High order structural image decomposition by using non-linear and non-convex regularizing objectives

Abdourrahmane Atto, Grégoire Mercier

HAL Id: hal-01159214

https://hal.science/hal-01159214

Submitted on 2 Jun 2015

High Order Structural Image Decomposition by Using Non-Linear and Non-Convex Regularizing Objectives

Abdourrahmane Mahamane ATTO[*,a,1], Grégoire MERCIER[b,2]

[a] *University Savoy Mont Blanc - Polytech Annecy-chambéry - LISTIC,*
Laboratory of Computer Science, Systems, Information and Knowledge Processing
BP 80439 - F-74944 Annecy-le-Vieux Cedex - FRANCE
[b] *Institut MINES-TELECOM - TELECOM Bretagne, UMR CNRS 3192 Lab-STICC, Université*
Européenne de Bretagne, Technopôle Brest-Iroise, CS 83818, 29238 Brest Cedex 3 - FRANCE

Abstract

The paper addresses structural decomposition of images by using a family of non-linear and non-convex objective functions. These functions rely on ℓ^p quasi-norm estimation costs in a piecewise constant regularization framework. These objectives make image decomposition into constant cartoon levels and rich textural patterns possible. The paper shows that these regularizing objectives yield image texture-*versus*-cartoon decompositions that cannot be reached by using standard penalized least square regularizations associated with smooth and convex objectives.

Keywords: Non-Convex Optimization, Regularization, Image decomposition.

1. Introduction

Image decomposition in cartoon and texture patterns plays an important role in High Definition (HD) image processing [1, 2, 3]. For instance, recent high definition image engines processes images by decomposing them into smooth/texture/color components, enhancing every component separately, then fusing the results. This yields high quality images and videos, even for images issued from a low definition camera.

---

[1]Abdourrahmane.Atto@univ-smb.fr, [*] Corresponding author
[2]Gregoire.Mercier@telecom-bretagne.eu

Cartoon-*versus*-texture image decomposition is possible *via* certain regularization objectives. Since regularization functions operate differently depending on the image content, the main issue raised by this decomposition is the selection of best regularizers for splitting a given image into its cartoon (smooth[3] or piecewise smooth) part and a purely textural[4] part.

Most of the texture-structure decomposition methods available in the literature pertain to the class of *least square* objectives (quadratic forms) in the sense that these methods are variants of *Penalized Least Squares* (PLS). The discrete form of a PLS objective can be written in the following form, involving a penalty function Q_λ:

$$||\mathbf{x}[\mathbf{i},\mathbf{r}]-\mathbf{y}||_2^2 + 2 \sum_{\ell=i-r}^{i+r} Q_\lambda(|y_\ell|), \tag{1}$$

where λ is a parameter for tuning data fidelity, smoothness and roughness. The PLS problem inherits the convexity [4, 5, 6, 7, 8, 9, 10, 11, 12, 13] or non-convexity of Q_λ [14, 15, 16, 17, 18, 19, 20, 21] (the quadratic term involved in this problem is a convex function). Both convex and non-convex forms have shown relevancy in many image processing applications involving

- denoising and deconvolution [4, 5, 8, 10, 14, 15, 22, 23, 2, 24];

- cartoon/texture decomposition and image inpainting [6, 7, 10, 25];

- inverse problems and compressive sampling reconstruction [26, 27, 28]; *etc.*

Note that a PLS problem can be expressed in the image input space [5, 8], the Fourier domain [7], [10], the wavelet domain [4, 11, 13] or *via* Meyer's oscillatory functions [6, 12].

Some image regularizing objectives that do not pertain to the PLS class are stack filters [29, 30, 31]. Stack filters include the classes of weighted medians and some morphological type filters. In particular, non-convex objectives have

---

[3]We use the following definition of smoothness: a smooth 2D function is a function having continuous derivatives up to certain order. This function is thus (at least) continuous everywhere. A 2D function that is smooth, excepted on 2D curves, is called piecewise smooth function.

[4]The textural part of the image is then defined as the complementary with respect to difference operator of the carton part. This includes noise, for noisy image acquisitions.

been considered in [31] for the design of stack filters. These objectives apply on a binary level image representation obtained from a combination of threshold operators over several positive boolean functions. The optimal solution for these stacked objectives is then determined as the solution of a linear program (see [31] for details). This optimal solution is mainly relevant for denoising purposes, see [29, 30, 31], among other references.

The objective functions discussed in this paper do not involve any quadratic term. In this respect, their cartoon-*versus*-texture image decompositions are expected to differ significantly from those issued from PLS objectives.

These objectives operate structural image decomposition thanks to fractional ℓ^p quasi norms. They are non-smooth by construction so as to make sharp cartooning effect and rich texture rendering possible. The insight is to approximate the cartoon part of an image as a piecewise constant field by means of functionals that promote the number of occurrences for any given value in the pixel neighborhood. In this respect, these objectives operate in a way such that image segmentation can be seen as a particular case of image filtering (the gap between filtering and segmentation can be filled by tuning 1 parameter, the order p of the ℓ^p quasi-norm used).

The paper will shows that choosing non-smooth objectives is essential for high level cartooning (few gray levels, sharp edges) of many images since smoothness constraint (for instance, the one induced by gradient concerns in solving PLS problems) is known to yield over-smoothed results. Section 2 introduces the non-smooth and non-convex objectives proposed in the paper and derives the mathematical characterization of their solutions. Section 3 addresses the non-convex and non-linear programming for computing these solutions and provides the interpretation of these solutions with respect to image analysis. Finally, Section 4 concludes the work.

2. Non-smooth and non-convex objective with minimum ℓ^p quasi norm error

2.1. Problem formulation

Let $\mathbf{I} = \{\mathbf{I}(m,n)\}_{1 \leqslant m \leqslant M, 1 \leqslant n \leqslant N}$ denotes a gray-level image. A $(2r_0 + 1) \times (2r_1 + 1)$ pixel neighborhood centered on pixel $\mathbf{I}(m,n)$ is the set

$$\{\mathbf{I}(m+j,n+k)\}_{\substack{j=-r_0,-r_0+1,\ldots,r_0 \\ k=-r_1,-r_1+1,\ldots,r_1}}.$$

3

This set is written hereafter in a vectored form by noting

$$\mathbf{x[i,r]} = (x_{i-r}, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_{i+r}) \in \mathbb{R}^{2r+1}$$

where $r = ((2r_0 + 1)(2r_1 + 1) - 1)/2$ and $x_i = \mathbf{I}(m,n)$ (for instance, $i = n + (m-1) \times N$ if we consider line-by-line image scan). This reshaping has no consequence with respect to the regularization problem considered in the paper.

The regularizing objectives studied in the paper are given by

$$f_p(t, \mathbf{x[i,r]}) = \|\mathbf{x[i,r]} - t\|_{\ell p}^p = \sum_{\ell=i-r}^{i+r} |x_\ell - t|^p, \tag{2}$$

where $\mathbf{x[i,r]}$ is the neighborhood set given above and $t \in \mathbb{R}$. The paper addresses 3 issues:

- characterizing function f_p given by Eq. (2),

- solving the minimization problem:

$$\alpha_p(\mathbf{x[i,r]}) = \arg_t \min f_p(t, \mathbf{x[i,r]}), \tag{3}$$

  and

- providing an interpretation of the solutions of the problem given by Eq. (3) in terms of image cartoon and textural contents.

Some particular solutions of Problem Eq. (3) are widely used statistics in image processing and analysis: α_2 is the sample mean[5] and α_1 is the *sample median*[6] [32], the latter being unique because we have considered an odd number of samples in $\mathbf{x[i,r]}$.

The main differences between objective f_p and the above PLS objectives are

- f_p does not involve a quadratic error term (difference with respect to both convex and non-convex PLS approaches).

---

[5]Function f_2 is differentiable and convex: solve $\frac{\mathrm{d}f_2}{\mathrm{d}t}(t) = 0$ to derive that α_2 is the mean value: $\alpha_2 = \left(\sum_{\ell=i-r}^{i+r} x_\ell\right)/(2r+1)$.

[6]The sample median is such that the median value is constrained to belong to the sample set.

- vector **y** involved in PLS objective of Eq. (1) is reduced to a constant sequence with constant element t in f_p objective. The search for this optimal constant is restricted to a small pixel neighborhood, expressing local variations with respect to a piecewise constant image model.

While ignoring quadratic error terms such as those involved in PLS, we cannot guarantee a reasonable mean squared error and the resulting image can "deviate" significantly from the original image in terms of the mean squared error. However, it is well known that quadratic loss criteria do not promote structural image perceptions whereas our goal is to point out and differentiate structural contents of a heterogeneous image in terms of cartoon/texture decompositions.

Problem Eq. (3) is non-convex in general for $0 \leqslant p < 1$. For instance, when $p = 0$ (the pseudo-norm ℓ^0, associated with a counting process is then used), solution α_0 of this problem is the sample mode, this mode being non-unique for distributions having multiple sample points with the same and most occurring value.

In what follows, we are interested in Problem Eq. (3) for $0 < p < 1$ and without any additional constraints such as the linearization performed in [31] (introduction of Boolean operators at stack levels). Sections 2.2 and 2.3 below derive the mathematical characterization of solutions to this problem. These characterizations highlight that the solutions to the Problem Eq. (3) pertain to the sample data observed, as in the case of order statistics. However, these characterizations also show that the solutions of Problem Eq. (3) are not order statistics in the sense that the regularizing cost acts by modifying order statistics depending on the number of occurrences of every value.

In the rest of the paper, we will assume that $x_i \leqslant x_j$ if $i < j$ (ordering operation on the vectored pixel neighborhood) since once the pixel neighborhood $\mathbf{x}[\mathbf{i}, \mathbf{r}]$ is extracted, the objective function f_p is invariant with respect to re-orderings of data in $\mathbf{x}[\mathbf{i}, \mathbf{r}]$).

