Non-convex Relaxation of Optimal Transport for Color Transfer Between Images

Julien Rabin, Nicolas Papadakis

Non-convex relaxation of optimal transport for color transfer between color images

Julien Rabin[*]
GREYC, Université de Caen

Nicolas Papadakis[†]
IMB, CNRS

Abstract

Optimal transport (OT) is at present a major statistical tool to measure similarity between features or to match and average features. However, a major drawback of this framework is the lack of regularity of the transport map and robustness to outliers. In practice, as it has been partially addressed in previous works, OT requires some relaxation and regularization to achieve these desirable properties. With such methods, as one feature can be matched to several ones, important interpolations between different features arise. This is not an issue for comparison purposes, but it involves strong and unwanted smoothing for transfer applications.

In this paper, we introduce a new regularized method based on a non-convex formulation that minimizes the transport dispersion by enforcing the one-to-one matching of features. After some illustrations with 1-D histogram matching, the interest of the proposed approach is demonstrated for color transfer. In our color transfer experiments, we show that the minimization of the transport dispersion combined with regularization enables to reduce color artifacts and color mixing.

1 Introduction

Many image processing applications require the modification or the prescription of some characteristics (such as colors, frequencies, patches, or wavelet coefficients) of a given image, while preserving other features. The statistics that have to be prescribed may come from prior knowledge, or more generally, are learned from an example. In such a case, another image is selected from a database to define a template.

Such a framework arises for image enhancement (considering either contrast, color, or geometry), inpainting, colorization of grayscale or infrared images, gammut mapping, tone mapping, color grading or color transfer. In this paper, we will focus on this last application in the context of histogram transfer between images.

[*] julien.rabin@unicaen.fr
[†] nicolas.papadakis@math.u-bordeaux1.fr

1.1 Color Transfer

Color transfer consists in modifying an image to match the color palette of another one, while preserving its geometry. In the literature, the different interpretations and definitions of color palettes have led to various algorithms. In the following, we only consider non supervised approaches.

Parametric modeling Following the seminal work of [22], various methods have first been designed to transfer some simple color statistics (*i.e.* the mean and standard deviation) in any color space. This strategy of parametric modeling has later been generalized for multi-variate Gaussian [14], and Gaussian mixtures [25].

Histogram modeling More general approaches match the complete empirical distribution of features from two images. When considering grayscale images, the problem of matching the gray levels is also known as 1-D *histogram specification*, and can be solved in linear time using look-up tables. Several works have been proposed to extend this framework for color histograms, using for instance 3-D cumulative histograms or 1-D ΔE-color index [8].

Histogram matching *via* Optimal transportation As pointed out by [15], the strong connection between histogram specification and the Monge-Kantorovich Optimal Transport framework [26] makes it possible to generalize the histogram specification principle to any multi-dimensional histogram matching problem. The Optimal Transport (OT) problem consists in estimating the map that transfers a source probability distribution onto a target one, while minimizing a given cost function. The corresponding transport cost –also known as the Wasserstein distance or the Earth Mover's distance– has been shown to produce state of the art results for the comparison of discrete statistical descriptors [23]. The transport cost is in practice obtained from the optimal transport map. This map is the key element to perform the transfer of color. Some approaches to find fast approximate solution of OT were investigated in [15, 21].

Spatial information Unfortunately, as pointed out by [18, 16], the exact transfer of color palette is generally not satisfying for practical applications in image processing. Indeed, the color distributions to be matched may have very different shapes, so that outliers generally appear in the processed image. Moreover, as the transfer is performed in the color space, it does not take into account the fact that coherent colors should be transferred to neighboring pixels, resulting in undesirable artifacts, such as JPEG compression blocks, enhanced noise, saturation, contrast inversion and color inconsistencies [15, 17, 24]. As a consequence, various models have been designed to consider the spatial nature of images [12] and to incorporate some regularity priors on the image domain, such as Total Variation [10]. Similarly to previous works on image enhancement, color transfer may be formalized [13, 20] as a variational problem in the image domain, in order to directly incorporate a spatial regularization of colors. A post-regularization of the image color is nevertheless required to restore the fine details and the grain of the original image [15, 17].

Approximate and regularized matching While spatial regularization manages to suppress small artifacts due to exact histogram specification, it cannot handle very strong artifacts due to an irregular transport map [20]. To tackle this problem, Fer-

radans *et al.* [6] introduced a new variational framework for the regularization of optimal assignment between point clouds. The exact matching constraint is relaxed to enforce robustness to outliers. Instead of providing one-to-one assignment, this relaxed approach enables one-to-many correspondences by using capacity variables. It is demonstrated that the additional introduction of regularity priors on the gradient of the average transport map produces smoother and more robust transport maps for image processing applications. This work has been recently extended in [19] for histogram matching problem, where spatial information is used to drive the regularization of color transfer, together with an automatic estimation of the relaxation capacity variables. Independently, Cuturi [5] also introduced a new variational framework for smooth optimal transport between histograms based on entropy prior. The proposed approach, which makes it possible to design very fast algorithms for intensive histogram comparison, also yields one-to-many correspondences between histogram bins, and is proven to be more robust to outliers.

