Patch-based Image Colorization

Aurélie Bugeau, Vinh-Thong Ta

Patch-based Image Colorization

Aurélie Bugeau and Vinh-Thong Ta
Univ. Bordeaux, LaBRI, UMR 5800, F-33400 Talence, France
{*aurelie.bugeau, vinh-thong.ta*}*@labri.fr*

Abstract

Image colorization consists in adding colors to grayscale images. Two approaches are mainly used: the first one consists in using manually pre-defined color inputs while the second consider an entire colored image as color example to transfer. The work presented here is in the second category. Indeed, we propose a simple patch-based image colorization based on an input image as a color example. First, we introduce a general colorization model in which many methods of literature can be casted within this framework. Second, we describe our method which is based on patch descriptors of luminance features and a color prediction model with a general distance selection strategy. We also propose to perform a Total Variation (TV) regularization on the colorized image to ensure the spatial color coherency of the final result. Finally, experiments show the potentiality of our proposition in order to automatically colorize grayscale images. Comparisons with methods from the literature are also provided.

1. Introduction

Image colorization is a process of adding colors to monochrome images. Methods proposed in literature can be divided into two main categories. Methods in the first category use manually pre-defined color inputs and mainly rely on optimization, variationnal or diffusion techniques (see for instance [4, 11] or more recently [6] and references therein). Methods in the second category are mainly based on one (or many) initial(s) colored image(s) example(s). Given a color and a grayscale input images, the colorization strategy tries to find, in the color input, the best color to transfer to the grayscale one. The color prediction is performed from the luminance channel of both images. In this context, one can mention the works of [10] that compute simple statistics from luminance images. Recently, [5] and [9] extend [10] to the case of multiple images. In [2], the color prediction is based on machine learning techniques, discrete optimization, continuous color refinement and image descriptors to capture textures or complex structures. One drawback of automatic colorization is the spatial coherency during the color transfer leading to possible inconsistent colorization in the final result.To overcome this limitation, [10] adds user interaction in the colorization process and [2] uses a parameter to enforce the local spatial relationship between pixels.

Main contributions. Inspired by the method of [10], we propose a novel patch-based colorization method. Contrary to [10], our method is completely unsupervised. The spatial coherency is ensured by a Total Variation (TV) regularization (see for instance [1] for more details on TV minimization). Another contribution is the definition of a general distance selection framework that can be extended to other image processing applications than image colorization. Patch-based image colorization has already been proposed, see for instance [8] and references therein, where patch features are computed in order to colorize black-and-white cartoons. Contrary to our proposition, [8] does not consider complex natural images and uses a different distance. Finally, in this work, only the case where one color image is given as input is considered, but our method can be easily extended to multiple images case.

Paper organization. This paper is organized as follows: Section 2 presents a general colorization scheme and explains [10] in this framework. Section 3 describes the different steps of our method and presents the general distance selection model. Section 4 explicits the patch-based descriptors and associated distances. Section 5 shows experimental results and comparisons.

2. General Colorization Scheme

In the rest of this paper, the color source and the grayscale target images will be noted as S and T, respectively. Many methods proposed in literature can be summarized as a general colorization scheme where different steps are performed. Indeed, we can divide these methods into three main steps: step (1): pre-process S and T, step (2): predict color from S and transfer to T step (3): post-process T.

Generally, step (1) consists in converting both S and T to a luminance-chrominance color space. The luminance channel of S is then modified in order to be comparable to the luminance of T. In step (2), for each pixel $p \in T$, the best pixel $q \in S$ is found by minimizing over q a distance on the luminance of the neighborhoods of p and q. In general, to speed up the process, the best q is only searched within a reduced number of samples of S. The chrominance of the source pixel is finally transfer to p. The last step (3) generally consists in a simple color space re-conversion or a color refinement [2].

One of the first colorization scheme of this type has been proposed in [10]. First, S and T are converted to the de-correlated $l\alpha\beta$ space [7] and a luminance remapping [3] is applied to the l channel of S in order to fit the T one. Next, approximately 200 pixels of S are selected on a randomly jittered grid. Their color prediction model is as follows: for each pixel $p \in T$, p is compared with the 200 samples. A distance based on the average of the standard deviation on a 5×5 neighborhood and the luminance of the pixel is computed. The best source pixel q is the one which minimizes this distance and the color transfer is performed as: $(p^\alpha, p^\beta) = (q^\alpha, q^\beta)$. Finally, the post-processing step consists in converting the target image from $l\alpha\beta$ to RGB color space. Similarly, the method proposed by [2] can also be casted in the same framework.

