Optimal transport and Wasserstein barycenters in computer vision, image and video processing.

Julie Delon
CNRS, LTCI, Télécom PARISTECH
MAP5, Université Paris Descartes

Workshop *Connections for Women on Optimal Transport*, Berkeley
August 22, 2013

Optimal transport and vision - Teaser

Earth Mover's Distance (EMD) between distributions.

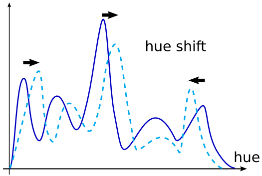
Popularized by Rubner, Guibas and Tomasi at the end of the 90's for image and texture retrieval [Rubner *et al.*, IJCV, 2000].

Very popular measure of similarity in computer vision :

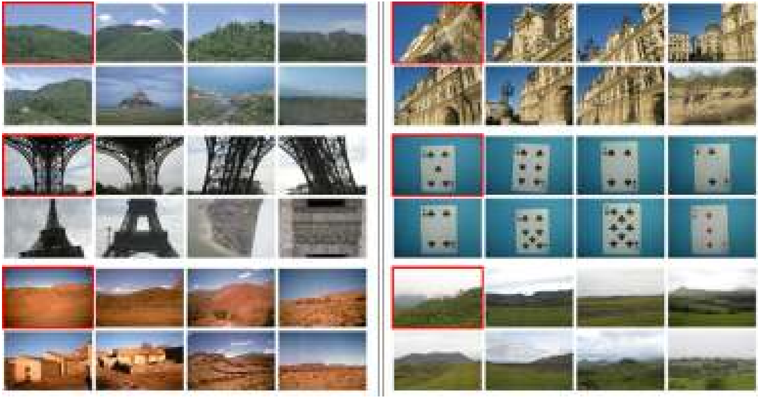
- Image retrieval [Rubner et al. 2000], [Hurtut et al. 2008], [Rabin et al. 2010]
- Object recognition, comparison of local descriptors [Ling et al. 2006], [Pele et al., 2008], [Rabin et al. 2008]
- Image registration and morphing [Haker et al. 2004], [Zhu et al. 2007]
- Junctions detection [Ruzon et al. 2001]

Reasons of popularity

- Robustness to small shifts in histograms
- Robustness to quantization



Optimal transport and vision - Teaser



[Hurtut, Gousseau, Schmitt], *Analyse et recherche d'oeuvres d'art 2D selon le contenu pictural*, 2007.

Optimal transport and image processing - Teaser

Matching distributions between images

- Contrast and color transfer [Delon 2004], [Pitie *et al.*, 2007], [Papadakis *et al.*, 2010], [Rabin *et al.*, 2011]
- Movie restoration [Delon, Desolneux, 2010]
- Texture synthesis and interpolation [Rabin *et al.*, 2011], [Ferrandans *et al.*, 2013].

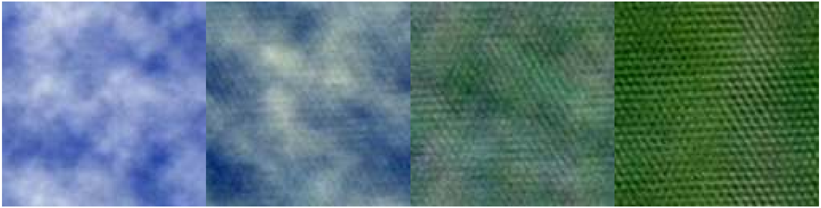
Optimal transport and image processing - Teaser

Matching distributions between images

- Contrast and color transfer [Delon 2004], [Pitie *et al.*, 2007], [Papadakis *et al.*, 2010], [Rabin *et al.*, 2011]
- Movie restoration [Delon, Desolneux, 2010]
- Texture synthesis and interpolation [Rabin *et al.*, 2011], [Ferrandans *et al.*, 2013].

Optimal transport and image processing - Teaser

Or computing average distributions between images



Images from [Ferrandans *et al.*, 2013]

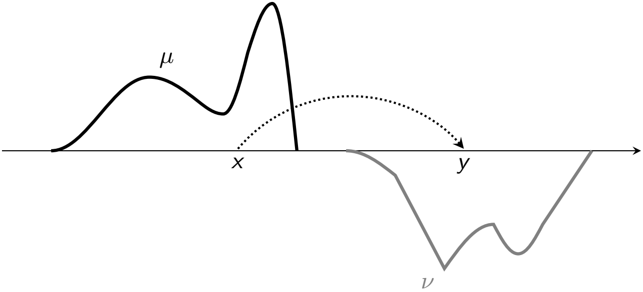
- Contrast and color transfer [Delon 2004], [Pitie *et al.*, 2007], [Papadakis *et al.*, 2010], [Rabin *et al.*, 2011]
- Movie restoration [Delon, Desolneux, 2010]
- Texture synthesis and interpolation [Rabin *et al.*, 2011], [Ferrandans *et al.*, 2013].

Notations

Two discrete probability measures μ and ν compactly supported, defined over $\mathcal{X}$. $c(x, y)$ cost of moving one unit mass from x to y.

$$\mathrm{MK}_c(\mu, \nu) = \inf_{\gamma \in \Pi(\mu,\nu)} \iint_{\mathcal{X}\times\mathcal{X}} c(x, y) d\gamma(x, y),$$

where $\Pi(\mu, \nu)$ is the set of all transport plans between μ and ν (probability measures on $\mathcal{X} \times \mathcal{X}$ with marginals μ and ν).



Fathers of optimal transport

- G. Monge, *Mémoire sur la théorie des déblais et des remblais*, 1781.
- L. Kantorovich, *On the transfer of masses*, 1942.

