Repositório ISCTE-IUL

Deposited in *Repositório ISCTE-IUL*:
2023-03-17

Deposited version:
Accepted Version

Peer-review status of attached file:
Peer-reviewed

Citation for published item:

Lopes, J. V. (2018). A data mining based methodology for the multidimensional study of public open spaces. In David Leite Viana, Franklim Morais and Jorge Vieira Vaz (Ed.), Formal methods in architecture and urbanism. Cambridge: Cambridge Scholars Publishing.

Further information on publisher's website:
http://www.cambridgescholars.com/formal-methods-in-architecture-and-urbanism

Publisher's copyright statement:
This is the peer reviewed version of the following article: Lopes, J. V. (2018). A data mining based methodology for the multidimensional study of public open spaces. In David Leite Viana, Franklim Morais and Jorge Vieira Vaz (Ed.), Formal methods in architecture and urbanism. Cambridge: Cambridge Scholars Publishing.. This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

Use policy

Creative Commons CC BY 4.0
The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in the Repository
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Serviços de Informação e Documentação, Instituto Universitário de Lisboa (ISCTE-IUL)
Av. das Forças Armadas, Edifício II, 1649-026 Lisboa Portugal
Phone: +(351) 217 903 024 | e-mail: administrador.repositorio@iscte-iul.pt
https://repositorio.iscte-iul.pt

A data mining based methodology for the multidimensional study of public open spaces

João V. LOPES

ISTAR, Instituto Universitário de Lisboa (ISCTE-IUL)
jvsls@iscte.pt

Abstract. Public open spaces can only be apprehended from multiple simultaneous perspectives. Urban morphology traditional descriptive methods have recognized limitations in relating the polymorphic and polysemantic nature of these spaces' attributes, derived from the different standpoints on their formal, historical and geographic idiosyncrasies. Identities and similarities may be disclosed by multivariate statistical analysis and data mining techniques by studying the relations between formal and intangible spatial properties in a multidimensional space. In an ongoing PhD research project we outline a method for the synchronic analysis and classification of the public open spaces, departing from a *corpus* of 126 Portuguese urban squares, whose analysis is intended to interactively (re)define it. Part of the work done so far is presented: (i) firming the concepts, criteria and attributes to extract; (ii) adaptation and/or creation of new analytical methods and tools; and (iii) research on multivariate analysis, data mining and data visualization techniques.

Keywords. Urban morphology, Urban design, Public open space, Parametric-algorithmic design, Data mining

1. INTRODUCTION

The identity of the public open space does not manifest itself only in its form and history. Nondiscursive attributes, in Hillier's words [1], cognitive and perceptive attributes not representable by traditional methods, as well as use, appropriation and environmental idiosyncrasies are part of it. Only the simultaneous consideration of such a heterogeneous set of characteristics may reveal its *identity*, as well as unsuspected familiarities. Given human limitation to apprehend spaces or problems with more than 3 or 4 dimensions [2], the combination of several theories and spatial and site analysis methods with data mining helps to overcome this inability. It allows the creation of new bottom-up knowledge from the very structure of the data itself and may help to promote the integration of urban morphology disciplines and design. The creation of more complex typological descriptions is essential in capturing desired qualities embedded in the urban structure, and so is its classification.

We present a methodology for the analysis and classification of public open space, focusing on the formal square, which is accepted as an individualizable urban element. Collecting contributions from the disciplines of urban morphology and site analysis, it aims to synchronically and multi-dimensionally characterize and classify these urban spaces. It resorts to multivariate statistical analysis and inductive patterns search techniques in large data sets by data mining.

We choose a corpus comprising a set of 126 Portuguese squares. This specific corpus [3] (Figure 1) has been selected because it characterizes a significant sample in the context of Portugal mainland, the information is comprehensive and available in digital format, the representation is systematic, consistent and updated; it offers a careful historical and architectural analysis in the perspective of the Italian school of urban morphology; and it has been a basis to other investigations. There are no spatial or formal data quantified, and that will be our first task.

Figure 1. Typical graphical representations of the *corpus*. Example: Giraldo square, Évora, in *Squares in Portugal* [3].

The overall research objectives are: (i) the design of an original multidimensional inductive method for the characterization and classification of public open spaces, able to gather descriptive and structural approaches in a single analytical process; (ii) to relate morphological features with measurable quality and performance attributes of urban spaces by datamining: and (iii) to conceive a urban design approach, based on concepts, workflows and digital tools structured in such a way to take advantage of the method proposed. This paper focus on the first objective and continues discussing: (2.) background review; (3.) methods and tools adopted, initial assumptions, definitions, and workflow; (4.) analytical algorithms for more complex attributes' extraction; and (5.) conclusion and future work.

2. BACKGROUND

Studies of urban morphology seldom carry out detailed analyses, and classification proposals, of urban environments and elements from different perspectives, at various scales and complexity levels. The latest of the disciplines in this field, space syntax [4], whose theories, models and tools will support us greatly, has within its best-known studies and applications the urban square (notably Trafalgar Square in London). Campos [5] and Campos and Golka [6] investigate the relationship between patterns of use, network configuration and visibility by studying the penetration of axial lines, the effect of isovists [7] and visual fields, through visual graph analysis (VGA) [8], on the space of London squares. Cutini [9] studies Tuscan historic squares (*piazzas*), focusing on the relationship between centrality, configuration and visual analysis, extending space syntax measures by creating a new compound VGA index that depicts the hierarchy of convex spaces in settlements.

