Variational Exemplar-Based Image Colorization

Aurélie Bugeau, Vinh-Thong Ta, Nicolas Papadakis

Variational Exemplar-based Image Colorization

Aurélie Bugeau, Vinh-Thong Ta, and Nicolas Papadakis

***Abstract*—In this paper, we address the problem of recovering a color image from a grayscale one. The input color data comes from a source image considered as a reference image. Reconstructing the missing color of a grayscale pixel is here viewed as the problem of automatically selecting the best color among a set of colors candidates while simultaneously ensuring the local spatial coherency of the reconstructed color information. To solve this problem, we propose a variational approach where a specific energy is designed to model the color selection and the spatial constraint problems simultaneously. The contributions of this paper are twofold: first, we introduce a variational formulation modeling the color selection problem under spatial constraints and propose a minimization scheme which allows computing a local minima of the defined non-convex energy. Second, we combine different patch-based features and distances in order to construct a consistent set of possible color candidates. This set is used as input data and our energy minimization allows to automatically select the best color to transfer for each pixel of the grayscale image. Finally, experiments illustrate the potentiality of our simple methodology and show that our results are very competitive with respect to the state-of-the-art methods.**

I. Introduction

IMAGE colorization consists in recovering a color image from a grayscale one with initial color input data. This application attracts a lot of attention in the image-editing community in order to restore or to colorize old grayscale movies or pictures. A first category of colorization algorithms requires a huge amount of user intervention to manually add initial colors to the grayscale image. The colorization process is then performed by propagating the input color data to the whole image. The color propagation is done by using methods such as quadratic optimization, diffusion or variational techniques (see for instance [1]–[7] and references therein). The main drawback of these methods is that the initialization of the colorization process is done completely manually.

Another colorization approach consists in transferring color from one (or many) initial color image considered as example. The color prediction is performed from the luminance (grayscale) channel of both images and the final color is computed with methods such as image statistics or machine learning approaches. For instance, in [8] (and later [9] that uses multiple images as references) the color transfer is performed by minimizing a distance based on simple statistics on luminance images. In [10], [11] and [12], more sophisticated methods are proposed based on numerous and complex steps such as image segmentation, classification, locally weighted histogram regression, machine learning techniques, color refinement and image descriptors to capture textures or complex structures. One drawback of such approaches for image colorization is the spatial coherency during the color transfer which can lead to possible inconsistent colorization in the final result. To overcome this limitation, [8] adds user interaction in the colorization process and [11] uses a parameter to enforce the local spatial relationship between pixels.

In this paper, we focus on this second category of methodology based on reference images. We propose a simple method with very few steps which simultaneously performs the color transfer while enforcing the spatial coherency of the reconstructed colors. The experiments presented in this paper (Section IV) show that our simple methodology can be competitive with the aforementioned more complex approaches.

Exemplar-based Methods: Most of exemplar-based colorization techniques make the assumption that pixels with similar intensities or similar neighborhood should have similar colors. Such approaches are also called exemplar-based (or patch-based) methods. They have recently received a lot of attention in the image processing community. Indeed, first introduced for texture synthesis [13], this concept has been derived for many applications such as image denoising or image inpainting where the corrupted or the missing data are restored from the values of the other pixels contained in the image. In the case of image denoising, the so-called NL-means filter [14], which corresponds to a weighted averaging filter [15], uses weights proportional to the similarity measure between the different patches of the image to restore. For image inpainting, the same idea is used. For a given missing pixel, the best corresponding pixel (or region) to copy/paste during the inpainting process is found by using exemplar-based similarities (see for instance in [13], [16] and references therein).

The main underlying idea behind all these exemplar-based approaches is that the best *a priori* for an image is the image itself.

For all the aforementioned applications, the common problem can be stated as follows. Given a point defined in a domain, how to select the best match(es) from a set of unordered examples (defined in a d-dimensional vector space) in order to obtain the best reconstruction of the missing data ? In the case of non-local image denoising, it consists in computing the weights between the different candidates in order to realize a weighted average. For image inpainting or image colorization, this problem corresponds to choosing a single candidate to realize the copy/paste operation.

One drawback of such approaches is that discontinuities or spatial incoherence in the reconstructed image may be visible

A. Bugeau is with University of Bordeaux, LaBRI, UMR 5800, F-33400 Talence, France. email: aurelie.bugeau@labri.fr. phone: +33(0) 5 4000 6900 fax: +33(0) 5 4000 6669

V.-T. Ta is with IPB, LaBRI, UMR 5800, F-33400 Talence, France email: vinh-thong.ta@labri.fr. phone: +33(0) 5 4000 6900 fax: +33(0) 5 4000 6669

N. Papadakis is with CNRS, IMB, UMR 5251, F-33400 Talence, France email: nicolas.papadakis@math.u-bordeaux1.fr phone: +33(0) 5 4000 2116 fax: +33(0) 5 4000 2123

since most of the algorithms are of greedy type and each pixel is processed independently. In the case of image inpainting, several methods have been proposed to overcome this issue by optimizing an energy defined over the whole missing area [17]–[20]. To the best of our knowledge, such strategy has not yet been studied in the specific context of image colorization.

Another particularity of exemplar-based methods is that they almost always use the SSD measure (sum of square differences) in order to compute a pixel-wise distance between the patches to compare. Nevertheless, natural images are composed of different textures and image structures which are not well captured by the SSD distance. To take into account the different type of information contained in natural images, it seems relevant to use and combine different patch-based features and their associated metrics.

Previous Works and Main Contributions: In a previous work [21], we have proposed an image colorization framework which combines different features and distances in order to select a set of different color candidates for each pixel of the grayscale image. For each pixel of the grayscale target image T, a set of possible colors is found in the color source image S. This set is constructed by comparing the luminance patches of both images according to different features and distances. One of the candidates is then selected by choosing the median value on the principal axis given by a dimensional reduction (PCA) on the different color candidates A color regularization step is finally performed to smooth this arbitrary point-wise choice and ensure the spatial coherency of the reconstruction.