Notations - Discrete measures

$$\mu = \sum_{i=1}^{M} s_i \delta_{p_i} \quad \text{and} \quad \nu = \sum_{j=1}^{N} d_j \delta_{q_j}$$

Transport plans γ in $\Pi(\mu, \nu)$ can be written $\gamma = \sum_{i,j} \gamma_{ij} \delta_{(p_i, q_j)}$, where the matrix $(\gamma_{ij})_{1 \le i \le M,\ 1 \le j \le N}$ satisfies the constraints

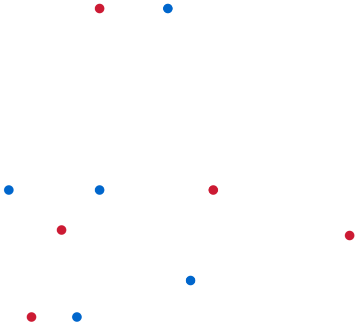
$$\gamma_{ij} \geq 0, \quad \sum_{j=1}^{M} \gamma_{ij} = s_i, \quad \sum_{i=1}^{N} \gamma_{ij} = d_j.$$

Linear Programming

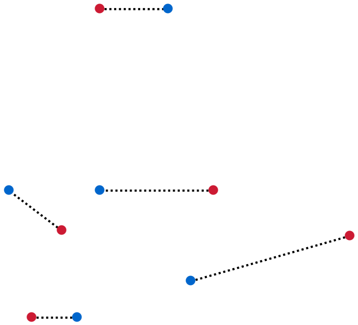
$$\mathrm{MK}_c(\mu, \nu) = \inf_{\gamma \in \Pi(\mu, \nu)} \sum_{i,j} c(p_i, q_j) \gamma_{ij}.$$

Unitary case : all masses s_i and d_j are equal $\rightarrow$ optimal assignment.

Notations - Discrete measures



Notations - Discrete measures

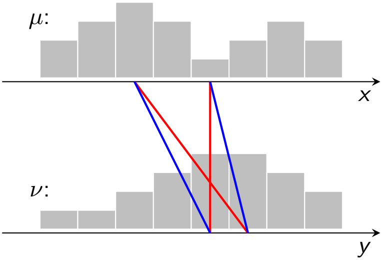


Part I

Applications of optimal transport in 1D: the line and the circle

The basic setting of transport optimization in 1D

Assume that c is a strictly convex and increasing function of the distance $|x-y|$ on $\mathbb{R}$.



If $x_1 < x_2$ and $y_1 < y_2$, $c(x_1, y_1) + c(x_2, y_2) < c(x_1, y_2) + c(x_2, y_1)$.
Thus, any optimal plan preserves the ordering of the points.

The solution to the transportation problem is given by the monotone rearrangement of μ onto ν.

M and N bins: $\quad O(M+N)$ operations

The basic setting of transport optimization in 1D

Cumulative distribution function of μ and its generalized inverse

$$F_\mu(t) = \mu[(-\infty, t]], \quad F_\mu^{-1}(v) = \inf\{t \in \mathbb{R};\ F_\mu(t) > v\},\ v \in [0,1].$$

Theorem *If c is a convex and increasing function of the distance $|x-y|$ on $\mathbb{R}$, a solution to the transportation problem is given by the monotone rearrangement of μ onto ν and*

$$\mathrm{MK}_c(\mu,\nu) = \int_0^1 c(F_\mu^{-1}(v), F_\nu^{-1}(v))dv.$$

Application to histogram specification

Notations:

- $u : \Omega \to \mathbb{R}$ a discrete image, $\Omega \subset \mathbb{Z}^2$ is bounded.
- $P = \{p_1, \dots, p_N\} \subset \mathbb{R}$ set of values possibly taken by $u(x)$. In practice, $P = \{0, \dots, 255\}$.

Histogram (grey level distribution) of u

$$h_u = \sum_{i=1}^{N} h_i \delta_{p_i}, \text{ where } h_i = \frac{1}{|\Omega|} |\{x \in \Omega;\ u(x) = p_i\}|.$$

H_u = cumulative distribution function of h_u.
For $x \in \Omega$, $H_u(u(x))$ is the rank of x in u, when all grey levels are ordered increasingly.

Application to histogram specification

Aim : an image u and a distribution μ being given, find a contrast change (*i.e.* an increasing function) g such that $h_{g(u)}$ is *as close as possible* to μ.

In the sense of optimal transport for a quadratic cost $c(x, y) = |x - y|^2$, it means that we want to minimize

$$\int_0^1 |H_{g(u)}^{-1}(t) - F_\mu^{-1}(t)|^2 dt,$$

The solution is given by

$$g(p_i) = \frac{1}{h_i} \int_{H_u[p_i]}^{H_u[p_{i+1}]} F_\mu^{-1}(t) dt.$$

- The values of g outside of $\{p_1, \dots, p_N\}$ have no incidence on the result.
- Solution generally proposed in the image processing literature :

$$g = F_\mu^{-1} \circ H_u.$$

Application to histogram specification

Histogram equalization obtained for μ uniform on $[0, 255]$

Application to histogram specification

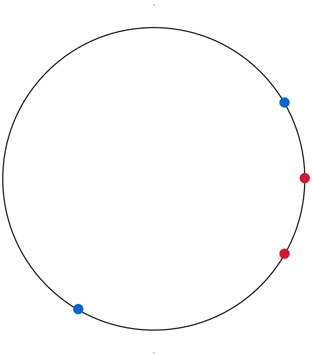
μ chosen as the grey level distribution of a second image.

Application to histogram specification

μ chosen as the grey level distribution of a second image.

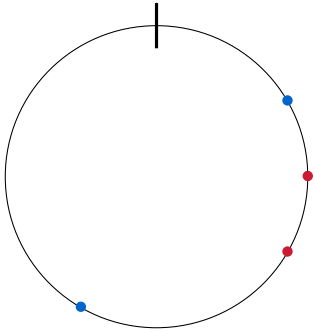
On the circle, convex cost

What happens on the circle ?



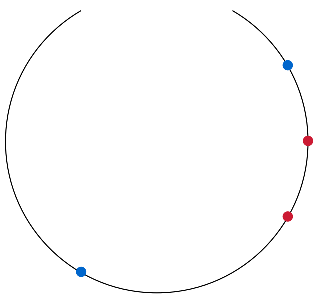
On the circle, convex cost

What happens on the circle ?



