Mathematical models in image processing

C.O.S. Sorzano [a], M.C. Pérez-Gandía [a], A. Otero-Quintana [b]

[a] Dpto. de Ing. de Sistemas Electrónicos y de Telecom. Escuela Politécnica Superior, Univ. San Pablo – CEU

coss.eps@ceu.es

[b] Dpto. de Ing. del Software y del Conocimiento. Escuela Politécnica Superior, Univ. San Pablo – CEU

Resumen

El procesamiento de imágenes es un caso particular del procesamiento de señales del cual la Ingeniería de Telecomunicación es quizás el mayor contribuyente. El procesamiento de imágenes está completamente permeado de técnicas matemáticas de mayor o menor complejidad utilizadas con el objetivo de resolver problemas relacionados con la reconstrucción y restauración de imágenes, extracción de bordes, modelado de formas, segmentación, registro, etc. Para ello se utilizan técnicas de análisis funcional, modelos estadísticos, optimización, ecuaciones en derivadas parciales, álgebra lineal, etc. En general es difícil encontrar un campo de la matemática que no haya sido utilizado para resolver cada uno de los problemas que aparecen en procesamiento de imágenes. Este artículo pretende introducir brevemente al lector en la importante fertilización cruzada que un problema como el procesamiento de imágenes ha tenido con las matemáticas.

Palabras clave

Procesado de imágenes, procesado de señales, matemática aplicada.

Abstract

Image processing is a particularization of signal processing of which Electrical Engineering is, maybe, the major contributor. Image processing is completely pervaded by more or less complex mathematical tools aimed at solving problems related to image reconstruction and restoration, boundary extraction, shape modeling, segmentation, registration, etc. For performing these tasks, mathematical tools are employed such as functional analysis, statistical models, optimization, partial differential equations, linear algebra, etc. In general, it is difficult to find a mathematical area that has not been used to solve each of the problems in image processing. This article briefly introduces the important cross-fertilization between the image processing problems and mathematics in general.

Keywords

Image processing, signal processing, applied mathematics.

1. Introduction

Image processing is a challenging engineering topic aimed at solving many real-life problems by means of images. Applications range from intelligent road vehicles "looking" into the road for possible dangers [1], automatic fruit quality assessment [2], medical imaging non-invasively looking for tumors inside the body [3], or the determination of the structure of a macromolecular complex [4]. All these problems share the commonality of trying to solve the problem at hand by extracting features from the image such as borders, dimensions, distances, textures, shapes, etc. This is done with the help of a computer and an algorithm that effectively carries out the job.

In general, all these tasks are grouped under a single name, image processing, which is no more than a particularization of a more general framework, signal processing. A signal can be mathematically modeled as a function $u: X \rightarrow Y$. The two most widely used spaces for X are $\mathbb{R}^n$ and $\mathbb{Z}^n$. If $n=1$, we speak of "signal processing". If $n=2$, we speak of "image processing" and $\mathbf{r} \in X$ represents the

spatial location at which a certain gray level or color is observed. For gray color images, Y represents the range of possible gray levels and is usually considered to be $\mathbb{R}$ or $\Delta\mathbb{Z} \cap [I_{\min}, I_{\max}]$ (where $\Delta \in \mathbb{R}$ is the quantization step, and $\left[I_{\min}, I_{\max}\right] \subseteq \mathbb{R}$ represents the minimum and maximum levels of the image). For a color image, the function is usually of the kind $u : X \to Y_{gray}^3$ where Y_{gray} is the range for a gray image. A well-known color image code is the RGB (Red-Green-Blue) in which each of the components of $u(\mathbf{r}) = (u_{red}(\mathbf{r}), u_{green}(\mathbf{r}), u_{blue}(\mathbf{r}))$ represents the intensity of a red light beam, a green light beam, and blue light beam, all of them superposed at the same point to produce a color.

