Position Fisher Information and the positional structural information

Among the basic cognitive functions of HVS, detection, recognition, and coarse localization are strongly conditioned by the individual experience. It seems plausible that the fine localization function is largely committed to low-level and inter-subjective functions of the visual sensory apparatus. The present approach starts from postulating that, under normal conditions, the HVS performs the fine localization of the observed objects with the best accuracy allowed by its physical macro-structure, considering the characteristics of the environment and of the human interaction with it. From estimation theory, the maximum accuracy attainable when measuring the fine position of patterns in background noise is pointed out by the Fisher Information about positional parameters and the Fisher Information inverse yields the minimum estimation variance [91]. Therefore, in this work, attention is focused to how blur discomfort depends on the undesired loss of the Positional Fisher Information (PFI) about observed patterns. In other studies, distance estimation through binocular vision requires accurate displacement estimation of patterns in the image plane [8]. Estimation theory provides a well-defined measure of the allowable accuracy of localization of an image pattern and suggests that the positional FI should be correlated with the subjective rating of image quality.

In the study of an image, in natural vision condition and no occlusions, i.e., in situations where no artificial additions appear (for instance, a title in an image), so that only natural images are considered (with spectrum equal to $1/\rho$), we can use the concept of the Fisher Information to find the global minimum (second order is sufficient). Fisher Information grows with the square of the image area and with the size of the samples N. Even in the case of Fisher Information close to zero, for the identification of the minimum we would always have information that grows more slowly than N, considering higher order components, not reachable with polynomial techniques.

As mentioned, when we talked about the ML model, we want to compare the portion of the selected distorted image with the corresponding reference image and the translation operation in one dimension can be written as follows:

$$f(x)+\frac{\partial f(x)}{\partial x}\Delta\mathrm{x} \tag{5.16}$$

We approximate $f(x)$ to order zero with a Gaussian curve and with Hermite-Gauss or Laguerre-Gauss functions for higher orders. As seen, up to a scale factor, they

form an orthogonal base and with the use of a Gaussian window, the spectrum drops as e^{-x^2}. Other structures do not allow such a rapid decrease of the spectrum even in the case in which the system is well identified (think, for example, of the Cauchy structure, in which the system, although very well identifiable, tends to a constant value or decreases very slowly).

To locate a pattern in an image, we can calculate the associated Fisher Information relating to position, rotation, and scale [69]. By defining a single vector $\theta = [b_1\ b_2\ \varphi\ a]$, Fisher Information J is defined as follows:

$$\mathbf{J}(\theta) = E\left\{\nabla_\theta \ln p(I;\theta)\left[\nabla_\theta \ln p(I;\theta)\right]^T\right\} \tag{5.17}$$

which we can divide into several contributions:

$$\mathbf{J} = \mathbf{J}^{(\widetilde{\mathbf{I}})} + \mathbf{J}^{(\mathbf{w})} + \mathbf{J}^{(\widetilde{\mathbf{I}},\mathbf{w})} + \left[\mathbf{J}^{(\widetilde{\mathbf{I}},\mathbf{w})}\right]^{\mathbf{T}} \tag{5.18}$$

where

$$\begin{aligned}\mathbf{J}^{(\widetilde{\mathbf{I}})} = \left(\frac{4}{N_0}\right)\iint \nabla_\theta\left\{I(\mathbf{x}) - \widetilde{I}\left(R_\varphi\left(\frac{\mathbf{x}-\mathbf{b}}{a}\right)\right)\right\}\times \\ \times\nabla_\theta\left\{I(\mathbf{x}) - \widetilde{I}\left(R_\varphi\left(\frac{\mathbf{x}-\mathbf{b}}{a}\right)\right)\right\}^T \left|w\left(R_\varphi(\mathbf{x}-\mathbf{b})\right)\right|^2 dx\end{aligned} \tag{5.19}$$

is the main information associated with the pattern we are looking for,

$$\begin{aligned}\mathbf{J}^{(\mathbf{w})} = \left(\frac{4}{N_0}\right)\iint \left[I(\mathbf{x}) - \widetilde{I}\left(R_\varphi\left(\frac{\mathbf{x}-\mathbf{b}}{a}\right)\right)\right]^{\mathbf{2}}\times \\ \times\nabla_\theta\left\{w\left(R_\varphi(\mathbf{x}-\mathbf{b})\right)\right\}\nabla_\theta\left\{w\left(R_\varphi(\mathbf{x}-\mathbf{b})\right)\right\}^T dx\end{aligned} \tag{5.20}$$

is the information associated with the particular windows adopted,

$$\begin{aligned}\mathbf{J}^{(\widetilde{\mathbf{I}},\mathbf{w})} = \left(\frac{4}{N_0}\right)\iint w\left(R_\varphi(\mathbf{x}-\mathbf{b})\right)\left[I(\mathbf{x}) - \widetilde{I}\left(R_\varphi\left(\frac{\mathbf{x}-\mathbf{b}}{a}\right)\right)\right]\times \\ \times\nabla_\theta\left\{I(\mathbf{x}) - \widetilde{I}\left(R_\varphi\left(\frac{\mathbf{x}-\mathbf{b}}{a}\right)\right)\right\}\nabla_\theta\left\{w\left(R_\varphi(\mathbf{x}-\mathbf{b})\right)\right\}^T dx\end{aligned} \tag{5.21}$$

represents the cross pattern/window information.

In all cases of interest, the pattern has a significant weight in the image and the value of the $\mathbf{J}^{(\widetilde{\mathbf{I}})}$ information is dominant over the other components:

$$\begin{aligned}\left|\left[\mathbf{J}^{(\widetilde{\mathbf{I}})}\right]_{q,p}\right| &\gg \left|\left[\mathbf{J}^{(\mathbf{w})}\right]_{q,p}\right| \\ \left|\left[\mathbf{J}^{(\widetilde{\mathbf{I}})}\right]_{q,p}\right| &\gg \left|\left[\mathbf{J}^{(\widetilde{\mathbf{I}},\mathbf{w})}\right]_{q,p}\right| \\ \mathbf{J} &\cong \mathbf{J}^{(\widetilde{\mathbf{I}})}\end{aligned} \tag{5.22}$$

Then, we write the $\mathbf{J}^{(\widetilde{\mathbf{I}})}$ matrix with respect to the parameters considered, to obtain $J_{\mathbf{b}}$, J_{φ} and J_a, which represent the information of each parameter brought by the noisy measurements:

$$\mathbf{J}^{(\widetilde{\mathbf{I}})} = \begin{bmatrix} J_{\mathbf{b}} & J_{\mathbf{b},\varphi} & J_{\mathbf{b},a} \\ J_{\mathbf{b},\varphi}^{\mathbf{T}} & J_{\varphi} & J_{\varphi,a} \\ J_{\mathbf{b},a}^{\mathbf{T}} & J_{\varphi,a}^{\mathbf{T}} & J_a \end{bmatrix} \tag{5.23}$$

From an energy point of view, we can write:

$$J_{\mathbf{b}} = \frac{4}{N_0} R_{\varphi} \mathcal{E}_{\nabla_{\widetilde{I}_x}} R_{\varphi}^T \tag{5.24}$$

$$J_{\varphi} = \frac{4}{N_0} a^2 \left(\mathcal{M}_{x_1,x_1}^{(0,2)} - 2\mathcal{M}_{x_1,x_2}^{(1,1)} + \mathcal{M}_{x_2,x_2}^{(2,0)} \right) \tag{5.25}$$

$$J_a = \frac{4}{N_0} \left(\mathcal{M}_{x_1,x_1}^{(2,0)} + 2\mathcal{M}_{x_1,x_2}^{(1,1)} + \mathcal{M}_{x_2,x_2}^{(0,2)} \right) \tag{5.26}$$

where

$$\mathcal{E}_{\nabla_{\widetilde{I}_x}} = \begin{bmatrix} \mathcal{E}_{\widetilde{I}_{x_1}} & \mathcal{E}_{\widetilde{I}_{x_1}\widetilde{I}_{x_2}} \\ \mathcal{E}_{\widetilde{I}_{x_1}\widetilde{I}_{x_2}} & \mathcal{E}_{\widetilde{I}_{x_2}} \end{bmatrix} \tag{5.27}$$

is the windowed energy tensor of $\nabla_{\widetilde{I}_x}$ and

$$\mathcal{M}_{x_q,x_p}^{(h,k)} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_1^h x_2^k \frac{\partial}{\partial x_q} \widetilde{I}(x_1,x_2) \frac{\partial}{\partial x_p} \widetilde{I}(x_1,x_2) \left| w(ax_1,ax_2) \right|^2 dx_1 dx_2 \tag{5.28}$$

are the windowed moments of partial derivatives $\widetilde{I}_{x_1}(x_1,x_2)$ and $\widetilde{I}_{x_2}(x_1,x_2)$.

In particular, the energy components can be expressed as:

$$\mathcal{E}_{\widetilde{I}_{x_1}} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left| \widetilde{I}_{x_1}(x_1,x_2) \right|^2 \left| w(ax_1,ax_2) \right|^2 dx_1 dx_2 \tag{5.29}$$

$$\mathcal{E}_{\widetilde{I}_{x_2}} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left| \widetilde{I}_{x_2}(x_1,x_2) \right|^2 \left| w(ax_1,ax_2) \right|^2 dx_1 dx_2 \tag{5.30}$$

$$\mathcal{E}_{\widetilde{I}_{x_1}\widetilde{I}_{x_2}} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \widetilde{I}_{x_1}(x_1,x_2) \widetilde{I}_{x_2}(x_1,x_2) \left| w(ax_1,ax_2) \right|^2 dx_1 dx_2 \tag{5.31}$$

Finally, in a more compact view, the positional FI carried by the image portion captured by the spot is calculated as

$$J(\mathbf{\Delta}) = \frac{\mathbf{T}_{\mathbf{\Delta}}}{\sigma_N^2} \tag{5.32}$$

where

$$\mathbf{T}_{\mathbf{\Delta}} = \begin{bmatrix} \mathcal{E}_{11} & \mathcal{E}_{12} \\ \mathcal{E}_{21} & \mathcal{E}_{22} \end{bmatrix} \tag{5.33}$$

is the *structure tensor* (ST) within the spot, defined by the following integrals:

$$\begin{aligned}
\mathcal{E}_{11} &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left[\frac{\partial i\,(p_1+\xi_1,p_2+\xi_2)}{\partial \xi_1}\right]^2 w(\xi_1,\xi_2)^2 d\xi_1 d\xi_2 \\
\mathcal{E}_{22} &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left[\frac{\partial i\,(p_1+\xi_1,p_2+\xi_2)}{\partial \xi_2}\right]^2 w(\xi_1,\xi_2)^2 d\xi_1 d\xi_2 \\
\mathcal{E}_{12} &= \mathcal{E}_{21} = \\
&\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left[\frac{\partial i\,(p_1+\xi_1,p_2+\xi_2)}{\partial \xi_1}\frac{\partial i\,(p_1+\xi_1,p_2+\xi_2)}{\partial \xi_2}\right] w(\xi_1,\xi_2)^2 d\xi_1 d\xi_2\ .
\end{aligned} \tag{5.34}$$

Being proportional to the ST, $J(\Delta)$ is symmetric and can be diagonalized by rotating the local coordinates into the coordinate system determined by the ST orthogonal eigenvectors. In addition, the eigenvalues of $J(\Delta)$, λ_1 and λ_2, represent the independent contributions to the FI along the two orthogonal axes in the eigenvector reference system. The total FI for the considered spot in the presence of additive WGRF is calculated as

$$\frac{\lambda_1+\lambda_2}{\sigma_N^2} = \frac{\mathcal{E}_{11}+\mathcal{E}_{22}}{\sigma_N^2} \tag{5.35}$$

This quantity is herein referred to as *Positional Structural Information* (PSI) carried by the generic spot.

ST is used in artificial vision: by exploiting its eigenvalues it is possible to detect fundamental local characteristics of an image (orientation, amplitude of strong edges, textures). In the analysis of textures, the ST is able to identify the dominant directions, in other cases, it can be used as a strong edge extractor. In video analysis, the ST detects the smallest eigenvector, which is useful for finding space-time direction with the least change. In recent years, ST has been used to guide diffusion processes to improve structural coherence.

We will retrieve this elegant formulation in Part IV, where an empirical estimator that works with strong edges and textures is presented.

Part III

A Mathematical Model of Blur Perception

Chapter 6

Predicting Blur Visual Discomfort by the Loss of Positional Information

6.1 Introduction to the methodologies adopted

6.1.1 Blur and discomfort

Blur is generally understood as a smoothing of the image, which is accompanied by a form of visual discomfort. This is *subjective blur*. Blur is generally modeled through low-pass filtering applied to the image. Through such a model it is possible to define an *objective blur*, and its *measure*. Two simple blur models are the circular aperture (sinc) and the Gaussian blur, both characterized by a single parameter.

The discomfort that matches viewing a blurry image is explained in several ways. Mostly, it is argued that the discomfort is felt because blur causes a deviation of the spectrum of the image projected on the retina with respect to a spectrum typical of real world images, the so-called "natural scenes".

One of the possible reasons is that the discomfort stems from the vain effort to eliminate blur through the eye accommodation system. From another point of view, the discomfort would be felt due to the inefficiency, caused by blur, of the image coding system towards the brain, which would cause metabolic overload. In both cases, we are dealing with physiological explanations of the phenomenon of discomfort.

6.1.2 The functional cognitive approach

The discomfort can also be explained as a consequence of a cognitive deficit. The discomfort would be due to the sensation of a loss of the ability to interact with the observed reality. According to this point of view, the discomfort due to blur would be measurable based on the loss of information about the content of the image. In this perspective, the goal is to define a function that links the blur level to the loss of the Fisher Information, without taking into consideration how its calculation is actually implemented by the Human Visual System (HVS).

This "cognitivist" and "functional" approach is inspired by the theory of vision set out by Marr in his book [60] and has already been used, for example, in the studies of image coding systems. In such cases, Shannon's Information theory and Kolmogorov's algorithmic Information theory were used. In the present context we consider the so-called Fisher Information, which is simple to measure and relates directly to blur.

Fisher Information indicates the degree of precision with which an observation system can measure the parameters of an observed system.

In our case, the observed system is the image projected on the retina. The essential parameters of interest are those that identify the position of objects.

6.1.3 The principle of efficiency and its consequences

Fundamental principle:

"The Human Visual System is efficient from the point of view of estimating the positional parameters of an object".

This principle is of fundamental importance. It implies that the HVS is able to calculate the Fisher Information contained in the image projected on the retina. Since the Fisher information of the observed image can be calculated objectively, it is possible to measure the Fisher Information that the subject perceives.

6.1.4 Measuring Positional Fisher Information using VRF

In the recent past, Fisher Information computation schemes have been defined and can be implemented through the so-called Laguerre-Gauss transform. It is achieved by means of a bank of spatial filters whose impulsive responses are constituted by Laguerre-Gauss functions, organized in a layered structure.

In particular, this scheme allows us to calculate the Positional Fisher Information (FPI) using only the first order Laguerre-Gauss functions. These functions can be interpreted as the real and the imaginary parts of a complex filter that calculates the smoothed complex gradient at each point of the image. The complex gradient is essentially a gradient defined in every direction by the Pythagorean theorem applied to the horizontal and vertical gradient considered as cathets. Its modulus is measured by the length of the hypotenuse. The smoothed complex gradient is the complex gradient calculated after the image has been smoothed by a Gaussian filter.

This filter performs the function of a Virtual Receptive Field (VRF), which calculates the Fisher Information loss regardless of how this function is actually performed by the system of the real receptive fields. The Fisher Information contained in an areola of the image (detail) is finally obtained by calculating the total energy of the complex gradient within the detail.

6.1.5 The Fisher Information loss

We accept the following hypothesis:
"A loss of the Fisher Information causes increasing cognitive discomfort with the loss itself."

Therefore, if the Fisher Information loss is measured, it is possible to think of measuring the discomfort. Fisher Information depends on the prevalence of the

contrast of the objects with respect to the background thermal noise. This ratio decreases in low light conditions, only partially compensated by the accommodation mechanisms, but it decreases mainly due to refractive defects, especially blur.

The eye mechanisms that give rise to blur are very simple compared to the enormous complexity of the HVS, and they can be described using parametric models suggested by physical optics. Using these models it is possible to objectively calculate the Fisher Information loss caused by blur.

6.1.6 The hypothesis of linearity

The relationship that links the growing discomfort to the Fisher Information loss is detectable only through experimental tests. However, an a-priori hypothesis can be advanced. The inverse of the Fisher Information represents the variance of the estimation error on the position for an efficient estimator. Therefore, its square root measures the uncertainty about the position. Since the reality perceived by the HVS is evidently represented in a 3-D Euclidean space, it is plausible that the discomfort, interpreted as the perception of the cost of the position error, is proportional to the increase in uncertainty.

Hence, it can be reasonably assumed that:
The inconvenience is proportional to the increase in uncertainty calculated using the Fisher Information.

6.1.7 The Fisher Equivalence

The formula of the expected discomfort through the Fisher Information loss due to blur is analytically very simple if blur is Gaussian. This simplicity translates into immediate interpretability and the possibility of deducing interesting, even if approximate, properties. This suggests replacing, when possible, the actual blur with a Gaussian blur that causes the same loss of information, which, therefore, is *Fisher Equivalent* to the original one.

For example, since the neural blur is Gaussian, it is immediate to measure the impact of the optical blur using the neural blur as a natural reference and unit of measure for blur.

6.2 The physical and perceptual effects of blur

It is possible to interpret blur as a decrease in information at high spatial frequencies [12]. There are many ways to quantify blur: we can think of the global statistical luminance or the slope of the local luminance gradients in the image. One way to characterize an image is its amplitude spectrum, which represents the relative amplitude of the information present in the entire image at different spatial scales. In one step, the amplitude of the harmonics falls with a $\frac{1}{f}$ pattern: higher harmonics have smaller amplitudes. Removing harmonics from this waveform has the effect of reducing the high spatial frequency content, decreasing local luminance gradients, and making the image look more like a sine wave.

Even the modification of the phase influences, in part, the gradients of the waveforms [3]: the image can have the same amplitude spectrum, but a flatter

gradient, if the spatial frequencies of the components are out of phase with respect to those in phase. The slope of the luminance gradients can be changed by reducing the amplitude at high spatial frequencies, thus increasing the slope of the amplitude spectrum, or by changing the phase relationships of the information in the image. Both the removal of spatial high frequencies and the phase shift in local luminance gradients contribute to the perception of blur. Therefore, if we have a visual discomfort due to a variation in the slope of the local luminance gradients, we can say that the subjective judgment is influenced by blur, as a result of amplitude and phase changes.

Understanding the cause of blur is simple in the case of artificial images, but very complex if we consider natural images. There are several techniques for adding blur to natural reference images. The use of a Gaussian filter has been used in the past to reproduce the effect of optical blur [9]; actually, it is a limited model because it does not add the phase correction factor, due to the diffraction of the light passing through the pupil and which modifies the image created on the retina. In addition, we have a phase inversion that produces stripes around the edges of the reproduced image, modifying the shape of the amplitude spectrum and, therefore, the local luminance gradients. Using a sinc filter we have a more realistic model that takes these corrections into account [68]. However, studies on subjective perception have shown that there is a sensitivity only to the spectrum of the amplitude of blur. The phase, although it modifies the edges, is not perceived as a nuisance. Basically, the blur transfer function does not have to be Gaussian and symmetrical; it can be asymmetrical and the eye would not distinguish it from a symmetrical one. These studies bring qualitative results: changing the phase, people's judgment remains the same.

Although, as mentioned, there are many descriptions of blur (lack of contrast, loss of high frequencies, smoothing of the edges), the work presented in this thesis includes all these definitions, demonstrating that blur depends on the energy of the gradient. Furthermore, we establish the fundamental concept of equivalence between Gaussian and asymmetric blur. Indeed, blur of the eye is unknown, it is not Gaussian, except when it is very small, because we also have astigmatism or other simultaneous defects. In the study of our model, we bring everything back to Gaussian and look for an equivalent Gaussian blur from the Fisher Information point of view. Therefore, it is possible to consider any blur perceptually equivalent to a Gaussian blur with a certain s_B standard deviation value (at this point, all the blur information focuses on this parameter).

Finally, we have two fundamental concepts:

- Fisher Information as a quantity that must be optimized by the eye to eliminate the discomfort. It is a quantity that we can maximize following a Gaussian model: any blur can be traced back to the Gaussian one from the Fisher Information point of view.

- The stimulus which gives a measure of perceived discomfort. The estimator calculates the amplitude of the gradients because the receptive fields are sensitive to this, not directly to the Fisher Information. The magnitude proposed in this study is the stimulus signal which mechanically regulates the pupil. For healthy subjects, this signal always tends to zero. In the case of

vision disturbances due to blur, the adjustment stimulus that is generated in the retina never returns to zero. Hence, the higher the sensation of discomfort is and the longer it persists, the more the disturbance increases. The estimator indicates the level of the error and, therefore, represents the level of the annoyance.

6.3 Visual discomfort and accommodation system

We can think of blur as a cause of a cognitive loss, and the discomfort as the immediate consequence of this loss. We look for a relationship between judgments of visual discomfort and changes in image content. Visual discomfort is represented by all the subjective side effects that occur in the visualization of certain stimuli. Changes in the content of spatial frequencies modify image quality judgment and deviations from natural image statistics cause visual discomfort [49]. Natural images are the basis for comparisons of corrupted images, as they have reliable statistical properties. In particular, Fourier analysis reveals that the amplitude of the luminance decreases with the increase of the spatial frequency: Fourier amplitude spectrum $\frac{1}{f^{\beta}}$, with $\beta \sim 1$ [76] and, therefore, we have more contrast energy at low frequencies than at high frequencies. The HVS is thought to be optimized to encode stimuli with the luminance statistics of natural images [30], and, indeed, visual discomfort could grow precisely from stimuli that do not have these natural statistics, which are not processed in the optimal way. Moreover, deviations from natural image statistics have been shown to influence subjective quality judgments [70]. The causes of the increase in visual discomfort can be either an increase in contrast energy, resulting in an increase in the amplitude spectrum, or an increase in β, which makes the shape of the amplitude spectrum steeper, with relatively lower information on the spatial frequency.

Therefore, in relating visual discomfort to blur we find objective techniques, in which a direct link is sought with a physical fact, in particular, with the change in contrast at each frequency, as perceived by the eye, and subjective techniques, in which mental mechanisms that give a measure of the discomfort in seeing a blurred image are considered. Actually, these mechanisms give no idea of what is happening in the eye. The discomfort associated with blur is often explained as a consequence of the concentration of the spatial energy spectrum of the perceived image into some bands, or as a byproduct of the discrepancy of this spectrum from the expected spectrum of natural images, as we said. Looking at a possible physical source of the blur discomfort, a first hypothesis is that discomfort is stimulated by the weak response of the accommodation system.

Alternative explanations addressed the mismatch of the spatial patterns with the expected ones. To understand what happens at the physiological level in the eye, it is useful to study the accommodation mechanism: this can be thought of as a computational increase necessary to perform the deblurring, in case the mechanical action has not brought the eye back to nominal conditions of vision (residual blur). Therefore, discomfort arises because the "microfluctuations" observed in the accommodation feedback signal become ineffective [17, 67, 59, 18]. For example, in digital photography we first try to focus mechanically, then, if there is still an

unresolved blur, we process the image afterwards.

A third hypothesis is that, when an image is correctly projected onto the retina, the receptive fields produce a parsimonious, *sparse* representation of this image [66]. The spatial spread caused by blur excites more receptors, producing a metabolic overload [49]. This post-processing is very expensive and causes visual discomfort, both for purely physiological reasons (increase in blood pressure in the eye) and for neuronal reasons, understood as an increase in energy consumption due to the greater number of neurons used and as anxiety (stress) created for failure to correct (there may be an effect on the higher semantic level). We can therefore think of statistical features strongly correlated with the feeling of blur.

Corrective answers provide a possible reason why blur could cause discomfort. One goal of corrective responses is maximizing the contrast of the image on the retina. We could think of refocusing a blurred image at an appropriate distance, but there would still be a limited understanding of what spatial frequency information is used by the corrective system [57]. There is also uncertainty as to whether the system uses local or global image statistics to achieve this. Among the basic cognitive functions of the HVS, detection, recognition, and coarse localization functions are strongly conditioned by the individual experience. Conversely, it seems plausible that the fine localization function is committed to stabler and inter-subjective functions of the HVS.

The fundamental assumption is that, under normal conditions, the HVS performs the fine localization of the observed objects with the best accuracy allowed by its physical macro-structure, given the characteristics of the environment and man's interaction. Indeed, it is known from the estimation theory that the maximum accuracy attainable when measuring the fine position of patterns in background noise is deduced by the Fisher Information (FI) about positional parameters. In fact, the FI inverse yields the minimum estimation variance [91].

Therefore, the focus in this thesis is on how the discomfort of blurring depends on the unwanted loss of PFI on observed patterns. Even if the present cognitive approach remains agnostic, it is compliant with the fact that blur discomfort concerns the regions of visual interest [87] and that blur is not always undesired or detrimental [85]. For instance, blur is sometimes a wanted effect in photography and microscopy.

Previous analyses of the blur perception phenomenon were mainly oriented to the study of the visual acuity, with specific attention to specific stimuli localized either in space, such as edges, lines, crosses, or in the spatial frequency domain, such as sinusoidal gratings, or even in both domains, such as Gabor wavelets. The model presented here is oriented to the evaluation of the visual impact of blur in the vision of natural scenes. To this purpose, a generic image projected on the retina is viewed as a realization of a random set, constituted by the universe of natural images, characterized by stable statistical features.

6.4 The Virtual Receptive Field model

The proposed approach is based on an abstract, functional model of the Receptive Fields (RF) of the HVS, referred to as VRF. The role of the VRF model is to extract the PFI as a measure of the pattern localizability loss.

For analytical convenience, blur is initially modeled as a Gaussian shaped isotropic blur. Then, the analysis is generalized to other types of blur, invoking a criterion of informational equivalence with respect to isotropic Gaussian blur under the PFI paradigm.

The VRF model is based on Harmonic Angular Functions (HAFs), which are basically different from the functions employed in the back-pocket model. The HAFs are polar-separable, and they have a polar-separable two-dimensional Fourier transform.

In particular, the VRF is complex valued in a different sense compared to the complex valued Gabor functions, whose paired phase/quadrature components are the real and imaginary parts of the complex envelope with respect to a spatial modulating grating harmonic carrier for a fixed orientation. The VRF have phase quadrature components that are geometrically orthogonal in the image plane. Therefore, they are steerable in azimuth by multiplication by a complex number [74].

The scope of this unusual approach is to capture the blur perception phenomenon with an essential, handy model. This approach provides a functional model of the discomfort phenomenon due to the optical blur, aimed to quantitatively predict psycho-physical findings irrespective of the underlying mechanisms.

To verify the limits of the present approach, the model-based discomfort predictions were first compared to empirical data about the subjective quality loss of blurred images, which is supposed to be highly correlated with the discomfort. These data are available in organized databases containing the results of experimental sessions finalized to Image Quality Assessment (IQA), conducted for the multimedia industry purposes [97, 10, 39]. Then the model is applied to empirical data about the visual discomfort. The results of these experiments confirm the validity of the approach.

As the luminance plays a dominant role for the localizability of patterns, for the sake of simplicity only the luminance component of the images is accounted for.

The retina in the *foveal vision* is abstractly modeled as a rectangular grid of receptors, whose position is individuated by the coordinate pair $\mathbf{p} \equiv (x_1, x_2)$. Receptors are regularly spaced *one arcmin* apart. The density of 60 *receptors/degree* assures that all the image information within the 30 (=60/2) *cycles/degree* spatial bandwidth is captured by the retina, according to the Nyquist sampling rule.

For any $\mathbf{p}$, the RF calculates a weighted sum of the luminance $\mathbf{I}(\mathbf{p})$ in a neighborhood of $\mathbf{p}$, yielding a *visual map* $y(\mathbf{p})$. This operation corresponds mathematically to a spatial *convolution*, indicated by the symbol $*$, between $I(\mathbf{p})$ and the visual map of a single lighting point in the dark, indicated by $h(\mathbf{p})$ and referred to as the Point Spread Function (PSF) of the RF:

$$y(\mathbf{p}) = I(\mathbf{p}) * h(\mathbf{p}) \,. \tag{6.1}$$

As we said, the RF model considered here is a harmonic angular filter (HAF). HAF filters are complex valued functions, i.e., they represent *pairs of real filters* [44].

With reference to the spatial frequency domain, defined by the horizontal and vertical frequencies f_1, f_2, and to the polar coordinates $\rho = \sqrt{f_1^2 + f_2^2}$ and $\vartheta = \arctan\left(\frac{f_2}{f_1}\right)$, where ρ [22] is referred to as the *radial frequency* and ϑ is the azimuth,

the Fourier spectrum $H(\rho, \vartheta)$ of the RF model is defined as

$$H(\rho, \vartheta) = j2\pi \left(\rho e^{-S_G^2 \rho^2} \cdot e^{j\vartheta} \right) \tag{6.2}$$

where j is the imaginary unit, and s_G is a parameter. This spectrum is *polar separable*, i.e., it is the product of a function of the radial frequency, and a function of the azimuth. One outstanding feature of this RF model is that it is also polar separable in the space domain [23]. In fact, taking the inverse Fourier transform of $H(\rho, \vartheta)$, with reference to the space polar coordinates $r = \sqrt{x_1^2 + x_2^2}$ and $\varphi = tg^{-1}\frac{x_2}{x_1}$, the PSF of the RF is:

$$h(r, \varphi) = -2\frac{\pi^3}{s_G^4} \left(r e^{-\frac{r^2 \pi}{s_G{}^2}} \cdot e^{j\varphi} \right) \tag{6.3}$$

which has the same shape of the VNTF, except for a scale factor.

This RF is used herein as a *functional* spatial vision model and, for this reason, it is referred to as VRF. For short, the term "VRF" will be used in the following to indicate also its PSF and its HAF shape, whereas its Fourier transform $H(\rho, \vartheta)$ will be referred to as Virtual Neural Transfer Function (VNTF), because it represents the *spatial frequency response* of the VRF.

The magnitude and the real components of the VRF are displayed in the upper row of Fig.6.1, where the ideal retinal grid is shown in the background. In the same figure, the magnitude, the real and the imaginary parts of the VNTF are displayed in the lower row.

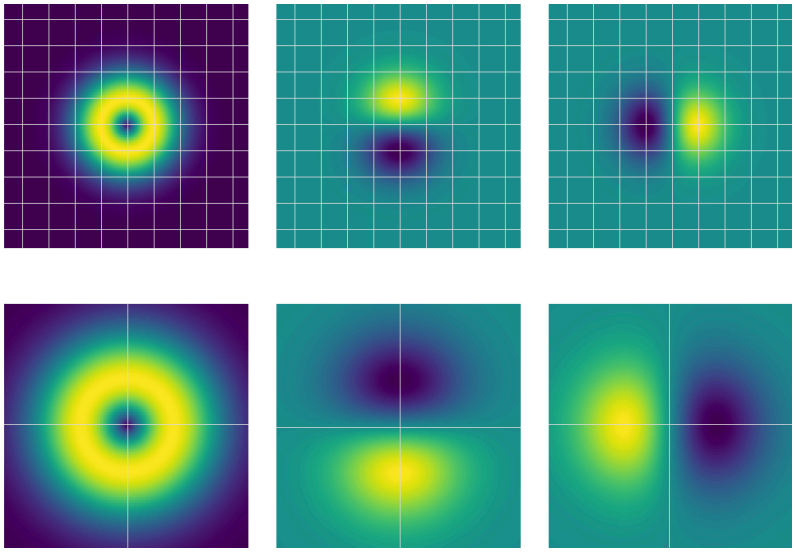

Figure 6.1. Upper row: the magnitude of the VRF, its real and imaginary parts referred to the ideal retina grid. Lower row: the magnitude of the VNTF, its real and imaginary parts in the spatial frequency plane. The vertical and horizontal spatial frequencies span the $(-30, 30)$ *cycles/degree* interval.

