

Measuring and Improving Image Resolution by Adaptation of the Reciprocal Cell

Andrés Almansa

Sylvain Durand

Bernard Rougé

July 5, 2002

Abstract

Traditionally, discrete images are assumed to be sampled on a square grid and from a special kind of band-limited continuous image, namely one whose Fourier spectrum is contained within the rectangular “reciprocal cell” associated with the sampling grid. With such a simplistic model, resolution is just given by the distance between sample points.

Whereas this model matches to some extent the characteristics of traditional acquisition systems, it doesn’t explain aliasing problems, and it is no longer valid for certain modern ones, where the sensors may show a heavily anisotropic transfer function, and may be located on a non-square (in most cases hexagonal) grid.

In this work we first summarize the generalizations of Fourier theory and of Shannon’s sampling theorem, that are needed for such acquisition devices. Then we explore its consequences: *(i)* A new way of measuring the effective resolution of an image acquisition system; *(ii)* A more accurate way of restoring the original image which is represented by the samples. We show on a series of synthetic and real images, how the proposed methods make a better use of the information present in the samples, since they may drastically reduce the amount of aliasing with respect to traditional methods. Finally we show how in combination with Total Variation minimization, the proposed methods can be used to extrapolate the Fourier spectrum in a reasonable manner, visually increasing image resolution.

Keywords: *Image resolution, acquisition, sampling, restoration, aliasing, Shannon.*

Contents

1	Introduction	3
2	Fourier Analysis in non-square grids	7
2.1	Sampling grids and tilings of the plane	7
2.2	Shannon's sampling theorem.	8
2.3	Choice of the reciprocal cell.	11
2.4	Critical Sampling.	12
3	Modeling of image acquisition systems	16
4	Effective Resolution	22
4.1	Nominal Resolution	22
4.2	Effective resolution	23
4.3	Information density.	25
5	Optimal Reciprocal Cell	28
6	Examples and Applications	32
7	Discussion and future work	42
8	Acknowledgements	42
A	Proof of Shannon's sampling theorem	43
A.1	Prerequisites	43
A.2	Shannon's sampling theorem	45
A.3	Suitability of the hypothesis in real systems	49
B	Proof of theorem 2	49
C	Sufficient conditions for admissibility	51

1 Introduction

When an analog image is converted into digital form it is sampled on a regular grid. If the image is not band-limited at the right frequency, this sampling operation produces a replication of higher frequency components into lower frequency components of the spectrum, a phenomenon known as “aliasing”. Aliasing artifacts in digital images and sampling grids other than the traditional square grid have received little attention in the mainstream literature on image processing.

Some textbooks on image processing like (Bracewell, 1995; Rosenfeld and Kak, 1982) do introduce the formalisms needed for sampling in general lattices, they describe the corresponding aliasing problems and the corresponding generalization of Shannon’s sampling theorem (in a similar fashion as we present it here). Most commonly, however, sampling and aliasing are only described in the case of rectangular grids and reciprocal cells (Gonzalez and Woods, 1992; Jähne, 1995; Russ, 1995) or at most in the case of quincunx grids and interlaced television signals (Jain, 1995). In any case in all these textbooks the treatment of non-orthogonal sampling grids and aliasing is limited to a description of the effects and the tools to analyze them, but rarely gives any practical means of measuring or reducing the amount of aliasing for a given sampling system, maybe with the exception of (Pratt, 1978) which gives different measures of resolution, aliasing and interpolation error, as well as a Wiener filtering technique which takes into account the aliasing noise. Otherwise, the degradation model assumed in the great majority of the image restoration literature, includes convolution with a band-limited kernel plus noise, and thus ignores the effects of aliasing. Discussions on how to limit the aliasing error are far more common in the computer graphics literature, where one has full control of the sampling system (most notably to its sampling rate and transfer function), thus leading to approaches that are in most cases not applicable in digital imaging, where both parameters are dictated by physical constraints of the imaging optics and the electronics of the sensors.

Concerning sampling on non-orthogonal grids, most textbooks dedicate a section to compare the properties of hexagonal and square grids, but then present the rest of the results only for the square grids. More particularly, concerning Fourier analysis in hexagonal grids, an important part of the literature started by (Mersereau, 1979) and followed by many others (Ehrhardt, 1993; Mersereau and Speake, 1983; Zapata and Ritter, 2000), deals with specialized FFT algorithms for the hexagonal sampling grids. Whereas this is necessary if we want to keep a rectangular periodization pattern, if we consider the image to be periodic along the same axes as the sampling grid, then we can still use the usual FFT (Morel and Ladjal, 1998) as we do in square grids. We believe that the little emphasis that has been given to this property as well as the lack of specialized literature discouraged the development of hexagonal image processing despite its advantages.

In fact, most authors agree on several advantages of the hexagonal grid over the more traditional rectangular grid. We summarize the most important ones here :

- *Sampling density:* under isotropic conditions a hexagonal grid requires 13.4% less samples than a square grid to represent the same amount of information.
- *Discrete topology:* we only need a single natural connectivity concept (6 neighbours), which is consistent with Jordan’s curve theorem, whereas in the square grid consistency with the continuous case obliges us to switch between 4-connectivity for the foreground and 8-connectivity for the background or conversely. This greatly simplifies many morphological operations.
- *Isotropic image analysis:* It has been shown that in the hexagonal grid we can construct more isotropic wavelet families with much better orientation selectivity (Cohen and Schlenker, 1993; Daubechies, Guskov, Schröder, and Sweldens, 1999; Jaffard, 1989; Kovacevic and Sweldens, 1998; Laine and Schuler, 1994; Meyer, 1990; Schuler and Laine, 1997; Simoncelli and Adelson, 1990a,b), more accurate distance transforms (Borgefors, 1989), isotropic 3-separable filters, etc.

For a more thorough discussion on the properties of hexagonal grids see (Almansa, Morel, and Rougé, 1999; Staunton, 1999).

Coming back to aliasing, despite the little importance given to the subject in the image processing literature, it has been recognized by several authors as an important problem, which significantly limits the quality of image data (Biberman, 1973). More recently, the importance of aliasing artifacts present in most DVD movies was analyzed (Moisan and Morel, 2001), and the physical constraints of any CCD array leading to a significant amount of aliasing were described in (Rougé, 1997). At the same time the latter work as well as (Latry and Rougé; Minho and Pokotilo, 1997; Rougé, 1998) propose a physical sampling device that allows to avoid the limitations of CCD arrays and reduce aliasing. Such sampling devices involve any kind of regular sampling grid (rectangular, hexagonal, or in general, any grid generated by two independent vectors on the plane), and may show heavily anisotropic transfer functions.

The main purpose of this work is to stress once more the importance of aliasing in imaging systems, provide ways of measuring to what an extent it may degrade the image quality, and provide practical methods to reduce the amount of aliasing of any sampling system, provided we know its three main characteristics: the geometry of the sampling grid, the transfer function, and the noise. The results can be used both by the designer of imaging systems who wants to choose between different possibilities the one which provides the best image-quality/cost ratio, as well as by the image processing engineer who wants to obtain the best possible reconstructed image from a given system.

We start by giving in section 2.2 a slightly generalized form of Shannon's sampling theorem, since:

1. It applies to any regular sampling grid, not necessarily rectangular.
2. Instead of the classical rectangular (Voronoi) reciprocal cell it allows more general shapes of this cell which is used both to express the band-limiting condition and the sinc-interpolation formula.

This form has long been known by the mathematicians and similar forms can be found in (Bony, 2001; Gasquet and Witomski, 1995; Morel and Ladjal, 1998; Rosenfeld and Kak, 1982), but its flexibility has not been fully exploited by the image processing community yet.

Next, we review in sections 2.3 and 2.4 the common notions of resolution, critical sampling and sampling efficiency, and extend them to the new formulation. We also present some simple examples on how to adapt the reciprocal cell to the image contents in the ideal case of strictly band-limited images and no noise.

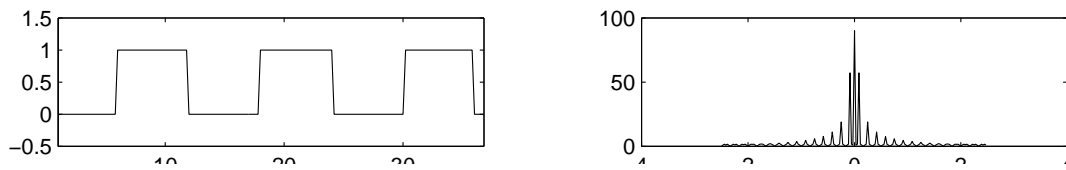
In section 3 we describe the main characteristics of some of the sampling systems mentioned above: both the classical CCD arrays, and the more anisotropic systems described in (Latry and Rougé; Minho and Pokotilo, 1997; Rougé, 1998). Then in sections 4 and 5 we extend to this more realistic case the ideas of resolution and optimal reciprocal cell that were introduced in sections 2.3 and 2.4 in the ideal band-limited case.

Finally in section 6 we present the results of our numerical experiments, commenting the quantitative comparison of the different systems according to the resolution-based measures that we introduced, and we propose an application of the optimal reciprocal cell to image restoration and spectral extrapolation, which considers not only convolution and noise, but also aliasing within the degradation model.

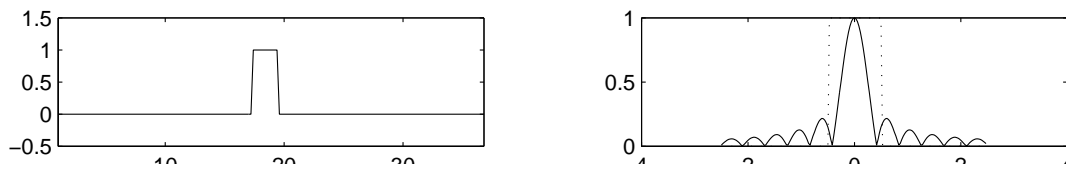
Before we start with the main presentation let us present a simple example that illustrates how important it may be to consider aliasing during restoration. We want to sample a square wave signal f with period $A = 12$, with a sampling system consisting of an aperture of $T \approx 2.38$, and a sampling rate of 1, which adds a white noise with standard deviation 0.01 (*i.e.* 1% of the wave's amplitude). Summarizing the sampled signal will be

$$g(k) = (f * h)(k) + n(k) \text{ for all } k \in \mathbb{Z} \quad (1)$$

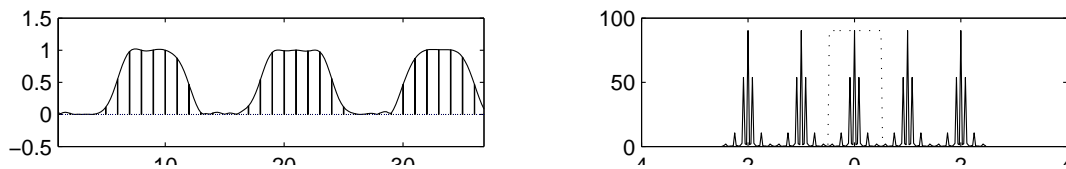
where the point-spread function is $h(t) = \mathbb{1}_{[-\frac{T}{2}, \frac{T}{2}]}(t)$. Figure 1 shows the different stages in the sampling of g and its restoration with a Wiener filter. Observe that the result is a catastrophic for the



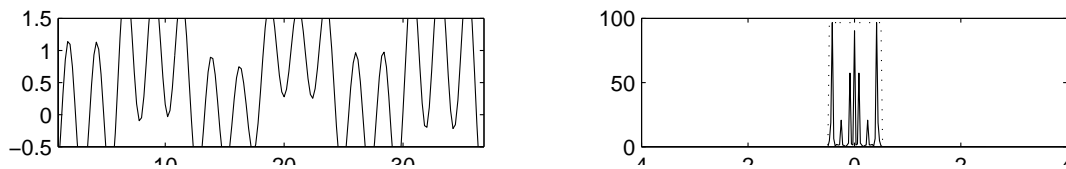
(a) original signal f and its Fourier transform



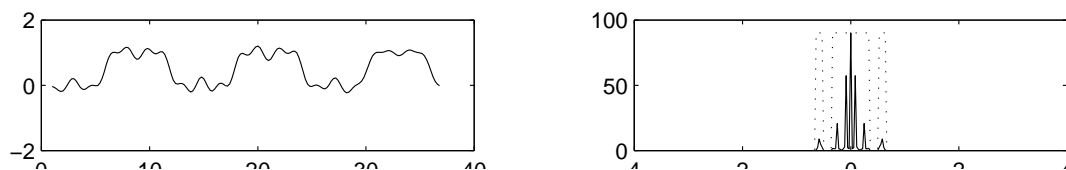
(b) pre-sampling filter h and transfer function \hat{h}



(c) sampled signal $g = f * h + n$ and sinc-interpolation



(d) Wiener-filter restoration f_1 with voronoi reciprocal cell



(e) Wiener-filter restoration f_2 with optimal reciprocal cell

Figure 1: Example of Wiener filter restoration with aliasing. In (d) the spectrum of the restored image is $\hat{f}_1(\xi) = \hat{g}(\xi)H(\xi)$ for $|\xi|$ smaller than the Nyquist frequency and zero otherwise. Here $H = \frac{H^*}{|H|^2 + \sigma^2}$ is the Wiener deconvolution filter and $H = \hat{h}$ is the modulation transfer function (MTF). Observe that this produces a bad result because we are deconvolving an aliased coefficient near the zero-crossing of H . In (e) we solve this problem by deconvolving $\hat{f}_2(\xi) = \hat{g}(\xi)H(\xi)$ for ξ within an optimal reciprocal cell adapted to the MTF, and $\hat{f}_2(\xi) = 0$ for ξ outside this reciprocal cell.

following reason, that is more clearly explained in figure 2: The MTF (the modulation transfer function *i.e.* the Fourier transform $H = \hat{H}$ of the blurring kernel h) has a zero crossing at a frequency $\xi_0 \approx 0.4$ just before the Nyquist frequency. If there was no aliasing, the corresponding Fourier coefficient of the sampled image $\hat{g}(\xi_0)$ should be nearly zero, plus noise. But since the MTF has a second lobe beyond the Nyquist frequency, some high frequency coefficients (see for instance the fifth peak at frequency $\xi_0 - 1 \approx -0.6$) are aliased near frequency ξ_0 ¹. Thus $\hat{g}(\xi_0)$ is far larger than the expected signal plus noise level, and when we try to deconvolve it by $\frac{H^*}{|H|^2 + \sigma^2}(\xi_0) \approx \frac{0.01}{0.01^2 + 0.01^2} = 50$ it explodes.

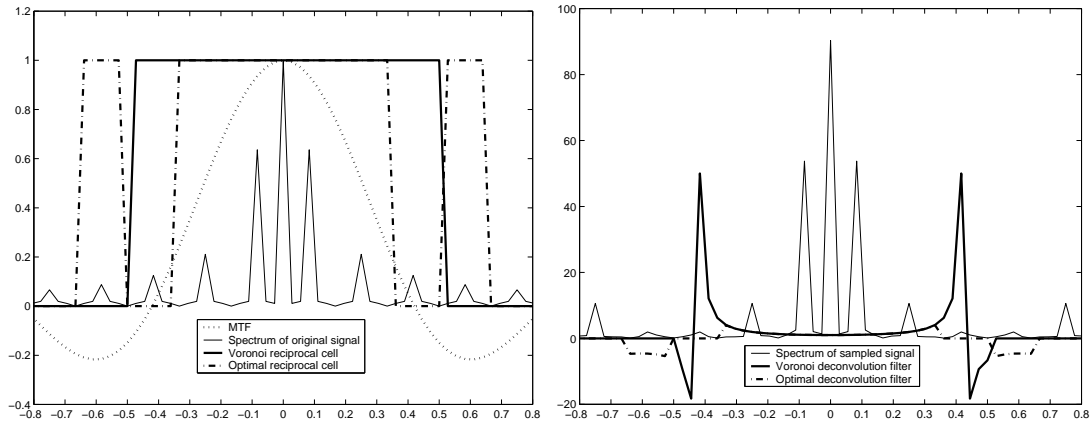


Figure 2: Example of Wiener filter restoration with aliasing (continued). Left: Closeup of the system’s transfer function \hat{h} , the spectrum \hat{f} of the original signal, the voronoi reciprocal cell (consisting of all frequencies smaller than the Nyquist frequency associated to the sampling rate), and the optimal reciprocal cell, adapted to the MTF. To the right we show the (periodic) spectrum of the sampled signal with the two Wiener deconvolution filters. Observe how the optimal reciprocal cell (and the corresponding deconvolution filter) avoids a region around the zero-crossing of \hat{h} at $\xi_0 \approx 0.4$ and substitutes it by the region around its alias at $\xi_0 - 1$, where the magnitude of \hat{h} is much larger.

This suggests that we should avoid deconvolving around ξ_0 , and do it rather around $\xi_0 - 1$ where the real information comes from. Put another way, we defined an alternative reciprocal cell, which is still compatible with Shannon’s sampling theorem, but is better adapted to our instrument’s transfer function. In the following sections we shall develop a general technique for choosing this reciprocal cell in an optimal way.

With the adapted Wiener deconvolution we obtain a reconstruction error $\frac{\|f_1 - f\|}{\|f\|}$ of 20%, which is only slightly better than the original 24% error in the sampled image g , but it is much better than the 146% reconstruction error of the original Wiener filter. The ringing artifacts in the result are mainly due to the Fourier coefficients that were lost between the two lobes of the optimal reciprocal cell. To avoid it we need to extrapolate the spectrum between these two lobes. In section 6 we propose to do so by minimizing the total variation.

¹ The magnitude of the non-aliased part of $\hat{g}(\xi_0)$ is $|\hat{f}(\xi_0)\hat{h}(\xi_0)| + \sigma = 0.84 * 0.01 + 0.01 = 0.018$, but the actual aliased coefficient is $\hat{g}(\xi_0) = 0.13$, seven times larger. The largest contribution to $\hat{g}(\xi_0)$ comes from frequency $\xi_0 - 1 \approx -0.6$, *i.e.* $|\hat{f}(\xi_0 - 1)\hat{h}(\xi_0 - 1)| = 0.58 * 0.20 = 0.12$

2 Review of Fourier Analysis in non-square grids

In this section we recall the mathematical basics of sampling theory in the ideal case of infinite spatial support and compact spectral support (band limited). In the following section 3 we will discuss how this modeling applies to real acquisition systems. The main elements of this review are contained in (Bony, 2001; Gasquet and Witomski, 1995; Morel and Ladjal, 1998), but here we give slightly different formulations which are better suited for defining our image resolution measures.

Section 2.1 defines the geometrical bases of sampling in non-orthogonal grids and section 2.2 provides a formal proof of Shannon's sampling theorem on such grids. A more rigorous proof, as well as different formulations depending on the degree of regularity of the image are given in appendix A.

Finally sections 2.3 and 2.4 review some common concepts related to image resolution in the ideal case of compact spectral support and no noise. These concepts will be generalized in sections 4 and 5 to the more realistic systems described in section 3.

2.1 Sampling grids and tilings of the plane

We will consider our discrete images to be defined on a *regular sampling grid*

$$\Gamma := \{n_1 e_1 + n_2 e_2 : n_1, n_2 \in \mathbb{Z}\} = \mathbb{Z}e_1 + \mathbb{Z}e_2 \quad (2)$$

where $\{e_1, e_2\}$ is a basis of \mathbb{R}^2 . To each sampling grid we shall associate a *dual grid*

$$\Gamma^* := \{n_1 e_1^* + n_2 e_2^* : n_1, n_2 \in \mathbb{Z}\} = \mathbb{Z}e_1^* + \mathbb{Z}e_2^* \quad (3)$$

where $\langle e_i^*, e_j \rangle = 2\pi \delta_{ij}$

This will be useful when we define the Fourier transform of a discrete image.

In the sequel we do not impose any further condition on the generating vectors e_1 and e_2 . In practice, however, only two such grids are used, namely:

$$\Gamma_4 = \mathbb{Z} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \mathbb{Z} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (\text{square grid})$$

$$\Gamma_6 = \mathbb{Z} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \mathbb{Z} \begin{pmatrix} 1/2 \\ \sqrt{3}/2 \end{pmatrix} \quad (\text{hexagonal grid}).$$

The reason is that these are the only ones which generate tilings of the plane by regular polygons, and show the highest number of symmetries. The corresponding dual grids are

$$\Gamma_4^* = 2\pi \left(\mathbb{Z} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \mathbb{Z} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) \quad (\text{dual square grid})$$

$$\Gamma_6^* = 2\pi \frac{2}{\sqrt{3}} \left(\mathbb{Z} \begin{pmatrix} \sqrt{3}/2 \\ -1/2 \end{pmatrix} + \mathbb{Z} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) \quad (\text{dual hexagonal grid})$$

as shown in figure 3.

Usually, sensor arrays are composed of a set of sensors all with the same shape and characteristics, with their centers lying on a regular grid, and covering the image plane without superposition. More precisely, given a set $D \subset \mathbb{R}^2$ and a grid Γ we shall say that:

- (Γ, D) is a *covering*², if $\bigcup_{\gamma \in \Gamma} (D + \gamma) = \mathbb{R}^2$
- (Γ, D) is a *packing*², if $\forall \gamma \in \Gamma, \gamma \neq 0 \implies D \cap (D + \gamma) = \emptyset$

² Here, set equality is to be understood in the weak sense or "almost everywhere", namely $A = B$ iff both sets $A \setminus B$ and $B \setminus A$ have zero measure.

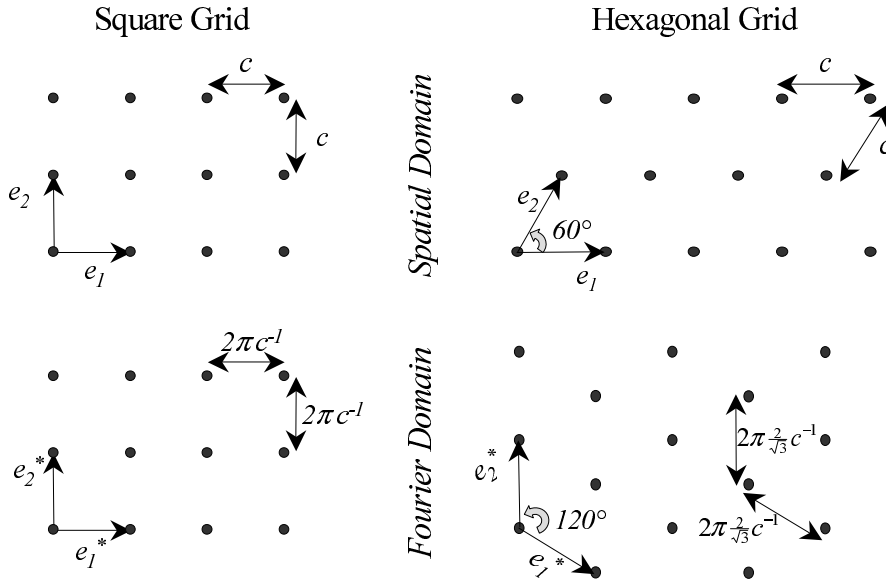


Figure 3: The most common sampling grids and their corresponding duals. The square and hexagonal grids can be singled out as the only two possibilities from a series of invariance properties (Almansa et al., 1999). As explained in section 2.2, when an image is sampled on a grid Γ , its Fourier transform is periodical with respect to its dual grid Γ^* , whose generating vectors are biorthogonal to those of Γ .

- (Γ, D) is a *tiling*, if it is both a packing and a covering.

Here $D + \gamma$ represents the translated set $\{x + \gamma : x \in D\}$. In addition we slightly relaxed the usual covering and packing conditions by considering the closure \overline{D} and the interior $\overset{\circ}{D}$ of the set D . We do so for clarity, to simplify the definitions of our sets D (otherwise, we need a convention to decide whether to include or not in D a point in the border ∂D).

