View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by HAL-Ecole des Ponts ParisTech

A spatial regularization approach for vector quantization

Caroline Chaux, Anna Jezierska, Jean-Christophe Pesquet, Hugues Talbot

To cite this version:

Caroline Chaux, Anna Jezierska, Jean-Christophe Pesquet, Hugues Talbot. A spatial regularization approach for vector quantization. Journal of Mathematical Imaging and Vision, Springer Verlag, 2011, 41, pp.23-38. <hal-00530369>

HAL Id: hal-00530369

https://hal.archives-ouvertes.fr/hal-00530369

Submitted on 28 Oct 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

manuscript No.
(will be inserted by the editor)

A spatial regularization approach for vector quantization

Caroline Chaux · Anna Jezierska · Jean-Christophe Pesquet · Hugues Talbot

Received: date / Accepted: date

Abstract Quantization, defined as the act of attributing a finite number of levels to an image, is an essential task in image acquisition and coding. It is also intricately linked to image analysis tasks, such as denoising and segmentation. In this paper, we investigate vector quantization combined with regularity constraints, a little-studied area which is of interest, in particular, when quantizing in the presence of noise or other acquisition artifacts. We present an optimization approach to the problem involving a novel two-step, iterative, flexible, joint quantizing-regularization method featuring both convex and combinatorial optimization techniques. We show that when using a small number of levels, our approach can yield better quality images in terms of SNR, with lower entropy, than conventional optimal quantization methods.

Keywords Vector quantization · Convex optimization · Combinatorial optimization · Proximal methods · Graph cuts · Image coding · Compression · Information theory · Entropy · Segmentation · Denoising · Regularization

1 Introduction

Quantization is a fundamental task in digital image processing and information theory [1]. It plays a prominent role in early processing stages such as image digitization, and it is essential in lossy coding. It bears close resemblance to high level tasks such as denoising, segmentation, and data classification. In particular, quantizing a grey scale image in Q levels can be viewed as a classification or segmentation of the image in Q areas following an intensity homogeneity criterion. Each segmented area then corresponds to a decision class of the quantizer.

A classical solution for designing an optimal quantizer of a monochrome image is provided by the celebrated Lloyd-Max (LM) algorithm [2,3]. An extension to the general vector case is the Linde–Buzo–Gray (LBG) algorithm [4]. The LBG algorithm proceeds iteratively by alternatively optimizing codevectors and decision classes so as to minimize a flexible quantization error measure. It is known to present good convergence properties in practice [5,6]. However, one drawback is the lack of spatial regularity of the quantized image. Spatially smooth properties may be useful in low-rate compression when using advanced coding algorithms (e.g based on run length, differential or multiresolution techniques), especially in the context of medical and low bit-rate video compression applications like compression of confocal laser scanning microscopy image sequences [7] or mobile television [8]. It may also be of interest for quantizing images featuring noise. In the latter case, quantization can be viewed as a means for denoising discrete-valued images that are piecewise constant.

Since the LBG algorithm is closely related to K-means, which are widely used in data classification, a possibility to enforce spatial smoothness of the quantized image would be to resort to fuzzy C-means clustering techniques and their extensions [9]. These algorithms are however based on local measures of smooth-

This work was supported by the Agence Nationale de la Recherche under grant ANR-09-EMER-004-03.

Université Paris-Est
Lab. Informatique Gaspard Monge, UMR CNRS 8049,
Champs-sur-Marne, 77454 Marne-la-Vallée, France
Tel.: +33 1 60 95 72 92
Fax: +33 1 60 95 77 55
{first.last}@univ-paris-est.fr

ness. Furthermore, an interesting approach was proposed by Alvarez *et al.* [10]. However, this method is based on reaction-diffusion PDEs and it addresses the quantization of grey-scale images, while our approach is more general and applicable to multicomponent images.

In this paper, we propose a quantization method that enforces some global spatial smoothness. This is achieved by introducing an adjustable regularization term in the minimization criterion, in addition to a quantization error measure. Similarly to the LBG algorithm, the optimal design of the quantizer is performed iteratively by alternating the minimization of a label field $i_{\mathcal{P}}$ and of a codebook $\boldsymbol{r}$. The latter minimization reduces to a convex optimization problem whereas the former is carried out by efficient combinatorial optimization techniques.

Section 2 describes the background of the work and introduces the notation used throughout the paper. The considered regularization approach is formulated in section 3. Section 4 describes the proposed quantizer design algorithm. Section 5 provides more details on the combinatorial optimization step. Finally, some simulation results are provided in section 6 both for grey scale and color images to show the effectiveness of the proposed quantization method before a conclusion is drawn in section 7.

2 Background

We consider the vector quantization of a multichannel image $f = \big(f(n,m)\big)_{(n,m)\in\mathbb{V}}$ where $\mathbb{V} = \{1,\ldots,N\} \times \{1,\ldots,M\}$ is the image support and, for every $(n,m) \in \mathbb{V}$,

$$f(n,m) = \big(f_1(n,m),\ldots,f_D(n,m)\big)^\top \in \mathbb{R}^D. \tag{1}$$

A similar notation will be used for the D-channel fields defined throughout the paper. Example of such multivariate data are complex-valued images ($D = 2$), color images ($D = 3$), multispectral images (D usually less than 10), hyperspectral images (D usually more than 10),... In the following, the vector quantizer will operate on each D-dimensional vector of pixel values. The case when $D = 1$ corresponds to a scalar quantization of a monochannel image.

In order to define such a vector quantizer, we introduce the following variables: Q is a positive integer, $\mathcal{P} = (\mathbb{D}_k)_{1\leq k\leq Q}$ is a partition of $\mathbb{V}$ and $\boldsymbol{r} = (r_1,\ldots,r_Q)$ is a matrix belonging to a nonempty closed convex subset C of $\mathbb{R}^{D\times Q}$. The role of this constraint set will be made more explicit in the next sections. The partition $\mathcal{P}$ can be characterized by the label image $\big(i_{\mathcal{P}}(n,m)\big)_{(n,m)\in\mathbb{V}} \in \{1,\ldots,Q\}^{N\times M}$, defined as: for every $(n,m) \in \mathbb{V}$ and $k \in \{1,\ldots,Q\}$,

$$i_{\mathcal{P}}(n,m) = k \quad \Leftrightarrow \quad (n,m) \in \mathbb{D}_k. \tag{2}$$

A vector quantized image over Q codevectors $r_1,\ldots,r_Q$ and associated with the partition $\mathcal{P}$ is then given by

$$q_{i_{\mathcal{P}},\boldsymbol{r}} = (r_{i_{\mathcal{P}}(n,m)})_{(n,m)\in\mathbb{V}} \in \{r_1,\ldots,r_Q\}^{N\times M}. \tag{3}$$

A numerical example is given below to better explain the relation between variables Q,$\boldsymbol{r}$, $i_{\mathcal{P}}$ and $q_{i_{\mathcal{P}},\boldsymbol{r}}$, which play a prominent role in the rest of the paper. For instance, if a quantization over 2 bits of a 3×3 monochannel image is performed, we have $N = M = 3$, $D = 1$, $Q = 4$, and we may have $\boldsymbol{r} = (1,4,9,10)$ and $i_{\mathcal{P}} = \left[\begin{smallmatrix}1&2&3\\1&4&1\\3&2&3\end{smallmatrix}\right]$, then $q_{i_{\mathcal{P}},\boldsymbol{r}} = \left[\begin{smallmatrix}1&4&9\\1&10&1\\9&4&9\end{smallmatrix}\right]$. Note that the $i_{\mathcal{P}}$ matrix values belong to the set $\{1,2,3,4\}$ which correspond to the set of labels and $q_{i_{\mathcal{P}},\boldsymbol{r}}$ matrix values belong to $\boldsymbol{r}$.

An "optimally" quantized image $q_{i_{\overline{\mathcal{P}}},\overline{\boldsymbol{r}}}$ of f is usually obtained by looking for $(i_{\overline{\mathcal{P}}},\overline{\boldsymbol{r}}) \in \{1,\ldots,Q\}^{N\times M} \times C$ solution to the following problem:

$$\underset{(i_{\mathcal{P}},\boldsymbol{r})\in\{1,\ldots,Q\}^{N\times M}\times C}{\text{minimize}} \quad \Phi(q_{i_{\mathcal{P}},\boldsymbol{r}}, f) \tag{4}$$

where $\Phi\colon (\mathbb{R}^D)^{N\times M} \times (\mathbb{R}^D)^{N\times M} \to\,]-\infty,+\infty]$ is some measure of the quantization error.

Standard choices for Φ are separable functions of the form

$$\big(\forall g = \big(g(n,m)\big)_{(n,m)\in\mathbb{V}} \in (\mathbb{R}^D)^{N\times M}\big)$$
$$\Phi(g,f) = \sum_{n=1}^{N}\sum_{m=1}^{M} \varphi_{n,m}\big(g(n,m), f(n,m)\big) \tag{5}$$

where, for every $(n,m) \in \{1,\ldots,N\} \times \{1,\ldots,M\}$, $\varphi_{n,m}\colon \mathbb{R}^D \times \mathbb{R}^D \to\,]-\infty,+\infty]$. For example, one can use:

- the matrix weighted quadratic norm

$$\varphi_{n,m}\big(g(n,m), f(n,m)\big) = \|g(n,m) - f(n,m)\|^2_{\Gamma_{n,m}} \tag{6}$$

  where $\Gamma_{n,m} \in \mathbb{R}^{D\times D}$ is a symmetric definite positive matrix and we have used the notation

$$(\forall a \in \mathbb{R}^D) \quad \|a\|_{\Gamma_{n,m}} = (a^\top \Gamma_{n,m} a)^{1/2}; \tag{7}$$

- the weighted ℓ_p norm measure ($p \in [1,+\infty[$)

$$\varphi_{n,m}\big(g(n,m), f(n,m)\big) = \sum_{d=1}^{D} \omega_d(n,m)|g_d(n,m) - f_d(n,m)|^p \tag{8}$$

  where $\omega_d(n,m) \in [0,+\infty[$. As a special case, a mean absolute error criterion is found when $p = 1$.

– the generalized Kullback-Leibler divergence

$$\varphi_{n,m}\big(g(n,m),f(n,m)\big)=\sum_{d=1}^{D}\kappa\big(g_d(n,m),f_d(n,m)\big) \tag{9}$$

where

$$\big(\forall(u,v)\in\mathbb{R}^2\big)\quad \kappa(u,v)=\begin{cases}-v\ln(u/v)+u-v & \\ \qquad\qquad \text{if } (u,v)\in\,]0,+\infty[^2 & \\ u \qquad\qquad \text{if } u\in[0,+\infty[\text{ and } v=0 & \\ +\infty \qquad\quad \text{otherwise.}\end{cases} \tag{10}$$

Maximum error measures may also be useful, which are expressed as

$$\big(\forall g=\big(g(n,m)\big)_{(n,m)\in\mathbb{V}}\in(\mathbb{R}^D)^{N\times M}\big)\quad \Phi(g,f)=\max_{\substack{1\le n\le N\\1\le m\le M}}\varphi_{n,m}\big(g(n,m),f(n,m)\big) \tag{11}$$

where, for every $(n,m)\in\mathbb{V}$, $\varphi_{n,m}\colon\mathbb{R}^D\times\mathbb{R}^D\to\,]-\infty,+\infty]$. For example, we can use the sup norm:

$$\varphi_{n,m}\big(g(n,m),f(n,m)\big)=\max_{1\le d\le D}|g_d(n,m)-f_d(n,m)|. \tag{12}$$

In this context, a numerical solution to problem (4) when $C=\mathbb{R}^{D\times Q}$ is provided by the LBG algorithm, the general form of which is recalled below.

