

Luminance-Chrominance Model for Image Colorization*

F. Pierre[†], J.-F. Aujol[‡], A. Bugeau[§], N. Papadakis[‡], and V.-T. Ta[¶]

Abstract. This paper provides a new method to colorize gray-scale images. While the computation of the luminance channel is directly performed by a linear transformation, the colorization process is an ill-posed problem that requires some priors. In the literature two classes of approach exist. The first class includes manual methods that need the user to manually add colors on the image to colorize. The second class includes exemplar-based approaches where a color image, with a similar semantic content, is provided as input to the method. These two types of priors have their own advantages and drawbacks. In this paper, a new variational framework for exemplar-based colorization is proposed. A nonlocal approach is used to find relevant color in the source image in order to suggest colors on the gray-scale image. The spatial coherency of the result as well as the final color selection is provided by a nonconvex variational framework based on a total variation. An efficient primal-dual algorithm is provided, and a proof of its convergence is proposed. In this work, we also extend the proposed exemplar-based approach to combine both exemplar-based and manual methods. It provides a single framework that unifies advantages of both approaches. Finally, experiments and comparisons with state-of-the-art methods illustrate the efficiency of our proposal.

Key words. colorization, optimization, non-local methods

AMS subject classifications. 68U10, 94A08, 65K10, 49M29

DOI. 10.1137/140979368

1. Introduction. The colorization of a gray-scale image consists of adding color information to it. It is useful in the entertainment industry to make old productions more attractive. The reverse operation is based on perceptual assumptions and is today an active research area [28], [13], [37]. Colorization can also be used to add information in order to help further analysis of the image by a user (e.g., sensor fusion [43]). It can also be used for art restoration; see, e.g., [17] or [41]. It is an old subject that began with the ability of screens and devices to display colors. A seminal approach consists in mapping each level of gray into a color-space [18]. Nevertheless, all colors cannot be recovered without an additional prior. In the existing approaches, priors can be added in two ways: with a direct addition of color on

*Received by the editors July 28, 2014; accepted for publication (in revised form) December 1, 2014; published electronically March 10, 2015. The research of the authors was supported by the French state, managed by the French National Research Agency (ANR) in the frame of the Investments for the Future Programme IdEx Bordeaux (ANR-10-IDEX-03-02).

<http://www.siam.org/journals/siims/8-1/97936.html>

[†]IMB, UMR 5251, Université de Bordeaux, F-33400 Talence, France, and IMB, UMR 5251, CNRS, F-33400 Talence, France, and LaBRI, UMR 5800, Université de Bordeaux, F-33400 Talence, France, and LaBRI, UMR 5800, CNRS, F-33400 Talence, France (fabien.pierre@math.u-bordeaux.fr).

[‡]IMB, UMR 5251, Université de Bordeaux, F-33400 Talence, France, and IMB, UMR 5251, CNRS, F-33400 Talence, France (jean-francois.aujol@math.u-bordeaux.fr, nicolas.papadakis@math.u-bordeaux.fr).

[§]LaBRI, UMR 5800, Université de Bordeaux, F-33400 Talence, France, and CNRS, LaBRI, UMR 5800, CNRS, F-33400 Talence, France (aurelie.bugeau@labri.fr).

[¶]LaBRI, UMR 5800, CNRS, F-33400 Talence, France, and LaBRI, UMR 5800, Bordeaux INP, F-33402 Talence, France (ta@labri.fr).

the image performed by a user or by providing a color image as an example (also called a *source* image). In the rest of this paper, we call the gray-scale image to colorize a *target*.

1.1. State-of-the-art. In the first category of methods, a user manually adds points of color (also called *scribbles*) on the target image. Numerous methods have been proposed based on this type of approach. For instance, the method of Levin, Lischinski, and Weiss [30] solves an optimization problem to diffuse the scribbles onto the target with the assumption that chrominances must have small variations if the luminance has small variations. Yatziv and Sapiro [42] propose a simple but fast method by using geodesic distances to blend the chrominances given by the scribbles. Heu et al. [22] use pixel priorities to ensure that important areas end up with the right colors. Other propagation schemes have been proposed: for instance, probabilistic distance transform [29], random walks [26], discriminative textural features [25], structure tensors [16], or nonlocal graph regularization [31]. As often described in the literature, with these color diffusion approaches, the contours are not well preserved. Quang, Kang, and Le [35] propose a variational approach in the chromaticity and brightness space to interpolate the missing colors. Reproducing kernel Hilbert spaces are used to perform a link between the chrominance and the brightness channels. In [15], the scribbles are automatically generated after segmenting the image and the user only needs to associate one color to each scribble. The colorization is further performed by computing quaternion wavelet phases so that the color is propagated along equal phase lines. The contours are therefore well preserved. As in the case with all manual methods, this latter approach suffers from the following drawbacks: if the target represents a complex scene, the segmentation might not be very accurate and the user interaction becomes very important.

In exemplar-based colorization methods, the color information is provided by a *source* color image selected by the user. The first exemplar-based method is the one proposed by Welsh, Ashikhmin, and Mueller [40] (derived from a texture synthesis algorithm [39]). It uses patch similarities in the colorization process. The authors of [40] also propose manual information (called *swatches*) to specify where to search patches in the source image. Di Blasi and Reforgiato [14] propose an improvement that accelerates the search of patches with tree-clustering. Chen and Ye [10] propose an improvement based on a Bayesian image matting. Generally, exemplar-based approaches suffer from spatial consistency problems since each pixel is processed independently. To overcome this limitation, several works use image segmentation to improve the colorization results. For instance, Irony, Cohen-Or, and Lischinski [23] propose to compute the best matches between the target pixels and regions in a presegmented source image. With these correspondences, microscribbles from the source are initialized on the target image and colors are finally propagated as in [30]. In [38], the authors use image segmentation to colorize cartoon images. Chia et al. [11] take advantage of the huge amount of images available on the Internet. Nevertheless, the user has to manually segment and label the objects of the target image. Next, for each labeled object, images with the same label are found on the Internet and used as source images. The image retrieval step relies on superpixels extraction [36] and graph-based optimization. Gupta et al. [20] extract different features from the superpixels [36] of the target image and match them with the source ones. The final color for each pixel is computed by optimizing a criterion imposing spatial consistency as in [30]. Charpiat, Hofmann, and Schölkopf [8] ensure spatial coherency without segmenting, but their method involves many complex steps.

1.2. The approach of Bugeau, Ta, and Papadakis [4]. Without requiring image segmentation or superpixels extraction, Bugeau, Ta, and Papadakis [4] compute a set of color candidates for each target pixel by matching patches with the source image using different features.

In the following, the target image is considered to correspond to the luminance channel Y of the YUV color-space. In order to preserve the initial gray-scale image content, colorization methods always impose that the luminance channel Y must be maintained. As most existing colorization methods do, [4] only computes the two chrominance channels (here U and V). Their method works as follows. First, the source image is converted into a luminance one. In order to be comparable to the target, an affine remapping [21] is applied to the source luminance. Next, for each pixel of the target image, the method extracts eight chrominance candidates c_i with $i = 1, \dots, 8$ from the source. To that end, the patch centered on each pixel in the target is compared to randomly chosen luminance patches in the source. (See [40] and [4] for more details.) Different features and different size of patches are used, with each configuration (eight in total) leading to one candidate. The best chrominance is finally chosen within the eight candidates with an energy-based method. To ensure the regularity of the resulting image u , the model includes a total variation (TV) regularization of the U and V channels.

The approach proposed by Bugeau, Ta, and Papadakis [4], which we call *chrominance model*, is invariant w.r.t. the scene illumination, but the method only retains the U and V values. This extraction is not hue consistent and can produce new colors unrelated to those of the source. Moreover, in their regularization algorithm, there is no coupling of the chrominance channels with the luminance leading to halo effects near strong contours. In order to reduce this effect, a strong regularization is needed, leading to the drab results mentioned in [4]. To overcome this limitation, a postprocessing is applied on the chrominance channels, but the contours are not always well preserved (see [33]). Another drawback is the computational time due to projections onto the simplex. Nevertheless, this approach is simple and provides promising results.

1.3. Contributions. The optimization algorithm of [4] is inspired by the one of Chambolle and Pock [6] that is dedicated to the convex problem. Some algorithms have been designed for nonconvex problem. The most adapted of them to our framework is probably PALM of Bolte, Sabach, and Teboulle [2], which is much slower than the algorithm we propose here to compute the minimizer of our functional. For the model of [4], it requires an internal loop which slows down the algorithm. To tackle this issue, it is preferable to design a new algorithm dedicated to this particular problem. The model proposed in this paper is inspired by the work of [4], but its formulation has advantages. The proposed functional is convex w.r.t. each of its variables. From this new model, we design a new algorithm, adapted from [6], whose convergence is theoretically proven. Notice that recently there have been some major advances in the minimization of convex optimization; see, e.g., [1] and references therein. However, these approaches rely on the technical Kurdyka–Łojasiewicz (KL) inequality (ensuring that the functional to minimize is not too flat around its minimizer(s)). A main advantage of the approach proposed here is that we do not need to check that the KL inequality is fulfilled.

The *main contributions* of this work are the improvements and the simplification of the model proposed by [4] and the introduction of a coupled regularization term that preserves

image contours during the colorization process. (These two points are presented in the first section.) Another contribution is the design of an algorithm and the proof of its convergence, although the functional to minimize is not convex (this is presented in section 2.3); the proposal of a second simpler algorithm is provided in section 2.4. The proposed model and its implementation are simple, and we can easily demonstrate its convergence to a critical point. Comparisons with several state-of-the-art methods in terms of colorization results are presented in section 3. For numerical experiments, we propose to use an approach combining both exemplar and manual priors. A preliminary version of this combination was published in [34]. This new paper presents in detail the theoretical framework of the variational model and the proof of convergence of the algorithm.

2. Variational image colorization model. This section describes the two parts of our approach. First, we detail the extraction of color candidates based on texture features. Second, we deal with the choice of the best candidate in order to provide a regular image.

2.1. Candidates extraction. In this section, we describe the method to extract candidates based on texture features. This method has been experimentally chosen in [4].

The target image is considered to correspond to the luminance channel Y of the YUV color-space. First, the source image is converted into a gray-scale one. An affine luminance remapping [21] is applied in order to match the histograms from the source and the target. Next, for each pixel of the target image, the method compares the patch centered on the pixel with some of the gray-scale version of the source. The comparison is performed with different metrics and patch sizes (in order to take into account the scale of textures) based on standard-deviation, discrete Fourier transform (DFT), and cumulative histogram. For each metric, the method retains the pixel in the source image having the nearest patch (see Figure 1). At the end of this step, each pixel of the target is associated with eight pixels from the source. The corresponding chrominances of the source form the set of color candidates. In practice, as in [33], we do not directly use the chrominances from the source. Indeed, they may have different luminances compared to the one of the target pixel. Instead, we first retain the RGB

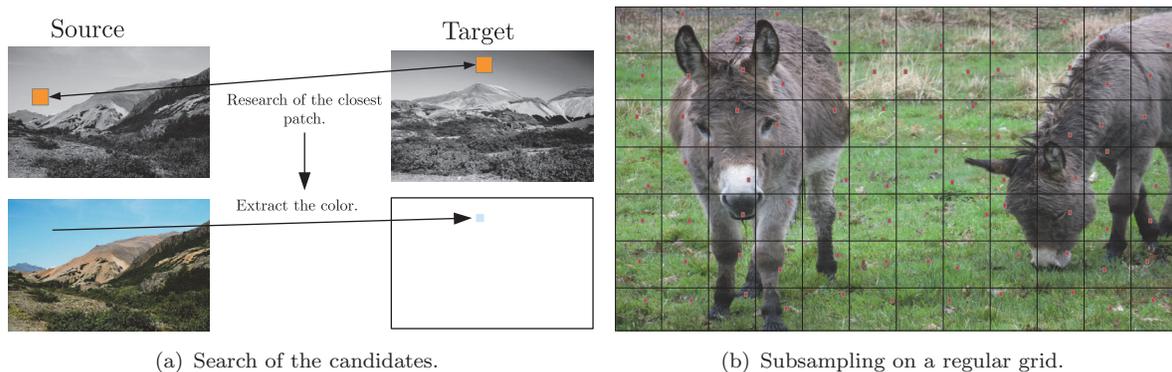


Figure 1. For each pixel of the target image, the method compares the patch centered on the pixel with ones of the gray-scale version of the source. Next, the method retains the color of the central pixel of the closest patch (see (a)). To speed up the algorithm, the search is not performed among all pixels but only on a subsampling (see (b)).

