Nonlocal Data Comparison and Seven Processing and Analysis Applications

Coloma Ballester

Workshop Nonlocal Methods for Data Processing and Analysis: Theory, Optimisation, and Applications

Politecnico di Milano. June 4th 2018

NoMADS

Nonlocal Data Comparison

Nonlocal Data Comparison, to be used to identify similarities

Nonlocal Data Comparison, to be used to identify similarities

Are there similarities in data? / in the real World?

Nonlocal Data Comparison, to be used to identify similarities

Are there similarities in data? / in the real World?

Natural images contain self-similarities, which are more evident on small neighborhoods (often called patches).

Human brain can easily compare and find similarities.

Human brain can easily compare and find similarities.

It can even remove the hedge and complete the Politecnico building by exploiting the self-similarity present in the scene and using appropriate perspective distorsions to adapt it to the point of view.

Ambitious goal

To understand what is the role of smoothness and self-similarity principles in modelling natural images or videos.

- **Analysis of Self-similarities + Redundancy + Regularity in images of real scenes.**

Goal: To compare in order to identify similarities

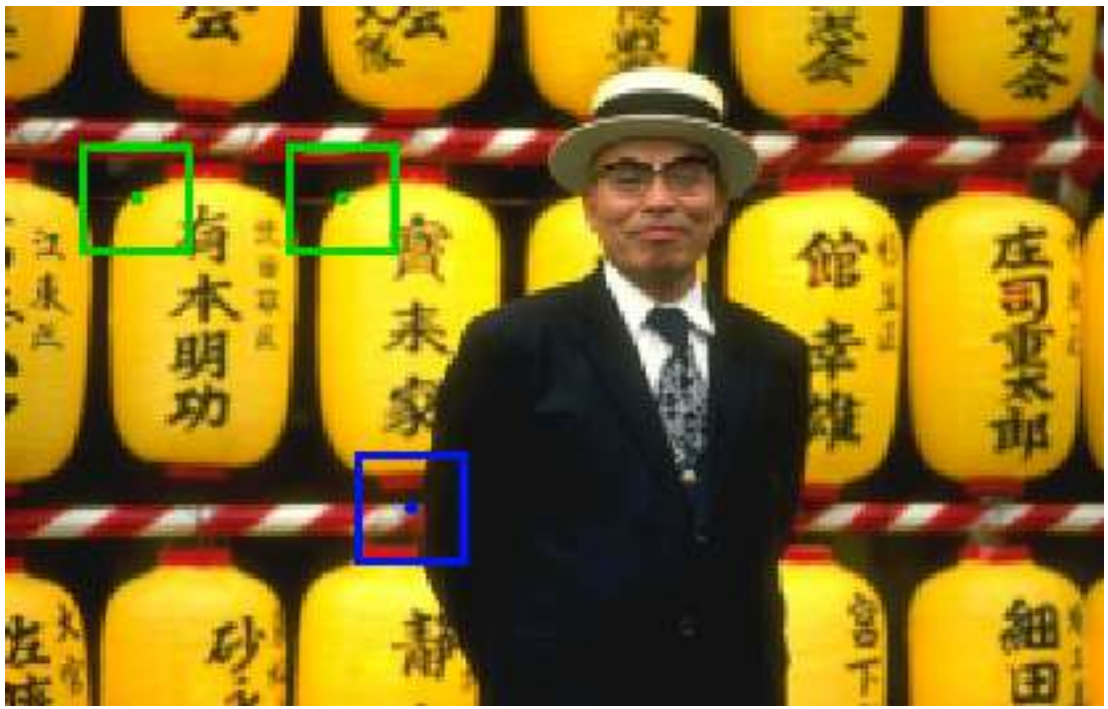
between several images or videos.

Goal: To compare in order to identify similarities

between several images or videos. Or within the same

Goal: To compare in order to identify similarities

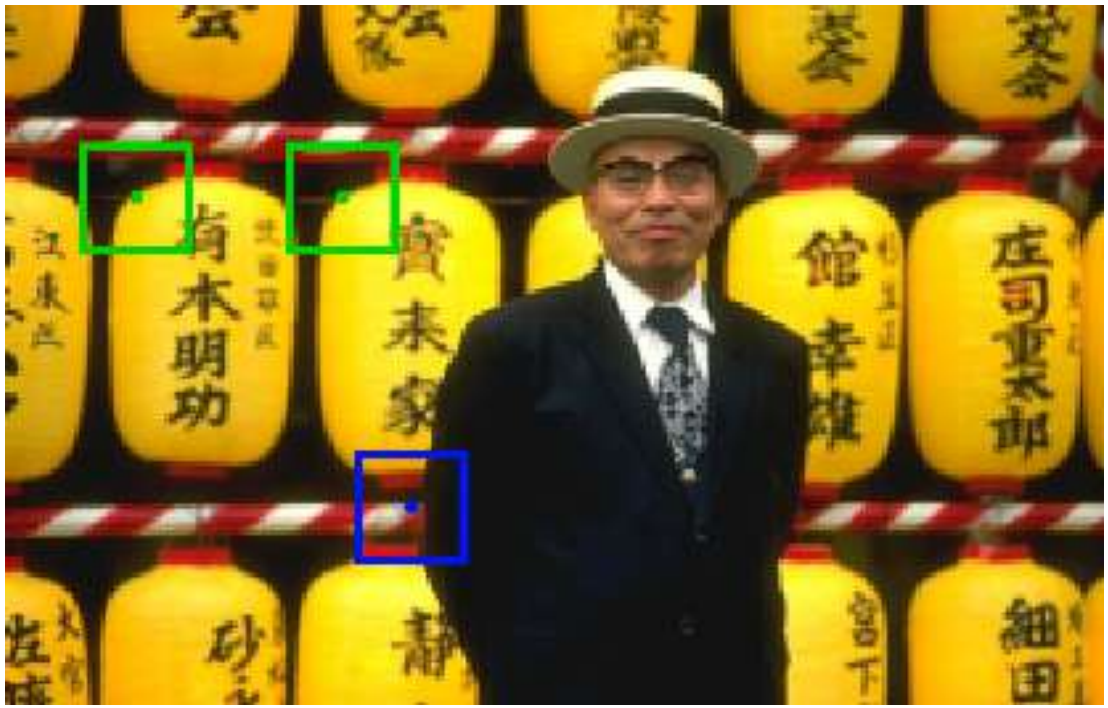
between several images or videos. Or within the same



- A common way to compare images is to compare **local neighborhoods**, or **patches**, around each pair of points in comparison.
  - **Patch-based approach: Nonlocal approach.**
  - Underlying assumption: image **self-similarity.**
-

Goal: To compare in order to identify similarities

between several images or videos. Or within the same



- A common way to compare images is to compare **local neighborhoods**, or **patches**, around each pair of points in comparison.
  - **Patch-based approach: Nonlocal approach.**
  - Underlying assumption: image **self-similarity.**
- Needed: An **appropriate similarity measure** or **comparison measure between patches** to analyse those self-similarities or similarities.

Patch-based image comparison

We consider a general definition for 'images', that is,

$$u : \Omega_u \subset \mathbb{R}^N \to \mathbb{R}^M,$$
$$v : \Omega_v \subset \mathbb{R}^N \to \mathbb{R}^M.$$

A **patch-based similarity measure** will be denoted by $\mathcal{D}$:

$$\begin{array}{rccl} \mathcal{D} : & \Omega_u & \times & \Omega_v \longrightarrow \mathbb{R} \\ & (x & , & y) \longrightarrow \mathcal{D}\left(\mathbf{p_u}(\mathbf{x}), \mathbf{p_v}(\mathbf{y})\right) \end{array}$$

where $\mathbf{p_u(x)}$: patch of u centered at x, $p_u(x) := p_u(x, \cdot)$, defined by $p_u(x, h) := u(x + h)$, $h \in \Delta$, the patch domain, often a square or a disc.

In short, denoted by

$$\begin{array}{rccl} \mathcal{D} : & \Omega_u & \times & \Omega_v \longrightarrow \mathbb{R} \\ & (x & , & y) \longrightarrow \mathcal{D}(\mathbf{x}, \mathbf{y}) \end{array}$$

Classical similarity measure between two (square) patches

- **Euclidean distance, in the discrete setting:**

$$\mathcal{D}(x,y) = \frac{1}{|\Delta|} \sum_{h \in \Delta} \Big(u(x+h) - v(y+h) \Big)^2,$$

where u and v are images, Δ denotes a patch centered at 0, and $|\Delta|$ its area.

- **In the continuous setting:**

$$\mathcal{D}(x,y) = \frac{1}{C} \int_{\mathbb{R}^2} \eta(h) \Big(u(x+h) - v(y+h) \Big)^2 \, dh,$$

where C is a normalization factor and $\eta(h)$ is either a characteristic function:

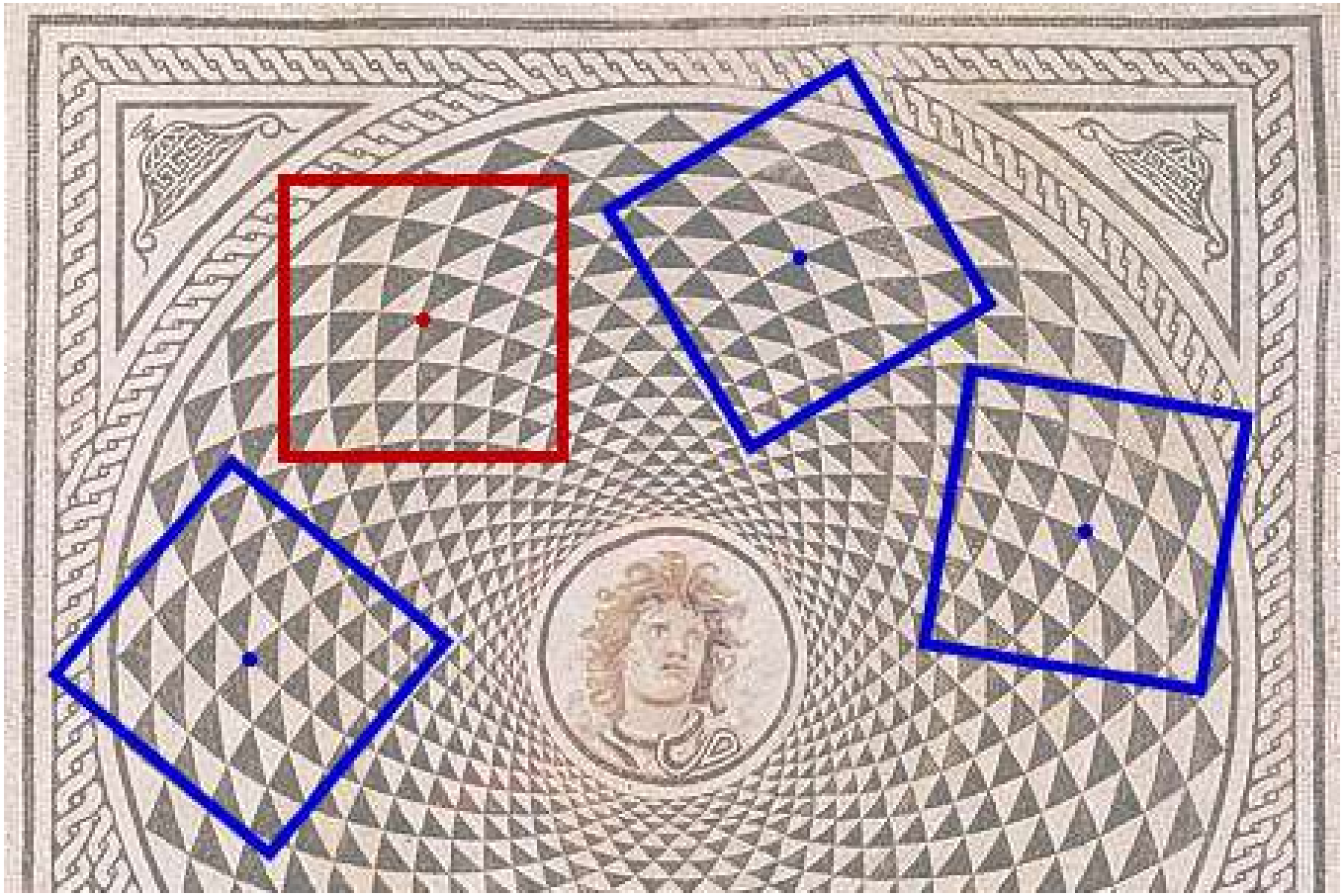
$$\eta(h) = \begin{cases} 1 & \text{if } h \in \Delta \\ 0 & \text{if } h \notin \Delta \end{cases}$$

or a windowing weighting function (e.g., a Gaussian) with effective support in Δ.

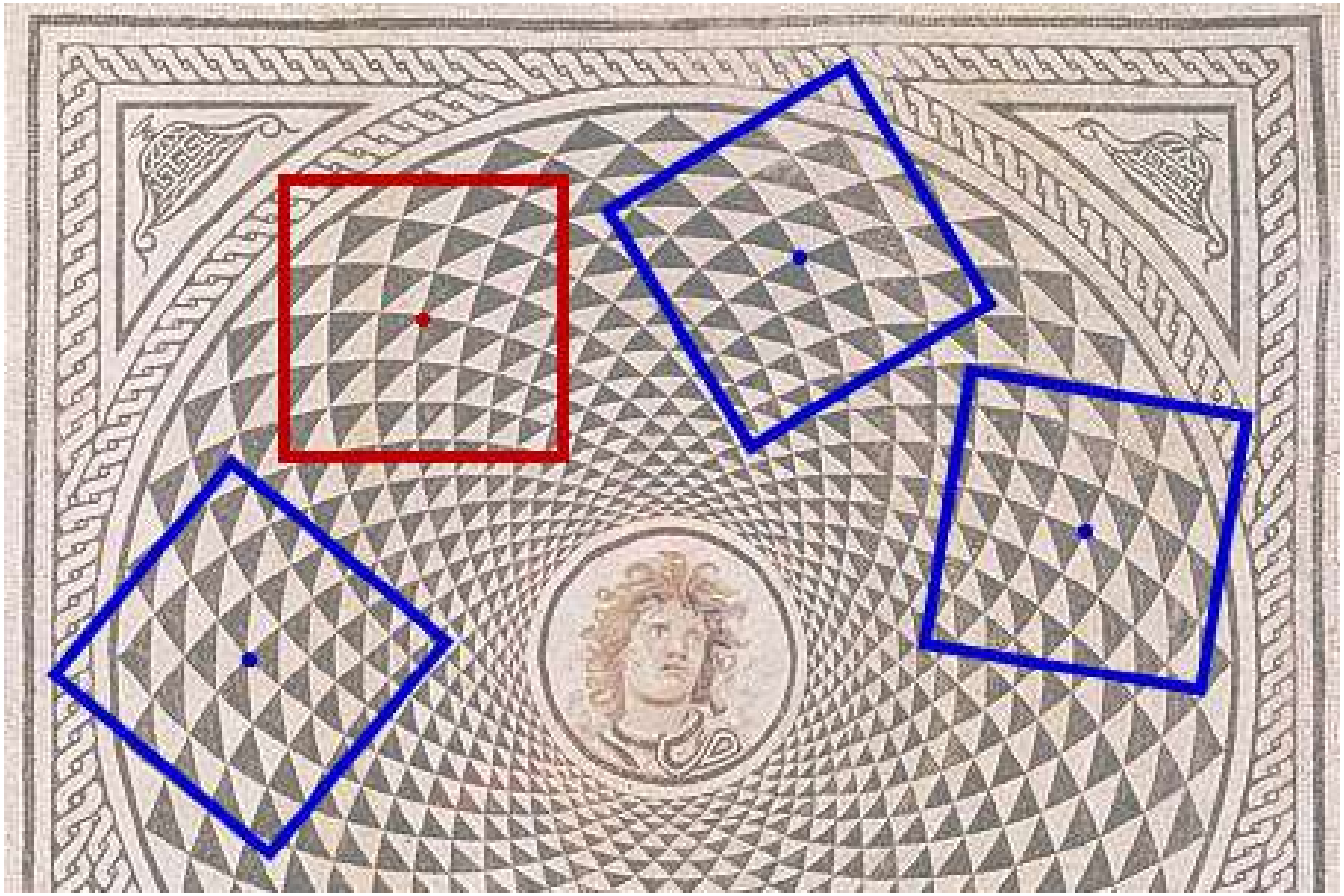
When just square patches and the Euclidean distance are not enough to compare and identify self-similarities




Intuition: patches related by a rotation



Intuition: patches related by a rotation



$$\text{Intuition:} \quad \mathcal{D}(x,y) = \frac{1}{C}\int_{\mathbb{R}^2} \eta(h)\Big(u(x+h) - u(y+R(y)h)\Big)^2\, dh$$

Self-similarity under perspective distortion

Self-similarity under perspective distortion

WANTED: shape-adaptive patches and an appropriate similarity measures to compare them.

Self-similarity under perspective distortion

WANTED: shape-adaptive patches and an appropriate similarity measures to compare them.

HOW: by endowing the image domain with Riemannian metrics and use an axiomatic approach providing appropriate comparison distances between patches that automatically and intrinsically adapt the size and shape of the patches being compared.

Nonlocal approach for self-similarity's understanding using Riemannian metrics and an axiomatic approach for image comparison

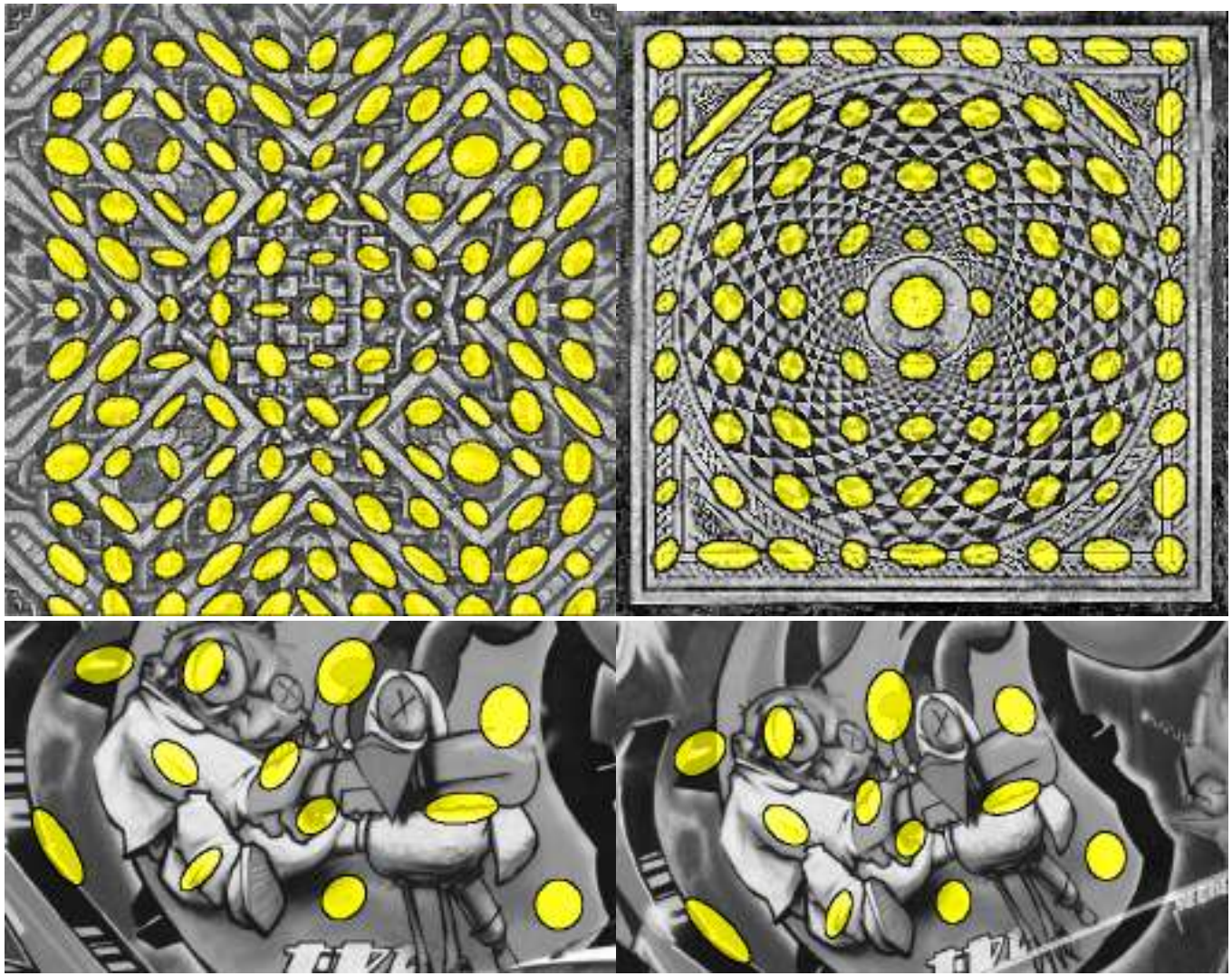
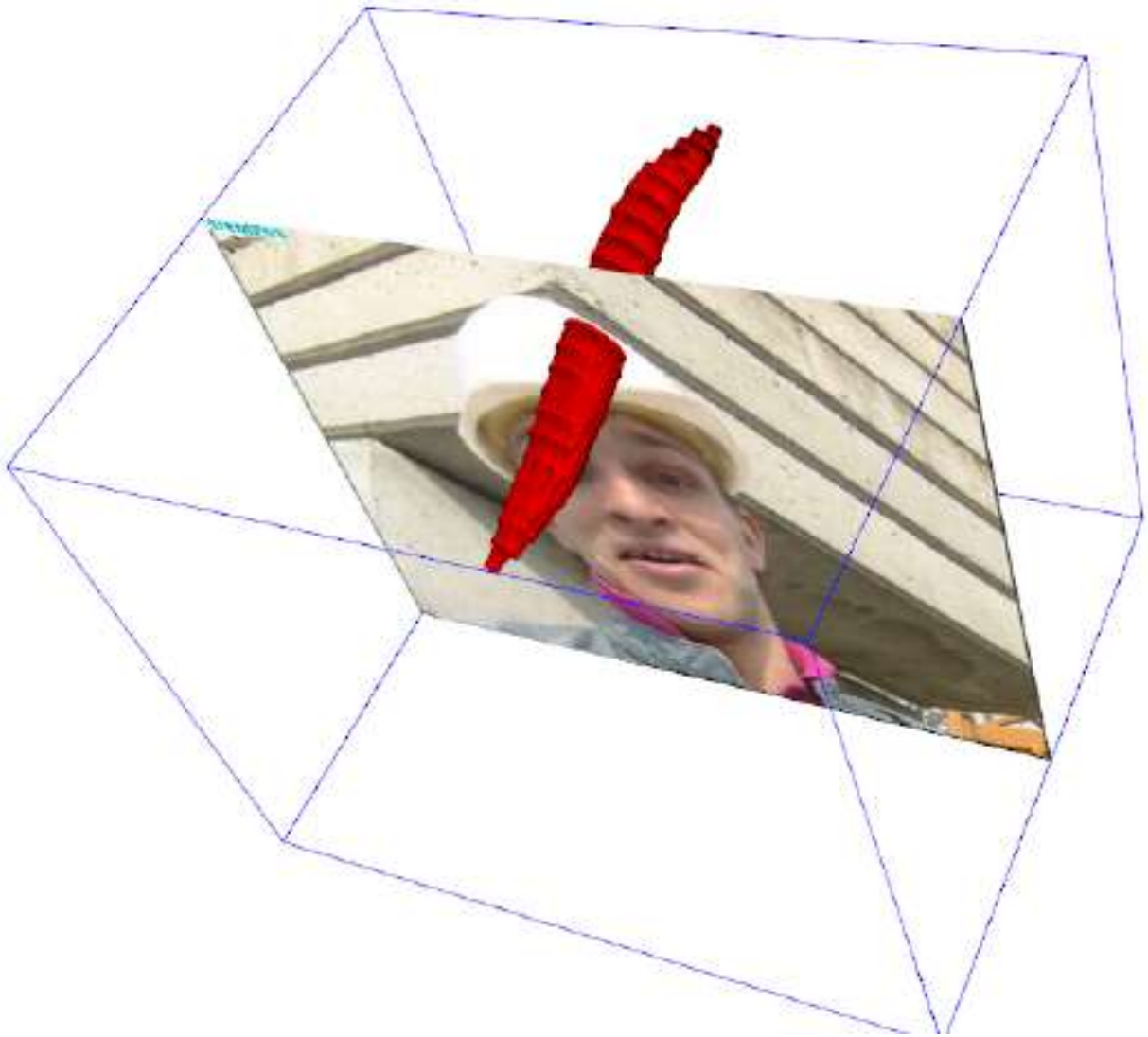


Image source: Vadim Fedorov

Affine covariant structure tensors as metrics on the image domain (1st row: metric balls are placed every 25 pixels. 2nd row: at corresponding ones).

