



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

ARCHIVIO ISTITUZIONALE
DELLA RICERCA

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Cortically Based Optimal Transport

This is the submitted version (pre peer-review, preprint) of the following publication:

Published Version:

Cortically Based Optimal Transport / Mattia Galeotti; Giovanna Citti; Alessandro Sarti. - In: JOURNAL OF MATHEMATICAL IMAGING AND VISION. - ISSN 0924-9907. - STAMPA. - 64:9(2022), pp. 1040-1057. [10.1007/s10851-022-01116-9]

Availability:

This version is available at: <https://hdl.handle.net/11585/907729> since: 2022-11-24

Published:

DOI: <http://doi.org/10.1007/s10851-022-01116-9>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).
When citing, please refer to the published version.

(Article begins on next page)

CORTICALLY BASED OPTIMAL TRANSPORT

MATTIA GALEOTTI, GIOVANNA CITTI, ALESSANDRO SARTI

ABSTRACT. We introduce a model for image morphing in the primary visual cortex V1 to perform completion of missing images in time. We model the output of simple cells through a family of Gabor filters and the propagation of the neural signal accordingly to the functional geometry induced by horizontal connectivity. Then we model the deformation between two images as a path relying two different outputs. This path is obtained by optimal transport considering the Wasserstein distance geodesics associated to some probability measures naturally induced by the outputs on V1. The frame of Gabor filters allows to project back the output path, therefore obtaining an associated image stimulus deformation. We perform a numerical implementation of our cortical model, assessing its ability in reconstructing rigid motions of simple shapes.

1. INTRODUCTION

The functional geometry of the visual cortex is a widely studied subject. It is known that cells of the primary visual cortex V1 are sensitive to specific features of the visual stimulus, like position, orientation, scale, colour, curvature, velocity and many others [17]. In the seventies the neurophysiologists Hubel and Wiesel discovered the modular organisation of the primary visual cortex [18], meaning that cells are spatially organized in such a way that for every point (x, y) of the retinal plane there is an entire set of cells, each one sensitive to a particular instance of the considered feature. This organisation corresponds to the so-called hypercolumnar structure. Hypercolumns of cells are then connected to each other by means of the horizontal connectivity, allowing cells of the same kind but sensitive to different points (x, y) of the stimulus to communicate. Hypercolumnar organization and neural connectivity between hypercolumns constitute the functional architecture of the visual cortex, that is the cortical structure underlying the low level processing of the visual stimulus. The mathematical modelling of the functional architecture of the visual cortex in terms of differential geometry was introduced in the seminal works of Hoffmann [15, 16], who proposed to model the hypercolumnar organization in terms of a fiber bundle. Many of such results dealing with differential geometry were given a unified framework under the new name of neurogeometry.

Petitot and Tondut [23], related the contact geometry introduced by Hoffmann with the geometry of illusory contours of Kanizsa [19]. The problem of completion of occluded object was afforded by computing geodesic curves in the contact structure.

Then, in [8] Citti and Sarti showed how the functional architecture could be described in terms of Lie groups structures. In particular, as proved by Daugman [11] the receptive profiles of simple cells can be modelled as a Gabor filter. Since these filters can be obtained via rotation and translation from a fixed one, the functional architecture of the whole family of simple cells have been described as the Euclidean motion group $SE(2)$ [8]. In presence of a visual stimulus $I : \mathbb{R}^2 \rightarrow [0, 1]$ on the retinal plane, the action of the whole family of simple cells is obtained by convolving the function I with the bank of Gabor filters. The output of the cells action will be a function $\mu : \mathbb{R}^2 \times S^1 \rightarrow \mathbb{R}$. The horizontal connectivity is strongly

anisotropic and it is modelled via a sub-Riemannian metric. Since it is very common that part of the visual input is occluded, the action of horizontal connectivity allows to complete the missing part by means of propagation in such a space, under the action of advection diffusion differential operators of Fokker-Planck type. Visual completion problems are then solved via geodesics or minimal surfaces. This approach was extended to scale in [29], to space-time in [4] and to frequency in [6]. In [13] and [30] the lifting has been extended to heterogeneous features defined in different groups. For an extended review on neurogeometry see [9].

In this paper we aim to reconsider the problem of completion of missing stimulus in time by means of morphing of one lifted cortical image in a different one in terms of optimal transport of a probability distribution in the functional geometry of the cortex. Two images can represent the same object at two different intervals of times, and different algorithms have been proposed to perform completion of missing images between the two. We recall the results of [32] and the model proposed by [33] on the image plane as a geodesic in the Wasserstein space.

We propose a cortical version of this phenomena, using geodesics in the cortical space endowed with the Wasserstein distance. We work in the manifold $M = \mathbb{R}^2 \times S^1 \times \mathbb{R}^+$ of cells sensible to position orientation and scale (see also [29]), and we develop a model that treats well shape rotations. We point out that the sub-Riemannian metric in $\mathbb{R}^2 \times S^1$ is important to keep together the shape of the object along the rotational movements. In fact the distance function in this metric approximates the statistical correlations of boundaries in natural images [28].

In order to obtain a metric deformation, we consider the positive and negative part of the output μ at a given time, normalized as two probability measures μ^+, μ^- over M . Following the papers [2, 3] of Ambrosio and Gigli, we consider the space $\mathcal{P}_2(X)$ of probabilities with finite 2-momentum over $X = \mathbb{R}^2 \times S^1 \times \mathbb{R}$, endowed with the associated Wasserstein distance; on this space we can find, for any pair of inputs I_0, I_1 taken at different times, a unique constant speed geodesic relying their associated output measures.

In particular, for any regular measure $\mu \in \mathcal{P}_2(X)$, we obtain Theorem 3.15, a generalization of previously known results, and it assures that for any measure $\nu \in \mathcal{P}_2(X)$ there exists a unique transport map T in the sense of Monge's formulation of optimal transport. Using this transport map, it is possible to give an explicit description of the constant speed geodesics in $\mathcal{P}_2(X)$; in the case of the measures induced by the output functions, this is done in Remark 4.4 and equation (6.3). Using the frame properties of the family of Gabor filters, we reconstruct with equation (5.5) a path of images I_t relying the initial and final input I_0, I_1 .

In the two final sections, we develop a numerical implementation of our model. In fact, the frame generated by the odd Gabor mother function is not invertible in a discrete setting, and in order to make our model workable, we use a Wavelet Gabor Pyramid generated by an odd and an even Gabor function. The outputs obtained via these frames are transported following the same approach detailed above. As shown in Section 9, our model allows the deformation of simple shapes through rotation and translation, preserving the basic structures in the treated pictures. As we discuss, the same preservation is not attained by the numerical 2-dimensional 'classical' implementation of optimal transport, meaning the implementation of the optimal transport between the inputs I_0, I_1 seen as measures on \mathbb{R}^2 and taking the square Euclidean distance as the cost function.

In Section 2 we introduce the operation associating to any input an output function via the convolution with the family of Gabor filters. In Section 3 the classical problem of optimal

transport is introduced, with the techniques necessary for a general solution. In Section 4 we describe the constant speed geodesics in $\mathcal{P}_2(X)$. In Section 5 we treat the properties of continuous frames such as the one of Gabor functions; they are fundamental in reconstructing a path of input images from the deformation path of output functions. Section 6 states that the measure deformation results are valid in our case. Finally, in Section 7 we find a constraining condition implying that the path of output functions μ_t is naturally induced by a path of input images I_t . In Section 8 we introduce the discrete setting and the Wave Gabor Pyramid. Finally in Section 9 we discuss the numerical implementation and its results.

2. FROM THE RETINA TO THE OUTPUT SPACE

We consider an input function $I: \mathbb{R}^2 \rightarrow [0, 1]$ in $L^2(\mathbb{R}^2)$. This functions models an input received in the retina plan and it induces an ouput function on the cortex. In order to introduce the output function, we recall the odd series of Gabor filters. We call Gabor mother filter the function

$$\psi_{0,0,0,1}: (\tilde{x}, \tilde{y}) \in \mathbb{R}^2 \mapsto e^{-(\tilde{x}^2 + \tilde{y}^2)} \cdot \sin(2\tilde{y}),$$

Moreover, we consider the roto-dilation defined by

$$A_{\theta, \sigma} := \sigma \cdot R_{\theta} = \sigma \cdot \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix},$$

for any $\theta \in S^1$ and $\sigma \in \mathbb{R}^+$. We also consider the application

$$A_{x,y,\theta,\sigma}: (\tilde{x}, \tilde{y}) \mapsto A_{\theta,\sigma}(\tilde{x}, \tilde{y}) + (x, y).$$

All this allows the definition of a family of Gabor filters

$$\psi_{x,y,\theta,\sigma}(\tilde{x}, \tilde{y}) := \frac{1}{\sigma^{3/2}} \cdot \psi_{0,0,0,1}(A_{\theta,\sigma}^{-1}(\tilde{x} - x, \tilde{y} - y)) = \frac{1}{\sigma^{3/2}} \cdot \psi_{0,0,0,1}(A_{x,y,\theta,\sigma}^{-1}(\tilde{x}, \tilde{y})).$$

In what follows we will denote the filter $\psi_{0,0,0,1}$ by ψ_0 when there is no risk of confusion. We consider the variety $M := \mathbb{R}^2 \times S^1 \times \mathbb{R}^+$ with its natural (Lebesgue) measure dk , this is the output space where we build the μ function induced by the input I . For any point $k = (x, y, \theta, \sigma) \in M$, we denote by ψ^k the Gabor filter $\psi_{x,y,\theta,\sigma}$.

Definition 2.1. Consider a point $k = (x, y, \theta, \sigma)$ on M , then the output function of a cell in response to the visual input I is

$$\begin{aligned} \mu(x, y, \theta, \sigma) &= \mu(k) := \langle I, \psi^k \rangle \\ &= \int_{\mathbb{R}^2} I(\tilde{x}, \tilde{y}) \psi_{x,y,\theta,\sigma}(\tilde{x}, \tilde{y}) d\tilde{x} d\tilde{y}. \end{aligned}$$

Remark 2.2. We work in the case $M = \mathbb{R}^2 \times S^1 \times \mathbb{R}^+$ and we put on M the Riemannian structure which endows the neurogeometry of the cortex. In particular the \mathbb{R}^2 factor is the retinal plan with the natural projection $M \rightarrow \mathbb{R}^2$, $\theta \in S^1$ is an angular parameter that encodes the border orientation in the processes of border recognition, while $\sigma \in \mathbb{R}^+$ is a scale parameter in the same process.

In order to define this metric g , consider the following four vector fields, in every point they span the M tangent bundle,

$$\begin{aligned} X_1 &= \cos \theta \cdot \partial_x + \sin \theta \cdot \partial_y, \\ X_2 &= \partial_\theta, \\ X_3 &= -\sin \theta \cdot \partial_x + \cos \theta \cdot \partial_y, \\ X_4 &= \partial_\sigma. \end{aligned}$$

Let g be the metric such that at any point $\frac{1}{\sqrt{\sigma}}X_1, \frac{1}{\sqrt{\sigma}}X_2, \sqrt{\sigma}X_3, \sqrt{\sigma}X_4$ is an orthonormal system. That is at every point the metric g is represented by the matrix

$$\tilde{g} = \begin{pmatrix} \sigma & 0 & 0 & 0 \\ 0 & \sigma & 0 & 0 \\ 0 & 0 & \frac{1}{\sigma} & 0 \\ 0 & 0 & 0 & \frac{1}{\sigma} \end{pmatrix}$$

For σ that tends to 0, g tends to the sub-riemannian structure treated in [8] (module rescaling), for $\sigma \rightarrow +\infty$ it approaches a hyperbolic metric.

In this setting the measure $\mu \cdot dk = \langle I, \psi^k \rangle \cdot dk$ on M , is $\langle I, \psi^k \rangle$ times the measure induced by the metric g .

As proved in Appendix A, the integral

$$\int_M \psi^k(\tilde{x}, \tilde{y}) dk$$

has finite value 0, independently of the pair (\tilde{x}, \tilde{y}) .