In addition we shall say that D is a *tile* of the grid Γ if (Γ, D) is a tiling. Similarly, we say that D is a *sub-tile* (resp. a *super-tile*) if (Γ, D) is a packing (resp. a covering). These three concepts, when applied to the dual grid Γ^* will also be useful to choose suitable spectral domains.

Figure 4 shows the only two tilings of the plane where the tile is a regular polygon. The triangular tile can only cover the plane if –in addition to the integer translations in Γ – we also allow an horizontal symmetry. In the sequel we shall mostly concentrate on the square and hexagonal sampling grids, but will occasionally use tiles for these grids, which are not regular polygons. Figure 5 shows an example, which illustrates how one can construct such tilings, without changing the sampling grid Γ .

2.2 Shannon's sampling theorem.

To each sampling grid we shall associate a *Dirac comb*

$$\Delta_{\Gamma} := \sum_{\gamma \in \Gamma} \delta_{\gamma} \quad (4)$$

composed of delta-functions, such that $\langle \delta_{\gamma}, f \rangle = \int f(x) \delta_{\gamma}(dx) = f(\gamma)$. Sampling on a grid Γ can then be expressed as simply multiplying by Δ_{Γ} .

A sampled image will be modeled as a Γ -discrete distribution, i.e. a tempered distribution g for which there exists a continuous function $f \in \mathcal{O}_M(\mathbb{R}^2)$ such that $g = \Delta_{\Gamma} \cdot f$. Similarly, we will say

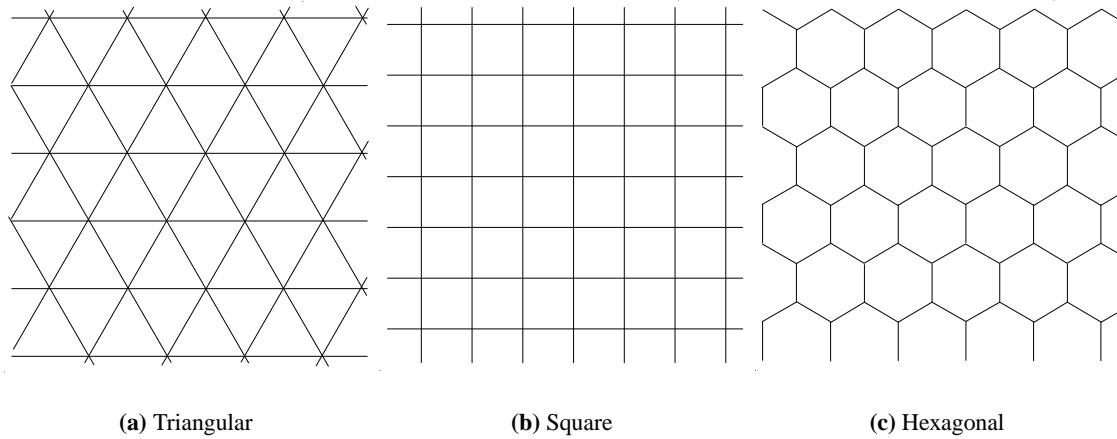


Figure 4: *The only three tilings of the plane by regular polygons.* Observe that the hexagonal tiling possesses the largest symmetry group. In addition only the square and hexagonal ones are tilings in the sense we just defined. The triangular tile requires, in addition to translations, a symmetry with respect to an horizontal axis in order to cover the whole plane. For this and other reasons the triangular tiling is not very useful in image processing.

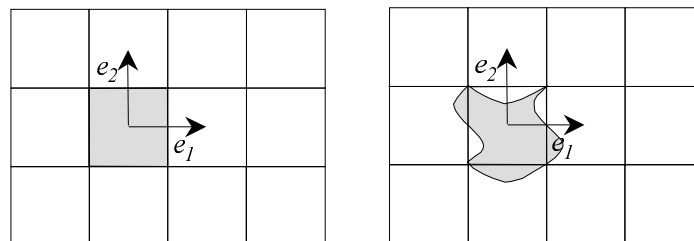


Figure 5: The shape of a tile can be modified, and still form a tiling with the same grid Γ .

that a tempered distribution g is Γ -periodic if it can be expressed as a convolution $g = \Delta_\Gamma * f$ for some tempered distribution with fast decay $f \in \mathcal{O}'_C(\mathbb{R}^2)$.

Observe that this definition is equivalent to the usual definition of periodicity for functions, namely that $g(x + \gamma) = g(x)$ for all $x \in \mathbb{R}^2$, $\gamma \in \Gamma$.

In fact if $g = \Delta_\Gamma * f$, then $g(x + \gamma_0) = \sum_{\gamma \in \Gamma} f(x + \gamma + \gamma_0) = \sum_{\gamma \in \Gamma} f(x + \gamma)$, since the mapping $\gamma \mapsto (\gamma + \gamma_0)$ is bijective in Γ . Conversely if $g(x + \gamma) = g(x)$ then $g = \Delta_\Gamma * (\chi \cdot g)$ where $\chi \in C_0^\infty(\mathbb{R}^2)$ is a partition of unity (i.e. such that $\sum_{\gamma \in \Gamma} \chi(x + \gamma) \equiv 1$).

It is easy to show that such a partition of unity exists. It suffices to start from a non-negative function $\varphi \in C_0^\infty$ which is strictly positive on a tile D of Γ , and then define $\chi(x) = \frac{\varphi(x)}{\sum_{\gamma \in \Gamma} \varphi(x + \gamma)}$.

We recall the continuous Fourier Transform of $f \in L^1$ and its inverse

$$\hat{f}(\xi) = \mathcal{F}(f)(\xi) := \int_{\mathbb{R}^2} e^{-i\langle x, \xi \rangle} f(x) dx \quad (5)$$

$$(2\pi)^2 f(x) = \overline{\mathcal{F}}(\hat{f})(x) := \int_{\mathbb{R}^2} e^{i\langle x, \xi \rangle} \hat{f}(\xi) d\xi \quad (6)$$

as well as some of its properties that we state without proof:

- $\overline{\mathcal{F}}(\mathcal{F}(f)) = (2\pi)^2 f$
- $\mathcal{F}(f * g) = \mathcal{F}(f) \cdot \mathcal{F}(g)$
- $\mathcal{F}(f \cdot g) = \frac{1}{(2\pi)^2} \mathcal{F}(f) * \mathcal{F}(g)$
- $\mathcal{F}(\Delta_\Gamma) = |D^*| \Delta_{\Gamma^*}$ where D^* is a tile of Γ^* .

These properties are not valid in general. The precise conditions, as well as some extensions of the Fourier transform and its properties to more general classes of functions and distributions is given in appendix A.1. Here we use these properties to give a formal derivation of the sampling theorem. For a rigorous proof and some variants of its formulation please refer to appendix A.2.

From the last two properties we deduce that the Fourier transform of a Γ -discrete image g is Γ^* -periodic:

$$\hat{g} = \mathcal{F}(\Delta_\Gamma \cdot f) = \frac{|D^*|}{(2\pi)^2} \Delta_{\Gamma^*} * \hat{f}. \quad (7)$$

Thus \hat{g} is completely determined if we know its value on a tile D^* of the dual grid Γ^* . (In fact, since D^* is a tile, for any $\xi \in \mathbb{R}^2$ we can find $\gamma^* \in \Gamma^*$ such that $(\xi + \gamma^*) \in D^*$, and therefore by periodicity $g(\xi) = g(\xi + \gamma^*) = g|_{D^*}(\xi + \gamma^*)$). If in addition $\hat{g}|_{D^*} = \hat{f}|_{D^*}$ then g is completely determined by $\hat{f}|_{D^*}$.

A slightly stronger condition than $\hat{g}|_{D^*} = \hat{f}|_{D^*}$ ensures the converse, i.e. that the continuous image f is completely determined by the discrete image $g = \Delta_\Gamma \cdot f$. It suffices to require that f be *band-limited*, i.e. that $\text{supp}(\hat{f}) \subseteq D^*$ with D^* a sub-tile of Γ^* .

In that case $\mathbb{1}_{D^*} \cdot (\Delta_{\Gamma^*} * \hat{f}) = \mathbb{1}_{D^*} \cdot \hat{f} = \hat{f}$.³ This allows us to recover \hat{f} from \hat{g} as follows:

$$\frac{(2\pi)^2}{|D^*|} \mathbb{1}_{D^*} \cdot \hat{g} = \frac{(2\pi)^2}{|D^*|} \mathbb{1}_{D^*} \cdot \left(\frac{|D^*|}{(2\pi)^2} \Delta_{\Gamma^*} * \hat{f} \right) = \mathbb{1}_{D^*} \cdot (\Delta_{\Gamma^*} * \hat{f}) = \hat{f}. \quad (8)$$

Finally we apply the Fourier inversion formula to obtain:

$$\frac{1}{|D^*|} \overline{\mathcal{F}}(\mathbb{1}_{D^*}) * g = f \quad (9)$$

This formally shows a generalized form of Shannon's sampling theorem:

³ Since D^* is a sub-tile of Γ^* , if $\xi \in D^*$ then $\xi + \gamma^* \notin D^*$ for $\gamma^* \in \Gamma^* \setminus \{0\}$. Now, since $\text{supp}(\hat{f}) \subseteq D^*$, only the term in $\gamma^* = 0$ is non-zero in $\sum_{\gamma^* \in \Gamma^*} \hat{f}(\xi + \gamma^*)$. This shows that $(\Delta_{\Gamma^*} * \hat{f})|_{D^*} = \hat{f}|_{D^*}$. Consequently $\mathbb{1}_{D^*} \cdot (\Delta_{\Gamma^*} * \hat{f}) = \mathbb{1}_{D^*} \cdot \hat{f}$. But again, since $\text{supp}(\hat{f}) \subseteq D^*$ we have $\mathbb{1}_{D^*} \cdot \hat{f} = \hat{f}$.

Theorem 1 (Shannon-Whittaker) Given a function $f \in L^2(\mathbb{R}^d)$, a d -dimensional sampling grid Γ and a compact reciprocal cell $D^* \subset \mathbb{R}^d$. If the following conditions are met

(S1) the reciprocal cell D^* is a sub-tile of the dual grid Γ^*

(S2) $\text{supp}(\hat{f}) \subseteq D^*$ (i.e. f is band-limited),

then f can be completely recovered from its samples in Γ , i.e. from

$$g = \Delta_{\Gamma} \cdot f. \quad (10)$$

In fact, $g \in l^2(\Gamma)$, and the reconstruction is given by the following convolution

$$f(x) = \sum_{\gamma \in \Gamma} g_{\gamma} s(x - \gamma) \quad (11)$$

of the sampled image g with the generalized sinc kernel $s = \frac{1}{S^*} \overline{\mathcal{F}}(\mathbb{1}_{D^*})$. This equality holds in general only in the sense of the L^2 norm. If in addition $g \in l^1(\Gamma)$, then there is pointwise equality and uniform convergence.

2.3 Choice of the reciprocal cell.

The main difference between this formulation of Shannon's sampling theorem and the more classical one is (apart from the fact that we are using a possibly non-orthogonal sampling grid) the introduction of one additional degree of freedom, namely the reciprocal cell D^* , which has to meet both conditions (S1) and (S2). As long as both conditions are met the reconstruction formula is valid, but we are free to choose any of them. In practice D^* is most commonly assumed to be in one of the two cases:

Spectral support of the image. That means we take $D^* = \text{supp}(\hat{f})$. This choice has the obvious advantage that it always satisfies (S2), but it presents several disadvantages. First there is no reason why this definition should lead to a convenient filter in the reconstruction formula (11). The alternative we discuss below addresses exactly this issue. Secondly, \hat{f} will likely have no compact support, so that actually we will be forced to consider an "essential support" where the energy of \hat{f} is large enough. We shall come back to this idea in the next section.

Voronoi reciprocal cell. The other canonical choice consists of choosing D^* to be the largest and most isotropic cell that satisfies (S1), namely we choose as a reciprocal cell the Voronoi set $D_{\text{Vor}}^* = \text{Vor}(\Gamma^*)$ of the dual grid, as defined below.

$$D_{\text{Vor}}^* = \text{Vor}(\Gamma^*) = \{\gamma^* \in \mathbb{R}^d : \forall \gamma_0^* \in \Gamma^*, \gamma_0^* \neq 0 \Rightarrow \|\gamma^*\| < \|\gamma^* + \gamma_0^*\|\} \quad (12)$$

It can be easily verified that D_{Vor}^* is an admissible tile for the dual grid Γ^* , hence it always satisfies condition (S1) in Shannon's sampling theorem. In addition its mathematical and numerical treatment is much easier. In fact, in dimension 2, the Voronoi cell can be expressed as the product of three one-dimensional indicator functions, thus the spatial-domain convolution is "three-separable" in the sense that it can be expressed as a sequence of three one-dimensional convolutions. (In the case of the square grid, the third convolution would be in the direction of the diagonal and becomes unnecessary). Figure 6 shows the Voronoi cell corresponding to the square and hexagonal grids, plus an intermediate case where the basis vectors of the grid form an angle of 80° .

Whereas the Voronoi reciprocal cell provides more convenient formulas and always satisfies (S1), it does not address the band-limiting condition (S2). In order to do so, it should be adapted to the spectral support of the image $\text{supp}(\hat{f})$. In fact as we will see in section 5, D_{Vor}^* is the most sensible

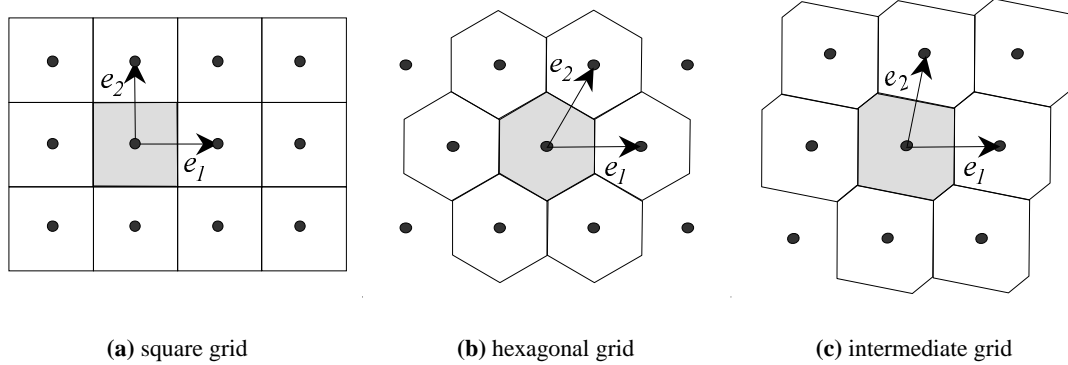


Figure 6: Voronoi reciprocal cell for three grids: square (generating vectors at 90°), hexagonal (60°), and an intermediate case where the generating vectors form an angle of 80° . Observe that the shape of this cell is in general a (not necessarily regular) hexagon, except in the square case. Hence the corresponding convolution filters in Shannon’s reconstruction formula are in general 3-separable, whereas in the square case they are 2-separable.

choice whenever \hat{f} is expected to be isotropic and radially decreasing, which is often the case in many imaging systems. However, many of the high resolution systems currently under development and study at CNES and other research centers, as well as some other more common situations do not satisfy these conditions (see section 3), which motivated to a large extent the present study.

But before going into the more complicated resolution measures, let us start with a few simple examples of systems that would prefer a Voronoi reciprocal cell, and others that would require another kind of D^* . These simple examples will be enough to show some of the central ideas in the rest of the paper. To do so we introduce a simple quantitative measure of sampling density and sampling efficiency, which is used to give a precise meaning to phrases like: “this is the best sampling grid”, or “this is the best reciprocal cell for this image”.

2.4 Critical Sampling.

Definition 1 (Sampling density) *The sampling density of a grid Γ is the number of samples per unit area, given by $\rho(\Gamma) = \frac{1}{\det(\Gamma)}$. Equivalently the sampling density can be computed as $\rho(\Gamma) = \frac{1}{|D|}$ for any admissible tile D .*

One would like to satisfy Shannon’s sampling theorem with as least samples as possible, *i.e.* with the lowest possible value of $\rho(\Gamma)$. The following definition expresses this desired limiting case:

Definition 2 (Critical sampling) Γ is a critical sampling grid for an image f iff

- (i) $(\Gamma^*, \text{supp}(\hat{f}))$ is a packing, and
- (ii) $\rho(\Gamma) = \inf \left\{ \rho(\Gamma_1) : (\Gamma_1^*, \text{supp}(\hat{f})) \text{ is a packing} \right\}$.

In such a case, since $(\Gamma^*, \text{supp}(\hat{f}))$ is a packing, there exists a reciprocal cell D^* satisfying both conditions in Shannon’s theorem. The trade-off between sampling density and the packing condition becomes clearer if we express the sampling density directly in terms of the dual grid. In fact from the definition of dual grid we observe that

$$\rho(\Gamma) = \frac{1}{2\pi} \det \Gamma^* = \frac{1}{(2\pi)^2 \rho(\Gamma^*)} = \frac{|D^*|}{(2\pi)^2} \quad (13)$$

for any admissible tile D^* of the dual grid Γ^* . So we want a dual grid with the smallest possible tile, which still contains $\text{supp}(\hat{f})$. The ideal situation is when $\text{supp}(\hat{f})$ is itself a tile, which leads us to the following definition.

Definition 3 (Sampling efficiency) *The sampling efficiency of a grid Γ with respect to an image f , which satisfy Shannon's theorem is defined as $\frac{|\text{supp}(\hat{f})|}{(2\pi)^2 \rho(\Gamma)}$ or equivalently as*

$$\frac{|\text{supp}(\hat{f})|}{|D^*|} \quad (14)$$

for any tile D^* of the dual grid Γ^* .

Thus, for a given spectral support, the sampling efficiency is inversely proportional to the sampling density, and we obtain critical sampling when the sampling efficiency is maximal among all grids satisfying Shannon's theorem. The sampling efficiency actually coincides with the *packing density*, which has been studied in a more general setting, and in more dimensions. In dimension two the main result states that the maximal packing density with a circular tile is attained for the hexagonal grid (even if we allow much more general transformations that just integer multiples of two generating vectors) (Bauer, 1981; Rogers, 1964; Wühtrich and Stucki, 1991). Restating this result in sampling terms we can say that when the spectral support of f is a circular domain, then the maximal sampling efficiency (critical sampling) is obtained when the dual sampling grid Γ^* is hexagonal. Figure 7 and table 1 illustrate how this happens and to what an extent the hexagonal sampling is more efficient than the square sampling in this case. From these figures we can conclude that for a circular spectral support satisfying Shannon's theorem requires 13.4% more samples in the case of a square grid than in the case of an hexagonal grid.

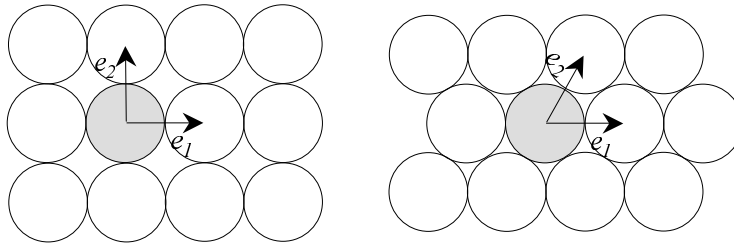


Figure 7: Denset packings of circular tiles for the square and hexagonal grids.

Grid	Packing Density	Covering Density
Triangular	$\frac{\pi}{3\sqrt{3}} \approx 60.46\%$	$\frac{4\pi}{3\sqrt{3}} \approx 241.84\%$
Square	$\frac{\pi}{4} \approx 78.54\%$	$\frac{\pi}{2} \approx 157.08\%$
Hexagonal	$\frac{\pi}{2\sqrt{3}} \approx 90.69\%$	$\frac{2\pi}{3\sqrt{3}} \approx 120.92\%$

Table 1: Maximal (minimal) densities of packings (coverings) by circles for the three regular grids.

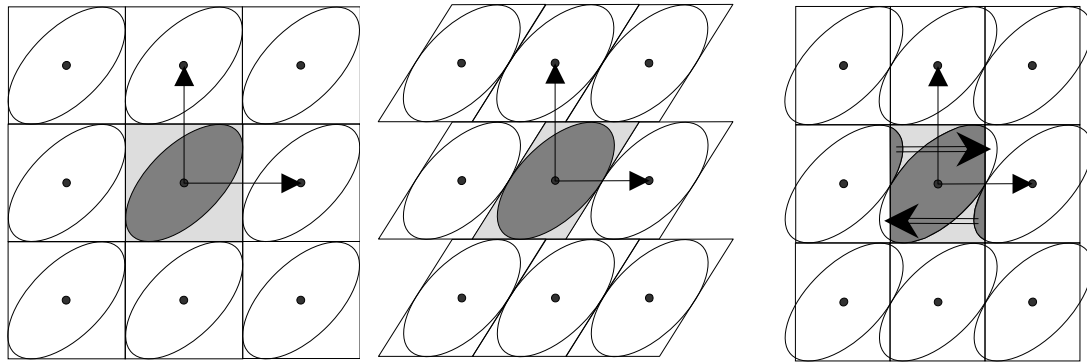
Observe that despite the fact that the square grid is not the most efficient sampling in this case, the square dual sampling grid shown in figure 7 is still the best possible choice among square grids. In this case we say that this square grid is critical within its geometry, but not globally critical:

Definition 4 Γ is a critical sampling grid for an image f within its geometry iff

- (i) $(\Gamma^*, \text{supp}(\hat{f}))$ is a packing, and
- (ii) $\rho(\Gamma) = \inf \left\{ \rho(\Gamma_1) : (\Gamma_1^*, \text{supp}(\hat{f})) \text{ is a packing, and } \exists a \in \mathbb{R}, \Gamma_1 = a\Gamma \right\}$.

In the examples discussed so far, in particular for both examples in figure 7, the Voronoi reciprocal cell is admissible in Shannon's theorem, even in the case of critical or partially critical sampling. And this, due to the isotropy of the spectral support.

This is no longer the case in the following example, where the spectral support of f is assumed to be elliptic with eccentricity 2 and with the major axis at 45° from the horizontal. Assume that we want to sample an image with such a support with a vertical/horizontal rectangular sampling grid. If we stick to Voronoi reciprocal cells, then our most efficient sampling will be far from critical (even within geometry), as shown in figure 8(a). We can obtain a more efficient sampling by relaxing the sampling rate in the horizontal direction, thus leading to the reciprocal cell in figure 8(b) with the shape of a parallelogram. This is still not critical, since we can still relax a little bit the vertical sampling rate (though with a more complicated D^*), but this is enough for illustration purposes.



(a) Voronoi reciprocal cell

(b) More efficient sampling

(c) Non-admissible reciprocal cell, which still allows an anti-aliasing

Figure 8: The figures above show the dual grid Γ and its generating vectors as well as the spectral support of the image to be sampled in dark-gray, and the chosen reciprocal cell in light-gray. We also show some aliases of the spectral support $\text{supp}(\hat{h}) + \gamma^*$ and the chosen reciprocal cell $D^* + \gamma^*$ for some $\gamma^* \in \Gamma^*$. Observe that Voronoi reciprocal cells (as in subfigure a) no longer yield a critical sampling when the spectral support is strongly anisotropic. To obtain a more efficient sampling we have to use a better adapted reciprocal cell like the parallelogram in subfigure (b). If displayed as-is (*i.e.* with no interpolation or with the usual sinc or zero-padding interpolation) the sampled image will show some aliasing artifacts as illustrated in subfigure c, but these can be solved by using an admissible reciprocal cell in the reconstruction formula. Equivalently, one can translate the aliased Fourier coefficients to their original location. The value of $\gamma^* \in \Gamma^*$ to be used for this translation can be deduced from the knowledge of the admissible reciprocal cell. Once the Fourier coefficients were translated to their correct location, we can resample at the rate given by the Voronoi reciprocal cell in subfigure (a).