Algorithm 1 (LGB Algorithm)
Fix $Q\in\mathbb{N}^*$ and $\boldsymbol{r}^{(0)}\in\mathbb{R}^{D\times Q}$.
For $\ell=0,1,\ldots$
$\quad i_{\mathcal{P}}^{(\ell)}\in\operatorname{Argmin}_{i_{\mathcal{P}}\in\{1,\ldots,Q\}^{N\times M}}\Phi(q_{i_{\mathcal{P}},\boldsymbol{r}^{(\ell)}},f)$
$\quad \boldsymbol{r}^{(\ell+1)}\in\operatorname{Argmin}_{\boldsymbol{r}\in\mathbb{R}^{D\times Q}}\Phi(q_{i_{\mathcal{P}}^{(\ell)},\boldsymbol{r}},f)$

For separable and maximum error measures (see (5) and (11)), the optimization of the label field at iteration ℓ then amounts to applying a nearest neighbour rule, i.e. finding $i_{\mathcal{P}}^{(\ell)}$ such that, for every $(n,m)\in\mathbb{V}$ and, for every $k\in\{1,\ldots,Q\}$, $i_{\mathcal{P}}^{(\ell)}(n,m)=k$ only if

$$(\forall k'\in\{1,\ldots,Q\})\quad \varphi_{n,m}\big(r_k,f(n,m)\big)\le\varphi_{n,m}\big(r_{k'},f(n,m)\big). \tag{13}$$

Note that, in general, $i_{\mathcal{P}}^{(\ell)}(n,m)$ is not uniquely defined since there may exist $k'\in\{1,\ldots,Q\}\setminus\{k\}$ such that $\varphi_{n,m}\big(r_k,f(n,m)\big)=\varphi_{n,m}\big(r_{k'},f(n,m)\big)$.

On the other hand, updating of the codebook at iteration ℓ is performed by computing the centroid of each region $\mathbb{D}_k^{(\ell)}$, $k\in\{1,\ldots,Q\}$. For the matrix weighted quadratic norm ((5) and (6)), we thus obtain the classical center of mass of $\mathbb{D}_k^{(\ell)}$:

$$r_k^{(\ell+1)}=\Big(\sum_{(n,m)\in\mathbb{D}_k^{(\ell)}}\Gamma_{n,m}\Big)^{-1}\Big(\sum_{(n,m)\in\mathbb{D}_k^{(\ell)}}\Gamma_{n,m}f(n,m)\Big). \tag{14}$$

For the mean absolute value criterion ((5) and (8) with $p=1$ and equal weights), $r_k^{(\ell+1)}$ is the vector median of the pixel values located in $\mathbb{D}_k^{(\ell)}$:

$$r_k^{(\ell+1)}=\big(\operatorname{median}\big\{f_d(n,m)\;\big|\;(n,m)\in\mathbb{D}_k^{(\ell)}\big\}\big)_{1\le d\le D}. \tag{15}$$

For the generalized Kullback-Leibler divergence ((5), (9) and (10)), we get

$$r_k^{(\ell+1)}=\frac{1}{\operatorname{card}\mathbb{D}_k^{(\ell)}}\sum_{(n,m)\in\mathbb{D}_k^{(\ell)}}f(n,m) \tag{16}$$

provided that $f\in([0,+\infty[^D)^{N\times M}$. For the sup norm ((11) and (12)), we have

$$r_k^{(\ell+1)}=\Big(\frac{\beta_{d,k}+\gamma_{d,k}}{2}\Big)_{1\le d\le D} \tag{17}$$

where $\beta_{d,k}=\min\big\{f_d(n,m)\;\big|\;(n,m)\in\mathbb{D}_k^{(\ell)}\big\}$ and $\gamma_{d,k}=\max\big\{f_d(n,m)\;\big|\;(n,m)\in\mathbb{D}_k^{(\ell)}\big\}$.

When a closed form expression of $r_k^{(\ell+1)}$ is not available, one may resort to numerical optimization algorithms [11] to compute centroids.

It can also be noticed that an alternative to the LBG algorithm is the dynamic programming approach proposed in [12] (see also [13,14] for more recent extensions) which features better global convergence properties. Generally, if LBG used, the final solution is suboptimal.

3 Considered design criterion

One drawback of the approach described in the previous section is that it does not guarantee any spatial homogeneity of the resulting quantized image. To alleviate this shortcoming, we propose to solve the following problem:

$$\underset{(i_{\mathcal{P}},\boldsymbol{r})\in\{1,\ldots,Q\}^{N\times M}\times C}{\text{minimize}}\quad \Phi(q_{i_{\mathcal{P}},\boldsymbol{r}},f)+\rho(i_{\mathcal{P}}) \tag{18}$$

where $\rho\colon\{1,\ldots,Q\}^{N\times M}\to\,]-\infty,+\infty]$ is some penalty function which is used to promote the spatial regularity of the label image. Note that an alternative approach for ensuring the smoothness of the quantized image would be to solve a problem of the form

$$\underset{(i_{\mathcal{P}},\boldsymbol{r})\in\{1,\ldots,Q\}^{N\times M}\times C}{\text{minimize}}\quad \Phi(q_{i_{\mathcal{P}},\boldsymbol{r}},f)+\widetilde{\rho}(q_{i_{\mathcal{P}},\boldsymbol{r}}) \tag{19}$$

where the regularization term $\widetilde{\rho}$ is now a function from $(\mathbb{R}^D)^{N\times M}$ to $]-\infty,+\infty]$. The latter problem appears however more difficult to solve than (18) since the regularization term in (19) is a multivariate function depending both on $i_{\mathcal{P}}$ and $\boldsymbol{r}$.

The existence of a solution to problem (18) is secured by the following result:

Proposition 1 *Assume that $\Phi(\cdot, f)$ is a lower-semi-continuous function and that one of the following conditions holds:*

(i) *$\Phi(\cdot, f)$ is coercive;*[1]
(ii) *C is bounded.*

Then, problem (18) *has a solution.*

Proof Let $i_{\mathcal{P}}$ be any given label field in $\{1,\ldots,Q\}^{N\times M}$. According to (3), $\boldsymbol{r}\mapsto q_{i_{\mathcal{P}},\boldsymbol{r}}$ is a linear operator, and consequently $\boldsymbol{r}\mapsto\Phi(q_{i_{\mathcal{P}},\boldsymbol{r}},f)$ is a lower-semicontinuous function. As a direct consequence of Weierstrass theorem [15], under Assumption (i) or (ii), there exists $\overline{\boldsymbol{r}}_{\mathcal{P}}\in C$ such that

$$\Phi(q_{i_{\mathcal{P}},\overline{\boldsymbol{r}}_{\mathcal{P}}},f)=\min_{\boldsymbol{r}\in C}\Phi(q_{i_{\mathcal{P}},\boldsymbol{r}},f). \tag{20}$$

Problem (18) can thus be reexpressed as

$$\underset{i_{\mathcal{P}}\in\{1,\ldots,Q\}^{N\times M}}{\text{minimize}}\quad \Phi(q_{i_{\mathcal{P}},\overline{\boldsymbol{r}}_{\mathcal{P}}},f)+\rho(i_{\mathcal{P}}). \tag{21}$$

The latter minimization can be performed by a search among a finite number of candidate values, so leading to an optimal label field $i_{\overline{\mathcal{P}}}$. Hence, $(i_{\overline{\mathcal{P}}},q_{i_{\overline{\mathcal{P}}},\overline{\boldsymbol{r}}_{\overline{\mathcal{P}}}})$ is a solution to problem (18).

Typical choices for ρ in (18) that can be made are the following:

– isotropic variation functions

$$\rho(i_{\mathcal{P}})=\mu\sum_{n=1}^{N-1}\sum_{m=1}^{M-1}\psi(\|\nabla i_{\mathcal{P}}(n,m)\|),\quad \mu\geq 0 \tag{22}$$

where $\nabla i_{\mathcal{P}}(n,m)=\big(i_{\mathcal{P}}(n+1,m)-i_{\mathcal{P}}(n,m),\ i_{\mathcal{P}}(n,m+1)-i_{\mathcal{P}}(n,m)\big)$ is the discrete gradient of $i_{\mathcal{P}}$ at location (n,m).

– anisotropic variation functions

$$\rho(i_{\mathcal{P}})=\mu\Big(\sum_{n=1}^{N-1}\sum_{m=1}^{M}\psi(|i_{\mathcal{P}}(n+1,m)-i_{\mathcal{P}}(n,m)|) +\sum_{n=1}^{N}\sum_{m=1}^{M-1}\psi(|i_{\mathcal{P}}(n,m+1)-i_{\mathcal{P}}(n,m)|)\Big),\quad \mu\geq 0. \tag{23}$$

[1] This means that $\lim_{\|g\|\to+\infty}\Phi(g,f)=+\infty$.

In the above two examples, ψ is a function from $[0,+\infty[$ to $]-\infty,+\infty]$. When ψ is the identity function, the classical isotropic or anisotropic total variations are obtained. A more flexible form is given by the truncated linear function [16] defined as

$$(\forall x\in[0,+\infty[)\qquad \psi(x)=\begin{cases}x & \text{if } x<\zeta\\ \zeta & \text{otherwise}\end{cases} \tag{24}$$

where $\zeta>0$ is the limiting constant. If $\psi=(\cdot)^2$, then a Tikhonov-like regularization is performed. Another interesting choice of ψ is the binary cost function (also named ℓ_0 criterion).

$$(\forall x\in[0,+\infty[)\qquad \psi(x)=\begin{cases}0 & \text{if } x=0\\ 1 & \text{otherwise.}\end{cases} \tag{25}$$

When ψ is a (strictly) increasing function, higher local differences of the label values entail a stronger penalization. For this behaviour to be consistent with the quantized image values, some ordering relation should typically exist between the codevectors. Hence, if $D=1$, a natural choice is to constrain the vector $\boldsymbol{r}$ to belong to the closed convex cone:

$$C=\big\{(s_1,\ldots,s_Q)\in\mathbb{R}^Q\ \big|\ s_1\leq\cdots\leq s_Q\big\}. \tag{26}$$

When $D>1$, the definition of C becomes more debatable since there exists no total order on $\mathbb{R}^D$. A possibility is to impose an artificial total order. In mathematical morphology, authors have proposed various lexicographic orderings [17,18] or bit-mixing [19] along space-filling (Peano-like) curves.

