Wasserstein Regularization of Imaging Problems

Julien Rabin, Gabriel Peyré

WASSERSTEIN REGULARIZATION OF IMAGING PROBLEMS

Julien Rabin

rabin@cmla.ens-cachan.fr

CMLA, ENS de Cachan
61, Avenue du Président Wilson
94230 Cachan, France

Gabriel Peyré

peyre@ceremade.dauphine.fr

CEREMADE, Univ. Paris-Dauphine
Place du Maréchal De Lattre De Tassigny
75775 Paris, France

ABSTRACT

This paper introduces a novel and generic framework embedding statistical constraints for variational problems. We resort to the theory of Monge-Kantorovich optimal mass transport to define penalty terms depending on statistics from images. To cope with the computation time issue of the corresponding Wasserstein distances involved in this approach, we propose an approximate variational formulation for statistics represented as point clouds.

We illustrate this framework on the problem of regularized color specification. This is achieved by combining the proposed approximate Wasserstein constraint on color statistics with a generic geometric-based regularization term in a unified variational minimization problem. We believe that this methodology may lead to some other interesting applications in image processing, such as medical imaging modification, texture synthesis, *etc*.

Index Terms— Variational model, Energy minimization, Image regularization, Gradient descent, color and contrast modification;

1. INTRODUCTION

This paper deals with the use of statistical priors for constrained functional minimization problems. More precisely, we take interest in the practical use of the Monge-Kantorovich optimal mass transportation theory [1] to enforce *statistical constraints* within a generic variational framework.

This problem arises in various image processing problems (such as medical imaging, texture synthesis, deconvolution, inpainting, *etc*), where the solution is known to follow some statistical properties (values, correlations, filter coefficients, ...). In this paper, we study the *color specification* of an image, which aims at modifying the color statistics of a given image to match some desired color distribution while preserving its geometrical information.

1.1. Previous work

Contrast and color modification This problem has received growing attention in the past few years, with various applications to image and movie enhancement such as in the movie industry, in medical or satellite imaging. Most approaches (see for instance [2, 3, 4, 5] for gray level or color modification) are based on histogram manipulation (*e.g.* histogram equalization, affine transformation, *etc*). Yet, as demonstrated in [6], these techniques often yields some unpleasant artefacts –such as noise enhancement, JPEG "bloc" effect, and loss of details–, for which [6] introduced a post-processing filter to remove them.

Variational regularization of color and contrast modification An interesting alternative to perform color modification and its regularization is to use a unified variational formulation, in which the two tasks are driven *simultaneously*. In such a variational setting, the functional to be minimized is obtained by combining several penalty terms: generic fidelity and regularization terms (*e.g.* based on the Euclidean norm, the Total Variation, the Sobolev norm, or the curvature of level set) and a supplementary term devoted to the desired contrast or color change (see *e.g.* [7, 8, 9]). As shown in these references, the intrinsic advantage of using this kind of variational framework is that it enables to preserve geometrical information of the original image and to control the regularity of the transformation so that it prevents the apparition of the aforementioned artefacts.

However, a major limitation of these techniques is the use of high-dimensional statistical constraints. Indeed, most of approaches have been proposed for one-dimensional statistics (for instance in [8] for local histogram equalization). Recently in [9], the difficulty of designing penalty term with multi-dimensional data such as color cloud is circumvented by the use of cumulated histograms.

The originality of our work is that we propose a new framework in which multi-dimensional constraints are easily embedded by making use of the Wasserstein metric.

1.2. Outline and contributions

We propose in section 2 a fully generic method to enforce statistical constraints with quadratic Wasserstein distances [1]. The problem statement is first formulated in the exact setting (§ 2.1 and 2.2), and then derived to an approximate setting (§ 2.3) for which a gradient based algorithm is designed to solve the corresponding minimization problem (§ 2.4). In Section 3, we illustrate the interest of the proposed framework for regularized color transfer between images.