Observe that a correct choice of D^* is necessary only when applying Shannon's theorem to visualize,

restore or interpolate the image. In the coding phase, any other reciprocal cell which forms a tiling would do just as well. The rectangular cell shown in figure 8(c) contains all the information about the original image. Some parts of the original spectral support were lost but they reappear as aliases in other regions, not used by the rest of the spectral support. If we display the sampled image as is, we would in fact see these aliasing artifacts, but since we know the shape of the spectral support, we can re-place each coefficient at its correct location. To do so we need some spectral space of course, and we have to oversample at the rate given by the Voronoi cell in figure 8(a). This is one of the main ideas that will be exploited in sections 4 through 6, but in a more realistic case, with noise, with unbounded spectral support but known decay rate, and some other more realistic details that will be the subject of the next section.

3 Modeling of image acquisition systems

In the previous section we assumed an ideal situation, but in real sampling systems present two limitations which force us to deal with approximations to Shannon's sampling theorem:

- *noise*, due to the fact that sampling systems can only make approximate measurements, and
- the fact that the analog image before sampling actually has *no compact spectral support*, and is neither *periodic* nor has an *infinite spatial support*.

In fact, a generic image sampling system can be viewed as a combination of:

- an optical system whose purpose is to focus the incoming light in the focal plane, and
- a set of sensors distributed on the focal plane which count the number of photons which meet this precise location during a certain period of time.

and the whole system might be moving during capture. The whole action of the sampling system can be modeled by the following formula:

$$g = \mathbb{1}_\Omega \cdot (\Delta_\Gamma \cdot (\overline{\mathcal{F}}(H) * f) + n) \quad (15)$$

where

- f represents the ideal image just before it enters the imaging system, this is assumed to be a radon measure;
- n represents the noise due to the imprecision in the measurements of the sensors;
- Γ represents the geometry of the array of sensors, which are assumed to be distributed on a regular grid;
- Ω represents the extent of the sensor array, which can never be infinite;
- H represents the transfer function of the instrument, which is a combination of several effects to be detailed later.

We observe that even in the absence of noise, and even if the transfer function H had a compact support, the resulting sampled image will have no compact spectral support, due to multiplication by $\mathbb{1}_\Omega$. Hence the periodization of the finite-extent sampled image g (needed to obtain a discrete spectrum) will introduce some aliasing. However this difficulty can be solved by a standard technique which consists of multiplying the whole image by a smooth window having a very smooth transition between 0 and 1 near $\partial\Omega$. With this technique this necessary amount of aliasing can be usually reduced to a level much below the noise level, so in the sequel we will assume a simplified image model:

$$g = \Delta_\Gamma \cdot (\overline{\mathcal{F}}(H) * f) + n \quad (16)$$

in which the sampled image g can be considered infinite or periodic.

Now let's consider the transfer function. It can be modeled as the product of three main effects:

$$H = H_{\text{sen}} H_{\text{mov}} H_{\text{opt}} \quad (17)$$

The first term is due to the fact that the sensors do not pick the value of f at one point but rather integrate all the photons that arrive at its sensitive region which usually has a square shape, thus:

$$FTM_{\text{sen}}(\xi, \eta) = \mathcal{F}(\mathbb{1}_{|x| < c/2} * \mathbb{1}_{|y| < c/2}) e^{-\beta c |\eta|} = 4c^2 \text{sinc}(\pi \xi c) \text{sinc}(\pi \eta c) e^{-\beta c |\eta|} \quad (18)$$

where c represents the size of the sensor. The exponential term, where $\beta \approx 0.9$ takes into account the conductivity between one sensor and its neighbour. The second term in equation (17) takes into account the fact that the system might be moving during capture. If it moves a distance d at direction \mathbf{v} then:

$$FTM_{\text{mov}}(\xi) = \mathcal{F}(\mathbb{1}_{|\langle \mathbf{x}, \mathbf{v} \rangle| < d/2}) = \text{sinc}(d\pi \langle \xi, \mathbf{v} \rangle). \quad (19)$$

Finally, the optical system itself is not perfect and it also convolves the signal by a transfer function which might be very complicated. For the satellite imaging systems under study at CNES the following model is quite accurate:

$$FTM_{\text{opt}}(\xi, \eta) = e^{-\alpha c \sqrt{u^2 + v^2}} \quad (20)$$

for values of α of about 2, which means that in the useful region the image is low-pass filtered much more by the sensors themselves than by the optical system. Figure 9 shows some level lines of H for three instances of the previously described imaging system, where the (square) sensors of size c are assumed to be distributed at regular spacings of c on a square grid: (a) ccd1: an ideal case with a still camera ($d = 0$) where the optics and sensor isolation are assumed to be perfect ($\alpha = \beta = 0$); (b) ccd2: a less ideal case, with a still camera, but more realistic values for optics and conductivity ($\alpha = 1.9$ and $\beta = 0.9$); (c) same situation as before but with the camera moving a distance $d = 2.35c$ along a direction \mathbf{v} which forms an angle of 11° with the horizontal.

Observe that all these systems produce highly aliased images. In figure 9(a) for instance, the transfer function only decayed from 1 to 0.6 at the border of the reciprocal cell. In order to reduce aliasing to a minimum we would ideally want the reciprocal cell to cover a larger region, *e.g.* up to the first zero-crossing of the sinc function. But that would require to double the sampling rate in both directions, which is not easy to do without changing the size of the sensors, because this would require the sensors to overlap.⁴

However, several mechanisms have been proposed at the French Spatial Agency (CNES), which allow to increase the sampling rate without reducing the size of the sensors. Some of them (called hypermode and supermode) have been implemented on the SPOT5 earth observation satellite (launched in April this year)(Latry and Rougé; Rougé, 1997). Other novel mechanisms for improving the sampling rate involve both square and hexagonal grids. The corresponding transfer functions for all these systems are shown in figures 10, 11 and 12.

None of these systems exactly satisfies Shannon's compact support condition, and therefore they produce aliased images. But for some of them (see for instance figure 12(c)) aliasing effects may be far below the noise level and thus imperceptible, meaning that we unnecessarily increased the sampling rate. In order to find the right balance between increasing resolution and reducing aliasing artifacts on one hand, and on the other reducing the sampling rate to a minimum (which is dictated by bandwidth limitations), we need to extend the notions of critical sampling, sampling density and sampling efficiency introduced in section 2.4 to the case of infinite spectral support.

In this sense, it has been proposed (Rougé, 1997) to consider the *essential spectral support* of the image to be sampled, namely:

$$\text{ess supp}(\hat{f}) := \left\{ \xi \in \mathbb{R}^2 : |\hat{f}| \geq \sigma|\Omega| \right\}. \quad (21)$$

i.e. the region over which the spectrum of the image is larger than the noise level (here we assumed a white iid Gaussian noise with variance σ^2 on an image domain Ω). Then we can apply the usual critical sampling and sampling efficiency definitions with the essential instead of the real spectral support. In (Durand, 1998) this criterion has been refined in order to take into account the signal to noise ratio at each location within the essential spectral support. The result is a resolution measure which is invariant

⁴ On the other hand reducing the size of the sensors wouldn't help, because this zero crossing would move to a higher frequency, and we would have the same aliasing problem, just at another scale.

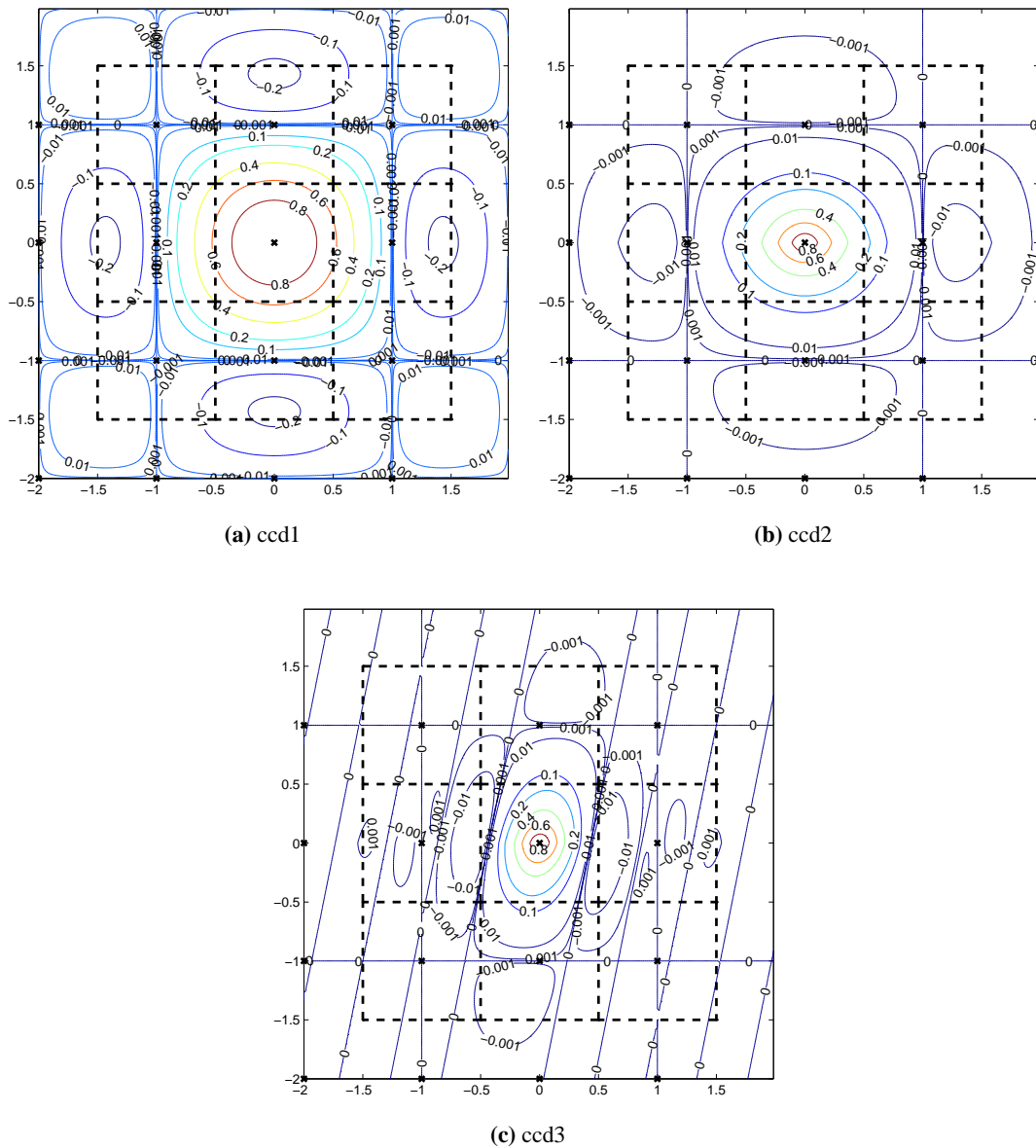


Figure 9: Modulation Transfer Functions and the corresponding Voronoi cells for three imaging systems composed of square CCD arrays. (a) Perfect optics and fixed camera; (b) more realistic optics; (c) the camera moves at uniform speed during capture. *Legend:* In all subfigures the dashed lines represent the boundaries of the Voronoi reciprocal cell and its aliases. The crosses represent the points in the dual (“aliasing”) grid Γ^* . The remaining curves are the level-lines of the MTF.

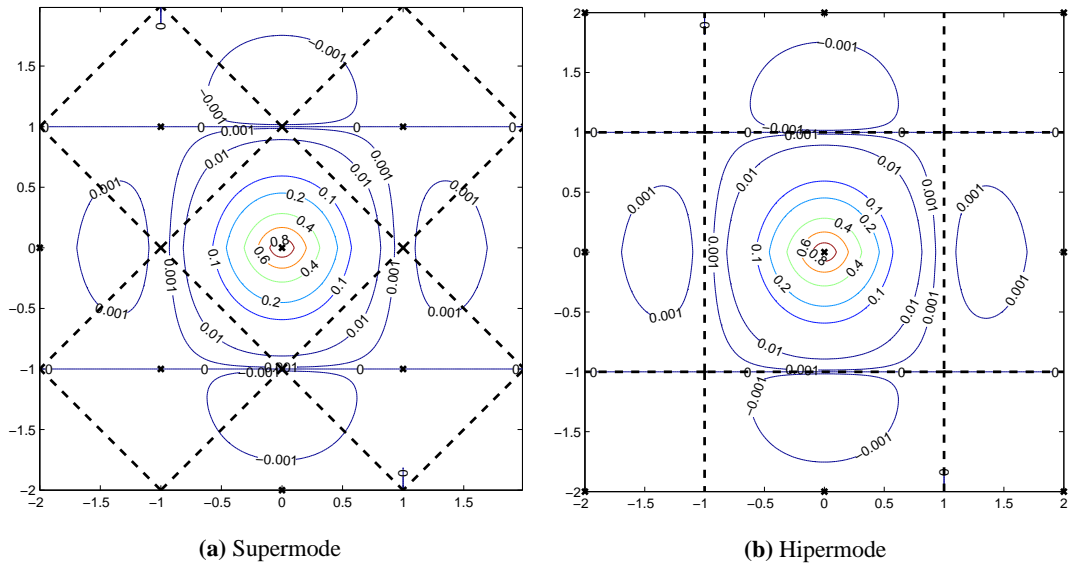


Figure 10: Modulation transfer function and the corresponding Voronoi cell for the Supermode sampling system used by the SPOT5 HRG earth observation satellite, and the Hipermode sampling system which has also been considered during its design. (See also the legend in figure 9.)

with respect to any linear post-processing that can be done to reduce noise or aliasing. In the next section we further develop these ideas, and propose a resolution measure which will also allow us to choose the best reciprocal cell (section 5).

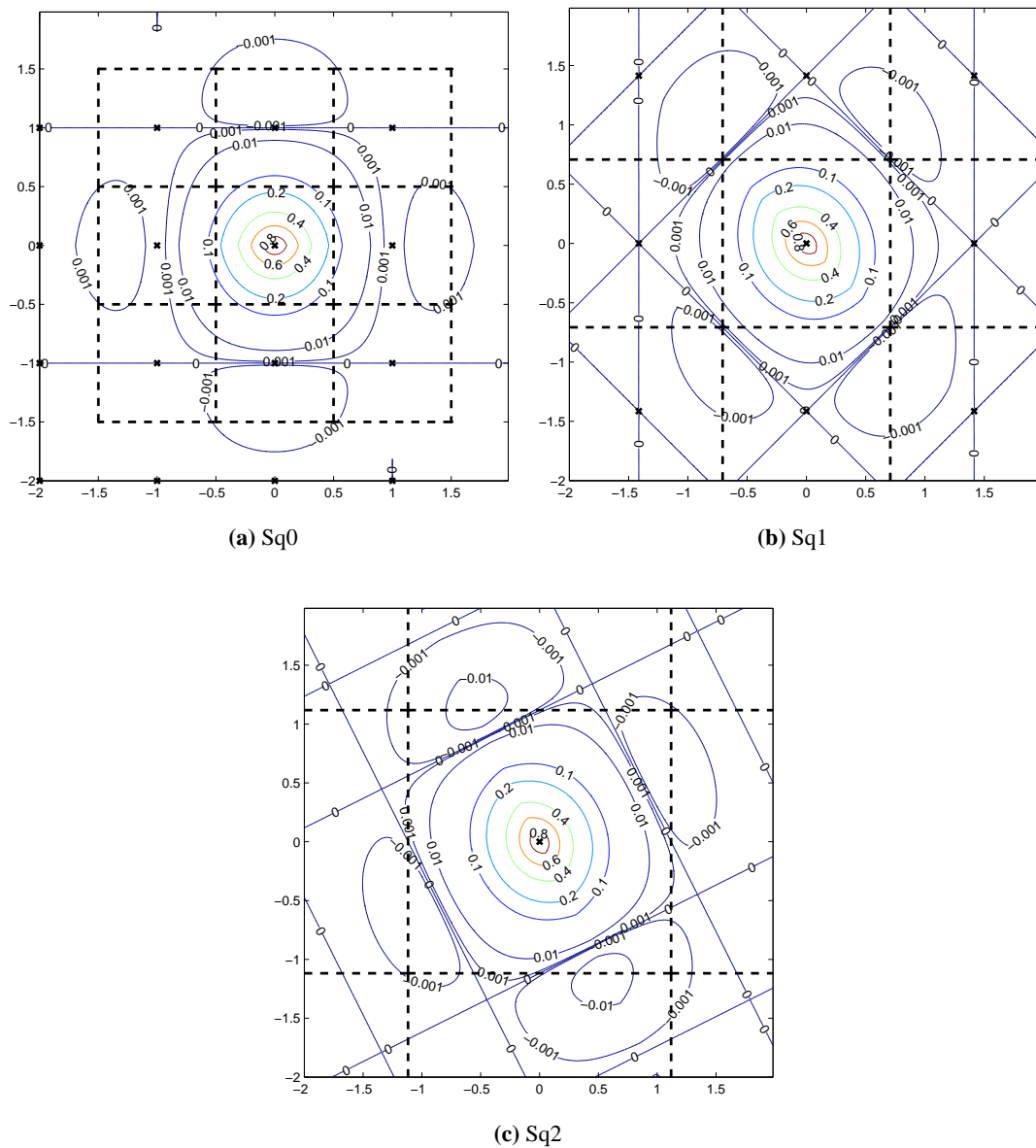


Figure 11: Modulation transfer function and the corresponding Voronoi cell for some high-resolution square sampling systems. (See also the legend in figure 9.)

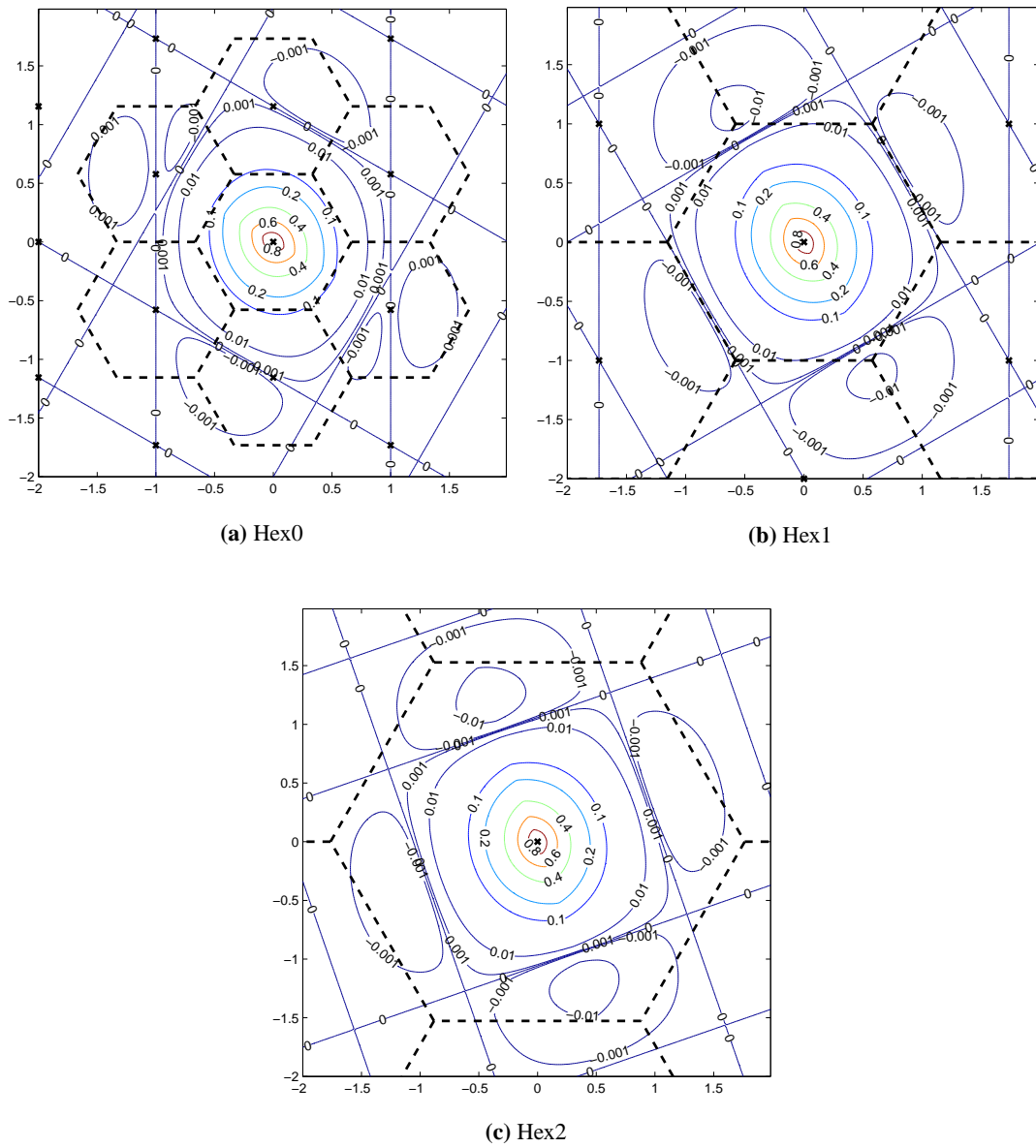


Figure 12: Modulation transfer function and the corresponding Voronoi cell for some high-resolution hexagonal sampling systems. (See also the legend in figure 9.)

4 Effective Resolution

4.1 Nominal Resolution

Usually resolution is defined in terms of dots-per-inch (dpi), lines-per-inch (lpi), etc. which measures the sampling density. This usual definition of resolution is what we shall call *spectral nominal resolution* $\sqrt{r_{\text{nom}}}$ (the reason for the word spectral will become clear in a minute), which we define precisely as the density of the sampling grid:

$$r_{\text{nom}} := \rho(\Gamma)$$

Note that r_{nom} is measured in terms of sampling points per unit area (e.g. dots per square inch). To obtain a dpi-style measure of resolution we simply use $\sqrt{r_{\text{nom}}}$. We also define a *spatial nominal resolution* which is proportional to the distance between neighboring samples, namely:

$$R_{\text{nom}} = r_{\text{nom}}^{-\frac{1}{2}}.$$

Usually we will measure these resolutions in terms of c where c^2 is the area covered by each of the sensors. Thus:

- a typical square CCD array (Sq0) has nominal resolution $r_{\text{nom}} = 1c^{-2}$, $R_{\text{nom}} = 1c$.
- a pair of CCD arrays in quincunx arrangement (SPOT5's supermode) has nominal resolution $r_{\text{nom}} = 2c^{-2}$, $R_{\text{nom}} = \frac{1}{\sqrt{2}}c$.
- a system of four CCD arrays (Hipermode) has nominal resolution $r_{\text{nom}} = 4c^{-2}$, $R_{\text{nom}} = \frac{1}{2}c$
- the hexagonal system Hex1 has $r_{\text{nom}} = 2\sqrt{3}c^{-2} \approx 3.46c^{-2}$, $R_{\text{nom}} = \frac{1}{\sqrt{2\sqrt{3}}}c$

Now consider a tile D^* of the dual grid Γ^* , for instance the Voronoi cell of Γ^* . Given the relationship between Γ and Γ^* we can show that the spectral nominal resolution

$$r_{\text{nom}} = \frac{1}{(2\pi)^2} \frac{1}{\rho(\Gamma^*)} = \frac{1}{(2\pi)^2} \int_{D^*} 1 \quad (22)$$

is in fact proportional to the area of the Voronoi cell of the dual grid Γ^* , whereas the spatial nominal resolution R_{nom}^2 is proportional to (the square root of) the area of the Voronoi cell of the sampling grid Γ (which explains the names we chose).

Considering noise and aliasing in the resolution measure. The spectrum of the sampled image \hat{g} may contain at different points very different amounts of noise and aliasing, which is not reflected by equation (22) which gives the same weight 1 to all Fourier coefficients of \hat{g} .