2.2. Solutions of Problem Eq. (3) are reached at the samples of $\mathbf{x}[\mathbf{i}, \mathbf{r}]$, when $0 < p < 1$

Theorem 1 *Assume that $x_k < x_{k+1}$. Function $f_p(\cdot, \mathbf{x}[\mathbf{i}, \mathbf{r}])$ defined by Eq. (2), when restricted to the set $]x_k, x_{k+1}[$, is concave on this interval. Its maximum is the unique real x^*, with $x_k < x^* < x_{k+1}$, satisfying:*

$$\sum_{\ell=i-r}^{k} \left| x_\ell - x^* \right|^{p-1} - \sum_{\ell=k+1}^{i+r} \left| x_\ell - x^* \right|^{p-1} = 0.$$

PROOF
The result follows by computing and checking the sign of the first two derivatives of $f_p(t, \mathbf{x[i,r]})$ for $t \in]x_k, x_{k+1}[$. □

The consequence of the concavity expressed in Theorem 1 is that in interval $[x_k, x_{k+1}]$, the <u>minimum</u> of function $f_p(\cdot, \mathbf{x[i,r]})$ is <u>either</u> $f_p(x_k, \mathbf{x[i,r]})$ <u>or</u> $f_p(x_{k+1}, \mathbf{x[i,r]})$:

$$\min_{t \in [x_k, x_{k+1}]} f_p(t, \mathbf{x[i,r]}) = \min\{f_p(x_k, \mathbf{x[i,r]}), f_p(x_{k+1}, \mathbf{x[i,r]})\}.$$

More precisely, by taking into account that

$$\min_{t \in]-\infty, x_{i-r}]} f_p(t, \mathbf{x[i,r]}) = f_p(x_{i-r}, \mathbf{x[i,r]})$$

and

$$\min_{t \in [x_{i+r}, +\infty[} f_p(t, \mathbf{x[i,r]}) = f_p(x_{i+r}, \mathbf{x[i,r]})$$

we derive from Theorem 1 that:

$$\min_{t \in \mathbb{R}} f_p(t, \mathbf{x[i,r]}) = \min\{f_p(x_k, \mathbf{x[i,r]})\}_{k=i-r,...,i+r} \tag{4}$$

and, for any $t \notin \mathbf{x[i,r]}$:

$$f_p(t, \mathbf{x[i,r]}) > \min\{f_p(x_k, \mathbf{x[i,r]})\}_{k=i-r,...,i+r}, \tag{5}$$

so that we can formalize:

Theorem 2 *The problem Eq. (3) is equivalent to:*

$$\beta_p(\mathbf{x[i,r]}) = \min\{f_p(x_k, \mathbf{x[i,r]})\}_{k=i-r,...,i+r}. \tag{6}$$

2.3. Characterization of solutions for the Problem Eq. (3) for $0 < p < 1$

From Theorem 2, we derive that function $f_p(\cdot, \mathbf{x[i,r]})$ reaches its minimum in the set $\mathbf{x[i,r]}$. Identifying the solutions of Problem Eq. (3) thus amounts to evaluating function f_p to the values available in this sample set. It follows that if we define

$$g_p(t, \mathbf{x[i,r]}) = \sum_{\ell=i-r}^{i+r} (a_{p,\ell} t + b_{p,\ell}) \, \mathbb{1}_{[x_\ell, x_{\ell+1}]}(t), \tag{7}$$

with

$$a_{p,\ell} = \begin{cases} \dfrac{f_p(x_{\ell+1},\mathbf{x[i,r]}) - f_p(x_\ell,\mathbf{x[i,r]})}{x_{\ell+1} - x_\ell} & \text{if} \quad x_\ell < x_{\ell+1} \\ 0 & \text{if} \quad x_\ell = x_{\ell+1} \end{cases} \tag{8}$$

and

$$b_{p,\ell} = \begin{cases} \dfrac{x_{\ell+1} f_p(x_\ell,\mathbf{x[i,r]}) - x_\ell f_p(x_{\ell+1},\mathbf{x[i,r]})}{x_{\ell+1} - x_\ell} & \text{if} \quad x_\ell < x_{\ell+1} \\ 0 & \text{if} \quad x_\ell = x_{\ell+1} \end{cases} \tag{9}$$

then we derive:

Theorem 3 *The problem Eq.* (3) *is equivalent to:*

$$\lambda_p\left(\mathbf{x[i,r]}\right) = \arg_t \min g_p(t,\mathbf{x[i,r]}), \tag{10}$$

where function g_p is defined by Eqs. (7), (8), (9) *above.*

Function g_p is a piecewise linear version of f_p: g_p can thus be decomposed into a linear programming problem involving $2r + 2$ linear constraints. In this respect, solutions of Problem (10) pertain to the vertices of the polyhedron of constraints (open polyhedron involving $2r + 1$ vertices $(x_\ell, g_p(x_\ell))$ for $\ell = i - r, \ldots, i + r$). Linear programming methods can thus be used to find solutions of Eq. (3). However, we are not concerned by such methods hereafter since r is small in our context (small pixel neighborhood): we evaluate function f_p at the vertices of this polyhedron and find the minimum by sorting the values of f_p on points $\{x_\ell\}_{\ell=i-r,\ldots,i+r}$.

In addition, we have:

Theorem 4 *Assume that:*

$$x_\ell \in \{0, 1, \ldots, D - 1\}, \ \text{for every } \ell = i - r, \ldots, i + r, \tag{11}$$

where $D \geqslant 2$ denotes the data dynamic (an 8-bit encoded image has a dynamic D=256). Let $0 < p \leqslant q \leqslant 1$. Then,

$$f_p(x_\ell, \mathbf{x[i,r]}) \leqslant f_q(x_\ell, \mathbf{x[i,r]}). \tag{12}$$

PROOF
On the one hand, if $\mathbf{x[i,r]}$ is a constant sequence, then we have: $f_p(x_\ell, \mathbf{x[i,r]})$ is zero for every $0 < p \leqslant 1$.

On the other hand, if there exist some $k, \ell \in \{i - r, \ldots, i + r\}$ such that $x_k \neq x_\ell$ and Eq. (11) *holds true, then we have:*

- *if $D = 2$, then $f_p(x_\ell, \mathbf{x}[\mathbf{i}, \mathbf{r}])$ counts the number of elements that are different from x_ℓ. In this respect, $f_p(x_\ell, \mathbf{x}[\mathbf{i}, \mathbf{r}]) = f_q(x_\ell, \mathbf{x}[\mathbf{i}, \mathbf{r}])$ for $p \neq q$.*

- *if $D > 2$, $f_p(x_\ell, \mathbf{x}[\mathbf{i}, \mathbf{r}])$, as a function of p, is derivable and has a non-negative derivative.*

This ends the proof. □

Since f_p is concave in all intervals $[x_\ell, x_{\ell+1}]$ for $\ell = i - r, i - r + 1, \ldots, i + r - 1)$, we derive that the piecewise linear functions g_p and f_1 verify: $g_p(t, \mathbf{x}[\mathbf{i}, \mathbf{r}]) \leqslant f_1(t, \mathbf{x}[\mathbf{i}, \mathbf{r}])$ for every p, with $0 < p \leqslant 1$. Since the minimum of f_1 is reached at x_i (sample median of $\mathbf{x}[\mathbf{i}, \mathbf{r}]$ by construction) and f_1 is convex, it follows that for many sample $\mathbf{x}[\mathbf{i}, \mathbf{r}]$ distributions, the sample median x_i inform us of the location of $\alpha_p(\mathbf{x}[\mathbf{i}, \mathbf{r}])$ when p is close to 1.

Furthermore, when p is close to 1, the extrema x_{i-r} and x_{i+r} are the worst candidates with respect to $\alpha_p(\mathbf{x}[\mathbf{i}, \mathbf{r}])$ location in general, except if these values have very large occurrences (extrema are not outliers if they have large occurrences).

One may think that Theorem 4 relates some order statistics…, however, value α_p is not an order statistic as highlighted in Figure 1. Indeed, value occurrences have a higher impact on the location of $\alpha_p(\mathbf{x}[\mathbf{i}, \mathbf{r}])$ than on the location of order statistics (see Figure 1).

The following theorem highlights the role played by sample x_k occurrences in the sample set $\mathbf{x}[\mathbf{i}, \mathbf{r}]$.

Theorem 5 *Let*

$$\mathscr{S}_k = \{\ell, i - r \leqslant \ell \leqslant i + r, x_\ell \neq x_k\}.$$

We have:

$$f_p(x_k, \mathbf{x}[\mathbf{i}, \mathbf{r}]) \leqslant f_p(x_m, \mathbf{x}[\mathbf{i}, \mathbf{r}]) + C_k |x_m - x_k|^p, \tag{13}$$

where $C_k = \#\mathscr{S}_k$ and $\#$ denotes set cardinality.

PROOF
We have

$$f_p(x_k, \mathbf{x}[\mathbf{i}, \mathbf{r}]) = \sum_{\ell = i-r}^{i+r} |x_\ell - x_k|^p = \sum_{\ell \in \mathscr{S}_k} |x_\ell - x_k|^p.$$

8

Ordered nearly uniform test data			$\alpha_{0.75}(\mathbf{x})$	$\alpha_1(\mathbf{x})$
Test 1	0 0 0 0 1 1 1 1 1 1	2	2 2 2 3 3 3 3 4 4 4	1	2
Test 2	0 0 1 1 1 1 1 1 2 2	2	2 2 2 3 3 3 3 4 4 4	2	2
Test 3	0 0 0 0 0 0 0 1 1 2	2	3 3 3 3 4 4 4 4 4 4	3	2

Functions $f_{0.75}(t,\mathbf{x})$ (blue), $g_{0.75}(t,\mathbf{x})$ (red) and $f_1(t,\mathbf{x})$ (black) for $t \in [-1,5]$
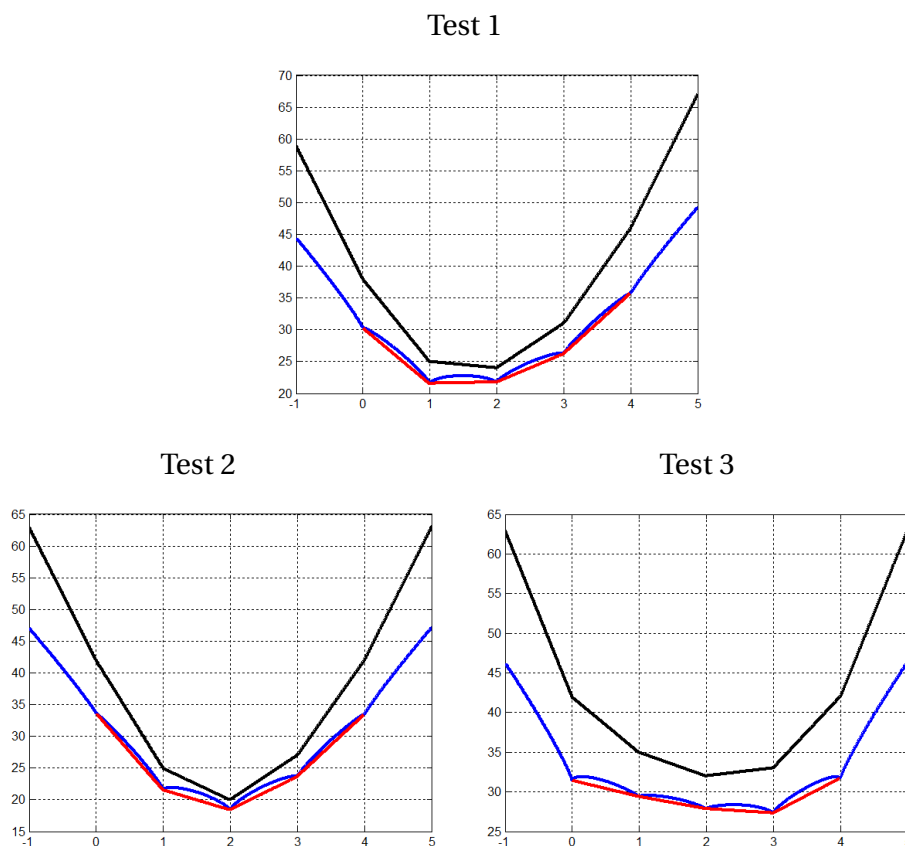

Figure 1: Test data $\mathbf{x}[\mathbf{i},\mathbf{r}]$, with $i = 11$ (median location) and $r = 10$ (samples on either side of the median), as well as their optimal α_p values when $p = 0.75$ and $p = 1$. Note that α_1 is the sample median of the test data. The test data are some non-negative integer values (rounded versions) generated from the uniform distribution on the interval $[0,4]$. The graphs represent function $f_{0.75}(t,\mathbf{x}[\mathbf{i},\mathbf{r}])$ (blue), $g_{0.75}(t,\mathbf{x}[\mathbf{i},\mathbf{r}])$ (piecewise linear version of f, in red) and $f_1(t,\mathbf{x}[\mathbf{i},\mathbf{r}])$ (in black). Depending on the occurrences of the different samples, the $\alpha_{0.75}$ can be the median (Test 2), smaller than the median (Test 1) or larger than the median (Test 3). Optima α_p provide us with a way of taking into account sample occurrences without deviating significantly from the median value.