In any dimension, including video



Appropriate metrics in video for video similarity measures

Image source: Patricia Vitoria

In any dimension, including video

Image source: Patricia Vitoria

In any dimension, including video

Image source: Patricia Vitoria

In any dimension, including video

Image source: Patricia Vitoria

In any dimension, including video

Image source: Patricia Vitoria

In any dimension, including video

Image source: Patricia Vitoria

In any dimension, including video

Image source: Patricia Vitoria

In any dimension, including video

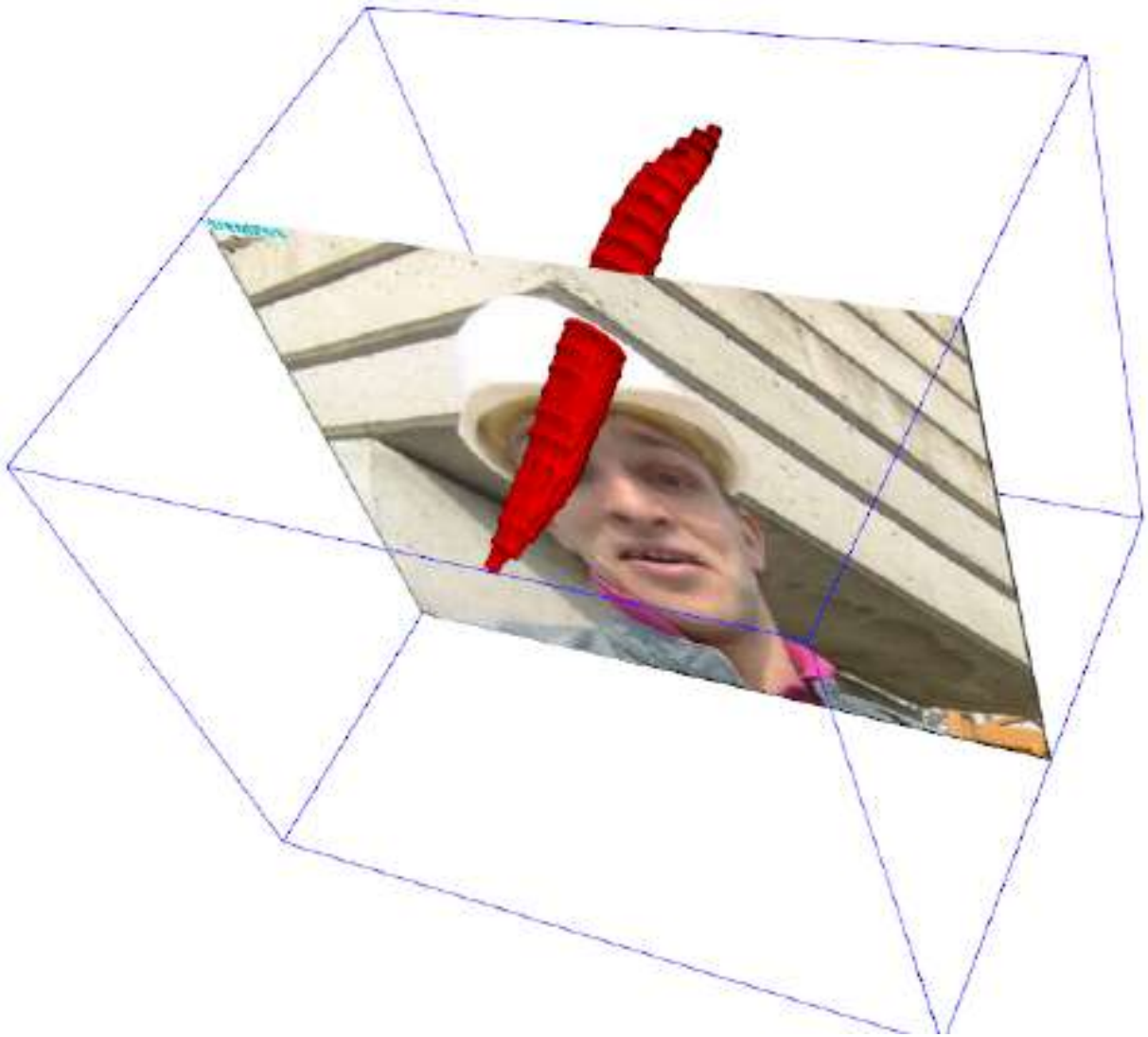
Image source: Patricia Vitoria

In any dimension, including video

Image source: Patricia Vitoria

In any dimension, including video

Image source: Patricia Vitoria



Appropriate metrics in video for video similarity measures

Image source: Patricia Vitoria

Joint work with:

Pablo Arias, Felipe Calderero, Vicent Caselles, Vadim Fedorov, Gabriele Facciolo, Gloria Haro, Vanel Lazcano, Maria Oliver, Roberto P.Palomares, Lara Raad, Rida Sadek, Patricia Vitoria.

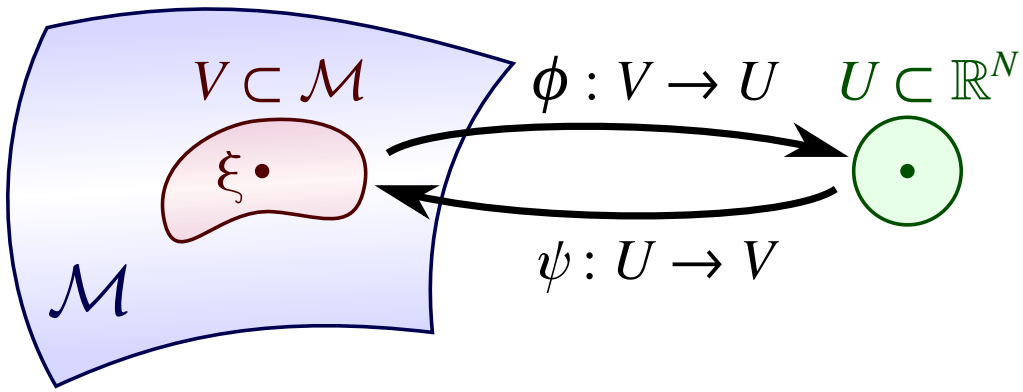
Outline

- **Nonlocal data comparison and similarity analysis**
  - **Multiscale analysis of similarities between images on Riemannian manifolds.** Axiomatic approach.
  - **The linear case.**
  - **The case of $\mathbb{R}^N$ endowed with a metric.**
  - **Affine covariant structure tensors as metrics.**
- **Seven applications:**
  1. Image and Video denoising.
  2. Exemplar-based inpainting.
  3. Image segmentation
  4. Video simplification/segmentation, spatio-temporal tubes for video analysis.
  5. Depth completion by a geodesic Biased AMLE method
  6. Motion inpainting by an image based geodesic AMLE method
  7. Dynamic shape disocclusion

Some notation and definitions

Riemannian Manifolds:

A subset $\mathcal{M} \subset \mathbb{R}^P$ is called a **smooth N-dimensional manifold** in $\mathbb{R}^P$ ($P \geq N$), if every point ξ of $\mathcal{M}$ has an open neighborhood $V \subset \mathcal{M}$ that is diffeomorphic to an open subset $U \subset \mathbb{R}^N$.



A **Riemannian manifold** is a smooth manifold equipped with a **Riemannian metric**, which provides smoothly varying choices of inner products on tangent spaces.

Some notation and definitions

- Let $(\mathcal{M}, g)$ be a **smooth Riemannian manifold** of dimension N.
- Given a point $\xi \in \mathcal{M}$, we denote by $T_\xi \mathcal{M}$ the tangent space to $\mathcal{M}$ at the point ξ. By $T_\xi^* \mathcal{M}$ we denote its dual space.
- Let $\xi \in \mathcal{M}$, $U \subseteq \mathbb{R}^N$ an open set containing 0, and $\psi : U \to \mathcal{N}$ be any coordinate system such that $\psi(0) = \xi$. For simplicity we shall denote by $\mathbf{G}(\xi)$ the (symmetric) matrix $(\mathbf{g_{ij}}(\xi))$, and by $\Gamma^{\mathcal{M},k}(\xi)$ the matrix $(\Gamma_{ij}^{\mathcal{M},k}(\xi))$, $i, j = 1, \ldots, N$ for each $k = 1, \ldots, N$.
- Rotations in the tangent space Let us define a rotation $R : T_\xi \mathcal{M} \to T_\xi \mathcal{M}$ as a linear map that satisfies

$$\langle Rv, Rw \rangle = \langle v, w \rangle \qquad \forall v, w \in T_\xi \mathcal{M}.$$

Notice that rotations satisfy $R^t G R = G$.

Some notation and definitions

The manifold $\mathcal{N} = \mathcal{M}_1 \times \mathcal{M}_2$

Let $(\mathcal{M}_i, g_i)$ be a smooth Riemannian manifold with metric g_i, $i = 1, 2$. Let $\Gamma^{(i)}$ be the connection (or Christoffel symbols) on $\mathcal{M}_i$.
We shall work with the manifold $\mathcal{N} = \mathcal{M}_1 \times \mathcal{M}_2$ with the metric $g = g_1 \times g_2$, so that $T_\xi \mathcal{N} = T_{\xi_1}\mathcal{M}_1 \times T_{\xi_2}\mathcal{M}_2$, $\xi = (\xi_1, \xi_2) \in \mathcal{M}_1 \times \mathcal{M}_2$.
If $(v_i, w_i) \in T_{\xi_1}\mathcal{M}_1 \times T_{\xi_2}\mathcal{M}_2$, $\xi = (\xi_1, \xi_2) \in \mathcal{M}_1 \times \mathcal{M}_2$, then we consider the metric

$$\langle (v_1, w_1), (v_2, w_2) \rangle_\xi = \langle v_1, v_2 \rangle_{\xi_1} + \langle w_1, w_2 \rangle_{\xi_2} = (G_1(\xi_1)v_1, v_2) + (G_2(\xi_2)w_1, w_2).$$

With a slight abuse of notation, let us write $G(\xi) = \mathrm{diag}(G_1(\xi_1), G_2(\xi_2))$.

Some notation and definitions

The manifold $\mathcal{N} = \mathcal{M}_1 \times \mathcal{M}_2$

Let $(\mathcal{M}_i, g_i)$ be a smooth Riemannian manifold with metric g_i, $i = 1, 2$. Let $\Gamma^{(i)}$ be the connection (or Christoffel symbols) on $\mathcal{M}_i$.
We shall work with the manifold $\mathcal{N} = \mathcal{M}_1 \times \mathcal{M}_2$ with the metric $g = g_1 \times g_2$, so that $T_\xi \mathcal{N} = T_{\xi_1}\mathcal{M}_1 \times T_{\xi_2}\mathcal{M}_2$, $\xi = (\xi_1, \xi_2) \in \mathcal{M}_1 \times \mathcal{M}_2$.
If $(v_i, w_i) \in T_{\xi_1}\mathcal{M}_1 \times T_{\xi_2}\mathcal{M}_2$, $\xi = (\xi_1, \xi_2) \in \mathcal{M}_1 \times \mathcal{M}_2$, then we consider the metric

$$\langle (v_1, w_1), (v_2, w_2) \rangle_\xi = \langle v_1, v_2 \rangle_{\xi_1} + \langle w_1, w_2 \rangle_{\xi_2} = (G_1(\xi_1)v_1, v_2) + (G_2(\xi_2)w_1, w_2).$$

With a slight abuse of notation, let us write $G(\xi) = \mathrm{diag}(G_1(\xi_1), G_2(\xi_2))$.

Gradient and Hessian: $D_{\mathcal{N}} C = (D_x C, D_y C)$, $D^2_{\mathcal{N}} C = \begin{pmatrix} D_{\mathcal{N},xx} C & D_{\mathcal{N},xy} C \\ D_{\mathcal{N},xy} C & D_{\mathcal{N},yy} C \end{pmatrix}$.

In coordinates, with $i, j, k \in \{1, \ldots, N\}$, $D^2_{\mathcal{N}} C = \begin{pmatrix} \frac{\partial^2 C}{\partial x^i \partial x^j} & \frac{\partial^2 C}{\partial x^i \partial y^j} \\ \frac{\partial^2 C}{\partial y^j \partial x^i} & \frac{\partial^2 C}{\partial y^i \partial y^j} \end{pmatrix} - \begin{pmatrix} \Gamma^{(1)k}(x) \frac{\partial C}{\partial x^k} & 0 \\ 0 & \Gamma^{(2)k}(y) \frac{\partial C}{\partial y^k} \end{pmatrix}$.

Key concept: A priori connections on $\mathcal{N} = \mathcal{M}_1 \times \mathcal{M}_2$

This is a key concept that allows to actually compare patches in both manifolds. It is an an operator that connects the tangent plane at both points in comparison.

An **a priori connection between both manifolds** is a field of linear maps which for each pair of points $(x, y) \in \mathcal{M}_1 \times \mathcal{M}_2$, maps isometrically the tangent space at x with the tangent space at y.

Key concept: A priori connections on $\mathcal{N} = \mathcal{M}_1 \times \mathcal{M}_2$

This is a key concept that allows to actually compare patches in both manifolds. It is an an operator that connects the tangent plane at both points in comparison.

An **a priori connection between both manifolds** is a field of linear maps which for each pair of points $(x, y) \in \mathcal{M}_1 \times \mathcal{M}_2$, maps isometrically the tangent space at x with the tangent space at y.

Definition. We say that $P(\xi)$, $\xi = (\xi_1, \xi_2) \in \mathcal{M}$, is an **a priori connection map in $\mathcal{N}$** if $P(\xi) : (T_{\xi_1}\mathcal{M}_1, G_1(\xi_1)) \to (T_{\xi_2}\mathcal{M}_2, G_2(\xi_2))$ is an isometry, i.e.

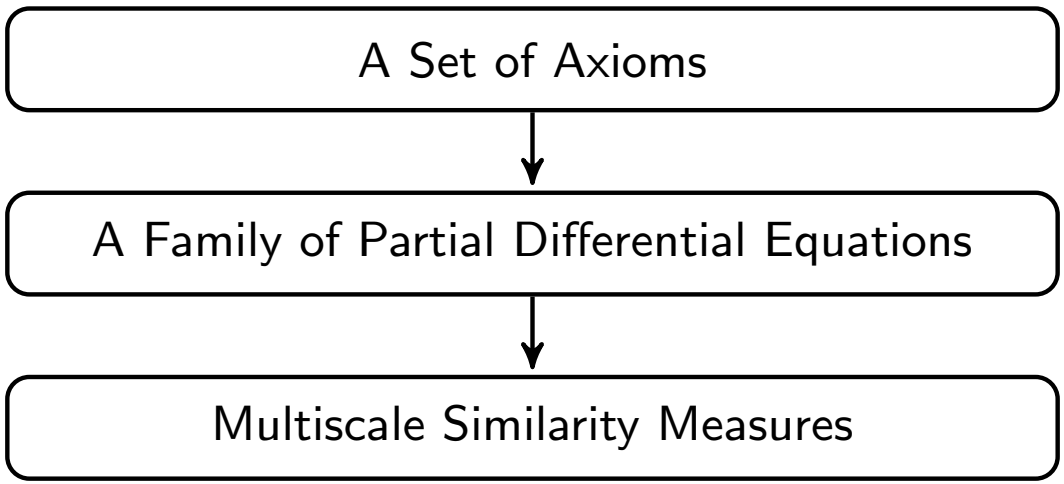
$$\langle P(\xi)v, P(\xi)w \rangle_{G_2(\xi_2)} = \langle v, w \rangle_{G_1(\xi_1)} \quad \forall v, w \in T_{\xi_1}\mathcal{M}_1,$$

and we assume also that the map is differentiable in ξ.

If $\mathcal{M}^1 = \mathcal{M}^2$ (orientable), $P(\xi)$ can be considered as an internal a priori connection (into itself) given from parallel transport between ξ_1 and ξ_2, which is an isometry.

Axiomatic approach

Considering images defined on Riemannian manifolds we exploit an axiomatic approach[1,2,3] to obtain multiscale affine invariant patch similarity measures.



---

[1] L. Alvarez, F. Guichard, P.L. Lions and J.M. Morel, *Axioms and fundamental equations of image processing*, Archive for Rational Mechanics and Analysis, 1993.

[2] F. Calderero and V. Caselles, *Multiscale Analysis for Images on Riemannian Manifolds*, SIAM J. Imaging Sciences 2014.

[3] C. Ballester, F. Calderero, V. Caselles and G. Facciolo, *Multiscale analysis of similarities between images on Riemannian manifolds*, SIAM Multiscale Modeling & Simulation, 2014.

Multiscale analyses

Definition

A "multiscale analysis" is a family of transforms T_t, $t \geq 0$, which, when applied to a given function u_0, yield a sequence of functions $u(t) = T_t u_0$ at different scales.

In our case of **multiscale analysis of image similarity measures**[†]:

Let $T_t : C_b(\mathcal{N}) \to C_b(\mathcal{N})$ be a nonlinear operator for any $t \geq 0$.

where:

$\mathcal{N} = \mathcal{M}_1 \times \mathcal{M}_2$ is a product manifold.

$C_b(\mathcal{N})$ denotes the space of bounded continuous functions on $\mathcal{N}$ with the maximum norm.

$\mathcal{D} \in C_b(\mathcal{N})$ denotes a similarity measure on $\mathcal{N}$.

[†] C. Ballester, F. Calderero, V. Caselles and G. Facciolo, *Multiscale analysis of similarities between images on Riemannian manifolds*, SIAM MMS, 2014.

The Set of Axioms

Let $T_t\mathcal{D}(\xi)$ be a multiscale analysis of image similarity measures, for all $\mathcal{D} \in C_b(\mathcal{N}), \xi = (\xi_1, \xi_2), t \geq 0$.

Architectural axioms:

- **Recursivity**: $T_0(\mathcal{D}) = \mathcal{D},\ T_h(T_t\mathcal{D}) = T_{t+h}\mathcal{D}_0, \forall t, h \geq 0$;
- **Infinitesimal generator**: independence of the choice of step h;
- **Regularity**: "continuity" of T_t;
- **Locality**;

Comparison principle: $T_t\mathcal{D} \leq T_t\widetilde{\mathcal{D}}$, $\forall t \geq 0$ and all $\mathcal{D} \leq \widetilde{\mathcal{D}}$;

Gray level shift invariance: $T_t(0) = 0,\ T_t(\mathcal{D} + \kappa) = T_t(\mathcal{D}) + \kappa$, $\forall \kappa \in \mathbb{R}$.

Multiscale analysis of image similarity measures

Theorem

Let T_t be a multiscale analysis satisfying all the **Architectural axioms**, and the **Comparison principle**. Then, there exists a function $F : SM_\xi(\mathcal{N}) \times T^*_\xi\mathcal{N} \times \mathbb{R} \times \mathcal{N} \to \mathbb{R}$ increasing with respect to its first argument such that

$$\frac{T_t(\mathcal{D},\psi) - (\mathcal{D},\psi)}{t} \to F(D^2(\mathcal{D}\circ\psi)(0), D(\mathcal{D}\circ\psi)(0), \mathcal{D}(\xi), \xi, G, \Gamma^k) \text{ in } C_b(\mathcal{N})$$

as $t \to 0+$, for all $\mathcal{D} \in C_b^\infty(\mathcal{N})$, ψ being a coordinate system around $\xi \in \mathcal{N}$. The function F is continuous in its first three arguments.

If we assume that T_t is **gray level shift invariant**, then the function F does not depend on $\mathcal{D}$.

The function **F is elliptic**, i.e., if $A_1, A_2 : T_\xi\mathcal{N} \to T^*_\xi\mathcal{N}$ are two matrices such that A_1, A_2 are symmetric and $A_1 \leq A_2$, $p \in T^*_\xi\mathcal{N}$, $c \in \mathbb{R}$, then

$$F(A_1, p, c, \xi, G, \Gamma^k) \leq F(A_2, p, c, \xi, G, \Gamma^k).$$

Theorem

Let T_t be a multiscale analysis satisfying all the **Architectural axioms**, the **Comparison principle**, and **Gray level shift invariance**.
If $\mathcal{D}(t,\xi) = T_t\mathcal{D}(\xi)$, then $\mathcal{D}$ is a viscosity solution of

$$\frac{\partial\mathcal{D}}{\partial t} = F(D^2_{\mathcal{N}}\mathcal{D}, D\mathcal{D}, \xi, G, \Gamma^k),$$

with $\mathcal{D}(0,\xi) = \mathcal{D}(\xi)$.

The linear case

Architectural axioms:

- **Recursivity**: $T_0(\mathcal{D}) = \mathcal{D},\ T_h(T_t\mathcal{D}) = T_{t+h}\mathcal{D}_0, \forall t, h \geq 0$;
- **Infinitesimal generator**: independence of the choice of step h;
- **Regularity**: "continuity" of T_t;
- **Locality**;

Comparison principle: $T_t\mathcal{D} \leq T_t\widetilde{\mathcal{D}}$, $\forall t \geq 0$ and all $\mathcal{D} \leq \widetilde{\mathcal{D}}$;

Gray level shift invariance;

Linearity: $T_t(a\mathcal{D} + b\widetilde{\mathcal{D}}) = aT_t(\mathcal{D}) + bT_t(\widetilde{\mathcal{D}}), \quad \forall a, b \in \mathbb{R}$.

The linear case

Theorem

Let T_t be a multiscale analysis on similarity functions satisfying the axioms: **all Architectural, Comparison principle, Gray level Shift invariance** and **Linear**. Then

$$\frac{\partial \mathcal{D}}{\partial t} = F(D^2_{\mathcal{N}}\mathcal{D}, \xi, G),$$

where

$$F(X, \xi, G) = c_{11}(\xi)\mathrm{Tr}(G_1^{-1}(\xi_1)X_{11}) + 2c_{12}(\xi, G)\mathrm{Tr}(\bar{D}_{12}I_1(\xi_1)^{-1}X_{12}) + c_{22}(\xi)\mathrm{Tr}(G_2^{-1}(\xi_2)X_{22}),$$

$\xi = (\xi_1, \xi_2)$, $\bar{D}_{12}$ is an isometry from $(T_{\xi_1}\mathcal{M}_1, G_1(\xi_1)) \to (T_{\xi_2}\mathcal{M}_2, G_2(\xi_2))$. The ellipticity of F implies that $c_{11}, c_{22} \geq 0$.

- the operators $c_{ii}(\xi)\mathrm{Tr}((G_i)^{-1}(\xi_i)X_{ii})$ are **multiples of the Laplace-Beltrami operator**.
- There are no first order terms in these operators. They cannot couple with vectors so that we have the invariance induced by the rotations of tangent planes.
- $2c_{12}(\xi, G)\bar{D}_{12}I_1(\xi_1)^{-1} = B_2(\xi_2)D'B_1(\xi_1)^t$ (where $B_1(\xi_1)$ and $B_2(\xi_2)$ are isometries) is related with the **a-priori connection**.

Practical situation: $\mathcal{M}_1 = (\mathbb{R}^N, g_1(x))$, $\mathcal{M}_2 = (\mathbb{R}^N, g_2(y))$

Let $x \in \mathcal{M}_1$, $y \in \mathcal{M}_2$.

$e^k = G_i(x)^{-1/2} f^k$ is a orthonormal basis of $(T_x\mathcal{M}_i, g_i(x))$, if f^k is an Euclidean orthonormal basis.

We can define $\mathbf{P(x,y)(v) = G_2(y)^{-1/2} G_1(x)^{1/2} v}$, $v \in \mathbb{R}^N$, as an **a priori connection of x and y.**

Then $|\mathbf{P(x,y)v}|_{\mathbf{g_2}} = |\mathbf{v}|_{\mathbf{g_1}}$ for all $(x, y) \in \mathbb{R}^{2N}$.

The PDE obtained is

$$\frac{\partial \mathcal{D}}{\partial \mathbf{t}} = \mathbf{a(x,y)} \boldsymbol{\Delta}_{\mathcal{M}\mathbf{x}} \mathcal{D} + \mathbf{2c_{12}(x,y)} \mathrm{Tr}(\mathbf{G_2(y)}^{-\frac{1}{2}} \mathbf{G_1(x)}^{-\frac{1}{2}} \mathbf{D_{xy}} \mathcal{D}) + \mathbf{c(x,y)} \boldsymbol{\Delta}_{\mathcal{M}\mathbf{y}} \mathcal{D},$$

where

$\Delta_{\mathcal{M}x}\mathcal{D} = \mathrm{Tr}(G^1(x)^{-1}(D_{xx}\mathcal{D}(x) - \Gamma^{(1)}(D\mathcal{D})(x)))$ (and similarly for the operator $\Delta_{\mathcal{M}y}$).

This will permit to construct also an operator in the case of video.

Linear multiscale analysis of similarities

Singularizing one of them[1], a **linear multiscale similarity measure** $\mathcal{D}(\mathbf{t}, \mathbf{x}, \mathbf{y})$ is a solution of the PDE

$$\frac{\partial \mathcal{D}}{\partial t} = \mathrm{Tr}\left(G_1(x)^{-1} D_x^2 \mathcal{D}\right) + 2\mathrm{Tr}\left(G_2(y)^{-\frac{1}{2}} G_1(x)^{-\frac{1}{2}} D_{xy} \mathcal{D}\right) + \mathrm{Tr}\left(G_2(y)^{-1} D_y^2 \mathcal{D}\right)$$

Linear multiscale analysis of similarities

Singularizing one of them[1], a **linear multiscale similarity measure** $\mathcal{D}(\mathbf{t}, \mathbf{x}, \mathbf{y})$ is a solution of the PDE

$$\frac{\partial \mathcal{D}}{\partial t} = \mathrm{Tr}(G_1(x)^{-1} D_x^2 \mathcal{D}) + 2\mathrm{Tr}(G_2(y)^{-\frac{1}{2}} G_1(x)^{-\frac{1}{2}} D_{xy} \mathcal{D}) + \mathrm{Tr}(G_2(y)^{-1} D_y^2 \mathcal{D})$$

That is:

$$\frac{\partial \mathcal{D}}{\partial t} = \mathrm{Tr}(G_1(x)^{-1} D_x^2 \mathcal{D}) + 2\mathrm{Tr}(\mathbf{P(x,y)} G_1(x)^{-1} D_{xy}^2 \mathcal{D}) + \mathrm{Tr}(G_2(y)^{-1} D_y^2 \mathcal{D}),$$

where $\mathbf{P(x,y) = G_2(y)^{-\frac{1}{2}} G_1(x)^{\frac{1}{2}}}$ is an a priori connection of x and y. It gives the tool to automatically and intrinsically transform the patches in comparison.