Color dispersion Due to the one-to-many relaxation and the fact that only the gradient of the *average transport* flow is penalized, the regularization does not prevent the transport map to associate very different colors to a single pixel or cluster. This leads to undesirable results such as color mixing or color inconsistencies in the modified image.

1.2 Contributions and outline

In this paper, we propose a new model that takes into account the aforementioned issues. We also relax and regularize the transport map and introduce a non-convex constraint that minimizes the variance of colors assigned to each cluster. We also rely on a fast proximal splitting algorithm in order to compute the transport map that is finally used for color transfer purposes.

The organization of the paper is as follows. Background on OT is given in Section 2. The proposed model is introduced in Section 3 and experimented in Section 4.

2 Color transfer *via* Optimal Transport

From now on, we refer to u as the input image to be modified, and to v as an exemplar image v provided by the user. For mandatory efficient computation and compact representation, we consider clustered feature distributions, which may be seen as multi-dimensional histograms or discrete probability distribution (for instance color palettes, patch dictionaries, wavelet coefficient histograms). We refer to $h_u = \sum_i h_u[i]\delta_{X_i}$ as the histogram of features $X := \{X_i \in \mathbb{R}^d\}_{i \leq n} \in \mathbb{R}^{n \times d}$ from the input image, so that $\sum_{i=1}^n h_u[i] = 1$. The histogram is thus composed of n features, each of them being of dimension d. In the same way, $h_v = \sum_{j=1}^m h_v[j]\delta_{Y_i}$ is the target distribution of the desired features $Y := \{Y_j \in \mathbb{R}^d\}_{j \leq m} \in \mathbb{R}^{m \times d}$. We do not assume anything about the way those histograms are built (uniform quantization, k-means, mean-shift, image segmentation, *etc*). This means that we have a quantized version of our image, for instance using Nearest-Neighbor interpolation (w.r.t a metric d, that will be the quadratic L_2 distance in the following):

$$\tilde{u}_i = X_{I(i)} \text{ where } I(i) = \operatorname{argmin}_I d(u(i), X_I).$$

2.1 Optimal transport of histogram (histogram specification)

The optimal transport of h_u onto h_v is obtained by estimating the optimal transport matrix:

$$P^\star \in \underset{P \in \mathcal{P}_{h_u,h_v}}{\operatorname{argmin}} \; C(P) := \langle P,\, C_{X,Y} \rangle \tag{1}$$

with the histogram constraint set $\mathcal{P}_{h_u,h_v} = \{\mathbb{R}^{n\times m}, P_{i,j} \in [0,1], P\mathbf{1}_m = h_u, P^T\mathbf{1}_n = h_v\}$ where $\mathbf{1}_N \in \mathbb{R}^N$ is the unit vector, and the cost matrix $C_{X,Y}$ is generally defined from quadratic distances $(C_{X,Y})_{i,j} = \|X_i - Y_j\|^2$.

Sinkhorn distance A variant approach, based on entropy regularization of this framework, has been studied by Cuturi in [5]

$$P^\star_\gamma \in \underset{P \in \mathcal{P}_{h_u,h_v}}{\operatorname{argmin}} \; C(P) - \gamma h(P) \quad \text{where } h(P) = -\sum_{i,j} P_{i,j} \log P_{i,j}. \tag{2}$$

This formulation enables the use of the fast Sinkhorn algorithm while providing smooth transfer matrix which turns out to be more robust to outliers. Observe that setting $\gamma = 0$ boils down to solve the original optimal transport problem (1), while using $\gamma = \infty$ yields $P^\star_\gamma = h_u h_v^T$ (uniform transport flow that maximizes entropy).

Optimal Mapping For transfer purposes, we need to compute a transfer map to change the statistical distribution of u accordingly to the transport matrix P. A naive approach would be to match directly centroid features between clusters, leading to quantization artifacts:

$$T^\star : \tilde{u}_i = X_{I(i)} \mapsto Y_J \text{ where } J \text{ is such that } J \in \operatorname{argmax}_j P^\star_{I(i),j}. \tag{3}$$

To reduce this effect, as proposed in [25] and recently in [19], spatial information can also be incorporated in a multivariate Gaussian mixture model to define the smoothed mapping

$$S : u_i \mapsto \frac{1}{W} \sum_i w_i(u_i) T(X_{I(i)}) \tag{4}$$

which average a given mapping T using adaptive weights functions $w_i(\cdot) = \exp\left(-\frac{1}{2}\|\cdot - X_i\|^2_{V_i}\right)$, and a normalization factor $W = \sum_i w_i$. Observe that this method uses estimated covariance matrices V_i of clusters in spatial and color product space.