Image processing addresses the problem of how to express image features like distances, textures, and borders in a mathematical way, and more importantly, how to detect them in a real image. A non-exhaustive list of image processing problems would comprise image denoising, image restoration, blind image deconvolution, image reconstruction, image enhancement, boundary extraction, shape modeling, image segmentation, image registration, motion analysis, volumes from images, feature recognition, etc. On the mathematical side, a list of techniques used would include functional analysis, statistical models, Monte Carlo simulations, optimization, partial differential equations, linear algebra, operator theory, graph theory, topology, differential geometry, discrete geometry, stochastic processes, set theory, level sets, fuzzy sets, fuzzy logic, etc. It would not be untrue to say that for any combination of mathematical tool and image processing problem there is at least 1 paper, and in most occasions, hundreds. A full coverage of the connection between applied mathematics and image processing would be out of the scope of a conference paper. The reader interested is referred to very good books in image processing such as [5-8]. Instead, in this short review, we show an example of imaging problem and how it involves many mathematical fields. In particular, we show the different approaches currently taken to denoise and restore images. For doing this, we need first to introduce the concepts of image observation model and image transforms.

2. Image observation model

The key step in many image processing problems is to state the way in which the image has been generated or degraded. This is called the image observation model. A very general model is the one given proposed in [6]. An ideal image $u(\mathbf{r}) : \mathbb{R}^2 \to \mathbb{R}$ is transformed into an observed image $v(\mathbf{r}) : \mathbb{R}^2 \to \mathbb{R}$ by linear and nonlinear transformations plus noise.

$$\begin{aligned} v(\mathbf{r}) &= g\left(w(\mathbf{r})\right) + \varepsilon(\mathbf{r}) \\ w(\mathbf{r}) &= \int_{\mathbb{R}^2} u(\mathbf{r}')h(\mathbf{r}, \mathbf{r}')d\mathbf{r} \\ \varepsilon(\mathbf{r}) &= f\left(g\left(w(\mathbf{r})\right)\right)\varepsilon_1(\mathbf{r}) + \varepsilon_2(\mathbf{r}) \end{aligned} \quad (1)$$

The linear transformation defined by $h(\mathbf{r}, \mathbf{r}') : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ is usually used to represent some blurring. If h fulfills $h(\mathbf{r}, \mathbf{r}') = h(\mathbf{r} - \mathbf{r}', \mathbf{0})$, then it is said that the system is spatially invariant and the integral used to compute w becomes a convolution integral. $g : \mathbb{R} \to \mathbb{R}$ is a nonlinear function usually representing nonlinear darkening or whitening of the image gray level. $\varepsilon(\mathbf{r})$ is a random process (i.e., a collection of random variables) formed by the combination of other two independent random processes, one depending on the signal through two nonlinear transformations (f and g) and one linear (h). If the autocorrelation of the random process $\varepsilon(\mathbf{r})$ is $\delta(\mathbf{r})$ (the Dirac impulse distribution), then the noise is said to be white since its Power Spectrum Density is constant (as the one of white light)[9]. White noise implies that the noise added in one position is independent of the noise added in another position. This may not be the general case.

Figure 1 shows an ideal image, the image after linear blurring to simulate the effect of an electron microscope and the image after blurring and noise addition.

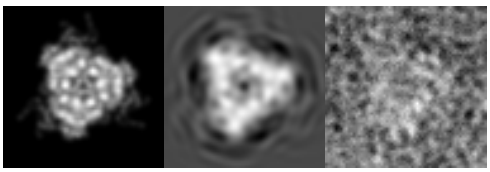

Fig. 1. Left: Ideal projection of the bacteriorhodopsin [10]. Middle: Linear convolution to simulate the effect of an electron microscope [4]. Right: Ideal image linearly convolved and with noise added [4].

3. Image transforms

The function $u(\mathbf{r})$ is assumed to belong to $L^2(\mathbb{R})$ (i.e., the space of all square integrable functions). The set of Dirac's delta functions is a basis of this space, and therefore $u(\mathbf{r})$ can be expressed as a linear combination of Dirac's delta functions, i.e.,

$$u(\mathbf{r}) = \int_{\mathbb{R}^2} u(\mathbf{r}')\delta(\mathbf{r}-\mathbf{r}')d\mathbf{r}' \qquad (2)$$