The computational complexity of solving this equation is determined by the product manifold $\mathcal{M}_1 \times \mathcal{M}_2$, thus is of order S^4, if each image is determined on a grid of size S^2.

---

[1] V. Fedorov, P. Arias, R. Sadek, G. Facciolo, and C. Ballester, *Linear Multiscale Analysis of Similarities between Images on Riemannian Manifolds: Practical Formula and Affine Covariant Metrics.* SIAM J. Imaging Sciences. 2015.

Approximate solution

For that reason, we used[1] the **WKB (Wentze-Kramers-Brillouin) approximation method** to find a solution of the previous PDE, and obtain the following applicable similarity measure

---

[1] V. Fedorov, P. Arias, R. Sadek, G. Facciolo, and C. Ballester, *Linear Multiscale Analysis of Similarities between Images on Riemannian Manifolds: Practical Formula and Affine Covariant Metrics*. SIAM J. Imaging Sciences. 2015.

Approximate solution

For that reason, we used[1] the **WKB (Wentze-Kramers-Brillouin) approximation method** to find a solution of the previous PDE, and obtain the following applicable similarity measure

$$\mathcal{D}(t,x,y) = \int_{\mathbb{R}^2} \eta_t(h) \left(u(x + G_1(x)^{-\frac{1}{2}} h) - v(y + G_2(y)^{-\frac{1}{2}} h) \right)^2 dh.$$

where $\eta_t(h)$ is an approximated geodesic weighting function, for instance

---

[1] V. Fedorov, P. Arias, R. Sadek, G. Facciolo, and C. Ballester, *Linear Multiscale Analysis of Similarities between Images on Riemannian Manifolds: Practical Formula and Affine Covariant Metrics.* SIAM J. Imaging Sciences. 2015.

Approximate solution

For that reason, we used[1] the **WKB (Wentze-Kramers-Brillouin) approximation method** to find a solution of the previous PDE, and obtain the following applicable similarity measure

$$\mathcal{D}(t,x,y) = \int_{\mathbb{R}^2} \eta_t(h) \left(u(x + G_1(x)^{-\frac{1}{2}} h) - v(y + G_2(y)^{-\frac{1}{2}} h) \right)^2 dh.$$

where $\eta_t(h)$ is an approximated geodesic weighting function, for instance

$$\mathcal{D}(t,x,y) = \frac{H}{\sqrt{t}} \int_{\mathbb{R}^2} e^{-d(x,x+G_1(x)^{-\frac{1}{2}} h)^2/t} \left(u(x + G_1(x)^{-\frac{1}{2}} h) - v(y + G_2(y)^{-\frac{1}{2}} h) \right)^2$$

where $d(x,x')$ represents a geodesic distance in $\mathcal{M}_1$.
$d(x, x + G_1(x)^{-\frac{1}{2}} h)^2$ could be approximated as
$\kappa_{spatial} \| G_1(x)^{-\frac{1}{2}} h \|^2 + \kappa_{color} |u(x) - u(x + G_1(x)^{-\frac{1}{2}} h)|^2$, $\kappa_{spatial}, \kappa_{color} > 0$.

---

[1] V. Fedorov, P. Arias, R. Sadek, G. Facciolo, and C. Ballester, *Linear Multiscale Analysis of Similarities between Images on Riemannian Manifolds: Practical Formula and Affine Covariant Metrics*. SIAM J. Imaging Sciences. 2015.

Approximate solution

For that reason, we used[1] the **WKB (Wentze-Kramers-Brillouin) approximation method** to find a solution of the previous PDE, and obtain the following applicable similarity measure

$$\mathcal{D}(t,x,y) = \int_{\mathbb{R}^2} \eta_t(h) \left(u(x + G_1(x)^{-\frac{1}{2}} h) - v(y + G_2(y)^{-\frac{1}{2}} h) \right)^2 dh.$$

where $\eta_t(h)$ is an approximated geodesic weighting function, for instance

$$\mathcal{D}(t,x,y) = \frac{H}{\sqrt{t}} \int_{\mathbb{R}^2} e^{-d(x,x+G_1(x)^{-\frac{1}{2}} h)^2/t} \left(u(x + G_1(x)^{-\frac{1}{2}} h) - v(y + G_2(y)^{-\frac{1}{2}} h) \right)^2$$

where $d(x,x')$ represents a geodesic distance in $\mathcal{M}_1$.
$d(x, x + G_1(x)^{-\frac{1}{2}} h)^2$ could be approximated as
$\kappa_{spatial} \| G_1(x)^{-\frac{1}{2}} h \|^2 + \kappa_{color} |u(x) - u(x + G_1(x)^{-\frac{1}{2}} h)|^2$, $\kappa_{spatial}, \kappa_{color} > 0$.

Last step: Which metrics g_1 and g_2?

---

[1] V. Fedorov, P. Arias, R. Sadek, G. Facciolo, and C. Ballester, *Linear Multiscale Analysis of Similarities between Images on Riemannian Manifolds: Practical Formula and Affine Covariant Metrics.* SIAM J. Imaging Sciences. 2015.

Which metrics g_1 and g_2?

• **Appropriate metrics defined on the data domain and depending on the data.**

Which metrics g_1 and g_2?

• Appropriate metrics defined on the data domain and depending on the data.

• One possibility is to choose the metrics such as the corresponding similarity measure is affine invariant.

By affine invariance in this setting we mean that the similarity values are invariant to affine warpings of any of the images.

Which metrics g_1 and g_2?

• Appropriate metrics defined on the data domain and depending on the data.

• One possibility is to choose the metrics such as the corresponding similarity measure is affine invariant.

By affine invariance in this setting we mean that the similarity values are invariant to affine warpings of any of the images.

• The corresponding patch-based similarity measure will be able to detect the scene similarities despite the view distortions.

Structure Tensors as Metrics on the image domain

The a priori connection will be then determined up to a rotation (for each pair (x, y)), which can be determined based on the data.

Affine covariant structure censor

Let u be a given image.

Definition

A structure tensor is said to be **affine covariant**, if for any affinity A

$$T_{u_A}(x) = A^t T_u(Ax) A,$$

where $u_A(x) := u(Ax)$ denotes the affinely transformed version of u.

Affine covariant structure censor

Let u be a given image.

Definition

A structure tensor is said to be **affine covariant**, if for any affinity A

$$T_{u_A}(x) = A^t T_u(Ax) A,$$

where $u_A(x) := u(Ax)$ denotes the affinely transformed version of u.

Given an affine covariant structure tensor $T_u(x)$ we can define a region (or patch) of "radius" r, centered at x, which is affine covariant as well

$$B_u(x, r) = \{y \in \mathbb{R}^2 : \langle T_u(x)(y - x), (y - x)\rangle \leq r^2\}.$$

(i.e., $B_{u_A}(x, r) = A^{-1} B_u(Ax, r)$)

Affine covariant structure censor

Let u be a given image.

Definition

A structure tensor is said to be **affine covariant**, if for any affinity A

$$T_{u_A}(x) = A^t T_u(Ax) A,$$

where $u_A(x) := u(Ax)$ denotes the affinely transformed version of u.

Given an affine covariant structure tensor $T_u(x)$ we can define a region (or patch) of "radius" r, centered at x, which is affine covariant as well

$$B_u(x, r) = \{y \in \mathbb{R}^2 : \langle T_u(x)(y - x), (y - x)\rangle \leq r^2\}.$$

(i.e., $B_{u_A}(x, r) = A^{-1} B_u(Ax, r)$)

Computation of structure tensors and patches

- Let $u : \Omega_u \to \mathbb{R}$ be a given image. Assume $\Omega_u = \mathbb{R}^2$, to simplify.

- For each $x \in \Omega_u$, let $T_u(x)$ be the **affine covariant structure tensor** of u at x defined by the following iterative algorithm, which also provides their corresponding **affine covariant neighborhoods**:

$$T_u^{(k)}(x) = \frac{\int_{B_{T_u^{(k-1)}}(x,r)} Du(y) \otimes Du(y)\, dy}{\mathrm{Area}(B_{T_u^{(k-1)}}(x,r))},$$

for $k \geq 1$, where

$$B_{T_u^{(0)}}(x,r) = \{y : |Du(x)(y-x)| \leq r\},$$

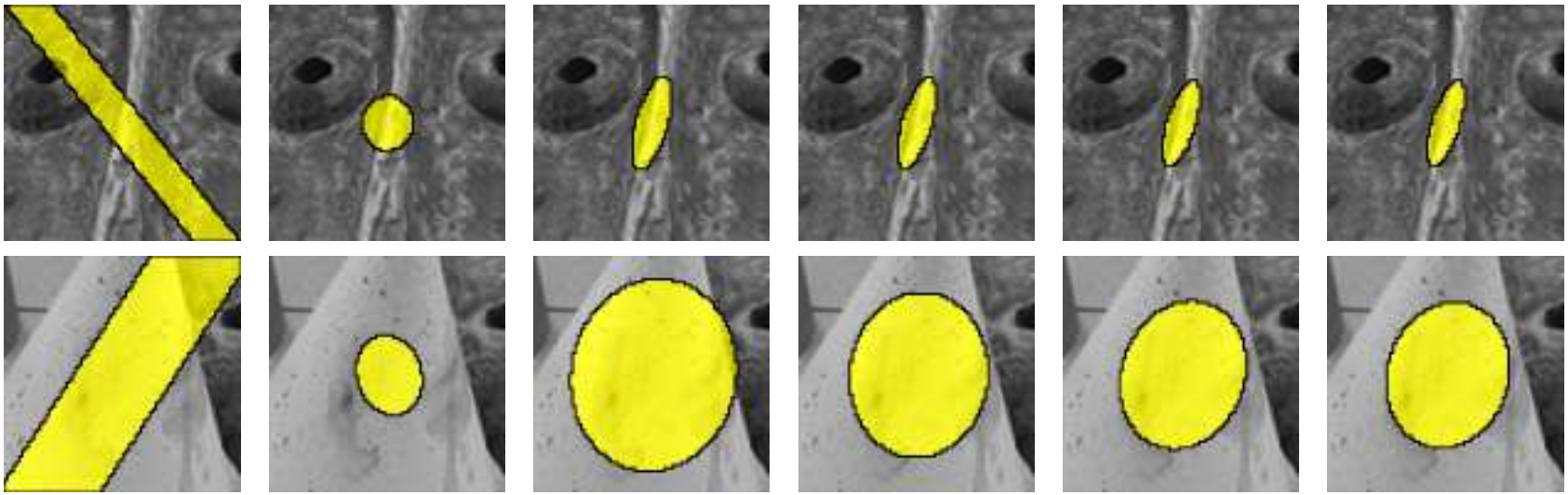
and

$$B_{T_u^{(k)}}(x,r) = \{y : \langle T_u^{(k)}(x)(y-x), (y-x)\rangle \leq r^2\}$$

for $k \geq 1$. We obtain a field of affine covariant tensors and neigborhoods.

V. Fedorov, P. Arias, R. Sadek, G. Facciolo, and C. Ballester, *Linear Multiscale Analysis of Similarities between Images on Riemannian Manifolds: Practical Formula and Affine Covariant Metrics.* SIAM J. Imaging Sciences. 2015.

Shape-adaptive Patches at Different Iterations



After approximate convergence, every 25 pixels:

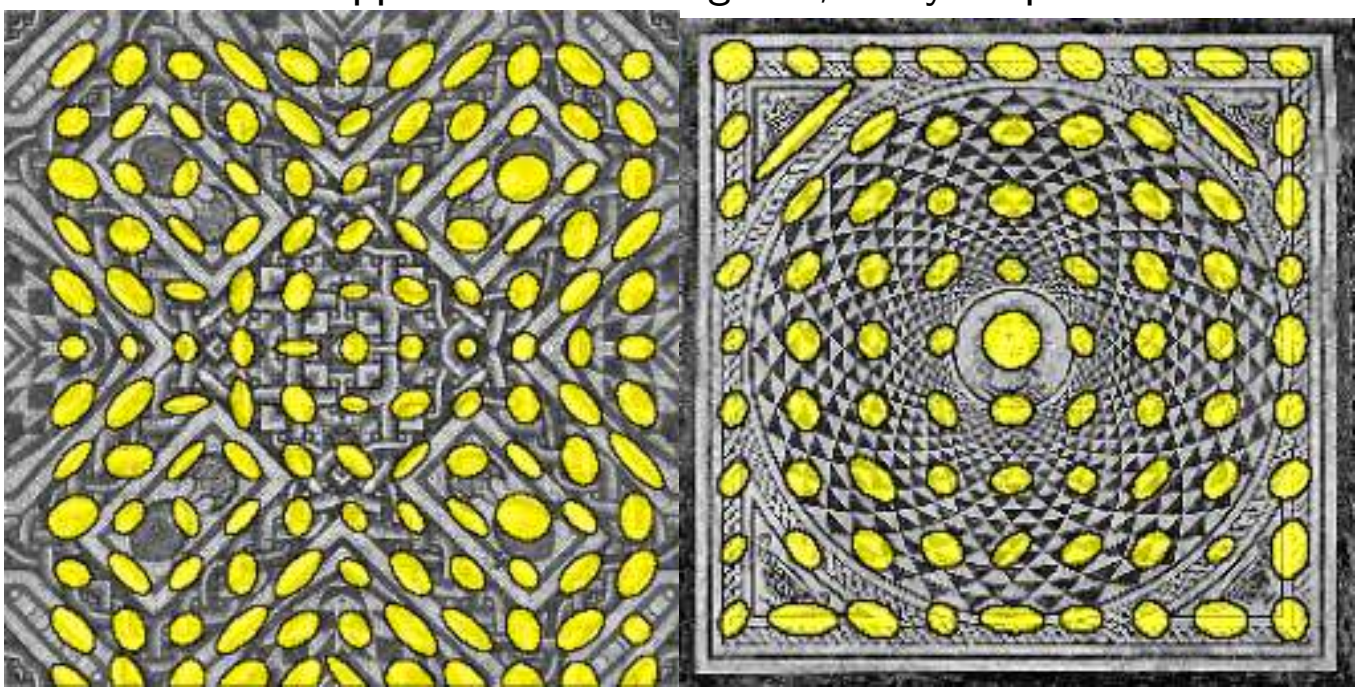


Image source: Vadim Fedorov

An afffine invariant patch similarity measure

Summarizing

- Let $u : \Omega_u \to \mathbb{R}$ and $v : \Omega_v \to \mathbb{R}$ be two given images, $\Omega_u, \Omega_v \subseteq \mathbb{R}^2$.

- For each $x \in \Omega_u$, let $T_u(x)$ be the affine covariant structure tensor of u at x. Similarly, let $T_v(y)$ be the affine covariant structure tensor of v at y, for all $y \in \Omega_v$.

- Consider the two Riemannian manifolds $\mathcal{M}_x = (\mathbb{R}^2, T_u)$, $\mathcal{M}_y = (\mathbb{R}^2, T_v)$ and the affine invariant patch distance

$$\mathcal{D}^{\mathsf{a}}(t, x, y) = \int_{\mathbb{R}^2} \eta_t(h) \left(u(x + T_u(x)^{-\frac{1}{2}} h) - v(y + T_v(y)^{-\frac{1}{2}} h) \right)^2 dh.$$

which approximates a solution of the PDE for the multiscale analysis of similarity measures satisfying a set of appropriate axioms (architectural axioms, comparison principle, gray level shift invariance and linearity):

$$\frac{\partial \mathcal{D}}{\partial t} = \mathrm{Tr}(T_u(x)^{-1} D_x^2 \mathcal{D}) + 2\mathrm{Tr}(P(x, y) T_u(x)^{-1} D_{xy} \mathcal{D}) + \mathrm{Tr}(T_v(y)^{-1} D_y^2 \mathcal{D})$$

where $P(x, y) = T_v(y)^{-\frac{1}{2}} T_u(x)^{\frac{1}{2}}$.

An afffine invariant patch similarity measure

The a priory connection allows to extract the affine distortion between corresponding patches, up to a rotation.

It can be shown that, if u and v differ in an affine transformation given by a $N \times N$ affinity matrix A, then

$$A = P_R(x, y) := T_v(y)^{-\frac{1}{2}} R T_u(x)^{\frac{1}{2}}$$

where $y = Ax$, R is an orthogonal matrix and $P_R(x, y)$ is an a priori connection.

An afffine invariant patch similarity measure

The a priory connection allows to extract the affine distortion between corresponding patches, up to a rotation.

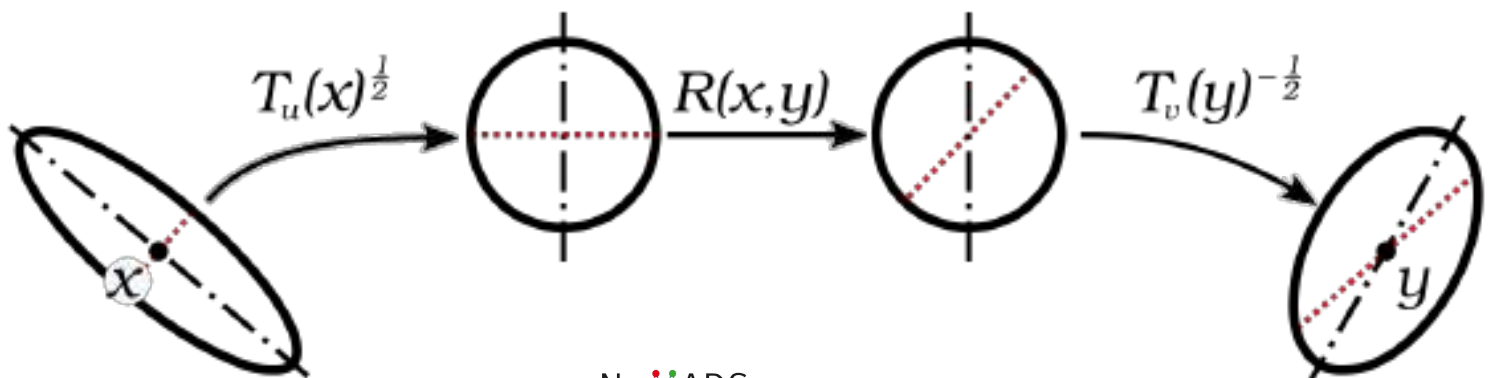
It can be shown that, if u and v differ in an affine transformation given by a $N \times N$ affinity matrix A, then

$$A = P_R(x, y) := T_v(y)^{-\frac{1}{2}} R T_u(x)^{\frac{1}{2}}$$

where $y = Ax$, R is an orthogonal matrix and $P_R(x, y)$ is an a priori connection.

In the **general case** we have two arbitrary points x and y on two images u and v, respectively:

$$P(x, y) = T_v(y)^{-\frac{1}{2}} R(x, y) T_u(x)^{\frac{1}{2}}.$$



Affine Invariant Patch Comparison; Geometrical meaning

In practice it is more suitable to split the rotation into two parts

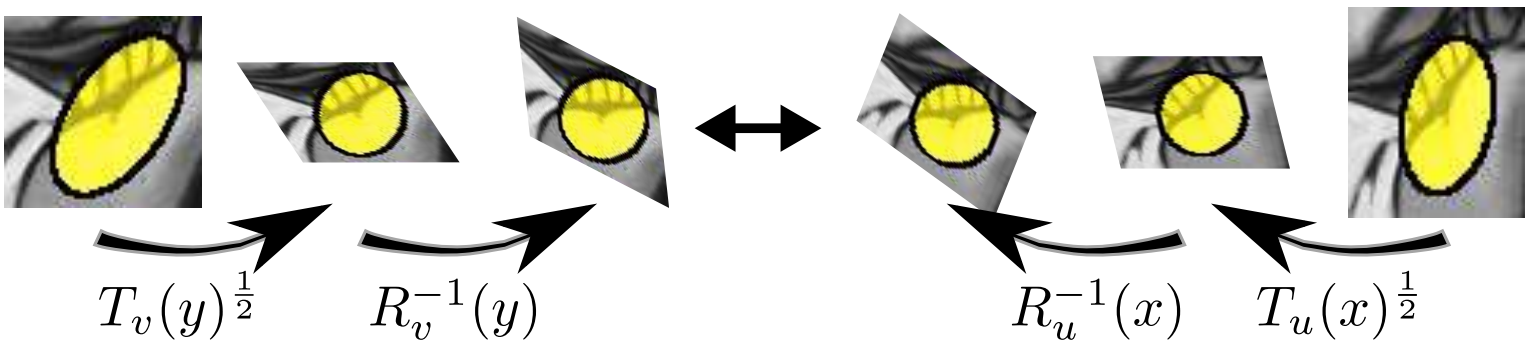
$$P(x,y) = T_v(y)^{-\frac{1}{2}} R_v(y) R_u^{-1}(x) T_u(x)^{\frac{1}{2}},$$

Affine Invariant Patch Comparison; Geometrical meaning

In practice it is more suitable to split the rotation into two parts

$$P(x,y) = T_v(y)^{-\frac{1}{2}} R_v(y) R_u^{-1}(x) T_u(x)^{\frac{1}{2}},$$

and to compare normalized patches

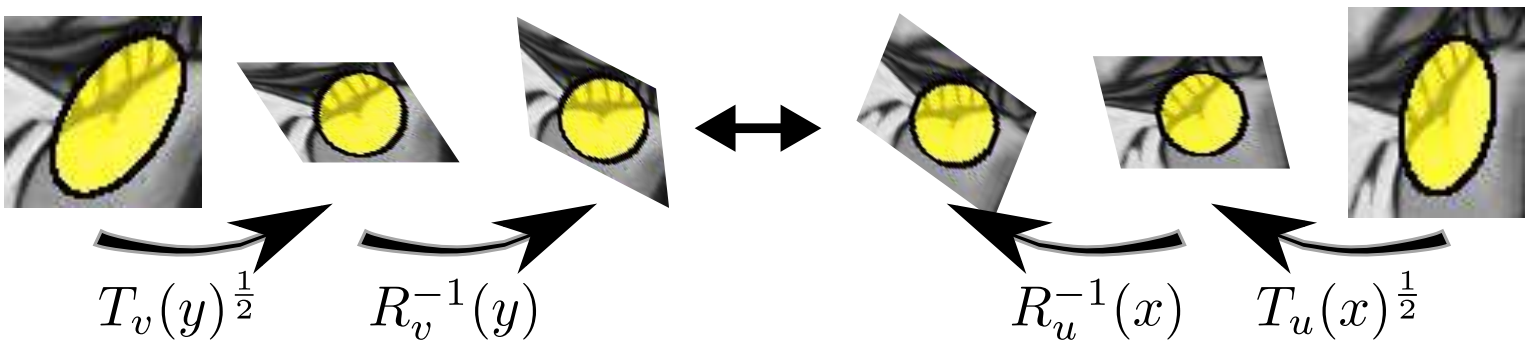


Affine Invariant Patch Comparison; Geometrical meaning

In practice it is more suitable to split the rotation into two parts

$$P(x,y) = T_v(y)^{-\frac{1}{2}} R_v(y) R_u^{-1}(x) T_u(x)^{\frac{1}{2}},$$

and to compare normalized patches



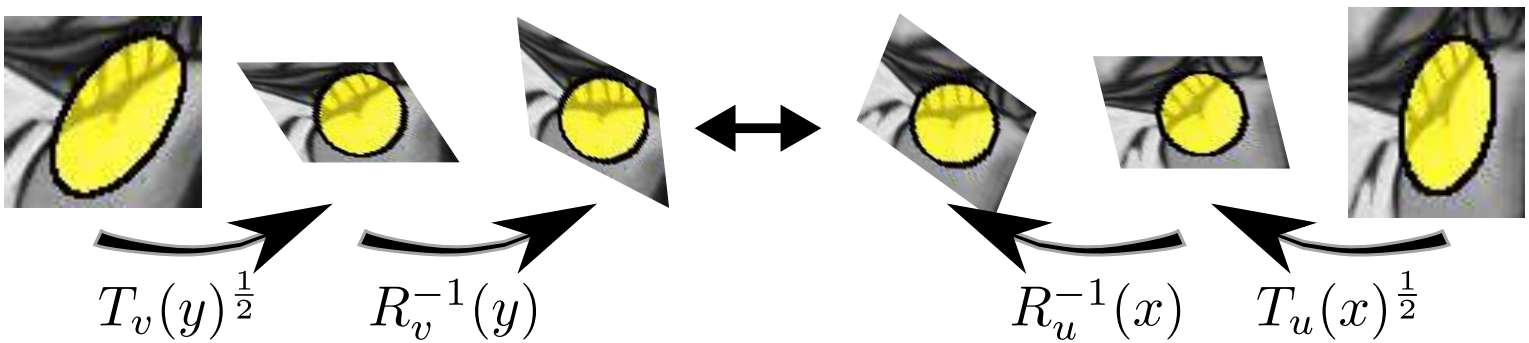
This is what happens in our similarity measure.