2.2 Optimal transport relaxation

As previously mentioned, one of the limitation of optimal transport (partially discussed in [18]) comes from the exact matching constraint which is not robust to outliers. To address this issue, a first attempt has been proposed in [23] and later in [3] to define optimal transport between unbalanced histograms. This idea has been generalized in [6], by making use of relaxed constraints for the optimal assignment problem, defining min/max capacities on the optimal flow. Rewriting this relaxed optimal transport problem in our context of histogram matching gives:

$$P^\star_\kappa \in \underset{P \in \mathcal{K}_{h_u,h_v}}{\operatorname{argmin}} \; C(P) \tag{5}$$

with the relaxed histogram constraint set $\mathcal{K}_{h_u,h_v} = \{\mathbb{R}^{n\times m}, P_{i,j} \in [0,1], P\mathbf{1}_m = h_u, \kappa_{min} h_v \leq P^T\mathbf{1}_n \leq \kappa_{max} h_v\}$. This model includes two vectorial parameters, $\kappa_{min} \leqslant 1 \in \mathbb{R}^m$ and $\kappa_{max} \geqslant 1 \in \mathbb{R}^m$, that control the proportion of the target histogram's bins that can be used by the color transfer. Outliers can therefore be taken into account by taking $\kappa_{min} < 1$. In this case, by setting $\kappa_{max} > 1$, some colors of the source palette will be used more frequently than in the original example image.

The limitation of this relaxed transport is that there is no statistical control of how "close" the transported histogram is over the source one and that it is very difficult to tune so many parameters by hand. In [19], the following extended model has been shown to tackle these two limitations, by including the calibration of the capacity parameters within the model:

$$(P,\kappa)^\star \in \underset{\substack{P\geq 0 \in \mathbb{R}^{n\times m} \ s.t.\ P\mathbf{1}_m = h_u \\ \kappa \geq 0 \in \mathbb{R}^m \ s.t.\ \langle \kappa, h_v\rangle \geq 1}}{\operatorname{argmin}} C(P) + \rho\|\kappa - \mathbf{1}_m\|_1. \tag{6}$$

2.3 Optimal transport regularization

Another limitation of the optimal transport framework is the lack of control over the regularity of the solution. In [6], the authors propose to measure the regularity of the average transfer mapping, as detailed in this section.

Average transfer First, one considers the following definition of *Posterior mean* to define a one-to-one mapping T from a transfer matrix P, instead of the NN mapping in Eq. (3):

$$T(X_i) = \overline{Y_i} = \frac{1}{\sum_{j=1}^m P_{ij}} \sum_{j=1}^m P_{ij} Y_j = \frac{1}{h_u(i)} \sum_j P_{i,j} Y_j = (D_{h_u} PY)_i, \tag{7}$$

where the normalization matrix D_{h_u} is diagonal: $(D_{h_u})_{ii} = h_u(i)^{-1}$.

Gradient of flow on graph The regularity of this average transfer map is then evaluated on a graph $\mathcal{G}_X = (I_X, E_X)$, built from the set of input features $\{X_i\}_i$. Denoting as $I_X = \{I_1, \cdots, I_n\}$ the set of nodes representing the features $\{X_1, \ldots X_n\}$, and $E_X \subset I_X^2$ the set of edges, the gradient $G_X V \in \mathbb{R}^{n\times n\times d}$ of a multi-valued function $V = \{V_i^l\} \in \mathbb{R}^{n\times d}$ on $\mathcal{G}_X$ is computed at point X_i as

$$(G_X V)_i = (w_{ij}(V_i - V_j))_{j\in E_X(i)} \in \mathbb{R}^{n\times d},$$

where the weight w_{ij} between features X_i and X_j relies on their similarity, *i.e.* using for instance $w_{ij} \propto \exp -d(X_i, X_j)$.

Regularized optimal transport problem The optimal transport matrix now solves the following problem

$$P^\star \in \underset{P\in\mathcal{K}_{h_u,h_v}}{\operatorname{argmin}} \ C(P) + \lambda\|G_X(D_{h_u}PY - X)\|, \tag{8}$$

where $\|G_X V\|$ can be interpreted as the TV norm of field V on the graph $\mathcal{G}_X$. The flow is taken as and $V = T(X) - X$ so that color translation are not penalized. With such regularization, artifacts or contrast inversion are also avoided.

Advantage and drawback The combination of problem (8) with relaxed formulation(6) yields smooth transport maps. However, as shown in the experimental section, the regularization of the *average* flow encourages transport between one cluster to many. In practice, this is undesirable for color transfer since the obtained prescribed colors are defined from linear combination of the target color palette, resulting in false colors artifacts and a lost of color contrast.