The classic methods of urban morphologic analysis, usually limited to the analysis of single or pairs of variables, so as to respond to human cognitive and perceptive limitations, restrict the simultaneous expression of features that give spaces their *uniqueness* and transform them into *places*. Gil et al. [10] compile the shortcomings of traditional typomorfological approaches: their time-consuming methods, restraining the amount of examples and dimensions, their relative opacity, subjectivity and dependence on the analyst's abilities and geographical/cultural contexts, questioning their reproducibility and generality. Some recent structural approaches try to escape these limitations by relating attributes and classifying examples through clustering in multivariate tri or quadrangular graphs [11].

The identified deficiencies can be handled by the use of new computational methods that allow for multidimensional analysis and typological classifications based on multivariate statistical models, exploratory data analysis (EDA) and machine learning. Most of these methods are currently included in the set of techniques associated with data mining.

Within urban morphological studies, they support analyses at different scales: from Laskari et al. [12] study on *urban identity through quantifiable attributes* on blocks' shape at district level, to street patterns in metropolitan areas [13]. Gil et al. [10] in an unsupervised classification of the urban fabric of two neighbourhoods of Lisbon, focusing on street and block elements, mention the possible integration of these techniques in design. Chazar and Beirão [14] point the potential of their expand to the analysis of non-formal spatial qualities, leading to a better understanding of the public open space morphology.

Although multidimensional analysis of the urban void are rarer, Laskari [15] analyses the blocks' residual void space in a neighbourhood of Athens, through a set of 13 properties and by different multidimensional methods. The syntactic analysis of convex spaces is being enriched by the ongoing work of Beirão et al. [16], which introduces a new urban void 3D analytic method and new classification and aggregation logics of its elementary units.

3. METHOD

Methodologically, this research collects concepts and tools from the urban morphology, GIS, algorithmic design/parametric urbanism and data mining fields.

The attributes are extracted through (virtual and real) surveys, and geographical and urban models based on the digital representations of the *corpus*. According to scale and goal we use analytic tools from geography (*QGIS*, www.qgis.org), space syntax (*Depthmap*, UCL) and parametric-algorithmic design (*Rhino/Grasshopper*, McNeel & Assoc.). The advantage of the latter, over other CAD and GIS platforms, in local scale urban modelling is pointed out by Hanzel [17]. Their associative and interactive 3D nature, rule base approach and data integration flexibility, promotes the synchronization between analysis, representation and design. As their processes are explicit, initial criteria and assumptions can be easily updated. For multidimensional analysis we resort to *RapidMiner* (RapidMiner GmbH), a popular data mining software, such as *Grasshopper*, based on *patterns* and a visual programming interface (VPI).

Data mining, at the intersection of artificial intelligence, machine learning, statistics and database systems, finds patterns and rules in large data sets via inductive methods. Its main objectives are prediction (classification of unknown cases and regression) and the discovery of new knowledge (finding unknown patterns in data) [18]. Among the various methods implemented in data mining we intend to start by exploring two of the most established in practice: principal component analysis (PCA) and clustering (*k-means* and *k-medoids* algorithms). The PCA analysis determines a smaller set of (artificial) variables that summarizes the original data with minimized loss of information and capable of revealing unsuspected relationships. Clustering is an unsupervised classification process that assigns objects to groups (clusters), so that the objects of each group are more similar to one another than with the objects of other groups. It aims at discovering natural groups of objects or variables, identifying extreme and prototypical cases and suggesting interesting hypotheses about relationships.

3.1. Analysis scales and boundaries

In order to adapt the nature of each recorded attribute to their characteristic spatial scale and representation, boundaries or spatial aggregation scales are defined. Despite the artificiality of their definition, given the *continuum* nature of the urban void and network, and the problems raised by sensitivity to frontier conditions, five operative boundary categories are proposed (Figure 2):

(1) **Site boundary**: determined by the 2D closed polygon formed by the extension of the facade segments and plots' limits that clearly form the space of the square. This determines factors related to shape, perimeter visibility/connectivity and public/private permeability at the most local scale.

(2) **Neighbourhood boundary**: determined by the streets bordering the adjacent blocks to the previous border within a minimum radius of 250m. It highlights the aggregation of features related to local urban indices, densities and VGA analysis, while ensuring contextual influence.

(3) **Settlement boundary**: determined by the limit of towns (or a minimum radius of 1,5km) and provides information on *centrality*, *intelligibility* and *synergy* features relating local to overall structures.

(4) **Regional boundary**: regional boundaries as defined by political/administrative or historical/natural frontiers.

(5) **National boundary**: the most generic boundary comprising all the examples and defined by the Portuguese mainland frontier. It depicts, along with (4), socio-cultural differences at a regional and national scale.