Affine Invariant Patch Comparison; Geometrical meaning

In practice it is more suitable to split the rotation into two parts

$$P(x,y) = T_v(y)^{-\frac{1}{2}} R_v(y) R_u^{-1}(x) T_u(x)^{\frac{1}{2}},$$

and to compare normalized patches



This is what happens in our similarity measure. It becomes
$(G_1^{-\frac{1}{2}}(x) = T_u(x)^{-\frac{1}{2}} R_u(x)^{-1}, \quad G_2^{-\frac{1}{2}}(y) = T_v(y)^{-\frac{1}{2}} R_v(y)^{-1})$

$$\mathcal{D}^{\mathsf{a}}(t,x,y) = \int_{\mathbb{R}^2} \eta_t(h) \left(u(x + T_u(x)^{-\frac{1}{2}} R_u(x)^{-1} h) - v(y + T_v(y)^{-\frac{1}{2}} R_v(y)^{-1} h) \right)^2 dh,$$

where η_t is an approximated geodesic weighting function or a Gaussian of variance t.

Also in video

Appropriate structure tensors as metrics in video $u(\mathbf{x}, t)$

$$T_u(\mathbf{x}, t) = \int_{\mathcal{E}_g((\mathbf{x},t),r)} D_{\mathbf{x}t}u((\mathbf{x}, t)+(\mathbf{y}, \tau)) \otimes D_{\mathbf{x}t}u((\mathbf{x}, t)+(\mathbf{y}, \tau))\mu(\mathbf{y}, \tau)|G|^{1/2}\, d\mathbf{y}d\tau.$$

where $\mathcal{E}_g((\mathbf{x}, t), r) = \left\{Y = (\mathbf{y}, s) : g_{ij}(\mathbf{x}, t)Y^i Y^j \leq r^2\right\}$, g is an initial metric on $\mathcal{M}$, $|G|$ denotes the determinant of the symmetric matrix $G = (g_{ij})$ and μ is a weight measure on $\mathcal{E}_g((\mathbf{x}, t), r)$, be either the usual Lebesgue measure or a weighting function. Our initial metric g is given by

$$g(\mathbf{x}, t)((\mathbf{y}, \tau), (\mathbf{y}, \tau)) = a(\mathbf{x}, t)(\mathbf{y} - \mathbf{v}(\mathbf{x}, t)\tau)^2 + b(\mathbf{x}, t)\tau^2,$$

that is, $G(\mathbf{x}, t) = \begin{pmatrix} aI & -a\mathbf{v} \\ (-a\mathbf{v})^t & b + a|\mathbf{v}|^2 \end{pmatrix}(\mathbf{x}, t)$, where $\mathbf{v}$ is the optical flow of the input video u, $(\mathbf{y}, \tau)$ denote coordinates in an infinitesimal neighborhood of $(\mathbf{x}, t)$ and the functions a and b are defined as $a(\mathbf{x}, t) = \alpha_1 + \alpha_2|\nabla_{\mathbf{x}}u|^p$, $b(\mathbf{x}, t) = \beta_1 + \beta_2(\partial_{\mathbf{v}}u)^p$, with $\alpha_1, \alpha_2, \beta_1, \beta_2 > 0$, $p > 0$ (usually $p = 1, 2$), and $\partial_v u = \mathbf{v} \cdot \nabla_x u + u_t$ denotes the convective derivative.

P. Vitoria, V. Fedorov and C. Ballester. *Spatio-temporal tube segmentation through a video metrics-based patch similarity measure.* IMVIP 2017.

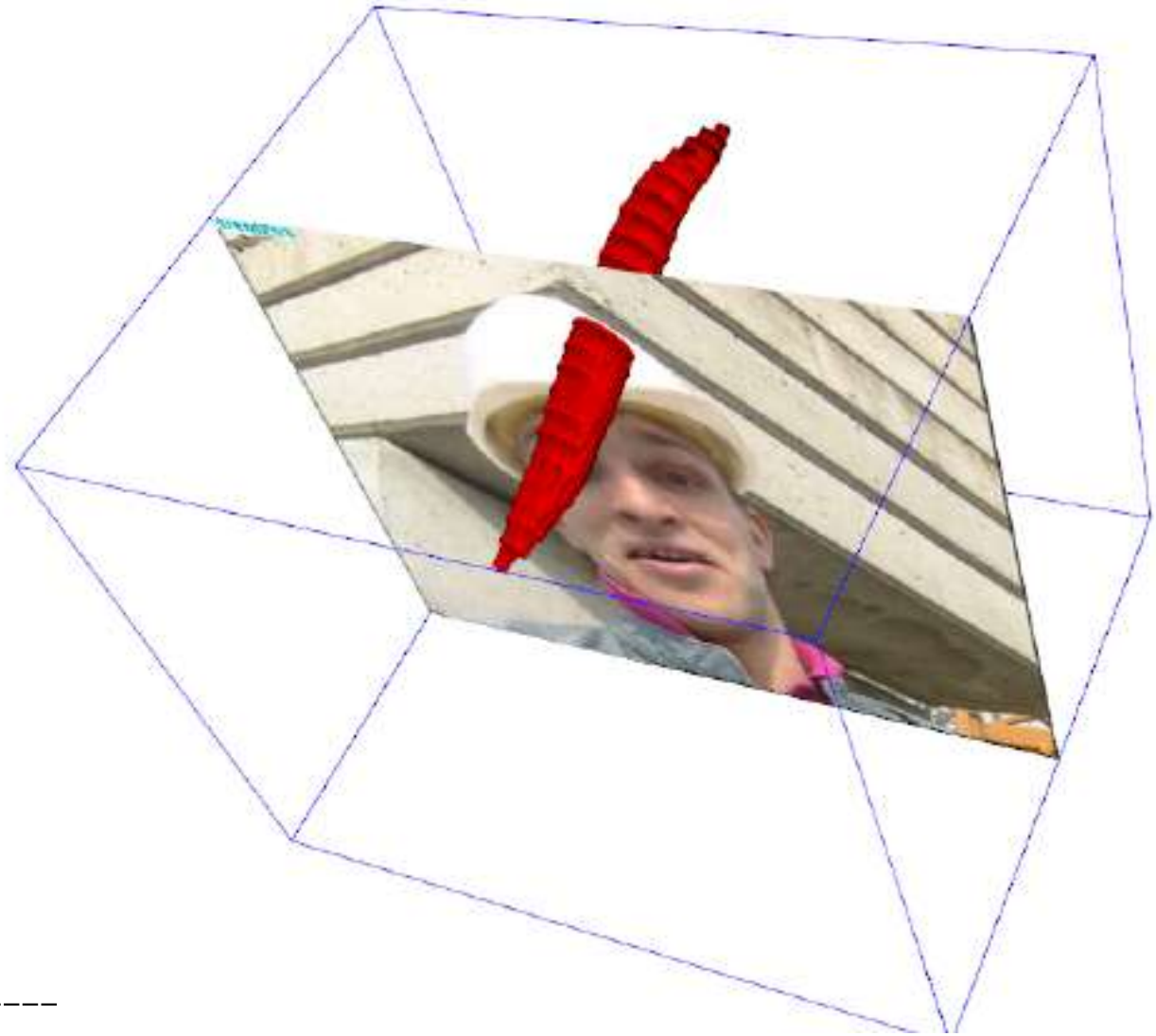
Iterative algorithm that simulates affine covariance

$$T_u^{(k)}(\mathbf{x},t) = \frac{\int_{B_{NT_u^{(k-1)}}((\mathbf{x},t),r)} D_{\mathbf{x}t}u((\mathbf{x},t)+(\mathbf{y},\tau)) \otimes D_{\mathbf{x}t}u((\mathbf{x},t)+(\mathbf{y},\tau))\mu(\mathbf{y},\tau)|G|^{1/2}\, d(\mathbf{y},\tau)}{\text{Volume}\left(B_{T_u^{(k-1)}}((\mathbf{x},t),r)\right)}$$

and

$$B_{T_u^{(k)}}((\mathbf{x},t),r) = \begin{cases} \left\{(\mathbf{y},\tau) : \langle G((\mathbf{x},t))\left((\mathbf{y},\tau)-(\mathbf{x},t)\right), \left((\mathbf{y},\tau)-(\mathbf{x},t)\right)\rangle \leq r^2\right\} & \text{when } k=0 \\ \left\{(\mathbf{y},\tau) : \langle T_u^{(k)}(\mathbf{x},t)\left((\mathbf{y},\tau)-(\mathbf{x},t)\right), \left((\mathbf{y},\tau)-(\mathbf{x},t)\right)\rangle \leq r^2\right\} & \text{when } k>0 \end{cases}$$

Structure tensors as metrics in video



P. Vitoria, V. Fedorov and C. Ballester. *Spatio-temporal tube segmentation through a video metrics-based patch similarity measure.* IMVIP 2017.

Structure tensors as metrics in video, synthetic example

Structure tensors as metrics in video, synthetic example

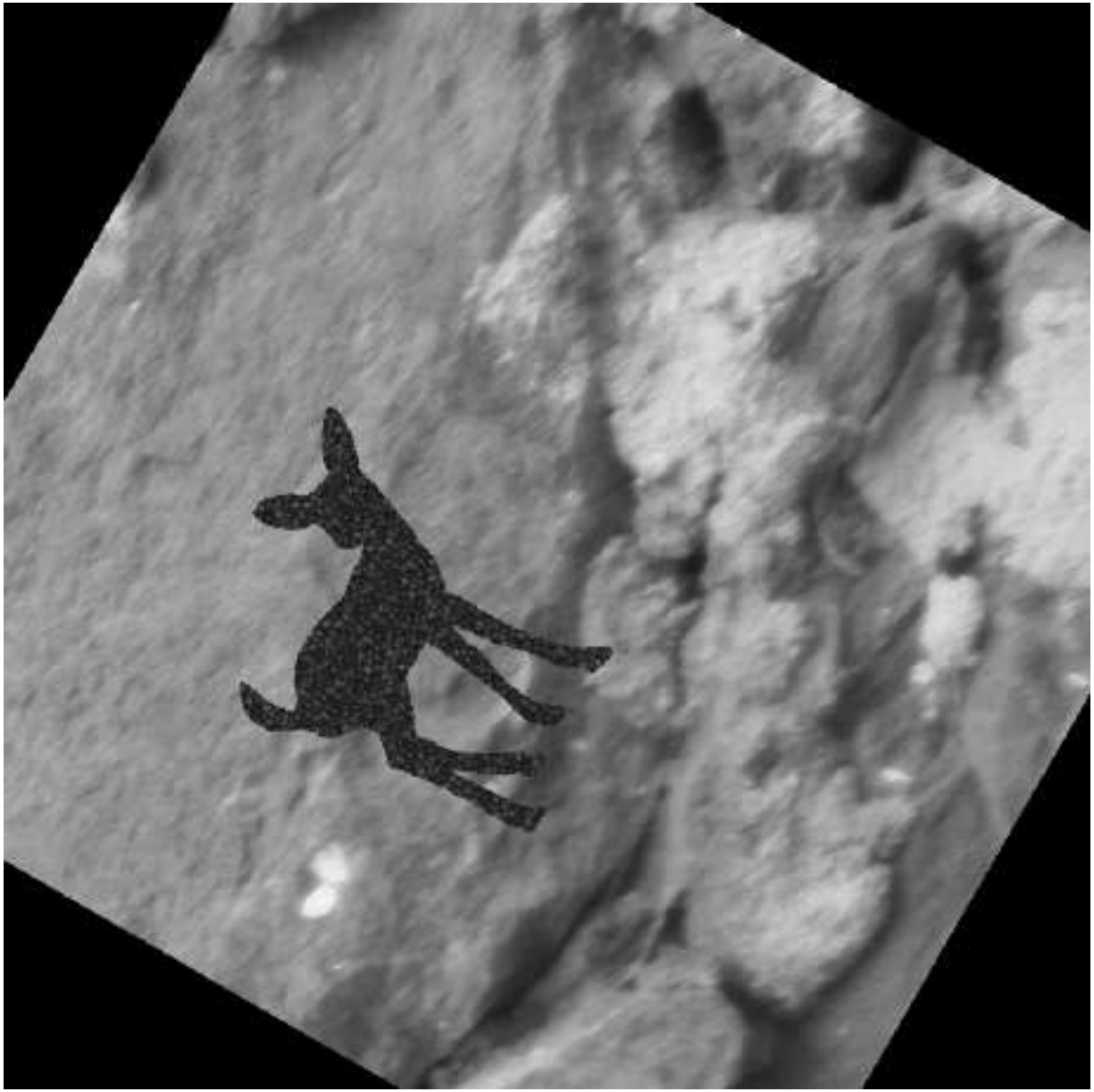
Structure tensors as metrics in video, synthetic example

Structure tensors as metrics in video, synthetic example

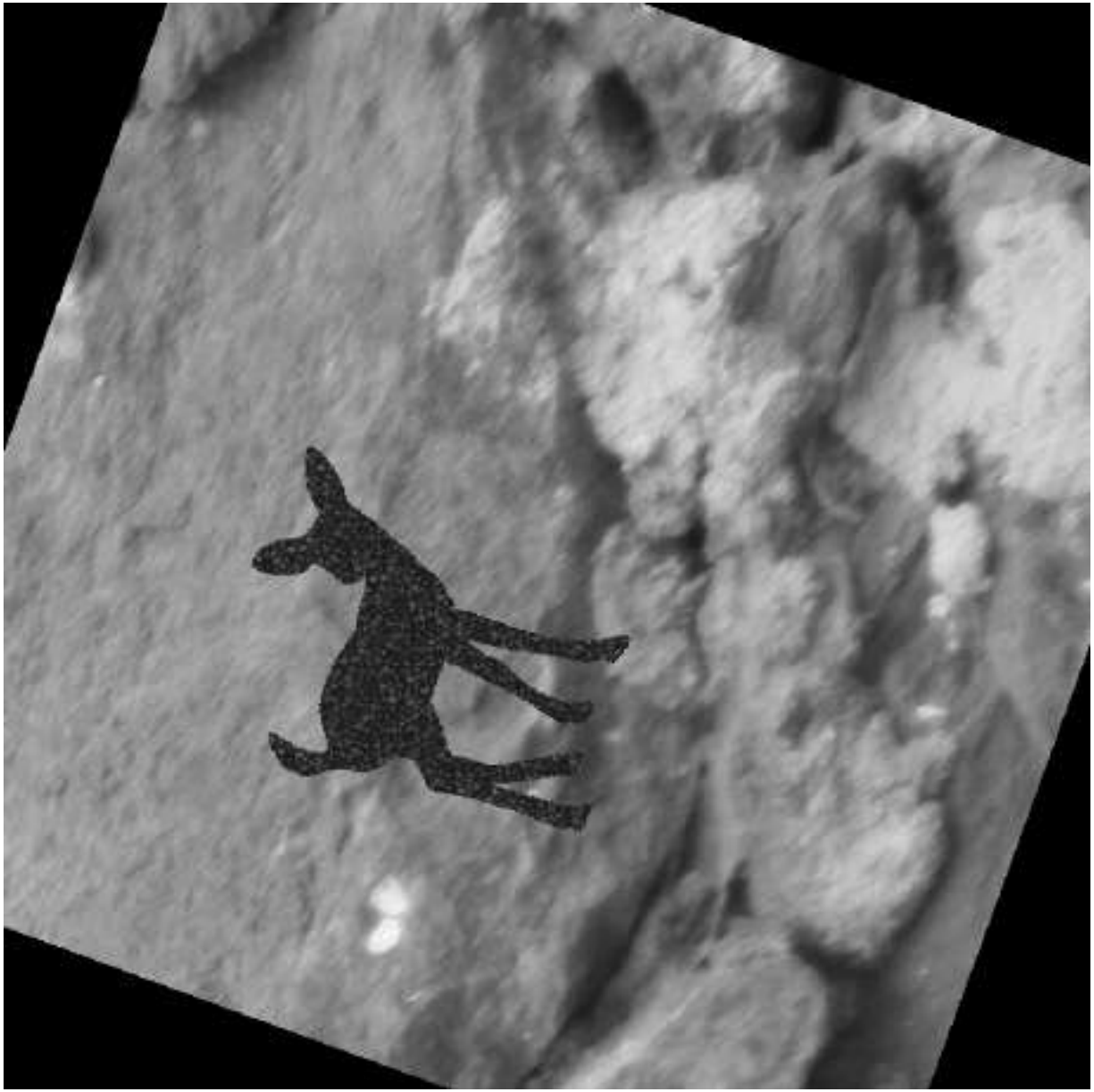
Structure tensors as metrics in video, synthetic example

Structure tensors as metrics in video, synthetic example

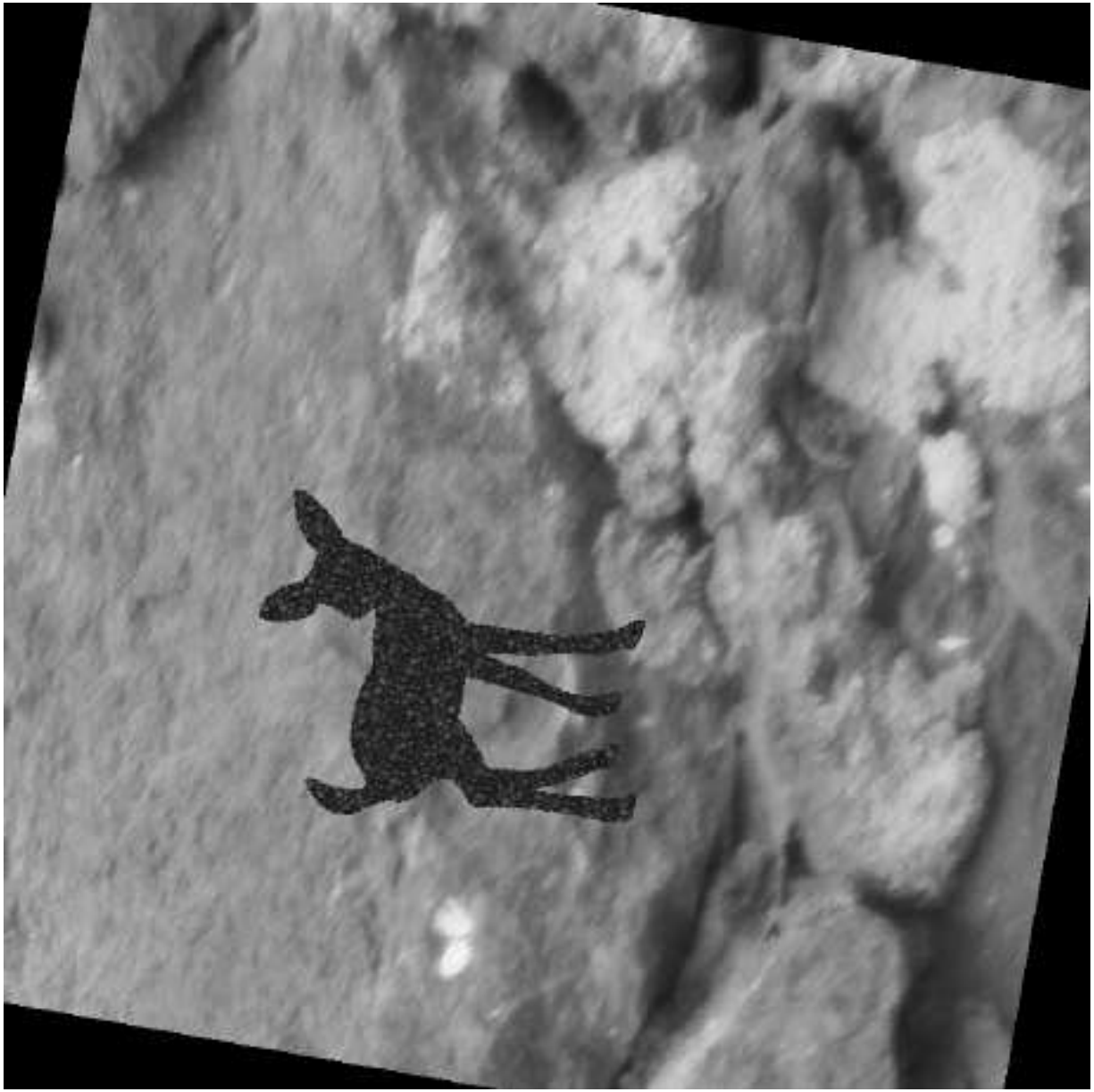
Structure tensors as metrics in video, synthetic example



Structure tensors as metrics in video, synthetic example



Structure tensors as metrics in video, synthetic example



Structure tensors as metrics in video, synthetic example

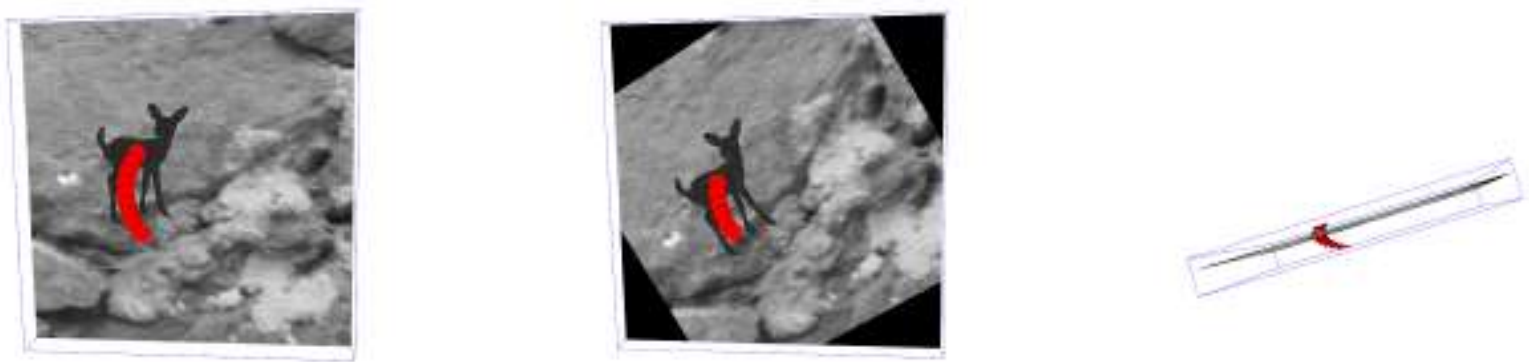


Figure : Visualization of 3D-shape adaptive patches (in red) corresponding to the tensor on a **point on the dear**.

Image source: Patricia Vitoria

Structure tensors as metrics in video

Figure : Visualization of a 3D-shape adaptative patch (red) $\mathcal{E}$ corresponding to the tensor on a **point on a ear of the deer**.

Image source: Patricia Vitoria

Structure tensors as metrics in video

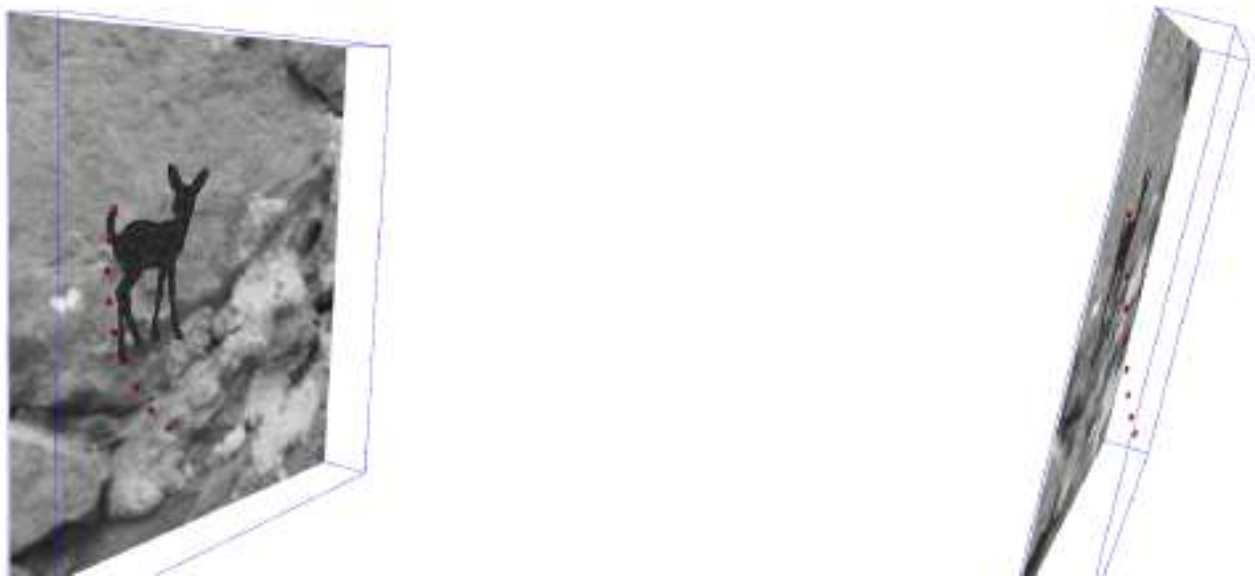


Figure : Visualization of a 3D-shape adaptative patch (red) $\mathcal{E}$ corresponding to the tensor on a **point on the tail of the deer**.