In this paper, we propose to solve this issue by incorporating information on the color transfer dispersion.

3 Non convex relaxation of color palette transport

The relaxation here considered is different from the one proposed in [6] where a capacity relaxation of the target histogram is considered. The closeness to the target histogram is here imposed through a data fidelity term which makes easier the control of the results while simplifying the projection onto the set of acceptable transport matrices.

By using linear programming to optimize the regularized problem ((6)) as in [6, 19], the dimension of the variables to estimate is greatly increased as one aditional variable is needed for each regularization constraint. Simplifications of the regularization term (through the mean transport and the use of divergence) are thus needed to reduce the complexity. Such regularizers limit the inter-cluster color dispersion but they induce the creation of new drab colors since an important interpolation of the target color palette may occur (i.e the intra-cluster color variance may be large with the one-to-many assignment). To cope with this issue, we here propose to penalize the dispersion of assigned colors with a non convex constraint. We also consider a different optimization tool which decreases the dimension of the problem with respect to linear programming, and we only deal with the estimation of the transport matrix.

3.1 Optimization problem

In order to deal with the aforementioned limitations which are intrinsic to the optimal transport framework, we propose the following relaxed and regularized optimal transport problem:

$$P^\star \in \underset{P \in \mathcal{P}_{h_u}}{\operatorname{argmin}} \left\{ \mathcal{E}(P) := C(P) + \rho F(P) + \lambda R(P) + \alpha D(P) \right\} \tag{9}$$

where:

- $\mathcal{P}_{h_u} = \{\mathbb{R}^{n \times m}, P \geq 0, P\mathbf{1}_m = h_u\}$ is the convex set of right stochastic matrices (where each row sums to the corresponding bin value in h_u).
- $C(P) = \langle C_{X,Y}, P \rangle$ is the linear cost matching function;
- $F(P)$ is the fidelity term w.r.t the target histogram h_v;
- $R(P)$ is the regularization prior on the average transport mapping;
- $D(P)$ is the dispersion term which measures the "sparsity" of average transport map;

In order to solve this problem very fast with projected gradient descent, we used smooth (differentiable) functions detailed hereafter. Observe that other choice would lead to different optimization algorithms.

Linear Constraint The linear constraint $P \in \mathcal{P}_{h_u}$ may be incorporated using an indicator function

$$\iota_{\mathcal{P}_{h_u}}(P) = \begin{cases} 0 & \text{if } P \in \mathcal{P}_{h_u} \\ +\infty & \text{otherwise} \end{cases}$$

The Euclidean orthogonal projector $\text{Proj}_{\mathcal{P}_{h_u}}$ onto the simplex $\mathcal{P}_{h_u}$ can be computed in $O(nm\log(m))$ by using, for each row of the matrix, a variant of the Euclidean projector onto the probability simplex (see for instance [4]).

Fidelity term As the set of acceptable transport matrices $\mathcal{P}_{h_u}$ does not anymore take into account the target distribution, we have to make sure that the transported histogram $P^T\mathbf{1}_n$ is close enough to the target histogram h_v. To do so, we rely on the Pearson's χ^2 statistics, which writes for any bistochastic matrix $P \in \mathcal{P}_{h_u,h_v}$

$$\begin{aligned} F(P) &= \tfrac{1}{2}\chi^2_{h_v}(P^T\mathbf{1}_n) = \tfrac{1}{2}\|D_{h_v}(P^T\mathbf{1}_n - h_v)\|^2 \\ &= \tfrac{1}{2}\sum_j \tfrac{1}{h_v[j]}\left(\sum_i P_{i,j} - h_v[j]\right)^2 = \tfrac{1}{2}\left\|\mathbf{1}_n^T P D_{h_v}^{1/2}\right\|^2 - \tfrac{1}{2}. \end{aligned} \tag{10}$$

We assume therefore from here, without loss of generality, that h_v has non empty bins. Observe that the corresponding fidelity term can be interpreted as a weighted L^2 metric, which further penalizes bins of the target histogram that has small values. This will prevent the model from using very rare features from the exemplar image. The gradient then reads

$$\nabla F(P) = \left[\frac{\sum_\ell P_{\ell,j}}{h_v[j]}\right]_{i,j} = \mathbf{1}_n \cdot (\mathbf{1}_n^T P D_{h_v}) = \mathbf{1}_{n\times n} P D_{h_v}. \tag{11}$$