In fact, this is the base in which it is usually expressed. However, we can change the basis of this space. For instance, complex exponential functions are also a basis of this space, and the function $u(\mathbf{r})$ can be expressed as

$$u(\mathbf{r}) = \int_{\mathbb{R}^2} \left\langle u(\mathbf{r}), e^{j\langle \boldsymbol{\omega},\mathbf{r}\rangle} \right\rangle e^{j\langle \boldsymbol{\omega},\mathbf{r}\rangle} d\boldsymbol{\omega} \qquad (3)$$

where $j=\sqrt{-1}$ and $\boldsymbol{\omega} \in \mathbb{R}^2$. $\langle \mathbf{x}, \mathbf{y} \rangle$ represents the dot product between vectors $\mathbf{x}$ and $\mathbf{y}$. For vectors in $\mathbb{R}^n$ it is computed in the usual way, and for functions it is computed as

$$\langle x(\mathbf{r}), y(\mathbf{r}) \rangle = \int_{\mathbb{R}^2} x(\mathbf{r}) y^*(\mathbf{r}) d\mathbf{r} \qquad (4)$$

where y^* denotes the complex conjugate of y. The projection of the function $u(\mathbf{r})$ onto each of the basis functions is called the Fourier transform of $u(\mathbf{r})$. The Fourier Transform of a signal is denoted as $FT\{u(\mathbf{r})\}$ or $u(\boldsymbol{\omega})$ and is defined as

$$FT\{u(\mathbf{r})\} = u(\boldsymbol{\omega}) \triangleq \left\langle u(\mathbf{r}), e^{j\langle \boldsymbol{\omega},\mathbf{r}\rangle} \right\rangle \qquad (5)$$

The Fourier Transform represents nothing more than a change of basis in the space $L^2(\mathbb{R})$. Eq. 3 is usually referred to as the Inverse Fourier Transform and is denoted by $FT^{-1}\{u(\boldsymbol{\omega})\}$.

The Fourier transform is not the only possible change of basis. In fact, recently the wavelet transform and other similar transforms (like the wavelet packets and the local cosine bases) [11] have been paid very much attention due to some desirable properties of the transformed image. In brief, 1D wavelet functions are functions $\psi : \mathbb{R} \to \mathbb{R}$ that have the property that the set of dilated and shifted versions of ψ can also span $L^2(\mathbb{R})$, i.e.,

$$u(r) = \int_{\mathbb{R}^2} C_{a,b} \left\langle u(r), \psi_{a,b}(r) \right\rangle \psi_{a,b}(r) da db \qquad (6)$$

where

$$\psi_{a,b}(r) = \frac{1}{a}\psi\left(\frac{r-b}{a}\right) \qquad (7)$$

and $C_{a,b}$ is some suitable constant depending on a and b. This is called the continuous wavelet transform since parameters a and b can take any real value. The extension to several dimensions is performed by tensor product in each of the directions.

However, this representation is overcomplete in the sense that there are many more $\psi_{a,b}(r)$ than

needed to construct a basis of $L^2(\mathbb{R})$. Alternatively, the continuous wavelet transform can be discretized so that parameters a and b are not continuous parameters but discrete ones. It can be shown that signal $u(r)$ can be decomposed as

$$u(r) = \sum_{(i,s)\in\mathbb{Z}^2} \left\langle u(r), \psi_{i,s}(r) \right\rangle \psi_{i,s}(r) \qquad (8)$$

where

$$\psi_{i,s}(r) = \frac{1}{\sqrt{2^s}} \psi\left(\frac{r - i2^s}{2^s} \right) \qquad (9)$$

Coefficients $\left\langle u(r), \psi_{i,s}(r) \right\rangle$ are referred to as the Discrete Wavelet Transform (DWT) of the function $u(r)$ and are generally denoted as

$$DWT\left\{u(r)\right\} = u(i,s) \triangleq \left\langle u(r), \psi_{i,s}(r) \right\rangle \qquad (10)$$

A wavelet decomposition is orthonormal if

$$\left\langle \psi_{i,s}(r), \psi_{i',s'}(r) \right\rangle = \delta(i - i')\delta(s - s') \qquad (11)$$

Orthonormal wavelets really constitute a basis of $L^2(\mathbb{R})$ and, therefore, efficiently transform the signal content.

The wavelet decomposition of a multidimensional signal, as is the case of an image, the multidimensional wavelet function is formed by the tensor multiplication of 1D wavelet functions [5]. The 2D wavelet transform of an image has four parameters: two location parameters ($\mathbf{i} \in \mathbb{Z}^2$), 1 scale parameter ($s \in \mathbb{Z}$), and 1 orientation parameter ($o \in \{ll, lh, hl, hh\}$) that accounts for how the tensor product was performed. In this paper, the 2D wavelet function of an image will be referred to as $u(\mathbf{i}, \mathbf{s})$ where $\mathbf{s} = (s \quad o)^t$.