A possible choice for C is the closed convex cone:

$$C=\big\{(s_1,\ldots,s_Q)\in\mathbb{R}^{D\times Q}\ \big|\ \theta(s_1)\leq\cdots\leq\theta(s_Q)\big\} \tag{27}$$

where

$$(\forall u\in\mathbb{R}^D)\qquad \theta(u)=\eta^\top u \tag{28}$$

and $\eta\in\mathbb{R}^D$. For example, for color images, by an appropriate choice of $\eta\in\mathbb{R}^3$ (possibly depending on the considered color system [20]), the function θ may serve to extract the luminance component of the codevectors.

More generally, the parameter vector $\eta\in\mathbb{R}^D$ may be obtained through a principal component analysis [21] of the original multichannel data. Note that, when the binary function in (25) is employed, the magnitudes of the local differences of the label fields have no influence as soon as they are nonzero. This means that ordering the codevectors does not appear useful in this case, and that one can set $C=\mathbb{R}^{D\times Q}$.

In addition to these considerations, when the regularization constant μ in (22) or (23) takes large values, solving (18) under the constraints modeled by (26) may

lead to very close or even equal values of codevectors. As a consequence, the readibility of the quantized image may be affected. In some applications, it may therefore be beneficial to redefine the constraint C in order to prevent this effect. When $D = 1$, the closed convex set C can thus be given by

$$C = \{(s_1, \ldots, s_Q) \in \mathbb{R}^Q \mid (\forall k \in \{1, \ldots, Q-1\})\ s_{k+1} - s_k \geq \delta\} \quad (29)$$

where $\delta \geq 0$. Similarly, when $D > 1$, we propose to set

$$C = \{(s_1, \ldots, s_Q) \in \mathbb{R}^{D\times Q} \mid (\forall k \in \{1, \ldots, Q-1\})\ \theta(s_{k+1} - s_k) \geq \delta\} \quad (30)$$

where $\delta \geq 0$ and θ is the function given by (28). Penalization of quantization values for being too close to each other was previously introduced in the energy model proposed by Alvarez *et al.* [10].

4 Proposed optimization method

Even if $\Phi(\cdot, f)$ and ρ are convex functions, problem (18) is a nonconvex optimization problem due to the fact that $i_{\mathcal{P}}$ belongs to a (nonconvex) set of discrete values. In order to solve numerically this problem, we propose to use the following alternating optimization algorithm:

Algorithm 2 (Proposed algorithm)
Fix $Q \in \mathbb{N}^*$ and $\boldsymbol{r}^{(0)} \in C$.
For $\ell = 0, 1, \ldots$
$\quad i_{\mathcal{P}}^{(\ell)} \in \operatorname{Argmin}_{i_{\mathcal{P}} \in \{1,\ldots,Q\}^{N\times M}} \Phi(q_{i_{\mathcal{P}}, \boldsymbol{r}^{(\ell)}}, f) + \rho(i_{\mathcal{P}})$
$\quad \boldsymbol{r}^{(\ell+1)} \in \operatorname{Argmin}_{\boldsymbol{r}\in C} \Phi(q_{i_{\mathcal{P}}^{(\ell)}, \boldsymbol{r}}, f)$

It is worth noticing that this algorithm constitutes an extension of the LBG algorithm (see Algorithm 1) which would correspond to the case when ρ is the null function and $C = \mathbb{R}^{D\times Q}$. Similarly to the LBG algorithm, under the assumptions of proposition 1, Algorithm 2 generates a sequence $(i_{\mathcal{P}}^{(\ell)}, \boldsymbol{r}^{(\ell+1)})_{\ell\in\mathbb{N}}$ such that $\big(\Phi(q_{i_{\mathcal{P}}^{(\ell)}, \boldsymbol{r}^{(\ell+1)}}, f) + \rho(i_{\mathcal{P}}^{(\ell)})\big)_{\ell\in\mathbb{N}}$ is a convergent decaying sequence. At each iteration ℓ, the determination of $i_{\mathcal{P}}^{(\ell)}$ given $\boldsymbol{r}^{(\ell)}$ is a combinatorial optimization problem for which there exist efficient solutions for particular choices of Φ and ρ, as explained in the next section.

In turn, if $\Phi(\cdot, f)$ is a convex function, the determination of $\boldsymbol{r}^{(\ell+1)}$ given $i_{\mathcal{P}}^{(\ell)}$ is a constrained convex optimization problem the solution of which can be determined numerically. For any given $i_{\mathcal{P}} \in \{1, \ldots, Q\}^{N\times M}$, let $L_{i_{\mathcal{P}}}$ be the linear operator defined as

$$L_{i_{\mathcal{P}}} \colon \mathbb{R}^{D\times Q} \to (\mathbb{R}^D)^{N\times M}, \quad \boldsymbol{r} \mapsto q_{i_{\mathcal{P}}, \boldsymbol{r}} \quad (31)$$

the adjoint of which is

$$L_{i_{\mathcal{P}}}^* \colon (\mathbb{R}^D)^{N\times M} \to \mathbb{R}^{D\times Q}, \quad g \mapsto \Big(\sum_{(n,m)\in\mathbb{D}_1} g(n,m), \ldots, \sum_{(n,m)\in\mathbb{D}_Q} g(n,m) \Big) \quad (32)$$

(with the convention $\sum_{(n,m)\in\varnothing} \cdot = 0$). Then,

$$L_{i_{\mathcal{P}}}^* L_{i_{\mathcal{P}}} \colon \mathbb{R}^{D\times Q} \to \mathbb{R}^{D\times Q}, \quad \boldsymbol{r} \mapsto \boldsymbol{r}\,\mathrm{Diag}(\mathrm{card}\,\mathbb{D}_1, \ldots, \mathrm{card}\,\mathbb{D}_Q). \quad (33)$$

In addition, let Θ be the linear operator defined as

$$\Theta \colon \mathbb{R}^{D\times Q} \to \mathbb{R}^{Q-1} \quad (34)$$

$$(s_1, \ldots, s_Q) \mapsto \big(\theta(s_2 - s_1), \ldots, \theta(s_Q - s_{Q-1})\big) \quad (35)$$

where θ is given by (28) (with $\eta = 1$ when $D = 1$). The set C defined in (29) or (30) is thus equal to $\Theta^{-1}([\delta, +\infty[^{Q-1})$. Hence, the problem of minimization of $\boldsymbol{r} \mapsto \Phi(q_{i_{\mathcal{P}}, \boldsymbol{r}}, f)$ over C can be reexpressed as

$$\underset{\boldsymbol{r}\in\mathbb{R}^{D\times Q}}{\text{minimize}} \quad \Phi(L_{i_{\mathcal{P}}}\boldsymbol{r}, f) + \iota_{[\delta,+\infty[^{Q-1}}(\Theta\boldsymbol{r}) \quad (36)$$

where ι_S denotes the indicator function of a set S, which is zero on S and equal to $+\infty$ on its complement. If we assume that $\Phi(\cdot, f)$ belongs to $\Gamma_0\big((\mathbb{R}^D)^{N\times M}\big)$, the class of lower-semicontinuous proper convex functions from $(\mathbb{R}^D)^{N\times M}$ to $]-\infty, +\infty]$, (36) can be solved through existing convex optimization approaches [11,22,23]. One possible solution is to employ the method proposed in [24] (hereafter called PPXA+) which constitutes an extension of the parallel proximal algorithm (PPXA) developed in [25] and of the simultaneous direction of multipliers method proposed in [26] (see also [27,28,29]).

Algorithm 3 (PPXA+ for solving (36))

Initialization
$(\omega_1, \omega_2, \omega_3) \in]0, +\infty[^3$
$t^{(1,0)} \in (\mathbb{R}^D)^{N\times M}, t^{(2,0)} \in \mathbb{R}^{Q-1}, \boldsymbol{s}^{(0)} \in \mathbb{R}^{D\times Q}$
$R = (\omega_1 L^*_{i_\mathcal{P}} L_{i_\mathcal{P}} + \omega_2 \Theta^*\Theta + \omega_3 I)^{-1}$
$\boldsymbol{r}^{(0)} = R\,(\omega_1 L^*_{i_\mathcal{P}} t^{(1,0)} + \omega_2 \Theta^* t^{(2,0)} + \omega_3 \boldsymbol{s}^{(0)})$

For $\ell = 0, 1, \ldots$
$p^{(1,\ell)} = \operatorname{prox}_{\frac{1}{\omega_1}\Phi(\cdot,f)}\big(t^{(1,\ell)}\big)$
$p^{(2,\ell)} = P_{[\delta,+\infty[^{Q-1}}\big(t^{(2,\ell)}\big)$
$\boldsymbol{c}^{(\ell)} = R\,(\omega_1 L^*_{i_\mathcal{P}} p^{(1,\ell)} + \omega_2 \Theta^* p^{(2,\ell)} + \omega_3 \boldsymbol{s}^{(\ell)})$
$\lambda_\ell \in]0, 2[$
$t^{(1,\ell+1)} = t^{(1,\ell)} + \lambda_\ell\big(L_{i_\mathcal{P}}(2\boldsymbol{c}^{(\ell)} - \boldsymbol{r}^{(\ell)}) - p^{(1,\ell)}\big)$
$t^{(2,\ell+1)} = t^{(2,\ell)} + \lambda_\ell\big(\Theta(2\boldsymbol{c}^{(\ell)} - \boldsymbol{r}^{(\ell)}) - p^{(2,\ell)}\big)$
$\boldsymbol{s}^{(\ell+1)} = \boldsymbol{s}^{(\ell)} + \lambda_\ell(2\boldsymbol{c}^{(\ell)} - \boldsymbol{r}^{(\ell)} - \boldsymbol{s}^{(\ell)})$
$\boldsymbol{r}^{(\ell+1)} = \boldsymbol{r}^{(\ell)} + \lambda_\ell(\boldsymbol{c}^{(\ell)} - \boldsymbol{r}^{(\ell)})$.

In the above algorithm, $\operatorname{prox}_{\frac{1}{\omega_1}\Phi(\cdot,f)}$ is the proximity operator of $\omega_1^{-1}\Phi(\cdot, f)$ [30] and $P_{[\delta,+\infty[^{Q-1}}$ is the projector onto $[\delta, +\infty[^{Q-1}$. Expressions of proximity operators for usual convex functions are listed in [31]. The convergence of the PPXA+ algorithm is guaranteed under weak assumptions.

Proposition 2 *Assume that*

(i) *there exists* $\underline{\lambda} \in]0, 2[$ *such that* $(\forall \ell \in \mathbb{N})\ \underline{\lambda} \leq \lambda_{\ell+1} \leq \lambda_\ell$.
(ii) *There exists* $\overline{\boldsymbol{r}} \in \mathbb{R}^{D\times Q}$ *such that*

$$L_{i_\mathcal{P}}\overline{\boldsymbol{r}} \in \operatorname{ri}\operatorname{dom}\Phi(\cdot, f) \quad \textit{and} \quad \Theta\overline{\boldsymbol{r}} \in]\delta, +\infty[^{Q-1} \tag{37}$$

where $\operatorname{dom}\Phi(\cdot, f)$ *is the domain of* $\Phi(\cdot, f)$ *and* $\operatorname{ri}\operatorname{dom}\Phi(\cdot, f)$ *is its relative interior.*

Then, the sequence $(\boldsymbol{r}^{(\ell)})_{\ell\in\mathbb{N}}$ *generated by Algorithm 3 converges to a solution to problem* (36).