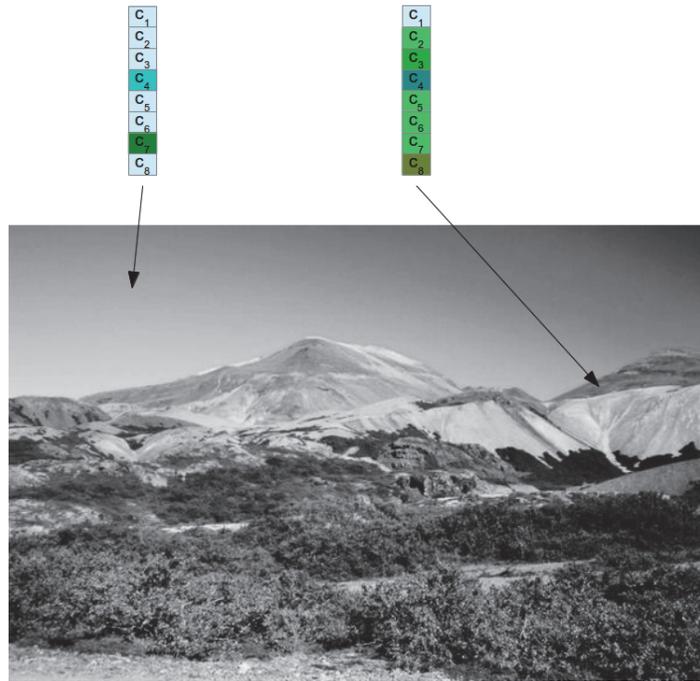


Figure 2. The first part of the method finds C candidates per pixel (here $C = 8$).

color and project it onto the right luminance using an oblique projection that preserves the hue [33].

With various sizes of patches and different features, [4] proposes to retain eight candidates per pixel. A choice has to be made among these candidates to colorize the image. Figure 2 shows an example of a target image and the representation of the set of eight candidates for two pixels. In this paper, the number of retained candidates is denoted by C .

2.2. The luminance-chrominance model. In this work, we propose to improve the chrominance model of [4] by introducing a different TV regularization term and by simplifying their work. Our model is able to couple the channels of luminance and chrominances. It will be called in what follows the *luminance-chrominance* model.

2.2.1. The proposed functional. To choose a candidate among those extracted, we now propose to minimize the following functional, where $u = (U, V)$ stands for chrominances to compute, c_i for the candidates, and $w = \{w_i\}$ with $i = 1, \dots, C$ for the candidate weights:

$$(2.1) \quad F(u, w) := TV_{\mathcal{C}}(u) + \frac{\lambda}{2} \int_{\Omega} \sum_{i=1}^C w_i \|u - c_i\|_2^2 + \chi_{\mathcal{R}}(u) + \chi_{\Delta}(w) .$$

To simplify the notation, the dependence of each values to the position of the current pixel is removed. For instance, the second term of (2.1) denotes $\int_{\Omega} \sum_{i=1}^C w_i(\omega) \|u(\omega) - c_i(\omega)\|_2^2 d\omega$.

This model is a classical one with a fidelity-data term $\int_{\Omega} \sum_{i=1}^C w_i \|u - c_i\|_2^2$ and a regularization term $TV_{\mathcal{C}}(u)$. Since the first step of the method extracts many candidates, we propose

averaging the fidelity-data term issued from each candidate. This average is weighted by w_i . Thus, the term

$$(2.2) \quad \int_{\Omega} \sum_{i=1}^C w_i \|u - c_i\|_2^2$$

connects the candidate color c_i to the color u that will be retained. The minimum of this term w.r.t. u is reached when u is equal to the weighted average of candidates c_i .

Since the average is weighted by w_i , these weights are constrained to be onto the probability simplex. This constraint is formalized by $\chi_{\Delta}(w)$ whose value is 0 if $w \in \Delta$ and $+\infty$ otherwise, with Δ defined as

$$(2.3) \quad \Delta := \left\{ (w_1, \dots, w_C) \text{ s.t. } 0 \leq w_i \leq 1 \text{ and } \sum_{i=1}^C w_i = 1 \right\}.$$

If $w_j = 1$ and $w_i = 0$ for $i \neq j$, then the minimum w.r.t. u is reached for $u = c_j$ and it provides a label. The joint minimization of the term (2.2) w.r.t. u and w provides a natural labeling, where c_i are labels. In addition, the characteristic function $\chi_{\mathcal{R}}(u)$ constrains the result to be in the set \mathcal{R} , which is the standard range for the chrominances.

In (2.1), $TV_{\mathcal{C}}$ is the regularization term which will be defined in (2.4). It favors the most regular solution among all the labelings. The parameter λ weights the influence of the regularization term of the model.

2.2.2. The coupled total variation. The chrominance model of [4] suffers from a lack of coupling in the regularization term, leading to halo effects in the colorization results. To cope with this issue, we introduce a TV regularization which is able to couple the chrominance channels with the luminance one. Although [33] and [32] propose a simple coupling performed by working directly in the RGB color-space, the convergence of these algorithms is not established. By coupling channels in the YUV color-space, we propose in this paper a model and an algorithm for which the convergence is proven.

Let $TV_{\mathcal{C}}$ be a *coupled* total variation defined as

$$(2.4) \quad TV_{\mathcal{C}}(u) = \int_{\Omega} \sqrt{\gamma \partial_x Y^2 + \gamma \partial_y Y^2 + \partial_x U^2 + \partial_y U^2 + \partial_x V^2 + \partial_y V^2},$$

where Y , U , and V are the luminance and chrominance channels. γ is a parameter which enforces the coupling of the channels. Some others total variation formulations have been proposed to couple the channels; see, for instance, [24] or [5].

As Figure 3 illustrates, this formulation naturally favors images where contours in chrominance channels are at the same location as the luminance one. For the sake of clarity, assume that there is a vertical contour in the Y channel, $\partial_x Y = a > 0$ and $\partial_y Y = 0$, and another one in the U channel such that $\partial_x U = b > 0$, $\partial_y U = \partial_x V = \partial_y V = 0$. If the two contours are at the same location, the value of the total variation is equal to $TV_{\mathcal{C}}(u) = \sqrt{\gamma a^2 + b^2}$, but if the contours have different locations, the value is equal to $TV_{\mathcal{C}}(u) = \sqrt{\gamma a^2} + \sqrt{b^2}$. Since $\sqrt{\gamma a^2 + b^2} < \sqrt{\gamma a^2} + \sqrt{b^2}$ when $\gamma > 0$, the minimization of $TV_{\mathcal{C}}$ encourages the values of U such that the contours in the chrominance channels are in the same location as the luminance

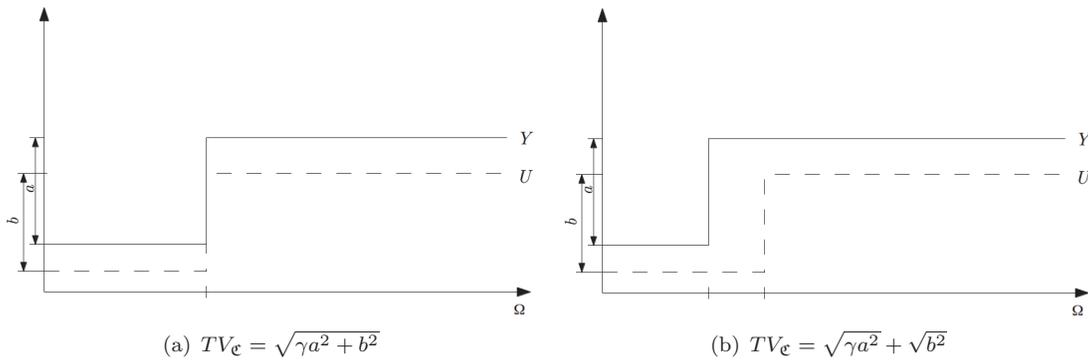


Figure 3. Illustration of two situations for contours in luminance and chrominance channels. Since $\sqrt{\gamma a^2 + b^2} < \sqrt{\gamma a^2} + \sqrt{b^2}$ when $\gamma > 0$, TV_{ϵ} favors coupling of channels, contrary to the TV on chrominance channels only, which values are equal for the situations (a) and (b).

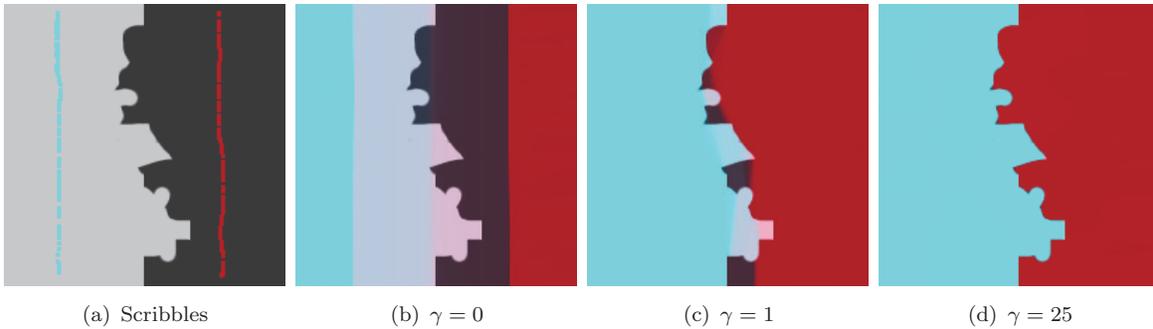


Figure 4. Comparison of the TV on chrominance channels only (b) and our TV_{ϵ} term (2.4) (c) and (d) with different values of γ .

one. In some cases, the minimization of TV_{ϵ} will prefer the shortest contour perimeter for the chrominance channels, rather than the coupling with the luminance one. Adding a high value for γ limits this problem by enforcing the coupling between luminance and chrominance channels.

To illustrate the advantage of TV_{ϵ} , we propose a reduced model that diffuses colors manually added by the user on some pixels of the gray-scale image. In other words, we use this coupled TV within a manual colorization method by minimizing the following functional:

$$(2.5) \quad \min_u TV_{\epsilon}(u) + \frac{\lambda}{2} \|\mathcal{M}(u - f)\|_2^2 + \chi_{\mathcal{R}}(u),$$

where f are chrominances defined by the user, \mathcal{M} is a mask which is equal to 1 if the user puts a color on the gray-scale image, and 0 otherwise. u is the chrominance to compute. λ is a parameter which rules the influence of the regularization. To solve this problem, it is possible to use the algorithm of [6]. Experiments on a toy example are presented in Figure 4. They illustrate the advantage of TV_{ϵ} compared to a TV on chrominance channels only (as proposed in [4]). With the TV on chrominance channels only (Figure 4(b)), no coupling with the luminance channel is available, and the transition between the red and the blue color is

computed without considering the main gray-scale image contour. With our coupled total variation (2.4) and $\gamma = 1$ (Figure 4(c)), the coupling is done, but the contour is not perfectly respected since the regularization term here favors the shortest contour perimeter on the chrominance channels. By increasing the γ value (Figure 4(d)), the contour is well preserved since we enforce the minimization to preserve the Y channel. In experimental results, we choose $\gamma = 25$.