This paper is an extension of the method proposed in [21] but relies on a framework that is fundamentally different. We address the exemplar-based image colorization problem under a variational point of view.

In this paper, contrary to [21], the candidate selection and color regularization problems are simultaneously solved through a variational energy minimization. Our new approach then avoids the arbitrary pre-selection of a single candidate. To that end, we design an energy whose minimizers automatically select the best color for each pixel from a set of candidates, while ensuring the spatial coherency of the reconstruction, as it enforces neighbor pixels to have similar colors in the final result.

Paper Organization: Section II reviews the general image colorization framework introduced in [21]. Section III presents a variational model for the automatic candidate selection and the related optimization algorithm. Experimental results and energy analysis are shown in Section IV and, finally, conclusions are given in Section V.

II. General Framework for Exemplar-based Image Colorization

Most of exemplar-based methods for image colorization proposed in the literature can be summarized into a general pipeline. We now detail this pipeline in order to introduce the proposed variational exemplar-based image colorization method. In the following S and T will respectively denote the color source image and the grayscale target image to colorize.

A. General Image Colorization Pipeline

Exemplar-based image colorization methods can be decomposed into three main steps.

1) Step–1, Pre-Process S and T: This step consists in converting both S and T to a luminance-chrominance color space such as $l\alpha\beta$ [8], Lab [11] or $YCbCr$ [2].

As mentioned in the introduction, exemplar-based image colorization methods rely on a copy/paste of the colors from the source image to the target image. More precisely, the chrominance channels will be transferred from one to the other. In order to select the color to transfer, the luminance of S and T should be comparable. Ideally, the transfer process should take into account the global difference in luminance of the two images, *i.e.*, these two images should have similar global statistics such as the mean and the variance. For this purpose, a luminance mapping can be performed [22].

2) Step–2, Predict Color from S and Transfer to T: In order to have acceptable computation time in this second step, only a reduced set $S_n = \{q_1, \dots, q_n\}$ of n possible candidate pixels are extracted from the source image S with $n \ll \#S$, where $\#S$ corresponds to the cardinality of the set S.

The second step aims at choosing for each pixel $p \in T$ the best pixel $\hat{q} \in S_n$ which minimizes a distance ρ on the luminance neighborhoods

$$\hat{q} = \arg\min_{q \in S_n} \rho(p, q) \ .$$

Finally, the chrominance of $\hat{q}$ is transferred to p, so that T becomes a final true luminance-chrominance image.

3) Step–3, Post-Process T: This last step generally consists in a simple color space re-conversion, *i.e.*, from the luminance-chrominance to the RGB color space. A color refinement step may also be realized in order to enhance the final visual quality of the result [11].

B. Proposed Image Colorization Pipeline

We now describe how each step of the pipeline is achieved in our image colorization approach.

1) Step–1: As done in [21], we propose to convert S and T to the convex YUV color space to obtain the luminance-chrominance decomposition $S = (Y_S, U_S, V_S)$ and $T = (Y_T, U_T, V_T)$. This empirical choice comes from our experiments. Indeed, by using non standard and non convex color spaces, such as the $l\alpha\beta$ [8], our experiments have shown that such spaces can lead to numerical issues. More precisely, ass the projection on non convex sets is not unique, variational methods can finally provide visually inconsistent colorization results.

Finally, similarly to [8], a luminance mapping [22] is performed. The luminance Y_S of the source image is modified in order to map the one of the target image Y_T so that the two luminance images become comparable.

2) Step–2: As described previously, only $n \ll \#S$ samples are randomly selected from S on a jittered grid leading to the reduced set of candidate pixels $S_n = \{q_1, \dots, q_n\}$. From now, only these n pixels of S will be used to colorize the target image T.

Our color prediction model consists in selecting the best chrominances to transfer to each pixel $p \in T$. Contrary to previous approaches in image colorization, we do not only predict one unique chrominance per pixel. Instead, several possible chrominances are kept according to different image features and distances. Natural images contain different types of complex structures, redundancies and textures. In order to perform colorization on natural images, every algorithm should integrate such information. By using several different features and distances, our process will adapt itself to the different properties of the considered image.

While many metrics to compare patches exist, we here only concentrate on $L = 3$ features f_l and their associated distances ρ_l with $l = \{1, 2, 3\}$. Each of these three features is computed for different size σ of patches. For each feature among all the possible sizes, the set of possible patch sizes is denoted as Σ_l (with $\sigma \in \Sigma_l$). The choice of the metrics and patch descriptors has been made experimentally, but any other feature could be integrated straightforward.

Each chosen feature captures a different characteristic of the grayscale image. Given a luminance image Y, $Y(p) = Y(x, y)$ is the luminance value of a pixel $p \in Y$. We denote a patch of size σ^2 centered at pixel $p = (x, y)$ as

$$P_\sigma(p) = \{Y(x + k, y + l)\} \quad \forall k, l = -\sigma/2, \ldots, \sigma/2$$

For any pixel $p \in Y$ and for different size σ of patches, we compute:

i) the variance

$$f_1(p, \sigma) = \nu(P_\sigma(p))$$

which differentiates constant and textured areas;

ii) the amplitude spectrum of the $2D$ discrete Fourier transform

$$f_2(p, \sigma, \xi) = \| \mathrm{DFT}(P_\sigma(p), \xi) \|_2$$

where $\xi = (\xi_1, \xi_2)$ states as the frequency. This feature gives information on the high frequencies of the most important structures;

iii) the cumulative histograms containing H bins with $\{h_1, \ldots, h_H\}$ and $\forall i = 1, \ldots, H$

$$f_3(p, \sigma, h_i) = \sum_{j=1}^{i} \sum_{q \in P_\sigma(p)} \frac{\delta(Y(q), h_j)}{\#P_\sigma(p)}$$

where $\delta(a, h_j) = 1$ if the color a belongs to the j-th bin h_j, *i.e.*, $a \in h_j$ and $\delta(a, h_j) = 0$ otherwise. This cumulative histogram allows to take into account the luminance distribution inside the patch.