Proof See [24].

5 Combinatorial partitioning

We now consider two combinatorial optimization methods for finding

$$i_{\widehat{\mathcal{P}}} \in \operatorname*{Argmin}_{i_\mathcal{P}\in\{1,\ldots,Q\}^{N\times M}} \Phi(q_{i_\mathcal{P},\boldsymbol{r}}, f) + \rho(i_\mathcal{P}) \tag{38}$$

for a given value of $\boldsymbol{r} \in C$. Here we seek to use standard methods in combinatorial optimization which have proved to be useful in applications to image processing. In this context, a common form for regularization problems is the following:

$$\operatorname*{minimize}_{i_\mathcal{P}\in\{1,\ldots,Q\}^{N\times M}} \widetilde{\Phi}(i_\mathcal{P}, f) + \rho(i_\mathcal{P}), \tag{39}$$

where $\widetilde{\Phi}\colon \{1,\ldots,Q\}^{N\times M} \times (\mathbb{R}^D)^{N\times M} \to]-\infty, +\infty]$ is a data fidelity function, ρ a regularization function, f the initial image and $i_\mathcal{P}$ the target discrete one. To formulate our problem in this framework, we need to introduce the auxiliary function

$$\begin{aligned}\chi_{\boldsymbol{r}}\colon \{1,\ldots,Q\}^{N\times M} &\mapsto \{r_1,\ldots,r_Q\}^{N\times M}\\ i_\mathcal{P} &\mapsto q_{i_\mathcal{P},\boldsymbol{r}}.\end{aligned}$$

Then, our problem becomes

$$\operatorname*{minimize}_{i_\mathcal{P}\in\{1,\ldots,Q\}^{N\times M}} \Phi(\chi_{\boldsymbol{r}}(i_\mathcal{P}), f) + \rho(i_\mathcal{P}). \tag{40}$$

Note that $\chi_{\boldsymbol{r}}$ is monotonic but nonlinear. Note further that the set $\{r_1,\ldots,r_Q\}$ changes at each iteration of the complete algorithm. However, during the regularization step, this set is fixed.

In this section, we use graph-cut based algorithms, which have proved to be effective in the context of smoothing, denoising and segmentation [32].

5.1 Method I - convex regularization term

Here we describe a way to formulate the problem as a globally optimal graph cut, inspired by the approach of Ishikawa *et al.* [33]. In this approach, we build a discrete graph that will allow us to represent the quantized and regularized version of our original image. Let us define the oriented, edge-weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ as follows:

(i) $\mathcal{V} = \mathbb{V} \times \{1,\ldots,Q\} \cup \{s, t\}$ the set of vertices quantized over Q levels, where $\mathbb{V}$ is the image support as defined in section 2. We add two special vertices, the *source* s and the *sink* t.
(ii) $\mathcal{E} = \mathcal{E}_D \cup \mathcal{E}_C \cup \mathcal{E}_P$, the set of edges. In the following we denote an oriented edge by $[a, b]$, with a and b the vertices it joins in the direction from a to b. We have :

(a) $\mathcal{E}_D = \bigcup_{v\in\mathbb{V}} \mathcal{E}_D^v$ the upward columns of the graph. For all $v \in \mathbb{V}$, let $h_{v,k}$ denote the node in column v and row k. A single column associated with pixel v is defined as
$\mathcal{E}_D^v = \{[s, h_{v,1}]\} \cup \{[h_{v,k}, h_{v,k+1}] \mid k \in \{1, ..., Q-1\}\} \cup \{h_{v,Q}, t\}$,
(b) $\mathcal{E}_C = \bigcup_{v\in\mathbb{V}} \mathcal{E}_C^v$ the downward columns of the graph, with
$\mathcal{E}_C^v = \{[h_{v,k}, h_{v,k+1}] \mid k \in \{1, ..., Q-1\}\}$,
(c) and the penalty edges of the graph are thus defined as
$\mathcal{E}_P = \{[h_{v,k}, h_{w,k}] \mid \{v, w\}$ neighbours in$\mathbb{V}$, $k \in \{1, ..., Q\}\}$.

The above graph is depicted in Fig. 1. In this figure, for simplicity we assume each pixel has only two neighbours, which allows us to represent the graph in a 2D planar layout. For actual 2D images, there exist many more penalty edges between all neighbours in $\mathbb{V}$. The graph layout is then non-planar, but remains similar. For 2D images, it is best to see the arrangement of v vertices as in the original images, with the column of penalty edges in an extra dimension.

If Φ is the separable function defined in (5) and ρ is the anisotropic TV in (23) where ψ is the identity function, we define the capacities (or weights) c of edges $[a, b] \in \mathcal{E}$ as follows:

(i) Links to the source have infinite capacity: $\forall v \in \mathbb{V}, c([s, h_{v,1}]) = +\infty$.
(ii) Data fidelity terms for any pixel $v \in \mathbb{V}$ is $\forall k \in \{1, \dots, Q-1\}, c([h_{v,k}, h_{v,k+1}]) = \varphi_v(r_k, f(v))$, $c([h_{v,Q}, t]) = \varphi_v(r_Q, f(v))$.
(iii) The capacity of downward columns is infinite to constrain a single cut per column: $\forall v \in \mathbb{V}, \forall k \in \{1 \dots Q-1\}, c([h_{v,k+1}, h_{v,k}]) = +\infty$.
(iv) The regularization term along the penalty edges of the graph is: for every $\{v, w\}$ neighbours in $\mathbb{V}, \forall k \in \{1, \dots, Q\}$, $c([h_{v,k}, h_{w,k}]) = \mu$

The above graph $\mathcal{G}$ has the same topology as the one proposed by Ishikawa and it can be extended to any convex function ψ [33]. The capacities of $\mathcal{E}$ are adjusted in such a way that a cut of $\mathcal{G}$ corresponds to the solution of (40), granted by the following result:

Proposition 3 *If ρ is the anisotropic TV in* (23) *where ψ is the identity function, then the min cut of $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is the globally optimal solution to* (40).

Proof This result is derived from the construction of the graph. First note that we build here a binary flow network with one source and one sink. Following Ishikawa, relying on the celebrated discrete maxflow/mincut theorem of Ford and Fulkerson [34], any binary cut that separates s and t along a series of edges, that can be interpreted as a solution $i_{\mathcal{P}}$. Indeed, the infinite capacity of the downward edges ensure a single cut edge in each column of the graph, and the infinite capacity of the upward $[s, h_{v,1}]$ edges for all v ensures that, in all columns, this cut will be located above one of the nodes corresponding to a level $k \in \{1, \dots, Q\}$. We can therefore associate the cut in column v with the value of the level immediately below the cut, and associate this with $i_{\mathcal{P}}(v)$. Recalling that all labels below the cut will have the same label as s, and all that above the cut the same label as t, the value of $i_{\mathcal{P}}$ at pixel v is the highest level l in column v of the graph that is labelled like the source s. Here, by convention, the source is labelled with 1 and the sink with 0. We can then write $i_{\mathcal{P}}(v) = \max\{k, h_{v,k} = 1\}$.

Now, the computation of the maxflow/mincut on this graph minimizes the energy of the cut, interpreted as the sum of two terms:

(i) since the downward constraint edges ensure a single cut edge along each column of the graph, this corresponds to contribution of the data fidelity term $\varphi_v(r_Q, f(v))$ to the total energy.
(ii) Similarly, we note that each penalty edges in $\mathcal{E}_P$ with capacity μ can be cut at most once. Let u and v be two neighbouring pixels in the graph. The cut along penalty edges between $i_{\mathcal{P}}(u)$ and $i_{\mathcal{P}}(v)$ crosses exactly as many penalty edges as there are quantization level differences between u and v. We note that this correspond to a contribution of $\mu|i_{\mathcal{P}}(u) - i_{\mathcal{P}}(v)|$ to the total energy.

Hence, the computation of the maxflow/mincut on this graph solves (40) exactly, in the case of (23), when ψ is the identity.

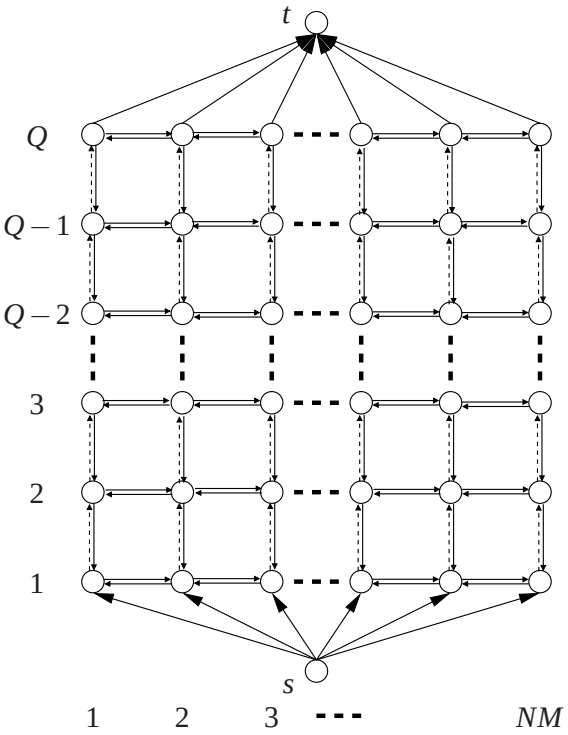


Fig. 1 Construction of the Ishikawa-like optimization graph. Arrows represent the edges $\mathcal{E}$ and circles the nodes in $\mathcal{V}$. Horizontal edges are in $\mathcal{E}_P$, the dotted upward vertical edges are in $\mathcal{E}_D$ and the plain downward vertical edges are in $\mathcal{E}_C$. Vertices s and t are respectively the source and the sink. All pixels in the image from 1 to NM are represented in the columns. In actual 2D images, there exist many more penalty edges $\mathcal{E}_P$ than depicted here: all those between neighbours in $\mathbb{V}$.

Remark 1

(i) It is also possible to solve this problem exactly in the case when ψ is convex and not necessarily the identity, by adding non-horizontal penalty edges [35], but we do not consider this case here, as $\psi = \mathrm{Id}$ is favorable when discontinuities exist in the original image.
(ii) In the case when the number of quantized levels Q is small (say between 1 and 32), the Ishikawa framework is very efficient.
(iii) As the dimensionality of the problem increases, so does the number of penalty edges in the graph. The cut is always an hypersurface of codimension 1.
(iv) Ishikawa recommends solving the maxflow/mincut by using a push-relabel algorithm, which makes perfect sense as the dimensionality increases, because these algorithms have an asymptotic complexity independent of the number of edges.