The magnitude of the VNTF frequency response versus the radial frequency is the same in any orientation. It is displayed in Fig.6.2.

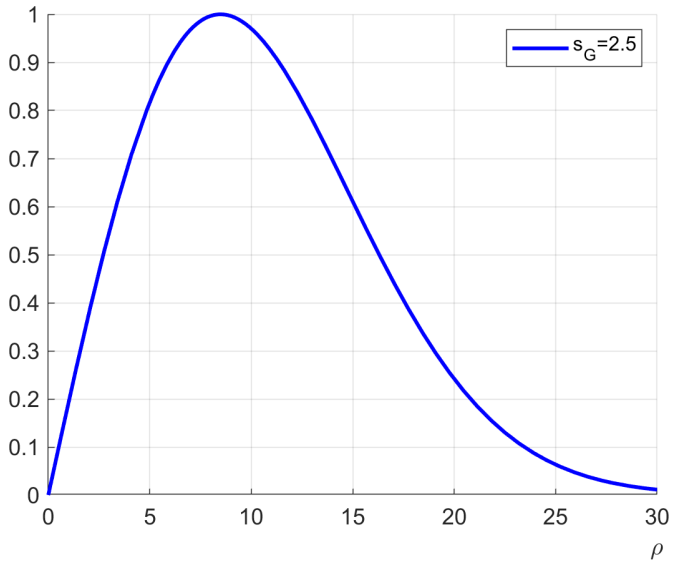

Figure 6.2. The radial frequency response magnitude of the VNTF for $s_G = 2.5$ *arcmin* normalized with respect to its maximum value.

Here, and from now on, s_G is assumed equal to $s_G = 2.5$ *arcmin*, unless otherwise noted. This choice sets the maximum of the radial frequency response magnitude at about 8.5 *cycles/degree* in the radial frequency, according to the experimental data provided in [11] and [107].

In the lowest spatial frequency range, the magnitude of the VNTF increases linearly. At higher spatial frequencies, the VNTF exhibits a soft-decaying low-pass behavior, reaching an attenuation of about 40 dB at 30 *cycles/degree* at the Nyquist frequency.

This behavior can be interpreted by regarding the VNTF as the cascade of two basic operators:

- an orientation selective *complex gradient* operator $\left(\frac{\partial}{\partial x_1} + j\frac{\partial}{\partial x_2}\right)$ [78], whose frequency response is obtained by the Fourier transform derivation rule:

$$j2\pi f_1 + j(j2\pi f_2) = j(2\pi\rho cos\vartheta + j2\pi\rho sin\vartheta) = j2\pi\rho e^{j\vartheta} \; ; \tag{6.4}$$

- a radial frequency selective *Gaussian smoothing* operator, represented by the frequency response:

$$G(\rho, \theta) = e^{-S_G^2\rho^2} \tag{6.5}$$

  which is responsible of a *neural blur*.

Therefore, the visual map $y(\mathbf{p})$ of the VRF is globally interpreted as a *complex, Gaussian-smoothed gradient field* associated to the retinal image. The parameter s_G will be referred to as the *spread* of the VRF, or as the *neural spread*.

Different from the *complex Gabor* functions, whose paired real components are aligned each other [24, 104], the paired real components of the VRF are mutually

orthogonal in the image plane (Fig.6.1). As such, the VRF is *steerable* [82] and more specifically *scalar steerable*, i.e., it rotates in azimuth by multiplication by a complex number.

$$\begin{array}{c} h(r,\varphi-\alpha)=h(r,\varphi)e^{j\alpha}\ ; \\ H(\rho,\vartheta-\alpha)=H(\rho,\vartheta)e^{j\alpha}\ . \end{array} \tag{6.6}$$

This implies that rotated version of the VRFs by a generic azimuth α are obtained by linear combinations of real components $Re\{h(r,\varphi)\}$ and $Im\{h(r,\varphi)\}$:

$$\begin{array}{c} Re\{h(r,\varphi-\alpha)\}=Re\{h(r,\varphi)\}cos\alpha-Im\{h(r,\varphi)\}sin\alpha\ ; \\ Im\{h(r,\varphi-\alpha)\}=Re\{h(r,\varphi)\}sin\alpha+Im\{h(r,\varphi)\}cos\alpha\ . \end{array} \tag{6.7}$$

The VRF model materializes into the visual map $y(\mathbf{p})$. In the correspondence of an *edge* of $I(\mathbf{p})$, the magnitude of $y(\mathbf{p})$ measures the *edge strength*, while the phase $tg^{-1}\left[\frac{Im\{y(\mathbf{p})\}}{Re\{y(\mathbf{p})\}}\right]$ indicates the orientation orthogonal to the edge [47, 45, 66]. In the example of Fig.6.3, the luminance of a retinal image (left) and its visual map in false color (right) are displayed. In the visual map the gradient strength is indicated by the luminance component whereas, for visual immediateness, only the direction of the gradient in the interval $[0,\pi)$ is indicated with the hue color component.

The visual map $y(\mathbf{p})$ is a *near-complete*, *sparse representation* of $I(\mathbf{p})$. In fact, except for its mean value, $I(\mathbf{p})$ can be fully recovered from $y(\mathbf{p})$ by *spectral inversion* (i.e., by division by $H(\rho,\vartheta)$ in the frequency domain).

Standard multichannel spatial vision models (see [79] for a historical account) follow a tomographic-like approach. Around any point, they analyze the image from a limited number of azimuthal views. For each view, they apply co-oriented filters tuned to different bands. Their outputs are then combined in different ways.

In comparison, around any point, the VRF performs a full-band *radial tomographic* analysis in every orientation, as described in [21, 44]. For the scope of this work, the outstanding advantage of the VRF is that its output (the visual map) allows straightforward computation of the PFI [69] as described in the next section.

The VRF is the simplest functional spatial vision model based on HAFs. It coincides with the first order component of the orthogonal family of the Laguerre Gauss (LG) functions [92, 46, 63] or, equivalently, of the 2D Hermite functions, which span the same signal space [61, 29]. Higher order LG analysis provides *functional* spatial vision models oriented to structures more complex that simple edges [28, 69]. HAF based *wavelets* can be also used for multiresolution analysis [46].

6.5 The Positional Fisher Information

The adoption of a polar separable complex-valued receptive field model, and of a visual information loss criterion, led to compact theoretical formulas for the prediction of the blur visual discomfort for natural scenes, exhibiting a good predictive power faced to several independent experimental data. The approach may provide coarse estimates of the discomfort caused by undesired blur, for optical correction in natural vision, and for calibration of image reproduction apparatus.

A *detail* $d_p(q)$ of a visual map $y(\mathbf{p})$ is formally defined as:

$$d_{\mathbf{p}}(\mathbf{q})=w_{\mathbf{p}}(\mathbf{q})\cdot y(\mathbf{p}-\mathbf{q}) \tag{6.8}$$

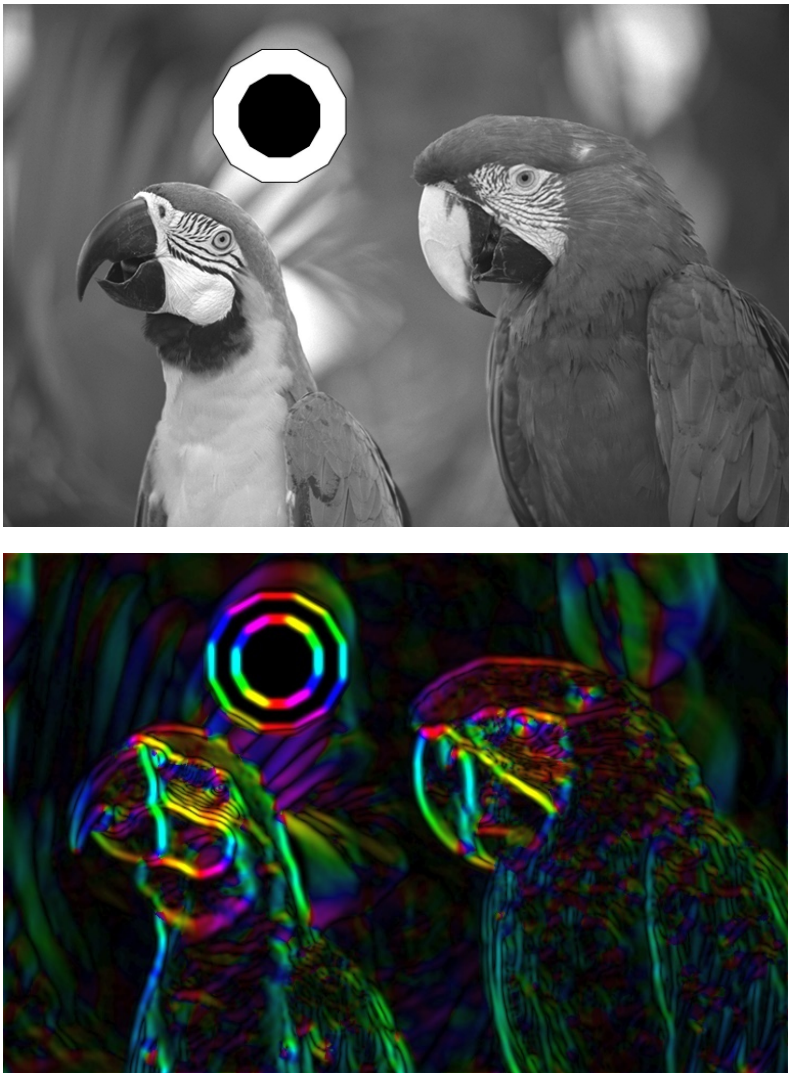

Figure 6.3. The luminance of an image (left) of the database LIVE [81] compared to the corresponding visual map $y(\mathbf{p})$ (right), where the magnitude of the edges is coded into luminance, and their direction, in the interval $[0, \pi)$, into hue. The hue/direction code is read in the edges of the upper-left superimposed polygonal.

where $w_{\mathbf{p}}(\mathbf{q})$ is a sampling window centered on $\mathbf{p}$.

A comprehensive calculus of the Fisher Information of a detail about its position, orientation, and scale, in the presence of a background Gaussian white noise, was provided in [69]. As specified in [28], the *total* PFI of a detail is calculated as:

$$\psi(\mathbf{p}) = \frac{\lambda(\mathbf{p})}{\sigma_V^2} \tag{6.9}$$

where $\lambda(\mathbf{p})$ is the smoothed gradient energy of the detail, computed as

$$\lambda(\mathbf{p}) = \sum_{\mathbf{q}} w_{\mathbf{p}}(\mathbf{q})^2 \left| y(\mathbf{p} - \mathbf{q}) \right|^2 \tag{6.10}$$

and σ_V^2 is the variance of the background noise[1].

The inverse square root of the PFI

$$e_{MIN}(\mathbf{p}) = \sqrt{\frac{1}{\psi(\mathbf{p})}} = \sqrt{\frac{\sigma_V^2}{\lambda(\mathbf{p})}} \tag{6.11}$$

represents the minimum standard deviation of the detail position error $e_{MIN}(\mathbf{p})$, achievable with an unbiased estimator, irrespective of the employed estimation method [91]. Thus, the quantity

$$\sqrt{\psi(\mathbf{p})} = \frac{1}{e_{MIN}(\mathbf{p})} \tag{6.12}$$

measures the *certainty* of the detail position in the visual plane. For a given amount of background noise, the higher the smoothed gradient energy, the higher the PFI, the greater the certainty about the detail position.

The smoothed gradient energy $\lambda(\mathbf{p})$ is significantly expressed by the (two-dimensional) Fourier transform of the detail. Applying the Parseval theorem (which equates the energy calculated in the space and in the spatial frequency domains) the PFI of the detail represented by its Fourier transform $D_{\mathbf{p}}(\rho, \vartheta)$ is

$$\psi(\mathbf{p}) = \frac{1}{\sigma_V^2} \int_0^{2\pi} \int_0^{+\infty} \rho^2 \left|G(\rho, \vartheta)\right|^2 \left|D_{\mathbf{p}}(\rho, \vartheta)\right|^2 \left|B(\rho, \vartheta)\right|^2 \rho d\rho d\vartheta \tag{6.13}$$

where $B(\rho, \vartheta)$ is the Optical Transfer Function (OTF), i.e., the spatial frequency response of the optical system, from the observed object to the retina [100]. Its inverse 2D transform is the Optical PSF $b(r, \varphi)$.

The overall OTF of a vision system, including the human eye, is a combination of the OTFs of cascaded subsystems including the OTF of correcting lenses, the OTF of an imaging system, the OTF of a display system, etc..

Under the hypothesis of linearity, the overall OTF is the product of the single OTFs. In other terms, the overall Optical PSF is the cascaded 2D convolution of the single PSFs.

[1]Strictly speaking, the PFI of a detail does not coincide with the PFI of the pattern "contained in" the detail. In fact, the window itself carries its own PFI. In the following, the latter contribution will be neglected, assuming that the window is so smooth that the information carried by its shape is small with respect to the information carried by the captured pattern.

Chapter 7

PFI Equivalence

7.1 The natural scene spectrum and the PFI equivalence

7.1.1 The PFI of natural scenes

For a generic natural image, the average PFI calculated on a group of N details visited during the visual exploration is:

$$\begin{aligned} &\frac{\sum_{\mathbf{p}} \lambda(\mathbf{p})}{N\sigma_V^2} = \\ &= \frac{1}{\sigma_V^2} \int_0^{2\pi} \int_0^{+\infty} \rho^2 \left|G(\rho,\vartheta)\right|^2 \left|D_N(\rho,\vartheta)\right|^2 \left|B(\rho,\vartheta)\right|^2 \rho d\rho d\vartheta \end{aligned} \tag{7.1}$$

where

$$\left|D_N(\rho,\vartheta)\right|^2 = \frac{1}{N}\sum_{\mathbf{p}} \left|D_{\mathbf{p}}(\rho,\vartheta)\right|^2 \tag{7.2}$$

is the average energy spectrum of the N visited details. The expected value Ψ of the PFI over the random set of natural images is defined as:

$$\begin{aligned} &\Psi \doteq E\left\{\frac{\sum_{\mathbf{p}} \lambda(\mathbf{p})}{N\sigma_V^2}\right\} = \\ &= \frac{1}{\sigma_V^2} \int_0^{2\pi} \int_0^{+\infty} \rho^2 \left|G(\rho,\vartheta)\right|^2 E\left\{\left|D_N(\rho,\vartheta)\right|^2\right\} \left|B(\rho,\vartheta)\right|^2 \rho d\rho d\vartheta \end{aligned} \tag{7.3}$$

where $E\left\{\left|D_N(\rho,\vartheta)\right|^2\right\}$ denotes the expected value of the energy spectrum of the visited details over the random set of natural images.

It is well known that the expected value of the energy spectrum of natural images is proportional to the inverse of the square of the radial frequency, i.e., to $\frac{1}{\rho^2}$. The generality of this spectral distribution is supported by theoretical arguments [83, 31, 33, 51, 90, 5].

Here, this property is attributed to the *average spectrum* of the *visited details*. Posing, for the sake of generality [88]

$$E\left\{\left|D_N(\rho,\vartheta)\right|^2\right\} = f(\vartheta)\frac{1}{\rho^2}\ ; \tag{7.4}$$

it follows that

$$\Psi = \frac{1}{\sigma_V^2} \int_0^{2\pi} f(\vartheta) \int_0^{+\infty} |G(\rho,\vartheta)|^2 \, |B(\rho,\vartheta)|^2 \, \rho d\rho d\vartheta \tag{7.5}$$

and, in the absence of blur:

$$\Psi_0 = \frac{1}{\sigma_V^2} \int_0^{2\pi} f(\vartheta) \int_0^{+\infty} |G(\rho,\vartheta)|^2 \, \rho d\rho d\vartheta \; . \tag{7.6}$$

In the case of isotropic Gaussian blur, the OTF is

$$B(\rho,\vartheta) = e^{-s_B^2 \rho^2} \tag{7.7}$$

where s_B is referred to as the *optical spread*. Therefore,

$$\Psi = \frac{F}{\sigma_V^2} \int_0^{+\infty} e^{-2(s_G^2+s_B^2)\rho^2} \rho d\rho \tag{7.8}$$

$$\Psi_0 = \frac{F}{\sigma_V^2} \int_0^{+\infty} e^{-2s_G^2\rho^2} \rho d\rho \tag{7.9}$$

where the coefficient F is

$$F = \int_0^{2\pi} f(\vartheta) \, d\vartheta \; . \tag{7.10}$$

Finally, from the equality

$$\int_0^{+\infty} e^{-2(s_G^2+s_B^2)\rho^2} \rho d\rho = \frac{1}{4(s_G^2+s_B^2)} \tag{7.11}$$

it follows that

$$\frac{\Psi}{\Psi_0} = \frac{s_G^2}{s_G^2+s_B^2} \; . \tag{7.12}$$

7.1.2 The PFI equivalence

The above result is so simple owing to the Gaussian shape of blur. However, having assumed that the blur discomfort depends only on the PFI loss, the discomfort due to different types of blur could be predicted by the same formula applying the concept of *PFI equivalence*.

Definition 7.1.1 (PFI Equivalence)**.** A blur characterized by the generic OTF $B(\rho,\vartheta)$ is said to be PFI equivalent to an isotropic Gaussian blur with standard deviation (spread) s_B if it yields the same expected PFI for natural images, i.e.:

$$\begin{aligned} &\int_0^{2\pi} \int_0^{+\infty} e^{-2s_G^2\rho^2} \, |B(\rho,\vartheta))|^2 \, \rho d\rho d\vartheta = \\ &= \int_0^{2\pi} \int_0^{+\infty} e^{-2(s_G^2+s_B^2)\rho^2} \rho d\rho d\vartheta \; . \end{aligned} \tag{7.13}$$

This equivalence criterion is intuitive. It equals the energies of the actual OTF and of an isotropic Gaussian OTF, both weighted by the squared magnitude of the VNTF. In particular, the PFI equivalence does not depend on the phase of the OTF.

An important example of PFI equivalence is the one of the out-of-focus blur, whose PSF is modeled as a cylinder of unitary volume and radius R. It is referred also to as the *disc blur*, or the *sinc blur*, and is characterized by the following OTF:

$$B(\rho, \vartheta) = 2\frac{J_1(2\pi\rho R)}{2\pi\rho R} \ . \tag{7.14}$$

Equating the PFI of the sinc blur and of the Gaussian isotropic blur yields

$$\begin{aligned} &4\int_0^{+\infty} \left(\frac{J_1(2\pi\rho R)}{2\pi\rho R}\right)^2 e^{-2s_G^2\rho^2}\rho d\rho = \\ &= \int_0^{+\infty} e^{-2(s_G^2+s_B^2)\rho^2}\rho d\rho = \frac{1}{4(s_G^2+s_B^2)} \ . \end{aligned} \tag{7.15}$$

The left side integral is not available in closed form. A careful numerical integration provides a value of the optical spread s_B of the isotropic Gaussian blur as a function of its PFI equivalent sinc blur of radius R, as plotted in Fig.7.1 and is roughly expressed by the rule $\frac{s_B}{R} \approx \frac{3}{8}$. In the same figure some subjective equivalence judgments averaged over a pool of six observers are also reported. These empirical data refer to synthetic images whose contours emulate the ones of natural images [68].

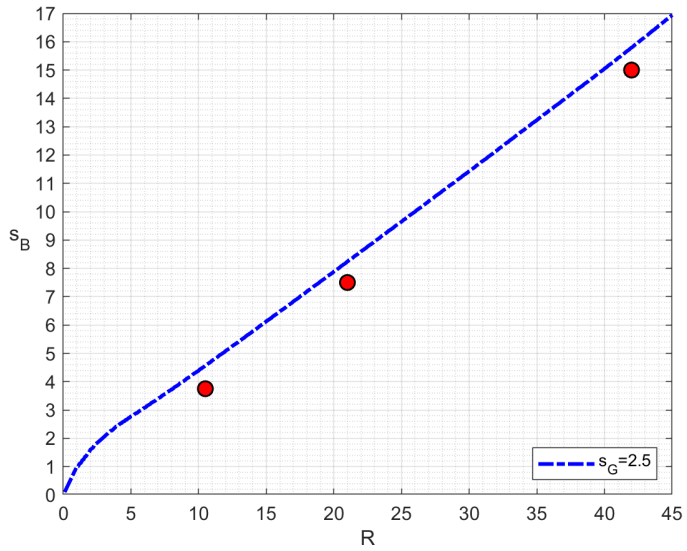

Figure 7.1. The radius of the sinc blur versus the spread of its PFI equivalent Gaussian blur. The red points indicate some average results of experiments conducted with synthetic images emulating natural images [68].

From a perceptual viewpoint, it appears that this equivalence works generally well. Here, to provide the reader with visual examples, the most critical cases in

the Tampere image database (TID2013) [75], are reported in Fig.7.2, including the non-natural image i25 as a benchmark. Notice that some grating patterns are cancelled out or amplified by the sidelobes of the OTF of the sinc blur in comparison to the Gaussian blur. Notice also that these effects are not present in the natural out-of-focus blur, owing to the apodization of the pupil, which attenuates the sidelobes [110].

Figure 7.2. Upper row: two original images. Second and third rows: Gaussian blurred image (left) and Fisher equivalent sinc blurred images (right). The values of R were chosen to put into evidence the effects of the spectral sidelobes of the sinc blur (see the encircled patterns). Natural image (i19 of the database): $R = 6$ *arcmin*, corresponding to $s_B = 3.1$ *arcmin*. Synthetic image (i25 of the database): $R = 7$ *arcmin*, corresponding to $s_B = 3.4$ *arcmin*.

A second example of PFI equivalence regards the non-isotropic Gaussian blur, referred to as *astigmatic* Gaussian blur. For the sake of simplicity, this equivalence is calculated here only for the case of isotropic image spectral energy distribution. Using for convenience the Cartesian coordinates, the OTF of this blur is:

$$B(f_1, f_2) = e^{-2(s_V^2 f_1^2 + s_H^2 f_2^2)} \tag{7.16}$$

where s_H and s_V are the horizontal and vertical optical spreads. A straightforward algebraic analysis shows that the astigmatic Gaussian blur is PFI equivalent to the isotropic Gaussian blur with spread

$$s_B = \sqrt{\sqrt{s_G^4 + s_G^2(s_H^2 + s_V^2) + s_V^2 s_H^2} - s_G^2}\,. \tag{7.17}$$

An example is provided in Fig.7.3. Two versions of an original image, respectively affected by an astigmatic Gaussian blur with $s_H = 4$, $s_V = 1$ and with $s_H = 1$, $s_V = 4$ are compared to the same image affected by their PFI equivalent isotropic Gaussian blur. Notice that the sea waves are better localized in presence of horizontal blur, while masts are better localized in presence of vertical blur. The average localizability loss is visually balanced by the isotropic Gaussian blur.

Figure 7.3. A unblurred image (upper left) and its astigmatic blurred versions. For $s_H = 4$, $s_V = 1$ (horizontal blur, upper right), and for $s_H = 1$, $s_V = 4$ (vertical blur, lower left). They are PFI equivalent to the isotropic Gassian blurred version shown in the lower right image with $s_B = 2.54$.

Chapter 8

Experimental Measures

8.1 Measuring blur discomfort

Consider the pooling Γ of the samples the observed details:

$$\Gamma = \frac{1}{N}\left[\sum_{\mathbf{p}}\sum_{\mathbf{q}}\sqrt{w_{\mathbf{p}}(\mathbf{q})}y\left(\mathbf{p}-\mathbf{q}\right)\right] \tag{8.1}$$

and the quantity

$$\Gamma\Gamma^{*} = \frac{1}{N^{2}}\left[\sum_{\mathbf{p}}\sum_{\mathbf{q}}\sqrt{w_{\mathbf{p}}(\mathbf{q})}y\left(\mathbf{p}-\mathbf{q}\right)\right]\left[\sum_{\mathbf{u}}\sum_{\mathbf{v}}\sqrt{w_{\mathbf{p}}(\mathbf{v})}y^{*}\left(\mathbf{u}-\mathbf{v}\right)\right]^{*}. \tag{8.2}$$

The terms of the cross products $y\left(\mathbf{p}-\mathbf{q}\right)\cdot y^{*}\left(\mathbf{u}-\mathbf{v}\right)$ of this sum become uncorrelated as far as the mutual distance of their locations exceeds the blur spread. For sufficiently high N, the contribution of these products to the above average tend to vanish, so that the following approximation holds

$$\Gamma\Gamma^{*} \simeq \sigma_{V}^{2}\Psi. \tag{8.3}$$

The consequence is that the computation of the uncertainty $\sqrt{\Psi}$ from $y\left(\mathbf{p}\right)$ is straightforward and simple:

$$\sqrt{\Psi} \propto \sum_{\mathbf{p}}\sum_{\mathbf{q}}\sqrt{w_{\mathbf{p}}(\mathbf{q})}\left|y\left(\mathbf{p}-\mathbf{q}\right)\right|. \tag{8.4}$$

As far as the knowledge of $\sqrt{\Psi_{0}}$ is concerned, for natural images it can be blindly extrapolated by means of simple operations [111]. However, it is conceivable that it is simply guessed from the subject's visual experience.

Following the above considerations, the amount of blur discomfort is assumed *proportional* to the *relative certainty loss* about the details position[1] defined as:

$$\varepsilon = \frac{\sqrt{\Psi_{0}}-\sqrt{\Psi}}{\sqrt{\Psi_{0}}} = 1-\sqrt{\frac{\Psi}{\Psi_{0}}}\,. \tag{8.5}$$

[1] An exponentiated version of such a measure was employed in [27].

This assumption is suggested by the belief that the perceived cost of wrong localization is proportional to the uncertainty of the Euclidean distances, at least for small errors.

Then, for the random set of natural images and for isotropic Gaussian blur, using (7.12) the relative certainty loss takes the form of the following a-dimensional *discomfort index*:

$$\varepsilon = 1 - \sqrt{\frac{\Psi}{\Psi_0}} = 1 - \sqrt{\frac{1}{1 + \left(\frac{s_B}{s_G}\right)^2}} \,. \tag{8.6}$$

This index ranges between 0 (in the absence of blur) to 1 (for diverging blur). It depends only on the *normalized blur*, i.e., the ratio between the optical spread on the retinal image and the neural spread of the VRF.

The optical blur spread s_B is subject to change by the action of the natural accommodation system and, in the context of a composite optical system, by the action of technical devices. The neural spread s_G plays instead the role of an inner reference. At a glance, it could be argued that it is a stable parameter. However, some experiments indicate that the s_G value is adaptive [105]. It appears that the visual adaption to a blurred image causes a dilation of the spread s_G, leading to a reduction of the normalized blur, so that the spectrum of the image looks wider and the image sharper.

The discomfort formula is now applied to the blur caused by out-of-focus condition of the eye optics, i.e., in *natural vision*. Using the geometrical arguments of [86], it is deduced that the radius R of the out-of-focus optical PSF in arcmin is

$$R = 1.71 p \, |D| \tag{8.7}$$

where p is the pupil diameter (in mm) and the D is the out-of-focus measure in diopter units (m^{-1}).

Applying the $\frac{s_B}{R} \approx \frac{3}{8}$ rule it follows that the spread of the PFI equivalent blur of the out-of-focus blur measured in diopters is approximated as

$$s_B \approx 0.64 p \, |D| \tag{8.8}$$

so that a coarse estimate of the blur discomfort in natural vision is:

$$\varepsilon = 1 - \sqrt{\frac{1}{1 + \left(0.64 \frac{p}{s_G} D\right)^2}} \,. \tag{8.9}$$

In the chart of Fig.8.1 this theoretical discomfort index ε is plotted versus D for different pupil diameters. The scale of ε is expressed in centesimal units.

This chart essentially says that, if the out-of-focus discomfort is proportional to the relative certainty loss, as assumed, it is not linear with the blur spread. Moreover, since the sensitivity of the discomfort index with respect to the normalized spread $\xi \doteq \frac{s_B}{s_G}$ is calculated as

$$\frac{d\varepsilon}{d\xi} = \xi \left[\frac{1}{1 + \xi^2}\right]^{\frac{3}{2}} \;; \tag{8.10}$$

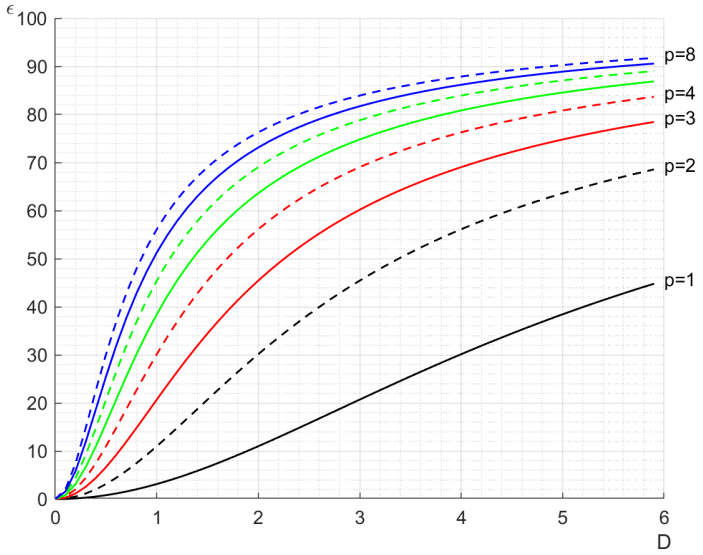

Figure 8.1. The value of the theoretical discomfort index in centesimal units versus the diopters (measured in m^{-1}) for different values of the pupil diameter (measured in mm). (A typical pupil diameter when reading at normal illumination is 3 mm).

the increment $\Delta\xi$ necessary to produce a given increment $\Delta\varepsilon$ is approximated as:

$$\Delta\xi = \frac{1}{\xi}\left[1+\xi^2\right]^{\frac{3}{2}}\Delta\varepsilon\ . \tag{8.11}$$

This increment exhibits a typical "dipper shape" [70, 84] as shown in Fig.8.2. The theoretical minimum occurs in the correspondence of the normalized spread value $\xi = \frac{s_B}{s_G} = \frac{1}{\sqrt{2}}$.

In the plot of Fig.8.3, the curves of $\frac{\Delta s_B}{s_G}$ versus $\frac{s_B}{s_G}$ for different values of $\Delta\varepsilon$ are plotted.

In the above formulas, diffraction and aberration contributions to blur discomfort are considered negligible. Non-negligible aberrations could be accounted for by adopting parametric models for characterizing their PSFs [32, 101], and calculating the PFI equivalent blur. A simple example is the aberration due to astigmatic Gaussian blur provided in Sect.4, where blur is characterized by the two parameters s_H and s_V.

8.2 Experimental verification

The above results about the phenomenon of blur discomfort for natural images are derived by principles and assumptions. To assess their effectiveness in the reality, the theoretical model was first verified against empirical ratings of the *subjective quality loss* of images caused by blur, which are argued to be strictly related to the blur discomfort. These data were released in response to the growing demand by the media industry for reliable automatic IQA through objective metrics. Four

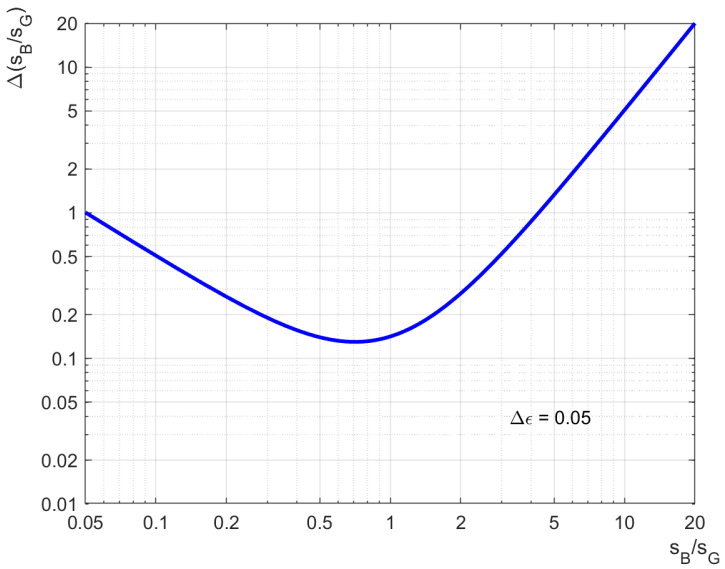

Figure 8.2. The theoretical increment $\Delta\left(\frac{s_B}{s_G}\right)$ versus $\frac{s_B}{s_G}$ for $\Delta\varepsilon = 0.05$ in a log/log scale. Curves for different values of $\Delta\varepsilon$ are obtained by vertical translation.