Remark 2.3. The mass of the output function μ on M , is proportional to the mass of I and therefore it is null. Indeed,

$$\begin{aligned} \int_M \mu(k) dk &= \int_M dk \int_{\mathbb{R}^2} I(\tilde{x}, \tilde{y}) \psi^k(\tilde{x}, \tilde{y}) d\tilde{x} d\tilde{y} \\ &= \int_{\mathbb{R}^2} I(\tilde{x}, \tilde{y}) d\tilde{x} d\tilde{y} \int_M \psi^k(\tilde{x}, \tilde{y}) dk = 0 \end{aligned}$$

by the Fubini's Theorem.

3. THE OPTIMAL TRANSPORT PROBLEM

We recall the classical Kantorovich's formulation of optimal transport. Our main reference is the Ambrosio-Gigli guide [3]. For any Polish space X (*i.e.* a complete and separable metric space) we denote by $\mathcal{P}(X)$ the set of Borel probability measures on X and by $\mathcal{B}(X)$ the set of Borel sets on X . Consider two Polish spaces X, Y , if $\mu \in \mathcal{P}(X)$ and $T: X \rightarrow Y$ is a Borel map, then we denote by $T_{\#}\mu \in \mathcal{P}(Y)$ the pushforward of μ through T , defined by

$$T_{\#}\mu(E) = \mu(T^{-1}E) \quad \forall E \in \mathcal{B}(Y).$$

Consider the natural product $X \times Y$ and its associated projections π^X, π^Y . Let $c: X \times Y \rightarrow \mathbb{R}$ be a Borel map called *cost* function, and consider two measures $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$.

Definition 3.1. An *admissible transport plan* between μ and ν is a measure $\gamma \in \mathcal{P}(X \times Y)$ such that $\pi_{\#}^X \gamma = \mu$ and $\pi_{\#}^Y \gamma = \nu$, or equivalently

$$\begin{aligned}\gamma(A \times Y) &= \mu(A) \quad \forall A \in \mathcal{B}(X) \\ \gamma(X \times B) &= \nu(B) \quad \forall B \in \mathcal{B}(Y).\end{aligned}$$

We denote the set of admissible transport plans between μ and ν by $\text{Adm}(\mu, \nu)$.

We want to minimize the integral

$$\int_{X \times Y} c(x, y) d\gamma(x, y)$$

for all the admissible transport plans between μ and ν . We say that γ is induced by a transport map if there exists a Borel map $T: X \rightarrow Y$ such that $\gamma = (\text{id} \times T)_{\#} \mu$, in that case

$$\int_{X \times Y} c(x, y) d\gamma = \int_X c(x, T(x)) d\mu.$$

An optimal transport plan is a transport plan γ that realizes the infimum above. It is known that such a minimizer exists under very general conditions. We denote the set of optimal plans by $\text{Opt}(\mu, \nu)$.

Proposition 3.2 (see [31, Theorem 4.1]). *Consider $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. If the cost function c is lower semicontinuous and bounded from below, then there exists an optimal plan γ for the functional*

$$\gamma \mapsto \int_{X \times Y} c(x, y) d\gamma(x, y),$$

among all $\gamma \in \text{Adm}(\mu, \nu)$.

We are interested in the cases where an optimal plan is induced by a transport map T . We state [3, Lemma 1.20], referring to Ambrosio-Gigli paper for a proof.

Lemma 3.3. *Consider $\gamma \in \text{Adm}(\mu, \nu)$. Then γ is induced by a map if and only if γ is concentrated in a measurable set $\Gamma \subset X \times Y$ such that for μ -a.e. x there exists only one $y = T(x)$ in $\Gamma \cap (\{x\} \times Y)$. In this case, $T(x)$ induces γ .*

In order to introduce the notion of c -concavity for some cost function c , and in order to show the existence and uniqueness of an optimal transport plan γ induced by a transport map T , we give some definitions following [3, Chap.1].

Definition 3.4 (Superdifferential). Consider M a riemannian manifold and any function $\varphi: M \rightarrow \mathbb{R}$, we define its superdifferential at any point $x \in M$,

$$\partial^+ \varphi(x) := \{dh(x) \in T_x^* M : h \in \mathcal{C}^1(M, \mathbb{R}), \varphi - h \text{ attains a local maximum at } x\}.$$

When there is no risk of confusion, we will denote by $\partial^+ \varphi$ the associated subspace of the total space $T^* M$. The subdifferential $\partial^- \varphi$ is defined analogously as the set of differentials $dh(p)$ where $\varphi - h$ attains a local minimum of $\varphi - h$.

Remark 3.5. Equivalently, $\partial^+ \varphi(x)$ is the set of vectors $v \in T_x M$ such that

$$\varphi(z) - \varphi(x) \leq \langle v, \exp_x^{-1}(z) \rangle + o(d(x, z)).$$

The same for $\partial^- \varphi(x)$ with inversed inequality. With this definition, $\partial^+ \varphi$ and $\partial^- \varphi$ are subspaces of TM .

It is well known that if φ is differentiable at $x \in M$, its superdifferential and subdifferential at x coincide and they contain only the φ gradient,

$$\partial^+ \varphi(x) = \partial^- \varphi(x) = \{\nabla \varphi(x)\}.$$

Consider two Polish spaces X, Y , and a cost function $c: X \times Y \rightarrow \mathbb{R}$.

Definition 3.6 (*c*-transforms). Consider a function $\varphi: X \rightarrow \mathbb{R} \cup \{\pm\infty\}$, its c_+ -transform $\varphi^{c+}: Y \rightarrow \mathbb{R} \cup \{\pm\infty\}$ is defined as

$$\varphi^{c+}(y) := \inf_{x \in X} c(x, y) - \varphi(x).$$

Analogously for any $\psi: Y \rightarrow \mathbb{R} \cup \{\pm\infty\}$, we can define its c_+ -transform $\psi^{c+}: X \rightarrow \mathbb{R} \cup \{\pm\infty\}$.

The c_- -transform of φ is $\varphi^{c-}: Y \rightarrow \mathbb{R} \cup \{\pm\infty\}$ defined as

$$\varphi^{c-}(y) := \sup_{x \in X} -c(x, y) - \varphi(x).$$

Analogously for the c_- -transform ψ^{c-} of ψ .

Definition 3.7 (*c*-concavity). We say that a function $\varphi: X \rightarrow \mathbb{R} \cup \{-\infty\}$ is *c*-concave if there exists $\psi: Y \rightarrow \mathbb{R} \cup \{-\infty\}$ such that $\varphi = \psi^{c+}$. Analogously we have a notion of *c*-convexity.

Definition 3.8 (Semiconcavity). A function $f: U \rightarrow \mathbb{R}$ whose domain is a convex subset U of a riemannian manifold M , is semiconcave with constant K if for every geodesic $\gamma: [0, 1] \rightarrow U$ and $t \in [0, 1]$, we have

$$t \cdot f(\gamma_0) + (1-t) \cdot f(\gamma_1) \leq f(\gamma_t) + \frac{1}{2} \cdot t(1-t)K \cdot d^2(\gamma_0, \gamma_1).$$

With the notation of [31, Chap.10], the definition above describes a semiconcave function with modulus $\omega(t) = K \frac{t^2}{2}$.

Remark 3.9. Observe that by [31, Equation (10.14)], if f is semiconcave, superdifferentiable at x and $q \in \partial^+ f(x)$, then

$$f(\exp_x w) - f(x) \leq \langle q, w \rangle - \frac{1}{2}K \|w\|^2.$$

Definition 3.10 (*c*-superdifferential). Consider $\varphi: X \rightarrow \mathbb{R} \cup \{-\infty\}$ a *c*-concave function, then its *c*-superdifferential $\partial^{c+}\varphi \subset X \times Y$ is defined as

$$\partial^{c+}\varphi := \{(x, y) : \varphi(x) + \varphi^{c+}(y) = c(x, y)\}.$$

We denote by $\partial^{c+}\varphi(x)$ the set of $y \in Y$ such that $(x, y) \in \partial^{c+}\varphi$. Analogously we can define the *c*-subdifferential $\partial^{c-}\varphi \subset X \times Y$.

Consider two probability measures $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. In the following we will consider a cost function $c: X \times Y \rightarrow \mathbb{R}$ such that there exists two functions $a \in L^1(\mu)$, $b \in L^1(\nu)$, respecting the inequality below.

$$(3.1) \quad c(x, y) \leq a(x) + b(y).$$

Theorem 3.11 ([3, Theorem 1.13]). Consider $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and $c: X \times Y \rightarrow \mathbb{R}$ a continuous and bounded from below cost function such that there exist two functions $a \in L^1(\mu)$ and $b \in L^1(\nu)$ verifying condition (3.1). Then there exists a *c*-concave function $\varphi: X \rightarrow \mathbb{R}$ such that $\varphi \in L^1(\mu)$, $\varphi^{c+} \in L^1(\nu)$ and for any optimal plan $\gamma \in \text{Opt}(\mu, \nu)$,

$$\text{supp}(\gamma) \subset \partial^{c+}\varphi.$$

Consider the manifold $X = \mathbb{R}^2 \times S^1 \times \mathbb{R}$ with the Lebesgue metric. Consider a function $\varphi: X \rightarrow \mathbb{R}$ and the square distance $c(x, y) = d^2(x, y)/2$ as cost function, then the following lemma and proposition state the link between the superdifferential $\partial^+\varphi$ and the c -superdifferential $\partial^{c+}\varphi$.

Lemma 3.12. *For any point $y \in X$, the function $d^2(-, y)/2$ is uniformly semiconcave on X .*

This is proven following the line of reasoning of [31, Third Appendix], because X is flat and therefore its sectional curvature is everywhere 0.

Proposition 3.13. *Consider a c -concave function $\varphi: X \rightarrow \mathbb{R}$, then it must be semiconcave. Furthermore, for any $x \in X$, $\exp_x^{-1}(\partial^{c+}\varphi(x)) \subset -\partial^+\varphi(x)$.*

Proof. This is a slight generalization of [3, Proposition 1.30], and we develop the same arguments. As stated by Lemma 3.12, for any $y \in X$ the distance function $d^2(-, y)/2$ is semiconcave, therefore by Remark 3.9 this implies

$$\frac{d^2(z, y)}{2} - \frac{d^2(x, y)}{2} \leq -\langle v, \exp_x^{-1}(z) \rangle + o(d(x, z)),$$

because if $v \in \exp_x^{-1}(y)$, then $-v$ is in the superdifferential of $d^2(-, y)/2$ at x .

If we take $d^2(x, y)/2$ as the cost function $c(x, y)$ and consider $y \in \partial^{c+}\varphi(x)$, therefore by definition $\varphi(z) - c(z, y) \leq \varphi(x) - c(x, y)$ for any $z \in X$. As a consequence

$$\varphi(z) - \varphi(x) \leq \frac{d^2(z, y)}{2} - \frac{d^2(x, y)}{2} \leq \langle -v, \exp_x^{-1}(z) \rangle + o(d(x, z)),$$

that means $-v \in \partial^+\varphi(x)$. □

Definition 3.14 (Regular measure). We say that a measure $\mu \in \mathcal{P}(X)$ is regular if it vanishes on the set of points of non differentiability of any semiconcave function $\varphi: X \rightarrow \mathbb{R}$.

For any Polish space (X, d) , we introduce the space of probability measures on X with finite 2-momentum with respect to d ,

$$\mathcal{P}_2(X) := \left\{ \mu \in \mathcal{P}(X) : \int d^2(x, x_0) d\mu < \infty \text{ for some, and thus any, } x_0 \in X \right\}.$$

We regard $\mathcal{P}_2(X)$ as a metric space with respect to the sup norm.

Theorem 3.15. *Consider the riemannian manifold X and a probability measure $\mu \in \mathcal{P}_2(X)$. If μ is regular, then for every $\nu \in \mathcal{P}_2(X)$ there exists only one transport plan from μ to ν and it is induced by a map T . If this is the case, the map T can be written as $x \mapsto \exp_x(-\nabla\varphi(x))$ for some c -concave function $\varphi: X \rightarrow \mathbb{R}$.*

Observe that this is a generalization of [3, Theorem 1.33] to the case of the non-compact riemannian manifold X . For another reference see also [14].

