Colour transfer in images and shape registration using Optimal transport and information theory

Prof. Rozenn Dahyot
Associate Professor
School of Computer Science & Statistics
Trinity College Dublin, Ireland
29/04/2019 Seminar @ University of Angers
1. Optimal Transport (OT)

2. Divergence L2

3. Applications:
   a. Colour transfer
   b. Shape registration

4. Final remarks on L2 or OT
   a. Machine learning
   b. Neural Networks

Video credit Gabriel Peyré:
https://twitter.com/gabrielpyre/status/979605863295053826
Optimisation with Optimal transport framework

Problem setting (unsupervised)

Consider an Euclidian space of dimension=1
e.g. $\mathbb{R}$
Optimisation with Optimal transport framework

Problem setting (unsupervised)

In that space $\mathbb{R}$, consider observations associated to a random variable $x$
In that same space $\mathbb{R}$, consider observations associated to another random variable $y$

$$y^{(1)}y^{(2)}y^{(3)}y^{(4)}y^{(5)}y^{(6)}y^{(7)}$$

In that space $\mathbb{R}$, consider observations associated to a random variable $x$

$$x^{(1)}x^{(2)}x^{(3)}x^{(4)}x^{(5)}x^{(6)}$$
In that same space \( \mathbb{R} \), consider observations associated to another random variable \( y \):

\[ y^{(1)}y^{(2)}y^{(3)} \quad y^{(4)}y^{(5)}y^{(6)}y^{(7)} \]

Find an estimate of the mapping (transfer) function \( \varphi \):

\[ y = \varphi(x) \]

In that space \( \mathbb{R} \), consider observations associated to a random variable \( x \):

\[ x^{(1)} \quad x^{(2)} \quad x^{(3)} \quad x^{(4)} \quad x^{(5)} \quad x^{(6)} \]
Optimisation with Optimal transport framework

Problem setting (unsupervised)

Find an estimate of mapping (transfer) function $\varphi$

$y = \varphi(x)$
Optimisation with Optimal transport framework

Problem setting (unsupervised)

For r.v. $y$
- pdf $g(y)$
- cdf $G(y)$

For r.v. $x$
- pdf $f(x)$
- cdf $F(x)$

Find an estimate of mapping (transfer) function $\varphi$

$y = \varphi(x)$
Kantorovitch proposed to find $p(x, y)$ (joint pdf) to relax Monge Optimal Transport formulation $y = \varphi(x)$.
Optimisation with Optimal transport framework
Problem setting (unsupervised)

• Monge-Kantorovitch’s cost function:

\[ \hat{\rho} = \arg \min \int \int c(x, y) \ p(x, y) \ dx \ dy \]

when \( c(x, y) = \|x - y\|^2 \) this is the Wasserstein distance.

• With deterministic coupling \( y = \varphi(x) \), the solution \( \varphi \) verifies the equation:

\[ f(x) = g(\varphi(x)) \quad |\nabla \varphi(x)| \]
Optimisation with Optimal transport framework
Problem setting (unsupervised)

- When $x \in \mathbb{R}$ and $y \in \mathbb{R}$ the solution $\varphi$ verifies:

\[
f(x) = g(\varphi(x)) \quad \varphi'(x)
\]

\[
F(x) = G(\varphi(x))
\]

Hence

\[
\varphi(x) = G^{-1} \circ F(x)
\]
Optimisation with Optimal transport framework
Problem setting (unsupervised)

\[ y = \varphi(x) = G^{-1} \circ F(x) \]

Application to Flicker removal in video:
• \( y \) is capturing pixel values in a reference image (target T) in a video
• \( x \) is representing a pixel value in another image (I) in that video
• Compute \( \varphi(x) \) with CDFs F and G
• All observations \( \{x^{(i)}\} \) (pixel values in T) associated with x are changed to \( \varphi(x^{(i)}) \) to create a recolored image R