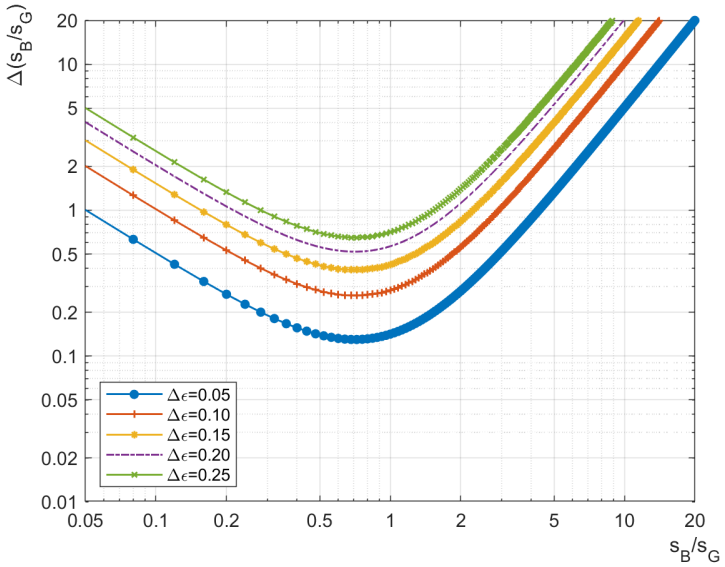

Figure 8.3. The theoretical increment of blur $\frac{\Delta s_B}{s_G}$ versus $\frac{s_B}{s_G}$ for different values of the discomfort index $\Delta\varepsilon$.

independent IQA databases were employed, based on different methodologies and protocols, and following different strategies to prevent biases and side effects.

IQA databases include images affected by Gaussian blur, which is considered sufficiently representative of the perceptual effect of blur in many technical applications.

Subsequently, the predictions of the model were compared to experimental data

where subjects were literally asked to rate "the *visual discomfort*" due to blur [71].

All these experiments do not account for blur discomfort secondary effects. Furthermore, in these experiments the eye optical blur of the observers is corrected if present. So, blur applied to the observed images emulates an undesired natural optical blur on the retina plane. In fact, under the hypothesis of linearity, the natural optical blur and the artificial blur applied to the observed images are interchangeable, because of the well-known property of cascaded convolutions.

Before presenting the experiments and discussing the results, it is essential to illustrate how the discomfort index must be scaled to fit the experimental settings of a database.

8.2.1 The Scaled Blur Discomfort Index

To this purpose, the following parametric "Scaled Blur Discomfort Index" (SBDI) is defined from (8.6):

$$SBDI \doteq a \left[1 - \sqrt{\frac{1}{1 + \left(\gamma^2 \frac{s_B}{s_G} \right)^2}} \right] \tag{8.12}$$

where the gain a fixes the *scoring scale* [80], and γ is the *Viewing Distance* (VD) defined as

$$\gamma \doteq \frac{\delta_0}{\delta} \tag{8.13}$$

where δ_0 is *nominal VD*, i.e., the distance from which the density of pixels projected on the retina matches the previously assumed density of the receptors (60/degree) and δ is the VD adopted in the experiment.

Differently stated, γ equals the ratio between the number of pixels viewed within one degree at distance δ_0 (60 pixels) and the number of pixels viewed within one degree at distance δ.

The role of the distance parameter is understood considering that the VRF spread on the image projected on the retina is $\frac{s_G}{\gamma}$, i.e., it increases proportionally with the VD. Conversely, the projection of the spread of blur applied to the observed image is γs_B, i.e., it is inversely proportional to the VD.

Unless specified, the parameters a and γ can be determined from data by regression.

8.2.2 The "subjective quality loss" experiments

The essential features of the employed IQA databases are illustrated below.

The LIVE IQA Database Release 2 (DBR2) [81] reports the quality ratings of 779 distorted versions of 29 reference images (included 145 blurred images) from about 23 subjects. Ratings of subjective quality loss with respect to reference images were expressed on a DMOS (Difference of Mean Opinion Score) scale ranging from 0 (perfect quality) to 100 (bad quality) using a *double stimulus* strategy.

The Tampere Image Database 2013 (TID2013) [75] contains 3000 distorted images, including 125 blurred images. Quality ratings were collected in five independent labs and on the internet using more than 300 subjects. They were asked to select the

best image between two distorted images in direct comparison to the reference image. The average quality scores were expressed on a Mean Opinion Score (MOS) scale ranging from 0 (bad quality) to 9 (perfect quality).

The Computational and Subjective Image Quality Database (CSIQ) [54] contains 30 reference images and 866 distorted versions, including 150 blurred images. The database includes 5000 ratings of 25 subjects, and the ratings, obtained by *comparative ratings between different images*, are reported in DMOS units.

The LIVE Multiply Distorted Image Quality Database (LIVE MD) [48] contains 15 reference images and 405 distorted images, including 45 blurred images, whose quality was rated by 37 subjects. The study was conducted using a *single stimulus with hidden reference* strategy, using DMOS scores.

All databases contain *natural images*, i.e., images representing *natural scenes*, except the image i25 of the TID2013 database (which was used as a benchmark in Fig.7.2).

In the graphs of Fig.8.4 the subjective quality loss prediction curves for the blurred images of the different databases are superposed to the empirical DMOS values, plotted versus the normalized spread s_B/s_G. These empirical DMOS data represent *average scores* of the pool of the observers. In addition, in the TID2013 and LIVE MD cases, averages with respect to the sample images are also indicated by filled circles.

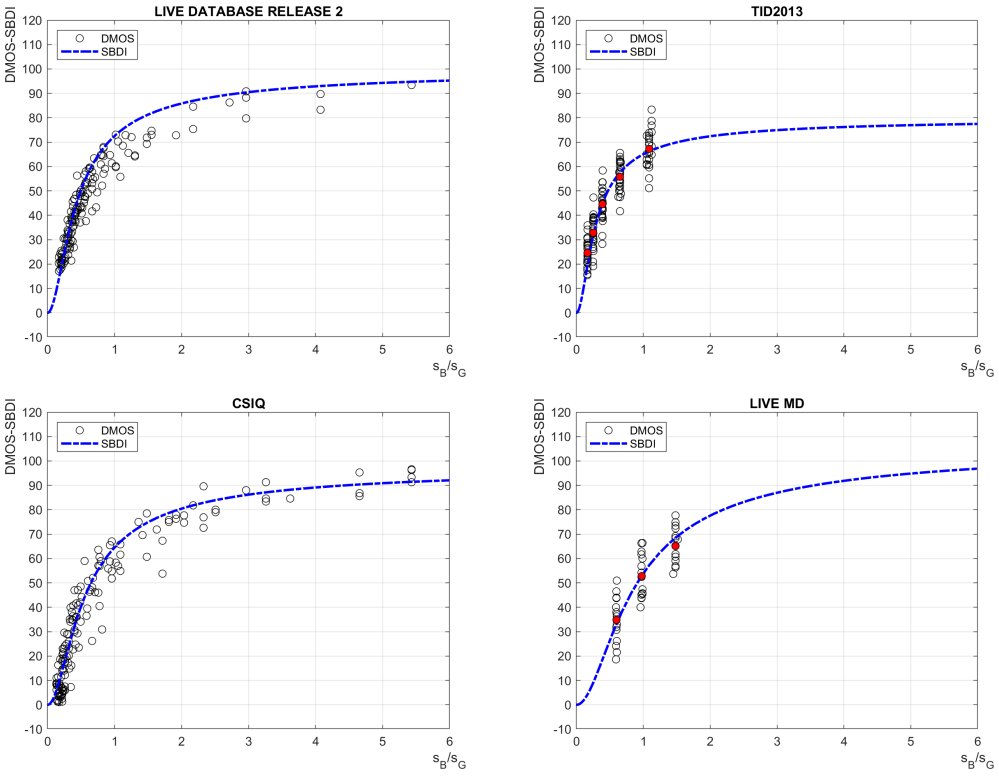

Figure 8.4. The empirical DMOS values of the subjective quality loss for the blurred images of the databases versus the normalized blur spread compared to the values predicted by the theoretical model (dashed curves). Averages with respect to images are indicated by the filled circles in the TID2013 and LIVE MD scatterplots.

In the TID2013 database, ratings are available as MOS values. The DMOS values were inferred considering that the best MOS ratings do not exceed 7.5 (see Fig.20 of [75]). Therefore, posing $MOS = 7.5$ in correspondence to $DMOS = 0$ and $MOS = 0$ in correspondence to $DMOS = 100$ yields

$$DMOS = (100/7.5) \cdot (7.5 - MOS) \, . \tag{8.14}$$

In the CISQ database, the DMOS was normalized between its minimum and maximum empirical value.

The blur values of the LIVE MD database were not available. They were estimated through a regularized spectral division of the blurred images with the unblurred ones.

The fitting of the empirical data with the theoretical prediction of the subjective quality loss is substantially linear, as evidenced by the scatterplots of the DMOS empirical ratings versus the predicted ones for all the images contained in the different databases (Fig.8.5). In the TID2013 and LIVE MD scatterplots, averages with respect to the sample images are indicated by filled circles.

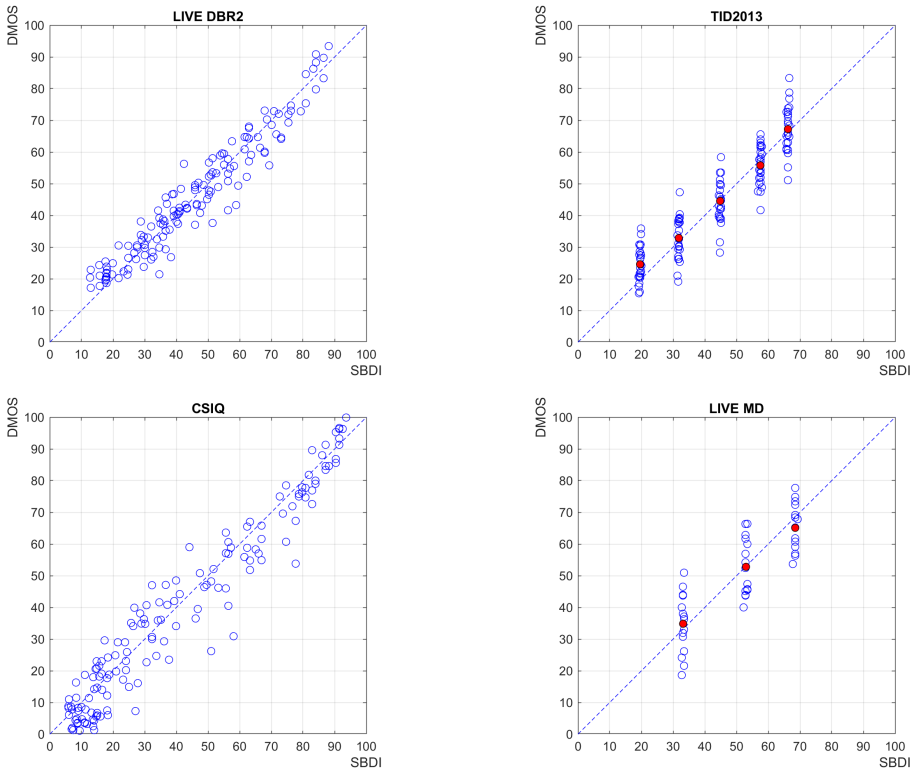

Figure 8.5. The scatterplots of the DMOS ratings versus the corresponding SBDI values predicted by the model for each blurred image in the different databases. Averages with respect to images are indicated by filled circles in the TID2013 and LIVE MD scatterplots.

Finally, in the Fig.8.6, a comprehensive scatterplot of the empirical data of the four databases is displayed.

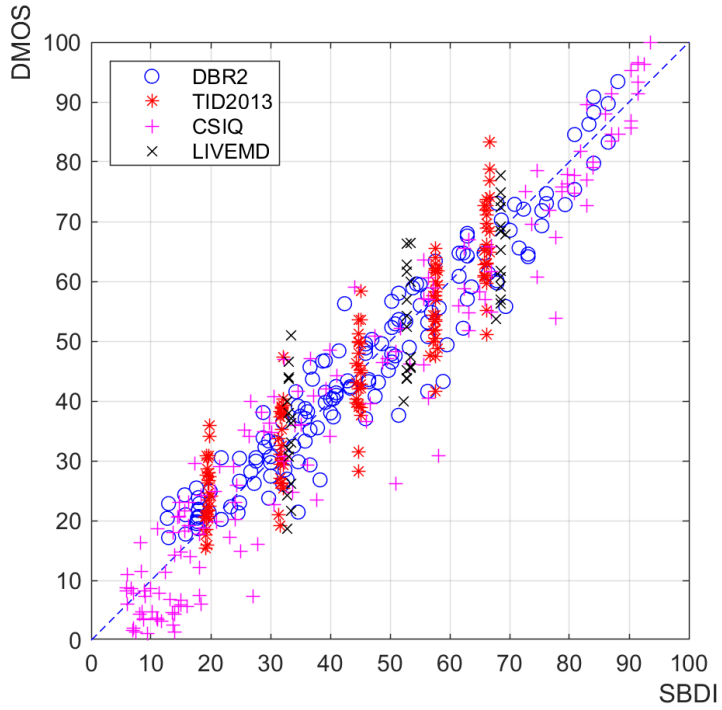

Figure 8.6. The cumulative scatterplot of all the ratings of the different databases versus their values predicted by the theoretical model.

In Tab.8.1 the most relevant data about the experimental validation of the model are resumed. The Pearson Linear Correlation Coefficient (PLCC) and the Root Mean Square Error (RMSE) between theoretical and empirical data are also provided. The "claimed" normalized VDs were calculated with the information about the experimental settings provided by the respective authors[2]. The sign "–" stands for "not available"[3].

Table 8.1. Summary of the experimental verification for the four databases and $s_G = 2.5$

DATABASE	a estimated	$\frac{\delta}{\delta_0}$ claimed	$\frac{\delta}{\delta_0}$ estimated	PLCC	RMSE
LIVE DBR2	93	0.46/0.57	0.53	0.96	5.44
TID2013	80	–	0.44	0.92	6.82
CSIQ	98	0.65	0.60	0.97	7.47
LIVE MD	107	0.84	0.76	0.83	8.57
ALL	–	–	–	0.95	6.44

[2]The VDs of CISQ and LIVE MD are slightly (10%) underestimated. The maximum absolute difference is within ten centimeters. However, the "physical" VD does also depend on the physical dimension of the pixels of the screen.

[3]The Authors are grateful to Prof. Ponomarenko for providing details about the blur settings in the TID2013 database in a personal communication.

8.2.3 The "blur discomfort" experiments

In [71] the results of some experiments aimed to investigate the relationship between visual discomfort judgments and image manipulations causing blur are reported. In particular "Experiment 3" regards *natural images*.

In these experiments sixty natural images were taken from a database whose images pertain to two general categories: distant natural scenes and closeups natural scenes. In particular, ten images from the first category whose spectral energy has a mean radial frequency decay of the kind $\left(\frac{1}{\rho^{\beta}}\right)^2$ with $\beta = 1.39$, and ten images from the second category, characterized by $\beta = 0.95$, were selected. The scope of this diversity was to see if visual discomfort judgments would depend on deviation from ideal spectral decay of natural scenes ($\beta = 1$).

During experimental sessions, thirteen subjects were asked to formulate *discomfort judgments* following a *pairwise comparison* strategy.

Differently from the preceding experiments, ratings are averaged not only over the pool of observers, but also over groups of images. One average regards the distant natural scenes, and the other one regards the closeups natural scenes. As in the TID2013 and in the LIVE MD experiments, few (three) Gaussian blur values were employed, with standard deviations 8, 16, and 32 *cycles/degree* in the spatial frequency domain, corresponding respectively to $s_B = 3.75$, $s_B = 1.875$ and $s_B = 0.9375$ *arcmin* in the spatial domain. In the Fig.8.7 the results of these experiments are reported, along with the discomfort predicted by the SBDI index. The results were expressed in the Thurnstone scale [89], according to the method employed for calculating the discomfort scores, and then converted into the discomfort scale of the SBDI by the following affine transformation:

$$SBDI = 50 \cdot (Thurnstone + 1) \; . \tag{8.15}$$

The VD, determined by regression, is $\frac{\delta}{\delta_0} = 0.6$, corresponding to $\gamma = 1.66$.

The scatterplots of the discomfort ratings versus the normalized Gaussian blur spread are displayed in Fig.8.7 for the groups of images contained in the database, along with the superimposed theoretical prediction curve. The triangles refer to the distant natural scene category, and the circles to the closeups scene category.

The fitting of the empirical data with the theoretical prediction of the subjective quality loss is substantially linear, as evidenced by the scatterplots of the empirical ratings versus the predicted ones for all the images contained in the database (Fig.8.8).

The Pearson Linear Correlation Coefficient (PLCC) between the average ratings and the predicted SBDI is 0.98. These results support the argued strict relationship between the subjective quality loss ratings and the visual discomfort ratings caused by blur.

8.3 Discomfort and quality loss

In principle, the visual discomfort and the quality loss are different concepts. The visual discomfort is a generic negative feeling caused by some visual stimuli. The

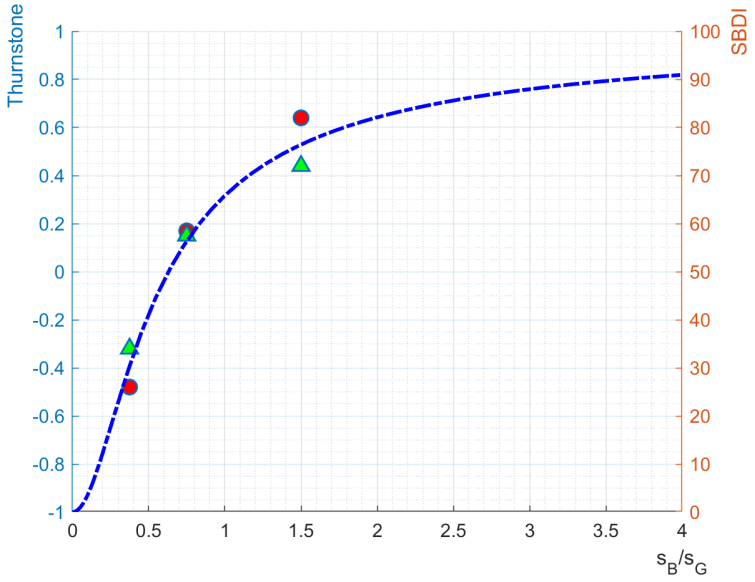

Figure 8.7. The scatterplots of the subjective discomfort ratings versus the normalized blur spread for the images in [71]. The triangles refer to the distant natural scene category, and the circles to the closeups scene category.

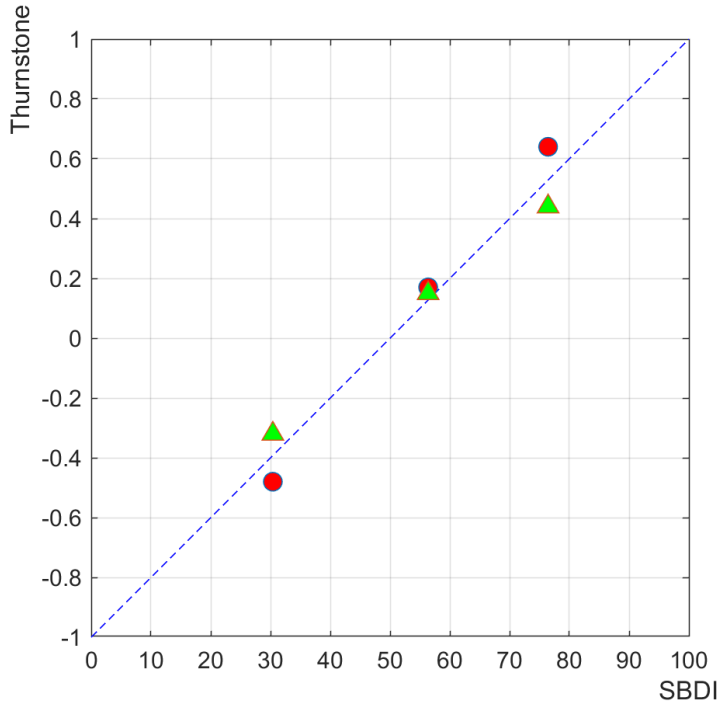

Figure 8.8. The subjective discomfort ratings versus their theoretical predictions for the images in [71]. The triangles refer to the distant natural scene category, and the circles to the closeups scene category.

image quality loss is the subjective impairment of a reproduced image with respect to an expected standard.

However, restricting the interest on blur, and interpreting it in terms of PFI loss, the two concepts significantly overlap. In fact, in the referred experiments, the eye optical blur was always corrected. Blur applied to the observed images emulates an undesired optical blur that cannot be compensated by accommodation [25].

Moreover, the subjective quality measured in experiments may be not indicative of the visual discomfort, because the quality is also influenced by the content of the images presented to the subjects. The visual discomfort should be considered as a basic inter-subjective aspect of a more complex phenomenon of quality loss perception. Such thesis is corroborated by the vertical clustering of the empirical data in the final scatterplots of the databases TID2013 and LIVE MD (see "Experimental verification" chapter). This occurs because few blur spread values were applied to the entire body of original images. In the LIVE DBR2 and in the CSIQ databases, this dependence is masked by the irregular variety of blur spread values applied to the images.

The present theoretical model is unable to explain these DMOS differences. On the other hand, being the subjective ratings averaged over different subjects, it is ascertained that they must depend on the different image content. Notably, some methods expressly designed for measuring the subjective image quality do correlate better with the DMOS data for blurred images with respect to the SBDI. This indicates that these methods implicitly account for content-related image features, other than blur.

To understand better this circumstance, in Fig.8.9 the points representing the DMOS values at different blur for the same image are linked together. It comes out that the ranking of the subjective image quality remains substantially the same for different values of blur.

The effects of the contents of the images will be better studied in the following chapters, completing the estimator with the analysis of the fourth order moments.

The PFI, upon which the present analysis is based, has the form of a signal to noise ratio, the signal being constituted by averages of gradient magnitude, and the noise by white Gaussian noise. The complex valued receptive field model VRF whitens the typical spectrum of natural images. This circumstance, along with the analytical simplicity of the Gaussian blur and of the Fisher equivalence criterion, has led to a formally compact model of the blur discomfort. The aim of these result is to contribute to verify in which sense and at which extent the blur visual discomfort can be regarded as an inter-subjective phenomenon. The model obtained may provide initial estimates of the blur discomfort for the design process of vision correcting systems, imaging, compression, and reproduction devices.

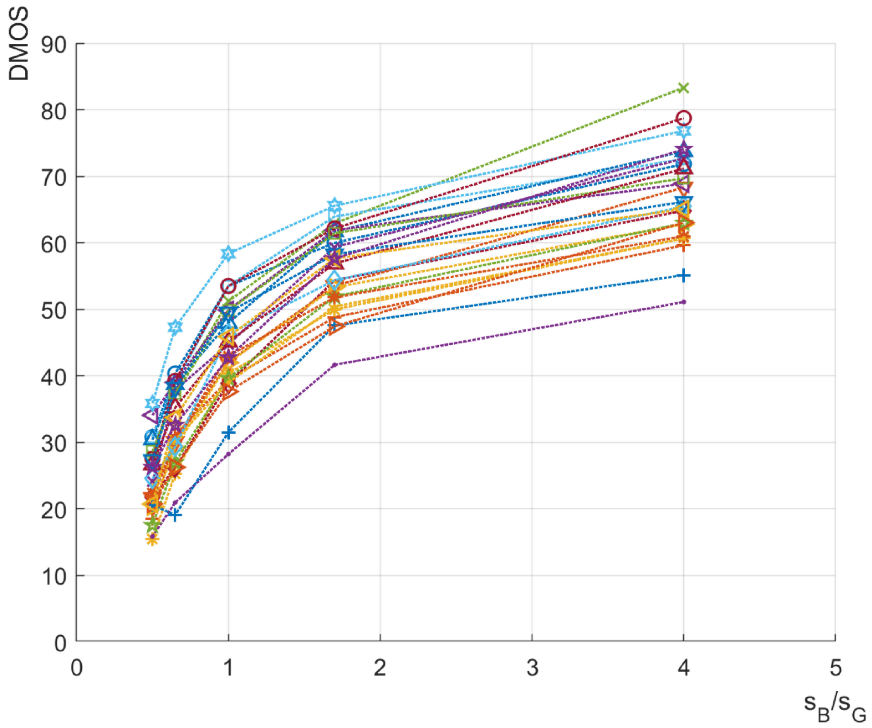

Figure 8.9. The DMOS values of the blurred images of the TID2013 database versus blur, showing the persistence of the DMOS ranking across different blur spreads.

Part IV

The Role of the Viewing Distance and the Diversity of Image Content

Chapter 9

Dependence on the Viewing Distance of the Theoretical DMOS

9.1 Full Reference IQA and subjective distance

Image Quality Assessment (IQA) is based on *subjective* as well as *objective* methods. The *subjective* quality of an image is defined as the average quality score assigned to it by a class of human subjects, usually expressed in the MOS (Mean Opinion Score) scale. Likewise, the *subjective quality loss* with respect to a pristine version of the same image considered of *perfect quality* (reference image) is expressed in DMOS (Difference of Mean Opinion Score) units.

The objective quality consists of the *algorithmic prediction* of the subjective quality based on measurable image features, expressed as well in MOS/DMOS units.

In the recent decades, many objective quality assessment methods have been proposed. Methods requiring a complete representation of the reference image are referred to as *Full Reference* (FR) IQA methods, while methods using incomplete representations of are referred to as *Reduced Reference* (RR). Methods based only on the knowledge of degraded images are finally referred to as *No Reference* (NR) methods. The present discussion is concerned with FR methods.

The typical scheme of objective FR IQA methods includes first a local comparative analysis of the pristine and of the degraded image stage, followed by a pooling stage over the whole image. The result is the so-called as IQA *metric* (sometimes referred as objective quality for short).

Formulation of existing metric was guided by different inspiring criteria. Some metrics measure the similarity among image signals (reproduction fidelity), possibly accounting for constraints suggested by more or less sophisticated models of the Human Visual System (HVS). Other classical metrics measure a loss of a somehow defined *visual information* caused by the image degradation, with reference to a *cognitive* interpretation of the HVS role. Mostly, these metrics are in the form of *scalar* metrics, even if some *vector* metrics have been adopted to distinguish among different types of image degradation.

The final step is the application of a *scoring function* whose goal is to convert

the IQA metric values into the subjective MOS/DMOS scales (for short, the result is sometimes referred to as subjective quality, subtending "estimates of").

The scoring function is typically formulated as a parametric function suggested by the aim to model threshold and saturation phenomena. Parameters are adjusted by non-linear regression using available empirical examples. Following the VQEG suggestions [93, 94], the scoring function for a scalar metric consists of a logistic function whose typical form is

$$\hat{m}(\zeta) = \beta_1 \left[\frac{1}{2} - \frac{1}{1 + e^{\beta_2(\zeta - \beta_3)}} \right] + \beta_4 \zeta + \beta_5 \tag{9.1}$$

where ζ is the metric value, $\hat{m}(\zeta)$ is the estimated DMOS value [81], and β_i are parameters usually adjusted minimizing the r.m.s. distance of empirical DMOS values from the estimated ones $\hat{m}(\zeta)$. Of course, the larger the number of parameters, the more accurate is the tracing of the specific training examples. But, on the other hand, the greater is its variability across different sets of examples (overfitting). The generalizability of the scoring function across different examples datasets decreases with the number of adjustable parameters [1]. As a matter of fact, it is hard to find default values for these parameters in the available literature. Even if it compensates for the variability of methodologies employed for subjective quality collection, example-based calibration prevents a fair comparison of quality measurement across applications. Notice that in [42] it reduces to the application of a S-shaped function regulated by three parameters only.

The scoring function depends critically on many factors, such as initialization of the non-linear optimization and experimental settings, display contrast, room illumination, angle of view, etc., other than by different protocols.

More evident, it does depend on the *Viewing Distance* (VD). In fact, the objective quality is calculated from the images reproduced on the *display*, while the subjective quality depends on the images perceived on the *retina*, whose scale depends on the VD. Unfortunately, the said parametric forms of scoring functions generally lack an explicit interpretation in terms of VD.

To attenuate the effect of the VD, in some IQA methods, namely the Multi-Scale Structural Similarity (MS-SSIM) [98] and the Feature Similarity Index (FSIM) [109], the dependence of their scoring function on the VD is attenuated by adjusting a mix of metrics calculated at different scales. Generalizing this approach, in [35] a preprocessing stage is applied to different IQA methods. Images are rescaled and filtered to emulate different VDs, and the corresponding metrics are combined using the criterion of optimizing the overall statistical performance. The VD was also included, as a learnable parameter, in a CNN based NR IQA method [10].

Here, following a quite opposite approach, the problem is tackled using a theoretical viewpoint, based on principles and reasonable hypotheses. The essential scope is to define an effective model of the rating function explicitly related to the VD.

To this purpose, the case of *blurred images* is adopted as a significant *benchmark*. In fact, in this case, the image degradation substantially depends on the VD, being scarcely influenced by additional environmental factors, such as room illumination or display contrast ratio. In addition, modeling the effects of blur is mathematically feasible, and relatively easy in the case of natural images.

The analysis conducted here stems from the overarching principle that the Human Vision System (HVS) is an optimal device for fine pattern localization, given the eye macro-structure [26]. This feature is especially important for stereoscopic vision. The outstanding consequence, drawn from information theory, is that the variance of fine localization of structures in observed images is objective, and calculable as a function of the degree of blur. It is approximated by the inverse of the Fisher Information about the position of observed patterns.

Based on this principle, and on well-known spectral properties of *natural* images, the rating function relating the degree of *Gaussian blur* to the subjective quality loss can be expressed in a closed form, where it was ascertained that this function fits well with empirical DMOS data provided by independent well known IQA databases.

In the past, various simplified modes of the HVS were adopted for IQA development. Here, an essentially abstract model of the HVS composed of the optical tract and of the neural tract (from the retina up to the visual cortex) is adopted. The last one is modelled by a single, complex valued, Virtual Receptive Field (VRF) accounting for pattern orientation selectivity and, at the same time, for 2D spatial frequency selectivity. It is viewed as the cascade of an ideal complex gradient operator, and a low-pass filter which produces a so-called *neural blur*. The VRF model allows straightforward calculation of the PFI of image details, and subsequent pooling.

The second, substantial simplification comes from restricting the analysis to the class of the so-called *natural scenes*, i.e., the images usually seen by subjects during their everyday experience directly or through imaging devices. In the present chapter, natural images are modeled as elements of a random set whose ensemble radial spectral energy distribution decays exponentially [88]. It is important to outline that in the IQA context other characterizations of natural images are used. See for instance the Natural Scene Statistics (NSS), employed in [80].