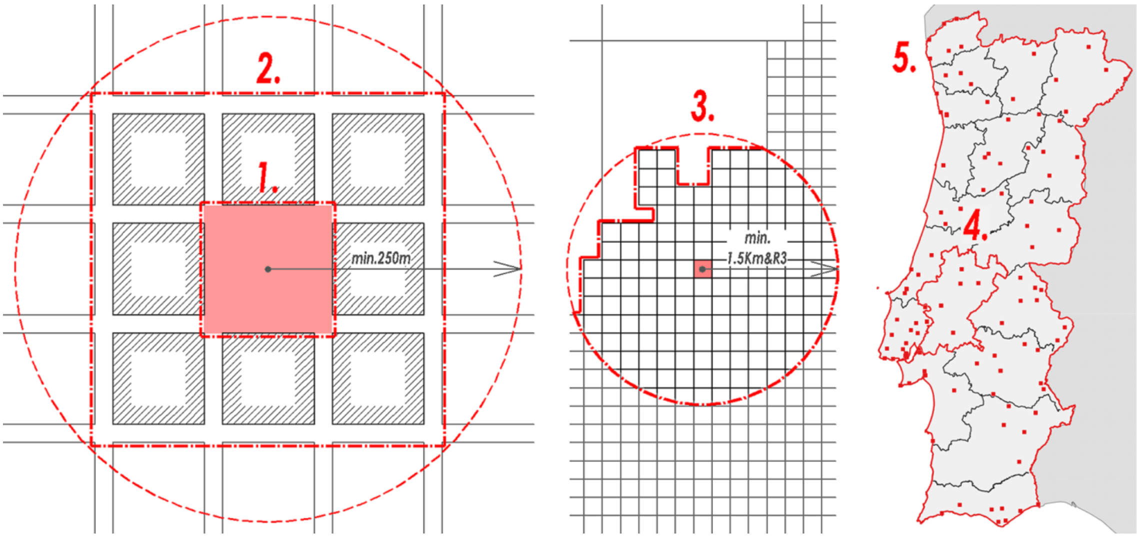

Figure 2. Analysis scales and boundaries: diagrammatic representations.

3.2. Spatial and formal attributes of the squares

The definition of the initial set of attributes to be quantified for each of the squares is founded on review and discussion of related work. We gather generic morphological, syntactic and environmental descriptions, local and global features from various disciplines as heterogeneity of attributes is essential to the proposed method. Its value lies in the ability to reveal non obvious correlations through dimensionality reduction and visualization of big data sets [18].

The attributes were divided into eight thematic groups (Figure 3, Table1), which reflect both the diversity of approaches and scales:

(1) Void shape. Attributes extracted from the two-dimensional representation of the square by its site boundary, analysed as a polygon, in order to extract geometric measures, ratios and shape factors.

(2) Vertical plane and permeability. Attributes related to the three-dimensional expression of the space perimeter and the facades, both their geometry and their behaviour as interface between public and private.

(3) Urban indices and density. Attributes that relate the area of the square to the area of the surrounding blocks and buildings in terms of their built area and footprint. It focuses on density measurements inspired by Spacematrix theory [11].

(4) Visibility and connectivity. Visibility properties, according to: (i) the distribution of connectivity along the square perimeter and the inter-visibility of the facades; (ii) VGA analysis limited by the neighbourhood boundary, aggregating the values inside the square perimeter; and (iii) the visibility from the exterior of the square, through the calculation of external isovists' overlapping area and perimeter.

(5) Urban system. This attribute theme focuses on the features of the squares' embedding urban system. Extraction of local and global syntactic values of the axial lines crossing the square, their lengths and geometric relationships.

(6) Use and appropriation. Attributes which classify buildings adjacent to the square into classes, according to their uses, and register the existence of exceptional buildings and characteristic elements of the urban squares (fountains, bandstands, pillories, kiosks, etc.).

(7) Environment. This group deals with the existence of natural elements and other environmental features (e.g. visible sky area percentage, solar orientation or the maximum topographic slope). These attributes, and the previous ones, are related to *urban quality potential,* which will have to be interpreted in their specific geographical contexts.

(8) Generic labels. These are essentially attributes related to geographical factors and territorial distribution, or some sort of *à priori* classification (labelling), whose correlation with the data can be tested or *machine learned*.

As the analysis progresses, the definition of the attributes may change and be optimized. In order to increase the expressiveness of the data its correlation and type of statistical aggregation shall be attested by EDA analysis and by the early modelling of the data mining algorithms.

