Generalized persistent homologies: G-invariant and multi-dimensional persistence 1 – Introduction and motivation

Massimo Ferri^{1,3} Patrizio Frosini^{1,3} Claudia Landi^{2,3}

¹Dipartimento di Matematica, Università di Bologna

²Dipartimento di Scienze e Metodi dell'Ingegneria, Università di Modena e Reggio Emilia

³ARCES, università di Bologna

massimo.ferri@unibo.it patrizio.frosini@unibo.it clandi@unimore.it

Homology: Theoretical and Computational Aspects, 9-13 February 2015

Outline of the course



- Introduction and motivation (Massimo Ferri)
- G-invariant persistent homology (Patrizio Fròsini)
- Multidimensional persistent homology (Claudia Landi)

Outline

- Shape?
- Main definitions
- Applications:
 - Face profiles
 - Hand-written letters and monograms
 - Sign alphabet
 - Tropical cyclones
 - Lithography in microelectronics
 - 3D models
 - Echocardiography
 - Leukocytes
 - Oncology:
 - Mouth cells
 - Melanocytic lesions
 - Hepatic cells

Outline

- Shape?
- Main definitions
- Applications:
 - Face profiles
 - Hand-written letters and monograms
 - Sign alphabet
 - Tropical cyclones
 - Lithography in microelectronics
 - 3D models
 - Echocardiography
 - Leukocytes
 - Oncology:
 - Mouth cells
 - Melanocytic lesions
 - Hepatic cells



"Quid est ergo tempus? si nemo ex me quaerat, scio; si quaerenti explicare velim, nescio"

"What is then time? when nobody asks me, I do know, but when asked to explain it, I don't know." (Augustine of Hippo)

Much the same could be said of "shape"!

A tentative way out: invariance under a given Transformation Group

















 $\begin{cases} x' = \cos \alpha x - \sin \alpha y + e \\ y' = \sin \alpha x + \cos \alpha y + f \end{cases}$



movement(direct congruence): det>0



 $\begin{cases} x' = \cos \alpha x - \sin \alpha y + e \\ y' = \sin \alpha x + \cos \alpha y + f \end{cases}$



congruence: nonvanishing det









affinity

 $ad - bc \neq 0$





$$\begin{cases} x' = ax + by + et \\ y' = cx + dy + ft \\ t' = gx + hy + kt \end{cases}$$



homography

 $adk_{_{13/80}} +$





homeomorphism





Homeomorphisms are important in everyday shape understanding







But here there is a homeomorphism too!



Linear invariance (e.g. affine, projective...) is too rigid.

Topological invariance (w.r.t. homeomorphisms) is too free.

On one hand, Algebraic Topology offers very powerful tools which formalize qualitative aspects of spaces. On the other hand, its "descriptors" are invariant under homeomorphisms, even under homotopy equivalences!



Should we then leave all hope of using mathematics in the description, analysis, synthesis, comparison, classification of shape?

NOT AT ALL!



An idea: data can be seen as functions (e.g. a picture is a function from a rectangle to \mathbb{R}^3), more precisely pairs (*X*,*f*) (called *size pairs*) of a topological space *X* and a function *f* defined on it.

Rather than defining *shape* itself, we may think of defining the concept of pairs (X,f) and (X,f') having the same shape within a certain group G of autohomeomorphisms of X if there is an element of G which composed with f yields f'.



We often probe (X,f) by applying an operator F (or several of them) which takes (X,f) to a pair (Y,g) – where Y may or may not coincide with X – selecting a particular feature of (X,f) depending on the context or on the application and representing the observer's viewpoint.

This is the idea behind all the applications that I am going to show.

The theoretical setting will be discussed by Patrizio Frosini.



(X,f)F(Y,g)(X = square, f = colour) $(Y = S^1, g = distance from centre)$

The function (the original *f* or the auxiliary *g*) yields a filtration of subspaces, on which we can use Algebraic Topology.

Outline

- Shape?
- Main definitions
- Applications:
 - Face profiles
 - Hand-written letters and monograms
 - Sign alphabet
 - Tropical cyclones
 - Lithography in microelectronics
 - 3D models
 - Echocardiography
 - Leukocytes
 - Oncology:
 - Mouth cells
 - Melanocytic lesions
 - Hepatic cells



A possible setting: persistent topology of a size pair $(X, \vec{\varphi})$

X is a topological space, $\overrightarrow{\varphi} : X \to \mathbb{R}^n$ a continuous map, called *measuring (filtering) function*.