Image source: Patricia Vitoria

Structure tensors as metrics in video

Structure tensors as metrics in video

Structure tensors as metrics in video

Structure tensors as metrics in video

Structure tensors as metrics in video

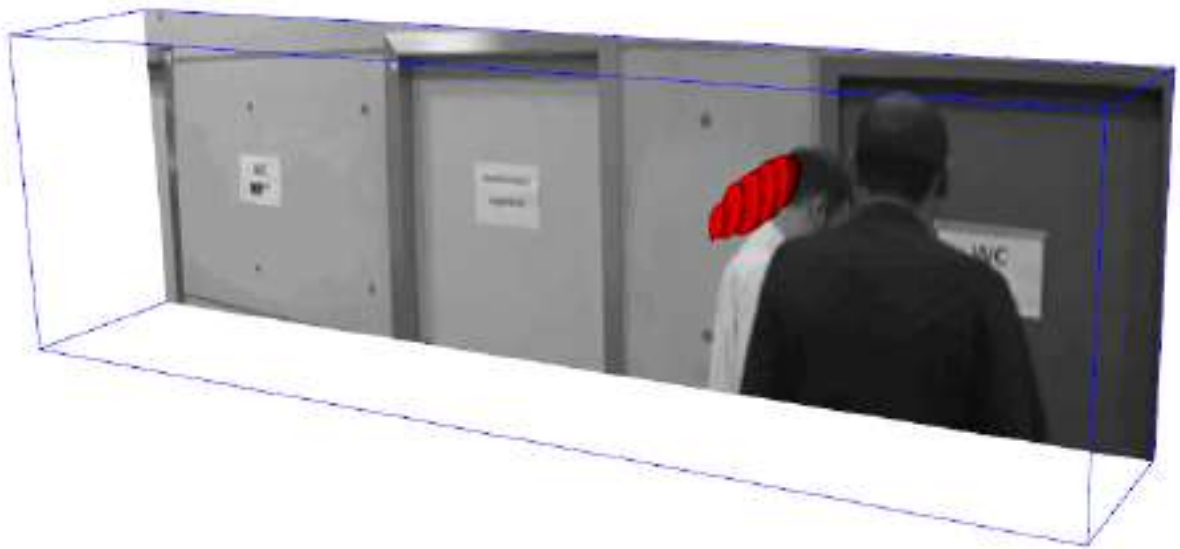
Structure tensors as metrics in video

Structure tensors as metrics in video

Structure tensors as metrics in video

Structure tensors as metrics in video

Structure tensors as metrics in video



First Row: visualization of the first frame and the patch corresponding to a point x that belongs to the hair of the white t-shirt's man.

Second row: Last three frames: once the man is occluded (last frame), the patch stops.

Applications

1. Image and Video denoising.
2. Exemplar-based inpainting.
3. Image segmentation
4. Video simplification/segmentation, spatio-temporal tubes for video analysis.
5. Depth completion by a geodesic Biased AMLE method
6. Motion inpainting by an image based geodesic AMLE method
7. Dynamic shape disocclusion.

1. Image and Video denoising

Denoising problem

Common noise model for images obtained with charge-coupled device (CCD):

$$\widetilde{u}(x) = u(x) + n(x)$$

where $\widetilde{u}(x)$ is a Poisson random variable, $u(x)$ is its mean, and $n(x)$ is added noise.

Using the Anscombe transform Poisson noise can be converted into Gaussian noise.

Reminder: $X_1, X_2, \ldots, X_n, \ldots$ i.i.d. random variables with zero mean, variance σ^2. **They represent noise** and σ^2 represents the **power of the noise**.
Let

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

Since $E(\overline{X}_n) = 0$, the variance of $\overline{X}_n$ is $E((\overline{X}_n - E(\overline{X}_n))^2) = \frac{\sigma^2}{n}$. **The basic principle for denoising: averaging reduces noise variance by a factor n.**

Reminder: $X_1, X_2, \ldots, X_n, \ldots$ i.i.d. random variables with zero mean, variance σ^2. **They represent noise** and σ^2 represents the **power of the noise**.
Let

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

Since $E(\overline{X}_n) = 0$, the variance of $\overline{X}_n$ is $E((\overline{X}_n - E(\overline{X}_n))^2) = \frac{\sigma^2}{n}$. **The basic principle for denoising: averaging reduces noise variance by a factor n.**

Assume now that for each $i = 1, 2, \ldots$, $U_i = a + X_i$, where a is a common value (representing the clean data), and X_i represents noise (i.i.d., zero mean, variance σ^2). Let us average U_i:

$$\overline{U}_n = \frac{U_1 + U_2 + \ldots + U_n}{n} = a + \frac{X_1 + X_2 + \ldots + X_n}{n} = a + \overline{X}_n.$$

Again, $E((\overline{U}_n - E(\overline{U}_n))^2) = E(\overline{X}_n^2) = \frac{\sigma^2}{n}$.
Then: If we have observed n values of a signal U_i and we are sure that these values share the same deterministic value a and differ only in the noise, its average is an estimation of a that reduces the noise power (variance) by a factor n.

Reminder: $X_1, X_2, \ldots, X_n, \ldots$ i.i.d. random variables with zero mean, variance σ^2. **They represent noise** and σ^2 represents the **power of the noise**.
Let

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

Since $E(\overline{X}_n) = 0$, the variance of $\overline{X}_n$ is $E((\overline{X}_n - E(\overline{X}_n))^2) = \frac{\sigma^2}{n}$. **The basic principle for denoising: averaging reduces noise variance by a factor n.**

Assume now that for each $i = 1, 2, \ldots$, $U_i = a + X_i$, where a is a common value (representing the clean data), and X_i represents noise (i.i.d., zero mean, variance σ^2). Let us average U_i:

$$\overline{U}_n = \frac{U_1 + U_2 + \ldots + U_n}{n} = a + \frac{X_1 + X_2 + \ldots + X_n}{n} = a + \overline{X}_n.$$

Again, $E((\overline{U}_n - E(\overline{U}_n))^2) = E(\overline{X}_n^2) = \frac{\sigma^2}{n}$.
Then: If we have observed n values of a signal U_i and we are sure that these values share the same deterministic value a and differ only in the noise, its average is an estimation of a that reduces the noise power (variance) by a factor n.
Difficulty: How do we ensure that the observed values share the same deterministic signal ?

Reminder: $X_1, X_2, \ldots, X_n, \ldots$ i.i.d. random variables with zero mean, variance σ^2. **They represent noise** and σ^2 represents the **power of the noise**.
Let

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

Since $E(\overline{X}_n) = 0$, the variance of $\overline{X}_n$ is $E((\overline{X}_n - E(\overline{X}_n))^2) = \frac{\sigma^2}{n}$. **The basic principle for denoising: averaging reduces noise variance by a factor n.**

Assume now that for each $i = 1, 2, \ldots$, $U_i = a + X_i$, where a is a common value (representing the clean data), and X_i represents noise (i.i.d., zero mean, variance σ^2). Let us average U_i:

$$\overline{U}_n = \frac{U_1 + U_2 + \ldots + U_n}{n} = a + \frac{X_1 + X_2 + \ldots + X_n}{n} = a + \overline{X}_n.$$

Again, $E((\overline{U}_n - E(\overline{U}_n))^2) = E(\overline{X}_n^2) = \frac{\sigma^2}{n}$.
Then: If we have observed n values of a signal U_i and we are sure that these values share the same deterministic value a and differ only in the noise, its average is an estimation of a that reduces the noise power (variance) by a factor n.
Difficulty: How do we ensure that the observed values share the same deterministic signal ?

For that we use **weights**.

Non-Local Means denoising

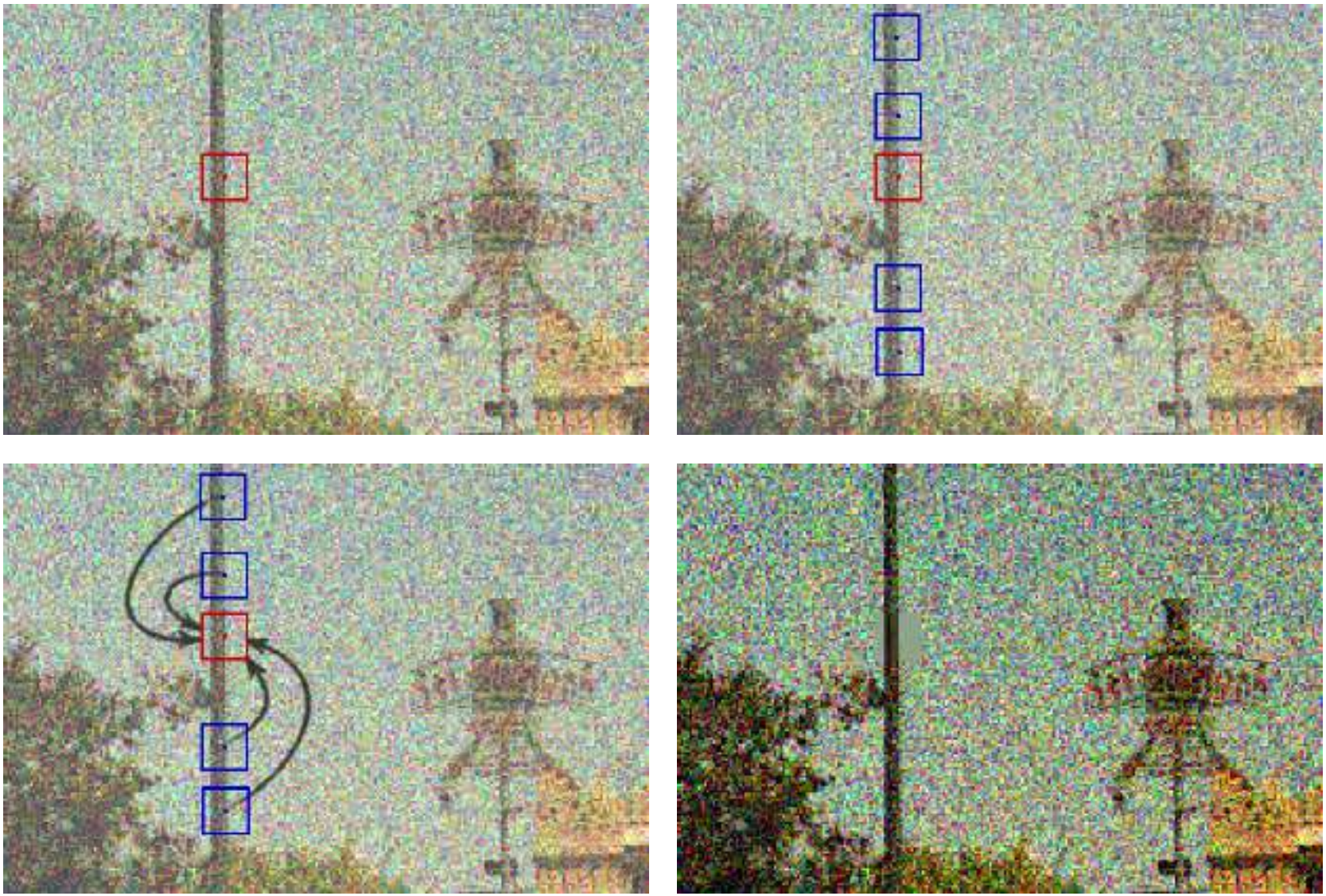
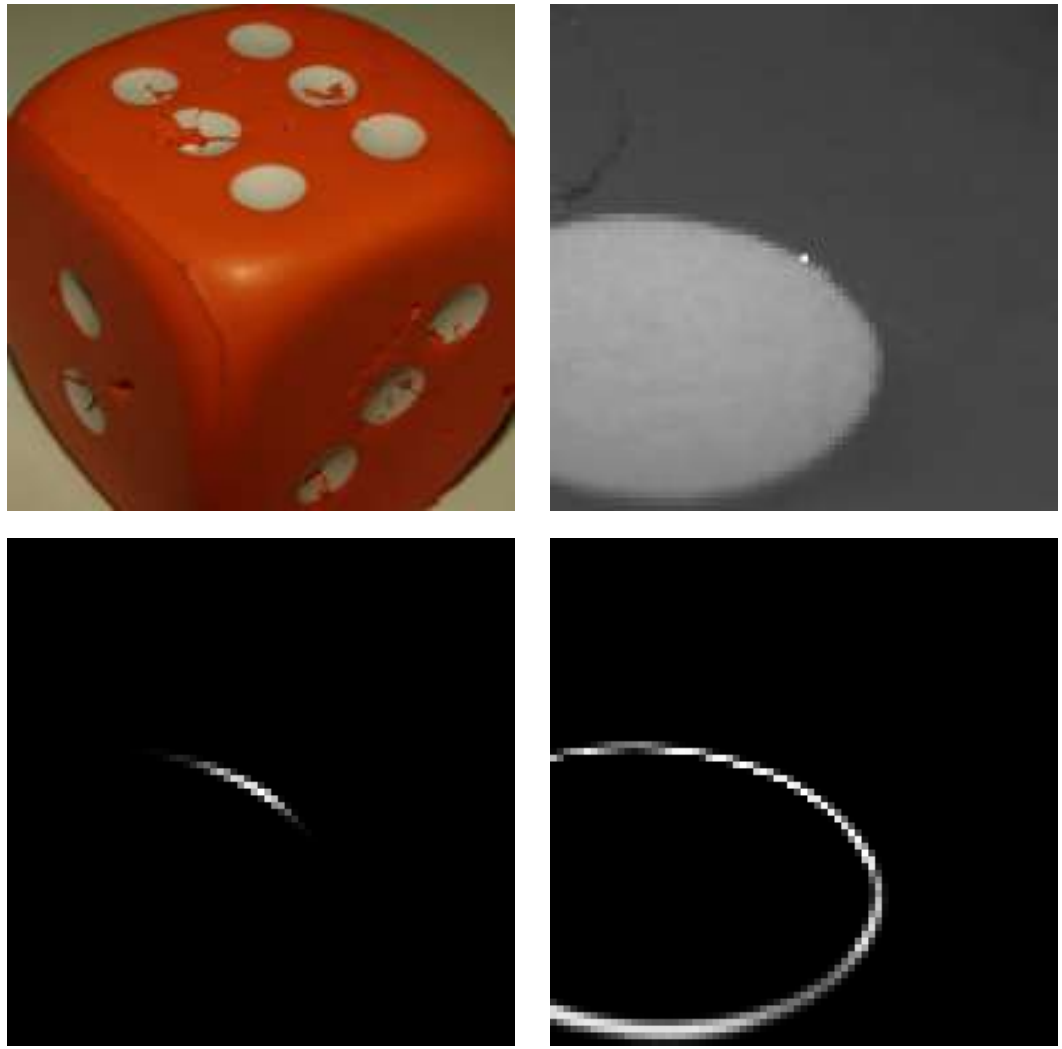


Image source: Vadim Fedorov

Non-Local Means denoising



V. Fedorov and C. Ballester. *Affine Non-local Means Image Denoising.* IEEE Transactions on Image Processing. 2017.

Non-Local Means denoising

Let $\mathbf{u}$ be a (noisy) color image in RGB color space.

Let $p_u(y)$ be a noisy patch of u at y, then the denoised patch is

$$\widehat{p}_u(x) = \frac{1}{\sum_y \mathcal{S}(x,y)} \sum_{y \in W(x)} \mathcal{S}(x,y) \cdot P(y,x) p_u(y),$$

where

$$\mathcal{S}(x,y) = e^{-\frac{\mathcal{D}^{\mathsf{a}}(t,x,y)}{\lambda^2}} \quad \text{and} \quad P(y,x) = T_u(x)^{-\frac{1}{2}} R_u(x)^{-1} R_u(y) T_u(y)^{\frac{1}{2}}.$$

$$\mathcal{D}^{\mathsf{a}}(t,x,y) =$$
$$\int_{\Omega_P} g_t(h) \left\| \mathbf{u}(x + T_u(x)^{-\frac{1}{2}} R_u(x)^{-1} h) - \mathbf{u}(y + T_u(y)^{-\frac{1}{2}} R_u(y)^{-1} h) \right\|_2^2 dh,$$

Image denoising

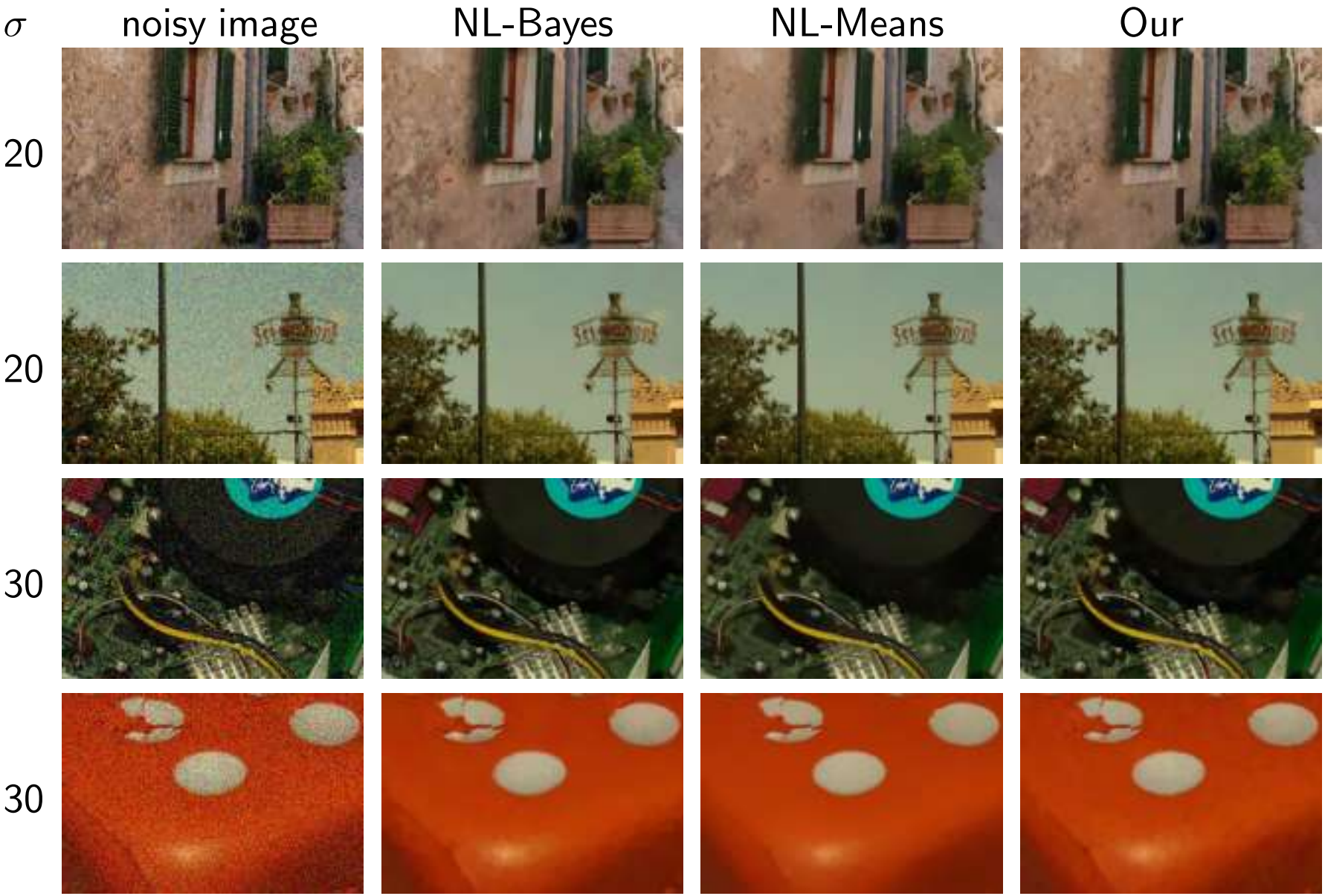
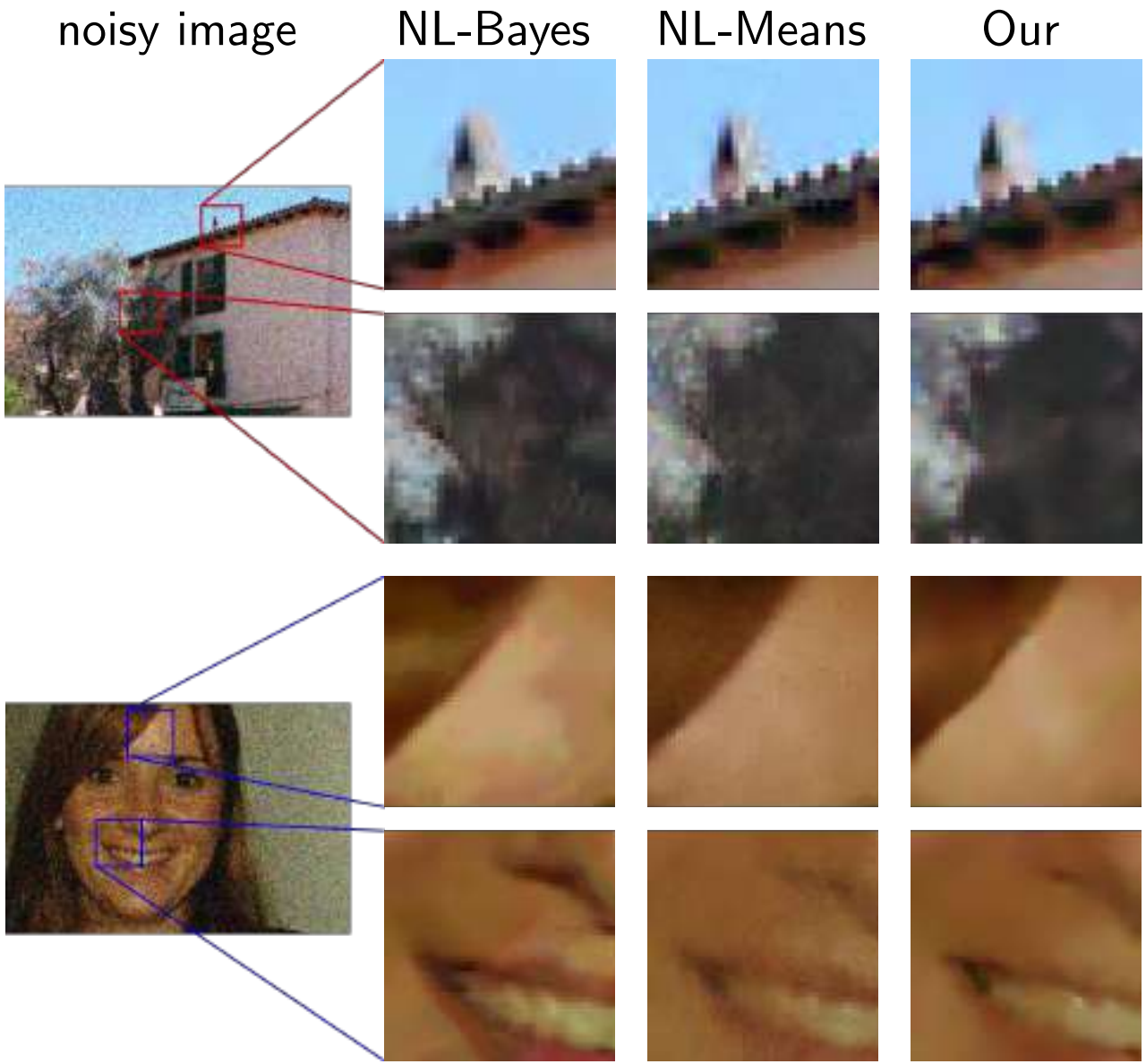


Image denoising

Image denoising



Quantitative comparison (PSNR)

Table : PSNR values for noise $\sigma = 2, 5, 10, 20, 30$ and 40

	NL-Bayes	NL-means	**our**	NL-Bayes	NL-means	**our**
	$\sigma = 2$			$\sigma = 20$		
Alley	**45.28**	42.68	43.37	**31.17**	29.98	30.15
Computer	**45.81**	43.93	44.67	**32.98**	31.67	32.03
Dice	**49.17**	48.12	48.22	**40.17**	38.31	39.62
Flowers	**47.75**	46.31	46.89	**36.14**	34.53	36.08
Girl	**47.67**	46.71	46.71	**38.62**	36.92	38.02
Traffic	**45.17**	43.45	44.00	**31.24**	30.14	30.46
Trees	**43.44**	42.15	42.62	**27.36**	26.37	26.40
Valldemossa	**45.07**	43.26	43.67	**29.72**	28.44	28.51
	$\sigma = 5$			$\sigma = 30$		
Alley	**39.14**	37.25	37.31	**29.15**	27.85	28.37
Computer	**40.54**	38.93	39.15	**30.68**	29.28	30.06
Dice	**46.02**	44.93	45.22	**37.95**	36.92	37.54
Flowers	**43.29**	42.17	42.74	33.85	32.35	**34.11**
Girl	**44.18**	43.36	43.37	**36.69**	35.58	36.25
Traffic	**39.39**	37.59	38.01	**29.03**	27.74	28.54
Trees	**36.54**	34.71	35.03	**25.03**	23.79	24.31
Valldemossa	**38.62**	35.96	36.70	**27.35**	25.89	26.46
	$\sigma = 10$			$\sigma = 40$		
Alley	**34.82**	33.64	33.55	**27.77**	26.48	27.11
Computer	**36.68**	35.54	35.40	**29.04**	27.61	28.57
Dice	**43.20**	41.92	42.63	**36.24**	35.26	36.05
Flowers	**39.53**	38.59	39.38	32.13	30.58	**32.70**
Girl	**41.43**	40.40	40.85	**35.06**	34.17	35.00
Traffic	**35.15**	34.05	34.10	**27.52**	26.23	27.17
Trees	**31.70**	29.59	30.31	**23.50**	22.42	22.88
Valldemossa	**33.96**	32.15	32.33	**25.81**	24.46	25.04

Video denoising

Figure : Left image: noisy frame from a sequence. Right image: corresponding denoised frame from a sequence.

P. Vitoria, Master thesis, 2017.

Video denoising

Figure : First column: Original frame. Second Column: noisy frame with $\sigma = 10$. Third Column: Denoised frame with $\sigma = 10$. Fourth Column: noisiy frame with $\sigma = 20$. Fifth Column: Denoised frame with $\sigma = 20$.