At this point, it is assumed that the perception of the quality loss follows the psycho-physical Weber-Fechner law with respect to PFI losses. It follows that an estimate of the presumed subjective quality loss in the case of *Gaussian blurred images* is expressed in a closed, unexpectedly simple form and constitutes the so-called *canonical* IQA method based on the *canonical metric*. This metric is constituted by the standard deviation of the Gaussian blur operator normalized with respect to the standard deviation of the Gaussian *neural blur* operator. The associated rating function conforms to the expected thresholding and saturating behavior, but it is characterized by two significant and operative parameters, the VD normalized with respect to a *nominal* VD (indicating the *resolution* of the display), and a *scoring gain* depending on conventional quality scores assigned to "anchor" images. After verification conducted on independent IQA databases, the canonical method exhibits a remarkable linearity versus empirical MOS/DMOS measurements reported in different independent IQA databases.

Subsequently, the applicability of the theoretical results to some most popular IQA FR methods is verified, with the scope of applying them to the prediction of their DMOS at various distances starting from the data available for a specific distance, without performing new experiments. The key result is that the calculus of the expected value of quality predicted with a scalar IQA metric can be decomposed into the cascade of an *intrinsic metric distortion*, typical of that metric, with the rating

function of the canonical method. This permits quality estimates parameterized only by the VD and the conventional scoring gain.

The final part of this chapter is devoted to the extension of these theoretical results to other types of image degradation, beyond the case of Gaussian blur. To this purpose, the relationship between metrics of some mainstream IQA methods and the canonical metric is discussed in some detail. The analysis starts from the basic consideration that usual IQA methods using *scalar metrics* are purposely designed to estimate the subjective image quality loss irrespective of the degradation cause. So, it is expected that different types of degradation will properly map onto the canonical metric provided that they are *perceptually and unambiguously equivalent*, in terms of image quality, to a certain level of normalized Gaussian blur. If this verifies, the a-priori linearity of the overall scoring function follows as a logical consequence.

Some mainstream IQA methods (VIF, MS-SSIM, FSIM, and GMSD) are specifically examined as possible candidates for rectification-free subjective quality prediction, showing the effectiveness of this approach for a technically significant class of image degradation. An exhaustive comparative analysis of performance with conventional rectification prone methods is provided. To this purpose the scatterplots of the DMOS actual data with respect to the estimates are shown and some statistical indices are reported in synoptic tables.

9.2 The retinal image versus the viewing distance

The retina in the foveal region is modeled as a distribution of light receptors whose position is individuated by the coordinate pair $\mathbf{p} \equiv (x_1, x_2)$. The projection of an image on the retina generates an image, whose luminance component is $I(\mathbf{p})$. The chrominance components are ignored here. It is assumed that $I(\mathbf{p})$ is sampled by a grid of receptors, whose average density is assumed as $60/degree$ of the angle of view. Since these receptors are not regularly placed on the retinal surface, invoking the generalized theory for non-baseband and nonuniform samples of [52], this amounts to say that $I(\mathbf{p})$ is correctly represented provided that its 2D bandwidth does not exceed $(+30, -30)$ $cycles/degree$, i.e., $(-1/2, +1/2)$ $cycles/arcmin$. This is the assumed retinal resolution, corresponds to the so-called *Snellen acuity*.

The screen is a rectangular array of pixels characterized by its *resolution* R expressed in $pixel/mm$, i.e., spaced $d = \frac{1}{R}$ mm apart. The image is projected onto the retina without loss of information and without redundancy if the spacing d matches the said $60/degree$ density of receptors. This occurs at a distance given by:

$$\delta_0 = \frac{1}{R \times tg(1)} \tag{9.2}$$

called hereafter *nominal VD*, where δ_0 is measured in $meters$, and angles in $arcmins$. In the ITU recommendations [40] this distance is referred to as the Design Viewing Distance (DVD), or Optimal Viewing Distance (OVD).

Specifically, for a screen of height H and characterized by L pixel rows *with reference to the center of the screen*:

$$\delta_0 = \frac{H}{L \times tg(1)} = \frac{H}{L} \times 3.438 \tag{9.3}$$

For instance, the nominal distance for a 4k, 32 inch screen is:

$$\delta_0 = \frac{H}{L} \times 3.438 = \frac{440}{2160} \times 3.438 = 0.7 \tag{9.4}$$

At nominal VD the spatial Fourier spectra of the images projected onto the retina are limited into the range $(-1/2, +1/2)$ *cycles/arcmin* $((-30, +30)$ *cycles/degree*), according to the Nyquist criterion. For a generic VD δ the spectrum of the retinal images lies within the band $(-\frac{1}{2}\frac{\delta}{\delta_0}, +\frac{1}{2}\frac{\delta}{\delta_0})$ *cycles/arcmin* $(-30\frac{\delta}{\delta_0}, +30\frac{\delta}{\delta_0})$ *cycles/degree*) both in horizontal and vertical directions. Thus, if the VD is less than the nominal one, the bandwidth of the sensed image shrinks. If the distance exceeds the nominal one, the bandwidth broadens.

9.3 The abstract model of vision

The HVS processes the visual information captured by the retina in successive stages of the neural tract connecting the output of the retina to the visual cortex, using *receptive fields* distributed across the visual field [36]. In the present work, no attempt is made to model the physical receptive fields.

Henceforth, here it is assumed that the eye accommodation mechanism properly works. More generally, that refraction defects are suitably corrected with lenses, so that the images are perfectly focused on the retina.

As far as the neural tract is concerned, it is abstractly modeled as the convolution of the sensed image $I(\mathbf{p})$ with an operator $h(\mathbf{p})$,

$$y(\mathbf{p}) = I(\mathbf{p}) * h(\mathbf{p}) \tag{9.5}$$

The output $y(\mathbf{p})$ will be referred to as the *visual map*. According to the general principles of Daugman [23], the operator $h(\mathbf{p})$ is an abstract *functional model* aimed to account for the *orientation selective behavior* and the spatial *frequency selective behavior* of the HVS.

This operator will be referred also as the VRF [26]. It is a complex valued operator, as a pair of real operators:

$$h(\mathbf{p}) = Re\{h(\mathbf{p})\} + jIm\{h(\mathbf{p})\}. \tag{9.6}$$

It is significantly described in the polar coordinates $r = \sqrt{x_1^2 + x_2^2}$ and $\varphi = tg^{-1}\frac{x_2}{x_1}$ as

$$h(r, \varphi) = \frac{r}{2\pi s_G^2} e^{-\frac{r^2}{2{s_G}^2}} e^{j\varphi} \tag{9.7}$$

where the parameter s_G is the *spatial spread* of the VRF. It is a *polar-separable* function, whose real components are orthogonal and also orthogonally oriented (see Fig.9.1 and Fig.9.2).

The VRF is a eigen-function with respect to 2D Fourier transformation. With reference to the polar coordinates in the frequency domain

$$\rho = \sqrt{f_1^2 + f_2^2} \tag{9.8}$$

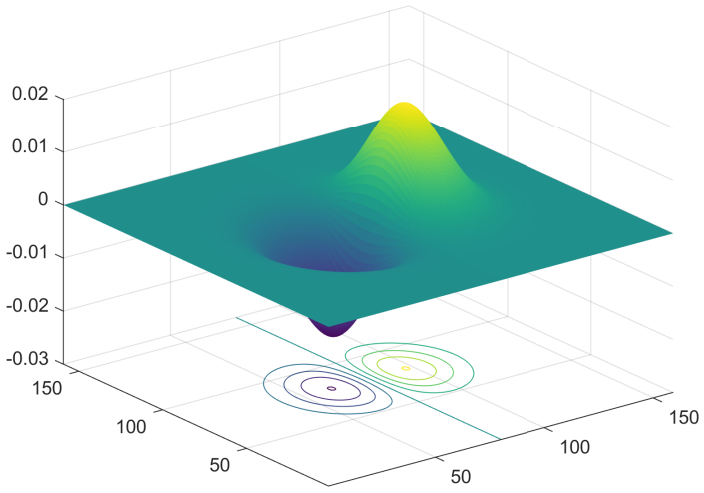

Figure 9.1. Laguerre real part of the Virtual Receptive Field (VRF) complex operator.

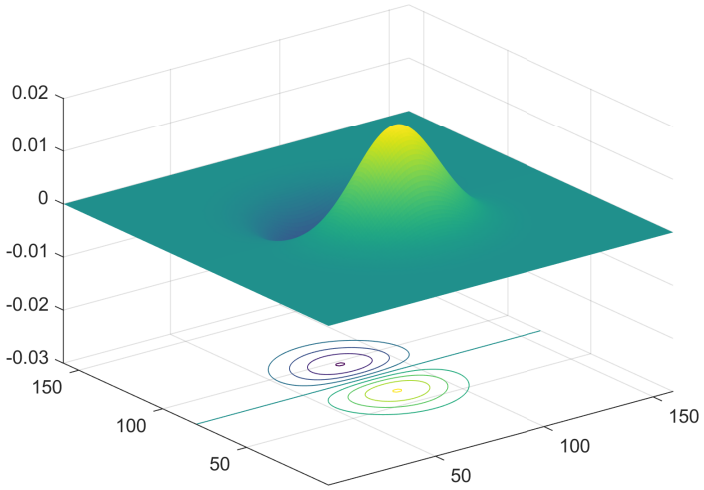

Figure 9.2. Laguerre imaginary part of the Virtual Receptive Field (VRF) complex operator.

$$\vartheta = tg^{-1}\frac{f_2}{f_1} \tag{9.9}$$

where ρ is the radial frequency, the Fourier transform of the VRF is

$$H(\rho,\vartheta) = j2\pi\rho e^{j\vartheta}e^{-S_G^2\rho^2} \tag{9.10}$$

referred in the sequel to as Virtual Neural Transfer Function (VNTF). Its magnitude is a model of the Contrast Sensitivity Function (CSF) [58], as shown in Fig.9.3.

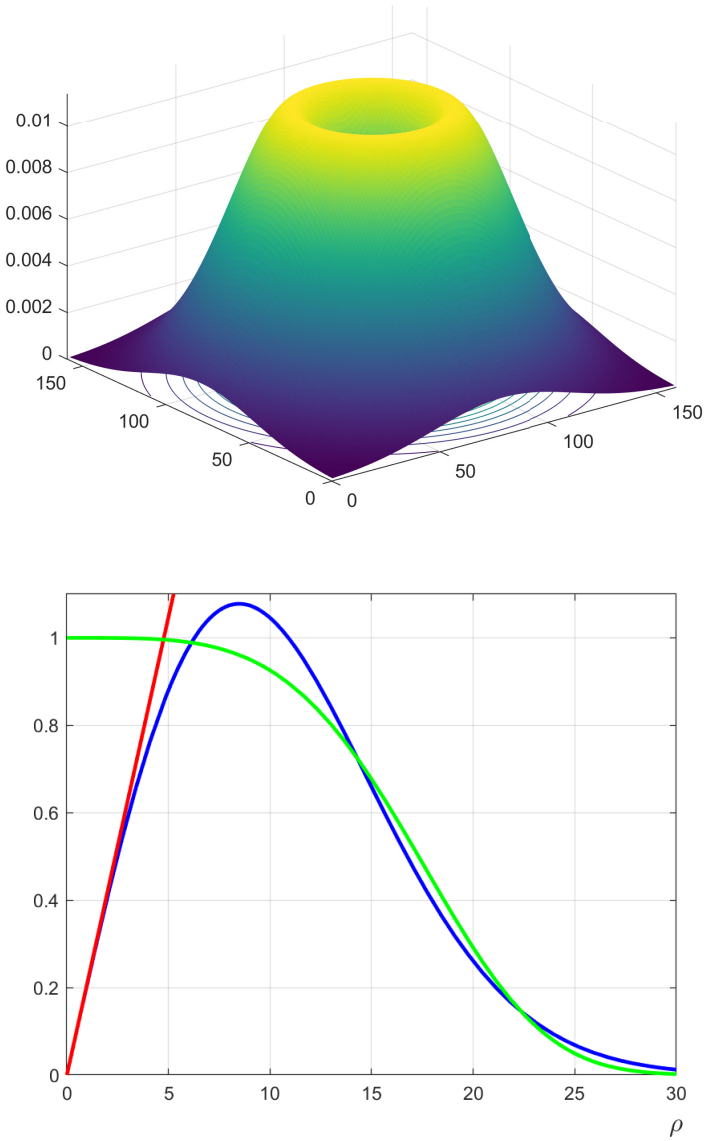

Figure 9.3. The radial frequency response magnitude of the VNTF for $s_G = 2.5$ *arcmin* in 3D representation (up) and splitted in its Gaussian and linear components (down). The spread $s_G = 2.5$ *arcmin* of the VRF sets the maximum of the VNTF at about 8.5 *cycles/degree*, and a loss of ∼40dB at 30 *cycles/degree*.

The radial frequency response of the VRF is a model the CSF (blue line), viewed as the cascade of an ideal complex gradient operator (red line) and of a low pass filter (green line).

The maximum of the VNTF along the radial frequency is posed at about 8.5 *cycles/degree*, according to the experimental data provided in the Fig.5 of [11]. This does correspond to a value $s_G = 2.5$ *arcmin*, and a loss of ∼40dB at the Nyquist frequency. The VRF is also interpreted as the cascade of two operators. With reference to the same Fig.9.3, the first one is the magnitude of $j2\pi\rho e^{j\vartheta}$ which

represents the (unsmoothed) *complex spatial gradient* operator [78] defined as

$$\nabla I\left(\mathbf{p}\right) \doteq \frac{\partial I\left(x_1, x_2\right)}{\partial x_1} + j\frac{\partial I\left(x_1, x_2\right)}{\partial x_2} \tag{9.11}$$

while the second factor

$$G(\rho, \vartheta) = e^{-S_G^2 \rho^2} \tag{9.12}$$

represents a Gaussian-shaped smoothing filter, interpreted as the source of a *neural blur*.

Thus, the whole VRF is interpreted as a complex, *smoothed* Gaussian gradient operator, acting as an edge extractor. Following the Marr's theory of vision [60], the visual map depicts a complex version of the image *skeleton*, where the magnitude indicates the edge strength, and the phase indicates the edge orientation.

Now, let the image on the display $I_D\left(\mathbf{p}\right)$ be a blurred version of a pristine (non-degraded) image $\widetilde{I}_D\left(\mathbf{p}\right)$, it is modeled as the convolution with the blur kernel $b\left(\mathbf{p}\right)$:

$$I_D\left(\mathbf{p}\right) = \widetilde{I}_D\left(\mathbf{p}\right) * b\left(\mathbf{p}\right) \tag{9.13}$$

so that the blurred image on the retina is modeled as

$$I(\mathbf{p}) = \widetilde{I}_D\left(\mathbf{p}\frac{\delta}{\delta_0}\right) * b\left(\mathbf{p}\frac{\delta}{\delta_0}\right) \tag{9.14}$$

and finally the visual map is modeled as

$$y\left(\mathbf{p}\right) = \widetilde{I}_D\left(\mathbf{p}\frac{\delta}{\delta_0}\right) * b\left(\mathbf{p}\frac{\delta}{\delta_0}\right) * h(\mathbf{p}) \tag{9.15}$$

9.3.1 The canonical IQA method

The accuracy of detail localization is a primary need of living beings. We assume then *as a principle* that the position of details in the observed images is determined with the maximum allowable accuracy, given some structural constraints. This means that such accuracy, measured by the inverse of the Fisher information, is intersubjective.

The Positional Fisher Information (PFI) of a detail of a pristine image and extracted by a window $w_{\mathbf{p}}(\mathbf{q})$, centered on $\mathbf{p}$ in presence of background white Gaussian noise with variance σ_V^2 is calculated in [69] as:

$$\widetilde{\psi}\left(\mathbf{p}\right) = \frac{\widetilde{\lambda}\left(\mathbf{p}\right)}{\sigma_V^2} \tag{9.16}$$

where

$$\widetilde{\lambda}\left(\mathbf{p}\right) = \sum_{\mathbf{q}} w_{\mathbf{p}}(\mathbf{q})^2 \left|\widetilde{y}\left(\mathbf{p}-\mathbf{q}\right)\right|^2 \tag{9.17}$$

is the energy of the detail. Likewise, for the same detail of a degraded version of the image we have

$$\psi\left(\mathbf{p}\right) = \frac{\lambda\left(\mathbf{p}\right)}{\sigma_V^2} \tag{9.18}$$

where

$$\lambda(\mathbf{p}) = \sum_{\mathbf{q}} w_{\mathbf{p}}(\mathbf{q})^2 \left| y(\mathbf{p}-\mathbf{q}) \right|^2 \tag{9.19}$$

Let us now calculate the PFI in the frequency domain. Here, a Gaussian blur applied to the observed image is described in polar frequency coordinates as

$$B(\rho,\vartheta) = e^{-s_B^2\rho^2} \tag{9.20}$$

where the parameter s_B will be referred as to the *spatial span* of the blur operator. Referring to the Fourier spectra of a generic detail, by the Parseval theorem:

$$\begin{aligned} \psi(\mathbf{p}) &= \frac{1}{\sigma_V^2}\int_0^{2\pi}\int_0^{+\infty} \left|D_{\mathbf{p}}(\rho)\right|^2 \rho^2 \left|G(\rho,\vartheta)\right|^2 \left|B(\rho,\vartheta)\right|^2 \rho d\rho d\vartheta \\ \widetilde{\psi}(\mathbf{p}) &= \frac{1}{\sigma_V^2}\int_0^{2\pi}\int_0^{+\infty} \left|D_{\mathbf{p}}(\rho)\right|^2 \rho^2 \left|G(\rho,\vartheta)\right|^2 \rho d\rho d\vartheta \end{aligned} \tag{9.21}$$

Let us now refer specifically to *natural images*. It is known that in this case [88]

$$E\left\{\left|D_{\mathbf{p}}(\rho,\vartheta)\right|^2\right\} = f(\vartheta)\frac{1}{\rho^2} \tag{9.22}$$

So, for the random set of natural images, the average PFI in presence of blur is calculated as

$$\Psi = E\left[\psi(\mathbf{p})\right] = K\int_0^{+\infty} \left|G(\rho,\vartheta)\right|^2 \left|B(\rho,\vartheta)\right|^2 \rho d\rho \tag{9.23}$$

and, in the absence of blur

$$\widetilde{\Psi} = E\left[\widetilde{\psi}(\mathbf{p})\right] = K\int_0^{+\infty} \left|G(\rho,\vartheta)\right|^2 \rho d\rho \tag{9.24}$$

where

$$K = \frac{1}{\sigma_V^2}\int_0^{2\pi} f(\vartheta) d\vartheta \tag{9.25}$$

From known results of integral calculus, we arrive at the following closed form results:

$$\Psi = K\int_0^{+\infty} e^{-2(s_G^2+s_B^2)\rho^2}\rho d\rho = \frac{K}{4\left(s_G^2+s_B^2\right)} \tag{9.26}$$

$$\widetilde{\Psi} = K\int_0^{+\infty} e^{-2s_G^2\rho^2}\rho d\rho = \frac{K}{4s_G^2} \tag{9.27}$$

Now, let us assume that the perceived loss of positional uncertainty is proportional to the objective one, considering the necessary consistency of the position estimates in the 3D Euclidean space during the interaction of the subject with its surrounding environment. According to classical psycho-physical laws, let us assume that the differential sensitivity to stimuli is inversely proportional to the size of the initial stimuli (Weber-Fechner law [26]). Identifying the stimulus as the positional uncertainty (see (9.26) and (9.27)), the following index is then proposed

as a psycho-physically grounded model of the quality loss perception caused by Gaussian blur applied to natural images:

$$\varepsilon(\xi) = \frac{\sqrt{\psi\,(p)} - \sqrt{\widetilde{\psi}\,(p)}}{\sqrt{\psi\,(p)}} = \frac{\sqrt{s_G^2 + s_B^2} - \sqrt{s_G^2}}{\sqrt{s_G^2 + s_B^2}} = 1 - \sqrt{\frac{1}{1+\xi^2}} \tag{9.28}$$

having introduced the Normalized Blur (NB) $\frac{s_B}{s_G} = \xi$, which is by fact the metric of this quality estimator. This index is straightforward mapped onto the DMOS scale (which assigns the value 0 to perfect quality, and 100 to the "worst" quality) by scaling it in centesimal units:

$$\hat{d(\xi)} = 100\left[1 - \sqrt{\frac{1}{1+\xi^2}}\right] \tag{9.29}$$

This theoretical DMOS estimator is plotted in Fig.9.4. It presents a characteristic sigmoidal behavior, with a threshold for small input values and a saturating behavior for large input. The psycho-physical plausibility of the proposed rating function is well supported by its compliance with the typical "dipper" shape [84] of the incremental blur producing a Just Noticeable Difference (JND) of the perceived quality. Specifically, as shown in [26], the theoretical minimum occurs in the correspondence of

$$\xi = \sqrt{\frac{1}{2}} \tag{9.30}$$

corresponding to $\hat{d}(\xi) = 18.4$.

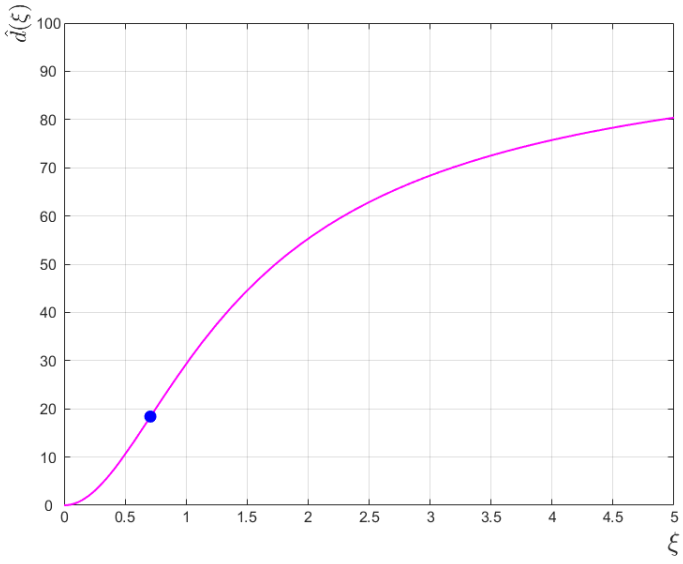

Figure 9.4. The rating function of the canonical DMOS estimator versus the metric ξ (in centesimal units DMOS). The blue dot indicates the point of maximum sensitivity of the HVS with respect to the variations in blur.

Still, this DMOS estimate remains undetermined. First, the concept of *worst quality* is undefined. To fix it, it is necessary to assign by convention the subjective quality in correspondence of a chosen value of the blur spread. This is done by introducing a *scoring gain* parameter Q so that

$$Q = \frac{d_A}{100}\left[1 - \sqrt{\frac{1}{1+\xi_A}}\right]^{-1} \tag{9.31}$$

where d_A is a DMOS value conventionally attributed to a selected image for ξ_A ("anchor").

9.4 The canonical IQA method at different VD

The above estimates refer to the nominal VD, i.e., to the case where $\delta = \delta_0$. Accounting for the scaling of the retinal image due to the actual visual distance we have

$$\begin{aligned} \Psi = K\int_0^{+\infty} e^{-2\left[\left(\frac{\delta}{\delta_0}s_G\right)^2+\left(\frac{\delta_0}{\delta}s_B\right)^2\right]\rho^2}\rho d\rho &= \frac{K}{4\left[\left(\frac{\delta}{\delta_0}s_G\right)^2+\left(\frac{\delta_0}{\delta}s_B\right)^2\right]} \\ \widetilde{\Psi} &= \frac{K}{4\left(\frac{\delta}{\delta_0}s_G\right)^2} \end{aligned} \tag{9.32}$$

So, posing for notational simplicity $\tau = \dfrac{\delta}{\delta_0}$ we finally obtain the following theoretical formula for the prediction of the DMOS of *blurred natural images*:

$$\hat{d}_{CAN}(Q,\tau,\xi) = 100 \times Q\left[1 - \sqrt{\frac{1}{1+\dfrac{\xi^2}{\tau^4}}}\right] \tag{9.33}$$

referred to as *canonical estimate* of the DMOS of Gaussian blur of natural images, characterized by Q and the dimensionless parameter τ. It provides the *theoretical expected value of the DMOS for the random set of natural images.*

Incidentally, let us observe that using if a set of empirical DMOS data is available for Gaussian blurred images Q and τ can be blindly estimated by regression. Provided that the distribution of the spectra of such images contained in the database is well balanced, i.e., that the empirical values of the DMOS scatter symmetrically around their expected value $\hat{d}_{CAN}(Q,\tau_1,\xi)$.

Applying for instance this method to the blurred images of the LIVE Database Release 2 (DBR2) [81], the Tampere Image Database (TID2013) [75], the LIVE Multi Distortion (MD) [48], and to the Computational and Subjective Image Quality Database (CSIQ) [54], it leads to the scatterplots of the DMOS empirical values versus their blur values (Fig.9.5) and to the scatterplots of the DMOS empirical

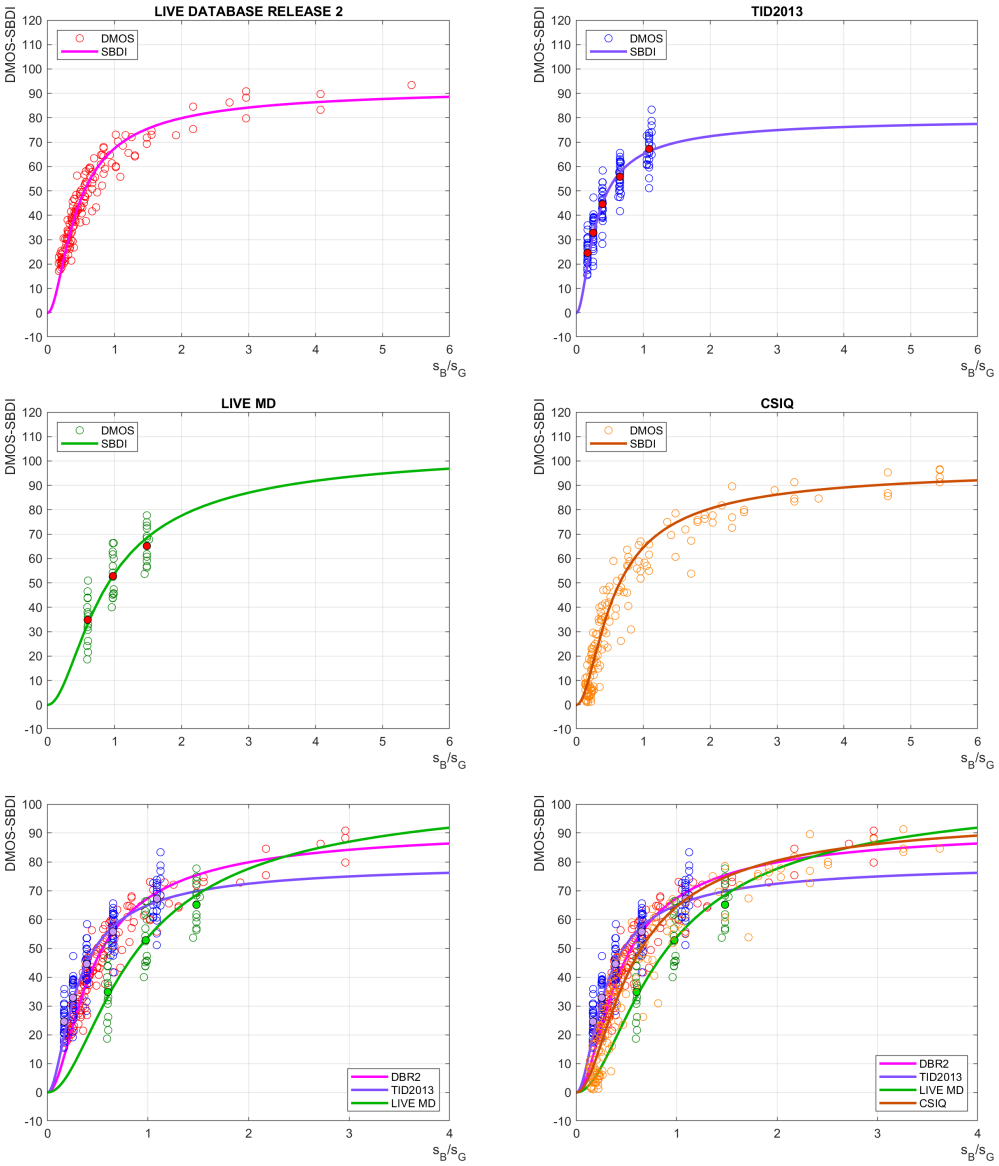

Figure 9.5. The canonical method (curves) and the DMOS values (scatterplots) versus the normalized blur spread for each database and in an overall comparative view.

values versus their estimates (Fig.9.6) obtained from the data of Tab.8.1 in Sect.8.2.2 [26].

The fit of theory based DMOS estimates with ground truth is quite good. However, one could notice that the scattering of DMOS data is greater with respect to some other methods, and specifically to the VIF method. This is explained by the fact that the canonical method, based on the measure of the PFI does not account at all for the image content. This is clearly revealed by the scattering of the plots for the TID2013 and the LIVE MD methods, where few distinct levels of blur are applied to the pristine images. Thus, the estimated DMOS values are vertically aligned, revealing the independent effects of blur and the image content on the

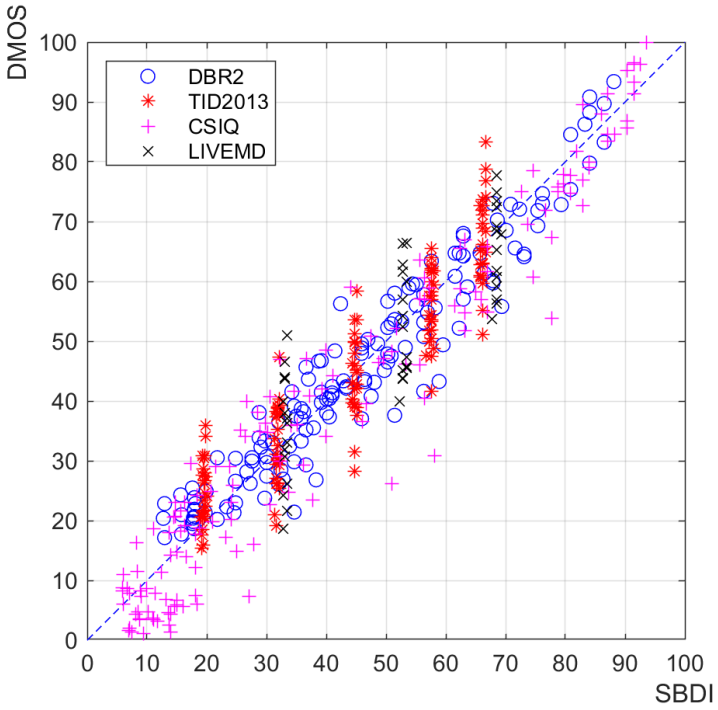

Figure 9.6. The scatterplots of the actual DMOS values versus the ones predicted by the canonical estimator for the IQA databases DBR2 (o), TID2013 (*), CISQ (+). LIVE MD (x). These scatterplots corroborate the theoretically predicted linearity of the canonical estimator for natural images.

perceived quality. The same does not occur for other methods where the image content plays a role.

9.5 The zooming function

Let us explicit how the theoretical estimate $\hat{d}_{CAN}(Q,\tau,\xi)$ depends on the VD using the interconversion formula

$$c_{CAN}(Q,\tau_1,\tau_2,\xi) = \frac{1-\sqrt{\dfrac{1}{1+\dfrac{\xi^2}{\tau_2^4}}}}{1-\sqrt{\dfrac{1}{1+\dfrac{\xi^2}{\tau_1^4}}}} \tag{9.34}$$

that will be referred to as the *zooming function*.