- $\vec{\varphi}$ provides the geometrical aspects;
- the choice of $\overrightarrow{\phi}$ conveys the subjective viewpoint of the observer.

Examples of measuring functions

- 1-dimensional: distance from center of mass, ordinate, curvature,
- multidimensional: color, coordinates, curvature and torsion, ...

. . .

Main definitions





How much do these two curves differ w.r.t. ordinate as a filtering function?

Our proposal: the minimum cost, in terms of f and f', of transforming one in the other.

24/80



Definition

The natural pseudo-distance between the size pairs (M, φ) and (N, ψ) is

$$\delta\left((\boldsymbol{M},\varphi),(\boldsymbol{N},\psi)\right) = \begin{cases} \inf_{\boldsymbol{h}\in H(\boldsymbol{M},\boldsymbol{N})} \max_{\boldsymbol{P}\in\boldsymbol{M}} \|\varphi(\boldsymbol{P}) - \psi(\boldsymbol{h}(\boldsymbol{P}))\|_{\infty}, \\ +\infty \quad \text{if } \boldsymbol{H}(\boldsymbol{M},\boldsymbol{N}) = \emptyset, \end{cases}$$

H(M, N) being the set of all the homeomorphisms between M and N.

P. Frosini, M. Mulazzani, Size homotopy groups for computation of natural size distances, Bull. of the Belgian Math. Soc. - Simon Stevin, 6 (1999), 455-464.

25/80



But the natural pseudodistance is difficult (or even impossible) to compute.

Therefore we need a computable lower bound for it.

Luckily, we have it: the *matching distance* (or *bottleneck distance*) between *Persistent Betti Number* functions of the size pairs.



We define the following relation $\leq (\prec)$ in \mathbb{R}^n : if $\vec{u} = (u_1, \ldots, u_n)$ and $\vec{v} = (v_1, \ldots, v_n)$, we write $\vec{u} \leq \vec{v}$ ($\vec{u} \prec \vec{v}$) if and only if $u_j \leq v_j$ $(u_j < v_j)$ for $j = 1, \ldots, n$. Let also Δ^+ be the set $\{(\vec{u}, \vec{v}) \in \mathbb{R}^n \times \mathbb{R}^n \mid \vec{u} \prec \vec{v}\}.$

We denote by $X\langle \vec{f} \leq \vec{u} \rangle$ the *lower level set* $\{p \in X \mid \overrightarrow{f}(p) \leq \overrightarrow{u} \}$.





For each $i \in \mathbb{Z}$, the *i*-th *Persistent Betti Number (PBN) function* of (X, \vec{f}) is $\rho_{(X, \vec{f}, i)} : \Delta^+ \to \mathbb{N}$ defined as $\rho_{(X, \vec{f}, i)}(\vec{u}, \vec{v}) = \dim(Imf_{\vec{u}}^{\vec{v}}), \ \vec{u} \prec \vec{v}$ with $f_{\vec{u}}^{\vec{v}} : H_i(X\langle \vec{f} \preceq \vec{u} \rangle) \to H_i(X\langle \vec{f} \preceq \vec{v} \rangle),$

where $f_{\vec{u}}^{\vec{v}}$ is the homomorphism induced by the inclusion map of lower level sets $X\langle \vec{f} \leq \vec{u} \rangle \subseteq X\langle \vec{f} \leq \vec{v} \rangle$

H. Edelsbrunner, D. Letscher and A. Zomorodian, *Topological Persistence and Simplification*, Proc. 41st Ann. IEEE Sympos. Found Comput. Sci. (2000), 454–463.

G. Carlsson and A. Zomorodian, *The Theory of Multidimensional Persistence*, Symposium on Computational Geometry, June 6–8, 2007, Gyeongiu, South Korea (2007) 184–193.

Main definitions



An easy example of 0-degree Persistent Betti Number function (also called *size function*) with $f : M \to \mathbb{R}$:



P. Frosini *Measuring shapes by size functions*, Proc. of SPIE, Intelligent Robots and Computer Vision X: Algorithms and Techniques, Boston, MA 1607 (1991).

29/80

Main definitions



All information carried by a size function can be condensed in the formal series of its *cornerpoints*



The *matching distance*

P. Frosini, C. Landi, Size functions and formal series, Appl. Algebra Engrg. Comm. Comput. 12 (2001), 327-349.



It turns out that:

$$d_{match}(\ell_{(\mathcal{M},\varphi)},\ell_{(\mathcal{N},\psi)}) \leq d((\mathcal{M},\varphi),(\mathcal{N},\psi))$$

i.e. the matching distance between size functions yields a lower bound to the natural pseudodistance.