3. Proposed Image Colorization Method

In this section, we describe our image colorization scheme based on patch features as pixels descriptors to capture image textures or complex structures.

Pre-processing Step. The proposed pre-processing step is similar to [10], *i.e.*, first, S and T are converted to a de-correlated color space. A luminance mapping [3] from T to S is also performed. Then, from S, n samples are randomly selected on a jittered grid. The main difference with [10] is the choice of the de-correlated color space. The transformation for RGB to $l\alpha\beta$ is non linear and therefore this space is non convex. In our experiments, we found that using such color space can lead to visually inconsistent results. Hence, we prefer working with the YUV color space that remains convex.

Distance Selection Framework and Color Prediction Model Let $F_l(p)$ be a general set of l general features that depends on a pixel p such that

$$F_l(p) = \{f_1(p), \dots, f_l(p)\}. \tag{1}$$

For a given feature $r = 1 \dots l$, let $d_r(p, q) = d_r(f_r(p), f_r(q))$ be an associated function that computes the distance between the features of p and q. Our color prediction model consists in predicting the best chrominance to transfer to a pixel $p \in T$ using the features space of S_n, $\{F_l(q_1), \dots, F_l(q_n)\}$.

We denote as $S_n = \{q_1, \dots, q_n\}$ the set of n random pixels of S. The color prediction framework starts by selecting, for each pixel $p \in T$, k candidates

$$\{\tilde{q}_{i,r} \in S_n, i = 1 \dots k\} \tag{2}$$

with respect to each feature $r = 1 \dots l$. These k source pixels are the ones leading to the smaller distance $d_r(p, q)$. By finding the k best source pixels for all the l features and distances, we finally obtain and the following corresponding chrominance set is

$$C_k = \{(\tilde{q}^U_{i,r}, \tilde{q}^V_{i,r}), i = 1 \dots k, r = 1 \dots l\}. \tag{3}$$

We now have a reduced set of possible chrominance for each pixel p of the grayscale image. The chrominance $(\hat{q}^U, \hat{q}^V)$ of the source pixel $\hat{q}$ that is finally transferred to the pixel p is the median of C_k. As C_k is a two-dimensional set, a PCA decomposition is computed so that the median of the set is the median of the projection of each chrominance onto the first principal component. Finally the color transfer is performed as $(p^U, p^V) = (\hat{q}^U, \hat{q}^V)$ and leads to the predicted image T_0.

Post-processing Step In order to obtain consistent and spatially coherent colorization, image T_0 is post-process with a TV regularization. Indeed, the result T_0 obtained with our color prediction model may not be smooth. Each pixel $p \in T_0$ is colorized without considering the result of its neighbors. To overcome this limitation, we take advantage of the recent advances in TV minimization. Using the notation $T = (T_Y, T_U, T_V)$, we minimize the following TV energy [1]:

$$\min_{T_U, T_V} \int_\Omega |\nabla T_U| + |\nabla T_V| + \frac{\lambda}{2}\|T - T_0\|^2, \tag{4}$$

where $\Omega \subset \mathbb{R}^2$ is the image domain, and T_0 is the colored image resulting from our color prediction model. The parameter λ weights the influence of each term. To optimize (4), we use the recent algorithm (ALG2) proposed in [1] based on convex optimization.

4. Patch-based Features and Distances

In the following, we describe the construction of features set (1) and associated distances.

Natural images contains different types of complex structures, redundancies and textures. In order to perform colorization on natural images, our method needs