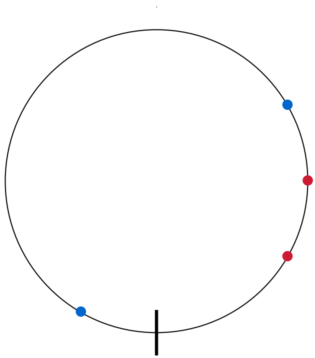
On the circle, convex cost

What happens on the circle ?



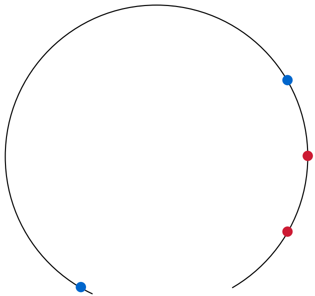
On the circle, convex cost

What happens on the circle ?



On the circle, convex cost

What happens on the circle ?



On the circle

[Rabin, Delon, Gousseau, 2011], [Delon, Salomon, Sobolesvskiĭ 2010]

Distance along the circle, points represented by coordinates on $[0,1)$:

$$d(x,y) = \min(|x-y|, 1-|x-y|).$$

Cumulative distribution functions defined on $[0,1)$ and extended to $\mathbb{R}$ by $F_\mu(y+1) = F_\mu(y) + 1$.

Theorem

Assume that $c = h(d)$, with h positive, increasing and convex. The optimal cost between μ and ν equals

$$\mathrm{MK}_c(\mu,\nu) = \inf_{\theta\in\mathbb{R}} \int_0^1 c(F_\mu^{-1}(v), (F_\nu^\theta)^{-1}(v))dv, \tag{1}$$

where $F_\nu^\theta = F_\nu - \theta$.

Complexity $O((M+N)\log\frac{1}{\varepsilon})$ to compute the infimum with a precision ε.

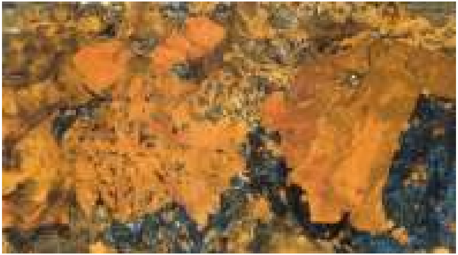
Application 1 : hue transfer between images

Using the optimal mapping to transfer the hue distribution from one image to the other

Whirlwind, Malyavin and *Jeunes filles au bord de la mer*, Puvis de Chavannes.

Application 1 : hue transfer between images

Using the optimal mapping to transfer the hue distribution from one image to the other

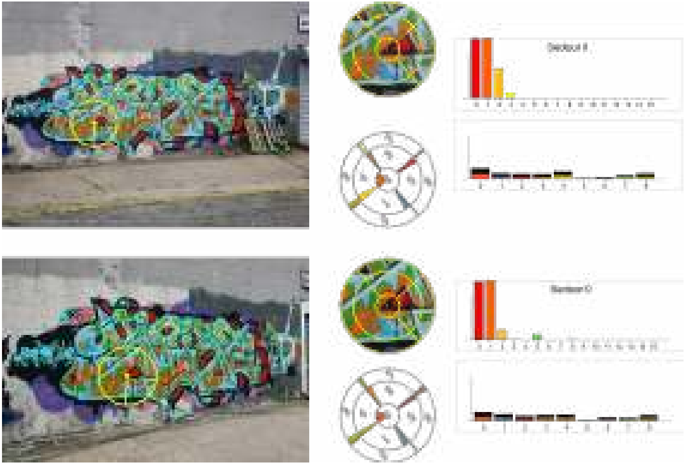


Whirlwind, Malyavin and *Jeunes filles au bord de la mer*, Puvis de Chavannes.

Application 2: matching of local descriptors

Geometry: SIFT-like descriptors [Lowe, 2004], [Rabin et al., 2009], composed of M circular histograms of gradient orientation, extracted from a localization grid.

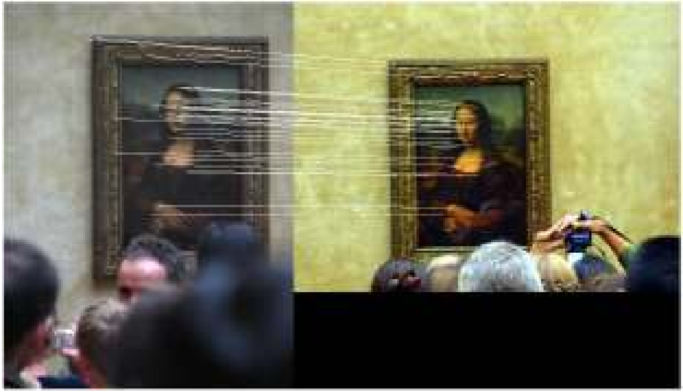
Color: similar descriptors with additional circular histograms of hue [Mazin, Delon, Gousseau, 2011].



Application 2: matching of local descriptors

[Mazin *et al.*, 2011]

Application 2: matching of local descriptors



[Mazin *et al.*, 2011]

Part II

Computing optimal transport in dimension ≥ 2

Computing optimal transport in dimension ≥ 2

No analytic formulation for $\mathrm{MK}_c(\mu, \nu)$ in dimension ≥ 2.

How can we match or interpolate distributions in this case ?



Some applications:

- Color transfer
- Feature matching
- Texture synthesis and mixing

Computing optimal transport in dimension ≥ 2

Solve the **linear program**

$$\min_{\gamma \in \Gamma} \sum_{i,j} c(p_i, q_j)\gamma_{ij},$$

where

$$\Gamma = \{\gamma,\ \gamma_{ij} \geq 0, \quad \sum_{j=1}^{M} \gamma_{ij} = s_i, \quad \sum_{i=1}^{N} \gamma_{ij} = d_j.\}$$

Standard algorithms for linear programming

- Simplex [Dantzig 1998] exponential complexity in the worst cases.
- Interior point algorithms [Karmarkar 1984] : polynomial.