Regularity term We consider a Tikhonov regularization of the gradient of a flow V, incorporating spatial information from the input feature distribution in the gradient operator G_X defined on the graph of clusters $\mathcal{G}_X$. We measure the gradient of the mean flow $V = D_{h_u}PY - X$ and therefore define

$$R(P) = \frac{1}{2}\|D_{h_u}^{-1}G_X(D_{h_u}PY - X)\|_2^2, \tag{12}$$

where the value of the gradient norm is weighted by the corresponding histogram bin value: $(D_{h_u}{}^{-1})_{ii} = h_u[i]$. The derivative is related to the graph-Laplacian $G_X^T G_X$:

$$\nabla R(P) = D_{h_u} G_X^T D_{h_u}^{-2} G_X (D_{h_u}PY - X)Y^T. \tag{13}$$

Dispersion term As we pointed out before, the regularized transfer induces a high variability of color assigned to each input color. We then consider the minimization of the variance of the flow. Denoting $\overline{Y} = D_{h_u}PY$ from Eq. 7, the intra-cluster variance of the color assigned to X_i is defined as:

$$\begin{aligned} \text{Var}(Y)_i &:= \left(\overline{\left(Y - \overline{Y}_i\right)^2}\right)_i = \tfrac{1}{h_u[i]}\sum_j P_{i,j}\|Y_j - \overline{Y}_i\|^2 \\ &= \left(\overline{Y^2}\right)_i - \overline{Y}_i^2 = \tfrac{1}{h_u[i]}\sum_j P_{i,j}\|Y_j\|^2 - \left\|\tfrac{1}{h_u[i]}\sum_j P_{i,j}Y_j\right\|^2. \end{aligned}$$

We therefore penalize this variance with respect to each cluster weight and thus obtain the functional term:

$$\begin{aligned} D(P) &:= \sum_i h_u[i] \operatorname{Var}(Y)_i = \sum_{i,j} P_{i,j}\|Y_j\|^2 - \sum_i \tfrac{1}{h_u[i]} \left\| \sum_j P_{i,j} Y_j \right\|^2 \\ &= \sum_{i,j} \left(P_{i,j}\|Y_j\|^2 - \tfrac{1}{h_u[i]} \sum_k P_{i,j} P_{i,k} Y_j^T Y_k \right) \\ &= \sum_{i,j} P_{i,j} Y_j^T \left(Y_j - \tfrac{1}{h_u[i]} \sum_k P_{i,k} Y_k \right) \\ &= \langle P,\, \mathbf{1}_n \operatorname{Diag}(YY^T)^T - D_{h_u}^{-1} PYY^T \rangle \end{aligned} \tag{14}$$

where $\operatorname{Diag} : \mathbb{R}^{n\times n} \mapsto \mathbb{R}^n$ is the diagonal extraction of a square matrix, that is used here to compute the norm vector $\operatorname{Diag}(YY^T) = (Y^T \odot Y^T)\mathbf{1}_d$. The derivative writes

$$\nabla D(P) = \mathbf{1}_n \operatorname{Diag}(YY^T)^T - 2D_{h_u}^{-1} PYY^T. \tag{15}$$

3.2 Algorithm

Because of the variance flow penalization term, the objective function is non-convex. Assuming that the functional of problem (9) satisfies the Kurdyka-Lojasiewicz property (namely that the functional is not too flat in the neighborhood of its local minima), the Forward-Backward algorithm can be used to find a critical point of this non-convex problem [2], since it contains a convex non-smooth term (the linear constraint $\iota_{\mathcal{P}_{h_u}}(P)$) and the sum of differentiable terms $G(P) = \langle C_{XY},\, P\rangle + \lambda R(P) + \rho F(P) + \alpha D(P)$. The gradient of G is L-lipschitz, where L is proportional to ρ, λ, α and m. A fast estimation gives us $L \leqslant \tilde{L} = \lambda\|G_X^T G_X\|_2\|YY^T\|_2 + \rho m + 2\alpha\|D_{h_u}^{-1}\|_2\|YY^T\|_2$. This constant can also be estimated empirically using a few random normalized matrices P. Then , initializing (for instance) $P = P^0$ as the optimal transport matrix, and taking $\tau < \frac{1}{\tilde{L}} \leqslant \frac{1}{L}$, the algorithm reads:

$$P^{k+1} = \operatorname*{Proj}_{\mathcal{P}_{h_u}} \left(P^k - \tau \left(C + \rho\nabla F(P^k) + \lambda\nabla R(P^k) + \alpha\nabla D(P^k) \right) \right),$$

A variant of the gradient descent acceleration of [9] using an inertial force has been proposed in [11]. We therefore rely on the following inertial algorithm

$$P^{k+1} = \operatorname*{Proj}_{\mathcal{P}_{h_u}} \left(P^k - \tau \left(C + \rho\nabla F(P^k) + \lambda\nabla R(P^k) + \alpha\nabla D(P^k) \right) + \beta(P^k - P^{k-1}) \right).$$

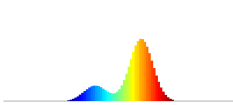
which converges to a local minima of the tackled problem by taking $\beta \in [0; 1[$ and $\tau < 2(1-\beta)/L$.