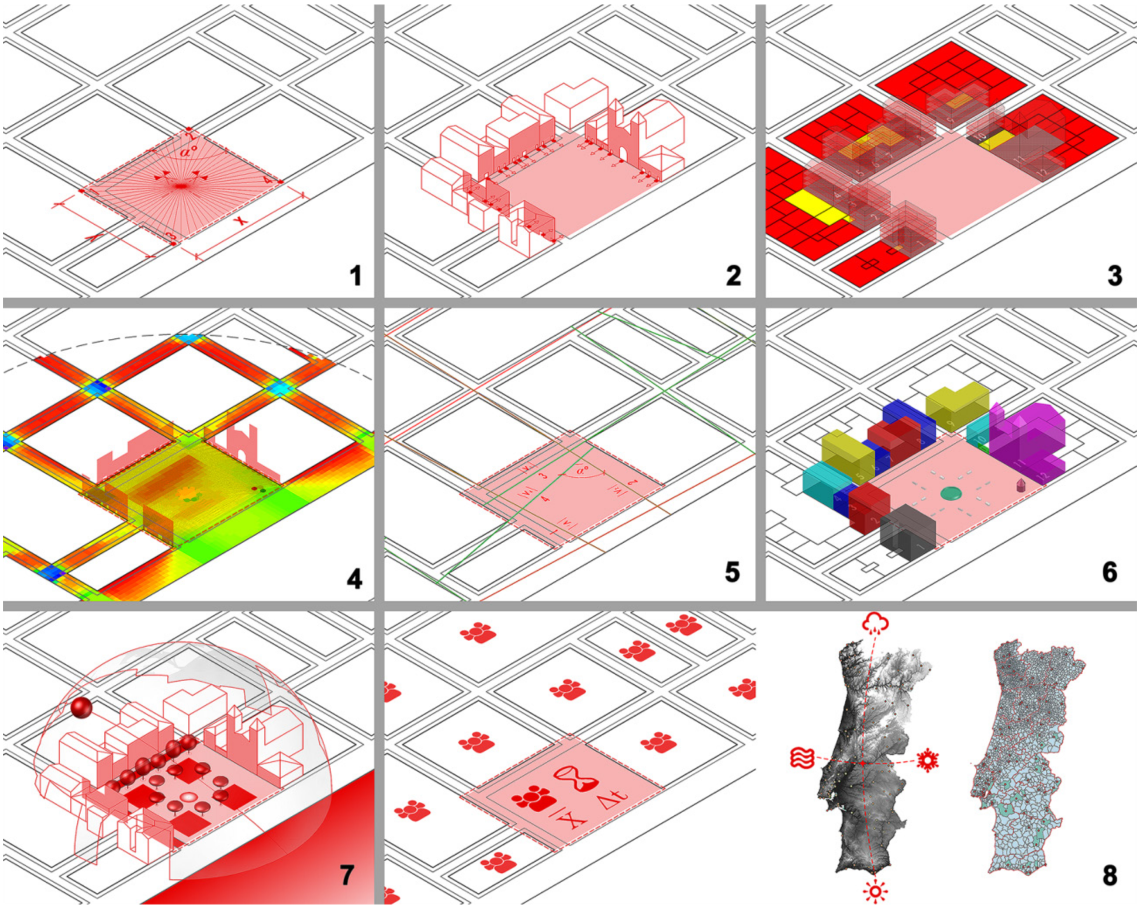

Figure 3. Attributes and attribute groups: diagrammatic representations.

#	Attribute	Code	Boundary	Unit	Data type	Main theory
	(1) VOID SHAPE					
1	**Area**	Vs A	Site	m2	Real	Shape
2	**Area/perimeter ratio**	Vs RP	Site	_	Real	Shape factor
3	**Nr of perimeter vertices**	Vs NV	Site	_	Integer	Shape
4	**Nr of internal islands**	Vs NI	Site	_	Integer	Shape
5	**Length of the longest side (angular deviation > 15°)**	Vs LL	Site	m	Real	Shape
6	**Aspect ratio (elongation)**	Vs AR	Site	_	Real	Shape factor
7	**Circularity ratio (isoperimetric quotient) (spikiness)**	Vs CI	Site	_	Real	Shape factor
8	**Entropy (symmetry)**	Vs EN	Site	_	Real	Shape factor/GIS
9	**Rectangularity ratio (rectangularity)**	Vs RR	Site	_	Real	Shape factor
10	**Perimeter fractal dimension-Hausdorff dim. (roughness)**	Vs FD	Site	_	Real	Shape factor
	(2) VERTICAL PLANE, AND PERMEABILITY					
11	**Perimeter/block frontage ratio**	Vp RB	Site	_	Real	Urban Morphology
12	**Facade area**	Vp FA	Site	m2	Real	Urban Morphology
13	**Maximum height of the facades**	Vp MAH	Site	m	Real	Urban Morphology
14	**Mode of the heights of the facades**	Vp MOH	Site	m	Real	Urban Morphology
15	**Entropy of the heights of the facades**	Vp EN	Site	_	Real	Shape factor
16	**Opening density (opening nr/m frontage)**	Vp OD	Site	_	Real	Space Syntax
	(3) URBAN RATIOS AND DENSITY					
17	**Nr of adjacent blocks (= nr of adjacent streets)**	De NBL	Site	_	Integer	Urban Morphology
18	**Nr of adjacent plots**	De NP	Site	_	Integer	Urban Morphology
19	**Nr of adjacent buildings**	De NBU	Site	_	Integer	Urban Morphology
20	**Urban square area/adjacent blocks area ratio**	De SBL	Neighbour.	_	Real	Spacematrix
21	**Urban square area/adjacent building footprint area ratio**	De SF	Neighbour.	_	Real	Spacematrix
22	**Urban square area/adjacent built area ratio**	De SBU	Neighbour.	_	Real	Spacematrix
	(4) VISIBILITY AND CONECTIVITY					
23	**Min nr of convex spaces**	Vc MC	Site	_	Integer	Space Syntax
24	**Perimeter mean connectivity value (mcv)**	Vc MCV	Site	_	Real	Space Syntax
25	**Vertical stand. deviation perimeter connectivity (v-value)**	Vc VV	Site	_	Real	Space Syntax
26	**Horizontal stand. deviation perimeter connectivity (h-value]**	Vc HV	Site	_	Real	Space Syntax
27	**Mean horizontal value (mhv)**	Vc MHV	Site	_	Real	Space Syntax
28	**Area overlapping isovists from the street (visual exposure)**	Vc VS	Neighbour.	_	Real	Isovist
29	**Urban square area/major isovist (360°) area ratio**	Vc PI	Neighbour.	_	Real	Isovist