Not tractable as soon as M and N increase, the number of variables and constraints being too large.

Computing optimal transport in dimension ≥ 2

Assignment algorithms

- Hungarian algorithm [Kuhn 1995]. $O(N^3)$.
- Auction algorithm [Bertsekas 1992], which solves the dual problem

$$\max_{v \in \mathbb{R}^N} \sum_{j=1}^{N} v_j + \sum_{i=1}^{N} \min_j (c(p_i, q_j) - v_j).$$

Complexity $O(N^{5/2} \log(N.C))$, where $C = \max_{i,j}(c(p_i, q_j))$. Used in cosmology. Still too greedy in its non sparse version to be used easily in computer vision.

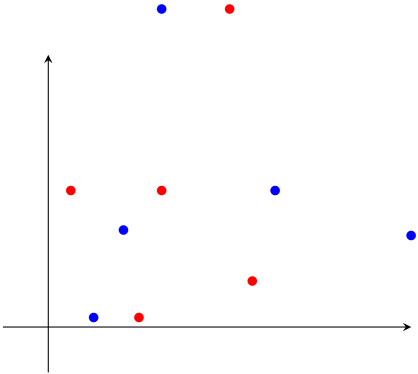
Approximation algorithms

- approximation using a wavelet decomposition [Shirdonkar *et al.*, 2008]
- embedding in a L^1 space [Indyk *et al.*, 2003], [Grauman *et al.*2004]
- well chosen ground distances [Ling *et al.*2007], [Pele *et al.*2009]

Approaches using fluid mechanics tools [Brenier, Benamou, 2000] ; [Papadakis, Peyré, Oudet, 2013]

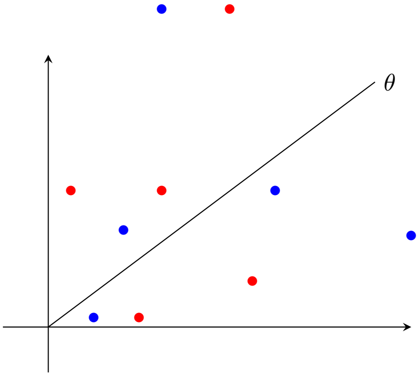
Sliced transportation for assignement [Rabin *et al.*, 2011]

$\mathcal{P}_N(\mathbb{R}^d)$ set of discrete probability measures on $\mathbb{R}^d$ made of N Dirac masses with equal weights.



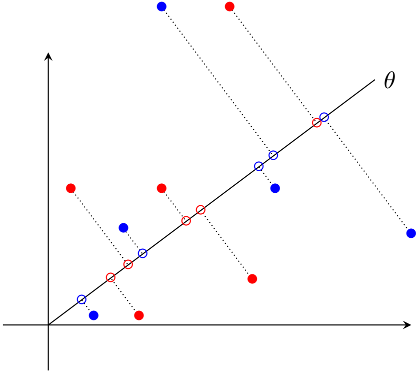
Sliced transportation for assignement [Rabin *et al.*, 2011]

$\mathcal{P}_N(\mathbb{R}^d)$ set of discrete probability measures on $\mathbb{R}^d$ made of N Dirac masses with equal weights.



Sliced transportation for assignement [Rabin *et al.*, 2011]

$\mathcal{P}_N(\mathbb{R}^d)$ set of discrete probability measures on $\mathbb{R}^d$ made of N Dirac masses with equal weights.



Sliced transportation for assignement [Rabin *et al.*, 2011]

$\mathcal{P}_N(\mathbb{R}^d)$ set of discrete probability measures on $\mathbb{R}^d$ made of N Dirac masses with equal weights.

Idea: Replace the classical transportation cost $\mathrm{MK}_c(\mu,\nu)$ by

$$\mathrm{SMK}(\mu,\nu) := \int_{\|\theta\|=1} MK_c\left((p_\theta)_\#\mu, (p_\theta)_\#\nu\right) d\theta,$$

where p_θ is the projection on the direction θ and $(p_\theta)_\#\mu$ is defined as

$$(p_\theta)_\#\mu = \frac{1}{N}\sum_{i=1}^{N} \delta_{<p_i,\theta>\cdot\theta}.$$

- $\sqrt{\mathrm{SMK}}$ is a distance over $\mathcal{P}_N(\mathbb{R}^d)$.
- The solution for each θ is given by the monotone rearrangement of $\{< p_i, \theta > \cdot\theta\}$ onto $\{< q_j, \theta > \cdot\theta\}$. Call σ_θ the corresponding optimal permutation.

Sliced transportation for assignement [Rabin *et al.*, 2011]

$\mathcal{P}_N(\mathbb{R}^d)$ set of discrete probability measures on $\mathbb{R}^d$ made of N Dirac masses with equal weights.

Idea: Replace the classical transportation cost $\mathrm{MK}_c(\mu, \nu)$ by

$$\mathrm{SMK}(\mu, \nu) = \int_{\|\theta\|=1} \sum_{i=1}^{N} \| < p_i - q_{\sigma_\theta(i)}, \theta > \|^2 d\theta.$$

- $\sqrt{\mathrm{SMK}}$ is a distance over $\mathcal{P}_N(\mathbb{R}^d)$.
- The solution for each θ is given by the monotone rearrangement of $\{< p_i, \theta > \cdot\theta\}$ onto $\{< q_j, \theta > \cdot\theta\}$. Call σ_θ the corresponding optimal permutation.