5.2 Method II - submodular regularization term

Since the method proposed in Section 5.1 works only for a convex function ψ, we propose to solve the general problem defined in (38) with the α-expansion algorithm [32], which has been proven to be very effective for some non-convex functions ψ such as the Potts model of (25). Though only a local minimum is then guaranteed, the resulting energy will be within a known factor of the global minimum energy [32]. Here we reintroduce the standard notation of α-expansions as we need to specify the capacities on the corresponding edges in the context of this article. Following Kolmogorov *et al.* [36], we build a directed graph for each quantization level, called α-expansion graph $\mathcal{G}_\alpha = (\mathcal{V}, \mathcal{E})$, defined as follows:

(i) $\mathcal{V} = V \cup \{\alpha, \overline{\alpha}\}$ is the set of vertices, with α and $\overline{\alpha}$ two special term nodes and $V = \{1, ..., NM\}$ is the set of image nodes ;
(ii) $\mathcal{E} = \mathcal{E}_V \cup \mathcal{E}_\mathcal{N}$ is the set of edges, defined as follows :
  (a) $\mathcal{E}_V = \bigcup_{v \in V} \{[\alpha, v], [v, \overline{\alpha}]\}$ is the set of edges between special term nodes and image nodes ;
  (b) $\mathcal{E}_\mathcal{N} = \bigcup_{\{u,v\} \text{ neighbours}}$ is the set of edges between neighbours and $\mathcal{N}$ is the set of neighbours pairs containing only ordered pairs u, v, i.e. such that $u < v$.
  (c) The capacity for all edges are given in Table 1.

Computing the max-flow/min-cost cut of $\mathcal{G}_\alpha$ separates vertices α and $\overline{\alpha}$ in such a way that the α region can only expand, hence the name of the algorithm. The value of the function associating new values to $i_\mathcal{P}$, based on cut of $\mathcal{G}_\alpha$, is called *"α-move of $i_\mathcal{P}$"* [16]. The algorithm is as follows:

Table 1 Capacities for the α-expansion graph of Fig 2.

edge	capacity [a]
$c([u, \overline{\alpha}])$	$\mathcal{R}\,(K_u) + \sum_{(u,v)\in\mathcal{N}} \mathcal{R}(A_{u,v} - C_{u,v}) + \sum_{(v,u)\in\mathcal{N}} C_{v,u}$
$c([\alpha, u])$	$\mathcal{R}(-K_u) + \sum_{(u,v)\in\mathcal{N}} \mathcal{R}(C_{u,v} - A_{u,v})$
$c([u, v])$	$\sum_{(u,v)\in\mathcal{N}} (B_{u,v} + C_{u,v} - A_{u,v})$

[a] The following notation is used:
$\mathcal{R}$ denotes the ramp function, i.e. $\mathcal{R}(x) = 0$ if $x \in (-\infty, 0)$ and $\mathcal{R}(x) = x$ if $x \in [0, +\infty)$
$K_u = \varphi_{n_u,m_u}(r_{i_\mathcal{P}(n_u,m_u)}, f(n_u, m_u)) - \varphi_{n_u,m_u}(r_\alpha, f(n_u, m_u))$
$A_{u,v} = \psi(|i_\mathcal{P}(n_u, m_u) - i_\mathcal{P}(n_v, m_v)|)$
$B_{u,v} = \psi(|i_\mathcal{P}(n_u, m_u) - \alpha|)$
$C_{u,v} = \psi(|\alpha - i_\mathcal{P}(n_v, m_v)|)$

Algorithm 4 (α-expansion algorithm)
Fix $i_\mathcal{P}^{(0)}$
For $\ell = 0, 1, \ldots$
$\quad \widetilde{\alpha}^{(\ell)} \in \mathrm{Argmin}_{\alpha\in\{1,\ldots,Q\}} \big\{\Phi(\chi_{\boldsymbol{r}}(\widehat{i}_\mathcal{P}), f) + \rho(\widehat{i}_\mathcal{P}) \mid$
$\quad \widehat{i}_\mathcal{P} = \alpha\text{-move of } i_\mathcal{P}^{(\ell)}\big\}$
$\quad i_\mathcal{P}^{(\ell+1)} = \widetilde{\alpha}^{(\ell)}\text{-move of } i_\mathcal{P}^{(\ell)}$

Proposition 4 *If* (38) *is submodular then it can be solved with the α-expansion algorithm.*

Proof It is shown in [36] that in order to employ the α-expansion algorithm, (38) has to satisfy the following conditions at iteration ℓ:

(i) (38) has a binary representation of the form:

$$\begin{aligned}\text{minimize } &\textstyle\sum_{u\in V} B_1^{(\ell)}(b_{(n_u,m_u)}) + \\ &\textstyle\sum_{\{u,v\}\text{neighbours}} B_2^{(\ell)}(b_{(n_u,m_u)}, b_{(n_v,m_v)}),\end{aligned} \tag{41}$$

where b is a binary field while $B_1^{(\ell)}$ and $B_2^{(\ell)}$ have binary arguments.
(ii) The binary representation b is graph-representable, which can be verified by testing if term $B_2^{(\ell)}$ satisfies the submodular inequality:

$$B_2^{(\ell)}(0,0) + B_2^{(\ell)}(1,1) \leq B_2^{(\ell)}(1,0) + B_2^{(\ell)}(0,1)\,. \tag{42}$$

We now propose the following binary formulation of (38) by defining :

$$B_1^{(\ell)}(b_{(n_u,m_u)}) = \varphi_{n_u,m_u}(r_{\widehat{i}_\mathcal{P}(n_u,m_u)}, f(n_u, m_u)) \tag{43}$$

and

$$\begin{aligned}B_2^{(\ell)}(b_{(n_u,m_u)}, b_{(n_v,m_v)}) = \\ \psi(|\widehat{i}_\mathcal{P}(n_u, m_u) - \widehat{i}_\mathcal{P}(n_v, m_v)|)\end{aligned} \tag{44}$$

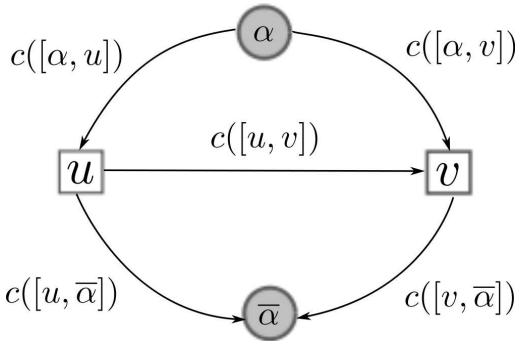


Fig. 2 Notations for the α-expansion graph, following Kolmogorov *et al.* [36]. Here we took a simplified 2-pixel neighbourhood. The cost (or capacity) between u and v is labelled as $c([u, v])$ for instance, and so on for all edges. The expressions for the capacity for all edges are given in Table 1.

where

$$\widehat{i}_{\mathcal{P}}(n_u, m_u) = \begin{cases} i_{\mathcal{P}}^{(\ell)}(n_u, m_u) & \text{if } b_{(n_u, m_u)} = 0 \\ \alpha & \text{if } b_{(n_u, m_u)} = 1. \end{cases} \quad (45)$$

More standard graph-cut formulations would only allow us to optimize (39). These formulations would be problematic because we would not be able to separate the two steps in the inner loop of Algorithm 2, and therefore no convergence property could be derived.

Assuming (38) submodular, then the terms of its binary representation defined in (44) satisfy (42). Furthermore, it is shown in [16] that for ψ defined as Potts model of (25) or the truncated linear function in (24), and when ρ is the anisotropic TV of (23), then this type of energy is indeed submodular. Consequently (38) can be solved with α-expansions.

Figure 2 provides an illustration of the notation for edge weights in a simplified situation. In order to solve problem (38) with the α-expansion algorithm, we propose to define the capacities c of edges $\mathcal{E}$ in the graph $\mathcal{G}_\alpha$ for all $\{u, v\}$ pairs of neighbours, as described in Table 1.

5.3 Other methods

Other combinatorial optimization methods might also be used. For instance, when minimizing *isotropic* TV as in (22), one might want to use Chambolle's algorithm [37]. Similarly to the Ishikawa framework, we would obtain the global optimum in this case also. Moreover *isotropic* TV minimization was recently discussed among others by Lellmann *et al.* [38], Trobin *et al.* [39] and Zach *et al.* [40]. One other possibility is the use of α - β generalized range moves algorithm, which is shown in [41] to be able to optimize a wider range of combinatorial energies than α-expansion method presented in Section 5.2. Furthermore, similar properties are held by the FastPD [42] and the PD3$_a$ [43] algorithms, both introduced by Komodakis *et al.*. Also worth mentioning is the Darbon and Sigelle method for levelable energies, introduced in [44], the Kolmogorov and Shioura primal and primal-dual algorithms and Zalesky's MSFM algorithm [45], since they are all faster than Ishikawa's approach, while still providing an exact solution for a similar class of functions. Our method can be also improved using higher order cliques, which already has been proven to provide effective filtering results [46]. It might be also possible to extend the quantization techniques proposed by Chambolle and Darbon in [47]. Also of interest would be to explore variants of anisotropic diffusion, and other combinatorial optimizers such as generalized Dirichlet solvers, which are naturally multilabel [48] and could provide much simpler algorithms.

6 Simulation examples

In this section we present four experiments in order to demonstrate the performance of our method in various scenarii. Both color and grey scale images are considered. For grey scale images, our approach is confronted with the LM method [3]. It is a fair comparison, since the same function Φ is used for both algorithms. Although sophisticated initialization procedures [49,50,51] can be employed for LM and our approach, the methods presented in the following were simply initialized with either uniform or cumulative histograms based decision levels. In the case of color images, we compared our method with: i) special case of LGB algorithm with Φ defined as ℓ_2 norm (K-means), ii) median cut [52] and iii) Wu's method [53]. Ximagic (http://www.ximagic.com) quantization package was used to generate results of K-means, Wu and median cut algorithms. Their performance is measured in terms of SNR between the original and quantized images and also by the Shannon entropy of order $(2, 2)$ (that is the entropy over image blocks of size 3×3). Note that, in all the following experiments, regularization functions are used corresponding to a 4-pixel neighbourhood (2 pixels in horizontal and 2 in vertical direction) in the employed graph cut techniques. They were implemented with the help of the publicly available library described in [54]. When running experiments using Algorithm 3, there are 4 parameters to set. We have set $\omega_1 = \omega_2 = \omega_3 = 1/3$ and λ_ℓ was fixed and equaled 1.5. The appropriate choice of parameter μ depends on the ratio between maximum values of Φ and ρ codomain,

the level of noise in original image and prior knowledge about the desired entropy of output images.

6.1 Low resolution quantization

First, we consider grey scale image quantization over $Q = 8$ levels. The combinatorial method described in Section 5.1 was used to find the global optimum of convex criterion (18) with function Φ defined as the ℓ_1 norm and function ρ defined by (23) where ψ is the identity. It was applied to 8 bit microscopy image of size 512×512 (from public domain, http://www.remf.dartmouth.edu), the fragment of which is shown in Fig. 3(a). Regularization parameter μ was hand-optimized to 10. Both methods, LM and ours, were initialized with uniform decision levels. In order to solve (36), Algorithm 3 was used. The convex set C is defined by (29), where $\delta = 12$. As expected, our results provide the best spatial smoothness among the considered methods, which is confirmed by the entropy equal to 0.58 bpp, while in case of LM it is equal to 0.84 bpp. In this example, it is shown that, in case of quantization with high level reduction, our method provides smaller entropy rate while maintaining the desired fidelity.