2.3. A first algorithm and its convergence. Recently, a primal-dual algorithm has been proposed by Chambolle and Pock with various applications to image processing [6]. It inspired us to provide a new algorithm dedicated to a particular class of problems based on a nonconvex functional. To apply this technique, we express our functional into a min-max problem. To this end, we rewrite the term $TV_{\mathfrak{C}}$ from (2.4) in a dual form. In a Hilbert space with norm $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$,

$$(2.6) \quad \|f\| = \max_{\|x\| \leq 1} \langle f|x \rangle,$$

which can also be rewritten as

$$(2.7) \quad \|f\| = \max_x \langle f|x \rangle - \chi_{\|x\|_2 \leq 1}.$$

In \mathbb{R}^6 , it becomes

$$(2.8) \quad \left\| \sqrt{\|q_1\|^2 + \|q_2\|^2 + \|q_3\|^2} \right\|_1 = \max_{p_1, p_2, p_3} \langle q_1, q_2, q_3 | p_1, p_2, p_3 \rangle - \chi_{\|(p_1, p_2, p_3)\|_2 \leq 1},$$

where $p, q \in \mathbb{R}^6$ and $p_j, q_j \in \mathbb{R}^2$, $j = 1, \dots, 3$.

Applied to $q_1 = \nabla U$, $q_2 = \nabla V$, and $q_3 = \gamma \nabla Y$ with $\gamma > 0$,

$$(2.9) \quad \left\| \sqrt{\|\nabla U\|^2 + \|\nabla V\|^2 + \gamma \|\nabla Y\|^2} \right\|_1 = \max_{p_1, p_2, p_3} \langle \nabla U, \nabla V | p_1, p_2 \rangle + \gamma \langle \nabla Y | p_3 \rangle - \chi_{\|(p_1, p_2, p_3)\|_2 \leq 1}.$$

Thus,

$$(2.10) \quad TV_{\mathfrak{C}}(U, V) = \max_{p=(p_1, p_2, p_3)} \langle \nabla U, \nabla V | p_1, p_2 \rangle + \gamma \langle \nabla Y | p_3 \rangle - \chi_{\|(p_1, p_2, p_3)\|_2 \leq 1}.$$

Minimizing (2.1) is equivalent to maximizing the dual model w.r.t. to the dual variable p and to minimizing it w.r.t. u and w .

Our problem (2.1) is a particular case of

$$(2.11) \quad \min_{u \in \mathcal{U}} \min_{w \in \mathcal{W}} \max_{p \in \mathcal{P}} \langle Ku | p \rangle - F^*(p) + G(u) + h(u, w) + H(w),$$

where $G : \mathcal{U} \rightarrow [0, +\infty)$, $F^* : \mathcal{P} \rightarrow [0, +\infty)$, $H : \mathcal{W} \rightarrow [0, +\infty)$, and $h : (\mathcal{U} \times \mathcal{W}) \rightarrow [0, +\infty)$ are proper lower semicontinuous functions. F^* , G , and H are convex, and h is convex w.r.t. each of its variables separately. K is a continuous linear mapping. $\forall u \in \mathcal{U}$, $h(u, \cdot) + H$ is lower semicontinuous, proper, and convex. This last statement can be replaced by a weaker one: in all the following, it is sufficient that $\forall u \in \mathcal{U}$, $h(u, \cdot) + H$ is lower semicontinuous, coercive, and proper. Remark that the functional is not convex w.r.t. (u, w) .

To apply this general model to (2.1), we can set $H = \chi_\Delta$, $K = \nabla$, $h(u, w) = \int_\Omega \sum_{i=1}^C w_i \|u - c_i\|_2^2$, and $F^* = \chi_{B_2(0,1)}$. Remark that h is a biconvex function (see, e.g., [19]). This function has many partial minima which are not generally global. We now define partial optimum.

Definition 2.1. *Let B be a biconvex set and $f : B \rightarrow \mathbb{R}$ be a given function and let $(x, y) \in B$. Then, (x, y) is called a partial optimum of f on B if*

$$(2.12) \quad f(x, y) \leq f(x, y) \quad \forall x \in B_y \text{ and } f(x, y) \leq f(x, y) \quad \forall y \in B_x$$

with $B_x := \{y \in Y : (x, y) \in B\}$ and $B_y := \{x \in X : (x, y) \in B\}$.

Since h is differentiable, we can define stationary points.

Definition 2.2. *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a given function, let $\zeta \in \mathbb{R}^n$, and let the partial derivatives of f in ζ exist. If $\nabla f(\zeta) = 0$, then ζ is called a stationary point of f .*

The following proposition provides a characterization of partial optimum.

Proposition 2.3. *Let $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ be a differentiable, biconvex function. Then a point $z \in \mathbb{R}^{n+m}$ is stationary if and only if z is a partial optimum.*

Since functional (2.11) is not convex, the definition of a saddle-point for the primal variables (u, w) and the dual one p does not make sense. So, let us extend the definition of saddle-points. Consider a generic saddle-point problem (2.11). For a given $w \in \mathcal{W}$, the reduced problem

$$(2.13) \quad \min_{u \in \mathcal{U}} \max_{p \in \mathcal{P}} \langle Ku|p \rangle - F^*(p) + G(u) + h(u, w)$$

is equivalent to the one of [6], and it admits a saddle-point (\hat{u}, \hat{p}) such that $\forall (u, p, w) \in \mathcal{U} \times \mathcal{P} \times \mathcal{W}$:

$$(2.14) \quad \langle Ku|\hat{p} \rangle - F^*(\hat{p}) + G(u) + h(u, w) \geq \langle K\hat{u}|p \rangle - F^*(p) + G(\hat{u}) + h(\hat{u}, w).$$

Moreover, for a given $u \in \mathcal{U}$, the problem

$$(2.15) \quad \min_{w \in \mathcal{W}} H(w) + h(u, w)$$

admits a minimizer \hat{w} such that $\forall w \in \mathcal{W}$,

$$(2.16) \quad H(w) + h(u, w) \geq H(\hat{w}) + h(u, \hat{w}).$$

Definition 2.4. *We call a pseudo saddle-point of the problem (2.11) a point $(\hat{u}, \hat{p}, \hat{w}) \in \mathcal{U} \times \mathcal{P} \times \mathcal{W}$ such that $\forall (u, p, w) \in \mathcal{U} \times \mathcal{P} \times \mathcal{W}$,*

$$(2.17) \quad \begin{aligned} &\langle Ku|\hat{p} \rangle - F^*(\hat{p}) + G(u) + h(u, w) + H(w) + h(u, w) \\ &\geq \langle K\hat{u}|p \rangle - F^*(p) + G(\hat{u}) + H(\hat{w}) + h(u, \hat{w}) + h(\hat{u}, w). \end{aligned}$$

The existence of such points is stated in Lemma A.1.

We call a critical pseudo saddle-point of the problem (2.11) a point $(\hat{u}, \hat{p}, \hat{w}) \in \mathcal{U} \times \mathcal{P} \times \mathcal{W}$ such that $\forall (u, p, w) \in \mathcal{U} \times \mathcal{P} \times \mathcal{W}$,

$$(2.18) \quad \begin{aligned} &\langle Ku|\hat{p} \rangle - F^*(\hat{p}) + G(u) + h(\hat{u}, w) + H(w) + h(u, \hat{w}) \\ &\geq \langle K\hat{u}|p \rangle - F^*(p) + G(\hat{u}) + H(\hat{w}) + 2h(\hat{u}, \hat{w}). \end{aligned}$$

2.3.1. General algorithm. Algorithm 1 is the most general primal-dual-like algorithm presented in this paper.

ALGORITHM 1. PRIMAL-DUAL-LIKE ALGORITHM.

```

1: for  $n \geq 0$  do
2:    $p^{n+1} \leftarrow \text{prox}_{\sigma F^*}(p^n + \sigma K \bar{u}^n)$ 
3:    $w^{n+1} \leftarrow \text{prox}_{\rho H + \rho h(\bar{u}^n, \cdot)}(w^n)$ 
4:    $u^{n+1} \leftarrow \text{prox}_{\tau G + \tau h(\cdot, w^{n+1})}(u^n - \tau K^* p^{n+1})$ 
5:    $\bar{u}^{n+1} \leftarrow 2u^{n+1} - u^n$ 
6: end for

```

Parameters ρ , τ , and σ are the time steps. For a function F proper, convex, and lower semi-continuous, $\text{prox}_F(\hat{p})$ is defined as $\text{argmin}_p \frac{\|p - \hat{p}\|_2^2}{2} + F(p)$; see, e.g., [12].

We recall two definitions used in Theorem 2.7. First, the definition of *isolated points* is presented.

Definition 2.5. *The point of a subset of \mathbb{R}^n are isolated if $\exists \varepsilon > 0$ such that $\forall x, y \in S$, $\|x - y\| \geq \varepsilon$.*

Next, the definition of a *cluster point* is presented.

Definition 2.6. *A cluster point of a sequence is the limit of one of its subsequence.*

Algorithm 1 converges to a critical pseudo saddle-point of the problem (2.11) under some conditions stated in Theorem 2.7.

Theorem 2.7. *Let $L = \|K\|$ and assume that the problem (2.11) admits a saddle-point $(\hat{u}, \hat{p}, \hat{w})$. Choose $\tau\sigma L^2 < 1$, $\rho > 0$ and let (u^n, p^n, w^n) be defined in Algorithm 1. For the sake of simplicity, assume that \mathcal{U} , \mathcal{P} , and \mathcal{W} are of finite dimension.*

(a) *Then $\forall n > 0$*

$$(2.19) \quad \begin{aligned} & \frac{\|p^n - \hat{p}\|_2^2}{2\sigma} + \frac{\|u^n - \hat{u}\|_2^2}{2\tau} + \frac{\|w^n - \hat{w}\|_2^2}{2\rho} \\ & \leq \beta \left(\frac{\|p^0 - \hat{p}\|_2^2}{2\sigma} + \frac{\|u^0 - \hat{u}\|_2^2}{2\tau} + \frac{\|w^0 - \hat{w}\|_2^2}{2\rho} \right), \end{aligned}$$

where $\beta \leq (1 - \tau\sigma L^2)^{-1}$. Thus, the sequence is bounded.

(b) *There exists a cluster point which is a fixed-point of Algorithm 1.*

(c) *Assume that the critical pseudo saddle-points are isolated. Thus there exists a fixed-point (u^*, p^*, w^*) such that the sequence (u^n, p^n, w^n) converges to it.*

The proof of convergence of Theorem 2.7, given in Appendix B, needs the existence of a pseudo saddle-point. The existence of such points for primal-dual problems of type (2.11) is stated in Lemma A.1.

The theoretical result of Theorem 2.7 is not directly applicable, but it provides a framework to propose a convergent algorithm studied in the next section. Indeed, in the case of the functional (2.1), the critical pseudo saddle-points are not isolated.

2.3.2. Application to the colorization problem. In our case, the hypotheses of Theorem 2.7 are not verified. The convergence of the sequence w^n to one of its fixed-points is

nevertheless demonstrated for the following model:

$$(2.20) \quad \min_w \min_u \max_p \langle Ku|p \rangle - F^*(p) + \frac{\lambda}{2} \int_{\Omega} \sum_{i=1}^C w_i \|u - c_i\|_2^2 + \chi_{\mathcal{E}}(w) + \chi_{\mathcal{R}}(u).$$

$\chi_{\mathcal{E}}$ is the indicator function of the canonical basis of \mathbb{R}^C . Its value is 0 if $w \in \mathcal{E}$, and $+\infty$ otherwise. This model differs from the original one (2.1) by the term $\chi_{\mathcal{E}}(w)$, which ensures that the cluster points of the sequence w^n produced by Algorithm 1 are isolated. In order to project a vector \tilde{w} onto \mathcal{E} , we have to compute a minimizer of $\|w - \tilde{w}\|_2^2 + \chi_{\mathcal{E}}(w)$. To that end, an index i corresponding of one of the minima of $(w_i - \tilde{w}_i)^2$ is first computed. The solution is then given by $(0, \dots, 1, 0, \dots, 0)$ with 1 in i th position. When there are at least two maxima, the algorithm chooses an arbitrary one among them.

The following lemma justifies the simplification of the model (2.1) into the model (2.20).

