How to discretize the Total Variation of an image?

Lionel Moisan[1,*]

[1] Paris Descartes University, MAP5, CNRS UMR 8145, 45 rue des Saints-Pères 75006 Paris, France.

Because they are based on finite differences, usual discretizations of the Total Variation lead to aliased images. We propose a new discretization called *spectral total variation* that agrees with Shannon sampling principles and produces images that can be exactly interpolated. The quality improvement is illustrated experimentally in the case of image deblurring.

1 Introduction

Originally introduced in image processing by Rudin, Osher and Fatemi in [8], Total Variation (TV) has been used in many applications in image processing ever since, in particular for image restoration, blind deconvolution [4], resolution enhancement [6], decompression [1], inpainting, etc. and still recently in $u + v$ image decomposition models inspired from Meyer [7]. In all of the applications mentioned above, TV is used as a regularization term that permits to select, among several competing solutions, the "most reasonable one". In the case of deblurring, for example, an observed image u_0 has to be deconvolved by a know kernel k, so that the resulting image u satisfy $u_0 \simeq k \star u$ up to the (unknown) noise. In order to enforce uniqueness and avoid severe ringing artifacts, the restored image u can be defined as the solution of

$$\text{(P):} \quad \arg\min_u \|k \star u - u_0\|^2 + \lambda TV(u), \qquad \text{where} \qquad TV(u) = \int_{\mathbb{R}^2} |Du| \qquad (1)$$

and λ is an hyper-parameter that has to be adjusted in function of the noise level.

Solving (P) numerically requires two different steps. First, a discrete version of (1) has to be defined for numerical images. Second, an algorithm has to be found to solve the discrete minimization problem obtained from (P). Whereas the second step has received a lot of attention in the past few years with the emergence of dual algorithms [2, 3, 5], there has not been much work about the discretization issue. However, as long as the algorithm used in the second step converges, the solution only depends on the discretization chosen, which is consequently a true modeling issue. In Section 2, we point out the drawbacks of classical discretizations, and propose a new one in Section 3. We then show experiment in the case of image deblurring.

2 Discrete Total Variation

Let us consider a discrete real-valued image u defined on $\{0, ..N - 1\}^2$ (N being a positive integer) and extended in some way to $\mathbb{Z}^2$ (for example, by assuming that u is N-periodic along both axes). The classical way to discretize (1) is to define the discrete TV

$$TV_d(u) = \sum_{0 \le k,l < N} |Du|_d(k,l) \qquad \text{with} \qquad |Du|_d(k,l) = \sqrt{\Big(u(k+1,l) - u(k,l)\Big)^2 + \Big(u(k,l+1) - u(k,l)\Big)^2}. \qquad (2)$$

There are other ways to define discrete TVs by means of finite differences, with more symmetric schemes (with 3, 4 or 8 neighbors), or absolute values (l^1 norm), but (2) is the simplest one, and has the advantage of being Euclidean, which is a requirement for Chambolle's fast algorithm [3]. The reason why (2) is considered as a discretization of (1) relies in the notion of consistency, well-known in numerical analysis: if we consider a regular function $U : \mathbb{R}^2 \to \mathbb{R}$ and its discretizations $(k,l) \mapsto U_h(k,l) = U(kh, lh)$ for $h > 0$, we have

$$h^{-1} \cdot |DU_h|_d(kh, lh) \longrightarrow |DU(x,y)| \qquad \text{as} \qquad h \to 0, \ kh \to x, \text{ and } lh \to y.$$

The idea of making the grid step h tend to zero is well-founded in numerical analysis, where it often corresponds to a discretization parameter that can be chosen to represent more or less precisely the continuous world. In image processing, however, the pixel size cannot be chosen in general, so that assuming that it is "sufficiently small" may not be relevant. Another reason why discrete TVs based on finite differences are questionable is that we would like to achieve sub-pixel accuracy in the representation of images, since many image processing algorithms (e.g. geometric transforms) use sub-pixel interpolation. From the point of view of Sampling Theory, (2) is not a good discretization of (1), since even if u has been correctly sampled from an band-limited function according to Shannon Theorem [9], the squares involved in $|Du|_d$ introduce high frequencies that can be represented only if the sampling step is halved (not to mention the square root). As a consequence, the estimate $|Du|_d$ is aliased, as well as any image produced by a minimization process involving TV_d.

* Corresponding author E-mail: moisan@math-info.univ-paris5.fr, Phone: +33 144 553 544, Fax: +33 142 864 144

3   Spectral Total Variation

To reconciliate TV with Shannon Theory, we could define the TV of a discrete image u as being the exact (continuous) TV of its Shannon interpolate U, defined by

$$U(x,y) = \sum_{-\frac{N}{2} < p,q \leq \frac{N}{2}} \hat{u}(p,q) \exp\left(-2i\pi \frac{px+qy}{N}\right), \quad \text{where} \quad \hat{u}(p,q) = \sum_{0 \leq k,l < N} u(k,l) \exp\left(2i\pi \frac{pk+ql}{N}\right). \quad (3)$$

However, since $TV(U)$ cannot be computed exactly, we use a Riemann sum with an oversampling factor n, and define the *spectral total variation* of u (of order $n \geq 1$) by

$$STV_n(u) = \frac{1}{n^2} \sum_{0 \leq k,l < nN} |DU| \left(\frac{k}{n}, \frac{l}{n}\right). \quad (4)$$

What we expect from $STV_n(u)$ is to yield a good approximation of $TV(U)$ for *any* image u, since this term is to be used as a regularization term. Proposition 3.1 below shows that $n = 1$ is not a good choice, because controlling the gradient norm of U only at grid points does not permit to control the gradient norm *between* the grid points.