Assume an image acquisition system is defined by a triplet (Γ, H, n) . Then, according to equation (16) the Fourier transform of the acquired image is:

$$\begin{aligned} \hat{g} &= \mathcal{F}(\Delta_\Gamma) * (H \cdot \hat{f}) + \hat{n} \\ &= |D^*| \Delta_{\Gamma^*} * (H \cdot \hat{f}) + \hat{n} \\ &= |D^*| \sum_{\omega_0 \in \Gamma^*} (H \cdot \hat{f})(\cdot + \omega_0) + \hat{n} \end{aligned} \quad (23)$$

which we can rewrite in terms of

$$G = |D^*| H \cdot \hat{f} \quad (24)$$

$$G_{\text{alias}}(\xi) = \sum_{\gamma^* \in \Gamma^* \setminus \{0\}} G(\xi + \gamma^*) \quad (25)$$

as follows:

$$\hat{g} = G + G_{\text{alias}} + \hat{n}. \quad (26)$$

This last expression shows clearly the aliasing and noise errors present in the sampled image g . Thus it is not fair to consider all pixels equal. In the next section we shall refine our resolution measure by substituting the 1 in equation (22) by a weighting function $w(G(\xi), G_{\text{alias}}(\xi), \hat{n}(\xi))$, which is close to 1 in the absence of aliasing and noise, and close to 0 when either aliasing and noise dominates over the signal.

4.2 Effective resolution

Once such a weighting function w is defined we can define the effective resolution as:

$$r_{\text{eff}}(\Gamma, D^*, H, n, f) := \frac{1}{(2\pi)^2} \int_{D^*} w(G(\xi), G_{\text{alias}}(\xi), \hat{n}(\xi)) d\xi. \quad (27)$$

But this definition has two problems. First, it depends on the image f *before sampling* which is unknown. And secondly it will give a different value for each realization of the image f and the noise n . We would be tempted to solve both problems by defining

$$r_{\text{eff}}(\Gamma, D^*, H, \mathcal{N}, \mathcal{I}) := \mathbb{E}\{r_{\text{eff}}(\Gamma, D^*, H, n, f)\} \quad (28)$$

Where the expected value is taken with respect to the random variables $n \sim \mathcal{N}$ and $f \sim \mathcal{I}$. We shall come back later to the problems posed by the computation of such expected value. For the moment let us concentrate on the choice of a suitable weighting function w .

Conditions on the weighting function w . A first condition we want to impose is that w be independent of any invertible⁵ linear filtering k that we can perform on the digitized image g . The reason for this is that it is very common to apply “anti-aliasing” or “denoising” filters to an image after sampling, in order to reduce these artifacts (in section 6 we give an example of such manipulations). However, we want our measure to reflect the resolution of the “best image” that we can obtain by such manipulations from the acquisition system.

In order to express this condition in terms of w observe that $\mathcal{F}(k * g) = \hat{k} \cdot G + \hat{k} \cdot G_{\text{alias}} + \hat{k} \cdot \hat{n} = \hat{k} \cdot \hat{g}$. Hence we require that:

$$w(\hat{k} \cdot G, \hat{k} \cdot G_{\text{alias}}, \hat{k} \cdot \hat{n}) = w(G, G_{\text{alias}}, \hat{n}), \text{ for all } k \quad (29)$$

or put another way, the weighting function only depends on *relative* noise and aliasing.

$$w(G, G_{\text{alias}}, \hat{n}) = w\left(1, \frac{G_{\text{alias}}}{G}, \frac{\hat{n}}{G}\right) = W\left(\frac{G_{\text{alias}}}{G}, \frac{\hat{n}}{G}\right). \quad (\text{W0})$$

Next, we don’t care about the *phase* of relative noise or aliasing, we just consider its *magnitude*.

$$W(a, b) = W(|a|, |b|). \quad (\text{W1})$$

Finally, we should impose a few conditions, to ensure that W actually behaves like a weighting function, which penalizes high levels of relative noise and aliasing:

$$W(a, b) \in [0, 1] \quad (\text{W2})$$

$$W(0, 0) = 1 \quad (\text{W3})$$

$$W(+\infty, b) = W(a, +\infty) = 0 \quad (\text{W4})$$

$$W(\cdot, b) \text{ non-increasing for all } b, W(a, \cdot) \text{ non-increasing for all } a. \quad (\text{W5})$$

⁵ Here we mean that k should be invertible in the sense that $\hat{k} \neq 0$ almost everywhere.

This still leaves a huge number of possibilities. For our experiments we considered the following two functions:

$$W(a, b) = \left(1 - \frac{a}{\theta_{\text{alias}}}\right)^+ \left(1 - \frac{b}{\theta_{\text{noise}}}\right)^+, \quad (30)$$

$$W(a, b) = \min\left\{\left(1 - \frac{a}{\theta_{\text{alias}}}\right)^+, \left(1 - \frac{b}{\theta_{\text{noise}}}\right)^+\right\}. \quad (31)$$

Observe that both functions satisfy a stronger condition than equation (W4). Namely, $W(a, b) = 0$ whenever $a \geq \theta_{\text{alias}}$ or $b \geq \theta_{\text{noise}}$, *i.e.* when either the relative aliasing or noise exceeds the specified thresholds. In principle we should set $\theta_{\text{alias}} \approx \theta_{\text{noise}} \approx 1$ meaning that whenever the aliasing or the noise are larger than the real signal, then we should ignore the corresponding Fourier coefficient because it contains almost no useful information. However, as we shall see in section 6, our experiments indicate that we obtain better results for $\theta_{\text{alias}} \approx 0.2$ and $\theta_{\text{noise}} \approx 5$, which may indicate that our perceptual system tolerates noise much better than aliasing artifacts.

Computing the expected value in equation (28). In order to compute the expected value in equation (28) we need suitable noise and image models. The noise model doesn't pose any serious problems, usually a white noise model $\mathcal{N} = N(0, \sigma)$ is accurate enough. But accurate statistical image models \mathcal{I} are unknown in general. In its absence we can compute the empirical expected value by taking the mean of equation (27) for a large number of images. But this is computationally very expensive, and its mathematical treatment is also much too complicated!!

A more sensible choice is to modify equation (28), by evaluating w not on a particular realization of the relative aliasing and noise, but on its expected value directly:

$$r_{\text{eff}}(\Gamma, D^*, H, \mathcal{N}, \mathcal{I}) := \frac{1}{(2\pi)^2} \int_{D^*} W \left(\sqrt{\frac{\mathbb{E}\{|G_{\text{alias}}(\xi)|^2\}}{\mathbb{E}\{|G(\xi)|^2\}}}, \sqrt{\frac{\mathbb{E}\{|\hat{n}(\xi)|^2\}}{\mathbb{E}\{|G(\xi)|^2\}}} \right) d\xi \quad (32)$$

Note that this expression is much simpler, since it doesn't take into account any possible dependency between $\hat{f}(\xi)$ and $\hat{f}(\xi')$ for $\xi \neq \xi'$, except for the term $\mathbb{E}\{|G_{\text{alias}}(\xi)|\}$ which considers the dependency between $\hat{f}(\xi)$ and its aliases $\hat{f}(\xi + \gamma^*)$ for all $\gamma^* \in \Gamma^*$. We will also simplify this term by substituting $\sqrt{\mathbb{E}\{|(H\hat{f})_{\text{alias}}(\xi)|^2\}}$ by $\sqrt{|HF|_{\text{alias}}^2(\xi)}$ where $F(\xi) = \sqrt{\mathbb{E}\{|\hat{f}(\xi)|^2\}}$. So, we define:

$$r_{\text{eff}}(\Gamma, D^*, H, N, F) := \frac{1}{(2\pi)^2} \int_{D^*} W \left(\sqrt{\frac{|HF|_{\text{alias}}^2(\xi)}{|HF|^2(\xi)}}, \sqrt{\frac{N^2(\xi)}{|HF|^2(\xi)}} \right) d\xi \quad (33)$$

which no longer depends on the image and noise models \mathcal{I}, \mathcal{N} , but just on (the square root of) the second order moments $F(\xi) = \sqrt{\mathbb{E}\{|\hat{f}(\xi)|^2\}}$ and $N(\xi) = \sqrt{\mathbb{E}\{|\hat{n}(\xi)|^2\}}$, for $n \sim \mathcal{N}$, $f \sim \mathcal{I}$. Now for a white noise model $\mathcal{N} = N(0, \sigma)$, we just have a constant $N(\xi) = \sigma$. (Note however that this study works equally well for any other zero-mean colored Gaussian noise model, which is determined by the function $N(\xi)$). Finally if the images are supposed to be in the BV space, a reasonable assumption is $F(\xi) \sim \frac{1}{|\xi|}$ except near the origin. In fact this assumption is quite accurate: by taking the empirical expected value $\mathbb{E}\{|\hat{f}(\xi)|^2\}$ for a large family of images, we obtain a function $F(\xi)$ whose shape is quite similar to that of $\frac{1}{|\xi|}$ (see figure 13).

We note that despite the many simplifications we did between equations (28) and (33), we find that both expressions yield experimentally comparable r_{eff} values, which indicates that these simplifications did not essentially affect the notion of resolution that we wanted to capture initially. We can conclude that the image model we used in equation (33), namely a coloured Gaussian noise, where $|\hat{f}(\xi)| \sim N(0, F(\xi)^2)$, is accurate enough for computing image resolution as defined in equation (28).

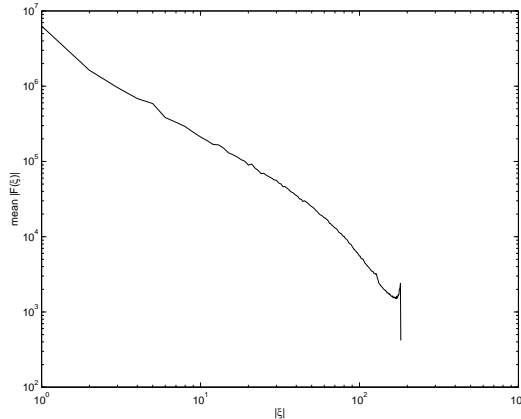


Figure 13: *Learning the decay rate of Fourier coefficients of natural images.* After computing the mean of $|\hat{f}|$ for several natural images f , we obtain a roughly rotationally invariant function F . Here we show the mean of this function over each ring where $r \leq |\xi| < r + 1$. The result shows that the Fourier coefficients decay like $|\hat{f}|(\xi) \approx C|\xi|^p$ with $1 \leq p \leq 2$, and typically $p \approx 1.6$.

4.3 Information density.

The previous definition of effective resolution was motivated by perceptual arguments on the influence of noise and aliasing artifacts, that lead us to a certain choice of weighting function W .

On the other hand we can attempt to measure the amount of information of the original image $h * f$ that is preserved by the samples g . So in this section we will choose as weighting function the mutual information between the Fourier coefficient $\hat{g}(\xi)$ of the sampled image and the corresponding Fourier coefficient of the original image $G(\xi) = H(\xi)\hat{f}(\xi)$. This choice leads to a measure of effective information density

$$\frac{1}{(2\pi)^2} I_{\text{eff}} := \int_{D^*} I(\hat{g}(\xi), G(\xi)) d\xi \quad (34)$$

which measures the number of bits of information (per unit area) of the original image that can be obtained from the sampled image. The reciprocal cell will be chosen, as in the previous section in such a way that (Γ^*, D^*) is a packing.

Observe that this packing condition means that, if all $\hat{f}(\xi)$ are independent then all $G(\xi)$ and $\hat{g}(\xi')$ are independent except when $\xi = \xi'$. Hence the integral $\int_{D^*} I(\hat{g}(\xi), G(\xi)) d\xi$ equals the mutual information $I(\hat{g}, G|_{D^*})$ between the sampled image (in the Fourier domain) and the Fourier transform of the original image $G = H\hat{f}$ restricted to the reciprocal cell D^* .

We could still refine this model by computing $I(\hat{g}, G)$ without restriction to a reciprocal cell, and without assuming the independence between the Fourier coefficients of \hat{f} , but for the moment we are interested in a simpler model which still allows some qualitative comparisons with our results of the previous section on resolution measures.

In any case the mutual information $I(\hat{g}(\xi), G(\xi))$ inside the integral (and hence the information density measure I_{eff}) also depends on the statistical model we choose for the image. The following theorem gives a general lower bound, as well as the value of I_{eff} for two simple image models. In order

to simplify the notation we express the theorem in terms of the following dimensionless variables:

$$a^2(\xi) = \frac{\text{Var}\{G_{\text{alias}}(\xi)\}}{\text{Var}\{G(\xi)\}} \quad (\text{relative aliasing}) \quad (35)$$

$$b^2(\xi) = \frac{\text{Var}\{\hat{n}(\xi)\}}{\text{Var}\{G(\xi)\}} = \frac{|N|^2(\xi)}{|HF|^2(\xi)} \quad (\text{relative noise}) \quad (36)$$

$$k^2(\xi) = \frac{\text{Cov}\{G(\xi), G_{\text{alias}}(\xi)\}^2}{\text{Var}\{G(\xi)\}\text{Var}\{G_{\text{alias}}(\xi)\}} \quad (\text{relative covariance}). \quad (37)$$

Observe that in the case where $\hat{f}(\xi)$ is independent of $\hat{f}(\xi')$ for all $\xi \neq \xi'$, we have a more convenient way of writing the relative aliasing, namely:

$$a^2(\xi) = \frac{|HF|_{\text{alias}}^2(\xi)}{|HF|^2(\xi)} \quad (38)$$

where $F(\xi) = \sqrt{\text{E}\{\hat{f}(\xi)\}}$ as in the previous section.

Theorem 2 *Given a sampled image g which relates to the original image $h*f$ through equations (16) and (26), the mutual information between $\hat{g}(\xi)$ and $\hat{G}(\xi) = H(\xi)\hat{f}(\xi)$ is always lower bounded by:*

$$I(\hat{g}(\xi), G(\xi)) \geq \frac{1}{2} \log_2 \left(1 + \frac{2^{2H(G(\xi))}}{2^{2H(G_{\text{alias}}(\xi))} + 2^{2H(\hat{n}(\xi))}} \right). \quad (39)$$

This lower bound is achieved when $G(\xi)$, $G_{\text{alias}}(\xi)$ and $\hat{n}(\xi)$ are all independent zero-mean Gaussian variables. In that case the mutual information reduces to:

$$I(\hat{g}(\xi), G(\xi)) = \frac{1}{2} \log_2 \left(1 + \frac{1}{a^2 + b^2}(\xi) \right). \quad (40)$$

Finally, if we refine the previous model, by considering the dependence between $G(\xi)$ and $G_{\text{alias}}(\xi)$ through its covariance, the mutual information is:

$$I(\hat{g}(\xi), G(\xi)) = \frac{1}{2} \log_2 \left(1 + \frac{1 + a^2 k^2 + 2ak}{a^2(1 - k^2) + b^2}(\xi) \right). \quad (41)$$

Non-independent Gaussian Modeling. Observe that in the last equation of the theorem, we considered the dependence between $G(\xi)$ and $G_{\text{alias}}(\xi)$, since both terms depend on the Fourier coefficients $\hat{f}(\xi + \gamma^*)$ (for $\gamma^* \in \Gamma^*$) of the original image, which are not necessarily independent. On the other hand we still consider $\hat{n}(\xi)$ to be independent of the other terms since it is usually a white noise.

Effective and Nominal Information Density. Finally our information density measure becomes, in the non-independent Gaussian case:

$$I_{\text{eff}}(\Gamma, D^*, H, N, F, k) = \frac{1}{(2\pi)^2} \int_{D^*} \frac{1}{2} \log_2 \left(1 + \frac{1 + a^2 k^2 + 2ak}{a^2(1 - k^2) + b^2}(\xi) \right) d\xi \quad (42)$$

which can be written as

$$I_{\text{eff}} = \frac{1}{(2\pi)^2} \int_{D^*} W_k(a(\xi), b(\xi)) d\xi \quad (43)$$

with $W_k(a, b) = \frac{1}{2} \log(1 + \frac{1 + a^2 k^2 + ak}{a^2(1 - k^2) + b^2})$. Observe that if we consider aliasing to be independent from the original image ($k = 0$), then $W_0(a, b) = \frac{1}{2} \log(1 + \frac{1}{a^2 + b^2})$. We observe that this function W_0 satisfies all

requirements (W0-W5) except for (W2) and (W3). Instead $W_0(a, b) \in [0, +\infty]$ with $W_0(0, 0) = +\infty$. But this is normal since we are not measuring resolution in terms of sampling per unit area, we are measuring information in terms of bits per unit area.

For $k \neq 0$, W_k still satisfies (W0,W1,W5) and it partially satisfies (W4). In fact $W_k(a, +\infty) = 0$ but $W_k(+\infty, b)$ may be positive. This happens for instance when $\text{Var}\{G(\xi)\} \ll \text{Var}\{G_{\text{alias}}(\xi)\}$. In this case we have infinite relative aliasing, but the sampled coefficient (essentially $G_{\text{alias}}(\xi)$) still gives some information about $G(\xi)$ through the covariance between $G(\xi)$ and $G_{\text{alias}}(\xi)$.

An ideal system with no aliasing ($H = \mathbb{1}_{D^*}$) and no noise $N(\xi) = 0$ would yield infinite information density, but as a reference to compare with our system, we can consider the “nominal information density” to be the one that is obtained by a semi-ideal system with no aliasing ($H = \mathbb{1}_{D^*}$) and a minimal amount of noise $N(\xi) = \varepsilon$.

$$I_{\text{nom}}(\Gamma, D^*, \varepsilon, F) := \frac{1}{(2\pi)^2} \int_{D^*} \frac{1}{2} \log_2 \left(1 + \frac{|F(\xi)|^2}{\varepsilon^2} \right) d\xi \quad (44)$$

Aliasing vs. Noise in Information and Resolution Measures. The previous definition of effective information density shows that for a given level of relative distortion d , if this distortion comes from (image-independent) noise, then information decreases to a larger extent than if it comes from (image-dependent) aliasing. In fact, in the first case, $a = 0$, $d = b$ and $W(a, b) = \frac{1}{2} \log_2(1 + \frac{1}{d^2})$, whereas in the second case $a = d$, $b = 0$ and $W(a, b) = \frac{1}{2} \log_2(1 + \frac{1+k^2d^2+kd}{1-k^2})$, which is larger.

With respect to the resolution measures introduced in the previous subsections, we observed the contrary: for a given level of relative distortion, if it comes from noise, resolution decreases less than if it comes from aliasing (cf. the discussion on the choice of parameters $\theta_{\text{noise}} \gg \theta_{\text{alias}}$).

This comes from the fact that image-independent distortions such as white noise are more easily ignored by our perceptual system than image-dependent distortions such as aliasing. From an information-theoretical point of view, however, the latter gives us more information than the former for reconstructing the original image, provided we know the statistics of the dependence between the distortion (aliasing) and the original image.

So, If we know how to correct aliasing artifacts from knowledge of image statistics, then it makes sense to introduce this statistical knowledge in our information measure. Otherwise, it is better to consider it as noise and eliminate it.

5 Optimal Reciprocal Cell

In the previous section, all measures of image resolution or information that we defined, depend upon the shape of the reciprocal cell D^* . In this section we address the problem of choosing such a reciprocal cell. However, once we defined our measures of image resolution and information, we just define the optimal reciprocal cells as those that maximize either resolution or information. This presents, however some technical difficulties, so we shall first introduce a general technique to construct tilings, and then we shall apply it to our information and resolution measures.

We start with the definition of the canonical reciprocal cell associated to a resolution measure u and an aliasing grid Γ^* .

Definition 5 (Canonical Tiling) Given Γ^* a regular grid in \mathbb{R}^d , and a function $u : \mathbb{R}^d \rightarrow \mathbb{R}$, the set

$$R_{\max} = \left\{ \xi \in \mathbb{R}^d : u(\xi) > u(\xi + \gamma) \text{ for all } \gamma \in \Gamma^*, \gamma \neq 0 \right\} \quad (45)$$

is called canonical tiling associated to the function and the grid.

Since we are selecting among all possible aliases of a frequency ξ , the one which gives the highest resolution $u(\xi + \gamma)$, it is quite clear that if R_{\max} is a tiling, then it maximizes $\int_R u(x) dx$ with respect to all packings R .

It is also clear that R_{\max} is a packing because we only select one point within the set $\{\xi + \gamma\}_{\gamma \in \Gamma^*}$.

Showing that the ‘‘canonical tiling’’ is actually a tiling and the only one which maximizes $\int_R u(x) dx$ is a little bit more technical, and requires some further properties that we state in the following definition.

Definition 6 Given a regular grid $\Gamma^* \subset \mathbb{R}^d$, a function $u : \mathbb{R}^d \rightarrow \mathbb{R}$ is called an admissible alias selector if it satisfies the following conditions

(A1) The associated canonical tiling R_{\max} is bounded.

(A2) For almost every $\xi \in \mathbb{R}^d$, there is a single $\gamma_0 \in \Gamma^*$ such that $u(\xi + \gamma_0) = \sup_{\gamma \in \Gamma^*} u(\xi + \gamma)$.

The purpose of condition (A1) is to provide reciprocal cells $\overline{R_{\max}}$ with compact support which allow us to apply Shannon’s sampling theorem. The purpose of condition (A2) is to avoid some situations that might lead R_{\max} to be a sub-tiling. In the sequel we will use the following (equivalent but more convenient) formulation for condition (A2):

(A2’) The associated anomaly set A_{u, Γ^*} defined below has zero measure.

$$A_{u, \Gamma^*} = \left\{ \xi \in \mathbb{R}^d : \text{card}(G_{\xi, u, \Gamma^*}) \neq 1 \right\}$$

$$\text{where } G_{\xi, u, \Gamma^*} = \left\{ \gamma_0 \in \Gamma^* : u(\xi + \gamma_0) = \sup_{\gamma \in \Gamma^*} u(\xi + \gamma) \right\}.$$

The examples below illustrate how we might fail to meet this condition.

Example. If u is Γ -periodic then R_{\max} is clearly empty. But an admissible alias selector cannot be Γ -periodic, since periodicity would imply that $G_{\xi, u, \Gamma^*} = \xi + \Gamma^*$ for any ξ . Thus the anomaly set would be all of \mathbb{R}^d .

Example. If $|u(x)|$ monotonically increases to u_{\max} when $|x| \rightarrow \infty$ without ever attaining this value, then R_{\max} is also empty. But an admissible alias selector will never show such a behaviour, since it would imply that $G_{\xi, u, \Gamma^*} = \phi$ for any ξ . Thus the anomaly set would be all of \mathbb{R}^d also in this case but for a different reason.

Appendix C gives some simpler sufficient conditions for a function u to be an admissible alias selector. However, the previous examples cover almost all anomalous cases.

Proposition 1 *Let Γ^* be a regular grid in \mathbb{R}^d . For any function $u : \mathbb{R}^d \rightarrow \mathbb{R}$, the associated canonical tiling R_{\max} is such that (Γ^*, R_{\max}) is a packing of \mathbb{R}^d .*

If in addition u is an admissible alias selector then (Γ^, R_{\max}) is also a covering and hence a tiling.*

Finally, such a tiling is the unique⁶ solution to the following optimization problem:

$$R_{\max} = \arg \sup_{\{R : (\Gamma^*, R) \text{ is a packing}\}} \int_R v(u(x)) dx \quad (46)$$

for any non-negative, strictly increasing function $v : [u_{\min}, \infty) \rightarrow [0, \infty)$ where $u_{\min} = \inf_x u(x)$.