The zooming function allows us to predict the subjective quality at any VD starting from the subjective quality at nominal distance, and vice versa, as:

$$\hat{d}_{CAN}(Q,\tau_2,\xi) = c_{CAN}(Q,\tau_1,\tau_2,\xi)\hat{d}_{CAN}(Q,\tau_1,\xi) \tag{9.35}$$

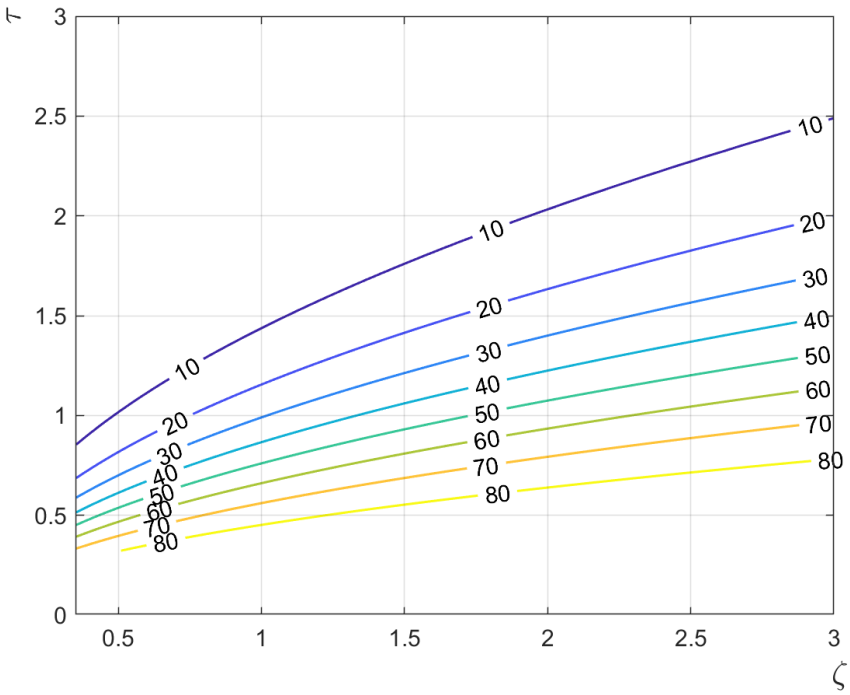

Figure 9.7. Iso-level contours as the ratio of the reference case and the nominal one with $\tau_1 = 1$. Variable ζ is used here instead of ξ following the generalization explained in section 9.7 ($\xi(\zeta)$).

This function is depicted in Fig.9.7 using iso-level contours for the reference case where the starting visual distance is the nominal one ($\tau_1 = 1$).

For VDs closer than the nominal ones, the effect of blur amplifies, and this amplification is more and more perceptually important as blur increases. For example, for $s_B = \sqrt{2}s_G$, at a distance one half of the nominal one (a distance adopted in the popular LIVE DBR2 database), the theoretically estimated DMOS for blurred natural images doubles with respect to the DMOS estimated for the nominal distance (see Fig.9.8).

Of course, for VDs farther than the nominal distance, the effect of blur diminishes with the distance.

It important to underline that the above formulas model the quality loss in a Full Reference scenario. They predict the quality loss caused by blur but are not indicative of the perceived quality in a No Reference scenario. In fact, in this case, the perceived quality may improve with the VD, depending on the content of the image. Thus, almost by principle, it is not correct to assess the performance of NR IQA methods using as reference the performance of FR IQA methods.

Finally, in Fig.9.9, it is highlighted how the theoretical estimate $DMOS(\tau, Q, \xi)$ depends on ξ (normalized s_B) for different τ (VD). It is possible to recognize the shape of the canonical estimator found in Part III for the viewing conditions of each database.

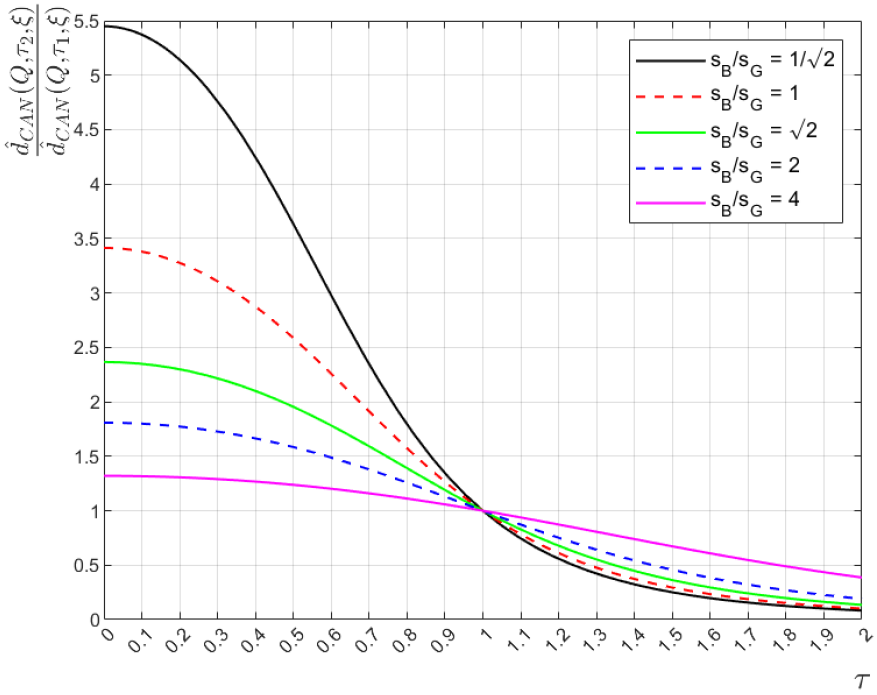

Figure 9.8. The theoretical estimate $DMOS(\tau, Q, \xi)$ depends on the VD using the interconversion formula (9.34), with the nominal distance as reference, where $\tau_1 = 1$ and $Q_1 = Q_2$ (same DB).

9.6 Behavior of common IQA methods versus the viewing distance

General purpose IQA methods are designed to consistently predict the subjective quality for different types of distortion. This implies that the rating functions for different degradations should overlap each other, and specifically with the blur rating function, that we have used as a benchmark.

To investigate this point, let us preliminarily observe how the rating functions for Gaussian blur modify across different databases. To this purpose, let us display the scatterplots of the DMOS estimates for Gaussian blurred images coming from different databases using different distances, for different IQA methods, along with their scoring functions. Gaussian blur distorsions are considered in Fig.9.10.

The methods give different results at different VDs, with the notable exception the FSIM [109], the MS-SSIM [95], and of the PSNR. This is justified because, as noted in introduction, the FSIM and the MS-SSIM are equipped with a rough pre-processing stage that adapt the resolution to the VD. As far as the PSNR is concerned, its behavior is intrinsically insensitive to the VD, because it does not account for any characteristic of the HVS. Instead, the VIF [80, 81], the GMSD [108] and the MAD [53] methods, are not robust across VD: their rating functions must be recalculated changing the VD.

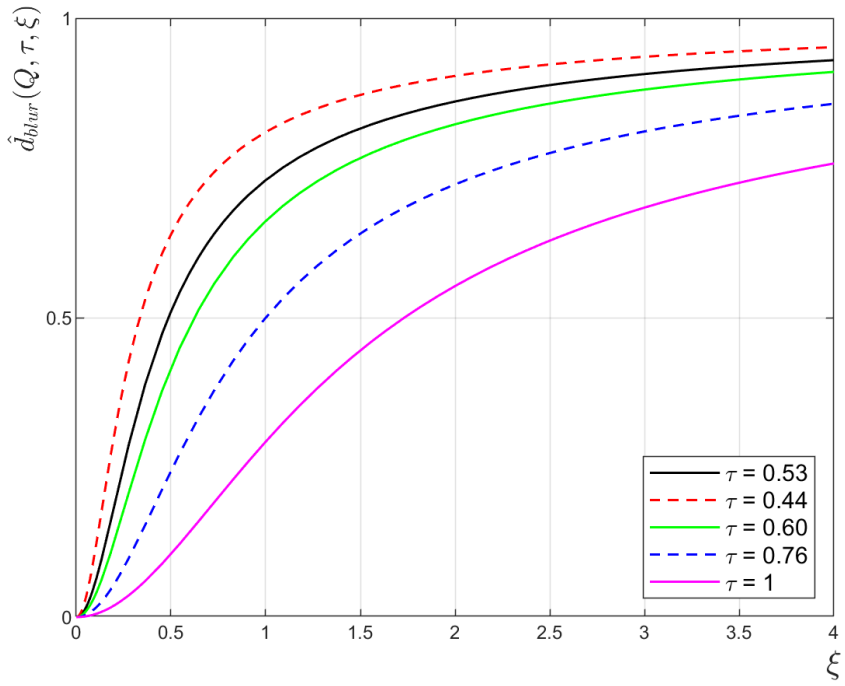

Figure 9.9. The theoretical estimate $DMOS(\tau, Q, \xi)$ depends on ξ (normalized s_B) for different τ (VD). In particular, $\tau_{DBR2} = 0.53$, $\tau_{TID2013} = 0.44$, $\tau_{CSIQ} = 0.60$ and $\tau_{LIVEMD} = 0.76$.

9.7 Generalizing the use of the zooming function

Let us verify at which amount the results obtained for the canonical method apply to other IQA metrics and to other distortions.

We make the hypothesis that the modification of the regression function across different databases, for other IQA method and for other distortions is described by the same mechanism of (9.35).

$$\hat{d}\,(Q, \tau_2, \zeta) = c_{CAN}\,[Q, \tau_1, \tau_2, \xi(\zeta)]\,\hat{d}\,(Q, \tau_1, \zeta) \tag{9.36}$$

where ζ is a generic IQA metric.

This hypothesis amounts to say that the modification of the regression function caused by a change of the VD is substantially independent of the metric.

To verify the hypothesis, it is necessary preliminarily to determine the function $\xi(\zeta)$ that align the metric ζ to the canonical metric ξ so that they refer to the same estimated DMOS, say $\hat{d}$.

To determine such a function, let us before invert the regression curve for the canonical method.

$$\xi(\hat{d}) = \tau^2 \sqrt{\frac{1}{\left[1 - \dfrac{\hat{d}}{100 \times Q}\right]^2} - 1} \tag{9.37}$$

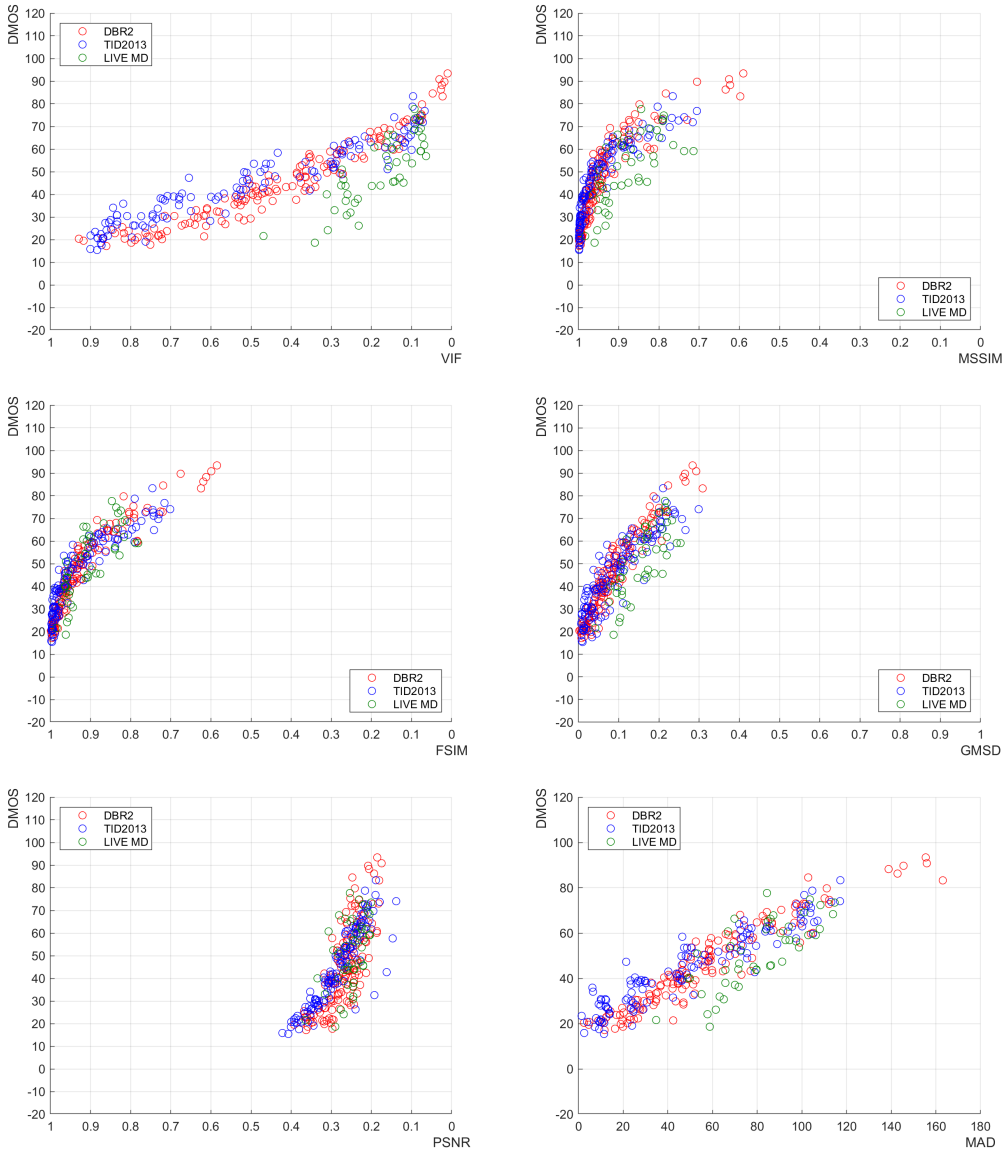

Figure 9.10. The scatterplots of the DMOS ratings versus the corresponding IQA methods values for each blurred image in the different databases.

Likewise, the equivalent value of the corresponding generic metric ζ that produces the same estimated DMOS is (empirically) determined by inverting the specific regression curve.

This allows us to determine the functional relationship $\xi(\zeta)$ and, by consequence, the proper value of ξ in the interconversion formula (9.37).

The inverse function $\zeta(\xi)$ is interpreted as the *intrinsic* distortion of the method based on this metric, and (9.37) can be viewed as a composition of the intrinsic distortion with the distortion caused by the change of the VD.

To corroborate the validity of the hypothesis, let us compare the regression curve of a database starting from the one predicted by the zooming function, by the

regression curves of the other databases using different VD. In particular, in Fig.9.11, LIVE DBR2 and TID2013 are considered with the VD varying from $\frac{\delta}{\delta_0} = 0.44$ (TID2013 distance, more linear) to $\frac{\delta}{\delta_0} = 0.76$ (LIVE MD distance, more bending). In addition to the zooming function obtained for each distance, we also have the displacement of the DMOS recalculated for the new VDs, useful for understanding the variance (as a continuous lines to highlight the range): the variance decreases as the distance increases, the curvature increases as the distance increases.

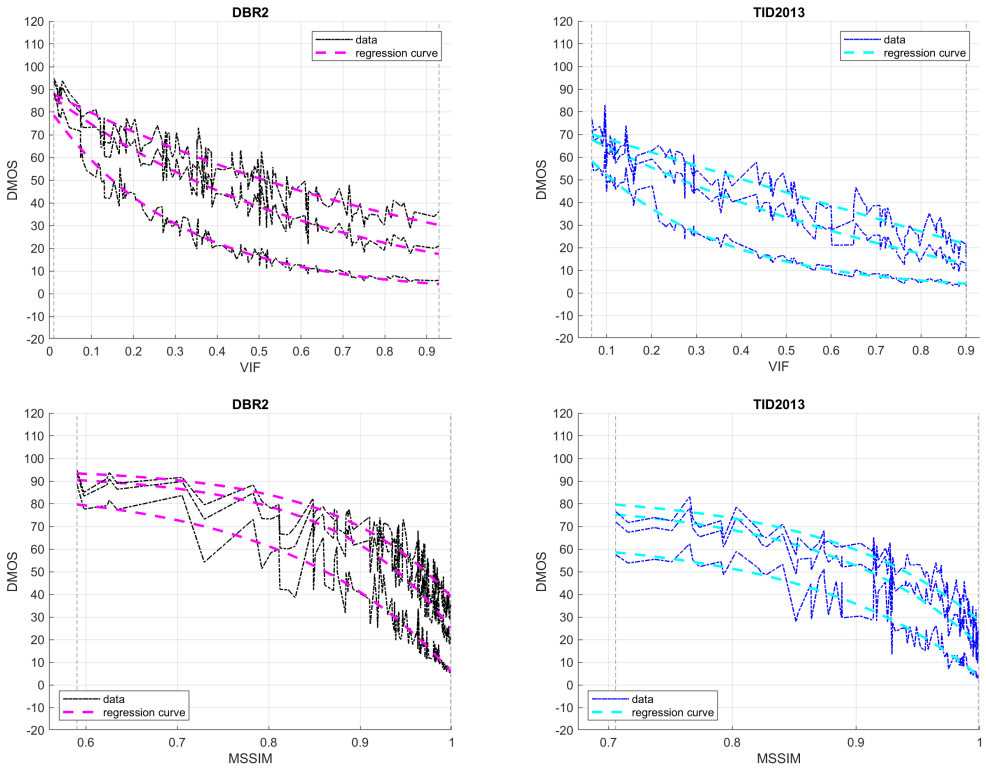

Figure 9.11. VIF and MS-SSIM methods with DMOS data at 3 different distances. Dotted lines: the regression curves predicted by the zooming function. Continuous lines: the raw data for the blurred images at different viewing conditions. The variance decreases as the distance increases and the curvature increases as the distance increases (from TID2013 ($\frac{\delta}{\delta_0} = 0.44$), LIVE DBR2 ($\frac{\delta}{\delta_0} = 0.53$) to LIVE MD ($\frac{\delta}{\delta_0} = 0.76$)).

The good matching of the predicted empirical regression curves with the actual ones does suggest that the zooming function can be used with a good approximation to the considered distortions.

9.8 Relationship among the canonical IQA method and other IQA methods

The scope of this section is to analyze the possibility of linearizing a-priori the behavior of a generic IQA method using the already known properties of the canonical method. For the sake of clarity, let us limit preliminarily to the case of

Gaussian blurred images.

Let us consider a generic IQA method based on the scalar metric ζ, and let us assume that this metric is *monotonically* related with the degree of Gaussian blur applied to images, i.e., with the canonical metric ξ.

Then, it is possible, almost in principle, to convert the metric ζ into the metric ξ. Such a conversion is obtained by searching for the function $\xi(\zeta, \tau)$ that satisfies the equality of the DMOS estimates:

$$\hat{d}_{CAN}(\xi, \tau) = \hat{d}_{IQA}(\zeta, \tau) \tag{9.38}$$

The function $\hat{d}_{CAN}(\xi, \tau)$ is directly available in an analytical closed form, but $\hat{d}_{IQA}(\zeta, \tau)$ is only available through the numerical calculus of this IQA metric. Therefore, we may approximate the conversion function $\xi(\zeta, \tau)$ using a semi-numerical method, calculating $\hat{d}_{IQA}(\zeta, \tau)$ for a *typical natural image*, i.e., an image $\widetilde{I}_{Typ}$ whose radial energy Fourier spectrum coincides with the average one of natural images, and interpolating it piecewise.

One could attempt to synthesize such an image, as exemplified in [68]. More pragmatically, we preferred to search for an exemplar in the LIVE MD dataset and selected the image "Iceroad" displayed in Fig.9.12.

Figure 9.12. The image "Iceroad" selected as a specimen of a *typical natural image*, i.e., an image possessing an energy Fourier spectrum close to the expected value of the energy Fourier spectrum of the random set of natural images.

A suite of fifty Gaussian blurred versions I_{Typ} of "Iceroad" (of which we report a few examples in Fig.9.13), whose estimated quality spanned the whole DMOS range, was prepared (see Fig.9.14). The resulting collection of points satisfying the equality is traced in Fig.9.15 for some mainstream IQA methods.

The semi-numerical solution for the function $\xi(\zeta, \tau)$ for the considered IQA method is formally provided in the form of the Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) Matlab® function and displayed in Fig.9.15 for a single value

Figure 9.13. Four blurred images from "Iceroad" with $s_B = 1$, $s_B = 3$, $s_B = 6$ and $s_B = 11$ (from top-left to down-right). Recall $\xi = s_B/s_G$, with $s_G = 2.5$ *arcmin*.

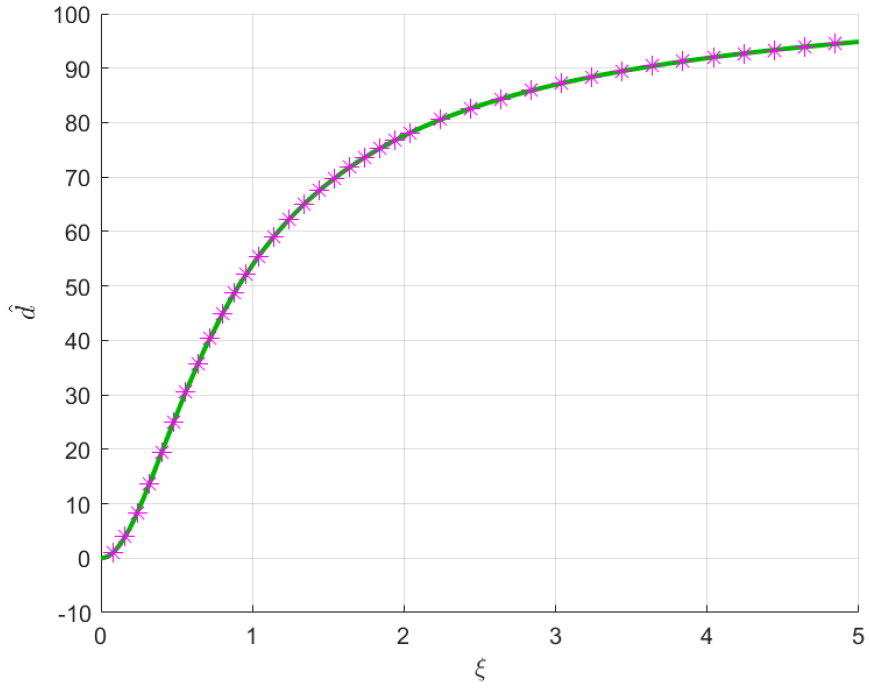

Figure 9.14. A suite of fifty Gaussian blurred versions I_{Typ} whose estimated quality spans the whole DMOS range. The magenta stars represent the blurred images and it is possible to see how well they fit the canonical model of the LIVE MD database (which contains the "Iceroad" image).

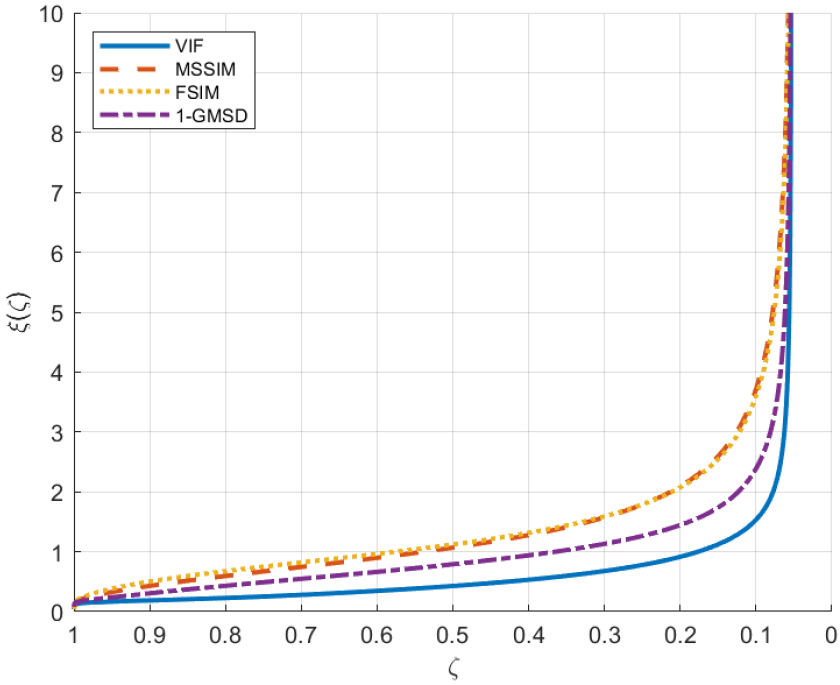

Figure 9.15. The *Equivalent Normalized Blur* function for the considered IQA method. The scales of the metrics of the MS-SSIM , FSIM and GMSD, at their own VD, are purposely aligned to the scale of the VIF metric to put into evidence their common asymptotic behavior.

of τ. Moreover, in Fig.9.16, the metric conversion function for the VIF method at different VDs are shown.

Let us observe that the fundamental behavior of the function $\xi(\zeta, \tau)$, referred also to as *Equivalent Normalized Blur* (ENB), is analogous for all the considered methods. It compresses the equivalent blur values for high degraded images, and conversely expand them for low degraded images.

Once mapped on the equivalent Gaussian blur axis ξ, a linear estimate of the DMOS Gaussian blurred natural images is finally obtained as

$$\hat{d}(\zeta) = 100 \times Q\left[1 - \sqrt{\frac{1}{1 + \frac{\xi(\zeta,\tau)^2}{\tau^4}}}\right] \tag{9.39}$$

To give a visual account of the actual linearity and of the statistical accuracy of this IQA method, in Fig.9.17 the scatterplots of the DMOS estimates for the canonical method against the same estimates obtained with the LVIF method are provided for Gaussian blur. Notice, in particular, that the obvious vertical alignments of the DMOS values in the case of the TID2013 database are destroyed by the conversion. In Fig.9.18, the scatterplots for the individual databases are given.

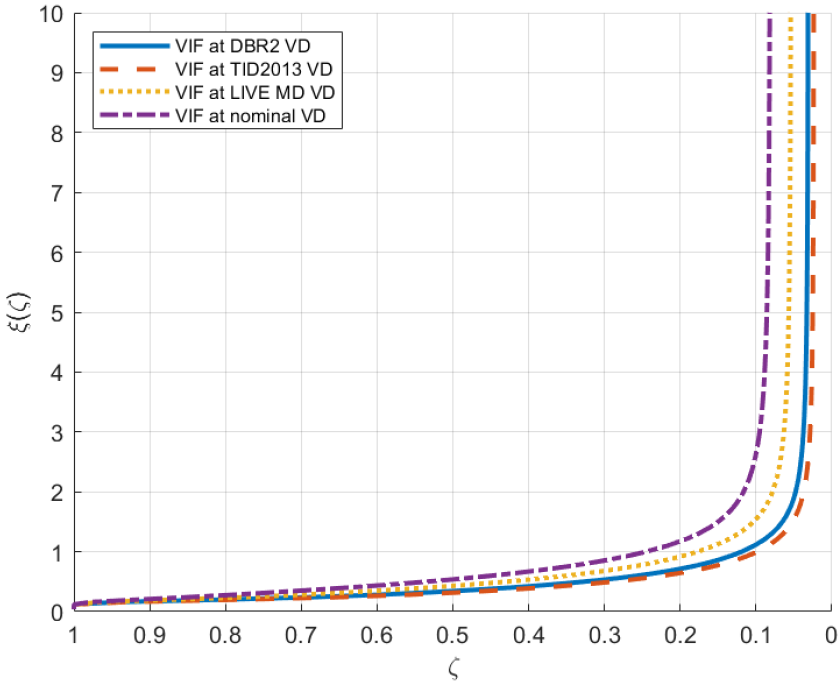

Figure 9.16. The metric conversion function for the VIF method at different VDs: $\tau = 0.53$, $\tau = 0.44$, $\tau = 0.60$, and $\tau = 0.76$, corresponding respectively to the LIVE DBR2, TID2013, CSIQ and LIVE MD VDs, and $\tau = 1$ for the nominal VD.

9.9 Extension of the method beyond Gaussian blur

The analysis conducted so far, limited to the case of Gaussian blurred natural images, is of limited practical interest. From an applicative viewpoint, it is important to have a method for assessing the quality of images subject to different types of degradation.

At a glance, it could appear that the conversion to the metric ξ of a generic IQA metric ζ is vane, since the metric ξ measures only one type of degradation, that cannot reflect for instance the effect of noise, which is neatly distinct with respect to blur from a perceptual viewpoint [62].

However, the metrics of some mainstream methods are purposely designed to measure a distance that is well correlated with the human quality score, producing the same MOS/DMOS even if its causes are *different* [96]. This aspect is also specifically discussed in [27] where two separate metrics are jointly employed and a two-dimensional scoring function is defined to achieve linearity of IQA estimates using a reduced number of adjustable parameters. Therein, it was outlined that the perceptual quality equivalence among blur and noise is *interpersonal* and then, assumed as constant. See also [16].

This implies that the conversion function $\xi(\zeta, \tau)$ makes sense even if the values of the metric ζ are not only determined by the presence of blur, but by other factors that produce the same values of $\hat{d}_{CAN}(\xi) = \hat{d}_{IQA}(\zeta)$. This is specifically the case of the considered metrics. For instance, in the VIF method the metric is computed as the contribution of many terms measuring losses of Shannon information, where the

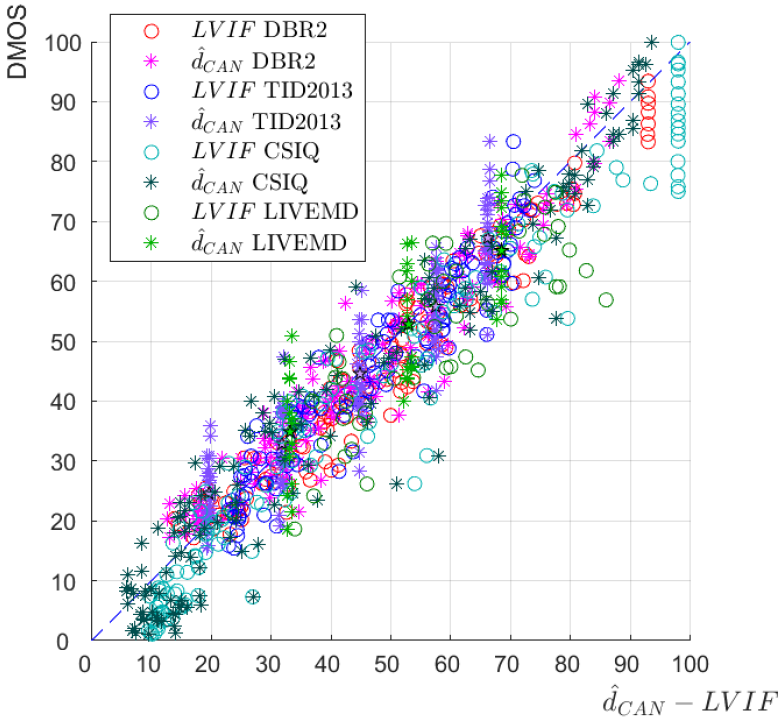

Figure 9.17. The scatterplots of the DMOS empirical measurements for blurred images versus the values predicted using respectively the canonical method (*) and the linearized VIF method (°) for different databases.

effect of increasing noise is equivalent to that caused by the reduction of the image signal energy due to blur.

A comprehensive look is given in Fig.9.19, where all distortions are considered for the LVIF method. More specifically, in Fig.9.20, an overall view of the only blur are proposed. In Tab.9.1 and Tab.9.2 are the respectively performance.