Lemma 2.8. *Assume that u^* is a uniform real-valued random variable over the set $[0, 255]^2$. Let us denote \mathcal{E} the canonical basis of \mathbb{R}^C .*

The set of minimizers of

$$(2.21) \quad \int_{\Omega} \sum_{i=1}^C w_i \|u^* - c_i\|_2^2 + \chi_{\Delta}(w)$$

is reduced to a point $w^(u^*)$ almost everywhere (a.e.).*

Moreover, the one of

$$(2.22) \quad \int_{\Omega} \sum_{i=1}^C w_i \|u^* - c_i\|_2^2 + \chi_{\mathcal{E}}(w)$$

*is reduced to a point $w^{**}(u^*)$ a.e. When these two minimizers are unique, $w^{**}(u^*) = w^*(u^*)$.*

Proof. u^* is in $[0, 255]^2$. We distinguish two cases.

Let us first consider the case when there exists i such that

$$(2.23) \quad \forall j \in \{1, \dots, C\} \text{ with } i \neq j, \|u^* - c_i\|_2^2 < \|u^* - c_j\|_2^2.$$

In this first case, the minimum of (2.21) and (2.22) is reached in $(0, \dots, 1, 0, \dots, 0)$ with 1 in i th position, and thus the minimizer of (2.21) is unique and is equal to the one of (2.22).

Let us now consider the event \mathcal{A} = “there exist at least two candidates c_i and c_j such that $\|u^* - c_i\|_2^2 = \|u^* - c_j\|_2^2 \leq \|u^* - c_k\|_2^2 \forall k \in \{1, \dots, C\}$.” In this case, neither minimizers of (2.21) nor the ones of (2.22) are necessarily unique. But, the set of such points u^* is the set of points in $[0, 255]^2$, which are at equal distance of two points of $[0, 255]^2$. It corresponds to the intersection of a line and $[0, 255]^2$ denoted by \mathcal{S} . Since \mathcal{S} is of measure 0, and the set of such \mathcal{S} is finite when the number of candidates is finite, \mathcal{A} is of measure 0. Thus, the event where neither minimizers of (2.21) nor the ones of (2.22) are necessarily unique is of probability 0. Finally, the minimizers of (2.21) and (2.22) are unique and equal a.e. ■

For our problem, the algorithm becomes Algorithm 2.

ALGORITHM 2. MINIMIZATION OF (2.20).

```

1: for  $n > 0$  do
2:    $p^{n+1} \leftarrow P_{\mathcal{B}}(p^n + \sigma \nabla \bar{u}^n)$ 
3:    $w^{n+1} \leftarrow P_{\mathcal{E}}(w^n - \rho \lambda (\|\bar{u}^n - c_i\|_2^2)_i)$ 
4:    $u^{n+1} \leftarrow P_{\mathcal{R}}\left(\frac{u^n + \tau (\operatorname{div}(p^{n+1}) + \lambda \sum_{i=1}^C w_i^{n+1} c_i)}{1 + \tau \lambda}\right)$ 
5:    $\bar{u}^{n+1} \leftarrow 2u^{n+1} - u^n$ 
6: end for

```

Parameters ρ , τ , and σ are the time steps. The operator div stands for the divergence. The algorithm requires the projection of the three estimated variables u , p , w . The projection $P_{\mathcal{R}}$ is necessary to ensure that the estimated image stays in the standard range of chrominance values \mathcal{R} . Precisely¹

$$(2.24) \quad P_{\mathcal{R}}(U, V) = (\max(\min(U, U_{\max}), U_{\min}), \max(\min(V, V_{\max}), V_{\min})).$$

Finally the projection of the dual variable $P_{\mathcal{B}}$ ensures the respect of the constraint $\chi_{\|(p_1, p_2, p_3)\|_2 \leq 1}$, by projecting p onto the L^2 unit ball. The following equation gives this projection for a pixel at position (l, k) :

$$(2.25) \quad P_{\mathcal{B}}(p^{l,k}) = \frac{\left(p_1^{l,k}, p_2^{l,k}, p_3^{l,k} - \sigma (\partial_x Y, \partial_y Y)^{l,k}\right)}{\max\left(1, \left\| \left(p_1^{l,k}, p_2^{l,k}, p_3^{l,k} - \sigma (\partial_x Y, \partial_y Y)^{l,k}\right) \right\|_2^2\right)}.$$

The projection $P_{\mathcal{E}}$ of a point onto a vector of the canonical basis is not well defined for all the points. For instance, the projection of $(1/2, 1/2, 0, \dots, 0)$ is not defined. We decide arbitrarily to retain the first closest vector of the canonical basis as the final projection when more than one vector is eligible.

$$(2.26) \quad P_{\mathcal{E}}(p_1, \dots, p_n) = \mathcal{E}_j \text{ with } j = \operatorname{argmin}_{i=1..n} p_i,$$

where \mathcal{E}_j denotes the j th vector of the canonical basis.

Notice that $\mathcal{U} = (\|u - c_i\|_2^2)_i$ is a vector of weights such that $\mathcal{U}_i = \|u - c_i\|_2^2$.

The following theorem ensures the convergence of (u^n, p^n, w^n) to a fixed-point with Algorithm 2.

Theorem 2.9. *Consider the application of Algorithm 2 to the problem (2.20).*

Let $H(w) = \chi_{\mathcal{E}}(w)$, $G(u) = \chi_{\mathcal{R}}(u)$, and $h(u, w) = \frac{\lambda}{2} \int_{\Omega} \sum_{i=1}^C w_i \|u - c_i\|_2^2$.

If the C candidates are all different (this is not a restricted hypothesis), the sequence (u^n, p^n, w^n) performed by Algorithm 2 converges to a critical pseudo saddle-point of the problem (2.20).

¹In (2.24), the values of U_{\max} , U_{\min} , V_{\max} , and V_{\min} are standard values equal, respectively, to 111.18, -111.18, 156.825, and -156.825.

Proof.

- (a) *Convergence of w^n .* Fixed-points of the algorithm are $(u^*, p^*, w^*) \in \mathcal{U} \times \mathcal{P} \times \mathcal{W}$ such that $\forall (u, p, w) \in \mathcal{U} \times \mathcal{P} \times \mathcal{W}$,

$$(2.27) \quad \begin{aligned} &\langle Ku|p^* \rangle - F^*(p^*) + G(u) + h(u^*, w) + H(w) + h(u, w^*) \\ &\geq \langle Ku^*|p \rangle - F^*(p) + G(u^*) + 2h(u^*, w^*) + H(w^*). \end{aligned}$$

We want to prove the convergence of the sequence w^n . To this end, we want to show that cluster points are isolated in the dimension of \mathcal{W} , i.e., we prove that points w^* are isolated such that $\exists (u^*, p^*) \in \mathcal{U} \times \mathcal{P}$ and (2.27) is verified.

Let us consider a larger set of points by particularizing $p = p^*$ and $u = u^*$. We will show that points (w^*) such that $\exists (u^*, p^*) \in \mathcal{U} \times \mathcal{P}$ and $\forall (u, w) \in \mathcal{U} \times \mathcal{W}$,

$$(2.28) \quad H(w) + h(u^*, w) \geq H(w^*) + h(u^*, w^*),$$

are isolated. We prove that minimizers of $\int_{\Omega} \sum_i w_i \|u^* - c_i\|_2^2 + \chi_{\mathcal{E}}(w)$ are isolated. These minimizers are on the canonical basis which constitutes a set of isolated points. Thus, points w^* verifying (2.27) are isolated. Finally, since w^n is on a finite set and $w^{n+1} - w^n \rightarrow 0$ when $n \rightarrow +\infty$, w^n converges.

- (b) *Convergence of (u^n, p^n, w^n) .* Since the weights w^n are in a discrete set and converge, they reach a fixed-point w^* from a certain rank. Algorithm 2 becomes the one of [6] applied to the saddle-point problem,

$$(2.29) \quad \min_u \max_p \langle Ku|p \rangle - F^*(p) + \frac{\lambda}{2} \int_{\Omega} \sum_{i=1}^C w_i^* \|u - c_i\|_2^2 + \chi_{\mathcal{R}}(u),$$

and converges to a fixed-point which is a critical pseudo saddle-point of the problem (2.20). ■

In practical experiments, in order to know if the convergence is reached, the number of changing weights between two iterations is computed. The convergence of the algorithm is performed when the number of changing weights is sufficiently small, e.g., less than 10 (see section 2.4).

2.3.3. Implementation details. We use the following definition for the discrete chrominance channels $u : \mathbb{R}^2 \rightarrow \mathbb{R}^2$:

$$(2.30) \quad u(k, l) = (U(k, l), V(k, l)) := (u_{k,l,1}, u_{k,l,2}).$$

The discrete gradient ∇ and divergence div operators are defined as in [3]. The following definitions stand for an image of size $N \times M$ with two channels.

Definition 2.10. *Let*

$$(2.31) \quad \nabla u_{k,l} = \begin{pmatrix} \begin{cases} u_{k+1,l,1} - u_{k,l,1} & \text{if } k < N, \\ 0 & \text{if } k = N, \end{cases} \\ \begin{cases} u_{k,l+1,1} - u_{k,l,1} & \text{if } l < M, \\ 0 & \text{if } l = M. \end{cases} \\ \vdots \\ \begin{cases} u_{k,l+1,2} - u_{k,l,2} & \text{if } l < M, \\ 0 & \text{if } l = M. \end{cases} \end{pmatrix}.$$

Next, we use the dual operator:

$$(2.32) \quad \operatorname{div}(p)_{k,l} = \begin{cases} p_{k,l}^1 - p_{k-1,l}^1 & \text{if } 1 < k < N, \\ p_{k,l}^1 & \text{if } k = 1, \\ -p_{k-1,l}^1 & \text{if } k = N \end{cases} + \cdots + \begin{cases} p_{k,l}^4 - p_{k,l-1}^4 & \text{if } 1 < l < M, \\ p_{k,l}^4 & \text{if } l = 1, \\ -p_{k,l-1}^4 & \text{if } l = M. \end{cases}$$

In the convex case, the behavior of primal dual algorithms has been studied in [6]. For one channel, the square of the operator norm of the divergence, neglecting borders effects, is equal to 8 (see [7, Remark, p. 92]). The convergence of this algorithm to a minimizer is verified for $\tau, \sigma > 0$ such that $\tau\sigma < 1/8$ (see [6, Theorem 1]). In the case of two color channels, neglecting borders effects, and simplifying to a square image,

$$(2.33) \quad \begin{aligned} \|\operatorname{div} p\|_2^2 &= \sum_{1 \leq k \leq N, 1 \leq l \leq M} (p_{k,l}^1 - p_{k-1,l}^1 + \cdots - p_{k,l-1}^4)^2 \\ &\leq 8 \sum_{1 \leq k \leq N, 1 \leq l \leq M} (p_{k,l}^1)^2 + \cdots + (p_{k,l-1}^4)^2 \\ &\leq 16\|p\|_2^2. \end{aligned}$$

Choosing $p_{k,l}^1 = p_{k,l}^2 = \cdots = p_{k,l}^4 = (-1)^{k+l}$, we obtain

$$(2.34) \quad \kappa = \|\operatorname{div}\| = \sup_{\|p\| \leq 1} \|\operatorname{div} p\|_2.$$

The norm of the divergence operator can be approximated with $\kappa = 16 - O(M.N)$. The value 16 is retained for the square of the norm of the operator div , equal to the gradient operator one (Theorem IV.5.6 in [27]). Finally, we set the time steps such that $16\tau\sigma < 1$ [7].

2.4. The choice of the parameter ρ . While Algorithm 2 converges to a critical point, the choice of the parameter ρ is not clear. Its influence on the algorithm is studied here, and we propose an asymptotic value.

We remark that in the hypotheses of Theorem 2.7, no bound is needed on the parameter ρ . A simplification of the algorithm is performed by taking this parameter as high as possible. If ρ is large enough, the problem

$$(2.35) \quad w^{n+1} = \operatorname{prox}_{\rho H + \rho h(\bar{u}^n, \cdot)}(w^n) = \operatorname{argmin}_w \|\tilde{w}^n - w\|_2^2 + \rho \left(\frac{\lambda}{2} \int_{\Omega} \sum_{i=1}^C w_i \|u - c_i\|_2^2 + \chi_{\mathcal{E}}(w) \right)$$

becomes

$$(2.36) \quad w^{n+1} = \operatorname{argmin}_w \int_{\Omega} \sum_{i=1}^C w_i \|u - c_i\|_2^2 + \chi_{\mathcal{E}}(w).$$

In the case of (2.36), the value of w^{n+1} does not depend on w^n and the final algorithm is reduced to Algorithm 3, where $C_{\bar{u}^n}$ stands for the closest candidate of \bar{u}^n in \mathbb{R}^2 .