4. Image denoising and restoration

Approaches to image denoising usually assume a simplified image formation model such as $v(\mathbf{r}) = u(\mathbf{r}) * h(\mathbf{r}) + \varepsilon(\mathbf{r})$ and aim at estimating $u(\mathbf{r})$ (we will refer to the estimate as $\hat{u}(\mathbf{r})$). Most of the times, the different techniques need to make some assumptions about the nature of $\varepsilon(\mathbf{r})$, the most common one that it is white Gaussian noise of zero mean and constant standard deviation. The different techniques for estimating $u(\mathbf{r})$ involve tools and formulations from different mathematical fields. In this paper we discuss the most popular ones, although the number of image denoising and image restoration algorithms is huge.

4.1. Transform based approaches

One of the most widely used techniques to image denoising is based on filtering its Fourier transform. The assumption behind this technique is that the Fourier transform of the image $u(\mathbf{r})$ is bandlimited (i.e., $u(\boldsymbol{\omega}) = 0$ for $\|\boldsymbol{\omega}\| > \omega_{\max}$ and $\omega_{\max} \in \mathbb{R}$). Considering a simplified model $v(\mathbf{r}) = u(\mathbf{r}) + \varepsilon(\mathbf{r})$, and taking into account that the Fourier transform is a linear operation, i.e., $v(\boldsymbol{\omega}) = u(\boldsymbol{\omega}) + \varepsilon(\boldsymbol{\omega})$, any value of $v(\boldsymbol{\omega})$ different from zero beyond $\omega_{\max}$ must come from the noise term. Hence, a simple way of removing some noise from the image is by setting to zero all $v(\boldsymbol{\omega})$ values beyond $\omega_{\max}$. This is known as low-pass filtering [5]. Defining a denoising function

$$g(\boldsymbol{\omega}) = \begin{cases} 1 & \|\boldsymbol{\omega}\| \le \omega_{\max} \\ 0 & \|\boldsymbol{\omega}\| > \omega_{\max} \end{cases}, \qquad (12)$$

low pass filtering can be performed as

$$\hat{u}(\mathbf{r}) = TF^{-1}\left\{ g(\boldsymbol{\omega}) v(\boldsymbol{\omega}) \right\} \qquad (13)$$

In practice, the design of function $g(\boldsymbol{\omega})$ is not so simple because our current design introduces artifacts in the image due to the avoidable discontinuity at $\|\boldsymbol{\omega}\| = \omega_{\max}$. The design of suitable continuous functions $g(\boldsymbol{\omega})$ is an active research topic and many solutions have been proposed. The reader interested is referred to [5].

Image denoising can also be performed in other transformed spaces. For instance, it has been observed that in the wavelet space, signal coefficients tend to be larger than noise coefficients. Thus, a possible way to remove noise is by setting to zero the small wavelet coefficients of $v(\mathbf{i},\mathbf{s})$. Defining the denoising function

$$g(v) = \begin{cases} v & v \geq v_{\min} \\ 0 & |v| < v_{\min} \end{cases}, \quad (14)$$

image denoising can be performed as

$$\hat{u}(\mathbf{r}) = DWT^{-1}\{g(v(\mathbf{i},\mathbf{s}))\} \quad (15)$$

This denoising strategy is called hard thresholding. Current research on this kind of denoising is focused in the selection of function $g(v)$ and the threshold $v_{\min}$. The reader interested is referred to [5].

The most common Fourier transform based approach to image restoration assumes an image observation model given by $v(\mathbf{r}) = u(\mathbf{r}) * h(\mathbf{r})$. Under these circumstances, it can be proved that $v(\boldsymbol{\omega}) = u(\boldsymbol{\omega})h(\boldsymbol{\omega})$. Hence, a possible image restoration method would be simply dividing the Fourier transform of the observed image by the Fourier transform of the linear degradation function, i.e.,

$$\hat{u}(\mathbf{r}) = FT^{-1}\left\{\frac{v(\boldsymbol{\omega})}{h(\boldsymbol{\omega})}\right\} \quad (16)$$

This is fine as far as $h(\boldsymbol{\omega}) \neq 0$, otherwise a different method must be employed since the division by zero would infinitely amplify the noise at that frequency.