Video denoising

		army	cooper	dog	truck	avg
$\sigma = 10$	ND-SAFIR	3.99	5.73	5.81	5.02	5.14
	VBM3D	**2.96**	4.31	4.48	**3.76**	**3.88**
	VBM4D	3.26	4.48	4.47	4.04	4.06
	AI-NLM3D	3.79	**3.98**	**4.22**	4.03	4.00
$\sigma = 20$	ND-SAFIR	5.42	9.63	7.80	7.90	7.69
	VBM3D	**4.42**	6.84	**6.37**	**5.70**	**5.83**
	VBM4D	5.02	7.15	6.46	6.32	6.24
	AI-NLM3D	7.33	**5.99**	6.50	6.02	6.46
$\sigma = 30$	ND-SAFIR	6.62	12.42	9.39	10.44	9.72
	VBM3D	**5.54**	8.87	**7.75**	**7.30**	**7.37**
	VBM4D	6.40	9.25	8.00	8.12	7.94
	AI-NLM3D	8.83	**7.96**	8.63	8.10	8.38

Figure : RMSE results for Color sequences: the values correspond to the RMSE (Averaged over the three channels) computed for the central frame of each sequence. The average RMSE for each method and each noise level is displayed in the last column.

2. Exemplar-based inpainting

Exemplar-based inpainting

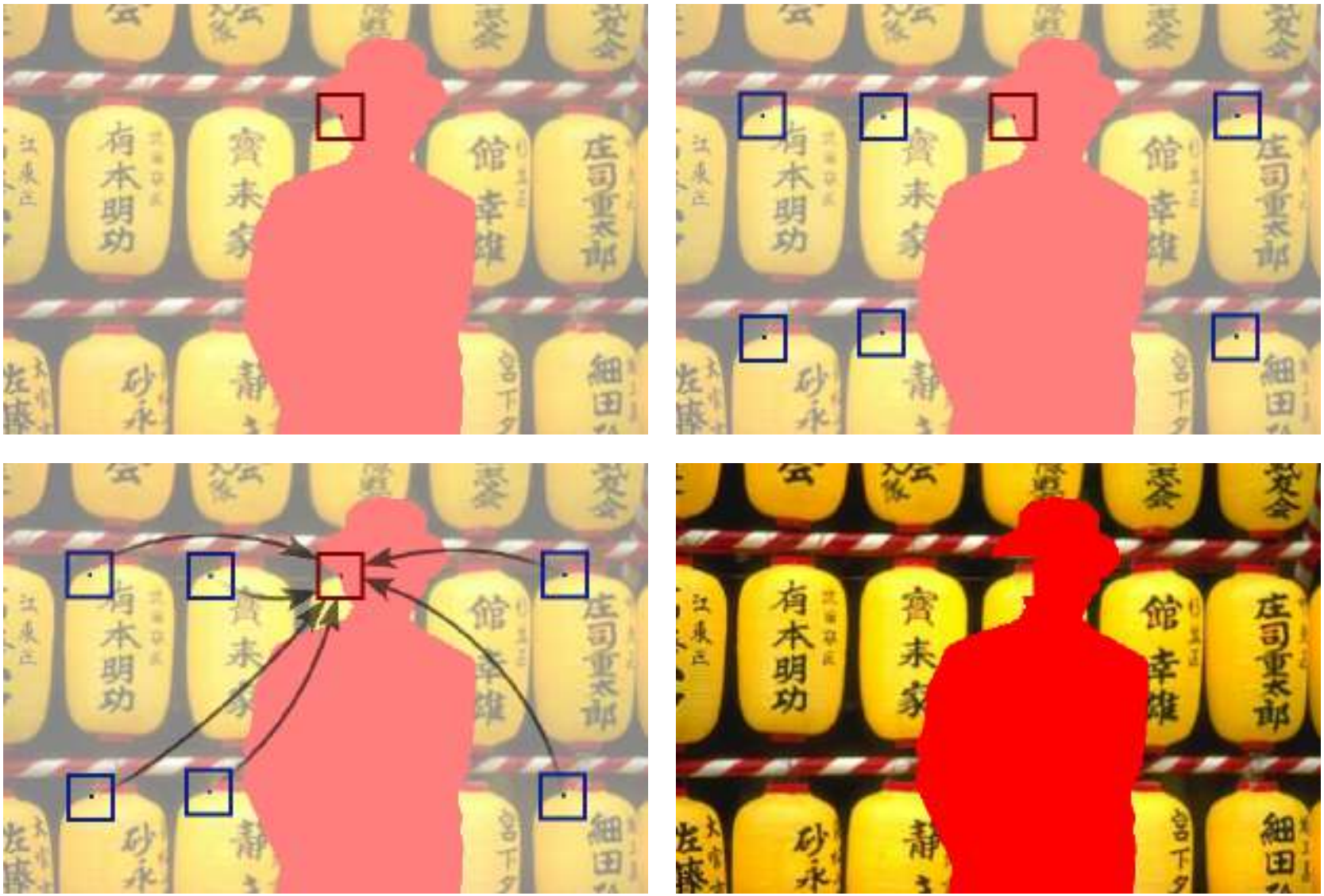
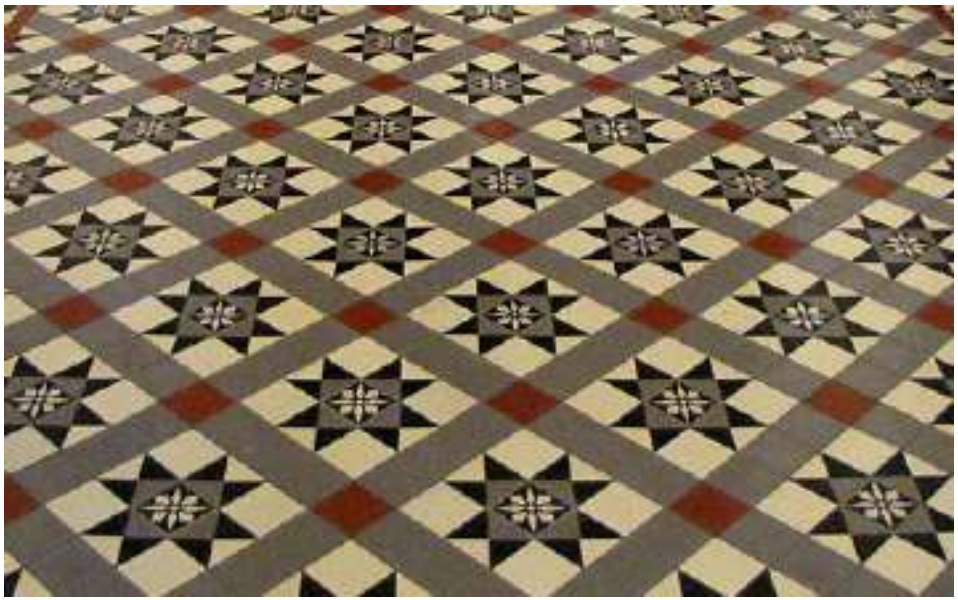
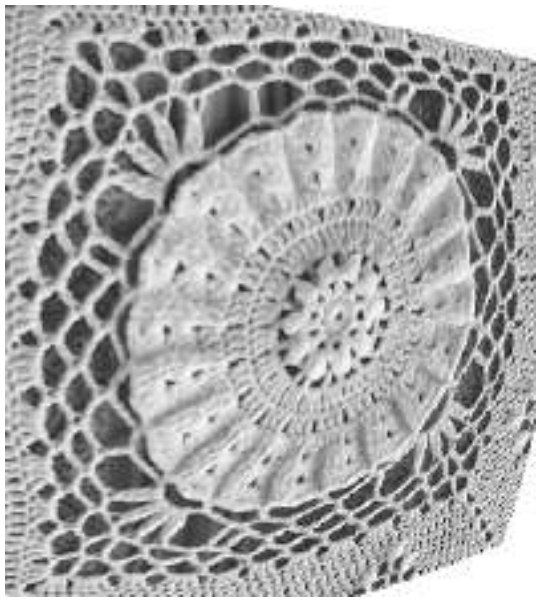


Image source: Vadim Fedorov

Now



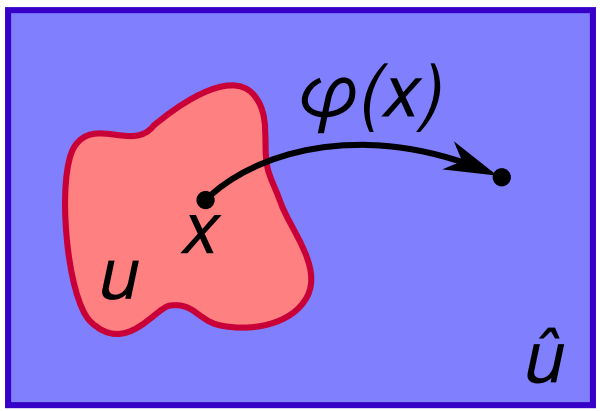


Affine-invariant self-similarity and its variational formulation for inpainting

We pose inpainting as minimization of the following energy functional:

$$E(u, \varphi) = \int_{\widetilde{O}} \mathcal{D}^{\mathsf{a}}\left(t, x, \varphi(x)\right)\, dx.$$

Here u is the unknown part and $\hat{u}$ is the known part of the image. $\widetilde{O}$ contains centers of incomplete patches, and φ is the correspondence map:

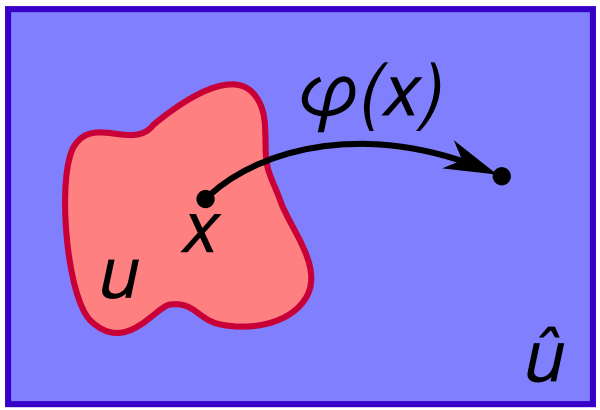


Affine-invariant self-similarity and its variational formulation for inpainting

We pose inpainting as minimization of the following energy functional:

$$E(u,\varphi) = \int_{\widetilde{O}} \mathcal{D}^{\mathsf{a}}\left(t,x,\varphi(x)\right)\, dx.$$

Here u is the unknown part and $\hat{u}$ is the known part of the image. $\widetilde{O}$ contains centers of incomplete patches, and φ is the correspondence map:

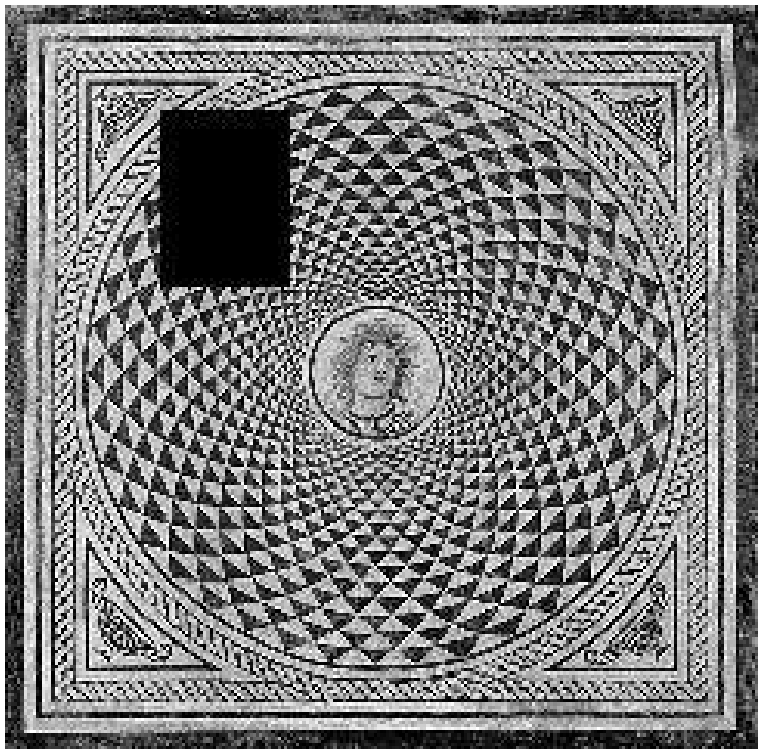


$$E(u,\varphi,G) =$$

$$\int_{\widetilde{O}}\int_{\Omega_p} g_t(h)\Big(u\big(x+G(x)^{-\frac{1}{2}}h\big)-\hat{u}\big(\varphi(x)+T_{\hat{u}}(\varphi(x))^{-\frac{1}{2}}R_{\hat{u}}(\varphi(x))^{-1}h\big)\Big)^2 dhdx$$

V. Fedorov, P. Arias, G. Facciolo, and C. Ballester, *Exemplar-Based Image Inpainting Using an Affine Invariant Similarity Measure.* International Joint Conference on Computer Vision, Imaging and Computer Graphics, 2016.

Inpainting results



Inpainting results

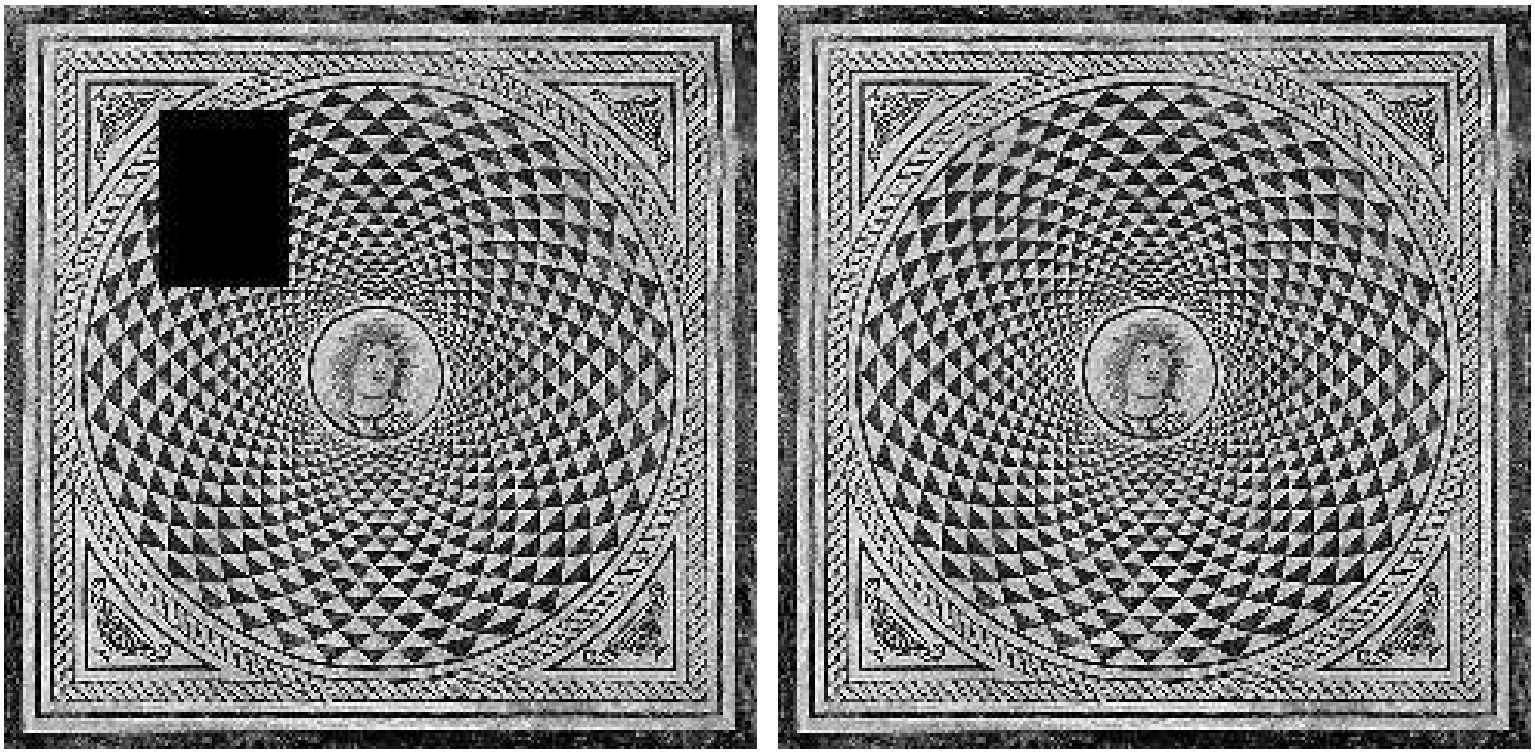


Image source: Vadim Fedorov

Inpainting results

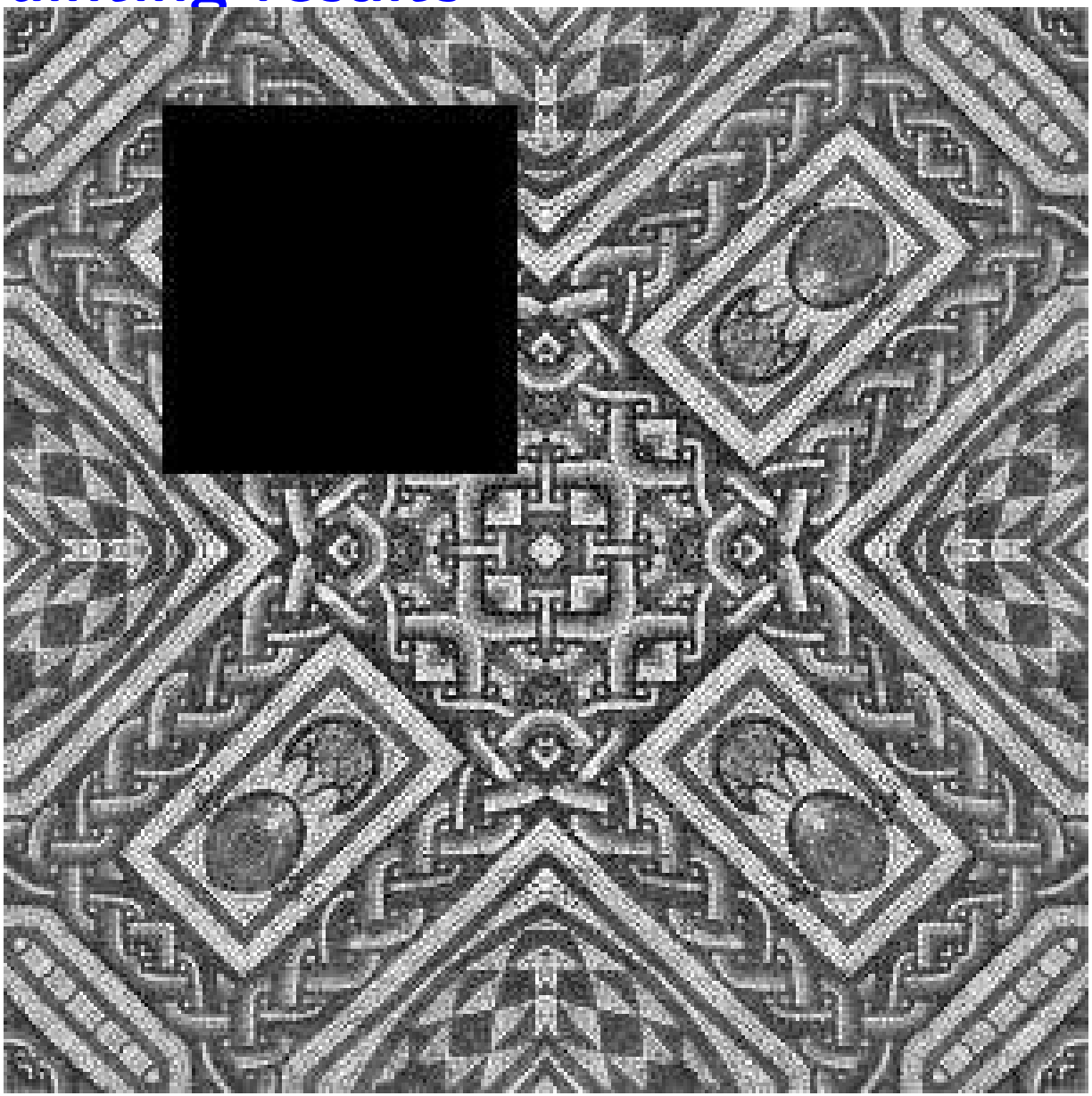


Image source: Vadim Fedorov

Inpainting results

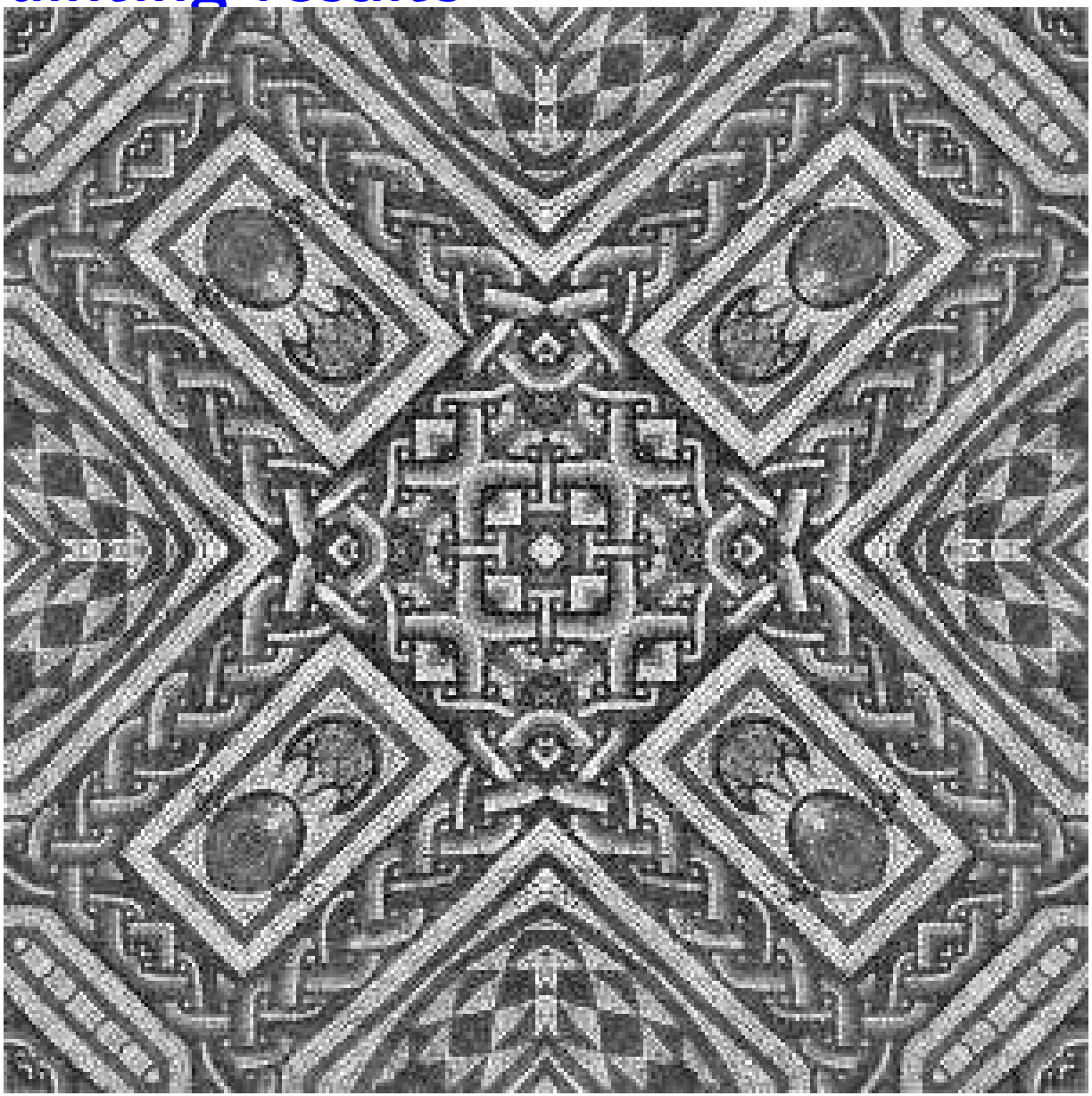


Image source: Vadim Fedorov

One view inpainted using another view

Figure : Source image, target image, [Wexler et al]-[Arias et al], and our method

One view inpainted using another view

Figure : Source image, target image, [Wexler et al]-[Arias et al], and our method

Projective transformation

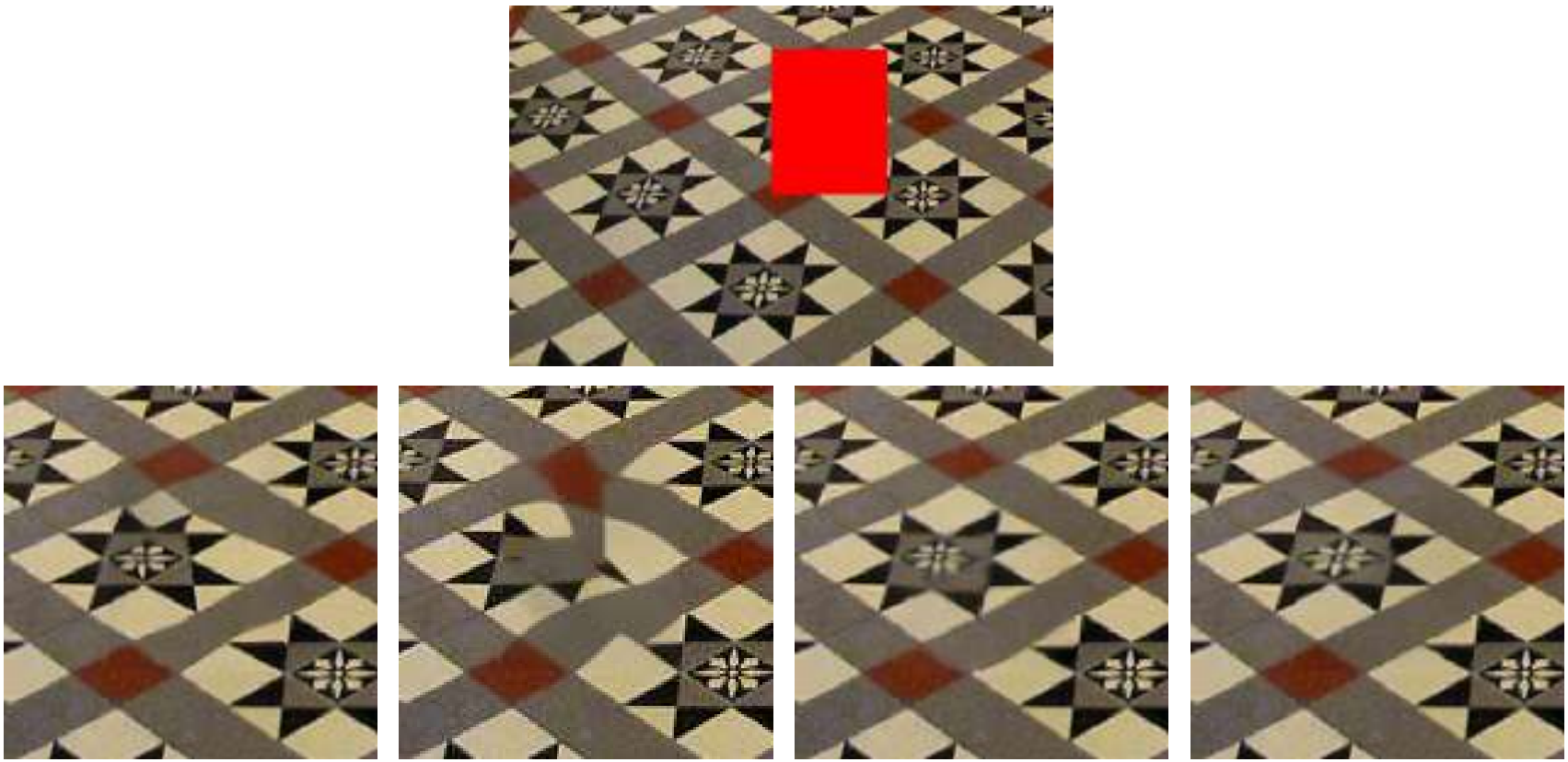


Figure : First row: image with the inpainting domain shown in red. Second row: [Wexler et al]-[Arias et al], method of Mansfield et al. (2011), method of Huang et al. (2014), and our method.