Thus,

$$f_p(x_k, \mathbf{x}[\mathbf{i}, \mathbf{r}]) \leqslant \sum_{\ell \in \mathscr{S}_k} |x_\ell - x_m|^p + \sum_{\ell \in \mathscr{S}_k} |x_m - x_k|^p .$$

On the one hand,

$$\sum_{\ell \in \mathscr{S}_k} |x_\ell - x_m|^p \leqslant f_p(x_m, \mathbf{x}[\mathbf{i}, \mathbf{r}]).$$

On the other hand,

$$\sum_{\ell \in \mathscr{S}_k} |x_m - x_k|^p = C_k |x_k - x_m|^p .$$

This ends the proof. □

3. Programming and Interpretation of Eq. (3) from a signal processing perspective

3.1. Programming the solution of Eq. (3) for $0 < p < 1$

Theorems 2 provides an equivalent formulation of the Problem Eq. (3): $\alpha_p(\mathbf{x}[\mathbf{i}, \mathbf{r}]) = \beta_p(\mathbf{x}[\mathbf{i}, \mathbf{r}])$.

While computing $\alpha_p(\mathbf{x}[\mathbf{i}, \mathbf{r}])$ from Eq. (3) is not straightforward, computing $\beta_p(\mathbf{x}[\mathbf{i}, \mathbf{r}])$ from Theorem 2 is straightaway: it involves computing $2r + 1$ values $f_p(x_k, \mathbf{x}[\mathbf{i}, \mathbf{r}])$ for $k = i - r, \dots, i + r$, sorting these values and retrieving the index/indices where the minimum value of f_p is reached.

The computational complexity induced by the direct calculation amounts to

- computing a symmetric matrix having zero diagonal and dimension $(2r + 1) \times (2r+1)$, the $r \times (2r+1)$ non-zero elements being those $|x_k - x_\ell|^p$ located at row/column k and column/row $\ell \neq k$,

- identifying the row (resp. column) with smallest sum of entries in a row-wise (resp. column-wise) summation.

In what follows, quantity α_p introduced above will be associated to image regularization and is computed by using the following recursive approach (notation $\aleph_p$):

$$\aleph_p[i] = \arg_t \min f_p\left(t, \{\aleph_p[i - r], \dots, \aleph_p[i - 1], x_i, x_{i+1}, \dots, x_{i+r}\}\right). \quad (14)$$

10

From regularization perspective, this recursive computation consists in replacing (over-writing) the central value $\mathbf{I}(m, n)$ by $\aleph_p[m, n]$ before sliding the window to process the next pixel. Note that when considering a line-by-line image scan, then over-writing involved in $\aleph_p$ computation is

$$\aleph_p[m, n] = \arg_t \min$$

$$f_p\left(t, \left\{\begin{array}{cccccc}
\aleph_{p(m-r_0,n-r_1)} & \cdots & \aleph_{p(m-r_0,n-1)} & \aleph_{p(m-r_0,n)} & \aleph_{p(m-r_0,n+1)} & \cdots & \aleph_{p(m-r_0,n+r_1)} \\
\vdots & & \vdots & \vdots & \vdots & & \vdots \\
\aleph_{p(m-1,n-r_1)} & \cdots & \aleph_{p(m-1,n-1)} & \aleph_{p(m-1,n)} & \aleph_{p(m-1,n+1)} & \cdots & \aleph_{p(m-1,n+r_1)} \\
\aleph_{p(m,n-r_1)} & \cdots & \aleph_{p(m,n-1)} & \mathbf{I}(m,n) & \mathbf{I}(m,n+1) & \cdots & \mathbf{I}(m,n+r_1) \\
\mathbf{I}(m+1,n-r_1) & \cdots & \mathbf{I}(m+1,n-1) & \mathbf{I}(m+1,n) & \mathbf{I}(m+1,n+1) & \cdots & \mathbf{I}(m+1,n+r_1) \\
\vdots & & \vdots & \vdots & \vdots & \cdots & \vdots \\
\mathbf{I}(m+r_0,n-r_1) & \cdots & \mathbf{I}(m+r_0,n-1) & \mathbf{I}(m+r_0,n) & \mathbf{I}(m+r_0,n+1) & \cdots & \mathbf{I}(m+r_0,n+r_1)
\end{array}\right\}^{\dagger}\right)$$

where † stands for vectorization of a pixel neighborhood (whatever the vectorization order since $\aleph_p$ is invariant with respect to set ordering). It is worth mentioning that one can also choose a *Peano Hilbert* scan of the image, that is: a continuous curve (1D) joining all pixels of $\mathbf{I}$. From such an association, the over-writing concerns pixel preceding position (m, n) with respect to Peano Hilbert 2D grid ordering of $\mathbf{I}$.

Operation Eq. (14) provides a value $\aleph_p$ that is optimal in the ℓ^p error-norm sense for representing a pixel neighborhood. Note that $\aleph_0$, the recursive mode, is robust to monotonic changes such as brightness and/or contrast variations. Note also that $\aleph_1$, the recursive sample median, ensuring the least absolute deviations, is robust to the presence outliers. The value $\aleph_p$ derived from Eq. (14), as a compromise between $\aleph_0$ and $\aleph_1$, is thus expected to be robust to both monotonic changes and the presence of outliers. As shown in Section 3 below, applying Eq. (14) to image $\mathbf{I}$, having size $M \times N$, amounts estimating $\mathbf{I}$ by a regularized image $\aleph_p$ with the same size and containing $q \ll M \times N$ robust measurements: each $\aleph_p[m, n]$ having a large occurrence Q_i, with $\sum_{i=1}^q Q_i = M \times N$. This corresponds to a cartoon-like part of an image that is estimated from the q more robust ℓ^p error-norm statistics of the image. Other approaches relating robust optimum can be found in [33].

In practice, Eq. (14) can lead to multiple solutions. When $\aleph_p$ does not reduce to a singleton, indetermination raised by multiple solutions can be removed by:

- increasing/reducing the window size associated with the pixel neighborhood,

- making the objective convex by adding an appropriate penalty function,

- sorting and selecting one among the solutions by specifying an additional criterion/constraint.

In the current implementation, we have used the third issue and we set the solution to be applicable (only if multiple solutions occur):

- the median of $\{\aleph_p[m, n]\}$ if $\#\{\aleph_p[m, n]\}$ is odd,

- the median of the set $\{\aleph_p[m, n]\} \cup \{\mathbf{I}(m, n)\}$ when $\#\{\aleph_p[m, n]\}$ is even.

In this respect, the optimum $\arg_s f_1(s, \arg_t \min f_p(t, \bullet))$ is always associated with one of the samples involved in the data (one amongst the observed values, a remarkable property of the approach discussed in the paper).
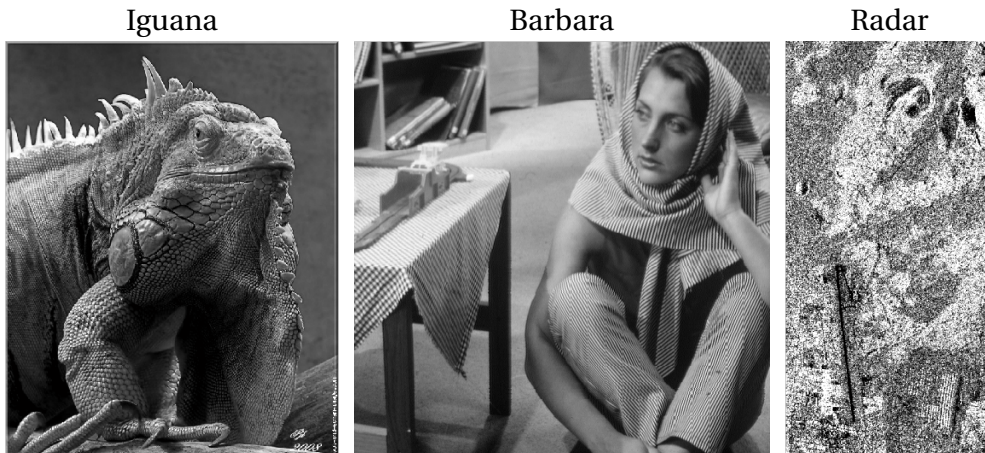
Iguana	Barbara	Radar

Figure 2: Test images: "Iguana" and "Barbara" are optical images. "Radar" is a synthetic aperture radar image.

3.2. Experimental results - Cartoon versus texture image decomposition

This section highlights the interest of exploiting non-convex and non-differentiable regularization functions in image cartoon/texture decomposition. A comparison is performed with respect to a convex regularization problem based on the *Total Variation* (TV) objective.

TV objectives are widely used for cartoon/texture decomposition. TV methods are based on the PLS problem introduced in Section 2.1: these methods involve an ℓ^2 based error-norm and an additional penalty function for tuning

the smoothness degree. TV cartoon/texture decomposition assumes an observation model having the form

$$\mathbf{x} = \mathbf{u} + \mathbf{v}$$

where $\mathbf{u}$ denotes the cartoon image and $\mathbf{v}$ is the texture image (noise is considered as a 'stochastic texture' in this paper and we do not consider an additional term for noise representation).