30	**Area of the visual integration core (visual integration >90%)**	Vc CO	Neighbour.	_	Real	Space Syntax VGA
31	**Mean visual connectivity (less prone to edge effect)**	Vc COM	Neighbour.		Real	Space Syntax VGA
32	**Mean visual clustering coefficient (less prone to edge effect)**	Vc COE	Neighbour.		Real	Space Syntax VGA
	(5) URBAN SYSTEM					
33	**Mean width of confluent streets**	Us SW	Neighbour.	m	Real	Urban Morph./Design
34	**Sum of the confluent axial lines length**	Us AL	Settlement	m	Real	Space Syntax
35	**Mean angle between confluent axial lines**	Us ANL	Settlement	Degree	Real	Space Syntax
36	**Maximum confluent axial lines global integration**	Us GI	Settlement	_	Real	Space Syntax
37	**Maximum confluent axial lines local integration (radius3)**	Us GI3	Settlement		Real	Space Syntax
38	**Maximum confluent axial lines global choice**	Us GC	Settlement	_	Real	Space Syntax
39	**Maximum confluent axial lines local choice (radius3)**	Us GC3	Settlement		Real	Space Syntax
40	**Mean intelligibility (local connectivity/global integration)**	Us IN	Settlement	_	Real	Space Syntax
41	**Mean synergy (local (r3) integration/global integration)**	Us SY	Settlement	_	Real	Space Syntax
	(6) USE AND APPROPRIATION					
42	**Adjacent building use (Church/historic building; Public building/Urban equipment; Comerce/Services; Residential)**	Uu BU	Neighbour.	_	Polynom.	Land use
43	**Urban elements (Bandstand; Pillory; Benches; Kiosk; Urban art; Fountain)**	Uu EL	Site	_	Polynom.	Urban design/Usage
	(7) ENVIRONMENT					
44	**Natural elements (Water; Green space; Trees)**	En NE	Site	_	Polynom.	Site analysis/Landscape
45	**Mean sky component (sky hemisphere visiblity %)**	En SK	Site	%	Real	Site analysis/Solar
46	**Area in constant shadow**	En SH	Site	_	Real	Site analysis/Solar
47	**Global solar orientation (wheghted segment contribution)**	En SO	Site	_	Real	Site analysis/Solar
48	**Maximum topographic slope**	En SL	Site	%	Real	Site analysis/Landscape
	(8) GENERIC LABELS					
49	**Toponymic type (square, churchyard, circus, etc)**	Ge TO	National	_	Polynom.	Urban Morphology
50	**Geografic/climatological region**	Ge GE	Regional	_	Polynom.	Geography
51	**Latitude (north/south differentiation)**	Ge LA	National	Degree	Real	Geography
52	**Longitude (coast/inland differentiation in Portugal)**	Ge LO	National	Degree	Real	Geography
53	**Site elevation**	Ge EL	Regional	m	Real	Geography
54	**Date (first historical reference: century)**	Ge DA	_	_	Date	Urban Morph./History
55	**Population density at square civil parishes' level**	Ge PD	Neighbour.	_	Real	Demography