4.2. Wiener filtering: an ideal statistical approach for restoration

Another famous approach to image restoration is the Wiener filtering which minimizes the variance of the estimate error [6]

$$\hat{u}^*(\mathbf{r}) = \arg\min E\left\{\left(\hat{u}(\mathbf{r}) - u(\mathbf{r})\right)^2\right\}, \quad (17)$$

where $E\{\cdot\}$ denotes the expectation operator, and it is assumed that the additive noise is a zero-mean random process. It is well-known that the solution of this optimization problem is

$$\hat{u}^*(\mathbf{r}) = E\{u(\mathbf{r}) \mid v\} \quad (18)$$

i.e., the conditional expectation of $u(\mathbf{r})$ given the whole observation v. However, computing this expectation is difficult in general since it is a nonlinear equation and because the probability distribution $f(u(\mathbf{r}) \mid v)$ is difficult to calculate. Here is where some assumptions are made about the *a priori* noise model (which allows for computing the *a posteriori* probabilities needed). A further simplification impose that the best estimate can be computed by linear filtering the observed image

$$\hat{u}^*(\mathbf{r}) = \int_{\mathbb{R}^2} v(\mathbf{r}')g(\mathbf{r},\mathbf{r}')d\mathbf{r} \quad (19)$$

and the problem now moves on how to design a good linear filter g. It can be proved [12] that the linear space-invariant solution of this problem is

$$g(\boldsymbol{\omega}) = \frac{h^*(\boldsymbol{\omega})}{|h(\boldsymbol{\omega})|^2 + \frac{1}{SNR(\boldsymbol{\omega})}}, \quad (20)$$

where h^* denotes the complex conjugate of h, and $SNR(\boldsymbol{\omega})$ is the Signal-to-Noise Ratio (SNR) at frequency $\boldsymbol{\omega}$ defined as

$$SNR(\boldsymbol{\omega}) = \frac{\left|u(\boldsymbol{\omega})\right|^2}{\left|\varepsilon(\boldsymbol{\omega})\right|^2}, \tag{21}$$

None of the quantities involved in the SNR are known, and they have to be estimated from the data itself. This is one of the main drawbacks of the Wiener filter as a practical filter. Despite this problem, Wiener filters are one of the most widely used techniques for image restoration and denoising.

4.3. Variational approaches

The total variation of a continuously differentiable function $u(\mathbf{r}) : \mathbb{R}^n \supseteq \Omega \to \mathbb{R}$ in an open-bounded set Ω is defined as

$$\begin{aligned} V_\Omega(u) &= \int_\Omega \left|\nabla u(\mathbf{r})\right| \\ &= \sup_{\substack{w \in C_c^1(\Omega,\mathbb{R}^m) \\ |w(\mathbf{r})| \le 1 \quad \forall \mathbf{r} \in \Omega}} \int_\Omega u(\mathbf{r}) \nabla \cdot w(\mathbf{r}) d\mathbf{r} \end{aligned} \tag{22}$$

where $C_c^1(\Omega, \mathbb{R})$ is the set of all continuously differentiable scalar functions with a compact support contained in Ω. The total variation measures the variability of function u within some region Ω. Variational approaches to image denoising minimize the variation of the restored function as long as the restored function is compatible with the data. In some way, this approach is similar to the maximum entropy one except that the functional to be optimized is different [13, 14]

$$\begin{aligned} &\hat{u}^*(\mathbf{r}) = \arg\min V(\hat{u}) \\ &s.t. \quad \left\|v(\mathbf{r}) - \hat{u}(\mathbf{r})\right\|^2 = \sigma_\varepsilon^2 \\ &\int_\Omega (v(\mathbf{r}) - \hat{u}(\mathbf{r})) d\mathbf{r} = 0 \end{aligned} \tag{23}$$

The norm in this case is the usual norm in L^2, i.e., $\left\| f(\mathbf{r}) \right\|^2 = \int_\Omega f^2(\mathbf{r}) d\mathbf{r}$.