Proof. In order to apply Theorem 3.11 we are going to verify that condition (3.1) is respected. In particular we want to show that taking $a(x) = d^2(x, x_0)$ and $b(y) = d^2(y, x_0)$ for any $x_0 \in M$, then $a \in L^1(\mu)$, $b \in L^1(\nu)$ and $c(x, y) = d^2(x, y)/2 \leq d^2(x, x_0) + d^2(y, x_0)$. The

inequality is proved by

$$\begin{aligned} d^2(x, x_0) + d^2(y, x_0) &\geq \frac{d^2(x, x_0) + d^2(y, x_0)}{2} + |d(x, x_0)d(y, x_0)| \\ &= \frac{1}{2}(d(x, x_0) + d(y, x_0))^2 \geq \frac{d^2(x, y)}{2}. \end{aligned}$$

To have $a \in L^1(\mu)$ (and therefore $b \in L^1(\nu)$) it suffices to have $\int_M d^2(x, x_0)\mu(x) < \infty$ but this is exactly the definition of $\mu \in \mathcal{P}_2(X)$.

Thus as a consequence of Theorem 3.11, there exists a c -concave function φ such that any optimal plan γ is concentrated on $\partial^{c+}\varphi$. By Proposition 3.13, φ is semiconcave and therefore differentiable μ -a.e. by μ -regularity. If φ is differentiable at x , then $\partial^+\varphi(x)$ is the singleton $\{\nabla\varphi(x)\}$, therefore $\partial^{c+}\varphi(x)$ is empty or equals $\exp_x(-\nabla\varphi(x))$. We define μ -a.e. the function $T(x) = \exp_x(-\nabla\varphi(x))$. As $\text{supp}(\gamma) \subset \partial^{c+}\varphi$, we must have that T induces γ , concluding the proof. \square

4. GEODESICS IN $\mathcal{P}_2(X)$

The introduction of the Wasserstein distance W_2 on \mathcal{P}_2 allows the definition of geodesics in this space. We show a theorem of existence and uniqueness for such a geodesic relying a pair of measures.

Definition 4.1. Consider a metric space (X, d) . A curve $\gamma: [0, 1] \rightarrow X$ is a constant speed geodesic if

$$d(\gamma_s, \gamma_t) = |t - s|d(\gamma_0, \gamma_1) \quad \forall t, s \in [0, 1].$$

We recall that (X, d) is called a geodesic space if for every $x, y \in X$ there exists a constant speed geodesic connecting them. We consider the metric space $\text{Geod}(X)$ of constant speed geodesics endowed with the sup norm.

On $\text{Geod}(X)$ we introduce for any $t \in [0, 1]$ the map $e_t: \text{Geod}(X) \rightarrow X$ such that

$$e_t: \gamma \mapsto \gamma_t.$$

Furthermore we define the Wasserstein distance associated to d on $\mathcal{P}_2(X)$.

Definition 4.2 (Wasserstein distance). If $\mu, \nu \in \mathcal{P}_2(X)$, then

$$W_2^2(\mu, \nu) := \inf_{\gamma \in \text{Adm}(\mu, \nu)} \int d^2(x, y)d\gamma.$$

The following theorem is proved in [3]. For any two probability measure μ_0, μ_1 on a Polish space, this gives a constant speed geodesic relying them.

Theorem 4.3. *If (X, d) is Polish and geodesic, then $(\mathcal{P}_2(X), W_2)$ is geodesic too. Furthermore, consider $\mu_0, \mu_1 \in \mathcal{P}_2(X)$ and a path $t \mapsto \mu_t \in \mathcal{P}_2(X)$ from μ_0 to μ_1 , then μ_t is a constant speed geodesic if and only if there exists $\boldsymbol{\mu} \in \mathcal{P}_2(\text{Geod}(X))$ such that $(e_0, e_1)_{\#}\boldsymbol{\mu} \in \text{Opt}(\mu_0, \mu_1)$ and $\mu_t = (e_t)_{\#}\boldsymbol{\mu}$.*

We consider on X the Lebesgue measure dk . The measures μ that we are going to treat, are always induced by density a.e.-continuous functions.

Remark 4.4. We know that the geodesic relying two points on a complete riemannian manifold X is almost everywhere unique (see for example [31]). This means that the set of pairs $(x, y) \in X^2$ such that the geodesic between them is unique, has full measure. The same is true for any measure μ on X if it is induced by a density a.e.-continuous function.

Therefore the maps e_t naturally induces almost everywhere a map that we denote in the same way:

$$e_t: X^2 \rightarrow X,$$

sending the pair (x, y) to the point γ_t where γ is the geodesic such that $\gamma_0 = x$ and $\gamma_1 = y$.

Consider two probability measures μ_0, μ_1 over X respecting the hypothesis of Theorem 3.15. Let T be the transport map between them, then as a consequence of Theorem 4.3 the (unique, coming from the uniqueness of T) geodesic between μ_0 and μ_1 can be written as

$$\mu_t = (e_t \circ (\text{id}, T))_{\#} \mu_0.$$

In what follows we will use the notation $e_t^{(T)} := e_t \circ (\text{id}, T)$, and therefore we will have $\mu_t = (e_t^{(T)})_{\#} \mu_0$.

5. RECONSTRUCTING THE VISUAL INPUT VIA THE GABOR FRAME

We are going to introduce the notion of continuous frame, this allows the reconstruction of a function I on the retinal plane, from the datum of an output function μ .

Definition 5.1. Consider a Hilbert space \mathcal{H} and a measure space M with a positive measure ρ . A continuous frame is a family of vectors $\{\psi^k\}_{k \in M}$ such that $k \mapsto \langle f, \psi^k \rangle$ is a measurable function on M for any $f \in \mathcal{H}$, and there exists $A, B > 0$ such that

$$A \cdot \|f\|^2 \leq \int_X |\langle f, \psi^k \rangle|^2 d\rho(k) \leq B \cdot \|f\|^2.$$

Consider $f, g \in \mathcal{H}$ and the mapping

$$h_f: g \mapsto \int_M \langle f, \psi^k \rangle \langle \psi^k, g \rangle d\rho(k).$$

This map is conjugated linear and moreover it is bounded. Indeed,

$$(5.1) \quad |h_f(g)|^2 \leq \int_M |\langle f, \psi^k \rangle|^2 d\rho(k) \cdot \int_X |\langle \psi^k, g \rangle|^2 d\rho(k) \leq B^2 \|f\|^2 \|g\|^2.$$

By the Riesz' representation theorem, there exists a unique element $\underline{h} \in \mathcal{H}$ which verifies $h_f = \langle \underline{h}, - \rangle$. We denote this element by $\int_X \langle f, \psi^k \rangle \psi^k d\rho(k)$.

We denote by $S: \mathcal{H} \rightarrow \mathcal{H}$ the operator

$$Sf := \int_M \langle f, \psi^k \rangle \psi^k d\rho(k), \quad \forall f \in \mathcal{H}.$$

Lemma 5.2. *the operator S is linear and*

- (1) *it is bounded and positive, with $\|S\| \leq B$;*
- (2) *it is invertible;*
- (3) *the family $\{S^{-1}\psi^k\}_{k \in M}$ is a continuous frame;*
- (4) *for any $f \in \mathcal{H}$,*

$$f = \int_M \langle f, \psi^k \rangle S^{-1}\psi^k d\rho(k) = \int_M \langle f, S^{-1}\psi^k \rangle \psi^k d\rho(k),$$

where the equality is intended in the weak sense.

For a proof of this see [7, §5.8].

From now on, we suppose $M = \mathbb{R}^2 \times S^1 \times \mathbb{R}^+$ with the measure ρ such that $d\rho(k) = \frac{dk}{\sigma^2}$. In our case, the vectors ψ^k are the Gabor filters indexed by the points $k = (x, y, \theta, \sigma) \in M$. Consider the usual scalar product $\langle \cdot, \cdot \rangle$ in $L^2(\mathbb{R}^2)$, then in Appendix B we prove that there exists a constant $C_\psi \in \mathbb{R}^+$ such that for any pair of inputs I, I' in $L^2(\mathbb{R}^2)$

$$(5.2) \quad \int_M \langle I, \psi^k \rangle \langle \psi^k, I' \rangle \frac{dk}{\sigma^2} = C_\psi \cdot \langle I, I' \rangle.$$

As a corollary,

$$(5.3) \quad SI = C_\psi \cdot I,$$

where the equality as to be intended in the weak sense, that is $h_I = \langle C_\psi I, - \rangle_{L^2}$. In Appendix C we prove that the equality (5.3) is also true in a much stronger sense.

We observe that $\langle I, \psi_{x,y,\theta,\sigma} \rangle$ is exactly the output function μ associated to I , see Definition 2.1. Therefore by Lemma 5.2 and equality (5.3),

$$(5.4) \quad I(\tilde{x}, \tilde{y}) = \frac{1}{C_\psi} \cdot \int_M \mu(x, y, \theta, \sigma) \cdot \psi^k(\tilde{x}, \tilde{y}) \cdot \frac{1}{\sigma^2} dx dy d\theta d\sigma.$$

In the next section, starting from two input functions I_0, I_1 we produce a path μ_t of output functions. In order to produce a path in the input space from μ_t , we define

$$(5.5) \quad I_t := \frac{1}{C_\psi} \cdot \int_M \mu_t(k) \cdot \psi^k \frac{dk}{\sigma^2}.$$

6. DEFORMATION OF THE OUTPUT

In this section we show the existence of a path relying two output functions μ_0, μ_1 . In particular for any output we obtain two probability measures from the positive and negative part of the output. Using the results of Section 3 and 4 we build the paths relying the associated measure, and conclude from there.

We introduce a new condition on the input image $I: \mathbb{R}^2 \rightarrow [0, 1]$. If μ is the output associated to I , that is

$$\mu(k) = \langle I, \psi^k \rangle.$$

We look at μ as a function defined on the whole $X = \mathbb{R}^2 \times S^1 \times \mathbb{R}$ but it is supported only on $M = \mathbb{R}^2 \times S^1 \times \mathbb{R}^+ \subset X$. We define the two functions $\tilde{\mu}^+ := \max(\mu, 0)$ and $\tilde{\mu}^- := \max(-\mu, 0)$. We impose the condition

$$(6.1) \quad \int_X d^2(k, 0) \tilde{\mu}^+(k) dk < \infty,$$

and the same for $\tilde{\mu}^-$.

Lemma 6.1. *If conditions (6.1) holds, then $\int d^2(k, k_0) \tilde{\mu}^+(k) dk$ is finite for any $k_0 \in X$.*

Proof. We start by observing that $\int_X \tilde{\mu}^+(k) dk < \infty$. Indeed, if we consider the compact $B = \{k \mid d(k, 0) \leq 1\}$ and the maximum b of $\tilde{\mu}^+$ over B , we have

$$\begin{aligned} \int_X \tilde{\mu}^+(k) dk &\leq \tilde{\mu}^+(B) \cdot b + \int_{k \notin B} \tilde{\mu}^+(k) dk \\ &\leq \tilde{\mu}^+(B) \cdot b + \int_{k \notin B} d^2(k, 0) \tilde{\mu}^+(k) dk < \infty. \end{aligned}$$

For any $k_0 \in X$, by the triangular inequality we have

$$d^2(k, k_0) \leq d^2(k, 0) + d^2(k_0, 0) + 2 \cdot |d(k, k_0) \cdot d(k_0, 0)| \quad \forall k \in X.$$

Therefore

$$\begin{aligned} \int_X d^2(k, k_0) \tilde{\mu}^+(k) dk &\leq \int_X d^2(k, 0) \tilde{\mu}^+(k) dk + d^2(k_0, 0) \cdot \int_X \tilde{\mu}^+(k) dk + \\ &\quad + 2 \cdot \int_X d(k, 0) \cdot d(k_0, 0) \tilde{\mu}^+(k) dk. \end{aligned}$$

The first two terms are clearly finite. Concerning the last one,

$$\begin{aligned} \int_X d(k, 0) \cdot d(k_0, 0) \tilde{\mu}^+(k) dk &\leq d(k_0, 0) \cdot \tilde{\mu}^+(B) + \int_{k \notin B} d(k, 0) \cdot d(k_0, 0) \tilde{\mu}^+(k) dk \\ &\leq d(k_0, 0) \cdot \tilde{\mu}^+(B) + d(k_0, 0) \cdot \int_{k \notin B} d(k, 0)^2 \tilde{\mu}^+(k) dk < \infty, \end{aligned}$$

and this concludes the proof. \square

Remark 6.2. For $I: \mathbb{R}^2 \rightarrow [0, 1]$ we defined above $\tilde{\mu}^+$ and $\tilde{\mu}^-$. Consider the coefficient

$$m := \int_M \tilde{\mu}^+(k) dk = - \int_M \tilde{\mu}^-(k) dk,$$

which is well defined (see Remark 2.3) and finite as a consequence of condition (6.1) (see the proof of Lemma 6.1). We renormalize, $\mu^+ := \tilde{\mu}^+/m$ and $\mu^- := \tilde{\mu}^-/m$.