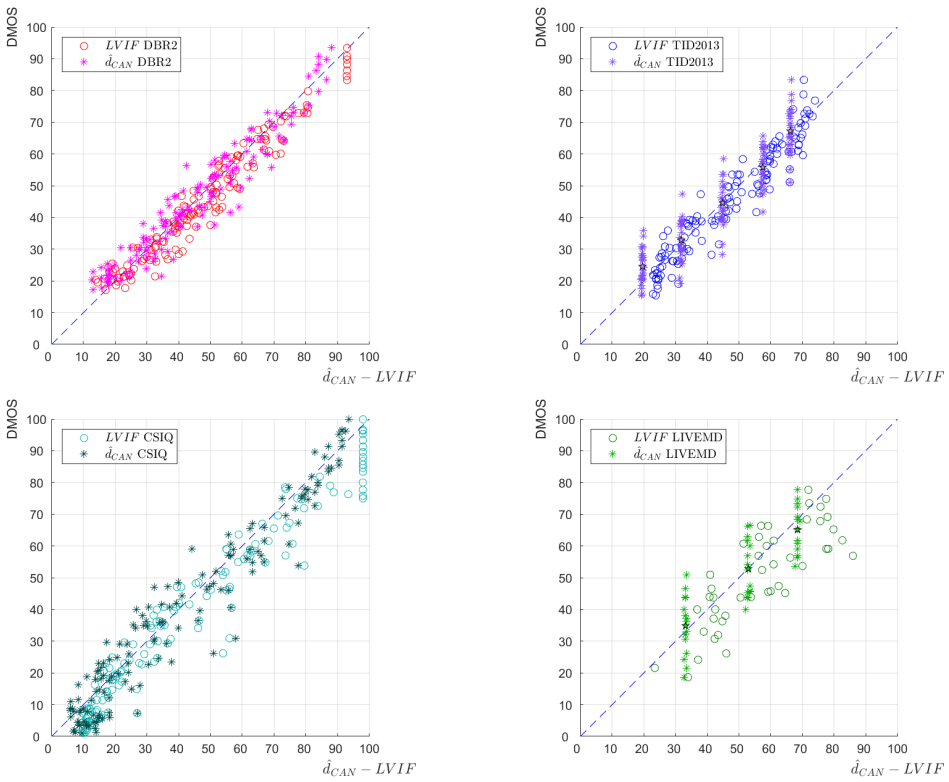

Figure 9.18. The scatterplots of the DMOS empirical measurements for blurred images versus the values predicted using respectively the canonical method (*) and the linearized VIF method (°) for individual databases.

9.10 Performance evaluation

The most fundamental difference between the conventional IQA methods considered above and their calibration free counterparts is that the former ones adapt a generic parameterized scoring function to the available examples to define a MOS/DMOS estimation rule minimizing empirical distances, while the latter ones adopt a scoring function parameterized only by the VD and the scoring gain to calculate the MOS/DMOS estimate as their expected value for natural images, using principles and theoretical arguments. By doing so, we trade adherence to particular empirical data for greater generalizability and for the ability to estimate subjective quality at different VDs.

The expected operational advantage of the calibration-free procedure is to allow comparison of MOS/DMOS estimates across different applications. The expected toll is a lower accuracy within contexts characterized by possibly unmodeled causes of quality loss.

To tangibly guess the entity of this trade-off, we present here a statistical performance comparison of some popular classical IQA methods already considered above, subject to final calibration according to VQEG recommendations, with their counterpart, indicated with the prefix L, LIQA methods, which do not make use of

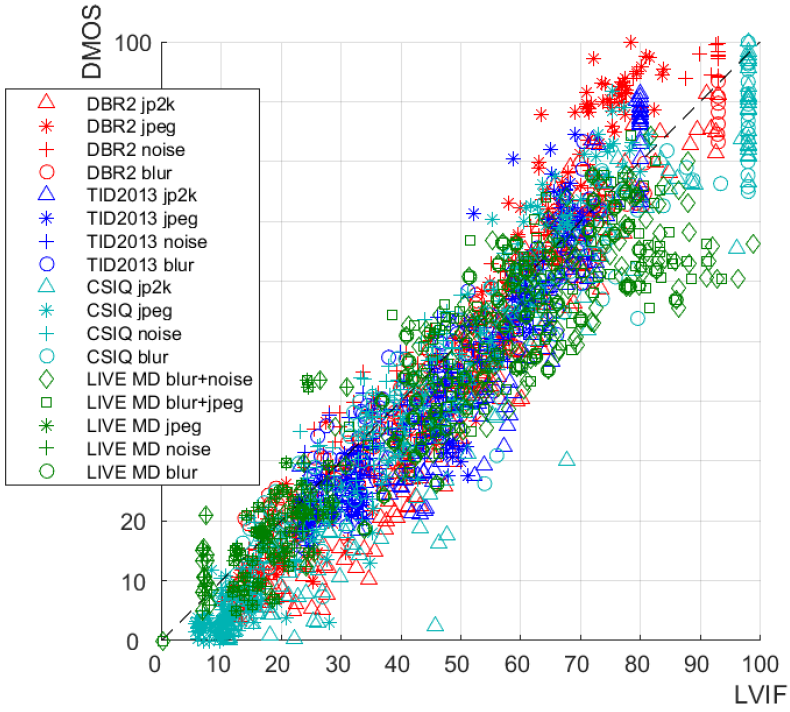

Figure 9.19. The scatterplots of the DMOS empirical measurements for all distorted images versus the values predicted using the linearized VIF method for different databases. Performance are in Tab.9.1

Table 9.1. Summary of the experimental verification for the **LVIF** (rectified IQA metrics) for DBR2, TID2013, CSIQ and LIVE MD, and for the overall case. **Four distortions subsets** for LIVE DBR2, TID2013 and CSIQ: blur, noise, JPEG, JPEG2000. **Five distortions subsets** for LIVE MD: blur, noise, JPEG, blur+JPEG, blur+noise.

Model	RMSE	SROCC	LCC
LVIF LIVE DBR2	8.6653	0.96536	0.95636
LVIF TID2013	7.0170	0.92782	0.92789
LVIF CSIQ	7.5789	0.95878	0.96336
LVIF LIVE MD	9.2065	0.92673	0.93604
LVIF OVERALL	8.3421	0.95350	0.94307

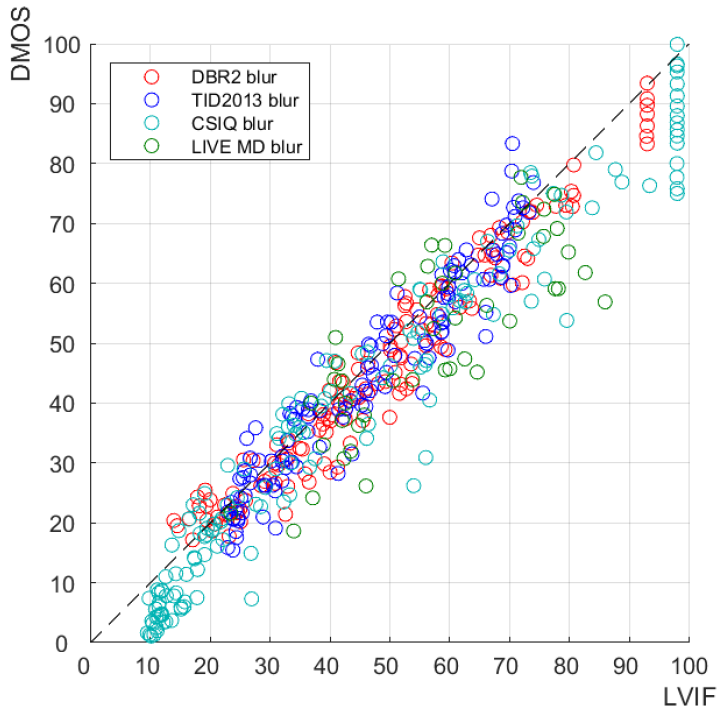

Figure 9.20. The scatterplots of the DMOS empirical measurements for blurred images versus the values predicted using the linearized VIF method for different databases. Performance are in Tab.9.2

Table 9.2. Summary of the experimental verification for the **LVIF** (rectified IQA metrics) for DBR2, TID2013, CSIQ and LIVE MD, and for the overall case. **Only blur** is considered for all databases.

Model	RMSE	SROCC	LCC
LVIF LIVE DBR2	4.3647	0.97222	0.97526
LVIF TID2013	4.9705	0.96529	0.95452
LVIF CSIQ	6.8154	0.97470	0.97409
LVIF LIVE MD	9.3008	0.78363	0.81243
LVIF OVERALL	6.2095	0.96525	0.96196

empirical calibration, whose estimate has the same form as the (9.39):

$$\hat{d}_{LIQA}(\zeta) = 100 \times Q \left[1 - \sqrt{\frac{1}{1 + \frac{\xi(\zeta,\tau)^2}{\tau^4}}} \right] \tag{9.40}$$

The present comparative analysis is limited to degradation types that appeared suitable for conversion to the canonical metric, namely, additive noise, JPEG and JPEG2000 compression, including some examples of *multiple degradations* [48]. To this purpose, four different well-known collections of examples have been employed, namely the above indicated subsets of the LIVE DBR2, the TID2013, the CISQ and the LIVE MD databases. These databases are built using independent protocols and methods for calculating the DMOS, which is used as a common measure of quality loss. Even if indications about the VD used in experiments are available, both the normalized VD τ and the scoring gain Q where independently estimated by regression of the available blur DMOS data, using the canonical estimator.

The results of this comparative analysis are reported in Tab.9.3 for the IQA methods equipped with the scoring functions [93, 94] calibrated on the empirical DMOS values of each database, and in Tab.9.4 for the LIQA methods. The selected statistical quality indices where the RMSE (indicative of the average distance between the actual DMOS values and the ones provided by the method), the SROCC (Spearman's Rank Correlation) [37] which reveals the monotonicity between predicted and actual DMOS values, and the PLCC (Pearson Linear Correlation Coefficient) which measures the linearity of the DMOS mapping.

These tables indicate that the DMOS estimates based on theoretical modeling are statistically slightly more dispersed than the ones determined using empirical fitting to specific data, using the five parameter curve of (9.1). However, this advantage does depend on the adopted fitting model. For instance, the three-parameter curve of [41] would exhibit in principle a larger RMSE in change of a better stability across applications, thus reducing this advantage over LIQAs.

On the other hand, the PLCC and the SROCC indices remain substantially the same for the IQAs and LIQAs methods.

Further indications come from the observation of the scatterograms of empirical DMOS values versus the estimated ones, reported in Fig.9.21,9.22,9.23,9.24. The most evident is that a numerical saturation occurs toward the extremely bad quality, in correspondence of the vertical asymptotes of the intrinsic distortion, of scarce practical relevance.

9.11 Implementation issues

The computational steps for the implementation of the a-priori Linearized methods are summarized as follows:

1. With reference to a typical natural image chosen as specimen whose spectral decay is invariant with respect to the VD stipulate a conventional DMOS for that image at some level of blur (*anchor* value). This determines the value of Q.

Table 9.3. Summary of the experimental verification for the **original IQA metrics** for DBR2, TID2013, CSIQ and LIVE MD. Four distortions subsets for LIVE DBR2, TID2013 and CSIQ: blur, noise, JPEG, JPEG2000. Five distortions subsets for LIVE MD: blur, noise, JPEG, blur+JPEG, blur+noise.

Model	RMSE	SROCC	LCC
VIF LIVE DBR2	7.3992	0.96535	0.96181
VIF TID2013	6.3344	0.92782	0.94021
VIF CSIQ	9.5763	0.92283	0.93089
VIF LIVE MD	7.3096	0.92199	0.94277
VIF OVERALL	–	–	–
MSSIM LIVE DBR2	8.4663	0.95307	0.94969
MSSIM TID2013	6.7297	0.91745	0.93224
MSSIM CSIQ	13.1528	0.88089	0.86502
MSSIM LIVE MD	10.0285	0.87118	0.88922
MSSIM OVERALL	–	–	–
FSIM LIVE DBR2	7.2289	0.96846	0.96358
FSIM TID2013	5.4113	0.95343	0.95673
FSIM CSIQ	10.5632	0.92742	0.91522
FSIM LIVE MD	7.9413	0.90831	0.93207
FSIM OVERALL	–	–	–
GMSD LIVE DBR2	6.9278	0.96888	0.96660
GMSD TID2013	5.5523	0.95059	0.95440
GMSD CSIQ	8.1502	0.95490	0.95044
GMSD LIVE MD	8.7612	0.89582	0.91666
GMSD OVERALL	–	–	–

2. Calculate the intrinsic distortion $\xi(\zeta;\tau)$ of the chosen conventional IQA method for the VD values of interest $\tau_1, \tau_2, ..., \tau_N$, thus obtaining a set of scoring functions for each VD normalized to δ_0. Since the exponential specimen image spectral decay is invariant with respect to VD, each image is resampled for each distance of interest. The image size scales up as τ^{-1}, i.e., it shrinks for a larger normalized VD.

3. Select the IQA method and the corresponding scoring functions $\xi(\zeta;\tau)$, for the selected values of τ.

4. Calculate the IQA metric ζ for the scaled original and distorted image pairs and use it in the calculus of $\hat{d}_{LIQA}(\zeta)$.

Since the step 2 is off-line, the computational cost reduces to that of the conventional IQA method.

Table 9.4. Summary of the experimental verification for the **rectified IQA metrics** for DBR2, TID2013, CSIQ and LIVE MD, and for the overall case. Four distortions subsets for LIVE DBR2, TID2013 and CSIQ: blur, noise, JPEG, JPEG2000. Five distortions subsets for LIVE MD: blur, noise, JPEG, blur+JPEG, blur+noise.

Model	RMSE	SROCC	LCC
LVIF LIVE DBR2	8.6653	0.96536	0.95636
LVIF TID2013	7.0170	0.92782	0.92789
LVIF CSIQ	7.5789	0.95878	0.96336
LVIF LIVE MD	9.2065	0.92673	0.93604
LVIF OVERALL	8.3421	0.9535	0.94307
LMSSIM LIVE DBR2	9.5823	0.95323	0.93910
LMSSIM TID2013	7.8579	0.91745	0.91394
LMSSIM CSIQ	8.5463	0.95445	0.95326
LMSSIM LIVE MD	13.8323	0.87709	0.84556
LMSSIM OVERALL	10.3989	0.92951	0.91107
LFSIM LIVE DBR2	8.8302	0.96851	0.95149
LFSIM TID2013	7.3283	0.95341	0.93291
LFSIM CSIQ	8.1802	0.96164	0.95810
LFSIM LIVE MD	12.0365	0.91389	0.88788
LFSIM OVERALL	10.0099	0.93155	0.91882
LGMSD LIVE DBR2	7.2629	0.96888	0.96478
LGMSD TID2013	6.7091	0.95059	0.93579
LGMSD CSIQ	7.7122	0.96909	0.96232
LGMSD LIVE MD	12.9031	0.90144	0.87629
LGMSD OVERALL	9.1525	0.94904	0.93314

The value of s_B is directly calculated from the blurred and reference images of the databases by straightforward spectral division (see Sect.10.5).

9.12 Conclusive remarks

Resuming, generalizability of results and comparability among different conventional IQA measurements is problematic. Moreover, the availability of IQA databases is inherently limited, and the risk of overfitting is always lurking. Using the theoretical results for the case of blur degradation, which is easy to analyze, an a-priori linearization procedure is derived. Results are then generalized invoking the perceptual equivalence, from the viewpoint of image fidelity, of different causes of degradation. This equivalence deserves some more comments.

This aspect is specifically discussed in [27] where two separate metrics are jointly

employed, and a two-dimensional scoring function is defined to achieve linearity of IQA estimates using a reduced number of adjustable parameters. Therein, it was outlined that the perceptual quality equivalence among blur and noise is interpersonal and then constant. See also [16].

With respect to conventional approaches, a-priori linearization presents the fundamental advantage of allowing for the quality prediction at any wanted VD. In other words, it may be used for calculating whole maps of estimated subjective quality, for instance in conference rooms. This possibility is precluded to conventional example calibrated methods, also because the scarcity of empirical data taken at different VDs [56].

Of course, the resulting rectification free method are not completely independent on the operative scenario since they do still depend on the parameters Q and τ. However, unlike parameters employed in conventional parametric a-posteriori calibration, these parameters have a direct operative meaning.

Let us observe, besides, that an *absolute version* of the calibration free methods is easily defined by fixing the parameters Q and τ. For instance, one could adopt for τ a VD corresponding to that of a specified specimen, and for Q a value which determines a DMOS for the specimen for a given normalized blur.

As a final remark, let us observe that, from a rigorous conceptual viewpoint, the rectification-free methods are not completely independent on empirical data since the settings of the employed IQA methods where already established through a lot of past experiments. However, it is implicitly assumed that these consolidated methods are not subject to significant modifications.

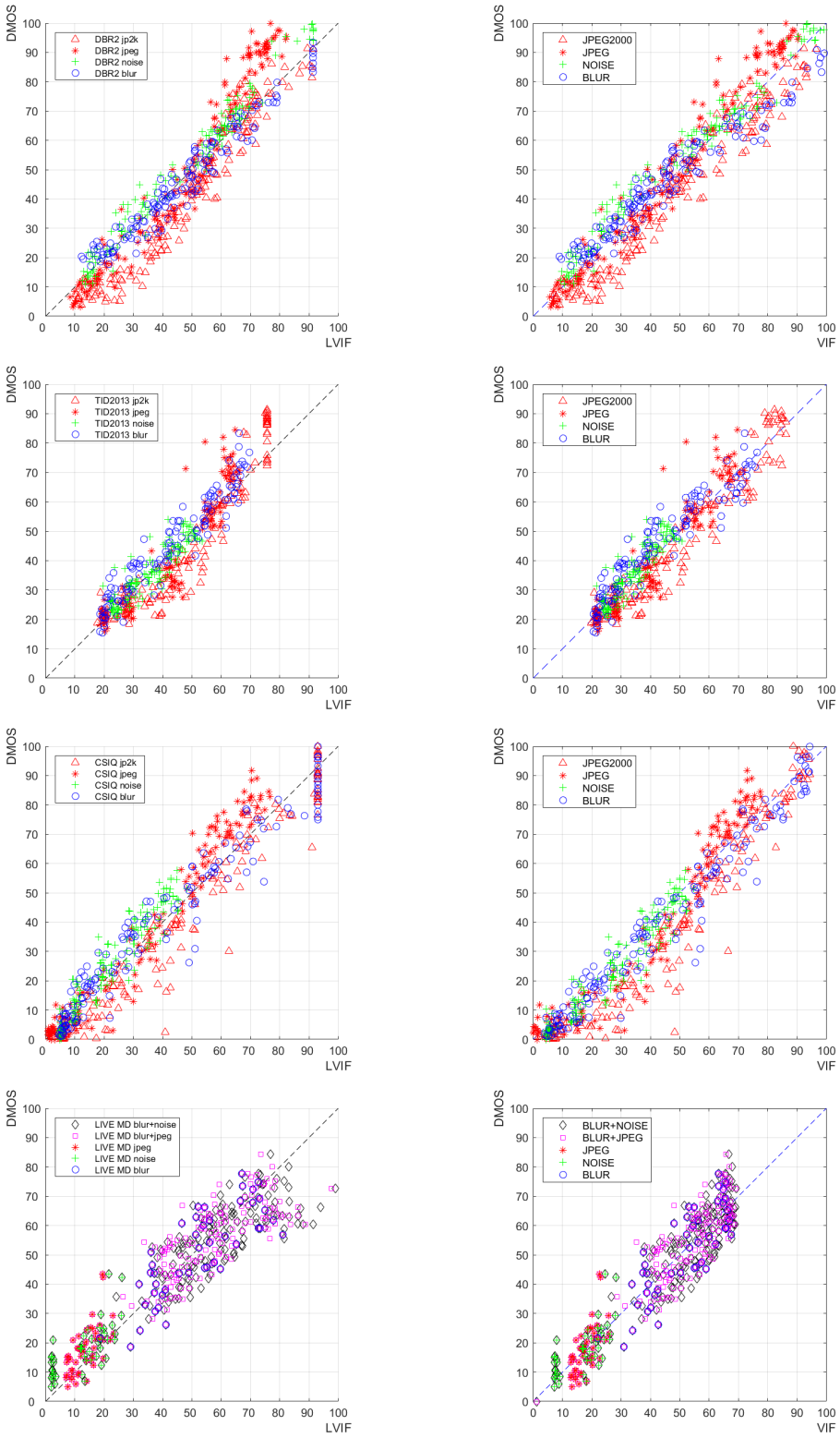

Figure 9.21. The scatterplots of the DMOS empirical measurements for blur, noise, JPEG, JPEG2000 distorsions versus the values predicted using respectively the linearized VIF method (left column) and the original IQA method (right column) for different databases.

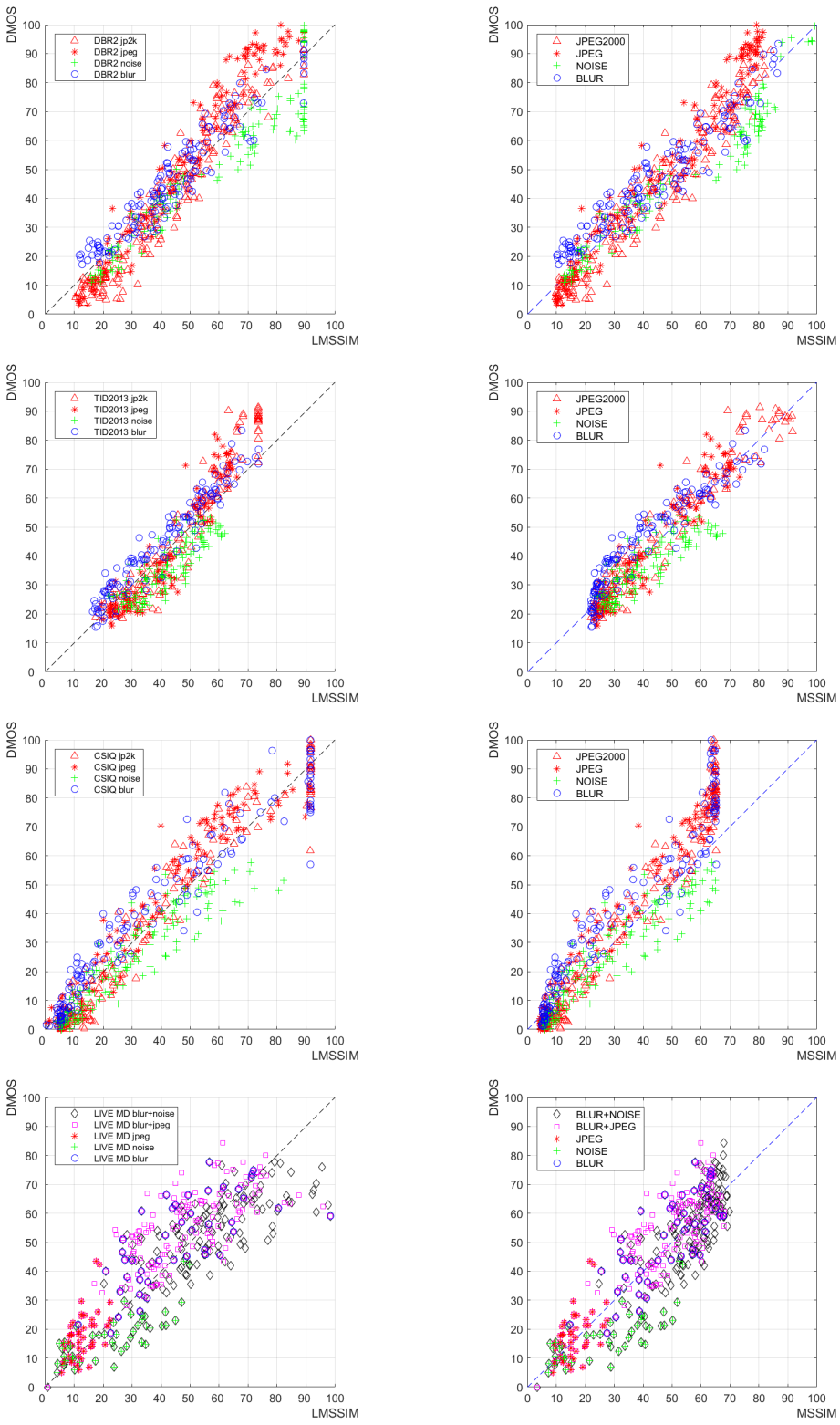

Figure 9.22. The scatterplots of the DMOS empirical measurements for blur, noise, JPEG, JPEG2000 distorsions versus the values predicted using respectively the linearized MS-SSIM method (left column) and the original IQA method (right column) for different databases.

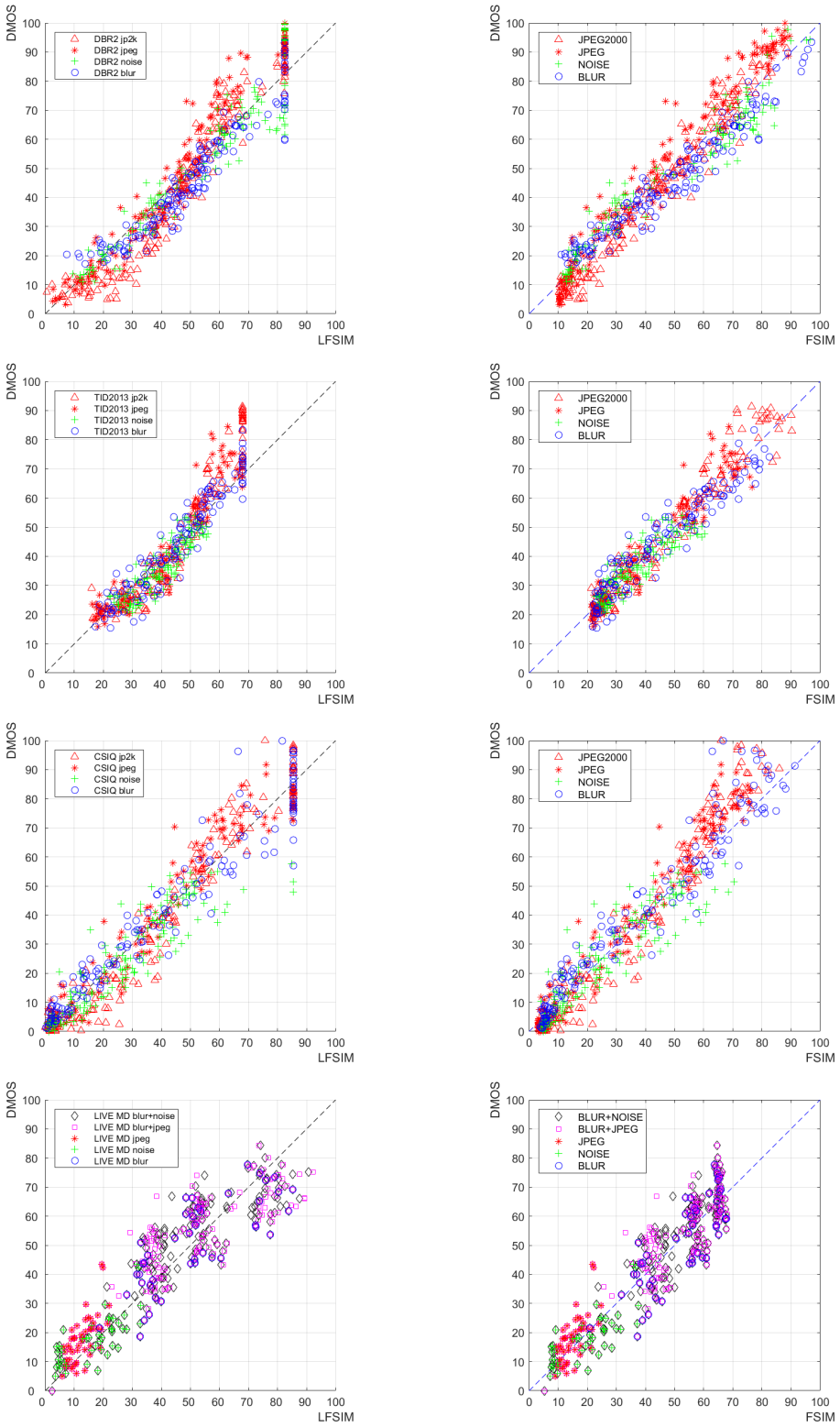

Figure 9.23. The scatterplots of the DMOS empirical measurements for blur, noise, JPEG, JPEG2000 distorsions versus the values predicted using respectively the linearized FSIM method (left column) and the original IQA method (right column) for different databases.

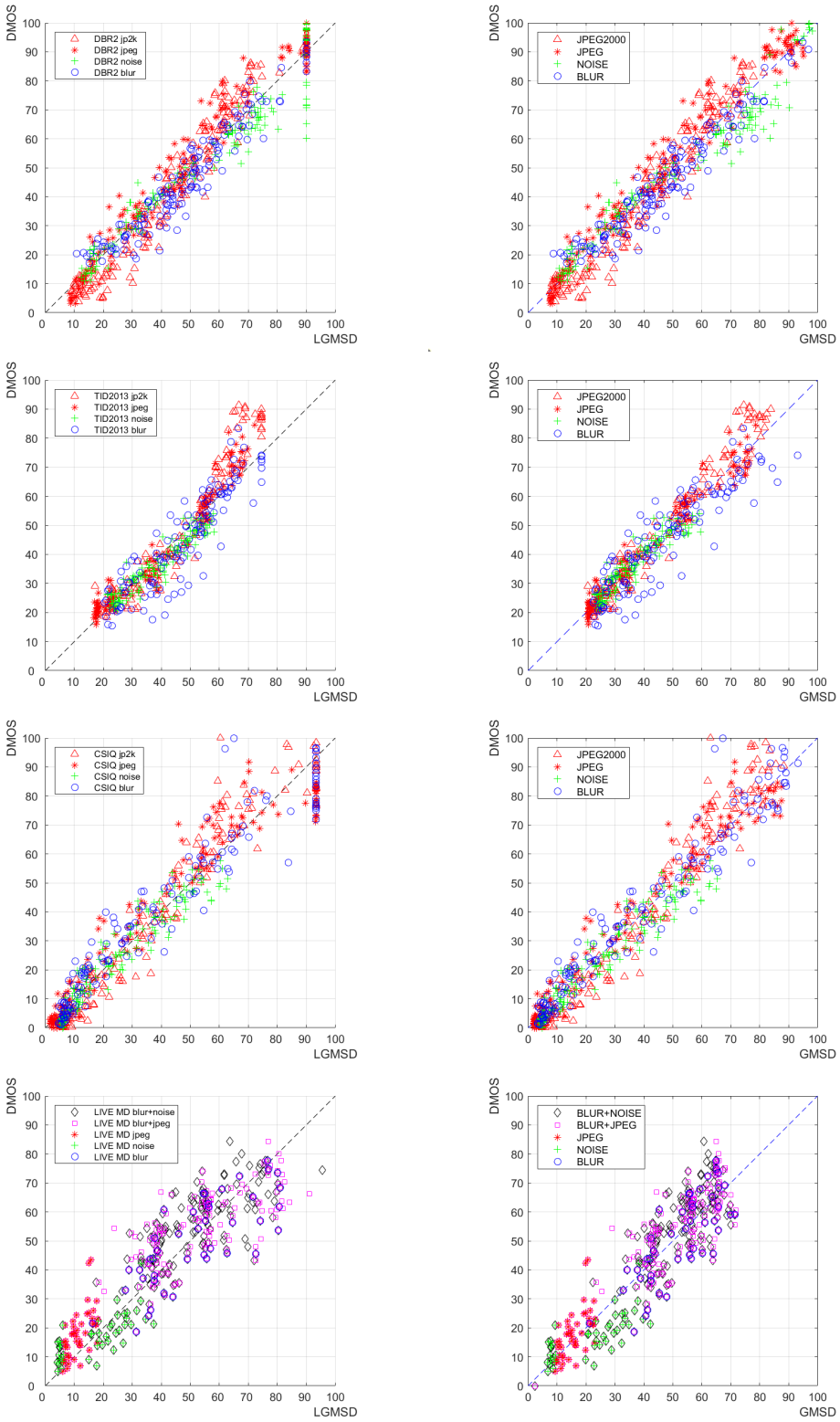

Figure 9.24. The scatterplots of the DMOS empirical measurements for blur, noise, JPEG, JPEG2000 distorsions versus the values predicted using respectively the linearized GMSD method (left column) and the original IQA method (right column) for different databases.