Optimal TV cartoon $\mathbf{u}^*$ and texture $\mathbf{v}^*$ images are defined to be solutions to the following minimization problem:

$$\begin{cases} \mathbf{u}^* = \arg_{\mathbf{u}} \min \mu T(\mathbf{x} - \mathbf{u}) + \lambda C(\mathbf{u}) \\ \mathbf{v}^* = \mathbf{x} - \mathbf{u}^* \end{cases} \tag{15}$$

where

- parameters λ, μ make tuning the amount of cartoon/texture possible,

- quantity $C(\mathbf{u})$ is a TV penalty function promoting cartoon features, this quantity being the energy of the gradient of $\mathbf{u}$ (note that $\mathbf{u}$ is assumed to follow from samples of a differentiable function):

$$C(\mathbf{u}) = \int_{\mathbb{R}} ||\nabla \mathbf{u}(t)||_2 \, \mathrm{d}t$$

and $T(\mathbf{x} - \mathbf{u}) = T(\mathbf{v})$,

$$T(\mathbf{v}) = \frac{1}{2} ||\mathbf{v}||_2^2$$

is an ℓ^2 error-norm representing the amount of extracted texture. The corresponding TV is called TV-ℓ^2, see [8] among other references.

Many variants of the above PLS based TV have been proposed in the literature. The variants include weighted Hilbert norms associated with Fourier (notation $\mathscr{F}$) domain representation and a selective filter G:

$$T(\mathbf{v}) = \frac{1}{2} ||W G \mathscr{F} \mathbf{v}||_2^2$$

where

- G is an all-pass filter and W is a fractional weight function having the form

$$W(\omega) = \frac{1}{\epsilon + ||\omega||}$$

in [10], the corresponding TV being called TV-H^{-1}, or

- G is associated to a Gabor filter (band-pass on texture frequencies when *a priori* knowledge is available on frequencies of interest associated with the textural information) in [7], the corresponding TV being called *TV Gabor Hilbert norm* or simply TV-G. It is usual considering in this TV model, a radial weight function

$$W(\omega) = 1 - e^{-\frac{(||\omega||-r)^2}{2\sigma^2}}.$$

The test images given in Figure 2 are selected to cover different texture types and different image acquisition modalities. "Iguana" and "Barbara" are optical images whereas "Radar" is a synthetic aperture radar image. "Iguana" image has a textured skin showing different type of cells and stripes on its body. In addition, some numerical characters and a signature have been affixed at the bottom-right of the image. "Barbara" image highlights grid-wise and linewise striped textures on clothes and tablecloth. In "Radar" image, the texture is a composition of forest, small urban structures and speckle (stochastic texture issued from reflection, refraction and diffraction of coherent radar waves on ground surface). The black line given in this radar image is an airstrip, the image also includes a volcano (top center of the image) and a part of the city of Goma in Congo (near the airstrip).

It is worth mentioning that performance evaluation depends on the *a priori* norm used: a quadratic regularization function yields optimal performances in the mean squared error sense. An absolute deviation strategy such as median regularization will yield optimal results when performance is evaluated with respect to ℓ^1 error-norm[7] and more generally, an ℓ^p regularization will lead to

---

[7]This equivalent is to say that *optimality is conditioned by the error-norm chosen for measuring performance*: the mean, the median and the mode filters are both optimal filters since these filters correspond respectively to the solutions $\aleph_2$ and $\aleph_1$ and $\aleph_0$. The difference between these optimal filters is the *a priori* error-norm used to assess their quality.

minimal ℓ^p error-norm by construction. Performance expressed in terms of PSNR is thus a biased performance measurement for the purpose of this section, the bias being related to the error-norm used for performance measurements (ℓ^2 for PSNR, which relates to *mean squared error*). Performance evaluation will thus be limited to visual perception of cartoon/texture when the decomposition is performed by using $\aleph_p$-value and the TV methods described above.

In all TV based experimental results, the regularization parameter λ is fixed at 1000 and different μ parameters are used. Among the effective methods for finding the optimal solution of this TV problem, we first consider the algorithm of [34] based on Split Bregman iterations, with no *a priori* on the texture of interest (in this respect, G is an all-pass filter).

Experimental results with different TV strategies (TV-G, TV-H^{-1} and TV-ℓ^2) are reported in Appendix A, in a context related to the definition of an ideal concept of cartoon.

Cartoon *versus* texture decompositions of the test images of Figure 2 are given for different $\aleph_p$-values and different TV μ regularization parameters in Figures 3, 4, 5, 6, 7, 8. For the sake of comparison, these images are displayed without histogram equalization (post-processing commonly used to enhance textural information). Indeed, histogram equalization performs adaptively with respect to the input image histogram and induces a biased comparison. Before comparing, it is worth highlighting that $\aleph_p$-value and TV capture different textural information and have different dynamics. Furthermore, the textural information quality is known to rely on personal perception: interpretation is left to the reader personal perception of cartoon and textures

"Iguana" cartoon images obtained by using $\aleph_p$-values, especially with 7×7 window size, are characterized by a high level cartoon description: all iguana cells and stripes have been removed from the cartoon part (see Figure 3) and placed in the textural part (see Figure 4). This yields a peeling-like effect of the iguana skin, with an important number of regions with constant values "under the skin" (cartoon) separated from the skin stripes and cells (texture). Results obtained from TV on "Iguana" are such that small μ parameters yield remaining textural cells in the cartoon images whereas large μ parameters tend to over-smooth the cartoon image (see Figure 3) without clearly highlighting the edges/segments between different regions. The best TV results are obtained for $\mu = 500$. The corresponding texture (see Figure 4) involves a smooth trend. In addition, both TV cartoon and texture images exhibit oscillating behavior for large μ parameters.

The same remarks as above hold true for "Barbara" image (see Figures 5, 6): while TV performs slightly better on this image than on "Iguana" image, TV performs less relevantly than $\aleph_p$-value for a complete extraction of textures having non-thin structures. Note also that TV continue to introduce oscillations in both cartoon and texture images. In contrast, textures captured by $\aleph_p$-value seem to be mainly associated with edges the image. For the specific case of "Barbara" image, substantial results (similar to the ones derived by $\aleph_p$-value) have been obtained in [6] and [35].

For "Radar" image, $\aleph_p$-value makes a segment-wise cartooning effect possible (see Figure 7) with no visible texture for large window sizes and when p decreases from 1 to 0.25: the deterministic structures and stochastic speckle involved in the original image are relocated in the texture image (see Figure 8). In contrast to this concise $\aleph_p$-value based cartoon/texture decomposition, TV "Radar" cartoon and texture images are impacted by a blurring effect (see Figure 7, 8). This blurring is probably due to the sensitivity of the TV to the smoothness degree of the input data, whereas "Radar" is highly heterogeneous. An adaptation of TV to multiplicative image models (case of "Radar" image) can be found in [36]. Note, however, that $\aleph_p$-value performs well on multiplicative type "Radar" image without the need of an adaptation.

When analyzing $\aleph_p$ based cartoon decompositions associated with different windows sizes and different p-values in Figures 3, 5, 7, we observe that no explicit limit from denoising to segmentation exists. The piecewise constant (segmentation-like) behavior induced by $\aleph_p$ regularization is illustrated in details in Figure 9 (line sections of test and cartoon images).

To conclude this section, a comparison of $\aleph_p$ and TV based decompositions is given in Figure 10 for a color image (richer visual rendering than a gray level image).

4. Conclusion

The paper has investigated the properties of some non-convex objectives for the decomposition of digital images in cartoon-*versus*-texture components. These objectives are associated with regularizing functions that can reach a high level of texture extraction and sharp cartoon edge preservation. The cartooning effect induced by these objectives ranges from denoising (small window sizes and when p parameter of the ℓ^p error-norm is close to 1) to image segmentation (large window sizes and when p is close to zero). Such a high

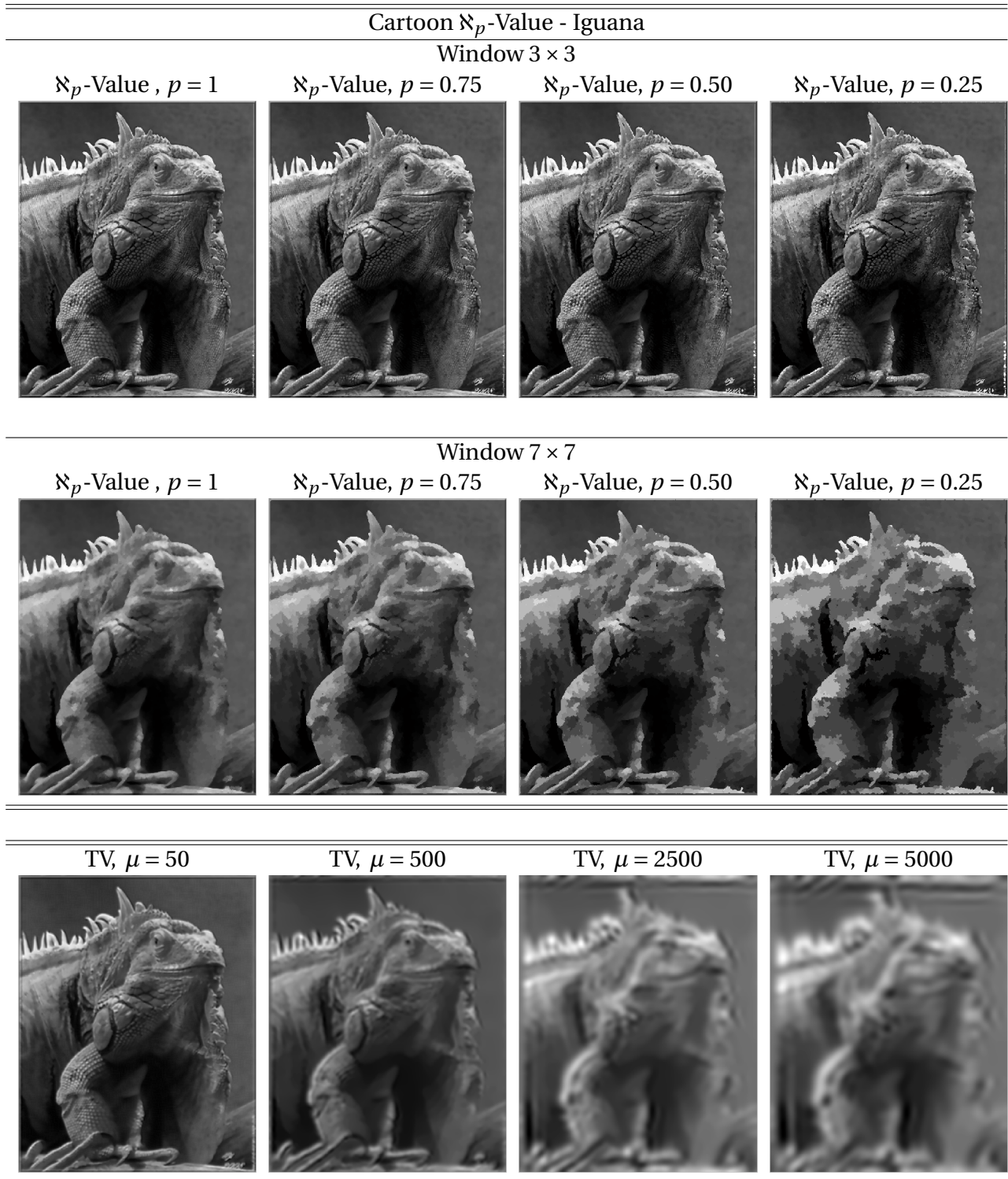| Cartoon $\aleph_p$-Value - Iguana			
Window 3×3			
$\aleph_p$-Value , $p = 1$	$\aleph_p$-Value, $p = 0.75$	$\aleph_p$-Value, $p = 0.50$	$\aleph_p$-Value, $p = 0.25$
Window 7×7			
$\aleph_p$-Value , $p = 1$	$\aleph_p$-Value, $p = 0.75$	$\aleph_p$-Value, $p = 0.50$	$\aleph_p$-Value, $p = 0.25$
TV, $\mu = 50$	TV, $\mu = 500$	TV, $\mu = 2500$	TV, $\mu = 5000$

Figure 3: Cartoon images obtained from different f_p objectives and different μ-parameters of TV on the Iguana image of Figure 2. The corresponding texture images are given in Figure 4.