2. A GENERIC FRAMEWORK FOR WASSERSTEIN CONSTRAINED VARIATIONAL PROBLEM

In this section, a general formulation of the problem is stated in § 2.1. In the next paragraph (§ 2.2) a brief overview is given on the definition and the time complexity of the quadratic Wasserstein distance when considering points clouds. An approximation of the Wasserstein distance and its numerical computation is then studied in § 2.3. Finally, a new formulation of variational problems with Wasserstein constraints and a gradient descent algorithm is proposed in § 2.4.

Research partially financed by the Centre National d'Études Spatiales (MISS project), the Office of Naval research (grant N00014-97-1-0839), by the European Research Council (advanced grant "Twelve labours"), and the French National Research Agency (ANR project NatImage).

2.1. Problem statement

Let $u : i \in \Omega \mapsto u_i = (u_R(i), u_G(i), u_B(i))^T \in \Gamma \subset \mathbb{R}^d$ be a discrete color image ($d = 3$), where $\Omega \subset \mathbb{Z}^2$ denotes the spatial domain (N-pixel grid) and Γ is the RGB color cube. For the sake of simplicity, we refer throughout the paper as $[u] := \{u_i\}_{i\in\Omega}$ the color statistics of image u, *i.e.* a points cloud with $N = |\Omega|$ points. In this paper, we aim at minimizing a large class of problems of the form

$$\min_{w\in\mathbb{R}^{d\times N}} \left\{ \mathcal{E}(w) = F(w,u) + \lambda_{\mathrm{R}}\, R(w) + \lambda_{\mathrm{S}}\, S([w],[v]) \right\} (*)$$

in which the function S is used to measure the adequacy of the statistics of image w with the given constraints $[v]$, where F is a fidelity term depending on the original image u, and where R is a generic regularization term depending only on the image w.

A solution of this problem may be found using for instance a gradient descent scheme, thus requiring the computation of the first derivative of S. In the following sections, we investigate which function S may be used in practice in the Monge-Kantorovich optimal transport framework to achieve the design of such gradient descent step.

2.2. Wasserstein distance

Definition In the discrete setting, when statistics are represented as point-clouds, the Wasserstein metric is equivalent to the optimal assignment problem [10]. The quadratic Wasserstein metric [1] between two d-dimensional N-point clouds $[u], [v] \in \mathbb{R}^{d\times N}$ is therefore the optimal permutation $\sigma \in \Sigma(\Omega)$ cost such that

$$W_2([u],[v]) = \min_{\sigma\in\Sigma(\Omega)} \left(\sum_{\ell\in\Omega} \|u_\ell - v_{\sigma(\ell)}\|^2 \right)^{\frac{1}{2}} \tag{1}$$

where $\|.\|$ is the Euclidean norm in $\mathbb{R}^d$, and where $\Sigma(\Omega)$ is the set of permutations of Ω.

Computation in the 1-D case When considering 1-D discrete distributions with N points, it is well known that the optimal assignment problem can be solved in $O(N \log N)$ operations *via* fast sorting algorithms. Indeed, the optimal assignment $\sigma^\star \in \Sigma(\Omega)$ of $[v]$ with $[u]$ minimizing (1) can be computed as:

$$\sigma^\star := \sigma_{[v]} \circ \sigma_{[u]}^{-1} \tag{2}$$

where $\sigma_{[u]}$ and $\sigma_{[v]}$ are the permutations that respectively sort the points of $[u]$ and $[v]$ in the same order.

Computation in the multi-dimensional case When considering the general case, the optimal assignment can be formulated as a linear program, which can be solved using dedicated methods (*e.g.* the Hungarian or the Auction algorithms) with at least a $O(N^{2.5}\log N)$ time complexity. Relaxation of the assignment problem makes it possible to use simplex or interior point methods, but a bi-stochastic matrix with N^2 entries has to be built, which requires large memory.

Moreover, we emphasize here that the computation of the derivative ∂W_2, as required by the energy minimization problem $(*)$, boils down to compute the optimal assignment of (1). Thus the computation of ∂W_2 has the same time complexity as W_2, which makes it practically unusable for imaging problems where N is very large.