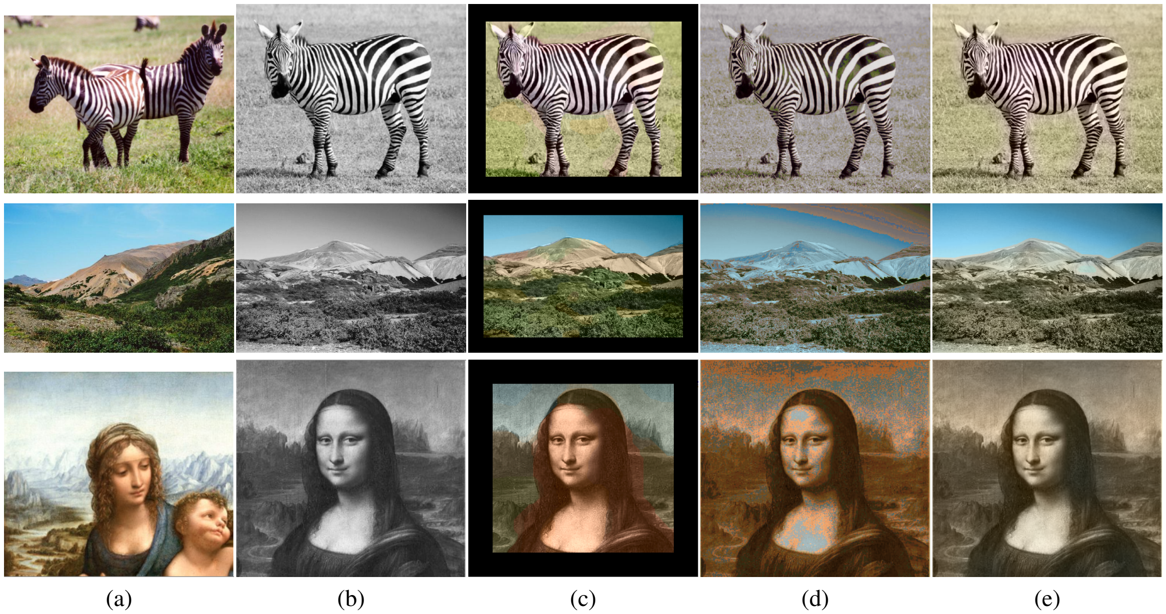


Figure 1. Comparisons with [2] and [10]. (a): source image S (b): target image T (c): [2] results (d): [10] results (e): our method.

to integrate such information. To capture these structure, we use patch-based features and distances computed from luminance images.

A patch of size $(2s+1)^2$ centered at pixel $p = (p_x, p_y)$ can be defined as $P_s(p) = \{L(p_x+i, p_y+j)\}$ for all $i, j = -s \dots s$. Given luminance image L, $L(p) = L(p_x, p_y)$ is the luminance value of $p \in L$.

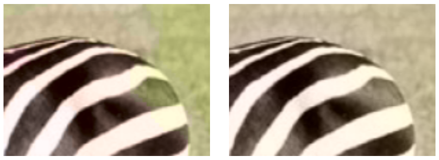
In order to capture natural image structures, we choose simple features (more complex features can be used) that characterize patch luminance, global statistics, orientation or redundancies. Here, we use three different features ($l = 3$), each one capturing a different characteristic of the luminance. For any pixel $p \in L$, we compute the usual variance $f_1(p) = \nu(P_s(p))$ which differentiates uniform areas and textured area. The second feature is the amplitude spectrum of the Discrete Fourier Transform $f_2(p) = \|\,\mathrm{DFT}(P_s(p))\|$ which give more information on the frequencies of the most important structures. Finally, the last one is the luminance histogram containing b bins $f_3(p) = h(P_s(p), b) = |P_s(p)|^{-1} \sum_{q \in P_s(p)} \delta(P_s(q), b)$, where $\delta(i,j) = 1$ if $i = j$, 0 otherwise. Notation $|A|$ denotes the cardinal of the set A. This histogram allows taking into account the values of the luminance inside the patch.

From the features space, we can now define the corresponding distances d_r with $r = \{1,2,3\}$ such that for two pixels p and q, $d_1(p,q) = |f_1(p) - f_1(q)|$, $d_2(p,q) = d_3(p,q) = d_n(p,q)$ where $d_n(p,q) = \sum_{i,j} |f_n(i) - f_n(j)|$ with $i = 1 \dots |f_n(p)|$, $j = 1 \dots |f_n(q)|$ and $n = \{2,3\}$. One can note that distance d_n corresponds to the standard sum of absolute differences for a given feature.

5. Experimental Results

In this section, we show several experimental results where all colorization are obtained with the same parameters, *i.e.*, the number of samples is approximatively 200 for each images, $k = 3$ in (2), $s = 5$ for the patch size, $\lambda = 2$ in (4) and $b = 8$ for the number of bins in histograms.