17

Texture ℵ$_p$-Value - Iguana			
Window 3×3			
ℵ$_p$-Value , $p = 1$	ℵ$_p$-Value, $p = 0.75$	ℵ$_p$-Value, $p = 0.50$	ℵ$_p$-Value, $p = 0.25$

Window 7×7

ℵ$_p$-Value , $p = 1$	ℵ$_p$-Value, $p = 0.75$	ℵ$_p$-Value, $p = 0.50$	ℵ$_p$-Value, $p = 0.25$

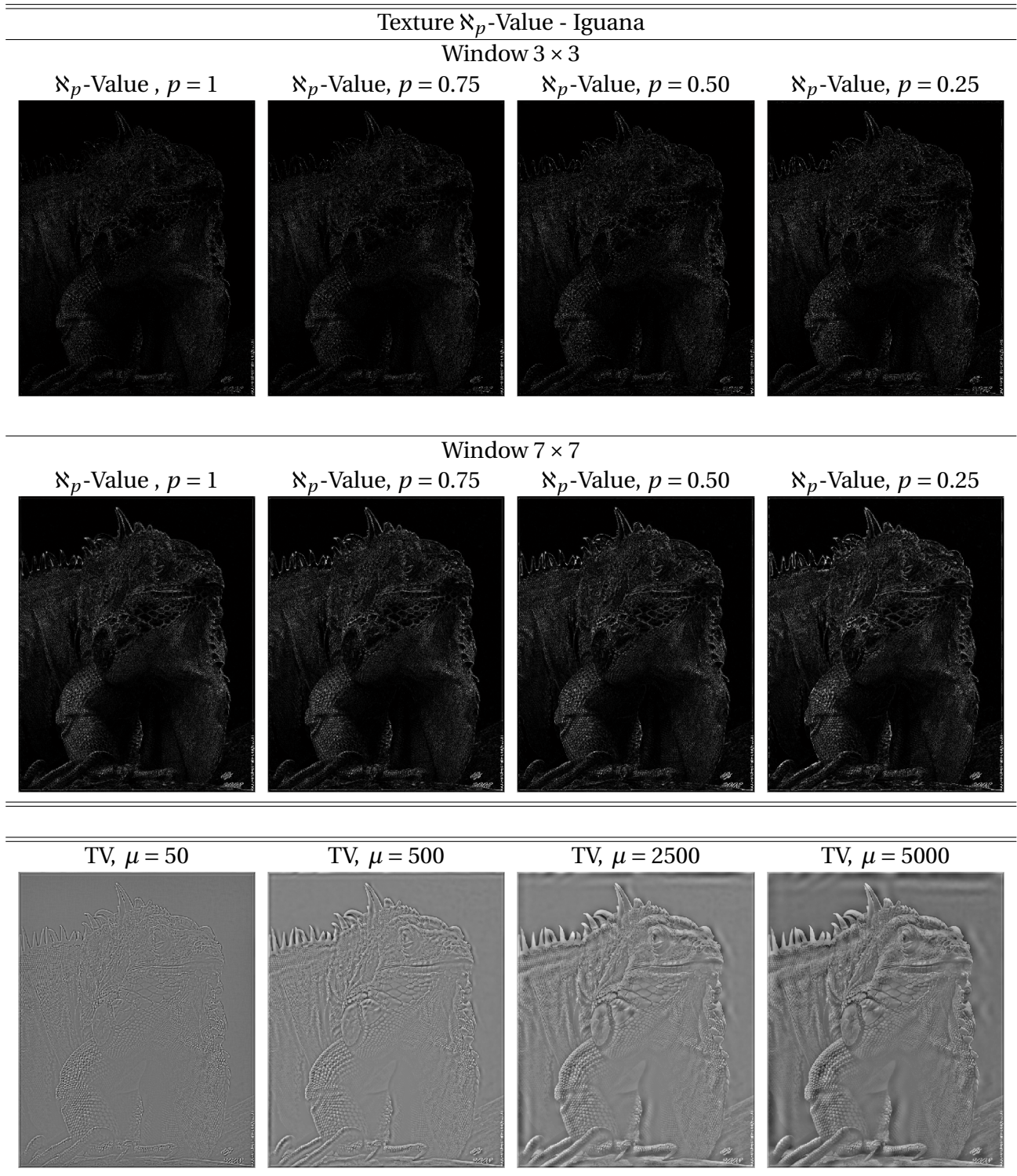| TV, $\mu = 50$	TV, $\mu = 500$	TV, $\mu = 2500$	TV, $\mu = 5000$

Figure 4: Texture images obtained from different f_p objectives and different μ-parameters of TV on the Iguana image of Figure 2. The corresponding cartoon images are given in Figure 3.

18

Cartoon ℵ_p-Value - Barbara			
Window 3 × 3			
ℵ_p-Value , $p = 1$	ℵ_p-Value, $p = 0.75$	ℵ_p-Value, $p = 0.50$	ℵ_p-Value, $p = 0.25$
Window 7 × 7			
ℵ_p-Value , $p = 1$	ℵ_p-Value, $p = 0.75$	ℵ_p-Value, $p = 0.50$	ℵ_p-Value, $p = 0.25$
TV, $\mu = 10$	TV, $\mu = 100$	TV, $\mu = 1000$	TV, $\mu = 10000$