This optimization problem is solved in the space of functions with bounded variations since they allow for regular solutions as well as for solutions with sharp edges. The bounded variation functional space is defined as

$$BV = \left\{ u(\mathbf{r}) \in L^1 : V(u) < \infty \right\} \tag{24}$$

The Euler-Lagrange equation associated to this problem is [14]

$$\frac{\partial}{\partial x}\left(\frac{\langle \nabla \hat{u}, x \rangle}{\left\|\nabla \hat{u}\right\|} \right) + \frac{\partial}{\partial y}\left(\frac{\langle \nabla \hat{u}, y \rangle}{\left\|\nabla \hat{u}\right\|} \right) - \lambda_1 - \lambda_2(\hat{u} - v) = 0 \tag{25}$$

which needs now to be solved numerically.

The variational approach can also be used for image restoration. For instance, assuming a linear convolution blurring, the corresponding variational problem would be [13]

$$\begin{aligned} &\hat{u}^*(\mathbf{r}) = \arg\min V(\hat{u}) \\ &s.t. \quad \left\|v(\mathbf{r}) - h(\mathbf{r}) * \hat{u}(\mathbf{r})\right\|^2 = \sigma_\varepsilon^2 \end{aligned} \tag{26}$$

The total variation can be generalized to any convex, positively one-homogeneous function that is 0 at the origin, $\phi : \mathbb{R}^n \to \mathbb{R}$,

$$V_\Omega(u) = \int_\Omega \phi\left(\nabla u(\mathbf{r})\right) \tag{27}$$

ϕ can be made anisotropic giving raise to whole plethora of anisotropic diffusion algorithms. In general, all the variational approaches boil down to a partial differential equation that needs to be solved numerically.

A different kind of variational approaches try to approximate the underlying continuous function $u(\mathbf{r})$ from observed noisy samples $v(\mathbf{r}_i)$. The idea is to find a smooth function that

approximates well the observed samples and at the same time is smooth. The smoothness measure provides the variational approach, while the fidelity to the observed data (in the previous examples treated as tight constraints) is handled as a soft constraint. The fact that the smooth function is not forced to pass through all observed measurements is supposed to be able to remove noise. The functional to minimize is

$$\hat{u}^*(\mathbf{r}) = \arg\min \int_\Omega \left\| D^p \hat{u}(\mathbf{r}) \right\|^2 d\mathbf{r} + \lambda \sum_i \left\| v(\mathbf{r}_i) - \hat{u}(\mathbf{r}_i) \right\|^2 \quad (28)$$

where λ is a user-provided parameter, D^p is the vector of all derivatives of order p. It is well-known that for $p = 2$, the solution is of the form [15]

$$\hat{u}^*(\mathbf{r}) = \sum_i w_i g\left(\left\| \mathbf{r} - \mathbf{r}_i \right\| \right) + l(\mathbf{r}) \quad (29)$$

where $l(\mathbf{r})$ is a polynomial of degree 1 lying in the kernel of the smoothness seminorm, i.e., $D^2 l(\mathbf{r}) = 0$, $\{w_i\}$ is a collection of weights to be determined, and $g(r) = r^2 \log r$ (this function is also known as a thin-plate spline). Although we already know the form of the solution, we still have to determine the weights $\mathbf{w} = \begin{pmatrix} w_1 & w_2 & \dots \end{pmatrix}^t$ and the polynomial coefficients. Let q_i be a basis of the kernel of D^2 (e.g., monomials in x and y), the polynomial can then be expressed as $l(\mathbf{r}) = \sum_{i=0}^{p-1} l_i q_i(\mathbf{r})$. Let us define the vector of polynomial coefficients $\mathbf{l} = \begin{pmatrix} l_1 & l_2 & \dots \end{pmatrix}^t$. The determination of vectors $\mathbf{l}$ and $\mathbf{w}$ is done with the help of a linear system of equations. Let B be a matrix such that $b_{ij} = g\left(\left\| \mathbf{r}_i - \mathbf{r}_j \right\| \right)$, Q be a matrix such that $q_{ij} = q_i(\mathbf{r}_j)$. Let $\mathbf{v}$ be a vector with all the observed values. Then, the linear equation system given by

$$\begin{pmatrix} B + \lambda I & Q \\ Q^t & 0 \end{pmatrix} \begin{pmatrix} \mathbf{w} \\ \mathbf{l} \end{pmatrix} = \begin{pmatrix} \mathbf{v} \\ \mathbf{0} \end{pmatrix}, \quad (30)$$

gives the solution to the smoothing problem (i.e., denoising) here defined [16].