Sliced transportation for assignment [Rabin *et al.*, 2011]

[Rabin, Peyré, Delon, Bernot 2011]

$$X = (X_1, \dots, X_N) \in (\mathbb{R}^d)^N \longrightarrow \mu_X = \frac{1}{N} \sum \delta_{X_i}.$$

Gradient descent applied to

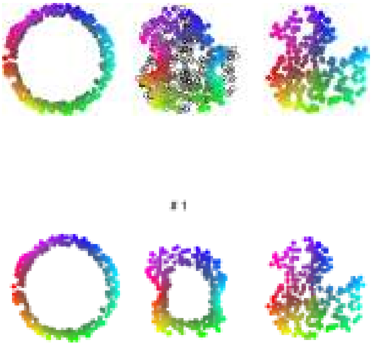
$$E_{SMK}(X) = \mathrm{SMK}(\mu_X, \nu),$$

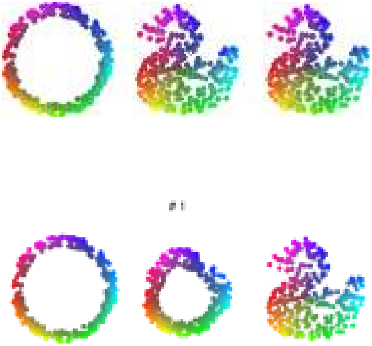
- E_{SMK} continuous, piecewise C^1,
- $E_{SMK}(X) = 0$ iff $\mu_X = \nu$,
- Only global minimizers ?
- In practice, convergence towards $X^{(\infty)} \sim \nu$. Starting from $X^{(0)}$ such that $\mu_{X^{(0)}} = \mu$, we obtain an assignment between μ and ν.
- Assignments obtained are generally "close" to the optimal assignment.

In practice: Set Θ of directions uniformly distributed over S^{d-1} and $E_{SMK_\Theta} : X \to \sum_{\theta \in \Theta} \sum_{i=1}^{N} \| < X_i - q_{\sigma_\theta(i)}, \theta > \|^2$.

Results with gradient descent

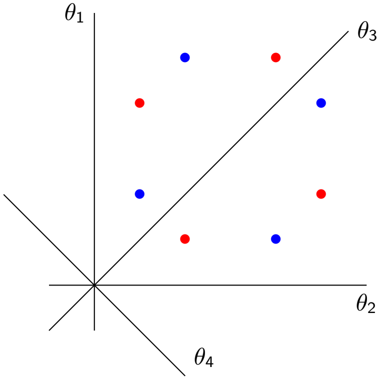
Projection results with respectively $|\Theta| = 2d$ and $|\Theta| = 100d$





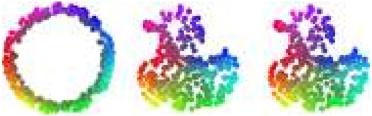
A stochastic gradient descent for the assignment problem

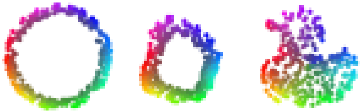
The set Θ must be chosen with caution.



A stochastic gradient descent for the assignment problem

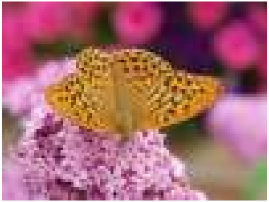
At each step k of the gradient descent, draw Θ_k at random. $|\Theta_k| = 10d$.



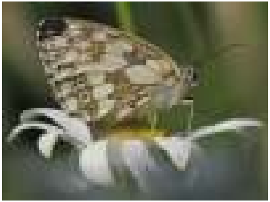


Application to color transfer

Two images u and v, apply the previous gradient descent to their color distributions.



Source image



Style image

Result

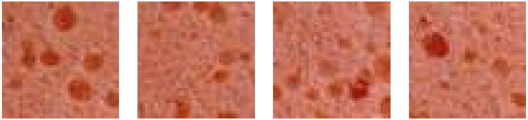
In practice, some sort of regularization is necessary.

Application to texture synthesis

u a color texture exemplar $u : \Omega \to \mathbb{R}^3$.

Goal: Generate a new random texture with the same visual aspect.

$\mapsto$

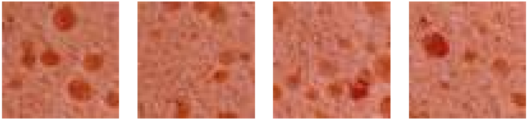


Application to texture synthesis

u a color texture exemplar $u : \Omega \to \mathbb{R}^3$.

Goal: Generate a new random texture with the same visual aspect.

$\mapsto$



Inspired by Heeger and Bergen algorithm: Texture synthesis through iterated projections on statistical sets.

Sketch of the algorithm: [Rabin *et al.*, SSVM'2011]

1– The texture exemplar u image is analyzed *via* its projection on a set of atoms (distribution of wavelet coefficients).

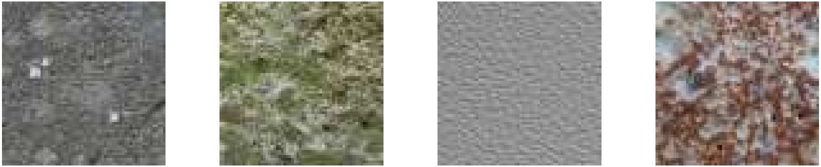
2– A random image is generated and analyzed, and its statistics are modified to match the ones of u thanks to the previous gradient descent.

3– The texture is synthesized by reconstruction

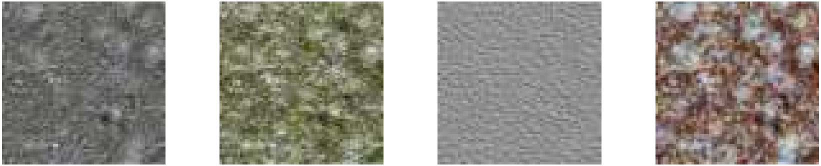
Application to texture synthesis

This approach succeeds to synthesize "micro-textures"

Exemplar textures



Synthesis



Part III

Kantorovich Barycenters

Kantorovich Barycenter

Kantorovich barycenter of L discrete distributions $(\nu^1, \dots \nu^L)$ in $\mathcal{P}_N(\mathbb{R}^d)$, for the positive weights $(\rho_1, \dots, \rho_L)$ s.t. $\sum \rho_l = 1$.

$$\inf_{\mu} \sum_{l=1}^{L} \rho_l \mathrm{MK}(\mu, \nu^l) \tag{2}$$

among all probability measures μ on $\mathbb{R}^d$.