ALGORITHM 3. MINIMIZATION OF (2.20).

- 1: $w = 1/C$ and $u^0 = \sum_{i=1}^C w_i c_i$
 - 2: $p^0 \leftarrow \nabla u$
 - 3: **for** $n \geq 0$ **do**
 - 4: $p^{n+1} \leftarrow P_{\mathcal{B}}(p^n + \sigma \nabla \bar{u}^n)$
 - 5: $u^{n+1} \leftarrow P_{\mathcal{R}}\left(\frac{u^n + \tau(\operatorname{div}(p^{n+1}) + \lambda C \bar{u}^n)}{1 + \tau\lambda}\right)$
 - 6: $\bar{u}^{n+1} \leftarrow 2u^{n+1} - u^n$
 - 7: **end for**
-

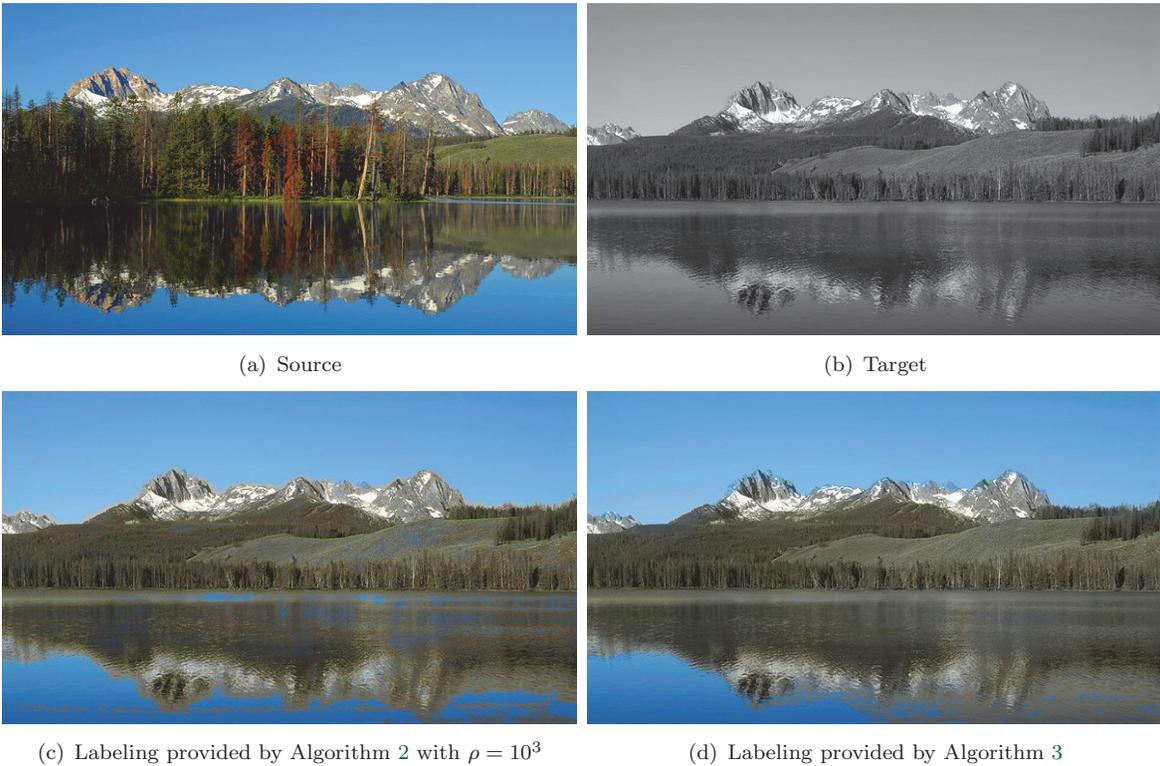


Figure 5. Comparison of the weights and the labeling ($\sum_{i=1}^C w_i c_i$) obtained with Algorithms 2 and 3.

The formulation of Algorithm 3 has the advantage to be compact and is easy to implement because it has one parameter less than Algorithm 2. We now propose to compare the impact of Algorithms 2 and 3 on the final labeling. All the results presented in this section have been performed with Figure 5(a) as the source image and Figure 5(b) as the target image.

We first provide a numerical study in order to show the influence of the parameter ρ and to demonstrate that Algorithm 3 is the most relevant choice for a regularized labeling. Figure 6 shows the number of changing weights between w^n and w^{n+1} for Algorithms 2 and 3. These

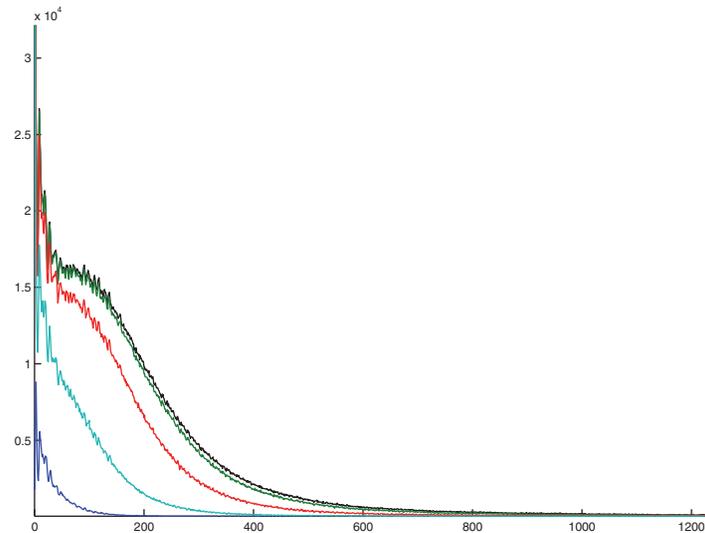


Figure 6. Number of weights changing during the iterations of Algorithm 2 with $\rho = 10^3$ (blue curve), $\rho = 10^4$ (cyan curve), $\rho = 10^5$ (red curve), $\rho = 10^6$ (green curve), and Algorithm 3 corresponding to $\rho = +\infty$ (black curve). At the convergence of the algorithms, the number of changes is very low.

curves are performed with images visible on Figure 5. Remark that, at each iteration, the weights are modified. In the case of $\rho = 10^3$ (blue curve) in Algorithm 2, the number of changing weights is low and becomes rapidly equal to 0. With $\rho = 10^6$ (green curve), the curve decreases more slowly. With Algorithm 3, the curve decreases slowly. The convergence of Algorithm 3 is therefore able to provide results more different from the initialization and has a higher dynamic. It means that the algorithm is able to change the label of a pixel in order to take into account the regularity of u^n . A high value of ρ is thus preferable. This value has no influence on the rate of decreasing of the global functional. Finally, after 5000 iterations, about 5 weights are modified per iteration. It represents only 0.0005% of the weights, confirming that w^n numerically converges.

We now propose to compare the quality of the results provided by Algorithm 2 with Algorithm 3. We analyze the difference between them in terms of labeling (obtained by $\sum_i w_i c_i$). To this end, we use $\rho = 10^3$ in Algorithm 2. Figure 5 shows the two labelings provided by the algorithms. We remark that Figure 5(c) has halos near strong contours and regularities on constant parts. We show that, contrary to Algorithm 3, Algorithm 2 is not able to fully take into account the regularization in the labeling.

Figure 7 shows the functional decreasing during the convergence of the algorithm. Here, the values of the functional are computed during the iterations. These values become asymptotically constant, which shows the numerical convergence of Algorithm 3.

3. Experimental results and discussions. In this section, we propose an experimentation of our method and a comparison with state-of-the-art methods. In all experiments, parameters will be $\sigma = 0.005$, $\tau = 5$, $\lambda = 0.005$, and $\gamma = 25$. The exemplar-based results are provided with eight candidates extracted with the method described in detail in [33].

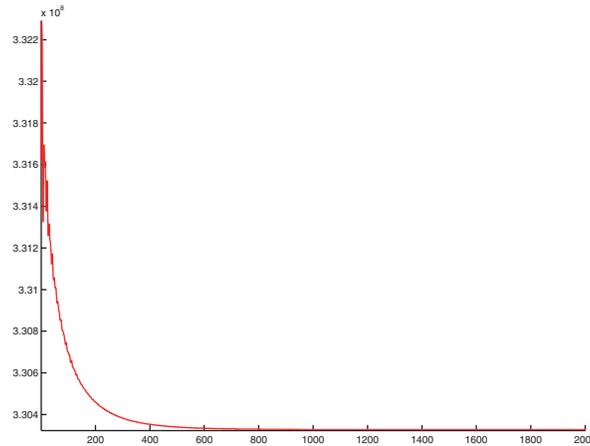


Figure 7. Value of the functional during the convergence of Algorithm 3.



Figure 8. Comparison of our method with the classical TV. Coupling the channel is useful in order to provide good results with the total variation.

3.1. Influence of the coupling on practical results. One of our contributions is the introduction of a novel TV regularization term. In Figure 8 we compare our model with a version without coupling. This result is provided by replacing $TV_{\mathcal{C}}$ by the classical total variation on chrominance channels, or by taking $\gamma = 0$ in our model. These results have been performed with 5(a) as the source image and 5(b) as the target image. We see that the result without coupling leads to a lot of halos. There are some melting of colors, and the image does not seem realistic.

3.2. Luminance-chrominance model vs. state-of-the-art exemplar-based methods. Figure 9 shows a comparison of exemplar-based colorization provided by the method of [20], [40], [4], and ourselves. The given source and target images are presented in the first and the second column. Figure 10 shows zooms on results. The results of Welsh, Ashikhmin, and Mueller [40] present an unrealistic color in the sky, due to texture features which are too simple to well differentiate all the parts of the image. In the second image, the methods of [20] and ourselves provide promising results. In the second image, the approach of [4] fails because the postpro-

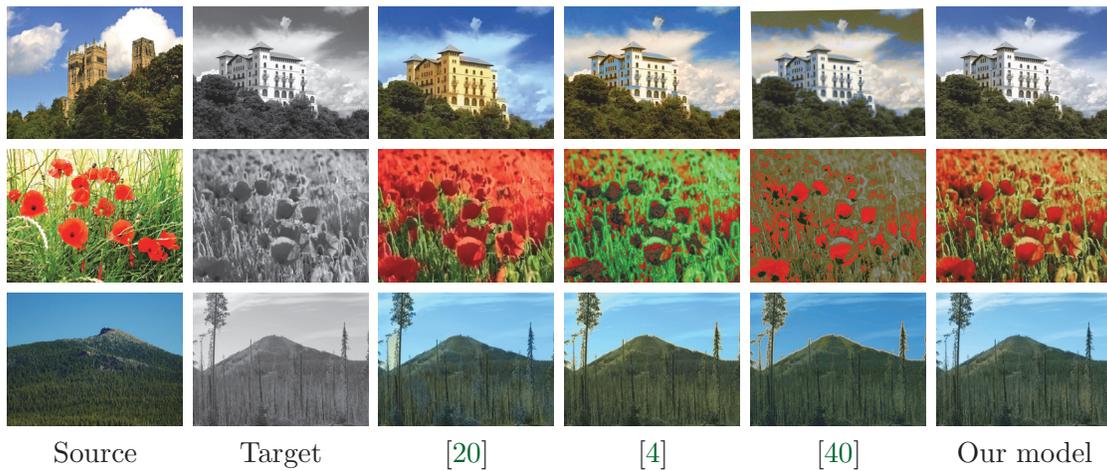


Figure 9. Comparison of our method with Gupta et al. [20], Bugeau, Ta, and Papadakis [4] and Welsh, Ashikhmin, and Mueller [40] for images with thin structures. On the left, the source image in color and the target image in gray-scale. Our results are comparable to [20], but our algorithm works without segmentation.