S. Biasotti, A. Cerri, P. Frosini, D. Giorgi, C. Landi *Multidimensional size functions for shape comparison* Journal of Mathematical Imaging and Vision 32 (2008), 161–179.

A. Cerri, B. Di Fabio, M. Ferri, P. Frosini, C. Landi *Betti numbers in multidimensional persistent homology are stable functions* Math. Meth. Appl. Sci. DOI: 10.1002/mma.2704 (2012).



The present evolution of Persistent Homology is in the computation of Persistent Betti Numbers when the filtering function has a multidimensional range. But

- There is no convenient information carrier like cornerpoints of the 1D case
- PBN's don't determine the Persistence Modules.

See the contribution of Claudia Landi in this course.

F. Cagliari, B. Di Fabio, and M. Ferri, *One-dimensional reduction of multidimensional persistent homology*, Proc. Amer. Math. Soc., 138(8) (2010), 3003–3017.

F. Cagliari, M. Ferri, L. Gualandri, C. Landi: *Persistence Modules, Shape Description, and Completeness*, Proc. CTIC 2012, Bertinoro, Italy, LNCS 7309, M. Ferri et al. eds, LNCS, vol. 7309 (2012), 148-156.



Invariance under particular subgroups of the group of homeomorphisms of a space CAN be taken into account: see the contribution of Patrizio Frosini in this course.

Outline

- Shape?
- Main definitions
- Applications:
 - Face profiles
 - Hand-written letters and monograms
 - Sign alphabet
 - Tropical cyclones
 - Lithography in microelectronics
 - 3D models
 - Echocardiography
 - Leukocytes
 - Oncology:
 - Mouth cells
 - Melanocytic lesions
 - Hepatic cells



Next, I will show applications - some recent, some very old - to concrete problems of classification or of retrieval.

All are based on a suitable choice of the auxiliary filtering function(s). Most could be enhanced by use of multidimensional or of G-invariant persistence.



VM group and Forensic Department (Bologna, Italy)

Profiles of standard police mug shots.

- •Goal: recognition in disguise
- •The space: rightmost contour C
- •Filtering functions:
 - Distances from six relevant points of profile
 - Angle between vertical and segment from rightmost point
 - Curvature

•One filtering function is defined on *CxC* and is the diameter of each pair of points






VM group (Bologna, Italy)

Letters written on a tablet (i.e. with time information).

- •The space: time interval
- •Filtering functions, computed in plane-time:
 - Distance of points from letter axis
 - Speed
 - Curvature
 - Torsion
 - (plane projection) distance from centre of mass



- Goals
 - Letter recognition
 - Writer recognition
- Fuzzy characteristic functions obtained from normalized inverse of distance
- Arithmetic average of characteristic functions

M. Ferri, S. Gallina, E. Porcellini, M. Serena, *On-line character and writer recognition by size functions and fuzzy logic*, Proc. ACCV '95, Dec. 5-8, Singapore, vol. 3 (1995), 622-626.



(Distance from axis)



Monograms (no time information).

- •Goal: personal identification
- •Space1: outline of monogram
- •Filtering function:
 - Distance from centre of mass
- •Space2: a segment (horizontal)
- •Filtering functions:
 - Number of black pixels along segments (3 directions)
 - Number of pixel-pixel black-white jumps (3 directions)















- Fuzzy characteristic functions obtained from normalized inverse of distance
- Weighted average of characteristic functions

Live demo performed at ACCV'98.

M. Ferri, P. Frosini, A. Lovato, C. Zambelli, *Point selection: A new comparison scheme for size functions (With an application to monogram recognition)*, Proc. ACCV'98, Hong Kong 8-10 Jan. 1998, Springer LNCS 1351 vol. 1 (1998), 329-337.



A. Verri (Genova, Italy)



C. Uras, A. Verri, *On the Recognition of the Alphabet of the Sign Language through Size Functions*, Proc. XII Int. Conf. IAPR, Jerusalem (1994), 334 – 338.



- Task: recognition. Signs performed with glove on uniform background.
- The space: horizontal baseline segment
- Filtering functions: for each point, maximum distance of a contour point within a strip of fixed width, with 24 different strip orientations.



Live demo performed at the 12th IAPR Conference.



S. Wang (Sherbrooke, Canada)

- The space: part of contour
- Filtering functions: distance from centre of mass if the point is above the minor (resp. major) axis of inertia.