Lens distortion

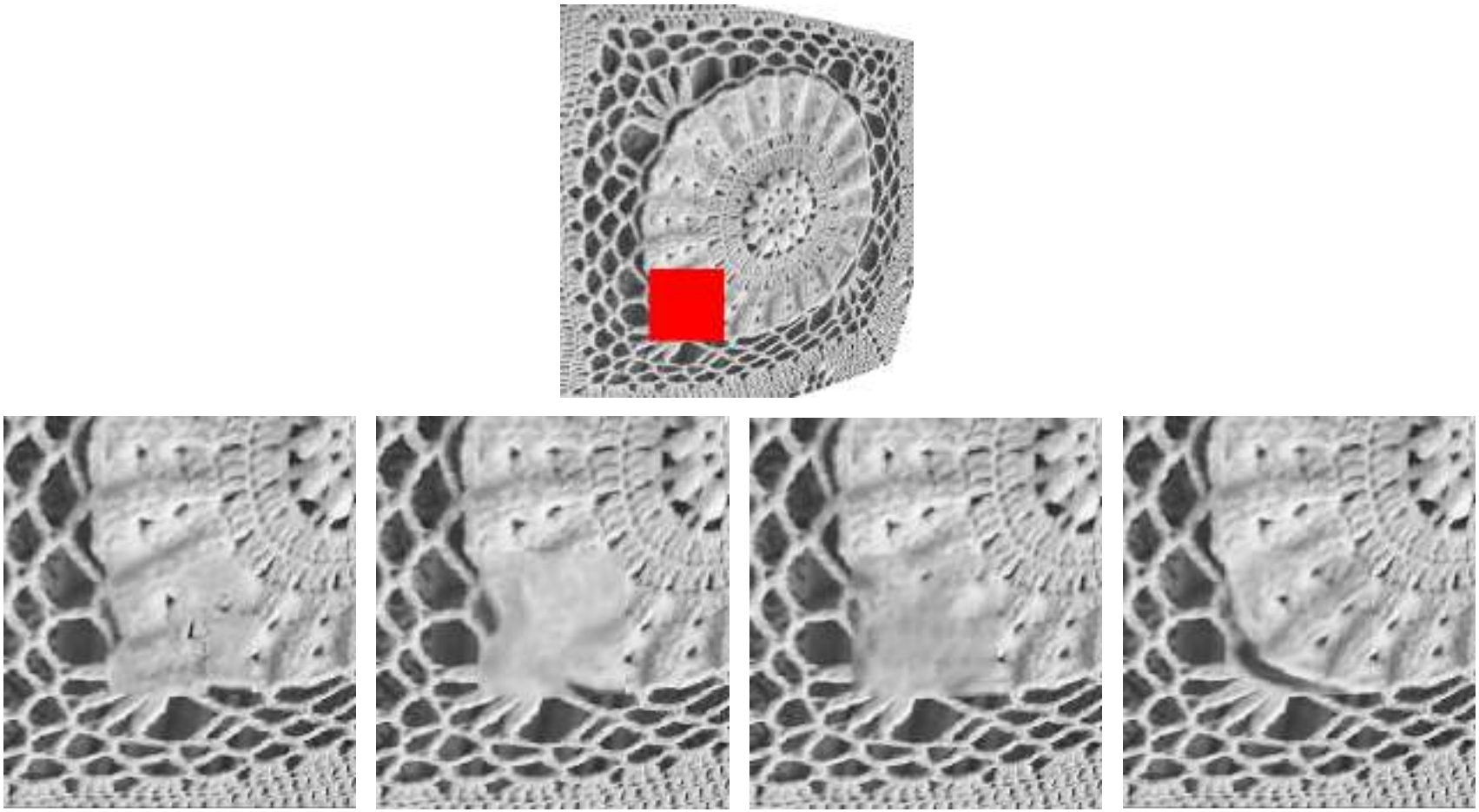


Figure : First row: image with the inpainting domain shown in red. Second row: [Wexler et al]-[Arias et al], method of Mansfield et al. (2011), method of Huang et al. (2014), and our method.

3. Image segmentation

Image segmentation into homogeneous textured regions

Segmentation: partition of an image into regions which share common features.

Goals

- Characterizing the image regions that have local homogeneous texture regardless of differences in the point of view or suffered local perspective or affine distortion.
- Obtaining a patch representative of the texture of each region.

How: a variational segmentation method that considers shape and size adaptive patches to characterize, in an affine way, the local structure of each homogeneous texture region of the image.

They are used in an $L1$-norm fidelity term and the total variation of fuzzy membership functions as relaxed length of the boundaries of the segmentation regions[1,2].

---

[1] M. Oliver, G. Haro, V. Fedorov and C. Ballester. *L1 Patch-based Image Partitioning into Homogeneous Textured Regions.* In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing. 2018.

[1] M. Oliver, SIAM Conference on Imaging Science, June 2018, Bologna (Italy) (Poster presentation) Best poster award (2nd position)

Image segmentation

Let $u : \Omega \to \mathbb{R}^M$ be an image, and let $T_u(x)$ be the spatially varying Riemannian metric given by the affine covariant tensor associated to u.

Let $\mathcal{P}_\mathbf{u}$ be the **set of all patches** obtained from u and defined using $T_u(x)$ at all x,

$$\mathcal{P}_\mathbf{u} = \{\mathbf{p_u}(\mathbf{x}), \mathbf{x} \in \mathbf{\Omega}\},$$

where $\mathbf{p}_u : \Omega \to L^q(\Delta_t)$ denote the function given by $\mathbf{p}_u(x) := \mathbf{p}_u(x, \cdot)$, where $\mathbf{p}_u(x, h) := u(x + T_u(x)^{-\frac{1}{2}} h)$, and $h \in \Delta_t$, a disc centered at the origin with radius proportional to t.

Let us notice that, thanks to the tensors, these elliptical patches can be considered defined on the *normalized disc* Δ_t.

From now on, with an abuse of notation, $\mathbf{p}_u(x)$ will denote the normalized discs.

Image segmentation

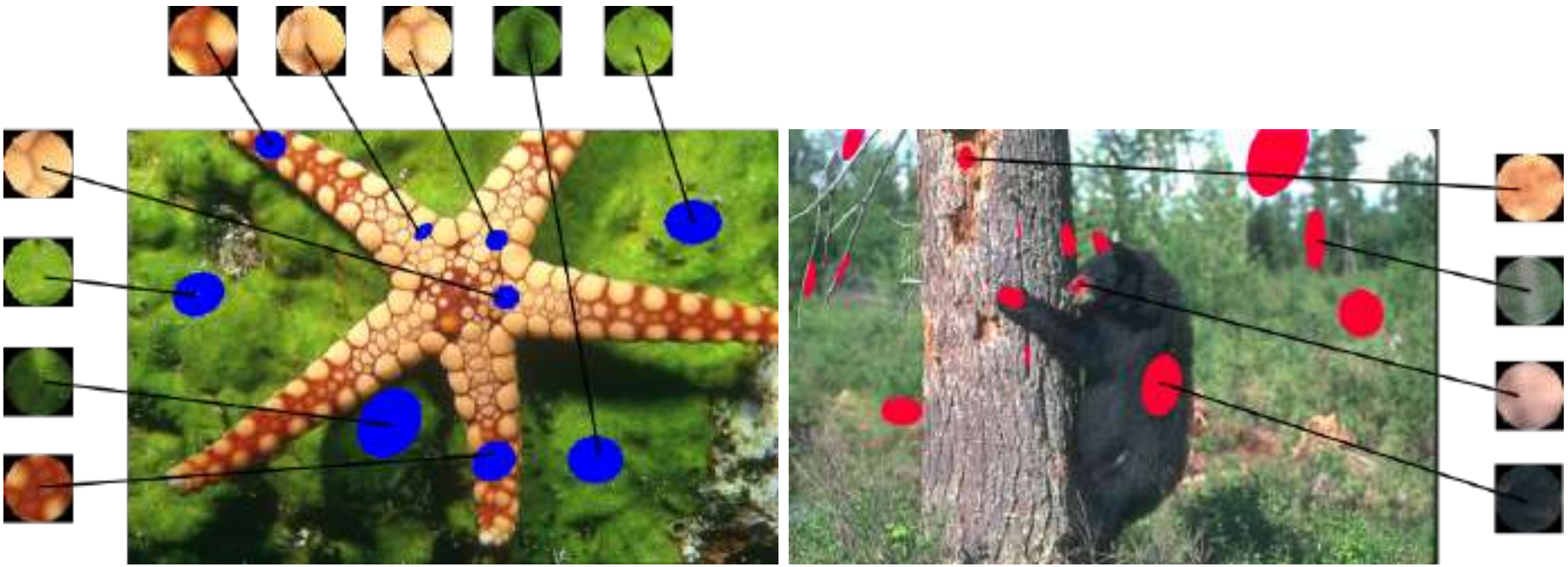


Image source: Maria Oliver

Image segmentation

Proposal [1]: to simplify the set of all patches $\mathcal{P}_u$ by estimating an optimal finite set of representative patches $\{\mathbf{p}_{\Omega_1}, \ldots, \mathbf{p}_{\Omega_N}\}$, where $\Omega = \cup_{i=1}^{N} \bar{\Omega}_i$ is a partition of the image domain into N disjoint open regions Ω_i, **containing the pixels with similar patches**, for $N \in \mathbb{N}$, with the energy:

$$E(\mathbf{p}, \chi) = \underbrace{\sum_{i=1}^{N} \int_{\Omega} |\nabla \chi_{\Omega_i}(x)| \, \mathrm{d}x}_{\text{Regularity term (boundary length)}} + \lambda \underbrace{\sum_{i=1}^{N} \int_{\Omega} \mathcal{D}_t^{\mathrm{a},1}\left(\mathbf{p}(x), \mathbf{p}_u(x)\right) \chi_{\Omega_i}(x) \mathrm{d}x}_{\text{Data term}}, \quad \lambda \geq 0$$

where $\mathbf{p} = \sum_i \mathbf{p}_{\Omega_i} \chi_{\Omega_i}$ is a piecewise constant patch function, $\chi_{\Omega_i} \in BV(\Omega; \{0,1\})$ the characteristic function of Ω_i, $\chi = (\chi_{\Omega_1}, \ldots, \chi_{\Omega_N})$ such that $\sum_i \chi_i(x) = 1, \forall x \in \Omega$, and

$$\mathcal{D}_t^{\mathrm{a},1}\left(\mathbf{p}_u(x), \mathbf{p}_u(y)\right) = \int_{\Delta_t} \eta_t(h) \left\| u(x + T_u(x)^{-\frac{1}{2}} h) - u(y + T_u(y)^{-\frac{1}{2}} h) \right\|_{L^1} \mathrm{d}h.$$

---

[1] M. Oliver, G. Haro, V. Fedorov and C. Ballester. *L1 Patch-based Image Partitioning into Homogeneous Textured Regions.* In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing. 2018.

Image segmentation

Proposal [1]: to simplify the set of all patches $\mathcal{P}_u$ by estimating an optimal finite set of representative patches $\{\mathbf{p}_{\Omega_1}, \ldots, \mathbf{p}_{\Omega_N}\}$, where $\Omega = \cup_{i=1}^{N} \bar{\Omega}_i$ is a partition of the image domain into N disjoint open regions Ω_i, **containing the pixels with similar patches**, for $N \in \mathbb{N}$, with the energy:

$$E(\mathbf{p}, \chi) = \underbrace{\sum_{i=1}^{N} \int_{\Omega} |\nabla \chi_{\Omega_i}(x)| \, \mathrm{d}x}_{\text{Regularity term (boundary length)}} + \lambda \underbrace{\sum_{i=1}^{N} \int_{\Omega} \mathcal{D}_t^{\mathrm{a},1}\left(\mathbf{p}(x), \mathbf{p}_u(x)\right) \chi_{\Omega_i}(x) \mathrm{d}x}_{\text{Data term}}, \quad \lambda \geq 0$$

where $\mathbf{p} = \sum_i \mathbf{p}_{\Omega_i} \chi_{\Omega_i}$ is a piecewise constant patch function, $\chi_{\Omega_i} \in BV(\Omega; \{0,1\})$ the characteristic function of Ω_i, $\chi = (\chi_{\Omega_1}, \ldots, \chi_{\Omega_N})$ such that $\sum_i \chi_i(x) = 1, \forall x \in \Omega$, and by analogy and by an abuse of notation we have denoted by $\mathcal{D}_t^{\mathrm{a},1}\left(\mathbf{p}(x), \mathbf{p}_u(x)\right)$ the patch similarity

$$\int_{\Delta_t} \eta_t(h) \left\| \mathbf{p}_{\Omega_i}(h) - u(x + T_u(x)^{-\frac{1}{2}} h) \right\|_{L^1} \mathrm{d}h,$$

for $x \in \Omega_i$.

---

[1] M. Oliver, G. Haro, M. Fedorov and C. Ballester. *L1 Patch-based Image Partitioning into Homogeneous Textured Regions*. In

Image segmentation

Relaxation: relax the characteristic functions to be fuzzy membership functions belonging to the set $\mathcal{C} = \Big\{(\omega_1, \ldots, \omega_N) \,|\, \omega_i \in BV(\Omega; [0,1]),\ 0 \leq \omega_i(x) \leq 1, \sum_{i=1}^{N} \omega_i(x) = 1, \forall x \in \Omega\Big\}$. $\omega_i(x)$ describes the fuzzy membership of a pixel x and can be understood as the probability that x belongs to the region Ω_i[2].

Let ω be $\omega = (\omega_1, \ldots, \omega_N)$ which is often denoted as an N-phase fuzzy membership function.

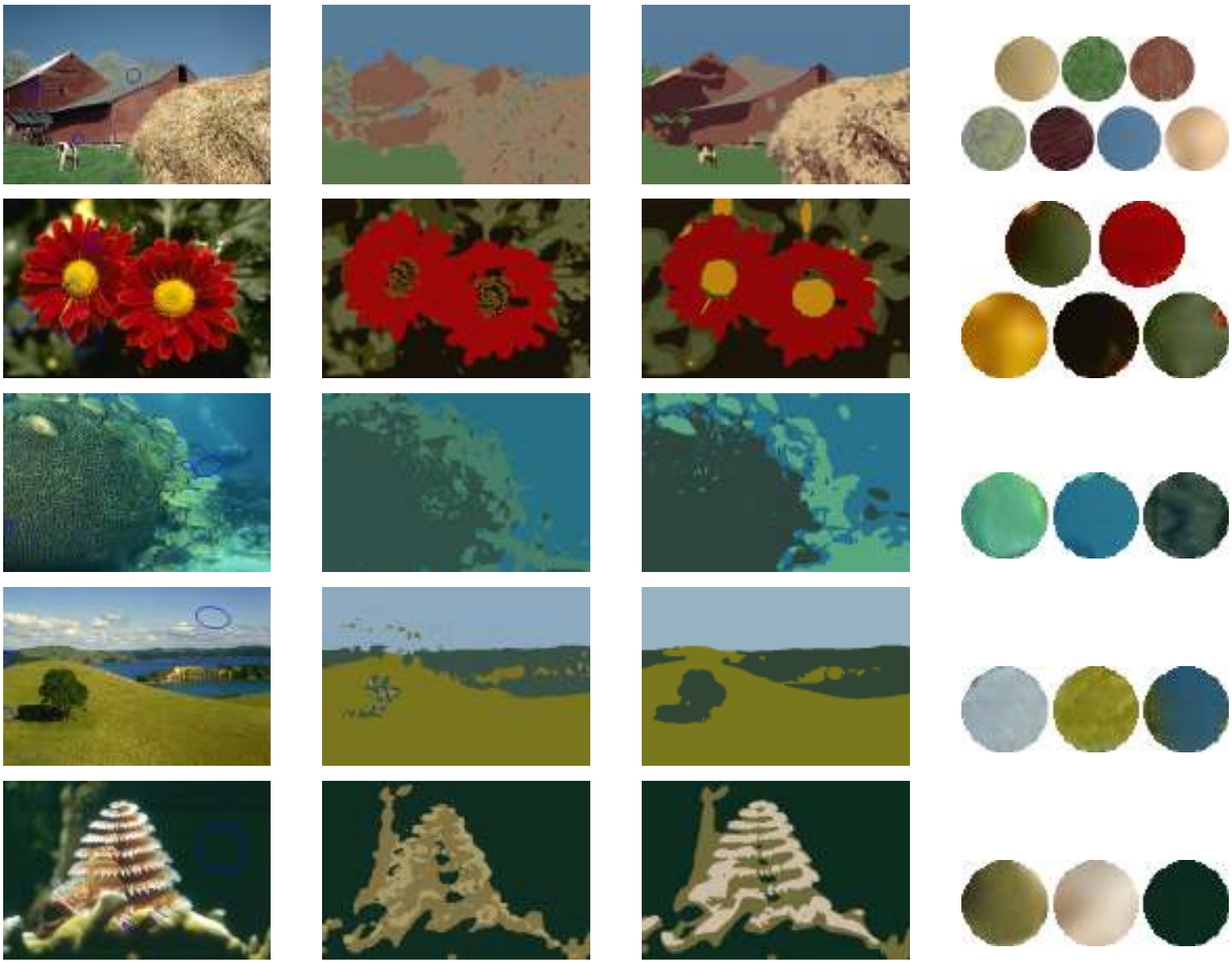
In this framework, the previous energy writes

$$\min_{(\mathbf{p},\omega)\in L^1(\Omega;L^1(\Delta_t))\times\mathcal{C}} \bar{E}(\mathbf{p},\omega) = \underbrace{\sum_{i=1}^{N} \int_\Omega |\nabla \omega_i(x)| \mathrm{d}x}_{E_s(\omega)} + \lambda \underbrace{\sum_{i=1}^{N} \int_\Omega \mathcal{D}_t^{\mathsf{a},1}\left(\mathbf{p}(x), \mathbf{p}_u(x)\right) \omega_i(x) \mathrm{d}x}_{E_d(\mathbf{p},\omega)}.$$

---

[2] F. Li, S. Osher, J. Qin, and M. Yan. *A multiphase image segmentation based on fuzzy membership functions and l1-norm fidelity*. J Sci Comput, 2016.

Input & patch	Li. et. al [**Li**]	Ours	Discs

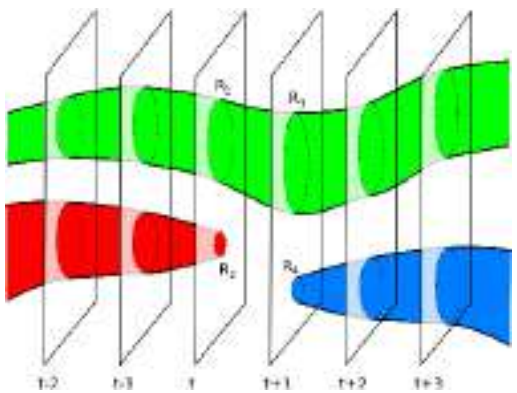


Input & patch	Li. et. al [**Li**]	Ours	Discs

4. Video simplification/segmentation, spatio-temporal tubes for video analysis

Spatio-temporal tube computation in video

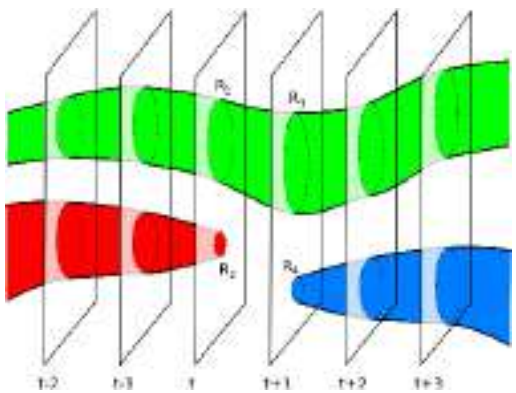
Spatio-temporal tube computation is an over-segmentation technique that group together into spatio-temporal regions pixels following a similarity criterion. The tube aims to represent the trajectory in the moving scene of each object.



Spatio-temporal tube computation in video

Spatio-temporal tube computation is an over-segmentation technique that group together into spatio-temporal regions pixels following a similarity criterion. The tube aims to represent the trajectory in the moving scene of each object.

Problem: Changes of position, illumination, and the interaction of the objects with the surrounding makes the problem even more challenging.

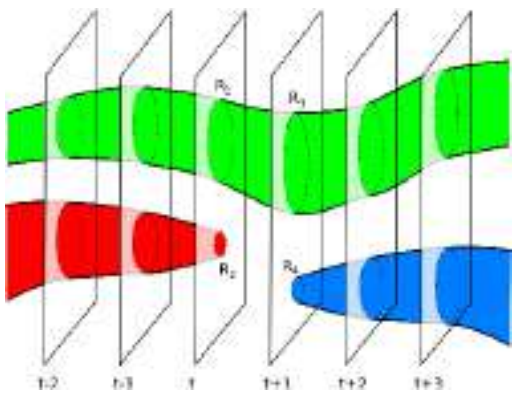


Spatio-temporal tube computation in video

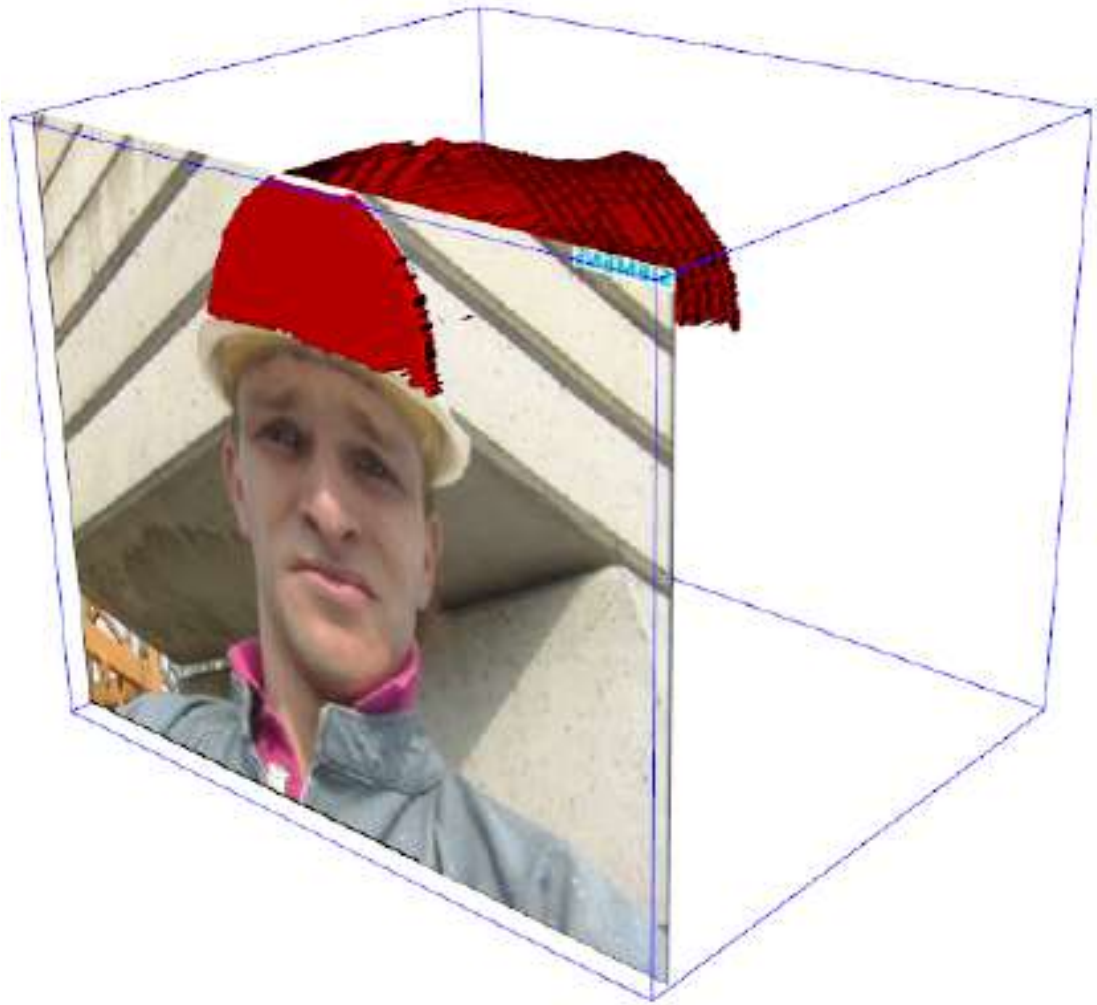
Spatio-temporal tube computation is an over-segmentation technique that group together into spatio-temporal regions pixels following a similarity criterion. The tube aims to represent the trajectory in the moving scene of each object.