From the features space, we define the associated distances ρ_l with $l = \{1, 2, 3\}$ such that for two pixels $p \in T$ and $q \in S_n$, we have:

$$\begin{aligned}
\rho_1(p, q, \sigma) &= |f_1(p, \sigma) - f_1(q, \sigma)| \,, \\
\rho_2(p, q, \sigma) &= \sum_{\xi} |f_2(p, \sigma, \xi) - f_2(q, \sigma, \xi)| \text{ and} \\
\rho_3(p, q, \sigma) &= \sum_{h=1}^{H} |f_3(p, \sigma, h) - f_3(q, \sigma, h)|.
\end{aligned}$$

As we are comparing 1-dimensional grayscale images, we can notice that the ℓ_1 distance ρ_3 on the cumulative histograms is equivalent to the ℓ_1 Wasserstein distance between the distributions.

From the introduced features and metrics, we can define the set of chrominances for each pixel $p \in T$ from which the final chrominance will be chosen. For each point $p \in T$, each feature $l = 1, \ldots, L$ and each size $\sigma \in \Sigma_l$, we select the candidate $q_{l,\sigma}$ such that:

$$q_{l,\sigma} = \underset{q \in S_n}{\arg\min}\, \rho_l(p, q, \sigma).$$

Therefore, by finding the best source point in S_n for all the L features and distances, we obtain the chrominance set $C(p)$ for each $p \in T$. As we only intend to colorize the chrominance channels U an V of the YUV image, the chrominance set is composed of $d = 2$-dimensional elements:

$$C(p) = \{(U(q_{l,\sigma}), V(q_{l,\sigma})) \,, \forall l = 1, \ldots, L \text{ and } \forall \sigma \in \Sigma_l\} \,.$$

For the sake of clarity, we rewrite this set as:

$$C(p) = \{c_i(p) = (U(q_i), V(q_i)) \,, \forall i = 1, \ldots, N\}$$

where $N = \sum_{l=1}^{L} \#\Sigma_l$. We now have a reduced set of possible chrominance for each pixel p of the grayscale image.

The best candidate for each point could be taken as the one having the smallest distance ρ_l. Unfortunately, as the distances are of different nature, they are not directly comparable. Each feature will allow one to measure different image characteristics and we have no prior on which feature should be used in a particular area. In [21], the problem of selecting the best color to transfer among the set of possible colors is solved, for each pixel independently, using a dimensional reduction (PCA) of the chrominance candidates $C(p)$. For each pixel, the median value on the main axis of the PCA is considered as the single color candidate. In this paper, in order to avoid such an arbitrary selection, we propose to use the spatial coherency to automatically chose the best candidate within a variational framework that will be presented in Section III.

3) Step–3: The post-processing step mainly aims at re-converting the luminance-chrominance image to the RGB color space. The generated color images are generally drab. In order to improve the shininess of the colored images, we here propose to first apply a mapping [22] of the chrominances from U_T to U_S and V_T to V_S before converting T to RGB space: the chrominance histograms are linearly shifted and scaled to fit the histograms of S. Obviously this solution is not ideal as there is no reasons for these histograms to be completely similar at the end of the process but it gave us satisfactory experimental results.

III. Variational Exemplar Selection Framework

In this section, we describe a variational method for solving the general following problem. Given a point $p \in T$, the problem is to select the best match from a set of unordered examples $C(p)$, defined in a d-dimensional vector space, in order to obtain a regular reconstruction of the missing information. In other words, we want to estimate $u : p \to u(p) \in C(p)$

while imposing spatial consistency between $u(p)$ and $u(q)$, with $q \in \mathcal{N}(p)$ where $\mathcal{N}(p)$ is the local neighborhood of p.

In the following, we define the variational model and describe how to integrate the spatial constraints within an energy functional. We next present the associated numerical scheme used to minimize the proposed functional.

A. General Variational Model

To address the problem of finding the best candidate for a pixel $p \in T$ while adding some spatial consistency, the strategy of [21] consists in first selecting the best candidate (in an arbitrary sense) among the set $C(p)$. Next, during the post-processing step, the solution is regularized to enforce the spatial constraints. Such approach highly depends on the chosen candidate and can induce blurry results. In this work, we propose to reformulate the problem into a single process, *i.e.*, without a first candidate selection step, the selection being induced by the regularization.

1) Data Modeling: In order to introduce the proposed data model, let us now detail what occurs for $d = 1$-dimensional candidates. It is well known that for an ordered set $A = \{a_1, \ldots, a_N\}$ with $a_i \in \mathbb{R}$ and $a_i \leq a_{i+1}$, we have:

$$\text{median}(A) = a_{[N/2]} = \min_{x \in \mathbb{R}} \sum_{i=1}^{N} ||x - a_i||_1 = \min_{x \in \mathbb{R}} \sum_{i=1}^{N} |x - a_i|,$$

where $[.]$ computes the integer part and $||.||_1$ stands for the ℓ_1 norm. Therefore selecting, for each point p, the median of the set $C(p)$ of size N is equivalent to minimizing the following energy:

$$E_1(u) = \sum_{p \in T} \sum_{i=1}^{N} |u(p) - c_i(p)| \tag{1}$$

where $c_i(p)$ is the i-th candidate of set $C(p)$ and u is the 1-dimensional vector to reconstruct. Notice that the number of candidates N could be different for each point p.

A related model has been considered in [23] for depth map fusion involving ordered depth candidates of dimension $d = 1$. Unfortunately, when working with $d >$ 1-dimensional data, the ℓ_1 norm separates the distance between dimensions and the median value does not correspond anymore to one of the candidates. This behavior is a main issue, as we would like to select a single element of the $d = 2$-dimensional candidates $C(p)$ for each point $p \in T$.