M.Handouyahia, D. Ziou, S. Wang, *Sign Language Recognition using Moment-Based Size Functions*, Proc of the Int'l Conf. on vision interface Kerkyra CRC Press (1999), 210–216.

M

D. Kelly (Maynooth, Ireland)

- The space: contour
- Filtering functions: distance from four lines.



Size Function $\ell_{in\theta}$ generated

D. Kelly, J. McDonald, T. Lysaght, Ch. Markham, *Analysis of Sign Language Gestures Using Size Functions and Principal Component Analysis*, Proc. IMVIP2008, Portrush, Northern Ireland (2008), 31-36.



S. Banderjee (Kolkata, India)

Sequences of satellite images of cloud systems.

- Goal: evaluate risk and intensity of forming hurricane
- The space: time interval
- Filtering functions: two characteristic measures of cyclones
 - Central Feature portion
 - Outer Banding Feature.

S. Banderjee, *Size Functions in the Study of the Evolution of Cyclones*, Int. J. Meteorology 36(358) (2011), 39-46.

Applications - Tropical cyclones







Applications - Lithography in microelectronics

A. Micheletti (Milano, Italy)

Goals:

- introducing a suitable distance to compare the shapes of the impressed structures;
- specifying confidence regions for the geometries impressed using standard process parameters;
- testing the effects of changing process ;
- looking for the most critical points (if any).

A. Micheletti, F. Terragni, M. Vasconi, *Statistical Aspects of Size Functions for the Description of Random Shapes: Applications to Problems of Lithography in Microelectronics*. In: Progress in Industrial Mathematics at ECMI 2006, Mathematics in Industry 12 (2008), 123-134.





Filtering function: distance from centre of mass

Applications - 3D models



3D meshes

- Goal: distinguish shapes by reciprocal distances
- The space: skeletal graph, in which each vertex corresponds to a region of the mesh.



S. Biasotti, D. Giorgi, M. Spagnuolo and B. Falcidieno. *Size functions for comparing 3D models*, Pattern Recognition 41 (2008), 2855-2873.

Applications - 3D models



- Filtering functions (defined on vertices):
 - area of the corresponding region R
 - minimum, maximum and average distance of the barycenter
 C of the triangles in R from the vertices of the region
 - lenghts of the lower and upper boundary components of R
 - sums of "pseudocone areas" relative to the boundary components.



Applications - 3D models



Two 3D models, their skeleton graphs, the corresponding PBN's relative to the filtering function "average distance".



Applications - Echocardiography

M

C. Lamberti (Bologna, Italy)

Time sequences of 3D echocardiograms of left ventricle.



Applications - Echocardiography

- Goal: recognition of dilated ventricle
- The space: time interval
- Filtering functions:
 - Major-minor axis ratio
 - Form factor
 - Elongation
 - Velocity of centre of mass
- Classification: matching distance, nearest neighbour, farthest neighbour, distance from central representatives.

Applications - Echocardiography



Receiver Operating Characteristic (ROC) curve: It plots Sensitivity vs. (1-Specificity)

Filtering function: Form factor





ROC curve with classification by Farthest neighbour

ROC curve with classification by Distance from central representative

M

VM group (Bologna, Italy)

- Goal: classification (up to confusion eosinophileneutrophile)
- The space: boundary of starlike hull of cell
- Filtering functions: along radii from centre of mass
 - Sum of grey tones
 - Max variation
 - Sum of pixel-pixel variations (in absolute value)
- Classification: distance from average function

M. Ferri., S. Lombardini ,C. Pallotti, *Leukocyte classification by size functions*, Proc. 2nd IEEE Workshop on Applications of Computer Vision, Sarasota, 1994 Dec. 5-7 (1994), 223-229





Neutrophile granulocyte



Eosinophile granulocyte



Basophile granulocyte



Lymphocyte



Monocyte







(Sum of grey tones)



Non-basophile granulocyte



Basophile granulocyte



Lymphocyte



Monocyte



A. Micheletti (Milano, Italy) and G. Landini (Birmingham, UK)

Nuclear profiles from images of cells of histological samples of "normal" or "tumor" zones of the epithelial tissues of the mouth.

- Goal: diagnosis
- Space: nucleus contour
- Filtering function: distance from centre of mass.

A. Micheletti, G. Landini, *Size Functions Applied to the Statistical Shape Analysis and Classification of Tumor Cells*. In: Progress in Industrial Mathematics at ECMI 2006, Mathematics in Industry 12 (2008), 538-542.