Figure 10. Zoom in on the third line of Figure 9.

cessing used in this method is not hue consistent. Due to the employed texture features, the method of [20] is not able to colorize the background. In the third image, the approaches of [40] and [4] present some problems near strong contours. With the coupled TV regularization, our method cannot produce halos near strong contours. Results of Gupta et al. [20] are of good quality, but their method fails on thin structures. This is due to the segmentation step. This last image is only well colorized by our approach. Finally, our method provides a colorization much faster than the one of [20], due to the lack of segmentation. It is also faster than the one of [4] that needs a projection onto the simplex, which is time consuming.

Figure 11 provides additional colorization results on different types of images. This figure clearly shows the good performance of our method to provide exemplar-based colorization.

4. User’s interactions. The proposed method, with (unoptimized) GPU implementation, can colorize an image of size 370×600 in approximately 1 second. This computation time allows us to propose an extension of our model by including user interactions and to provide a near real-time image colorization. This section presents how such interactions, represented by the user’s scribbles, can be directly introduced in our model.

The scribbles can either be given by the user before or added in an interactive or an iterative way. When a source is provided, the first step consists in extracting for each pixel a set of eight candidates [4] and the weights are initialized as $w = 1/C$ with C the number

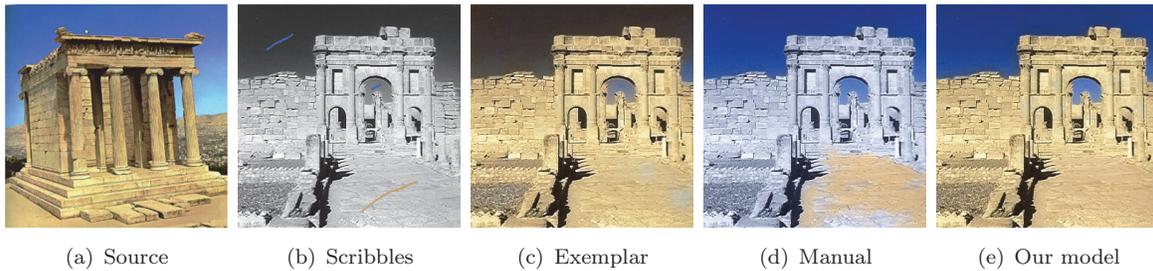


Figure 12. First example of image colorization with multiple priors. (a) Source image, (b) target image, (c) exemplar-based result, (d) scribbles-based result, (e) with both priors.

colorization result provided by our model. In this figure, the sky is not correctly colorized since it appears brown instead of blue as in the ruins’ main door. Moreover, blue colors appear on the floor. Figure 12(d) shows the corrections of the user where three scribbles are added in order to correct the first (exemplar-based) colorization result (Figure 12(c)). Figure 12(e) illustrates the advantage of the proposed extended image colorization since the user with less effort obtained the desired result. Finally, this result also highlights that our model is well adapted to preserve the color contours.

Figure 13 presents more results and illustrates the advantage of using our image colorization model as compared to only using a source (fourth column) or some scribbles (fifth column). Colorization results of the last column of Figure 13 are clearly better than the ones obtained with one prior only. This experiment also highlights that old photographs and faces are known to be hard to colorize as remarked, e.g., in [9]. Indeed, old photographs contain a lot of noise and the texture is usually degraded. Face images contain very smooth parts, e.g., the skin and the background is rarely suitable. Nevertheless, very promising results are obtained with our method. Moreover, the additional prior given by the scribbles of the user does not only have a local effect. Indeed, in the last result of Figure 13, the blue scribble needed in order to colorize the sky through the arch also improves the sky color at the bottom left of the image.

5. Conclusion. In this paper, a variational image colorization model is proposed. Our variational model includes a total variation term which couples luminance and chrominances channels. With this representation, the contours of the colorized image are well preserved. An algorithm is provided and its convergence is theoretically and numerically proven. Its implementation is fast enough to propose an interactive method in which the user can draw scribbles in order to provide colors not present in the source or to correct the results from the exemplar-based approach. We plan in the future to improve the results of the exemplar-based methods by studying new features to compare source and target patches. Next, we will explore quantitative metrics to measure the quality of the results. Finally, the extension to video colorization will be addressed.

Appendix A. The generic algorithm. In this paper, we give theoretical elements ensuring the convergence of our algorithm with some assumptions on the critical points.

Algorithm 1 is recalled in Algorithm 4. Algorithm 4 differs from the original one of [6]. First, the minimization problem is convex w.r.t. each variable, but it is not convex w.r.t. the couple (u, w) . Second, there is no relaxation on the variable w .

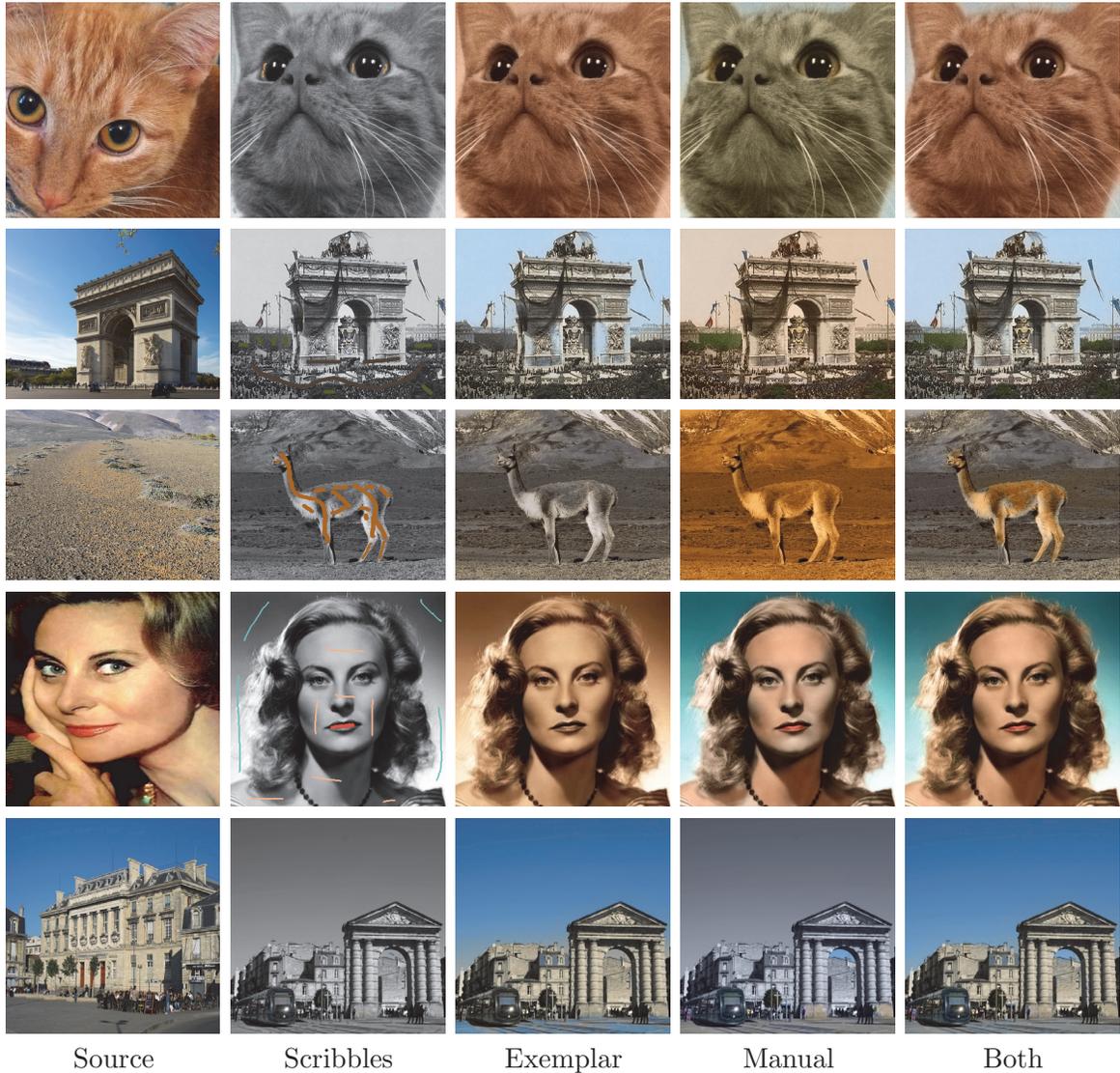


Figure 13. Advantage of the proposed extended image colorization as compared to only exemplar-based or scribble-based information. From left to right: the source, the target with the scribbles added by the user, the result with only the source, the result with only the scribbles, and the result with our approach.

This first result is a key ingredient to show the convergence of Algorithm 4.

Lemma A.1. *There exists $(\hat{u}, \hat{p}, \hat{w})$ verifying (2.17).*

Proof. $\forall w \in \mathcal{W}$ the problem (2.13) admits a saddle-point (\hat{u}, \hat{p}) . Thus, there exists (\hat{u}, \hat{p}) such that $\forall (u, p, w) \in \mathcal{U} \times \mathcal{P} \times \mathcal{W}$:

$$(A.1) \quad \langle Ku|\hat{p} \rangle - F^*(\hat{p}) + G(u) + h(u, w) \geq \langle K\hat{u}|p \rangle - F^*(p) + G(\hat{u}) + h(\hat{u}, w).$$

Equation (2.17) is obtained by summing (2.14) and (2.16) applied to \hat{u} . ■

ALGORITHM 4. PRIMAL-DUAL-LIKE ALGORITHM.

```

1: for  $n \geq 0$  do
2:    $p^{n+1} \leftarrow \text{prox}_{\sigma F^*}(p^n + \sigma K \bar{u}^n)$ 
3:    $w^{n+1} \leftarrow \text{prox}_{\rho H + \rho h(\bar{u}^n, \cdot)}(w^n)$ 
4:    $u^{n+1} \leftarrow \text{prox}_{\tau G + \tau h(\cdot, w^{n+1})}(u^n - \tau K^* p^{n+1})$ 
5:    $\bar{u}^{n+1} \leftarrow 2u^{n+1} - u^n$ 
6: end for

```

Theorem A.2. *Fixed-points of Algorithm 4 verify (2.18), i.e., fixed-points of the algorithm are critical pseudo saddle-points.*

Proof. Let us consider (u^*, p^*, w^*) a fixed-point of the algorithm. We recall a characterization of the proximal operator of a convex function f on a Hilbert space E :

$$(A.2) \quad r = \text{prox}_f(s) \Leftrightarrow \forall t \in E, \langle t - r | s - r \rangle + f(r) \leq f(t).$$

We write three inequalities from the three lines:

$$(A.3) \quad p^* = \text{prox}_{\sigma F^*}(p^* + \sigma K u^*).$$

With the characterization (A.2), we then obtain $\forall p \in \mathcal{P}$

$$(A.4) \quad \langle p - p^* | p^* + \sigma K u^* - p^* \rangle + \sigma F^*(p^*) \leq \sigma F^*(p).$$

Simplifying, we obtain

$$(A.5) \quad \langle p - p^* | K u^* \rangle + F^*(p^*) \leq F^*(p).$$

Assuming that

$$(A.6) \quad w^* = \text{prox}_{\rho H + \rho h(u^*, \cdot)}(w^*),$$

we obtain $\forall w \in \mathcal{W}$

$$(A.7) \quad H(w^*) + h(u^*, w^*) \leq H(w) + h(u^*, w).$$

This equation holds true even if w^* is not unique.

Recalling that, the third step of Algorithm 4 reads

$$(A.8) \quad u^* = \text{prox}_{\tau G + \tau h(\cdot, w^*)}(u^* - \tau K^* p^*),$$

and we have $\forall u \in \mathcal{U}$

$$(A.9) \quad -\langle u - u^* | K^* p^* \rangle + G(u^*) + h(u^*, w^*) \leq G(u) + h(u, w^*).$$

Summing now (A.5), (A.7), and (A.9) leads to $\forall (u, p, w) \in \mathcal{U} \times \mathcal{P} \times \mathcal{W}$

$$(A.10) \quad \begin{aligned} & \langle K u | p^* \rangle - F^*(p^*) + H(w) + G(u) + h(u^*, w) + h(u, w^*) \\ & \geq \langle K u^* | p \rangle - F^*(p) + H(w^*) + G(u^*) + h(u^*, w^*) + h(u^*, w^*). \end{aligned}$$

Fixed-points of the algorithm are thus critical pseudo saddle-points. ■

Appendix B. Proof of Theorem 2.7. We now give the proof of Theorem 2.7, inspired from the one of Theorem 1 of [6].