2.3. Approximation using Sliced Wasserstein Distance

Sliced Wasserstein Definition Based on the approximation of the Wasserstein metric proposed in [11] and recently applied to shape recognition in [12], we define the *Sliced Wasserstein distance* as the sum of 1-D optimal assignment costs

$$\begin{aligned} \mathrm{SW}_2([u],[v])^2 &:= \frac{1}{|\Psi|}\sum_{\theta\in\Psi} W_2\left([u]_\theta,[v]_\theta\right)^2 \ , \qquad (3)\\ &= \frac{1}{|\Psi|}\sum_{\theta\in\Psi} \min_{\sigma_\theta\in\Sigma(\Omega)} \sum_{\ell\in L} \left|\, \langle\, u_\ell - v_{\sigma_\theta(\ell)}\,,\, \theta \,\rangle\, \right|^2 , \end{aligned}$$

where θ is a unit vector of from the set Ψ of directions sampled over the unit sphere $\mathbb{S}^{d-1}$ in $\mathbb{R}^d$, where $\langle ., \theta\rangle$ is the Euclidean scalar product according to direction θ. We denote by $\sigma_\theta^\star$ the optimal assignment of the 1-D distribution $[u]_\theta = \{\ \langle\, u_\ell\,,\,\theta\,\rangle\ \}_{\ell\in\Omega}$ with $[v]_\theta = \{\ \langle\, v_\ell\,,\,\theta\,\rangle\ \}_{\ell\in\Omega}$ which is computed using Eq. (2). The SW_2 distance can thus be estimated faster than W_2 in $O(|\Psi| N \log N)$ operations, so as its gradient which is defined in the next paragraph.

Sliced Wasserstein Gradient It can be shown[1] under mild assumptions that the first derivative of SW_2 exists and is easy to compute.

Proposition 1 Let $[u], [v]$ be two discrete distributions of $\mathbb{R}^{d\times N}$ and assume that points $\{u_\ell \in \mathbb{R}^d\}_{\ell\in\Omega}$ are pairwise different, i.e. $u_\ell \neq u_m\, , \forall\, (\ell, m) \in \Omega\times\Omega \setminus \{\ell\}$. *Let $\Psi \in \mathbb{S}^{d-1}$ be a discrete set of orientations s.t. $\forall\,\theta \in \Psi$, $\langle\,\theta\,,\,u_\ell - u_m\,\rangle \neq 0\, \forall\,(\ell,m) \in \Omega\times\Omega \setminus \{\ell\}$. Then, the derivative of functional $\mathrm{SW}_2([u],[v])$ according to the point of $[u]$ with index ℓ exists. Moreover, once the 1-D optimal assignments $\{\sigma_\theta^\star\}_{\theta\in\Psi}$ have been computed, the gradient of SW_2^2 is defined as follows, $\forall\,\ell\in\Omega$*

$$\frac{\partial \mathrm{SW}_2([u],[v])^2}{\partial u_\ell} := \frac{2}{|\Psi|}\sum_{\theta\in\Psi} \langle\, u_\ell - v_{\sigma_\theta^\star(\ell)}\,,\,\theta\,\rangle\,\theta\ . \tag{4}$$

2.4. Proposed Solution to the Wasserstein Constrained Problem (*) for Regularized Color Transfer

The Sliced Wasserstein Regularization Problem Using the proposed Sliced Wasserstein distance and its derivative formulas (3) and (4), we may rewrite problem $(*)$ as:

$$\min_{w\in\mathbb{R}^{d\times N}} \left\{ \mathcal{E}(w) = F(w,u) + \lambda_{\mathrm{R}}\, R(w) + \frac{\lambda_{\mathrm{S}}}{2}\, \mathrm{SW}_2([w],[v])^2 \right\} (\star)$$

where the quadratic Sliced Wasserstein distance is used as a statistical fidelity term.