Applications - Oncology: mouth cells B = centre ofmass of the white edge detection pixels .B 1.2⁺ 11 $\varphi = dist.$ from 0.9 y the centre 0.0 Algorithm of mass (discrete size 0.7 functions) 0.6 P, 0.8 0.7 8.0 0.9 1.1 12 х

Applications - Oncology: mouth cells





Applications - Oncology: mouth cells



Probability density functions of cornerpoints

Applications - Oncology: mouth cells



Classification with cornerpoints	% unclassified	% of correct recognition
Tumor cells	7.9%	70.12%
Normal cells	6.4%	86.56%

Applications - Oncology: melanocytic lesions

VM group (Bologna, Italy) and I. Stanganelli (Ravenna, Italy)

Melanocytic lesion images acquired under polarized light and mild magnification.

- Goal: distinction between naevus and melanoma
- Problems:
 - No template for either class
 - Various diagnostic criteria
 - Morphological analysis not always sufficient
 - Processing speed compatible with medical consulting room environment

M. Ferri, I. Stanganelli, *Size functions for the morphological analysis of melanocytic lesions,* Int. J. Biomed. Imaging 2010 (2010), Article ID 621357, doi:10.1155/2010/621357

Applications - Oncology: melanocytic lesions

We concentrated on the search for asymmetries of

- boundary shape
- color distribution
- pattern distribution.

This is performed for each lesion as follows.

We take a bundle of 45 lines through the center of mass, and for each we compare the two halves of the lesion, separated by the line.

Instead of making a *geometric* comparison (e.g. by superimposition) we performed a *qualitative* comparison by computing the distance of 0-PBN's of the two halves.

The chosen measuring functions are:

- distance from the splitting line
- sum of luminance along segments
- sum of color variations along segments.

Then, the distances from all splittings are gathered into a function (called *A-curve*) of which some classical invariants (one for each A-curve) are computed.



An image and one of its splittings

The A-curve of the image (filt. fct.: luminance)

From this curve the software extracts *min, max, average, min* plus the value at 90° from *min, integral, first moment, variation, min derivative, max derivative, integral of absolute value of derivative, variation of absolute value of derivative.*

A Support Vector Machine with a 3rd order kernel is fed with these numbers, computed for each filtering function.
Applications - Oncology: melanocytic lesions

The vectors also contain three more parameters: area, perimeter, and a bumpiness measure coming from the 0-PBN's of the whole lesion, with distance from center of mass as the filtering function.



Applications - Oncology: melanocytic lesions V

Experimentation:

The data set contains 50 melanomas and 927 naevi.

Receiver Operating Characteristic (ROC) curve



We are presently developing - with the Romagna Tumor Research Institute and CaMi srl - a retrieval system for melanocytic lesions, as a support for the dermatologist and the general practitioner.

To this goal we are working with new filtering functions and with advanced retrieval methods.



G. Carlsson (Stanford, USA)

CT scans of hepatic cells

- Main goal: classification of lesions
- Secondary goal: comparison of 1D and 2D persistence
- Space: the image rectangle
- Filtering function: the pair (grey tone, distance from cell boundary)

A. Adcock, D. Rubin, G. Carlsson, *Classification of hepatic lesions using the matching metric,* Comput. Vis. Image Und. 121 (2014), 36-42.

76/80

Applications - Hepatic cells

M

- Lesions:
 - cysts
 - metastases
 - hemangiomas
 - hepatocellular carcinomas (HCC)
 - focal nodules
 - abscesses
 - neuroendocrine neoplasms (NeN)
 - a laceration
 - a fat deposit.

Applications - Hepatic cells





(a) Cyst



 $(d) \ \mathrm{HCC}$



(g) NeN



(b) Metastasis



(e) Focal Nodule



(h) Laceration



(c) Hemangioma



(f) Abscess



(i) Fat Deposit

Applications - Hepatic cells



Different classification experiments have been performed:

- All lesion types (Full)
- Hepatocellular carcinomas, hemangiomas, cysts, metastases (HcHeCM)
- hemangiomas, cysts, and metastases (HeCM)
- cysts and metastases (CM)

SVM Classification accuracies for 1D and 2D filtrations.

Filtration	Full (%)	HcHeCM (%)	HeCM (%)	CM (%)
1D (intensity)	55.30	59.66	63.89	80.00
2D	66.67	72.27	80.56	85.56



massimo.ferri@unibo.it patrizio.frosini@unibo.it clandi@unimore.it

http://vis.dm.unibo.it

Thank you for your attention!