Remarks:

- The minimizers of (2) do not always belong to $\mathcal{P}_N(\mathbb{R}^d)$. They can have N^L atoms.
- A minimizer can be found by linear programming, but the number of variables and constraints is prohibitive for real applications.
- In practice, for coherence reasons, we have to constrain the solution to be in $\mathcal{P}_N(\mathbb{R}^d)$. The problem can be recast as a **NP-hard** integer program...

Kantorovich Barycenter - Easy cases

Two cases with (semi-) explicit solutions:

- 1D case
- Gaussian case

1D case

Midway cumulative histogram between distributions $\nu^1, \dots \nu^L$ for weights $\rho_1, \dots \rho_L$

$$H_{midway} = \left(\sum \rho_i H_i^{-1} \right)^{-1}.$$

Application to image comparison

Application to image comparison

Application to image comparison

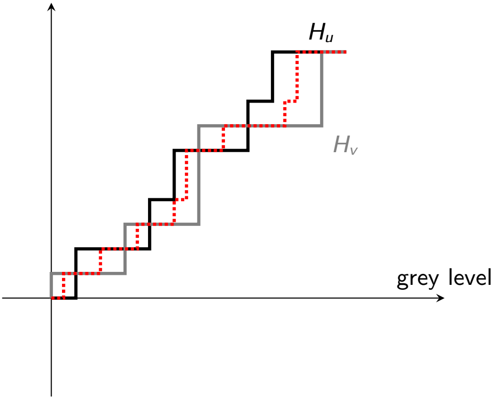
Application to image comparison

Midway histogram between h_u and h_v obtained for $L = 2$ and $\rho_1 = \rho_2 = \frac{1}{2}$:

$$H_{midway} = \left(\frac{H_u^{-1} + H_v^{-1}}{2} \right)^{-1}.$$

Midway equalization between u and v, consists in applying the contrast changes

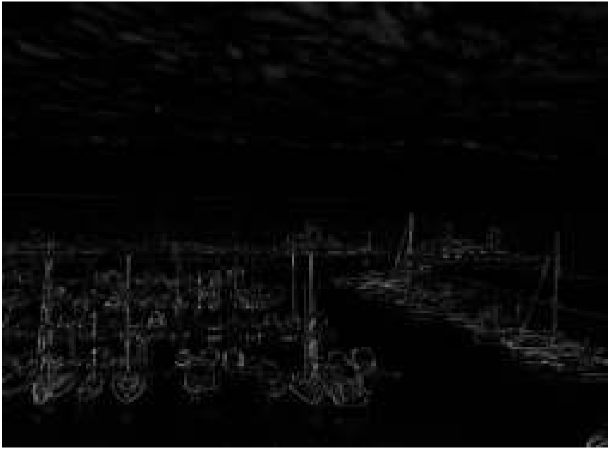
$$g_u = H_{midway}^{-1} \circ H_u \quad \text{and} \quad g_v = H_{midway}^{-1} \circ H_v$$



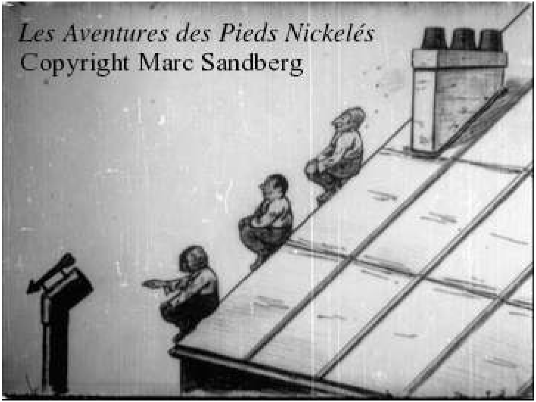
Application to image comparison

Application to image comparison

Application to image comparison



Flicker reduction



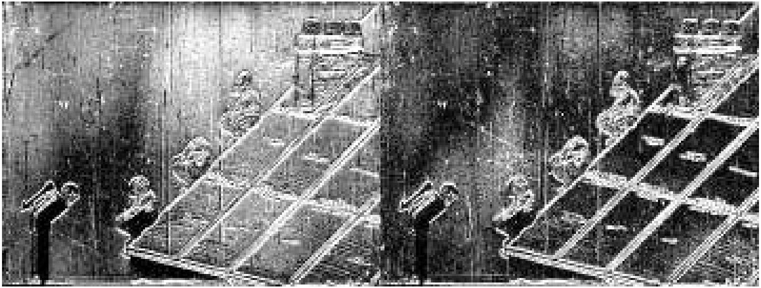
Les Aventures des Pieds Nickelés, Emile Cohl/Eclair, 1917-1918 (copyright: Marc Sandberg).

Complete sequences available at http://www.tsi.enst.fr/~delon/Demos/Flicker_stabilization/.

Extension to flicker reduction

Global correction [Delon, IEEE IP, 2006]

$$\mathrm{STE}[u_t](x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int e^{-(t-s)^2/2\sigma^2} \underbrace{H_s^{-1} \circ \underbrace{H_t(u_t(x))}_{\text{rank of } x \text{ in } u_t}}_{\text{grey level of the pixel having the same rank in } u_s} ds.$$



Global reduction not always enough !

Extension to flicker reduction

- $u_t(\Lambda_x)$ patch centered at x, and H_{t,Λ_x} cumulative distribution function of this patch,
- motion estimation : $\varphi_{t,s}(x) = \mathrm{Argmin}_{y \in W} D(u_t(\Lambda_x), u_s(\Lambda_y))$, where D is robust to contrast changes and W is a search window.