Instead of relying on the median operator through the ℓ_1 norm, we rather consider the ℓ_2 norm to have a metric that compares the d-dimensions conjointly. In order to have a smooth derivable data term, we consider the ℓ_2^2 norm. Our problem of selecting one single candidate per point can be re-casted as follows with $i = 1, \ldots, N$:

$$E_2(u) = \sum_{T} \min_i \|u - c_i\|_2^2 = \sum_{T} \min_i \sum_{k=1}^{d} \|u^k - c_i^k\|_2^2. \tag{2}$$

In the right side of the equation, u^k and c_i^k stand for the k^{th} dimension of the data. In the image colorization problem, $c_i^0(p) = U(q_i)$ and $c_i^1(p) = V(q_i)$.

2) Data Term Relaxation: In order to minimize the data term (2), we now consider a relaxation by increasing the state dimension with a new variable $w = \{w_i(p), i = 1, \ldots, N\}$ which represents the probability of choosing the i-th candidate for each point $p \in T$, such that

$$\sum_{T} \min_i \|u - c_i\|_2^2 = \min_{w \in \overline{\mathcal{W}}} \sum_{T} \sum_{i} w_i \|u - c_i\|_2^2. \tag{3}$$

where w has binary values and is defined in the set

$$\overline{\mathcal{W}} = \left\{ w(p) = \{w_i(p)\}_{i=1}^{N},\ w_i : T \to \{0; 1\},\ \text{s.t.} \sum_{i=1}^{N} w_i(p) = 1 \right\}.$$

As $\overline{\mathcal{W}}$ is not a convex set, we consider a relaxation and now define $w = (w_1, \ldots, w_N)$ in the convex set, such as

$$\mathcal{W} = \left\{ w(p) = \{w_i(p)\}_{i=1}^{N},\ w_i : T \to [0; 1],\ \text{s.t.} \sum_{i=1}^{N} w_i(p) = 1 \right\}.$$

Nevertheless, since we can not ensure that the minima of our data model will be reached for binaries values, we add an additional non convex term that enforces the binary nature of w:

$$E_3(u, w) = \sum_{T} \sum_{i=1}^{N} w_i \|u - c_i\|_2^2 + \frac{\alpha}{2} \sum_{T} \sum_{i=1}^{N} w_i (1 - w_i) \tag{4}$$

with $\alpha > 0$. This additional term will enforce the binary conditions as it has two minima in 0 and 1 for w with values in the interval $[0; 1]$.

3) Spatial Constraints and TV Regularization: The energy (4) has trivial minima in its current form. In this paper, we focus on the image colorization problem which involves consistent and spatially coherent image results. To model this behavior, a Total Variation (TV) based regularization of the unknown variable u is introduced in the energy to smooth the solution while keeping sharp discontinuities when required by the data. The final variational model for automatic candidate selection is given by

$$\begin{aligned} E(u, w) = \sum_{T} \|Gu\|_1 + \frac{\lambda}{2} \sum_{T} \sum_{i=1}^{N} w_i \|u - c_i\|_2^2 \\ + \frac{\alpha}{2} \sum_{T} \sum_{i=1}^{N} w_i (1 - w_i) \end{aligned} \tag{5}$$

where G is the discrete gradient operator defined w.r.t the chosen neighborhood $\mathcal{N}(p)$ with $p \in T$. The parameters λ and α weight the influence of each energy term.

4) Links with the Approach of Arias et al.*:* Our problem as strong connections with the method of [17] dedicated to patch selection for image inpainting. Using our notations, the energy they propose reads:

$$\sum_{p \in T} \sum_{i=1}^{N} w_i \|u(p) - c_i(p)\|_{2,P}^2 - \alpha \sum_{T} \sum_{i=1}^{N} w_i \log(w_i). \tag{6}$$

The first term involves a spatial consistency since $\|.\|_{2,P}$ states as the ℓ_2 norm defined on a patch P centered on the pixel $p \in$

T. The second term allows to maximize the weight entropy and enforces the weights to be binary.

One can note that, in the model (6), the candidates $c_i(p)$ for a given pixel $p \in T$ are chosen within the target image T itself on the contrary of our approach where another image is taken as reference.

The interesting aspect of the model (6) is that explicit expressions can be obtained to alternatively update u and w (recovering the NL-means weights for w [14]). However, in our current problem involving a TV-based regularization of u, we no longer have an explicit expression for u when computing the associated Euler-Lagrange equations.

We will therefore have to rely on gradient descent based minimization schemes. In the same way, as we would like to recover binary weights, the explicit update of w seems unadapted to our problem. Hence, we will also consider a gradient descent strategy on w.

In this context, the weight entropy model leads to numerical issues since the derivation of this term gives $1+\log(w_i)$ and tends to infinity when w_i tends to 0. An $\epsilon > 0$ parameter should therefore be added to prevent numerical instabilities. Notice that we tested this entropy term. It can give satisfactory results but it involves a difficult tuning of the time-steps parameters in an iterative update process. As a consequence, we rather consider our own non-convex term $w_i(1-w_i)$.

B. Energy Minimization

In order to minimize (5), we first recall the dual formulation of the TV:

$$\sum_T |Gu|_1 = \max_{z\in\mathcal{B}} \sum_T uG^*(y),$$

where G^* is the transpose of the discrete operator G and the dual variable z at point $p \in T$ belongs to $\mathbb{R}^{2d}$ if $u(p) \in \mathbb{R}^d$. The convex set $\mathcal{B}$ is then defined, with $k = 1, \ldots, d$, as:

$$\mathcal{B} = \left\{ z = (z_{k,1}, z_{k,2}), z_{k,l} : T \to \mathbb{R}, \text{ s.t.} \sum_{\mathrm{k=1}}^{\mathrm{d}} \sum_{\mathrm{l=1}}^{2} \mathrm{z}_{\mathrm{k,l}}^2 \leq 1 \right\}.$$

We end up with the following non convex problem:

$$\min_{u\in\mathcal{A}, w\in\mathcal{W}} \max_{z\in B} \sum_T uG^*z + \frac{\lambda}{2} \sum_T \sum_{i=1}^N w_i \|u - c_i\|^2 + \frac{\alpha}{2} \sum_T \sum_{i=1}^N w_i(1-w_i) \tag{7}$$

where the unknown variable u is assumed to take its values in a predefined convex set $\mathcal{A}$ characterized with minimal and maximal values for each dimension d:

$$\mathcal{A} = \{u = \{u^k\}, k = 1 \ldots d \text{ and } u^k : T \to [m^k; M^k] \subset \mathbb{R}\}.$$

For instance, in the case of color images in YUV space, $m = \{0, -0.436, -0.615\}$ and $M = \{1, 0.436, 0.615\}$.