Therefore for any function $\mu: M \rightarrow \mathbb{R}$ there exists two probability densities μ^+ and μ^- such that

$$\mu = m \cdot (\mu^+ - \mu^-),$$

where m is a positive coefficient and $\mu^+ \cdot \mu^- \equiv 0$. This means that μ^+ and μ^- are the positive and negative part of μ , renormalized in order to become probability densities.

Given two inputs I_0, I_1 , we define respectively μ_i^+, μ_i^- and m_i for $i = 0, 1$. Using the equality (5.4) we obtain,

$$\begin{aligned} I_0 &= C_\psi^{-1} \cdot m_0 \cdot \left(\int_M \mu_0^+(k) \psi^k \frac{dk}{\sigma^2} - \int_M \mu_0^-(k) \psi^k \frac{dk}{\sigma^2} \right) \\ I_1 &= C_\psi^{-1} \cdot m_1 \cdot \left(\int_M \mu_1^+(k) \psi^k \frac{dk}{\sigma^2} - \int_M \mu_1^-(k) \psi^k \frac{dk}{\sigma^2} \right). \end{aligned}$$

We consider the complete riemannian variety $X = \mathbb{R}^2 \times S^1 \times \mathbb{R}$ with the Lebesgue metric dk . In particular we consider μ_i^+ and μ_i^- , for $i = 0, 1$, as measures on X even if they are defined over $M \subset X$. The function μ_i^\pm is identified with $\mu_i^\pm \cdot dk$ if $\sigma > 0$ and with the null measure elsewhere. As a consequence of condition (6.1), $\mu_i^+, \mu_i^- \in \mathcal{P}_2(X)$ for $i = 0, 1$.

We consider the pairs μ_0^+, μ_1^+ and μ_0^-, μ_1^- , by Theorem 4.3 and Remark 4.4, there exists two constant speed geodesics μ_t^+ and μ_t^- .

Remark 6.3. In particular by Theorem 3.15 there exists a transport map T^+ such that $\mu_1^+ = T^+_{\#} \mu_0^+$ and the same for the negative part with a transport map T^- . Therefore, $\mu_t^+ = \left(e_t^{(T^+)}\right)_{\#} \mu_0^+$ and $\mu_t^- = \left(e_t^{(T^-)}\right)_{\#} \mu_0^-$.

Remark 6.4. By definition of e_t and of the transport map, the measures μ_t^{\pm} are null outside M for any $t \in [0, 1]$, therefore we can always look at them as measures in $\mathcal{P}_2(M)$.

Furthermore, we point out that by construction μ_t^+ and μ_t^- are absolutely continuous with respect to the Lebesgue measure dk . In the following we will use the notation μ_t^+ and μ_t^- indistinctly for the measures and for the density functions (defined over M) when there is no risk of confusion.

We consider a linear variation of the mass m , this means that we define a varying coefficient

$$(6.2) \quad m_t := m_0(1 - t) + m_1 t \quad \forall t \in [0, 1]$$

We define the path of output functions using these coefficients,

$$(6.3) \quad \mu_t := m_t \cdot \left(\left(e_t^{(T^+)}\right)_{\#} \mu_0^+ - \left(e_t^{(T^-)}\right)_{\#} \mu_0^- \right).$$

Finally from Equation (5.5) we obtain a path of input functions from I_0 to I_1 .

7. CONSTRAINING THE OUTPUT

In this section we introduce a useful tool to describe the geodesic μ_t , the so called weak riemannian structure of $(\mathcal{P}_2(X), W_2)$, the space of probabilities endowed with the Wasserstein distance.

If μ_t is an absolutely continuous curve in $\mathcal{P}_2(X)$ (with respect to the Wasserstein distance), consider a time dependent vector field v_t on TX such that the following *continuity equation* is verified in the sense of distributions,

$$(7.1) \quad \frac{d}{dt} \mu_t + \nabla \cdot (v_t \mu_t) = 0.$$

For the proof of the following theorem we refer again to [3].

Theorem 7.1 (see [3, Theorem 2.29]). *If X is a smooth complete riemannian manifold without boundary, then*

- (1) *for every absolutely continuous curve $\mu_t \in \mathcal{P}_2(X)$ there exists a Borel family of vector fields v_t such that $\|v_t\|_{L^2(\mu_t)} \leq |\dot{\mu}_t|$ for a.e. t and the continuity equation (7.1) is satisfied (in the sense of distributions);*
- (2) *if (μ_t, v_t) satisfies (7.1) and $\int_0^1 \|v_t\|_{L^2(\mu_t)} dt$ is finite, then μ_t is an absolutely continuous curve (up to a negligible set of points) and $|\dot{\mu}_t| \leq \|v_t\|_{L^2(\mu_t)}$ for a.e. $t \in [0, 1]$.*

Remark 7.2. We recall the Benamou-Brenier formula proved for example at [3, Proposition 2.30] and stating that the minimization problem solved by a geodesic relying $\mu_0, \mu_1 \in \mathcal{P}_2(X)$ can be reformulated in terms of the vector field v_t . In particular we have

$$W_2(\mu_0, \mu_1) = \inf \int_0^1 \|v_t\|_{L^2(\mu_t)} dt,$$

where the infimum is taken among all weakly continuous distributional solutions of the continuity equation for (μ_t, v_t) .

As a direct consequence of Theorem 7.1, for every absolutely continuous curve μ_t in $\mathcal{P}_2(X)$, there exists a family of vector fields (v_t) verifying the continuity equation and such that $\|v_t\|_{L^2(\mu_t)} = |\dot{\mu}_t|$ for a.e. t . This family is not unique in general, but it is unique if we define as follows the tangent space to $\mathcal{P}_2(X)$ where the vector fields must live in.

Definition 7.3. If $\mu \in \mathcal{P}_2(X)$ then the tangent space to $\mathcal{P}_2(X)$ is defined as

$$\begin{aligned} T_\mu \mathcal{P}_2(X) &:= \overline{\{\nabla \varphi : \varphi \in \mathcal{C}_c^\infty(X)\}}^{L^2(\mu)} \\ &= \left\{ v \in L^2(\mu) : \int \langle v, w \rangle d\mu = 0, \forall w \in L^2(\mu) \text{ s.t. } \nabla \cdot (w\mu) = 0 \right\}. \end{aligned}$$

Therefore for any absolutely continuous curve μ_t in $\mathcal{P}_2(X)$, we have an associated vector field v_t . In particular we have it for the geodesics μ_t^+ and μ_t^- obtained via two inputs I_0, I_1 (see Remark 6.3). Observe that both μ_t^+ and μ_t^- are constant speed geodesics in $(\mathcal{P}_2(X), W_2)$ therefore they are absolutely continuous curves, and so Theorem 7.1 applies to them.

We denote by v_t^+ and v_t^- the vector fields associated respectively to μ_t^+ and μ_t^- . Moreover, we define the normalized image

$$J_t := \frac{I_t}{m_t},$$

so that

$$(7.2) \quad J_t = \int_M \frac{\mu_t(k)}{m_t} \psi^k \frac{dk}{\sigma^2} = \int_M (\mu_t^+ - \mu_t^-) \psi^k \frac{dk}{\sigma^2}.$$

We are interested in the existence of a family J_t of inputs $\mathbb{R}^2 \rightarrow [0, 1]$ that relies J_0 to J_1 , such that J_t is in $\mathcal{C}^1([0, 1]; L^2(\mathbb{R}^2))$ and $\frac{\mu_t(k)}{m_t} = \langle J_t, \psi^k \rangle$ for any $k \in M$.

In order to find such a path, we observe that if it exists, then

$$\frac{d}{dt} \frac{\mu_t(k)}{m_t} = \left\langle \frac{dJ_t}{dt}, \psi^k \right\rangle \quad \forall k \in M \subset X.$$

From the continuity equation we know that

$$\begin{aligned} \frac{d\mu_t^+}{dt} &= -\nabla \cdot (v_t^+ \mu_t^+) \\ \frac{d\mu_t^-}{dt} &= -\nabla \cdot (v_t^- \mu_t^-). \end{aligned}$$

We define $v_t := v_t^+ - v_t^-$, and therefore

$$\frac{d}{dt} \frac{\mu_t}{m_t} = -\nabla \cdot \left(v_t \frac{\mu_t}{m_t} \right).$$

Indeed, by Theorem 7.1, $v_t^+ \in L^2(\mu_t^+)$ and $v_t^- \in L^2(\mu_t^-)$. Therefore it is possible to extend both vector fields to the whole M in such a way that $v_t^+ \cdot \mu_t^- \equiv 0$ and $v_t^- \cdot \mu_t^+ \equiv 0$. Then we have,

$$\frac{d}{dt} \frac{\mu_t(k)}{m_t} = -\nabla \cdot \left(v_t \langle J_t, \psi^k \rangle \right) = -\left\langle J_t, v_t(k) \cdot \nabla \psi^k + \psi^k \cdot (\nabla \cdot v_t(k)) \right\rangle \quad \forall k \in M \subset X.$$

For an opportune vector field $\alpha \in TM$ we have

$$(7.3) \quad \nabla \psi^k = \psi^k \cdot \alpha(k).$$

In particular if we use the notation $(\tilde{x}_k, \tilde{y}_k) = \sigma^{-1} R_{-\theta}(\tilde{x} - x, \tilde{y} - y)$, where $k = (x, y, \theta, \sigma) \in M$, it is straightforward to verify that

$$\begin{aligned}\alpha^x &= 2 \left(\sigma^{-2}(\tilde{x} - x) + \frac{\sigma^{-1} \sin \theta}{\tan(2\tilde{y}_k)} \right) \\ \alpha^y &= 2 \left(\sigma^{-2}(\tilde{y} - y) - \frac{\sigma^{-1} \cos \theta}{\tan(2\tilde{y}_k)} \right) \\ \alpha^\theta &= \frac{2\sigma^{-1}}{\tan(2\tilde{y}_k)} (\cos \theta(\tilde{x} - x) + \sin \theta(\tilde{y} - y)) \\ \alpha^\sigma &= 2 \left(\sigma^{-3} \|\tilde{v} - v\|^2 + \frac{\sigma^{-2}(\sin \theta(\tilde{x} - x) - \cos \theta(\tilde{y} - y))}{\tan(2\tilde{y}_k)} - \frac{3\sigma^{-1}}{4} \right).\end{aligned}$$

In particular for any k , the vector field α is well defined almost everywhere on \mathbb{R}^2 .

Therefore for a.e. $t \in [0, 1]$ and every $k \in M$,

$$\left\langle \frac{dJ_t}{dt}, \psi^k \right\rangle = - \left\langle J_t(\alpha \cdot v_t + \nabla \cdot v_t), \psi^k \right\rangle.$$

Thus, if $\alpha \cdot v_t + \nabla \cdot v_t$ is independent of the variable k as a function $\mathbb{R}^2 \rightarrow \mathbb{R}$, as $(\psi^k)_k$ is a frame, this implies

$$(7.4) \quad \frac{dJ_t}{dt} = -J_t(\alpha \cdot v_t + \nabla \cdot v_t).$$

This last equality is true in the weak sense but also in the (stronger) sense showed in Appendix C.

In order to state our last theorem, we introduce two additional conditions. First we impose that the inputs I_0, I_1 are null outside a compact subset of \mathbb{R}^2 . This simply says that the images we are treating are limited in space.

We also impose that the vector fields take their values in a Sobolev space.