Notice that with Algorithm 4, all the following holds true even if $\text{prox}_{\rho H + \rho h(\bar{u}, \cdot)}$ is an arbitrary point of the set of minimizers of $\frac{\|w - w^n\|_2^2}{2\rho} + H(w) + h(\bar{u}^n, w)$.

Proof.

(a) *Bound of the sequence.* We write the three iterations in the general form:

$$(B.1) \quad \begin{aligned} p^{n+1} &= \text{prox}_{\sigma F^*}(p^n + \sigma K\bar{u}), \\ w^{n+1} &= \text{prox}_{\rho H + \rho h(\bar{u}, \cdot)}(w^n), \\ u^{n+1} &= \text{prox}_{\tau G + \tau h(\cdot, w^{n+1})}(u^n - \tau K^*\bar{p}), \end{aligned}$$

\bar{u} and \bar{p} denote a relaxation of the previous iterates. This relaxation will be different for the variables u and w . Due to the convexity of the functions, $\forall (u, p, w) \in \mathcal{U} \times \mathcal{P} \times \mathcal{W}$ we deduce the following three inequalities:

$$(B.2) \quad \left\{ \begin{aligned} F^*(p) &\geq F^*(p^{n+1}) + \left\langle \frac{p^n - p^{n+1}}{\sigma} |p - p^{n+1}\right\rangle + \langle K\bar{u} | p - p^{n+1} \rangle, \\ H(w) + h(w, \bar{u}) &\geq H(w^{n+1}) + h(w^{n+1}, \bar{u}) + \left\langle \frac{w^n - w^{n+1}}{\rho} |w - w^{n+1}\right\rangle, \\ h(w^{n+1}, u) + G(u) &\geq G(u^{n+1}) + h(w^{n+1}, u^{n+1}) \\ &\quad + \left\langle \frac{u^n - u^{n+1}}{\tau} |u - u^{n+1}\right\rangle + \langle K(u - u^{n+1}) | \bar{p} \rangle. \end{aligned} \right.$$

We detail the computation of the second inequality below. Since

$$(B.3) \quad w^{n+1} = \text{prox}_{\rho H + \rho h(\bar{u}, \cdot)}(w^n),$$

from the relation (A.2), we have $\forall w \in \mathcal{W}$

$$(B.4) \quad \langle w - w^{n+1} | w^n - w^{n+1} \rangle + \rho H(w^{n+1}) + \rho h(\bar{u}, w^{n+1}) \leq \rho H(w) + \rho h(\bar{u}, w).$$

Simplifying by $\rho > 0$, we obtain the second line of (B.2).

Summing the three inequalities (B.2) and using Cauchy–Schwarz inequality, we obtain

$$(B.5) \quad \begin{aligned} &\frac{\|p - p^n\|_2^2}{2\sigma} + \frac{\|u - u^n\|_2^2}{2\tau} + \frac{\|w - w^n\|_2^2}{2\rho} \\ &\geq [\langle K u^{n+1} | p \rangle - F^*(p) + G(u^{n+1}) + H(w^{n+1}) + h(\bar{u}, w^{n+1}) + h(u^{n+1}, w^{n+1})] \\ &\quad - [\langle K u | p^{n+1} \rangle - F^*(p^{n+1}) + G(u) + H(w) + h(u, w^{n+1}) + h(\bar{u}, w)] \\ &\quad + \frac{\|p - p^{n+1}\|_2^2}{2\sigma} + \frac{\|u - u^{n+1}\|_2^2}{2\tau} + \frac{\|w - w^{n+1}\|_2^2}{2\rho} \\ &\quad + \frac{\|p^n - p^{n+1}\|_2^2}{2\sigma} + \frac{\|u^n - u^{n+1}\|_2^2}{2\tau} + \frac{\|w^n - w^{n+1}\|_2^2}{2\rho} \\ &\quad + \langle K(u^{n+1} - \bar{u}) | p^{n+1} - p \rangle - \langle K(u^{n+1} - u) | p^{n+1} - \bar{p} \rangle. \end{aligned}$$

We choose $\bar{p} = p^{n+1}$ and $\bar{u} = 2u^n - u^{n-1}$. The last line of (B.5) becomes

$$\begin{aligned}
& \langle K(u^{n+1} - \bar{u})|p^{n+1} - p \rangle - \langle K(u^{n+1} - u)|p^{n+1} - \bar{p} \rangle \\
&= \langle K((u^{n+1} - u^n) - (u^n - u^{n-1}))|p^{n+1} - p \rangle \\
&= \langle K(u^{n+1} - u^n)|p^{n+1} - p \rangle - \langle K(u^n - u^{n-1})|p^n - p \rangle \\
&= \langle K(u^n - u^{n-1})|p^{n+1} - p^n \rangle \\
&\geq \langle K(u^{n+1} - u^n)|p^{n+1} - p \rangle - \langle K(u^n - u^{n-1})|p^n - p \rangle \\
&\quad - L\|u^n - u^{n-1}\|_2\|p^{n+1} - p^n\|_2,
\end{aligned}
\tag{B.6}$$

where $L = \|K\|$. As for $a, b \in \mathbb{R}$ and $\alpha > 0$, the inequality

$$2ab \leq \alpha a^2 + b^2/\alpha \tag{B.7}$$

holds, and we obtain

$$-L\|u^n - u^{n-1}\|_2\|p^{n+1} - p^n\|_2 \leq \frac{L\alpha\tau}{2\tau}\|u^n - u^{n-1}\|_2^2 + \frac{L\sigma}{2\alpha\sigma}\|u^n - u^{n-1}\|_2^2. \tag{B.8}$$

Choosing $\alpha = \sqrt{\sigma/\tau}$ and summing (B.5) and (B.8)

$$\begin{aligned}
& \frac{\|p - p^n\|_2^2}{2\sigma} + \frac{\|u - u^n\|_2^2}{2\tau} + \frac{\|w - w^n\|_2^2}{2\rho} \\
&\geq [\langle Ku^{n+1}|p \rangle - F^*(p) + G(u^{n+1}) + H(w^{n+1}) + h(\bar{u}, w^{n+1}) + h(u^{n+1}, w^{n+1})] \\
&\quad - [\langle Ku|p^{n+1} \rangle - F^*(p^{n+1}) + G(u) + H(w) + h(u, w^{n+1}) + h(\bar{u}, w)] \\
&\quad + \frac{\|p - p^{n+1}\|_2^2}{2\sigma} + \frac{\|u - u^{n+1}\|_2^2}{2\tau} + \frac{\|w - w^{n+1}\|_2^2}{2\rho} + (1 - \sqrt{\sigma\tau}L)\frac{\|p^n - p^{n+1}\|_2^2}{2\sigma} \\
&\quad + \frac{\|u^n - u^{n+1}\|_2^2}{2\tau} - \sqrt{\sigma\tau}\frac{\|u^n - u^{n-1}\|_2^2}{2\tau} + \frac{\|w^n - w^{n+1}\|_2^2}{2\rho} \\
&\quad + \langle K(u^{n+1} - u^n)|p^{n+1} - p \rangle - \langle K(u^n - u^{n+1})|p^n - p \rangle.
\end{aligned}
\tag{B.9}$$

Now, summing with n from 0 to $N-1$, it follows that $\forall u, p$, and w with the convention that $u^{-1} = u^0$ and $w^{-1} = w^0$

$$\begin{aligned}
& \sum_{n=1}^N \{[\langle Ku^{n+1}|p \rangle - F^*(p) + G(u^{n+1}) + H(w^{n+1}) + h(\bar{u}, w^{n+1}) + h(u^{n+1}, w^{n+1})] \\
&\quad - [\langle Ku|p^{n+1} \rangle - F^*(p^{n+1}) + G(u) + H(w) + h(u, w^{n+1}) + h(\bar{u}, w)]\} \\
&\quad + \frac{\|p - p^N\|_2^2}{2\sigma} + \frac{\|u - u^N\|_2^2}{2\tau} + \frac{\|w - w^N\|_2^2}{2\rho} \\
&\quad + (1 - \sqrt{\sigma\tau}L)\sum_{n=1}^{N-1}\frac{\|p^n - p^{n-1}\|_2^2}{2\sigma} + (1 - \sqrt{\sigma\tau}L)\sum_{n=1}^{N-1}\frac{\|u^n - u^{n-1}\|_2^2}{2\tau} \\
&\quad + \sum_{n=1}^{N-1}\frac{\|w^n - w^{n-1}\|_2^2}{2\rho} + \frac{\|u^N - u^{N-1}\|_2^2}{2\tau} + \frac{\|w^N - w^{N-1}\|_2^2}{2\tau} \\
&\leq \frac{\|p - p^0\|_2^2}{2\sigma} + \frac{\|u - u^0\|_2^2}{2\tau} + \frac{\|w - w^0\|_2^2}{2\rho}.
\end{aligned}
\tag{B.10}$$

From (B.7) and using Cauchy–Schwarz,

$$(B.11) \quad \langle Ku^N - u^{N-1} | p^N - p \rangle \leq \frac{\|u^N - u^{N-1}\|_2^2}{2\tau} + \tau\sigma L^2 \frac{\|p - p^N\|_2^2}{2\sigma},$$

and we write the inequality

$$(B.12) \quad \begin{aligned} & \sum_{n=1}^N \{ [\langle Ku^{n+1} | p \rangle - F^*(p) + G(u^{n+1}) + H(w^{n+1}) + h(\bar{u}, w^{n+1}) + h(u^{n+1}, w^{n+1})] \\ & \quad - [\langle Ku | p^{n+1} \rangle - F^*(p^{n+1}) + G(u) + H(w) + h(u, w^{n+1}) + h(\bar{u}, w)] \} \\ & + (1 - \tau\sigma L^2) \frac{\|p - p^N\|_2^2}{2\sigma} + \frac{\|u - u^N\|_2^2}{2\tau} + \frac{\|w - w^N\|_2^2}{2\rho} \\ & + (1 - \sqrt{\sigma\tau}L) \sum_{n=1}^{N-1} \frac{\|p^n - p^{n-1}\|_2^2}{2\sigma} + (1 - \sqrt{\sigma\tau}L) \sum_{n=1}^{N-1} \frac{\|u^n - u^{n-1}\|_2^2}{2\tau} \\ & + \sum_{n=1}^{N-1} \frac{\|w^n - w^{n-1}\|_2^2}{2\rho} + \frac{\|w^N - w^{N-1}\|_2^2}{2\rho} \\ & \leq \frac{\|p - p^0\|_2^2}{2\sigma} + \frac{\|u - u^0\|_2^2}{2\tau} + \frac{\|w - w^0\|_2^2}{2\rho}. \end{aligned}$$

Choosing (u, p, w) as a pseudo saddle-point of the problem and denoting it by $(\hat{u}, \hat{p}, \hat{w})$, we have

$$(B.13) \quad \begin{aligned} & \langle Ku^{n+1} | \hat{p} \rangle - F^*(\hat{p}) + G(u^{n+1}) + H(w^{n+1}) + h(\bar{u}, w^{n+1}) + h(u^{n+1}, w^{n+1}) \\ & \geq \langle K\hat{u} | p^{n+1} \rangle - F^*(p^{n+1}) + G(\hat{u}) + H(\hat{w}) + h(\hat{u}, w^{n+1}) + h(\bar{u}, \hat{w}). \end{aligned}$$

Thus, the two first lines of (B.12) are nonnegative and the first point of the theorem follows because $\tau\sigma L^2 < 1$.