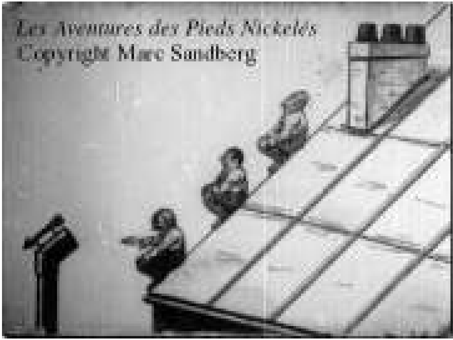
Local correction [Delon, Desolneux, 2010]

$$\mathrm{LStab}[u_t](x) = \frac{1}{Z_{t,x}} \int_s e^{-(t-s)^2/2\sigma^2} w_{t,x}(s, \varphi_{t,s}(x)) \quad \underbrace{H^{-1}_{s,\Lambda_{\varphi_{t,s}(x)}}(\underbrace{H_{t,\Lambda_x}(u_t(x))}_{\text{rank of } x \text{ in } u_t(\Lambda_x)})}_{\text{grey level with the same rank in } u_s(\Lambda_{\varphi_{t,s}(x)})} ds,$$

where

- $w_{t,x}(s, y) = e^{-D^2(u_t(\Lambda_x), u_s(\Lambda_y))/h^2}$, h a parameter of the method,
- $Z_{t,x}$ is a normalization factor.

Flicker reduction



Before restoration

After restoration

Barycenters between two Gaussian distributions

Two Gaussian distributions on $\mathbb{R}^M$, $\mu_0 = \mathcal{N}(m_0, \Sigma_0)$ and $\mu_1 = \mathcal{N}(m_1, \Sigma_1)$.

OT distance between μ_0 and μ_1 (for a quadratic ground cost) [Dowson, Landau, 1982]

$$d(\mu_0, \mu_1)^2 = \|m_0 - m_1\|^2 + \mathrm{tr}\left(\Sigma_0 + \Sigma_1 - 2\Sigma_{0,1}\right),$$

with $\Sigma_{0,1} = \left(\Sigma_1^{1/2}\Sigma_0\Sigma_1^{1/2}\right)^{1/2}$.

Geodesic Interpolation

$$\forall t \in [0,1],\ \mu_t = \arg\min_{\mu} \quad (1-t)d(\mu_0,\mu)^2 + td(\mu_1,\mu)^2.$$

Proposition

If $\ker(\Sigma_0) \not\subset \ker(\Sigma_1)^{\perp}$ and $\mathrm{rank}(\Sigma_0) \geq \mathrm{rank}(\Sigma_1)$, the unique Gaussian OT-geodesic between μ_0 and μ_1 is $\mu_t = \mathcal{N}(m_t, \Sigma_t)$, where

$$\forall t \in [0,1], \begin{cases} m_t = (1-t)m_0 + tm_1 \\ \Sigma_t = ((1-t)Id + t\Pi)\,\Sigma_0\,((1-t)Id + t\Pi) \end{cases}$$

Where $\Pi = \Sigma_1^{1/2}\Sigma_{0,1}^{+}\Sigma_1^{1/2}$, with $\Sigma_{0,1}^{+}$ the Moore-Penrose pseudo-inverse of $\Sigma_{0,1}$.

Barycenters between L Gaussian distributions

Gaussian distributions on $\mathbb{R}^M$, $\mu_i = \mathcal{N}(m_i, \Sigma_i)$ and weights ρ_i.

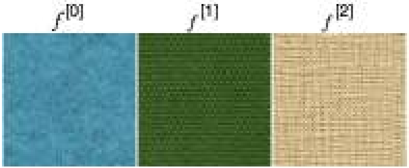
Proposition [Agueh, Carlier, 2010]
If at least one of the Σ_i has full rank, then the OT barycenter μ_t exists, is unique and is a Gaussian process $\mathcal{N}(m_t, \Sigma_t)$ where $m_t = \sum \rho_i m_i$ and Σ_t is the unique solution of the fixed point equation

$$\Sigma_t = \sum \rho_i (\Sigma_t^{1/2} \Sigma_i \Sigma_t^{1/2})^{1/2}.$$

Illustration from [Ferradans et al., 2013]

Texture mixing with Gaussian texture models [Ferradans et al., 2013]

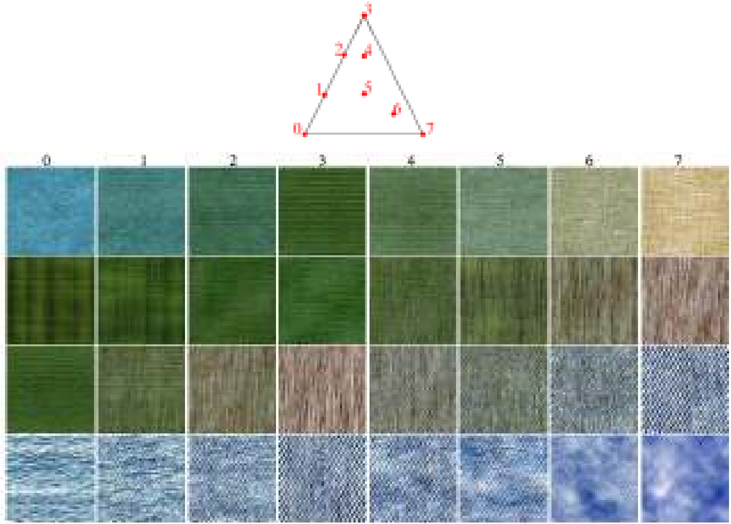
Goal: synthetize the geodesics and barycenters between several texture images $f_i \in \mathbb{R}^N$, $i = 1 \dots L$, with N the number of pixels.



Idea: model textures as realizations of Gaussian random fields $\mu_i = \mathcal{N}(m_i, \Sigma_i)$;

1. Analyse: compute the parameters of these fields from the input texture images f_i. Assuming stationarity, learn m_i (constant) and Σ_i (cyclic matrix) from each input texture f_i.
2. Compute Barycenters of these distributions.
3. Synthesis: synthetize the corresponding textures. A realization of $\mathcal{N}(m, \Sigma)$ is obtained by computing $AW + m$, where $W \sim \mathcal{N}(0, \mathrm{Id})$ and A a matrix such that $\Sigma = AA^*$.