4 Experiments

4.1 Regularized 1-D histogram matching

Figure 1 illustrates the interest of the proposed approach to cancel the effect of transport dispersion. The first row shows input and target histograms, respectively in Fig. 1a and 1b. The colormap in Fig. 1a is arbitrary, and used to illustrate the optimal transport map in Fig. 1b. Observe that the mapping is highly irregular, sending neighboring colors (orange) in two distant locations, as illustrated in Fig. 1f. Figures 1c and Figures 1d, respectively proposed in [6] and [5], provides more regular mapping of *the average*

flow, as illustrated in Fig. 1g and 1h, but the colors are spread out to many different locations. The proposed model, illustrated in Fig. 1e, relax the matching constraint while controlling the variance of the mapping and the χ^2 statistics with the desired output histogram. This leads to sparser transport flow (Fig. 1i) and almost no color mixing. The matrices P of transport flows computed with the different methods are illustrated in Figures 1f to 1i. The null values P_{ij} are displayed in white color. These matrices describe how the mass from the bins of the input histogram (in rows) are sent to the bins of the target histogram (in columns). As a bin $h_u(i)$ corresponds to the line i of the matrix, the variance of this bin is larger when a lot of columns j are active (i.e $P_{ii} > 0$) and correspond to very different features Y_j.

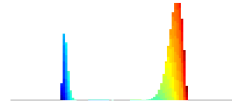


(a) Input histogram

(b) Output histogram using OT

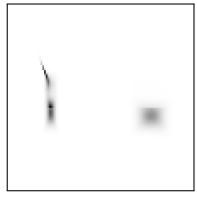
(c) Gradient-based Regularization [6]

(d) Entropy-based Regularization [5]

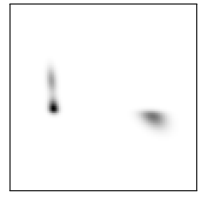


(e) Proposed relaxation model

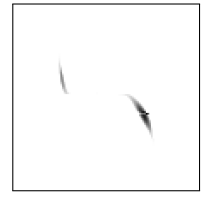
(f) OT flow



(g) Gradient regularized OT flow [6]



(h) Entropy regularized OT flow [5]



(i) Proposed model for transport flow

Figure 1: ***Illustration with 1D histograms of various models for regularized optimal transport.*** *(Please, refer to the text for details)*

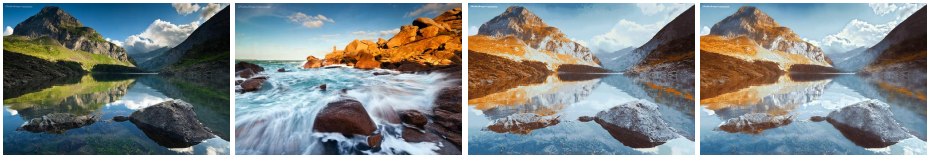
4.2 Color transfer

The first step of our color transfer process consists in defining the source and target sets X, Y, which involves spatio-color clustering on the input image u and the exemplar one v, respectively. Here, the clustering is performed using the fast super-pixels method [1], with the default regularization parameter 0.02 and a raw 20×20 seed initialization. These clusters are then used to build a weighted graph $(\omega_{i,j})$ and define the transport cost matrix C_{XY} that are involved in the minimization of the non-convex functional (9). The color transfer is finally applied using the estimated relaxed and regularized transport map.

As we work at a super-pixel scale to speed-up the OT computation, the last step of the proposed approach is to synthesize a new image w from the source image u using the new color palette. Like [25], we use maximum likelihood estimation to incorporate geometrical information from the source image u into the synthesis process. In order to restore the sharp details from the original image that may have been lost in the process,

we make use of the NLMR filter from [17]. Note that this step can be approximated and speeded up using the real time guided filter proposed by [7]. More details can be found in [19].

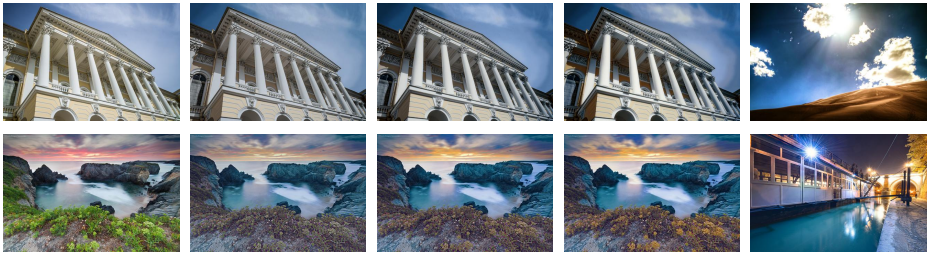
Experiments As illustrated in Figure 2, when the histograms of the two images have very different shapes, the classic OT color transfer create a lot of artifacts (Fig. 2c). The original colors are better recovered with increasing penalization of the color variance (Fig. 2e, 2f, 2g, 2h). Such property is illustrated in other examples in Figures 3 and 4. When no penalization is applied to the color variance (*i.e.* $\alpha = 0$), it corresponds to the model of [19]. By monitoring the capacity of the target histogram and regularizing the average flow [19], the synthesized images look more plausible (3b) but they contain new drab colors (that do not exist in the target image) and they are over-smoothed. On the other hand, the transfer is far better when the color variance is penalized (with high values of α). In this case, the final images only contain the colors of the target images.