Chapter 10

Diversity of Image Content in the Early Vision

10.1 The raw early sketch

The image content can be thought of as an edge distribution above a certain threshold of visibility. Natural images are composed for the most part by isolated, long, strong, unidirectional edges, and textures (which are composed of weak edges, close to each other, multi-directional, above the noise level, but partially unsolved by the eye), and flat areas which do not carry information and which are also covered by noise. Referring to the VRF model studied in Part III, we now want to understand how these different objects behave with respect to the loss of the PFI. The strong edge is what attracts our attention most [60, 83]. The eye has the ability to compensate for losses up to a certain level of added blur. We want to understand in what proportion and with what impact these different objects represent the contents of an image and contribute to the estimation of subjective quality. Furthermore, we know that if the edge is already partially blurred in the reference image, the task of judging the degradation becomes more difficult.

A strong edge carries Fisher Information in the direction orthogonal to the edge itself (locally it is like a degenerate saddle point). Its degradation due to blur occurs, therefore, only on the cross-section.

The visible textures, that can be seen as edges close to each other, have a much more uniform distribution of Fisher Information on space (distributed in two directions, not separable), and much more variable from the point of view of subjective judgment. With closer edges, the PFI increases, we have two components, and the eye cannot immediately perceives how much and in what direction the portion of image with textures is blurred. Our personal judgment is much more complex and causes large variations in DMOS values.

It has been verified experimentally that removing the weakest edges from the estimate up to a certain threshold, the spread is reduced. The optimal search for the threshold comes from the theoretical estimator studied in Part III and we call it the *natural vision threshold*, dependent on the degree of blur applied. A stable judgment is, therefore, based mainly on the dominant edges.

Starting from a visual inspection of some images of the LIVE MD database,

we can see the proportion and magnitude of strong edges and textures. In the left image of Fig.10.1, we have a dominance of strong edges, which however are already slightly out-of-focus in the original image, due to the imprecision of the depth of field shooting system (see the seat and the profile of the face). We also edges that are weaker but very focused, with strong changes of direction (see hair with more brightness) and textures of low intensity (see darker hair). In the right image of Fig.10.1, we have an example of extended textures (the animals themselves, the field in the foreground and the background), with a few more strong edges (contours of the animal and the strong edge of land-sea separation). We will resume these visual inspections and give them more emphasis in the *Certainty map* section (sec.10.4), where we will explain the isoluminance filter that gives immediate feedback on contents.

Figure 10.1. Two example of the dominance of strong edges (left image) and the dominance of textures (right image). In the second rows, it is used the Canny operator for zero-crossing, only to give a first evidence of the edges and textures.

10.2 Empirical estimation of the DMOS

Starting from the theoretical SBDI estimator already found in Part III, the new estimator must also take into account the variation in subjective quality due to the contents of the image. So, it is possible to define the empirical DMOS as

$$DMOS_{emp} = \alpha(1 - \sqrt{Q}) \tag{10.1}$$

where α is the gain parameter that fixes the *scoring scale* of the database considered, and

$$Q = \frac{1}{N_\Omega} \sum_{p \subset \Omega} \frac{\lambda(\mathbf{p})}{\widetilde{\lambda}(\mathbf{p})} \tag{10.2}$$

N_Ω is the set of points where $\widetilde{y}(\mathbf{p})$ exceeds a minimum value, and

$$\begin{aligned} \lambda(\mathbf{p}) &= \sum_{\mathbf{q}} w_{\mathbf{p}}(\mathbf{q})^2 \left| y_\delta(\mathbf{p} - \mathbf{q}) \right|^2 , \\ \widetilde{\lambda}(\mathbf{p}) &= \sum_{\mathbf{q}} w_{\mathbf{p}}(\mathbf{q})^2 \left| \widetilde{y}_\delta(\mathbf{p} - \mathbf{q}) \right|^2 . \end{aligned} \tag{10.3}$$

$y_\delta(\mathbf{p})$ is the *visual map* (see (9.5)) for the blurred image, calculated at the actual distance, and $\widetilde{y}_\delta(\mathbf{p})$ for the reference image.

For isolated edges, let us consider, without loss of generality, a vertical edge placed on $x_2 = 0$, Gaussian smoothed at nominal VD:

$$\widetilde{y}(x) = \delta(x_1) * F^{-1}(e^{-s_G^2 \rho^2}) \tag{10.4}$$

Its Fourier transform is

$$Y(f) = \delta(f_2) \cdot e^{-s_G^2(f_1^2 + f_2^2)} = \delta(f_2) \cdot e^{-s_G^2 f_1^2} \tag{10.5}$$

We have for the reference

$$\widetilde{y}(x_1, 0) = \frac{\sqrt{\pi}}{s_G} e^{-\frac{\pi^2 x_1^2}{s_G^2}} \tag{10.6}$$

and for the blurred version

$$y(x_1, 0) = \frac{\sqrt{\pi}}{\sqrt{s_G^2 + s_B^2}} e^{-\frac{\pi^2 x_1^2}{s_G^2 + s_B^2}} \tag{10.7}$$

Therefore, the ratio of the profiles is

$$\frac{y(x_1, 0)}{\widetilde{y}(x_1, 0)} = \frac{\frac{\sqrt{\pi}}{\sqrt{s_G^2 + s_B^2}} e^{-\frac{\pi^2 x^2}{s_G^2 + s_B^2}}}{\frac{\sqrt{\pi}}{s_G} e^{-\frac{\pi^2 x^2}{s_G^2}}} = \frac{s_G}{\sqrt{s_G^2 + s_B^2}} = \sqrt{\frac{1}{1 + \frac{s_B^2}{s_G^2}}} \tag{10.8}$$

This result is useful as a reference and it is the starting point for the certainty map in section 10.4. In the first column of Fig.10.2, it is shown an isolated, vertical edge in teal color (purple color is an effect of the window of the operator adopted). Moreover, in the second column, it is possible to appreciate the difference between an isolated edge (left of the image) compared to edges that are closer to each other (right of the image). The color changes according to the proximity of the edges: edges closest to each other take on a red color in the map (greater degradation due to blur), more isolated edges keep the teal color, an index of resistance to blur.

From a theoretical point of view, in a single resolution case

$$s_G^{(\delta)} = s_G \frac{\delta}{\delta_0} \,, \tag{10.9}$$

with reference to the Structure Tensor (see section 5.3)

$$J = \begin{bmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{bmatrix} \tag{10.10}$$

The PFI in presence of additive white Gaussian noise of variance σ_N^2 is

$$\frac{\lambda_1 + \lambda_2}{\sigma_N^2} \tag{10.11}$$

where λ_1 and λ_2 represent the contributions to the PFI along the orthogonal axes of the eigenvectors of the Structure Tensor.

In general, the loss of information due to blur is determined by both the eigenvalues.

In the case where the image is 1D, typically correspondent to isolated edges, the loss of positional information due to blur is substantially determined by the highest eigenvalue, since the effect of blur along the orthogonal direction is small.

The loss of positional information is defined as:

$$Q = \frac{1}{N_\Omega} \sum_{p \subset \Omega} \frac{\lambda_1(\mathbf{p})}{\widetilde{\lambda}_1(\mathbf{p})} + \frac{1}{N_\Omega} \sum_{p \subset \Omega} \frac{\lambda_2(\mathbf{p})}{\widetilde{\lambda}_2(\mathbf{p})} \tag{10.12}$$

where the second term is small in correspondence of edges.

In the frequency domain, the correct sampling of the images is assumed, according to the Nyquist criterion. At nominal VD δ_0, the spatial Fourier spectra are limited into the the range $(-\frac{1}{2}, +\frac{1}{2})$ *cycles/arcmin* $((-30, +30)$ *cycles/degree*$)$. For closer VDs, the spectra of the retinal images range in the band from $-\frac{1}{2}\frac{\delta}{\delta_0}$ to $+\frac{1}{2}\frac{\delta}{\delta_0}$ *cycles/arcmin* $((-30\frac{\delta}{\delta_0}, +30\frac{\delta}{\delta_0})$ *cycles/degree*$)$. Using a N points FFT representation, the spectra range from $-\frac{N}{2}\frac{\delta}{\delta_0}$ to $+\frac{N}{2}\frac{\delta}{\delta_0}$.

Let us now think of the map defined as:

$$\begin{aligned} M(\mathbf{p}) &= \frac{|y(\mathbf{p})|}{|\widetilde{y}(\mathbf{p})|} \,, \quad if \ \frac{|y(\mathbf{p})|}{|\widetilde{y}(\mathbf{p})|} \leq 1 \,, \\ M(\mathbf{p}) &= 0 \,, \qquad\qquad elsewhere \,. \end{aligned} \tag{10.13}$$

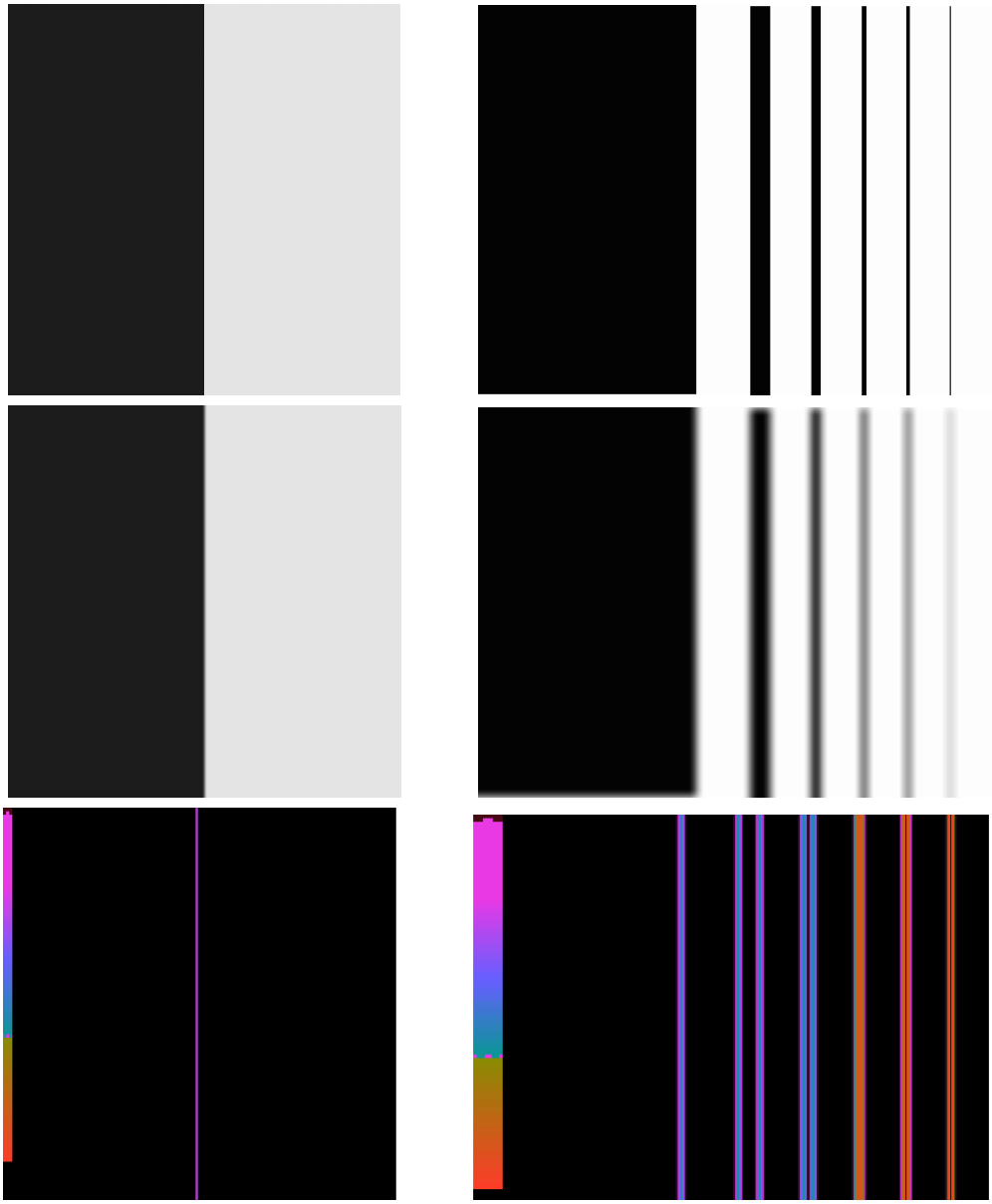

Figure 10.2. Upper row: two original images. Second rows: Gaussian blurred images with $s_B = 3.7$. Lower rows: certainty maps (see also section 10.4). The images in the left column correspond to an isolated edge in teal color. The images in the right column correspond to several edges and the color changes according to the proximity of the edges: edges closest to each other take on a red color in the map (greater degradation due to blur), more isolated edges keep the teal color, an index of resistance to blur.

A good separation threshold between strong edges, which give the main Fisher information contribution in the estimation, and weak edges or textures, is what we can define as *natural vision threshold*:

$$\frac{y(x_1,0)}{\widetilde{y}(x_1,0)} = \sqrt{\frac{1}{1+\frac{s_B^2}{s_G^2}}} \tag{10.14}$$

The empirical linear estimator is obtained from the portion of the image where the natural and strong edges dominate:

$$M(\mathbf{p}) \geqslant \sqrt{\frac{1}{1+\frac{s_B^2}{s_G^2}}} \tag{10.15}$$

so, the empirical SBDI is defined as:

$$SBDI_{empirical} = \alpha(1-\sqrt{Q}) \tag{10.16}$$

where α is the gain parameter that fixes the *scoring scale* of the database, and

$$\begin{aligned} &Q = \frac{1}{N_\Omega}\sum_{p\subset\Omega}\frac{\lambda(\mathbf{p})}{\widetilde{\lambda}(\mathbf{p})}\,, \quad if\ M(\mathbf{p}) \geqslant \sqrt{\frac{1}{1+\frac{s_B^2}{s_G^2}}}\,, \\ &Q = 0\,, \qquad\qquad\qquad elsewhere\,. \end{aligned} \tag{10.17}$$

10.3 Experimental results and considerations

The performance of the empirical estimator found in the previous section, the empirical SBDI, are compared here with those of the IQA methods known in literature. For the evaluation of the goodness of an estimator, two fundamental requirements are the linearity and the r.m.s. of the scatterplot of a metric with respect to the DMOS values. In the reference IQA estimators, these requirements are satisfied by applying neural rectification as a post-processing. In particular, linearization was performed by a parameter (blur) monotonic logistic function, trained by the BRLS modified Newton algorithm [73] over 250k runs, to minimize the risk of stop into local minima. Good performance evaluation indices are MSE, SROCC (The Spearman rank-order correlation measures the relative monotonicity between predictions and DMOS values), and LCC (The Pearson correlation coefficient measures how good is the correlation of the predictions with the DMOS values): the lower their values are, more uniform and linear will be the scatterplot of the estimator. Here, we consider both the best performing and most used estimators in the real applications. As an experimental verification, the metrics related to the following databases are compared for the IQA Full Reference scenario: LIVE DBR2 [81], TID2013 [75] and LIVE MD [48]. The performance of the empirical SBDI are compared to the theoretical SBDI (found in Part III), the VIF [80, 81], the MS-SSIM [95], the FSIM [109] and the GMSD [108] methods.

As explained in section 8.2.2, in the TID2013 database, ratings are available as MOS values, so the DMOS values were inferred considering that the best MOS ratings do not exceed 7.5. Therefore, posing $MOS = 7.5$ in correspondence to $DMOS = 0$ and $MOS = 0$ in correspondence to $DMOS = 100$ yields $DMOS = (100/7.5) \cdot (7.5 - MOS)$.

Tab.10.1 shows the performances of the estimators for LIVE DBR2, Tab.10.2 for TID2013 and Tab.10.3 for LIVE MD. Fig.10.4 shows the scatterplot of the empirical SBDI estimator for the databases considered. Finally, in Tab.10.4, the overall performances for the combined databases (LIVE DBR2, TID2013 and LIVE MD) are shown.

It is possible to appreciate how the empirical SBDI has performances comparable to those of the best estimators for each DB (VIF for LIVE DBR2 and LIVE MD, MS-SSIM for TID2013) and for overall performance (VIF). It is also important to point out that performance is compared between a linear estimator and IQA estimators that need neural regression. Furthermore, the simplicity of implementation allows us to save costs and resources compared to the well-known IQA estimators, which are multi-channel and multi-resolution.

Table 10.1. Summary of the experimental verification for the **LIVE DBR2** database for Gaussian blur.

Model	RMSE	SROCC	LCC
VIF	**4.1546**	**0.9722**	**0.9744**
MSSIM	5.4069	0.9538	0.9562
FSIM	4.2168	0.9708	0.9736
GMSD	5.0932	0.9567	0.9612
$\text{SBDI}_{theoretical}$	5.4318	0.9579	0.9604
$\text{SBDI}_{empirical}$	**4.4068**	**0.9686**	**0.9715**

Table 10.2. Summary of the experimental verification for the **TID2013** database for Gaussian blur.

Model	RMSE	SROCC	LCC
VIF	5.0236	0.9653	0.9533
MSSIM	**4.4638**	**0.9676**	**0.9633**
FSIM	4.9398	0.9551	0.9549
GMSD	6.9378	0.9114	0.9089
$\text{SBDI}_{theoretical}$	6.8326	0.9339	0.9188
$\text{SBDI}_{empirical}$	**4.5994**	**0.9653**	**0.9630**

Table 10.3. Summary of the experimental verification for the **LIVE MD** database for Gaussian blur.

Model	RMSE	SROCC	LCC
VIF	**8.1955**	**0.7836**	**0.8376**
MSSIM	9.4893	0.7684	0.7745
FSIM	8.8431	0.7693	0.8078
GMSD	9.2812	0.7728	0.7856
$\text{SBDI}_{theoretical}$	8.6268	0.8276	0.8301
$\text{SBDI}_{empirical}$	**8.6194**	**0.7925**	**0.8274**

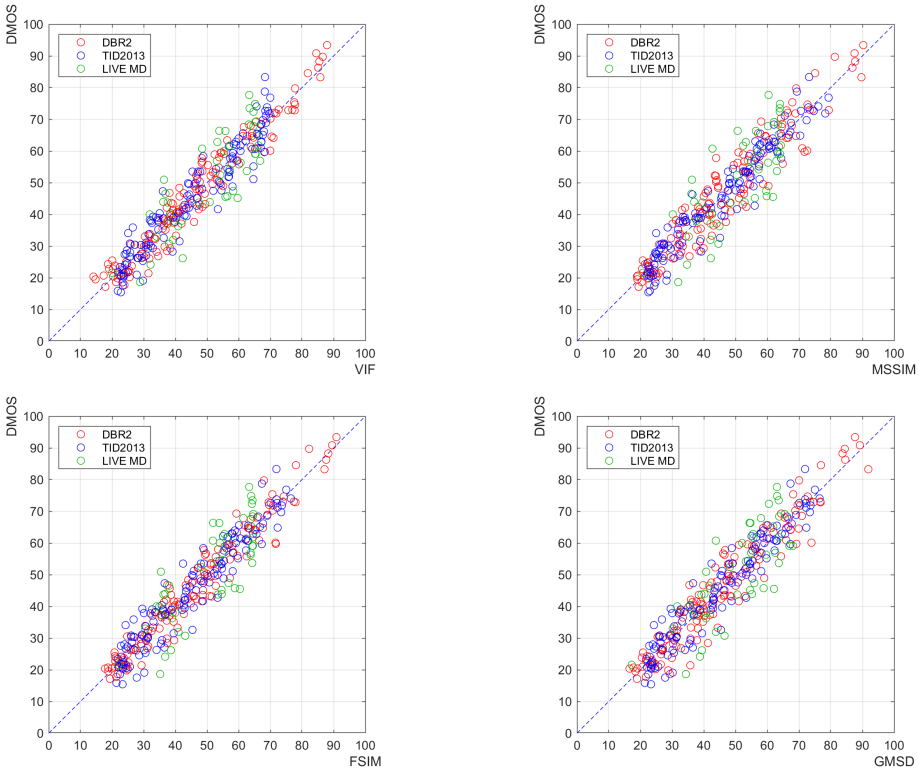

Figure 10.3. The scatterplot of the DMOS ratings versus the corresponding **IQA methods** for the blurred image in LIVE DBR2, TID2013 and LIVE MD databases. Linearization was performed by a parameter (blur) monotonic logistic function, trained by the BRLS modified Newton algorithm over 250k runs.

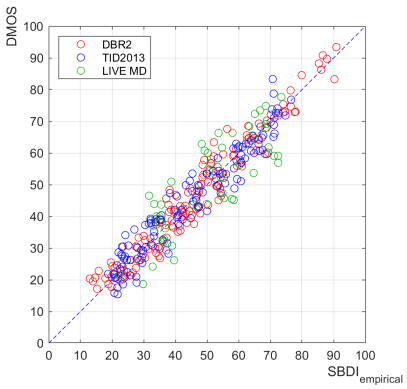

Figure 10.4. The scatterplot of the DMOS ratings versus the corresponding **SBDI empirical** values predicted by the model for the blurred image in LIVE DBR2, TID2013 and LIVE MD databases. RMSE=5.3081, SROCC=0.9524, LCC=0.9537.

Table 10.4. Summary of the experimental verification for the **DBR2, TID2013 and LIVE MD** databases for Gaussian blur.

Model	RMSE	SROCC	LCC
VIF	**5.2667**	**0.9533**	**0.9531**
MSSIM	5.8504	0.9410	0.9418
FSIM	5.3889	0.9498	0.9509
GMSD	5.8249	0.9425	0.9423
$\text{SBDI}_{empirical}$	**5.2865**	**0.9526**	**0.9540**

10.4 Certainty map for blur degradation

Recalling the equation 10.13, we now define the *certainty map* as:

$$\begin{aligned} M(\mathbf{p}) &= \frac{|y(\mathbf{p})|}{|\widetilde{y}(\mathbf{p})|}\,, \quad if\ \frac{|y(\mathbf{p})|}{|\widetilde{y}(\mathbf{p})|} \leq 1\,, \\ M(\mathbf{p}) &= 0\,, \qquad\qquad elsewhere\,. \end{aligned} \tag{10.18}$$

The *weighted certainty map* is then obtained by multiplication by the local contrast given by

$$MW(\mathbf{p}) = M(\mathbf{p}) \cdot f\left(\frac{|\widetilde{y}(\mathbf{p})|}{|\widetilde{y}_{MAX}|}\right)\,. \tag{10.19}$$

$MW(\mathbf{p})$ has been *conventionally* mapped to colors by attributing warm colors to small values (edges close together) and cold colors to large values (isolated edges), referring to the value that $M(\mathbf{p})$ assumes at ideal edges in the original image, a *nominal* value which is:

$$\bar{M}(\mathbf{p}) = \sqrt{\frac{s_G^2}{s_G^2 + s_B^2}} \tag{10.20}$$

$f\left(\frac{|\widetilde{y}(\mathbf{p})|}{|\widetilde{y}_{MAX}|}\right)$ follows the logarithmic law $log\left(1 + \frac{|\widetilde{y}(\mathbf{p})|}{|\widetilde{y}_{MAX}|}\right)$ to enhance the intensity of stronger edges.

We can therefore divide the image into a set of Ω_C points where the losses are lower than the nominal one due to the fact that there are already blurs in the original image (cold area):

$$M(\mathbf{p}) \geqslant \sqrt{\frac{s_G^2}{s_G^2 + s_B^2}} \tag{10.21}$$

and a set of Ω_H points where the losses are higher than the nominal one because they are not isolated straight edges (hot area):

$$M(\mathbf{p}) < \sqrt{\frac{s_G^2}{s_G^2 + s_B^2}} \tag{10.22}$$

Furthermore, this can also happen if there is an *edge enhancement* in the original image. This occurs in some images of the LIVE MD database, causing a bias. In the case of images or portions of the image already blurred in the reference, the effect of blur is less and in the certainty map we will see a color tending to purple, which indicates greater certainty. If the reference images are perfectly focused, we will see the isolated and strong edges of a teal color, as they are in the range of natural losses. In fact, we know from the theory that if we only had isolated and strong edges, perfectly in focus, the image spectrum goes as $1/\rho^2$ (see Part III) and in our map this theoretical threshold coincides with the teal color.

In Fig.10.5 and Fig.10.6, in the colorbar on the left of the image, we can distinguish a purple dotted line that represents the theoretical index (SBDI found in Part III) useful for distinguishing the threshold of natural losses and dependent on the level of blur s_B (it is the reference color and will always be teal). On the other hand, the dark dotted line represents the subjective judgment given to the image (the DMOS values). In the case of the image in the left column of Fig.10.5, for example, the subjective judgment is better than that estimated by the theoretical index. This indicates that the image is rated better than it really is, as shown by the dominance of the purple color. On the contrary, in the image of the right column, we note how the subjective judgment is worse than the theoretical one. The image is perceived very degraded, and the dominance of red gives the visual feedback. In Fig.10.6, we have two images with no dominance of red or purple. Let's see how, as the blur level increases from $s_B = 1.5$ to $s_B = 3.7$, the image degradation increases, in particular, where the edges are less strong and close to each other. On the contrary, we appreciate how strong and isolated edges are less affected by the added blur and keep their purple color, a sign of high certainty.

Figure 10.5. Upper row: two original images. Second rows: Gaussian blurred images. Lower rows: certainty maps. The images in the left column correspond to an image judged with very low DMOS (high quality), on the other hand, in the right column with very high DMOS (low quality). It is possible to appreciate a high dominance of purple which represents greater certainty, in the image with higher quality (left column), and a high dominance of reds which represents greater loss, in the image with worse quality (right column). The isoluminance colors maintain a constant intensity at the same edge level, so, only the weight defines the local contrast. The blur levels are $s_B = 2.4$ for the left column image and $s_B = 2.5$ for the right column image.

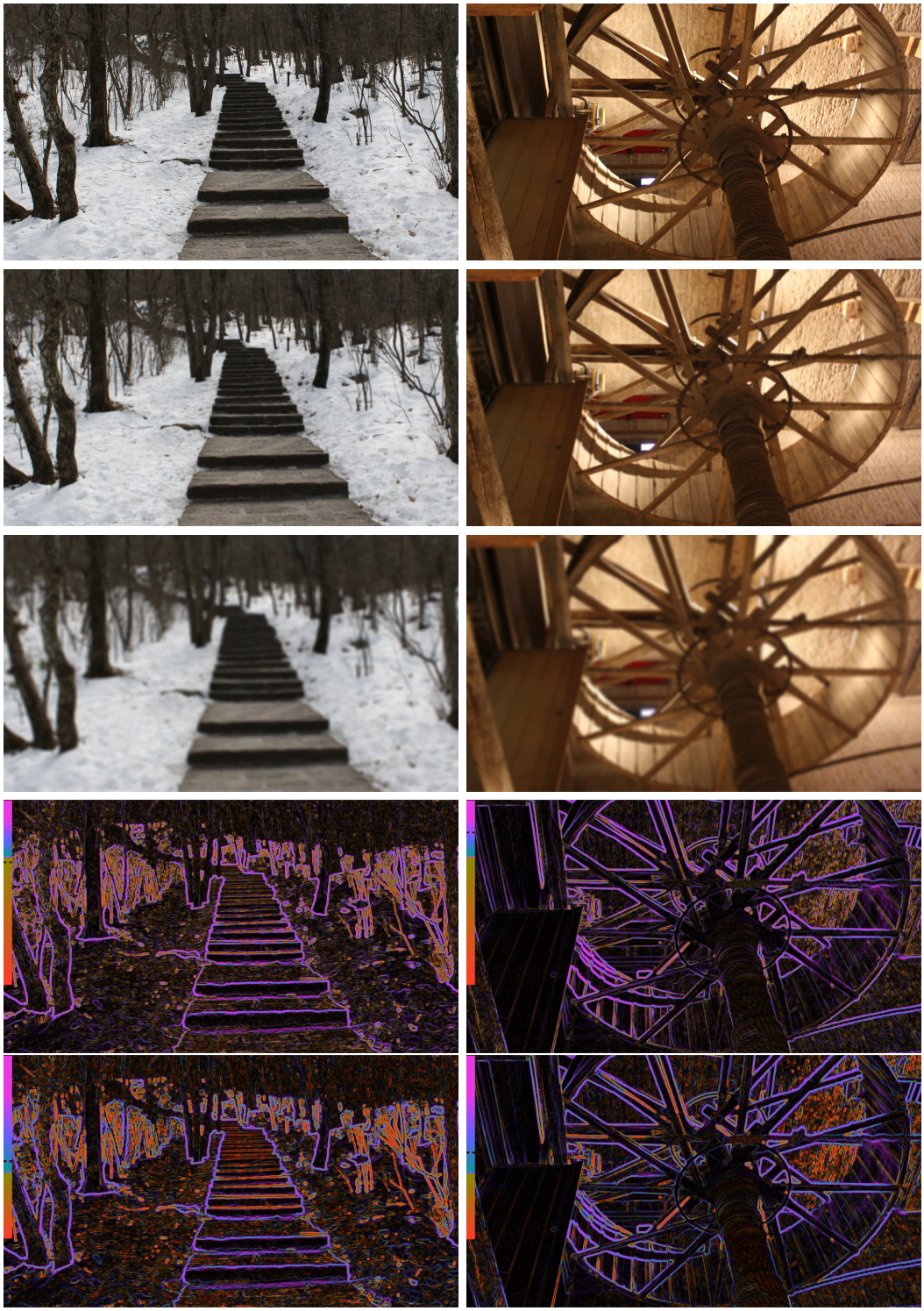

Figure 10.6. Upper row: two original images. Second rows: Gaussian blurred images with $s_B = 1.5$. Third rows: Gaussian blurred images with $s_B = 3.7$. Lower rows: certainty maps for $s_B = 1.5$ and $s_B = 3.7$. As the blur level increases from $s_B = 1.5$ to $s_B = 3.7$, image degradation (red color) increases where the edges are less strong and close to each other. On the contrary, strong and isolated edges are less affected by the added blur (purple color).

10.5 Empirical estimation of the blur value

At this point, it could be argued that to find the proposed empirical estimator, it is necessary to know the value of the added blur (s_B), different for each image, required to have the natural vision threshold. To overcome this limitation, working on spectral analysis, it is possible to estimate this parameter with good precision starting from the real part of the ratio of the Fourier transform of the blurred image and the Fourier transform of the reference image.

$$I\left(\mathbf{k}\right) \equiv \left(2\pi\right)^{-1} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{i\mathbf{kx}} I\left(\mathbf{x}\right) d\mathbf{x} \tag{10.23}$$

$$\widetilde{I}\left(\mathbf{k}\right) \equiv \left(2\pi\right)^{-1} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{i\mathbf{kx}} \widetilde{I}\left(\mathbf{x}\right) d\mathbf{x} \tag{10.24}$$

$$P(\mathbf{k}) \equiv Re\left(\frac{I(\mathbf{k})}{\widetilde{I}(\mathbf{k})}\right) \tag{10.25}$$

where $P(\mathbf{k})$ is the profile of the ratio.

In Fig.10.7, 3D images of this ratio are shown for images of the LIVE DBR2 database with different blur levels.