- (b) *The existence of a cluster point which is a critical pseudo saddle-point.* Since \mathcal{U} , \mathcal{P} , and \mathcal{W} are of finite dimensions, the first point of the theorem means that the sequence (u^n, p^n, w^n) produced by the algorithm is bounded. Therefore, it admits a subsequence $(u^{n_k}, p^{n_k}, w^{n_k})$ which converges to (u^*, p^*, w^*) . The relation (B.12) (taking a pseudo saddle-point, using the positivity of the terms and observing the convergence of series) implies that

$$(B.14) \quad \lim_n (u^n - u^{n-1}) = \lim_n (p^n - p^{n-1}) = \lim_n (w^n - w^{n-1}) = 0.$$

Thus u^{n_k-1} , u^{n_k-1} , and u^{n_k-1} converge, respectively, to u^* , p^* , and w^* . It follows that (u^*, p^*, w^*) is a fixed-point of the algorithm and a critical pseudo saddle-point of the problem.

Remark that this last point proves the existence of a critical pseudo saddle-point of the problem.

- (c) *Convergence.* The cluster points are critical pseudo saddle-point thanks to Theorem A.2, and thus they are isolated. Hence, the sequence converges to a critical pseudo saddle-point according to Proposition C.1 and (B.14). ■

Appendix C. Useful proposition. The next proposition is useful to prove the convergence of a sequence in the general case, but it can be ignored for the particular case of Algorithm 2.

Proposition C.1. *Let $(u^n)_n$ be a sequence in a space of finite dimension, such that*

- u^n is bounded;
- cluster points of u^n are isolated;
- $\lim_n u^{n+1} - u^n = 0$.

Then the sequence $(u^n)_n$ converges.

Proof. By contradiction, let us assume that there are at least two cluster points. Since cluster points are isolated, $\exists A > 0$ such that if d and e are two cluster points, $\|d - e\| > 2A$.

Since the sequence of differences between two terms converges to 0, $\exists n_0$ such that $\forall n > n_0$, $\|u^{n+1} - u^n\|_2^2 \leq A/4$.

Let us denote as S the set of cluster points of u^n . We define

$$(C.1) \quad \mathcal{V} := \bigcup_{s \in S} B(s, M)$$

and consider $B(u^0, R)$ such that $(u^n)_n \in B(u^0, R)$ and $\mathcal{V} \subset B(u^0, R)$.

An infinite number of terms of the sequence u^n belong to $\mathcal{K} := B(u^0, R) \setminus \mathcal{V}$ because $\|u^{n+1} - u^n\|_2^2$ converges to 0. Indeed, if not, an infinite number of terms of the sequence are located at the interior of one of the balls $B(s, R)$. So, either the sequence converges, which proves the theorem, or there is a cluster point at a distance strictly less than A of another cluster point, which is impossible.

If we consider the sequence $(v^n)_n = (u^n)_n \cap \mathcal{K}$, it contains an infinite number of terms in a bounded set $B(u^0, R)$. Thus, it admits a cluster point which is a cluster point of u^n . However, as cluster points of u^n are in \mathcal{V} , this leads to a contradiction. Thus, the initial hypothesis is rejected and the sequence converges. ■

Acknowledgments. J.-F. Aujol is a member of Institut Universitaire de France. The authors would like to thank Raj Kumar Gupta for providing the images presented in [20].

REFERENCES

- [1] H. ATTOUCH, J. BOLTE, AND B. SVAITER, *Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward-backward splitting, and regularized Gauss–Seidel methods*, Math. Program., 137 (2013), pp. 91–129.
- [2] J. BOLTE, S. SABACH, AND M. TEBoulLE, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Math. Program., 146 (2014), pp. 459–494.
- [3] X. BRESSON AND T. F. CHAN, *Fast dual minimization of the vectorial total variation norm and applications to color image processing*, Inverse Probl. Imaging, 2 (2008), pp. 455–484.
- [4] A. BUGEAU, V.-T. TA, AND N. PAPADAKIS, *Variational exemplar-based image colorization*, IEEE Trans. Image Process., 23 (2014), pp. 298–307.
- [5] V. CASELLES, G. FACCILOLO, AND E. MEINHARDT, *Anisotropic Cheeger sets and applications*, SIAM J. Imaging Sci., 2 (2009), pp. 1211–1254.
- [6] A. CHAMBOLLE AND T. POCK, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145.
- [7] A. CHAMBOLLE, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vision, 20 (2004), pp. 89–97.

- [8] G. CHARPIAT, M. HOFMANN, AND B. SCHÖLKOPF, *Automatic image colorization via multimodal predictions*, in Proceedings of the European Conference on Computer Vision, D. Forsyth, P. Torr, and A. Zisserman, eds., Marseille, France, 2008, Springer, New York, pp. 126–139.
- [9] T. CHEN, Y. WANG, V. SCHILLINGS, AND C. MEINEL, *Grayscale image matting and colorization*, in Proceedings of the Asian Conference on Computer Vision, 2004, pp. 1164–1169.
- [10] Y. CHEN AND X. YE, *Projection onto a simplex*, preprint, arXiv:1101.6081, 2011.
- [11] A. Y.-S. CHIA, S. ZHUO, R. K. GUPTA, Y.-W. TAI, S.-Y. CHO, P. TAN, AND S. LIN, *Semantic colorization with internet images*, ACM Trans. Graphics, 30 (2011), p. 156.
- [12] P. L. COMBETTES AND V. R. WAJS, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul., 4 (2005), pp. 1168–1200.
- [13] M. CUI, J. HU, A. RAZDAN, AND P. WONKA, *Color-to-gray conversion using isomap*, The Visual Computer, 26 (2010), pp. 1349–1360.
- [14] G. DI BLASI AND D. REFORGIATO, *Fast colorization of gray images*, Eurographics Italian, (2003).
- [15] X. DING, Y. XU, L. DENG, AND X. YANG, *Colorization using quaternion algebra with automatic scribble generation*, in Advances in Multimedia Modeling, Lecture Notes in Comput. Sci. 713, 2012, pp. 103–114.
- [16] M. S. DREW AND G. D. FINLAYSON, *Improvement of colorization realism via the structure tensor*, Internat. J. Image Graphics, 11 (2011), pp. 589–609.
- [17] M. FORNASIER, *Nonlinear projection recovery in digital inpainting for color image restoration*, J. Math. Imaging Vision, 24 (2006), pp. 359–373.
- [18] R. C. GONZALES AND P. WINTZ, *Digital Image Processing*, 2nd ed., Addison-Wesley Longman, Boston, 1987.
- [19] J. GORSKI, F. PFEUFFER, AND K. KLAMROTH, *Biconvex sets and optimization with biconvex functions: A survey and extensions*, Math. Methods Oper. Res., 66 (2007), pp. 373–407.
- [20] R. KUMAR GUPTA, A. Y.-S. CHIA, D. RAJAN, E. S. NG, AND H. ZHIYONG, *Image colorization using similar images*, in Proceedings of the ACM International Conference on Multimedia, 2012, pp. 369–378.
- [21] A. HERTZMANN, C. E. JACOBS, N. OLIVER, B. CURLLESS, AND D. H. SALESIN, *Image analogies*, in Proceedings of the ACM Conference on Computer Graphics and Interactive Techniques, 2001, pp. 327–340.
- [22] J. HEU, D.-Y. HYUN, C.-S. KIM, AND S.-U. LEE, *Image and video colorization based on prioritized source propagation*, in Proceedings of the IEEE International Conference on Image Processing, 2009, pp. 465–468.
- [23] R. IRONY, D. COHEN-OR, AND D. LISCHINSKI, *Colorization by example*, in Proceedings of the Eurographics Conference on Rendering Techniques, Eurographics, 2005, pp. 201–210.
- [24] S. H. KANG AND R. MARCH, *Variational models for image colorization via chromaticity and brightness decomposition*, in IEEE Trans. Image Process., 16 (2007), pp. 2251–2261.
- [25] M. KAWULOK, J. KAWULOK, AND B. SMOLKA, *Discriminative textural features for image and video colorization*, IEICE Trans. Inform. Systems, 95-D (2012), pp. 1722–1730.
- [26] T. H. KIM, K. M. LEE, AND S. U. LEE, *Edge-preserving colorization using data-driven random walks with restart*, in Proceedings of the IEEE International Conference on Image Processing, 2010, pp. 1661–1664.
- [27] A. KOLMOGOROV AND S. FOMINE, *Éléments de la Théorie des Fonctions et de l'Analyse Fonctionnelle*, 2nd ed., Mir, Moscow, 1977.
- [28] G. R. KUHN, M. M. OLIVEIRA, AND L. A. F. FERNANDES, *An improved contrast enhancing approach for color-to-grayscale mappings*, The Visual Computer, 24 (2008), pp. 505–514.
- [29] P. LAGODZINSKI AND B. SMOLKA, *Digital image colorization based on probabilistic distance transformation*, in Proceedings of the 50th International Symposium ELMAR, Vol. 2, 2008, pp. 495–498.
- [30] A. LEVIN, D. LISCHINSKI, AND Y. WEISS, *Colorization using optimization*, ACM Trans. Graphics, 23 (2004), pp. 689–694.
- [31] O. LEZORAY, V.-T. TA, AND A. ELMOATAZ, *Nonlocal graph regularization for image colorization*, in Proceedings of the 19th International IEEE Conference on Pattern Recognition, 2008, p. 1–4.
- [32] F. PIERRE, J.-F. AUJOL, A. BUGEAU, N. PAPADAKIS, AND V.-T. TA, *Exemplar-based colorization in RGB color space*, in Proceedings of the IEEE International Conference on Image Processing, 2014; also available online from <https://hal.archives-ouvertes.fr/hal-01019627/document>.

- [33] F. PIERRE, J.-F. AUJOL, A. BUGEAU, AND V.-T. TA, *Hue Constrained Image Colorization in the RGB Space*, preprint, 2014; also available online from <https://hal.archives-ouvertes.fr/hal-00995724/document>.
- [34] F. PIERRE, J.-F. AUJOL, A. BUGEAU, AND V.-T. TA, *A unified model for image colorization*, in Proceedings of the Color and Photometry in Computer Vision (ECCV Workshop), 2014, pp. 1–12; also available online from <https://hal.archives-ouvertes.fr/hal-01038055/document>.
- [35] M. H. QUANG, S. H. KANG, AND T. M. LE, *Image and video colorization using vector-valued reproducing kernel Hilbert spaces*, *J. Math. Imaging Vision*, 37 (2010), pp. 49–65.
- [36] X. REN AND J. MALIK, *Learning a classification model for segmentation*, in Proceedings of the IEEE International Conference on Computer Vision, 2003, pp. 10–17.
- [37] M. SONG, D. TAO, C. CHEN, J. BU, AND Y. YANG, *Color-to-gray based on chance of happening preservation*, *Neurocomputing*, 119 (2013), pp. 222–231.
- [38] D. SÝKORA, J. BURIÁNEK, AND J. ŽÁRA, *Unsupervised colorization of black-and-white cartoons*, in Proceedings of the International Symposium on Non-photorealistic Animation and Rendering, ACM, 2004, pp. 121–127.
- [39] L.-Y. WEI AND M. LEVOY, *Fast texture synthesis using tree-structured vector quantization*, in Proceedings of the ACM Computer Graphics and Interactive Techniques, ACM Press/Addison-Wesley, New York, 2000, pp. 479–488.
- [40] T. WELSH, M. ASHIKHMIN, AND K. MUELLER, *Transferring color to greyscale images*, *ACM Trans. Graphics*, 21 (2002), pp. 277–280.
- [41] P. A. MARKOWICH W. BAATZ, M. FORNASIER, AND C.-B. SCHÖNLIEB, *Inpainting of ancient Austrian frescoes*, in Bridges Leeuwarden: Mathematics, Music, Art, Architecture, Culture, R. Sarhangi and C. H. Séquin, eds., Tarquin Publications, London, 2008, pp. 163–170; also available online from <http://archive.bridgesmathart.org/2008/bridges2008-163.html>.
- [42] L. YATZIV AND G. SAPIRO, *Fast image and video colorization using chrominance blending*, *IEEE Trans. Image Process.*, 15 (2006), pp. 1120–1129.
- [43] Y. ZHENG AND E. A. ESSOCK, *A local-coloring method for night-vision colorization utilizing image analysis and fusion*, *Information Fusion*, 9 (2008), pp. 186–199.