(a) Input image (b) Exemplar image (c) Raw OT transfer (d) Post-process of (c)

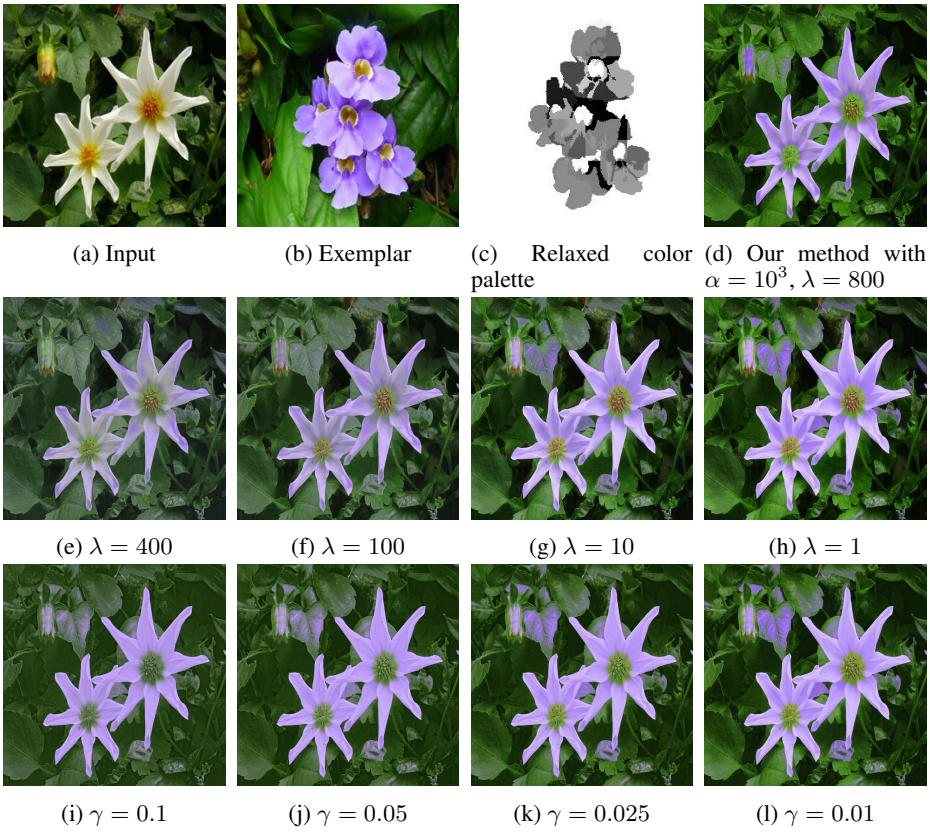
(e) $\alpha = 0$ [19] (f) $\alpha = 10$ (g) $\alpha = 30$ (h) $\alpha = 50$

Figure 2: ***Illustration of the penalization of transport variance for color transfer preservation.*** *Colors of image (a) are modified using image (b) as a template. Image (c) illustrates the result obtained from of optimal transfer, without any regularization. A post-processing (d) may applied to remove small artifacts, but large color inconsistencies still occur. The adaptive color transfer approach recently proposed in [19], which mixes capacity relaxation and spatial regularization yields better results (e). However, final colors may be washed-out due to the mixing of colors. In the proposed model, the parameter α directly controls the amount of transport dispersion, as illustrated here in (f), (g) and (h).*



(a) Input (b) $\alpha = 0$ [19] (c) $\alpha = 10$ (d) $\alpha = 100$ (e) Exemplar

Figure 3: ***Color transfer.*** *The colors of the exemplar images (column (e)) are transfered to the input images (column (a)).*



(a) Input (b) Exemplar (c) Relaxed color palette (d) Our method with $\alpha = 10^3$, $\lambda = 800$

(e) $\lambda = 400$ (f) $\lambda = 100$ (g) $\lambda = 10$ (h) $\lambda = 1$

(i) $\gamma = 0.1$ (j) $\gamma = 0.05$ (k) $\gamma = 0.025$ (l) $\gamma = 0.01$

Figure 4: ***Comparison of the proposed method with entropic regularization [5] and spatial regularization [19].*** *First row shows the obtained color transfer (d) of (a) using (b), and the relaxed color palette (c) (the darker, the less a color is used). The second row illustrates the tradeoff between spatial regularity and color fidelity when considering the method proposed in [19], depending on regularization parameter λ (i.e. with $\alpha = 0$ in (9)). In the last row, a similar tradeoff is obtained with the method of [5], when tuning the regularization parameter γ in (2).*

5 Conclusion and future work

In this paper, we have proposed a method for transferring color between images using relaxed and regularized optimal transport. Our model involves a non-convex constraint that minimizes the dispersion of the relaxed transport and prevents from creating new drab colors. Further improvements will concern the use of faster optimization tools and the incorporation of high-order moments (such as the covariances of the transfered clusters) into the final synthesis.