In the second example, we show that a similar behaviour is obtained for different choices of Φ, regularity criterion and combinatorial method. This time, the number of quantization levels is $Q = 32$, function (18) is specified by Φ defined as the squared ℓ_2 norm and ρ defined by (23) where ψ is the binary cost-function (25). It is applied to the color-image of size 256×256, which is shown in Fig. 4(a). Fig. 4(e) presents the results when μ is set to 25 and in Fig. 4(f), when it is set to 50. The difference between the two presented images (Fig. 4(e) and Fig. 4(f)) is not significant but highlights the visual influence of parameter μ. The criterion (18) was minimized by using the modified α-expansion graph described in Section 5.2, which was initialized with $\boldsymbol{r}^{(0)}$ obtained by median cut algorithm. Image pixels were mapped into the XYZ image space [55]. Similarly to the previous example, Fig. 4 shows that a better spatial smoothness is obtained with the proposed approach. This is also verified by inspecting the entropy value, which in our case is equal to 1.06 bpp for $\mu = 25$, and 1.00 bpp for $\mu = 50$, whereas in the case of Wu, K-means and Median-cut the entropies are equal to 1.18 bpp, 1.14 bpp, 1.19 bpp, respectively.

6.2 Quantization in the presence of noise

Next, we present the performance of our method in the presence of noise. Note that here function ϕ is chosen based on two noise models, i.e. ℓ_2 for Gaussian and ℓ_1 for Laplacian noise. Firstly, the problem of grey scale image quantization over 16 levels is investigated. The image of size 256×256, shown in Fig. 5(b), is corrupted by zero-mean i.i.d. Laplacian noise with standard deviation 9. Quantization is performed using Algorithm 2. The method described in Section 5.2 is used to minimize energy (18), where Φ is defined as the ℓ_1 norm and ρ is given by (23) with ψ taken as the truncated linear function (24), where the limiting constant is set to $\zeta = 3$. The associated regularization parameter μ was experimentally chosen equal to 6. Both methods, LM and ours, were initialized with cumulative histogram based decision levels. Problem (36) was solved by using Algorithm 3. The convex set C is defined by (29), where $\delta = 1$. The proposed approach shows satisfactory results when dealing with Laplacian noise: i) the visual effect of the noise is reduced (see Fig. 5(d)), ii) the SNR, which was equal to 22.7 dB for the noisy image increases to 24.6 dB, and iii) the entropy is only 0.96 bpp. In case of LM (see Fig. 5(c)), the SNR is equal 22.4 dB and the entropy is 1.41 bpp. In this example, we show that, in case of quantization in the presence of noise, our method reconstructs the original image, while performing image quantization.

Similar properties have been observed for $D > 1$. To illustrate this fact, the quantization over 16 quantization levels of a 300×300 color image is presented in Fig. 6. Zero-mean Gaussian noise with standard deviation 20 was added to the image presented in Fig. 6(a) (source: photo by Neon JA, colored by Richard Bartz / Wikimedia Commons). This image was transformed from the RGB space into a more appropriate one, using the linear transformation defined by the matrix of its PCA (Principal Component Analysis) components. Then, the total order of quantization levels along the principal component is chosen, which corresponds to $\eta^\top = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$ The convex set C is defined by (29). Since the probability of merging codevectors is negligible in three-channel color space, the associated parameter δ was set to 0. The resulting image (see Fig. 6(f)) was obtained by minimizing energy (18), which was initialized with decision levels computed by the median cut method. Function Φ was defined by (6) and ρ by (23) where ψ is identity and μ is equal to 250. The algorithm described in Section 5.1 was used for computing $i_{\mathcal{P}}^{(\ell)}$. One can observe that the noise has been highly reduced in our result (Fig. 6(f)), while the K-means method (Fig. 6(d)) preserved noise in the images. This is also verified by SNR values which is equal to 13.8 dB for our method and 10.6 dB, 10.4 dB and 9.8 dB for the K-means, Wu and median-cut, respectively. The difference is even greater in terms of entropy: our method

led to 0.79 bpp and the other ones to 1.48 bpp. Additionally, the quantization result for the original image is presented. Our result (Fig. 6(e)) was obtain with the same algorithm settings as described above except μ, which here is equal to 30 and of course the PCA parameters, which were computed from the original image. Our method performs the required quantization and provides an interesting tradeoff between precision and smoothness, which is validated with an SNR of 18.5 dB and an entropy of 0.9 bpp. In contrast, K-means (Fig. 6(c)) achieved a SNR = 20.2 dB and an entropy = 1.1 bpp.

6.3 Note about computation time

The time complexity of Algorithm 2 is equal to the product of the complexity of each iteration and the complexity of the number of iterations ℓ. The bound on ℓ is not known a priori. Our observation suggests that it is a function of the weight of smoothness term μ, number of quantization levels Q, and the spatial entropy of original image f. Moreover, there may be small differences in the number of iterations, depending on the choice of the combinatorial optimization method. For instance, the first problem described in Section 6.1, which was solved with an Ishikawa-like graph, converges in 18 iterations. In contrast, using the α-expansion algorithm (Algorithm 4), it converges in only 16. In practice the number of iterations never seems to exceed 50 for grey-scale and 200 for color images. By analyzing the inner loop of Algorithm 2, one can observe that the complexity of step 1 is greater than the one of step 2. Thus, the computation time of each iteration is strongly dominated by the cost of step 1, namely finding $i_{\mathcal{P}}^{(\ell)}$. Note that Algorithm 3 is run only if matrix $\boldsymbol{r}$ derived from a centroid rule does not belong to C, so usually its influence on the overall time complexity of Algorithm 2 is small for grey-scale images. It becomes more important for multi-channel images. Generally, the cost of combinatorial graph-cut based methods depends on $|\mathbb{V}|$ and on the number of quantization levels Q. More precisely, the relabeling algorithm finds solution for single graphs in polynomial time $O(|\mathcal{V}|^3)$, where $|\mathcal{V}|$ is equal to $Q \times |\mathbb{V}|$ for the method described in Section 5.1 and to $|\mathbb{V}|$ for the method described in Section 5.2. However, Algorithm 4 (in Section 5.2) requires solving many different graphs independently, so its computation cost increases linearly with the number of quantization levels Q. It is worth noting that some recently published extensions of the α-expansion algorithm are faster. In particular, Lipitsky *et al.* presented the LogCut and Fusion move methods that lead to nearly logarithmic growth [56], e.g. for $Q = 256$ the algorithm converges approximately 10 times faster. A similar acceleration was obtained by the FastPD algorithm introduced by Komodakis *et al.* [42] and analyzed by Kolmogorov in [57]. Likewise, recently introduced primal and primal-dual algorithms by Kolmogorov *et al.* in [58] may be an alternative for the method described in Section 5.1. They offer a significant improvement in terms of time efficiency. For instance, our first problem described in Section 6.1 solved with the method described in Section 5.1 takes 37 seconds, while using Kolmogorov's primal only algorithm, it takes only 12 seconds. As an alternative to the methods presented in Section 5, one may adopt these novel methods. Constant progress in the efficiency of graph-cut algorithms makes our approach increasingly competitive with the ones that do not feature a smoothness constraint. Nonetheless, our method may take significantly more time than the use of basic quantization methods (for details see Table 2). The tests were performed single-threaded, on a computer with a 2.5GHz Intel Xeon processor, in the RedHat Enterprise Linux 5.5 environment, using the GCC compiler version 4.1 in 64-bit mode.

	Ex1	Ex2	Ex3	Ex4
No. of iter.	18	183 (113)	6	51 (42)
Time [s]	37	1618 (1015)	27	808 (625)

Table 2 Iteration number and computation time for all examples presented in the paper. The values in the brackets for Ex2 concerns case with $\mu = 50$, and without brackets with $\mu = 25$. The values in and without brackets for Ex4 concerns case without and with noise, respectively.

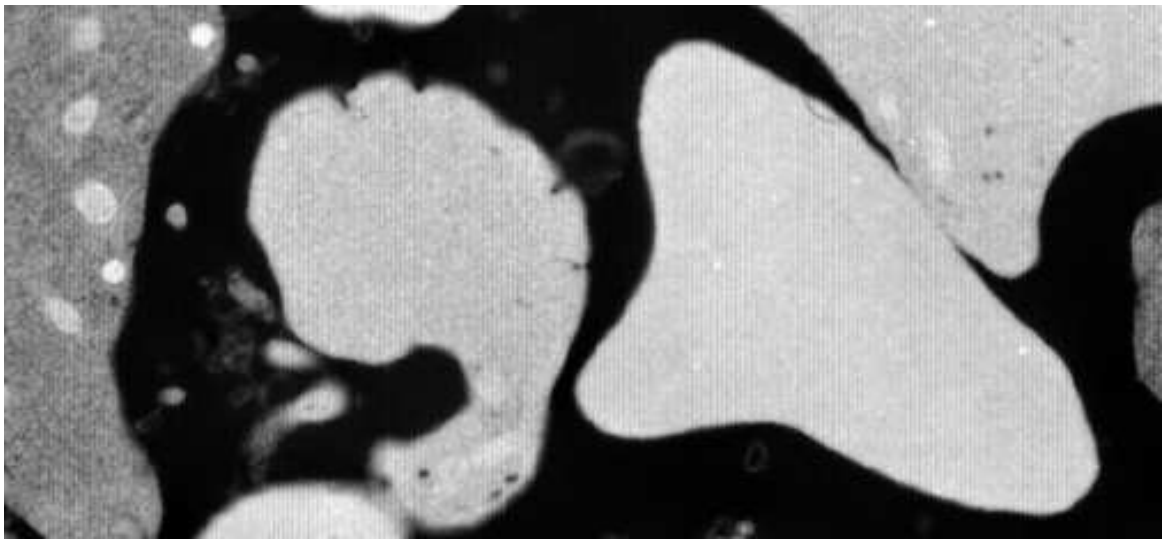
7 Conclusion

In this paper, we have proposed a new quantization method based on a two-step procedure intertwining a convex optimization algorithm for quantization level selection and a combinatorial regularization procedure. Unlike classical methods, the proposed approach allows us to enforce a tunable spatial regularity in the quantized image. We have also shown that both grey scale and color images can be processed. As shown by our simulation results, the proposed approach leads to promising results, in particular in the presence of noise. As future work, we plan to explore isotropic regularization methods, to adapt and implement faster combinatorial algorithms and to take advantage of this method in various applications such as image compression and multispectral/hyperspectral imaging.