Theorem 7.4. *Consider two inputs $I_0, I_1: \mathbb{R}^2 \rightarrow [0, 1]$ in $L^2(\mathbb{R}^2)$ null outside a compact subset of \mathbb{R}^2 , their associated output functions μ_0, μ_1 , the absolutely continuous curve μ_t relying μ_0 to μ_1 defined in (6.3) and the associated (unique) vector field v_t . Moreover, take the vector field α defined in (7.3).*

We suppose that

$$v \in L^1([0, 1]; W^{1, \infty}(X, \mathbb{R}^4)).$$

If for any $t \in [0, 1]$ the following equality is verified

$$(7.5) \quad \nabla(\alpha \cdot v_t + \nabla \cdot v_t) = 0,$$

then there exists a path $I_t \in C^1([0, 1]; L^2(\mathbb{R}^2))$ relying I_0 to I_1 such that $\mu_t(k) = \langle I_t, \psi^k \rangle$ for any $k \in M$.

Proof. We use the notation $u_t := \alpha \cdot v_t + \nabla \cdot v_t$ and observe that by (7.5), u_t is independent of the point k . As a consequence of this independence, u_t can be defined also where α is singular. We consider the differential equation

$$\frac{dJ_t}{dt} = -u_t \cdot J_t,$$

and the coefficient m_t as defined in (6.2). Given the initial function $J_0 := I_0/m_0$, the solution to the equation above is

$$J_t := J_0 \cdot e^{-h_t},$$

where $h_t := \int_0^t u_s ds \forall t \in [0, 1]$ is a primitive of u_t , and it exists as a consequence of $v_t \in W^{1,\infty}$. The function J_0 is null outside a compact subset $K \subset \mathbb{R}^2$, and h_t is continuous therefore limited over K . This implies $J_t \in L^2(\mathbb{R}^2)$ for any $t \in [0, 1]$.

We define $I_t := m_t \cdot J_t$ and

$$\nu_t(k) := m_t \cdot \langle J_t, \psi^k \rangle = \langle I_t, \psi^k \rangle.$$

By construction, for any $t \in [0, 1]$ ν_t satisfies a.e. the continuity equation (7.1) with respect to the vector field v_t . Moreover, $\nu_0 = \frac{\mu_0}{m_0} = \mu_0^+ - \mu_0^-$. The solution to the continuity equation under this conditions is unique for absolutely continuous measures (see [1]). Therefore $\nu_t = \frac{\mu_t}{m_t}$ a.e. for a.e. $t \in [0, 1]$. We observe that this also proves that $m_1 J_1 = I_1$. \square

8. DISCRETE MODEL

In order to produce an implementation of a transport model in the space $M = \mathbb{R}^2 \times S^1 \times \mathbb{R}^+$, we have to work in a discrete setting, and therefore it is not possible to use the frame ψ^k introduced above, because it is in fact a frame only when the index k varies along all M .

We consider instead a discrete frame that produce a new set of output functions. The frame we consider is the so called Gabor Wavelet Pyramid (see for instance [20]) that relies again in the Gabor complex mother function

$$e^{-\tilde{x}^2 - \tilde{y}^2} \cdot e^{2\pi i \omega \tilde{y}},$$

but we split it in its real and complex component, defining two sets of filter functions associated to the two following mother functions

$$\begin{aligned} \psi_e(\tilde{x}, \tilde{y}) &:= e^{-\tilde{x}^2 - \gamma \cdot \tilde{y}^2} \cdot \cos(2\pi \omega \tilde{y}) \\ \psi_o(\tilde{x}, \tilde{y}) &:= e^{-\tilde{x}^2 - \gamma \cdot \tilde{y}^2} \cdot \sin(2\pi \omega \tilde{y}). \end{aligned}$$

As we have seen, in the case of the continuous frame, the elements of the frame are obtained via the action of the Heisenberg group on the mother function. In the case of wavelets the acting group is the affine group. We recall that by $\psi_{e,\theta}$ and $\psi_{o,\theta}$ we mean the same functions with the rotation by θ applied, that is

$$\begin{aligned} \psi_{e,\theta}(\tilde{x}, \tilde{y}) &= \psi_e(R_{-\theta}(\tilde{x}, \tilde{y})) \\ &= \psi_e(\cos(\theta)\tilde{x} + \sin(\theta)\tilde{y}, -\sin(\theta)\tilde{x} + \cos(\theta)\tilde{y}), \end{aligned}$$

and analogously for ψ_o .

Definition 8.1. Given two real numbers a_0, b_0 and a positive integer number d , we consider $\theta_0 := \pi/d$ and $\theta_\ell = \ell \cdot \theta_0$ for any ℓ positive integer. Then we can define the wavelets associated to ψ_e and ψ_o . For any $n, k, \ell, j \in \mathbb{Z}_{\geq 0}$,

$$\begin{aligned} \psi_e^{n,k,\ell,j}(\tilde{x}, \tilde{y}) &= \frac{1}{a_0^j} \cdot \psi_{e,\theta_\ell} \left(\frac{\tilde{x}}{a_0^j} - nb_0, \frac{\tilde{y}}{a_0^j} - kb_0 \right) \\ \psi_o^{n,k,\ell,j}(\tilde{x}, \tilde{y}) &= \frac{1}{a_0^j} \cdot \psi_{o,\theta_\ell} \left(\frac{\tilde{x}}{a_0^j} - nb_0, \frac{\tilde{y}}{a_0^j} - kb_0 \right) \end{aligned}$$

As developed in [10] by Daubechies for one variable wavelets, and generalized to two variable wavelets by Lee in [21], for an oportune choice of $a_0, b_0 \in \mathbb{R}$, and $n, k, \ell, j \in \mathbb{Z}_{\geq 0}$, the wavelets

above form a frame, that is there exist real numbers $A, B > 0$ such that for any $f \in L^2(\mathbb{R})$ we have

$$A \cdot \|f\|^2 \leq \sum_{n,k,\ell,j} |\langle f, \psi_e^{n,k,\ell,j} \rangle|^2 + \sum_{n,k,\ell,j} |\langle f, \psi_o^{n,k,\ell,j} \rangle|^2 \leq B \cdot \|f\|^2,$$

where the scalar product is the classical

$$\langle f, \psi \rangle = \int f(\tilde{x}, \tilde{y}) \cdot \psi(\tilde{x}, \tilde{y}) d\tilde{x} d\tilde{y}.$$

As a consequence it is possible to reconstruct the function, as

$$(8.1) \quad f = C \cdot \left(\sum_{n,k,\ell,j} \langle f, \psi_e^{n,k,\ell,j} \rangle \cdot \psi_e^{n,k,\ell,j} + \sum_{n,k,\ell,j} \langle f, \psi_o^{n,k,\ell,j} \rangle \cdot \psi_o^{n,k,\ell,j} \right),$$

where C is a constant and the equality is true in the weak sense. Observe that the splitting of the mother function in its even and odd part allows to consider only ‘half’ of the S^1 circle, or equivalently it allows to use an equipartition of the space of directions $S^1/\{\pm 1\}$.

In our setting, we consider in this case four output functions obtained by convoluting any input $I: \mathbb{R}^2 \rightarrow [0, 1]$ with the frame above. Then, we apply the same procedure of Section 6, that is we use optimal transport between functions on M to build a path relying the initial output to the final output.

Instead of working over the whole M , we consider a compact subset

$$M_c := [0, D]^2 \times S^1 \times [\sigma_{\min}, \sigma_{\max}],$$

where $[0, D]^2$ represents the portion of the plane which is registered on the retina, while $[\sigma_{\min}, \sigma_{\max}]$ is the interval where the scale parameter σ varies.

Remark 8.2. The idea behind the Gabor Wavelet Pyramid is that we define the output functions indexed by a discrete subset of M_c . We use the notation $\llbracket 0, N \rrbracket$ for the subset $\{0, 1, \dots, N\} \subset \mathbb{Z}_{\geq 0}$ for any positive integer N . We set $a_0 = \sigma_{\min}$ and $j_\star = \log_{a_0}(\sigma_{\max})$. For any $j \in \llbracket 1, j_\star \rrbracket$ we consider the discrete subset

$$(b_0 a_0^j) \cdot \left[\left[0, \left\lfloor \frac{D}{b_0 a_0^j} \right\rfloor \right] \right]^2 \times \theta_0 \cdot [1, d] \times \{a_0^j\} \subset [0, D]^2 \times S^1 \times \{a_0^j\},$$

thus defining different strata of a discrete subset of the whole M_c .

Remark 8.3. We observe that any wavelet $\psi_e^{n,k,\ell,j}$ (the same is true for ψ_o) is centered at $(nb_0 a_0^j, kb_0 a_0^j)$. Furthermore, given two inputs I_0, I_1 , we can define the following functions on M_c , for $i = 0, 1$,

$$\begin{aligned} \tilde{\mu}_{i,e}^+(nb_0 a_0^j, kb_0 a_0^j, \theta_\ell, a_0^j) &= \max \left(\langle I_i, \psi_e^{n,k,\ell,j} \rangle, 0 \right) \\ \tilde{\mu}_{i,e}^-(nb_0 a_0^j, kb_0 a_0^j, \theta_\ell, a_0^j) &= - \min \left(\langle I_i, \psi_e^{n,k,\ell,j} \rangle, 0 \right) \end{aligned}$$

and analogously for $\tilde{\mu}_{i,o}^+$ and $\tilde{\mu}_{i,o}^-$. We normalize these functions in order to obtain the probabilities $\mu_{i,e}^\pm$ for $i = 0, 1$ (and analogously for $\mu_{i,o}^\pm$).

In order to build the transport path μ_t for any of the pairs of output functions above, we focus on the distance over M_c . In fact, instead of working with the distance treated in the previous sections, another distance turns out to be more efficient in our setting, and at the same time it is a distance equivalent to the previous one.

Definition 8.4. Two distances d_1, d_2 over the space Ω are said to be equivalent if for any compact set $K \subset \Omega$, there exists a constant C such that

$$\frac{1}{C} \cdot d_2(p_1, p_2) \leq d_1(p_1, p_2) \leq C \cdot d_2(p_1, p_2) \quad \forall p_1, p_2 \in \Omega.$$

They are locally equivalent if for any $p_0 \in \Omega$ there exists a neighborhood U of p_0 such that d_1 and d_2 are equivalent over U .

The main reference for the equivalence result we are going to use is [22]. With our notation (see Remark 2.2), we consider in every point the metric

$$\tilde{g}(x, y, \theta, \sigma) := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{h_1^2}{\sigma^2} & 0 \\ 0 & 0 & 0 & \frac{h_2^2}{\sigma^2} \end{pmatrix}$$

and we denote by Y_1, Y_2, Y_3, Y_4 the orthonormal basis at every point defined by $Y_1 = X_1$ and $Y_2 = X_2$, while $Y_3 = \frac{\sigma X_3}{h_1}$ and $Y_4 = \frac{\sigma X_4}{h_2}$. As we are interested in constant coefficient flows, consider

$$Y = c_1 Y_1 + c_2 Y_2 + c_3 Y_3 + c_4 Y_4,$$

with $c_i \in \mathbb{R}^+$ for $i = 1, \dots, 4$. In order to evaluate the distance d_c between two points $p_0 = (x_0, y_0, \theta_0, \sigma_0)$ and $p_1 = (x_1, y_1, \theta_1, \sigma_1)$, we consider the constant coefficient vector field Y that induces a curve $p_t: [0, 1] \rightarrow M$ relying them. We develop these evaluations in Appendix D.

Proposition 8.5 (See, [22, Theorem 2]). *The distance induced by the riemannian metric \tilde{g} and the distance d_c are locally equivalent.*

In the next section we are going to show an implementation of our transport model based on the distance d_c .

Remark 8.6. From the previous results we know that there exists four unique paths $\mu_{t,e}^\pm, \mu_{t,o}^\pm$ relying the associated probabilities. Using the fact that the chosen set of filters is a frame, we reconstruct the intermediary images, by defining

$$\begin{aligned} I_t := & \sum_{n,k,\ell,j} m_{e,t}(\mu_{t,e}^+ - \mu_{t,e}^-)(nb_0 a_0, kb_0 a_0, \theta_\ell, a_0^j) \cdot \psi_e^{n,k,\ell,j} \\ & + \sum_{n,k,\ell,j} m_{o,t}(\mu_{t,o}^+ - \mu_{t,o}^-)(nb_0 a_0, kb_0 a_0, \theta_\ell, a_0^j) \cdot \psi_o^{n,k,\ell,m}. \end{aligned}$$

Where the coefficients $m_{e,t}$ and $m_{o,t}$ depend on the masses of $\tilde{\mu}_{i,e}^\pm$ and $\tilde{\mu}_{i,o}^\pm$ for $i = 0, 1$.