The problem (7) is solved with a primal-dual approach [24] allowing us to compute a local minima of the energy function. The final process is summed up in Algorithm 1. We considered separate implicit gradient descent schemes over the variables u ad w, through the computation of the corresponding proximal operators. This approximation nevertheless leads to convergence in our experiments

Algorithm 1 Solving problem (7)

Initialize $w^0 = \frac{1}{N}$, $u^0 = \sum_{i=1}^N w_i^0 c_i$, $z^0 = 0$, $\epsilon \to 0$ and choose $0 < \tau_w < \frac{1}{2\alpha}$ and $\tau_u, \tau_z > 0$ such that $\tau_u\tau_z < \frac{1}{8}$ [24]

while $||u^k - u^{k-1}|| > \epsilon$ **do**

$$z^{k+1} = P_{\mathcal{B}}\left(z^k + \tau_z G u^k\right)$$
$$w_i^{k+1} = P_{\mathcal{W}}\left(\frac{w_i^k - \tau_w(\lambda\|u - c_i\|^2 + \alpha)}{1 - 2\alpha\tau_w}\right),$$
$$u^{k+1} = P_{\mathcal{A}}\left(\frac{u^k + \tau_u(G^* z^{k+1} + \lambda \sum_{i=1}^M w_i^k c_i)}{1 + \tau_u\lambda}\right)$$

end while

The projection operator over the convex set $\mathcal{B}$ reads: $P_{\mathcal{B}}(z) = z/\max(1, \|z\|_2^2)$ with $\|z\|_2^2 = \sum_{k=1}^d \sum_{l=1}^2 z_{k,l}^2$. The projection over the set $\mathcal{W}$ is obtained via the method of [25] that allows to project onto a simplex. The projection on the set $\mathcal{A}$ is simply: $P_{\mathcal{A}}(u^k) = \min(M^k, \max(m^k, u^k))$.

Let us now give the last details necessary to implement the proposed algorithm.

In order to take into account the $2D$ domain of T of size $width \times height$, the spatial operators G and G^*, in Algorithm 1, are defined as follows. The vectors of matrices $G = [G_x; G_y]^t$ and G^* should satisfy $\langle G_x u, z_1\rangle + \langle G_y u, z_2\rangle = \langle u, (G_x^* z_1 + G_y^* z_2)\rangle$. Hence, one can consider a 4-neighborhood $\mathcal{N}(p)$ and the finite differences $Gu(i,j) = [u_x^+(i,j), u_y^+(i,j)]^t$, with $1 \leq i \leq width$ and $1 \leq j \leq height$. With this notation, the first term reads

$$u_x^+(i,j) = \begin{cases} u(i+1,j) - u(i,j) & \text{if } i < width \\ 0 & \text{if } i = width \end{cases}.$$

We then obtain $G^*z = (z_1)_x^- + (z_2)_y^-$, where for the first dimension

$$z_x^-(i,j) = \begin{cases} z(i,j) & \text{if } i = 1 \\ z(i,j) - z(i-1,j) & \text{if } 1 < i < width \\ -z(i-1,j) & \text{if } i = width \end{cases}.$$

IV. Experimental Results

Before presenting several colorized images, we provide an experimental analysis of the minimization of energy (5). All the experiments have been run with the same parameters, *i.e.*, $\tau_u = 0.5$, $\tau_w = 1$, $\tau_z = 0.9/(8\tau_u)$ in Algorithm 1 and $\lambda = 1$ and $\alpha = 0.5$ in (5). For the parameters concerning the colorization method, we have fixed the number of samples to $n = 100$ for each images, $H = 16$ for the number of histograms bins and $N = 8$ for the number of candidates per pixel. For the patches size, we use the following values: $\Sigma_1 = \{3, 5\}$, $\Sigma_2 = \Sigma_3 = \{7, 9, 11\}$.

A. Experimental Feature Analysis

In this section, we show experimentally that the same feature is not adapted to all kind of images.

For each result, only one feature with one patch size is used. In order to observe the real influence of each feature, we do not regularize the colorization results. Hence, as there is only one candidate and no regularization, the variational model is not used. Figure 1 presents the results for the following features/distances: variance, discrete Fourier transform, Wasserstein and the SSD distances. The first feature is computed with 3×3 patches and the three others with 11×11 patches. This experiment shows the interest of combining different features. We can also observe that the candidates provided by SSD are often not well suited for colorization applications. That is the reason why we did not considered this patch metric in the rest of our experiments.