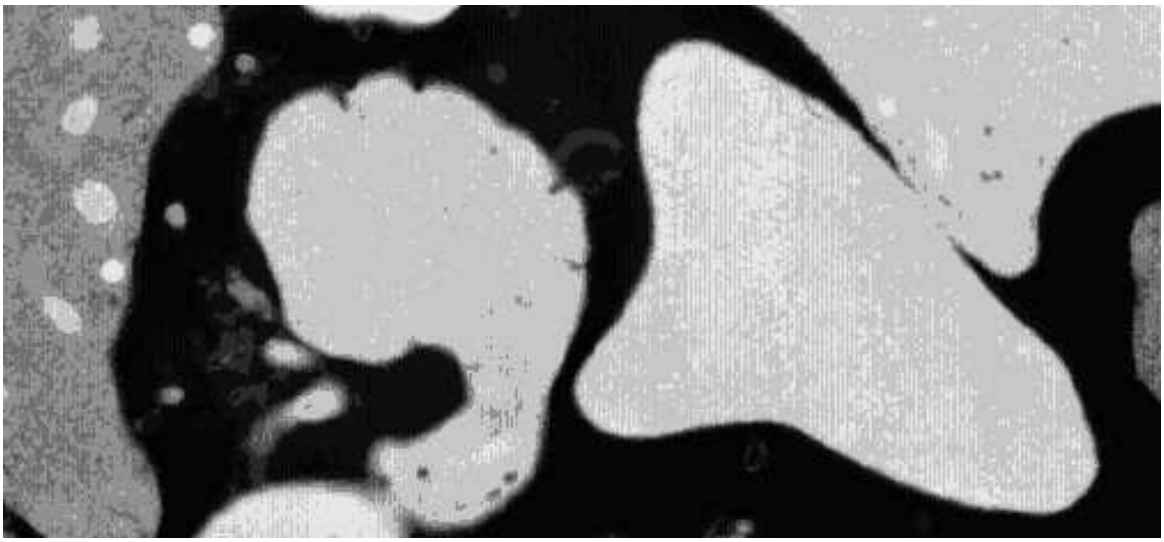
References

1. A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, MA, US, 1992.
2. J. Max. Quantizing for minimum distortion. *IRE Trans. on Inform. Theory*, 6(1):7–12, Mar. 1960.
3. S. Lloyd. Least squares quantization in PCM. *IEEE Trans. Inform. Theory*, 28(2):129–137, Mar. 1982.
4. Y. Linde, A. Buzo, and R. Gray. An algorithm for vector quantizer design. *IEEE Trans. Commun.*, 28(1):84–95, Jan. 1980.
5. X. Wu. On convergence of Lloyd's method I. *IEEE Trans. Inform. Theory*, 38(1):171–174, Jan. 1992.
6. Q. Du, M. Emelianenko, and L. Ju. Convergence of the Lloyd algorithm for computing centroidal voronoi tessellations. *SIAM J. Numer. Anal.*, 44(1):102–119, Feb. 2006.
7. V. Arya, A. Mittal, and R.C. Joshi. An efficient coding method for teleconferencing and medical image sequences. *International Conference on Intelligent Sensing and Information Processing*, pages 8–13, 2005.
8. R. Kawada, A. Koike, and Y. Nakajima. Prefilter control scheme for low bitrate tv distribution. *IEEE International Conference on Multimedia and Expo*, pages 769–772, 2006.
9. K. Chuang, H. Tzeng, S. Chen, J. Wu, and T. Chen. Fuzzy C-means clustering with spatial information for image segmentation. *Comput. Med. Imag. Graph.*, 30:9–15, 2006.
10. L. Alvarez and J. Esclarín. Image quantization using reaction-diffusion equations. *SIAM J. Appl. Math.*, 57(1):153–175, 1997.
11. S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, England, Mar. 2004.
12. J. D. Bruce. Optimum quantization. Technical Report 429, Massachusetts Institute of Technology, Research Laboratory of Electronics, Cambridge, Massachusetts, Mar. 1965.
13. X. Wu and J. Rokne. An $O(KN \log N)$ algorithm for optimum k-level quantization on histograms of n points. In *ACM Annual Computer Science Conference*, pages 339 – 343, Louisville, Kentucky, Jul. 1989.
14. X. Wu. Optimal quantization by matrix searching. *Journal of algorithms*, 12(4):663–673, 1991.
15. R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Springer, Oxford Oxfordshire, 2004.
16. O. Veksler. *Efficient graph-based energy minimization methods in computer vision*. PhD thesis, Cornell University, Ithaca, NY, USA, 1999.
17. C. Vertan, V. Popescu, and V. Buzuloiu. Morphological like operators for color images. In *Proc. Eur. Sig. and Image Proc. Conference*, Trieste, Italy, Sep., 10-13 1996.
18. H. Talbot, C. Evans, and R. Jones. Complete ordering and multivariate mathematical morphology. In *ISMM '98: Proceedings of the fourth international symposium on Mathematical morphology and its applications to image and signal processing*, pages 27–34, Norwell, MA, USA, 1998. Kluwer Academic Publishers.
19. J. Chanussot and P. Lambert. Bit mixing paradigm for multivalued morphological filters. In *Image Processing and Its Applications, 1997., Sixth International Conference on*, volume 2, pages 804–808, Dublin, Jul. 1997.
20. B. Hill, Th. Roger, and F. W. Vorhagen. Comparative analysis of the quantization of color spaces on the basis of the CIELAB color-difference formula. *ACM Trans. Graph.*, 16(2):109–154, 1997.
21. C. Eckart and G. Young. The approximation of one by another of lower rank. *Psychometrika*, 1(3):211–218, Sep. 1936.
22. P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. Springer-Verlag, New York, 2010.
23. J. B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms*. Grundlehren 305, 306. Springer Verlag, 1993.
24. J.-C. Pesquet. A parallel inertial proximal optimization method. Preprint, 2010.
25. P. L. Combettes and J.-C. Pesquet. A proximal decomposition method for solving convex variational inverse problems. *Inverse Problems*, 24(6), Dec. 2008.
26. S. Setzer, G. Steidl, and T. Teuber. Deblurring Poissonian images by split Bregman techniques. *J. Vis. Comm. Image Repr.*, 21(3):193–199, 2010.
27. T. Goldstein and S. Osher. The split Bregman method for L1-regularized problems. *SIAM J. Imaging Sciences*, 2(2):323–343, Apr. 2009.
28. M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo. Fast image recovery using variable splitting and constrained optimization. *IEEE Trans. Image Process.*, 19:2345–2356, 2010.
29. M. Afonso, J. Bioucas-Dias, and M. Figueiredo. A fast algorithm for the constrained formulation of compressive image reconstruction and other linear inverse problems. In *Proc. Int. Conf. Acoust., Speech Signal Process.*, Dallas, USA, Mar. 2010.
30. J. J. Moreau. Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France*, 93:273–299, 1965.
31. C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs. A variational formulation for frame based inverse problems. *Inverse Problems*, 23:1495–1518, Jun. 2007.
32. Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Int.*, 23(11):1222–1239, Nov. 2001.
33. H. Ishikawa and D. Geiger. Mapping image restoration to a graph problem. In *IEEE-EURASIP Workshop Nonlinear Signal Image Process.*, pages 189–193, Antalya, Turkey, Jun. 20-23 1999.
34. J. L. R. Ford and D. R. Fulkerson. *Flows in Networks*. Princeton University Press, Princeton, NJ, 1962.
35. H. Ishikawa. Exact optimization for Markov random fields with convex priors. *IEEE Trans. Pattern Anal. Mach. Int.*, 25(10):1333–1336, Oct. 2003.
36. V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Trans. Pattern Anal. Mach. Int.*, 26(2):147–159, 2004.
37. A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.*, 20(1–2):89–97, 2004.
38. J. Lellmann, J. Kappes, J. Yuan, F. Becker, and C. Schnörr. Convex multi-class image labeling by simplex-constrained total variation. In *SSVM '09: Proceedings of the Second International Conference on Scale Space and Variational Methods in Computer Vision*, pages 150–162, Berlin, Heidelberg, 2009. Springer-Verlag.
39. W. Trobin, T. Pock, D. Cremers, and H. Bischof. Continuous energy minimization via repeated binary fusion. In *ECCV*, Part IV, pages 677–690, Marseille, France, Oct. 12-18 2008.
40. M. Frahm J. M. Zach, C. Niethammer. Continuous maximal flows and wulff shapes: Application to MRFs. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1911–1918, Miami, FL, 2009.
41. O. Veksler. Graph cut based optimization for MRFs with truncated convex priors. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, Minneapolis, MN, 2007.
42. N. Komodakis, G. Tziritas, and N. Paragios. Performance vs computational efficiency for optimizing single and dynamic

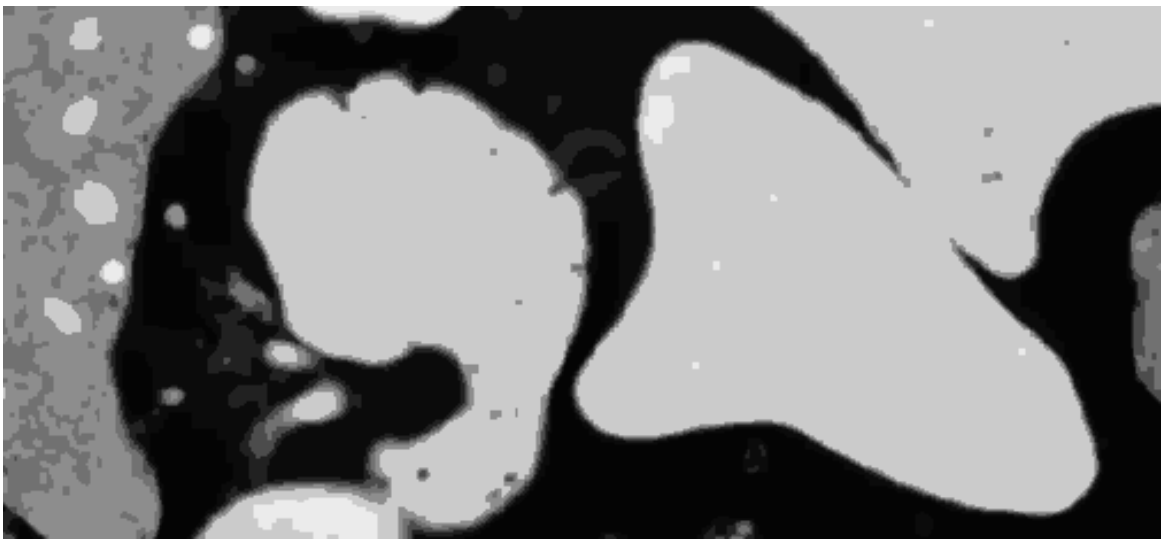
MRFs: Setting the state of the art with primal-dual strategies. *Computer Vision and Image Understanding*, 112(1):14–29, 2008.
43. N. Komodakis and G. Tziritas. Approximate labeling via graph cuts based on linear programming. *IEEE Trans. Pattern Anal. Mach. Int.*, 29(8):1436–1453, 2007.
44. J. Darbon and M. Sigelle. Image restoration with discrete constrained total variation part ii: Levelable functions, convex priors and non-convex cases. *JMIV*, 26(3):277 – 291, Dec. 2006.
45. B. Zalesky. Efficient determination of Gibbs estimators with submodular energy functions. http://arxiv.org/abs/math/0304041, 2003.
46. H. Ishikawa. Higher-order clique reduction in binary graph cut. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2993–3000, Miami, FL, 2009.
47. A. Chambolle and J. Darbon. On total variation minimization and surface evolution using parametric maximum flows. *Int. J. Comput. Vision*, 84(3):288 – 307, 2009.
48. C. Couprie, L. Grady, L Najman, and H. Talbot. Power watersheds: A new image segmentation framework extending graph cuts, random walker and optimal spanning forest. In *Proceedings of the 11th IEEE International Conference on Computer Vision (ICCV)*, pages 731–738, Kyoto, Japan, 2009.
49. X. Wu. On initialization of Max's algorithm for optimum quantization. *IEEE Trans. on Communications*, 38(10):1653–1656, Oct. 1990.
50. Z. Peric and J. Nikolic. An effective method for initialization of Lloyd-Max's algorithm of optimal scalar quantization for Laplacian source. *Informatica*, 18(2):279–288, 2007.
51. I. Katsavounidis, C.-C. J. Kuo, and Z. Zhang. A new initialization technique for generalized Lloyd iteration. *IEEE Sig. Proc. Let.*, 1(10):144–146, Oct. 1994.
52. P. Heckbert. Color image quantization for frame buffer display. *SIGGRAPH Comput. Graph.*, 16(3):297–307, Jul. 1982.
53. X. Wu. Color quantization by dynamic programming and principal analysis. *ACM Transactions on Graphics*, 11(4):348–372, 1992.
54. Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Mach. Int.*, 26:359–374, 2001.
55. Y. Ohno. CIE fundamentals for color measurements. In *IS&T NIP16 Conference*, Vancouver, Canada, Oct. 16-20 2000.
56. V.S. Lempitsky, C. Rother, S. Roth, and A. Blake. Fusion moves for markov random field optimization. *IEEE Trans. Pattern Anal. Mach. Int.*, 32(8):1392–1405, 2010.
57. V. Kolmogorov. A note on the primal-dual method for semi-metric labeling problem. Technical report, UCL, 2007.
58. V. Kolmogorov and A. Shioura. New algorithms for convex cost tension problem with application to computer vision. *Discrete Optimization*, 6(4):378–393, 2009.