Here we are interested in color transfer application, for which we choose to restrict our attention to the following penalty terms:

- a *fidelity term* F defined as the sum of the quadratic loss and a level set consistency term already considered in [7, 9]

$$F(w,u) = \sum_{i\in\Omega} \left\{ \frac{\lambda_L}{2}\,\|w_i - u_i\|^2 - \lambda_{LS}\,\langle\,\nabla w_i\,,\,\frac{\nabla u_i}{\|\nabla u_i\|}\,\rangle \right\}$$

- a *regularization term* R defined as the color Total Variation [13] penalty (TV):

$$R(w) = \|w\|_{\mathrm{TV}} = \sum_{i\in\Omega} \|\nabla w_i\|\, ,$$

[1]The proof is omitted here due to the lack of space.

where the first derivative operator ∇ is defined for color image as:

$$\nabla u_i = \left((\nabla u_R(i))^T, (\nabla u_G(i))^T, (\nabla u_B(i))^T \right)^T \in \mathbb{R}^6 \ .$$

Forward-Backward Proximal Iterations To solve Problem ($\star$), which is a non-convex minimization problem, we use a *forward-backward proximal* scheme to converge to a local minimum of energy $\mathcal{E}$.

Starting from $w^{(0)} := u$, the update of the image $w^{(k)}$ at iteration k and point of coordinate $i \in \Omega$ depends on the two following Forward (F) and Backward (B) steps:

$$\begin{cases} w_i^{(k+\frac{1}{2})} = & w_i^{(k)} - \tau \left(F'(w^{(k)}, u)(i) + \lambda_{\mathrm{S}} \frac{\partial \mathrm{SW}_2(w^{(k)}, [v])^2}{\partial w_i^{(k)}} \right) \text{(F)} \\ w_i^{(k+1)} = & \operatorname{prox}_{\tau \cdot \lambda_R R} \left(w^{(k+\frac{1}{2})} \right)(i) \qquad \text{(B)} \end{cases}$$

where in our case the gradient of the fidelity term F is defined as

$$F'(w^{(k)}, u)(i) = \lambda_L (w_i^{(k)} - u_i) + \lambda_{LS} \operatorname{div} \frac{\nabla u_i}{\|\nabla u_i\|} \ ,$$

where the gradient of the sliced Wasserstein distance is computed using Proposition 1, and eventually the proximal operator of the backward step is a ROF denoising [13] which is computed using Chambolle's iterative scheme [14] recalled in the next paragraph. Note that the divergence operator div is here applied on 6-dimensional vector-valued images.

Computation of the proximal operator The proximal operator in (B), when defined as ROF denoising [13], may be computed with

$$\operatorname{prox}_{\lambda \|\cdot\|_{\mathrm{TV}}}(w) = \underset{z \in \mathbb{R}^{d \times N}}{\operatorname{argmin}} \frac{1}{2} \|z - w\|^2 + \lambda \|z\|_{TV} = w + \nabla^*(v^*),$$

where $\nabla^* = -\operatorname{div}$ is the dual operator of the discrete gradient ∇ and v^* is the vector field solution in $\mathbb{R}^{2d \times N}$ of the following dual constrained problem [14]:

$$v^* \in \underset{v \in \mathbb{R}^{2d \times N}, \|v\|_\infty \leq \lambda}{\operatorname{argmin}} \|f - \operatorname{div} v\|^2 \text{ with } \|v\|_\infty = \max_{i \in \Omega} \|v_i\| \ .$$

The solution can be computed using a gradient descent

$$v^{(t+1)} = \operatorname{Proj}_{\|\cdot\|_\infty \leq \lambda} \left\{ v^{(t)} + \rho \nabla (\operatorname{div}(v^{(\ell)}) - f) \right\} \ , \qquad (5)$$

with $\rho \leq \frac{2}{\|\operatorname{div} \circ \nabla\|} = \frac{1}{4d}$ and $\operatorname{Proj}_{\|\cdot\|_\infty \leq \lambda} : v_i \mapsto \frac{v_i}{\|v_i\|} \min(\|v_i\|, \lambda)$.

3. APPLICATION TO SIMULTANEOUS COLOR TRANSFER AND REGULARIZATION BETWEEN IMAGES

This section is devoted to the experimental study of the proposed gradient descent algorithm for color transfer between images.

Experimental Settings The desired color statistics $[v]$ are defined here by the color cloud of an image v chosen by the user. For all experiments, the algorithm is run with $\rho = \frac{1}{12}$ and $\tau = 10^{-1}$, using the following set of parameters: number of directions for the sliced Wasserstein distance $|\Psi| = 10$ in Formula (4) and the weights of the color transfer energy ($\star$) $\lambda_L = 0.1$, $\lambda_R = \lambda_{LS} = 0.5$ and $\lambda_W = 1$.