Table 1. Attributes metadata.

3.3. Workflow

The proposed workflow (Figure 4) includes the gathering and preparation of information, building the models and analytical algorithms, the attributes extraction, tables' construction and the data storage in a *PostgreSQL* relational database, a scalable central repertoire aiming at streamline the workflow.

The data mining and visualization of statistical information will be made with *RapidMiner*, and their spatial mapping in *Rhino/Grashopper* and *QGIS*. Data visualization will be essential to the critical interpretation by field experts from which may result the redefinition of the initial assumptions and the process itself.

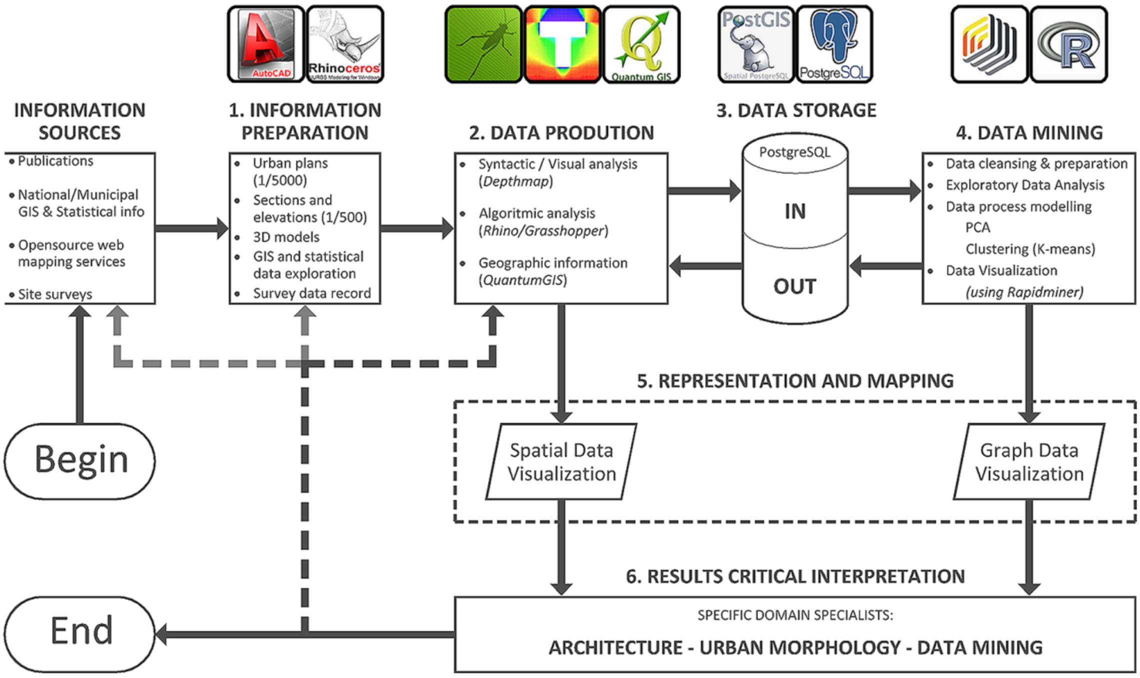

Figure 4. Workflow diagram and tools.

4. ANALYTICAL ALGORITHMS

Some of the squares' simple attributes (lengths, areas, counts) are extracted directly from the 2D and 3D models without further processing, and composed attributes (ratios, densities, shape factors) are calculated within the database. We briefly describe some of the developed algorithms requiring more complex and/or specific modelling (Figure 5):

4.1. Fractal dimension (Hausdorff dimension by a *coast of Britain-like* process) (Figure 5-1)

The Hausdorff dimension, also called *covering dimension*, is one of the various mathematical definitions of fractal dimension. This is a statistical index of the roughness of a pattern and its definition is based on the basic principle of the change in detail with scaling. It corresponds to the limit of the slope for the regression line of a *log vs log* graph of a function *size vs scale* [19]. In our model the pattern is the square perimeter; size is the number of equal segments (rulers) forming a polyline whose vertices slide on the perimeter line; and scale is the length of these rulers. This last factor has to be carefully considered as detail is finite in the graphical representations. The approximation to the Hausdorff dimension is an iterative process in which rulers double their length in each iteration, on a scale between 0.5m and 64.0m considered appropriate to the squares' plan scale (1/1000). This is totally done in *Grasshopper* including the calculation of the regression line and its slope: the estimated fractal dimension.

This method is implemented alternatively to the box-counting method as it is computationally less expensive and insensitive to the shape spatial orientation. However in cases of extreme convulsion it is sensitive to the direction of the rulers. In general, for non-mathematic fractals the Hausdorff and box-counting dimensions coincide (https://en.wikipedia.org/wiki/Hausdorff_dimension).

4.2. Perimeter Connectivity (Figure 5-2)

Based on work by Psarra and Grajewski [20] and Laskari et al [12], we implemented in *Grasshopper* an algorithm that allows the characterization of shapes by analysing the connectivity of their perimeters. A set of measures are defined that allow the description of shapes as *patterns of stability and differentiation,* *rhythm and repetition* and their understanding *beyond the conventional characterization of its geometric order*, quoting Psarra and Grajewski. The measures, based on visibility values between segments discretizing the shape perimeter, as plotted in a graph, are: total mean connectivity value (*mcv*); vertical standard deviation (*v-value*), the standard deviation from mcv; horizontal standard deviation (*h-value*), the standard deviation of the distance between subsequent points of mcv value; to which Laskari adds: mean horizontal value (*mhv*), the average value of that distance.

Though these measures are concerned with shape description, we intend to extend the use of this model to assess the intervisibility between the space defining facades. It is expected a high degree of correlation with some squares' shape factors. The models created have a typical resolution of 1.5 m and calculations are made integrally in the *Grasshopper*.

4.3. Shape diameters, radii and diagonals (Figure 5-3)

The outlined algorithm is capable of finding the major segment inscribed within the perimeter of the square (major diameter), and the major and minor ones perpendicular to it (minor diameters), as well as the lowest absolute diagonal originating from a vertex. Deals with convex and concave spaces and is sensitive to the existence of *islands*. Contrary to axial lines of space syntax and the concept of the diagonal of a polygon in discrete geometry [22], this model allows the diagonals to graze the sides of the polygon.

The algorithm also calculates the longest axis, passing through the centroid of the square, and its perpendicular, as well the radii centroid-vertices, and variable radial sampling from a selected point. These values allow the determination of attributes related to shape factors and moments whose calculations are mainly taken from the fields of GIS and image analysis.

4.4. Internal and external isovists. Isovist fields (Figure 5-4)

Strictly geometrical 2D isovists (without radial sampling) are created. Distinction between isovist polygon segments representing obstacles and occlusion lines is made, and the calculation of their geometric based measures, such as occlusion and drift [7], can be made rigorously. The use of *Grasshopper* allows an accurate and interactive control of isovists location, resolution and extension. Isovist fields can be created step-by-step for points in a grid, using the software animated sliders and the data recording component, illustrating the possibility of performing a simple *foreach-loop* without resorting to advanced scripting. However, it is computationally expensive and not presented as a substitute for isovists by radial sampling or their calculation by specialized programs like *Depthmap* or *Syntax2D* (University of Michigan).

4.5. Facades, building heights and solar orientation (Figure 5-5)

The area of the facades bounding the space is extracted from the existing conventional 3D CAD models of the *corpus* through automatic selection of their surfaces, which are unfolded for the purpose of visualization and mapping. The facades' height is characterized by its maximum, mode and entropy values of its distribution, respectively accounting for the presence of landmarks, a predominant height and its complexity. An attribute characterizing solar orientation is given by the weighted mean value of orientation (gradient from south=1 to north=0) for the square boundary segments, where the normalized segment lengths are the weights. Values closer to 1 indicate predominant good solar orientation for the typical latitudes of Portugal, expressed as a potential since it considers the entire perimeter even if not built.