Observe that if we do not require uniqueness of R in the sense of maximization of $\int_R v(u(x)) dx$ we can avoid the technical hypothesis that u is an admissible alias selector by modifying the definition of R_{\max} as follows:

$$R_{\max} = \{ \xi \in \mathbb{R}^d : \text{for all } \gamma \in \Gamma^*, \gamma \neq 0 \implies \begin{aligned} &u(\xi) > u(\xi + \gamma) \text{ or} \\ &u(\xi) = u(\xi + \gamma) \text{ and } \xi \prec (\xi + \gamma) \} \end{aligned}$$

where \prec is any arbitrary order in \mathbb{R}^d (e.g. compare first the magnitude, then the angle to the first axis, etc.). With this definition proposition 1 still holds (except for the uniqueness of R_{\max}), and we do not need the hypothesis that u be an admissible alias selector. Then instead of uniqueness, we resolved the ambiguity in the maximization by introducing an arbitrary ordering of the frequency space. Here we judged uniqueness as an important property, and it is easy to show that the functions we considered are admissible alias selectors. For this reason we keep the original definition of R_{\max} .

Proof To prove that (Γ^*, R_{\max}) is a packing, consider $x \in R_{\max}$. Then for all $\gamma \in \Gamma^* \setminus \{0\}$ we have $u(x) > u(x + \gamma)$. This means that $(x + \gamma) \notin R_{\max}$, so $x \notin (R_{\max} + \gamma)$, and we conclude that (Γ^*, R_{\max}) is a packing.

To prove that (Γ^*, R_{\max}) is a covering consider any $x \in \mathbb{R}^d \setminus A_{u, \Gamma^*}$. Because $x \notin A_{u, \Gamma^*}$, we have that G_{x, u, Γ^*} consists of a single element γ_x and $(x + \gamma_x) \in R_{\max}$. Since A_{u, Γ^*} has zero measure this means that for almost every point $x \in \mathbb{R}^d$, there is $\gamma_x \in \Gamma^*$ such that $x \in (R_{\max} + \gamma)$, so (Γ^*, R_{\max}) is a covering.

To show the last part of the proposition, observe first that for all $x \in R_{\max}$, $v(u(x)) > 0$. In fact, since v is increasing, $v(u(x)) = 0$ would mean $u(x) = u_{\min}$, and consequently $u(x) \leq u(y)$ for all y , and in particular for $y = x + \gamma$, with $\gamma \in \Gamma^*$. Hence $v(u(x)) = 0$ implies $x \notin R_{\max}$, which proves that for $x \in R_{\max}$, $v(u(x)) > 0$.

Next consider any packing (Γ^*, R) , and any point $x \in R \setminus R_{\max}$. Since (Γ^*, R_{\max}) is a tiling there is some $\gamma_x \in \Gamma^*$ such that $(x + \gamma_x) \in R_{\max}$, except if $x \in N_1$ where N_1 has zero measure. Note that $\gamma_x \neq 0$ (otherwise $x \in R_{\max}$ would contradict our initial assumption) and the packing condition of R implies that $(x + \gamma_x) \notin R$ except for $x \in N_1 \cup N_2$, where N_2 has zero measure. If we call τ the mapping $x \mapsto (x + \gamma_x)$ then τ maps $(R \setminus R_{\max}) \setminus (N_1 \cup N_2)$ (which we shall abbreviate as R_1) into a subset of $(R_{\max} \setminus R)$. Now from the definition of R_{\max} we deduce that $u(\tau(x)) > u(x)$, and since v is increasing

⁶ up to a set of zero measure

$v(u(\tau(x))) > v(u(x))$. Therefore

$$\int_R v(u(x))dx - \int_{R_{\max}} v(u(x))dx = \int_{R \setminus R_{\max}} v(u(x))dx - \int_{R_{\max} \setminus R} v(u(x))dx \quad (47)$$

$$= \int_{(R \setminus R_{\max}) \setminus (N_1 \cup N_2)} v(u(x))dx - \int_{R_{\max} \setminus R} v(u(x))dx \quad (48)$$

$$\leq \int_{\tau((R \setminus R_{\max}) \setminus (N_1 \cup N_2))} v(u(y))dy - \int_{R_{\max} \setminus R} v(u(x))dx \quad (49)$$

$$\leq 0. \quad (50)$$

Since in equation (49) the integration domain $\tau(R_1) \subseteq R_{\max}$, and we have $v(u(y)) > 0$, so the first inequality is tight iff $\tau(R_1)$ has zero measure. From the construction of τ and the fact that N_1 and N_2 have zero measure, we conclude that $R \setminus R_{\max}$ has zero measure.

Then the second inequality is tight only if $((R_{\max} \setminus R) \setminus \tau(R_1))$ has zero measure. Using the previous condition this means that $(R_{\max} \setminus R)$ has zero measure.

So R is optimal iff it coincides with R_{\max} at almost every point. \square

The easiest way to define a tiling of the spectral domain, which produces little aliasing, is the following. Consider $u(\xi) = \text{Var}\{G(\xi)\}$ and define

$$D_{\text{tiling}}^* = \{\xi : \text{Var}\{G(\xi)\} > \text{Var}\{G(\xi + \gamma^*)\} \text{ for all } \gamma^* \in \Gamma^*, \gamma^* \neq 0\}. \quad (51)$$

According to the previous proposition then, $(\Gamma^*, D_{\text{tiling}}^*)$ is a tiling of the spectral domain. In addition, by its construction, it is a tiling that concentrates as much of the energy of the original image G as possible. In fact it maximizes $\int_{D^*} \text{Var}\{G(\xi)\}d\xi$ among all packings (Γ^*, D^*) . This tiling will be useful in the sequel, when we consider the resolution-optimizing reciprocal cell.

Similarly, we can apply the proposition to $u(\xi) = \frac{1+a^2k^2+ak}{a^2(1-k^2)+b^2}(\xi)$ and $v(u) = \log_2(1+u)$ to define the information maximizing reciprocal cell as:

$$D_{\text{I-opt}}^* := \left\{ \xi : \frac{1+a^2k^2+ak}{a^2(1-k^2)+b^2}(\xi) > \frac{1+a^2k^2+ak}{a^2(1-k^2)+b^2}(\xi + \gamma^*) \text{ for all } \gamma^* \in \Gamma^*, \gamma^* \neq 0 \right\}. \quad (52)$$

This reciprocal cell is in fact, according to proposition 1, a tiling that maximizes the effective information density (for a given sampling grid, image and noise model), among all packings of the spectral domain:

$$D_{\text{I-opt}}^* = \arg \sup_{\{D^*: (\Gamma^*, D^*) \text{ is a packing}\}} I_{\text{eff}}(\Gamma, D^*, H, N, F, k). \quad (53)$$

Its definition differs from D_{tiling}^* , even in the Gaussian independent case where $k = 0$. Nevertheless, our experiments show that in practice both criteria produce almost identical reciprocal cells (see figures 14 to 17).

We could define the resolution-optimal reciprocal cell in an similar way, but then we would not obtain a unique solution. This is due to the fact that in our effective resolution measure we included a thresholding operation that may prevent the support of $v \circ u$ to form a covering with Γ . Hence we can no longer apply proposition 1 to $u(\xi) = W(a(\xi), b(\xi))$, $v = Id$.

However we can consider a limiting case:

$$D_{\text{R-opt}}^* := \{\xi : a(\xi) < \theta_{\text{alias}} \text{ and } b(\xi) < \theta_{\text{noise}}\} \quad (54)$$

and characterize all resolution-optimizing reciprocal cells in terms of $D_{\text{R-opt}}^*$ and D_{tiling}^* , as follows:

Proposition 2 Consider W in equations (30-31) and assume that $\theta_{\text{alias}} \leq 1$. Then the maximum of

$$\int_{D^*} W(a(\xi), b(\xi)) d\xi$$

among all packings (Γ^*, D^*) , is attained whenever $D_{\text{R-opt}}^* \subseteq D^* \subseteq D_{\text{tiling}}^*$. (Inclusions are in the weak, “almost everywhere” sense, i.e. up to a set of zero measure).

Proof First we show that $D_{\text{R-opt}}^* \subseteq D_{\text{tiling}}^*$. In fact $\xi \in D_{\text{R-opt}}^*$ iff:

$$\begin{aligned} & \theta_{\text{alias}} > a(\xi) \text{ and } \theta_{\text{noise}} > b(\xi) \\ \Rightarrow & \text{Var}\{G(\xi)\} > \theta_{\text{alias}}^{-2} \text{Var}\{G_{\text{alias}}(\xi)\} \geq \text{Var}\{G_{\text{alias}}(\xi)\} \\ & \geq \text{Var}\{G(\xi + \gamma^*)\} \text{ for all } \gamma^* \in \Gamma^*, \gamma^* \neq 0 \\ \Rightarrow & \xi \in D_{\text{tiling}}^*. \end{aligned} \tag{55}$$

Now observe that $D_{\text{R-opt}}^* = \text{supp}(W(a(\xi), b(\xi)))$. This means that the (non-constrained) maximum of $\int_{D^*} W(a(\xi), b(\xi)) d\xi$ is attained for any $D^* \supseteq D_{\text{R-opt}}^*$. If in addition $D^* \subseteq D_{\text{tiling}}^*$ then (Γ^*, D^*) is a packing and D^* solves the (constrained) maximization problem. \square

From the proof it is clear that $(\Gamma^*, D_{\text{R-opt}}^*)$ is a packing, and it becomes a tiling only if $\theta_{\text{alias}} = 1$, θ_{noise} is sufficiently large and the maximal alias $\text{Var}\{G(\xi + \gamma_{\text{max}}^*)\}$ is larger than the squared sum of the remaining aliases $\sum_{\gamma^* \in \Gamma^* \setminus \{0, \gamma_{\text{max}}^*\}} \text{Var}\{G(\xi + \gamma^*)\}$. Thus in general it will not be a tiling and all reciprocal cells D^* such that $D_{\text{R-opt}}^* \subseteq D^* \subseteq D_{\text{tiling}}^*$ will maximize r_{eff} among all possible packings. Thus we shall keep the minimal resolution-optimizing reciprocal cell as a reference.

Finally let’s observe that if both the image model F and the transfer function H are isotropic, then G is also isotropic and it is easy to verify that in this case both the information-maximizing reciprocal cell $D_{\text{I-opt}}^*$ and D_{tiling}^* coincide with the Voronoi reciprocal cell D_{Vor}^* .

6 Examples and Applications

Figures 14, 15, 16 and 17 show the optimal reciprocal cells that were obtained for the different sampling systems that were presented in section 3. In all simulations we assumed a signal to noise ratio of 24 dB, *i.e.* the standard deviation of the noise is $\sigma = \|f\|/256$. Both the first and second column show the boundary of D_{tiling}^* (in black) and the boundary of $D_{\text{R-opt}}^*$ (in white) for $\theta_{\text{alias}} = \theta_{\text{noise}} = 0$. The difference between these two columns is the background information that is showed: In the first column the gray-level represents the relative aliasing in log-scale $\log_{10} a(\xi)$, whereas the first column shows the relative noise in log scale $\log_{10} a(\xi)$.

Finally the third column shows the information-optimal reciprocal cell $D_{\text{I-opt}}^*$, in the case where the Fourier coefficients of the image $\hat{f}(\xi)$ are modeled as independent, zero-mean Gaussian variables. Even though the definition of $D_{\text{I-opt}}^*$ is not exactly the same as that of $D_{\text{R-opt}}^*$, the resulting regions are hardly distinguishable in all cases.

Observe how, even in a simple situation like ccd3 that may arise even with a digital camera, the optimal reciprocal cell might consist of more than one connected component.

System	R_{nom} (c)	R_{eff} (c)	I_{nom} (bits/ c^2)	I_{eff} (bits/ c^2)	A (%)	B (%)	$\frac{R_{\text{nom}}}{R_{\text{eff}}}$ (%)	$I_{\text{eff}} R_{\text{nom}}^2$ (bits/sample)
ccd1	1.0000	1.2601	8.18	2.45	32.98	0.38	79.36	2.45
ccd2	1.0000	1.1703	6.86	3.48	23.58	1.23	85.44	3.48
ccd3	1.0000	1.1525	5.63	3.95	23.14	5.72	86.77	3.95
Sq0	1.0000	1.1463	6.66	3.91	20.81	1.57	87.24	3.91
Hex0	0.9306	1.0380	7.44	5.14	18.84	1.97	89.65	4.45
Supermode	0.7071	0.7843	9.51	8.76	11.91	9.78	90.16	4.38
Sq1	0.7071	0.7761	9.99	8.91	13.68	8.68	91.11	4.46
Hex1	0.5373	0.6716	11.40	11.35	3.81	18.37	80.00	3.28
Hipermode	0.5000	0.7093	10.30	10.27	2.42	18.31	70.50	2.57
Sq2	0.4472	0.6653	11.68	11.67	1.67	18.37	67.22	2.33
Hex2	0.3517	0.6626	11.82	11.82	0.00	18.24	53.09	1.46

Table 2: Quantitative evaluation of the sampling system $(H, n, \Gamma, D_{\text{R-opt}}^*)$ with $\theta_{\text{alias}} = \theta_{\text{noise}} = 1$, and for $(H, n, \Gamma, D_{\text{I-opt}}^*)$ with independent Gaussian model for the image Fourier coefficients, and signal to noise ratio of 24 dB.

Once we obtained the optimal reciprocal cells we can compute the corresponding effective resolution and information density measures. These are displayed in table 2, along with some further information which gives some indication of sampling efficiency and quality of the different systems.

The first two columns show the spatial nominal and effective resolutions R_{nom} and R_{eff} , both expressed as multiples of the size c of the sensors. Observe that at equal sampling rate (equal R_{nom}), the first four systems have quite different effective resolution. In particular, the case of the CCD array which is blurred due to a moving camera (ccd3) performs much better than the ideal CCD array (ccd1) where the only low-pass filter is performed by the sensors themselves. This is due to the high amount of aliasing present in ccd1, which is not that much the case in ccd3. This can be observed in columns 4 and 5 which show the average amount of relative aliasing and noise: $A = \int_{D^*} a(\xi) d\xi / |D^*|$ and $B = \int_{D^*} b(\xi) d\xi / |D^*|$ respectively. The sampling efficiency in terms of resolution is measured by the ratio $R_{\text{nom}}/R_{\text{eff}}$ which expresses how much of the nominal resolution (sampling rate) is effectively attained by this system.

Column four shows the information density measure I_{eff} , expressed in terms of bits of mutual information per unit area, where the unit area c^2 is the surface of a single sensor. This is to be compared with the nominal information density I_{nom} as defined in equation (44). Note that unlike nominal resolution I_{nom} takes into account the amount of noise and the decay rate of Fourier coefficients in the image. We can also compare I_{eff} to the information density one would obtain in an even more

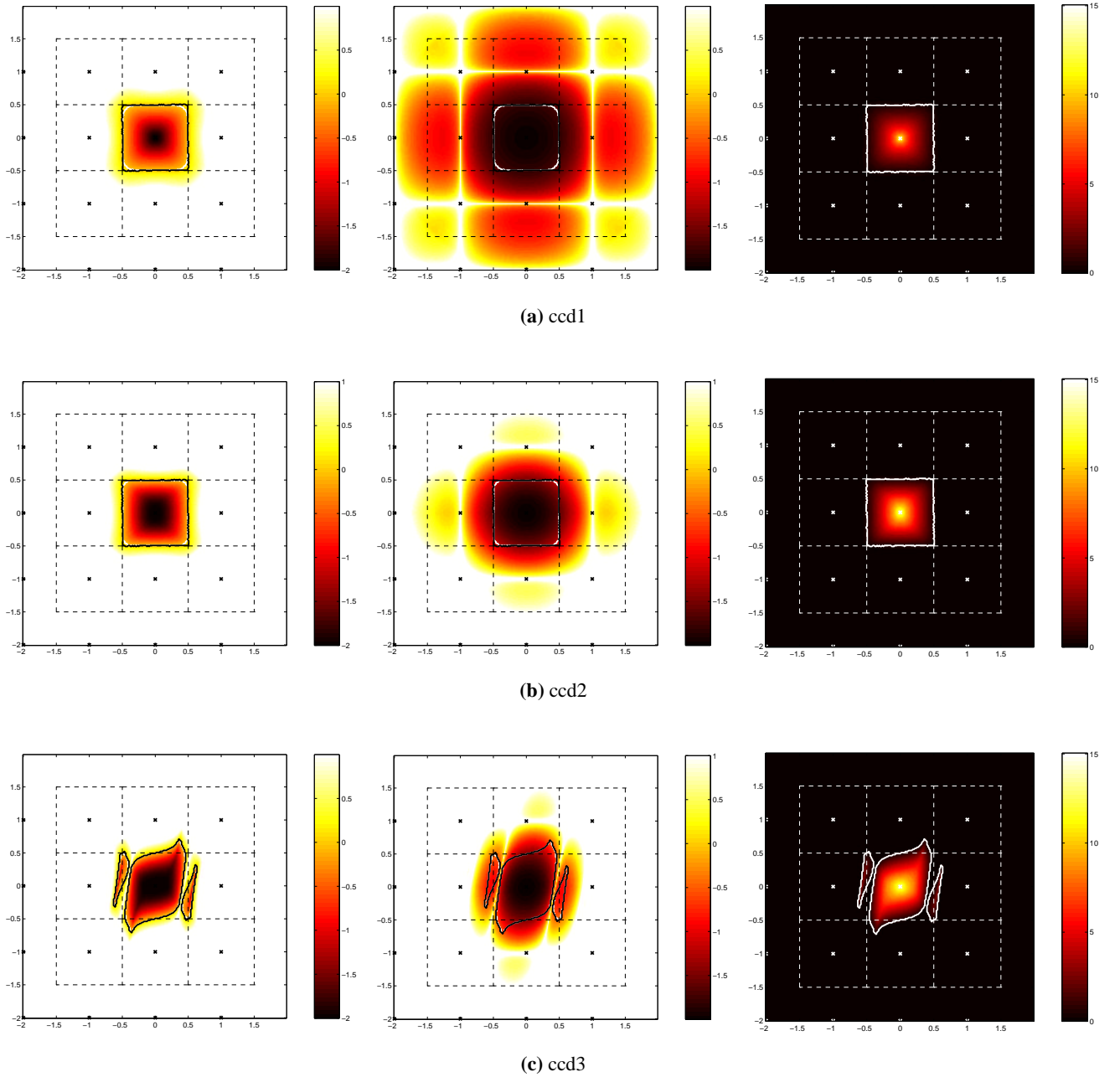


Figure 14: Optimal reciprocal cells for the three CCD sampling systems in figure 9.

Legend: In all figures above, the dashed lines represent the boundaries of the Voronoi reciprocal cell and its aliases. The crosses represent the points in the dual (“aliasing”) grid Γ^* . The remaining curves are the level-lines of the MTF.

In the first and second columns the thick black curve represents the boundary of the optimal tiling D_{tiling}^* , and the thick white line represents the boundary of the resolution-optimal reciprocal cell $D_{\text{R-opt}}^*$. In the last column the thick white line represents the boundary of the information-optimal reciprocal cell $D_{\text{I-opt}}^*$.

The gray-scale represents: in the first column the value of the relative aliasing $\log_{10} a(\xi)$ in logarithmic scale; in the second column the value of the relative noise $\log_{10} b(\xi)$ also in logarithmic scale; in the third column the value of the mutual information density $\log_2(1 + \frac{1}{a^2 + b^2})(\xi)$ measured in bits/ c^2 .

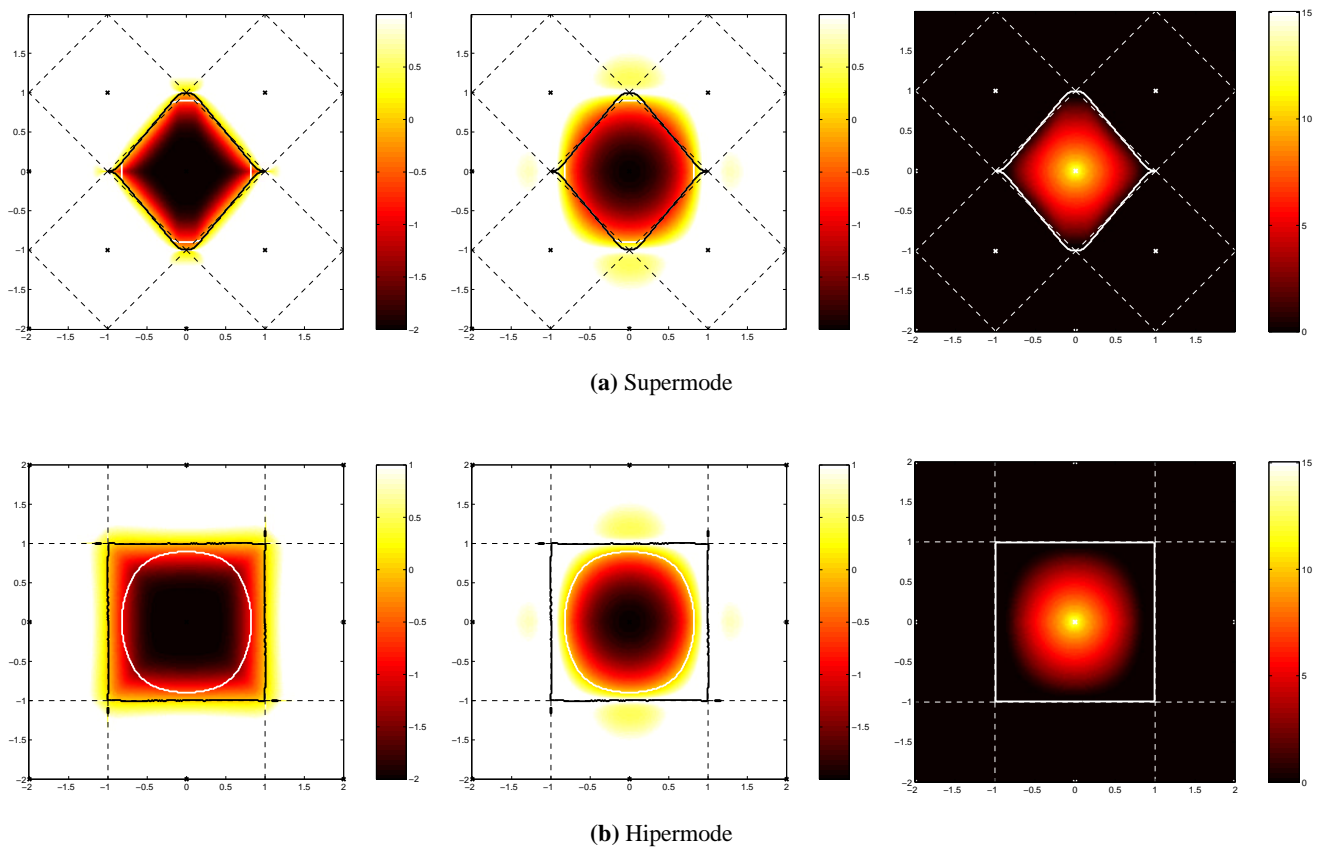


Figure 15: *Optimal reciprocal cells* for the sampling systems in figure 10. (See also the legend in figure 14.)