Results Figure 1 exhibits several color transfer examples with the proposed approach. To show the role of the fidelity and the penalty terms in the energy ($\star$), we also display the results of the "raw color transfer" obtained with the same gradient descent algorithm –with identical direction set Ψ– where $\lambda_F = \lambda_R = 0$ (*i.e.* without regularization).

As expected, the regularized color transfer approach enables to considerably reduced the presence of artefacts (compression blocks, noise enhancement, details reduction) which arises when directly matching the statistics (the raw color transfer). This application also demonstrates that the proposed use of the Wasserstein constraints within a variational framework yields for color transfer as good results as for contrast modification (*i.e.* with only 1-D constraint), but with higher statistical constraints.

4. CONCLUSION AND FUTURE WORK

In this paper, a new framework has been introduced to combine classical variational-based regularization techniques with high-dimensional statistical constraints. Such approach is likely to find some other applications in image processing and computer vision (*e.g.* medical imaging, texture synthesis).

Several extensions of this work are foreseen, such as the use of more complex statistical features than color (cross-correlation, patches, ...) and the regularization of image sequence.

5. REFERENCES

[1] C. Villani, *Topics in optimal transportation*, American Math. Soc., 2003.

[2] J. Morovic and P.L. Sun, "Accurate 3d image colour histogram transformation," *Pat. Recog. Letters*, vol. 24, no. 11, pp. 1725–1735, July 2003.

[3] A. J. McCollum and W. F. Clocksin, "Multidimensional histogram equalization and modification," in *CIAP*, 2007, pp. 659–664.

[4] F. Pitié, A. C. Kokaram, and R. Dahyot, "Automated colour grading using colour distribution transfer," *Comput. Vis. Image Underst.*, vol. 107, pp. 123–137, 2007.

[5] F. Pitié and A. Kokaram, "The Linear Monge-Kantorovitch Colour Mapping for Example-Based Colour Transfer," in *Proc. of CVMP'06*, 2006.

[6] J. Rabin, J. Delon, and Y. Gousseau, "Regularization of Transportation Maps for Color and Contrast Transfer," in *Proc. of the IEEE Int. Conf. on Image Processing (ICIP)*, 2010.

[7] C. Ballester, V. Caselles, L. Igual, J. Verdera, and B. Rougé, "A variational model for p+xs image fusion," *Int. J. Comput. Vision*, vol. 69, no. 1, pp. 43–58, 2006.

[8] M. Bertalmío, V. Caselles, E. Provenzi, and A. Rizzi, "Perceptual color correction through variational techniques," *IEEE Transactions on Image Processing*, vol. 16, no. 4, pp. 1058–1072, 2007.

[9] N. Papadakis, E. Provenzi, and V. Caselles, "A variational model for histogram transfer of color images," *IEEE Transaction on Image Processing*, vol. 19, no. 12, November 2010.

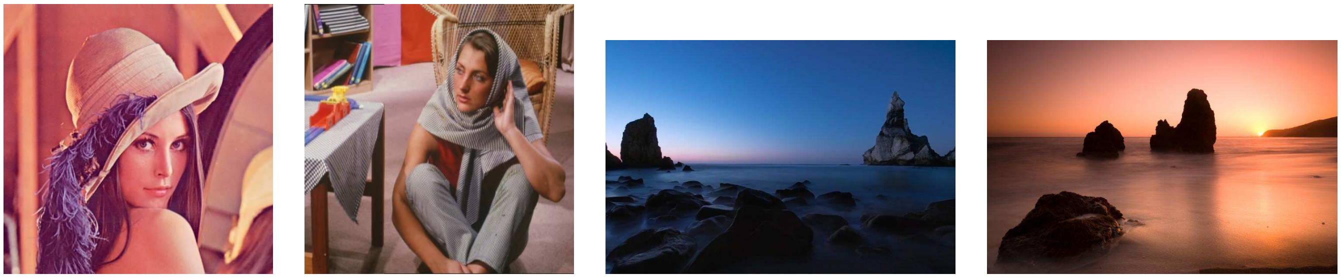
[10] R. Burkard, M. Dell'Amico, and S. Martello, *Assignment Problems*, SIAM, 2009.