4.4. Algebraic approaches

If the support of the image $u(\mathbf{r})$ is a finite subset of $\mathbb{Z}^2$, then $u(\mathbf{r})$ can be represented as vector simply by lexicographically writing its components. In these conditions, it can be proved that the convolution with h can be computed as a matrix multiplication by the appropriate block-Toeplitz matrix H [12]. Hence, the simplified image formation model can be expressed as

$$\mathbf{v} = H\mathbf{u} + \boldsymbol{\varepsilon}, \quad (31)$$

Many matrix approaches can be followed now. Some of them are based on the minimization of a certain functional. Some others are based on some "reasonable" iterative algorithm that has been later proved to converge to a desired solution.

The image denoising and restoration problem can be expressed in matrix form as

$$\begin{aligned} &\hat{\mathbf{u}}^* = \arg\min \left\| W\hat{\mathbf{u}} \right\|^2 \\ &s.t. \left\| \mathbf{v} - H\hat{\mathbf{u}} \right\|^2 \le \varepsilon^2 \end{aligned}, \quad (32)$$

where W is a weight matrix. This problem formulation is known as the Tikhonov-Miller regularization. The solution of this optimization problem is the solution of the equation system [6]

$$\begin{aligned} &\hat{\mathbf{u}}^* = (\gamma W^t W + H^t H)^{-1} H^t \mathbf{v} \\ &\gamma^2 \left\| \left(H (W^t W)^{-1} H^t + \gamma I \right)^{-1} \mathbf{v} \right\|^2 - \varepsilon^2 = 0 \end{aligned} \quad (33)$$

If there is no blurring ($H = I$), and W is computed from a discrete version of the

Laplacian, then the equation system is a two-dimensional version of a smoothing spline [6].

If the image formation model is further simplified to $\mathbf{v} = H\mathbf{u}$, then the restored image could be found as the solution of the least-squares (LS) problem

$$\hat{\mathbf{u}}^* = \arg\min \left\| \mathbf{v} - H\hat{\mathbf{u}} \right\|^2 \tag{34}$$

It is well-known that the solution of this LS problem uses the pseudoinverse of H

$$\hat{\mathbf{u}}^* = H^+\mathbf{v} = (H^t H)^{-1} H^t \mathbf{v} \tag{35}$$

Some of the algebraic solutions proposed for the denoising/restoration problem make use of iterative equations that look reasonable and, later proving, that they converge to a solution having some desirable properties. This is the case of the Landweber or van Cittert iterative equation [12]

$$\begin{aligned} \hat{\mathbf{u}}^{(0)} &= \alpha \mathbf{v} \\ \hat{\mathbf{u}}^{(t)} &= \alpha \mathbf{v} + (I - \alpha H)\hat{\mathbf{u}}^{(t-1)} \end{aligned} \tag{36}$$

If there is no noise, this recursive equation can be shown to converge to

$$\hat{\mathbf{u}}^* = H^{-1}\mathbf{v} \tag{37}$$

This is the matrix homolog of the division by $h(\boldsymbol{\omega})$ presented at the end of Section 4.1. Hence, the inverse of matrix H does not exist if the Fourier transform of $h(\mathbf{r})$ contain zeros. Alternatively, the iterative step may be changed to

$$\hat{\mathbf{u}}^{(t)} = \hat{\mathbf{u}}^{(t-1)} + \alpha H^* \left(\mathbf{v} - H\hat{\mathbf{u}}^{(t-1)} \right) \tag{38}$$

which is known to converge to the pseudoinverse solution [12].

.