9. IMPLEMENTATION

In order to implement our model, we code a Gabor Wavelet Pyramid as sketched in the previous section, and evaluate the transport maps using a Sinkhorn's algorithm of the kind

treated by Peyré and Cuturi [25]. In discrete setting, the transport plan between two functions u_1, u_2 , is a matrix P . In particular, if the two functions are represented as vectors of dimension m , the matrix P has dimension $m \times m$ and verifies the property

$$(9.1) \quad P\mathbf{1}_m = u_1, \quad P^T\mathbf{1}_m = u_2$$

where $\mathbf{1}_m$ is the vector of length m whose coordinates are all 1, and therefore (9.1) means that $P \in \text{Adm}(u_1, u_2)$.

We recall that the transport plan P between two functions u_1, u_2 minimizes the product $\langle P, C \rangle$ where C is the cost matrix associated to our setting, and P respects condition (9.1). In the case of the Sinkhorn's algorithm, with regularization coefficient ε , the minimized quantity is

$$(9.2) \quad \langle P, C \rangle - \varepsilon H(P),$$

where $H(P)$ is the entropy of P . The uniqueness of the solution is in fact true for any strictly concave function H (see [12] for a wider analysis on entropy regularizations). A well known result [25, §4] states that if $L_C(u_1, u_2)$ is the minimum reached by $\langle P, C \rangle$ under condition (9.1), and $L_C^\varepsilon(u_1, u_2)$ is the minimum of (9.2) under the same condition, then

$$L_C^\varepsilon(u_1, u_2) \xrightarrow{\varepsilon \rightarrow 0} L_C(u_1, u_2).$$

We observe that the plan achieving $L_C(u_1, u_2)$ is not always unique in the discrete setting, while the one achieving $L_C^\varepsilon(u_1, u_2)$ is in fact unique as proved for example by [25, Proposition 4.3]. Moreover, if P^ε is the optimal plan for L_C^ε , then $P^\varepsilon \rightarrow P^*$ where P^* is the maximum entropy plan among those achieving the optimum L_C .

In order to compute the geodesic between $\mu_{0,e}^+$ and $\mu_{1,e}^+$ we define the optimal plan for this case

$$P_e^+ := \operatorname{argmin}_{P \in \text{Adm}(\mu_{0,e}^+, \mu_{1,e}^+)} (\langle P, C \rangle - \varepsilon \cdot H(P)).$$

and consider the meshgrid of Remark 8.2. The pseudocode for the computation is the following.

Algorithm 1 Computation of the geodesic $\mu_{t,e}^+$

Input the optimal plan P_e^+ ; $\sigma_{\min}, \sigma_{\max}, b_0 \in \mathbb{R}^+$; $D, d \in \mathbb{Z}_{>0}$;
 Set $\mu_{t,e}^+ \equiv 0$;
 Set $a_0 = \sigma_{\min}$;
 Set $j_\star = \log_{a_0}(\sigma_{\max})$;
for $j, j' = 1, \dots, j_\star$ **do**
 for $\ell, \ell' = 1, \dots, d$ **do**
 for $n, k = 0, \dots, \left\lfloor \frac{D}{b_0 a_0^j} \right\rfloor$ and $n', k' = 0, \dots, \left\lfloor \frac{D}{b_0 a_0^{j'}} \right\rfloor$ **do**
 Set $p, q = (nb_0 a_0^j, kb_0 a_0^j, \ell \theta_0, a_0^j), (n' b_0 a_0^{j'}, k' b_0 a_0^{j'}, \ell' \theta_0, a_0^{j'})$;
 Take the geodesic γ relying p to q and its point γ_t ;
 $\mu_{t,e}^+(\gamma_t) := \mu_{t,e}^+(\gamma_t) + P_e^+(p, q)$;
 end for
end for
end for
Output $\mu_{t,e}^+$.

We do the analogously for $\mu_{t,e}^-$ and $\mu_{t,o}^\pm$. In order to develop the code we implemented various algorithms developed in [25], in particular we used part of the code developed by the same authors for the Sinkhorn’s algorithm and disposable online [24].

In our simulation, we set $a_0 = 2, b_0 = 1$ and as already said $D = 32$ or 64 . We also set $\ell = 8$ which is coherent with previous results on the orientation sampling in primates (see [5, 27]). Furthermore, we set $\sigma_{\min} = 1, 1244$ and $\sigma_{\max} = \sigma_{\min} \cdot D$. The pseudocode for the complete simulation is the following.

Algorithm 2 Computation of I_t

Input $h_1, h_2, b_0, \sigma_{\min}, \sigma_{\max}, \varepsilon \in \mathbb{R}^+$; $N_{\text{iter}}, D, d \in \mathbb{Z}_{>0}$; $I_0, I_1: \llbracket 1, D \rrbracket \rightarrow [0, 1]$;
 Read the two input image I_0, I_1 that are $D \times D$ matrices (with usually $D = 32, 64$);
 Evaluation of the cost matrix c (see Appendix D) over the discrete subset of M_c built in Remark 8.2;
 Evaluation of $\mu_{i,e}^\pm$ and $\mu_{i,o}^\pm$ (see Remark 8.3);
 Evaluation of P_e^\pm and P_o^\pm via the Sinkhorn’s algorithm with N_{iter} iterations and ε regularization coefficient;
 Evaluation of $\mu_{t,e}^\pm$ and $\mu_{t,o}^\pm$ via Algorithm 1;
 Evaluation of I_t via the formula of Remark 8.6;
Output I_t .

The images we consider are simple shapes of the letters ‘T’ and ‘E’, the second ones rotated of about $\frac{\pi}{4}$ counterclockwise. We show the implementation for this input in order to emphasize the ability of our model to reconstruct rotational displacement. We also add an hammer-type shape, in order to show that the implementation works well also in the case of a multi-scale object, that is shapes with different thickness along the figure are also well preserved.

Furthermore, we compare this numerical cortical-style implementation with a classical 2-dimensional planar regularized optimal transport implementation, following the general theory treated for example in [25, §3], and evaluating the transport matrix again via the Sinkhorn’s algorithm. In this case we take the two inputs I_0, I_1 and the probabilities $\nu_0, \nu_1 \in \mathcal{P}(\llbracket 1, D \rrbracket)$ obtained by normalizing the inputs. The cost function is the quadratic cost $c: \llbracket 1, D \rrbracket^2 \rightarrow \mathbb{R}^+$ such that $c(i, j) = (i - j)^2$.

We point out that we applied a smoothing threshold in order to reduce the blur due to the entropic regularization in both our implementations. In particular, we passed the pictures through the sigmoid

$$\Sigma(z) = \frac{1}{1 + e^{-k \cdot (z - z_0)}}$$

where $y \in [0, 1]$ is the pixel intensity and we choose $k = 30$ and $z_0 = 0.65$.

In these reconstructed images obtained with the cortical model, we observe the conservation of the image structure along the rotational movement, while in the standard 2-d optimal transport the basic shape is lost. Therefore, lifting the input via Gabor filters and moving the output function through optimal transport tools, seems to allow an effective image deformation preserving the fundamental aspects of rigid rotational motion.

If we consider a rigid translation, our procedure works well in the case of simple shapes that are moved along their principal direction, as it is the case for the ‘T’ shape in Figure 6.

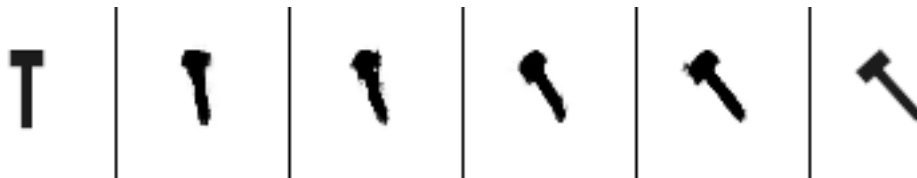


FIGURE 1. Reconstruction of missing images with the parameters $D = 64$, $h_1 = 0.7$, $h_2 = 5$, $N_{\text{iter}} = 1000$ and $\varepsilon = 0.04$. We also processed the image applying a smooth threshold via the Σ sigmoid function.



FIGURE 2. Reconstruction of missing images in the case of the hammer shape, with the parameters $D = 64$, $h_1 = 0.7$, $h_2 = 5$, $N_{\text{iter}} = 1000$ and $\varepsilon = 0.04$. We also processed the image applying a smooth threshold via the Σ sigmoid function.



FIGURE 3. Reconstruction of missing images via the application of a planar optimal transport in the case of the hammer shape, with cost induced by the square of the euclidean distance, and using the the Sinkhorn's algorithm with $N_{\text{iter}} = 10000$ and $\varepsilon = 0.01$. We also processed the image through the sigmoid Σ .

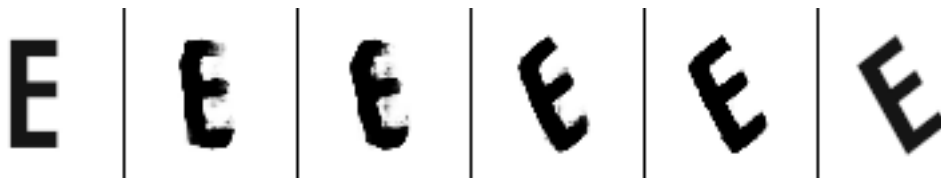


FIGURE 4. Reconstruction of missing images with the parameters $D = 64$, $h_1 = 0.7$, $h_2 = 5$, $N_{\text{iter}} = 1000$ and $\varepsilon = 0.04$. We also processed the image through the sigmoid Σ .

10. CONCLUSIONS AND FUTURE DEVELOPMENTS

We formulated the general theory of a lifting of retinal inputs in a 4-dimensional cortical space, and of the time completion between two cortical outputs μ_0, μ_1 (corresponding to inputs I_0, I_1 in the retinal plane). We did the lifting through Gabor filters, and we considered the frame conditions that allow to project cortical measures back to the retinal space, thus

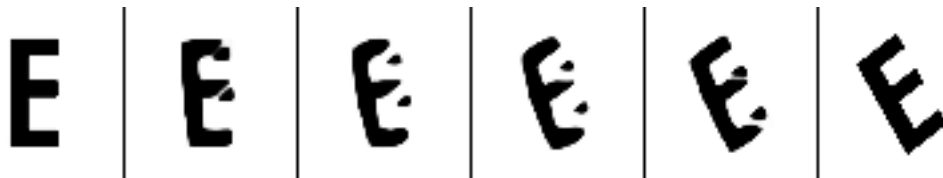


FIGURE 5. Reconstruction of missing images via the application of a planar optimal transport, with cost induced by the square of the euclidean distance, and using the the Sinkhorn’s algorithm with $N_{\text{iter}} = 10000$ and $\varepsilon = 0.01$. We also processed the image through the sigmoid Σ .

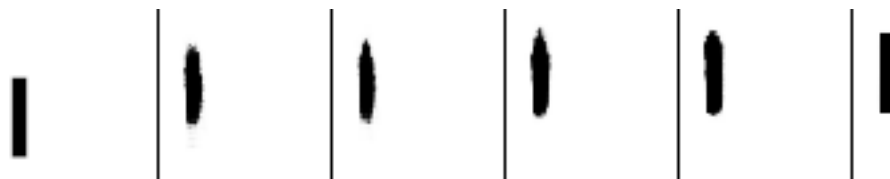


FIGURE 6. Reconstruction of missing images in the case of the ‘I’ shape, with the parameters $D = 64$, $h_1 = 0.7$, $h_2 = 5$, $N_{\text{iter}} = 1000$ and $\varepsilon = 0.01$. We also processed the image through the sigmoid Σ .

obtaining from the completion paths μ_t^\pm on the cortical space, a retinal path I_t between the original inputs.