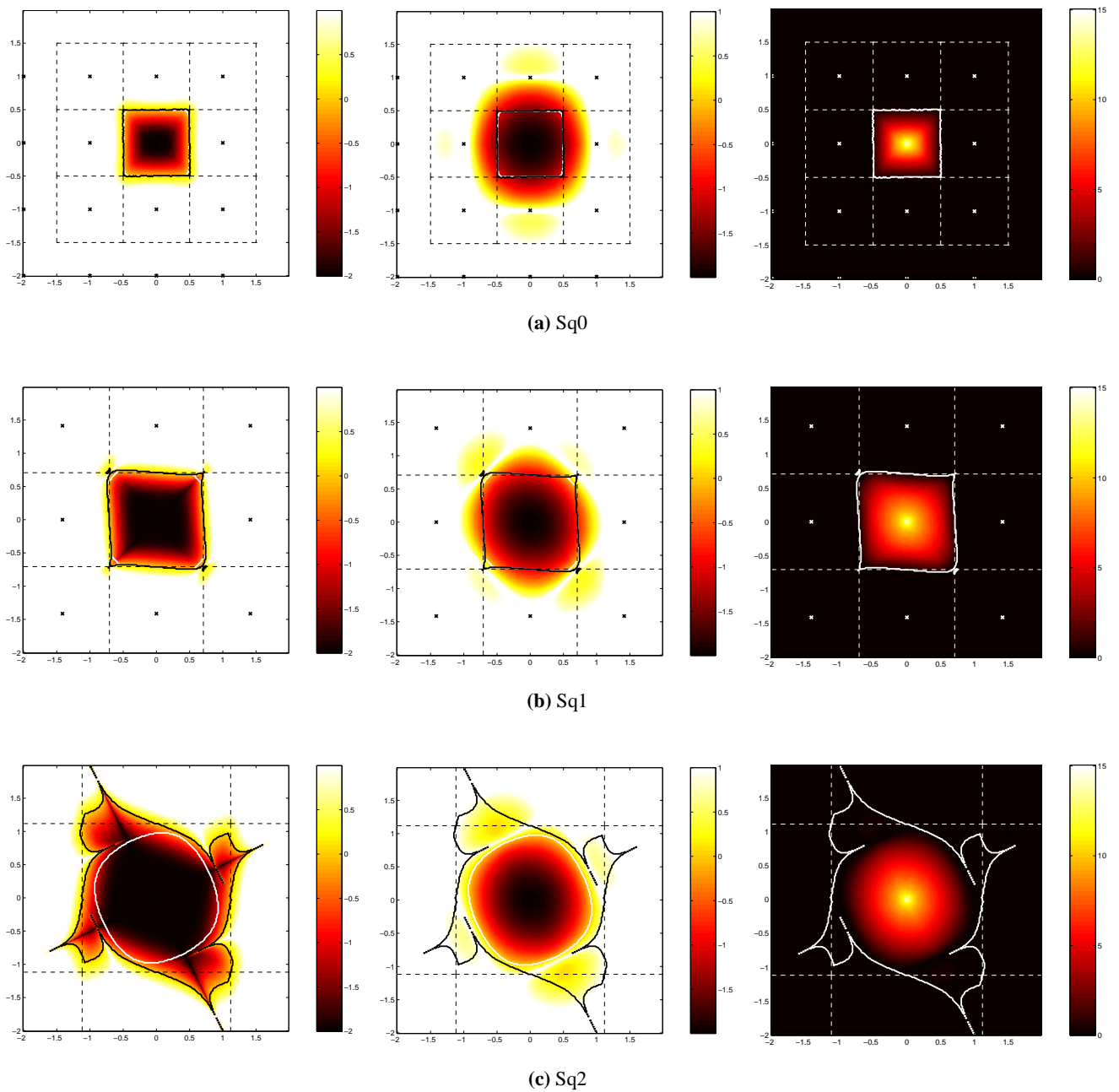


Figure 16: *Optimal reciprocal cells* for the three Square sampling systems in figure 11. (See also the legend in figure 14.)

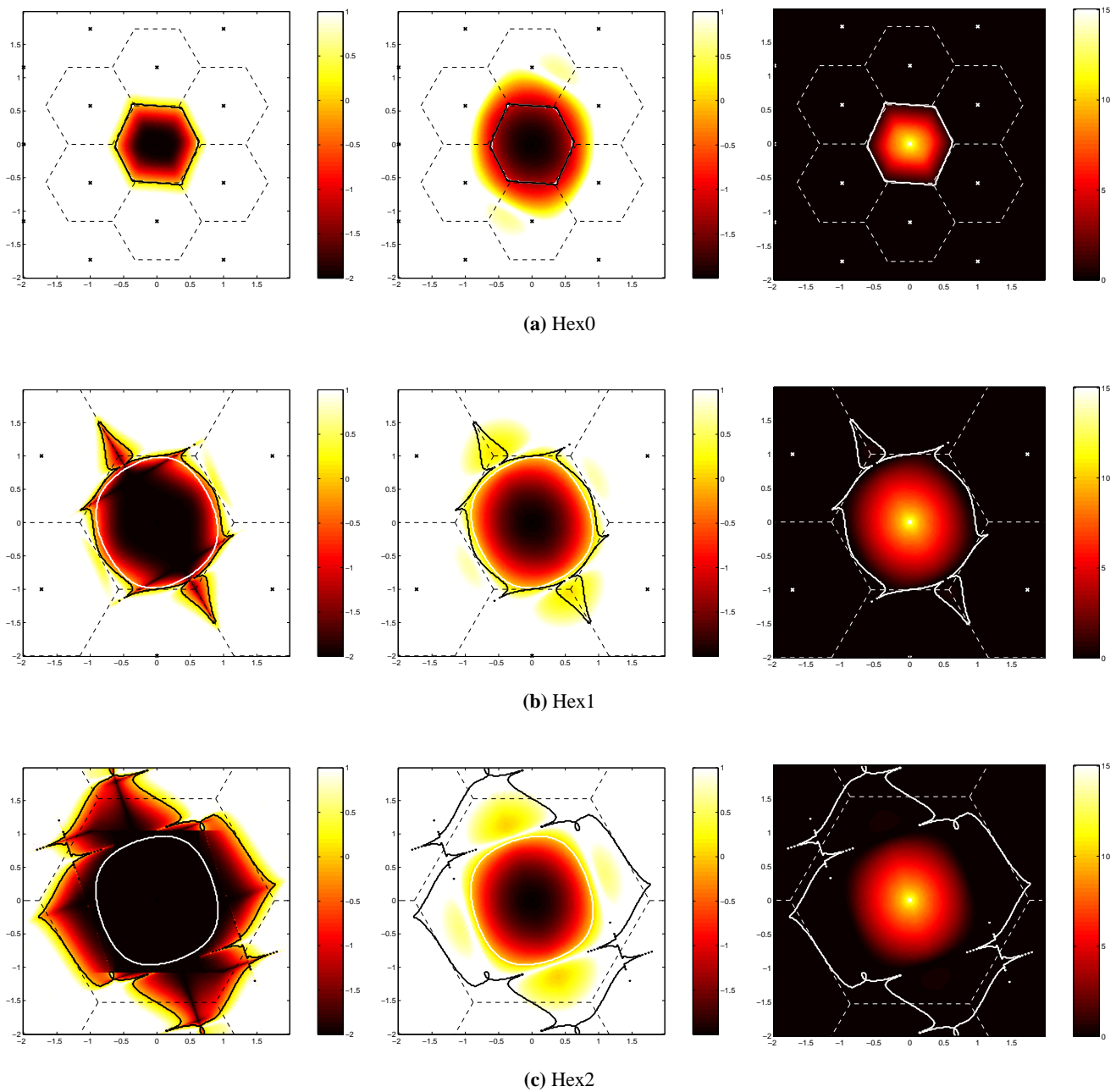


Figure 17: *Optimal reciprocal cells* for the three hexagonal sampling systems in figure 12. (See also the legend in figure 14.)

ideal situation where we do not only have an ideal transfer function $H = \mathbb{1}_{D_{\text{Vor}}^*}$, but also an ideal decay rate of Fourier coefficients, *i.e.* $F(\xi) = \text{constant}$. In such an ideal case the information density would be $I_{\text{ideal}} = \frac{|RC|}{(2\pi)^2} \frac{1}{2} \log_2(1 + \text{SNR})$. Since in our case $\text{SNR} = 24$ dB, we would have $I_{\text{ideal}} = \frac{|RC|}{(2\pi)^2} 8$ bits = $R_{\text{nom}}^2 8$ bits, and we would at most have 8 bits of mutual information per sample. Hence another way of representing the efficiency of sampling is the product $I_{\text{eff}} R_{\text{nom}}^2$ (last column) which represents the number of bits of mutual information per sample and is bounded by 8. Observe that both the effective resolution and information measures yield Supermode and Sq1 as the most effective sampling systems. Resolution may be further increased, but the cost in terms of sampling rate per increase in resolution will start growing faster beyond this point.

Finally we show an unexpected application of these results. We started with the definition of a resolution measure. To do so, we realized that we needed to adapt the reciprocal cell. Otherwise, our resolution measure would be under-estimated. Now, once we found the optimal reciprocal cell, we should use it for restoration as well. The procedure is as follows:

From the discrete image g obtained by equation (16) we compute the discrete Fourier transform \hat{g} . In order to remove aliasing from this image without losing any good information, we need to oversample. We do so either by adding zeros between the samples g before the Fourier transform, or equivalently by periodically extending \hat{g} . Then we multiply \hat{g} by $\mathbb{1}_{D_{\text{R-opt}}^*}$ and take the inverse Fourier transform to obtain a zoomed version of g with some aliased or noisy coefficients removed and some aliased coefficients re-placed at their correct locations.

This actually improves the image but in most cases the difference is hardly perceptible. In order to observe the full advantage of this procedure we need to deconvolve the transfer function. To do so we applied Total Variation minimization, which is specially well suited for this kind of problems where the spectrum is known on a region which consists of several connected components (Malgouyres, 1999; Malgouyres and Guichard, 2001). In fact this is a good way of choosing the unknown Fourier coefficients in a way that minimizes the ringing artifacts that would otherwise be created by the gaps between two connected components.

Specifically, the usual method of deconvolution and spectral extrapolation by TV minimization consist of finding u , such that it minimizes

$$\int |\nabla u(x)| dx + \lambda \int (h * u - g)^2 dx. \quad (56)$$

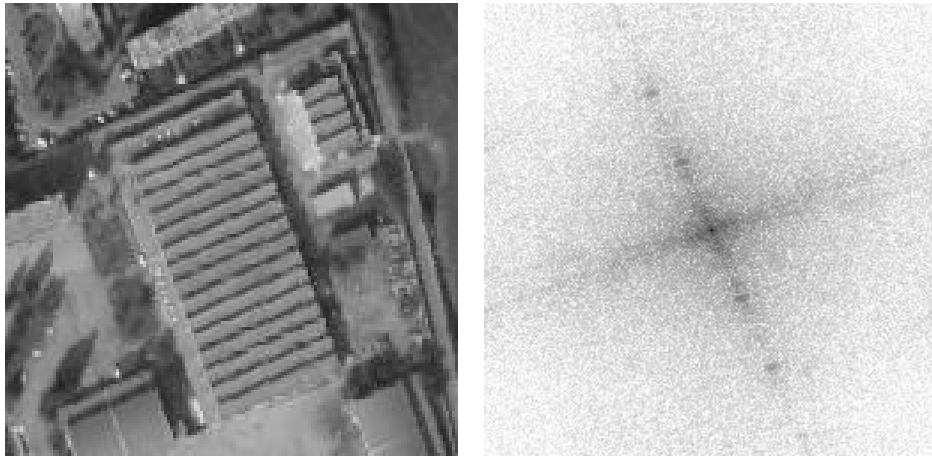
If we rewrite the data-fitting term in the Fourier domain we obtain

$$u = \arg \inf_u \int |\nabla u(x)| dx + \lambda \int_{D_{\text{Vor}}^*} (H\hat{u} - \hat{g})^2 d\xi. \quad (57)$$

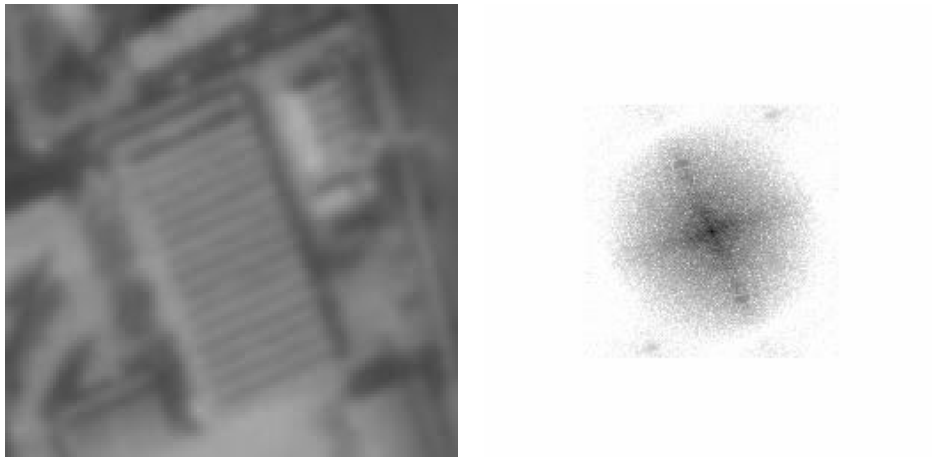
The modification we propose is to integrate the data-fitting term over the $D_{\text{R-opt}}^*$ (after periodic extension of \hat{g}) instead of over D_{Vor}^* (which would be equivalent to integrate it in the spatial domain as in equation (56)):

$$u = \arg \inf_u \int |\nabla u(x)| dx + \lambda \int_{D_{\text{R-opt}}^*} (H\hat{u} - \hat{g})^2 d\xi. \quad (58)$$

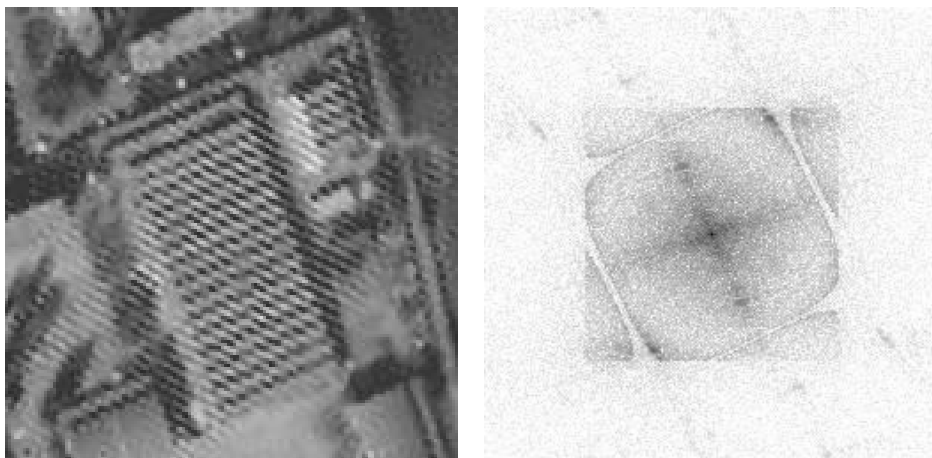
Figure 19 shows the result of this operation on a sample image. Note how the proposed method produces a correct deconvolution, and even allows to extrapolate the spectrum, whereas the naïve application of equation (56) produces a catastrophic result. The reason is that the high-energy second harmonic of the roof texture which falls outside of D_{Vor}^* , say at $\xi \notin D_{\text{Vor}}^*$ is translated (due to sampling and aliasing) to a region very near one of the zero-crossings of H , say at $\xi + \gamma^* \in D_{\text{Vor}}^*$. This means that $\hat{g}(\xi + \gamma^*) \approx H(\xi)\hat{f}(\xi)$ since $H(\xi + \gamma^*) \approx 0$. Hence equation (56), which doesn't know about aliasing, tries to deconvolve the high-energy aliased coefficient $\hat{g}(\xi + \gamma^*)$ with $H(\xi + \gamma^*)$ which is almost zero,



(a) Original image and its Fourier spectrum



(b) Sampled image zoomed by zero-padding



(c) Deconvolved by Total Variation

Figure 18: Sampling with the Sq2 system and deconvolution by Total Variation minimization. Observe how the aliased second harmonic of the roof's texture explodes when deconvolved by a wrong Fourier coefficient of the transfer function.

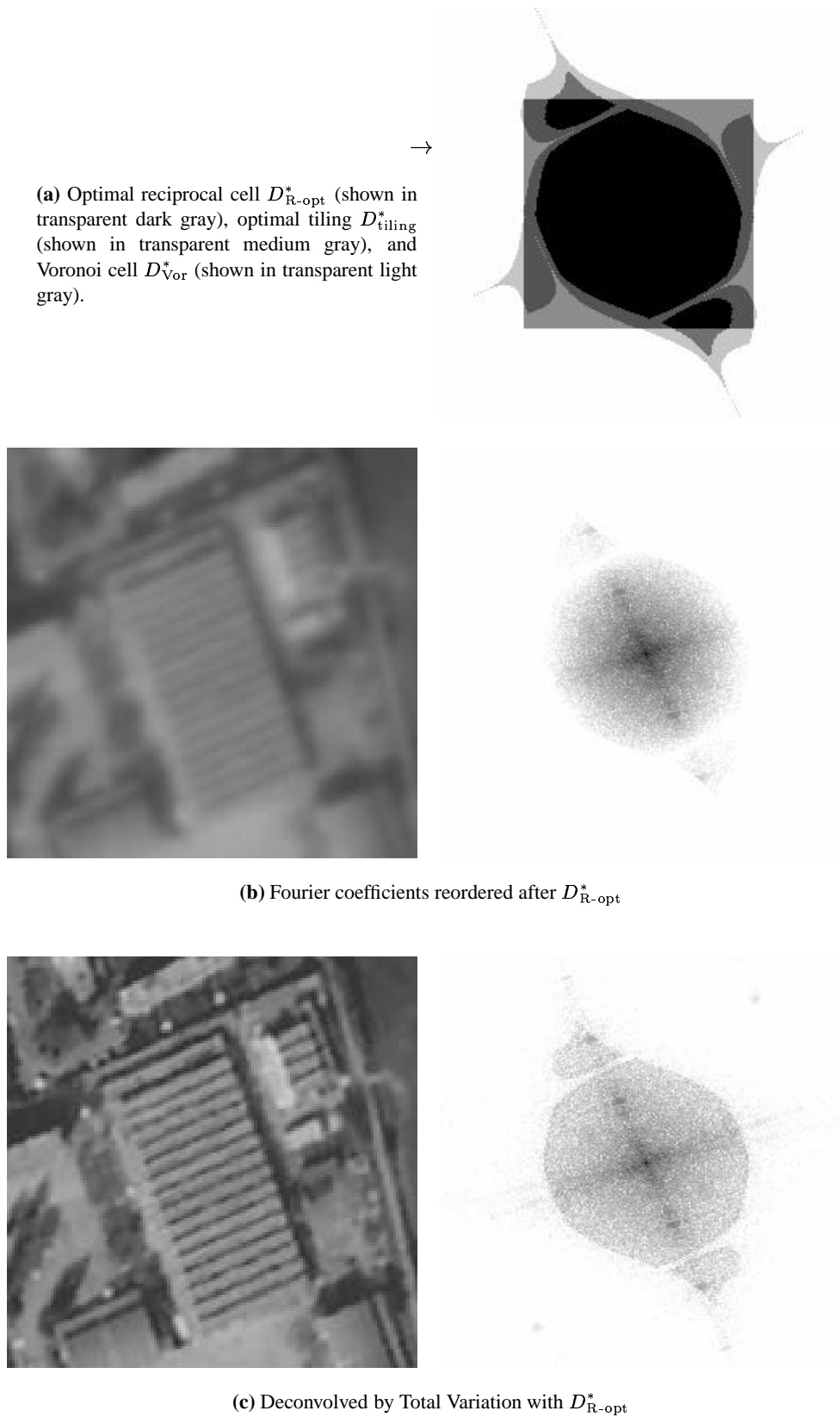
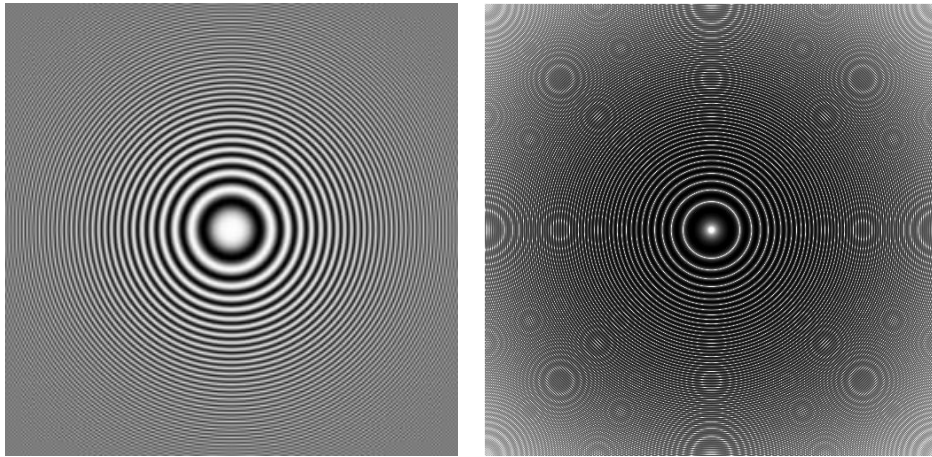


Figure 19: Deconvolution of the image in figure 18 by Total Variation minimization with the data-fitting term integrated over the optimal reciprocal cell D_{R-opt}^* . Observe how the errors in the previous figure are reduced by placing aliasing-dominated regions at their correct locations in the Fourier spectrum.



(a) Original image and its Fourier spectrum

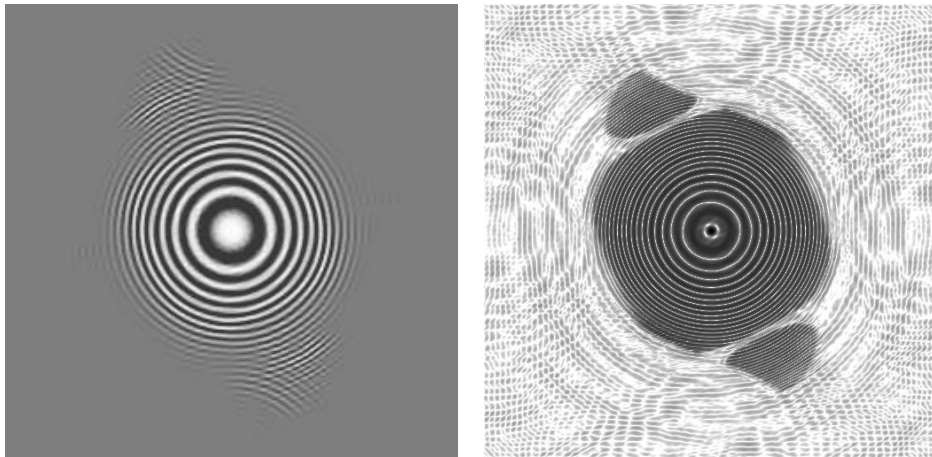
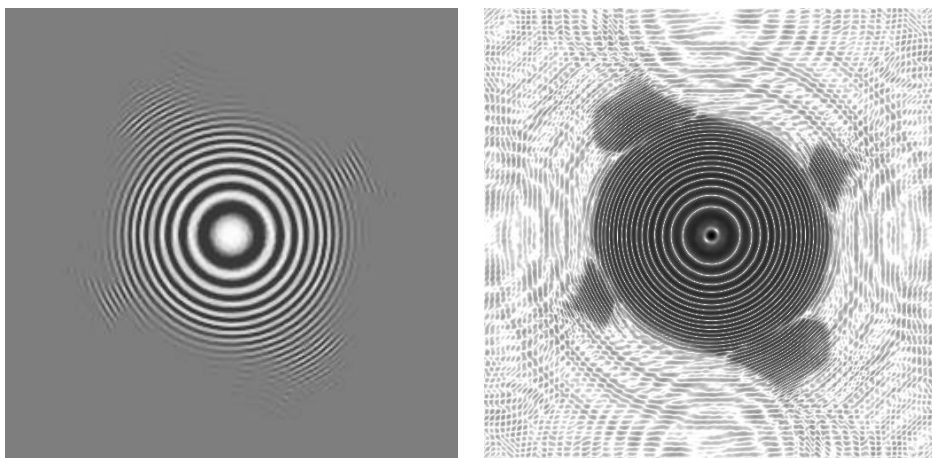
(b) Deconvolved by Total Variation with D_{R-opt}^* , $\theta_{alias} = \theta_{noise} = 1$ (c) Deconvolved by Total Variation with D_{R-opt}^* , $\theta_{alias} = 0.3$ and $\theta_{noise} = 3$

Figure 20: Deconvolution by Total Variation minimization with the data-fitting term integrated over the optimal reciprocal cell D_{R-opt}^* , for two different values of the thresholds. Observe that to completely eliminate visible aliasing we need an aliasing threshold θ_{alias} smaller than one. On the other hand, deconvolved noise is less visible or better dealt with by TV and we can tolerate higher values of the threshold θ_{noise} without any serious visible artifacts.

producing the instability that we can observe in the result. Equation (58) instead, deconvolves $\hat{g}(\xi + \gamma^*)$ with $H(\xi)$, thus producing the correct result, closer to $\hat{f}(\xi)$.