4.6. Sky factor and 3D solid isovists (Figure 5-6)

This algorithm calculates the percentage of the visible area of the sky, from a point at eye level, without discretizing the sky vault. The process entails the creation of *3D solid isovists*, by the extrusion of the visible facades´ surfaces to a *vantage* point, and their subtraction from a hemispherical surface representing the sky. The measures are expressed in percentage of the area of the initial hemisphere which represents an unobstructed horizon.

Intercepting the resulting surface with another one representing the solar movement for a given period and the local latitude, insolation can be approximately calculated. This algorithm records the average percentage of the squares sky's visible area and the area permanently in shadow, using a grid of points (sensors) spaced 1.5 m. The calculation is done step-by-step for each of the points as described above in the construction of isovists.

Other indicators can be drawn from this model relating properties of the 3D isovists or the sky map shape, such as the *Gibsonian* inspired spherical metrics for the urban open space created by Teller [21], and exemplified by that author for a set of European historical squares.

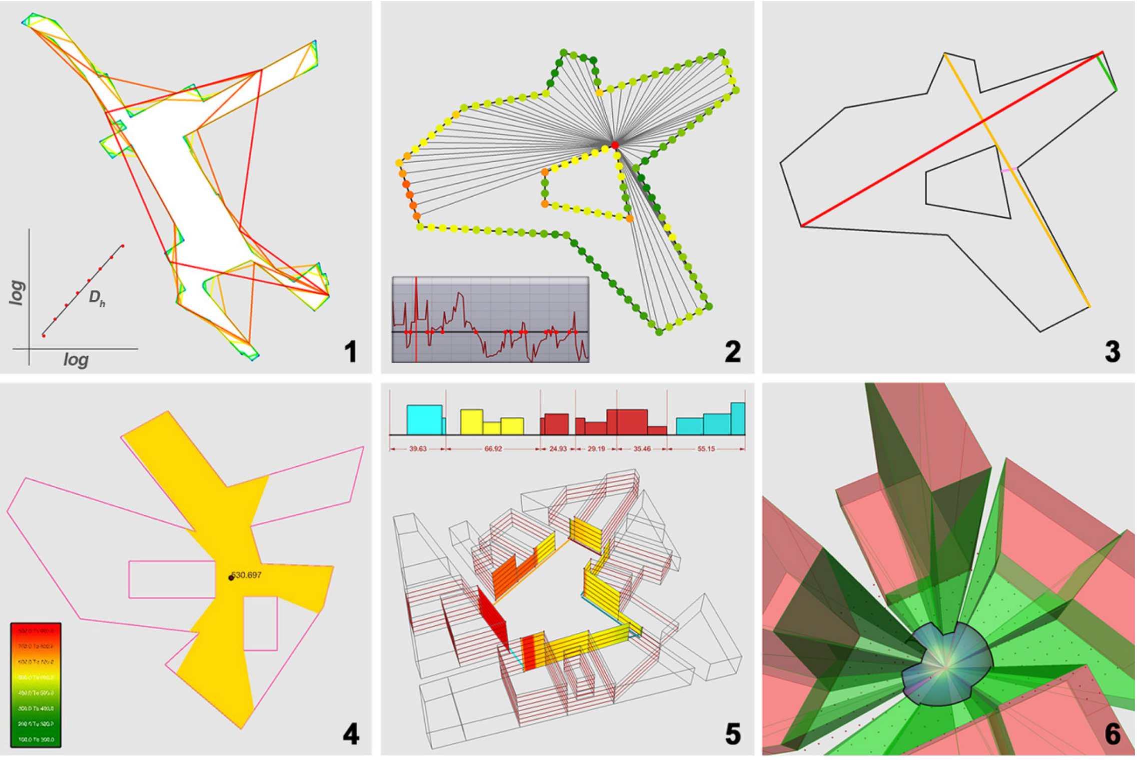

Figure 5. Analytical algorithms.

Other *Grasshopper* definitions where devised dealing with the creation and maintenance of *PostgreSQL* databases from inside *Grasshopper's* modelling environment, using Nathan Miller's *Slingshot!* component. This allows the interoperability between CAD, GIS and data mining platforms, without resorting to intermediate file formats like *shapefiles*. Spatial analysis software *Depthmap* could only be indirectly integrated in this database workflow by parsing *csv* data files. Using this method, and by means of string manipulation, a *Grasshopper* definition allows *Depthmap's* axial and VGA analyses reconstruction, and data and geometry simultaneous manipulation, within the algorithmic/parametric environment.

5. CONCLUSION AND FUTURE WORK

We have presented a scheme for the construction of a multidimensional method specifically designed for uniquely identify and classify public open spaces, through the expressiveness of the data itself, and having the formal urban square at its starting point. This method entails:

- A comprehensive set of *perspectives* and scales of analysis;
- Using the entire *spectrum* of spatial dimensions;
- The synthesis of physical, spatial and immaterial attributes in a single descriptive vector;
- The determination of the correlations between attributes and the minimum *individualizing* attribute set (correlation analysis; dimensionality reduction);
- The bottom-up classification of the examples by natural clustering of the data.