Proposition 3.1 *There is no constant C such that for any N and any discrete $N \times N$ image u,*

$$\int |DU|(x,y)\,dx\,dy \leq C \cdot STV_1(u),$$

where U is the Shannon interpolate of u given by (3).

Using $n = 2$, we obtain a new TV discretization that brings several improvements to classical schemes: independence with respect to the grid, compatibility with Shannon interpolation and sub-pixel operations (e.g. registration or stereo correlation), at the cost of a systematic use of Fourier Transforms, which is a common requirement in deconvolution problems. From a numerical point of view, one can use gradient descent schemes (∇STV_n is easily computed from (4)), or Chambolle's fast algorithm by means of the discrete gradient and divergence derived from (3). Fig. 1 below illustrates the ability of the spectral total variation to produce aliasing-free images, contrary to classical schemes based on finite differences.
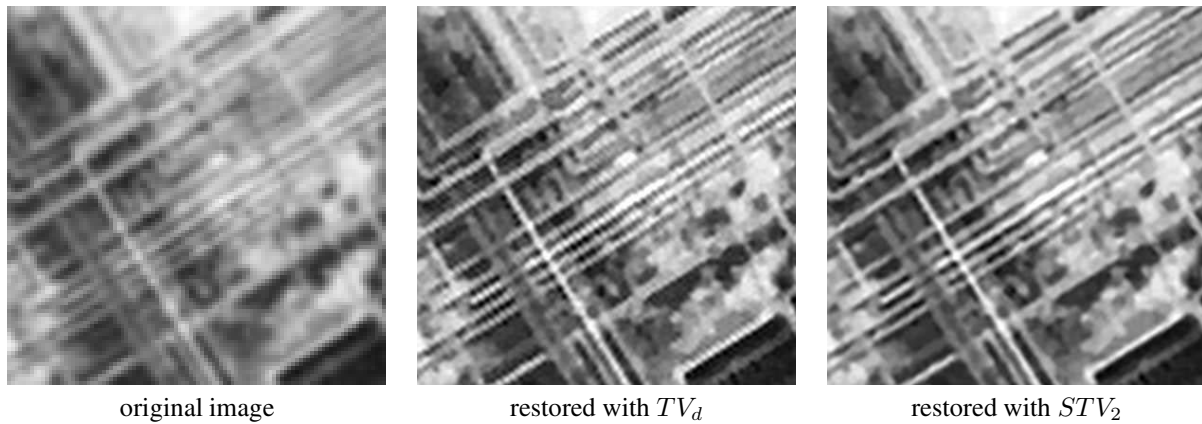

| original image | restored with TV_d | restored with STV_2 |

Fig. 1 An original image (left) undergoing Gaussian blur and white noise is deconvolved by means of TV regularization (P), then zoomed by a factor 4 using Shannon interpolation. With a finite difference TV discretization (middle), aliasing appear along contours. By reconciliating TV restoration and Shannon sampling Theory, the *spectral total variation* (right) avoids aliasing artifacts and yields a better quality image.

References

[1] F. Alter, S. Durand, J. Froment, "Adapted Total Variation for Artifact Free Decompression of JPEG Images", *J. Math. Imaging and Vision*, **23**(2), 199–211 (2005).
[2] J.L. Carter, "Dual methods for total variation-Based image restoration", Ph.D. thesis, U.C.L.A., 2001.
[3] A. Chambolle, "An Algorithm for Total Variation Minimization and Applications", *J. Math. Imaging and Vision*, **20**, 89–97 (2004).
[4] T. Chan, C. Wong, "Total Variation Blind Deconvolution", *IEEE Transactions on Image Processing*, **7**(3), 370–375 (1998).
[5] T.F. Chan, G.H. Golub, P. Mulet, "A nonlinear primal-dual method for total variation-based image restoration", *SIAM J. Sci. Comput.*, **20**(6), 1964–1977 (1999).
[6] F. Guichard, F. Malgouyres, "Total Variation based interpolation", *Proc. European Signal Processing Conf.*, **3**, 1741-1744 (1998).
[7] Y. Meyer, "Oscillating patterns in image processing and in some nonlinear evolution equations", (Lewis Memorial Lectures) University Lecture Series, 22, AMS, Providence, RI, 2001.
[8] L.I. Rudin, S. Osher, E. Fatemi, "Nonlinear total variation based noise removal algorithms", *Physica D*, **60**, 259–268 (1992).
[9] C.E. Shannon, "A mathematical theory of communication", *Bell System Technical Journal*, **27**, 379–423 and 623–656 (1948).