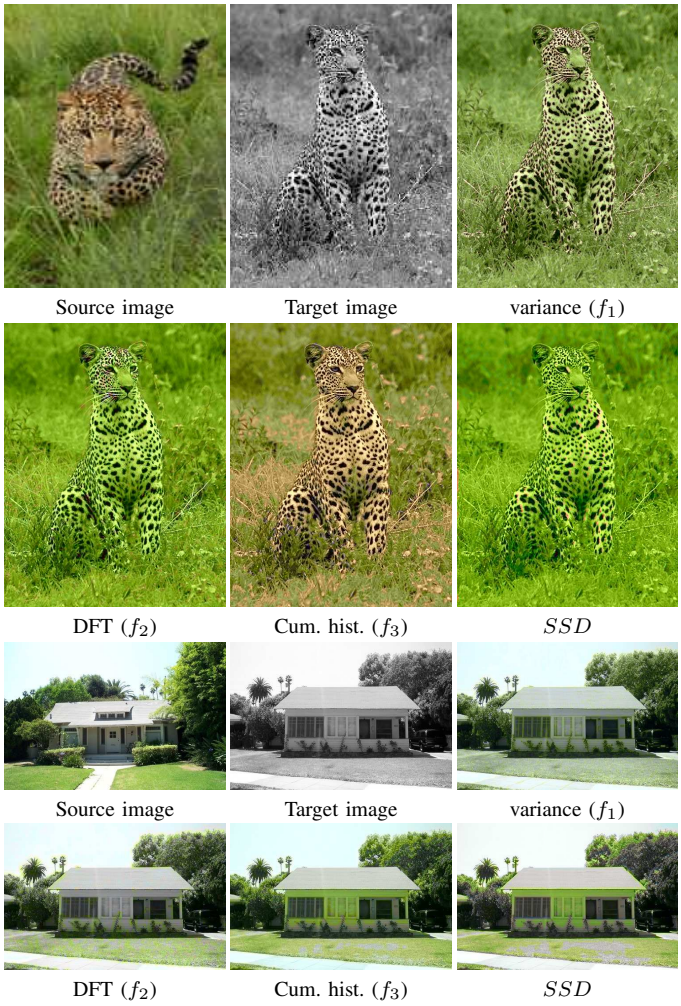


Fig. 1. Influence of the features on the colorization result.

B. Energy Minimization Analysis

In the following experiment, we propose to analyze the values of the energy as well as the values of the weights w_i on a colorization example. The same image is used as the source and the target (Figures 2(a)-(b)) leading to the same luminance images.

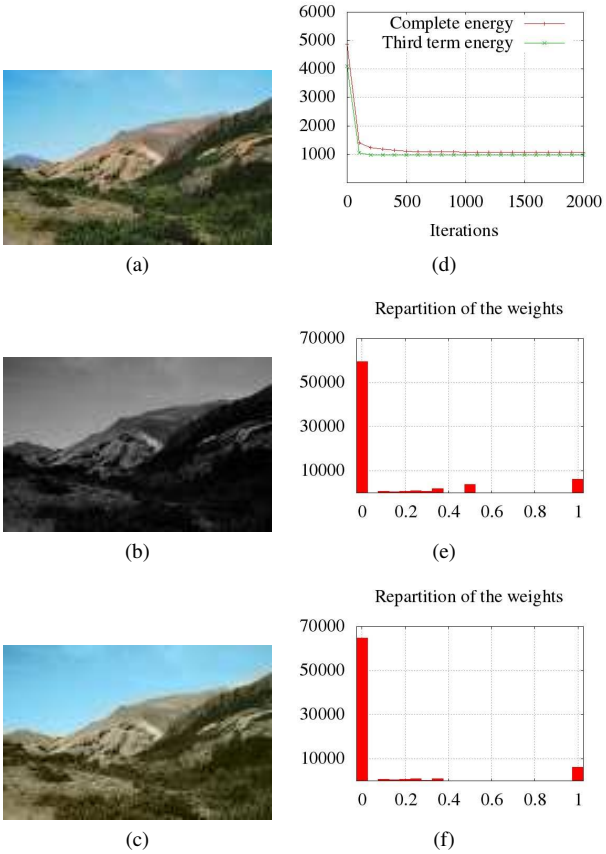


Fig. 2. Analysis of the weights and the energy (see text for details).

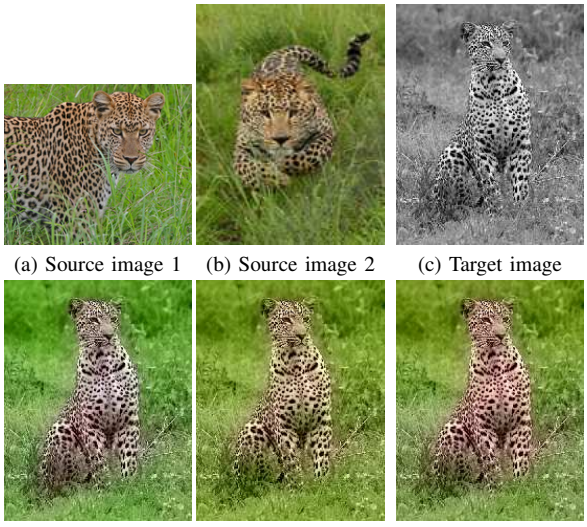
By using only 100 sample points in the source image, we are able to recover an image (Figure 2(c)) that is visually close to the original one (Figure 2(a)). The corresponding energy is presented in Figure 2(d). As one can see, the energy decreases as expected but the third term in energy (7), that enforces the binary nature of the weights, does not reach 0 with the chosen parameters. The final distribution of the weights is shown in Figure 2(e). While at initialization the distribution is uniform (see Algorithm 1), after convergence, we get about 88% of the weights equal to 0. It corresponds to 68% of the pixels having correctly selected one candidate at convergence. For the 32% of pixels left, we can observe that, for a given pixel p, the candidates with $w_i(p) \neq 0$ may corresponds to the same chrominance candidates due to the restricted initial set of color examples ($n = 100$). The consequence is that the different metrics may select the same sample. By removing these redundancies, we finally have 90% (Figure 2(f)) of the pixels that have correctly chosen a single color candidate.

Moreover, we have observed that by increasing the α and τ_w parameters, the third term tends to 0 as expected. The reconstructed image is nevertheless not visually correct with this setting. Indeed, since the initial energy is not convex, the minimization tends to reach a binary local minimum very quickly which does not always correspond to a good candidate satisfying the spatial constraints.

C. Colorization Results

This section presents several colorization results. It is important to notice that all the results presented in this section are obtained with the same set of parameters as described at the beginning of Section IV.