We obtained the cortical paths via methods of optimal transport, where the cost function is the squared distance on the cortical space. We implemented these tools using a Sinkhorn’s algorithm on a discretized version of the cortical manifold, and via a Gabor Wavelet Pyramid system we also implemented the retinal path I_t . We tested positively our model on rigid rotational movements of multi-scale shapes, verifying the shape conservation.

The Sinkhorn’s algorithm does not minimize the original cost $\langle c, \gamma \rangle$, but a regularized version of it with an entropic correction $\langle c, \gamma \rangle - \varepsilon \cdot h(\gamma)$. If γ^ε is a solution to the regularized problem, we know that $\gamma^\varepsilon \xrightarrow{\varepsilon \rightarrow 0} \gamma^*$ where γ^* is a solution of the original problem. The difference between γ^ε and γ^* is what induces a blurring on the resulting moving shapes in our implementation. We de-blurred the images by filtering them through a sigmoid function.

Knowing that $\|\gamma^\varepsilon - \gamma^*\|$ is controlled by ε , in future works we intend to follow the Rigollet-Weed model [26] and use another optimal transport technique to do the de-blurring. Indeed, we can suppose that the noise factor σ^2 in [26] depends directly on ε , and from there we can develop the Wasserstein distance minimization as conceived in the work above.

In the future we also intend to improve the representation of translational movements. In our actual implementations, translation works correctly particularly along the boundary directions. In order to extend our model by considering spatio-temporal Gabor filtering, as done for example in [4], we could improve translation in the direction ortogonal to boundaries, still implementing the optimal transport tools considered in this work and thus preserving the boundary shapes of the retinal inputs.

Anyway, the problem with displacements orthogonal to the object boundaries is linked also to the structure of the Gabor Wavelet Pyramid: in order to optimize the sampling, the pyramid frame is usually built by considering σ values that are “far” from 0, this emphasizes the boundary constrains. We will work in frame buildings that allow for different σ ranges.

APPENDIX A.

In this first appendix, we prove that the integral

$$(A.1) \quad \int_M \psi^k(\tilde{x}, \tilde{y}) dk,$$

is well defined for any pair $(\tilde{x}, \tilde{y}) \in \mathbb{R}^2$ and has finite value 0, independently from the pair (\tilde{x}, \tilde{y}) .

Lemma A.1. *For any pairs (\tilde{x}, \tilde{y}) and (\tilde{x}', \tilde{y}') in \mathbb{R}^2 , we have*

$$\int_M \psi^k(\tilde{x}, \tilde{y}) dk = \int_M \psi^k(\tilde{x}', \tilde{y}') dk.$$

Proof. If we consider the variable change $(x, y) \mapsto (x - \tilde{x}, y - \tilde{y})$, then

$$\begin{aligned} \int_M \psi^k(\tilde{x}, \tilde{y}) dk &= \int_M \frac{1}{\sigma^{3/2}} \psi_0(A_{\theta, \sigma}^{-1}(\tilde{x} - x, \tilde{y} - y)) dx dy d\theta d\sigma \\ &= \int_M \frac{1}{\sigma^{3/2}} \psi_0(A_{\theta, \sigma}^{-1}(-x, -y)) dx dy d\theta d\sigma \\ &= \int_M \psi^k(0, 0) dk, \end{aligned}$$

and this prove the independence from (\tilde{x}, \tilde{y}) . \square

In order to prove the convergence of (A.1), we rewrite it by making the change to polar coordinates $(\tilde{x}, \tilde{y}) \mapsto (r, \alpha)$, such that

$$\tilde{x} = r \cos(\alpha), \quad \tilde{y} = r \sin(\alpha).$$

In order to have a simpler notation, we write $(\tilde{x}_\theta, \tilde{y}_\theta) = R_\theta(\tilde{x}, \tilde{y})$, that is $\tilde{y}_\theta = r \sin(\alpha + \theta)$. In this notation we have

$$\begin{aligned} \int_M \psi^k(0, 0) dk &= \int_M \frac{1}{\sigma^{3/2}} \cdot e^{-\sigma^{-2}(\tilde{x}^2 + \tilde{y}^2)} \sin(-2\sigma^{-1}\tilde{y}_\theta) d\tilde{x} d\tilde{y} d\theta d\sigma \\ &= \int_M \frac{r e^{-\sigma^{-2}r^2}}{\sigma^{3/2}} \sin\left(-\frac{2r}{\sigma} \sin(\alpha + \theta)\right) dr d\alpha d\theta d\sigma. \end{aligned}$$

Proposition A.2. *The following integral converges to 0,*

$$\int_M \psi^k(0, 0) dk.$$

Proof. We rewrite again the integral as

$$\int_{S^1 \times \mathbb{R}^+} d\theta \sigma^{1/2} d\sigma \int_{S^1 \times \mathbb{R}^+} \frac{1}{\sigma} dr d\alpha \frac{r e^{-\sigma^{-2}r^2}}{\sigma} \sin\left(-\frac{2r}{\sigma} \sin(\alpha + \theta)\right).$$

We consider the coordinate changes $\alpha' = \alpha - \theta$ and $s = \frac{r}{\sigma}$, obtaining the integral

$$(A.2) \quad \int_{S^1 \times \mathbb{R}^+} 2\pi \sigma^{1/2} d\sigma \int_{S^1 \times \mathbb{R}^+} ds d\alpha' s e^{-s^2} \sin(-2s \sin \alpha').$$

We have

$$\left| \int_{S^1 \times \mathbb{R}^+} ds d\alpha' s e^{-s^2} \sin(-2s \sin \alpha') \right| \leq 2\pi \int_0^\infty s e^{-s^2} ds = \pi,$$

therefore the internal integral in (A.2) has finite value. As the sinus is an odd function, its value is 0. Therefore by the Fubini's theorem we have $\int_M \psi^k(0,0) = 0$ too. \square

APPENDIX B.

The goal of this appendix is to prove that the integral

$$(B.1) \quad C_\psi = \int_{\mathbb{R}^2} \frac{|\widehat{\psi}_0(\xi)|^2}{|\xi|^2} d\xi$$

is finite and for any $f, g \in L^2(\mathbb{R}^2)$,

$$(B.2) \quad \int_M \frac{dk}{\sigma^2} \langle f, \psi^k \rangle \langle \psi^k, g \rangle = C_\psi \cdot \langle f, g \rangle.$$

Observe preliminarily that $\widehat{\psi}_0(0) = 0$. Indeed,

$$\int_{\mathbb{R}^2} e^{-\tilde{x}^2 - \tilde{y}^2} \sin(2\tilde{y}) d\tilde{x} d\tilde{y} = 0,$$

because the sinus is an odd function.

Lemma B.1. *The integral (B.1) is finite.*

Proof. We use the notation $\xi = (\xi_1, \xi_2)$. Moreover, we observe $\sin(2\tilde{y}) = \frac{1}{2i}(e^{2i\tilde{y}} - e^{-2i\tilde{y}})$. Therefore, we have

$$\widehat{\psi}_0(\xi) = \int_{\mathbb{R}^2} \psi_0(\tilde{x}, \tilde{y}) e^{-2\pi i(\tilde{x}, \tilde{y}) \cdot \xi} d\tilde{x} d\tilde{y} = \int_{\mathbb{R}^2} e^{-\tilde{x}^2 - \tilde{y}^2 - 2\pi i \xi_1 \tilde{x} - 2\pi i \xi_2 \tilde{y}} \cdot \frac{e^{2i\tilde{y}} - e^{-2i\tilde{y}}}{2i} d\tilde{x} d\tilde{y}.$$

By completing the squares, we obtain

$$\widehat{\psi}_0(\xi) = \frac{1}{2i} e^{-\pi^2 \xi_1^2} \int_{\mathbb{R}^2} e^{-(\tilde{x} + i\pi \xi_1)^2} \left(e^{-(\pi+1)^2 \xi_2^2} e^{-(\tilde{y} + i(\pi+1)\xi_2)^2} - e^{-(\pi-1)^2 \xi_2^2} e^{-(\tilde{y} + i(\pi-1)\xi_2)^2} \right) d\tilde{x} d\tilde{y}.$$

For $c \in \mathbb{R}$, we denote by $D(c)$ the value

$$D(c) := \int_{\mathbb{R}} e^{-(\tilde{x} + ic)^2} d\tilde{x},$$

which is known to be finite. Therefore

$$\widehat{\psi}_0(\xi) = \frac{1}{2i} e^{-\pi^2 \xi_1^2} D(\pi \xi_1) \left(e^{-(\pi+1)^2 \xi_2^2} D((\pi+1)\xi_2) - e^{-(\pi-1)^2 \xi_2^2} D((\pi-1)\xi_2) \right).$$

We know that $D(0) = \sqrt{\pi}$, the integral of the Gaussian function. Therefore the development of $|\widehat{\psi}_0|$ around 0 is

$$|\widehat{\psi}_0(\xi)| = 2\pi^2 \xi_2^2 + o(|\xi|^3).$$

Moreover $D(c) \leq D(0)$ for any $c \in \mathbb{R}$, then at infinity the function behaves as the difference of two Gaussians. This proves that

$$\int_{\mathbb{R}^2} \frac{|\widehat{\psi}_0(\xi)|^2}{|\xi|^2} d\xi$$

is a convergent integral. \square

Proposition B.2. *The equality (B.2) is satisfied for any pair of functions $f, g \in L^2(\mathbb{R}^2)$.*

Proof. We start by observing that, for any $k = (x, y, \theta, \sigma)$,

$$(B.3) \quad \widehat{\psi}^k(\xi) = \sigma^{1/2} \cdot \widehat{\psi}_0(\sigma R_{-\theta}\xi) \cdot e^{-2\pi i((x,y)\cdot\xi)}.$$

Indeed, with the coordinate change $(\tilde{x}', \tilde{y}') = \sigma^{-1}R_{-\theta}(\tilde{x} - x, \tilde{y} - y)$ we have

$$\begin{aligned} \widehat{\psi}^k(\xi) &= \int_{\mathbb{R}^2} \psi_0(\sigma^{-1}R_{-\theta}(\tilde{x} - x, \tilde{y} - y)) e^{-2\pi i((\tilde{x}, \tilde{y})\cdot\xi)} d\tilde{x}d\tilde{y} \\ &= \left(\int_{\mathbb{R}^2} \psi_0(\tilde{x}', \tilde{y}') e^{-2\pi i((\tilde{x}', \tilde{y}')\cdot(\sigma R_{-\theta}\xi))} \sigma^{1/2} d\tilde{x}' d\tilde{y}' \right) e^{-2\pi i((x,y)\cdot\xi)}, \end{aligned}$$

and therefore we obtain (B.3).

Moreover we observe that

$$\int_{\mathbb{R}^2} \widehat{\psi}^k(\xi) \cdot \overline{f(\xi)} d\xi = \int_{\mathbb{R}^2} \sigma^{1/2} \cdot \widehat{\psi}_0(\sigma R_{-\theta}\xi) \cdot \overline{f(\xi)} \cdot e^{-2\pi i((x,y)\cdot\xi)} d\xi$$

is the Fourier transform of $F_f(\xi) := \sigma^{1/2} \cdot \widehat{\psi}_0(\sigma R_{-\theta}\xi) \cdot \overline{f(\xi)}$ evaluated in (x, y) .

In order to prove (B.2), we observe that by the Plancherel Theorem,

$$\begin{aligned} \int_M \langle f, \psi^k \rangle \langle \psi^k, g \rangle \frac{dk}{\sigma^2} &= \int_M \frac{d\theta d\sigma}{\sigma^2} \left(\int f(\xi) \overline{\widehat{\psi}^k(\xi)} d\xi \right) \left(\int \widehat{\psi}^k(\xi) \overline{g(\xi)} d\xi \right) \\ &= \int_M \frac{d\theta d\sigma}{\sigma^2} \widehat{F}_f(x, y) \overline{\widehat{F}_g(x, y)} d\xi \\ &= \int_M \frac{d\theta d\sigma}{\sigma^2} \sigma \cdot f(\xi) \cdot \overline{g(\xi)} \cdot \left| \widehat{\psi}_0(\sigma R_{-\theta}\xi) \right|^2 d\xi. \end{aligned}$$

By Fubini's Theorem we have

$$\begin{aligned} \int_M \langle f, \psi^k \rangle \langle \psi^k, g \rangle \frac{dk}{\sigma^2} &= \int_{\mathbb{R}^2} d\xi f(\xi) g(\xi) \int_0^{2\pi} \int_0^\infty \frac{d\theta d\sigma}{\sigma} \left| \widehat{\psi}_0(\sigma R_{-\theta}\xi) \right|^2 \\ &= \int_{\mathbb{R}^2} d\xi f(\xi) g(\xi) \int_{\mathbb{R}^2} \frac{\left| \widehat{\psi}_0(h) \right|^2}{|h|} dh \\ &= \langle f, g \rangle \cdot C_\psi, \end{aligned}$$

where we used the change of coordinates $(\theta, \sigma) \mapsto h = \sigma R_{-\theta}\xi$. □

APPENDIX C.