Despite the research project being in an early stage, there is confidence in the validity of the approach, as some general points seem clear:

- Human limitations in handling, visualizing and discovering patterns in multiple dimensional data sets, is greatly assisted, or only possible, by data mining.
- The potential of data mining in urban morphology and design is mostly unexplored.
- Multidimensional analysis in urban morphology research is recent but its advantages are well documented.
- Heterogeneous knowledge can be associated and used to classify public open space from simultaneous perspectives, and correlate it with quality attributes.

Problems encountered relate to limitations in the cartographic sources, the difficulty in completing information for some remote settlements and the poor scalability of the algorithms created. These are intended as learning and mock-up tools for testing models that would have to be implemented through high performance programming languages like *Python* or *C*.

In future work the potential of the method presented points in four directions. First, to extend the example set to other types of contemporary public open spaces that challenge the formal concept of urban square. Second, to deepen the potential of data mining in urban analysis, exploring other dimensionality reduction and learning (supervised and unsupervised) techniques. Third, to ascertain the receptivity of data mining in the urban morphology community, submitting the results to criticism. And fourth, probably the most significant assignment, to correlate spatial and formal attributes with urban quality or real performance measures through data mining techniques, in order to explore the combination of this analytical method with generative design processes.

Acknowledgments

The author thanks his PhD advisors Prof.s Alexandra Paio, José Beirão e Eliana Pinho, the collaboration of FORMA URBIS Lab (FAUL), and Prof.s Dias Coelho, Sérgio Fernandes and Luís Nunes. Research supported by FCT grant SFRH/BD/95148/2013.

References

[1] B. Hillier, *Space is the machine: a configurational theory of architecture*. Cambridge: Cambridge University Press, 1996.
[2] G. Halford, R. Baker, J. McCredden, and J. Bain, “How many variables can humans process?,” *Psychol. Sci.*, vol.16, no.1, pp.70–76, 2005.
[3] C. Coelho, *A praça em Portugal. Inventário de espaço público*. Lisbon: DGOTDU, 2008.
[4] B. Hillier and J. Hanson, *The Social Logic of Space*. Cambridge: Cambridge University Press, 1984.
[5] A. Campos, “Strategic spaces: patterns of use in public squares of the city of London,” in *Space Syntax-1st International Symposium Proceedings*, 1997, pp.26.1–26.11.
[6] A. Campos and T. Golka, “Public Spaces Revised: of the relationship between Patterns of Stationary Activities and Visual fields,” in *Proceedings, 5th International Space Syntax Symposium*, 2005.
[7] M. Benedikt, “To take hold of space: isovists and isovist fields,” *Environ. Plan. B*, vol.6, no.1, pp.47–65, 1979.
[8] A. Turner, M. Doxa, D. O’Sullivan, and A. Penn, “From isovists to visibility graphs: A methodology for the analysis of architectural space,” *Environ. Plan. B*,vol.28, no.1, pp.103–121, 2001.
[9] V. Cutini, “Lines and squares: Towards a configurational approach to the morphology of open spaces,” in *Proceedings 4th International Space Syntax Symposium, London*, 2003.
[10] J. Gil, J. Beirão, N. Montenegro, and J. Duarte, “On the discovery of urban typologies: Data mining the many dimensions of urban form,” *Urban Morphol.*, vol.16, no.1, pp.27–40, 2012.
[11] M. Berghauser Pont and P. Haupt, *Spacematrix: Space, Density and Urban Form*. Rotterdam: NAi Pub., 2010.
[12] A. Laskari, S. Hanna, and C. Derix, “Urban Identity Through Quantifiable Spatial Attributes,” *Des. Comput. Cogn.*, vol.2008 SRC , pp.615–634, 2008.
[13] M. Serra, J. Gil, and P. Pinho, “Unsupervised Classification of Evolving Metropolitan Street Patterns,” in *9th International Space Syntax Symposium*, 2013.
[14] A. Chaszar and J. Beirão, “Feature Recognition and Clustering for Urban Modelling-Exploration and Analysis in GIS and CAD,” in *Open Systems: Proceedings of the 18th Internationa CAADRIA Conference*, 2013, pp.601–610.
[15] A. Laskari, “Multidimensional Comparative Analysis for the Classification of Residual Urban Voids,” in *Proceedings of the 32nd eCAADe Conference*, 2014, pp.283–292.
[16] J. Beirão, A. Chaszar, and L. Cavic, “Convex-and Solid-Void Models for Analysis and Classification of Public Spaces,” in *Proceedings of the 19th CAADRIA International Conference*, 2014, pp.253–262.
[17] M. Hanzl, “Modelling of Public Spaces,” in *Proceedings of the 31st eCAADe Conference*, 2013, vol.1, pp.319–327.
[18] I. Witten and E. Frank, *Data mining : practical machine learning tools and techniques*. Boston: Morgan Kaufman, 2005.
[19] P. Frankhauser, *La Fractalité des structures urbaines*. Paris: Anthropos, 1994.
[20] S. Psarra and T. Grajewski, “Describing Shape and Shape Complexity Using Local Properties,” in *3th International Space Syntax Symposium*, 2001.
[21] J. Teller, “A Spherical Metric for the Field-Oriented Analysis of Complex Urban Open Spaces,” *Environ. Plan. B*, vol.30, no.3 , pp.339–356, Jun.2003.