Figure 3 compares our approach with [8], [11], [10] and [21] on two images. Results with [8] have been obtained by implementing the fully-automatic method (no swatches) while the results of [11], [10] and [21] have been directly extracted from the original papers. Our method globally leads to results of the same quality as the ones of [11] or [10][1]. The main advantage of our approach is that it requires less pre-processing steps. By zooming on the images, we can observe that our results are smoother. For instance, the house is less green with our method. Furthermore, contrary to [11], we are able to process the whole image and not only the inner part. Also we want to highlight that the result of [11] presented here has been obtained with a set of multiple reference images. Compared to our previous work in [21], our results are more shiny thanks to the post-processing step and we correctly recover the blue color in the sky of the house image thanks to the variational candidate selection. The second example in Figure 3 presents a comparison with a result extracted from [10]. In their paper, the authors designed a method for image colorization relying on a predefined image segmentation. Each pixels in the target image only searches for its best match inside one segmented area of the source image. Without needing this first segmentation step, our method is able to recover an image that is visually comparable. For both examples, we clearly outperformed the results of [8].



(a) Source image 1 (b) Source image 2 (c) Target image

(d) Result with (a) (e) Result with (b) (f) Result with (a)&(b)

Fig. 4. Influence of the source image on the colorization results.

Figure 4 illustrates the influence of the source images. Obviously, the choice of input data has an impact on the colorization results. As illustrated in this example, depending on the initial image, the result can be different, *e.g.* the grass or the leopard body. To overcome this limitation, one could use many source images (as was done in [9] for image colorization and in [26] for image completion). Figure 4(f) shows a colorization results with the two initial examples as input. This result clearly demonstrates the benefit of the multiple image approach. Figure 5 presents other multiple source image examples in order to colorize grayscale data from the three independent images. The first example is inspired from [11] and a comparison can be found in the associated paper.

(a) Source images

(b) Target image (c) Result with (a)

(d) Source images

(e) Target image (f) Result with (d)

Fig. 5. Colorization with multiple source images.

Figure 6 presents other colorization results. They demonstrate the capacity of our method to colorize different images with exactly the same parameters and without any user intervention (except for the choice of the source image). Finally, the two last examples of Figure 6 are about colorizing old postcard or photos from more recent pictures. Old media have the particularity to be dark and often contain a lot of noise. This makes the colorization process complicated as the features in old and new images do not have the same properties.

[1]The images have been extracted from the article [10] and the website [11] and thus may not have the original quality.

Fig. 3. Comparisons with other methods [11] (that uses multiple reference images), [8] [21], and [10] (see text for more details).

D. Summary of limitations

The proposed method for image colorization lead to very satisfactory results. Nevertheless, it still has several limitations. First, it requires to choose carefully the source images. Secondly, the three chosen features are not well suited to all kind of images. In particular, it seems that some effort should be given to find the good features for old medias. Also, final images are post-processed with luminance remapping which may sometimes create unwanted colors. Finally, one drawback concerns the computational time. Depending on the size of the image, the number of features used and the number of candidates per pixel, the computational time required by all the steps of the proposed colorization pipeline can go from 1 to 10 minutes. For instance, the total processing time for the 500×332 image of the house (Figure 3) is about 6 minutes on a standard linux computer. This point can be easily fixed with code optimization and parallel programming.

V. Conclusion and Future Works

In this work, a variational approach for image colorization has been proposed. First, several possible candidates are computed using different features and associated metrics. The final color is next obtained by solving a variational model which allows selecting automatically the best candidate while adding some regularization in the solution. Experimental results show that our approach produces visually plausible colorization results and can be competitive with respect to the complex methods proposed in the literature.

As future works, we would like to improve the colorization result to overcome the de-saturation effect of the final result. More concern should indeed be given to the reason why all colorization methods produce drab images. We will also include other complex descriptors or features to improve

the characterization of complex image structures. Finally, the extension to video colorization could be considered.

REFERENCES

[1] A. Levin, D. Lischinski, and Y. Weiss, "Colorization using optimization," in *Proc. SIGGRAPH*, 2004, pp. 689–694.

[2] L. Yatziv and G. Sapiro, "Fast image and video colorization using chrominance blending," *IEEE TIP*, vol. 15, no. 5, pp. 1120–1129, 2006.

[3] S. H. Kang and R. March, "Variational models for image colorization via chromacity and brightness decomposition," *IEEE TIP*, vol. 16, no. 09, pp. 2251–2261, 2007.

[4] O. Lézoray, V.-T. Ta, and A. Elmoataz, "Nonlocal graph regularization for image colorization," in *Proc. ICPR*, 2008.

[5] Q. Luan, F. Wen, D. Cohen-Or, L. Liang, Y.-Q. Xu, and H.-Y. Shum, "Natural image colorization," in *Eurographics Symposium on Rendering*, J. Kautz and S. Pattanaik, Eds., 2007.

[6] M. H. Quang, H. K. Sung, and Le, "Image and video colorization using vector-valued reproducing kernel hilbert spaces," *JMIV*, vol. 37, pp. 49–65, 2010.

[7] A. Buades, B. Coll, J. Lisani, and J.-M. Morel, "Conditional image diffusion," *Inverse Problems and Imaging*, vol. 1, no. 4, pp. 593–608, 2007.

[8] T. Welsh, M. Ashikhmin, and K. Mueller, "Transferring color to greyscale images," in *Computer graphics and interactive techniques*, 2002, pp. 277–280.

[9] Y. Morimoto, Y. Taguchi, and T. Naemura, "Automatic colorization of grayscale images using multiple images on the web," in *Proc. SIGGRAPH*, 2009, pp. 59–59.

[10] R. Irony, D. Cohen-Or, and D. Lischinski, "Colorization by example," in *Proc. EGSR*, 2005, pp. 201–210.