Video credit Sigmedia team in collaboration with Irish Film Institute
http://www.sigmedia.tv/Research
RORY O’MORE: http://ifiplayer.ie/rory/
Optimisation with Optimal transport framework
Problem setting (unsupervised)

- When $x \in \mathbb{R}^d$ and $y \in \mathbb{R}^d$ with $d > 1$ the solution $\varphi$ is difficult to find!

$$f(x) = g(\varphi(x)) \quad |\nabla \varphi(x)|$$

Our Iterative Distribution Transfer (IDT) algorithm use the simple solution $y = \varphi(x) = G^{-1} \circ F(x)$ iteratively instead. Kullback-Leibler (KL) divergence between pdfs $f$ and $g$ is observed decreasing during IDT.

Automated Colour Grading using Colour Distribution Transfer
Example

\( x \in \mathbb{R}^2 \) and
\( y \in \mathbb{R}^2 \)
Divergences between pdfs $f$ and $g$

Bregman divergence [36]

Density power divergence [35]

A general divergence family [45]

The minimum divergence estimator is a M-estimator [37, 38] only if $\phi = 0$ or $1$

Windham divergence [46]

IDT algorithm for OT cost function (TCD-PHD2007)

L2 distance [5, 34]

KL divergence (maximum likelihood)

Correlation [4, 33]

Figure credit:
Robust Point Set Registration Using Gaussian Mixture Models,
B. Jian & B.C. Vermuri (2011)
IEEE transactions on Pattern Analysis and Machine Intelligence DOI:10.1109/TPAMI.2010.223
Optimisation with minimizing L2
Problem setting (unsupervised)

\[ y(1)y(2)y(3) \quad y(4)y(5)y(6)y(7) \]
\[ x(1) \quad x(2) \quad x(3) \quad x(4) \quad x(5) \quad x(6) \]

In this framework, the two datasets allow to provide two pdf candidates for the r.v. \( y \)
Optimisation with minimizing L2
Problem setting (unsupervised)

Euclidian distance between pdfs:

\[ \|f_\theta - g\|^2 = \int (f_\theta(y) - g(y))^2 \, dy = \|f_\theta\|^2 - 2\langle f_\theta | g \rangle + \|g\|^2 \]

Used for parameter estimation:

\[ \hat{\theta} = \arg \min \|f_\theta - g\|^2 \]

Equivalent to (L2E)

\[ \hat{\theta} = \arg \min \{ \|f_\theta\|^2 - 2\langle f_\theta | g \rangle \} \]
Optimisation with minimizing L2
Problem setting (unsupervised)

- $f_\theta$ and $g$ modelled as GMMs
- Transfer function $\varphi_\theta$ applied to the means of Gaussians in GMM $f_\theta$
- $\varphi_\theta$ rigid or non-rigid parametric function (e.g. Thin Plate Splines) controlled by $\theta$
Optimisation with minimizing L2
Problem setting (unsupervised)

\[ f_\theta(y) = \frac{1}{n_x} \sum_{i=1}^{n_x} \frac{1}{\sqrt{2\pi} h} \exp \left( -\frac{\|y - \varphi_\theta(x^{(i)})\|^2}{2h^2} \right) \]

\[ g(y) = \frac{1}{n_y} \sum_{k=1}^{n_y} \frac{1}{\sqrt{2\pi} h} \exp \left( -\frac{\|y - y^{(k)}\|^2}{2h^2} \right) \]

\[ \langle f_\theta | g \rangle = \frac{1}{n_y n_x} \sum_{i=1}^{n_x} \sum_{k=1}^{n_y} \frac{1}{2\sqrt{\pi} h} \exp \left( -\frac{\|y^{(k)} - \varphi_\theta(x^{(i)})\|^2}{4h^2} \right) \]

\[ \|f_\theta\|^2 = \langle f_\theta | f_\theta \rangle = \frac{1}{n_x n_x} \sum_{i=1}^{n_x} \sum_{k=1}^{n_x} \frac{1}{2\sqrt{\pi} h} \exp \left( -\frac{\|\varphi_\theta(x^{(k)}) - \varphi_\theta(x^{(i)})\|^2}{4h^2} \right) \]
In practice, K-means clustering is used to reduce the number of Gaussians in $f_\theta$ and $g$.