References

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE TPAMI*, 34(11):2274–2282, 2012.

[2] H. Attouch, J. Bolte, and B. Svaiter. Convergence of descent methods for semi-algebraic and tame problems. *Mathematical Programming*, 137(1-2):91–129, 2013.

[3] J.-D. Benamou. Numerical resolution of an "unbalanced" mass transport problem. *M2AN Math. Model. Numer. Anal.*, 37(5):851–868, 2003.

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

[5] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems (NISP'13)*, pages 2292–2300, 2013.

[6] S. Ferradans, N. Papadakis, J. Rabin, G. Peyré, and J.-F. Aujol. Regularized discrete optimal transport. In *SSVM'13*, pages 428–439, 2013.

[7] K. He, J. Sun, and X. Tang. Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6):1397–1409, 2013.

[8] J. Morovic and P.-L. Sun. Accurate 3d image colour histogram transformation. *Pattern Recognition Letters*, 24(11):1725–1735, 2003.

[9] Y. E. Nesterov and A. S. Nemirovsky. *Interior Point Polynomial Methods in Convex Programming : Theory and Algorithms*. SIAM Publishing, 1993.

[10] M. Nikolova, Y.-W. Wen, and R. H. Chan. Exact histogram specification for digital images using a variational approach. *JMIV*, 46(3):309–325, 2013.

[11] P. Ochs, Y. Chen, T. Brox, and T. Pock. ipiano: Inertial proximal algorithm for nonconvex optimization. *SIAM Journal on Imaging Sciences*, 7(2):1388–1419, 2014.

[12] N. Papadakis, A. Bugeau, and V. Caselles. Image editing with spatiograms transfer. *IEEE Trans. on Image Processing*, 21(5):2513–2522, May 2012.

[13] N. Papadakis, E. Provenzi, and V. Caselles. A variational model for histogram transfer of color images. *IEEE Trans. on Image Proc.*, 20(6):1682–1695, 2011.

[14] F. Pitié, A. C. Kokaram, and R. Dahyot. N-dimensional probability density function transfer and its application to colour transfer. In *ICCV*, pages 1434–1439, 2005.

[15] F. Pitié, A. C. Kokaram, and R. Dahyot. Automated colour grading using colour distribution transfer. *CVIU*, 107:123–137, 2007.

[16] T. Pouli and E. Reinhard. Progressive color transfer for images of arbitrary dynamic range. *Comput. Graph.*, 35(1):67–80, Feb. 2011.

[17] J. Rabin, J. Delon, and Y. Gousseau. Removing artefacts from color and contrast modifications. *IEEE Trans. On Image Processing*, 20(11):3073–3085, 2011.

[18] J. Rabin, J. Delon, and Y. Gousseau. Transportation distances on the circle. *Journal of Mathematical Imaging and Vision*, 41(1-2):147–167, 2011.

[19] J. Rabin, S. Ferradans, and N. Papadakis. Adaptive color transfer with relaxed optimal transport. In *IEEE International Conference on Image Processing*, ICIP'14, 2014.

[20] J. Rabin and G. Peyré. Wasserstein regularization of imaging problem. In *IEEE Int. Conf. on Image Processing (ICIP'11)*, ICIP'11, pages 1541–1544, 2011.

[21] J. Rabin, G. Peyré, J. Delon, and M. Bernot. Wasserstein barycenter and its application to texture mixing. In *SSVM'11*, volume 6667, pages 435–446, 2011.

[22] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley. Color transfer between images. *IEEE trans. on Computer Graphics and Apps.*, 21(5):34–41, 2001.

[23] Y. Rubner, C. Tomasi, and L. Guibas. A metric for distributions with applications to image databases. In *ICCV'98*, International Conference on Computer Vision, pages 59–66, 1998.

[24] Z. Su, K. Zeng, L. Liu, B. Li, and X. Luo. Corruptive artifacts suppression for example-based color transfer. *Multimedia, IEEE Transactions on*, 16(4):988–999, June 2014.

[25] Y.-W. Tai, J. Jia, and C.-K. Tang. Local color transfer via probabilistic segmentation by expectation-maximization. CVPR '05, pages 747–754, 2005.

[26] C. Villani. *Topics in Optimal Transportation*. American Mathematical Society, 2003.