Figure 5: Cartoon images obtained from different f_p objectives and different μ-parameters of TV on the Barbara image of Figure 2. The corresponding texture images are given in Figure 6.
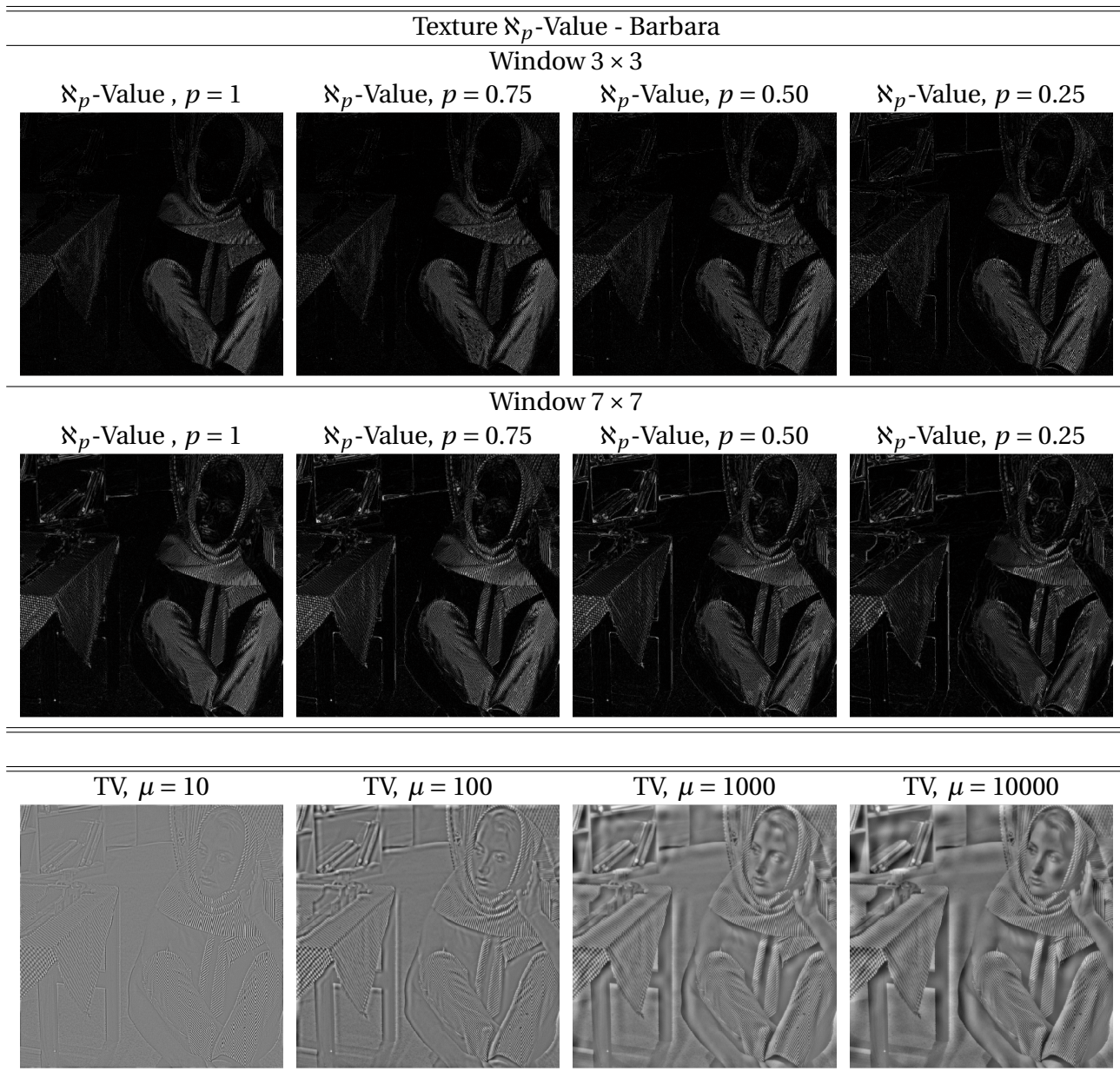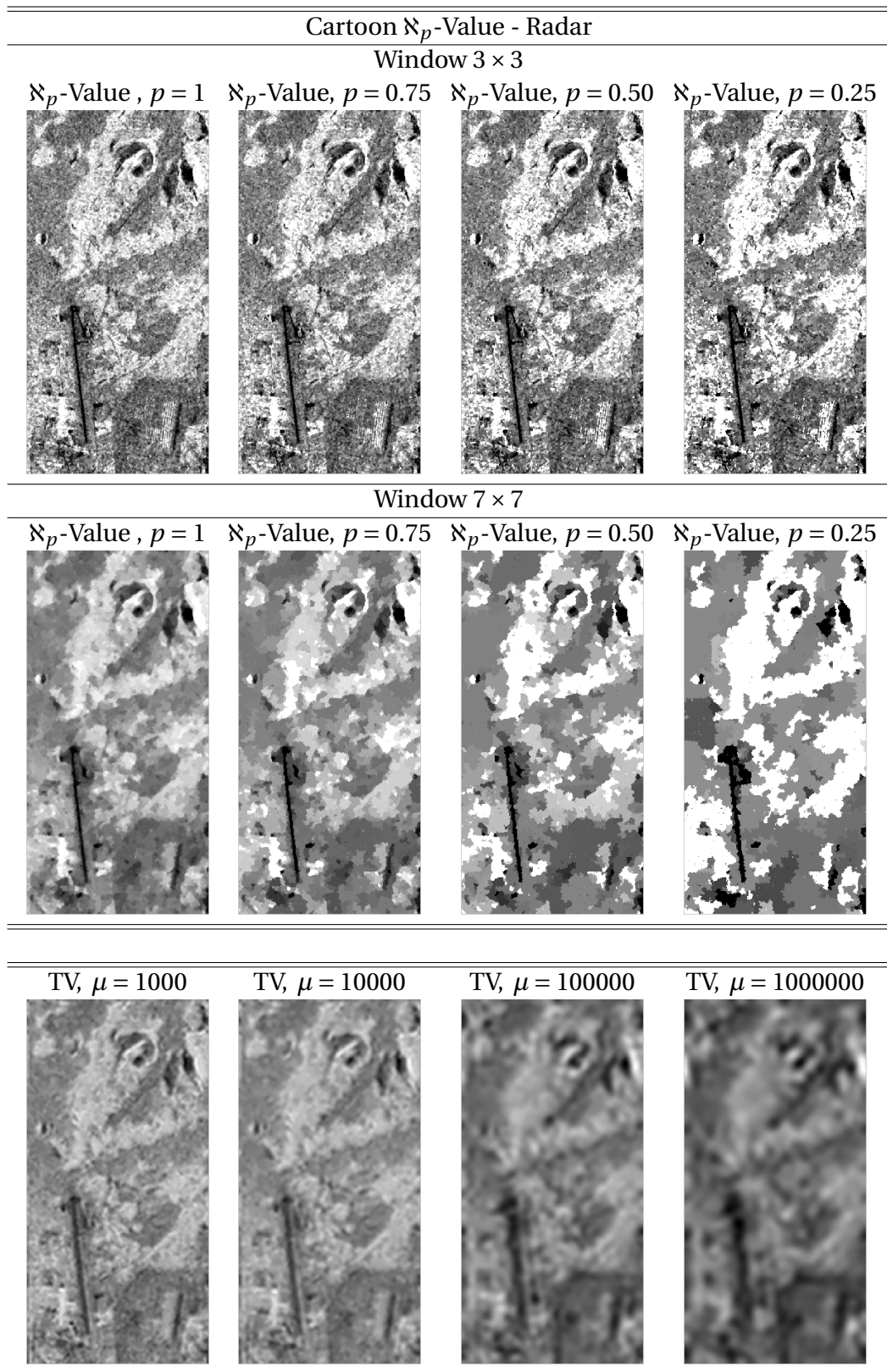
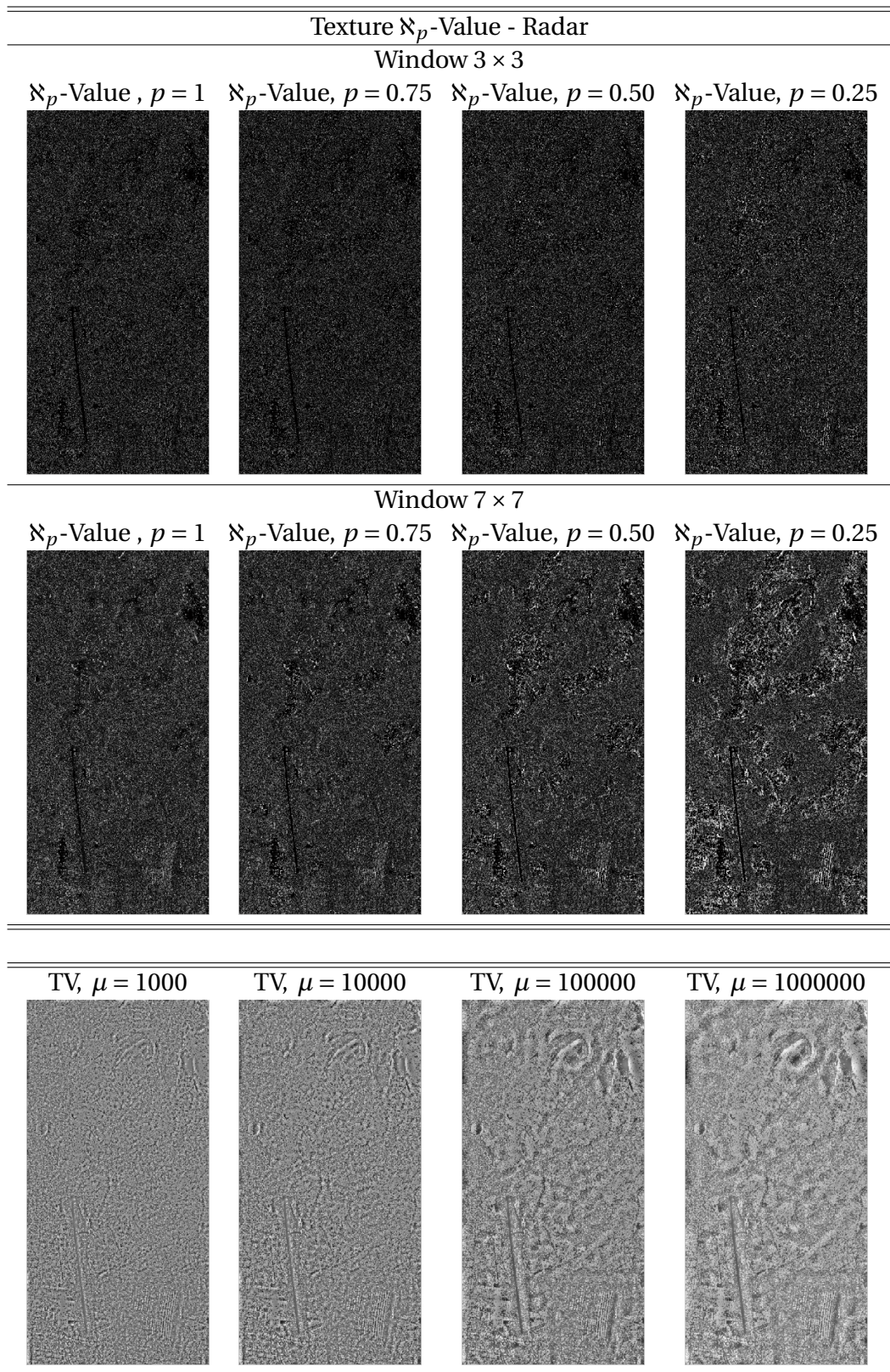| Texture ℵ_p-Value - Barbara			
Window 3×3			
ℵ_p-Value , $p = 1$	ℵ_p-Value, $p = 0.75$	ℵ_p-Value, $p = 0.50$	ℵ_p-Value, $p = 0.25$
Window 7×7			
ℵ_p-Value , $p = 1$	ℵ_p-Value, $p = 0.75$	ℵ_p-Value, $p = 0.50$	ℵ_p-Value, $p = 0.25$
TV, $\mu = 10$	TV, $\mu = 100$	TV, $\mu = 1000$	TV, $\mu = 10000$

Figure 6: Texture images obtained from different f_p objectives and different μ-parameters of TV on the Barbara image of Figure 2. The corresponding cartoon images are given in Figure 5.

Cartoon $\aleph_p$-Value - Radar			
Window 3×3			
$\aleph_p$-Value , $p = 1$	$\aleph_p$-Value, $p = 0.75$	$\aleph_p$-Value, $p = 0.50$	$\aleph_p$-Value, $p = 0.25$
Window 7×7			
$\aleph_p$-Value , $p = 1$	$\aleph_p$-Value, $p = 0.75$	$\aleph_p$-Value, $p = 0.50$	$\aleph_p$-Value, $p = 0.25$
TV, $\mu = 1000$	TV, $\mu = 10000$	TV, $\mu = 100000$	TV, $\mu = 1000000$

Figure 7: Cartoon images obtained from different f_p objectives and different μ-parameters of TV on the Radar image of Figure 2. The corresponding texture images are given in Figure 8.

Texture $\aleph_p$-Value - Radar

Window 3×3

$\aleph_p$-Value , $p = 1$	$\aleph_p$-Value, $p = 0.75$	$\aleph_p$-Value, $p = 0.50$	$\aleph_p$-Value, $p = 0.25$

Window 7×7

$\aleph_p$-Value , $p = 1$	$\aleph_p$-Value, $p = 0.75$	$\aleph_p$-Value, $p = 0.50$	$\aleph_p$-Value, $p = 0.25$

TV, $\mu = 1000$	TV, $\mu = 10000$	TV, $\mu = 100000$	TV, $\mu = 1000000$

Figure 8: Texture images obtained from different f_p objectives and different μ-parameters of TV on the Radar image of Figure 2. The corresponding cartoon images are given in Figure 7.