[11] J. Rabin, G. Peyré, J. Delon, and M. Bernot, "Wasserstein Barycenter and its Application to Texture Mixing," in *Proc. of the Int. Conf. on Scale Space and Variational Methods in Computer Vision*, 2011.

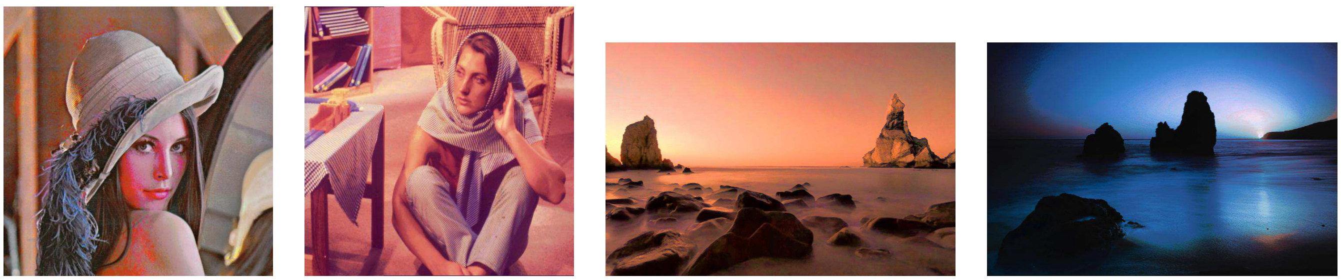
[12] J. Rabin, G. Peyré, and L. D. Cohen, "Geodesic Shape Retrieval via Optimal Mass Transport," in *Proc. ECCV'10*, 2010, vol. V, pp. 771–784.

[13] L.I. Rudin, S.J. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

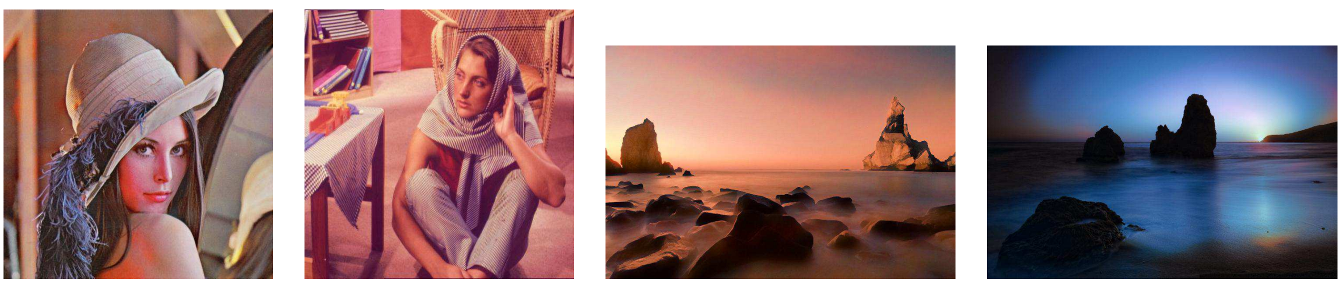
[14] Antonin Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imaging Vis.*, vol. 20, pp. 89–97, January 2004.



(a) Original image (b) Desired color distribution (c) Original image (d) Desired color distribution



(e) Raw color transfer (f) Raw color transfer (g) Raw color transfer (h) Raw color transfer



(i) Regularized color transfer (j) Regularized color transfer (k) Regularized color transfer (l) Regularized color transfer

(m) Original image (n) Desired color distribution (o) Raw color transfer (p) Regularized color transfer

(q) Original image (r) Desired color distribution (s) Raw color transfer (t) Regularized color transfer

Fig. 1. **Illustration of color transfer using a variational framework with Wasserstein constraint.**
(For a good visualization of the images, the reader is encouraged to use the electronic version of this paper)