In this appendix we prove that equality (5.2) holds not only in the weak sense, but also in a much stronger version. In particular, for any $\sigma_1, \sigma_2, B \in \mathbb{R}^+$ and $f \in L^2(\mathbb{R}^2)$, we define

$$\int_0^{2\pi} d\theta \int_{\sigma_1}^{\sigma_2} \frac{d\sigma}{\sigma^2} \int_{\|(x,y)\| \leq B} dx dy \langle f, \psi^k \rangle \psi^k$$

as the unique element in $L^2(\mathbb{R}^2)$ whose inner product with any $g \in L^2(\mathbb{R}^2)$ is

$$\int_0^{2\pi} d\theta \int_{\sigma_1}^{\sigma_2} \frac{d\sigma}{\sigma^2} \int_{\|(x,y)\| \leq B} dx dy \langle f, \psi^k \rangle \langle \psi^k, g \rangle.$$

Proposition C.1. *For any $f \in L^2(\mathbb{R}^2)$,*

$$\lim_{\substack{\sigma_1 \rightarrow 0 \\ \sigma_2, B \rightarrow \infty}} \left\| f - C_\psi^{-1} \int_0^{2\pi} \int_{\sigma_1}^{\sigma_2} \int_{\|(x,y)\| \leq B} \langle f, \psi^k \rangle \psi^k \frac{1}{\sigma^2} dx dy d\theta d\sigma \right\| = 0.$$

Proof. The term we are estimating is

$$\sup_{\|g\|=1} \left| \left\langle f - C_\psi^{-1} \int_0^{2\pi} \int_{\sigma_1}^{\sigma_2} \int_{\|(x,y)\| \leq B} \langle f, \psi^k \rangle \psi^k \frac{dk}{\sigma^2}, g \right\rangle \right|$$

which is bounded above by

$$\sup_{\|g\|=1} \left| C_\psi^{-1} \int_{k \in U_{\sigma_1, \sigma_2, B}} \langle f, \psi^k \rangle \langle \psi^k, g \rangle \frac{dk}{\sigma^2} \right|,$$

where $U_{\sigma_1, \sigma_2, B} \subset M$ is defined by

$$U_{\sigma_1, \sigma_2, B} = \{k \in M \mid \sigma < \sigma_1 \text{ or } \sigma > \sigma_2 \text{ or } \|(x, y)\| > B\}.$$

By Cauchy-Schwarz inequality this is again bounded by

$$\leq \sup_{\|g\|=1} \left| C_\psi^{-1} \int_{k \in U_{\sigma_1, \sigma_2, B}} |\langle f, \psi^k \rangle|^2 \frac{dk}{\sigma^2} \right|^{1/2} \cdot \left| C_\psi^{-1} \int_M |\langle g, \psi^k \rangle|^2 \frac{dk}{\sigma^2} \right|^{1/2}.$$

The second term is $\|g\| = 1$ by equality (B.2). The first term tends to 0 for $\sigma_1 \rightarrow 0$ and $\sigma_2, B \rightarrow \infty$, because the infinite integral $\int_M |\langle f, \psi^k \rangle|^2 \frac{dk}{\sigma^2}$ converges to $\|f\|^2$. \square

APPENDIX D.

In order to evaluate the distance d_c between two points $p_0 = (x_0, y_0, \theta_0, \sigma_0)$ and $p_1 = (x_1, y_1, \theta_1, \sigma_1)$, we search for the constant coefficients that allows a vector field Y whose flow relies p_0 to p_1 .

We start by observing that

$$\dot{\theta} = c_2,$$

therefore $c_2 = \theta_1 - \theta_0$ and

$$\theta_t = c_2 \cdot t + \theta_0.$$

Regarding the σ variable, we have

$$\dot{\sigma} = \frac{c_4}{h_2} \cdot \sigma,$$

therefore $c_4 = \ln\left(\frac{\sigma_1}{\sigma_0}\right) \cdot h_2$ and

$$\sigma_t = \sigma_0 \cdot e^{\frac{c_4 t}{h_2}}.$$

In what follows we will use the notation $\tilde{c}_4 := \frac{c_4}{h_2}$.

Regarding the first two coordinates

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} \cos \theta & -\frac{\sigma}{h_1} \cdot \sin \theta \\ \sin \theta & \frac{\sigma}{h_1} \cdot \cos \theta \end{pmatrix} \begin{pmatrix} c_1 \\ c_3 \end{pmatrix}.$$

We denote by S_t the matrix obtained by integrating the matrix above up to time t , therefore

$$S_t = \begin{pmatrix} S_t^{11} & S_t^{12} \\ S_t^{21} & S_t^{22} \end{pmatrix}$$

with

$$\begin{aligned}
S_t^{11} &= \left[\frac{\sin(\theta_t)}{c_2} \right]_0^t \\
S_t^{12} &= \left[\frac{1}{h_1 \cdot (\tilde{c}_4^2 + c_2^2)} \cdot (c_2 \cdot \sigma_t \cos(\theta_t) - \tilde{c}_4 \cdot \sigma_t \sin(\theta_t)) \right]_0^t \\
S_t^{21} &= \left[\frac{-\cos(\theta_t)}{c_2} \right]_0^t \\
S_t^{22} &= \left[\frac{1}{h_1 \cdot (\tilde{c}_4^2 + c_2^2)} \cdot (\tilde{c}_4 \cdot \sigma_t \cos(\theta_t) + c_2 \cdot \sigma_t \sin(\theta_t)) \right]_0^t.
\end{aligned}$$

We obtain

$$\begin{pmatrix} c_1 \\ c_3 \end{pmatrix} = S_1^{-1} \begin{pmatrix} x_1 - x_0 \\ y_1 - y_0 \end{pmatrix}$$

and

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = S_t \begin{pmatrix} c_1 \\ c_3 \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}.$$

For any two points as above, their distance is

$$d_c(p_1, p_0)^2 = c_1^2 + c_2^2 + c_3^2 + c_4^2.$$

The datasets and codes generated during and analysed during the current study are available from the corresponding author on reasonable request.

REFERENCES

- [1] Luigi Ambrosio and Gianluca Crippa. *Existence, Uniqueness, Stability and Differentiability Properties of the Flow Associated to Weakly Differentiable Vector Fields*, pages 3–57. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.
- [2] Luigi Ambrosio and Nicola Gigli. Construction of the parallel transport in the wasserstein space. *Methods Appl. Anal.*, 15(1):1–30, 03 2008.
- [3] Luigi Ambrosio and Nicola Gigli. *A User’s Guide to Optimal Transport*, pages 1–155. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
- [4] Davide Barbieri, Giovanna Citti, Giacomo Cocci, and Alessandro Sarti. A cortical-inspired geometry for contour perception and motion integration. *Journal of mathematical imaging and vision*, 49(3):511–529, 2014.
- [5] Davide Barbieri, Giovanna Citti, and Alessandro Sarti. How uncertainty bounds the shape index of simple cells. *The Journal of Mathematical Neuroscience*, 4(1):1–15, 2014.
- [6] Emre Baspinar, Alessandro Sarti, and Giovanna Citti. A sub-riemannian model of the visual cortex with frequency and phase. *arXiv preprint arXiv:1910.04992*, 2019.
- [7] O. Christensen. *An Introduction to Frames and Riesz Bases*. Applied and Numerical Harmonic Analysis. Birkhäuser Boston, 2002.
- [8] G. Citti and A. Sarti. A cortical based model of perceptual completion in the roto-translation space. *Journal of Mathematical Imaging and Vision*, 24(3):307–326, May 2006.
- [9] Giovanna Citti and Alessandro Sarti. *Neuromathematics of vision*, volume 32. Springer, 2014.
- [10] I. Daubechies. The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory*, 36(5):961–1005, 1990.
- [11] John G Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *JOSA A*, 2(7):1160–1169, 1985.

- [12] Simone Di Marino and Augusto Gerolin. Optimal transport losses and sinkhorn algorithm with general convex regularization. *arXiv preprint arXiv:2007.00976*, 2020.
- [13] Marta Favali, Samaneh Abbasi-Sureshjani, Bart ter Haar Romeny, and Alessandro Sarti. Analysis of vessel connectivities in retinal images by cortically inspired spectral clustering. *Journal of Mathematical Imaging and Vision*, 56(1):158–172, 2016.
- [14] Wilfrid Gangbo and Robert J. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996.
- [15] William C Hoffman. The lie algebra of visual perception. *Journal of mathematical Psychology*, 3(1):65–98, 1966.
- [16] William C Hoffman. The visual cortex is a contact bundle. *Applied Mathematics and Computation*, 32(2-3):137–167, 1989.
- [17] David H Hubel. Eye, brain and vision, volume 22 of scientific american library. *Scientific American Press, New York*, 1988.
- [18] David Hunter Hubel and Torsten Nils Wiesel. Ferrier lecture-functional architecture of macaque monkey visual cortex. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 198(1130):1–59, 1977.
- [19] Gaetano Kanizsa. Grammatica del vedere. saggi su percezione e gestalt. *Il Mulino, Bologna*, 1980.
- [20] Kendrick N Kay, Thomas Naselaris, Ryan J Prenger, and Jack L Gallant. Identifying natural images from human brain activity. *Nature*, 452(7185):352–355, 2008.
- [21] Tai Sing Lee. Image representation using 2d gabor wavelets. *IEEE Transactions on pattern analysis and machine intelligence*, 18(10):959–971, 1996.
- [22] Alexander Nagel, Elias M Stein, Stephen Wainger, et al. Balls and metrics defined by vector fields i: Basic properties. *Acta Mathematica*, 155:103–147, 1985.
- [23] Jean Petitot and Yannick Tondut. Vers une neurogéométrie. fibrations corticales, structures de contact et contours subjectifs modaux. *Mathématiques et sciences humaines*, 145:5–101, 1999.
- [24] G. Peyré and M. Cuturi. Computational optimal transport, web site of the book “computational optimal transport”. <https://github.com/optimaltransport>. Accessed: 2021-06-21.
- [25] G. Peyré and M. Cuturi. *Computational Optimal Transport: With Applications to Data Science*. Foundations and trends in machine learning. Now, the essence of knowledge., 2019.
- [26] Philippe Rigollet and Jonathan Weed. Entropic optimal transport is maximum-likelihood deconvolution. *Comptes Rendus Mathématique*, 356(11-12):1228–1235, 2018.
- [27] Dario L Ringach. Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of neurophysiology*, 2002.
- [28] Gonzalo Sanguinetti, Giovanna Citti, and Alessandro Sarti. A model of natural image edge co-occurrence in the rototranslation group. *Journal of vision*, 10(14):37–37, 2010.
- [29] Alessandro Sarti, Giovanna Citti, and Jean Petitot. The symplectic structure of the primary visual cortex. *Biological Cybernetics*, 98(1):33–48, Jan 2008.
- [30] Alessandro Sarti, Giovanna Citti, and David Piotrowski. Differential heterogenesis and the emergence of semiotic function. *Semiotica*, 2019(230):1–34, 2019.
- [31] C. Villani. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2008.
- [32] Yonatan Wexler, Eli Shechtman, and Michal Irani. Space-time video completion. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 1, pages I–I. IEEE, 2004.
- [33] Lei Zhu, Yan Yang, Steven Haker, and Allen Tannenbaum. An image morphing technique based on optimal mass preserving mapping. *IEEE transactions on image processing*, 16(6):1481–1495, 2007.