22
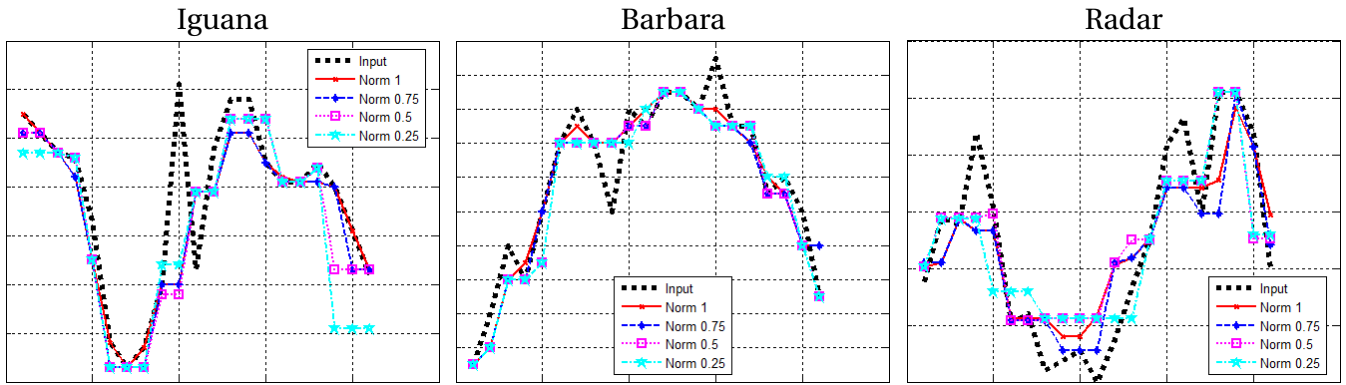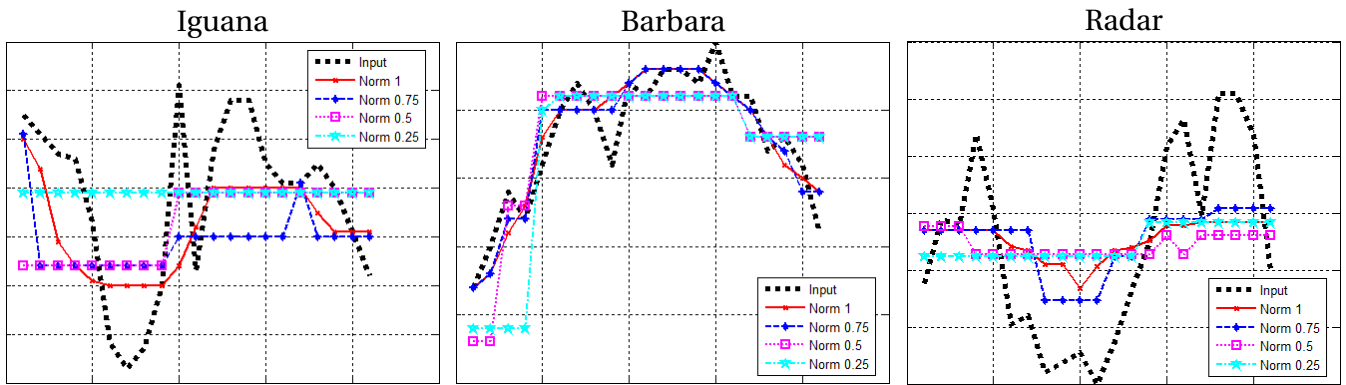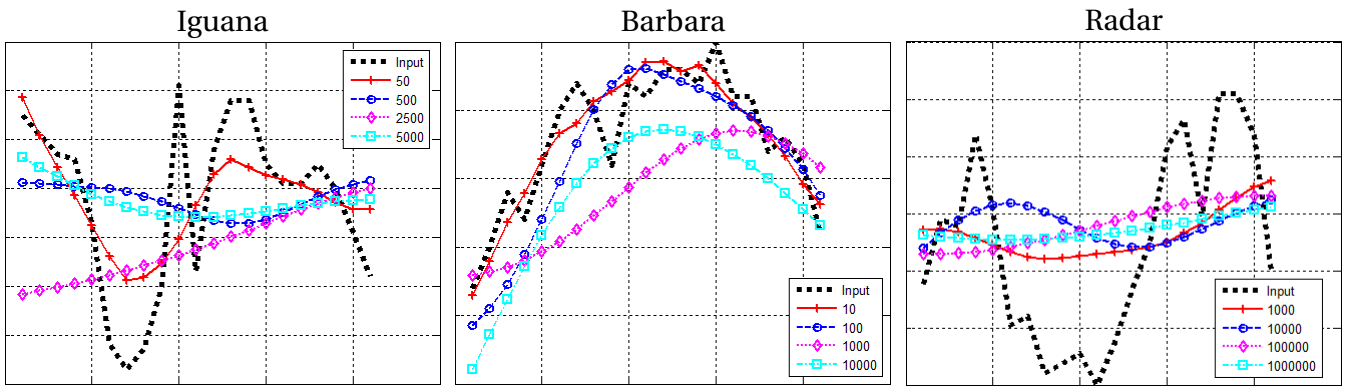
Figure 9: Line sections of the original test images in dashed-black-color (legend: Input), their cartoon $\aleph_p$-Value versions (legend: Norm p) and their TV versions for different μ-parameter (legend: value of parameter μ). The cartoon effect induced by the f_p objective results in piecewise constant curves whereas TV yields smooth curves).
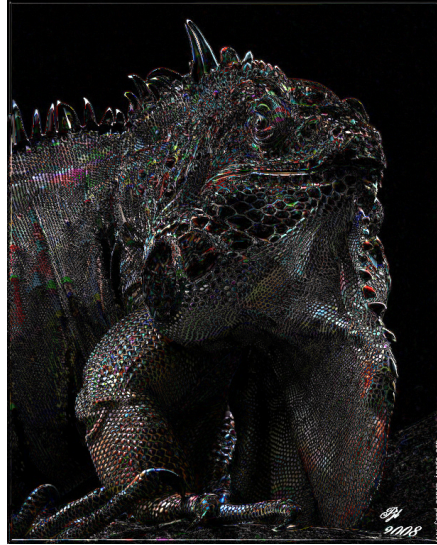
Iguana Original

Cartoon ℵ$_p$-Value, $p = 0.9$

Texture ℵ$_p$-Value, $p = 0.9$

Cartoon TV, $\lambda = 1000$, $\mu = 500$

Texture TV, $\lambda = 1000$, $\mu = 500$
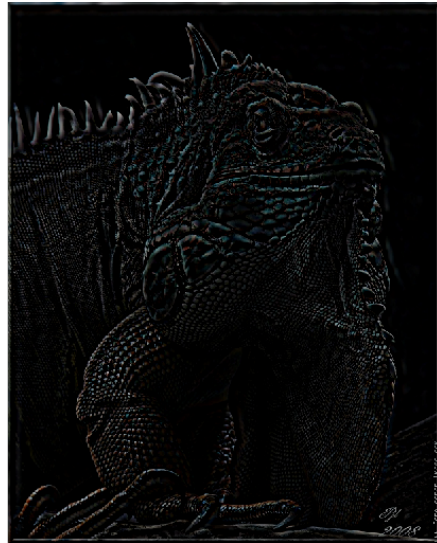
Figure 10: Global comparison of ℵ$_p$-Value and TV cartoon/texture decompositions on a color image. The window size used for ℵ$_p$-Value is 5 × 5 pixels. ℵ$_p$-Value cartoon is sharp and significant textural information has been extracted. TV cartoon representation is limited by the smoothness degree (over-smoothing occurs due to a differentiability assumption).

level cartooning and texture extraction can benefit modern high definition image processing which raises the need of separating an image into fundamental textured and cartoon parts, processing each part separately and fusing the corresponding processed images for increasing definition.

A first extension of this work concerns weighted functionals such as the ones involved in TV-G and TV-H^{-1}. Another extension of this work concerns multi-level/multi-scale decompositions such those described in [37, 38, 39, 40]. As far as the theoretical point of view is concerned, a non-straightforward extension of this work is a direct seeking of an ideal cartoon model where total variation is constrained by predefined pixel levels and a weight function associated with the number of occurrences of observed levels. An heuristic motivation of this issue is given in Appendix A.

Acknowledgment

[1] R. Wu, W. Wang, Y. Yu, Optimized synthesis of art patterns and layered textures, IEEE Transactions on Visualization and Computer Graphics 20 (3) (2014) 436–446.

[2] J. Bioucas-Dias, M. A. T. Figueiredo, Multiplicative noise removal using variable splitting and constrained optimization, IEEE Transactions on Image Processing 19 (7) (2010) 1720–1730.

[3] Z. Xu, H. Chen, S.-C. Zhu, J. Luo, A hierarchical compositional model for face representation and sketching, IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (6) (2008) 955–969.

[4] A. M. Atto, D. Pastor, G. Mercier, Wavelet shrinkage: unification of basic thresholding functions and thresholds, Signal, Image and Video Processing 5 (1) (2011) 11–28. doi:10.1007/s11760-009-0139-y.

[5] R. Chan, Y. Dong, M. Hintermüller, An efficient two-phase L1-TV method for restoring blurred images with impulse noise, IEEE Transactions on Image Processing 19 (4).

[6] A. Buades, T. Le, J. M. Morel, L. Vese, Fast cartoon + texture image filters, IEEE Transactions on Image Processing 19 (8) (2010) 1978–1986.

[7] J.-F. Aujol, G. Gilboa, T. Chan, S. Osher, Structure-texture image decomposition, modeling, algorithms, and parameter selection, International Journal of Computer Vision 67 (1) (2006) 111–136.

[8] J.-F. Aujol, G. Gilboa, Constrained and snr-based solutions for TV-hilbert space image denoising, J. Math. Imaging Vis. 26 (1-2) (2006) 217–237.

[9] J.-L. Starck, M. Elad, D. Donoho, Image decomposition via the combination of sparse representations and a variational approach, IEEE Transactions on Image Processing 14 (10) (2005) 1570–1582.

[10] S. Osher, A. Solé, L. Vese, Image decomposition and restoration using total variation minimization and the H1, Multiscale Modeling & Simulation 1 (3) (2003) 349–370.

[11] M. Wakin, J. Romberg, H. Choi, R. Baraniuk, Image compression using an efficient edge cartoon + texture model, in: Data Compression Conference, 2002. Proceedings. DCC 2002, 2002, pp. 43–52.

[12] Y. Meyer, Oscillating Patterns in Image Processing and Nonlinear Evolution Equations: The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures, American Mathematical Society, Boston, MA, USA, 2001.

[13] A. Antoniadis, J. Fan, Regularization of wavelet approximations, Journal of the American Statistical Association 96 (455) (2001) 939 – 955.

[14] M. Hintermuller, T. Wu, Nonconvex TV^q-models in image restoration: Analysis and a trust-region regularization–based superlinearly convergent solver, SIAM Journal on Imaging Sciences 6 (3) (2013) 1385–1415.

[15] P. Ochs, A. Dosovitskiy, T. Brox, T. Pock, An iterated l1 algorithm for non-smooth non-convex optimization in computer vision, in: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, CVPR '13, IEEE Computer Society, Washington, DC, USA, 2013, pp. 1759–1766.

[16] R. Chartrand, Nonconvex splitting for regularized low-rank + sparse decomposition, IEEE Transactions on Signal Processing 60 (2012) 5810–5819.

[17] X. Chen, Smoothing methods for nonsmooth, nonconvex minimization, Mathematical Programming 134 (1) (2012) 71–99.

[18] K. Bredies, D. A. Lorenz, Regularization with non-convex separable constraints, Inverse Problems 25 (8) (2009) 085011.

[19] M. Nikolova, M. K. Ng, S. Zhang, W.-K. Ching, Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization, SIAM Journal on Imaging Sciences 1 (1) (2008) 2–25.

[20] M. Grasmair, M. Haltmeier, O. Scherzer, Sparse regularization with ℓ^q penalty term, Inverse Problems 24 (2008) 055020 (13pp).

[21] R. Chartrand, Exact reconstruction of sparse signals via nonconvex minimization, IEEE Signal Processing Letters 14 (10) (2007) 707–710.

[22] N. Pustelnik, C. Chaux, J. Pesquet, Parallel proximal algorithm for image restoration using hybrid regularization, Image Processing, IEEE Transactions on 20 (9) (2011) 2450 – 2462.