Figure 1 compares our approach with [10] and [2]. Our method globally leads to result of the same quality as the ones with [2]. The images for this method have been directly extracted from the website associated to [2]. However, as shown in Figure 2 and also on the Joconde images, one can see that our approach produces more plausible colors thanks to smoothing effect of regularization post-processing. Indeed, as shown in Figure 1(c), method proposed by [2] fails to capture image edges. Results with [10] have be obtained by implementing the fully-automatic method (no swatches). For the images presented in Figure 1(d), this method clearly produces inconsistent colorization. In particular, for the landscape, the luminance of the sky in the two input images is different and a distance measure that re-

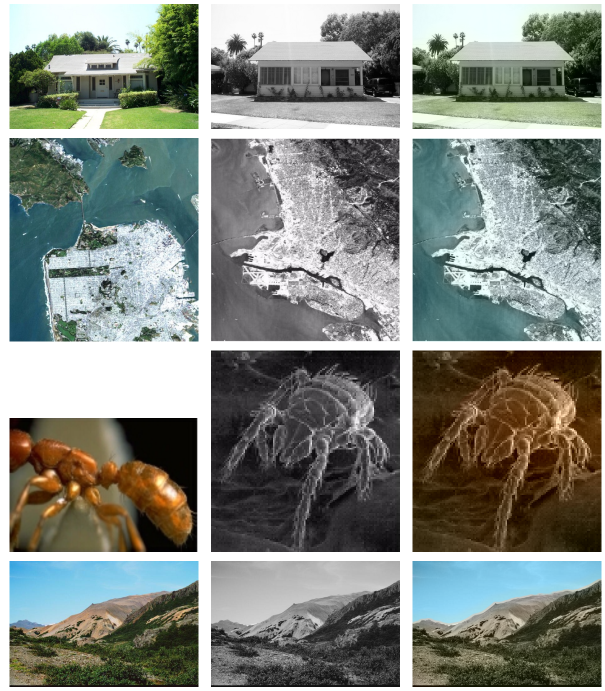


(a) (b)

Figure 2. Zoom on the zebra image. (a) [2] result (b) our method.

lies on the luminance values produces a bad distance estimation. For the Joconde image, one can also see the saturation effect of the colorization probably due to the $l\alpha\beta$ color space.

Finally, Figure 3 presents other colorization results. They demonstrate the capacity of our method to colorize different type of complex natural images with exactly the same parameters and without any user intervention. The last row of Figure 3 shows an interesting result since in this experiment we use the same image as source and target leading to the same luminance images. It demonstrates the importance of the luminance mapping in the pre-processing step in order to obtain satisfying colorization results



(a) (b) (c)

Figure 3. Results on different images. (a) source image S (b) target image T (c) our method (see text for more details).

6. Conclusion and Future Works

In this work, a simple patch-based image colorization method and a general distance selection framework for color prediction have been proposed. Experimental results show that our approach produces visually plausible colorization results and can be competitive with methods proposed in literature.

As future works, we can for instance improve the colorization result to overcome the de-saturation effect of the final result. We can also study other complex descriptors or features to improve the characterization of complex image structure. Finally, the extension to video colorization could be considered.

References

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

[2] G. Charpiat, M. Hofmann, and B. Schölkopf. Automatic image colorization via multimodal predictions. In *Proc. ECCV*, pages 126–139, 2008.

[3] A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin. Image analogies. In *Proc. SIGGRAPH*, pages 327–340, 2001.

[4] A. Levin, D. Lischinski, and Y. Weiss. Colorization using optimization. In *Proc. SIGGRAPH*, pages 689–694, 2004.

[5] Y. Morimoto, Y. Taguchi, and T. Naemura. Automatic colorization of grayscale images using multiple images on the web. In *Proc. SIGGRAPH*, pages 59–59, 2009.

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

[7] D. L. Ruderman, T. W. Cronin, and C.-C. Chiao. Statistics of cone responses to natural images: implications for visual coding. *J. Optical Soc. of America*, 15(8):2036–2045, 1998.

[8] D. Sýkora, J. Buriánek, and J. Žára. Colorization of black-and-white cartoons. *Image and Vision Computing*, 23:767–782, 2005.

[9] A. Torralba, R. Fergus, and W. T. Freeman. 80 million tiny images: a large database for non-parametric object and scene recognition. *IEEE PAMI*, 30(11):1958–1970, 2008.

[10] T. Welsh, M. Ashikhmin, and K. Mueller. Transferring color to greyscale images. In *Computer graphics and interactive techniques*, pages 277–280, 2002.

[11] L. Yatziv and G. Sapiro. Fast image and video colorization using chrominance blending. *IEEE TIP*, 15(5):1120–1129, 2006.