$$\hat{\theta} = \arg \min \{ \|f_\theta\|^2 - 2\langle f_\theta | g \rangle \}$$

$\|f_\theta\|^2$ corresponds to $-\log$ (Renyi Entropy of $f_\theta$)

So the estimation constrains the resulting pdf $f_\theta$ to maximise entropy of $f_\theta$, while maximizing overlap between $f_\theta$ and $g$. 
Application: Colour transfer
Observations are in 3D colour space

\((x^{(i)})\)

\((y^{(j)})\)
L2 Divergence for Robust Colour Transfer
Mairead Grogan and Rozenn Dahyot,
Computer Vision and Image Understanding (2019)
DOI:10.1016/j.cviu.2019.02.002
User Interaction for Image Recolouring using L2
M. Grogan, R. Dahyot & A. Smolic
Conference on Visual Media Production (2017 – Awarded Best paper)
DOI:10.1145/3150165.3150171
Application: Shape Registration

Shape representation

\[
\begin{align*}
\{(x^{(i)})\} \\
\text{or} \\
\{(x^{(i)}, u^{(i)})\}
\end{align*}
\]

\[
\begin{align*}
\chi \in \mathbb{R}^2, \quad u \in S, \\
\|u\| = 1
\end{align*}
\]

\[
\hat{y}_1 = \varphi_{\theta_1}(x)
\]

\[
\hat{y}_2 = \varphi_{\theta_2}(x)
\]

\[
\{(y_1^{(j)}, v_1^{(j)})\}
\]

\[
\{(y_2^{(j)}, v_2^{(j)})\}
\]
Without normal vectors

\[ \hat{y}_1 = \varphi_{\theta_1}(x) \]

\[ \hat{y}_2 = \varphi_{\theta_2}(x) \]

Failed!
With normal vectors

\[ \hat{y}_1 = \varphi_{\theta_1}(x) \]

\[ \hat{y}_2 = \varphi_{\theta_2}(x) \]

Success!
Machine Learning : OT Vs L2

Terminology unsupervised, semi-supervised and supervised

• Supervised scenario: When observations are paired

\[ \{(x^{(i)}, y^{(i)})\} \]

L2/L2E

• Unsupervised scenario: When observations are not paired

\[ \{(x^{(i)})\} \quad \{(y^{(j)})\} \]

OT

L2/L2E

\[ \text{machine} \quad y = \varphi(x) \]

\[ \text{dim}(x) = \text{dim}(y) \]
Sliced Wasserstein Generative Models (2019)

OT & DNNs
Iterative Distribution Transfer (IDT) == Sliced Wasserstein Distance (SWD)

Optimal transport solution with IDT algorithm https://github.com/frcs/colour-transfer
Summary

**OT**  Francois Pitié (TCD-PhD2007)
1. Cost function defined with OT
2. Iterative distribution transfer (IDT) algorithm
3. Estimate a non parametric mapping function $\varphi$
4. Shown to reduce on average at each step the KL divergence

**L2**  Mairead Grogan (TCD-PhD2017)
1. Cost function defined as **L2E distance**
2. This cost function include a constraint on the estimated pdf $f_\theta$ to maximize Entropy
3. Simulated Annealing algorithm
4. Estimate a parametric mapping function $\varphi_\theta$
5. Many tricks used for modelling the pdf $f$ and $g$ as GMMs

[DOI:10.1016/j.cviu.2006.11.011](https://github.com/frcs/colour-transfer)
Any questions?
Thank you!

https://www.scss.tcd.ie/Rozenn.Dahyot/