[23] J.-C. Baritaux, K. Hassler, M. Unser, An efficient numerical method for general regularization in fluorescence molecular tomography, Medical Imaging, IEEE Transactions on 29 (4) (2010) 1075 – 1087. doi:10.1109/TMI.2010.2042814.

[24] J. P. Oliveira, J. M. Bioucas-Dias, M. A. T. Figueiredo, Adaptive total variation image deblurring: A majorization-minimization approach, Signal Processing 89 (9) (2009) 1683 – 1693.

[25] M. Elad, J.-L. Starck, P. Querre, D. Donoho, Simultaneous cartoon and texture image inpainting using morphological component analysis (mca), Applied and Computational Harmonic Analysis 19 (3) (2005) 340 – 358, computational Harmonic Analysis - Part 1.

[26] G. Peyré, Best basis compressed sensing, IEEE Transactions on Signal Processing 58 (5) (2010) 2613 – 2622.

[27] D. L. Donoho, Compressed sensing, IEEE Transactions on Information Theory 52 (12) (2006) 1289 – 1306.

[28] E. J. Candes, T. Tao, Decoding by linear programming, IEEE Transactions on Information Theory 51 (12) (2005) 4203 – 4215.

[29] M. E. Buemi, A. C. Frery, H. S. Ramos, Speckle reduction with adaptive stack filters, Pattern Recognition Letters 36 (0) (2014) 281 – 287.

[30] R. Lukac, B. Smolka, K. N. Plataniotis, A. N. Venetsanopoulos, Selection weighted vector directional filters, Computer Vision and Image Understanding 94 (1-3) (2004) 140 – 167, special Issue: Colour for Image Indexing and Retrieval.

[31] C. Savin, M. Ahmad, M. Swamy, Lp norm design of stack filters, IEEE Transactions on Image Processing 8 (12) (1999) 1730–1743.

[32] B. Justusson, Median filtering: Statistical properties, in: Two-Dimensional Digital Signal Prcessing II, Vol. 43 of Topics in Applied Physics, Springer Berlin Heidelberg, 1981, pp. 161–196.

[33] D. Bertsimas, D. Brown, C. Caramanis, Theory and applications of robust optimization, SIAM Review 53 (3) (2011) 464–501.

[34] J. Gilles, S. Osher, Bregman implementation of Meyer's g-norm for cartoon+ textures decomposition, UCLA CAM Report.

[35] G. Evangelopoulos, P. Maragos, Image decomposition into structure and texture subcomponents with multifrequency modulation constraints, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2008, pp. 1–8.

[36] G. Aubert, J. Aujol, A variational approach to removing multiplicative noise, SIAM Journal on Applied Mathematics 68 (4) (2008) 925–946.

[37] G. Gilboa, A total variation spectral framework for scale and texture analysis, SIAM Journal on Imaging Sciences 7 (4) (2014) 1937–1961.

[38] M. Hidane, O. Lézoray, A. Elmoataz, Nonlinear multilayered representation of graph-signals, Journal of Mathematical Imaging and Vision 45 (2) (2013) 114–137.

[39] J. Gilles, Multiscale texture separation, Multiscale Modeling & Simulation 10 (4) (2012) 1409–1427.

[40] E. Tadmor, S. Nezzar, L. Vese, A multiscale image representation using hierarchical (BV, L2) decompositions, Multiscale Modeling & Simulation 2 (4) (2004) 554–579.

Appendix A. Ideal cartoon, L^p spaces and smoothness

What is cartoon? How to define a texture? The computer vision and image processing literature have proposed several definitions of these image components from regularity (cartoon) and intensity variations with pseudo-periodicity assumptions (texture).

The "regularity-*versus*-variations" assumption on cartoon and texture parts has naturally led researchers to use gradient methods for finding cartoon as an optimal solution of a split-functional defined on Hilbert spaces (see Section 3.2 for details on TV based approaches). Note that one difficulty induced by gradient methods is the non-trivial choice of gradient descent steps: these steps depend on the regularity of the experimental objective, the one associated with an arbitrary input image. However, this is not the main limitation of gradient based objectives in effective cartoon identification: a fine tuning of gradient steps is always possible, at the cost of a slow convergence rate.

The main limitation of gradient based methods is probably the differentiability constraint imposed to the cartoon image. This differentiability assumption does not comply with the perception of an ideal cartoon image: Figure A.11 (left image, "Snow White") provides an example of an ideal cartoon image: there is no visible texture on this image whereas the image presents many singularities in terms of sharp edges and contours. A cartoon image can thus be a non-continuous (non-smooth) 2D field.

"Snow White" ideal cartoon image	"Snow White" with texture (hairs)

Figure A.11: [Left] painted cartoon image without texture surface (edges and contours are the sole singularities for this image). [Right] cartoon-textured image where a realistic hair texture has been added to the left image (edges and hair texture are the singularities of the latter image).

We might think that a piecewise smooth/constant field model is more appropriate than a smooth one (smoothness assumes at least continuity everywhere). Indeed, an ideal edge/contour model is that of a surface step function: this function is not continuous at all and thus, cannot be smooth everywhere. Note that "Snow White" image is made from a dozen of colors associated with image partition in about thirty regions. This reinforces our feeling on the relevancy of a piecewise constant field model with small number of gray levels for defining ideal cartoon model.

Some 2D fields are universally accepted as textured elements: natural hairs for instance. In order to derive a test image with identified cartoon and texture parts, we have removed the painted cartoon hairs of "Snow White" image and replaced them by natural hairs. The result is given in Figure A.11 (right side).

If one models the cartoon part of Figure A.11 (right) as a smooth, differentiable field, then iterative optimization procedures based on this smoothness assumption will always find smooth solutions. Different TV methods involving this smoothness assumption have been applied to the decompose "Snow White" (the one with textured hairs) image in cartoon and texture parts and the results are given in Figure A.12. Comparison is provided with $\aleph_p$-values in the same figure. Note that certain TV based algorithms find texture over some constant fields of 'Snow White" image.

From the results, we suggest that forcing a "piecewise constant" or a "piecewise smooth" behavior is essential for accurate cartoon definition. When reasoning in $\mathbb{R}^2$ instead of a discrete field, such a cartoon image is a piecewise smooth field associated with two types of singularities: countable union of singletons (Type-1 singularities) and curved $\mathbb{R}^2$ singularities (Type-2).

Type-1 singularities are cartoon-negligible whereas Type 2 singularities are not cartoon-negligible since they are associated with cartoon frontiers. Type-2 singularities (curves of $\mathbb{R}^2$) play an important role in visual perception: they cannot be considered as negligible sets since they represent edges of cartoon objects. Note that the latter singularities are negligible in standard integration theory and in TV decomposition functionals C and T (as a consequence) whereas the contrary is suggested by the above remark. This opens some new prospects regarding TV image decomposition: building cartoon functionals that involves a mixture of regular and curvi-singular functionals.
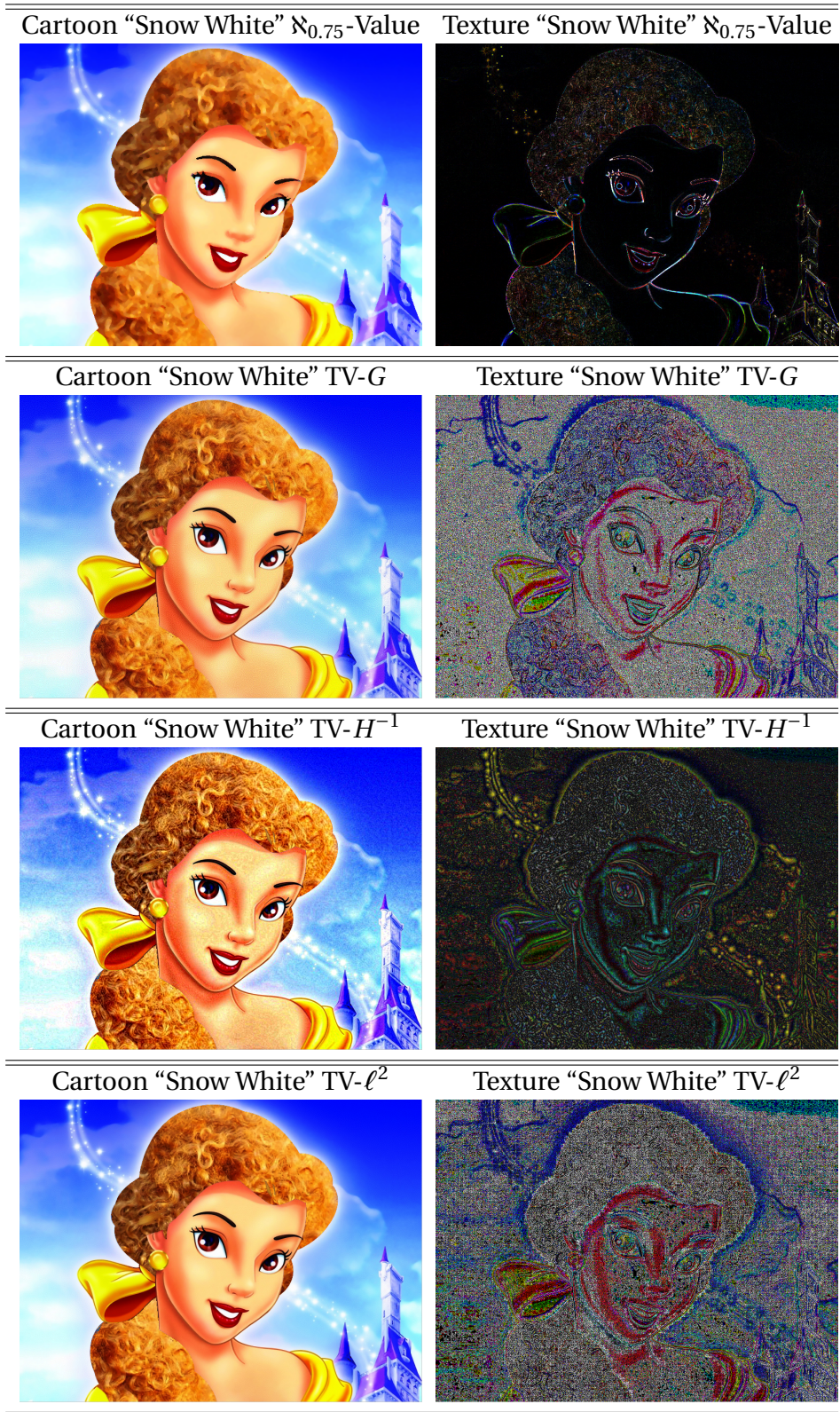
Figure A.12: Cartoon-*versus*-Texture decompositions of 'Snow White' image of Figure A.11 (right). Decomposition associated with $\aleph_{0.75}$-value find zeros over constant cartoon fields whereas certain TV based algorithms find texture over some of these constant cartoon fields.