This kind of error is not very common in more traditional systems where the transfer function is quite isotropic, and hence (according to the last observation of the previous section) the optimal reciprocal cell is included in it. Furthermore this effect can be reduced by adding a threshold, which avoids to deconvolve coefficients where H is below the threshold. But this is not a very efficient solution, because it consists of treating aliasing as if it was white noise, and for highly anisotropic transfer functions with many zero-crossings, it may force us to use a quite high threshold, thus forcing us to unnecessarily throw away a lot of information.

Finally we show an example which illustrates the different roles played by the thresholds θ_{alias} and θ_{noise} used in the resolution measure. Figure 20 shows a synthetic image which was chosen to better illustrate our results. It is a radial chirp image centered at x_0 , where the local frequency at a point x is around $x - x_0$. Thus, The image and its spectrum have a very similar shape. We simulated a sampling of this image with the Sq2 system, and we applied the Equation (58) with two different values of R-opt. In the first case we set $\theta_{\text{alias}} = \theta_{\text{noise}} = 1$ as before, and in the second case we set $\theta_{\text{alias}} = 0.3$ whereas $\theta_{\text{noise}} = 3$, *i.e.* we tolerate higher distortions from noise than from aliasing. The result, as anticipated at the end of section 4.3, confirms that this is a sensible choice, because we perceive better distortions due to image-correlated aliasing than distortions due to image-independent noise. Another reason may be that TV too may do a better job at filtering white noise than aliasing errors.

7 Discussion and future work

In this work we propose a new way of measuring the effective resolution of an image acquisition system, which takes into account the noise level, and the geometric distribution of sensors and their transfer functions. This definition is motivated by modern imaging systems, which may consist of hexagonal or square grids of sensors, and may show a better resolution in certain preferred directions.

Our experiments suggest that the proposed measure can not only be used to compare the resolution of different imaging systems, but also to avoid aliasing artifacts that may cause serious problems in restoration. When this restoration is done in conjunction with total variation minimization in order to avoid ringing by fill-in the spectrum between connected components of the optimal reciprocal cell, we actually extrapolate the spectrum in a reasonable manner with a visual effect of sharpness and increased resolution.

However this spectrum extrapolation should not be confused with other “super-resolution” methods like in (Freeman, Jones, and Pasztor, 2001; Romberg, Choi, and Baraniuk, 2001), where a very detailed knowledge of image statistics is used to produce plausible interpolations. In our work we used a minimal knowledge of image statistics, namely the decay rate of Fourier coefficients, but we used a very detailed knowledge of the image acquisition device, its geometry, transfer function and noise. On the other hand this kind of knowledge is usually not used in super-resolution algorithms, in any case not the aliasing part.

Thus there is a potential to integrate both approaches to obtain better super-resolution methods which exploit both detailed image statistics and image acquisition models. The mutual information density measure and the corresponding reciprocal cell provide a first indication on how these approaches could be integrated. We calculated this measure to include up to second order statistics (through the relative covariance k between pairs of Fourier coefficients), but this model can be refined as much as we want. Nevertheless, image statistics are more relevant when expressed in more local bases, such as wavelets or wavelet packets bases as in (Romberg et al., 2001), and this would oblige us to approximate the effect of the transfer function and aliasing in such bases. Such an integration of both approaches to super-resolution will be the subject of future research.

8 Acknowledgements

We would like to thank Francois Malgouyres, Vicent Caselles, and Andrés Solé for useful discussions on total variation restoration and for generously providing the sources of the corresponding software which allowed us to introduce the necessary modifications to test our method. We also had fruitful discussions with Lionel Moisan and Lenny Rudin on this subject, which is gratefully acknowledged. Finally, we would like to specially thank Jean-Michel Morel for his support and guidance throughout this project.

A Proof of Shannon's sampling theorem

In this section we give a more rigorous proof of Shannon's sampling theorem. The main elements of the theory and the proof were obtained from (Bony, 2001; Gasquet and Witomski, 1995; Morel and Ladjal, 1998). We rewrite the proofs here to unify the notation, and because we state the theorem under slightly different hypothesis which are more useful for the purposes of this article. In (Bony, 2001) for instance we can find all the elements of Fourier and distribution theory in multiple dimension, but not a proof of the sampling theorem. On the other hand (Gasquet and Witomski, 1995) gives a proof of the sampling theorem in L^2 but in dimension 1. Finally (Morel and Ladjal, 1998) constructs the theory of Fourier series in several dimensions and gives a proof of the sampling theorem under conditions very similar to 5.

In the next section we recall the basic elements of the theory that we will use in section A.2 to give the proof. Finally in section A.3 we discuss to what an extent the hypothesis we assumed on the continuous image are verified by real systems.

A.1 Basic results from Fourier analysis and distribution theory

The Fourier transform and its conjugate are defined initially for $f \in L^1(\mathbb{R}^d)$ as

$$\hat{f}(\xi) = \mathcal{F}(f)(\xi) := \int_{\mathbb{R}^2} e^{-i\langle x, \xi \rangle} f(x) dx \quad (59)$$

$$\overline{\mathcal{F}}(F)(x) := \int_{\mathbb{R}^2} e^{i\langle x, \xi \rangle} F(\xi) d\xi. \quad (60)$$

The Fourier transform maps L^1 to continuous functions which tend to 0 at infinity, but $\mathcal{F}(f)$ is not necessarily in L^1 . Whenever $\hat{f} \in L^1$ we have the inversion formula

$$\overline{\mathcal{F}}(\mathcal{F}(f)) = (2\pi)^d f, \text{ if } f \in L^1, \hat{f} \in L^1.$$

On the contrary the Schwartz class \mathcal{S} of functions with fast decay at infinity and an infinite number of continuous derivatives is closed under \mathcal{F} , so:

$$f \in \mathcal{S} \implies \mathcal{F}(f) \in \mathcal{S} \wedge \overline{\mathcal{F}}(\mathcal{F}(f)) = (2\pi)^d f.$$

The Schwartz class allows us to extend the definition of the Fourier transform to more general classes. First we can extend it to the class of tempered distributions \mathcal{S}' , so that for $u \in \mathcal{S}'$, the Fourier transform is $\hat{u} = \mathcal{F}(u)$ such that

$$\begin{aligned} \forall \varphi \in \mathcal{S}, \langle \mathcal{F}(u), \varphi \rangle &= \langle u, \mathcal{F}(\varphi) \rangle \\ \langle \overline{\mathcal{F}}(u), \varphi \rangle &= \langle u, \overline{\mathcal{F}}(\varphi) \rangle \end{aligned}$$

The definition is consistent with the definition in $L^1 \subset \mathcal{S}$, and it isometrically maps \mathcal{S}' into itself, and satisfies the Fourier inversion formula, but now in the sense of tempered distributions:

$$f \in \mathcal{S}' \implies \mathcal{F}(f) \in \mathcal{S}' \wedge \quad (61)$$

$$\overline{\mathcal{F}}(\mathcal{F}(f)) = (2\pi)^d f. \quad (62)$$

The extension of the Fourier transform to \mathcal{S}' provides an extension to $L^2(\mathbb{R}^d) \subset \mathcal{S}'(\mathbb{R}^d)$. Since \mathcal{S} is complete and dense in L^2 we can show that this is the only extension of $\mathcal{F} : \mathcal{S} \rightarrow \mathcal{S}$. We can check that

this extension maps L^2 to itself, that the map is continuous, that it preserves the L^2 norm up to a factor $(2\pi)^{\frac{d}{2}}$ and that it satisfies the Fourier inversion formula (in the sense of the L^2 norm):

$$f \in L^2 \implies \mathcal{F}(f) \in L^2 \wedge \tag{63}$$

$$\overline{\mathcal{F}(\mathcal{F}(f))} = (2\pi)^d f \wedge \tag{64}$$

$$\|\mathcal{F}(f)\|_{L^2} = (2\pi)^{\frac{d}{2}} \|f\|_{L^2}$$

The Fourier transform in L^2 is closely related to the more elementary notion of Fourier series of periodic functions. Any Γ -periodic function $f \in L^2_\Gamma$ admits a Fourier series development

$$f(x) = \sum_{k \in \Gamma^*} c_k e^{i\langle k, x \rangle} \tag{65}$$

where equality holds in the sense of the $L^2(D)$ norm and the Fourier coefficients are unique and defined by

$$c_k = \frac{1}{S} \int_D e^{-i\langle x, k \rangle} f(x) dx$$

for any tile D of Γ , where $S = |D|$ denotes the area of such any such tile. The link with the Fourier transform of distributions is given by:

$$\hat{u} = (2\pi)^d \sum_{k \in \Gamma^*} c_k(u) \delta_k$$

which is valid for any Γ -periodic function u .

Similarly, convolution is defined by the usual formula for $f, g \in L^1(\mathbb{R}^d)$

$$(f * g)(x) = \int_{\mathbb{R}^d} f(x - y)g(y)dy$$

and then extended in several ways. The main difficulty is that the decay rate at infinity of both terms need to compensate for the integral to converge. For our purposes it suffices to know that it is well defined in the following cases

$$\begin{aligned} \forall u \in \mathcal{E}', \forall \varphi \in C_0^\infty, (u * \varphi) \in C_0^\infty \text{ with} \\ \text{supp}(u * \varphi) = \text{supp}(u) + \text{supp}(\varphi) \end{aligned} \tag{66}$$

$$\begin{aligned} \forall u \in \mathcal{E}', \forall v \in \mathcal{S}', (u * v) \in \mathcal{S}' \\ \forall u \in \mathcal{O}'_C, \forall v \in \mathcal{S}', (u * v) \in \mathcal{S}' \\ \forall u \in \mathcal{S}, \forall v \in \mathcal{S}', (u * v) \in \mathcal{O}_M. \end{aligned}$$

Our notation follows (Bony, 2001): \mathcal{E}' denotes the compact support distributions, \mathcal{O}'_C the distributions with fast decay at infinity, and \mathcal{O}_M the functions having slow growth at infinity and an infinite number of continuous derivatives. Observe that we have the following inclusions relationships between these spaces:

$$\begin{array}{ccccccc} C_0^\infty & \longrightarrow & \mathcal{S} & \longrightarrow & \mathcal{O}_M & \longrightarrow & C^\infty \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ \mathcal{E}' & \longrightarrow & \mathcal{O}'_C & \longrightarrow & \mathcal{S}' & \longrightarrow & \mathcal{D}' \end{array} \tag{67}$$

In addition $\mathcal{S} \subset L^p \subset \mathcal{S}'$ for all $p \geq 1$. In all the previous cases where the convolution is well defined, its Fourier transform exchanges with convolution in the following manner:

if u and v satisfy one of the following conditions:

$$\begin{aligned} u &\in \mathcal{E}' \wedge v \in \mathcal{S}' \\ u &\in \mathcal{O}'_C \wedge v \in \mathcal{S}' \\ u &\in \mathcal{S} \wedge v \in \mathcal{S}' \\ \text{then } \mathcal{F}(u * v) &= \mathcal{F}(u) \cdot \mathcal{F}(v). \end{aligned} \quad (68)$$

which is also valid for $\overline{\mathcal{F}}$ instead of \mathcal{F} .

The multiplication of two distributions is not always well defined either. For our purposes it suffices to know that we can safely multiply a tempered distribution by a slowly growing regular function

$$f \in \mathcal{O}_M \wedge g \in \mathcal{S}' \implies f \cdot g \in \mathcal{S}' \quad (69)$$

and that the Fourier transform maps distributions with compact support into \mathcal{O}_M :

$$f \in \mathcal{E}' \implies \mathcal{F}(f) \in \mathcal{O}_M. \quad (70)$$

These two results allow to obtain the dual form of equation (68), namely:

if u and v satisfy one of the following conditions:

$$\begin{aligned} \hat{u} &\in \mathcal{E}' \wedge v \in \mathcal{S}' \\ u &\in \mathcal{O}_M \wedge \hat{u} \in \mathcal{O}'_C \wedge v \in \mathcal{S}' \\ u &\in \mathcal{S} \wedge v \in \mathcal{S}' \\ \text{then } \mathcal{F}(u \cdot v) &= \frac{1}{(2\pi)^d} \mathcal{F}(u) * \mathcal{F}(v). \end{aligned} \quad (71)$$

To conclude we state two more results that we will need in the following section. The first is valid for $f \in \mathcal{S}'$, $f \in \mathcal{S}$ or $f \in L^2$, with the corresponding meaning of equality in each case:

$$\mathcal{F}(f(x)e^{-i\langle x, \omega \rangle}) = \tau_\omega \hat{f} \quad (72)$$

$$\mathcal{F}(\Delta_\Gamma) = S^* \Delta_{\Gamma^*}. \quad (73)$$

In the last line $S^* = |D^*|$ denotes the area of any tile D^* of Γ^* .

A.2 Shannon's sampling theorem

The formal proof given in section 2.2 is actually only valid if we impose a strong regularity and fast decay on the analog image f :

Theorem 3 (Shannon-Whittaker in \mathcal{S}) *Given a function $f \in \mathcal{S}(\mathbb{R}^d)$ (with fast decay and an infinite number of continuous derivatives), a d -dimensional sampling grid Γ and a compact reciprocal cell $D^* \subset \mathbb{R}^d$. If the following conditions are met*

(S1) (D^*, Γ^*) is a packing

(S2) $\text{supp}(\hat{f}) \subset D^*$ (i.e. f is band-limited),

then f can be completely recovered from its samples in Γ , i.e. from

$$g = \{f_\gamma\}_{\gamma \in \Gamma}. \quad (74)$$

In fact the reconstruction is given by the following convolution

$$f = \frac{1}{S^*} \overline{\mathcal{F}}(\mathbb{1}_{D^*}) * g \quad (75)$$

of the sampled image g with the generalized sinc kernel $s = \frac{1}{S^*} \overline{\mathcal{F}}(\mathbb{1}_{D^*})$.

Proof Since $f \in \mathcal{S}$ (which is included in \mathcal{O}_M) and $\Delta_\Gamma \in \mathcal{S}'$ the product $g = \Delta_\Gamma \cdot f$ defining the sampled image g makes sense and is in \mathcal{S}' according to (69).

Now again, since f has compact spectral support we can apply (71) in the case when $u = f$ is such that $\hat{u} \in \mathcal{E}'$ and $v = \Delta_\Gamma \in \mathcal{S}'$ to obtain: $\mathcal{F}(f \cdot \Delta_\Gamma) = \frac{1}{(2\pi)^d} \mathcal{F}(f) * \mathcal{F}(\Delta_\Gamma)$. Finally applying (73) to compute the Fourier transform of the Dirac comb the previous equation becomes Poisson's formula:

$$\mathcal{F}(f \cdot \Delta_\Gamma) = \frac{S^*}{(2\pi)^d} \hat{f} * \Delta_{\Gamma^*},$$

where both members are tempered distributions, and equality is to be understood in this sense. Multiplying both sides by $\mathbb{1}_{D^*} \in \mathcal{O}_M$ we still get the following equality in \mathcal{S}'

$$\mathcal{F}(f \cdot \Delta_\Gamma) \cdot \mathbb{1}_{D^*} = \frac{S^*}{(2\pi)^d} (\hat{f} * \Delta_{\Gamma^*}) \cdot \mathbb{1}_{D^*} = \frac{S^*}{(2\pi)^d} \hat{f}. \quad (76)$$

To show the last equality observe that (S1) and (S2) ensure that there is no aliasing, in other words, that $\text{supp}(\hat{f} * \delta_{\gamma^*}) \cap D^* = \emptyset$ for any $\gamma^* \in \Gamma^* \setminus 0$. Thus $(\hat{f} * \Delta_{\Gamma^*}) \cdot \mathbb{1}_{D^*} = \sum_{\gamma^* \in \Gamma^*} (\hat{f} * \delta_{\gamma^*}) \cdot \mathbb{1}_{D^*}$ actually has only one non-zero term, namely when $\gamma^* = 0$.

Finally we only have to compute the inverse Fourier transform on both sides of (76). According to the Fourier inversion formula (62) the right-hand side gives $S^* f$, and to compute the left-hand side we apply once again (71) to obtain

$$\frac{1}{(2\pi)^d} \overline{\mathcal{F}}(\mathcal{F}(g)) * \overline{\mathcal{F}}(\mathbb{1}_{D^*}) = S^* f.$$

Observe that we can apply (71) for $u = \mathcal{F}(g)$ and $v = \mathbb{1}_{D^*} \in \mathcal{S}'$, because $g \in \mathcal{O}'_C$. In fact, g is clearly a distribution and it has a fast decay because so does $f \in \mathcal{S}$.

To complete the proof we only need to apply once more the Fourier inversion formula to obtain $\overline{\mathcal{F}}(\mathcal{F}(g)) = (2\pi)^d g$ and rearrange the terms in the last equation. \square

Note that this proof shows the interpolation equation (75) only in the sense of \mathcal{S}' (tempered distributions). We shall see later that it is also valid pointwise with uniform convergence and with less restrictive hypothesis on the decay rate of the continuous signal f at infinity. Part of the proof, however, will require results from Fourier series in L^2 .

We can still give a proof which is completely based on distribution arguments if we trade some decay speed in f for a slightly over-critical sampling which allows for interpolation kernels s with a more regular spectrum. This form of Shannon's sampling theorem is also interesting for the numerical implementation, since the more regular form of \hat{s} means a faster decay rate for s and hence faster approximate interpolation.

Theorem 4 (Shannon-Whittaker in \mathcal{O}_M) Given a function $f : \mathbb{R}^d \rightarrow \mathbb{R}$, a d -dimensional sampling grid Γ and a compact reciprocal cell $D^* \subset \mathbb{R}^d$. If

(S1) (D^*, Γ^*) is a packing

(S2') $\text{supp}(\hat{f}) \subset \overset{\circ}{D^*}$ (i.e. f is band-limited, and D^* is slightly over-critical),

then f can be completely recovered from its samples in Γ , i.e. from

$$g = \Delta_\Gamma \cdot f. \quad (77)$$

In fact the reconstruction is given by the following convolution

$$f = \frac{1}{S^*} \overline{\mathcal{F}}(\varphi) * g \quad (78)$$

of the sampled image g with any kernel $s = \frac{1}{S^*} \overline{\mathcal{F}}(\varphi)$ satisfying the following conditions:

(S3) $\varphi \in C_0^\infty$, $\text{supp}(\varphi) \subseteq D^*$, $\varphi|_{\text{supp}(\hat{f})} \equiv 1$.

Such a kernel always exists but is not necessarily unique.

Proof Since f has compact spectral support, according to equation (70), it has an infinite number of continuous derivatives and grows slowly, i.e. $f \in \mathcal{O}_M$. Since in addition $\Delta_\Gamma \in \mathcal{S}'$ the product $g = \Delta_\Gamma \cdot f$ defining the sampled image g makes sense and is in \mathcal{S}' according to (69).

Then we derive Poisson's formula exactly as before:

$$\mathcal{F}(f \cdot \Delta_\Gamma) = \frac{S^*}{(2\pi)^d} \hat{f} * \Delta_{\Gamma^*}.$$

Now let's construct φ satisfying (S3). Let $K = \text{supp}(\hat{f})$ denote the compact spectral support of f . From (S2) there is a positive distance $d(\partial D^*, K) = \varepsilon > 0$ between the border of D^* and K . Let $\psi \in C_0^\infty$ be a test function with support in a zero-centered ball of radius $\varepsilon/2$, and let K_ε be the dilation of K by the same ball. Then it is easy to see from (66) that $\varphi = \psi * \mathbb{1}_{K_\varepsilon}$ satisfies all conditions in (S3).

Multiplying both sides of the Poisson formula by $\varphi \in C_0^\infty$ we still get the following equality in \mathcal{S}'

$$\mathcal{F}(f \cdot \Delta_\Gamma) \cdot \varphi = \frac{S^*}{(2\pi)^d} (\hat{f} * \Delta_{\Gamma^*}) \cdot \varphi = \frac{S^*}{(2\pi)^d} \hat{f}. \quad (79)$$

The last equality is shown in exactly the same manner as before. The only non-zero term in the convolution is when $\gamma^* = 0$ because $\text{supp}(\varphi) \subseteq D^*$, and this term is $(\hat{f} * \delta_0) \cdot \varphi = \hat{f}$ because $\varphi|_{\text{supp}(\hat{f})} \equiv 1$.

Finally we only have to compute the inverse Fourier transform on both sides of (79). According to the Fourier inversion formula (62) the right-hand side gives $S^* f$, and to compute the left-hand side we apply once again (71) to obtain

$$\frac{1}{(2\pi)^d} \overline{\mathcal{F}}(\mathcal{F}(g)) * \overline{\mathcal{F}}(\varphi) = S^* f.$$

Observe that this time we can apply (71) for $v = \mathcal{F}(g)$ and $u = \varphi$, but the image and the kernel play inverse roles as in the previous case. As for the sampled image g the best thing we can say is that it grows slowly, but it is not in \mathcal{O}_M since it is a distribution. So $g \in \mathcal{S}'$ is a tempered distribution and so is $v = \mathcal{F}(g) \in \mathcal{S}'$ according to (61). Thus the necessary decay rate for the formula to hold is provided by $u = \varphi \in C_0^\infty \subset \mathcal{S}$.

To complete the proof we only need to apply once more the Fourier inversion formula to obtain $\overline{\mathcal{F}}(\mathcal{F}(g)) = (2\pi)^d g$ and rearrange the terms in the last equation. \square

Finally we provide an optimal formulation we know of Shannon's sampling theorem in the case of critical sampling. It is optimal in the sense that it holds with the weakest hypothesis, and that it provides the strongest results.

Theorem 5 (Shannon-Whittaker in L^2) Given a function $f \in L^2(\mathbb{R}^d)$, a d -dimensional sampling grid Γ and a compact reciprocal cell $D^* \subset \mathbb{R}^d$. If the following conditions are met

(S1) (D^*, Γ^*) is a packing

(S2) $\text{supp}(\hat{f}) \subset D^*$ (i.e. f is band-limited),

then f can be completely recovered from its samples in Γ , i.e. from

$$g = \{f(\gamma)\}_{\gamma \in \Gamma}. \quad (80)$$

In fact, $g \in l^2(\Gamma)$, and the reconstruction is given by the following convolution

$$f(x) = \sum_{\gamma \in \Gamma} g_\gamma s(x - \gamma) \quad (81)$$

of the sampled image g with the generalized sinc kernel $s = \frac{1}{S^*} \overline{\mathcal{F}}(\mathbb{1}_{D^*})$. This equality holds in general only in the sense of the L^2 norm. If in addition $g \in l^1(\Gamma)$, then there is pointwise equality and uniform convergence.

Proof Since f has compact spectral support, $\hat{f} \in \mathcal{E}'$ and we deduce Poisson's formula in the same manner as in theorem 3

$$\mathcal{F}(f \cdot \Delta_\Gamma) = \frac{S^*}{(2\pi)^d} \hat{f} * \Delta_{\Gamma^*},$$

where both members are tempered distributions, and equality is to be understood in this sense.