[11] G. Charpiat, M. Hofmann, and B. Schölkopf, "Automatic image colorization via multimodal predictions," in *Proc. ECCV*, 2008, pp. 126–139.

[12] S. Liu and X. Zhang, "Automatic grayscale image colorization using histogram regression," *Pattern Recognition Letters*, pp. 1673–1681, 2012.

[13] A. Efros and T. Leung, "Texture synthesis by non-parametric sampling," in *Proc. ICCV*, 1999, pp. 1033–1038.

[14] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. CVPR*, vol. 2, June 2005, pp. 60 – 65.

[15] ——, "Image denoising methods : A review," *SIAM Review*, vol. 52, pp. 113–147, 2010.

[16] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE TIP*, vol. 13, no. 9, pp. 1200–1212, 2004.

[17] P. Arias, G. Facciolo, V. Caselles, and G. Sapiro, "A variational framework for exemplar-based image inpainting," *IJCV*, vol. 93, no. 3, pp. 319–347, 2011.

[18] A. Bugeau, M. Bertalmio and, V. Caselles, and G. Sapiro, "A comprehensive framework for image inpainting," *IEEE TIP*, vol. 19, no. 10, pp. 2634 –2645, 2010.

[19] N. Komodakis and G. Tziritas, "Image completion using efficient belief propagation via priority scheduling and dynamic pruning," *IEEE TIP*, vol. 16, no. 11, pp. 2649–2661, nov. 2007.

[20] Y. Wexler, E. Shechtman, and M. Irani, "Space-time video completion," *IEEE PAMI*, vol. 3, no. 29, pp. 1463–476, march 2007.

[21] A. Bugeau and V.-T. Ta, "Patch-based image colorization," in *Proc. ICPR*, 2012.

[22] A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin, "Image analogies," in *Proc. SIGGRAPH*, 2001, pp. 327–340.

[23] G. Graber, T. Pock, and H. Bischof, "Online 3d reconstruction using convex optimization," in *ICCV Workshops*, 2011, pp. 708–711.

[24] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *JMIV*, vol. 40, pp. 120–145, 2011.

[25] Y. Chen and X. Ye, "Projection onto a simplex," *ArXiv e-prints*, jan 2011.

[26] A. Torralba, R. Fergus, and W. T. Freeman, "80 million tiny images: a large database for non-parametric object and scene recognition," *IEEE PAMI*, vol. 30, no. 11, pp. 1958–1970, 2008.

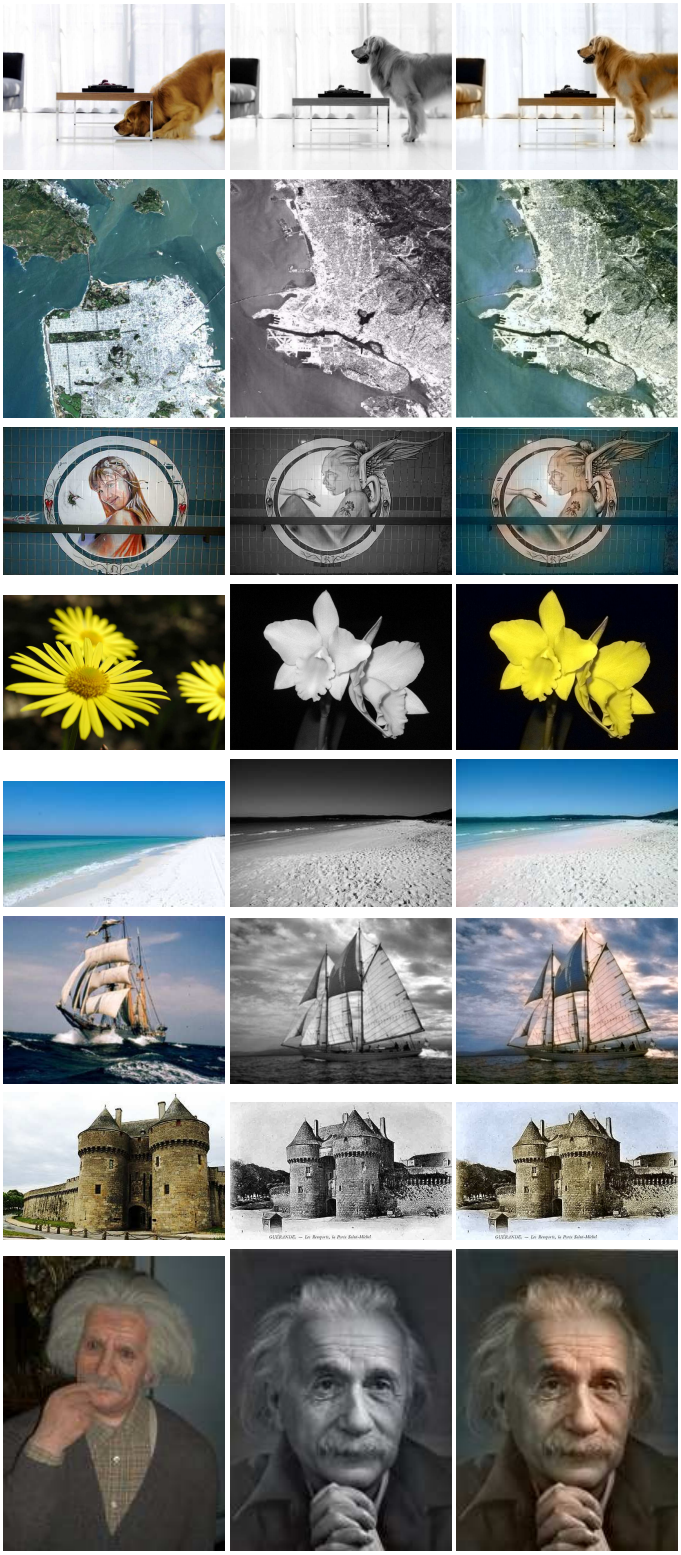


Fig. 6. Colorization results. First column: source images. Second column: target images. Last column: results with our method.