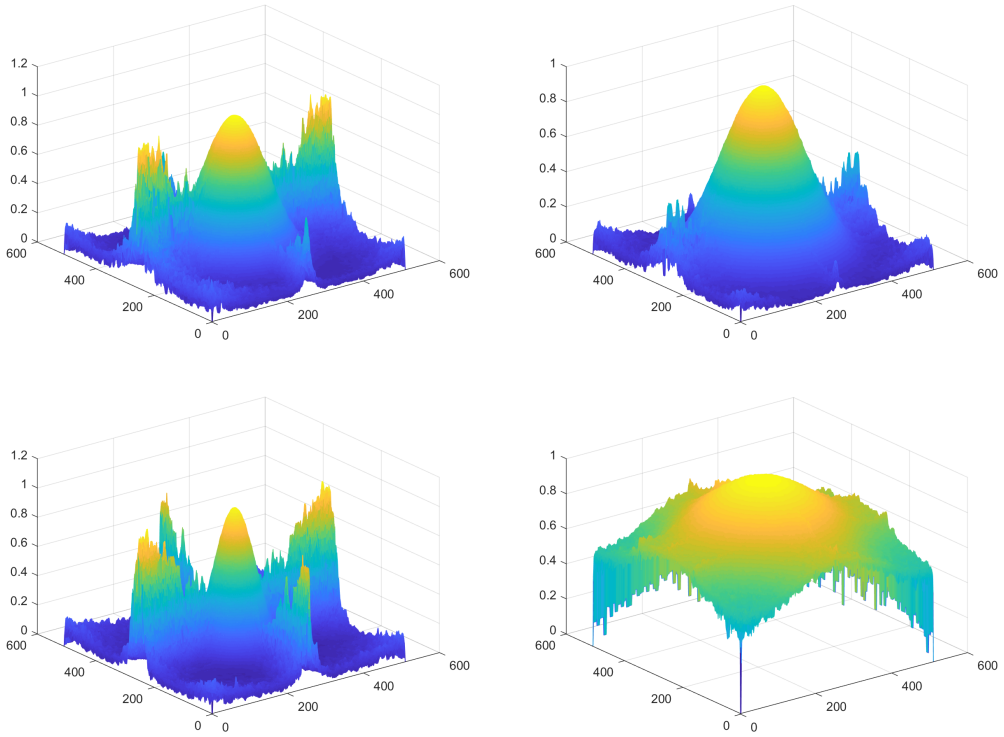

Figure 10.7. 3D representation of the real part of the ratio of the Fourier transform of the blurred image and the Fourier transform of the reference image for different images and blur levels.

In Fig.10.8, a median profile is shown for s_B that increases from 0.5, 0.65, 1 to 1.7. A median profile is defined as:

$$\bar{P}(k) = median\left[Re\left(\frac{I(\mathbf{k})}{\widetilde{I}(\mathbf{k})}\right)\right] \tag{10.26}$$

where $Re\left(\frac{I(\mathbf{k})}{\widetilde{I}(\mathbf{k})}\right)$ is a vector representing 128 viewing angles (rotations around the axis passing between 0 and the center of the Gaussian peak). The theoretical Hermite-Gauss kernel is shown with a black curve and the empirical result is in a blue curve.

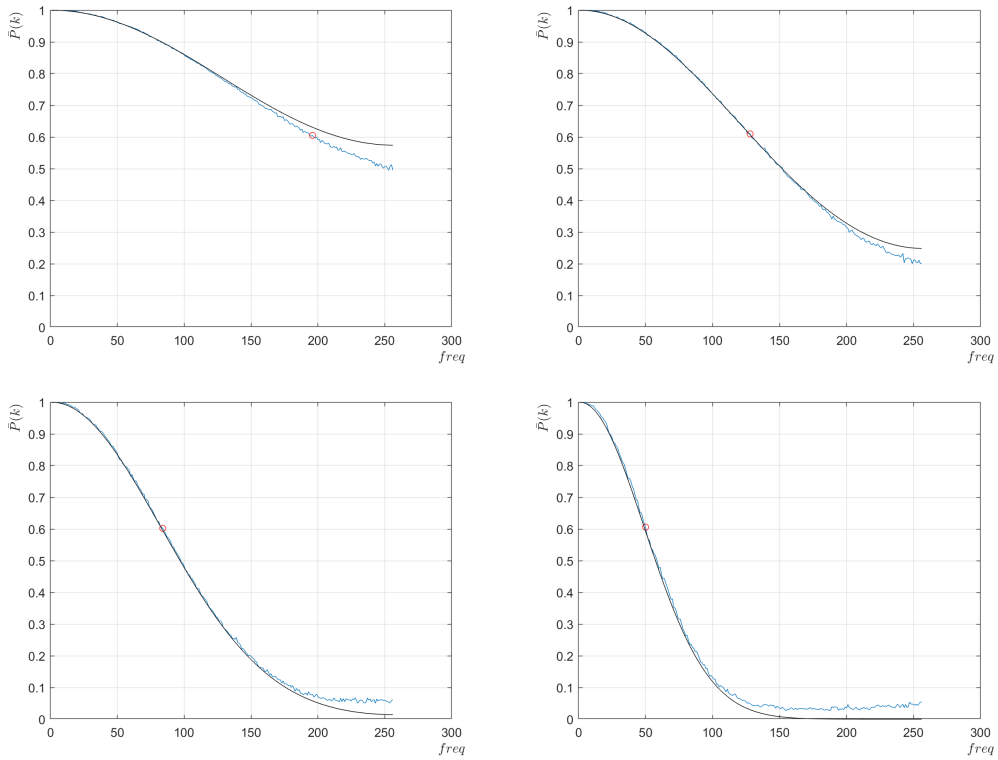

Figure 10.8. The median profile for s_B that increases from 0.5 (upper-left image), 0.65 (upper-right image), 1 (lower-left image) to 1.7 (lower-right image).

By selecting the frequency point corresponding to $\bar{P}(k) \equiv 0.6$, it is possible to obtain the estimated s_B as:

$$\hat{s}_B = \frac{N_{FFT}/2}{\pi \times \text{freq}_{0.6}} \tag{10.27}$$

where N_{FFT} is the number of points for the FFT transform and $\text{freq}_{0.6}$ is the frequency value corresponding to $\bar{P}(k) \equiv 0.6$.

In Fig.10.9, the scatterplots of the reference s_B from LIVE DBR2 and TID2013 (the only databases that provide these data) and the estimated s_B are shown[1].

[1]In TID2013 database, the particular case with $s_B = 4$ has a kernel error in the reference values. Hence, the estimated s_B are the correct ones (communication with the author Ponomarenko by e-mail on November 11, 2019).

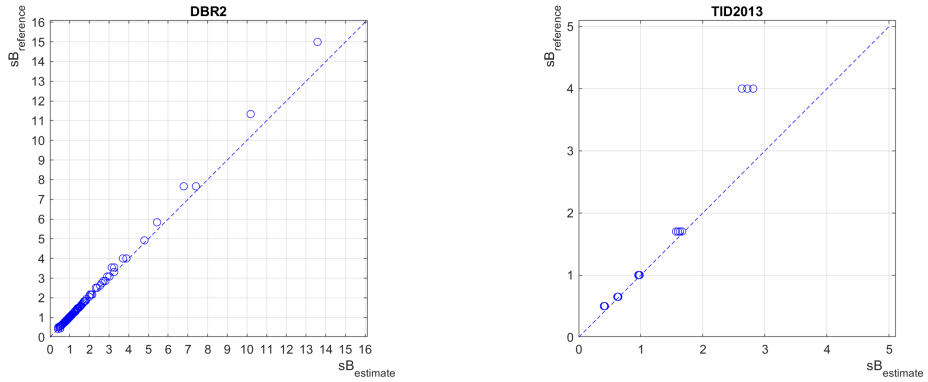

Figure 10.9. The scatterplots of the reference s_B versus the estimate s_B for the blurred image in LIVE DBR2 and TID2013 databases.

In Tab.10.5, the performances of the s_B estimators for the LIVE DBR2 and TID2013 are shown.

Table 10.5. Summary of the experimental verification for the s_B estimation for the LIVE DBR2 and TID2013 databases.

Database	RMSE	SROCC	LCC
LIVE DBR2	0.2158	0.9994	0.9992
TID2013	0.5751	0.9829	0.9817

Conclusions and Remarks

In this thesis, a new model of blur based on the loss of Positional Fisher Information (PFI) has been presented. The loss of Fisher Information has been shown to be directly linked to visual discomfort due to blur and the "Scaled Blur Discomfort Index" (SBDI) has been obtained from a theoretical model with two parameters: the Viewing Distance (VD) parameter of the experiment (γ or its inverse τ) and the amplification coefficient (the scoring scale Q) of the DB. A useful application has been provided to map the loss of a generic Optical Transfer Function (OPT) to the Gaussian case and the range of validity for the sinc and astigmatic cases has been given. The adoption of a polar separable complex-valued receptive field model, and of a visual information loss criterion, predicts well the blur visual discomfort for natural scenes and reflects the cognitive deficiencies of fine localization of the objects in a natural image. This abstract, functional model may provide coarse estimates of the discomfort caused by undesired blur, for optical correction in natural vision, and for calibration of image reproduction apparatus.

With respect to conventional empirical approaches, a-priori linearization has been presented with the advantage of predicting the quality at any wanted VD. This possibility is precluded to conventional example-calibrated methods, also because of the scarcity of empirical data taken at different VD. Let us underline that in the present approach we are not estimating the quality of specific images, but the quality of realizations of a relatively tiny random set of images, satisfying recognized ensemble properties of Gaussian blurred natural images as a function of the normalized blur spread, and parameterized by the VD and an unpredictable multiplicative parameter Q, the so called rating gain. Moreover, the linearity of such estimates is derived from principles assumed as universally valid, namely the optimality of Human Visual System (HVS) with respect to structure localization and from the Weber-Fechner psycho-physical law. Several DMOS maps have been provided to calculate the new Linearized Image Quality Assessment (LIQA) metrics at the desired VD. Of course, the resulting rectification free method are not completely independent on the operative scenario, since they do still depend on the parameters Q and τ. However, unlike parameters employed in conventional parametric a-posteriori calibration, these parameters have a direct operative meaning. Furthermore, from a strict conceptual viewpoint, the rectification-free methods are not absolutely independent of empirical data, since the settings of the employed IQA methods were already established through a lot of past experiments. It is implicitly assumed here that these consolidated methods are not subject to significant modifications.

Finally, an empirical estimator has been presented based on the selection of details with stronger, isolated and longer edges, which takes better account of the

eye behavior: it is known, indeed, that the difference with the real cases of natural scenes directly projected onto the retina (not through a display) is that the eye does not always see each and every details in the image (isolated gradients that are too small and not identifiable) or the displacement due to fact that the loss of the Fisher Information is so complex that noise is replaced by other noise (textures). Therefore, we need to adjust the metric with the HVS. In addition to the estimator, a filter has been provided to clearly highlight where the losses due to blur in an image are and what intensity they have.

The next step will be the study of a model that takes into account the contents of the images that can be extended to all distortions. This approach will be critical to provide a scientific justification to the variation in subjective quality values already found in the images of the databases considered. A new definition of a noise model is also needed. The logarithmic function used in many performing estimators for the noise metric is not justifiable for a good theoretical model and lacks theoretical principles. It is known, indeed, that blur and noise form an orthogonal basis to model apparently more complex distortions. By combining the two models, it will be possible to have a performing and linear estimate of the complete set of distortions for all databases. Other important aspects so far left out will need to be further analyzed to include them in the model: the maximum luminance of the display and its dynamic range expressed as the ratio between the luminances of pure white and pure black in a dark environment; ambient illumination; extension of operator to a 3-D space by adding a third orthogonal axis (coding-related distortions and artifacts).

Bibliography

[1] Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19** (1974), 716. doi:10.1109/TAC.1974.1100705.

[2] Athar, S. and Wang, Z. A comprehensive performance evaluation of image quality assessment algorithms. *IEEE Access*, **7** (2019), 140030. doi:10.1109/ACCESS.2019.2943319.

[3] Badcock, D. R. How do we discriminate relative spatial phase? *Vision Research*, **24** (1984), 1847. doi:https://doi.org/10.1016/0042-6989(84)90017-8.

[4] Baroncini, V., Capodiferro, L., Di Claudio, E. D., and Jacovitti, G. The polar edge coherence: A quasi blind metric for video quality assessment. In *2009 17th European Signal Processing Conference*, pp. 564–568 (2009).

[5] Bell, A. J. and Sejnowski, T. J. The independent components of natural scenes are edge filters. *Vision Research*, **37** (1997), 3327 . doi:https://doi.org/10.1016/S0042-6989(97)00121-1.

[6] Berry, R. H., Hobson, M. P., and Withington, S. Modal decomposition of astronomical images with application to shapelets. *Monthly Notices of the Royal Astronomical Society*, **354** (2004), 199. doi:10.1111/j.1365-2966.2004.08180.x.

[7] Blakemore, C. and Campbell, F. W. On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *The Journal of Physiology*, **203** (1969), 237. doi:https://doi.org/10.1113/jphysiol.1969.sp008862.

[8] Blonstein, S. D. and Huang, T. S. Error analysis in stereo determination of 3-D point positions. *IEEE Trans. on Pattern Analysis amd Machine Intelligence*, **9** (1987), 752.

[9] Bocheva N, M. L. Model of visual localization. *Acta Neurobiol Exp*, **53** (1993), 377.

[10] Bosse, S., Maniry, D., Muller, K., Wiegand, T., and Samek, W. Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Transactions on Image Processing*, **27** (2018), 206.

[11] Campbell, F. W. and Green, D. G. Optical and retinal factors affecting visual resolution. *The Journal of physiology*, **181** (1965), 576.

[12] Campbell, F. W., Howell, E. R., and Johnstone, J. R. A comparison of threshold and suprathreshold appearance of gratings with components in the low and high spatial frequency range. *The Journal of Physiology*, **284** (1978), 193. doi:https://doi.org/10.1113/jphysiol.1978.sp012535.

[13] Campbell, F. W. and Kulikowski, J. J. Orientational selectivity of the human visual system. *The Journal of Physiology*, **187** (1966), 437. doi:https://doi.org/10.1113/jphysiol.1966.sp008101.

[14] Campbell, F. W. and Robson, J. G. Application of fourier analysis to the visibility of gratings. *The Journal of Physiology*, **197** (1968), 551. doi:https://doi.org/10.1113/jphysiol.1968.sp008574.

[15] Campisi, P. and Scarano, G. A multiresolution approach for texture synthesis using the circular harmonic functions. *IEEE Trans. on Image Processing*, **11** (2002), 37.

[16] Caron, A. L. and Jodoin, P.-M. Image multidistortion estimation. *IEEE Transactions on Image Processing*, **20** (2011), 3442. doi:10.1109/TIP.2011.2159233.

[17] Charman, W. N. and Heron, G. Microfluctuations in accommodation: an update on their characteristics and possible role. *Ophthalmic and Physiological Optics*, **35** (2015), 476. doi:10.1111/opo.12234.

[18] Cholewiak, S. A., Love, G. D., and Banks, M. S. Creating correct blur and its effect on accommodation. *Journal of Vision*, **18** (2018), 1. doi:10.1167/18.9.1.

[19] Chow, L. S. and Paramesran, R. Review of medical image quality assessment. *Biomedical Signal Processing and Control*, **27** (2016), 145. Available from: https://www.sciencedirect.com/science/article/pii/S1746809416300180, doi:https://doi.org/10.1016/j.bspc.2016.02.006.

[20] Conlon, E. G., Lovegrove, W. J., Chekaluk, E., and Pattison, P. E. Measuring visual discomfort. *Visual Cognition*, **6** (1999), 637. doi:10.1080/135062899394885.

[21] Cusani, R. and Jacovitti, G. A double tomographic approach to the estimation and classification of single objects. In *International Conference on Acoustics, Speech, and Signal Processing,*, pp. 1626–1629 vol.3 (1989). doi:10.1109/ICASSP.1989.266757.

[22] Daugman, J. G. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research*, **20** (1980), 847. doi:https://doi.org/10.1016/0042-6989(80)90065-6.

[23] Daugman, J. G. Six formal properties of two-dimensional anisotropic visual filters: structural principles and frequency/orientation selectivity. *IEEE Trans. Syst., Man, Cybern.*, **13** (1983), 882.

[24] Daugman, J. G. Quadrature-phase simple-cell pairs are appropriately described in complex analytic form. *J. Opt. Soc. Am. A*, **10** (1993), 375. doi:10.1364/JOSAA.10.000375.

[25] Dehnert, A., Bach, M., and Heinrich, S. P. Subjective visual acuity with simulated defocus. *Ophthalmic and Physiological Optics*, **31** (2011), 625. doi:https://doi.org/10.1111/j.1475-1313.2011.00857.x.

[26] Di Claudio, E. D., Giannitrapani, P., and Jacovitti, G. Predicting blur visual discomfort for natural scenes by the loss of positional information. *Vision Research*, **189** (2021), 33. Available from: https://www.sciencedirect.com/science/article/pii/S0042698921001826, doi:https://doi.org/10.1016/j.visres.2021.07.018.

[27] Di Claudio, E. D. and Jacovitti, G. A detail-based method for linear full reference image quality prediction. *IEEE Transactions on Image Processing*, **27** (2018), 179.

[28] Di Claudio, E. D., Jacovitti, G., and Laurenti, A. Maximum Likelihood orientation estimation of 1-D patterns in Laguerre-Gauss subspaces. *IEEE Trans. on Image Processing*, **19** (2010), 1113.

[29] Di Claudio, E. D., Jacovitti, G., and Laurenti, A. On the interconversion between Hermite and Laguerre local image expansions. *IEEE Transactions on Image Processing*, **20** (2011), 3553.

[30] Field, D. J. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, **4** (1987), 2379. doi:10.1364/JOSAA.4.002379.

[31] Field, D. J. and Brady, N. Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. *Vision Research*, **37** (1997), 3367 . doi:https://doi.org/10.1016/S0042-6989(97)00181-8.

[32] Goodman, J. W. Introduction to fourier optics. *Introduction to Fourier optics, 3rd ed., by JW Goodman. Englewood, CO: Roberts & Co. Publishers, 2005*, **1** (2005).

[33] Graham DJ, F. D., Chandler DM. Can the theory of whitening explain the center-surround properties of retinal ganglion cell receptive fields? *Vision Res*, **46** (2006), 2901.

[34] Graham N, N. J. Detection of grating patterns containing two spatial frequencies: a comparison of single-channel and multiple-channels models. *Vision Res*, **11** (1971), 251.

[35] Gu, K., Liu, M., Zhai, G., Yang, X., and Zhang, W. Quality assessment considering viewing distance and image resolution. *IEEE Transactions on Broadcasting*, **61** (2015), 520. doi:10.1109/TBC.2015.2459851.

[36] Hubel, D. H. *Eye, Brain, and Vision (Scientific American Library, No 22)*. W. H. Freeman (1995).

[37] Huber, P. *Robust Statistics*. John Wiley, New York (1981).

[38] Ingrid, L. M. . I. W. . F. M. . H. Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of Imaging Science and Technology*, **53** (2009), 30201.

[39] ITU. *Subjective video quality assessment methods for multimedia applications*. International Telecommunication Union (2008).

[40] ITU, T. S. S. O. Recommendation ITU-R BT.2022 general viewing conditions for subjective assessment of quality of sdtv and hdtv television pictures on flat panel displays (2012).

[41] ITU, T. S. S. O. ITU-T recommendation P.10 vocabulary for performance and quality of service, amendment 5 (2016).

[42] ITU, T. S. S. O. ITU-T series J cable networks and transmission of television, sound programme and other multimedia signals, measurement of the quality of service, part 3 (2016).

[43] Jacovitti, G. and Cusani, R. Image features extraction by radial tomographic analysis. In *Signal Processing V: Theories and Applications*, pp. 1011–1014 (1990).

[44] Jacovitti, G. and Cusani, R. Local angular spectrum analysis of 2d signals. In *Fifth ASSP Workshop on Spectrum Estimation and Modeling*, pp. 54–58 (1990).

[45] Jacovitti, G. and Neri, A. Bayesian removal of coding-block artifacts in the harmonic-angular filters features domain. In *Proc. SPIE, Nonlinear Image Processing VI*, vol. 2424, pp. 139–150. San Jose, CA, USA (1995).

[46] Jacovitti, G. and Neri, A. Multiresolution circular harmonic decomposition. *IEEE Trans. Signal Processing*, **48** (2000), 3242.

[47] Jacovitti, R. C. G. Contour coding of still pictures by harmonic angular filters. In *Proc. Int. Conf. Digital Signal Processing*, pp. 362–365 (1991).

[48] Jayaraman, D., Mittal, A., Moorthy, A. K., and Bovik, A. C. Objective quality assessment of multiply distorted images. In *2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pp. 1693–1697 (2012).

[49] Juricevic, I., Land, L., Wilkins, A., and Webster, M. A. Visual discomfort and natural image statistics. *Perception*, **39** (2010), 884.

[50] Kayargadde, V. and Martens, J. Perceptual characterization of images degraded by blur and noise: model. *J Opt Soc Am A Opt Image Sci Vis*, **13** (1996), 1178.

[51] Kuang, X., Poletti, M., Victor, J. D., and Rucci, M. Temporal encoding of spatial information during active visual fixation. *Curr Biol*, **22** (2012), 510. doi:10.1016/j.cub.2012.01.050.

[52] Landau, H. Sampling, data transmission, and the Nyquist rate. *Proceedings of the IEEE*, **55** (1967), 1701. doi:10.1109/PROC.1967.5962.

[53] Larson, E. C. and Chandler, D. M. Full-reference image quality assessment and the role of strategy: The most apparent distortion. online supplement. Available from: vision.okstate.edu/mad/.

[54] Larson, E. C. and Chandler, D. M. Most apparent distortion: full-reference image quality assessment and the role of strategy. *J. Electron. Imaging*, **9** (2010), 011006/1.

[55] Lin, W. and Jay Kuo, C. C. Perceptual visual quality metrics: A survey. *J. Vis. Comun. Image Represent.*, **22** (2011), 297?312. Available from: https://doi.org/10.1016/j.jvcir.2011.01.005, doi:10.1016/j.jvcir.2011.01.005.

[56] Liu, X., Pedersen, M., and Hardeberg, J. Y. Cid:iq – a new image quality database. In *Image and Signal Processing* (edited by A. Elmoataz, O. Lezoray, F. Nouboud, and D. Mammass), pp. 193–202. Springer International Publishing, Cham (2014). ISBN 978-3-319-07998-1.

[57] MacKenzie KJ, W. S., Hoffman DM. Accommodation to multiple-focal-plane displays: Implications for improving stereoscopic displays and for accommodation control. *Journ. Vision*, **10** (2010), 22.

[58] Mannos, J. and Sakrison, D. The effects of a visual fidelity criterion of the encoding of images. *IEEE Trans. on Information Theory*, **20** (1974), 525. doi:10.1109/TIT.1974.1055250.

[59] Marin-Franch, I., Aguila-Carrasco, A. J. D., Bernal-Molina, P., Esteve-Taboada, J. J., Lopez-Gil, N., Montes-Mico, R., and Kruger, P. B. There is more to accommodation of the eye than simply minimizing retinal blur. *Biomed. Opt. Express*, **8** (2017), 4717. doi:10.1364/BOE.8.004717.

[60] Marr, D. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. The MIT Press (2010).

[61] Martens, J. B. The Hermite transform - theory. *IEEE Trans. on Acoustic, Speech and Signal Processing*, **38** (1990), 1595.

[62] Martens, J. B. and Kayargadde, V. Image quality prediction in a multidimensional perceptual space. In *Proc. of the 1996 IEEE Int. Conf. on Image Processing*, vol. 1, pp. 877 – 880 (1996).

[63] Massey, R. and Refregier, A. Polar shapelets. *Monthly Notices of the Royal Astronomical Society*, **363** (2005), 197.

[64] Massey, R., Refregier, A., and Bacon, D. Shapelets ?multiple multipole? shear measurement. *Proceedings of the International Astronomical Union*, **2004** (2004), 31?36. doi:10.1017/S1743921305001766.

[65] Massey, R., Refregier, A., Conselice, C. J., and Bacon, D. J. Image simulation with shapelets. *Monthly Notices of the Royal Astronomical Society*, **348** (2004), 214. doi:10.1111/j.1365-2966.2004.07353.x.

[66] McIlhagga, W. H. and May, K. A. Optimal edge filters explain human blur detection. *Journal of Vision*, **12** (2012), 9. doi:10.1167/12.10.9.

[67] Metlapally, S., Tong, J. L., Tahir, H. J., and Schor, C. M. Potential role for microfluctuations as a temporal directional cue to accommodation. *Journal of Vision*, **16** (2016), 19. doi:10.1167/16.6.19.

[68] Murray, S. and Bex, P. Perceived blur in naturally contoured images depends on phase. *Frontiers in Psychology*, **1** (2010), 185. doi:10.3389/fpsyg.2010.00185.

[69] Neri, A. and Jacovitti, G. Maximum likelihood localization of 2-D patterns in the Gauss-Laguerre transform domain: theoretic framework and preliminary results. *IEEE Trans. on Image Processing*, **13** (2004), 72. doi:10.1109/TIP.2003.818021.

[70] O'Hare, L. and Hibbard, P. B. Spatial frequency and visual discomfort. *Vision Research*, **51** (2011), 1767. doi:https://doi.org/10.1016/j.visres.2011.06.002.

[71] O'Hare, L. and Hibbard, P. B. Visual discomfort and blur. *Journal of Vision*, **13** (2013), 7. doi:10.1167/13.5.7.

[72] Panci, G., Campisi, P., Colonnese, S., and Scarano, G. Multichannel blind image deconvolution using the bussgang algorithm: Spatial and multiresolution approaches. *IEEE Trans. on Image Processing*, **12** (2003), 1324.

[73] Parisi, R., Claudio, E. D. D., Orlandi, G., and Rao, B. A generalized learning paradigm exploiting the structure of feedforward neural networks. *IEEE Trans. on Neural Networks*, **7** (1996), 1450.

[74] Perona, P. Deformable kernels for early vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17** (1995), 488.

[75] Ponomarenko, N., et al. Image database TID2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication*, **30** (2015), 57 . doi:https://doi.org/10.1016/j.image.2014.10.009.

[76] Redies C, D. J., Hasenstein J. Fractal-like image statistics in visual art: similarity to natural scenes. *Spat Vis*, **21** (2007), 137. doi:10.1163/156856807782753921.

[77] Refregier, A. Shapelets ? I. A method for image analysis. *Monthly Notices of the Royal Astronomical Society*, **338** (2003), 35. doi:10.1046/j.1365-8711.2003.05901.x.

[78] Reisert, M. and Burkhardt, H. Complex derivative filters. *IEEE Trans. on Image Processing*, **17** (2008), 2265.

[79] Schütt, H. H. and Wichmann, F. A. An image-computable psychophysical spatial vision model. *Journal of Vision*, **17** (2017), 12. doi:10.1167/17.12.12.

[80] Sheikh, H. R. and Bovik, A. C. Image information and visual quality. *IEEE Trans. on Image Processing*, **15** (2006), 430.

[81] Sheikh, H. R., Sabir, M. F., and Bovik, A. C. A statistical evaluation of recent Full Reference image quality assessment algorithms. *IEEE Trans. on Image Processing*, **15** (2006), 3441. Available from: live.ece.utexas.edu/research/quality/subjective.htm.

[82] Simoncelli, E. P., Freeman, W. T., Adelson, E. H., and Heeger, D. J. Shiftable multi-scale transforms. *IEEE Trans. Information Theory*, **38** (1992), 587. Special Issue on Wavelets.

[83] Simoncelli, E. P. and Olshausen, B. A. Natural image statistics and neural representation. *Annual Review of Neuroscience*, **24** (2001), 1193. doi:10.1146/annurev.neuro.24.1.1193.

[84] Solomon, J. A. The history of dipper functions. *Attention, Perception, Psychophysics*, **71** (2009), 435.

[85] Sprague, W. W., Cooper, E. A., Reissier, S., Yellapragada, B., and Banks, M. S. The natural statistics of blur. *J Vis*, **16** (2016), 23.

[86] Strasburger, H., Bach, M., and Heinrich, S. P. Blur unblurred - a mini tutorial. *i-Perception*, **9** (2018), 2041669518765850. PMID: 29770182. doi:10.1177/2041669518765850.

[87] Taylor, B. P., CP. On the number of perceivable blur levels in naturalistic images. *Vision Res*, **115** (2015), 142.

[88] Torralba, A. and Oliva, A. Statistics of natural image categories. *Network: Computation in Neural Systems*, **14** (2003), 391. PMID: 12938764. doi:10.1088/0954-898X_14_3_302.

[89] Tsukida, K. and Gupta, M. R. How to analyze paired comparison data. *Technical Report UWEETR-2011-0004*, (2011).

[90] Van Der Schaaf, A. and Van Hateren, J. H. Modelling the power spectra of natural images: statistics and information. *Vision Res*, **36** (1996), 2759. doi:10.1016/0042-6989(96)00002-8.

[91] Van Trees, H. L. *Detection, Estimation, and Modulation Theory: Radar-Sonar Signal Processing and Gaussian Signals in Noise*. Krieger Publishing Co., Inc., USA (1992). ISBN 0894647482.

[92] Victor, J. D. and Knight, B. W. *Simultaneously band and space limited functions in two dimensions, and receptive fields of visual neurons*, chap. -, pp. 375–420. Springer Applied Mathematical Sciences Series. Springer-Verlag, New York, E. Kaplan, J. Marsden and K. R. Sreenivasan edn. (2003).

[93] VQEG. Final report from the video quality experts group on the validation of objective models of video quality assessment. Tech. rep., Video Quality Experts Group, http://www.vqeg.org (2000).

[94] VQEG. Final report from the video quality experts group on the validation of objective models of video quality assessment, phase II. Tech. rep., Video Quality Experts Group, http://www.vqeg.org (2003).

[95] Wang, Z. IW-SSIM: Information content weighted structural similarity index for image quality assessment (2011). Available from: `https://ece.uwaterloo.ca/~z70wang/research/iwssim/`.

[96] Wang, Z. and Bovik, A. C. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE Signal Processing Magazine*, **26** (2009), 98. `doi:10.1109/MSP.2008.930649`.

[97] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. on Image Processing*, **13** (2004), 600.

[98] Wang, Z., Simoncelli, E., and Bovik, A. Multi-scale structural similarity for image quality assessment. In *Proc. of the 37th IEEE Asilomar Conference on Signal, Systems and Computer*, vol. 2, pp. 1398–1402. IEEE Comp. Soc. (2003).

[99] Wang, Z. and Simoncelli, E. P. Local phase coherence and the perception of blur. In *Advances in Neural Information Processing Systems 16* (edited by S. Thrun, L. K. Saul, and B. Scholkopf), pp. 1435–1442. MIT Press (2004).

[100] Watson, A. B. A formula for the mean human optical modulation transfer function as a function of pupil size. *J Vis.*, **13** (2013), 18.

[101] Watson, A. B. Computing human optical point spread functions. *Journal of Vision*, **15** (2015), 26. `doi:10.1167/15.2.26`.

[102] Watson, A. B. and Ahumada, A. J. Blur clarified: A review and synthesis of blur discrimination. *Journal of Vision*, **11** (2011), 10. `doi:10.1167/11.5.10`.

[103] Watson, A. B. and Ahumada, J., Albert J. A standard model for foveal detection of spatial contrast. *Journal of Vision*, **5** (2005), 6. `doi:10.1167/5.9.6`.

[104] Watson, A. B. and Solomon, J. A. Model of visual contrast gain control and pattern masking. *J. Opt. Soc. Am. A*, **14** (1997), 2379. doi:10.1364/JOSAA.14.002379.

[105] Webster, G. M. W. S., MA. Neural adjustments to image blur. *Nat Neurosci.*, **5** (2002), 839. doi:10.1038/nn906.

[106] Wilkins, A. A physiological basis for visual discomfort: Application in lighting design. *Lighting Research & Technology*, **48** (2016), 44. doi:10.1177/1477153515612526.

[107] Williams, D. R. Visibility of interference fringes near the resolution limit. *J. Opt. Soc. Am. A*, **2** (1985), 1087. doi:10.1364/JOSAA.2.001087.

[108] Xue, W., Zhang, L., Mou, X., and Bovik, A. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Trans. on Image Processing*, **23** (2014), 684. doi:10.1109/TIP.2013.2293423.

[109] Zhang, L., Zhang, L., Mou, X., and Zhang, D. FSIM: A feature similarity index for image quality assessment (13). Available from: http://sse.tongji.edu.cn/linzhang/IQA/FSIM/FSIM.htm.

[110] Zhang, X., Ye, M., Bradley, A., and Thibos, L. Apodization by the stiles–crawford effect moderates the visual impact of retinal image defocus. *J. Opt. Soc. Am. A*, **16** (1999), 812. doi:10.1364/JOSAA.16.000812.

[111] Zhuo, S. and Sim, T. Defocus map estimation from a single image. *Pattern Recognition*, **44** (2011), 1852. Computer Analysis of Images and Patterns. doi:https://doi.org/10.1016/j.patcog.2011.03.009.