Now observe that the right-hand side of Poisson's formula $F = \frac{S^*}{(2\pi)^d} \hat{f} * \Delta_{\Gamma^*}$ is a Γ^* -periodic function in $L^2(R)$ for any tile R of Γ^* (this is due to the fact that $f \in L^2(\mathbb{R}^d)$ and hence also $\hat{f} \in L^2(\mathbb{R}^d)$, according to (63)). Therefore F admits a Fourier series development:

$$F(\xi) = \sum_{\gamma \in \Gamma} c_\gamma e^{i\langle \xi, \gamma \rangle} \quad (82)$$

where $\{c_{-\gamma}\}_\Gamma$ is in $l^2(\Gamma)$, and the equality is valid in the $L^2(R)$ for any tile R of Γ^* .

On the other hand the left-hand side of the Poisson formula can be written:

$$F(\xi) = \sum_{\gamma \in \Gamma} f(\gamma) e^{-i\langle \gamma, \xi \rangle}. \quad (83)$$

So the uniqueness of the Fourier series development (65) means that $c_{-\gamma} = f(\gamma) = g_\gamma$, and we conclude that the sampled image $g \in l^2(\Gamma)$ is square-summable.

Multiplying F by $\mathbb{1}_{D^*} \in L^2$ we still get the following equality in $L^2(\mathbb{R}^d)$

$$\sum_{\gamma \in \Gamma} f(\gamma) \mathbb{1}_{D^*}(\xi) e^{-i\langle \gamma, \xi \rangle} = F(\xi) \mathbb{1}_{D^*}(\xi) = \frac{S^*}{(2\pi)^d} ((\hat{f} * \Delta_{\Gamma^*}) \cdot \mathbb{1}_{D^*})(\xi) = \frac{S^*}{(2\pi)^d} \hat{f}(\xi). \quad (84)$$

To show the last equality observe as in theorem 3 that (S1) and (S2) ensure that there is no aliasing.

Finally we only have to compute the inverse Fourier transform on both sides of (76). According to the Fourier inversion formula (64) the right-hand side gives $S^* f$, and to compute the left-hand side we use the continuity of $\overline{\mathcal{F}}$ (63) to obtain

$$\overline{\mathcal{F}}\left(\sum_{\gamma \in \Gamma} f(\gamma) \mathbb{1}_{D^*}(\xi) e^{-i\langle \gamma, \xi \rangle}\right) = \sum_{\gamma \in \Gamma} f(\gamma) \overline{\mathcal{F}}(\mathbb{1}_{D^*}(\xi) e^{-i\langle \gamma, \xi \rangle}).$$

Using equation (72) we substitute $\overline{\mathcal{F}}(\mathbb{1}_{D^*}(\xi)e^{-i\langle \gamma, \xi \rangle}) = \overline{\mathcal{F}}(\mathbb{1}_{D^*})(x - \gamma)$, to obtain:

$$f(x) = \frac{1}{S^*} \sum_{\gamma \in \Gamma} g_\gamma \overline{\mathcal{F}}(\mathbb{1}_{D^*})(x - \gamma)$$

which is equivalent to (81), and the convergence of the series towards f holds in the sense of the $L^2(\mathbb{R}^d)$ norm. If in addition $g \in l^1(\Gamma)$ (i.e. $\sum_\gamma |g_\gamma| < +\infty$) then, since $s = \frac{1}{S^*} \overline{\mathcal{F}}(\mathbb{1}_{D^*})$ is bounded⁷, we also have $\sum_{|\gamma| < n} |g_\gamma s(x - \gamma)| \leq \sum_{|\gamma| < n} |g_\gamma| = c_n \rightarrow 0$ (where the sequence c_n is independent of x). Therefore the series converges uniformly to a function h which is continuous over \mathbb{R}^d . So, the series converges also to h in $L^2(J)$ for any bounded interval J . We conclude that $h = f$ almost everywhere over \mathbb{R}^d , which means that $f(x) = h(x)$ for all $x \in \mathbb{R}^d$, since f and h are continuous. \square

A.3 Suitability of the hypothesis in real systems

As we explained in section 3 in real image acquisition systems the continuous image f to be sampled can be modeled as a convolution:

$$f = h * O \quad (85)$$

where O represents the landscape, i.e. the photons that meet the imaging system, and the kernel h depends on the filtering performed by the optics of the acquisition device, the sensors and the movement of the whole system. Thus O can be modeled as a positive Radon measure with compact support, and hence $O \in \mathcal{E}'$. As for the kernel h , despite the approximate models introduced in section 3, it is commonly accepted that it has compact spectral support. However this support is usually much larger than its essential support (beyond the noise level), so from an engineering point of view this remark is not very useful. But it is useful for checking the validity of the theorem for real images.

Now since $\hat{h} \in \mathcal{E}'$ we conclude that $\hat{f} = \hat{h} \cdot \hat{O}$ is also in \mathcal{E}' according to (68), and $f \in \mathcal{O}_M$. So this is enough to satisfy the conditions of theorem 4 with over-critical sampling.

To meet the hypothesis of the two other versions we shall need to assume some more regularity on the kernel h . For theorem 4 we shall need to assume that the transfer function $\hat{h} \in C_0^\infty$ is a test function. Then according to (68) we have $\hat{f} = \hat{h} \cdot \hat{O}$ with $\hat{O} \in \mathcal{O}_M$. Therefore f has both compact support and an infinite number of derivatives, so $\hat{f} \in C_0^\infty \subset \mathcal{S}$, and $f \in \mathcal{S}$ according to (A.1), so f meets the hypothesis in theorem 4.

Finally, for theorem 5 it suffices to assume that \hat{h} is continuous (and in addition of compact support as before). So it is in $L^1 \cap L^2$, and $\hat{h} \in \mathcal{O}_M \cap L^2$ according to equations (63) and (70). We can conclude that $f = h * O$ is C^∞ that it tends to zero at infinity and $f \in L^2$.

B Proof of theorem 2

The proof is a simple application of the following results from information theory, which can be found for instance in (Cover and Thomas, 1991).

Lemma 1 *Given any three random variables X_1, X_2, X_3*

$$H(X_1 + X_2 + X_3) \geq H(Y_1 + Y_2 + Y_3) \text{ where } Y_i \text{ are indep. normal} \quad (86)$$

$$\text{with } H(X_i) = H(Y_i)$$

$$\text{and = iff } X_i \text{ are indep. normal.}$$

⁷ In fact $|s(x)| \leq \frac{1}{S^*} \int_{\mathbb{R}^d} |\mathbb{1}_{D^*}(\xi)e^{-i\langle \xi, x \rangle}| d\xi \leq \frac{1}{S^*} \int_{D^*} 1 d\xi = \frac{|D^*|}{S^*} \leq 1$, where the last inequality holds because D^* is a sub-tile of Γ^* .

Lemma 2 Given any pair of random variables X_1, X_2 :

$$H(X_1 + X_2 | X_1) = H(X_2 | X_1) \leq H(X_2) \quad (87)$$

with = iff X_1, X_2 indep.

Lemma 3 The entropy of any normally distributed random vector \mathbf{X} with zero mean and covariance matrix \mathbf{K} is

$$H(\mathbf{X}) = \frac{1}{2} \log_2 (2e\pi |\mathbf{K}|) \quad (88)$$

where $|\mathbf{K}|$ denotes the determinant of \mathbf{K} .

In particular, for a zero-mean, normally distributed random variable X the entropy is:

$$H(X) = \frac{1}{2} \log_2 (2e\pi \text{Var}\{X\}). \quad (89)$$

Lemma 4 Given any normally distributed random vector \mathbf{X} with zero mean and covariance matrix \mathbf{K} , the sum $\sum_i X_i$ is also zero-mean, normally distributed, with variance:

$$\text{Var}\left\{\sum_i X_i\right\} = \mathbf{1}^T \mathbf{K} \mathbf{1} \quad (90)$$

where $\mathbf{1}$ is a column vector of ones having the same size as \mathbf{X} .

In particular, given two independent, zero-mean, normally distributed random variables Y_1, Y_2 , the sum $Y_1 + Y_2$ is also zero-mean normally distributed with $\text{Var}\{Y_1 + Y_2\} = \text{Var}\{Y_1\} + \text{Var}\{Y_2\}$.

Proof (of theorem 2)

The mutual information is defined in terms of entropy H as follows:

$$\begin{aligned} I(\hat{g}(\xi), G(\xi)) &= H(G(\xi) + G_{\text{alias}}(\xi) + \hat{n}(\xi)) - \\ &\quad H(G(\xi) + G_{\text{alias}}(\xi) + \hat{n}(\xi) | G(\xi)) \\ &= H(G(\xi) + G_{\text{alias}}(\xi) + \hat{n}(\xi)) - \\ &\quad H(G_{\text{alias}}(\xi) + \hat{n}(\xi) | G(\xi)). \end{aligned} \quad (91)$$

If we apply lemmas 1 and 2 to each term of the previous equation with $X_1 = G(\xi)$, $X_2 = G_{\text{alias}}(\xi)$ and $X_3 = \hat{n}(\xi)$ we obtain the following lower bound:

$$I(\hat{g}(\xi), G(\xi)) \geq H(Y_1 + Y_2 + Y_3) - H(G_{\text{alias}}(\xi)) \quad (92)$$

with Y_i independent and normally distributed. Hence we can apply lemmas 4 and 3 to obtain:

$$\begin{aligned} H(Y_1 + Y_2 + Y_3) &= \frac{1}{2} \log_2 (2\pi e (\text{Var}\{Y_1 + Y_2 + Y_3\})) \\ &= \frac{1}{2} \log_2 (2\pi e (\text{Var}\{Y_1\} + \text{Var}\{Y_2\} + \text{Var}\{Y_3\})) \\ &= \frac{1}{2} \log_2 \left(2^{2H(Y_1)} + 2^{2H(Y_2)} + 2^{2H(Y_3)} \right) \\ &= \frac{1}{2} \log_2 \left(2^{2H(G(\xi))} + 2^{2H(G_{\text{alias}}(\xi))} + 2^{2H(\hat{n}(\xi))} \right) \end{aligned} \quad (93)$$

where in the last equation we used the fact that Y_i in lemma 2 are such that $H(Y_i) = H(X_i)$. Finally if we combine this last expression with $H(G_{\text{alias}}(\xi)) = \frac{1}{2} \log_2 (2^{2H(G_{\text{alias}}(\xi))})$, we obtain equation (39). Then equation (40) follows directly by applying lemma 3 in the scalar case.

To prove the third part of the theorem we go back to equation (91). To develop the first term, in that equation, observe that the covariance matrix of $(G(\xi), G_{\text{alias}}(\xi), \hat{n}(\xi))$ consists of a 2 by 2 diagonal block (corresponding to $(G(\xi), G_{\text{alias}}(\xi))$), and a 1 by 1 diagonal block (corresponding to $\hat{n}(\xi)$). Hence:

$$\begin{aligned} \mathbf{1}^T \mathbf{K} \mathbf{1} &= \text{Var}\{G(\xi)\} + \text{Var}\{G_{\text{alias}}(\xi)\} \\ &\quad + 2\text{Cov}\{G(\xi), G_{\text{alias}}(\xi)\} + \text{Var}\{\hat{n}(\xi)\} \end{aligned} \quad (94)$$

and we can develop the first term as follows:

$$\begin{aligned} H(G(\xi) + G_{\text{alias}}(\xi) + \hat{n}(\xi)) &= \frac{1}{2} \log(2e\pi (\text{Var}\{G(\xi)\} + \text{Var}\{G_{\text{alias}}(\xi)\} \\ &\quad + 2\text{Cov}\{G(\xi), G_{\text{alias}}(\xi)\} + \text{Var}\{\hat{n}(\xi)\})) \end{aligned} \quad (95)$$

To develop the second term in equation (91) we use the definition of conditional entropy, and then lemma 3, to obtain:

$$\begin{aligned} H(G_{\text{alias}}(\xi) + \hat{n}(\xi) | G(\xi)) &= H(G(\xi), G_{\text{alias}}(\xi) + \hat{n}(\xi)) - H(G(\xi)) \\ &= \frac{1}{2} \log(2e\pi (\text{Var}\{G(\xi)\} \text{Var}\{G_{\text{alias}}(\xi) + \hat{n}(\xi)\} \\ &\quad - \text{Cov}\{G(\xi), G_{\text{alias}}(\xi) + \hat{n}(\xi)\}^2)) \\ &= \frac{1}{2} \log(2e\pi (\text{Var}\{G(\xi)\} (\text{Var}\{G_{\text{alias}}(\xi)\} + \text{Var}\{\hat{n}(\xi)\}) \\ &\quad - \text{Cov}\{G(\xi), G_{\text{alias}}(\xi)\}^2)). \end{aligned} \quad (96)$$

In the last line we applied lemma 4 to the independent Gaussian variables $G_{\text{alias}}(\xi), \hat{n}(\xi)$ together with the fact that $\text{Cov}\{G(\xi), G_{\text{alias}}(\xi) + \hat{n}(\xi)\} = \text{Cov}\{G(\xi), G_{\text{alias}}(\xi)\}$ due to the independence of $G_{\text{alias}}(\xi)$ and $\hat{n}(\xi)$.

Finally we combine the two terms in equation (91) using the previous dimensionless variables to obtain equation (41) which is always well defined, since $k(\xi) \in [0, 1]$, because $(\text{Var}\{G(\xi)\} \text{Var}\{G_{\text{alias}}(\xi)\} - \text{Cov}\{G(\xi), G_{\text{alias}}(\xi)\}^2)$ must be non-negative, since it is the determinant of the (positive semi-definite) covariance matrix of $(G(\xi), G_{\text{alias}}(\xi))$. \square

C Sufficient conditions for admissibility

Definition 6 for an admissible alias selector is quite general but a bit technical. Here we present a set of sufficient conditions that may be easier to verify. In particular they can be easily checked for the function u we apply it to, namely $u(\xi) = \frac{1}{2} \log_2(1 + \frac{1}{a(\xi)^2 + b(\xi)^2})$ when a and b are the relative aliasing and noise measures for a given transfer function and image and noise models.

Lemma 5 *Let Γ^* be a regular grid in \mathbb{R}^d . For any function $u : \mathbb{R}^d \rightarrow \mathbb{R}$, satisfying the following properties*

- (i) $u(x) \geq 0, \forall x \in \mathbb{R}^d$
- (ii) $u(x) \rightarrow 0$, when $|x| \rightarrow \infty$
- (iii) $(\text{supp}(u), \Gamma^*)$ is a covering
- (iv) The periodicity set P_{u, Γ^*} of u and Γ^* defined below has zero measure

$$P_{u, \Gamma^*} = \left\{ x \in \mathbb{R}^d : \exists \gamma \in \Gamma^* \setminus \{0\}, u(x) = u(x + \gamma) > 0 \right\}$$

(iv) the associated optimal tiling R_{\max} is bounded

then u is an admissible alias selector.

Proof

For each point x consider the set G_{x,u,Γ^*} in the definition of an admissible alias selector (that we abbreviate in the sequel by G_x , let's characterize the anomaly set by the following case discussion.

Case 1: If G_x consists of a single element γ_x , then x is not in the anomaly set A_{u,Γ^*} .

Case 2: We cannot have $G_x = \phi$ because this would mean that there is a sequence $\{\gamma_n\}$ of elements in Γ^* , such that $\lim u(x + \gamma_n) = u_x$, and the limit u_x is strictly larger than any $u(x + \gamma_n)$. Since γ takes values in Γ^* this means that $\{\gamma_n\}$ is unbounded, so according to (ii) we have $\lim u(x + \gamma_n) = u_x = 0$. Since (i) imposes u to be non-negative this means that $u(x + \gamma_n) = u_x = 0$ for all n , which gives an infinite number of elements in G_x , contradicting our initial supposition that $G_x = \phi$.

Case 3: Similarly, if $G_x = \{\gamma_n : n \in \mathbb{N}\}$ has an infinite number of different elements we conclude that $u_x = 0$. Let's call the set of such points zero-set of u and Γ^* :

$$Z_{u,\Gamma^*} = \{x : u_x = 0\}.$$

Clearly any point $y \in \bigcup_{\gamma \in \Gamma^*} (\text{supp}(u) + \gamma)$ is excluded from the zero-set, i.e. $y \notin Z_{u,\Gamma^*}$. Therefore, (iii) implies that the zero-set Z_{u,Γ^*} has zero measure.

Case 4: Finally if G_x contains a finite number of elements but more than one, then x belongs to the periodicity set of u and Γ^* .

We conclude that the anomaly set is the union of the zero set and the periodicity set, both of zero measure. So the anomaly set $A_{u,\Gamma^*} = P_{u,\Gamma^*} \cup Z_{u,\Gamma^*}$ also has zero measure. \square

More generally, observe that in proposition 1 the anomaly set can as well be divided into two sets: A^+ consisting of all points x such that G_x has more than one element, and A^0 consisting of all points x such that G_x is empty. Observe that we could tolerate that A^+ has a positive measure, but not that A^0 has a positive measure. In the latter case there is no maximal tiling, whereas in the former case we have several possible maximal tilings. Indeed for any $x \in A^+$, to build a maximal tiling R_{\max} , we can substitute x by $(x + \gamma)$ for any $\gamma \in G_x$, without changing the value of the integrand in $\int_{R_{\max}} v(u(x)) dx$.

References

- A. Almansa, J.M. Morel, and B. Rougé. Hexagonal sampling in image processing: A review. Technical report, CMLA, ENS Cachan, August 1999.
- B. Bauer. *Rappresentazione di Vettori in Spazi ad n Dimensioni*. PhD thesis, Università degli Studi di Milano, 1981.
- Lucien M. Biberman. *Perception of Displayed Information*. Optical Physics and Engineering. Plenum Press, New York, London, 1973.
- Jean-Michel Bony. *Cours d'analyse. Théorie des distributions et analyse de Fourier*. Les Éditions de l'École Polytechnique, 2001.
- G. Borgefors. Distance transformations in hexagonal grids. *Pattern Recognition Letters*, (9):97–105, 1989.
- Ronald N. Bracewell. *Two-Dimensional Imaging*. Signal Processing Series. Prentice Hall, Englewood Cliffs, New Jersey 07632, 1995.
- A. Cohen and J.M. Schlenker. Compactly supported bidimensional wavelet bases with hexagonal symmetry. *Constructive Approximation*, (9):209–236, 1993.
- Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications. John Wiley and sons, 1991.
- I. Daubechies, I. Guskov, P. Schröder, and W. Sweldens. Wavelets on irregular point sets. *Phil. Trans. Royal Society of London A*, 1999. Submitted.
- Sylvain Durand. Evaluation quantitative de la qualité image. Technical report, CMLA, ENS Cachan, 1998.
- J.C. Ehrhardt. Hexagonal fast fourier transform with rectangular output. *IEEE Transactions on Signal Processing*, 41(3):1469–1472, March 1993.
- William T. Freeman, Thonis R. Jones, and Egon C. Pasztor. Example-based super-resolution. Technical report, MERL – A Mitsubishi Electric Research Laboratory, <http://www.merl.com>, August 2001.
- Claude Gasquet and Patrick Witomski. *Analyse de Fourier et Applications*. Masson, 1995.
- Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Addison Wesley, 1992.
- Stéphane Jaffard. *Construction et propriétés des bases d'ondelettes. Remarques sur la contrôlabilité exacte*. PhD thesis, École Polytechnique, May 1989.
- Bernd Jähne. *Digital Image Processing*. Springer-Verlag, 3rd edition, 1995.
- Anil K. Jain. *Fundamentals of Image Processing*. Prentice Hall, 1995.
- J. Kovacevic and W. Sweldens. Wavelet families of increasing order in arbitrary dimensions. *IEEE Transactions on Image Processing*, 1998. Submitted.
- A. F. Laine and S. Schuler. Hexagonal wavelet representations for recognizing complex annotations. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 740–745, Los Alamitos, CA, USA, June 1994. IEEE Computer Society Press. ISBN 0-8186-5825-8.

- Christophe Lathy and Bernard Rougé. Spot5 thr mode. Internal Report, CNES, Toulouse, France.
- François Malgouyres. *Augmentation de la résolution d'images satellitaires: Thorie variationnelle et applications*. PhD thesis, Université de Paris IX – Dauphine, 1999.
- François Malgouyres and Frederic Guichard. Edge direction preserving image zooming: a mathematical and numerical analysis. *Journal on Numerical Analysis*, 39(1):1–37, 2001.
- R.M. Mersereau. The processing of hexagonally sampled two-dimensional signals. *IEEE Proceedings*, 67:930–949, 1979.
- R.M. Mersereau and T.C. Speake. The processing of periodically sampled multidimensional signals. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 31(1):188–194, February 1983.
- Yves Meyer. *Wavelets and operators*, volume 2. Hermann, 1990.
- Alexey A. Minho and Sergey A. Pokotilo. Nonuniform quantization of the image in infrared remote sensing systems: Ways and devices for its elimination. In *Proceedings of SPIE*, volume 3063, pages 312–320, 1997.
- Lionel Moisan and Jean-Michel Morel. Les défauts du dvd sous l'oeil de l'équation. *La Recherche*, (342):56–57, May 2001.
- J.M. Morel and S. Ladjal. Analyse de fourier et théorie de shannon. Analyse de Fourier et traitement d'images. Journées X-UPS 1998. Centre de Mathématiques – Ecole Polytechnique. France., 1998.
- William K. Pratt. *Digital Image Processing*. John Wiley and Sons, 1978.
- C.A. Rogers. *Packing and Covering*. Cambridge University Press, 1964.
- J.K. Romberg, Hyeokho Choi, and Richard Baraniuk. Bayesian tree-structured image modeling using wavelet-domain hidden markov models. *IEEE Transactions on Image Processing*, 10(7):1056–1068, July 2001.
- Azriel Rosenfeld and Avinash C. Kak. *Digital Picture Processing*, volume 1 of *Computer Science and Applied Mathematics*. Academic Press, second edition, 1982.
- Bernard Rougé. Théorie de la chaîne image optique et restauration à bruit final fixé. Mémoire en vue de l'obtention de l'habilitation à diriger des recherches. Option: Mathématiques appliquées, May 1997.
- Bernard Rougé. Théorie de l'échantillonnage et satellites d'observation de la terre. Analyse de Fourier et traitement d'images. Journées X-UPS 1998. Centre de Mathématiques – Ecole Polytechnique. France., 1998.
- John C. Russ. *The image processing handbook*. CRC, 2nd edition, 1995.
- S. Schuler and A. Laine. Hexagonal qmf banks and wavelets. In M. Akai, editor, *Time-Frequency and Wavelet Transforms in Biomedical Engineering*, chapter ? IEEE Press, 1997.
- Eero P. Simoncelli and Edward H. Adelson. Subband image coding with hexagonal quadrature mirror filters. In *Picture Coding Symposium*, Cambridge, MA., March 1990a.
- E.P. Simoncelli and E.H. Adelson. Subband transforms. In John W. Woods, editor, *Subband Coding*, chapter 4, pages 143–192. Kluwer Academic Publishers, 1990b. Also available from MIT Media Laboratory as Technical Report #137, <ftp://whitechapel.media.mit.edu/pub/tech-reports/TR-137.ps.Z>.

- R.C. Staunton. Hexagonal sampling in image processing. *Advances in Imaging and Electron Physics*, 107:231–307, 1999.
- C.A. Wühtrich and P. Stucki. An algorithmic comparison of square- and hexagonal-based grids. *CVGIP: Graphical Models and Image Processing*, 53(4):324–339, 1991.
- Jaime L. Zapata and Gerhard X. Ritter. Fast fourier transform for hexagonal aggregates. *Journal of Mathematical Imaging and Vision*, 12:183–197, 2000.