4.5. Maximum entropy: an information theoretic approach

Coming back to the matrix image observation model $\mathbf{v} = H\mathbf{u} + \boldsymbol{\varepsilon}$, the image can be restored in a maximum entropy sense. Entropy is a measure of the information provided by a probability distribution. For a discrete distribution p, the entropy $H(p)$ is defined as

$$H(p) = \sum_i p_i \log \frac{1}{p_i} \tag{39}$$

where p_i range over all the probabilities defined by the probability distribution [6]. Considering the restored image as a probability distribution, one way to find the restored image is by searching for the image providing maximum information and that is compatible with the observed measurements

$$\hat{\mathbf{u}}^* = \arg\max H(\hat{\mathbf{u}}) \quad s.t. \frac{1}{2}\left\| \mathbf{v} - H\hat{\mathbf{u}} \right\|^2 = \sigma_\varepsilon^2 \tag{40}$$

for some specified σ_ε^2. The solution of this problem is the solution of the equation [6]

$$\hat{\mathbf{u}}^* = \exp\left(-1 - \lambda H^t (\mathbf{v} - H\hat{\mathbf{u}}^*) \right) \tag{41}$$

where λ is the Lagrange multiplier so that the constraint $\frac{1}{2}\left\| \mathbf{v} - H\hat{\mathbf{u}} \right\|^2 = \sigma_\varepsilon^2$ is met.

4.6. Conclusions

As has been shown in this paper, image processing, and in general, signal processing is a fertile application field in which many mathematical tools find their way to applications. In fact, in many applications it is difficult to delineate the difference between applied mathematics and signal processing. In the example presented, image denoising and restoration, we have shown that different solutions come from very different mathematical perspectives. There is none that can be considered absolutely better than the rest, and usually which is the best technique depends on the data at hand. We hope that this paper encourage

mathematicians looking for applied fields, as well as practitioners looking for a more grounded algorithms.

Bibliography

[1] M. Bertozzi, A. Broggi, M. Cellario, A. Fascioli, P. Lombardi, and M. Porta, "Artificial Vision in Road Vehicles," vol. 90, pp. 1258-1271, 2002.

[2] M. A. Shahin, E. W. Tollner, R. W. McClendon, and H. R. Arabnia, "Apple classification based on surface bruises using image processing and neural networks," *Trans. of the American Society of Agricultural Engineers*, vol. 45, pp. 1619-1627, 2002.

[3] I. Bankman and S. Morcovescu, *Handbook of medical imaging*: Academic Press, 2000.

[4] J. J. Fernández, C. O. S. Sorzano, R. Marabini, and J. M. Carazo, "Image processing and three-dimensional reconstruction in electron microscopy of biological specimens," *IEEE Signal Processing Magazine*, vol. 23, pp. 89--94, 2006.

[5] R. C. Gonzalez and R. C. Woods, *Digital Image Processing*: Prentice-Hall, 2001.

[6] A. K. Jain, *Fundamentals of digital image processing*: Prentice-Hall, 1989.

[7] N. Paragios, Y. Chen, and O. Faugeras, *Handbook of mathematical models in computer vision*: Springer, 2006.

[8] J. C. Russ, *The image processing handbook*. Boca Raton, FL, USA: CRC Press, 1999.

[9] A. Papoulis, *Probability, Random Variables and Stochastic Processes*: McGraw Hill, 1991.

[10] T. A. Ceska, R. Henderson, J. M. Baldwin, F. Zemlin, E. Beckmann, and K. Downing, "An atomic model for the structure of bacteriorhodopsin, a seven-helix membrane protein," *Acta Physiol. Scand. Suppl.*, vol. 607, pp. 31-40, 1992.

[11] S. G. Mallat, *A wavelet tour of signal processing*: Academic Press, 1999.

[12] J. Biemond, R. L. Lagendijk, and R. M. Mersereau, "Iterative methods for image deblurring," *Proc. of the IEEE*, vol. 78, pp. 856-882, 1990.

[13] T. Chan, S. Esedoglu, F. Park, and A. Yip, "Total variation image restoration: overview and recent developments," in *Handbook of mathematical models in computer vision*, N. Paragios, Y. Chen, and O. Faugeras, Eds.: Springer, 2006, pp. 17-31.

[14] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259-268, 1992.

[15] G. Wahba, *Spline models for observational data*. Philadelphia, PA: SIAM, 1990.

[16] M. Arigovindan, M. Sühling, P. Hunziker, and M. Unser, "Variational image reconstruction from arbitrarily spaced samples: a fast multiresolution spline solution," *IEEE Trans. Image Processing*, vol. 14, pp. 450-460, 2005.