Problem: Changes of position, illumination, and the interaction of the objects with the surrounding makes the problem even more challenging.

Solution Affine invariant 3D patch similarity measure.



Video segmentation, spatio-temporal tubes



Example of Spatio-temporal tube (following the helmet)

P. Vitoria, V. Fedorov and C. Ballester. *Spatio-temporal tube segmentation through a video metrics-based patch similarity measure.* IMVIP 2017.

Video segmentation, spatio-temporal tubes

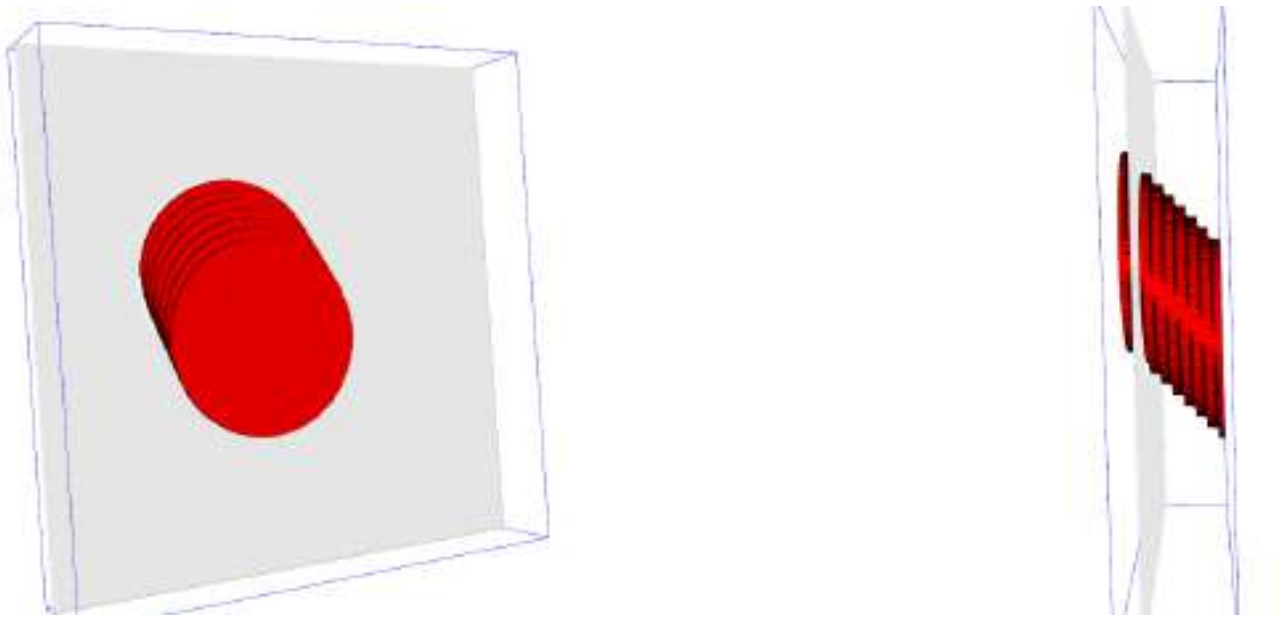


Figure : Synthetic sequences. A ball falling down: Tube computation from a white point inside the ball (seen as a disc on each temporal frame).

Video segmentation, spatio-temporal tubes

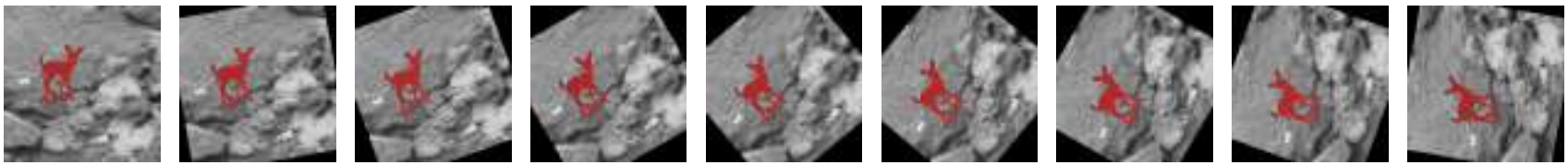


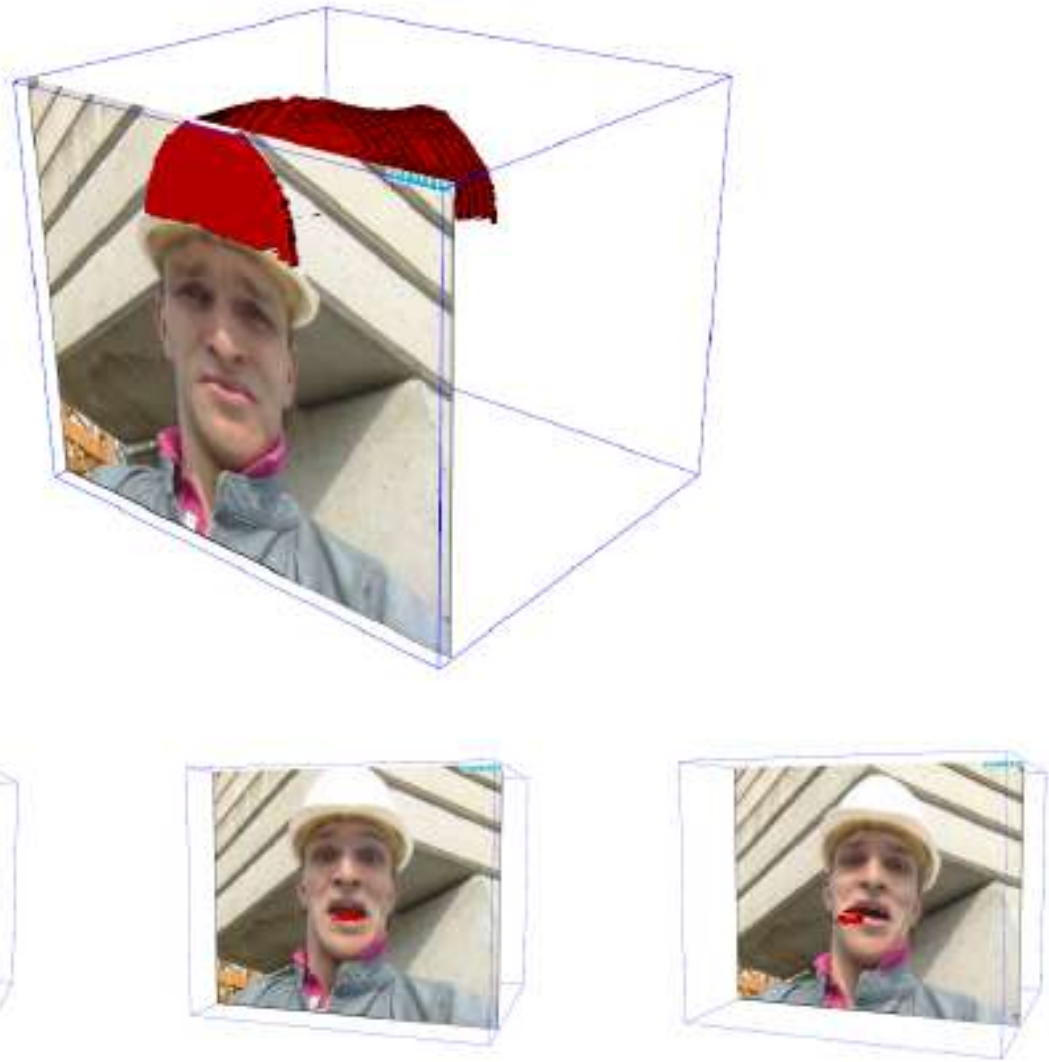
Figure : Synthetic sequences. Tube computation from a point inside of a deer that rotates each frame by 10 degrees counterclockwise.

Video segmentation, spatio-temporal tubes

Figure : Real sequences. Tube of the right ball.

Video segmentation, spatio-temporal tubes

Figure : Real sequences. Tube computation from a white point inside of the "WC" sign. From left to right: Frame 4, 10 and side view. Notice, that when the sign is occluded by the black T-shirt's man, the tube also gets reduced. The tube groups together all the pixels with similar texture, for that reason the letter "WC" remains outside of the region/tube.



Spatio-temporal tubes following the helmet (first image) and the tongue of the man (second to fourth image). The tongue is occluded in the first frame (second image) and disoccluded in the fourth frame (third image) and again disappear in the tenth frame (fourth image) when only the background of the mouth is visible.

6. Motion inpainting by an image based geodesic AMLE method

Motion inpainting

Optical flow or motion inpainting is a pervasive problem:

- semantic video analysis,
- video editing,
- to optical flow estimation in occlusion and disocclusion regions,
- video inpainting; cinema post-production (elimination of unwanted moving objects),
- automatic assistance of sensor-based optical flow estimation (e.g., in Kitti Vision benchmark)

Motion inpainting

Optical flow or motion inpainting is a pervasive problem:

- semantic video analysis,
- video editing,
- to optical flow estimation in occlusion and disocclusion regions,
- video inpainting; cinema post-production (elimination of unwanted moving objects),
- automatic assistance of sensor-based optical flow estimation (e.g., in Kitti Vision benchmark)

Goal: Given a video and an incomplete motion field, an automatic method to complete it.

Motion inpainting

Optical flow or motion inpainting is a pervasive problem:

- semantic video analysis,
- video editing,
- to optical flow estimation in occlusion and disocclusion regions,
- video inpainting; cinema post-production (elimination of unwanted moving objects),
- automatic assistance of sensor-based optical flow estimation (e.g., in Kitti Vision benchmark)

Goal: Given a video and an incomplete motion field, an automatic method to complete it.

How: Each 2D frame domain is endowed with a Riemannian metric based on the video values and the missing optical flow are recovered by solving the Absolutely Minimizing Lipschitz Extension (AMLE) partial differential equation on the 2D Riemannian manifold from the known values on the boundary of the interpolation domain[1,2].

[1] V. Lazcano, PhD thesis *Some Problems in Depth Enhanced Video Processing*. Universitat Pompeu Fabra. February 2016.

[2] M. Oliver, L. Raad, C. Ballester and G. Haro. *Motion inpainting by an image-based geodesic amle method*. ICIP 2018.

Motion inpainting

Let $I(\mathbf{x}, t)$ be a video defined on $\Omega \times \{1, \dots T\}$, where $\Omega \subset \mathbb{R}^2$ denotes the image frame domain and $\{1, \dots T\}$ is the set of discrete times

Let $\mathbf{v} = (v_1, v_2)$ be the optical flow of the video I representing the apparent motion between a pixel $\mathbf{x} \in \Omega \setminus \Omega_0(t)$ at time t and the corresponding at time $t + 1$.

We assume that, at time t, $\mathbf{v}(\mathbf{x}, t)$ is unknown on a region $\Omega_0(t) \subset \Omega$ whose boundary, denoted by $\partial\Omega_0$, consists of a finite union of smooth curves and possibly isolated points.

We endow Ω, at each time t, with a metric $g(t)$. Let $\mathcal{M}(t) = (\Omega, g(t))$ be the corresponding Riemannian manifold.

We propose to **complete $\mathbf{v}$ in $\Omega_0(t)$ with the motion field $(\mathbf{u_1}, \mathbf{u_2})$** such that u_1 and u_2 are solutions, respectively, of the geodesic AMLE, given by the PDE

$$\boldsymbol{\Delta_{\infty,g}\mathbf{u} = 0 \quad \text{in } \Omega_0(t) \quad \text{s.t} \quad \mathbf{u}|_{\partial\Omega_0(t)} = \mathbf{v_i},}$$

for $i = 1, 2$, respectively. We also use Newmann boundary conditions on $\partial\Omega$.

Motion inpainting

Here we have denoted

$$\Delta_{\infty,g} u := D^2_{\mathcal{M}} u \left(\frac{\nabla_{\mathcal{M}} u}{|\nabla_{\mathcal{M}} u|}, \frac{\nabla_{\mathcal{M}} u}{|\nabla_{\mathcal{M}} u|} \right) \tag{1}$$

where $\nabla_{\mathcal{M}} u$ and $D^2_{\mathcal{M}} u$ denote, respectively, the gradient and the Hessian on the manifold. To simplify, we have omitted the dependence on t of g and $\mathcal{M}$.

When g is the Euclidean metric, the operator in (1) is known as the infinity Laplacian.

The metric g can be affine covariant structure tensors or, e.g., taking into account spatial distances and photometric similarities.

Motion inpainting

Frame 16

Ground truth optical flow with the inpainting region indicated with a box

Ours[2] (EPE: 1.5582)

EpicFlow[1] (EPE: 2.8584)

Some optical flow inpainting results for a frame of the temple_3 sequence of MPI-Sintel benchmark.

---

[1] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid. *Epicflow: Edge-preserving interpolation of correspondences for optical flow*. CVPR 2015.

[2] M. Oliver, L. Raad, C. Ballester and G. Haro. *Motion inpainting by an image-based geodesic amle method*. ICIP 2018.

The geodesic AMLE on a finite graph

We solve the AMLE equation on the manifold with an efficient numerical algorithm which is based on the eikonal operators for nonlinear elliptic PDEs on a finite graph [Oberman'2005], [Manfredi et al'2015].

In particular, we consider the discrete grid of Ω as the nodes of a finite weighted graph G. If $\mathbf{x}$ and $\mathbf{y}$ are neighbouring pixels, its distance is denoted by $d(\mathbf{x}, \mathbf{y})$. We test also

$$d_1(\mathbf{x}, \mathbf{y}) = \sqrt{(1-\lambda)\|I(\mathbf{x}, t) - I(\mathbf{y}, t)\|^2 + \lambda\|\mathbf{x} - \mathbf{y}\|^2}$$
$$d_2(\mathbf{x}, \mathbf{y}) = (1-\lambda)\|I(\mathbf{x}, t) - I(\mathbf{y}, t)\| + \lambda\|\mathbf{x} - \mathbf{y}\|$$

where $\lambda \in [0, 1]$. We also include

$$d_3(\mathbf{x}, \mathbf{y}) = (1-\lambda)\|I(\mathbf{x}, t) - I(\mathbf{y}, t)\|^2 + \lambda\|\mathbf{x} - \mathbf{y}\|^2$$

which is a semimetric.

Given a path $\gamma = \{\gamma(i)\}_{i=0}^{m}$ on the graph G joining two points, $\mathbf{x} = \gamma(0)$ and $\mathbf{y} = \gamma(m)$, its length is defined as usual by $L^g(\gamma) = \sum_{i=0}^{m-1} d(\gamma(i), \gamma(i+1))$. Given any two points $\mathbf{x}$ and $\mathbf{y}$ on the grid, then the geodesic distance $d^g(\mathbf{x}, \mathbf{y})$ is

$$d^g(\mathbf{x}, \mathbf{y}) = \inf\{L^g(\gamma) : \gamma \text{ is a curve joining } \mathbf{x} \text{ and } \mathbf{y}\}.$$

This distance can be computed using Dijkstra's algorithm.

The geodesic AMLE on a finite graph

Given a point $\mathbf{x}$ on the grid, let $\mathcal{N}(\mathbf{x})$ be a neighborhood of $\mathbf{x}$. The **positive and negative eikonal operators** [Oberman'2005], [Manfredi et al'2015] are

$$\|\nabla u(\mathbf{x})\|^+_{\mathcal{M}} = \max_{\mathbf{y}\in\mathcal{N}(\mathbf{x})} \frac{u(\mathbf{y})-u(\mathbf{x})}{d^g(\mathbf{x},\mathbf{y})},$$

$$\|\nabla u(\mathbf{x})\|^-_{\mathcal{M}} = \min_{\mathbf{z}\in\mathcal{N}(\mathbf{x})} \frac{u(\mathbf{z})-u(\mathbf{x})}{d^g(\mathbf{x},\mathbf{z})}.$$

The discrete infinity Laplacian corresponds to

$$\Delta_{\infty,g} u(\mathbf{x}) = \frac{\|\nabla u(\mathbf{x})\|^+_{\mathcal{M}} + \|\nabla u(\mathbf{x})\|^-_{\mathcal{M}}}{2}.$$

We solve $\Delta_{\infty,g} u = 0 \quad \text{in } \Omega_0(t) \quad \text{s.t} \quad u|_{\partial\Omega_0(t)} = v_i$ with the iterative discrete scheme

$$u^{k+1}(\mathbf{x}) = \frac{d^g(\mathbf{x},\mathbf{z})u^k(\mathbf{y}) + d^g(\mathbf{x},\mathbf{y})u^k(\mathbf{z})}{d^g(\mathbf{x},\mathbf{z}) + d^g(\mathbf{x},\mathbf{y})}$$

where $\mathbf{y}$ and $\mathbf{z}$ are the pixels providing the maximum and minimun of the eikonals. This scheme is applied for $\mathbf{x} \in \Omega_0(t)$, initializing $u^0(\mathbf{x}) = 0$ in that case and keeping the values of $v_1(\mathbf{x})$, respectively $v_2(\mathbf{x})$, on the known region $\Omega \setminus \Omega_0$ for all k.

The geodesic AMLE on a finite graph

Given a point $\mathbf{x}$ on the grid, let $\mathcal{N}(\mathbf{x})$ be a neighborhood of $\mathbf{x}$. The **positive and negative eikonal operators** [Oberman'2005], [Manfredi et al'2015] are

$$\|\nabla u(\mathbf{x})\|^+_{\mathcal{M}} = \max_{\mathbf{y}\in\mathcal{N}(\mathbf{x})} \frac{u(\mathbf{y})-u(\mathbf{x})}{d^g(\mathbf{x},\mathbf{y})},$$

$$\|\nabla u(\mathbf{x})\|^-_{\mathcal{M}} = \min_{\mathbf{z}\in\mathcal{N}(\mathbf{x})} \frac{u(\mathbf{z})-u(\mathbf{x})}{d^g(\mathbf{x},\mathbf{z})}.$$

The discrete infinity Laplacian corresponds to

$$\Delta_{\infty,g} u(\mathbf{x}) = \frac{\|\nabla u(\mathbf{x})\|^+_{\mathcal{M}} + \|\nabla u(\mathbf{x})\|^-_{\mathcal{M}}}{2}.$$

We solve $\Delta_{\infty,g} u = 0 \quad \text{in } \Omega_0(t) \quad \text{s.t} \quad u|_{\partial\Omega_0(t)} = v_i$ with the iterative discrete scheme

$$u^{k+1}(\mathbf{x}) = \frac{d^g(\mathbf{x},\mathbf{z})u^k(\mathbf{y}) + d^g(\mathbf{x},\mathbf{y})u^k(\mathbf{z})}{d^g(\mathbf{x},\mathbf{z}) + d^g(\mathbf{x},\mathbf{y})} \quad \text{(convergent scheme [Oberman'2005])}$$

where $\mathbf{y}$ and $\mathbf{z}$ are the pixels providing the maximum and minimun of the eikonals. This scheme is applied for $\mathbf{x} \in \Omega_0(t)$, initializing $u^0(\mathbf{x}) = 0$ in that case and keeping the values of $v_1(\mathbf{x})$, respectively $v_2(\mathbf{x})$, on the known region $\Omega \setminus \Omega_0$ for all k.

Appl. to optical flow inpainting in occlusion areas

video frame

inpainting mask

ground truth

$d_1(\mathbf{x}, \mathbf{y})$

$d_2(\mathbf{x}, \mathbf{y})$

$d_3(\mathbf{x}, \mathbf{y})$

Figure : Comparison of three tested possibilities for the metric illustrated in an experiment of completion of the optical flow in the occlusion areas (white regions in image b).

(2 appl) inpainting in larger holes and sparse-to-dense optical flow estimation

Figure : Comparison of different motion completion algorithms in two different

Optical flow inpainting in larger holes

Frame 02 seq. 10

Groundtruth

Inpainting

Figure : A result on Kitti dataset that contains large holes.

SINTEL	Ours - EPE	EF - EPE
occlusions	**5.4198**	6.8797
sparse 1%	**0.7061**	1.8532
sparse 5%	**0.4340**	1.4199
sparse 30%	**0.2241**	1.1212
sparse DM	4.4404	**4.1507**
sparse DM (gt)	**2.1360**	2.3802
hole	**1.7208**	1.9587

Table : Comparison of the EPE for EpicFlow and our method in different situations. The comparison is done over all the training Sintel dataset.

MIDDLEBURY	Ours - EPE	EF - EPE
sparse 1%	**0.1979**	0.3105
sparse 5%	**0.1053**	0.2426
sparse 30%	**0.0567**	0.1801
sparse DM	0.9216	**0.8112**
sparse DM (gt)	**0.2049**	0.2789

Table : Idem. The comparison is done in the Middlebury dataset on the optical flow corresponding to frames 10 and 11, where the groundtruth is available.

The experiments show that in general our results outperform those of EpicFlow which has become a reference method for optical flow estimation and a standard technique for post-processing an estimated and filtered optical flow.

5. Depth completion by a geodesic Biased AMLE method

Depth completion by a geodesic Biased AMLE method

We apply it to the context of depth estimation in images or videos where large regions of incomplete depth data often appear due to unreliable data, occlusions, or adquisition failure.

We endow the image or video domain with an anisotropic metric g and interpolate the missing information by solving the **Biased Absolutely Minimizing Lipschitz Extension (bAMLE)** in the manifold, that is,

$$D^2_{\mathcal{M}} u \left(\frac{\nabla u}{|\nabla u|_\xi}, \frac{\nabla u}{|\nabla u|_\xi} \right) + \beta |\nabla u|_\xi = 0,$$

or

$$\Delta_{\infty,g} u + \beta \, |\nabla u|_\xi = 0.$$

where $\beta \geq 0$ and $\xi \in \mathcal{M}$. The bAMLE ($\beta > 0$) favours the extension of data to large regions and it is an exponential cone interpolator which offers an interesting alternative since the value at points expands and produces a smoother profile, which makes it more adapted to depth interpolation in real scenes.

---

[1] V. Lazcano, PhD thesis *Some Problems in Depth Enhanced Video Processing.* Universitat Pompeu Fabra. February 2016.

[2] V. Lazcano, F. Calderero, and C. Ballester. *Interpolation in Manifolds and Applications to Depth Interpolation in Images and Videos.* In ICM 2018.

The geodesic Biased AMLE on a finite graph

Given a point $\mathbf{x}$ on the grid, let $\mathcal{N}(\mathbf{x})$ be a neighborhood of $\mathbf{x}$.
Using the **positive and negative eikonal operators** [Oberman'2005], [Manfredi et al'2015], the **discrete version of the Biased AMLE equation** is

$$\frac{\|\nabla u(x)\|_x^+ + \|\nabla u(x)\|_x^-}{2} + \beta \left| \|\nabla u(x)\|_x^+ \right| = 0$$

with $\beta > 0$. It can be rewritten

$$\beta_+ \|\nabla u(x)\|_x^+ + \beta_- \|\nabla u(x)\|_x^- = 0, \qquad (2)$$

where $\beta_+ > \beta_-$ (actually, $\beta_+ = \frac{1}{2} + \beta \text{sign}(\|\nabla u(x)\|_x^+)$ and $\beta_- = \frac{1}{2}$). Notice that, if $\beta_+ = \beta_-$, it results in a multiple of the infinity Laplacian.

By introducing these expressions into (2) we obtain the numerical scheme for the discrete biased AMLE

$$u^{k+1}(x) = \frac{\beta_+ d_{xz} u^k(y) + \beta_- d_{xy} u^k(z)}{\beta_+ d_{xz} + \beta_- d_{xy}}, \qquad k = 0, 1, ...$$

where $\mathbf{y}$ and $\mathbf{z}$ are the pixels providing the maximum and minimun of the eikonals.

Depth completion (image case)

Figure : (a) Color reference image. (b) Initial depth data obtained by a depth sensor. (c) Result obtained by the biased AMLE.

Thank you for your attention

http://gpi.upf.edu/
email: coloma.ballester@upf.edu