Texture mixing with Gaussian texture models [Ferradans et al., 2013]



General case - Sliced Kantorovich Barycenter [Rabin et al., 2011]

Replace the previous definition by

$$\inf_{\mu \in \mathcal{P}_N(\mathbb{R}^d)} \sum_{l=1}^{L} \rho_l \mathrm{SMK}_\Theta(\mu, \nu^l).$$

Gradient descent on the energy

$$bar_{SMK}(X) = \sum_{l=1}^{L} \rho_l \mathrm{SMK}_\Theta(\mu_X, \nu^l).$$

Step k of the gradient descent (still with $H = \sum_{\theta \in \Theta} \theta\theta^T$)

$$X^{(k+1)} = X^{(k)} - \lambda H^{-1} \nabla bar_{\mathrm{SMK}_\Theta}(X^{(k)}).$$

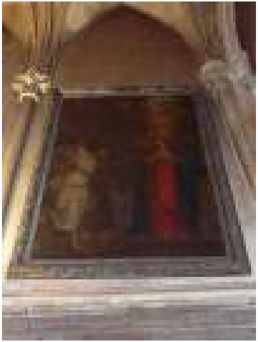
Kantorovich Barycenter [Rabin et al., 2011]

Application to color midway

Color harmonization of several images

▷ Step 1: compute Sliced Kantorovich Barycenter of color statistics;

▷ Step 2: projection of each image onto the Barycenter.



Application to color midway

Color harmonization of several images

▷ Step 1: compute Sliced Kantorovich Barycenter of color statistics;

▷ Step 2: projection of each image onto the Barycenter.

Part IV

Regularized optimal transport

Optimal transport for image modifications. Shortcomings.

Shortcomings of contrast and color modification : $T(u)$ image obtained after color modification.

- Noise enhancement: the variance of the noise in $T(u)$ is amplified.


Optimal transport for image modifications. Shortcomings.

Shortcomings of contrast and color modification : $T(u)$ image obtained after color modification.

- Noise enhancement: the variance of the noise in $T(u)$ is amplified.
- Detail loss: contrast reduction between u and $T(u)$.

Optimal transport for image modifications. Shortcomings.

Shortcomings of contrast and color modification : $T(u)$ image obtained after color modification.

- Noise enhancement: the variance of the noise in $T(u)$ is amplified.
- Detail loss: contrast reduction between u and $T(u)$.
- Compression artifacts: when the image u is highly compressed and when pixels with a very similar color are mapped to different colors.

Regularizing transportation maps. Variational approaches.

Variational problem

$$\inf_u \quad \underbrace{F(u)}_{\text{data fidelity term}} \quad + \quad \underbrace{R(u)}_{\text{regularization term}} \quad + \quad \underbrace{W_2(h_u,\nu)}_{\text{Transportation cost between } h_u \text{ and } \nu} \quad ,$$

Case of color transfer between an image u_0 and another image v

- ν is the color distribution of v;
- $R(u) = \mathrm{TV}(u)$, typically;
- $F(u) = \|u - u_0\|^2 - \lambda < \nabla u, \frac{\nabla u_0}{\|\nabla u_0\|} >$ (the second term ensures the consistency between the level sets of u_0 and those of u).

Still complex to minimize.
To make the problem computationally tractable, a possibility is to replace W_2 by a simpler term, for instance

- its sliced version [Rabin, Peyré, 2011],
- a term relying on cumulative histograms [Papadakis, Provenzi, Caselles, 2011].

Regularizing transportation maps. Post-processing solution.

Transportation Map Regularization [Rabin, Delon, Gousseau, 2011]: if $T(u)$ is the color-modified image, regularize the difference $\mathcal{M}(u) = T(u) - u$ conditionally to the source image u.

Inspired by previous works on denoising: Neighborhood filter [Yarovslavski 1985] and Non-Local means [Buades, Coll and Morel 2005]

$$\mathrm{TMR}_u[T(u)] := \; u + \mathrm{Y}_u\left[\mathcal{M}(u)\right] \; = \; \underbrace{\mathrm{Y}_u\big[T(u)\big]}_{\text{filtering of } T(u)} + \underbrace{u - \mathrm{Y}_u[u]}_{\text{Details from } u}$$

where Y_u is the following guided filter

$$\mathrm{Y}_u\left[w\right](x) = \frac{1}{C(x)} \int_{y \in \mathcal{N}(x)} w(y) e^{-\frac{||u(x)-u(y)||^2}{\sigma^2}} dy.$$

Regularizing transportation maps. Post-processing solution.

Regularizing transportation maps. Post-processing solution.

Regularizing transportation maps. Post-processing solution.

Regularizing transportation maps. Post-processing solution.

Source Image

Style Image

Result of color transfer

Regularizing transportation maps. Post-processing solution.

Source Image

Style Image

Result after regularization

Beyond regularization. Transport relaxation.

Beyond regularization. Transport relaxation.

Beyond regularization. Transport relaxation.

Beyond regularization. Transport relaxation.

Relaxation + regularization in [Ferradans et al., 2013].

Regularization not possible if one want to maintain a bijective assignment between distributions ?

A bijective assignment is usually not necessary or desirable in image processing problems. Relaxing this constraint permits to handle mass differences in the distribution modes (for instance color distributions).

Relaxation

Assignment problem between two sets of points μ and ν replaced by

$$\inf_{\gamma \in \Gamma(\mu,\nu)} \sum_{i,j} c(p_i, q_j)\gamma_{ij}.$$

where $\Gamma(\mu,\nu)$ is the set of matrices γ satisfying the constraints

$$\gamma_{ij} \geq 0, \quad \sum_{j=1}^{M} \gamma_{ij} = 1, \quad \sum_{i=1}^{N} \gamma_{ij} \leq \alpha, \text{ with } \alpha > 1.$$

Conclusion

Natural framework for image processing and modeling with statistical constraints

- generalization of histogram specification to color images; can be embedded in a variationnal formulation;
- texture synthesis;
- Enables to define barycenters between distributions
  - midway image processing, flicker reduction;
  - texture mixing;
- Robust distances in computer vision.
  - image retrieval;
  - object recognition.

Need some form of relaxation or regularization in practice.