(a)



(b)



(c)

Fig. 3 Figures (a,b,c) illustrate a fragment of the original image, LM and our results, respectively. Note that LM retained acquisition vertical artifacts, which are absent in our result.

(a)

(b)

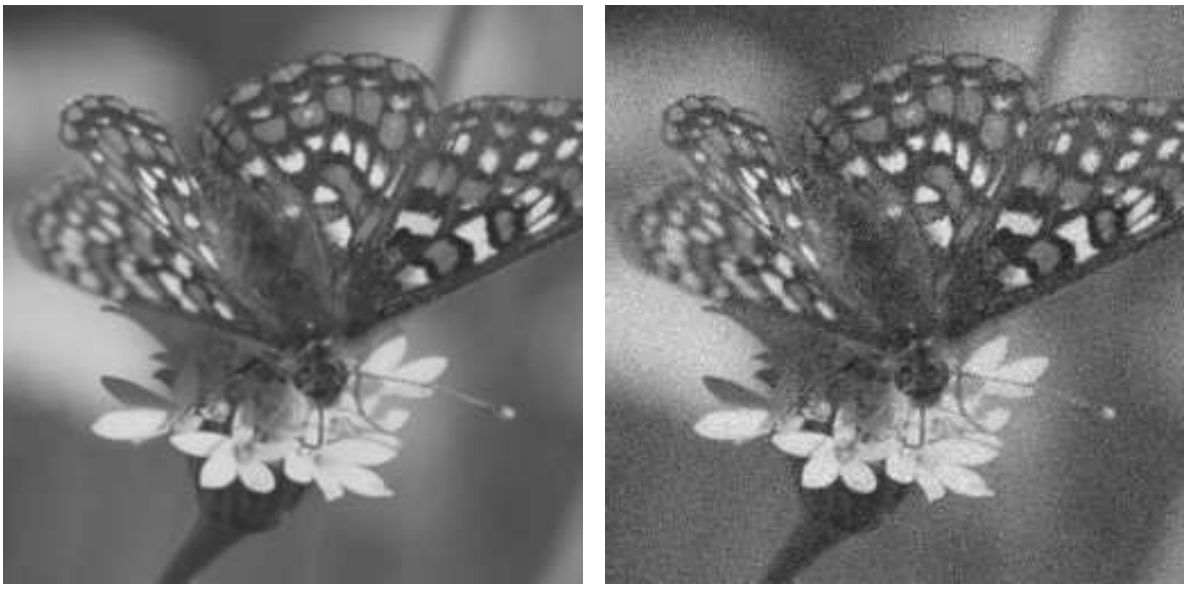
(c)

(d)

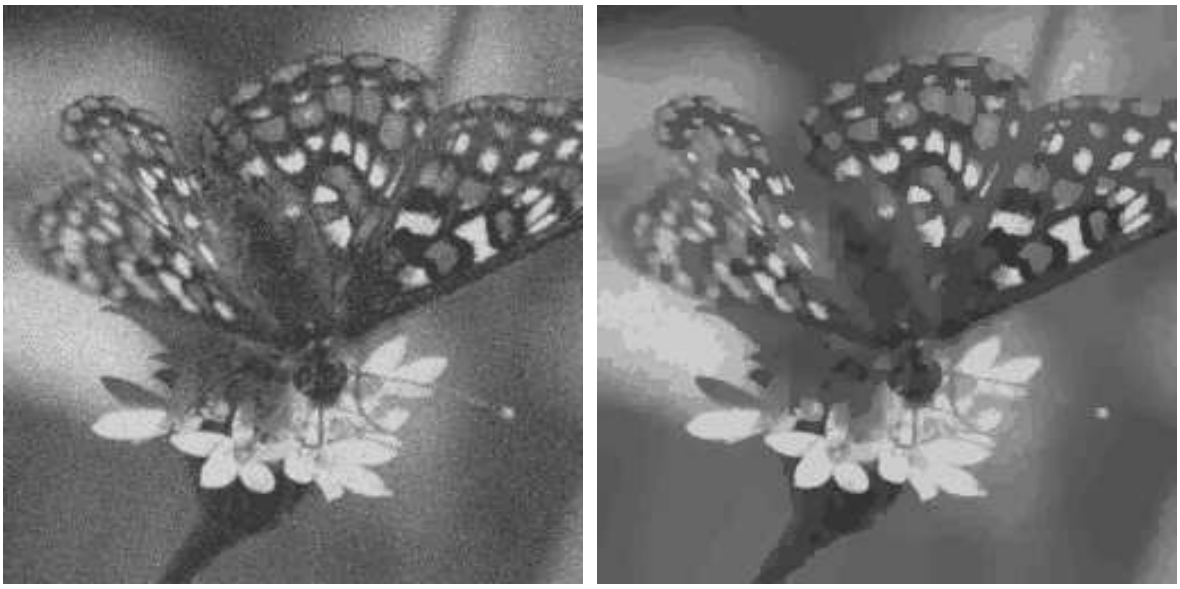
(e)

(f)

Fig. 4 (a) is the original image, (b,c,d) the Wu, K-means and median-cut results respectively ; (e,f) show our result for μ equal 25 and 50 respectively. Note that there are many isolated small regions in (b,c,d), while both (e) and (f) feature only smooth large regions, retaining global aspect nonetheless.

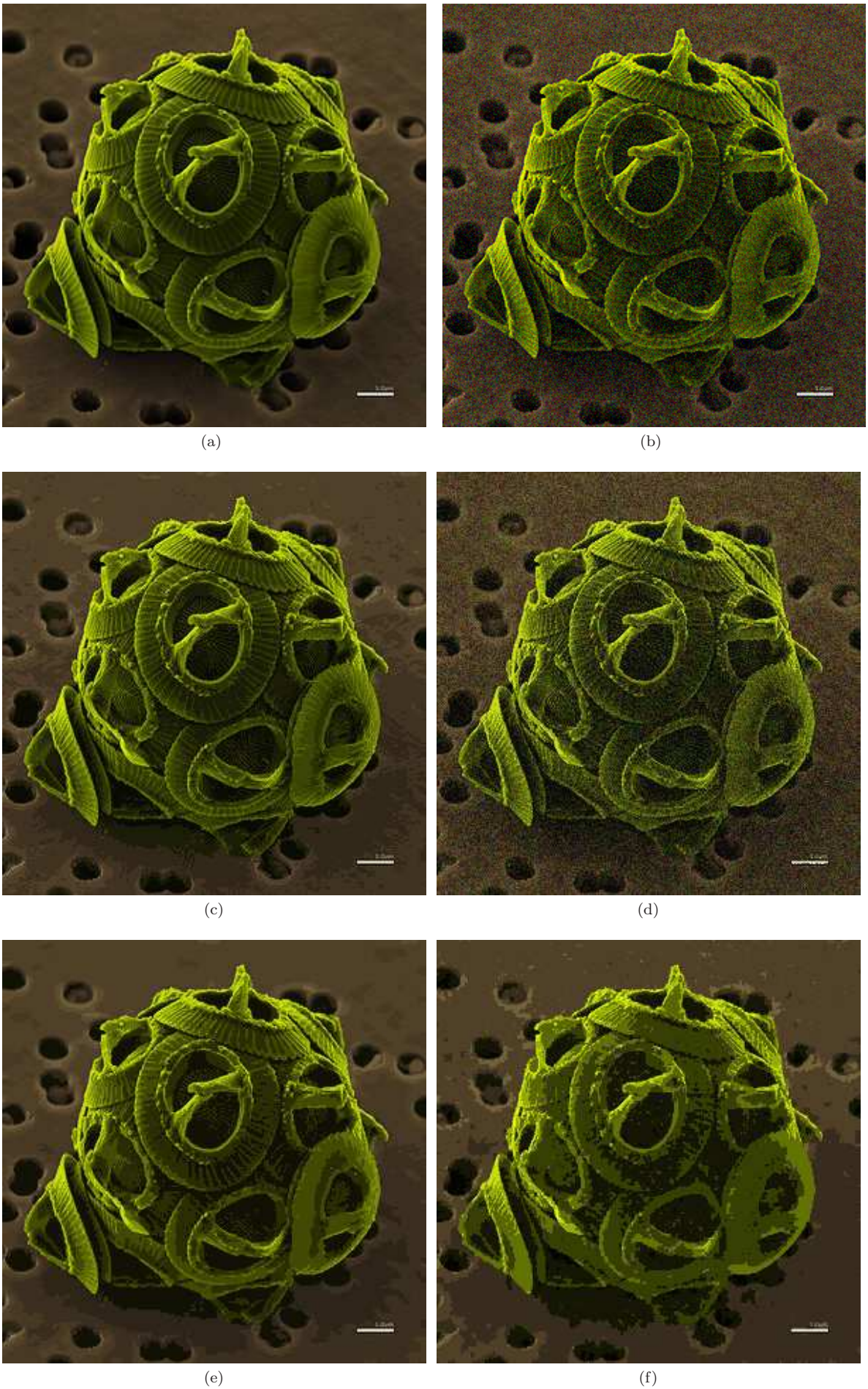


(a) (b)



(c) (d)

Fig. 5 (a) original image, (b) noisy version, and (c,d) LM and our result, respectively.



(a) (b)

(c) (d)

(e) (f)

Fig. 6 (a) original image, (b) its noisy version, (c,e) and (d,f) K-means and our result for clear and noisy case, respectively.